SatMAE++: Self-supervised Learning for Multi-Spectral Satellite Imagery Analysis

1. Implementation Details

SatMAE++ thus takes the concept and applies it to the success of MAE architecture in processing multi-spectral satellite imagery. The actual implementation will contain:

> Multi-spectral data of the masked satellite images thus forms the input for the encoder, which then compresses this information into latent representations. In this respect, the decoder attempts to reconstruct the input by making predictions over the masked patches. Self-Supervised Pretraining : This model is pre-trained with a mask ratio of 50percent on images that are multi-spectral. The model learns representations without being helped by any labeled data through the prediction of masked patches. Fine-tuning: The fine-tuning of the model takes place after pretraining on downstream tasks like land-cover classification as well as object detection. Training Information: Pretraining is conducted over 800 epochs with a batch size of 256 and an initial learning rate of 1e-4 by utilizing AdamW as an optimizer. For fine-tuning, there are 100 epochs with a lower learning rate of 1e-5.

2. Dataset Description

Dataset: The dataset used for training and evaluation is the Functional Map of the World (fMoW) dataset. fMoW contains multi-spectral satellite imagery from various Earth observation satellites, covering 5 spectral bands: Red, Green, Blue (RGB), Near-Infrared (NIR), and Shortwave Infrared (SWIR).

The dataset comprises over 1 million images, representing 62 classes of functional land use and infrastructure types, such as airports, schools, wind farms, and industrial sites. The images span multiple geographical regions and various time periods, capturing a wide range of environmental and weather conditions, seasonal changes, and human development activities.

Preprocessing:

- **Normalization:** Each spectral band is normalized based on its min-max values to ensure consistency across images.
- **Data Augmentation:** Augmentations applied to the dataset include random crops, horizontal and vertical flips, rotations, and spectral channel swaps. This improves the model's robustness to variations in satellite imagery.
- **Resolution:** The spatial resolution of the images varies from 0.3m to 1.2m per pixel, depending on the satellite source and the region of interest.

Preprocessing:

- **Normalization:** Each spectral band is normalized based on its min-max values.
- **Data Augmentation:** Augmentations include random crops, rotations, and spectral channel swaps to enhance robustness.