MAE Galaxy Classifier Project Report

Arin Idhant

Environment Setup

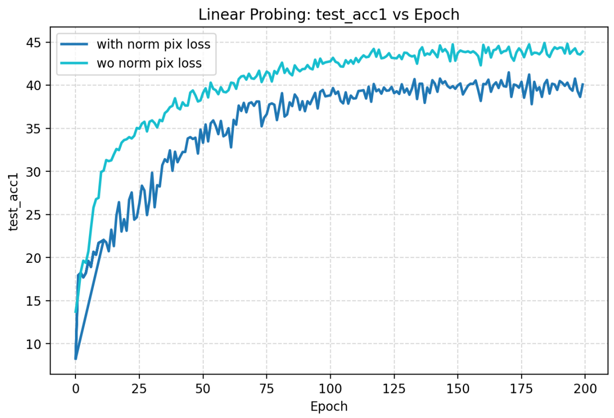
For this project, I implemented Meta's Masked Autoencoder (MAE) to pre-train a Vision Transformer (ViT Base Patch 16). The ViT-Base model was chosen because our dataset was relatively small, and training a larger model would have required more computational resources. Instructions to run the code are in Install.md. **Code Link.**

Pre-training Experiments

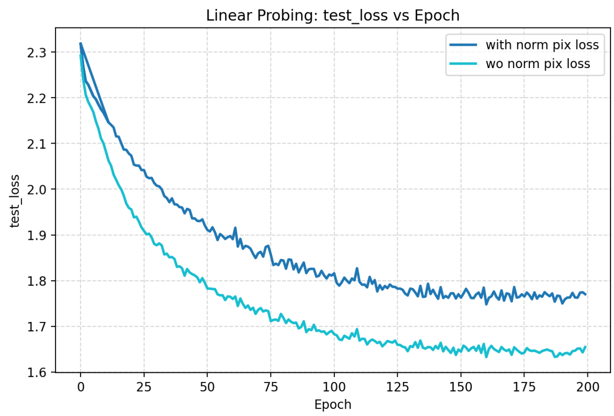
Loss: Normalized Pixel Loss vs Mean Squared Error

In the project, I compare the difference between Mean Squared Error Loss and Normalized Pixel Loss. The latter normalizes the target image patch so that brightness has a lower influence on the final reconstruction loss. For our galaxy image, it appears that the brightness values are important in classifying the galaxies and normalizing them migh result in loosing out on intensity features.

Below are the results from the linear probe accuracy:



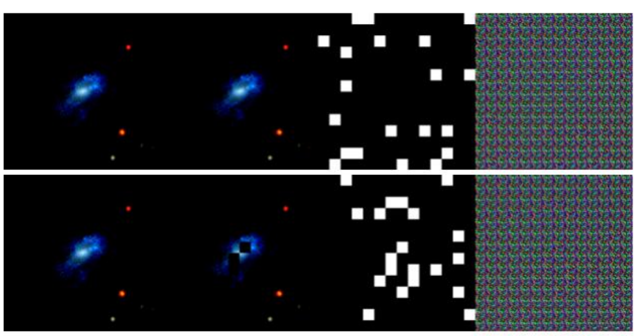
Test Accuracy Comparison



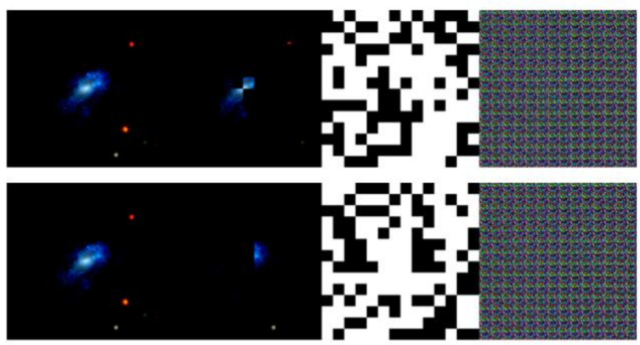
Test Loss Comparison

Patching Ratio and Distribution

The MAE paper uses a 0.75 masking ratio, which works for ImageNet. However, for our sparse galaxy images, masking empty space doesn't help the model learn much. I tried lower masking ratios and also experimented with non-uniform random masking by biasing the sampling toward the image center.



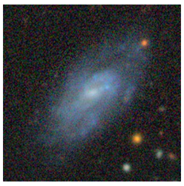
(below) Center Biased Random Masking. Mask Ratio = 0.1



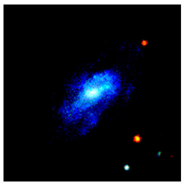
(below) Center Biased Random Masking. Mask Ratio = 0.6

Augmentations

Some of the augmentations used in the MAE paper seemed ill-suited for our dataset. Particularly, random cropping proved to be detrimental to learning, as the crop ratio would frequently crop out important details about the Galaxy. Instead, I end up selecting data augmentations like center cropping and rotations (since our galaxies are located in the center of the images). Furthermore, I experimented with normalization and gamma compression to bring out the details in our image.

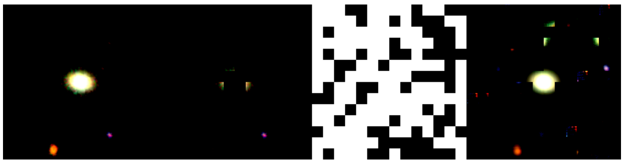


Without Normalization



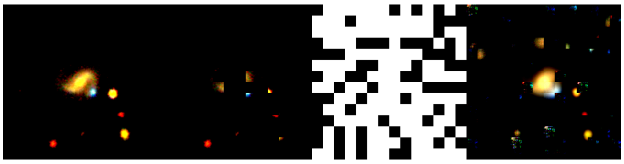
With Normalization

Below are examples of reconstructed images from the normalized dataset, with and without gamma compression:

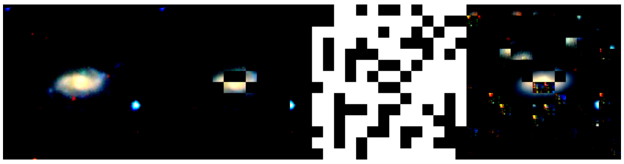


Without Gamma

Without Gamma



With Gamma



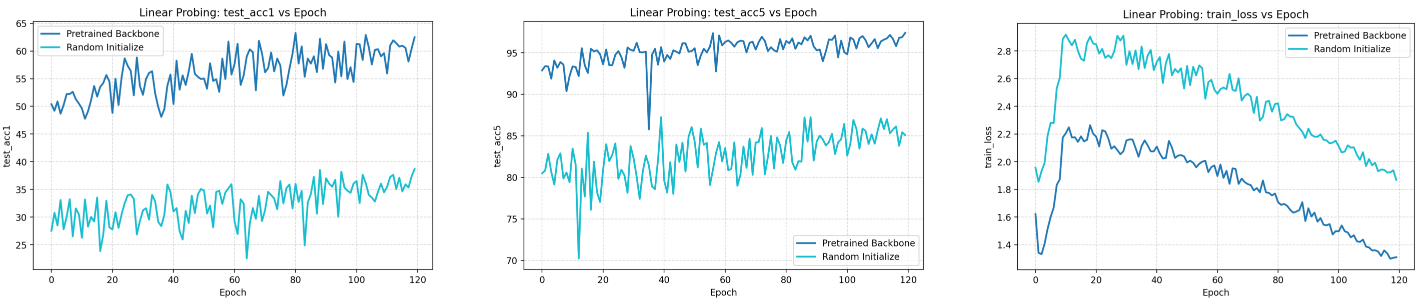
With Gamma

Although this made the images visually clearer, it did not significantly improve training performance.

Linear Probing

To evaluate the learned features, I used a linear probe for classification. Using the Adam optimizer and my customized augmentations helped improve the accuracy of the linear probe.

Below are results comparing my pretrained backbone to a randomly initialized ViT. After 200 epochs, I achieved a maximum accuracy of 69.06%.



Linear Probe Performance Comparison

Conclusion

This project explored how MAE pre-training and tailored augmentations can improve feature learning for galaxy classification.