A Machine-learning Approach to Enhancing eROSITA Observations

John Soltis1, Michelle Ntampaka2,1, John F. Wu2,1, John ZuHone3, August Evrard4, Arya Farahi5, Matthew Ho6,7, and Daisuke Nagai8

1 Department of Physics & Astronomy, Johns Hopkins University, Baltimore, MD 21218, USA
2 Space Telescope Science Institute, Baltimore, MD 21218, USA
3 Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA
4 Departments of Physics and Astronomy and Leinweber Center for Theoretical Physics, University of Michigan, Ann Arbor, MI 48109, USA
5 Departments of Statistics and Data Science, University of Texas at Austin, Austin, TX 78705, USA
6 W. M. Keck Observatory, Hawaii, USA
7 NSF AI Planning Institute for Physics of the Future, Carnegie Mellon University, Pittsburgh, PA 15213, USA
8 Department of Physics, Yale University, New Haven, CT 06520, USA

Received 2022 August 1; revised 2022 October 6; accepted 2022 October 13; published 2022 November 21

Abstract

The eROSITA X-ray telescope, launched in 2019, is predicted to observe roughly 100,000 galaxy clusters. Follow-up observations of these clusters from Chandra, for example, will be needed to resolve outstanding questions about galaxy cluster physics. Deep Chandra cluster observations are expensive, and it is unfeasible to follow up every eROSITA cluster, therefore the objects that are chosen for follow-up must be chosen with care. To address this, we have developed an algorithm for predicting longer-duration, background-free observations, based on mock eROSITA observations. We make use of the hydrodynamic cosmological simulation Magneticum, simulate eROSITA instrument conditions using SIXTE, and apply a novel convolutional neural network to output a deep Chandra-like “super observation” of each cluster in our simulation sample. Any follow-up merit assessment tool should be designed with a specific use case in mind; our model produces observations that accurately and precisely reproduce the cluster morphology, which is a critical ingredient for determining a cluster’s dynamical state and core type. Our model will advance our understanding of galaxy clusters by improving follow-up selection, and it demonstrates that image-to-image deep learning algorithms are a viable method for simulating realistic follow-up observations.

Unified Astronomy Thesaurus concepts: Convolutional neural networks (1938); Neural networks (1933); Galaxy clusters (584); Observational cosmology (1146); Astronomical methods (1043); X-ray astronomy (1810); X-ray surveys (1824); Astronomy image processing (2306); Astronomy data analysis (1858)

1. Introduction

Galaxy clusters are the most massive gravitationally bound objects in the universe. They consist of scores to hundreds to thousands of galaxies in a common dark matter halo. Galaxies and the intracluster medium (ICM) form the ordinary baryonic matter component of these structures and emit light across the electromagnetic spectrum, allowing us to observe them. Through bremsstrahlung, collisional excitation, recombination radiation, and two-photon emission processes, the ICM produces X-ray photons, allowing for X-ray observations of clusters. Galaxy clusters are an important probe of dark matter (e.g., Clowe et al. 2006), and they are of special interest to cosmologists because they are the high-density peaks of the present-day universe and are sensitive to the underlying cosmological model (see Pratt et al. 2019 for a recent review).

The sensitivity of galaxy clusters to cosmological parameters makes them excellent probes of cosmology. The abundance as a function of mass and redshift provides us with information about the underlying cosmological model, in particular the matter density, \( \Omega_m \), and the amplitude of the matter fluctuations, \( \sigma_8 \) (Allen et al. 2011; Kravtsov & Borgani 2012; Pratt et al. 2019). Mass estimations of a population of clusters with a well-understood selection function can be used to construct a halo mass function (e.g., Tinker et al. 2008; Bočquet et al. 2016), which can be used to constrain cosmological models. X-ray observations are especially useful for mass estimation, because they provide low-scatter proxies of cluster mass (e.g., Kravtsov et al. 2006). However, mass estimations of clusters are reliant on mass proxies, which may result in biased estimates of the true mass (e.g., Nagai et al. 2007; Lau et al. 2013; Nelson et al. 2014; Biffi et al. 2016; Shi et al. 2016; Barnes et al. 2021).

The dynamical state of a cluster, which is a function of its mass accretion history, can substantially bias mass estimation (e.g., Lau et al. 2009; Nelson et al. 2014; Shi et al. 2015). To accurately correct the level of bias introduced into mass estimates, some understanding of the dynamical state of the clusters is needed. Conveniently, a substantial mass accretion history often leaves noticeable signals. It has a measurable impact on the radial density profile of the cluster outskirts (Diemer & Kravtsov 2014), the clumpiness of the clusters (Nagai & Lau 2011), and their morphology (Evrard et al. 1993). If one can control for the mass accretion rate, mass bias can be reduced. High-angular resolution, long-duration X-ray observations of galaxy clusters can provide information to characterize the dynamical states of clusters.

Galaxy cluster core astrophysics is also an area of active inquiry. Populations of clusters can be categorized according to the apparent cooling properties of their cores, ranging from cool-core clusters to non-cool-core clusters (Jones & Forman 1984). The origins of these different cluster types are
not fully understood, with the proposed theories requiring revision following improved observations (see Fabian 1994; McNamara & Nulsen 2012; Inoue 2022, and references therein for reviews). Detailed X-ray imaging of galaxy cluster cores is necessary to better understand the core dynamics.

eROSITA (the extended ROentgen Survey with an Imaging Telescope Array; Merloni et al. 2012) will provide an all-sky X-ray survey, and is expected to detect ~100,000 clusters (Pillepich et al. 2018). eROSITA’s observations are complementary to existing observatories, like Chandra. Whereas Chandra’s high-angular resolution observations offer detailed spatial information about individual clusters, eROSITA’s all-sky survey allows for a well-modeled selection function of the underlying galaxy cluster population. Improving our understanding of galaxy clusters will require leveraging both instruments, using eROSITA’s observed cluster populations to discover cluster populations of interest that are suitable for detailed follow-up observations. However, eROSITA will provide us with a plethora of potential follow-up candidates, far surpassing the ~10^5 large extended sources observed by Chandra (the Chandra Source Catalog release 2; Evans et al. 2019, 2020). Given Chandra’s operating constraints, such an increase in the number of observable clusters will be too large for us to conduct follow-up observations on any more than a small fraction of them. Future observers will need to carefully select follow-up candidates from the eROSITA survey.

The enormous disparity between Chandra and eROSITA in the number of galaxy clusters they can observe necessitates a follow-up merit assessment tool. In this work, we present a tool that, given an eROSITA observation, provides a prediction of a background-free, long-duration, follow-up observation. This tool illustrates the suitability of deep learning algorithms for follow-up merit assessment, by predicting a high-quality observation from a lower-quality observation more accurately and precisely than the original observation or a simple non-deep learning prediction method.

There is a long history of astronomers developing tools to improve the resolutions and the signal-to-noise ratios of their images (e.g., Richardson 1972; Lucy 1974; Cornwell & Evans 1985), and machine learning offers modern techniques for addressing this problem. For any resolution- or signal-boosting tool to be useful, it must appropriately capture the relevant galaxy cluster properties, and the relevant cluster properties are dependent on the science use case. Cluster shape is one example: morphology is an observable indicator of a cluster’s mass distribution, ellipticity, and substructure, and it is correlated to the cluster’s dynamical state (Melot et al. 2001; Rasia et al. 2013; Parekh et al. 2015; Losivari et al. 2017; Chen et al. 2019; Lau et al. 2021) and core type (Santos et al. 2008). Cluster shape measurements are also useful for estimating cluster mass (Green et al. 2019), and, for these reasons, morphological parameters such as the cluster’s surface brightness concentration (Santos et al. 2008), asymmetry (Lotz et al. 2004), and smoothness (Lotz et al. 2004) are relevant for identifying potential clusters for follow-up observations; see Rasia et al. (2013) and Ghirardini et al. (2022) for descriptions of common morphology parameters. Morphologically accurate prediction images will allow observers to efficiently select clusters based on dynamical state and core type, which will not only improve our understanding of those key cluster properties, but also advance our understanding of physics more broadly. For example, the selection of clusters by dynamical state is important for the study of dark matter (e.g., Eckert et al. 2022).

In addition to morphology, we evaluate our model’s capacity for denoising observations and predicting the total flux of the observation. Like most forms of astronomical observation, X-ray observations suffer from foreground and background contamination. While both eROSITA-like and Chandra-like observation strategies suffer from these issues, eROSITA’s shorter observation time and poorer angular resolution make it more challenging to disentangle signal from noise. Our follow-up merit assessment tool compensates for these differences, by reducing the background and differentiating between the emission from active galactic nuclei (AGNs) and the ICM.

Our tool uses a novel image-to-image convolutional neural network (CNN) trained on observations of simulated galaxy clusters from the hydrodynamic cosmological simulation Magnetic (Dolag et al. 2016). We accessed Magnetic via the Cosmological Web Portal (Ragagnin et al. 2017), which simulates mock cluster observations with the PHOX algorithm (Biffi et al. 2012, 2013). We used SIXTE to simulate realistic eROSITA observations (Dauser et al. 2019). A machine-learning algorithm, which by its nature learns from inputted data, is appropriate for our task, because it is the most capable of learning complicated nonlinear signals and correlations in data, such as we would expect to exist between eROSITA-like observations and the underlying astrophysical sources being observed (see Schmidhuber 2014 for a review of deep learning). We choose to use a CNN, in particular, because of its capability of learning localized patterns in inputted data, like gradients, textures, and patterns. CNNs have been critical to advances in image-to-image prediction and analysis (e.g., Ronneberger et al. 2015; Johnson et al. 2016). CNNs have been applied to a variety of problems in astronomy, like cosmic web simulations (Rodríguez et al. 2018), exoplanet atmospheres (Zingales & Waldmann 2018), image reconstruction (Flamary 2016), and image denoising (Vojtekova et al. 2021). In applying deep learning methods to galaxy cluster observations using cosmological simulations, we build on an established and proven practice (see, e.g., Ntampaka et al. 2015; Green et al. 2019; Ntampaka et al. 2019). Our work is also complementary to image-to-image neural network galaxy cluster cosmology work, like the Sunyaev–Zel’dovich effect image emulator used in Rothchild et al. (2022) and the XMM-Newton X-ray super-resolution and denoising algorithms developed in Sweere et al. (2022).

We present a machine-learning approach to enhancing eROSITA images to assess them for follow-up. This paper is organized as follows. In Section 2, we describe the simulated data that were used (Section 2.1), the structure of the algorithm (Section 2.2), and the manner in which it was trained (Section 2.3). In Section 3, we describe the performance of our model for each of the morphological parameters (concentration: Section 3.1; asymmetry: Section 3.2; and smoothness: Section 3.3), total flux (Section 3.4), background reduction (Section 3.5), and mass dependence (Section 3.6). In Section 4, we discuss the implications and limitations of our model. Section 5 presents the conclusion.

2. Methods

2.1. Data

Machine-learning methods, in combination with hydrodynamic cosmological simulations, offer a powerful tool for galaxy cluster science. Machine-learning methods, especially the CNN variant that we develop in Section 2.2, are exceptional
at learning complicated patterns in multidimensional data. The methods that we use require “labeled” data, meaning that we need many realistic observations of galaxy clusters paired with matching, longer-duration, background-free observations, which we refer to as “super observations.” Moreover, the choice of data determines the utility of the algorithm, meaning that we need realistic observations of accurately simulated clusters. We choose to use hydrodynamic cosmological simulations because they easily provide pairs of simulated eROSITA observations and super observations, while also closely matching the observed properties of AGNs and the ICM (see, e.g., Hirschmann et al. 2014; Rasia et al. 2015; Dolag et al. 2016).

This research requires a hydrodynamic cosmological simulation that is large enough to generate a diverse sample of clusters, with high-enough resolution to accurately model cluster substructures. To satisfy these constraints, we use the hydrodynamic cosmological simulation Magneticum (Dolag et al. 2016). Specifically, we use Box2/hr (Hirschmann et al. 2014), a 352 Mpc/h sized box with $2 \times 1584^3$ particles. It has a dark matter particle mass $M_{\text{DM}} = 6.9 \times 10^8 M_{\odot}$ and a gas particle mass $M_{\text{gas}} = 1.4 \times 10^8 M_{\odot}$, and provides 6927 clusters at redshift $z = 0.07$ with masses above $10^{13} M_{\odot}$, with a variety of mass accretion histories. The simulation uses cosmology constraints from Komatsu et al. (2011); i.e., a total matter energy density, $\Omega_m = 0.272$ with 16.8% baryons; a cosmological constant, $\Omega_\Lambda = 0.728$; the Hubble constant, $H_0 = 70.4$; a spectral index of the primordial power spectrum, $n_s = 0.963$; and an amplitude of matter fluctuations, $\sigma_8 = 0.809$.

Machine-learning algorithms are inherently data-driven, and therefore careful consideration must be given to data selection. To avoid biasing our model toward the more plentiful low-mass, low-redshift clusters, we choose a roughly uniform mass and redshift distribution of clusters. We do so by subsampling the available low-mass, low-redshift clusters. Each cluster is included in the data set only once, from a unique line of sight. Our data set has 3285 galaxy cluster observations. The observations have a depth of 10 Mpc, and include emission from the galaxy cluster, nearby neighboring galaxy clusters, and nearby AGNs. The emission from AGNs is simulated as detailed in Biffi et al. (2018). The observed galaxy clusters have a mass range of $3.16 \times 10^{13} M_{\odot}$ to $1.17 \times 10^{15} M_{\odot}$ and a redshift range of 0.07 to 0.47.

The goal of our work is to create an algorithm that is capable of predicting a high-quality observation from a lower-quality observation. To do this, we must train the algorithm using pairs of observations. The galaxy cluster observations are therefore split in two categories, mock eROSITA observations and super observations. The mock eROSITA observations begin with a field of view and a resolution matching that of eROSITA, with the field of view being roughly 1° and the pixel size being $9^\prime 6$. The observation time of these images is 2 ks. The expected background particle emission, instrument response, and point-spread function are simulated using the SIXTE software (Dauser et al. 2019). The super observations have the same detector area, field of view, and resolution as the eROSITA observations, but are background-free, lack any instrument response or point-spread function, and have an observation time of 10 ks. Because our overarching goal is to accurately predict follow-up observations including line-of-sight AGN sources, AGN sources within 10 Mpc of the central galaxy cluster are included in our super observations.

A background-subtracted eROSITA observation set, which we refer to as eROSITA-NR, is used as a baseline for comparing the prediction effectiveness to that of a non-machine-learning method. The per-pixel background is defined as the mean of all nonzero pixels in an annulus with an inner radius of 140 pixels and a width of 5 pixels. This annulus range is sufficiently large to be exterior to all of the $R_{500c}$ radii of the clusters in our data set. This is important, because we use the $R_{500c}$ radius, which is the radius in which the mass density of the cluster is 500 times the critical density of the universe, as a measure of the extent of the cluster. After the subtraction, all pixels with flux values less than zero are set to zero.

Each observation is divided into three energy bands, corresponding to soft X-rays (0.5–1.2 keV), medium X-rays (1.2–2.0 keV), and hard X-rays (2.0–7.0 keV). These bands are chosen following the definitions of the Chandra/ACIS science and source detection energy bands, and also to take advantage of the different spectral behaviors of AGN, ICM, and eROSITA particle backgrounds. The frequency of the photons is known exactly for the super observations, but for the mock eROSITA observations, the photons are sorted by their observed eROSITA-defined channel number (the pulse height amplitude, or PHA, channel). The soft, medium, and hard X-rays are divided by the PHA bands 74–177, 178–274, and 275–722, respectively. Due to memory constraints, we only used the inner $256 \times 256$ pixels of each image. This reduces the field of view to 40/96, but leaves the resolution unchanged. A summary of the observation image information is shown in Table 1. An example observation for each energy band and observation type is shown in Figure 1.

### 2.2. Algorithm

A CNN is a class of machine-learning algorithm that is often used for image processing tasks. They make use of what are called “convolutional filters.” In the two-dimensional case, these can be understood as two-dimensional matrices, where each element of the matrix is a parameter fitted during training. The images are convolved by sliding these filters over the

| Observation            | Field of View (Arcminutes) | Pixel Size (Arcseconds) | Exposure Time (s) | Background (Soft/Medium/Hard) (Counts/Pixel) |
|------------------------|----------------------------|-------------------------|-------------------|---------------------------------------------|
| eROSITA                | 40.96°                     | $9^\prime 6 \times 9^\prime 6$ | 2000              | 0.02/0.02/0.1                              |
| Super observation      | 40.96°                     | $9^\prime 6 \times 9^\prime 6$ | 10,000            | 0                                           |

Note. All observation types have the same field of view and the same angular resolution. They differ only in the observing time and background level. Super observations are ideal observations, being background-free, with long exposure times, perfect instrument response, and no point-spread function. Background counts are estimates from simulated blank-sky observations that were put through our pipeline.

---

https://cxc.harvard.edu/csc/columns/ebands.html

---
image, taking the dot product of the matrix and a given region of the image. The sizes of these filter matrices and the method of sliding them over an image are hyperparameters that are determined by the user. During training, these filters transition from random realizations to relevant feature detectors, commonly detecting edges, textures, and other patterns. By stacking layers of these filters on top of each other, the algorithm becomes capable of detecting more complex large-scale features, such as faces or animals. See Lecun et al. (2015) for a review of deep learning and CNNs.

Our model is engineered to focus on accurately probing multiple length scales simultaneously. The images in our data set contain three primary components: the ICM, AGNs, and the background, all of which are limited in size to length scales smaller than the entirety of the image. These small characteristic length scales encouraged us to abandon more traditional algorithm architectures, like the UNET (Ronneberger et al. 2015), which rely on reducing images into a small set of globally important features. Our model instead primarily consists of two layers that are split into three paths. Each path has different sized filters (1 × 1, 3 × 3, or 5 × 5), each with 200 filters. We constructed and trained our model using the Python (Van Rossum & Drake 2009) module Tensorflow (Abadi et al. 2015).

In the first layer, an input image observation is fed directly into each path. After having the filters applied to it, the image is compressed back into a three-band image, using 1 × 1 filters, thus leaving us with three different three-band images. These images are concatenated together to form a single nine-band image. This image is then also reduced into a three-band image,

![Figure 1. Example observations of a sample galaxy cluster. The cluster shown has M_{500c} = 8.2 × 10^{13} M_\odot and is at z = 0.07. The rows correspond to the soft, medium, and hard X-ray energy bands. The columns correspond to the different observation types (see Section 2.1): from left to right, the eROSITA mock image used as the machine-learning input ("eROSITA"); a background-subtracted image ("eROSITA-NR"); the machine-learning output ("Prediction"); and the ground truth long-duration mock observation ("Super Observation"). The flux color mapping is in log space, with the same color scaling for each image, from a color minimum of 0 photons per pixel to a color maximum of 10 photons per pixel. The predictions of the machine-learning method developed in this research visibly outperform the eROSITA and background-subtracted eROSITA observations. While the soft-band eROSITA observation captures most of the shape of the central cluster and the merging cluster, the substructure of the central cluster and the neighboring AGN are not obviously distinguishable. The soft-band prediction, on the other hand, shows most of this information clearly. This superior performance is most visible in the hard band, where eROSITA is less sensitive and more noisy. The hard-band eROSITA observation displays the core of the central cluster in a field of background emission, whereas the prediction clearly shows the full shape of the central and merging cluster. For a more quantitative comparison of the results, see Section 3.](image)
using $1 \times 1$ filters. This composite three-band image is then fed into another layer of the same structure as the first, albeit with differently trained weightings. The outputs of each of these paths in the second layer are then concatenated with the three-band output of the corresponding path of the first layer. Filters are applied to these concatenations, so that they are transformed into three-band images. These three resulting three-band images are then themselves concatenated with the initial input. The resulting 12-band image is then subsequently reduced to a three-band image output. Our resulting model has 48,687 trainable parameters. A diagram of this algorithm is shown in Figure 2 and is described in Table 2.

We choose these three different filter sizes to correspond to the relevant image properties. Different astrophysical objects and features have different spatial and spectral properties; e.g., AGN flux is spatially compact and is brighter in hard X-ray, while ICM flux is spatially correlated on larger scales and is dominant in soft X-ray. The $1 \times 1$ filter path examines purely spectral information, comparing the ratio of the fluxes in each energy band, which we believe improves background

---

**Figure 2.** A diagram of the CNN architecture. We input a $256 \times 256$ three-band pixel image. The three bands correspond to the three X-ray energy bands used, which are chosen to leverage both eROSITA’s energy-dependent sensitivity and the difference between AGN and ICM spectra. A further discussion of the energy bands is found in Section 2.1. The number of bands is shown in each box, either in parentheses or as the leftmost number. This is equal to the number of filters applied to the output of the preceding step (the steps are linked by arrows). The three paths, each corresponding to a different filter size, are shown in different colors. Their filter sizes are the two rightmost numbers. From top to bottom, the sizes of the filters in each path are: $1 \times 1$ (red), $3 \times 3$ (blue), and $5 \times 5$ (green). The dotted arrows indicate that a concatenation occurs before the filters are applied. This is done to probe different relevant length scales of an image, without compressing the $256 \times 256$ pixel size of the image. Instead, images are repeatedly reduced to three-band images, to reduce the memory burden.

**Table 2**

A Table of the CNN Algorithm

| Layer Name | Layer Type | Preceding Layer(s) | Filter # | Kernel Size | Activation |
|------------|------------|--------------------|----------|-------------|------------|
| x0         | Input      | None               | None     | None        | None       |
| x1         | Conv2D     | x0                 | 200      | $1 \times 1$| LeakyReLU  |
| x3         | Conv2D     | x0                 | 200      | $3 \times 3$| LeakyReLU  |
| x5         | Conv2D     | x0                 | 200      | $5 \times 5$| LeakyReLU  |
| x1b        | Conv2DTranspose | x1 | 3      | $1 \times 1$| LeakyReLU  |
| x3b        | Conv2DTranspose | x3 | 3      | $1 \times 1$| LeakyReLU  |
| x5b        | Conv2DTranspose | x5 | 3      | $1 \times 1$| LeakyReLU  |
| x_concat   | Concat     | x1b, x3b, x5b     | None     | None        | None       |
| z0         | Conv2DTranspose | x_concat | 3      | $1 \times 1$| LeakyReLU  |
| z1         | Conv2D     | z0                 | 200      | $1 \times 1$| LeakyReLU  |
| z3         | Conv2D     | z0                 | 200      | $3 \times 3$| LeakyReLU  |
| z5         | Conv2D     | z0                 | 200      | $5 \times 5$| LeakyReLU  |
| z1b        | Concat     | x1, z1             | None     | None        | None       |
| z3b        | Concat     | x3, z3             | None     | None        | None       |
| z5b        | Concat     | x5, z5             | None     | None        | None       |
| z1c        | Conv2DTranspose | z1b | 3      | $1 \times 1$| LeakyReLU  |
| z3c        | Conv2DTranspose | z3b | 3      | $1 \times 1$| LeakyReLU  |
| z5c        | Conv2DTranspose | z5b | 3      | $1 \times 1$| LeakyReLU  |
| z_concat   | Concat     | x0, z1c, z3c, z5c | None     | None        | None       |
| Output     | Output (Conv2DTranspose) | z_concat | 3      | $1 \times 1$| Linear     |

**Note.** A $256 \times 256$ three-band pixel image is fed into the algorithm. Through an interconnected series of convolution and transpose convolution layers, we achieve accurate predictions of cluster morphology, while maintaining a smaller model. More details are available in Section 2.2 and Figure 2.
suppression and AGN identification, as both of these image features have limited spatial correlations but unique spectral behaviors. The $3 \times 3$ and $5 \times 5$ filter paths then ought to identify spatial correlations from the ICM, with two length scales chosen to account for both the substructure and the varying pixel size of the clusters, given their mass distribution and redshift.

2.3. Training

Standard supervised machine-learning training involves inputting data into an algorithm and then comparing the output (i.e., the prediction) to the label of the input (i.e., the truth value of the property of interest). The comparison is computed using a loss function. The weights of the algorithm are then changed to minimize the output of the loss function. In our case, the inputs are the mock eROSITA observations. The labels are the super observations. The loss function is a linear combination of the mean absolute error (i.e., the mean absolute difference between the pixels of the prediction image and the super observation) and the “morphology loss,” defined as the linear combination of the mean absolute error of three morphology parameters: surface brightness concentration, asymmetry, and smoothness. We choose this loss function in order to emphasize the properties of the cluster that we view as being the most important. By minimizing the morphology loss of the algorithm in training, we improve the morphology parameters derived from the predictions produced by the algorithm. We also tried a perceptual loss function, inspired by Johnson et al. (2016), using the third layer of the VGG19 network (Simonyan & Zisserman 2014). This is discussed in Section 4.

For the loss function, we use “fixed” versions of the morphology parameters, wherein either a fixed radius (in pixels) or the entire image is used to calculate the parameter. This is used in the algorithm due to its simplicity and consistency; it does not require information about the redshift or the size of the central cluster. The fixed parameters are calculated as follows. The fixed surface brightness concentration is the ratio of the fluxes within 10 pixels and 100 pixels of the center of the image. The fixed asymmetry parameter is calculated using the absolute difference of the full image and the same image rotated 180°, and is equal to the sum of the pixel values of this difference image normalized by the total flux of the original image. The fixed smoothness is calculated by applying an 11 pixel boxcar smoothing to the full original image, calculating the absolute difference between the smoothed image and the original image, summing the total flux of the difference image, and then normalizing that by the total flux of the original image. The fixed concentration, asymmetry, and smoothness parameters are described in Equations (1), (2), and (3), respectively. $X$ is the observation image, $X_{180}$ is the observation image rotated 180°, $\hat{X}$ is the smoothed observation image, and $F$ is the total flux within a radius $r$ (if $r$ is unstated, the full image is used), where $r$ is in units of pixels. Examples of morphology parameter extremes, albeit for the $R_{500c}$ versions described in Section 3, are shown in Figure 3.

\[
A = \frac{F(|X - X_{180}|)}{F(X)}.
\]

\[
S = \frac{F(|X - \hat{X}|)}{F(X)}.
\]

Our training set—the data that we set aside specifically for training the weights of the algorithm—constitutes 80% of the full data set. 10% of the remaining data forms our validation set. The validation set is used to evaluate the training progress, in order to determine when to stop the training. The algorithm is saved after each epoch when it achieves a minimal validation loss. This procedure is repeated until the validation loss minimum is stable for over 100 epochs. The final 10% of the data forms the test set, which is used to analyze the efficacy of the fully trained algorithm. The full data set is shuffled prior to partitioning.

3. Results

Prediction algorithms like the one we are proposing must have clearly defined use cases. To illustrate the ways in which our model might benefit observers, we have defined several important predictive capabilities. These metrics are informed by the nature of the problem that we seek to solve: a follow-up observation of an initial eROSITA observation will have a higher signal-to-noise ratio, a more physically accurate and more constrained luminosity profile, more visible substructure, and a more definite shape. A prediction algorithm must therefore seek to increase the signal-to-noise ratio in the input image; increase the brightness of astrophysical sources, while differentiating between extended sources and point-spread function–blurred AGNs; and predict the location and brightness of real substructure. Similarly, the enhanced properties of the prediction must provide useful information for survey selection. That is why we choose to prioritize morphology parameters, which provide useful information about the properties of the cluster, especially those properties—such as dynamical state, mass, and core type—that are of particular importance to galaxy cluster research. Note that while we prioritize morphologically accurate observations, we do not intend our prediction observations to be used for morphology measurements directly. Morphology is simply a metric that we use to grade image accuracy (see Section 4 for a discussion about assessing image accuracy and morphology).

Our intention is to aid in the selection of follow-up candidates, not to replace follow-up observations. Therefore, predictions must be probable enough to aid the discernment of which clusters merit following up. Metrics like those that have previously been mentioned provide an understanding of the accuracy of the predictions. An example prediction observation is shown in Figure 1. Example observations of similar mass clusters with very different morphology parameters are shown in Figure 3. In the following subsections, we examine the predicting power of our trained model for the three key morphology parameters (concentration: Section 3.1; asymmetry: Section 3.2; and smoothness: Section 3.3), total flux (Section 3.4), and background reduction (Section 3.5). We also evaluate the precision and accuracy of the predictions as a function of mass (Section 3.6). We use the $R_{500c}$ versions of the morphology parameters. This formulation of the morphology parameters is more commonly used and more physically meaningful than the fixed version of the morphology.
parameters that are used in the loss function, but it requires information about the cluster radius and redshift. Our true morphology parameters are calculated in observations with AGNs, but the level of contamination from AGNs in the soft band, where the ICM dominates, is minimal.

3.1. Concentration

Surface brightness concentration is a morphological parameter that estimates how centrally concentrated the mass of a galaxy cluster is, and it is a key probe of a variety of cluster properties, including mass error (Green et al. 2019), core type (Santos et al. 2008), and dynamical state (Rasia et al. 2013; Parekh et al. 2015; Lovisari et al. 2017). Given its ubiquity and usefulness as a metric, concentration is an important parameter for our prediction images to be able to accurately replicate. Our $R_{500c}$ concentration parameter is modeled on the variant used by Lovisari et al. (2017) and Green et al. (2019). The $R_{500c}$ surface brightness concentration calculates the ratio of the fluxes within $0.1 \times R_{500c}$ and within $R_{500c}$ of each cluster. The concentration varies from 0, a minimally concentrated cluster, to 1, a maximally concentrated cluster. Equation (4), shown below, describes the concentration calculation. Here, $F$ denotes the total flux within a pixel radius $r$:

$$C = \frac{F(r \leq 0.1 \times R_{500c})}{F(r \leq R_{500c})}.$$  

True concentration is the concentration derived from the super observations. We find that the concentration values derived from the machine-learning prediction observations are consistently closer to the true values, compared to the concentrations derived from the eROSITA or eROSITA-NR images. In our test set data, we find that the predicted $R_{500c}$ concentrations differed from the truth value by $(\Delta C_{\text{soft}}, \Delta C_{\text{med}}, \Delta C_{\text{hard}}) = (0.02^{+0.05}_{-0.04}, 0.08^{+0.10}_{-0.07}, -0.03^{+0.08}_{-0.09})$, with the $+/-$ values indicating the 84th and 16th percentile values in our test set results. The predicted soft-band concentrations therefore have a median difference that is roughly three times smaller than the eROSITA concentrations, with smaller scatter. This superiority is a reflection of the trained model’s ability to simultaneously reduce background, while boosting signal. This becomes especially apparent in the concentration parameters of the hard band, where high background dramatically degrades the accuracy. Here, the median difference of the prediction

---

**Figure 3.** Soft-band observations of roughly equal-mass clusters with varying $R_{500c}$ morphology parameters. All clusters have a redshift of 0.07. Overlaid in blue is the $R_{500c}$ radius. For the leftmost column, the smaller blue circle is $0.1 \times R_{500c}$, which is used to calculate the concentration. Left column: an example of low (top) and high (bottom) concentration. Center column: an example of low (top) and high (bottom) asymmetry. Right column: an example of low-valued (top) and high-valued (bottom) smoothness. The larger values in the smoothness parameter indicate a peaky flux distribution. The bottom cluster is strongly concentrated (the same cluster as the bottom left), leading to a higher smoothness value. The top cluster is clearly in a merger, which leads to a flatter flux distribution, meaning a more smoothly filled $R_{500c}$ radius. Note that the extremes in the morphology parameter space tend to overlap; i.e., more centrally concentrated clusters tend to be more symmetric. For example, the same cluster is shown in the bottom left (highly concentrated) and top center (highly symmetric). Similarly, very asymmetric clusters tend to be less centrally concentrated. The top left cluster (less concentrated) and bottom center cluster (highly asymmetric) are the same.
observation–derived concentration is half that of the background-subtracted eROSITA concentration, the next most accurate concentration for that energy band, with half as much scatter.

Plots of the concentration results as a function of the true (super-observation) concentrations are shown in Figure 4. In the top plots, we show the median difference between the concentration derived by our prediction, the eROSITA observation, the eROSITA-NR observation, and the super observation for the corresponding bins shown in the histograms directly below the plots. The shaded regions reflect the range of values in 68% of the data in a given bin. The prediction-derived concentrations have consistently lower bias, and the results from the data have consistently lower scatter across the energy bands and true concentrations. Moreover, the bias and scatter of the concentration predictions are relatively consistent across the true concentration bins, unlike the eROSITA-derived results. Our results are evidence that a machine-learning prediction model is better suited to accurately and precisely predicting the concentrations of a diverse population of clusters than either the native eROSITA observations or non-machine-learning improvement. See Table 3 for a full quantitative comparison of the results.

### 3.2. Asymmetry

The asymmetry of a cluster, meaning its deviation from circularity in its two-dimensional profile, is another key morphological parameter. Several metrics have been devised to probe asymmetry, including ellipticity and photon asymmetry (see Ghirardini et al. 2022). Our $R_{500c}$ asymmetry metric is modeled on the variants used by Rasia et al. (2013) and Green et al. (2019). We choose this formulation because of its simple implementation, usefulness as a probe of cluster dynamical state (Rasia et al. 2013), and informativeness in cluster mass estimation (Green et al. 2019). Symmetric clusters are less likely to be disturbed, and thus estimations of their mass are probably less biased. Asymmetry is an obvious selection criterion for a variety of cosmological uses, therefore accurately predicting asymmetry is necessary for our model to be useful.

The $R_{500c}$ asymmetry is calculated using the same procedure as the fixed asymmetry discussed in Section 2.3, but uses only pixels within $R_{500c}$. The asymmetry varies from 0, perfectly symmetric, to 2, maximally asymmetric. Equation (5), shown below, describes the calculation for asymmetry. Here, $F$ denotes the total flux, $X$ is the observation image, $X_{180}$ is the observation image rotated 180°, and $r$ is the radius within which the flux is calculated:

$$A = \frac{F(|X - X_{180}|; r \leq R_{500c})}{F(X; r \leq R_{500c})}. \quad (5)$$

The predicted asymmetry values are more closely correlated with the super observation–derived asymmetries than those from eROSITA or eROSITA-NR are. Asymmetry, compared to concentration, is poorly constrained in the eROSITA observations, and is biased relative to super observation–derived asymmetries. This is unsurprising, since asymmetry probes the less luminous outskirts of galaxy clusters, and is therefore more sensitive to background and short observation times. Despite this limitation, the prediction observation–derived asymmetries are minimally biased across energy bands. In our test set data, we found that the predicted $R_{500c}$ asymmetries differed from the truth value, as calculated using the super observations, by $(\Delta A_{\text{soft}}, \Delta A_{\text{med}}, \Delta A_{\text{hard}}) = (0.03^{+0.16}_{-0.12}, -0.04^{+0.16}_{-0.18}, 0.06^{+0.20}_{-0.24})$.

![Figure 4. $R_{500c}$ surface brightness concentration calculated for our full data set, as described in Section 3.1. The top plots show the difference between the calculated concentrations from the eROSITA, eROSITA-NR (eROSITA with background subtracted; see Section 2.1), and prediction observations, compared to the true concentration. The true morphological parameters are calculated using the super observations. The results are binned by their true concentration value, with the number per bin shown in the histograms in the bottom panels. The medians of the data in each bin (points) and the middle 68% (error bars) are shown. For all energy bands, the prediction observation–derived surface brightness concentrations are the most accurate. The median difference and scatter of those predictions appear to be independent of the true concentration, implying that our prediction observations are valid for a diverse population of clusters.](image-url)
Table 3
A Table of the Median and 68% Intervals for the Difference between the eROSITA/eROSITA-NR Prediction-derived Values and the Super Observation-derived Values for the Morphology Parameters (Concentration, Asymmetry, and Smoothness), Total Flux, and Background Removal

| Observation          | Band  | Concentration | Asymmetry | Smoothness | Total Flux | Background |
|----------------------|-------|---------------|-----------|------------|------------|------------|
| Training + Validation|       |               |           |            |            |            |
| eROSITA              | Soft  | −0.07±0.05    | 0.19±0.17 | 0.23±0.15  | 0.72±0.45  | 1115±24    |
|                      | Soft  | 0.14±0.13     | 0.12±0.14 | 0.04±0.09  | −0.06±0.04 | 10±4       |
| Prediction           | Soft  | 0.02±0.04     | 0.04±0.13 | 0.04±0.09  | −0.06±0.05 | 138±19     |
| eROSITA              | Medium| −0.09±0.07    | 0.16±0.21 | 0.12±0.17  | 0.71±0.27  | 122±54     |
|                      | Medium| 0.14±0.15     | 0.19±0.31 | 0.25±0.22  | −0.01±0.04 | 11±3       |
| Prediction           | Medium| 0.08±0.08     | −0.04±0.19| −0.02±0.12 | 0.00±0.05  | 100±15     |
| eROSITA              | Hard  | −0.38±0.22    | 0.37±0.21 | 0.04±0.33  | 3.41±0.66  | 6530±1038  |
|                      | Hard  | −0.05±0.18    | 0.20±0.33 | 0.24±0.34  | 0.1±0.21   | 311±22     |
| Prediction           | Hard  | −0.09±0.09    | 0.05±0.18 | −0.01±0.09 | 0.01±0.04  | 152±24     |
| Test Data            |       |               |           |            |            |            |
| eROSITA              | Soft  | −0.07±0.05    | 0.17±0.17 | 0.22±0.15  | 0.72±0.49  | 1107±238   |
|                      | Soft  | 0.14±0.13     | 0.15±0.39 | 0.39±0.26  | −0.07±0.18 | 10±4       |
| Prediction           | Soft  | 0.02±0.05     | 0.02±0.16 | 0.02±0.06  | −0.07±0.05 | 137±16     |
| eROSITA              | Medium| −0.09±0.07    | 0.17±0.22 | 0.12±0.17  | 0.71±0.29  | 1216±194   |
|                      | Medium| 0.14±0.15     | 0.13±0.29 | 0.25±0.27  | −0.01±0.05 | 12±3       |
| Prediction           | Medium| 0.08±0.01     | −0.04±0.18| −0.02±0.14 | 0.01±0.06  | 100±15     |
| eROSITA              | Hard  | −0.37±0.2     | 0.37±0.23 | 0.02±0.39  | 3.4±0.66   | 6519±209   |
|                      | Hard  | −0.07±0.15    | 0.26±0.38 | 0.21±0.39  | 0.09±0.05  | 307±33     |
| Prediction           | Hard  | −0.03±0.08    | 0.06±0.24 | −0.03±0.11 | 0.00±0.06  | 151±19     |

Note. See Section 3 for a description of each evaluation metric. See Section 2.1 for information about the observation types and energy bands. Results for both the training + validation combined data set and the test set are shown. The prediction-derived results for the test set are consistent with those from the training + validation combined data set, suggesting that the trained model has not over-fit the training data.

Like concentration, the strongest relative performance of the trained model’s predictions for asymmetry occurs in the hard band. Here, the median difference is ~six times smaller, with comparable scatter, compared to the eROSITA results, and ~four times smaller, with nearly half as much scatter, as the eROSITA-NR results. Plots of the asymmetry results as a function of the true (super-observation) asymmetries are shown in Figure 5. In the top plots, we show the median difference between the super-observation asymmetry and the asymmetry derived by our prediction, the eROSITA observation, or the eROSITA-NR observation. The results are binned by the super-observation-derived asymmetry, with histograms of the bins shown directly below the plots. The shaded regions reflect the range of values in 68% of the data in a given bin. Excluding the smallest-asymmetry bin, which is too underpopulated to derive meaningful results, the bias in the asymmetry predictions is consistently lower than that of the other results. The asymmetry data set is not balanced across asymmetry parameter space, which is probably the cause of the slight dependence on true asymmetry. Very symmetric clusters are generally under-represented in our data set, and therefore our model will struggle to predict them accurately. This could be corrected with access to more data. See Table 3 for more quantitative information about the results.

3.3. Smoothness

As they grow, galaxy clusters accrete nearby dark matter and baryons. This process results in clumpy substructure within the cluster, which can be visible in X-ray observations. The smoothness morphology parameter is a measure of the amount of substructure in a galaxy cluster, which, in turn, provides information about the dynamical state of the cluster (Rasia et al. 2013). Like the other morphology parameters, smoothness also provides important information for cluster mass estimation (Green et al. 2019). We adopt the smoothness parameter as defined in Green et al. (2019), which is similar to the fluctuation parameter used in Rasia et al. (2013).

To calculate the $R_{500c}$ smoothness, like the fixed smoothness discussed in Section 2.3, we again apply an 11 pixel boxcar smoothing scale, but only to the pixels within $R_{500c}$ of the center. Equation (6), shown below, describes the calculation for smoothness. Here, $F$ denotes the total flux, $X$ is the observation image, $\tilde{X}$ is the observation image after boxcar smoothing has been applied, and $r$ is the radius within which the flux is calculated:

$$S = \frac{F((X - \tilde{X}); r \leq R_{500c})}{F(X); r \leq R_{500c}}.$$  (6)

As is the case with all the key morphology parameters, the predicted smoothness values are more closely correlated with the super-observation-derived smoothness than those from eROSITA or eROSITA-NR are. In our test set data, we found that the predicted $R_{500c}$ smoothness differed from the truth value, as calculated using the super observations, by $(\Delta S_{\text{true}}, \Delta S_{\text{med}}, \Delta S_{\text{band}}) = (0.03±0.06, -0.02±0.14, -0.03±0.11)$. See Table 3 for more information.
Plots of the smoothness results as a function of the true (super-observation) smoothness are shown in Figure 6. In the top plots, we show the median difference between the smoothness derived by our prediction, the eROSITA observation, the eROSITA-NR observation, and the super observation for the corresponding bins shown in the histograms directly below the plots. The shaded regions reflect the range of values in 68% of the data in a given bin. The plots illustrate the low bias and low scatter of the prediction-derived results in the data set across energy bands and the smoothness parameter space. As in the case of the asymmetry results, the strength of the results shown is limited by the unbalanced coverage of the smoothness parameter space. Very smooth clusters, and very clumpy clusters in soft X-ray, are not well represented. Machine learning is by nature data-driven, so a lack of data in this parameter space could result in more inaccurate results.

### 3.4. Total Flux

Do prediction observations conserve the total flux of the observation? Because morphology parameters all involve normalization by the total flux, they shed little light on that question. Conservation of the total flux is a potentially important feature for prediction observations, however, and merits investigation. We perform a simple test for this and find that the prediction observations do preserve the flux, and that they do so with less scatter than the background-subtracted eROSITA observations. See Figure 7 for a plot of the results. Figure 7 illustrates the power of our model in conserving the total flux, especially with regard to the soft band, where the scatter in the difference between the predicted and the true total flux is negligible. The results are also listed in Table 3.

### 3.5. Background Reduction

Observations of galaxy clusters, like all astronomical observations, are hindered by various forms of background. Background in an observation has multiple sources. Background is generated both by non–galaxy cluster X-ray sources, particles, and as a product of the instrument itself. One potential use of an observation prediction model is to reduce background wherever possible. The machine-learning model can achieve this by utilizing mult+wavelength information, as the cluster signal-to-background ratio of eROSITA is higher at lower frequencies. Additionally, as the model learns the general shape of galaxy clusters and AGNs, it can better suppress pixels that do not conform to their luminosity profiles. It is important to reiterate that what the model produces is only a prediction, so some background reduction will be inaccurate.

We can quantify the background reduction by comparing our background-free super observations to our eROSITA-like input observations and the prediction observations. We apply a Gaussian smoothing filter to the super observations in order to mimic the effects of both the point-spread function and the difference in the random realizations of photons between observations. We then catalog the pixels with zero flux in the smoothed images. For the rest of our observations, we define the background as the total flux of those pixels. Plots of the background results are shown in Figure 8. Table 3 quantifies the results. We find that background subtraction, as described in Section 2.1, outperforms our model’s background suppression in the soft- and medium-energy bands. It does so at a cost to the accuracy of the $R_{500c}$ morphology parameters, however, especially asymmetry and smoothness (see Figures 5 and 6). The model outperforms subtraction in the hard-energy band, though, where the eROSITA observations are especially background-dominated. This advantage in performance could
be useful for observations that are reliant on the hard band, like AGN-focused observations. Note that our mock eROSITA observations only included simulated particle background, not X-ray foreground, or other X-ray background sources.

3.6. Mass Dependence

In addition to checking for bias in the morphological accuracy of our prediction observations as a function of true morphology, we also examined whether there were any trends as a function of cluster mass. We found no obvious mass dependence in the soft band, which is the band most useful for measuring cluster morphology. The trends that were apparent in the higher-energy bands were very small, and the scatter in the difference between the true and predicted parameters was roughly consistent across the energy bands. These results suggest that the model predicts morphologically accurate and precise observations for the clusters within the mass range of our data sets (i.e., $\sim 10^{13}$ to $10^{15} M_\odot$). See Figure 9 for a visualization of the results.

4. Discussion

Our model demonstrates the powerful potential of machine learning for assessing the merit of follow-up observations. We have shown that our prediction images preserve morphology, which correlates to important galaxy cluster properties like dynamical state and core type. This is important, because to resolve outstanding questions about these properties, we need high-resolution, long-duration observations from instruments like Chandra, which have limited availability. The model’s useful prediction observations will allow galaxy cluster observers to effectively and efficiently evaluate which eROSITA-observed clusters merit follow-up observations. Predicted observations are not a replacement for follow-up observations, nor are predicted morphology parameters a replacement for real morphology parameter measurements, but predictions are useful for determining whether an eROSITA-observed cluster is likely to have a property of interest (e.g., a cool core) that merits a follow-up observation. Machine-learning models are most effective when they are used in an ethical way, with an understanding of their strengths and limitations. There are ways in which an informed user might adapt or improve a model. Below, we discuss the strengths and limitations of our model, and the ways in which one might overcome those limitations.

4.1. Classification

A strength of our model, as we have shown, is that our prediction observations are more morphologically accurate than either eROSITA or eROSITA-NR observations. It is worth noting, however, that this relative accuracy can be applied in a variety of unexpected ways. For example, imagine that an observer wishes to perform follow-up observations on only the most highly disturbed clusters. To do so, they would like to determine which clusters are in the 90th percentile of asymmetry. We can test this premise on our own data, by
examining whether the 10% most asymmetric clusters according to the prediction observations match the 10% most asymmetric clusters as determined by the super observations. In doing so, we find that our trained model can determine whether a cluster is among the 10% most asymmetric clusters with a true positive rate (TPR) of 59% and a false-positive rate (FPR) of 5%. In other words, 59% of the clusters in the 90th percentile of true asymmetries are found in the 90th percentile of predicted asymmetries. Of the clusters that are not in the 90th percentile of true asymmetries, only 5% are found in the 90th percentile of predicted asymmetries. To identify the 90th percentile of soft-band concentration or soft-band smoothness, the TPRs and FPRs improve to TPR_C = 82%, FPR_C = 2%, TPR_S = 73%, and FPR_S = 3%, respectively. The number of possible classifications is infinite, and we cannot provide an exhaustive list of the TPR and FPR rates, but we encourage users to recognize this potential application of our model.

4.2. Domain Shift

A drawback of using simulated data is the problem known as domain or distributional shift (see Section 7 of Amodei et al. 2016 for a discussion). Simulated data will invariably differ from real data. These differences, rooted in both the computational constraints of simulating a universe as well as our still limited understanding of important cosmological and astrophysical phenomena, will limit the utility of a model that has been trained solely on simulated data. The morphology parameters, by their nature, are deeply intertwined with the baryonic physics of galaxy clusters (e.g., Lau et al. 2011, 2012; Chen et al. 2019; Machado Poletti Valle et al. 2021). These physics are the most difficult to simulate, and the least well understood in modeling, and therefore they often vary from cosmological simulation to cosmological simulation. This increases the likelihood that biases induced by domain shift will be present when our model is applied to real data, or to data from a different cosmological simulation. In addition to simulation-related biases, we are limited by the morphological parameter space of our data set. Clusters at the extremes of the parameter space—e.g., very smooth clusters—are relatively unrepresented. In the regions of cluster morphology parameter space with limited training data (see the bottom plots of Figures 4, 5, and 6), one should use the assessment tool’s predictions with caution.

A potential solution to this problem is to make use of what is known as transfer learning. We can retrain a model that has previously been trained on simulated observations on pairs of observations (e.g., eROSITA and Chandra observations of the same clusters). This technique can correct the distributional shift between different data sets—in this case, simulated data and real data. It has the additional advantage of requiring many fewer real data observation pairs than would be required to completely train a new model on real data. This technique has been applied to a variety of problems in astronomy (e.g., Ackermann et al. 2018; Domínguez Sánchez et al. 2019; Pérez-Carrasco et al. 2019).

4.3. Redshift, Observing Time, and Resolution Dependence

We are not the first to recognize the importance of morphology in sample selection. For example, Mantz et al. (2015) developed symmetry–peakiness–alignment morphology
specifically to provide a more robust characterization of cool-core clusters across different redshifts and observing instruments. While we opt for an alternative construction of morphology for computational reasons, our concentration–asymmetry–smoothness morphology provides similar information about cool cores, and we characterize robustness issues relating to observing instrument and redshift. In the case of redshift, we find no substantial redshift bias in concentration, asymmetry, or smoothness. This is true across all energy bands, but especially in the soft band. We have also performed a minor analysis of the effects of different observing instruments ourselves, investigating the changes in the morphology parameter values for an individual cluster as a function of resolution and observing time (see Figure 10). We find that the parameter values do change as a function of observing time and resolution. However, our results also suggest that the uncertainty of the morphology parameters is minimal, even for lower observing times and coarser resolutions. This suggests that given a representative sample of clusters observed by a single instrument, the morphology parameters of a cluster still provide useful information when they are compared to the distribution of morphology parameters of clusters observed by that instrument. Given the number of clusters that eROSITA is expected to observe, this constraint is not problematic. Moreover, while morphology parameters that are derived from the observations of one instrument do not directly map onto those derived from the observations of another, the relationship between morphology and observing instrument can still be characterized.

Our own model is proof that a machine-learning model can learn the relationship between the morphology parameters of 2 ks and 10 ks observations. This suggests that machine-learning models could characterize the relationship between morphology parameters that are derived from the observations of one instrument and those that are derived from the observations of another. With this in mind, we advise that this or any other follow-up merit assessment tool should be developed by giving careful consideration to its intended use. One use case that we envision for our tool is informing observers as to whether an eROSITA-observed cluster has an extreme morphology relative to other eROSITA-observed clusters (e.g., very asymmetric versus very symmetric). This information would be useful for determining whether a cluster has a dynamical state or core type that merits selection for follow-up. On the other hand, if one intends to predict Chandra-resolution morphology parameters exactly, then one should train a model using Chandra-resolution images as the truth images, instead of eROSITA-resolution images, as we did.

4.4. Alternative Metrics and Models

In our work, we have aimed to create predictions that build on established techniques in the field of cluster cosmology. For computational and scientific reasons, we opted for the simple but well-established concentration, asymmetry, and smoothness parameters. While there is strong evidence of the correlation of these parameters with our properties of interest (e.g., Rasia et al. 2013; Parekh et al. 2015; Lovisari et al. 2017; Green et al. 2019), these are not the only morphology parameters used in cluster

![Figure 8. Background calculated for our full data set, as described in Section 3.5. The top plots show the background (B) from the eROSITA, eROSITA-NR, and prediction observations, plotted against the eROSITA observation background (B\textsubscript{eROSITA}). Results are binned by their eROSITA background, as shown in the histogram plots in the bottom panels. The medians of the data in each bin and the regions representing 68% of the results for the bins are plotted above. The background subtraction method, described in Section 2.1, outperforms the prediction in terms of background reduction for the soft- and medium-energy bands. The prediction observations have lower background in the hard band, however, likely due to the weak signal in the eROSITA observations at energies above 2 keV. Moreover, the prediction observations have the benefit of consistent background across energy bands, with only a weak dependence on the input observation’s background. It should also be noted that the background subtraction method, while better at reducing background, also removes signal, and has worse morphology parameter accuracy.](image-url)
Figure 9. The morphology parameter results are shown above. The morphology parameters, as derived from mock eROSITA, eROSITA-NR (background-subtracted), and prediction observations, are compared to the true values, derived from the super observations. The top multicolored plots show the difference between the observed parameters and the true parameters, binned by mass. The bins are roughly equal in population. The points and error bars show the medians and the 16th to 84th percentile ranges of the results, and are plotted in the middle of the mass bin ranges. The bottom orange plots illustrate the true parameters as a function of mass. The prediction-derived constraints, shown in green, with median values marked by triangles, are clearly superior to the alternatives. This is especially apparent in the soft band, where the median difference and scatter are small for all of the parameters. These results suggest that our model produces morphologically accurate and precise observations without a noticeable mass bias for clusters between $\sim 10^{13}$ and $10^{15} M_{\odot}$.
cosmology, and the statistical properties of these parameters are not well understood. This is especially true of the asymmetry parameter, as there exists plentiful literature on anisotropy tests and measures with better-characterized statistical properties (for examples and reviews of this topic, see Mardia & Jupp 1999; Feigelson & Babu 2012; Pewsey et al. 2013; Baddeley et al. 2015; Rajala et al. 2018). While an analysis of the correlation of alternate anisotropy parameters to cluster properties of interest is outside the scope of this work, we encourage others to explore applying new measures of anisotropy to galaxy cluster cosmology.

Morphology is not the only potentially relevant observable for the outstanding questions in galaxy cluster cosmology. Other relevant properties, such as X-ray scaling relations (e.g., Lx–Tx, Lx–M, and Tx–M), X-ray surface brightness profiles, reconstructions of three-dimensional gas and temperature profiles, or hydrostatic mass profiles, might be of interest to X-ray observers. We chose morphology parameters because they are computationally easy to calculate (which is very relevant for training a model), applicable to observations of individual clusters, and valid for the available simulation data. There is a wide parameter space of valid galaxy cluster properties to investigate; and, moreover, our research is of potential interest to astronomy beyond X-ray galaxy cluster observation. We encourage others to explore this ample parameter space, but we have opted to fully explore what we view as the most promising avenue for success.

We chose to focus on the morphological accuracy of observations, and therefore designed our model particularly for that purpose. Alternative algorithm architectures are also possible and ought to be considered, depending on the priorities of the user. We considered different loss functions, including a perceptual loss function, inspired by Johnson et al. (2016), which uses the output of the third layer of the VGG19 network (Simonyan & Zisserman 2014). An example prediction from that model is shown in Figure 11. The perceptual loss function model produced prediction observations that better preserved concentration in the soft band than the morphology loss model. Moreover, the prediction images from the perceptual loss model are arguably more realistic, appearing smoother and less noisy, while clearly preserving substructure and AGNs. However, the prediction observations did not preserve asymmetry or smoothness effectively, nor did they preserve concentration in the hard band. We valued the morphology loss model’s preservation of asymmetry and smoothness more than the improvements in concentration and appearance that the perceptual loss model offered, but we recognize that other observers might have different priorities. In addition, we considered using a standard UNET architecture (Ronneberger et al. 2015), but it did not function well with our chosen loss function. The choice of the evaluation metric that is used to test the utility of the trained model is also important. We chose to focus on morphology parameters, but many image-to-image machine-learning models were instead designed to produce realistic-looking images that could fool human observers (see Dahl et al. 2017 for examples of different image accuracy metrics, including human evaluation). Machine learning offers a wide parameter space in terms of algorithm hyperparameters, and we do not argue that the model that we have presented is the most optimal. Instead, we argue that it illustrates that a machine-learning approach for follow-up merit assessment is not only possible, but also potentially more powerful than alternative solutions.

5. Conclusion

Galaxy clusters are an important probe of cosmology and useful laboratories for understanding physics. Key cluster properties, like core cooling and dynamical state, remain poorly
understood and are in need of further study. Detailed follow-up observations provide insights to these properties, but resources for follow-up are limited, and an efficient method for evaluating the merit of these observations is needed. Machine learning offers one such method.

Our follow-up merit assessment tool can predict background-free, long-duration observations with accurate and precise morphology parameters. Given morphology’s correlation with the aforementioned cluster properties—which are important in their own right as well as useful for selecting clusters for dark matter studies and mass estimation—morphologically accurate observations will aid follow-up selection. Our model will therefore advance our understanding of galaxy cluster internal physics, galaxy cluster cosmology, and cosmology more broadly.

Additionally, our work illustrates how follow-up merit assessment might be approached for a variety of different observational goals. Our model has been designed to prioritize morphological accuracy, but the model could be trained as needed to address different properties. In working on this problem, we explored a wide range of models, many with their own strengths and deficiencies. We encourage observers to strongly consider the priorities of their own observations when designing a follow-up merit assessment tool, then tailor it to their specific needs. When utilized appropriately, machine learning can be an incredibly powerful tool for advancing galaxy cluster cosmology.

The material presented is based upon work supported by NASA under award No. 80NSSC22K0821. This research project was conducted using computational resources at the Maryland Advanced Research Computing Center (MARCC). We thank the MARCC support staff for invaluable help. We thank Christian Kirsch, Ole König, Maximilian Lorenz, and

Figure 11. Plots of different observations of the same cluster. From left to right: mock eROSITA inputs, morphology loss predictions, perceptual loss predictions, and the super-observation targets. The perceptual loss images performed worse at matching the morphology parameters of the super observations, so we instead presented our morphology loss model. The perceptual loss model produces more visually appealing, and arguably more realistic-looking, predictions than the morphology loss model. The perceptual loss model also offers superior accuracy in the soft-band concentration parameter. The perceptual loss model introduces nonphysical structures into the image, and its prediction observations have worse morphological accuracy in asymmetry and smoothness relative to the morphology loss model. We note this to emphasize the important decisions that one must take in designing, training, and evaluating an algorithm. Our model has been designed with a specific use case in mind, but other follow-up merit assessment algorithms might choose different algorithm architecture or accuracy metrics.
