
Reconstructing the local density field with combined convolutional and point cloud architecture

Baptiste Barthe-Gold

Ecole Polytechnique
Rte de Saclay, 91120 Palaiseau, France
baptiste.barthe-gold@polytechnique.edu

Nhat-Minh Nguyen 

Center for Data-Driven Discovery
Kavli IPMU, UTIAS, The University of Tokyo
Kashiwa, Chiba 277-8583, Japan
nhat.minh.nguyen@ipmu.jp

Leander Thiele 

Center for Data-Driven Discovery
Kavli IPMU, UTIAS, The University of Tokyo
Kashiwa, Chiba 277-8583, Japan
leander.thiele@ipmu.jp

Abstract

We construct a neural network to perform regression on the local dark-matter density field given line-of-sight peculiar velocities of dark-matter halos, biased tracers of the dark matter field. Our architecture combines a convolutional U-Net with a point-cloud DeepSets. This combination enables efficient use of small-scale information and improves reconstruction quality relative to a U-Net-only approach. Specifically, our hybrid network recovers both clustering amplitudes and phases better than the U-Net on small scales.

1 Introduction

One key objective in observational cosmology is the inference of the local matter density field in the late-time universe. Given that the matter distribution is dominated by the invisible dark matter, inference must rely on indirect probes. One such probe is galaxy peculiar velocity. Peculiar velocity is the velocity component of galaxies induced by large-scale matter clustering (separated from the smooth Hubble flow component due to cosmic expansion).

Unlike the formation and clustering of galaxies, the peculiar velocities of galaxies are directly sourced by gravitational interactions of the matter distribution itself, hence a fundamentally unbiased tracer of the matter density field in the local universe, out to the distance of a few hundred Megaparsecs (Mpc), though only the radial, i.e. line-of-sight, component of galaxy peculiar velocities can be measured from their Doppler redshift.

Peculiar velocity determinations rely on direct distance measurements, for example via the fundamental plane, Tully-Fisher relations or supernovae Ia standard candles. Reconstruction of the local density field from peculiar velocities has a venerable history. Traditionally, linear methods have been employed, i.e. relations of the form $\delta = Mv$, where δ is the (over)density field $\rho/\bar{\rho} - 1$, v is the array of line-of-sight velocities, and M is a linear operator. The direct inversion technique is an application

of linear theory (written in Fourier conjugate \mathbf{k} -space and restricted to redshift $z = 0$ for simplicity):

$$\delta_L(\mathbf{k}) = -\frac{i}{Hf} \mathbf{k} \cdot \mathbf{v}_{\text{LOS}}(\mathbf{k}), \quad (1)$$

where H is the Hubble parameter and f is the linear growth rate of the density field δ , while \mathbf{v}_{LOS} denotes the line-of-sight velocity vector. In the context of Bayesian inference, assuming a Gaussian prior on δ , one obtains the alternative linear technique, namely Wiener filtering [1, 2, 3].

Linear reconstruction is necessarily limited to large scales (the regime where $\delta \ll 1$) and thus cannot optimally use the dense information available in modern peculiar velocity data. In order to improve the reconstruction on nonlinear scales, sampling of initial conditions through a forward model has been developed [e.g. 4, 5, 6]. The computational expense of running a gravity solver in each sampling step and the approximations in the solver limit forward modeling in resolution. Furthermore, being able to run only a few chains within reasonable computational budgets makes it difficult to explore systematic effects [7].

Therefore, machine learning has been introduced as a possible improvement. Recent works use a convolutional neural network to perform the mapping from input data to the density field. Some methods only use the observed (galaxy) density field as input, while others also include the gridded line-of-sight velocities [8, 9].

Our contribution aims to assess improvements on this method in two aspects. First, directly passing the gridded radial velocity appears suboptimal given the inductive bias of convolutional networks. Indeed, except for boundary effects the point of the convolutional network is to exploit translational invariance for weight sharing. In order to improve the match between inductive bias and problem formulation, we pass the linear reconstruction δ_L from Eq. (1). Second, working solely with a grid misses small-scale information, particularly in high-density regions. In our convolutional setup, any vorticity information is also neglected. More involved models have been developed to improve small-scale behavior, such as using graph neural networks along with a learned assignment scheme [10]. We propose another method and add a second component to the architecture that is responsible for recovering the small-scale features. This component is chosen as a point-cloud architecture such that it can be evaluated directly on the local set of surrounding galaxies.

We restrict ourselves to application in numerical simulations in this work. Evaluation on observational data such as Cosmicflows-4 [11] requires careful treatment of complicated systematic effects in the training data. Furthermore, for such application we will need to include stochasticity in the network architecture (presumably in a generative fashion). For simplicity, we only consider regression of the mean in this work, noting that the inclusion of stochasticity should be relatively straightforward. In addition, we attempt to prevent information about the tracer density field from informing the prediction, in order to isolate and focus on the information content in peculiar velocities.

After introducing the network and training method in Sec. 2, we will show in Sec. 3 that the neural network improves reconstruction quality on nonlinear scales, i.e. $k = 0.1 - 1 \, h\text{Mpc}^{-1}$.

2 Method

2.1 Data

We train on the high-resolution version of the Quijote fiducial simulations [12] at redshift $z = 0$. The single-redshift approximation is highly accurate and sufficient for the typical size of peculiar velocity surveys (see above).

In this work, we assume that the background cosmology is known and fixed, hence the same cosmology for all simulations. Note that this is a standard assumption in most (if not all) reconstructions [e.g. 13, 5, 14] and their derived astrophysics applications [e.g. 15, 16, 17]. In future works, especially for a joint inference of cosmological parameters and the matter density field, we will additionally consider simulations with varying cosmology.

We further approximate galaxies as dark matter halos, identified with Rockstar [18], a friends-of-friend halo-finding algorithm in 6D phase-space. For each of the 100 simulation boxes with side length $1 \, h^{-1}\text{Gpc}$, we randomly pick a certain halo as the observer's host halo, then proceed to collect all neighboring halos within a sphere of $200 \, h^{-1}\text{Mpc}$ radius. This radius roughly corresponds to

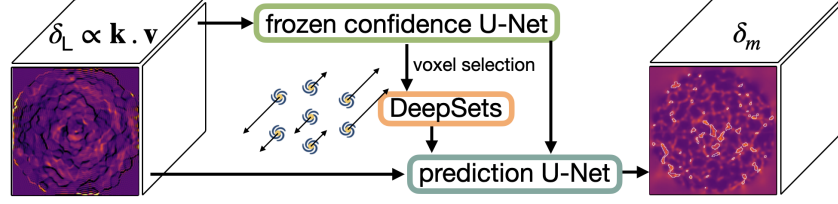


Figure 1: Schematic representation of our architecture. The backbone is a U-Net evaluated on the velocity divergence. Additional small-scale information is provided by a DeepSets point-cloud architecture which is evaluated locally. Due to the high cost of the DeepSets evaluation, a pretrained and then frozen confidence U-Net is used to select a small percentage of voxels for which the DeepSets evaluation is deemed worth the expense.

the maximum range of current peculiar velocity surveys. We thus use $\sim 30,000$ halos per training sample, and generate 256 of those training samples per simulation, which results in a dataset of 25,600 items. We split those into training, validation, and test sets in an 80/10/10 proportion among the 100 Quijote simulations, thereby preventing information leakage.

Constructing the peculiar velocity field from halo peculiar velocities requires first constructing their momentum grid and then dividing by their density grid. Care usually must be taken during the latter to avoid spurious bias due to the sparsity of the halos. The resulting peculiar velocity field is, in any case, inevitably biased low in amplitude. To this end, one might attempt to correct for such bias analytically; here, however, we take a pragmatic approach and smooth the momentum and density fields with a Gaussian kernel after assignment. For small values of the kernel size, the direct inversion δ_L shows large artifacts. On the other hand, a large kernel size leads to loss of small-scale information. We thus stack three choices of kernel size (1, 2, $4 h^{-1} \text{Mpc}$) as separate channels in the input to the convolutional network and let the learning algorithm find the optimal weighting. We use a 128^3 regular grid, corresponding to a cell spacing of $\approx 3.23 h^{-1} \text{Mpc}$.

2.2 Network Architecture

Our architecture is a combination of convolutional and point cloud, as we illustrate schematically in Fig. 1.

For the point-cloud part, we use a DeepSets architecture [19] which operates on the halos in a $10 h^{-1} \text{Mpc}$ radius around a given voxel. The DeepSets uses the relative position and line-of-sight velocity of the halos as point-level data while the global position of the respective voxel and mean line-of-sight velocity in the set are treated as global features.

Computationally, it is challenging to evaluate the DeepSets across the entire volume (it would mean 128^3 separate DeepSets computations for each training sample). We thus introduce a trick to make the problem manageable. We train a separate two-headed U-Net (the “confidence network”) to predict both the expectation value $\bar{\delta}$ and the error σ in the style of a moment network [20]. The confidence network is trained with a v loss [21] with $\beta = 0.5$. Only for the $N_{\text{DS}} = 5 \times 10^4$ voxels (about 2.4% of volume) with the highest predicted error σ do we evaluate the DeepSets. Empirically, the convergence in N_{DS} is quite good and we do not expect substantial improvement from assigning more resources to the DeepSets evaluation.

For the convolutional part, we use a U-Net [22] with LeakyReLU activations and layer normalization. The input channels consist of:

- the direct inversion linear density field δ_L for three different smoothing scales (as described above),
- the position of the voxel relative to the observer,
- the DeepSets output (four channels in our setup), and
- the frozen confidence network output $\bar{\delta}$.

Table 1: The Mean Squared Error (MSE) obtained by different models evaluated on the test set.

Model	MSE
Direct inversion ($1 h^{-1}\text{Mpc}$ smoothing)	4.94
Direct inversion ($4 h^{-1}\text{Mpc}$ smoothing)	4.48
3D Wiener filter	3.80
Normal U-Net	3.20
Confidence U-Net (μ prediction)	3.26
U-Net + DeepSets (50k voxels)	2.99

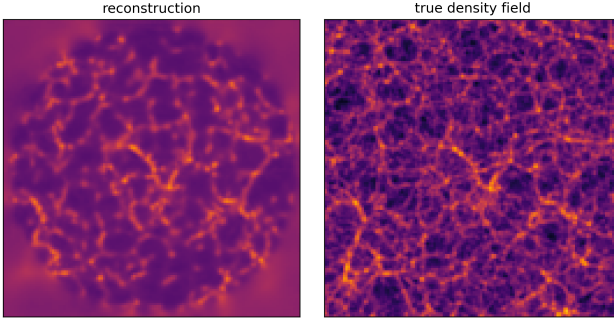


Figure 2: The same density field slice through the reconstruction (left) and the truth (right), on logarithmic color scale.

The output y of the U-Net is transformed as $\hat{\delta} = \text{ReLU}(\sinh(y)) - 1$ in order to better match the dynamic range of the density field: many values of order unity, but much larger in high-density regions. A similar transformation was found beneficial in Ref. [23].

2.3 Network Training

The training procedure is split into two stages, first training the confidence U-Net model by itself and then training the full model using the confidence U-Net as one of the input channels. The main drawback of this method is having to train two models separately.

We optimize the field-level mean-squared error (or β -NLL loss for the confidence model) using the AdamW optimizer [24] with a one-cycle learning rate scheduler [25].

The model is trained using 24 Nvidia A100 GPUs. Each model is trained until a convergence of the validation loss, usually reached at around 200 epochs. The training time for the whole model is on the order of 30 hours.

Table 1 reports the MSE for all models considered, when evaluated on the test set. Most models yield quantitative gains over the baseline U-Net. The only exception is the confidence U-Net although not really unexpected, given its β -NLL objective does not optimize the MSE. Note that the confidence U-Net only serves as one input channel for the final DeepSets-augmented model, which is trained separately and achieves the lowest MSE among all methods.

3 Results

We begin by presenting our model output in an example from the test set, displayed in Fig. 2. The circular geometry in the reconstructed field is due to the selection of tracer halos within a $200 h^{-1}\text{Mpc}$ radius. We observe good agreement between the reconstruction and the underlying true density field over a range of scales. Of course, the reconstructed field loses fidelity on small scales due to the finite tracer density and lack of transverse modes. Outside the region covered by tracers, the model mostly falls back to the mean, although thanks to the learned cosmic web prior it does predict some vague structures like the density in the lower left.

Next, we quantify the performance of the model using two-point statistics. The cross-correlation and transfer function are shown in Fig. 3. There, the baseline linear reconstruction via Wiener filtering is actually optimistic since it uses the full 3-dimensional velocities. In a realistic setup with only the line-of-sight component available, the quality of the linear reconstruction would be even worse.

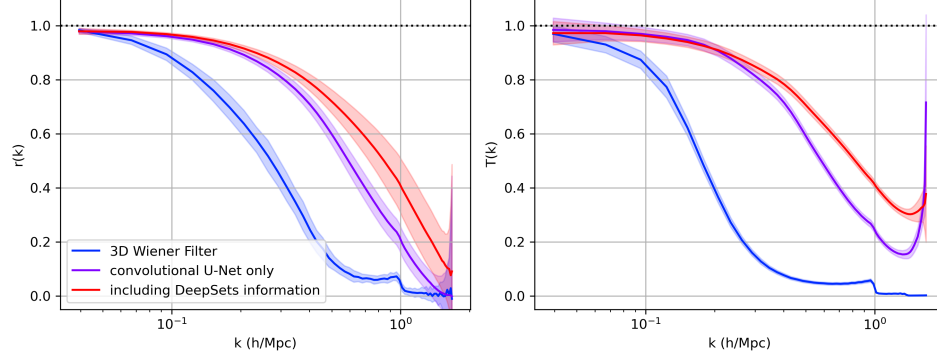


Figure 3: Performance evaluation in two-point statistics. The left panel shows the cross-correlation coefficient, while the right panel shows the transfer function. Blue is the 3D Wiener filter (which is an upper bound on linear reconstruction), magenta shows the reconstruction with a convolutional U-Net only, and red shows our complete architecture which includes small-scale information from the DeepSets component. The shaded areas correspond to the standard deviation of the metric across the whole testing dataset.

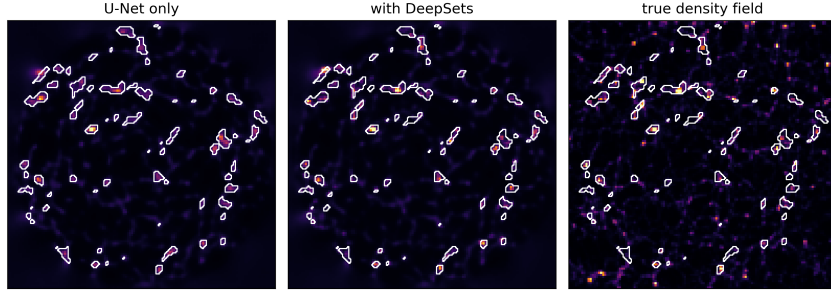


Figure 4: Visual illustration of the effect of the small-scale DeepSets information. Compared to Fig. 2, the color scale concentrates on the high-density regime. The contours represent approximately the regions which the confidence network picked for DeepSets evaluation.

Our U-Net improves substantially on the linear reconstruction both in cross-correlation and transfer function. Even better results are obtained when the sub-grid and vorticity information is included with the DeepSets. The improvement from DeepSets occurs for scales smaller than $k \sim 0.1 - 0.2 hMpc^{-1}$ depending on whether cross-correlation or transfer function are considered.

In order to gain more intuition about the action of the DeepSets, we show an example in Fig. 4. There, the white contours approximately delineate the regions where the confidence network returned the highest uncertainty and which are thus passed on for DeepSets evaluation. We observe that the confidence network predominantly picks out the high-density regions. The correlation coefficient between σ and the underlying δ is greater than 0.8 beyond $k = 1 hMpc^{-1}$. This is physically expected and implies that the DeepSets with fixed receptive field have a good chance to improve the prediction thanks to higher tracer density. Within the regions impacted by the DeepSets, the prediction of high-density small-scale features is improved. However, some density peaks were completely missed by the confidence network, implying that there could be room for improvement.

4 Conclusions

We have presented a hybrid convolutional and point-cloud neural network trained to recover the local density field from peculiar velocities. The machine learning reconstruction dramatically improves upon linear Wiener filtering. Our results indicate that even with moderate tracer density (below Cosmicflows-4 density) the extra small-scale information recovered by the point-cloud network is significant. This result could change once observational errors are folded in, which is the subject of

future work. With multiple surveys gathering data on peculiar velocities, the future seems bright for machine learning solutions to the problem. The combination of a U-Net with a confidence network-controlled DeepSets could find application in other problems in cosmology and beyond.

Acknowledgements

BBG thanks Ben Horowitz for helpful discussions on the Wiener filter implementation. BBG thanks the International Laboratory for Astrophysics, Neutrino and Cosmology Experiments (ILANCE) to organize research internships for Master students at the Center for Data-driven Discovery (CD3), Kavli IPMU (WPI), where this work was conducted. NMN acknowledges support from the Japan Foundation for Promotion of Astronomy Research Grant and the Japan Foundation for Promotion of Science (JSPS) KAKENHI Grant Number 25K23373. LT is supported by JSPS KAKENHI Grant 24K22878.

References

- [1] Saleem Zaroubi, Yehuda Hoffman, and Avishai Dekel. Wiener Reconstruction of Large-Scale Structure from Peculiar Velocities. *Astrophys. J.*, 520(2):413–425, August 1999.
- [2] Hélène M. Courtois, Yehuda Hoffman, R. Brent Tully, and Stefan Gottlöber. Three-dimensional Velocity and Density Reconstructions of the Local Universe with Cosmicflows-1. *Astrophys. J.*, 744(1):43, January 2012.
- [3] Yehuda Hoffman, Aurelien Valade, Noam I. Libeskind, Jenny G. Sorce, R. Brent Tully, Simon Pfeifer, Stefan Gottlöber, and Daniel Pomarède. The large-scale velocity field from the Cosmicflows-4 data. *Mon. Not. Roy. Astron. Soc.*, 527(2):3788–3805, January 2024.
- [4] Huiyuan Wang, H. J. Mo, Xiaohu Yang, Y. P. Jing, and W. P. Lin. ELUCID—Exploring the Local Universe with the Reconstructed Initial Density Field. I. Hamiltonian Markov Chain Monte Carlo Method with Particle Mesh Dynamics. *Astrophys. J.*, 794(1):94, October 2014.
- [5] J. Jasche and G. Lavaux. Physical Bayesian modelling of the non-linear matter distribution: New insights into the nearby universe. *Astron. Astrophys.*, 625:A64, May 2019.
- [6] James Prideaux-Ghee, Florent Leclercq, Guilhem Lavaux, Alan Heavens, and Jens Jasche. Field-based physical inference from peculiar velocity tracers. *Mon. Not. Roy. Astron. Soc.*, 518(3):4191–4213, January 2023.
- [7] Nhat-Minh Nguyen, Fabian Schmidt, Guilhem Lavaux, and Jens Jasche. Impacts of the physical data model on the forward inference of initial conditions from biased tracers. *J. Cosmol. Astropart. Phys.*, 2021(3):058, March 2021.
- [8] Punyakoti Ganeshiah Veena, Robert Lilow, and Adi Nusser. Large-scale density and velocity field reconstructions with neural networks. *Monthly Notices of the Royal Astronomical Society*, 522(4):5291–5307, April 2023.
- [9] Sungwook E. Hong, Donghui Jeong, Ho Seong Hwang, and Juhan Kim. Revealing the Local Cosmic Web from Galaxies by Deep Learning. *Astrophys. J.*, 913(1):76, May 2021.
- [10] Yurii Kvasiuk, Jordan Krywonos, Matthew C. Johnson, and Moritz Münchmeyer. Reconstruction of continuous cosmological fields from discrete tracers with graph neural networks, 2024.
- [11] R. Brent Tully, Ehsan Kourkchi, Hélène M. Courtois, Gagandeep S. Anand, John P. Blakeslee, Dillon Brout, Thomas de Jaeger, Alexandra Dupuy, Daniel Guinet, Cullan Howlett, Joseph B. Jensen, Daniel Pomarède, Luca Rizzi, David Rubin, Khaled Said, Daniel Scolnic, and Benjamin E. Stahl. Cosmicflows-4. *Astrophys. J.*, 944(1):94, February 2023.
- [12] Francisco Villaescusa-Navarro, ChangHoon Hahn, Elena Massara, Arka Banerjee, Ana Maria Delgado, Doogesh Kodi Ramanah, Tom Charnock, Elena Giusarma, Yin Li, Erwan Allys, Antoine Brochard, Cora Uhlemann, Chi-Ting Chiang, Siyu He, Alice Pisani, Andrej Obuljen,

- Yu Feng, Emanuele Castorina, Gabriella Contardo, Christina D. Kreisch, Andrina Nicola, Justin Alsing, Roman Scoccimarro, Licia Verde, Matteo Viel, Shirley Ho, Stephane Mallat, Benjamin Wandelt, and David N. Spergel. The Quijote Simulations. *Astrophys. J. Suppl.*, 250(1):2, September 2020.
- [13] Guilhem Lavaux. Bayesian 3D velocity field reconstruction with VIRBIUS. *Mon. Not. Roy. Astron. Soc.*, 457(1):172–197, March 2016.
 - [14] Guilhem Lavaux, Jens Jasche, and Florent Leclercq. Systematic-free inference of the cosmic matter density field from SDSS3-BOSS data. *arXiv e-prints*, page arXiv:1909.06396, September 2019.
 - [15] Nhat-Minh Nguyen, Jens Jasche, Guilhem Lavaux, and Fabian Schmidt. Taking measurements of the kinematic Sunyaev-Zel’dovich effect forward: including uncertainties from velocity reconstruction with forward modeling. *J. Cosmol. Astropart. Phys.*, 2020(12):011, December 2020.
 - [16] Eleni Tsaprazi, Nhat-Minh Nguyen, Jens Jasche, Fabian Schmidt, and Guilhem Lavaux. Field-level inference of galaxy intrinsic alignment from the SDSS-III BOSS survey. *J. Cosmol. Astropart. Phys.*, 2022(8):003, August 2022.
 - [17] D. J. Bartlett, A. Kostić, H. Desmond, J. Jasche, and G. Lavaux. Constraints on dark matter annihilation and decay from the large-scale structure of the nearby Universe. *Phys. Rev. D*, 106(10):103526, November 2022.
 - [18] Peter S. Behroozi, Risa H. Wechsler, and Hao-Yi Wu. The ROCKSTAR Phase-space Temporal Halo Finder and the Velocity Offsets of Cluster Cores. *Astrophys. J.*, 762(2):109, January 2013.
 - [19] Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Ruslan Salakhutdinov, and Alexander Smola. Deep sets, 2018.
 - [20] Niall Jeffrey and Benjamin D. Wandelt. Solving high-dimensional parameter inference: marginal posterior densities & Moment Networks. *arXiv e-prints*, page arXiv:2011.05991, November 2020.
 - [21] Maximilian Seitzer, Arash Tavakoli, Dimitrije Antic, and Georg Martius. On the pitfalls of heteroscedastic uncertainty estimation with probabilistic neural networks. In *International Conference on Learning Representations*, 2022.
 - [22] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.
 - [23] Leander Thiele, Francisco Villaescusa-Navarro, David N. Spergel, Dylan Nelson, and Annalisa Pillepich. Teaching Neural Networks to Generate Fast Sunyaev-Zel’dovich Maps. *Astrophys. J.*, 902(2):129, October 2020.
 - [24] Ilya Loshchilov and Frank Hutter. Fixing weight decay regularization in adam, 2018.
 - [25] Leslie N. Smith and Nicholay Topin. Super-Convergence: Very Fast Training of Neural Networks Using Large Learning Rates. *arXiv e-prints*, page arXiv:1708.07120, August 2017.