Galaxy shape measurement with convolutional neural networks

Dezső Ribli¹*, László Dobos¹ and István Csabai¹
¹Eötvös Loránd University, Institute of Physics, Budapest, Hungary

ABSTRACT

We present our results from training and evaluating a convolutional neural network (CNN) to predict the shapes of galaxies from wide field survey images. As “ground truth”, we use conventional shape measurements from an overlapping superior, deeper survey with less sky coverage. In particular, we predict galaxy shapes from images of the DR1 data release of Dark Energy Survey (DES) using shape measurements from the Canada-France Hawaii Telescope Lensing Survey (CFHTLenS). We demonstrate that CNN predictions from DES images reproduce the results of CFHTLenS at bright magnitudes and show a significantly higher correlation with CFHTLenS at fainter magnitudes than DES results from conventional image analysis. The more accurate shape measurements with the CNN increase the effective galaxy density of DES by 35% compared to the official DES Y1 metacalibration catalogue and by 150% compared to the im3shape catalogue. Prediction of shape parameters with a CNN is also extremely fast, it takes only 0.2 milliseconds per galaxy, improving 4 orders of magnitude over model fitting. Our proposed setup is applicable to current and next generation weak lensing surveys where higher quality “ground truth” shapes can be measured in dedicated deep fields.

Key words: gravitational lensing: weak – techniques: image processing – cosmology: dark matter

1 INTRODUCTION

Light follows space-time geodesics and gets deflected by the gravitational potential of the matter density field, resulting in gravitational lensing. Detecting the effects of gravitational lensing in galaxy images caused by foreground mass concentrations allows us to indirectly observe the elusive, and apparently very abundant dark sector of the Universe. Measuring the total matter distribution at different redshifts through its lensing signal offers a unique window to the evolution of the dark energy dominated late Universe, which is complementary to other observations (Kilbinger 2015).

When the lensing effect of foreground matter is strong enough, it can dramatically alter the appearance of background sources, making the lensing signal detectable even for individual galaxies. More diffuse dark matter, however, only distorts the shapes of background objects with a small, approximately linear shear, rendering the lensing signal indistinguishable from the intrinsic shape variations of the lensed galaxies. The shapes of galaxies in the same cosmic neighborhood are similarly distorted when their light travels through the same matter, making cosmic shear statistically measurable from an ensemble of galaxies. The spatial correlations in the apparent shapes of galaxies introduced by the large scale structure of the Universe allow us to indirectly map the distribution of dark matter (Kilbinger 2015).

Since the first detection of gravitational lensing due to large scale structure about two decades ago (Bacon et al. 2000; Kaiser et al. 2000; Van Waerbeke et al. 2000; Wittman et al. 2000) cosmic shear measurements have matured into a potent probe of cosmology. Weak lensing survey volumes increased constantly, with the COSMOS field and Hubble Space Telescope images (Schrabback et al. 2010) followed by CFHTLenS ¹ which was the first major weak lensing survey pushing the number of galaxies with high-quality shape measurements to the millions by covering an area of 154 square degrees at a resolved galaxy density of 17 per square arc minute. The shapes of hundreds of millions of galaxies are being measured by ongoing weak lensing surveys, such as the Dark Energy Survey (DES) ², the Kilo-Degree Sur-

* E-mail: dkrib@caesar.elte.hu (DR)

¹ http://www.cfhtlens.org
² https://www.darkenergysurvey.org

© 2019 The Authors
The complex ellipticity parameter

$$\epsilon_1 + i \epsilon_2 = \frac{a - b}{a + b} e^{i 2 \theta}$$

where $a$ is the length of the semi-major axis, $b$ is the length the semi-minor axis and $\theta$ is the position angle of a perfect ellipse.

The bulk of the cosmic shear signal is carried by the shapes of small and faint galaxies typically those with sizes of a few arcsecs and $\approx 24$ magnitudes (Kilbinger 2015). Weak lensing surveys, therefore, need to attempt to estimate the shapes of galaxies which only cover a few pixels and have very low signal to noise ratios ($S/N$).

The redshifts of galaxies also need to be determined for the optimal extraction of the cosmological signal, but this is not possible with spectroscopy due to the vast number of galaxies used, therefore weak lensing surveys rely on photometric redshift estimates from multi-band observations. The shapes of galaxies are highly correlated in different colors and Jarvis & Jain (2008) showed in simulations that combining information from multiple bands can improve the signal to noise ratio of shape measurements, and effectively increase the number of useful of galaxies, a claim also demonstrated in observations (Zuntz et al. 2018). On the other hand, CFHTLenS and KiDS surveys preferred to reserve the best observing conditions for a single band which is then used alone for the shape estimation (Erben et al. 2013; de Jong et al. 2017).

Shape measurements are complicated due to the smearing of original images with the point spread function (PSF) originating from the telescope, the camera and the atmosphere for ground-based observations. The PSF itself has a coherent spatially and temporally variable anisotropy which mimics the effect of gravitational lensing, making shape measurements especially problematic for ground-based surveys. Spurious correlations in the lensing signal due to the PSF must be detected and corrected by cross-correlating the observed galaxy shears with the ellipticity of the PSF (Heymans et al. 2012).

Shape measurement algorithms fall in two main categories, the first approach attempts to directly measure the shapes of galaxies through second-order moments of the surface brightness profile (Kaiser et al. 1995). Another approach, dominantly used by recent weak lensing surveys, uses model fitting, where parametric surface brightness model profiles are convolved with the PSF and compared to the measurements (Kuijken 1999; Miller et al. 2007; Kitching et al. 2008; Miller et al. 2013; Kuijken et al. 2015; Hildebrandt et al. 2017; Fenech Conti et al. 2017; Zuntz et al. 2018). Model fitting is preferred over measuring moments due to the effective handling of the PSF especially during a joint fit over multiple exposures (Heymans et al. 2012).

Even the model fitting approach which uses a forward convolution of the model with a PSF cannot eliminate its effect on shape measurements. Faint galaxies, with signal to noise ratios ($S/N$) around 10 or less, or small galaxies with shapes comparable to the PSF suffer from a significant noise bias introduced by the PSF via skewing the likelihood surface towards zero and towards the ellipticity of the PSF itself (Bernstein & Jarvis 2002; Kitching et al. 2010a; Melchior & Viola 2012; Refregier et al. 2012; Kacprzak et al. 2012; Miller et al. 2013). The bias in ellipticity or shear measurements is usually described with a linear function following Heymans et al. (2006)

$$\epsilon_i^{true} = (1 + m_i) \epsilon_i^{obs} + c_i$$

where the intercept $c_i$ is called the additive bias, and the excess slope $m_i$ is the multiplicative bias, $i$ denotes the component of the ellipticity (1,2). The additive bias can be corrected using measurement data (Heymans et al. 2012), but the multiplicative term generally needs to be remedied very carefully with simulations.

The multiplicative bias strongly depends on the observed $S/N$ and size of galaxies and the precise relationship is generally identified through estimating the properties of millions of galaxy images simulated with sophisticated tools (Rowe et al. 2014) and known shape parameters (Miller et al. 2013; Jarvis et al. 2016; Fenech Conti et al. 2017; Zuntz et al. 2018; Pujol et al. 2019). These simulations need to be specifically tailored to fit the details of the surveys, and the calibration process has to be conducted for every single survey or data release. After the precise nature of the bias is established, empirical corrections of the measured galaxy ellipticities are applied to negate the bias. A new inventive version of this methodology called “self-calibration” first estimates this bias of single galaxies using a simulated version of the galaxy with the same parameters. Self-calibration was successfully applied in the KiDS450 shape catalogue, however, regular calibration was also used to further mitigate the bias (Fenech Conti et al. 2017). Another interesting novel approach is METACALIBRATION which only uses observation data to correct for noise bias (Huff & Mandelbaum 2017; Sheldon & Huff 2017). In METACALIBRATION the shape of the original galaxy is measured and compared to measurements made on versions of the galaxy which are deconvoluted with the PSF, sheared with a small factor and reconvoluted again with the PSF. The differences of the measured shapes define the response of shape measurement to a known shear. The METACALIBRATION scheme was successfully applied to produce the DES Y1 shape catalogue (Zuntz et al. 2018).

Selection biases also appear when galaxies are preferentially detected depending on the alignment of their shapes compared to the anisotropy of the PSF or the lensing shear.
(Kaiser 2000; Hirata & Seljak 2003). A version of the selection bias may also affect detected galaxies due to the fact that the ellipticity likelihood surface is narrower for galaxies if the intrinsic and PSF ellipticities are aligned. Methods which assign an inverse-variance weight to shape measurements based on the width of the likelihood surface (Miller et al. 2013; Fenech Conti et al. 2017), favor galaxies parallel to the PSF, creating a selection bias in the shear.

Shape measurement approaches which assume simplified galaxy surface brightness profiles potentially suffer from model bias if they are not able to capture the shapes of complex galaxies (Voigt & Bridle 2010; Melchior et al. 2010). Model bias is expected to have only a minor contribution in ground-based observations (Miller et al. 2013; Mandelbaum et al. 2015). It was explored in detail for model fitting shape measurements (Zuntz et al. 2013; Kacprzak et al. 2014).

Biases in the shape measurements could systematically alter the cosmology inferred from weak lensing measurements and the community understandably paid particular attention to the question of systematic biases, also in the form of organized challenges (Heymans et al. 2006; Massey et al. 2007; Bridle et al. 2009, 2010; Kitching et al. 2010b,a; Simet et al. 2014; Mandelbaum et al. 2015). However, shape measurements not only need to be unbiased but they also need to have small variance in order to reduce statistical uncertainty, a quantity often expressed in terms of an effective galaxy surface density (Heymans et al. 2012; Chang et al. 2013). Large efforts are ongoing and planned (HSC, LSST, Euclid, WFIRST) to conduct very deep surveys which aim to radically increase the galaxy surface density and the statistical power of measurements demonstrating the importance of reducing the variance of shape measurements (Takada 2010; Chang et al. 2013; Amendola et al. 2018). Increased galaxy densities enable more precise measurements of the cosmic shear at high angular resolutions, allowing the extraction of non-Gaussian information, which can further empower weak lensing measurements (Ribli et al. 2019a).

The variance of shape measurements received much less attention than their bias, probably because one might think that model fitting with forward convolution is the optimal method and only incremental improvements can be expected. We also think that in measurements with completely well-known noise and signal characteristics model fitting is optimal, however, the noise in galaxy shape measurements is known to be not a completely independent shot noise (Mandelbaum et al. 2015) and model fitting approaches need to restrict and simplify the surface brightness profiles (Miller et al. 2013).

Therefore we think it is indeed possible that methods which can utilize the true characteristics of the noise together with more realistic surface density profiles could decrease the variance of galaxy shape measurements. The effects of systematic biases in current surveys are significantly smaller than statistical uncertainties (Mandelbaum et al. 2015; Zuntz et al. 2018; Fenech Conti et al. 2017), therefore shape measurements with smaller variance could reduce the uncertainty of shear measurements. Even if biases are not completely eradicated, higher statistical power also helps to deepen the understanding of these biases.

Our study focuses on reducing the statistical uncertainty, the variance of shape estimates, via convolutional neural networks (CNN) which recently revolutionized the field of computer vision and reached the level of human accuracy in image classification (He et al. 2015). CNNs are sophisticated machine learning models which are able to learn from a large number of labeled images, and apart from image classification they also excel in image deconvolution or “super-resolution” in the presence of noise for everyday images (Xu et al. 2014) or microscopy (Wang et al. 2019). We construct a large labeled dataset using images from a wide survey, DES DR1 (Abbott et al. 2018; Morganson et al. 2018), and high quality galaxy shape measurements from a significantly deeper survey with overlapping footprint, CFHTLenS (Hildebrandt et al. 2012; Erben et al. 2013; Heymans et al. 2013; Miller et al. 2013). The CNN is trained to predict the ellipticities of galaxies measured by the deeper survey using the images of the wide survey as inputs. Trained on real images of galaxies, the CNN can potentially learn the characteristic of noise and implicit models and priors of galaxy surface densities, possibly overcoming the simplifications of model fitting approaches. One could also try to create a training dataset from simulated galaxies, however, we prefer to use observational data for this present study, which allows us to make no compromises in this respect. Another strength of CNNs is that they make use of massively parallel computer hardware (GPU, TPU) which allow sub-millisecond execution times per galaxy, improving orders of magnitudes upon model-fitting approaches.

Multiple studies explored recently machine learning and CNNs in problems related to weak gravitational lensing or the estimation of the properties of galaxies. Dieleman et al. (2015) used custom rotation invariant CNNs for galaxy morphology prediction using “ground truth” values determined by citizen scientist in the Galaxy Zoo project (Lintott et al. 2008). CNNs were used to infer the morphological parameters of simulated galaxies (Tuccillo et al. 2017), with an improved catalogue released for SDSS (Dominguez-Sánchez et al. 2018). Tewes et al. (2019) use machine learning to predict shear from on the outputs of conventional shape fitting algorithms complementary to our approach. Herbel et al. (2018) use CNNs for the estimation of PSF shape parameters to enable fast PSF modeling, and Springer et al. (2018) train CNNs to directly predict galaxy shear in simulations for galaxy cluster lensing and apply it to observational data.

In §2 we describe the dataset used for training and testing, and §3 the details of the CNN architecture and the training process are explained. In §3 we compare the estimated galaxy shapes to the “ground truth” measured by CFHTLenS and the DES Y1 shape catalogues which use model fitting. We evaluate the accuracy of the CNN and the model fitting approaches depending on various measured factors such as magnitude, shape, color or PSF size. We calculate effective galaxy surface densities and we also test the accuracy of the CNN depending on the number of galaxies used for training it.

2 DATA

We select galaxies from the DES Y1 shape catalogues ([IMJSHAPE,METACALIBRATION]) (Zuntz et al. 2018) which are also found in the overlapping W4 region of CFHTLenS. The DES Y1 shape catalogues are cross-matched with the CFHTLenS shape catalogue (Erben et al. 2013) using Sky-
Figure 1. Example galaxies with DES (left) and deeper CFHTLenS (right) images in the $i$-band. CFHTLenS images have higher S/N and they allow more accurate shape measurements of the same galaxies. The galaxies have $i$-band magnitudes (22.05, 22.09, 22.18), top to bottom.

and Y3 DES operations concentrated on other regions of the sky (Diehl et al. 2016).

We use highly confident ellipticity parameters measured with lensfit using the CFHTLenS $i$ band images (Miller et al. 2013) as the “ground truth” labels to train our neural network. The CFHTLenS survey is approximately 1 magnitude deeper than DES DR1 (Erben et al. 2013; Abbott et al. 2018), which allows much better shape measurements for faint galaxies. The difference in image quality for faint galaxies is demonstrated visually with examples on [Fig. 1].

We compare the performance of our CNN to the official shape DES Y1 catalogues (Zuntz et al. 2018). The im3shape catalogue uses the similarly named maximum likelihood method (Zuntz et al. 2013) and relies only on the $r$ band (Jarvis et al. 2016), while the metacalibration catalogue uses joint fitting in the $ri$z bands with the ngmix engine (Sheldon 2015). We use the metacalibration catalogue for most of our experiments, because we find it to be significantly more accurate than the im3shape catalogue, the superiority of the metacalibration catalogue is also discussed in Zuntz et al. (2018).

For calibration purposes, the “accuracy” of shape measurement algorithms is often characterized with the values of the multiplicative and the additive bias parameters. However, in this study we set out to reduce the variance, the statistical uncertainty of shape measurements, therefore we choose a metric which reflects the covariance of the “ground truth” and the estimated ellipticities: the Pearson correlation coefficient.

$$\rho(x, y) = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y}$$  \hfill (3)

We do acknowledge the immense importance of bias correction, which must be thoroughly performed for credible shear measurements, but we expect that the CNN shape estimates can just as well be calibrated using a large number of realistic image simulations or metacalibration as any other well-behaving shape estimator. For the galaxy level comparisons performed in the present study, bias corrections are negligible compared to intrinsic ellipticities of galaxies. Note, that the “ground truth” ellipticities and the baseline catalogues are calibrated (Miller et al. 2013; Zuntz et al. 2018).
We split the dataset into a 70% training and a 30% test set based on the position on the sky [Fig. 2], in order to evaluate the capabilities of our CNN to make predictions on a new region, other than the one used for training. Generalization to other regions of the sky is essential in our proposed scheme where the model is trained and evaluated on the small deep fields of surveys but deployed to predict on the whole survey.

3 CONVOLUTIONAL NEURAL NETWORK

We design a custom CNN architecture specifically for the task, building on successful image classification CNNs (Krizhevsky et al. 2012; Simonyan & Zisserman 2014; Szegedy et al. 2015; He et al. 2016; Redmon & Farhadi 2017). The network consists of subsequent sliding window filter matching operations, called convolutional layers, which can powerfully express the translation invariance of image data. The width of convolutional filters is 3 × 3 pixels for the majority of layers, as this filter width makes the most effective use of parameters (Simonyan & Zisserman 2014). Convolutional layers with at least 256 filters are preceded by a bottleneck 1 × 1 “convolutional” layer which compresses the input representations to half the number of filters to save computation time and to reduce the number of overall parameters. These bottleneck layers are found in almost every single modern CNN architecture (Szegedy et al. 2015; He et al. 2016; Redmon & Farhadi 2017).

Each convolutional layer, except for the last one, is followed by batch normalization (Ioffe & Szegedy 2015), which rescales the activations in the previous layer in order to stabilize and facilitate training. Batch normalizations are followed by non-linear activation functions called Rectified Linear Units (ReLU), which take the form \( \max(0, x) \). The flat part introduces non-linearity, which allows the neural network to model complex non-linear functions, while the linear region provides stable, non-vanishing gradients when propagated through very deep networks (Krizhevsky et al. 2012). After every 2-3 convolutional layers, the representations are spatially downsampled with max-pooling, which replaces 2 × 2 blocks with their maximum values. Downsampling helps to aggregate localized lower level information into higher semantic levels with less accurate spatial information.

When the spatial extent of the convolutional layers becomes very small (3 × 3) we average all activations along the spatial dimensions, creating a single one-dimensional representation. Finally, a linear layer with 2 outputs predicts the ellipticity components of the galaxy. The outline of the neural network is detailed in [Tab. 1].

The inputs to the neural network are 50 × 50 pixel gizY 5-band postage stamp images with pixel values rescaled to have zero mean and unit standard deviation in every band. We do not incorporate pixel weights or masks into our inputs, however, we note that it would be straightforward to simply stack these as additional channels of the input image.

During training, we augment the dataset with random horizontal and vertical flips and random transpositions to combat overfitting. Naturally, the “ground truth” ellipticities of the galaxies are transformed accordingly during augmentation. The augmentation scheme creates an 8 × larger dataset, however, the new examples are not independent. The applied transformations enable the neural network to learn additional symmetries of the dataset alongside the translation invariance represented with convolutions. An interesting branch of research attempts to create neural networks with built-in representations for rotational symmetries (Cohen & Welling 2016; Kondor et al. 2018), however, these works have not completely matured yet. We only use augmentation during training, but not during testing, as we find no significant improvement from test time augmentation, and it makes inference significantly slower.

We train the neural network with stochastic gradient descent optimization with an initial learning rate of 0.005. We train for 40 complete iterations on the training dataset, called an epoch, and we decrease the learning rate tenfold after the 20th and the 30th epoch to enable convergence by the end of the training schedule. We use a minibatch size of 512, and the traversal order or the training dataset is shuffled before each epoch to create a varying composition of minibatches in each iteration, which has a regularizing effect when batch normalizations are used (Ioffe & Szegedy 2015).

We do not attempt to exhaustively optimize the training process of the neural network by varying hyper-parameters, neither do we attempt to find the optimal CNN architecture. We find that our basic setup works reasonably well, and we delegate the fine-tuning of the details to later works.

Our CNN does not make use the known PSF during predictions, although one could possibly stack another channel with a PSF model and hope that the CNN figures the corrections out by itself. Rather, we leave the PSF correction to a well-controlled post-processing step, in order to have complete control over systematics. We find that the dependence of the error of the predicted galaxy ellipticities on the PSF anisotropy is well characterized with a linear relationship, which we correct for, using the PSF anisotropy parameters in the DES Y1 shape catalogues. Note, that this correction is very small compared to the overall variance of shape estimation, and it absolutely does not affect our results, when
4 RESULTS

First, we directly compare the galaxy ellipticities predicted by the CNN and the measurements of the DES METACALIBRATION catalogue to the “ground truth” ellipticities obtained from CFHTLenS. For this purpose, we select a relatively high-quality sample in the test set where the \( S/N \) measured by DES is larger than 15, in order to omit galaxies from DES which are measured with low confidence. This selection is made to show the DES estimates in their best possible form. Heatmaps of the true and the estimated values reveal that the predictions of the CNN show significantly smaller variance than the DES catalogue with no visible bias [Fig. 3]. Multiplicative and the additive biases are shortly inspected in Appendix §A.

Next, we investigate the dependence of accuracy on the brightness of galaxies. We use all galaxies selected for cosmology in the DES METACALIBRATION catalogue, and measure the correlation between the estimated and the CFHTLenS ellipticities after dividing them into red magnitude bins. We also include the DES IM3SHAPE catalogue into this analysis on galaxies where the estimated weight is larger than 0 in the IM3SHAPE catalogue. Note, that for each of the following experiments we use the predictions made after training on the whole training set, and we only divide galaxies into different bins for the evaluations. We use the magnitudes for binning because the \( S/N \) ratios are only available for the METACALIBRATION catalogue. As the main result of our study, we find that the DES catalogues are just as accurate as the CNN for bright galaxies, but the CNN becomes significantly superior as noise begins to dominate the images. The CNN is able to maintain the same accuracy as the DES METACALIBRATION catalogue for even 0.5 – 1 magnitude fainter objects. The IM3SHAPE catalogue deteriorates even faster than METACALIBRATION, as it is expected for a method which uses less bands therefore less information in the fit (Jarvis & Jain 2008). The imperfect shape predictions for faint galaxies the DES catalogues are due to the noise and not differences in the surveys and tools, demonstrated by the fact that galaxies brighter than 21.5 magnitudes are almost perfectly estimated by these catalogues [Fig. 4]. This threshold is notable as that is where simple fitting algorithms start to break down in the Y1 data (Tarsitano et al. 2018).

As a logical consequence of the result above, the higher accuracy of the CNN at the same magnitudes translates to more galaxies measured with the same accuracy. Scanning different thresholds in magnitude we calculate the galaxy surface densities and the corresponding Pearson correlations when evaluating galaxies brighter than the threshold value. We find that the CNN is able to measure more than 60% more galaxies than the DES METACALIBRATION catalogue at a given required accuracy [Fig. 4]. Our results indicate that shape estimation with a CNN enables weak lensing surveys to increase their galaxy density by more than 60% via algorithmic post-processing, without further observations.

Additionally, we evaluate the impact of the size of galaxies on the accuracy of shape estimates, for the set of galaxies with \( S/N > 15 \). The CNN is more accurate for any galaxy size [Fig. 5], and both the CNN and the DES catalogue start comparing the shapes of individual galaxies, without ensembling for shear estimation. We find no bias related to the size of the PSF.

Our results are indistinguishable for the two ellipticity components, therefore we only show figures and results for the first ellipticity component.

The neural networks are implemented in the Keras framework, training and evaluation were performed on a machine with 3 NVIDIA V100 GPUs. We make the source code for our approach publicly available on GitHub.

---

\[ \text{https://keras.io} \]

\[ \text{https://github.com/riblidesso/shearNN} \]
to fail for galaxies which are smaller than 1 arcsec, which is hardly surprising given that this radius is comparable to the seeing which is (0.6 – 0.8) arcsec, and it is less than 4 times the size of a pixel.

We also find that the accuracy of the CNN is stable for different PSF levels, and it consistently outperforms the DES catalogue regardless of seeing [Fig. 5]. Note that our test region has strongly bimodal PSF due to suboptimal observation conditions for one pointing, which is clearly visible in the spatial dependence of the PSF (not shown).

The CNN learns the galaxy surface density profiles from the data itself, without strong simplifying assumptions, which is expected to give an advantage when predicting the shapes of complex spiral or irregular galaxies. To test this idea we selected galaxies with peak photometric redshifts between 0.6 – 0.8 according to the CFHTLenS photometric redshift (Chang et al. 2013). We quantify the expected gain in the statistical uncertainty of shear measurements from the more accurate shapes provided by the CNN using the effective surface density introduced by (Chang et al. 2013).

\[ N_{\text{eff}} = \frac{1}{\Omega} \sum \frac{\sigma_{\epsilon}^2}{\sigma_i^2 + \sigma_{\epsilon}^2} \]

Figure 4. Left: Faint galaxies are estimated more accurately by the CNN than DES shape catalogues. Pearson correlation between “ground truth” and estimated ellipticities are shown in different $r$ band magnitude bins. The histogram of all galaxies are shown in light grey, and the ones selected for cosmology analysis in DES are shown with dark grey. Right: The CNN is able to measure the shapes of significantly more galaxies at a given required Pearson correlation. The plotted lines are sequences with different $r$-band magnitude cuts indicated by the colorbar. The CNN is able to measure 1.6–2.4x more galaxies than the DES catalogue with the same overall correlation, depending on the magnitude threshold.

Figure 5. The CNN estimates galaxy shapes more accurately than the DES metacalibration catalogue regardless of galaxy size (left), psf size (center) or galaxy $r - i$ color (right). The histogram of all galaxies are shown in light grey, and the ones selected for cosmology analysis in DES are shown with dark grey. The $r - i$ color dependence is only shown for galaxies with peak photometric redshifts between 0.6 – 0.8.

Galaxy shape measurement with convolutional neural networks
Figure 6. The CNN increases the effective galaxy density of the DES survey by 35% compared to the DES METACALIBRATION catalogue and by 150% compared to the DES IM3SHAPE catalogue. The largest gain in effective surface density comes from galaxies between 22.5 – 23 magnitudes.

\[ \text{N}_{\text{eff}} = \frac{\text{mag}}{\sigma_{\text{SN}}} \]

where \( \Omega \) is the survey area, \( \sigma_{\text{SN}} \) is the shape noise due to the intrinsic ellipticities, here 0.225, and \( \sigma_e \) is the variance of shape measurements. Similarly to Jarvis et al. (2016) we calculate \( \sigma_e \) empirically, we divide galaxies into 0.2 magnitude bins, and calculate the variance from the prediction errors in these bins. We use each galaxy with \( S/N > 0 \) for this analysis to make it most comparable among the methods without relying on specific cuts in IM3SHAPE of the METACALIBRATION catalogues. For the IM3SHAPE catalogue we also require the original weights to be higher than 0.

We find that the CNN reaches 35% higher effective galaxy density than the METACALIBRATION catalogue, and 150% higher than IM3SHAPE [Fig.6]. Due to the fact that fainter galaxies have smaller weights and therefore their contribution diminishes the 35% increase in effective galaxy density is smaller than the actual number of galaxies measured with the same accuracy as shown in [Fig. 4]. Interestingly the CNN does not gain most of its advantage from the very faintest galaxies, but instead for DES images the main contribution comes from galaxies with \( r \) magnitudes between 22.5 and 23 [Fig.6].

Our training dataset contains approximately \( 10^5 \) galaxies, and it may prove to be hard to collect such a large training dataset with reliable “ground truth” ellipticities. Therefore we investigate the performance depending on the number of galaxies for training the CNN, by randomly excluding galaxies from the training set. We maintain the same number of minibatch iterations regardless of the final training data size, in order to allow convergence for small training datasets. We find that the CNN is able to match the performance of the DES catalogues with as few as 2000 galaxies in the training set, and the performance plateaus around \( 4\times10^4 \) galaxies [Fig. 7]. A peculiar flat region also appears between \( 2 \times 10^3 - 10^4 \) galaxies, which might arise due to the suboptimal training schedule for some training set size, however, we can not optimize the training schedule for every single training size value tested.

Finally, we show examples of galaxies where the CNN estimated shapes significantly more accurately than the DES catalogue [Fig. 8]. The strong noise which possibly misleads the DES shape fitting algorithms is generally visible on these examples, and it is hard to see the subtle clues which allowed the CNN to miraculously recover the shapes of galaxies from these noisy observations.

Figure 7. The accuracy of CNN galaxy shape estimation depending on the size of the training dataset. The CNN surpasses the DES catalogues with as little as 2000 galaxies in the training set, and approximately \( 4\times10^4 \) are needed for maximal efficiency. The error bars represent the minimum and the maximum values for 4 runs with different random training subsamples.

5 DISCUSSION

We present a novel setup to create a training and testing dataset from only observational data using a wider but shallower survey to obtain postage stamp images of galaxies, and a narrow but deeper survey to measure the “ground truth” shapes of galaxies. Such a training and testing setup will also be possible to construct for any ongoing or planned weak lensing survey with dedicated deep fields such as the Deep Drilling Field for LSST, the Euclid Deep Fields the Subaru Deep Field and the WFIRST Deep Fields. Space telescope data from Hubble and WJT could also be used to provide “ground truth” ellipticities.

We propose convolutional neural networks for galaxy shape estimation for weak lensing, and we show that the CNN estimates the shapes of faint galaxies significantly more accurately in terms of variance than model fitting methods used in the DES Y1 catalogues. We also demonstrate that the superior accuracy of the CNN translates to 35% larger effective galaxy densities for DES than achieved by the METACALIBRATION catalogue and 150% larger than IM3SHAPE, via only algorithmic post-processing of data and no further observations. We find that the largest contribution to effective
Figure 8. Example galaxy shape measurements and $i$-band images from both DES (top) and CFHTLenS (bottom) for galaxies where the CNN is more accurate than the DES catalogue. The ellipticities are represented with a line at the bottom of the figures. The width of the grey band represents an ellipticity of 0.8.

Galaxy density does not come from the faintest galaxies but from between $22.5 - 23$ magnitudes.

The CNN is able to surpass the DES catalogues using a training sample counting as few as 2000 galaxies and it reaches peak performance around $4 \times 10^5$ galaxies, which should be the lowest target for the size of a training dataset for a shape estimator CNN.

Model fitting procedures are very resource intensive, which already warrant simplifications in current surveys (Fenech Conti et al. 2017; Zuntz et al. 2018), and it will be an even bigger problem to tackle the large scale, deep surveys in the near future which are expected to detect an order of magnitude more galaxies. CNNs offer a super-fast alternative which could significantly facilitate the data analysis of very large weak lensing surveys. Our CNN is able to estimate the shape of a galaxy in 0.2 milliseconds, improving approximately four orders of magnitude compared to model fitting (Zuntz et al. 2018), allowing the shapes of whole galaxy sample of LSST to be measured in less than a week on a single machine with 3 NVIDIA V100 GPUs.

We expect that CNNs should also be useful for outlier and catastrophic fitting error detection even when trained with “ground truth” shape data from the wide survey itself instead of deeper “ground truth” data, serving as a quick quality control tool.

For the demonstration purpose of the current study we use the openly released and processed coadded images from DES DR1, however, weak lensing surveys generally evaluate multi-epoch data with a joint fit over different exposures (Miller et al. 2013; Zuntz et al. 2018) as the discontinuities and the interpolation of the PSF creates artifacts in the coadded images which could be problematic for shear estimation (Miller et al. 2013). A joint prediction using multiple exposures could be simply stacked in different input channels for a joint fit, as a CNN can easily use images with any number of input channels.

The accuracy of CNN shape estimators should be tested on simulations too in order to improve our understanding of the specific points such as noise, shape models and priors which contribute to the superior performance of the CNN compared to model fitting approaches. The insights learned from such a study could even be used to improve model fitting approaches, as it is sometimes possible to extract meaningful and interpretable knowledge from the inspection of a CNN (Ribli et al. 2019b).

Systematic biases must be very carefully investigated and corrected when applying shape estimators on observational data to extract cosmic shear, however, this is not unique for the CNN, and we expect no more difficulties in its calibration compared to model fitting approaches (Fenech Conti et al. 2017). METACALIBRATION is agnostic to the shear measurement algorithm (Huff & Mandelbaum 2017; Sheldon & Huff 2017) and the CNN estimator should be straightforward to embed in that scheme, but realistic image simulations could also be used to calibrate the CNN just as other estimators.

Finally, note that in our proposed scheme, galaxy shape estimation with CNNs cannot completely replace model fitting approaches, as the training procedure relies on high-quality shape measurements from a deeper survey which must be performed with conventional methods.

ACKNOWLEDGEMENTS

This work was partially supported by the National Research, Development and Innovation Office of Hungary via grant
REFERENCES

Abbott T. M. C., et al., 2018, ApJS, 239, 18
Amendola L., et al., 2018, Living Reviews in Relativity, 21, 2
Bacon D. J., Refregier A. R., Ellis R. S., 2000, MNRAS, 318, 625
Bernstein G. M., Jarvis M., 2002, AJ, 123, 583
Bridle S., et al., 2009, Annals of Applied Statistics, 3, 6
Bridle S., et al., 2010, MNRAS, 405, 2044
Chang C., et al., 2013, MNRAS, 434, 2121
Cohen T., Welling M., 2016, in International conference on machine learning. pp 2990–2999
Diehl H. T., et al., 2016, in Observatory Operations: Strategies, Processes, and Systems VI. p. 99101D, doi:10.1117/12.2233157
Dieleman S., Willett K. W., Dambre J., 2015, MNRAS, 450, 1441
Domínguez Sánchez H., Huertas-Company M., Bernardi M., Tuccillo D., Fischer J. L., 2018, MNRAS, 476, 3661
Erbens T., et al., 2013, MNRAS, 433, 2545
Fenech Conti I., Herbonnet R., Hockstra H., Mertens J., Miller L., Viola M., 2017, MNRAS, 467, 1627
He K., Zhang X., Ren S., Sun J., 2015, in Proceedings of the IEEE international conference on computer vision. pp 1026–1034
He K., Zhang X., Ren S., Sun J., 2016, in Proceedings of the IEEE conference on computer vision and pattern recognition. pp 770–778
Herbel J., Kacprzak T., Amara A., Refregier A., Lucchi A., 2018, J. Cosmology Astropart. Phys., 7, 054
Heymans C., et al., 2006, MNRAS, 368, 1323
Heymans C., et al., 2012, MNRAS, 427, 146
Heymans C., et al., 2013, MNRAS, 432, 2433
Hildebrandt H., et al., 2012, MNRAS, 421, 2355
Hildebrandt H., et al., 2017, MNRAS, 465, 1454
Hirata C., Seljak U., 2003, MNRAS, 343, 459
Huff E., Mandelbaum R., 2017, arXiv e-prints, Ioffe S., Szegedy C., 2015, arXiv preprint arXiv:1502.03167
Jarvis M., Jain B., 2008, J. Cosmology Astropart. Phys., 1, 003
Jarvis M., et al., 2016, MNRAS, 460, 2245
Kacprzak T., Zuntz J., Rowe B., et al., 2012, MNRAS, 427, 2711
Kacprzak T., Bridle S., Rowe B., Voigt L., Zuntz J., Hirsch M., MacCrann N., 2014, MNRAS, 441, 2528
Kaiser N., 2000, ApJ, 537, 555
Kaiser N., Squires G., Broadhurst T., 1995, ApJ, 449, 460
Kaiser N., Wilson G., Luppiño G. A., 2000, arXiv Astrophysics e-prints,
Kilbinger M., 2015, Reports on Progress in Physics, 78, 086901
Kitching T. D., Miller L., Heymans C. E., van Waerbeke L., Heavens A. F., 2008, MNRAS, 390, 149
Kitching T., et al., 2010a, arXiv e-prints, Kitching T., et al., 2010b, Gravitational Lensing Accuracy Testing 2010 (GREAT10) Challenge Handbook. (arXiv:1009.0779)
Kondor R., Lin Z., Trivedi S., 2018, in Advances in Neural Information Processing Systems. pp 10138–10147
Krizhevsky A., Sutskever I., Hinton G. E., 2012, in Advances in neural information processing systems. pp 1097–1105
Kuijken K., 1999, A&A, 352, 355
Kuijken K., et al., 2015, MNRAS, 454, 3500
Lintott C. J., et al., 2008, MNRAS, 389, 1179
Mandelbaum R., et al., 2015, MNRAS, 450, 2963
Massey R., et al., 2007, MNRAS, 376, 13
Melchior P., Viola M., 2012, MNRAS, 424, 2757
Milan L., Böhmert A., Lombardi M., Bartelmann M., 2010, A&A, 510, A75
Miller L., Kitching T. D., Heymans C., Heavens A. F., van Waerbeke L., 2007, MNRAS, 382, 315
Miller L., et al., 2013, MNRAS, 429, 2858
Morganson E., et al., 2018, PASP, 130, 074501
Pujol A., Kilbinger M., Sureau F., Bobin J., 2019, A&A, 621, A2
Redman J., Farahdi A., 2017, arXiv preprint Refregier A., Kacprzak T., Amara A., Bridle S., Rowe B., 2012, MNRAS, 425, 1951
Ribili D., Armin Pataki B., Zorrilla Matilla J. M., Hsu D., Haiman Z., Csabai I., 2019a, arXiv e-prints,
Ribili D., Pataki B. A., Csabai I., 2019b, Nature Astronomy, 3, 93
Rowe B., et al., 2014, arXiv:1407.7676 [astro-ph]
Schrabback T., et al., 2010, A&A, 516, A63
Sheldon E., 2015, Astrophysics Source Code Library
Sheldon E. S., Huff E. M., 2017, ApJ, 841, 24
Simet M., Mandelbaum R., Rowe B., Great3 Collaboration 2014, 223, 255.17
Simonyan K., Zisserman A., 2014, arXiv preprint arXiv:1409.1556
Springer O. M., O’Leck E. O., Weiss Y., Merten J., 2018, arXiv e-prints,
Szegedy C., et al., 2015.
Takada M., 2010, in AIP Conference Proceedings. pp 120–127
Tarsitano F., et al., 2018, MNRAS, 481, 2018
Tewes M., Kuntzer T., Nakajima R., Courbin F., Hildebrandt H., Schrabback T., 2019, A&A, 621, A36
Tuccillo D., Huertas-Company M., Decenciè re E., Velasco-Fuero S., 2017, [10.1017/S1743921317000552], 325, 191
Van Waerbeke L., et al., 2000, A&A, 358, 30
Voigt L. M., Bridle S. L., 2010, MNRAS, 404, 458
Wang H., et al., 2019, Nat. Methods, 16, 103
Wittman D. M., Tyson J. A., Kirkman D., Dell’Antonio I., Bernstein G., 2000, Nature, 405, 143
Xu L., Ren J. S., Liu C., Jia J., 2014, in Advances in Neural Information Processing Systems. pp 1790–1798
Zuntz J., Kacprzak T., Voigt L., Hirsch M., Rowe B., Bridle S., 2013, MNRAS, 434, 1604
Zuntz J., et al., 2018, MNRAS, 481, 1149
de Jong J. T. A., et al., 2017, A&A, 604, A134

APPENDIX A: NOISE BIAS

We test the systematic bias of the CNN shape estimates in the form of a multiplicative and an additive component, $m_i$ and $c_i$ for both ellipticity components $(1, 2)$. We divide the test galaxies into $10$ bins based on their $r$-band magnitudes, and we fit a linear function to estimate the multiplicative and the additive bias in each bin [Fig. A1]. We find that the CNN shape estimates show a regular noise bias pattern with a similar or smaller magnitude than uncalibrated model fitting approaches (Heymans et al. 2012; Miller et al. 2013; Jarvis et al. 2016; Fenech Conti et al. 2017; Zuntz et al. 2018). The bias shown by the neural network should be possible to calibrate with simulated galaxies, or meta-calibration, in order to produce shear estimates without significant noise bias.

This paper has been typeset from a TEX/LATEX file prepared by the author.
Figure A1. Multiplicative $m_{1,2}$ (top) and additive $c_{1,2}$ (bottom) noise biases in different magnitude bins. The range of the bias in the uncorrected estimates is similar to the bias seen for model fitting approaches. The error bars show the standard deviations of the fitted bias parameters.