UNIVERSITY OF TARTU

Measuring Testis Tubule Wall Thickness in Histopathology Images

Arnel Pällo

Supervisor: Dmytro Fishman, PhD

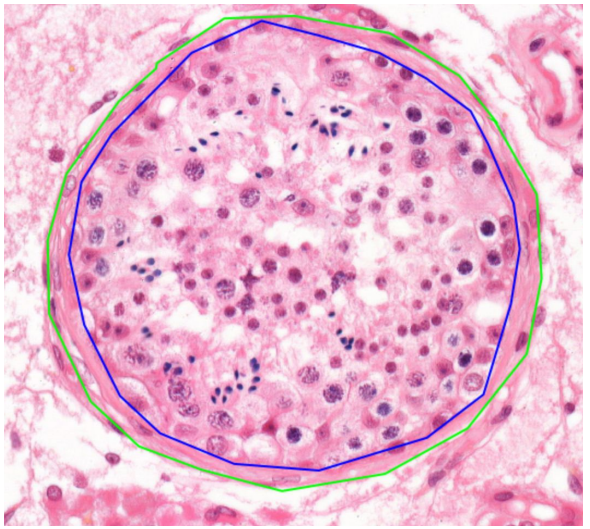
Infertility

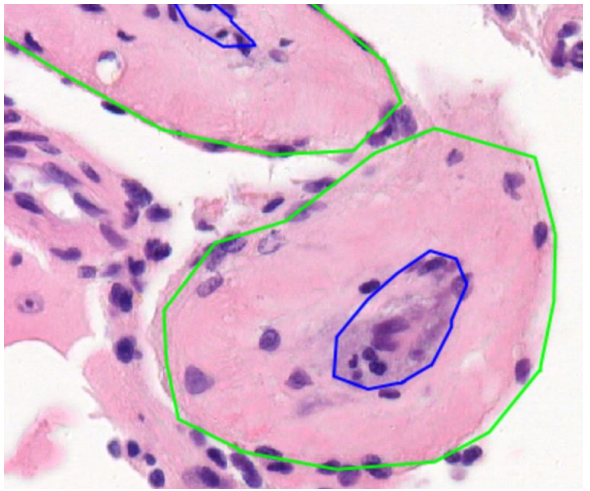
- Affects 8–12% of couples
- 40-50% due to the “male factor”

One Particular Cause

- Too thick tubule walls
  - Sperm is locked in

- Diagnosed by analysing microscopy images
  - Done manually
  - Takes 30 minutes
  - Does not scale



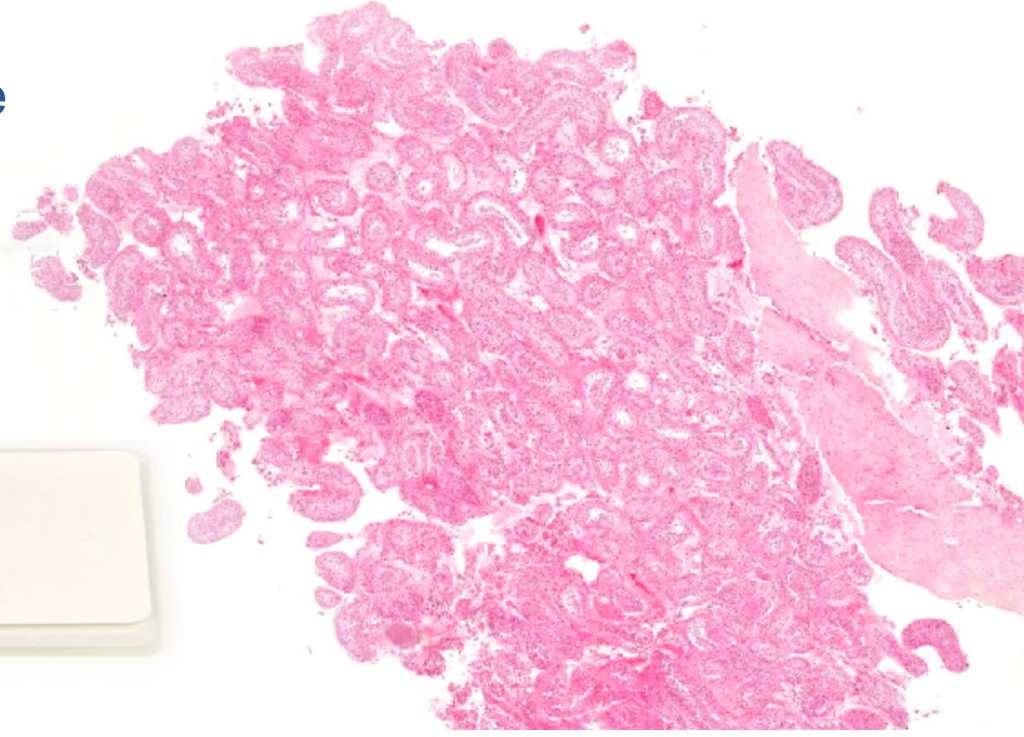


Can we automate this?

A software pipeline to identify tubules, measure walls' thicknesses and report results

Analysing Microscopy Images

- Tissue samples processed into glass slides
- Scanned under a microscope
- 50 000 megapixel images



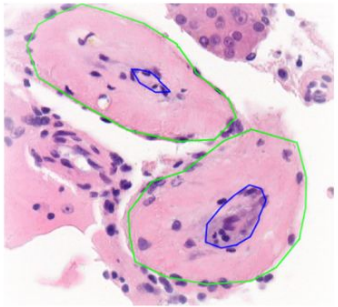
Prior Work

- Many models for medical image segmentation
  - U-Net most popular

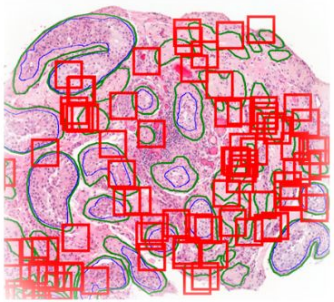
- No papers specifically focusing on wall thickness measurement

- Few resources on how to create a full pipeline

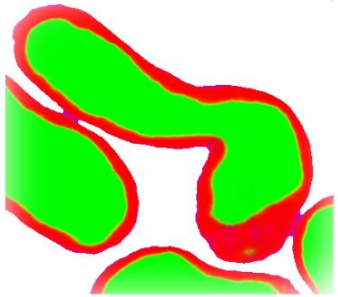
Pipeline Overview



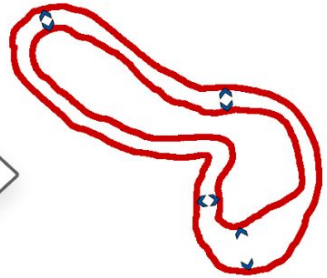
50K megapixel microscopy images



Generate small patches



Train segmentation model

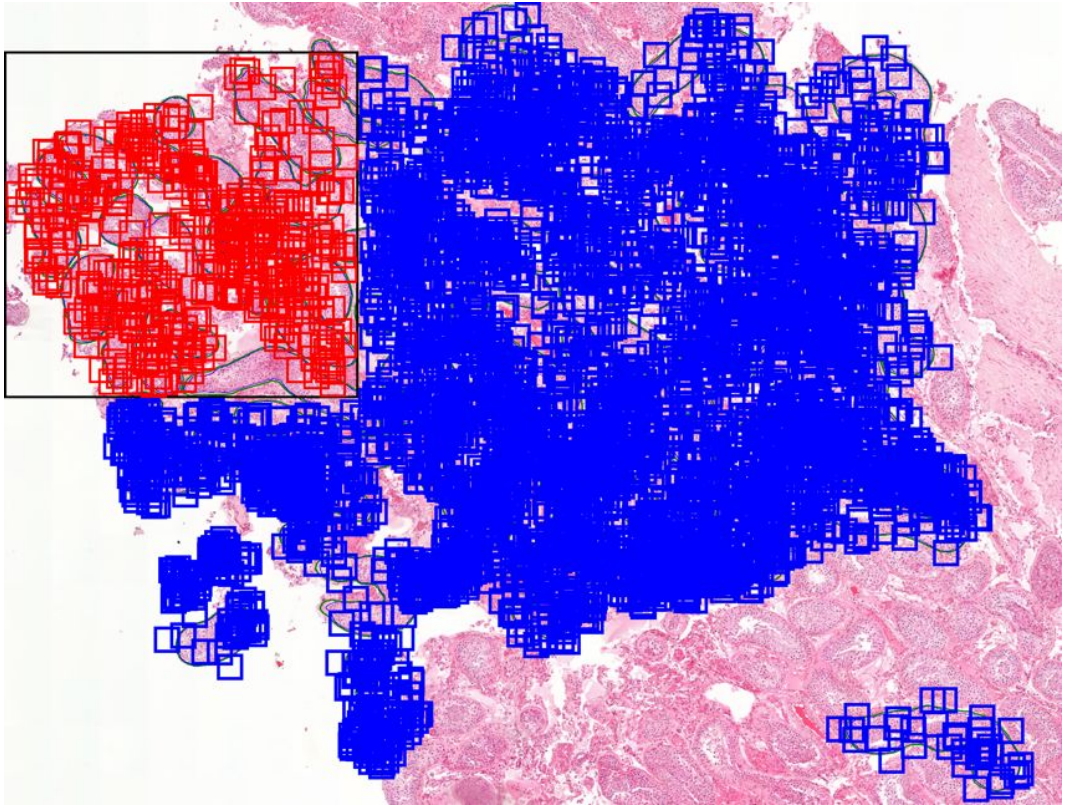


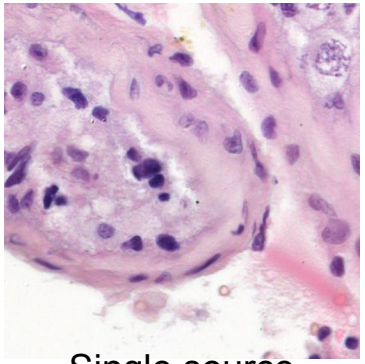
Measure border widths on segmentation maps

Dataset

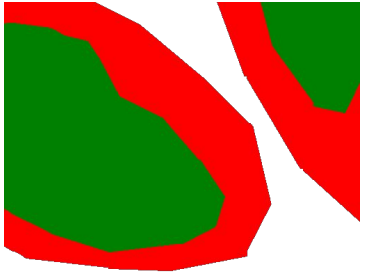
- 3 microscopy images from East-Tallinn Central Hospital
  - 50K MP Digital scans in MIRAX file format
  - Ground-truth annotations in GeoJSON format

Generating Train Patches



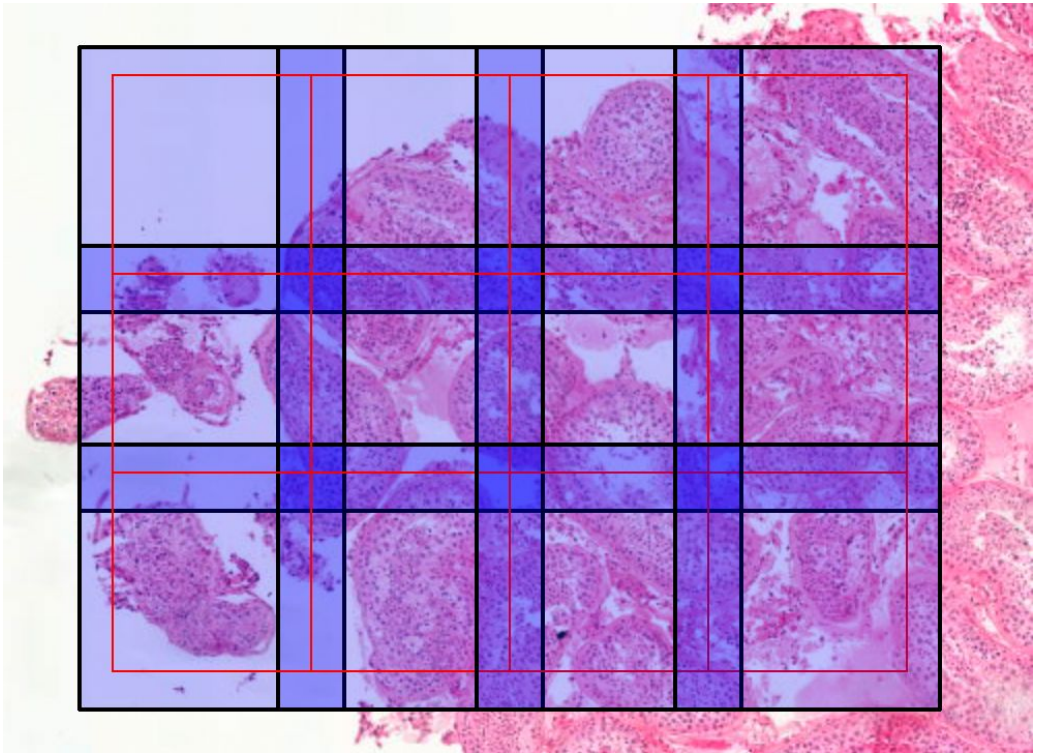


Single source patch (512 px)



Ground truth segmentation map

Generating Patches from Grid



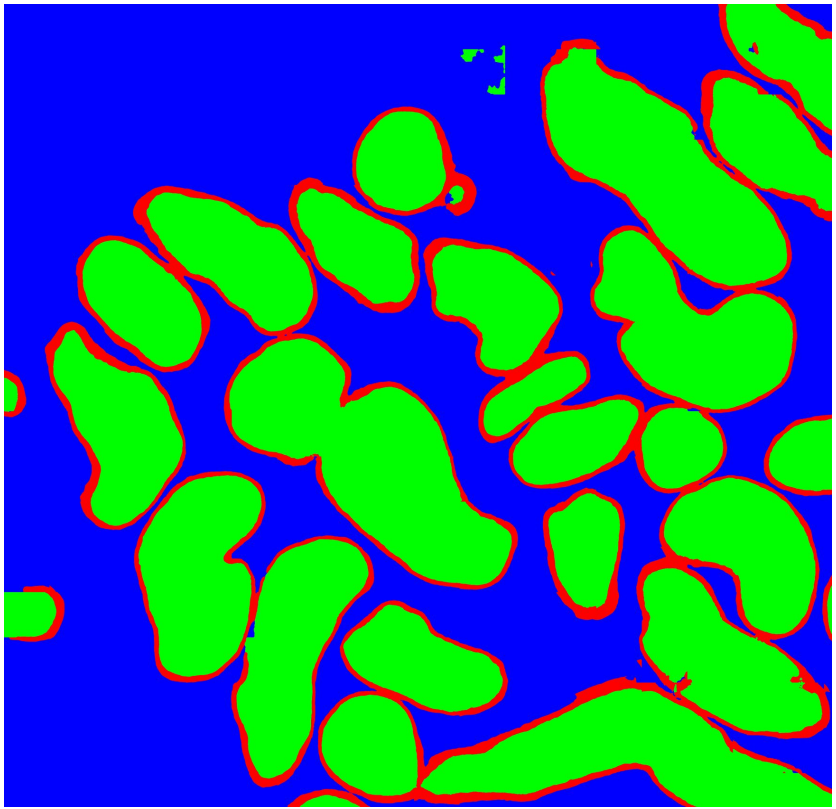
Inference: generate patches form grid with overlap

Training

- U-Net with resnet-34 backbone
  - Categorical cross-entropy loss
  - Image pixel values normalised to 0…1
- Trained for 100 epochs

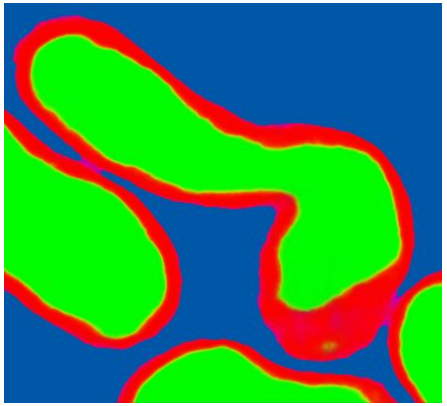
Reconstruction & Cleanup

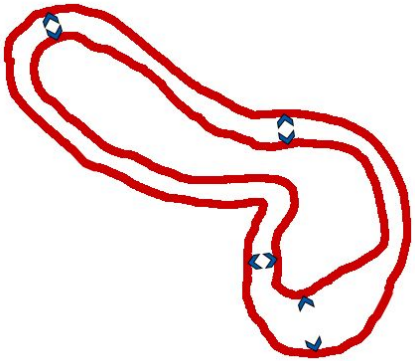
- Put pieces back together
  - Discard overlapping regions
- Post-Processing:
  - Thresholding
  - Labelling
  - Morphological operations



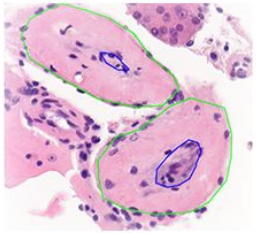
Edge Detection and Distance Measurement

- Detect edges with binary dilation
  - tubule body boundary line
  - background boundary line
- Measure distance between the lines

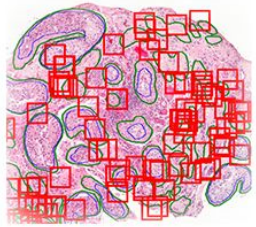




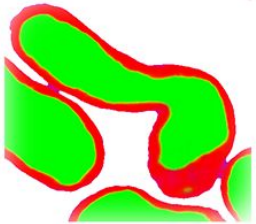
Complete Pipeline



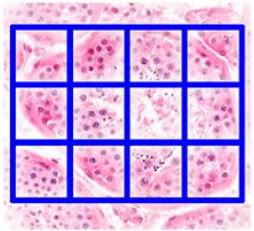
50K megapixel microscopy images



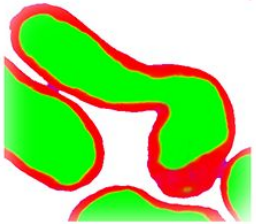
Training: generate random patches from slides



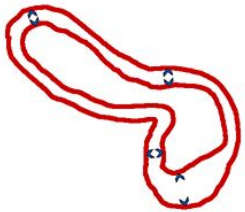
Train U-Net model to segment tubules



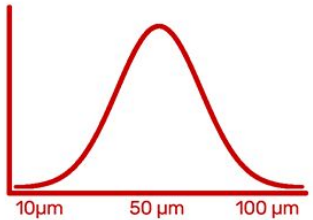
Inference: generate patches from grid with overlap



Segment patches with trained U-Net model

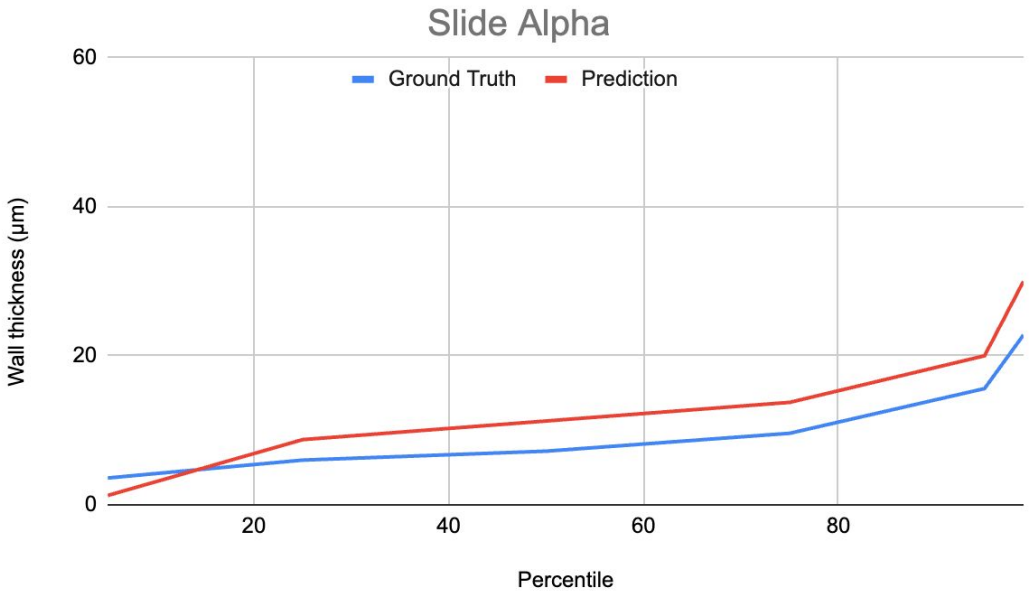


Measure border widths



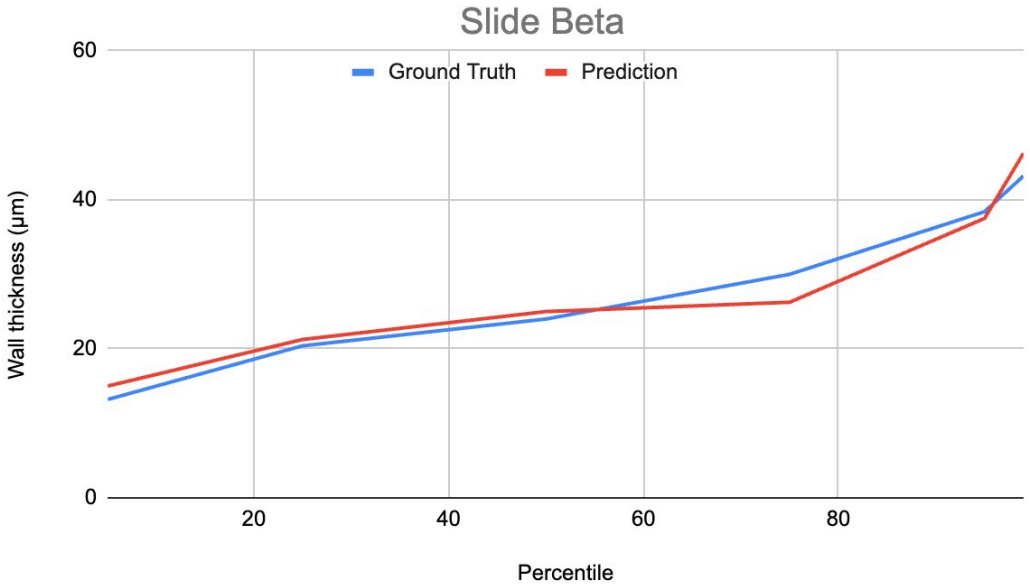
Report Results

Measurement Accuracy - Slide Alpha



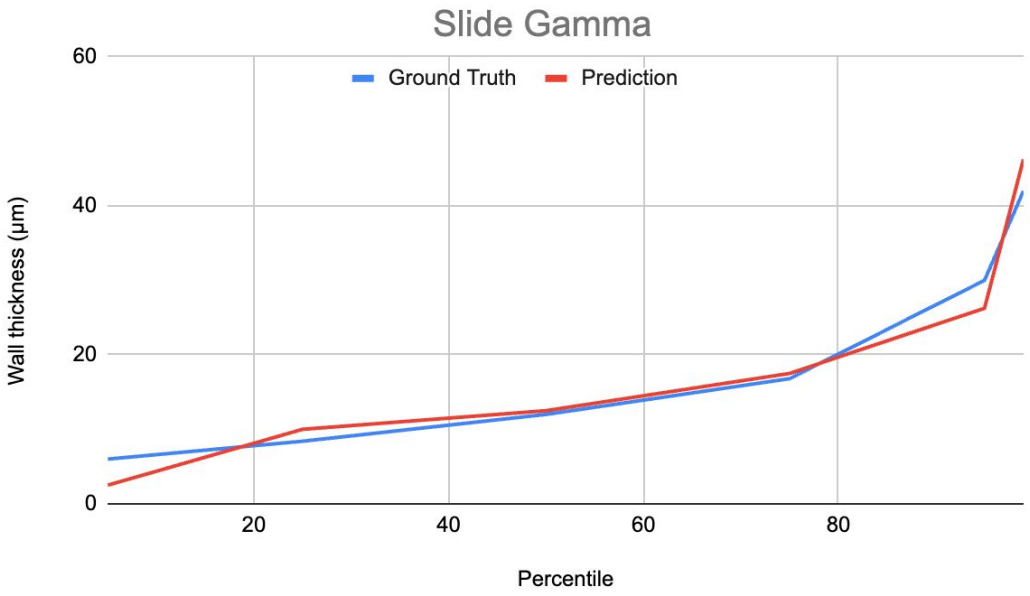
Mean (ground truth): 7 µm; Mean (prediction): 11 µm; RMSE 3.705

Measurement Accuracy - Slide Beta



Mean (ground truth): 24 µm; Mean (prediction): 25 µm; RMSE 2.294

Measurement Accuracy - Slide Gamma



Mean (ground truth): 12 µm; Mean (prediction): 13 µm; RMSE 1.049

Hyperparameters

- Tested various pre-processing hyperparameters
  - zoom level
  - RGB vs grayscale
  - min. tubule area in every patch
- Tested different patch overlap during inference
  - 25 % vs 50 %

Experiment Results

- The model kept working
- None of the hyperparameter values was clearly outperforming others in every slide
  - Different settings worked better on different slides
  - Differences barely statistically significant
- Consider ensembling methods

Discussion & Conclusions

- The results are promising
  - Segmentation model successful
  - Software detected enlarged wall thicknesses on slide beta
- Contributions
  - Pre-processing
  - Thickness measurements
  - The whole pipeline

Going forward

- Report measurements tubule-by-tubule
- Adjust post-processing aggressiveness
- Investigate model ensembling
- Prepare for medical adoption
  - Train with more slides
  - Integrate with image data warehouse and/or scanners
  - Develop graphical user-interface

UNIVERSITY OF TARTU

Thank You!

UNIVERSITY OF TARTU
1632

Opponent's questions

1) What was the selection criteria for choosing the training data (e.g. Figure 3)?

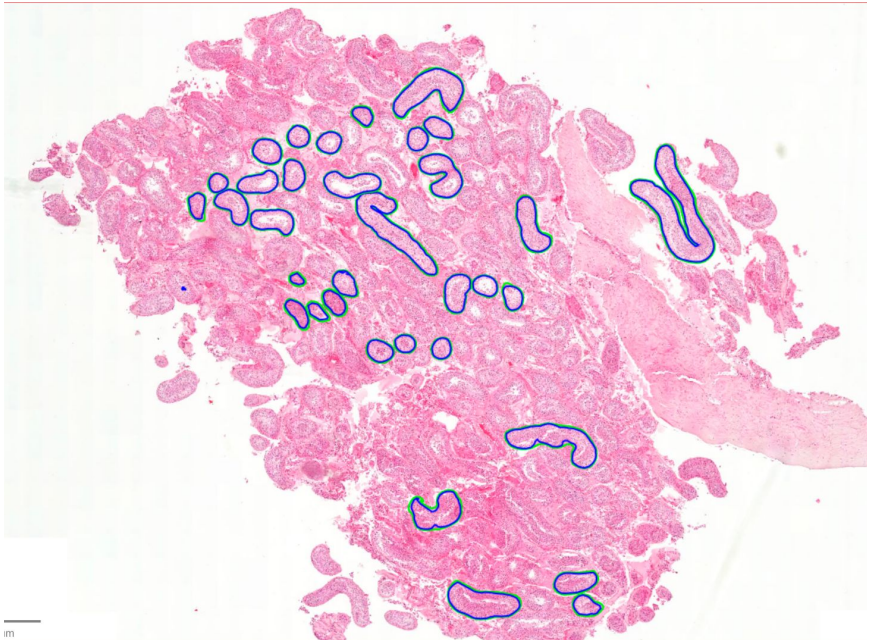
Training Data

- We received three annotated slides from East-Tallinn Central Hospital
- The patients were chosen randomly

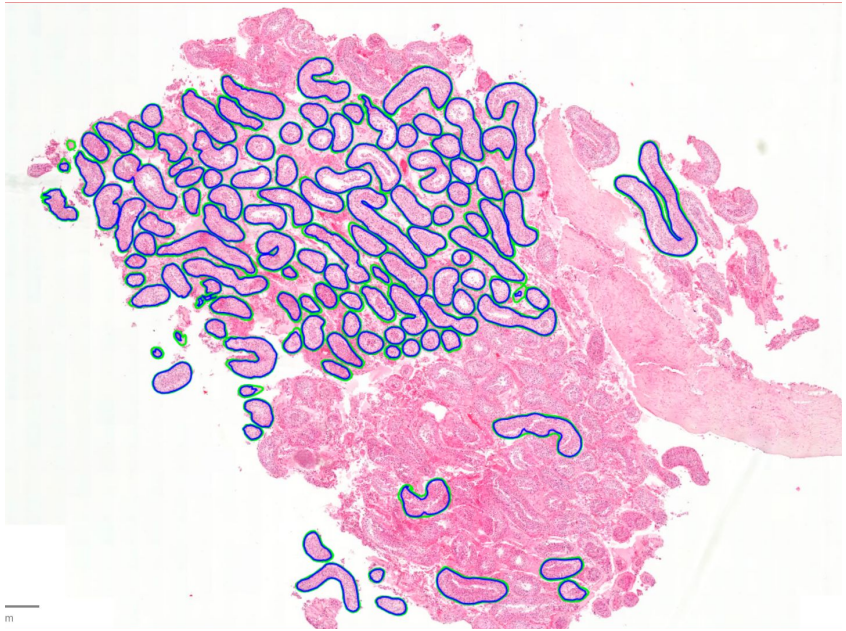
- One region on every slide was designated as test region. Manually picked and defined the area. Should contain about 20% of tubules
- Ideal world: completely separate slide for validation

2) The third tissue slide you received did not have complete annotations. You claim that you continued annotating it and after 12 hours of effort the slide was still less than 50% complete.

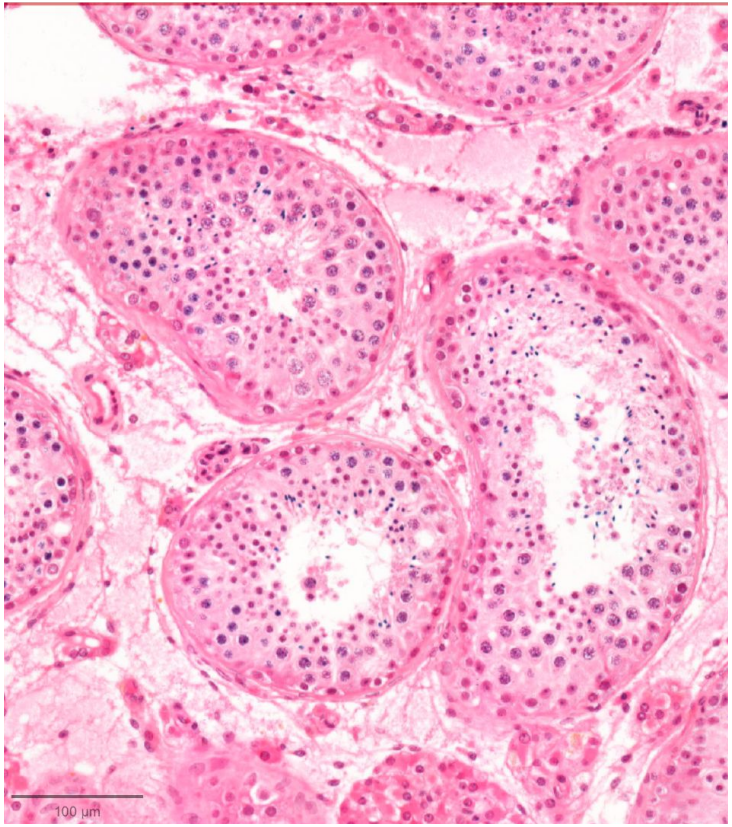
Did you end up using the annotations you generated? What portion did they make of the training data? Is the quality of your annotations at the same level with the clinical collaborators?

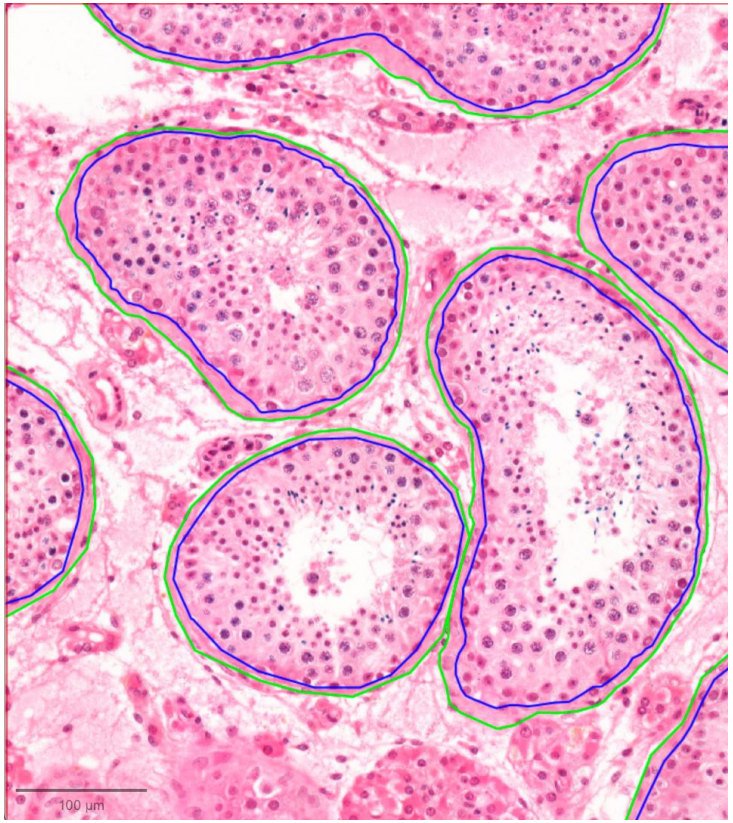


34 annotations



120 annotations


100 µm

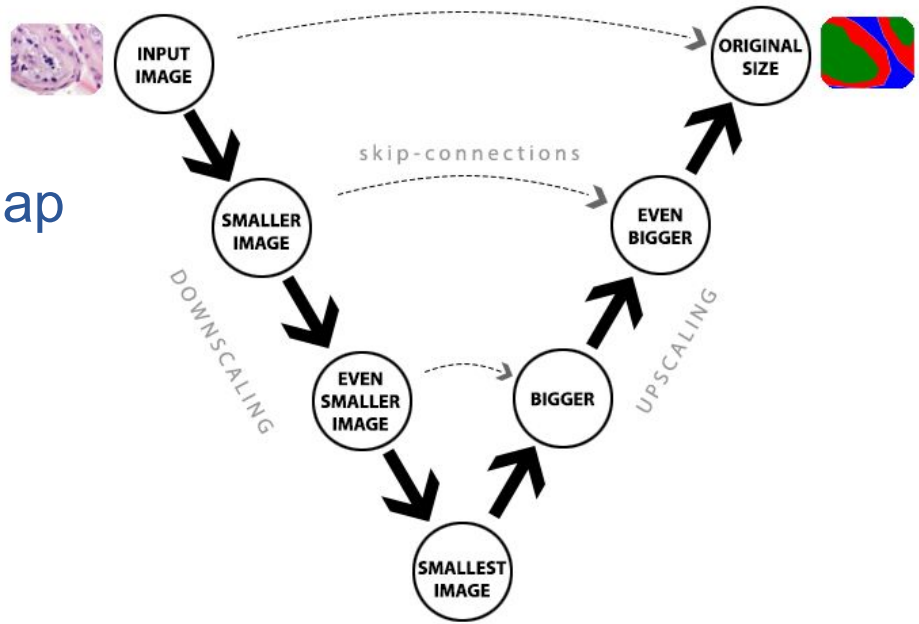

100 µm

3) What is the function of the encoder sub-network (backbone)?

How would using a different backbone affect the accuracy of the segmentation?

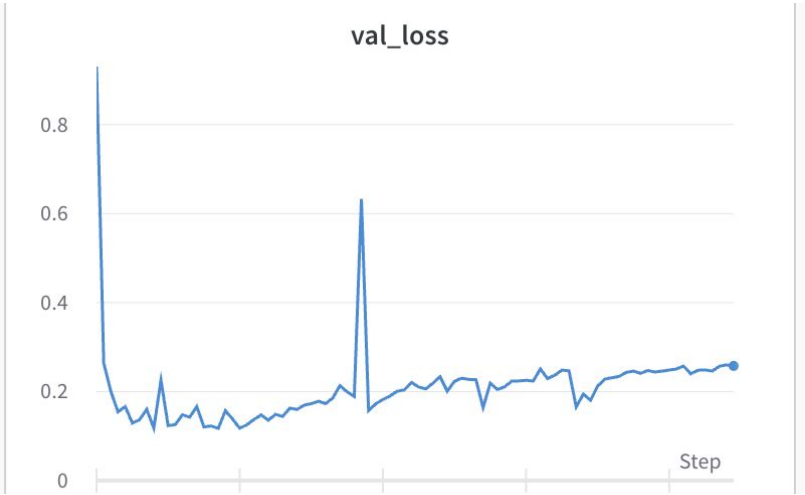
Encoder & Decoder

- Encoder extracts relevant features
  - compresses/simplifies data
  - learns patterns
- Decoder builds segmentation map
  - using the extracted features



Different Backbone

- Likely little to no effect
  - resnet-34 already overfits
  - pre-processing and data quantity are the differentiating factors

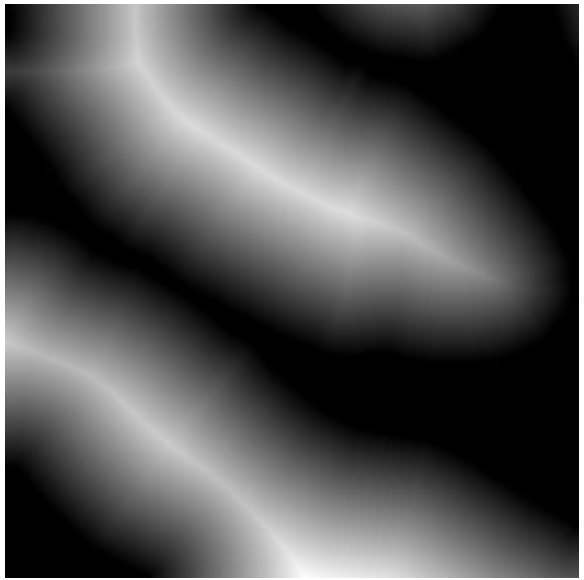


ResNet34 vs ResNet18

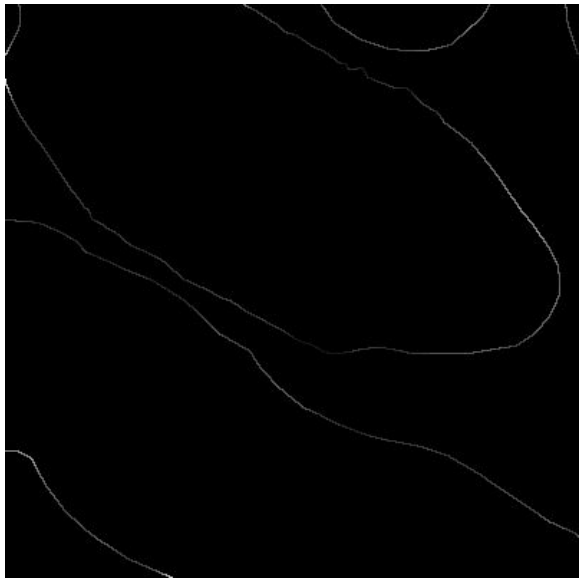
4) How did the algorithm used by the distance-maps library to measure the distance between the borders differ from yours?

If your algorithm was that much faster, did it come at the price of accuracy?

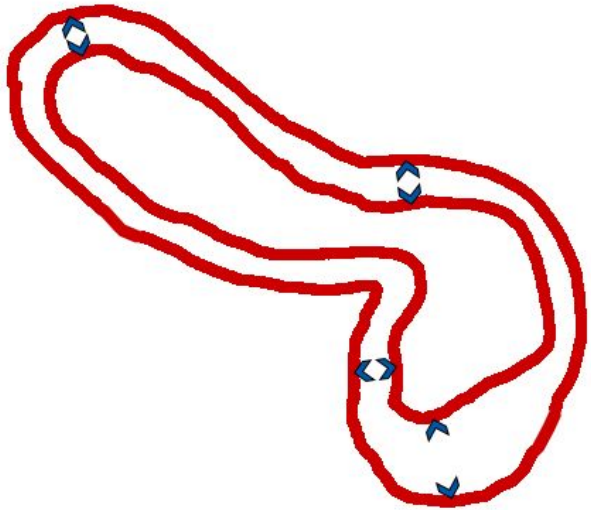
Could it handle the artefacts generated by the Otsu thresholding perhaps that your algorithm could not?



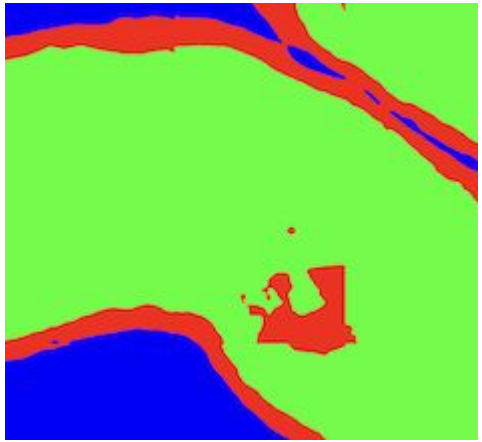
Distance-maps measures every pixel to the nearest “positive” pixel



Detect tubule edges, use it as a mask for the distance map. Remaining non-zero values used as a result

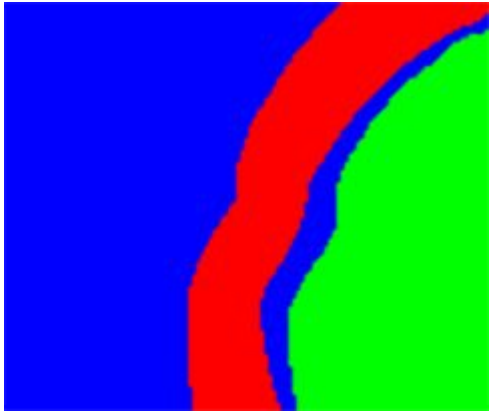


Detect tubule and background edges. Measure distance at every point.



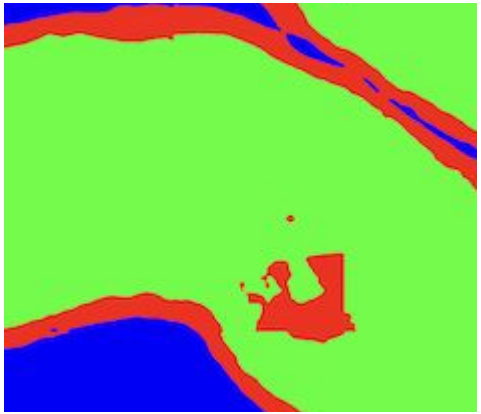
Tubule edge detected in the middle of a tubule.

Distance to background is long

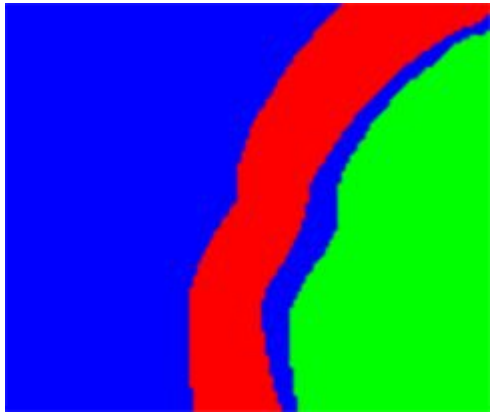


Distance from tubule edge to background is zero

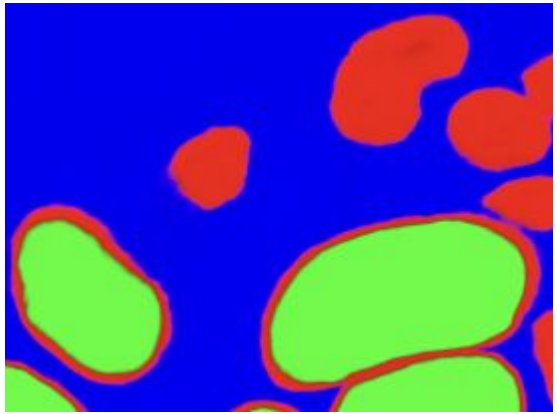
Measure “outside in”



Solves the issue

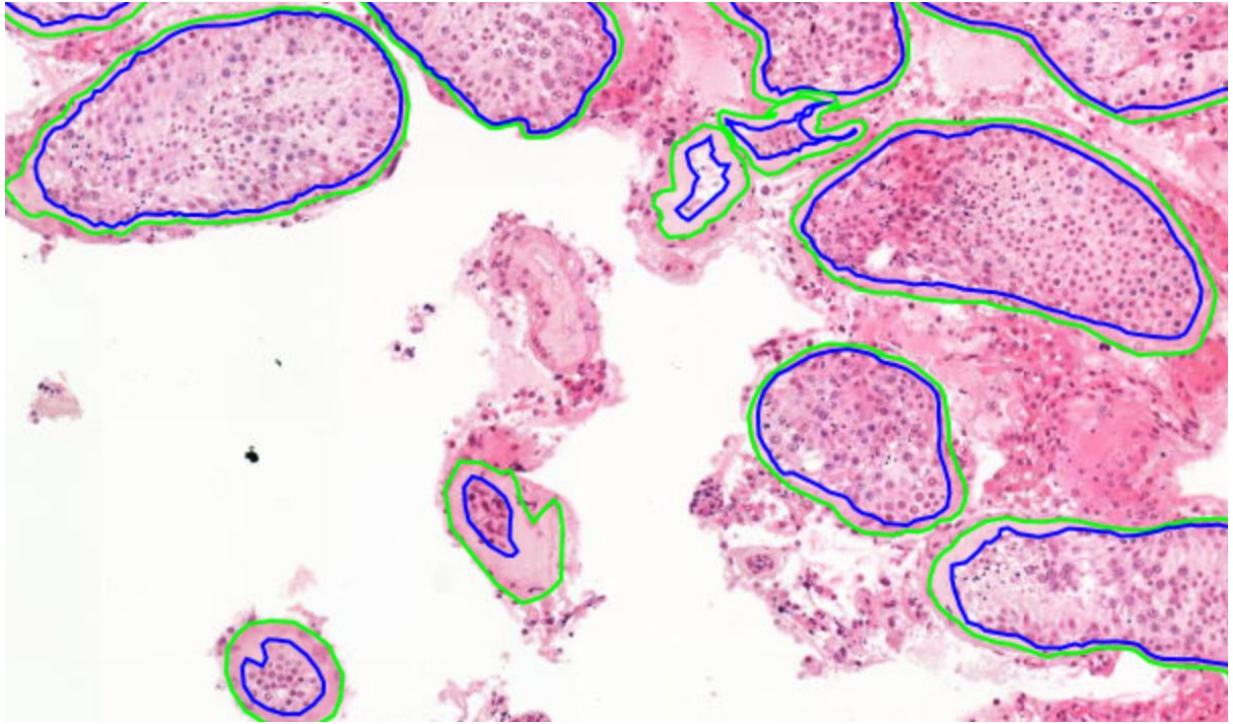


Generates large and small values. Average closer to correct



Introduces new error mode for fully hyalinized tubules

5) Besides testing different hyperparameters, are there any other methods you could use in the pre-processing part to get a better segmentation map?



Sparse region on the histopathology slide

Sampling of training data

6) What are the steps required for this software pipeline to be used by medical facilities on real patients?

Steps Required

- Train with more slides
- Integrate with image data warehouse and/or scanners
- Develop graphical user-interface