Comparing a New Implementation of SegFormer to UNet for Pathological Scan Segmentation

An Implementation and Evaluation Study within an Applied Digital Pathology AI Platform

BA Defense – 18.08.2023

Director: Prof. Dr. Nassir Navab

Supervisors: Vanessa Gonzalez, Kai Standvoss

Presenter: Carlotta Sophia Hölzle

Technische Universität München + Johns Hopkins Whiting School of Engineering

Collaborators

aignostics

AI-POWERED PRECISION
DIAGNOSTICS FOR PATHOLOGY

Technische Universität München

+

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Motivation

H&E segmentation important for cancer diagnosis, treatment, and research

- Manual segmentation time consuming & highly subjective to annotator style
- CNN-based methods, particular UNet architectures, accurate automated segmentations

  \+ Fine-grained local features

  − Lack global attention

  − Problems with robustness

  − Hit performance roof

Motivation

- Success of Vision Transformer in natural image segmentation task
  - Self-attention to include long-range dependencies
  - Self-attention mechanism to strongen the robustness of models to input perturbation

Hybrid Vision Transformer are based on self-attention and include to some degree convolutional layers

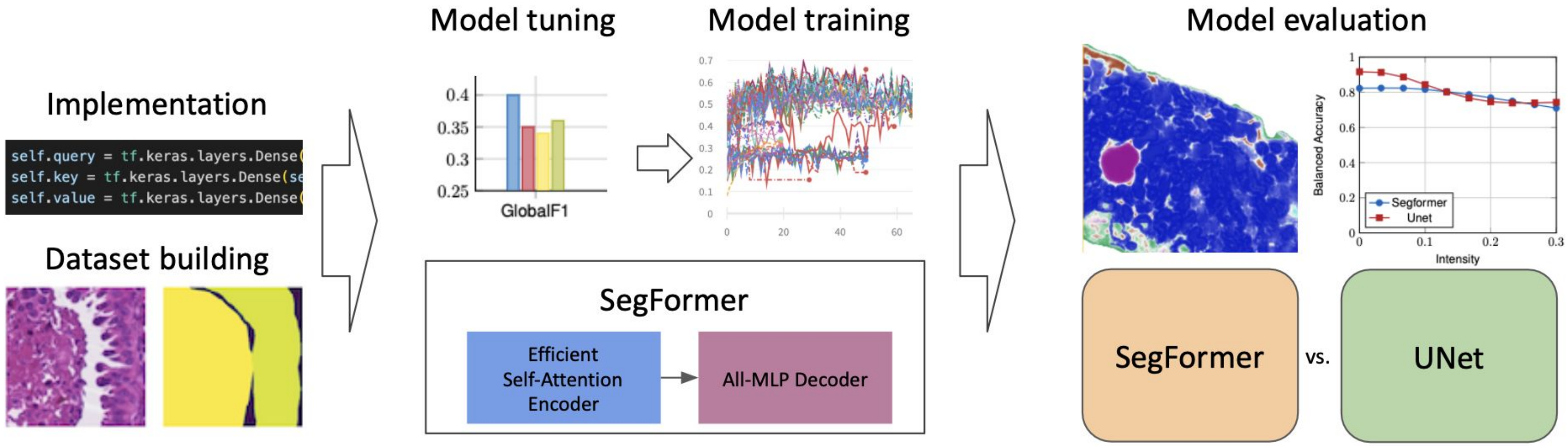
- Complex, computational expensive hybrid implementation promising results for H&E segmentation

Goal

Problem: Need for computational efficient, low-complex Transformer model for H&E segmentation

Research Question: How does SegFormer, a computationally efficient Transformer-based model, requiring no post- and pre-processing, perform within the context of H&E segmentation

Research Design:



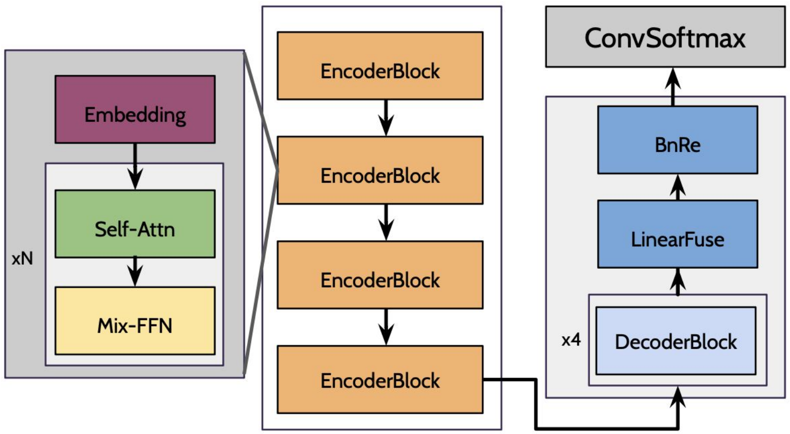
Contributions

1. <u>Integration</u> of the SegFormer architecture, for easy and continuous deployment for new H&E segmentation tasks
2. Determining optimal architecture and ML <u>parameters for best performance</u>
3. <u>Quantitative and Qualitative performance</u> comparison of SegFormer and a reference UNet
4. Investigation into the impact of various <u>input augmentation</u> and perturbation techniques on performance
5. Assessment of <u>inter-annotator variability</u> towards the segmentation performance and generalizability.
6. Consideration of the <u>applicability</u> of SegFormer for H&E segmentation in <u>applied pathology</u>, including economic implications and overall segmentation quality

[1] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "Segformer: Simple and efficient design for semantic segmentation with transformers," *Advances in Neural Information Processing Systems*, vol. 34, pp. 12 077–12 090, 2021.

Motivation - SegFormer[1]

+ Global attention:
  - receptive field not upper bound by the last stage receptive field as in CNNs
+ Fusion of global and local features:
  - hierarchical attention Encoder
  - MixFFN
+ No Positional Encoding (PE):
  - adaptability to diverse input shapes without compromising performance
+ Reduction of computational complexity:
  - reduction of Key and Value vector in self-attention calculation
  - lightweight all-MLP Decoder



|----- EncoderBlock-----
| Embedding -> N x Layer -> LayerNorm

|----- Layer -----
| LayerNorm -> Attention -> LayerNorm -> MixFFN

|----- Attention -----
| SelfAttention -> Dense

|----- MixFFN -----
| Dense -> Conv2D_{3x3} -> GeLU -> Dense

|----- DecoderBlock -----
| Dense -> UpSampling

Parameter Testing & Quantitative Performance Evaluation

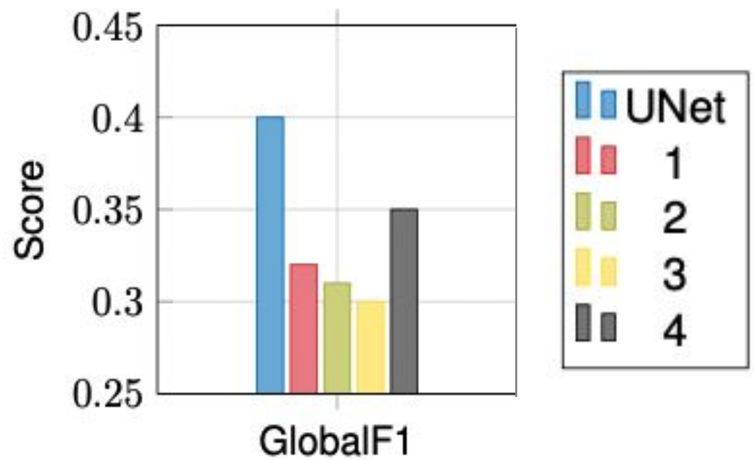


Figure 5.1 Compared Performance of Tested SegFormer-b0 Parameters

Run	lr	wd	λ	Global Macro F1
1	1e-04	1e-06	0.9	0.32
2	1e-06	1e-04	0.1	0.31
3	1e-04	1e-06	0.9	0.30
4	5e-04	1e-06	0.9	0.35

Table 5.1 Impact of Learning Rate, Weight Decay, λ of Combined Loss on Global Macro F1 of SegFormer-b0

$$\textit{Combined Loss Function} = \lambda * \textit{CrossEntropy} + (1-\lambda) * \textit{Dice}$$

Parameter Testing & Quantitative Performance Evaluation

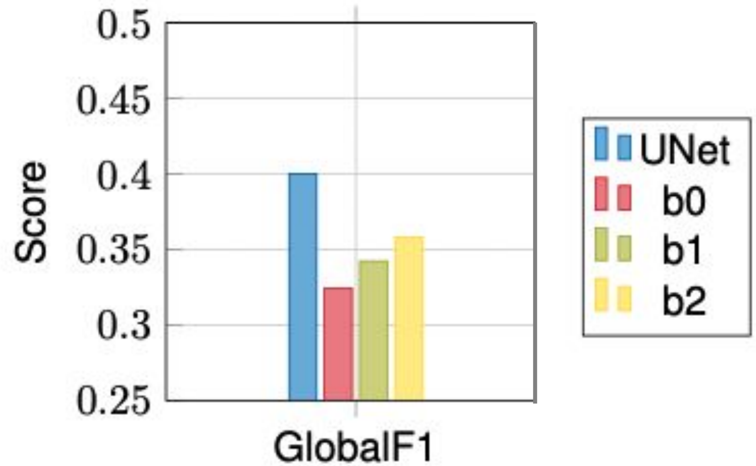


Figure 5.2 Impact of SegFormer Network Size on Performance

Variant	Layers per Encoder Block	Encoder Block Dimension	Decoder Dimension	Nr. Parameters
b0	[2, 2, 2, 2]	[32, 64, 160, 256]	256	3.7 million
b1	[2, 2, 2, 2]	[64, 128, 320, 512]	256	14.0 million
b2	[3, 4, 6, 3]	[64, 128, 320, 512]	768	25.4 million
b3	[3, 4, 18, 3]	[64, 128, 320, 512]	768	45.2 million
b4	[3, 8, 27, 3]	[64, 128, 320, 512]	768	62.6 million
b5	[3, 6, 40, 3]	[64, 128, 320, 512]	768	82.0 million

Table 3.1 Parameters for SegFormer Network Variants

Robustness Analysis

Input Augmentations Tested:

1. Brightness
2. Contrast
3. Hue
4. Gaussian Noise & Blur
5. H&E specific input perturbations
    1. H Noise & E Noise
    2. Increase & Decrease H & E Channel

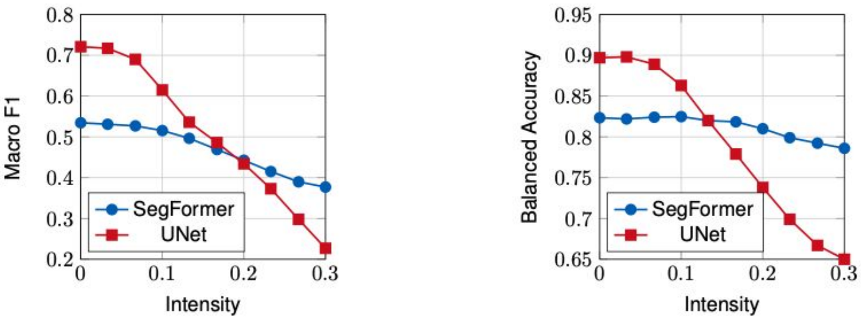


Figure 5.15 Impact of Hematoxylin Noise on Performance of SegFormer and UNet

Tested Robustness of 2 SegFormer-b1 with different λ values

Findings:

- Generally SegFormer higher robustness than UNet
- Constant higher robustness with H&E specific augmentation
- λ in loss function low impact on robustness
    - λ = 0.1 marginally more robust

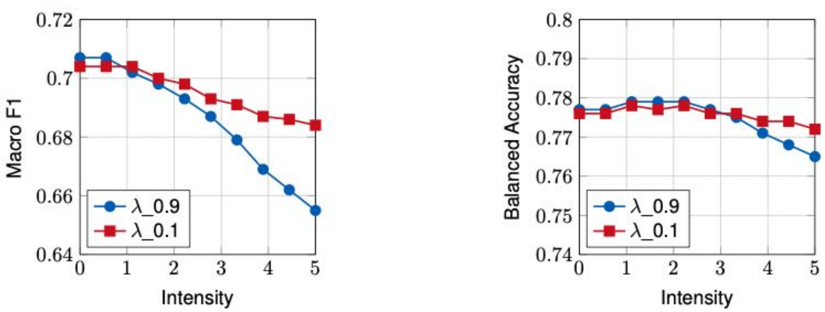


Figure 5.25 Impact of Gaussian Blur on SegFormer Performance

Qualitative Analysis

Results:

1. Patching Characteristics
2. Inconsistent predictions across boundaries

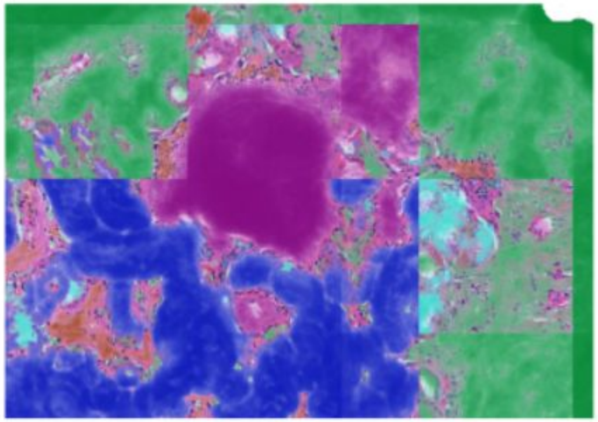


Figure 5.29 Patching Characteristics in Prediction

Explanation:

- Pipeline optimized for UNet -> Overlapping patch merging
- High share of Cross Entropy loss
- Never sees the full slide

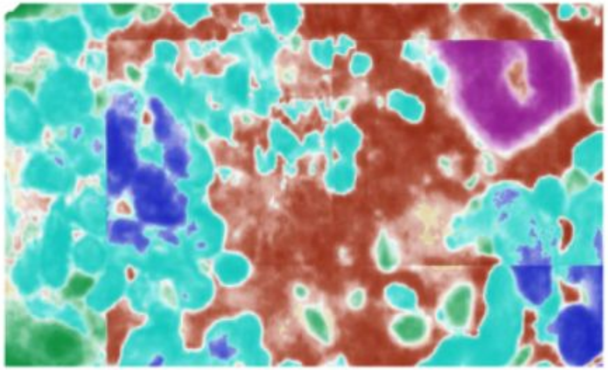
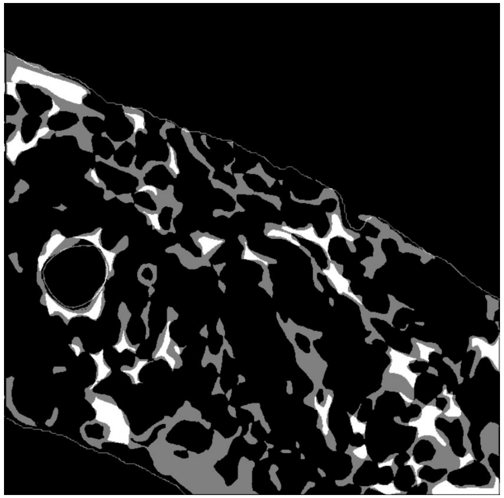
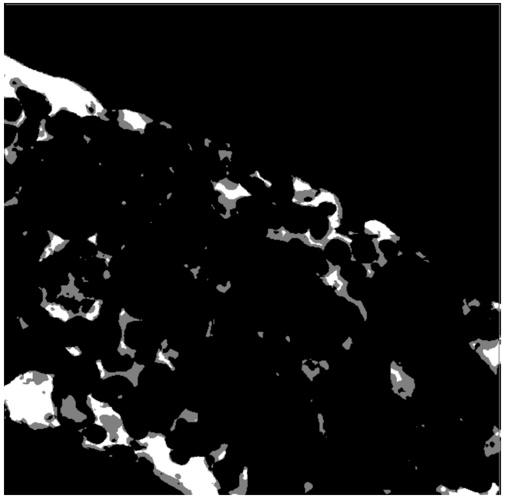


Figure 5.30 Inconsistent Segmentation Prediction Across Patch Boundaries

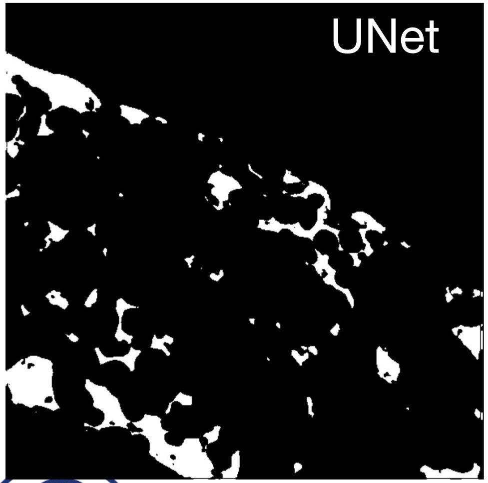
Inter-Annotator Variability

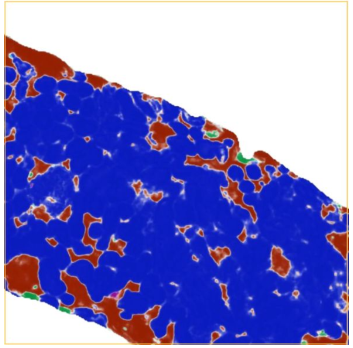


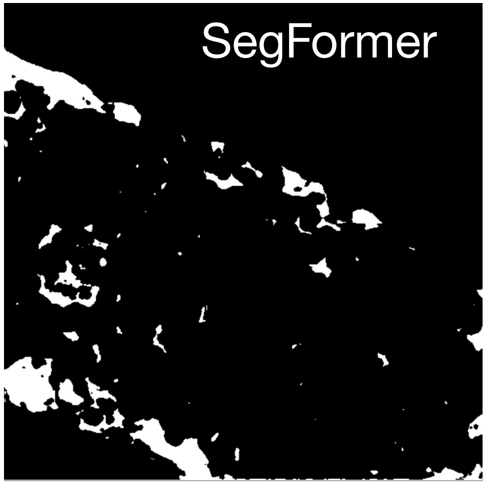
Annotations:
White = as fibrosis label in dataset
Black = as non-fibrosis label in dataset
Gray = as fibrosis and non-fibrosis in dataset

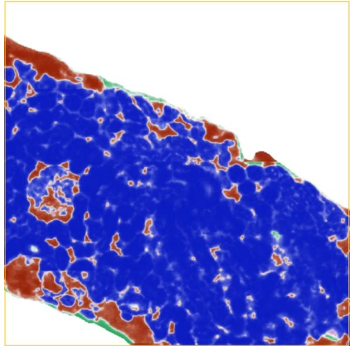


Predictions:
Black = agree on non-fibrosis
White = agree on fibrosis
Gray = disagree on fibrosis









Conclusion

- Implementation of novel architecture into applied pathological pipeline
  - Some necessary adjustments & optimization necessary
- Similar performance larger datasets
- Difficulties to match performance on smaller datasets
  - Known issue with Transformer models
- Qualitative analysis showed struggles with complex features
- Similar quantitative & qualitative performance on inter-annotator variability experiments
- Analysis of biases from single annotation style inconclusive
  - Both architectures learned better of a certain annotator style
- SegFormer high resilience to input perturbations

Future work

- More ground truth labels for extensive comparison

- Training on whole slide H&E images to better capitalize on self-attention mechanism

- More in-depth research into computationally efficient and low complexity Transformer-based methods for automated histophathological segmentation

Thank you!

Technische Universität München

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Backup slides

Motivation

Success of Vision Transformers in natural semantic image segmentation

- self-attention mechanism allows models to selectively focus on different parts of the input
  - relevant for H&E segmentation: information spread out across the input

Hybrid Vision Transformers to include convolutional layers:

- H&E slides have fine-grained & complex features
  - local attention needed

Parameter Testing & Quantitative Performance Evaluation

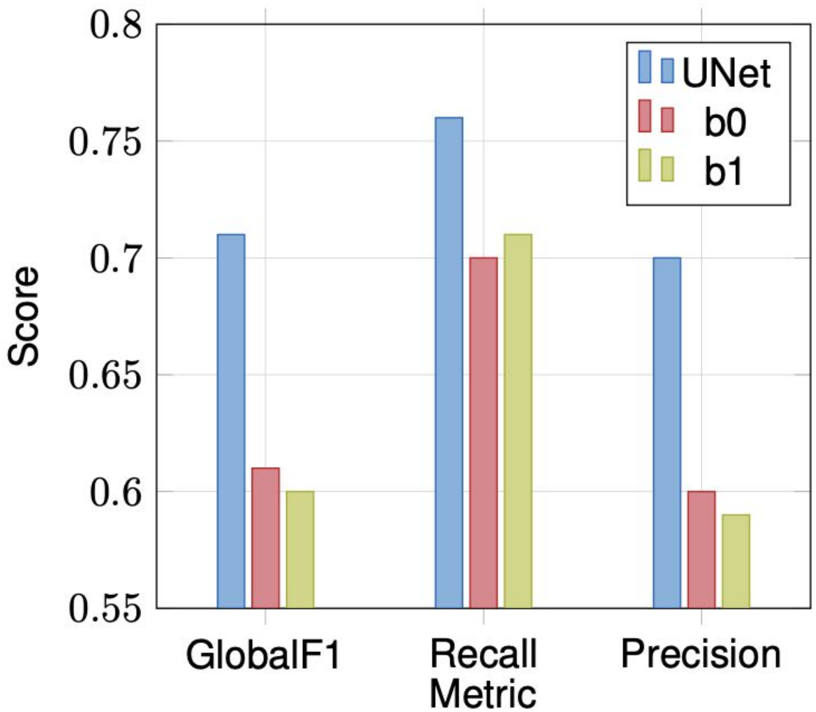


Figure 5.7 Influence of Network Size with Dataset 4 on Performance

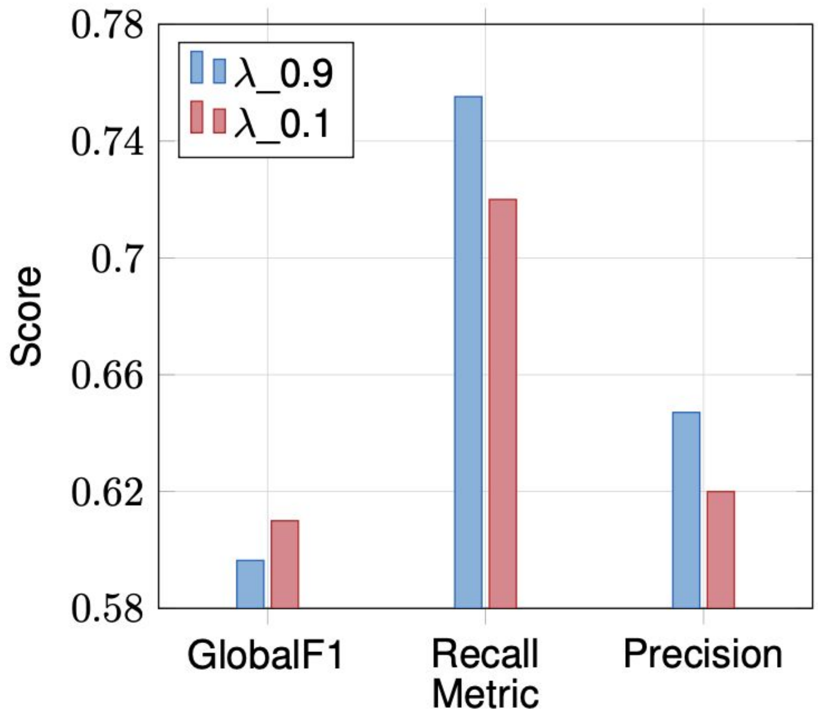
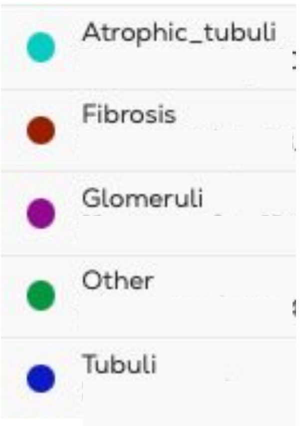
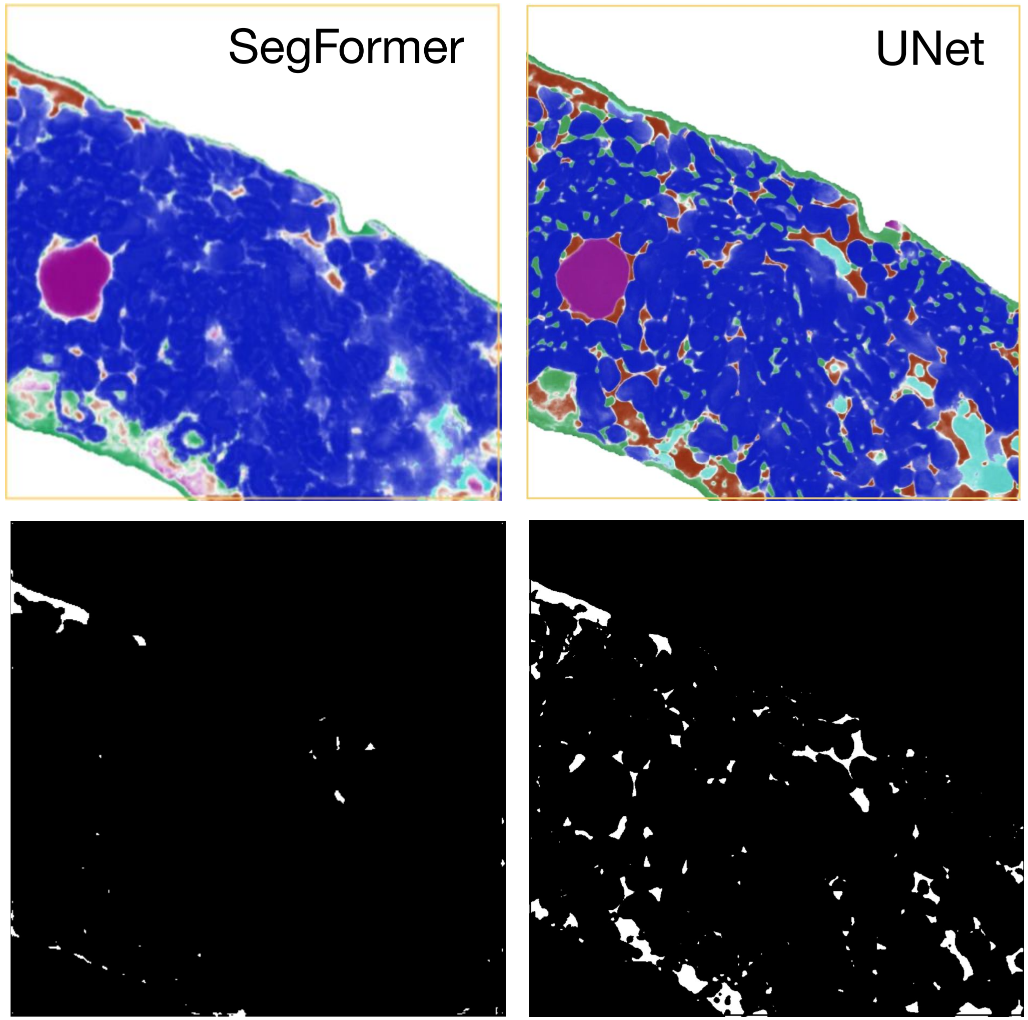


Figure 5.8 Influence of λ in the Combined Loss with Dataset 4 on Performance of SegFormer-b1

Qualitative Analysis

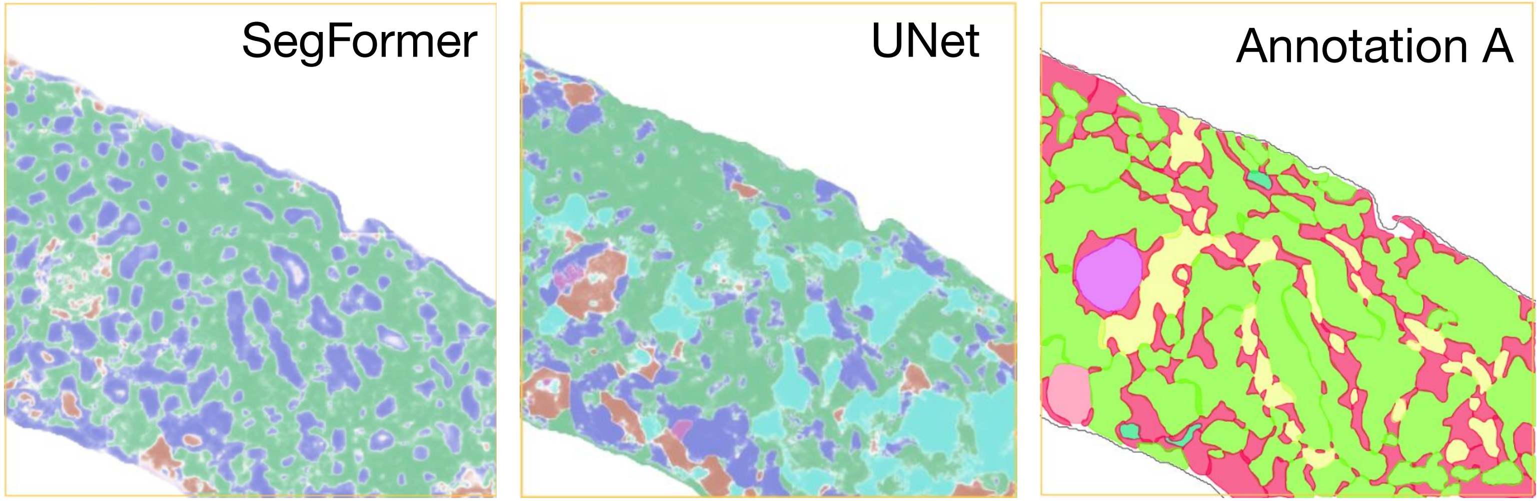




Absence of details in SegFormer prediction, favoring broader contextual data over complex specifics

Predictions:
White = model predicts fibrosis
Black = models predicts non-fibrosis

Single-Style Annotations



Analysis of the Applicability to an Industrial Setting

- comparable resource demands for the two smaller network sizes
- suitability for H&E segmentation in industrial setting is debatable
  - considerable demands for training data
  - lower performance
- uncertain regions & high robustness require deeper consideration to leverage their potential benefits