CHALLENGES FOR DARK ENERGY SCIENCE: COLOR GRADIENTS AND BLENDED OBJECTS

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF PHYSICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Sowmya Kamath
March 2020

This dissertation is online at: http://purl.stanford.edu/rx238kt9962

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Patricia Burchat, Primary Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Steven Allen

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Sarah Church

Approved for the Stanford University Committee on Graduate Studies.

Stacey F. Bent, Vice Provost for Graduate Education

This signature page was generated electronically upon submission of this dissertation in electronic format. An original signed hard copy of the signature page is on file in University Archives.

Abstract

Some of the most exciting scientific challenges at present are in understanding the nature of dark matter and dark energy. There are many different cosmic probes of dark matter and dark energy physics, of which weak gravitational lensing is the most direct probe of the mass distribution in the Universe. The Legacy Survey of Space and Time (LSST) , which will encompass billions of galaxies imaged with the Vera Rubin Observatory, will dramatically improve the statistical power of weak lensing observations. This increased statistical sensitivity means that potential systematic biases must be carefully identified, quantified, and minimized. This thesis is dedicated to addressing two such systematic biases for weak lensing with the LSST: galaxy color gradients and blending.

As well as being sheared by gravitational lensing, the observed shapes of galaxies are distorted by the atmosphere, the telescope optics and sensor effects – the effects of all of which is captured in the point spread function (PSF). If these effects are wavelength dependent and if the galaxy has a spectral energy distribution (SED) that varies across its profile – referred to as "color gradients" – then the distortions will be different at different points on the galaxy. PSF corrections that are applied assuming a constant SED can lead to systematic bias in the shape measurements. In the first part of this thesis, we predict the bias due to color gradients using simulations of parametric galaxies and then extend the study to realistic galaxy images. We quantify the expected impact on weak lensing measurements.

Due to the increased depth to which LSST will observe, as well as the significant smearing of the image due to the atmospheric PSF, a significant fraction of the lensed galaxy images will overlap with images of other objects, affecting the accuracy of flux and shape measurements. This effect, called "blending", is the focus of the second part of this thesis. We describe the procedure to measure the impact of blending on detection algorithms in the current LSST Science Pipeline – specifically the impact of galaxies that go undetected due to blending (unrecognized blends). We investigate two approaches for detecting such galaxies from their model residual images: in the first approach, we train a Convolutional Neural Network (CNN) to detect unrecognized blends; in the second, we apply an iterative detection approach by running the science pipeline twice. Lastly, we quantify the increase in detection efficiency and accuracy with these two approaches and comment on the potential impact on measurements of dark energy.

Acknowledgments

Six years is a long time to be worrying about tiny effects that could impact results from a telescope that hasn't even started observing yet. But knowing that LSST has been a dream in the making for about 40 years now, one can truly appreciate the attention to tiny systematic effects. It has been an interesting journey to see LSST being built from the ground up and to see the Dark Energy Science Collaboration (DESC) grow. This journey would be incomplete without acknowledging all the people who have helped, guided, supported, and made it a wonderful experience.

First and foremost I would like to thank my advisor Pat Burchat for all the mentoring, encouragement, kindness and patience any graduate student could hope for. Thank you Josh Meyers, for basically teaching me how to code and for being a constant source of guidance throughout my thesis. I am grateful to Claire Hébert, my fellow graduate student and companion in navigating through PhD life. Thank you – past and honorary members of the Burchat group: Pierre-François Léget, Ismael Mendoza and Mandeep Gill for all the discussions and the fun conversations. The amount of time we have collectively stared at plots, trying to reason and understand what we see, while also being on the lookout for bugs in the code, has been very valuable in the writing of this thesis.

I am grateful to David Kirkby for his help and feedback on all things blended. I am thankful to Peter Melchior for suggesting the idea of identifying unrecognized blends from `SCARLET` residual images. Thank you to the other creators and maintainers of `SCARLET` for being very helpful and responsive: Fred Moolkamp and Rémy Joseph. Thank you WL and `GalSim` experts Rachel Mandelbaum and Mike Jarvis. Thank you Javier Sanchez for always being a ping away when I needed help with my simulations.

I would also like to acknowledge my reading committee, Professor Sarah Church and Professor Steven Allen, for their feedback both on this thesis and research in general. I would like to acknowledge other KIPAC faculty who have been a part of this journey. I would like to thank the departmental administrators, both from KIPAC and the Physics Department. I am thankful to the people making sure the computing facilities at NERSC, Sherlock, SLAC, CC-IN2P3 run smoothly, so I could do my research.

I would like to acknowledge my fellow KIPAC and/or DESC people whose company has made it all very enjoyable. Thanks Anna, my first friend here and for getting through all the coursework together. PF, eagerly waiting for a ride on your boat! Other people who have made my PhD life enjoyable: Ji Won, Lori, Jamie, Erfan, Divy, Mike, Chris, Adam, JB, Bela, Emily, Dong Woo, Melisa, Sean, Nicole, Ronen, Kate, Stan, Edgard, Emil, Diogo, Chaya, David, Violet, Maria.

I am grateful to all my physics and astronomy professors in school and college, particularly Lawrence Pinto, Anandamayee Tej, Anand Narayanan, Sameer Mandal, and Maheshwar Gopinathan.

None of this would have been possible were it not for my family. Thank you Amma, my mom, for always supporting and encouraging me. Thank you Aanu, Akka, Jeeju, Shwe, B, Sidd, Manganna, Gondu and Chintu. Thank you to my partner in crime since birth, Indu. Thank you Nono and Guha for all the good times. Thank you for everything, my CCD, my binary star, Basheer.

Contents

Abstract iv

Acknowledgments v

1 Introduction 1

- 1.1 Our understanding of the Universe 1
- 1.2 Standard Model of Cosmology (ΛCDM) 2
  - 1.2.1 Measuring cosmic expansion 3
  - 1.2.2 Measuring growth of large scale structure 5
- 1.3 Observational probes of cosmology 8
- 1.4 Survey science era of cosmology: LSST 9
- 1.5 Basics of weak lensing 11
  - 1.5.1 Photometric redshifts for weak lensing measurements 12
  - 1.5.2 Cosmic shear formalism 13
  - 1.5.3 Shear estimation from galaxy shapes 13
- 1.6 Observational systematic effects 14
  - 1.6.1 Shear calibration for the LSST 15
- 1.7 Motivation for this thesis 16
  - 1.7.1 Color gradients 16
  - 1.7.2 Blending 17

2 Color Gradient Bias 18

- 2.1 Introduction 18
- 2.2 Origin of color gradient bias 20
  - 2.2.1 Chromatic PSF 20
  - 2.2.2 Effective PSF for correction 21
  - 2.2.3 Weight function in shape measurements 22
  - 2.2.4 Illustration of color gradient bias for a simple bulge + disk galaxy 22
- 2.3 Isolating and quantifying CG bias 24

2.3.1 Isolating shear bias due to CG . . . 25
2.3.2 Creating equivalent galaxy with no color gradient . . . 25
2.3.3 Moment-based shape estimation with `GalSim` . . . 27
2.4 Estimating CG bias with parametric galaxies . . . 28
2.4.1 Observing conditions . . . 28
2.4.2 Parametric galaxy with color gradients . . . 29
2.4.3 Reference parametric bulge + disk galaxy with extreme color gradients . . . 30
2.4.4 CatSim catalog of parametric galaxies . . . 35
2.5 Estimating CG bias with chromatic real galaxies . . . 38
2.5.1 `ChromaticRealGalaxy` model . . . 39
2.5.2 Limitations of using real galaxy images . . . 40
2.5.3 Estimating CG bias with AEGIS galaxies . . . 42
2.6 Limitations of this study . . . 47
2.7 Conclusions . . . 48

3 The Blending Challenge for LSST **50**
3.1 Overview . . . 50
3.1.1 Why blending is a challenge for LSST . . . 50
3.1.2 Impact of blending on LSST weak lensing measurements . . . 52
3.2 Addressing the blending challenge with the LSST Science Pipeline . . . 56
3.2.1 Detection . . . 56
3.2.2 Different types of blending detection errors . . . 57
3.2.3 Deblending . . . 58
3.2.4 Measurement . . . 60

4 Unrecognized Blends **64**
4.1 Impact of unrecognized blends on cosmological inference . . . 64
4.2 Unrecognized blends in the current LSST Science Pipeline . . . 65
4.3 Dataset . . . 65
4.3.1 `WeakLensingDeblending` package . . . 66
4.3.2 Definition of a blended group . . . 66
4.3.3 Rendering images of the simulated blend scenes . . . 71
4.3.4 Matching detections to true objects . . . 72
4.4 Detection with the LSST Science Pipeline . . . 73
4.5 Results . . . 74
4.5.1 Potential surrogates for characterizing blending . . . 77
4.5.2 Characteristics of undetected galaxies in unrecognized blends . . . 78
4.6 Conclusion . . . 79

5 `ResidualDetectron`: Object Detection with Neural Networks 81
5.1 Past work on neural networks in astronomy . . . 81
5.2 Overview of Neural Networks . . . 83
  5.2.1 Backpropagation . . . 85
  5.2.2 Stochastic gradient descent . . . 85
  5.2.3 Activation function . . . 85
  5.2.4 Regularization techniques . . . 85
  5.2.5 Convolutional neural networks . . . 86
  5.2.6 Pooling layers . . . 87
  5.2.7 Batch Normalization . . . 88
5.3 Object detection with neural networks . . . 88
5.4 Improving object detection with `ResidualDetectron` . . . 90
  5.4.1 Overview of Mask R-CNN . . . 90
5.5 `ResidualDetectron` network architecture . . . 92
  5.5.1 Network Input . . . 92
  5.5.2 Region proposal network . . . 93
  5.5.3 Extracting regions of interest from feature maps . . . 103
  5.5.4 Head branch . . . 105
  5.5.5 Network training . . . 106
  5.5.6 Loss function . . . 106
5.6 Data augmentation for training . . . 107

6 Identifying Unrecognized Blends from Residual Images 108
6.1 Methodology . . . 108
  6.1.1 Modeling the scene with `SCARLET` . . . 109
  6.1.2 Assumptions made in this study . . . 110
6.2 Datasets . . . 110
  6.2.1 Test data . . . 111
  6.2.2 Assessing detection performance . . . 111
6.3 Residual detection with `ResidualDetectron` . . . 112
  6.3.1 Network input . . . 112
  6.3.2 Network output . . . 113
  6.3.3 Training the network . . . 116
  6.3.4 Results . . . 120
6.4 Iteratively detecting sources with LSST Science Pipeline . . . 125
  6.4.1 Difference between iterative `LSP` detection and `ResidualDetectron` . . . 126
  6.4.2 Results . . . 127
6.5 Comparisons of detection methods . . . 131

6.6 Potential limitations of the `ResidualDetectron` method 133
6.7 Future work 135
6.8 Conclusion 136

7 Conclusions 138
7.1 Color gradients 138
7.2 Blending 139
7.3 Final Thoughts 140

A Testing `ChromaticRealGalaxy` 141
A.1 Testing CRG with the reference galaxy with extreme color gradients 141
A.2 Testing CRG with CatSim galaxies 145

B Blending Tool Kit: Framework for testing algorithms 150

Acronyms 152

Symbols 153

Glossary 154

List of Tables

1.1 Parameters of ΛCDM model with Planck CMB data 8

2.1 Parameters of simulated galaxies for bulge/disk components in CG bias study 30
2.2 Results for multiplicative shear bias statistics for $m_{\mathrm{CG}}(z)$, due to color gradient . . . 49

4.1 Survey parameters for LSST *ugrizy* simulations 71

6.1 Description of parameters relevant to the `SEP` source detection algorithm and values used in this study. 118
6.2 Parameters of the LSST Science Pipeline (`LSP`) detection algorithm. 126
6.3 Summary of detection performance of the four detection algorithms analyzed: `LSP`, `SourceExtractor`, `ResidualDetecron`, and iterative `LSP`. 136

A.1 HST and LSST parameters. 143

List of Figures

1.1 Type Ia supernovae Hubble measurement 6

1.2 Growth of structure at large scales in the universe 7

1.3 Forecast of dark energy constraints with the LSST 10

1.4 LSST 10-year r-band depth 11

1.5 Weak lensing transformation applied to a circular source 14

1.6 Illustration of weak gravitational lensing and other observational effects impacting the observed image of a galaxy 15

2.1 Image of galaxy with color gradients 19

2.2 Illustration of origin of CG bias 23

2.3 Illustration of technique for isolating CG shear measurement bias 24

2.4 Schematic diagram of process to create an equivalent galaxy with no CG 27

2.5 Transmission probabilities for LSST r and i bands and HST V and I bands 29

2.6 Test CG bias definition 31

2.7 m_{CG} for reference galaxy with extreme-CG galaxy as a function of SED and redshift 33

2.8 m_{CG} for reference galaxy with extreme-CG galaxy as a function of weight size 34

2.9 m_{CG} for reference galaxy with extreme-CG galaxy as a function of PSF size and chromaticity 35

2.10 Distributions of galaxy parameters for CatSim galaxies in CG bias analysis 37

2.11 Histograms of CG multiplicative shear bias in the r and i bands for CatSim galaxies 38

2.12 Example of HST V- and I-band galaxy images simulated as they would be seen by LSST in r and i bands, using `ChromaticRealGalaxy` 39

2.13 Flow chart describing the method for measuring CG bias for real galaxies from the AEGIS survey. 42

2.14 Distributions of AEGIS galaxy parameters in CG bias study 44

2.15 HST SNR of AEGIS galaxies 45

2.16 Dependence of estimated CG bias of AEGIS galaxies on HST SNR 46

2.17 Results of CG bias estimates for high SNR AEGIS galaxies 47

2.18 Estimated CG bias of AEGIS galaxies in LSST r and i bands 47

3.1 Impact of observing depth on blending. 51
3.2 Impact of seeing on blending 52
3.3 Fraction of seriously blended galaxies expected for LSST 54
3.4 Images of unrecognized blends that are resolvable from space but not from ground . 55
3.5 Schematic diagram of detection decision criterion to identify sources in blends 57
3.6 Schematic of different ways blending can impact detection 58
3.7 Illustration of SDSS deblender 62
3.8 Example of source separation by `SCARLET` 63

4.1 Examples of blended groups simulated from cosmoDC2 catalog 67
4.2 Distribution of group size for cosmoDC2 catalog 68
4.3 Distributions of intrinsic properties of galaxies in selected subsets of the cosmoDC2 catalog 70
4.4 Schematic of algorithm for matching true objects to detections 72
4.5 Efficiency matrix for correctly detecting sources with `LSP` 75
4.6 Distributions of properties of galaxies detected and undetected with `LSP` 76
4.7 Probability density of a gold sample galaxies that are detected or not by `LSP` 78
4.8 Distributions of differences in parameters for an undetected galaxy and the closest matched object detected by `LSP` 79

5.1 Example of a fully connected network with one hidden layer 84
5.2 Example of a convolutional layer 87
5.3 Scene of blended galaxies imaged by Hyper Suprime-Cam 89
5.4 Illustration of segmentation of a scene 92
5.5 Example of six-band residual and model image input to `ResidualDetectron` 93
5.6 Example of anchor boxes in object detection 94
5.7 Example of anchor boxes in `ResidualDetectron` 95
5.8 Anchor boxes defined in `ResidualDetectron` 96
5.9 Example of anchor box and refinements on image and feature map 97
5.10 Comparison of pyramidal structure of feature extraction between deep CNN and FPN 98
5.11 Building blocks of a feature pyramid network 99
5.12 Feature extraction with the FPN in `ResidualDetectron` 101
5.13 Schematic of object detection with RPN 102
5.14 Schematic of the head branch of `ResidualDetectron` 105

6.1 Example of using `SCARLET` residual images to identify unrecognized blends 109
6.2 Pixel values in `SCARLET` model and residual image input to `ResidualDetectron` . . . 114

6.3 Impact of `ResidualDetectron` score threshold on the detection precision and recall 115
6.4 Two-galaxy blend example of input and target prediction of `ResidualDetectron` . . 116
6.5 Schematic of `ResidualDetectron` workflow 117
6.6 Example of successful detection by `ResidualDetectron` 121
6.7 Example of failed detection by `ResidualDetectron` 122
6.8 Efficiency matrix for correctly detecting sources with SEP coadd and `ResidualDetectron` predictions 123
6.9 Efficiency matrix for `ResidualDetectron` detections for correctly detected sources with matching algorithm 2 and all detected sources 125
6.10 Schematic of iterative detection with `LSP` 126
6.11 Example of successful iterative detection by the LSST Science Pipeline 128
6.12 Example of failed iterative detection by the LSST Science Pipeline 129
6.13 Efficiency matrix for correctly detected sources with `LSP` alone, and iterative `LSP` . . 130
6.14 Efficiency matrices for iterative `LSP` detections for correctly detected sources with matching algorithm 2 and for all detected sources 131
6.15 Cumulative counts of galaxies correctly detected with different algorithms 132
6.16 Limitations of `SCARLET` for modeling galaxy spiral arms 134
6.17 Extinction of light from background galaxy in an occulting pair. 135

A.1 Flowchart of testing CRG using reference galaxy with extreme-CG 142
A.2 CG bias of reference galaxy with CRG 144
A.3 Studying impact of SNR on CG bias of reference galaxy drawn with CRG 144
A.4 Flowchart of testing CRG using CatSim galaxies 146
A.5 HST and LSST SNR distributions for CatSim galaxies 147
A.6 Distribution of estimated CG bias in r and i bands of CatSim galaxies drawn with CRG 148
A.7 Results of testing CRG to estimate CG bias with CatSim galaxies 149

B.1 Schematic of `BlendingToolKit` workflow 151

Chapter 1

Introduction

1.1 Our understanding of the Universe

Until the 1900s, Newton's theory of gravity – the universal attractive force – formed the basis of our understanding of the universe. By 1915, it was realized that Newtonian gravity was a limiting case of Einstein's general theory of relativity (GR) at low velocities and weak space-time curvature. A significant implication of GR (and Newtonian gravity) was that for a homogeneous universe made of *only* matter and radiation, the attractive force of gravity would lead matter to pull in on itself, causing the cosmic expansion to slow down. This was contradictory to the experimental evidences in the mid to late 1990s, that showed instead that the universe is expanding at an increasing rate, leading to the hypothesis of a mysterious "Dark Energy".

The first direct evidence of the accelerated expansion of the universe was the consensus in results from two different collaborations (Riess et al., 1999; Perlmutter et al., 1999), which each observed distant supernovae and measured both how far they were (from their brightness) and how fast they were receding from us (from their redshift). The fact that distant objects are receding had already been established: measurements of galaxy redshifts by Hubble showed that they recede at speeds proportional to their distance from us (Hubble, 1929). Additionally, the discovery of the cosmic microwave background (CMB) was evidence for the universe being born in a hot big bang, cooling as it expands (Dicke et al., 1965; Penzias & Wilson, 1965). The supernovae results showed that the expansion of the universe is currently accelerating, requiring the existence of a new ingredient, beyond GR, which we dub "dark energy".

Discoveries that shed light on the nature of dark energy could have profound implications on our understanding of the physical universe. Even the simplest mathematical solution requires the presence of a new component of the universe, called the cosmological constant and denoted by Λ. Alternatively, it could also imply a breakdown of GR, opening up possibilities of new physics at cosmological scales.

While at present we have conclusive evidence of the existence of dark energy, a better understanding of its properties requires sub-percent precision. To address this objective, several Stage IV experiments, as defined by the Dark Energy Task Force (DETF) (Albrecht et al., 2006), including the Dark Energy Spectroscopic Instrument (DESI)[1] (DESI Collaboration et al., 2016), Euclid[2] (Laureijs et al., 2010) and WFIRST[3] (Akeson et al., 2019), are underway with a goal of measuring the dark energy density today as well as its evolution. One such ambitious experiment is Vera C. Rubin Observatory[4], which will generate the Legacy Survey of Space and Time (LSST)[5]. Extending over a 10-year period, the LSST will provide petabytes of information about billions of galaxies, enabling precision measurements of the cosmos. To reach the high statistical power that the LSST dataset will allow, we must achieve a deep understanding of potential biases from systematic effects and achieve stringent limits on uncertainties related to the effects. This thesis is devoted to tackling two such systematic effects.

This chapter is divided as follows. We first describe our current understanding of the universe and provide the theoretical background that will help the reader fully appreciate the significance of the dark energy riddle and the approaches to solving it. We then describe how the LSST will use several observational probes in this quest to understand dark energy. We focus on one such observational probe – gravitational weak lensing (WL) measurements. In the remainder of the introduction, we describe the requirements on WL measurements that are necessary for achieving the science objectives. We conclude with a brief transition to two observational systematic effects that are the focus of the thesis: biases introduced by galaxy color gradients and by overlapping images of objects (known as "blending").

1.2 Standard Model of Cosmology (ΛCDM)

In this section, we briefly summarize the current standard model of the universe. For a pedagogical overview, the reader may wish to refer to the review article by Weinberg et al. (2013), or the textbooks by Dodelson (2003), Ryden (2003), and Carroll (2004).

Our cosmological models are based on the assumption that, on large scales ($\approx$100 Mpc), the universe is isotropic and homogeneous – i.e., there is no preferred direction or location in the universe. The universe came into existence with the hot big bang about 13.7 billion years ago, with subsequent cosmic inflation lasting for about 10^{-33} seconds. Atomic nuclei formed within the first minute. About 370,000 years later, recombination led to photons decoupling from electrons, producing the cosmic microwave background (CMB) radiation.

The current standard model of cosmology is the ΛCDM model, in which the universe is composed

---

[1] https://www.desi.lbl.gov/
[2] http://euclid-ec.org
[3] https://www.nasa.gov/wfirst
[4] https://www.vro.org
[5] https://www.lsst.org

of three main components, listed here in order of decreasing contributions to the energy density of the universe (given as percentages of the total energy density). First is the cosmological constant, Λ, associated with dark energy ($\approx$68%). Second is the cold dark matter (CDM, 27%), which interacts only through gravity. Third is ordinary matter, which accounts for less than 5%. The "cold" of CDM is due to observational evidence that requires dark matter to be cold or non relativistic, in order to clump efficiently and form the large-scale structure we observe. If high temperature (relativistic) particles dominated the dark matter, the large-scale structure would not be consistent with what we see today. The evolution of the universe is understood as a balancing act between cosmic acceleration and structure formation.

1.2.1 Measuring cosmic expansion

Cosmic expansion is parametrized by the scale factor a, which measures the increase in spatial distances as the universe expands. For a hypothetical grid for which the grid lines (and the points at which they cross) remain at fixed positions in space as the universe expands, the comoving distance between two grid points measures the distance between the grid coordinates. However, the physical distance between the two points is proportional to the scale factor and increases with time. By definition $a = 1$ corresponds to the present, implying that in the past $a < 1$.

Light traveling through expanding space is also affected: since galaxies are receding from us, light with wavelength emitted at λ_{emit} will be observed by us at a wavelength $\lambda_{\text{obs}} = \lambda_{\text{emit}}/a$. This stretching of the wavelength is a useful observational tool to measure the scale factor by comparing the observed wavelength of spectral features to their rest-frame wavelengths.

An important observational measurement related to a is the cosmological redshift z:

$$1 + z = \frac{\lambda_{\text{obs}}}{\lambda_{\text{emit}}} = \frac{1}{a}. \tag{1.1}$$

For example, $z = 0$ corresponds to the present ($a = 1$). A redshift of $z = 1$ corresponds to $a = 1/2$; i.e., when the light was emitted, each volume of space was a factor of 2 smaller in each dimension compared to its size today.

The expansion of a homogeneous and isotropic universe is governed by the Friedmann equation (see for example Equation 1.2 in Dodelson (2003)):

$$\frac{H^2(z)}{H_o^2} = \Omega_m(1+z)^3 + \Omega_r(1+z)^4 + \Omega_\phi \frac{u_\phi(z)}{u_\phi(z=0)}, \tag{1.2}$$

where $H(z)$ is the Hubble parameter, defined as $H(z) \equiv \dot{a}/a$, and Ω_m, Ω_r, and Ω_ϕ are the present-day energy densities of matter, radiation, and a generic form of dark energy, each expressed as a fraction of the critical density $\rho_{\rm cri}$, for which the universe is “flat”:

$$\Omega_x = \frac{u_x}{\rho_{\rm cri}c^2},\ \rho_{\rm cri} = \frac{3H_0^2}{8\pi G}. \tag{1.3}$$

The different scaling of each term with redshift indicates that the fractional energy density contributed by each component itself evolves over time. If dark energy is described by a “cosmological constant” (Λ), then, by definition, the energy density Ω_Λ does not evolve with time.

For a flat universe, the sum of all densities is equal to the critical density. The departure from flatness is parametrized by the spatial curvature density Ω_k defined as,

$$\Omega_k \equiv 1 - \Omega_m - \Omega_r - \Omega_\phi. \tag{1.4}$$

$\Omega_k = 0$ for a flat universe. $\Omega_k < 0$ implies the universe is closed (i.e., space has positive curvature) and $\Omega_k > 0$ results in an open universe (i.e., space has negative curvature).

The source of cosmic acceleration is ascribed to dark energy, whether it be a cosmological constant Λ with constant energy density, or an unknown scalar field ϕ with a slowly varying energy density. The equation of state of a scalar field with energy density u and pressure p can be characterized by the dimensionless ratio $w \equiv p/u$. The ΛCDM model describes dark energy as a constant cosmological constant with $w = -1$ for all time.

To test for other theories of dark energy, a linear parameterization of the redshift dependence of the equation of state parameter for dark energy is commonly used:

$$w(a) = w_0 + (1-a)w_a, \tag{1.5}$$

where w_0 is the value of w at present and w_a determines the evolution of w with time. The evolution of the dark energy density is described by this equation (see for example Equation 22 in Weinberg et al. (2013)):

$$\frac{u_\phi(z)}{u_\phi(z=0)} = \exp\left[3\int_0^z [1+w(z')]\frac{dz'}{1+z'}\right]. \tag{1.6}$$

For constant w, this reduces to

$$\frac{u_\phi(z)}{u_\phi(z=0)} = (1+z)^{3(1+w)}. \tag{1.7}$$

The common cosmological distance measures $D(z)$ are co-moving distance $D_C(z)$, luminosity distance $D_L(z)$ and angular distance $D_A(z)$, defined for a flat universe as,

$$
\begin{aligned}
D_C(z) &= \frac{c}{H_0}\int_0^z dz' \frac{H_0}{H(z')} && (1.8)\\
D_A(z) &= \frac{D_C(z)}{1+z} && (1.9)\\
D_L(z) &= D_C(z) \times (1+z). && (1.10)
\end{aligned}
$$

We see from Equation 1.2 and Equation 1.8 that measuring $H(z)$ and $D(z)$ constrains the dark energy equation-of-state parameter $w(z)$ and energy density Ω_ϕ. Some astrophysical probes can be used to measure $H(z)$ directly while others measure the redshift-distance relation. An example of the latter is the Type Ia supernovae measurement, briefly discussed in Section 1.1. The height and shape of a Type Ia supernova light curve is correlated to its intrinsic peak luminosity. A comparison of the observed luminosity of a supernova to its predicted intrinsic luminosity is used to compute the distance to the source, while spectroscopic measurements produce redshifts estimate. Measurements of distance as a function of redshift can help constrain different cosmologies. Figure 1.1 shows measurements of Type 1a supernovae that were instrumental in proving the existence of dark energy. The distance is estimated as the difference between the observed and intrinsic magnitudes[6] (or distance modulus).

1.2.2 Measuring growth of large scale structure

The fractional energy densities Ω_x denote the spatial average and do not describe the "lumpiness" of matter that we see around us on large scales (e.g., the observed groups or clusters of galaxies that are hosted by dark matter clumps). To explain the lumpiness, we need to look at the history of structure formation.

Inhomogeneities in the density of the matter distribution in the universe, which were initially small (of order 1 part in 10^5) but have grown due to the gravitational attraction of dark matter, are understood to be the seeds for all structure growth. The inflationary model of the universe, supported by CMB anisotropy measurements, predict that these tiny perturbations were described by a Gaussian random field. For small density fluctuations, linear perturbation theory describes the density contrast δ with a linear growth function $G(t)$ as

$$\delta(\mathbf{x},t) = \frac{\rho(\mathbf{x},t)-\bar{\rho}(t)}{\bar{\rho}(t)} = \delta(\mathbf{x},t_i)\frac{G(t)}{G(t_i)}, \qquad (1.11)$$

where $\bar{\rho}(t)$ is the mean matter density at time t, $\mathbf{x}$ is the comoving spatial coordinate, and t_i is an arbitrarily chosen initial time.

In GR-based linear perturbation theory, the growth function is described by(see Equation 14 in

---

[6] In astronomy, brightness is measured in a unit called "magnitude" that is proportional to the log of the flux.

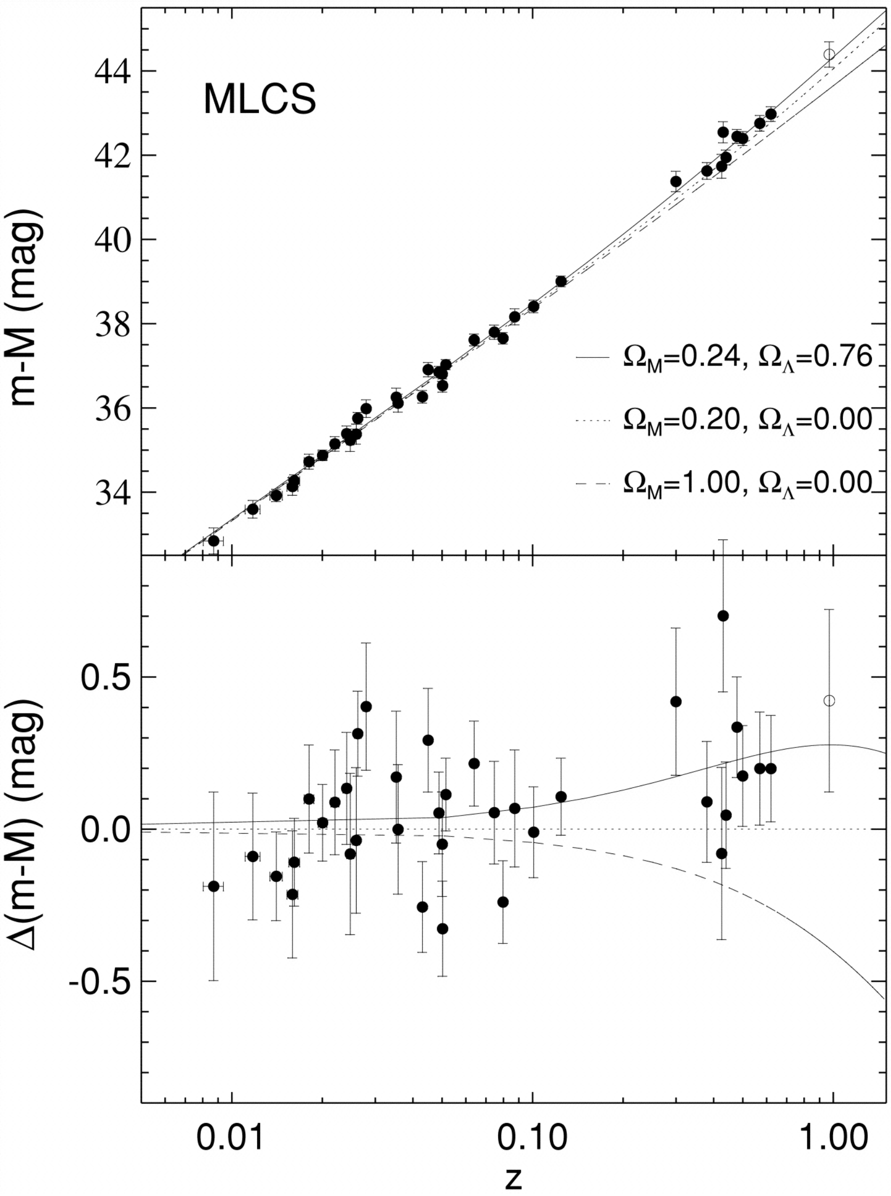


Figure 1.1: Hubble diagram plotting distance-redshift relation to constrain dark energy. The redshift is plotted along the horizontal axis. The distance estimated from the difference between the observed and intrinsic magnitudes (distance modulus) is plotted on the vertical axis. The points denote measured values from observations. The three curves show the distance-redshift relation for three cosmologies with different values of Ω_Λ and Ω_m. The bottom panel shows the difference in distance between the measured values (points) or two cosmologies (solid and dashed curves) and those predicted by a cosmology with no dark energy ($\Omega_\Lambda = 0$) and $\Omega_m = 0.2$. The stronger consistency of the points with the solid curve show evidence for a non-zero dark energy component. Figure is from Riess et al. (1998).

(Weinberg et al., 2013))

$$\ddot{G} + 2H(z)\dot{G} - \frac{3}{2}\Omega_m H_0^2(1+z)^3 G = 0. \tag{1.12}$$

The initial small density fluctuations were random but correlated – i.e., the fluctuation at one point in space is statistically related to another point according to their separation distance. The two-point correlation function of the density perturbation field, with power spectrum $P_\delta(k)$, is given by

$$\langle \delta(k)\delta^*(k) \rangle = (2\pi)^3 P_\delta(k)\delta_D^3(k - k'), \tag{1.13}$$

where k and k' are two Fourier modes, and δ_D is the Dirac delta function. The overall amplitude of the spectrum is parametrized by σ_8, the variance in the matter density in spheres of radius $8h^{-1}$Mpc.

While the gravitational interaction between dark matter particles is the primary reason for structure formation, the actual shape of the matter power spectrum is impacted by baryonic physics and non-linear processes in the later stages of the history of the universe. The linear growth function for two different cosmologies and their impact on structure formation are shown in Figure 1.2.

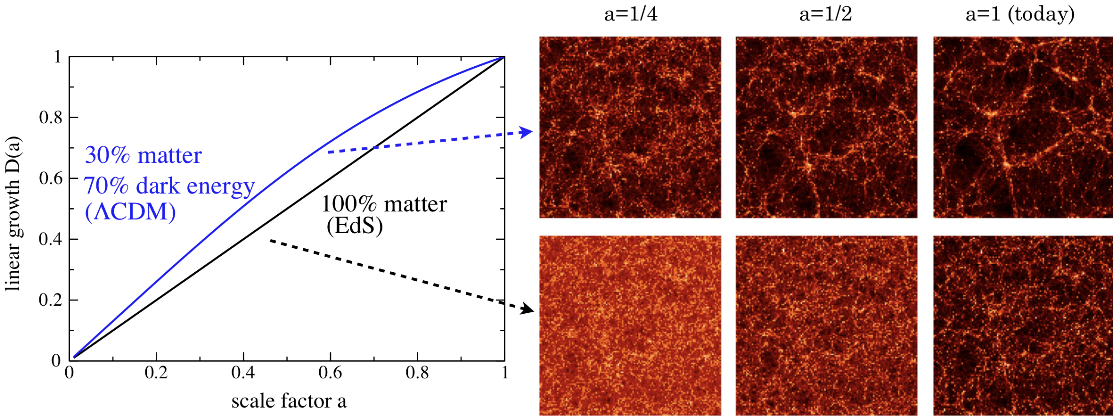


Figure 1.2: Growth of structure at large scales in the universe. Left: Linear growth function G ($D(a)$ in the figure) as a function of scale parameter a for two different cosmologies, ΛCDM (blue line) and Einstein-de Sitter (EdS) model with $\Omega_\lambda = 0$ (black line). Right: Snapshots from numerical (N-body) simulations by the Virgo consortium(Jenkins et al., 1998), at different timescales of the universe with ΛCDM (top row) showing a larger amplitude of density fluctuations in the past in ΛCDM (top row) and EdS (bottom row) models. Figure is from Huterer et al. (2015).

The general approach to measuring dark energy involves constraining $H(z)$, $D(z)$, and $G(z)$ in order to constrain $w(z)$. CMB anisotropy measurements play an important part in constraining nuisance parameters like Ω_m, Ω_k, and σ_8. CMB measurements are incorporated as priors on parameters that are then marginalized over during analysis.

Table 1.1: Summary of current best estimates of ΛCDM model parameters with 68% confidence limits. Each estimate comes from combining Planck CMB temperature, polarization, and lensing power spectra, and BAO results, except w_0, which is estimated from Planck + Type Ia supernovae measurements. These values are derived in Planck Collaboration et al. (2018).

Parameter	Description	Value
H_0	Hubble constant [$\mathrm{kms^{-1}Mpc^{-1}}$]	67.7 ± 0.4
Ω_Λ	dark energy density	0.688 ± 0.006
Ω_m	matter density	0.311 ± 0.006
Ω_k	curvature parameter	0.0007 ± 0.0002
σ_8	RMS matter fluctuation	0.810 ± 0.006
t_0	Age of the universe [Gyr]	13.8 ± 0.02
w_0	dark energy equation of state	-1.03 ± 0.03

1.3 Observational probes of cosmology

To understand the nature of dark energy we must address the following with observational probes:

1. Determine whether the cosmic acceleration is consistent with being due to a cosmological constant ($w = -1$). If not, then measure the time evolution of the dark energy equation of state parameter, $w(a)$.
2. By comparing the effect of dark energy on cosmic acceleration and growth of structure, search for potential failures of GR.

The history of cosmic expansion is imprinted on a number of astrophysical observations we can make from Earth. Of these, the five observational probes with the most promise to constrain dark energy parameters identified by the LSST Dark Energy Science Collaboration (The LSST Dark Energy Science Collaboration, 2019) are:

- Weak Lensing: analysis of observed galaxy shapes, which are distorted by the bending of light due to fluctuations in matter density, allows measurements of the growth rate of cosmic structure.
- Large Scale Structure: measurements of galaxy positions allows constraints on the large-scale power spectrum of the spatial distribution of matter as a function of redshift. Large scale structure probes include baryonic acoustic oscillations (BAO) which also allows the measurement of redshift-distance relation.
- Galaxy Clusters: the spatial density, distribution, and masses of galaxy clusters as a function of redshift are sensitive to the expansion history and growth of structure.

- Type Ia supernova: measurements of redshift-distance relation from light curve analysis allows measurements of cosmic expansion(see Figure 1.1 for an example of this).
- Strong Lensing: the angular displacement, morphological distortion, and time delay for the multiple images by gravitational lensing of massive foreground objects allows measurements of the evolution of cosmic geometry.

1.4 Survey science era of cosmology: LSST

In order to constrain the power of different probes for measuring the nature of dark energy, the Dark Energy Task Force (Albrecht et al., 2006) developed a quantitative "figure of merit" (FoM) corresponding to the reciprocal of the area enclosing the 95% confidence limit for measurements in the $w_0 w_a$ plane. One of their findings was that no single observational technique could produce the desired FoM, leading to the conclusion that a combination of techniques is necessary. A combined analysis of the observational probes will increase the statistical precision and be more robust to systematic errors. To this end, several Stage IV experiments[7] were proposed. One such proposal was for a dedicated ground-based, multi-band, wide-field imaging survey: the Large Synoptic Survey Telescope (LSST) (LSST Science Collaboration et al., 2009; Ivezić et al., 2019).

The predicted power to constrain dark energy parameters with the 10-year LSST data set, as reported in the LSST DESC Science Requirements Document (The LSST Dark Energy Science Collaboration et al., 2018), is shown in Figure 1.3. The contours show the forecasts for the expected precision for LSST measurements of the dark energy parameters with each of the five probes described in Section 1.3 and the joint forecast from their combined analysis.

The LSST, now renamed the Legacy Survey of Space and Time, will be based on observations at the NSF (National Science Foundation) Vera C. Rubin Observatory[8] at Cerro Pachón in Chile, using the 3.2 gigapixel DOE (Department of Energy) LSST Camera. The LSST will observe through the Simonyi Survey Telescope, a three-mirror system with an 8.4-m primary mirror, enabling a $10\,\text{deg}^2$ field of view optimized for the ability to survey the entire night sky (visible from the Southern Hemisphere) every three nights. The Rubin Observatory is expected to see first light in 2022 and will continue observing for the LSST until 2032.

Over its 10 year observing period, the telescope is expected to survey $\approx 20,000\,\text{deg}^2$ of the night sky, with each sky location expected to be visited about 100 times per year. A visit consists of two consecutive 15-second exposures (although a single 30-second exposure is also being considered). Rubin Observatory will observe the sky in six photometric bands in the wavelength range 320 to

---

[7] This refers to the highest of the four stages of experiments with increasing ability to constrain dark energy studied by the Dark Energy Task Force.

[8] In honor of Vera Rubin, whose pioneering work on measuring galaxy rotation curves led to the discovery of dark matter. This is the first observatory to be named after a woman.

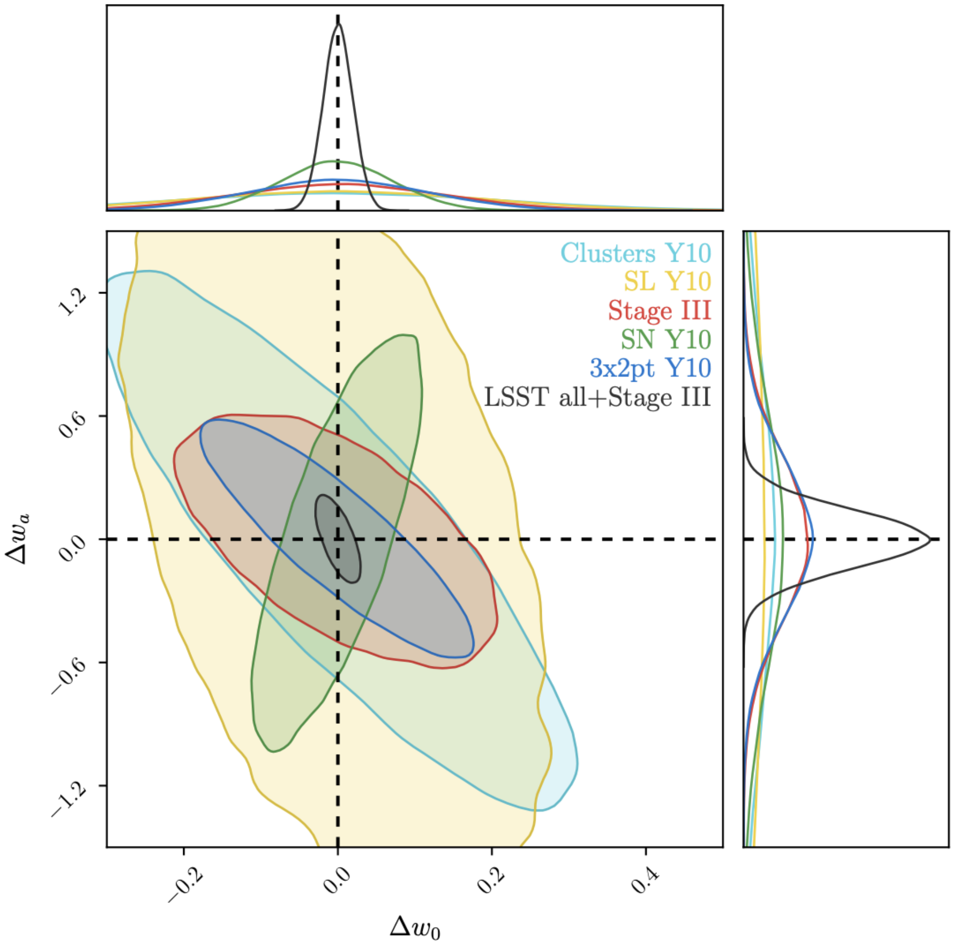


Figure 1.3: Forecast of dark energy constraints with the 10-year LSST survey. The constraining power is computed for each of the five probes described in Section 1.3 individually: weak lensing and large scale structure (3×2 pt), galaxy clusters (Clusters), Type Ia supernova (SN), and strong lensing (SL). The forecast for the joint analysis, in black, includes Stage III priors. The red contour shows the expected precision from Stage III surveys. Each contour shows the 68% confidence interval of the dark energy equation of state parameters w_a and w_o (see Equation 1.5). Figure from The LSST Dark Energy Science Collaboration et al. (2018).

1050 nm: *ugrizy*. The predicted distribution of coadded depth observed by the LSST r-band in 10 years is shown in Figure 1.4.

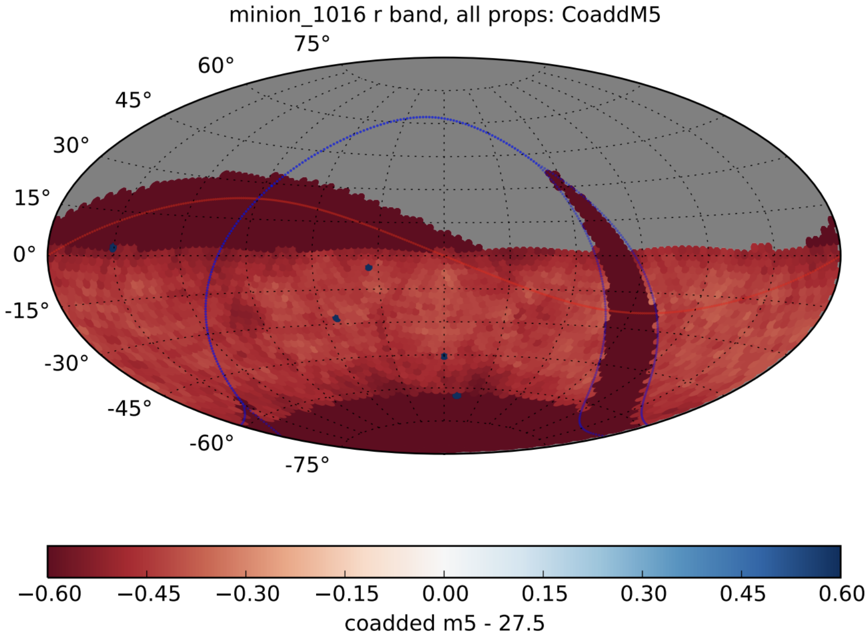


Figure 1.4: The coadded 5σ depth for point sources in the r-band across the sky by the end of 10 years for the LSST. Figure from LSST Science Collaboration et al. (2017).

Along with constraining dark matter and dark energy, the LSST has three other primary science drivers (LSST Science Collaboration et al., 2009):

1. Taking an inventory of the solar system: The LSST will map the Kuiper Belt Objects (KBOs) and identify potentially hazardous asteroids.
2. Exploring the transient optical sky: Transient objects like gamma-ray bursts, microlensing events, etc., are identified by comparing nightly images. Community-wide notification will be sent within minutes of identification, so that events can be followed up by other telescopes.
3. Mapping the Milky Way: LSST will map the 3-D shape and extent of the halo of our Galaxy. This will help better understand its structure and accretion history.

1.5 Basics of weak lensing

In this section we summarize the theory of the weak lensing (WL) probe, including an overview of the sources of systematic uncertainty and the requirements on their maximum size, for accurate cosmological analysis. For a more detailed coverage of the subject the reader may refer to several excellent sources of documentation such as Schneider (2005), Mandelbaum (2018), Seitz & Schneider (1997), and Bartelmann & Schneider (2001).

In describing weak lensing measurements, the Dark Energy Task Force (Albrecht et al., 2006) states that, "*If* the systematic errors are at or below the level asserted by the proponents, it is likely to be the most powerful individual Stage-IV technique and also the most powerful component in a multi-technique program". These systematic errors may stem from either observational or astrophysical effects.

Light emitted from background galaxies is continuously deflected as it travels through the gravitational potential of the matter density field (this includes both dark and ordinary matter). This "weak" lensing produces coherent distortion and alignment of images of distant galaxies. An observer can measure correlations in these distortions to determine the statistical properties of the intervening mass distribution as a function of redshift. These perturbations are small ($\approx 0.5\%$) when compared to the initially random orientations of galaxies ($\approx 30\%$), thereby requiring a large number of galaxies for meaningful cosmological measurements.

Summary statistics are extracted for two types of WL measurements: cosmic shear, which correlates galaxy shapes with galaxy shapes; and galaxy-galaxy lensing, which correlates galaxy positions with galaxy shapes. A cosmological analysis is performed by combining the WL summary statistics with the summary statistics of galaxy clustering (correlating galaxy positions with galaxy positions) to form a set of three two-point functions or 3×2pt analysis.

Constraining cosmology from WL measurements involves controlling both astrophysical and observational systematic effects. Astrophysical effects include intrinsic galaxy alignments, effects of baryonic feedback processes, the non-linear evolution of the matter density field, etc.. Our main focus in this thesis is to understand and mitigate observational systematic effects.

The two primary observables in WL measurements are 1) galaxy shapes, from which we infer the shear signal, and 2) the redshift of the galaxies, inferred from multi-band photometry.

1.5.1 Photometric redshifts for weak lensing measurements

In this section, we briefly discuss photometric redshift estimation and systematic uncertainties produced by them. There are two steps to redshift estimation: 1) measure photometric flux in the observing bands, and 2) use multi-band photometry to compute the photometric redshift.

Galaxy fluxes are measured for the six LSST photometric bands within a fixed aperture, while matching the PSF variations across the bands. Errors in PSF modeling and aperture matching can introduce errors in the photometric flux estimates. Correlations in redshift errors due to observational or astrophysical effects such as galactic extinction can also introduce biases. With multi-band photometry, the six-band spectral energy distribution (SED) is computed for each galaxy.

There are two common methods to convert SEDs into photometric redshift estimates:

1. Template-fitting: A set of templates for galaxy SEDs are used to predict the galaxy photometry as a function of redshift, by comparing them to the observed photometry. However, issues arise due to insufficiency of the templates to accurately describe the full span of the real data.
2. Machine learning methods: These can empirically learn the relationship between photometry and redshift based on a training sample. The drawback however is that biases can be introduced if the training data does not match the observed galaxies.

Both methods require spectroscopic redshifts for training or calibration and therefore are limited by insufficient spectroscopic redshift samples representative of redshifts at LSST observing depths.

The LSST WL analysis places two stringent limitations on the uncertainty associated with photometric redshift. The first is that the uncertainty on the mean redshift in each tomographic bin should not exceed $0.001(1+z)$. The second is that the uncertainty in the width of the redshift distribution in the tomographic bin should not exceed $0.003(1+z)$.

1.5.2 Cosmic shear formalism

Weak lensing distortions can be described by the linear transformation between the unlensed (x_u, y_u) and lensed (x_l, y_l) coordinates[9]:

$$\begin{pmatrix} x_u \\ y_u \end{pmatrix} = \begin{pmatrix} 1-\gamma_1-\kappa & -\gamma_2 \\ -\gamma_2 & 1+\gamma_1-\kappa \end{pmatrix} \begin{pmatrix} x_l \\ y_l \end{pmatrix} = (1-\kappa) \begin{pmatrix} 1-g_1 & -g_2 \\ -g_2 & 1+g_1 \end{pmatrix} \begin{pmatrix} x_l \\ y_l \end{pmatrix} \tag{1.14}$$

The stretching in galaxy images is described by the complex lensing shear $\gamma = \gamma_1 + \mathrm{i}\gamma_2$, or reduced shear $g = \gamma/(1-\kappa)$, where the convergence κ describes the change in size and brightness of the lensed objects.

The effect of shear and convergence on a circular source is shown in Figure 1.5. Since galaxies are not intrinsically round, the observed shapes are a combination of intrinsic ellipticity and shear.

1.5.3 Shear estimation from galaxy shapes

The shape of a galaxy with flux profile $I(\mathbf{x})$ can be computed from its second moments, defined as

$$Q_{ij} = \frac{\int d^2x I(\mathbf{x}) W(\mathbf{x}) x_i x_j}{\int d^2x I(\mathbf{x}) W(\mathbf{x})}, \quad i,j \in 1,2, \tag{1.15}$$

where $W(\mathbf{x})$ is a weighting function. The galaxy shape can be expressed as a complex ellipticity e:

$$e = e_1 + \mathrm{i}e_2 = \frac{Q_{11} - Q_{22} + 2\mathrm{i}Q_{12}}{Q_{11} + Q_{22} + 2(Q_{11}Q_{22} - Q_{12}^2)^{1/2}} \, . \tag{1.16}$$

---

[9] The origins of the coordinate systems are at the positions of the galaxy.

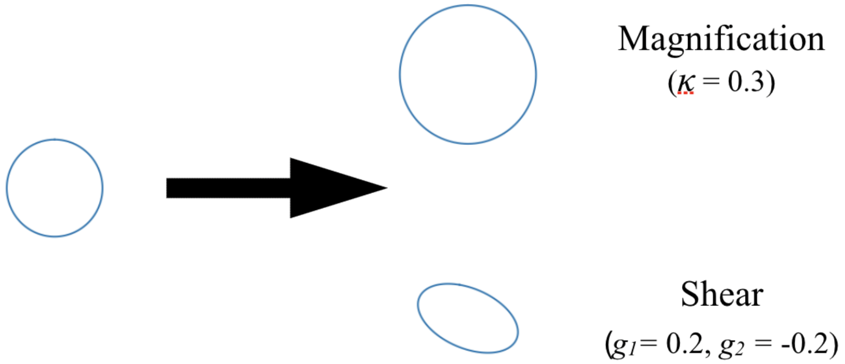


Figure 1.5: Impact of weak lensing on a circular source (left) when transformed according to Equation 1.14. In the absence of shear, a convergence κ only modifies the radius of the circle. A reduced shear g results in an ellipse with an orientation angle.

The reduced shear applied to an unlensed source with intrinsic shape e^{intr} is related to the observed ellipticity e by the transformation

$$e = \frac{e^{\text{intr}} + g}{1 + g^* e^{\text{intr}}}, \quad \text{if } |g| \leq 1. \tag{1.17}$$

Assuming that the intrinsic orientation of the galaxies is random, the mean of their observed sheared ellipticities is the reduced shear, $\langle e \rangle = g$. Because the weak lensing shear is so small compared to the intrinsic, randomly oriented galaxy ellipticities (shape noise), averaging over very large ensembles of galaxies is necessary to achieve small statistical uncertainties.

1.6 Observational systematic effects

We list below a few of the significant observational systematic effects in shear estimation with LSST. For a more exhaustive list refer to Mandelbaum et al. (2018) and The LSST Dark Energy Science Collaboration (2019)

- Point spread function (PSF): One of the primary sources of WL systematic uncertainties is the distortion in galaxy shapes by the PSF, caused by atmospheric seeing (for ground based surveys) and telescope effects. If the PSF distortions are insufficiently corrected, they can introduce a) multiplicative biases that depend on the galaxy properties, and b) an additive bias due to PSF anisotropies. Errors in the model for the size and shape of the PSF result in multiplicative and additive shear biases, respectively. PSF modeling involves two parts:

modeling the PSFs using bright stars, and interpolating the star PSFs to the positions of the observed galaxies on the focal plane.

- Detector effects: Detector non-idealities, such as the brighter-fatter effect, can affect the interpolation of star PSFs to model faint galaxy PSFs. Detector defects that correlate with position or galaxy orientation can induce spurious coherent shear signals or photometry errors.
- Coaddition: Multiple exposures at different pointing positions are combined to reach the full survey depth. This introduces challenges of specifying well-defined PSFs at each point on the coadd, further complicated by chip edge effects and blending.

In Figure 1.6, we show an example of the distortion of a galaxy's shape due to WL and the impacts of observational effects (the point spread function, pixelization and noise) on the observed image of the galaxy.

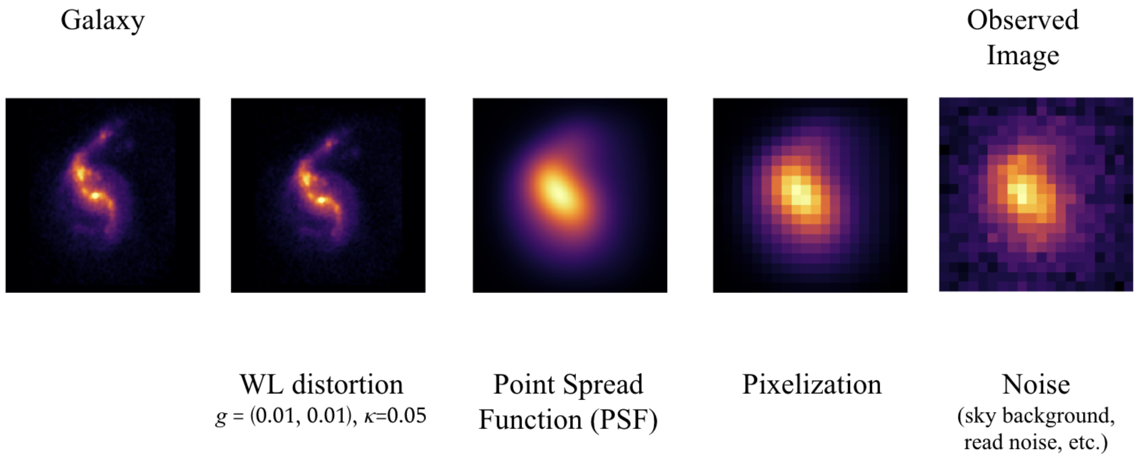


Figure 1.6: Illustration of weak gravitational lensing and other observational effects impacting observations of galaxies by the LSST. The image of a galaxy undergoes distortion from weak lensing, loss of resolution due to smearing by the PSF and pixelization effects, and noise in each pixel, to produce the noisy observed image.

1.6.1 Shear calibration for the LSST

Deviations from the perfect shear measurement are commonly parameterised with multiplicative bias (m) and additive bias (c) terms (Heymans et al., 2006; Massey et al., 2013) as

$$\hat{g} = (1 + m)g + c, \tag{1.18}$$

where $\hat{g}$ is the shear estimated from an ensemble of galaxies with true shear g.

As documented in secs. 5.2, D2.1 and D2.3 in version 1 of the LSST Dark Energy Science Collaboration (DESC) Science Requirements Document (SRD) (LSST DESC et al., 2018), the requirement on the total systematic uncertainty in the redshift-dependent shear calibration is that it not exceed 0.003. Combined with other requirements in the LSST DESC SRD, LSST will then be able to achieve its design constraints on dark energy (LSST DESC et al., 2018; Albrecht et al., 2006).

The SRD analysis uses five photometric redshift bins, each with a width of $\Delta z = 0.2$, in the redshift range $0.2 \leq z \leq 1.2$. For simplicity, a linear parameterization is assumed for the redshift dependence of the multiplicative shear bias, $m(z)$:

$$m\left(z\right) = m_z \left(\frac{2z - z_{\mathrm{max}}}{z_{\mathrm{max}}}\right) + m_{\mathrm{avg}}\,, \tag{1.19}$$

where z_{max} is the redshift value at the center of the highest redshift bin, $m_{\mathrm{avg}} = m(z_{\mathrm{max}}/2)$ is the average value of m in the range $z \in [0, z_{\mathrm{max}}]$, and $2m_z = m(z_{\mathrm{max}}) - m(0)$ is the total variation in m in the range $z \in [0, z_{\mathrm{max}}]$. The 0.003 requirement is then on the total systematic uncertainty in the redshift-dependent shear calibration m_z due to all contributions to multiplicative shear bias. In principle, there could be a higher-order dependence (i.e., non-linear) of $m(z)$ on z. In the absence of an explicit requirement for the case of non-linear dependence of $m(z)$ on z, we map the requirement of $2(0.003) = 0.006$ on the uncertainty on $2m_z = m(z_{\mathrm{max}}) - m(0)$ onto a requirement on the uncertainty on the maximum span of $m(z)$ over the redshift range [0, 1.2].

1.7 Motivation for this thesis

Because of the stringent requirements on systematic effects for WL measurements with LSST, effects that were not of major concern to previous WL surveys, such as DES or KiDS, are significant for LSST. The impacts of these systematic effects are calculated with simulations. If the biases or their uncertainties are found to be comparable to the LSST requirements, then methods to mitigate or correct them must be developed. We focus on two such systematic effects in this thesis: color gradients and blending. While the two effects are not related to each other, we use common techniques and simulations to measure their impact.

1.7.1 Color gradients

High resolution images of galaxies show that their spectral energy distributions vary across the galaxy profile. This color gradient is generally ignored in photometric measurements, where the measured flux values are ascribed to the entire galaxy profile. However, the distortions in the observed galaxy images caused by physical effects that are chromatic in nature, such as atmospheric turbulence, diffraction due to the limited aperture size, etc., will be affected by galaxy color gradients. This motivates our study to measure the impact of ignoring color gradients while modeling PSF corrections for weak lensing measurements.

In Chapter 2, we describe in detail our study of the origin, effects and measurements of color gradient bias using parametric simulations and real galaxy images. This work has been published in Kamath et al. (2020).

1.7.2 Blending

LSST measurements are based on the analysis of three-dimensional cosmological processes using two-dimensional observed images. Thus, it is possible for objects at different distances along our line-of-sight to appear to be overlapping when imaged by the telescope. This superposition of object images is called "blending" and poses a very significant challenge in the detection, separation and measurement of sources. Existing algorithms are known to be inadequate at LSST depths, opening possibilities for testing new and non-standard solutions, including machine-learning approaches that leverage recent advances in "deep learning".

The second half of this thesis is devoted to understanding the impact and challenges associated with blending, along with an overview of existing efforts to mitigate the effects (Chapter 3), with special emphasis on unrecognized blends (Chapter 4), and our proposed method to improve object detection with neural networks (Chapter 5) and iterative detection and measurement techniques (Chapter 6).

Chapter 2

Color Gradient Bias

2.1 Introduction

In addition to cosmic shear, the observed shapes of galaxies are distorted by the atmosphere, the telescope optics, and sensor effects – all of which contribute to the point-spread function (PSF) . The observed images of the galaxies are thus a convolution of the true image and the observing PSF. This makes it vital for weak lensing measurements to model the convolving PSF and correct for it using stars. Additional complexity arises when the distortions caused by the PSF depend on the observing wavelength (Cypriano et al., 2010). The effects of chromatic atmospheric PSFs on weak lensing measurements have been studied by Meyers & Burchat (2015) who conclude that while the predicted shear bias is significant compared to the LSST requirements, the bias can be reduced sufficiently to meet the LSST requirements if the PSF is corrected using multi-band photometry and machine learning techniques. However, these corrections are based on the assumption that the spatial and wavelength dependencies of the galaxy surface brightness profile are independent. If instead the galaxy has a SED that varies across its profile – a "color gradient (CG)" – then the distortions due to the chromatic PSF will be different for different points on the galaxy. An example image of a galaxy with color gradients is shown in Figure 2.1. PSF corrections that are applied assuming a position-*independent* SED do not correct for potential color gradient effects. Thus, it is necessary to quantify the size of this bias compared to the LSST requirements on bias for shear estimators.

Previous studies of shear bias due to color gradients have focused on the space-based Euclid survey (Amiaux et al., 2012), where the wavelength dependence of the PSF is dominated by diffraction, and the wavelength range for the (single) optical filter is ≈ 550 nm to 900 nm. These studies have concluded that shear bias due to color gradients in surveys like Euclid can be significant compared to the requirements on systematic errors (Voigt et al., 2012; Semboloni et al., 2013; Er et al., 2018). We use and extend a method similar to Semboloni et al. (2013) (hereafter S13) to estimate the impact

Figure 2.1: High resolution image of the Whirlpool Galaxy (Messier 51) illustrating color gradients. The spectral energy distribution of the light varies across the galaxy profile. (Image:NASA/ESA/Hubble Heritage Team).

of color gradient (CG) bias on weak lensing measurements with the LSST, where the wavelength dependence of the PSF is dominated by atmospheric effects, and the wavelength range for each filter is $\approx 150\,\mathrm{nm}$.

We first use a reference galaxy with extreme color gradients to demonstrate the presence of CG bias as well as its dependence on atmospheric seeing and shape measurement algorithms. We then measure the expected size of this bias on LSST weak lensing measurements by extending the study to a more representative sampling of color gradients. This is done with two different approaches, each with advantages and disadvantages:

(a) We assemble a parametric catalog of galaxies with a range of color gradients that could be observed with LSST. This method allows us to simulate galaxy images with infinite signal-to-noise ratio and measure bias from CG only. However, the accuracy of the prediction for real surveys depends on the extent to which the simulation represents color gradients in real galaxies.

(b) We use observed high-resolution galaxy images to estimate color gradients. This method allows us to leverage information from real galaxies, potentially leading to more realistic bias estimates. However, when the input galaxy images are noisy (as they are for real galaxies), the effect of noise cannot be removed completely and the estimated bias is a combination of CG bias and residual noise bias (see Refregier et al. (2012), for example, for details on impact of noise bias in shear measurements).

The parametric galaxy simulations for method (a) are generated from the LSST Catalog Simulator, CatSim (Connolly et al., 2014), which contains astrophysical sources with properties that are representative of what the LSST will observe at its ten-year coadded depth[1]. Real galaxy images for method (b) were obtained from HST V- and I-band observations in the All-wavelength Extended Groth strip International Survey (AEGIS)[2].

2.2 Origin of color gradient bias

Shear measurements require accurate estimates of galaxy shapes corrected for PSF distortions. Galaxy color gradients can produce errors in shape measurements if the following conditions hold:

1. The PSF is wavelength dependent.
2. The PSF correction is performed assuming that the galaxy has a uniform SED across its profile.
3. A spatial weight function is used in the estimation of shapes.

We explore in detail below how the three conditions produce CG bias. In this study, we restrict ourselves to analyzing the effects of CG bias for moment-based shape measurement algorithms that use weight functions matched to galaxy profiles. However, color gradients have also been shown to lead to shape measurement errors when "fitting methods" are used, as described in Voigt et al. (2012) where the profiles themselves weight the different regions of the image. Both Voigt et al. (2012), using fitting methods, and Semboloni et al. (2013), using moment-based shape estimation, showed that the CG bias could be substantial in the Euclid survey – exceeding nominal requirements for the multiplicative bias in the shear.

2.2.1 Chromatic PSF

For this analysis we approximate a chromatic PSF, $\Pi(\vec{x}, \lambda)$, by a Gaussian profile with a wavelength-dependent size:

$$\sigma_{\mathrm{PSF}}(\lambda) = \sigma^{o}\left(\frac{\lambda}{\lambda^{o}}\right)^{\alpha}. \tag{2.1}$$

---

[1] https://www.lsst.org/scientists/simulations

[2] http://aegis.ucolick.org

The scaling exponent α is determined by the origin of the dominant source of the chromatic PSF. In the case of space-based telescopes (e.g., HST, Euclid, WFIRST), the PSF chromaticity arises primarily from diffraction due to the finite aperture: $\alpha \approx +1.0$ (Cypriano et al., 2010). For ground-based telescopes like the LSST, the chromaticity is primarily determined by the Kolmogorov turbulence in the atmosphere: $\alpha \approx -0.2$ (Fried, 1966). The PSF size at a reference wavelength λ^o is defined as σ^o.

2.2.2 Effective PSF for correction

The observed image of the galaxy in the ith band, $I_i(\vec{x})$, can be described as the convolution (denoted by $*$) of the galaxy profile $f(\vec{x}, \lambda)$ with the chromatic PSF $\Pi(\vec{x}, \lambda)$, weighted by the ith band transmission function $T_i(\lambda)$ (which includes contributions from the atmosphere, optics, filter, and sensor) and integrated over wavelength:

$$I_i(\vec{x}) = \int f(\vec{x}, \lambda) * \Pi(\vec{x}, \lambda) T_i(\lambda)\, d\lambda. \tag{2.2}$$

Since the PSF is chromatic, its effect on the observed galaxy shape will depend on the observing wavelengths. Similarly, any PSF correction to be applied to a galaxy shape will require knowledge of the galaxy SED over the observing band. If color gradients are ignored then the PSF correction will assume an "effective" SED that is uniform across the galaxy profile in place of the true SED of the galaxy that is position dependent. The effective SED $S_{\mathrm{eff}}(\lambda)$ is defined as the spatially integrated flux of the galaxy as a function of wavelength:

$$S_{\mathrm{eff}}(\lambda) = \int f(\vec{x}, \lambda)\, d\vec{x}. \tag{2.3}$$

The PSF for correction will also be an "effective" PSF in the ith band, $\Pi_{\mathrm{eff},i}(\vec{x})$, corresponding to the image of a point source with SED $S_{\mathrm{eff}}(\lambda)$:

$$\Pi_{\mathrm{eff},i}(\vec{x}) = \int S_{\mathrm{eff}}(\lambda) \Pi(\vec{x}, \lambda) T_i(\lambda)\, d\lambda, \tag{2.4}$$

where we have implicitly evaluated the convolution with a delta-function point source. Semboloni et al. (2013) pointed out that color gradients do not produce a bias if the integrals are allowed to extend to infinity in each direction, which is impossible in practice. All realistic methods use some kind of spatial weighting (even if only to impose zero weight beyond the region of the "postage stamp"), which means that different regions of the galaxy with different colors and PSFs are weighted differently, introducing a potential bias.

2.2.3 Weight function in shape measurements

We described how galaxy shapes are measured from galaxy images in Section 1.5.3, quantifying the ellipticities from the second moments with Equation 1.16. Assuming that the galaxy intrinsic ellipticities are randomly oriented, reduced shear g is estimated by averaging over an ensemble of galaxies. We employ a technique to suppress shape noise without generating a large number of galaxies by simulating six galaxies with equidistant intrinsic ellipticity on a ring around 0 – "ring test" (Nakajima & Bernstein, 2007) to measure shear (see, for example, section 6.1 in Meyers & Burchat (2015)).

When computing second moments, pixel values at larger distances from the centroid have more impact on the measurement than pixel values close to the centroid. In real galaxy shape measurements, the moments are computed from noisy images, where the noise will dominate the pixel values (Melchior et al., 2011). Thus in order to prevent noise divergence in the calculation of second moments, it is common practice to employ weight functions of finite width to limit the integration. A weight function whose centroid, size, and ellipticity are matched to the source galaxy image optimizes the significance of the measurement. A Gaussian weight is generally preferred due to its rapid convergence to zero at large radii and the absence of singularities, as well as general mathematical convenience (Hirata & Seljak, 2003).

By its very nature, the weight function gives more significance to the central region of the galaxy. For a galaxy with a spatially independent SED, this does not produce an error as long as one correctly accounts for the weight function in the shape estimate. However, if there exists a spatial dependence of the galaxy SED, regions with different colors are weighted differently, potentially resulting in a bias in shape estimation that depends on the size and profile of the weight function and the color gradient.

A possible method for reducing the impact of CG bias would be to weight the galaxy profile in the computation of the effective SED (Equation 2.3) with the same weight function used to measure the galaxy shape. We do not explore this here because, in this study, our goal is to understand and quantify the size of shear measurement bias due to color gradients assuming we do not attempt to correct for galaxy CG; we do not explore optimal methods for mitigating the bias if it is necessary to do so.

2.2.4 Illustration of color gradient bias for a simple bulge + disk galaxy

We illustrate in Figure 2.2 the impact of the aforementioned three conditions by comparing the measured shape of a galaxy with and without color gradients. The top row depicts this measurement for a galaxy with no CG when observed with a chromatic PSF and with moments computed using a weight function. The true galaxy (green ellipse) is convolved by the PSF (green circle) and integrated over the observing bandpass to produce the observed galaxy image. The radius of the PSF circle is proportional to the PSF size. As a result of convolution with the PSF, the observed galaxy image appears rounder and larger than the true galaxy. To obtain the correct measured shape the effect of the PSF must be measured. In the absence of CG, the effective SED is the same as the uniform galaxy SED and the effective PSF correctly encapsulates the PSF distortions. The dotted yellow circle denotes the weight function for computing moments, which are corrected using the effective PSF in order to retrieve the correct measured shape (grey ellipse).

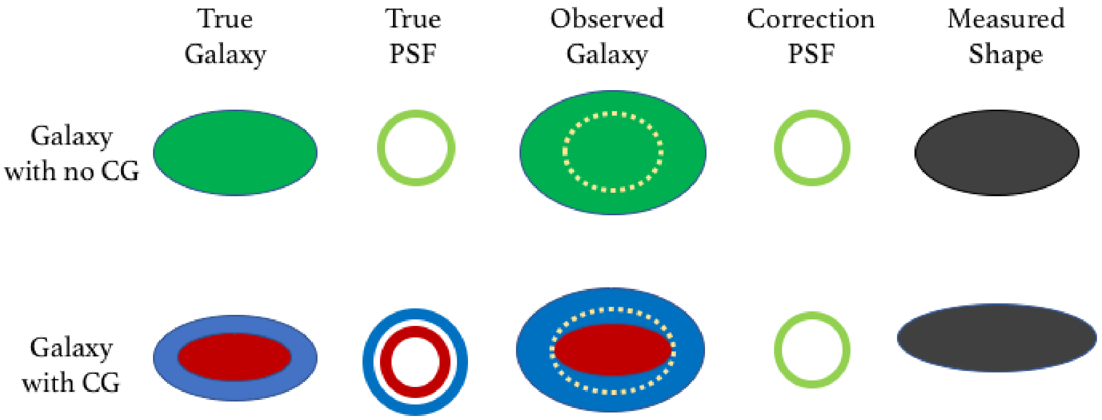


Figure 2.2: Schematic depicting the error in the measured shape of a galaxy with color gradients caused by a wavelength-dependent PSF combined with a weight function for computing moments. The galaxy profiles are denoted as filled ellipses, PSFs as circles, and the weight functions as yellow-dotted ellipses. See Section 2.2.4 in text for details.

The bottom row illustrates the same process but for a galaxy with color gradients. For simplicity we demonstrate this for a galaxy with radially dependent color gradient, where the galaxy has a compact central bulge (red ellipse) and an extended disk (blue ellipse). Both the bulge and the disk have the same centroid and shape, but have different size and color with the bulge being smaller and redder than the disk. Since the size of the PSF is color dependent, the distortion caused by the PSF on the two components is different. We assume here that the PSF is smaller for longer wavelengths (as expected for the LSST). Therefore the bulge PSF (red circle) is smaller than the PSF acting on the disk (blue circle). The observed image will be the sum of the convolution of the two components with their respective PSFs. Since the disk PSF is larger, it will have a stronger "rounding" effect on the disk component in comparison to the smaller bulge PSF, making the disk rounder than the bulge in the observed image. The effective SED is an average of the disk and bulge SED, resulting in the effective PSF having a size in between the bulge and disk PSFs.

As before, the weight function is the yellow-dotted circle. Since the weight function effectively gives more importance to the center, the bulge will be weighted more than the disk in the shape measurement. The correction PSF (effective PSF) is larger than the actual PSF that acted on the bulge. The PSF correction step will thus overestimate the ellipticity of the galaxy.

The extent of this error in the measured shape depends on the intrinsic color gradient of the galaxy. Color gradient depends on galaxy morphology, which is correlated with redshift (La Barbera et al., 2010; Kennedy et al., 2016). Therefore, color gradients can lead to redshift-dependent shape measurement biases.

2.3 Isolating and quantifying CG bias

Shape measurement algorithms can lead to a range of different shape measurement errors that depend on noise, galaxy properties, etc.; see Mandelbaum et al. (2018). To isolate color gradient bias, we use a technique described in S13: compare the shear measured from a galaxy with CG to that of an "equivalent" galaxy with no CG.

The technique is illustrated in Figure 2.3. We create a pair of galaxies – one galaxy with a spatially dependent SED (on the left), and an equivalent galaxy with uniform SED (on the right). The SED and surface brightness profile (SBP) of the equivalent galaxy is chosen so that the two galaxies appear identical when convolved with the same chromatic PSF and observed through the same filter (top row). This leads to all non-CG biases being the same for the two galaxies. As illustrated in the bottom row in the figure, if we apply the same shear to each galaxy and convolve it with the same chromatic PSF, the measured shear is no longer the same. We quantify the CG shear bias as this difference between the measured shear estimators (ellipticities). More specifically, in the absence of shear although the observed image of the galaxy with CG and the equivalent galaxy are the same, the SBP for the two galaxies are not identical. Therefore, the response to the same shear g will be different for the two galaxies, causing the PSF-convolved images and the shear measured from the two observed images to be different. The difference in the two shear values ($\hat{g}_{\mathrm{CG}}$ and $\hat{g}_{\mathrm{no\ CG}}$) is an estimate of the CG bias.

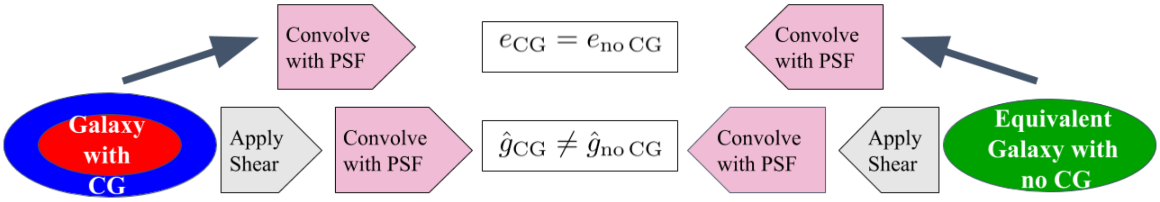


Figure 2.3: Illustration of technique for isolating shear measurement bias due to only color gradients (CG). Top row: The PSF-convolved "equivalent" galaxy with no CG (right) has, by construction, the same shape as the PSF-convolved galaxy with CG (left). Bottom row: In the presence of an applied shear, the measured shapes are no longer equal, isolating CG shear bias.

2.3.1 Isolating shear bias due to CG

When a small shear g is applied to a galaxy with no CG, the measured shear, $\hat{g}_{\text{no CG}}$, can be approximated as

$$\hat{g}_{\text{no CG}} = (1 + m_o)g + c_o, \tag{2.5}$$

where m_o and c_o are the multiplicative and additive bias terms. We assume that when the same small shear g is applied to a galaxy with CG, the bias on the measured shear, $\hat{g}_{\text{CG}}$, due to color gradients can be encapsulated by two new terms m_{CG} and c_{CG}:

$$\hat{g}_{\text{CG}} = (1 + m_o + m_{\text{CG}})g + (c_o + c_{\text{CG}}). \tag{2.6}$$

If the galaxy with no CG is the equivalent galaxy described earlier then taking the difference of the two measured shears will remove all other systematic bias contributions isolating contributions from color gradients only:

$$\Delta g = \hat{g}_{\text{CG}} - \hat{g}_{\text{no CG}} = m_{\text{CG}} g + c_{\text{CG}}. \tag{2.7}$$

2.3.2 Creating equivalent galaxy with no color gradient

The method described in Section 2.3.1 for isolating CG bias assumes we can create an equivalent galaxy with no CG with the properties shown in Figure 2.3. We describe below a method adapted from Section 2 of S13 to create an equivalent galaxy with no CG.

In the absence of color gradients, the galaxy SBP can be factored into a product of spatial and spectral components $a(\vec{x})$ and $S(\lambda)$:

$$f(\vec{x}, \lambda) = a(\vec{x})\, S(\lambda). \tag{2.8}$$

The true image of the galaxy, $I^o(\vec{x})$, through the transmission function $T(\lambda)$ can be written as

$$I^o(\vec{x}) = \int f(\vec{x}, \lambda)\, T(\lambda)\, d\lambda. \tag{2.9}$$

However, as shown in Equation 2.2, the observed galaxy is a convolution of the true galaxy and the PSF, in the observed band:

$$I^{\text{obs}}(\vec{x}) = \int f(\vec{x}, \lambda) * \Pi(\vec{x}, \lambda)\, T(\lambda)\, d\lambda. \tag{2.10}$$

In Fourier space, convolution becomes multiplication and the expression for the observed galaxy image becomes

$$\begin{aligned}\tilde{I}^{\mathrm{obs}}(\vec{k}) &= \int \tilde{a}(\vec{k})\, S(\lambda)\, \tilde{\Pi}(\vec{k},\lambda)\, T(\lambda)\, d\lambda \\ &= \tilde{a}(\vec{k}) \int S(\lambda)\, \tilde{\Pi}(\vec{k},\lambda)\, T(\lambda)\, d\lambda, \qquad (2.11)\end{aligned}$$

while the unconvolved image in Fourier space is

$$\begin{aligned}\tilde{I}^{\mathrm{o}}(\vec{k}) &= \int \tilde{f}(\vec{k},\lambda)\, T(\lambda)\, d\lambda \\ &= \int \tilde{a}(\vec{k})\, S(\lambda)\, T(\lambda)\, d\lambda. \qquad (2.12)\end{aligned}$$

Solving Equation 2.11 for $\tilde{a}(\vec{k})$ and substituting in Equation 2.12 we get

$$\tilde{I}^{o}(\vec{k}) = \int \frac{\tilde{I}^{\mathrm{obs}}(\vec{k})}{\int S(\lambda)\, \tilde{\Pi}(\vec{k},\lambda)\, T(\lambda) d\lambda} S(\lambda)\, T(\lambda)\, d\lambda. \qquad (2.13)$$

The denominator is the effective wavelength-independent PSF, $\Pi^{\mathrm{eff}}(\vec{x})$, computed in Equation 2.4, where the effective SED is identical to the uniform galaxy SED. Thus, the SBP of the galaxy at a particular wavelength λ_{ref} can be written as

$$\tilde{f}^{o}(\vec{k},\lambda_{\mathrm{ref}}) = \frac{S(\lambda_{\mathrm{ref}})\, \tilde{I}^{\mathrm{obs}}(\vec{k})}{\tilde{\Pi}^{\mathrm{eff}}(\vec{k})}. \qquad (2.14)$$

In other words, in the absence of color gradients the galaxy SBP at a given wavelength is the PSF convolved galaxy image, deconvolved by the effective PSF.

We now extend the above analysis to a galaxy with CG. We approximate the SBP of the equivalent galaxy as Equation 2.14 applied to the observed image of a galaxy with CG with the same effective SED but no color gradient. Thus the SBP of the effective galaxy with no CG is

$$\tilde{f}_{\mathrm{no\,CG}}(\vec{k},\lambda) = \frac{S_{\mathrm{eff}}(\lambda) \tilde{I}^{\mathrm{obs}}_{\mathrm{CG}}(\vec{k})}{\tilde{\Pi}_{\mathrm{eff}}(\vec{k})}, \qquad (2.15)$$

where $S_{\mathrm{eff}}(\lambda)$ is the effective SED computed in Equation 2.3. In the absence of an applied shear, the observed image corresponding to a galaxy with SBP of Equation 2.15 is identical to $\tilde{I}^{\mathrm{obs}}_{\mathrm{CG}}(\vec{k})$. A flowchart illustrating the entire method is also shown in Figure 2.4.

We use the modular galaxy simulation toolkit `GalSim`[3] (Rowe et al., 2015) to perform the integrations, convolutions and deconvolutions described above to simulate both the galaxy with CG and the equivalent galaxy with no CG. While the galaxy image with CG, $I^{\mathrm{obs}}_{\mathrm{CG}}(\vec{x})$ in Equation 2.2, is

---

[3] https://github.com/GalSim-developers/GalSim

drawn[4] with the LSST pixel scale, the images in Fourier space, $\tilde{I}^{\mathrm{obs}}_{\mathrm{CG}}(\vec{k})$ and $\tilde{\Pi}_{\mathrm{eff}}(\vec{k})$ in Equation 2.15, are drawn with four times the resolution in order to reduce errors produced by finite pixel size during deconvolution.

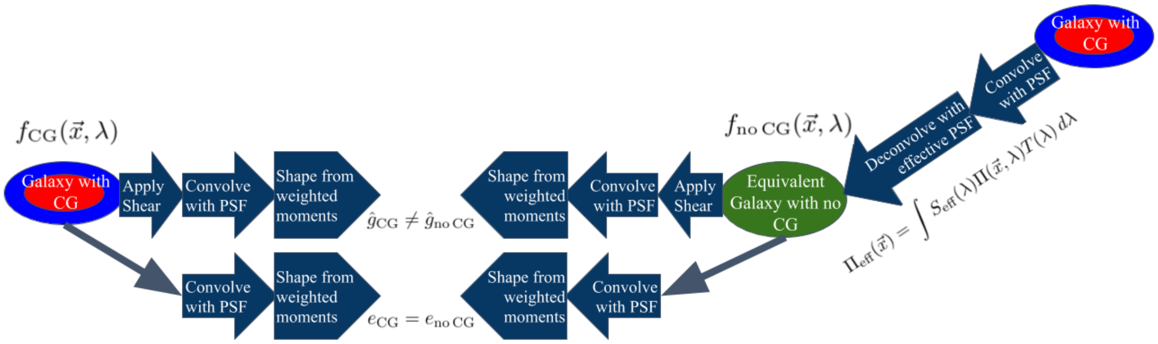


Figure 2.4: Schematic showing the complete process of creating an equivalent galaxy with no CG (green ellipse) from a galaxy with CG (multi-colored ellipse). See Section 2.3.2 in text for details.

2.3.3 Moment-based shape estimation with `GalSim`

`GalSim` is used to perform shape estimation, in addition to the image simulation described above. The shapes are computed with "adaptive" moments that use an elliptical Gaussian weight function, the shape of which is matched to that of the PSF-convolved galaxy image. The measured moments of the galaxy are corrected for the PSF to produce the ellipticity of the intrinsic galaxy image (Bernstein & Jarvis, 2002).

For most of this analysis, we use the `GalSim` implementation of the "re-Gaussianization" algorithm (`REGAUSS`), which treats the deviations of the PSF from Gaussianity perturbatively in real space, without first measuring moments (Hirata & Seljak, 2003).

We found that, when compared to other shape estimation algorithms in `GalSim`, the `REGAUSS` algorithm makes more accurate shape measurements. The error in estimated shape is approximately 10^{-6} for a simulated Gaussian galaxy with a Gaussian PSF drawn with LSST pixel scale; the bias is larger for more complex situations.

Our method (described above) of using an equivalent galaxy without color gradients to isolate the contribution from color gradients reduces the impact of the shape estimation algorithm on the estimated CG bias. For the reference galaxy simulated in a redshift range of 0-1.2, we estimate the CG bias using three other adaptive moment estimation and PSF correction algorithms implemented in GalSim: `KSB`[5] (Kaiser, Squires, & Broadhurst, 1995), `LINEAR` (Hirata & Seljak, 2003) and `BJ` (Bernstein & Jarvis, 2002). The difference in CG bias estimates for `REGAUSS`, `KSB` and `BJ` is less than

---

[4] Draw refers to the `GalSim` `drawImage` function, which draws an object profile onto an image.

[5] The `GalSim` `KSB` algorithm is a specific implementation of the KSB method (Kaiser et al., 1995; Luppino & Kaiser, 1997), as described in Appendix C of Hirata & Seljak (2003).

1×10^{-4} for all redshifts. For `LINEAR` and `REGAUSS`, the difference is also less than 1×10^{-4}, except for $z \leq 0.2$, when the difference in bias is just under 4×10^{-4}.

As discussed in Section 2.2.3, the weight function used in the shape measurement has a very significant effect on CG bias. Therefore, it is important to study the effect of the size of the weight function on the measured CG bias. The implementation of `REGAUSS` in `GalSim` does not have the provision to switch off or fix the size of the weight function computed with adaptive moments. Therefore, to study the impact of the weight function size, we use the `GalSim` implementation of the `KSB` shape measurement algorithm where the size of the circular weight function can be fixed by the user.

While we found that `REGAUSS` robustly yields correct results for noise-free parametric simulations and real galaxies with high signal-to-noise ratio (SNR)[6], we found a high failure rate for low SNR galaxies. This is because the adaptive moment algorithm is not robust to noise and therefore does not always converge to a solution. This resulted in elimination of a significant fraction of small, faint galaxies from our analysis sample. Although these galaxies are not likely to be used in weak lensing analyses, we recommend that shape estimation methods that are more robust to noise be used in future CG bias studies.

2.4 Estimating CG bias with parametric galaxies

We first evaluate the impact of color gradients on shear measurements with simulated parametric galaxies. Since the galaxies can be simulated with no noise, this enables us to isolate the effect of CG bias without being contaminated by noise bias.

2.4.1 Observing conditions

The galaxy images are simulated in the two main LSST lensing bands, r and i, with the measurements performed in the r band, unless otherwise noted. In Figure 2.5, we show the transmission functions for the LSST r and i bands[7], as well as the HST V and I bands used in the analysis of AEGIS galaxies, described below.

As described in Equation 2.1, we model the LSST PSF as a circular Gaussian profile with a wavelength-dependent size that is seeing limited with scaling exponent $\alpha = -0.2$. This value is valid for purely Kolmogorov turbulence – i.e., an infinite outer scale – in the limit of long exposure times. For this analysis we set the PSF size to $\sigma^o = 0.297$ arcsec (FWHM=0.7 arcsec) at $\lambda^o = 550\,\mathrm{nm}$. This value was chosen to match the expected LSST median zenith seeing in the r band (Ivezić et al., 2019). We study the impact of the scaling exponent and PSF size on measured bias. We limit the scope

---

[6] We use the same definition of SNR as that used in the GREAT3 challenge (Mandelbaum et al., 2014), where SNR is defined as the ratio of measured flux to uncertainty on the flux within an elliptical Gaussian aperture matched to the size and shape of the PSF-convolved galaxy image.

[7] https://github.com/lsst/throughputs/tree/master/baseline

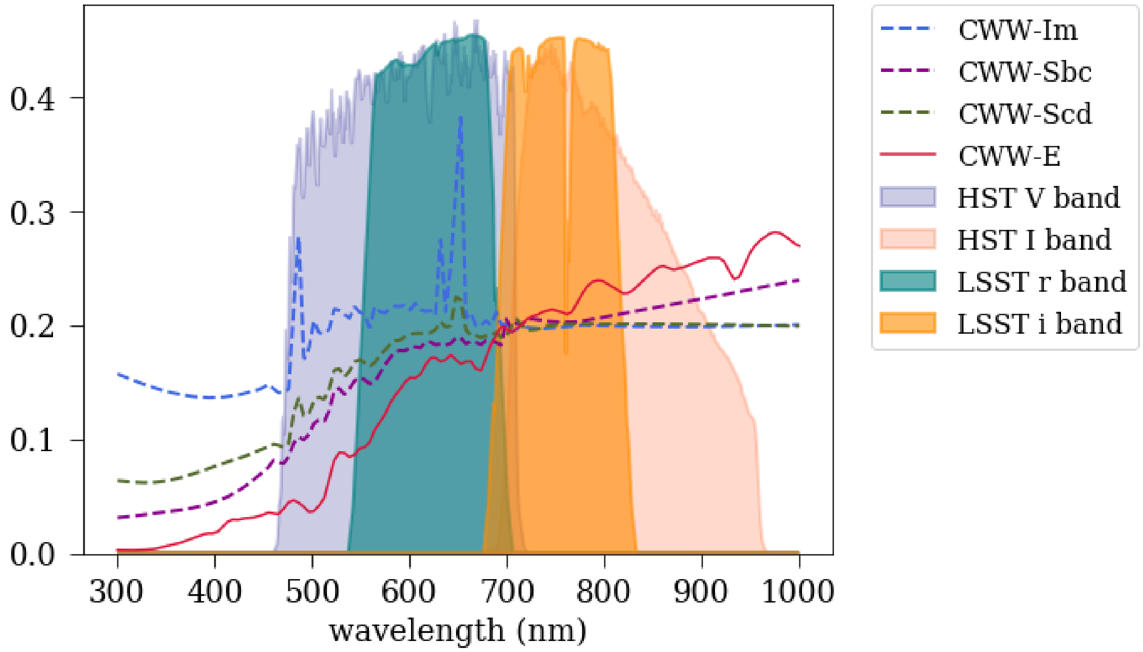


Figure 2.5: Transmission probabilities for LSST r and i bands and HST V and I bands. Also shown are three types of SEDs: CWW-Im, CWW-Sbc, and CWW-Scd spectra (dashed lines), assigned as disk SEDs in the CG analysis of the reference galaxy, and CWW-E spectrum (solid line), assigned as the bulge SED. The SEDs shown here have been normalized to 0.2 at 550 nm and redshifted to 0.3.

of this study to only wavelength-dependent PSF size and do not include any wavelength-dependent PSF ellipticity effects, such as atmospheric differential chromatic refraction (DCR).

2.4.2 Parametric galaxy with color gradients

We model galaxies with CG as concentric bulge + disk components with different SEDs, with the following SBP:

$$f_{\rm CG}(\vec{x},\lambda) = a_b(\vec{x})S_b(\lambda) + a_d(\vec{x})S_d(\lambda), \tag{2.16}$$

where $a_{b,d}(\vec{x})$ are the spatial profiles and $S_{b,d}(\lambda)$ the SEDs of the bulge and disk components denoted by the subscripts b and d, respectively. We use this parameterization of galaxies with CG for two different samples: 1) the reference galaxy with extreme color gradients and 2) the CatSim catalog of galaxies.

2.4.3 Reference parametric bulge + disk galaxy with extreme color gradients

We first demonstrate the presence of CG bias for a simulated reference galaxy with extreme color gradients: a superposition of a small red bulge and an extended blue disk. The parameters of the galaxy are chosen to be identical to the "B" type galaxy in S13, which allows us to compare results for the same galaxy model. Indeed, we find that our estimates of CG shear bias for the reference galaxy with a Euclid PSF and bandpass are in agreement with the values estimated in S13.

2.4.3.1 Parameters for simulating extreme-CG galaxy

The spatial components of the bulge and disk SBP, $a_b(\vec{x})$ and $a_d(\vec{x})$, are modeled as Sersic profiles (Sérsic, 1963) with indices $n_s = 1.5$ [8] and 1.0 for the bulge and disk, respectively. The bulge is compact with a half-light radius (HLR) of 0.17 arcsec. The spatial component of the bulge accounts for 25% of the total galaxy flux. The extended disk is much larger with HLR of 1.2 arcsec. Both, the bulge and disk, have the same centroid positions and ellipticities, $e^{\text{intr}} = (0.3, 0.3)$.

The SEDs of the bulge and disk are shown in Figure 2.5 with the redder bulge assigned a CWW-E spectrum (solid red line) and the disk a CWW-Im spectrum (dashed blue line), which has strong emission lines. CWW here refers to template SEDs created from spectroscopic data of nearby galaxies of different types measured by Coleman et al. (1980) and extended below 1400 Å and beyond 10000 Å by Bolzonella et al. (2000) using evolutionary models of Bruzual A. & Charlot (1993). To further investigate the impact of intrinsic galaxy color gradients, we measure the CG bias for the reference galaxy with extreme color gradients at different redshifts and for two additional disk SEDs: CWW-Sbc and CWW-Scd (dashed magenta and green lines, respectively). The SEDs in Figure 2.5 are normalized to 0.2 at 550 nm in the rest frame and then redshifted to 0.3. The bulge and disk parameters are summarized in Table 2.1.

Table 2.1: Parameters of simulated galaxies for bulge/disk components .

Name	Sersic index	SED	HLR (arcsec)	Flux proportion (%)
Reference galaxy with extreme color gradients	1.5/1	E/(Im, Scd, Sbc)	0.17/1.2	25/75
CatSim galaxies (mean)	4/1	Template SEDs	0.2/0.4	20/80

Note. Values correspond to bulge/disk components of the reference galaxy and CatSim galaxies. The entries for the CatSim galaxies are the mean bulge and disk parameters of the entire sample. The CatSim SEDs are drawn from an ensemble of template SEDs.

---

[8] We chose the bulge Sersic index to match the value in S13, rather than using the more conventional value $n_s = 4$. We find that the predicted CG shear bias m_{CG} is up to 26% smaller in magnitude when the reference galaxy bulge Sersic index is taken to be 4 rather than 1.5.

2.4.3.2 Verification of method to isolate CG bias

We verify our proposed method for isolating CG bias described in Section 2.3 by simulating an equivalent galaxy with no CG, for the reference galaxy, and then applying the same shear and PSF to both the original galaxy and the equivalent galaxy. The shears recovered with a ring test for each galaxy are compared. If this difference in shears, Δg, shows a linear dependence on the applied shear then we have confirmation that the method described in Section 2.3.1 is able to isolate CG bias. The slope and intercept of the linear model, $\Delta g = m_{\mathrm{CG}} g + c_{\mathrm{CG}}$, correspond to multiplicative and additive bias, respectively, from color gradients only.

The galaxy shapes are estimated with the `REGAUSS` method of `GalSim`. Since the bias values are found to be similar for the two shear components g_1 and g_2, we report results here for only g_1.

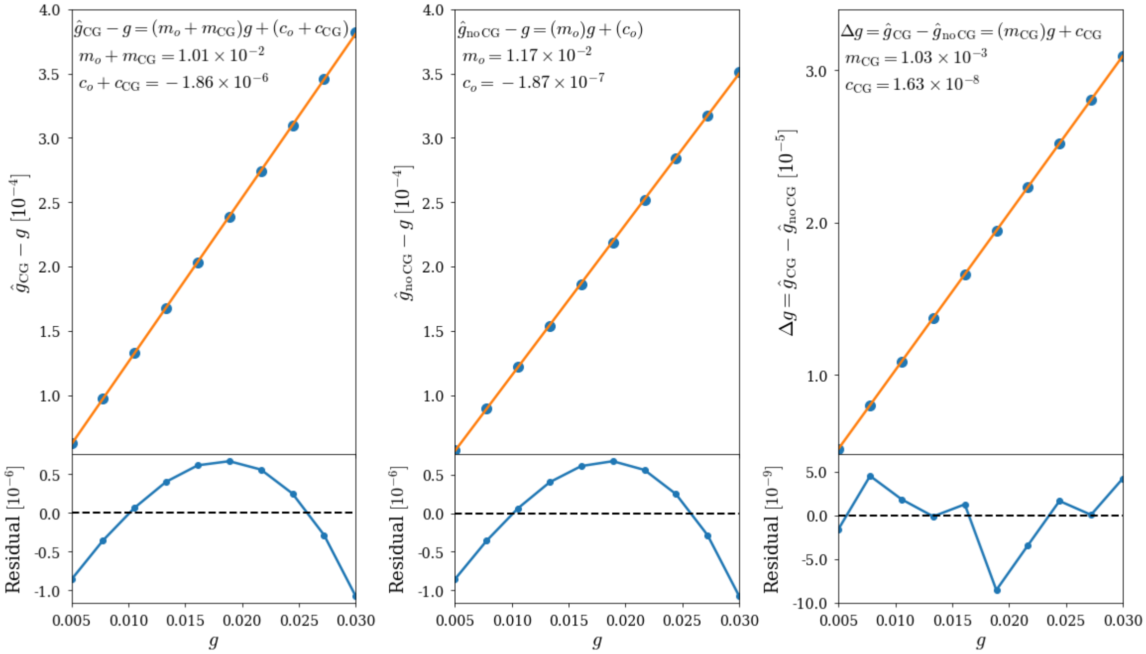


Figure 2.6: Tests of the validity of the definition of multiplicative and additive shear biases m_{CG} and c_{CG} (Section 2.3.1) for simulated LSST galaxies. The horizontal axes correspond to applied shear g. The vertical axis in the top panels corresponds to (left) the difference between measured and applied shear for a galaxy with CG, (center) the difference between measured and applied shear for the equivalent galaxy with no CG, and (right) the difference between measured shears for galaxies with and without CG. The blue dots correspond to these differences and the orange lines to linear fits. The values of the multiplicative and additive biases from the fits are shown near the top of each panel. In the bottom plot of each panel, we show the difference between the linear fit and the data points.

In the top left and center panels of Figure 2.6, we plot as a function of applied shear g the difference between measured shear $\hat{g}_{\mathrm{CG}}$ and applied shear for a galaxy with CG (left), and the difference between measured shear $\hat{g}_{\text{no CG}}$ and applied shear for the equivalent galaxy with no CG (center). The blue dots correspond to these differences and the red line corresponds to the result of a linear fit to these values. The bottom plot in each panel corresponds to the difference between the linear fit and the points. We see that for this range of applied shear ($g \leq 0.03$), the measured shear is linearly dependent on the applied shear. The multiplicative and additive biases, corresponding to the slope ($m = m_o + m_{rmCG}$) and intercept ($c = c_o + c_{rmCG}$) of the fits, are found to be $m = 1.01 \times 10^{-2}$ and $c = -1.86 \times 10^{-7}$ for the original galaxy with CG, and $m_o = 1.17 \times 10^{-2}$ and $c_o = -1.87 \times 10^{-7}$ for the equivalent galaxy with no CG.

We proposed in Section 2.3.1 that m_o includes the shear bias from all sources except CG and that the bias from CG can be isolated as $m_{rmCG} = m - m_o$. We test this hypothesis by plotting in the right panel of Figure 2.6 the difference $\Delta g = \hat{g}_{\text{CG}} - \hat{g}_{\text{no CG}}$ as a function of applied shear. The small differences between the linear fit and the points, shown in the bottom right panel, validate the linear relationship assumed in Equation 2.7. The slope of the linear fit is an estimate of bias from only CG: $m_{\text{CG}} = 1.03 \times 10^{-3}$. The sign of m_{CG} is positive; i.e., the measured shear is larger than the true shear, in agreement with the example in Section 2.2.4.

We now apply this method to study the dependence of CG bias on intrinsic color gradient and observing conditions. Due to the small magnitudes of the additive bias in our simulations, which is expected for a circular PSF with circularly symmetric wavelength-dependence, we focus our analysis on studying the impact of color gradients on only the shear multiplicative bias.

2.4.3.3 Results of CG bias measurements on extreme-CG galaxy

To illustrate the impact of intrinsic galaxy color gradients on CG bias we estimate the CG bias for the reference galaxy with extreme color gradients. The estimated bias values here are not representative of the bias expected for shear measurements from all galaxies seen by LSST; rather, we expect these bias values to be outliers for a more representative sample.

We plot in Figure 2.7 the estimated value of m_{CG} for three different disk SEDs, with the same bulge SED (E), for different redshifts. The CG biases are estimated from the shear recovered when a constant shear of $g = (0.01, 0.01)$ is applied to the galaxy. The bias is larger for the Im disk SED, which has emission lines. Since CG bias arises due to the difference in bulge and disk color in the observing bandpass, galaxies with the same bulge and disk SEDs can have different CG bias at different redshifts depending on their SED profiles. The yellow dashed lines show the LSST DESC requirement of 0.003 on the total systematic uncertainty in the redshift-dependent shear calibration. The estimated CG bias for all three disk SEDs and redshifts are observed to be lower than the LSST requirement, with the mean CG bias at all redshifts and disk SEDs being $m_{\text{CG}} = 8.65 \times 10^{-4}$.

In S13, the magnitude of the CG bias for the reference galaxy in the same redshift range, with Euclid's expected PSF size and wavelength dependence, was estimated to be in the range $4 \times 10^{-4} - 1.5 \times 10^{-3}$, which is not dissimilar to our predictions for LSST for this extreme case. The observing bandpass for Euclid covers a wider wavelength range (550nm-900nm) than those for LSST and the slope of the PSF chromaticity is steeper, leading to higher sensitivity to color gradients in Euclid. However, the Euclid PSF is smaller (FWHM=0.15 arcsec) than the LSST PSF, offsetting the other effects.

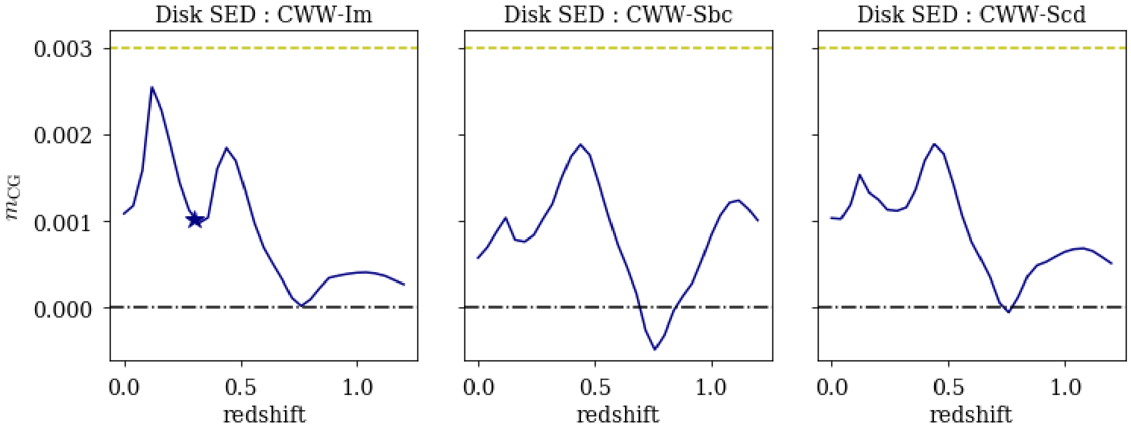


Figure 2.7: Multiplicative shear bias due to CG, m_{CG}, as a function of redshift for a reference galaxy with CWW-E bulge SED and three different disk SEDs: CWW-Im, CWW-Sbc, CWW-Scd (left to right). The star denotes the CG bias for the galaxy with CWW-Im disk SED at redshift of 0.3.

2.4.3.4 Dependence on weight function

As described in Section 2.2.3, CG bias occurs in the presence of a weight function that assigns more weight to certain parts of the galaxy image. In the absence of a weight function, we expect the CG bias to be zero (Semboloni et al., 2013). To study the effect of our choice of weight function, we estimate CG bias while varying the size of the circular Gaussian weight function using the KSB method for galaxy shape measurement in `GalSim`. The results of the analysis for the reference galaxy with Im disk SED at redshift 0.3 is shown in Figure 2.8. The values on the horizontal axis correspond to the ratio of the HLR of the weight function to the bulge+disk galaxy HLR (0.949 arcsec). As expected, the CG bias decreases with increasing size of the weight function with the bias approaching zero for large weight sizes. The yellow circle denotes the CG bias for a Gaussian weight function with HLR equal to the size of the galaxy as used in S13. For comparison we show that the CG bias computed with REGAUSS for the reference galaxy (blue star) is larger. This is because the `GalSim` REGAUSS algorithm uses adaptive moments to match an elliptical Gaussian weight to the galaxy image. Since the reference galaxy is the sum of two Sersics, the Gaussian profile matched by REGAUSS has an HLR = 0.7 arcsec, smaller than the PSF convolved galaxy HLR of 0.949 arcsec. Thus this analysis uses a smaller weight function than in S13, leading to larger biases.

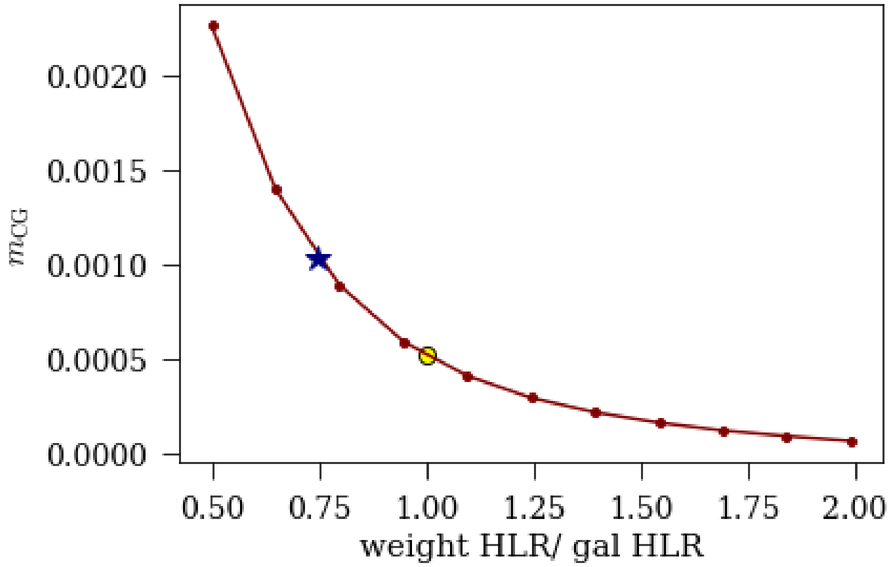


Figure 2.8: Multiplicative shear bias from CG, $m_{\rm CG}$, as a function of the ratio of weight function size to galaxy size (red dots) with shape measurements performed using the `GalSim` implementation of KSB shape estimation algorithm. We show the value of the CG bias when the weight function is computed with adaptive moments (blue star) and when the weight function HLR is equal to the galaxy HLR (yellow circle).

2.4.3.5 Dependence on PSF

We illustrate the effect of the PSF chromaticity on CG bias in Figure 2.9. The left panel shows the dependence on the PSF size scaling exponent α. A scaling exponent of zero is an achromatic PSF and thus has no CG bias. The sign of the exponent determines the sign of the bias, with bias being positive for a negative exponent (ground-based survey) and negative for positive exponents (space-based surveys). The CG bias is highly sensitive to the chromatic nature of the PSF, with a change of 0.1 in the value of α changing the CG bias by 20% of the LSST requirement. The panel on the right in Figure 2.9 shows the dependence of CG bias on the size of the PSF. For small PSF sizes, the image distortions and PSF corrections are small, resulting in smaller CG bias. When the PSF is larger than the bulge ($\sigma \approx 0.15$ arcsec), the CG bias comes into play and $m_{\rm CG}$ remains constant even at larger PSF sizes. The blue star is the same as in Figure 2.7 and shows the bias for the reference galaxy PSF with $\alpha = -0.2$ and $\sigma^o = 0.297$ arcsec at $\lambda^o = 550\,$nm.

We also found that the predicted CG shear bias $m_{\rm CG}$ was up to 20% smaller when the reference galaxy is convolved with a Kolmogorov PSF, rather than a Gaussian PSF, with the same full width at half maximum (FWHM) value and the same wavelength-dependent scaling exponent. Since the aim of this analysis is to place conservative upper limits on CG bias estimates, we use the nominal Gaussian PSF for the rest of the analysis.

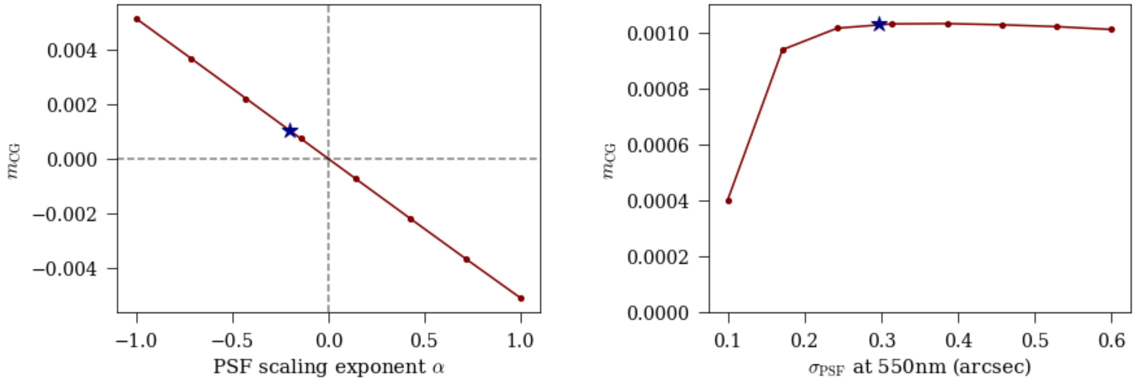


Figure 2.9: Left: Multiplicative shear bias from CG, $m_{\rm CG}$, as a function of PSF chromatic scaling exponent α. Right: $m_{\rm CG}$ as a function of PSF size. The stars correspond to the PSF parameters used in the reference galaxy analysis: $\alpha = 0.2$ and $\sigma^o = 0.297$ arcsec at $\lambda^o = 550$ nm.

2.4.4 CatSim catalog of parametric galaxies

Although the CG bias for the reference galaxy computed above is small, it is not negligible in comparison to the LSST requirement for *all* systematic effects. Since we also observe a large dependence of CG bias on the galaxy intrinsic properties, CG bias cannot be outright ignored. Therefore, to more accurately estimate the expected CG bias for LSST, we calculate the bias using a catalog of galaxy characteristics (redshift, size, magnitude and shape) expected in a 10-year LSST weak lensing dataset.

We use a simulated catalog prepared by the LSST Catalog Simulator, CatSim (Connolly et al., 2014), containing astrophysical sources with properties that are representative of what the LSST will observe to its coadded depth. The catalog covers a 4.5 × 4.5 degree footprint on the sky with realistic galaxy morphologies, apparent colors and spatial distributions, and redshifts. The galaxy simulation is based on dark matter halos from the Millennium Simulation (Springel et al., 2005) and a semi-analytic baryon model described in De Lucia et al. (2006). Galaxy morphologies are modeled using two Sersic profiles, where the bulge and disk components have Sersic indices 4 and 1, respectively. For all sources, an SED is fit to the galaxy colors using a sophisticated fitting algorithm, independently for the bulge and disk, that includes inclination-dependent reddening. The catalog contains galaxies with r-band AB magnitudes brighter than 28.

2.4.4.1 Parameters of CatSim galaxies

For our analysis, we select galaxies from 858,502 galaxies in the one-square-degree CatSim catalog. The selected galaxies satisfy the following criteria, where the number in parentheses is the number of galaxies satisfying the cumulative criteria:

1. Composed of bulge and disk components – i.e., there is a non-zero color gradient (245,359).
2. Redshift less than 1.2 (76,201).
3. i-band magnitude < 25.3 (58,310).
4. Bulge and disk semi-major axis each less than 3 arcseconds (≈ 15 pixels)[9] (57,943).

The `GalSim` `REGAUSS` shape measurement algorithm fails for a fraction of galaxies with CG or their equivalent galaxies with no CG. While the CG bias estimates shown in this section are for noise-free CatSim simulations, we also study the impact of Poisson noise in the image on our CG bias estimates, as explained in detail in Section 2.5.2 and Appendix A.2. As described earlier in Section 2.3.3, `GalSim` shape estimation methods have a high failure rate for small noisy galaxy images. We exclude such galaxies from our final sample if the shape measurement failed for any of the six galaxy shape measurements in the ring test of the galaxy with CG or its equivalent galaxy with no CG. Our final results are based on 45,534 CatSim galaxies where the shape measurement was always successful.

In Figure 2.10 we show the distribution of the intrinsic parameters of the selected galaxies. The top-left panel shows a histogram of the galaxy AB magnitudes in the i and r bands. With the criteria being applied to select galaxies in the "Gold Sample" ($i < 25.3$) (LSST Science Collaboration, 2009), the sample includes galaxies with r-band magnitudes up to 26.5. The top-right plot shows the distribution of galaxy redshifts. The bottom-left panel shows the distribution of bulge and disk sizes, with mean HLR of 0.2 and 0.4 arcsec, respectively. A large fraction of the selected galaxies are small, with a large number of galaxies with HLR comparable to an LSST pixel (0.2 arcsec). Most galaxy SBPs are composed of a small compact bulge and an extended disk. The CatSim galaxy bulges have a steeper profile in the center when compared to the reference galaxy, while the disks are smaller on average (see Table 2.1). The distribution of the difference between the disk and bulge magnitudes is shown in the bottom-right panel. Most of the galaxy flux is contained in the disk, with the disk being on average 2.5 magnitudes brighter than the bulge.

The galaxies selected here are not completely representative of the galaxies in the LSST lensing sample. However, the criteria are conservative in that they are not expected to reduce the estimated CG bias on LSST shear estimates. While the criteria could potentially introduce selection bias, the focus of this study is to isolate the impact of only color gradients.

2.4.4.2 Results of CG shear bias measurements for CatSim galaxies

For each galaxy in the CatSim catalog we draw an equivalent galaxy with no CG and measure the shear response when a shear of $g = (0.01, 0.01)$ is applied to both galaxies. A histogram of the estimated CG bias m_{CG} in the r and i bands is shown in Figure 2.11, with their means shown

---

[9] We impose this maximum size to avoid `GalSim` run-time errors caused by Fourier transform arrays being too large in the `ChromaticRealGalaxy` modeling described later in Section 2.5.1.

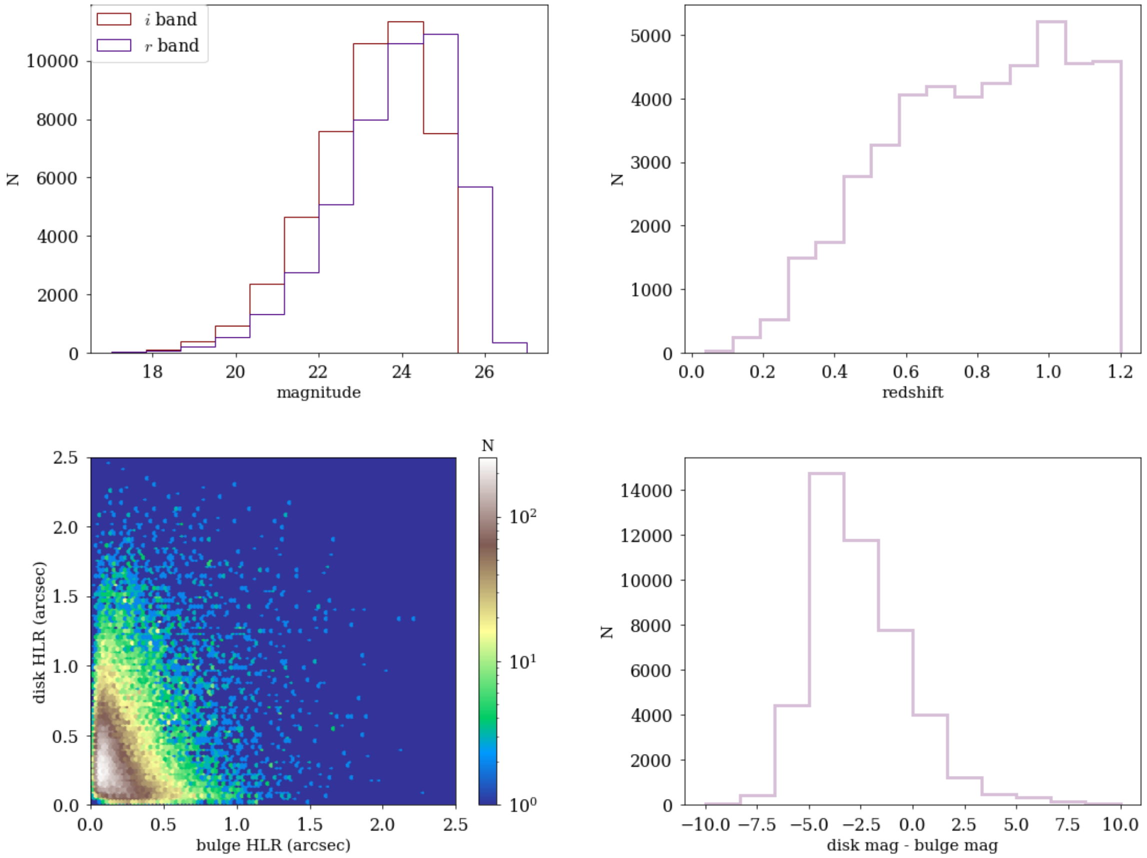


Figure 2.10: Distributions of galaxy parameters for selected CatSim galaxies. Top left: *i*- and *r*-band AB magnitudes. Top right: Galaxy redshifts. Bottom left: Half-light radii (in arc seconds) of the disk and bulge components. Bottom right: Difference between disk and bulge magnitudes.

as dashed vertical lines. While the means of the distributions are consistent with zero, there exist galaxies in the tails of the distribution with large bias. Approximately 5% of the galaxies have $|m_{\text{CG}}| \geq 10^{-3}$. For these galaxies, the bulge and disk have similar brightness, but different colors. The bulge-to-disk flux ratio of the entire sample is ≈ 0.1, compared to $\approx$0.4 for the galaxies in the tails and 0.33 for the extreme CG galaxies described in Section 2.4.3.1. The median difference in V-I color of bulge and disk components is $\approx$0.4 for the entire sample, compared to $\approx$0.7 for the galaxies in the tails. The mean and median multiplicative CG biases are $(2.69 \pm 0.03) \times 10^{-5}$ and 6.17×10^{-5} in r band, and $(3.72 \pm 0.01) \times 10^{-5}$ and 2.58×10^{-5} in i band – in each case, much smaller than LSST requirements.

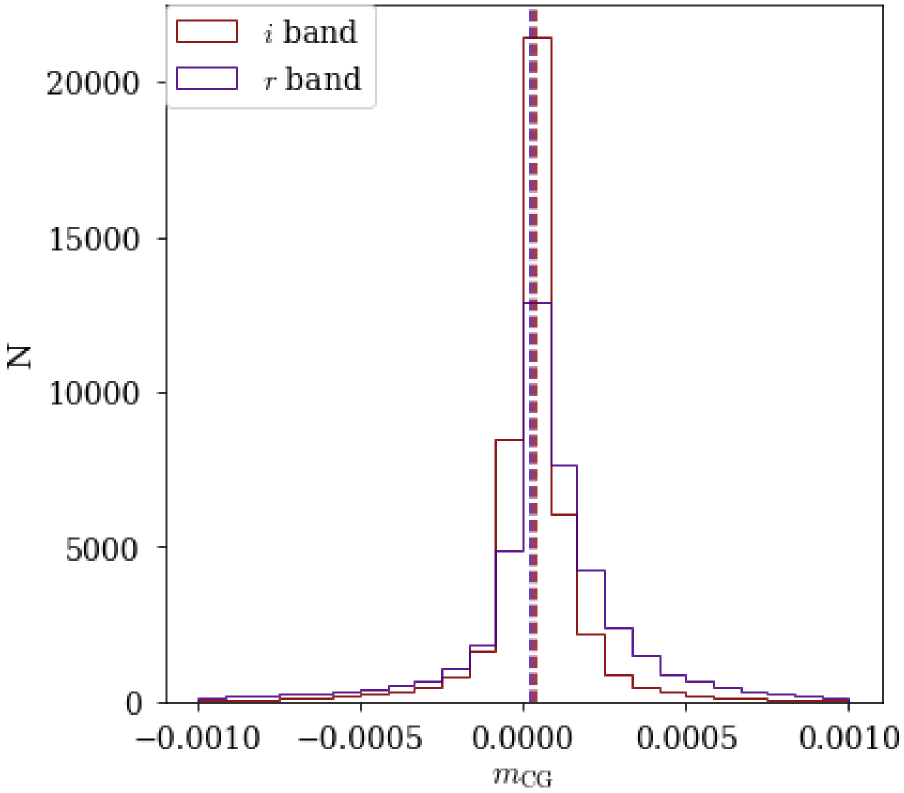


Figure 2.11: Histograms of CG multiplicative shear bias in the r and i bands for CatSim galaxies. The almost completely overlapping dashed vertical lines denote the mean in each band.

2.5 Estimating CG bias with chromatic real galaxies

To place more robust limits on the expected CG shear bias for LSST we estimate CG bias for a large sample of realistic galaxies with color gradients. We use galaxy images from the Hubble Space Telescope (HST) to obtain realistic galaxy morphologies and then draw them as seen by LSST using the `ChromaticRealGalaxy` module in `GalSim`.

In a separate note (Kamath et al., 2019), we describe the data reduction process used to produce postage-stamp images of 27k galaxies observed by the All-Wavelength Extended Groth strip International Survey (AEGIS) (Davis et al., 2007) in the HST V (F606W) and I (F814W) bands. Our goal is to use the V- and I-band images of galaxies in the AEGIS catalog to create their profiles in r and i LSST bands, while preserving their chromatic features. We can thus estimate the CG bias when these galaxies are imaged by LSST.

In Figure 2.12, we show an example of a pair of galaxies in the HST V and I bands of the AEGIS catalog, rendered by `ChromaticRealGalaxy` to simulate how they would appear in the LSST r and

i bands.

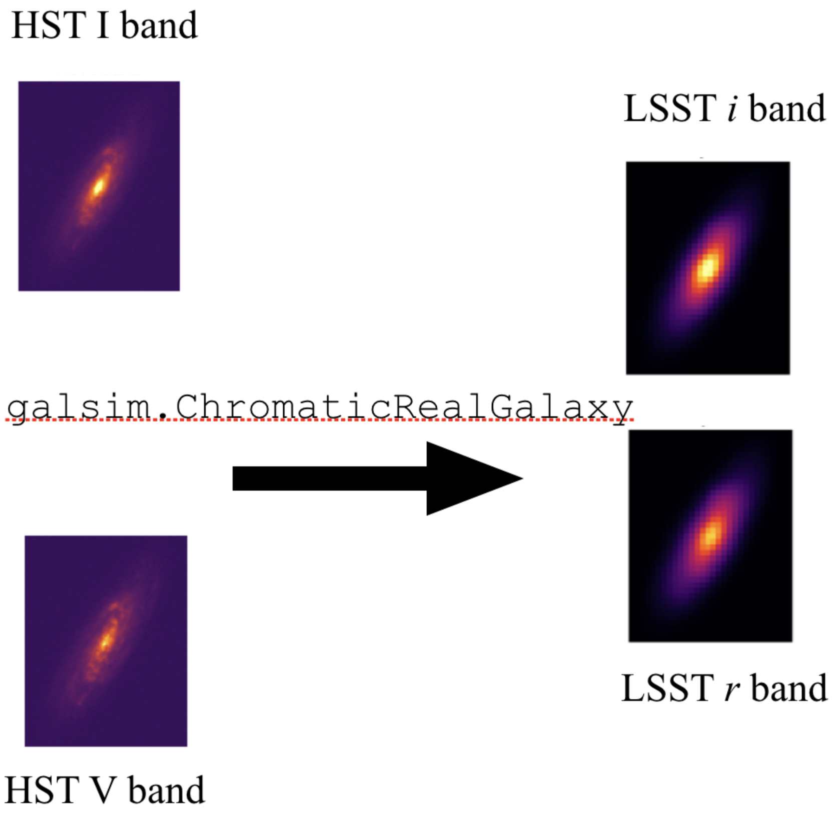


Figure 2.12: Example of HST V- and I-band galaxy images (on the left) and images produced with `ChromaticRealGalaxy` to simulate how they would appear in the LSST r and i bands (on the right).

2.5.1 `ChromaticRealGalaxy` model

`ChromaticRealGalaxy`[10] models multi-band images of individual galaxies as chromatic PSF convolutions (and integrations over wavelength) with a sum of profiles separable into spatial and spectral components. This decomposition can be thought of as a constrained chromatic deconvolution of the multi-band images by the associated PSFs.

The pre-convolution chromatic surface brightness profile of the galaxy, $f_{\mathrm{CG}}(\vec{x}, \lambda)$, is modeled as a sum of two or more separable chromatic surface brightness profiles, each with a particular asserted

---

[10] http://github.com/GalSim-developers/GalSim/blob/master/devel/modules/CGNotes.pdf

SED:

$$f_{\mathrm{CG}}(\vec{x},\lambda) \equiv \sum_j a_j(\vec{x}) S_j(\lambda), \tag{2.17}$$

where $S_j(\lambda)$ is the jth SED asserted as part of the decomposition, and $a_j(\vec{x})$ is the spatial component of the jth separable chromatic profile.

The observed image in the ith band is

$$\begin{aligned} I_i(\vec{x}) &= \int T_i(\lambda)\big[\Pi(\vec{x},\lambda) * f(\vec{x},\lambda)\big]\, d\lambda + \eta_i(\vec{x}) \\ &= \int T_i(\lambda) \sum_j S_j(\lambda)\,[\Pi(\vec{x},\lambda) * a_j(\vec{x})]\, d\lambda + \eta_i(\vec{x}), \end{aligned}$$

where $\eta_i(\vec{x})$ corresponds to (potentially spatially correlated) Gaussian noise in the ith band image. The noise $\eta_i(\vec{x})$ is related to the noise covariance function via $\langle \eta_i(\vec{x}_l)\eta_i(\vec{x}_m)\rangle = \xi_i(\vec{x}_l - \vec{x}_m)$, where angle brackets indicate averaging over realizations of the noise.

The convolution is easier to work with in Fourier space where it becomes a mode-by-mode product. The model in Fourier space is

$$\begin{aligned} \tilde{I}_i(\vec{k}) &= \int T_i(\lambda) \sum_j S_j(\lambda)\tilde{\Pi}(\vec{k},\lambda)\tilde{a}_j(\vec{k})\, d\lambda + \tilde{\eta}(\vec{k}) \\ &= \sum_j \big[\int T_i(\lambda)S_j(\lambda)\tilde{\Pi}(\vec{k},\lambda)\, d\lambda\big]\tilde{a}_j(\vec{k}) + \tilde{\eta}(\vec{k}) \\ &= \sum_j \tilde{\Pi}^{\mathrm{eff}}_{ij}(\vec{k})\tilde{a}_j(\vec{k}) + \tilde{\eta}(\vec{k}), \end{aligned} \tag{2.18}$$

where the effective PSF for the ith band and jth SED is

$$\tilde{\Pi}^{\mathrm{eff}}_{ij}(\vec{k}) = \int T_i(\lambda)S_j(\lambda)\tilde{\Pi}(\vec{k},\lambda)\, d\lambda. \tag{2.19}$$

`ChromaticRealGalaxy` solves for the (complex-valued) $\tilde{a}_j(\vec{k})$ while propagating the statistics of the noise.

2.5.2 Limitations of using real galaxy images

We use `ChromaticRealGalaxy` (CRG) to model real HST galaxies in an LSST observing band and then convolve the images with the expected LSST PSF and add the sky noise expected at 10-year LSST depth. We assume here that the observing conditions for HST and LSST are known perfectly. For real images where the SED of the galaxies are unknown, we approximate the SEDs of the CRG components as polynomials. Thus, the accuracy with which the CRG algorithm reproduces chromatic features depends on how different the bandpasses of input and target surveys are, and the assumed model of the position-dependent SEDs. Since we use only two images of the HST galaxy (V and I band), the SEDs are assumed to be polynomials of order 0 and 1 (linear SED).

The simulations of parametric galaxies analyzed in Section 2.4 are noise free. Bias from CG is computed as the difference in the shear estimator measured from a galaxy with CG and the equivalent galaxy with no CG. The assumption was that biases from all sources other then color gradients would act equally on both measurements and thus cancel in the difference, giving the bias from CG only. However, each real galaxy image includes noise in the signal in each pixel[11] (pixel noise) while the image of the equivalent galaxy with no CG has different noise in each pixel. This results in unequal noise bias in the shear measurements for the two galaxies. Therefore, the difference in shear measurements (Equation 2.7) no longer isolates bias from only CG; rather it is a combination of CG bias and residual noise bias.

Therefore, before using CRG to model real HST galaxies, we test CRG on simulated parametric galaxies where the truth is known. The detailed procedure used to test the impacts of the linear approximation for the SED and pixel noise while using CRG to model the reference galaxy and the CatSim galaxies is described in Appendix A. We summarize the results here.

1. For reference galaxy with extreme color gradients the redshift-averaged error in CG shear bias caused by modeling the galaxy SEDs as linear is $\mathcal{O}(10^{-4})$; however, the linear SED model tends to smooth the variation in bias with redshift, leading to a total variation in bias over the redshift range $[0, 1.2]$ that is a factor of ≈ 2 less for the linear SED compared to the true SED.

2. For simulated HST-like CatSim galaxies (modeled by CRG), the value of the CG bias estimate exhibits a strong dependence on SNR. The measured bias diverges from the noise-free CG bias estimate by $\mathcal{O}(10^{-3})$ for HST I-band SNR ≤ 100.

3. For simulated HST-like CatSim galaxies with I-band SNR > 200, the mean bias is $m_{\mathrm{CG}} = (-2.53 \pm 0.27) \times 10^{-4}$ in the r band and $m_{\mathrm{CG}} = (0.47 \pm 0.10) \times 10^{-4}$ in the i band. The mean CG bias for the noise-free parametric simulations of the same sample of galaxies is $m_{\mathrm{CG}} = (-0.02 \pm 0.03) \times 10^{-4}$ in the r band and $m_{\mathrm{CG}} = (0.25 \pm 0.02) \times 10^{-4}$ in the i band.

Thus, by using noisy images of CatSim galaxies, we find that our method of using an "equivalent galaxy" to minimize the impact of effects other than color gradients does not completely eliminate the impact of "pixel noise" for smaller values of SNR. Therefore, we expect the results from (noisy) AEGIS galaxies to be impacted by residual noise bias for small SNR. For high SNR galaxies in CatSim, the impact of pixel noise is statistically significant; however, this residual noise bias is smaller than the requirement on multiplicative bias by at least a factor of a few in r band, and by

---

[11] We assume that fluctuations in pixel values are dominated by Poisson fluctuations in a large number of detected electrons so that we can neglect the relatively small noise contributions expected from dark current and readout noise.

an order of magnitude in i band. Therefore, we conclude that, for high SNR AEGIS galaxies, our method will be sensitive to color gradient biases that are significant compared to the requirements.

2.5.3 Estimating CG bias with AEGIS galaxies

Having demonstrated that `ChromaticRealGalaxy` is able to reproduce color gradient bias results for galaxies with high SNR, we now apply our analysis to real galaxies with color gradients. A flowchart depicting the methodology to estimate CG bias from real HST galaxies in the AEGIS survey is shown in Figure 2.13. The V- and I-band images of galaxies that pass certain selection criteria are input to `ChromaticRealGalaxy` to model the galaxy SBP. Along with the PSF-convolved images, CRG also requires as input the noise correlation functions and PSF images in the V and I bands. The modeled chromatic SBP of each galaxy is then used to estimate CG bias when seen by LSST as described in Section 2.2.1.

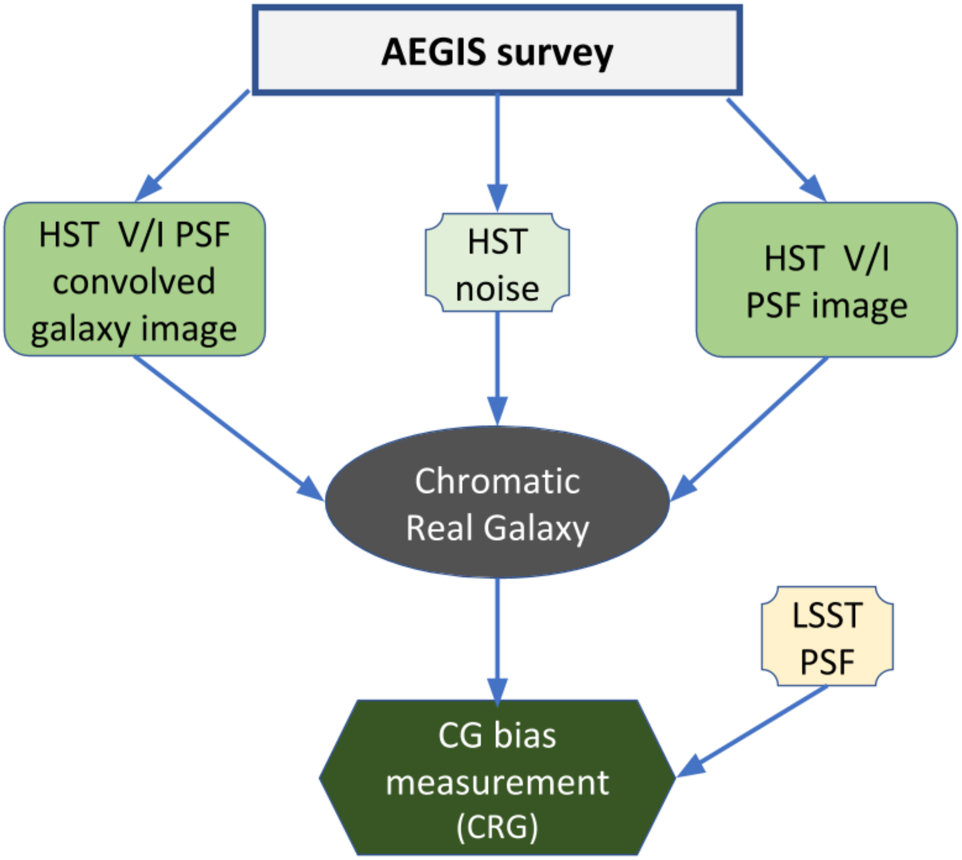


Figure 2.13: Flow chart describing the method for measuring CG bias for real galaxies from the AEGIS survey.

2.5.3.1 Creating the AEGIS catalog

The AEGIS imaging data is composed of a mosaic pattern of $21 \times 3 = 63$ contiguous HST "tiles" covering an effective area of $10.1' \times 70.5' = 710.9$ arcmin^2, with a four pointing dither pattern for each tile (Davis et al., 2007). The dithered observations for each pointing and band were combined ("drizzled") using `MultiDrizzle` to produce coadded images with a pixel scale of 0.03 arcsec (Rhodes et al., 2007).

The procedure used to create the catalog, based on Emi (2018), is similar to that used for the weak lensing catalog for the HST ACS COSMOS survey (Leauthaud et al., 2007). We use version 2.8.6 of the `SExtractor`[12] (Bertin & Arnouts, 1996) package to produce a source catalog of positions and various photometric parameters. Detection was performed on a coaddition image of both bands, and photometric measurements then performed on each band at the previously detected locations. Unreliable regions such as tile boundaries, diffraction spikes and "ghosts" due to internal reflections are masked. CRG modeling requires postage stamp images of individual galaxies. Thus if flux from a neighboring object lies within a stamp, it is replaced with noise. If the overlap with the neighbor is too significant, then the galaxy is not selected for analysis.

The final catalog consists of 26,517 galaxies that satisfy the following criteria:

- The object was detected in both V and I bands.
- The object is not classified as a star in either band.
- The object is not in a masked region in either band.
- The object has magnitude brighter than 25.2 in the I band.
- The postage stamp does not contain flux from a neighboring object (no significant overlap).

For 24,635 of these galaxies, shapes were successfully measured in the ring test for the galaxy and for its equivalent with no CG.

We model the PSF by comparing stars in a given HST tile to star fields drawn from `Tiny Tim` ray-tracing software (Krist, 1993). The variations in the size and ellipticity of the PSF across the focal plane of the HST ACS is dominated by the effective focus, which changes due to thermal expansion and contraction of the HST. `Tiny Tim` simulates the variation of the PSF across the field for different focal lengths. The PSF variation across an image is characterized by comparing the measured PSF ellipticity for stars in a field to the `Tiny Tim` predictions for different focus offsets, and finding the focus that best matches the measured and predicted PSFs across the field (Leauthaud et al., 2007). The PSF for each galaxy is taken to be the `Tiny Tim` model PSF image whose location in the `Tiny Tim` grid is closest to that galaxy's position in the focal plane, for the best-fit focus offset.

---

[12] https://www.astromatic.net/software/sextractor

The correlated noise in the two bands is estimated from empty regions in the AEGIS fields. A detailed description of how individual galaxy images, their corresponding PSFs and the noise correlation functions are determined can be found in the note described in Section 2.5. We also include in our final catalog the spectroscopic redshifts for 3763 matched galaxies from the DEEP2 galaxy redshift survey, Data Release 4 (Newman et al., 2013).

2.5.3.2 Characteristics of AEGIS catalog

Characteristics of the selected AEGIS galaxies are shown in Figure 2.14. The left panel shows the distribution of magnitudes in V and I bands. The selection criteria ensure that all galaxies have I < 25.2. The center panel shows the distribution of galaxy HLR in arcseconds; 32% of the galaxies have HLR less than or equal to the LSST pixel scale of 0.2 arcsec. The magnitude and HLR were estimated with `SExtractor`. The right panel shows the redshift distribution for the 14% of the galaxies for which spectroscopic redshift estimates are available in the DEEP2 Galaxy Redshift Survey.

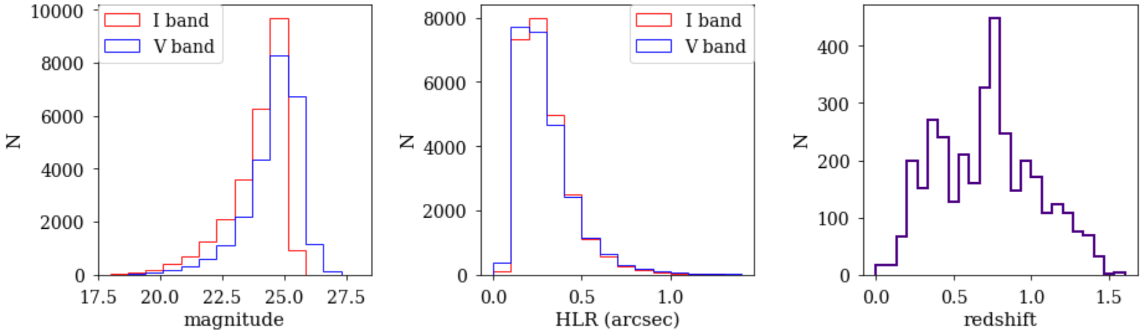


Figure 2.14: Distributions of AEGIS galaxy parameters. Left: V- and I-band magnitude. A selection criterion of I$<$ 25.2 was applied in the generation of the catalog. Center: Half-light radius in arcseconds. Right: Spectroscopic redshift for 14% of galaxies in the survey.

Since CG bias estimates show a large dependence on galaxy SNR, we show the distribution of galaxy SNR in HST V and I bands in Figure 2.15 (top left panel); the mean SNRs are 76 and 88 in V and I bands, respectively. The top-right panel in Figure 2.15 shows the correlation of I-band SNR with magnitude. A selection criterion of HST I-band SNR > 200 excludes galaxies with I-band magnitude fainter than 23.2. In the bottom panels, we plot SNR in I band as a function of V$-$I color (left), and redshift (right). The selection cut of I-band SNR > 200 eliminates 97% of galaxies with $z > 1.2$ from the sample.

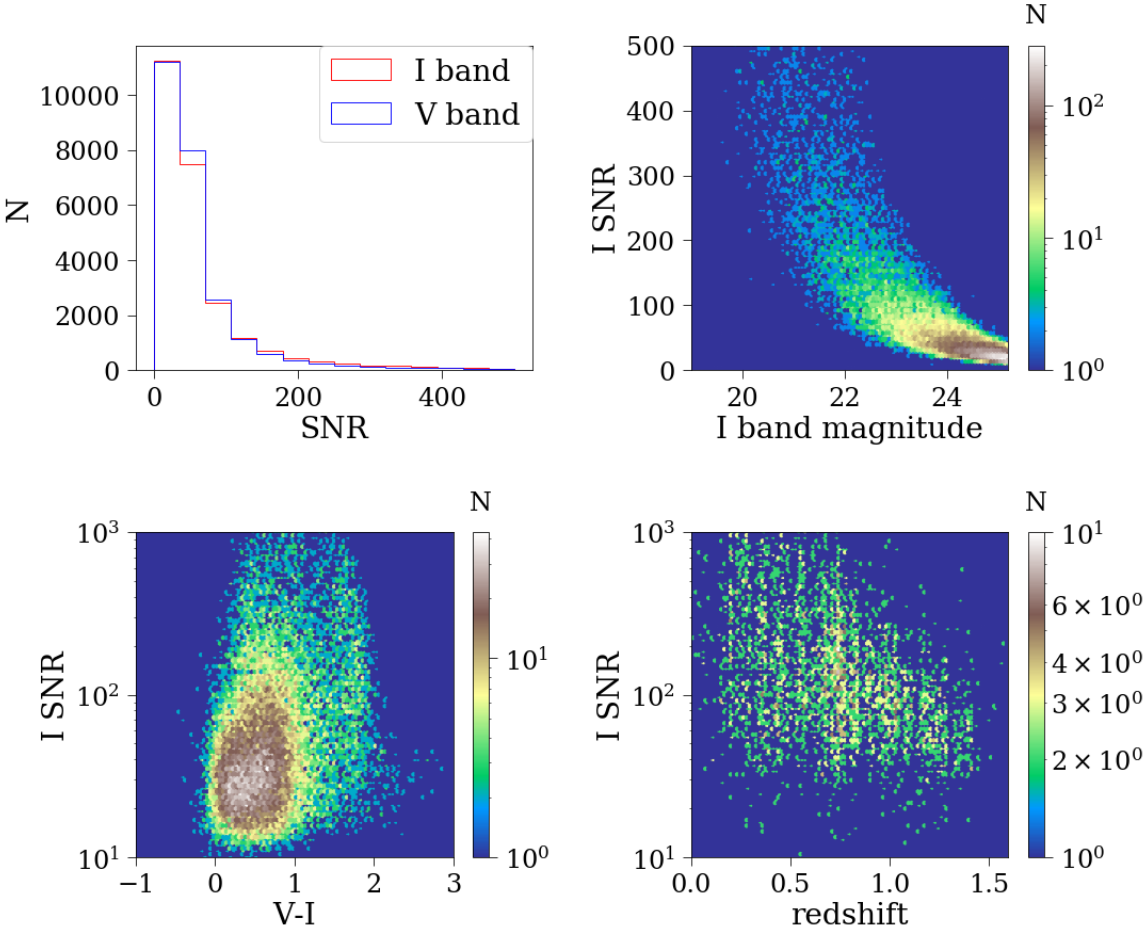


Figure 2.15: HST SNR of AEGIS galaxies. Top left: SNR for HST V and I bands. Top right: I-band SNR versus I-band magnitude. Bottom left: I-band SNR versus V−I color (left). Bottom right: I-band SNR versus redshift for the 14% of galaxies for which spectroscopic redshift estimates are available.

2.5.3.3 Results of CG bias analysis of AEGIS galaxies

The results of the CG bias analysis of AEGIS galaxies when observed in the LSST r band are shown in Figure 2.16. The plot on the left shows the distribution of estimated bias. The mean multiplicative shear bias is $m_{\mathrm{CG}} = (-3.7 \pm 1.6) \times 10^{-3}$ for all galaxies in the sample. However, as shown in the center panel, the large spread in the estimated bias is due to the low SNR galaxies. In the right panel, we show the mean value of m_{CG} for galaxies with I-band SNR greater than a minimum value ranging from 0 to 500. The dashed black horizontal line corresponds to zero bias. As we exclude low SNR galaxies from the sample, the measured CG bias values approach zero. The error bars correspond to the statistical uncertainties on the mean. The mean bias for the 1900 galaxies with

I-band SNR > 200 is $m_{\rm CG} = (0.4 \pm 2.2) \times 10^{-4}$. As emphasized in Section 5.2, this estimated bias is due to both CG bias and residual noise bias of $\mathcal{O}(10^{-4})$.

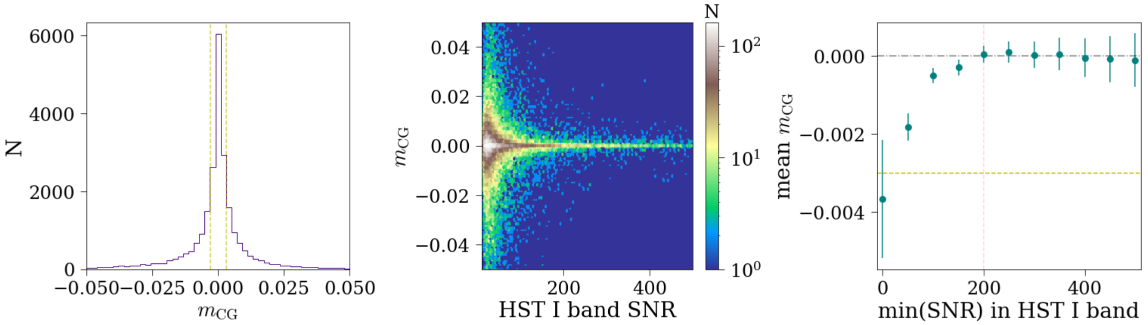


Figure 2.16: Left: Histogram of estimated multiplicative shear bias due to CG, $m_{\rm CG}$, for r-band images computed with `ChromaticRealGalaxy` from all AEGIS galaxies. Center: $m_{\rm CG}$ as a function of HST I-band SNR. Right: Mean of estimated CG bias for galaxies for which SNR is greater than min(SNR). At low SNR, the magnitude of the estimated CG bias is dominated by residual noise bias. The yellow dashed line shows the LSST DESC requirement of 0.003 on the total systematic uncertainty in the redshift-dependent shear calibration.

In Figure 2.17, we show the dependence of the CG bias estimate $m_{\rm CG}$ on SNR, V−I color, and redshift for galaxies with I-band SNR > 200. The SNR and color dependence is depicted for the 1900 galaxies with I-band SNR > 200, while the redshift dependence is plotted for 1043 of these galaxies for which spectroscopic redshift estimates are available. The dots denote the mean value for each bin while the error bars denote the statistical uncertainty on the mean. The CG bias values do not show a statistically significant dependence on galaxy color. The bias values show a small dependence on SNR; however, the magnitudes of the mean biases are lower than the LSST requirement.

Similar to the results for CatSim galaxies in Section 2.4.4, the CG biases computed here for the AEGIS galaxies could be affected by selection effects leading to a sample that is not representative of the LSST sample that would be selected for a shear analysis. Overcoming this limitation will require further studies of galaxy CG.

We compare CG bias estimates for the AEGIS galaxies when observed in LSST i and r bands in Figure 2.18. We show bias estimates for 24,635 galaxies from the AEGIS sample for which bias estimates were successfully computed in both bands. The left panel shows the distribution of bias estimates. The i-band CG bias estimate distribution has smaller tails than the r band, with a mean $m_{\rm CG} = (1.7 \pm 0.5) \times 10^{-3}$ for the entire sample. The right panel plots the mean CG bias for galaxies above a minimum I-band SNR. The mean bias quickly converges to zero, once the low SNR galaxies are excluded from the sample. For galaxies with I-band SNR > 200, the mean multiplicative shear bias in the i band is $m_{\rm CG} = (7.8 \pm 3.6) \times 10^{-5}$.

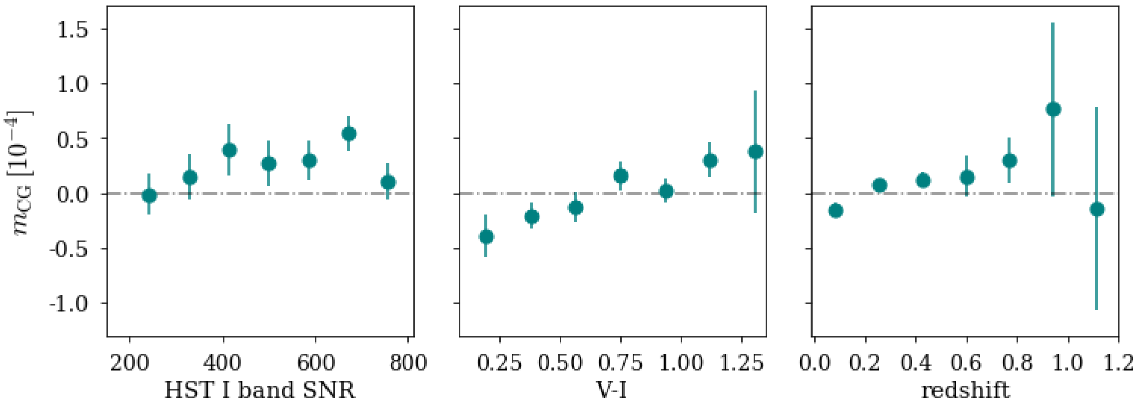


Figure 2.17: Mean CG bias estimates for r-band images generated from AEGIS galaxies with I-band SNR > 200 as a function of HST I-band SNR (left) and V$-$I color (center) for all galaxies, and as a function of redshift (right). The dependence on SNR and color is shown for 1900 galaxies while the redshift dependence is shown for the 1043 galaxies for which spectroscopic redshift estimates are available.

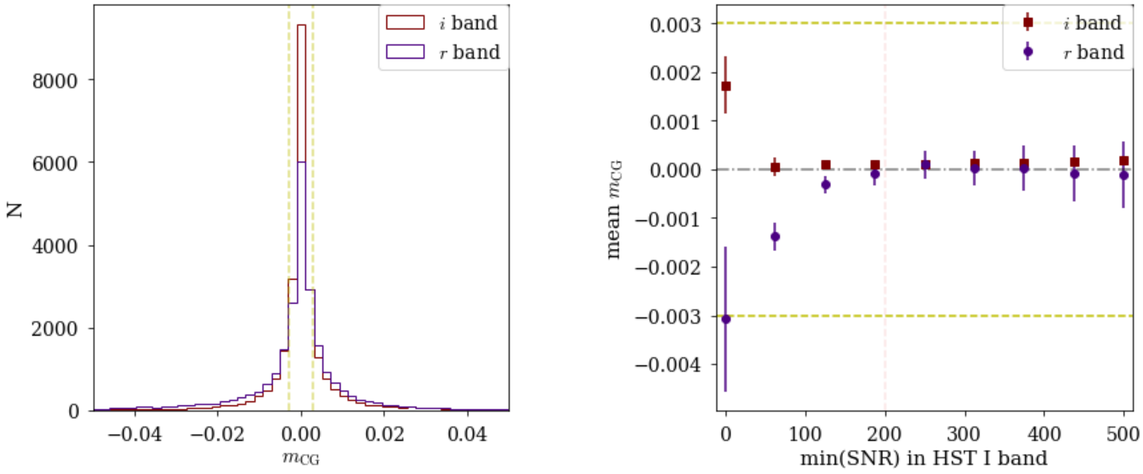


Figure 2.18: Left: Histogram of multiplicative shear bias from CG, m_{CG}, in the i band (red) and r band (blue) for AEGIS galaxies. Right: Mean m_{CG} in the i band (red) and r band (blue) for galaxies with I-band SNR $>$ min(SNR). The yellow dashed lines show the LSST DESC requirement of 0.003 on the total systematic uncertainty in the redshift-dependent shear calibration.

2.6 Limitations of this study

We summarize here the limitations of this analysis and how it could be extended to future studies.

1. Our method for isolating CG bias breaks down when galaxy SBPs are modeled from noisy images. This limits our analysis of real images to high SNR galaxies. Due to this limitation,

we are only able to measure CG bias for real galaxies with redshift < 1.2. Estimating CG bias for higher redshift galaxies will require a statistically significant sample of high SNR real images or high fidelity simulations of noise-free galaxies with realistic morphologies.

2. Our methodology assumes a linear response to the applied shear, which is true for weak lensing measurements. However, for lensing by galaxy clusters and in the strong lensing regime when this approximation no longer holds, color gradient bias could result in non-negligible biases in shear estimates.

3. We studied one source of PSF chromaticity: wavelength-dependent PSF size. However, several chromatic effects, including differential chromatic refraction, refraction in optics, and sensor effects, can also produce a wavelength-dependent PSF ellipticity.

4. The fidelity of color gradient models in the samples of galaxies used in this study is limited. In `ChromaticRealGalaxy`, the SEDs for the AEGIS galaxies, as would be seen by LSST, are modeled as simple linear functions of wavelength because only two HST bands are available. The parametric galaxies in the CatSim sample are described as simple Sersic bulge + disk profiles. As our knowledge of color gradients as a function of redshift improves, future studies could use more realistic CG models, including active galactic nuclei, star forming regions, dust lanes, etc.

2.7 Conclusions

We extend the method employed in Semboloni et al. (2013) for Euclid-type observations to isolate the effect of color gradients on weak lensing shear measurements for observations in the LSST r and i bands. The bias originates when shapes of galaxies with color gradients, observed with a chromatic PSF, are measured with some form of spatial weighting. The bias depends only weakly on the size of the PSF, but depends linearly on the magnitude of the exponent α in its wavelength dependence λ^{α} (Figure 2.9). The bias is very sensitive to the size of the weight function (Figure 2.8). We estimate CG shear bias for three samples as observed with LSST: reference parametric galaxies with large color gradients, an ensemble of parametric galaxies with more realistic color gradients, and galaxies with realistic morphology. The results are summarized in Table 2.2.

For noise-free parametric galaxy simulations, the value of half the maximum span of $m_{\mathrm{CG}}(z)$ in the redshift range [0, 1.2] is $\leq 1.5 \times 10^{-3}$ for the reference galaxies with extreme color gradients and $\leq 10^{-4}$ for CatSim galaxies. For input AEGIS galaxies with pixel noise, the estimated bias shows a strong dependence on SNR due to contributions from effects other than CG. However, for AEGIS galaxies with HST I-band SNR > 200, the magnitude of the mean estimated bias in the LSST r band is $(0.4 \pm 2.2) \times 10^{-4}$, while the value of half the maximum span of $m_{\mathrm{CG}}(z)$ in the redshift range [0, 1.2] is $\leq 1.5 \times 10^{-4}$. Therefore, for both the noise-free parametric galaxies and for the AEGIS galaxies with SNR > 200, the half-maximum span is less than the LSST full-depth requirement of 0.003 on the total systematic uncertainty in the redshift-dependent shear calibration m_z (Equation 1.19). This result is important because shear estimation depends on accurate knowledge of the PSF – even for shear estimation techniques that self-calibrate responsivity to shear (e.g., Metacal, Huff & Mandelbaum, 2017) – and the effects of color gradients on the per-galaxy PSF could be very challenging to predict and correct. We are thus optimistic that color gradients will not be a source of limiting systematic uncertainty for LSST; however, the limitations listed in Section 2.6 should be investigated as more observations and studies of color gradients in galaxies become available.

Table 2.2: Results for multiplicative shear bias statistics for $m_{\mathrm{CG}}(z)$, due to color gradients.

Sample	Half maximum span of $m_{\mathrm{CG}}(z)$ in r band	Mean m_{CG} in r band	Mean m_{CG} in i band
Reference galaxy with extreme CG (parametric)	1.5×10^{-3} (see Figure 2.7)	$(8.7 \pm 0.6) \times 10^{-4}$	–
CatSim galaxies (parametric)	$\leq 10^{-4}$ (see Figure A.7)	$(0.268 \pm 0.003) \times 10^{-4}$	$(0.372 \pm 0.001) \times 10^{-4}$
AEGIS galaxies (CRG)	$\leq 1.5 \times 10^{-4}$ (see Figure 2.17)	$(0.4 \pm 2.2) \times 10^{-4}$	$(0.8 \pm 0.4) \times 10^{-4}$

Chapter 3

The Blending Challenge for LSST

3.1 Overview

The greater observing depths of LSST will lead to an increase in the number of galaxies that are observable, as well as higher sensitivity to low surface brightness features of extended objects, that were previously too faint to detect. A resulting effect is that a significant fraction of these observed objects overlap with each other. This "blending" places additional limitations on our ability to correctly identify and measure the observed objects. In this chapter we discuss the challenge of blending for the LSST as well the current and future plans to tackle it.

We will follow the LSST convention of referring to the astrophysical objects as "objects" and the detections and measurements made as corresponding to "sources". Each source may be defined as a single component or multiple components; e.g., bulge + disk sources.

3.1.1 Why blending is a challenge for LSST

The problem of blending is not new and has always plagued astronomical studies (Beard et al., 1990; Jarvis & Tyson, 1981), particularly in crowded fields at low galactic latitudes (Irwin, 1985; Stetson, 1987; Mighell, 1989). However, in shallow weak lensing surveys, galaxy blends were rare enough to just be omitted from the study without impacting the statistical precision significantly. For example, in the Canada-France-Hawaii Telescope Lensing Survey (CFHTLenS) WL survey ($i \leq 24.7$), galaxies with overlapping isophotes were excluded from analysis (Miller et al., 2013). But with the deeper imaging and higher statistical sensitivity expected from the LSST, this solution is no longer viable. Close pairs are more prevalent in high-density regions. Therefor excluding them from the analysis could introduce a few-percent scale-dependent bias in shear-shear correlations (Hartlap et al., 2011).

The two primary reasons why a significant fraction of galaxies will be blended in LSST images are its increased observing depth and smearing of images due to atmospheric turbulence.

3.1.1.1 Increased depth of observation

Using simulations Chang et al. (2013) estimated that the fraction of blended objects in optical surveys would increase as surveys probe fainter limiting magnitudes. LSST will observe objects to limiting magnitudes of $r \leq 27.5$ and $i \leq 26.8$ (Ivezic et al., 2008). This is significantly deeper than previous ground based surveys. DES (Dark Energy Survey Collaboration et al., 2016) with a limiting magnitude of $i \approx 24$ has about 30% of galaxies in their WL survey impacted by blending (Samuroff et al., 2018). For the Hyper Suprime-Cam (HSC) (Aihara et al., 2018) wide field survey, with a higher limiting magnitude $i \approx 26$, about 58% of galaxies are in blended groups (Bosch et al., 2018).

Figure 3.1 shows simulated images of a 5×5 arcmin patch of sky observed in the i-band with LSST 10-year depth and DES 5 year depth in the left and right panels, respectively.

Even by eye one can clearly see that a large number of fainter galaxies not visible in the DES field will be observed by LSST. A large number of source that appear isolated in the DES field overlap with their neighbors in the LSST image. While the HST deep fields have observed at such depths, LSST being a ground based survey has the additional complexity caused by atmospheric seeing.

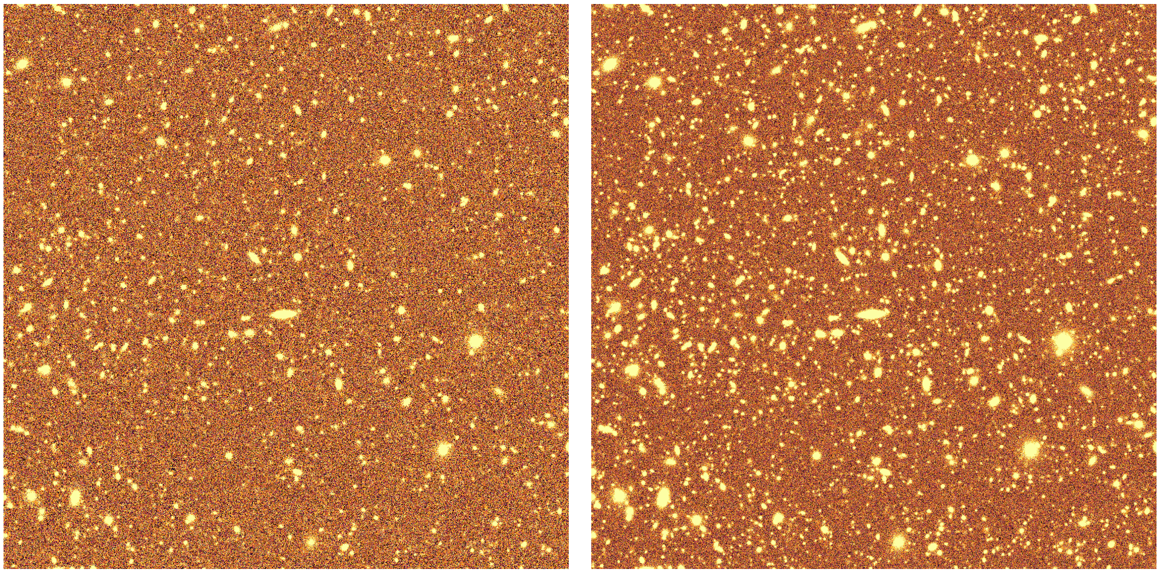


Figure 3.1: Simulated 5×5 arcmin field as would be observed in the i-band by DES (left) and LSST (right). A large number of sources not visible in the DES image are observed in the LSST image making blending a significant challenge for LSST.

3.1.1.2 Larger atmospheric PSF

The smearing of light by the PSF exacerbates the blending problem, making it more challenging for ground-based surveys than space-based telescopes. The fiducial *i*-band PSF of the LSST is 0.7 arcsec. A large number of galaxies seen by LSST will be small with a size comparable to the PSF. This results in a large number of objects(14%) that are well resolved in HST deep images, to be inseparable from their neighbor when observed by LSST with lower resolution (Dawson et al., 2016).

This is demonstrated in Figure 3.2 using simulated images of a two-galaxy blend when observed by LSST with different atmospheric seeing. The intrinsic parameters of the galaxy are kept the same as we vary just the size of the Kolmogorov PSF that the galaxies are convolved with. The PSF FWHM is shown at the top of each image. The two galaxies that are well separated in the first panel (low seeing) appear to overlap more and more with larger seeing making it harder to even resolve the blend as two objects at very large seeing.

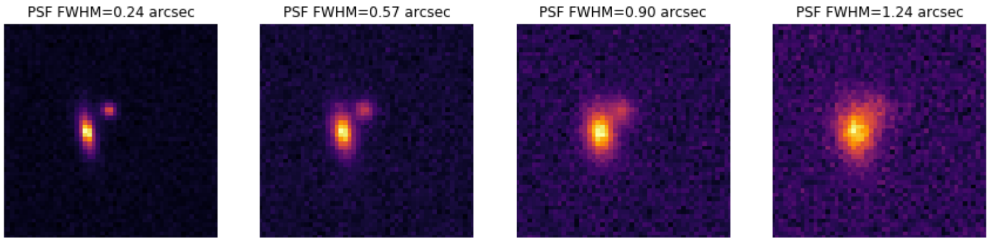


Figure 3.2: Simulated images of two galaxies seen by LSST with increasing seeing from left to right. As the PSF FWHM increases blends that are well resolved are increasingly harder to separate.

3.1.2 Impact of blending on LSST weak lensing measurements

While some degree of blending will be experienced by all science studies with LSST, its most significant impact will be experienced in Weak lensing (WL) measurements due to the stringent requirements on systematic biases. In this section we describe the different ways in which blending can impact WL measurements.

We classify blending into two domains, depending on the extent of overlap between the objects. The first is "recognized blends" where the peak brightness of two or more overlapping objects is well enough separated so that each object is detected. In such cases the performance of the deblending algorithm can be assessed to determine the presence of systematic biases that can be mitigated, for example, by removing blended objects with large overlaps.

On the other end of the blending spectrum are "unrecognized blends" where objects overlap to such large extents that the detection algorithm is unable to recognize them as different sources. Since we are not able to identify these in images, it is only possible to quantify their rate of occurrence, and apply analysis-level mitigation strategies. Studying the impact of these unrecognized blends in LSST observations and developing methods to reducing their occurrences is the aim of the second part of this thesis. While we discuss unrecognized blends in more detail in the following chapter (Chapter 4), we describe their impact of blending on WL measurements here.

As described in Section 1.5, WL information arises from correlations between galaxy shapes as a function of redshift. Thus, blending could introduce biases in cosmological results from LSST WL measurements by impacting either or both shear estimation and photometric redshift estimates (Mandelbaum, 2018).

3.1.2.1 Impact of blending on shear estimation

One way of quantifying the uncertainties in shear estimates and the statistical power of the dataset is with the effective number density, n_{eff}, defined to be the number density of perfectly measured galaxies that would contribute the same amount of statistical sensitivity as the (imperfectly) measured ensemble of galaxies (Albrecht et al., 2006). Blending leads to smaller values of n_{eff} in two ways: by increasing the uncertainty in shape estimation and by reducing the number of detected sources.

One of the earliest studies quantifying the impact of several survey factors for the LSST observations (including blending) was by Chang et al. (2013) using galaxy simulations. The study did not distinguish between recognized and unrecognized blends, but classified galaxies with a center-to-center distance d less than a certain threshold (that depends on the deblending and measurement algorithm) as seriously blended objects that would not be deblended in the processing of LSST data. The fraction of galaxies thus classified as blends is shown in Figure 3.3 as a function of the real number density of galaxies for three different values of d. The study did not include the impact of measurement errors due to blending of galaxies with separation larger than d. For a fiducial distance of about twice the width of the PSF or $d \approx 2$ arcsec, in a typical WL field at moderate galactic latitude $b \approx 60$ deg, the authors estimate that rejection of seriously blended objects would lead to a ≈15% decrease in n_{eff} for the LSST.

Prior studies on unrecognized blends include comparing Hubble Space Telescope (HST) Advanced Camera for Surveys (ACS) observations to galaxies observed with Subaru SuprimeCam at LSST seeing and depths (Dawson et al., 2016)[1] in a high density cluster field. Figure 3.4 shows four such unrecognized blends which were detected as a single source in the Subaru i-band (left) but were in fact multiple objects when seen in the HST field (right). The ellipticity of detected sources are overlaid on both images with green denoting HST measurement and red the measurement from the Subaru composite source. The shapes of the unrecognized blends were found to be different from the true galaxy shapes. Unrecognized blends were measured to have an rms ellipticty that is 13%

---

[1] Unrecognized blends were called ambiguous blends in the study.

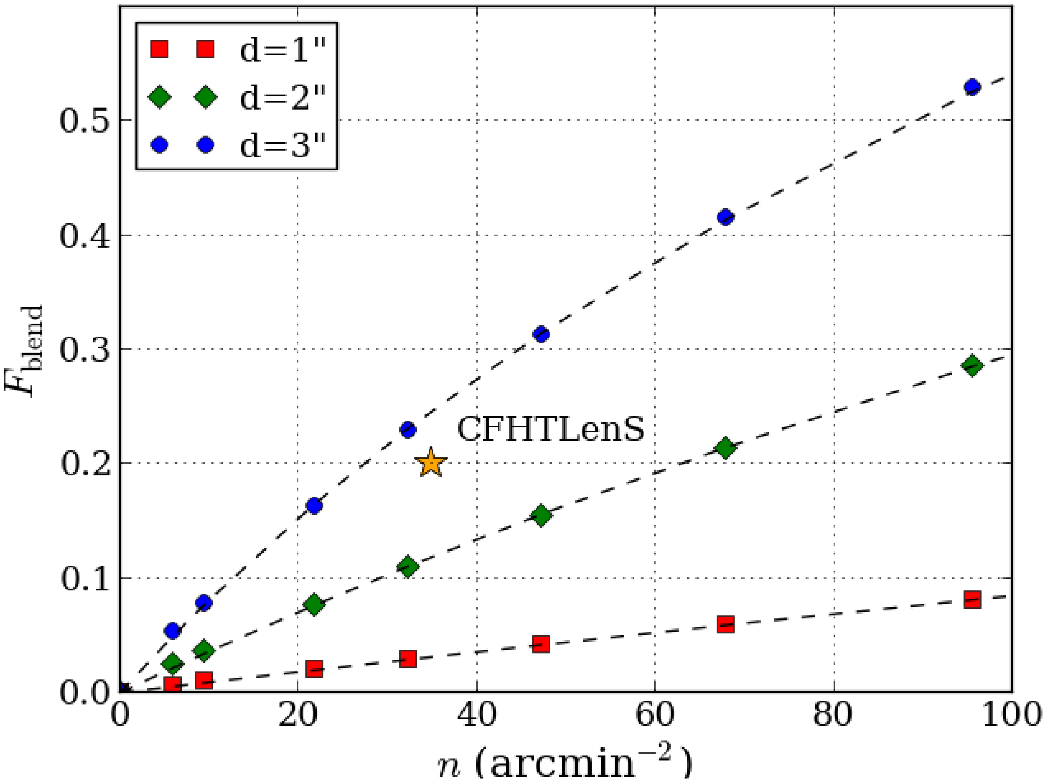


Figure 3.3: Fraction of galaxies seen LSST that have neighbors with a centre-to-centre distance less than d, as a function of the galaxy number density. The fraction of seriously blended galaxies increases with the number density of the objects as well as the detection/measurement algorithm's ability to distinguish sources. Figure is from Chang et al. (2013).

larger than non-blended galaxies. They conclude that for a survey with a limiting magnitude of $i \approx 27$, unrecognized blends would lead to 14% increase in shear noise, translating to 12% decrease in $n_{\rm eff}$. If a magnitude selection of $i < 25.3$ is made, then there would still a 7% increase in shear noise (or a 7% decrease in $n_{\rm eff}$), implying that despite them not being used in science analysis, it is important to detect fainter sources since they can still impact the results through blending.

3.1.2.2 Impact of blending on photometric redshift measurements

Blending could impact photometric redshifts by affecting photometric measurements and color measurements, thereby introducing errors in SED estimates. A study of synthetic galaxies injected into real HSC images, found that the degree to which an object is blended with other objects impacts the overall photometric performance of the HSC pipeline with highly blended objects tending to have

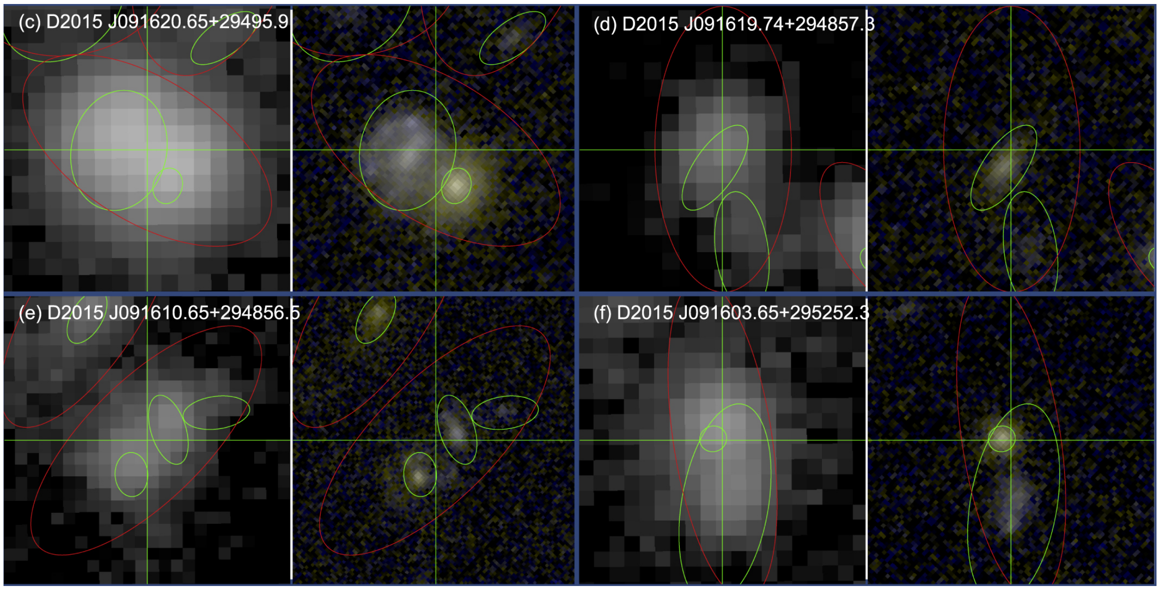


Figure 3.4: Four scenes of unrecognized blends that are resolvable from space but not from ground. For each of blends, the Subaru *i*-band image (left) is shown alongside the HST color image (right). The ellipses show the observed object ellipticities (red = Subaru, green = HST). Figure is from Dawson et al. (2016).

larger photometric uncertainties, systematically underestimated fluxes, and slightly biased colors (Huang et al., 2018).

The impact is even more severe for unrecognized blends whose composite flux measurements could lead to entirely incorrect, and possibly unrealistic, SEDs , which may not give a correct photo-z. Thus, even calibrating an algorithm to estimate photometric redshifts will be challenging if the rate of unrecognized blends in the training sample (usually spectroscopic redshift) is different from the general galaxy population that LSST will observe (Mandelbaum, 2018).

In addition to photometric measurements, a unique challenge is posed in how unrecognized blends must be interpreted during calibration of systematic effects in shear estimators. If the objects with overlapping images are at the same redshift, then they will have been affected by the same shear, thereby making it possible to properly calibrate shear estimates for the combined (non-deblended) object (Sheldon et al., 2019). However, if the objects are at different redshifts, it is unclear how the shear estimation would be assigned to a redshift bin and calibrated (Mandelbaum, 2018).

In addition to the above, blending in crowded fields will also make it very hard to perform background estimation and subtraction. Blending could also introduce shear dependent-selection effects. A studying on measuring impact of blending on galaxy cluster shear estimates shows that blending between the cluster galaxies and background galaxies causes a 0.5% decrease in measured reduced shear (Fu et al. 2020, in prep).

3.2 Addressing the blending challenge with the LSST Science Pipeline

A description of the entire LSST science pipeline LSST Science Pipeline (`LSP`)[2], from image cleaning to science measurements, is beyond the scope of this work. For this we refer the reader to LSST Science Collaboration et al. (2017) and LSST Science Collaboration (2009). We describe here some of the primary algorithms that are important to blending challenges, along with a comparison to some contemporary algorithms. The `LSP` builds on an algorithmic and conceptual foundation inherited from the Sloan Digital Sky Survey (SDSS) Photo Pipeline (Lupton et al., 2001) and is currently the optical imaging data processing pipeline for the Subaru Telescopes Hyper Suprime-Cam (HSC) instrument Bosch et al. (2018).

The current pipeline plan is to approach blending challenge in three (possibly inter dependent) stages: detection, source separation or “deblending”, and measurement. The three stages do not have to be sequential, can be an iterative, and pairs of these operations could be done simultaneously (The LSST Dark Energy Science Collaboration, 2019). Note that the description below is based on the most recent LSST Data Management Science Pipelines Design document v 4.2, (LDM-151) [3], and in all likelihood will evolve before being applied to process real LSST images.

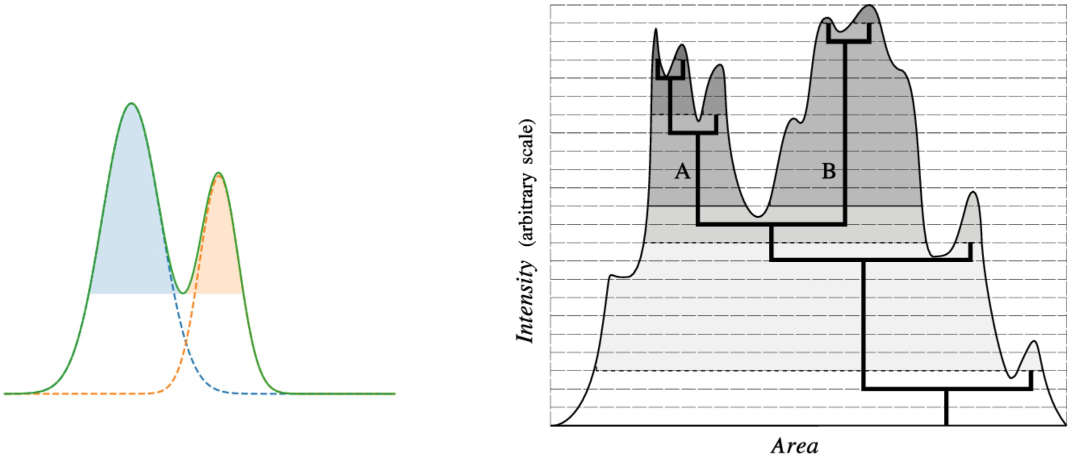
3.2.1 Detection

Before the sources can be separated and measured, we first need to know how many sources are there and where they are. This is traditionally done by identifying regions above a certain background threshold and then finding the approximate positions of sources within them. We show this with an example of a cross section along the shortest line between the centers of two overlapping neighbors in the left panel of Figure 3.5. For simplicity we ignore background noise and assume both the overlapping objects as Gaussians. The combined surface brightness profile (green curve) is the sum of SBP of the two sources (blue and orange). The size of the through between the sources as well as the integrated flux above the through region (shaded region) determines if the blend would be identified as one or two sources. In the commonly used detection algorithm `SourceExtractor` (Bertin & Arnouts, 1996), the integrated pixel intensity of each peak above the trough is required to be greater than a certain fraction of the total intensity of the composite object for it to be considered a blend of two objects. For more than two peaks a “tree” of similar decision criterion is used for every trough encountered, as shown in the right panel of Figure 3.5. The detected region is re-thresholded at multiple levels (horizontal dashed lines) exponentially spaced between the background threshold and its peak value. Then the algorithm goes downwards, from the tips of branches to the trunk, and decides at each junction whether it shall extract two (or more) objects or continue its way down.

---

[2] https://docushare.lsstcorp.org/docushare/dsweb/Get/LPM-17

[3] https://www.lsst.org/scientists/publications/science-requirements-document



(a) Schematic of detection decision for two Gaussian blends

(b) Schematic of detecting sources with `SourceExtractor` with thresholding (Bertin & Arnouts, 1996).

Figure 3.5: Schematic diagram of detection decision criterion to identify sources in blends

While source separation performed with simultaneous fit would also estimate the centroid locations, it would require knowledge of number of sources and a preliminary guess of the centroids. The step of predicting the number of sources and their preliminary centroids are thus classified here as the detection step.

3.2.2 Different types of blending detection errors

Depending on the configuration, blending can introduce errors in object detection in several ways. We demonstrate this for the simple case of a two-galaxy blend in the schematic shown in Figure 3.6 where the true overlapping objects are shown as red and blue ellipses and the black-dashed ellipse denotes the detected source. The first two overlaps represent (a) recognized and (b) unrecognized blends, respectively, that have been already described in Section 3.1.2. Blending can also introduce detection errors like shredding shown in panel (c) of Figure 3.6 where the two-galaxy blend is detected as three objects.

The first three scenes in Figure 3.6 depict blends in which each galaxy would have been correctly detected if it were isolated. Blending can also lead to a configuration like that in Figure 3.6(d) where the each of the two galaxies was below the detection threshold, but the flux of the combined source was higher than the threshold leading to the scene being detected as one source. Detection errors in recognized blends can manifest as incorrect centroid estimation, while the other configurations lead to incorrect number of detections.

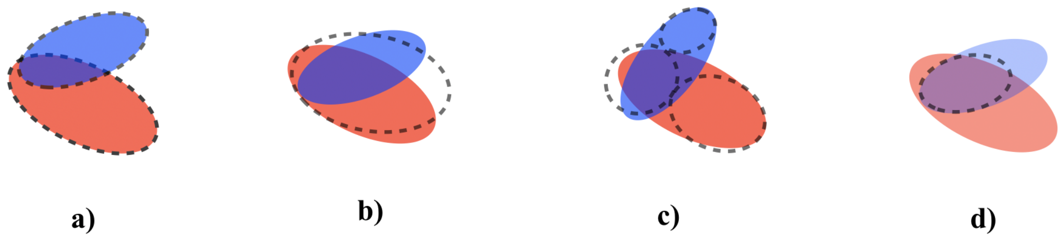


Figure 3.6: Schematic of different ways blending can impact detection: a) recognized blends, b) unrecognized blends c) shredding, and d) blend induced detection

3.2.2.1 Source detection in the LSST science pipeline

After background subtraction, the images are convolved with a smoothing filter and contiguous pixels above a certain threshold are defined as a "Footprint". The smoothing filter is the transpose of the PSF[4]. Peaks identified in the Footprint are assumed to correspond to a distinct astrophysical object that should be separated from its neighbors by the deblender. The `LSP` does not use color information during detection, but rather detection is performed in each band and the detected centers are then merged together, merging co-located peaks to eliminate redundant detections.

3.2.3 Deblending

Deblending can be defined as the process of constructing individual object images or models where a portion of the flux in each pixel is assigned to each blended object. There are two ways this can be done; one is with algorithm separates per pixel flux between overlapping objects and produces postage stamps for each source; measurements could then be performed on individual postage stamps as though it were isolated. The second approach merges the deblending and measurement steps by modeling the blend scene as parametric objects with the goal of the fit to get the parameters of the individual sources. The current `LSP` follows the first approach, while the simultaneous fit algorithm `lensfit` has been used in analyzing CFHTLens images (Miller et al., 2013). The detection and deblending steps defined in the `SourceExtractor` algorithm would be classified as a detection step according to our definition. `SourceExtractor` does not divide the flux in a pixel between sources but only assigns a pixel to one source. It does not output debelended images of the individual blend object, but instead only provides segmentation maps that do not overlap.

We continue with the SDSS nomenclature of defining the initial blend image as the "parent" and

---

[4] This is the optimal filter for detection of isolated point sources when the noise is dominated by sky background (Bosch et al., 2018)

the deblended sources as "children". A blend is thus represented in `LSP` as a Footprint containing multiple peaks and the job of the deblender algorithm is to apportion the flux in that Footprint to the peaks, creating a child image for each peak that can be measured as a distinct source.

3.2.3.1 SDSS Photo deblender

The fundamental ansatz in the SDSS deblender is that objects have approximate 180 deg rotational symmetry about their peak flux position (Lupton, 2005). Thus, in a blend if one portion of the object's profile is contaminated by flux from its neighbor, then the assumption is that profile can be recovered by looking at the opposite side of the peak. The different steps of the algorithm can be summarized as:

1. Find peaks in the blend. In the top-left panel (a) of Figure 3.7 a one-dimensional toy model shows a blended image where three peaks have been identified. The dashed lines show the true profiles of the objects in the blend, while the solid black line is the composite profile.
2. For each child construct a template by comparing the intensities of pairs of pixels symmetrically placed about the peak, and replacing both by the lower of the two. The parent image, $I(x)$, is decomposed as the sum of products of weights, w_r, and templates, T_r, of r (here 3) children.

$$I(x) \approx \sum_{r=1}^{n} w_r T_r(x). \tag{3.1}$$

   The templates of the children are shown in different colors in panel (b).
3. Solve for weights by minimizing E:

$$E \equiv \sum_{x} (I(x) - \sum_{r=1}^{n} w_r T_r(x))^2 \tag{3.2}$$

4. Pixel flux of every r'th child, $C_r(x)$, is assigned as

$$C_r(x) = I(x,y) \frac{w_r T_r(x)}{\sum_{r=1}^{n} w_r T_r(x)} \tag{3.3}$$

5. The pixel intensities for each child are thus estimated as shown in panel (c).

The deblending process is performed independently on all bands at the merged center locations identified in the detection step.

While the deblender works well for two galaxy blends, the assumption of inferring flux from the opposite region fails for three or more objects in a line (as in Figure 3.7). When building the template for the middle peak, the minimum in Equation 3.1 will have to choose between pixels that are both affected by neighbors, yielding a poor template. The difference in blend image and the sum of deblended sources is plotted in panel (d), showing non zero residuals for the central object. While the central object was a monotonically decreasing smooth profile, it's deblended image has smaller peaks on either side of the center. While this was not a significant source of error in SDSS, this problem is dramatically more prevalent in HSC, because blends are both more common and complex at more depths (Bosch et al., 2018). Thus, new deblending algorithms are being tested for analyzing HSC and LSST images. The algorithm showing the most promise is the multi-band deblender `SCARLET`

3.2.3.2 Source separation with `SCARLET`

`SCARLET` (Melchior et al., 2018) is a source separation framework for multi-band images, based on a generalization of the Non-negative Matrix Factorization to alternative and several simultaneous constraints. `SCARLET` describes the blend scene to be a superposition of multiple components, each with a spatially compact support and a spectrum that is constant over that component. An example of `SCARLET` deblending of a two-galaxy blend scene is shown in Figure 3.8. The top panel shows an image of a two-galaxy blend (left), its `SCARLET` model (middle), and the residual (right). The true image of each isolated galaxy is shown in the left column of the bottom panel. The galaxies are modeled with both spatial and spectral support, which obeys certain constraints: the spectral support (right column) must be non-zero and the spatial support (center column) must be monotonically decreasing from the center of the galaxy and rotationally symmetric.

Another multi-band deblender similar to `SCARLET` is the Multi-band morpho-Spectral Component Analysis Deblending Tool (MuSCADeT) (Joseph et al., 2016), which uses morphological dictionaries such as wavelets and SEDs to separate images of galaxies with different colors.

In the DES Y1 cosmological analysis, deblending was performed by the multi-object fitting (MOF) deblender, where each source is fit replacing others with noise iteratively until the collective residual reaches certain convergence criteria (Drlica-Wagner et al., 2018).

3.2.4 Measurement

We classify source measurement as any step where an algorithm that characterizes the properties of individual detected sources is run. These include centroid, shape and photometric measurements.

3.2.4.1 Measurement with noise replacement in LSP

`LSP` performs measurements in each band on both the parent and child object. Measurements on the children are performed using a noise replacement method as follows:

1. All pixel values in the footprint in the image are replaced with random Gaussian noise with the same variance as the original noise in those pixels.

2. The deblended pixels for a particular source to be measured are inserted back into the image (replacing the noise pixels).
3. Measurement algorithms are run on this image.
4. The above steps are repeated for all sources in the blend.

The process of replacing neighboring pixels by noise leads to the low-signal-to-noise outer parts of extended objects, that were not assigned to any source, being measured multiple times.

In addition to the current implementation of `LSP`, other recently developed techniques are being tested for potential application to WL measurements, including the self-calibrating shear measurement algorithm `Metacalibration` (Huff & Mandelbaum, 2017) and its extension to object detection in blends, `Metadetection` (Sheldon et al., 2019). In `Metadetection`, an artificial shear is applied to images of small regions of sky, followed by detection and measurement algorithms applied to the sheared images, in order to calculate the shear response for particular algorithms.

We emphasize that the steps described above outline the current plan to tackle the blending challenge for LSST and may evolve in the future. Rigorous tests are being performed on the pipeline with simulations (Korytov et al., 2019; Sánchez et al., 2020) and real galaxy images from HSC, while newer algorithms are being developed to improve performance. While recognized blends by themselves pose a significant challenge to LSST measurements, the focus of the rest of this thesis will be on measuring the rate of unrecognized blends in object detection and developing techniques to reduce their frequency.

The remainder of this thesis focuses on understanding, measuring and reducing instances of unrecognized blends. By their very definition we would not be able to identify the unrecognized blends in real LSST images. The only way to limit their impact is to study the extent of their occurrences and calibrate their impact on estimation of cosmological parameters. In Chapter 4, we discuss in detail the performance of the current `LSP` detection algorithm and study the extent and impact of unrecognized blends using simulated galaxy blend scenes.

Improvements in computer architecture and developments in machine learning have led to a more recent significant increase in applying such solutions to astronomy problems. We thus explore one such solution called `ResidualDetectron`, where convolutional neural networks are trained to identify unrecognized blends. We describe the network model in detail in Chapter 5 and test it on synthetic galaxy images in Chapter 6.

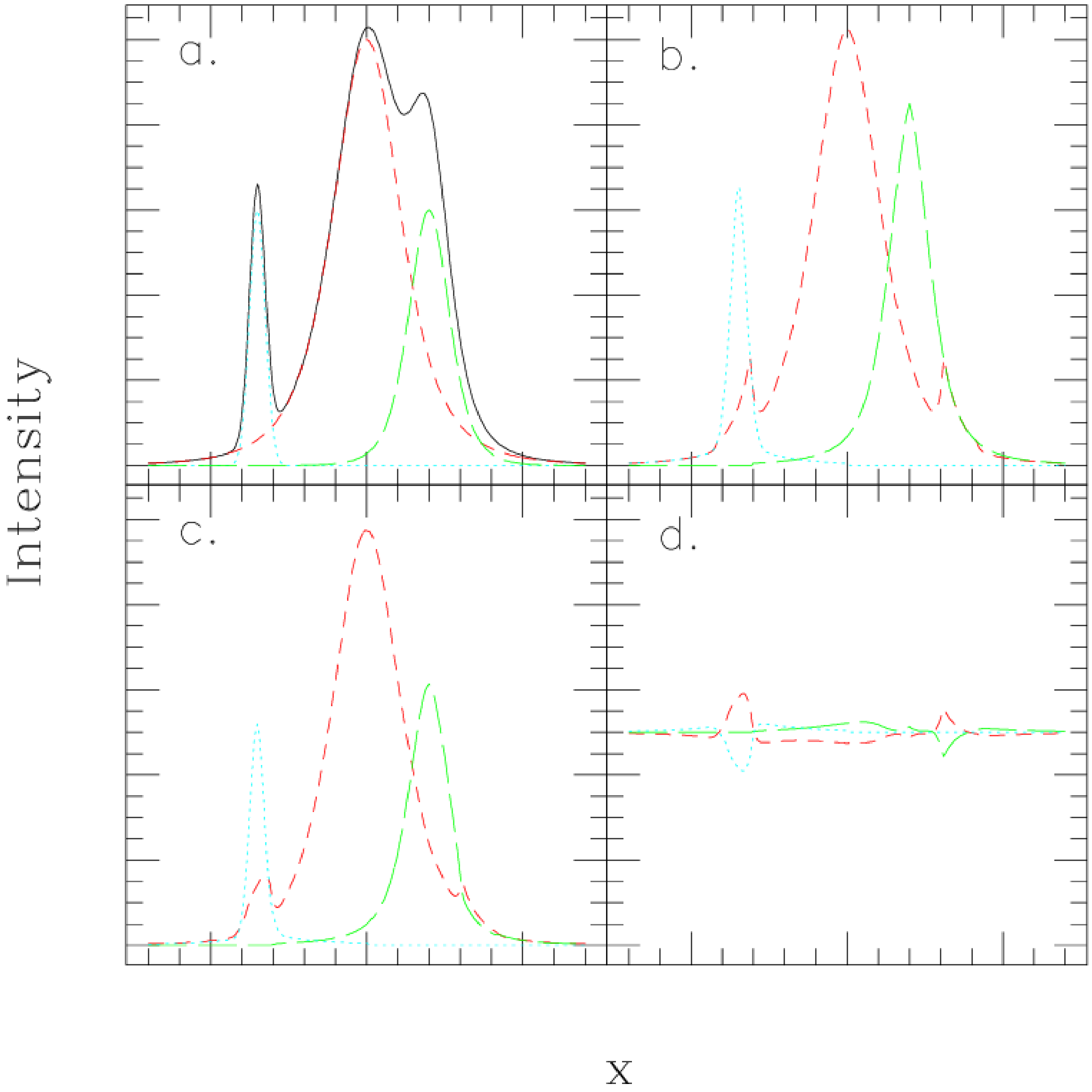


Figure 3.7: A simple 1-dimensional example of deblending with the SDSS Photo Algorithm. Panel a shows the blended object (solid) and the three components that make it up; from left to right they are referred to as star, galaxy1, and galaxy2. Panel b shows the corresponding templates; see text for details. Panel c shows the three deblended children; and panel d shows the difference between the three input components and the output of the deblender (Lupton, 2005).

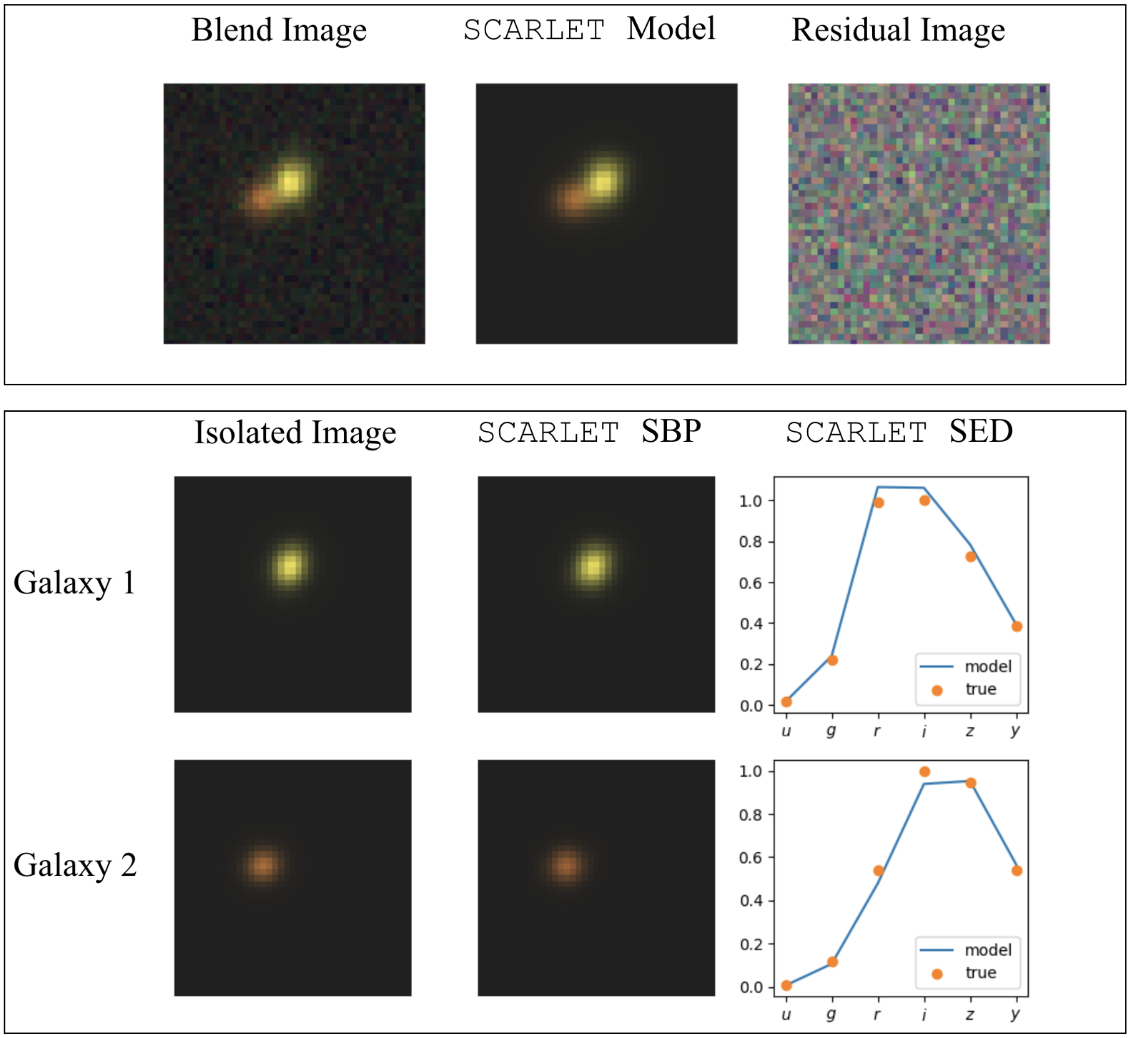


Figure 3.8: Example of source separation in a blend scene by `SCARLET`. Top: Image of a two-galaxy blend (left), its `SCARLET` model (middle) and the residual image (right). Bottom: `SCARLET` modeled SBP and SED for each galaxy. The left column shows the true isolated image of galaxies 1 and 2 while the center column shows the `SCARLET` modeled SBP for each. The right column shows the relative flux in each band for the true SED (orange dots) and the `SCARLET` model for the SED (blue curve).

Chapter 4

Unrecognized Blends

We defined "unrecognized blends" in Section 3.1.2 as blends where one or more objects were not detected because of blending with its neighbor. While the term "unrecognized blends" is not common in optical astronomy, the issue has been studied extensively in radio and infrared astronomy where the term confusion limit is used to describe the case in which superimposed signals from faint unresolved sources that cluster on the scale of the observing beam contribute noise in the image (Condon, 1974; Scheuer, 1974). Similar studies in the optical imaging of extragalactic sky found a significant contribution (29 mag arcsec^{-2}) from unresolved faint blue galaxies to the extragalactic background light (Tyson, 1995). In this thesis we focus on unrecognized blends in which galaxies that could have otherwise been detected as separate objects are instead detected as a single unresolved source by the current LSST Science Pipeline.

4.1 Impact of unrecognized blends on cosmological inference

We described in Section 3.1.2.1 how unrecognized blends can increase uncertainties in shear estimates. In addition they can introduce challenges in shear bias calibration if the unrecognized blend is composed of galaxies in different redshift bins. Preliminary studies by Nourbakhash et al. (in prep.) show that unrecognized blends in the WL analysis sample could introduce biases in shear measurements, that can propagate to cosmological inferences. The study identified unrecognized blends using a distance criterion where a pair of galaxies with centroids closer than a threshold (2 arcsec) is labeled as an unrecognized blend. The shapes of galaxies were used to compute 3×2 correlation functions for two cases: with blending (i.e., including unrecognized blends) and without blending (i.e., with only isolated galaxies). Preliminary results from an analysis of the difference in the 3×2 correlation functions for the two cases indicate that unrecognized blends could have a significant impact on the inference of cosmological parameters.

The study used a simple distance threshold as the criterion for a pair of galaxies to be an unrecognized blend; however, intrinsic properties of the galaxies, as well as the blend configuration, can significantly impact detection in a blend scene. Another limitation of the current study is that it uses the true redshift of the detected source for analysis and does not include errors in photometric redshifts due to blending. As described in Section 3.1.2.2, unrecognized blends can introduce a bias in photometric redshift estimation if there is a mismatch in the population distribution of the calibration spectroscopic redshift sample and that which LSST will observe. This makes it important that we first identify the population of galaxies that are part of unrecognized blends when observed by the LSST and analyzed with the current LSST Science Pipeline (`LSP`).

4.2 Unrecognized blends in the current LSST Science Pipeline

We can characterize unrecognized blends in ground-based imaging with either simulations, where we know the “truth”, or by comparing real ground-based images with high-resolution space-based images (see Dawson et al. (2016)). While we lose on “realness” of images with simulations, we gain in statistical precision since deep space-based images are limited in area. Pixel level simulations from numerical catalogs (e.g., BUZZARD (DeRose et al., 2019), CATSIM (Connolly et al., 2014)), combined with pixel level simulations such as DESC Data Challenge 1 (Sánchez et al., 2020), provide datasets that can be used for validation and verification of the processing and analysis pipeline. In this thesis work, we utilize simulations for our study of unrecognized blends.

In this section, we describe how we use simulated synthetic blend scenes to measure the performance of the detection algorithm in the current LSST Science Pipeline (`LSP`) and to study the characteristics of unrecognized blends.

4.3 Dataset

To simulate blend scenes, we use a synthetic catalog of extragalactic objects, called cosmoDC2[1] (Korytov et al., 2019), which contains the true positions and intrinsic parameters of the overlapping galaxies. CosmoDC2 was developed for the second LSST DESC Data Challenge to produce a dataset of objects representative of those that LSST will see. The cosmoDC2 catalog is based on an N-body (gravity-only) cosmological simulation, called the “Outer Rim” run, which was carried out using the Hybrid/Hardware Accelerated Cosmology Code (Habib et al., 2016) along with data products from the smaller AlphaQ simulation and the UniverseMachine (Behroozi et al., 2019) for the semi-analytic and empirical modeling components, respectively. Each galaxy is characterized by properties that include stellar mass, broadband filter magnitudes, weak lensing shear, separate coarse-grained SEDs for disk and bulge components of the image, and half-light radius.

For our unrecognized blend study, we use a portion of the first of the two-stage release of the

---

[1] https://github.com/LSSTDESC/cosmodc2

cosmoDC2 catalog, which is public[2]. The public catalog encompasses 440 deg^2 of the sky to a redshift depth of 3, and is complete to a magnitude depth of 28 in the r-band. We use a 24 deg^2 slice of the cosmoDC2 catalog with right ascension between 52 and 58 deg and declination between -32 and -28 deg. The galaxies are rendered as parametric bulge + disk profiles in the six *ugrizy* LSST observing bands.

We use the galaxy parameters from the cosmoDC2 catalog but do not use the images rendered in Data Challenge 2. We also do not include shear information from the cosmoDC2 catalog. Future studies of the impact of unrecognized blends on 3×2 point correlation functions could be performed with the `LSP` detection on the DC2 image simulations for the entire 440 deg2 catalog, including weak lensing effects.

From the ≈70 million galaxies in the 24 deg^2 slice of the cosmoDC2 catalog defined above, we identify galaxies that are part of a blend (as and perform LSP detection on postage stamp images of these blend scenes.

4.3.1 `WeakLensingDeblending` package

To render images of both isolated objects (no blending) and scenes with blended objects, based on the galaxy parameters in the DC2 catalog, we use the simulation functions from the `WeakLensingDeblending` package[3]. This package, which is based on `GalSim`, generates postage stamp images of PSF-convolved galaxy images. The relevant simulation parameters describing the LSST survey are sensor pixel size, nominal full-depth exposure time in each band, level of sky background, quality of atmospheric seeing, and overall spectral throughput and detector response ("zero points"). We model only a mean sky level, mean seeing, etc., over an entire survey. For more details see Sanchez, J. et al. (in prep).

4.3.2 Definition of a blended group

Following the definition implemented in the `WeakLensingDeblending` package, we define a blended group – or a "blend" – as a group of overlapping objects, where a group is defined as the minimal set of galaxies for which at least one pixel of each galaxy contains flux from another galaxy in the group. Since theoretically the Sersic profiles of the galaxies extend to infinity, the spatial extent to which a galaxy profile is rendered is limited based on the flux density. In particular, each galaxy's surface brightness profile (SBP) is simulated out to a limiting isophote for which the expected flux per pixel is above a threshold equal to the standard deviation of the expected sky background. We define groups in the i-band image.

We set the galaxy flux threshold equal to the sky noise, rather than the default value of 5% of the sky noise in the `WeakLensingDeblending` package, because we found that the lower default threshold

---

[2] https://portal.nersc.gov/project/lsst/cosmoDC2/_README.html

[3] The `WeakLensingDeblending` code is at `https://github.com/LSSTDESC/WeakLensingDeblending` with documentation at `http://weaklensingdeblending.readthedocs.io`.

led to inclusion of a significant number of well-separated galaxies that were successfully detected as distinct sources. Measurements of these successfully detected sources might still be biased due to their neighbors, which is why it would be important to include them in studies of the impact of blends on measurements in recognized blends. However, since it is not our goal here to address the issue of errors in measurements from blending (rather, we limit ourselves to detection issues), these well-separated galaxies are not relevant for our study. Thus, we place the more stringent limit on what constitutes a blend. The higher per-pixel threshold used here helps ensure that only galaxies that significantly overlap are included in the study. Examples of blended groups with three and nine galaxies are shown in the left and right panels of Figure 4.1, respectively.

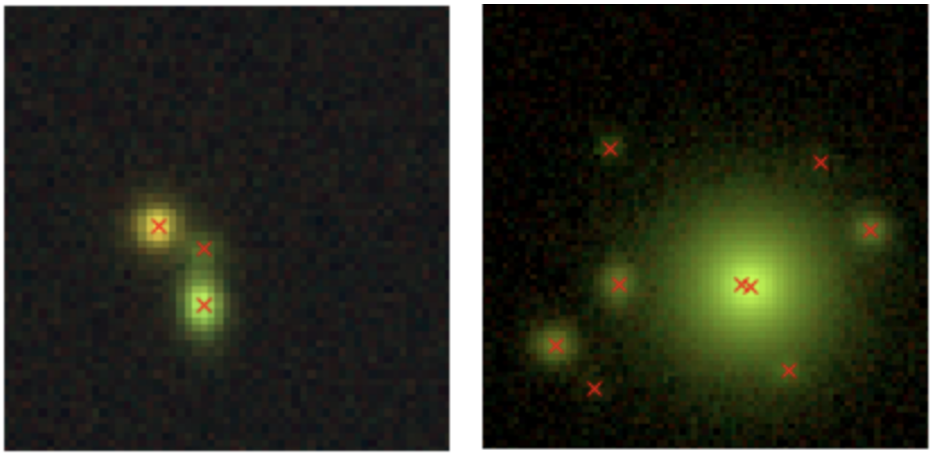


Figure 4.1: Examples of blended groups with three galaxies (left) and nine galaxies (right), based on galaxy parameters from cosmoDC2 catalog. The location of each galaxy centroid is denoted by a red ×.

The distribution of number of galaxies per group or “group size” for the 24-deg^2 cosmoDC2 catalog is shown in the top-left panel of Figure 4.2. Almost 90% of the groups formed by the ≈15 million rendered galaxies are of group size one; i.e., they consist of a single isolated galaxy. The most common group size is two with 1,945,804 galaxies (or 13% of all galaxies) found in these two-galaxy blends. Large groups are much less frequent.

To avoid excessive computational loads, we limit the group size of blends in our study to avoid rendering very large blend scenes. Because the spatial extent of the group grows with group size, and since ≈99% of groups have 10 or fewer galaxies, we include only blends with group size between 2 and 10 galaxies for our study of unrecognized blends. We show the distribution of group size for only these groups in the top-right panel of Figure 4.2.

In the bottom left panel of Figure 4.2, we show the 2D distribution of $d_{\max}$ versus group size, where $d_{\max}$ is the maximum distance between the position of each galaxy center and the mean

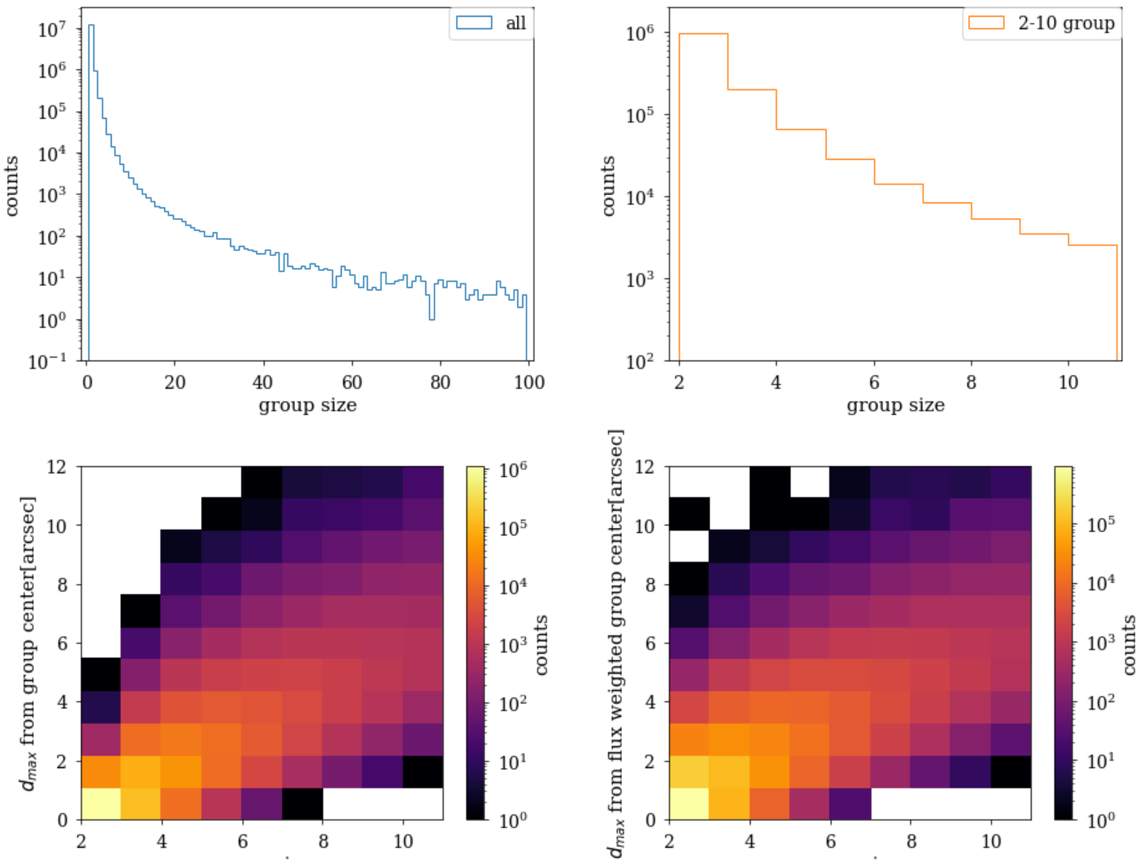


Figure 4.2: Distributions of group size for cosmoDC2 catalog. Top-left: distribution of group size of galaxy blends in 24-deg^2 cosmoDC2 catalog. Top-right: distribution of group size for groups containing 2 to 10 galaxies, used in the study of unrecognized blends. Bottom-left: maximum center-to-center distance between centroid of the group and the center of each galaxy in the group ($d_{\max}$) versus group size. Bottom-right: distance $d_{\max}$ computed from the flux-weighted group centroid.

position of all galaxy centers in the group. As expected, blends with fewer galaxies have smaller values of $d_{\max}$. Larger groups can be composed of galaxies that are ≈10 arcsec (≈50 pixels) away from the group centroid. Since the larger brighter galaxies are more likely to overlap with their neighbors, we also show the distribution of $d_{\max}$ versus group size computed from the flux-weighted centroid of the group in the bottom-right panel of Figure 4.2. Since $d_{\max} \lesssim 12$ arcsec for most groups with ten or fewer galaxies, we simulate the blended group centered in a postage stamp image of size 128×128 pixels (25.6×25.6 arcsec).

Distributions of intrinsic properties of galaxies in the cosmoDC2 catalog are shown in Figure 4.3. The distribution of i-band magnitude for all galaxies in the cosmoDC2 catalog is shown in the top-left panel (blue histogram). While we do not place an explicit selection criterion on the galaxy magnitude for our analysis, the per-pixel flux threshold of one-sigma sky noise eliminates all galaxies fainter than 27.3 mag in the i-band. Also overlaid is the distribution of i-band magnitude of galaxies in groups of sizes 2 to 10 (orange) and isolated galaxies (green). The top-right panel of Figure 4.3 shows the distributions of galaxy size $\sigma_m = |Q^{0.5}|$, where Q is the second moments of the galaxy, for the same three sets.

The distributions of i-band magnitude and second-moment size σ_m show that our test sample is depleted in both large bright galaxies and small faint galaxies. This is because the largest, brightest galaxies will overlap with more galaxies making them more likely to belong to a group with more than 10 galaxies. On the other hand, the small faint galaxies are less likely to overlap with another galaxy, thereby remaining isolated.

The bottom-left panel of Figure 4.3 shows the distribution of galaxy redshifts for all galaxies in the cosmoDC2 catalog (blue). We do not see a significant redshift-dependent selection effect depending on whether a galaxy is isolated or in a group of size 2 to 10. The bottom-right panel of Figure 4.3 shows the distribution of the purity parameter[4], which measures the extent of blending and has a value between 0 and 1. This dimensionless parameter quantifies the degree of overlap as the ratio of flux of a single galaxy to the total flux in all galaxies with which it overlaps (see Sanchez et al. (in prep) for details). A purity value of 1 implies no overlapping pixels – i.e., the object is isolated – as evident in the purity distribution of isolated galaxies in the bottom-right panel of Figure 4.3. Lower values of purity indicate increased overlap with a neighboring object or objects. Since the impact of blending depends on the extent of overlap as well as the intrinsic galaxy parameters (primarily size and flux), purity is defined so that for two overlapping objects the purity of the brighter will be higher than the fainter object.

To study the extent of unrecognized blends for the LSST, we perform LSP detection on a test sample of 100,000 randomly selected[5] blends with 2-10 galaxies per group.

---

[4] The purity was computed by `WeakLensingDeblending` package with galaxy SBP rendered to the per-pixel threshold equal to the sky background noise

[5] It is important to randomize the test sample of blended groups since the entries in the cosmoDC2 catalog are ordered by healpix numbers and in redshift slices of size 1. Identifying a portion of the blend groups for analysis without random sampling can introduce spurious selection effects since the `WLD` labels groups based on the brightest

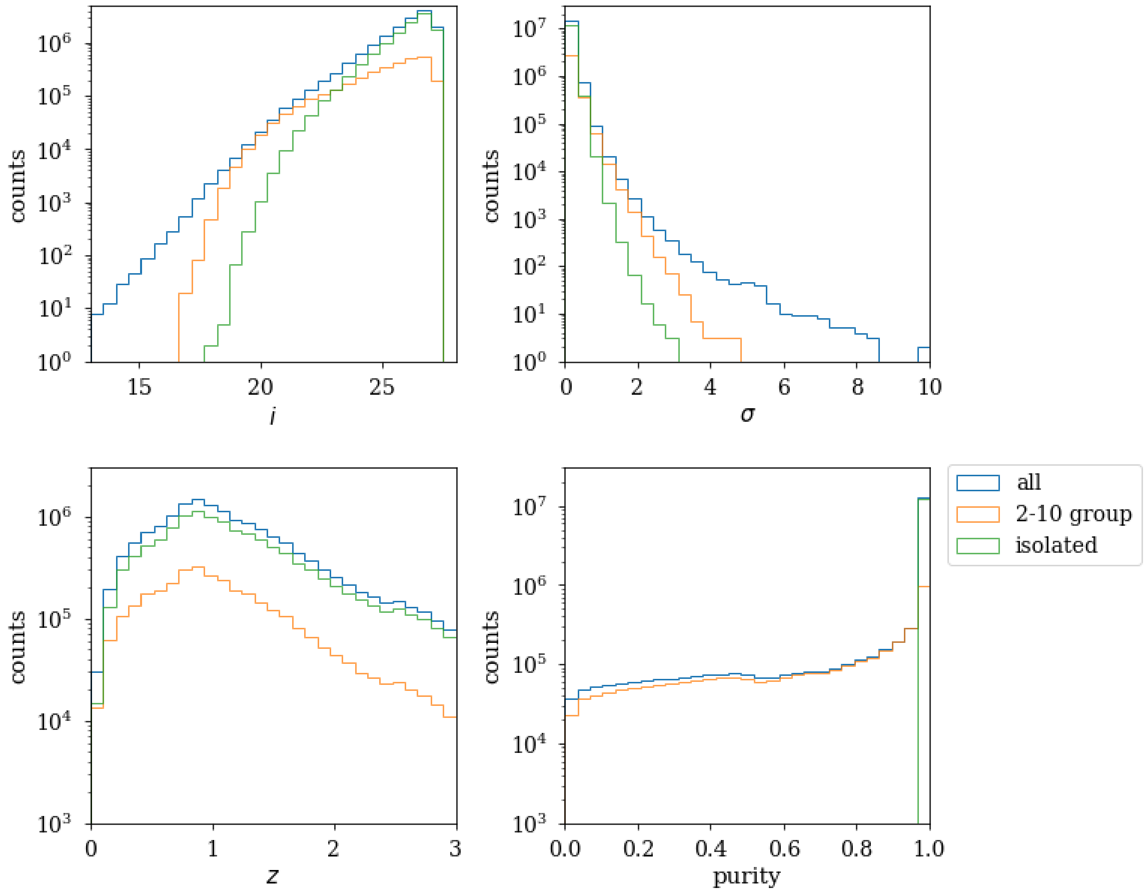


Figure 4.3: Distribution of intrinsic properties for all cosmoDC2 catalog galaxies in the groups defined in Section 4.3.2 (blue), galaxies in groups of size 2 to 10 (orange), and isolated galaxies (green). The intrinsic galaxy properties shown are i-band magnitude (top-left), galaxy size σ_m computed from the determinant of second moment (top-right), redshift (bottom-left), and purity (bottom-right).

4.3.3 Rendering images of the simulated blend scenes

As described above, each blend scene consists of two to ten overlapping PSF-convolved galaxy images drawn on a postage stamp of size 128×128 pixels, with each galaxy profile extending to 5% of the noise level of the sky background[6]. The PSF is a circular profile modeled as a convolution of an atmospheric PSF with an instrumental PSF. The atmospheric component of the PSF is a Kolmogorov profile, with a different size in each filter, based on the fiducial seeing expected for the LSST as described in Ivezic et al. (2008)[7]

The instrumental component of the PSF is modeled as an airy disk profile that encapsulates the diffraction limit PSF for a circular aperture with a central obscuration. For each filter, we use the central wavelength of the band to define the diffraction limit for the primary mirror diameter of 8.36 m, effective area of 32.4 m^2, and 64% obscuration (see Sanchez et al. (in prep) for details).

We add to each pixel Poisson noise corresponding to the sky background brightness level described in Ivezic et al. (2008). The 10-year mean sky-background levels and PSF sizes in each band are shown in Table 4.1.

Table 4.1: Survey parameters for LSST *ugrizy* simulations. The parameter values correspond to the mean values expected for the 10-year LSST survey (see Ivezic et al. (2008)).

Band	Exposure Time (seconds)	Sky Brightness (mag arcsec^{-2})	PSF FWHM (arcseconds)	Zero point (es^{-1})	Extinction
u	1680	22.9	0.77	9.16	0.451
g	2400	22.3	0.73	50.7	0.163
r	5520	21.2	0.70	43.7	0.1
i	5520	20.5	0.67	32.4	0.07
z	4800	19.6	0.65	22.7	0.043
y	4800	18.6	0.63	10.6	0.138

The galaxies are simulated as a single long-exposure image with parameters described in Table 4.1 rather than a co-add of multiple short-exposure images that the real LSST 10-year images would be. Multiple short exposures would result in the PSF and sky background levels not being identical for all the blend images. We mimic this effect by introducing small variations between groups in the the PSF FWHM (≈4%) and exposure time (≈1%). [8] The `LSP` detection is performed on these "background-subtracted" images of the blend scene.

---

galaxy in the group.

[6] Note that the per-pixel threshold of 100% of the sky background is used to define what constitutes the blend group. The `LSP` detection is performed on the blend images where each galaxy profile is rendered out to 5% of the sky noise.

[7] We assume a zenith atmospheric seeing with an airmass of 1.2.

[8] Variations were randomly sampled from uniform distribution of $[-50, 50]$ seconds for exposure time, and $[-0.05, 0.5]$ arcseconds for the PSF FWHM, in each band.

4.3.4 Matching detections to true objects

We calculate the distance between the positions of the detections and the positions of the true object centers using two object-matching algorithms. We work in pixel coordinates and calculate the distance as the Euclidean norm. In Figure 4.4, we illustrate the two matching algorithms with a cartoon example of a galaxy with true center indicated by the red ×, and detected center location shown as the green O.

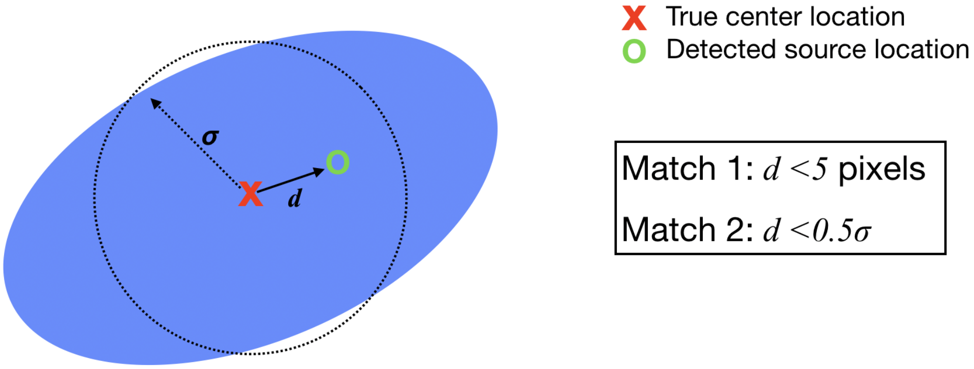


Figure 4.4: Schematic of algorithm for matching true objects to detections. The blue ellipse corresponds to a galaxy with true center located at the red ×. The green circle (○) corresponds to the location of a detected source. The Euclidean distance between the true and detected center locations is denoted by d. The size of the galaxy profile, σ, is denoted by the dashed-black circle.

- Matching algorithm 1: We compute the Euclidean distance d between the true and predicted (detection) centers. For each true object, the detected center with the smallest value of d is assigned as a match, if d is less than 5 pixels.
- Matching algorithm 2: We compute a "normalized" distance d_{norm} by dividing the distance d described in algorithm 1 by the size, σ, of the PSF-convolved true object. The galaxy size is estimated from its second moments computed in the i-band image, and corresponds to the quantity r_{sec} described in Appendix A1 of Chang et al. (2013). The PSF-convolved galaxy size is the square root of the sum in quadrature of r_{sec} and the PSF size calculated from its second

moments. For each true object, the detection with the smallest value of d_{norm} is matched, if the normalized distance d_{norm} is less than 0.5.

Each detection is matched to no more than one true object, but a true object can be matched to any number of true detections. An *undetected object* is defined to be a true object that has not been matched to any detection. If the center of an undetected object is closer than one σ to the location of a detected source, then we classify the system as an *unrecognized blend.* A detection with no true object match is classified as a *spurious detection.* A true object matched to multiple detections is defined as "shredding".

Matching algorithm 1 compares only the center-to-center distance, while Matching algorithm 2 takes into account the size of the galaxy. Neither algorithm takes into account the ellipticity of the galaxy SBP or the total flux.

For alternate matching algorithms the reader can refer to the friends-of-friends algorithm (Berlind et al., 2006), or the preferential matching of brighter objects within a matching radius, which is expected to be implemented in the DES Y3 run of Balrog (Suchyta et al., 2016).[9]

4.4 Detection with the LSST Science Pipeline

The 100,000 simulated images of blended groups, with 2-10 galaxies per group, are analyzed with the LSST Science Pipeline (`LSP`). We described briefly in Section 3.2 how detection, deblending and measurement are performed with the current LSP. The LSP identifies peaks above the detection threshold in the background-subtracted image and computes the "footprint" of the parent object, followed by the deblender that divides the flux in the parent pixels among the deblended children. In this study, we compare the true centers of a galaxy from our input cosmoDC2 catalog, to the measured centers of the deblended children[10].

The LSP detection is performed on a naive multiband-coadd image of the blend scene, where we simply sum the flux per pixel in the six LSST observing bands. Similarly, a naive multiband-coadd image of the PSF and noise variance is also constructed. As described in Section 3.2.2.1 the LSP uses the PSF model as a smoothing kernel during detection. Assuming that the noise is Poisson (which is a good assumption in our case), the noise variance image for a given band is simply the blend scene image before background subtraction.

---

[9] Balrog is a software package that characterizes measurement biases with fake objects embedded in real images.

[10] We included in the analysis `LSP` output catalog entries that satisfy `deblend_nChild` == 0 and `base_SdssCentroid_flag` == `False`. The predicted centers are (`base_SdssCentroid_x`, `base_SdssCentroid_y`) from the `LSP` output catalog.

4.5 Results

We discuss here results for the `LSP` detection algorithm applied to 100,000 simulated images of blended groups with 2-10 galaxies per group. In order to compare to the `LSP` detection performance in the absence of blending, we also perform detection on single galaxies (one galaxy randomly selected from each of 40,000 blend groups) and rendered as isolated objects. Out of the 261,148 galaxies in the 100,000 blend groups, 123,308 ($\approx$ 47%) were correctly detected, while $\approx$ 98% of the isolated galaxy images have a *correct* detection.

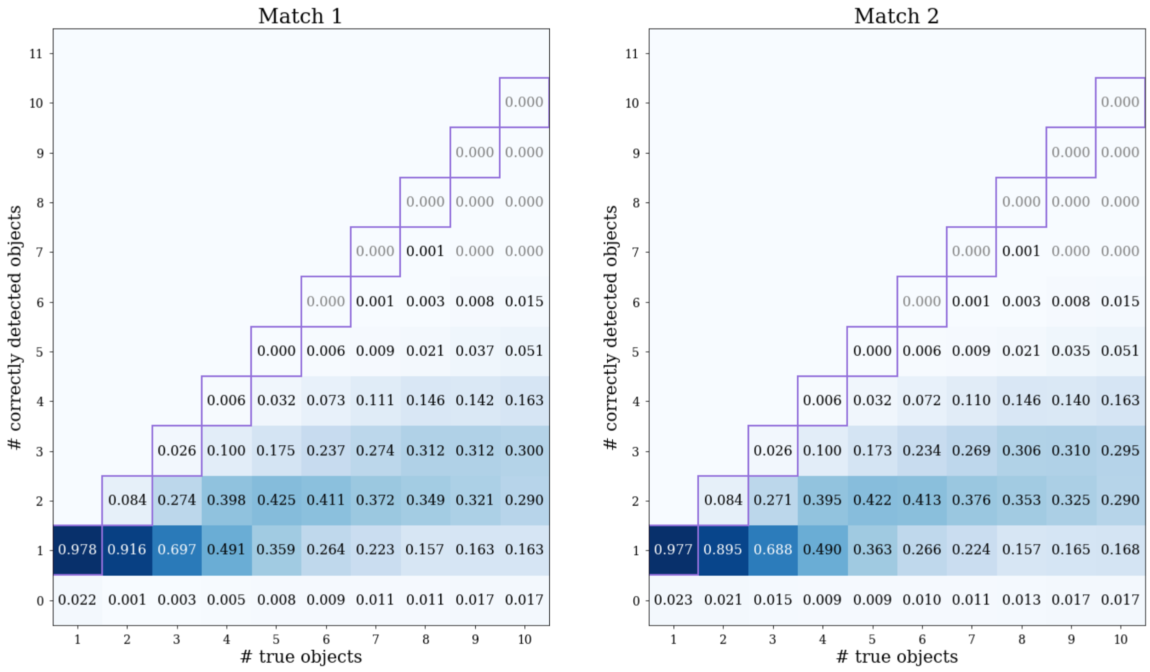
The efficiency for the `LSP` detection algorithm is shown in Figure 4.5 as a function of group size. We define the detection efficiency as the fraction of groups in which a particular number of sources are correctly detected, for each group size. The left and right panels show the detection efficiencies when the detection is matched to the true object with matching algorithms 1 and 2, respectively.

Figure 6.8 shows the detection efficiency matrix for the galaxies in groups with group size ranging from 1 to 10. Each column of the matrix sums to 1. For example, for a two-galaxy blend it is possible to have non-zero values in the elements of three rows in the matrix, corresponding to 0, 1, or 2 detected sources. For these two-galaxy blends, matching algorithm 2 (right plot) predicts that the `LSP` detection correctly identified both objects $\approx$8.4% of the time, correctly detected only one of the two objects $\approx$89.5% of the time, and failed to correctly detect either in 2.1% of cases. For an ideal detection algorithm that detects all the objects correctly, each diagonal element (purple boxes) would have the value 1 and all off-diagonal elements would be 0.

The values of corresponding elements of the efficiency matrices in the left and right panels of Figure 6.8 are mostly very similar indicating that the distance metric used in the matching algorithm does not have a significant influence on the evaluation of detection performance. One significant difference between the two algorithms is for two-galaxy blends, where $\approx$ 2% of groups have no correct detections with matching algorithm 2, but have at least one correct detection with matching algorithm 1. The galaxies in the two-galaxy blends with no correct detections with algorithm 2 are mostly small faint galaxies, with a median size (σ) of 3.3 pixels and median i-band magnitude equal to 25.7.

With both matching algorithms the isolated galaxies were correctly detected $\approx$98% of the time. The values of the diagonal elements for all other group sizes are much smaller because we deliberately selected groups of galaxies that have significant overlap. We would expect the diagonal values to be closer to 1 if we were to use a less stringent definition of overlap in defining a group – i.e., by using a lower value for the per-pixel threshold described in Section 4.3.2. For the rest of this thesis we will use matching algorithm 1 as the default matching algorithm, unless specified otherwise.

We show in the top-left panel of Figure 4.6 the distribution of i-band magnitude for galaxies in groups of size 2 to 10 that are detected (blue histogram) and undetected (orange histogram). As expected the undetected galaxies are, on average, fainter than the detected galaxies; the median



Match 1

	1	2	3	4	5	6	7	8	9	10
11										
10										0.000
9									0.000	0.000
8								0.000	0.000	0.000
7							0.000	0.001	0.000	0.000
6						0.000	0.001	0.003	0.008	0.015
5					0.000	0.006	0.009	0.021	0.037	0.051
4				0.006	0.032	0.073	0.111	0.146	0.142	0.163
3			0.026	0.100	0.175	0.237	0.274	0.312	0.312	0.300
2		0.084	0.274	0.398	0.425	0.411	0.372	0.349	0.321	0.290
1	0.978	0.916	0.697	0.491	0.359	0.264	0.223	0.157	0.163	0.163
0	0.022	0.001	0.003	0.005	0.008	0.009	0.011	0.011	0.017	0.017

Match 2

	1	2	3	4	5	6	7	8	9	10
11										
10										0.000
9									0.000	0.000
8								0.000	0.000	0.000
7							0.000	0.001	0.000	0.000
6						0.000	0.001	0.003	0.008	0.015
5					0.000	0.006	0.009	0.021	0.035	0.051
4				0.006	0.032	0.072	0.110	0.146	0.140	0.163
3			0.026	0.100	0.173	0.234	0.269	0.306	0.310	0.295
2		0.084	0.271	0.395	0.422	0.413	0.376	0.353	0.325	0.290
1	0.977	0.895	0.688	0.490	0.363	0.266	0.224	0.157	0.165	0.168
0	0.023	0.021	0.015	0.009	0.009	0.010	0.011	0.013	0.017	0.017

Figure 4.5: Efficiency matrix for correctly detecting sources with `LSP`. The true and detected sources are matched with matching algorithm 1 (left) and matching algorithm 2 (right). Each element in the efficiency matrix corresponds to the efficiency for correctly detecting different numbers of sources in a group (vertical axis) for different numbers of true sources in the group (horizontal axis). The values in each column sum to 1.

values for detected and undetected galaxies are 23.9 and 26.3, respectively.

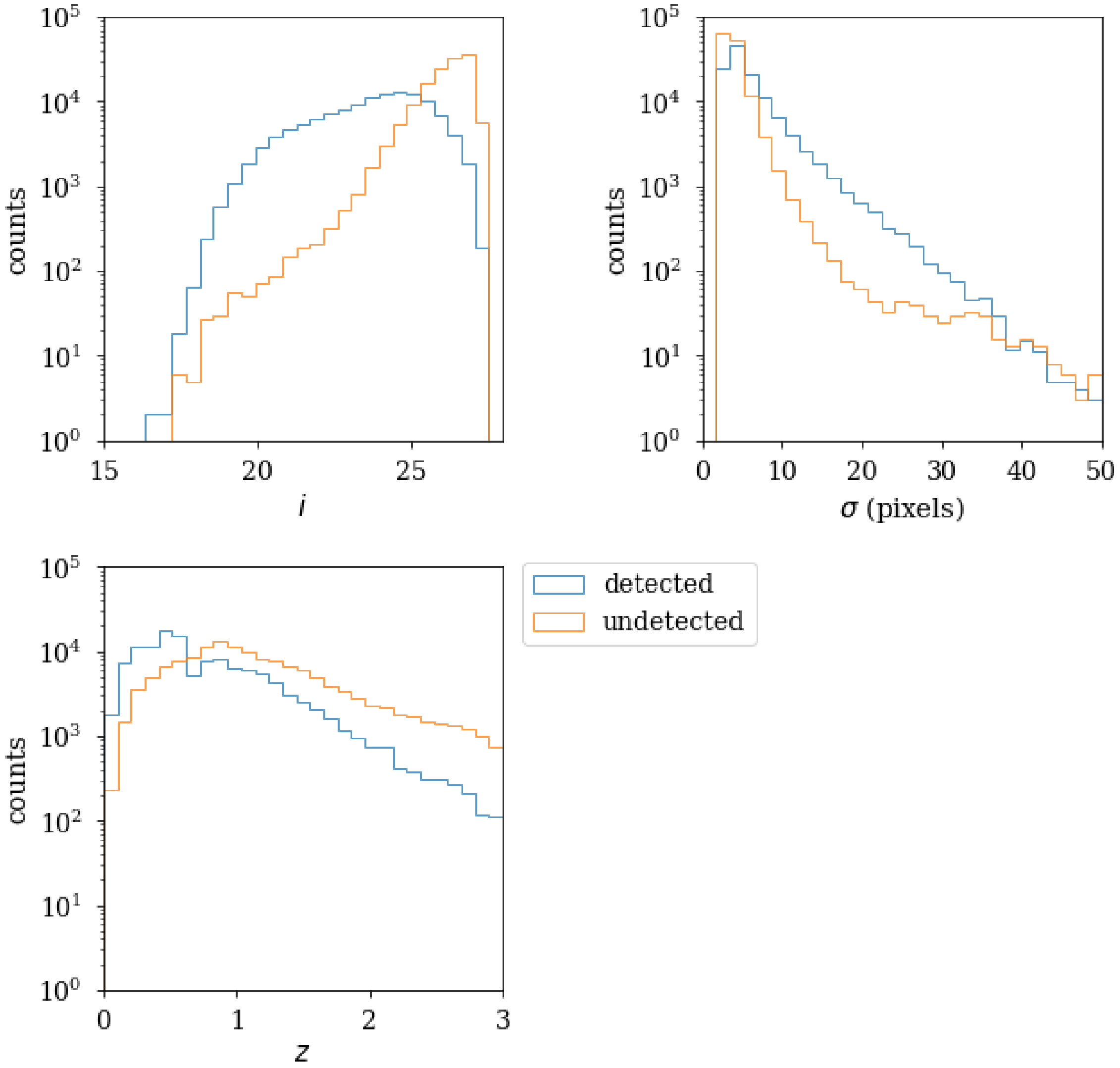


Figure 4.6: Distribution of intrinsic galaxy properties for the set of objects in blended groups of size 2 to 10 analyzed with `LSP`, for detected (blue) and undetected (orange) galaxies. Top-left: i-band magnitude. Top-right: PSF-convolved galaxy size (in pixels). Bottom-left: redshift.

.

The distribution of galaxy size for detected and undetected galaxies is shown in the top-right panel of Figure 4.6. As described above in Section 4.3.4, we define the PSF-convolved galaxy size σ as the the square root of the sum in quadrature of r_{sec} and the PSF size, where the galaxy size r_{sec} is measured from second moments. Almost 98% of galaxies with σ <3 pixels (0.6 arcsec) are undetected by `LSP`.

Because of correlations between redshift and both galaxy size and magnitude, the relative probability of whether a galaxy is detected or not depends on redshift as we can see in the bottom-left panel of Figure 4.6. The high redshift galaxies, which have more lensing information, are also more likely to be undetected.

4.5.1 Potential surrogates for characterizing blending

We explored the potential for a set of galaxy parameters to predict the extent of blending and act as a surrogate that could be used to perform catalog-level studies of the impact of unrecognized blends, for example, on measurements of cosmological parameters. We cosidered the following set of intrinsic properties defined for each galaxy:

- **Distance to nearest neighbor** (d_{neighbor}): We define this as the center-to-center angular distance between a galaxy and its nearest neighbor.
- **Purity**: As described in Section 4.3.2, purity is a dimensionless parameter that quantifies the degree of overlap of each object with neighboring objects.

In order to minimize the impact of *very* faint galaxies – i.e., galaxies that are too faint to be useful in measurements of cosmic shear – on our measures of the ability of a set of galaxy parameters to predict whether blends will be recognized or not, we use only galaxies that meet the criteria for the LSST "gold sample" – i.e., galaxies brighter than 25.3 magnitude in the i band.

We show in Figure 4.7 the distributions of the two blending parameters (d_{neighbor} and purity) for objects in the LSST gold sample that are either detected (blue) or undetected (orange) with the `LSP` detection algorithm. The black-dashed line in the left plot denotes $d_{\text{neighbor}} = 10$ pixels or 2 arcsec. We see that if a galaxy is within 2 arcsec of its neighbor, it is slightly more likely to be undetected; however, d_{neighbor} by itself is not strongly correlated with detection probability. The median values of d_{neighbor} for detected and undetected galaxies are 6.2 and 5.5 pixels, respectively.

We see from the purity distribution in the right panel of Figure 4.7 that, compared to distance to nearest neighbor, purity is more strongly correlated with whether or not a galaxy is detected. Approximately 83% of galaxies with high purity (> 0.98) are correctly detected, for this sample of groups. The median values for the purity of detected and undetected galaxies are 0.99 and 0.61, respectively.

The distributions in Figure 4.7 illustrate that no single parameter can be used as a surrogate to accurately predict whether a galaxy will be detected or not. In other words, we cannot define a selection criterion based on purity or d_{max} alone to identify unrecognized blends. While we do not explore it here, future studies can investigate the potential for machine learning solutions that use these blending-related parameters along with intrinsic galaxy parameters as multi-dimensional input features to predict the probability that a galaxy in a blended group will be detected.

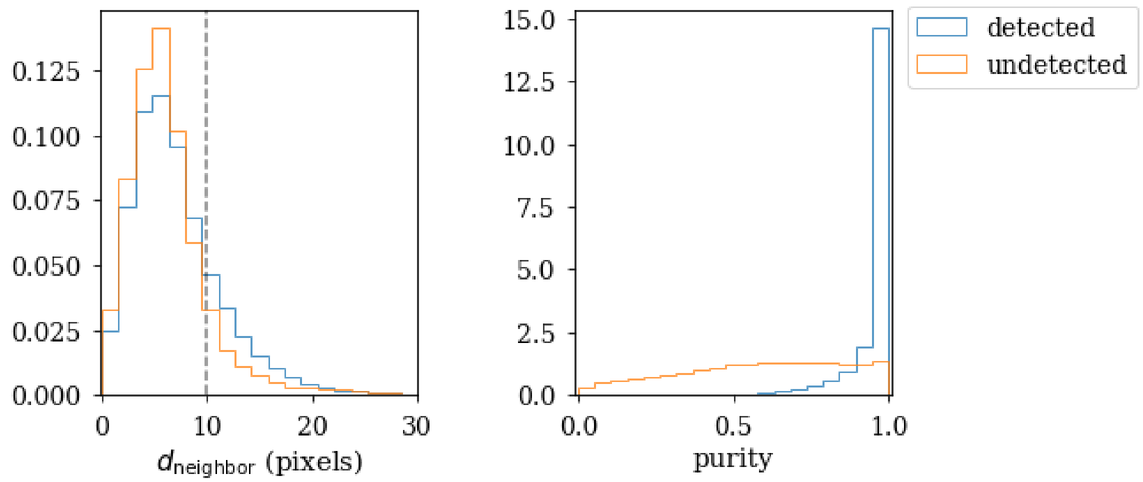


Figure 4.7: Distributions of distance to nearest neighbor (left) and purity (right) for galaxies that satisfy the criteria for the LSST gold sample (i-band magnitude ¡ 25.3), when they are detected (blue histogram) or undetected (orange histogram) by `LSP`.

4.5.2 Characteristics of undetected galaxies in unrecognized blends

We expect that the likelihood of a galaxy being in an unrecognized blend is not uniform across all galaxy properties. For example as seen in Figure 4.6, small faint galaxies are more likely to be undetected in a blended group. Hence, unrecognized blends could introduce selection effects in the WL analysis sample.

We infer the properties of unrecognized blends by comparing several true galaxy properties of an undetected source to those of its closest neighbor in the unrecognized blend. We define a group with at least one undetected object to be an "unrecognized blend" if an undetected object is within one σ distance from the center of a detected object. This is similar to the matching algorithm 2 criteria described in Section 4.3.4, but with a threshold of one σ rather than 0.5σ. Of the 100,000 blended groups in our test dataset, 69,629 are identified as unrecognized blends.

In Figure 4.8, we compare galaxies that are undetected by `LSP` and to their closest matched detected neighbor in the unrecognized blend. The left panel shows the distance d_{match} (in arcsec) between each undetected object and its matched neighbor that was detected by `LSP`. We see from Figure 4.8 that most unrecognized blends ($\approx$ 91%) contain a pair of galaxies with center-to-center distance less than 10 pixels (2 arcsec).

The center and right panels show the difference in magnitude and redshift of the undetected galaxy and the closest detected neighbor. The asymmetry about 0 in the distribution of $\Delta i = i_{\mathrm{undet}} - i_{\mathrm{det}}$ indicates that the undetected galaxy is more likely to be dimmer. The median value of Δi is 2.2 magnitudes. The asymmetry about 0 in the distribution of $\Delta z = z_{\mathrm{undet}} - z_{\mathrm{det}}$ illustrates

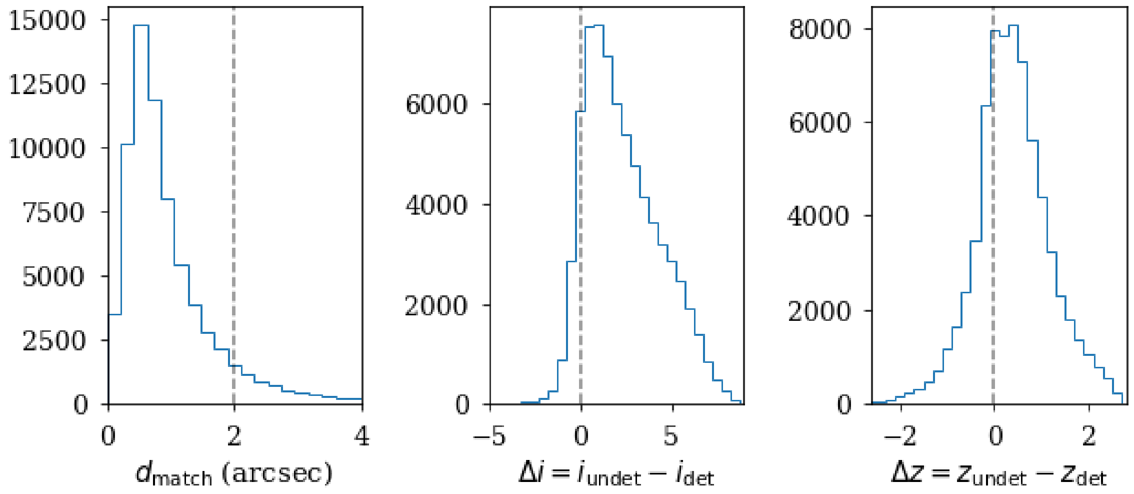


Figure 4.8: Distributions of differences in galaxy parameters for each undetected galaxy and the closest galaxy matched to a source detected by `LSP`, in unrecognized blends with a total of 2 to 10 galaxies. The left panel shows the distance between the center locations of the detected and undetected galaxies. The center and right panels show the difference in the true galaxy magnitudes and redshifts for the pairs.

that the undetected object is more likely to be at higher redshift, as expected given that redshift and magnitude are positively correlated. We see that most unrecognized blends are due to a faint source at higher redshift being undetected because of a brighter neighbor at lower redshift.

The standard deviation of the Δz distribution is 0.78, much larger than the redshift bin size of 0.2 expected to be used in the LSST cosmic shear analysis. This highlights the importance of identifying the presence of unrecognized blends, even if they are then excluded from the final analysis sample.

We described in Section 3.1.2 how unrecognized blends in the same redshift bins would not bias the cosmological analysis since both objects are sheared by approximately the same intervening matter. However, if the objects are at different redshifts then their lensing shear is different and even self-calibrating algorithms such as `MetaCal` would not be able to measure the correct shear. Thus, unrecognized blends could be a source of systematic bias for the LSST WL measurements – biases that may need to be calibrated with external datasets such as high fidelity simulations or deep space-based images.

4.6 Conclusion

We described in this chapter how blending can impact the performance of detection algorithms because of unrecognized blends. While the undetected sources, which are mostly faint, will not be included in the weak lensing analysis catalogs, their presence in a blend with a brighter detected objects can bias cosmological results because of their impact on both cosmic shear estimators and photometric redshift estimators. Therefore it is important to identify unrecognized blends.

By using relatively high per-pixel flux thresholds (equal to the sky noise) to define the extent of a galaxy, we were able to identify blended groups with significant galaxy overlaps such that $\approx 70\%$ of the groups are unrecognized blends. $\approx 80\%$ of the undetected galaxies in these unrecognized blends are in a different LSST redshift bin than their closest matched neighbor.

We measured the performance of the `LSP` detection algorithm for blended groups of galaxies, with a focus on the rate of unrecognized blends. We did not measure the impact on shape and photometric flux measurements. Photometric redshift estimates may be significantly impacted by the presence of undetected sources in flux measurements. To understand the full impact of unrecognized blends on the inference of cosmological parameters, future studies will need to be extended to the evaluation of the resulting errors in shear and photo-z estimates. Future studies could quantify the impact of identifying unrecognized blends and possibly eliminating them from cosmological studies. A study similar to the analysis briefly discussed in Section 4.1 could quantify the reduction of systematic errors in two-point correlation functions.

Chapter 5

`ResidualDetectron`: Object Detection with Neural Networks

In this chapter we describe a convolutional neural network (CNN) model, `ResidualDetectron`, which we developed to improve object detection by identifying unrecognized blends. We first discuss, in Section 5.1, the role of neural networks in astronomy, particularly for object detection. We then give a brief introduction to the theory of neural networks, with special emphasis on CNNs in Section 5.2. We introduce the CNN model, `ResidualDetectron`, in Section 5.4 and describe its architecture in detail in Section 5.5.

5.1 Past work on neural networks in astronomy

Since the 1990s, neural networks have been used to address several astronomical problems (Miller, 1993; Tagliaferri et al., 2003). Some of the earliest applications were in adaptive optics for wave front correction (Lloyd-Hart et al., 1992; Hege, 1992) and cosmic ray detection (Murtagh & Adorf, 1992). There have also been numerous applications to star/galaxy classification using fully connected neural networks (Odewahn et al., 1992; Bertin, 1994; Serra-Ricart et al., 1996), which have had moderate success. Advances in *deep* convolutional neural networks (Lecun et al., 1998) led to improvements in star/galaxy classification (Kim & Brunner, 2017). Deep neural networks have increasingly been used in astronomical studies such as galaxy morphology classification (Dieleman et al., 2015), gravitational lens finding and analysis (Hezaveh et al., 2017; Petrillo et al., 2017; Lanusse et al., 2018), modeling quasar micro-lensing light curves (Vernardos & Tsagkatakis, 2019), exoplanet detection (Pearson et al., 2018), and photometric redshift estimation (Pasquet et al., 2019). The past four years have seen implementations of other neural network architectures for solving astronomical challenges; for example, variational autoencoders (VAE) (Kingma & Welling, 2013) have been used to generate realistic galaxy images (Ravanbakhsh et al., 2016) and generative adversarial networks (GAN) (Goodfellow et al., 2014) have been used to perform deconvolution (Schawinski et al., 2017).

Neural networks are particularly useful in cases where parametric solutions are difficult to define. Thus, it is not surprising that there have been several recent attempts to use CNNs to tackle blending challenges, where the nature of the blends themselves depend on the complex morphologies of the overlapping sources. CNNs have shown promise for detecting and segmenting sources, as well as performing star/galaxy classification on relatively low density fields in DES images (~ 30 galaxies arcmin^{-2}) (Burke et al., 2019), and performing morphological classification (disk, spheroid, irregular, and point source/compact classes) on the CANDELS (Grogin et al., 2011) high resolution images (Hausen & Robertson, 2019). However, these fields are not representative of the high density, low resolution images from LSST with its increased observing depth and larger PSF exacerbating the blending problem, as described in Section 3.1.1.

While many of these past studies have been performed with real survey images, some studies are based on synthetic two-galaxy blends, where two simulated (isolated) galaxy images are superimposed with variable offsets between their centroids. These synthetic blends may not reflect the true blend configurations in real survey fields, but they are a useful way of approximating real blends while generating large training sets of images. Also, using real images would require a detection/deblending algorithm (usually `SourceExtractor`) to generate the training data targets, thereby propagating any limitations of that algorithm to the training set. This problem can be avoided with synthetic blends because the true configuration is already known. In some studies based on synthetic two-galaxy blends, one galaxy is always placed at the center of the postage stamp in order to break the degeneracy between the two objects while labeling their network outputs.

Using synthetic blends of two overlapping galaxies from the Galaxy Zoo project (Lintott et al., 2008), Reiman & Göhre (2019) demonstrated that GANs could be trained to produce deblended images of both galaxies. The study used blends created by selecting the pixel-wise maximum pixel intensity between the two galaxy images, which is not physically motivated. High resolution images of occulting galaxies show that galaxies are semi transparent objects with dust extinction (Holwerda et al., 2007) in the foreground galaxy partially blocking light in the background galaxy.

Another study by Boucaud et al. (2019) used deep networks to estimate flux and segmentation maps on single-band images of blends constructed by adding two isolated CANDELS galaxies. Their dataset did not include highly overlapping or unrecognized blends, and was representative of the galaxy density in the Euclid space telescope. For these condition, the network was successful in making photometric measurements within an uncertainty of ≈ 0.2 magnitudes.

The above two studies used high resolution, high S/N HST images and were aimed at tackling blending problems in space-based images. An ongoing study of using VAEs to produce deblended images of the central galaxy (Doux, C. et al. in prep) uses parametric galaxy simulations of two overlapping transparent galaxies observed at 10-year LSST depths (i.e., with ground-based PSFs).

The authors demonstrated that the VAE network could be trained to predict ellipticities of the central galaxy with a median error of ±0.01. However, in the training and test samples, one galaxy is always centered in the postage stamp image, and traning and testing is applied to only this galaxy; therefore, these studies inherently assume that an initial detection step has already been performed. When the test galaxy is not perfectly centered, the median ellipticity error increases by a factor of ∼5.

5.2 Overview of Neural Networks

Artificial neural networks originated as simplified mathematical models of the human brain. The first layer of the network is genrally called the input layer and all intermediate layers are called the hidden layers, while the last layer is called the output layer. The earliest networks were fully connected networks where, as the name suggests, the components of an input vector are connected to all the "neurons" in the hidden layers. This is illustrated in Figure 5.1 for the example of a two-component input vector $\vec{x} = (x_1, x_2)$ connected to neurons N_1, N_2, N_3 to output $\vec{y} = (y_1)$. The synapses connecting the neurons are characterized by weights linking the input to the neuron and the neuron to the output (indicated by solid lines). These weights determine the importance of each input to the neuron. The weighted output of the neurons passes through an activation function that limits the output of the neuron, typically to the range [0, 1] (Csaji, 2001).

The mathematical theory of neural networks (NN) is based on the universal approximation theorem, which states that a feed-forward network with a single hidden layer containing a finite number of neurons can approximate continuous functions, under mild assumptions on the activation function (Cybenko, 1989; Hornik, 1991; Kratsios, 2019). With advances in network architectures, recent NNs resemble their namesake neurons of the brain less and less, and have evolved from simple fully connected layers to complex models such as a CNN, recurrent neural network (RNN) (Rumelhart et al., 1986), VAE, GAN, etc..

While a complete and detailed overview of neural networks is beyond the scope of this thesis, a brief description of the theory required for understanding the network architecture of `ResidualDetectron` is provided in this section. We will discuss the fundamentals of NN and feed-forward networks, with particular emphasis on CNNs. For a more detailed understanding of the subject, the reader can refer to several excellent resources such as Goodfellow et al. (2016) or Schmidhuber (2014).

Neural networks can be best described as a series of non-linear mappings between the input data x and output y. Each layer consists of a linear transformation, like matrix multiplication, followed by a non-linear activation function. The layers are arranged in a "feed-forward" manner, where the output of one layer is fed as input to the next layer. An n-layer NN can thus be represented as

$$y = \sigma^n(W^n \otimes \sigma^{n-1}(W^{n-1} \otimes \sigma^{n-2}(...\sigma^1(W^1 \otimes x...)))), \tag{5.1}$$

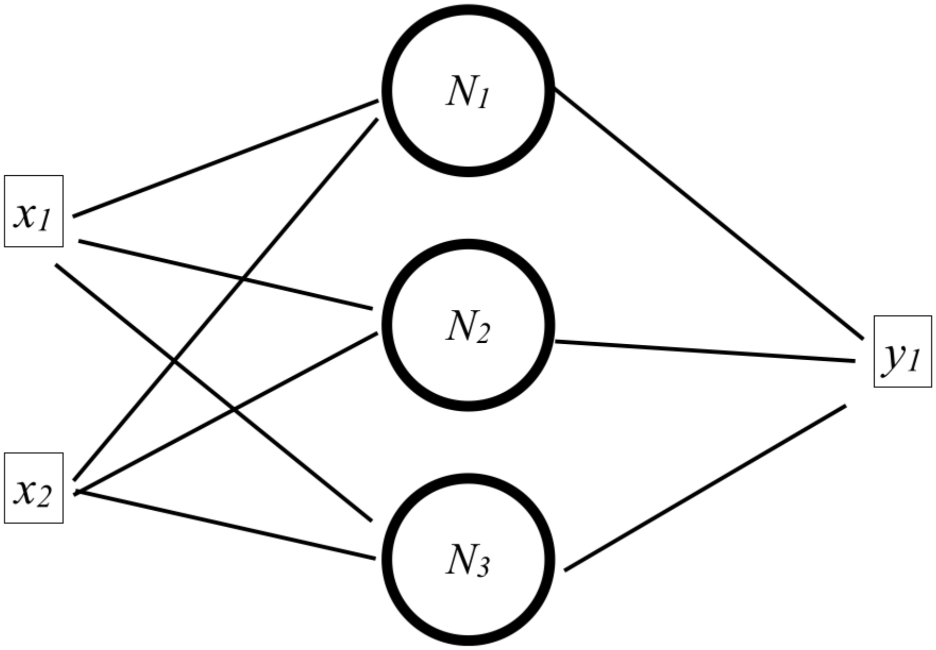


Figure 5.1: Example of a fully connected network with one hidden layer. The inputs $\vec{x} = (x_1, x_2)$ are connected to each of the neurons N. The neuron here includes the linear matrix multiplication and the non-linear activation function.

where W^ℓ are the weights of layer ℓ and σ^ℓ its activation function. The output of a layer ℓ is thus $\sigma^\ell\Big(W^\ell \otimes (\text{output from layer } \ell - 1)\Big)$. $\otimes$ represents matrix multiplication, which could correspond to matrix-vector multiplication in the case of fully connected networks or convolution by a matrix filter in the case of CNNs. By utilizing multiple layers, a neural network is able to adapt to produce more and more complicated mappings between the inputs and outputs. The output prediction $\hat{y}$ of the final layer is compared to the ground truth y to compute a user defined loss function that is optimized during the training process. During training the weights are updated to best match the desired output, using backpropagation of errors and gradient descent algorithms. Gradient descent helps ensure that the weights are updated in a direction that minimizes the loss function. If the training data is comprehensive and the number of weights in the hidden layers sufficient, then the network can learn a non-linear mappings to help make accurate predictions.

5.2.1 Backpropagation

The information from input x propagating through the hidden layers to the predicted output $\hat{y}$ is called the forward propagation step. The backpropagation algorithm (Rumelhart et al., 1986) allows information from the loss function $\mathcal{L}$ to flow backward through the network in order to compute the gradient of the loss function (Goodfellow et al., 2016). The gradient is efficiently computed with respect to the weights of the network by first computing the partial derivative of $\mathcal{L}$ with respect to weights in the output layer and then propagating them backwards via the chain rule of calculus.

5.2.2 Stochastic gradient descent

The simplest version of gradient descent is to change each weight W by an amount proportional to the accumulated $\partial\mathcal{L}/\partial W$. Thus, at each step in the training process the weights are updated by a value,

$$\Delta W = -\alpha \frac{\partial \mathcal{L}}{\partial W}, \tag{5.2}$$

where α is a user defined parameter (called the learning rate) that determines the speed of convergence: too small a value will produce small weight updates requiring more gradient descent steps to converge, while setting it to a value that is too large leads to large weight updates potentially causing the network to overshoot the desired minimum in the loss function.

Computing the loss and gradients for the entire dataset and then updating the weights would be extremely time consuming, slowing the network learning. Thus, to speed up the learning process, while still ensuring that the network converges to a solution, a stochastic approximation of gradient descent optimization is used where the weight updates are estimated from a randomly selected subset of the data (called a mini-batch) (Bottou & Bousquet, 2007).

5.2.3 Activation function

Activation functions introduce non-linearity in the network, without which all the linear matrix operations across the layers would collapse to a single-layer linear operation. Every activation function takes a single number and performs a certain fixed mathematical operation on it. The type of activation function to use depends on the output expected. For example, if a layer output between 0 and 1 is desired then a sigmoid function might be useful; for a desired output in the range [-1 1], we could use the hyperbolic tangent function.

5.2.4 Regularization techniques

One of the dangers of using neural networks is that a large number of trainable parameters might result in over-fitting for the training data. One way to avoid this is with regularization, which imposes certain prior distributions on the model parameters. The most commonly used method is

L2 regularization, which adds the L2 norm (Euclidean norm) of the weight terms as an additional term in the loss function. The L2 regularization loss $\mathcal{L}_{\mathrm{L2}}$ can be written as

$$\mathcal{L}_{\mathrm{L2}} = \frac{\lambda}{2} \sum_{\text{all weights}} |W^2| \tag{5.3}$$

$$= \frac{\lambda}{2} W^T W, \tag{5.4}$$

where λ is a user defined weight decay parameter that determines the amount of regularization.

With regularization, the weight update at each training step is:

$$\Delta W = -\alpha \frac{\partial \mathcal{L}}{\partial W} - \alpha \frac{\lambda}{2} \frac{\partial W^T W}{\partial W} \tag{5.5}$$

$$= -\alpha \Big(\frac{\partial \mathcal{L}}{\partial W} + \lambda W \Big) \tag{5.6}$$

Thus, the network weights are prevented from becoming too large in magnitude; thereby no component of an input training vector can have undue importance on the prediction. The penalty the regularization loss term would lead to the network to take into account all the network connections to small amounts rather than a few of them strongly.

5.2.5 Convolutional neural networks

A Convolutional Neural Network (CNN) is a class of NN most commonly used in image analysis because it uses information from local patterns to extract features from the image (Lecun et al., 1998). It enforces translational invariance by sharing weights (that describe the convolution kernel) across the entire image. Each successive layer receives information from a small local neighborhood of the previous layer, producing a kind of receptive field. The weight sharing reduces the number of free parameters, thereby speeding up training and reducing opportunities for over-fitting.

The forward pass of a CNN involves sliding (or convolving) a filter across the spatial dimensions of the input image, and computing the dot product of the filter and the input at each position. As the filter slides over the image, a 2-dimensional activation map is generated (Lecun et al., 2010).

A cartoon image of a convolutional layer is shown in Figure 5.2. Each filter slides over the spatial dimensions of the input image to produce the feature map. Convolution by multiple such filters introduces a depth dimension to the feature map.

It is important to note that the connections are local only in the spatial dimensions and are always made full in the entire depth dimension. For example, an input RGB image with 3 color channels (depth dimension) when convolved with a filter 5×5 is in fact being convolved with a $5 \times 5\times$ 3 matrix filter with $5 \times 5 \times 3 = 75$ weights.

The output of a convolutional layer is determined by three hyperparameters: depth, stride and zero padding. The depth corresponds to the number of filters we will use, each learning to look for a

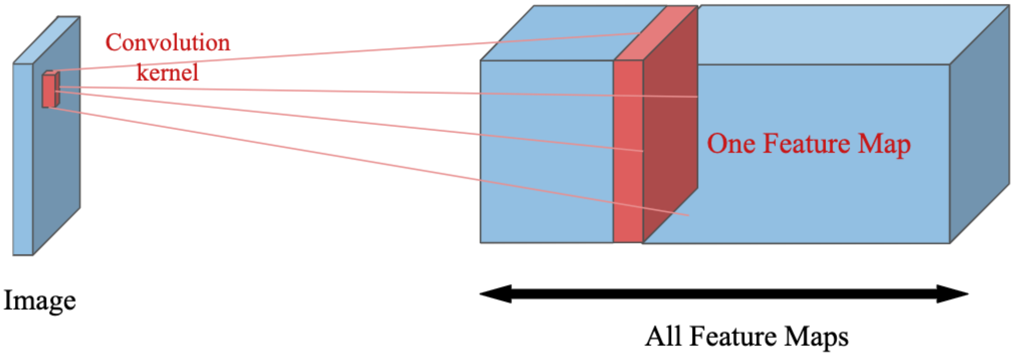


Figure 5.2: Example[1] of a convolutional layer. A convolution filter (small red cube) acts on the image to produce a feature map. Each pixel in the feature map contains contributions from a small local receptive field of the previous layer. Using multiple filters produces the depth dimension of the feature map.

different pattern in the input. The stride determines the number of pixels the filter slides over. For example, when the stride is 2 the filter jumps 2 pixels at a time while sliding. The images can be padded with zeros at the edges before convolution in order to make them larger so that the feature maps include contributions even when the filter is centered on an edge pixel (Simard et al., 2003).

Each convolutional layer creates abstractions based on information from the previous layers. The first layer usually learns low level features like edges, while the deeper layers find patterns in these features, thereby enabling more complex decisions. Eventually a convolutional feature map is produced with spatial dimensions much smaller than the original image, but greater depth. This feature map encodes information about the image in its depth layers while preserving relative location of the target objects in the image. For example, if there was a red square on the top left of the image and a convolutional layer is activated by the red square, then the information for that red square would still be at the top-left of the feature map produced by that layer.

5.2.6 Pooling layers

labelsec:pooling Pooling layers are commonly inserted between convolutional layers in deep networks in order to reduce the spatial size of feature maps by down-sampling, thereby reducing the number of parameters required by each successive layer. For example, 2×2 pooling will reduce the size of the feature map by a factor of 2 in both height and width dimensions. The most common pooling functions are `max` and `average` where the down-sampling replaces the four pixels by the maximum among them, or by the average value, respectively.

5.2.7 Batch Normalization

During training, as the network parameters of the previous layers change, the distribution of activations that a layer receives also changes. Weight optimization becomes more difficult if each layer must also take into account changing scales in its input due to updates of previous layers, slowing the training process. Normalizing activation outputs before they are input to the next layer, called batch normalization (Ioffe & Szegedy, 2015), has been found to improve training speed, performance, and stability of NNs (Santurkar et al., 2018). The normalization step fixes the batch mean and batch variance of each layer's input across an input batch.

5.3 Object detection with neural networks

Object detection is an active area of research in Computer Vision (CV) largely due to CNNs. There are several potential advantages of CNNs over conventional algorithms for object detection in astronomy.

- Non-parametric solutions: Neural networks can describe a problem with non-linear models, which may be well suited to detection problems in astronomical images. For example, a NN could learn to detect spiral galaxies, while they would be difficult to define with a parametric model. As described in Section 5.2.5, the architecture of CNNs enables them to learn spatial features from the input without requiring them to be pre-defined in the model.

- Use of color: CNNs are designed to naturally use 3 or 4 channels as input with the network utilizing information in each channel as well as similarities and differences between them. Traditional detection algorithms like the LSST science pipleine (LSP) and `SourceExtractor` (see Section 3.2.1) work on single-channel images (either a single band or a coadd of multiple bands), and thus their performance is limited. Figure 5.3 is an image observed by HSC that illustrates overlapping galaxies with different colors. Looking at this image, even our brain uses color to differentiate the galaxies.

Most CV solutions have been developed to tackle object detection in photographic images of ordinary "earthly" objects. Thus there are several fundamental challenges in applying these methods to astronomical images:

1. Lack of sharp well-defined edges: The weights that the CNNs optimize during training learn to pick out certain features and patterns in the image, with object boundaries (or edges) being particularly useful. Galaxies and stars, which have continuous brightness profiles, pose a challenge in detecting and assigning regions to a particular source.

2. Transparency: Unlike the targets for most CV tasks, galaxies are not opaque. Therefore, a foreground galaxy does not completely occlude a background galaxy; rather the pixel values in

Figure 5.3: A scene of blended galaxies imaged by the Hyper Suprime-Cam. The overlapping galaxies have different colors making multi-channel information helpful in the detection of blended objects. The overlapping galaxies are semi-transparent, do not have sharp edges, and have a large dynamic range of brightness and sizes, making astronomical images a challenge for CV applications, which are generally developed for earthly objects. Figure from Hyper Suprime-Cam Team.

the region of overlap will include a contribution from the background galaxy that depends on the foreground galaxy transparency. This make it difficult to demarcate overlapping regions. In addition, flux in the overlapping region to depends on the intrinsic profile of the sources as well as the configuration of their overlap.

3. Range of object sizes: The objects of interest in astronomical images can vary in size from a few to hundreds of pixels. Although star profiles are determined by the PSF, their apparent sizes are directly proportional to their brightness. The most important galaxies in the WL survey analysis are small faint galaxies; however, large bright objects can overlap with a large number of sources and lead to significant blending effects. Therefore, it is important to identify both large and small sources to both recognize that image are blended and to avoid selection effects.

4. Brightness: There is a large dynamic range in brightness of astronomical objects in the same field, with pixel values of some bright object being orders of magnitude larger than that for

fainter objects. Since we are more interested in small faint sources, these objects will have a S/N much smaller than training sets typically used for CV tasks.

While machine learning approaches to detection and deblending may not be mature enough to be included in the LSST Science pipeline for initial image processing (for example, the first year of data), these approaches are likely to play a significant role in future LSST operations. The work presented here is a proof of concept demonstration of using neural networks to improve object detection.

5.4 Improving object detection with `ResidualDetectron`

In order to reduce the number of undetected sources due to blending, we develop an iterative method of detection to identify unrecognized blends from residual images – i.e., from the difference between the original image and a model for any detected sources. This involves running two detection algorithms: a primary detection step that identifies peak positions that can then be used to model the scene, followed by a secondary detection on the residual image using a neural network model, which we call `ResidualDetectron`. In theory one could iterate multiple times with the possibility of adding new source on each iteration, to better model the scene. We limit ourselves to only one iterations to reduce computational costs. Note that this iterative detection is different from the iterative fitting method employed in the multi-object fitting (MOF) deblender (Samuroff et al., 2018) described in Section 3.2.3, where each source is fit replacing others with noise iteratively until the collective residual reaches certain convergence criteria.

The `ResidualDetectron` network developed in this thesis is a modification of the Tensorflow and Keras implementation (Abdulla, 2017) of the Mask R-CNN architecture (He et al., 2017) developed at Facebook AI Research (FAIR) to perform object detection, classification and image segmentation.

> Descriptions of significant changes to the Mask R-CNN architecture that have been made in the implementation of `ResidualDetectron` are highlighted in gray boxes in this chapter.

5.4.1 Overview of Mask R-CNN

Mask R-CNN is the most recent advancement in the category of "region-based" convolutional neural network detection algorithms, concisely known as R-CNN (Girshick et al., 2013). Mask R-CNN is a progeny of the Faster R-CNN architecture (Ren et al., 2015), which itself is a "faster" evolution of the Fast R-CNN model (Girshick, 2015). Apart from improvements in detection performance, one of the key distinguishing features of the Mask R-CNN network compared to its progenitors is that the network produces an instance segmentation mask for each detected source, in parallel with performing classification and bounding box prediction.

Mask R-CNN is based on a two-stage approach to finding regions that contain a target object, where the first stage generates a sparse set of candidate proposals that should detect all the target objects while filtering out the majority of negative locations using a fast, but possibly imprecise, algorithm. In the second stage, another network filters the negatives from the input sparse candidates to predict the target objects. One-stage detectors, such as `You Only Look Once (YOLO)` (Redmon et al., 2015) and `Single Shot MultiBox Detector (SSD)` (Liu et al., 2015), on the other hand, are generally faster and simpler, but have trailed in accuracy compared to the two-stage detectors (Lin et al., 2017).

5.4.1.1 The “mask” in Mask R-CNN

Segmentation is the procedure of classifying individual pixels in an image as belonging to a target object. In *semantic* segmentation, pixels are identified as belonging to a class of target objects, but individual instances of objects in that class are not distinguished. Mask R-CNN performs *instance* segmentation in which pixels belonging to each detection of a target object in the class are identified.

To illustrate this we show an example of segmentation maps in Figure 5.4. The left panel is the input image from which we wish to make a segmentation map. The middle panel shows the semantic segmentation of the scene. All pixels in the scene have been classified as different classes, such as table (white), chair (blue), etc.. However each instance of a class has not been differentiated; e.g., all pixels belonging to the class of chair have been marked as blue. Therefore, if pixels of multiple objects belonging to the same class are contiguous, this method would fail to recognize them as separate objects. The instance segmentation map of the scene is shown in the panel at the right. Each object of the chair class has been correctly identified and marked as a different instance of the class. This “mask” correctly identifies the instances, even when segments are contiguous and when instances occlude each other. The primary advantage of instance segmentation is that it permits occlusion of objects. Thus, architectures like Mask R-CNN that permit occlusion are more suited for blending problems.

For our use case of identifying undetected sources in blended images, the segmentation mask is of limited use since our goal is to identify the `presence` of an undetected source. However, even though we do not use the mask branch of the Mask R-CNN model, we benefit from the fact that the architecture is able to correctly detect overlapping objects. We also limit the classification branch of the network to only predict whether a bounding box contains a source or not. We do this because, once again, our goal here is to detect whether an unrecognized blend is present or not. Future implementations could extend this network to a classification problem as well, providing prior information to the next iteration of the deblender on what kind of objects are in the scene or a inform a template model.

Figure 5.4: Illustration of segmentation of a scene. Left: input image. Middle: semantic segmentation of the scene. All pixels belonging to the class chair are labeled blue. Each instance of the chair class is not distinguished from the others. Right: correct instance segmentation, in which instances are correctly assigned within contiguous segments and across occluded regions. Figure from Silberman et al. (2014).

5.5 `ResidualDetectron` network architecture

`ResidualDetectron` takes residual images as input, and outputs square bounding boxes at the positions of the sources that were undetected in the primary detection step. The first stage of the `ResidualDetectron` architecture is called a Region Proposal Network (RPN), which proposes candidate object bounding boxes. The second stage or the “head”, extracts features from each candidate box and performs classification and bounding-box regression. The RPN comprises of a convolutional backbone architecture that performs feature extraction over an entire image. The proposed regions that are further analyzed in the second stage are called Regions of Interest (ROI).

5.5.1 Network Input

As described above `ResidualDetectron` is trained to predict undetected objects from the residual image of the blend scene. We will discuss in detail in Section 6.1 the motivation and methodology of producing the residual images. In this chapter we discuss the network architecture without going into the details of the input residual image.

Each blend image is modeled by the multi-band deblender `SCARLET` (see Section 3.2.3.2) as the sum of extended sources at each detected location . The residual image is the difference between the blend image and the model image. In addition to the residual image, the model image of the scene is also input to the network. Similar to Chapter 4, the blend scenes are rendered on postage stamps of size 128×128 pixels. The residual and model images are rendered in each of the six LSST bands *ugrizy*. An example of the $128 \times 128 \times 12$ (six-band residual + six-band model) image input to `ResidualDetectron` is shown in Figure 5.5.

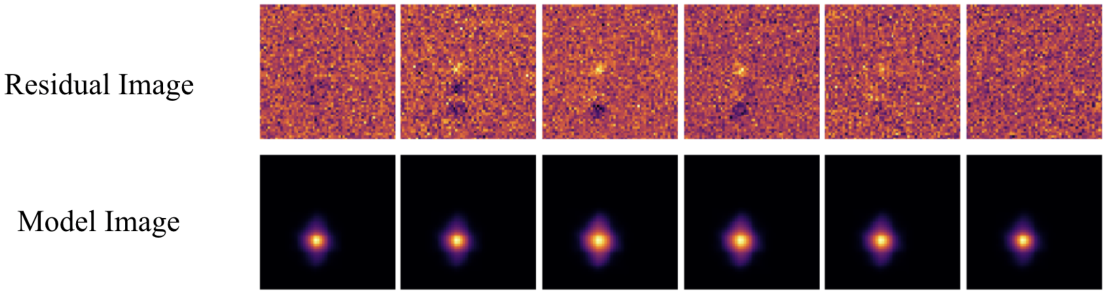


Figure 5.5: Example of six-band residual (top row) and model (bottom row) images input to `ResidualDetectron`.

5.5.2 Region proposal network

The first stage of `ResidualDetectron`, called a region proposal network (RPN), is a neural network that scans the image in a sliding-window fashion and finds areas that contain objects. Importantly, the task of the RPN is not to identify the object, but instead to identify ROIs that are likely to contain an object.

5.5.2.1 Anchor boxes

The architecture of CNNs requires fixed dimensions for the input images and the output of each layer. This is particularly restrictive for detection problems in which the number of sources in an image is not known a priori and can vary from image to image. One way around this problem is the use of “anchor boxes”, which are boxes of variable known sizes spaced uniformly across the image.

An example of this is shown in the left panel of Figure 5.6 where anchor boxes of four different scales and with three different aspect ratios are centered on an image of size 128 × 128. The right panel of Figure 5.6 shows square anchor boxes of size 4 pixels uniformly spaced across the image. For visual clarity we chose a smaller anchor box here because the larger boxes would overlap.

Using anchor boxes for object detection switches the problem from attempting to identify all sources in the image to a classification problem to determine whether a given anchor box contains an object or not. However, this raises other issues since an anchor box could contain more than one object, or more than one anchor box could enclose a single object. These problems are mitigated by using anchor boxes of varying sizes as shown in Figure 5.6.

An anchor box is matched to the ground truth bounding box for a target object based on its Intersection over Union (IoU), also called the Jaccard index. IoU is, as the name suggests, the ratio of the area of the intersection of two boxes to the area of their union. All anchor boxes with IoU $>$ 0.5 are marked as positive identifications while those with IoU $<$ 0.3 are marked as

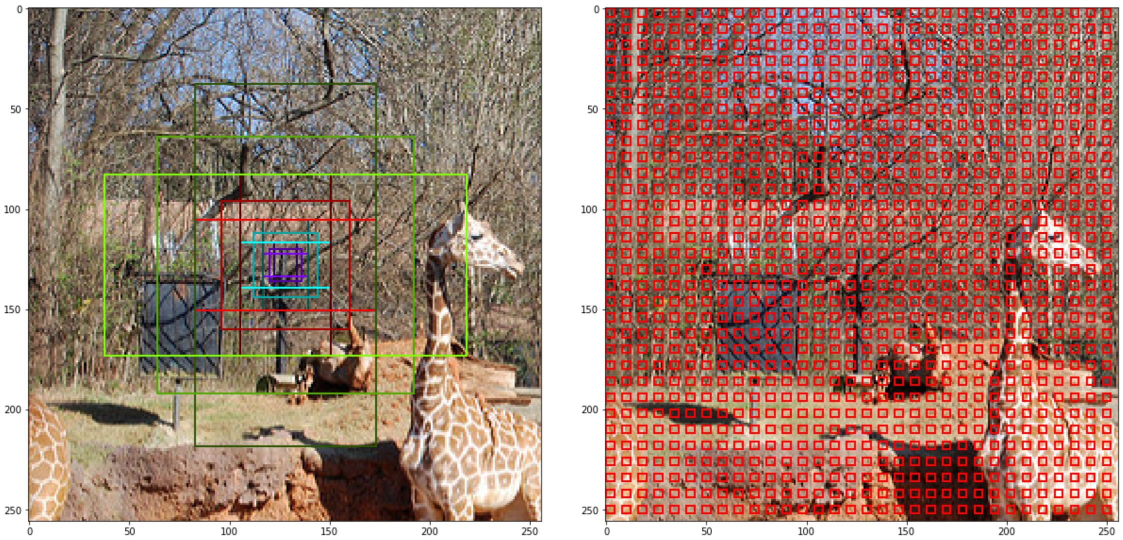


Figure 5.6: Example of use of anchor boxes in object detection. Left: anchor boxes of four different scales (16, 32, 64 and 128 pixels) and three aspect ratios (1:2, 1:1 and 2:1), centered on the image. Right: square anchor boxes of size 4 pixels spaced uniformly across the image.

negative. If no anchor box has IoU greater than 0.5, then the anchor with the highest IoU value is marked as positive. The remaining anchor boxes are marked as neutral and do not contribute to the training. We do not alter the IoU threshold values of 0.5 and 0.3 defined in the original Mask R-CNN architecture. During training the RPN aims to identify all the positive boxes and pass them to the next stage.

The left panel in Figure 5.7 shows an example target bounding box (green-solid square) overlaid on the input image. Also shown are 10 random anchor boxes of different scales. Since these anchors do not overlap with the target their IoU is less than the threshold of 0.3 and they are marked as negative anchor boxes. The right panel in Figure 5.7 shows a zoom-in on the central region of the image. Five anchor boxes that overlap with the target box are shown as dashed squares, with their IoU with the target bounding box shown at the bottom of each anchor box. Since none of the IoUs are greater than 0.5, the anchor with the highest IoU – i.e., the yellow anchor box with IoU equal to 0.49 – is marked positive.

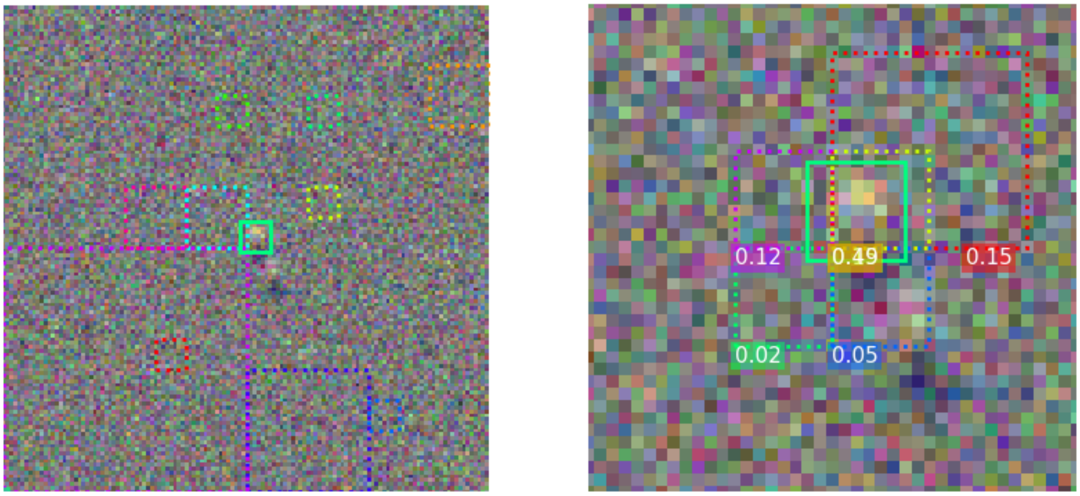


Figure 5.7: Example of anchor boxes in `ResidualDetectron`. Left: Ten randomly selected anchor boxes overlaid on a residual image. Right: zoom-in on central region of the left panel. Anchor boxes that overlap with the target bounding box are shown as dashed squares, with their IoU values indicated at the bottom of each box. The bounding box corresponding to the target undetected object is shown as a green square (solid outline) in both panels.

The original Mask R-CNN architecture was designed with five anchor boxes of scales 32, 64, 128, 256 and 512 pixels, each with multiple aspect ratios (1:2, 1:1 and 2:1). As described in Section 6.3.2, since our target boxes are always square in shape, we use only square anchor boxes in `ResidualDetectron`. Also, our targets are primary small undetected galaxies with size comparable to the PSF ($FWHM \approx 0.7$ arcseconds). Since the LSST pixel scale is 0.2 arcseconds, these objects will be only a couple of pixels across. Therefore we choose smaller anchor boxes in `ResidualDetectron`, with 4, 8, 16, and 32 pixels on a side, as shown in Figure 5.8. This can be compared to the larger anchor boxes in the default implementation shown in Figure 5.6 with sizes 16, 32, 64, and 128 pixels.

Another major difference is that we do not allow any spacing between the grid positions of anchor boxes of a given scale and aspect ratio as shown in the right panel of Figure 5.6. Also, we ensure that anchor boxes of a given scale cover the entire image uniformly and do not leave gaps in coverage like those at the bottom and right edge of the left panel in Figure 5.6. The anchor boxes in `ResidualDetectron` form a tight, non-overlapping grid-like pattern without gaps.

The first task of the RPN is to identify whether an anchor box contains an object or not – i.e., to classify each anchor box as a positive or negative detection. If an anchor box does contain an

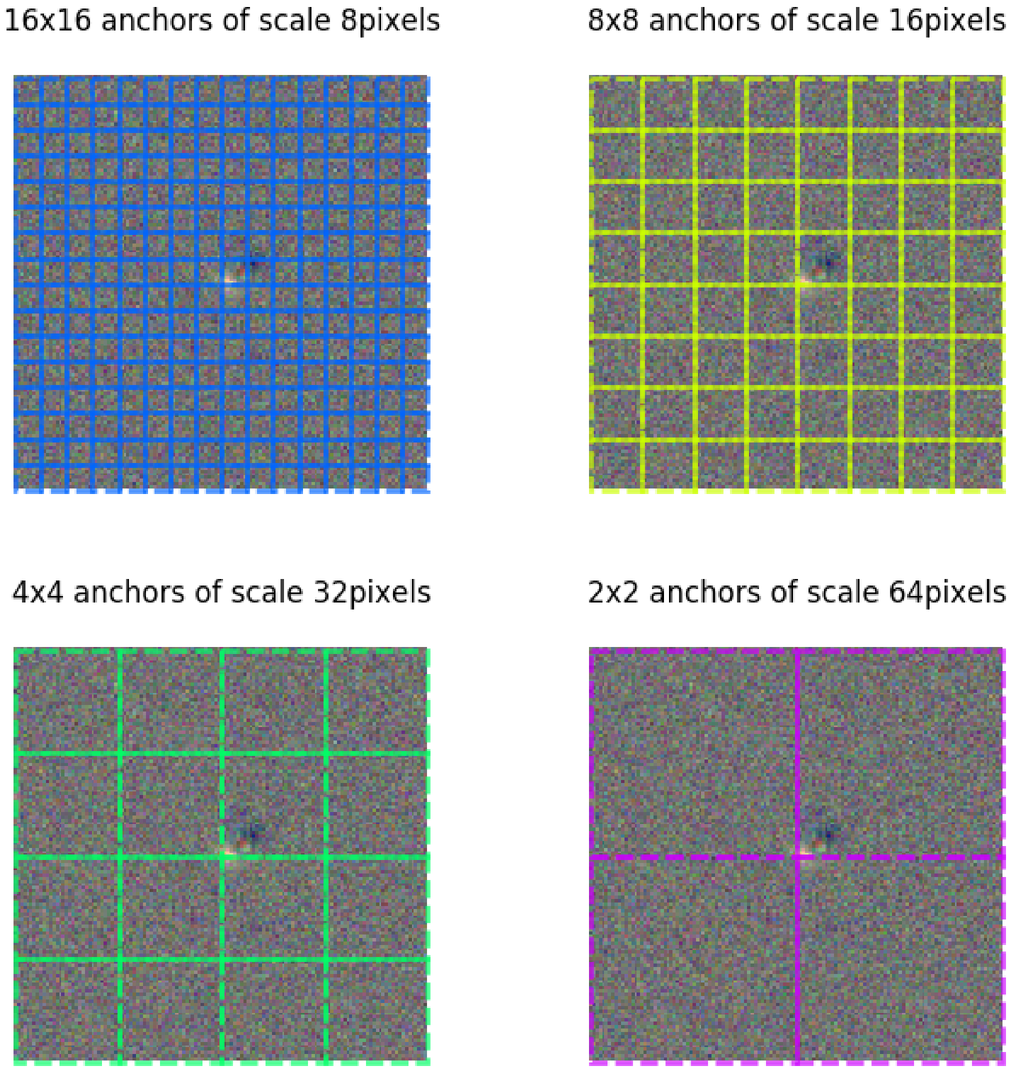


Figure 5.8: `ResidualDetectron` performs detection with four layers of FPN (see Section 5.5.2.2) on 256, 64, 16, and 4 anchor boxes with scales 8, 16, 32,and 64 pixels, respectively.

object, then the next task is to modify the location and size of the anchor box to match the target ground truth box. These refinements to the anchor box parameters are expressed in the following form:

$$\begin{aligned}\Delta x &= (x - x_a)/h_a \\ \Delta y &= (y - y_a)/h_a \\ \Delta h &= \log(h/h_a) \end{aligned} \tag{5.7}$$

where x (x_a) and y (y_a) are the x and y coordinates of the center of the ground truth box (anchor box) and h (h_a) is its size. Thus, the RPN needs to predict the Δ values for an anchor box so that it matches the ground truth. The left panel of Figure 5.9 shows the Δ values that will match the anchor box (dashed) to the ground truth (solid).

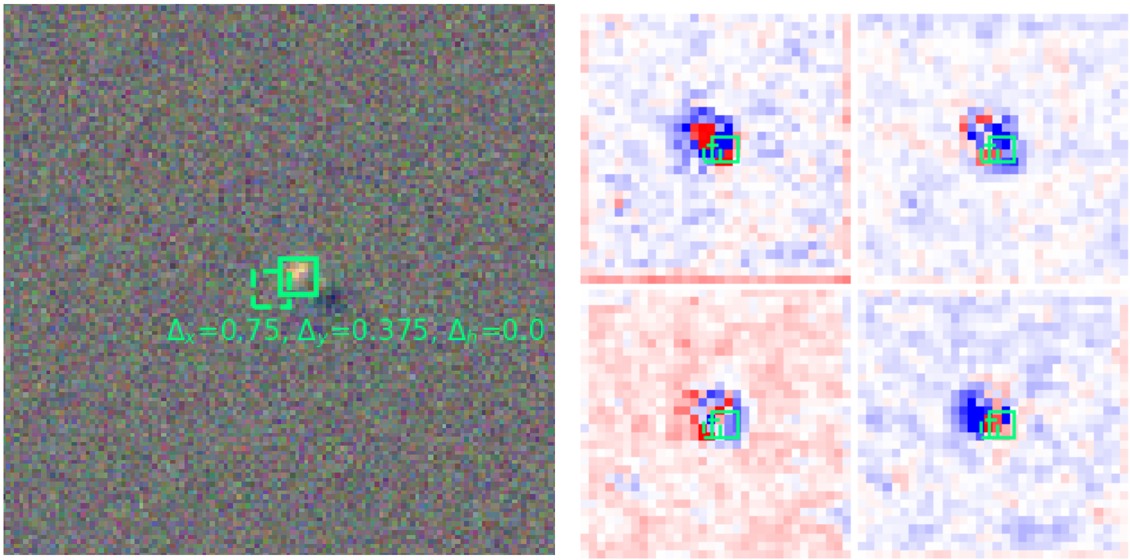


Figure 5.9: Example of anchor box (dashed) and refinements to match target bounding box of undetected object (solid). Left: the residual image of a two-galaxy blend scene with one undetected object. The refinements to match the anchor box (Δ_x, Δ_y, Δ_h) to the target box are also shown. Right: four random feature maps from the FPN P2 layer (see Section 5.5.2.2 and Figure 5.13 for more details about this feature map). Overlaid on each map is the anchor box and the target bounding box. The relative position of the bounding boxes remain the same in both the input image (left) and the feature maps (right).

Although the anchor boxes are shown drawn on the input image itself in the right panel of Figure 5.7, the RPN does not actually scan over the input image directly. Rather, it scans over the backbone feature maps using a network called a feature pyramid network (FPN) (Lin et al., 2016), which we describe in the next section. The right panel of Figure 5.9 shows the anchor boxes overlaid on four feature map layer images. Note that the feature map is a 32×32 image while the input image here was 128×128. The size of the anchor boxes have also been reduced proportionally so that the box represents the same relative region in the feature maps.

5.5.2.2 Feature pyramid network

While translational invariance is a key feature in the success of CNNs, they suffer from not being scale invariant. As described in Section 5.2.5, a deep CNN converts images into feature maps that contain information useful for the desired task. The network gradually shrinks the feature map size while increasing the depth dimension with the inclusion of more layers. This can be seen in the left panel of Figure 5.10, which shows feature maps generated by a deep CNN. The bottom layer is the input image and as we move to the top, each successive layer is a feature map produced deeper in the network. These feature maps have different spatial resolutions and semantic information. For example, the shallower layers have higher spatial resolution but only contain low-level structures, which limits their accuracy in object recognition. Deeper layers, however, are designed to detect high-level features, and therefore have more semantic information, but with lower spatial resolution. In some architectures, features from only the deepest layer used to make output predictions (Lin et al., 2016), as shown in the left panel of Figure 5.10. Since the spatial footprint of objects will be significantly smaller in the deepest layers, such networks perform better for detecting larger objects, potentially introducing selection effects based on object size. The "feature pyramid network" (FPN) architecture overcomes this by introducing additional information pathways while also using all the feature map layers for making predictions, as illustrated by the pyramid of feature maps on the right side in the right panel of Figure 5.10.

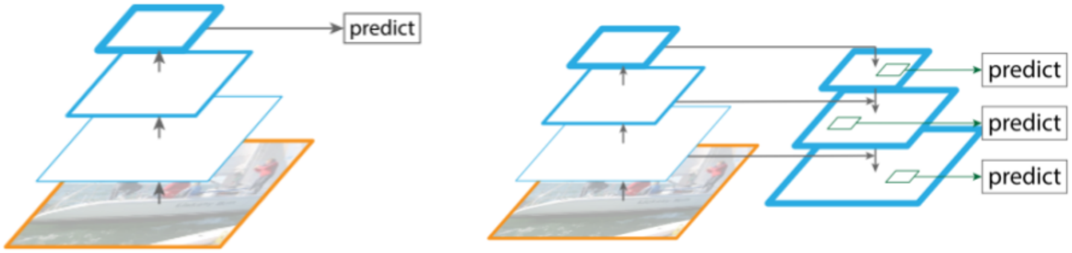


Figure 5.10: Left: Feature map of deep CNN. Right: Feature map of a feature pyramid network (FPN). Figure from Lin et al. (2016).

The FPN architecture primarily consists of three parts as shown in more detail in Figure 5.11: 1) a bottom-up pathway, 2) a top-down pathway, and 3) a lateral connection between the two. The bottom-up pathway (left pyramid) is a feed-forward computation of a backbone deep CNN with the spatial resolution of features decreasing by two with each successive layer as one goes deeper in the network. The top-down pathway (right pyramid) produces higher resolution features by up-sampling by two the spatially coarser, but semantically stronger, feature maps from higher pyramid levels. Each lateral connection then merges feature maps of the same spatial size from the bottom-up pathway and the top-down pathway. A 1×1^2 convolution of the bottom-up feature map is performed to ensure that the feature maps have the same dimensions so that they can then be added element-wise to produce the feature map for that layer.

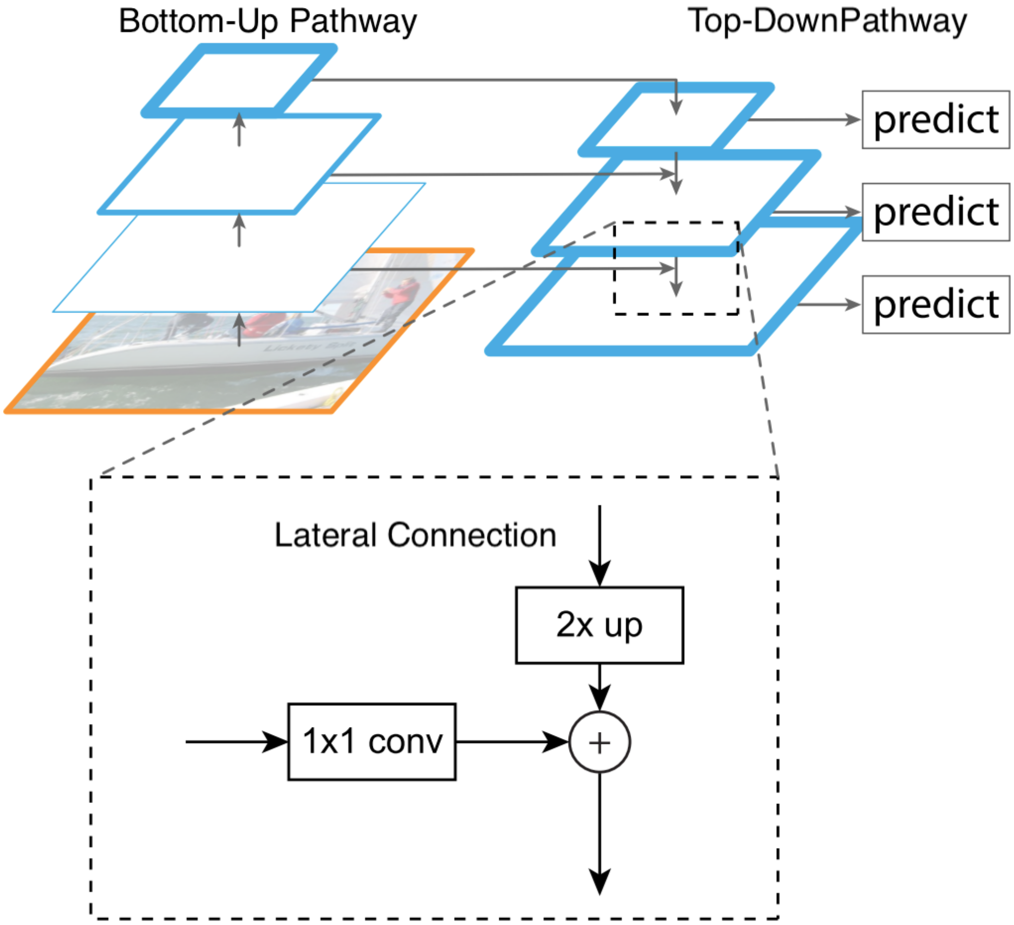


Figure 5.11: Building blocks of a feature pyramid network (FPN) showing the feature map from a bottom-up network connected to a top-down pyramid by a lateral element-wise addition function. Figure from Lin et al. (2016).

Unlike deep CNNs where predictions are based only on the deepest layer, FPN predictions are based on all levels of the feature pyramid, since each level has access to both lower and higher level features. Thus, FPN improves feature extraction, thereby leading to increased object detection

---

[2] Convolution by a 1×1 kernel, identical to a fully connected layer, is generally used to obtain an output layer of a desired depth without altering the spatial dimensions.

performance.

In Section 5.5.2.3 below, we will describe in detail the ResNet backbone network – i.e., how each layer of the bottom-up pathway is created. Here we will simply represent the feature maps in each of the bottom-up layers as R_n, where $n \in \{1, 2, 3, 4, 5\}$ is the number of the layer in the hierarchical model. A schematic of the FPN in `ResidualDetectron` is shown in Figure 5.12. The first row shows the bottom-up pathway features R_n. The input image of size $128 \times 128{\times}12$ goes through successive layers of the backbone to generate feature maps that have decreasing spatial dimensions but increasing feature depth. The spatial dimension is halved with each successive layer, such that the final layer, R_5, has dimensions $4 \times 4{\times}2048$.

As we go down the top-down pathway, we up-sample the previous layer by a factor of 2 using nearest-neighbor up-sampling. All the C_n layers have depth dimension of 256 and each lower layer increases the spatial dimension by two. A 1×1 convolution of R_5 converts the $4 \times 4{\times}2048$ feature map to C_5 with dimensions $4{\times}4{\times}256$. The 1×1 convolution is necessary to convert the 2048 feature depth dimension of R_5 to match the 256 dimension of the top-down layers. A lateral connection merges R_4 and C_5 to produce C_4, and similarly for C_3 and C_2. A 3×3 convolution is applied to all merged layers to reduce the aliasing effects produced by merging with an up-sampled layer. This produces the pyramid feature map, P_n, for the $n \in \{2, 3, 4, 5\}$ layers. P_6, which is generated by max pooling of P_5, is used only to predict RoIs, and is not used later in the head branch.

The feature pyramid network is not an object detector by itself, but rather a feature detector that works with an object detector. The FPN produces multiple feature map layers, $[P_2, P_3, P_4, P_5, P_6]$, and feeds them into another detector that uses convolutions to detect objects and generate RoI. The feature map size for layer $n \in \{2, 3, 4, 5, 6\}$ is $h_n = \{32, 16, 8, 4, 2\}$. The same detection network is applied to all feature maps. An example of an FPN detection network is shown in Figure 5.13 for one layer of the feature map with dimension $h_n \times h_n \times 256$. The object detection network applies 3×3 convolutions over the feature map followed by two separate 1×1 convolutions: one for class predictions and another for bounding box regression.

A $(h_n \times h_n \times 256)$ feature map would generate two $h_n \times h_n$ matrices at the end of the classification branch, one for the positive identification class (contains object) and the other for the negative class (background). The values corresponding to each of the $h_n \times h_n$ pixels in the two matrices are converted to two scores using the "softmax" function or the normalized exponential function. The softmax function on network prediction $z = (z_1.z_2, ...z_K)$ for K classes ($K = 2$ in our case) is defined as

$$s(z_i) = \frac{e^{z_i}}{\sum_1^K e^{z_k}}, \tag{5.8}$$

for $i = (1, 2, ..K)$.

Since we have only two classes, corresponding to positive and negative identification, the class-prediction network outputs two scores corresponding to each pixel in the feature map. The positive class score p can be interpreted as the probability of the prediction being an object (and $1 - p$ the

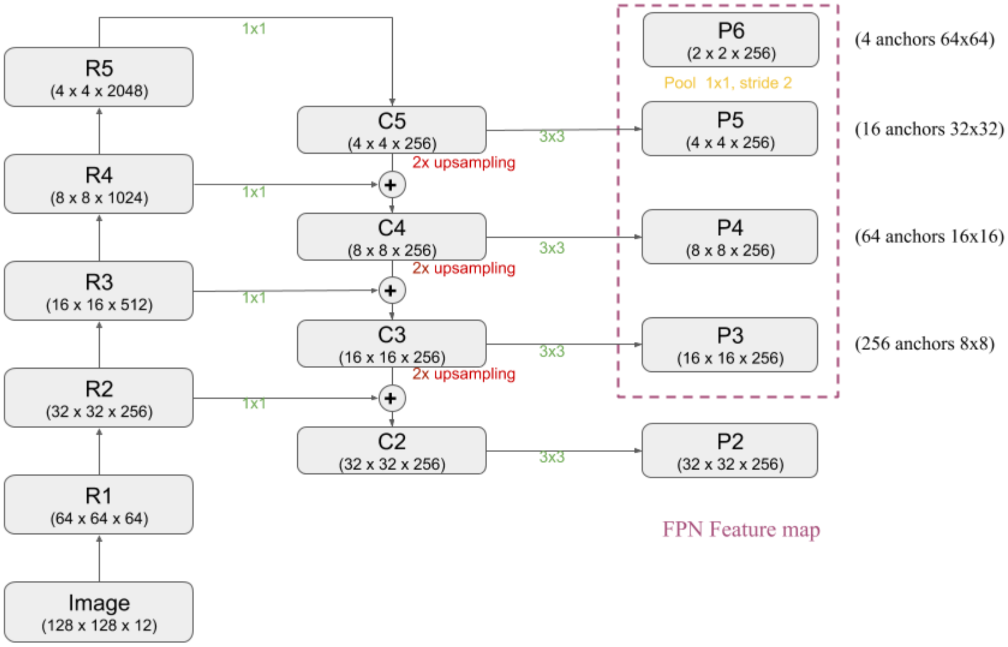


Figure 5.12: Feature extraction with FPN in `ResidualDetectron`. The left side of the schematic illustrates the bottom-up pathway shown as feature maps generated by the ResNet backbone. The middle column shows the top-down pathway with layers labeled C_n for layer $n \in \{2, 3, 4, 5\}$. The lateral connections merge the two pathways to produce the FPN feature maps for the different layers P_n, shown in the right column. The text on the far right gives the number and size of the anchor box corresponding to each layer.

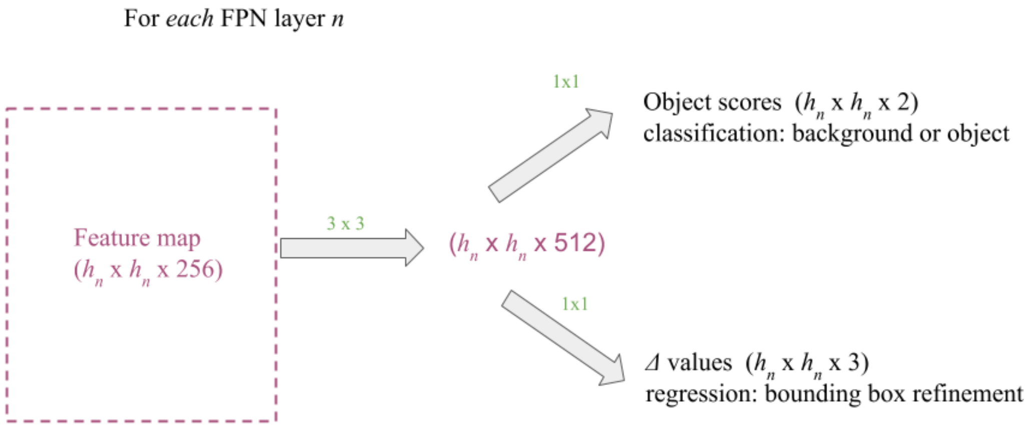


Figure 5.13: Schematic of object detection with the region proposal network (RPN). A CNN applied to the feature map generates two outputs: a score classifying the anchor box as containing the object or not, and the bounding box refinements (Δx, Δy, Δh) to the position and size of the anchor boxes.

probability of it being background). The two scores are thus in the range $[0, 1]$ and sum to 1. A higher positive class score would indicate that the corresponding anchor box contains an object, while the inverse would predict no object.

The second output of the RPN detector is the set of refinements to the position and size of the anchor boxes required to match the target ground truth box, as described in Equation 5.7.

Each layer in the pyramid feature map corresponds to anchor boxes of a particular size while each pixel in the classification and box refinement matrix corresponds to a particular anchor-box location. The anchor-box size for each FPN layer is shown in the rightmost column of Figure 5.12. For example, the largest feature map P_2, corresponds to anchor boxes that are 4 pixels on a side. Each of the 1024 pixels in the classification network of P_2 predicts whether the anchor box corresponding to that location contains an object or not.

5.5.2.3 Network backbone: ResNet

The choice of backbone architecture for the FPN depends in part on the the user's computing resources. The original Mask R-CNN model used ResNet-101, which is an implementation of the residual neural network (ResNet) (He et al., 2015), a popular architecture for computer vision tasks.

ResNet was a break through architecture due to the presence of "skip connections" that helped solve the problem of vanishing gradients. Deep networks are difficult to train because, as the gradient is back-propagated to the earlier layers, repeated multiplication makes the gradient very small. As a result, as the network goes deeper, its performance begins to saturate or can even rapidly degrade. ResNet overcomes this challenge by introducing "residual blocks"[3] that include skip connections. These shortcut connections simply perform an identity mapping, and their outputs are added to the outputs of the stacked layers. The full ResNet architecture stacks residual blocks made up of multiple 3×3 convolutional layers. Periodically, the number of filters is doubled and the feature map is down-sampled spatially by a factor of 2.

A common practice in computer-vision neural network models is to use "transfer learning", where a known architecture with pre-trained weights is defined as the backbone of the network and only the final layers of the model are modified and retrained to perform the particular task at hand. This backbone model is usually pre-trained on a large dataset for which benchmarks for model performance have been set. For example, weights in the Mask R-CNN ResNet-101 model were trained on Microsoft COCO (Common Objects in Context) (Lin et al., 2014), a dataset of 328,000 3-channel (RGB) images with 2,500,000 labeled instances of objects with bounding boxes. The advantage of transfer learning is that the network already knows how to recognize or identify features. Good results can be achieved by changing and re-training only the final layers on the user-specific dataset.

> ResNet-101 is a state-of-the-art classification network that was trained on 3-channel images. However, the `ResidualDetectron` we are developing uses 12-channel images (6-band residual and 6-band `SCARLET` model images). Therefore, we do not use transfer learning from the original Mask R-CNN model in the training of our network. Instead, we train the network from scratch.
>
> ResNet-101, with its 101 layers containing ≈ 40 million trainable parameters, is very difficult and computationally expensive to fully train. Therefore, we use a smaller network, ResNet-41, with ≈ 18 million free parameters, as our backbone architecture.

5.5.3 Extracting regions of interest from feature maps

The region proposal network (RPN) object detector generates bounding boxes corresponding to proposed regions of interest (RoI) that are to be passed to the second stage detector or head branch layers. The corresponding RoI features are extracted from the RPN feature maps. However, the FPN produces four feature maps corresponding to the four layers described in Section 5.5.2.2, with different spatial dimensions. We follow the method used in He et al. (2017) and assign an RoI of

---

[3] Not to be confused with residual image input to `ResidualDetectron`, which is the difference between the blend image and its modeled scene.

size h (on the input image to the network) to the P_n level in our feature pyramid using the formula:

$$n = 4 + \log_2(h/224), \tag{5.9}$$

such that an RoI of size 224 pixels is assigned to P_4. Smaller and larger RoIs will be mapped to the lower $[P_2, P_3]$ or higher $[P_2]$ layers.

Each proposed region is analyzed by the same head branch network; therefore the inputs must be the same size. However, the features extracted corresponding to the RoI proposals are of different spatial dimensions and hence must be resized before analysis with the head branch. Regions from the feature map corresponding to each RoI proposal are extracted and resized to 7×7 pixels using bilinear sampling. This is done with TensorFlows `crop_and_resize` function, a numerically efficient approximation of the "RoI align" procedure in the original Mask R-CNN model (He et al., 2017). In contrast to sampling at one location, as in our model, RoI align requires dividing each of the 7×7 RoIs into four bins. The interpolated values at each of four regular locations is sampled followed by a pooling layer to generate the 7×7 head input. See He et al. (2017) for more details.

We pick the 100 anchors from the RPN predictions that are most likely to contain objects, and refine their location and size. If several anchors overlap significantly (determined by a threshold defined in the NMS operation) we keep the one with the highest positive class score and discard the rest; this step is referred to as "non-maximum suppression" (NMS) and is described in detail in the next section.

5.5.3.1 Non-maximum suppression

Non-maximum suppression (NMS) takes the list of proposals sorted by score and iterates over the sorted list, discarding those proposals that have an IoU with a higher score proposal, larger than some predefined threshold . This threshold value is important since setting it too high would result in multiple predictions for the same source (shredding). On the other hand, a low threshold would fail to separate closely space objects. This was one of the hyperparametrs tuned during training step. The RoIs that satisfy the NMS threshold are passed to the next stage of detection.

Finally, the 10 positive RoIs with the highest overlap (IoU) with a ground truth target are passed to the head branch to perform classification and bounding box refinement. In addition, 20 negative RoIs, defined as RPN proposals with IoU< 0.5[4] with every ground truth box, are also passed to the branch. This ensures a 33% positive sample for the head classifier to train on. In the second stage, a convolutional neural network is applied on each of the proposals. In the Mask R-CNN model (He et al., 2017), the head branch network performs classification, bounding-box regression and binary segmentation mask prediction for each proposal. As described above, `ResidualDetectron` does not perform mask prediction.

---

[4] Note that this IoU threshold is different from the parameter used to assign the target bounding box to an anchor box.

5.5.4 Head branch

Thirty 7 × 7×256 features extracted by the RPN are input to the head branch, where they are either classified as a positive detection and their bounding box refined, or else discarded as background. A schematic of the head branch is illustrated in Figure 5.14. The head branch is composed of two fully connected layers that reduce the 7 × 7×256 feature map to a 1-d vector of length 1024. This feature vector is then passed through both the classification and the bounding box branches. The head branch then identifies the bounding boxes with positive prediction scores. If there are more than 10, then only the 10 with the highest scores are retained

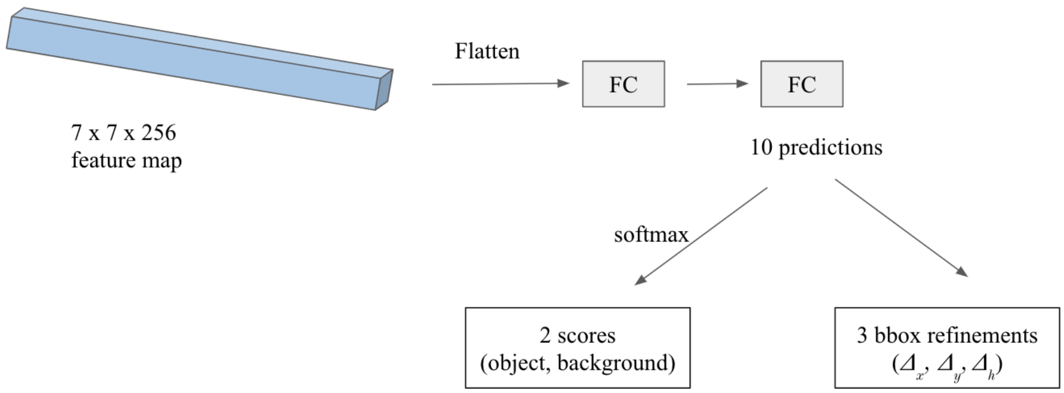


Figure 5.14: Schematic of the head branch illustrating how classification and bounding box regression are performed simultaneously. Each extracted ROI proposal passes through the head branch where fully connected (FC) layers convert the 7 × 7×256 feature map region into a 1-d vector of length 1024. This vector is then analyzed by the classification branch, which computes two scores: one corresponding to the object, the other to background. The scores must sum to 1. The vector is also analyzed by the bounding box regression branch that predicts refinements to the ROI box in order to best match the target.

5.5.4.1 Classification branch

The classification branch consists of a fully connected layer followed by a softmax layer (see Equation 5.8) that predicts two scores denoting the confidence of the network in classifying it as an object or as background. During evaluation only predictions with object score greater than 0.96 are classified as a prediction.

5.5.4.2 Bounding box regression branch

A second fully connected layer in the head branch reduces the 1024 feature vector to the three predicted Δ refinements needed for the RoI to best match the target bounding box.

5.5.4.3 Mask branch

As mentioned at the beginning of the chapter, we do not use the mask head for our residual detection method. This essentially makes this network resemble the Faster R-CNN, where the first stage RPN is followed by classification and bounding box prediction in the head branch. However, some of the improvements of Mask R-CNN over Faster R-CNN are the RoI-align algorithm to generate RoI and the FPN to improve detection at multiple scales. Thus, despite performing only classification and bounding box predictions like Faster R-CNN – and *not* using the definitive “Mask” in Mask R-CNN – we continue to use the Mask R-CNN architecture and refer to the network as such[5].

5.5.5 Network training

Training the complete model requires four different losses: two for the RPN and two for the head branch. While the RPN and the head branch can be trained independently, we choose to train the whole network together, end to end. It is important to note that gradients from RPN do not propagate to the head branch and that the loss for the RPN and the head branch are computed independently.

5.5.6 Loss function

The RPN and the head branch have similar multi-task loss functions $\mathcal{L}$, defined by,

$$\mathcal{L} = \mathcal{L}_{\text{cls}} + \mathcal{L}_{\text{box}}, \tag{5.10}$$

where $\mathcal{L}_{\text{cls}}$ is the classification loss and $\mathcal{L}_{\text{box}}$ is the bounding-box loss.

The classification branch computes a score p corresponding to its confidence that the object is positively identified and not background. The classification loss, $\mathcal{L}_{\text{cls}}$, is a binary log loss, $\mathcal{L}_{\text{cls}}(p) = -\log(p)$ if the prediction is correct, and $\mathcal{L}_{\text{cls}}(p) = -\log(1-p)$ if the prediction is incorrect.

The bounding box loss, $\mathcal{L}_{\text{box}}$, is evaluated over the true bounding box refinement targets $\boldsymbol{\Delta_i} = (\Delta_x, \Delta_y, \Delta_h)$ as described in Equation 5.7 and the network predicted bounding box refinements, $\boldsymbol{\Delta^p} = (\Delta_x^p, \Delta_y^p, \Delta_h^p)$ as,

$$\mathcal{L}_{\text{box}} = \sum_{i \in x,y,h} \text{smooth}_{L_1}(\Delta_i^p - \Delta_i). \tag{5.11}$$

---

[5] The most accurate name for the network would be Faster R-CNN with FPN and RoIAlign

The $\text{smooth}_{\text{L1}}$ loss function (Huber loss) is defined as,

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2, & \text{if } |x| < 1. \\ |x| - 0.5, & \text{otherwise.} \end{cases} \tag{5.12}$$

This loss function is a robust L1 loss that is less sensitive to outliers than the L2 loss. $\mathcal{L}_{\text{box}}$ is evaluated only over boxes that correspond to positive ($p > 0.5$) predictions.

In addition to the regular losses ($\mathcal{L}$ in Equation 5.10), the L2 regularization loss ($\mathcal{L}_{\text{L2}}$, see Section 5.2.4) is also computed for all the layers, with a weight decay of $\lambda = 0.0001$.

5.6 Data augmentation for training

As mentioned above, larger training data helps neural networks avoid over-fitting. One of the common techniques to increase the size of the training set without actually generating more data is to perform data augmentation. Since our images have rotational invariance, for every input image we produce three more images (90 degrotations clockwise and counter-clockwise, and mirror image). Thus, our training dataset becomes four times bigger at negligible computational cost.

Now that we have described the model, we test its performance in detecting unrecognized objects in blended groups of objects and compare it to other detection methods in Chapter 6.

Chapter 6

Identifying Unrecognized Blends from Residual Images

As described in Section 5.4 (Improving object detection with `ResidualDetectron`), the aim of this study is to identify unrecognized blends – i.e., cases for which one or more objects in the blend scene is not detected – using the residual image calculated as the difference between the original image and the model generated by the `SCARLET` multi-band deblender. When the blend scene includes undetected sources, the `SCARLET` deblender is unable to correctly model the scene, leading to patterns in the residual image. Our goal is to use these patterns to detect and locate the undetected sources.

6.1 Methodology

In Figure 6.1, we illustrate our method using a postage stamp image of a two-galaxy blend. The left panel shows a simulated image of two overlapping galaxies at 10-year LSST depth with the true centroid of each galaxy shown as a red ×. The RGB channels for the image correspond to the LSST *gri* bands. The detection algorithm of the LSST Science Pipeline (`LSP`) (Section 3.2.2.1) was able to detect only one source, with centroid shown as the green circle in the center panel. When modeled by `SCARLET` as consisting of only one source, the scene is described by the profile shown in the image in the center panel. The difference between the true image and the `SCARLET` model is the residual image, shown in the right panel. Since `SCARLET` tried to model the scene as one symmetric, monotonically decreasing source, the model overestimates the flux diametrically opposite the input model center and does not include the flux of the undetected source. This results in a dipolar pattern in the residual image. Since the two galaxies have different colors, the residual dipoles are also colored. These residual patterns are what we utilize to provide information about the presence of the undetected source.

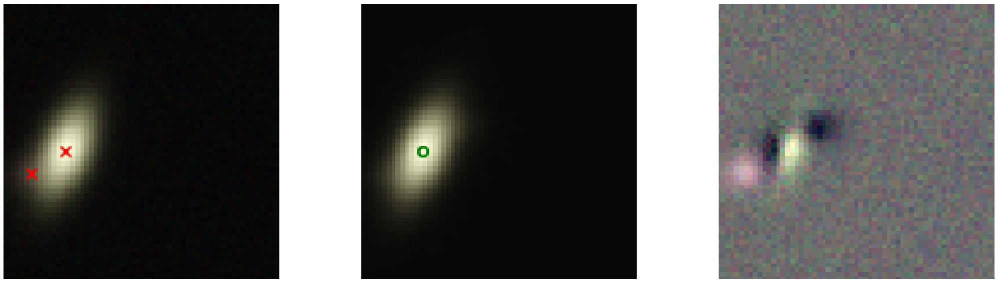


Figure 6.1: Example of using `SCARLET` residuals to identify unrecognized blends. A postage stamp image of a simulated two-galaxy blend scene (left) when modeled by `SCARLET` as one source (center) produces the pattern in the residual image (right). The true centroid of each of the two galaxies is shown as a red ×. The green circle is the centroid position of the single detected galaxy predicted by the LSST Science Pipeline detection algorithm. The RGB channels correspond to LSST *gri* bands.

In this study, we implement and compare two different techniques to identify the unrecognized blends using residual images:

1. We design and train a new deep neural network (`ResidualDetectron`) to identify the undetected galaxies.
2. We iteratively apply an existing detection algorithm, first on the original blend image and then on the residual image.

6.1.1 Modeling the scene with `SCARLET`

While the general workings of `SCARLET` were described in detail in Section 3.2.3.2, we describe here how `SCARLET` is used in this residual detection study. The `SCARLET` algorithm decomposes the multi-band image of the blend scene as a sum of sources with non-negative flux. For this study, we model each source as an extended object with two components[1], each with a different SED and a non-parametric morphology, but with the same centers. We provide as input to `SCARLET` the *ugrizy* image of the scene along with the centers of sources predicted by a detection algorithm. `SCARLET` models the scenes as sources centered at these points. Input centers act only as starting points for the fit, and can change throughout the fitting process. While the centers are free to move during the fit, new sources are not created. Each source is forced to be symmetric about and monotonically decreasing from its center. We run the `SCARLET` fit until either a relative convergence error of 0.0015 is attained, or the number of iterations exceeds 200.

---

[1] If `SCARLET` is unable to model the sources as two components, the scene is modeled as comprising of one-component sources.

6.1.2 Assumptions made in this study

This study is a proof of concept demonstration of using residual images to detect unrecognized blends. In the absence of an existing exhaustive catalog of real images of tagged unrecognized blends, we use synthetic blends to simulate the blending in LSST images. The assumptions made are as follows:

- Surface brightness profile (SBP) of each galaxy is described by parametric bulge+disk+AGN profiles. While the galaxy parameters are derived from observed images, we do not use the real galaxy images themselves.
- We assume that the blending effect is largely due to smearing of the galaxy SBP by the PSF. In particular, we ignore possible occlusion (see Section 6.6) of one galaxy by the other since, in the absence of a PSF, the galaxy images would be largely non-overlapping. In particular, we assume that the observed flux in each pixel due to the galaxies in the blend is simply the sum of fluxes from each object.
- We ignore possible systematic effects due to errors in PSF modeling (including chromatic PSFs and galaxy color gradients).
- We assume accurate coaddition of exposures (see Section 1.6) by simulating images for a single exposure corresponding to the full 10-year depth. The true PSF image is used by `SCARLET` during deblending.

6.2 Datasets

We use simulated datasets for this study in order to provide a controlled environment to test systematic effects. The parameters used to simulate galaxies in the training and test datasets are from one of two catalogs: the CatSim catalog of objects (described in Chapter 2); and the catalog produced for the LSST DESC Data Challenge 2 (cosmoDC2) (described in Chapter 4). While the CatSim and cosmoDC2 catalogs extend to a magnitude of 30 in i band, we restrict our simulations to a magnitude of 28 in order to prevent extremely faint objects (that would not be detected, even in isolation) from dominating our training and test datasets. Galaxies are simulated as the sum of disk (Sersic $n = 1$) and bulge (Sersic $n = 4$) components with some galaxies also containing an AGN component modeled as a point source.

The postage stamps of the blend scenes are simulated with the galaxy simulation package `GalSim`. In order to provide a uniform framework to simulate, test and compare performance metrics of different algorithms, we developed a `Python` package called `BlendingToolKit` (`btk`)[2]. All simulations and analysis described henceforth were performed with the `btk`, details of which are described in

---

[2] https://github.com/LSSTDESC/BlendingToolKit

Appendix B. The `btk` generates (on the fly) a batch of postage stamps of overlapping galaxies that can then be analyzed with any user-defined detection, deblending, and/or measurement algorithms. We use the simulation functions from the `WeakLensingDeblending` package (see Section 4.3.1 to read in the CatSim or DC2 galaxy parameters, and generate postage stamp images of PSF-convolved galaxy images. For more details see Sanchez, J. et al. (in prep).

6.2.1 Test data

We will measure the performance of two-step residual detection methods on the same dataset used to study unrecognized blends in Section 4.3. The test dataset is composed of 100,000 groups of galaxies with 2 to 10 galaxies per blend, where a group is defined as the minimal set of galaxies for which at least one pixel of each galaxy contains flux from another galaxy in the group above a defined threshold (see Section 4.3.2 for more details).

We set the size of the input images to 128 × 128 pixels (25.6×25.6 arcsec^2). Including group sizes larger than 10 could result in some galaxies lying near or beyond the edge of the stamp. Future models can be trained with larger stamps, or on blends of larger group sizes, without altering the network architecture. However, for this proof of concept study, we limit our training and test data to blends with no more than 10 galaxies.

Similar to Section 4.3, the galaxies are simulated as a single long-exposure image with parameters described in Table 4.1. Once again small variations are introduced to the fiducial 10-year LSST PSF FWHM and exposure time to ensure that the observing conditions are not identical for all blend images. This ensures that the neural network does not use the information of constant PSF and background during training.

6.2.2 Assessing detection performance

The performance of our detection method is assessed by comparing the position coordinates for each detection to the centroids of the true objects. To make this comparison, we merge the set of position coordinates from the primary detection step and the set predicted from the residual image. In some cases, significant residuals in pixels near a true source that is detected in the primary detection step cause this source to be detected again in the residual detection step. To remove these duplicate detections, we drop residual detections that lie within one pixel of a detection in the primary detection step.

We employ the two matching algorithms described in Section 4.3.4 to match a detected source center to a true object centroid – i.e., two methods in which a true object is matched to a detection if it is closer than a defined threshold, with the methods using two different definitions for distance. See Section 4.3.4 for more details.

6.3 Residual detection with `ResidualDetectron`

The architecture of the neural network model was described in detail in Chapter 5. The six-band *ugrizy* images of the blend scene, $I(x, y, \lambda)$, are first modeled by `SCARLET` to produce six model images, $M(x, y, \lambda)$. The model images have zero background beyond the footprints of the modeled galaxies. The six `SCARLET` models and the six residual images are input to the network, which predicts bounding boxes at the locations of the undetected objects (the targets).

6.3.1 Network input

The network uses the six residual and six `SCARLET` model images to predict undetected source locations. Because the pixel values of the images (expressed in electron counts) are noisy and have a large dynamic range, two transformations are applied to these pixel values before they are input to the network:

- The residual pixel values are normalized by dividing by the estimated pixel noise.
- A nonlinear mapping is applied to the `SCARLET` model image.

6.3.1.1 Normalized residual ("pull") image

The pixel values in the blend image include Poisson noise corresponding to the mean sky background, B_i, and the signal itself. Since the exposure times and sky levels vary between the bands as described in Table 4.1, the mean sky noise is different in each band. To account for this noise in the residual images, we divide each pixel value by the estimated noise for that pixel. This normalized residual is sometimes called "pull". We then use this image of residual pulls as input to the network.

Specifically, in the ith band, the pull image $P_i(\vec{x})$ is computed from the residual image $R_i(\vec{x})$ as,

$$P_i(\vec{x}) = \frac{R_i(\vec{x})}{\sqrt{B_i + I_i(\vec{x})}}, \tag{6.1}$$

where $I_i(\vec{x})$ is the blend image and B_i is the sky noise.

6.3.1.2 `SCARLET` model image

The `SCARLET` models for the primary detected galaxies, expressed in electron counts per pixel, have a large dynamic range. Neural network training is more efficient and robust if the input pixel values are all of order 1. Therefore, we apply a non-linear mapping to the `SCARLET` model image, that follows this transformation function, motivated by Lupton et al. (2004):

$$M_i^{\text{map}}(\vec{x}) = \frac{1}{Q} \sinh^{-1}\left(\frac{QM_i(\vec{x})}{\mathcal{S}}\right), \tag{6.2}$$

where $M_i(\vec{x}, \lambda)$ and $M_i^{\mathrm{map}}(\vec{x})$ are the pixel values for the original and transformed `SCARLET` models, respectively, in the ith band. The symbols Q and $\mathcal{S}$ represent mapping parameters. We find that the values $Q = 2731^3$ and $\mathcal{S} = 0.5$ lead to spreads that are roughly of order 1 and not more peaked than the original distributions.

6.3.1.3 Network input images

In the end, our input to the network is a set of twelve 128×128 pixel images $[P_i,\, M_i^{\mathrm{map}}]$, comprising of the six-band residual pull images and the six-band transformed `SCARLET` model images. The distributions of pixel values for a sample of 128 blends are shown in Figure 6.2. The top two panels show the distributions of the raw pixel values (over all six bands). The spikes at 0 in the distributions of pixel values for the `SCARLET` model (left panels in Figure 6.2) are due to the pixels that are beyond the footprints of the `SCARLET` model sources. The bottom two panels show the distributions of pixel values after the transformations described in equations Equation 6.1 and Equation 6.2. While the raw pixel values are of order 10^4 and have a large dynamic range, the pixel values after transformation have a lower spread – of order 1. The lower spread leads to smaller changes in the network weights between updates. In addition, training is not overly influenced by outlier input values (see Section 5.2.7). This leads to faster and more robust training of the neural network.

6.3.2 Network output

The network is trained to process the twelve channel input images and return a square bounding box centered at the location of each undetected source. Each dimension of the bounding box is equal in value to twice the size of the PSF-convolved galaxy, σ, defined in the matching algorithm 2 described in Section 4.3.4.

In some cases, all the objects in the blend scene are detected in the primary detection step and the residual image does not contain any undetected sources. However, the network architecture requires at least one target bounding box per input. To satisfy this constraint, we introduce a target bounding box of 1×1 pixel at the top-left corner for all input during training. During testing time we ignore any prediction at this location. Since the blends are centered on a postage stamp that is generally much larger than the spatial extent of the blend to avoid clipping at edge boundaries, it is very unlikely that a galaxy SBP would extend up to the edge of the top-left corner. Therefore ignoring detections with bounding boxes of size 1×1 pixels at the top-left corner of the stamp does not impact our results.

Each bounding box is accompanied by a score that indicates the network confidence in the prediction being an undetected source (see Section 5.5.4.1 for details on how this is generated). While

---

[3] This is the standard deviation of the pixel values in all the `SCARLET` model images in a validation set made of 1,000 blend scenes.

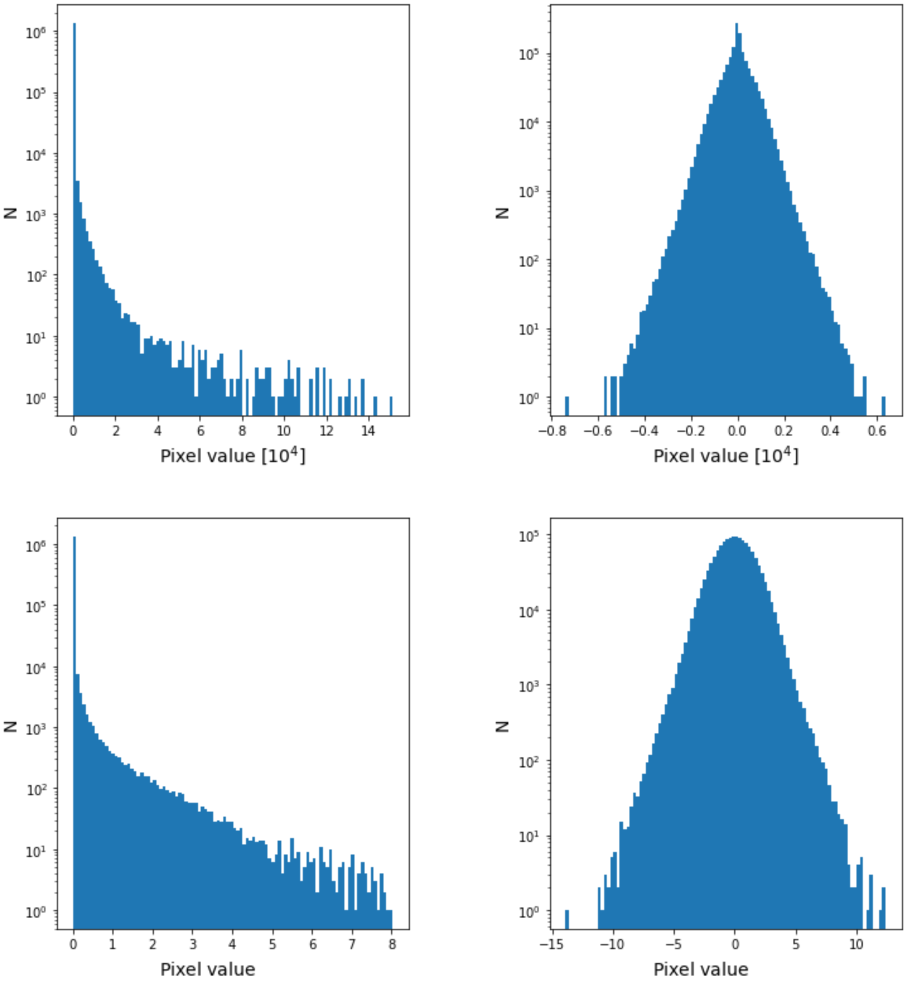


Figure 6.2: Distributions of pixel values in all six bands of the `SCARLET` model (left panels) and residual images (right panels) for 32 blends. The top two distributions correspond to the raw pixel values without transformations applied, while the bottom distributions show the pixel values after applying the transformations described in Equation 6.1 and Equation 6.2 in Section 6.3.1. Note that the scales on the horizontal axes in the top plots are multiplied by 10^4.

being similar to probability in that the value lies between 0 and 1, the score is not mathematically equal to the probability that the prediction is accurate; rather, it is a useful although not statistically accurate indicator of the confidence that it is correct. During the network training stage this score is used in the log loss of the classification branch described in detail in Section 5.5.6.

As described in Section 5.5.4.1, for each prediction `ResidualDetectron` also outputs a confidence score. By eliminating low-confidence predictions we can eliminate predictions that are likely false positives. A good threshold score optimizes two detection metrics – precision and recall.

Precision is defined as the fraction of all predictions that are correct. A lower precision indicates a higher false-positive rate. Recall (also called completeness) is the fraction of the targets that were correctly predicted. An ideal detection algorithm would have precision and recall values of 1.

Figure 6.3 shows the precision and recall values for the predictions as a function of the threshold for the minimum score. During the test stage we consider only predictions with scores over 0.96 (black-dashed line) to assess network performance.

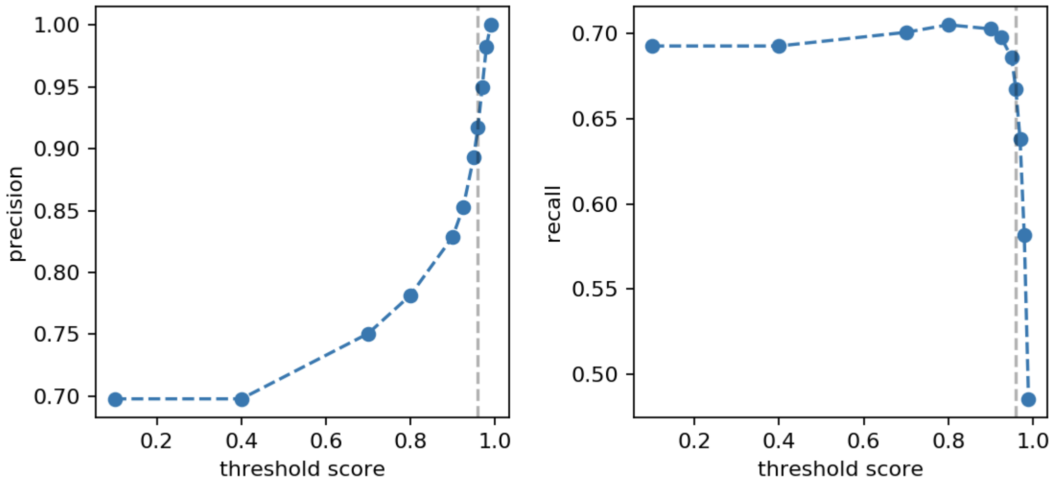


Figure 6.3: Impact of `ResidualDetectron` score threshold on the detection precision and recall. `ResidualDetectron` detection precision (left) and recall (right) as a function of score threshold. The black-dashed line shows the threshold score of 0.96 used to eliminate low-confidence predictions during the testing phase.

In Figure 6.4, we show an example set of raw images (top row) and transformed images (bottom row) for a two-galaxy blend with only one object detected in the primary detection step. Each RGB image corresponds to an LSST *gri* image. The images in the top row correspond to the blend image (left), the `SCARLET` model (center), and the residual image (right). Those in the bottom row correspond to the `SCARLET` model with the non-linear transformation described in equation Equation 6.2 applied and the residual image normalized to the pixel noise as described in Equation 6.1. The dashed green square in the bottom images represents the bounding box centered at the location of the undetected source; the network will be trained to predict the location and size of this box.

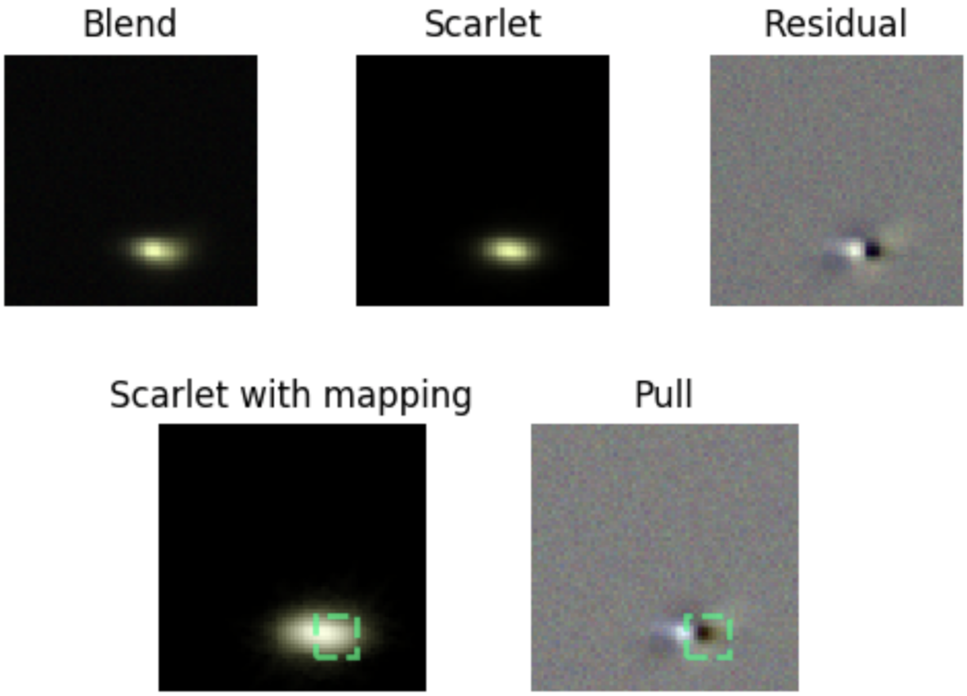

Blend
Scarlet
Residual
Scarlet with mapping
Pull

Figure 6.4: Top: Example of an unrecognized two-galaxy blend image (left) that was detected as one source and modeled as such by `SCARLET` (middle). The undetected source is clearly visible in the residual image (right). Bottom: `SCARLET` model image after the non-linear scaling of flux described in Equation 6.2 (left), and the residual pull image defined in Equation 6.1 (right) that are the input images to the `ResidualDetectron` detection algorithm. Overlaid on the two input images is the target bounding box that denotes the location and size of the undetected object, which the network will be trained to predict.

6.3.3 Training the network

The goal of the training stage is to learn the weights in each layer of the network model, that will generate a set of bounding boxes that best match the target boxes. A schematic of the training process for residual detection is shown in Figure 6.5, where ellipses denote images, rectangles denote catalogs or tables, and boxes with rounded corners denote processes. The training proceeds through these steps:

1. Run a primary detection algorithm on the multiband-coadd image of the blend scene.

2. Compare detected object centers to true source centers using a matching algorithm and identify the undetected sources. For this study we use only matching algorithm 1 to identify undetected objects in the primary detection step.
3. Model the six-band blend images with `SCARLET` using only the source centers predicted in the primary detection method in Step 1.
4. Subtract the `SCARLET` model from the blend image to get six residual images.
5. Run `ResidualDetectron` with the residual pull and transformed `SCARLET` model images as input: $[P_i, M_i^{\mathrm{map}}]$.
6. Calculate the value of the loss function (see Section 5.5.6) by comparing the target and predicted output bounding boxes.
7. Update the network weights to output bounding boxes that best locate the undetected sources identified in step 2. The weights are updated using the stochastic gradient descent method described in Section 5.2.2.

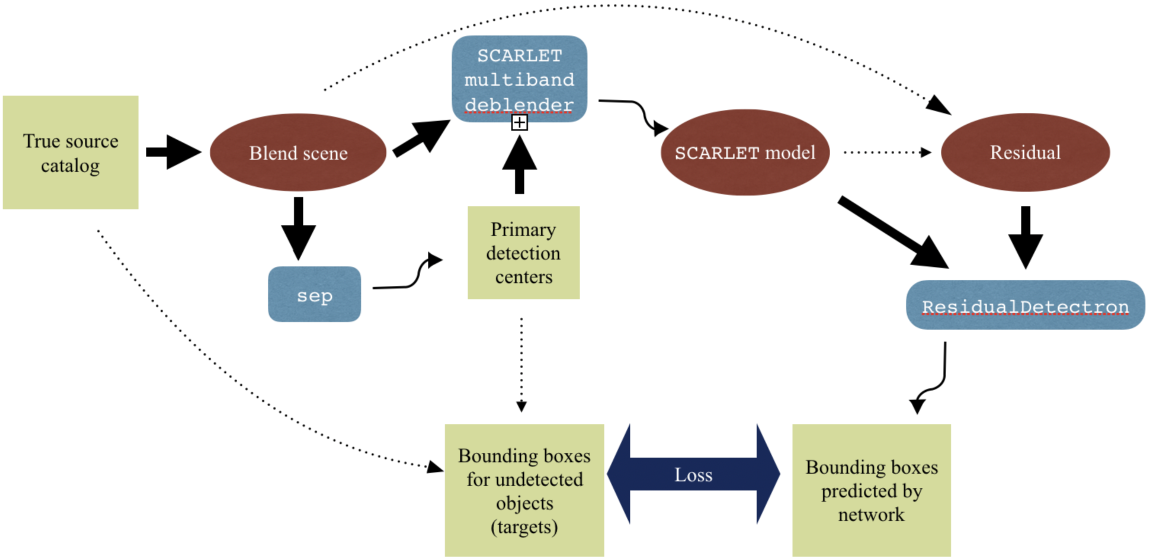


Figure 6.5: Schematic of the `ResidualDetectron` workflow. The ellipses denote images, the rectangles denote catalogs or tables, and the boxes with rounded corners denote processes.

During the training process, we use as the primary detection algorithm the method implemented in `SourceExtractor` (Bertin & Arnouts, 1996) rather than that implemented in the LSST Science Pipelines (`LSP`). This choice is based solely on practical computational constraints related to the fact that the network training is enormously faster when implemented on GPUs[4]. While `SourceExtractor` is simple to install and run, at the time this study was done, installing `LSP` on machines with GPUs was not yet practical. During the test stage, we compare the `ResidualDetectron` performance (after running `SourceExtractor` for primary detection) with `LSP` detection performance, with only primary detection and with the `LSP` algorithm run iteratively on the residual image.

We detect the sources using the Source Extraction and Photometry (`SEP`)[5] (Barbary, 2018) Python library, which makes the core algorithms of `SourceExtractor` available as stand-alone functions and classes. The `SEP` parameter descriptions and the settings that we use for source detection are given in Table 6.1.

Table 6.1: Description of parameters relevant to the `SEP` source detection algorithm and values used in this study.

Name	Description	Parameter value
`thresh`	Threshold pixel value for detection	$1.5\sigma_{\rm noise}$
`minarea`	Minimum number of pixels required for an object	5
`deblend_nthresh`	Number of thresholds used for object deblending	32
`deblend_cont`	Minimum contrast ratio used for object deblending	0.005
`bw, bh`	Size of background boxes in pixels	64

Since we have twelve images, $[P_i,\, M_i^{\rm map}]$, as input to our network, rather than only three RGB bands, as used to train the ResNet backbone in Mask R-CNN (see Section 5.5.2.3), we cannot use transfer learning to initialize the weights of the network model. Instead, our networks are trained from scratch with the weights randomly set at the beginning of the training[6]. Our test sample as described above consists of groups of overlapping galaxies. Since this yields a limited sample of undetected objects, and we want to avoid over-fitting the network for this particular dataset, we first train the network on artificial blends created by randomly selecting galaxies from the CatSim catalog and drawing them on a postage stamp with a distribution of distances between galaxy centers that we control to produce a higher rate of undetected sources than realistically expected by LSST. Although these ensembles of blends are not realistic, they can provide a good starting point for networks to learn to identify signatures of undetected sources from residual images.

For more efficient optimization of the weights, we train the network in three stages, using datasets of increasing complexity as described in the next sections. The weights from the end of one stage are set as the starting point for training in the next stage. For each of the four stages, the loss function is calculated for a validation set of images that is similar to, but statistically independent of, the

---

[4] For network training, we used GPUs available on the Stanford University Sherlock computing cluster.

[5] https://github.com/kbarbary/sep

[6] The CNN weights are randomly initialized using the Xavier/Glorot normal initialization method (Glorot & Bengio, 2010).

training set used for that stage.

6.3.3.1 Stage 1: Closely spaced two-galaxy blends close to center of postage stamp

The Dawson et al. (2016) study on unrecognized blends found that $\approx 75\%$ of unrecognized blends in deep images are composed of only two true objects. Therefore, we begin our network training by introducing residuals images with zero or one undetected source. These training images contain one or two true galaxies per postage stamps. For the two-galaxy images, the angular distance between the galaxies was randomly selected from a uniform distribution over a small enough range to generate a training sample of blends with a significant probability of containing an undetected source. The x and y coordinates of each galaxy center relative to the postage stamp center were selected from a uniform distribution in the range $[-d, d]$, where d is one-fifth the postage stamp size.

We include postage stamps with only one galaxy in order for the network to learn the profiles of isolated sources with no undetected source in the residual image. Since training on only residual images with undetected sources could lead the network to always expect an undetected source, including these single-galaxy images correctly modeled by `SCARLET` is important for reducing the number of false positives predicted by the network. Similarly not all the two-galaxy blends contain an undetected source.

6.3.3.2 Stage 2: Closely spaced non-centered two-galaxy blends with larger galaxies

While the Stage 1 galaxies were centered on the postage stamp, we now reduce this limitation, and allow our two-galaxy blends to lie further away from the center of the stamps.

Since the CatSim galaxy catalog contain more small galaxies than large ones, we increase the probability of sampling large galaxies for the blends. Two galaxies are randomly sampled from the CatSim catalog and placed on the postage stamp. The x and y coordinates of the two galaxy centers are selected from a uniform distribution between $[-6, 6]$ pixels from the center of the postage stamp.

6.3.3.3 Stage 3: Closely spaced multi-galaxy blends

Here we increase the maximum number of galaxies per blend by including blends with up to 10 galaxies. The number of CatSim galaxies in a scene is selected randomly between 1 and 10 with uniform probability. The displacements of the x and y coordinates of the galaxy centroids relative to the center of the postage stamp are selected randomly from a uniform distribution between [-s, s] where $s = (\text{stamp size}/20)\sqrt{\text{N}} = 2.56\sqrt{\text{N}}$ arcsec ($\approx 13\sqrt{N}$ pixels), where N is the number of galaxies in the blend.

6.3.3.4 Stage 4: Blends identified as groups in the DC2 catalog

In the final stage of learning, we train the network on a set of 10,000 groups of galaxies in the cosmoDC2 catalog identified by the `WeakLensingDeblending` package described in Section 6.2.1. This set is statistically similar to the test set (see Section 6.2.1) that we will use to assess the performance of the network.

6.3.4 Results

An example of a successful detection by `ResidualDetectron` is shown in Figure 6.6. The top left panel of the figure shows a two-galaxy blend scene with the centroid of each galaxy identified with a red ×. The `sep` detection algorithm is able to find a single galaxy at the position marked with a green circle. The scene is modeled as a single galaxy with `SCARLET`, producing the image in the top right. The residual pull image is shown in the bottom panels. A bounding box indicating the true location and size of the undetected object is represented as the red box in the lower right panel. Overlaid on the same panel is the bounding box predicted by the network, shown as a dashed-green square, and the network confidence score for the prediction. Since the value is greater than our threshold of 0.96, we classify this prediction as a success – i.e., the network has correctly identified an undetected object from the residual.

We also observe cases where the network is unable to identify undetected objects or where it makes spurious predictions. An example of `ResidualDetectron` failing to identify an undetected source is shown in Figure 6.7. Once again, the top panels show the two-galaxy blend scene image (left) where only one source was detected, and its `SCARLET` model image (right). The bounding box corresponding to the undetected object is shown as the red square in the bottom-right panel of Figure 6.7. The network failed to make a prediction and was unable to identify the undetected object.

We compare the performance of the detection algorithms on a common test dataset of 100,000 groups of blend scenes with group sizes ranging from 2 to 10, and 40,000 isolated galaxies (group size = 1). The source centers detected in the primary and residual detection steps are combined to generate a merged list of detected source centers (see Section 6.2.2). These combined source positions are then compared to the positions of the true object centers using the two matching algorithms described in Section 6.2.2.

As in Section 4.4, we first measure the detection efficiency – the fraction of groups in which a particular number of sources are correctly detected – for each group size. The panels in Figure 6.8 show the detection efficiency matrix for the galaxies in groups with group size ranging from 1 to 10. Each column of the matrix sums to 1. The left panel of Figure 6.8 shows the detection efficiency for predictions made by `sep`, while the right panel shows the detection efficiency for merged centers predicted by `ResidualDetectron`.

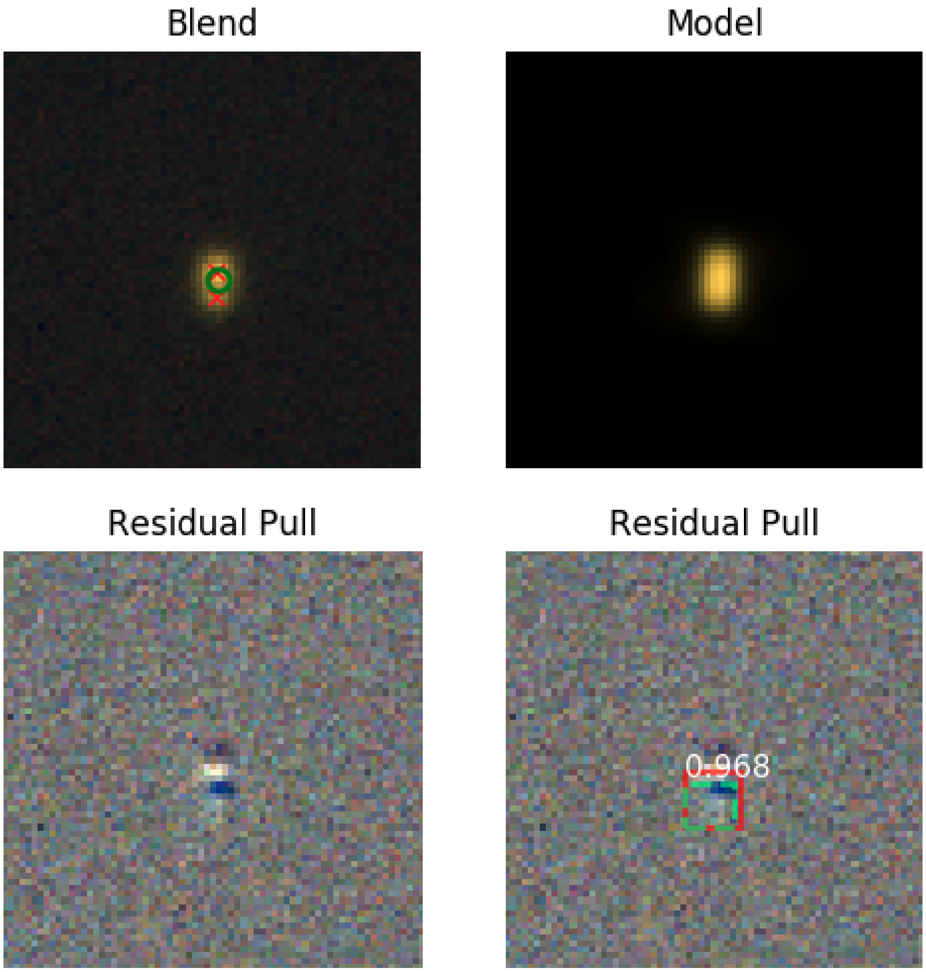


Figure 6.6: Example of successful detection by `ResidualDetectron`. Top left: Image for a two-galaxy blend. The location of each galaxy centroid is indicated with a red ×. The green circle marks the location of a single source detected by `sep` in the primary detection step. Top right: `SCARLET` model for the single detected source. Bottom: Residual pull images. The red box in the lower right image corresponds to the bounding box for the undetected galaxy. The green dashed box corresponds to the single source successfully identified by `ResidualDetectron` in the iterative detection step based on the residual pull image and the `SCARLET` model.

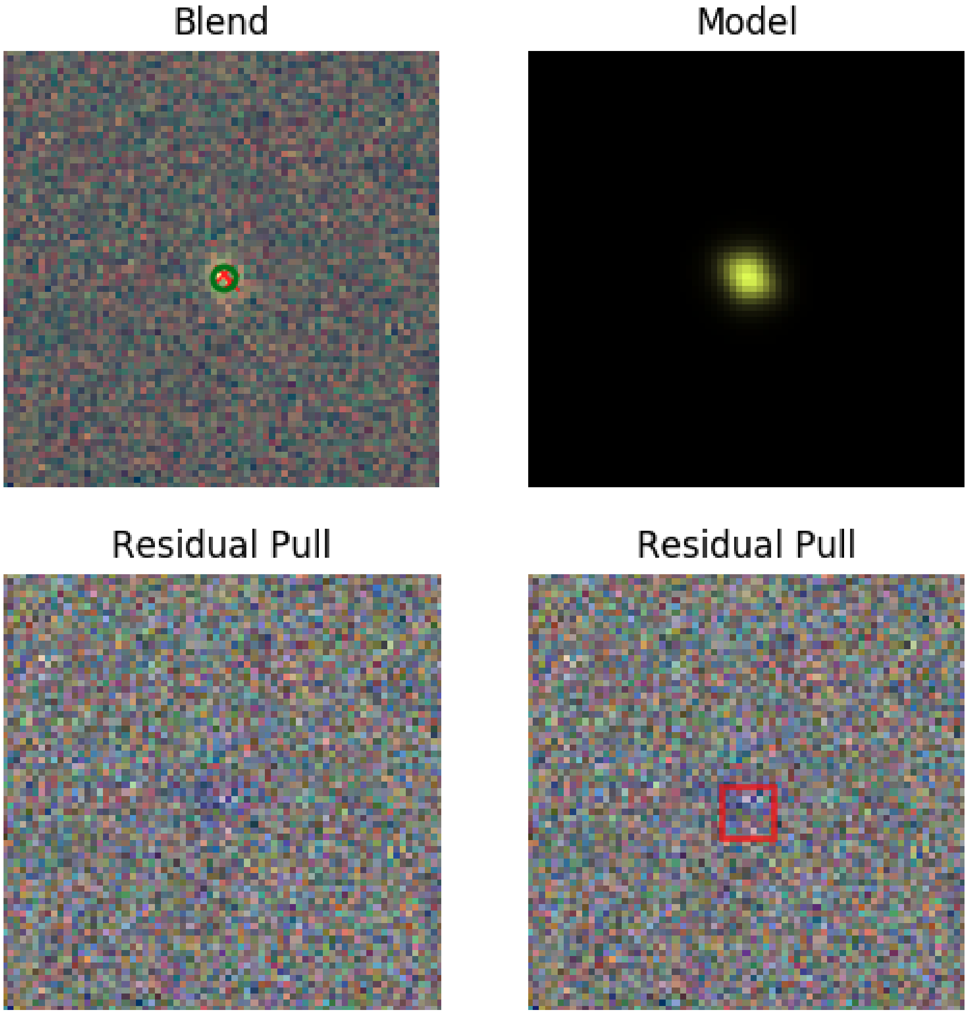
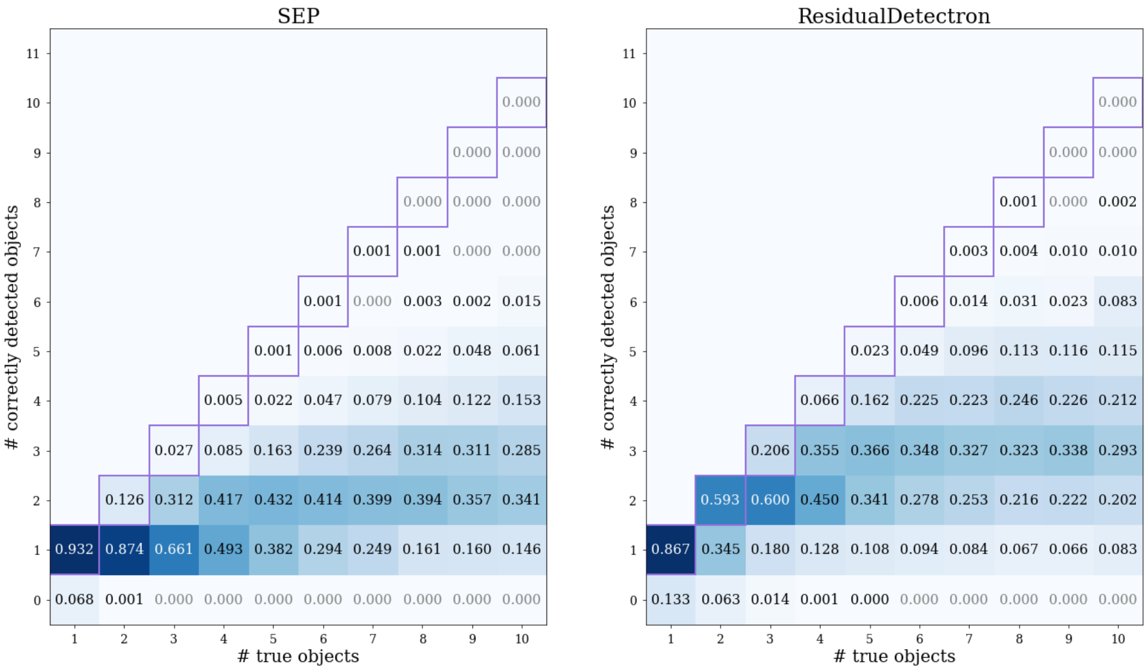


Figure 6.7: Example of failed detection by `ResidualDetectron`. Top left: Image for a two-galaxy blend. The location of each galaxy centroid is indicated with a red ×. The green circle marks the location of a single source detected by `sep` in the primary detection step. Top right: `SCARLET` model for the single detected source. Bottom: Residual pull images. The red box in the lower right image corresponds to the bounding box for the undetected galaxy. `ResidualDetectron` failed to predict the existence of this undetected object.



SEP

# correctly detected objects \ # true objects	1	2	3	4	5	6	7	8	9	10
11										
10										0.000
9									0.000	0.000
8								0.000	0.000	0.000
7							0.001	0.001	0.000	0.000
6						0.001	0.000	0.003	0.002	0.015
5					0.001	0.006	0.008	0.022	0.048	0.061
4				0.005	0.022	0.047	0.079	0.104	0.122	0.153
3			0.027	0.085	0.163	0.239	0.264	0.314	0.311	0.285
2		0.126	0.312	0.417	0.432	0.414	0.399	0.394	0.357	0.341
1	0.932	0.874	0.661	0.493	0.382	0.294	0.249	0.161	0.160	0.146
0	0.068	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

ResidualDetectron

# correctly detected objects \ # true objects	1	2	3	4	5	6	7	8	9	10
11										
10										0.000
9									0.000	0.000
8								0.001	0.000	0.002
7							0.003	0.004	0.010	0.010
6						0.006	0.014	0.031	0.023	0.083
5					0.023	0.049	0.096	0.113	0.116	0.115
4				0.066	0.162	0.225	0.223	0.246	0.226	0.212
3			0.206	0.355	0.366	0.348	0.327	0.323	0.338	0.293
2		0.593	0.600	0.450	0.341	0.278	0.253	0.216	0.222	0.202
1	0.867	0.345	0.180	0.128	0.108	0.094	0.084	0.067	0.066	0.083
0	0.133	0.063	0.014	0.001	0.000	0.000	0.000	0.000	0.000	0.000

Figure 6.8: Efficiency matrix for correctly detected sources with Source Extraction and Photometry (`SEP`) (left) and `ResidualDetectron` (right). Each element in the efficiency matrix corresponds to the efficiency for correctly detecting different numbers of sources in a group (vertical axis) for different numbers of true sources in the group (horizontal axis). The values in each column sum to 1. The true centers and detection locations were matched with matching algorithm 1.

For two-galaxy blends in our test set, `ResidualDetectron` correctly detected both objects ≈59% of the time, correctly detected one of the two objects ≈34% of the time, and failed to correctly detect either in 6% of cases. Because we show only the fraction of *correctly* detected sources in Figure 6.8, the detection efficiency matrix has no entries in the upper left half. As expected, the `ResidualDetectron` performance is better for smaller blends; many sources in large groups with multiple overlapping objects remain undetected even with `ResidualDetectron`. We see that `ResidualDetectron` is able to identify objects that are undetected by `SEP`. For example, two-galaxy blends are ≈ 5 times more likely to be correctly identified as such by `ResidualDetectron` (59% efficiency) than by `SEP` alone (13% efficiency).

Of particular interest is the row corresponding to 1 correctly detected object when the group contains 2 or more true objects since these are the detections for which we might incorrectly *assume* we can estimate shear and photometric redshifts without any impact of blending. We see that values of these elements in the efficiency matrix are much lower for `ResidualDetectron` than for `SEP`.

One expected drawback of any iterative detection algorithm is the potential for a higher false positive rate due to “shredding”. For example, if `ResidualDetectron` detects an object that was already (correctly) detected by `SEP`, then the additional detection would not be counted as correctly detected. This type of shredding in the residual detection stage is the reason for the row corresponding to zero correct detections being larger for `ResidualDetectron` than for `SEP`. This also causes a reduction in correctly detected isolated objects in `ResidualDetectron` where ≈ 6% of blends are now either shredded or include spurious predictions. In order to reduce multiple predictions of an object already detected, we remove duplicate predictions within 1 pixels as described in Section 6.2.2. Future studies could explore opportunities to improve detection performance by increasing value to match the PSF size(≈ 3 pixels).

In Figure 6.8 the true centers and detected locations were matched with matching algorithm 1. We show the detection efficiency when matching algorithm 2 is used in the left panel of Figure 6.9.

The correctly detected sources matched with either algorithms do not include spurious detections and shredded objects. In the right panel of Figure 6.9, we show the detection efficiency for *all* detections – not only for the correctly matched objects – for all predictions made by `ResidualDetectron`.

We see that non-zero values for elements of the detection efficiency matrix for all detections can extend into the upper-left corner of the matrix, since the number of detections can be larger than the number of true objects in the case of shredding. Approximately 14% of isolated galaxy images are shredded with `ResidualDetectron` predicting either 3 or 2 sources. Approximately 19% of three-galaxy blends were predicted to be 3 sources.

While the recall values of predictions by `ResidualDetectron` for the blended galaxies is 0.674, better than 0.485 for `sep`, the false positives lead to a precision value of 0.932, which is smaller than the precision of 0.99 for `sep`.

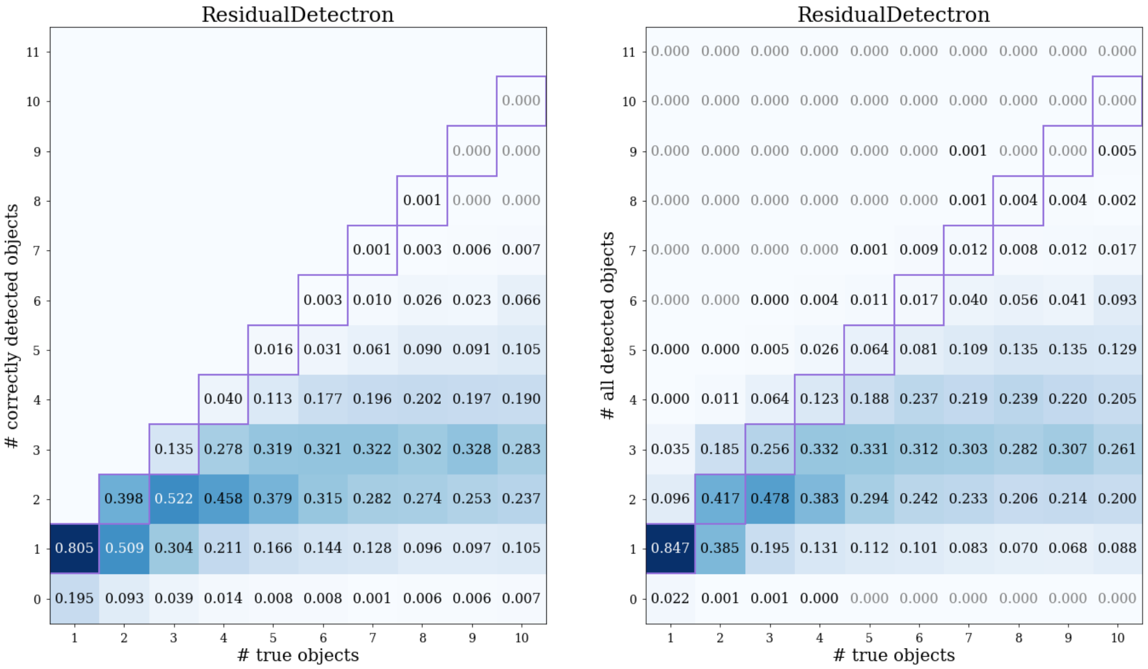


Figure 6.9: Left: efficiency matrix for correctly detecting sources with `ResidualDetectron` when detections are matched with matching algorithm 2. Right: efficiency matrix for all detections `ResidualDetectron`. This includes spurious detections and shredded objects.

6.4 Iteratively detecting sources with LSST Science Pipeline

As described in Section 3.2.2.1 and Section 3.2.3.2, the current LSST Science Pipeline (`LSP`) performs detection and deblending in two separate steps: the detection algorithm identifies peaks above a certain threshold and the multi-band deblender models the scene with `SCARLET`. We introduce an additional step in this pipeline to identify any undetected source from the residual image.

While `LSP` will perform detection in each band separately and then merge the centers, we limit this study to monochromatic detections in the multiband-coadd image.

We illustrate in Figure 6.10 the following steps for iterative detection with `LSP`:

1. Run primary `LSP` detection algorithm on the multiband-coadd image of the blend scene.
2. Use `SCARLET` to model the scene in the six-band blend image as sources with centers predicted by the detection algorithm in step 1.
3. Subtract the `SCARLET` model from the original blend image to produce the residual image.
4. Run the `LSP` detection algorithm again on the multiband-coadd of the residual image.

The parameters of the `LSP` detection algorithm are shown in (Table 6.2) and are the same for both the primary and iterative detection steps.

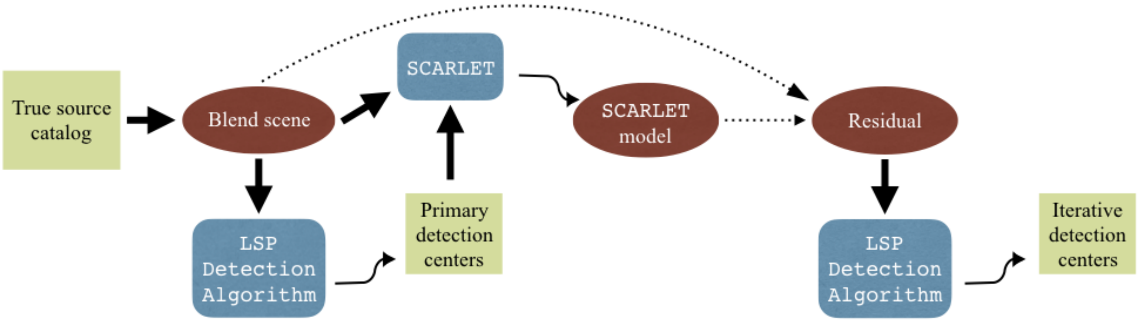


Figure 6.10: Schematic of iterative detection with `LSP`. The ellipses denote images, the yellow squares denote catalogs or tables, and the blue boxes denote a process.

Table 6.2: Parameters of the LSST Science Pipeline (`LSP`) detection algorithm.

Name	Description	Parameter Value
`thr_value`	Threshold pixel value for detection	$5\sigma_{\text{noise}}$
`min_pix`	Minimum number of pixels required to be considered a source	1
`bkg_bin_size`	Binning size of local background in pixels	32

As described in Section 6.2.2, we combine the sources identified in the primary and iterative detection stages in order to get a merged set of detected object locations. The merged set of object locations is compared to the locations of the true object centers to evaluate detection accuracy.

6.4.1 Difference between iterative `LSP` detection and `ResidualDetectron`

While iterative `LSP` detection and `ResidualDetectron` are both two-step detection techniques to identify undetected objects from residual images, there are key differences between the two:

- The `LSP` detection algorithm operates on either a single-band image or a multiband-coadd image of multiple bands. In either case, the color information is not used in detecting sources. In the `ResidualDetectron`, on the other hand, all six bands are used for both the residual image and the `Scarlet` model.
- The `LSP` algorithm is designed to detect *positive* peaks in an image. Since the residual image has contributions due to differences between the intrinsic galaxy surface brightness profile and the `SCARLET` model, it may contain local maxima due to inaccurate modeling, as well as

upward statistical fluctuations due to noise. These can lead to a high false prediction rate when iterative detection is applied. In the `ResidualDetectron`, on the other hand, the CNN learns the characteristic signature of an undetected source, which typically exhibits a "dipole" pattern with a positive peak and a negative valley.

6.4.2 Results

An example of a successful detection with the iterative `LSP` method is depicted in Figure 6.11. The top-left panel shows a monochromatic six-band coadd image of a two-galaxy blend scene. The true center of each object is indicated with a cyan ×. For this image, the primary `LSP` detection identified only one source, marked with a green circle. The `SCARLET` model of the scene is shown in the top-right panel and the residual image is shown in the bottom panels. The iterative `LSP` detection step correctly predicts one source shown to be centered at the green diamond. The true center of the undetected source is at the center of the blue square.

As described above, a limitation of iterative detection of `LSP` is that it simply identifies positive peaks and interprets each peak as a new source. The residual image could contain local maxima, not centered on the undetected object. This can lead to false predictions by producing shredded objects as illustrated in Figure 6.12, where the primary detected source was identified as an additional source during the iterative detection step.

To estimate the performance of iterative `LSP` detection, both single-step and iterative `LSP` detection were run on the test dataset described in Section 6.3.4. The results are shown in Figure 6.13 as a matrix of efficiencies for correctly detecting sources (described in Section 6.3.4). The left panel of Figure 6.13 shows the detection efficiency of single-step `LSP` detection while the right panel shows the efficiency with the iterative step. The ability of the iterative detection step to identify objects that were undetected in the primary detection step is reflected in the higher efficiency values along or close to the diagonal of the matrix for `LSP` iterative detection. As expected, the detection performance is better for groups with fewer sources in the blend; a significant fraction of sources in groups with larger numbers of sources remain undetected, even with the iterative step.

The detection efficiencies of single-step `LSP` and `sep` are different, as evidenced by comparing the left panels of Figure 6.13 and Figure 6.8. `sep` performs better on blends, with $\approx$13% of two-galaxy blends correctly detected in comparison to $\approx$8% for `LSP`. On the other hand, `LSP` correctly detected $\approx$98% of the isolated galaxies, higher than the $\approx$93% detection efficiency for `sep`.

The detection efficiency for iterative `LSP` (with matching algorithm 2) to identify correct detections is shown in the left panel of Figure 6.14. The right panel of Figure 6.14 shows the fractions of predictions for all numbers of detected objects as a function of group size. This matrix includes unsuccessful predictions such as spurious detections and shredded objects.

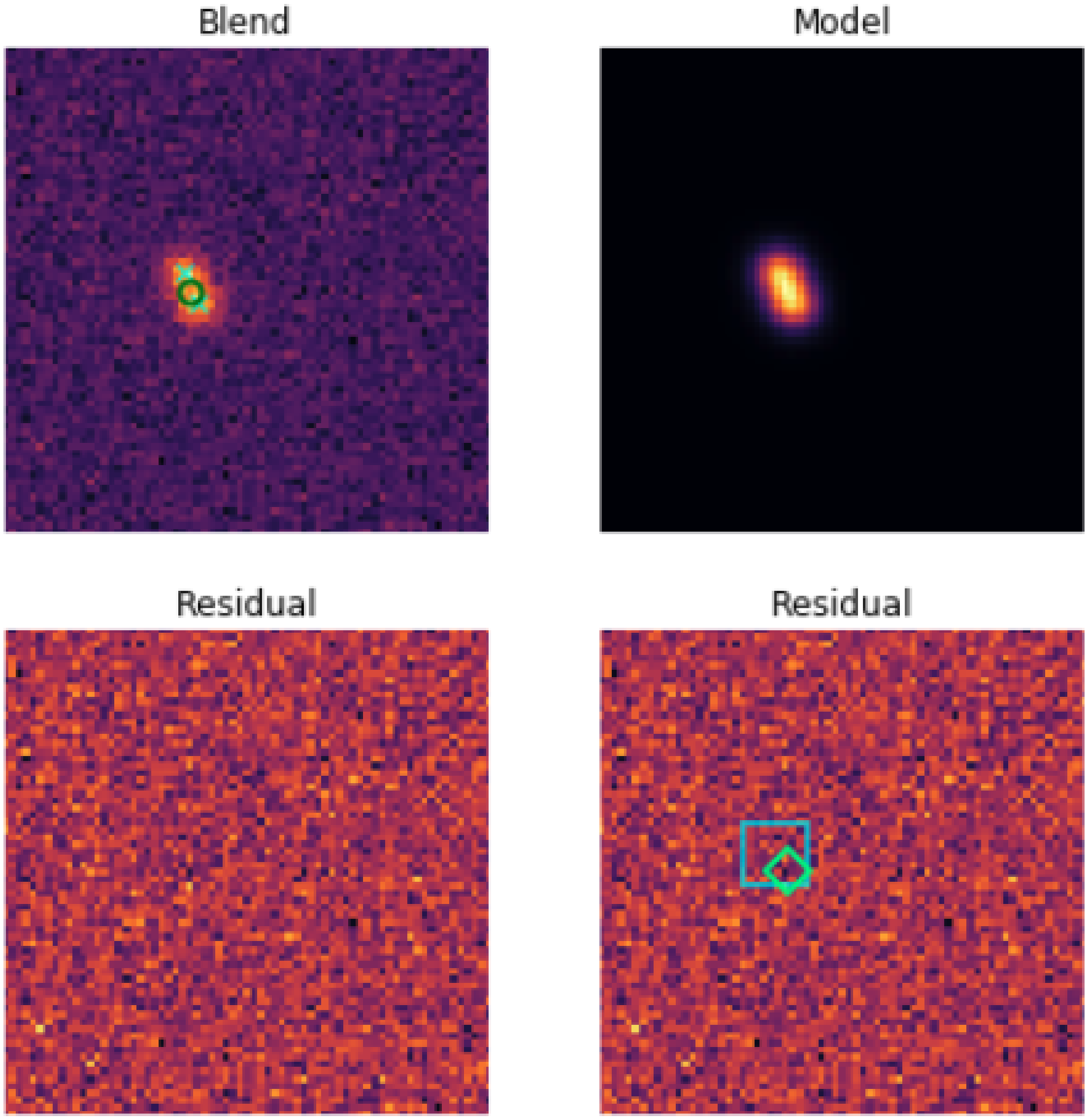


Figure 6.11: Example of a successful iterative detection by the LSST Science Pipeline. Each cyan × denotes the true centers of one of the two galaxies in the blend. The dark green circle indicates the location of a single source detected by the `LSP` in the primary detection step. In the iterative detection step on the residual image (bottom panel), the previously undetected source (blue square) was successfully detected (green diamond).

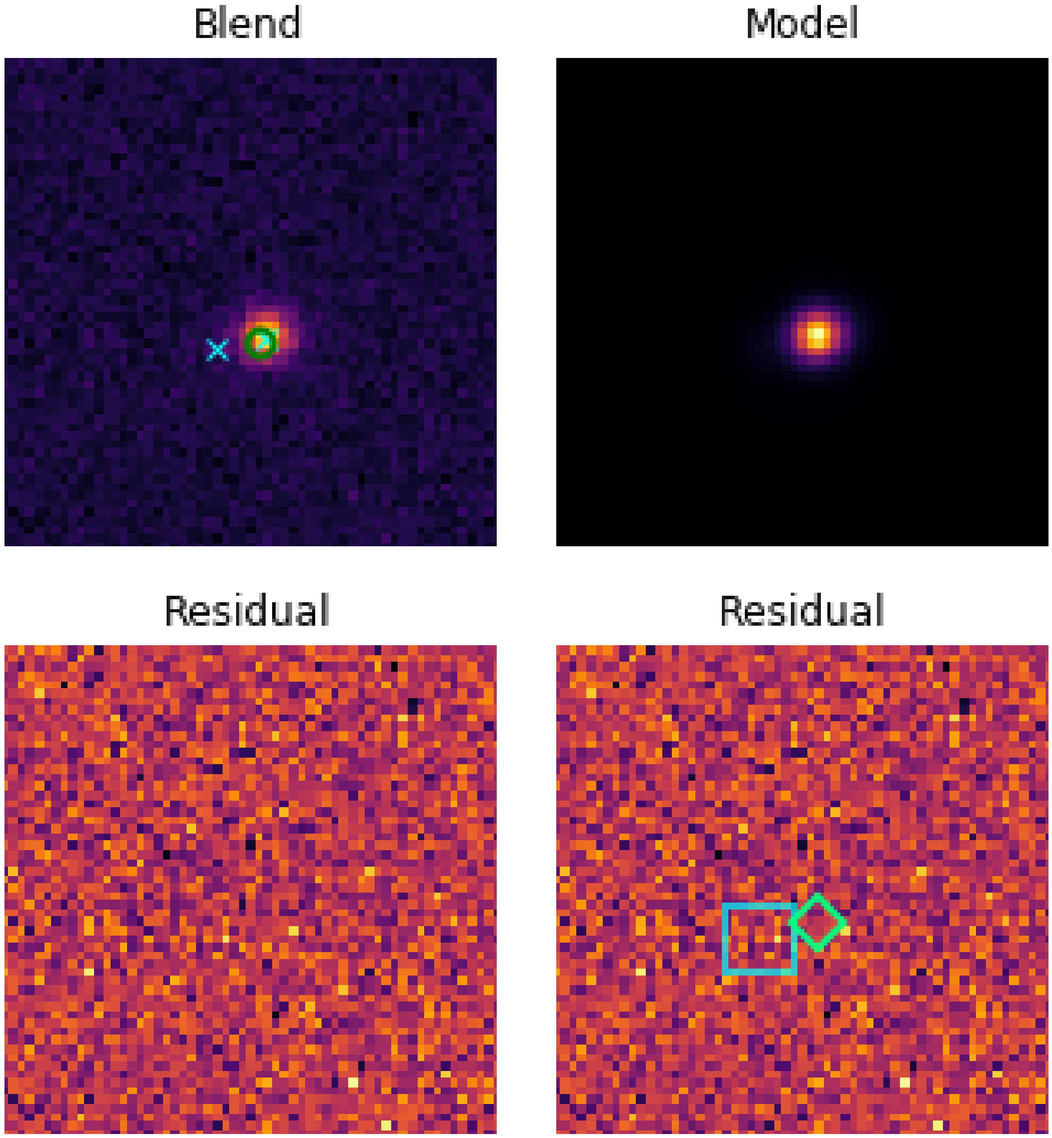


Figure 6.12: Example of failed iterative detection by the LSST Science Pipeline. The iterative `LSP` detection step fails to predict the undetected object (blue square).

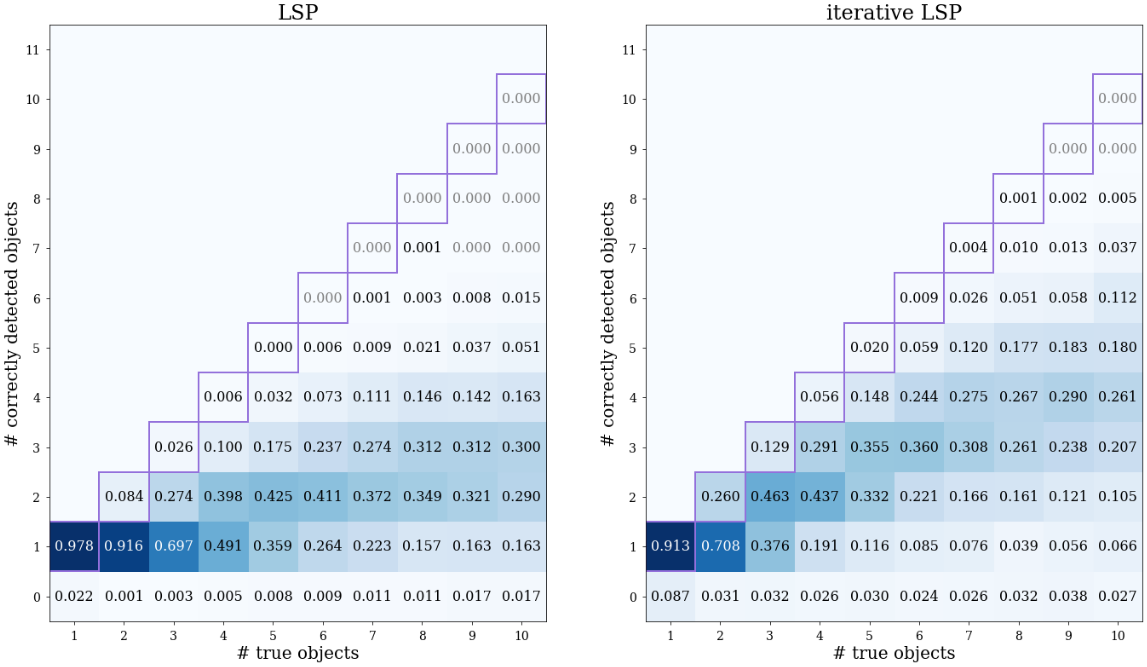


Figure 6.13: Efficiency matrix for correctly detected sources with single-step `LSP` (left) and `LSP` run iteratively on the original image and then the residual image (right). Each element in the efficiency matrix corresponds to the efficiency for correctly detecting different numbers of sources in a group (vertical axis) for different numbers of true sources in the group (horizontal axis). The values in each column sum to 1. The true centers and detected locations were matched with matching algorithm 1.

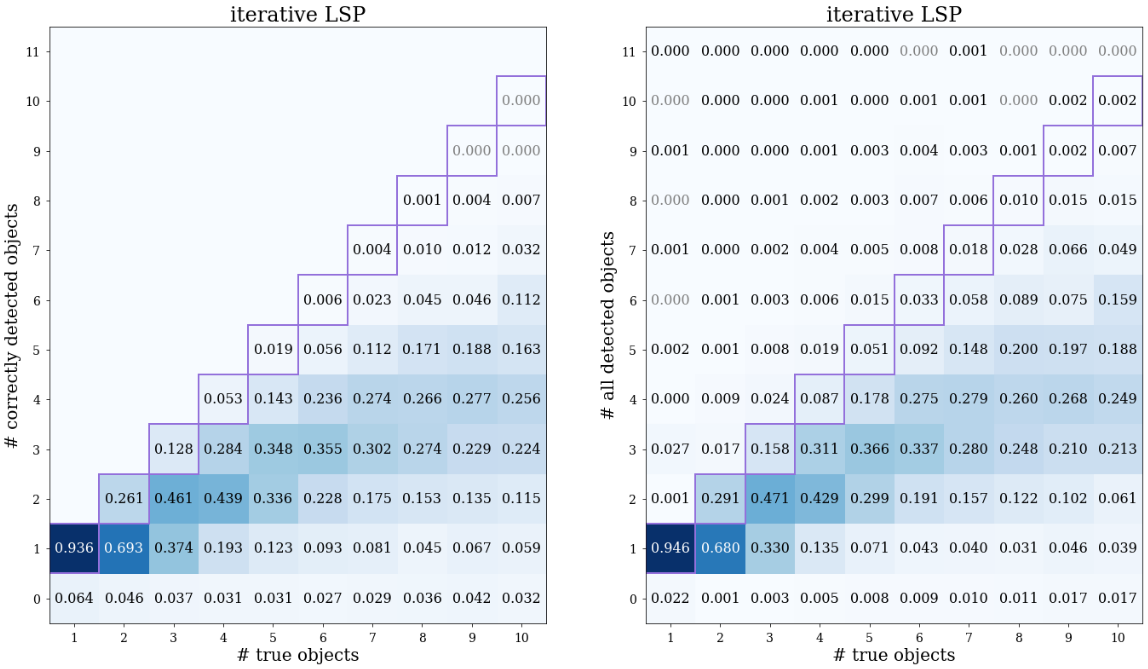


Figure 6.14: Efficiency matrices for iterative `LSP`. Left: efficiency matrix for correctly detected sources when matched with matching algorithm 2. Right: efficiency matrix for all detections. This includes spurious detections and shredded objects.

6.5 Comparisons of detection methods

We compare the detection efficiency of the four different algorithms in Figure 6.15 where we display the cumulative number counts as a function of different intrinsic galaxy properties. The dashed-purple curve (labeled “True”) corresponds to the distribution for all input galaxies. The other curves show the distributions for the four detection algorithms: `ResidualDetectron`, iterative `LSP`, single-pass `LSP`, and `SourceExtractor` in orange, green, red and brown colors, respectively. The galaxy parameter values on the horizontal axis of all three panels correspond to the true values of the detected galaxy.

The top-left panel of the Figure 6.15 shows the cumulative increase in the number of correctly detected objects with increasing magnitude (decreasing brightness). As expected, the detection efficiency decreases with increasing magnitude for all four methods; fainter sources are more difficult to detect due to their lower SNR. The residual detection methods `ResidualDetectron` and iterative `LSP` correctly detect more objects than one-step detection for i magnitude greater than ≈ 25, with the `ResidualDetectron` correctly detecting more sources than iterative `LSP` for i magnitude greater than ≈ 26.

The top-right and bottom-left panels of Figure 6.15 show the distribution of cumulative counts

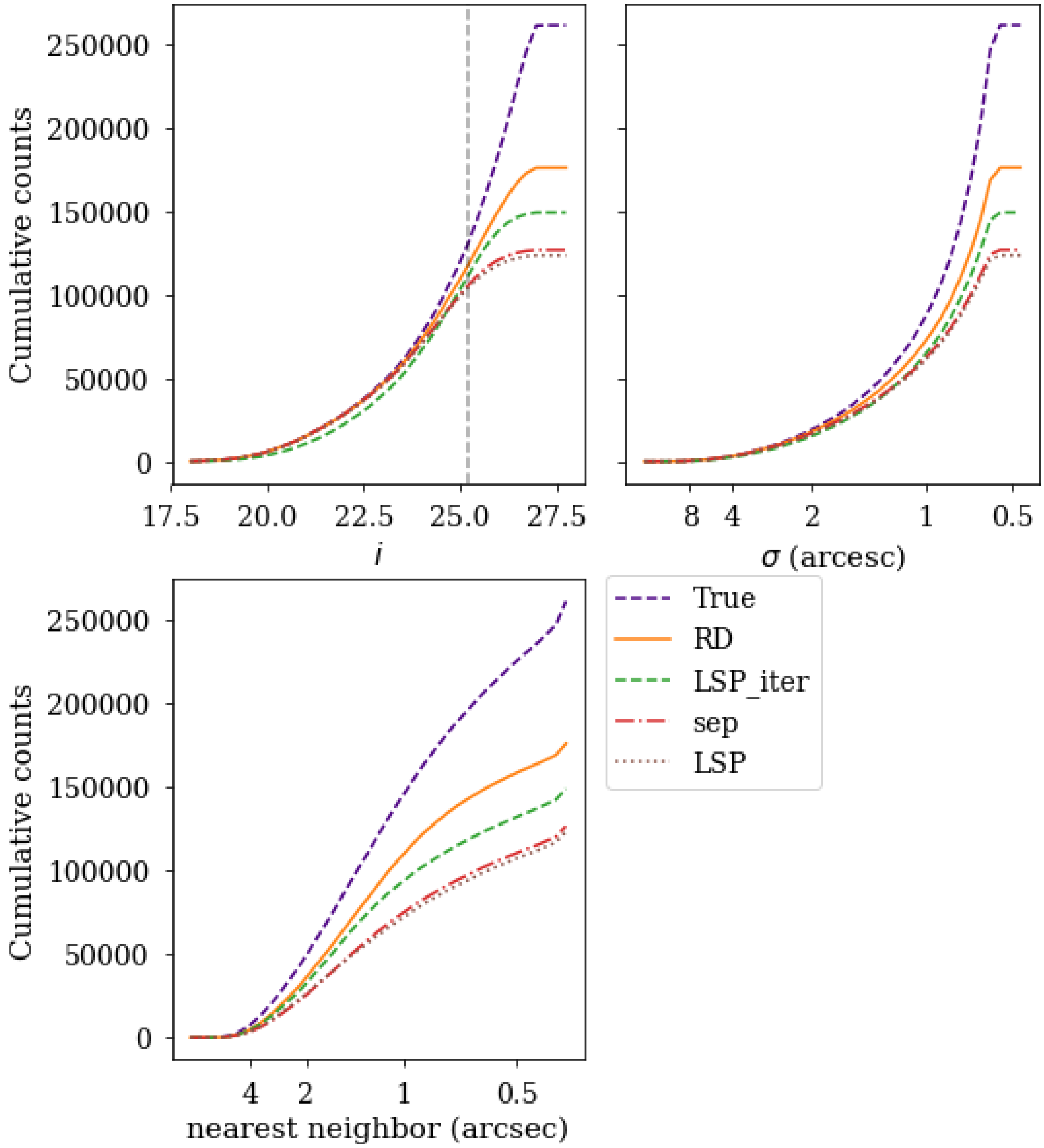


Figure 6.15: Cumulative counts of all galaxies (dashed-purple curve) and correctly detected galaxies for four different detection algorithms: `ResidualDetectron` run on six-band residual images after Source Extraction and Photometry (`SEP`), LSST Science Pipeline (`LSP`) detection run iteratively on the `LSP` residual image, `sep`, and single-step `LSP` detection. The distribution for i-band magnitude, PSF-convolved galaxy size σ, and distance to nearest neighbor for all correctly detected objects is shown for correctly detected objects in the top-left, top-right and bottom-left panels, respectively.

of detected galaxies as a function of PSF-convolved galaxy size σ and distance to nearest neighbor, respectively.

6.6 Potential limitations of the `ResidualDetectron` method

In this proof-of-concept study of residual detection methods, we assume that pixel residuals arise because `SCARLET` was not given the correct number (and location) of sources to model the scene and this modeling error alone causes it to incorrectly assign flux to each detected source. However, this is not an accurate assumption if `SCARLET` incorrectly models the scene for any other reasons. Any limitation in the `SCARLET` model of the scene will impact the viability of a residual detection technique for identifying undetected sources from residual images – especially if these modeling inaccuracies are not included in the training data. We identify some potential limitations here.

- **Complex galaxy morphology.** The simulated scenes in our training and test samples correspond to (transparent) parametrized bulge+disk galaxy profiles, with sky noise. In `SCARLET`, the scene is modeled as overlapping (transparent) two-component extended sources, with constraints on symmetry, monotonicity and sparsity of each galaxy profile. In other words, the deblending algorithm is able to accurately model sources (up to the impact of sky noise) *if* there are no unrecognized blends.

  Any deviation of galaxy profiles from these assumptions will produce residuals. If a real dataset includes galaxies with star-forming regions, for example, the deblender would likely not model these regions correctly, leading to a source of residuals that is not included in our training and test sets. As an example, we show in Figure 6.16 the limitations of the current version of `SCARLET` for modeling galaxy spiral arms.

  In our study, we did not include complex galaxy morphologies such as irregular galaxies or galaxies with star-forming regions. Since it has not been firmly established how the LSST Science Pipeline will incorporate these complexities in the deblending pipeline, we did not include them in our study.

- **Occultation of background galaxies.** Studies of high resolution HST images of overlapping galaxies show that dust from the foreground galaxies can obstruct light from the background galaxy along our line of sight (Holwerda et al., 2007). This “occultation” of galaxies leads to the flux from the background galaxy to be attenuated, with the amount of attenuation depending on the band. This extinction can break the assumption of spatial symmetry.

  A high resolution HST image of an occulting galaxy pair (Holwerda et al., 2008) is shown in Figure 6.17. Light from the large spiral galaxy in the background has undergone extinction from the dust lanes of the smaller foreground galaxy. At present, `SCARLET` does not model galaxy occultation and we do not include the effect in simulating our blend scenes. In a real

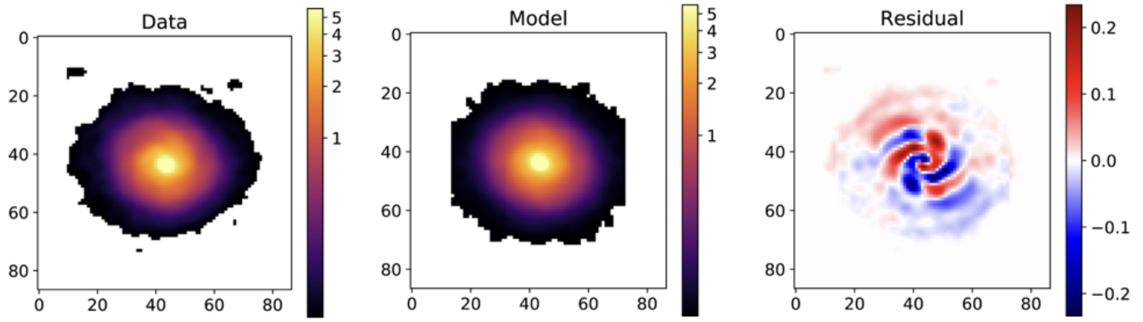


Figure 6.16: Example showing limitations of the current version of `SCARLET` for modeling a spiral galaxy from the HST COSMOS dataset. Left: Image of a spiral galaxy input to `SCARLET`. Center: The model generated by `SCARLET` with symmetry and monotonicity constraints. Right: The residual image (true image − model) reveals spiral structure that is not included in the `SCARLET` model. Figure is from Melchior et al. (2018).

dataset, galaxy occultation will lead to a source of residuals that is not included in our training and test sets.

- **Impact of stars.** We limited this study to blends produced by only overlapping galaxies. In the future, the impact of stars on detection and on the modeling of blended scenes should be studies. Our methods would need to be modified so that `SCARLET` models stars as point sources and not as extended objects, to reduce modeling errors.

- **Image artifacts.** Our simulations do not include a multitude of issues that are present in real images such as ghosting, diffraction spikes, cosmic rays and pixel saturation. Future models could incorporate masks to limit the effect of these artifacts.

- **Edge effects.** In simulating postage stamp images for the training and test sets, we did not include objects that were close enough to an edge that the galaxy image would be partially clipped. In real images, these galaxies are not likely to be accurately modeled, introducing another source of residuals.

- **Astrophysical and cosmological effects:** The galaxy simulations used in this study do not include galaxy shape correlations due to cosmic shear or intrinsic alignments, and lensing magnification. These effects can impact blending rates – albeit at levels that are not significant for this study.

- **Small-scale structure:** The simulations do not include small scale structures such as dwarf galaxies. Therefore, their impacts on blending are not included in the study.

Figure 6.17: Example showing the extinction of light from a large background spiral galaxy by a smaller foreground galaxy, in an occulting pair. Figure from Holwerda et al. (2008).

6.7 Future work

The limitations described above could be addressed in the future with higher fidelity simulations and ongoing efforts to improve `SCARLET` modeling. With improvements in the modeling algorithms and more informed priors on the model constraints, `SCARLET` will be able to better model galaxies with complex morphologies. However, for now we limited our study to modeling scenes that `SCARLET` would otherwise have correctly modeled if not for blending. Future studies will need to incorporate these complex morphologies in the training and test sample before they can be expected to perform well on real images.

Blended scenes with known fake sources injected into real images, as in Huang et al. (2018), could also be used in the future to produce a dataset of more realistic blends.

The `ResidualDetectron` was trained to predict bounding boxes around undetected galaxies. Generating target bounding boxes for training with simulated datasets is straightforward since the true source locations and sizes are known for each source. This task gets complicated when one uses real galaxy images. In addition, images for real galaxies will contain a number of features that `SCARLET` does not currently model correctly – e.g., features of irregular galaxies, star-forming regions and occultation . It will be difficult to produce large training and test datasets for which each of these components are marked.

An alternative to limiting the model to only identifying undetected source positions would be to design a network that generates `SCARLET` model images in two passes – first to model the scene using the source locations identified in the primary detection step, and then to model the scene with these sources plus the new sources identified by `ResidualDetectron`. The target of the network would be to best match this second pass model to the input blend image scene. This method, however, would require some clever optimizations in the neural network architecture and in the `SCARLET` algorithm in order to be computationally feasible.

6.8 Conclusion

In this chapter we explored two techniques to improve galaxy detection using residual images – one iteratively applies the LSST Science Pipeline `LSP` detection algorithm and the other uses a trained convolutional neural network `ResidualDetecron` for the second detection step. We use simulated galaxy blend scenes to test the two methods and compare them to primary detections made with `SourceExtractor` and `LSP`. As expected, each of these residual detection methods are found to identify galaxies that are missed in the single-step detection algorithms. `ResidualDetecron` successfully learned features corresponding to unrecognized blends in the residual images and outperformed iterative `LSP` detection, which only performs peak identification. The performance of the four detection algorithms is summarized in Table 6.3.

Table 6.3: Summary of detection performance of the four detection algorithms analyzed: `LSP`, `SourceExtractor`, `ResidualDetecron`, and iterative `LSP`.

Algorithm	Blended galaxies		Isolated galaxies	
	Precision	Recall	Precision	Recall
`LSP`	0.998	0.472	0.999	0.978
`SourceExtractor`	0.999	0.485	0.999	0.932
`ResidualDetecron`	0.932	0.674	0.795	0.867
Iterative `LSP`	0.891	0.571	0.863	0.913

Among the four detection algorithms, `ResidualDetecron` is able to achieve the highest recall for detections on blended galaxies, correctly identifying ≈67% of the galaxies. As expected, both the residual detection methods, `ResidualDetecron` and iterative `LSP`, produce more false positives than the single-step detection methods.

The lower precision for primary `SourceExtractor` detections on isolated objects translates into lower precision and recall values for `ResidualDetecron` detections as well. Future studies could test the performance of `ResidualDetectron` with `LSP` as the primary detection algorithm for a better comparison between `LSP`, iterative `LSP` and `ResidualDetectron`.

We limited this study to a proof of concept demonstration of using CNNs to improve object detection. Our network was trained and tested on simulated galaxy images. Future studies could retrain `ResidualDetecron` with more realistic galaxy images, and also including galaxy blends from overlap with stars.

Chapter 7

Conclusions

One of the yet-to-be-solved puzzles of the 21st century in cosmology is the nature of dark energy – the source of accelerating expansion that our universe is experiencing at present. As discussed in Chapter 1, constraining dark energy parameters using observational probes requires high precision measurements of astrophysical sources. The Legacy Survey of Space and Time (LSST) will play an important role in constraining the dark energy parameters. This thesis is devoted to a deeper understanding of two potential sources of observational systematic effects in LSST weak lensing (WL) measurements.

7.1 Color gradients

We discussed in the introduction to Chapter 2 the importance of correcting for PSF distortions in WL measurements and the potential significance of the effects of galaxy color gradients (CG) in these corrections. We described how the combination of a chromatic PSF and a spatial weight function in shape estimates can lead to incorrect shear estimates. We then introduced our method to isolate any bias from color gradients – by comparing the shear estimated from a galaxy with CG to the shear estimated from an equivalent galaxy without CG. We quantified the CG bias as multiplicative and additive terms in an assumed linear relation between true and observed shear values. We first studied the magnitude of CG bias for parametric galaxies with extreme color gradients. The CG bias was shown to depend on the size of the weight function and the chromaticity of the observing PSF. While the estimated CG bias was smaller than the requirements on LSST shear measurements from *all* systematic effects, it was not negligible. To beter understand the expected CG bias for the LSST, we measured the bias on two ensembles of galaxies – one using noise-free parametric simulations and the other using noisy real galaxies.

CG bias was estimated for ensembles of simulated parametric galaxies from the CatSim catalog. The value of half the maximum span of the multiplicative CG bias, $m_{\text{CG}}(z)$, in the redshift range

[0, 1.2], was found to be $\leq 10^{-4}$. This is an order of magnitude smaller than the LSST full-depth requirement of 0.003 on the total systematic uncertainty in the redshift-dependent shear calibration m_z (Equation 1.19). We study the impact of pixel noise on the ability of our method to isolate CG bias by comparing m_{CG} estimated from simulations with and without noise. We observe a strong dependence on the galaxy SNR due to contributions from effects other than CG.

We also estimated CG bias from real galaxy images in the AEGIS survey rendered as seen by LSST. For AEGIS galaxies with HST I-band SNR > 200, half of the maximum span of $m_{\mathrm{CG}}(z)$ in the redshift range [0, 1.2] was $\leq 1.5 \times 10^{-4}$.

We conclude from this study that for both the noise-free parametric galaxies and for the high-SNR AEGIS galaxies for which our method is sensitive to CG bias, the half-maximum span is at least a factor of two less than the LSST full-depth requirement. These results are important because shear estimation depends on accurate knowledge of the PSF and the effects of color gradients on the per-galaxy PSF could be very challenging to predict and correct. We are thus optimistic that color gradients will not be a source of limiting systematic uncertainty for LSST.

7.2 Blending

The unprecedented depth of LSST observations introduces the problem of blending in which the observed images of objects overlap with each other. We discussed in Chapter 3 how systematic errors can be introduced during detection, deblending and measurement of these blended objects. We then focused on one subset of blended objects called unrecognized blends, where the detection algorithm fails to detect one or more sources in a blend.

In Chapter 4, we measured the performance of the LSST Science Pipeline (`LSP`) detection algorithm for blended groups of galaxies, with a focus on the rate of unrecognized blends. Measurements on our test dataset of highly overlapping galaxies ($\approx 70\%$ unrecognized blends) showed that $\approx 80\%$ of the undetected galaxies in these unrecognized blends are in an LSST redshift bin that is different than that of their closest matched neighbor.

We discussed in Chapter 6 the residual image of a blend scene defined as the difference between the observed image and the deblended image, which can be used to inform the presence of an undetected source. We proposed and studied two methods to identify unrecognized blends by analyzing the residual images.

First, we developed and successfully trained a convolutional neural network (CNN), called `ResidualDetectron`, on the residual images for the `SCARLET` multi-band deblender, for simulated galaxy blends, to identify the center and size of undetected objects. The trained CNN model was able to identify $\approx 50\%$ of the two-galaxy groups that were unrecognized blends in the primary (single-step) detection. `ResidualDetectron` leads to a detection "recall" (or completeness) value of 57% in blends of two to ten galaxies, compared to 48% recall for the primary `SourceExtractor` detection algorithm.

We also proposed a second method in which we use the LSST science pipeline (LSP) iteratively, once on the input image and again on the `SCARLET` residual image. Iterative LSP was also able to identify undetected sources from residuals. However, because LSP detection is a peak-finding algorithm, any local maxima in the residual above the detection threshold was predicted as a source, thereby increasing the false-positive rate.

The detection efficiencies measured for `ResidualDetectron` and iterative LSP were both better than one-stage detection with LSP, with the CNN-based `ResidualDetectron` providing lower false positives and higher efficiency than iterative LSP. This demonstrates that the performance of the LSST detection pipeline could be significantly improved – particularly for identifying unrecognized blends – by including residual detection in the analysis pipeline.

7.3 Final Thoughts

This is going to be a very interesting decade for cosmology. LSST is expected to see first light by 2021. Within the first ten years of observations, it will map the sky with combined depth and sky coverage never achieved before, generating an enormous volume of data. The statistical precision expected from measurements of billions of observed galaxies leaves very little room for error on the cosmological parameter values.

In the words of the 1817 British Parliamentarian William Lamb, "the possession of great power necessarily implies great responsibility." The high statistical power expected from LSST requires that all systematic uncertainties and biases are lower than a threshold defined so that they do not dominate the total uncertainty. This makes it crucial that observational systematic effects that were previously ignored are carefully studied for the LSST. This thesis is an attempt to quantify and/or remedy two such systematic effects.

While we can conclude from our study of galaxy color gradients that these will *not* contribute significant systematic biases in cosmic shear measurements, the effects of blending remain a challenge for the deep LSST images. We showed that performance in the first step in the blending challenge – source detection – can be improved by analyzing residual images. Even if successful deblending of some of these detected blends is not possible, the knowledge of their existence can inform cosmological analyses and reduce systematic errors. We showed that our proposed neural network (NN) model, `ResidualDetectron`, and the iterative LSP detection method are able to identify unrecognized blends, with the NN model outperforming LSP. The ability of neural networks to utilize abstract features in images makes them a powerful tool for problems without parametric solutions. We predict that NN algorithms used to augment the standard algorithms will be a valuable tool for the analysis of LSST images at ten-year depths.

Appendix A

Testing `ChromaticRealGalaxy`

In this section we describe the tests conducted to estimate the accuracy of the `ChromaticRealGalaxy` (CRG) algorithm for reproducing chromatic features from noisy images. We simulate parametric galaxies as they would be observed by HST (henceforth called "HST-like" galaxies) – i.e., convolved with an HST-like chromatic PSF and observed in the HST V and I bands with correlated pixel noise. `ChromaticRealGalaxy` is then used to model the SBP of the HST-like galaxies in order to estimate the CG bias when observed by LSST. We then compare this estimated CG bias to the CG bias estimated in Section 2.4 from noise-free parametric galaxies simulated directly in LSST bands.

A.1 Testing CRG with the reference galaxy with extreme color gradients

We first test the ability of CRG to model the reference galaxy with extreme color gradients from V- and I-band HST-like images of the galaxy. In particular, we investigate two potential limitations:

1. Impact of imperfect SEDs: For real images where the SEDs of the galaxies are unknown, we approximate the SEDs as linear. We test the impact of these approximate SEDs by comparing the estimated CG bias when the CRG is modeled with linear SEDs and with true SEDs.
2. Impact of pixel noise: The simulations analyzed in Section 2.4 were noise free. However, real galaxy images include pixel noise. We test the impact of this noise by comparing CG bias estimates from noisy and noise-free HST-like images.

A flowchart describing the CRG test with the reference galaxy with extreme color gradients is shown in Figure A.1. The red dashed box A shows the method used in Section 2.4 to estimate CG bias of the reference parametric galaxy when seen by LSST. Boxes 1 and 2 show the CRG tests to study the impact of imperfect SEDs and pixel noise, respectively.

For Box 1, the HST-like images of the reference galaxy in V and I bands are input to CRG along with the effective HST PSF images. Two chromatic SBP of the galaxy are produced: one for CRG modeled with the true bulge and disk SEDs and the other with linear SEDs. Each of the modeled CRG profiles are then used to estimate CG bias similar to Box A.

We study the impact of noise in the input HST images (Box 2) by drawing the HST-like reference galaxy images with pixel noise in V and I bands. These noisy images are then input to CRG to model the chromatic SBP with linear SEDs. The chromatic SBPs are then used to estimate CG shear bias. Since the analysis would likely depend on the galaxy S/N, we perform the study for varying noise levels for the input HST-like images.

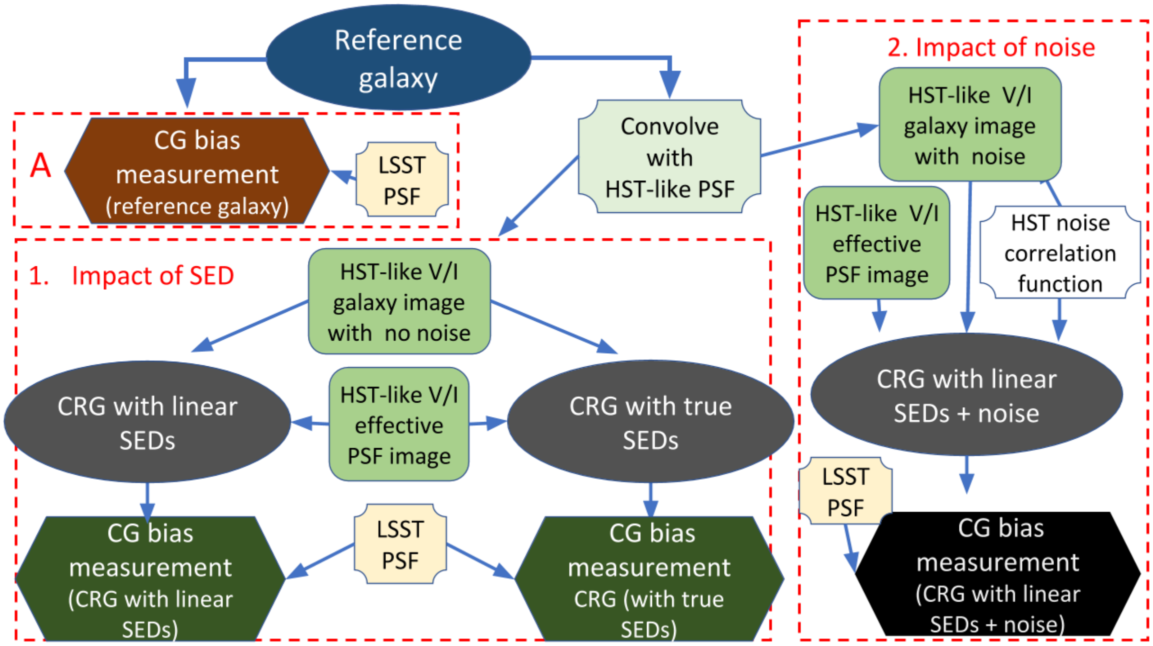


Figure A.1: Flowchart describing the method for testing the impact of (1) the linear approximation for the SED and (2) pixel noise, on the CRG model using HST-like images of the reference parametric galaxy with extreme color gradients. The CG bias estimated from 1 and 2 are then compared to the estimates from noise-free parametric simulations in A. See Appendix A.1 in text for details.

A.1.0.1 Parameters of HST-like images

The bulge and disk SBPs have the same parameters as described in Section 2.4.3.1. The HST and LSST parameters used in the simulations are summarized in Table A.1. The HST-like images are drawn with a pixel scale of 0.03 arcsec and the LSST images with pixel scale 0.2 arcsec. The chromatic PSFs are assumed to have a Gaussian profile with a wavelength-dependent size as described in Equation 2.1. The chromatic HST PSF is assumed to be diffraction-limited with wavelength-dependent size scaling exponent $\alpha = +1.0$. The PSF size at a reference wavelength of $\lambda^o = 806\,\text{nm}$ is set equal to $\sigma^o = 0.071\,\text{arcsec}$, which is the mean value from fits to real I-band HST PSF images. The HST noise is modeled using the same noise correlation function described in Section 2.5.3.2.

Table A.1: HST and LSST parameters.

Survey	Observing bands	α	PSF σ^o (arcsec)	PSF λ^o (nm)	Pixel size (arcsec)	Collecting area (m^2)
HST	V/I	+1.0	0.071	806	0.03	4.44
LSST	r/i	−0.2	0.297	550	0.2	32.4

Note. The PSF is modeled as a circular Gaussian with size σ^o at reference wavelength λ^o. The wavelength-dependent PSF size is determined by the scaling exponent α (see Equation 2.1).

A.1.0.2 Results of CRG tests with reference galaxy with extreme color gradients

In Figure A.2 we show the estimated multiplicative shear bias due to CG, m_{CG}, as a function of redshift for the reference galaxy with the three different disk SEDs (CWW-Im, CWW-Sbc, CWW-Scd) and a common CWW-E bulge SED. The estimated bias m_{CG} from parametric simulations is shown as the solid blue line, identical to Figure 2.7. We show the bias values measured by modeling the reference galaxy using `ChromaticRealGalaxy` with the true SED (orange squares) and with a linear SED (teal dots). As expected, the biases estimated for the parametric simulation are in excellent agreement with the CRG-modeled SBP with true SEDs. However, the predicted biases for CRG with linear SEDs differ from these values, revealing the limitations in using zero and first-degree polynomials to reproduce the nonlinear features in the bulge and disk SEDs. In particular, CRG with a linear SED tends to smooth the variation in bias with redshift. The difference between the redshift-averaged CG bias estimated with the linear SED and with the true SED is $\mathcal{O}(10^{-4})$, and the estimated biases satisfy the LSST DESC requirement of 0.003 on the total systematic uncertainty in the redshift-dependent shear calibration (yellow dashed lines) for both the linear and true SED models, for the three disk SEDs. However, the total variation in CG bias over the redshift range $[0, 1.2]$ is a factor of 2.6, 1.9 and 2.5 less when calculated with the linear SED rather than the true SED, for the CWW-Im, CWW-Sbc and CWW-Scd disk SEDs, respectively.

To assess the impact of noise we simulate HST-like images for the reference galaxy SBP with E-type bulge and Im-type disk SED, at a redshift of 0.3, but with different noise levels. As illustrated in Box 3 of Figure A.1, we add HST-like noise to the simulated galaxy images before they are modeled with CRG. Since our noise-free analysis already showed that the predicted biases for CRG with true SEDs are identical to those for the parametric simulations, we perform the noise study for CRG with linear SEDs only. Since our ultimate aim is to model AEGIS galaxies with CRG, we add noise

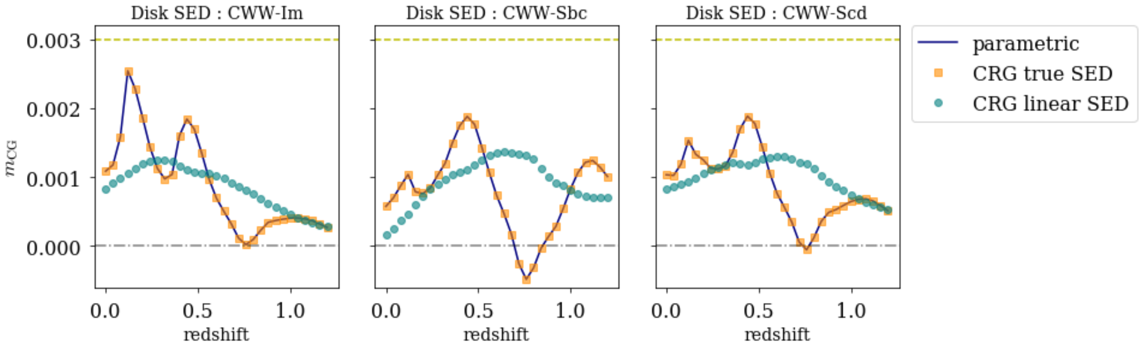


Figure A.2: Multiplicative shear bias due to CG, $m_{\rm CG}$, as a function of redshift for a reference galaxy with CWW-E bulge spectrum and three different disk SEDs (from left to right: CWW-Im, CWW-Sbc, CWW-Scd), for three types of noise-free simulations: parametric (solid blue line), CRG with the true SED (orange squares), and CRG with linear SEDs (teal dots). The yellow dashed line shows the LSST DESC requirement of 0.003 on the total systematic uncertainty in the redshift-dependent shear calibration.

levels similar to those in the AEGIS images.

The noisy reference galaxy is simulated 10,000 times, each with a different correlated noise realization to obtain an SNR level drawn randomly from the AEGIS catalog. The sampled I-band SNR values are shown in the left panel of Figure A.3. The CG bias is then estimated using the CRG model of each of the noisy images.

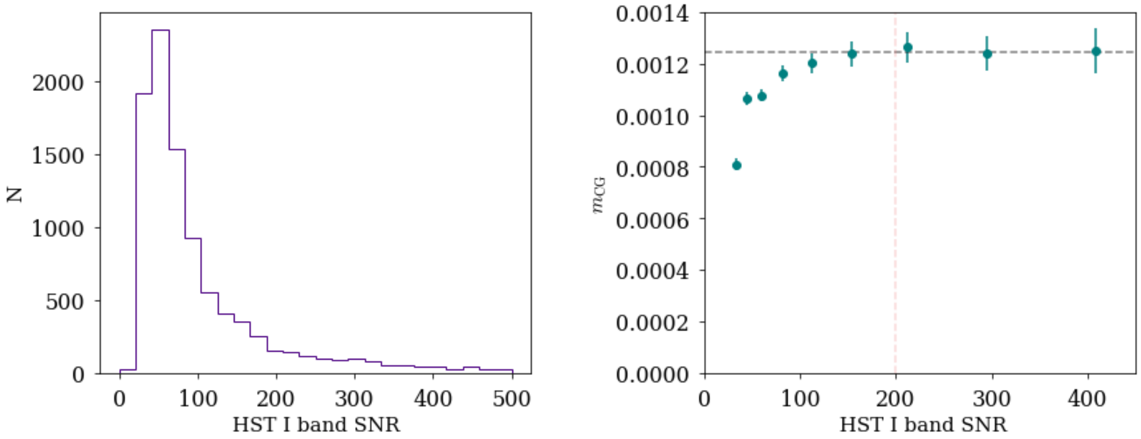


Figure A.3: Left: Distribution of I-band SNR of 10,000 randomly selected galaxies from the AEGIS catalog. Right: CG shear bias as a function of I-band SNR of the simulated HST-like images of the reference galaxy with extreme color gradients at a redshift of 0.3, modeled with CRG. The points correspond to the mean $m_{\rm CG}$ in each SNR bin and error bars show the statistical uncertainty on the mean. The black dashed line corresponds to the value of $m_{\rm CG}$ computed from the noise-free image of the galaxy modeled by CRG with linear SEDs ($m_{\rm CG} = 1.24 \times 10^{-3}$).

The measured biases are binned by SNR in the right panel of Figure A.3. The solid dots denote the mean binned values of the measured CG bias and the error bars correspond to the statistical uncertainty on the means. The estimated CG bias shows a dependence on SNR even though all points correspond to the same galaxy SBP and thus have the same intrinsic color gradient. The dashed line at +0.00124 corresponds to the estimated CG bias for the reference galaxy at a redshift of 0.3 drawn with no noise and modeled by CRG with linear SEDs. At high SNR, the mean value of the m_{CG} estimate approaches this noise-free CG bias. However, at low SNR the estimate diverges from the noise-free CG bias indicating that our methodology to isolate CG bias is not robust against noise bias. Hence we are limited to estimating CG bias for real galaxies with HST I-band SNR greater than 200.

A.2 Testing CRG with CatSim galaxies

We extend the CRG tests in the presence of noise to an ensemble of galaxies by repeating our analysis in Box 2 of Figure A.1 to galaxies in the CatSim catalog with a range of bulge and disk parameters. Since we concluded above that our methodology is applicable to only high SNR AEGIS galaxies, we compute the SNR of the HST-like galaxy as well as the SNR at 10-year LSST depth.

A flowchart describing the procedure is shown in Figure A.4. The left side of the flowchart illustrates the method used in Section 2.4.4.2 to compute CG bias for the parametric CatSim galaxies as measured by LSST and does not involve modeling with CRG. We also draw the galaxies in the r band with sky noise corresponding to the 10-year LSST depth and compute their SNRs.

The right side of the flowchart (in green) illustrates how we test CRG with HST-like CatSim galaxy images as input. The V- and I-band galaxy images are input to CRG, along with the HST effective PSF image and the correlation function for pixel noise, to model a chromatic SBP. CG bias is then estimated for these galaxies when observed in LSST r and i bands. For each noisy HST-like CatSim galaxy image, the V- and I-band SNRs are also computed.

A.2.0.1 Estimating SNR of CatSim galaxies

The results of the SNR calculations for the galaxies are shown in Figure A.5. The top panels show the computed SNR for the CatSim galaxies when seen by HST V and I bands (left) and LSST r and i bands (center). 53% of the galaxies have I-band SNR less than 200. We plot the measured LSST i-band SNR against the true i-band CatSim catalog magnitude in the top right panel; the lower magnitude (i.e., brighter) galaxies have higher SNR as expected. In the bottom left plot, we compare HST I-band SNR and LSST i-band SNR. We do not expect the SNR in the two bands to be the same since the bandpasses, exposure times, collecting areas, and noise levels are different (see Table A.1). However, since we have added a constant noise to all the HST and LSST simulations individually, we do expect the two SNRs to be linearly related. We observe that LSST i-band SNR

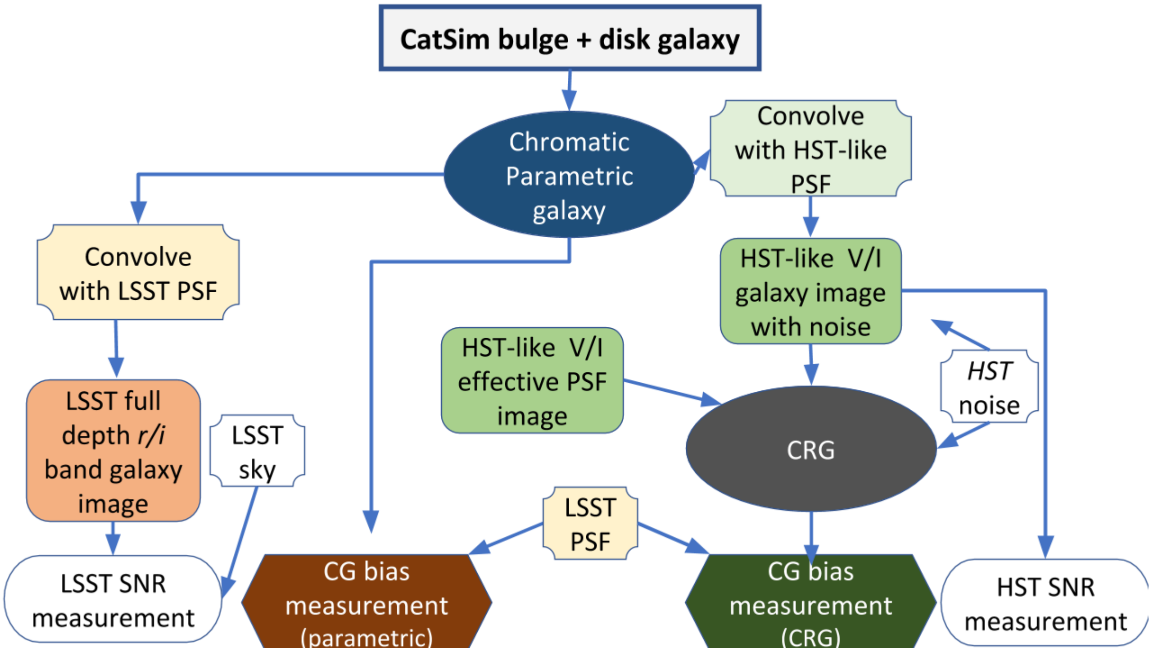


Figure A.4: Flowchart describing the method used to test CRG by measuring CG shear bias with galaxies parameterized in CatSim. See Appendix A.2 in text for details.

is ≈ 0.7× HST I-band SNR. A selection criterion of I-band SNR > 200 for the AEGIS galaxies effectively eliminates all galaxies with LSST i-band SNR less than ≈ 80.

In the final two panels, we show the correlation between I-band SNR and redshift (center) and $r-i$ color (right). The magnitudes, redshifts and colors are from the CatSim catalog and are noise free. High redshift galaxies have, on average, smaller SNRs, which is expected since the higher redshift galaxies tend to have lower observed flux. As a consequence, our CG bias estimates from AEGIS galaxies are not as reliable for higher redshift galaxies. There appears to be a weak correlation between SNR and color.

A.2.0.2 CG shear bias estimates from noisy CatSim galaxies with CRG

Figure A.6 plots the results from the CG bias estimates for all the noisy HST-like CatSim galaxies modeled by CRG and observed in LSST r band (top) and i band (bottom). For reference we also show the estimates from noise-free parametric simulations (blue). The panels on the left show histograms of the estimated CG bias from noisy HST-like CatSim images modeled by CRG with true SEDs (orange) and linear SEDs (teal). While the noise-free parametric galaxies exhibit small biases, the estimates from noisy CRG-modeled images show large tails due to contributions from pixel noise. This is more evident in the right panels where we plot the mean CG bias estimates for galaxies with SNR higher than min(SNR) in the HST I band. Noise impacts the CG bias estimates for both linear

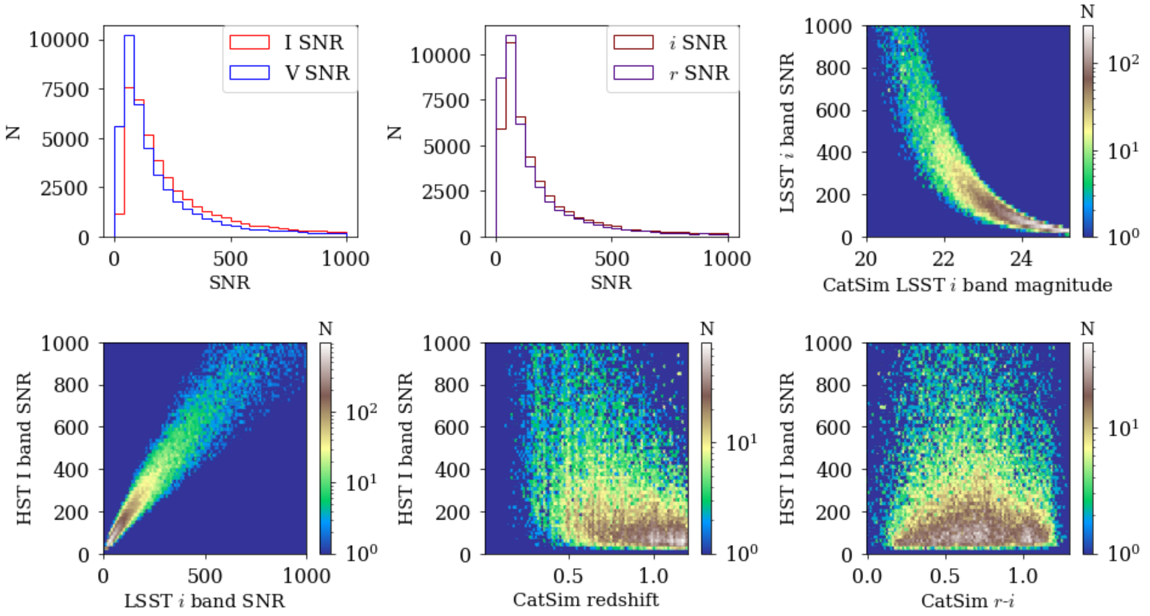


Figure A.5: HST and LSST SNR distributions for CatSim galaxies. Histograms of SNR for HST V and I bands (top left) and LSST r and i bands (top center). Calculated LSST i-band SNR versus i-band magnitude (top right), and HST I-band SNR versus LSST i-band SNR (bottom left). HST I-band SNR versus redshift (bottom center) and LSST $r-i$ color (bottom right).

and true SED models at low SNR. With no SNR cut on the HST-like images, the mean CG bias estimates from CRG galaxies modeled with linear SEDs in r band is $m_{\text{CG}} = (-1.7 \pm 0.5) \times 10^{-3}$ while the median of the distribution is -4.1×10^{-6} with a median absolute deviation (MAD) of 3.3×10^{-3}. The mean and median bias estimates in the i band are 3.2×10^{-4} and 4.1×10^{-6}, respectively.

The pixel noise causes the bias estimated from HST-like galaxies modeled by CRG with true SEDs to be more negative in both bands, while the impact of imperfect SEDs causes the bias to be more negative in the r band and positive in the i band. The impact of noise bias on the CG bias estimates is less pronounced in the i band. This is because the galaxy images have higher SNR in i band than r band as shown in Figure A.5. As more low-SNR galaxies are excluded, the bias measured with the CRG method approaches the value determined with the parametric analysis. For galaxies with HST I-band SNR > 200, the mean color gradient bias is $m_{\text{CG}} = (-2.5 \pm 0.3) \times 10^{-4}$ in the r band and $m_{\text{CG}} = (0.47 \pm 0.10) \times 10^{-4}$ in the i band. The LSST requirement on redshift dependence of the shear bias from all systematic effects is shown as the dashed yellow line.

Since the ability of `ChromaticRealGalaxy` to model chromatic features and accurately estimate the magnitude of CG bias depends on the intrinsic properties of the galaxies, we investigate the dependence of the estimated CG bias on the galaxy color and redshift. In Figure A.7, we show the

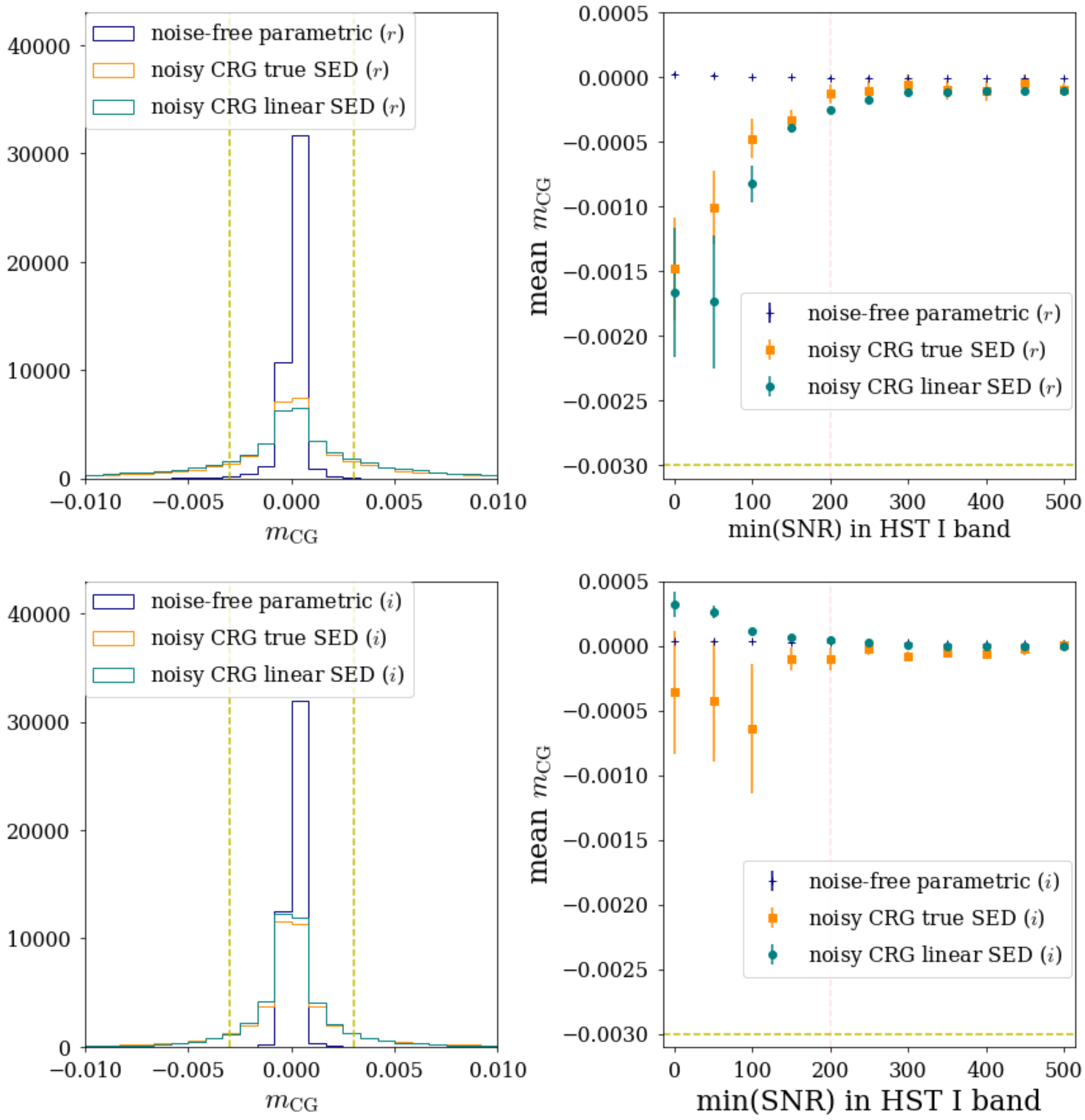


Figure A.6: Left panels: Distribution of m_{CG} in the r band (top) and i band (bottom) for CatSim galaxies. Right panels: Mean m_{CG} in the r band (top) and i band (bottom) for galaxies with I-band SNR > min(SNR). Results from parametric simulations are shown for true SEDs in blue, for CRG with linear SEDs in teal, and CRG with true SED input in orange. The yellow dashed lines show the LSST DESC requirement of 0.003 on the total systematic uncertainty in the redshift-dependent shear calibration.

dependence of estimated CG bias on $r - i$ color (left) and redshift (right). The top plots correspond to noise-free parametric simulations of CatSim galaxies. The bottom plots illustrate the impact of applying three different minimum SNR cutoffs. The dots denote the difference in mean m_{CG} for noise-free simulations (top panel) and mean m_{CG} for HST-like galaxies modeled with CRG (Δm_{CG}), with HST I-band SNR >100, 200 and 1000. The difference in the mean bias decreases with increasing SNR cutoffs. Selection effects due to the SNR cutoffs are found to bias the binned means of CG bias estimates for noise-free parametric simulations by $\mathcal{O}(10^{-4})$. Therefore, our method of using `ChromaticRealGalaxy` to model galaxies observed by LSST from noisy HST images is able to reproduce CG bias with an sensitivity of $\mathcal{O}(10^{-4})$ for galaxies with HST I-band SNR>200.

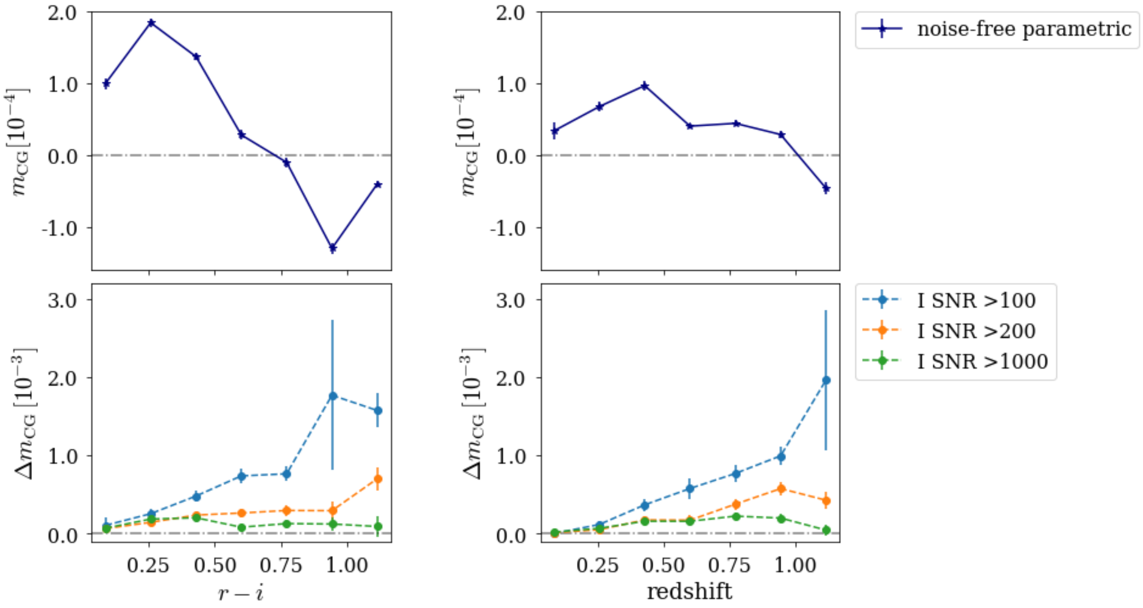


Figure A.7: Multiplicative shear bias from CG, m_{CG}, as a function of $r-i$ color (left) and redshift (right) for CatSim galaxies. Top: m_{CG} estimates from noise-free parametric simulations. Bottom: Difference in mean m_{CG} estimated from noise-free parametric simulations and HST-like galaxies modeled with CRG (Δm_{CG}) for three different HST I-band SNR cutoffs. The error bars correspond to the statistical uncertainties on the binned means.

Appendix B

Blending Tool Kit: Framework for testing algorithms

In this appendix we describe the framework developed to generate images of blends and evaluate algorithm performance. The aim of the framework is so that users can run any detection, deblending and/or measurement algorithms of their choice on the same input images so as to facilitate ease of comparing performances. A schematic describing the workflow is shown in Figure B.1. The blue boxes denote `btk` functions, while grey ellipses are user inputs.

The parts of the workflow are described here:

1. `config`: Set parameter values to create postage stamps: size of stamp, number of objects per blend, and how many stamps are to be drawn in one batch i.e., the number of stamps `btk` will produce together in a single batch. The observing survey name (e.g., LSST, DES), the name of the input catalog to draw objects from, and the names of the observing bands are also specified here.

2. `load_catalog`: A `CatSim`-like catalog from which to sample galaxies. This step includes an option to input user-defined selection criteria.

3. `create_blend_generator`: The sources for the blend scene in a postage stamp are defined here. Based on the user-defined sampling function, galaxies from the input catalog are sampled to make a "true sources" catalog for each blend scene. A `Python` generator is created that returns a new blend catalog every time it is called.

4. `create_observing_generator`: Creates a `Python` class that contains information about the observing conditions such as PSF parameters, noise level, etc. The default is to use the full-depth values corresponding to the survey named in `config`. Alternately, the user can input a

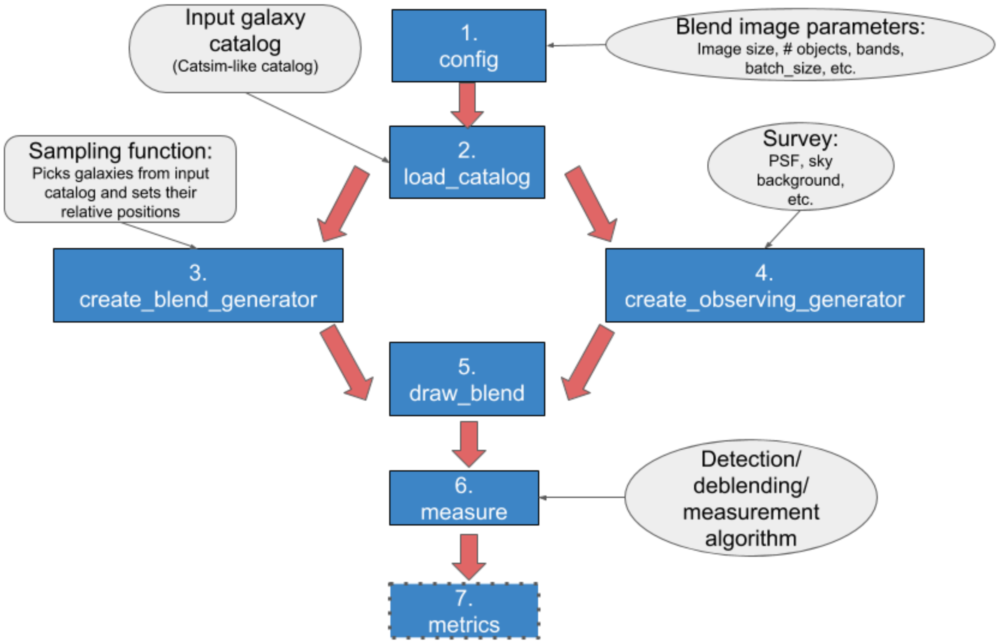


Figure B.1: Schematic of the `BlendingToolKit` workflow.

function to define these parameter values. A `Python` generator is created that returns a new set of observing conditions every time it is called.

5. `draw_blend`: Simulates scene of overlapping objects, convolved by the PSF and, if specified, with pixel noise applied in all the observing bands defined in `config`. If specified, an "isolated image of each object is also drawn without pixel contributions from other objects, in each band.
6. `measure`: A user input class to perform detection, deblending and/or measurement to run on the postage stamp images generated by `draw_blend`.
7. `metrics`: Compares the true centroids, shapes, and/or flux values to values predicted by the user-input algorithm in `measure`.

Acronyms

`LSP` LSST Science Pipeline. xi, 56, 65, 108, 125, 126, 139

`SEP` Source Extraction and Photometry. 118, 123, 132

CG color gradient. 18

CNN convolutional neural network. 81, 83, 86, Glossary: CNN

FWHM full width at half maximum. 34, Glossary: FWHM

HLR half-light radius. 30, Glossary: HLR

PSF point-spread function. 18, 110, 141, Glossary: PSF

S/N signal-to-noise ratio. 28, 82, 90, 142, Glossary: SNR

SBP surface brightness profile. 24, 66, 110

SED spectral energy distribution. 12, 18, 109, Glossary: SED

WL Weak lensing. 52

Symbols

B_i Sky background in the ith band. 112

$I_i(\vec{x})$ Observed image in the ith band. 112

$M_i^{\mathrm{map}}(\vec{x})$ `SCARLET` model image with asinh mapping in the ith band. 113

$P_i(\vec{x})$ Residual pull image in the ith band. 112

$\Pi(\vec{x}, \lambda)$ point-spread function (PSF). 20

α Scaling exponent governing PSF chromaticity. 21

e Complex ellipticity spinor of sheared galaxy shape. 13

g Complex reduced shear. 24

r_{sec} Galaxy size computed from second moments as described in A1 of Chang et al. (2013). 72, 76

Glossary

`SCARLET` Multi-band source separation algorithm. 60, 92

CatSim Catalog Simulator: Catalog managed by the LSST Systems Engineering group. 20, 110

CNN convolutional neural network: class of deep neural networks, most commonly applied to analyzing visual imagery. 81

cosmoDC2 Catalog of extra-galactic objects produced for the LSST DESC Data Challenge 2. 110

deblending Separating flux in the pixels of two or more overlapping sources.. 58

FWHM full width at half maximum: For circularly symmetric SBP, diameter of circle at which the surface brightness is half the peak surface brightness.. 34

HLR half-light radius: Radius within which half the galaxy flux is contained. 30

PSF Point-spread function: Response of imaging system to a point source. 18, 153

SED spectral energy distribution: Energy emitted as a function of wavelength. 12

SNR signal-to-noise ratio: Ratio of measured flux to uncertainty on the flux. 28

Bibliography

Abdulla, W. 2017, Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow, `https://github.com/matterport/Mask_RCNN`

Aihara, H., Arimoto, N., Armstrong, R., et al. 2018, The Hyper Suprime-Cam SSP Survey: Overview and survey design, PASJ, 70, S4

Akeson, R., Armus, L., Bachelet, E., et al. 2019, The Wide Field Infrared Survey Telescope: 100 Hubbles for the 2020s, arXiv e-prints, arXiv:1902.05569

Albrecht, A., Bernstein, G., Cahn, R., et al. 2006, Report of the Dark Energy Task Force, arXiv e-prints, astro

Amiaux, J., Scaramella, R., Mellier, Y., et al. 2012, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 8442, Proc. SPIE, 84420Z

Barbary, K. 2018, SEP: Source Extraction and Photometry, ascl:1811.004

Bartelmann, M., & Schneider, P. 2001, Weak gravitational lensing, Phys. Rep., 340, 291

Beard, S. M., MacGillivray, H. T., & Thanisch, P. F. 1990, The Cosmos system for crowded-field analysis of digitized photographic plate scans., MNRAS, 247, 311

Behroozi, P., Wechsler, R. H., Hearin, A. P., & Conroy, C. 2019, UNIVERSEMACHINE: The correlation between galaxy growth and dark matter halo assembly from z = 0-10, MNRAS, 488, 3143

Berlind, A. A., Frieman, J., Weinberg, D. H., et al. 2006, Percolation Galaxy Groups and Clusters in the SDSS Redshift Survey: Identification, Catalogs, and the Multiplicity Function, ApJS, 167, 1

Bernstein, G. M., & Jarvis, M. 2002, Shapes and Shears, Stars and Smears: Optimal Measurements for Weak Lensing, AJ, 123, 583

Bertin, E. 1994, Classification of Astronomical Images with a Neural Network, Ap&SS, 217, 49

Bertin, E., & Arnouts, S. 1996, SExtractor: Software for source extraction., A&AS, 117, 393

Bolzonella, M., Miralles, J.-M., & Pelló, R. 2000, Photometric redshifts based on standard SED fitting procedures, A&A, 363, 476

Bosch, J., Armstrong, R., Bickerton, S., et al. 2018, The Hyper Suprime-Cam software pipeline, PASJ, 70, S5

Bottou, L., & Bousquet, O. 2007, in Proceedings of the 20th International Conference on Neural Information Processing Systems, NIPS'07 (USA: Curran Associates Inc.), 161–168

Boucaud, A., Huertas-Company, M., Heneka, C., et al. 2019, Photometry of high-redshift blended galaxies using deep learning, arXiv e-prints, arXiv:1905.01324

Bruzual A., G., & Charlot, S. 1993, Spectral evolution of stellar populations using isochrone synthesis, ApJ, 405, 538

Burke, C. J., Aleo, P. D., Chen, Y.-C., et al. 2019, Deblending and classifying astronomical sources with Mask R-CNN deep learning, MNRAS, 490, 3952

Carroll, S. M. 2004, Spacetime and geometry. An introduction to general relativity

Chang, C., Jarvis, M., Jain, B., et al. 2013, The effective number density of galaxies for weak lensing measurements in the LSST project, MNRAS, 434, 2121

Coleman, G. D., Wu, C.-C., & Weedman, D. W. 1980, Colors and magnitudes predicted for high redshift galaxies, ApJS, 43, 393

Condon, J. J. 1974, Confusion and Flux-Density Error Distributions, ApJ, 188, 279

Connolly, A. J., Angeli, G. Z., Chandrasekharan, S., et al. 2014, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9150, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 14

Csaji, B. 2001, Approximation with Artificial Neural Networks, M.S.'Thesis, Dept. Science, Eotvos Lorand Univ., Budapest, Hungary

Cybenko, G. 1989, Approximation by superpositions of a sigmoidal function, Mathematics of Control, Signals and Systems, 2, 303

Cypriano, E. S., Amara, A., Voigt, L. M., et al. 2010, Cosmic shear requirements on the wavelength dependence of telescope point spread functions, MNRAS, 405, 494

Dark Energy Survey Collaboration, Abbott, T., Abdalla, F. B., et al. 2016, The Dark Energy Survey: more than dark energy - an overview, MNRAS, 460, 1270

Davis, M., Guhathakurta, P., Konidaris, N. P., et al. 2007, The All-Wavelength Extended Groth Strip International Survey (AEGIS) Data Sets, ApJ, 660, L1

Dawson, W. A., Schneider, M. D., Tyson, J. A., & Jee, M. J. 2016, The Ellipticity Distribution of Ambiguously Blended Objects, ApJ, 816, 11

De Lucia, G., Springel, V., White, S. D. M., Croton, D., & Kauffmann, G. 2006, The formation history of elliptical galaxies, MNRAS, 366, 499

DeRose, J., Wechsler, R. H., Becker, M. R., et al. 2019, The Buzzard Flock: Dark Energy Survey Synthetic Sky Catalogs, arXiv e-prints, arXiv:1901.02401

DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016, The DESI Experiment Part I: Science,Targeting, and Survey Design, arXiv e-prints, arXiv:1611.00036

Dicke, R. H., Peebles, P. J. E., Roll, P. G., & Wilkinson, D. T. 1965, Cosmic Black-Body Radiation., ApJ, 142, 414

Dieleman, S., Willett, K. W., & Dambre, J. 2015, Rotation-invariant convolutional neural networks for galaxy morphology prediction, MNRAS, 450, 1441

Dodelson, S. 2003, Modern cosmology

Drlica-Wagner, A., Sevilla-Noarbe, I., Rykoff, E. S., et al. 2018, Dark Energy Survey Year 1 Results: The Photometric Data Set for Cosmology, ApJS, 235, 33

Emi, B. 2018, Expanding the Chromatic Range of Galaxies for Weak Gravitational Lensing Simulation, Stanford Undergraduate Research Journal, 15, 105

Er, X., Hoekstra, H., Schrabback, T., et al. 2018, Calibration of colour gradient bias in shear measurement using HST/CANDELS data, MNRAS, 476, 5645

Fried, D. L. 1966, Optical Resolution Through a Randomly Inhomogeneous Medium for Very Long and Very Short Exposures, Journal of the Optical Society of America (1917-1983), 56, 1372

Girshick, R. 2015, Fast R-CNN, arXiv e-prints, arXiv:1504.08083

Girshick, R., Donahue, J., Darrell, T., & Malik, J. 2013, Rich feature hierarchies for accurate object detection and semantic segmentation, arXiv e-prints, arXiv:1311.2524

Glorot, X., & Bengio, Y. 2010, Understanding the difficulty of training deep feedforward neural networks, Journal of Machine Learning Research - Proceedings Track, 9, 249

Goodfellow, I., Bengio, Y., & Courville, A. 2016, Deep Learning (MIT Press), `http://www.deeplearningbook.org`

Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al. 2014, in Advances in Neural Information Processing Systems 27, ed. Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, & K. Q. Weinberger (Curran Associates, Inc.), 2672–2680

Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, CANDELS: The Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey, ApJS, 197, 35

Habib, S., Pope, A., Finkel, H., et al. 2016, HACC: Simulating sky surveys on state-of-the-art supercomputing architectures, New A, 42, 49

Hartlap, J., Hilbert, S., Schneider, P., & Hildebrand t, H. 2011, A bias in cosmic shear from galaxy selection: results from ray-tracing simulations, A&A, 528, A51

Hausen, R., & Robertson, B. 2019, Morpheus: A Deep Learning Framework For Pixel-Level Analysis of Astronomical Image Data, arXiv e-prints, arXiv:1906.11248

He, K., Gkioxari, G., Dollár, P., & Girshick, R. 2017, Mask R-CNN, arXiv e-prints, arXiv:1703.06870

He, K., Zhang, X., Ren, S., & Sun, J. 2015, Deep Residual Learning for Image Recognition, arXiv e-prints, arXiv:1512.03385

Hege, E. K. 1992, in European Southern Observatory Conference and Workshop Proceedings, Vol. 39, European Southern Observatory Conference and Workshop Proceedings, 599

Heymans, C., Van Waerbeke, L., Bacon, D., et al. 2006, The Shear Testing Programme - I. Weak lensing analysis of simulated ground-based observations, MNRAS, 368, 1323

Hezaveh, Y. D., Perreault Levasseur, L., & Marshall, P. J. 2017, Fast automated analysis of strong gravitational lenses with convolutional neural networks, Nature, 548, 555

Hirata, C., & Seljak, U. 2003, Shear calibration biases in weak-lensing surveys, MNRAS, 343, 459

Holwerda, B. W., Keel, W. C., & Bolton, A. 2007, Spiral Disk Opacity from Occulting Galaxy Pairs in the Sloan Digital Sky Survey, AJ, 134, 2385

Holwerda, B. W., Keel, W. C., Williams, B., Dalcanton, J. J., & de Jong, R. S. 2008, An extended dust disk in a spiral galaxy; An occulting galaxy pair in ANGST, arXiv e-prints, arXiv:0810.2646

Hornik, K. 1991, Approximation capabilities of multilayer feedforward networks, Neural Networks, 4, 251

Huang, S., Leauthaud, A., Murata, R., et al. 2018, Characterization and photometric performance of the Hyper Suprime-Cam Software Pipeline, PASJ, 70, S6

Hubble, E. 1929, A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae, Proceedings of the National Academy of Science, 15, 168

Huff, E., & Mandelbaum, R. 2017, Metacalibration: Direct Self-Calibration of Biases in Shear Measurement, arXiv e-prints, arXiv:1702.02600

Huterer, D., Kirkby, D., Bean, R., et al. 2015, Growth of cosmic structure: Probing dark energy beyond expansion, Astroparticle Physics, 63, 23

Ioffe, S., & Szegedy, C. 2015, Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, arXiv e-prints, arXiv:1502.03167

Irwin, M. J. 1985, Automatic analysis of crowded fields., MNRAS, 214, 575

Ivezic, Z., Tyson, J. A., et al. 2008, LSST: from Science Drivers to Reference Design and Anticipated Data Products, ArXiv e-prints, arXiv:0805.2366

Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, LSST: From Science Drivers to Reference Design and Anticipated Data Products, ApJ, 873, 111

Jarvis, J. F., & Tyson, J. A. 1981, Focas: faint object classification and analysis system., AJ, 86, 476

Jenkins, A., Frenk, C. S., Pearce, F. R., et al. 1998, Evolution of Structure in Cold Dark Matter Universes, ApJ, 499, 20

Joseph, R., Courbin, F., & Starck, J. L. 2016, Multi-band morpho-Spectral Component Analysis Deblending Tool (MuSCADeT): Deblending colourful objects, A&A, 589, A2

Kaiser, N., Squires, G., & Broadhurst, T. 1995, A Method for Weak Lensing Observations, ApJ, 449, 460

Kamath, S., Meyers, J., & Burchat, P. 2019, Reducing AEGIS galaxies in two filters, doi:10.5281/zenodo.3590100

Kamath, S., Meyers, J. E., Burchat, P. R., & (LSST Dark Energy Science Collaboration. 2020, Shear Measurement Bias Due to Spatially Varying Spectral Energy Distributions in Galaxies, ApJ, 888, 23

Kennedy, R., Bamford, S. P., Häußler, B., et al. 2016, Galaxy And Mass Assembly (GAMA): understanding the wavelength dependence of galaxy structure with bulge-disc decompositions, MNRAS, 460, 3458

Kim, E. J., & Brunner, R. J. 2017, Star-galaxy classification using deep convolutional neural networks, MNRAS, 464, 4463

Kingma, D. P., & Welling, M. 2013, Auto-Encoding Variational Bayes, arXiv e-prints, arXiv:1312.6114

Korytov, D., Hearin, A., Kovacs, E., et al. 2019, CosmoDC2: A Synthetic Sky Catalog for Dark Energy Science with LSST, ApJS, 245, 26

Kratsios, A. 2019, Universal Approximation Theorems, arXiv e-prints, arXiv:1910.03344

Krist, J. 1993, in Astronomical Society of the Pacific Conference Series, Vol. 52, Astronomical Data Analysis Software and Systems II, ed. R. J. Hanisch, R. J. V. Brissenden, & J. Barnes, 536

La Barbera, F., De Carvalho, R. R., De La Rosa, I. G., et al. 2010, Spider. IV. Optical and Near-infrared Color Gradients in Early-type Galaxies: New Insight into Correlations with Galaxy Properties, AJ, 140, 1528

Lanusse, F., Ma, Q., Li, N., et al. 2018, CMU DeepLens: deep learning for automatic image-based galaxy-galaxy strong lens finding, MNRAS, 473, 3895

Laureijs, R. J., Duvet, L., Escudero Sanz, I., et al. 2010, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7731, The Euclid Mission, 77311H

Leauthaud, A., Massey, R., Kneib, J.-P., et al. 2007, Weak Gravitational Lensing with COSMOS: Galaxy Selection and Shape Measurements, ApJS, 172, 219

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. 1998, Gradient-based learning applied to document recognition, Proceedings of the IEEE, 86, 2278

Lecun, Y., Kavukcuoglu, K., & Farabet, C. 2010, 253–256

Lin, T.-Y., Dollár, P., Girshick, R., et al. 2016, Feature Pyramid Networks for Object Detection, arXiv e-prints, arXiv:1612.03144

Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. 2017, Focal Loss for Dense Object Detection, arXiv e-prints, arXiv:1708.02002

Lin, T.-Y., Maire, M., Belongie, S., et al. 2014, Microsoft COCO: Common Objects in Context, arXiv e-prints, arXiv:1405.0312

Lintott, C. J., Schawinski, K., Slosar, A., et al. 2008, Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey, MNRAS, 389, 1179

Liu, W., Anguelov, D., Erhan, D., et al. 2015, SSD: Single Shot MultiBox Detector, arXiv e-prints, arXiv:1512.02325

Lloyd-Hart, M., Wizinowich, P., McLeod, B., et al. 1992, First Results of an On-Line Adaptive Optics System with Atmospheric Wavefront Sensing by an Artificial Neural Network, ApJ, 390, L41

LSST DESC, Mandelbaum, R., Eifler, T., et al. 2018, The LSST DESC Science Requirements Document v1, ArXiv e-prints, arXiv:1809.01669v1

LSST Science Collaboration. 2009, LSST Science Book, Version 2.0, ArXiv e-prints, arXiv:0912.0201

LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, LSST Science Book, Version 2.0, arXiv e-prints, arXiv:0912.0201

LSST Science Collaboration, Marshall, P., Anguita, T., et al. 2017, Science-Driven Optimization of the LSST Observing Strategy, arXiv e-prints, arXiv:1708.04058

Luppino, G. A., & Kaiser, N. 1997, Detection of Weak Lensing by a Cluster of Galaxies at $z = 0.83$, ApJ, 475, 20

Lupton, R. 2005, SDSS Image Processing I : The Deblender

Lupton, R., Blanton, M. R., Fekete, G., et al. 2004, Preparing Red-Green-Blue Images from CCD Data, PASP, 116, 133

Lupton, R., Gunn, J. E., Ivezić, Z., Knapp, G. R., & Kent, S. 2001, Astronomical Society of the Pacific Conference Series, Vol. 238, The SDSS Imaging Pipelines, ed. J. Harnden, F. R., F. A. Primini, & H. E. Payne, 269

Mandelbaum, R. 2018, Weak Lensing for Precision Cosmology, ARA&A, 56, 393

Mandelbaum, R., Rowe, B., Bosch, J., et al. 2014, The Third Gravitational Lensing Accuracy Testing (GREAT3) Challenge Handbook, ApJS, 212, 5

Mandelbaum, R., Lanusse, F., Leauthaud, A., et al. 2018, Weak lensing shear calibration with simulations of the HSC survey, MNRAS, 481, 3170

Massey, R., Hoekstra, H., Kitching, T., et al. 2013, Origins of weak lensing systematics, and requirements on future instrumentation (or knowledge of instrumentation), MNRAS, 429, 661

Melchior, P., Moolekamp, F., Jerdee, M., et al. 2018, SCARLET: Source separation in multi-band images by Constrained Matrix Factorization, Astronomy and Computing, 24, 129

Melchior, P., Viola, M., Schäfer, B. M., & Bartelmann, M. 2011, Weak gravitational lensing with DEIMOS, MNRAS, 412, 1552

Meyers, J. E., & Burchat, P. R. 2015, Impact of Atmospheric Chromatic Effects on Weak Lensing Measurements, ApJ, 807, 182

Mighell, K. J. 1989, Accurate stellar photometry in crowded fields., MNRAS, 238, 807

Miller, A. S. 1993, A review of neural network applications in Astronomy, Vistas in Astronomy, 36, 141

Miller, L., Heymans, C., Kitching, T. D., et al. 2013, Bayesian galaxy shape measurement for weak lensing surveys - III. Application to the Canada-France-Hawaii Telescope Lensing Survey, MNRAS, 429, 2858

Murtagh, F. D., & Adorf, H. M. 1992, in Data Analysis in Astronomy, 103–111

Nakajima, R., & Bernstein, G. 2007, Shear Recovery Accuracy in Weak-Lensing Analysis with the Elliptical Gauss-Laguerre Method, AJ, 133, 1763

Newman, J. A., Cooper, M. C., Davis, M., et al. 2013, The DEEP2 Galaxy Redshift Survey: Design, Observations, Data Reduction, and Redshifts, ApJS, 208, 5

Odewahn, S. C., Stockwell, E. B., Pennington, R. L., Humphreys, R. M., & Zumach, W. A. 1992, Automated Star/Galaxy Discrimination With Neural Networks, AJ, 103, 318

Pasquet, J., Bertin, E., Treyer, M., Arnouts, S., & Fouchez, D. 2019, Photometric redshifts from SDSS images using a convolutional neural network, A&A, 621, A26

Pearson, K. A., Palafox, L., & Griffith, C. A. 2018, Searching for exoplanets using artificial intelligence, MNRAS, 474, 478

Penzias, A. A., & Wilson, R. W. 1965, A Measurement of Excess Antenna Temperature at 4080 Mc/s., ApJ, 142, 419

Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, Measurements of Ω and Λ from 42 High-Redshift Supernovae, ApJ, 517, 565

Petrillo, C. E., Tortora, C., Chatterjee, S., et al. 2017, Finding strong gravitational lenses in the Kilo Degree Survey with Convolutional Neural Networks, MNRAS, 472, 1129

Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2018, Planck 2018 results. VI. Cosmological parameters, arXiv e-prints, arXiv:1807.06209

Ravanbakhsh, S., Lanusse, F., Mandelbaum, R., Schneider, J., & Poczos, B. 2016, Enabling Dark Energy Science with Deep Generative Models of Galaxy Images, arXiv e-prints, arXiv:1609.05796

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. 2015, You Only Look Once: Unified, Real-Time Object Detection, arXiv e-prints, arXiv:1506.02640

Refregier, A., Kacprzak, T., Amara, A., Bridle, S., & Rowe, B. 2012, Noise bias in weak lensing shape measurements, MNRAS, 425, 1951

Reiman, D. M., & Göhre, B. E. 2019, Deblending galaxy superpositions with branched generative adversarial networks, MNRAS, 485, 2617

Ren, S., He, K., Girshick, R., & Sun, J. 2015, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, arXiv e-prints, arXiv:1506.01497

Rhodes, J. D., Massey, R. J., Albert, J., et al. 2007, The Stability of the Point-Spread Function of the Advanced Camera for Surveys on the Hubble Space Telescope and Implications for Weak Gravitational Lensing, ApJS, 172, 203

Riess, A. G., Filippenko, A. V., Li, W., & Schmidt, B. P. 1999, Is there an Indication of Evolution of Type IA Supernovae from their Rise Times?, AJ, 118, 2668

Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant, AJ, 116, 1009

Rowe, B. T. P., Jarvis, M., Mandelbaum, R., et al. 2015, GALSIM: The modular galaxy image simulation toolkit, Astronomy and Computing, 10, 121

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. 1986, Learning representations by back-propagating errors, Nature, 323, 533

Ryden, B. 2003, Introduction to cosmology

Samuroff, S., Bridle, S. L., Zuntz, J., et al. 2018, Dark Energy Survey Year 1 results: the impact of galaxy neighbours on weak lensing cosmology with IM3SHAPE, MNRAS, 475, 4524

Sánchez, F. J., Walter, C. W., Awan, H., et al. 2020, The LSST DESC Data Challenge 1: Generation and Analysis of Synthetic Images for Next Generation Surveys, arXiv e-prints, arXiv:2001.00941

Santurkar, S., Tsipras, D., Ilyas, A., & Madry, A. 2018, How Does Batch Normalization Help Optimization?, arXiv e-prints, arXiv:1805.11604

Schawinski, K., Zhang, C., Zhang, H., Fowler, L., & Santhanam, G. K. 2017, Generative adversarial networks recover features in astrophysical images of galaxies beyond the deconvolution limit, MNRAS, 467, L110

Scheuer, P. A. G. 1974, Fluctuations in the X-ray background, MNRAS, 166, 329

Schmidhuber, J. 2014, Deep Learning in Neural Networks: An Overview, arXiv e-prints, arXiv:1404.7828

Schneider, P. 2005, Weak Gravitational Lensing, arXiv e-prints, astro

Seitz, C., & Schneider, P. 1997, Steps towards nonlinear cluster inversion through gravitational distortions. III. Including a redshift distribution of the sources., A&A, 318, 687

Semboloni, E., Hoekstra, H., Huang, Z., et al. 2013, On the shear estimation bias induced by the spatial variation of colour across galaxy profiles, MNRAS, 432, 2385

Serra-Ricart, M., Gaitan, V., Garrido, L., & Perez-Fournon, I. 1996, Faint object classification using Artificial Neural Networks., A&AS, 115, 195

Sérsic, J. L. 1963, Influence of the atmospheric and instrumental dispersion on the brightness distribution in a galaxy, Boletin de la Asociacion Argentina de Astronomia La Plata Argentina, 6, 41

Sheldon, E. S., Becker, M. R., MacCrann, N., & Jarvis, M. 2019, Metadetection: Mitigating Shear-dependent Object Detection Biases with Metacalibration, arXiv e-prints, arXiv:1911.02505

Silberman, N., Sontag, D., & Fergus, R. 2014, in Computer Vision – ECCV 2014, ed. D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Cham: Springer International Publishing), 616–631

Simard, P. Y., Steinkraus, D., & Platt, J. C. 2003, in Proceedings of the Seventh International Conference on Document Analysis and Recognition - Volume 2, ICDAR '03 (Washington, DC, USA: IEEE Computer Society), 958–

Springel, V., White, S. D. M., Jenkins, A., et al. 2005, Simulations of the formation, evolution and clustering of galaxies and quasars, Nature, 435, 629

Stetson, P. B. 1987, DAOPHOT: A Computer Program for Crowded-Field Stellar Photometry, PASP, 99, 191

Suchyta, E., Huff, E. M., Aleksić, J., et al. 2016, No galaxy left behind: accurate measurements with the faintest objects in the Dark Energy Survey, MNRAS, 457, 786

Tagliaferri, R., Longo, G., Milano, L., et al. 2003, Neural neZtworks in astronomy, Neural Networks, 16, 297 , neural Network Analysis of Complex Scientific Data: Astronomy and Geosciences

The LSST Dark Energy Science Collaboration. 2019, LSST DESC Science Roadmap, Immediately after the executive summary, there is a brief ”How to use this document” guide that concisely describes how to navigate it. Please start there to find your way through the document!, doi:10.5281/zenodo.3588457

The LSST Dark Energy Science Collaboration, Mandelbaum, R., Eifler, T., et al. 2018, The LSST Dark Energy Science Collaboration (DESC) Science Requirements Document, arXiv e-prints, arXiv:1809.01669

Tyson, J. A. 1995, in Extragalactic Background Radiation Meeting, 103–133

Vernardos, G., & Tsagkatakis, G. 2019, Quasar microlensing light-curve analysis using deep machine learning, MNRAS, 486, 1944

Voigt, L. M., Bridle, S. L., Amara, A., et al. 2012, The impact of galaxy colour gradients on cosmic shear measurement, MNRAS, 421, 1385

Weinberg, D. H., Mortonson, M. J., Eisenstein, D. J., et al. 2013, Observational probes of cosmic acceleration, Phys. Rep., 530, 87