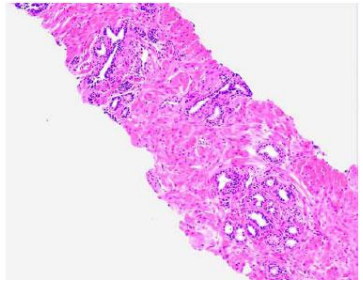
Automatic Prostate Cancer Classification using Deep Learning

Ida Arvidsson
Centre for Mathematical Sciences, Lund University, Sweden

Outline

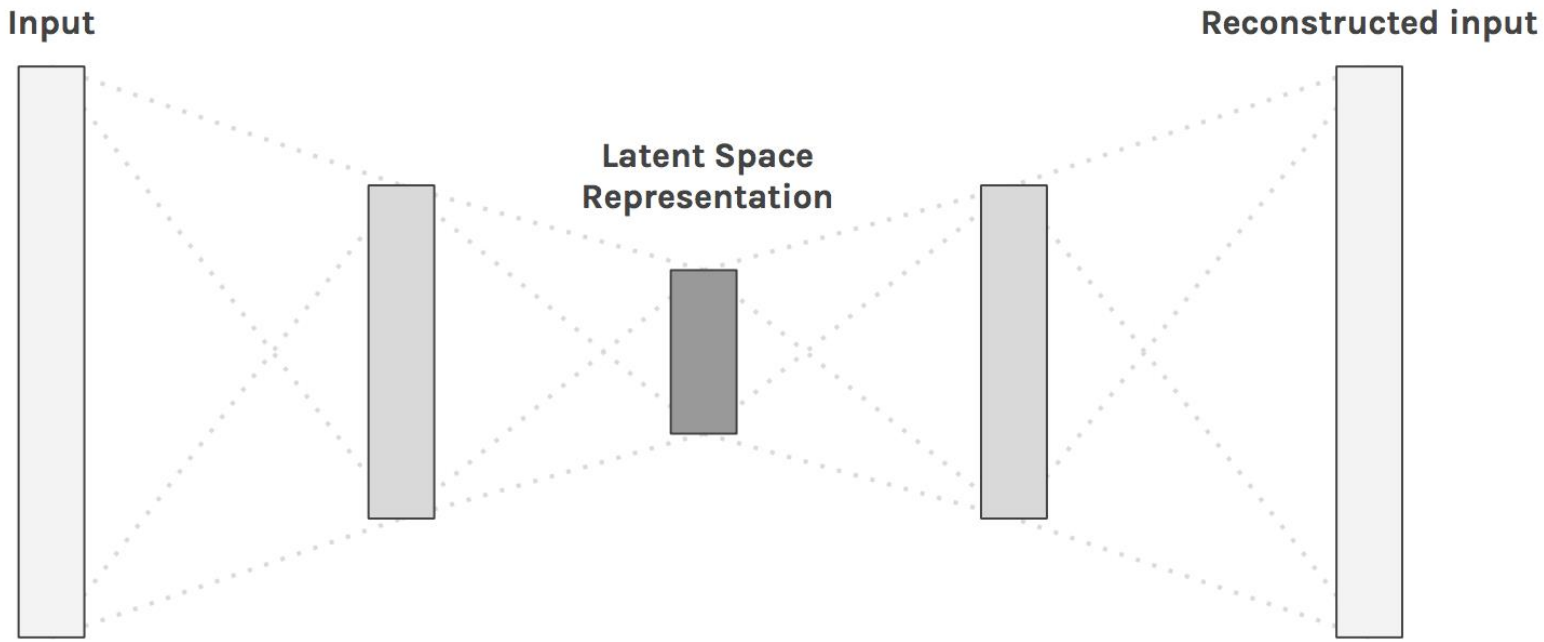
- Autoencoders, theory
- Motivation, background and goal for prostate project
- Data, problems specific to digital pathology
- Classification of patches
  - Augmentation techniques
  - Attempts to improve the generalization performance, using e.g. autoencoders and cycle GANs
- Semantic segmentation



LUND UNIVERSITY

Autoencoder

- Neural networks that aim to reproduce their input
  - Trivial if no constraints on the network
  - Interesting when the network is constrained

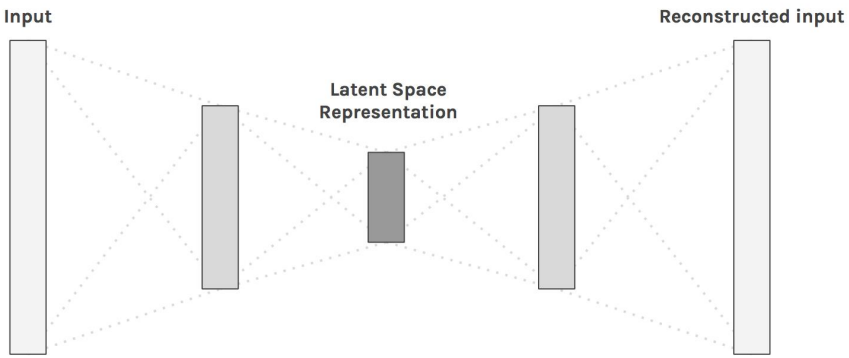


LUND UNIVERSITY

Autoencoder

- Find a useful encoding, $h = f(x)$, of data x in an unsupervised manner
- Train using the encoder $h = f(x)$ and a decoder $\hat{x} = g(h)$
- Loss is some measure comparing the input to the reconstruction; $L(x, \hat{x})$

- If f and g are linear and L is the mean squared error, the autoencoder learns to span the same subspace as PCA
- If they are nonlinear, they can learn a more powerful nonlinear generalization of PCA



LUND UNIVERSITY

Autoencoder, variants

To avoid that the autoencoder learns the identity function

- Undercomplete autoencoder: make sure h has a smaller dimension than x
- Denoising autoencoder: corrupt the data, or activations h, with random noise
- Regularized autoencoder: add regularizing term to the cost function, e.g. $\|h\|_1$
- Contractive autoencoder: add regularizing term to the cost function, e.g. $\|\Delta_x h\|$, i.e. the code h becomes invariant to small perturbations in x

Other variants

- Variational autoencoder: generative model

Autoencoder, application

Geoffrey E Hinton and Ruslan R Salakhutdinov. “Reducing the dimensionality of data with neural networks”. In: Science 313.5786 (2006), pp. 504–507.

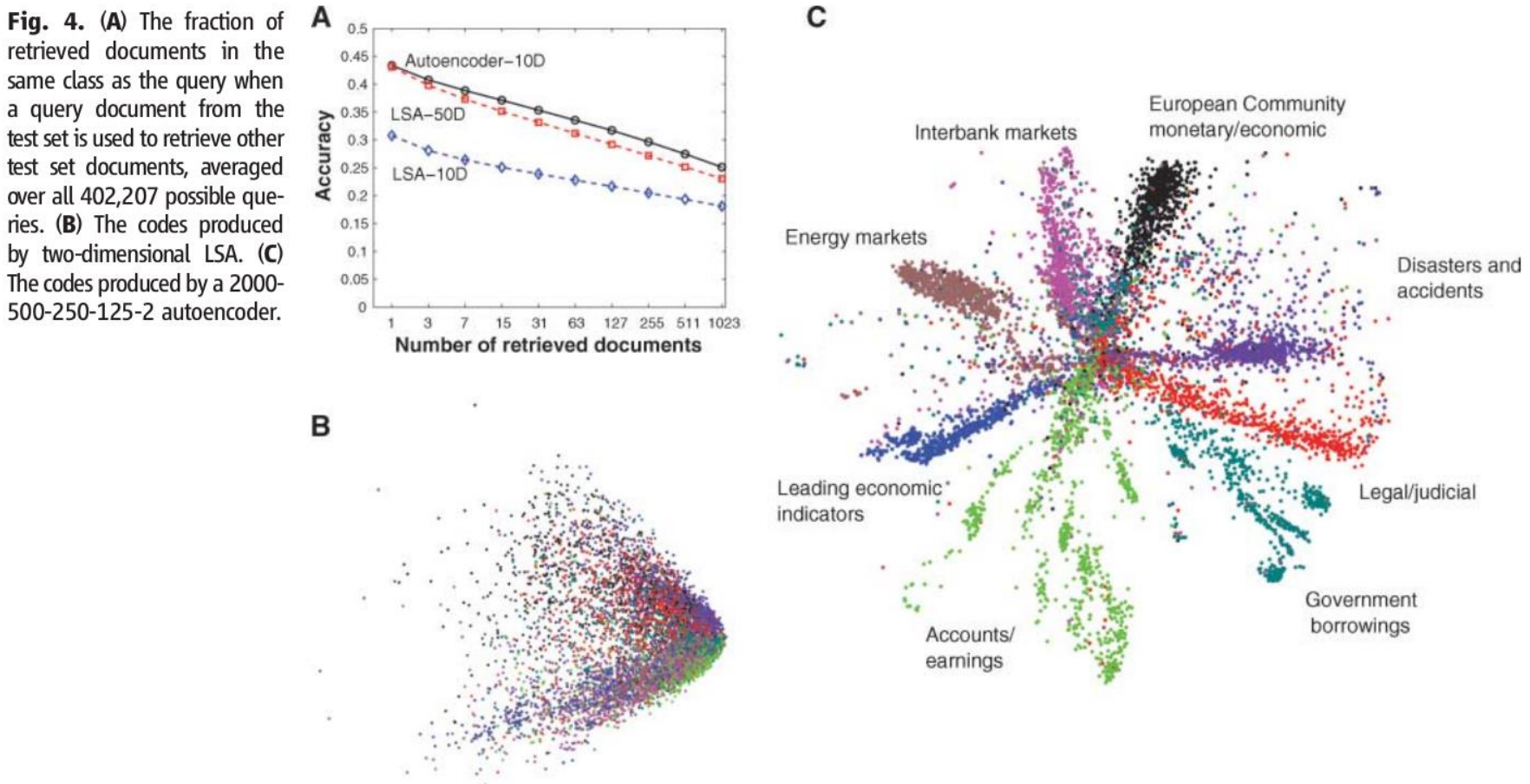


Fig. 4. (**A**) The fraction of retrieved documents in the same class as the query when a query document from the test set is used to retrieve other test set documents, averaged over all 402,207 possible queries. (**B**) The codes produced by two-dimensional LSA. (**C**) The codes produced by a 2000-500-250-125-2 autoencoder.

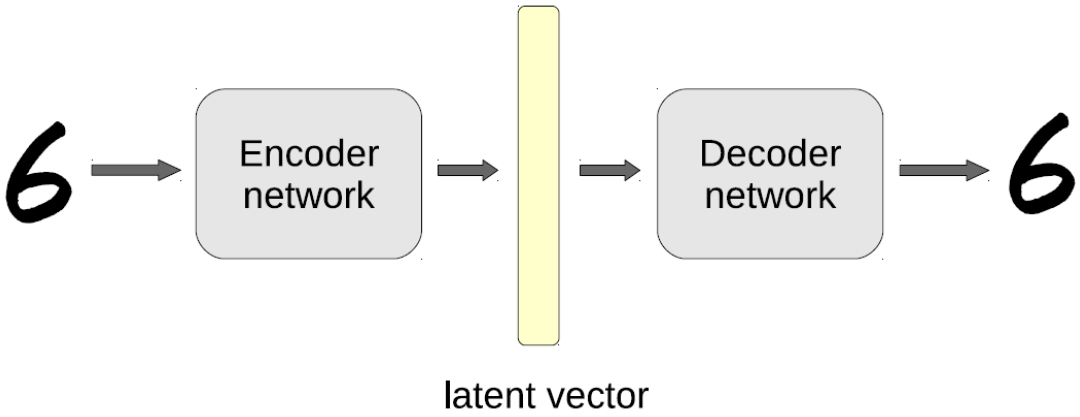
LUND UNIVERSITY

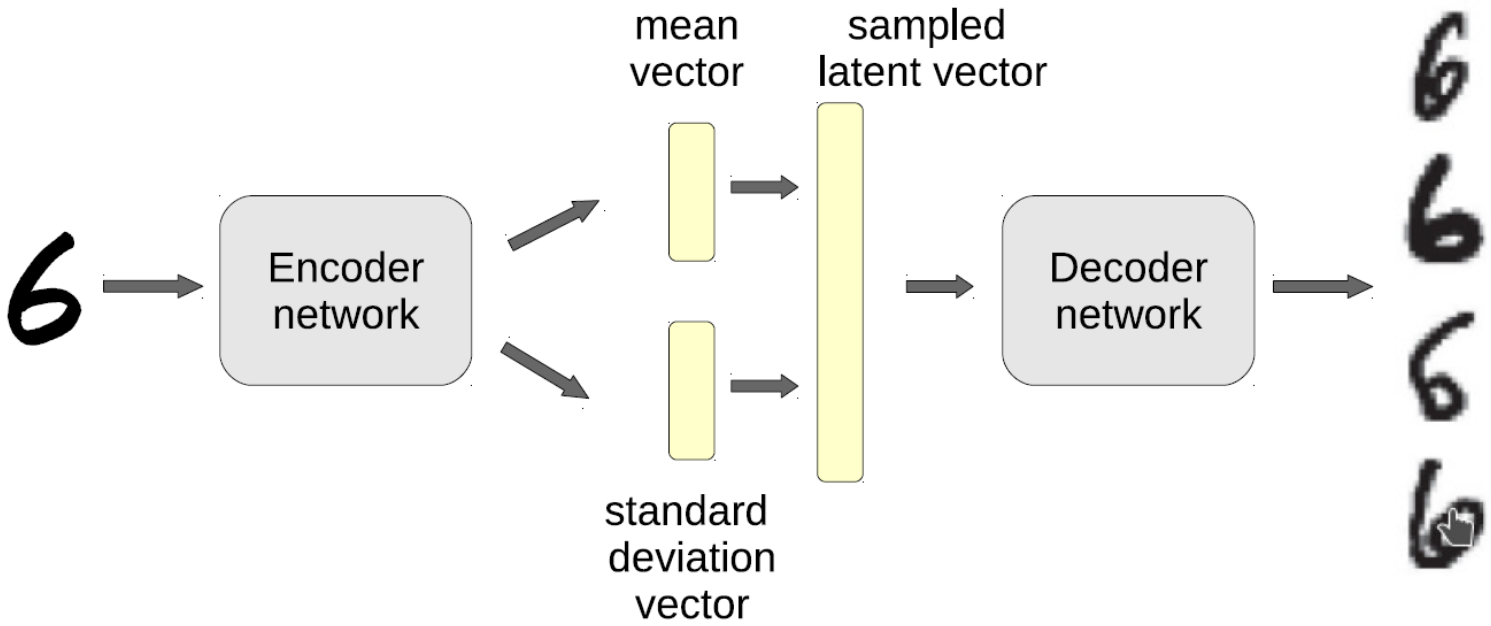
Autoencoder, applications

- Dimensionality reduction (nonlinear PCA)
- Feature extraction
- Unsupervised pretraining
- Manifold learning
- Data denoising
- Missing data
- Detect outliers

- Generative model

Variational Autoencoder





References, learn more

- Chapter 14 in the book Deep Learning by Bengio et al. https://www.deeplearningbook.org/
- Chapter 20.10.3 in the book Deep Learning (variational autoencoders)

- https://www.youtube.com/watch?v=s96mYcicbpE
- https://www.youtube.com/watch?v=FzS3tMl4Nsc

LUND UNIVERSITY

Prostate Cancer

Predicted incidence of prostate cancer in 2030:
1,7 million (2012: 1,1 million) worldwide

Up to 10 000 new cases/ year in Sweden

Mortality - approx 2500
1/6 men will be diagnosed with PCa

Estimated New Cases

		Males
Prostate	220,800	26%
Lung & bronchus	115,610	14%
Colon & rectum	69,090	8%
Urinary bladder	56,320	7%
Melanoma of the skin	42,670	5%
Non-Hodgkin lymphoma	39,850	5%
Kidney & renal pelvis	38,270	5%
Oral cavity & pharynx	32,670	4%
Leukemia	30,900	4%
Liver & intrahepatic bile duct	25,510	3%
All Sites	**848,200**	**100%**

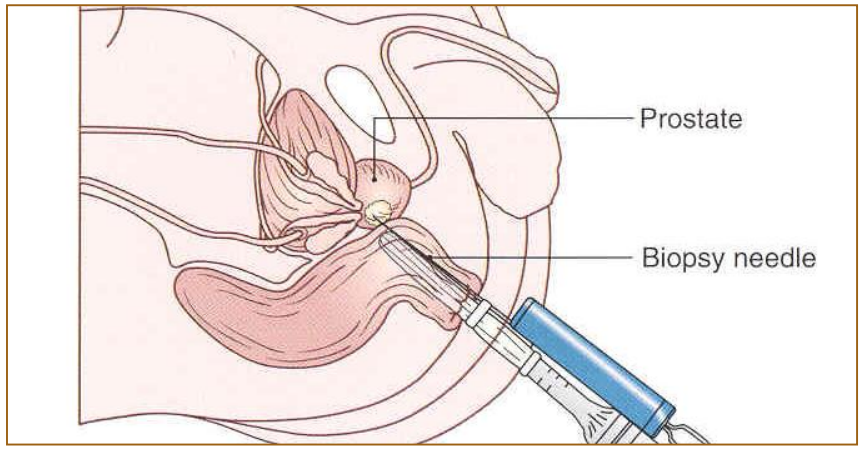
Estimated Deaths

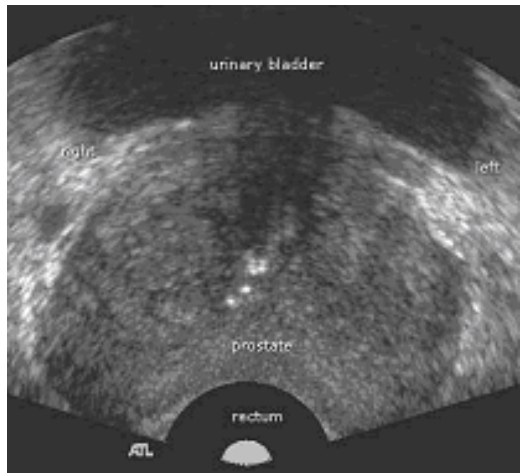
		Males
Lung & bronchus	86,380	28%
Prostate	27,540	9%
Colon & rectum	26,100	8%
Pancreas	20,710	7%
Liver & intrahepatic bile duct	17,030	5%
Leukemia	14,210	5%
Esophagus	12,600	4%
Urinary bladder	11,510	4%
Non-Hodgkin lymphoma	11,480	4%
Kidney & renal pelvis	9,070	3%
All Sites	**312,150**	**100%**

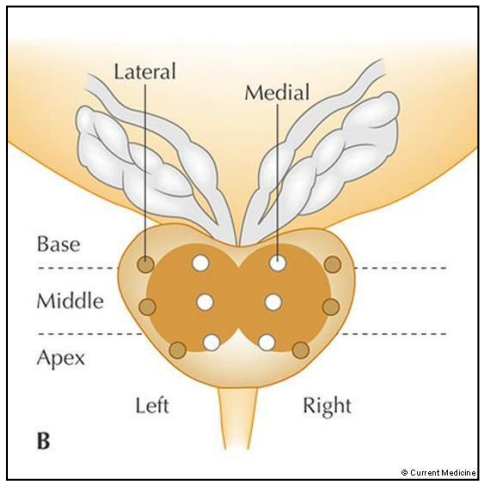
Siegel et al, Cancer Statistics 2015, CA CANCER J CLIN 2015;65:5–29

LUND UNIVERSITY

Transrectal ultrasonography (TRUS)
> 10 core biopsies







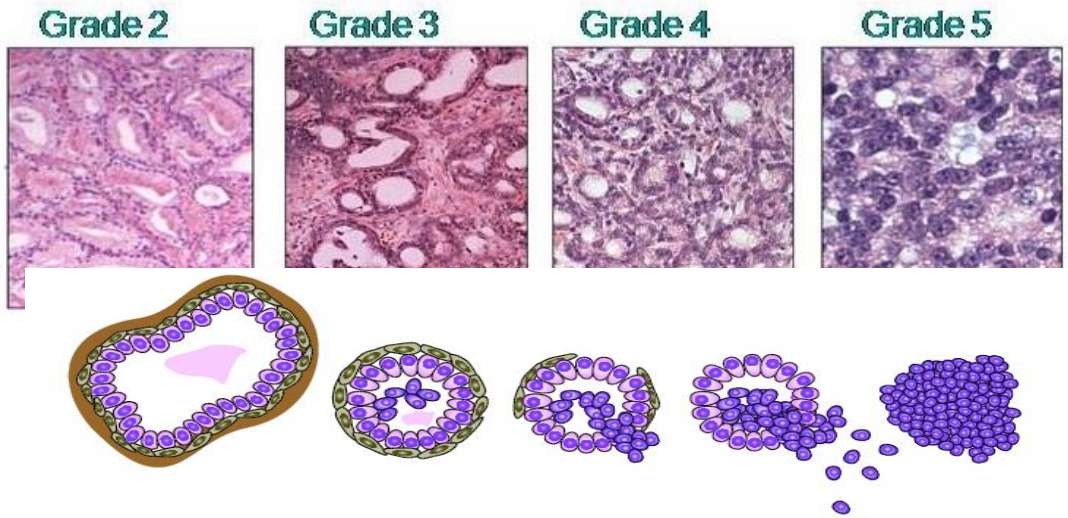
Biopsies: > 700 / year

Prostatectomies: > 400

Dept of Urology at Skåne University Hospital

LUND UNIVERSITY

Gleason grade / score



Donald F Gleason 1966

Consensus 2005

Upgrade 2010

The best prognostic and predictive biomarker in Pca
… but it is subjective and with low reproducibility
…attemps to standardize …

EQUALIS

SVENSK FÖRENING FÖR PATOLOGI

KVAST
Kvalitets- och standardiseringskommittén.

LUND UNIVERSITY

Gleason grade / score

Scandinavian Journal of Urology. 2013; Early Online, 1–8

Interobserver variability in the pathological assessment of radical prostatectomy specimens: Findings of the Laparoscopic Prostatectomy Robot Open (LAPPRO) study

JOSEFIN PERSSON[1], ULRICA WILDERÄNG[2], THOMAS JIBORN[1], PETER N. WIKLUND[3], JAN-ERIK DAMBER[4], JONAS HUGOSSON[5], GUNNAR STEINECK[2,6], EVA HAGLIND[7] & ANDERS BJARTELL[1]

LAPPRO study:

- Systematic variation in GS assessment
- Individual differences in judgement between pathologists

Table II. Interobserver variability in Gleason score (GS) between the local pathologists (LP) and the reference pathologists at central review (CR).

GS	GS (no. CR)				Exact concordance	Undergrading by LP vs CR	Overgrading by LP vs CR
(no. LP)	5–6	3 + 4	4 + 3	8–10	(%)	(%)	(%)
5–6 (117)	91	23	3	0	77.8	22.2	0.0
3 + 4 (107)	49	41	14	3	38.3	15.9	45.8
4 + 3 (43)	7	12	18	6	41.9	14.0	44.2
8–10 (16)	3	1	4	8	50.0	0.0	50.0
All (283)	150	77	39	17	55.8	17.3	26.9

Need for standardization

Need for faster diagnostics in Standardiserat Vårdförlopp ”SVF”

Lack of pathologists

Digital pathology

LUND UNIVERSITY

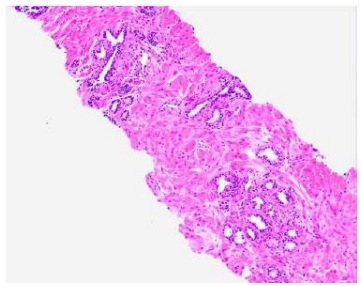
Project development

- Mathematical Imaging Group
  - Ida Arvidsson, Anders Heyden, Kalle Åström, Niels Chrisitan Overgaard
- Dept. of Translational Medicine
  - Anders Bjartell, Agnieszka Krzyzanowska
- Skåne University Hospital, dept. of Urology, dept. of Pathology
  - Felicia-Elena Marginean, Athanasios Simoulis
- Sectra
  - Claes Lundström, Erik Sjöblom

LUND UNIVERSITY

Goal

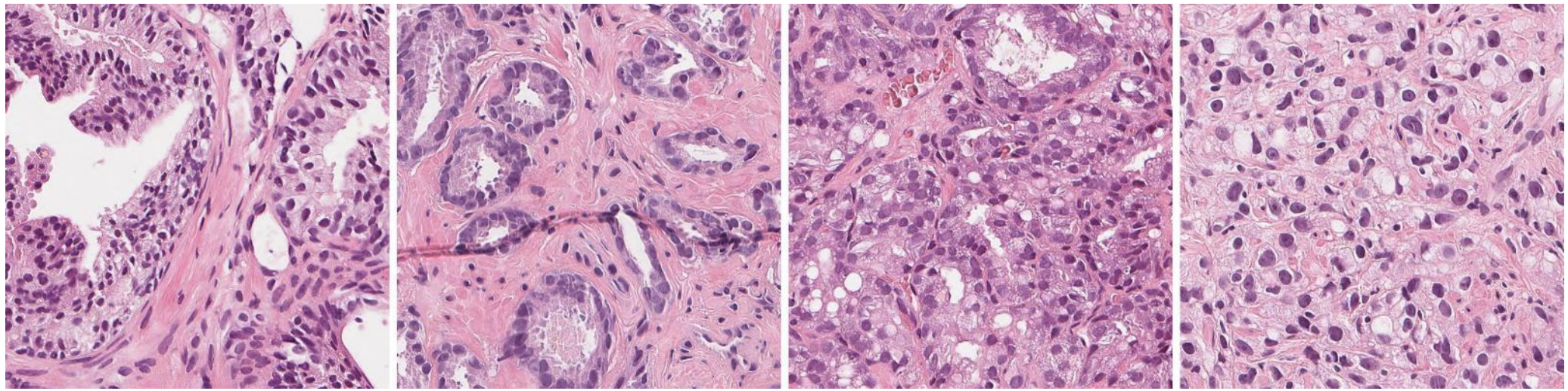
Develop an algorithm for automatic Gleason grading, that is able to classify images from different sites although the training data is limited.



LUND UNIVERSITY

Gleason Grading

- Prostate cancer is diagnosed based on the Gleason grade
- Manual inspection of prostate biopsies
- Stained with haematoxylin and eosin (H&E)

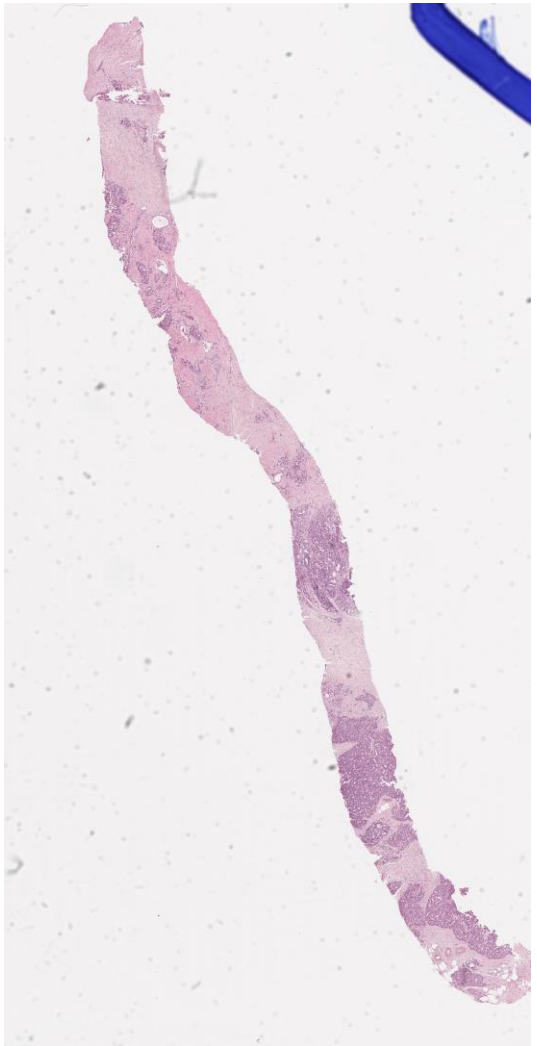


Benign *Gleason grade 3* *Gleason grade 4* *Gleason grade 5*

LUND UNIVERSITY

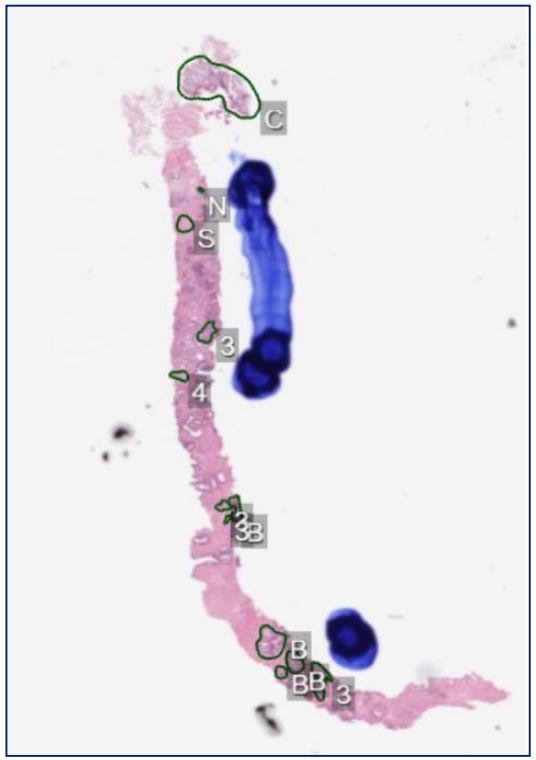
Data

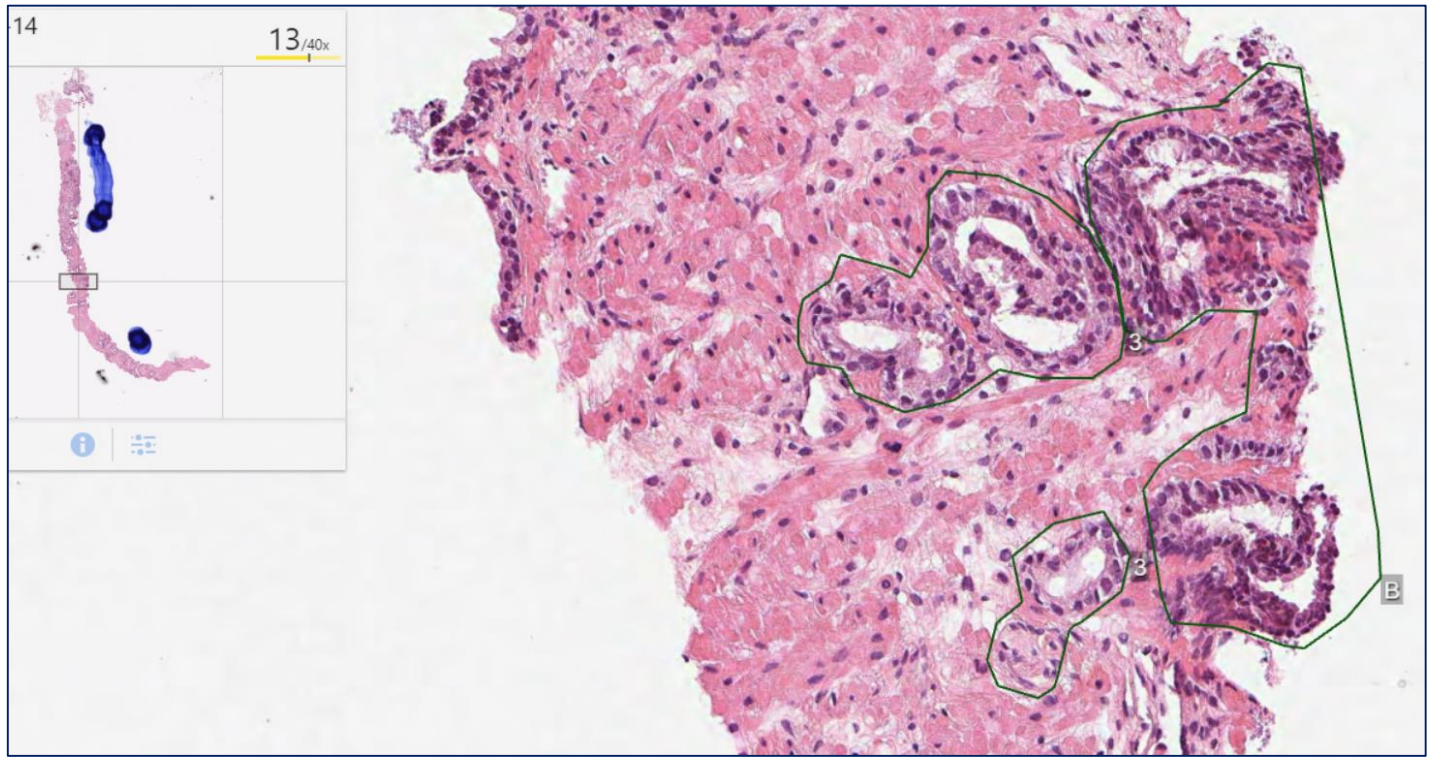
- Images captured at 40x, ~ 30000*60000 pixels, 0.2 GB
- Want to detect malignant areas, typically between
  20*20 pixels (single cell) and
  3000*3000 pixels (width of the tissue sample)



LUND UNIVERSITY

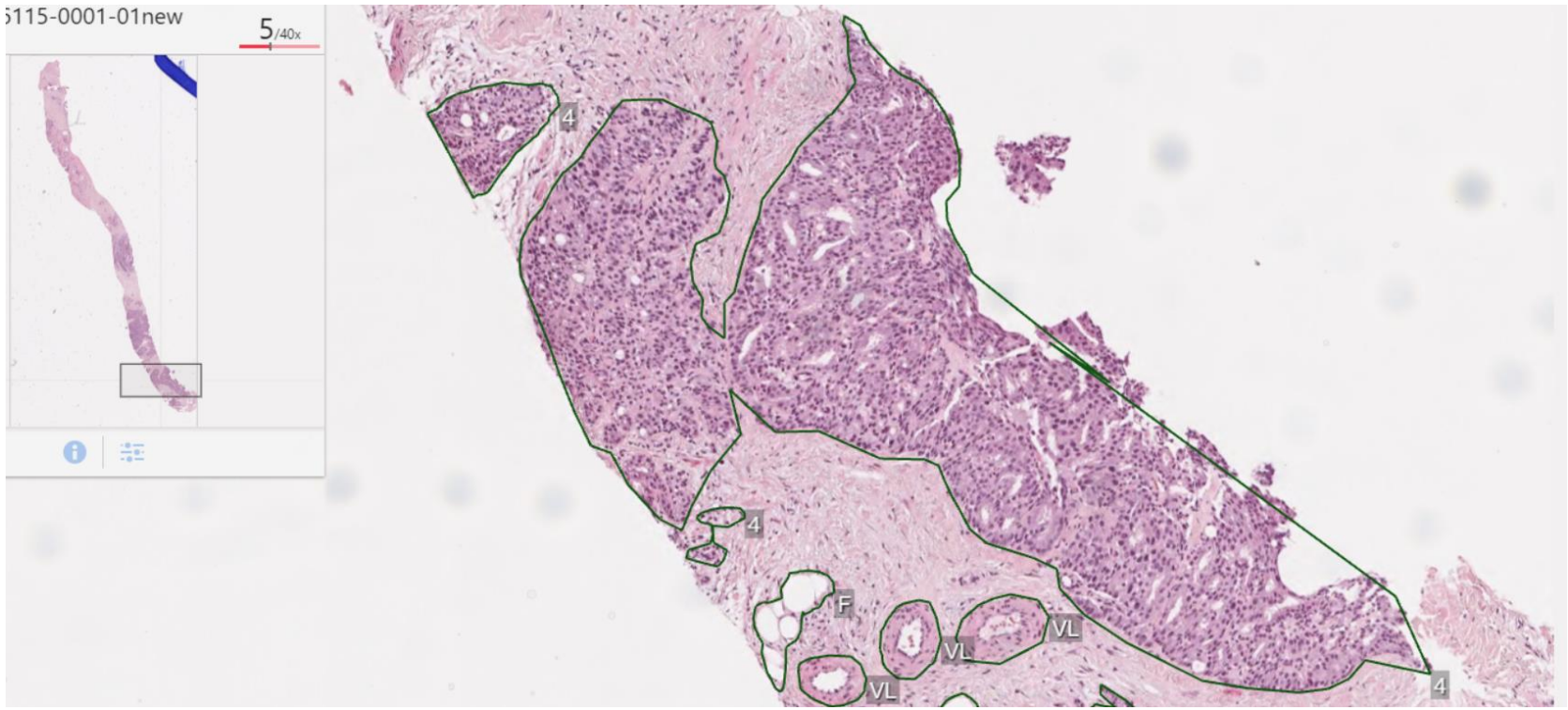
Data





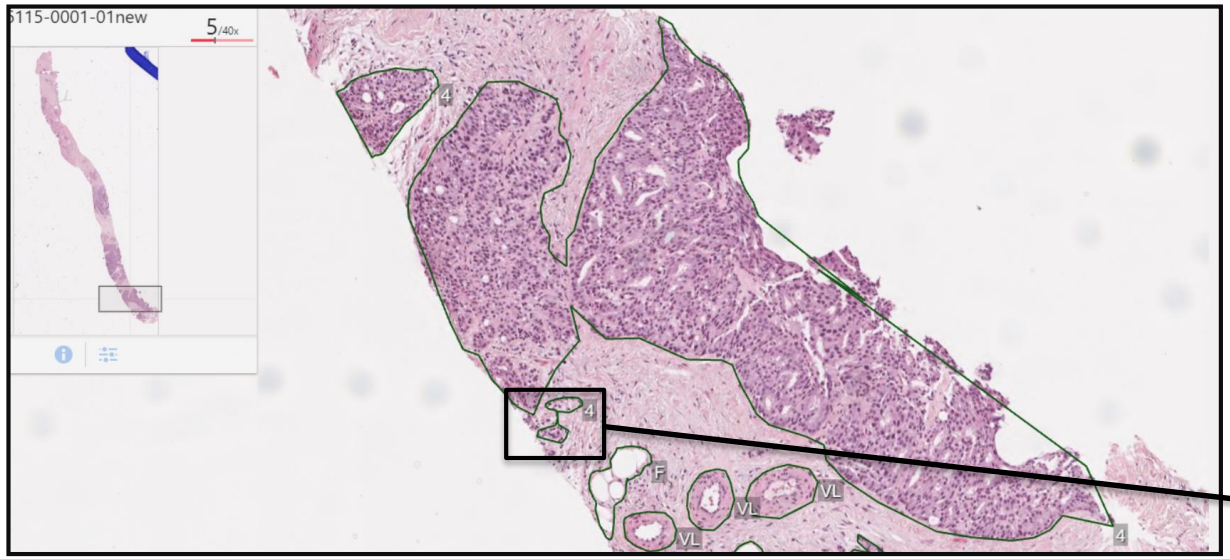
LUND UNIVERSITY

Gleason Grading, Examples

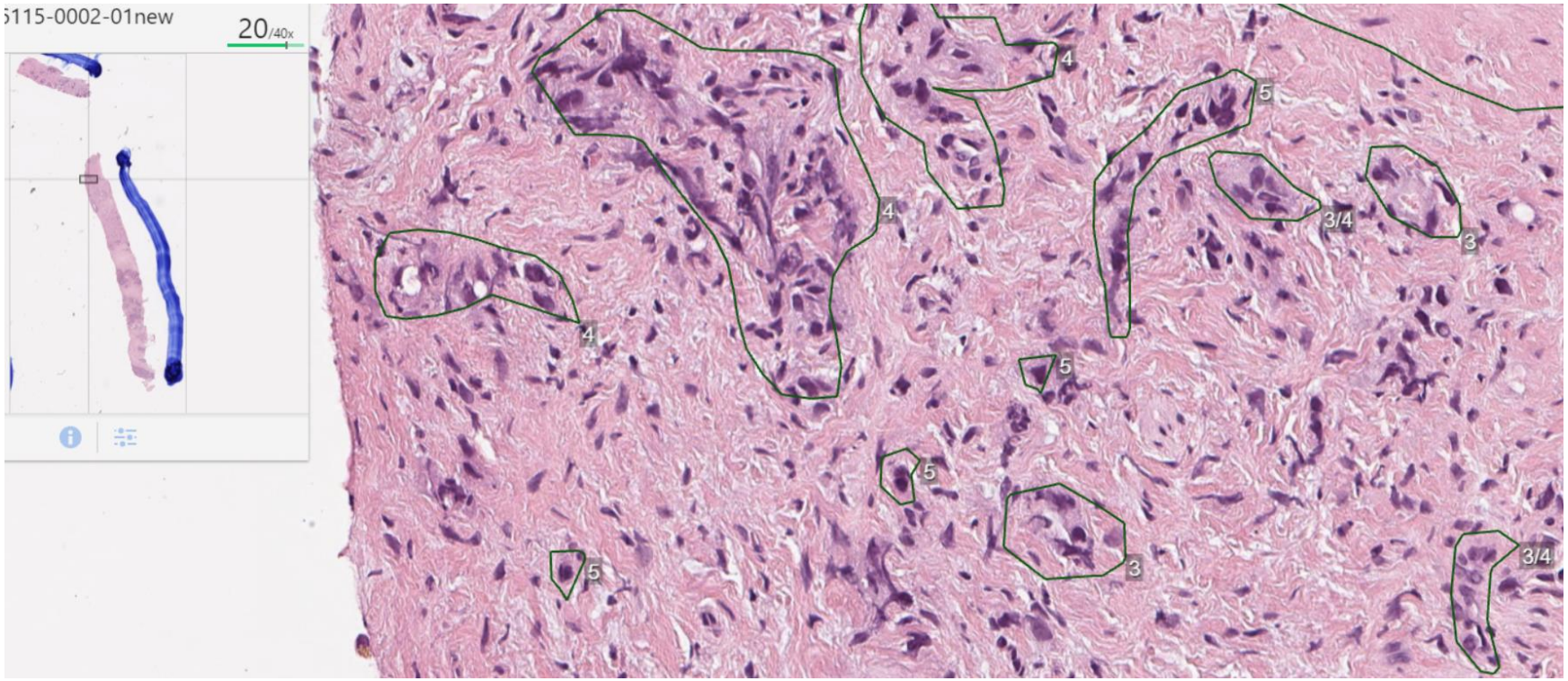


LUND UNIVERSITY

Gleason Grading, Examples

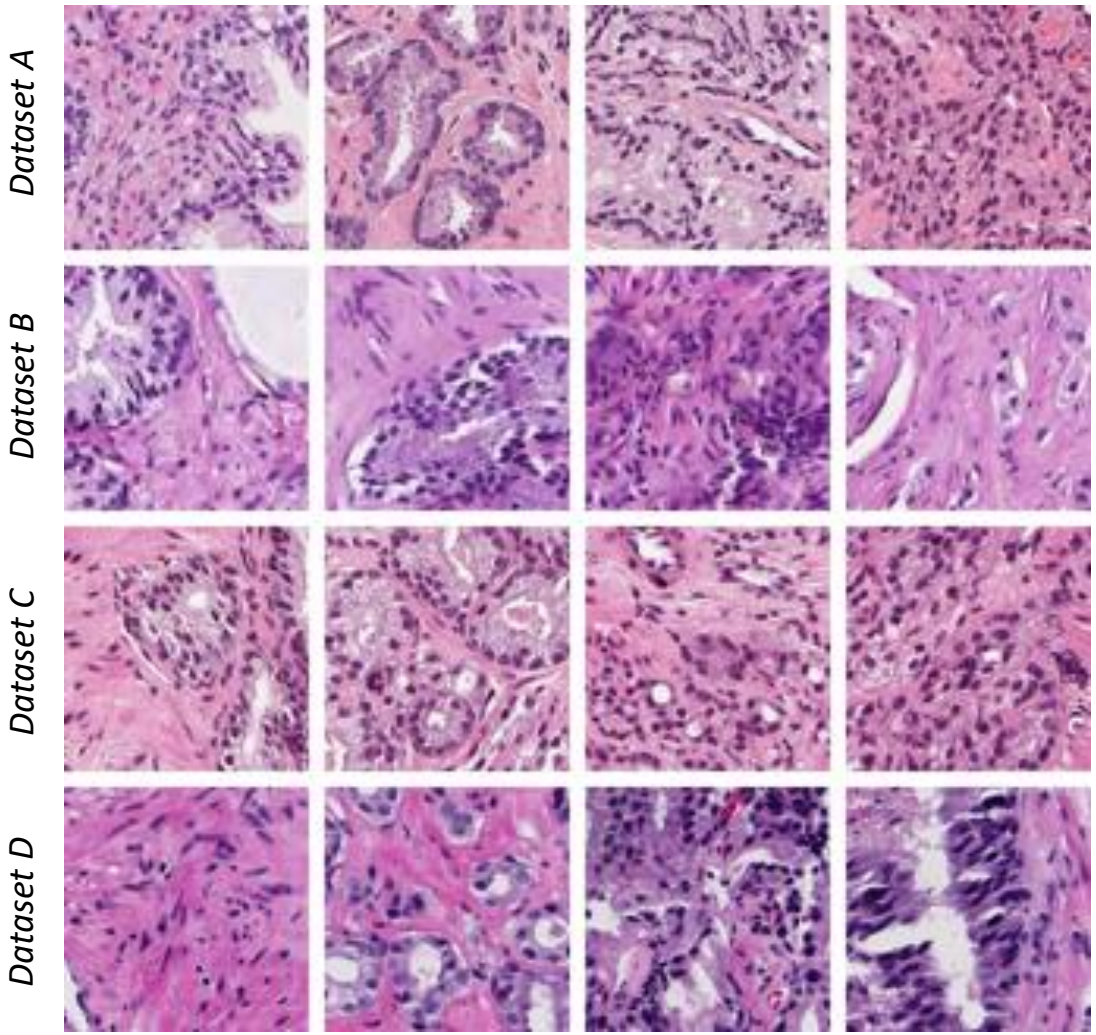


Gleason Grading, Examples



LUND UNIVERSITY

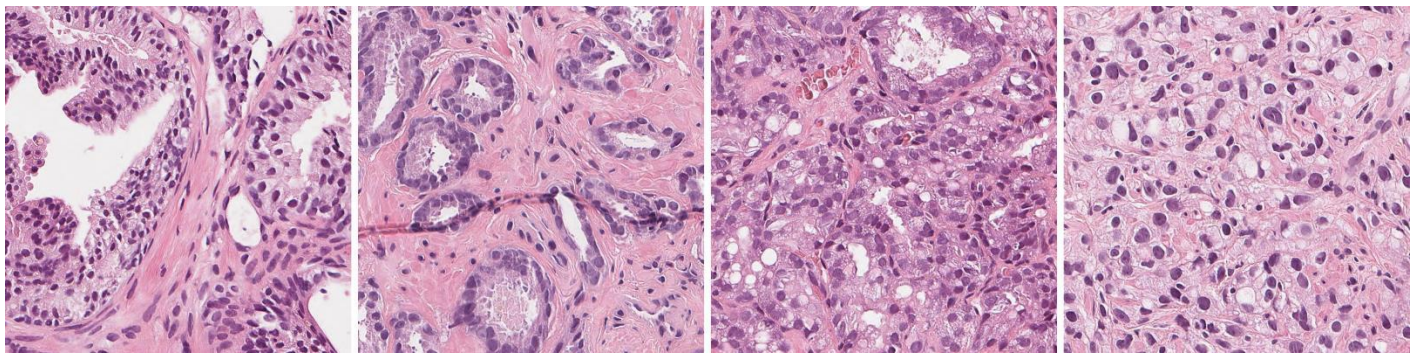
Stain Variations between Sites



LUND UNIVERSITY

Data

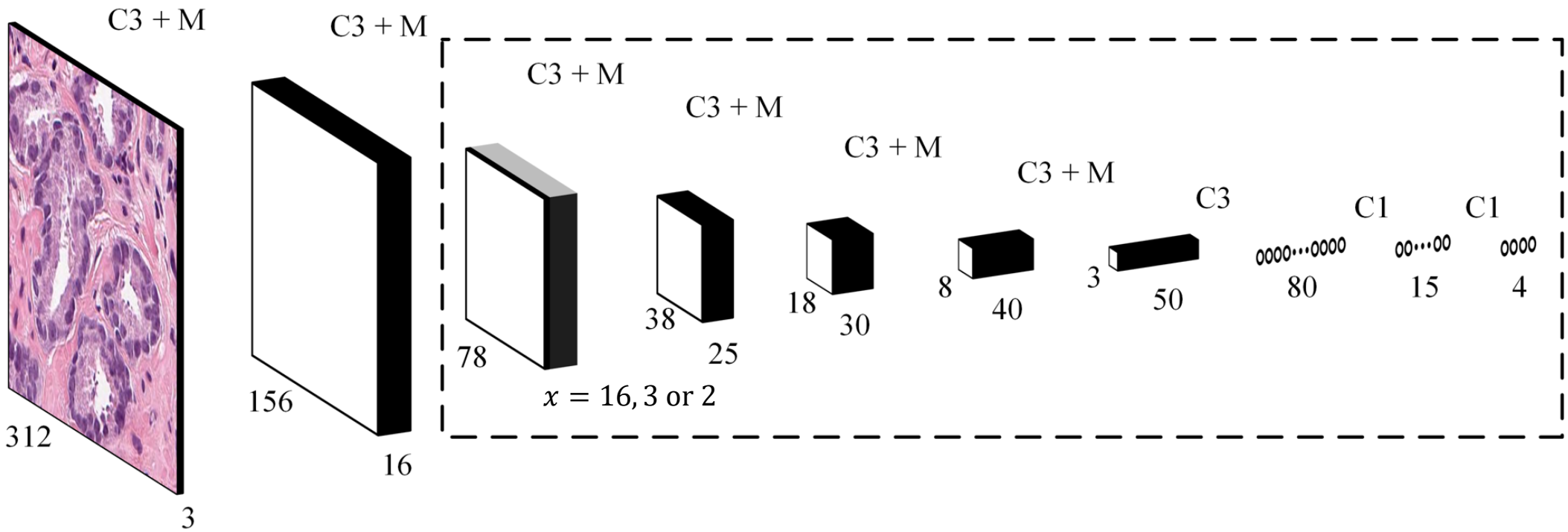
Dataset	Number of slides	Number of annotated areas				Number of extracted patches			
		Benign	G3	G4	G5	Benign	G3	G4	G5
A train	109	1902	545	637	268	348504	224808	315312	186672
A test	-	633	181	212	89	9679	1396	2249	1115
B validation	55	117	48	52	9	3628	879	487	42
C validation	16	54	9	189	17	1635	296	2087	335
D validation	50	29	304	209	8	977	1909	1103	47



Benign *Gleason 3* *Gleason 4* *Gleason 5*

LUND UNIVERSITY

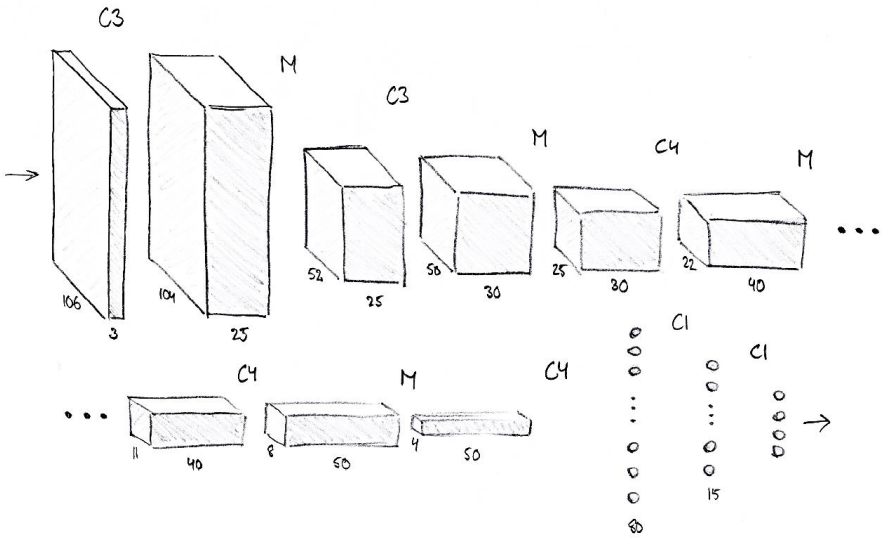
Classifier



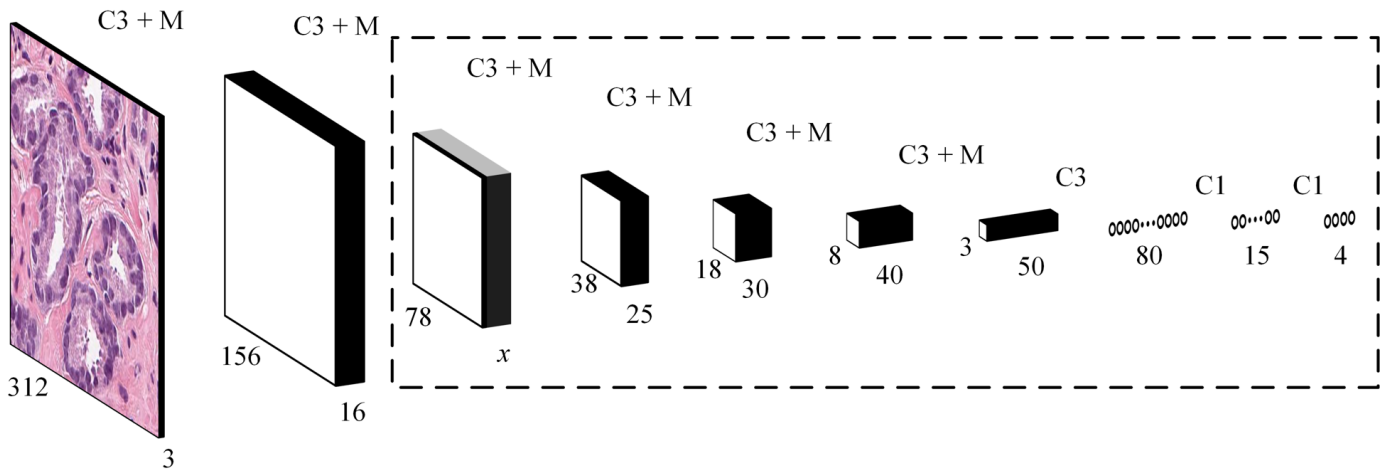
C3: 3 × 3 convolution
C1: 1 × 1 convolution (i.e. fully connected)
M: 2 × 2 maxpooling

LUND UNIVERSITY

Classifiers, CNNs trained from scratch

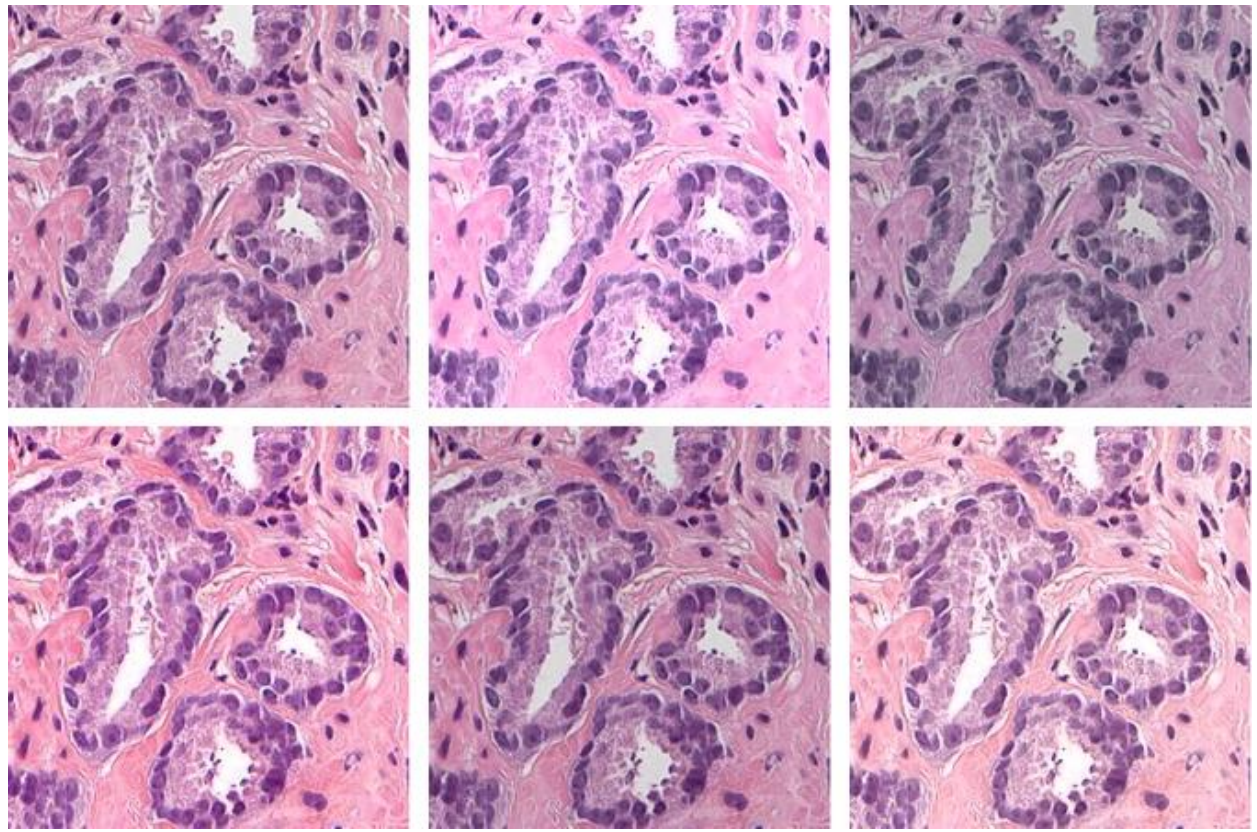


Function	Number of filters	Size
conv.	16	3x3
max pool.		2x2
conv.	16	3x3
max pool.		2x2
conv.	32	3x3
max pool.		2x2
conv.	32	3x3
max pool.		2x2
conv.	64	4x4
max pool.		2x2
conv.	64	3x3
conv.	128	4x4
dropout		
conv.	32	1x1
dropout		
conv.	4	1x1



Color Augmentation

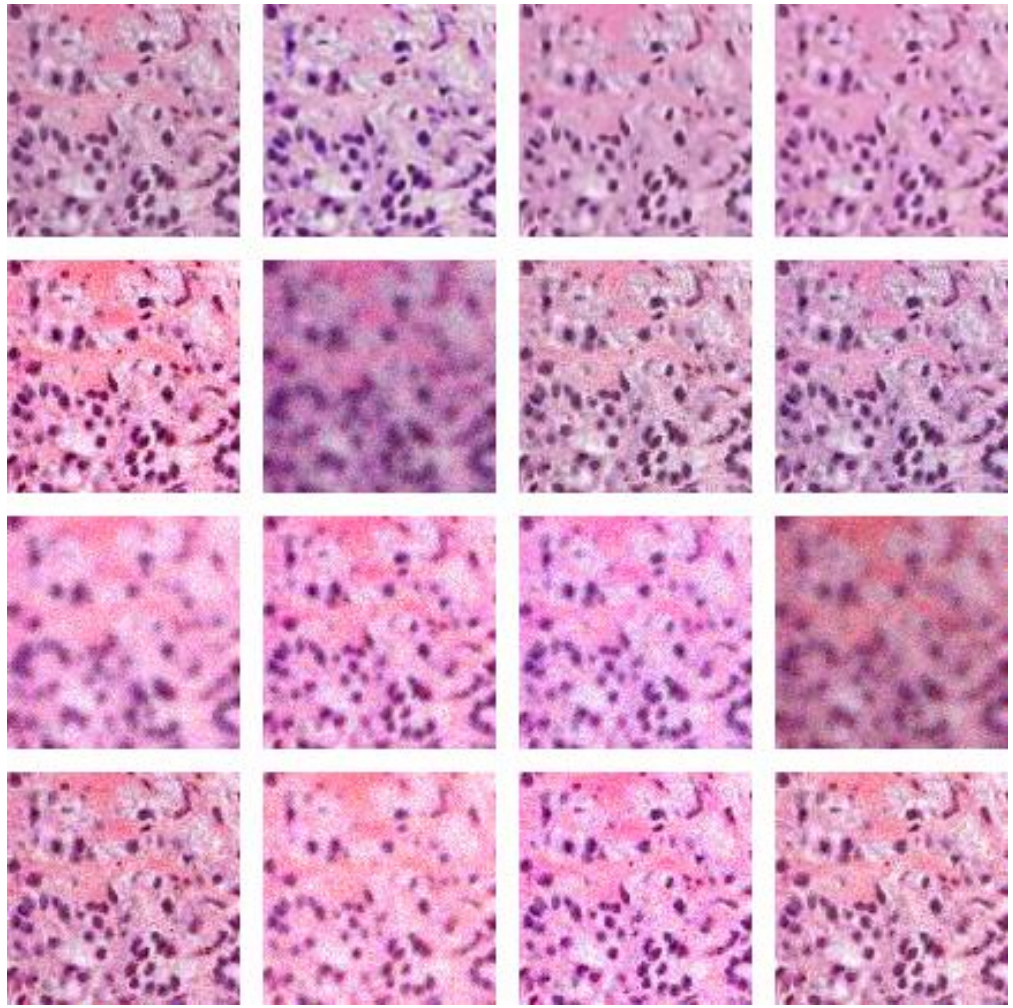
- $RGB \rightarrow HSV$
- $H' = H \cdot random([1/1.04, 1.04])$
- $S' = S \cdot random([1/1.25, 1.25])$
- $V' = V \cdot random([1/1.25, 1.25])$
- $H'S'V' \rightarrow R'G'B'$



LUND UNIVERSITY

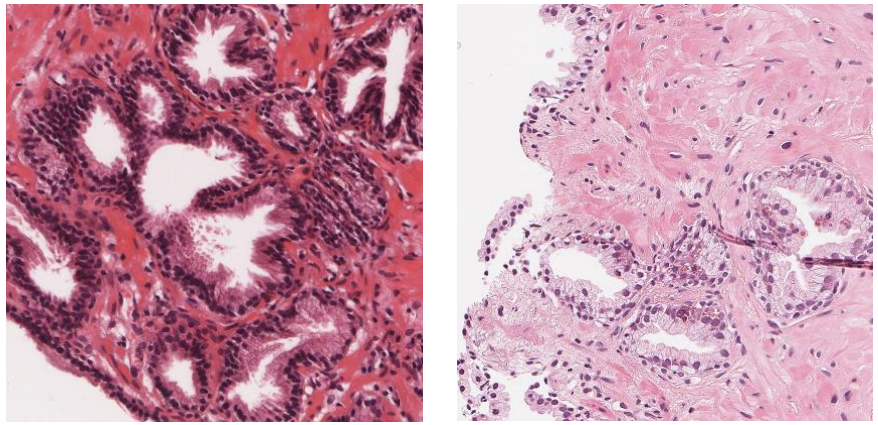
Augmentation

- Blurring, intensity clipping, adding noise, color augmentation



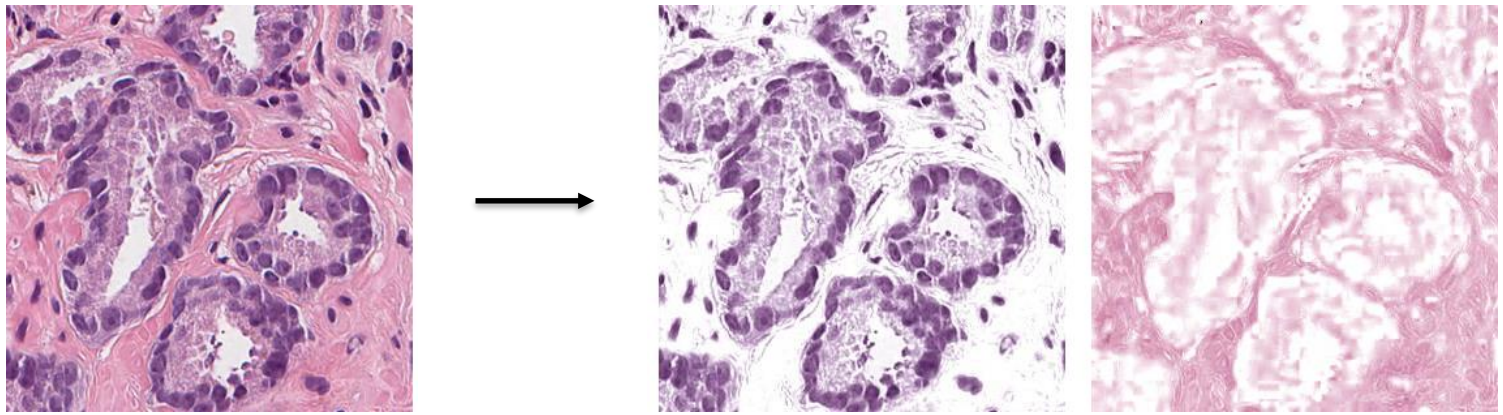
LUND UNIVERSITY

Stain Separation and Normalization



Example of stain variation

Method from: A. Vahadane, et al., *Structure preserving color normalization and sparse stain separation for histological images*, IEEE transactions on medical imaging, vol. 35, no. 8, pp. 1962–1971, 2016.



Digital stain separation

LUND UNIVERSITY

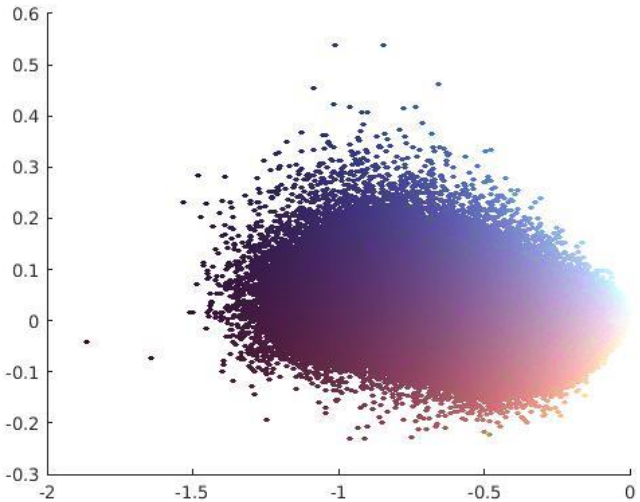
Stain Separation and Normalization

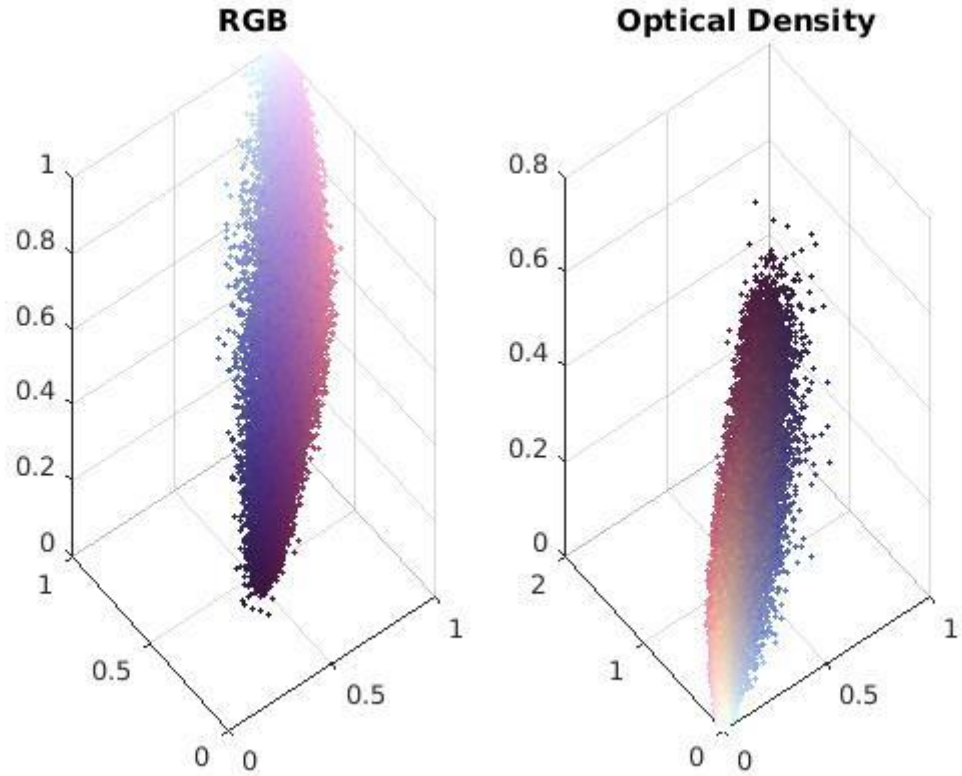
- Beer-Lamberts law

  $I_c = I_0 e^{-\alpha C}$

- Optical density $OD = -\ln\left(\frac{I_C}{I_0}\right) = \alpha C$,

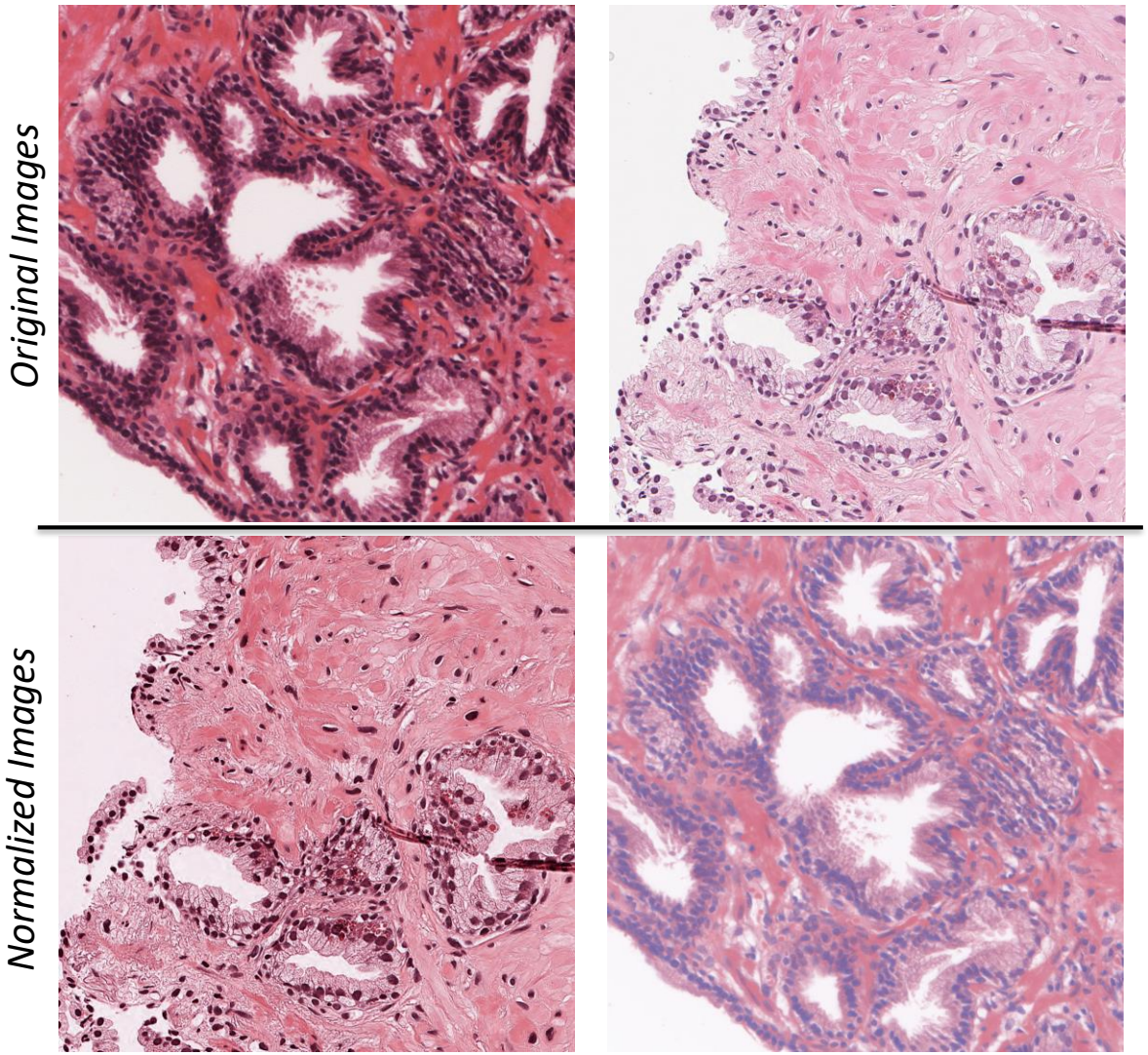
  thus $OD \sim C$





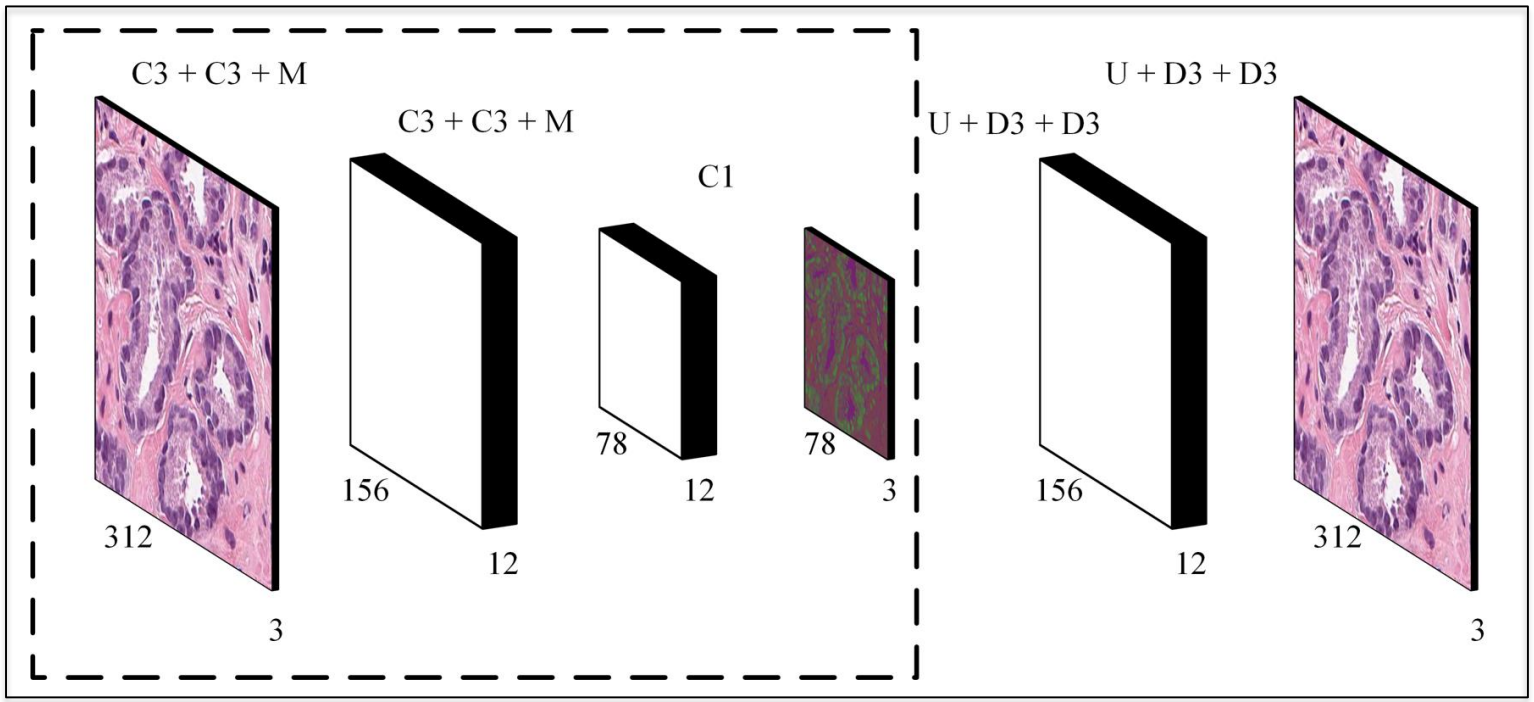
LUND UNIVERSITY

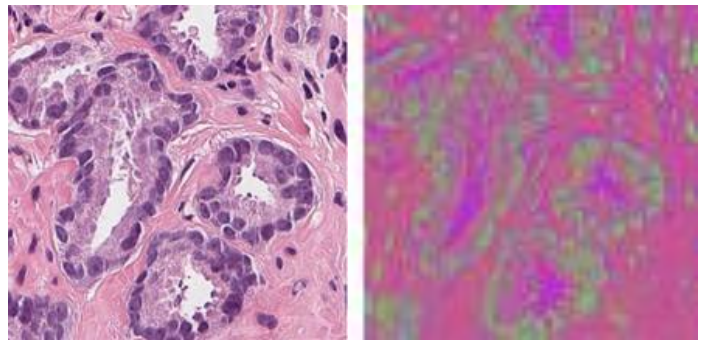
Stain Separation and Normalization



LUND UNIVERSITY

Autoencoder



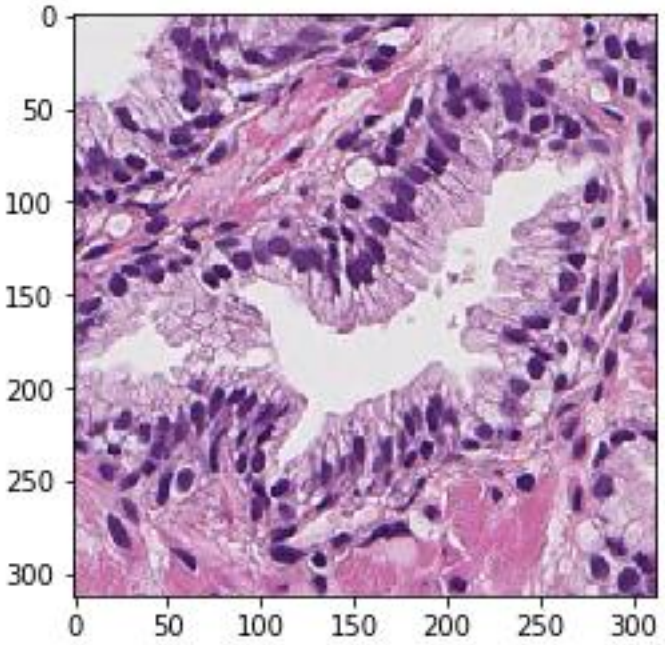


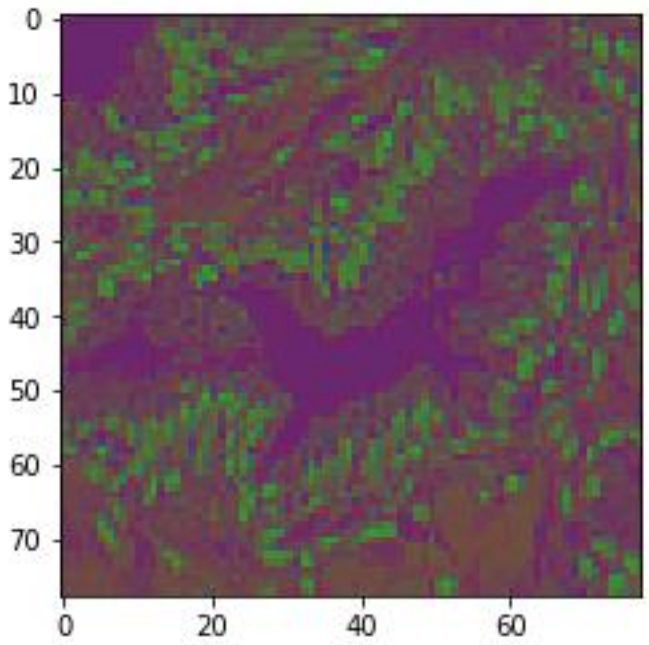
Original image and its encoded version

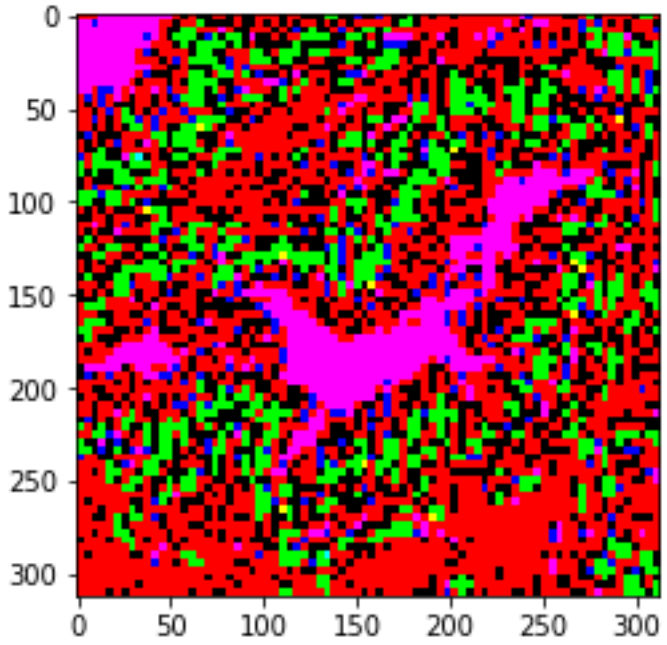
C3: 3 × 3 convolution
C1: 1 × 1 convolution
D3: 3 × 3 transposed convolution
M: 2 × 2 maxpooling
U: 2 × 2 upsampling

LUND UNIVERSITY

Autoencoder







LUND UNIVERSITY

Results

Result for Gleason grading as well as for benign versus malignant in parenthesis.

Input size & preprocessing	Test accuracy (%)	Validation accuracy (%)	
	Dataset A	Dataset B	Dataset C
20X – color augmentation	**81** (**95**)	46 (76)	50 (92)
5X	73 (94)	46 (80)	45 (88)
5X – color augmentation	79 (95)	42 (81)	49 (92)
5X – stain normalization	78 (94)	48 (79)	**53** (89)
5X – autoencoder	75 (92)	50 (83)	47 (90)
5X – autoencoder, color augmentation	75 (93)	**53** (**83**)	50 (**94**)

Arvidsson, I., Overgaard, N. C., Marginean, F. E., Krzyzanowska, A., Bjartell, A., Åström, K., & Heyden, A. (2018, April). Generalization of prostate cancer classification for multiple sites using deep learning. In *IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), 2018* (pp. 191-194).

LUND UNIVERSITY

Results and Conclusion

- Higher resolution improves performance
- Color augmentation is beneficial in all cases
- Stain normalization improves performance for validation datasets
- Best generalization performance when the autoencoder is used as preprocessing
- Not satisfactory validation results for Gleason grading

Input size & preprocessing	Test acc. (%)	Validation acc. (%)	
	Dataset A	Dataset B	Dataset C
20X – color augmentation	**81** (**95**)	46 (76)	50 (92)
5X	73 (94)	46 (80)	45 (88)
5X – color augmentation	79 (95)	42 (81)	49 (92)
5X – stain normalization	78 (94)	48 (79)	**53** (89)
5X – autoencoder	75 (92)	50 (83)	47 (90)
5X – autoencoder, color augm.	75 (93)	**53** (**83**)	50 (**94**)

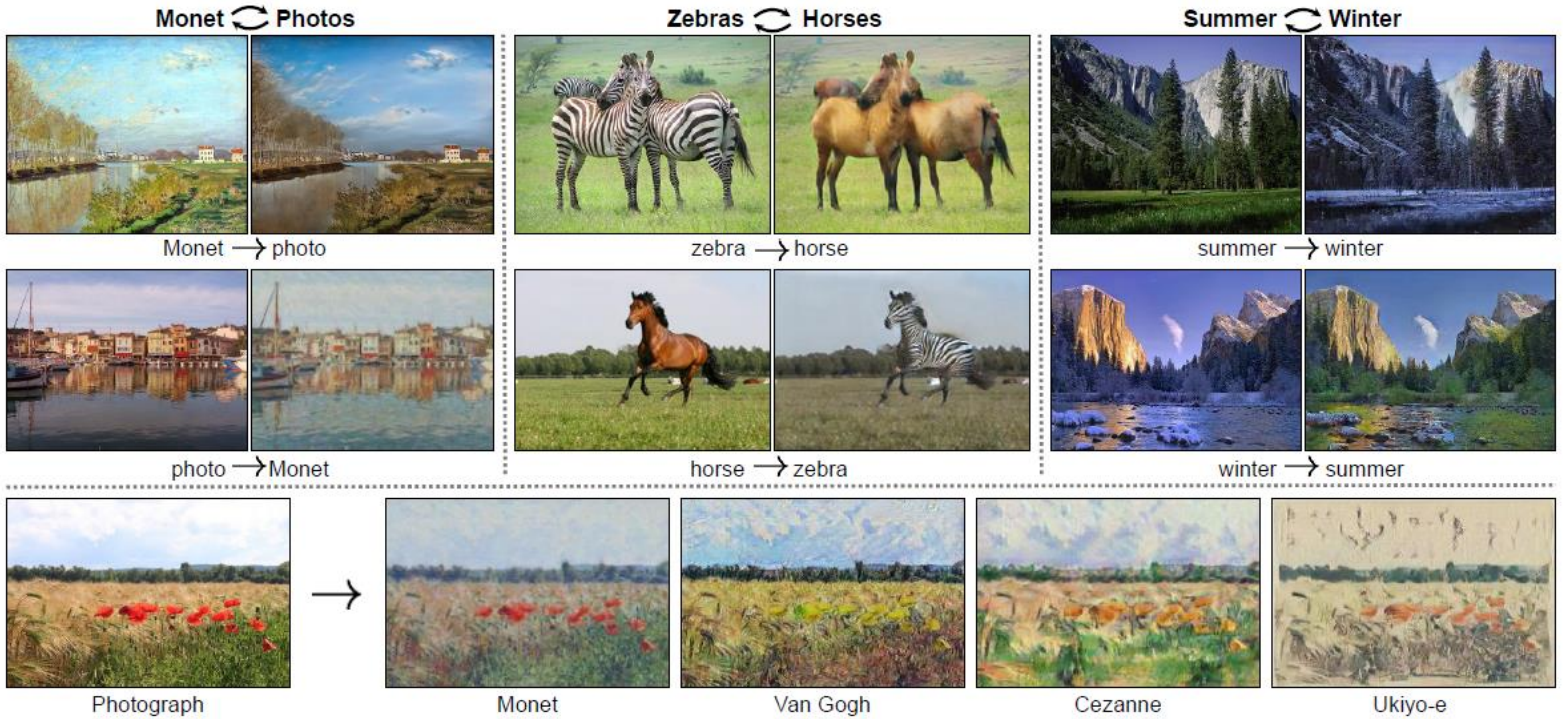
LUND UNIVERSITY

Cycle Generative Adversarial Networks

ICCV 2017

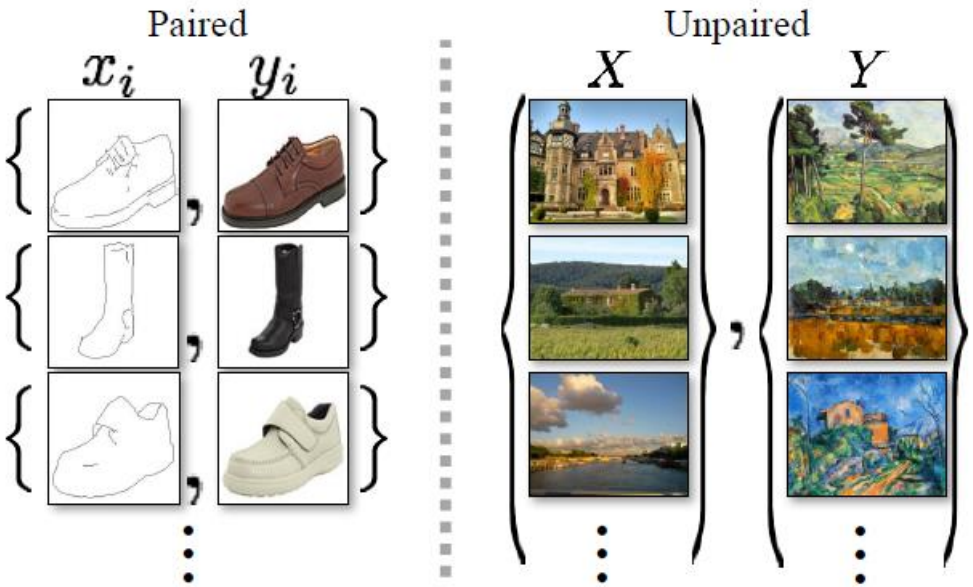
Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks

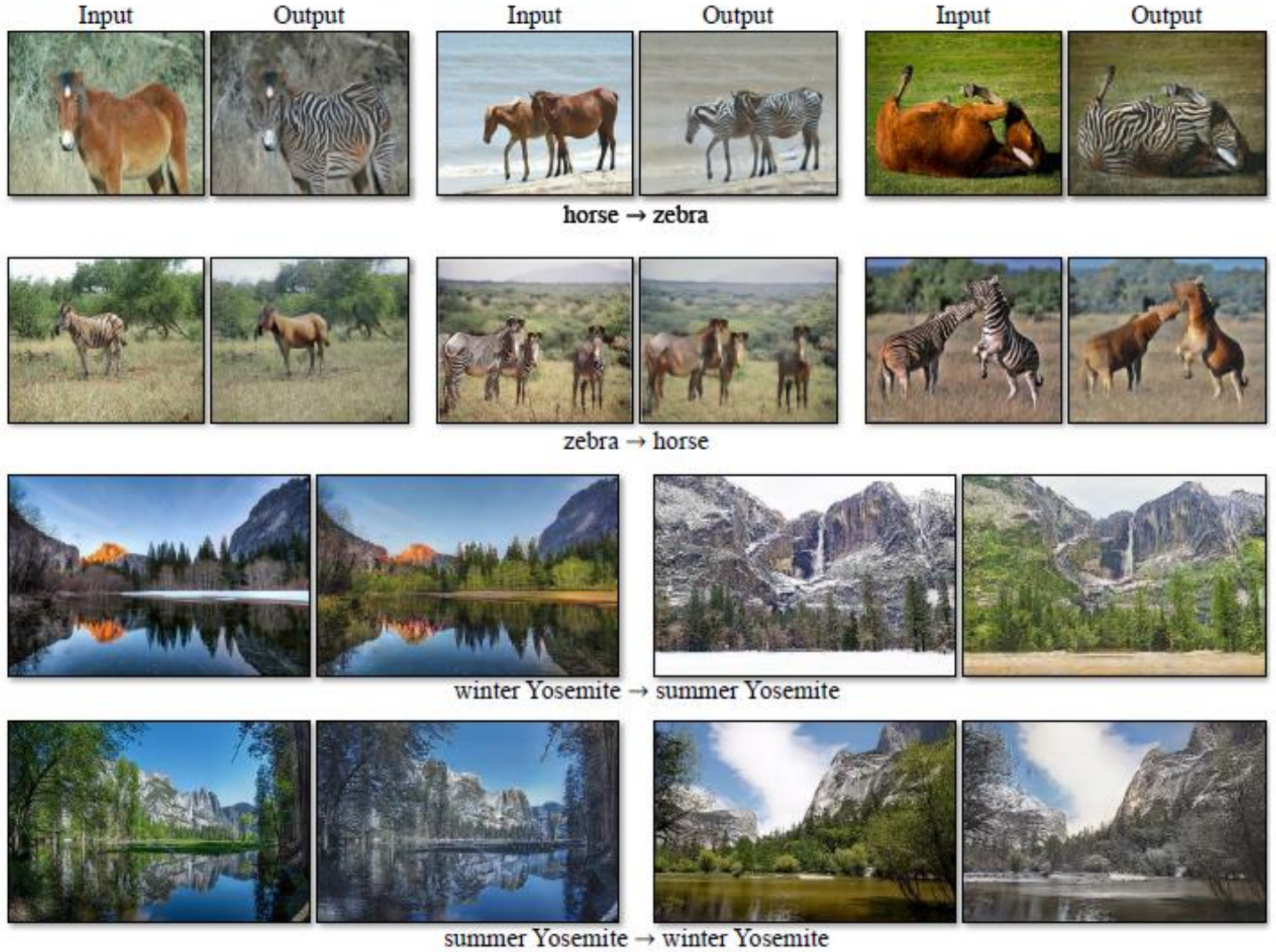
Jun-Yan Zhu* Taesung Park* Phillip Isola Alexei A. Efros
Berkeley AI Research (BAIR) laboratory, UC Berkeley



LUND UNIVERSITY

Cycle Generative Adversarial Network





Cycle Generative Adversarial Networks

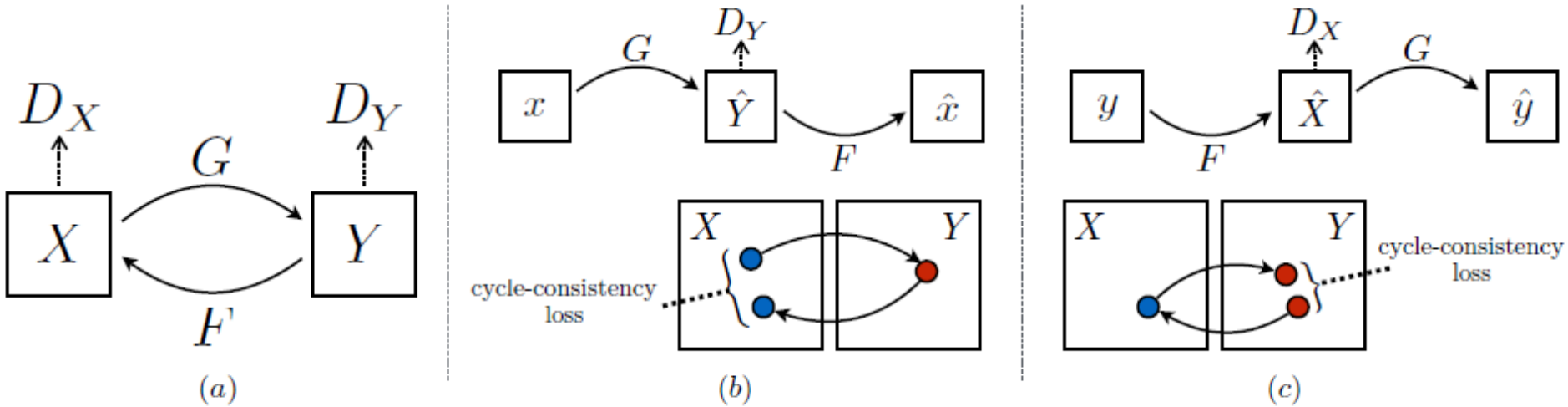
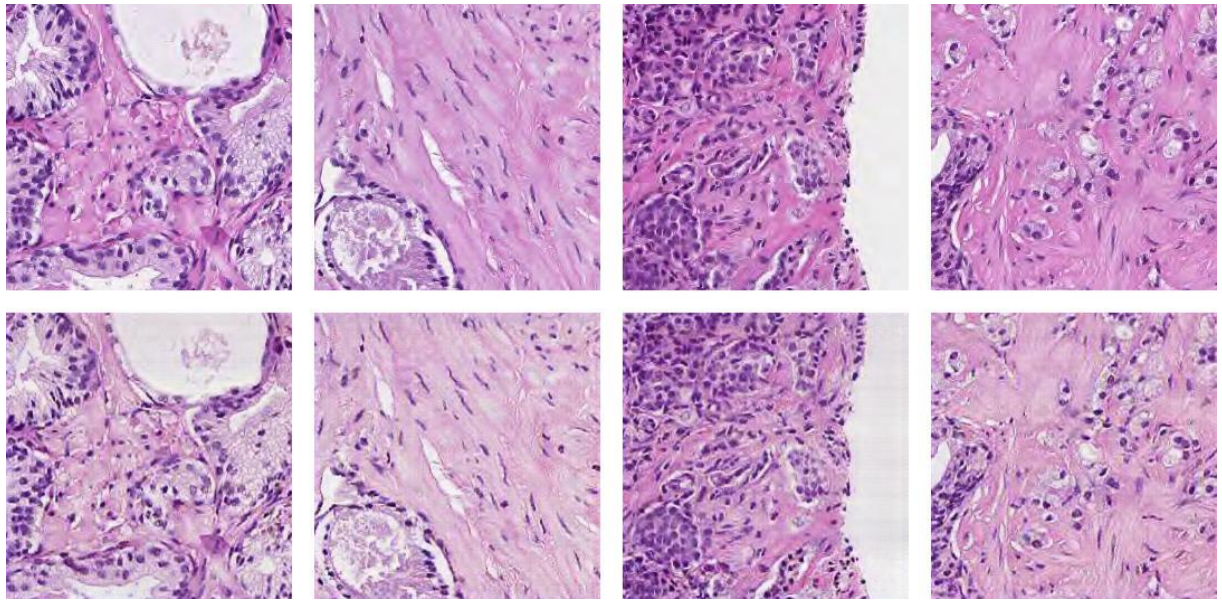


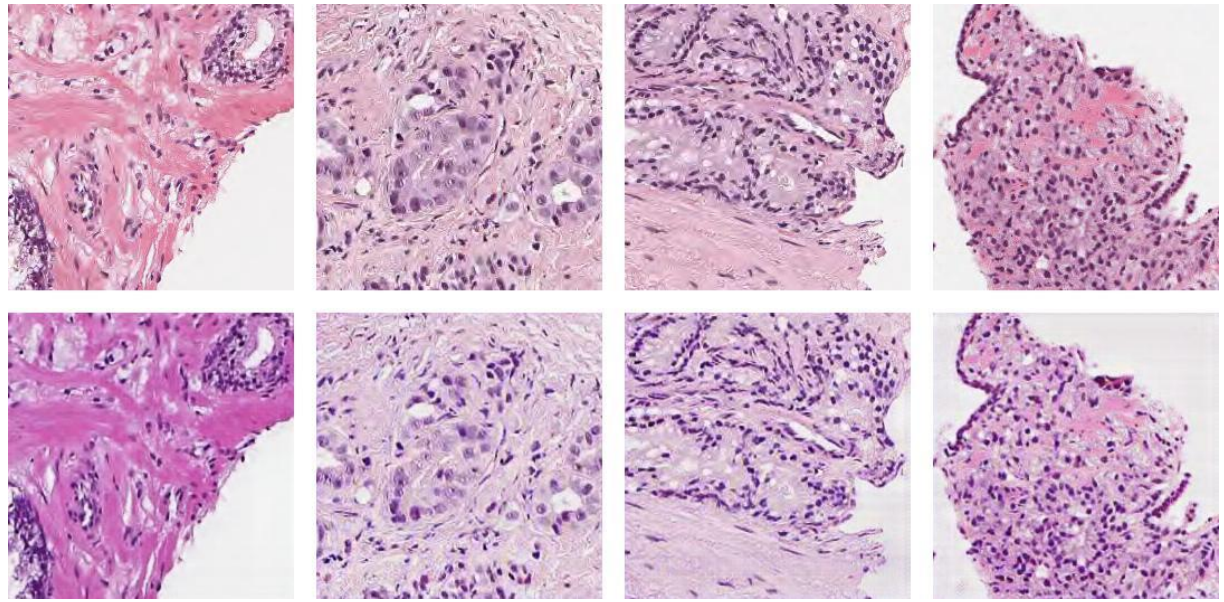
Figure 3: (a) Our model contains two mapping functions $G : X \rightarrow Y$ and $F : Y \rightarrow X$, and associated adversarial discriminators D_Y and D_X. D_Y encourages G to translate X into outputs indistinguishable from domain Y, and vice versa for D_X and F. To further regularize the mappings, we introduce two *cycle consistency losses* that capture the intuition that if we translate from one domain to the other and back again we should arrive at where we started: (b) forward cycle-consistency loss: $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$, and (c) backward cycle-consistency loss: $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$

LUND UNIVERSITY

Cycle GAN, Helsingborg-Malmö



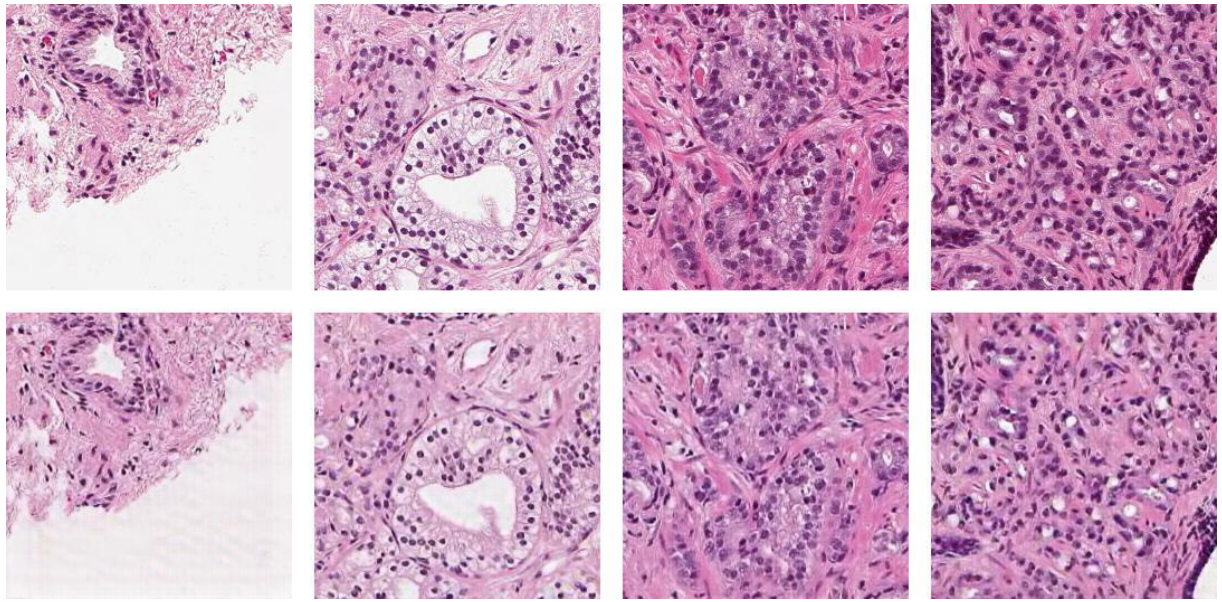
Top row: Original images from Helsingborg
Bottom row: Transformed to Malmö



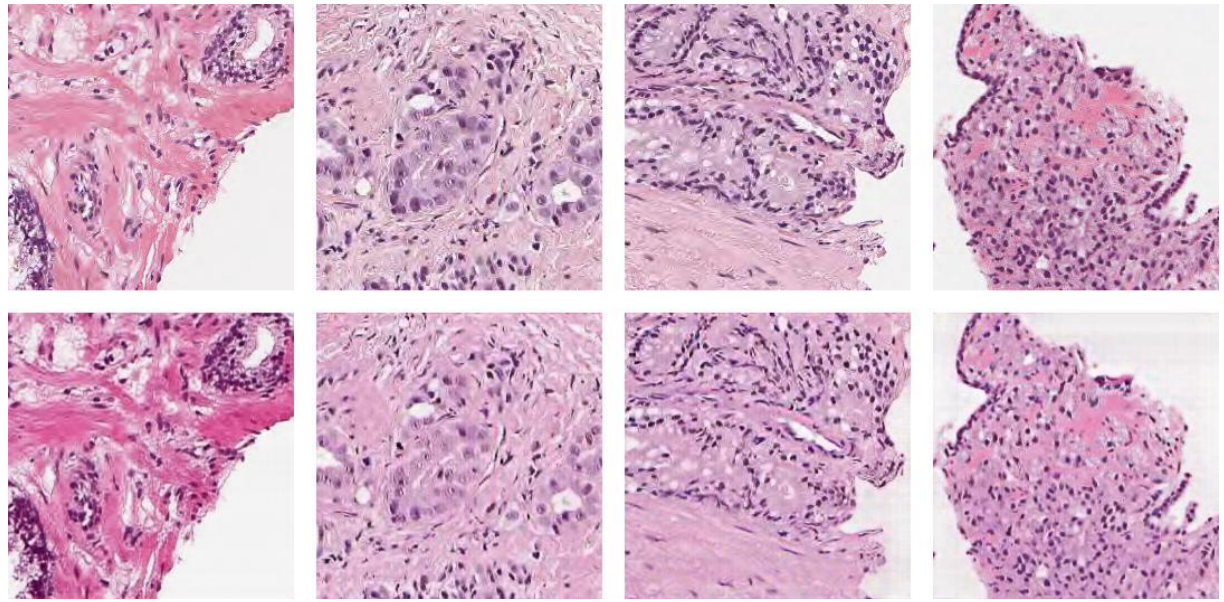
Top row: Original images from Malmö
Bottom row: Transformed to Helsingborg

LUND UNIVERSITY

Cycle GAN, Linköping-Malmö



Top row: Original images from Linköping
Bottom row: Transformed to Malmö



Top row: Original images from Malmö
Bottom row: Transformed to Linköping

LUND UNIVERSITY

Cycle GAN, results

Table 3. Accuracy for Gleason grading when using different augmentation techniques, where aug includes both blurring, intensity clipping and adding noise. Results in parentheses is the result when first transforming the images to dataset A using cycle GANs.

Augmentation technique	Test accuracy (%)	Validation accuracy (%)		
	Dataset A	Dataset B	Dataset C	Dataset D
Nothing	77	46 (48)	57 (58)	50 (55)
Color	77	55 (51)	57 (62)	59 (56)
Color, aug	77	59 (57)	59 (61)	56 (60)
Cycle GAN	74	44 (44)	52 (49)	48 (51)
Cycle GAN, color	76	55 (53)	57 (55)	50 (56)
Cycle GAN, color, aug	71	46 (50)	54 (59)	47 (51)

LUND UNIVERSITY

Cycle GAN, example of failed transformation

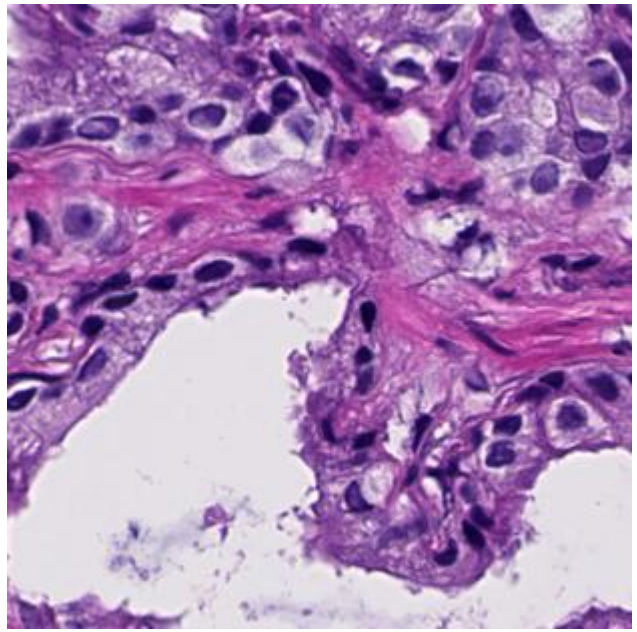
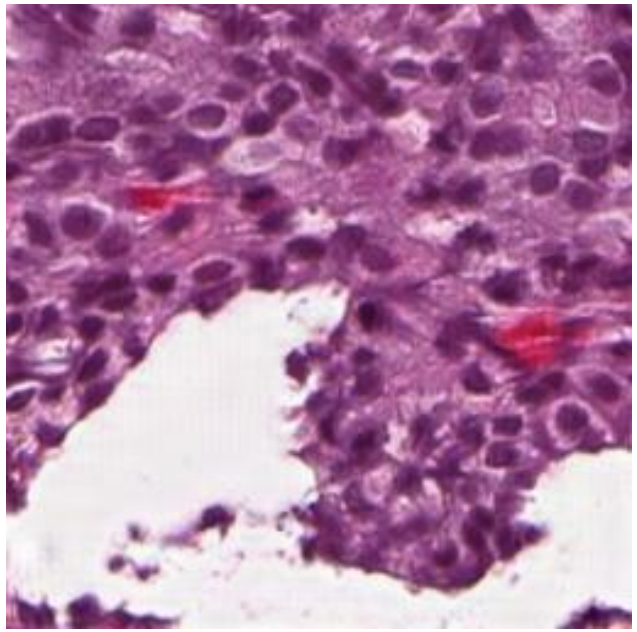


Image from Malmö …



… transformed to Linköping

LUND UNIVERSITY

Training of the Networks

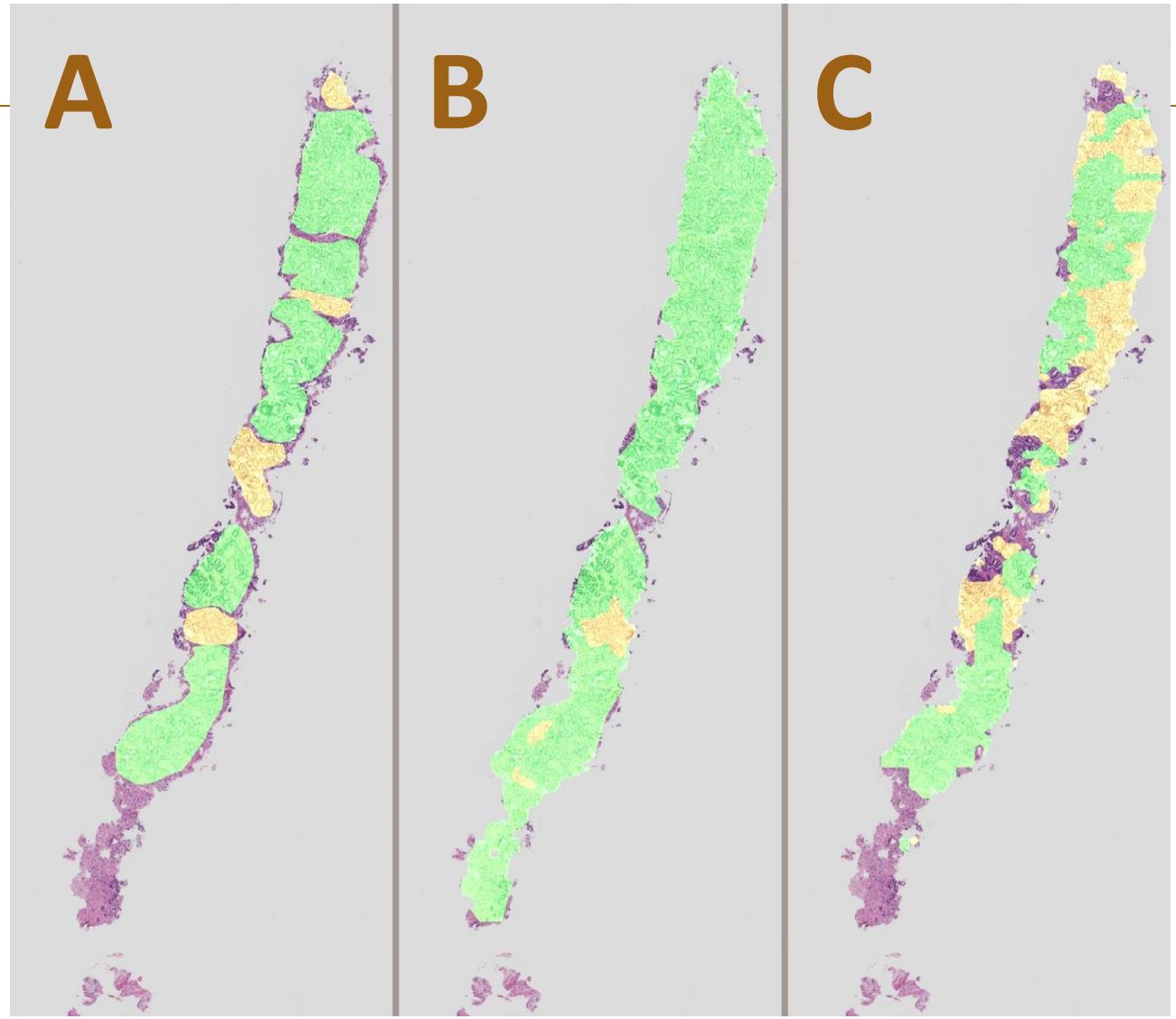
- Keras with the TensorFlow backend
- Adam optimizer
- Computers with GPU; GeForce GTX 970 4GB and TITAN X 12GB (Pascal)

LUND UNIVERSITY

Results

- A – pathologist 1
- B – pathologist 2
- C – algorithm

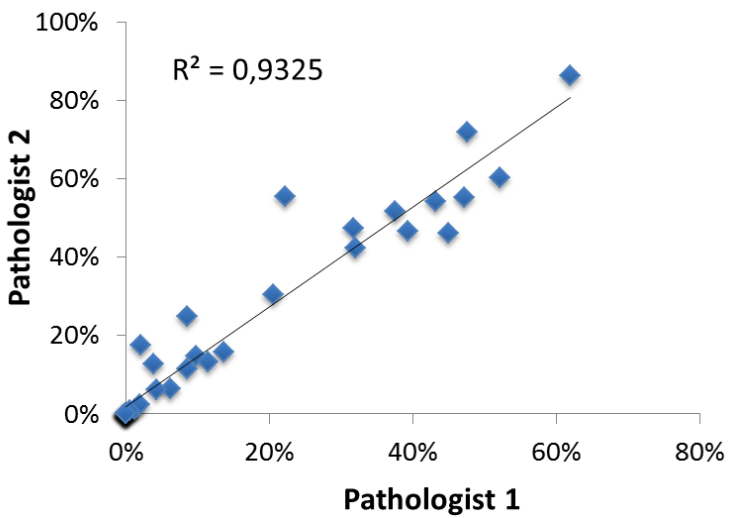
- Green – G3
- Yellow – G4

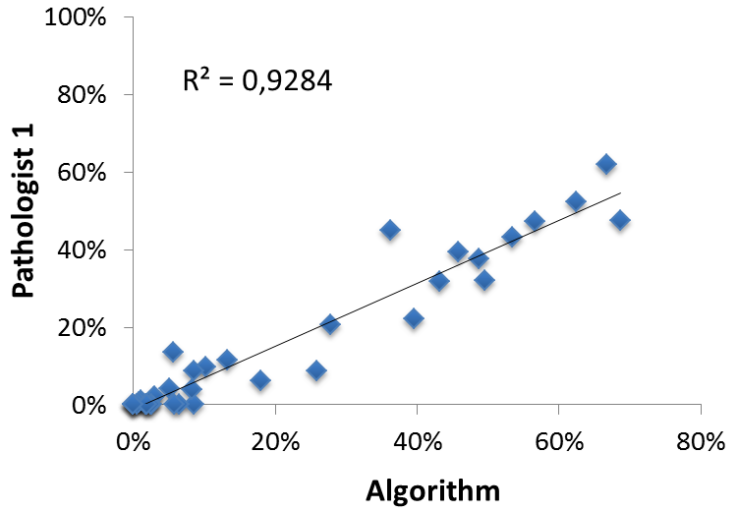


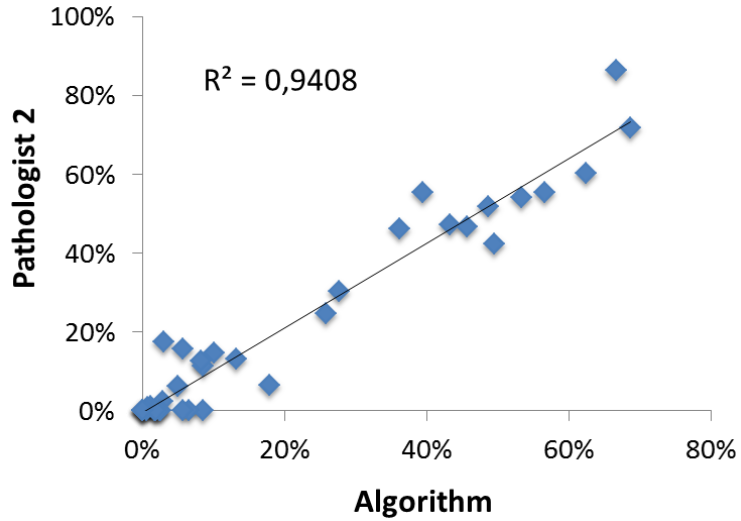
LUND UNIVERSITY

Results

Area covered by cancer in each biopsy







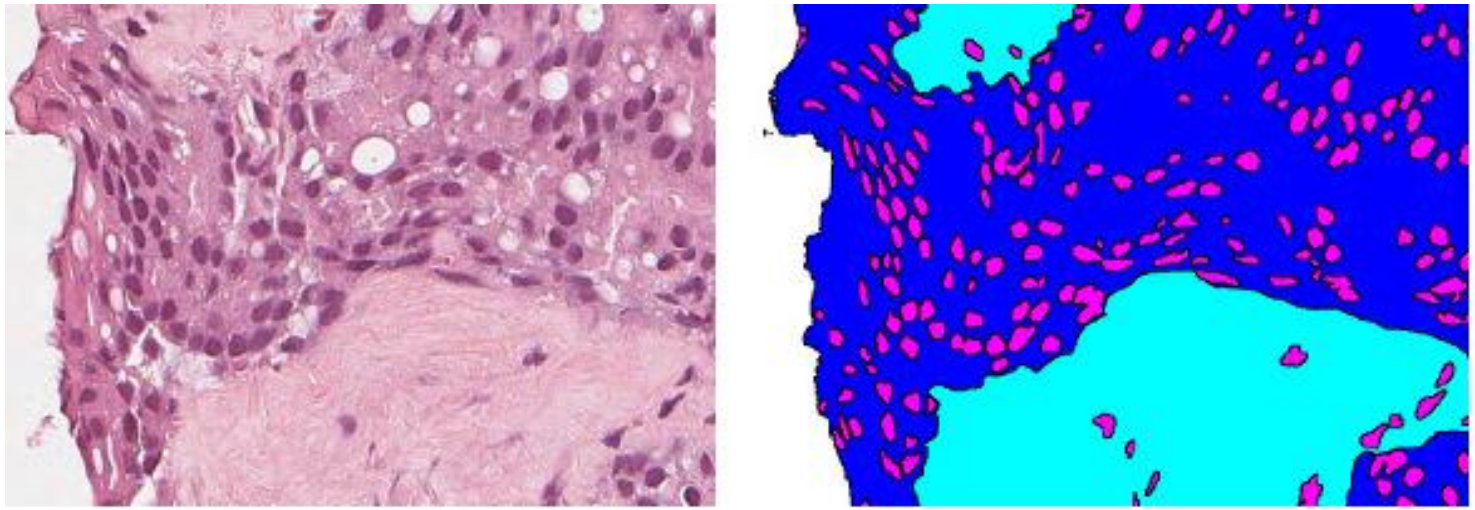
Correlation (R^2) of biopsy area

	Alg vs Path1	Alg vs Path2
cancer	0,928	0,941
GS3	0,789	0,753
GS4	0,627	0,430
GS5	0,329	0,832

LUND UNIVERSITY

Semantic Segmentation using Deep Learning

- Semantic segmentation of the H&E stained tissue into four components:
  - Background - white
  - Stroma – light blue
  - Epithelial Cytoplasm – dark blue
  - Nuclei – Purple

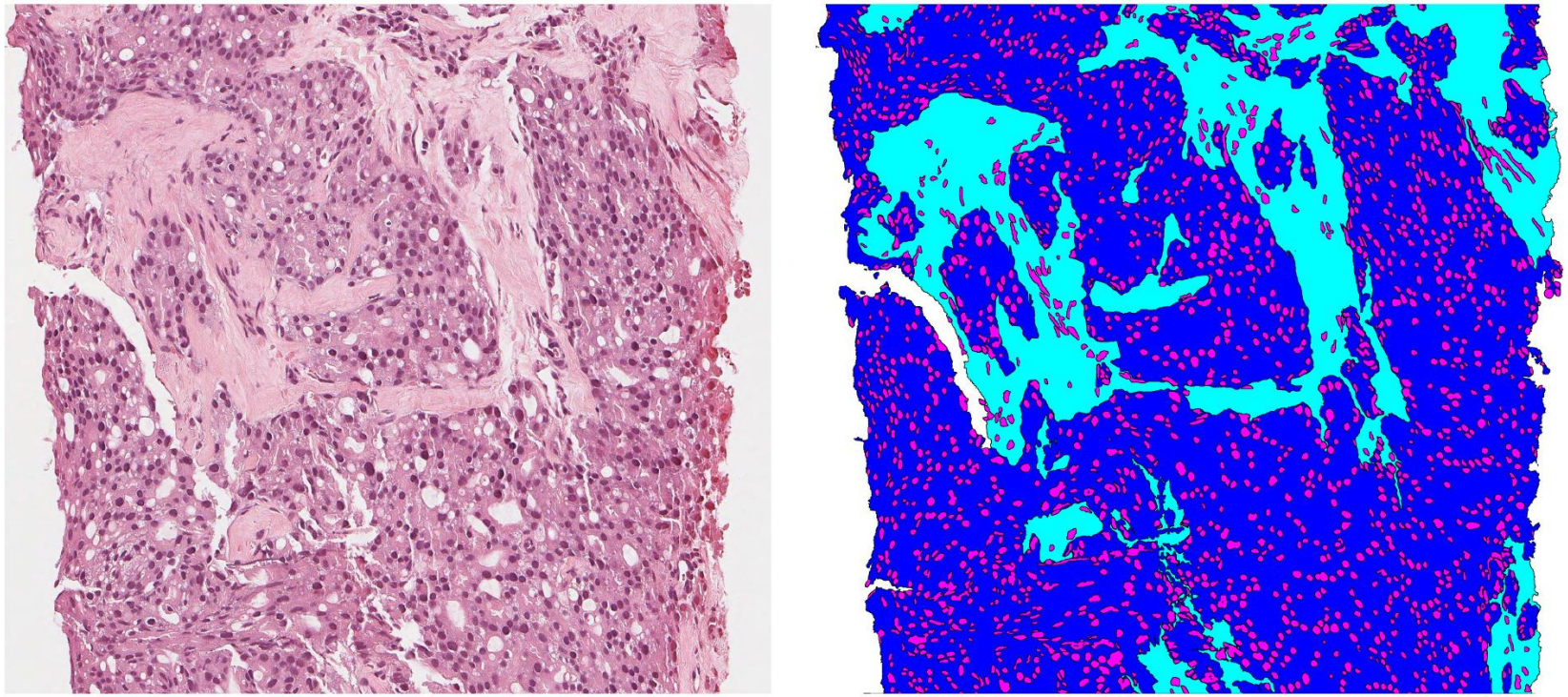


Example of data and corresponding ground truth

Isaksson, J., Arvidsson, I., Åström, K., & Heyden, A. (2017, May). Semantic segmentation of microscopic images of H&E stained prostatic tissue using CNN. In *International Joint Conference on Neural Networks (IJCNN), 2017* (pp. 1252-1256). IEEE.

Dataset

- Images in 20x magnification
- Manually annotated ground truth for each pixel
- Two pixel wide border between all adjacent classes

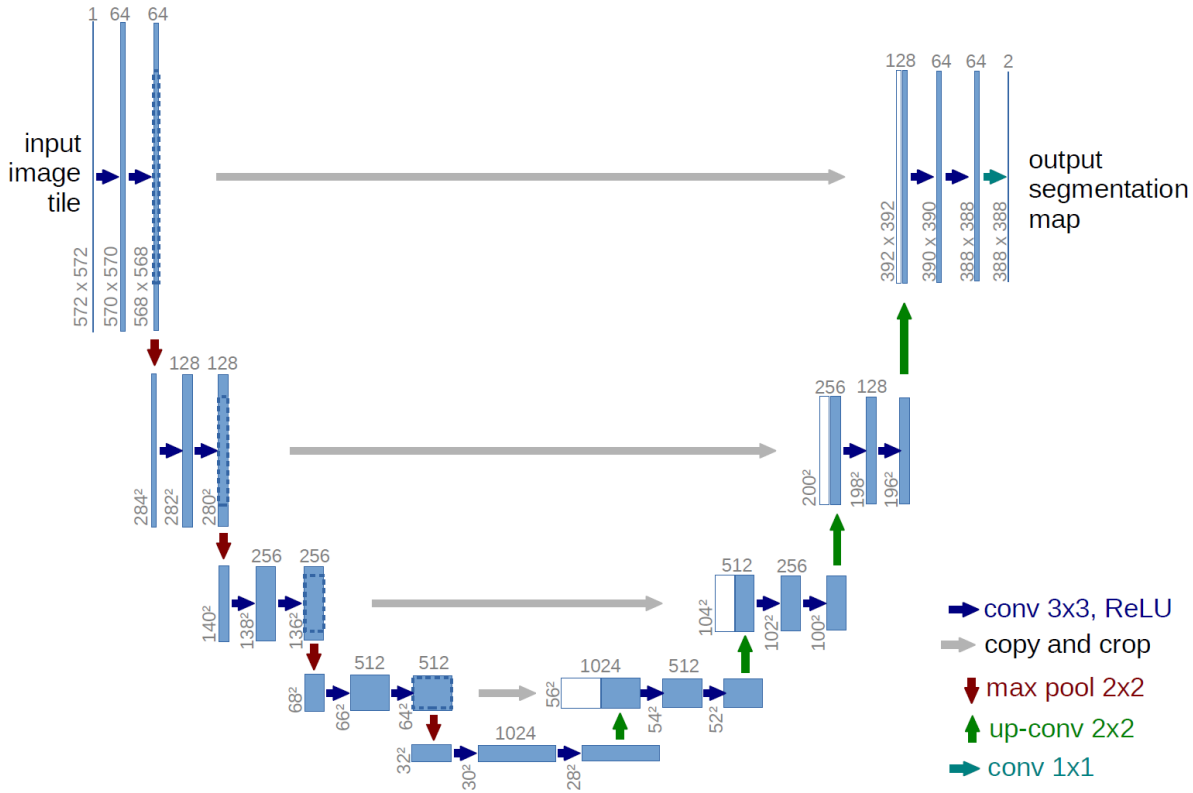


Example of data and corresponding ground truth

LUND UNIVERSITY

Semantic Segmentation using Deep Learning, Network Design: U-net*

- Semantic segmentation using a CNN with up-convolution layers
- Outperformed prior best method on ISBI challenge 2012
- Won the ISBI cell tracking challenge 2015



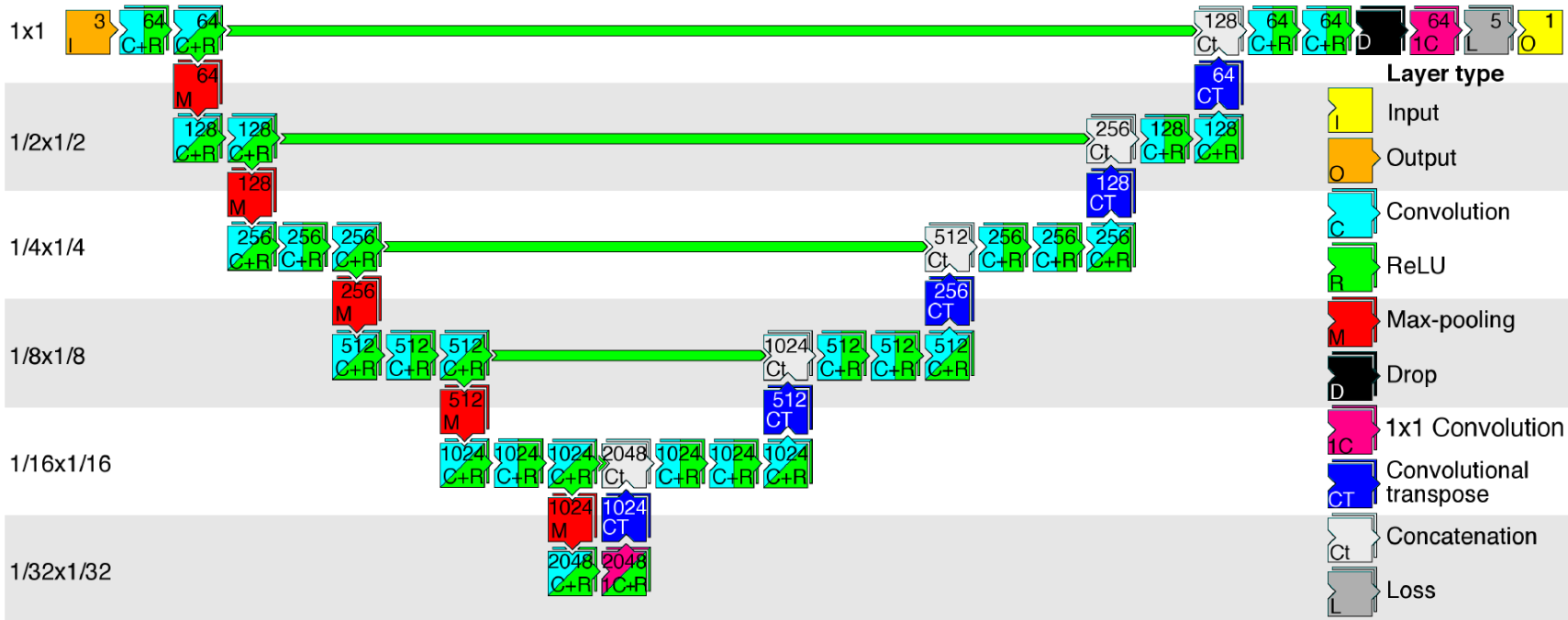
*Original U-net architecture, proposed by O. Ronneberger et. al**

* O. Ronneberger, P. Fischer, and T. Brox
U-net: Convolutional networks for biomedical image segmentation
in MICCAI, pp. 234–241, Springer, 2015.

LUND UNIVERSITY

Semantic Segmentation using Deep Learning, Network Design

- Differences from the U-Net design:
  - Segmentation into four classes instead of two
  - Zero-padding, gives same size of the output as the input
  - One more layer



Our network design

LUND UNIVERSITY

Training Our Network

- Implemented using MatConvNet
- Trained from scratch on patches of size 250 x 250 pixels
- Data augmentation by
  - Distortion by fixed displacement map
  - Flipping
  - Rotating

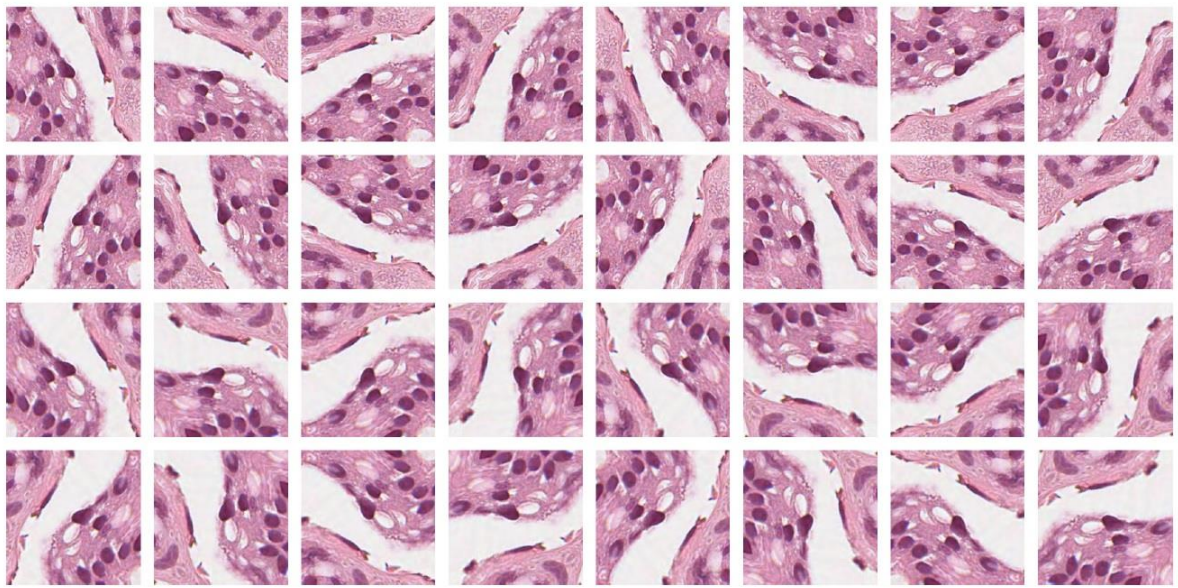


Illustration of the performed data augmentation

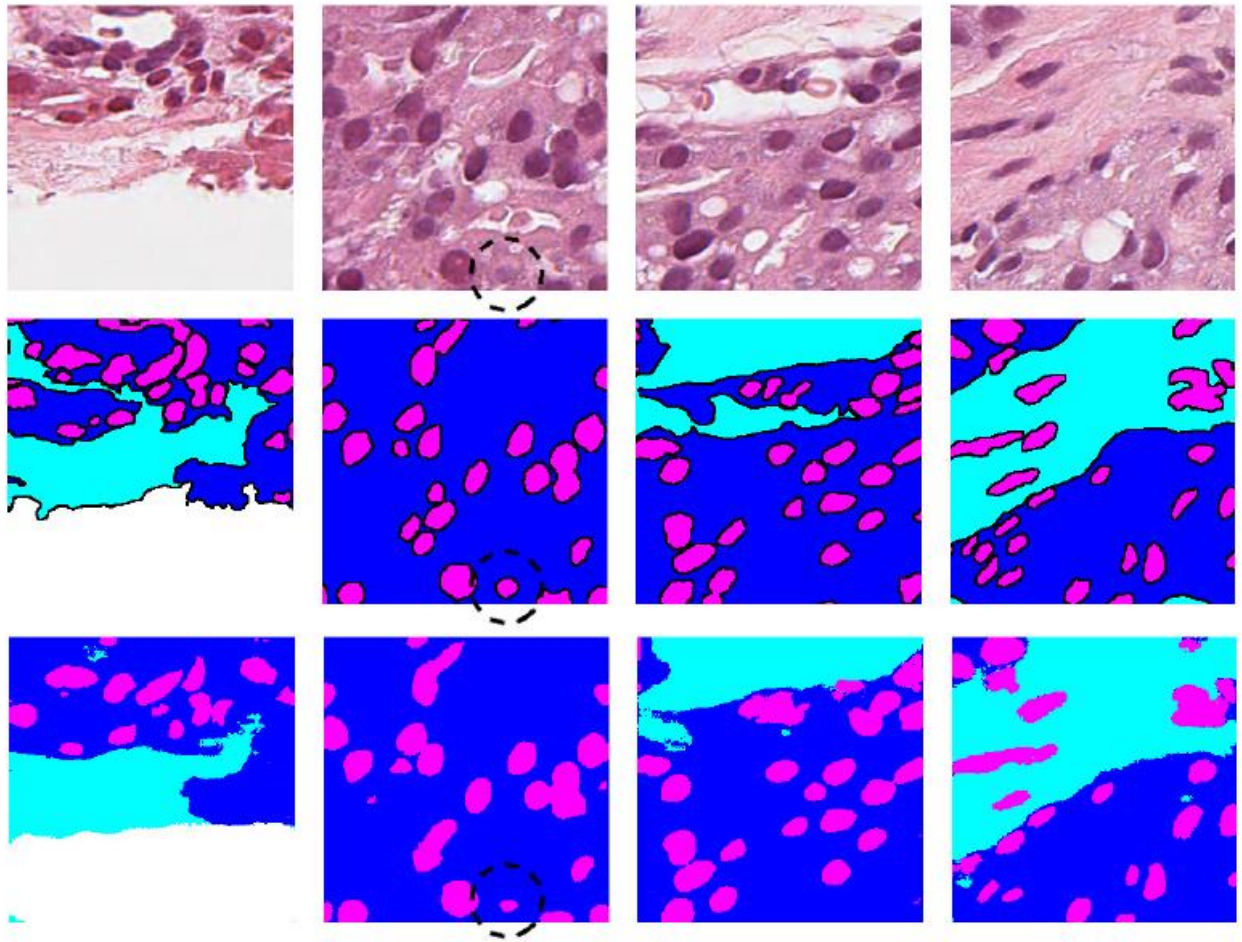
LUND UNIVERSITY

Results, Semantic Segmentation

Original Images

Ground Truths

Results



LUND UNIVERSITY

Result, Semantic Segmentation

Intersection over union

Background	Stroma	Epithelial Cytoplasm	Nuclei	Mean
84%	75%	88%	72%	80%

- For the task of nuclei segmentation we get similar results as previous works
- No previous work on segmentation of stroma and epithelial cytoplasm has been found

Ground truth statistics, image percentage of each class

Background	Stroma	Epithelial Cytoplasm	Nuclei
11.1%	20.0%	49.5%	10.9%

LUND UNIVERSITY

Results and Discussion, Accuracy

Confusion matrix

		Prediction			
		Background	Stroma	Epithelial Cytoplasm	Nuclei
Ground Truth	Background	0.84	0.00	0.07	0.09
	Stroma	0.00	0.78	0.01	0.21
	Epithelial Cytoplasm	0.00	0.01	0.85	0.14
	Nuclei	0.00	0.01	0.03	0.96

Ground truth statistics, image percentage of each class

Background	Stroma	Epithelial Cytoplasm	Nuclei
11.1%	20.0%	49.5%	10.9%

LUND UNIVERSITY

Summary and Future Work

- Classification of patches into four classes using CNN
- Semantic segmentation into four classes using a CNN

- Large variations in staining from different labs/hospitals
- Data from several hospitals
- Data augmentation and invariance
- Full image segmentation and classification
- (How should the methods be used, UI, field-testing)

LUND UNIVERSITY

Acknowledgements

eSSENCE

AIDA, Analytic Imaging Diagnostic Arena, medtech4health

Vinnova (2015-04740)

Felicia-Elena Marginean, Athanasios Simoulis and Agnieszka Krzyzanowska have scanned and annotated the images. Images from Linköping University Hospital were provided by Claes Lundström. Software support was provided by Sectra.

LUND UNIVERSITY

LUND
UNIVERSITY