Deep learning dark matter map reconstructions from DES SV weak lensing data

Niall Jeffrey,[1]⋆ François Lanusse[2], Ofer Lahav[1], Jean-Luc Starck[3]

[1] *Department of Physics & Astronomy, University College London, Gower Street, London, WC1E 6BT, UK*

[2] *Berkeley Center for Cosmological Physics and Department of Physics, University of California, Berkeley, CA 94720*

[3] *AIM, CEA, CNRS, Université Paris-Saclay, Université Paris Diderot, Sorbonne Paris Cité, F-91191 Gif-sur-Yvette, France*

ABSTRACT

We present the first reconstruction of dark matter maps from weak lensing observational data using deep learning. We train a convolution neural network (CNN) with a Unet based architecture on over 3.6×10^5 simulated data realisations with non-Gaussian shape noise and with cosmological parameters varying over a broad prior distribution. We interpret our newly created DES SV map as an approximation of the posterior mean $P(\kappa|\gamma)$ of the convergence given observed shear. Our DeepMass† method is substantially more accurate than existing mass-mapping methods. With a validation set of 8000 simulated DES SV data realisations, compared to Wiener filtering with a fixed power spectrum, the DeepMass method improved the mean-square-error (MSE) by 11 per cent. With N-body simulated MICE mock data, we show that Wiener filtering with the optimal known power spectrum still gives a worse MSE than our generalised method with no input cosmological parameters; we show that the improvement is driven by the non-linear structures in the convergence. With higher galaxy density in future weak lensing data unveiling more non-linear scales, it is likely that deep learning will be a leading approach for mass mapping with Euclid and LSST.

Key words: gravitational lensing: weak – large-scale structure of Universe– methods: statistical

1 INTRODUCTION

The evolving cosmological density field is rich in information about the cosmological model of the Universe, its unknown parameters, and cosmic web-dependent astrophysics. Though the largest fraction of the density is invisible dark matter, the gravitational lensing effect of galaxies can be used to infer fluctuations in the total foreground matter distribution. Accurate "mass maps" will be essential for the science goals of the upcoming LSST survey and the ESA Euclid mission.

The maps considered in this paper are of the two-dimensional convergence, κ, a weighted projection of the matter density field in the foreground of the observed galaxies. Recovering the convergence from the measured galaxy shapes, known as observed shear $\gamma_{\rm obs}$ in the weak lensing regime, is an ill-posed inverse problem, troubled by survey masks (missing data) and galaxy "shape noise".

A typical principled approach to reconstructing more

accurate mass maps in the presence of noisy, masked shear data is to use physically motivated priors. In Jeffrey et al. (2018b) it was shown that using either Gaussian priors or "halo model" sparsity priors for κ improved the accuracy of the reconstructions with Dark Energy Survey Science Verification (DES SV) data. Implemented methods include using log-normal (Böhm et al. 2017) priors or E-mode priors (Mawdsley et al. 2019).

However, all of these priors take functional forms that only approximate the true object of interest, the prior on the convergence field $P(\kappa|\mathcal{M})$ (with model assumptions $\mathcal{M}$). These approximations are necessary because we cannot represent the probability distribution of the non-linear density field in closed form. For example, we cannot characterise it uniquely in terms of its moments (Carron & Szapudi 2017). Even if the true, unapproximated prior were available, evaluation via direct calculation would likely be intractable.

Fortunately we can still draw realisations of convergence maps from the prior distribution $P(\kappa)$ in the form of simulations, which provides opportunity to a new generation of methods based on deep learning. Such an approach has been simultaneously proposed by Shirasaki et al. (2018), where a

---

⋆ E-mail: niall.jeffrey.15@ucl.ac.uk
† github.com/NiallJeffrey/DeepMass

conditional adversarial network was used to learn a mapping from noisy convergence maps to an estimate of the noise-free convergence.

In this work, we propose a deep learning method to estimate the posterior mean of the convergence map from observed weak lensing shear measurements. In section 4 we demonstrate our method on simulations and DES SV data.

2 WEAK GRAVITATIONAL LENSING

2.1 Shear and convergence

Given a distribution of source galaxies n in radial comoving distance ω, the convergence at position $\vec{\phi}$ on the sky is given by a weighted integral of the density

$$\kappa(\vec{\phi}) = \frac{3H_0^2\Omega_m}{2}\int_0^\infty\Big[\int_0^\omega \mathrm{d}\omega'\frac{\omega'(\omega-\omega')}{\omega}\frac{\delta(\vec{\phi},\omega')}{a(\omega')}\Big]n(\omega)\mathrm{d}\omega\ , \tag{1}$$

where H_0 is the present value of the Hubble parameter, a is the cosmological scale factor, Ω_m is the matter density parameter, and δ is the overdensity.

We express the linear data model in matrix notation,

$$\gamma = \mathbf{A}\kappa + \mathbf{n}\ , \tag{2}$$

where $\mathbf{n}$ is a vector of noise per pixel. The matrix operator acting on the convergence $\mathbf{A}\kappa$ is the shear contribution due to lensing (Bartelmann & Schneider 2001). In this formulation, the elements γ are the complex shear measurements binned into angular pixels in a two-dimensional image format.

We do not take into account the second order effects of reduced shear (Schneider & Seitz 1995), flexion (Bacon et al. 2006) or intrinsic alignments (Kirk et al. 2015). However, the deep learning approach taken in this paper is extremely flexible; as long as an effect can be modelled and included in the training data, it will be taken into account in the mass map reconstruction. This is not generally true of other methods. For example, flexion requires reformulations of methods (e.g. Lanusse et al. 2016). Additionally, noise per pixel is invariably approximated as Gaussian, which we do not assume in our deep learning approach.

2.2 Previous mapping approaches

The original mass mapping approach by Kaiser & Squires (1993) was a direct deconvolution. In practice Kaiser-Squires (KS) inverts the matrix $\mathbf{A}$ in Fourier space, where the matrix is diagonal. As this deconvolution is across a finite space, the edges of the data and internal masks introduce artefacts. KS is further troubled by the noise term in equation 2, which it does not take into account.

In a Bayesian framework we may wish to consider the posterior distribution of the convergence κ conditional on the observed shear γ

$$P(\kappa|\gamma,\mathcal{M}) = \frac{P(\gamma|\kappa,\mathcal{M})\ P(\kappa|\mathcal{M})}{P(\gamma|\mathcal{M})}\ , \tag{3}$$

The denominator $P(\gamma)$ is a Bayesian evidence term conditional on model $\mathcal{M}$. The first factor of the numerator is the

likelihood $P(\gamma|\kappa,\mathcal{M})$, which encodes our noise model. The second term is the prior $P(\kappa|\mathcal{M})$, a possible selection of which was discussed in section 1.

If we believe a realisation of the convergence κ is a realisation of Gaussian random field, then the form of $P(\kappa)$ would be Gaussian. If the noise per pixel is Gaussian then the likelihood is also Gaussian, which results in a posterior distribution with both the mean and maximum given by the Wiener filter:

$$\hat{\kappa}_\mathrm{w} = \mathbf{W}\gamma = \mathbf{S}_\kappa\mathbf{A}^\dagger\big[\mathbf{A}\mathbf{S}_\kappa\mathbf{A}^\dagger + \mathbf{N}\big]^{-1}\ \gamma\ , \tag{4}$$

where $\mathbf{S}_\kappa = \langle\kappa\kappa^\dagger\rangle$ and $\mathbf{N} = \langle\mathbf{n}\mathbf{n}^\dagger\rangle$ are the signal and noise covariance matrices respectively (Wiener 1949, Zaroubi et al. 1995, Jeffrey et al. 2018a). The signal covariance in harmonic space is diagonal for isotropic fields. On the sphere, its elements are given by the κ power spectrum, $C_\kappa(\ell)$.

This Gaussian distribution is only approximately true for large scales where Gaussianity persists from the early Universe. On smaller scales, non-Gaussianity grows due to non-linear structure formation, which results in the cosmic web of the late Universe.

3 DEEP LEARNING MAPS

3.1 Convolution neural networks

We take a standard deep learning approach. We seek an approximation $\mathcal{F}_\Theta$ to the function that maps the pixelised shear to the convergence map

$$\hat{\kappa} = \mathcal{F}_\Theta(\gamma)\ , \tag{5}$$

where the parameters of the function Θ are to be learned (Goodfellow et al. 2016). We learn these parameters by minimising a mean-square-error (MSE) cost function

$$J(\Theta) = ||\mathcal{F}_\Theta(\gamma) - \kappa_\mathrm{true}||_2^2\ , \tag{6}$$

evaluated on a set of training data which consists of pairs of realistic shear and "truth" (noise-free) convergence maps. If the training data "truth" maps are drawn from a prior distribution $P(\kappa)$, and the corresponding noisy shear map is drawn from the likelihood $P(\gamma|\kappa)$, this MSE cost function corresponds to $\mathcal{F}_\Theta(\gamma)$ being a mean[1] posterior estimate (Jaynes 2003), such that $\hat{\kappa}$ is approximating:

$$\hat{\kappa} = \mathcal{F}_\Theta(\gamma) = \int \kappa\ P(\kappa|\gamma)\ \mathrm{d}\kappa\ . \tag{7}$$

We use a deep convolution neural network (CNN) to approximate the function $\mathcal{F}_\Theta$, where the parameters Θ are primarily elements of learned filters in convolutional layers. CNNs are particularly suited for two-dimensional image or one-dimensional time series data with translation invariant features in the underlying signal.

The CNN is a series of iteratively computed layers. At a given layer j the signal $\mathbf{x}_j$ is computed from the previous layer $\mathbf{x}_j = \rho\mathbf{M}_j\mathbf{x}_{j-1}$, with linear operator (e.g. convolution) $\mathbf{M}_j$ and nonlinear *activation* function ρ (LeCun et al.

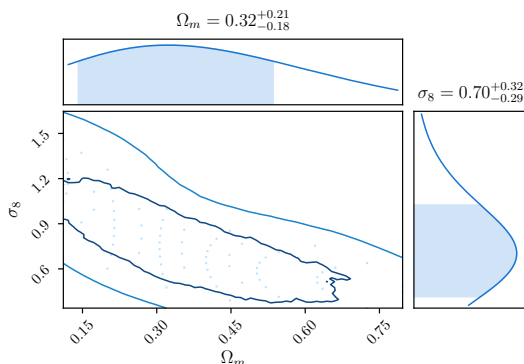[1] The mean posterior is not generally the *maximum a posteriori*

Figure 1. Prior range of cosmological parameters Ω_m and σ_8 of the training data. Simulations were run at the marked points.

1990, Mallat 2016). The output of a layer is sometimes called a *feature map*.

Due to their additional layers, deep architectures are often able to learn features with greater complexity than shallow architectures and therefore can better approximate the target function. For a general overview of deep learning and neural networks we recommend Goodfellow et al. (2016).

3.2 DeepMass architecture

Our DeepMass architecture is based on the Unet (Ronneberger et al. 2015), which has a so-called *expanding path* and *contracting path*. The DeepMass contracting path differs from the original Unet: usually convolutions and activation are followed by a *max pooling* operation to downsample the images, whereas we use *average pooling* (Géron 2017). With each downsampling operation, the images decrease in resolution, but the 3×3 filters cover more angular size of the image. The convolution after a pooling operation therefore has a receptive field that covers larger physical features in the convergence κ map.

There are similarities between Unet architectures and sparse recovery methods. These consider representations where the solution is sparse and employ transforms which are fixed (e.g. Fourier, wavelets) or learned from data, and optimisation is solved using proximal theory (Starck et al. 2015). The Unet expanding and contracting path are very similar to synthesis and analysis concepts in sparse representations. This has motivated the use of wavelets to implement the Unet average pooling and the expanding path (Ye et al. 2018; Han & Ye 2018). There are nevertheless significant differences: Unets can learn rich sets of features (corresponding to sparse dictionaries) from large training data sets, and the CNN implementation of non-linearity.

We differ from the original Unet by not using padding in the convolutional layers, as the edge of our data mask is already many pixels away from the edge of the square image. This choice means that output of a convolution has the same image dimensions as the input.

The full architecture and code can be seen online: DeepMass[†]. We have added *Batch Normalization* layers (Ioffe & Szegedy 2015) after each convolutional layer; without this, training often became stuck in local minima of the cost function with respect to the parameters Θ. For all layers, except for the final, we use the rectified linear unit (ReLU) activation. In the final layer we use a sigmoid func-

tion, which forces the output to be between 0 and 1 (inputs and outputs are correspondingly rescaled).

For simplicity and memory efficiency, we aimed to work with real (32-bit) numbers, thus necessitating an initial operation acting on the complex shear γ. The best results came from using a fixed Wiener filter operation before the first convolution (rather than KS, as might be expected). This is equivalent to the first layer having $\mathbf{M}_{j=0} = \mathbf{W}$ and $\rho = 1$, with no free parameters. We could also interpret the Unet after the initial Wiener operation as $\mathcal{G}_\Theta$ where $\mathcal{F}_\Theta(\gamma) = \mathcal{G}_\Theta(W(\gamma))$. The Wiener filter used a power spectrum with cosmological parameters σ_8 and Ω_m fixed at the mean of the marginal posterior distributions from DES Y1 analysis (Abbott et al. 2018). The flat sky power spectrum was an average of 102 power spectra of projected patches.

3.3 Training data

3.3.1 L-PICOLA *simulations*

The training data is derived from 74 independent dark matter simulations, with each simulation covering an octant of the sky. The simulations used a standard flat ΛCDM cosmological model with $H_0 = 70$ km Mpc^{-1}s^{-1}. The scalar spectral index and baryon density were fixed at $n_s = 0.95$ and $\Omega_b = 0.044$ respectively. The values of Ω_m and the amplitude parameter σ_8 are distributed on a non-Euclidean grid with distances between points giving a density according to our prior $P(\sigma_8, \Omega_m)$ as shown in figure 1. Weak lensing constraints are most sensitive to combinations of this pair of parameters, so we avoid overfitting to a single cosmology by varying them in the training data.

To generate a convergence map from a simulation, the matter particles were binned using the HEALPix (Górski et al. 2005) pixelisation of the sphere with NSIDE=2048 in comoving radial shells of 50 Mpc/h. The density ρ map in a given redshift was converted into an overdensity $\delta = \rho/\bar{\rho} - 1$ using the average density in the shell $\bar{\rho}$. The convergence was calculated per pixel using equation 1. We wish to have the $n(z)$ in the lensing kernel match the DES SV data (section 4.1), which we approximate by summing the individual posterior redshift distributions per galaxy from the BPZ photometric redshift code (Coe et al. 2006). The convergence maps were downgraded to NSIDE=1024.

The dark matter simulations are generated using the L-PICOLA code (Howlett et al. 2015), which is based on the COLA (Tassev et al. 2013) algorithm. This uses a combination of second-order Lagrangian perturbation theory (2LPT) and a Particle-Mesh (PM) which requires fewer time steps than "full" N-body (e.g. Gadget Springel 2005) and therefore can generate simulations more quickly. This allows more training data to be generated in a given amount of compute time.

We used a 1250 Mpc/h comoving simulation box, 768^3 particles, and a 1536^3 grid. A $z < 1.6$ lightcone was generated with up to four box replicates, using 30 time steps from $z = 20$. The initial conditions used Eisenstein & Hu 1999 for the linear matter power spectrum.

The drawback of this approach is the accuracy of the dark matter distribution. The finite spatial resolution and fewer timesteps used by the COLA method particularly affects small distance scales. Our experiments have shown a suppression of the L-PICOLA power spectrum at scales of
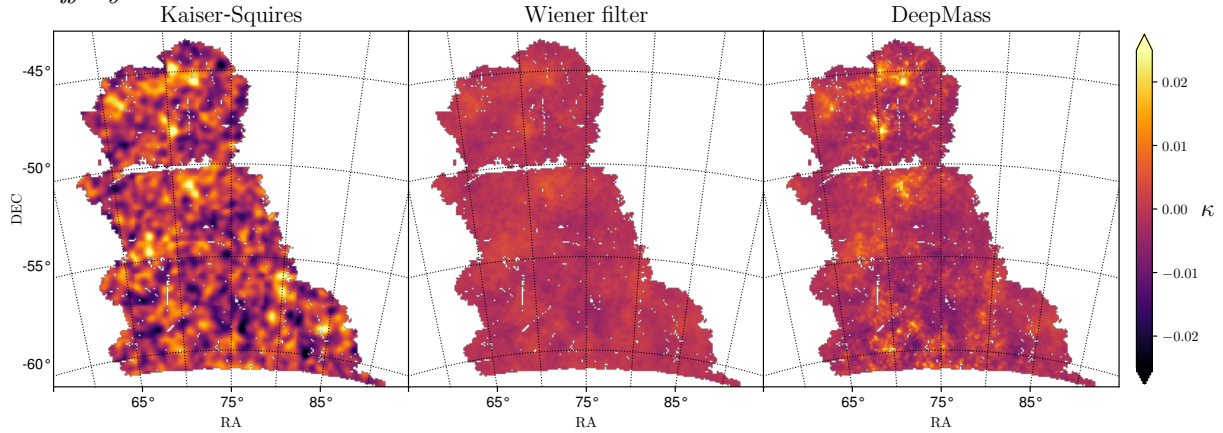
Figure 2. Convergence κ reconstruction from DES SV observational data with: KS, Wiener filtering, and DeepMass.
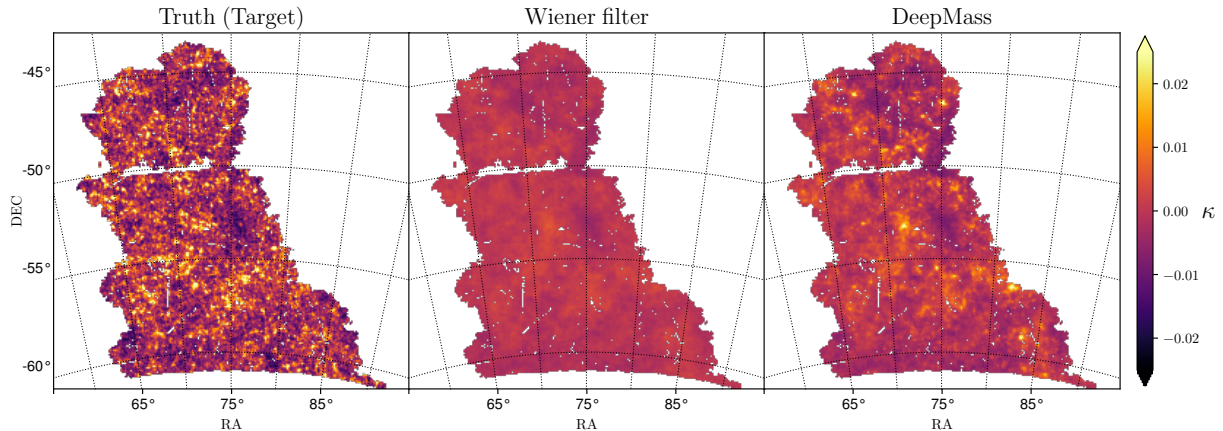

Figure 3. Example L-PICOLA validation simulation (*left*) and the corresponding Wiener (*centre*) and DeepMass (*right*) reconstructions.

$\ell > 700$ of order 10 per cent (relative to NICAEA[2] (Kilbinger et al. 2009) theory), as is expected with COLA methods. We correct the power of the L-PICOLA convergence by estimating the smooth part of the $C_\kappa(\ell)$ using a polynomial order 1 Savgol filter with window size 91 for each convergence map and reweighting spherical harmonics. Using the ratio of NICAEA and only the smooth part of the measured simulation power spectrum ensures that the natural fluctuations inherent in $C_\kappa(\ell)$ for a given realisation are preserved.

3.3.2 Training images

From the 74 independent HEALPix convergence maps over an octant of the sky, we generate 376,684 DES SV mock data realisations. A given realisation is generated from the HEALPix convergence map by randomly choosing a position on the sphere, applying a uniform random rotation between 0 and 360 deg, and extracting a square patch using a gnomonic projection with 256^2 pixels of size 4.5^2 arcmin^2. If the generated image has pixels outside the octant, it is rejected. The rotation step is not to make the reconstruction rotation invariant, which happens naturally as $P(\kappa)$ is isotropic by the cosmological principle, but it is to augment the training data and learn $\mathcal{F}_\Theta$ better.

From the projected square κ convergence map, the complex noise-free shear map is generated using the **A** matrix from equation 2. The mask is applied and a random shape noise map is added. The noise map is generated by randomly

shuffling the positions of galaxies in the original catalogue; this keeps the density of galaxies the same, but destroys the coherent lensing signal. This way we forward model the non-Gaussian noise inherent in the data (something that other methods do not do).

4   RESULTS

4.1 Dark Energy Survey SV data

DES is a ground based photometric galaxy survey, observing in the southern sky from the 4m Blanco telescope in Chile with five photometric filters (Flaugher et al. 2015). The SV (A1) data[3] come from an initial run of 139 deg^2, but with depth of approximately that of the full 6 year survey (Chang et al. 2015). We make a redshift cut of $0.6 < z_{\mathrm{mean}} < 1.2$, where z_{mean} is the mean of the z posterior per galaxy. Data selection choices match Jeffrey et al. (2018b), though some maps appear different due to changes in pixel size and flat-sky projection.

In figure 2 we apply KS, Wiener filtering, and the trained DeepMass CNN. Kaiser-Squires uses a 10 arcmin Gaussian smoothing as in Jeffrey et al. (2018b). The Wiener filtering uses a power spectrum with Ω_m and σ_8 at the mean of their respective marginal posterior distributions from the Year 1 DES cosmology result (Abbott et al. 2018). The

DeepMass CNN was trained using the Adam optimiser with a learning rate $= 1 \times 10^{-5}$ for 20 epochs (retraining over the full training set). The final Wiener and DeepMass maps were smoothed with a Gaussian of $\sigma = 2.25$ arc min (half pixel size) to remove very small scale artefacts arising from the HEALPix projection.

The DeepMass reconstruction clearly shows more non-linear structure than the Wiener filter. Individual peaks, which are suppressed by Wiener filtering, are resolved by DeepMass.

4.2 Validation on simulations

Out of the original generated training images (section 3.3), 8000 were kept and not used for training to be used as validation. One such example can be seen in figure 3, with the corresponding Wiener filter and DeepMass reconstructions. As with the reconstruction from observational data, DeepMass can be seen to recover the non-linear (cosmic-web) structure better than Wiener filtering. Compared to Wiener filtering, the MSE over all 8000 maps is improved using DeepMass by 11 per cent.

In Jeffrey et al. (2018b), using a "halo-model" sparsity prior did not outperform Wiener filtering in terms of MSE, so we expect DeepMass MSE to outperform GLIMPSE. However, MSE minimisation relates just to the posterior mean, so alternative metrics remain to be explored.

Using 18 non-overlapping mock DES SV data from the MICE (Fosalba et al. 2015) simulations we apply a Wiener filter with an optimal power spectrum calculated using the known cosmological parameters (not available in real data applications). Nevertheless, without using the known cosmological parameters as input, DeepMass still recovers maps with an average of 2 per cent better MSE.

Furthermore, the MICE cosmological parameters lead to a relatively low power and fewer high variance structures above the signal-to-noise level. The largest improvement over Wiener filtering comes when there are more non-linear (non-Gaussian) structures. With the same MICE simulations, restricting ourselves to pixels where the truth is greater than two standard deviations from the mean $\kappa > 2\sigma$, compared to Wiener filtering, DeepMass improves the MSE by 8 per cent. As is to be expected, therefore, DeepMass improves over Wiener filtering due to its ability to reconstruct the non-linear structures in the cosmological signal.

5 CONCLUSION

With DeepMass, we have presented a deep learning method to reconstruct convergence κ maps from shear measurements. With DES SV, we have shown the mass map reconstruction with deep learning from observational data.

By training with simulations over a broad prior distribution of cosmological parameters, we have a generalised method which needs no input cosmological parameters. This method has shown substantial improvement over Wiener filtering both qualitatively (by eye) and quantitatively (11 per cent MSE reduction on the validation data). The flexible approach also takes into account non-Gaussian noise in the weak lensing data. As our simulated training data are samples drawn from the prior $P(\kappa)$, the approach has a principled Bayesian interpretation, without the need for evaluation of closed-form priors.

The quality of the reconstruction with these initial experiments, and its flexibility, makes the deep learning approach a preeminent candidate for mass mapping with future weak lensing surveys.

REFERENCES

Abbott T. M. C., et al., 2018, Phys. Rev. D, 98, 043526
Bacon D. J., Goldberg D. M., Rowe B. T. P., Taylor A. N., 2006, MNRAS, 365, 414
Bartelmann M., Schneider P., 2001, Phys. Rep., 340, 291
Böhm V., Hilbert S., Greiner M., Enßlin T. A., 2017, Phys. Rev. D, 96, 123510
Carron J., Szapudi I., 2017, MNRAS, 469, 2855
Chang C., et al., 2015, Physical Review Letters, 115, 051301
Coe D., Benítez N., Sánchez S. F., Jee M., Bouwens R., Ford H., 2006, AJ, 132, 926
Eisenstein D. J., Hu W., 1999, ApJ, 511, 5
Flaugher B., et al., 2015, AJ, 150, 150
Fosalba P., Gaztañaga E., Castander F. J., Crocce M., 2015, MNRAS, 447, 1319
Géron A., 2017, Hands-on machine learning with Scikit-Learn and TensorFlow. O'Reilly Media
Goodfellow I., Bengio Y., Courville A., 2016, Deep Learning. MIT
Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759
Han Y., Ye J. C., 2018, IEEE transactions on medical imaging, 37, 1418
Howlett C., Manera M., Percival W. J., 2015, A&C, 12, 109
Ioffe S., Szegedy C., 2015, arXiv preprint arXiv:1502.03167
Jaynes E. T., 2003, Probability theory: the logic of science. CUP
Jeffrey N., Heavens A., Fortio P., 2018a, A&C, 25, 230
Jeffrey N., et al., 2018b, MNRAS, 479, 2871
Kaiser N., Squires G., 1993, ApJ, 404, 441
Kilbinger M., et al., 2009, A&A, 497, 677
Kirk D., et al., 2015, Space Sci. Rev., 193, 139
Lanusse F., Starck J.-L., Leonard A., Pires S., 2016, A&A, 591,A2
LeCun Y., Boser B. E., Denker J. S., Henderson D., Howard R. E., Hubbard W. E., Jackel L. D., 1990, in Advances in neural information processing systems. pp 396–404
Mallat S., 2016, Philosophical Transactions of the Royal Society of London Series A, 374, 20150203
Mawdsley B., et al., 2019, arXiv e-prints, p. arXiv:1905.12682
Ronneberger O., Fischer P., Brox T., 2015, CoRR, abs/1505.04597
Schneider P., Seitz C., 1995, A&A, 294, 411
Shirasaki M., Yoshida N., Ikeda S., 2018, arXiv e-prints, p. arXiv:1812.05781
Springel V., 2005, MNRAS, 364, 1105
Starck J.-L., Murtagh F., Fadili J., 2015, Sparse image and signal processing: Wavelets and related geometric multiscale analysis. CUP
Tassev S., Zaldarriaga M., Eisenstein D. J., 2013, J. Cosmology Astropart. Phys., 2013, 036
Wiener N., 1949, Extrapolation, interpolation, and smoothing of stationary time series. Vol. 7, MIT
Ye J. C., Han Y., Cha E., 2018, SIAM Journal on Imaging Sciences, 11, 991
Zaroubi S., Hoffman Y., Fisher K. B., Lahav O., 1995, ApJ, 449, 446