Weak lensing shear estimation beyond the shape-noise limit: a machine learning approach

Ofer M. Springer,[1,2]* Eran O. Ofek,[2] Yair Weiss[1] and Julian Merten[3,4]

[1]*School of Computer Science and Engineering, The Hebrew University of Jerusalem, Jerusalem, Israel*
[2]*Benoziyo Center for Astrophysics, Weizmann Institute of Science, 76100 Rehovot, Israel*
[3]*INAF-OAS Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, Via Gobetti 93/3, 40129, Bologna, Italy*
[4]*Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK*

ABSTRACT

Weak lensing shear estimation typically results in per galaxy statistical errors significantly larger than the sought after gravitational signal of only a few percent. These statistical errors are mostly a result of *shape-noise* — an estimation error due to the diverse (and a-priori unknown) morphology of individual background galaxies. These errors are inversely proportional to the limiting angular resolution at which localized objects, such as galaxy clusters, can be probed with weak lensing shear. In this work we report on our initial attempt to reduce statistical errors in weak lensing shear estimation using a machine learning approach — training a multi-layered convolutional neural network to directly estimate the shear given an observed background galaxy image. We train, calibrate and evaluate the performance and stability of our estimator using simulated galaxy images designed to mimic the distribution of *HST* observations of lensed background sources in the CLASH galaxy cluster survey. Using the trained estimator, we produce weak lensing shear maps of the cores of 20 galaxy clusters in the CLASH survey, demonstrating an RMS scatter reduced by approximately 26% when compared to maps produced with a commonly used shape estimator. This is equivalent to a survey speed enhancement of approximately 60%. However, given the non-transparent nature of the machine learning approach, this result requires further testing and validation. We provide python code to train and test this estimator on both simulated and real galaxy cluster observations. We also provide updated weak lensing catalogues for the 20 CLASH galaxy clusters studied.

Key words: galaxy clusters – weak gravitational lensing – shear estimation – machine learning

1 INTRODUCTION

In weak lensing, the small gravitational deflection of light in its path from the source to the observer enables the measurement of the intervening mass distribution. Prime examples in which weak lensing is applied are in the measurement of cosmic shear, where the large-scale structure is of interest, and in measurements of galaxy cluster lensing, where the mass distribution is comparably more localized (see Schneider 2006, for a review of weak lensing applications).

This weak deflection of light induces a small distortion that maps the source plane coordinates $\vec{\theta}'$ to the image plane coordinates $\vec{\theta}$. The distortion $\vec{\theta}'(\theta)$ may be locally approximated as an affine transformation with a Jacobian

$$\frac{\partial \vec{\theta}'}{\partial \vec{\theta}} = \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix} = (1 - \kappa) \begin{pmatrix} 1 - g_1 & -g_2 \\ -g_2 & 1 + g_1 \end{pmatrix}, \quad (1)$$

where κ is known as the *convergence* and parametrizes the uniform stretching of the image, γ_i are the two components of *shear* and $g_i = \gamma_i/(1 - \kappa)$ are the *reduced shear*. Weak lensing shear estimation is typically concerned with measuring the reduced shear in images of faint, high-redshift galaxies and with deducing the lensing mass distribution from these estimates (see Bartelmann & Schneider 2001, for a derivation of weak lensing formalism).

An ideal estimator of g_i is unable to access the pre-lens galaxy image and therefore has to rely on the (true) probability distribution $P(g_i|x)$ of the shear given galaxy image x. Here x represents galaxy image pixel responses as observed,

* E-mail: springer@phys.huji.ac.il (OMS)

including the effects of galaxy population morphology, lens distortion, point-spread-function (PSF), band throughputs and noise — only to name a few known observational factors.

To date, many shear estimation techniques have been developed and studied and these have been primarily focused on reducing various types of systematic errors (e.g. biases due to PSF correction, noise, or source population variation). For a recent review of shear estimation methods and the different sources of estimation bias see Mandelbaum et al. (2015) and Massey et al. (2012).

Less attention has been given to the issue of reducing the statistical errors of the shear estimator that result from the wide variation in galaxy shapes. In cluster lensing, the number of background sources is often very limited. The angular resolution of galaxy cluster weak lensing maps is typically limited by the density of available background sources (on which shear can be estimated) and such maps are often computed by locally averaging the shear estimates of only a few neighbouring galaxies at each point in the field. For a given set of background sources, reducing the statistical error (or RMS scatter) of the shear estimator would therefore directly translate to a linear enhancement of the angular resolution of the map (at a pre-selected level of map noise).

In galaxy cluster lensing studies, two shear estimation methods that have been widely used are the KSB method (Kaiser et al. 1994) for ground based observations and the RRG method (Rhodes et al. 2000) for space based observations. These estimators essentially measure the two component ellipticity e of the galaxy intensity profile from its weighted quadrupole moments, correcting for the weight profile and PSF in different ways. After calibrating a proper *shear susceptibility factor G* (see Rhodes et al. 2000; Leauthaud et al. 2007), the calibrated ellipticity $\hat{\epsilon} = e/G$ is used as an unbiased local estimate of the reduced shear $g = \langle \hat{\epsilon} \rangle$. Note that although different definitions of ellipticity exist in the literature, here $\hat{\epsilon}$ has been scaled to match the scale of the reduced shear g defined in equation 1. The intrinsic per-component scatter $\sigma_{\hat{\epsilon}}$ of $\hat{\epsilon}$ (at constant g), resulting from the natural diversity of galaxy intensity profiles (and not from various measurement errors), is known as *shape-noise* and is generally considered to be a lower-limit of the statistical error that quadrupole moment based shear estimators can reach. The actual value of the shape-noise $\sigma_{\hat{\epsilon}}$ is somewhat dependent on source galaxy population (see Leauthaud et al. 2007, where it is found to be ~0.25 and ~0.30 at the lower and upper magnitude cuts of a specific space-based survey, respectively).

In this work we aim to measure shear with a statistical error lower than $\sigma_{\hat{\epsilon}}$ by using additional morphological information, visible in the source galaxy stamp, beyond its quadrupole moments. The premise of our work is that the apparent structure of galaxies isn't fully described by their ellipticity parameters and that quadrupole-based estimators may therefore be suboptimal with respect to statistical error. Shear estimators that measure higher order intensity moments (above quadrupole) and attempt to model their joint statistics (see e.g. Refregier 2003; Bernstein et al. 2016) could also potentially utilize substructure to achieve this goal, although, to the best of our knowledge, these estimators have not yet demonstrated to do so.

The idea that there is additional information to be harnessed for weak lensing measurements beyond the imaged quadrupole moments has been suggested before. Additional observables such as radio polarization were used by Brown & Battye (2011) to constrain the intrinsic morphological orientation of galaxies. Resolved spectroscopic kinematic maps were used by Blain (2002); Morales (2006); Huff et al. (2013) and more recently de Burgh-Day et al. (2015), to reduce the shape noise in measurements of disk galaxies. Dimensionality reduction techniques of existing multi-band photometric data, when used in addition to quadrupole moment ellipticities, have also been shown to carry some additional information (Croft et al. 2017; Niemi et al. 2015). In contrast, in this work, we use only a single optical image as an observable and seek new image features that may be used to infer shear.

Our approach to this problem is based on a discriminative machine learning model — a multi-layered (or *deep*) *convolutional neural network* trained to directly estimate shear (and not ellipticity) from the observed galaxy stamps. This approach relies on the availability of a large dataset of simulated galaxy stamps, having known shear and simulated observational conditions matching those of the galaxy clusters under study. We stress that this requirement is both challenging and crucial for properly training and validating a shear estimator in such a machine learning approach.

This paper is structured as follows: In §2 we describe the *Hubble Space Telescope* (*HST*) observations, both of the survey in which we measure shear and of the survey from which we generate simulated training and calibration data for the estimator. In §3 we describe the steps taken to match the simulations to the galaxy cluster observations and to generate a larger and richer training dataset that would promote certain useful invariance properties of the learnt estimator. In §4 we describe the machine learning model used and its training procedure. In §5 we calibrate biases and measure the model performance using simulated data. In §6 we use the learnt model to measure shear in real images of galaxy clusters, also assessing bias (relative to RRG estimates) and statistical performance on the cluster data. We describe our code and data release in §7 and conclude in §8.

2 OBSERVATIONS

In this section we describe the *HST* observations and input catalogues used in this work. These served both to produce simulations, based on real galaxy images drawn from COSMOS observations (Scoville et al. 2007; Koekemoer et al. 2007), and to perform weak lensing analysis of the CLASH galaxy cluster data (Postman et al. 2012). We also describe here the additional photometric measurements and cuts performed to supplement the input catalogues. These were needed to enable the matching of the COSMOS and CLASH distributions in the simulation phase discussed in the following section.

Table 1. Properties of the 20 CLASH clusters included in our analysis. Here z_{clust} are the cluster redshifts provided by Postman et al. (2012), A_{eff} is the fully exposed angular area in $\prime^2$ (see §2.1.3), N_{wl} is the number of sources in the full weak lensing catalogue of Merten et al. (2015), N_{cut} is the number of sources in our subset catalogue after the additional cuts of §2.1.3, β_{eff} is the mean lensing efficiency (see Bartelmann & Schneider 2001, §4.3.2 for a definition of lensing efficiency, or *lens strength*) of sources in the cut (β_{eff} assumes a flat cosmological model similar to a WMAP7 cosmology as in Komatsu et al. 2011, with $\Omega_m = 0.27$, $\Omega_\Lambda = 0.73$ and a Hubble constant of $h = 0.7$) and n_{cut} is the number density of the cut sources in $\prime^{-2}$.

Name	z_{clust}	A_{eff} $\prime^2$	N_{wl}	N_{cut}	β_{eff}	n_{cut} $\prime^{-2}$
Abell 1423	0.213	14.4	807	442	0.738	30.7
Abell 209	0.206	14.4	832	323	0.759	22.5
Abell 2261	0.224	17.5	725	441	0.710	25.2
Abell 383	0.187	16.4	796	574	0.767	34.9
MACSJ0329-02	0.45	15.2	493	359	0.593	23.6
MACSJ0416-24	0.396	14.5	551	326	0.632	22.5
MACSJ0429-02	0.399	14.4	654	393	0.594	27.2
MACSJ0647+70	0.584	17.1	773	384	0.484	22.5
MACSJ1115+01	0.352	14.4	491	331	0.645	23.0
MACSJ1149+22	0.544	13.7	844	642	0.460	46.9
MACSJ1206-08	0.44	14.8	581	345	0.600	23.4
MACSJ1311-03	0.494	14.4	447	288	0.503	20.0
MACSJ1720+35	0.391	16.6	635	331	0.626	19.9
MACSJ1931-26	0.352	14.3	709	515	0.595	36.0
MACSJ2129-07	0.57	14.3	853	528	0.519	37.0
MS 2137.3-2353	0.313	14.4	801	399	0.695	27.6
RXJ1347-1145	0.451	14.3	633	339	0.577	23.8
RXJ1532.9+3021	0.345	16.2	508	310	0.648	19.2
RXJ2129+0005	0.234	14.4	624	311	0.701	21.6
RXJ2248-4431	0.348	15.1	598	366	0.651	24.2

2.1 CLASH data

The premise of our work is that galaxy shape estimation methods may not fully capture the lensing shear signal in an observational regime where a large fraction of the lensed sources are well resolved. We therefore focus our attention on space based observations, specifically the galaxy cluster core observations available in CLASH (Postman et al. 2012). The full CLASH data set[1] provides *HST* observations in 16 bands for 25 galaxy clusters, enabling also a robust estimation of photometric redshifts for the detected background sources (produced by the CLASH collaboration using the BPZ algorithm of Benitez 2000). Of these 25 clusters, five could not be included in our analysis, either due to a missing F625W filter image (Abell 611, MACSJ0744+39, MACSJ1423+24), missing distortion parameters (CLJ1226+3332) or a guide star astrometric failure (MACSJ0717+37). The 20 clusters we do include are listed in Table 1.

In this work, the CLASH observations serve a dual purpose. First, we would like our simulations to be grounded to a specific observational setting. Second, we would like to demonstrate that our estimator, which is trained and tested on simulated examples, is also able to estimate shear on real galaxy cluster observations, producing results which are at least consistent with those obtained from existing shear estimation methods.
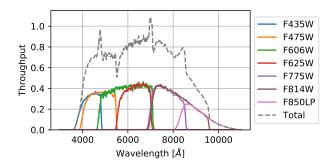
[1] https://archive.stsci.edu/prepds/clash/

Figure 1. Throughputs of the 7 ACS optical filters used in this work. We also plot the throughput of the effective filter resulting from their addition in the *ACS total* images.

The RRG (Rhodes et al. 2000) shape catalogue produced for the weak lensing analysis of the CLASH *HST* observations by Merten et al. (2015) is the baseline we choose to compare our method to. The RRG method was specifically designed to reduce shape measurement errors for well resolved galaxies and is commonly used in weak lensing analyses of *HST* observations. Using the full CLASH catalogue, the shape catalogue of Merten et al. (2015) was produced by first selecting background sources using BPZ photometric redshifts. PSF corrected RRG shapes were measured for each background source on 0.03″ pixel-scale co-added images. This was done per filter, per visit, allowing dense TINY-TIM PSF models to be estimated in the field of view and reference frame of each co-added image (Krist et al. 2011; Rhodes et al. 2007). These shape measurements were then rotated to a J2000 north-up reference frame and combined per background source by computing *signal-to-noise* (S/N) weighted averages.

2.1.1 ACS total images

For our own weak lensing analysis of the CLASH data, we choose to estimate shear on the same set of filters used by Merten et al. (2015) but using the 0.065″ pixel-scale, north-up, MOSAICDRIZZLE images made available by the CLASH collaboration. Although this pixel-scale is considered sub-optimal for weak lensing analysis (see Rhodes et al. 2007), this choice allows us to reduce the effects of noise correlations in the simulations as discussed in §3. To simplify the current analysis and to maximize S/N, we also choose to measure shear in each cluster on a single total filter-co-added image. We define this image and its accompanying inverse variance image below.

Although these two choices may introduce additional systematic and statistical errors (e.g. co-addition of images having different PSF) in our weak lensing analysis of the CLASH clusters (§6), they do not affect the validity of the performance metrics evaluated on the simulations (§5). Moreover, in our CLASH analysis, after calibrating a linear bias model relative to the baseline RRG shears, we demonstrate that our results are both consistent with those of the baseline and show a relative reduction in statistical errors comparable to the one we measure on simulations. In §8 we discuss ways in which these design choices can be avoided in future studies.
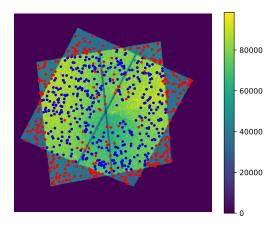
Figure 2. Inverse variance weight map for representative CLASH cluster Abell 1423, with weak lensing catalogue overlaid. Blue (red) points mark locations of the original (excluded) background sources. A minimum mean weight threshold of $5 \times 10^4 \, [\mathrm{ADU}^{-2}]$ was chosen to exclude underexposed sources.

We now describe how these *total images* were computed. Starting with the available optical filter exposures (see Figure 1) taken by the *Advanced Camera for Surveys* (*ACS*; Ford et al. 2003) on-board the *HST*, CLASH performed image reduction, alignment, co-addition and produced per filter images using the MOSAICDRIZZLE pipeline at an output pixel-scale of $0.065''$. Each such filter image is accompanied by an inverse variance image which encodes at each pixel the noise from accumulated dark current, detector readout, and photon noise from the background, but not the additional Poisson noise from the astronomical sources in the image (see Koekemoer et al. 2011, §5.8.1). We denote the set of $n = 7$ filter images by I_i and the inverse variance images by V_i^{-1}. The total images and total inverse variance images that are used in this work (for each cluster field) are then simply

$$I_{\mathrm{total}} \;=\; \frac{1}{n} \sum_i I_i, \tag{2}$$

$$V_{\mathrm{total}}^{-1} \;=\; n^2 \left(\sum_i V_i \right)^{-1}. \tag{3}$$

We choose to use these total images because they have a spatially constant and well defined spectral throughput (see *total* in Figure 1), as opposed to the *inverse variance weighted* total images made available by the CLASH collaboration (Postman et al. 2012).

2.1.2 Photometry

In addition to the per band photometric catalogue of Postman et al. (2012) we measure photometry in the aforementioned ACS total images. Using SEXTRACTOR version 2.19.5 (Bertin & Arnouts 1996), we detect sources and measure full-widths at half-maximum (FWHM), magnitudes (M_{total}), and half-light flux-radii ($r_{1/2}$) on `AUTO` apertures. We use the configuration parameters used by Postman et al. (2012), and the ACS total inverse variance images (defined in §2.1.1) as

weight maps. These measurements are then assigned to the weak lensing catalogue of Merten et al. (2015) by image coordinate matching of the two catalogues. Approximately 87% of the weak lensing sources, which were originally detected using additional bands (see the ACS+IR detection image used by Postman et al. 2012), are matched to sources within $0.65''$ in the ACS total images in this way. For presentation purposes, we define the zero-points of the ACS total image magnitudes, M_{total}, such that $\langle M_{\mathrm{total}} \rangle = \langle M_{\mathrm{F814W}} \rangle$ on all matched sources in the weak lensing catalogues, and M_{F814W} are the isophotal F814W magnitudes of Postman et al. (2012).

2.1.3 Point source detection and additional cuts

To reduce lensing signal dilution due to incorrect inclusion of Milky Way stars in the background source catalogue of Merten et al. (2015), we detect point sources using the method described by Leauthaud et al. (2007) and Bardeau et al. (2005). This method makes use of the fact that sufficiently bright stars are usually well separated from galaxies in the radius-magnitude and the μ_{max}-magnitude planes. Here, μ_{max} is the magnitude of the maximal pixel in the detected aperture of a source. In Figure 3 we show the distribution of the background sources in these two planes. Sources belonging to either one of the grey regions in Figure 3 are considered to be stars and excluded from our weak lensing catalogue. Leauthaud et al. (2007) found that separation using the two planes (radius-magnitude and μ_{max}-magnitude) is overall consistent to about 1% and 2% at F814W magnitudes below 24 and 25 respectively, with the difference between the two methods mainly arising from a misclassification of objects in the radius-magnitude plane due to the presence of close pairs. We use the two planes in tandem in the following way. First, the clearly separated horizontal locus is selected in the radius-magnitude plane by selecting lenient upper bounds for the radius and magnitude at which the two populations are well separated (allowing for some point sources to be missed). Using this radius-magnitude selected population of point sources, their locus of points in the μ_{max}-magnitude plane is selected by fitting the intercept and slope in the μ_{max}-magnitude plane and manually setting an upper bound for μ_{max} (and a horizontal extent for this stripe) at which the populations are well separated. This allows us to use the μ_{max}-magnitude plane for the eventual classification while reducing the number of manually adjusted degrees of freedom.

We measure the median FWHM of these point sources to be $0.121''$ and use this width (and an isotropic Gaussian profile) as the model PSF in our simulations.

To reduce the variation in the noise properties of the sources in our weak lensing catalogue we additionally exclude sources located in regions of the field that were not fully exposed (see Figure 2). We do this by cutting out sources for which the mean value of the inverse variance image in a 32 pixel wide square aperture is below $5 \times 10^4 \, [\mathrm{ADU}^{-2}]$. We also remove sources overlapping the bad-pixel mask of any one of the optical filters.
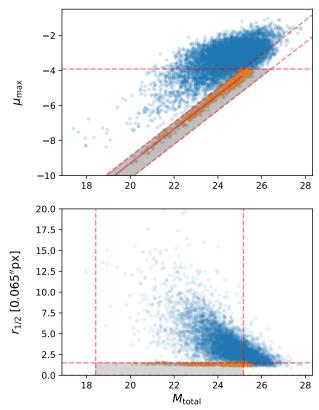
Figure 3. Point source detection (orange points) in the combined set of background source catalogues of Merten et al. (2015), for the CLASH clusters listed in Table 1. Point sources were selected using the procedure of Leauthaud et al. (2007) and Bardeau et al. (2005), by first manually selecting a segment in $r_{1/2} - M_{\rm total}$ space (bottom) for which the horizontal branch is well separated. This population was then used to fit a slope and intercept in $\mu_{\rm max} - M_{\rm total}$ space (top). After also manually selecting a well separating maximal $\mu_{\rm max}$ cutoff, the final set of point sources was defined to be those contained in either one of the enclosed regions (grey).

2.1.4 Correlated background estimation

It is well known that the co-addition of images by the Mo-saicDrizzle pipeline results in local noise correlations in the output images. To faithfully reproduce the statistics of the CLASH observations in the simulations, we measure the local covariance properties of the background pixels. Using the SEGMENTATION and -BACKGROUND *check images* produced by SExtractor while processing each field, we extract all background subtracted 7×7 pixel patches which do not overlap detected apertures and which survive the same noise exclusion criteria listed in §2.1.3. Denoting the 49-dimensional column vector encoding the i'th such patch by p_i, then the empirical covariance matrix of m such patches is (see Figure 4)

$$\Sigma_{\rm BG} = \frac{1}{m-1} \sum_{i=1}^{m} p_i p_i^T .$$ (4)

Computing a matrix square root of $\Sigma_{\rm BG}$ and extracting the central row of the resulting matrix,

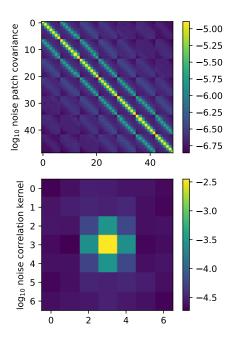$$K_{\rm BG} = \left[\sqrt{\Sigma_{\rm BG}} \right]_{i=25},$$ (5)

Figure 4. Background noise covariance $\Sigma_{\rm BG}$ computed from background subtracted 7 × 7 pixel patches not overlapping detected sources (top), and noise reproducing kernel $K_{\rm BG}$ (bottom). Convolving $K_{\rm BG}$ with standard normal noise images produces correlated noise having the same local covariance as measured by $\Sigma_{\rm BG}$.

we obtain a vector encoding a 7 × 7 kernel. This kernel is then used in §4.2 to generate additive correlated noise having the same covariance properties as the typical CLASH background, by convolving $K_{\rm BG}$ with independent identically distributed (iid) standard normal noise image pixels.

2.2 COSMOS data

For the simulations (based on real galaxy images) of the following section we use the two magnitude limited subsets of the *HST* COSMOS survey (Scoville et al. 2007; Koekemoer et al. 2007), $M_{\rm F814W} < 23.5, 25.2$ magnitude, distributed with the GalSim[2] package (Mandelbaum et al. 2012; Rowe et al. 2015). These two datasets consist of random subsets of the full catalogue of Leauthaud et al. (2007), containing 56,062 and 87,798 sources respectively, with 24,868 sources contained in their intersection. The datasets include galaxy stamps extracted from registered, undistorted, sky-subtracted images, co-added in an unrotated reference frame at an 0.03″ pixel-scale using the MultiDrizzle pipeline (Koekemoer et al. 2003). This was done after correcting for the effects of charge transfer inefficiency (CTI) using the method of Massey et al. (2009). Point source detection was performed to remove stars from the source catalogues (Leauthaud et al. 2007; Bardeau et al. 2005). The datasets additionally include TinyTim PSF models

---

[2] https://github.com/GalSim-developers/GalSim

(Krist et al. 2011; Rhodes et al. 2007) estimated at the extracted source positions as well as a SExtractor (Bertin & Arnouts 1996) photometric catalogue, including measurements of F814W AB magnitudes on `AUTO` apertures and half-light-radii (`FLUX_RADIUS` with `PHOT_FLUX_FRAC` set to 0.5). For a more complete description of the image processing stages, additional catalogue cuts and the photometric measurements performed, the reader is referred to Mandelbaum et al. (2012) and references therein.

3   SIMULATIONS

We simulate realistic galaxies using the GalSim package (Rowe et al. 2015) implementing the Reconvolution algorithm of Mandelbaum et al. (2012). GalSim uses the aforementioned COSMOS datasets to render realistic galaxies as they would appear, post shear, in wider PSF observing conditions. Although we required simulations similar to those produced for the realistic, space based, constant shear (RSC) branch of the GREAT3 challenge (Mandelbaum et al. 2014), these were not suitable due to a *shape-noise cancellation* procedure employed there (the inclusion of 90° rotated pairs in each constant shear field, to reduce the effects of shape-noise). Additionally, our aim was to match the distribution of the CLASH observations as closely as possible. We do so with regards to noise properties (§3.1) and the source magnitude-radius distributions (§3.2).

3.1   S/N trade-offs, rescaling and cut-off

In this work, we simulate observations which are intended to match a target population of galaxy images (the CLASH background galaxies) using a source population of galaxy images (the COSMOS in-the-field galaxies). As we discuss in the following section, we choose to match the simulations to the target population by focusing on their distribution in magnitude-radius space. We also require the simulations to reproduce the S/N conditions in the observations. The COSMOS stamps already contain background noise, at a level and having correlations different from the ones we wish to simulate in CLASH. Training our model on simulated galaxies containing this type of correlated noise leads to poor generalization to a validation set and to the real CLASH dataset. Reproducing the S/N conditions was therefore necessary and required particular treatment.

GalSim's *noise whitening* and *noise symmetrization* features are intended to allow control of the covariance properties of the additive background noise of the simulated target stamps. In the version of GalSim used in this study (version 1.4.4) we found that both *noise whitening* and *noise symmetrization* resulted in correlated noise in the target stamps which depended on the applied shear. Recently Sheldon & Huff (2017) reported difficulties in performing model bias calibration using these features due to what we believe are similar causes.

Unable to reproduce the intended background noise in the stamps at their original scale (using GalSim's *noise whitening* and *noise symmetrization*), we chose instead to reduce the scale of the stamps and thereby reduce the level of per pixel noise leaking into the simulations. Additionally, in those regions of the stamp where flux from the galaxy was

low relative to the noise level (we chose a particular per pixel S/N cutoff of 3.5), pixel values were nullified. These rescaled and masked COSMOS stamps were then treated as noiseless observations to which the target (pseudo-random) correlated background noise (having the CLASH noise statistics as measured in §2.1.4) was added.

This procedure introduces a trade-off between an image coordinate rescaling factor α, and the extent to which each galaxy aperture is masked (which should both preferably be avoided as much as possible: α should preferably be close to unity to avoid a galaxy age mismatch and the aperature should be as large as possible to avoid a discontinuity at the galaxy outskirts). As discussed above, this trade-off is due to the fact that the target stamp S/N increases with decreasing rescaling factor α. In this section we illustrate and quantify this trade-off and report on the particular choices made in our own simulations.

Assuming an isotropic Gaussian profile for the galaxy intensities in the target images

$$I(r) = \frac{F}{2\pi\sigma^2} e^{-\frac{r^2}{2\sigma^2}}, \tag{6}$$

parametrized by a total galaxy flux F and a Gaussian width σ. Using the Gaussian half-light-radius, $r_{1/2} = \sqrt{2\ln 2}\,\sigma$, we can express the cutoff radius of a particular galaxy, r_c, at which $I(r_c) = I_c$, as follows

$$r_c^2 = \left[\ln\left(\frac{F}{r_{1/2}^2}\right) - \ln I_c + \ln\left(\frac{\ln 2}{\pi}\right)\right] \frac{r_{1/2}^2}{\ln 2}. \tag{7}$$

We also assume the following scaling relation between the background RMS in the source and target images

$$\sigma_{\text{bg,t}} = \frac{\alpha P_t}{P_s} \sigma_{\text{bg,s}} \tag{8}$$

where P_s and P_t are the pixel-scales of the source and target stamps respectively. This is consistent with the way we scale surface brightness, and assumes the background noise in the source pixels is uncorrelated. Equation 8 however does not take the effects of the target PSF into account. Defining the per pixel cutoff signal-to-noise ratio in the target images, $\text{SN}_c \equiv I_c/\sigma_{\text{bg,t}}$, then for pixel-scales $P_s = 0.03''$, $P_t = 0.065''$, a measured mean background RMS in the source COSMOS stamps of $\sigma_{\text{bg,s}} \approx 2.9 \times 10^{-3}$, and the requirement $\text{SN}_c = 3.5$, we find that the cutoff intensity is $I_c \approx 2.2 \times 10^{-2}\,\alpha$. Using equation 7 and the $F - r_{1/2}$ distribution of the COSMOS datasets (scaled to the target stamps), we plot in Figure 5 the proportion of galaxies for which at least the inner half-light-radius will not be masked (those having $r_c > r_{1/2}$), as a function of the rescaling factor α. Finally, we choose a somewhat arbitrary rescaling factor $\alpha = 0.542$ so that a large fraction of the COSMOS samples (99% of the $M_{\text{F814W}} < 23.5$ dataset and 92% of the $M_{\text{F814W}} < 25.2$ dataset) have $r_c > r_{1/2}$. Moreover, the very low surface brightness galaxies in COSMOS, which are severely masked, are rarely included in the resampled distribution discussed in §3.2.

To summarize this section – this choice of geometric rescaling factor, between source stamps and target stamps, allowed us to control the noise properties of the simulations
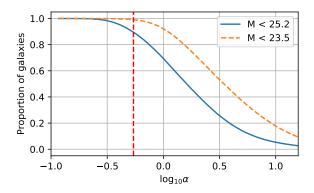
Figure 5. Proportion of galaxies in the COSMOS $M_{F814W} <$ 23.5, 25.2 magnitude datasets for which the inner half-light-radius remains unmasked as a function of the image coordinate rescaling factor α. The vertical line marks the value $\alpha = 0.542$ chosen for our simulations.

(mask pixels having S/N below 3.5) while minimizing the extent to which stamps are masked. In the following section, the CLASH magnitude-radius distributions are then matched by resampling the simulated distribution. This is done after taking this rescaling factor into account.

3.2 Matching magnitude-radius distributions

The distribution of observed galaxies is highly diverse and dependent on some well known parameters. Examples are the dependence of morphology on age, the dependence of spectra on redshift, background levels and their effect on S/N conditions, only to name a few. Attempting to simulate the distribution of CLASH *total ACS* filter stamps using the rescaled COSMOS F814W filter observations, and unable to do so perfectly, we choose to focus on matching the COSMOS and CLASH distributions only in the observed magnitude-radius plane. We find that this choice captures much of the variation in appearance (see Figure 6). Treating the rescaled COSMOS stamps (see §3.1) as noiseless observations and adding stochastic correlated background noise (as measured on CLASH in §2.1.4), results in simulations which also reproduce the observed S/N distribution.

We perform this matching by resampling the rescaled (as in §3.1) COSMOS stamps. In Figure 7 we show the CLASH observations in the $\log_{10} r_{1/2} - M_{total}$ plane (top left), the distribution of the rescaled COSMOS $M_{F814W} <$ 25.2 dataset (top right) and the distribution of this resampled COSMOS dataset (bottom left). The rescaling of COS-MOS included resizing by the geometric factor $\alpha = 0.542$ and translating the COSMOS distribution in the positive horizontal direction to maximize the overlap with the CLASH distribution (after matching zeropoints and scaling COS-MOS flux by α^2, this amounted to an extra magnitude shift of $\Delta M = 1.236$). We plot the original CLASH and resampled COSMOS marginal distributions in Figures 8 and 9. Aside from a slight lack of simulated galaxies at the high $r_{1/2}$ tail, we see that the distributions match well overall.

4 MODEL

We approach the weak lensing shear estimation problem as a high-dimensional regression task, where given a 32×32 pixel stamp of the observed galaxy intensities, encoded as a 1024-dimensional vector denoted x, we want to predict the two component vector g parametrizing the reduced shear it has undergone. To this end, we wish to find a function $\hat{g} = \Phi(x; w)$ parametrized by a constant weight vector w, that would map observations x to consistent and efficient estimates $\hat{g}$ of the true reduced shears g, for various pairs (x, g) arising from a joint distribution $(x, g) \sim \mathcal{D}_{train}$. We discuss the class of functions $\{\Phi(x; w)\}_w$ which we consider in §4.1, and the way in which we search for appropriate weights in §4.2. Finally in §4.3 we discuss how imposing certain symmetry properties on the model further enhances its performance.

The class of functions we use, the optimization algorithm we use to search for a good set of weights, and the way we do this directly from a large set of *training examples*, are all drawn from an approach to machine learning known as *deep learning* (see LeCun et al. 2015). This approach has in recent years led to considerable progress in the field of computer vision, most notably in visual object classification and detection tasks. Typically, weak lensing shear estimation methods are designed by directly defining a procedure that localizes galaxies (finds their centroid and effective radius) and then measures certain morphological features of the observed galaxies that correlate to the reduced shear they have undergone. The way in which localization is performed, the particular way in which galaxy shape is measured (e.g. RRG's weighted moment approach), the way in which background subtraction is performed, as well as various other observational conditions (e.g. per galaxy S/N), all affect the results of the shear estimation process. The typical approach to these complicating factors is to use matching simulations to calibrate the effects these design choices and observational conditions have on estimation bias (see e.g. Fenech Conti et al. 2017; Samuroff et al. 2017; Zuntz et al. 2018; Mandelbaum et al. 2018). In this work the simulations are used not only to calibrate bias, but also to directly learn the estimation algorithm itself using a large set of training examples drawn from the simulations. This is motivated by the success of this type of *end-to-end learning* in finding models that perform better than their manually designed counterparts.

The model we describe in this section, as well as its optimization and testing procedures were implemented using the TENSORFLOW[3] and KERAS[4] open source deep learning libraries (Abadi et al. 2016; Chollet et al. 2015). The reader is referred to §7 for further implementation details.

4.1 Architecture

The regression function that we seek, $\Phi(x; w)$, has a particular form known as a multi-layered (or *deep*) neural network (DNN), or, more specifically, a deep *convolutional neural network* (CNN). This type of function can be expressed as
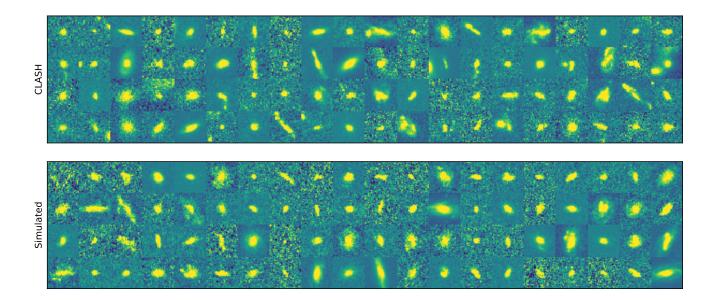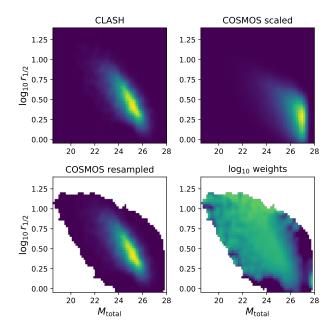

Figure 6. Random subsets of 32×32 pixel galaxy stamps drawn from the CLASH ACS total image dataset (top) and from the COSMOS based simulations (bottom). Each stamp is shown here after subtracting its original mean and dividing by its standard deviation.

Figure 7. Distribution of the CLASH sources (top left), rescaled COSMOS $M_{\rm F814W} < 25.2$ dataset (top right) and this rescaled + resampled COSMOS dataset (bottom left) in the $\log_{10} r_{1/2} - M_{\rm total}$ plane. The resampling weights, matching COSMOS to CLASH, are shown on the bottom right.

a composition of n intermediate functions (or *layers*),

$$\Phi(x; w) = \Phi_n \circ \Phi_{n-1} \circ \ldots \circ \Phi_1(x), \qquad (9)$$

$$\Phi_i \circ \Phi_{i-1}(\cdot) \equiv \Phi_i(\Phi_{i-1}(\cdot; w_{i-1}); w_i), \qquad (10)$$

$$w \equiv \{w_i\}_{i=1}^n. \qquad (11)$$

Here we used "$\circ$" to denote function composition and "$\cdot$" to denote unnamed function arguments (e.g., the composition
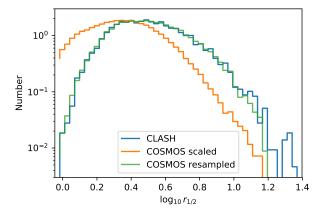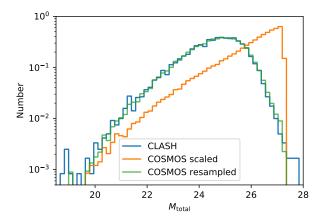

Figure 8. Normalized histograms in $\log_{10} r_{1/2}$ bins of the CLASH sources, the rescaled COSMOS $M_{\rm F814W} < 25.2$ dataset and the rescaled + resampled COSMOS $M_{\rm F814W} < 25.2$ dataset.

of functions $f(\cdot)$ and $g(\cdot)$ is then $h = f \circ g$ such that $h(x) = f(g(x))$ for some input x).

Each intermediate function $\Phi_i(\cdot; w_i)$, parametrized by its own set of weights w_i, transforms one intermediate representation x_i of the original input to the following

$$x_{i+1} = \Phi_i(x_i; w_i). \qquad (12)$$

The i'th *layer* of a DNN refers either to x_i, or to $\Phi_i(\cdot; w_i)$, depending on the context. The specification of the functions $\Phi_i(\cdot; w_i)$, for yet unspecified values of the weights, constitutes the *architecture* of the DNN. In our CNN, these layer functions are either *convolutional* or *affine*. In Table 2 we list the particular six-layer architecture used in this work. This architecture, computed left-to-right and top-to-bottom, transforms a 1024-dimensional input stamp to a two-dimensional shear estimate $\hat{g}$. The architecture consists

Figure 9. Normalized histograms in M_{total} bins of the CLASH sources, the rescaled COSMOS $M_{\text{F814W}} < 25.2$ dataset and the rescaled + resampled COSMOS $M_{\text{F814W}} < 25.2$ dataset.

Table 2. Layers of the CNN architecture used in this work, computed left-to-right and top-to-bottom. Overall this architecture has 107,866 parameters. See Appendix A for a definition of the layers and operations mentioned here.

Layer operations	Output Shape	Params.
Input	$32 \times 32 \times 1$	-
Conv $(5,5) \times 64$, Stride $(3,3)$, ReLU	$10 \times 10 \times 64$	1,664
Conv $(1,1) \times 40$, ReLU	$10 \times 10 \times 40$	2,600
Conv $(1,1) \times 10$, ReLU	$10 \times 10 \times 10$	410
Affine (100), ReLU, Dropout	100	100,100
Affine (30), ReLU, Dropout	30	3,030
Affine (2)	2	62

of three *convolutional layers* followed by three *affine layers* (refer to Appendix A for a definition of *convolutional* and *affine layers*, as well as some other operations mentioned in Table 2). Table 2 defines the dimensions (n, m) and number (k) of convolution kernels used in a convolutional layer by Conv $(n, m) \times k$ and the following spatial downsampling operation by Stride (Δ_x, Δ_y). Affine layers, having output dimension l, are denoted by Affine (l). Table 2 also lists, for each layer, the dimensions of the intermediate representation x_i, as well as the number of scalar parameters included in the weight vector w_i. In total, our model's architecture includes $107\,866$ parameters which are subsequently optimized (or trained), using the simulation training set, in §4.2. We remark here that at the present time, there is no rigorous way of finding an optimal architecture for a particular deep learning problem. Our architecture was chosen after testing only a few alternatives. It is therefore likely that other architectures, leading to better performance, can be found.

4.2 Training and data augmentation

For the purpose of optimizing the model, we use the standard mean-squared-error (MSE) *loss function*, evaluated on a distribution of (stamp, shear) pairs

$$L(w) = \mathop{\mathbb{E}}_{(x,g) \sim \mathcal{D}_{\text{train}}} \left[||\Phi(x; w) - g||^2 \right], \tag{13}$$

where $\mathcal{D}_{\text{train}}$ is the distribution we draw our training samples from. The global minimum w^* of such a loss function, for an expressive enough model, should result in an estimator $\Phi(x; w^*) = \mathbb{E}[g|x]$, which produces the conditional mean of the shear g given the observation x (see e.g. Bishop 1994). In practice, the optimization is performed using a finite set of training examples, using the specific architecture defined in §4.1 and using an iterative stochastic optimization algorithm known as RMSProp (Tieleman & Hinton 2012), which is not guaranteed to produce a global minimum. The training set we use at this stage differs from the simulations as defined in §3 in the following ways:

(i) Both the resampled and the original COSMOS datasets discussed in §3 have a large over-abundance of high magnitude galaxies. We found that training using a more homogeneous magnitude distribution enhances the model's performance on the CLASH matched simulations. We homogenize the training set by sampling an equal number of galaxies in each of the original COSMOS M_{F814W} magnitude segments: $[16, 19.5], [19.5, 21.5], [21.5, 23.5]$.

(ii) The scale of simulated stamps is randomly and uniformly augmented in the range $\pm 15\%$.

(iii) The post shear centroid position of simulated stamps is randomly and uniformly augmented in the range $\pm 7\,\text{px}$ both in the x and y directions.

(iv) Stamps are randomly mirrored (with probability 0.5), and rotated randomly and uniformly by an angle $\theta \in [0, 2\pi]$.

The augmentations described here allow us to enlarge the effective size of the training and test sets and also encourage the resulting model to be invariant to the above transformations, specifically to translation (see §5.2). Additionally, without enlarging the training data using this data augmentation procedure the resulting models tend to over-fit the training dataset and perform poorly on the test dataset. This property, known as *over-fitting* (or lack of generalization), commonly occurs in machine learning models having a large number of parameters and a relatively small number of training examples. This is why these augmentations are necessary. Note that although the training set includes augmentations that may be leading to an "improper" population resampling, the test set, with which we calibrate and evaluate the performance of our model in §5, does not include these type of augmentations (i - iii) and is properly sampling the population as defined in §3 in these respects.

Table 3 gives a quantitative summary of how the original COSMOS $M_{\text{F814W}} < 23.5, 25.2$ magnitude datasets were split and enlarged into the final training and test sets. Both the training stamps and the test stamps are normalized, before the CNN operates on them, by subtracting the stamp's per-pixel mean and dividing by its standard deviation (see e.g. Figure 6). Doing this both assists in the convergence of the optimization and has the added benefit of making the trained model invariant to the subtracted background level. Correctly estimating the background level is a difficult task in itself and an incorrect estimate can in turn affect RRG moment estimates (see discussion in §5.2). Having the shear estimator insensitive to constant background levels is therefore an advantageous property.

The CNN estimator is also completely invariant to flat

Table 3. Simulated datasets generated by augmenting mutually exclusive training and test random subsets of the original COSMOS $M_{\mathrm{F814W}} < 23.5, 25.2$ sources, as discussed in §4.2.

	Original		Aug. (i-iv)	Aug. (iv)	
	Train	Test	Train	Train	Test
COSMOS 25.2	65 756	21 994	1.5M	308K	308K
COSMOS 23.5	22 423	-	750K	-	-

back- ground levels subtracted from the stamp. As discussed in §4.2, this is due to the way stamps are normalized before the CNN operates on them. Accurately measuring background levels in galaxy clus- ter lensing is particularly challenging due to the spatially varying foreground intra-cluster-light (ICL). Attempts have recently been made to better model the ICL and correct for its systematic effect on background source magnitude and photometric redshift measurements (see Molino et al. 2017; Gruen et al. 2018). We expect an incorrect background estimation to also affect RRG shear estimates through its affect on the SExtractor segmentation process (and half-light-radius estimation), as well as through the weighted mo- ments evaluated by RRG. This insensitivity of the CNN estimator to constant background levels is therefore an advantageous property.

At the final stage of training, we set all model weights constant and fit the last affine layer by linear-least-squares regression between the 30-dimensional representation at layer five and the 2-dimensional output at layer six. This is done using the training set in which only augmentation (iv) was performed. This allows us to tune the model to the CLASH matched distribution. Additionally, we multiply all weights in layer six by $[1 - L(w^*)/\mathrm{Var}(g)]^{-1}$, where w^* are the optimized weights and $L(w^*)$ is the MSE loss evaluated on the CLASH matched training set at w^*. This is done to correct for the effects of *regression attenuation* as discussed in Appendix B.

4.3 Enforcing shear symmetries

The components of the true reduced shear transform as $[g_1, g_2]^T \rightarrow R(2\theta)[g_1, g_2]^T$ in response to a rotation of the image coordinates by an angle θ, and as $[g_1, g_2] \rightarrow [g_1, -g_2]$ under reflection. Although we train the model in §4.2 using a distribution of samples containing both mirrored and rotated copies of the original COSMOS datasets, the optimization does not necessarily converge to a model estimator having these precise transformation properties. We impose these transformation properties on our trained estimator in the following way

$$\tilde{g}_{\mathrm{CNN}} \equiv \phi(x; w^*) = \frac{1}{8} \sum_{i \in \{-1,1\}} \sum_{j=0}^{3} t_{i,j}(\Phi(T_{i,j}(x); w^*)). \quad (14)$$

Here the function $T_{i,j}(x)$ vertically flips the stamp x, depending on the value of i, and rotates it by an angle of $\pi j/2$. The function $t_{i,j}(g)$ performs the appropriate shear transformation (as discussed above). Using $\phi(x; w^*)$ instead of $\Phi(x; w^*)$ as our estimator results in an additional overall reduction of ~2.5% in the RMS errors measured in §5.3.

5   PERFORMANCE ON SIMULATED DATA

Based on the simulations described in §3.2, we measure the performance of an estimator $\hat{g}_e$ using the RMS error $\sigma_{e,i}$,

$$\sigma_{e,i}^2 = \mathop{\mathbb{E}}_{(x,g) \sim \mathcal{D}_{\mathrm{test}}} \left[\left(\hat{g}_{e,i} - g_i \right)^2 \right], \quad (15)$$

$$\hat{g}_{e,i} = (\tilde{g}_{e,i} - c_{e,i})/(1 + m_{e,i}), \quad (16)$$

where we denoted the true simulated shears by g, the uncalibrated estimates by $\tilde{g}_{e,i}$, the calibrated estimates by $\hat{g}_{e,i}$, the type of estimator used by subscript $e \in \{\mathrm{CNN}, \mathrm{RRG}\}$, the reduced shear component by subscript $i \in \{1, 2\}$, and the specific test distribution we evaluate performance on using $\mathcal{D}_{\mathrm{test}}$. Here we use our own implementation of the RRG algorithm as it is described in Rhodes et al. (2000). The PSF we use in the simulations, for all generated galaxies is an isotropic Gaussian profile having a FWHM of 0.121″ as discussed in 2.1.3. We evaluate the performance on $\mathcal{D}_{\mathrm{test}}$ only after calibrating an affine bias model for each estimator and component of shear separately,

$$(m_{e,i}, c_{e,i}) =$$
$$\mathop{\arg\min}_{(m_{e,i}, c_{e,i})} \mathop{\mathbb{E}}_{(x,g) \sim \mathcal{D}_{\mathrm{test}}} \left[\left(\tilde{g}_{e,i} - c_{e,i} - (1 + m_{e,i})g_i \right)^2 \right]. \quad (17)$$
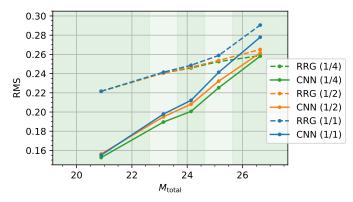
We do not use a separate calibration set to do this, but instead use the same test set both to calibrate biases and then to measure RMS errors. By repeated simulation of this calibration process we find for the RMS error estimates (see Table 4) statistical uncertainties of $\pm 1.0 \times 10^{-3}$ and $\pm 8.6 \times 10^{-4}$ for the RRG and CNN estimators respectively and systematic uncertainties below $+5 \times 10^{-5}$.
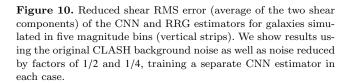
By varying the test set $\mathcal{D}_{\mathrm{test}}$, we measure how each estimator's performance varies with varying source population magnitude (§5.1) and assess the sensitivity of the estimators to centroid misalignment and to an improper background subtraction (§5.2). In §5.3 we discuss the overall RMS errors and biases for the CNN estimator when evaluated on the full population of the CLASH matched simulations.

5.1   Magnitude dependence

In Figure 10 we measure performance in five M_{total} magnitude bins. We do this with background noise simulated using the original noise reproducing kernel K_{BG} (see §2.1.4), and using kernels $K_{\mathrm{BG}}/2$ and $K_{\mathrm{BG}}/4$ to simulate lower noise conditions. This generates stamps with additive background noise having the original CLASH standard deviation, and standard deviations reduced by factors of $1/2$ and $1/4$. Different CNN models were trained for each noise level. We see that the RMS errors of both the RRG and CNN estimators rise with magnitude and that higher background noise levels lead to higher RMS errors, mainly at the fainter end. The RRG error curve also appears to converge, at low noise levels, to a bottom envelope.

In the brighter galaxies, more substructure is typically visible (see Figure 6). This substructure is what presumably provides the CNN estimator with additional shear information. The RRG estimator only has access to the second order surface brightness moments and is therefore limited in its statistical efficiency by their distribution.

Figure 10. Reduced shear RMS error (average of the two shear components) of the CNN and RRG estimators for galaxies simulated in five magnitude bins (vertical strips). We show results using the original CLASH background noise as well as noise reduced by factors of 1/2 and 1/4, training a separate CNN estimator in each case.

5.2 Sensitivity to alignment and background subtraction

In §5.1 we did not include the effects that centroid determination has on shear RMS errors. The centroid used there was the one provided by the COSMOS sample. This sample has higher resolution and higher S/N than the CLASH sample we eventually simulate and should therefore have a more accurate centroid measurement (visible in Figure 6). Here we would like to separately measure the effects a random misalignment of the galaxy centroid has on each estimator. We do this by producing test samples in which the centroid of each galaxy is randomly perturbed uniformly in a range having a total extent which is a particular proportion of the galaxy's half-light-radius, both in the x and y directions. This is done post shear. The results are presented in Figure 11. We see that the RMS errors of the RRG estimator rise significantly above a misalignment (range) of ~0.5. The performance of the CNN estimator, on the other hand, is barely affected by perturbations of the centroid in this range. This insensitivity of the CNN estimator is a result of the translational augmentations performed when producing the training set (see §4.2).

The CNN estimator is also completely invariant to flat background levels subtracted from the stamp. As discussed in §4.2, this is due to the way stamps are normalized before the CNN operates on them. Accurately measuring background levels in galaxy cluster lensing is particularly challenging due to the spatially varying foreground intra-cluster-light (ICL). Attempts have recently been made to better model the ICL and correct for its systematic effect on background source magnitude and photometric redshift measurements (see Molino et al. 2017; Gruen et al. 2018). We expect an incorrect background estimation to also affect RRG shear estimates through its affect on the SEXTRACTOR segmentation process (and half-light-radius estimation), as well as through the weighted moments evaluated by RRG. This
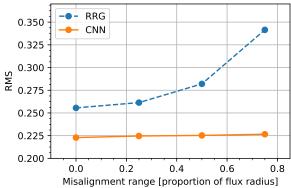

Figure 11. Reduced shear RMS error (average of the two shear components) of the CNN and RRG estimators for galaxies simulated with an increasingly perturbed centroid.

Table 4. Calibration parameters and RMS errors evaluated on the full CLASH matched test set. Here c_i are the additive biases, m_i are the multiplicative biases and σ_i are the RMS errors of the bias corrected estimates $\hat{g}_{e,i}$ of estimators $e \in \{\text{CNN}, \text{RRG}\}$.

	c_1	m_1	c_2	m_2	σ_1	σ_2
RRG	0.001	0.119	0.000	0.121	0.257(1)	0.259(1)
CNN	0.000	0.005	0.000	0.005	0.225(1)	0.224(1)

insensitivity of the CNN estimator to constant background levels is therefore an advantageous property.

5.3 Overall performance and bias

In Table 4 we summarize the calibration parameters and performance of the RRG and CNN estimators when evaluated on the full CLASH matched test set. As discussed in 5.2, these simulations do not account for a more realistic stamp misalignment or an incorrect background estimation. The simulated shear components in this test set were independently drawn such that $g_i \sim U(-0.2, 0.2)$. Overall the RRG estimator has an RMS error approximately 15% higher compared to that of the CNN estimator, at all tested true shear bins. In Figure 12 we measure the RMS errors and biases in 16 separate g_1 and g_2 true shear bins, each containing approximately 12,500 samples. We see that the absolute gap in RMS errors between the two methods is approximately constant as a function of shear, with both methods showing higher RMS errors at lower absolute shears. The bias (presented at five times the amplitude for clarity) appears to be consistent with being zero at this confidence level for each of the tested true shear bins.

6 CLASH WEAK LENSING ANALYSIS

To measure the statistical performance of the estimators on the CLASH observations we make use of the angular correlations in the reduced shear fields in each cluster. We first compute in §6.1 fields smoothed at various length scales. We then measure the covariance properties of the smooth fields for the two estimators. This enables us in §6.2 to calibrate a
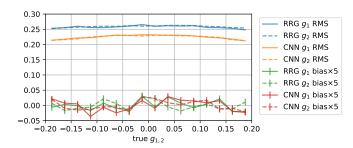
Figure 12. RMS errors and biases in 16 separate g_1 and g_2 true shear bins for the bias corrected RRG and CNN estimators. Each bin contains approximately 12,500 samples from the full CLASH matched test set. The bias correction here is constant and is the one presented in Table 4. Note that bias here is presented at 5 times the amplitude for clarity.

relative affine bias model between the CNN and RRG estimates on this dataset. In §6.3 we correct for this bias and use a held out set of background galaxies to measure the scatter between the per galaxy point estimates and the smooth fields. This provides us with an evaluation of the per galaxy statistical errors of the RRG and CNN estimators on the CLASH observations. Our final product for the 20 CLASH clusters is then a shear catalogue having RMS errors reduced by approximately 26%, as well as smooth shear fields having enhanced angular resolutions (at a set noise level).

6.1 Smooth field regression

For a given galaxy cluster, we measure the reduced shear field $g_{i,e,\lambda}(\vec{\theta})$, smoothed at angular scale λ

$$g_{i,e,\lambda}(\vec{\theta}) \quad = \quad \sum_{j=1} \hat{g}_{i,e}^{(j)} \delta(\vec{\theta} - \vec{\theta}_j) * S_\lambda(\vec{\theta}), \tag{18}$$

$$S_\lambda(\vec{\theta}) \quad \equiv \quad \frac{1}{\sqrt{2\pi}\lambda} \exp\left(-\frac{||\vec{\theta}||^2}{2\lambda^2}\right), \tag{19}$$

where $\vec{\theta}$ denotes sky coordinates, $\hat{g}_{i,e}^{(j)}$ is the reduced shear estimate of component i computed with estimator e of background galaxy j located at $\vec{\theta}_j$ and $S_\lambda(\vec{\theta})$ is a Gaussian smoothing kernel having width λ. We define spatial convolution of two images $f(\vec{\theta})$ and $h(\vec{\theta})$ as

$$(f * h)(\vec{\theta}) = \int d^2\theta' f(\vec{\theta}')h(\vec{\theta} - \vec{\theta}'). \tag{20}$$

In Figure 13 we show the resulting reduced shear field component maps g_i, for the RRG and CNN estimators at a smoothing length of $\lambda = 19.5''$ for cluster MACSJ1720+35. We also present the field standard deviations σ_{g_i} estimated using 100 bootstrap resampling iterations of the full background galaxy catalogue. The CNN smooth shear fields for the full set of CLASH clusters are presented in Figure 17.

6.2 Relative bias calibration

To allow a comparison of the relative performance of the RRG and CNN estimators on real observations we first wish to calibrate an additional relative affine bias

Table 5. Calibration parameters for the CNN estimator on the CLASH dataset. Here c_i are the additive biases and m_i are the multiplicative biases that transform (according to Equation 16) the simulation calibrated CNN estimator such that on the CLASH dataset it is unbiased relative to the RRG catalogue.

c_1	m_1	c_2	m_2
0.000 ± 0.002	0.048 ± 0.024	0.005 ± 0.002	0.057 ± 0.024

on the CLASH observations. To do so we utilize the per-cluster shear fields $g_{i,e,\lambda}(\vec{\theta})$ and bootstrap error estimates mentioned in the previous section. In these aforementioned bootstrap iterations, we estimate the full 2×2 covariance matrix, $\Sigma_{i,\lambda}(\vec{\theta})$, at each point in the field. For each component of shear, these matrices measure both the variances of each estimator at angular position θ, as well as their mutual covariances there. We randomly sample the shear fields uniformly at 1/10'th the density of each cluster's available background sources. We do this to minimize the effect noise correlations between the sampled points would have on (underestimating) the bias parameter confidence intervals. Figure 14 illustrates this process by showing the sparse random sample evaluated in cluster MACSJ1720+35. Using the sparse samples from the full set of CLASH clusters we fit the relative bias model using an error-in-variables likelihood model (Hogg et al. 2010, §7). The resulting relative bias parameters are presented in Table 5. Figure 15 shows density plots of the RRG and CNN estimates for the full set (over all galaxy clusters) of sparse samples, after applying this additional calibration to the CNN shears.

6.3 Statistical error analysis

We wish to measure the statistical errors of each of the (mutually calibrated) estimators in the CLASH fields. To do so we first evaluate the following scatter between the smooth shear field and the shear measured at a particular background source

$$\sigma^2_{\text{field-source},i,e}(\lambda) = \mathop{\mathbb{E}}_k \left\{ \mathop{\mathbb{E}}_{j \in S_k} \left[\left(g^{\overline{S}_k}_{i,e,\lambda}(\vec{\theta}_j) - \hat{g}^{(j)}_{i,e} \right)^2 \right] \right\}. \tag{21}$$

Here the index k denotes the CLASH cluster, S_k is a randomly selected 50% hold-out set of background sources in this cluster and $g^{\overline{S}_k}_{i,e,\lambda}(\vec{\theta}_j)$ are the smooth shear fields evaluated on S_k's complement. Due to this random split of the background sources, we can assume that the smooth field noise and the per source noise are statistically independent. Using 100 bootstrap resampling iterations on the held-in background sources $\overline{S}_k$, we estimate the following average field variance at the held-out source positions

$$\sigma^2_{\text{field},i,e}(\lambda) = \mathop{\mathbb{E}}_k \left\{ \mathop{\mathbb{E}}_{j \in S_k} \left[\text{Var} \left(g^{\overline{S}_k}_{i,e,\lambda}(\vec{\theta}_j) \right) \right] \right\}. \tag{22}$$

Finally, we measure the overall statistical error of estimator e by subtracting the above two error terms

$$\sigma^2_{\text{source},i,e}(\lambda) \cong \sigma^2_{\text{field-source},i,e}(\lambda) - \sigma^2_{\text{field},i,e}(\lambda) \tag{23}$$
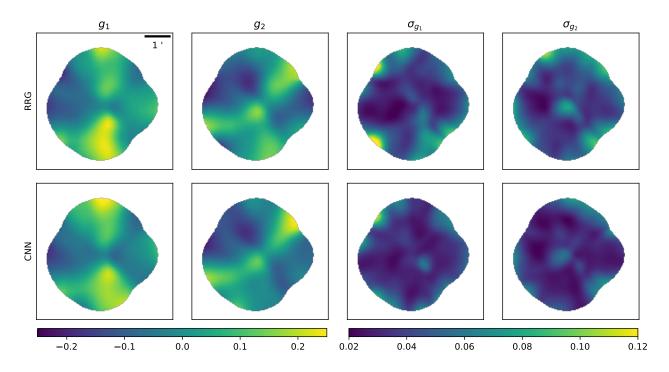
Figure 13. Reduced shear field components g_i, and standard deviations σ_{g_i}, for the RRG and CNN estimators computed using a smoothing length of $\lambda = 19.5''$ in representative CLASH cluster MACSJ1720+35.
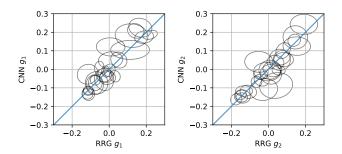

Figure 14. CNN vs. RRG shear estimates sampled at a random set of locations (at 31 points) in the CLASH cluster MACSJ1720+35 field. At each point and for each estimator and shear component, smooth field estimates as well as the 1σ bootstrap confidence ellipse is shown (a smoothing length of $\lambda = 19.5''$ was used here).

Figure 15. Density plots of RRG and (mutually calibrated) CNN smooth ($\lambda = 19.5''$) shear field sparse samples, for the combined set of CLASH galaxy clusters.

Figure 16 shows the dependence of these three statistical errors on the smoothing length. We see that both $\sigma_{\mathrm{field\text{-}source},e}(\lambda)$ and $\sigma_{\mathrm{field},e}(\lambda)$ fall with increasing smoothing length λ. This is qualitatively expected due to the reduced finite-sample-noise in the smooth field estimates at greater values of λ. The statistical errors of the per source estimates, $\sigma_{\mathrm{source},e}(\lambda)$, are approximately constant in λ. This is consistent with the fact that the per source estimate errors should be statistically independent of the field estimate errors. Overall we find RMS errors of 0.286 ± 0.003 for the RRG estimator and 0.227 ± 0.006 for the CNN estimator on the CLASH data, a reduction of approximately 26% in statistical errors. The RMS errors we measure for the RRG estimator are in agreement with the shape-noise Leauthaud et al. (2007) measures using RRG
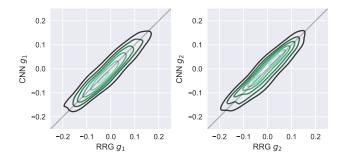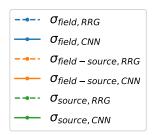
for a population having similar magnitudes. There, shape-noise was estimated using in-the-field galaxies, unaffected by cluster lensing shear. The σ_{field} curves in Figure 16 also illustrate the resulting enhancement in the effective angular resolution of the smooth field estimates at a set level of noise.

7 CODE, DATA AND REPRODUCIBILITY

We provide online[5] the python code and data used in this work. This includes procedures to match and simulate a target background source distribution, the training and testing procedures we used on the simulated data, as well as our weak lensing analysis pipeline. To enable the reader to fully
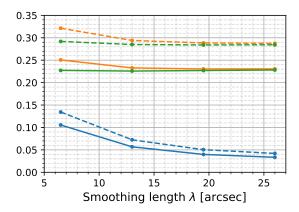
---

[5] http://github.com/ofersp/wlenet

Figure 16. Statistical errors of the smooth field estimates (σ_{field}), the mean scatter between the field and per source estimates ($\sigma_{\mathrm{field-source}}$) and the statistical errors of the per source estimates (σ_{source}). Here we present the average standard deviation of the two shear components as a function of field smoothing length λ for the RRG and CNN estimators.

reproduce our work, we release the datasets used to train and test the models described, the specification of the neural network architectures as well as our resulting trained model weights. Regarding the 20 CLASH cluster cores analysed in this work, we publish a set of tables with the updated point source classification described in §2.1.3, updated weak lensing shear catalogues (per background source) produced by our CNN model. Additionally we provide the weak lensing fields (per cluster) evaluated in §6.1 at various smoothing length scales. Finally, we publish a set of jupyter notebooks[6] (Kluyver et al. 2016) with which we performed the analysis and generated the figures included in this paper.

8 DISCUSSION

We present a machine learning process that allowed us to train a model to perform weak lensing shear estimation directly from simulated examples. This process includes: Simulating galaxy stamps having known shear and an observational distribution matching that of the target stamps, training a *deep convolutional neural network* on the simulated data to produce a model and an assessment of the model performance on both the simulated data and actual galaxy cluster field observations.

This approach is novel in several respects: The model is trained to directly estimate lensing shear and not galaxy

---

[6] https://jupyter.org

shapes, it is not based on hand-crafted features but instead attempts to extract all relevant information from the galaxy observations (in the limits of the model architecture) and it is designed to be invariant to the stamp centroid and background estimates.

We dealt with the following challenges that were specific to our approach: Avoiding simulation artefacts which would otherwise deteriorate model generalization to actual observations, enlarging and augmenting the training data to promote wanted invariance properties at the learning phase and enforcing known discrete symmetries of lensing shear at the inference phase. In the observed CLASH fields, an additional challenge was to perform bias calibration and statistical error analysis where the true per-stamp shears are unknown.

The results of our learnt estimator appear to be consistent with those of commonly used quadrupole moment estimators and show enhanced statistical performance. We find this to be the case both on the simulations and in the real observations where the reduction in RMS errors, compared to the RRG estimator, is at the level of 26%. This is equivalent to a relative improvement in survey speed of approximately 60%. This serves as initial indication that what is commonly referred to as the *shape-noise limit* can likely be overcome, at least to some extent, by using the joint statistics of higher order intensity moments of high-redshift galaxy images (above the second order moment statistics tapped by RRG).

These results however require additional testing and validation due to the non-transparent nature of the machine learning approach. Although we have made great efforts to rid our simulations of artefacts, these may still exist in our training and test datasets and could potentially be providing the learnt estimator with clues as to what the simulated shears were — clues that would not be available in real observations. Additional simulations by independent means would allow us to mostly rule out this possibility. Furthermore, studying the error sensitivity (statistical and systematic) of our estimator to additional distributional factors, such as galaxy morphology (using e.g. the Shapelet decomposition of Refregier 2003, at increasing order), S/N levels and PSF parameters, is still required. This would enable us to disentangle the various sources of statistical error (e.g. morphology from Poisson noise) and to control potential systematic biases when applying this learnt estimator to different real observing conditions.

Using high dimensional statistics to estimate shear can either be done explicitly, by modelling the conditional distribution (conditioned on shear) of post-shear galaxy images — this is known as a *generative approach* to machine learning, or implicitly, by learning to fit a function that directly maps post-shear images of galaxies to their true shear, in a statistically consistent and efficient way — known as a *discriminative approach*. This later approach is the one we eventually followed in this work. This choice requires us to have good simulations at our disposal (Rowe et al. 2015) but allows us to avoid analytically modelling the action of shear on the pre-sheared galaxy image distribution. Training our model in a way that promotes invariance to certain types of transformations is also simplified by
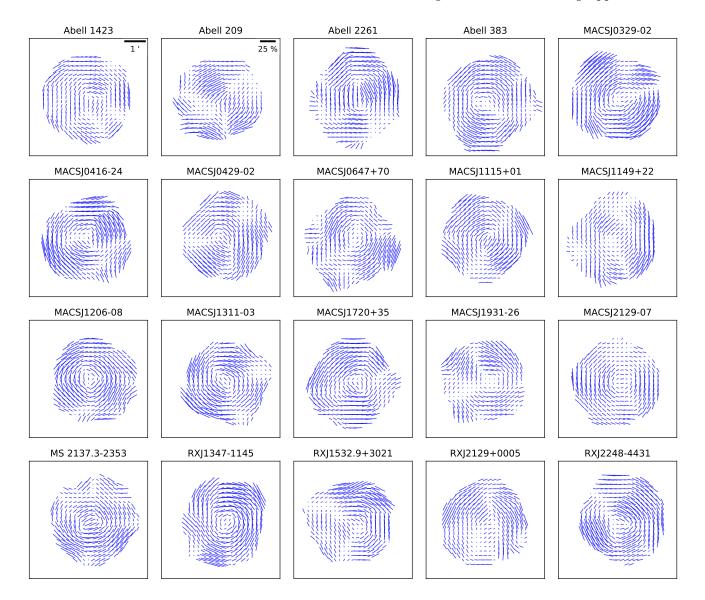
Figure 17. Reduced shear fields, $g_{i,\lambda}(\vec{\theta})$, for the CNN estimator, computed using a smoothing length of $\lambda = 19.5''$ in the 20 CLASH clusters analysed. In all clusters J2000 north is up and both the angular scales and the shear scales are identical. An angular scale-bar corresponding to $1'$ is shown next to Abell 1423 and a scale-bar corresponding to a reduced shear of 25% is shown next to Abell 209. The lengths and directions of the sticks, at each point, correspond to the magnitude and angle ϕ of the complex representation $g_{1,\lambda} + i g_{2,\lambda} = |g_\lambda| e^{2i\phi}$ of the reduced shear field.

the discriminative approach — achieving this only requires us to simulate these transformations when preparing the training data.

We list the following caveats of our approach:

• Our estimator does not presently deal with PSF variation. We believe this does not affect the validity of our statistical error analysis, particularly on the simulated test data, but could in principle be leading to local biases of the estimator within the observed field-of-view in the CLASH data — possible biases which we have not accounted for. We believe this issue can be overcome by training a model to infer shear given both a galaxy stamp and a local estimate of the PSF, such as those made available by the TINYTIM

PSF model of the *HST* fields (Krist et al. 2011). We defer this issue to future work.

• Assuming a total galaxy number density of 37 arcmin^{-2} (see Chang et al. 2013) we estimate that in the well resolved CLASH data ~4% should contain blends within a $32 \times 32\,\mathrm{px}$ stamp, although the vast majority of these blends would be of all-background galaxies. For all-background blends, a similar shear would affect the whole stamp, as is the case in a similar proportion of the stamps in the COSMOS based simulations. In addition, blends including foreground galaxies typically manifest as constant or nearly constant additive background levels within the $32 \times 32\,\mathrm{px}$ background stamps. As discussed in §5.2, our estimator is expected to be robust to this type of blending. Stamps with blends that include relatively small foreground galaxies may introduce

some systematic biases which we have not accounted for. Although we do not expect blending to introduce a significant effect in the CLASH observational regime, simulating these foreground-background blends in the training data could allow us to both train the CNN estimator to be robust to blending issues and to assess its sensitivity to blending. We defer this refinement of our simulation process to future work.

• Our simulations are lacking in that we do not match the simulated and observed populations in redshift, but only in magnitude-radius space (see §3.2). This choice could in principle be overcome, to a large extent, when deeper observations of in-the-field galaxies become available. This is due to the fact that the higher the S/N of the base dataset used by GalSim, the closer the image rescaling factor (α in §3.1) can be to unity. The availability of deeper observations at the training phase, particularly when provided in the specific band we later wish to measure weak lensing shear in, will also allow us to avoid the band mismatch we opt for in §2.1.1.

• The reliance of our approach on high-fidelity image simulations adds difficulty to its application. This is similarly the case with conventional simulation-based shear calibration techniques and the potential advantage of data-based calibration (e.g. Huff & Mandelbaum 2017).

• When producing the training data for our learnt estimator we performed a set of transformations, or augmentations, in addition to the simulated shear. The choice of augmentations was not particularly tuned to the CLASH dataset which we later ran our estimator on, but these were chosen after we were exposed to the CLASH data. As previously discussed, testing our model on additional simulations as well as on real galaxy cluster observations are needed to further validate our approach.

Our technique was developed and tested in the regime of well-resolved space-based images. This was done with particular aim at galaxy cluster weak lensing measurements. The applicability of this technique to the type of ground-based images used in cosmic shear measurements is currently not known and remains to be studied. It is likely that the lack of visible substructure in the ground-based image regime would lead to little gain for this approach.

Finally, we would like to stress the point raised in §4.1 — we are currently unaware of any justified strategies to choose an optimal model architecture, model weight optimization scheme and training data augmentation scheme. We would like to encourage the reader to use the code and data released in this work as a basis for future experimentation, hopefully, reaching lower statistical errors and further enhancing resolving power in real galaxy cluster weak lensing observations.

APPENDIX A: CNN LAYERS

In the architecture of §4.1, a *convolutional layer*, denoted $\mathtt{Conv}(n, m) \times l$, transforms an input tensor $I_{x,y,c}$, having dimensions $n_i \times m_i \times l_i$, to an output tensor $O_{x,y,c}$ having dimensions $n_o \times m_o \times l$

$$O_{x,y,c} = B_c + \sum_{x'=1}^{n} \sum_{y'=1}^{m} \sum_{c'=1}^{l_i} F_{x',y',c'}^{(c)} I_{x+x',y+y',c'} , \tag{A1}$$

where $w_{\mathrm{conv}} = (B_d, F_{x,y,c}^{(d)})$ are the bias and filter weight parameters associated with this layer.

An *affine layer*, denoted $\mathtt{Affine}(l)$, transforms an input tensor I_k (or a tensor $I_{x,y,c}$ that has been flattened appropriately), having dimension n_i, to an output tensor O_j having dimension l

$$O_j = B_j + \sum_{k=1}^{n_i} A_{jk} I_k , \tag{A2}$$

where $w_{\mathrm{affine}} = (B_j, A_{jk})$ are the bias and linear weight parameters of this layer.

Convolutional layers, in our architecture, also perform the $\mathtt{Stride}(\Delta x, \Delta y)$ operation, which downsamples an input tensor $I_{x,y,c}$ to an output tensor $O_{x,y,c}$ in the following way

$$O_{x,y,c} = I_{x,y,c} \downarrow_{\Delta x, \Delta y} . \tag{A3}$$

Both convolutional and affine layers, in our architecture, perform the point-wise non-linear operation defined by $O = \mathtt{ReLU}(I) = \max(0, I)$.

Finally, a training technique that promotes generalization (enhances the performance of the resulting model on the test set) known as Dropout (Srivastava et al. 2014) is employed in some of the layers of Table 2.

APPENDIX B: REGRESSION ATTENUATION

Here we show that performing linear least squares regression in the inverse direction, that is, from noisy features to clean signal, produces a biased estimator (a bias known in statistics as *regression attenuation*). Under some model assumptions regarding the joint probability of noise and signal we find a simple correction factor that transforms the inverse regression estimator to the conditionally unbiased linear estimator having least variance.

In this appendix we use $x \in \mathbb{R}$ to denote the signal we wish to estimate and $y \in \mathbb{R}^d$ to denote the d-dimensional noisy feature vector which we measure. To simplify the discussion we assume the signal x has a prior distribution having zero mean, $\mathbb{E}[x] = 0$. The components of reduced shear

in §4.2 are represented here by x and this zero mean assumption is compatible with how the reduced shear was prepared in the training data there. Our model for how the noisy features are generated is the following

$$y = sx + \eta, \tag{B1}$$

where $s \in \mathbb{R}^d$ is some set of constant coefficients and $\eta \in \mathbb{R}^d$ is noise uncorrelated with the signal, $\mathbb{E}[x\eta] = 0$. We also assume the noise has zero mean for every x, $\mathbb{E}[\eta|x] = 0$. We found that both of these properties approximately hold in §4.2 for the noisy features after a constant mean is subtracted. We now confine our estimator $\hat{x}$ of the signal x to be some linear function of the noisy features y,

$$\hat{x} = a^T y, \tag{B2}$$

for some set of constant coefficients $a \in \mathbb{R}^d$.

For a set of N samples of signals $X = (x_1, x_2, \cdots, x_N)^T$ and matching features $Y = (y_1, y_2, \cdots, y_N)^T$ we can write equation B1 in matrix form as

$$Y = Xs^T + W, \tag{B3}$$

with $W = (\eta_1, \eta_2, \cdots, \eta_N)^T$. Performing linear-least-squares regression from the noisy features to the clean signal produces the following estimator coefficients

$$a_{\mathrm{MMSE}} = \arg\min_a L(a) = \arg\min_a \sum_{i=1}^{N} \left(a^T y_i - x_i \right)^2. \tag{B4}$$

This is the minimum-mean-squared-error (MMSE) estimator which can be shown to have the explicit form

$$a_{\mathrm{MMSE}} = \left(Y^T Y \right)^{-1} Y^T X \tag{B5}$$

when $Y^T Y$ is invertible. For a sufficiently large number of samples, this can be shown to be

$$a_{\mathrm{MMSE}} = \frac{1}{N} C_y^{-1} Y^T X \tag{B6}$$

$$= \frac{1}{N} C_y^{-1}(sX^T + W^T)X \tag{B7}$$

$$= C_y^{-1}\left(\frac{1}{N} sX^T X + \frac{1}{N} W^T X\right) \tag{B8}$$

$$= C_y^{-1} \{\mathrm{Var}(x)s + \mathbb{E}[x\eta]\} \tag{B9}$$

$$= \mathrm{Var}(x)C_y^{-1}s, \tag{B10}$$

where C_y is the $d \times d$ covariance matrix of y.

Different estimator coefficients would have resulted if we would have looked for the conditionally unbiased linear estimator — the $\hat{x}$ for which

$$\forall x \quad x = \mathbb{E}[\hat{x}|x] = a^T s x, \tag{B11}$$

having minimal variance

$$\mathrm{Var}(\hat{x}) = \mathbb{E}\left[(\hat{x} - \mathbb{E}[\hat{x}])^2\right] = a^T C_y a. \tag{B12}$$

These two last requirements lead to the following constrained minimization problem

$$\underset{a}{\text{minimize}} \quad a^T C_y a$$

$$\text{subject to} \quad a^T s = 1,$$

for constant C_y and s. Using the method of Lagrange multipliers, this can be shown to have the solution

$$a^* = \frac{C_y^{-1}s}{s^T C_y^{-1}s}. \tag{B13}$$

We see that the ratio of these two sets of estimator coefficients, B6 and B13, is the following scalar factor

$$\frac{a^*}{a_{\mathrm{MMSE}}} = \frac{\mathrm{Var}(x)^{-1}}{s^T C_y^{-1} s}. \tag{B14}$$

In the large sample number limit, this ratio is

$$\frac{a^*}{a_{\mathrm{MMSE}}} = \left[1 - \frac{L(a_{\mathrm{MMSE}})}{\mathrm{Var}(x)}\right]^{-1}. \tag{B15}$$

REFERENCES

Abadi M., et al., 2016, in OSDI. pp 265–283

Bardeau S., Kneib J.-P., Czoske O., Soucail G., Smail I., Ebeling H., Smith G., 2005, Astronomy & Astrophysics, 434, 433

Bartelmann M., Schneider P., 2001, Physics Reports, 340, 291

Benitez N., 2000, The Astrophysical Journal, 536, 571

Bernstein G. M., Armstrong R., Krawiec C., March M. C., 2016, Monthly Notices of the Royal Astronomical Society, 459, 4467

Bertin E., Arnouts S., 1996, Astronomy and Astrophysics Supplement Series, 117, 393

Bishop C. M., 1994

Blain A. W., 2002, The Astrophysical Journal Letters, 570, L51

Brown M. L., Battye R. A., 2011, The Astrophysical Journal Letters, 735, L23

Chang C., et al., 2013, Monthly Notices of the Royal Astronomical Society, 434, 2121

Chollet F., et al., 2015, Keras, https://keras.io

Croft R. A., Freeman P. E., Schuster T. S., Schafer C. M., 2017, Monthly Notices of the Royal Astronomical Society, 469, 4422

Fenech Conti I., Herbonnet R., Hoekstra H., Merten J., Miller L., Viola M., 2017, Monthly Notices of the Royal Astronomical Society, 467, 1627

Ford H. C., et al., 2003, in Future EUV/UV and Visible Space Astrophysics Missions and Instrumentation. pp 81–95

Gruen D., et al., 2018, arXiv preprint arXiv:1809.04599

Hogg D. W., Bovy J., Lang D., 2010, arXiv preprint arXiv:1008.4686

Huff E., Mandelbaum R., 2017, arXiv preprint arXiv:1702.02600

Huff E. M., Krause E., Eifler T., George M. R., Schlegel D., 2013, arXiv preprint arXiv:1311.1489

Kaiser N., Squires G., Broadhurst T., 1994, arXiv preprint astro-ph/9411005

Kluyver T., et al., 2016, in ELPUB. pp 87–90

Koekemoer A. M., Fruchter A. S., Hook R., Hack W., 2003, in HST Calibration Workshop: Hubble after the Installation of the ACS and the NICMOS Cooling System. p. 337

Koekemoer A. M., et al., 2007, The Astrophysical Journal Supplement Series, 172, 196

Koekemoer A. M., et al., 2011, The Astrophysical Journal Supplement Series, 197, 36

Komatsu E., et al., 2011, ApJS, 192, 18

Krist J. E., Hook R. N., Stoehr F., 2011, in Optical Modeling and Performance Predictions V. p. 81270J

LeCun Y., Bengio Y., Hinton G., 2015, nature, 521, 436

Leauthaud A., et al., 2007, The Astrophysical Journal Supplement Series, 172, 219

Mandelbaum R., Hirata C. M., Leauthaud A., Massey R. J., Rhodes J., 2012, Monthly Notices of the Royal Astronomical Society, 420, 1518

Mandelbaum R., et al., 2014, The Astrophysical Journal Supplement Series, 212, 5

Mandelbaum R., et al., 2015, Monthly Notices of the Royal Astronomical Society, 450, 2963

Mandelbaum R., et al., 2018, Monthly Notices of the Royal Astronomical Society, 481, 3170

Massey R., Stoughton C., Leauthaud A., Rhodes J., Koekemoer A., Ellis R., Shaghoulian E., 2009, Monthly Notices of the Royal Astronomical Society, 401, 371

Massey R., et al., 2012, Monthly Notices of the Royal Astronomical Society, 429, 661

Merten J., et al., 2015, The Astrophysical Journal, 806, 4

Molino A., et al., 2017, Monthly Notices of the Royal Astronomical Society, 470, 95

Morales M. F., 2006, The Astrophysical Journal Letters, 650, L21

Niemi S.-M., Kitching T. D., Cropper M., 2015, Monthly Notices of the Royal Astronomical Society, 454, 1221

Postman M., et al., 2012, The Astrophysical Journal Supplement Series, 199, 25

Refregier A., 2003, Monthly Notices of the Royal Astronomical Society, 338, 35

Rhodes J., Refregier A., Groth E. J., 2000, The Astrophysical Journal, 536, 79

Rhodes J. D., et al., 2007, The Astrophysical Journal Supplement Series, 172, 203

Rowe B., et al., 2015, Astronomy and Computing, 10, 121

Samuroff S., et al., 2017, Monthly Notices of the Royal Astronomical Society, 475, 4524

Schneider P., 2006, in , Gravitational Lensing: Strong, Weak and Micro. Springer, pp 269–451

Scoville N., et al., 2007, The Astrophysical Journal Supplement Series, 172, 38

Sheldon E. S., Huff E. M., 2017, The Astrophysical Journal, 841, 24

Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R., 2014, The Journal of Machine Learning Research, 15, 1929

Tieleman T., Hinton G., 2012, COURSERA: Neural networks for machine learning, 4, 26

Zuntz J., et al., 2018, Monthly Notices of the Royal Astronomical Society, 481, 1149

de Burgh-Day C. O., Taylor E. N., Webster R. L., Hopkins A. M., 2015, Monthly Notices of the Royal Astronomical Society, 451, 2161

This paper has been typeset from a TeX/LaTeX file prepared by the author.