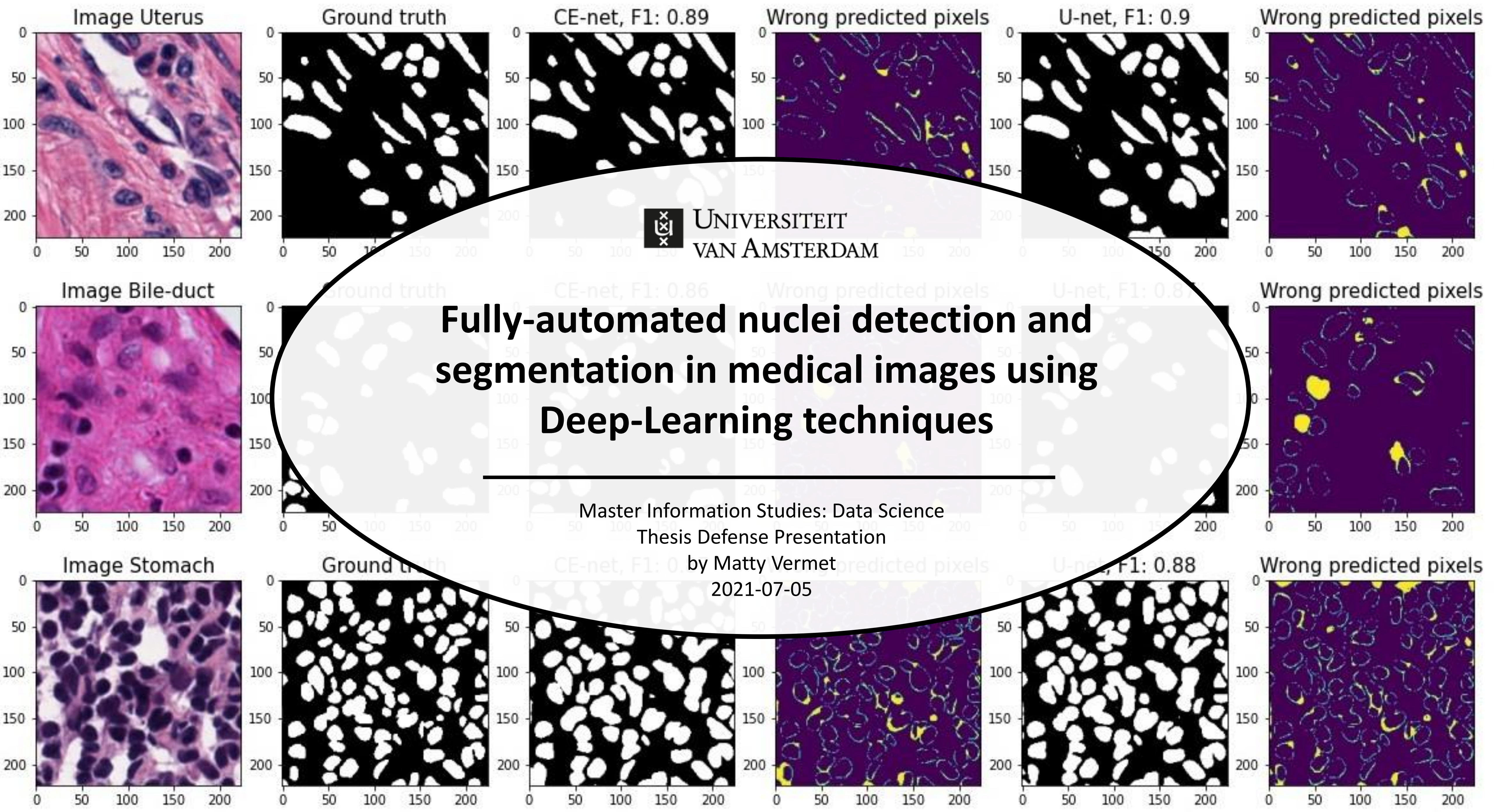


UNIVERSITEIT VAN AMSTERDAM

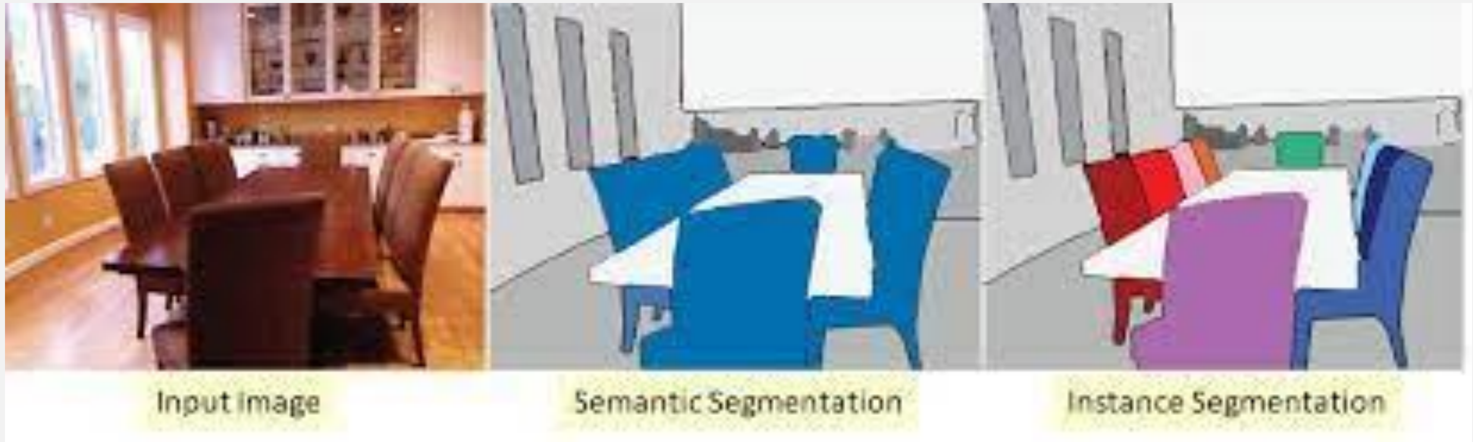
Fully-automated nuclei detection and segmentation in medical images using Deep-Learning techniques

Master Information Studies: Data Science
Thesis Defense Presentation
by Matty Vermet
2021-07-05

Outline

- Introduction
- Methods
- Results
- Discussion
- Conclusion

Introduction

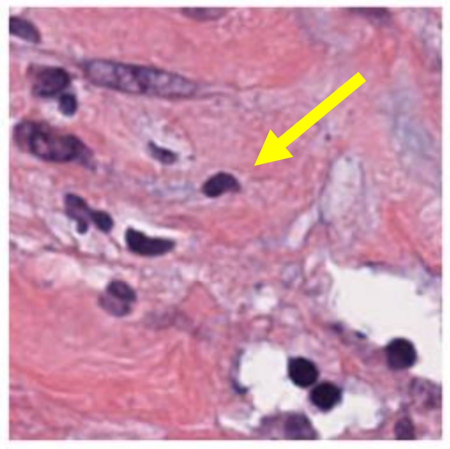


- **Image segmentation:** a digital image is partitioned into multiple segments that capture particular objects in a set of pixels

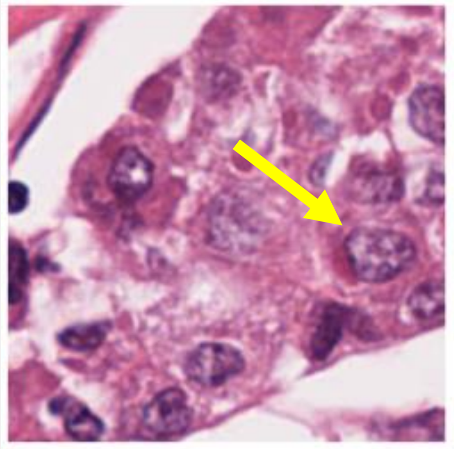
- **A nucleus (NL *celkern*)** is the core of a cell which carries the DNA that contains genetic information about the organism.
- **Why is it important?**
  Cell-type classification (e.g. cancer cells)
  Distribution of certain cell-types
  Phenotype analysis
- Implications and difficulties: different images, shapes, colors, etc. Generalization problem

Introduction

Testis tissue image



Liver tissue image



- **Image segmentation:** a digital image is partitioned into multiple segments that capture particular objects in a set of pixels

➢ **A nucleus (NL *celkern*)** is the core of a cell which carries the DNA that contains genetic information about the organism.

➢ **Why is it important?**
Cell-type classification (e.g. cancer cells)
Distribution of certain cell-types
Phenotype analysis

- **Implications and difficulties:** different images, shapes, colors, etc. Generalization problem

Introduction

- **Image segmentation:** a digital image is partitioned into multiple segments that capture particular objects in a set of pixels
- **A nucleus (NL *celkern*)** is the core of a cell which carries the DNA that contains genetic information about the organism.
- **Why is it important?**
  Cell-type classification (e.g. cancer cells)
  Distribution of certain cell-types
  Phenotype analysis

➢ **Implications and difficulties**: different images, shapes, colors, etc. Generalization problem

Research Questions

How does CE-Net perform in detection and segmentation of nuclei compared to the medical baseline U-Net?

1. How sensitive are the models to outliers and noise?
2. Are the F1-scores or training times of the models influenced by a decrease in training samples?
3. How robust are the predictions of both models?
4. How is CE-Net performing in the segmentation of nuclei compared to U-Net looking at the F1-score and the Jaccard index scores?
5. How accurate is CE-Net in detecting the nuclei compared to U-Net looking at the Cell-counting accuracy metric?
6. Are the results on the PanNuke dataset of this research similar to the results of Gamper et al. (2020), when comparing the Panoptic Quality metric?

Methods

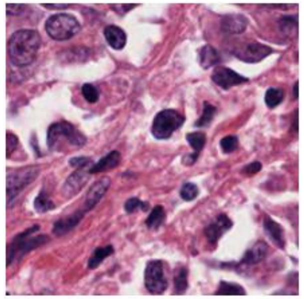
Data

PanNuke dataset by Gamper et al (2019)

- Combination of four different datasets
- > 7000 images + corresponding masks
- 19 different tissue types
- 5 different nuclei types: Neoplastic, Non-Neo Epithelial, Inflammatory, Connective, and Dead nuclei

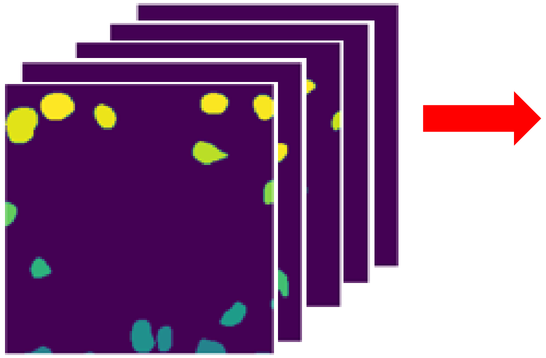
Image

256 X 256

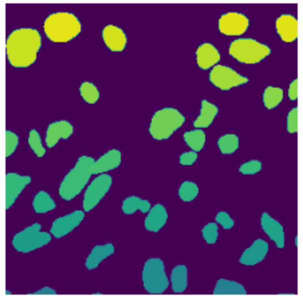


Masks

256 X 256 X 6



256 X 256



Methods

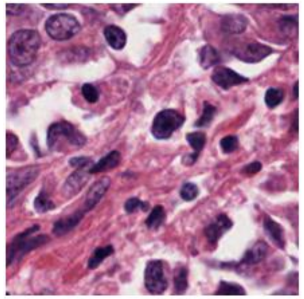
Data

PanNuke dataset by Gamper et al (2019)

- Combination of four different datasets
- > 7000 images + corresponding masks
- 19 different tissue types
- 5 different nuclei types: Neoplastic, Non-Neo Epithelial, Inflammatory, Connective, and Dead nuclei

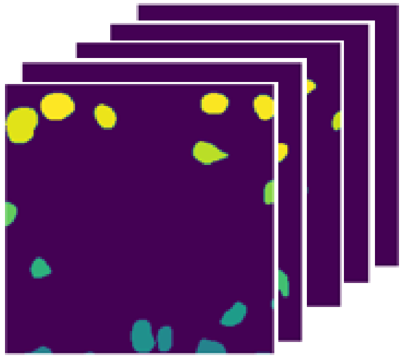
Image

256 X 256

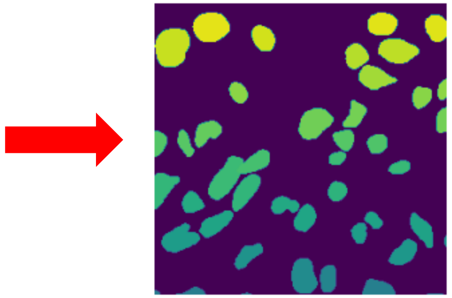


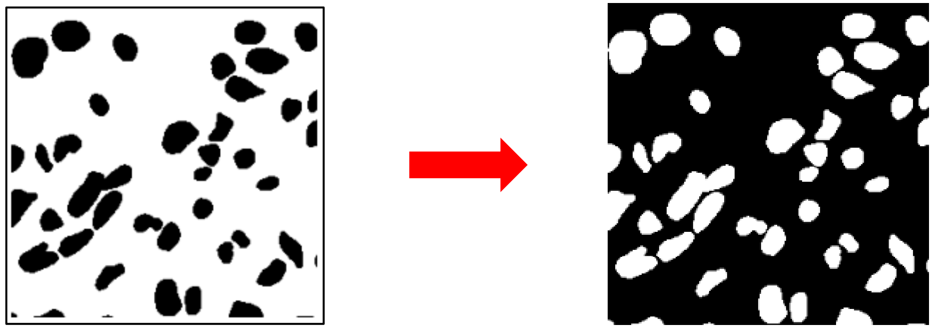
Masks

256 X 256 X 6



256 X 256

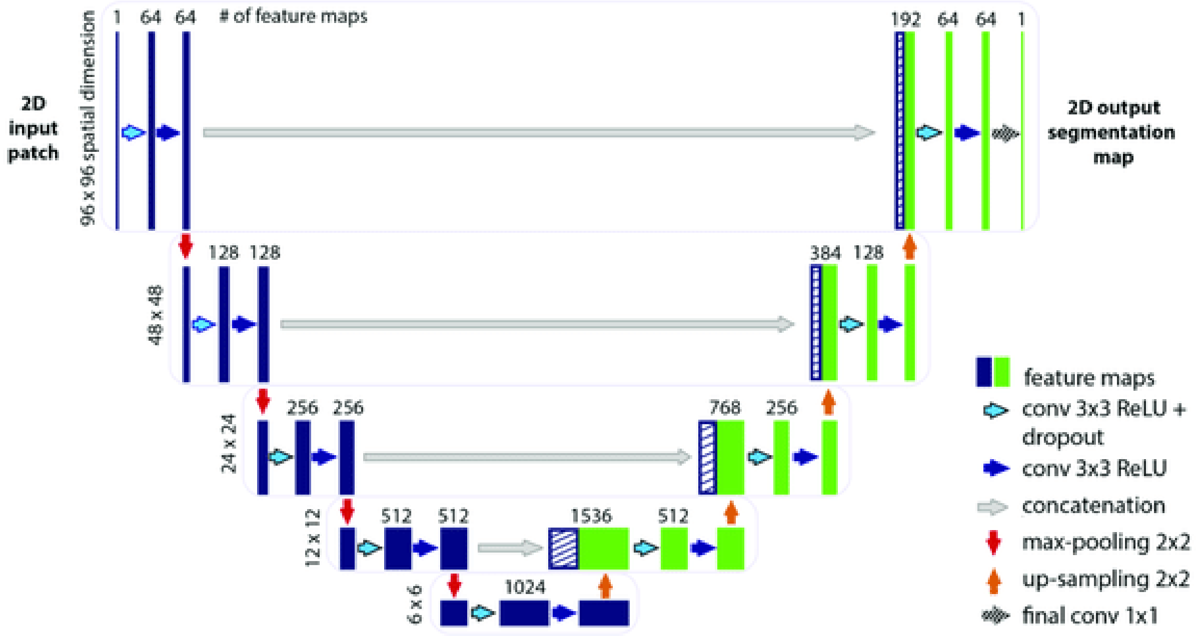




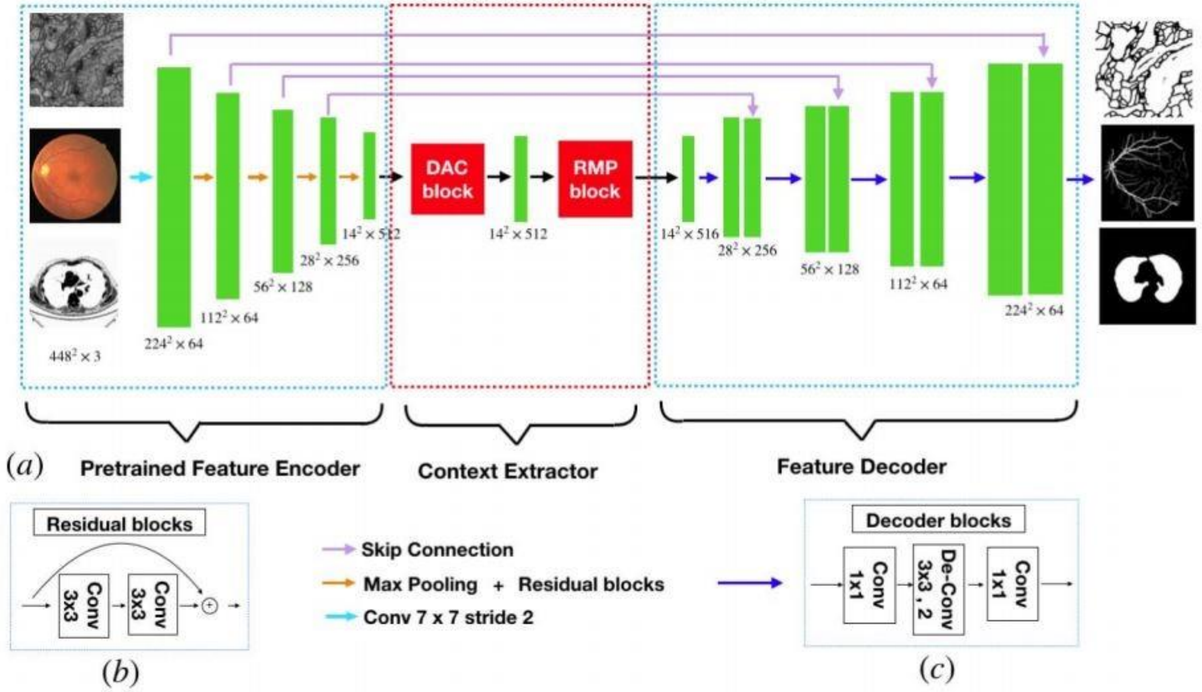
Methods

Models

U-Net



CE-Net



Methods

Evaluation Metrics

Pixel-level:

1. **F1 Score** is a measure of test's accuracy. And is the harmonic mean of the precision and recall.
2. Jaccard index

Cell-level:

1. Cell-counting Accuracy
2. Panoptic Quality

TP = true positives
FP = false positives
FN = false negatives

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

$$F1 = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \qquad (1)$$

Methods

Evaluation Metrics

Pixel-level:

1. **F1 Score** is a measure of test's accuracy. And is the harmonic mean of the precision and recall.
2. **Jaccard index:** is a ratio of the number of matching pixels to the total number of both matching pixels and mismatching pixels.

Cell-level:

1. Cell-counting Accuracy
2. Panoptic Quality

TP = true positives
FP = false positives
FN = false negatives

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

$$F1 = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \tag{1}$$

$$Jaccard\ Index = \frac{TP}{TP + FP + FN} \tag{2}$$

Methods

Evaluation Metrics

Pixel-level:

1. F1 Score
2. Jaccard index

Cell-level:

1. **Cell-counting Accuracy** calculates the difference between the number of nuclei in the predicted mask and ground truth mask.
2. Panoptic Quality

$$Nuclei\ counting\ Accuracy = 1 - MAE$$

$$MAE(nr\ of\ nuclei) = \frac{\sum_{i=1}^{n} \frac{|G_i - P_i|}{G_i}}{n} \qquad (3)$$

G = ground truth
P = predicted

Methods

Evaluation Metrics

pixel level:

1. F1 Score
2. Jaccard index (IoU)

Cell-level:

1. Cell-counting Accuracy
2. **Panoptic Quality** is a combination of the detection quality (F1-score) and the segmentation quality. The TP, FP, FN of segments (so nuclei) instead of pixels.

TP = if IoU of nucleus > 0.5
FP or FN = if IoU of nucleus < 0.5

$$PQ = \underbrace{\frac{\sum_{(p,g)\epsilon TP} \mathrm{IoU}(p,g)}{\cancel{TP}}}_{\text{Segmentation Quality(SQ)}} \times \underbrace{\frac{\cancel{TP}}{TP + \frac{1}{2}FP + \frac{1}{2}FN}}_{\text{Detection Quality(DQ)}} \qquad (4)$$

$$F1 = \frac{TP}{TP + \frac{1}{2}(FP + FN)}$$

Example in an image:
3 TP nuclei (IoU 0.6)
1 FP or FN nuclei

SQ = (0.6+0.6+0.6) / 3 = 0.6
DQ = 3 / 3+(0.5x1)+(0.5x0) = 0.86

PQ = 0.6 x 0.86 = **0.52**

Results

1. How sensitive are the models to outliers and noise?

Table 2. The table shows the sensitivity to outliers and noise of U-Net model and CE-Net model.

Removal of:	CE-Net		U-Net	
	F1-score	Jaccard	F1-score	Jaccard
	test set		test set	
-	0.63	0.49	0.66	0.52
Noise	0.73	0.60	0.77	0.64
Outliers area	0.62	0.48	0.66	0.52
Outliers PCA	0.70	0.58	0.74	0.62
Noise + all outliers	**0.74**	**0.61**	**0.79**	**0.67**

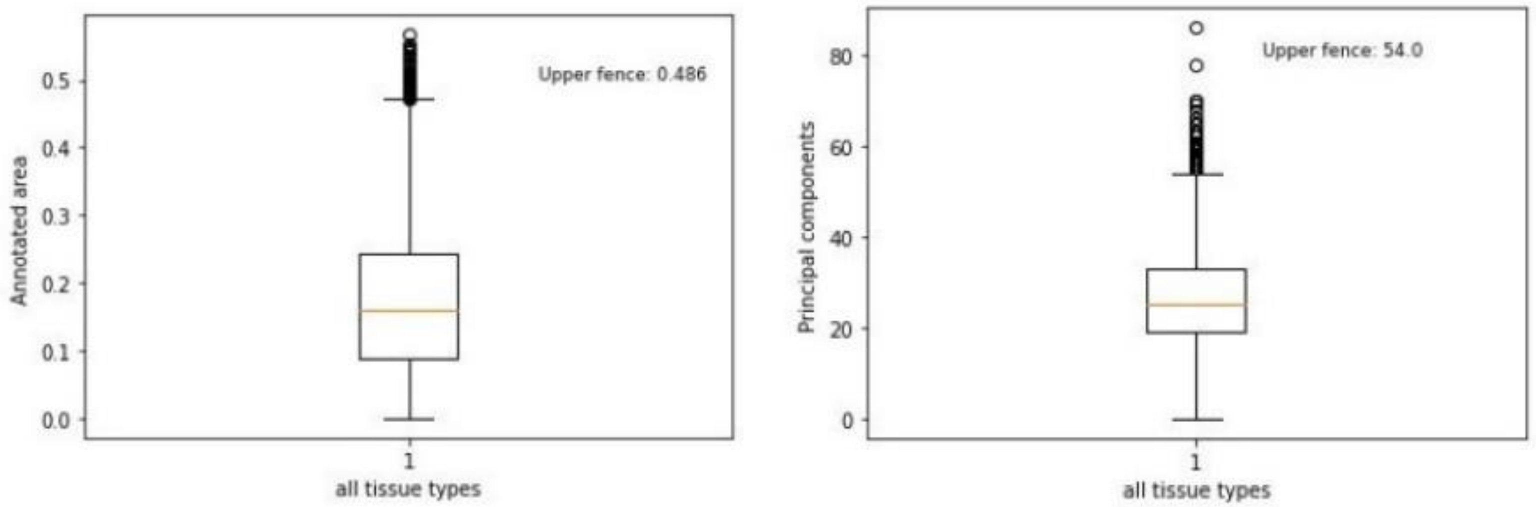


Figure 6. Outlier detection based on the total annotated area of the images and the principal component analysis.

Results

2. Are the F1-scores or training times of the models influenced by a decrease in training samples?

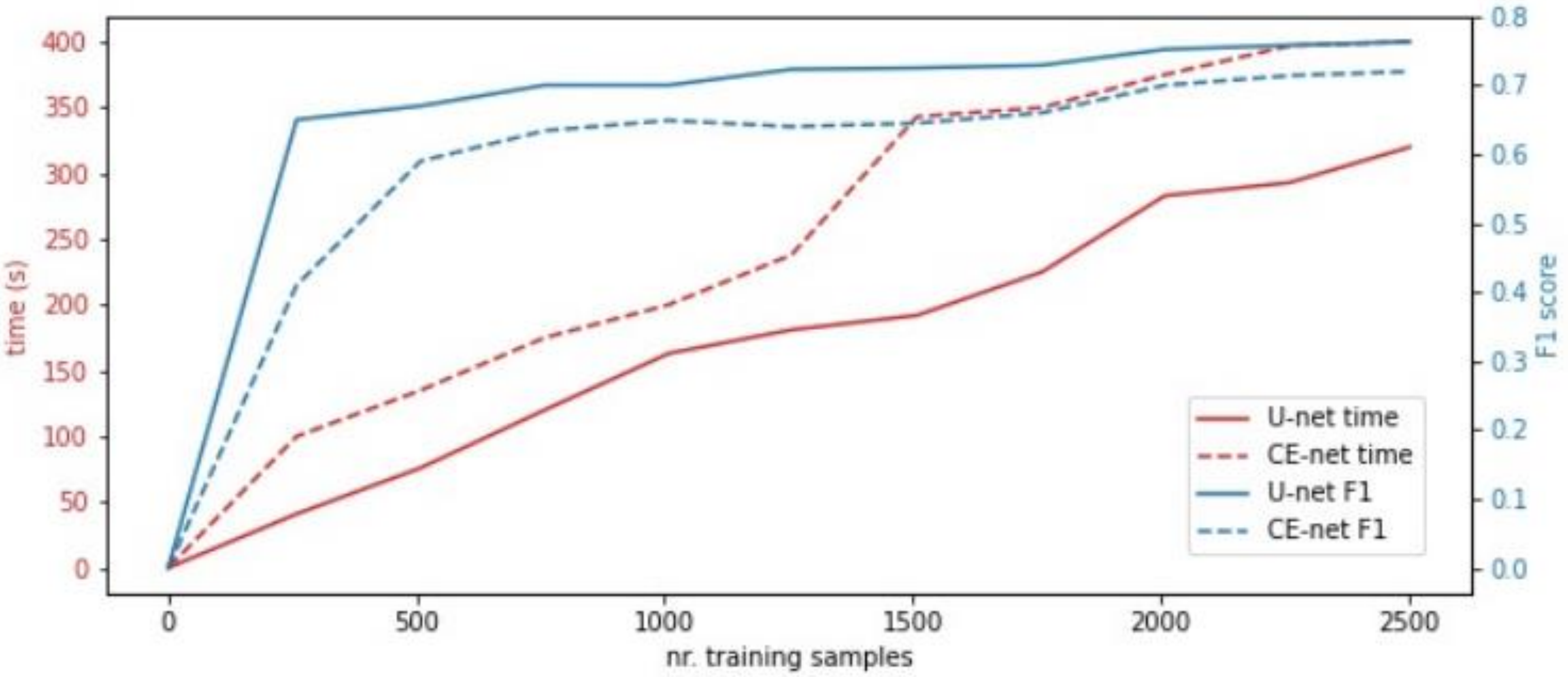


Figure 7. The figure shows the effect of increasing the training samples on the training time (red) and F1-score (blue) for U-Net (solid lines) and CE-Net (dashed lines).

Results

3. How robust are the predictions of both models?

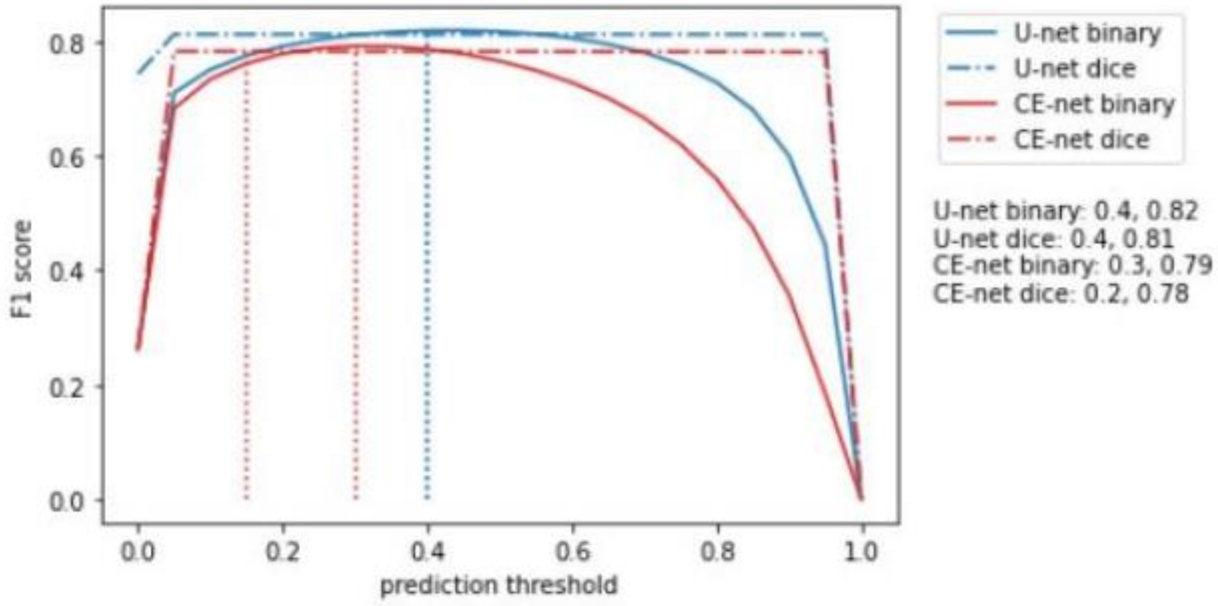


Figure 8. Prediction threshold for the test-set in relation to the F1-score for U-Net and CE-Net to compare robustness of the predictions.

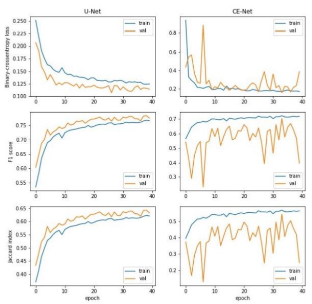
Results

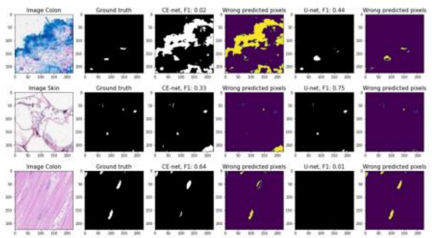
4. How is CE-Net performing in the segmentation of nuclei compared to U-Net looking at the F1-score and the Jaccard index scores?
5. How accurate is CE-Net in detecting the nuclei compared to U-Net looking at the Cell-counting accuracy metric?
6. Are the results on the PanNuke dataset of this research similar to the results of Gamper et al., when comparing the Panoptic Quality metric?

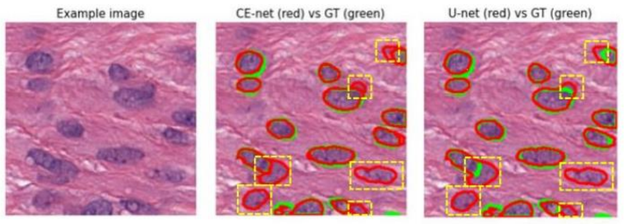
Table 3. Performances comparison between U-Net, CE-Net, and a non-trainable baseline where all pixels are predicted as nuclei (mean / standard deviation).

Loss-function		Jaccard Index	F1 score	Nuclei-Counting Accuracy	Panoptic Quality
		test set	test set	test set	test set
U-Net	Binary crossentropy	**0.70 / 0.13**	**0.82 / 0.11**	0.69 / 0.77	**0.67 / 0.12**
	Dice loss	0.70 / 0.14	0.81 / 0.12	**0.69 / 0.75**	0.65 / 0.13
CE-Net	Binary crossentropy	0.67 / 0.14	0.79 / 0.14	0.55 / 0.82	0.64 / 0.13
	Dice loss	0.66 / 0.15	0.78 / 0.14	0.65 / 0.80	0.62 / 0.14
Non-trainable baseline: all nuclei		0.16 / 0.10	0.26 / 0.14	-	-

Discussion



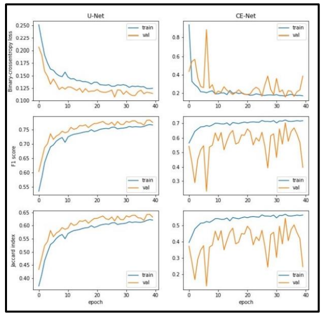


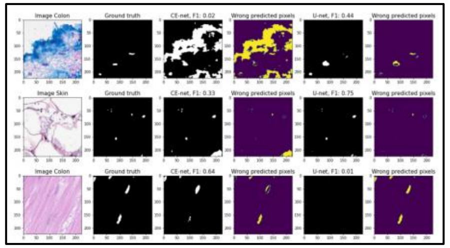


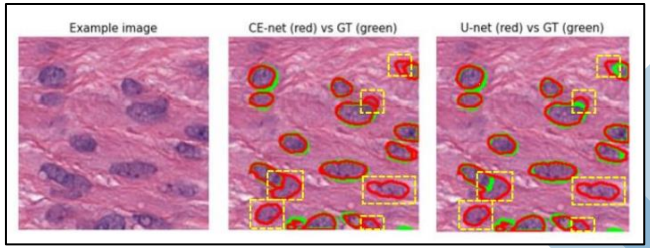
- Predicted masks:
  - CE-Net finds wrong patterns
  - CE-Net can capture nuclei that are not obvious
- Hyperparameter tuning:
  - Learning rate schedular
  - Weight decay
  - Add dropout layers
  - Momentum
- Potential errors in data:
  - Check significance of errors to the outcomes of the models
  - Apply models to different dataset

Why is CE-Net not outperforming U-Net as was expected?

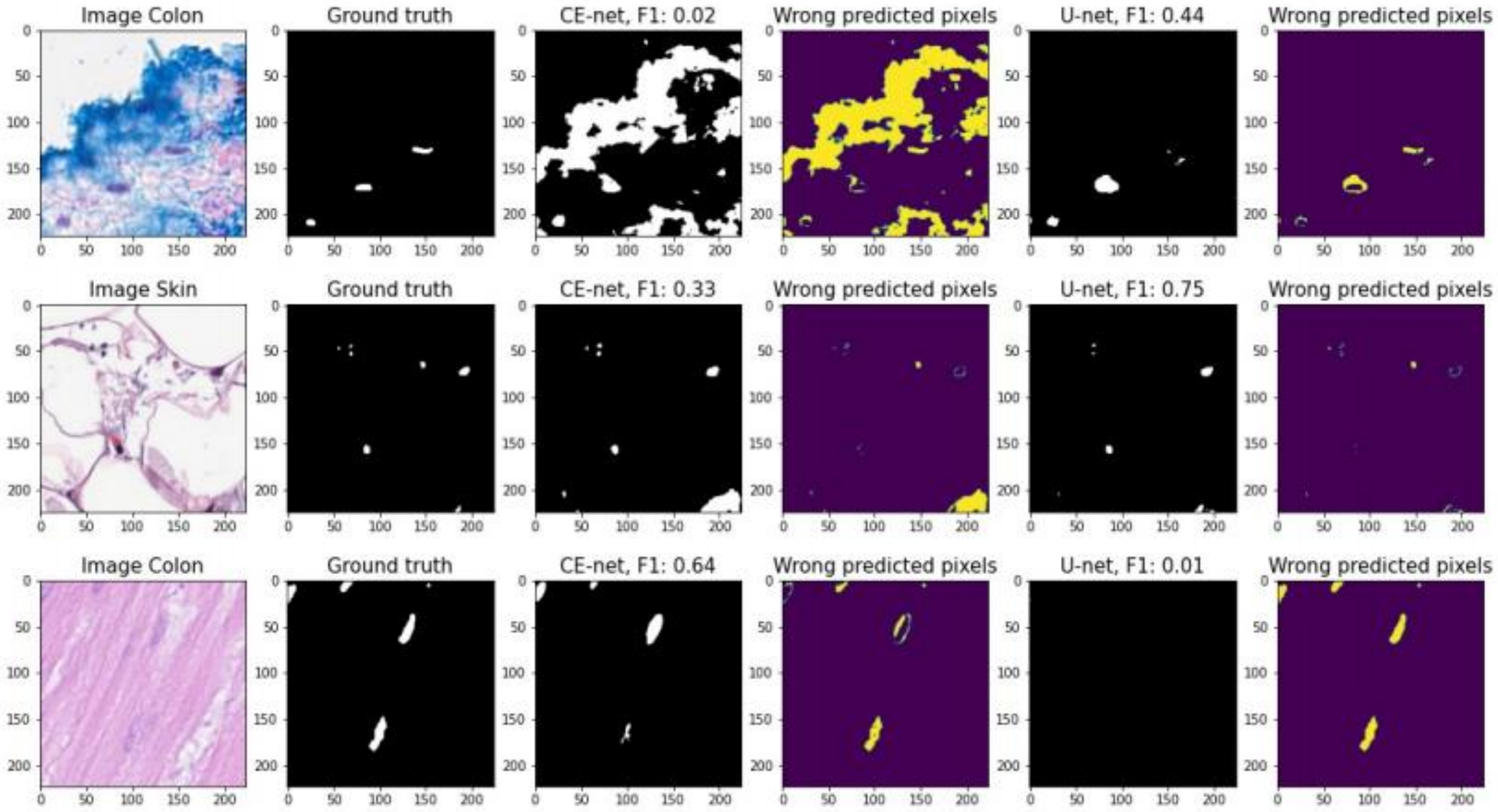
Discussion



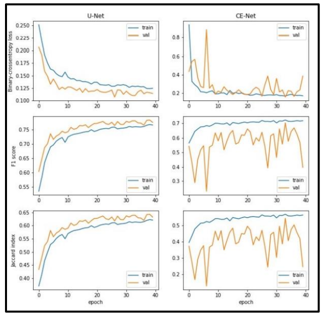


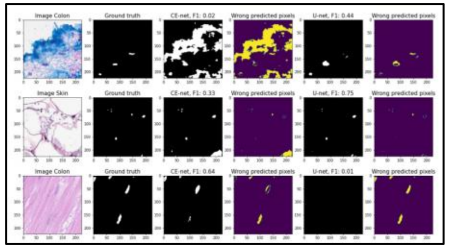


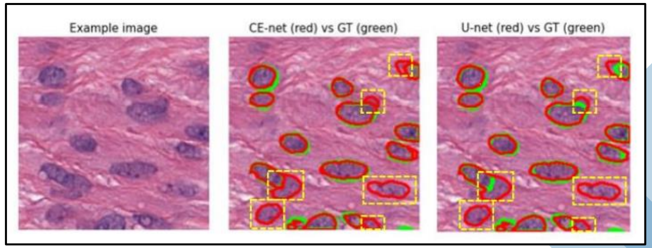
- Predicted masks:
  - CE-Net finds wrong patterns
  - CE-Net can capture nuclei that are not obvious
- Hyperparameter tuning:
  - Learning rate schedular
  - Weight decay
  - Add dropout layers
- Potential errors in data:
  - Check significance of errors to the outcomes of the models
  - Apply models to different dataset

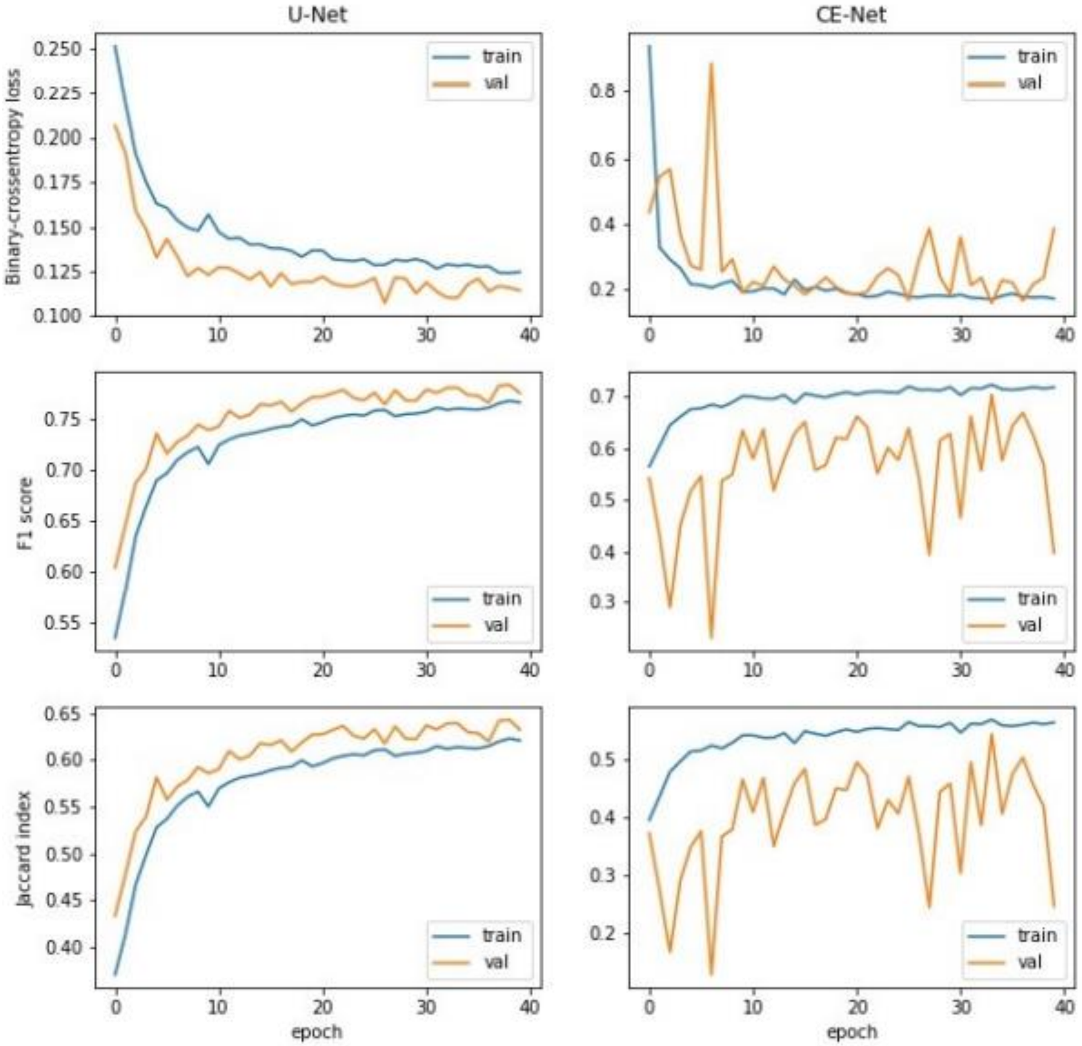


Discussion



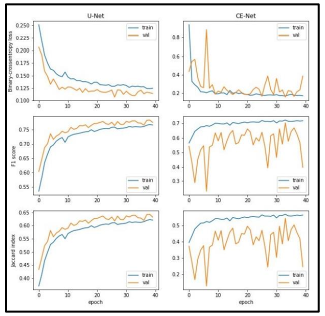


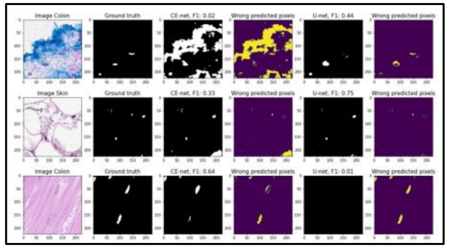


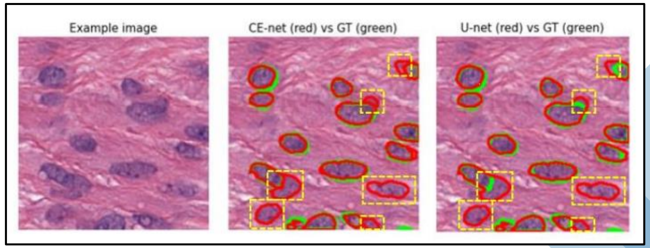


- Predicted masks:
  - CE-Net finds wrong patterns
  - CE-Net can capture nuclei that are not obvious
- ➢ Hyperparameter tuning:
  - Learning rate schedular
  - Weight decay
  - Add dropout layers
- Potential errors in data:
  - Check significance of errors to the outcomes of the models
  - Apply models to different dataset

Discussion

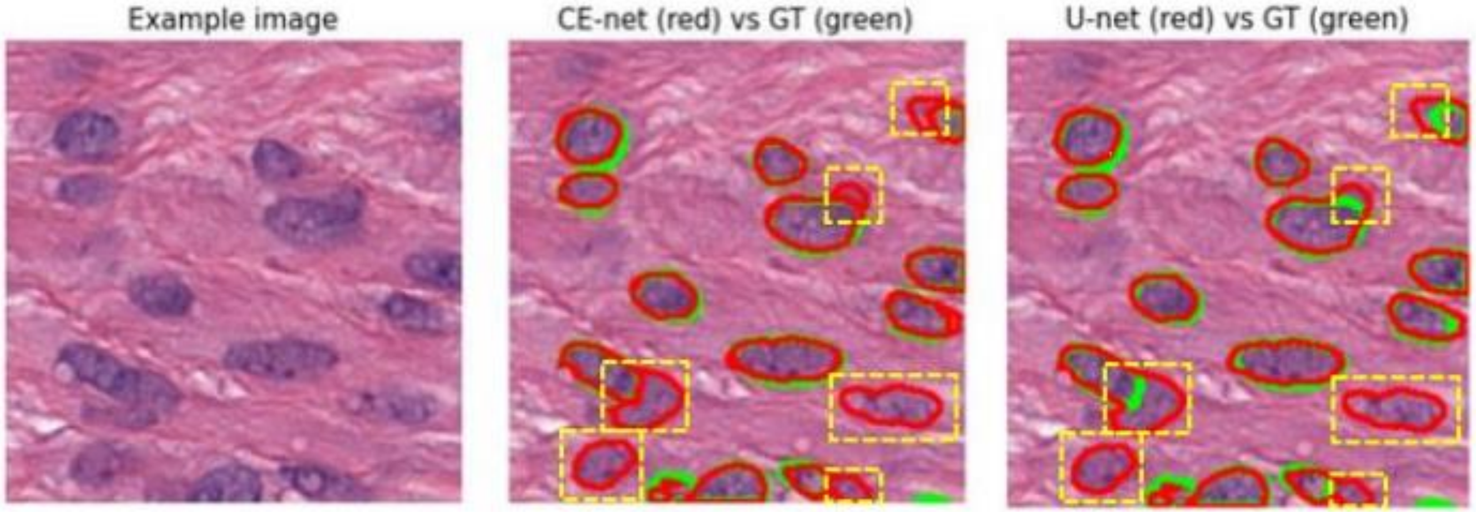






- Predicted masks:
  - CE-Net finds wrong patterns
  - CE-Net can capture nuclei that are not obvious
- Hyperparameter tuning:
  - Learning rate schedular
  - Weight decay
  - Add dropout layers

➢ Potential errors in data:
  - Check significance of errors to the outcomes of the models
  - Apply models to different dataset



Conclusion
CE-Net vs U-Net

U-Net:

- ✓ has higher outcomes on all evaluation metrics
- ✓ is able to handle a smaller training set (<500)
- ✓ is more robust as it is less responsive to different prediction thresholds

There is potential of CE-Net to capture complex patterns

Tuning the hyperparameters could reduce overfitting and improve the scores

Significance of the errors should be investigated

Thank you for listening