A Deep Learning Approach to Galaxy Cluster X-ray Masses

Michelle Ntampaka[1, 2], J. ZuHone[3], D. Eisenstein[1], D. Nagai[4], A. Vikhlinin[1], L. Hernquist[1], F. Marinacci[1], D. Nelson[5], R. Pakmor[5], A. Pillepich[6], P. Torrey[7], and M. Vogelsberger[8]

[1]Center for Astrophysics | Harvard & Smithsonian [2]Harvard Data Science Initiative [3]Smithsonian Astrophysical Observatory [4]Yale University [5]Max-Planck-Institut für Astrophysik [6]Max-Planck-Institut für Astronomie [7]University of Florida [8]Massachusetts Institute of Technology

HDSI | Harvard Data Science Initiative

CENTER FOR ASTROPHYSICS HARVARD & SMITHSONIAN

Abstract

We present a machine-learning approach for estimating galaxy cluster masses from *Chandra* mock images. We utilize a Convolutional Neural Network (CNN), a deep machine learning tool commonly used in image recognition tasks. The CNN is trained and tested on our sample of 7,896 *Chandra* X-ray mock observations, which are based on 329 massive clusters from the *IllustrisTNG* simulation. Our CNN learns from a low resolution spatial distribution of photon counts and does not use spectral information. Despite our simplifying assumption to neglect spectral information, the resulting mass values estimated by the CNN exhibit small bias in comparison to the true masses of the simulated clusters (-0.02 dex) and reproduce the cluster masses with low intrinsic scatter, 8% in our best fold and 12% averaging over all. In contrast, a more standard core-excised luminosity method achieves 15-18% scatter. We interpret the results with an approach inspired by Google DeepDream and find that the CNN ignores the central regions of clusters, which are known to have high scatter with mass.

Introduction: Galaxy Clusters as a Cosmological Probe

Galaxy clusters are gravitationally bound systems that contain hundreds or thousands of galaxies in dark matter halos of mass $\gtrsim 10^{14}\,M_\odot$. They are massive and rare, and their abundance is sensitive to the underlying cosmological model. Utilizing cluster abundance as a cosmological probe requires a large cluster sample with a well-defined selection function and a way to connect cluster observations to the underlying cluster mass with low intrinsic scatter.

Methods: Mock Chandra Observations

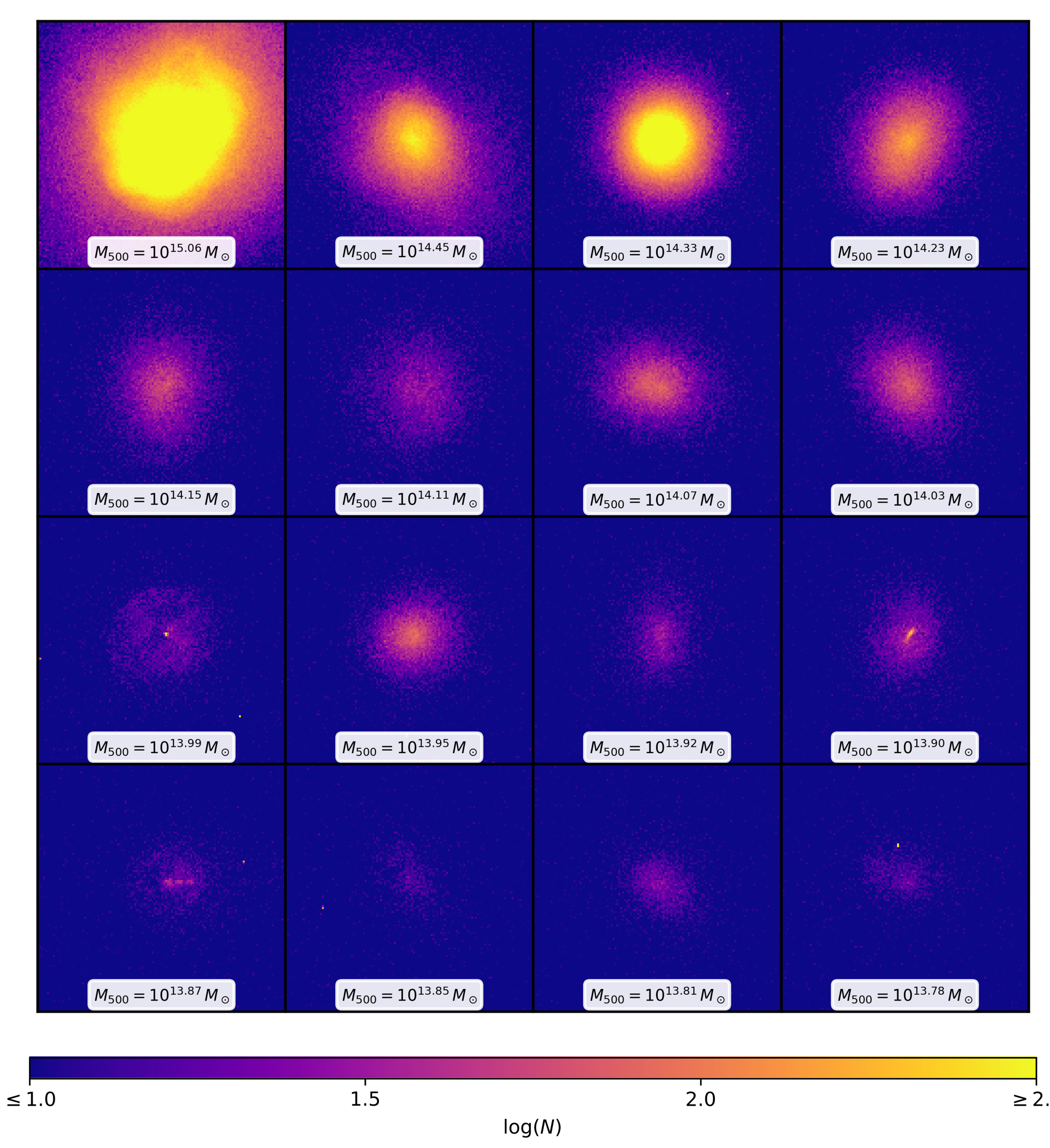

Figure: A sample of 16 of the 7,896 mock X-ray cluster observations created with pyXSIM software (ZuHone et al., 2014) applied to the *IllustrisTNG* cosmological hydrodynamical simulation (Springel et al., 2018). The mock observations emulate 100ks *Chandra* observations that have been degraded to 128×128 postage stamp images for one broad $0.5 - 7$keV energy band; shown are the number of photons, N, in each pixel for this mock observation.

Methods: Convolutional Neural Network

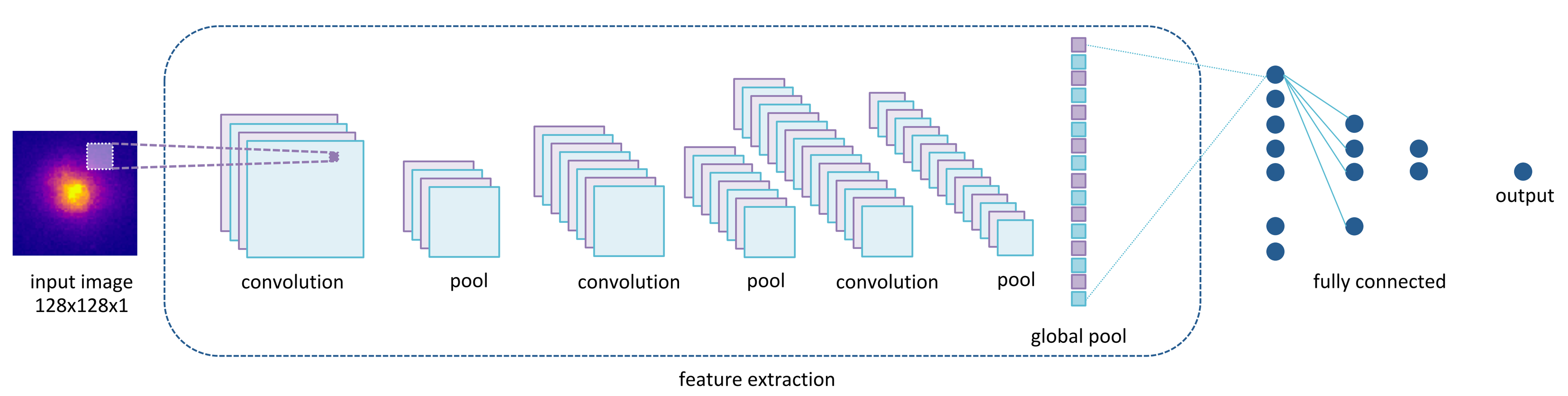

Figure: Convolutional neural networks employ a many-layered network to identify features of an input image, typically for image recognition. Deep convolutional neural networks (CNNs) have been used successfully in computer vision and image recognition. The method utilizes many layers of information processing, with multiple modules, each consisting of a convolutional layer and a pooling layer, and modules stacked to form a deep network. The convolutional layer averages the information within a rectangular window, while the pooling layer subsamples from the convolutional layer. Our model utilizes three layers of convolution and pooling, followed by three fully connected layers (see Ntampaka et al., 2018, for full model details). Results were produced through ten-fold crossvalidation.

References

ZuHone, J. A., Biffi, V., Hallman, E. J., et al. 2014, in Proceedings of the 13th Python in Science Conference, ed. Stéfan van der Walt & James Bergstra, 103 – 110
Springel, V., Pakmor, R., Pillepich, A., et al. 2018, MNRAS, 475, 676
Ntampaka, M., ZuHone, J., Eisenstein, D., et al. 2018, arXiv e-prints, arXiv:1810.07703
Maughan, B. J. 2007, ApJ, 668, 772
Mantz, A. B., Allen, S. W., Morris, R. G., & von der Linden, A. 2018, MNRAS, 473, 3072
Mordvintsev, A., Olah, C., & Tyka, M. 2015, in Google AI Blog
Chollet, F. 2016, How Convolutional Neural Networks See the World

Results: Cluster Mass Predictions

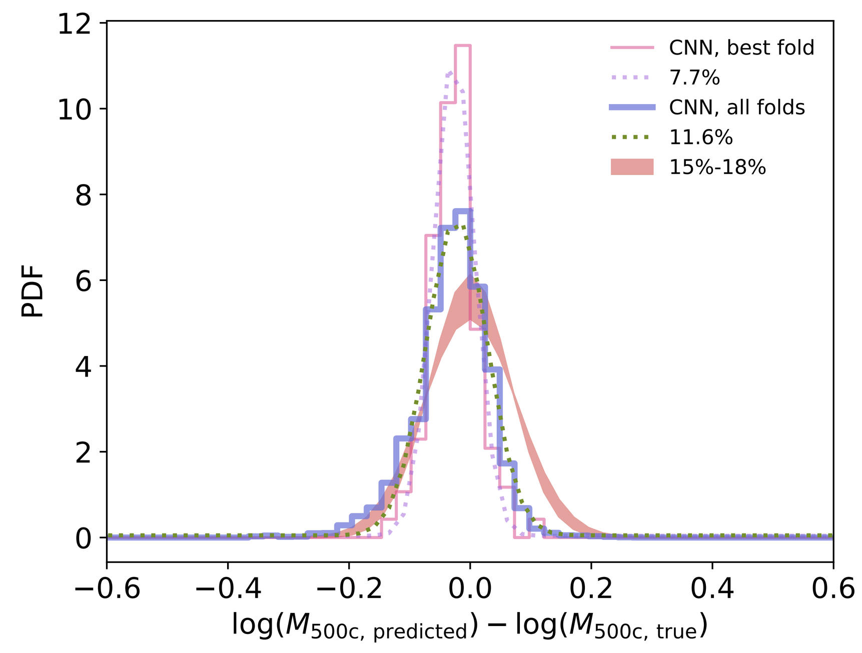

Figure: PDF of mass error (blue solid) given by $\log(M_{\rm predicted}) - \log(M_{\rm true})$. The full sample error distribution best fit Gaussian (green dash) has 11.6% intrinsic scatter, while the best-fit fold has 7.7% intrinsic scatter (pink solid and purple dash). Core-excised L_X-based methods that use a single measure of cluster luminosity typically achieve a $15\% - 18\%$ scatter (red band, Maughan, 2007; Mantz et al., 2018), though it is somewhat larger ($15\% - 23\%$) in the sample of simulated clusters. The Y_X technique, which uses a the full spatial and spectral observation, can yield a tighter $5\% - 7\%$ scatter. Our CNN approach uses low-resolution spatial information to improve mass estimates over one based on a single summary parameter, L_X.

Interpretation

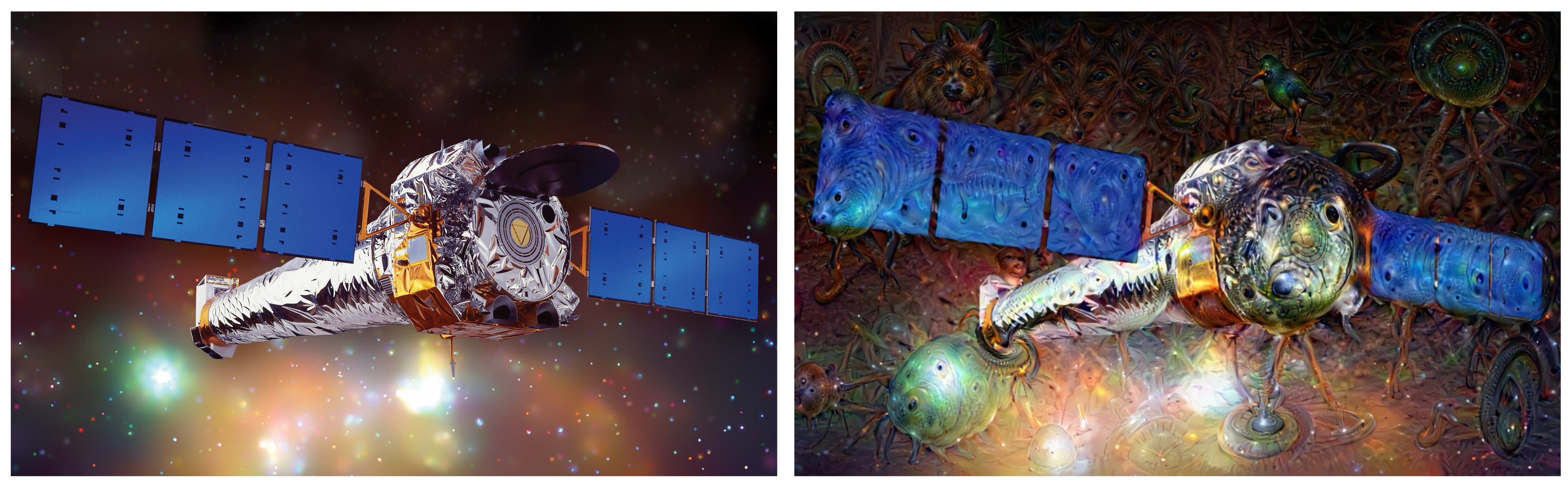

Figure: Convolutional Neural Networks are notoriously difficult to interpret. To understand the method, we use an approach inspired by DeepDream (e.g. Mordvintsev et al., 2015) which, for an image classifier, asks the question "What changes in the input image will result in a significant change in the classification of this image?" In contrast, our implementation regresses an output mass label, so we ask "What changes in the input cluster image will result in a mass change of this image?" Original image (left) credit: NASA/CXC/NGST, public domain.

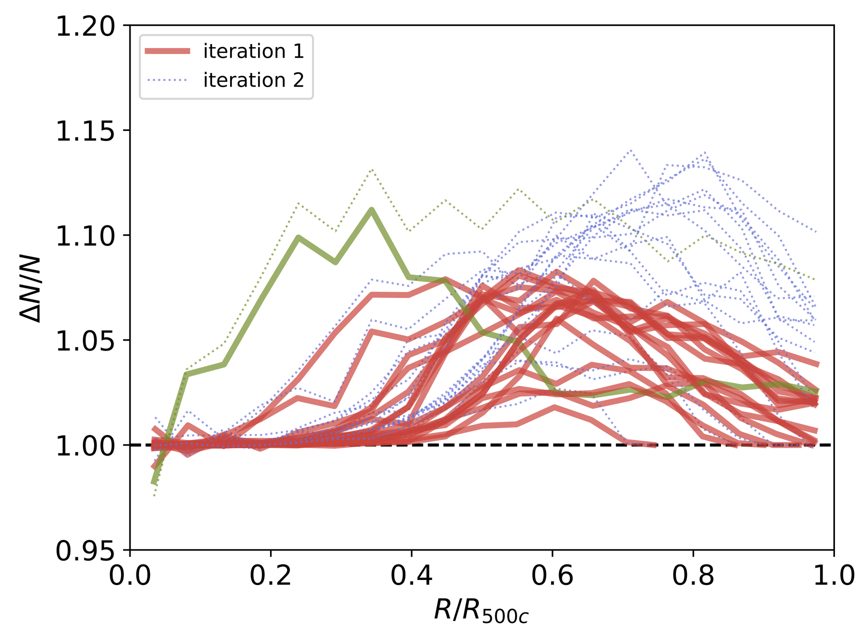

Figure: The fractional change in photons ($\Delta N/N$) as a function of projected distance from the cluster center (R/R_{500c}) for a sample of test clusters. The first iteration (solid red) adds photons beyond $\approx 0.2\,R_{500c}$, while the second iteration (blue dotted) increases the photon count at larger radii. This suggests that the CNN has learned to excise cores, which have been shown to have large scatter with M_{500c}. One notable exception (green) occurs when the CNN misidentifies off-center structure as the core. Interpretation tools such as this one can be used to interpret a CNN. For more details on visualizing filters of CNNs implemented in Keras, see Chollet (2016).

Summary

Convolutional neural networks, a class of machine learning image recognition algorithms, can reduce scatter in galaxy cluster mass estimates. For 100ks *Chandra* mock observations, they reduce scatter by a factor of ~2 or more. The model is not a black box; it produces interpretable results.

Ntampaka, et al., ApJ, Volume 876, Issue 1, article id. 82, 7 pp. (2019).