Learnable wavelet neural networks for cosmological inference

Christian Pedersen [1 2]   **Michael Eickenberg** [3]   **Shirley Ho** [1 2]

Abstract

Convolutional neural networks (CNNs) have been shown to both extract more information than the traditional two-point statistics from cosmological fields, and marginalise over astrophysical effects extremely well. However, CNNs require large amounts of training data, which is potentially problematic in the domain of expensive cosmological simulations, and it is difficult to interpret the network. In this work we apply the learnable scattering transform, a kind of convolutional neural network that uses trainable wavelets as filters, to the problem of cosmological inference and marginalisation over astrophysical effects. We present two models based on the scattering transform, one constructed for performance, and one constructed for interpretability, and perform a comparison with a CNN. We find that scattering architectures are able to outperform a CNN, significantly in the case of small training data samples. Additionally we present a lightweight scattering network that is highly interpretable.

1. Introduction

The process of extracting information from cosmological fields is a fundamental component of modern cosmology. The early Universe, observed as the Cosmic Microwave Background, is well described by a Gaussian random field, meaning the two-point function (or the power spectrum) contains all relevant information. However the growth of non-linear structure means this is not the case for probes of the late-time Universe, and it has been shown that significant information lies outside the two-point statistics in this regime. In the advent of several upcoming large surveys of late-time structure growth such as Euclid (Laureijs et al., 2011), DESI (DESI Collaboration, 2016), Roman ob-

servatory (Spergel et al., 2015), Rubin observatory (LSST Science Collaboration, 2009), it is becoming increasingly important to consider methods of cosmological inference that go beyond the power spectrum, in order to maximise the scientific yield from these datasets.

The situation is further complicated by the fact that on small scales, astrophysical phenomena such as supernovae and AGN feedback affect the clustering of observational tracers. Recent work has demonstrated that neural networks are able to optimally marginalise over these effects (Villaescusa-Navarro et al., 2020), and that convolutional neural networks (CNNs) are able to extract significantly more information from cosmological fields than the power spectrum (**?**Gupta et al., 2018; Villaescusa-Navarro et al., 2021b;a; Lu et al., 2022) . However CNNs suffer from two potential pitfalls. First, the large numbers of parameters involved in deep convolutional networks require large amounts of training data to optimise. This is a significant consideration in the context of late-universe cosmology, where the hydrodynamical simulations required to model the non-linear baryon density field are extremely computationally expensive. Second, CNNs also suffer from a lack of interpretibility.

In this work we attempt to address these issues by presenting two models for cosmological inference based on the scattering transform (Mallat, 2011). Scattering transforms use a cascade of analytic wavelet transforms followed by complex modulus nonlinearities to construct descriptors of the input signal that behave smoothly to small deformations and that can be made to exhibit known symmetries of the data. Scattering transforms and the related wavelet phase harmonics have been successfully applied to cosmological parameter estimation from different types of input fields (Cheng et al., 2020; Allys et al., 2020). Recent work has demonstrated that learning the filter parameters can provide performance gains in the regime of small datasets (Gauthier et al., 2021). Using the suite of CAMELs simulations (Villaescusa-Navarro et al., 2021a), we investigate two kinds of scattering networks, one designed for performance and one designed for interpretability, and perform a comparison with a traditional CNN.

---

[*]Equal contribution [1]Department of Physics, New York University, New York City, USA [2]Centre for Computational Astrophysics, Flatiron Institute, New York City, USA [3]Centre for Computational Mathematics, Flatiron Institute, New York City, USA. Correspondence to: Christian Pedersen <c.pedersen@nyu.edu>.
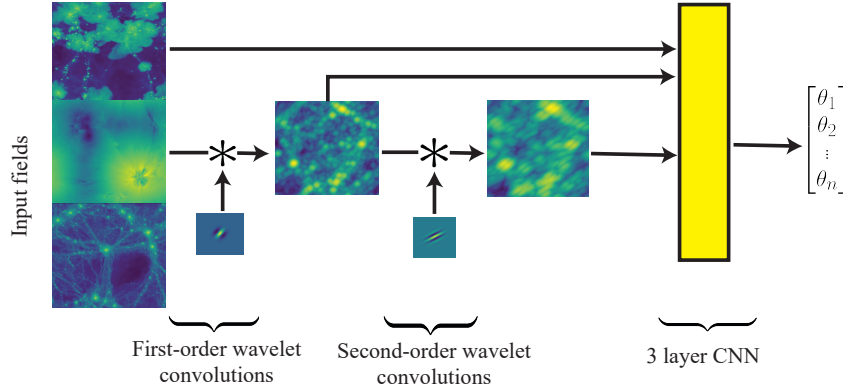
Figure 1. Architecture for the *scattering network* (SN). The input fields are convolved with a set of 8 first order wavelet filters. These convolved fields are then convolved with a further set of 8 more second order filters. During each wavelet convolution the field is downsampled by a factor of 2. We include residual connections, so the zeroth and first order fields are smoothed and downsampled to match the resolution of the second order fields. These are then concatenated and passed to a 3 layer CNN, which provides the model output. The 4 parameters describing each wavelet filter are included in the backprop during training.

2. Method

2.1. Simulations

We use the publicly available CAMELs simulations (Villaescusa-Navarro et al., 2021a). Whilst we refer the reader to (Villaescusa-Navarro et al., 2021a) for a full description, we briefly overview the most relevant aspects. We restrict our analysis to the subset of CAMELs simulations that use the IllustrisTNG simulation code (Weinberger et al., 2017; Pillepich et al., 2018). The dataset includes a suite of 1,000 simulations varying 6 parameters: Ω_m and σ_8 to describe cosmology, and A_{SN1}, A_{SN2}, A_{AGN1}, A_{AGN2} that control the efficiency of the supernova and AGN feedback. Each simulation produces 13 different fields describing various properties such as gas temperature, total matter density, and magnetic fields. We use projected 2D slices of the 3D $25 \times 25 \ h^{-3} \ \text{Mpc}^3$ simulation boxes. Each slice is $5 \ h^{-1}$ Mpc thick, so with 5 slices along each axis, this produces a total set of 15,000 maps available for each field, each with a resolution of 256×256 pixels. These maps are then divided into training, test and validation sets at a fractional split of 0.9, 0.05 and 0.05 respectively.

2.2. Models

We investigate three models. As our baseline CNN, we use a model with 12 convolutional layers, each with batch normalisation and Leaky ReLu activations. Following (Villaescusa-Navarro et al., 2022) we train moment neural networks presented in (Jeffrey & Wandelt, 2020) to infer the mean (μ_i) and variance (σ_i) of the posterior for the 6 parameters ($\theta_i = \{\Omega_{m,i}, \sigma_{8,i}, A_{\text{SN1},i}, A_{\text{SN2},i}, A_{\text{SN1},i}, A_{\text{AGN1},i}, A_{\text{AGN2},i}\}$)

describing each simulation. Therefore we define our loss function:

$$
\begin{aligned}
\mathcal{L} &= \sum_{i=1}^{6} \log \left(\sum_{j \in \text{batch}} (\theta_{i,j} - \mu_{i,j})^2 \right) \\
&+ \sum_{i=1}^{6} \log \left(\sum_{j \in \text{batch}} \left((\theta_{i,j} - \mu_{i,j})^2 - \sigma_{i,j}^2 \right)^2 \right) \quad (1)
\end{aligned}
$$

We also consider two models based on the scattering transform. First, we introduce the *scattering network* (SN), shown in figure 1. The first two layers are composed of convolutions with banks of 8 wavelet filters. We use Morlet wavelets, defined as

$$
\psi_{\sigma,\phi,\xi,\gamma}(u) = e^{-\|D_\gamma R_\phi(u)\|^2/(2\sigma^2)} \left(e^{i\xi u'} - \beta \right), \quad (2)
$$

where β is a normalisation constant to ensure that the wavelet integrates to 0 over the spatial domain, $u' = u_1 \cos \phi + u_2 \sin \phi$, R_ϕ is the rotation matrix of angle ϕ and $D_\gamma = \begin{pmatrix} 1 & 0 \\ 0 & \gamma \end{pmatrix}$. We therefore consider 4 parameters that modify the Morlet wavelet: the orientation is controlled by ϕ, the spatial frequency by ξ, the Gaussian envelope is determined by σ, and the aspect ratio is set by γ. At initialisation, we randomly sample these parameters from $\phi \sim U[0, 2\pi]$, $\xi \sim U[0.5, 1]$, $\sigma \sim \log(U[\exp 1, \exp 5])$, and $\gamma \sim U[0.5, 1.5]$. These filter parameters are modified by gradient descent as the network trains, and so are optimised to extract information from the fields they are operating on. Unlike in (Gauthier et al., 2021), we use different wavelet filters for first- and second-order scattering, allowing the network more flexibility to optimise these parameters.
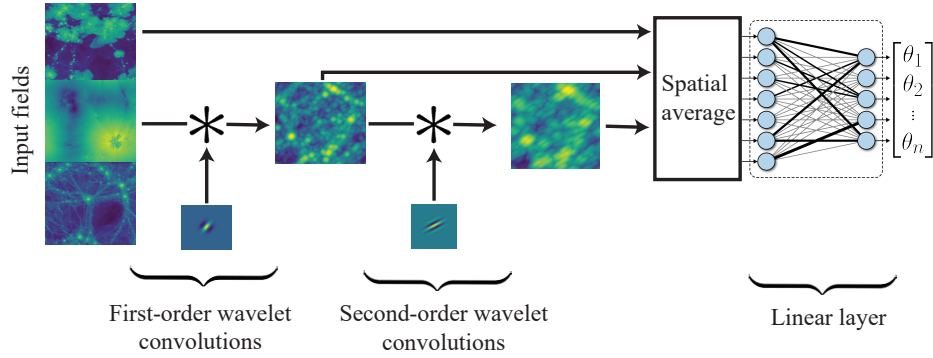
Figure 2. Model architecture for the *interpretable network* (IN). The first two layers are identical to fig 1, except now the output fields are spatially averaged, providing a single vector of 73 values. A single linear layer is used on this vector to predict the model output.

At each convolution, the input field is smoothed and down-sampled by a factor of 2, so the second-order convolved fields have a resolution of 64×64. We include residual connections, so the zeroth and first-order fields are smoothed and downsampled to match this resolution and concatenated with the second order fields. This produces a total of $1 + 8 + 64 = 73$ fields output by the scattering layers. Finally these fields are passed to a shallow CNN with 3 convolutional layers, each with batch normalisation and ReLu activations. Finally we use one fully connected layer generating the output. We train this network on the same set of 11 fields as the "MultiField" (MF) results from (Villaescusa-Navarro et al., 2021a), which includes all simulated fields except the dark matter mass and velocity fields.

Second, we also introduce an extremely lightweight model, the "Interpretable network" (IN) shown in figure 2. The first two layers of this network are identical to the SN, except now the 73 output fields of the scattering layers are spatially averaged to form a single vector of 73 numbers. We then use a single linear layer on this vector to obtain parameter values. For the sake of simplicity in interpreting this network, we train this model on only the cold dark matter mass field (M_{cdm}) to produce the results in section 3.2.

For all models, we use batch gradient descent with Adam optimiser and a batch size of 32. Hyperparameters are optimised using the `optuna` package using 30 trials. We vary independently the learning rate for the scattering and neural layers, and in the case of the SN and CNN, we vary the number of convolutional filters.

3. Results

3.1. Scattering network

We show a comparison of model performance between the baseline CNN and SN in table 1. We focus on the accuracy

MODEL	$\Delta\Omega_m(\%)$	$\Delta\sigma_8\,(\%)$	VALID. LOSS
CNN (10K)	6.31	4.02	-11.90
SN (10K)	4.24	3.42	-12.18
CNN (5K)	6.61	5.15	-11.18
SN (5K)	5.29	4.15	-11.58
CNN (1K)	10.07	8.99	-9.82
SN (1K)	6.32	4.33	-10.62
SN (1K M_{cdm})	2.48	1.84	-3.83
IN (1K M_{cdm})	11.50	2.79	-3.33

Table 1. Comparison of model performance for datasets of sizes 10,000, 5,000 and 1000 simulated maps. We show results for the empirical accuracy on Ω_m and σ_8 averaging over the test set, and the validation loss. For the final two rows, we use only the M_{cdm} fields. For simplicity, for these last examples we only model the mean of the posterior, i.e. the loss function is only the first term in equation 1, which is the reason for the large difference in stated values.

of the predicted values for Ω_m and σ_8, when averaging over the test set. We also show the validation loss as evaluated by equation 1. The two models are compared for 3 different dataset sizes: $10,000$, $5,000$ and $1,000$, in order to evaluate how performance scales with the size of the training set. These numbers include test and validation sets. We find significantly better performance from the SN when using smaller datasets. This is consistent with a similar comparison performed in (Gauthier et al., 2021) on the CIFAR-10 dataset. The SN model has approximately an order of magnitude less parameters than the CNN, with $\sim 2 \times 10^6$ versus 2×10^7, and is therefore able to optimise effectively on smaller datasets. In addition, our results indicate that even as we scale to larger datasets, the SN performance mildly outperforms the CNN.
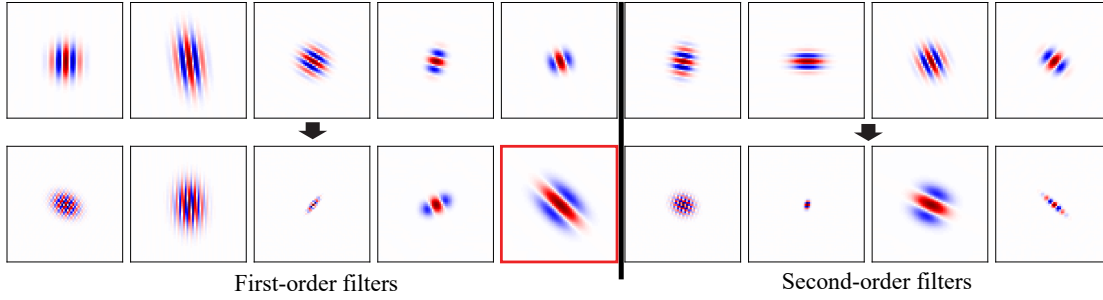
Figure 3. Wavelet filters before (top) and after (bottom) training. We show only the 9 most relevant fields after applying L_1-penalised linear regression to the IN described in figure 2. We find that these 9 filters plus the zeroth order filters are able to retain 98% of the model performance. All second order fields selected by the network were the product of a single first order wavelet, highlighted in red. For visualisation purposes we have magnified the spatial size of the wavelets by a factor of 4 in this figure.
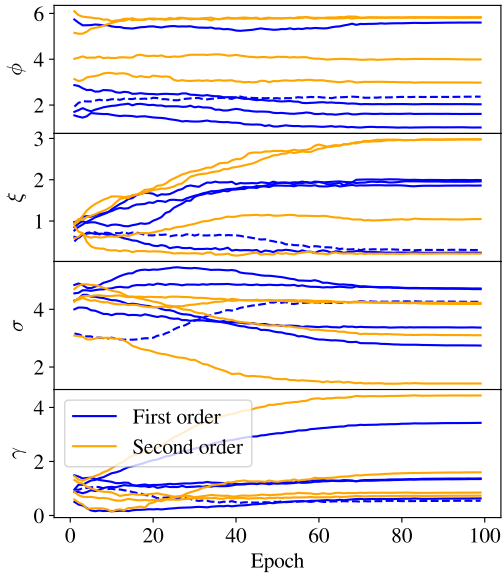

Figure 4. Evolution of the wavelet parameters describing each filter shown in figure 3 during training. First order wavelets are shown in blue and second order wavelets are shown in orange. The first order filter used to propagate fields to the second order filters is shown in blue dashed lines.

3.2. Interpretable network

At the bottom of table 1 we compare the performance of the SN and IN when using a set of $1,000$ maps of the M_{cdm} field. First, we find that the SN is able to predict Ω_m and σ_8 to a remarkable degree of accuracy when using only $1,000$ maps. This is potentially due to the fact that when operating on only one type of input field, the wavelet parameters are able to better extract information from this field. Whilst the difference in performance on Ω_m is very large, the IN is

able to predict σ_8 to 2.79% accuracy despite having only 654 parameters.

Finally, after training the filters for the IN, we use L_1-penalised linear regression to isolate the most significant filters in the model. We find that using only 10 of the 73 elements of the final linear layer, we are able to retain 98% of the model performance of the IN. These 10 fields correspond to the zeroth order field, 5 first-order wavelets, and 4 second-order wavelets. We show these 9 wavelets before and after training in figure 3. Interestingly, all 4 of the second-order fields are downstream from the same first-order wavelet, which is highlighted in red. It is possible that the network has chosen to optimise this filter to operate in conjunction with the second order fields.

In figure 4, we show the evolution of the parameters describing these 9 wavelet filters during training of the IN. Whilst the orientations of the wavelets do not change significantly, the frequency and aspect ratios are clearly strongly driven by the input fields to different values than they were initialised with.

4. Conclusion

We performed a comparison between a learnable scattering transform and a CNN for the purposes of inferring cosmological parameters and marginalising over astrophysical effects. We find a scattering network is able to outperform a CNN, with the performance difference significantly increasing for smaller training set sizes. We also present a lightweight interpretable scattering network that is able to find a sparse wavelet compression of the input fields.

References

Allys, E., Marchand, T., Cardoso, J.-F., Villaescusa-Navarro, F., Ho, S., and Mallat, S. New interpretable statistics for large-scale structure analysis and generation. *Physical Review D*, 102(10), nov 2020. doi: 10.1103/physrevd.102.103506. URL https://doi.org/10.1103%2Fphysrevd.102.103506.

Cheng, S., Ting, Y.-S., Mé nard, B., and Bruna, J. A new approach to observational cosmology using the scattering transform. *Monthly Notices of the Royal Astronomical Society*, 499(4):5902–5914, oct 2020. doi: 10.1093/mnras/staa3165. URL https://doi.org/10.1093%2Fmnras%2Fstaa3165.

DESI Collaboration. The DESI Experiment Part I: Science,Targeting, and Survey Design. *arXiv e-prints*, art. arXiv:1611.00036, October 2016.

Gauthier, S., Thérien, B., Alsène-Racicot, L., Chaudhary, M., Rish, I., Belilovsky, E., Eickenberg, M., and Wolf, G. Parametric Scattering Networks. *arXiv e-prints*, art. arXiv:2107.09539, July 2021.

Gupta, A., Zorrilla Matilla, J. M., Hsu, D., and Haiman, Z. Non-Gaussian information from weak lensing data via deep learning. , 97(10):103515, May 2018. doi: 10.1103/PhysRevD.97.103515.

Jeffrey, N. and Wandelt, B. D. Solving high-dimensional parameter inference: marginal posterior densities & Moment Networks. *arXiv e-prints*, art. arXiv:2011.05991, November 2020.

Laureijs et al. Euclid Definition Study Report. *arXiv e-prints*, art. arXiv:1110.3193, October 2011.

LSST Science Collaboration. LSST Science Book, Version 2.0. *arXiv e-prints*, art. arXiv:0912.0201, December 2009.

Lu, T., Haiman, Z., and Zorrilla Matilla, J. M. Simultaneously constraining cosmology and baryonic physics via deep learning from weak lensing. , 511(1):1518–1528, March 2022. doi: 10.1093/mnras/stac161.

Mallat, S. Group Invariant Scattering. *arXiv e-prints*, art. arXiv:1101.2286, January 2011.

Pillepich, A., Springel, V., Nelson, D., Genel, S., Naiman, J., Pakmor, R., Hernquist, L., Torrey, P., Vogelsberger, M., Weinberger, R., and Marinacci, F. Simulating galaxy formation with the IllustrisTNG model. , 473(3):4077–4106, January 2018. doi: 10.1093/mnras/stx2656.

Spergel et al. Wide-Field InfrarRed Survey Telescope-Astrophysics Focused Telescope Assets WFIRST-AFTA 2015 Report. *arXiv e-prints*, art. arXiv:1503.03757, March 2015.

Villaescusa-Navarro, F., Wandelt, B. D., Anglés-Alcázar, D., Genel, S., Zorrilla Mantilla, J. M., Ho, S., and Spergel, D. N. Neural networks as optimal estimators to marginalize over baryonic effects. *arXiv e-prints*, art. arXiv:2011.05992, November 2020.

Villaescusa-Navarro, F., Anglés-Alcázar, D., Genel, S., Spergel, D. N., Li, Y., Wandelt, B., Nicola, A., Thiele, L., Hassan, S., Zorrilla Matilla, J. M., Narayanan, D., Dave, R., and Vogelsberger, M. Multifield Cosmology with Artificial Intelligence. *arXiv e-prints*, art. arXiv:2109.09747, September 2021a.

Villaescusa-Navarro, F., Genel, S., Angles-Alcazar, D., Spergel, D. N., Li, Y., Wandelt, B., Thiele, L., Nicola, A., Zorrilla Matilla, J. M., Shao, H., Hassan, S., Narayanan, D., Dave, R., and Vogelsberger, M. Robust marginalization of baryonic effects for cosmological inference at the field level. *arXiv e-prints*, art. arXiv:2109.10360, September 2021b.

Villaescusa-Navarro, F., Genel, S., Anglés-Alcázar, D., Thiele, L., Dave, R., Narayanan, D., Nicola, A., Li, Y., Villanueva-Domingo, P., Wandelt, B., Spergel, D. N., Somerville, R. S., Zorrilla Matilla, J. M., Mohammad, F. G., Hassan, S., Shao, H., Wadekar, D., Eickenberg, M., Wong, K. W. K., Contardo, G., Jo, Y., Moser, E., Lau, E. T., Machado Poletti Valle, L. F., Perez, L. A., Nagai, D., Battaglia, N., and Vogelsberger, M. The CAMELS Multifield Data Set: Learning the Universe's Fundamental Parameters with Artificial Intelligence. , 259(2):61, April 2022. doi: 10.3847/1538-4365/ac5ab0.

Weinberger, R., Springel, V., Hernquist, L., Pillepich, A., Marinacci, F., Pakmor, R., Nelson, D., Genel, S., Vogelsberger, M., Naiman, J., and Torrey, P. Simulating galaxy formation with black hole driven thermal and kinetic feedback. , 465(3):3291–3308, March 2017. doi: 10.1093/mnras/stw2944.