Accounting for Spatial Variability with the Histogram of Oriented Gradients Based Masking Improves Performance of Masked Autoencoder over Hyperspectral Satellite Imagery (Student Abstract)

Tanjim Bin Faruk, Abdul Matin, Shrideep Pallickara, Sangmi Lee Pallickara

Department of Computer Science, Colorado State University
Fort Collins, 80526, USA
{Tanjim.Faruk, Abdul.Matin, Shrideep.Pallickara, Sangmi.Pallickara}@colostate.edu

Abstract

Masked autoencoders employ random masking to effectively reconstruct input images using self-supervised techniques, which allows for efficient training on large datasets. However, the random masking strategy does not adequately tap into information encapsulated within high-dimensional hyperspectral satellite imagery that is used in several domains. We propose a novel masking strategy, *HOGMAE*, based on the Histogram of Oriented Gradients that incorporates rich information inherent within satellite images during the mask creation step. Our experiments, over a hyperspectral satellite dataset, demonstrate the effectiveness of our methodology.

Introduction

Masked autoencoders (MAE) apply self-supervised techniques to capture latent representations from input images. MAEs accomplish this by randomly masking a significant portion of the input image and reconstructing the original image from the masked image. MAE operates on images with shape $C \times H \times W$, where C represents the number of spectral channels or bands, H is the height and W is the width. MAE resizes the input image into a set of contiguous, non-overlapping square patches of size of P^2. This results in a sequence of patches $N \times P^2C$, where $N = \frac{HW}{P^2}$ is the number of patches. A fraction r of the N tokens is masked and the remaining unmasked tokens are fed to the encoder. The decoder operates on all N patch tokens and reconstructs the input image using a specified loss function, computed only on masked patches.

To reduce computational overloads, the original MAE architecture operates over only a small portion of visible patches using a random masking strategy. Although the random masking performs effectively for datasets such as ImageNet , where the objects have mostly uniform shapes and sizes, it is less effective for satellite images that exhibit highly diverse patterns. Satellite images, in particular, can encompass a wide range of landscape and topographical features, and the relationships between pixels are not always uniform. To effectively capture spatial patterns and variations while ensuring high model performance, the masking strategy should be complemented with a more targeted approach specifically suited to satellite imagery.

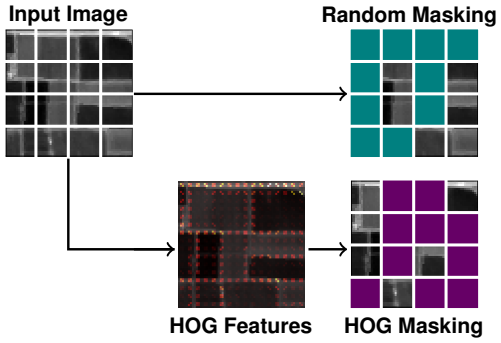

Figure 1: Contrasting *random masking* vs *HOG-based masking*. Our HOG-based masking preferentially selects areas with higher spatial variability for encoder learning.

Recent work on MAE masking strategies has explored using attention maps from pre-trained vision transformers (Chen et al. 2023; Zhu et al. 2023a). However, this approach requires an additional pre-training step to generate these attention maps, as the transformer's parameters must be learned before the attention patterns can be utilized. Other approaches such as SpectralMAE (Zhu et al. 2023b) have experimented with fixed masking strategies during fine-tuning stages, but this requires knowing in advance which spectral bands need to be reconstructed - a limitation in real-world scenarios where band failures may be unpredictable. Additionally, most of these attention-based methods have only been evaluated on traditional RGB image datasets. While some studies have applied MAE to satellite imagery, they have largely relied on simple random masking strategies (Lin et al. 2023; Cong et al. 2024). In contrast, our HOG-based masking approach does not require any pre-training or prior knowledge of which bands will need reconstruction, making it more practical for real-world hyperspectral satellite imagery applications.

HOG Based Masking Strategy

Our methodology preferentially selects masked tokens based on how they contribute to spatial variability. This preferential masking boosts learning by guiding the model to focus on areas with higher information density. Rather than as-

signing each patch an equal probability during the selection process, we boost priorities for some patches over others. We compute the spatial variability of each patch based on the Histogram of Oriented Gradients (HOG) (Dalal and Triggs 2005) and inform our masking strategy based on these computed features. HOG tracks the occurrence of gradient orientations in localized portions of an image. Therefore, patches with higher HOG values are associated with greater amounts of visual information in the image, such as shape and edges. This allows the model to prioritize patches with higher variability for retention during the masking phase, facilitating the encoder's learning over those areas.

Since HOG is typically applied on a single channel, we apply Principal Component Analysis (PCA) (computed using qrpca (S. de Souza et al. 2022)) to reduce the spectral dimensionality and identify the channel image patch with the highest variance across the channels. The HOG features are then extracted at the patch level. We sort the patches based on the HOG features and categorize them into *low* and *high* groups based on an adjustable split ratio. Next, we generate the mask by randomly selecting patches from both groups with a bias towards lower variability patches.

In particular, we assign two separate local masking ratios, one for highly variable patches and another for lower variability patches. We dynamically adjust the local ratios such that during training, more patches are masked from the low variability group and the number of masked tokens selected from the two groups maintains the specified (and configurable) masking ratio r. Our proposed HOG-based masking strategy is contrasted with the random masking strategy in figure 1.

Performance Benchmarks & Analysis

Dataset We retrieved hyperspectral satellite images from the Environmental Mapping and Analysis Program (EnMAP). The spatial resolution of the images is 30m, and each hyperspectral image comprises 224 bands. The final dataset comprised 218 bands after removing bands with invalid data. To facilitate training, we created 64×64 spatial resolution tiles from the original hyperspectral images. The entire dataset contains 22078 such tiles, divided into a training dataset with 12466 tiles, a validation dataset of 2078 tiles, and a test dataset with 7534 tiles. The test dataset is further divided into two groups: Test dataset 1 with 6234 tiles and Test dataset 2 with 1300 tiles. Test dataset 2 contains *"unseen"* areas, which represent geospatial regions that the model was not exposed to during training.

Evaluation Metrics We used Mean Absolute Error, Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index (SSIM) to evaluate the reconstruction accuracy. SSIM is used to assess the quality of images by contrasting structural similarity between the reference image and its altered representation. PSNR measures the quality of compressed images compared to their original versions.

Experiment We trained the models for 300 epochs and set the masking ratio to 75% for both masking strategies. We used the Mean Absolute Error as the reconstruction loss function.

Masking Strategy	Test Dataset 1			Test Dataset 2		
	MAE	PSNR	SSIM	MAE	PSNR	SSIM
Random	0.0126	34.39	0.8582	0.0136	34.22	0.8466
HOG	**0.0106**	**35.94**	**0.8899**	**0.0112**	**36.08**	**0.8872**

Table 1: Test Accuracy Comparison of Masking Techniques

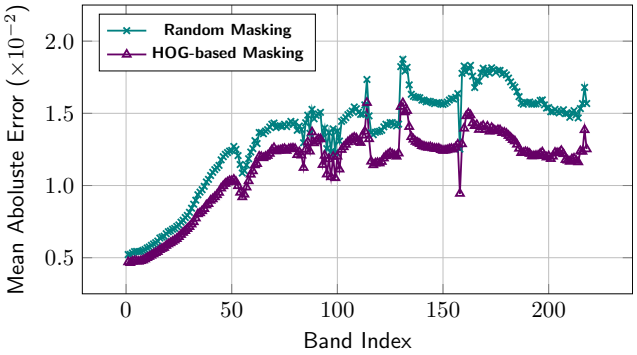

Figure 2: Bandwise Accuracy for Unseen Spatial Region

Results

HOG-based masking outperforms random masking on both partitions of the test dataset; please see table 1. The improvement is particularly pronounced for test dataset 2, where the reconstruction accuracy for HOG-based masking achieves a 17.6% reduction in MAE loss, a 5.4% increase in PSNR, and a 4.8% increase in SSIM. Figure 2 illustrates that our HOG masking strategy is able to significantly minimize the reconstruction errors across all bands for the unseen region (test dataset 2) compared to random masking. Our HOG-based strategy also demonstrates more effectiveness with higher bands which exhibit greater spectral variability and pose significant reconstruction challenges compared to lower bands.

Conclusions

Our HOG-guided dynamic masking strategy improves the performance of masked autoencoders on hyperspectral imagery. Because our masking strategy prioritizes patches with higher spatial variability, it is able to leverage rich information inherent within satellite images. Our experiments validate the suitability of our methodology and demonstrate that the HOG-based masking method outperforms random masking. This also underscores the potential of our proposed strategy in applications that leverage such imagery, for example, in agriculture, forestry, and other terrestrial processes.

Acknowledgements

This research was supported by the National Science Foundation (1931363, 2312319), the National Institute of Food Agriculture (COL014021223), and an NSF/NIFA Artificial Intelligence Institutes AI-CLIMATE Award [2023-03616].

References

Chen, H.; Zhang, W.; Wang, Y.; and Yang, X. 2023. Improving Masked Autoencoders by Learning Where to Mask. In *PRCV 2023*.

Cong, Y.; Khanna, S.; Meng, C.; Liu, P.; Rozi, E.; He, Y.; Burke, M.; Lobell, D. B.; and Ermon, S. 2024. SatMAE: pretraining transformers for temporal and multi-spectral satellite imagery. In *NeurIPS '22*.

Dalal, N.; and Triggs, B. 2005. Histograms of oriented gradients for human detection. In *CVPR'05*.

Lin, J.; Gao, F.; Shi, X.; Dong, J.; and Du, Q. 2023. SS-MAE: Spatial–Spectral Masked Autoencoder for Multisource Remote Sensing Image Classification. *IEEE TGRS*.

S. de Souza, R.; Quanfeng, X.; Shen, S.; Peng, C.; and Mu, Z. 2022. qrpca: A package for fast principal component analysis with GPU acceleration. *Astronomy and Computing*, 41: 100633.

Zhu, H.; Chen, Y.; Hu, G.; and Yu, S. 2023a. Information-density Masking Strategy for Masked Image Modeling. In *2023 IEEE ICME*.

Zhu, L.; Wu, J.; Biao, W.; Liao, Y.; and Gu, D. 2023b. SpectralMAE: Spectral Masked Autoencoder for Hyperspectral Remote Sensing Image Reconstruction. *Sensors*, 23(7).