Ashkan Zarkhah
610399196

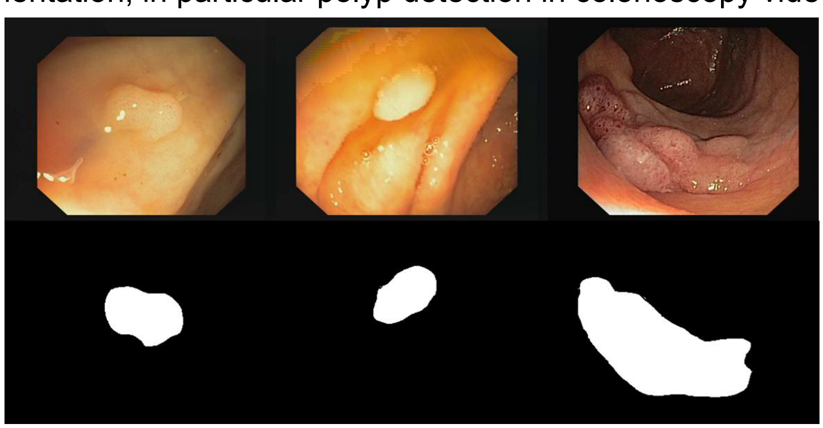
Digital Image Processing Project Report

Abstract:

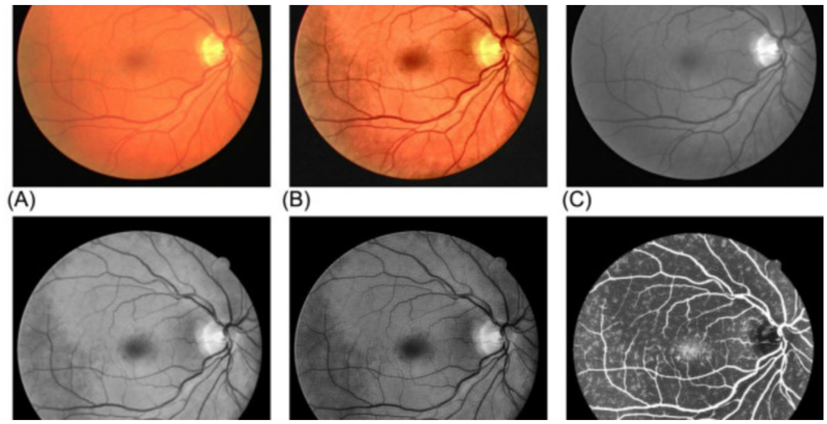
In this project, we focus on the TransNetR (Transformer-based Residual Network) model as described in the provided paper (the paper can be found alongside this report). We aim to preserve the model's architecture and measure its performance across different datasets. The original paper trained the model on the Kvasir SEG dataset for polyp segmentation. Our objective is to determine if the model performs well only on the Kvasir SEG dataset or if it generalizes well to other polyp segmentation datasets. We first train the model on the CVC-ClinicDB dataset to assess its performance across different polyp segmentation datasets. Since both Kvasir SEG and CVC-ClinicDB are relatively large datasets with simple masks, we also evaluate the model on the Retinal Blood Vessel dataset, which has a smaller size and more complex masks.

Datasets:

- **CVC-ClinicDB:** CVC-ClinicDB is an open-access dataset of 612 images with a resolution of 384×288 from 31 colonoscopy sequences. It is used for medical image segmentation, in particular polyp detection in colonoscopy videos.



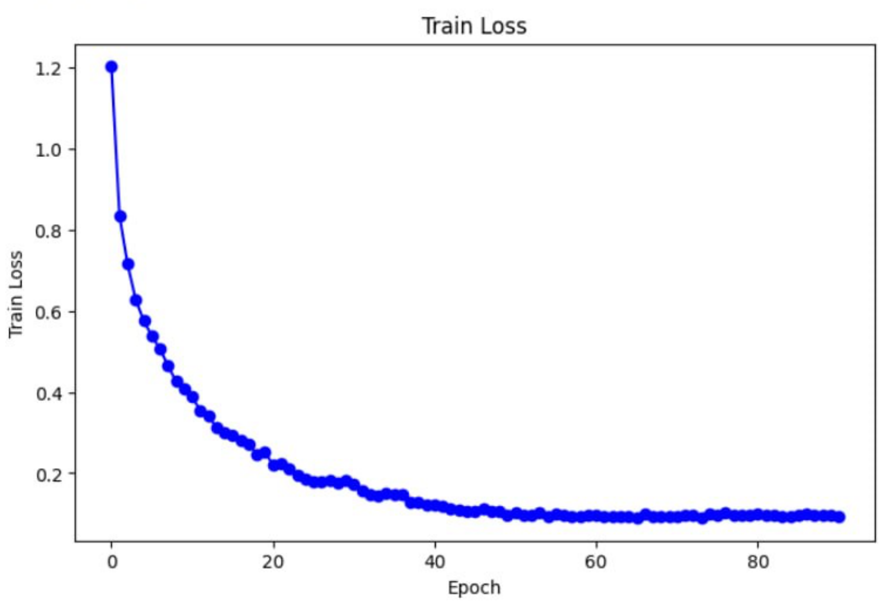
- **Retinal Blood Vessel:** Retinal Blood Vessel dataset contains a comprehensive collection of retinal fundus images, meticulously annotated for blood vessel segmentation. The dataset comprises a total of 100 high-resolution retinal fundus images captured using state-of-the-art imaging equipment.



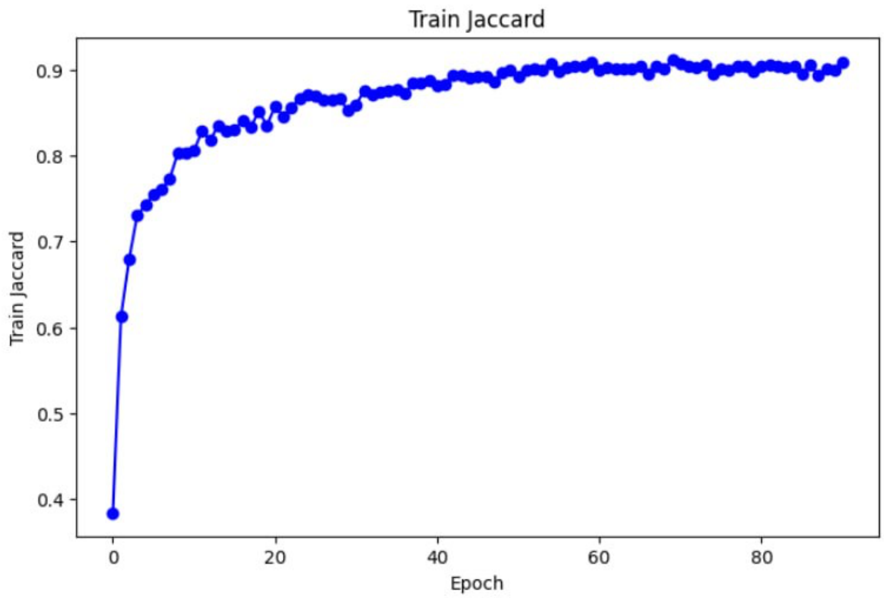
Training:

We trained the model on both datasets using the exact architecture from the original paper, adjusting only the training and test utilities to accommodate the new datasets. For each dataset, we trained the model from scratch to ensure independent and comparable results.

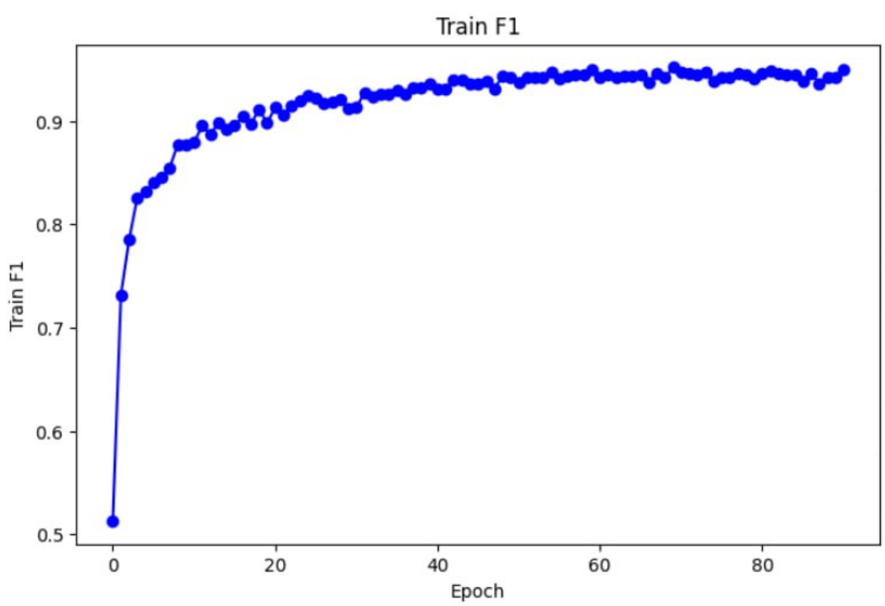
- **CVC-ClinicDB:** In training the model on the CVC-ClinicDB dataset, the training loss decreases smoothly and it reaches its minimum amount, in about epoch number 60.



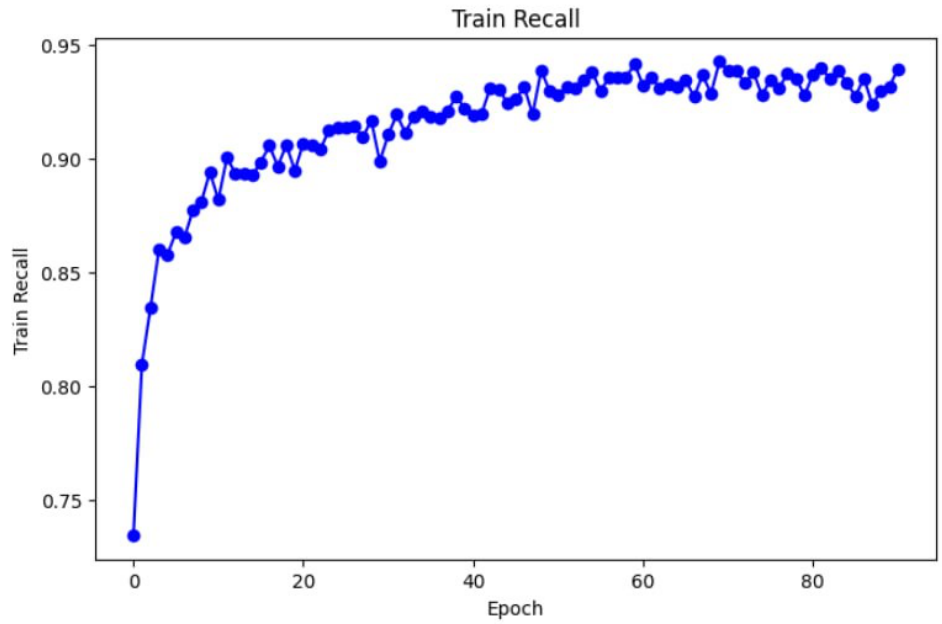
The training Jaccard coefficient increases very much in the first epochs, but it keeps to increase until the epoch number 50, and it stays approximately the same since.



The training F1-score behaves a lot like the Jaccard coefficient, but it reaches the most in epoch number 70.

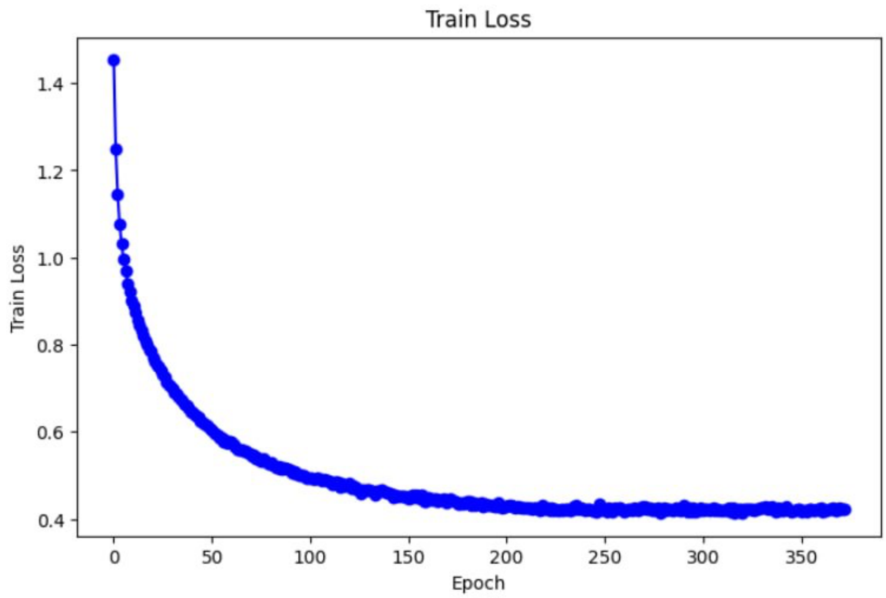


The training recall varies more than the other parameters, but the variance is very small in the last 30 epochs and can be assumed it has reached its maximum state.

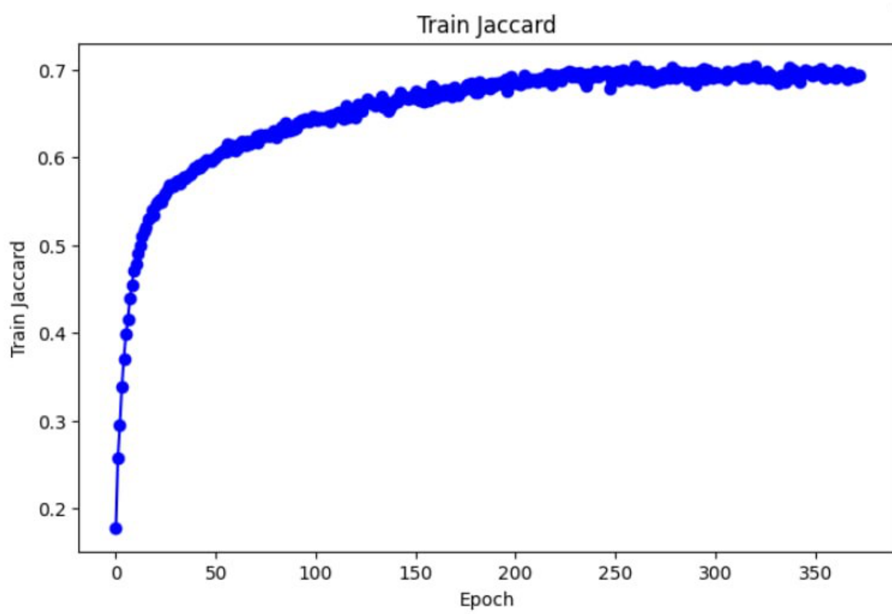


All the results together, shows that 100 epochs has been a sufficient number of epochs for learning the model on the CVC-ClinicDB dataset.

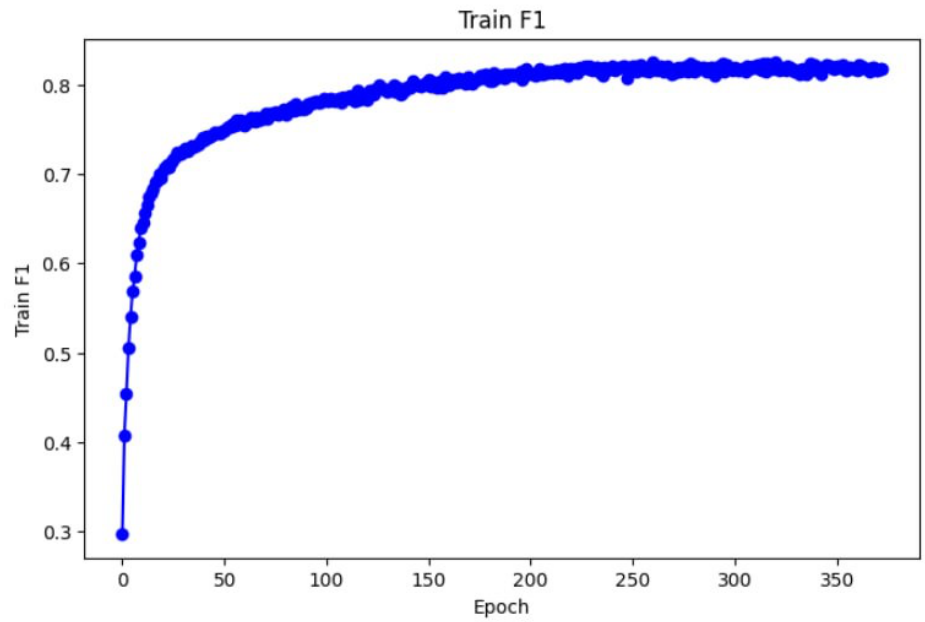
- **Retinal Blood Vessel:** In training the model on the Retinal Blood Vessel dataset, due to the smaller set of data, we decided to learn the model for 500 epochs, but it reached an early stopping in the epoch number 373, which guarantees that the model is learned completely.
  The training loss decreases and reaches its minimum amount even smoother than the previous training, and it has kept a approximately same value for the last 100 epochs.



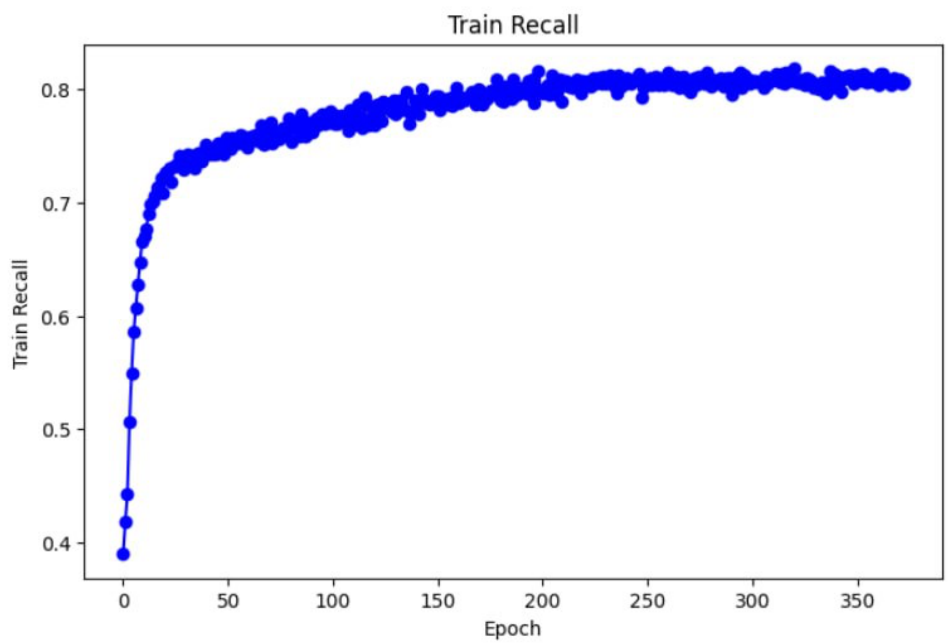
The training Jaccard coefficient increases very much in the first epochs, but it keeps to increase until the epoch number 250, and it stays approximately the same since. It is notable that the training Jaccard coefficient is also less variant than the previous learning.



The training F1-score behaves a lot like the Jaccard coefficient, and it is again much less variant than the previous learning.



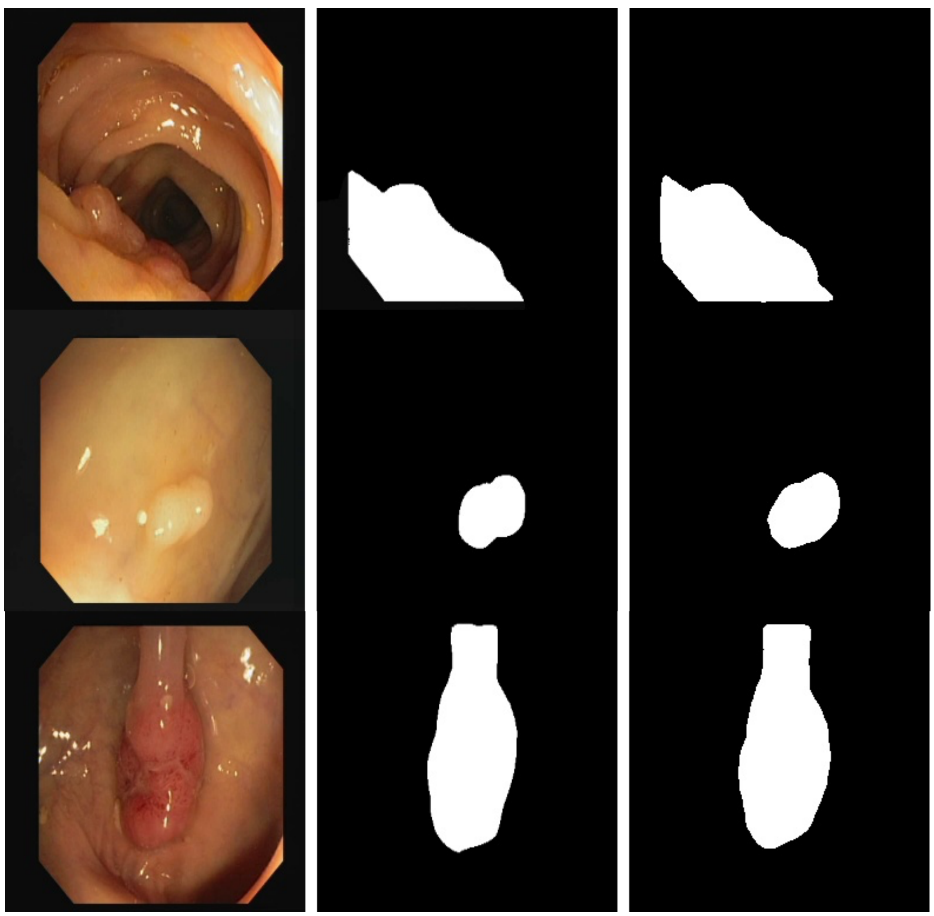
The training recall increases very much in the first 20 epochs and it keeps increasing until the last 50 epochs. Similar to the previous learning, this learning has also a variant training recall, but the variance is even smaller.



All the results together, shows that 373 epochs has been a sufficient number of epochs for learning the model on the Retinal Blood Vessel dataset.

Testing:

- **CVC-ClinicDB:** On the 122 test images, the model has been able to reach the evaluation metrics of:
  - Jaccard Coefficient: 0.8534
  - F1-score: 0.9079
  - Recall: 0.9077
  - Precision: 0.9297
  - Accuracy: 0.9881
  - F2-score: 0.9069
  - HD: 3.2200
  - Mean FPS: 80.88709162660159



- **Retinal Blood Vessel:** On the 20 test images, the model has been able to reach the evaluation metrics of:
    - Jaccard Coefficient: 0.6302
    - F1-score: 0.7728
    - Recall: 0.8001
    - Precision: 0.7510
    - Accuracy: 0.9613
    - F2-score: 0.7884
    - HD: 5.5486
    - Mean FPS: 83.48393787127408

Comparing Results:

The results indicate that the TransNetR model performs better and faster on larger datasets with simpler masks. It struggles with predicting tiny and disconnected segments and requires more data, not more epochs, to improve performance. To compare the datasets with other models, we have:

- **CVC-ClinicDB:** First we compare the dataset with the results achieved from the original paper on a similar but different dataset.

	Recall	**Precision**	**F2-score**
CVC-ClinicDB	0.9077	0.9297	0.9069
Kvasir-SEG	0.8843	0.9073	0.8744

As we can see, the model has worked even better on the new dataset, and it is safe to assume it works very good on the polyp segmentation task.

Now to compare the TransNetR model with different models on this dataset we have:

Model	**F1-Score**
DUCK-Net	0.9684
EMCAD	0.9521
RaBiT	0.951
UGCANet	0.950
MSRF-Net	0.9420
KDAS	0.925
TransFuse-S	0.918
TransNetR	**0.9079**
ResUNet++ + TTA	0.9017
PraNet	0.8990
U-Net	0.8230
ResUNet++	0.7955
U-Net++	0.7940

As we can see, TransNetR has achieved better result than the similar models, but there are other models with better architecture designed for this dataset.

- **Retinal Blood Vessel:** As for the Retinal Blood Vessel dataset, we do not have any similar datasets learned on this model, so we only compare the TransNetR model with other models on this dataset.

Model	F1-Score
FR-UNet	0.8316
SA-UNet	0.8263
LadderNet	0.8202
ConvMixer	0.8245
ConvMixer-Light	0.8215
TransNetR	**0.7728**
DR_2021	0.75

As we can see, TransNetR is not ideally suited for unconnected masks and small datasets but still delivers results that are convincing and close to the average for state-of-the-art models.

Summary:

In this project, we evaluated the performance of the TransNetR (Transformer-based Residual Network) model across different medical image segmentation datasets. The original paper trained the model on the Kvasir SEG dataset for polyp segmentation. We extended this work by testing the model on the CVC-ClinicDB dataset for polyp segmentation and the Retinal Blood Vessel dataset for blood vessel segmentation, which has much more precise masks but a smaller amount of data. As a result, we found that the TransNetR model is designed to identify simpler masks and, with more data, it can learn much faster than similar models.

All the code can be found in the “transnetr-on-different-datasets.ipynb” file included with this report. The original paper is available in the “TransNetR.pdf” file, and all the predicted masks can be found in the results folder included with this report.