서울여자대학교 소프트웨어융합학과 김세희

Can SegFormer be a True Competitor to U-Net for Medical Image Segmentation?

T.Sourget et al.

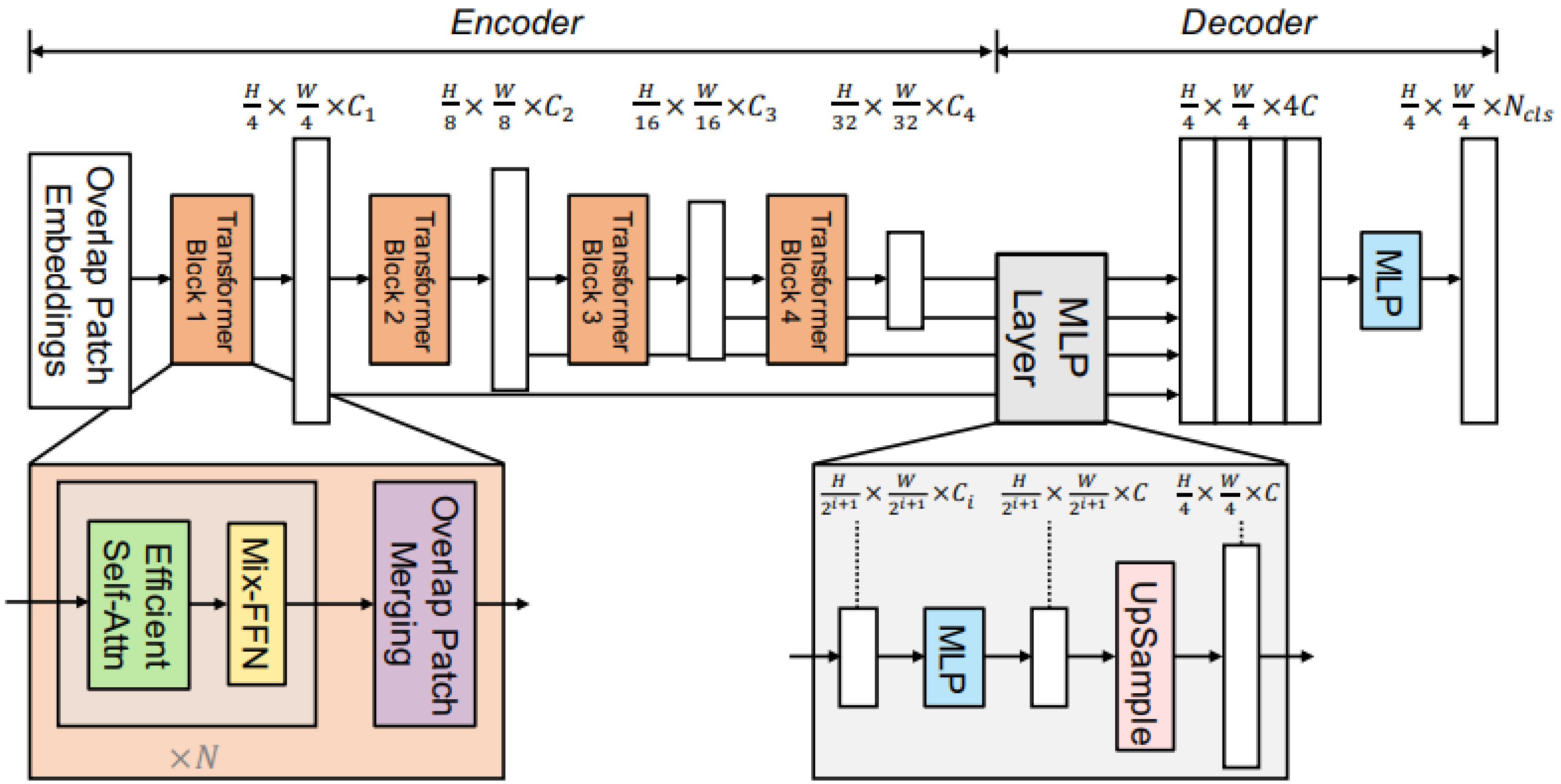
2024.01

1. Introduction

- ✓ Since 2015, the U-Net has been established as the state of the art model for medical image segmentation.
- ✓ Transformer-based architectures have been proposed on image processing tasks, starting with the Vision Transformer.
- ✓ Several hybrid models combining the architectures of CNNs and Transformers have been proposed, but these models are often highly complex and have millions of parameters, requiring a significant amount of time for training. (SETR, PVT)

2. SegFormer

❖ Structure



2. SegFormer

❖ Why did they choose SegFormer?

1. Encoder produces multiple level feature maps and Decoder use it as input. The model is able to capture both high and low resolution information.

2. SegFormer is less complex than other transformer based architectures.
   - ✓ All-MLP Decoder
   - ✓ Mix-FFN

3. SegFormer requires less data to be trained.

4. SegFormer can be applied to real-time application.

3. Experiments

❖ Datasets: CAMUS, Polyp, Instruments

Table 1: Dataset size

Dataset size	Train	Valid	Test
Camus	500	450	50
Polyp	640	160	200
Instrument	377	95	118

✓ CAMUS

- Size: 256 x 256
- Argumentation: random rotation between -10° and 10°

✓ Polyp, Instruments

- Size: 224 x 224
- Argumentation: random rotation between 0° and 180°

3. Experiments

❖ Protocol

Dataset	Model	Optimizer	Loss Function	Learning Rate
CAMUS	U-Net	Adam	Cross Entropy	1e-3
	SegFormer			1e-4
Polyp & Instruments	U-Net	AdamW	average of Cross-Entropy and Dice	1e-4
	SegFormer			1e-4

3. Experiments

❖ Protocol

✓ Models

1. Original U-Net(31M)

2. U-Net Lite(3.7M): Reduced the number of filters from [64,128,256,512,1024] to [22,44,88,176,352]

3. SegFormer(3.7M)

4. Pretrained SegFormer(3.7M): Applied transfer learning by using encoder's weights trained on ImageNet-1k.

4. Results and Discussion

❖ Segmentation accuracy

✓ The average Dice Scores for each dataset

Table 2: Results: Average Dice scores of U-Net and SegFormer of 3 datasets: CAMUS, Polyp and Instrument. * indicates that the score is significantly different from that of UNet ($p<0.05$).For graphical representation see Figure 4

		U-Net	U-Net Lite	SegFormer	SegFormer pre-trained
Pre-trained?		No	No	No	Yes
# param		31M	3.7M	3.7M	3.7M
CAMUS	Endo	0.90	0.90	0.89	**0.91***
	Epi	0.80	0.79	0.81	**0.83***
	Atrium	0.83	0.84	0.81	0.85
Polyp		0.74	0.67	0.60	**0.83***
Instrum		0.79	0.75	0.82	**0.92***

4. Results and Discussion

❖ Segmentation accuracy

✓ The average Dice Scores for each dataset

Title Suppressed Due to Excessive Length 7

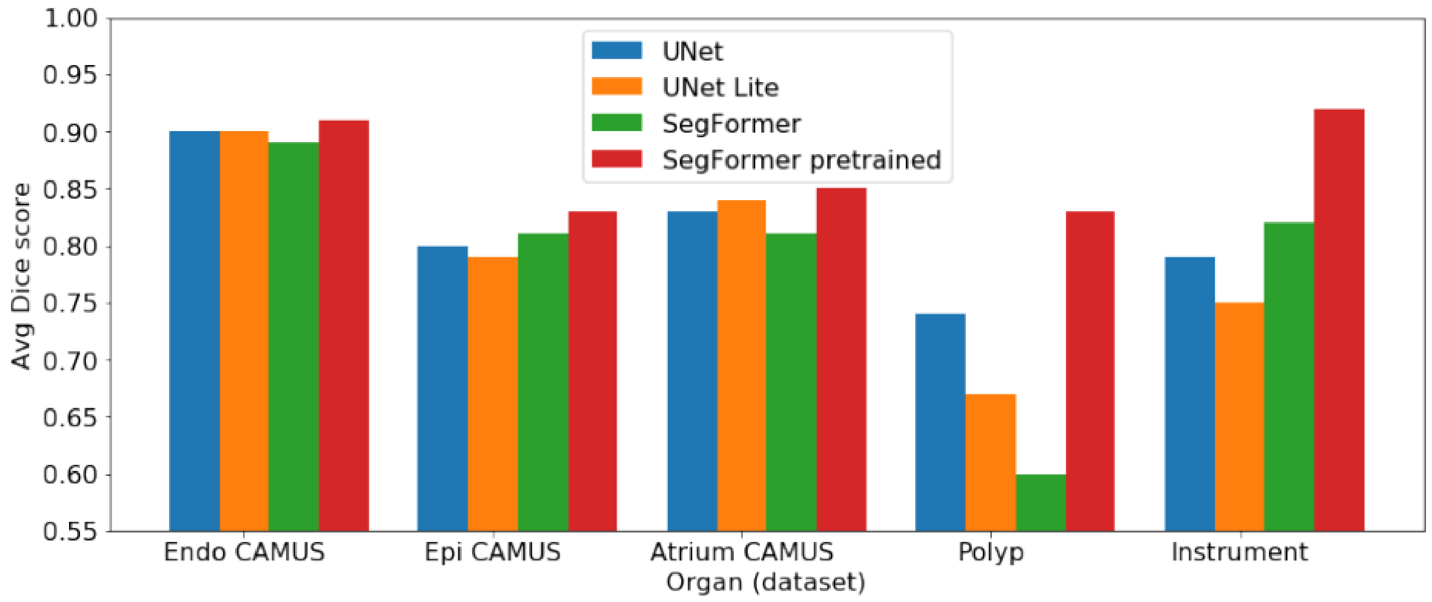
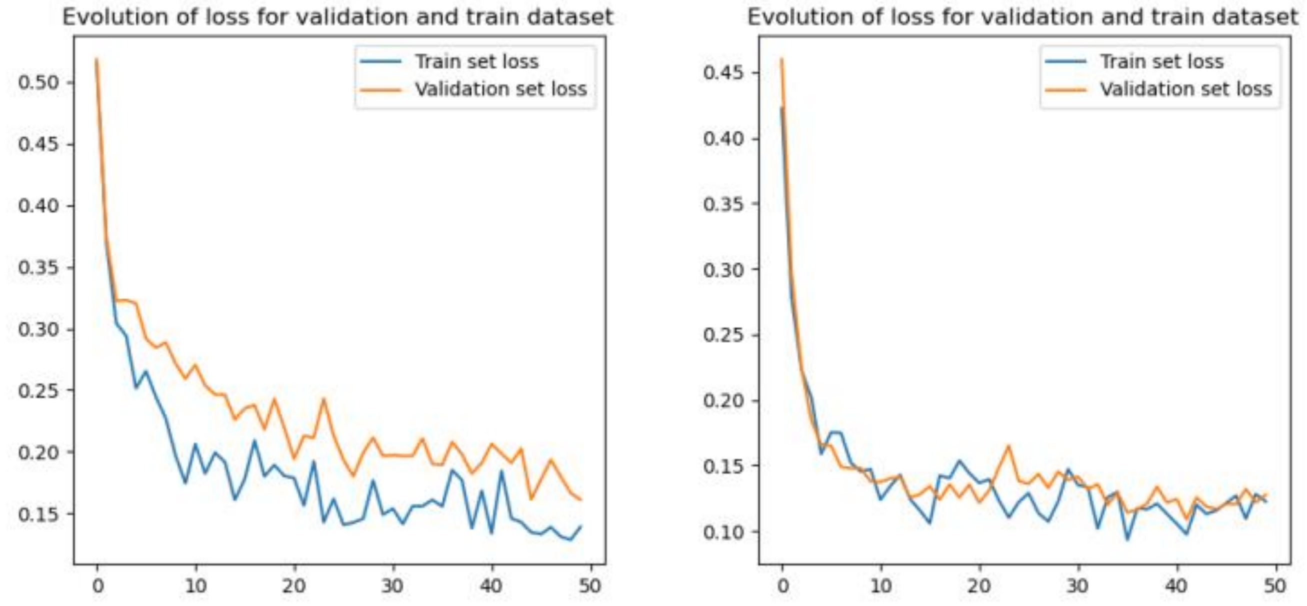


Fig. 4: Average Dice scores of U-Net and SegFormer on test sets of 3 datasets: CAMUS, Polyp and Instrument. Corresponds to results in Table 2.

4. Results and Discussion

❖ Segmentation accuracy

- ✓ Evolution of loss during training for SegFormer



(a) SegFormer (trained from scratch)   (b) pre-trained SegFormer

Fig. 2: Evolution of loss during training for SegFormer with and without pre-training on ImageNet-1K

4. Results and Discussion

❖ Segmentation accuracy

✓ Comparing segmentation results

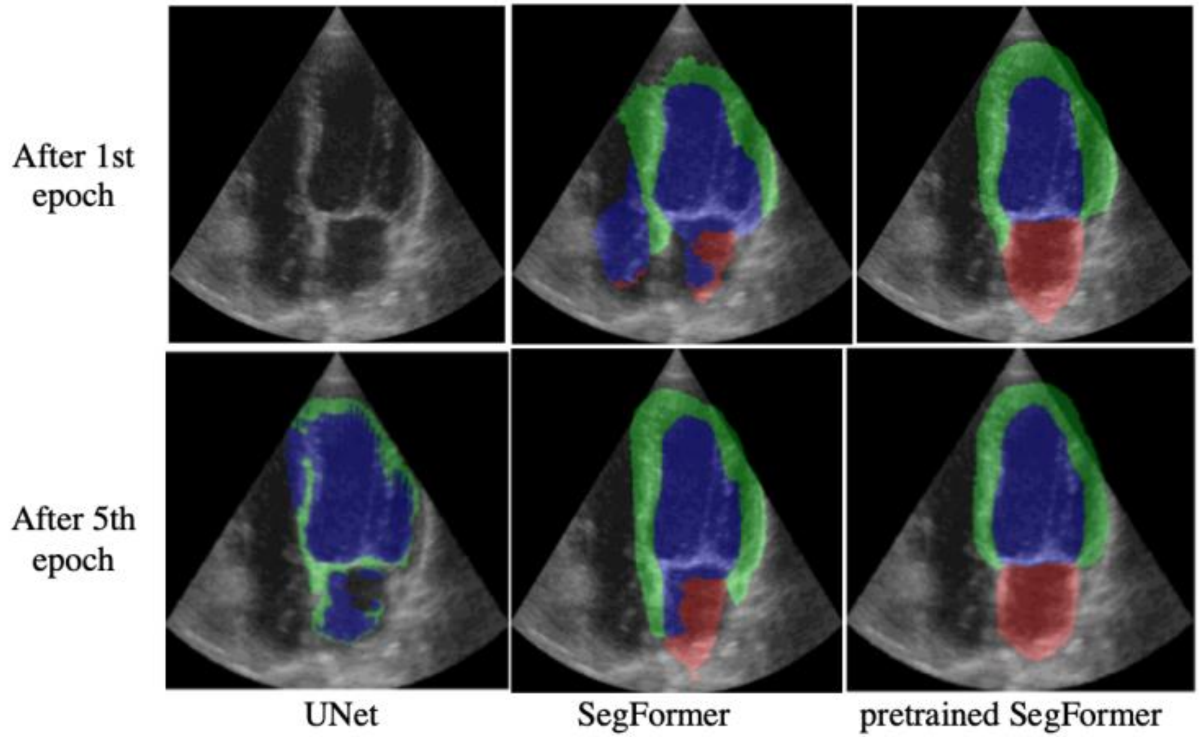


Fig. 3: Comparing segmentation results between U-Net, SegFormer and pre-trained SegFormer after the 1st and 5th epoch

4. Results and Discussion

❖ Training time

Table 3: Drop in training time of U-Net Lite, SegFormer and pre-trained SegFormer with respect to U-Net's training time.

Model	U-Net Lite	SegFormer	SegFormer pre-trained	Epochs
CAMUS	-57.5%	-49.7%	-53.0%	50
Polyp	-51.2%	-62.3%	-65.1%	80
Instrum	-40.4%	-46.4%	-46.9%	80

4. Conclusion

- ✓ On every task, pre-trained SegFormer-B0 obtained on par or better results than U-Net, and in less training time than the original U-Net thanks to its light architecture.

- ✓ We have shown that even if transformers usually need more data to be trained, they can still be applied on medical imaging tasks and obtained better results with limited dataset, taking advantage of transfer learning.