An improved cosmological parameter inference scheme motivated by deep learning

Dezső Ribli, Bálint Ármin Pataki and István Csabai

Dark matter cannot be observed directly, but its weak gravitational lensing slightly distorts the apparent shapes of background galaxies, making weak lensing one of the most promising probes of cosmology. Several observational studies have measured the effect, and there are currently running\textsuperscript{6,22} and planned efforts\textsuperscript{3,4} to provide even larger and higher-resolution weak lensing maps. Owing to nonlinearities on small scales, the traditional analysis with two-point statistics does not fully capture all of the underlying information\textsuperscript{7}. Multiple inference methods have been proposed to extract more details based on higher-order statistics\textsuperscript{8–13}, peak statistics\textsuperscript{14–17}, Minkowski functionals\textsuperscript{18–20} and recently convolutional neural networks\textsuperscript{21,22}. Here we present an improved convolutional neural network that gives significantly better estimates of the $\Omega_m$ and $\sigma_8$ cosmological parameters from simulated weak lensing convergence maps than state-of-the-art methods and that is also free of systematic bias. We show that the network exploits information in the gradients around peaks, and with this insight we have constructed an easy-to-understand and robust peak-counting algorithm based on the steepness of peaks, instead of their heights. The proposed scheme is even more accurate than the neural network on high-resolution noiseless maps. With shape noise and lower resolution, its relative advantage deteriorates, but it remains more accurate than peak counting.

Following the idea and using the simulation data from a recent study\textsuperscript{18} we created an improved convolutional neural network (CNN) architecture (see details in the Methods) that is able to recover cosmological parameters more accurately from simulated weak lensing maps. The input of the network is a set of mock convergence ($\kappa$) maps generated by ray-tracing $N$-body simulations with 96 different values for the matter density $\Omega_m$ and the scale of the initial perturbations normalized at the late Universe, $\sigma_8$ (see refs\textsuperscript{18,19} for details of the weak lensing map generation), the outputs of the network were the predicted cosmological parameters. The modifications of the CNN mostly consisted of adding further activations, increasing the number of filters, and introducing a regular block structure, following successful computer vision models\textsuperscript{20,21}.

Our CNN’s parameter estimation accuracy—on previously unseen lensing maps—beats all the state-of-art approaches for the complete parameter range and, more importantly, it is free of systematic bias as shown in Fig. 1 and summarized in Table 1. The CNN architecture used in the previous work\textsuperscript{18} shows large errors and strong bias in the predicted predictions, even close to the fiducial cosmological parameters ($\Omega_m=0.26$, $\sigma_8=0.8$). The improved CNN architecture predicts both $\Omega_m$ and $\sigma_8$ parameters with $\approx 2\times$ smaller errors than peak counting in the full parameter range, with no bias, demonstrating that CNNs are indeed capable of extracting significantly more information from weak lensing maps than standard approaches.

Despite the superb accuracy we demonstrated, the peculiarities of neural networks warrant extraordinary caution when trying to infer credible physical parameters from measurement data with a CNN, which was only trained on simulated data. Interesting results in the context of image recognition caution that neural networks may not be as robust as the regular tool-set of a cosmologist. It is possible to engineer malicious, imperceptible perturbations of images that completely fool a CNN\textsuperscript{22} and it was shown that unexpected inputs to deep neural networks are most likely to be processed incorrectly, and that the behaviour of a CNN is only reasonable on a thin manifold encompassing the training data\textsuperscript{23}. It is also important to keep in mind that parameter inference with neural networks is conceptually different from the established approaches in weak lensing, which rely on directly comparing simulation data and measurements, through reduction to the power spectrum or peak statistics. A CNN, on the other hand, learns to approximate a numerical function that directly maps the very high dimensional space of measurements into the final parameter space. One consequence of this process is that without explicit data comparison the estimation of the goodness of fit through residuals is not possible. It may be possible to overcome these hurdles through careful investigations, however, we chose to follow a different path in our study.

Although the complex interplay of millions of parameters in a CNN simply cannot be fully comprehended, through careful investigation of the internal parameters of the trained neural network, we can gain insights into its workings. The internal representations of neural networks often have human-understandable interpretations, high-level feature maps frequently learn to selectively detect complex concepts on the images such as legs, wheels or faces\textsuperscript{23}. We attempted to go even further, not only to find meaningful internal representations but to use them as a hint, and build an easy-to-understand and robust estimation method. To make the interpretation of the CNN’s weights easier, we trained a different CNN with a larger ($7 \times 7$) kernel size in the first layer, on lensing maps resized to 2 arcmin pixel size, similar to the resolution expected from observations. The inspection of the kernels immediately revealed that the neural network discovered some interesting and familiar concepts from the training data (Fig. 2).

The neural network learned to use a kernel strikingly similar to a 2D discrete (negative) Laplace operator that basically calculates the difference of the peaks and the surrounding pixel values. With the middle element scaled to 1 the learned kernel is the following:

\[
K \approx \begin{bmatrix}
-0.05 & -0.25 & -0.06 \\
-0.21 & 1 & -0.29 \\
-0.07 & -0.15 & -0.22
\end{bmatrix}
\]  

Department of Physics of Complex Systems, ELTE Eötvös Loránd University, Budapest, Hungary. *e-mail: csabai@complex.elte.hu

Department of Physics of Complex Systems, ELTE Eötvös Loránd University, Budapest, Hungary. *e-mail: csabai@complex.elte.hu
Using this hint, we considered the most isotropic discrete Laplace operator, which is very close to the learned kernel, and a simpler version with zeros in the corners:

\[
L_1 = \frac{-10}{3} \begin{bmatrix} -0.05 & -0.2 & -0.05 \\ -0.2 & 1 & -0.2 \\ -0.05 & -0.2 & -0.05 \end{bmatrix},
\]

\[
L_2 = \begin{bmatrix} 0 & -0.25 & 0 \\ -0.25 & 1 & -0.25 \\ 0 & -0.25 & 0 \end{bmatrix}
\]

Another interesting kernel learned by the neural network is very similar to one of the Roberts cross kernels (Fig. 2), which approximate the gradient of an image:

\[
R_x = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix},
\]

\[
R_y = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix},
\]

\[
G = \sqrt{G_x^2 + G_y^2},
\]

where the Roberts cross kernels are denoted by \(R_x, R_y\), the computed gradients with \(G_x, G_y\) and the magnitude of the gradient with \(G\).

Table 1 | Overview of prediction errors for the CNN and different peak-counting schemes

| Peak counting | \(\Omega_m\) (RMSE) \(\times 10^3\) | \(\sigma_8\) (RMSE) \(\times 10^3\) |
|---------------|----------------------------------|----------------------------------|
| CNN \(^\dagger\) | 11.2 ± 0.2 | 22.8 ± 0.2 |
| CNN (this work) | 35.1 ± 0.2 | 40.3 ± 0.3 |
| Laplace v1 | 5.5 ± 0.1 | 13.5 ± 0.1 |
| Laplace v2 | 4.7 ± 0.1 | 11.6 ± 0.1 |
| Roberts cross | 4.6 ± 0.1 | 10.9 ± 0.1 |
| Peak counting (noisy, \(n_g = 8\)) | 4.3 ± 0.1 | 9.7 ± 0.1 |
| Laplace (fiducial) | \(\Omega_m\) | \(\sigma_8\) (fiducial) |
| Peak counting (noisy, \(n_g = 8\)) | 43.9 ± 0.8 | 64.4 ± 0.7 |
| Sobel filter (noisy, \(n_g = 8\)) | 30.3 ± 0.3 | 56.7 ± 0.6 |
| Peak counting (noisy, \(n_g = 26\)) | 19.6 ± 0.4 | 34.5 ± 0.4 |
| Sobel filter (noisy, \(n_g = 26\)) | 14.8 ± 0.2 | 30.9 ± 0.5 |

The evaluation criterion of predictions is the RMSE or predicted values (see Methods), standard deviations are estimated from 10,000 bootstrap samples. Results on the noisy maps are calculated at the fiducial cosmological model (&\(\Omega_m = 0.26, \sigma_8 = 0.8\)), with a simulated footprint of 450 deg\(^2\). The two pairs of noise and angular resolution parameters used are: (\(A_{\text{pix}} = 4\text{ arcmin}^2, n_g = 8\text{ arcmin}^{-2}\)) and (\(A_{\text{pix}} = 1\text{ arcmin}^2, n_g = 26\text{ arcmin}^{-2}\)). Boldface entries show the most accurate results.
Kernels (2) and (3) calculate the difference of peak values and their surroundings, or the gradients around a peak, therefore, they potentially describe the steepness of peaks. Naturally, the gradients at the peaks are 0, thus steepness can be described using the magnitudes of gradient values around peaks. We found that steepness is strongly correlated with the heights of the peaks, as shown previously, but it may contain more information than height, which is used in the standard peak-counting method for parameter estimation in weak lensing. A previous study has shown that $\Omega_m$ and $\sigma_8$ cosmological parameters have substantial effect on the stacked tangential shear profile of peaks, and their steepness, and another study showed that using the gradient moments of the field improved predictions of cosmological parameters, which may explain why these kernels were discovered and used by the neural network.

Based on these insights we can conclude that the neural network achieves better results than peak counting in part by using representations based on the steepness of the peaks. Beyond understanding the success of the CNN, the steepness of peaks can be used as a simple and robust descriptor, without the rest of the network and its countless other parameters. With the kernels (2) and (3) learned from the CNN we created an algorithm that uses the distribution of peak steepness values instead of the peak heights, which is used in the original scheme.

The results achieved with the peak-steepness scheme are shown in Fig. 2 and summarized in Table 1. Using the kernels suggested by the CNN, prediction errors were reduced over twofold for both $\Omega_m$ and $\sigma_8$ parameters compared with the original peak-counting algorithm. Another rather surprising result is that the peak-steepness scheme even surpassed the neural network’s performance. The accurate results indicate that the kernels extracted from the neural network were indeed responsible for its high accuracy, and it was possible to combine the best of both approaches in the peak-steepness method. The gain over the neural network may be explained by the fact that a CNN is not able to explicitly construct histograms of data values, or perform likelihood analysis, which could turn out to be the best approach in this case.

The weak lensing maps used in this study have a very high angular resolution (0.2 arcmin), which is not reachable in experiments due to the low density of observable galaxies, therefore, we evaluated the reconstruction errors on maps with reduced angular resolutions (Fig. 2). The peak-counting scheme based on steepness continues to predict $\Omega_m$ and $\sigma_8$ parameters more accurately than peak height at lower angular resolutions too. The results indicate that the gradients around peaks contain additional information compared with the height of the peaks, even for extended distances.

The intrinsic ellipticity of galaxies, ‘shape noise’, has a profound effect on observations and dominates over the lensing signal. Since the two filters described previously, Laplace and Roberts cross, were learned from noiseless maps they are not robust to noise, thus a new method to estimate the steepness of peaks in a more noise-resistant way is needed. Gradient estimation with Sobel filters is known to be robust in the presence of noise, therefore, we implemented a peak-counting version using these filters. Similarly to the two Roberts cross kernels, the two Sobel filters calculate the gradients in the x and y directions, and the magnitude of the gradient can be calculated in the same way. The method based on Sobel filters is essentially the same as the one based on the Robert cross kernels, with the difference that gradients are calculated from a larger area, which leads to more robust estimates, however, it also results in a loss of resolution, therefore Sobel filters are not optimal in the absence of noise:

$$S_x = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}, \quad S_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

(4)
σn(right) distributions were calculated at 4 different cosmological parameters, with shape noise corresponding to predictions. The distribution of peak steepness values exhibits a shift that is absent in the case of peak heights. The peak height (left) and steepness Peak steepness distributions show larger separation between different cosmologies than peak height values, which results in more accurate predictions. To simulate a realistic measurement, we evaluated predictions with peak heights and Sobel-filter-based peak steepness on lensing maps corresponding to the fiducial cosmologies with additional shape noise and reduced angular resolution. For each prediction we averaged the histograms of 37 individual maps, resulting in an approximately 450 square-degree simulated footprint, and we measured the RMSE of the estimated parameters from 10,000 realizations of noise and randomly selected 37 maps.

In the first set-up, which mimics near-future surveys, LSST and EUCLID3,4, we resized the maps to an angular resolution of 1 arcmins and added Gaussian shape noise at an effective galaxy density n_g = 26 arcmin−2. Peak steepness was more accurate than height both when predicting Ω_m and σ_8 (Table 1).

Our second set-up, which is comparable to currently running observations, KiDS and DES1,2, had an angular resolution of 2 arcmins and a galaxy density n_g = 8 arcmin−2. While the advantage of peak steepness deteriorated compared with the first case, it was still more accurate than peak height in terms of RMSE both when predicting Ω_m and σ_8 (Table 1).

The two scenarios demonstrate that peak steepness with Sobel filtering may be more accurate than peak height even in the presence of shape noise in relevant observational conditions. To gain insight into the mechanism of the peak steepness counting, we evaluated the mean histograms of the peak heights, and the Sobel-filter-based steepness values for 4 simulations with different Ω_m but similar σ_8 parameters, and 4 simulations with different σ_8 but similar Ω_m parameters (Fig. 3). The height of the distribution of steepness values seems to decrease similarly to peak counting with higher Ω_m values, and, in addition, the distribution of peak steepness values significantly shifts towards higher values with higher Ω_m parameters. Peak height and steepness distributions share the Ω_m–σ_8 degeneracy, and very similar effects appear on the histograms when changing σ_8 values (Fig. 3).

Deep CNNs are promising new tools for the analysis of 2D or 3D scientific datasets, and the adopters of this technology need to be aware that small details may create large differences in the quality of predictions. The complexity of choices makes working with neural networks more like an art with no simple recipe for success.

In many cases, it may be hard to obtain truly credible physical parameters with a CNN through direct inference from measurement data, thus we expect that our approach of building simple and robust descriptors based on the insights gained from interrogating a neural network may be applicable to other scientific machine-learning studies.

Peak counting based on the steepness of peaks is significantly more accurate than peak counting based on the height of peaks on noiseless, high-resolution convergence maps. With shape noise and lower resolution its relative advantage deteriorates but it remains more accurate. Our results indicate that peak counting based on the steepness of peaks has the potential to tighten the constraints of both Ω_m and σ_8 cosmological parameters compared with established methods. Improved parameter constraints from future surveys could alleviate or strengthen the tension between estimates gained from local and cosmic microwave background measurements. The proposed scheme’s efficiency on measurement data needs to be evaluated in future studies.

Methods

Data. Weak lensing convergence maps were generated with ray tracing from 96 cosmological N-body simulations, with 512 realizations from each simulation14.
these realizations were shown to be quasi-independent\textsuperscript{5}. Individual simulations had different pairs of $\Omega_m$ and $\sigma_8$ cosmological parameters. The parameters were most densely sampled around a `fiducial model' with $\Omega_m = 0.26$, $\sigma_8 = 0.8$, for more details see ref.\textsuperscript{3}. Each simulation had the same initial condition, therefore cosmic variance is neglected, and prediction accuracies may be systematically overestimated. As we considered the comparison of different inference methods, the following one of the simulations ($\Omega_m = 0.285$, $\sigma_8 = 1.134$) was not used as it was an obvious outlier based on ref.\textsuperscript{3} and our findings.

**Evaluation.** Predictions were evaluated using the most common evaluation criterion used for regression, the RMSE of the predicted values:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i} (y_{\text{predicted}} - y_{\text{true}})^2}$$

where $N$ denotes the number of predictions and $y$ is the quantity predicted. The uncertainties of the RMSE values were evaluated with the standard deviation of the RMSE values using 10,000 bootstrap samples. Note that this estimation ignores the variability in the models and therefore it underestimates the uncertainty in the presence of shape noise.

**Training neural networks.** The lensing maps were split into a training, validation and test set as 60%, 10% and 30%. Each 1,024 x 1,024 pixel map was tiled into 16 smaller images. During training, the network handled the tiles individually, and during prediction, the inferred values of the 16 tiles were averaged in order to obtain a final prediction for a whole map. The networks were trained for 5 epochs with Adam optimizer. The optimizer's parameters were $\beta_1 = 0.9$, $\beta_2 = 0.999$, where $\beta_1$ is the learning rate, $\beta_2$, and $\beta_3$ is the exponential decay rate for the weighted sum of the previous gradients' first and second moment. The loss was mean squared error (MSE) for the modified architecture and mean absolute error (MAE) for the architecture from ref.\textsuperscript{5}, as it was used in their analysis.

**Improved neural network architecture.** Instead of building a neural network from scratch, we decided to use the model from the previous work\textsuperscript{5} as a starting point, in order to highlight the important differences that radically improve performance. The architectural guidelines were based on the work of successful computer vision models\textsuperscript{17–19}. First, we made sure that each convolution was followed by an activation layer. Second, an unusually low number of filters in the first layer might cause under-fitting, thus we increased the number of filters in the first layer from 4 to 32. Third, we introduced a regular block structure, made of subsequent 2 convolutions and a spatial downsampling operation, which is a common structure\textsuperscript{20,21}. The 5 convolutional blocks have 32, 64, 128, 128 and 128 filters, and the blocks are followed by 3 dense layers with 256, 256 and 2 filters. Note that although our model starts with more filters, it has fewer parameters overall (2,649,088 versus 1,467,618) due to less extensive dense layers.

Apart from improvements in the architecture, we also decided to simplify the network to show which features are not absolutely necessary for good performance. We changed the LeakyReLU operations to simpler ReLU operations. With a large amount of simulation data available, we also decided to drop the Dropout layers from the network. The MAE loss function was also replaced by the more commonly used MSE cost function.

**Interpretation of the weights learned by the CNN.** To detect strong signals in the first layer of the CNN we trained another neural network with larger (7 x 7) kernel size on noiseless maps, whereas in the presence of noise the difference is reduced, as described previously\textsuperscript{3}. The approach is almost the same as peak counting, the peaks are still located on the original maps, however, we use the histograms of the peak height values. For the discrete Laplace filter, the calculation is straightforward, we convolve the original map with the filter, and take the values at the position of the peaks. In the case of the Roberts cross kernel, we apply the $R_x$ and $R_y$ filter on the image and calculate the magnitude of the gradient in the 4 adjacent 2 x 2 pixel blocks around the peak. For the histograms, we used the sum of the calculated 4 magnitudes. One small difference compared with peak heights is that calculated values are positive by definition at the local peaks, therefore we adjusted bins to run from 0 to 0.22 for the new values. Otherwise, the number and width of bins remained the same as for peak counting for the Laplace filter and the Roberts cross kernels. We verified that a similar bin shift for the original peak counting scheme does not explain the different results, and small differences in the bin width or the range do not noticeably alter the results in any schemes.

**Reducing angular resolution.** When testing on lower angular resolutions, we resized the weak lensing maps to a resolution that is an integer times the original resolution ($3, 2, 1$). In this decreasing sampling scheme, the new pixel values were calculated as the means of the pixel values in the area corresponding to the new pixel in order to avoid artefacts, such as the moiré pattern in the low-resolution maps. Reducing angular resolution this way does not destroy gradients like a Gaussian blur used in ref.\textsuperscript{5}, thus the modified peak-counting schemes can still work on reduced-resolution maps. We expect that measurement data can be handled in a similar way, which does not destroy differences among neighbouring pixels. Gradient values were observed to become larger at lower resolutions, therefore we increased the bin width linearly from 0.01 to 0.02, when changing resolution from 0.02 to 2.0 arcmins.

**Peak steepness counting with Sobel filters.** Sobel filters calculate the gradients of an image just as the Roberts cross kernels, thus the calculations are very similar. These kernels use information from a larger area, and they produce more robust gradient estimates, making them very useful in the presence of shape noise. We convolved the image with $S_x$ and $S_y$ Sobel filters to obtain the gradients $G_x, G_y$ and calculate the magnitude of the gradient as $G = \sqrt{G_x^2 + G_y^2}$. For the histograms, we used the mean of the gradient values calculated in the 8 adjacent pixels around peaks. In the case of the 2 arcmin resolution maps, and $n = 8$, we used 23 bins with equal width, which run from 0.3 to 1.1, and in the case of the 1 arcmin resolution maps, and $n = 26$, we used 23 bins with equal width, which run from 0.4 to 1.1.

**Shape noise.** The noise emerging from the intrinsic ellipticity of galaxies was modelled with a Gaussian noise in pixels\textsuperscript{10}:

$$\sigma_{\text{shape}} = \sigma \sqrt{\frac{A}{A_{\text{pix}}}}$$

where the ellipticity dispersion of galaxies $\sigma = 0.4, A_{\text{pix}}$ is the area of a pixel and $n$ is the surface density of galaxies. The two pairs of parameters used in the study were ($A_{\text{pix}} = 4 \text{ arcmin}^2, n = 8 \text{ arcmin}^{-2}$) and ($A_{\text{pix}} = 1 \text{ arcmin}^2, n = 26 \text{ arcmin}^{-2}$).

**Source code.** The source code used in this study is available online at https://github.com/ribidezio/peak_steepeastness.

**Data availability.** The data that support the plots within this paper and other findings of this study are available from the corresponding author upon reasonable request.

Received: 15 June 2018; Accepted: 12 September 2018; Published online: 29 October 2018

**References**

1. Hildebrandt, H. et al. KiDS-450: cosmological parameter constraints from tomographic weak gravitational lensing. Mon. Not. R. Astron. Soc. 465, 1454–1498 (2016).
2. Abbott, T. et al. Dark Energy Survey year 1 results: cosmological constraints from galaxy clustering and weak lensing. Phys. Rev. D 98, 043526 (2018).
3. Ivezic, Z. et al. LSST: from science drivers to reference design and anticipated data products. Preprint at http://arXiv.org/abs/1105.2366 (2008).
4. Laureijs, R. et al. Euclid definition study report. Preprint at http://arXiv.org/abs/1103.3193 (2011).
5. Kilbinger, M. Cosmology with cosmic shear observations: a review. Rep. Prog. Phys. 78, 086901 (2015).
6. Takada, M. & Jain, B. Three-point correlations in weak lensing surveys: model predictions and applications. *Mon. Not. R. Astron. Soc.* **344**, 857–886 (2003).

7. Fu, L. et al. CFHTLenS: cosmological constraints from a combination of cosmic shear two-point and three-point correlations. *Mon. Not. R. Astron. Soc.* **411**, 2725–2743 (2014).

8. Dietrich, J. & Hartlap, J. Cosmology with the shear-peak statistics. *Mon. Not. R. Astron. Soc.* **402**, 1049–1058 (2010).

9. Kratochvil, J. M., Haiman, Z. & May, M. Probing cosmology with weak lensing peak counts. *Phys. Rev. D* **81**, 043519 (2010).

10. Marian, L., Smith, R. E., Hilbert, S. & Schneider, P. The cosmological information of shear peaks: beyond the abundance. *Mon. Not. R. Astron. Soc.* **432**, 1338–1350 (2013).

11. Shan, H. et al. Weak lensing mass map and peak statistics in Canada–France–Hawaii Telescope Stripe 82 survey. *Mon. Not. R. Astron. Soc.* **441**, 2725–2743 (2014).

12. Liu, J. et al. Cosmology constraints from the weak lensing peak counts and the power spectrum in CFHTLenS data. *Phys. Rev. D* **91**, 063507 (2015).

13. Kaprzaek, T. et al. Cosmology constraints from shear peak statistics in Dark Energy Survey Science Verification data. *Mon. Not. R. Astron. Soc.* **463**, 3653–3673 (2016).

14. Petri, A., Haiman, Z., Hui, L., May, M. & Kratochvil, J. M. Cosmology with Minkowski functionals and moments of the weak lensing convergence field. *Phys. Rev. D* **88**, 125002 (2013).

15. Kratochvil, J. M. et al. Probing cosmology with weak lensing Minkowski functionals. *Phys. Rev. D* **85**, 103513 (2012).

16. Shirasaki, M. & Yoshida, N. Statistical and systematic errors in the measurement of weak-lensing Minkowski functionals: application to the Canada-France-Hawaii Lensing Survey. *Astrophys. J.* **786**, 43 (2014).

17. Schmelzle, J. et al. Cosmological model discrimination with Deep Learning. Preprint at http://arXiv.org/abs/1707.05167 (2017).

18. Gupta, A., Matilla, J. M. Z., Hsu, D. & Haiman, Z. Non-Gaussian information from weak lensing data via deep learning. *Phys. Rev. D* **97**, 103515 (2018).

19. Matilla, J. M. Z., Haiman, Z., Hsu, D., Gupta, A. & Petri, A. Do dark matter halos explain lensing peaks? *Phys. Rev. D* **94**, 083506 (2016).

20. Kuzhelevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. In *Adv. Neural. Information Processing Syst.* 1097–1105 (2012).

21. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. Preprint at http://arXiv.org/abs/1409.1556 (2014).

22. Goodfellow, I. J., Shlens, J. & Szegedy, C. Explaining and harnessing adversarial examples. Preprint at http://arXiv.org/abs/1412.6572 (2014).

23. Zeiler, M. D. & Fergus, R. Visualizing and understanding convolutional networks. In *European Conference on Computer Vision (ECCV)* 818–833 (Springer, Cham, 2014).

24. Lindeberg, T. Scale-space for discrete signals. *IEEE Trans. Pattern. Anal. Mach. Intell.* **12**, 234–254 (1990).

25. Roberts, L. G. *Machine Perception of Three-Dimensional Solids*. PhD thesis, Massachusetts Institute of Technology (1963).

26. Chang, C. et al. The effective number density of galaxies for weak lensing measurements in the LSST project. *Mon. Not. R. Astron. Soc.* **434**, 2121–2135 (2013).

27. MacCrann, N., Zuntz, J., Bridle, S., Jain, B. & Becker, M. R. Cosmic discordance: are Planck CMB and CFHTLenS weak lensing measurements out of tune? *Mon. Not. R. Astron. Soc.* **451**, 2877–2888 (2015).

28. Petri, A., Haiman, Z. & May, M. Sample variance in weak lensing: how many simulations are required? *Phys. Rev. D* **93**, 063524 (2016).

Acknowledgements
This work was partially supported by National Research, Development and Innovation Office of Hungary via grant OTKA NN 114560 and the National Quantum Technologies Program. The authors thank Z. Haiman and J. M. Z. Matilla for making available the simulated weak lensing maps used in this study.

Author contributions
I.C., D.R. and B.A.P. contributed to the conception and design of the study; B.A.P. performed the training and evaluation of neural networks; D.R. conducted the experiments with peak steepness. All authors reviewed the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Reprints and permissions information is available at www.nature.com/reprints.
Correspondence and requests for materials should be addressed to I.C.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2018