applied sciences

MDPI

Article

Deep Learning Based Real Age and Gender Estimation from Unconstrained Face Image towards Smart Store Customer Relationship Management

Md. Mahbubul Islam and Joong-Hwan Baek *

Department of Electronics and Information Engineering, Korea Aerospace University, Goyang 10540, Korea; mahbubcse@cu.ac.bd

* Correspondence: jhbaek@kau.ac.kr

Featured Application: A smart store customer relationship management system to estimate the customer's age and gender for simplifying the shopping experience by facilitating personalized product recommendation and advertisement to promote the smart trading along with developing an inventory for future business promotion.

Abstract: The COVID-19 pandemic markedly changed the human shopping nature, necessitating a contactless shopping system to curb the spread of the contagious disease efficiently. Consequently, a customer opts for a store where it is possible to avoid physical contacts and shorten the shopping process with extended services such as personalized product recommendations. Automatic age and gender estimation of a customer in a smart store strongly benefit the consumer by providing personalized advertisement and product recommendation; similarly, it aids the smart store proprietor to promote sales and develop an inventory perpetually for the future retail. In our paper, we propose a deep learning-founded enterprise solution for smart store customer relationship management (CRM), which allows us to predict the age and gender from a customer's face image taken in an unconstrained environment to facilitate the smart store's extended services, as it is expected for a modern venture. For the age estimation problem, we mitigate the data sparsity problem of the large public IMDB-WIKI dataset by image enhancement from another dataset and perform data augmentation as required. We handle our classification tasks utilizing an empirically leading pre-trained convolutional neural network (CNN), the VGG-16 network, and incorporate batch normalization. Especially, the age estimation task is posed as a deep classification problem followed by a multinomial logistic regression first-moment refinement. We validate our system for two standard benchmarks, one for each task, and demonstrate state-of-the-art performance for both real age and gender estimation.

Keywords: deep learning; age estimation; gender estimation; smart store; COVID-19

Citation: Islam, M.M.; Baek, J.-H. Deep Learning Based Real Age and Gender Estimation from Unconstrained Face Image towards Smart Store Customer Relationship Management. *Appl. Sci.* **2021**, *11*, 4549. https://doi.org/10.3390/app11104549

Received: 1 April 2021
Accepted: 11 May 2021
Published: 17 May 2021

1. Introduction

Amid the COVID-19 pandemic situation, a customer prefers a store where it is possible to avoid contact with staff and stay for a short period of time while shopping. A smart store is a trading store equipped with smart technologies where a customer is able to do shopping from kiosks without the assistance of staff and not being checked out by a cashier. An automated store with recent technologies grants retailers to know more about the customers, product preferences, and their shopping behavior. A smart store can use artificial-intelligence based customer management systems to extract customer information in real-time and can provide the best product recommendations by analyzing the customer information for steering additional trades in real-time. A smart store can help the purchaser's preferences by knowing their age and gender. Deep learning-based smart store management systems can arrange their store by placing items alongside to promote cross-selling based on customers demographic choices.

Owing to the uprising trends of intelligent systems, there is an increased demand for automatic human demographic extraction from face images. The estimation of human demographics such as age and gender from face images is a very promising and challenging task in academia and industry. Application where age and gender estimation can play a useful part include (i) access control [1], e.g., curbing the entry of an underaged person to sensible items from vending machines or to an event where only people of specific gender can join; (ii) human–computer interaction (HCI) [2,3], e.g., providing a different product advertisement or offer by looking at the gender and age of a person automatically; (iii) law enforcement [4], e.g., a criminal demographic estimation can help the law enforcement agency to find out the suspects more proficiently from previous records; (iv) surveillance [5], e.g., an automated system recognizing unattended minors to some unexpected places and times; (v) electronic customer relationship management [6], e.g., companies may use internet-based platforms to interact with customers to perceive their preferences and customize their store products. Accordingly, many traditional retail systems migrate to the intelligent system with the recent development of technology such as smart store retail management. To ease the shopping process and the future retail of smart stores, the age and gender estimation of a customer is indispensable. Similarly, for welcoming salutations, the content and manner are quite different based on age and gender.

In order to achieve these tasks, major challenges arise due to the unconstrained real-world face images which are captured in a different angle, pose, and background. A significant amount of research has been conducted to estimate the age from a face image having the form of real or biological age estimation. This age and gender estimation research spans decades, as summarized in large studies [7–11]. Each of these face analysis tasks (age and gender estimation) are sought to solve distinct research problems through a variety of techniques [12–17]. The facial attribute information such as age and gender are already being predicted using facial landmark information [15–21]. Spotting accurate facial landmarks is in itself a challenging issue. Localization of the facial points is heavily complicated in some imaging conditions, e.g., when the face is occluded by something, rotated extremely, complex facial expressions, and the resolution of the image is very low. Analogously, landmark extraction from faces is practically impossible when the imaging environment is in the far-field.

It is worth bearing in mind that the real-world smart applications for age and gender estimation need to tackle faces having certain unconstrained environments, like improperly aligned or having unusual pose and expressions. Therefore, under these circumstances, prior to input a face to the age and gender estimation system, a face should be detected first and, in a next step, properly aligned. Despite the recent progress made in [7,22,23] in the context of handling faces in the wild, the accurate prediction of age and gender remains a challenging problem due to the limited and constrained image datasets. Hence, a shallow network was proposed to classify the age and gender in [23]. In [24], authors utilized the benefit of a manageable deep neural network to train with a large and diverse face image dataset. A robust face detection and alignment operation was performed over the in-the-wild face images that play a noticeable role in the overall performance. Although this work achieves very good performance for the real age estimation task, this network is biased to the classes that belong to adult people because of the large inter-class sample variation problem that exists in the training dataset.

In this paper, we approach an integrated framework for human age and gender estimation being motivated by the contemporary deep learning-based advancement in the associated research on age and gender estimation [23,24]. We demonstrate better results for age estimation by a substantial margin compared to the state-of-the-art (SOA) approach [24] with necessary improvement in the training data and imposing more regularization techniques (i.e., batch normalization, data augmentation) in the network. Subsequently, we achieve better results in anticipation of gender classification compared to the SOA method [23]. The contribution of this work is summarized as follows:

- We propose an automatic age and gender estimation system for the customers of a smart store (e.g., Amazon Go, SmartMart) to affluence the offline smart shopping and to update the future stock by analyzing the customer demographics (i.e., age and gender) due to the new shopping nature amid the Covid-19 pandemic situation.
- We handle the data sparsity problem (see Section 3.1.1) that exists in the publicly available in-the-wild face image dataset IMDB-WIKI [24].
- We consider both age and gender estimation tasks as a classification problem and deploy the ImageNet pre-trained model VGG-16 [25], although real age estimation is basically a regression problem. We address the age estimation problem like [24], with an effective change in the dataset by making it almost balanced and introduce a batch normalization layer to speed up the learning process along with stable performance because a smart store needs a robust system.
- For the comparable results, we evaluate our model on the constrained and specific aged people image dataset Morph [1]. Subsequently, the challenging Adience [9] image dataset is used to evaluate the gender estimation performance. Our approach marginally outperforms the state-of-the-art methods in both age and gender estimation tasks.

The anatomy of the rest of the paper is as follows: Section 2 elucidates the literature regarding age and gender estimation. Section 3 presents our proposed method. The insight of the experiments and attained results are presented in Section 4. Section 5 carries out the comparative discussion regarding age and gender estimation results, and Section 6 concludes this research.

2. Related Work

This section succinctly reviews the associated works of age and gender estimation. Although a significant amount of literature is already available related to these topics, we try to provide a superficial outline in what way these tasks were approached earlier by other researchers.

2.1. Real Age Estimation

Age estimation is a long-studied research topic among computer vision researchers. Most of the researchers considered human age estimation as either a classification or regression problem. In the case of age classification, age is coupled with a specific range or age group. On the other hand, age regression is a single value estimated for a person. However, it is very challenging to estimate an exact age due to diversity in the aging process across different ages [26]. Furthermore, for accurate age estimation, the model needs a huge amount of correctly labeled face data.

Many of the early age estimation methods used hand-crafted facial features in the constrained imaging conditions. A survey of such methods was reported in [6] and a recent survey of age estimation including all approaches of the last decades can be found in [27]. A method that extracts the geometric features from the face and calculates ratios among the facial features to estimate the age is presented in [28]. Initially, the face wrinkles are detected and localized, then the size and distances of the facial features are measured and finally the face is classified into different age categories. A similar approach as presented in [28] is proposed in [29] by modeling growth-related shape variations observed in human faces considering anthropometric evidence. This work was limited to a certain age. The abovementioned methods are unsuitable for images in-the-wild due to the necessity of accurate localization of facial features.

A couple of subspace methods were introduced in [30,31], where aging features were extracted from an aging pattern representative subspace and a robust regressor was used to predict the face ages. Although the aforementioned methods achieved excellent performances compared to previous cutting-edge methods, some limitations are exposed by these algorithms. Their system worked well only with frontal and properly aligned images. The algorithms proposed by these researchers are not well suited for practical applications where the input images might be collected in an unconstrained environment.

There are some methods where face images are represented using spatially localized facial patches. In [32,33], the patch distribution was represented by exploiting probabilistic Gaussian mixture models [34]. A robust descriptor was used in place of pixel patches in [32]. Later on, the Hidden-Markov-Model [35] was introduced instead of the Gaussian mixture model (GMM) for representation of face patch distribution [36].

A significant amount of research employed different robust image descriptors as an alternative to the local image intensity patch for the age estimation task. Gao et al. [37] used Gabor filters along with a Fuzzy-LDA classifier where one face belongs to multiple age classes. Similarly, Biologically-Inspired-Features in [38], and Local Binary Patterns (LBP) were presented in [39].

The age estimation problem was considered as a regression problem by Fu et al. [40] or as a classification problem using a quadratic function, shortest distance, and neural network-based classifiers in [41]. The popular regression techniques reported by the researchers are Support Vector Regression (SVR) [42], Partial Least Squares (PLS) [43], Canonical Correlation Analysis (CCA) [44]. Accordingly, Nearest Neighbor (NN) and Support Vector Machines (SVM) [45] are the most used classification techniques.

Next, we choose a couple of real age estimation methods to describe those that are most related to our suggested method. Guo et al. [30] presented a learning scheme to draw aging features named manifold learning and utilized SVRs with local adjustment for age prediction, Han et al. [46] extracted features using boosting algorithms and formed a hierarchical approach for classification between age group and regression inside a group (DIF). Geng et al. [31] introduced an aging pattern subspace (AGES) wherefrom features were extracted and performed regression for age estimation. Zhang and Yeung [47] handled age estimation as a multi-task problem based on the warped Gaussian process (MTWGP) where common features were shared among the tasks. Chen et al. [48] introduced a mapping between the cumulative attribute space and low-level sparse features for age regression. Chang et al. [49] formed the age labels into binary groups that formed subproblems and imposed a cost on each subproblem. Thus, they ranked the ordinal hyperplanes based on classification cost for age estimation, while Guo and Mu [50] used a canonical correlation analysis and partial least squares founded methods to perform feature projection and estimate human traits jointly.

Recently, the biologically inspired CNN models were successfully deployed for the age estimation task. Yi et al. [51] deployed a multiscale CNN. Wang et al. [52] used features from the intermediate layer of CNN rather than top layer features and performed manifold learning. Rothe et al. [22] incorporated a deep CNN for extracting features and real age regression as an estimate using SVR. In [24], they used a deeply learned CNN model from large in-the-wild image data and performed age regression through classification. A hybrid system was introduced in [53], where the CNNs were used for face feature extraction and an extreme learning machine (ELM) for the classification task. A lightweight CNN network with mixed attention mechanism for low end devices was proposed in [54], where the output layer was fused by classification and regression approach. Another multi-task learning approach merging classification and regression concepts to fit the age regression model with heterogeneous data with the help of two different techniques for partitioning data towards classification was proposed in [55]. To resolve the problem of data disparity and ensure the generality of the model, a very recent method is proposed by Kim et al. [56] where a cycle generative adversarial network-based race and age image transformation method is used to generate sufficient data for each distribution. All of the aforementioned CNN-based systems are evaluated on the basis of the common dataset Morph [1] for age estimation. To the best of our knowledge, [24,53] demonstrate state-of-the-art results. A comparison table of different age estimation methods mostly related with our experiments are summarized in Table 1.

Table 1. Comparison between top performing age estimation algorithms.

Research	Model	Strengths	Weakness
[24]	Deep Expectation, Vgg16 and Classification	- Perform age regression through classification without facial landmarks - Achieve state of the art performance on real age and winner of LAP challenge 2015 on apparent age estimation	- Do not maintain class-wise data balancing during training - Huge noisy data in the proposed dataset
[31]	Aging pattern subspace and classification	- Reconstruct unseen face image and a complete aging face database is not necessary - Performance is comparable to the human observer	- Highly relies on landmark Detecting algorithms - Limited performance on early childhood faces
[48]	Cumulative attribute space and Support vector regression	- Cumulative attributes capture the shape and texture features from a human face compared to aging processes-Significantly performed even if the training data is sparse and imbalanced	- Sensitive in the sense of feature inconsistency and impreciseness
[49]	Ranking CNN	- Estimate age based on the age distinctive feature ordering information, ranking the age labels and thus avoid the binary decision for each label	- Slower compared to cumulative attribute space model and less accurate with sparse data
[50]	CCA & PLS for dimensionality reduction and Regression	- Able to estimate three traits jointly from three dimensions - Implicitly benefited from the gender and race features for age estimation	- The methods used for dimension reduction is very sensitive to the identity of the features between sets.
[54]	Attention based ShuffleNet and Hybrid model	- Perform classification and regression together and used a light-weight network for low end devices	- Not compact enough to deploy in the embedded devices
[56]	Cycle GAN, CNN	- Resolve data imbalance and network over-fitting by generating sufficient images using CGAN	- This image generation seems inefficient for the noisy image dataset
[57]	Group-n encoding, series of binary subproblems, and Decoding	- Utilize the adjacent ages information as shared class features and decompose as binary subproblems simplifies the age estimation processing - Cost-sensitive learning helps to reduce the data imbalance	- Intra and inter group formation is quite random - Encoding and decoding does not have too much impact on the performance compared to computing cost
[58]	Ordinal Regression and CNN	- Transformed to binary subproblems and multiple output CNN maintain the correlations between tasks - Proposed an Asian Face Age Dataset with more than 160 k images	- Low robustness - Cannot handle multicollinearity between independent features
Ours	Deep first-moment refinement, MTCNN, vgg16	- Resolve the data sparsity problem exist in the DEX method and specify the target application where to deploy our model	- Not trained with enough balance data due to very less images for the child's and elders

2.2. Gender Estimation

A lot of progress has been achieved in the gender estimation topic, but it is still a challenging problem in the real-world environment. The literature about gender estimation comes under the umbrella of the authors of [59,60]. Here, we will discuss some of those methods where the well-known classifiers are used for the gender estimation. As one of the very early methods, the authors of [61] used a fully connected two-layer neural network that learned from a limited number of near-frontal face images for gender classification. In [62], SVM classifiers were directly applied to image intensities. Similarly, AdaBoost was introduced instead of SVM classifier by keeping the same working pipeline [63]. Later on, a viewpoint-invariant model for age and gender estimation was suggested by Toews and Arbel [64] which is robust to local scale rotations.

A combination of human knowledge and a gait information-based gender classification system was provided by Yu et al. [65]. An unconstrained face image benchmark for gender classification along with a high classification accuracy was presented in [9]. Khan et al. [66] formed a semantic pyramid by extracting features from the full and upper body together with face regions from the image to recognize the gender and action. This method does not depend on the annotation value of a person's face and upper body to extract semantic features. In [67], the authors proposed a model where the first name of a person is used as a special feature and associates a name with facial appearance to recognize the gender of that person. At the same time, the authors showed that their method achieved higher accuracy in the task of gender recognition and demonstrated the potential in the use of face verification. In recent times, a generic framework for age and gender estimation was proposed in [46], where a hierarchical estimator was modeled based on extracted biologically inspired features. Besides, this method was formed to detect low-quality images due to a poor image background.

The above efforts regarding gender estimation contributed a lot in this research area. However, the lion's share of these methods is only suitable for the applications with constraint images or have higher computational costs. Recently, a deep CNN-based approach was presented in [68]. It was pretrained with a huge unconstrained dataset and then fine-tuned on two other datasets to achieve a very good accuracy in the gender estimation task. Although, their method states a high accuracy, a lot of pretraining is required prior to evaluating the system. In our paper, we propose a system that will work comparatively well with unconstrained imaging conditions.

3. Proposed Method

Our proposed method is utilized during the course of our experiments for both age and gender classification. Our approach is inspired by the research advancement in the computer vision fields, such as image classification [48,69,70], object detection [71], age estimation [23,24], and gender classification [23] fueled by deep learning. At the very beginning of our age and gender estimation process, we ensure a class-wise balance of the image samples for training. To do this task, we down-sample those classes up to a specified threshold where the number of images is huge, then up-sample other classes by importing images from another dataset built on the same setup and perform data augmentation where necessary. Additionally, we manually filter out those images that seem wrongly annotated using human visual perception. In the later stages, before feature extraction, we detect the face from raw faces and prepare the images for network input performing some preprocessing tasks like rescaling and resizing. We use the same CNN structure for the feature extraction and the Softmax layer for classification output regardless of the estimation task, but a further formulation is performed in case of age estimation. We calculate the expected value over the Softmax probabilities for age regression. Each step of the proposed approach shown in Figure 1 is depicted thoroughly in this section.

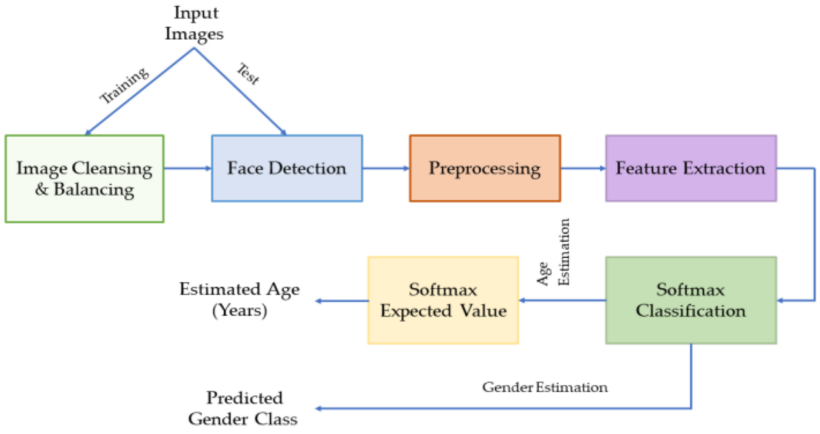

Figure 1. Schematic diagram of the proposed age and gender estimation method.

In Figure 2, we present the process diagram of the smart store customer relationship management system based on our approach. When a customer approaches the smart shelf, the camera attached with the shelf automatically captures the image of the customer and the installed age and gender estimation model will predict the demographics of the respective customer. During the interaction of the customer and automated system, the customer will get the personalized product recommendation, advertisement, and offer. In a smart store enterprise solution, several outlets are operating simultaneously under the same setup. Therefore, the customer data will be stored in a local server that exists on the outlet and to the central server from the different outlets through a communication network.

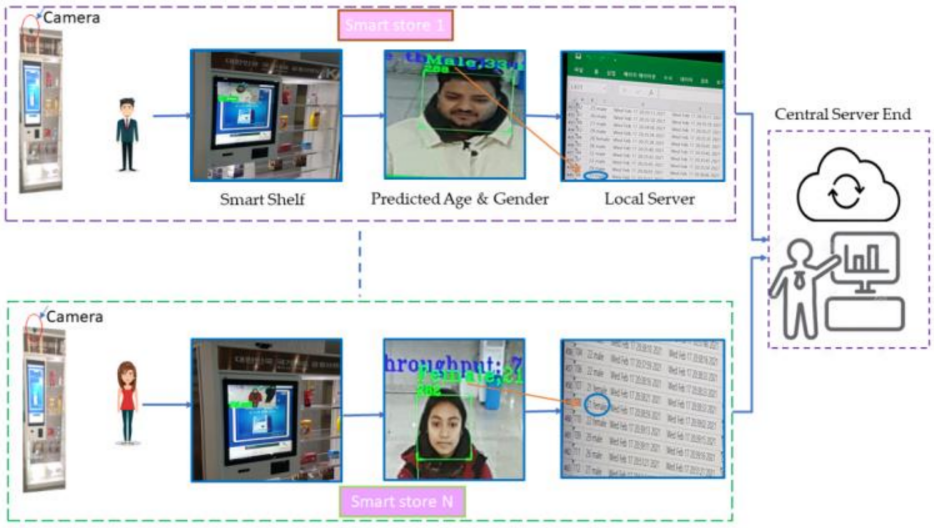

Figure 2. The process diagram of a smart store customer relationship management system.

3.1. Age Estimation

3.1.1. Data Cleansing & Balancing

Real age and gender estimation from a face image is a very complex problem acknowledged by computer vision researchers. Especially, age estimation from the face image is still an open complex problem within the research community. It is a well-known fact that a very deep neural network is needed to solve a complex problem. Accordingly, training a very deep network requires a huge amount of training images. Otherwise, a network will overfit if we fail to provide an optimum amount of training images to that network. To overcome the overfitting problem of the network, the first and foremost step is the data adequacy for training the deployed network. In reality, there are only a few datasets that have a very rich number of face images under an unconstrained environment. In

our experiment, we have used the richest in-the-wild image dataset IMDB-WIKI [24] for training the deep network. Despite the IMDB-WIKI dataset is rich concerning the number of images, data sparsity is huge in this dataset. In the case of the age estimation task, we observed that this dataset consists of a huge number of images for people aged between 15–65 compared to children and elderly people. In this situation, if we train the model with this class sparsity, it is impossible to get optimized results for the class which is imbalanced in real-time as the model never gets a sufficient look at the underlying class. We mitigate the data sparsity problem by taking the following scheme:

- Randomly choose the number of samples from the class which has sufficient observations so that the comparative ratio among the class will be retained;
- Manually filter out the wrongly annotated samples from each class that is shown in Figure 3
- For the classes of fewer samples, we first enhance the data from another benchmark dataset, the Adience dataset [9], and perform necessary offline data augmentation operations to make the class balance. The performed data augment operations, such as right flipping, rotation in the angle between -30° to 30° with the steps of 5°, scaling, and adding noise to a certain probability.

Figure 3. Some of the wrongly annotated image samples in the existing IMDB-WIKI benchmark dataset.

During training, we perform online data augmentation by rescaling the input image into 256×256 pixels and taking a center crop of 224×224 pixels from the 256×256 size image and pass it on for the training. This can alleviate the over-fitting problem of the network and enhances the robustness of the model. Through empirical observation, it is evident that after these operations, the training effect of the network is better, and the age and gender estimation accuracy of the final model is higher. In Figure 4, we show the resulting data distribution among the underlying classes before and after we introduce parity in the number of image samples of the combined IMDB-WIKI dataset.

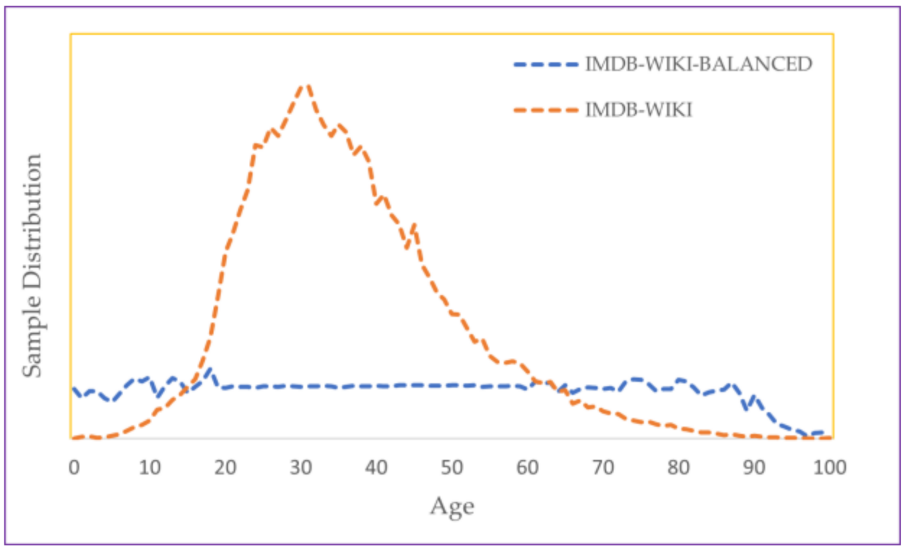

Figure 4. Class-wise sample distribution of the IMDB-WIKI dataset prior to and following parity among classes.

3.1.2. Regression through Classification

Age estimation is basically a regression problem, as age is a continuous value rather than a set of discrete classes. The deployed pre-trained model VGG-16 architecture is applied for the ImageNet classification task where the output layer consists of 1000 neurons normalized using the Softmax function, one for each of the object classes. In practice, we replace the last layer with only one output neuron and employ a Euclidean loss function for the regression task. It is unfortunate that, if we train a CNN solely for any regression task the model experienced a high error because of the instability when handling outliers. As a result, the network is facing the difficulty of poor convergence due to high gradients and predictions become unstable.

Under these circumstances, we handle the age estimation task through a classification approach by discretizing the ages into K categories. Following this procedure, we learn our CNN model for age classification and quantify the regression value from the expected value formulated using the Softmax-probabilities that belong to the K neurons, as shown in Equation (1).

$$E = \sum_{i=0}^{|k-1|} y_i.p_i \tag{1}$$

where k stands for 101 age categories, $y_i \in [0, 100]$, and $0 \leq i \leq 100$. p_i denotes the Softmax-normalized output probability of neuron i. The experimental results show that this formula increases the robustness during training and the prediction accuracy during testing.

3.2. Face Detection & Alignment

Face detection from the human face is naturally a very challenging task due to a lot of variation in appearances and external factors. Face detection is a necessary first step in the age and gender prediction system where discriminative facial features make the decision. The number of datasets used for our age and gender estimation task comprises in-the-wild face images. In the very beginning, we need to detect the face from the raw input image and then align the face for the training part as well as testing. An ideal input image should be of approximately identical size, centered, rotated to a normalized position, and with a minimum background. We opt for the robust Deformable Parts Model (DPM) [72] based face detection algorithm [73] to find the location and size of the face on the IMDB-WIKI [24] images. Similarly, a deep cascaded multi-task framework [74] is adopted for face detection on Adience images which exploits the inherent correlation between detection and alignment to boost up their performance. The face detection procedure using a multi-task cascaded convolutional neural network (MTCNN) is presented in Figure 5. Deep CNNs are powerful enough to handle small alignment errors and that is why we focused on a robust face detector with a marginal up-frontal rotation for alignment, as proposed in [24]. Consequently, the age and gender estimation tasks show improved performance with the detected face rather than the entire image.

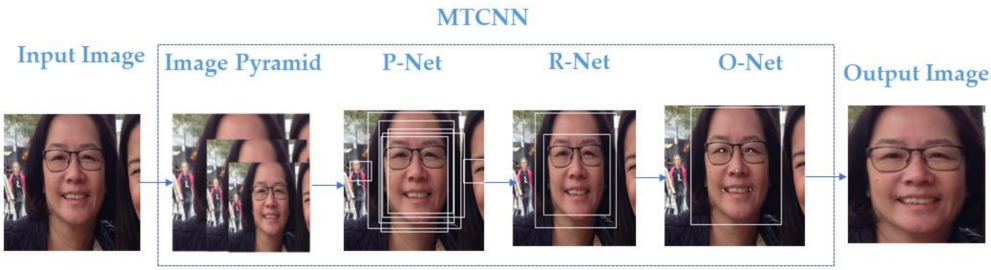

Figure 5. Face detection process over raw face images using MTCNN.

Our chosen face detectors are able to detect a face perfectly in most cases, although unsuccessful in some face images. The failure case is handled by providing the entire image as the face. If we consider some extra context around the face, it also helps improve

the classification performance. Therefore, the detected face is extended by adding a 40% margin on all sides. To ensure the same position of the detected face in the image the border pixels are simply repeated when there is no context on some sides of the too-large faces. The image is squeezed to 256 × 256 pixels to maintain the aspect ratio of the resulting images. Finally, the data augmentation operation described in [24] is performed to prepare the CNN input image of 224 × 224 pixels.

3.3. Scratch Model

We employ a convolutional neural network to predict the age and gender of a human exclusively from a single face image. This network uses an aligned face with the background as input and outputs a real predicted age or corresponding gender class. In our system, we use a popular pre-trained CNN architecture, named VGG-16 [25]. The intuitions behind choosing this architecture are (1) very intense but tractable network, (2) top performer of the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [70], (3) maintains very good performance along with prediction time compared to other pre-trained networks, and (4) for the classification task, the publicly available pre-trained models concede very good starts for training.

The VGG-16 network is composed of 16 layers where 13 layers are convolution layers, and the remaining 3 layers are dense. This network is reasonably deeper than the previous popular network AlexNet [69]. This network differs from AlexNet due to the use of fixed size convolution filter 3 × 3 and 2 × 2 size max-pool kernels with a stride of 2 instead of the much larger filter size of 11 × 11, 5 × 5 along with a stride of 4. Hence, multiple stacked 3 × 3 size kernel enables the network to learn more complex features at a lower cost than one large size convolution filter, although the network depth increases. In our approach, we incorporate batch normalization before the rectified linear unit activation function to reduce network overfitting, generalization error and expedite network convergence. We perform our experiments for age and gender estimation with the convolutional neural network model proposed in [25]. It is worth mentioning that the pre-trained CNN model is fine-tuned with publicly available face image benchmark dataset to adapt with face image related to age and gender estimation task. Lastly, our network is further tuned with the actual dataset on which we evaluate our model. The fine-tuning permits the CNN model to extract the detail features, the variations, and the bias from every dataset that helps to boost the performance. The underlying CNN architecture for the age and gender estimation tasks is shown in Figure 6. The summary of the deployed CNN architecture is presented in Table 2.

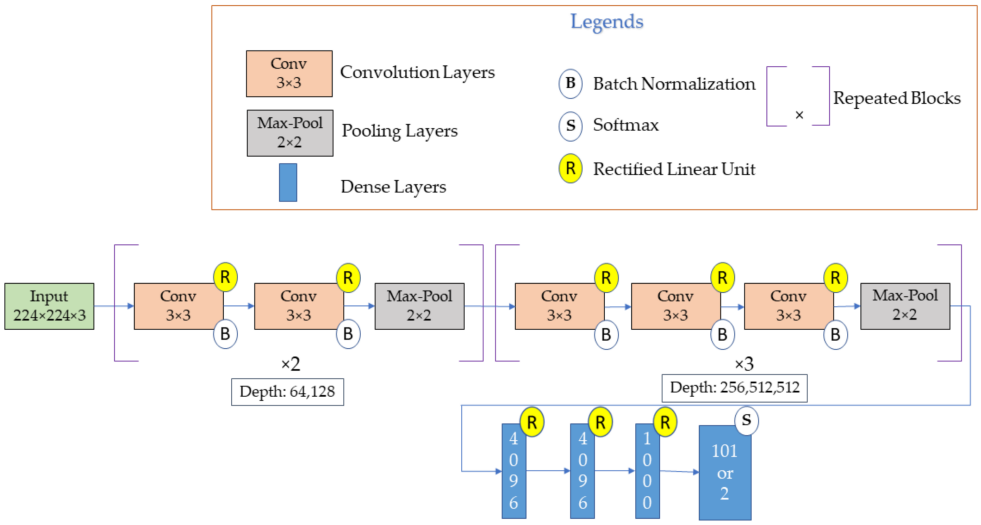

Figure 6. The deployed CNN architecture for the age and gender estimation.

Table 2. The details of our age and gender estimation network.

Layers	Feature Map Size	Filter Size	Stride	Parameters
Input	224 × 224 × 3			
CONV1-1	224× 224×64	3 × 3	1 × 1	1792
*BR				128
CONV1-2		3 × 3	1 × 1	36,928
*BR				128
MaxPool	112 × 112 × 64	2 × 2	2 × 2	
CONV2-1	112 × 112 × 128	3 × 3	1 × 1	73,856
*BR				256
CONV2-2		3 × 3	1 × 1	147,584
*BR				256
MaxPool	56 × 56 × 128	2 × 2	2 × 2	
CONV3-1	56 × 56 × 256	3 × 3	1 × 1	295,168
*BR				512
CONV3-2		3 × 3	1 × 1	590,080
*BR				512
CONV3-3		3 × 3	1 × 1	590,080
*BR				512
MaxPool	28 × 28 × 256	2 × 2	2 × 2	
CONV4-1	28 × 28 × 512	3 × 3	1 × 1	1,180,160
*BR				1024
CONV4-2		3 × 3	1 × 1	2,359,808
*BR				1024
CONV4-3		3 × 3	1 × 1	2,359,808
*BR				1024
MaxPool	14 × 14 × 512	2 × 2	2 × 2	
CONV5-1	14 × 14 × 512	3 × 3	1 × 1	2,359,808
*BR				1024
CONV5-2		3 × 3	1×1	2,359,808
*BR				1024
CONV5-3		3 × 3	1 × 1	2,359,808
*BR				1024
MaxPool	7 × 7 × 512	2 × 2	2 × 2	
FC1	4096			102,764,544
Dropout				
FC2				16,781,312
Dropout				
FC3	1000			4,097,000
Dropout				
FC4	101 or 2			101,101
		Total Trainable Parameters		138,467,093

* BR indicates Batch Normalization Layer and Rectified Linear Unit layer together.

3.4. Performance Metric

For the quantitative evaluation of our age and gender estimation experiments, we use different evaluation metrics. Mean Absolute Error (MAE) and Cumulative Score (CS) measures are used for evaluating the age prediction task whereas Accuracy (ACC) is used for the gender estimation task. Besides accuracy, for evaluation of a classification model, two other measures, the positive predictive value (PPV) and the true positive rate (TPR) are considered as a performance metric for our experiments. To illustrate the performance metrices for the binary classification problem, the terms needed to form the equation are summarized in Table 3.

Table 3. The background parameters to describe the performance metrices for the classification model.

		Actual Class	
		Male	**Female**
Predicted Class	Male	True Positive (TP)	False Positive (FP)
	Female	False Negative (FN)	True Negative (TN)

MAE: The estimated age is reported as the mean absolute error (MAE) in years. It computes the mean of the absolute error between the predicted and ground truth age. MAE is considered as the de facto standard for measuring age estimation performance because the lion's share of the literature used it for the model evaluation. The MAE is calculated using Equation (2).

$$\text{MAE} = \frac{\sum_{i=1}^{n} \left| y_i^p - y_i^{gt} \right|}{N} \tag{2}$$

where N is the number of images that belongs to the test set, y_i^{gt} denotes the ground truth age, and y_i^p denotes the predicted age of the ith image.

Cumulative Score (CS): A cumulative score is the number of test images having an absolute error that is no larger than a threshold value t over the total number of test images. The equation used for calculating the cumulative score is given below:

$$\text{CS} = \frac{N_{ae<t}}{N} * 100\% \tag{3}$$

where $t \in [0, 100]$ and $N_{ae<t}$ represents the quantity of images from a test set that possesses an absolute error less than the specified threshold value. The total number of test images are denoted as N.

Accuracy (ACC): Accuracy is the number of correct predictions made by the model in relation to its overall prediction. It is a good measure when the target variable classes in the data are nearly balanced. The corresponding formula for accuracy is presented in Equation (4).

$$\text{ACC} = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

Precision (PPV): Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. The formula used for quantifying the precision is shown in Equation (5).

$$\text{PPV} = \frac{TP}{TP + FP} \tag{5}$$

Recall (TPR): Recall is the ratio of correctly predicted positive observations to all observations in the actual positive class. The standard method for measuring the recall is presented in Equation (6).

$$\text{TPR} = \frac{TP}{TP + FN} \tag{6}$$

3.5. Softmax Classification

We primarily consider the age and gender estimation task as a deep classification problem; hence, a Softmax activation function is applied in the final layer to normalize the arbitrary real value output of the network into probabilities for the predicted age and gender classes. The normalized output produced by the Softmax function ranges between 0 to 1 and the overall component sum is 1. The Softmax activation function generates the probabilities for every labeled class through Equation (7) mentioned below:

$$S(v)_i = \frac{e^{v_i}}{\sum_{j=1}^{N} e^{v_j}} \tag{7}$$

where i denotes the current element index of the input vector v, N is the total number of classes of the specified task and all v values are the components of the input vector.

4. Experiments and Results

At the beginning of this section, we discuss the implementation details of the age and gender estimation task. Next, we introduce the datasets used for both tasks. Subsequently, we report both quantitative and qualitative results observed following the experiments. Finally, we discuss the results at the end of this section.

4.1. Implementation Details

We trained our deployed CNN model separately based on the age and gender classification task. We handle both tasks as a classification approach. The models are trained using the deep learning framework PyTorch [75] developed by Facebook's artificial research lab. The training was performed on Nvidia GeForce GTX 1080 Ti GPU that consists of 3584 CUDA cores with 11 GB of video memory. Training on the large IMDB-WIKI datasets took almost one day whereas the fine-tuning on the smaller dataset only required a couple of hours.

In the case of age classification, we calculate the expected value from the Softmax probabilities belonged to output neurons, similar to what was considered in [24]. On the contrary to [24], we train the model with a balanced IMDB-WIKI dataset and introduce the batch normalization layer to reduce the training time. We consider every age as an individual class that ranges from 0 to 100. For all experiments regarding age estimation, the CNN is initialized with the weights trained on ImageNet [76]. This pre-trained model is then further trained on the IMDB-WIKI image dataset for classification with 101 output neurons. Finally, the CNN is fine-tuned on the test dataset.

For the gender classification, we report the gender class of the neuron carries the highest probability. We first deployed the pre-trained model trained on ImageNet. In the next step, we fine-tuned the pre-trained model with the real-world face image dataset Adience [9] with two output neurons and the performance is reported from the test split of the Adience dataset.

The training set consists of 80% of the images from the dataset and 20% is reserved for the testing. Further 90% of images from the training set are used for learning the weights and the rest of the images are used as validation set during the training phase. Every experiment begins in conjunction with pre-trained ImageNet weights from [25]. When fine-tuning the pre-trained network with a smaller dataset, the learning rate of 0.001 remains fixed except for the last layer. The last layer weights are initialized randomly as the number of output neurons are changing. We used the Adam optimizer with the setting of momentum 0.9 and a weight decay rate of 5×10^{-4}. We adjust the learning rate by a factor of 10 after every 30 epochs.

4.2. Datasets

In this paper, we use four different datasets for real age and gender estimation. We first introduce the datasets with a description of their specifications. Figure 7 represents exemplar images for each dataset used for the age estimation experiment and sample images for gender prediction are presented in Figure 8. Table 4 shows the size of each dataset with its properties. For the age classification, we trained our model with the IMDB-WIKI dataset and fine-tuned it with the MORPH dataset. We evaluate our age estimation model with the MORPH dataset. In the gender estimation task, we train and evaluate our model with the Adience dataset.

Figure 7. Exemplar images with real (biological) age from each dataset used in the age estimation experiment.

Figure 8. Exemplar images from the Adience dataset used for the gender estimation experiment.

Table 4. The particulars of the datasets used for our experiments with the corresponding training and testing split.

Dataset	Number of Images	Number of Subjects	Training Images	Testing/ Validation Images	Experiment	Age Range
IMDB	460,723	20,284	85,532	21,400	Real Age Estimation	0–100
WIKI	62,328				Real Age Estimation	0–100
MORPH	55,134	13,618	4380	1095	Real Age Estimation	16–77
Adience	26,580	2284	13,116	4372	Age Range/ Estimation	-

IMDB-WIKI: IMDB-WIKI is the biggest face image dataset labeling for age and gender that is free for public use. This dataset contains images with real age annotation in the range 0–100 and a total of 523,051 celebrity face images. The images were crawled from the IMDb website and Wikipedia. The age of a person is calculated based on the date of birth and the timestamp when the photo was taken (crawled from the sources). Among the half-a-million images, around 460 k images of 20,284 subjects are collected from IMDb and the rest, 62 k images are crawled straightaway from Wikipedia. A lot of images of this dataset are substandard images for training, such as humorous images, sketch images, severely occluded, full-length images, images containing multiple subjects, and blank images. For our experiment, firstly, we consider single-person images and then remove the wrongly annotated faces from individual classes. Thenceforth, we perform class-wise down-sampling until a threshold is basically formed by considering the low sample classes existing in the dataset. Finally, we enhance the image of the lower sampled classes from other similar datasets and perform data augmentation to make a nearly balanced

dataset. As a result, the total number of images that belong to the balanced dataset is 107 k and we use approximately 85 k images for training which is 80% of the balanced IMDB-WIKI dataset.

MORPH: The Craniofacial Longitudinal Morphological Face Database (Morph) is the most used dataset for real age estimation. It is a publicly available facial aging benchmark with about 55,000 facial images from more than 13,000 subjects. MORPH comprises 46,645 images of males and 8487 images of females with an age range from 16 to 77 years. For our age estimation experiments, we adopt the setup often used in the literature [22,24,30,48,49,52], where a subset of Caucasian people's images is used for the experiment. The system is evaluated by taking 20% of the images from this subset while the remaining 80% are used for network training. Although these works [50,77] do not follow the same setup while experimenting, we still report their result because of using the same benchmark.

Adience: Adience is a collection of face images from real-world and unconstrained imaging conditions. This dataset signifies all the aspects that are expected from an image collected from challenging real-world scenarios. There are face images that were uploaded to the Flickr website from smartphones without any filtering. Adience images, therefore, display a high-level of variations in noise, pose, and appearance, among others. The entire collection of the Adience dataset comprises nearly 26 K face images with 2288 distinct subjects. We used this dataset only for gender estimation for the sake of state-of-the-art comparison.

The age distribution among the datasets is presented in Figure 9. A large amount of variation is observed in the distribution curve. In the Morph dataset, two dense regions are observed in the early 20s and 40s. It seems that the images comprised in this academic Morph dataset were collected from two different data sources. From the distribution curve, it is clear that Wikipedia contains a long tail for elderly people whereas IMDB shows a peak in the young and middle adulthood. The IMDB and WIKI datasets maintain the image ratio of about 8 to 1. That is why the combined IMDB-WIKI dataset follows an analogous distribution to the IMDB dataset.

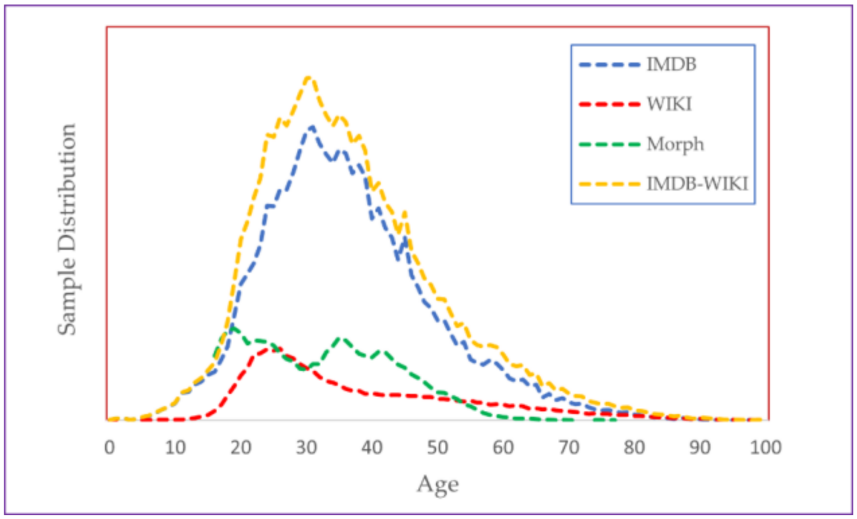

Figure 9. Distribution of age samples among the datasets experimented in our system.

4.3. Results

The quantitative results of our proposed human biological age and gender estimation system are reported in this section.

4.3.1. Age Estimation

We reported our age estimation results in MAE. We evaluated our system on the Morph dataset for estimating the real/biological age of a person. The Morph dataset has become one of the standard benchmarks for the real age estimation over the last few years.

We compare our results with the classic and state-of-the-art age estimation methods, such as Deep Expectation (DEX) [24], Ordinal Hyperplanes Ranker (OHRank) [49], AGES [31], AGE group-n encoding (AGEn) [57], OR-CNN [58], and Compact yet efficient Cascade Context-based Age Estimation (C3AE) [78] on the Morph dataset as shown in Table 5. As per the comparison table, the proposed method has a beneficial impact in estimating the age of a person over the same dataset and demonstrates better results than the traditional as well as deep learning-based age estimation models. The qualitative results of our model for the Morph dataset are presented in Figure 10.

Table 5. Result (MAE) comparison for real (biological) age estimation for the Morph dataset.

Methods	MAE
Rothe et al., (2016) [22]	3.45
DEX [24]	3.25
DEX (IMDB-WIKI)	2.68
AGES [31]	8.83
CA-SVR [48]	5.88
OHRank [49]	6.07
Guo and Mu (2014) [50]	3.92 *
Yi et al., (2014) [51]	3.63 *
Liu et al. [54]	2.68
Liu et al. [55]	2.32 *
Kim et al. [56]	4.29 *
AGEn [57]	2.93
AGEn (IMDB-WIKI)	2.52
OR-CNN [58]	3.27
C3AE [78]	2.78
Ours	**2.42**

* indicates different split.

Figure 10. A pictorial presentation of some of the test images with predicted age of the Morph dataset evaluated by our model.

From the comparison presented in Table 2, it can be concluded that the resulting MAE of our method for the Morph dataset is 2.42, and the results were meaningfully improved in comparison with the other hand-crafted feature-based age estimation approaches such as OHRank, and also exceeded the deeply learned models, namely DEX, Ranking-CNN, and C3AE.

As stated in the calculation procedure of the cumulative score (CS), the CS values for the Morph dataset under distinct error thresholds are plotted in Figure 11. From the figure, a steady growth of the CS value is observed if the allowable error thresholds increase. We plot the cross-entropy training and validation loss curve during the age estimation task in Figure 12. We stopped our training when the validation loss was increasing constantly.

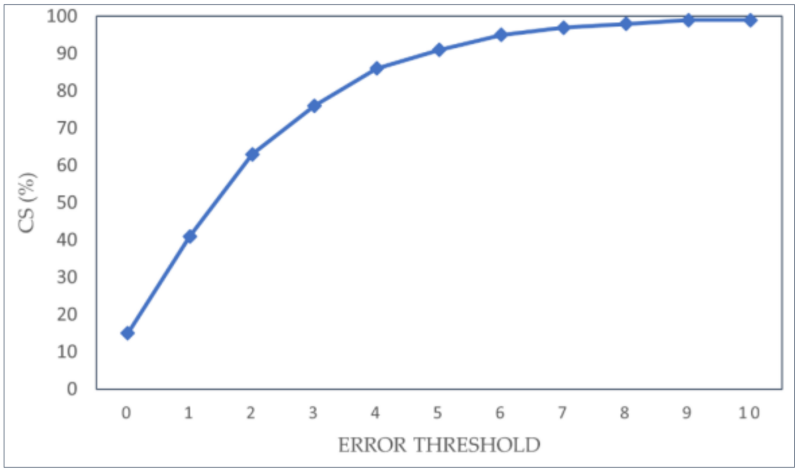

Figure 11. Cumulative score (CS) curve for the Morph dataset for the age estimation task.

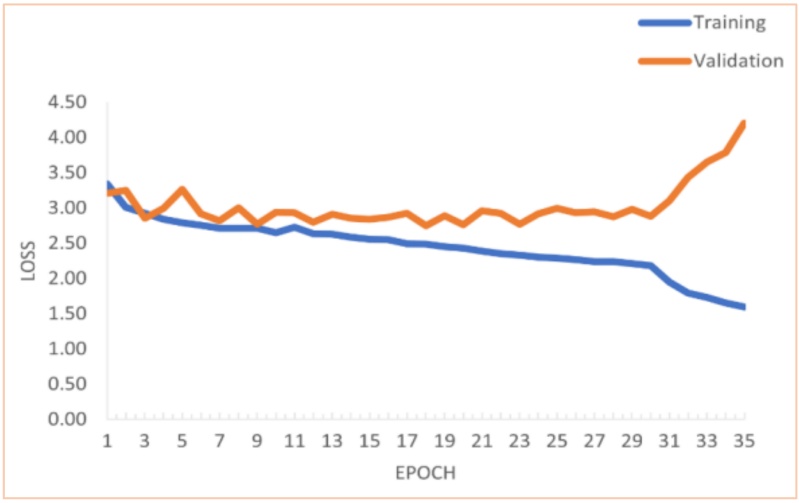

Figure 12. Graph of the cross-entropy loss results for the age estimation task.

In Table 6, we present the insightful parameters that lead the whole real age estimation experiments. In this comparative analysis, two factors (e.g., balance dataset, batch normalization) basically make the differences in the overall performance of the model. From the table, it is evident that when we train the model with a balanced dataset along with introducing a batch normalization layer in the deployed network, a good performance is obtained compared to the other experiment setup.

Table 6. The insights of the real age estimation experiments.

Pretrained on IMDB-WIKI	Fine-Tuned on Morph	Image Pre-Processing	Batch Normalization	Pretraining on Balance Dataset	MAE
Yes	Yes	Yes	Yes	Yes	2.42
Yes	Yes	Yes	Yes	No	2.96
Yes	Yes	Yes	No	Yes	3.75

4.3.2. Gender Estimation

We reported our gender estimation results in the form of classification accuracy (ACC). We evaluated our system for the Adience dataset for estimating the gender of a person. We assessed our model with multiple splits of the cross-validation protocol and present

the mean value of the performance metrices belonging to gender classification in Table 7. In Figure 13, we present the best ROC (receiver operating characteristics) curves with corresponding area under the curve (AUC) scores for the gender estimation results.

Table 7. The gender estimation model performance for the Adience dataset based on different cross-validation splits.

Number of Split (K)	ACC	TPR	PPV
5	97.41	97.21	97.22
4	93.36	92.97	92.63
3	96.99	96.45	97.06

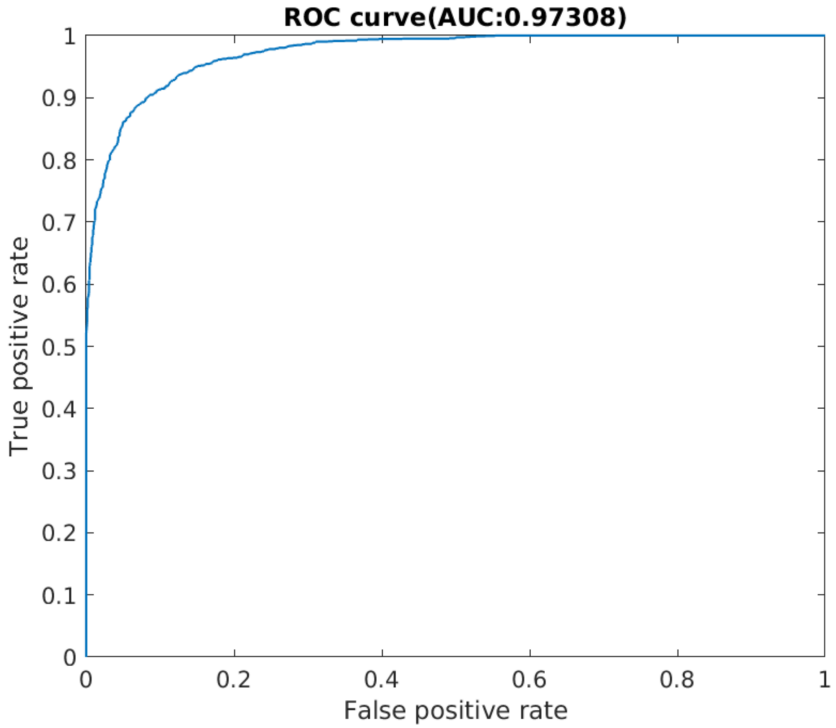

Figure 13. ROC curves for gender estimation results for the Adience benchmark.

It is very hard to compare our system with other works fairly because the validation protocol and image settings are varying among the approaches. We compared our reported results with state-of-the-art methods with corresponding classification accuracy in Table 8. In Figure 14, we have shown misclassified exemplar images from the Adience benchmark.

Table 8. Comparative gender estimation empirical results (ACC) for the Adience dataset.

Methods	ACC (%)
Levi et al. [23]	86.8
CNNs-EML [53]	77.8
Olatunbosun et al. [68]	96.2
Lapuschkin et al. [79]	85.9
Hassner et al. [80]	79.3
Khan et al. [81]	89.7
Ours	**97.28**

Figure 14. A pictorial presentation of some of the misclassified test images from the Adience dataset evaluated by our model. (**Top row**): Female subjects are misclassified as males. (**Bottom row**): Male subjects are misclassified as females.

5. Discussion

The proposed approach for estimating the real age and gender of a person demonstrates state-of-the-art results for the Morph dataset in anticipation of real age and the Adience dataset for gender estimation. We surpass the age estimation results marginally compared to the state-of-the-art DEX [24] method by lessening the data sparsity problem that exists in their approach. Our approach for the age estimation reduces the mean average error reported in the SOA method, i.e., the DEX fine-tuned for the IMDB-WIKI dataset and without fine-tuning by 0.26 years and 0.57 years, respectively. The system reveals that pre-training on a balanced dataset along with sufficient training data boosts the system performance reasonably. Our target application is a real-time system. Therefore, we incorporate the batch normalization layer before the non-linear RELU layer in the model to achieve a quicker convergence. In Table 3, it is empirically demonstrated that training the CNN model using a balanced dataset with batch normalization takes the lead in the result table. For the gender estimation, we used the ImageNet pre-trained model as the age estimation task without pre-training on the IMDB-WIKI dataset. We fine-tune and evaluate the model with the heterogenous Adience dataset and achieve SOA results. We achieve 5% more accuracy than the state-of-the-art approach [81].

In future research, we will try to enrich the balanced dataset because the amount of data in every class is not sufficient enough to train a very deep model and secure an enviable performance for a complex problem like real age estimation. We will try to devise an optimal network for the age and gender estimation from masked face images since smart store customers must wear a face mask amid the pandemic situation.

6. Conclusions

In this paper, we propose an automatic age (biological) and gender estimation system for the promising smart store enterprise which is a modern venture in the retail industry. This automated system can extract the human demographics necessary to provide the customer a very good shopping experience that results in a boost of offline smart store sales. In addition, this enterprise solution can ease the shopping process and shorten the shopping time for the consumers. Although recent methods show their potentials for the problem of age and gender estimation, the best works focused on constrained image benchmarks. As a result, these methods are not robust enough for the application involving real-world images. Most recently, some of the researchers learned to apply their model utilizing unconstrained image datasets, but these models are biased for the early and middle adulthood classes due to image sparsity problems in the dataset. In our paper, we resolve the data sparsity problem that exists in the state-of-the-art real age estimation approach DEX [24] by constructing class-wise data parity and incorporate batch normalization concepts that jointly improve the previous results marginally. We follow the

same setup for the gender estimation task and achieve substantially improved results over SOA methods regarding this task.

Author Contributions: Conceptualization, M.M.I. and J.-H.B.; methodology, M.M.I.; software, M.M.I.; validation, M.M.I. and J.-H.B.; formal analysis, M.M.I.; investigation, M.M.I; data curation, M.M.I.; writing—original draft preparation, M.M.I.; writing—review and editing, J.-H.B. and M.M.I.; visualization, M.M.I.; supervision, J.-H.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by the GRRC program of Gyeonggi province [GRRC Aviation 2017-B04, Development of Intelligent Interactive Media and Space Convergence Application System].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The authors have used publicly archived IMDB-WIKI and Adience dataset for the experiments. The IMDB-WIKI dataset is available in [24]. The Adience dataset is available in [9].

Acknowledgments: We would like to acknowledge Korea Aerospace University with much appreciation for its ongoing support to our research. We are thankful to authors who make their datasets publicly available to pave the way to research in this area.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ricanek, K.; Tesafaye, T. Morph: A Longitudinal Image Database of Normal Adult Age-Progression. In Proceedings of the Seventh International Conference on Automatic Face and Gesture Recognition, Southampton, UK, 10–12 April 2006; pp. 341–345. [CrossRef]
2. Geng, X.; Zhou, Z.-H.; Zhang, Y.; Li, G.; Dai, H. Learning from Facial Aging Patterns for Automatic Age Estimation. In Proceedings of the ACM International Conference on Multimedia, Santa Barbara, CA, USA, 23–27 October 2006; pp. 307–316. [CrossRef]
3. Minear, M.; Park, D.C. A lifespan database of adult facial stimuli. *Behav. Res. Methods Instrum. Comput.* **2004**, *36*, 630–633. [CrossRef] [PubMed]
4. Choi, S.E.; Lee, Y.J.; Lee, S.J.; Park, K.R.; Kim, J. Age Estimation Using a Hierarchical Classifier Based on Global and Local Facial Features. *Pattern Recognit.* **2011**, *44*, 1262–1281. [CrossRef]
5. Song, Z.; Ni, B.; Guo, D.; Sim, T.; Yan, S. Learning Universal Multi-View Age Estimator Using Video Context. In Proceedings of the International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 241–248. [CrossRef]
6. Fu, Y.; Guo, G.; Huang, T.S. Age synthesis and estimation via faces: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 1955–1976. [CrossRef] [PubMed]
7. Panis, G.; Lanitis, A.; Tsapatsoulis, N.; Cootes, T.F. Overview of research on facial ageing using the FG-NET ageing database. *IET Biom.* **2016**, *5*, 37–46. [CrossRef]
8. Chen, B.C.; Chen, C.S.; Hsu, W.H. Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset. *IEEE Trans. Multimed.* **2015**, *17*, 804–815. [CrossRef]
9. Eidinger, E.; Enbar, R.; Hassner, T. Age and gender estimation of unfiltered faces. *IEEE Trans. Inf. Forensics Secur.* **2014**, *9*, 2170–2179. [CrossRef]
10. Han, H.; Otto, C.; Jain, A.K. Age Estimation from Face Images: Human vs. Machine Performance. In Proceedings of the 2013 International Conference on Biometrics (ICB), Madrid, Spain, 4–7 June 2013; pp. 1–8. [CrossRef]
11. Guo, G. Human Age Estimation and Sex Classification. In *Video Analytics for Business Intelligence*; Shan, C., Porikli, F., Xiang, T., Gong, S., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; Volume 409, pp. 101–131. [CrossRef]
12. Asthana, A.; Zafeiriou, S.; Cheng, S.; Pantic, M. Robust Discriminative Response Map Fitting with Constrained Local Models. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Portland, OR, USA, 23–28 June 2013; pp. 3444–3451. [CrossRef]
13. Belhumeur, P.N.; Jacobs, D.W.; Kriegman, D.J.; Kumar, N. Localizing parts of faces using a consensus of exemplars. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 2930–2940. [CrossRef]
14. Cao, X.; Wei, Y.; Wen, F.; Sun, J. Face alignment by explicit shape regression. *Int. J. Comput. Vis.* **2014**, *107*, 177–190. [CrossRef]
15. Dantone, M.; Gall, J.; Fanelli, G.; Van Gool, L. Real-Time Facial Feature Detection Using Conditional Regression Forests. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012. [CrossRef]
16. Saragih, J.M.; Lucey, S.; Cohn, J.F. Deformable model fitting by regularized landmark mean-shift. *Int. J. Comput. Vis.* **2011**, *91*, 200–215. [CrossRef]

17. Xiong, X.; De la Torre, F. Supervised Descent Method and Its Applications to Face Alignment. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Portland, OR, USA, 23–28 June 2013; pp. 532–539. [CrossRef]
18. Davies, G.; Ellis, H.; Shepherd, J. Perceiving and remembering faces. *Am. J. Psychol.* **1983**, *96*, 151–154. [CrossRef]
19. Sinha, P.; Balas, B.; Ostrovsky, Y.; Russell, R. Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proc. IEEE* **2006**, *94*, 1948–1962. [CrossRef]
20. Gross, R.; Matthews, I.; Baker, S. Generic vs. person specific active appearance models. *Image Vis. Comput.* **2005**, *23*, 1080–1093. [CrossRef]
21. Zhang, Y.; Xu, T. Landmark-Guided Local Deep Neural Networks for Age and Gender Classification. *J. Sens.* **2018**, *2018*, 5034684. [CrossRef]
22. Rothe, R.; Timofte, R.; Gool, L.V. Some Like It Hot-Visual Guidance for Preference Prediction. In Proceedings of the 29th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016. [CrossRef]
23. Levi, G.; Hassner, T. Age and Gender Classification Using Convolutional Neural Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 34–42. [CrossRef]
24. Rothe, R.; Timofte, R.; Gool, L.V. Deep expectation of real and apparent age from a single image without facial landmarks. *Int. J. Comput. Vis.* **2018**, *126*, 144–157. [CrossRef]
25. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In Proceedings of the 3rd International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015.
26. Dib, M.Y.E.; Onsi, H.M. Human age estimation framework using different facial parts. *Egypt. Inform. J.* **2011**, *12*, 53–59. [CrossRef]
27. Atallah, R.R.; Kamsin, A.; Ismail, M.A.; Abdelrahman, S.A.; Zerdoumi, S. Face recognition and age estimation implications of changes in facial features: A critical review study. *IEEE Access* **2018**, *6*, 28290–28304. [CrossRef]
28. Young, H.K.; da Vitoria Lobo, N. Age Classification from Facial Images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 21–23 June 1994; pp. 762–767. [CrossRef]
29. Ramanathan, N.; Chellappa, R. Modeling Age Progression in Young Faces. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, NY, USA, 17–22 June 2006; pp. 387–394. [CrossRef]
30. Guo, G.; Fu, Y.; Dyer, C.R.; Huang, T.S. Image-based human age estimation by manifold learning and locally adjusted robust regression. *IEEE Trans. Image Process.* **2008**, *17*, 1178–1188. [CrossRef]
31. Geng, X.; Zhou, Z.-H.; Smith-Miles, K. Automatic Age Estimation Based on Facial Aging Patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 368. [CrossRef]
32. Yan, S.; Liu, M.; Huang, T.S. Extracting Age Information from Local Spatially Flexible Patches. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, NV, USA, 31 March–4 April 2008; pp. 737–740. [CrossRef]
33. Yan, S.; Zhou, X.; Liu, M.; Hasegawa-Johnson, M.; Huang, T.S. Regression from Patch-Kernel. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008; pp. 1–8. [CrossRef]
34. Fukunaga, K. *Introduction to Statistical Pattern Recognition*; Academic Press: Cambridge, MA, USA, 1990.
35. Rabiner, L.; Juang, B.-H. An introduction to hidden markov models. *IEEE ASSP Mag.* **1986**, *3*, 4–16. [CrossRef]
36. Zhuang, X.; Zhou, X.; Hasegawa-Johnson, M.; Huang, T. Face Age Estimation Using Patch-Based Hidden Markov Model Supervectors. In Proceedings of the International Conference on Pattern Recognition, Tampa, FL, USA, 8–11 December 2008. [CrossRef]
37. Gao, F.; Ai, H. Face Age Classification on Consumer Images with Gabor Feature and Fuzzy LDA Method. In *Advances in Biometrics*; Tistarelli, M., Nixon, M.S., Eds.; Springer: Berlin/Heidelberg, Germany, 2009; Volume 5558, pp. 132–141. [CrossRef]
38. Guo, G.; Mu, G.; Fu, Y.; Dyer, C.; Huang, T. A Study on Automatic Age Estimation Using a Large Database. In Proceedings of the International Conference on Computer Vision, Kyoto, Japan, 29 September–2 October 2009; pp. 1986–1991. [CrossRef]
39. Yang, Z.; Ai, H. Demographic Classification with Local Binary Patterns. In Proceedings of the International Conference on Biometrics (ICB), Seoul, Korea, 27–29 August 2007. [CrossRef]
40. Fu, Y.; Huang, T.S. Human age estimation with regression on discriminative aging manifold. *IEEE Trans. Multimed.* **2008**, *10*, 578–584. [CrossRef]
41. Lanitis, A.; Draganova, C.; Christodoulou, C. Comparing different classifiers for automatic age estimation. *IEEE Trans. Syst. Man Cybern.* **2004**, *34*, 621–628. [CrossRef] [PubMed]
42. Drucker, H.; Burges, C.J.C.; Kaufman, L.; Smola, A.J.; Vapnik, V. Support Vector Regression Machines. In Proceedings of the International Conference on Neural Information Processing Systems, Denver, CO, USA, 2–5 December 1996; pp. 155–161.
43. Geladi, P.; Kowalski, B.R. Partial least-squares regression: A tutorial. *Anal. Chim. Acta* **1986**, *185*, 1–17. [CrossRef]
44. Hardoon, D.R.; Szedmak, S.; Shawe-Taylor, J. Canonical correlation analysis: An overview with application to learning methods. *Neural Comput.* **2004**, *16*, 2639–2664. [CrossRef] [PubMed]
45. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
46. Han, H.; Otto, C.; Liu, X.; Jain, A.K. Demographic estimation from face images: Human vs. machine performance. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1148–1161. [CrossRef]
47. Zhang, Y.; Yeung, D.Y. Multi-Task Warped Gaussian Process for Personalized Age Estimation. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 2622–2629. [CrossRef]

48. Chen, K.; Gong, S.; Xiang, T.; Change Loy, C. Cumulative Attribute Space for Age and Crowd Density Estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 2467–2474. [CrossRef]
49. Chang, K.Y.; Chen, C.S.; Hung, Y.P. Ordinal Hyperplanes Ranker with Cost Sensitivities for Age Estimation. In Proceedings of the CVPR 2011, Colorado Springs, CO, USA, 20–25 June 2011. [CrossRef]
50. Guo, G.; Mu, G. A framework for joint estimation of age, gender and ethnicity on a large database. *Image Vis. Comput.* **2014**, *32*, 761–770. [CrossRef]
51. Yi, D.; Lei, Z.; Li, S.Z. Age Estimation by Multi-Scale Convolutional Network. In Proceedings of the Asian Conference on Computer Vision (ACCV), Singapore, 1–5 November 2014. [CrossRef]
52. Wang, X.; Guo, R.; Kambhamettu, C. Deeply-Learned Feature for Age Estimation. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 5–9 January 2015; pp. 534–541. [CrossRef]
53. Duan, M.; Li, K.; Yang, C.; Li, K. A hybrid deep learning CNN–ELM for age and gender classification. *Neurocomputing* **2018**, *275*, 448–461. [CrossRef]
54. Liu, X.; Zou, Y.; Kuang, H.; Ma, X. Face Image Age Estimation Based on Data Augmentation and Lightweight Convolutional Neural Network. *Symmetry* **2020**, *12*, 146. [CrossRef]
55. Liu, N.; Zhang, F.; Duan, F. Facial Age Estimation Using a Multi-Task Network Combining Classification and Regression. *IEEE Access* **2020**, *8*, 92441–92451. [CrossRef]
56. Kim, Y.H.; Nam, S.H.; Park, K.R. Enhanced Cycle Generative Adversarial Network for Generating Face Images of Untrained Races and Ages for Age Estimation. *IEEE Access* **2021**, *9*, 6087–6112. [CrossRef]
57. Tan, Z.; Wan, J.; Lei, Z.; Zhi, R.; Guo, G.; Li, S.Z. Efficient Group-n Encoding and Decoding for Facial Age Estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 2610–2623. [CrossRef]
58. Niu, Z.; Zhou, M.; Wang, L.; Gao, X.; Hua, G. Ordinal Regression with Multiple Output CNN for Age Estimation. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 4920–4928. [CrossRef]
59. Makinen, E.; Raisamo, R. Evaluation of gender classification methods with automatically detected and aligned faces. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 541–547. [CrossRef]
60. Reid, D.; Samangooei, S.; Chen, C.; Nixon, M.; Ross, A. Soft Biometrics for Surveillance: An overview. In *Handbook of Statistics, Machine Learning: Theory and Applications*; Elsevier: Amsterdam, The Netherlands, 2013; Volume 31, pp. 327–352. [CrossRef]
61. Golomb, B.A.; Lawrence, D.T.; Sejnowski, T.J. Sexnet: A Neural Network Identifies Sex from Human Faces. In Proceedings of the 1990 Conference on Advances in Neural Information Processing Systems, Denver, CO, USA, 26–29 November 1990; pp. 572–577.
62. Moghaddam, B.; Yang, M.-H. Learning gender with support faces. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 707–711. [CrossRef]
63. Baluja, S.; Rowley, H.A. Boosting sex identification performance. *Int. J. Comput. Vis.* **2006**, *71*, 111–119. [CrossRef]
64. Toews, M.; Arbel, T. Detection, localization, and sex classification of faces from arbitrary viewpoints and under occlusion. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 1567–1581. [CrossRef]
65. Yu, S.; Tan, T.; Huang, K.; Jia, K.; Wu, X. A study on gait-based gender classification. *IEEE Trans. Image Process.* **2009**, *18*, 1905–1910. [CrossRef]
66. Khan, F.S.; van de Weijer, J.; Anwer, R.M.; Felsberg, M.; Gatta, C. Semantic pyramids for gender and action recognition. *IEEE Trans. Image Process.* **2014**, *23*, 3633–3645. [CrossRef]
67. Chen, H.; Gallagher, A.; Girod, B. The Hidden Sides of Names—Face modeling with first name attributes. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 1860–1873. [CrossRef]
68. Agbo-Ajala, O.; Viriri, S. Deeply Learned Classifiers for Age and Gender Predictions of Unfiltered Faces. *Sci. World J.* **2020**, *2020*, 1289408. [CrossRef]
69. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet Classification with Deep Convolutional Neural Networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS), Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105. [CrossRef]
70. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis. IJCV* **2015**, *115*, 211–252. [CrossRef]
71. Girshick, R.B.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Seg-Mentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587. [CrossRef]
72. Felzenszwalb, P.F.; Girshick, R.B.; McAllester, D.; Ramanan, D. Object detection with discriminatively trained part-based models. *IEEE Trans. Pattern Anal. Mach. Intell. TPAMI* **2010**, *32*, 1627–1645. [CrossRef] [PubMed]
73. Mathias, M.; Benenson, R.; Pedersoli, M.; Gool, L.V. Face Detection without Bells and Whistles. In Proceedings of the IEEE European Conference on Computer Vision (ECCV), Zurich, Switzerland, 6–12 September 2014; pp. 720–735. [CrossRef]
74. Zhang, K.; Zhang, Z.; Li, Z.; Qiao, Y. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Process. Lett.* **2016**, *23*, 1499–1503. [CrossRef]

75. Paszke, A.; Gross, S.; Chintala, S.; Chanan, G.; Yang, E.; DeVito, Z.; Lin, Z.; Desmaison, A.; Antiga, L.; Lerer, A. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In Proceedings of the 33rd Conference on Neural Information Processing Systems (NIPS 2019), Vancouver, BC, Canada, 8–14 December 2019.
76. Deng, J.; Dong, W.; Socher, R.; Li, L.; Li, K.; Li, F.-F. ImageNet: A Large-Scale Hierarchical Image Database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255. [CrossRef]
77. Huerta, I.; Fernández, C.; Prati, A. Facial Age Estimation through the Fusion of Texture and Local Appearance Descriptors. In Proceedings of the IEEE European Conference on Computer Vision (ECCV), Zurich, Switzerland, 6–12 September 2014; pp. 667–681. [CrossRef]
78. Zhang, C.; Liu, S.; Xu, X.; Zhu, C. C3AE: Exploring the Limits of Compact Model for Age Estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 12579–12588. [CrossRef]
79. Lapuschkin, S.; Binder, A.; Muller, K.-R.; Samek, W. Understanding and Comparing Deep Neural Networks for Age and Gender Classification. In Proceedings of the IEEE International Conference on Computer Vision Workshop, Venice, Italy, 22–29 October 2017; pp. 1629–1638. [CrossRef]
80. Hassner, T.; Harel, S.; Paz, E.; Enbar, R. Effective Face Frontalization in Unconstrained Images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015 ; pp. 4295–4304. [CrossRef]
81. Khan, K.; Attique, M.; Syed, I.; Sarwar, G.; Irfan, M.A.; Khan, R.U. A Unified Framework for Head Pose, Age and Gender Classification through End-to-End Face Segmentation. *Entropy* **2019**, *21*, 647. [CrossRef]