



Learning to predict the cosmological structure formation

Siyu He^{a,b,c,1}, Yin Li^{d,e,f}, Yu Feng^{d,e}, Shirley Ho^{a,b,c,d,e,1}, Siamak Ravanbakhsh^g, Wei Chen^c, and Barnabás Póczos^h

^aDepartment of Physics, Carnegie Mellon University, Pittsburgh, PA 15213; ^bMcWilliams Center for Cosmology, Carnegie Mellon University, Pittsburgh, PA 15213; ^cCenter for Computational Astrophysics, Flatiron Institute, New York, NY 10010; ^dBerkeley Center for Cosmological Physics, University of California, Berkeley, CA 94720; ^ePhysics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720; ^fKavli Institute for the Physics and Mathematics of the Universe, University of Tokyo Institutes for Advanced Study, The University of Tokyo, Chiba 277-8583, Japan; ^gComputer Science Department, University of British Columbia, Vancouver, BC V6T1Z4, Canada; and ^hMachine Learning Department, Carnegie Mellon University, Pittsburgh, PA 15213

Edited by Neta A. Bahcall, Princeton University, Princeton, NJ, and approved May 23, 2019 (received for review December 17, 2018)

Matter evolved under the influence of gravity from minuscule density fluctuations. Nonperturbative structure formed hierarchically over all scales and developed non-Gaussian features in the Universe, known as the cosmic web. To fully understand the structure formation of the Universe is one of the holy grails of modern astrophysics. Astrophysicists survey large volumes of the Universe and use a large ensemble of computer simulations to compare with the observed data to extract the full information of our own Universe. However, to evolve billions of particles over billions of years, even with the simplest physics, is a daunting task. We build a deep neural network, the Deep Density Displacement Model (D³M), which learns from a set of prerun numerical simulations, to predict the nonlinear large-scale structure of the Universe with the Zel'dovich Approximation (ZA), an analytical approximation based on perturbation theory, as the input. Our extensive analysis demonstrates that D³M outperforms the second-order perturbation theory (2LPT), the commonly used fast-approximate simulation method, in predicting cosmic structure in the nonlinear regime. We also show that D³M is able to accurately extrapolate far beyond its training data and predict structure formation for significantly different cosmological parameters. Our study proves that deep learning is a practical and accurate alternative to approximate 3D simulations of the gravitational structure formation of the Universe.

cosmology | deep learning | simulation

Astrophysicists require a large amount of simulations to extract the information from observations (1–8). At its core, modeling structure formation of the Universe is a computationally challenging task; it involves evolving billions of particles with the correct physical model over a large volume over billions of years (9–11). To simplify this task, we either simulate a large volume with simpler physics or a smaller volume with more complex physics. To produce the cosmic web (12) in large volume, we select gravity, the most important component of the theory, to simulate at large scales. A gravity-only *N*-body simulation is the most popular and effective numerical method to predict the full 6D phase-space distribution of a large number of massive particles whose position and velocity evolve over time in the Universe (13). Nonetheless, *N*-body simulations are relatively computationally expensive, thus making the comparison of the *N*-body-simulated large-scale structure (of different underlying cosmological parameters) with the observed Universe a challenging task. We propose to use a deep model that predicts the structure formation as an alternative to *N*-body simulations.

Deep learning (14) is a fast-growing branch of machine learning, where recent advances have led to models that reach and sometimes exceed human performance across diverse areas, from analysis and synthesis of images (15–17), sound (18, 19), text (20, 21), and videos (22, 23) to complex control and planning tasks as they appear in robotics and game play (24–26). This new paradigm is also significantly impacting a variety of domains

in the sciences, from biology (27, 28) to chemistry (29, 30) and physics (31, 32). In particular, in astronomy and cosmology, a growing number of recent studies are using deep learning for a variety of tasks, ranging from analysis of cosmic microwave background (33–35), large-scale structure (36, 37), and gravitational lensing effects (38, 39) to classification of different light sources (40–42).

The ability of these models to learn complex functions has motivated many to use them to understand the physics of interacting objects, leveraging image, video, and relational data (43–53). However, modeling the dynamics of billions of particles in *N*-body simulations poses a distinct challenge.

In this work, we show that a variation on the architecture of a well-known deep-learning model (54) can efficiently transform the first-order approximations of the displacement field and approximate the exact solutions, thereby producing accurate estimates of the large-scale structure. Our key objective is to prove that this approach is an accurate and computationally efficient alternative to expensive cosmological simulations, and, to this end, we provide an extensive analysis of the results in the following section.

Significance

To understand the evolution of the Universe requires a concerted effort of accurate observation of the sky and fast prediction of structures in the Universe. *N*-body simulation is an effective approach to predicting structure formation of the Universe, though computationally expensive. Here, we build a deep neural network to predict structure formation of the Universe. It outperforms the traditional fast-analytical approximation and accurately extrapolates far beyond its training data. Our study proves that deep learning is an accurate alternative to the traditional way of generating approximate cosmological simulations. Our study shows that one can use deep learning to generate complex 3D simulations in cosmology. This suggests that deep learning can provide a powerful alternative to traditional numerical simulations in cosmology.

Author contributions: S. He, Y.L., S. Ho, and B.P. designed research; S. He, Y.L., Y.F., and S. Ho performed research; S. He, Y.L., S.R., and B.P. contributed new reagents/analytic tools; S. He, Y.L., Y.F., and W.C. analyzed data; and S. He, Y.L., Y.F., S. Ho, S.R., and W.C. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Data deposition: The source code of our implementation is available at <https://github.com/siyucosmo/ML-Recon>. The code to generate the training data is available at <https://github.com/rainwoodman/fastpm>.

¹To whom correspondence may be addressed. Email: shirleyho@flatironinstitute.org or siyuh@andrew.cmu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1821458116/-DCSupplemental.

Published online June 24, 2019.

The outcome of a typical N -body simulation depends on both the initial conditions and on cosmological parameters which affect the evolution equations. A striking discovery is that the Deep Density Displacement Model (D^3M), trained by using a single set of cosmological parameters, generalizes to new sets of significantly different parameters, minimizing the need for training data on a diverse range of cosmological parameters.

Setup

We build a deep neural network, D^3M , with similar input and output of an N -body simulation. The input to our D^3M is the displacement field from the Zel'dovich Approximation (ZA) (55). A displacement vector is the difference of a particle position at target redshift $z = 0$ —i.e., the present time—and its Lagrangian position on a uniform grid. ZA evolves the particles on linear trajectories along their initial displacements. It is accurate when the displacement is small; therefore, ZA is frequently used to construct the initial conditions of N -body simulations (56). As for the ground truth, the target displacement field is produced by using FastPM (57), a recent approximate N -body-simulation scheme that is based on a particle-mesh (PM) solver. FastPM quickly approaches a full N -body simulation with high accuracy and provides a viable alternative to direct N -body simulations for the purpose of our study.

A significantly faster approximation of N -body simulations is produced by second-order Lagrangian perturbation theory (2LPT), which bends each particle's trajectory with a quadratic correction (58). The 2LPT is used in many cosmological analyses to generate a large number of cosmological simulations for comparison of the astronomical dataset against the physical model (59, 60) or to compute the covariance of the dataset (61–63). We regard 2LPT as an effective way to efficiently generate a relatively accurate description of the large-scale structure, and therefore we select 2LPT as the reference model for comparison with D^3M .

We generate 10,000 pairs of ZAs as input and accurate FastPM approximations as the target. We use simulations of 32^3 N -body particles in a volume of $128 h^{-1} \text{Mpc}$ (600 million light years, where $h = 0.7$ is the Hubble parameter). The particles have a mean separation of $4 h^{-1} \text{Mpc}$ per dimension.

An important choice in our approach is training with a displacement field rather than a density field. Displacement field Ψ and density field ρ are two ways of describing the same distribution of particles. And an equivalent way to describe a density field is the overdensity field, defined as $\delta = \rho/\bar{\rho} - 1$, with $\bar{\rho}$ denoting the mean density. The displacement field and overdensity field are related by Eq. 1.

$$\begin{aligned} \mathbf{x} &= \Psi(\mathbf{q}) + \mathbf{q} \\ \delta(\mathbf{x}) &= \int d^3q \delta_D(\mathbf{x} - \mathbf{q} - \Psi(\mathbf{q})) - 1. \end{aligned} \quad [1]$$

When the displacement field is small and has zero curl, the choice of overdensity vs. displacement field for the output of the model is irrelevant, as there is a bijective map between these two representations, described by the equation:

$$\Psi = \int \frac{d^3k}{(2\pi)^3} e^{i\mathbf{k}\cdot\mathbf{q}} \frac{i\mathbf{k}}{k^2} \delta(\mathbf{k}). \quad [2]$$

However, as the displacements grow into the nonlinear regime of structure formation, different displacement fields can produce identical density fields (e.g., ref. 64). Therefore, providing the model with the target displacement field during the training eliminates the ambiguity associated with the density field. Our inability to produce comparable results when using the density

field as our input and target attests that relevant information resides in the displacement field (*SI Appendix, Fig. S1*).

Results and Analysis

Fig. 1 shows the displacement vector field as predicted by D^3M (*Left*) and the associated point-cloud representation of the structure formation (*Right*). It is possible to identify structures such as clusters, filaments, and voids in this point-cloud representation. We proceed to compare the accuracy of D^3M and 2LPT compared with ground truth.

Point-Wise Comparison. Let $\Psi \in \mathbb{R}^{d \times d \times d \times 3}$ denote the displacement field, where d is the number of spatial-resolution elements in each dimension ($d = 32$). A natural measure of error is the relative error $|\hat{\Psi} - \Psi_t|/|\Psi_t|$, where Ψ_t is the true displacement field (FastPM), and $\hat{\Psi}$ is the prediction from 2LPT or D^3M . Fig. 2 compares this error for different approximations in a 2D slice of a single simulation. We observe that D^3M predictions are very close to the ground truth, with a maximum relative error of 1.10 over all 1,000 simulations. For 2LPT, this number is significantly higher at 4.23. In average, the result of D^3M comes with a 2.8% relative error, while for 2LPT, it equals 9.3%.

Two-Point Correlation Comparison. As suggested by Fig. 2, the denser regions seem to have a higher error for all methods—that is, more nonlinearity in structure formation creates larger errors for both D^3M and 2LPT. The dependence of error on scale is computed with two- and three-point correlation analysis.

Cosmologists often use compressed summary statistics of the density field in their studies. The most widely used of these statistics are the two-point correlation function (2PCF) $\xi(r)$ and its Fourier transform, the power spectrum $P_{\delta\delta}(k)$:

$$\begin{aligned} \xi(|\mathbf{r}|) &= \langle \delta_A(\mathbf{r}') \delta_B(\mathbf{r}' + \mathbf{r}) \rangle, \\ P_{\delta\delta}(|\mathbf{k}|) &= \int d^3\mathbf{r} \xi(r) e^{i\mathbf{k}\cdot\mathbf{r}}, \end{aligned} \quad [3]$$

where the ensemble average $\langle \rangle$ is taken over all possible realizations of the Universe. Our Universe is observed to be both homogeneous and isotropic on large scales—i.e., without any special location or direction. This allows one to drop the dependencies on \mathbf{r}' and on the direction of \mathbf{r} , leaving only the amplitude $|\mathbf{r}|$ in the final definition of $\xi(r)$. In the second equation, $P_{\delta\delta}(k)$ is simply the Fourier transform of $\xi(r)$ and captures the dispersion of the plane-wave amplitudes at different scales in the Fourier space. \mathbf{k} is the 3D wavevector of the plane wave, and its amplitude k (the wavenumber) is related to the wavelength λ by $k = 2\pi/\lambda$. Due to isotropy of the Universe, we drop the vector form of \mathbf{r} and \mathbf{k} .

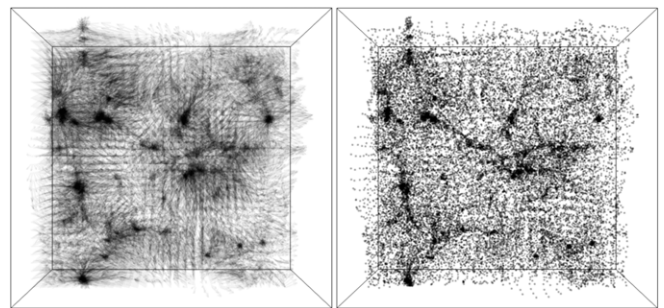


Fig. 1. The displacement vector field (*Left*) and the resulting density field (*Right*) produced by D^3M . The vectors in *Left* are uniformly scaled down for better visualization.

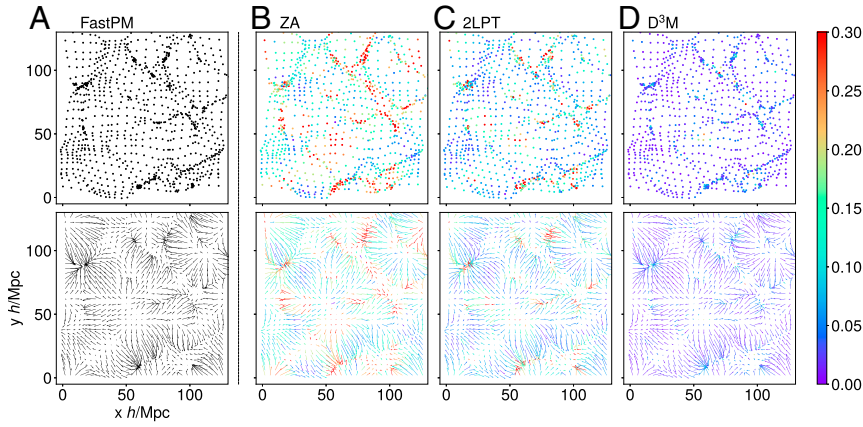


Fig. 2. The columns show 2D slices of full-particle distribution (*Upper*) and displacement vector (*Lower*) by various models: FastPM, the target ground truth, a recent approximate N -body simulation scheme that is based on a PM solver (A); ZA, a simple linear model that evolves particle along the initial velocity vector (B); 2LPT, a commonly used analytical approximation (C); and deep-learning model (D^3M) as presented in this work (D). While FastPM (A) served as our ground truth, B–D include color for the points or vectors. The color indicates the relative difference $(q_{\text{model}} - q_{\text{target}})/q_{\text{target}}$ between the target location (A) or displacement vector and predicted distributions by various methods (B–D). The error bar shows that denser regions have a higher error for all methods, which suggests that it is harder to predict highly nonlinear region correctly for all models: D^3M , 2LPT, and ZA. Our model D^3M has the smallest differences between predictions and ground truth among the above models B–D.

Because FastPM, 2LPT, and D^3M take the displacement field as input and output, we also study the two-point statistics for the displacement field. The displacement power spectrum is defined as:

$$P_{\Psi\Psi}(k) = \langle \Psi_x(k)\Psi_x^*(k) \rangle + \langle \Psi_y(k)\Psi_y^*(k) \rangle + \langle \Psi_z(k)\Psi_z^*(k) \rangle. \quad [4]$$

We focus on the Fourier-space representation of the two-point correlation. Because the matter and the displacement power spectrum take the same form, in what follows, we drop the subscript for matter and displacement field and use $P(k)$ to stand for both matter and displacement power spectrum. We use the transfer function $T(k)$ and the correlation coefficient $r(k)$ as metrics to quantify the model performance against the ground truth (FastPM) in the two-point correlation. We define the transfer function $T(k)$ as the square root of the ratio of two power spectra,

$$T(k) = \sqrt{\frac{P_{\text{pred}}(k)}{P_{\text{true}}(k)}}, \quad [5]$$

where $P_{\text{pred}}(k)$ is the density or displacement power spectrum as predicted by 2LPT or D^3M , and, analogously, $P_{\text{true}}(k)$ is the ground truth predicted by FastPM. The correlation coefficient $r(k)$ is a form of normalized cross-power spectrum,

$$r(k) = \frac{P_{\text{pred} \times \text{true}}(k)}{\sqrt{P_{\text{pred}}(k)P_{\text{true}}(k)}}, \quad [6]$$

where $P_{\text{pred} \times \text{true}}(k)$ is the cross-power spectrum between 2LPT or D^3M predictions and the ground-truth (FastPM) simulation result. The transfer function captures the discrepancy between amplitudes, while the correlation coefficient can indicate the discrepancy between phases as functions of scales. For a perfectly accurate prediction, $T(k)$ and $r(k)$ are both 1. In particular, $1 - r^2$ describes stochasticity, the fraction of the variance in the prediction that cannot be explained by the true model.

Fig. 3A shows the average power spectrum, transfer function $T(k)$, and stochasticity $1 - r^2(k)$ of the displacement field and the density field over 1,000 simulations. The transfer function of

density from 2LPT predictions is 2% smaller than that of FastPM on large scales ($k \approx 0.05 \text{ hMpc}^{-1}$). This is expected since 2LPT performs accurately on very large scales ($k < 0.01 \text{ hMpc}^{-1}$). The displacement transfer function of 2LPT increases >1 at $k \approx 0.35 \text{ hMpc}^{-1}$ and then drops sharply. The increase of the 2LPT displacement transfer function is because 2LPT overestimates the displacement power at small scales (e.g., ref. 65). There is a sharp drop of power near the voxel scale because smoothing over voxel scales in our predictions automatically erases power at scales smaller than the voxel size.

Now, we turn to the D^3M predictions: Both the density and displacement transfer functions of the D^3M differ from 1 by a mere 0.4% at scale $k \lesssim 0.4 \text{ hMpc}^{-1}$, and this discrepancy only increases to 2% and 4% for density field and displacement field, respectively, as k increases to the Nyquist frequency at 0.7 hMpc^{-1} . The stochasticity hovers at $\sim 10^{-3}$ and 10^{-2} for most scales. In other words, for both the density and displacement fields, the correlation coefficient between the D^3M predictions and FastPM simulations, all the way down to small scales of $k = 0.7 \text{ hMpc}^{-1}$, is $>90\%$. The transfer function and correlation coefficient of the D^3M predictions show that it can reproduce the structure formation of the Universe from large to semilinear scales. D^3M significantly outperforms our benchmark model 2LPT in the two-point function analysis. D^3M only starts to deviate from the ground truth at fairly small scales. This is not surprising, as the deeply nonlinear evolution at these scales is more difficult to simulate accurately and appears to be intractable by current analytical theories (66).

Three-Point Correlation Comparison. The three-point correlation function (3PCF) expresses the correlation of the field of interest among three locations in the configuration space, which is equivalently defined as bispectrum in Fourier space. Here, we concentrate on the 3PCF for computational convenience:

$$\zeta(r_1, r_2, \theta) = \langle \delta(\mathbf{x})\delta(\mathbf{x} + \mathbf{r}_1)\delta(\mathbf{x} + \mathbf{r}_2) \rangle, \quad [7]$$

where $r_1 = |\mathbf{r}_1|$ and $r_2 = |\mathbf{r}_2|$. Translation invariance guarantees that ζ is independent of \mathbf{x} . Rotational symmetry further eliminates all direction dependence except dependence on θ , the angle between \mathbf{r}_1 and \mathbf{r}_2 . The multipole moments of $\zeta(r_1, r_2, \theta)$,

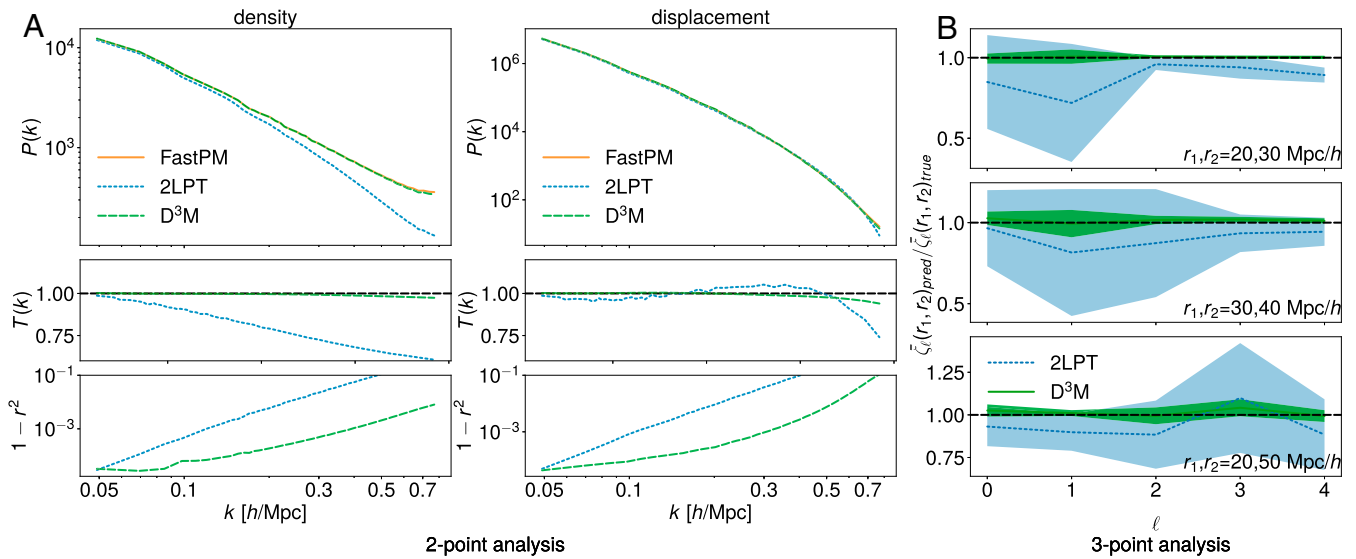


Fig. 3. (A) Displacement and density power-spectrum of FastPM (orange), 2LPT (blue), and c (green) (Top); transfer function—i.e., the square root of the ratio of the predicted power-spectrum to the ground truth (Middle); and $1 - r^2$, where r is the correlation coefficient between the predicted fields and the true fields (Bottom). Results are the averaged values of 1,000 test simulations. The transfer function and correlation coefficient of the D³M predictions are nearly perfect from large to intermediate scales and outperform our benchmark 2LPT significantly. (B) The ratios of the multipole coefficients ($\zeta_\ell(r_1, r_2)$) (to the target) of the two 3PCFs for several triangle configurations. The results are averaged over 10 test simulations. The error bars (padded regions) are the SDs derived from 10 test simulations. The ratio shows that the 3PCF of D³M is closer than 2LPT to our target FastPM with lower variance.

$\zeta_\ell(r_1, r_2) = (2\ell + 1) \int d\theta P_\ell(\cos\theta) \zeta(r_1, r_2, \theta)$, where $P_\ell(\cos\theta)$ is the Legendre polynomial of degree ℓ , can be efficiently estimated with pair counting (67). While the input (computed by ZA) do not contain significant correlations beyond the second order (power spectrum level), we expect D³M to generate densities with a 3PCF that mimics that of ground truth.

We compare the 3PCF calculated from FastPM, 2LPT, and D³M by analyzing the 3PCF through its multipole moments $\zeta_\ell(r_1, r_2)$. Fig. 3B shows the ratio of the binned multipole coefficients of the two 3PCFs for several triangle configurations, $\bar{\zeta}_\ell(r_1, r_2)_{\text{pred}} / \bar{\zeta}_\ell(r_1, r_2)_{\text{true}}$, where $\bar{\zeta}_\ell(r_1, r_2)_{\text{pred}}$ can be the 3PCF for D³M or 2LPT and $\bar{\zeta}_\ell(r_1, r_2)_{\text{true}}$ is the 3PCF for FastPM. We used 10 radial bins with $\Delta r = 5 h^{-1} \text{Mpc}$. The results are averaged over 10 test simulations, and the error bars are the SD. The ratio shows that the 3PCF of D³M is closer to FastPM than 2LPT, with smaller error bars. To further quantify our comparison, we calculate the relative 3PCF residual defined by

$$\begin{aligned} & \text{3PCF relative residual} \\ &= \frac{1}{9 \times N_r} \sum_{\ell=0}^8 \sum_{r_1, r_2} \frac{|\zeta_\ell(r_1, r_2)_{\text{pred}} - \zeta_\ell(r_1, r_2)_{\text{true}}|}{|\zeta_\ell(r_1, r_2)_{\text{true}}|}, \quad [8] \end{aligned}$$

where N_r is the number of (r_1, r_2) bins. The mean relative 3PCF residual of the D³M and 2LPT predictions compared with FastPM are 0.79% and 7.82%, respectively. The D³M accuracy on 3PCF is also an order of magnitude better than 2LPT, which indicates that the D³M is far better at capturing the non-Gaussian structure formation.

Generalizing to New Cosmological Parameters

So far, we train our model using a “single” choice of cosmological parameters $A_s = 2.142 \times 10^{-9}$ (hereafter $A_0 = 2.142 \times 10^{-9}$) and $\Omega_m = 0.3089$ (68). A_s is the primordial amplitude of the scalar perturbation from cosmic inflation, and Ω_m is the fraction of the total energy density that is matter at the present time, and we will call it matter density parameter for short. The

true exact value of these parameters is unknown, and different choices of these parameters change the large-scale structure of the Universe; Fig. 4.

Here, we report an interesting observation: The D³M trained on a single set of parameters in conjunction with ZA (which depends on A_s and Ω_m) as input can predict the structure formation for widely different choices of A_s and Ω_m . From a computational point of view, this suggests a possibility of producing simulations for a diverse range of parameters, with minimal training data.

Varying Primordial Amplitude of Scalar Perturbations A_s . After training the D³M using $A_s = A_0$, we change A_s in the input of our test set by nearly one order of magnitude: $A_s = 1.8A_0$ and $A_s = 0.2A_0$. Again, we use 1,000 simulations for analysis of each test case. The average relative displacement error of D³M remains $<4\%$ per voxel (compared with $<3\%$ when train and test data have the same parameters). This is still well below the error for 2LPT, which has relative errors of 15.5% and 6.3% for larger and smaller values of A_s , respectively.

Fig. 5A shows the transfer function and correlation coefficient for both D³M and 2LPT. The D³M performs much better than 2LPT for $A_s = 1.8A_0$. For small $A_s = 0.2A_0$, 2LPT does a better job than D³M predicting the density transfer function and correlation coefficient at the largest scales; otherwise, D³M predictions are more accurate than 2LPT at scales larger than $k = 0.08 h \text{Mpc}^{-1}$. We observe a similar trend with 3PCF analysis: The 3PCFs of D³M predictions are notably better than 2LPT ones for larger A_s , compared with smaller A_s , where it is only slightly better. These results confirm our expectation that increasing A_s increases the non-linearity of the structure-formation process. While 2LPT can predict fairly well in linear regimes, compared with D³M, its performance deteriorates with increased nonlinearity. It is interesting to note that the D³M prediction maintains its advantage, despite being trained on data from more linear regimes.

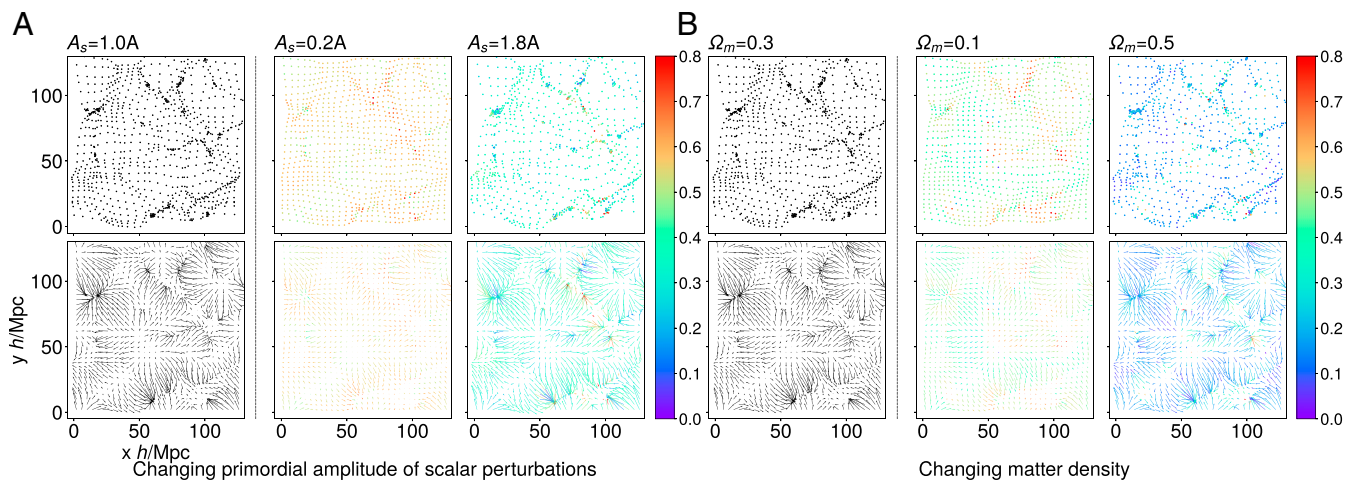


Fig. 4. We show the differences of particle distributions and displacement fields when we change the cosmological parameters A_s and Ω_m . (A) The error bar shows the difference of particle distribution (Upper) and displacement fields (Lower) between $A_s = A_0$ and the two extremes for $A_s = 0.2 A_0$ (Center) and $A_s = 1.8 A_0$ (Right). (B) A similar comparison showing the difference of the particle distributions (Upper) and displacement fields (Lower) for smaller and larger values of $\Omega_m \in \{0.1, 0.5\}$ with regard to $\Omega_m = 0.3089$, which was used for training. While the difference for the smaller value of A_s (Ω_m) is larger, the displacement for the larger A_s (Ω_m) is more nonlinear. This nonlinearity is due to concentration of mass and makes the prediction more difficult.

Varying Matter Density Parameter Ω_m . We repeat the same experiments, this time changing Ω_m to 0.5 and 0.1, while the model is trained on $\Omega_m = 0.3089$, which is quite far from both of the test sets. For $\Omega_m = 0.5$, the relative residual displacement errors of the D³M and 2LPT averaged over 1,000 simulations are 3.8% and 15.2%, and for $\Omega_m = 0.1$, they are 2.5% and 4.3%. Fig. 5 C and D show the two-point statistics for density field predicted by using different values of Ω_m . For $\Omega_m = 0.5$, the results show that the D³M outperforms 2LPT at all scales, while for smaller $\Omega_m = 0.1$, D³M outperforms 2LPT on smaller scales ($k > 0.1 \text{ hMpc}^{-1}$). As for the 3PCF of simulations with different values of Ω_m , the mean relative 3PCF residual of the D³M for $\Omega_m = 0.5$ and $\Omega_m = 0.1$ are 1.7% and 1.2%, respectively, and for 2LPT, they are 7.6% and 1.7%, respectively. The D³M prediction performs better at $\Omega_m = 0.5$ than $\Omega_m = 0.1$. This is again because the Universe is much more nonlinear at $\Omega_m = 0.5$ than $\Omega_m = 0.1$. The D³M learns more nonlinearity than is encoded in the formalism of 2LPT.

Conclusions

To summarize, our deep model D³M can accurately predict the large-scale structure of the Universe as represented by FastPM simulations, at all scales, as seen in Table 1. Furthermore, D³M learns to predict cosmic structure in the nonlinear regime more accurately than our benchmark model 2LPT. Finally, our model generalizes well to test simulations with cosmological parameters (A_s and Ω_m) significantly different from the training set. This suggests that our deep-learning model can potentially be deployed for a range of simulations beyond the parameter space covered by the training data (Table 1). Our results demonstrate that the D³M successfully learns the nonlinear mapping from first-order perturbation theory to FastPM simulation beyond what higher-order perturbation theories currently achieve.

Looking forward, we expect that replacing FastPM with exact N -body simulations would improve the performance of our method. As the complexity of our D³M model is linear in the

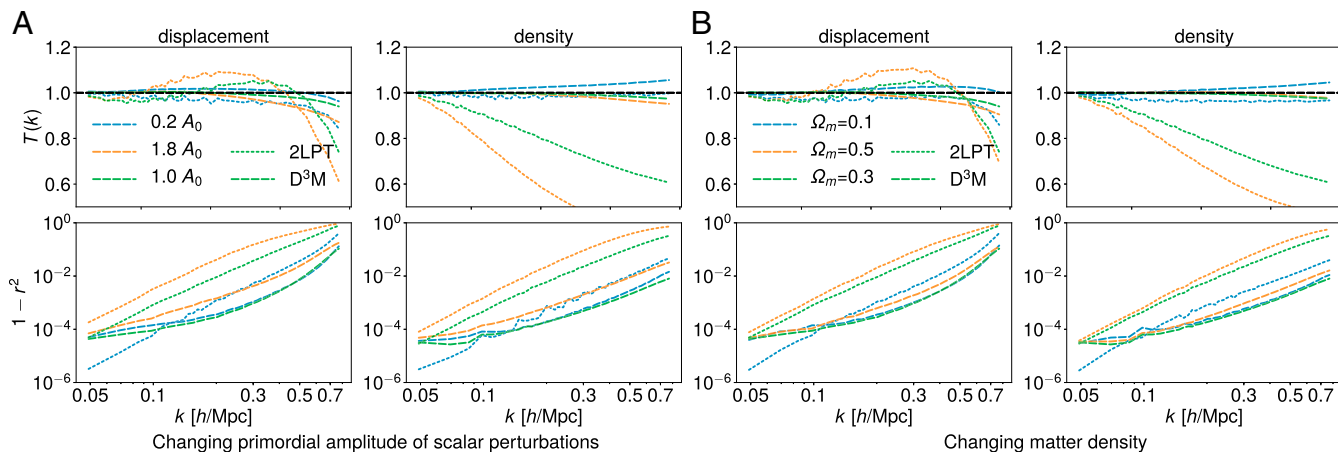


Fig. 5. Similar plots as in Fig. 3A, except that we test the two-point statistics when we vary the cosmological parameters without changing the training set (which has different cosmological parameters) or the trained model. We show predictions from D³M and 2LPT when tested on different A_s (A) and Ω_m (B). We show the transfer function—i.e., the square root of the ratio of the predicted power spectrum to the ground truth (Upper)—and $1 - r^2$, where r is the correlation coefficient between the predicted fields and the true fields (Lower). The D³M prediction outperforms 2LPT prediction at all scales except in the largest scales, as the perturbation theory works well in linear regime (large scales).

Table 1. A summary of our analysis

Data	Point-wise	$T(k)k = 0.11$	$r(k)k = 0.11$	$T(k)k = 0.50$	$r(k)k = 0.50$	3PCF
Test phase						
2LPT density	N/A	0.96	1.00	0.74	0.94	0.0782
D ³ M density	N/A	1.00	1.00	0.99	1.00	0.0079
2LPT displacement	0.093	0.96	1.00	1.04	0.90	N/A
D ³ M displacement	0.028	1.00	1.00	0.99	1.00	N/A
$A_s = 1.8A_0$						
2LPT density	N/A	0.93	1.00	0.49	0.78	0.243
D ³ M density	N/A	1.00	1.00	0.98	1.00	0.039
2LPT displacement	0.155	0.97	1.00	1.07	0.73	N/A
D ³ M displacement	0.039	1.00	1.00	0.97	0.99	N/A
$A_s = 0.2A_0$						
2LPT density	N/A	0.99	1.00	0.98	0.99	0.024
D ³ M density	N/A	1.00	1.00	1.03	1.00	0.022
2LPT displacement	0.063	0.99	1.00	0.95	0.98	N/A
D ³ M displacement	0.036	1.00	1.00	1.01	1.00	N/A
$\Omega_m = 0.5$						
2LPT density	N/A	0.94	1.00	0.58	0.87	0.076
D ³ M density	N/A	1.00	1.00	1.00	1.00	0.017
2LPT displacement	0.152	0.97	1.00	1.10	0.80	N/A
D ³ M displacement	0.038	1.00	1.00	0.98	0.99	N/A
$\Omega_m = 0.1$						
2LPT density	N/A	0.97	1.00	0.96	0.99	0.017
D ³ M density	N/A	0.99	1.00	1.04	1.00	0.012
2LPT displacement	0.043	0.97	1.00	0.97	0.98	N/A
D ³ M displacement	0.025	0.99	1.00	1.02	1.00	N/A

The unit of k is $h\text{Mpc}^{-1}$. N/A, not applicable.

number of voxels, we expect to be able to further improve our results if we replace the FastPM simulations with higher-resolution simulations. Our work suggests that deep learning is a practical and accurate alternative to the traditional way of generating approximate simulations of the structure formation of the Universe.

Materials and Methods

Dataset. The full simulation data consists of 10,000 simulations of boxes with ZA and FastPM as input–output pairs, with an effective volume of $20 (\text{Gpc}/h)^3$ ($190 \times 10^9 h^3$), comparable to the volume of a large spectroscopic sky survey like Dark Energy Spectroscopic Instrument or EUCLID. We split the full simulation dataset into 80%, 10%, and 10% for training, validation, and test, respectively. We also generated 1,000 simulations for 2LPT for each set of tested cosmological parameters.

Model and Training. The D³M adopts the U-Net architecture (54) with 15 convolution or deconvolution layers and $\sim 8.4 \times 10^6$ trainable parameters. Our D³M generalizes the standard U-Net architecture to work with 3D data (69–71). The details of the architecture are described in the following sections, and a schematic figure of the architecture is shown in *SI Appendix, Fig. S2*. In the training phase, we use the Adam Optimizer (72) with a learning rate of 0.0001, and first- and second-moment exponential decay rates equal to 0.9 and 0.999, respectively. We use the mean-squared error as the loss function (Loss Function) and L2 regularization with regularization coefficient 0.0001.

Details of the D³M Architecture. The contracting path follows the typical architecture of a convolution network. It consists of two blocks, each of which consists of two successive convolutions of stride 1 and a down-sampling convolution with stride 2. The convolution layers use $3 \times 3 \times 3$ filters with a periodic padding of size 1 (Padding and Periodic Boundary) on both sides of each dimension. Notice that at each of the two down-sampling steps, we double the number of feature channels. At the bottom of the D³M, another two successive convolutions with stride 1 and the same periodic padding as above are applied. The expansive path of our D³M is an inverted version of the contracting path of the network. (It includes two repeated applications of the expansion block, each of which consists of one up-sampling–transposed convolution with stride 1/2 and two successive con-

volution of stride 1. The transposed convolution and the convolution are constructed with $3 \times 3 \times 3$ filters.)

We take special care in the padding and cropping procedure to preserve the shifting and rotation symmetry in the up-sampling layer in expansive path. Before the transposed convolution, we apply a periodic padding of length 1 on the right, down, and back sides of the box [padding = (0,1,0,1,0,1) in pytorch], and after the transposed convolution, we discard one column on the left, up, and front sides of the box and two columns on the right, down, and back sides [crop = (1,2,1,2,1,2)].

A special feature of the D³M is the concatenation procedure, where the up-sampling layer halves the feature channels and then concatenates them with the corresponding feature channels on the contracting path, doubling the number of feature channels.

The expansive building block then follows a $1 \times 1 \times 1$ convolution without padding, which converts the 64 features to the final 3D displacement field. All convolutions in the network except the last one are followed by a rectified linear unit activation and batch normalization.

Padding and Periodic Boundary. It is common to use constant or reflective padding in deep models for image processing. However, these approaches are not suitable for our setting. The physical model we are learning is constructed on a spatial volume with a periodic boundary condition. This is sometimes also referred to as a torus geometry, where the boundaries of the simulation box are topologically connected—that is, $x_{i+L} = x_i$, where i is the index of the spatial location, and L is the periodicity (size of box). Constant or reflective padding strategies break the connection between the physically nearby points separated across the box, which not only loses information but also introduces noise during the convolution, further aggravated with an increased number of layers.

We find that the periodic padding strategy significantly improves the performance and expedites the convergence of our model, comparing to the same network using a constant padding strategy. This is not surprising, as one expects that it is easier to train a model that can explain the data than to train a model that does not.

Loss Function. We train the D³M to minimize the mean square error on particle displacements

$$\mathcal{L} = \frac{1}{N} \sum_i (\hat{\Psi}_i - \Psi_{t,i})^2, \quad [9]$$

where i labels the particles and N is the total number of particles. This loss function is proportional to the integrated squared error, and by using a

Fourier transform and Parseval's theorem, it can be rewritten as

$$\int (\hat{\Psi} - \Psi_t)^2 d^3q = \int |\hat{\Psi} - \Psi_t|^2 d^3k = \int d^3k \left(|\Psi_t|^2 (1 - T)^2 + 2 |\hat{\Psi}| |\Psi_t| (1 - r) \right), \quad [10]$$

where \mathbf{q} is the Lagrangian space position, and \mathbf{k} its corresponding wavevector. T is the transfer function defined in Eq. 5, and r is the correlation coefficient defined in Eq. 6, which characterize the similarity between the predicted and true fields, in amplitude and phase, respectively. Eq. 10 shows that our simple loss function jointly captures both of these measures: As T and r approach 1, the loss function approaches 0.

Data Availability. The source code of our implementation is available at <https://github.com/siyucosmo/ML-Recon>. The code to generate the training data is also available at <https://github.com/rainwoodman/fastpm>.

ACKNOWLEDGMENTS. We thank Angus Beane, Peter Braam, Gabriella Contardo, David Hogg, Laurence Levasseur, Pascal Ripoche, Zack Slepian, and David Spergel for useful suggestions and comments; Angus Beane for comments on the paper; and Nick Carriero for comments on Center for Computational Astrophysics (CCA) computing clusters. The work was supported partially by the Simons Foundation. The FastPM simulations were generated on the computer cluster Edison at the National Energy Research Scientific Computing Center, a US Department of Energy Office of Science User Facility operated under Contract DE-AC02-05CH11231. The training of the neural network model was performed on the CCA computing facility and the Carnegie Mellon University AutonLab computing facility. The open-source software toolkit nbodykit (73) was used for the clustering analysis. Y.L. was supported by the Berkeley Center for Cosmological Physics and the Kavli Institute for the Physics and Mathematics of the Universe, established by the World Premier International Research Center Initiative of the MEXT, Japan. S. Ho was supported by NASA Grants 15-WFIRST15-0008 and Research Opportunities in Space and Earth Sciences Grant 12-EUCLID12-0004; and the Simons Foundation.

1. M. Colless *et al.*, The 2dF galaxy redshift survey: Spectra and redshifts. *Mon. Not. R. Astron. Soc.* **328**, 1039–1063 (2001).
2. D. J. Eisenstein *et al.* SDSS-III: Massive spectroscopic surveys of the distant universe, the Milky Way galaxy, and extra-solar planetary systems. *Astron. J.* **142**, 72 (2011).
3. H. D. Jones *et al.* The 6dF galaxy survey: Final redshift release (DR3) and southern large-scale structures. *Mon. Not. R. Astron. Soc.* **399**, 683–698 (2009).
4. J. Liske *et al.*, Galaxy and Mass Assembly (GAMA): End of survey report and data release 2. *Mon. Not. R. Astron. Soc.* **452**, 2087–2126 (2015).
5. M. Scodreggio *et al.* The VIMOS Public Extragalactic Redshift Survey (VIPERS). Full spectroscopic data and auxiliary information release (PDR-2). arXiv:1611.07048 (21 November 2016).
6. Ž. Ivezić *et al.*, LSST: From science drivers to reference design and anticipated data products. arXiv:0805.2366 (15 May 2008).
7. L. Amendola *et al.*, Cosmology and fundamental physics with the Euclid satellite. *Living Rev. Relativ.* **21**, 2 (2018).
8. D. Spergel *et al.*, Wide-Field InfraRed Survey Telescope—Astrophysics Focused Telescope Assets WFIRST-AFTA 2015 report. arXiv:1503.03757 (12 March 2015).
9. T. MacFarland, H. M. P. Couchman, F. R. Pearce, J. Pichlmeier, A new parallel P³M code for very large-scale cosmological simulations. *New Astron.* **3**, 687–705 (1998).
10. V. Springel, N. Yoshida, S. D. M. White, GADGET: A code for collisionless and gasdynamical cosmological simulations. *New Astron.* **6**, 79–117 (2001).
11. J. S. Bagla, TreePM: A code for cosmological N -body simulations. *J. Astrophys. Astron.* **23**, 185–196 (2002).
12. J. R. Bond, L. Kofman, D. Pogosyan, How filaments of galaxies are woven into the cosmic web. *Nature* **380**, 603–606 (1996).
13. M. Davis, G. Efstathiou, C. S. Frenk, S. D. M. White, The evolution of large-scale structure in a universe dominated by cold dark matter. *Astrophys. J.* **292**, 371–394 (1985).
14. Y. LeCun, Y. Bengio, G. Hinton, Deep learning. *Nature* **521**, 436–444 (2015).
15. G. Huang, Z. Liu, L. Van Der Maaten, K. Q. Weinberger, “Densely connected convolutional networks” in *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, Piscataway, NJ, 2017), pp. 2261–2269.
16. T. Karras, T. Aila, S. Laine, J. Lehtinen, Progressive growing of GANs for improved quality, stability, and variation. arXiv:1710.10196 (27 October 2017).
17. I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A. C. Courville, “Improved training of Wasserstein GANs” in *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, I. Guyon *et al.*, Eds. (Neural Information Processing Systems Foundation, Inc., San Diego, CA, 2017), pp. 5767–5779.
18. A. Van Den Oord *et al.*, WaveNet: A generative model for raw audio. arXiv:1609.03499 (12 September 2016).
19. D. Amodei *et al.*, “Deep speech 2: End-to-end speech recognition in English and Mandarin” in *Proceedings of the 33rd International Conference on Machine Learning*, M. F. Balcan, K. Q. Weinberger, Eds. (Association for Computing Machinery, New York, NY, 2016), pp. 173–182.
20. Z. Hu, Z. Yang, X. Liang, R. Salakhutdinov, E. P. Xing, Toward controlled generation of text. arXiv:1703.00955 (2 March 2017).
21. A. Vaswani *et al.*, “Attention is all you need” in *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, I. Guyon *et al.*, Eds. (Neural Information Processing Systems Foundation, Inc., San Diego, CA, 2017), pp. 5998–6008.
22. E. Denton, R. Fergus, Stochastic video generation with a learned prior. arXiv:1802.07687 (21 February 2018).
23. J. Donahue *et al.*, “Long-term recurrent convolutional networks for visual recognition and description” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, Piscataway, NJ, 2015), pp. 2625–2634.
24. D. Silver *et al.*, Mastering the game of Go with deep neural networks and tree search. *Nature* **529**, 484–489 (2016).
25. V. Mnih *et al.*, Human-level control through deep reinforcement learning. *Nature* **518**, 529–533 (2015).
26. S. Levine, C. Finn, T. Darrell, P. Abbeel, End-to-end training of deep visuomotor policies. *J. Mach. Learn. Res.* **17**, 13d34–1373 (2016).
27. T. Ching *et al.*, Opportunities and obstacles for deep learning in biology and medicine. *J. R. Soc. Interf.* **15**, 20170387 (2018).
28. B. Alipanahi, A. Delong, M. T. Weirauch, B. J. Frey, Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nat. Biotechnol.* **33**, 831–838 (2015).
29. M. H. S. Segler, M. Preuss, M. P. Waller, Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* **555**, 604–610 (2018).
30. J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, G. E. Dahl, Neural message passing for quantum chemistry. arXiv:1704.01212 (4 April 2017).
31. G. Carleo, M. Troyer, Solving the quantum many-body problem with artificial neural networks. *Science* **355**, 602–606 (2017).
32. C. Adam-Bourdarios *et al.*, “The Higgs boson machine learning challenge” in *Proceedings of the NIPS 2014 Workshop on High-Energy Physics and Machine Learning* (Neural Information Processing Systems Foundation, Inc., San Diego, CA, 2015), pp. 19–55.
33. S. He, S. Ravanbakhsh, S. Ho, “Analysis of cosmic microwave background with deep learning” in *Proceedings of the 33rd International Conference on Machine Learning* (Journal of Machine Learning Research, 2016), Vol. 48. (2018).
34. N. Perraudin, M. Defferrard, T. Kacprzak, R. Sgier, DeepSphere: Efficient spherical convolutional neural network with HEALPix sampling for cosmological applications. arXiv:1810.12186 (29 October 2018).
35. J. Caldeira *et al.*, DeepCMB: Lensing reconstruction of the cosmic microwave background with deep neural networks. arXiv:1810.01483 (2 October 2018).
36. S. Ravanbakhsh *et al.*, Estimating cosmological parameters from the dark matter distribution. arXiv:1711.02033 (6 November 2017).
37. A. Mathuriya *et al.*, CosmoFlow: Using deep learning to learn the universe at scale. arXiv:1808.04728 (14 August 2018).
38. Y. D. Hezaveh, L. P. Levasseur, P. J. Marshall, Fast automated analysis of strong gravitational lenses with convolutional neural networks. *Nature* **548**, 555–557 (2017).
39. F. Lanusse *et al.*, CMU DeepLens: Deep learning for automatic image-based galaxy-galaxy strong lens finding. *Mon. Not. R. Astron. Soc.* **473**, 3895–3906 (2018).
40. N. Kennamer, D. Kirkby, A. Ihler, F. J. Sanchez-Lopez, “ContextNet: Deep learning for star galaxy classification” in *Proceedings of the 35th International Conference on Machine Learning*, J. Dy, A. Krause, Eds. (Proceedings of Machine Learning Research, Journal of Machine Learning Research, 2018), vol. 80, pp. 2582–2590.
41. E. J. Kim, R. J. Brunner, Star-galaxy classification using deep convolutional neural networks. *Mon. Not. R. Astron. Soc.* **464**, 4463–4475 (2016).
42. M. Lochner, J. D. McEwen, H. V. Peiris, O. Lahav, M. K. Winter, Photometric supernova classification with machine learning. *Astrophys. J. Suppl. Ser.* **225**, 31 (2016).
43. P. W. Battaglia, J. B. Hamrick, J. B. Tenenbaum, Simulation as an engine of physical scene understanding. *Proc. Natl. Acad. Sci. U.S.A.* **45**, 18327–18332 (2013).
44. P. Battaglia *et al.*, “Interaction networks for learning about objects, relations and physics” in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, D. D. Lee, U. von Luxburg, R. Garnett, M. Sugiyama, I. Guyon, Eds. (Association for Computing Machinery, New York, NY, 2016), pp. 4502–4510.
45. R. Mottaghi, H. Bagheri-zhad, M. Rastegari, A. Farhadi, “Newtonian scene understanding: Unfolding the dynamics of objects in static images” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, Piscataway, NJ, 2016), pp. 3521–3529.
46. M. B. Chang, T. Ullman, A. Torralba, J. B. Tenenbaum, A compositional object-based approach to learning physical dynamics. arXiv:1612.00341 (1 December 2016).
47. J. Wu, I. Yildirim, J. J. Lim, B. Freeman, J. Tenenbaum, “Galileo: Perceiving physical object properties by integrating a physics engine with deep learning” in *Proceedings of the 28th International Conference on Neural Information Processing Systems*, C. Cortes, D. D. Lee, M. Sugiyama, R. Garnett, Eds. (Association for Computing Machinery, New York, NY, 2015), pp. 127–135.
48. J. Wu, J. J. Lim, H. Zhang, J. B. Tenenbaum, W. T. Freeman, “Physics 101: Learning physical object properties from unlabeled videos” in *Proceedings of the British Machine Vision Conference 2016*, R. C. Wilson, E. R. Hancock, W. A. P. Smith, Eds. (BMVA Press, Durham, UK, 2016), pp. 39.1–39.12.
49. N. Watters *et al.*, “Visual interaction networks: Learning a physics simulator from video” in *Advances in Neural Information Processing Systems* (Neural Information Processing Systems Foundation, Inc., 2017), vol. 30, pp. 4539–4547.
50. A. Lerer, S. Gross, R. Fergus, Learning physical intuition of block towers by example. arXiv:1603.01312 (3 March 2016).

51. P. Agrawal, A. V. Nair, P. Abbeel, J. Malik, S. Levine, "Learning to poke by poking: Experiential learning of intuitive physics" in *Advances in Neural Information Processing Systems*, D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, R. Garnett, Eds. (Neural Information Processing Systems Foundation, Inc., 2016), vol. 29, pp. 5074–5082.
52. K. Fragkiadaki, P. Agrawal, S. Levine, J. Malik, Learning visual predictive models of physics for playing billiards. arXiv:1511.07404 (23 November 2015).
53. J. Tompson, K. Schlachter, P. Sprechmann, K. Perlin, Accelerating Eulerian fluid simulation with convolutional networks. arXiv:1607.03597 (13 July 2016).
54. O. Ronneberger, P. Fischer, T. Brox, "U-Net: Convolutional networks for biomedical image segmentation" in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, N. Navab, J. Hornegger, W. Wells, A. Frangi, Eds. (Lecture Notes in Computer Science, Springer, Cham, Switzerland, 2015), vol. 9351, pp. 234–241.
55. Y. B. Zel'dovich, Gravitational instability: An approximate theory for large density perturbations *Astron. Astrophys.* **5**, 84–89 (1970).
56. M. White, The Zel'dovich approximation. *Mon. Not. R. Astron. Soc.* **439**, 3630–3640 (2014).
57. Y. Feng, M.-Y. Chu, U. Seljak, P. McDonald, FASTPM: A new scheme for fast simulations of dark matter and haloes. *Mon. Not. R. Astron. Soc.* **463**, 10.1093/mnras/stw2123 (2016).
58. T. Buchert, Lagrangian theory of gravitational instability of Friedman-Lemaître cosmologies—a generic third-order model for nonlinear clustering. *Mon. Not. R. Astron. Soc.* **267**, 811–820 (1994).
59. J. Jasche, B. D. Wandelt, Bayesian physical reconstruction of initial conditions from large-scale structure surveys. *Mon. Not. R. Astron. Soc.* **432**, 894–913 (2013).
60. F.-S. Kitaura, The initial conditions of the Universe from constrained simulations. *Mon. Not. R. Astron. Soc.* **429**, L84–L88 (2013).
61. K. S. Dawson et al., The Baryon oscillation spectroscopic survey of SDSS-III. *Astron. J.* **145**, 10 (2013).
62. K. S. Dawson et al., The SDSS-IV extended Baryon oscillation spectroscopic survey: Overview and early data. *Astron. J.* **151**, 44 (2016).
63. DESI Collaboration et al., The DESI Experiment part I: Science, targeting, and survey design. arXiv:1611.00036 (31 October 2016).
64. Y. Feng, U. Seljak, M. Zaldarriaga, Exploring the posterior surface of the large scale structure reconstruction. *J. Cosmol. Astropart. Phys.* **7**, 043 (2018).
65. K. C. Chan, Helmholtz decomposition of the Lagrangian displacement (2014) *Phys. Rev. D* **89**, 083515.
66. A. Perko, L. Senatore, E. Jennings, R. H. Wechsler, Biased tracers in redshift space in the EFT of large-scale structure. arXiv:1610.09321 (28 October 2016).
67. Z. Slepian, D. J. Eisenstein, Computing the three-point correlation function of galaxies in $\mathcal{O}(N^2)$ time. *Mon. Not. R. Astron. Soc.* **454**, 4142–4158 (2015).
68. Planck Collaboration et al., Planck 2015 results. XIII. Cosmological parameters *Astron. Astrophys.* **594**, A13 (2016).
69. F. Milletari, N. Navab, S.-A. Ahmadi, V-Net: Fully convolutional neural networks for volumetric medical image segmentation. arXiv:1606.04797 (15 June 2016).
70. P. Berger, G. Stein, A volumetric deep convolutional neural network for simulation of mock dark matter halo catalogues. *Mon. Not. R. Astron. Soc.* **482**, 2861–2871 (2019).
71. M. A. Aragon-Calvo, Classifying the large scale structure of the universe with deep neural networks. arXiv:1804.00816 (3 April 2018).
72. D. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv:1412.6980 (22 December 2014).
73. N. Hand et al., nbodykit: An open-source, massively parallel toolkit for large-scale structure. *Astron. J.* **156**, 160 (2018).