



IMT Atlantique

Bretagne-Pays de la Loire

École Mines-Télécom

UE DEEP Learning 2023

TransUNet: Transformers Make Strong
Encoders for Medical Image Segmentation

Santiago Quinteros
Sofia Pistone

Index

1. The model

1.1 The model

2.2 Starting point

2.3 Objective

2. Tests

2.1 How to test

2.2 Test for different number of epochs

2.3 Test for different number of skips

2.4 The impact of the input image size (resolution)

2.5 Influence of Patch Size

3. Results



IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

1. The model

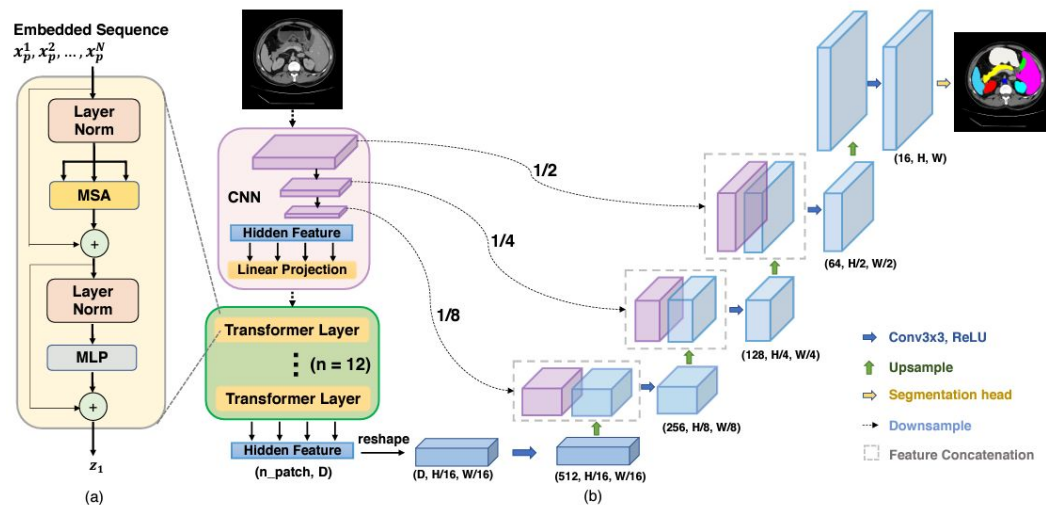


IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

1. The model:

1.1 TransUNet

4



Concept: Using transformers as encoders in a UNet

Why: allows modeling explicit long-range relation, which is not the case for CNNs

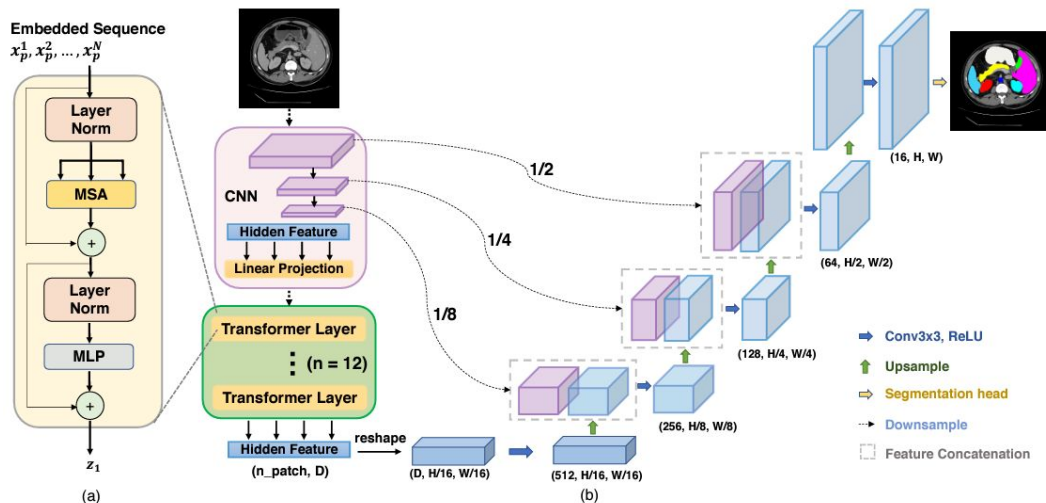
1. The model:

5

1.1 The relevance of skip connections

Skip connections allow a direct flow of information from the encoding layers to the decoding layers.

This helps to ensure that detailed, low-resolution information that is lost during dimensionality reduction in the encoding layers can be recovered in the decoding layers.



More skip connections =
+ more preservation of local
details and features

- Increase in the complexity of
the architecture.

Data Set:

- **Characteristics:** multi-organ labeled CT images of 512x512x(85~198).
- **Training command:**

```
CUDA_VISIBLE_DEVICES=0 python train.py --dataset Synapse --vit_name R50-ViT-B_16
```

- **Problems:**
 - Running out of RAM.
 - Training takes a whole class to run.

- **Solution:** Reduce the batch_size to 6 reduced the epochs to 15

1. Contexte :

2.2 Starting point

7

- **Test command:**

```
python test.py --dataset Synapse --vit_name R50-ViT-B_16
```

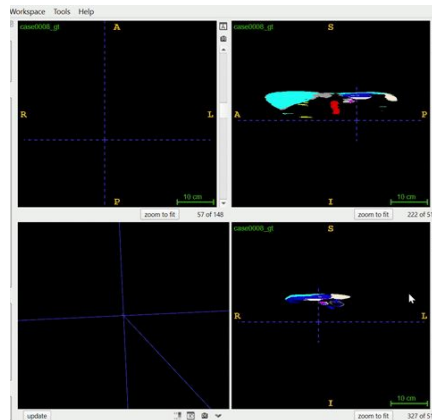
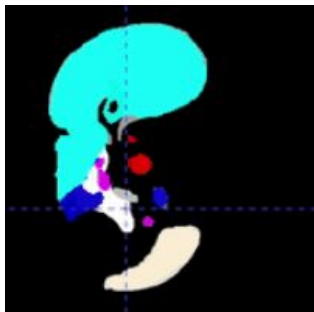
- **Problem:**
 - Tests take a lot of space in the Hard drive
- **Solution:** migration to Google colab.

1. Context :

2.3 Objective

First inference successfully generated!

But the results are poor due to the low number of epochs (15) and the default parameters



Objective: find the best hyper-parameters of the model

- to understand the network
- to have the best segmentations with limited time and without a super computer.

2. Tests



IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

We made a script that performs numerous tests of your choice without the need for a user to run them manually.

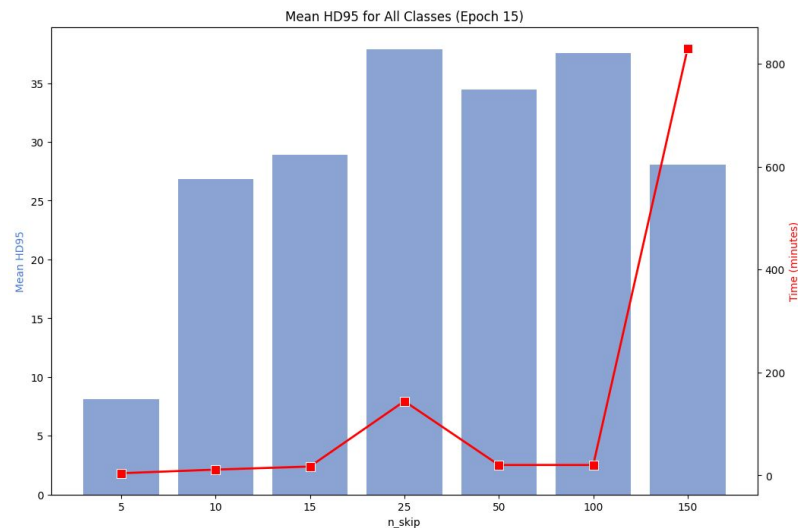
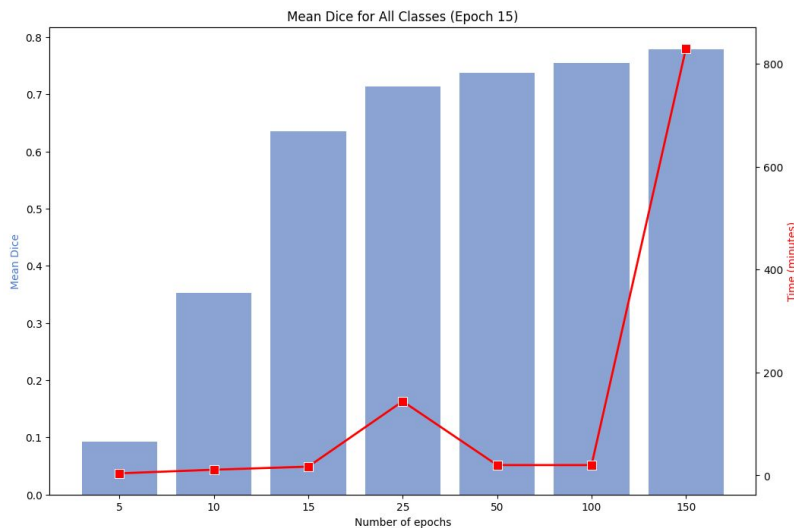
We made a script that saves the relevant information and automatically deletes the old models to avoid space problems.

We change the output so that instead of being done by console, it is done in a csv file and the graphics are done easily.

2. Tests:

2.2 Test for different number of epochs

11

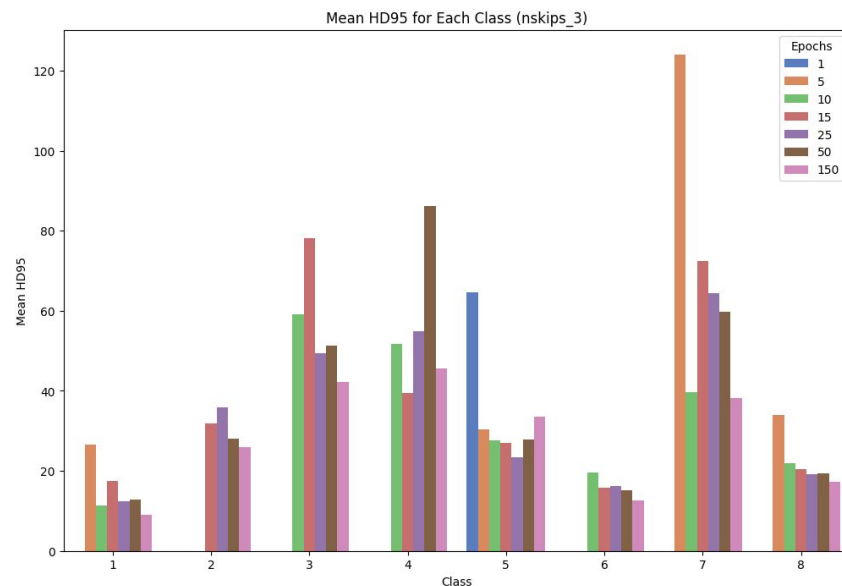
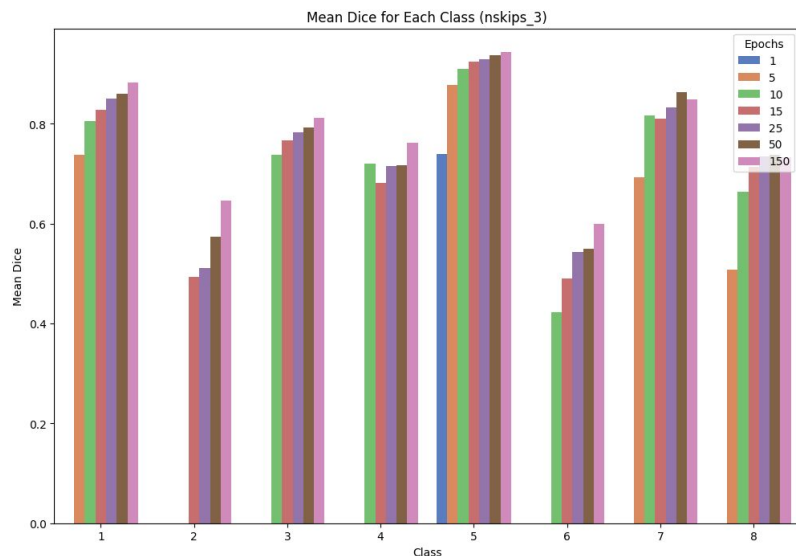


As expected, the dice score improved as we added epochs, but also considerably the training time. The difference in dice score between 15 epochs and 150 is not crucial, so we will do the tests with 15 epochs

2. Tests:

2.2 Test for different number of epochs

12



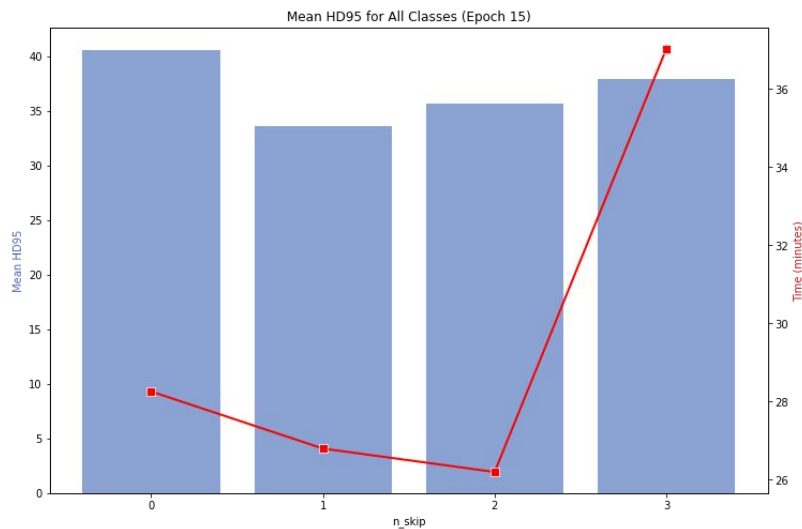
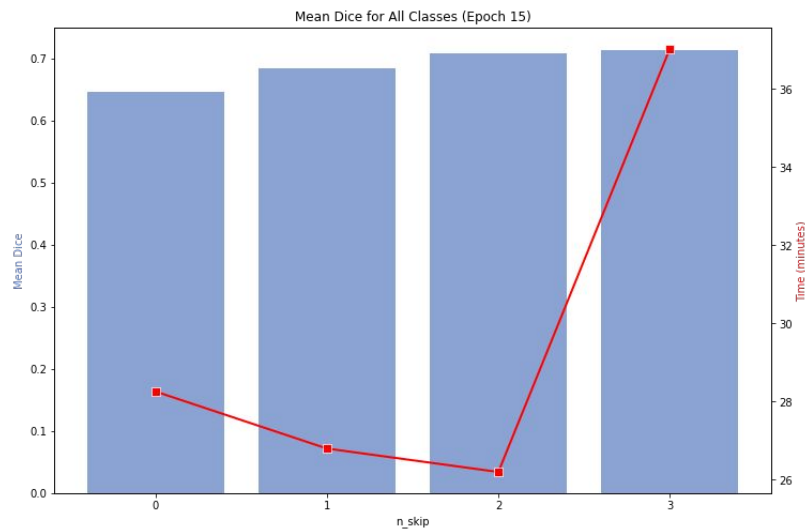
Evidently the dice_score was lower the higher the number of epochs.
Likewise the Hausdorff distance was lower the higher the number of epochs.

We note that the dice score is not much lower for 15 epochs than for 150 epochs and takes much less time. This is why we will use 15 epochs for the following tests.

2. Tests:

2.3 Test for different number of skip connections

13



As expected, the dice score improved as we added skip connections.

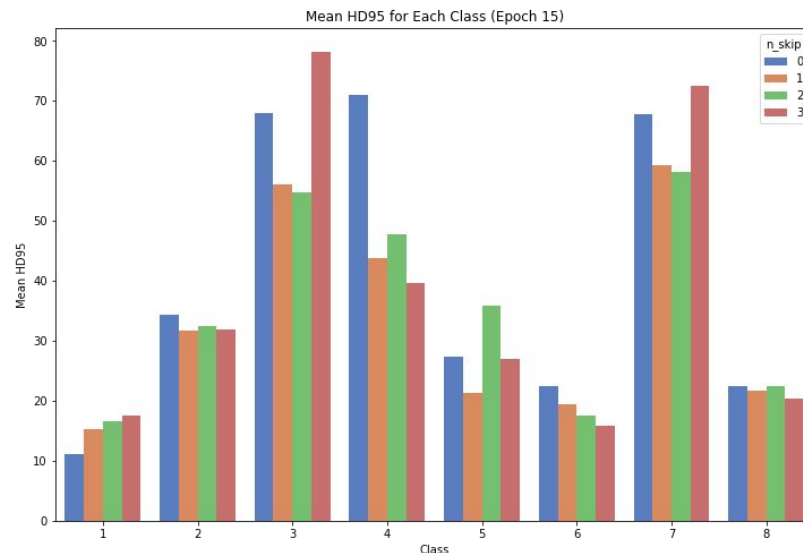
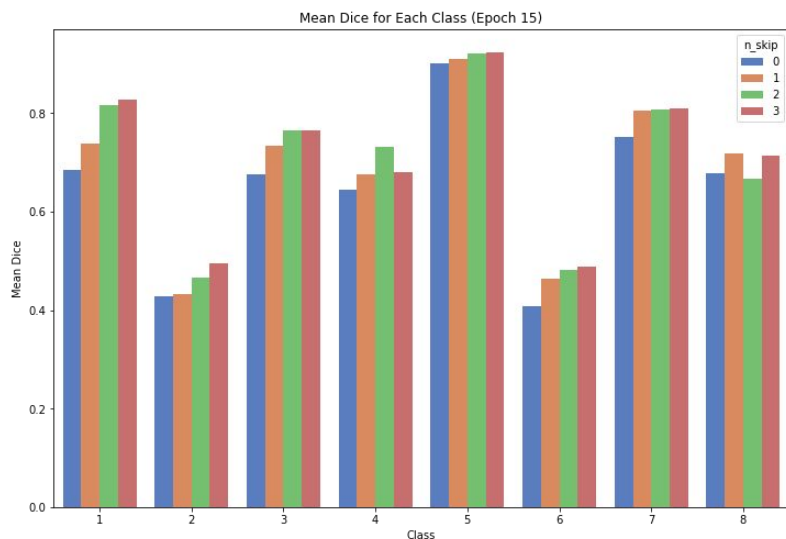
But training takes twice as long.

Adding more skip connections could be a good idea, although the training time would increase considerably.

2. Tests:

14

2.3 Test for different number of skip connections



Using 3 skip connections improves segmentation by 10%. This shows the great value of this approach versus using a transformer and decoding directly.

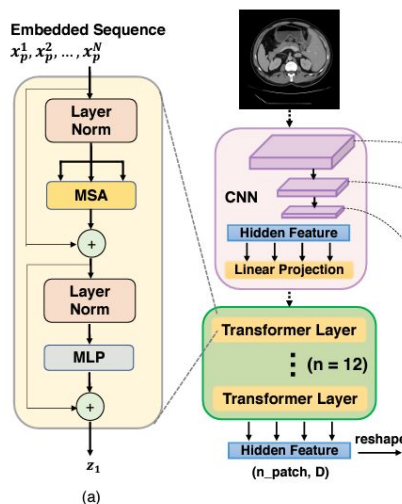
2.4 The impact of the input image size (resolution)

Until now we have been using a lower resolution than the original for faster processing.

It is expected that using larger images will give better results, however, the higher the resolution, the more features the model has to take into account, the more RAM it consumes and the longer it takes to train.

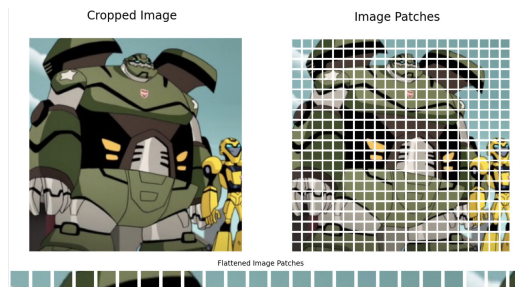
Indeed, we couldn't finish this test as RAM memory reached its maximum capacity.

Possible improvement for the tests:



Dividing the image into patches allows the use of transformers to capture the relationships between different parts of the image.

Using larger patches captures less detail but greatly simplifies the model allowing for faster training.



3. Results



IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

3. Results :

Interactive visual comparison for patient 38

18

Ground truth

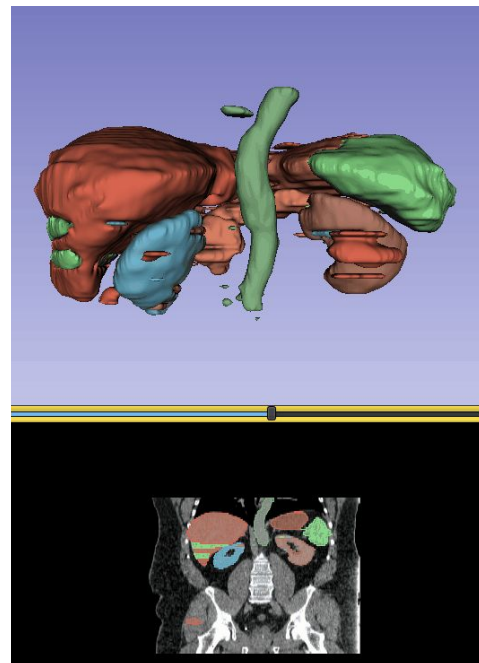


École Mines-Télécom

150 epochs 3 skip



20 epochs 3 skip



20 epochs 0 skip

