Report for the course Machine Learning for Signal Processing

Erik Deinzer
Mat. Nr. 2222615

05 May 2025

Contents

1 Introduction

The goal of this project was to apply patch classification on the EuroSAT Dataset. For this application, a small Framework for classification-pipelines was build and is explained in the following. The framework aims to be easily extendable to other tasks such as detection, segmentation and other usecases.

The developed framework follows a config-based approach - the complete training, testing and validation pipeline can be set up by using configuration dicts. This allows easy hyperparameter tuning and model development. Detailed explaination of the usability of the framework are found in the *README.md*.

1.1 EuroSAT Dataset

The EuroSAT dataset[1] is an collection of Sentinel-2 satellite images designed for land-use and land-cover classification. It comprises 27,000 labeled patches (64×64 pixels) covering ten classes such as annual crops, forests, highways, and residential areas. Each image patch contains thirteen spectral bands (including RGB, near-infrared and shortwave infrared), allowing models to learn from both visual and multispectral information. The tiles are $640m \times 640m$ in size. In the development of the project, the focus shifted to the use of RGB-features only since they already achieve very good results.

1.2 ImageNet-Tiny

ImageNet-Tiny[2] is a reduced version of the original ImageNet dataset, commonly used to speed up experiments while retaining the core characteristics of large-scale pretraining. In this project, ImageNet-Tiny refers to a subset of ImageNet where each class contains a limited number of samples (500 per class) across a smaller number of total classes (200 classes), and all images are downscaled (to 64×64).
The goal of using ImageNet-Tiny is to enable faster training and evaluation of models under constrained compute settings, while still leveraging the representational richness of ImageNet. Despite its smaller size, pretraining on ImageNet-Tiny can still provide useful feature representations for downstream tasks such as land cover classification. In our case, it serves as a lightweight proxy for full ImageNet pretraining, allowing us to assess the transferability of general visual features to the EuroSAT domain. The reason it was used was the size of the dataset, it only uses 2 GB instead of 1.3 TB of hard disk memory.

2 Methods

The task was to develop a truncated ResNet-18 architecture with equal or less than 1.5M parameters and to further process the Resnets outputs for patch classification.

2.1 ResNets

Residual **Net**works are a classic neural network architecture used in deep-learning scenarios. Their key innovation is the *residual block*, which incorporates a shortcut (identity) connection that skips one or more convolutional layers. Instead of directly learning an underlying mapping $H(x)$, a residual block learns the *residual function*

$$F(x) \;=\; H(x) - x,$$

so that the block's output is

$$y \;=\; F(x) + x.$$

This formulation makes it easier for gradients to flow backward through many layers, since the identity shortcut provides an unobstructed path for gradient propagation.
Since state-of-the-art solutions in Computer Vision almost always use ResNets as a backbone to a classification head, this idea was also applied in this project. A Feed Forward Network (a Network of fully connected layers)

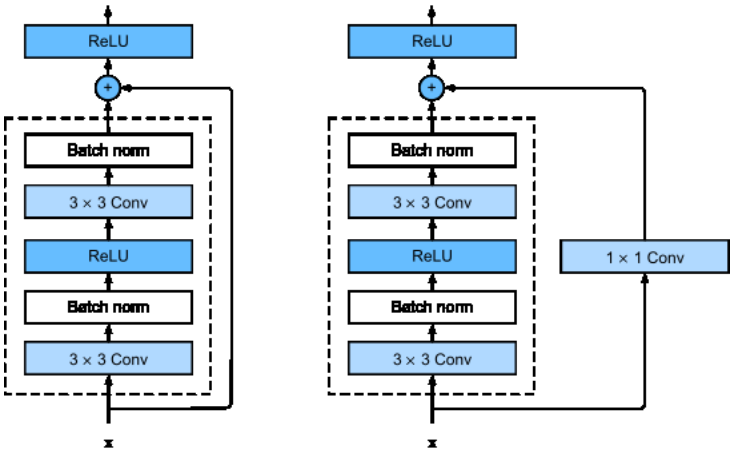


Figure 1: ResNet BasicBlocks [3]

2.2 Feed Forward Networks

The feed forward network is the basic type of neural network architecture as it is solely a connection of linear layers and in our case $ReLU$ activation functions.

- **Input layer:** Receives the input feature vector of dimension f_{in} and expands to an output dimension of $f_{\text{hidden\_dims}}$
- **Hidden layers:** An alternating connection of linear layers with in-and output dimension $f_{\text{hidden\_dims}}$ and a activation function of ones choosing (in our case ReLU).
- **Output layer:** Receives the last hidden layers feature vector of dimension $f_{\text{hidden\_dims}}$ and expands to an output dimension of $f_{\text{out\_dims}}$, in our case $f_{\text{out\_dims}} = N_{\text{classes}} = 10$

2.3 Models

With this, the connected models add up to the following diagram:
The Configurations are . Those configurations are applied on a 128×128 feature map, which demanded

Table 1: Model Configuration Summary

Name	**idims**	**odims**	**bd**	**rn**$_{odims}$	**hd**	**Pretrained**	**Dropout**	**N**$_{bb}$	**N**$_{h}$
EuroSAT	3	10	12	64	1024	False	0.2	1.58M	4.27M
ImageNet	3	200	12	64	1024	False	0.2	1.58M	4.28M
EuroSAT*	3	10	12	64	1024	ImageNet	0.2	1.58M	4.27M

a resize before application. Early stopping at 0.05 was applied on the validation loss (absolute). The stopping configuration can be easily set and controlled, in particular whether it should be applied to validation loss, mAP or F1 score. The exact configurations can be checked in the repository.

3 Results

Table 2: Model Performance Comparison

Name	**mAP**	**F1 Score**
EuroSAT	0.983	0.941
ImageNet	0.281	0.272
EuroSAT*	**0.986**	**0.953**

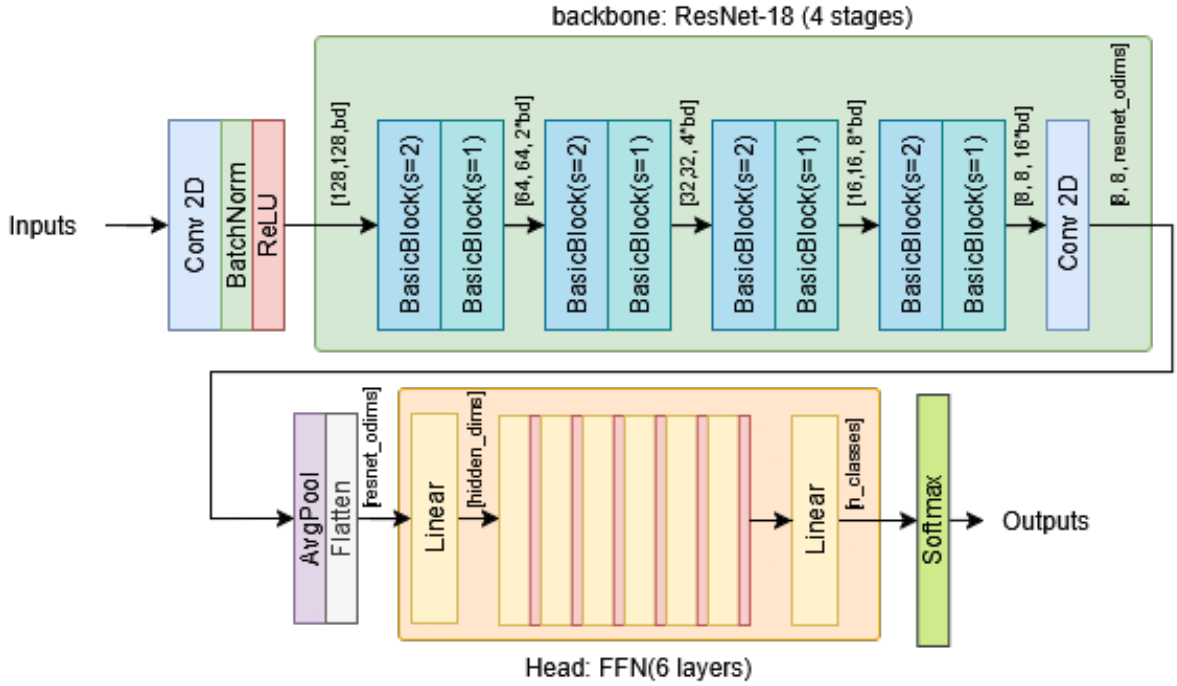
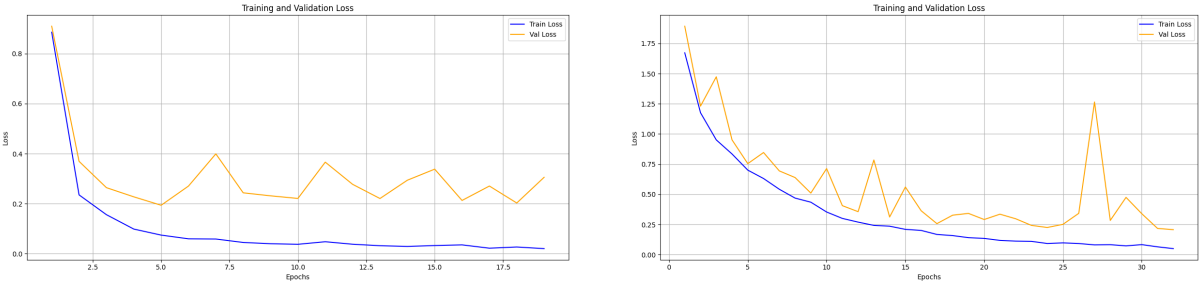


Figure 2: Model architecture

As shown in Table 2, models trained directly on EuroSAT data achieve very high performance, with a mean average precision (mAP) of 0.983 and an F1 score of 0.941. The EuroSAT* model, which leverages a backbone pretrained on ImageNet, shows slightly improved metrics: a mAP of 0.986 and F1 score of 0.953. These results suggest that while the base model already performs strongly, transfer learning can still contribute marginal but meaningful gains.
Interestingly, the model trained solely on ImageNet performs poorly on this domain-specific task, with a mAP of only 0.281 and an F1 score of 0.272. This highlights the importance of domain relevance in feature representations—ImageNet pretraining alone does not yield satisfactory performance for remote sensing classification without fine-tuning. Furthermore, this may be an indication of a too small backbone for classification in a highly multiclass setting (200 classes in ImageNet Tiny in comparison to 10 in EuroSAT).
In addition to better final metrics, the EuroSAT* model benefits from faster and more stable convergence, as shown in Figure 3. It reaches near-optimal performance in fewer epochs and exhibits smoother loss curves, indicating more reliable gradient updates and a less noisy training process.

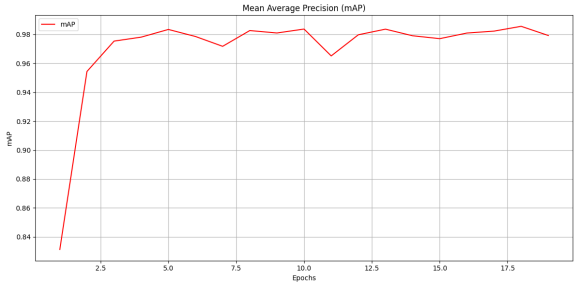


((a)) EuroSAT* Training/Validation Loss

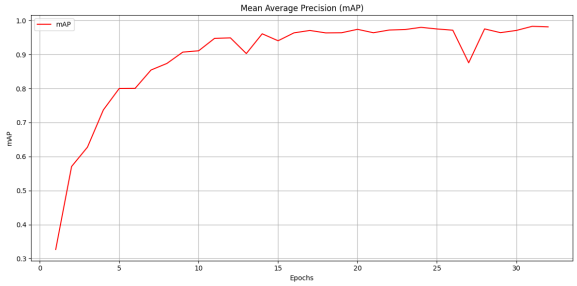
((b)) EuroSAT Training/Validation Loss

Figure 3: Comparison of training and validation losses.

Figure 4 further supports this, showing the progression of mAP scores throughout training. The EuroSAT* model consistently outperforms the baseline across epochs, suggesting that the pretrained backbone provides more discriminative features early in training.
Similarly, Figure 5 shows the F1 score progression, with the pretrained model not only achieving a higher



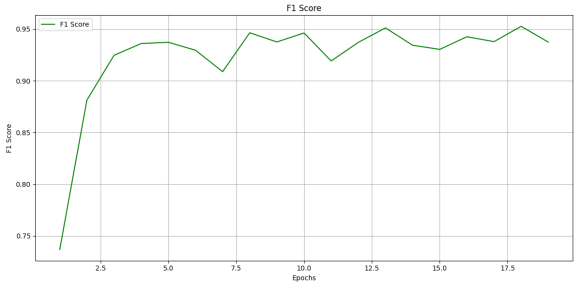
((a)) EuroSAT* mAP



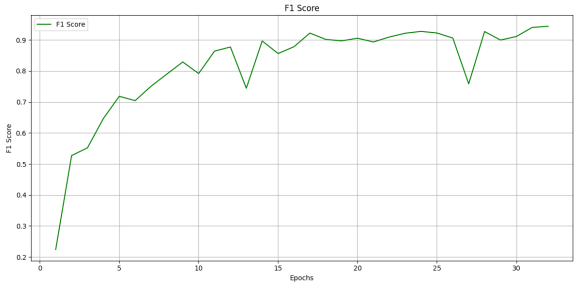
((b)) EuroSAT mAP

Figure 4: Mean Average Precision (mAP) over training epochs.

final score but also demonstrating more consistent learning behavior. This stability is particularly valuable in settings with limited compute resources, as it can reduce the need for extensive hyperparameter tuning.



((a)) EuroSAT* F1 Score



((b)) EuroSAT F1 Score

Figure 5: F1 Score progression during training.

Overall, these results demonstrate that while the EuroSAT dataset enables very high accuracy even from scratch, incorporating pretrained features—particularly those trained on large-scale datasets like ImageNet—can still enhance both performance and training efficiency. The benefits, although negliglible in final metrics, are especially evident in the early phases of training and in the consistency of learning dynamics.
The final results lead to an almost unit-matrix style confusion matrix Interestingly, this Confusion Matrix was achieved after 5 epochs of training, showing a fast learning progress.
Tiles of Annual Crop (class 0) tend to be detected as either Annual Crop (86% of cases), but are sometimes confused with Permanent Crop (class 6) or Pasture (class 5) tiles. This is probably caused by similar layouts of crop fields and the geological similarity as well as the availability of transitional land covers in the dataset - land stock is often located near to crop fields.

4 Discussion

The results of this study demonstrate that high performance can be achieved on the EuroSAT dataset even when training a model from scratch, with a mean average precision (mAP) of 0.983 and an F1 score of 0.941. However, incorporating a backbone pretrained on ImageNet (EuroSAT*) led to measurable improvements in both accuracy and training dynamics, pushing mAP to 0.986 and F1 to 0.953. Although these gains may appear modest, they are significant given the already strong baseline and the low overall error margin.
The benefits of transfer learning were especially noticeable in terms of convergence speed and training stability. The pretrained model required fewer epochs to reach optimal performance and exhibited smoother loss and metric curves. This indicates that leveraging generic visual features learned from large-scale natural image datasets like ImageNet can accelerate learning, even when applied to the remote sensing domain.

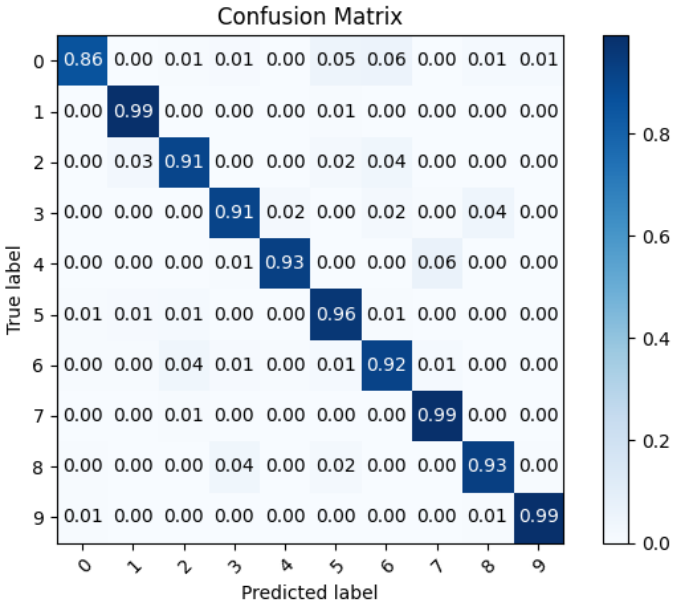


Figure 6: Confusion matrix for EuroSAT*

The model sometimes misclassifies between visually similar classes such as *Permanent Crop* and *Annual Crop*, due to their shared texture and color characteristics. Second, classification errors also occur in tiles with mixed or transitional land cover, where the label represents the dominant class but the image contains substantial portions of other categories. These observations highlight the limitations of fixed-size patch classification and the challenges of label ambiguity in remote sensing data.
Moreover, the performance disparity between the model trained solely on ImageNet labels and those fine-tuned on EuroSAT underscores the importance of domain-specific training. Features learned from natural scenes do not readily generalize to satellite imagery without adaptation, reinforcing the need for targeted datasets and task-specific fine-tuning.
In future work, incorporating richer spectral information (e.g., full Sentinel-2 bands), applying multi-label or segmentation approaches, and using context-aware architectures like transformers could help address current failure cases and further improve classification robustness. Using the transformer architecture would allow stronger performance on ImageNet and probably even more accurate final results.

References

[1] Apollo2506. Eurosat dataset (rgb). `https://www.kaggle.com/datasets/apollo2506/eurosat-dataset`, 2023. Accessed: 2025-06-15.

[2] Umang Jain. Tiny imagenet-200 dataset. `https://www.kaggle.com/datasets/umangjjw/tinyimagenet200`, 2022. Accessed: 2025-06-15.

[3] Aston Zhang, Zachary C Lipton, Mu Li, and Alexander J Smola. *Dive into Deep Learning*. Cambridge University Press, 2020. `https://d2l.ai/chapter_convolutional-modern/resnet.html`.