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

Attention-based U-Net for segmentation

Deep Learning - Topic 7

Emanuelle PASSOS MARTINS, Nathan YGÉ

28/11/2024

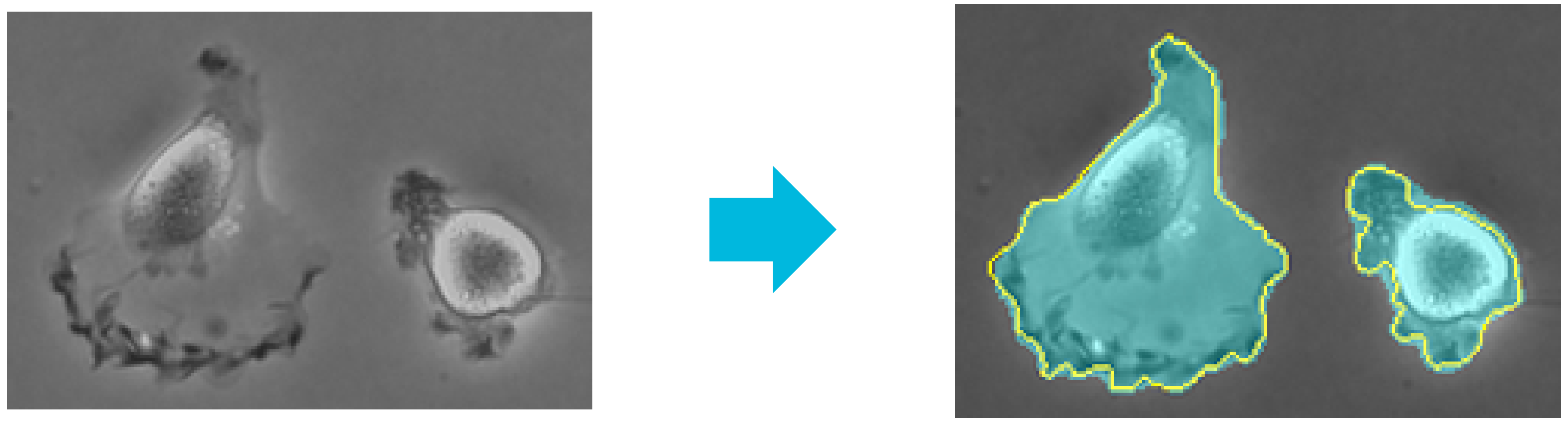
SUMMARY

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

1. Introduction

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

Segmentation of medical images with deep learning:



image

classification
(pixel by pixel)

- CNN-based architectures, such as U-Net, have demonstrated strong performance in automating this task.
- Variations in organ size, shape, and location across patients
  Cascaded frameworks:
  - Define a region of interest (ROI)
  - Make predictions on this ROI

- **High computational demand due to full-image processing!**

Manual annotation

a tedious and error-prone task

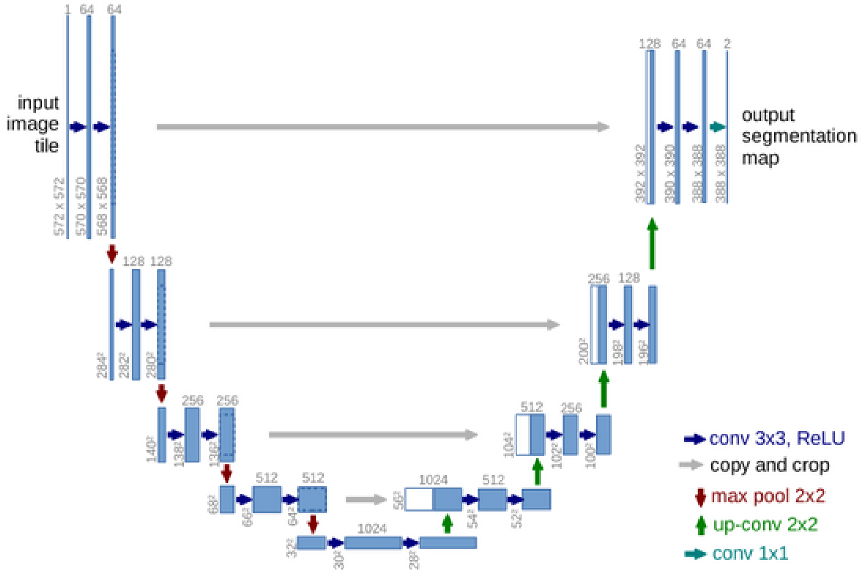


Figure 1 - Example of U-Net architecture

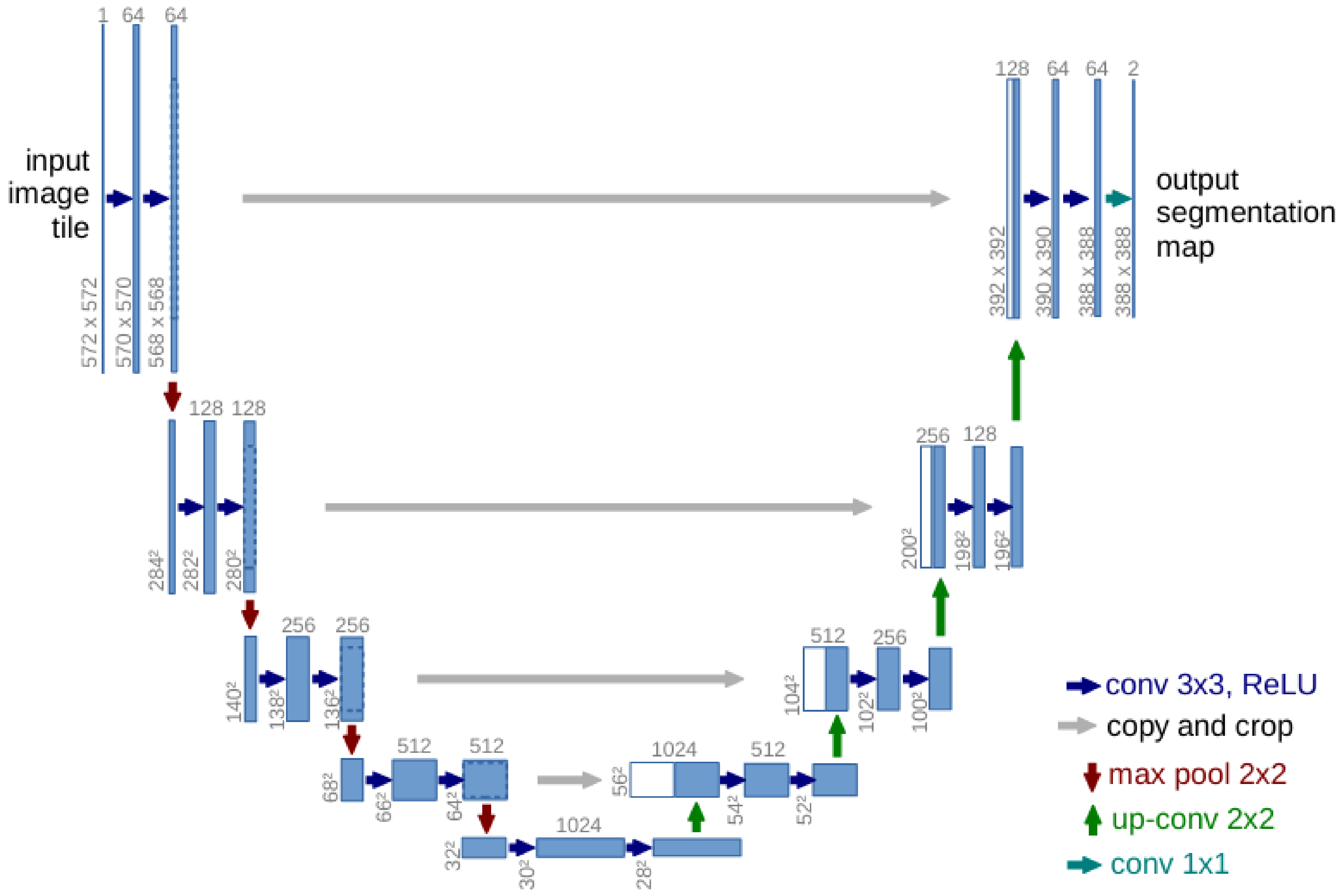
How does the U-Net work?

Contracting path

- Convolutional layers
- Activation function (ReLU)
- Pooling layers

Expanding path

- Up-convolutional layers
- Concatenation / skip connections
- Convolutional layers + ReLU
- Final 1x1 convolution layer, followed by an activation layer (sigmoid)



Attention U-Net

Attention gates (AGs) **automatically adapt their attention toward target structures!**

- No large number of parameters nor computational demand

The AG receives two main inputs:

- Feature map from the encoder (high detail)
- Gating signal from the decoder (lower resolution)
  - “where to look” without getting stuck on small irrelevant details

These two signals are combined in each AG to generate an attention map, filtering the relevant pixels.

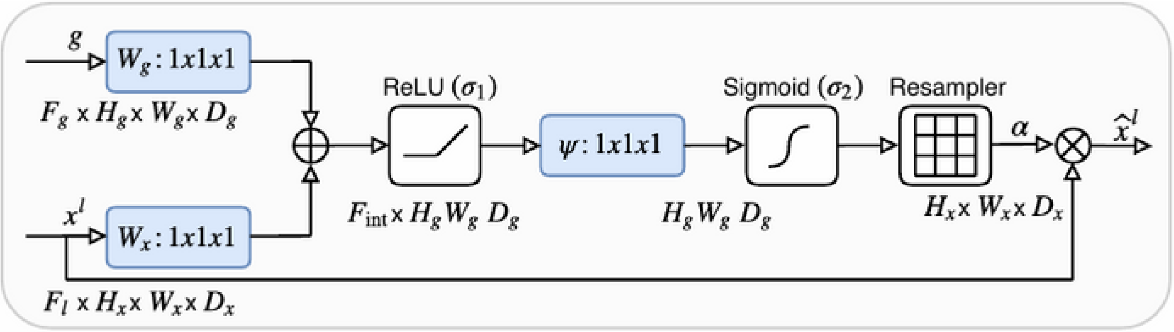


Figure 2 - Attention gate

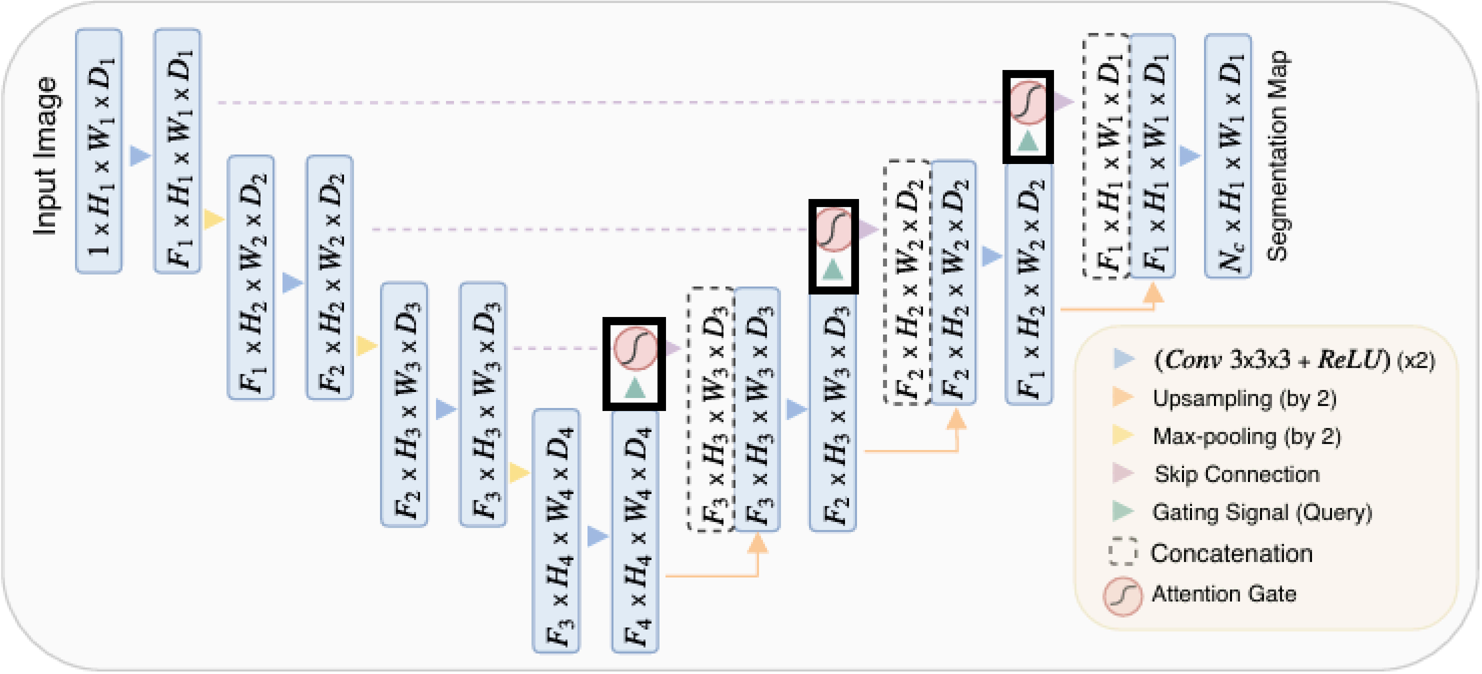


Figure 3 - Attention U-Net segmentation model

2. Methodology

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

Preprocessing

1 split

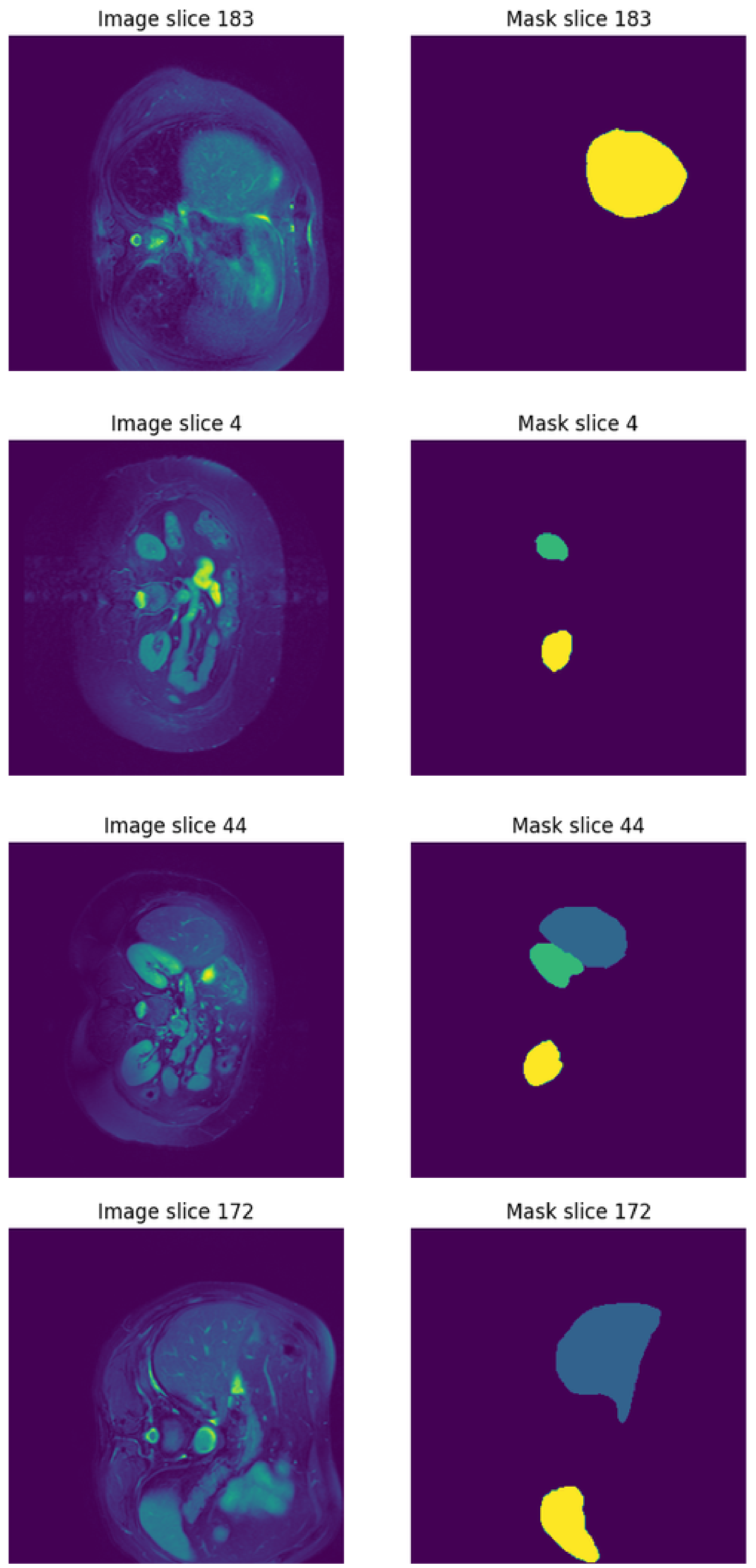
- **20** T2 MRI abdominal images + masks
  - Train: 12 patients
  - Validation: 4 patients
  - Test: 4 patients

2 data transformation: 2D into 3D slices

3 Train: slice weights based on the present classes to correct class imbalance

- 5 classes
  - Background
  - Liver
  - Left kidney
  - Right kidney
  - Spleen

```
Fréquence des pixels dans le masque 3D complet:
Classe 0.0: 2227580 pixels (94.4171 %)
Classe 80.0: 100875 pixels (4.2756 %)
Classe 160.0: 11883 pixels (0.5037 %)
Classe 240.0: 11851 pixels (0.5023 %)
Classe 255.0: 7107 pixels (0.3012 %)
```

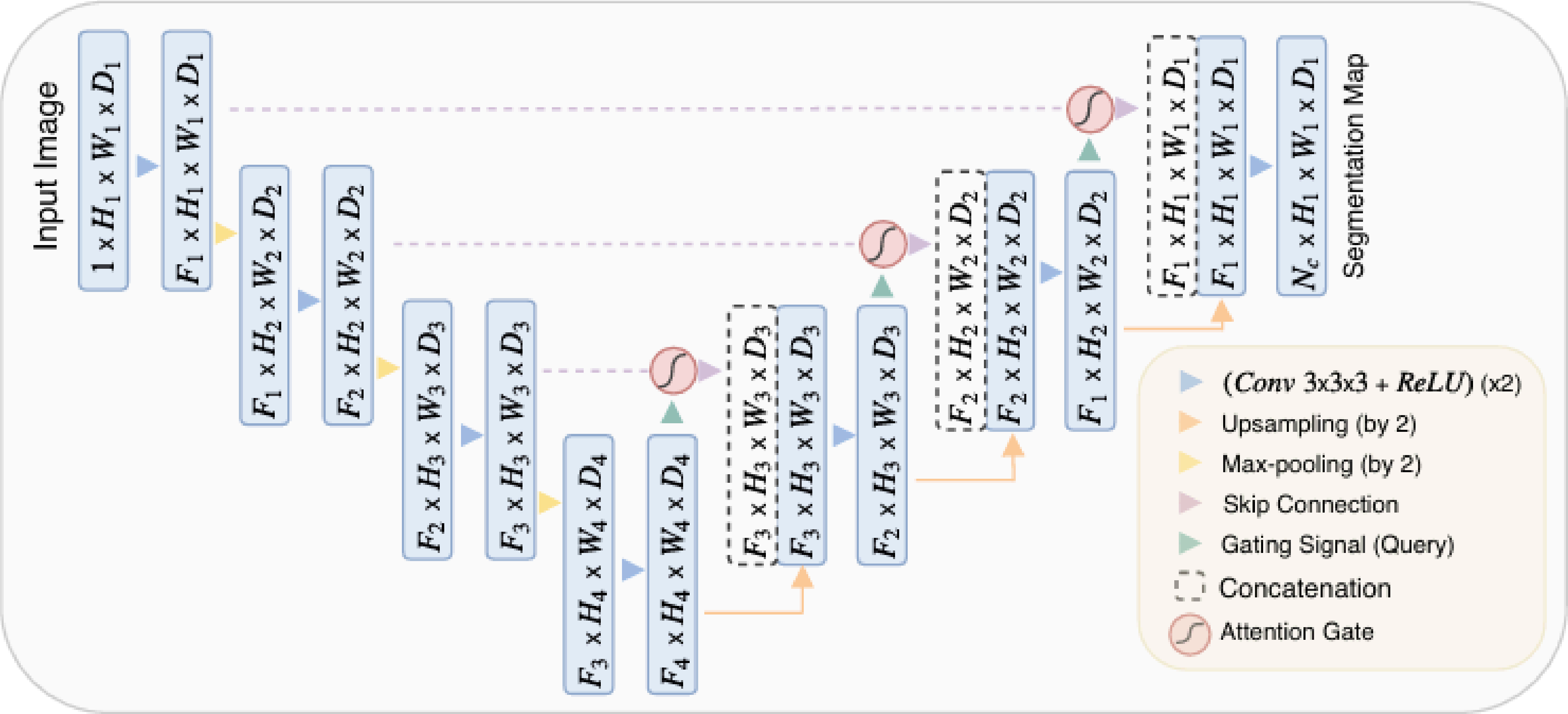


Training + Evaluation

- U-Net
- U-Net with attention gate

Best configuration:

- Epochs = 40
- Batch size = 4
- Patience = 10
- Optimizer = Adam
- Learning rate = 1e-4



Loss

- Cross-entropy
- Focal loss
- Dice loss

Metrics

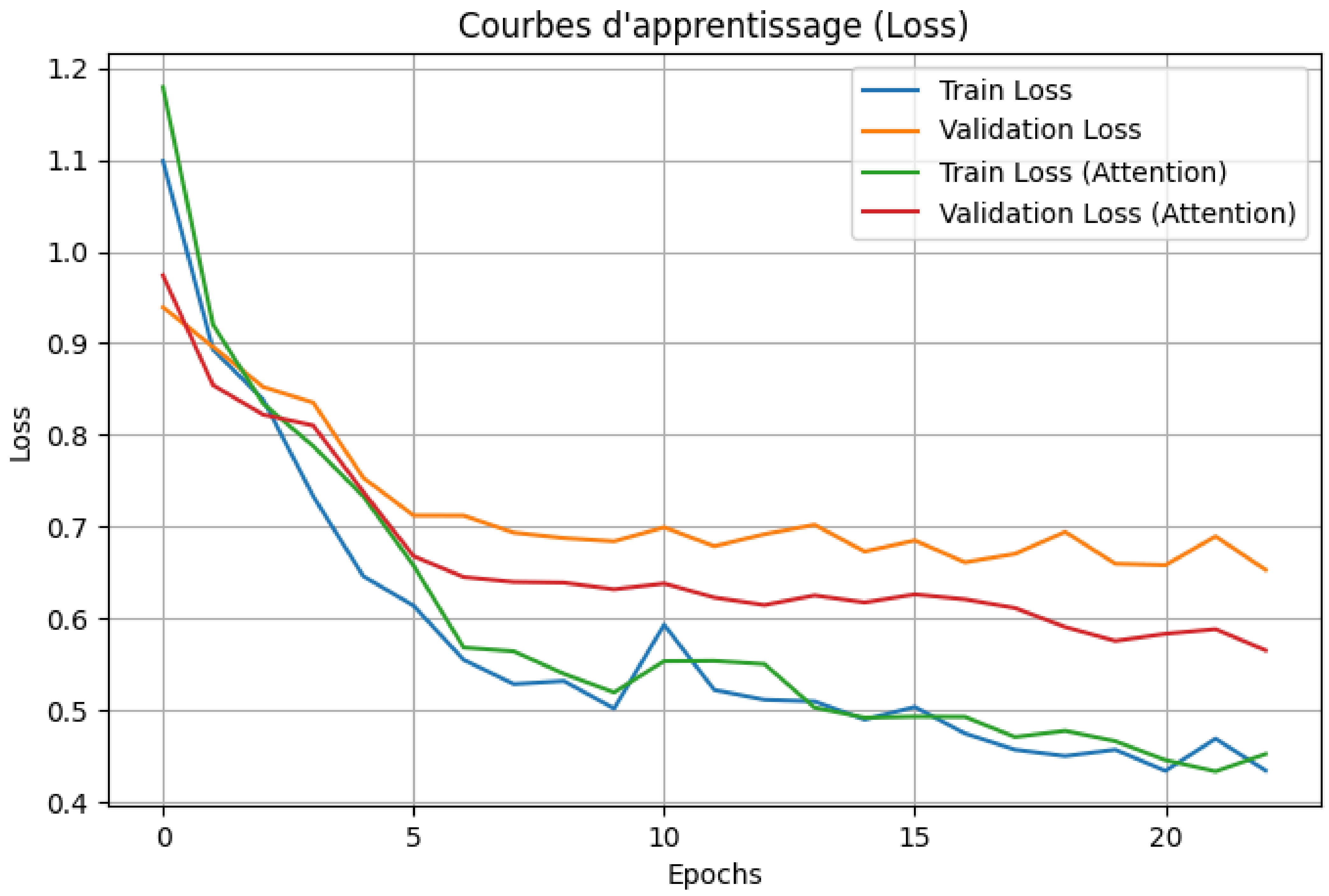
- Dice score (per class, multiclass)
- Average Symmetric Surface Distance (ASSD)
- Haussdorf distance

3. Results

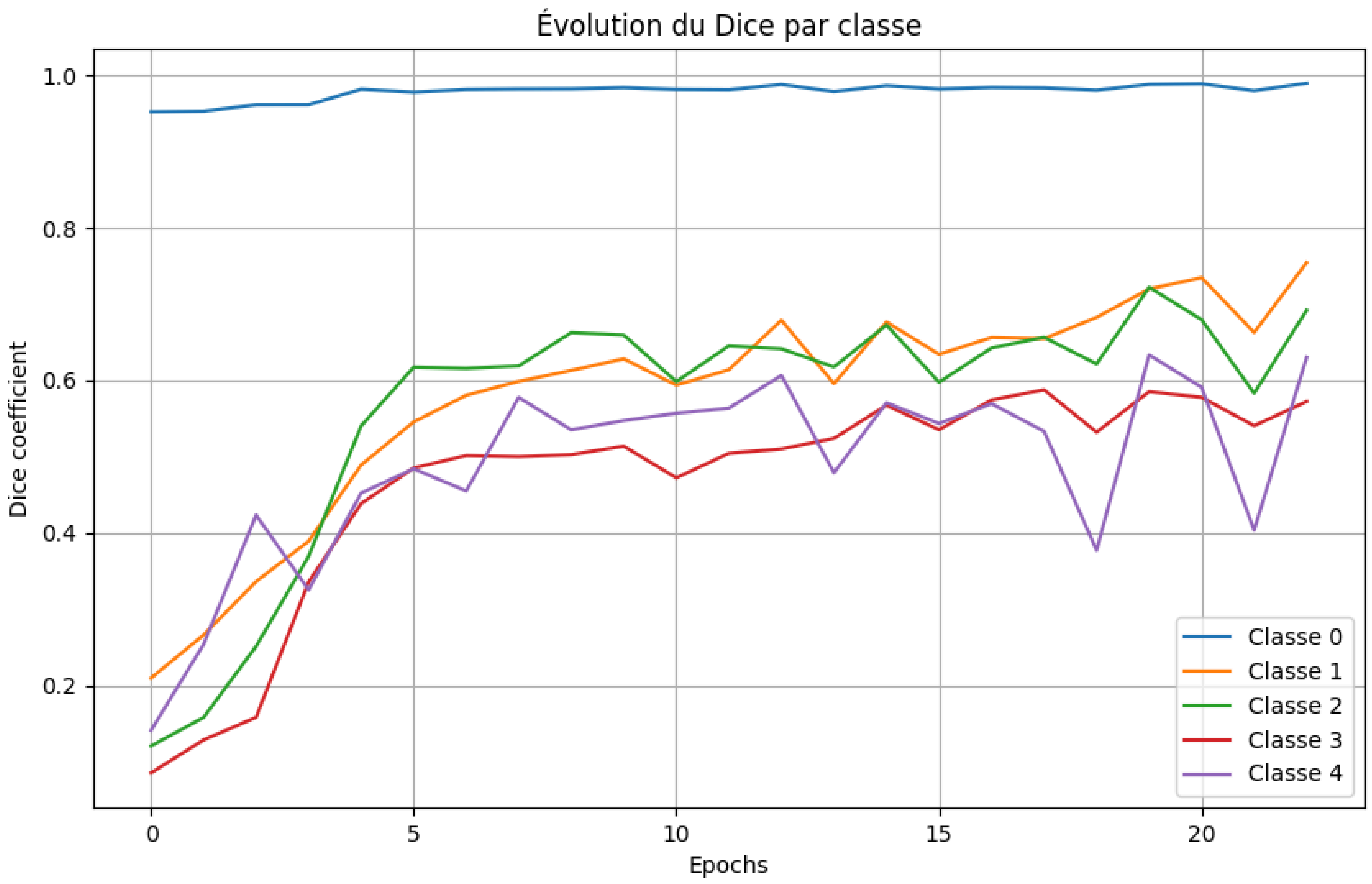
IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

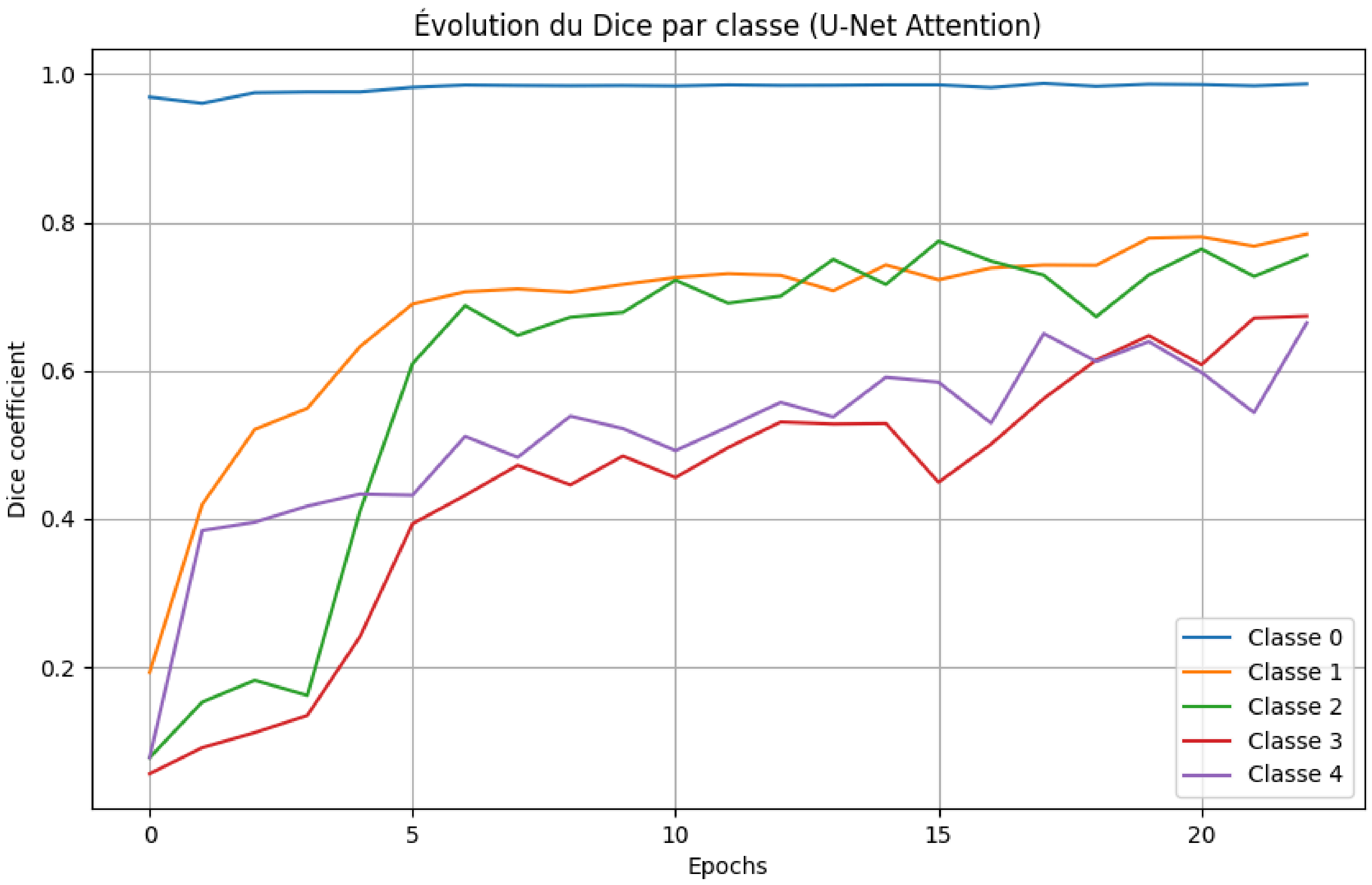
Train and validation losses

it has **similar train loss**
BUT
different validation loss meaning it generalize better



Evolution of dice per class

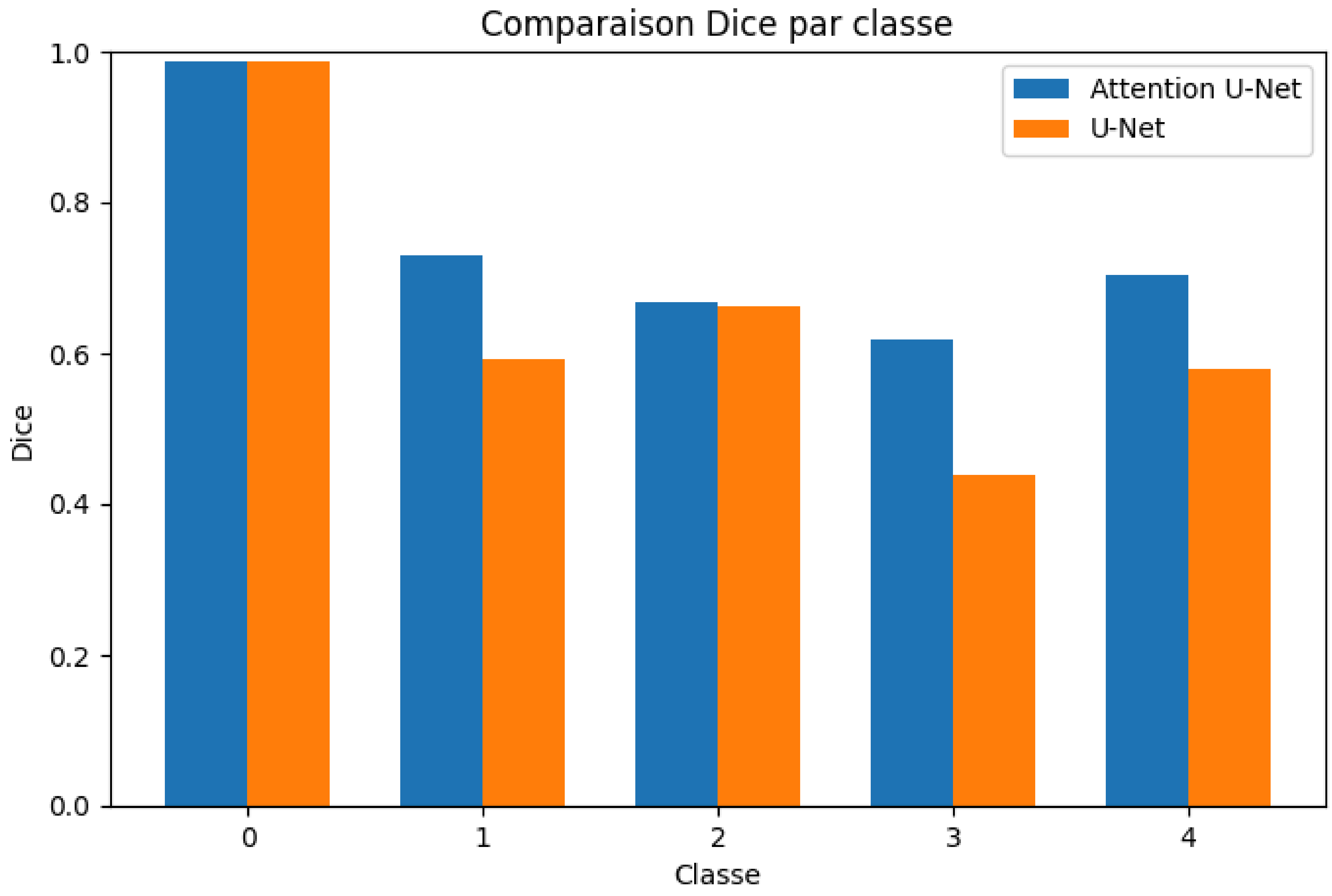




Evaluation metric DICE

to see in **global** (across all pixels of a class) **if the predicted masks are efficient**
BUT
not looking in local regions

Does not capture local boundary errors or small misalignments

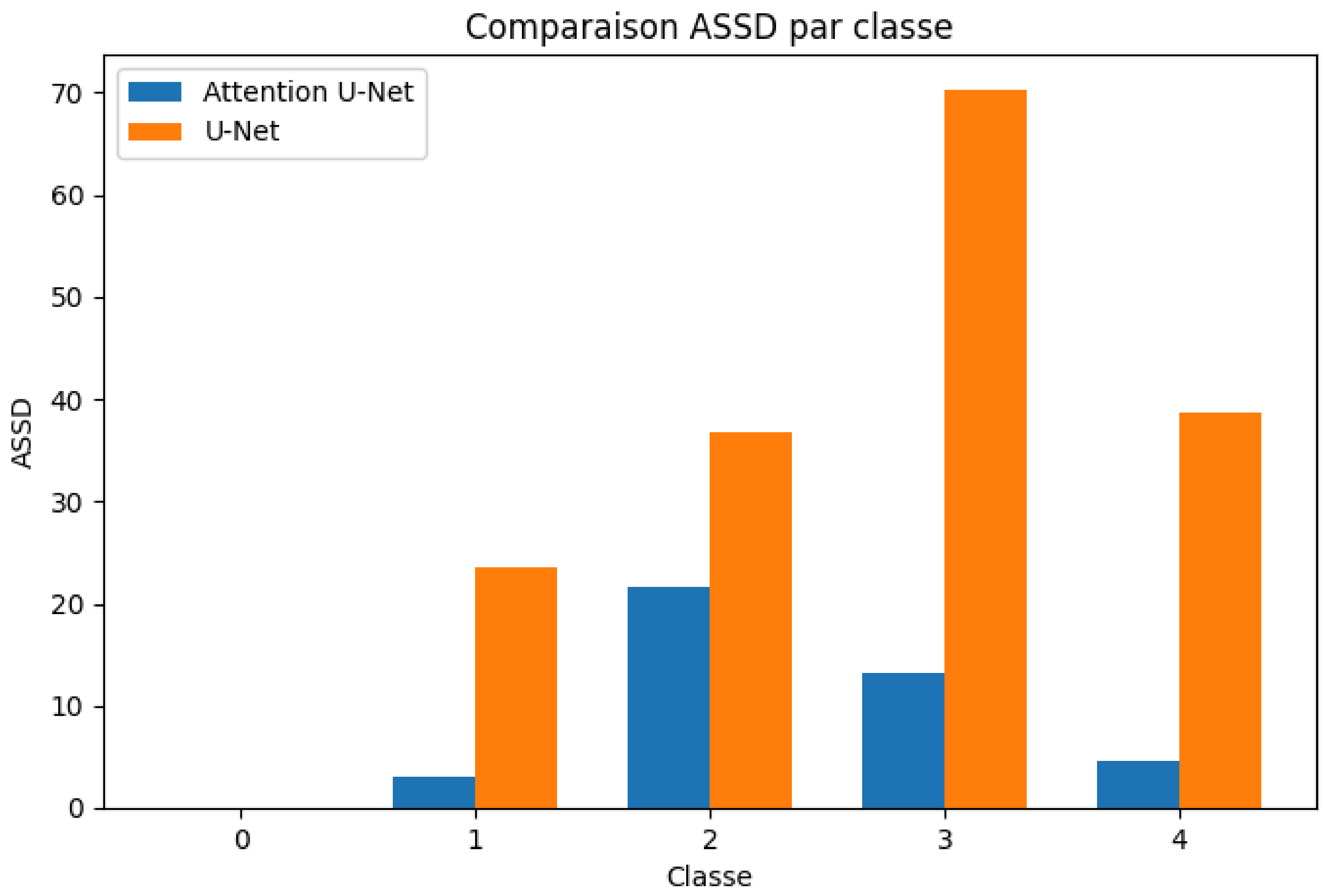


Results

Evaluation metric ASSD

It computes the **average distance between the edges of the ground truth and the predicted mask** so it is looking specifically on the edges **(contours)**
BUT
it is an average

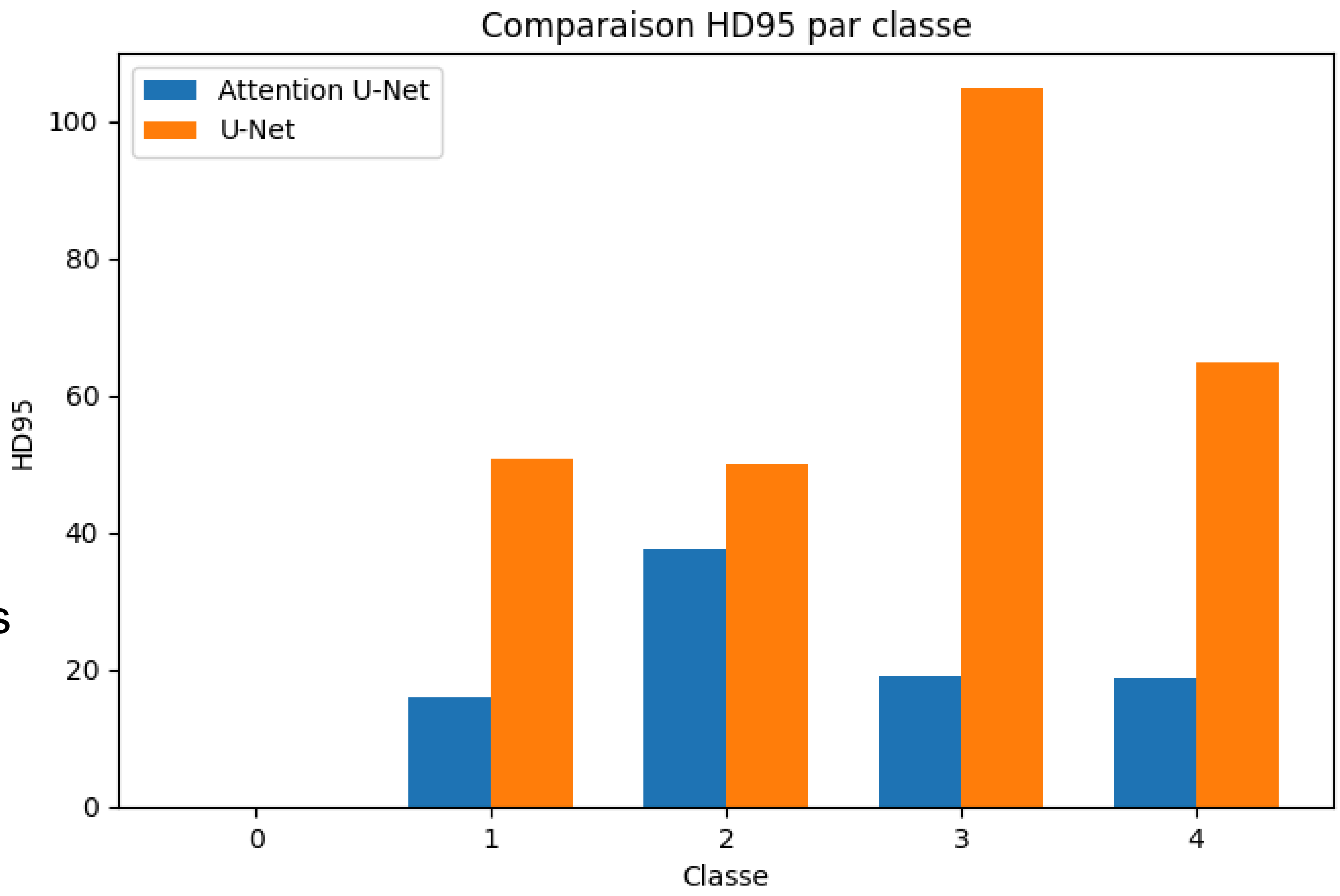
Lower values indicate better alignment



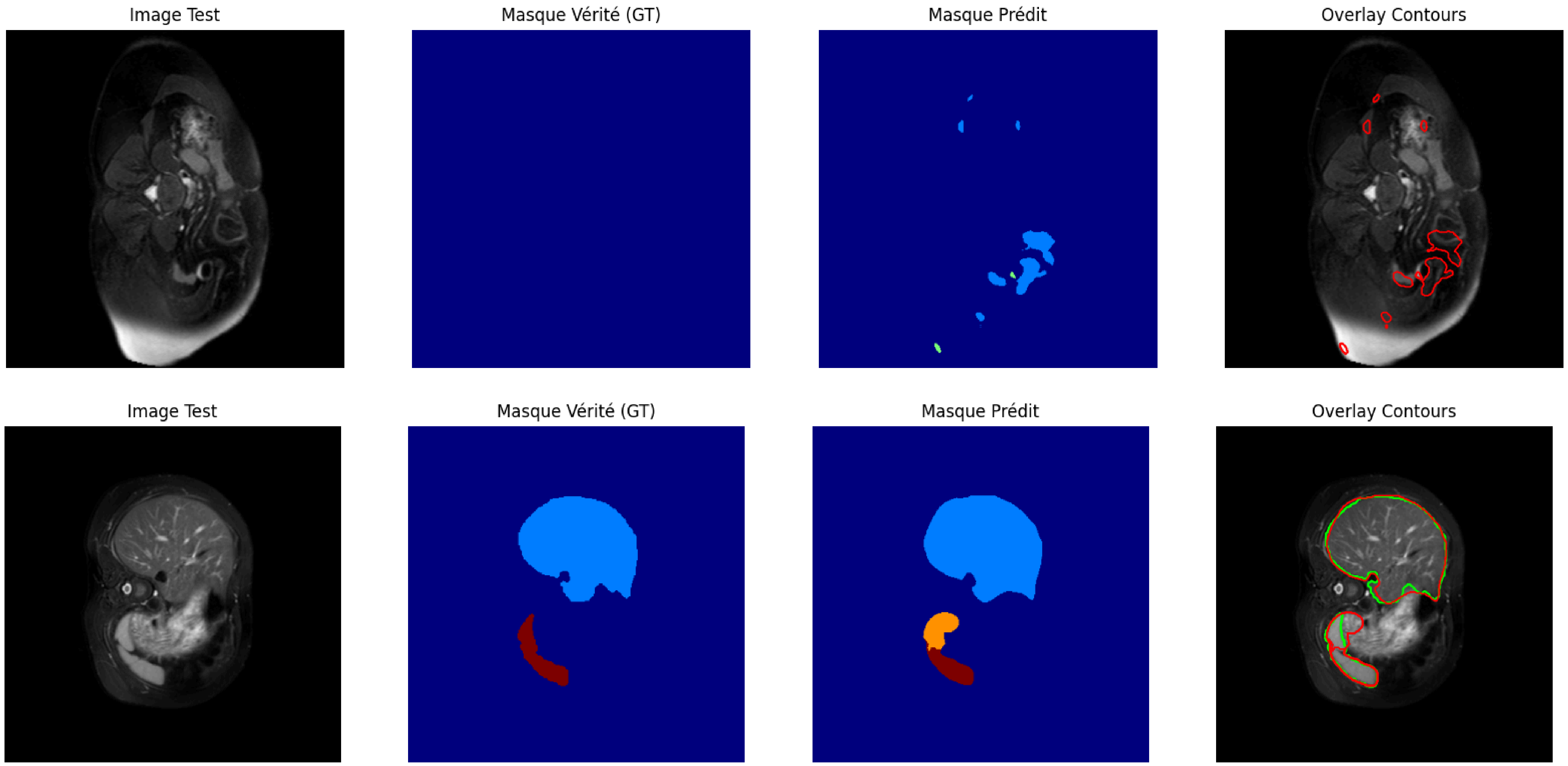
Evaluation metric HD95

It returns the 95th percentile of the **distances between the edges of the ground truth and the predicted mask**
SO
it gives the worst case

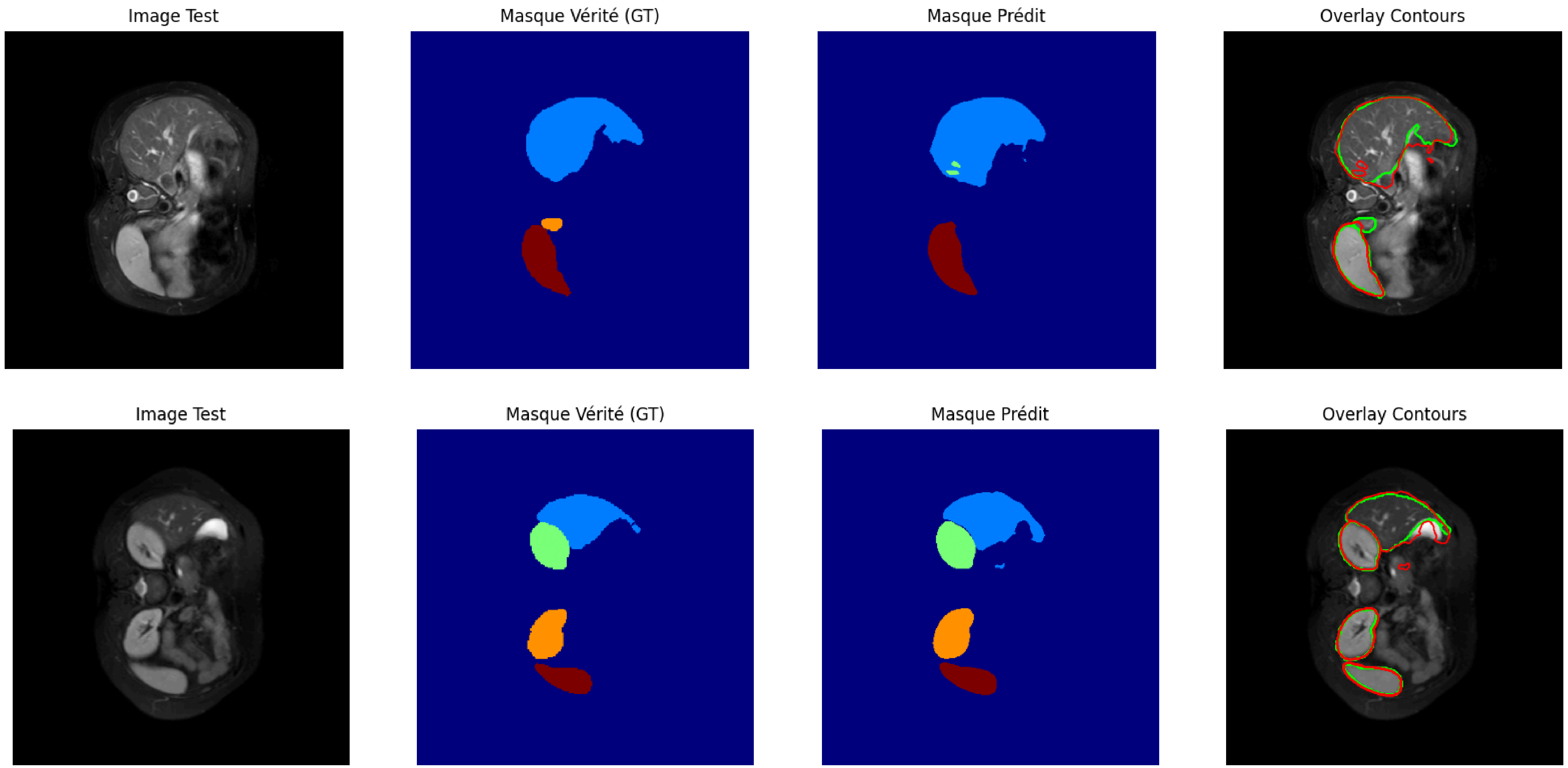
It reduces the influence of small outliers



Predicted mask of the U-Net



Predicted mask of the Attention U-Net



4. Conclusion

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

Faced challenges

- **Computational power and time requirements**

Strategy: using GPU and playing with the parameters.

- **Well organizing implementation details**

Strategy: implementing the model itself was only one part of the work, before and after training, several adjustments were necessary, like taking image and mask together to properly split the dataset, handling the much larger number of background pixels, etc.

- Optimization of the model to get better performances

Strategy: trying different parameters and combinations, such as slice weight, combined loss, reducing batch size, and using different metrics to evaluate.

Proposed improvements

- **Change the complexity of the attention gate ?** as we saw how much it affects the results

- **Change the evaluation metric ?** Use Tversky loss (focuses on FN)

- **Try with a better GPU to increase the batch size ?**

- **Apply data augmentation ?** due to the lack of data

- **Use an unsupervised network to pre-train the encoder ?** because we used only half the dataset

5. References

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

References

O. Oktay et al., “Attention U-Net: Learning Where to Look for the Pancreas”, 20 mai 2018, arXiv: arXiv:1804.03999. doi: 10.48550/arXiv.1804.03999.

O. Ronneberger, P. Fischer, et T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation”, in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, vol. 9351, N. Navab, J. Hornegger, W. M. Wells, et A. F. Frangi, Éd., Cham: Springer International Publishing, 2015, p. 234-241. doi: 10.1007/978-3-319-24574-4_28.

Merci beaucoup pour votre attention !

Avez-vous des questions ?

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