Optical Wavelength Guided Self-Supervised Feature Learning For Galaxy Cluster Richness Estimate

Gongbo Liang[1,2], Yuanyuan Su[1], Shen-Chieh Lin[1], Yu Zhang[1], Yuanyuan Zhang[3], Nathan Jacobs[1]

[1]University of Kentucky [2]Eastern Kentucky University [3]Fermi National Accelerator Laboratory

Abstract

Clusters of galaxies are the largest gravitationally-bound systems in the Universe. Their optical richness (λ) is crucial for understanding the co-evolution of galaxies and large-scale structures in modern astronomy and cosmology.

Challenges

The measurement of λ is confronted by foreground and background objects, leading to the uncertainty of the location of a given galaxy along the line of sight. The membership of cluster galaxies can be determined with their spectroscopic redshifts. But spectroscopic surveys are expensive. Photometric surveys in multiple wavelength ranges, i.e., multi-band optical imaging, are much more achievable. The imaging modality is efficient for detecting galaxy clusters, but the richness of a galaxy cluster cannot be directly estimated.

Method

We propose a two-phase CNN model with a self-supervised training strategy that estimates λ from multi-band optical images. We first use the optical wavelength band to train an image feature extractor through a band classification task. Since the band labels are known, no manually annotated data is needed at this stage. Then, a λ estimation model can be built upon the pre-trained feature extractor.

Result

We demonstrated the proposed method on the Sloan Digital Sky Survey (SDSS). The evaluation result shows that our method significantly improves the λ estimate while reducing the need for labeled training data by up to 60%.

Project Page: www.gb-liang.com/owg

Architecture

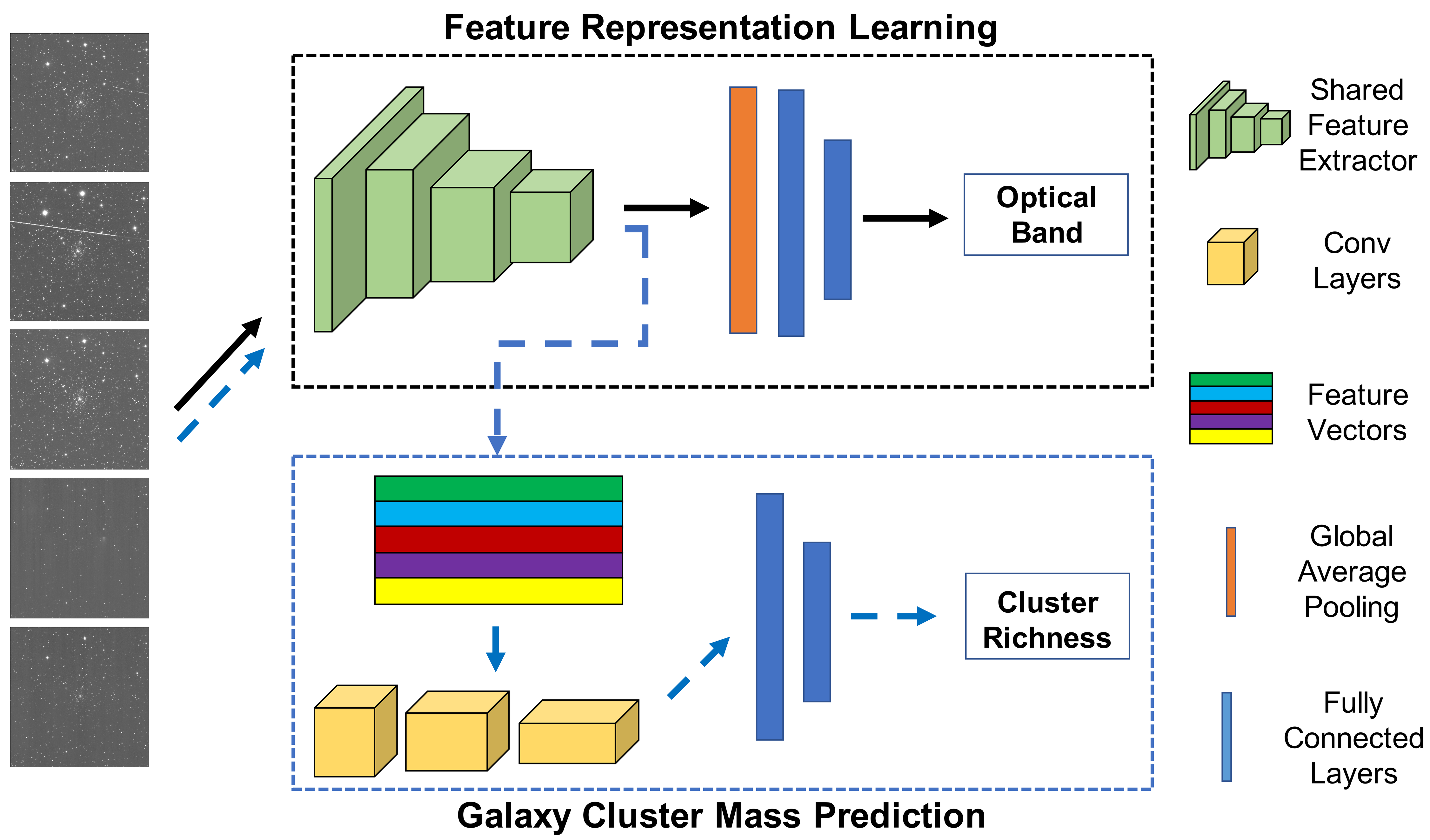

Figure 1: Two branches self-supervised learning architecture: 1) feature extractor via optical band classifications (solid black line); 2) galaxy cluster richness estimation (dashed blue line).

Galaxy Cluster Richness Estimate

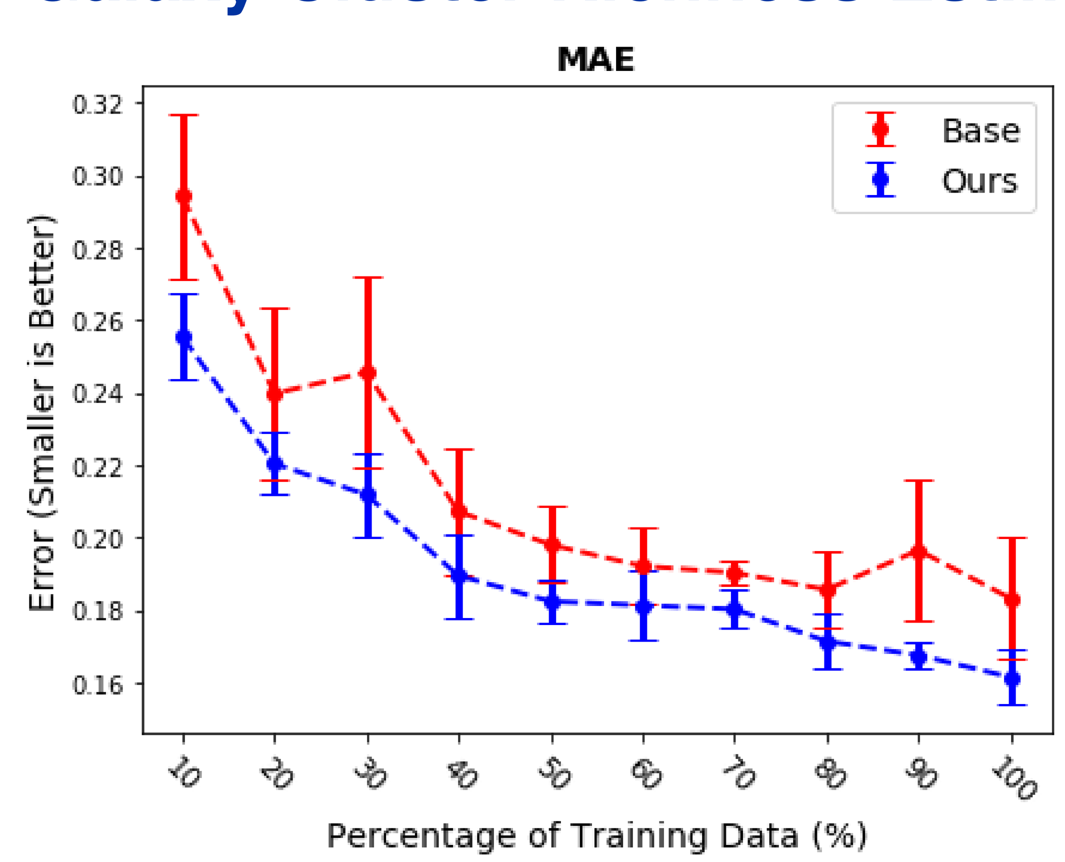

Figure 2: The mean absolute error (MAE) of Base model and the proposed method (Ours).

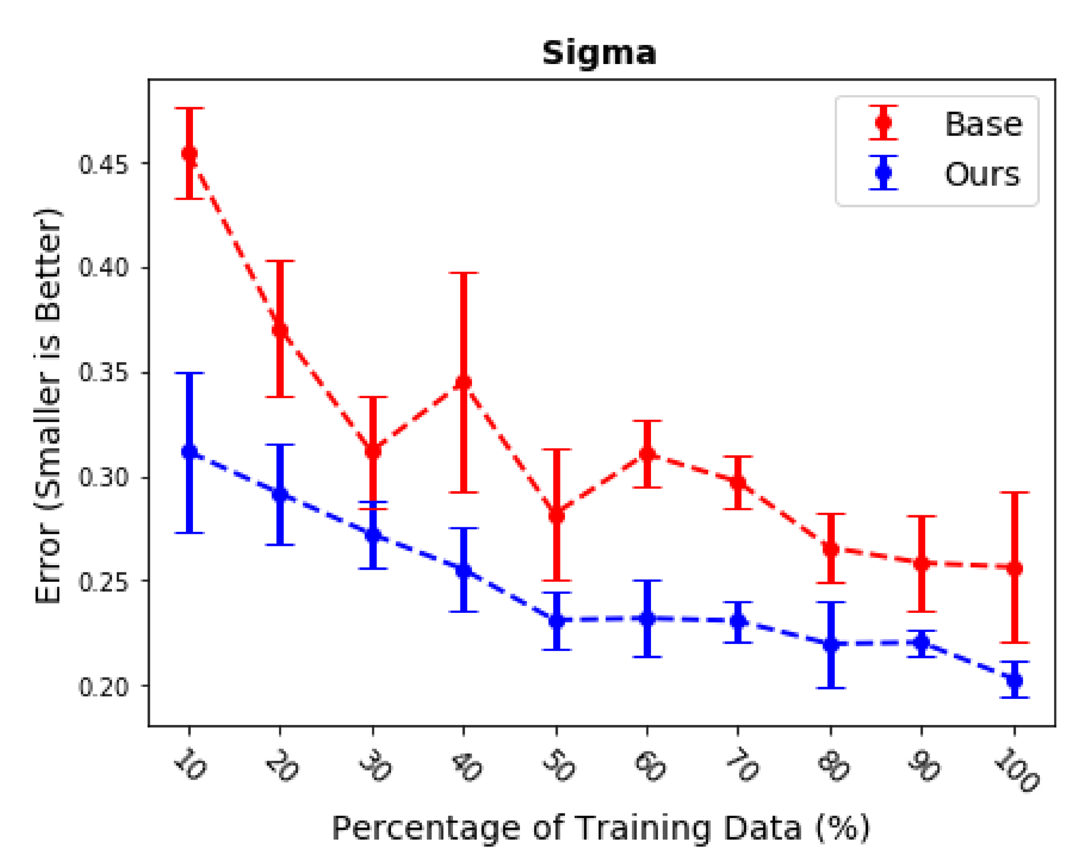

Figure 3: Intrinsic scatter between theground-truth richness and predicted richness (Sigma) of Base model and the proposed method (Ours).

Evaluation Method

We evaluate the proposed method based on the λ estimation and the degree of need for labeled instances for training. We use the MAE and intrinsic scatter between the ground-truth richness and predicted richness (Sigma) as the evaluation metrics for λ estimation. We compare the proposed method (Ours) against a baseline model (Base), which has the same architecture as Ours but without pre-training. We train multiple models for Base and Ours using different percentages of the training data (1% to 100%) with10-fold cross-validation. Each model was trained for three trials.

Pre-Trained Feature

We evaluate the feature extractor by applying an occlusion test on the pre-trained optical band classification network. We found that a significant performance change often happens when occluding foreground stars or member galaxies. This systematic phenomenon may indicate such information is critical to the classification decision. Thus, we believe the pre-trained features may well represent both foreground stars and member galaxies. Conceptually speaking, galaxy richness estimation is a process of separating member galaxies from other objects in the same image. The ability to represent member galaxies and foreground stars is an important criterion that leads to λ estimate.

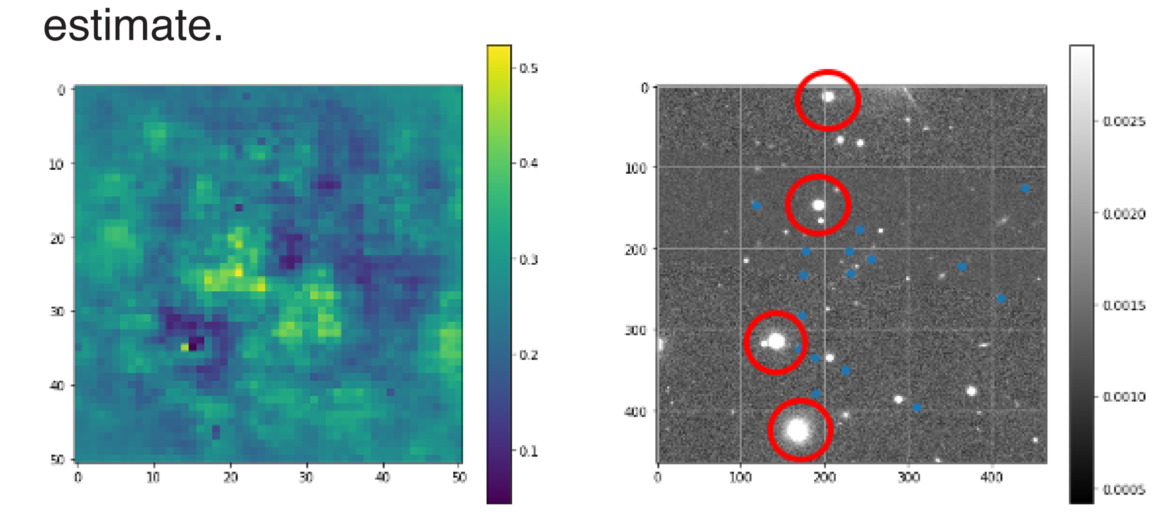

Figure 4: An occlusion testing example. Left: an optical image, Right: occlusion map, Blue dot: member galaxy, Red circle: foreground star.