

SatMAE: Pre-training Transformers for Temporal and Multi-Spectral Satellite Imagery

Name Praanshu Gautam Roll No-210728

1. Dataset Description

The **Functional Map of the World - Sentinel-2 (fMoW-Sentinel)** dataset consists of image time series captured by the Sentinel-2 satellite. These images correspond to locations from the Functional Map of the World (fMoW) dataset, spanning multiple time points. The locations are categorized into 62 different types of building and land use, as defined by the fMoW dataset. The images have a spatial resolution of 10 meters and are generated as cloud-free composites over 90-day intervals, with each image containing one channel for each of the 13 bands of the Sentinel-2 surface reflectance dataset.

The dataset is divided into training, validation, and test sets according to the original fMoW splits. Specifically, it contains:

- 712,874 images in the training set,
- 84,939 images in the validation set,
- 84,966 images in the test set.

The metadata for these splits is provided in the following CSV files: `train.csv`, `val.csv`, and `test_gt.csv`.

The fMoW-Sentinel dataset is derived from two data sources, each governed by its own license:

- The *Functional Map of the World Challenge Public License* applies to the metadata, including the image locations and categories.

2.Implementation Details

Create a separate google colab file and clone the repo. Add all the datasets by mounitng it on GDrive .We can then use the pretrain files to pretrain the command.Finetuning the files can increase the accuracy but the result is insignificant as compared to a computational power. One should use a Kaggle environment for as it does not requires the dataset to be loaded again and again. Keep in mind to never change versions as in many cases it did not run.

1 3

Results Our method SatMAE+Group+IM achieves the decent accuracy all be it at at the cost of increased computational power, outperforming supervised training from scratch and ImageNet-initialized backbones . ImageNet initializations may be less useful than in fMoW-RGB given the larger distributional shift to multi-spectral input data. We also note the effectiveness of grouping channels over processing all bands only at the patch embedding level (i.e. SatMAE+Stack).