Learning with Limited Supervision in Medical Imaging

Giacomo Boracchi,
Masaryk University, Brno, May 23th 2025

https://boracchi.faculty.polimi.it/

Giacomo Boracchi

- Mathematician (Università Statale degli Studi di Milano 2004),
- PhD in Information Technology (DEIB, Politecnico di Milano 2008)
- Associate Professor since 2019 at DEIB (Computer Science), Polimi

Research Interests are mathematical and statistical methods for:

- Image / Signal analysis and processing
- Unsupervised learning, change / anomaly detection

Major Courses:

- Mathematical Models and Methods for Image Processing (MSc, Polimi)
- Artificial Neural Networks and Deep Learning (MSc Polimi, Bocconi)
- Advanced Deep Learning Models And Methods (PhD, Polimi)
- Computer Vision (MSc Bocconi, USI 2020)

The Team

We are 3 faculties, 10 PhD students, 1 Research Assistant... and 20+ MSc students!

Giacomo Boracchi

Luca Magri

Federica Arrigoni

Riccardo Margheritti

Michele Craighero

Edoardo Peretti

Andrea Diecidue

Roberto Basla

Rakshith Madhavan

Andrea Ferraris

Diego Martin

Luca Alessandrini

Andrea Marelli

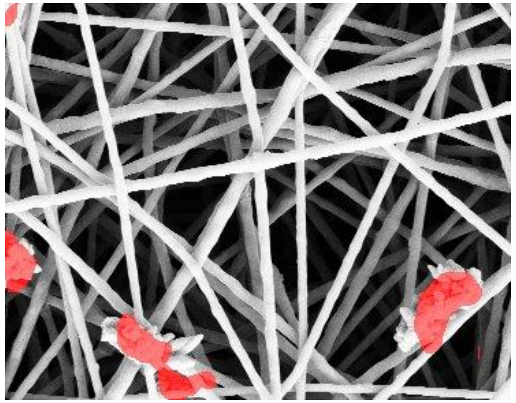
Carlo Sgaravatti

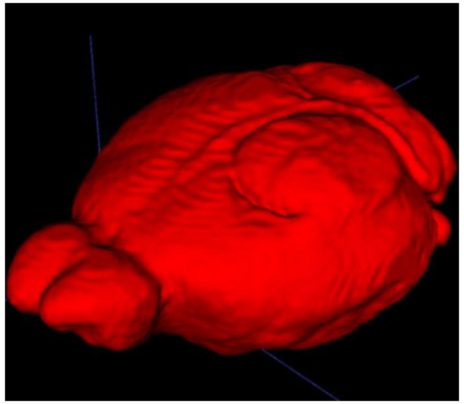
Computer Vision Research

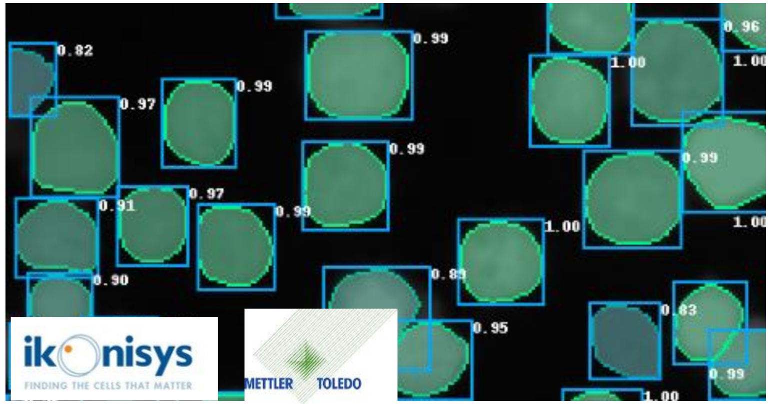
Giacomo Boracchi, Luca Magri, Federica Arrigoni

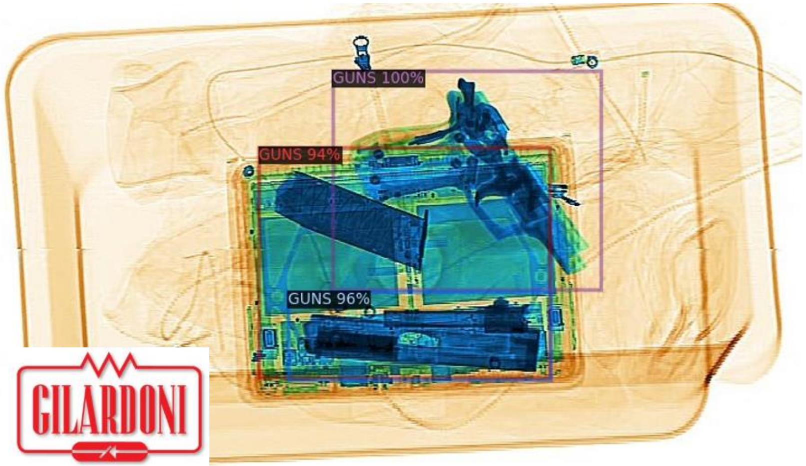
Research themes and relevant projects:

- **Deep Learning for Visual Recognition:** Object Detection and Segmentation (CCD, SPAD, X-ray, 2D/3D Medical, Aerial Images), Anomaly Detection (images, signals), Learning under limited supervision.









GILARDONI

PERIVALLON

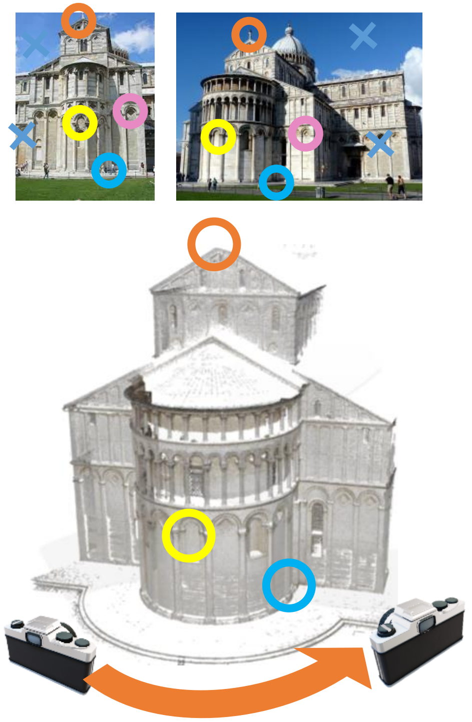
with Prof. Piero Fraternali

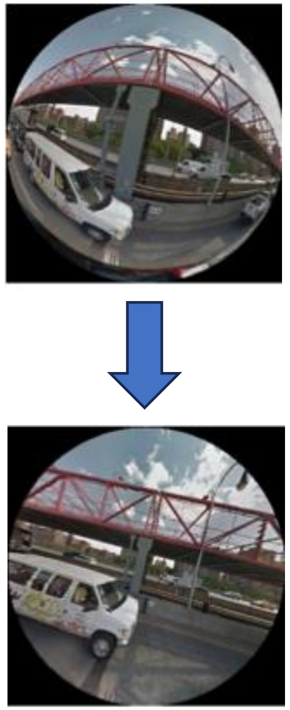
Computer Vision Research

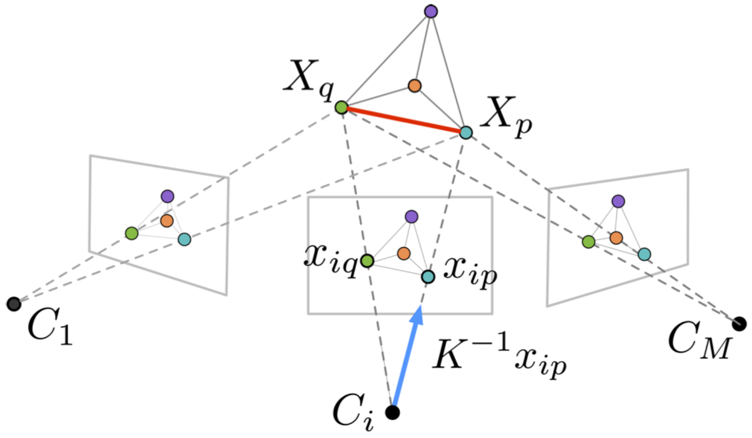
Giacomo Boracchi, Luca Magri, Federica Arrigoni

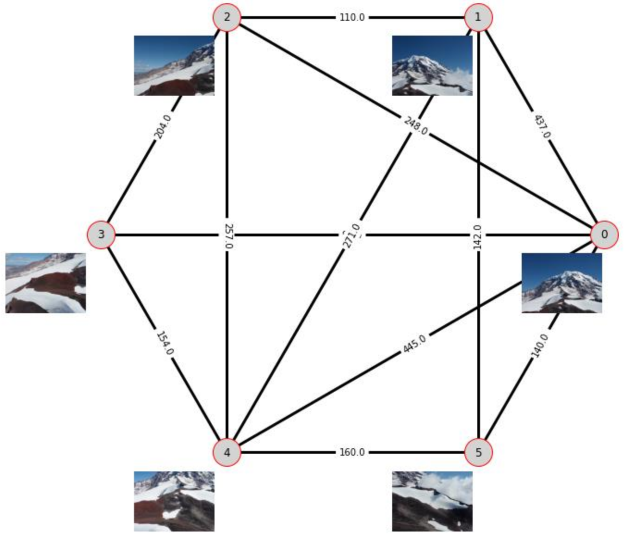
Research themes and relevant projects:

- **Deep Learning for Visual Recognition:** Object Detection and Segmentation (CCD, SPAD, X-ray, 2D/3D Medical, Aerial Images), Anomaly Detection (images, signals), Learning under limited supervision.
- **Multi-view Geometry:** 3D Reconstruction, Calibration (Conventional / Event Cameras, X-ray, LiDar, hybrid systems), 6DOF Pose Estimation, Foundational Research.







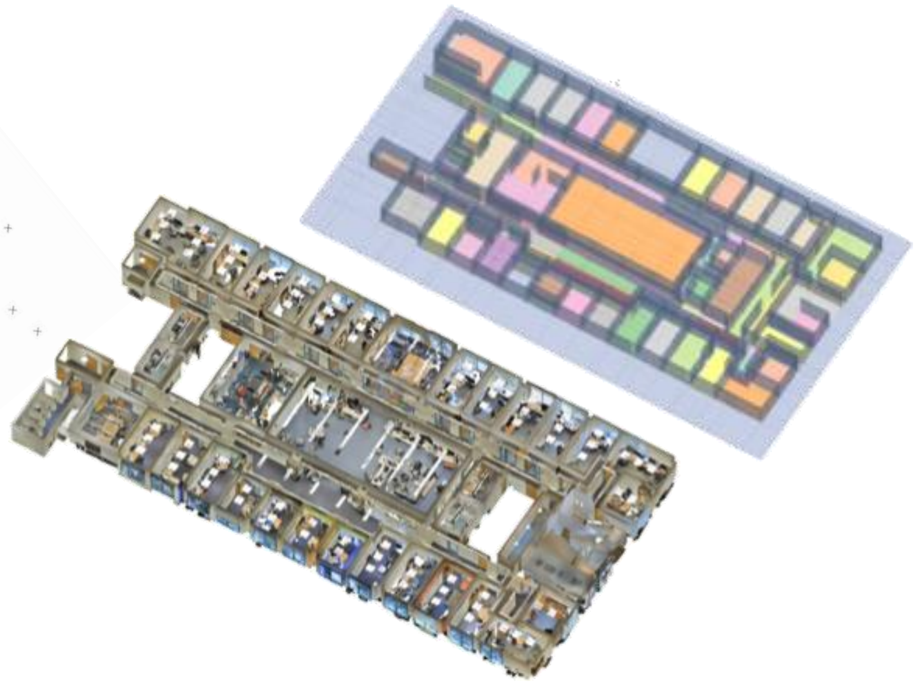


Computer Vision Research

Giacomo Boracchi, Luca Magri, Federica Arrigoni

Research themes and relevant projects:

- **Deep Learning for Visual Recognition:** Object Detection and Segmentation (CCD, SPAD, X-ray, 2D/3D Medical, Aerial Images), Anomaly Detection (images, signals), Learning under limited supervision.
- **Multi-view Geometry:** 3D Reconstruction, Calibration (Conventional / Event Cameras, X-ray, LiDar, hybrid systems), 6DOF Pose Estimation, Foundational Research.
- **Pattern Recognition:** Clustering, Robust Model Fitting (Scan2BIM, Template Detection), Quantum Computer Vision.

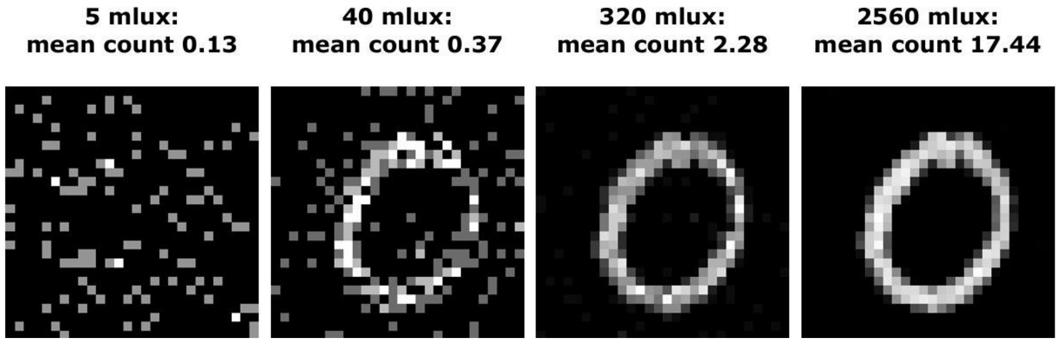


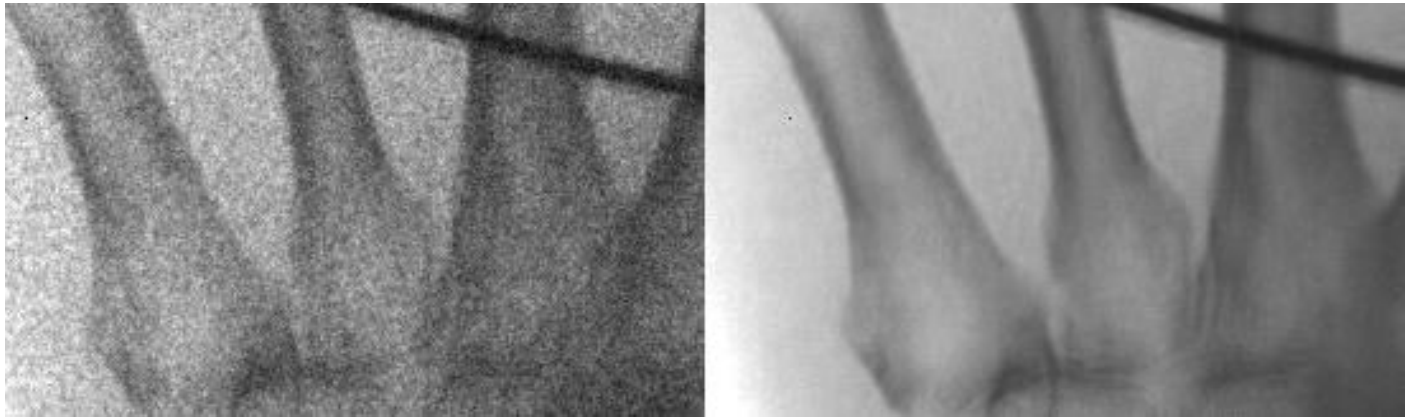
Computer Vision Research

Giacomo Boracchi, Luca Magri, Federica Arrigoni

Research themes and relevant projects:

- **Deep Learning for Visual Recognition:** Object Detection and Segmentation (CCD, SPAD, X-ray, 2D/3D Medical, Aerial Images), Anomaly Detection (images, signals), Learning under limited supervision.
- **Multi-view Geometry:** 3D Reconstruction, Calibration (Conventional / Event Cameras, X-ray, LiDar, hybrid systems), 6DOF Pose Estimation, Foundational Research.
- **Pattern Recognition:** Clustering, Robust Model Fitting (Scan2BIM, Template Detection), Quantum Computer Vision.
- **Image/Signal Processing and Analysis:** Image Enhancement and Restoration (X-ray, CCD, SPAD), Change and Outlier Detection (Optical fiber signals)



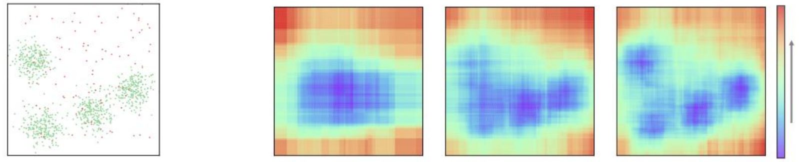


Computer Vision Research

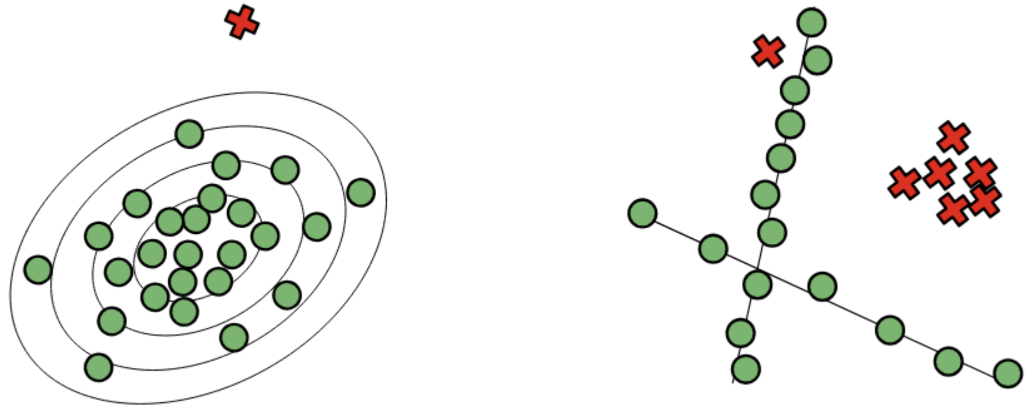
Giacomo Boracchi, Luca Magri, Federica Arrigoni

Research themes and relevant projects:

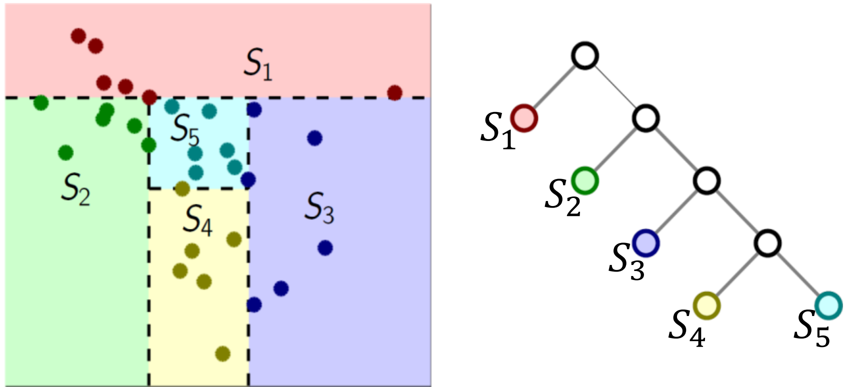
- **Deep Learning for Visual Recognition:** Object Detection and Segmentation (CCD, SPAD, X-ray, 2D/3D Medical, Aerial Images), Anomaly Detection (images, signals), Learning under limited supervision.
- **Multi-view Geometry:** 3D Reconstruction, Calibration (Conventional / Event Cameras, X-ray, LiDar, hybrid systems), 6DOF Pose Estimation, Foundational Research.
- **Pattern Recognition:** Clustering, Robust Model Fitting (Scan2BIM, Template Detection), Quantum Computer Vision.
- **Image/Signal Processing and Analysis:** Image Enhancement and Restoration (X-ray, CCD, SPAD), Change and Outlier Detection (Optical fiber signals)
- **Change and Anomaly Detection:** Design of change-detection tests for high-dimensional data (QuantTree), Sequential Monitoring, Datastream Mining, Anomaly Detection (tree based methods).



(a) *Data stream $\boldsymbol{x}_1, \dots, \boldsymbol{x}_t \in \mathbb{R}^d$.* **(b)** *Anomaly scores s at different time instants t, from left to right.*



(a) *Density.* **(b)** *Structure.*



Computer Vision Research

Giacomo Boracchi, Luca Magri, Federica Arrigoni

Research themes and relevant projects:

- **Deep Learning for Visual Recognition:** Object Detection and Segmentation (CCD, SPAD, X-ray, 2D/3D Medical, Aerial Images), Anomaly Detection (images, signals), Learning under limited supervision.
- **Multi-view Geometry:** 3D Reconstruction, Calibration (Conventional / Event Cameras, X-ray, LiDar, hybrid systems), 6DOF Pose Estimation, Foundational Research.
- **Pattern Recognition:** Clustering, Robust Model Fitting (Scan2BIM, Template Detection), Quantum Computer Vision.
- **Image/Signal Processing and Analysis:** Image Enhancement and Restoration (X-ray, CCD, SPAD), Change and Outlier Detection (Optical fiber signals)
- **Change and Anomaly Detection:** Design of change-detection tests for high-dimensional data (QuantTree), Sequential Monitoring, Datastream Mining, Anomaly Detection (tree based methods).

Relevant results achieved:

- Selected Industrial Partners: STMicroelectronics (Quality Inspection), Gilardoni Raggi X (Explosive and Weapon Detection), Cisco Photonics (Change/Anomaly Detection), Huawei (SPAD imaging).
- Fundrising in tech-transfer projects: more than 1.5M€ in research projects and 0.7M€ in PhD grants.
- Publications: 200+ papers, including 25 Q1 journal articles , 26 top-tier conference papers, 6 patents

Research collaborations

Tampere University

mpi max planck institut informatik

CTU CZECH TECHNICAL UNIVERSITY IN PRAGUE

ETH zürich

UNIVERSITÀ DEGLI STUDI DI UDINE

MUNI Masaryk University

UNIVERSITÀ DEGLI STUDI DI BERGAMO

ISTITUTO DI RICERCHE FARMACOLOGICHE MARIO NEGRI · IRCCS

ULB UNIVERSITÉ LIBRE DE BRUXELLES

NORCE

FONDAZIONE IRCCS ISTITUTO NAZIONALE DEI TUMORI

Our Industrial Reach

	Industry 4.0	Medical	Security
Deep Learning for Visual Recognition	ST life.augmented, CISCO, faital, ML cube	ikonisys FINDING THE CELLS THAT MATTER, METTLER TOLEDO	GILARDONI
Computer Vision and Pattern Recognition	CISCO, MS Meccanica del Sarca, T&O ITALIA, WE-WEAR the future of fashion, DANIELI	digitec DIGITAL INNOVATION	Cleafy
Image Processing and Analysis	HUAWEI, GILARDONI, LISALAB, CISCO	digitec DIGITAL INNOVATION, ST life.augmented	GILARDONI

Our reach

	Industry 4.0	Medical	Security
Deep Learning for Visual Recognition	Silicon Wafer Manufacturing OTDR Event Detection Defect Detection	Histological Image Segmentation Cell Segmentation	Hazard detection
Computer Vision and Pattern Recognition	Template detection Defect Detection in Point Clouds 3D Body scanner OCM Anomaly Detection	RGB-D and X-ray Calibration	Data stream monitoring
Image Processing and Analysis	Noise Modeling in SPAD Image Enhancement RON time series monitoring Battery management system	Video Denoising Online ECG Monitoring	Explosive detection Image Enhancement

Signals/Streams

- OTDR Event Detection
- OCM Anomaly Detection
- RON time series monitoring
- Online ECG Monitoring
- Data stream monitoring
- Battery management system

Images

- Silicon Wafer Manufacturing
- Histological Image Segmentation
- Noise Modeling in SPAD
- RGB-D and X-ray Calibration
- Video Denoising
- Image Enhancement
- Explosive detection
- Prohibited Item detection
- Template detection
- Defect Detection
- Cell Segmentation

3D data

- Silicon Wafer Manufacturing
- Defect Detection in Point Clouds
- 3D Body scanner
- RGB-D and X-ray Calibration

TRL ≥ 7

Silicon Wafer Manufacturing

OTDR Event Detection

OCM Anomaly Detection

Template detection

Explosive detection

Image Enhancement

TRL = 5

Online ECG Monitoring

Battery management system

Histological Image Segmentation

RGB-D and X-ray Calibration

Video Denoising

Ongoing

Time Series Generation

RON time series monitoring

Noise Modeling in SPAD

Defect Detection

Cell Segmentation

This Talk

- We have a background from image processing / computer vision methods, we are not expert in medical imaging itself.
- We got involved with medical companies / research institutes to solve of their specific problems.
- We typically use methods/principles that are otherwise known but
  - never used for solving the specific medical imaging task
  - Cannot be used “off-the shelf”
- I present these collaborations and the way we overcame the data-scarcity problenm.

An Illustrative Case Study Histological Image Segmentation

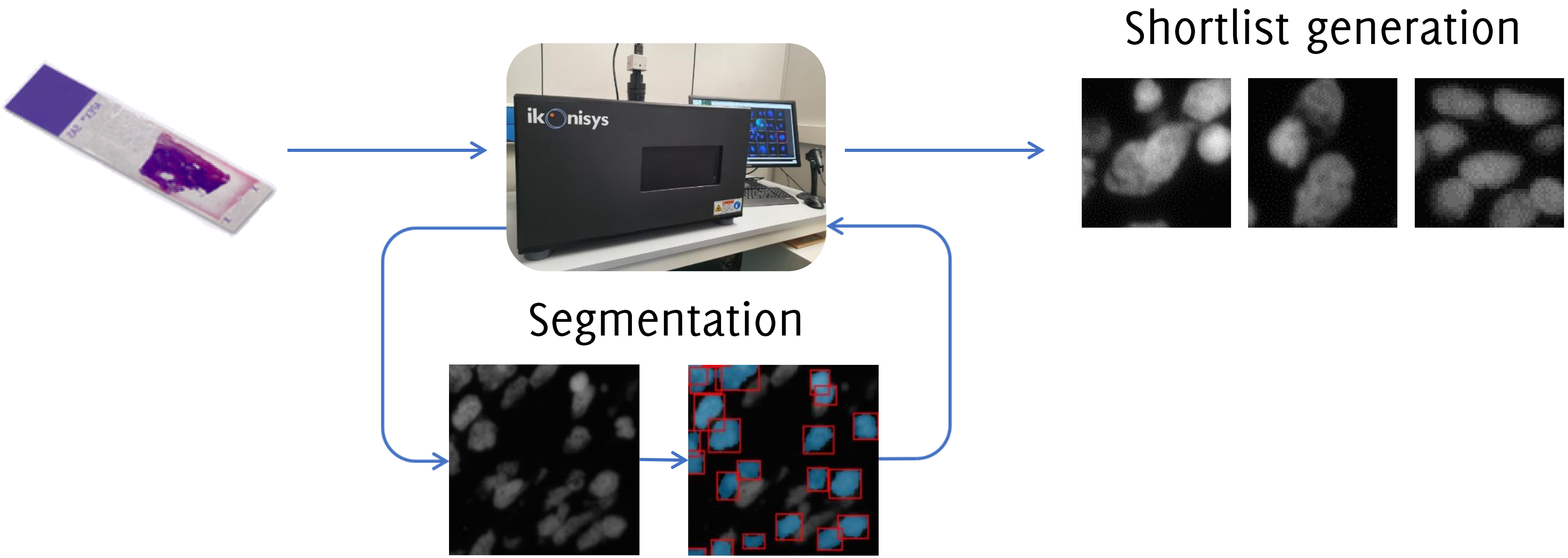
Instance segmentation of Nuclei

ikonisys
FINDING THE CELLS THAT MATTER

Roberto Basla

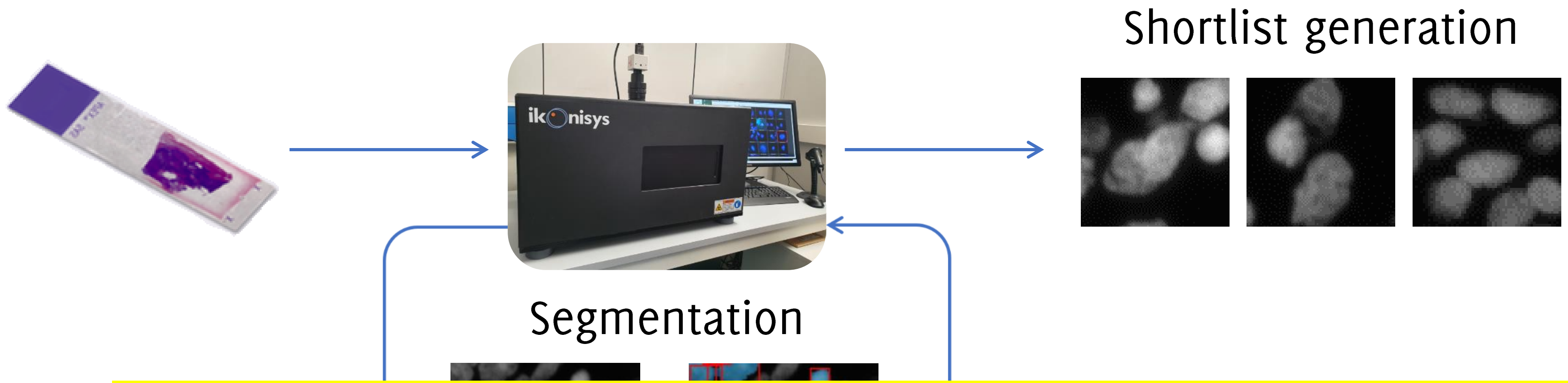
Context: Histological (fluoroscopy) imaging

The Ikoniscope provides a shortlist of interesting nuclei to be presented to the medical professional.



Context: Histological (fluoroscopy) imaging

The Ikoniscope provides a shortlist of interesting nuclei to be presented to the medical professional.



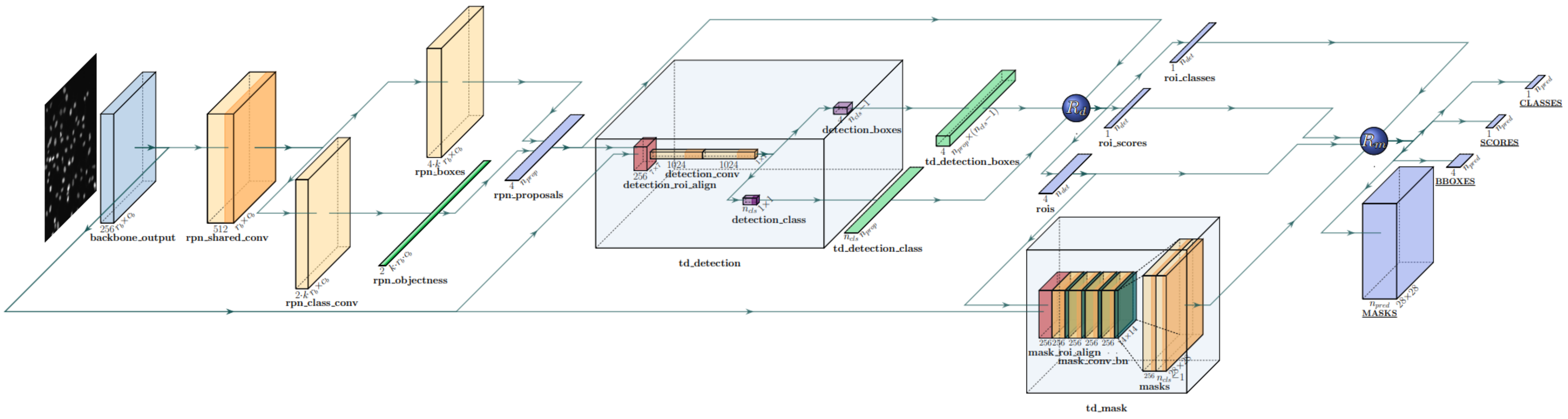
Segmentation was performed with expert-driven methods that poorly generalize to

- **Different tissues**
- **Different types of treatments for the sample**

Solution: Deep learning framework

We have designed a deep learning framework composed of a custom implementation of two (alternative) solutions

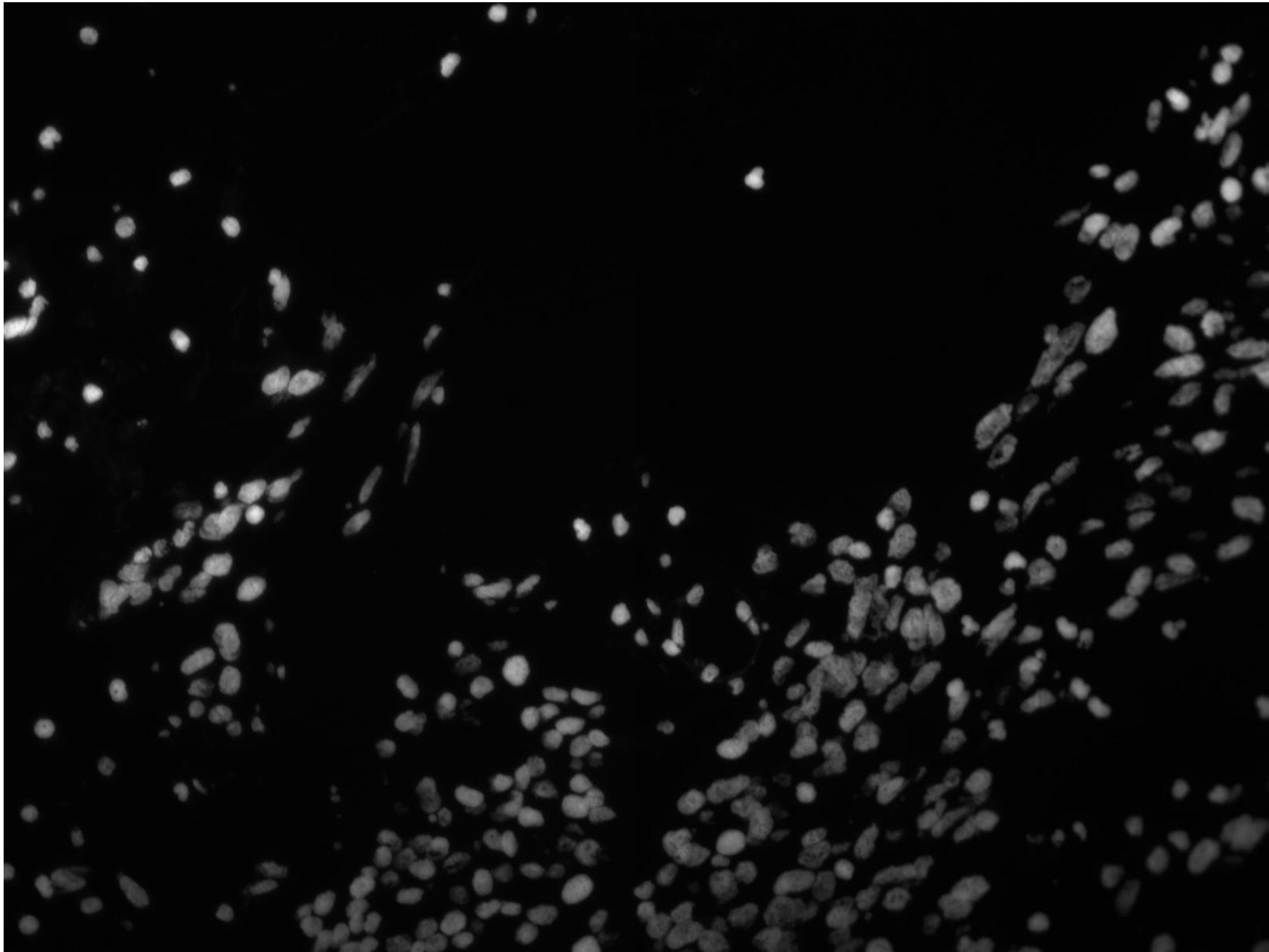
- Mask R-CNN
- Hover-Net



Graham, S., Vu, Q. D., Raza, S. E. A., Azam, A., Tsang, Y. W., Kwak, J. T., & Rajpoot, N. (2019). Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. Medical image analysis, 58, 101563.

Input Image

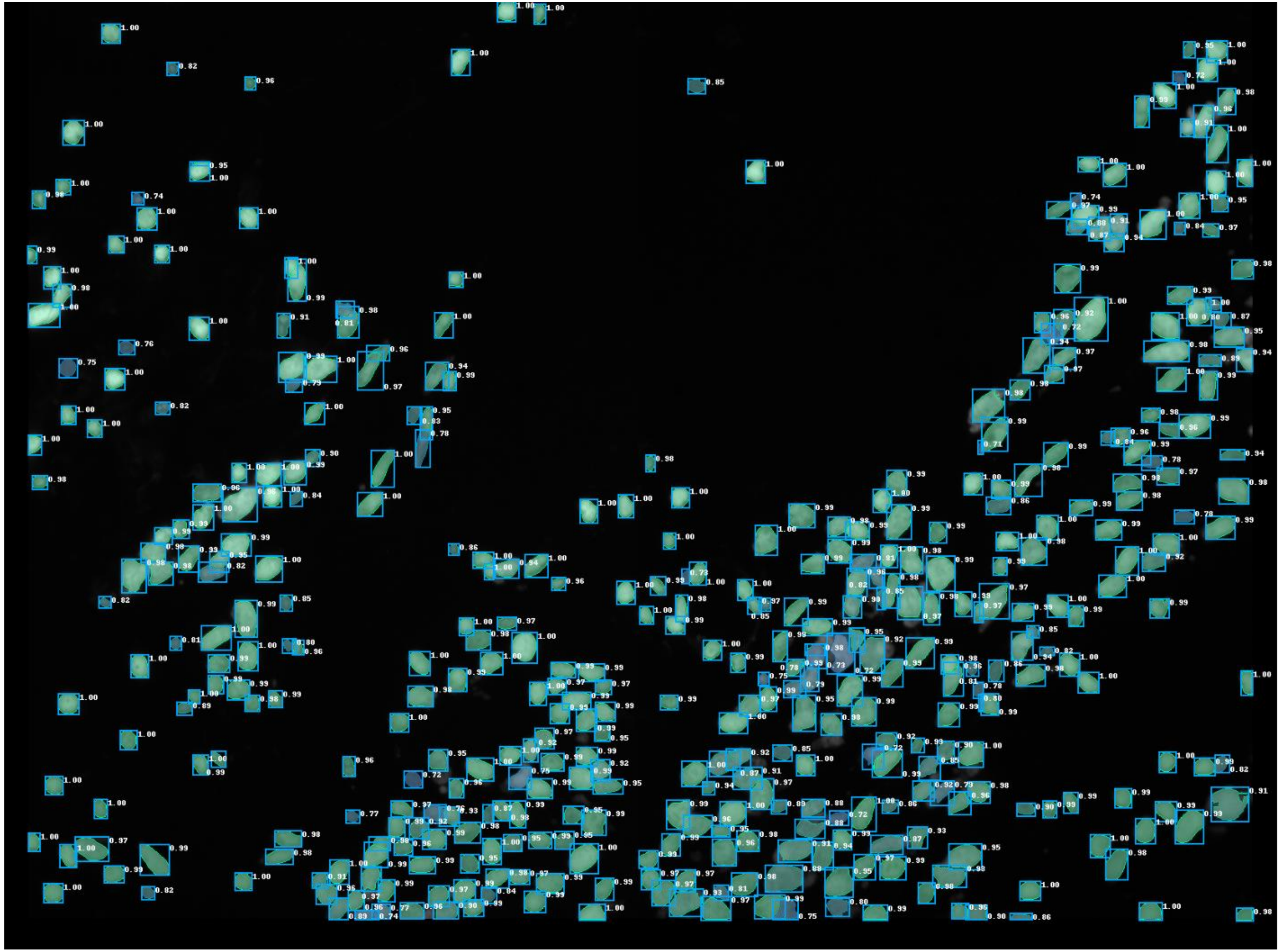
Input: an histological image



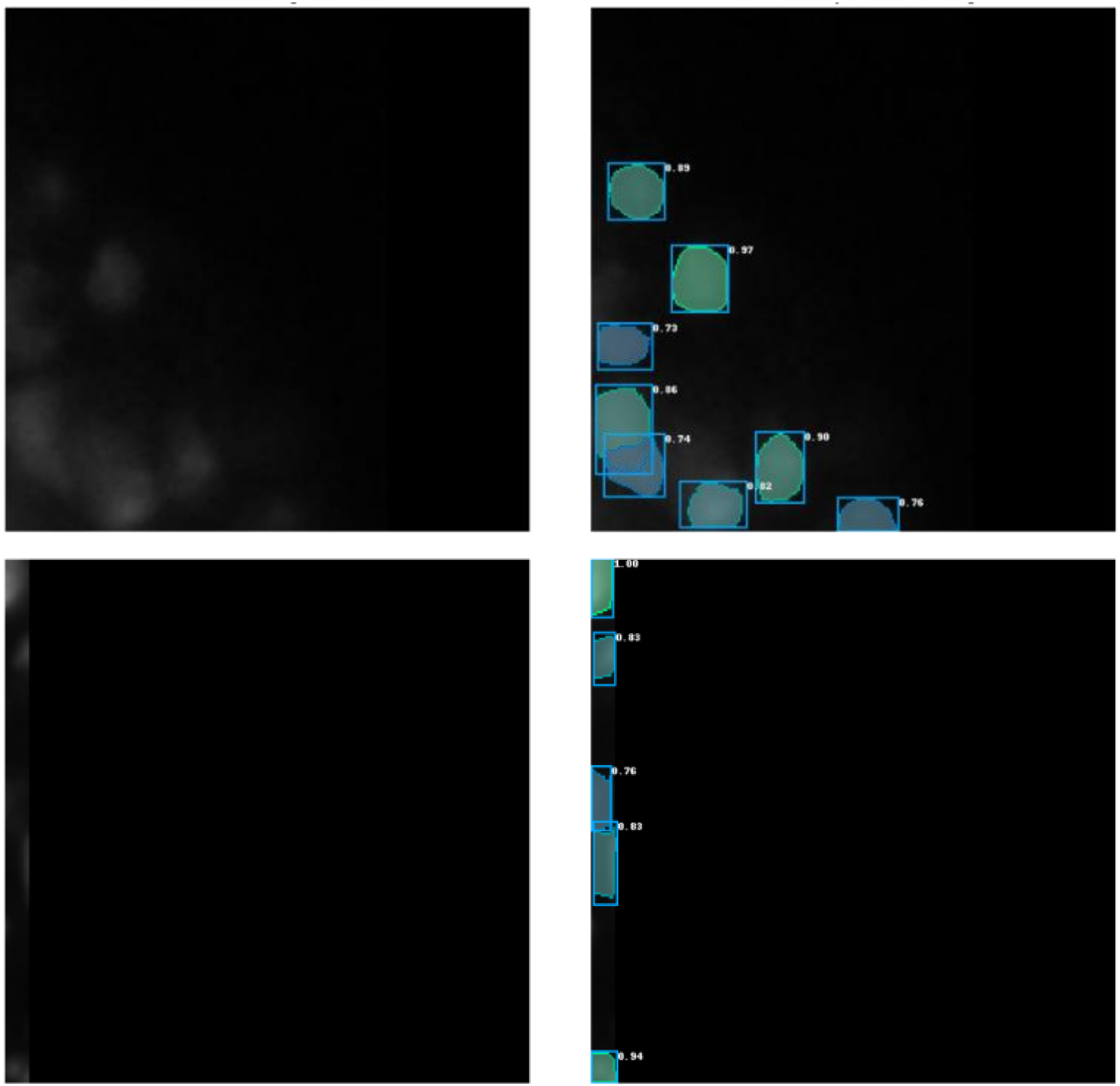
Output Image: Instance Segmentation

Output: each cell is expected to be associated to an individual segment

Possibly class information can be associated to each mask

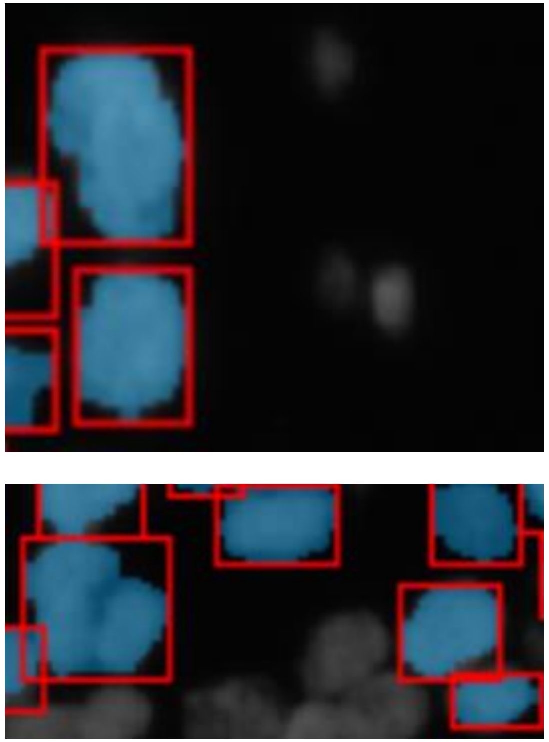


Very effecitve indeed!

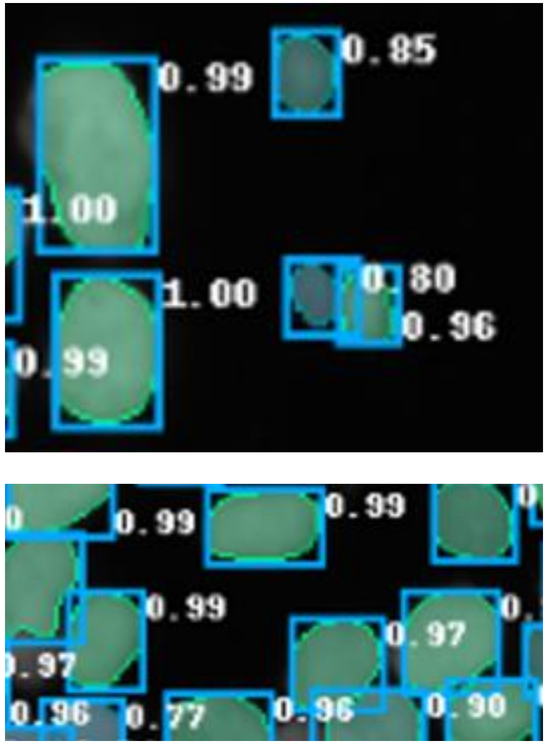


Comparison with expert-driven segmentation

- We released the trained model with interfaces to communicate with the Ikoniscope instrument.
- The model provided by us displays **better performance** than previous geometrical segmentation techniques,
- The new model solves the task of **instance segmentation**, enabling the identification of **overlapping nuclei**.



Geometrical approach



Our model

Gathering Training Data: a Well-Known Problem

- Gathering unlabeled data is relatively easy, **gathering annotations is not**
- In medical images (more than in natural images) it is often important to **quantitatively assess areas** (nr of pixels) covered by a specific class, not just to assign image-wise labels.
- **Annotating** images for **segmentation** is incredibly **time consuming**
- Annotations required an **histopathologist**, which are costly and difficult to gather
- There is a high risk of receiving **inconsistent labelling** (annotator fatigue?)

Solution: Training on publicly available dataset

Image from a private dataset

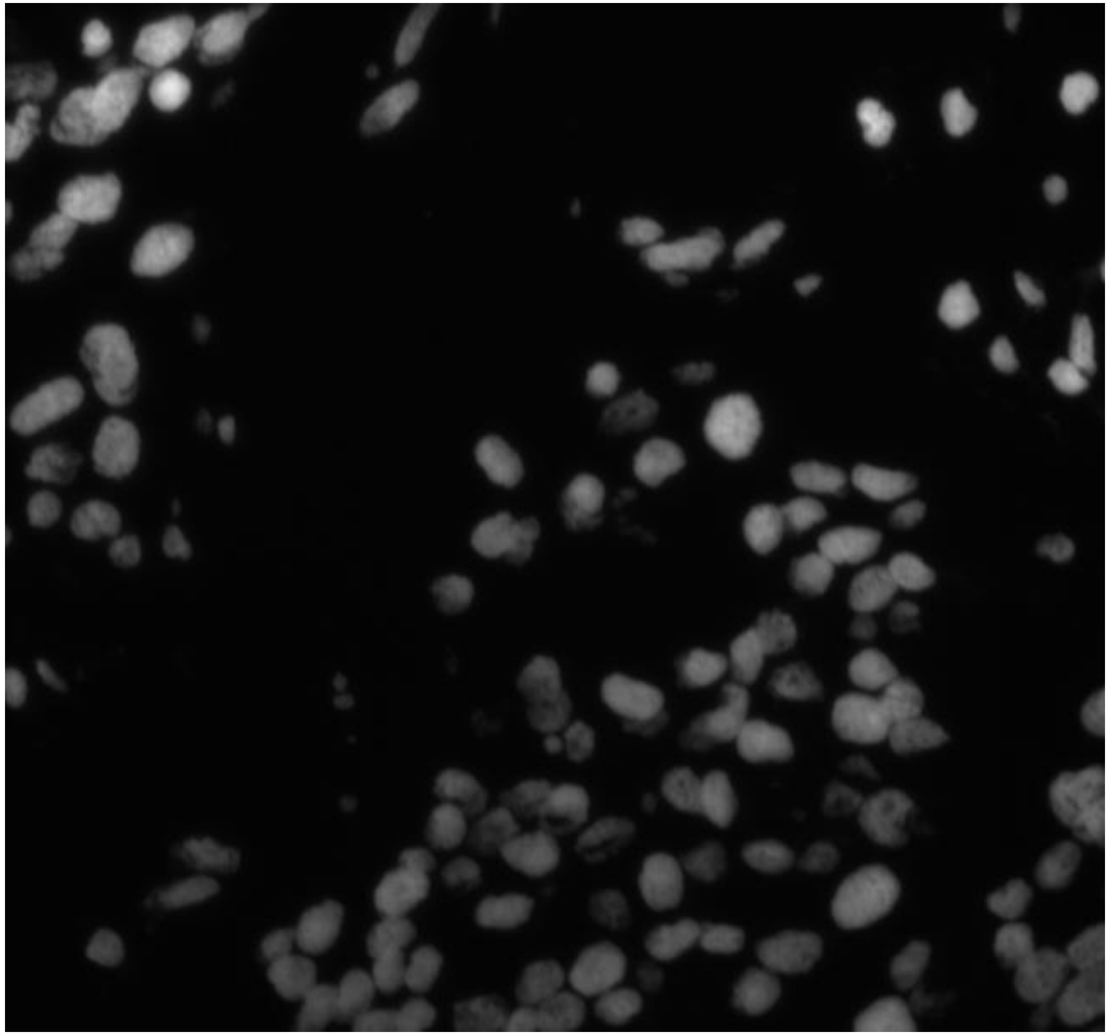
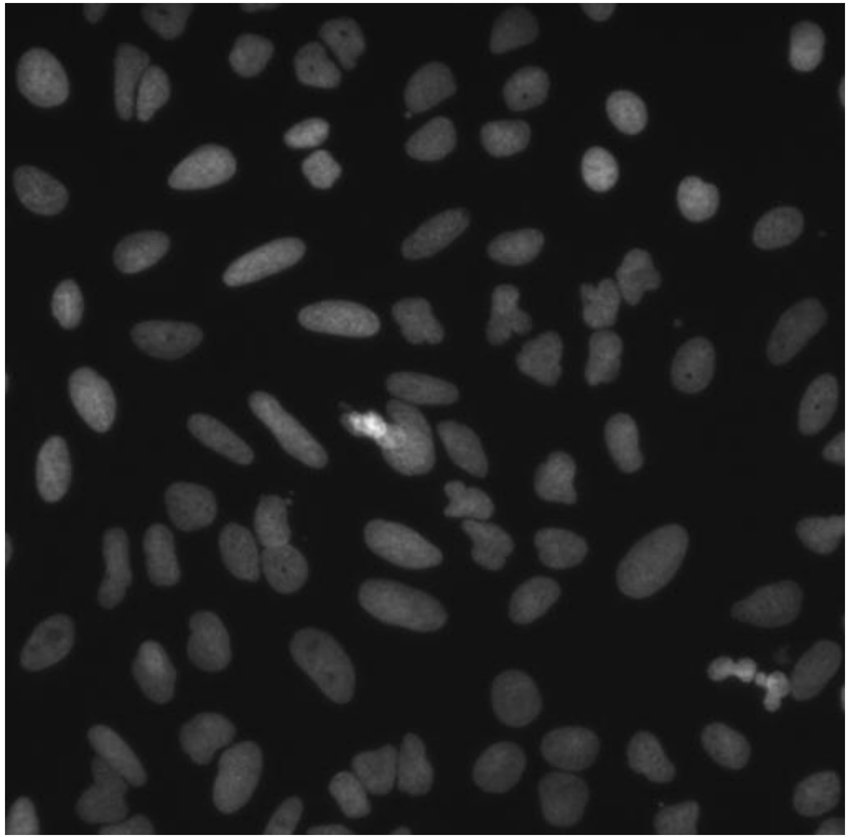


Image from data bowl



Solution: Training on publicly available dataset

Image from a private dataset

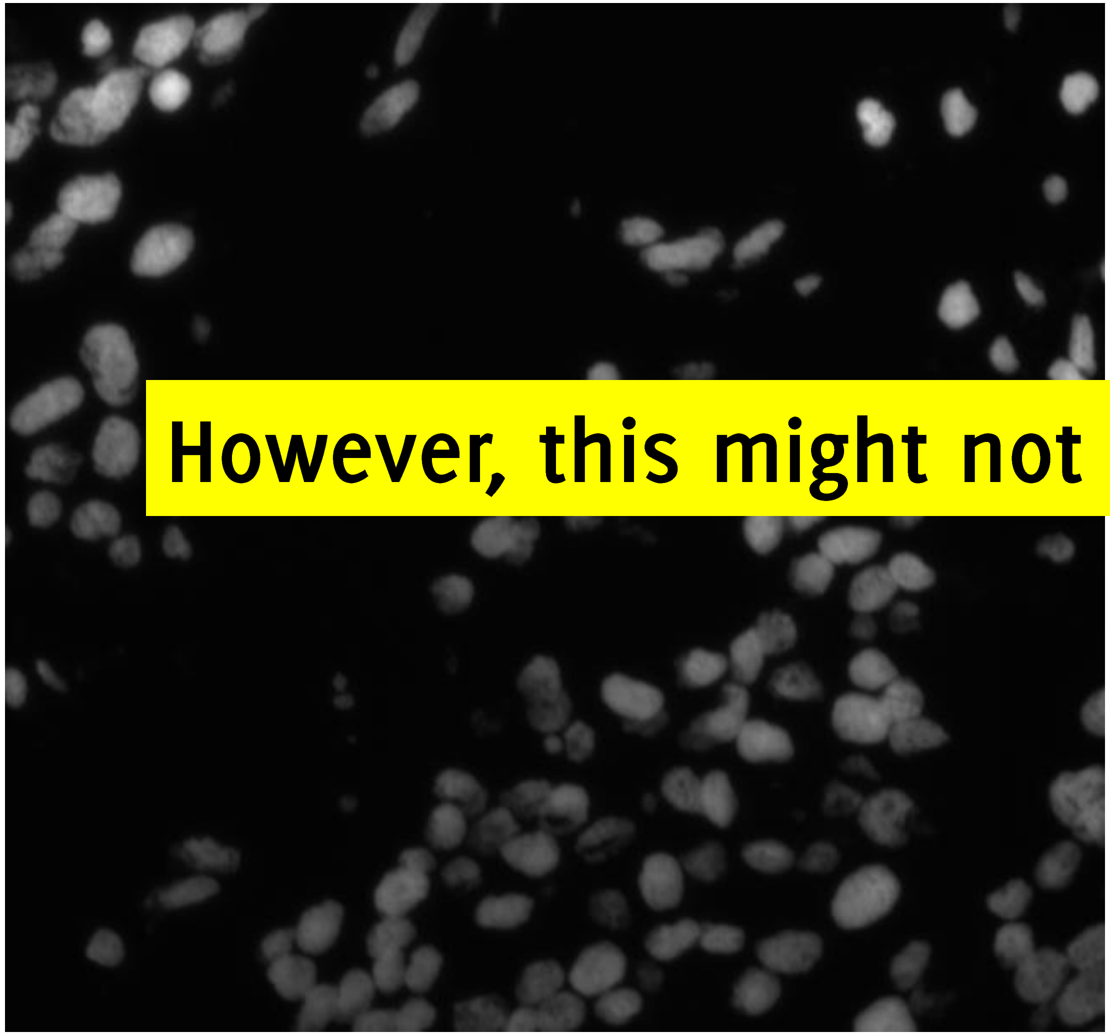
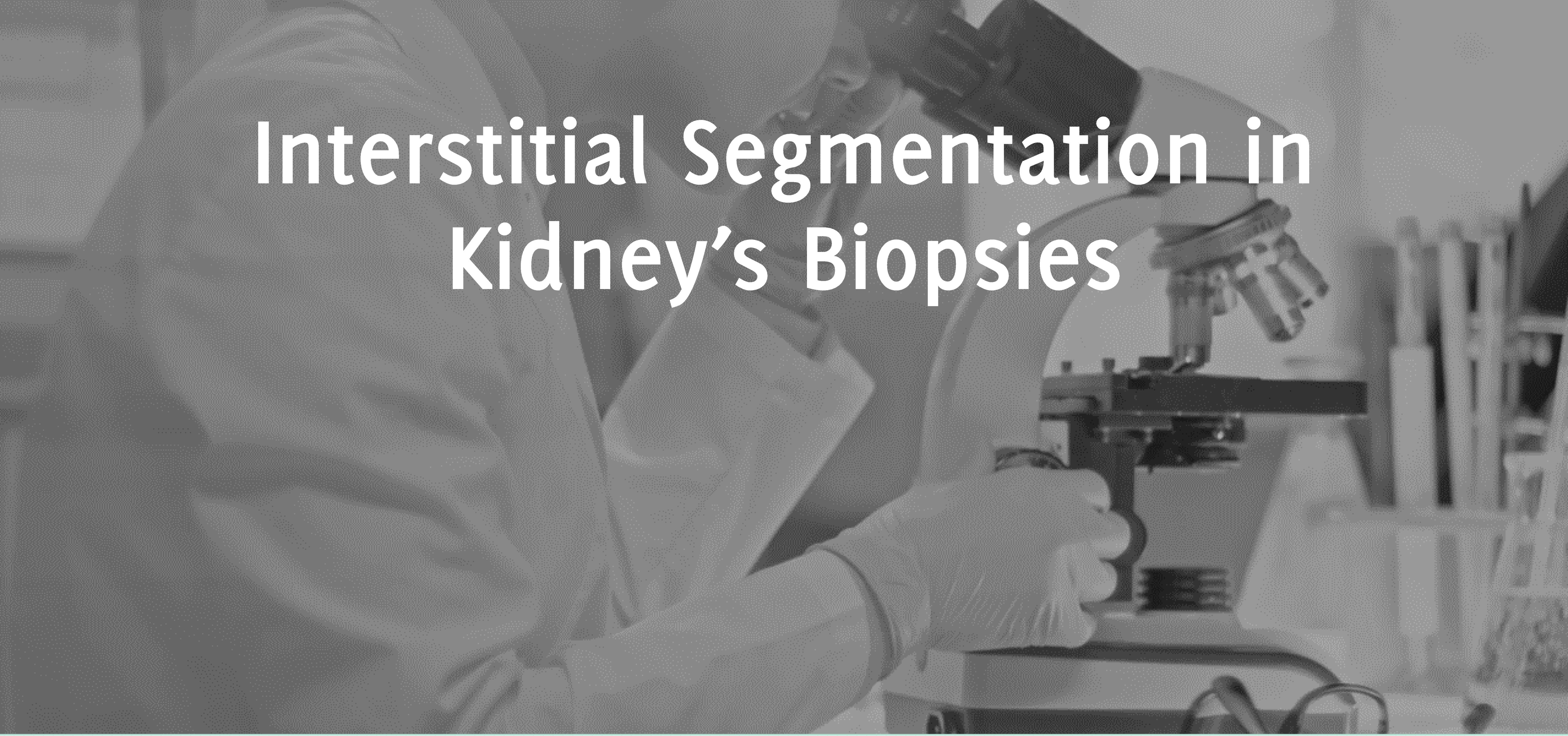


Image from data bowl

However, this might not always be a viable option…

Interstitial Segmentation in Kidney's Biopsies



ISTITUTO DI RICERCHE FARMACOLOGICHE MARIO NEGRI · IRCCS

Deep Learning for Visual Recognition

Patient-specific fine-tuning with scribbles supervision and uncertainty weighting for semantic segmentation: application to kidney biopsies

Alfonso Catelli[a], Ettore Lanzarone[b,*], Sebastiano Rossi[a], Giulia Villa[c], Anna Caroli[c], Paola Rizzo[d], Elisa Scalco[e] and Giacomo Boracchi[a]

[a]*Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB), Politecnico di Milano, Milan, Italy*
[b]*Department of Management, Information and Production Engineering (DIGIP), University of Bergamo, Dalmine (Bg), Italy*
[c]*Department of Biomedical Engineering, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Bergamo, Italy*
[d]*Department of Molecular Medicine, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Bergamo, Italy*
[e]*Institute of Biomedical Technologies (ITB), National Research Council (CNR), Segrate (Mi), Italy*

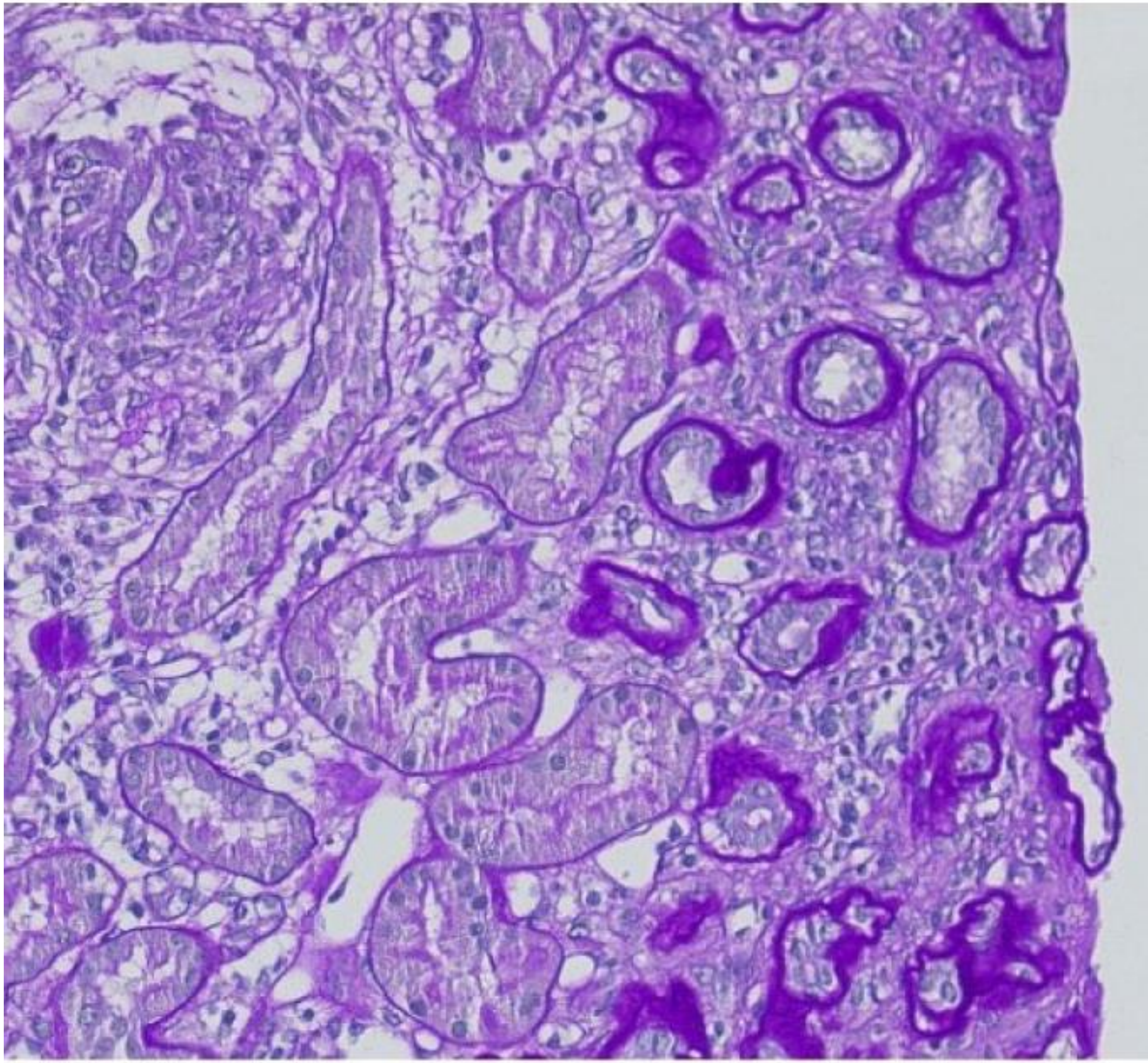
Almost ready for submission…

Kidney's Biopsies

The **area of renal peritubular interstitium** (the inter-tubular, extra-glomerular, and extra-vascular space) **correlates with** the evolution of **Chronic Kidney's Disease** (CKD).

This is considered a **biomarker** to:

- assess the **progression** of tumors and pathologies.
- assess the **effectiveness** of **treatments**.



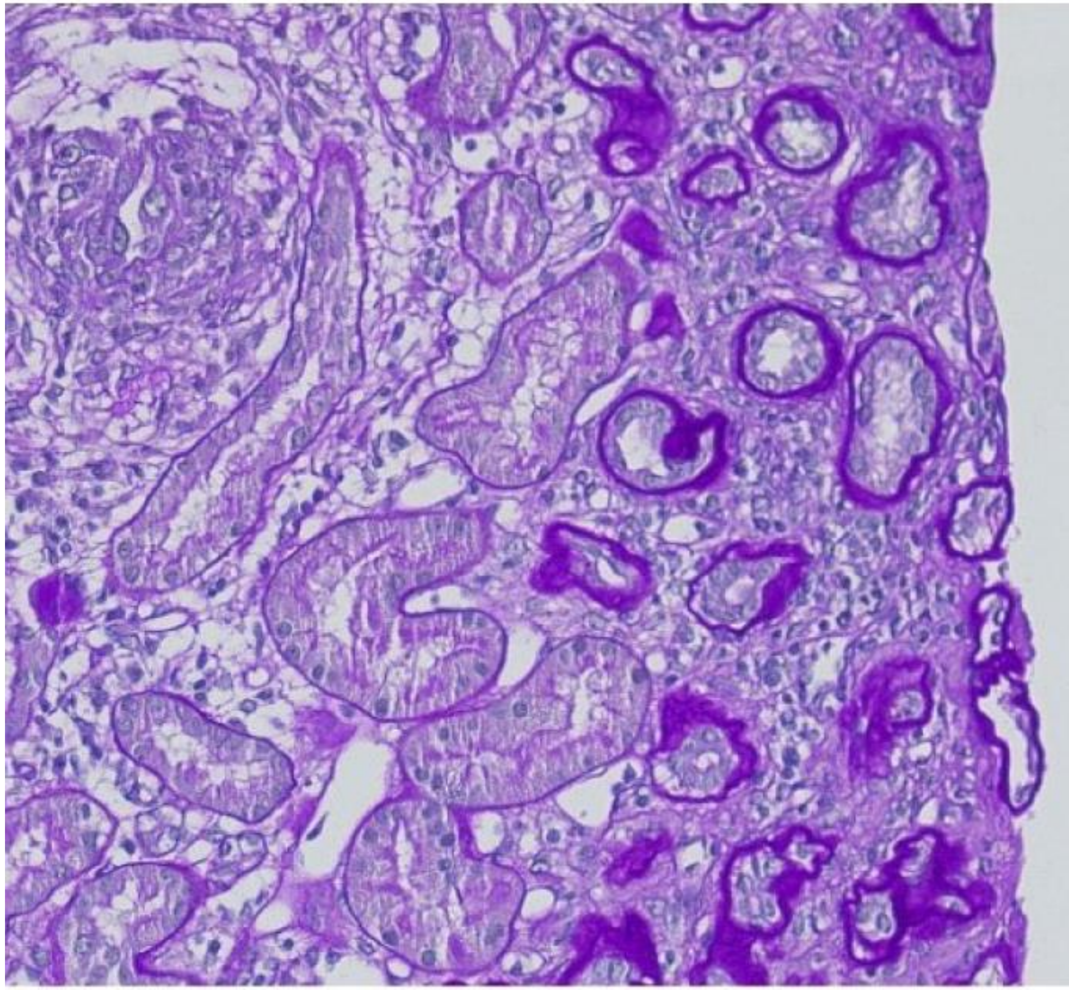
Histological Images and CKD

The area of interstitium can be manually measured / segmented by operators, however this is a task:

- **Labor intensive**
- **Time consuming**
- Very **subjective**

Deep Neural Networks for image segmentation are **very appealing as it is:**

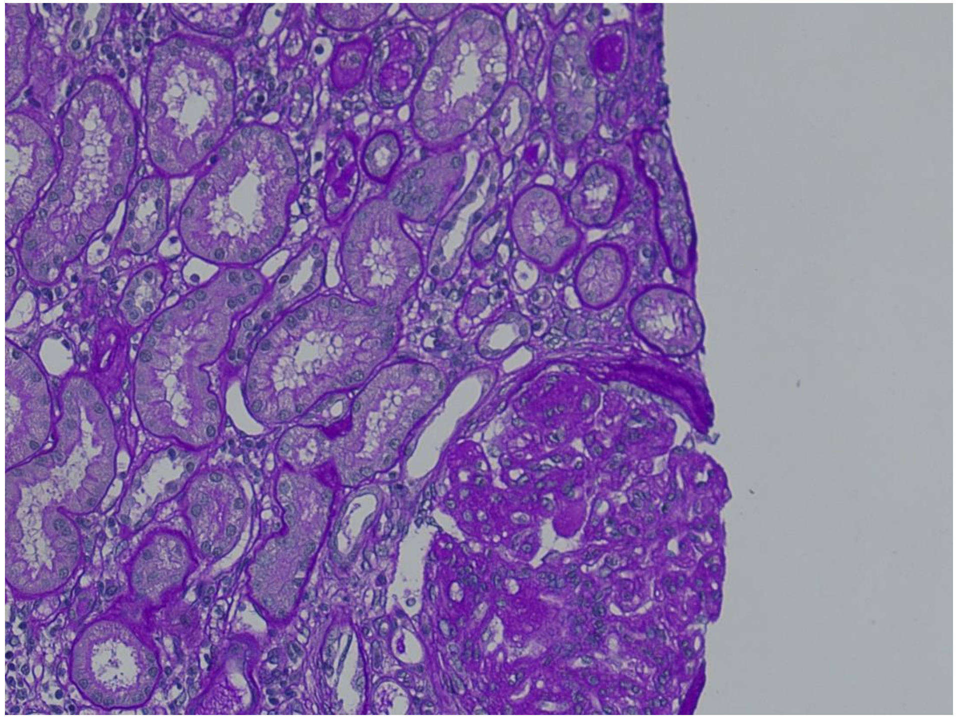
- Automated processing,
- Repeatable results
- Still, requires annotations!



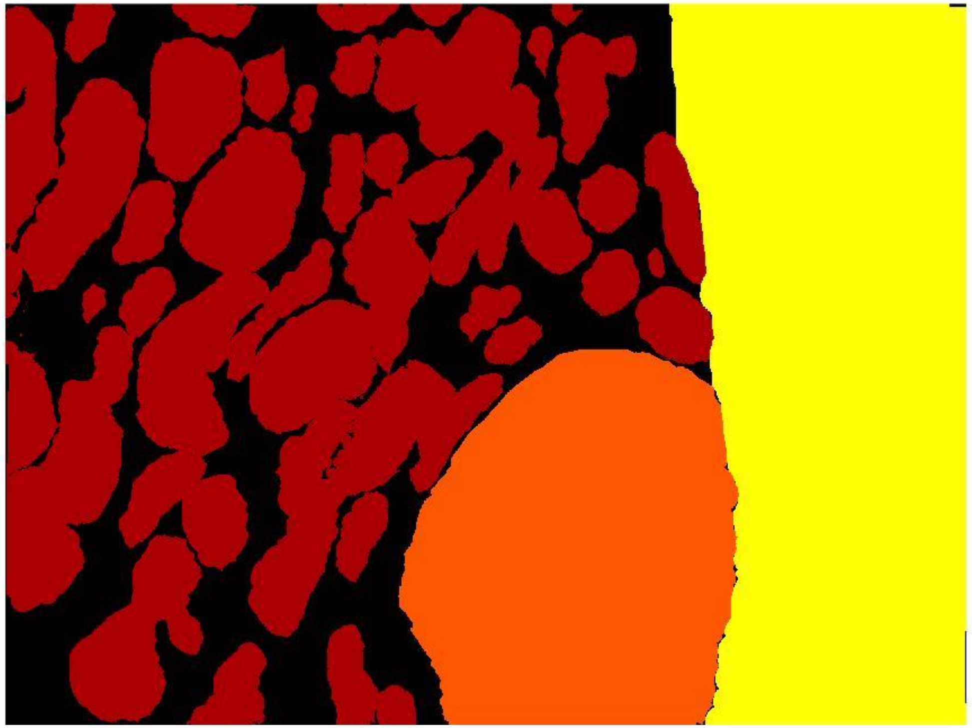
Problem Formulation

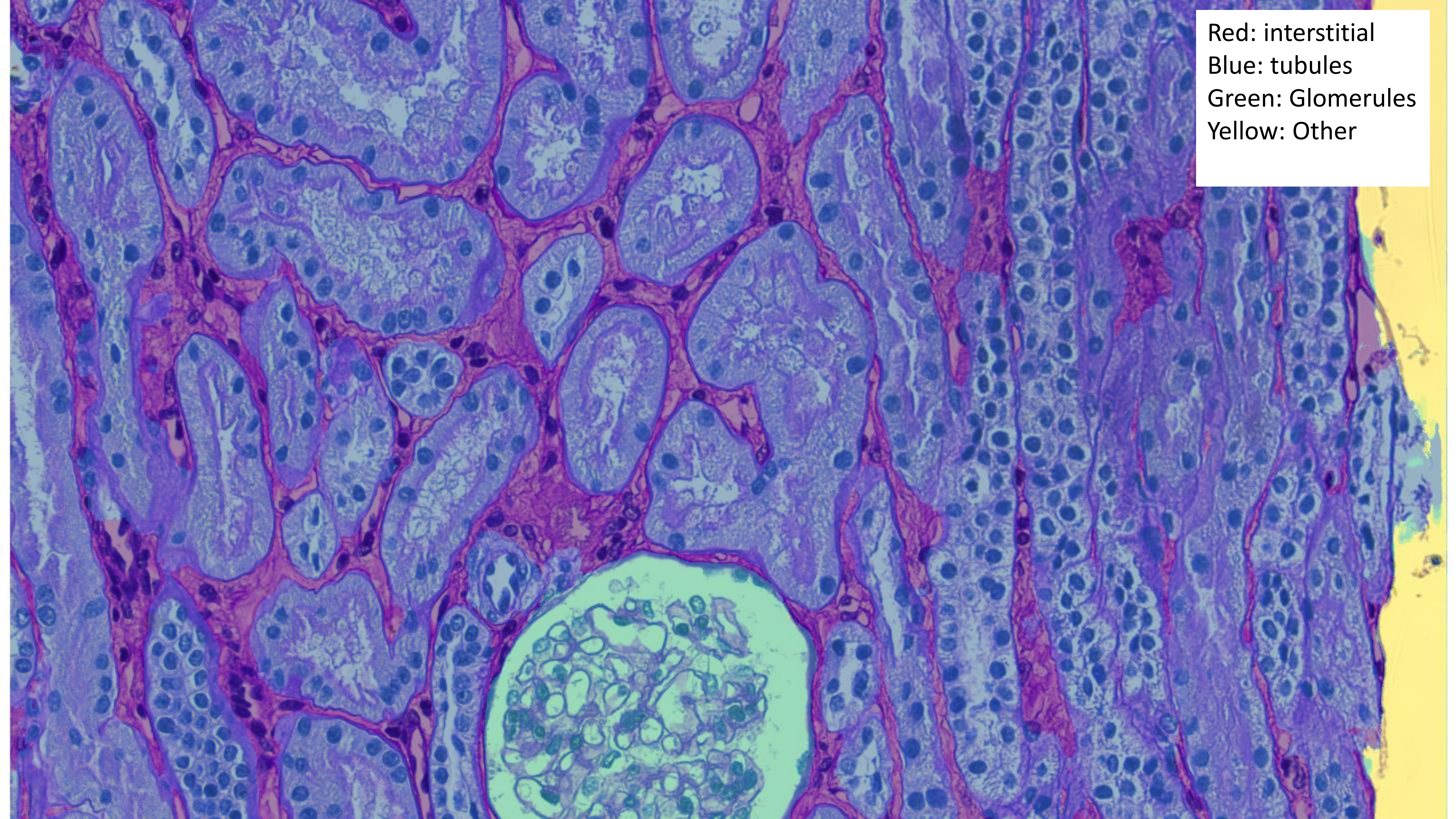
- Semantic segmentatation of kidney biopsises (1920 x 2560 x 3 images)
- 4 classes: *'Interstitial' (black), 'Tubules' (red), 'Glomerolus' (orange), 'Other' (yellow)*

Original image



Ground truth



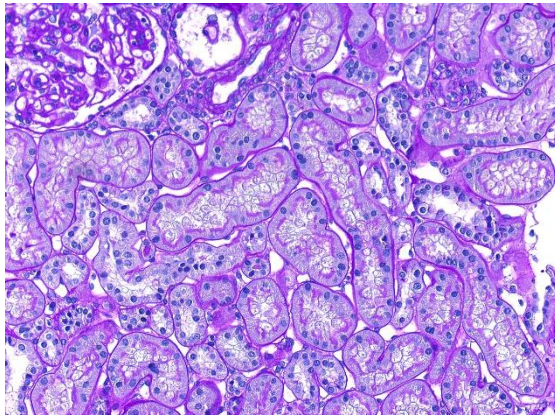

Red: interstitial
Blue: tubules
Green: Glomerules
Yellow: Other

Intrinsic Challenges

- Many souces of **anatamical variability**
  - Different pathologies
  - Different conditions even within the same pathology group

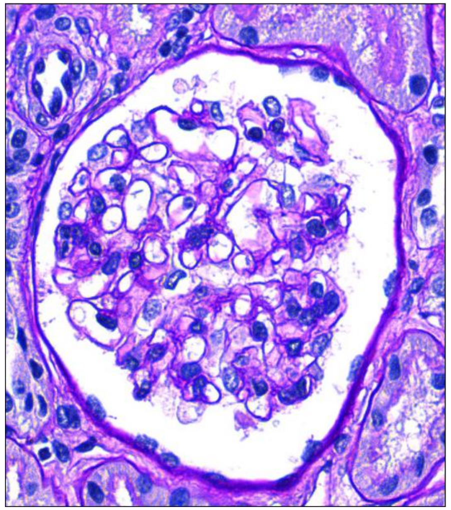
- **Few images** available with the corresponding **dense annotations.**
  - the average **annotation time** was about 1.5 hours per image.
  - rare pathologies involved

Image

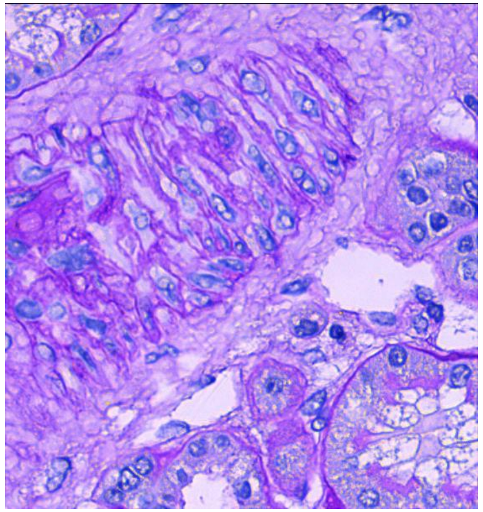


Dense Annotations

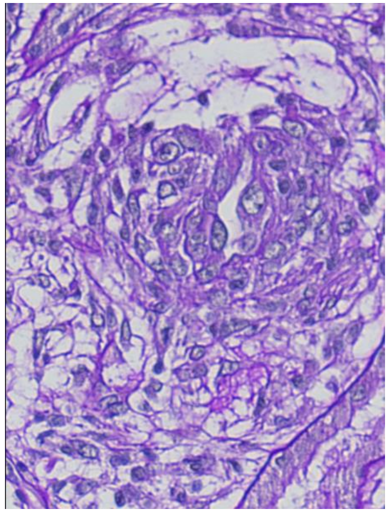
Healthy Subject



Healthy Subject



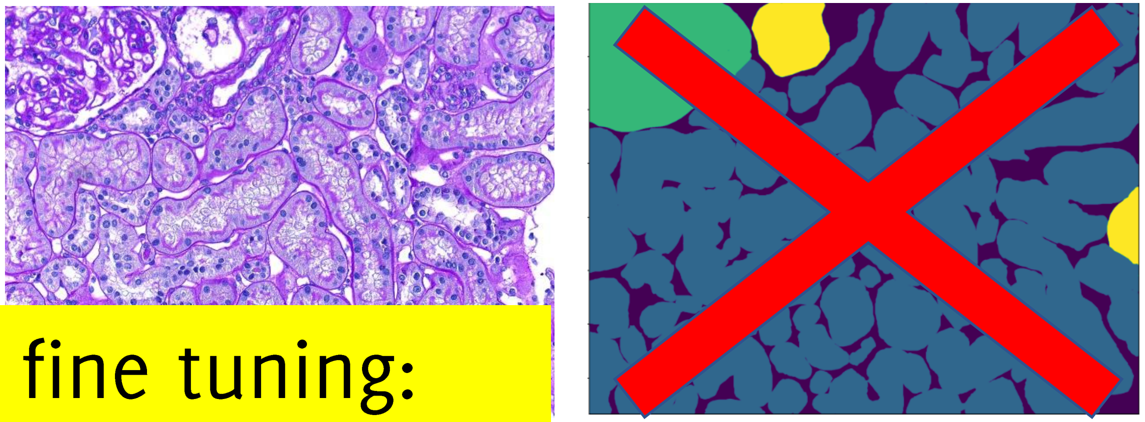
Patient



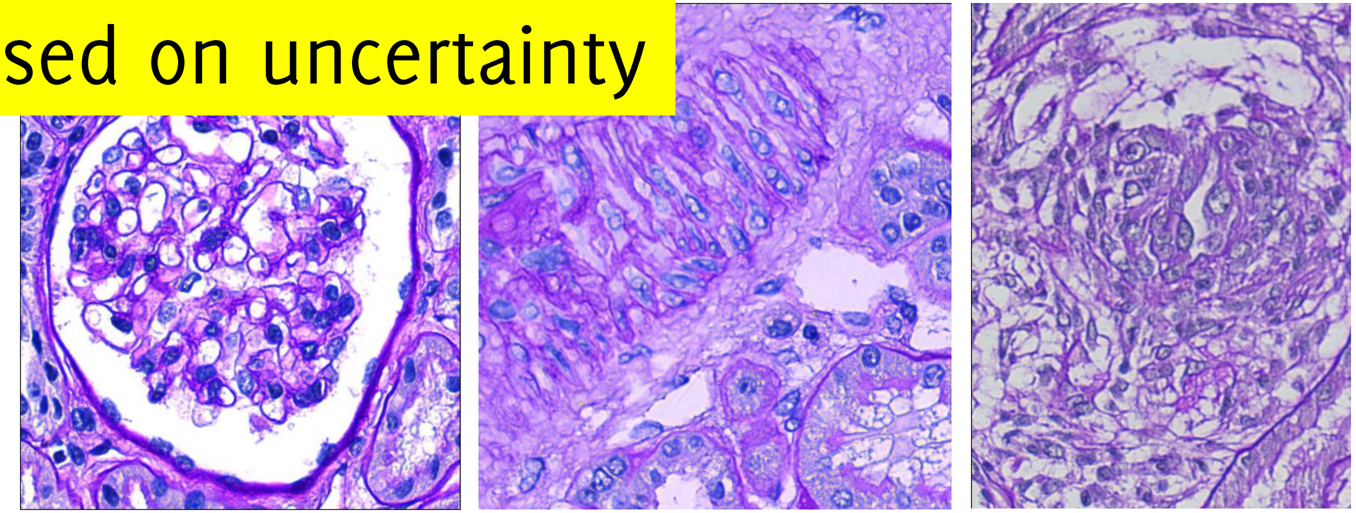
Intrinsic Challenges

- Many souces of **anatamical variability**
  - Different pathologies
  - Different conditions even within the same pathology group

- **Few images** availab corresponding **dens**
  - the average **annota** about 1.5 hours per image.
  - rare pathologies involved



Two solutions for fine tuning:
- sparse annotations
- weights based on uncertainty

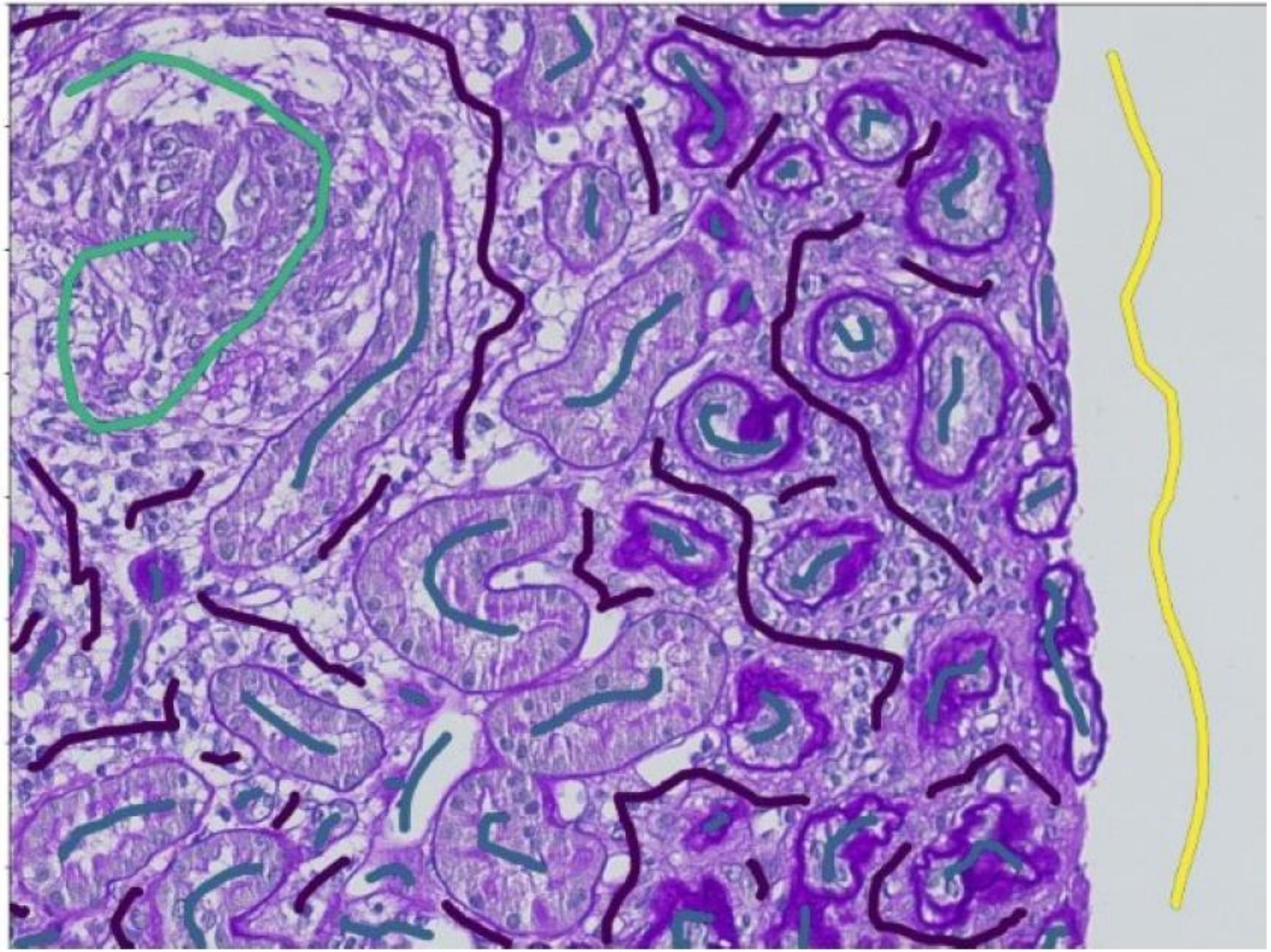


Fine tuning using sparse annotations

Solution Idea: Weak Annotations

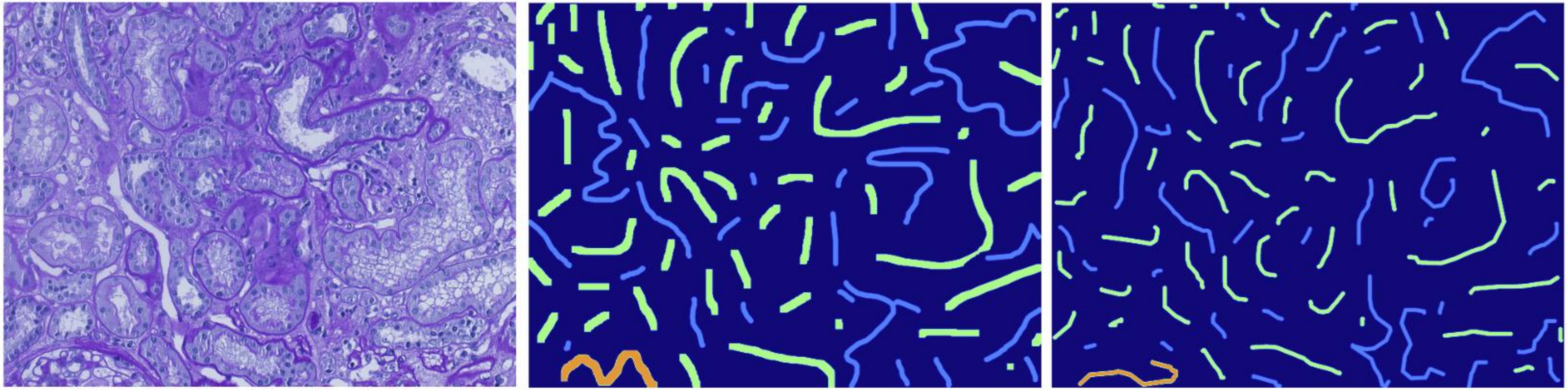
Adopt **sparse and fast-to-obtain annotations** to fine tune a general model on each specific patient / pathological condition

Scribbles can take less than 5 minutes to prepare (vs 1.5hrs of dense annotations)



Wang, G., Li, W., Zuluaga, M.A., Pratt, R., Patel, P.A., Aertsen, M., Doel, T., David, A.L., Deprest, J., Ourselin, S., et al., 2018. Interactive medical image segmentation using deep learning with image-specific fine tuning. IEEE transactions on medical imaging 37, 1562–1573

Two scribble annotators with different styles



(a) (b) (c)

Figure 3: Sample histological image (a) and corresponding annotations from the main (a) and the alternative (b) annotators.

Scribble Fine Tuning Pipeline

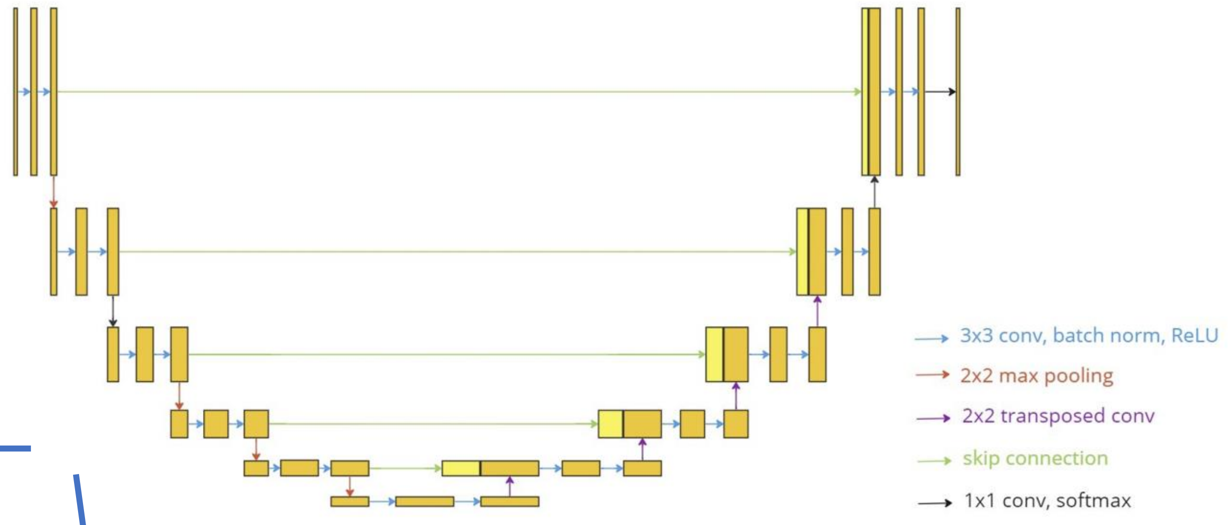
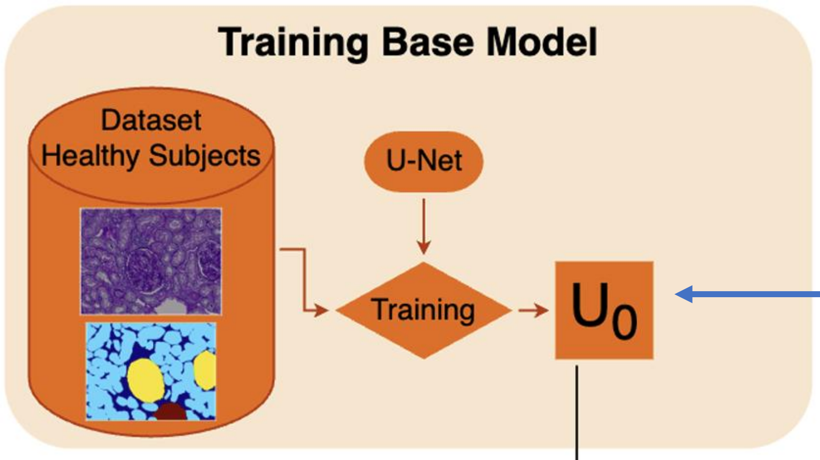


Figure 5: Adopted U-Net architecture.



Dense annotations

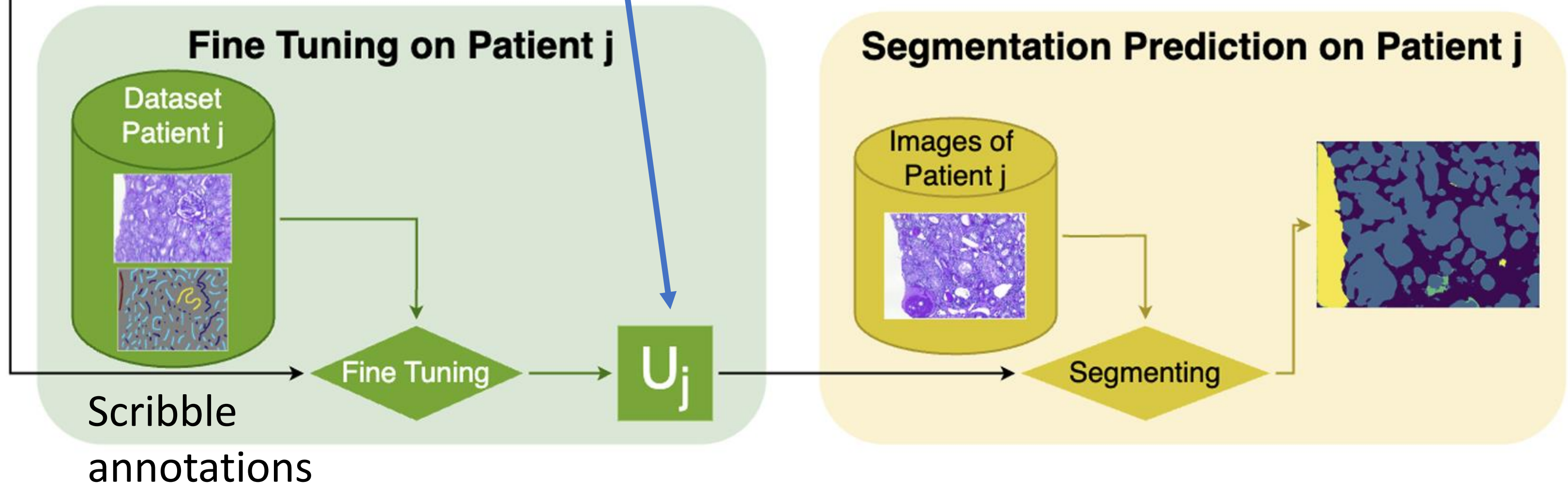


Figure 1: Proposed pipeline for training the initial network U_0, fine-tuning the patient-specific network U_j, and predicting the segmentation of kidney structures.

Patients and Controls

- 11 Controls (healthy individuals) with dense annotations, overall 32 dense annotated images.
- 20 Patients affected by different conditions. Each has scribble annotations on 7 images, dense annotations for testing only on 3 images.

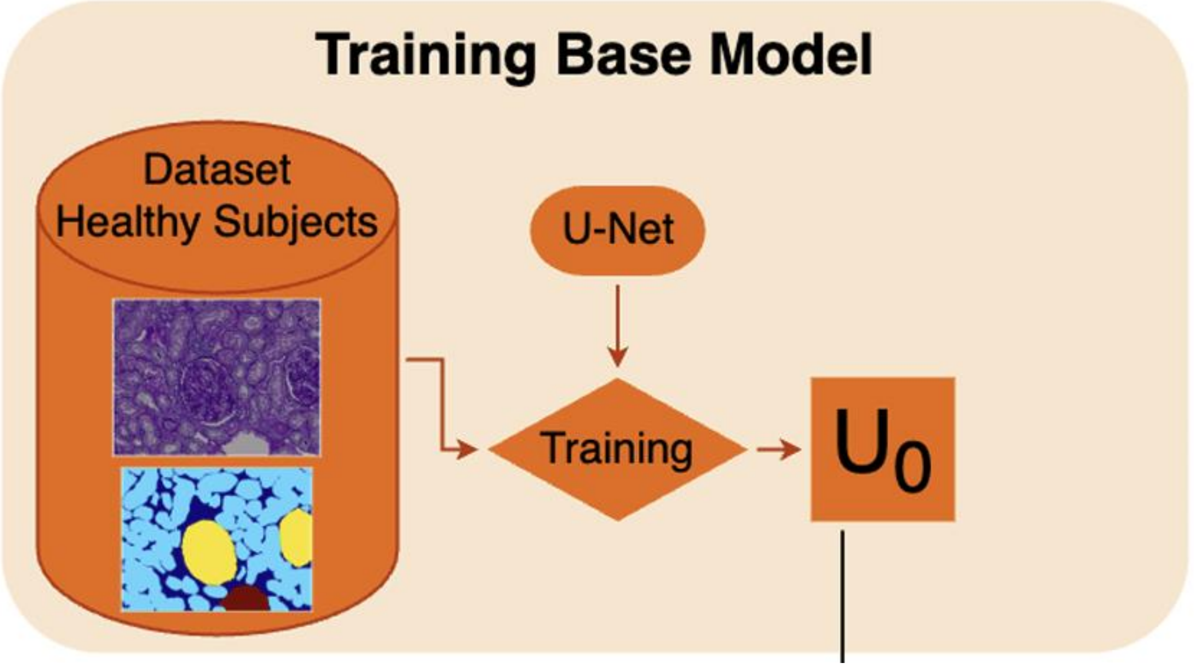
Pathological condition	**Patient IDs**	**Number of patients**
Membrano-proliferative glomerulonephritis Membrano-proliferative glomerulonephritis (early stage)	p01, p02, p03, p04, p14 p19	6 (30%)
ANCA-associated glomerulonephritis	p05, p06, p08, p09	4 (20%)
Tubulo-interstitial nephritis	p16	1 (5%)
Minimal change disease	p13, p18	2 (10%)
IgA nephropathy (mild) IgA Nephropathy (mild – moderate) IgA Nephropathy (moderate) IgA Nephropathy (moderate – severe)	p07 p12 p10, p11 p15	5 (25%)
Minimal non-specific abnormalities	p17, p20	2 (10%)
Total		20 (100%)

Table 1
Distribution of pathological conditions in the dataset.

U_0 training

The network U_0 is trained on 11 healthy individuals with dense annotations, using overall 32 images

- Class weights to counteract class imbalance
- Data augmentation (rotation, zoom, brightness, horizontal, vertical flip)
- Patch-wise training (the network is fully convolutional, tested on arbitrary image sizes)
- Categorical cross-entropy loss (Dice ++)



U_j fine tuning

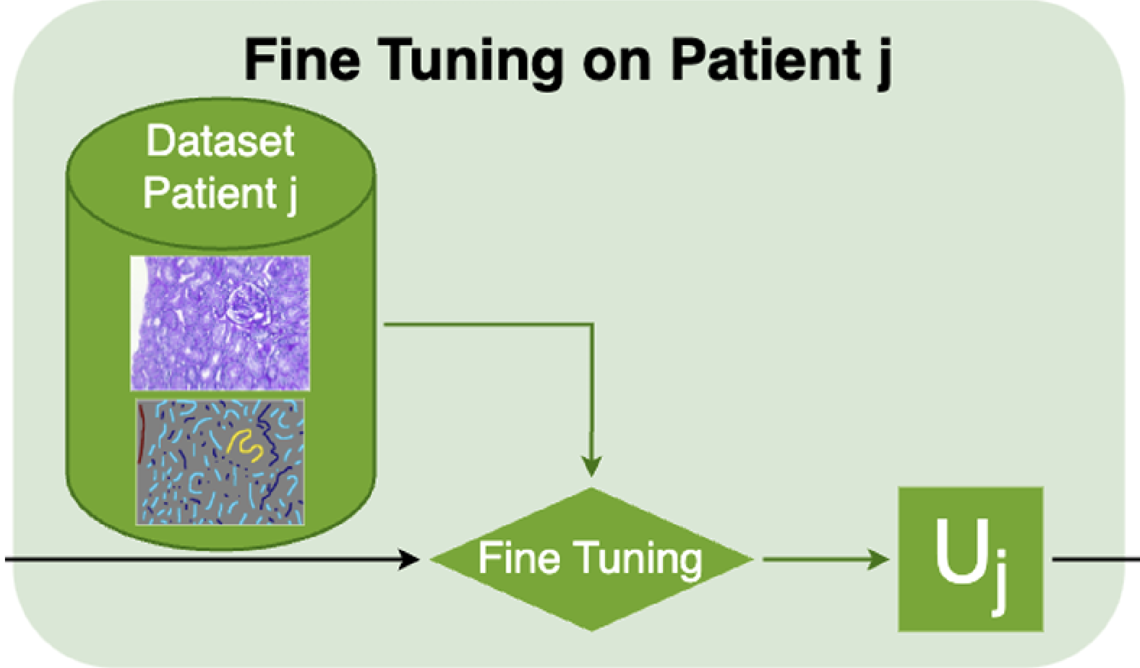
On each patient we fine tune U_0 on 7 scribble annotated images

- Class weights:

$$w(i,j) = \begin{cases} \min\left(\frac{|S|}{C(i,j)}, K\right), (i,j) \in S \\ 0, \qquad\qquad\qquad\qquad (i,j) \notin S \end{cases}$$

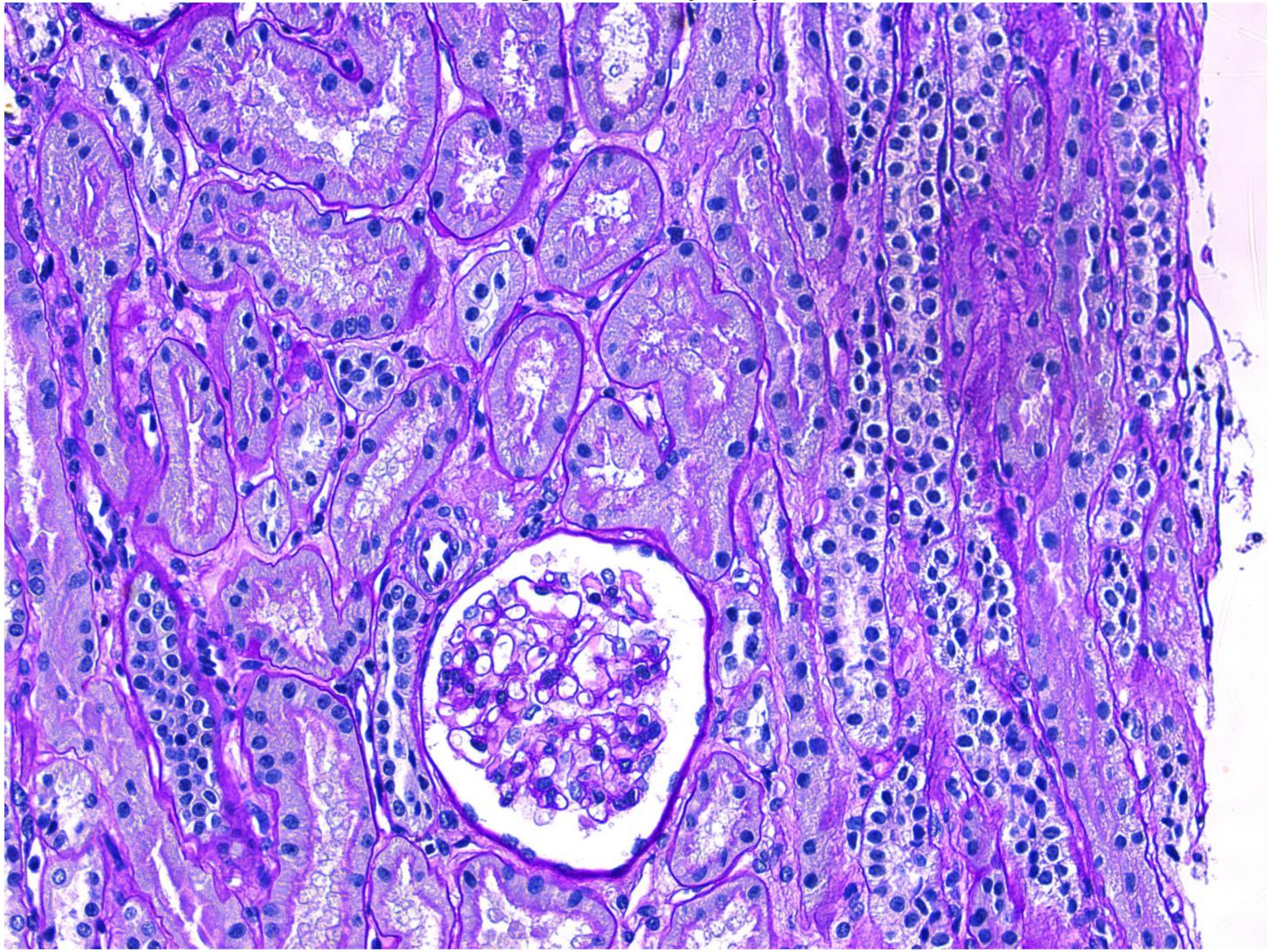
Where S are the scribble annotations, $C(i,j)$ is the class support, K a maximum weight

- Pixels not included in any scribble are completely ignored



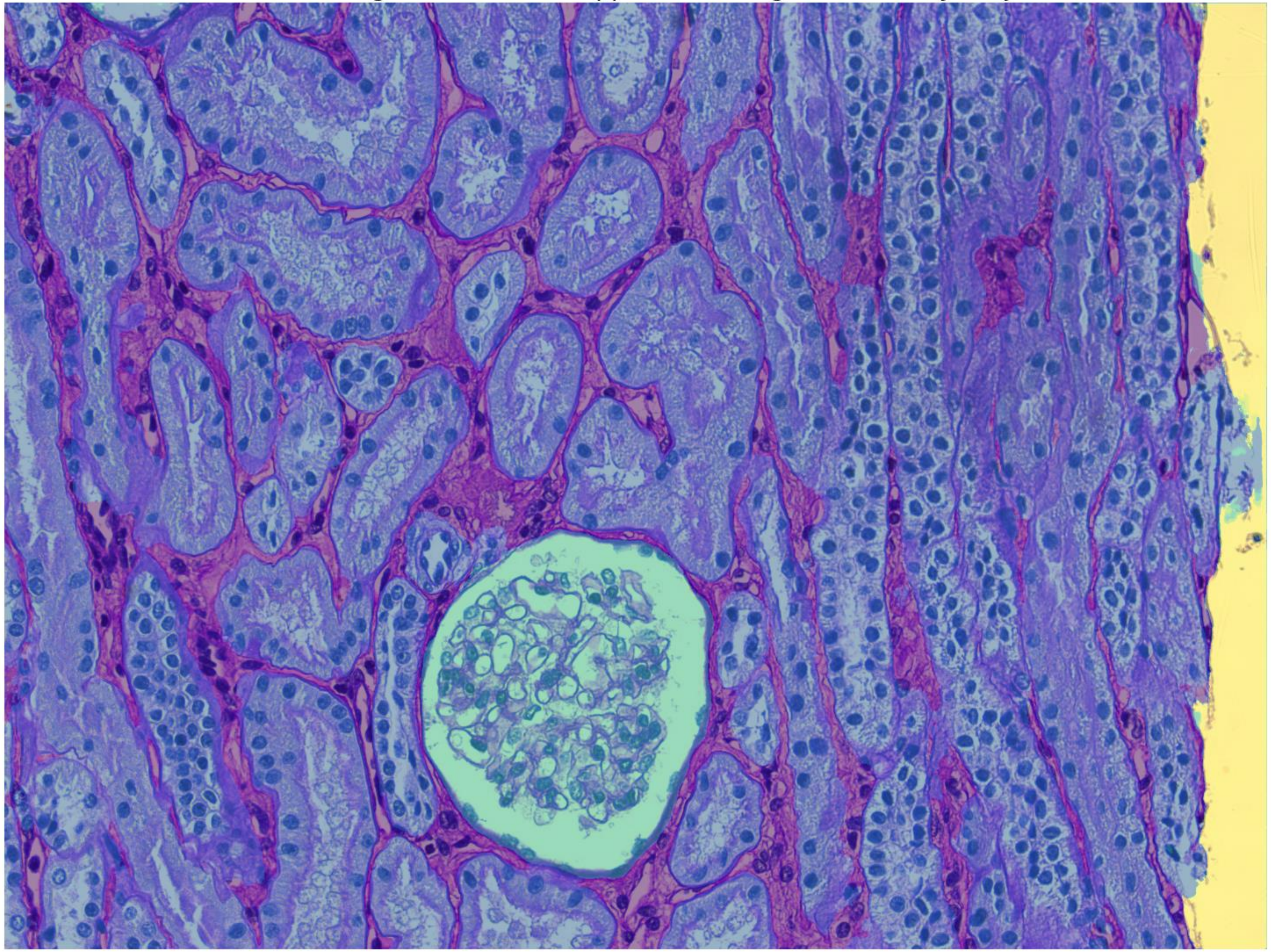
Test Image of a Healty Subject

Image of a healthy subject



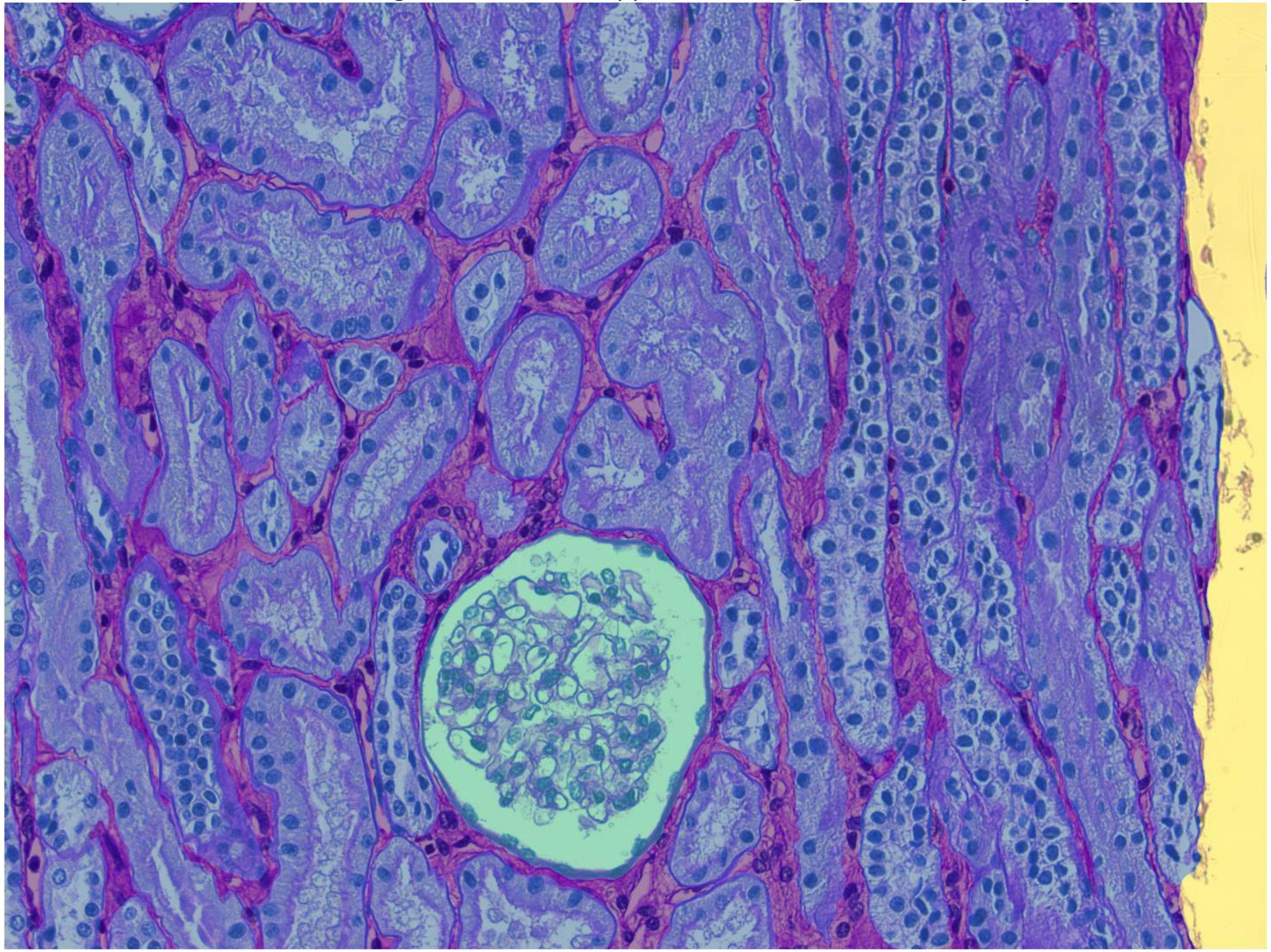
Predicted Segmentation by U_0 of a Healthy Subject

Predicted segmentation overlapped with image of a healthy subject



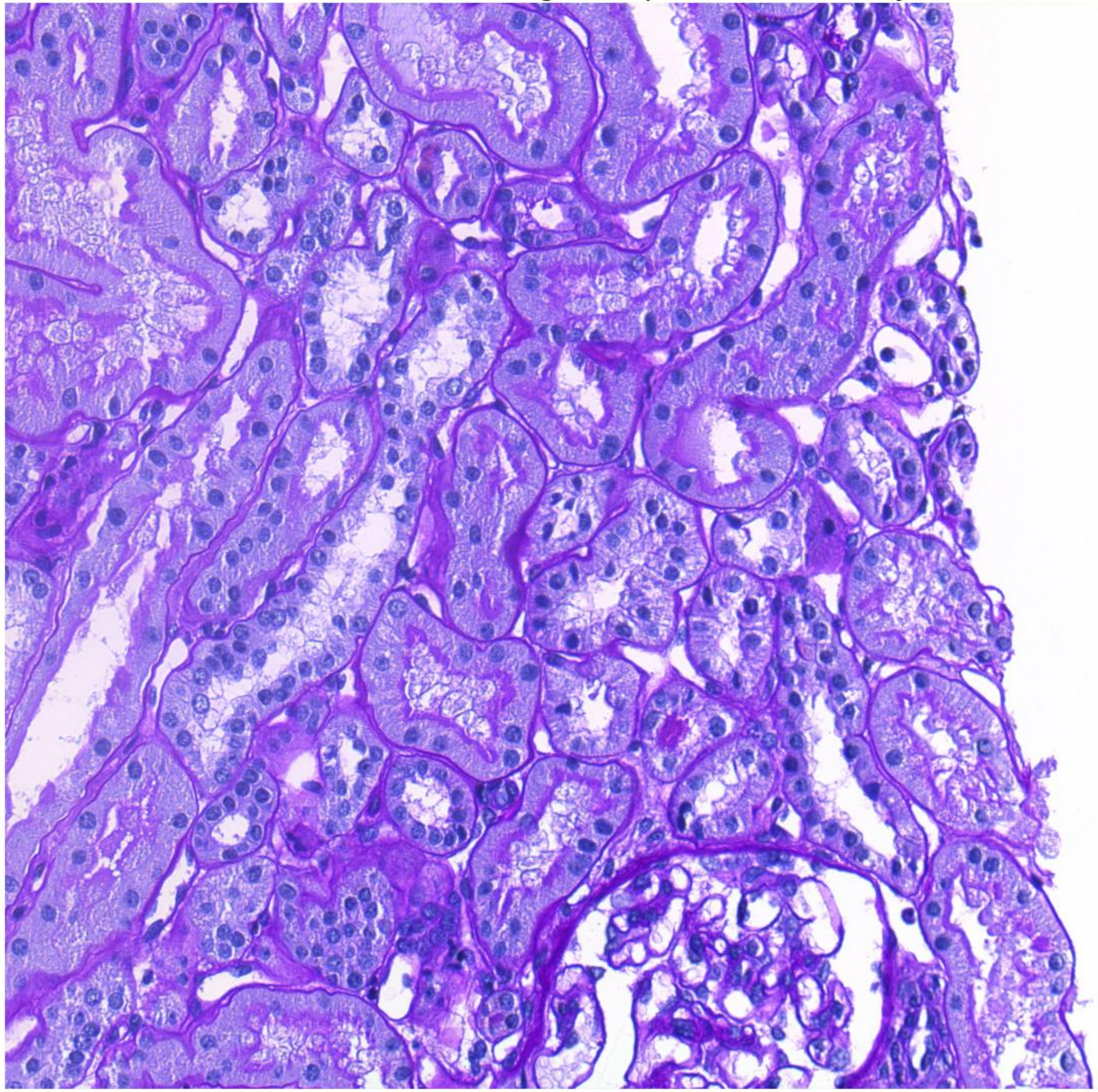
GT Segmentation of a Healthy Subject

Ground truth segmentation overlapped with image of a healthy subject



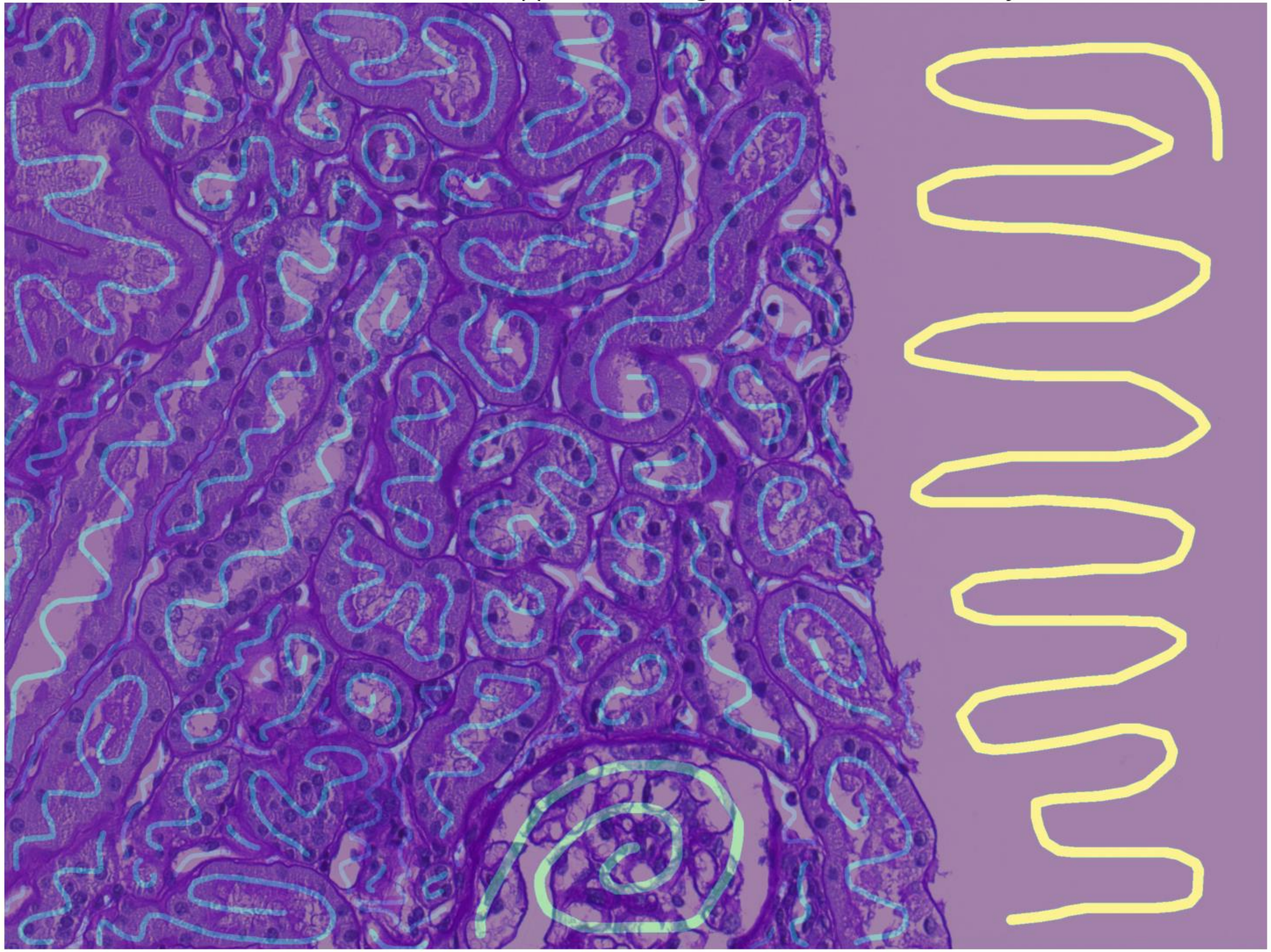
Test image of a Patient

Image of a patient affected by PAS



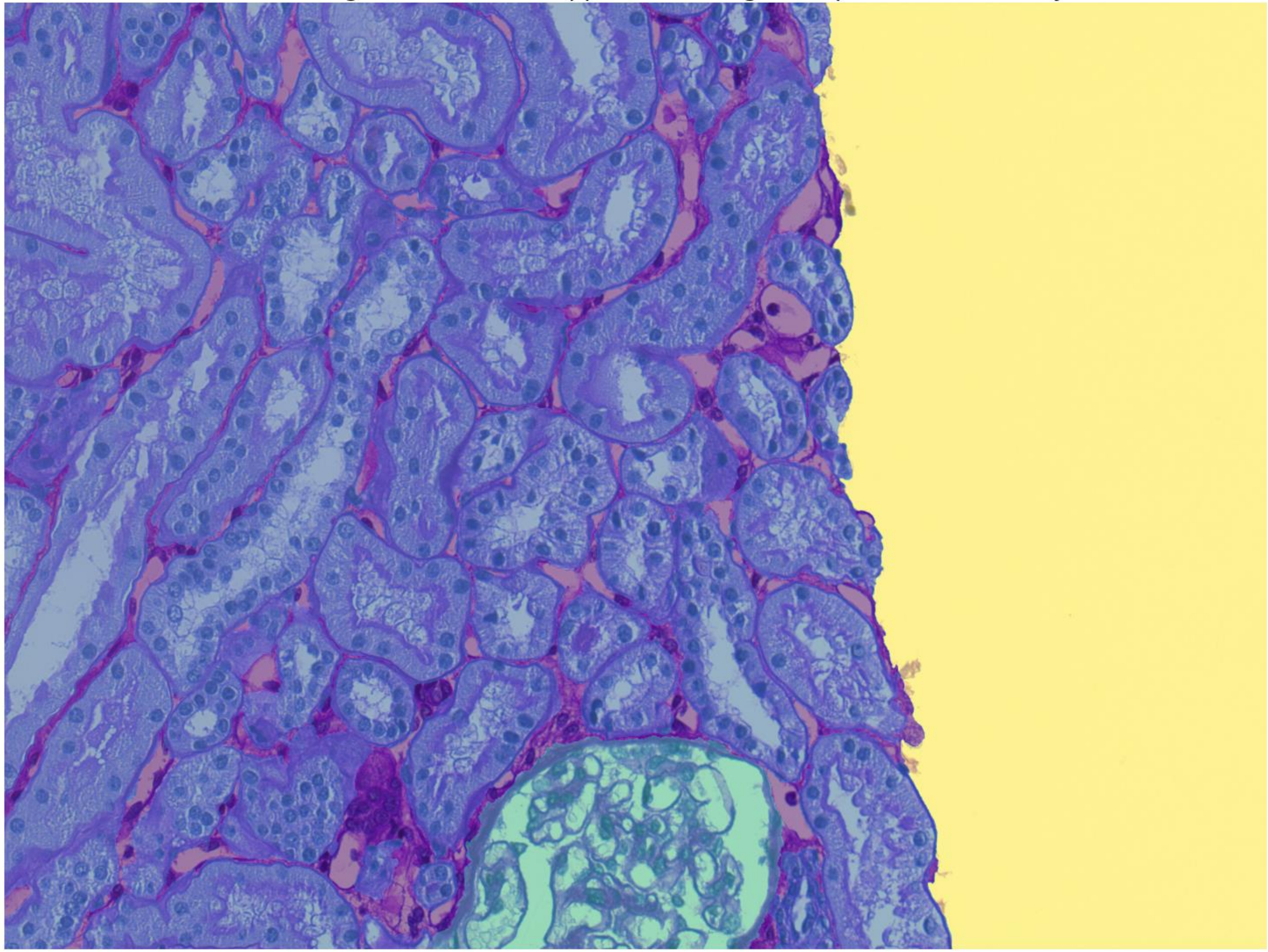
Scribble Annotations for fine Tuning U_0

Scribble annotation overlapped with image of a patient affected by PAS



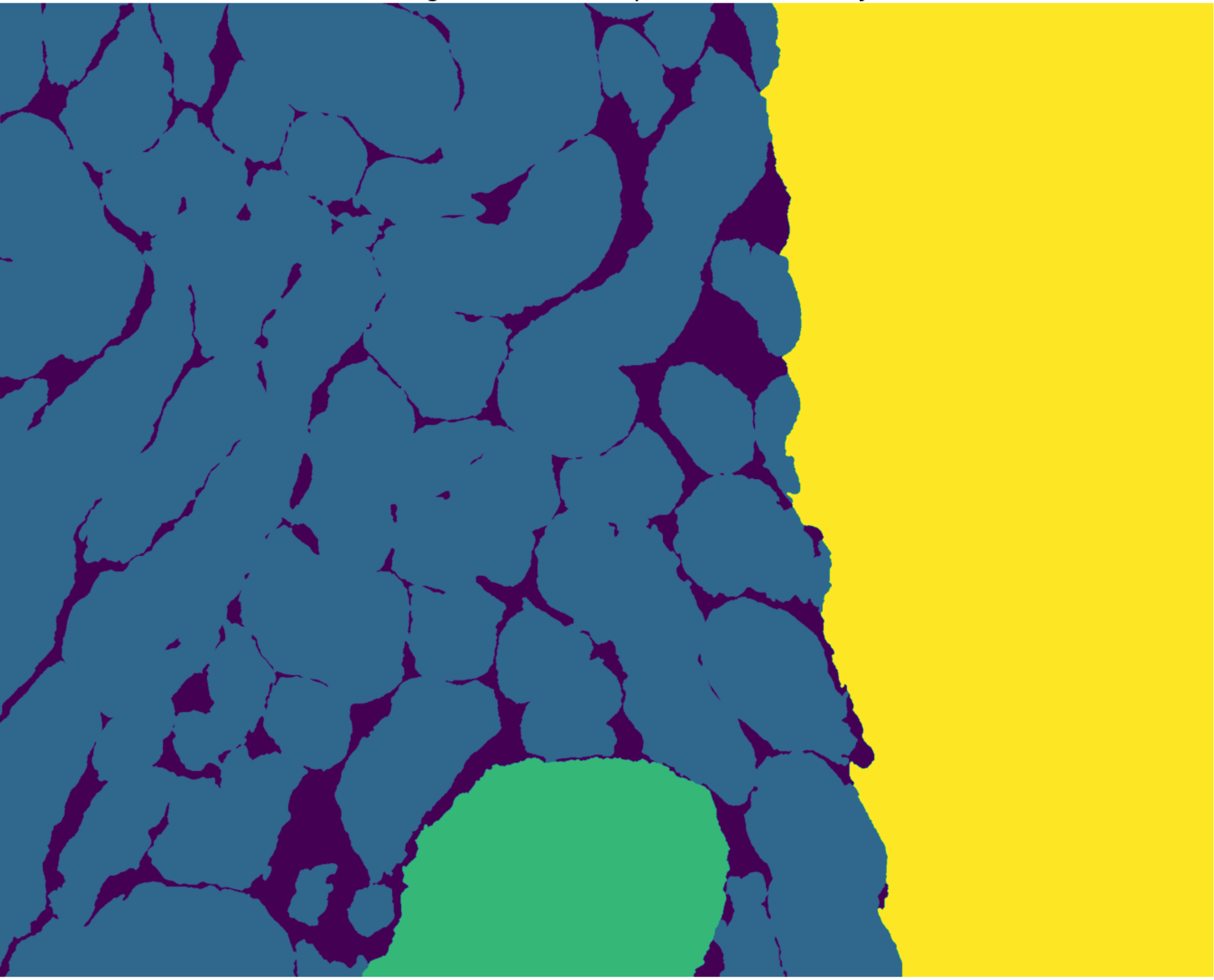
GT Segmentation of a Patient

Ground truth segmentation overlapped with image of a patient affected by PAS

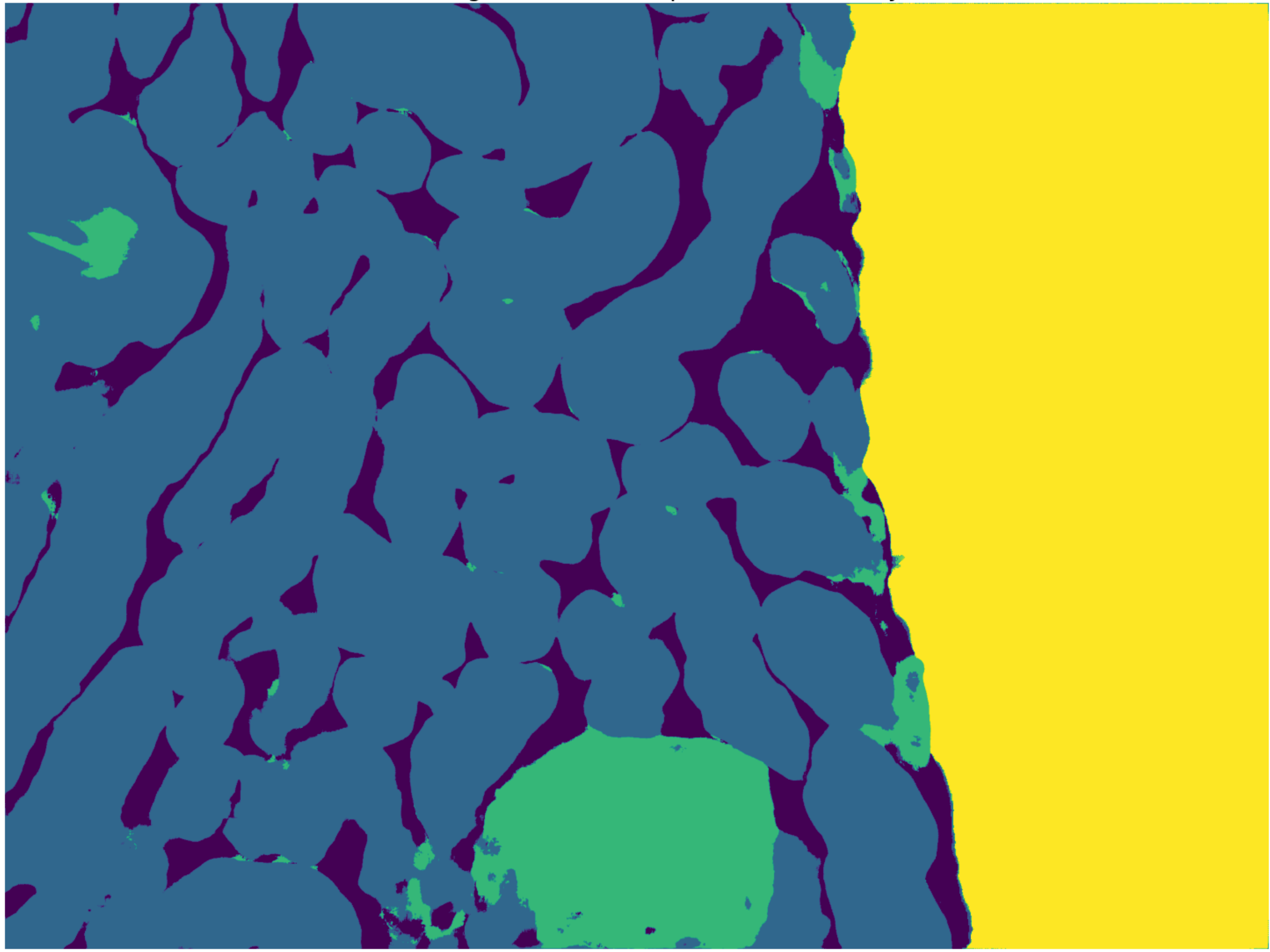


Predictions vs GT for a Patient

Ground truth segmentation of a patient affected by PAS

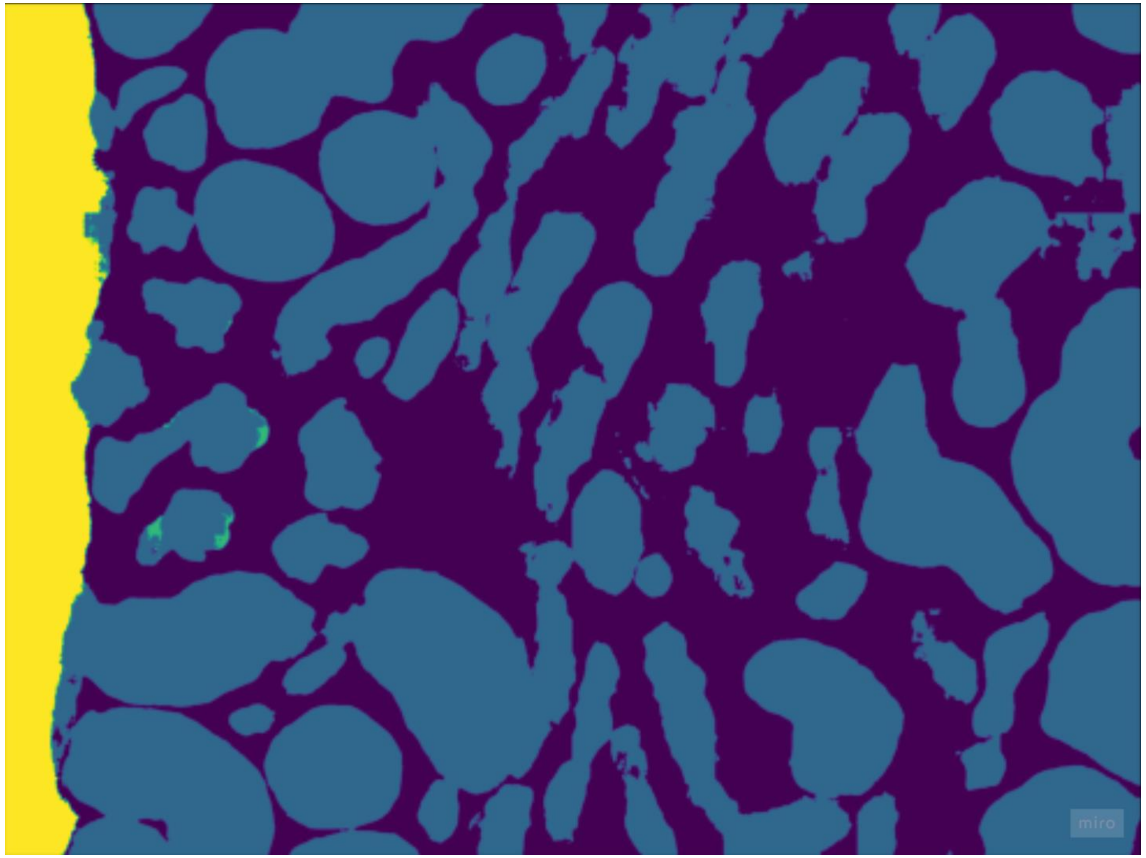


Predicted segmentation of a patient affected by PAS

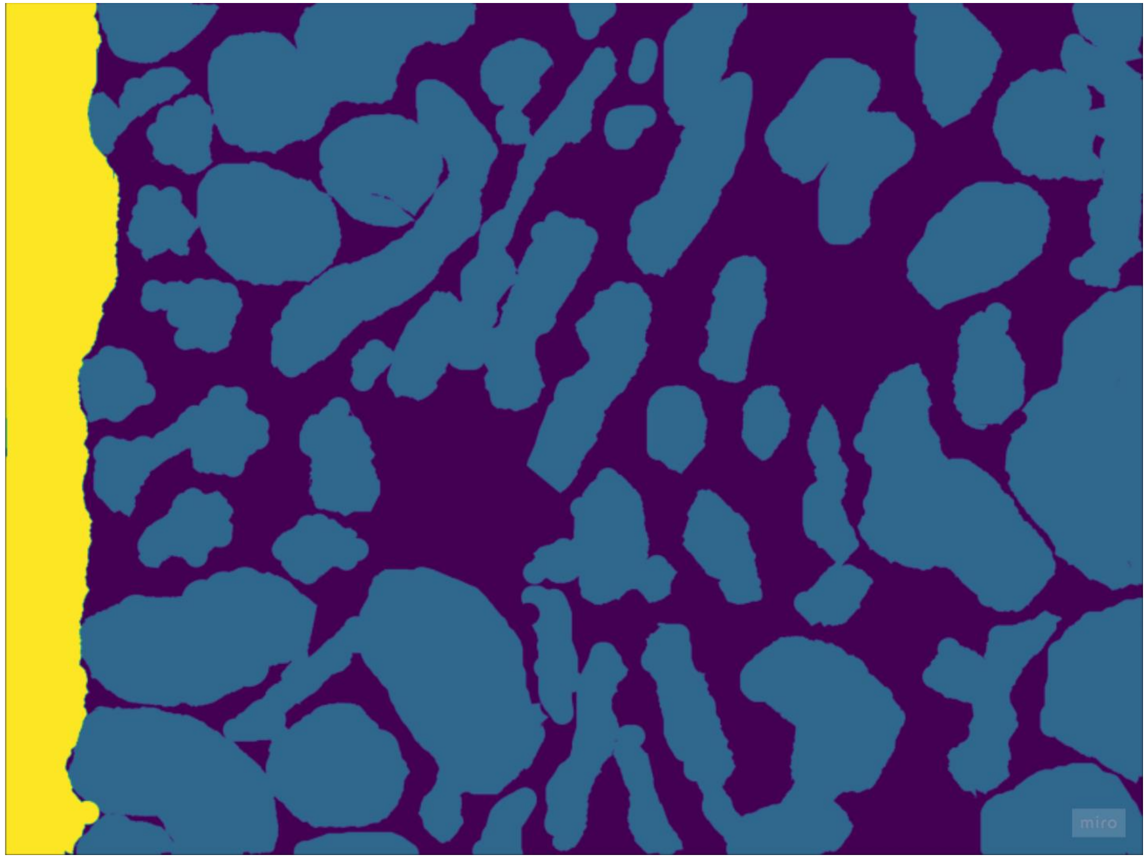


Qualitative Performance

Ground Truth



Predictions from Tuning using sparse annotations



Quantitative Assessment, avg. class. error on all the classes

Fine tuning on scribble annotations is in general beneficial (Wilcoxon test p-value 0.035)

- in 14 out of 20 patients it is beneficial (in particular p05)
- When it is not beneficial, losses are minima.

Too sparse annotations (50% of scribbles) are detrimental.

Scribbles from the second annotator are less informative (fine tuning detrimental)

Patient	Initial network U_0	Fine-tuned U_j 100% scribbles main annotator	50% scribbles main annotator	100% scribbles secondary annotator
p01	0.900	0.912	0.876	0.908
p02	0.837	0.868	0.826	0.875
p03	0.671	0.646	0.665	0.586
p04	0.796	0.806	0.666	0.805
p05	0.515	0.797	0.733	0.636
p06	0.674	0.785	0.798	0.783
p07	0.659	0.687	0.655	0.662
p08	0.884	0.868	0.754	0.560
p09	0.603	0.627	0.643	0.628
p10	0.818	0.830	0.820	0.808
p11	0.562	0.537	0.649	0.435
p12	0.812	0.838	0.597	0.578
p13	0.922	0.919	0.860	0.916
p14	0.680	0.709	0.647	0.666
p15	0.594	0.701	0.727	0.610
p16	0.741	0.746	0.818	0.593
p17	0.866	0.853	0.837	0.838
p18	0.705	0.728	0.683	0.724
p19	0.808	0.759	0.564	0.695
p20	0.865	0.872	0.609	0.819
Median [interquartile range]	0.769 [0.668; 0.844]	0.791 [0.707; 0.857]	0.705 [0.649; 0.819]	0.681 [0.606; 0.811]

Table 2

Mean dice scores for all classes obtained by the deterministic network U_0, the networks U_j fine-tuned using 100% of the main annotator's scribbles, the networks U_j fine-tuned using 50% of the main annotator's scribbles, and the networks fine-tuned U_j using 100% of the secondary annotator's scribbles.

Quantitative Assessment on Interstitial Class

Fine tuning on scribble annotations is in general beneficial (Wilcoxon test p-value 0.005)

Patient	Initial network U_0	Fine-tuned U_j 100% scribbles main annotator	50% scribbles main annotator	100% scribbles secondary annotator
p01	0.702	0.818	0.760	0.809
p02	0.640	0.722	0.755	0.744
p03	0.800	0.789	0.838	0.686
p04	0.802	0.828	0.806	0.833
p05	0.870	0.910	0.871	0.907
p06	0.852	0.843	0.870	0.817
p07	0.748	0.841	0.748	0.782
p08	0.697	0.797	0.718	0.473
p09	0.847	0.783	0.777	0.788
p10	0.876	0.809	0.866	0.858
p11	0.783	0.734	0.735	0.726
p12	0.797	0.848	0.865	0.644
p13	0.757	0.820	0.772	0.820
p14	0.759	0.829	0.853	0.784
p15	0.799	0.832	0.746	0.674
p16	0.653	0.723	0.798	0.618
p17	0.615	0.708	0.719	0.697
p18	0.839	0.858	0.853	0.867
p19	0.793	0.802	0.630	0.732
p20	0.742	0.862	0.717	0.802
Median [interquartile range]	0.788 [0.732; 0.811]	0.819 [0.788; 0.842]	0.775 [0.743; 0.853]	0.783 [0.694; 0.818]

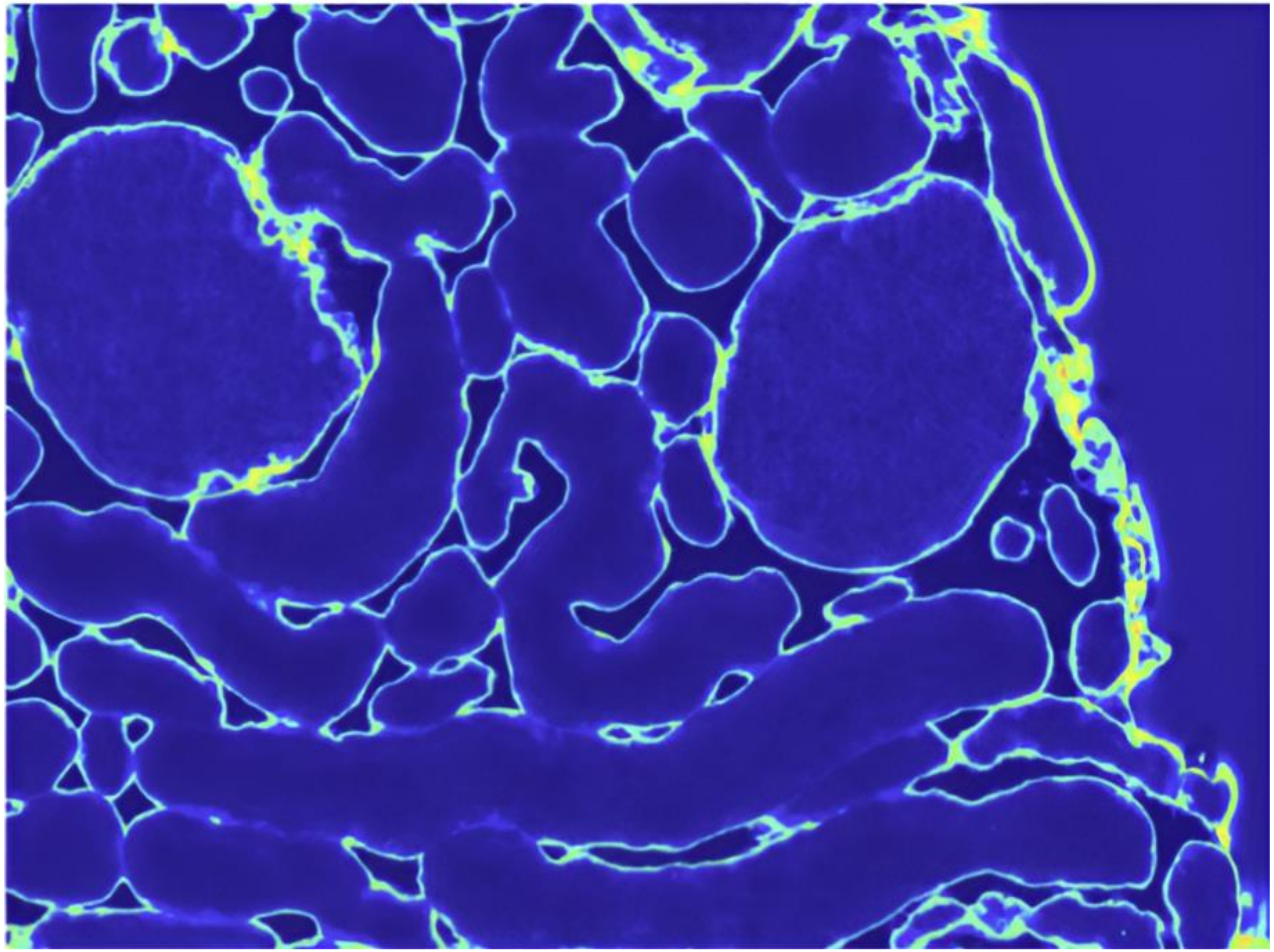
Table 3
Dice scores for the class c_{RPI} obtained using the deterministic network U_0, the networks U_j fine-tuned using 100% of the main annotator's scribbles, the networks U_j fine-tuned using 50% of the main annotator's scribbles, and the networks U_j fine-tuned using 100% of the secondary annotator's scribbles.

Fine tuning using uncertainty-based weights

Solution Idea: Uncertainty Estimation

Along with class posterior, networks can be modified to provide estimates of uncertainty in their predictions.

We integrate uncertainty estimates from MCD in the fine-tuning on scribble annotations



Wang, G., Li, W., Zuluaga, M.A., Pratt, R., Patel, P.A., Aertsen, M., Doel, T., David, A.L., Deprest, J., Ourselin, S., et al., 2018. Interactive medical image segmentation using deep learning with image-specific fine tuning. IEEE transactions on medical imaging 37, 1562–1573

Dropout

Dropout randomly set to zero connections in the network.

- **Active during training** to reduce the risk of overfitting: the network is seen as training an ensemble of models.
- **Disabled during inference**: the network returns an aggregation of individual estimates.

We add two dropout layers to the U-net architecture U_0^D and follow the same training on dense annotations and fine-tuning on scribbles as for U_0.

Scribble + MC Dropout Fine Tuning Pipeline

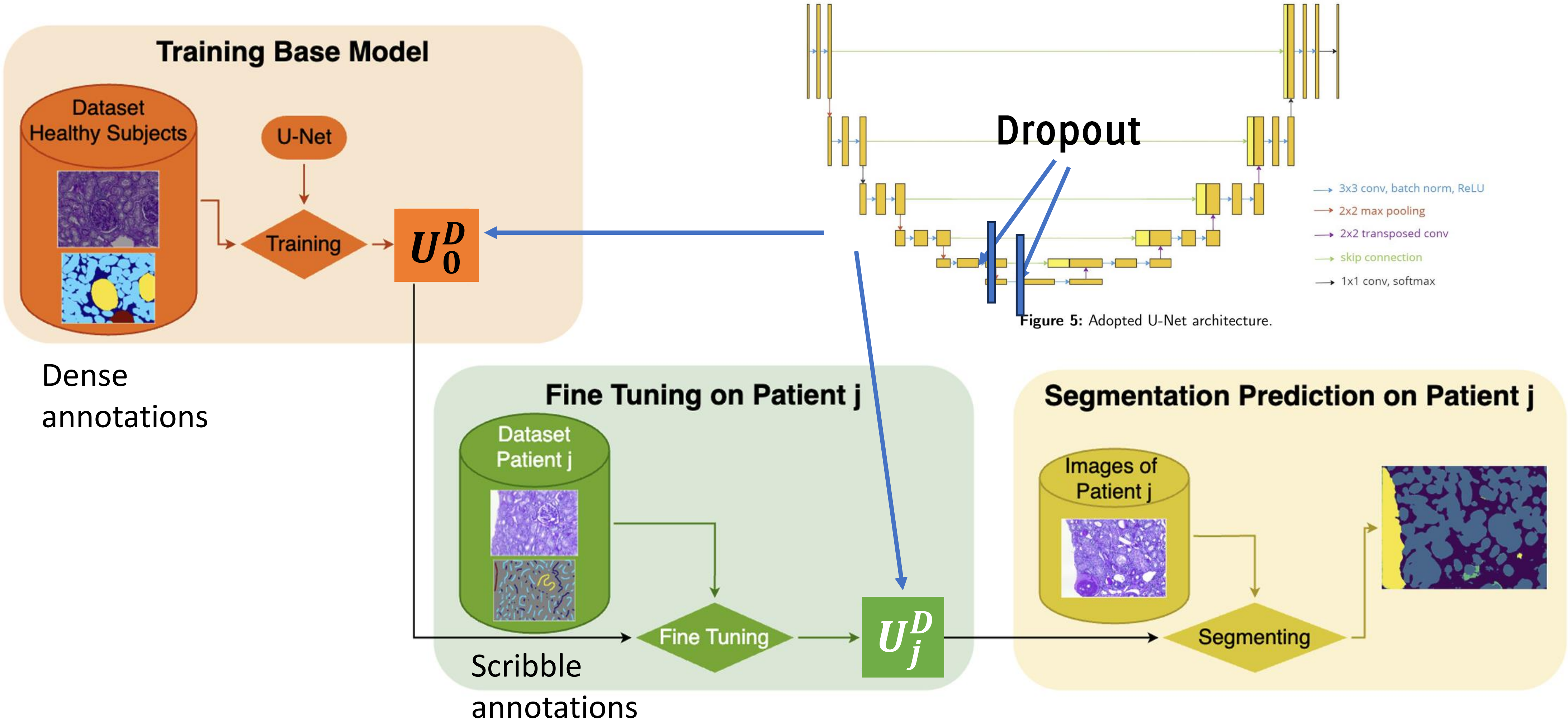


Figure 1: Proposed pipeline for training the initial network U_0, fine-tuning the patient-specific network U_j, and predicting the segmentation of kidney structures.

Uncertainty Estimation in Bayesian Networks

Bayesian networks keep **dropout** active during **inference**

During the t-th call of the network U_0^D on the same test image we obtain

$$U_0^D(I) \mapsto p_{c,t}, c = 1, . . , 4, t = 1, \ldots, T$$

- The network U_0^D **returns different outputs $\boldsymbol{p_t}$** when fed with the same input I.
- The network U_0^D output is given by the average posterior $\bar{p}_{c,t}(i, j)$ over all the network calls.

Uncertainty Estimation in Bayesian Networks

We can compute the network uncertainty for each class c, as

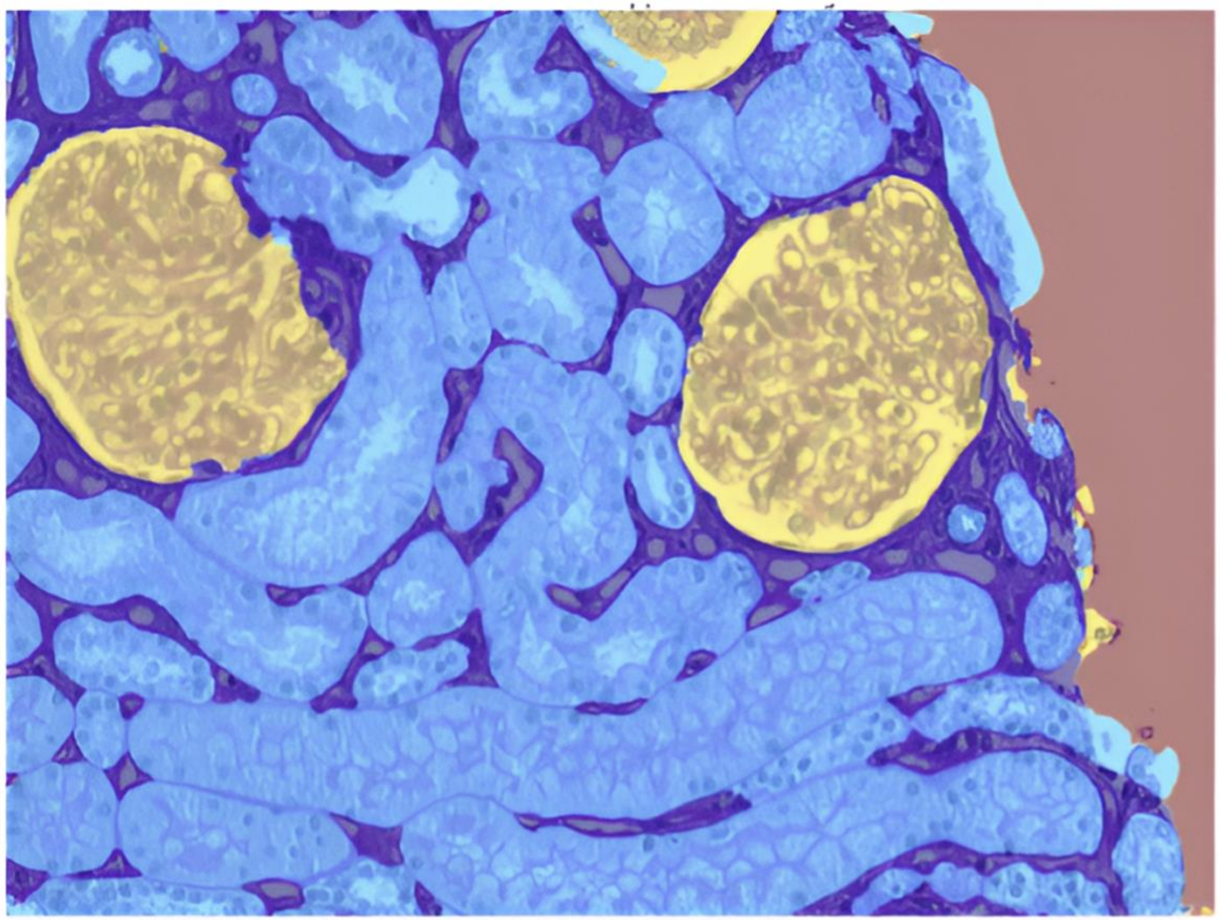
$$\mathcal{V}_c(i,j) = \mathcal{A}_c(i,j) + \mathcal{E}_c(i,j)$$

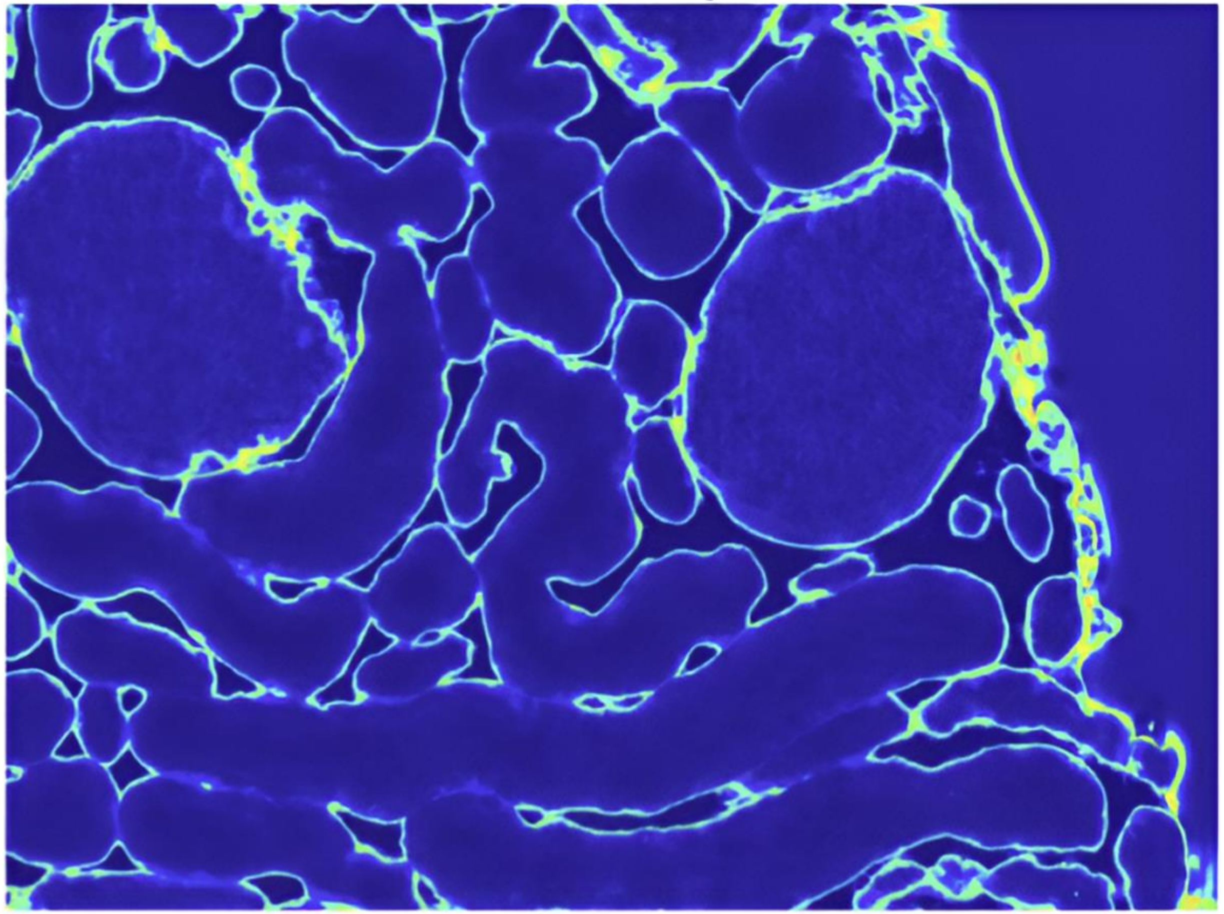
$$\mathcal{V}_c(i,j) = \frac{1}{T}\sum_t p_{c,t}(i,j)(1 - p_{c,t}(i,j)) + \frac{1}{T}\sum_t \left(p_{c,t}(i,j) - \bar{p}_{c,t}(i,j)\right)^2$$

- $\mathcal{A}_c$ aleatoric uncertainty: due to noise in the inference data, unavoidable.
- $\mathcal{E}_c$ epistemic uncertainty: due to model knowledge and can be reduced by providing more training data.

Uncertainty Estimation on an Mild Pathology

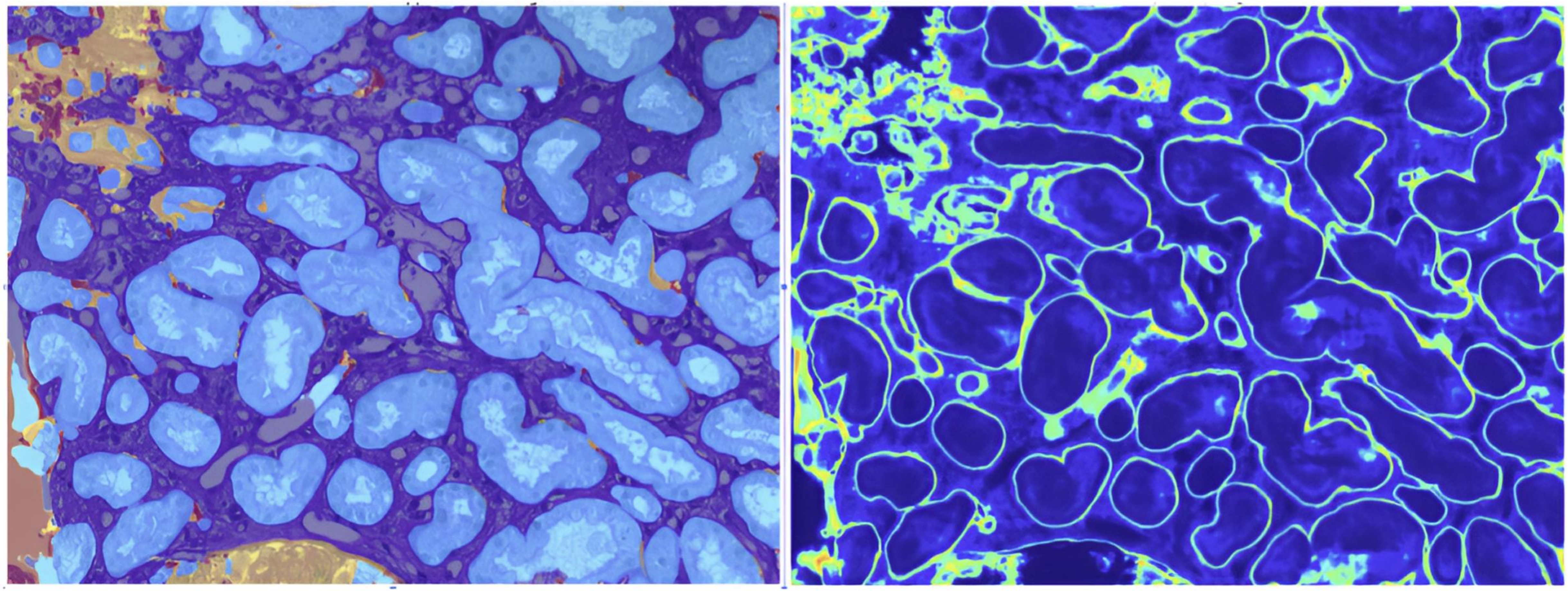
High uncertainity pixels are close to cell boundaries





Uncertainty Estimation on a Severe Pathology

On the patient U_0^D is way more uncertain, with larger uncertainty estimates.
We use uncertainty to **weight more errors on scribbles** overlapping to **uncertain regions**



Fine tuning using uncertainty estimates

The network is fine tuned using weights that uses as class weights

$$w_c(i,j) = \mathcal{V}_c(i,j)$$

- That are conveniently clipped to 1 when they exceed 0.8.
- Uncertanty weights are used only on scribble annotations.

Quantitative Assessment, avg. class. error on all the classes

Fine tuning using uncertainty weights provide superior improvements than before.

However, U_0^D performance are lower than U_0.

Patient	Initial network U_0^D	Fine-tuned U_j^D 100% scribbles main annotator	Fine-tuned U_j^D 100% scribbles secondary annotator	Fine-tuned all patients
p01	0.816	0.887	0.839	0.437
p02	0.734	0.848	0.814	0.531
p03	0.606	0.659	0.665	0.370
p04	0.648	0.649	0.680	0.415
p05	0.491	0.756	0.777	0.411
p06	0.600	0.807	0.785	0.432
p07	0.615	0.635	0.593	0.500
p08	0.849	0.856	0.826	0.556
p09	0.586	0.627	0.634	0.333
p10	0.716	0.757	0.789	0.348
p11	0.508	0.621	0.601	0.354
p12	0.693	0.799	0.686	0.445
p13	0.880	0.894	0.822	0.465
p14	0.716	0.713	0.796	0.397
p15	0.569	0.748	0.232	0.467
p16	0.719	0.775	0.799	0.419
p17	0.836	0.824	0.825	0.381
p18	0.692	0.688	0.734	0.525
p19	0.807	0.698	0.750	0.426
p20	0.811	0.849	0.825	0.547
Median [interquartile range]	0.705 [0.605; 0.808]	0.757 [0.681; 0.830]	0.781 [0.676; 0.816]	0.429 [0.393; 0.476]

Table 4
Mean dice scores for all classes obtained using the Bayesian network U_0^D, the networks U_j^D fine-tuned using 100% of the main annotator's scribbles, the networks U_j^D fine-tuned using 100% of the secondary annotator's scribbles, and the networks fine-tuned using all patients (main annotator).

Quantitative Assessment, avg. class. error interstitial class

The performance improvement is even more consistent (19 out of 20) through the patients and apparent

U_j^D still outperforms U_j in 9 out of 20 cases

Patient	Initial network U_0^D	Fine-tuned U_j^D 100% scribbles main annotator	Fine-tuned U_j^D 100% scribbles secondary annotator	Fine-tuned all patients
p01	0.779	0.786	0.715	0.640
p02	0.619	0.741	0.673	0.689
p03	0.685	0.840	0.852	0.747
p04	0.597	0.813	0.805	0.750
p05	0.403	0.864	0.881	0.703
p06	0.680	0.872	0.869	0.667
p07	0.782	0.799	0.614	0.595
p08	0.764	0.779	0.648	0.721
p09	0.717	0.846	0.864	0.585
p10	0.759	0.888	0.839	0.667
p11	0.483	0.742	0.633	0.594
p12	0.761	0.833	0.756	0.731
p13	0.789	0.792	0.650	0.630
p14	0.724	0.854	0.847	0.553
p15	0.585	0.861	0.526	0.714
p16	0.667	0.830	0.811	0.575
p17	0.688	0.700	0.674	0.493
p18	0.795	0.820	0.821	0.734
p19	0.837	0.762	0.796	0.667
p20	0.819	0.851	0.805	0.739
Median [interquartile range]	0.721 [0.655; 0.780]	0.825 [0.784; 0.852]	0.801 [0.667; 0.841]	0.667 [0.595; 0.724]

Table 5
Dice scores for the class c_{RPI} obtained using the Bayesian network U_0^D, the networks U_j^D fine-tuned using 100% of the main annotator's scribbles, the networks U_j^D fine-tuned using 100% of the secondary annotator's scribbles, and the networks fine-tuned using all patients (main annotator). Summaries are in terms of median and interquartile range because the data are not normally distributed.

Conclusions

- **Patient-specific fine tuning** is beneficial
  - Experiments comparing «all-patients» fine-tuning vs patient-specific fine tuning demonstrates this
  - This effect underlines the major differences in patterns among patients
- Our solution addresses key challenges:
  - Training under **limited annotated images**
  - **Adaptation** to different visual characteristics of pathologies
  - Uncertainty estimates.
- **Uncertainty maps** were estimated from an expert who routinely engages in diagnosis and treatment planning based on these images.
  - In a **qualitative comparison high-uncertainty images are deemed the most difficult from the expert to segment as well**
- Limitation: severe dependance on annotation style and amount

3D Segmentation on Mice Brain (MRI)

3D CNN for skull stripping and lesion assessment

ISTITUTO DI RICERCHE FARMACOLOGICHE MARIO NEGRI · IRCCS

Marcello De Salvo

Under Preparation…

Three Dimensional Convolutional Neural Networks for Automated Lesions and Regions Segmentation in Rodents with Traumatic Brain Injury

*Marcello De Salvo[a,b], #*Federico Moro[b], Luther Loose[a], Edoardo Mazzone[a], Edoardo Micotti[b], Farima Nasrallah[c], Virginia F.J. Newcombe[d], #Giacomo Boracchi[a], #Elisa R. Zanier[b]

These authors contributed equally to this work

[a]*Politecnico di Milano, Piazza Leonardo da Vinci, 32, 20133, Milan, Italy*

[b]*Istituto di Ricerche Farmacologiche Mario Negri, Via Mario Negri, 2, 20156, Milan, Italy*

[c]*Queensland Brain Institute, The University of Queensland, Saint Lucia, QLD 4067, Australia*

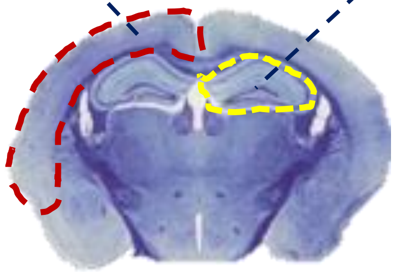
[d]*Department of Medicine, University of Cambridge, Cambridge, CB2 0QQ, UK*

Context

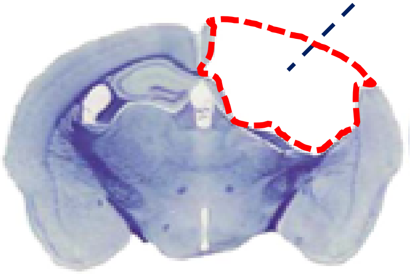
Research collaboration with a laboratory investigating recovery from Traumatic Brain Injuries (TBI) via pre-clinical studies.

Goal: Design tools to automatically process brain MRI of mice/rats under study. These are useful to compute biomarkers to assess the recovery.

Rodents Model
(this is not a NN…)



Healty brain

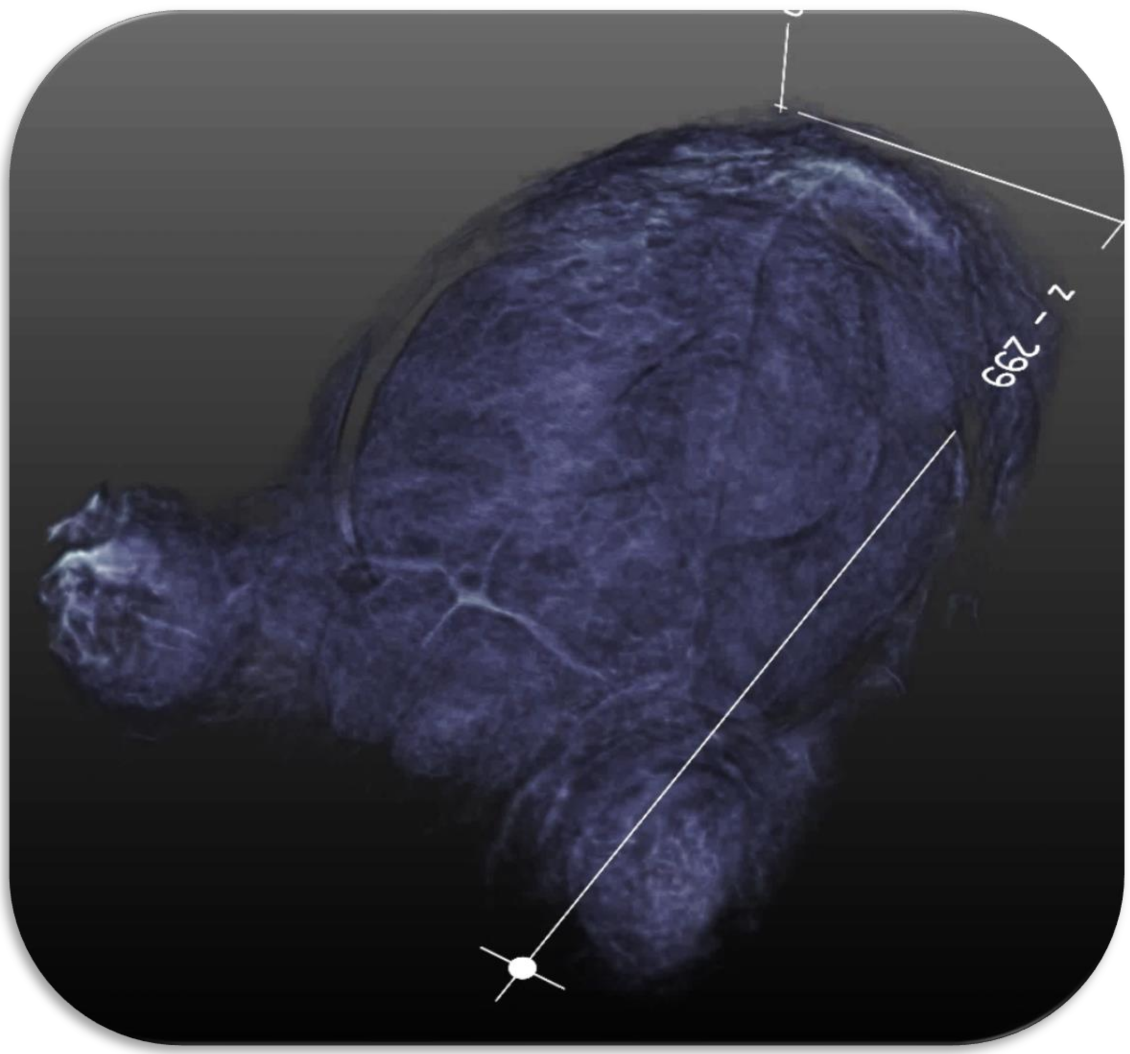


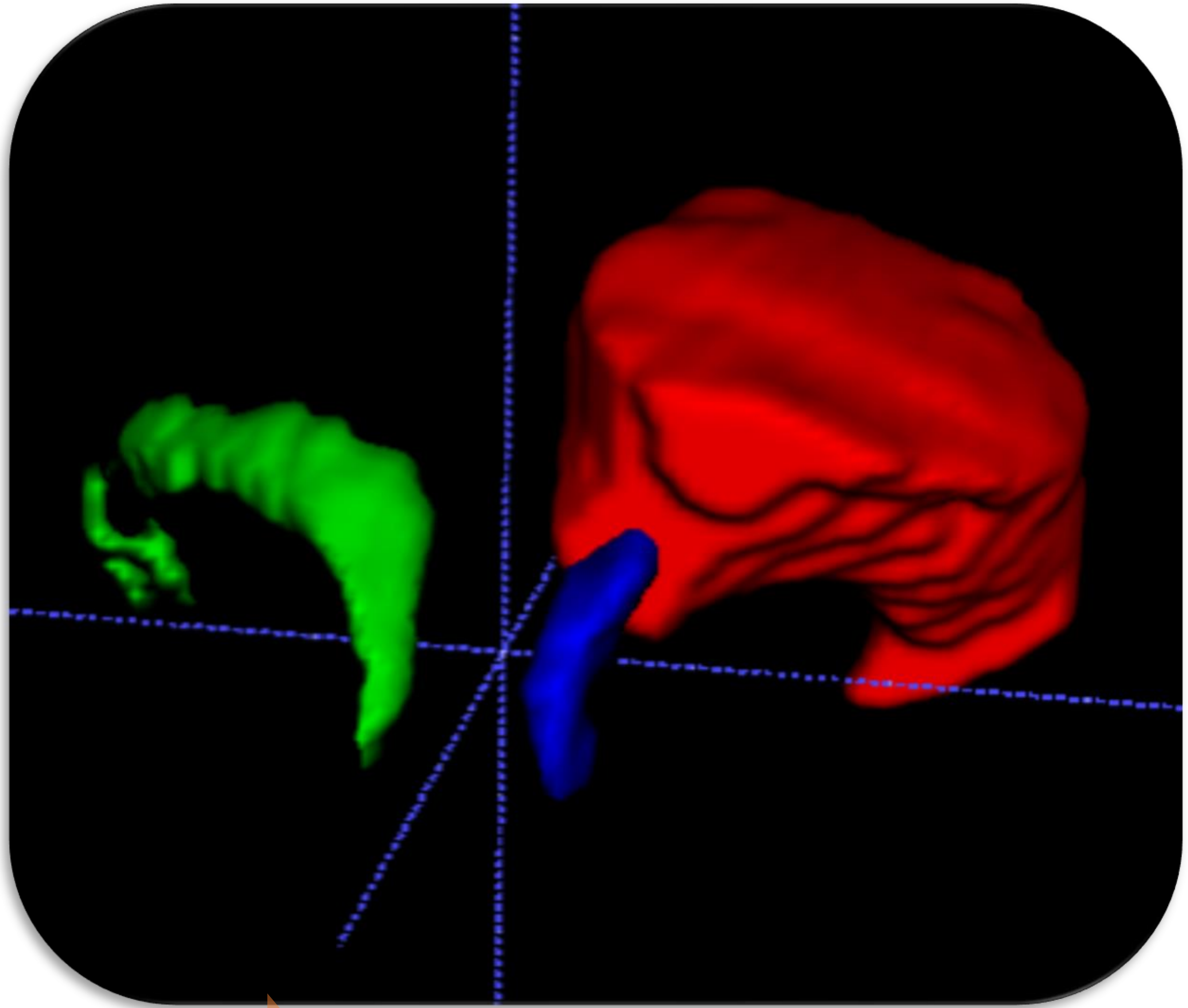
TBI

MRI

Assessing 3D
Lesion Volume

Problem Formulation





3D image from MRI $I \in \mathbb{R}^{+H\times W\times D\times C}$
Each voxel contains signal intensity

$f: \mathbb{R}^{+\,H\times W\times D\times C} \rightarrow K^{\,H\times W\times D}$

3D Semantic segmentation mask
$\hat{\Delta} \in K^{H\times W\times D}$

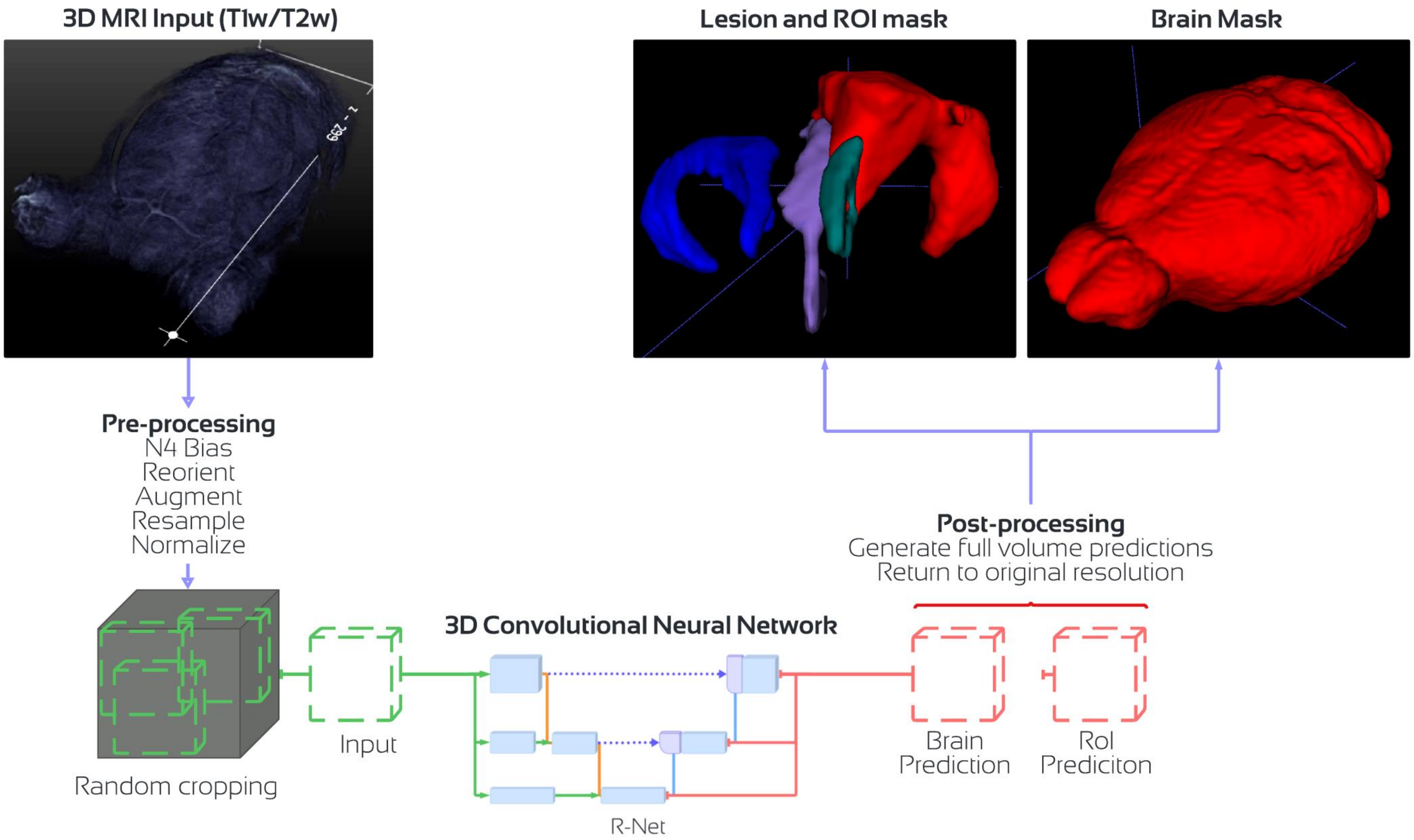
Challenges

- **Atalas-based approaches are not viable**.
- **Shortage of annotated data**:
    - 3D segmentation is extremely time consuming
    - Require experts' annotation
    - Lack of annotated training data or pre-trained models on mice/rats
- **Domain shift w.r.t humans**, where there are multiple annotated data
    - Lesions modify the shape of atlas and prevent using standard approaches.
    - Lesion is typically not a target class in datasets
    - Different types of MRI (FLASH, Rare)
    - Different species (mice/rats)

Idea of the Solution to compensate for annotation shortage

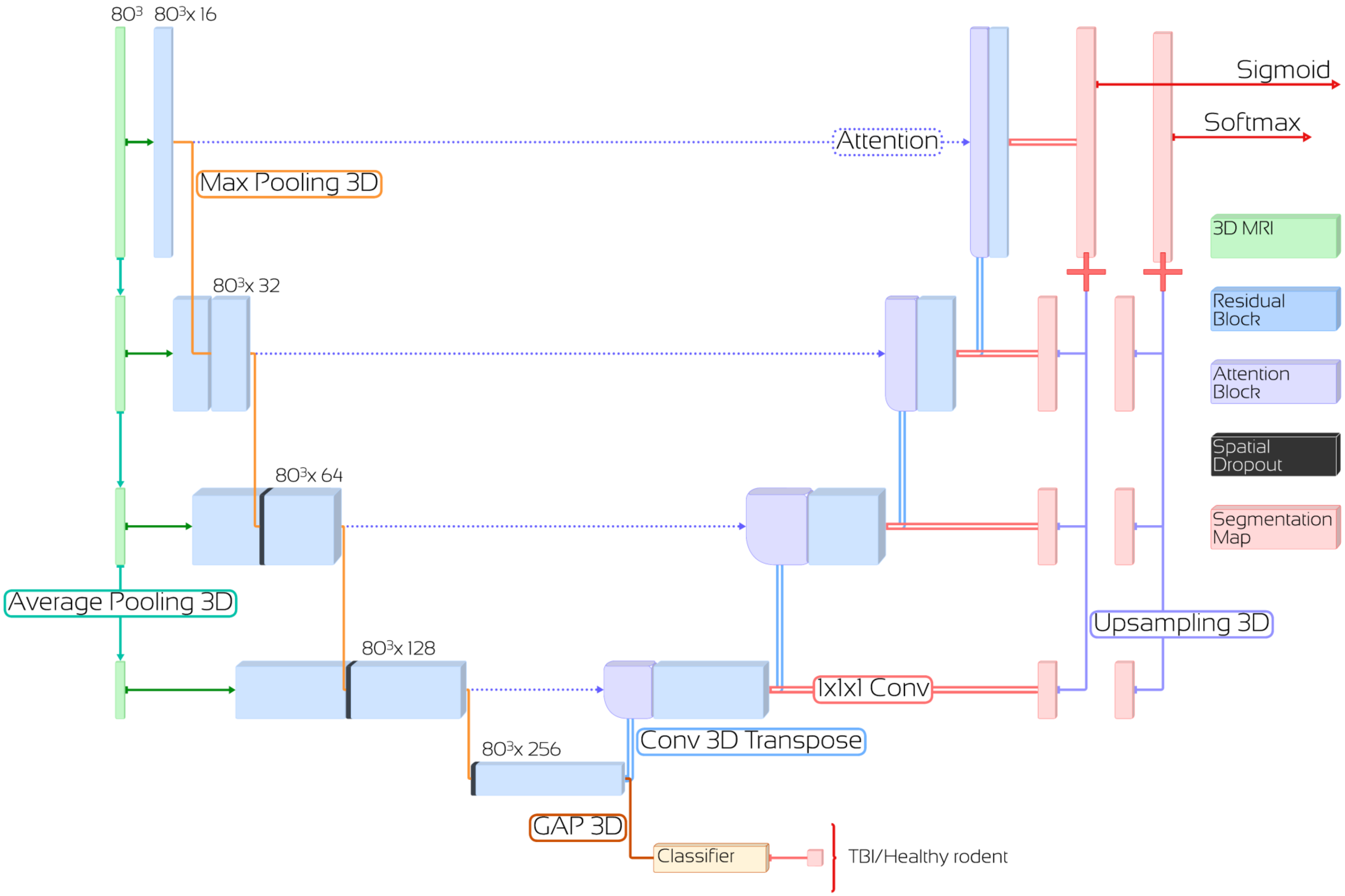
- Formulating auxiliary learning problems where it is possible to get supervision
  - Segmentation of additional regions, not only the target ones, but also some for which we can get auxiliary annotated data.
  - Skull-Stripping as an additional learning task.
  - Classifier in the latent space (TBI/healty).
- Two-headed segmentation network trained on multiple data sources.
- Combination of different loss functions.

R-Net 3D overview

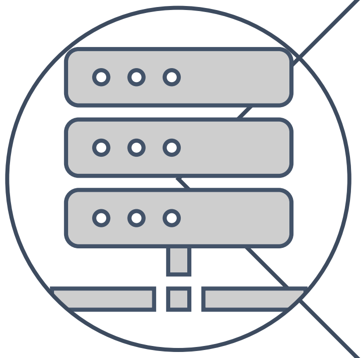


R-Net 3D architecture

- 3D U-net layout
- Three output branches:
  - Lesion (softmax)
  - Skull-stripping (sigmoid)
  - TBI/Healthy rodent classifier
- Multi-task learning loss



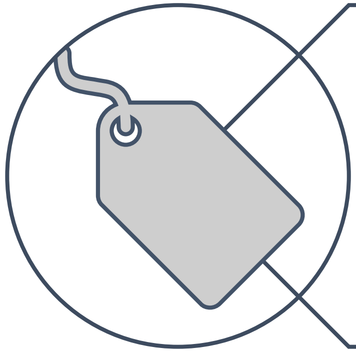
3D U-Net Skull Stripping



Dataset

- 48 FLASH
- 38 RARE
- 25 DTI

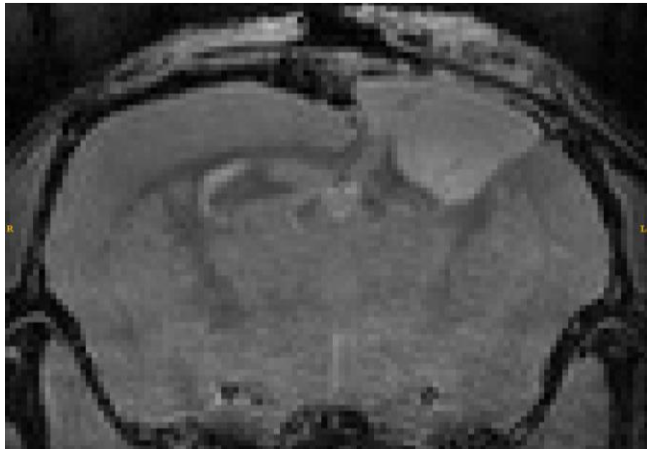
Provided with N4 Bias field correction



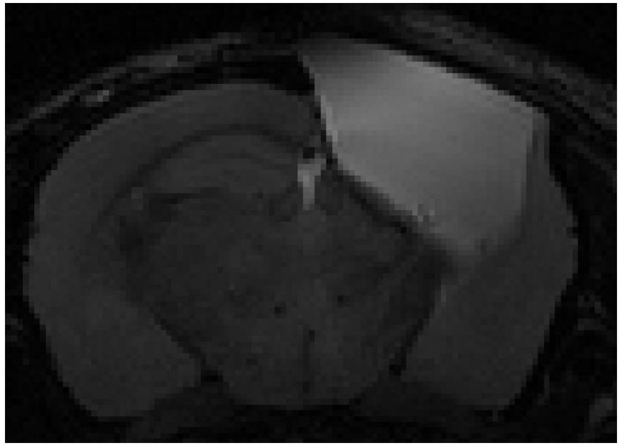
Binary segmentation

0 – Background
1 – Brain tissue

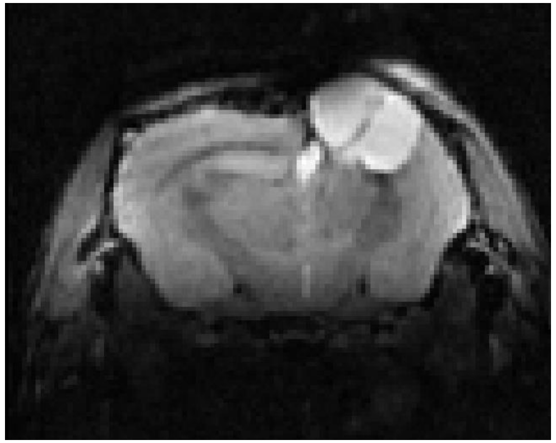
FLASH – 0.1^3mm – 110x80x300



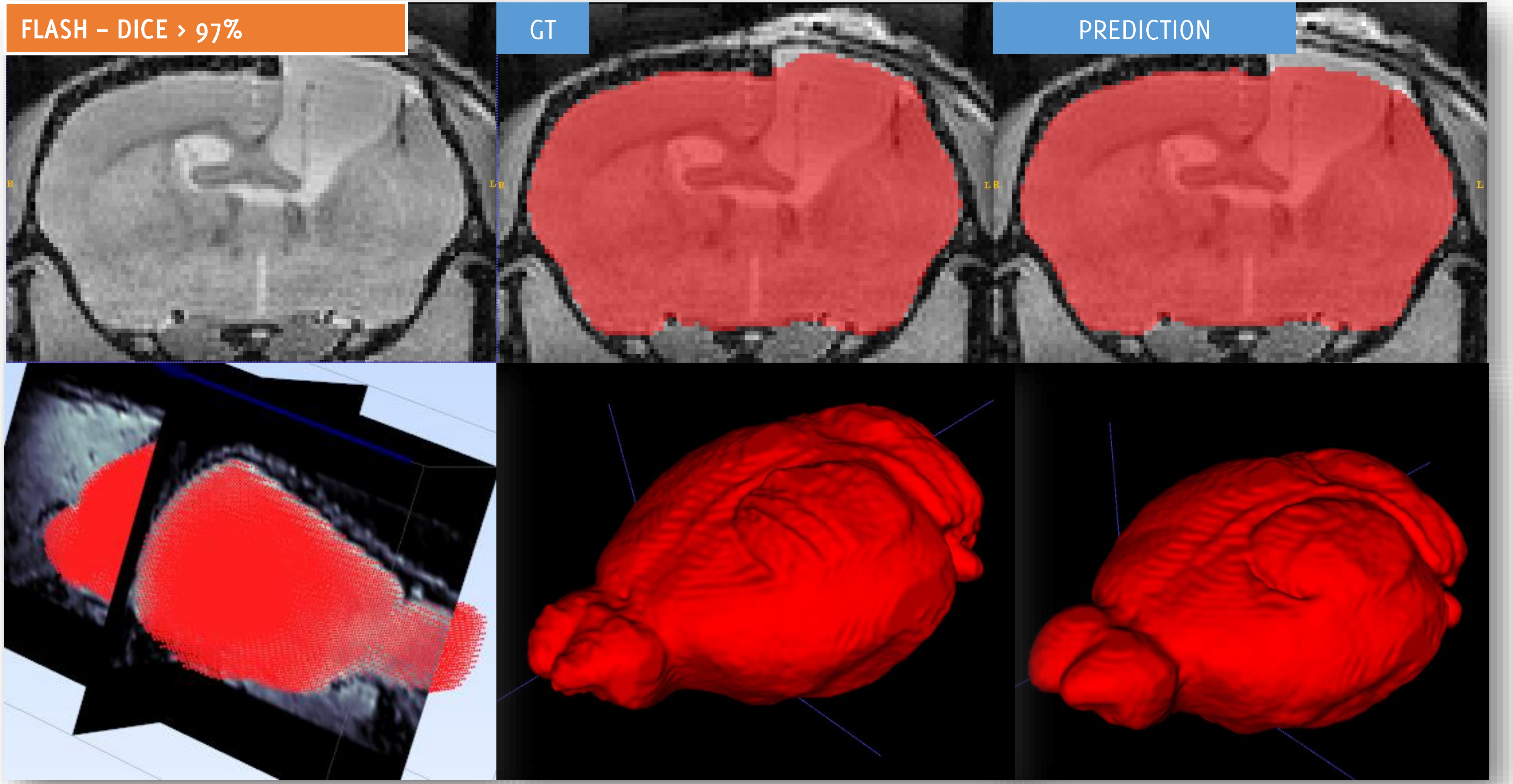
RARE – 0.1, 0.1, 0.3 mm – 150x150x37



DTI – 0.12, 0.12, 0.3 mm – 120x120x28



3D U-Net Skull Stripping



3D U-Net RARE Lesion Segmentation

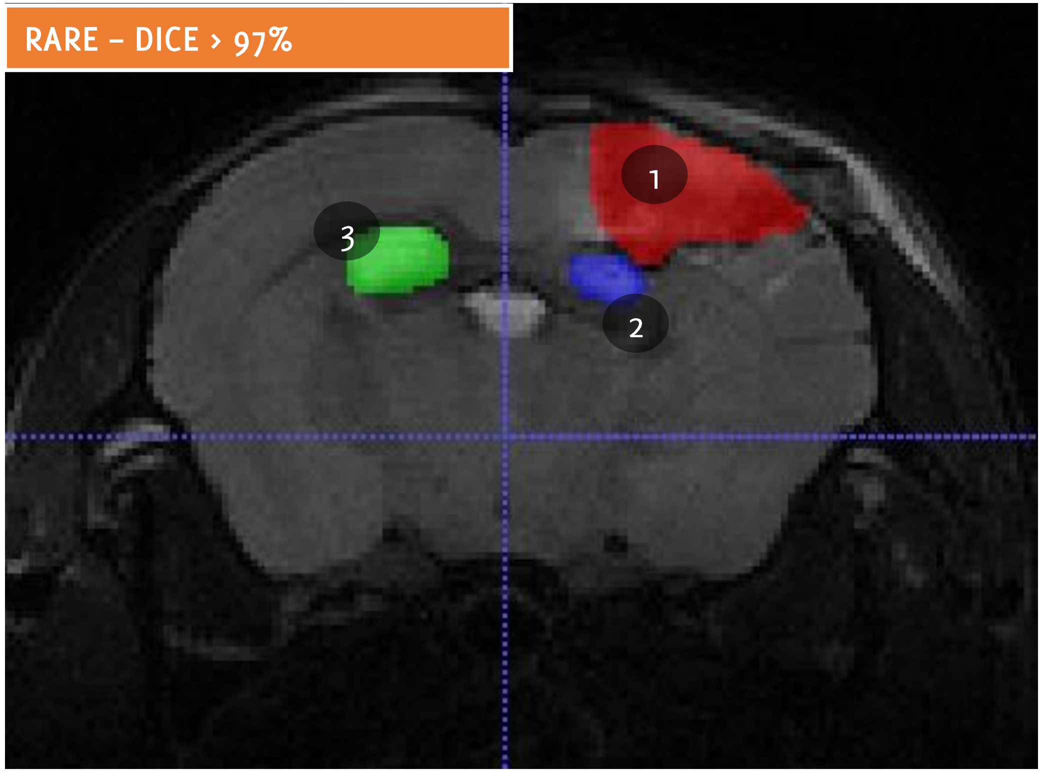
Dataset

- 145 RARE

Provided with N4 Bias field correction

Four Classes

0 – Background
1 – **Lesion**
2 – Ipsi Ventricle
3 – Contra Ventricle



3D U-Net RARE + FLASH Lesion Segmentation

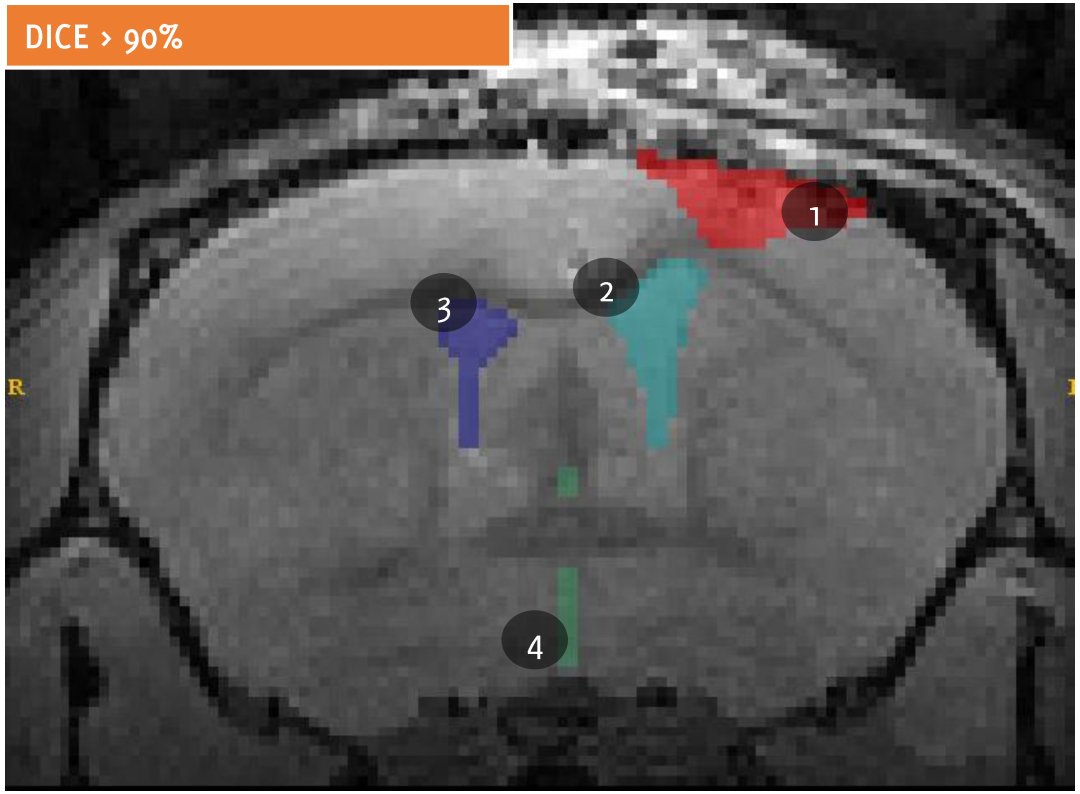
Dataset

- 33 FLASH
- 16 RARE

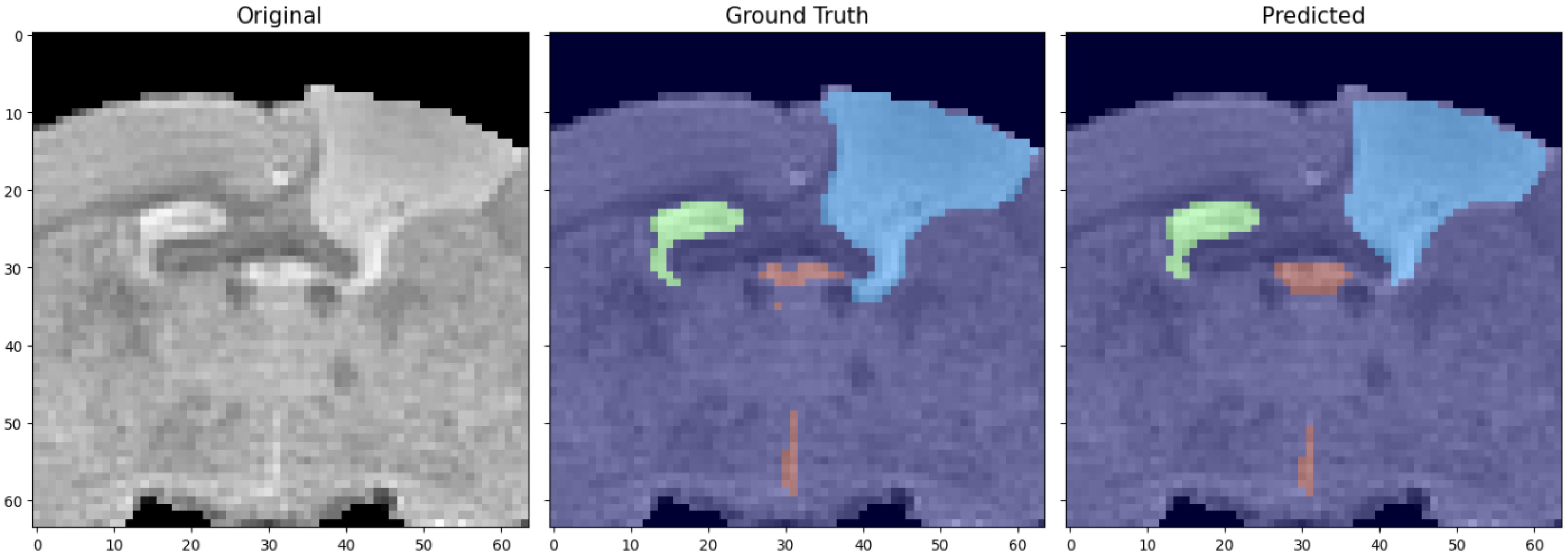
Provided with N4 Bias field correction

Four Classes

0 – Background
1 – **Lesion**
2 – Ipsi Ventricle
3 – Contra Ventricle
4 – Third Ventricle



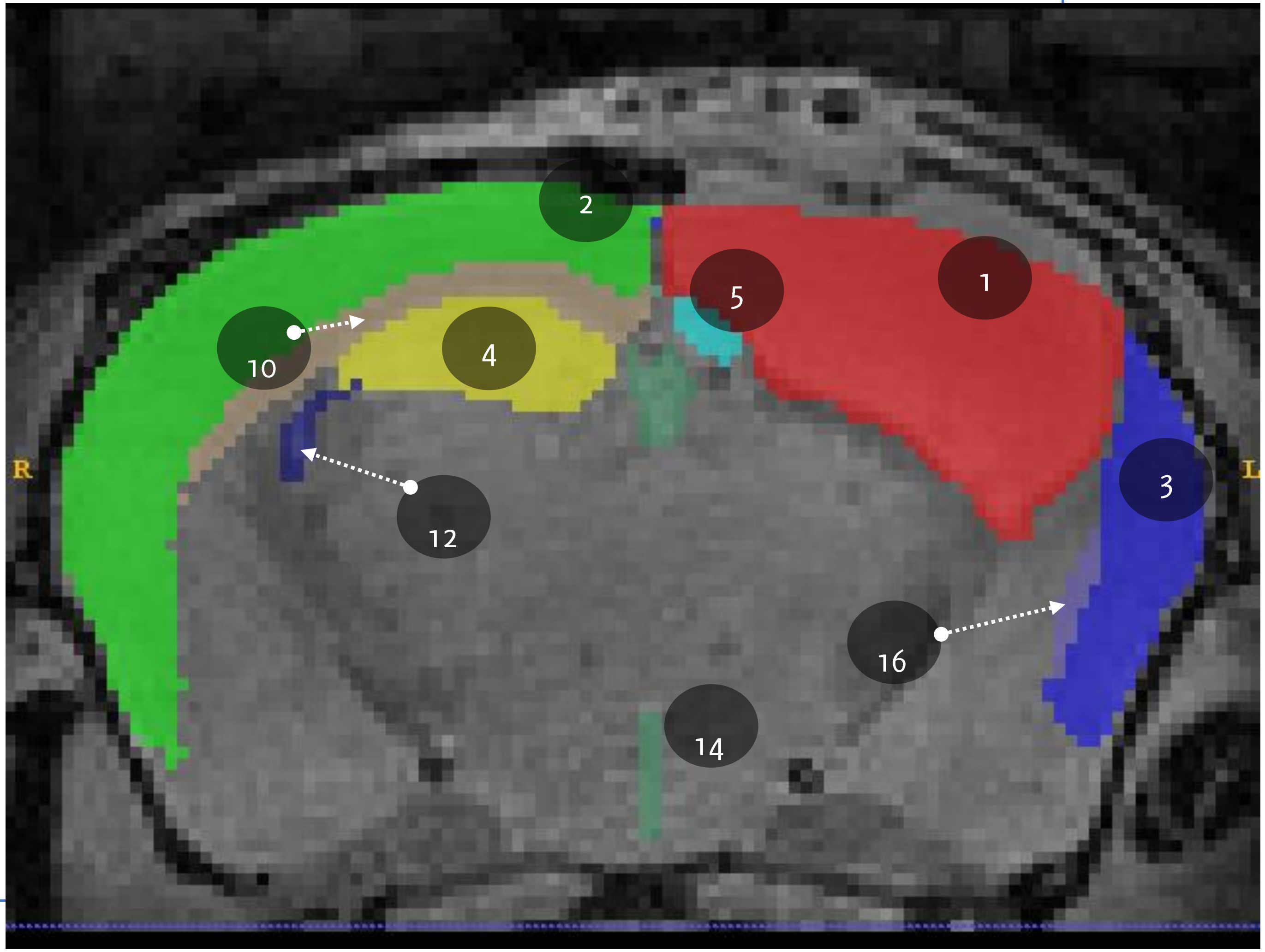
3D U-Net FLASH Lesion Segmentation



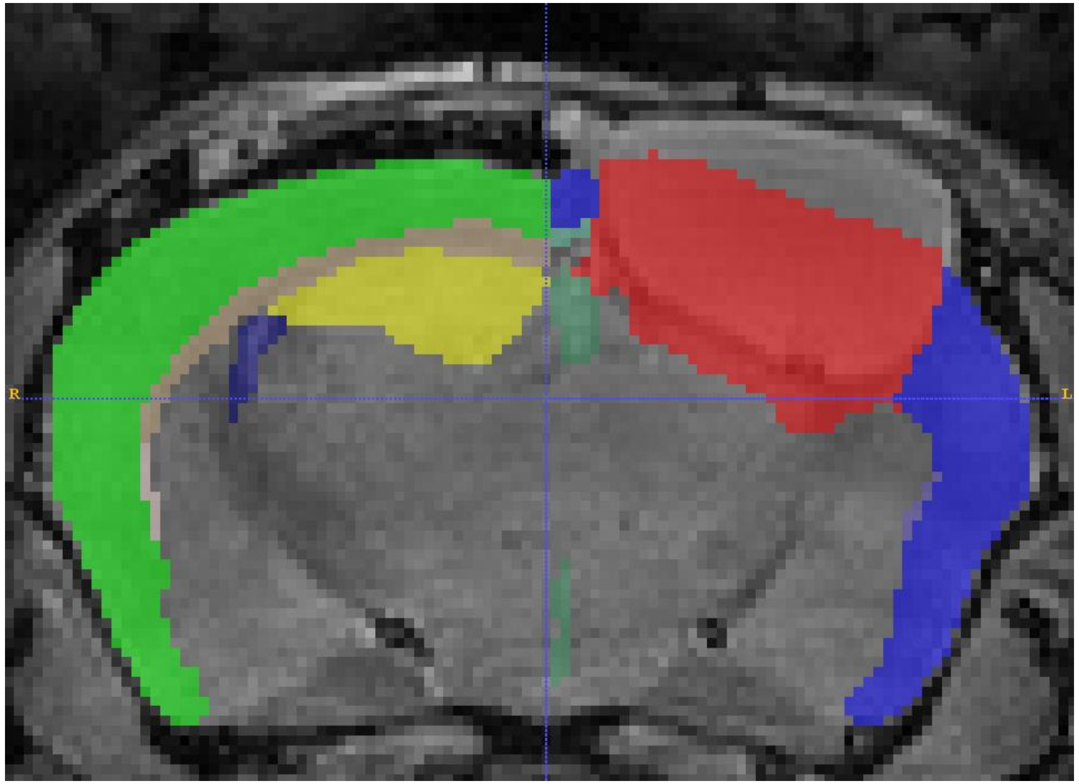
3D U-Net Flash Roi Segmentation

Different Annotations: 9 regions from healthy individuals, 4 regions from TBI

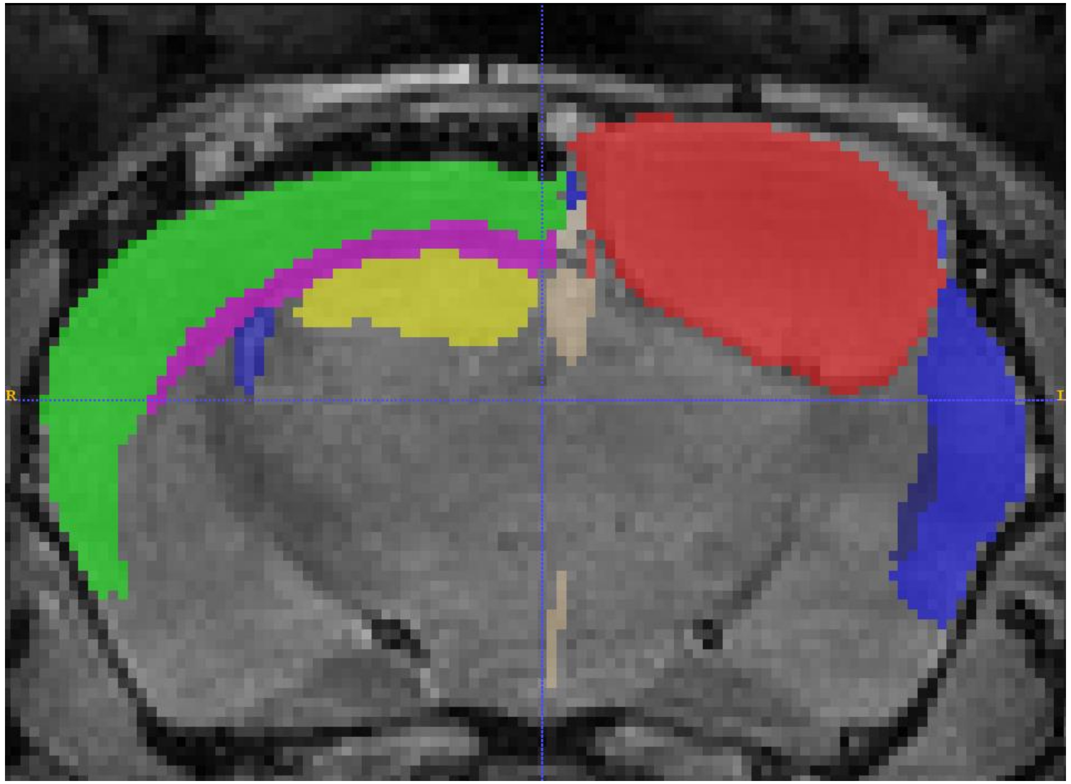
- Background (label 0)
- **Lesion** (red, label 1)
- Cortex contra (green, label 2)
- Cortex ispi (blue, label 3)
- Hippo contra (yellow, label 4)
- Hippo ipsi (light blue, label 5)
- CC contra (light brown, label 10)
- CC ipsi (light green, label 11)
- Ventricle contra (dark blue, label 12)
- Ventricle ipsi (light blue, label 13)
- Third Ventricle (green, label 14)
- EC contra (pink, label 15)
- EC ipsi (purple, label 16)



3D U-Net Flash Roi Segmentation

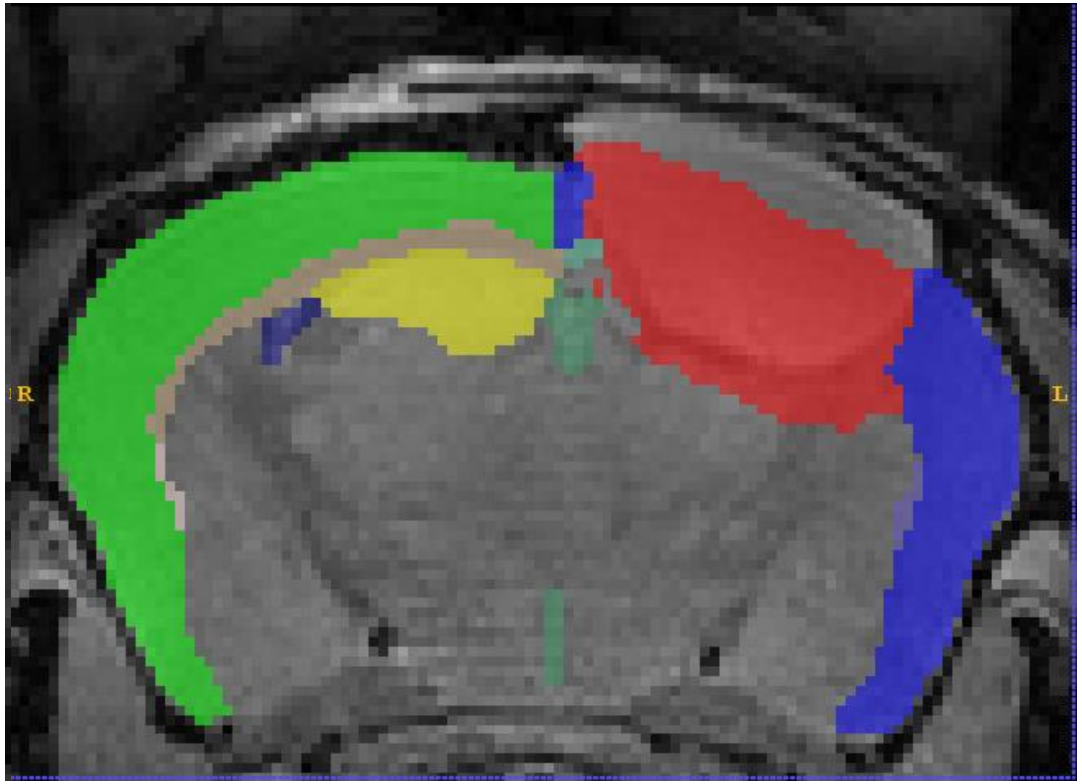


Multi Atlas

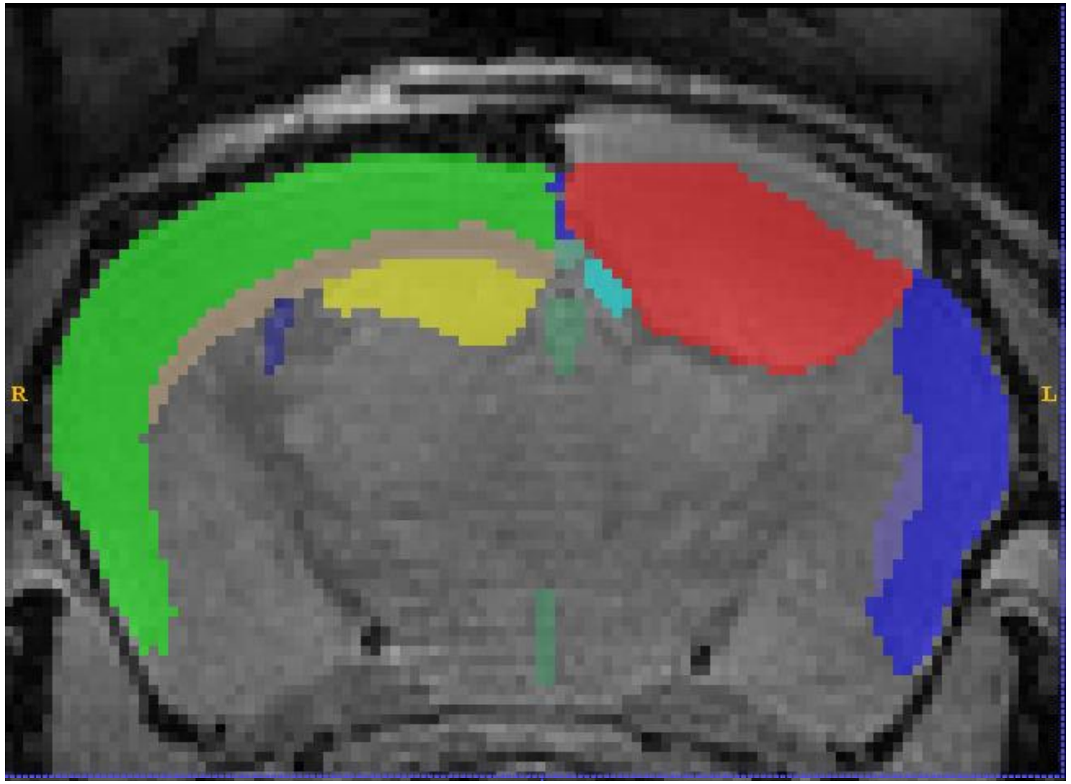


CNN

3D U-Net Flash Roi Segmentation



Multi Atlas



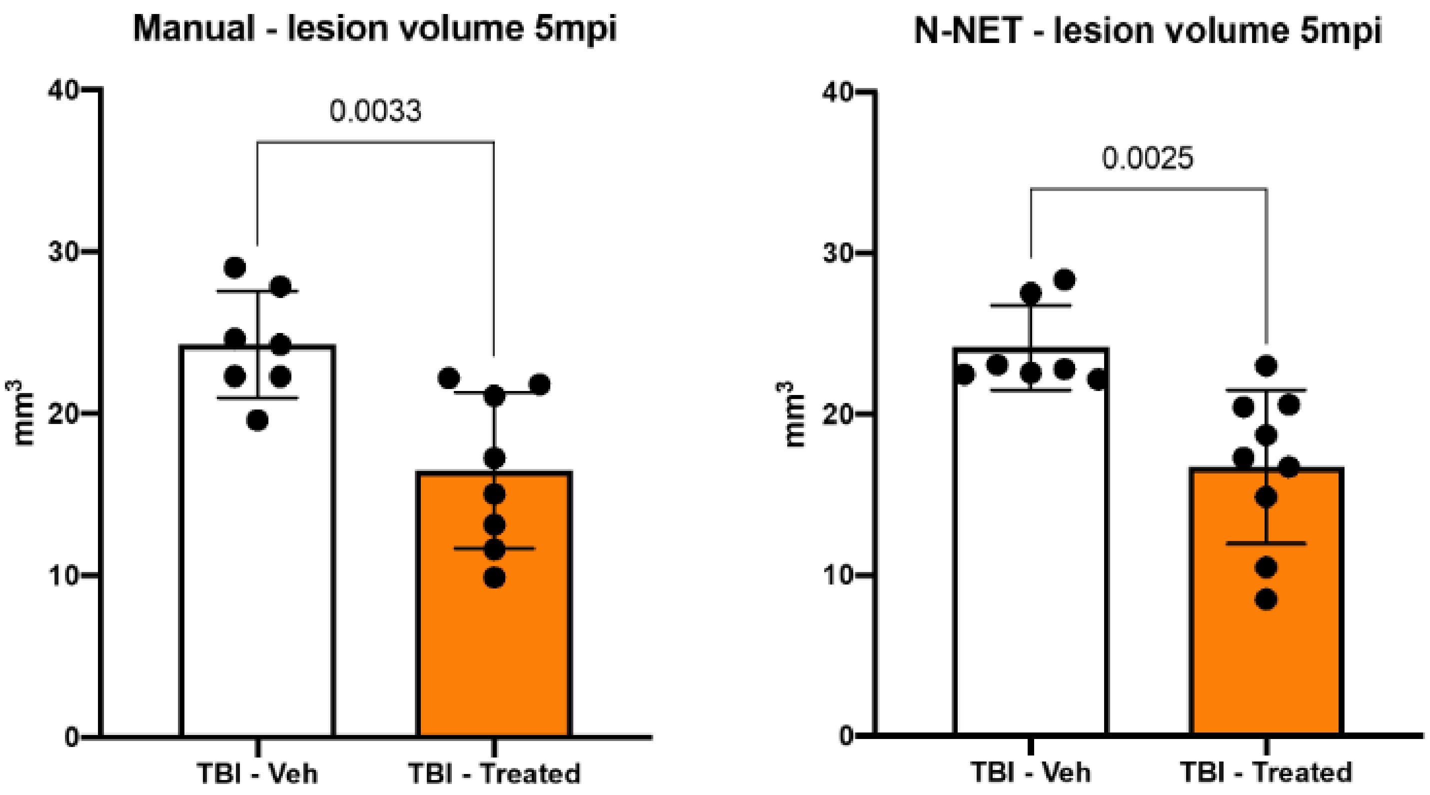
CNN

Volume Lesion Estimation

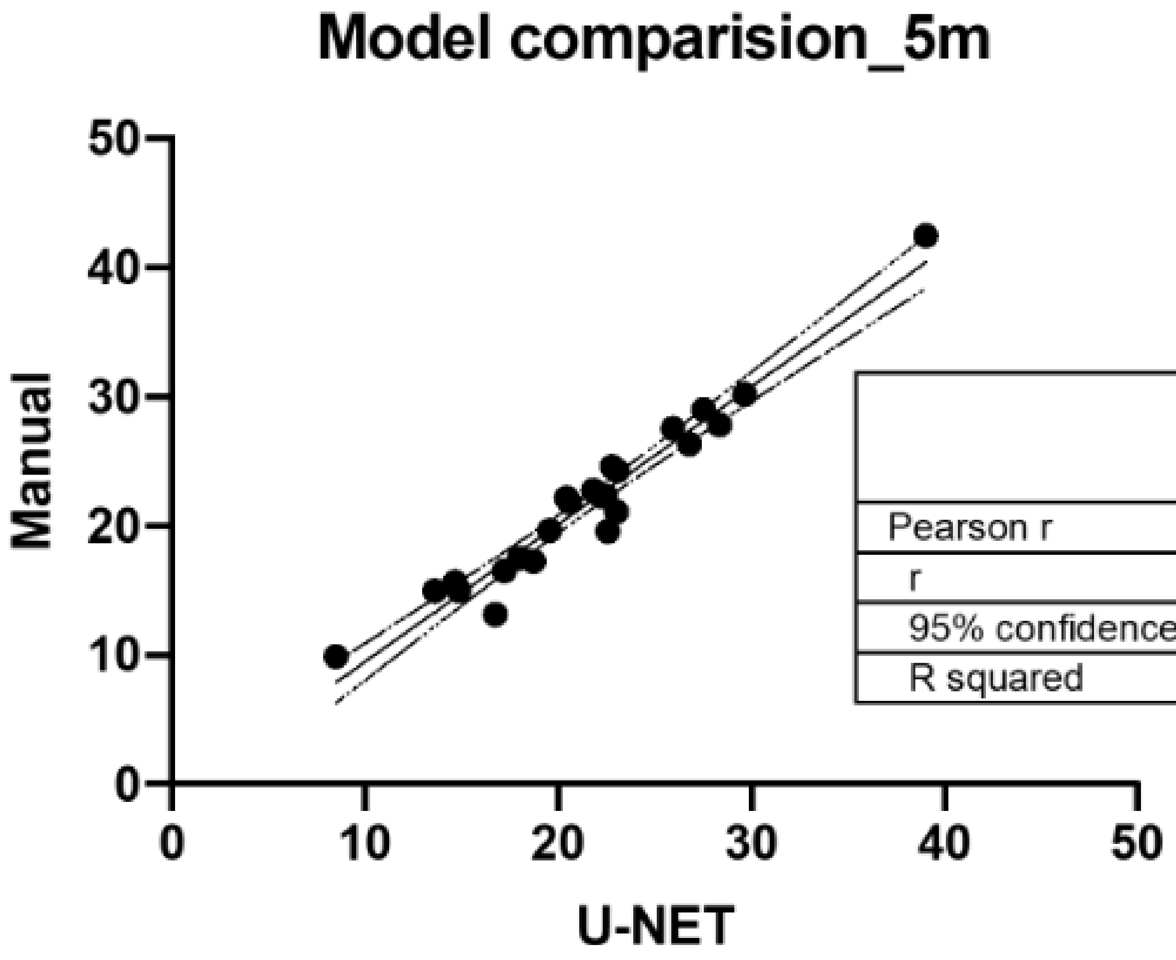
Tested on 16 extra mice, divided in two groups:

- TBI treated
- TBI control (no treatment)

Comparison against manual annotations



R-3Net and manual annotations are correlated



	U-NET vs. Manual
Pearson r	
r	0.9741
95% confidence interval	0.9400 to 0.9889
R squared	0.9488

An Expert-driven Data Generation Pipeline for Histological Images

Roberto Basla, Loris Giulivi, Luca Magri, Giacomo Boracchi,
"An expert-driven data generation pipeline for histological images"
International Symposium on Biomedical Imaging, May 2024

Roberto Basla

AN EXPERT-DRIVEN DATA GENERATION PIPELINE FOR HISTOLOGICAL IMAGES

Roberto Basla, Loris Giulivi, Luca Magri, Giacomo Boracchi

{name.surname}@polimi.it DEIB, Politecnico di Milano, Italy

ABSTRACT

Deep Learning (DL) models have been successfully applied to many applications including biomedical cell segmentation and classification in histological images. These models require large amounts of annotated data which might not always be available, especially in the medical field where annotations are scarce and expensive. To overcome this limitation, we propose a novel pipeline for generating synthetic datasets for cell segmentation. Given only a handful of annotated images, our method generates a large dataset of images which can be used to effectively train DL instance segmentation models. Our solution is designed to generate cells of realistic shapes and placement by allowing experts to incorporate domain knowledge during the generation of the dataset.

Index Terms— Instance Segmentation, Data Generation, Deep Learning.

can only be applied to already-existing samples resulting in a limited increase of variability. Image generation, instead, has the potential to obtain a large amount of diverse data, enabling more effective model training. On the flip side, this also requires generating annotations (here also referred to as *blobs*) that are pixel-wise consistent with generated samples.

A few efforts [2, 3] have been made towards generating both image and GT. These rely on DL models like Generative Adversarial Networks (GANs) [4] that, while providing good results, do not enable to steer the image generation towards images featuring desired properties like the cell distribution and spacing. Other works break down the generation problem to make it more controllable, but are limited to re-using cell masks from real data [2], or generating blobs at random [5], yielding potentially unrealistic results. Other approaches extract blobs from real images and place them over an empty canvas to create the image mask [6, 7]. Lastly, works such as

Data Generation for Instance Segmentation

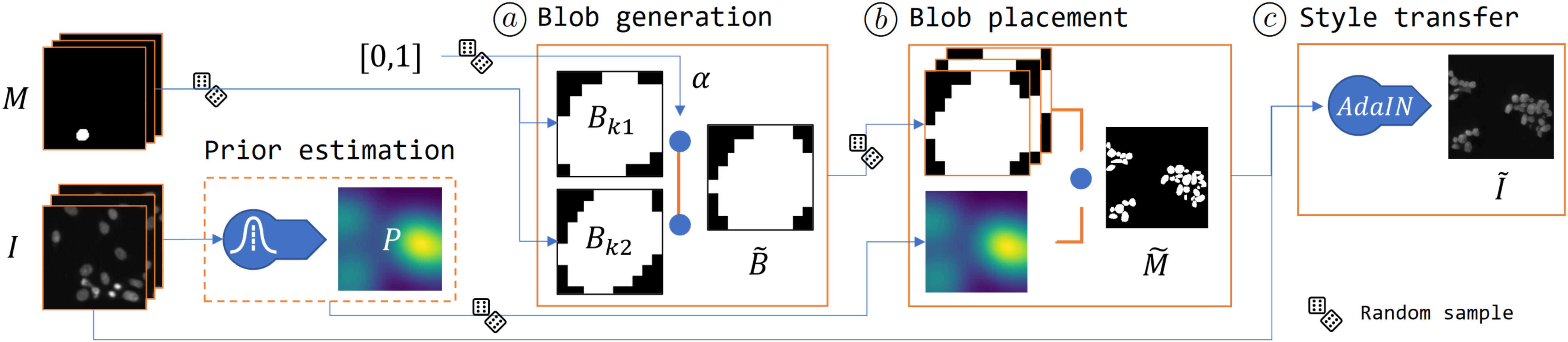
Goal: Train instance segmentation networks for different tissues

Challenges:

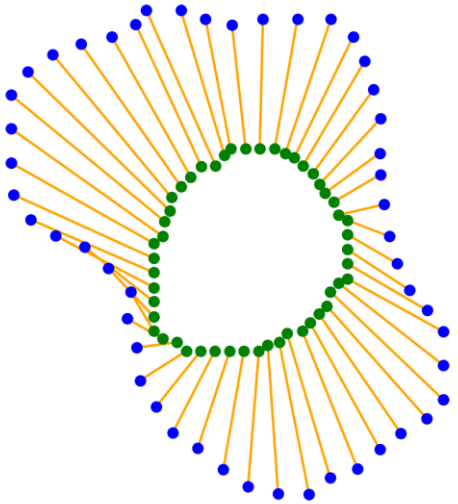
- Limited availability of annotated data, in particular for different treatment of the nuclei, different types of cells, pathologies…
- Traditional Data Augmentation might not significantly increase the variability of the dataset in very low data regimes.
- Data Generation might result in inconsistent images

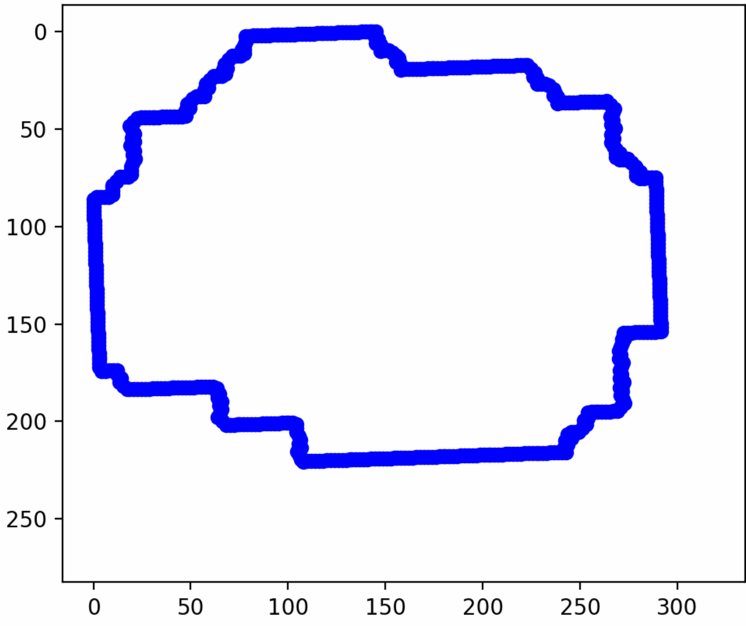
Solution: our pipeline generate realistic images to be regulated by expert-tunable parameters, starting from very few annotations.

Our data generation pipeline: Blob Generation



We first generate a set of artificial blobs $\tilde{B}$ (a) by interpolating boundary points of available masks.



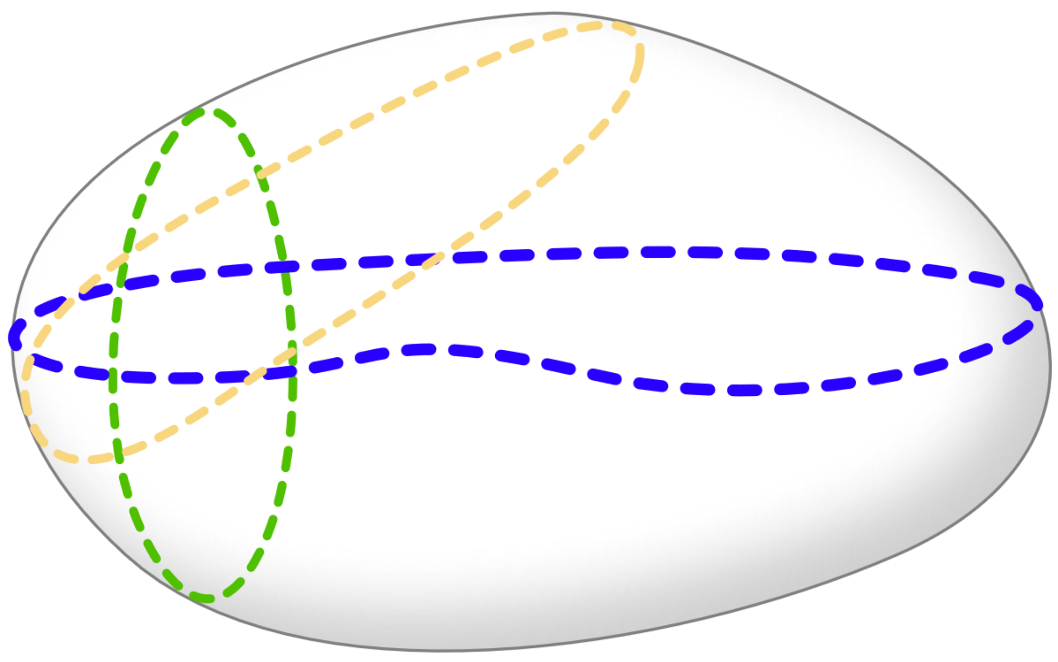


Rationale: Blob Appearance

We want to preserve two key realism aspects: the **appearance of blobs and their positioning over the image.**

To preserve blob shapes, we assume *homotopic equivalence* between different sections (or *projections*) of the same 3D nucleus

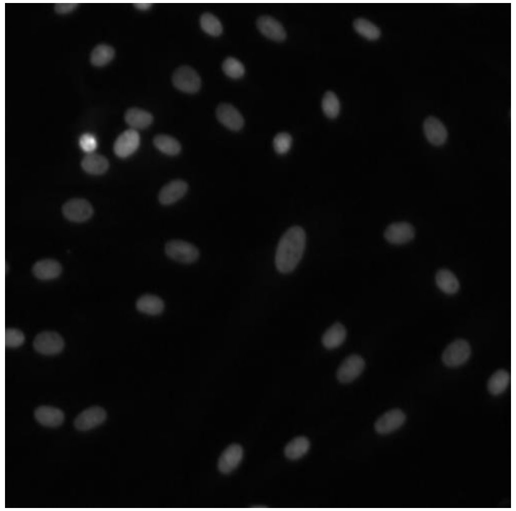
As blobs are obtained as sections of similar nuclei, we generate blobs by interpolating blob boundaries. The texture is instead preserved by a style transfer Neural Network.

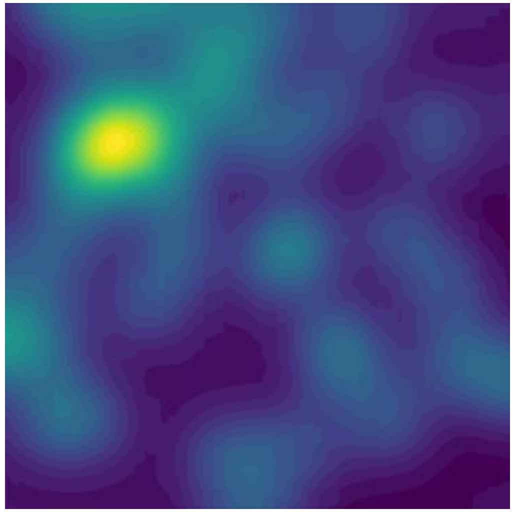


Rationale: Blob Placement

We want to preserve two key realism aspects: the **appearance of blobs and their positioning over the image.**

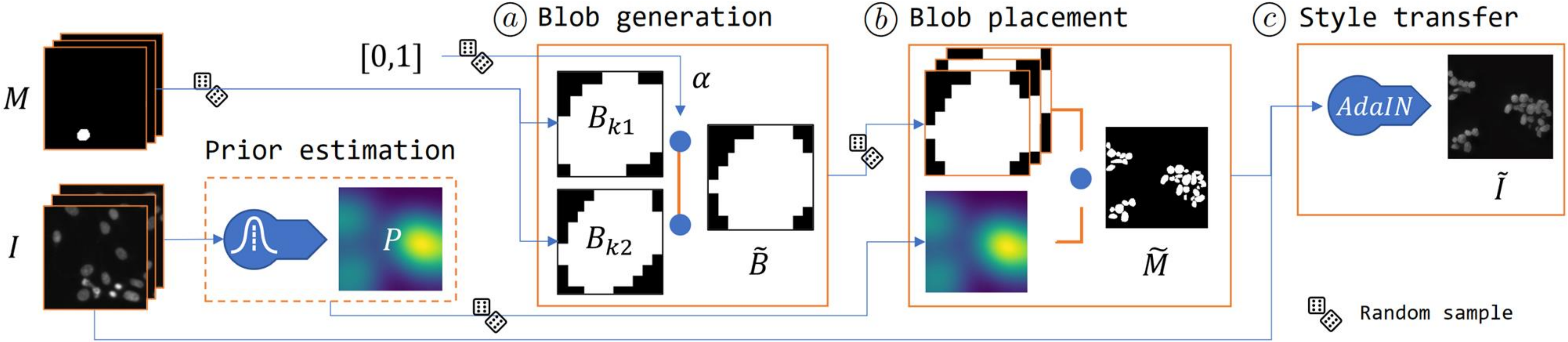
- To preserve the distribution, we perform a greedy optimization procedure on a distribution parametrized over the available real images (obtained by blurring).



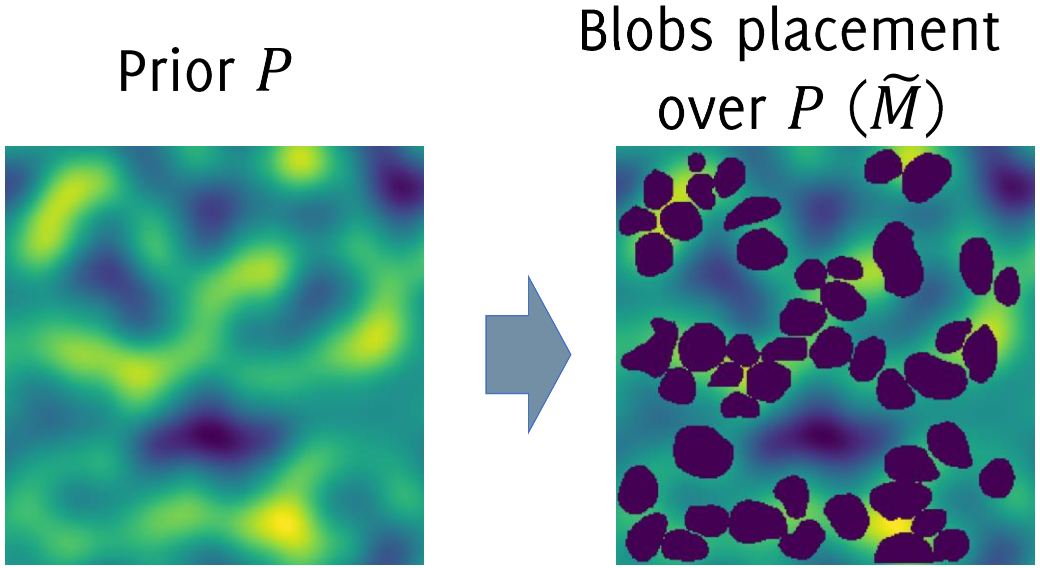


Distribution Parameters

Our data generation pipeline: Blob Placement

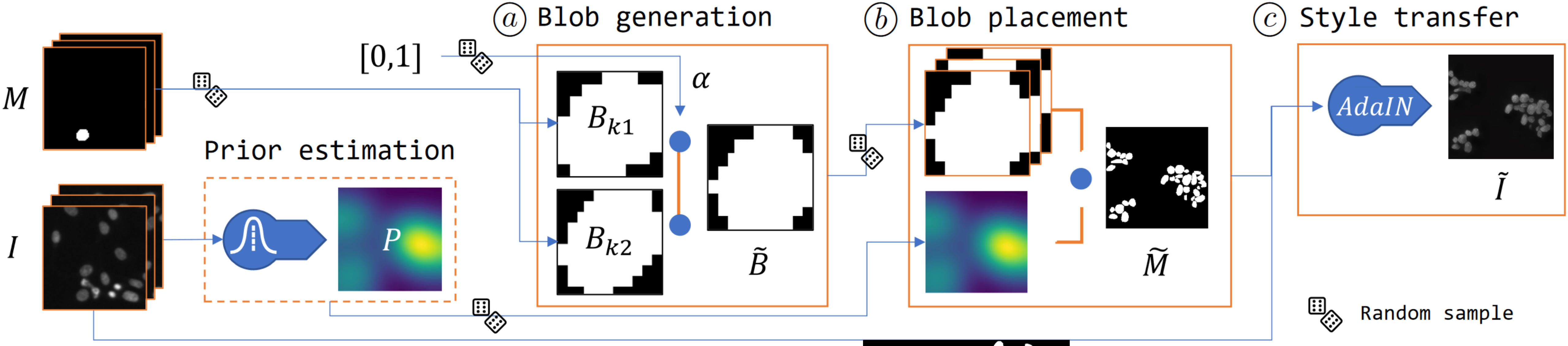


Then, we obtain our Ground Truth $\widetilde{M}$ (b) by greedily placing our blobs to maximize the coverage of a defined density distribution P. P can be either expert-defined or, as in our case, modeled from data using Perlin Noise.

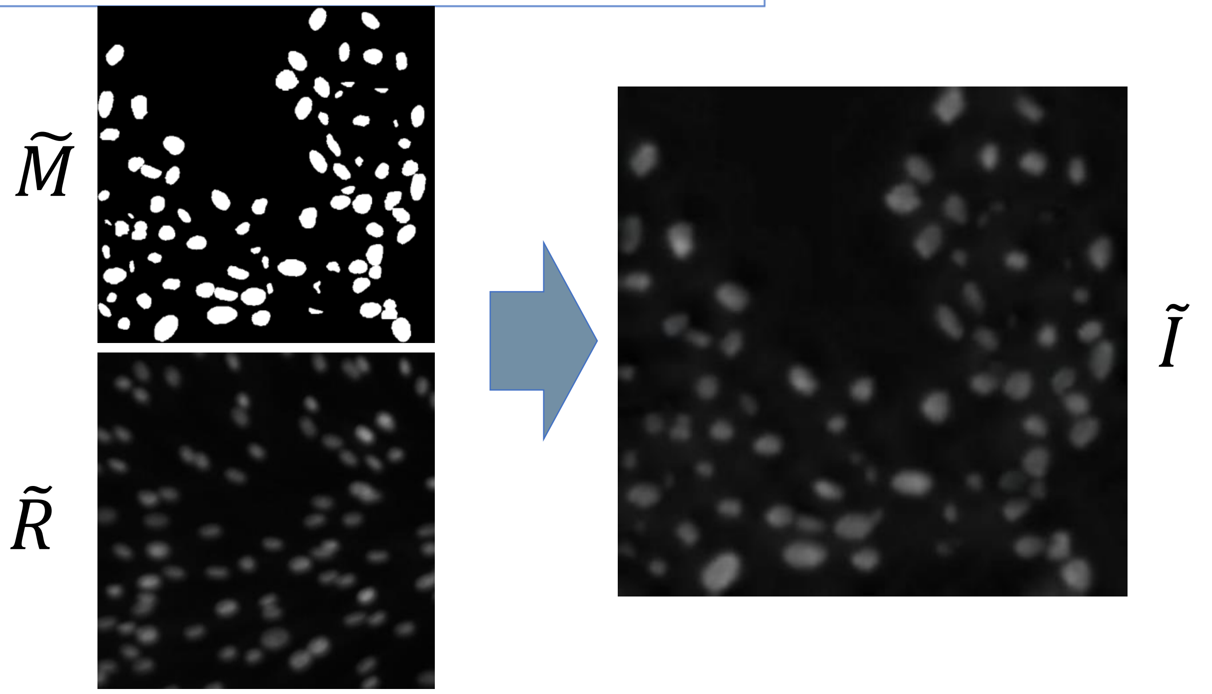


- Ken Perlin, “Improving noise,” in Proceedings of the 29th annual conference on Computer graphics and interactive techniques, 2002

Our data generation pipeline: Style Transfer

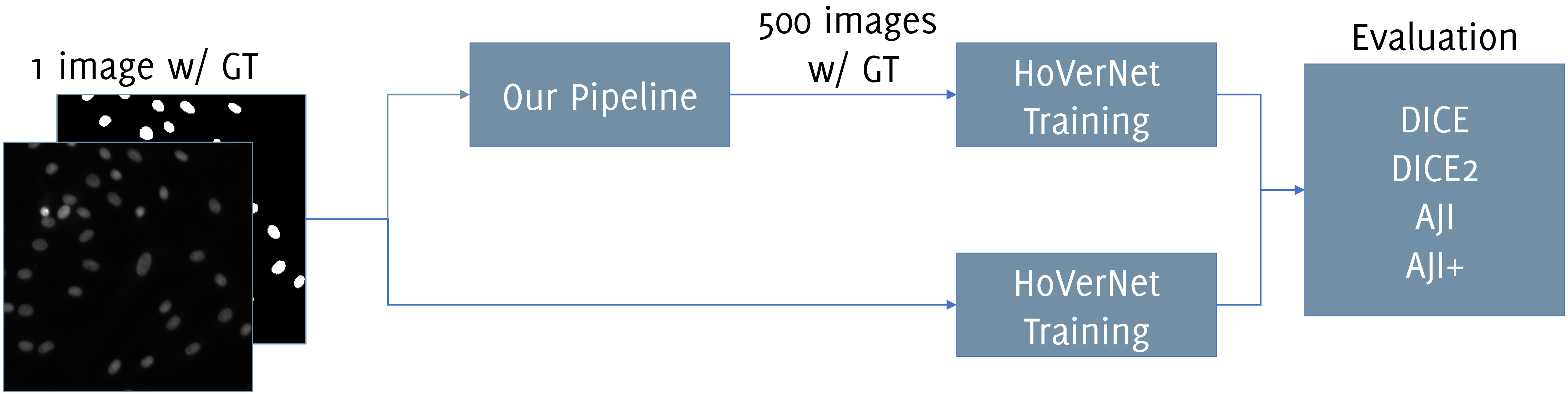


Finally, we perform style transfer (c) to transform the ground truth $\widetilde{M}$ into a histological image $\tilde{I}$ according to a reference style $\tilde{R}$.



Experiments

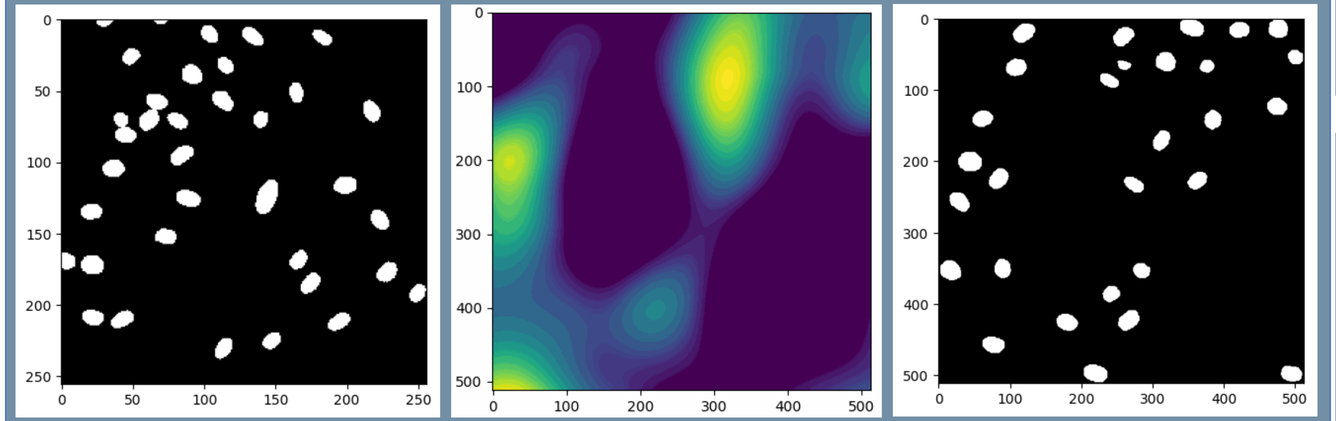
We tested our method by generating datasets using an increasing number of images/nuclei. We compared the results obtained by training a SOTA instance segmentation model (HoVerNet) on both generated datasets and the images used for generation.

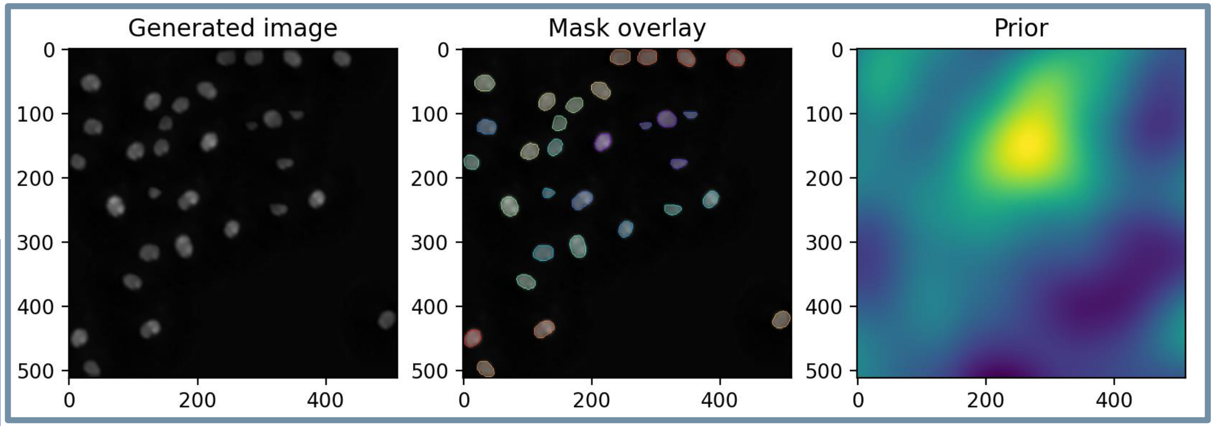


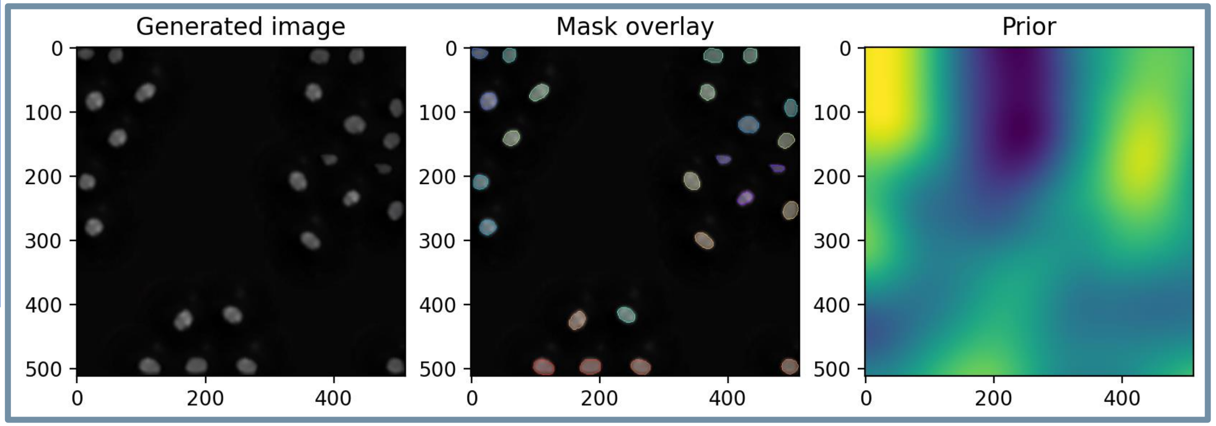
Graham, Simon, et al. "Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images." *Medical image analysis*, 2019

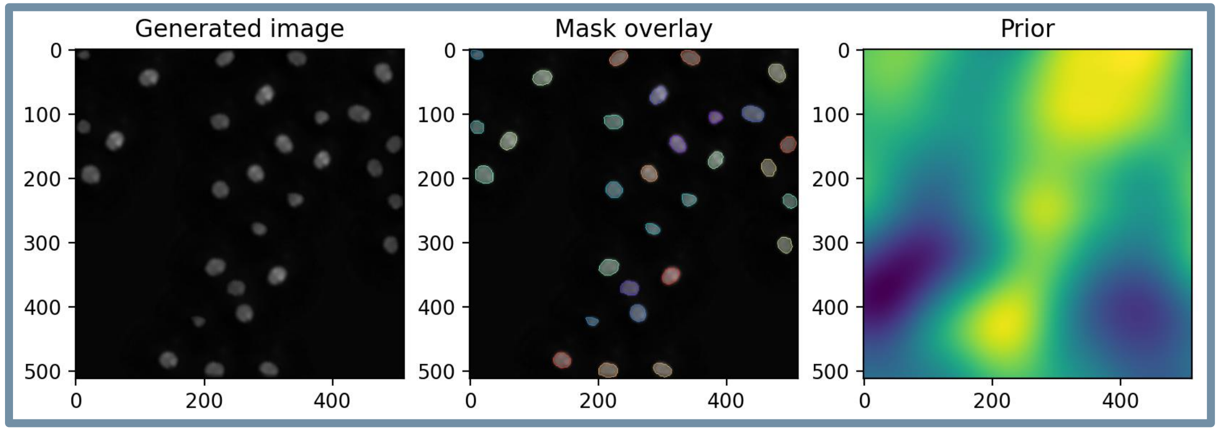
Image Generation

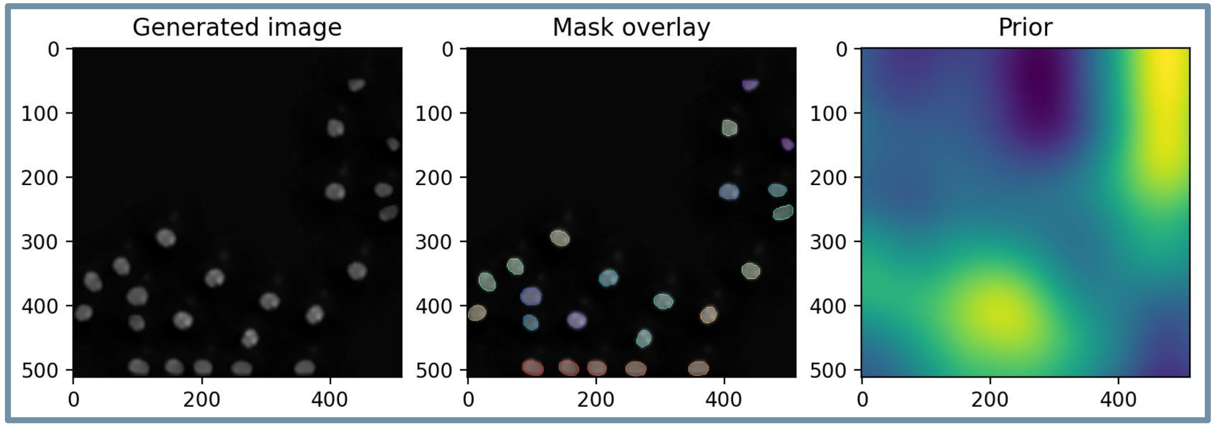
Placement example





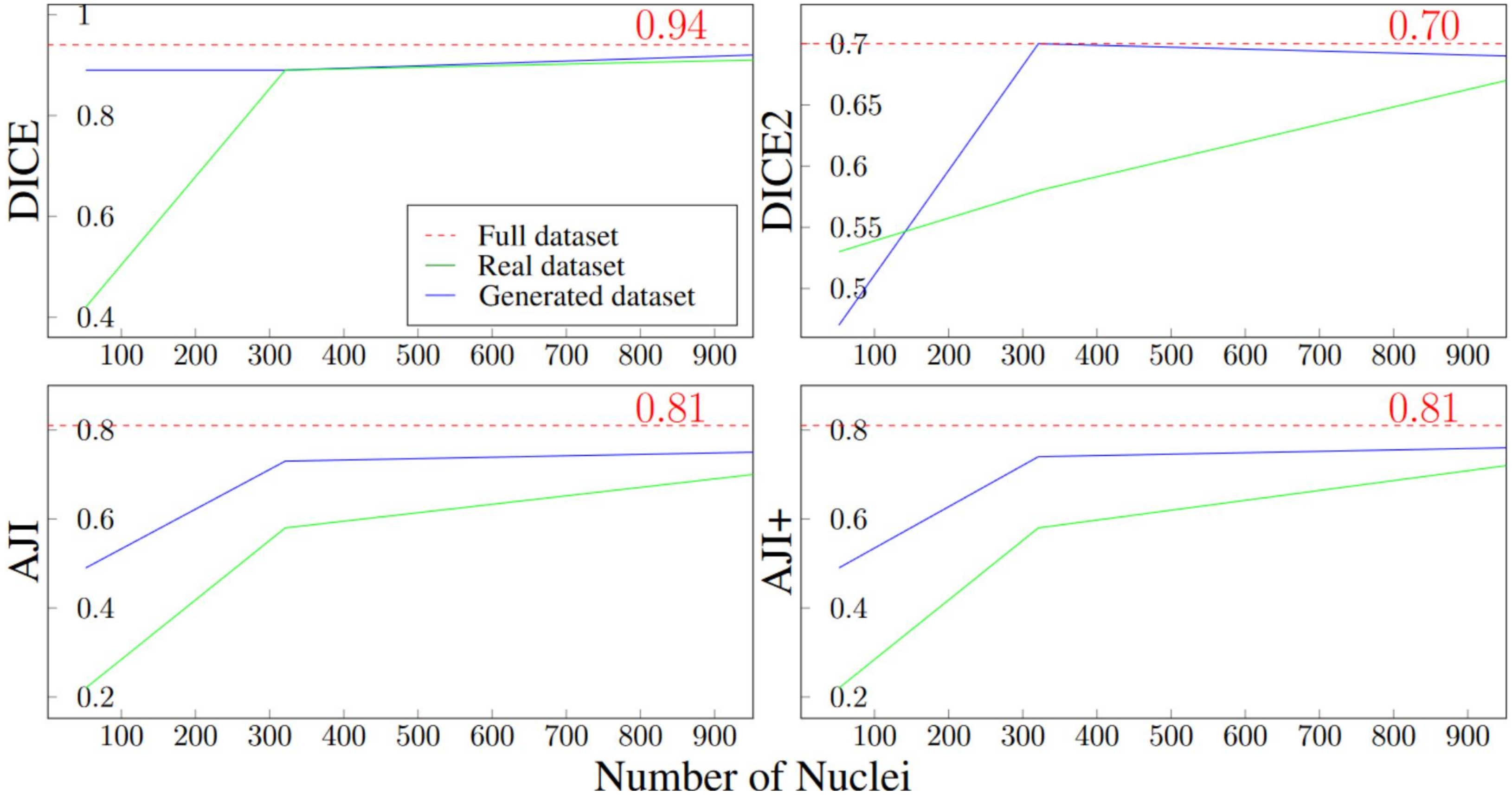




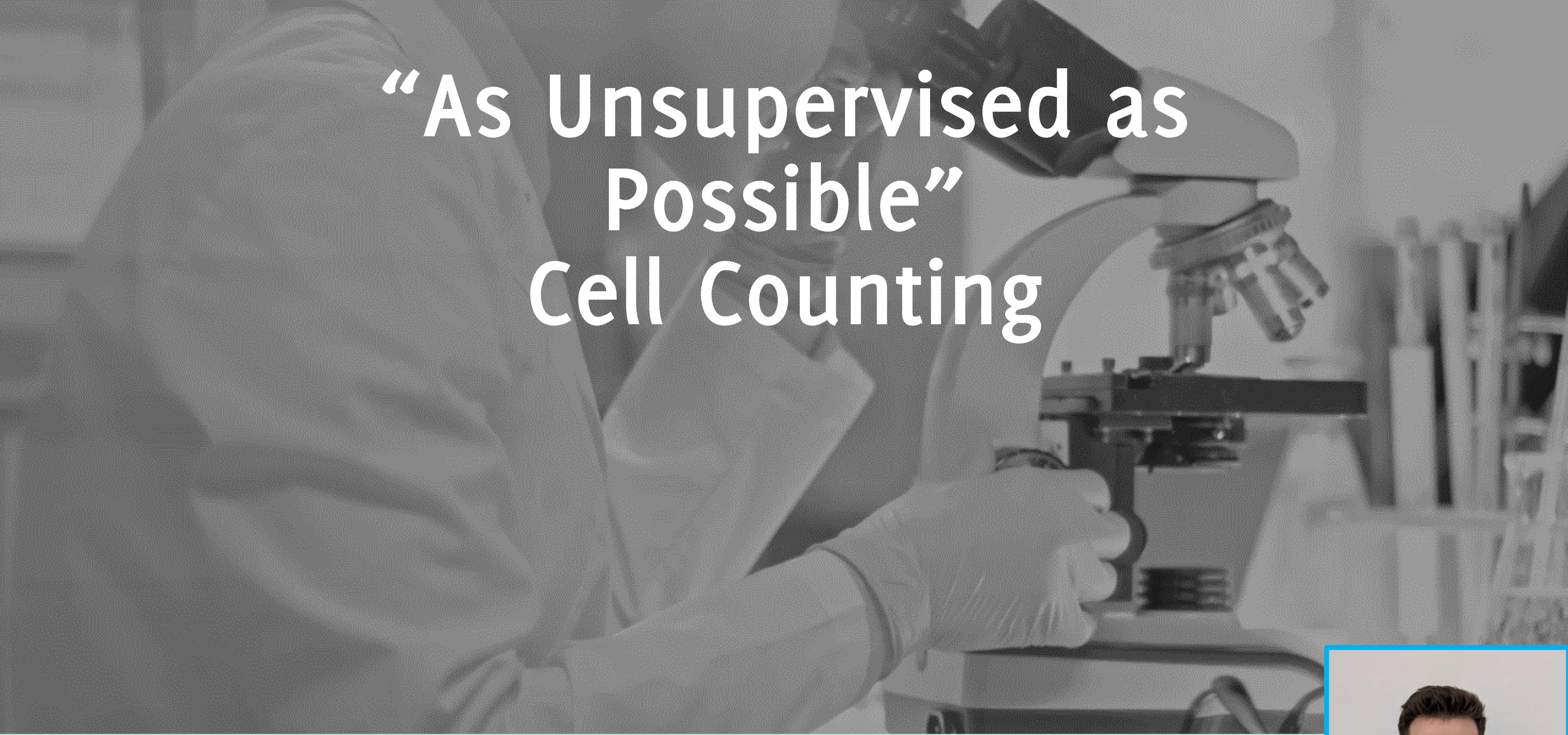


Results

A State-of-the-Art Instance Segmentation model can increase its performance in very low-data regimes (down to 1 annotated image with ~30 nuclei). Approaching the performance of models trained on a whole dataset (~700 images).



“As Unsupervised as Possible” Cell Counting



Luca Alessandrini

Context – Cell Counting

Need of precise density estimation

Goal:

- Count the number of cells, to infer the density, and segment them, to infer properties

Challenges:

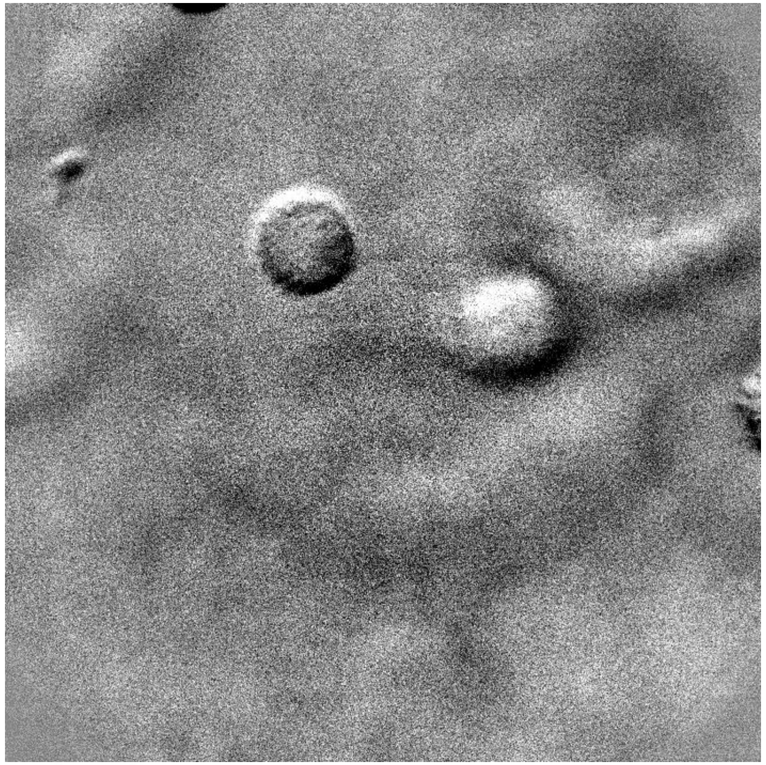
- No segmentation annotations provided
- Image at different densities may exhibit different properties
  - Illumination
  - Scattering

Context – Cell Counting

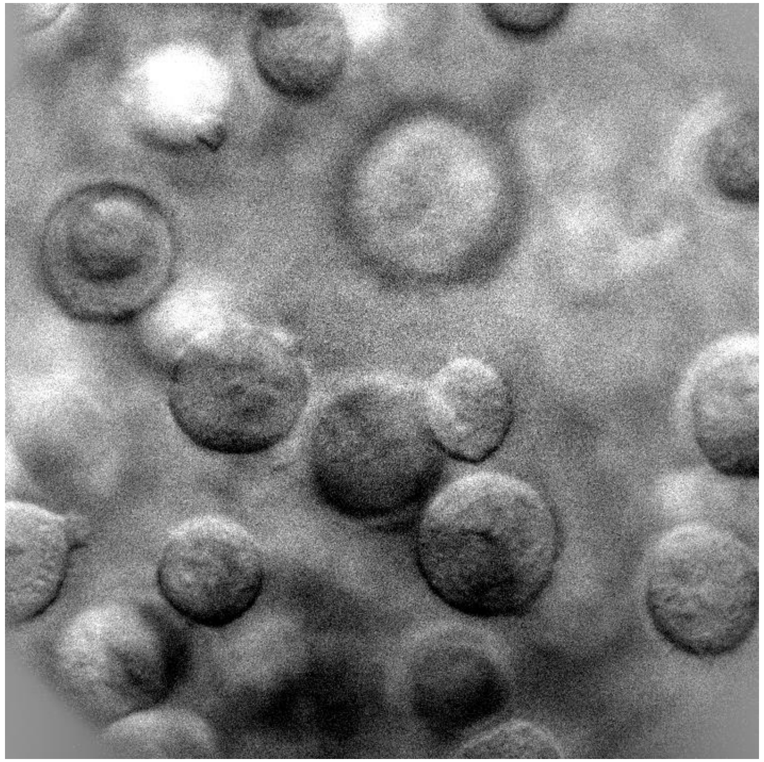
Cells have a completely different appearance than the Ikonisys ones, these are «farmed» in a bioreactor

Example at three density levels:

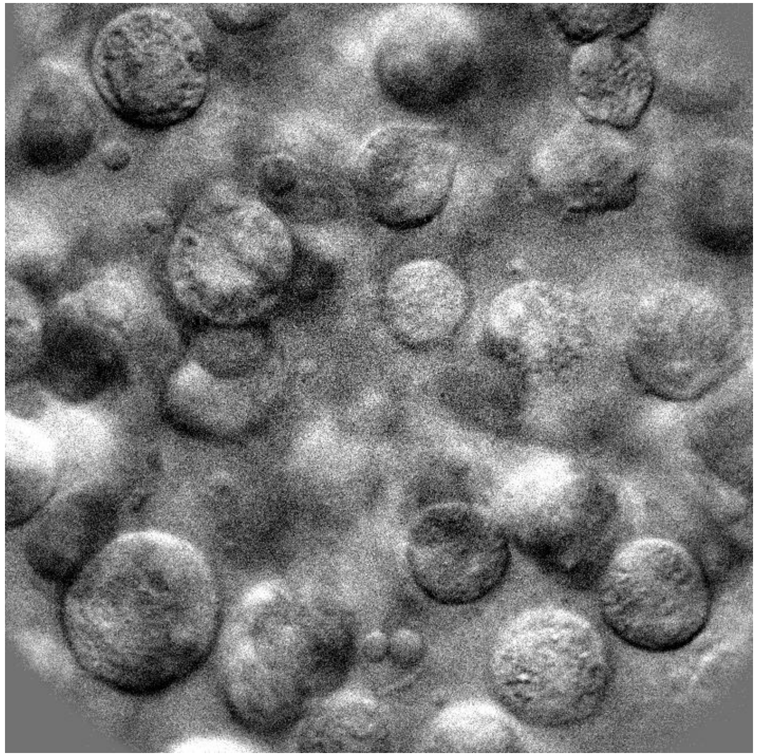
Low



Medium



High

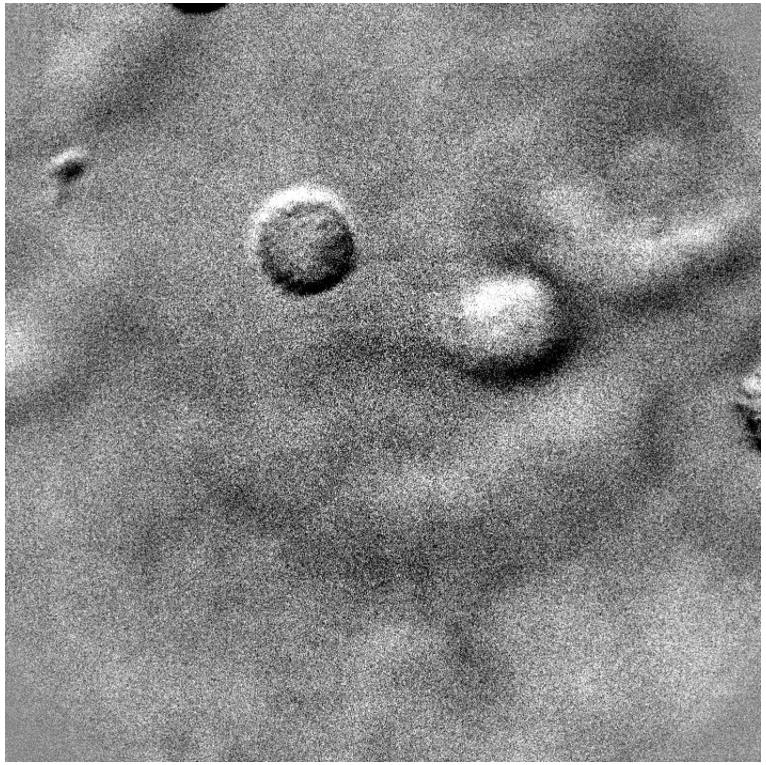


Research Directions

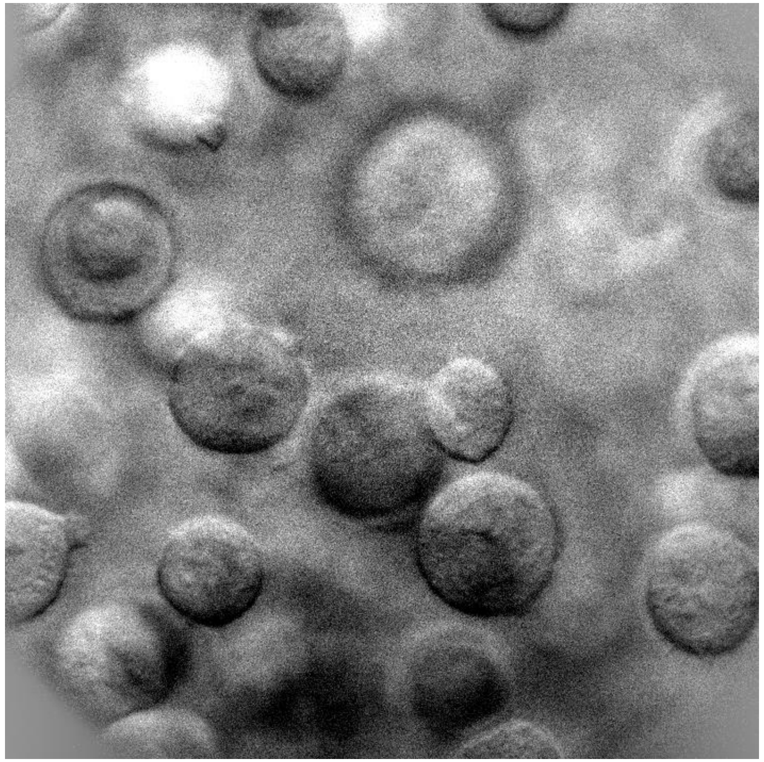
Develop weakly-supervised segmentation tools on low densities and use this for fine tuning a detection network as the density grow.

Still work in progress, hopefully to be published

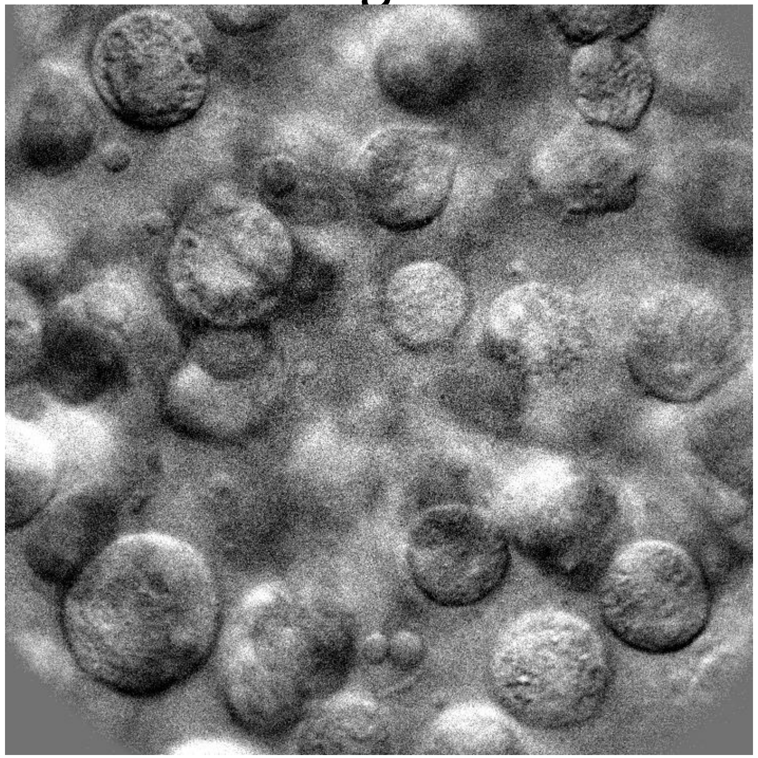
Low



Medium



High

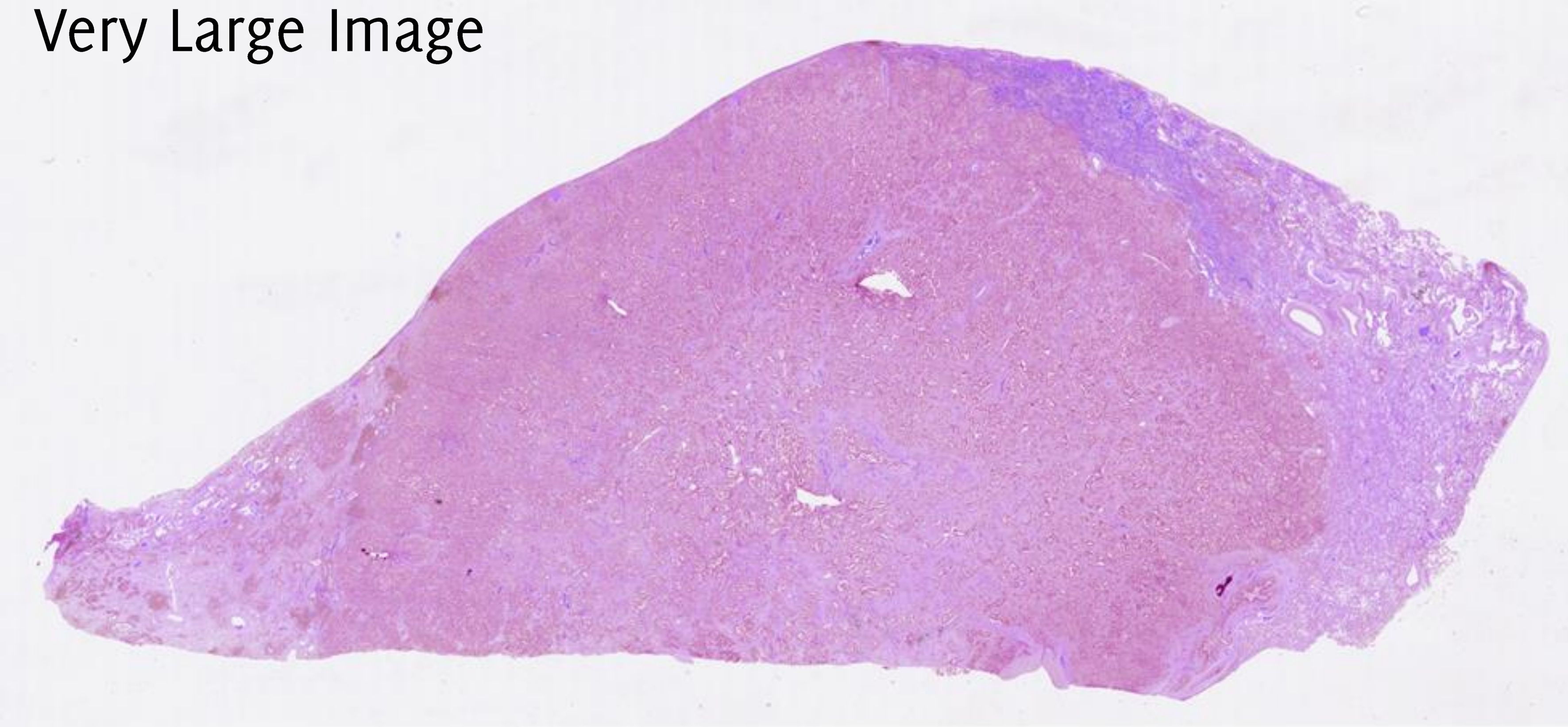


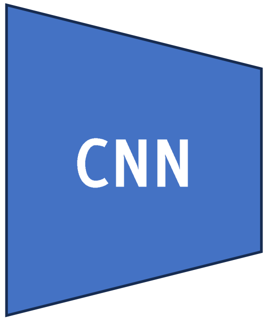
Multiple Instance Learning in Digital Pathology

FONDAZIONE IRCCS
ISTITUTO NAZIONALE
DEI TUMORI

MIL for Digital Pathology

- Whole Slide Images (WSIs) are extremely large and usually come with **slide-level labels** (e.g., responder vs. non-responder)





Responder/
Non responder

Life expectation

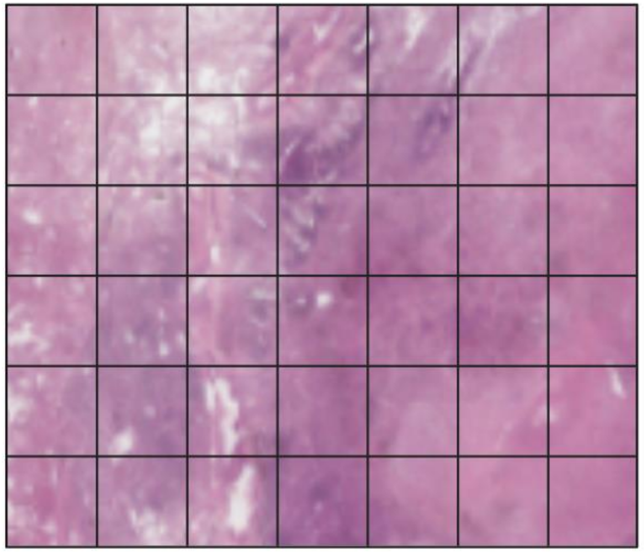
Challenges in Digital Pathology

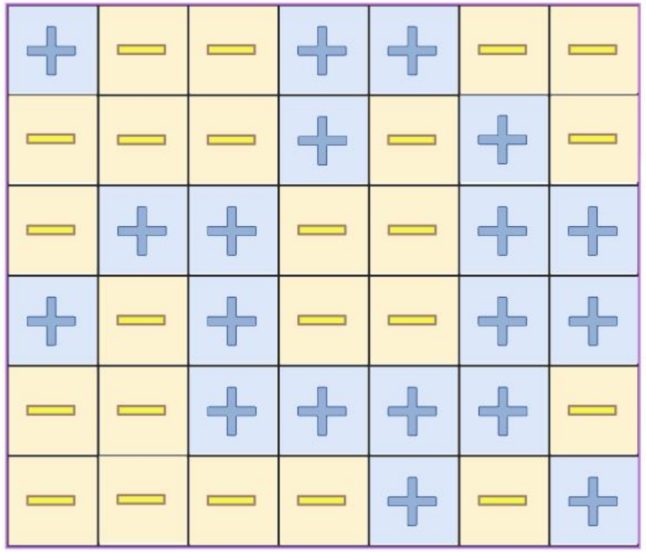
Challenges:

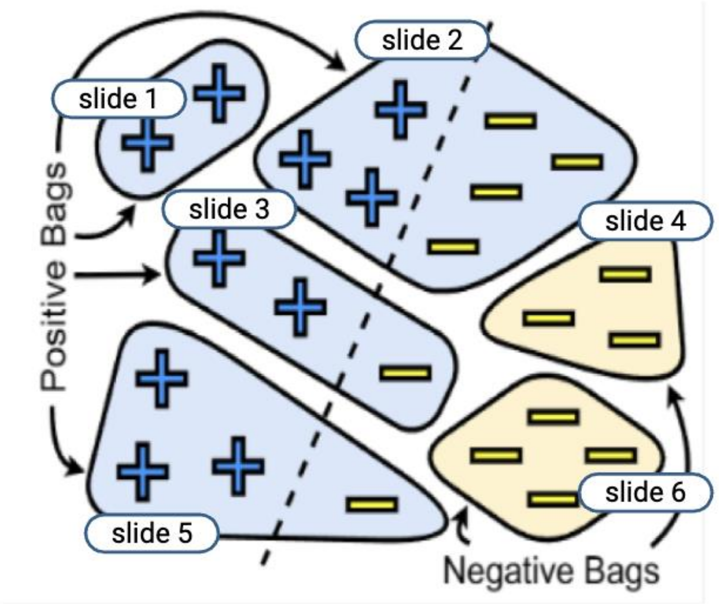
- Images are very large
- Little supervision
- It is often not possible to map to WSI labels (set labels) to portions ot the image (instance labels)
- Set-learning settings, but set-learning models are meant for way smaller inputs

MIL for Digital Pathology

- Whole Slide Images (WSIs) are extremely large and usually come with **slide-level labels** (e.g., responder vs. non-responder)
- WSIs are divided into smaller image **tiles** (patches), which are treated as ***instances*** within a ***bag*** (**the WSI**)

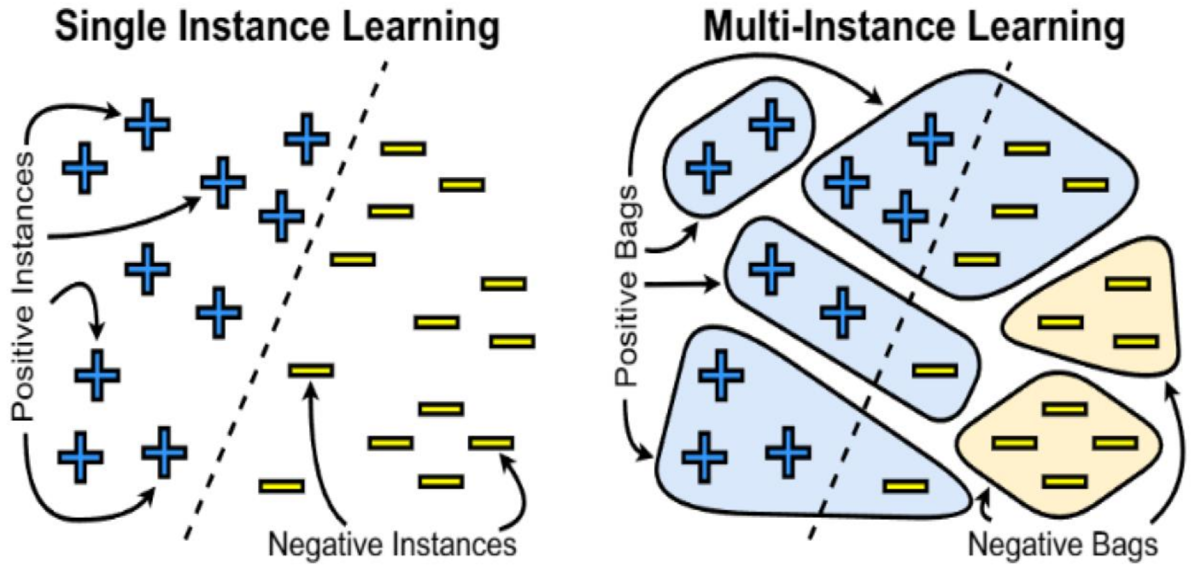






Multiple Instance Learning (MIL)

- A weakly supervised learning approach where data is grouped into bags of instances
- Labels are provided only at the bag level, not for individual instances
- A bag is labelled positive if at least one instance is positive; otherwise, it is labelled negative
- **Instance-level labelling is not done manually** - the model infers which instances are positive or negative during training



Concluding Remaks

Conclusions

Automated tools for processing biomedical images calls for Computer Vision methods that:

- Solve different problems than natural images (segmentation, MIL, way more popular here!)
- Face severe shortage of annotation
- Face severe domain shift problems on very specific tasks

Different strategies to counteracting lack of annotation:

- Sparse supervision
- Multi-Task Learning (auxiliary problems)
- (Controlled) data generation
- … and of course leveraging task-agnostic self-supervised vision encoders.

Chen, R.J., Ding, T., Lu, M.Y. et al. Towards a general-purpose foundation model for computational pathology. Nat Med 30, 850–862 (2024)