Exploring Transfer Learning Using Segment Anything Model in Optical Remote Sensing

Image Classification and Semantic Segmentation Case Studies

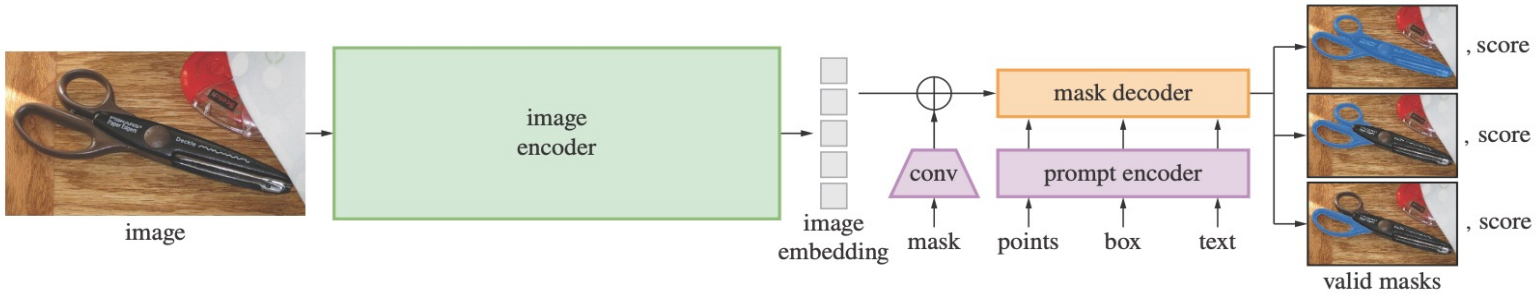
Mohanad Albughdadi, Vasileios Baousis, Tolga Kaprol, Armagan Karatosun, and Claudio Pisa

European Center for Medium Range Weather Forecasts (Bonn, Reading)

Firstname.lastname@ecmwf.int

ECMWF

Segment Anything Model (SAM)



- It is a foundation model for prompt image segmentation.
- It enables powerful generalization in zero-shot and few-shots learning.
- It consists of three main components:
  - Image encoder: a masked autoencoder (MAE); Scalability and strong pretraining; Relies on ViT architecture.
  - Prompt encoder: encodes prompts (points or bounding boxes) using their positional encoding.
  - Light-weight mask decoder: maps prompt embeddings and image embeddings with the output mask; learns class tokens; Upsamples the mask using transposed convolutions.

ECMWF [1] A. Kirillov, E. Mintun *et al.*, “Segment anything,” *arXiv e-prints*, pp. arXiv–2304, 2023.

Out of the Box Integration of SAM with Remote Sensing Images

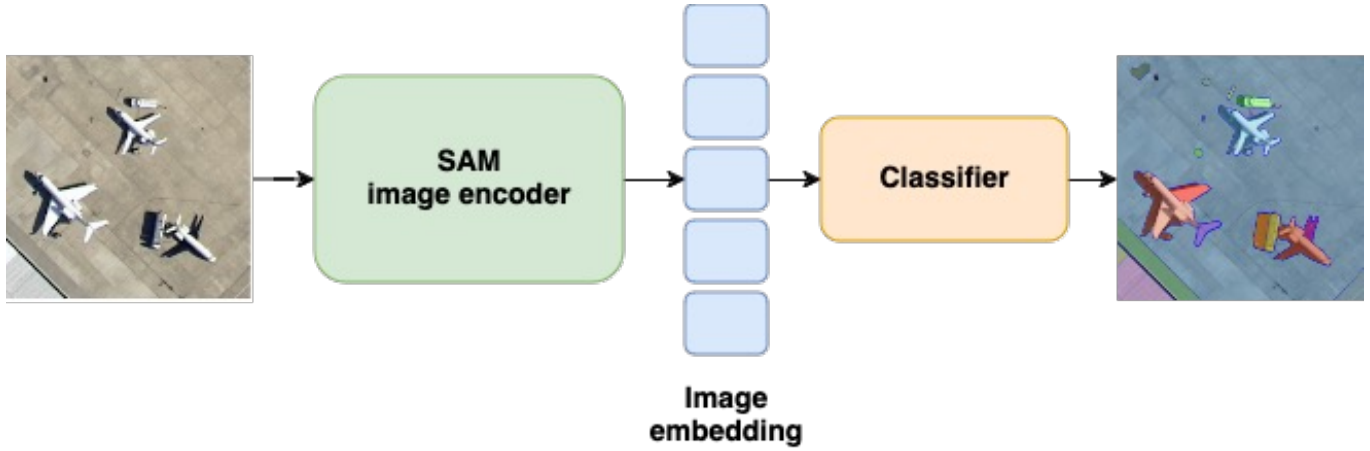
- Prompt segmentation using spatial prompts:
  - Sentinel-1
  - VHR aerial images
  - Sentinel-2

SAM Transfer Learning

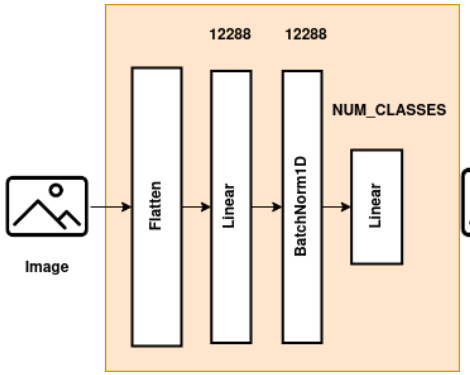
- Transfer learning is an ML technique to transfer the knowledge from the source domain to the target domain, which relaxes the assumption of the training dataset and test dataset must be i.i.d, it also allow training models with fewer amounts of training data.

- The encoder component from the SAM model is used to learn useful image embeddings.

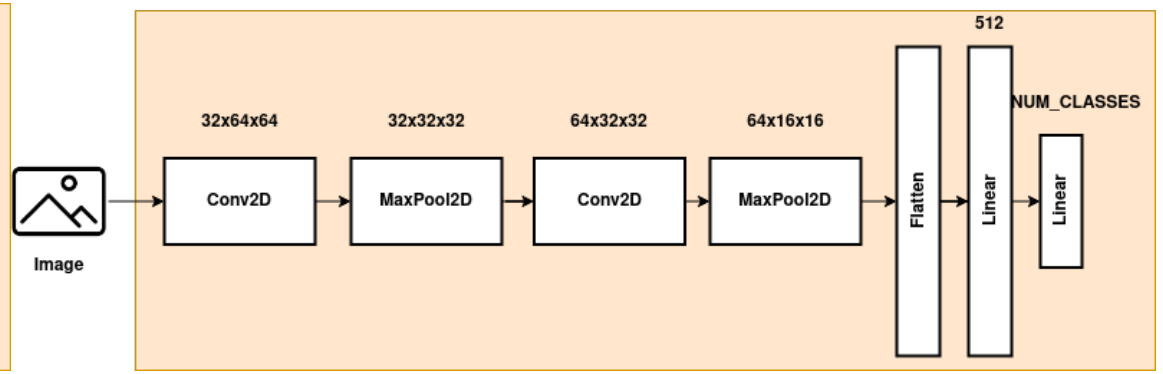
- A classifier is then trained on top of the frozen layers of SAM encoder.



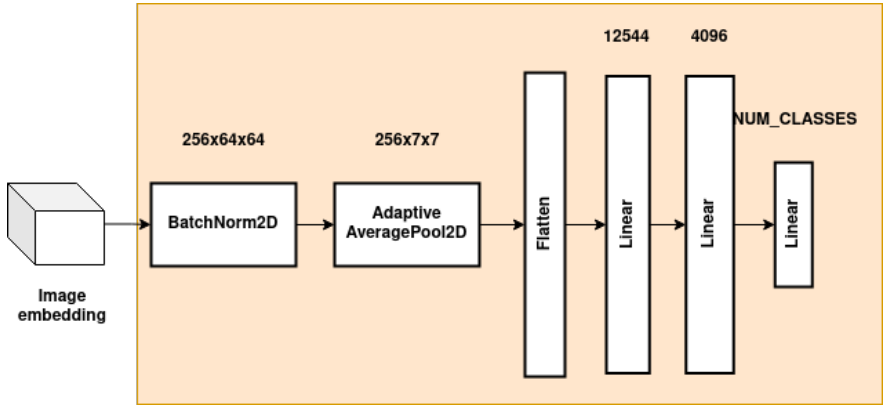
Case Study #1: Image Classification (Models)



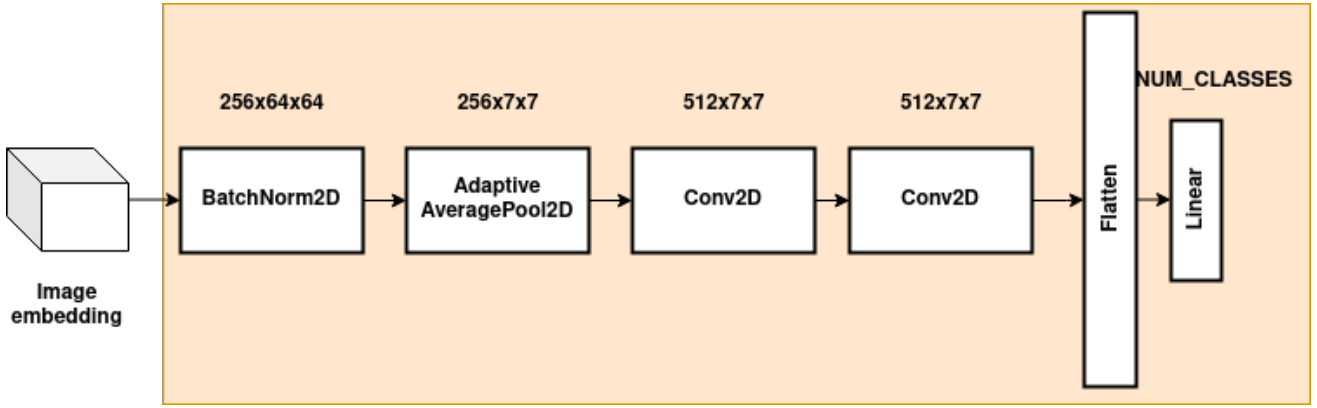
Linear



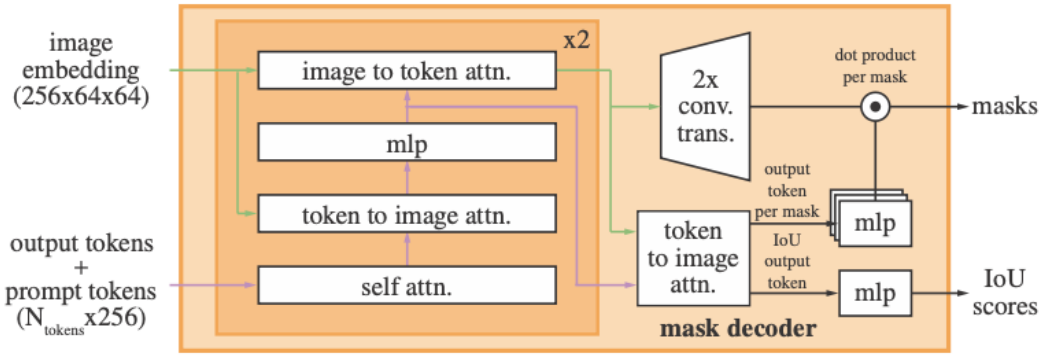
Convolutional



SAM-LC



SAM-CONV



SAM-ViT [1]

[1] A. Kirillov, E. Mintun *et al.*, “Segment anything,” *arXiv e-prints*, pp. arXiv–2304, 2023.

Case Study #1: Image Classification (Datasets)

- EuroSAT [2]:
  - 27000 labeled Sentinel-2 patches
  - HR10m
  - 64x64 in the visible part of the spectrum (RGB)
  - Each patch is distributed into one of 10 distinct classes of land cover (multiclass)

Class	Samples	Class	Samples
Annual Crop	3000	Highway	2500
Forest	3000	Industrial	2500
Herbaceous Vegetation	3000	Pasture	2500
Permanent Crop	2500	Residential	3000
River	2500	Sea Lake	3000

[2] P. Helber, B. Bischke *et al.*, “Introducing Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification,” in *IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2018, pp. 204–207.

Case Study #1: Image Classification (Experiments - EuroSAT)

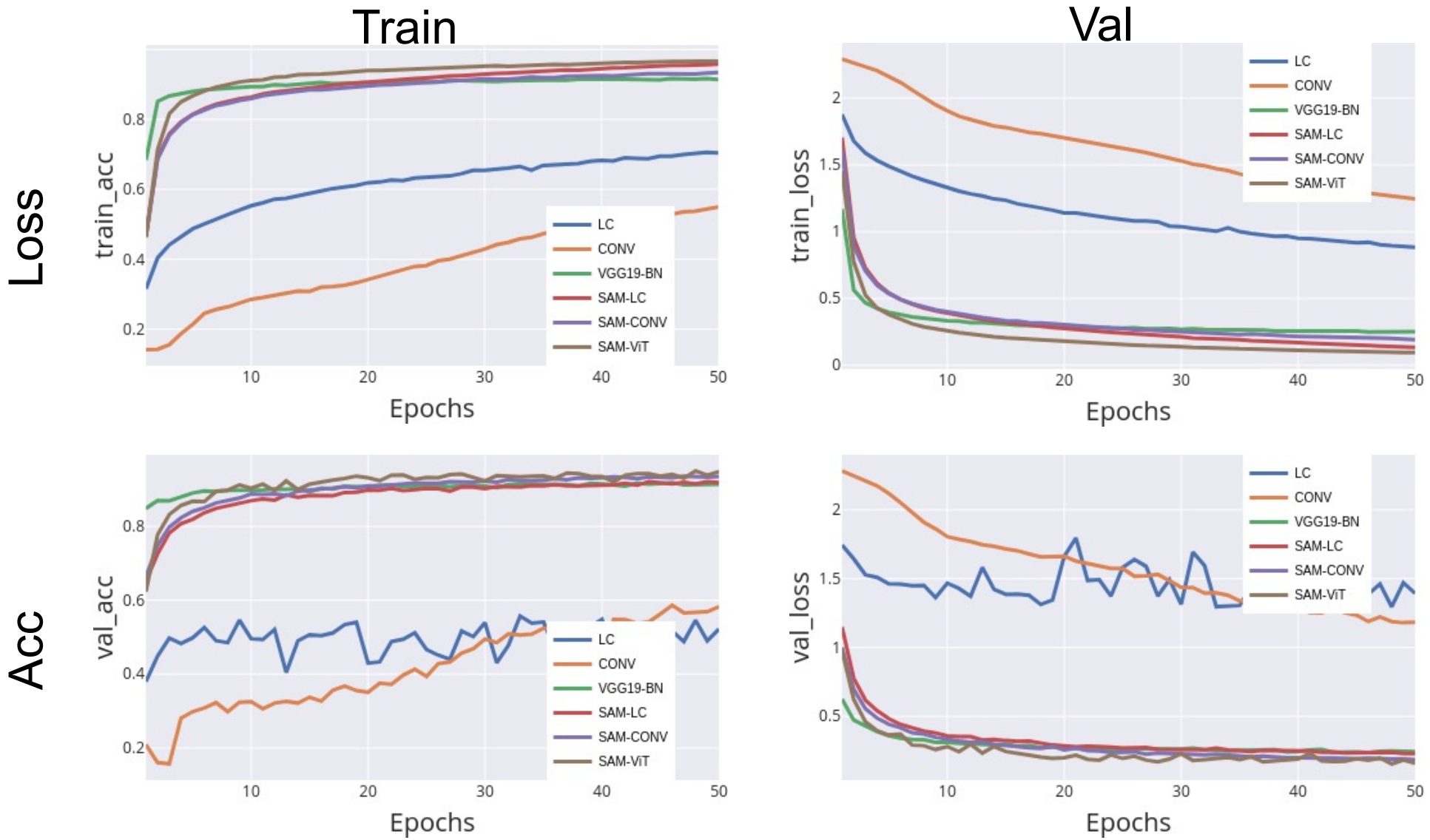


Figure: Train and val loss and accuracy using a train-val split of 80-20.

Case Study #1: Image Classification (Experiments - EuroSAT)

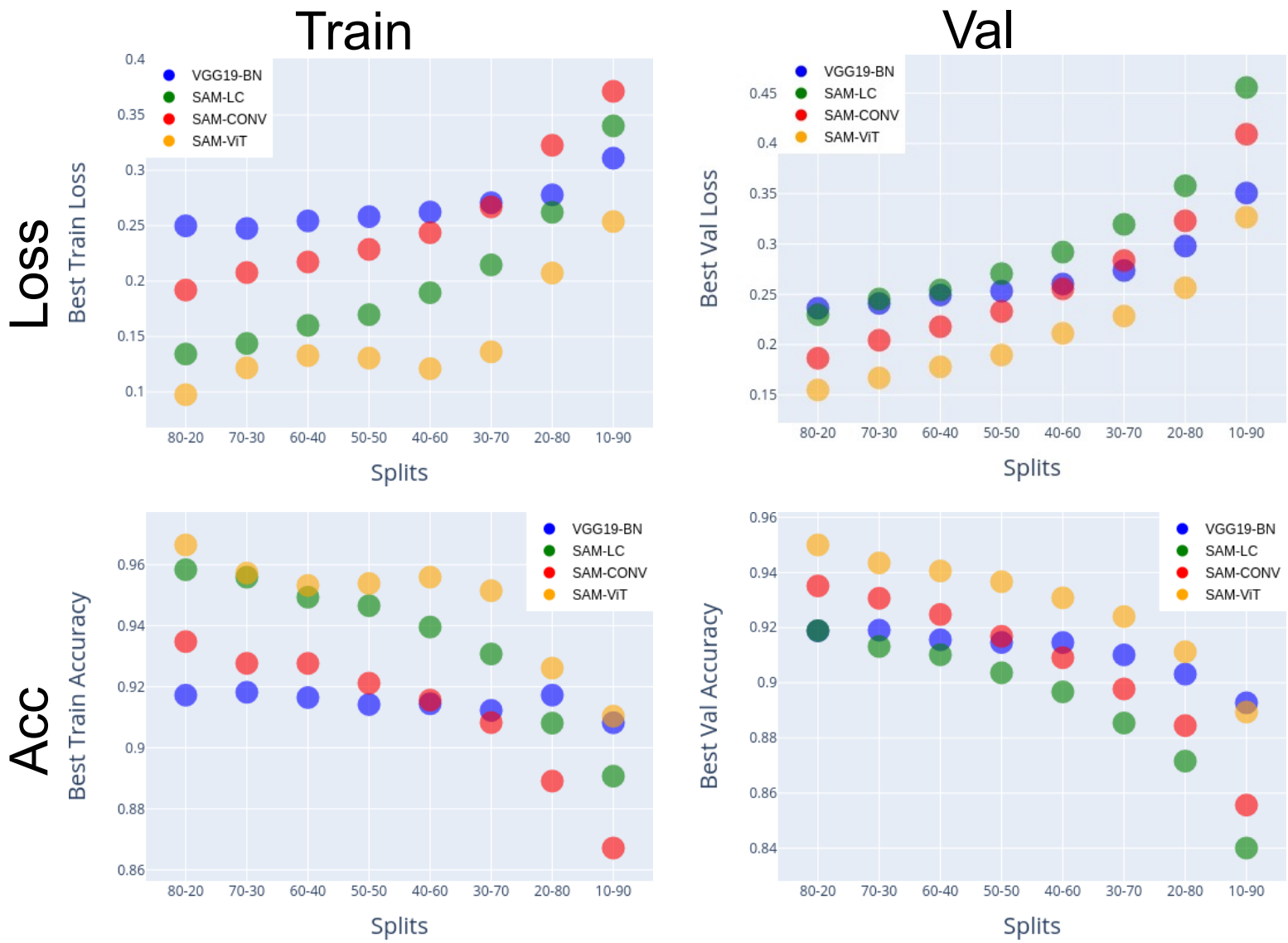
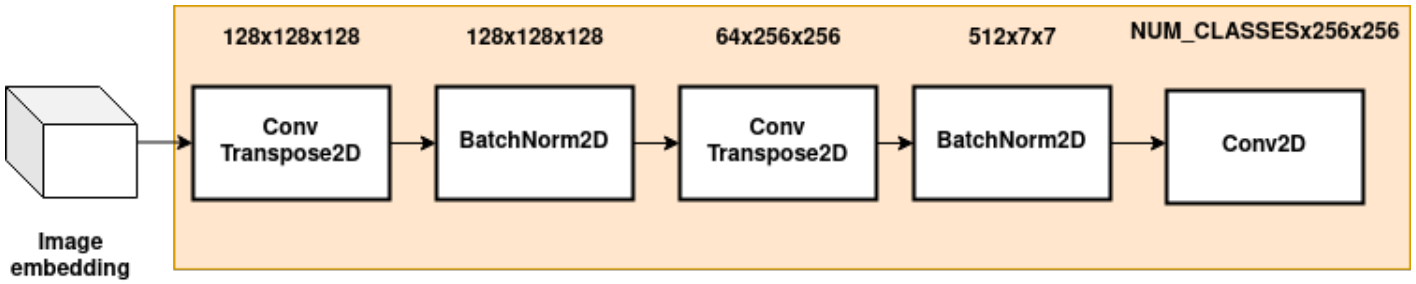
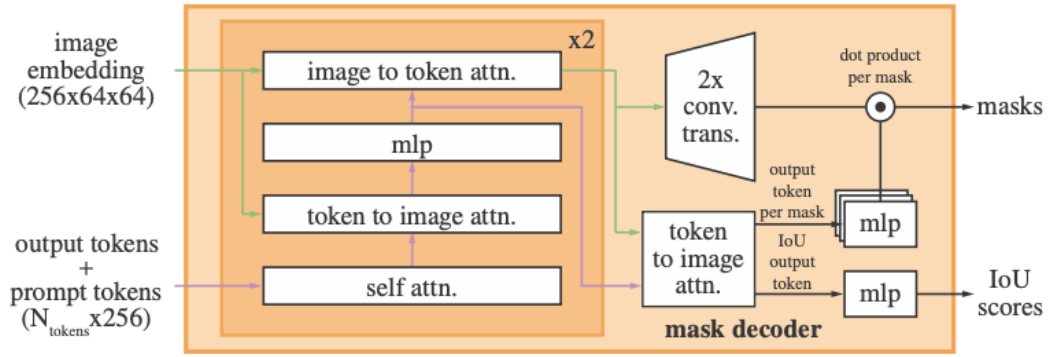


Figure: Performance of the competing models for different train-validation splits of the EurotSAT dataset.

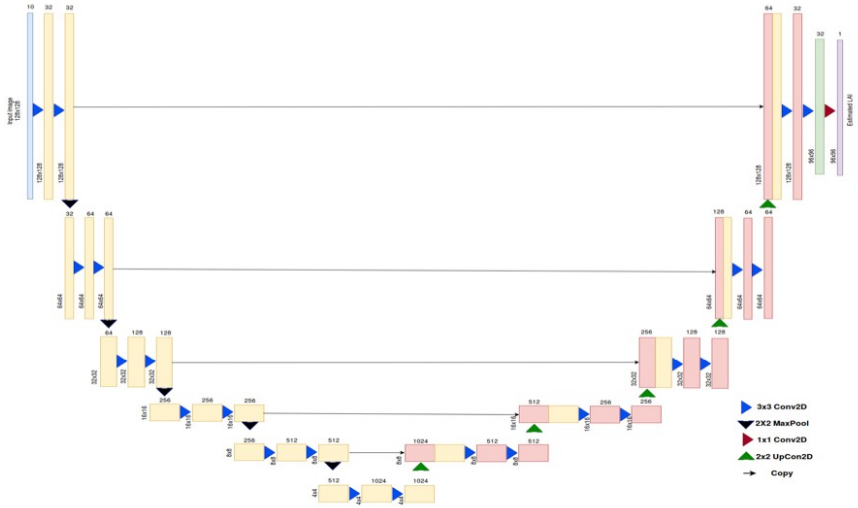
Case Study #2: Semantic Segmentation (Models)



SAM-CONV



SAM-ViT



UNet

Case Study #2: Semantic Segmentation (Dataset)

- The Road dataset [3] was used to explore SAM transfer learning for semantic segmentation.
- This dataset contains 224 RGB images.
- Spatial resolution is 1.2 m.
- The dataset is divided into training, validation and testing sets of sizes 160, 20, and 44, respectively.

Figure: Examples of the Road dataset along with the corresponding segmentation masks

[3] G. Cheng, Y. Wang *et al.*, "Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 6, pp. 3322– 3337, 2017.

Case Study #2: Semantic Segmentation (Experiments)

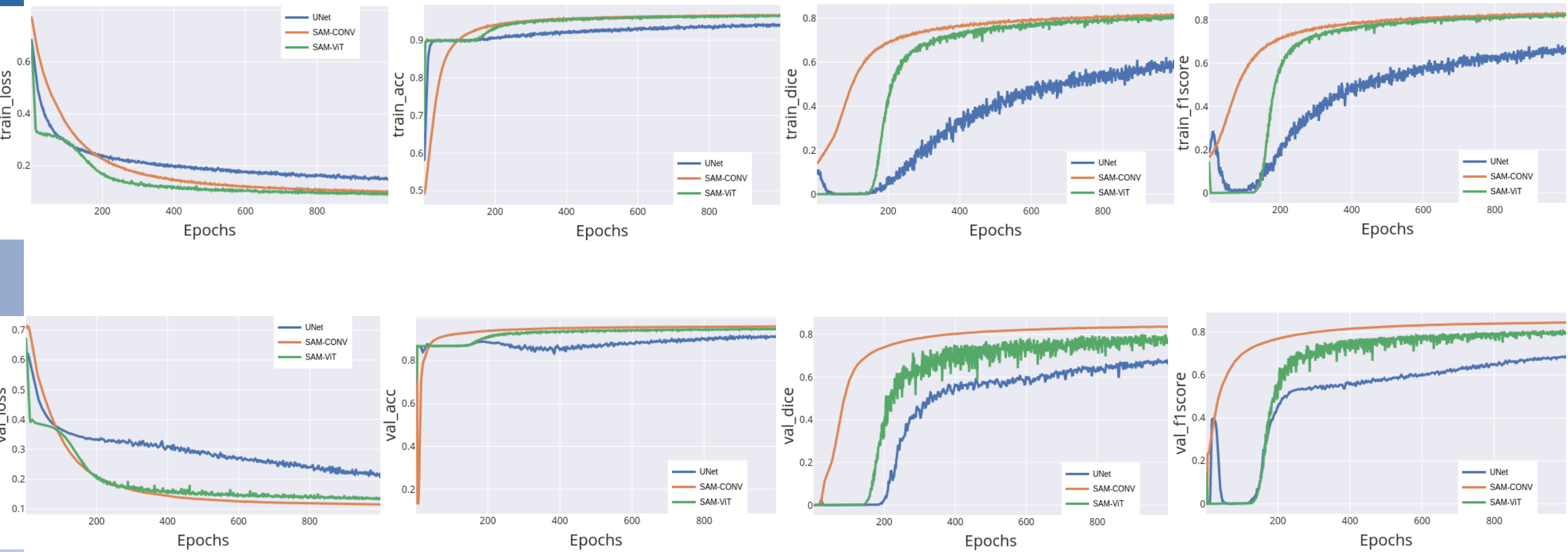


Figure: Training and validation curves of the competing algorithms on the Road dataset.

Case Study #2: Semantic Segmentation (Experiments)

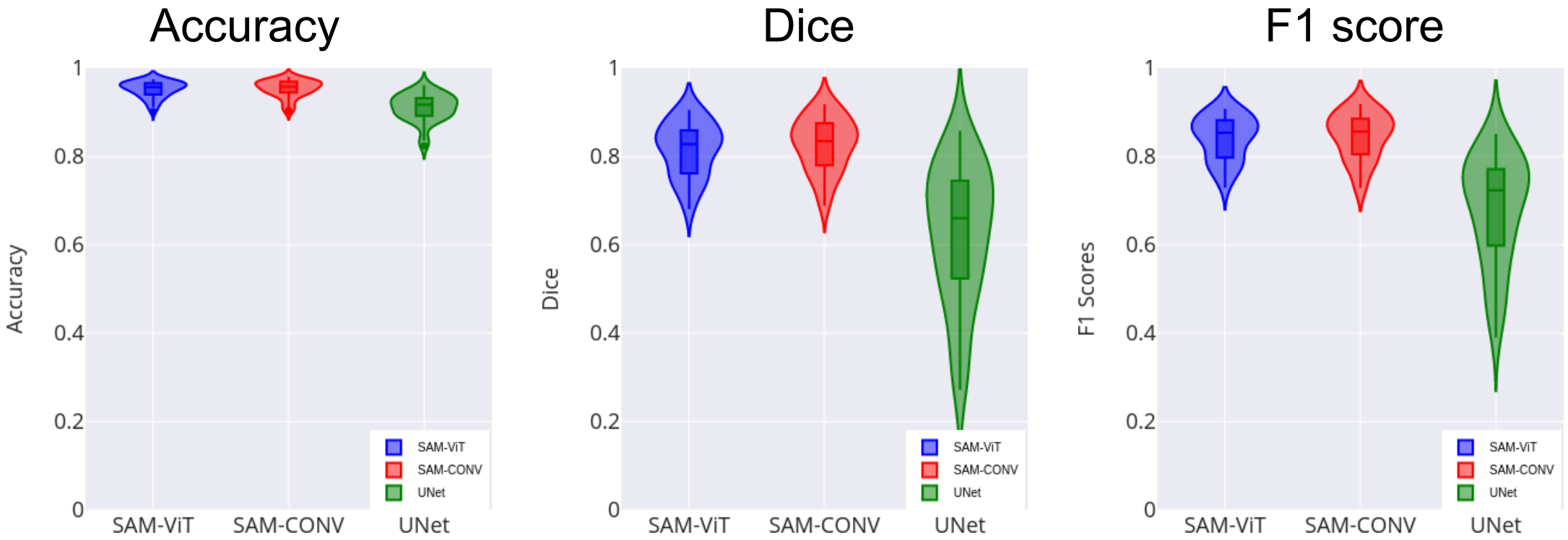


Figure: Accuracy, Dice and F1-score distribution for the semantic segmentation task using the Road test dataset.

Case Study #2: Semantic Segmentation (Experiments)

Figure: Segmentation examples from the Road test dataset.

Conclusion

- This study evaluated transfer learning using SAM for remote sensing tasks (image classification and semantic segmentation).
- For image classification, the SAM based models were compared to baseline models and transfer learning from the VGG19 model.
- For semantic segmentation, SAM-based models were compared to training a UNet model from scratch.
- The study highlights SAM's efficacy in transfer learning, with SAM-ViT excelling in image classification and SAM-CONV leading in semantic segmentation
- Future research to focus on comparing SAM's transfer learning with other computer vision and geospatial foundation models such as DINO and IBM NASA Geospatial and assessing its efficiency in different tasks.

Thank you for your attention!

Funded by the European Union

Work funded by the EU Horizon Europe grant 101060784 (EO4EU).

EO4EU
Earth Observation Data & Service Accessibility

https://www.eo4eu.eu