

...

Transfer Learning for CNN and Transformers architectures

Jakub Sawicki, Katsiaryna Bokhan, Milanna
Pahasian, Rafał Pyzowski

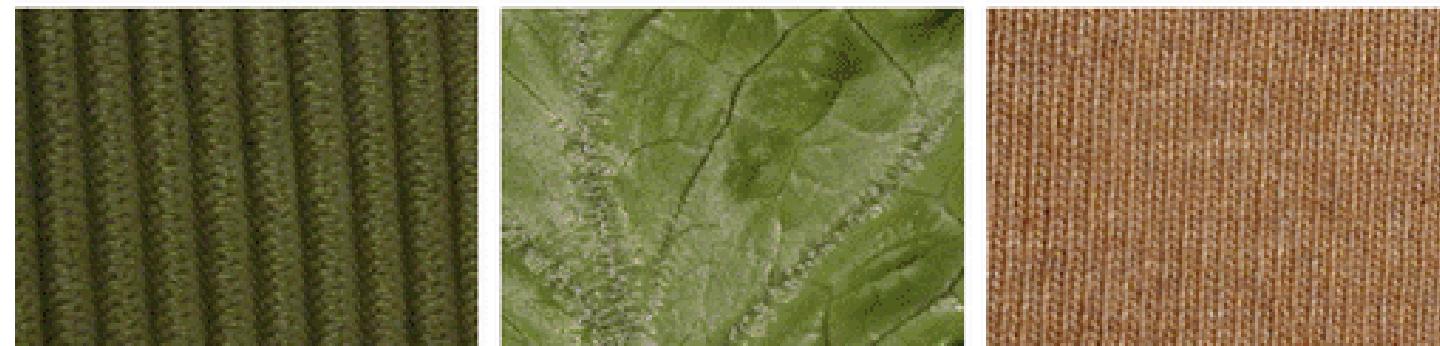
Task

- Train CNN, Transformers models on DTD dataset
- Use transfer learning to train these models on RSSCN7 dataset
- Find CNN, Transformers models trained on ImageNet dataset
- Use transfer learning to train them on RSSCN7 dataset
- Compare results
- XAI

...

Data

DTD: 5640 texture images, 47 categories



ImageNet: 14,197,122 annotated images

RSSCN7: 2800 satellite images, 7 categories



CNN

**University at Buffalo
School of Engineering and Applied Sciences**

**CSE 676 - Deep Learning
Spring 2023**

We found the published work on CNN based image classification for DTD dataset. It has some basics of CNN architecture and a simple CNN model. The accuracy of this model on a validation set for DTD is around 40% and its performance on RSSCN7 data is poor. We tried to replace given CNN with LeNet-5.

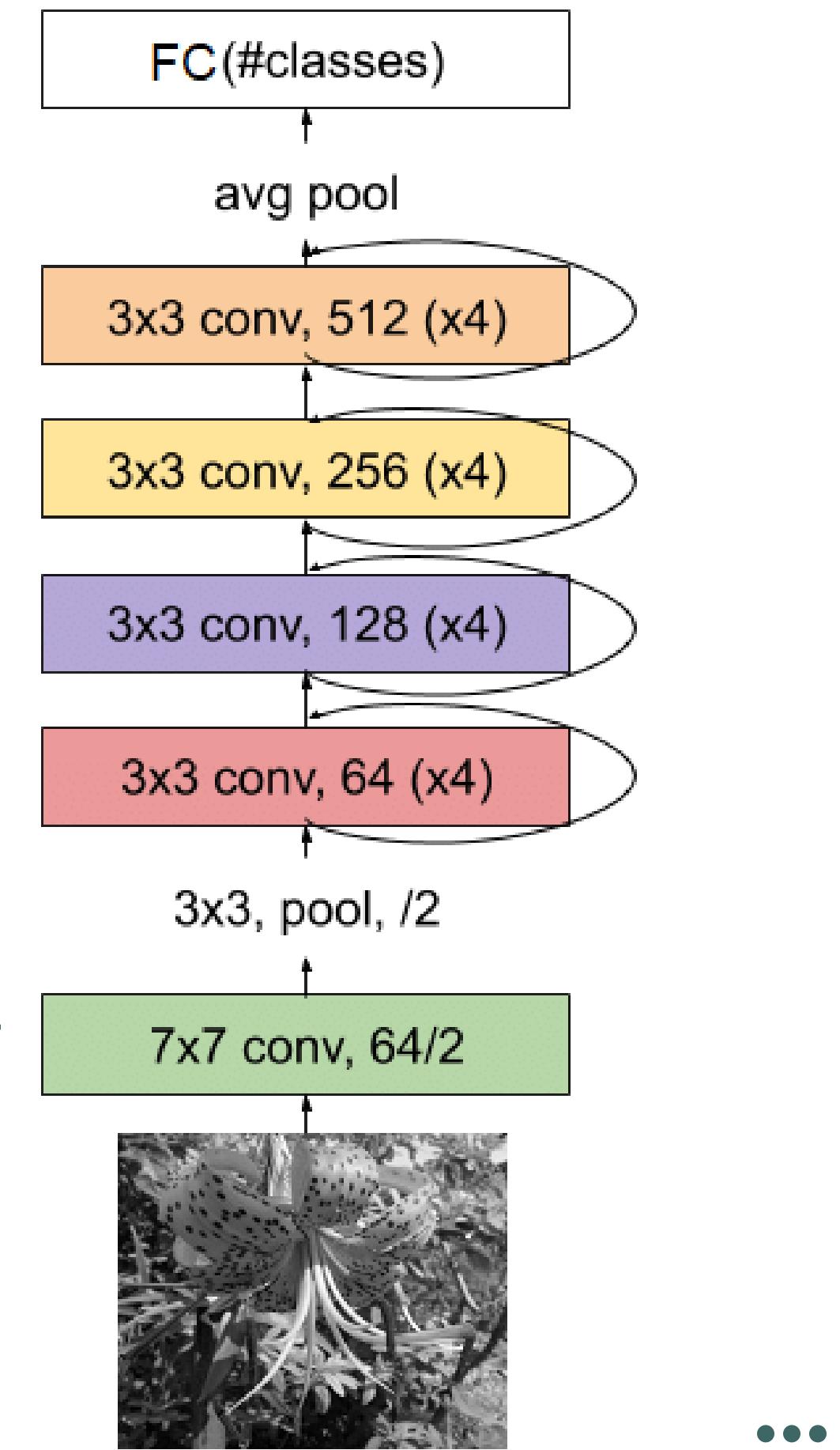
...

ResNet 18

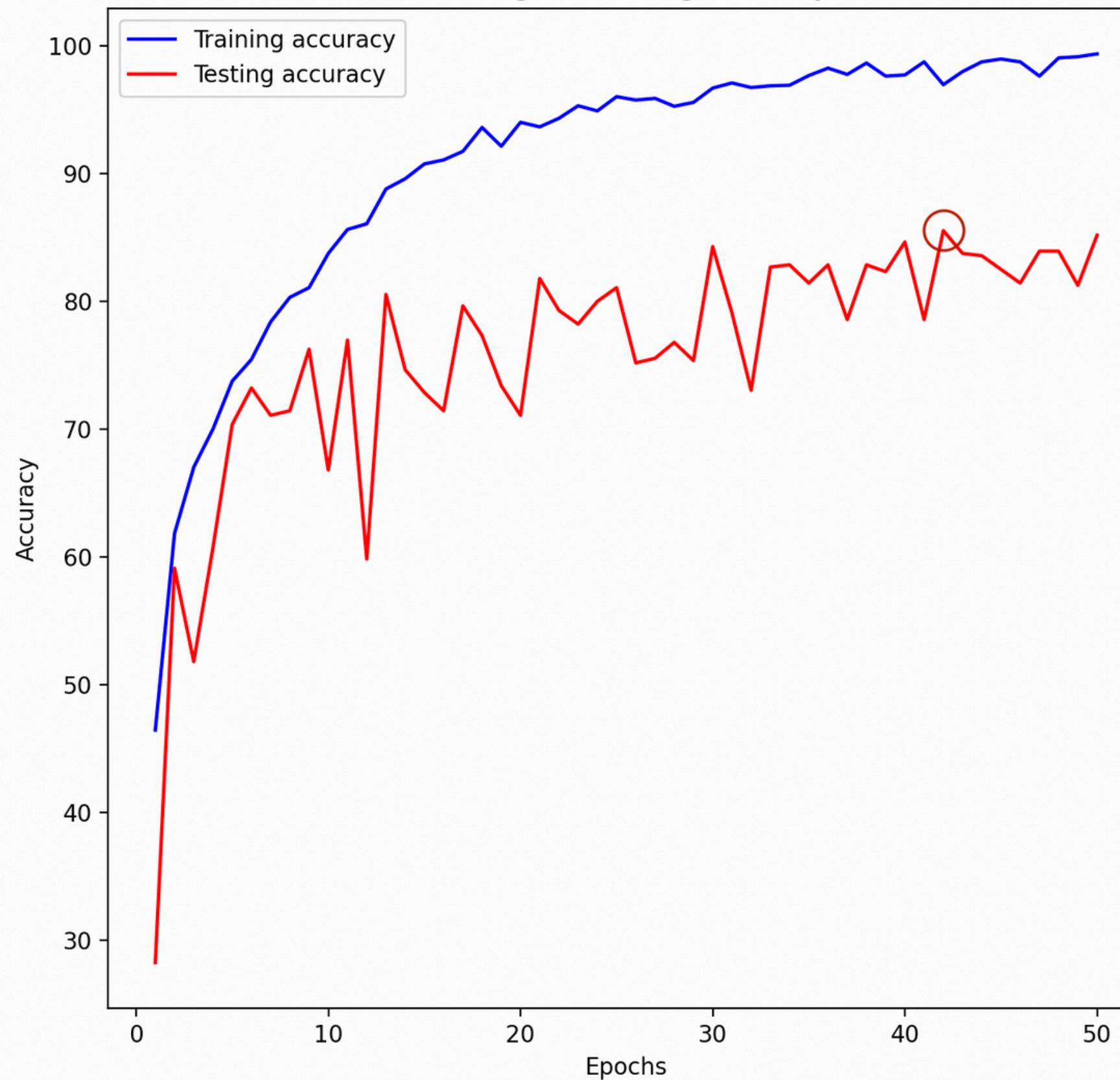
We decided to use ResNet18 architecture with predictions vector output.

The problem with images size was solved by cropping them. Additionally data was transformed and normalized. Data augmentation used.

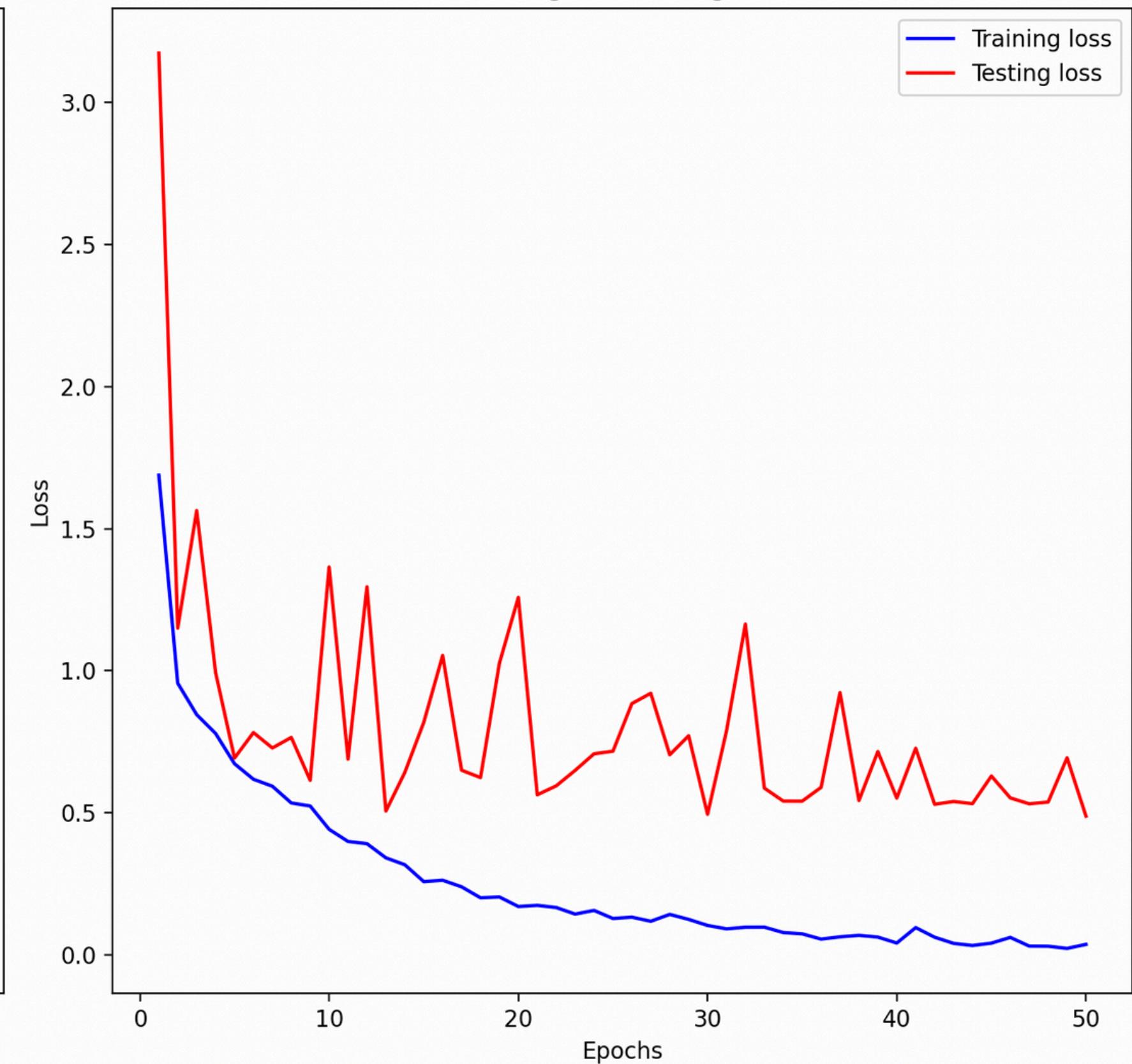
The model was trained on DTD for 90 epochs(Accuracy 52%). We used transfer learning for RSSCN7 dataset. The model was trained for an optimal number of epochs - 42 with results:
Loss: 0.5297
'Accuracy': 85.5357%



Training and testing accuracy



Training and testing loss



...

ResNet18 (ImageNet)

We used pre-trained on ImageNet ResNet18 model to compare results.

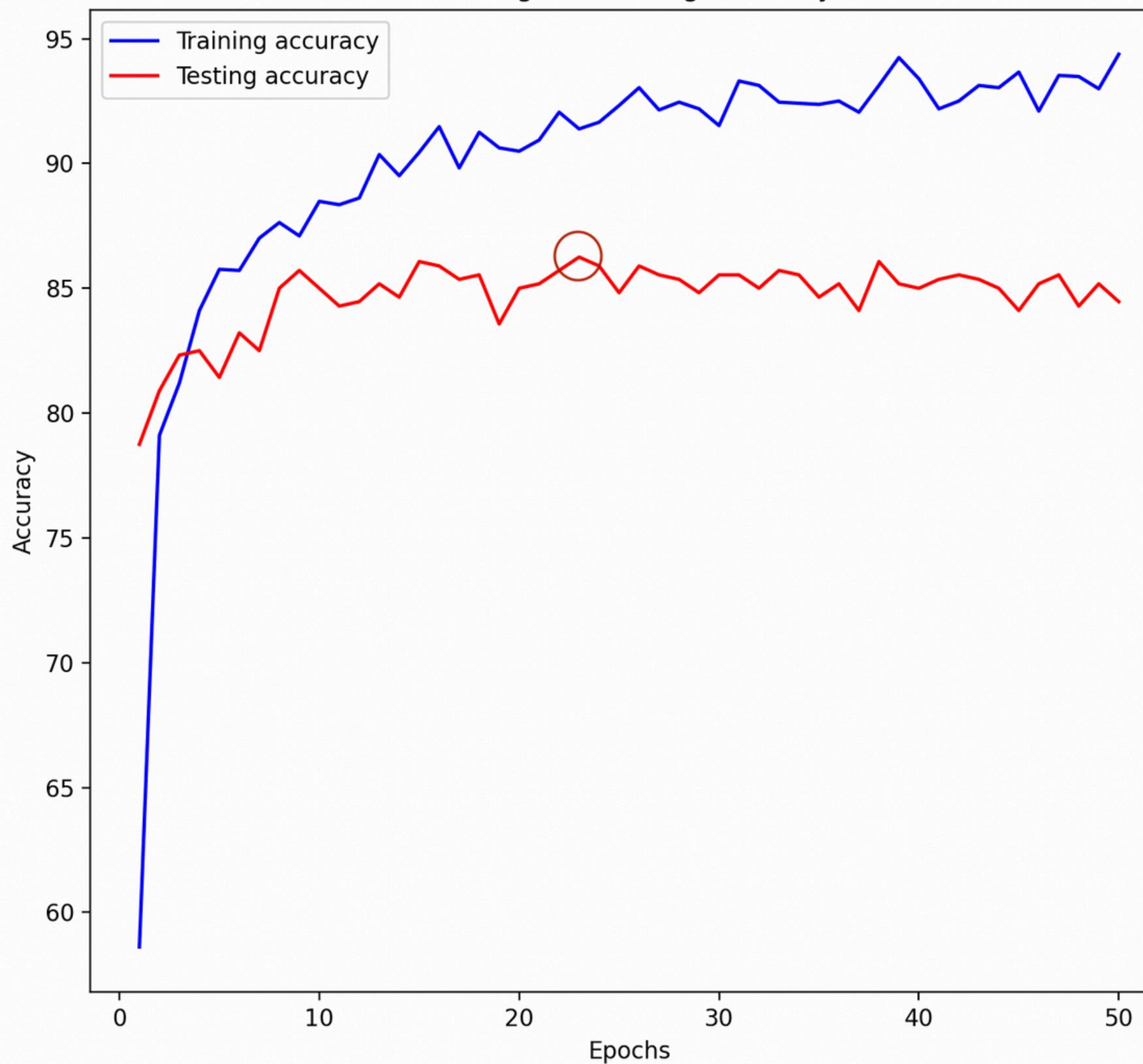
ResNet18 accuracy for Imagnet: Top 1 69.754%, Top 5 89.088%

The transfer learning model was trained for an optimal number of epochs - 23 with results:

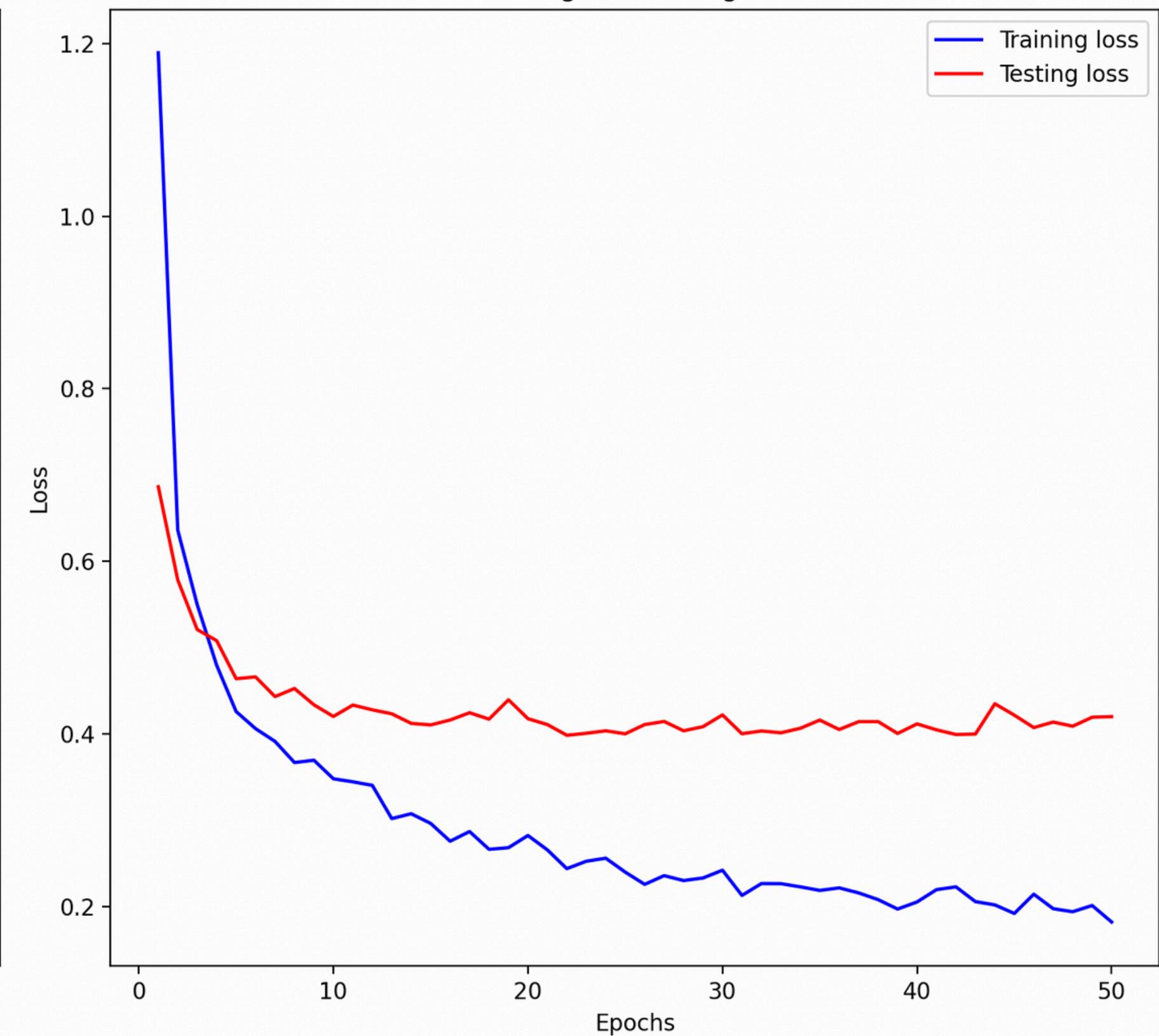
Loss: 0.4009

'Accuracy': 86.2500%

Training and testing accuracy



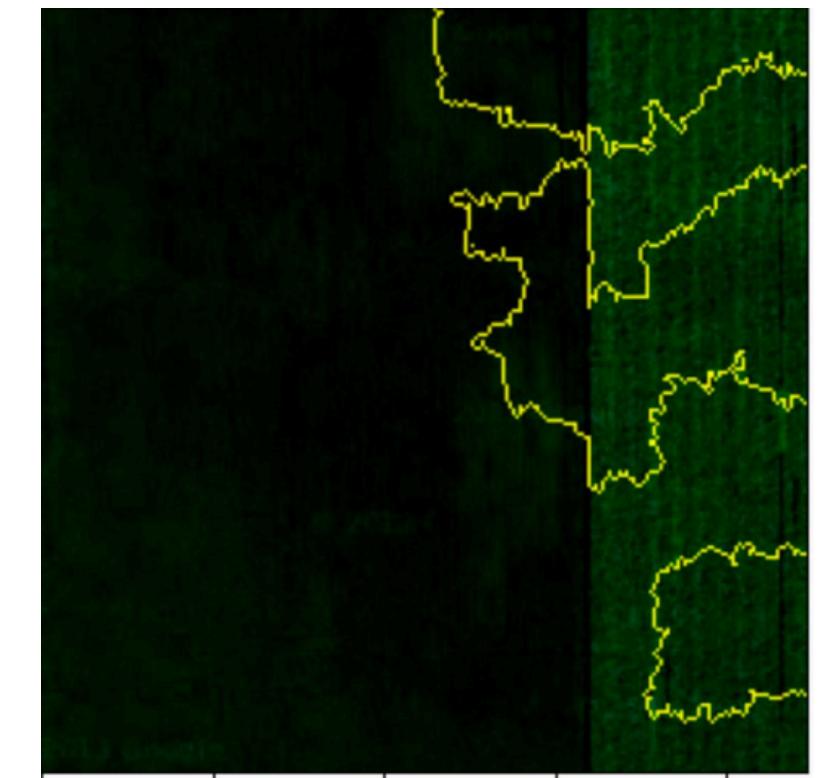
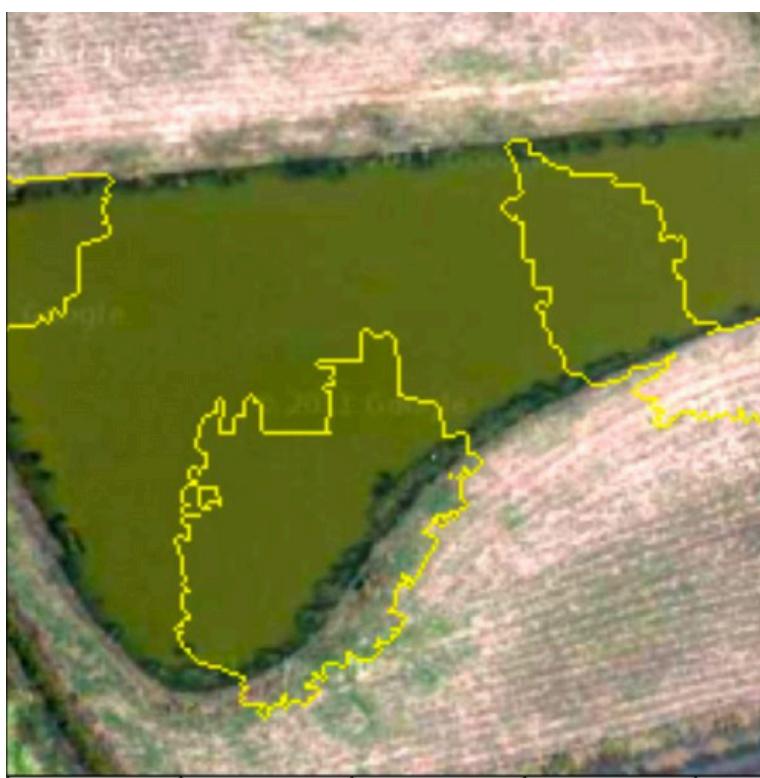
Training and testing loss



Components with the greatest influence on a particular model prediction.

- Generating samples
- Predicting samples
- Highlighting key regions

XAI (LIME)



ViT

Architecture based on a GitHub repository.

12 multi head self attention blocks.

Image resolution: 224x224

...

Training from scratch

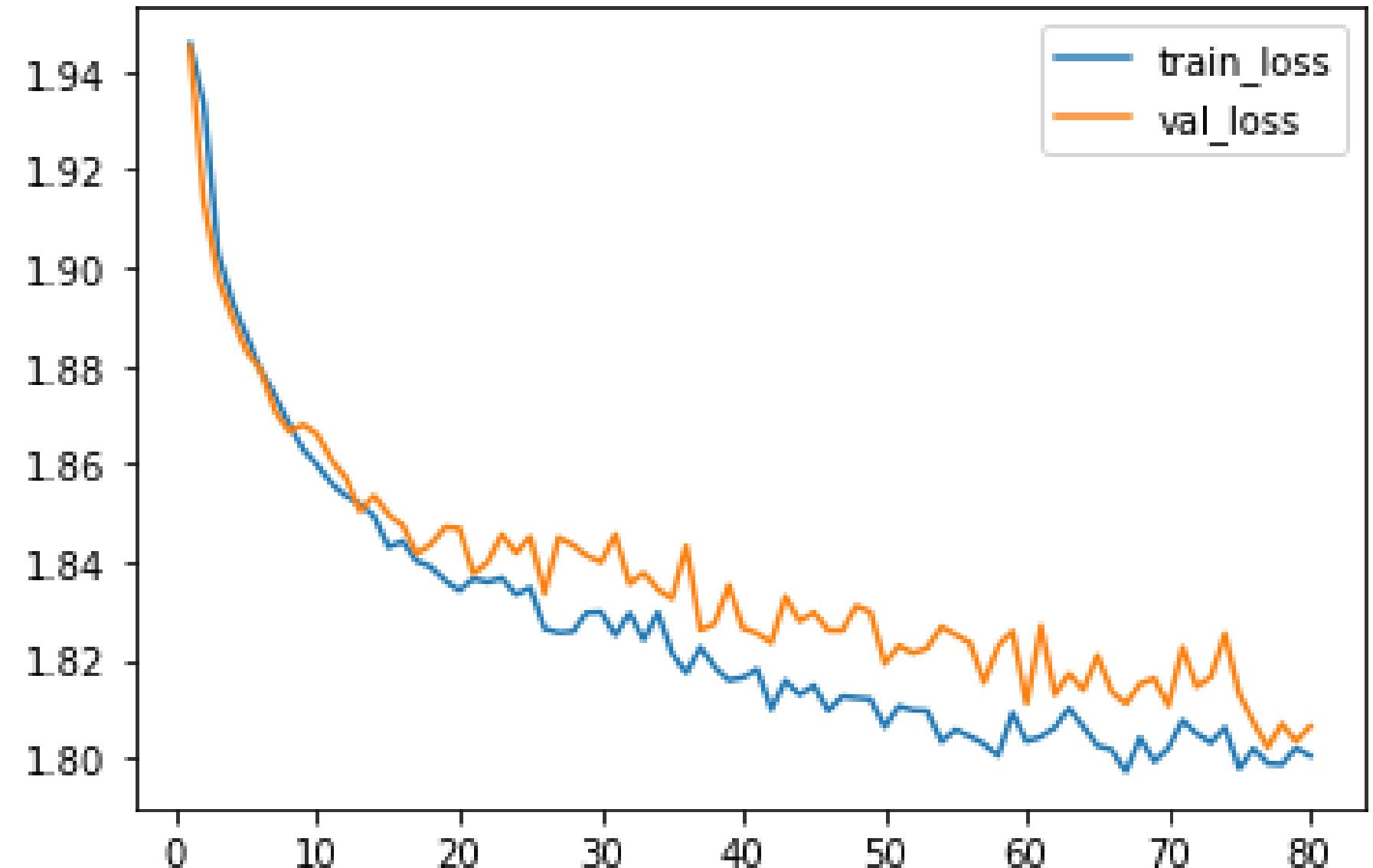
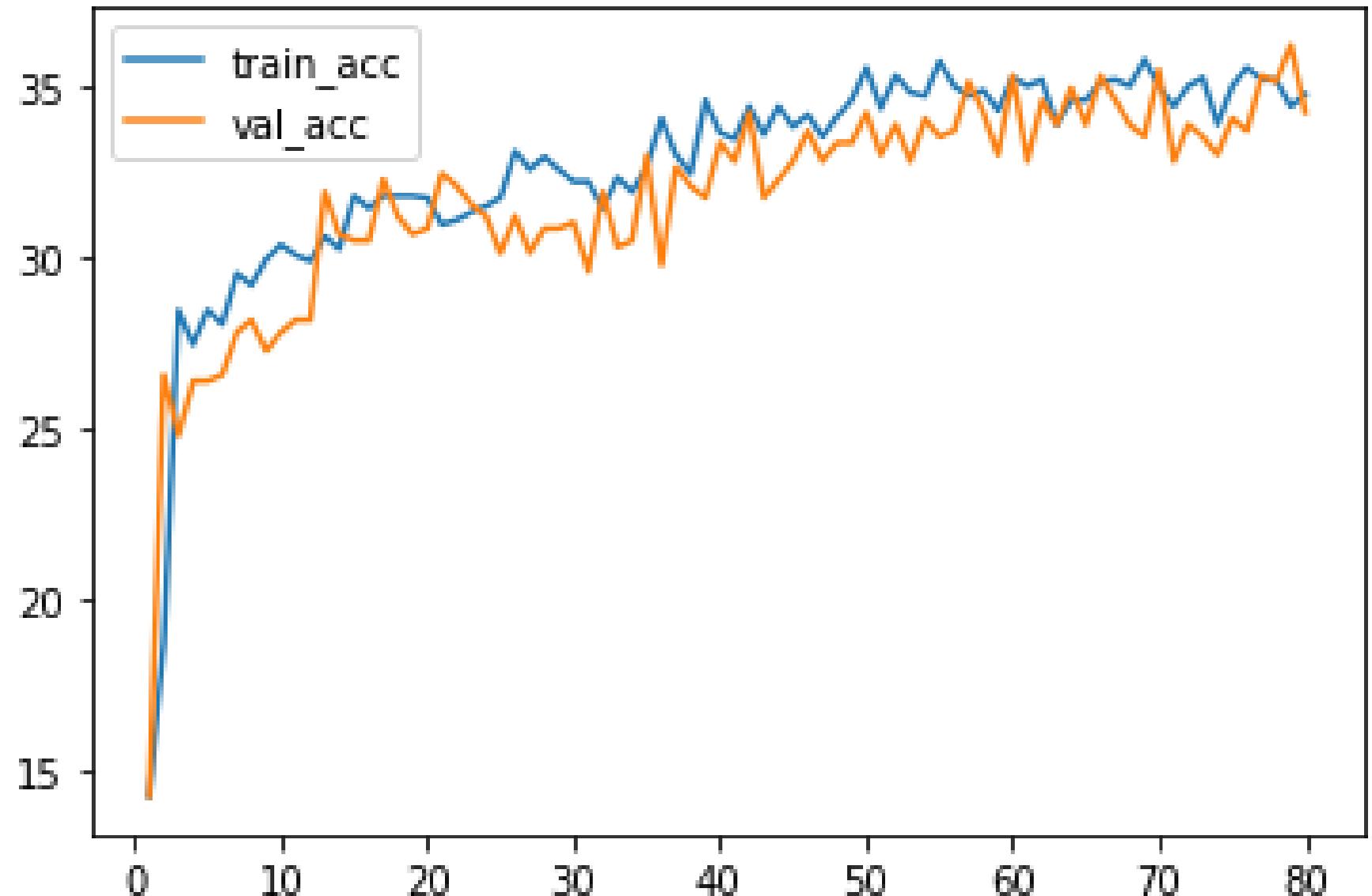
Pre-training on multi-label classification task.

The model was trained on DTD for 40 epochs. We used transfer learning for RSSCN7 dataset. The model was trained for 80 epochs with results:

Binary cross entropy loss: 1.78

Accuracy on test data: 38.57%

...



Fine-tuning

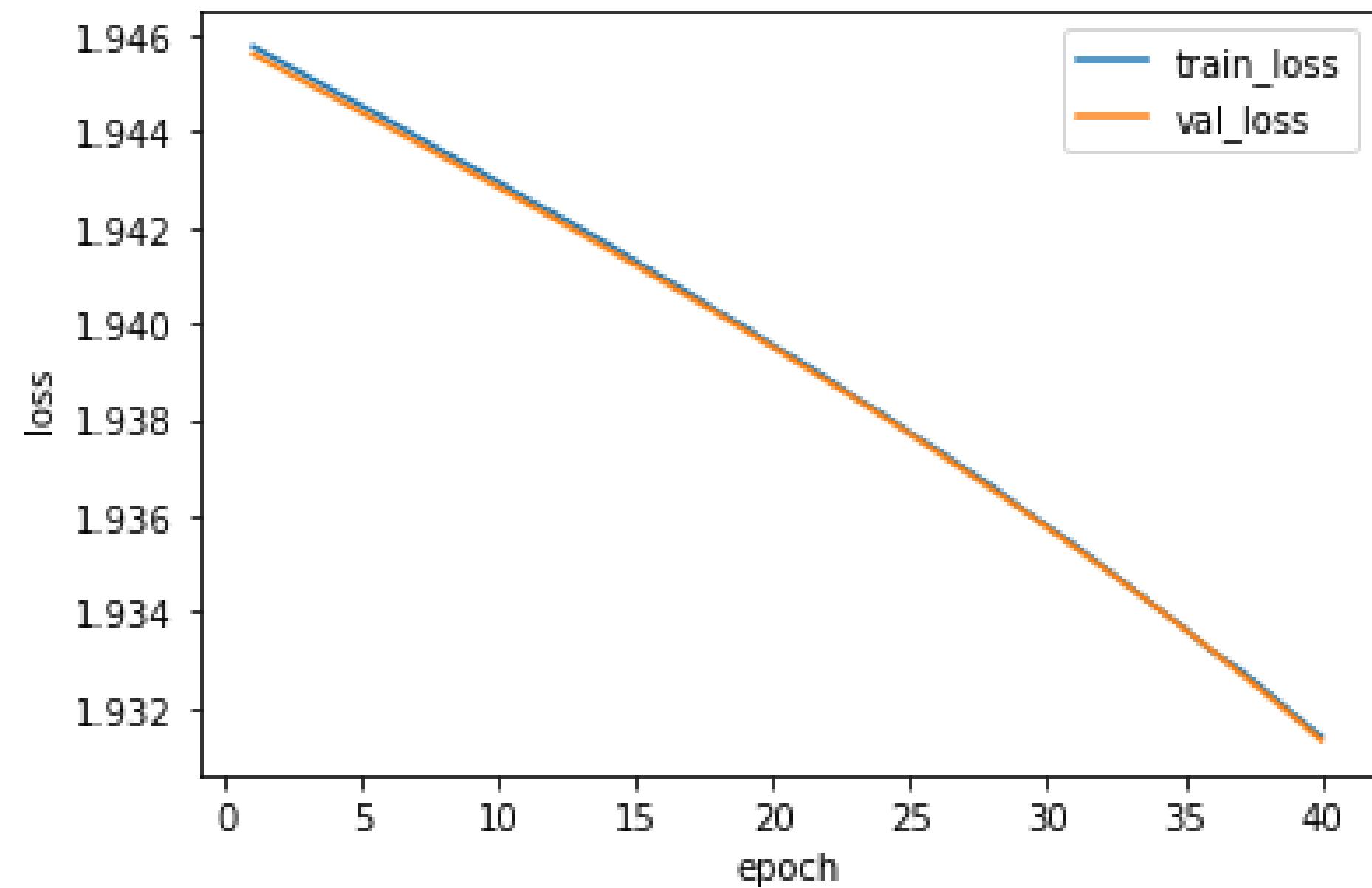
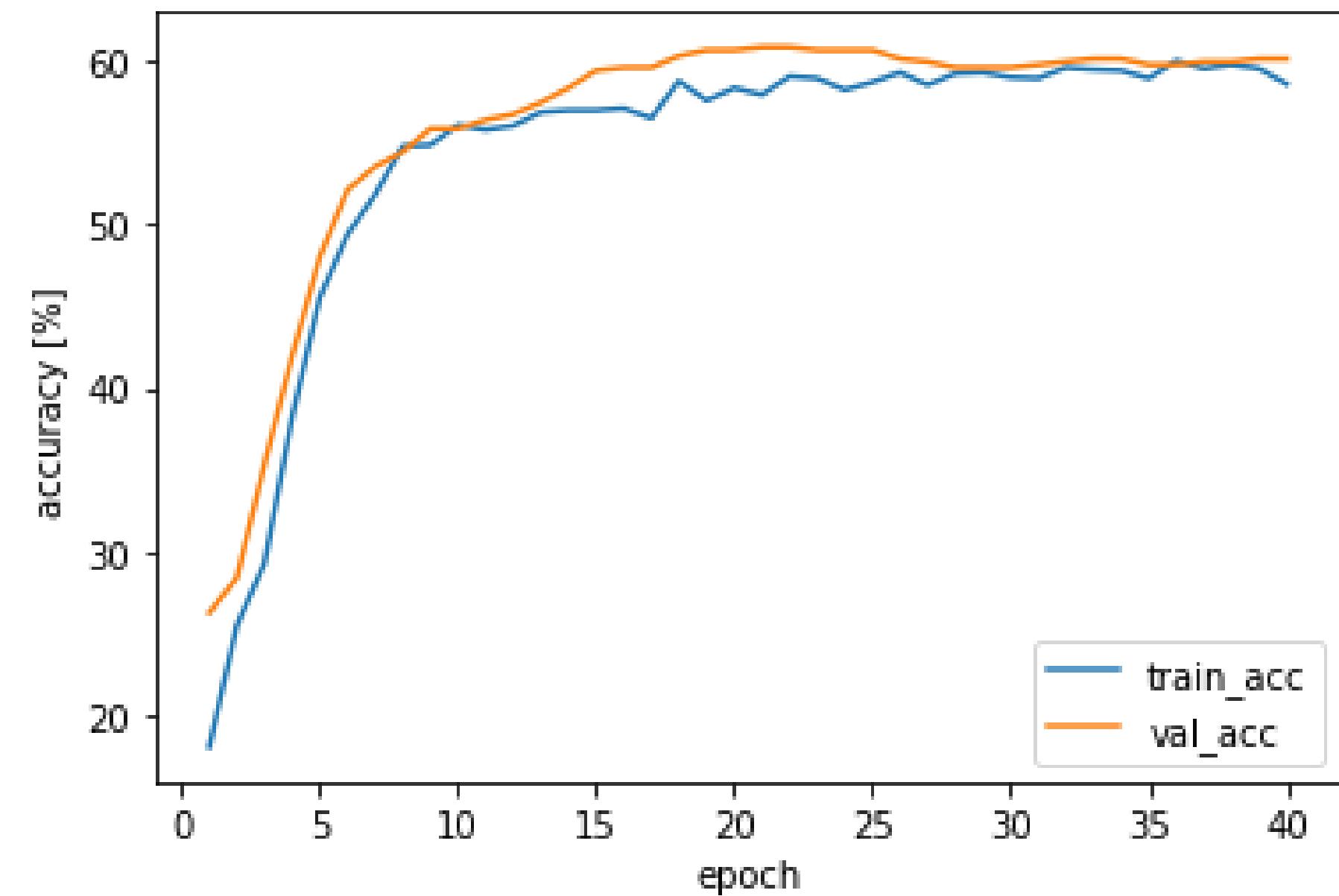
Training model taken from GitHub repository.

Epochs: 40

Average test cross entropy loss: 1.93

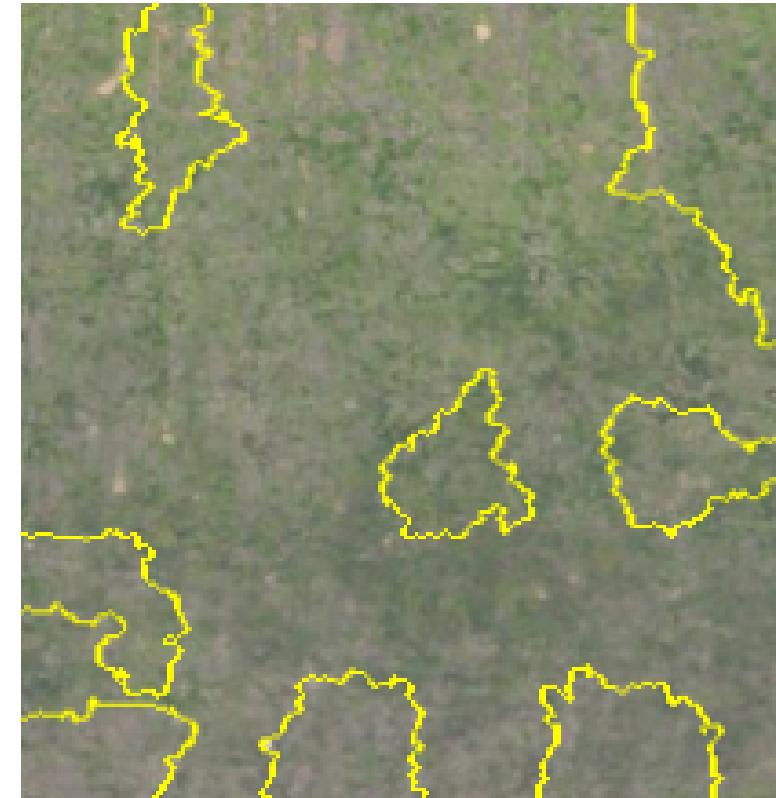
Average test accuracy: 60.71%

...

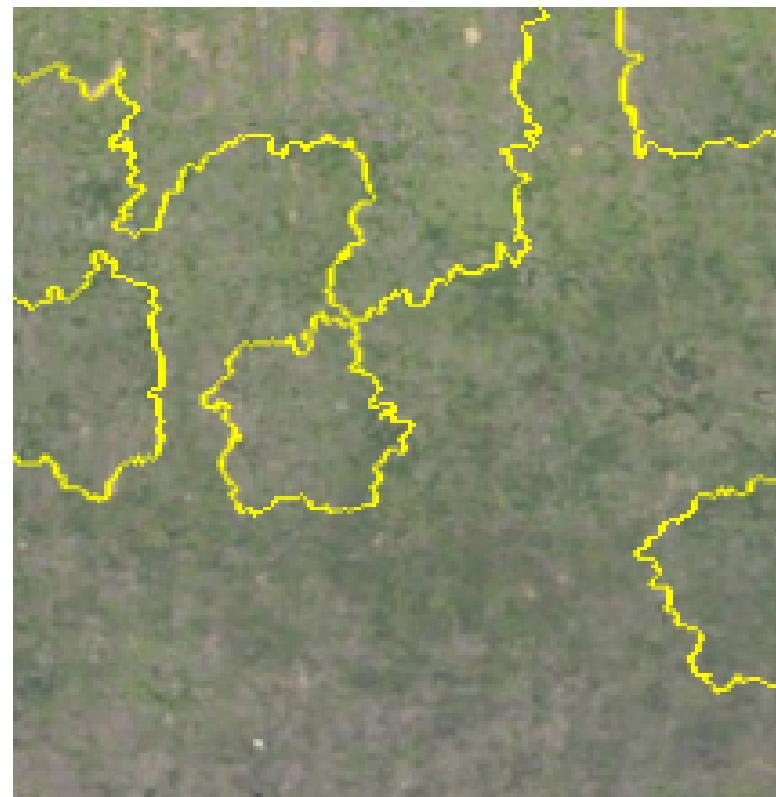


XAI (LIME)

pre-trained on DTD



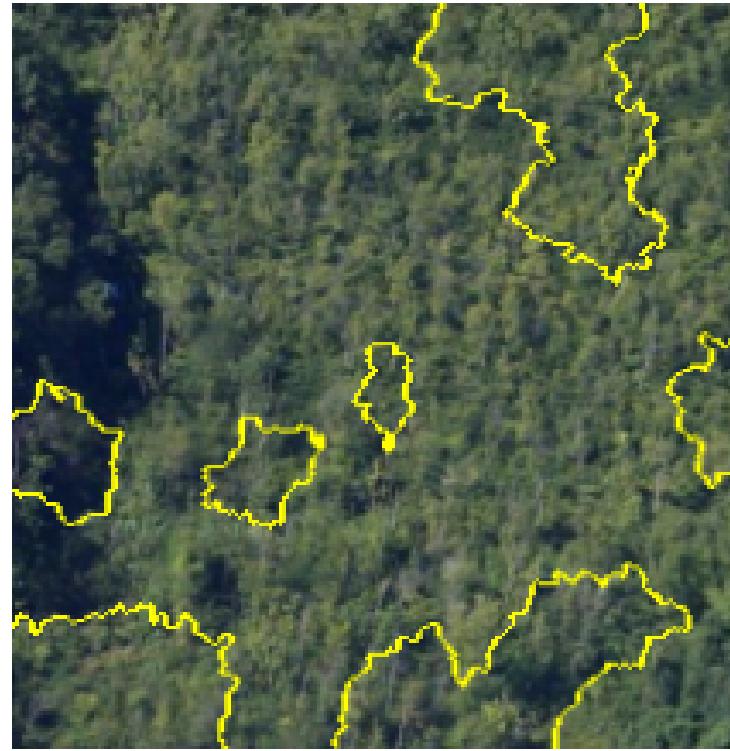
pre-trained on
ImageNet



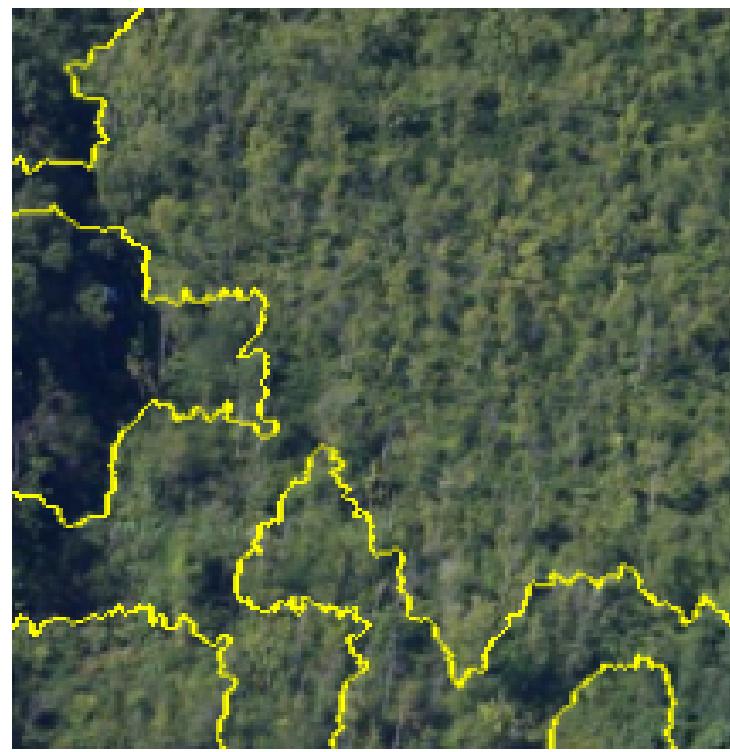
...

XAI (LIME)

pre-trained on
DTD



pre-trained on
ImageNet



Literature

DTD: <https://www.robots.ox.ac.uk/~vgg/data/dtd/>

ImageNet: <https://www.image-net.org/>

RSSCN7: <https://github.com/palewithout/RSSCN7>

CNN based image classification for DTD dataset:

<https://github.com/sakshamsds/describable-textures-dataset-classification>

<https://github.com/lukemelas/PyTorch-Pretrained-ViT>