ESA-NASA International Workshop on AI Foundation Models for EO

orion lab

Investigating Redundancy in Remote Sensing Images and its Implications to Foundation Models

A. Papazafeiropoulos, N. Bountos, I. Papoutsis, Orion Lab, National Technical University of Athens & National Observatory of Athens
tpapazafeiropoulos@ntua.gr

This project has received funding from the European Union’s Horizon Europe research and innovation programme under Grant Agreement No 101130544

ThinkingEarth

Motivation

Example of Redundancy

Redundant information: the information which is not necessary for our task

mask 95

mask 85%

mask 75%

original

Kaiming, He, et al.“Masked Autoencoders Are Scalable Vision Learners’’ arXiv:2111.06377v3 (2021).

RSLab
Remote Sensing Laboratory
National Technical University of Athens

Sources of Redundancy in EO Imagery

EO imagery shows unique characteristics compared to natural images:

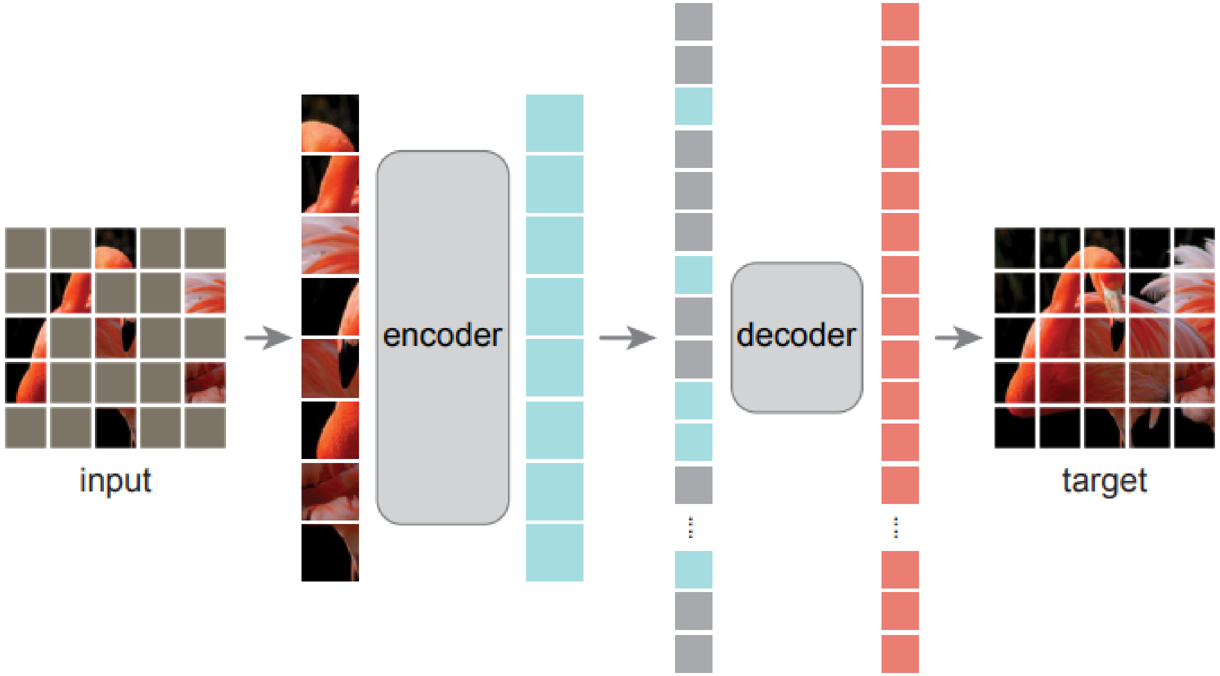
- Multi-scale scenes
- No clear “background”; every pixel contains information
- High spatio-temporal variability (seasons, climate, geography)

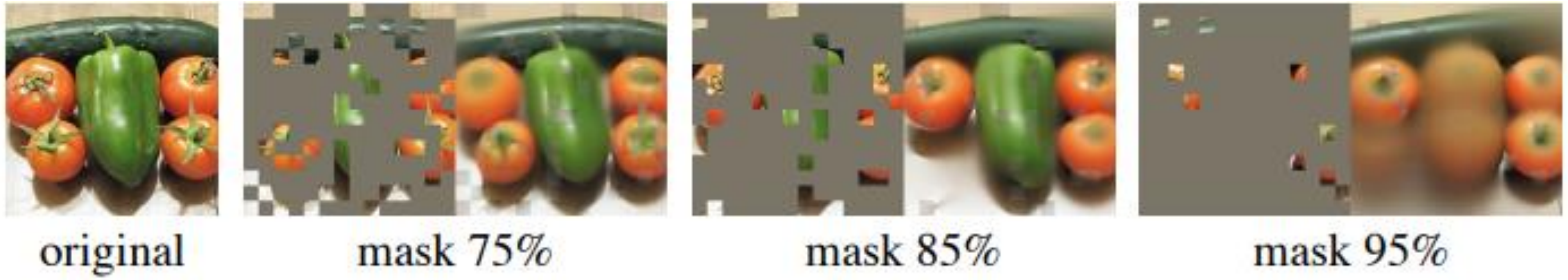
Why Remote Sensing Is So Redundant:

- Large Spatial Coverage
- Repetitive Patterns Over Space & Time
- Varying Spatial Resolution & Sensor Modalities

RSLab
Remote Sensing Laboratory
National Technical University of Athens

Motivation: MAE are scalable vision learners





Kaiming, He, et al.“Masked Autoencoders Are Scalable Vision Learners’’ arXiv:2111.06377v3 (2021).

Key hypothesis - Redundancy as a feature

Questions:

- Assuming the existence of Redundancy:
  - How can we exploit it, instead of ignoring it?
  - What happens when we eliminate redundancy?
  - How does this affect model performance, especially across different downstream tasks?

Our Proposition:

- EO imagery carries **inherent “redundancy”** that may be:
  - **Proven** to exist
  - **Masked out** to focus on information-rich regions

Key Hypothesis:

- By **quantifying and exploiting** this redundancy in EO imagery, we can understand the domain better and, therefore, unlock more robust and scalable EO foundation models.

Proposed Approach

Overview

Tasks:

- Multilabel Classification
- Image Segmentation

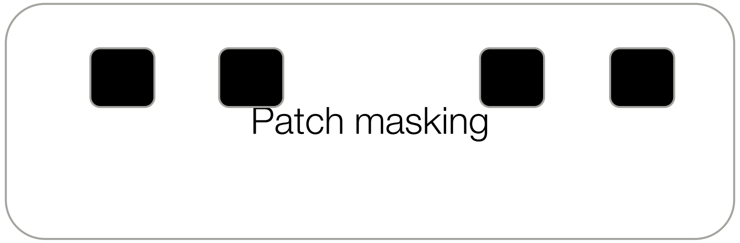
Backbones:

- RViT: Redundancy aware Vision Transformer for classification
- RUPerNet: UPerNet-style model adapted for redundancy-aware segmentation

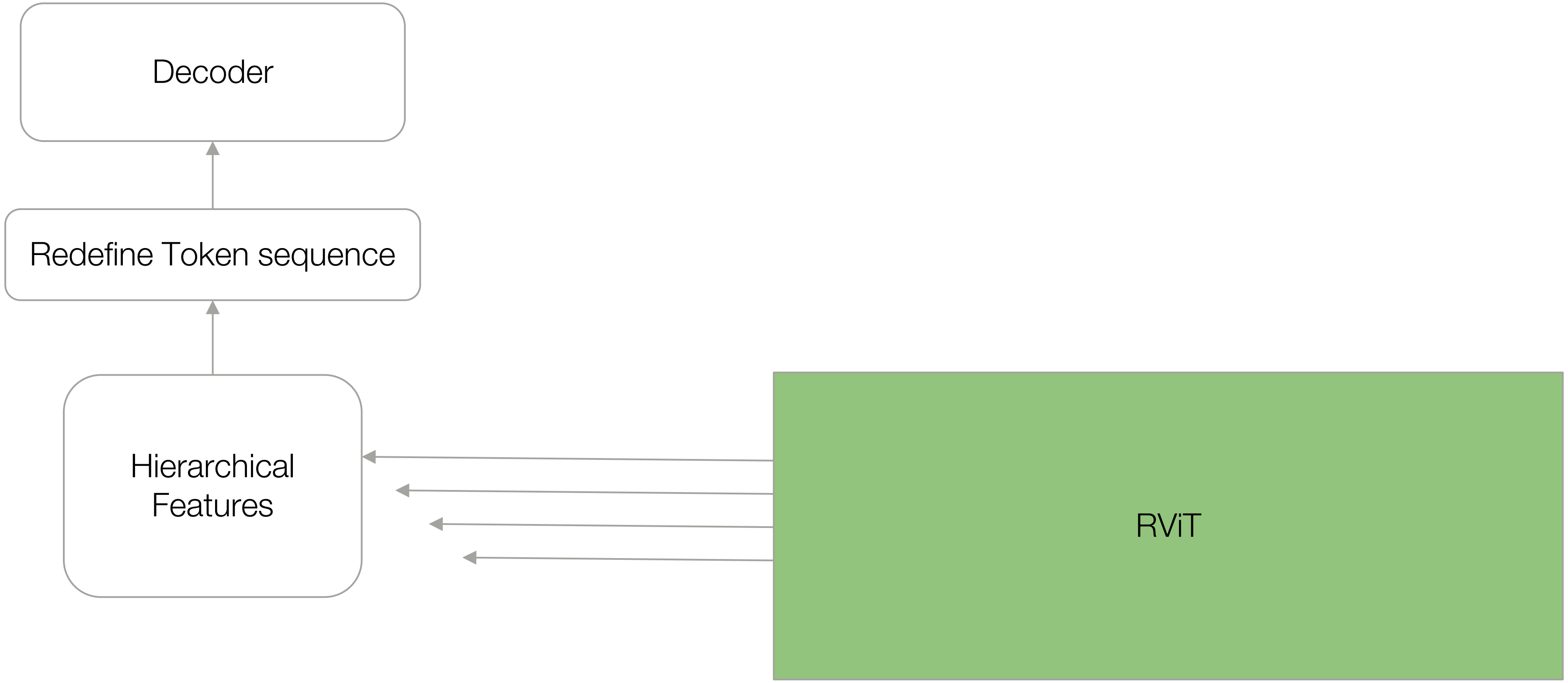
RViT - Redundancy aware ViT

- Train ViTs with the most information-rich patches → Remove “redundant” information via masking

- Masking strategies (per-sample):
  - Dynamic → cosine sim. Threshold
    - Varying num patches per sample
  - Static → Top-k% of the least similar patches
  - Random



RUPerNet



Tete, Xiao, et al. "Unified Perceptual Parsing for Scene Understanding", arXiv:1807.10221 (2018)

Evaluation Framework

Overview

Tasks:

- Multilabel Classification
- Image Segmentation

Backbones:

- RViT: Redundancy aware Vision Transformer for classification
- RUPerNet: UPerNet-style model adapted for redundancy-aware segmentation

Datasets and Metrics:

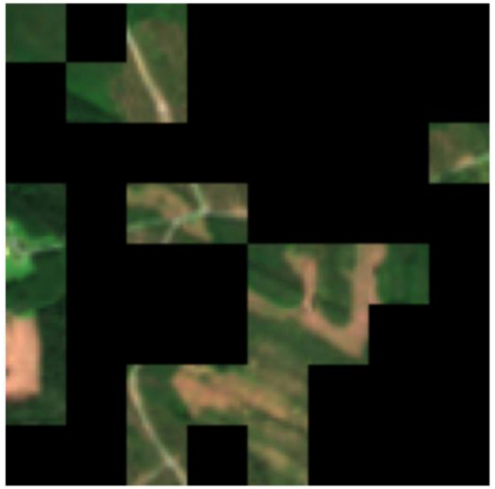
- BigEarthNet: macro Multilabel Average Precision, weighted Multilabel F1-score, macro Multilabel F1-score
- MLRSNet: micro Multilabel Average Precision, weighted Multilabel F1-score, micro Multilabel F1-score
- Woody: micro Multiclass F1-score, macro Multiclass F1-score, weighted Jaccard index
- Flair: micro Multiclass F1-score, macro Multiclass F1-score, weighted Jaccard index

Datasets

Dataset	Input Modality	Sensor	ML Problem	Num of Classes	EO task	Spatial Resolution	Image Size	Coverage
BigEarthNet	MS/SAR	S1, S2	Multi-label Classification	19	LULC Classification	10m	120x120	Europe
MLRSNet	RGB	Multi Sensor	Multi-label Classification	60	Semantic Scene Understanding	≈ 10-0.1m	256x256	Global
Woody	RGB	Aerial	Image Segmentation	4	Tree species detection	50cm	224x224	Chile
Flair	RGB/NIR/DEM	Aerial	Image Segmentation	19	LULC Classification	20cm	512x512	France

Examples

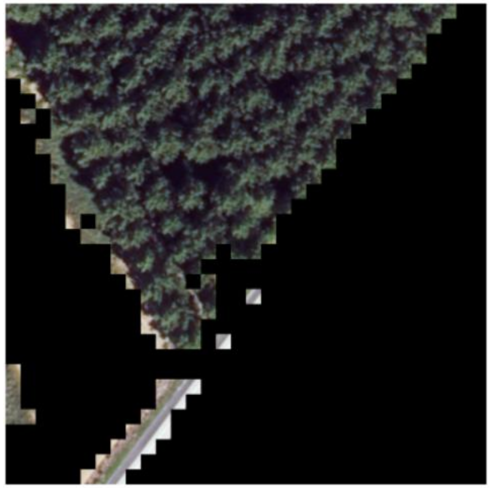
BigEarthNet



mask top65%

original Image

MLRSNet



mask top60%

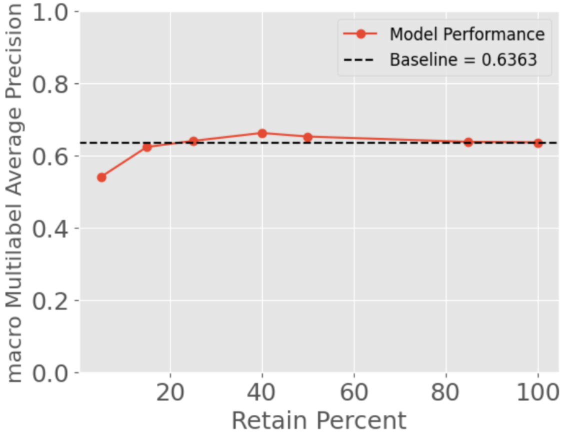
original Image

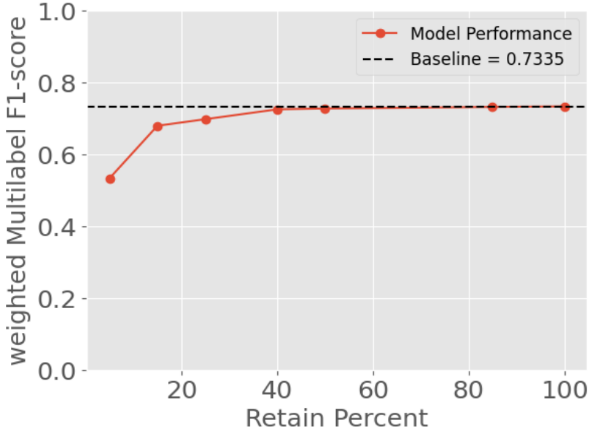
RSLab
Remote Sensing Laboratory
National Technical University of Athens

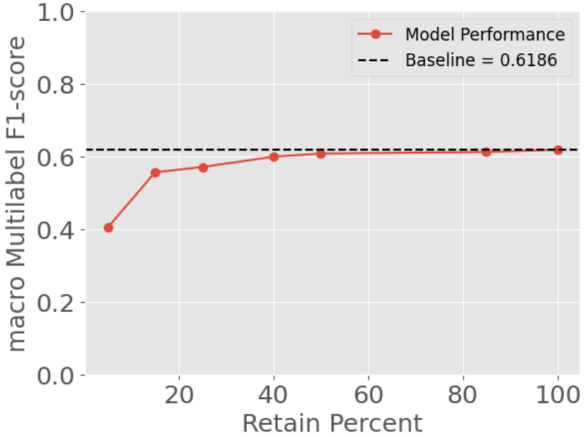
Preliminary Results

Multilabel Classification Results

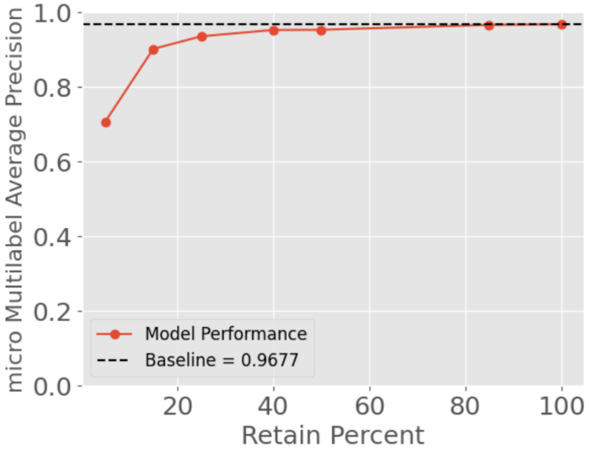
- BigEarthNet, top-k%, Vit-tiny architecture pretrained on ImageNet

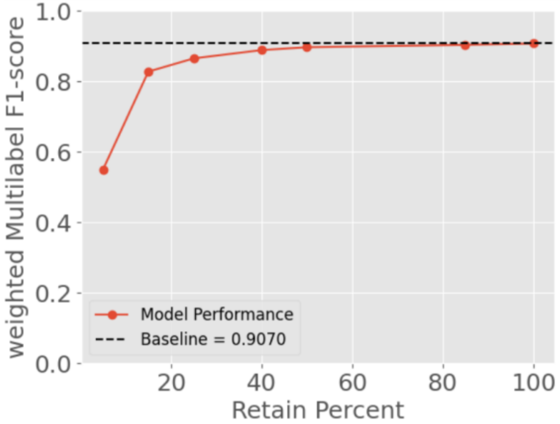






- MLRSNet, top-k%, Vit-tiny architecture pretrained on ImageNet





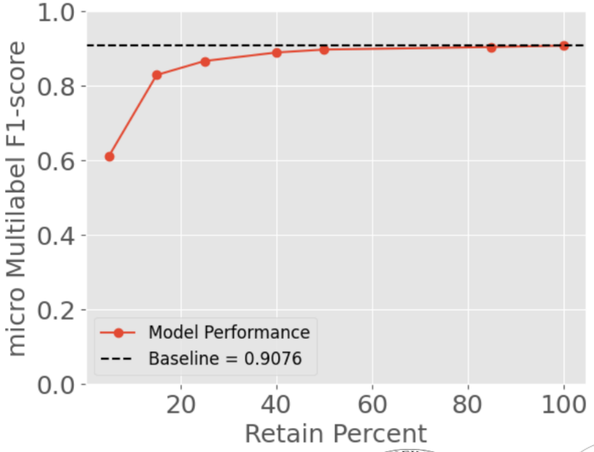
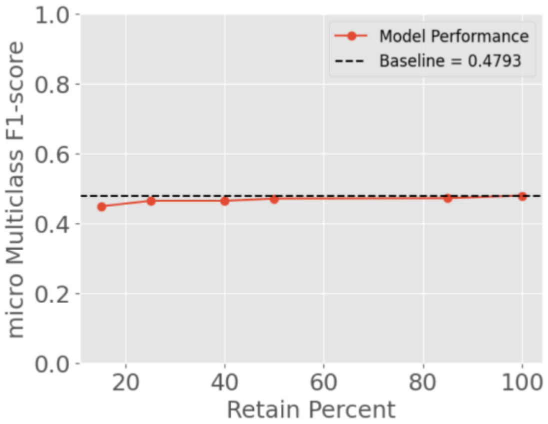
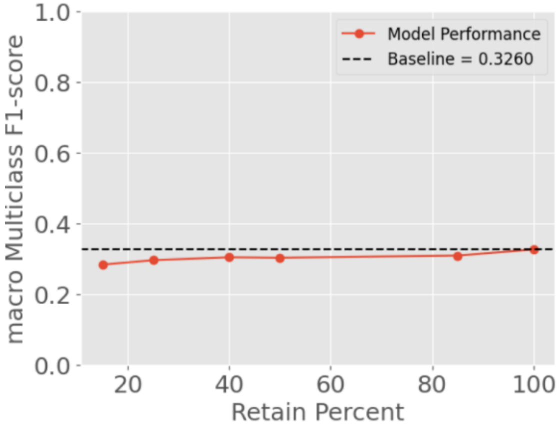
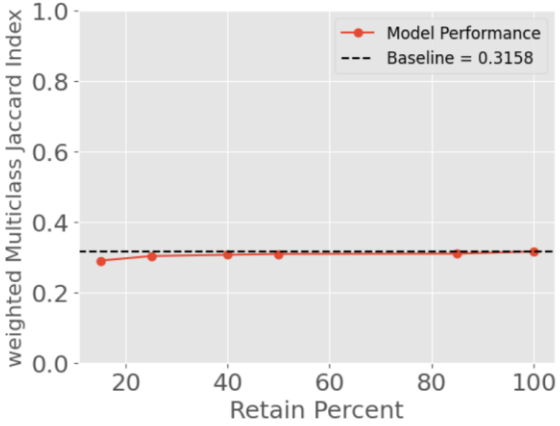


Image Segmentation Results

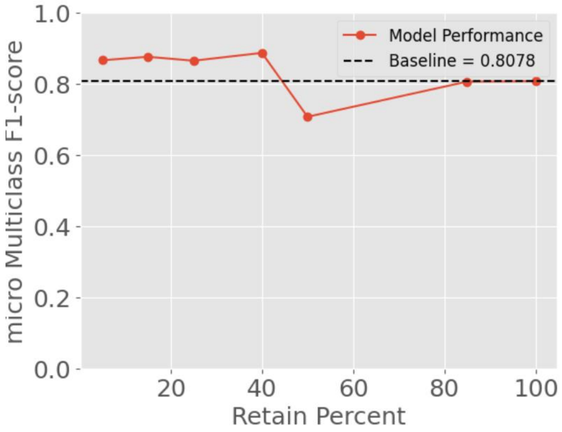
- Flair, top-k%, Vit-tiny architecture pretrained on ImageNet

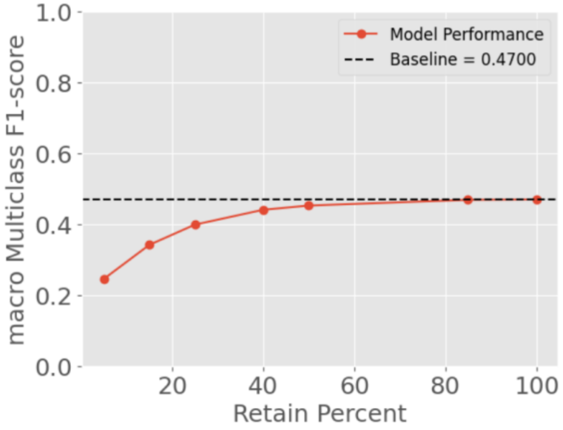


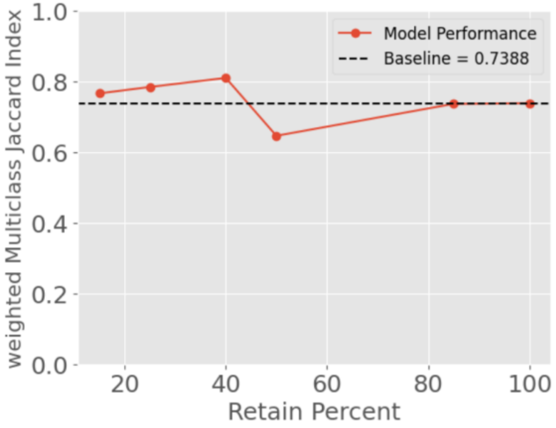




- Woody, top-k%, Vit-tiny architecture pretrained on ImageNet







Implications & Future Steps

Implications to Foundation Models

Unlocking Efficiency without Sacrificing Performance:

- Robust Performance under Heavy Masking
  Minimal precision, accuracy and F1-score loss up to 60% masking in all tasks and in some cases up to 85% masking.

- Path to Efficient Large-Scale Models
  Pruning redundant patches at the sample level significantly reduces compute and memory needs, making it feasible to train much larger transformers on EO data.

- Patch-Level Focus for Smarter Learning
  Concentrating on the most informative regions push models toward learning richer, more generalizable features, emphasizing quality of input over quantity.

Future steps

What’s next?

- Deepen Analysis of Preliminary Findings
  - Develop insight-driven masking strategies and validate across additional EO datasets
- Fine-Tune Segmentation Tasks
  - Perform hyperparameter optimization on RUPerNet
- Quantify Efficiency Gains
  - Systematically report memory footprint and training speed-up under varying mask ratios
- Explore Research Questions:
  - Can we achieve comparable downstream performance when pretraining on smaller, information-rich subsets of EO data?
  - How does per-sample redundancy correlate with generalization across tasks, modalities, and geographies?

RSLab
Remote Sensing Laboratory
National Technical University of Athens

Q&A