University of Amsterdam

Amsterdam UMC
Universitair Medische Centra

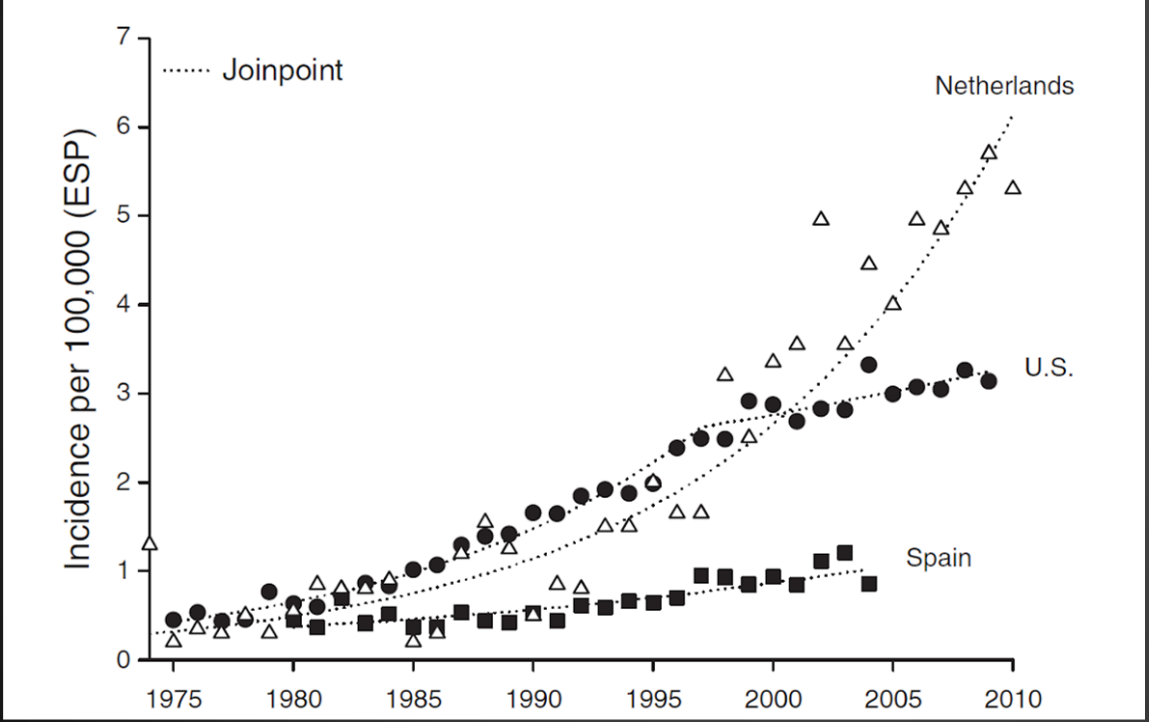
U-net Ensembles for Accurate Risk Stratification of Precursor Lesions of Esophageal Adenocarcinoma

Bryan Cardenas Guevara* & Vivian van Oijen*

*Graduate School of Informatics, University of Amsterdam

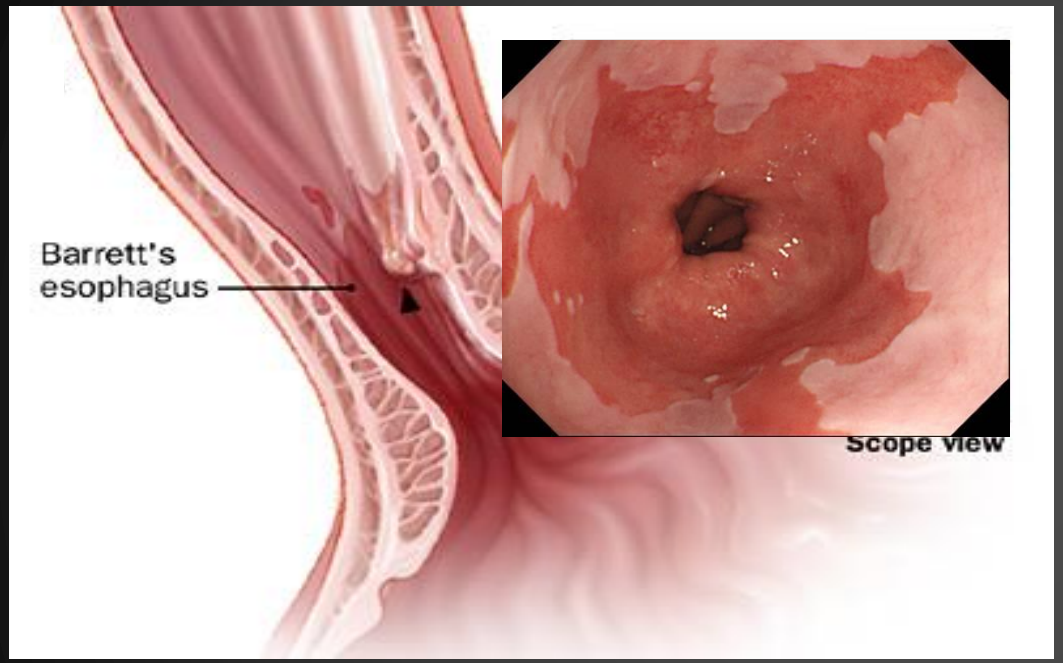
Incidence Rate of Esophageal Adenocarcinoma

- Sharp increase in the incidence rate



Am J Gastroenterol 2014; 109; 336-343

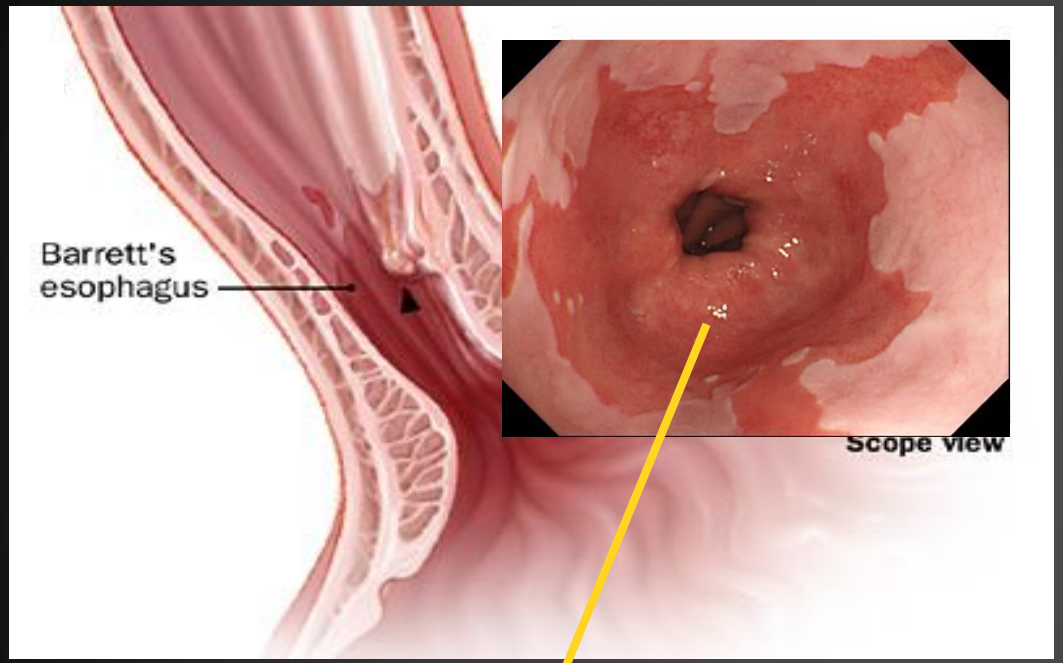
Barrett's Esophagus



Gastro Esophageal Reflux disease Barrett's Esophagus

At risk to develop Adenocarcinoma
Endoscopic surveillance
with biopsies for risk stratification

Barrett's Esophagus

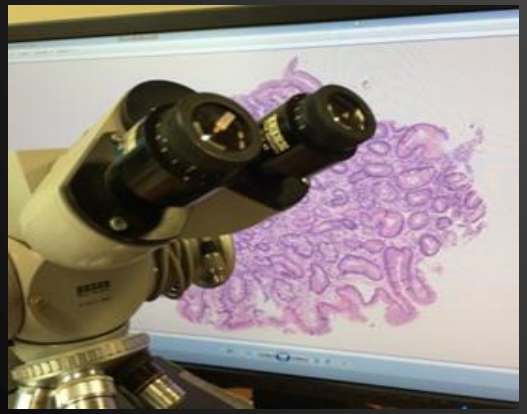


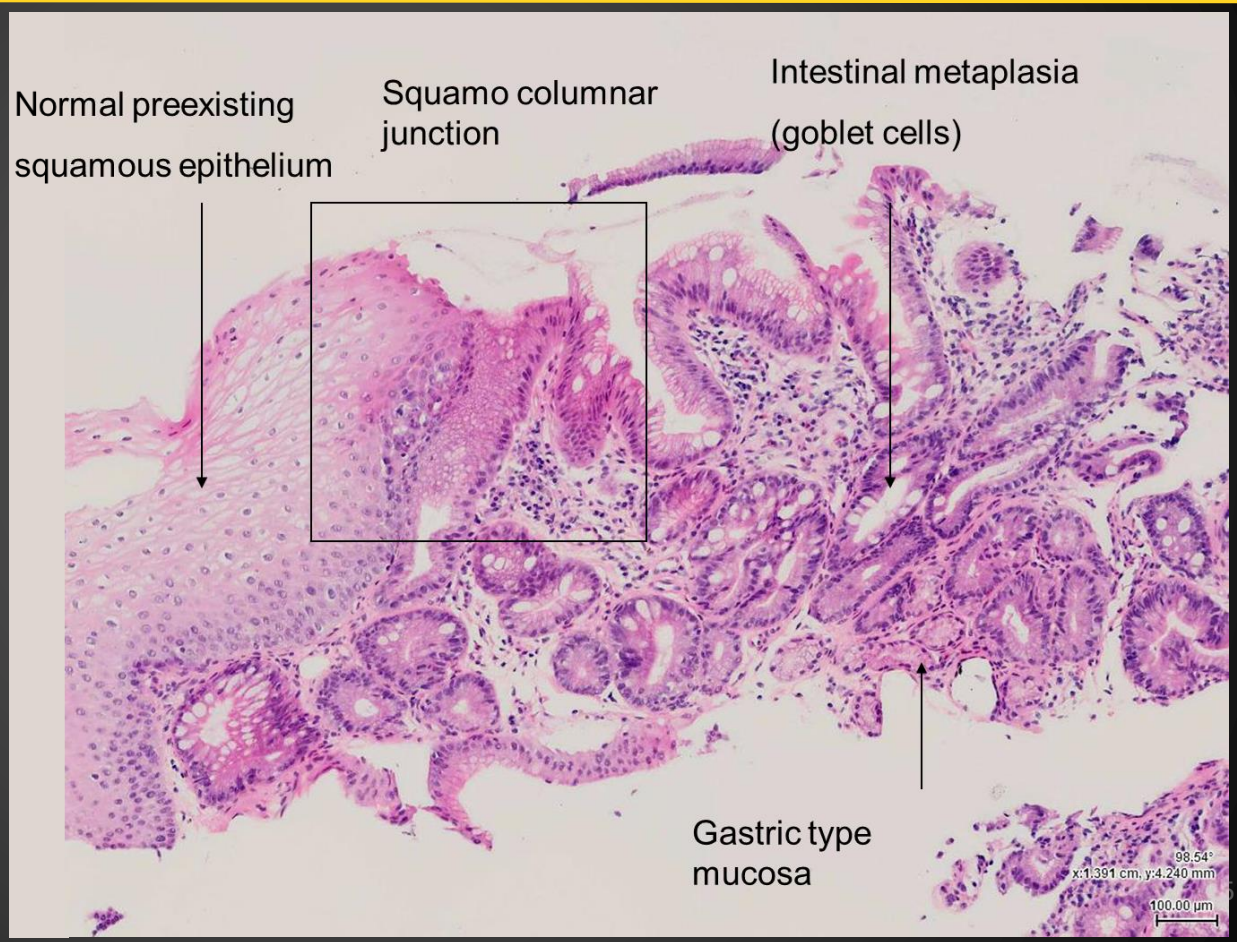
Pink discoloration

Gastro Esophageal Reflux disease Barrett's Esophagus

At risk to develop Adenocarcinoma
Endoscopic surveillance
with biopsies for risk stratification

Barrett's on Biopsy





From Barrett's Esophagus to Adenocarcinoma

A proportion of patients with Barrett's Esophagus develop Adenocarcinoma

Risk of progression of Non-Dysplastic Barrett's Epitheleum to cancer is low:

- **~0.3 - 0.6 %** per year

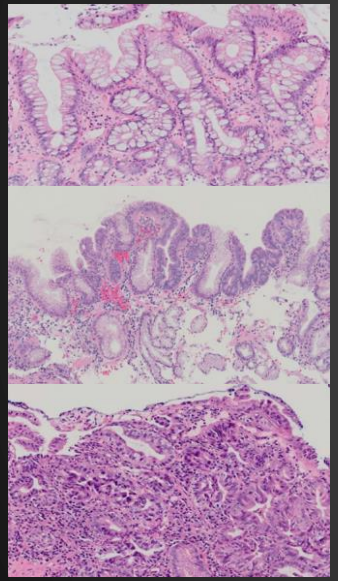
Hvid-Jensen F. NEJM 2011, de Jonge P. Gut 2014; Dutch Barrett guideline

From Barrett's Esophagus to Adenocarcinoma

A proportion of patients with Barrett's Esophagus develop Adenocarcinoma

Risk of progression of Non-Dysplastic Barrett's Epitheleum to cancer is low:

- **~0.3 - 0.6 %** per year



No Dysplasia (NDBE)	**Low interval of endscopic surveillance**
Low Grade Dysplasia	**High interval of endoscopic surveillance -- Endoscopic ablation therapy**
High Grade Dysplasia	**Endoscopic ablation therapy -- surgical therapy**

CANCER

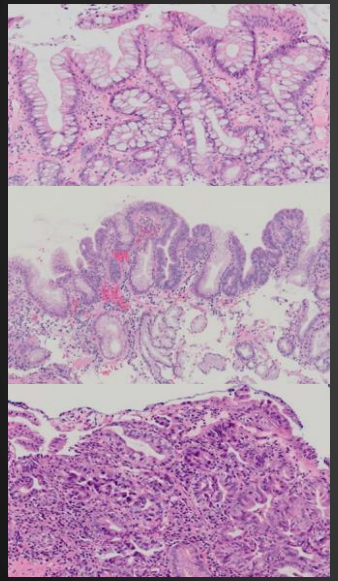
Hvid-Jensen F. NEJM 2011, de Jonge P. Gut 2014; Dutch Barrett guideline

From Barrett's Esophagus to Adenocarcinoma

A proportion of patients with Barrett's Esophagus develop Adenocarcinoma

Risk of progression of Non-Dysplastic Barrett's Epitheleum to cancer is low:

- **~0.3 - 0.6 %** per year



No Dysplasia (NDBE)	**Low interval of endscopic surveillance**
Low Grade Dysplasia	**High interval of endoscopic surveillance -- Endoscopic ablation therapy**
High Grade Dysplasia	**Endoscopic ablation therapy -- surgical therapy**

CANCER

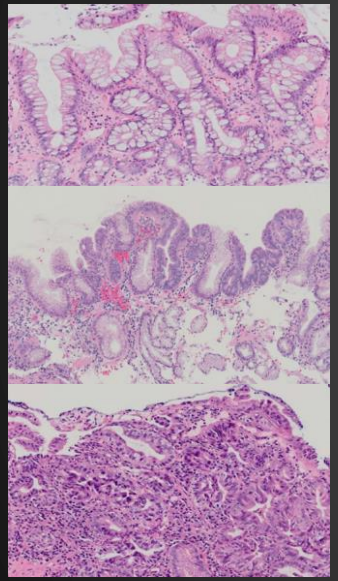
Hvid-Jensen F. NEJM 2011, de Jonge P. Gut 2014; Dutch Barrett guideline

From Barrett's Esophagus to Adenocarcinoma

A proportion of patients with Barrett's Esophagus develop Adenocarcinoma

Risk of progression of Non-Dysplastic Barrett's Epitheleum to cancer is low:

- **~0.3 - 0.6 %** per year



No Dysplasia (NDBE) — **Low interval of endscopic surveillance**

Low Grade Dysplasia — **High interval of endoscopic surveillance -- Endoscopic ablation therapy**

High Grade Dysplasia — **Endoscopic ablation therapy -- surgical therapy**

CANCER

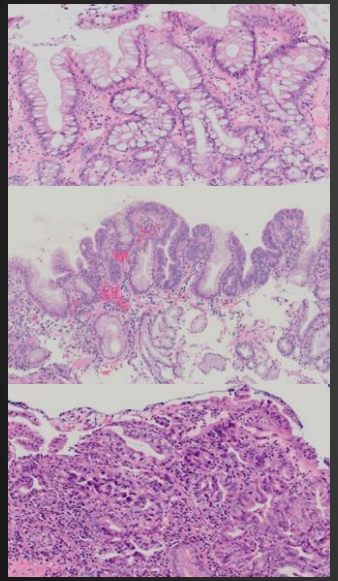
Hvid-Jensen F. NEJM 2011, de Jonge P. Gut 2014; Dutch Barrett guideline

From Barrett's Esophagus to Adenocarcinoma

A proportion of patients with Barrett's Esophagus develop Adenocarcinoma

Risk of progression of Non-Dysplastic Barrett's Epitheleum to cancer is low:

- **~0.3 - 0.6 %** per year



No Dysplasia (NDBE)	**Low interval of endscopic surveillance**
Low Grade Dysplasia	**High interval of endoscopic surveillance -- Endoscopic ablation therapy**
High Grade Dysplasia	**Endoscopic ablation therapy -- surgical therapy**

CANCER

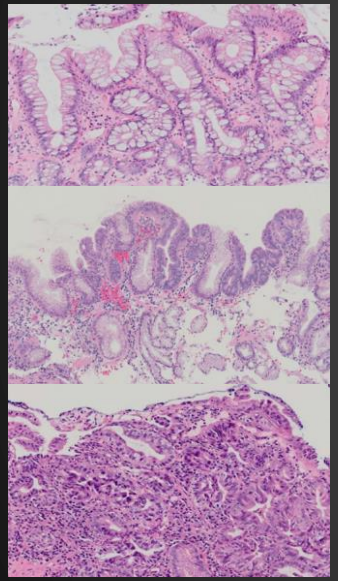
Hvid-Jensen F. NEJM 2011, de Jonge P. Gut 2014; Dutch Barrett guideline

From Barrett's Esophagus to Adenocarcinoma

A proportion of patients with Barrett's Esophagus develop Adenocarcinoma

Risk of progression of Non-Dysplastic Barrett's Epitheleum to cancer is low:

- **~0.3 - 0.6 %** per year



No Dysplasia (NDBE)	**Low interval of endscopic surveillance**
Low Grade Dysplasia	**High interval of endoscopic surveillance -- Endoscopic ablation therapy**
High Grade Dysplasia	**Endoscopic ablation therapy -- surgical therapy**

CANCER

Hvid-Jensen F. NEJM 2011, de Jonge P. Gut 2014; Dutch Barrett guideline

Interobserver Variation in Low Grade Dysplasia

Gastroenterology 2017;152:564–570

Discordance Among Pathologists in the United States and Europe in Diagnosis of Low-Grade Dysplasia for Patients With Barrett's Esophagus

Prashanth Vennalaganti,[1,2] Vijay Kanakadandi,[1,2] John R. Goldblum,[3] Sharad C. Mathur,[4] Deepa T. Patil,[3] G. Johan Offerhaus,[5] Sybren L. Meijer,[6] Michael Vieth,[7] Robert D. Odze,[8] Saligram Shreyas,[1,2] Sravanthi Parasa,[1,2] Neil Gupta,[9] Alessandro Repici,[10] Ajay Bansal,[1,2] Titi Mohammad,[1,2] and Prateek Sharma[1,2]

American Journal of Gastroenterology
© 2008 by Am. Coll. of Gastroenterology
Published by Blackwell Publishing
ISSN 0002-9270
doi: 10.1111/j.1572-0241.2008.02020.x

ORIGINAL CONTRIBUTIONS CME

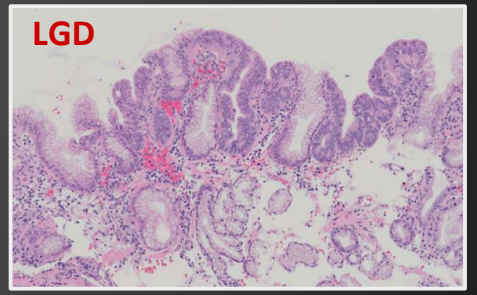
Pathology

Poor Interobserver Agreement in the Distinction of High-Grade Dysplasia and Adenocarcinoma in Pretreatment Barrett's Esophagus Biopsies

Erinn Downs-Kelly, D.O.,[1] Joel E. Mendelin, M.D.,[1] Ana E. Bennett, M.D.,[1] Elias Castilla, M.D.,[1] Walter H. Henricks,[1] Lynn Schoenfield, M.D.,[1] Marek Skacel, M.D.,[1] Lisa Yerian, M.D.,[1] Thomas W. Rice, M.D.,[2] Lisa A. Rybicki, M.S.,[3] Mary P. Bronner, M.D.,[1] and John R. Goldblum, M.D.[1]
[1]*Cleveland Clinic Departments of Anatomic Pathology,* [2]*Thoracic Surgery, and* [3]*Quantitative Health Sciences, Cleveland, Ohio*

Observer variability has led to need for expert review

All cases of Barrett's dysplasia should be confirmed by a second **EXPERT** gastro-intestinal pathologist



Curvers W. Am J Gastroenterol 2010

Aim of Study

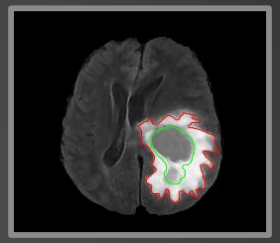
Reliable and Reproducible diagnosis of dysplasia in patients with Barrett's Esophagus

Aim of Study

Reliable and Reproducible diagnosis of dysplasia in patients with Barrett's Esophagus

Deep Learning Methods with promising results in

- Brain Tumor Detection

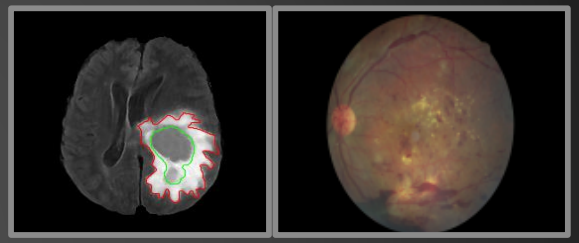


Aim of Study

Reliable and Reproducible diagnosis of dysplasia in patients with Barrett's Esophagus

Deep Learning Methods with promising results in

- Brain Tumor Detection
- Diabetic Retinopathy classification

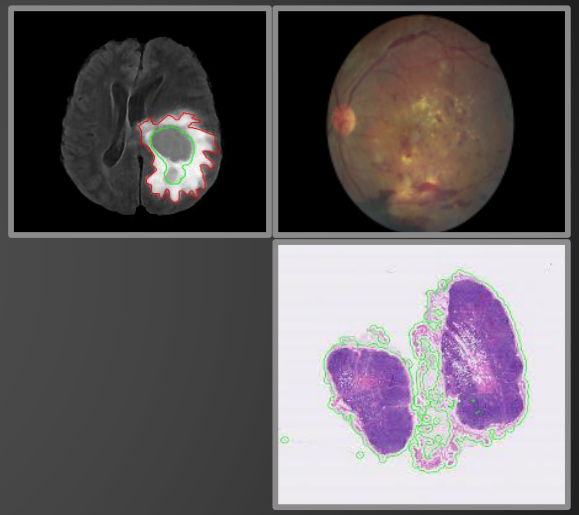


Aim of Study

Reliable and Reproducible diagnosis of dysplasia in patients with Barrett's Esophagus

Deep Learning Methods with promising results in

- Brain Tumor Detection
- Diabetic Retinopathy classification
- Detecting breast cancer metastasis (Camelyon16/17)

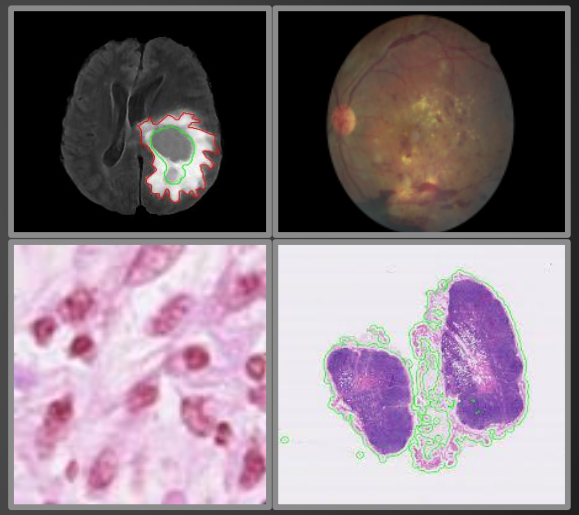


Aim of Study

Reliable and Reproducible diagnosis of dysplasia in patients with Barrett's Esophagus

Deep Learning Methods with promising results in

- Brain Tumor Detection
- Diabetic Retinopathy classification
- Detecting breast cancer metastasis (Camelyon16/17)
- Microscopic nuclei classification

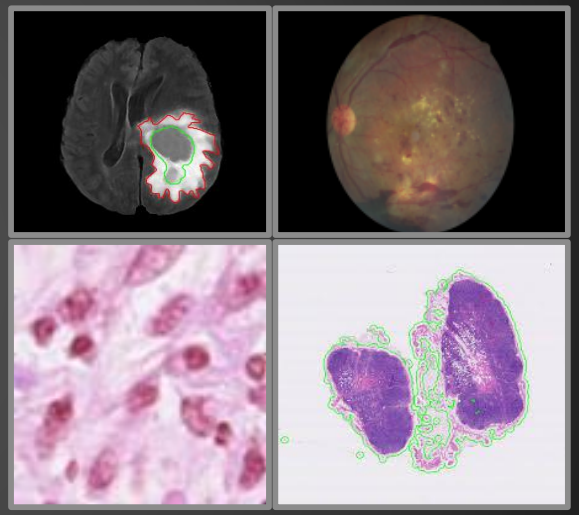


Aim of Study

Reliable and Reproducible diagnosis of dysplasia in patients with Barrett's Esophagus

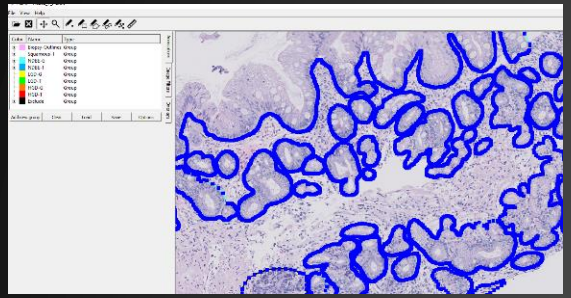
Deep Learning Methods with promising results in

- Brain Tumor Detection
- Diabetic Retinopathy classification
- Detecting breast cancer metastasis (Camelyon16/17)
- Microscopic nuclei classification



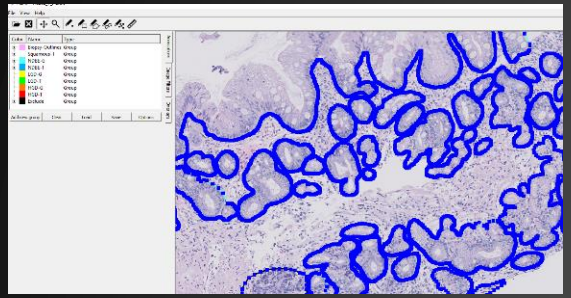
Lack of research in the segmentation of dysplasia - carcinoma sequence in Barrett's Esophagus

The Work Flow

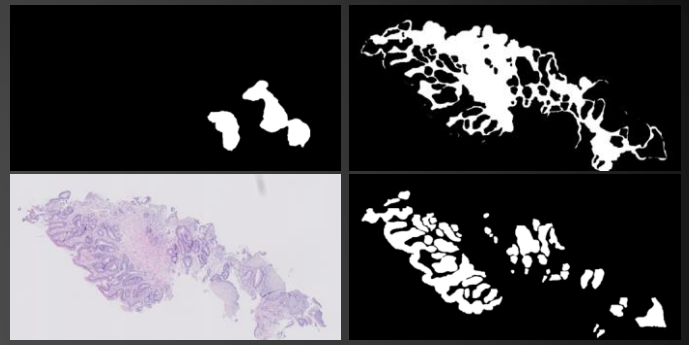


1. Annotate slides in high detail for:
    - Squamous epithelium (Sq)
    - Non dysplastic epithelium (NDBE)
    - Low grade dysplasia (LGD)
    - High Grade dysplasia (HGD)

The Work Flow

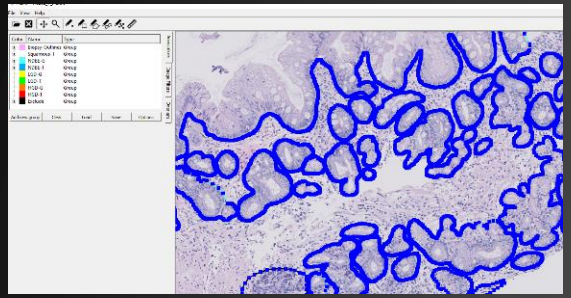


2. Create the masks



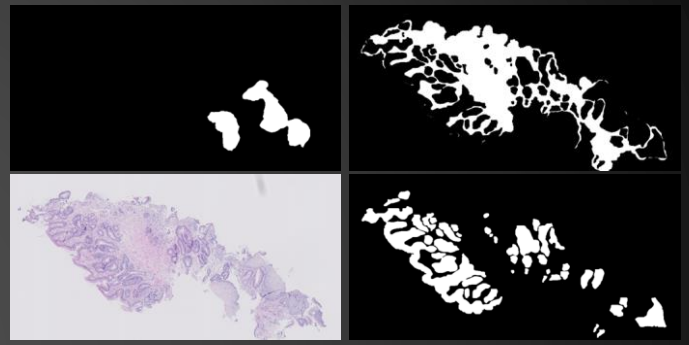
1. Annotate slides in high detail for:
   - Squamous epithelium (Sq)
   - Non dysplastic epithelium (NDBE)
   - Low grade dysplasia (LGD)
   - High Grade dysplasia (HGD)

The Work Flow



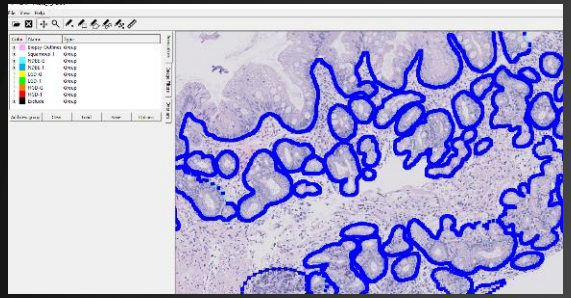
1. Annotate slides in high detail for:
   - Squamous epithelium (Sq)
   - Non dysplastic epithelium (NDBE)
   - Low grade dysplasia (LGD)
   - High Grade dysplasia (HGD)

2. Create the masks



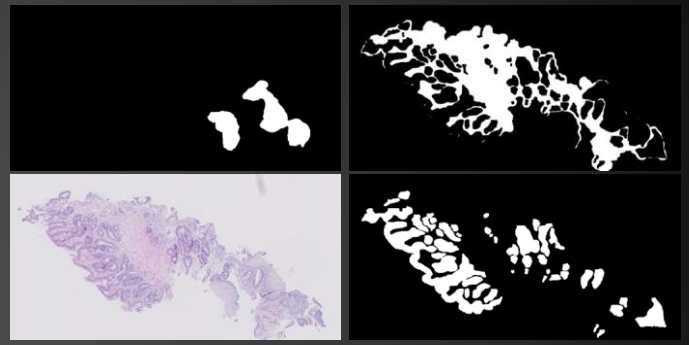
3. Extract Patches (512 x 512)

The Work Flow



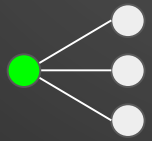
1. Annotate slides in high detail for:
   - Squamous epithelium (Sq)
   - Non dysplastic epithelium (NDBE)
   - Low grade dysplasia (LGD)
   - High Grade dysplasia (HGD)

2. Create the masks

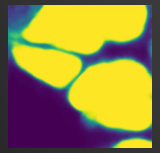


3. Extract Patches (512 x 512)

4. Artificial Neural Network



Predictions



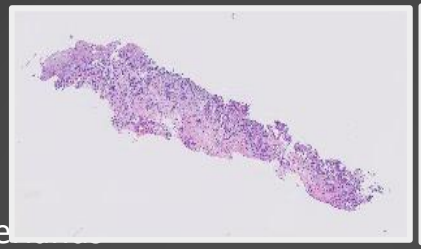
The Data: Whole Slide Image

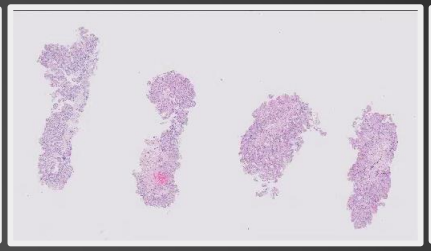
114 H&E stained whole slide images

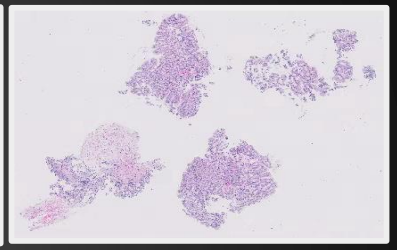
- At 20x magnification:
  - 2 pixels = 1μ
- 94 different patients
- Four Different Labs

  From the Nethe

- ~3 Biopsies per slide







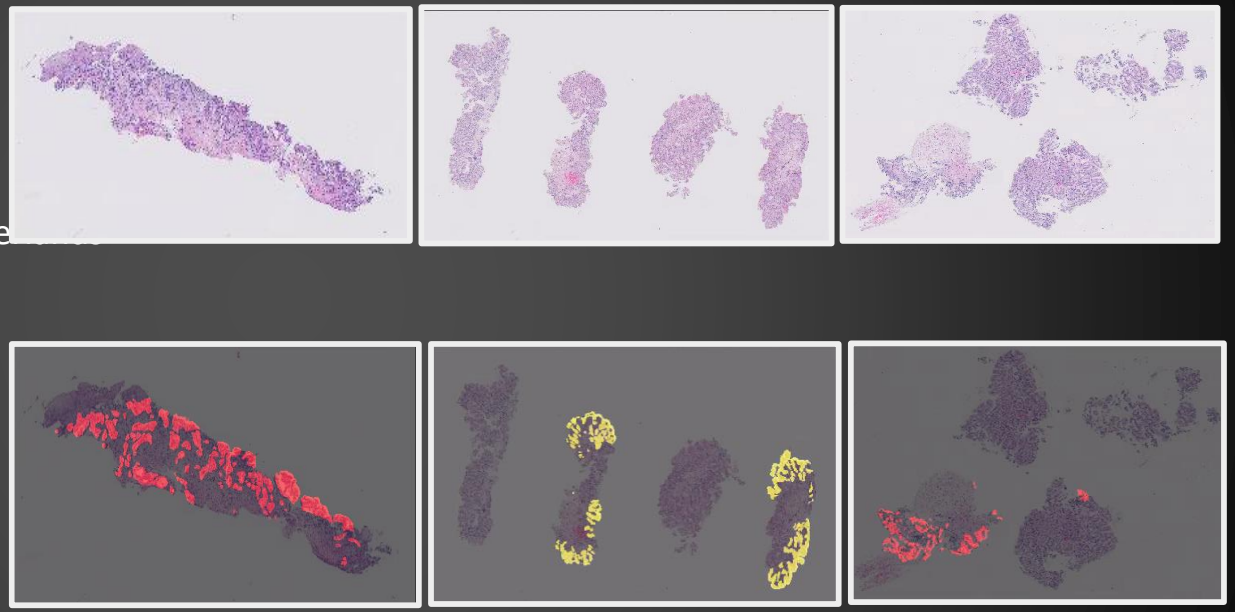
The Data: Whole Slide Image

114 H&E stained whole slide images

- At 20x magnification:
  - 2 pixels = 1μ
- 94 different patients
- Four Different Labs

  From the Nethe

- ~3 Biopsies per slide



High Grade Dysplasia (HGD)

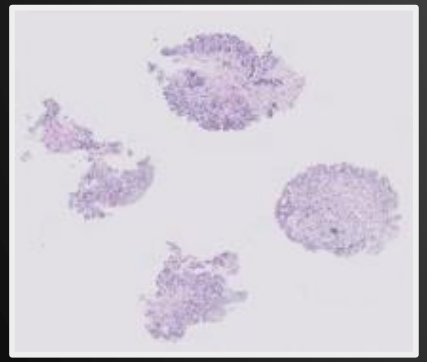
Low Grade Dysplasia (LGD)

Background (B)

The Data: Partitioning

114 whole slide images

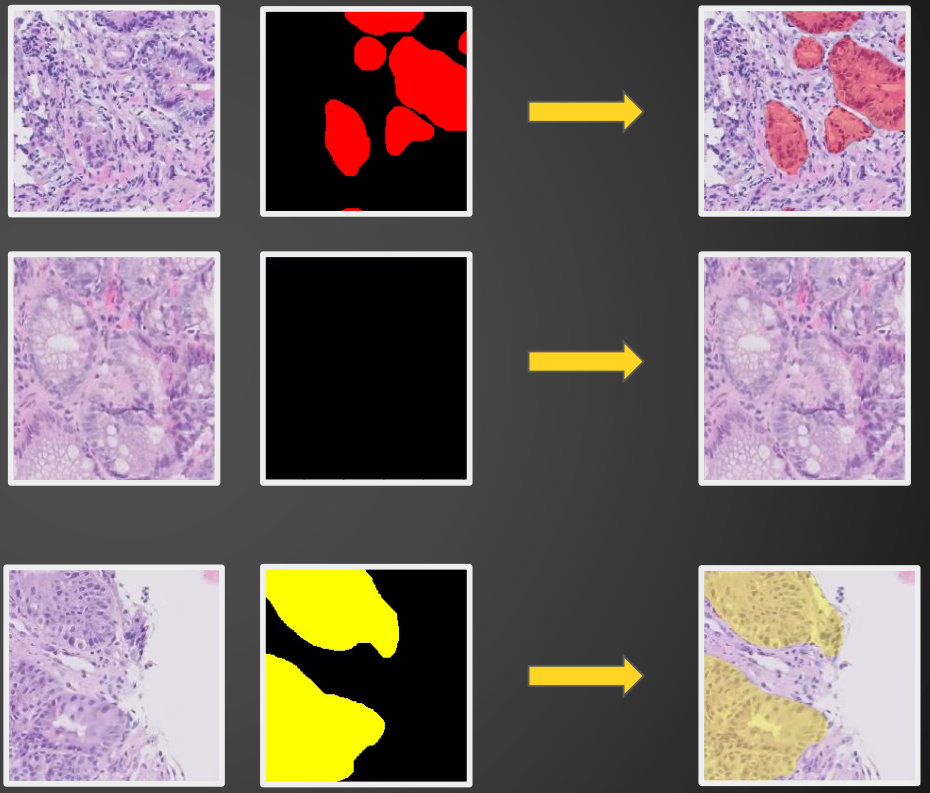
- 93 used for training
- 21 for testing

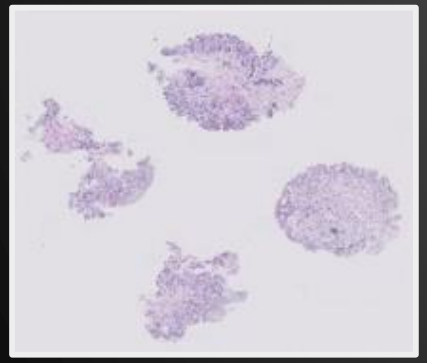


The Data: Partitioning

114 whole slide images

- 93 used for training
- 21 for testing





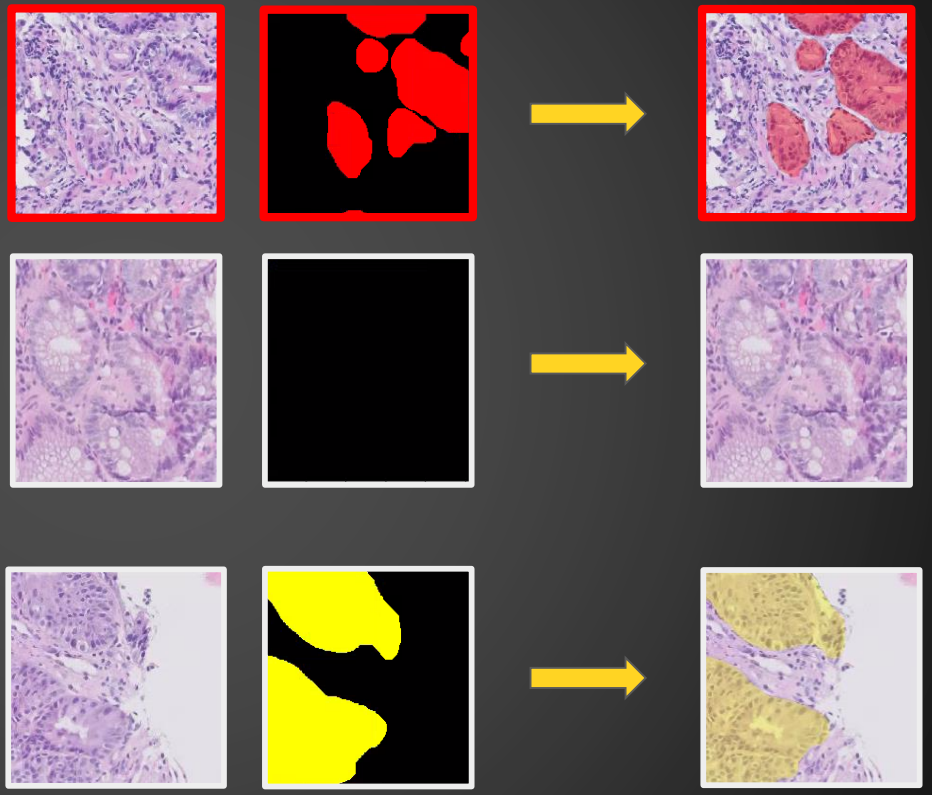
The Data: Partitioning

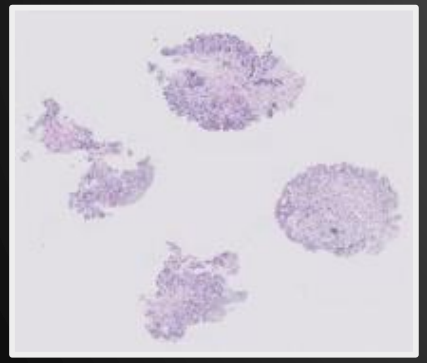
114 whole slide images

- 93 used for training
- 21 for testing

Slicing resulted in ~5000 patches
For training and ~2000 for testing

- 512 x 512 x 3
- One half contains Dysplastic regions





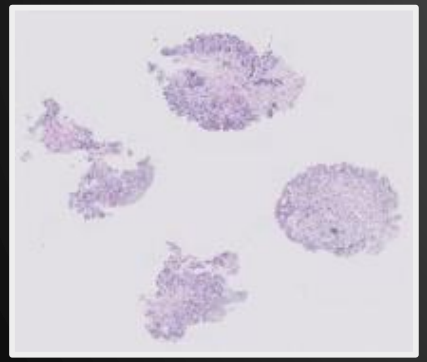
The Data: Partitioning

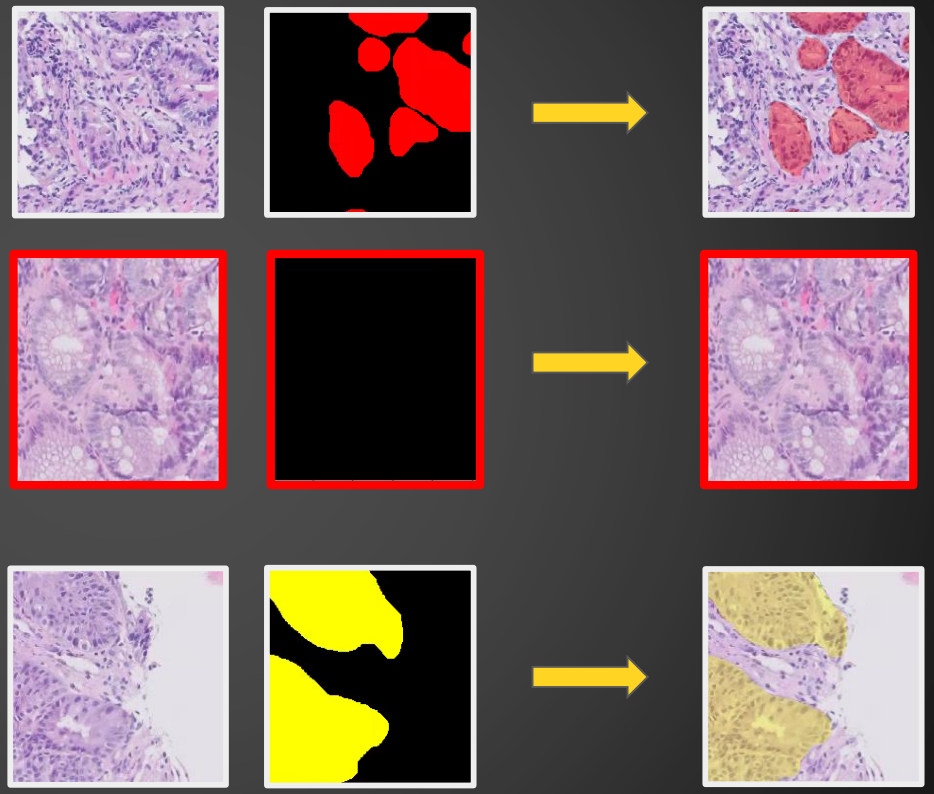
114 whole slide images

- 93 used for training
- 21 for testing

Slicing resulted in ~5000 patches
For training and ~2000 for testing

- 512 x 512 x 3
- One half contains Dysplastic regions

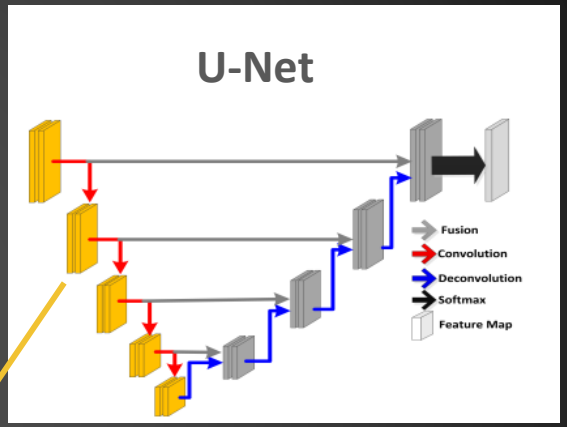




The Neural Network

U-Net Architecture

- Encoding done by DenseNet-161 and ResNet-50



Encoding

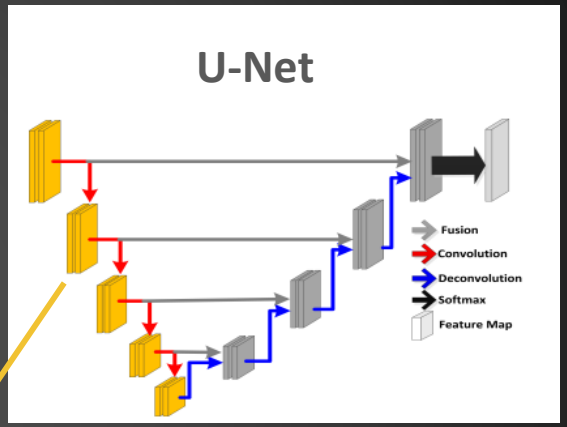
The Neural Network

U-Net Architecture

- Encoding done by DenseNet-161 and ResNet-50

Ensemble of 10 Imagenet pre-trained U-nets

- 5 networks based on DenseNet-161
- 5 networks based on ResNet-50



Encoding

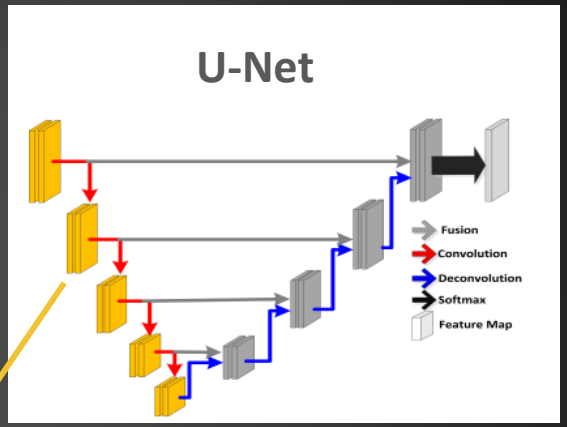
The Neural Network

U-Net Architecture

- Encoding done by DenseNet-161 and ResNet-50

Ensemble of 10 Imagenet pre-trained U-nets

- 5 networks based on DenseNet-161
- 5 networks based on ResNet-50

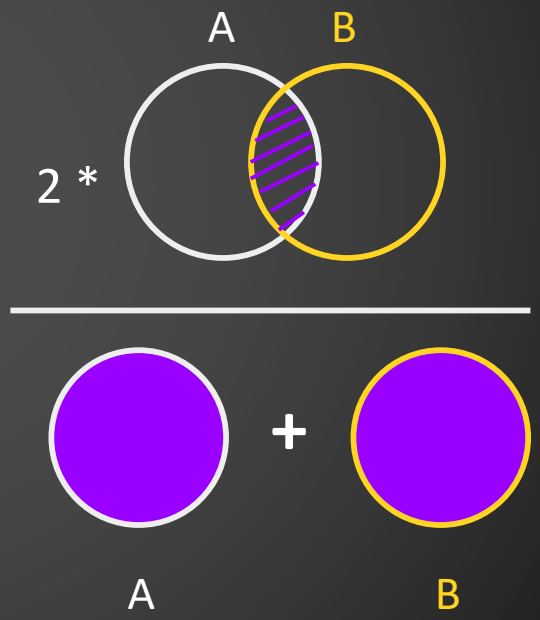


Encoding

Evaluation Metric and Annotations Considered

Main evaluation metric:

$$\text{Dice-Score} = F_1\text{-Score} = \frac{2|A \cap B|}{|A| + |B|}$$

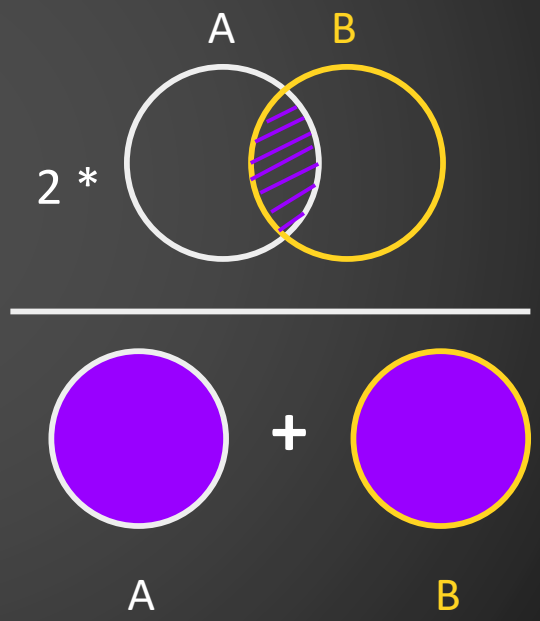


Evaluation Metric and Annotations Considered

Main evaluation metric:

$$\text{Dice-Score} = F_1\text{-Score} = \frac{2|A \cap B|}{|A| + |B|}$$

Weighted average Dice score



Evaluation Metric and Annotations Considered

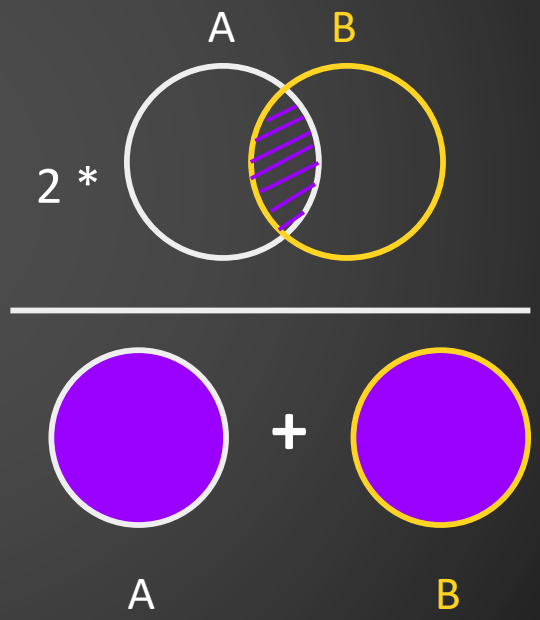
Main evaluation metric:

$$\text{Dice-Score} = F_1\text{-Score} = \frac{2|A \cap B|}{|A| + |B|}$$

Weighted average Dice score

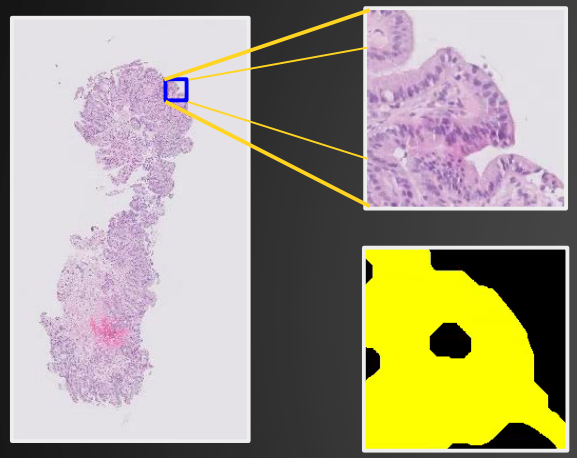
Low Grade Dysplasia vs Background

High Grade Dysplasia vs Background

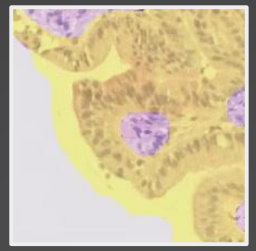


Patch Feeding the Neural Network

Extracted Patch



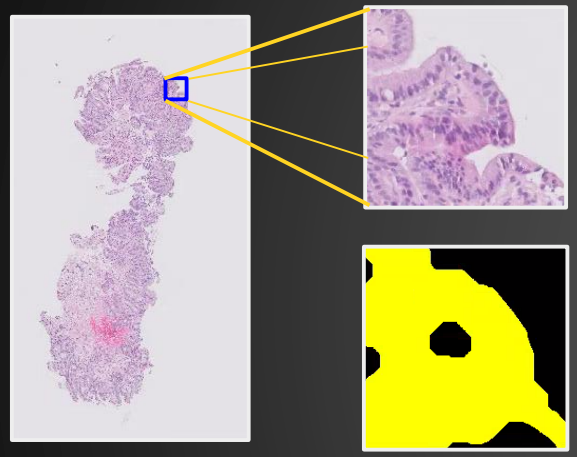
Data Augmentation



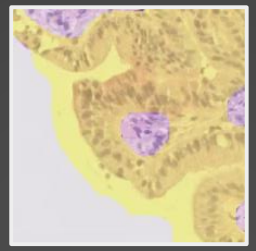
- Rotations
- Flips
- Random Noise
- Random Crops

Patch Feeding the Neural Network

Extracted Patch

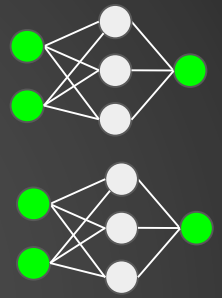


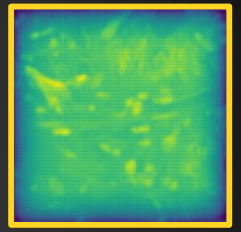
Data Augmentation



- Rotations
- Flips
- Random Noise
- Random Crops

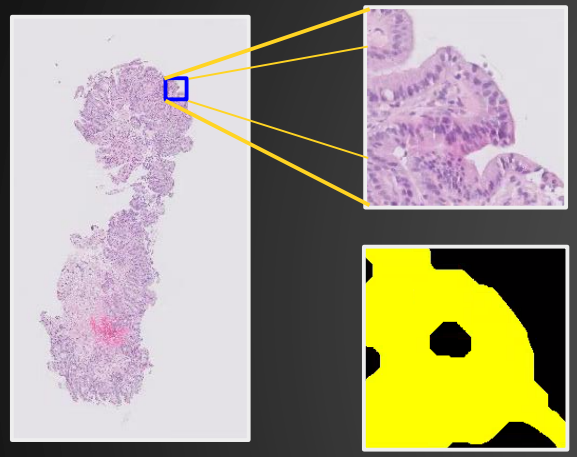
Neural Networks



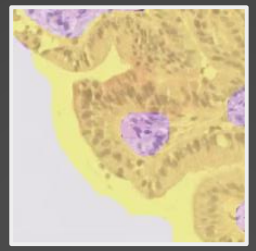


Patch Feeding the Neural Network

Extracted Patch

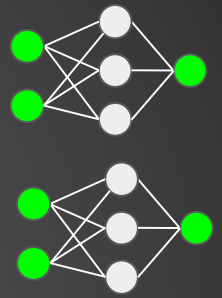


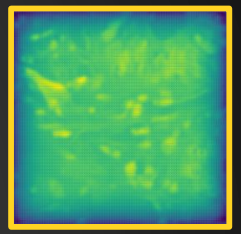
Data Augmentation



- Rotations
- Flips
- Random Noise
- Random Crops

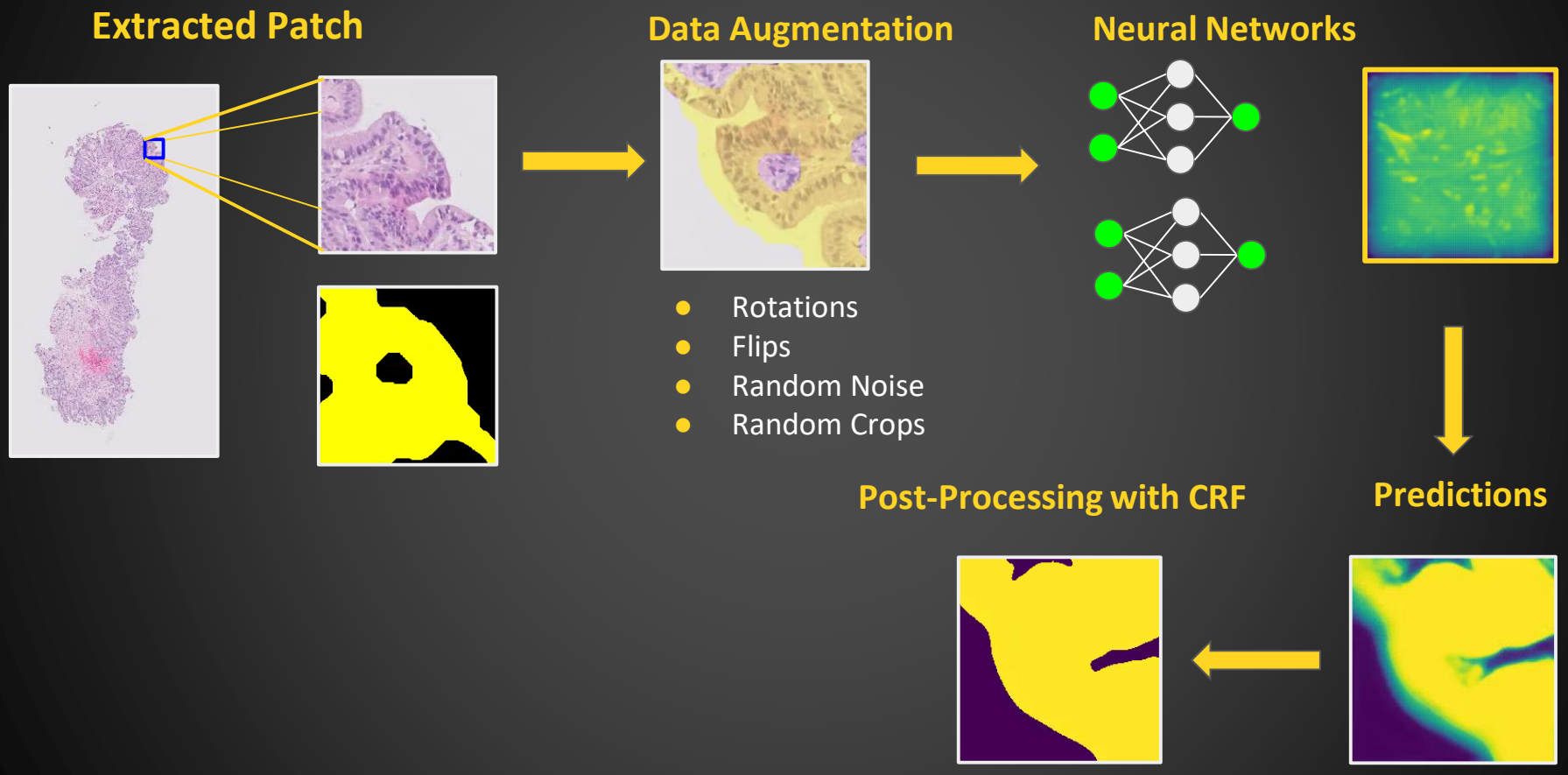
Neural Networks



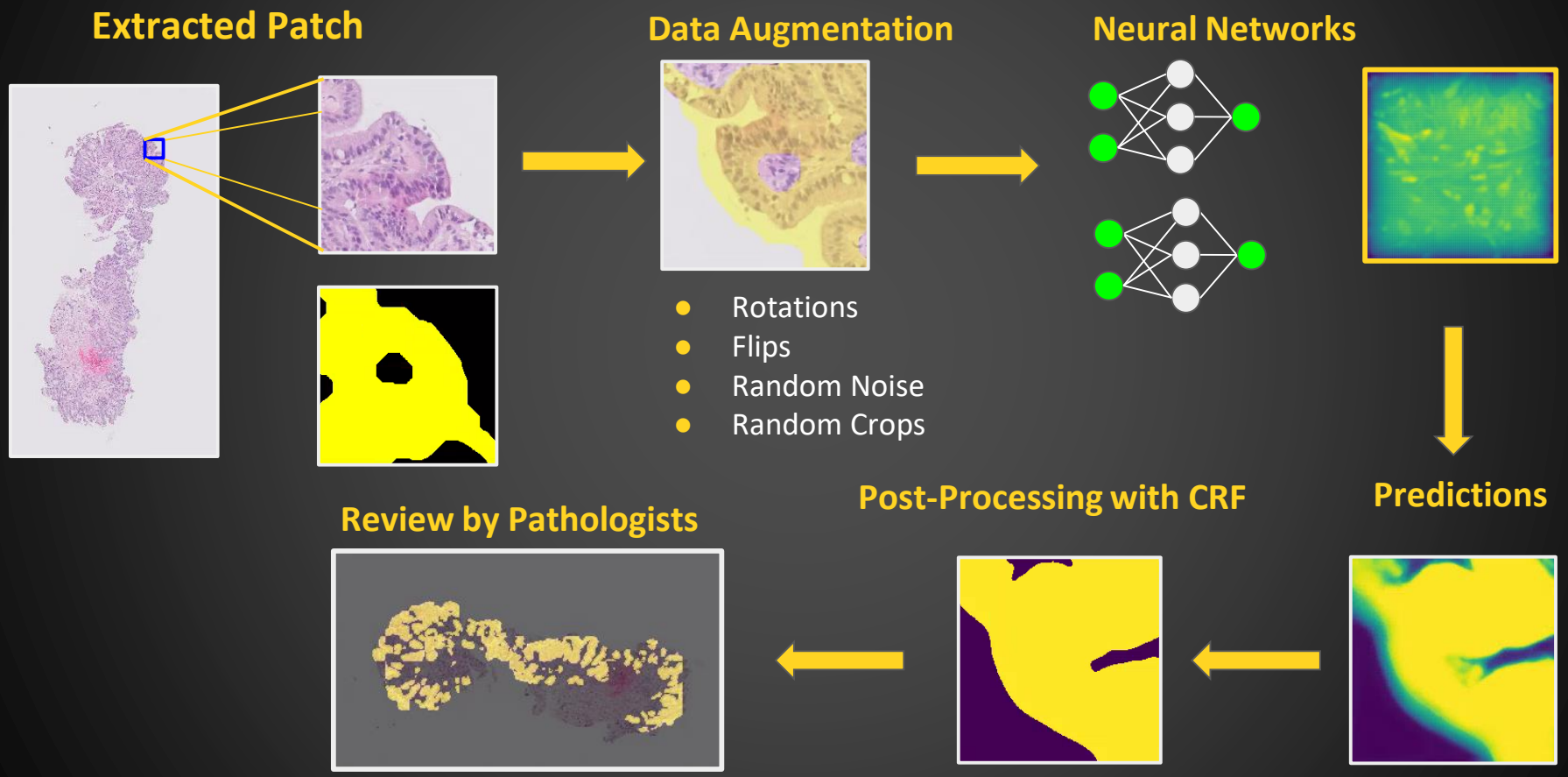


Predictions

Patch Feeding the Neural Network



Patch Feeding the Neural Network

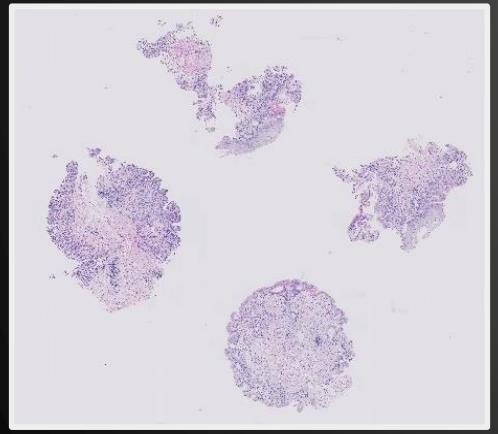


Results: High Grade Dysplasia

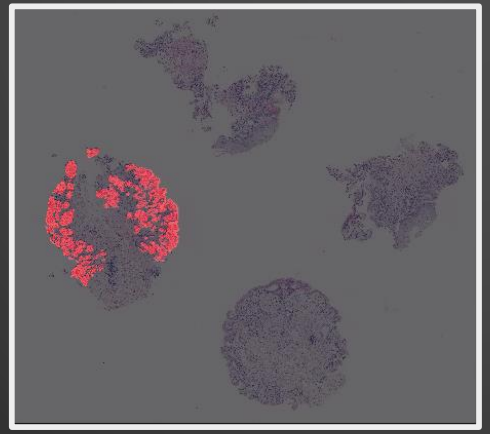
Weighted Dice-Score of **0.83** on the held out set

	F1-Score	Precision	Recall	Pixel-wise accuracy
Background	0.96	0.97	0.95	-
High Grade Dysplasia	0.80	0.76	0.85	-
Weighted Score	0.83	0.8	0.87	0.94

WSI



Mask

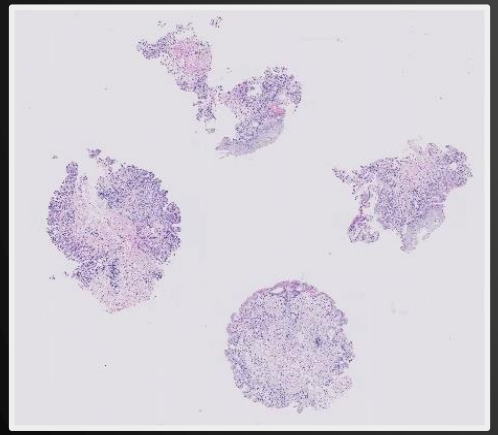


Results: High Grade Dysplasia

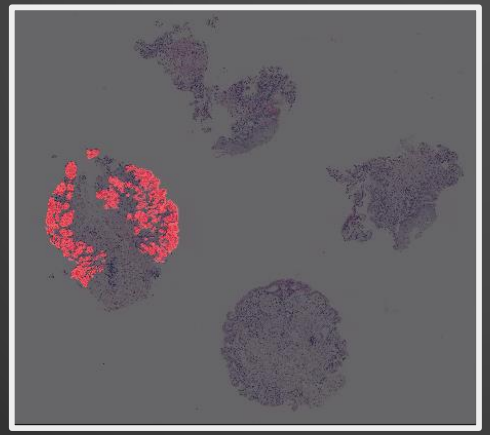
Weighted Dice-Score of **0.83** on the held out set

	F1-Score	Precision	Recall	Pixel-wise accuracy
Background	0.96	0.97	0.95	-
High Grade Dysplasia	0.80	0.76	0.85	-
Weighted Score	0.83	0.8	0.87	0.94

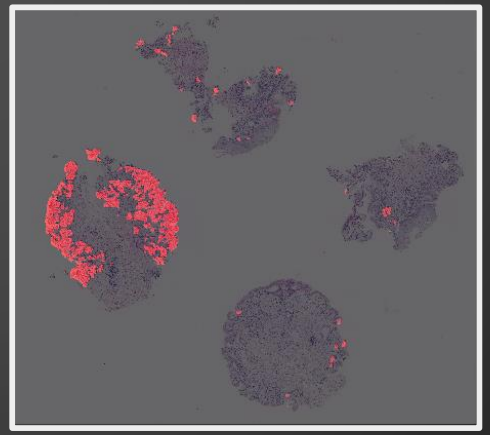
WSI



Mask



Prediction



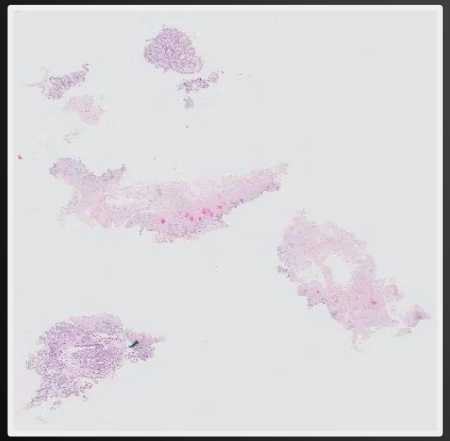
WSI
Weighted Dice-Score:
0.865

Results: Low Grade Dysplasia

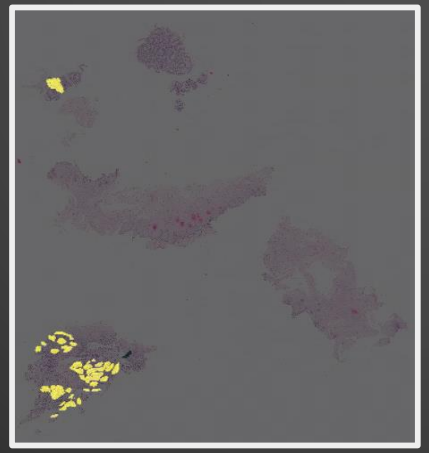
Weighted Dice-Score of **0.74** on the held out set

	F1-Score	Precision	Recall	Pixel-wise accuracy
Background	0.91	0.92	0.90	-
Low Grade Dysplasia	0.67	0.64	0.70	-
Weighted Score	0.74	0.72	0.76	0.86

WSI



Mask

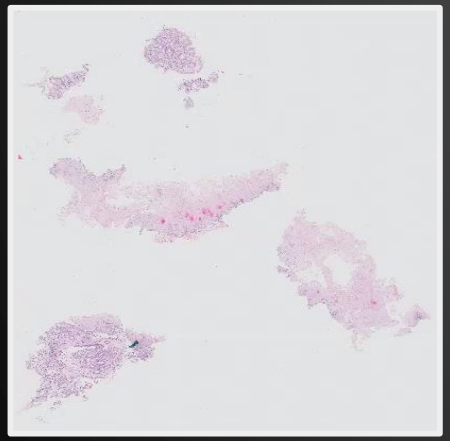


Results: Low Grade Dysplasia

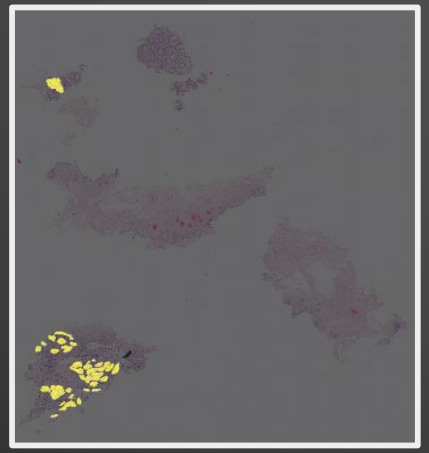
Weighted Dice-Score of **0.74** on the held out set

	F1-Score	Precision	Recall	Pixel-wise accuracy
Background	0.91	0.92	0.90	-
Low Grade Dysplasia	0.67	0.64	0.70	-
Weighted Score	0.74	0.72	0.76	0.86

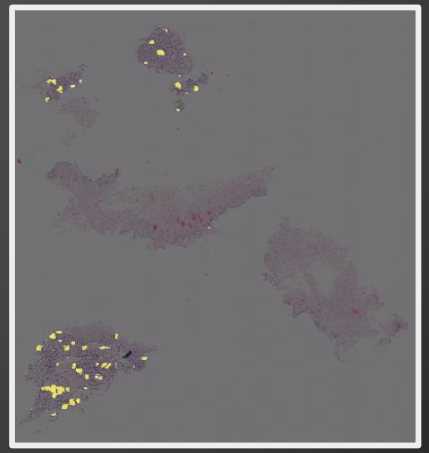
WSI



Mask



Prediction



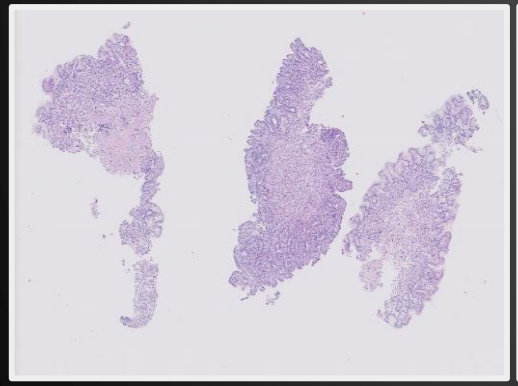
WSI
Weighted Dice-Score:
0.451

Results: Low Grade Dysplasia

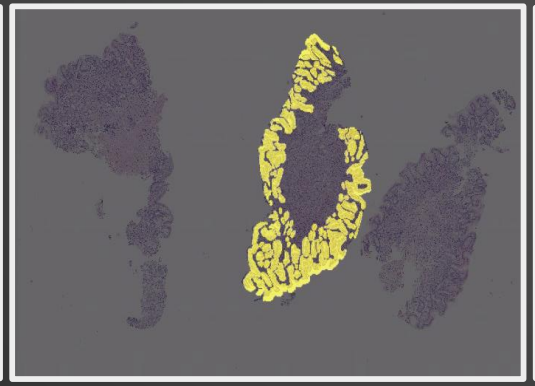
Weighted Dice-Score of **0.74** on the held out set

	F1-Score	Precision	Recall	Pixel-wise accuracy
Background	0.91	0.92	0.90	-
Low Grade Dysplasia	0.67	0.64	0.70	-
Weighted Score	0.74	0.72	0.76	0.86

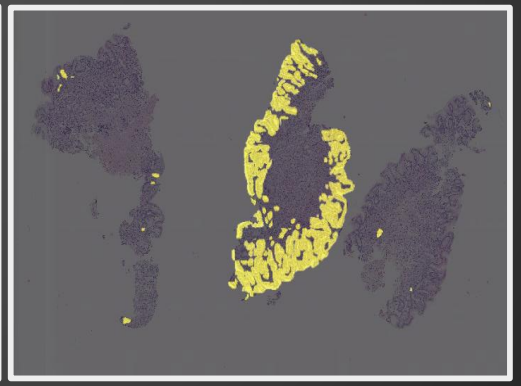
WSI



Mask



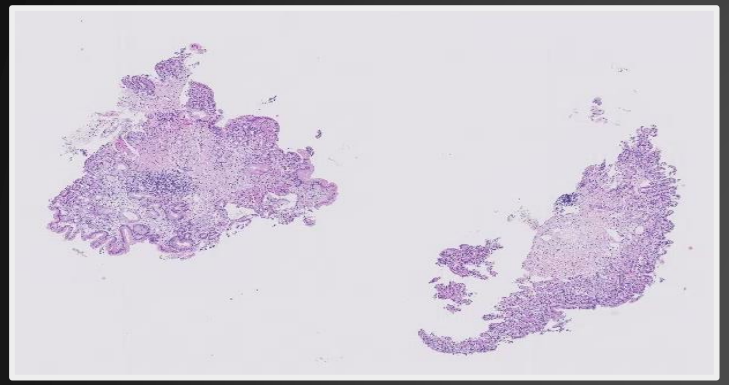
Prediction



WSI
Weighted Dice-Score:
0.94

Variability in the Predictions

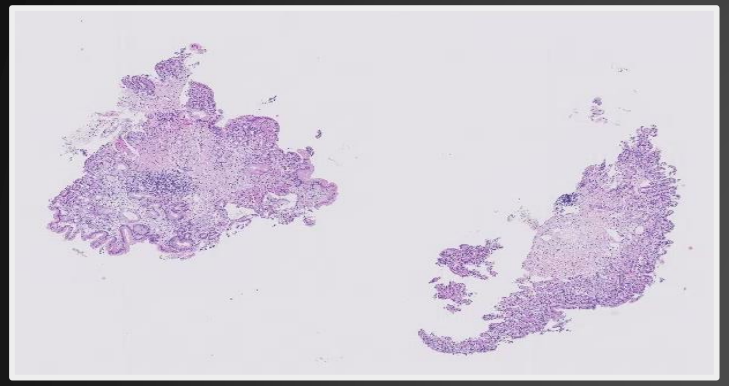
WSI



Mask for High Grade Dysplasia

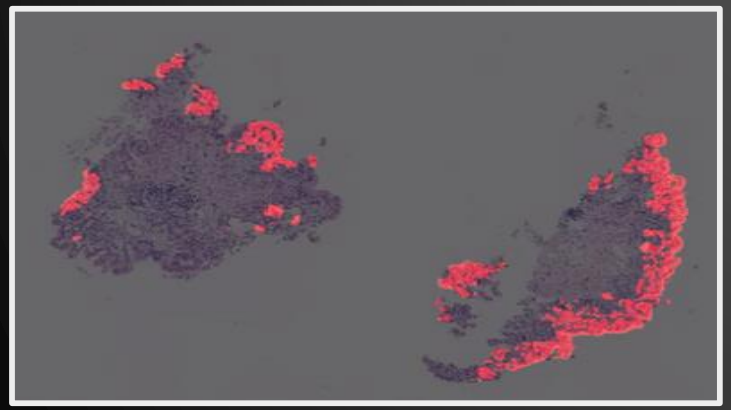
Variability in the Predictions

WSI



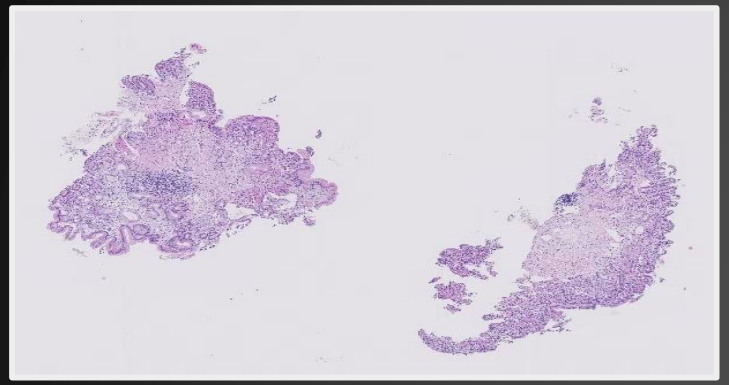
Mask for High Grade Dysplasia

Model Prediction for HGD



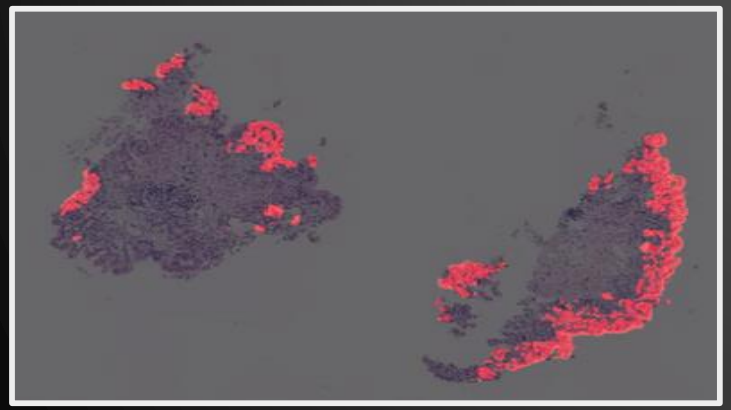
Variability in the Predictions

WSI

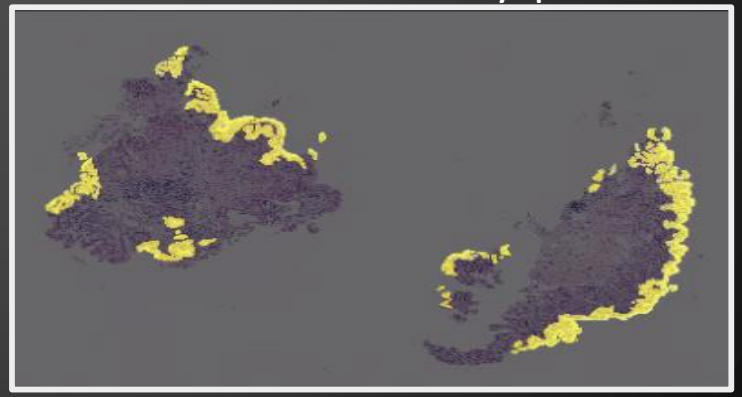


Mask for High Grade Dysplasia

Model Prediction for HGD

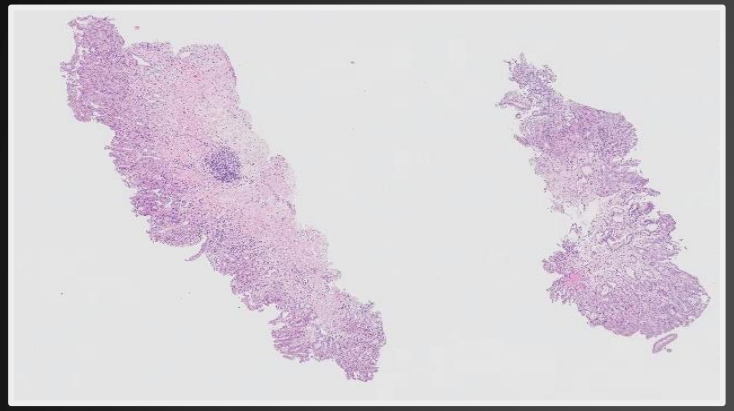


Mask for Low Grade Dysplasia

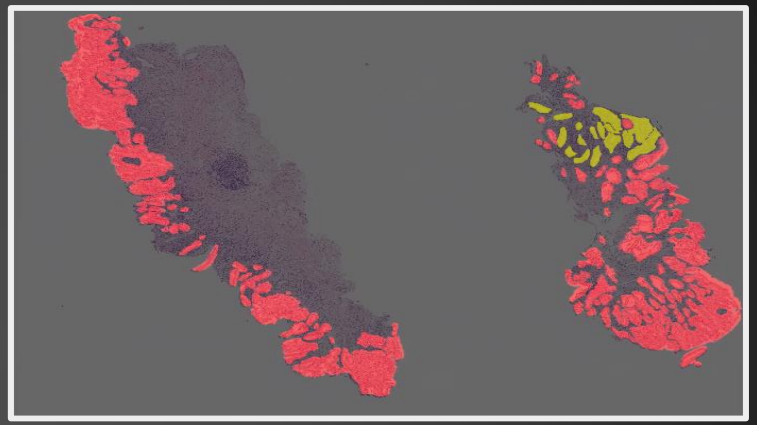


Variability in the Predictions

WSI

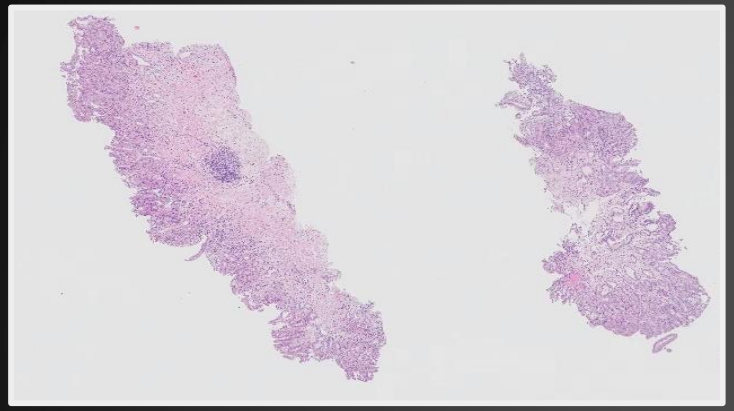


Mask for HGD and LGD

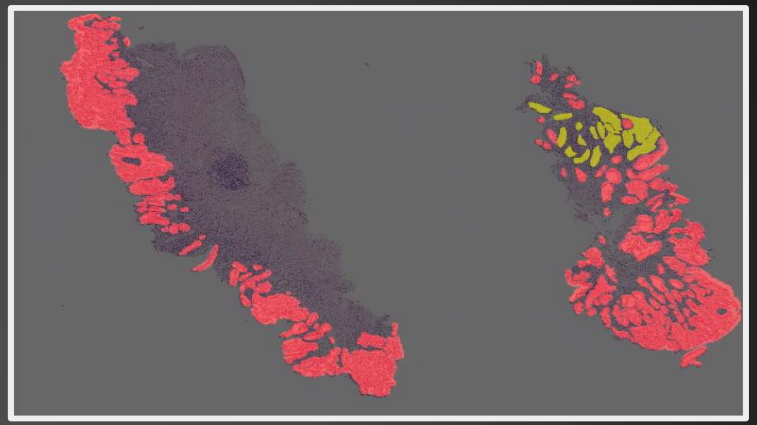


Variability in the Predictions

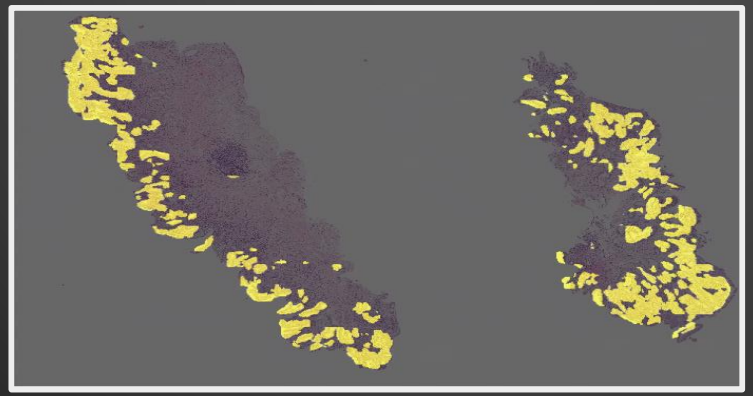
WSI



Mask for HGD and LGD



Model Prediction for LGD



Summarization

Good results with the models trained on HGD

Both models make a lot of false positives

The models have more uncertainty regarding LGD

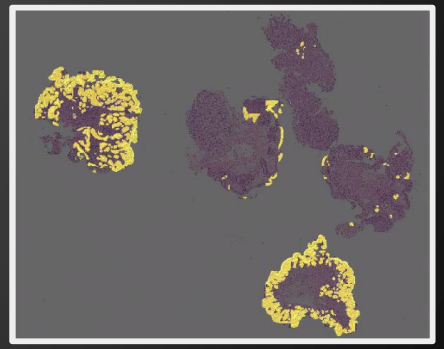
Summarization

Good results with the models trained on HGD

Both models make a lot of false positives

The models have more uncertainty regarding LGD

Visually Useful Predictions

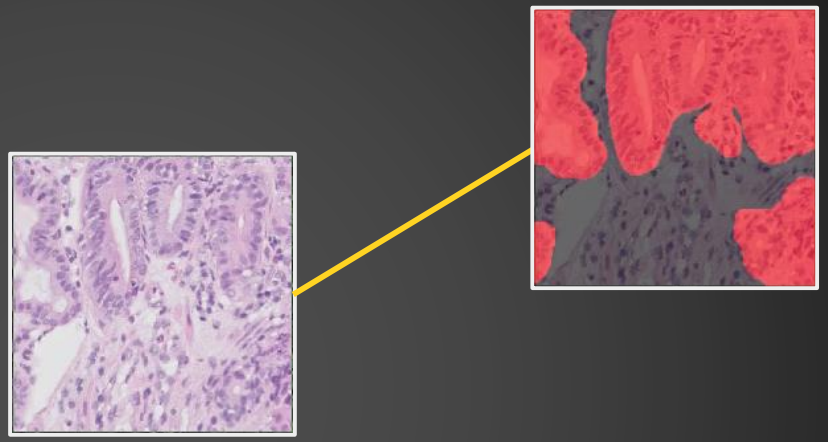


Improvements

Annotations could be coarse

Improvements

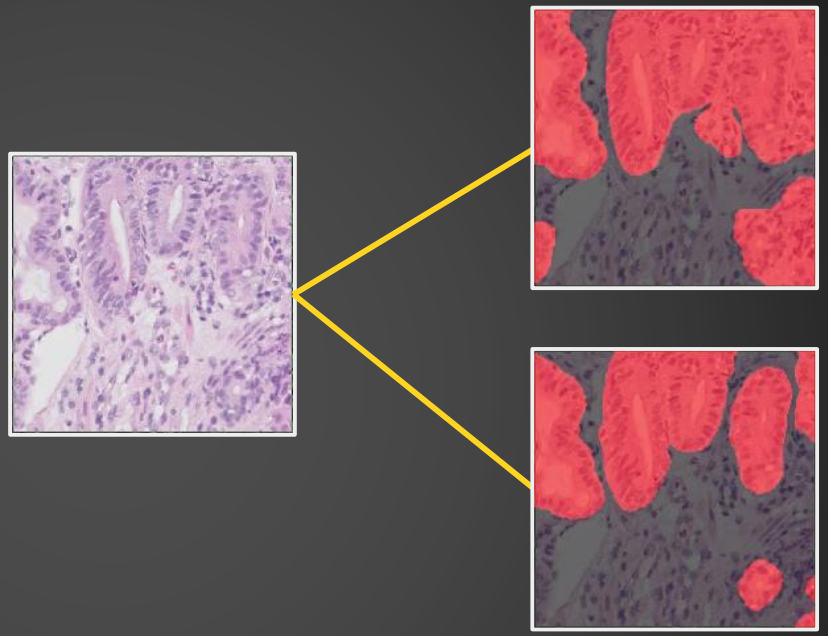
Annotations could be coarse



Mask

Improvements

Annotations could be coarse



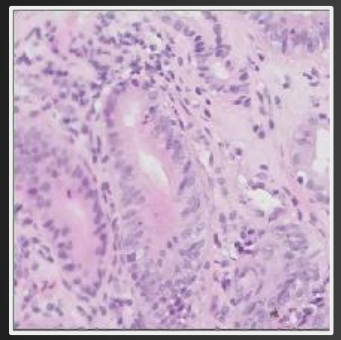
Mask

Prediction

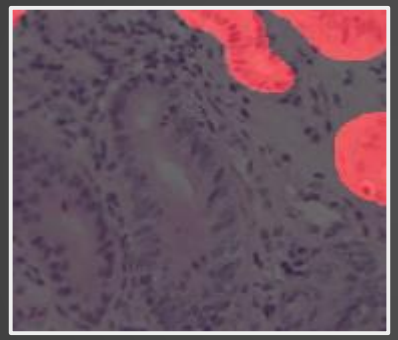
Improvements

Annotations could be coarse

Patch



Mask

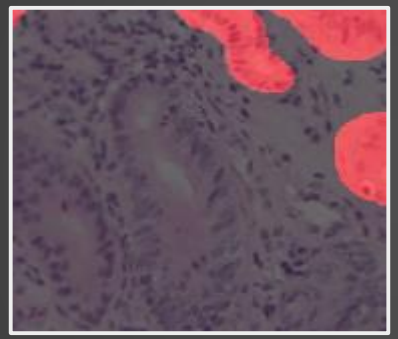


Improvements

Annotations could be coarse

Patch

Mask

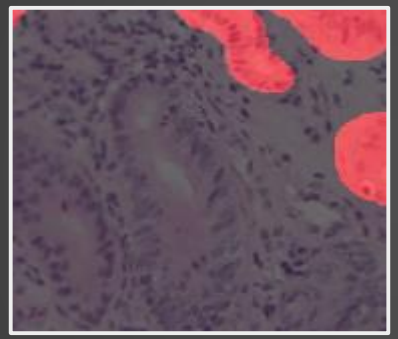


Improvements

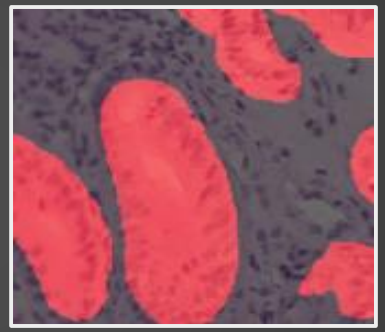
Annotations could be coarse

Patch

Mask



Prediction

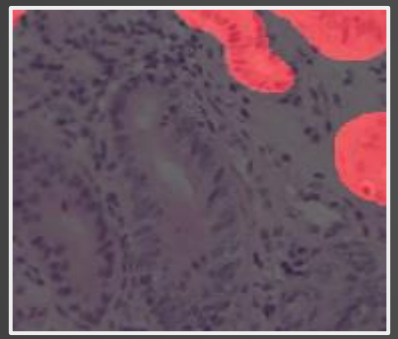


Improvements

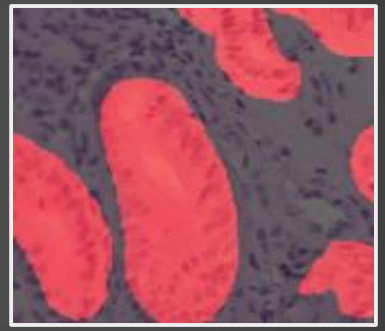
Annotations could be coarse

Patch

Mask



Prediction



More data!

Multiple annotators to isolate regions of interest: Inject Expert Knowledge

References

- Shidan Wang et al, *Pathology Image Analysis Using Segmentation Deep Learning Algorithms*, The American Journal of Pathology, V. 189, 9, 2019

- G Litjens et al: *A survey on deep learning in medical image analysis*, Arxiv 2017

- Olga Russakovsky eta l., ImageNet Large Scale Visual Recognition Challenge, CoRR, 2014

- *Gao Huang et al., Densely Connected Convolutional Networks, CoRR, 2016*

- *Hao Dong et al, automatic Brain Tumor Detection and Segmentation Using U-net Based Fully Convolutional Networks, CoRR 2018*

- *Kaiming He et al., Deep Residual Learning for Image Recognition, CoRR, 2015*

- *S.M.S Islam et al, Deep Learning based Early Detection and grading of Diabetic Retinopathy Using Retinal Fundus Images, MIND, 2019*

- *Yun Liu et al, Artificial intelligence-Based Breast cancer Nodal metastasis Detection, APLM, 2018*

- *Ehsteshami Bejnordi et al., Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women with Breast Cancer, JAMA, 318(22), 2199-2210, 2017*

References

- *Md. Z. Alom et al., Microscopic Nuclei Classification, Segmentation and Detection with improved Deep Convolutional Neural Network Approaches, 2018*

- *Qikui Zhu et al., Deeply-Supervised CNN for Prostate Segmentation, CoRR, 2018*

- *Andreas Maier et al., A gentle introduction to deep learning in medical image processing, zemedi, V. 29, I. 2 86-101, 2019*

- *Shidan Wang et al., Pathology Image Analysis Using Segmenation Deep Learning Algorithms, AJPath, V.189, I.9 1686-1698, 2019*

- *Luis A. de Souza Jr. et al., A survey on Barrett's Esophagus analysis using machine learning, CBM, 96, 203-213, 2018*

- *Olaf Ronneberger et al., U-Net Convolutional Networks for Biomedical Image Segmentation, CoRR, 2015*

- *Tsung Yi Lin et al., Focal Loss for Dense Object Detection, CoRR, 2017*

Acknowledgments

Amsterdam Machine Learning Lab, University of Amsterdam

Bryan Cardenas Guevara
Vivian van Oijen
Daniel Worrall

Dept. of Biomedical Engineering and Physics
Amsterdam University Medical Center

Daniel de Bruin
Ilaria Jansen
Marit Lucas
Henk Marquering

Dept. of Pathology,
Amsterdam University Medical Center

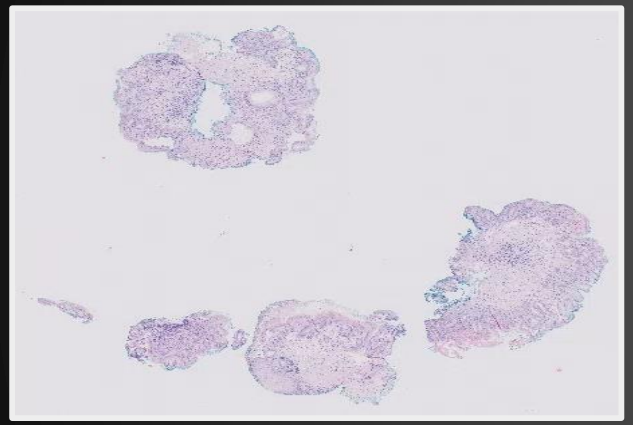
Kiki de Laat
Dilara Savci-Heijink
Myrtle van der Wel
Onno de Boer
Sybren Meijer

THE END

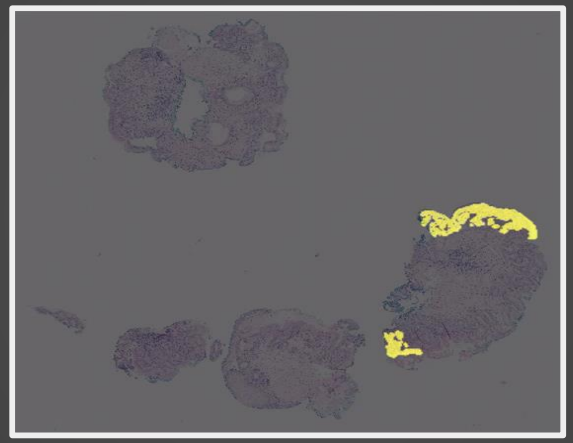
UNUSED SLIDES

Variability in the Predictions

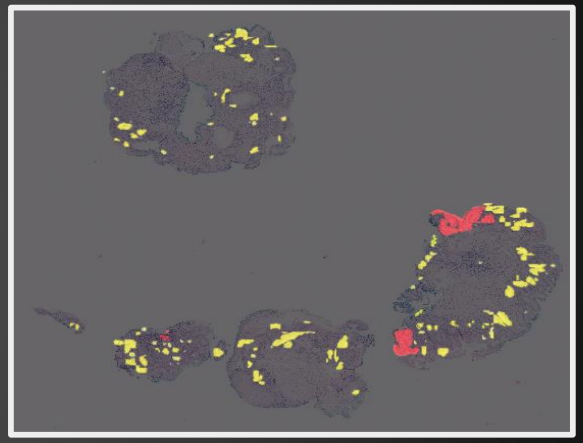
WSI



Mask for HGD and LGD

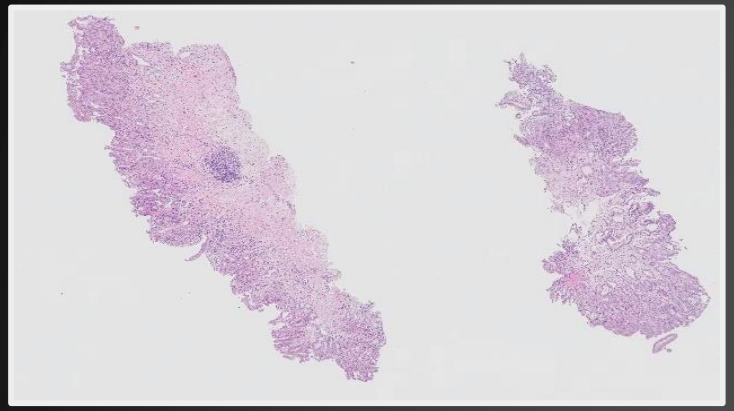


Prediction

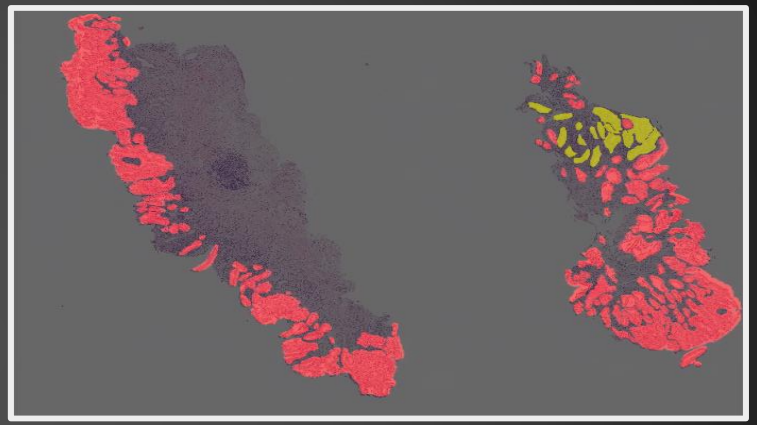


Variability in the Predictions

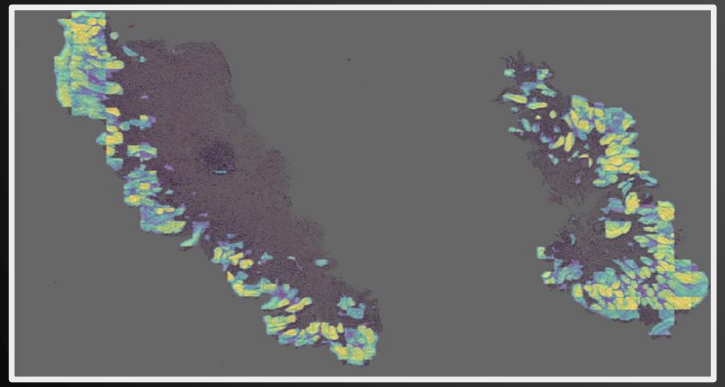
WSI



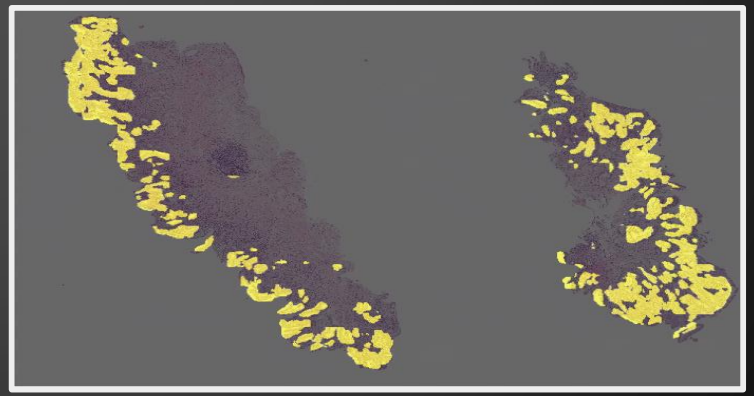
Mask for HGD and LGD



Probabilities for LGD



Thresholded Prediction for LGD

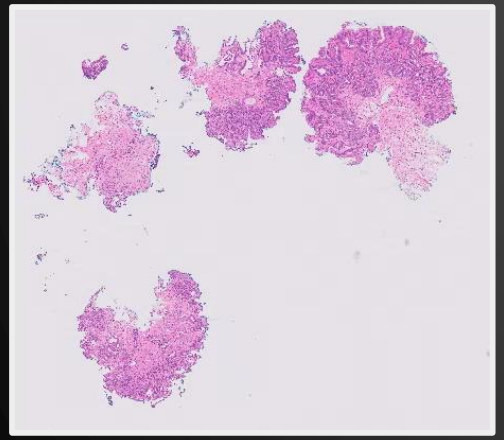


Results: High Grade Dysplasia

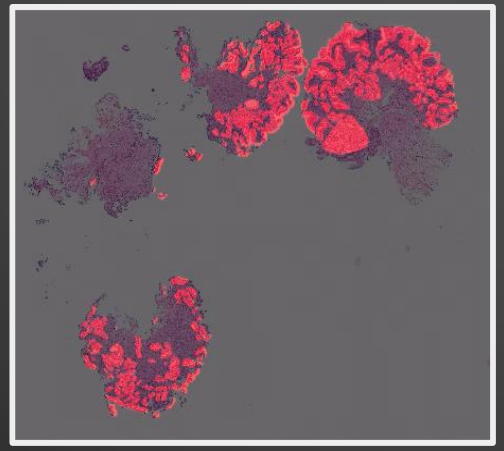
Weighted Dice-Score of **0.83** on the held out set

	F1-Score	Precision	Recall	Pixel-wise accuracy
Background	0.96	0.97	0.95	-
High Grade Dysplasia	0.80	0.76	0.85	-
Weighted Score	0.83	0.8	0.87	0.94

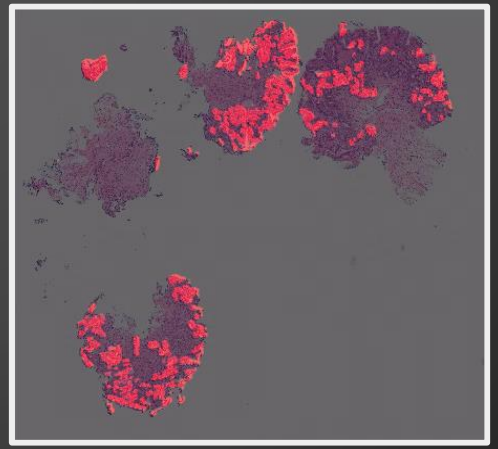
WSI



Mask

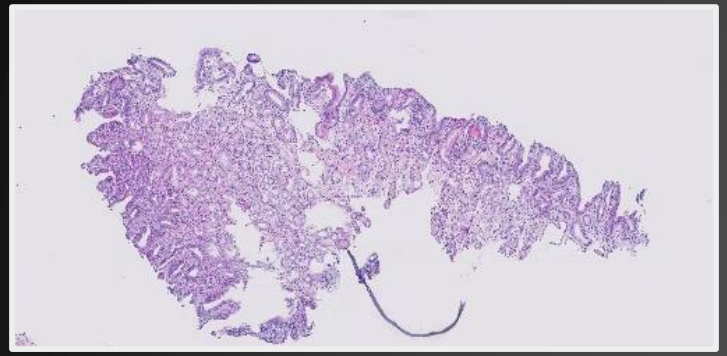


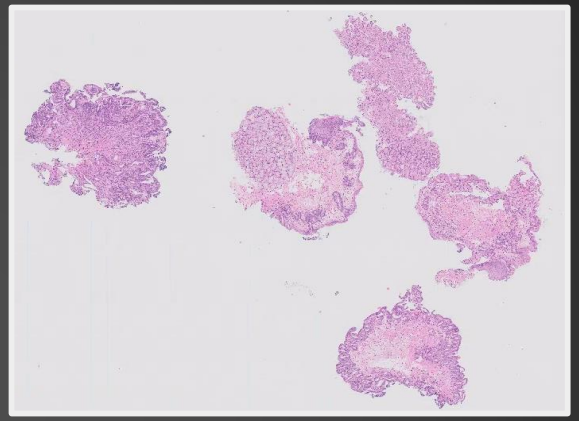
Prediction

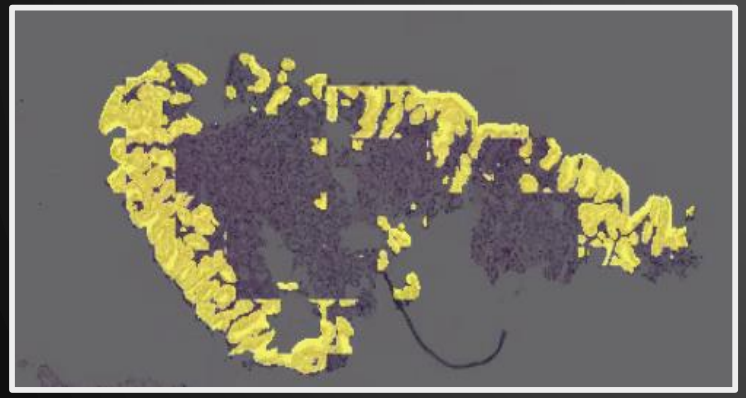


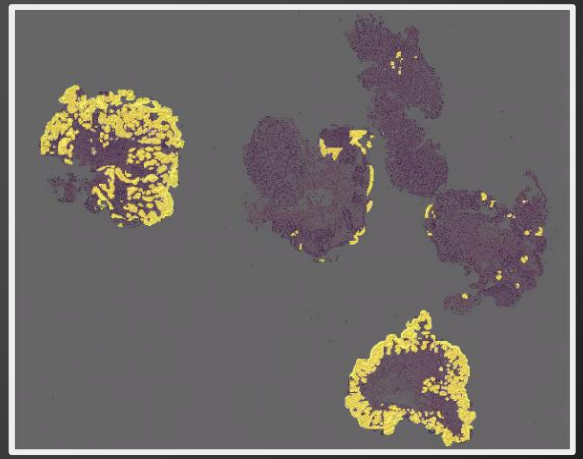
WSI
Weighted Dice-score:
0.754

Remarks



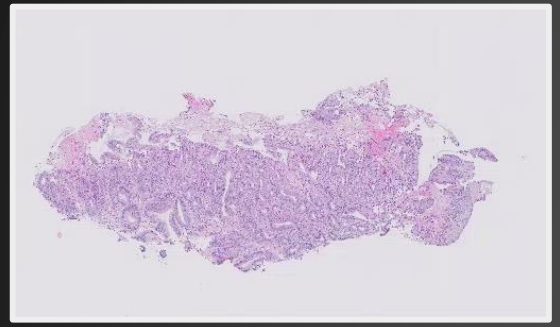




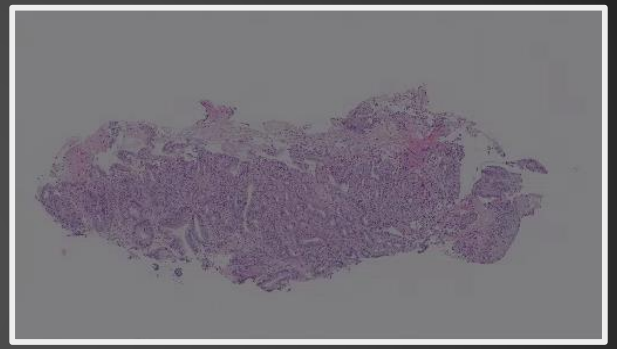


Variability in the Predictions

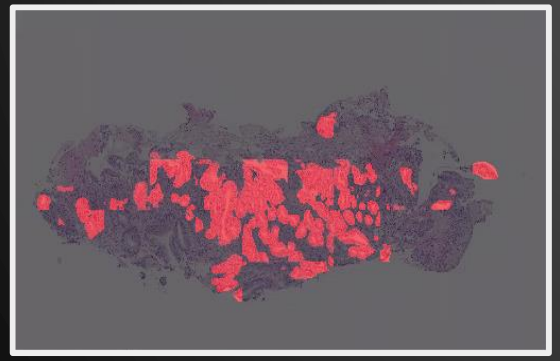
WSI



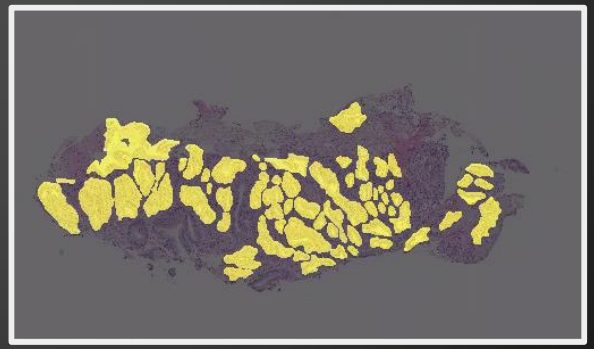
Mask for High Grade Dysplasia



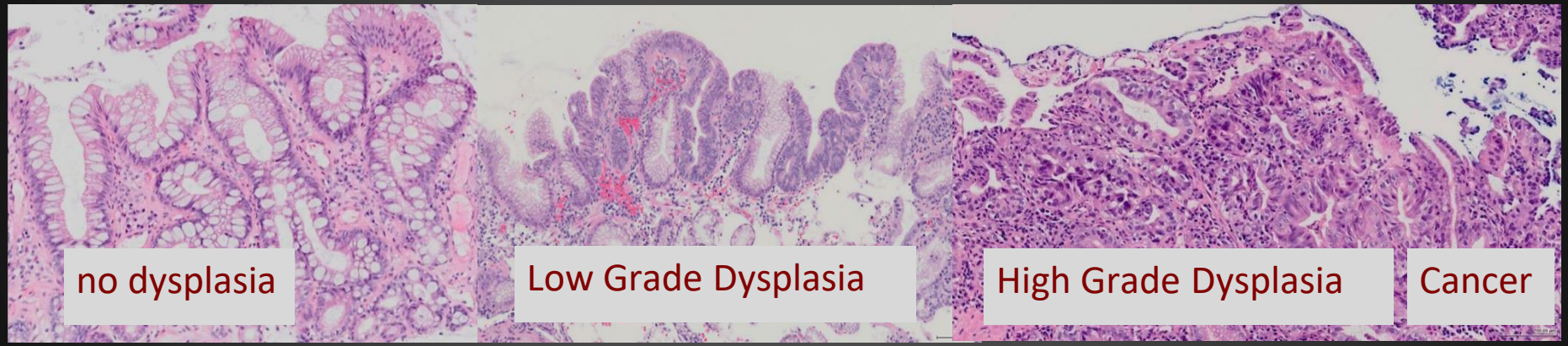
Model Prediction



Mask for Low Grade Dysplasia



Metasplasia - Dysplasia to Adenocarcinoma Sequence



Hvid-Jensen F. NEJM 2011, de Jonge P. Gut 2014; Dutch Barrett guideline

Aim of Study

Reliable and Reproducible diagnosis of dysplasia in patients with Barrett's Esophagus

Deep Learning Methods with promising results in

- Brain Tumor Detection

Hao Dong et al, automatic Brain Tumor Detection and Segmentation Using U-net Based Fully Convolutional Networks, CoRR 2018

- Diabetic Retinopathy classification

S.M.S Islam et al, Deep Learning based Early Detection and grading of Diabetic Retinopathy Using Retinal Fundus Images, MIND, 2019

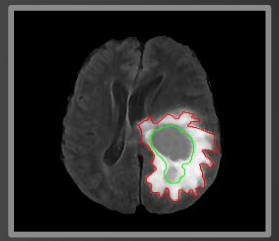
- Detecting breast cancer metastasis (Camelyon16/17)

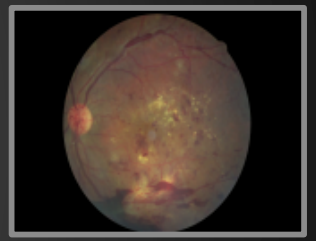
Yun Liu et al, Artificial intellignece-Based Brast cancer Nodal metastasis Detection, APLM, 2018

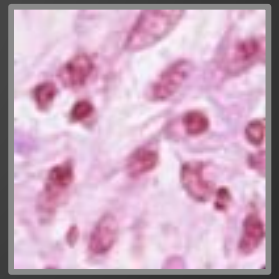
Ehsteshami Bejnordi et al., Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women with Breast Cancer, JAMA, 318(22), 2199-2210, 2017

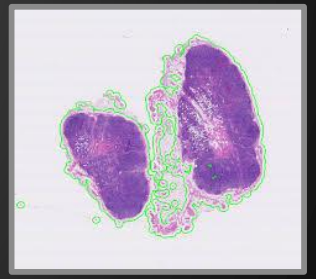
- Microscopic nuclei classification

Md. Z. Alom et al., Microscopic Nuclei Classification, Segmenation and Detection with improved Deep Convolutional Neural Network Approaches, 2018









Lack of research in the segmentation of dysplasia - carcinoma sequence in Barrett's Esophagus

The Neural Network

U-Net Architecture

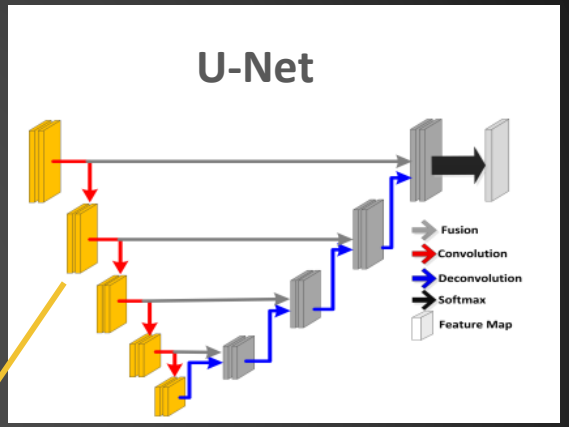
- Encoding done by DenseNet-161 and ResNet-50

Gao Huang et al., Densely Connected Convolutional Networks, CoRR, 2016

Kaiming He et al., Deep Residual Learning for Image Recognition, CoRR, 2015

Ensemble of 10 Imagenet pre-trained U-nets

- 5 networks based on DenseNet-161
- 5 networks based on ResNet-50



Encoding

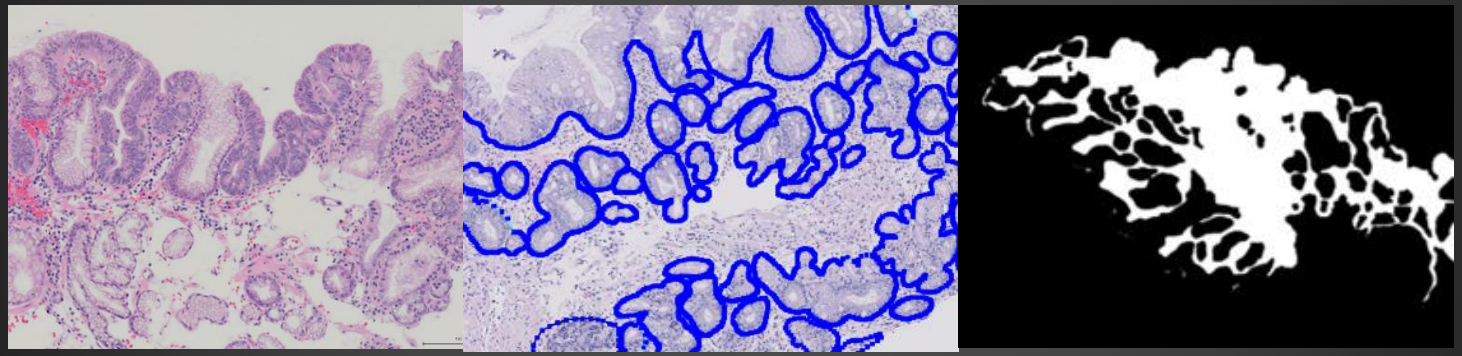
Qikui Zhu et al., Deeply-Supervised CNN for Prostate Segmentation, CoRR, 2018

Olaf Ronneberger et al., U-Net Convolutional Networks for Biomedical Image Segmentation, CoRR, 2015

Cross Entropy Loss and Focal Loss

Tsung Yi Lin et al., Focal Loss for Dense Object Detection, CoRR, 2017

Annealed Learning rate between 2e-5 and 8e-5



Results: Full Dysplasia

Weighted Dice-Score of **0.77** on the held out set

	F1-Score	Precision	Recall	Pixel-wise accuracy
Background	0.95	0.99	0.91	-
Dysplasia	0.69	0.56	0.90	-
Weighted Score	0.77	0.69	0.90	0.91