

*Seminar at Mila, Université de Montréal
November 14th, 2018, Montreal, Canada*

Connecting MRI physics and A.I. to advance neuroimaging

Julien Cohen-Adad, PhD

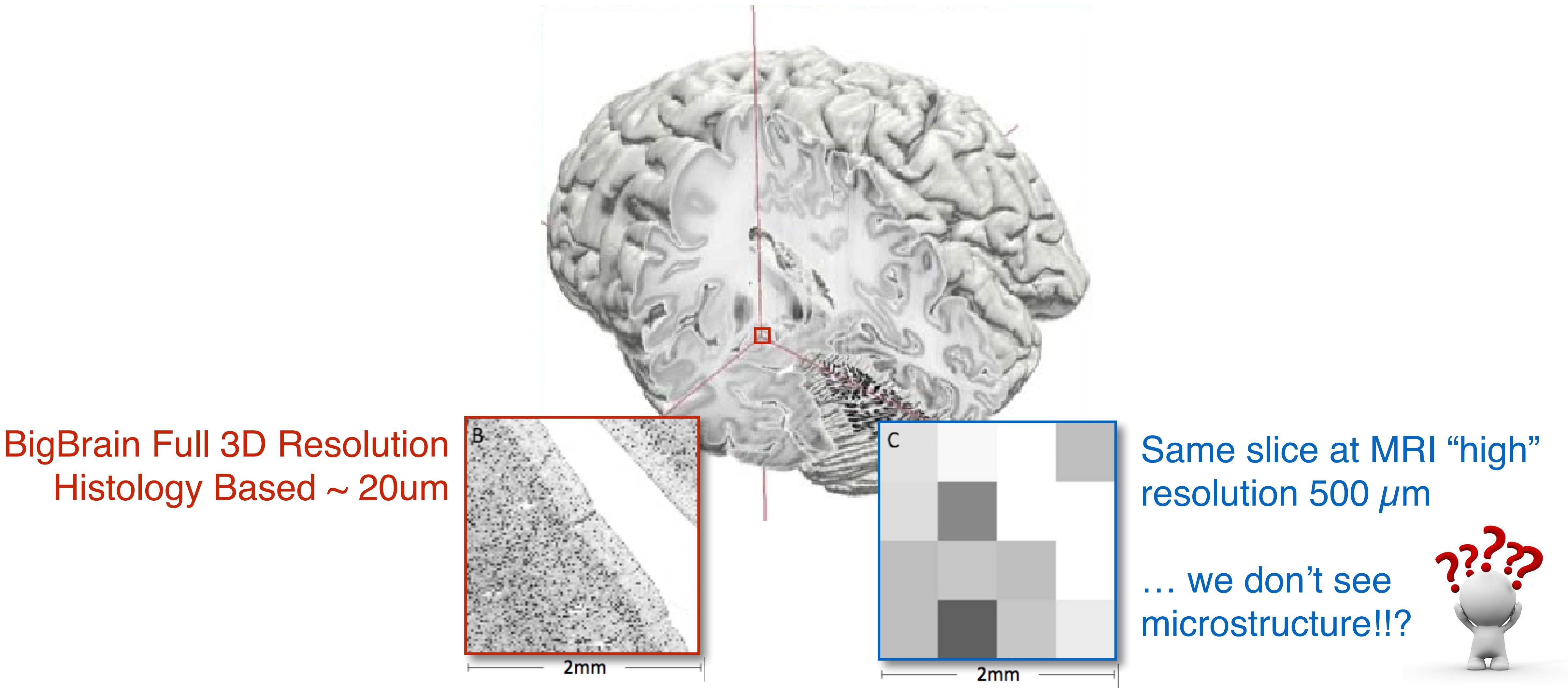
Associate Professor, Ecole Polytechnique de Montreal

Associate Director, Functional Neuroimaging Unit, University of Montreal

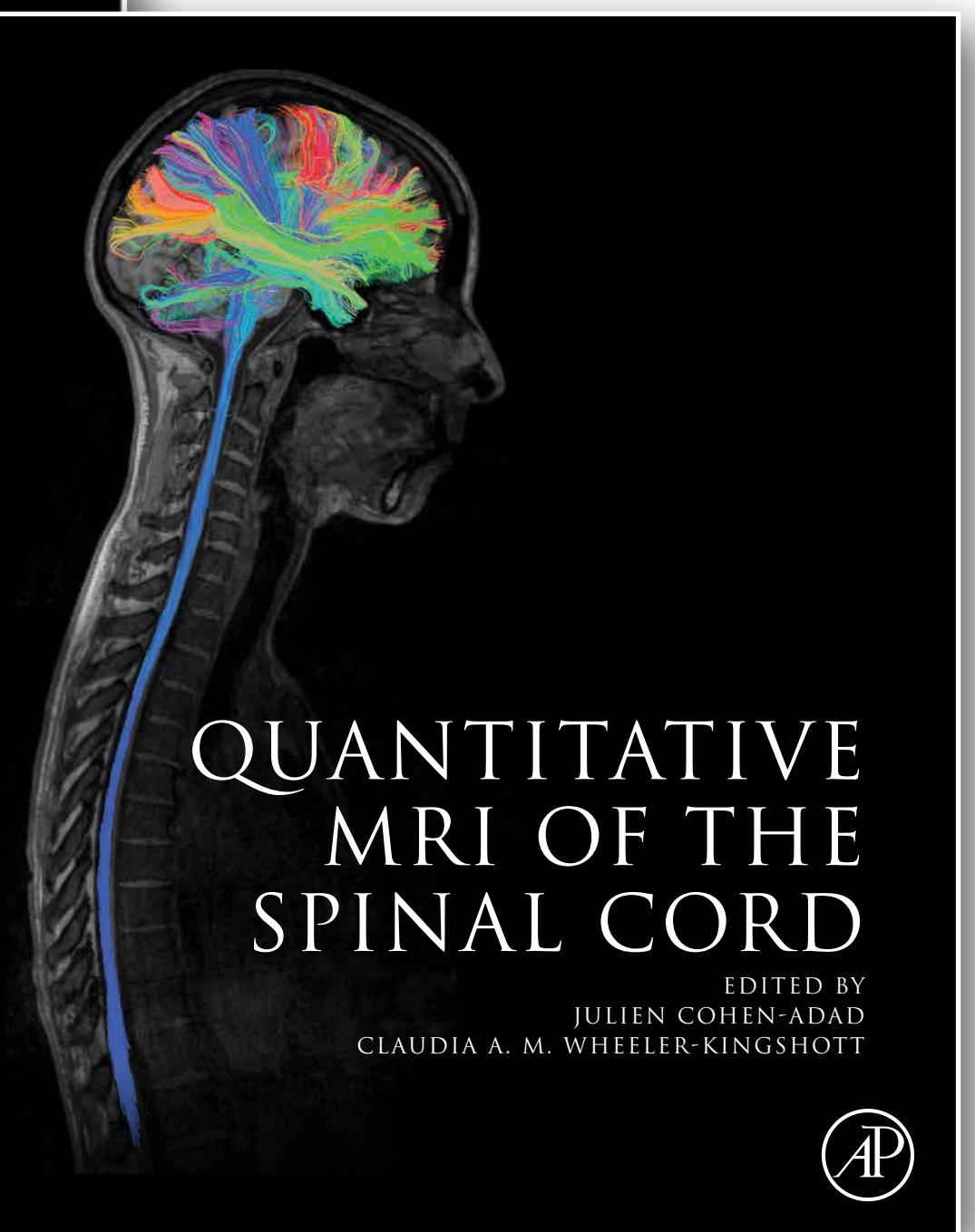
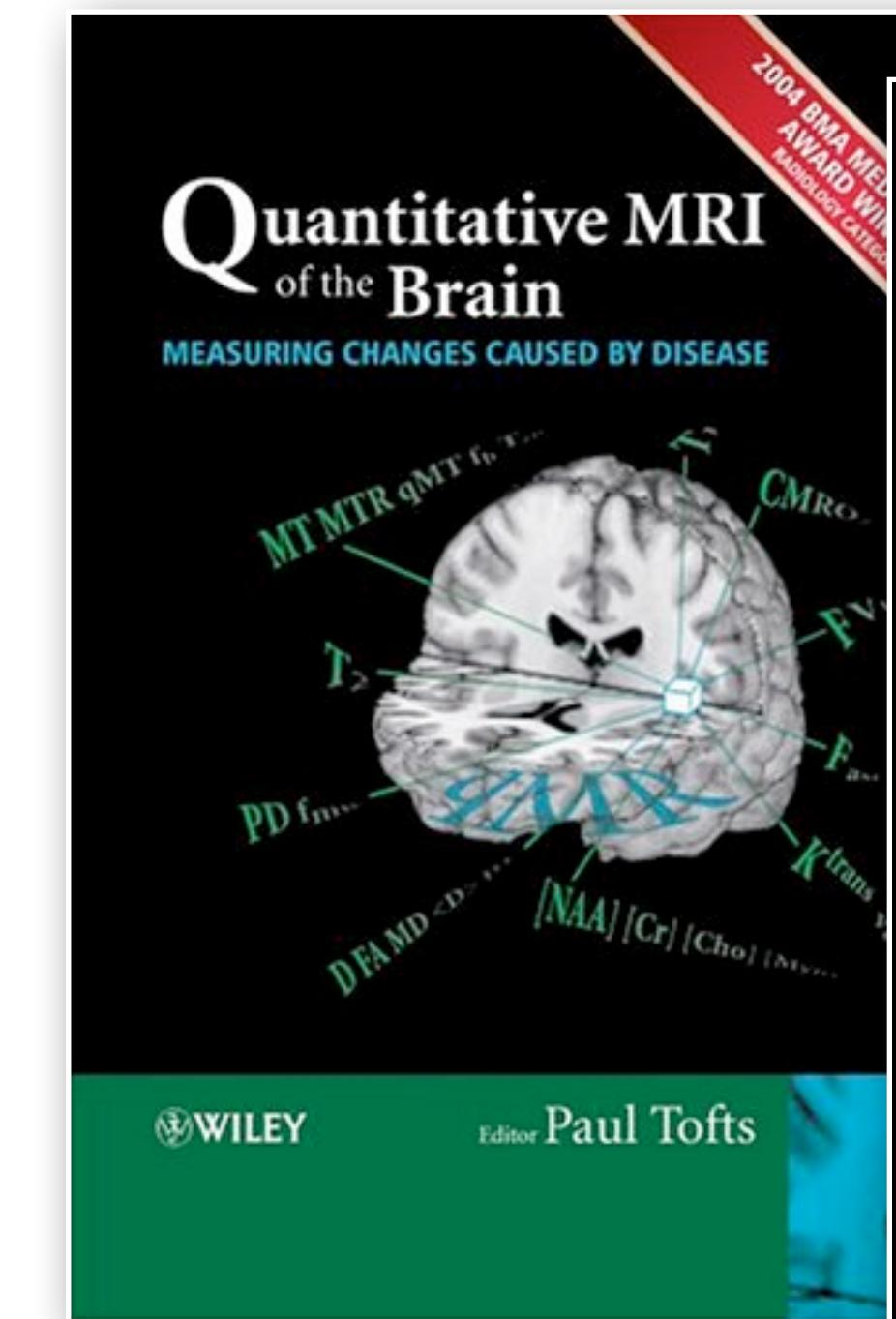
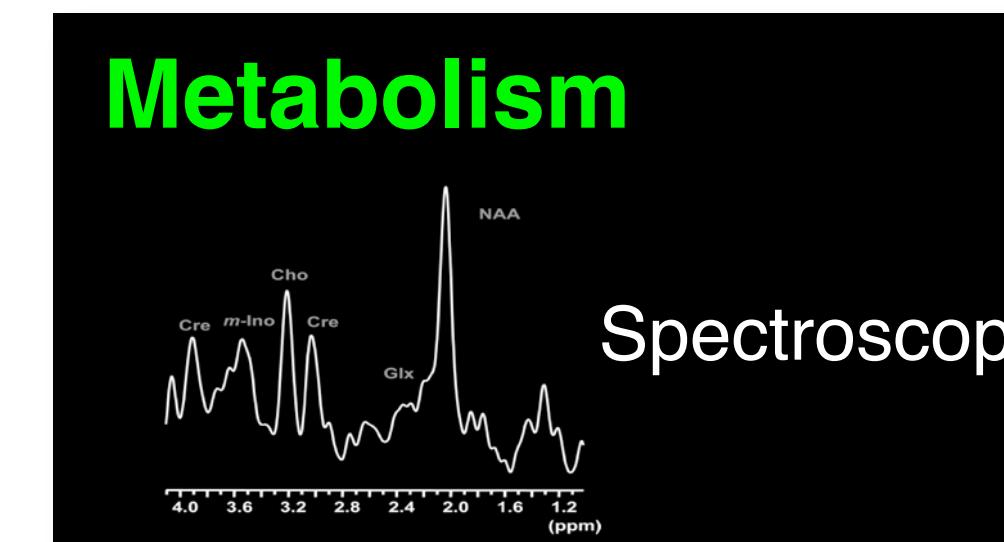
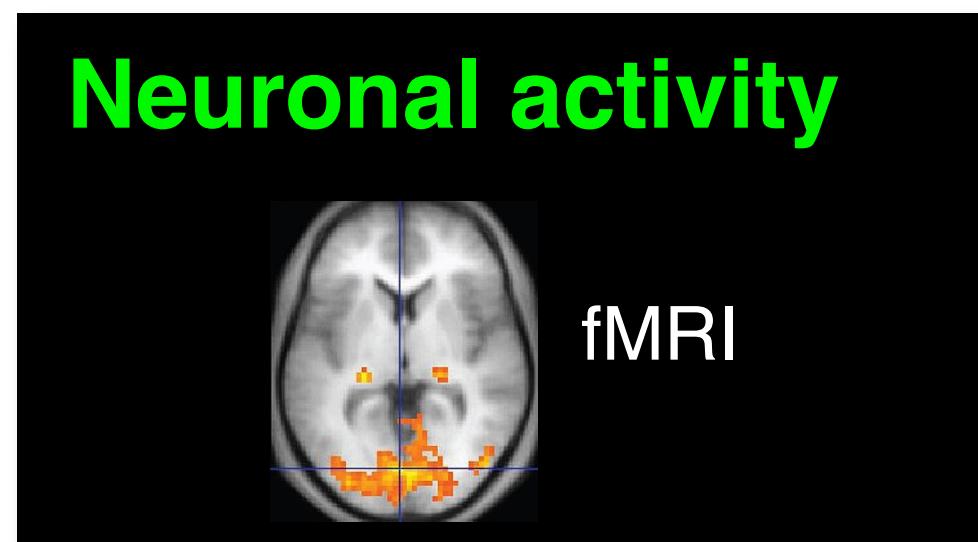
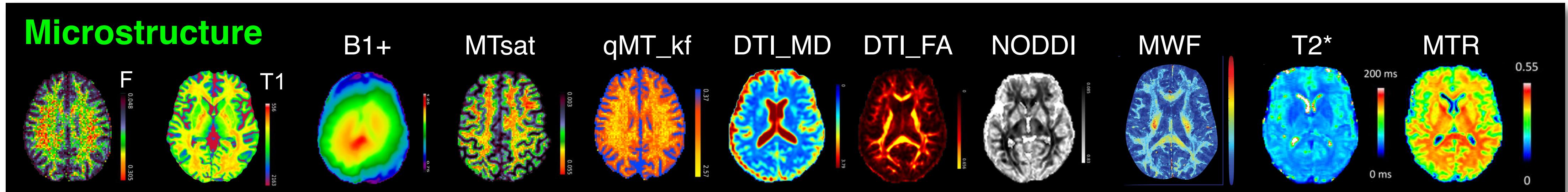
Canada Research Chair in Quantitative Magnetic Resonance Imaging



In vivo histology with MRI: Is it possible?



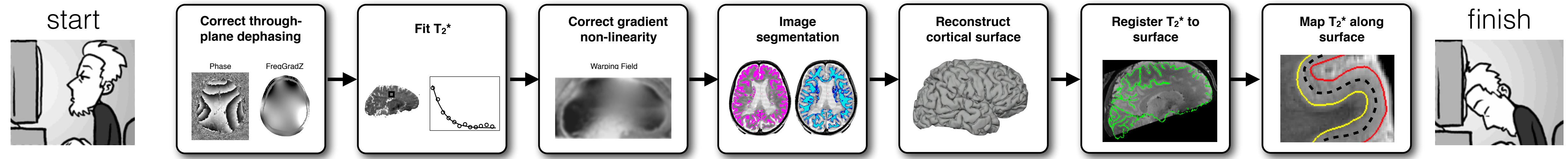
Quantitative MRI is an “In vivo microscope”



What is qMRI used for?

- Diagnosis/prognosis traumas, neurological diseases, cancers
- Objective biomarkers for testing new drugs

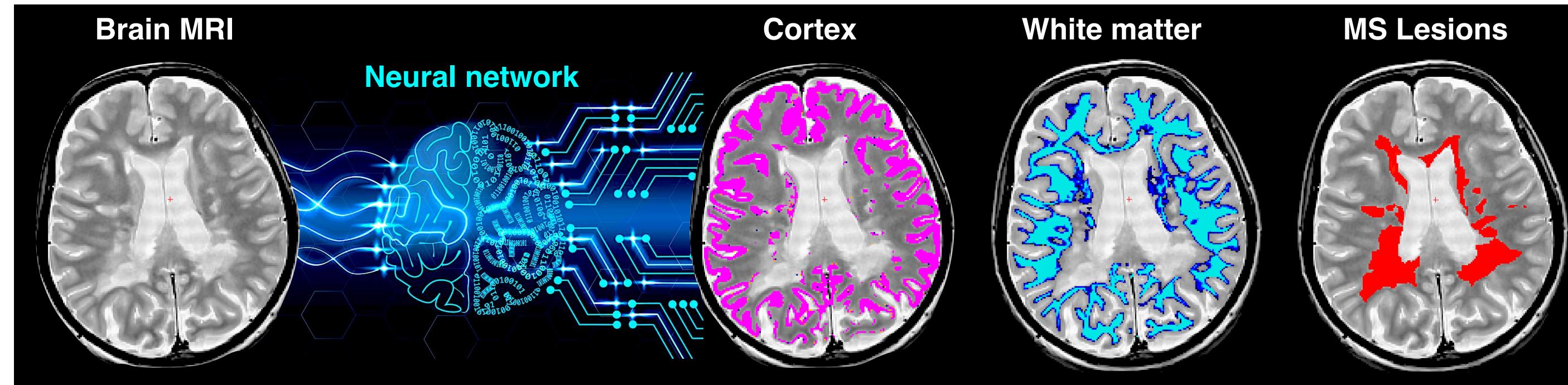
Problem: Data processing is long and tedious



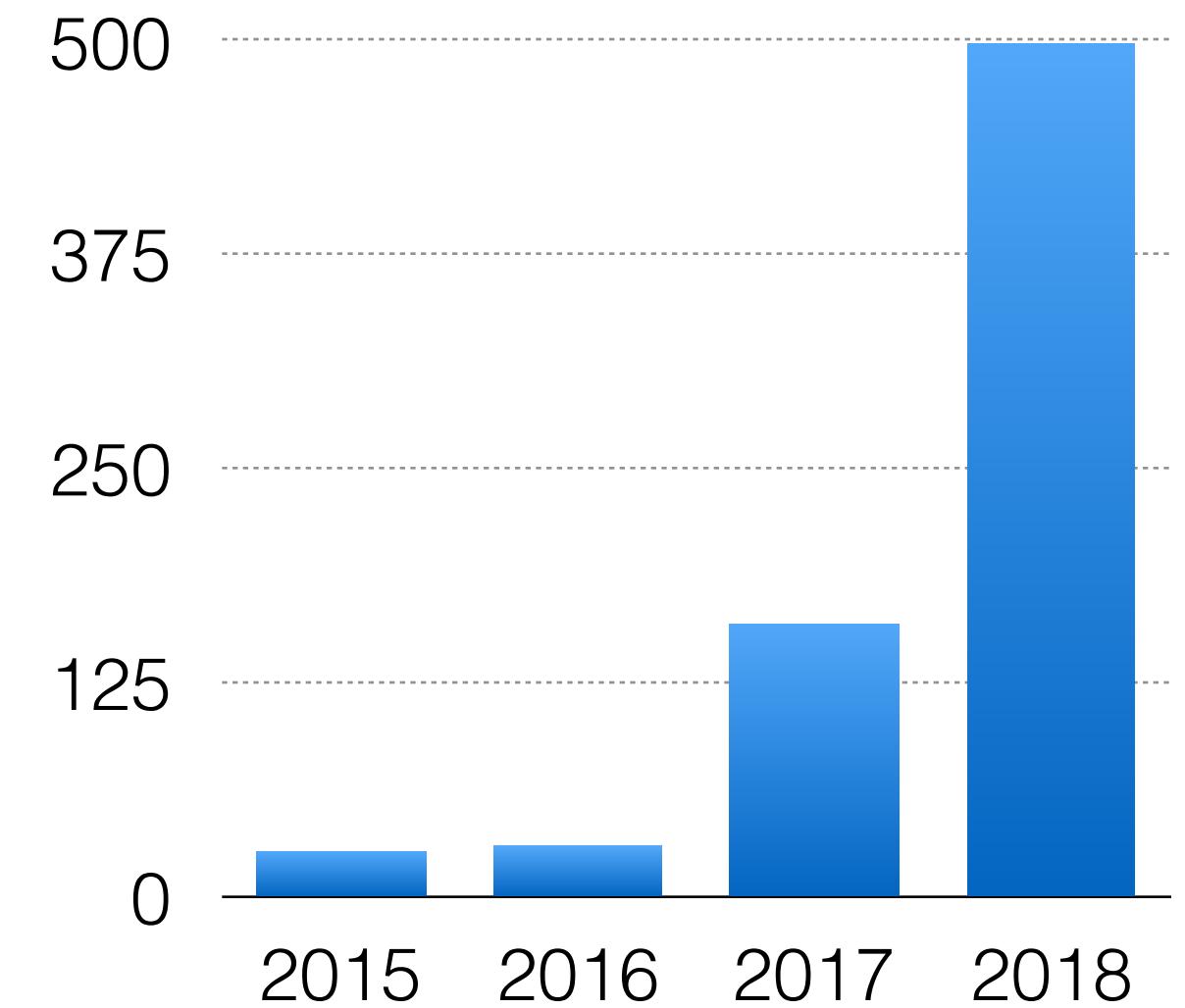
- However, quantitative MRI data require complex analysis pipelines that are often **executed manually** and hence suffer from **poor reproducibility**.
- Lack of replicability and transparency contributes to the “**neuroskeptic**” trends [1]

Deep learning to the rescue

- DL can be leveraged to achieve key tasks:
 - Segmentation of healthy structures (cortex, spinal cord, etc.)
 - Segmentation of pathological structures (lesions, tumors, etc.)
 - Categorization (diagnosis, prognosis)
 - Identify novel biomarkers

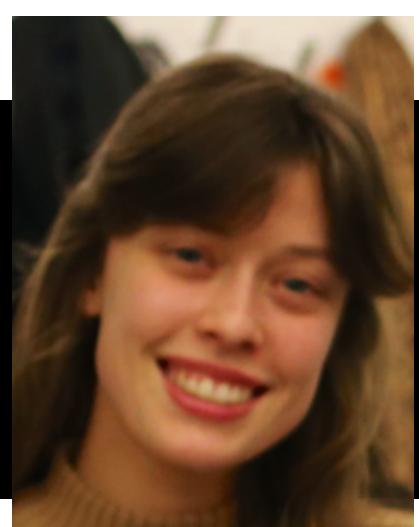


Number of abstracts on Deep Learning at the main MRI conference (ISMRM)



Examples of ongoing projects in our lab

Gray matter atrophy in ALS



Marie-Eve
Paquin

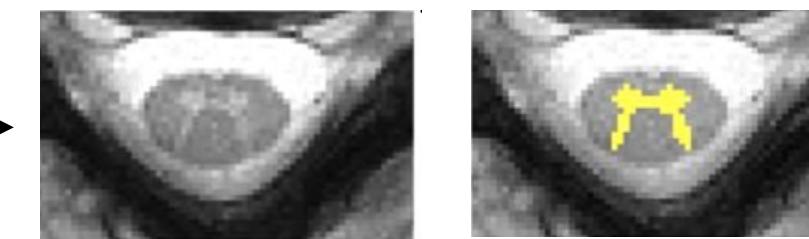
Charley
Gros



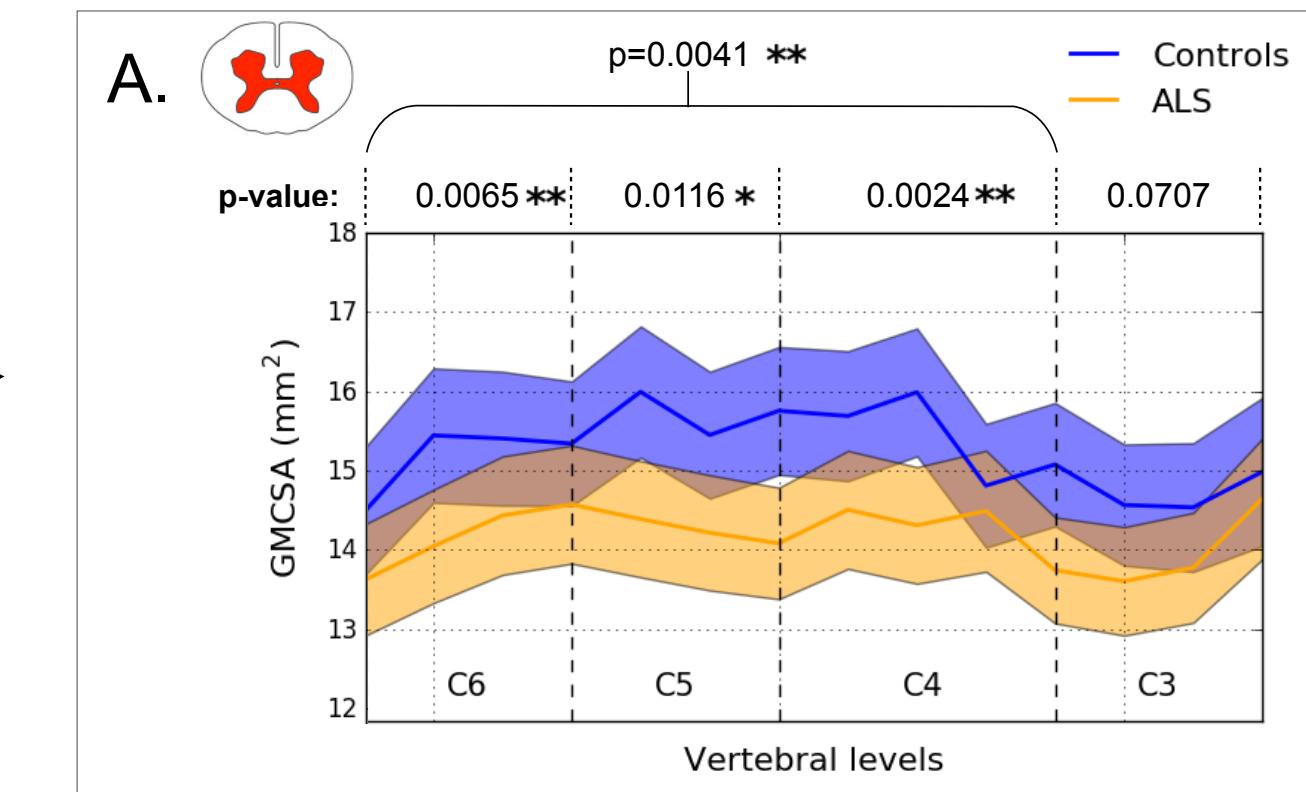
We need better biomarkers of motoneuron degeneration for earlier diagnosis and treatment monitoring



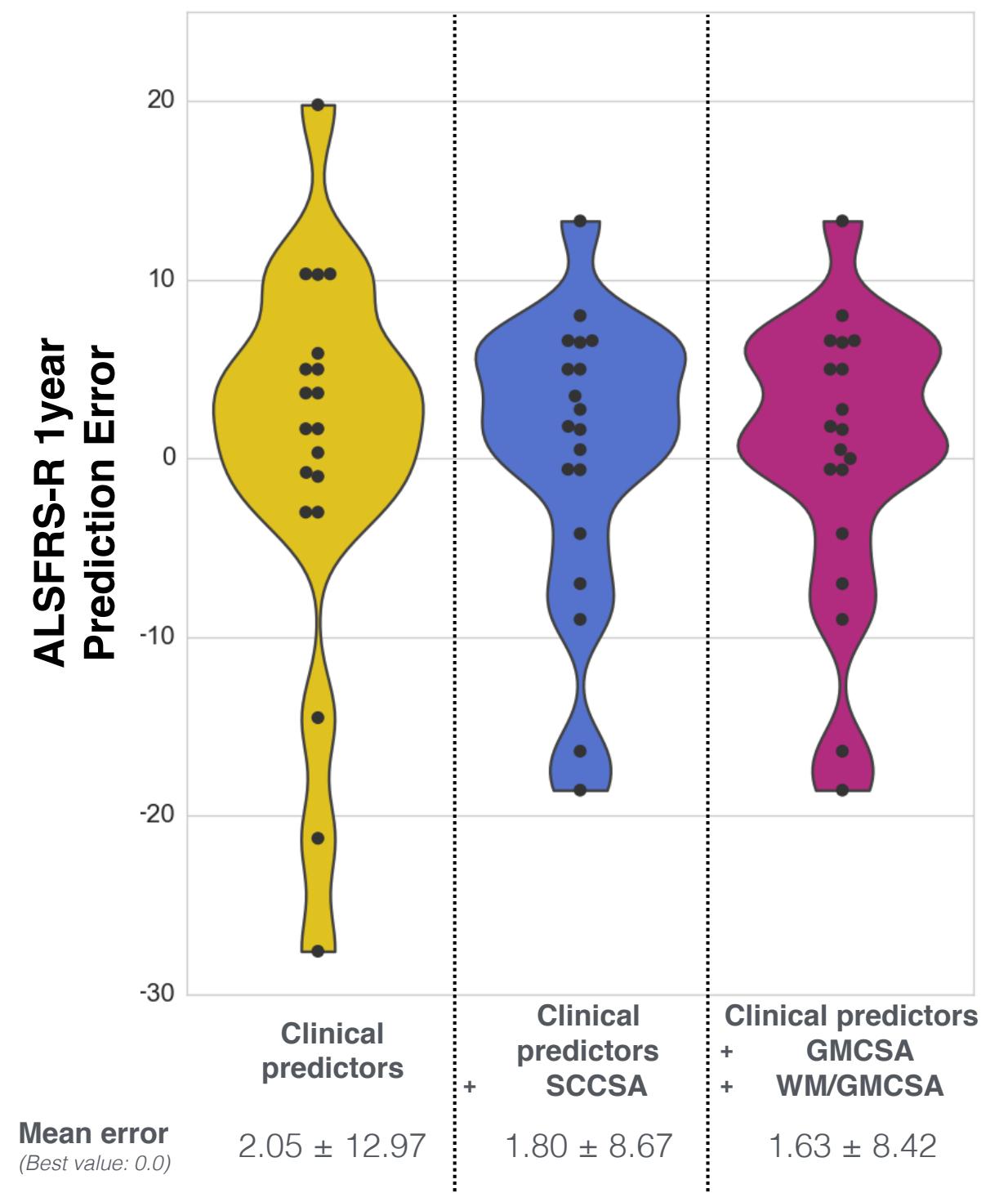
Spinal cord Gray Matter segmentation



GM atrophy is a good biomarker for ALS diagnosis



Prediction at 1 year

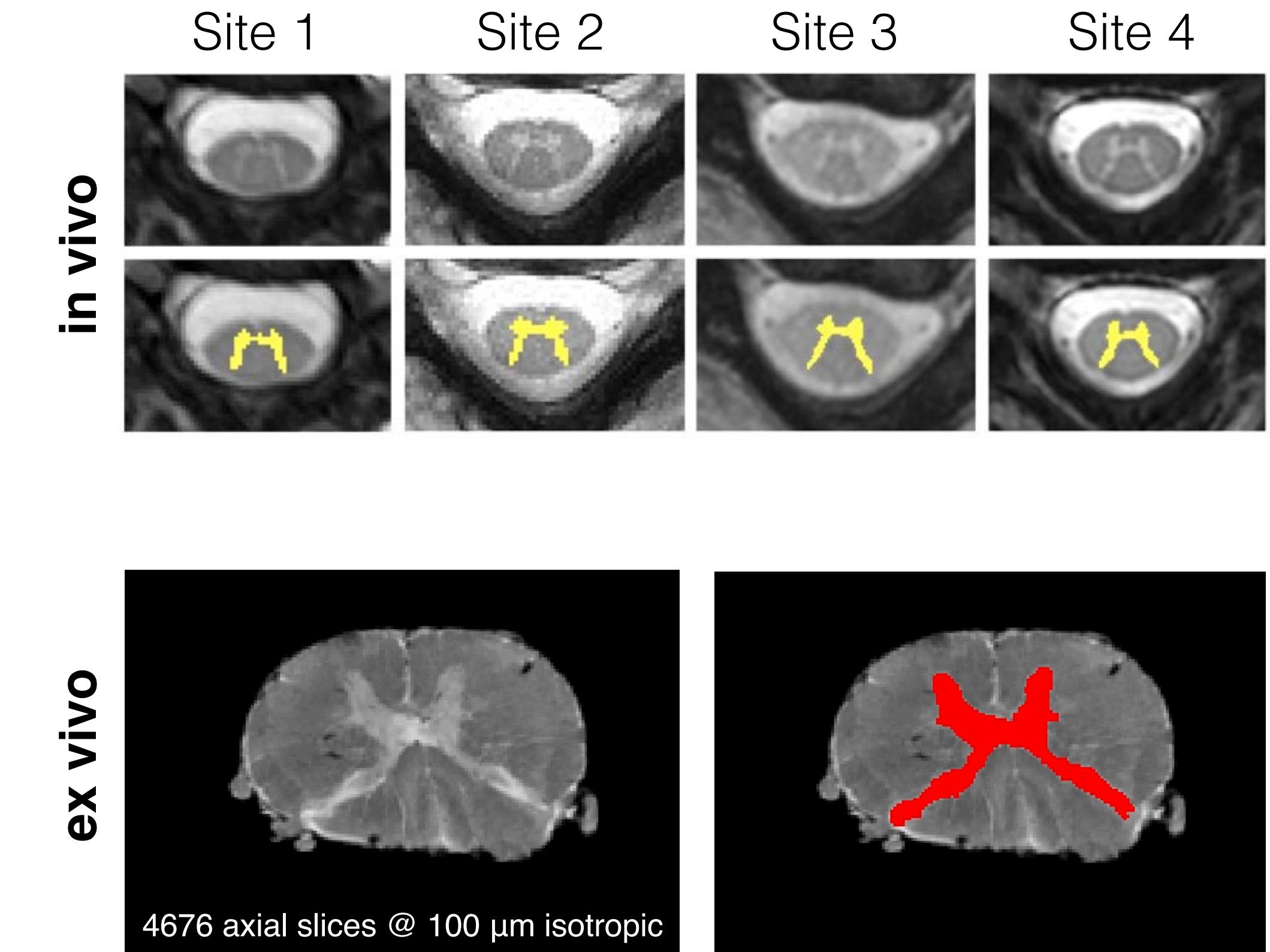
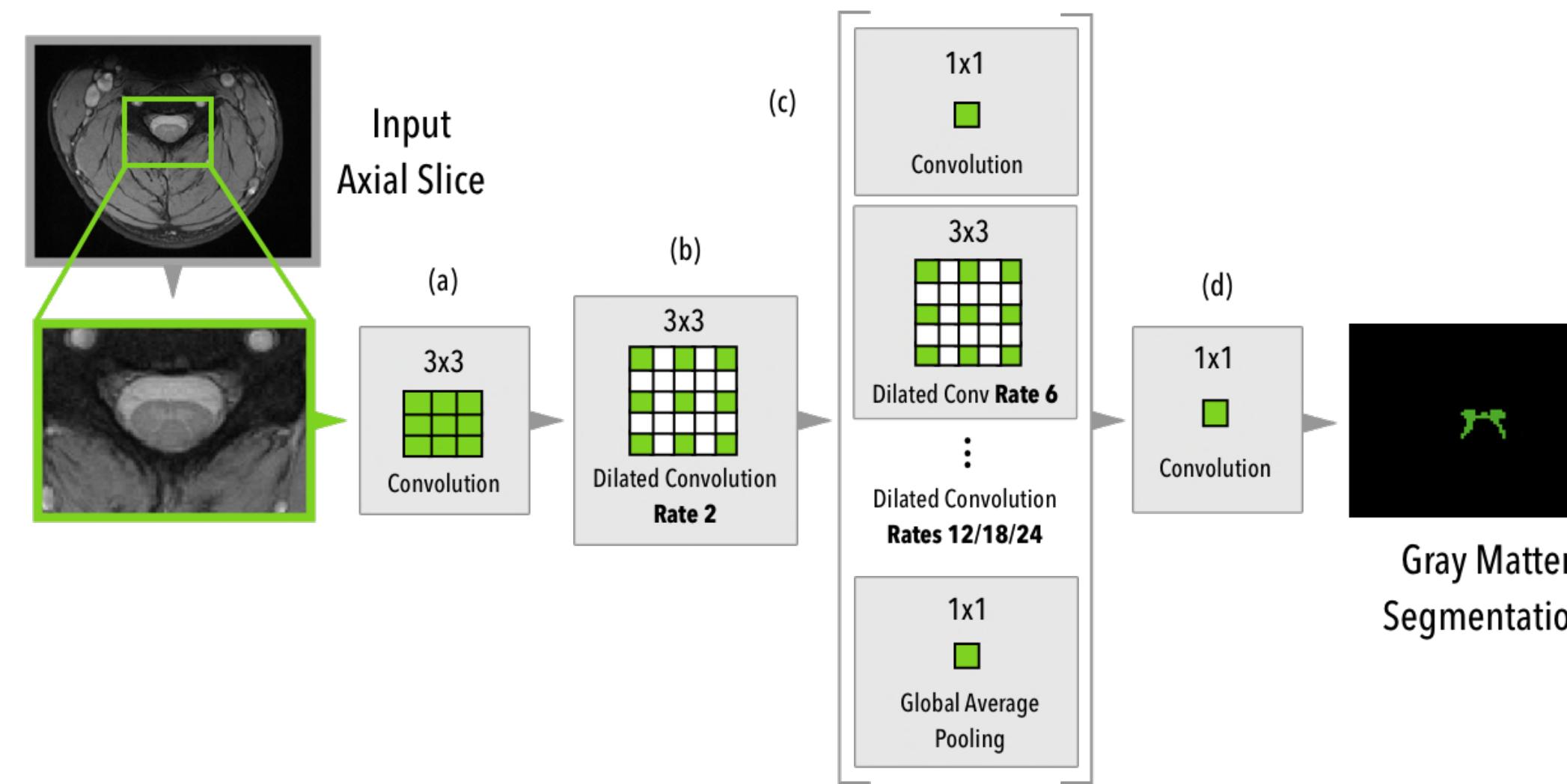


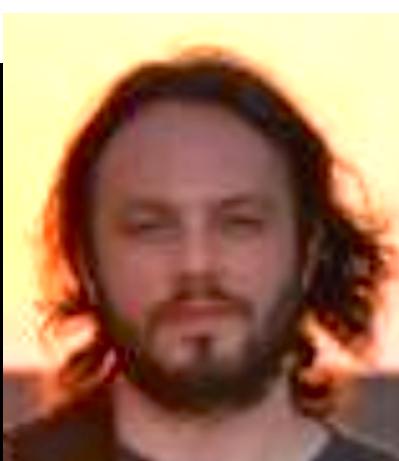
Spinal cord gray matter segmentation using deep dilated convolutions [1]



Christian
Perone

- **Context:** U-Net popular for segmentation. However, most of the information is contained in low-level layers.
- **Idea:** Atrous Spatial Pyramid Pooling [2] add holes in convolution kernels, providing parameter reduction while increasing the effective receptive field.
- **Methods:** Data from GM challenge [3], Training (N=40), Testing (N=40)

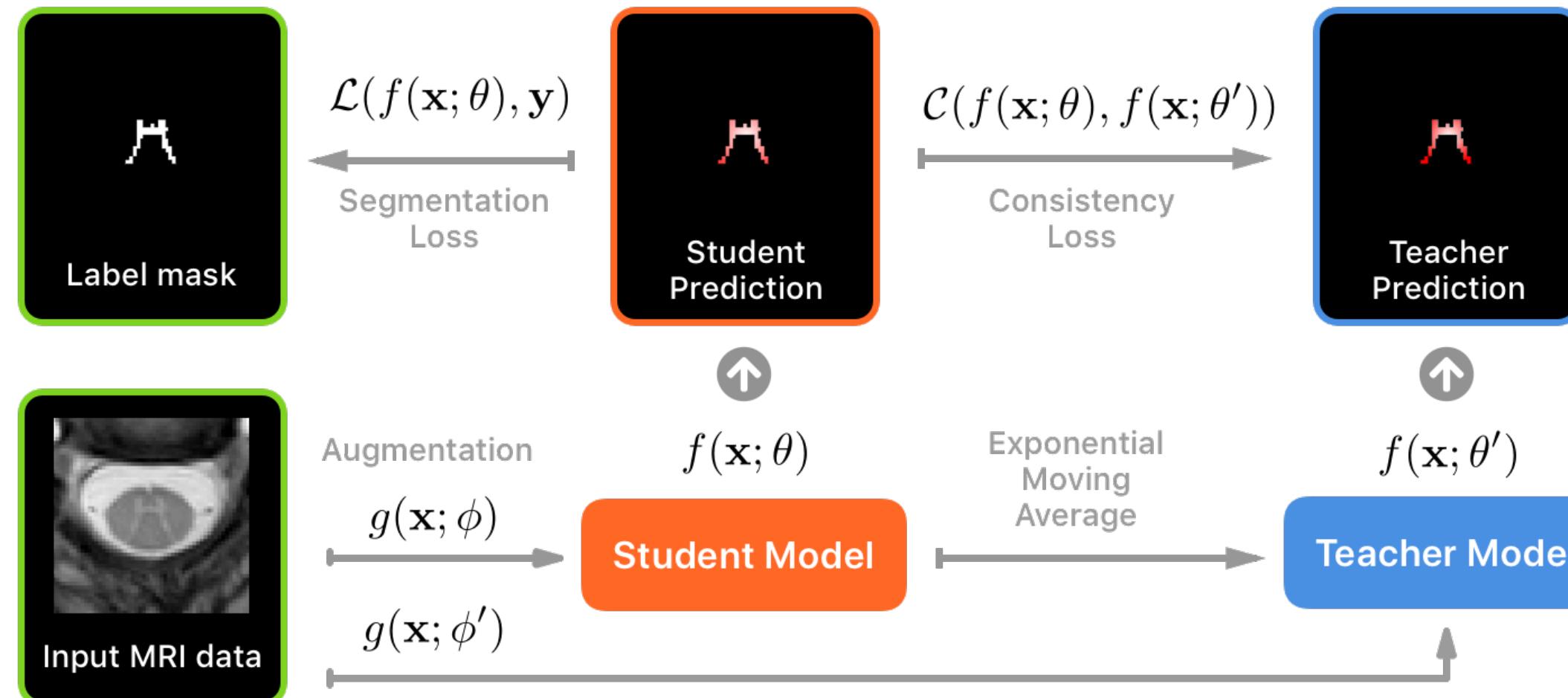




Deep semi-supervised segmentation with weight-averaged consistency targets [1]

Christian
Perone

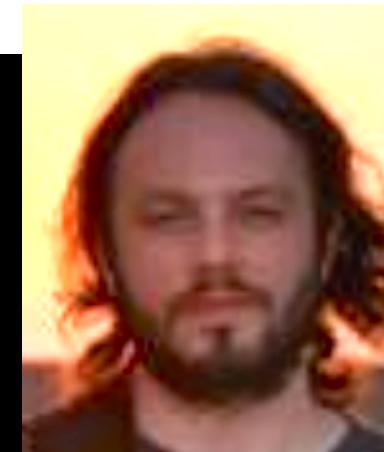
- **Context:** We have tons of data, but we lack annotations/labeling
- **Idea:** Extend semi-supervised learning with *Mean Teacher* method [2] to segmentation task
- **Methods:** Mean Teacher with weights obtained from exponential moving average
 - **Data:** GM challenge [3]. Labeled data: *Training*: 8 subjects; *Validation*: 8 subjects ; Unlabeled data: 40 subjects
 - U-Net (15 layers); Data augmentation: applied *before* Student model (on image) and *after* Teacher prediction (on segmentation)



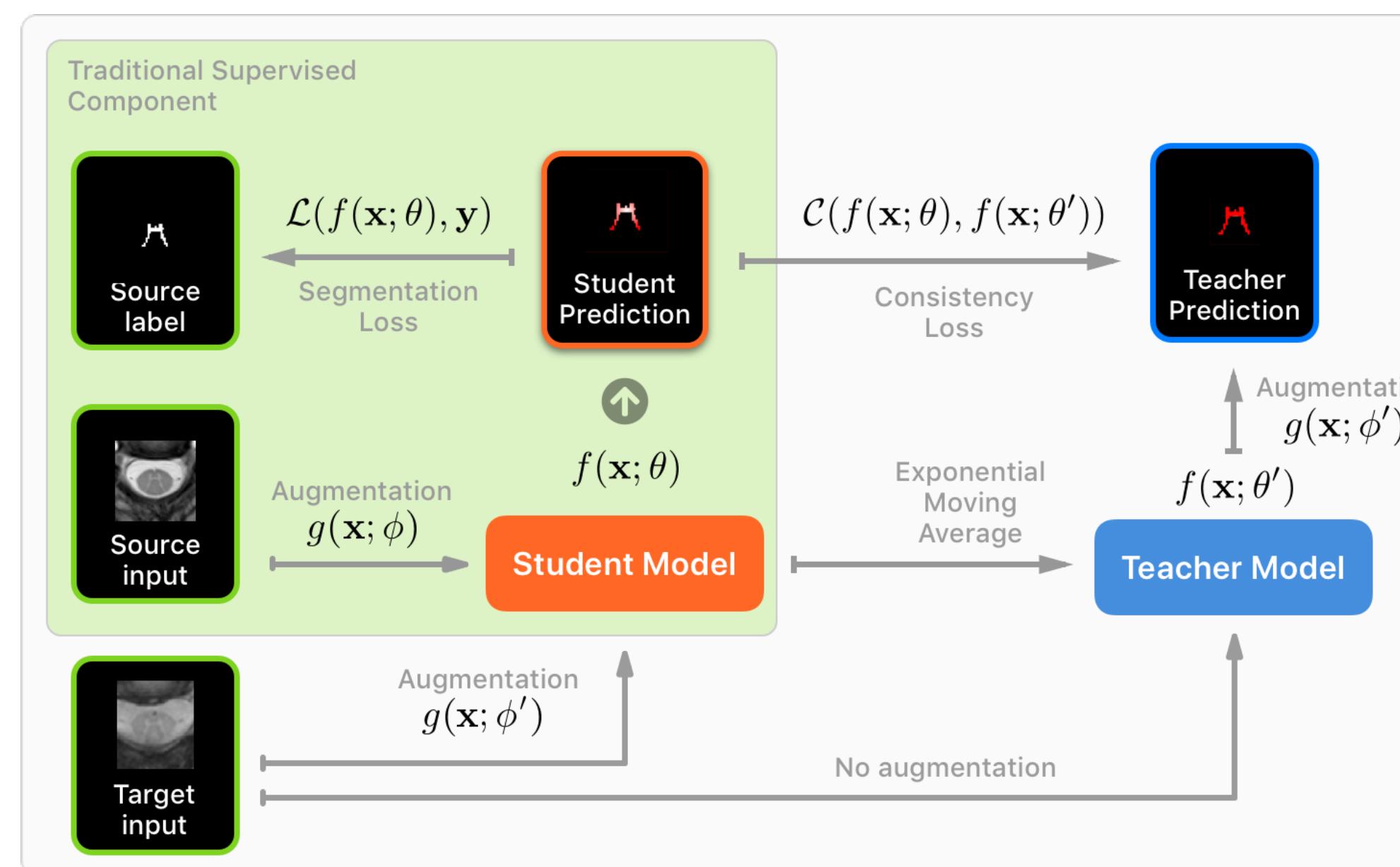
Results: Overall better performance for semi-supervised approach

	Dice	mIoU	Accuracy	Precision	Recall	Specificity
Supervised	67.915 (0.313)	53.679 (0.327)	99.745 (0.005)	57.948 (0.788)	92.495 (0.907)	99.775 (0.010)
Semi-supervised	70.209 (0.229)	55.509 (0.253)	99.792 (0.003)	64.732 (0.773)	86.112 (0.936)	99.846 (0.006)

Unsupervised domain adaptation for medical imaging segmentation with self-ensembling [1]

Christian
PeronePedro
Ballester

- **Context:** MRI data features vary across centers (contrast, resolution, etc.).
- **Idea:** Combine Mean Teacher method with unsupervised domain adaptation [2] for segmentation task.



Results: Specific improvement of the domain-adaptation approach (cannot be explained only by the EMA of unlabeled data).

Evaluation	Adaptation	Dice	mIoU	Recall	Precision	Specificity	Hausdorff
Center 3	Baseline	82.81 ± 0.33	71.05 ± 0.36	90.61 ± 0.63	77.09 ± 0.34	99.86 ± 0.0	2.14 ± 0.02
	Center 3	84.72 ± 0.18	73.67 ± 0.28	87.43 ± 1.90	83.17 ± 1.62	99.91 ± 0.01	2.01 ± 0.03
	Center 4	84.45 ± 0.14	73.30 ± 0.19	87.13 ± 1.77	82.92 ± 1.76	99.91 ± 0.01	2.02 ± 0.03
Center 4	Baseline	69.41 ± 0.27	53.89 ± 0.31	97.22 ± 0.11	54.95 ± 0.35	99.70 ± 0.00	2.50 ± 0.01
	Center 3	73.27 ± 1.29	58.50 ± 1.57	94.92 ± 1.48	60.93 ± 2.51	99.77 ± 0.03	2.36 ± 0.06
	Center 4	74.67 ± 1.03	60.22 ± 1.24	93.33 ± 1.96	63.62 ± 2.42	99.80 ± 0.02	2.29 ± 0.05



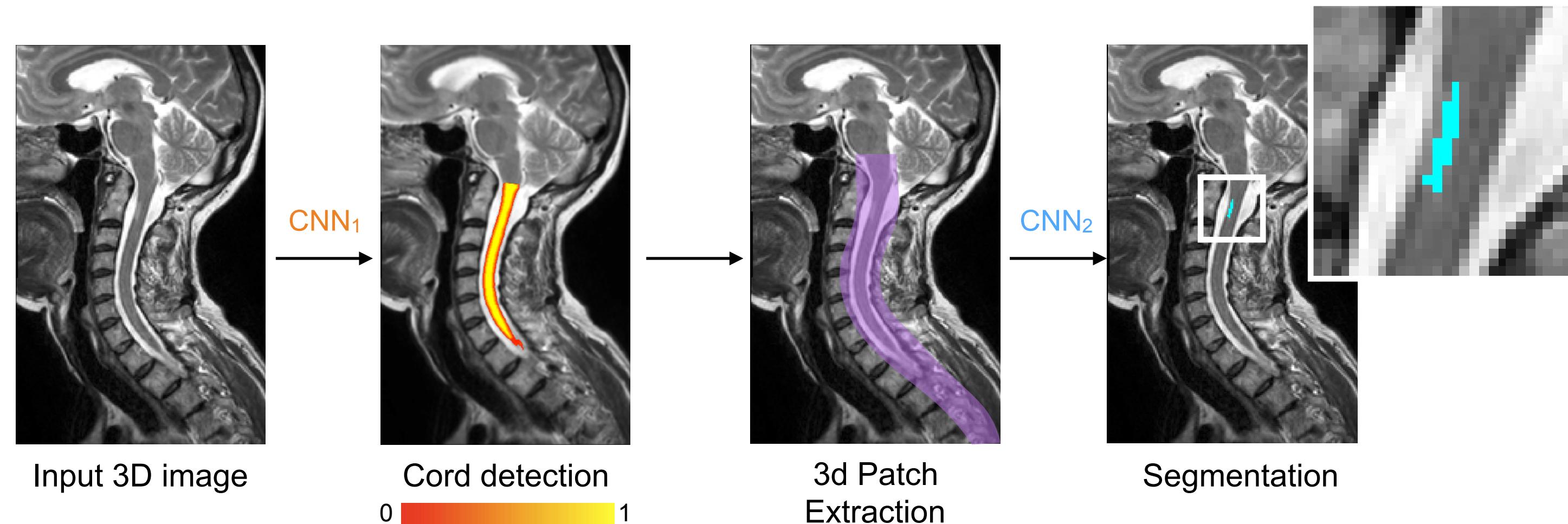
Segmentation of MS lesions with DL [1]

Charley Gros

- **Context:** Segmentation of MS lesions from MRI data. **Challenge:** large class imbalance.

- **Methods:**

- Cascade of Convolutional Neural Networks: **(1)** 2d-CNN₁ (cord detection with subsequence cropping to address class imbalance) and **(2)** 3d-CNN₂ (MS lesion segmentation).
- Implemented in Keras, available in SCT package [\[2\]](#)
- Data: N=750 patients from 12 MS clinical centers, manual labelling by 8 neuroradiologists



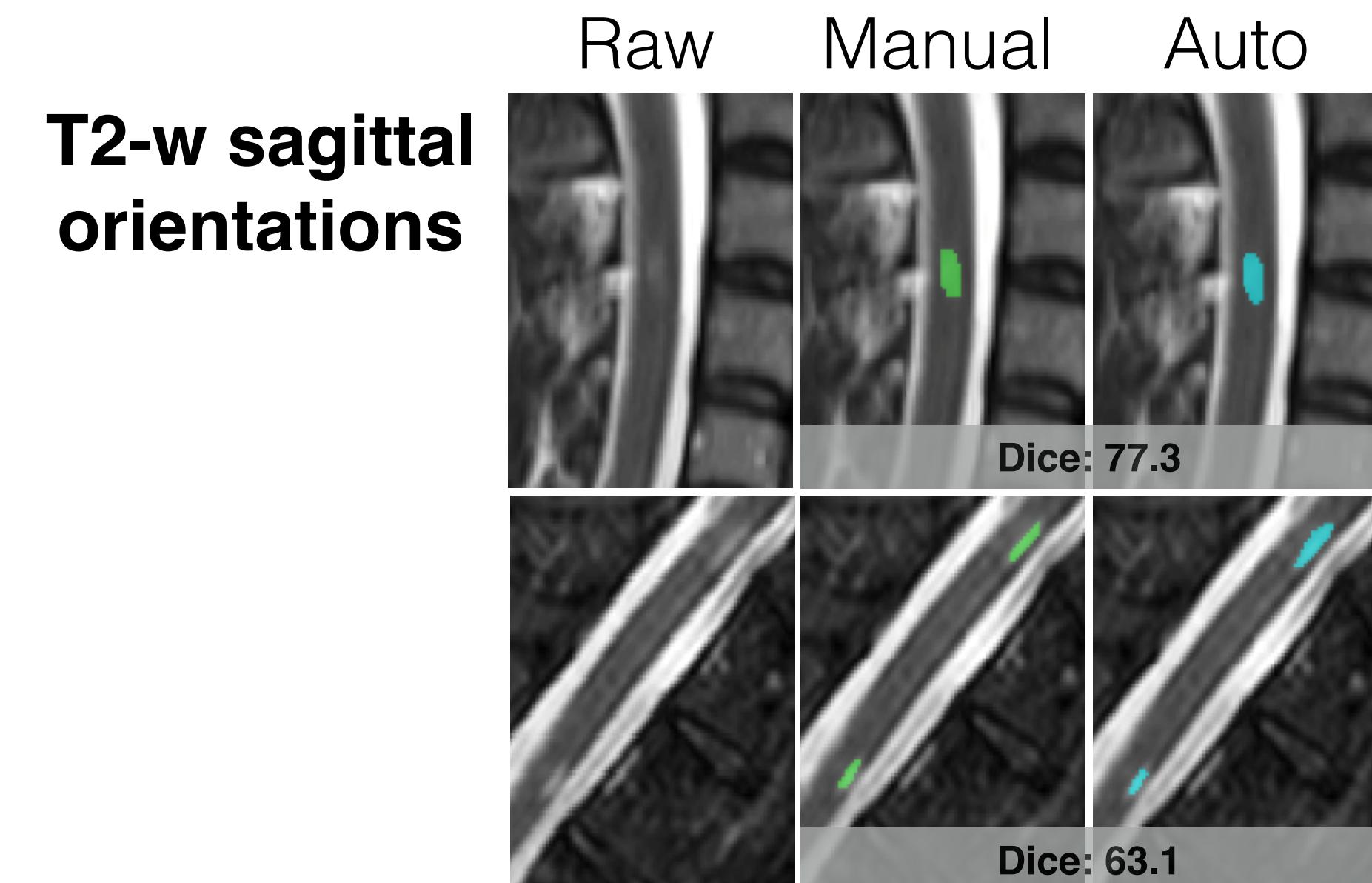
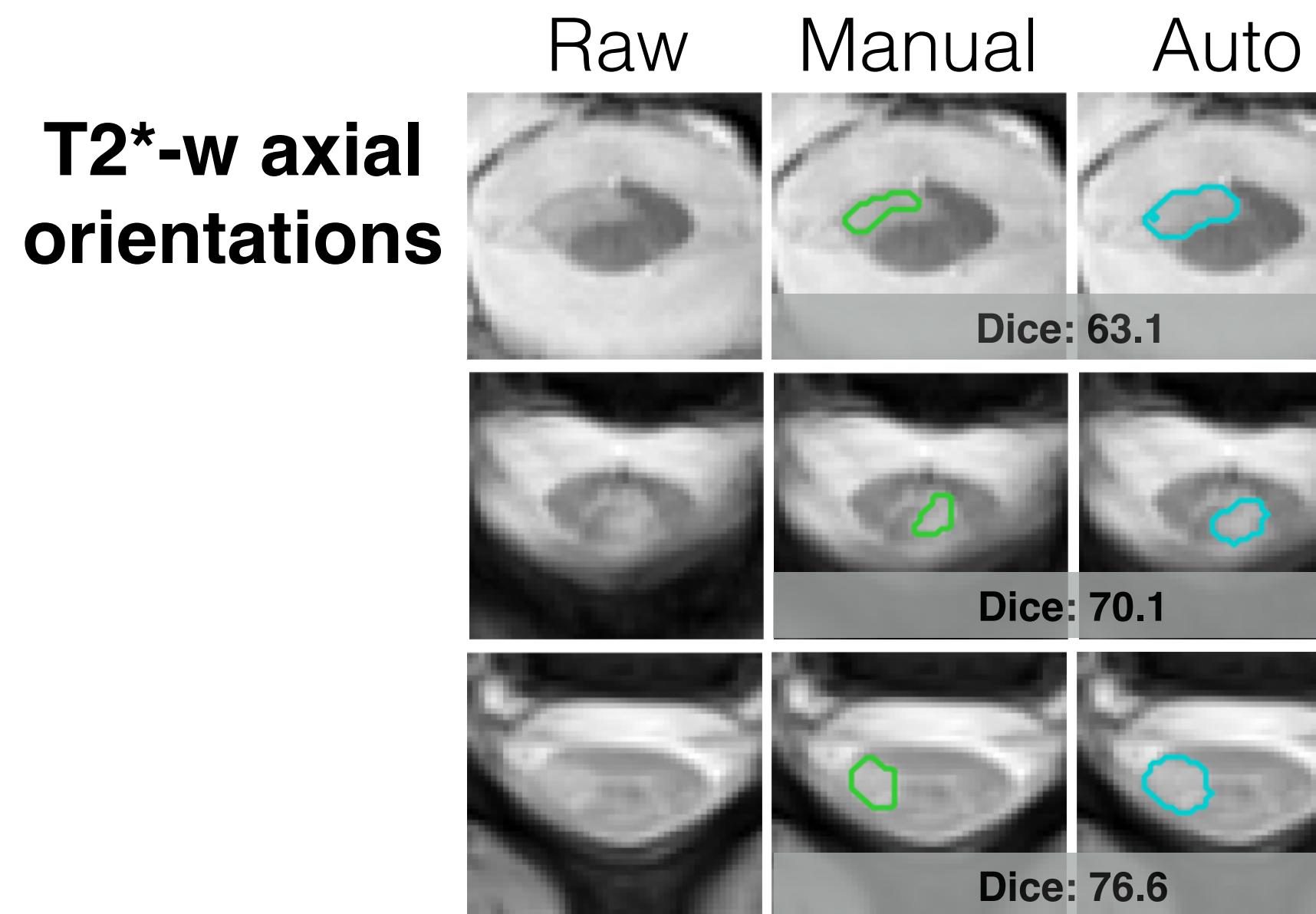


Segmentation of MS lesions with DL [1]

Charley Gros

- **Results:**

- Overall good sensitivity (85%) but 30% false positive rate (i.e. spurious lesions)
- Robust to image contrast and orientation.

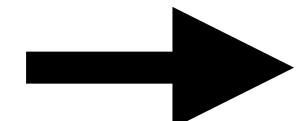
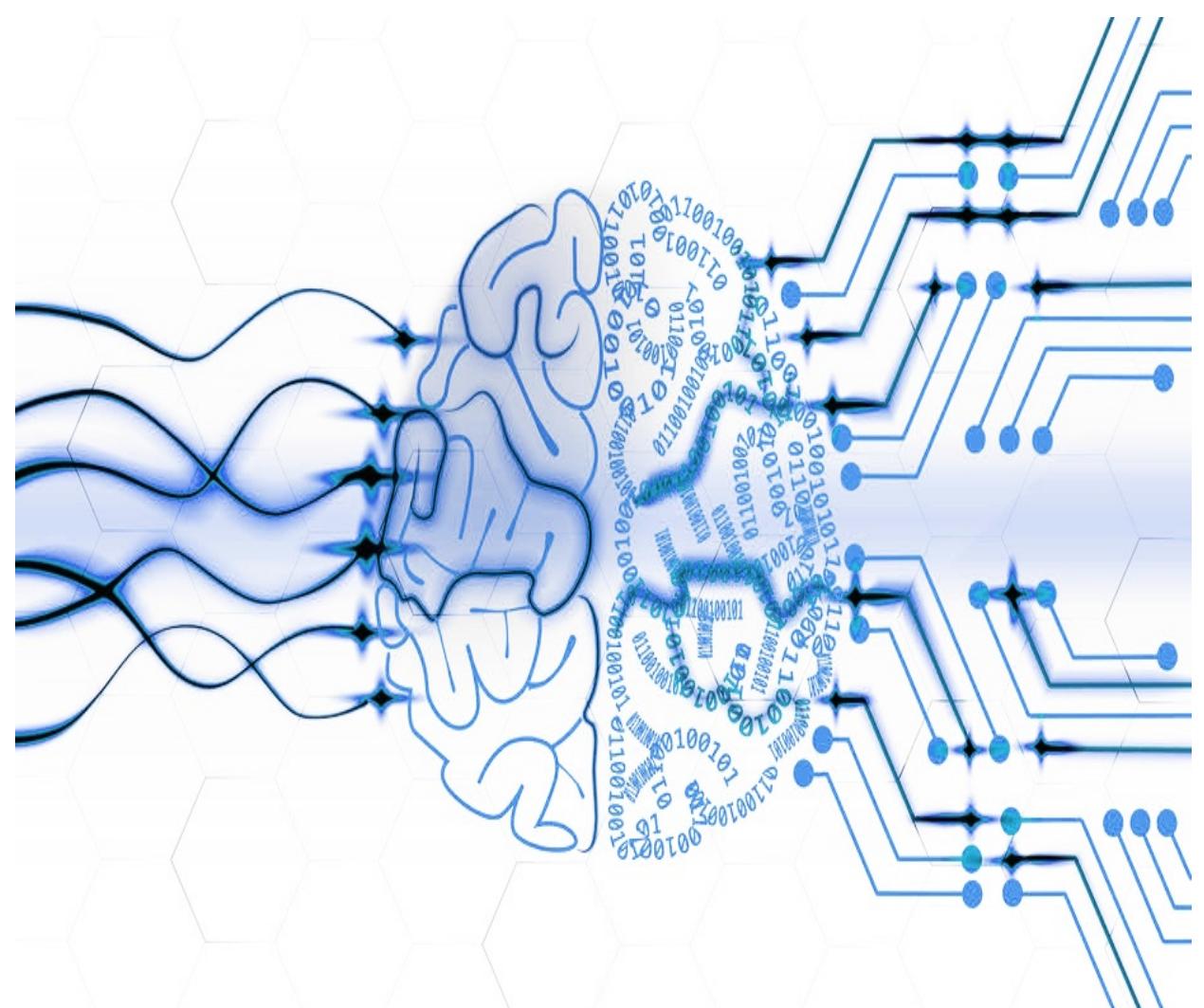


qMRI

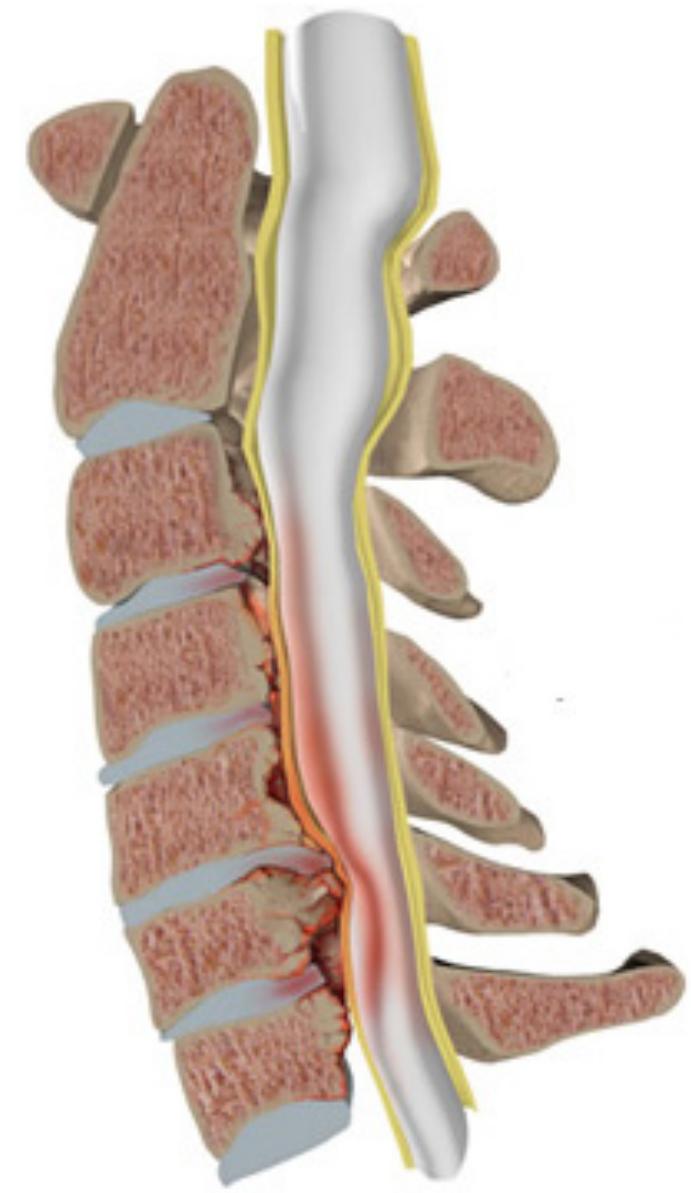


+

A.I.



Prognosis in DCM



Collaborators: Dr. Fehlings (Toronto)
Dr. Weber (Montreal)

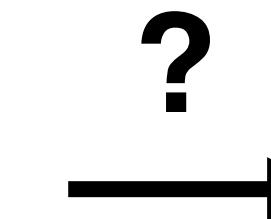


DCM: Prognosis with deep learning

Lucas
Rouhier

Matthieu
Parizet

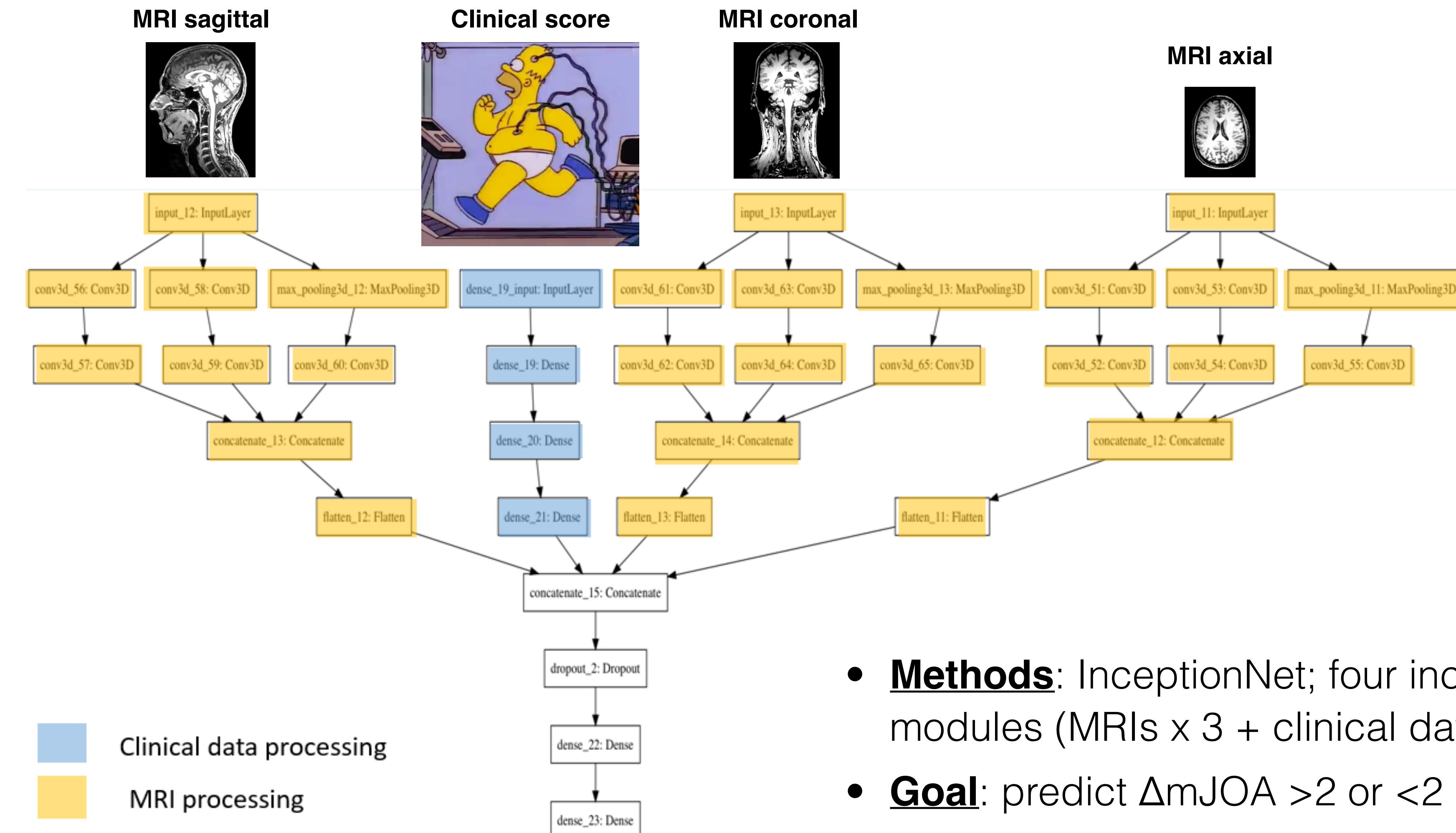
- **Context:** Patients with non-traumatic cord compression (disc bulging) → pain, motor deficits
- **Goal:** Predict the outcome of these patients to help decide whether to operate or not.
- **Data:** MRIs of ~500 patients from the AO Spine myelopathy database.
 - Clinical Scores before and 6 months after surgery (mJOA)



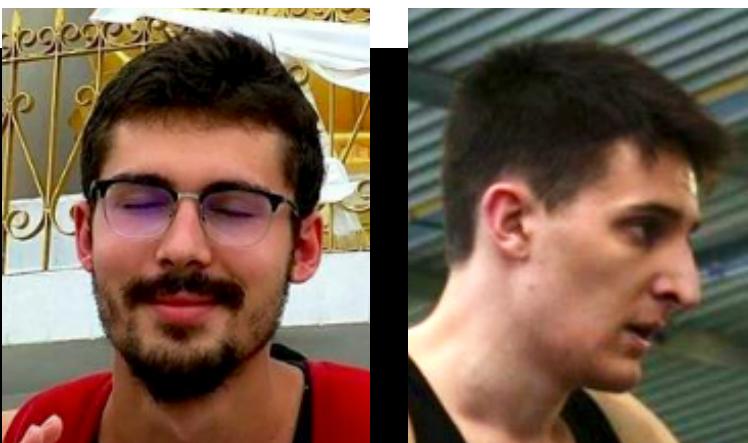
*to operate or
not to operate?*



DCM: Prognosis with deep learning



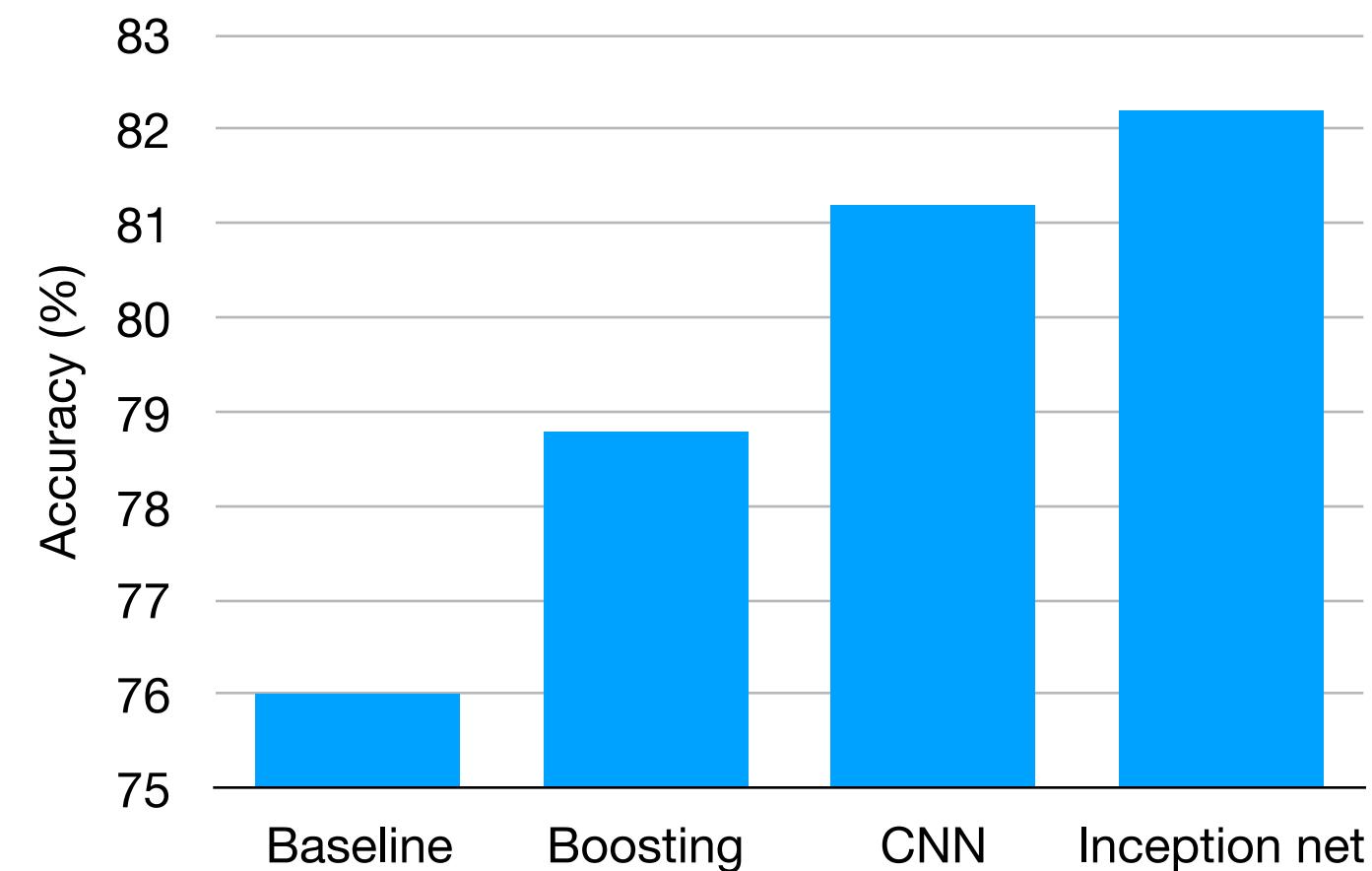
- **Methods:** InceptionNet; four inception modules (MRIs x 3 + clinical data)
- **Goal:** predict $\Delta\text{mJOA} > 2$ or < 2



DCM: Prognosis with deep learning

Lucas
RouhierMatthieu
Parizet

- Accuracy: up to 82.3% vs. 76% with clinical data only



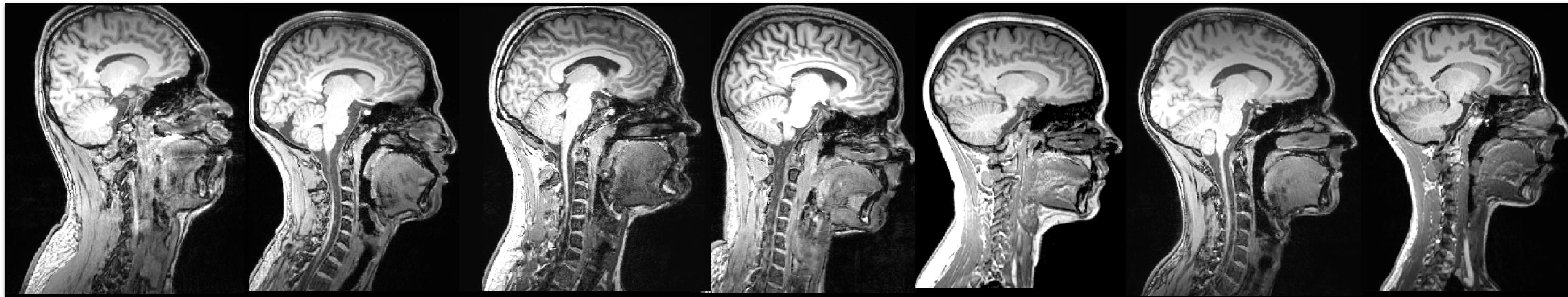
- Acquisition protocol **very** heterogenous, reducing performance of the network.
- **Ongoing work:** (i) Use processing tools to train networks on segmented images (reduce dimensionality), (ii) Explore Inception layers and residual connections to improve generalizability and (iii) introduce qMRI metrics.

Outstanding issues

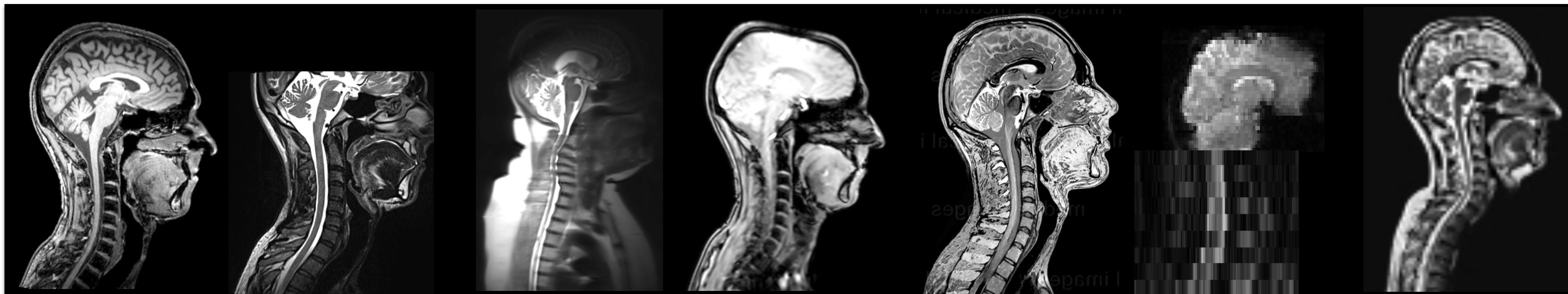
- Code rarely publicly available. When it is, maintenance/commenting is of poor quality.
- Low amount of data and manual labels are available (IRB & Privacy).
- Variety of input image resolution (2mm, 1mm, 0.5mm) —> difficult to set proper batch size
- Training with multiple contrasts: Some datasets might be missing
- Published models validated in well-curated single-center datasets.
They usually fail when applied to other centers (a.k.a. Real life data!).

Intra- vs. Inter-study variability

People are different... but not that much!

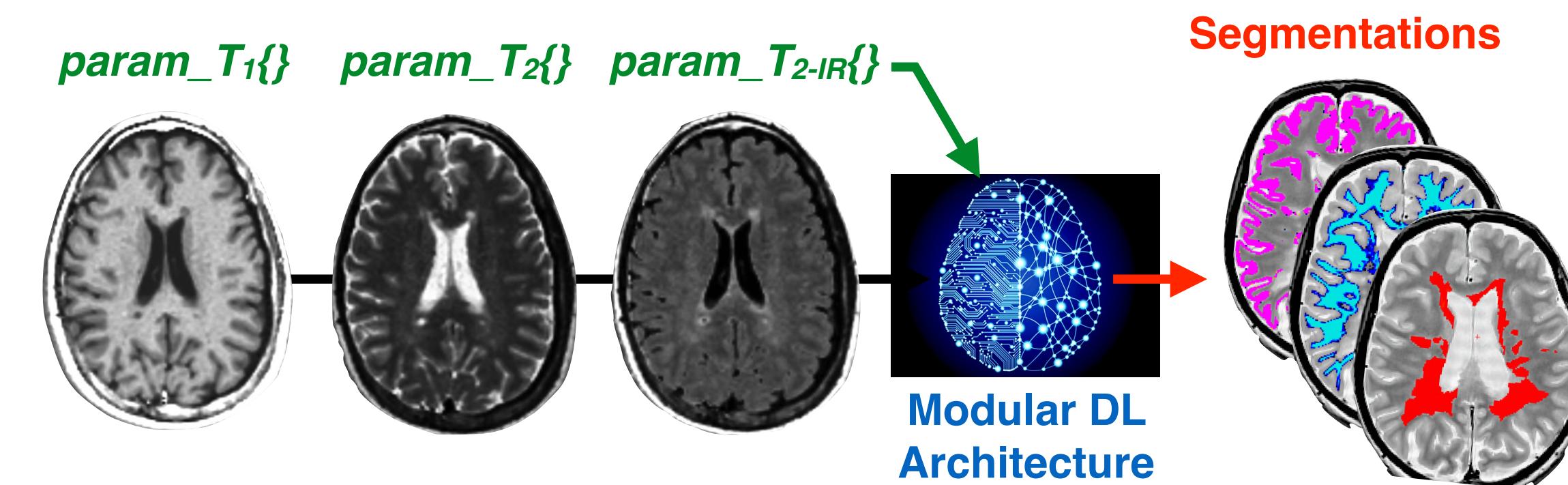


Imaging parameters, MRI model, etc. create the largest variability:

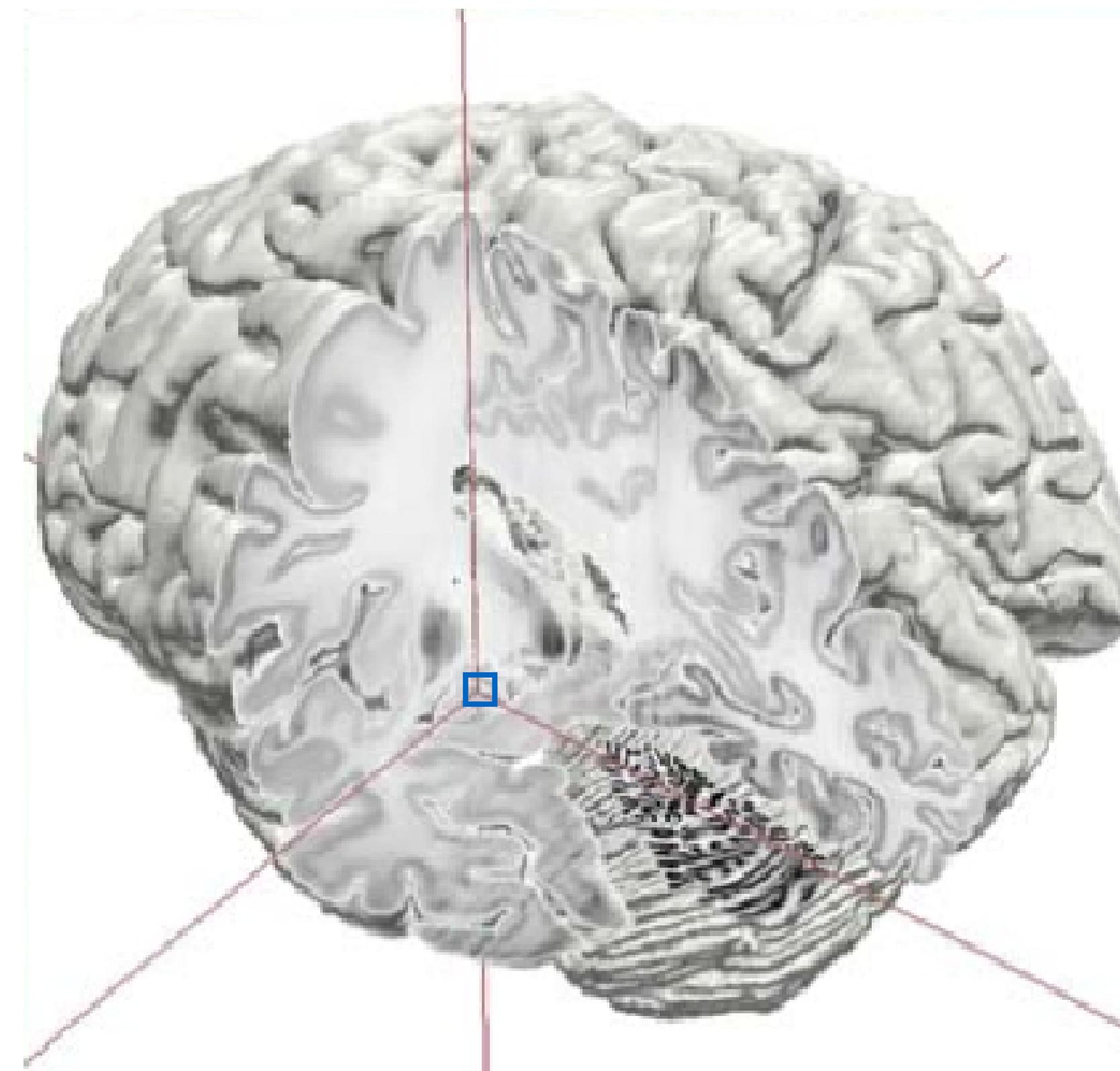


Possible avenues

- **GANs**: Treat each clinical center as a “domain”, then create a function that can map data from one center to another.
 - Problem #1: Image features not only varies between centers, but also across a large number of acquisition parameters.
 - Problem #2: Domain adaptation can introduce bias [1]
- **Multiple contrasts**: Within a typical MR session, multiple contrasts are acquired (e.g., T1w, T2w). We should use all the available contrasts to improve the robustness of DL tasks, but also deal with “missing” contrast for generalizability [2].
- **Inject priors from MR physics**: Learn image features based on acquisition parameters that have an impact on the image (TR, TE, flip angle, inversion recovery pulse, saturation band, magnetization transfer pulse, etc.)

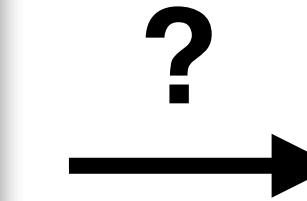
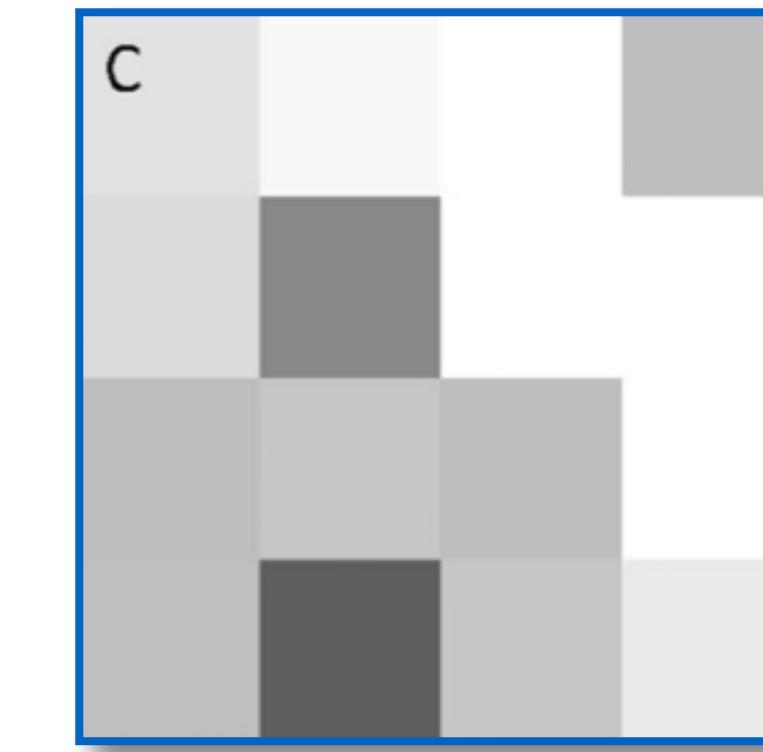


In vivo histology with MRI: Myth or reality?

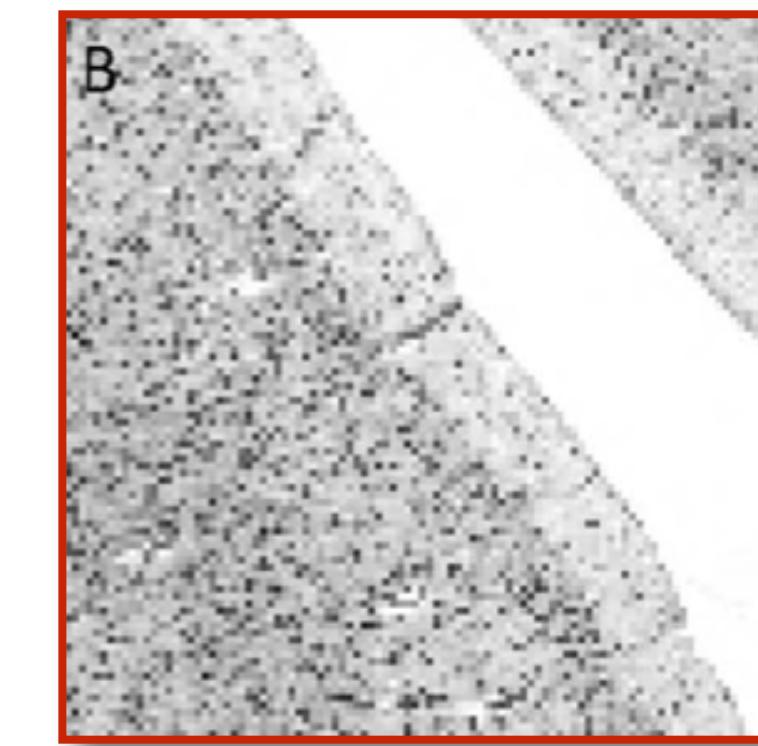


Validation is required!

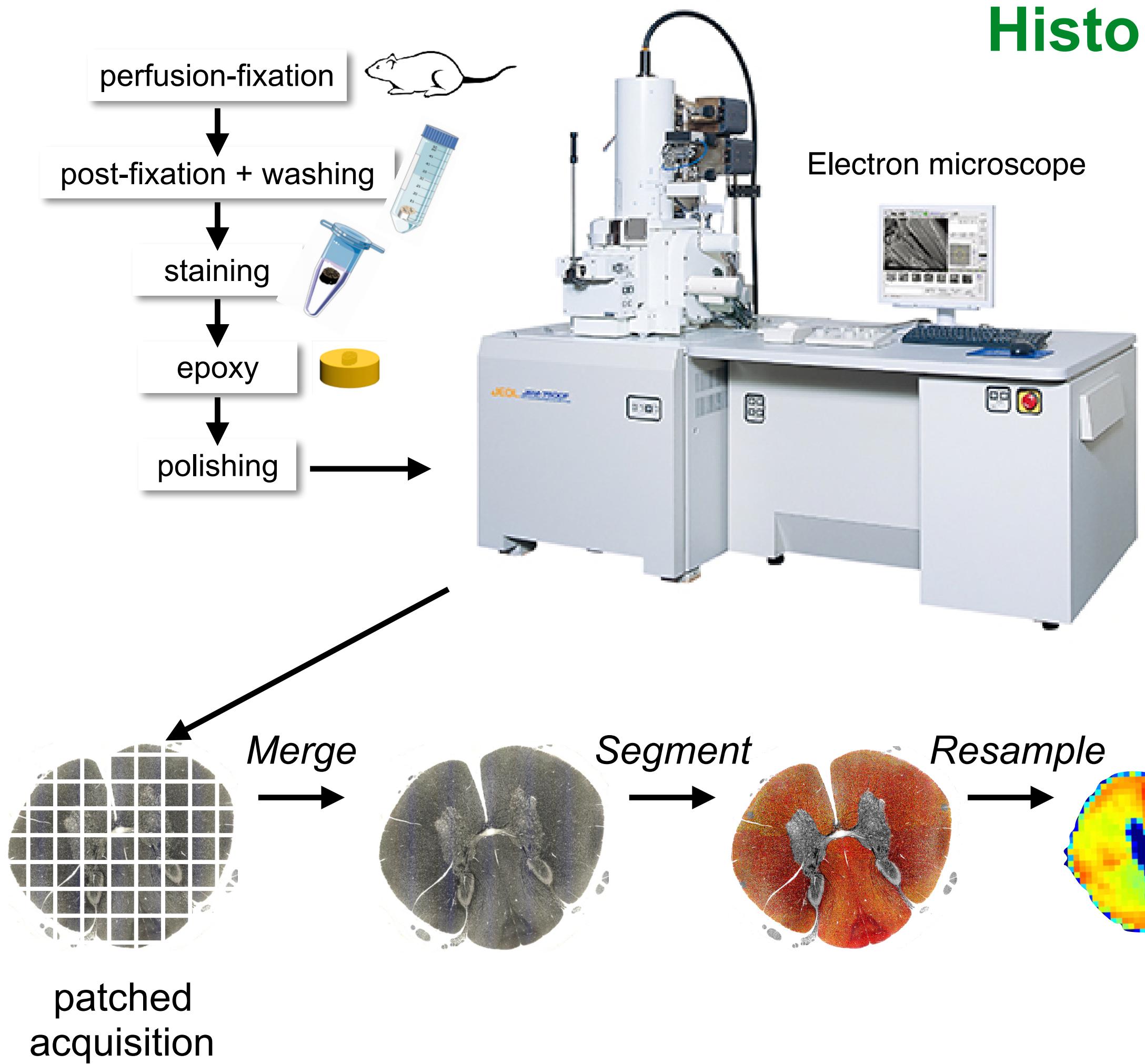
Fancy MRI
technique



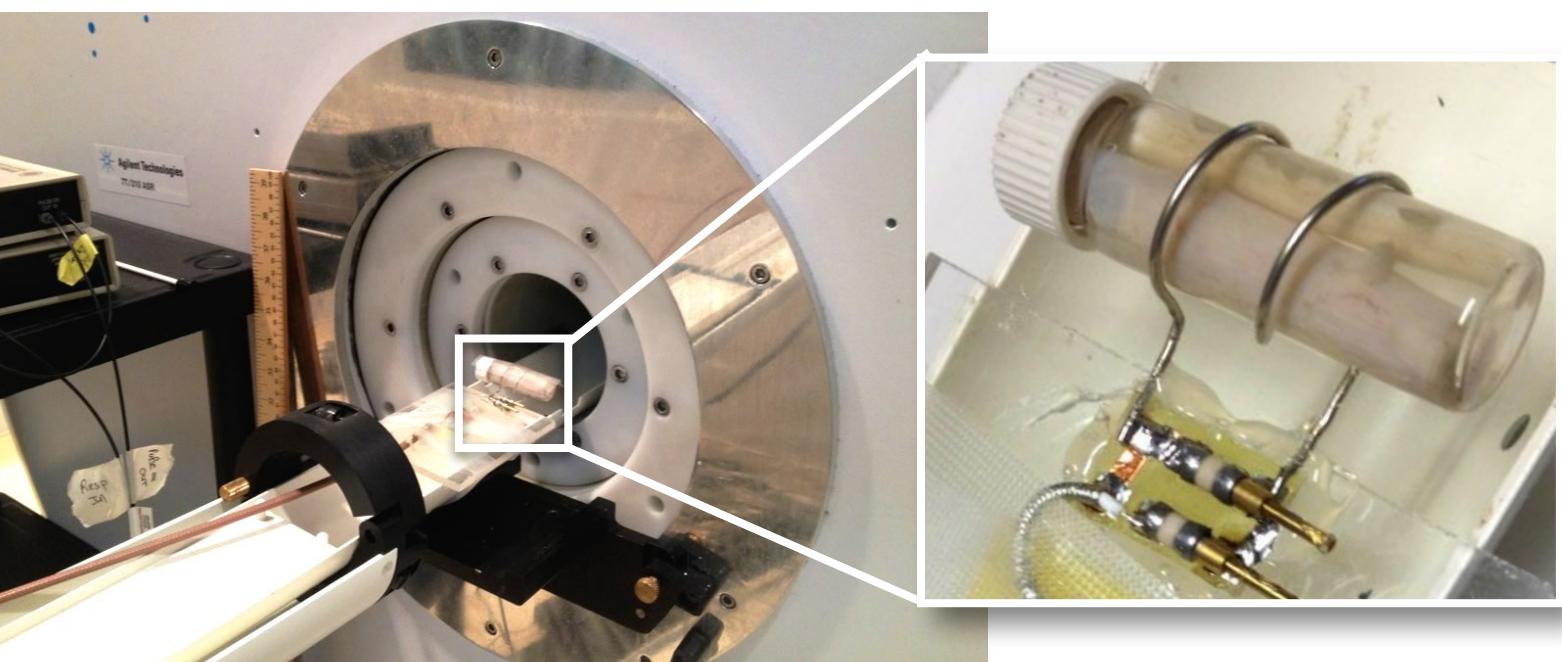
Ground-truth
histology



Histology

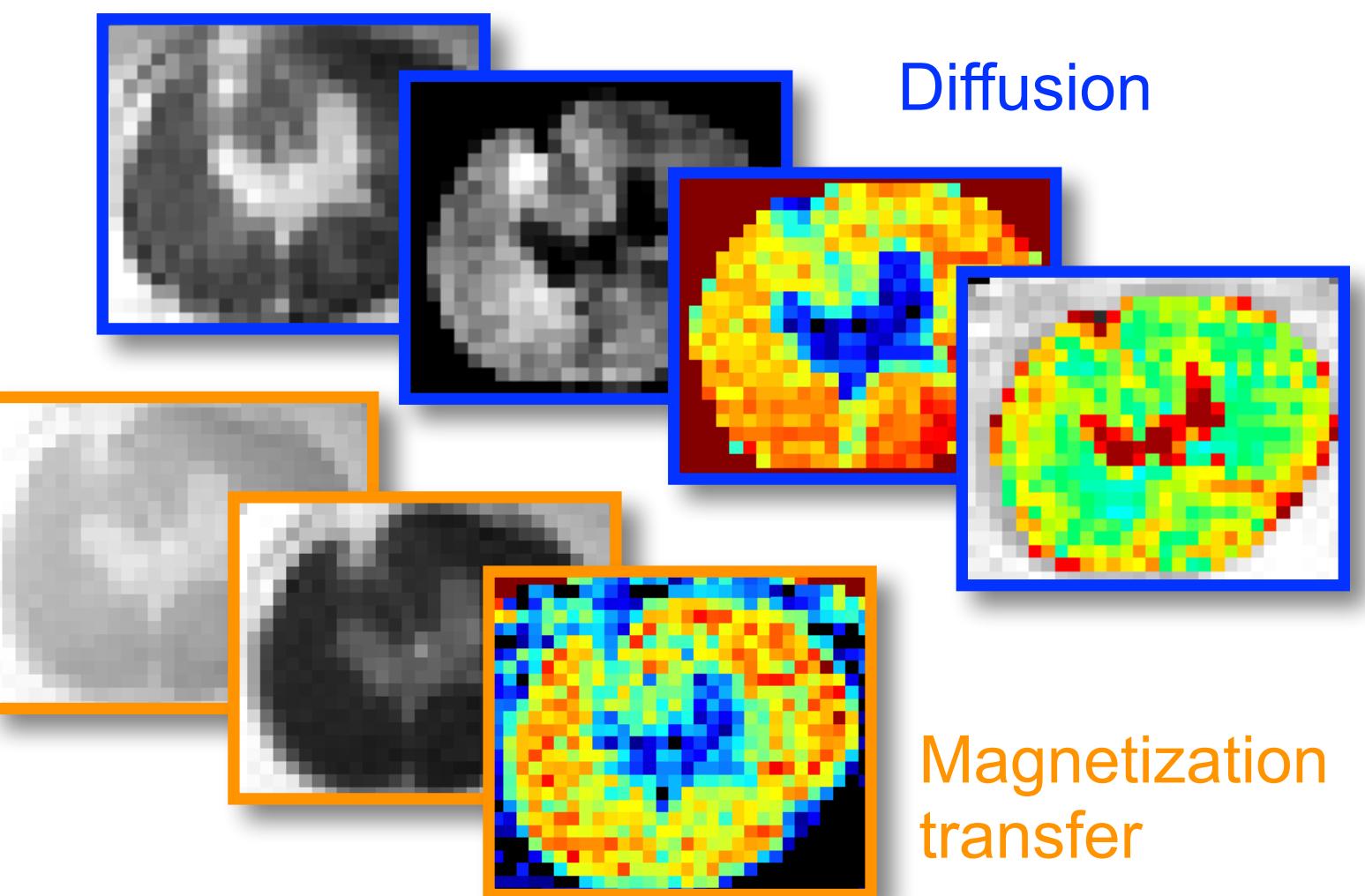


MRI

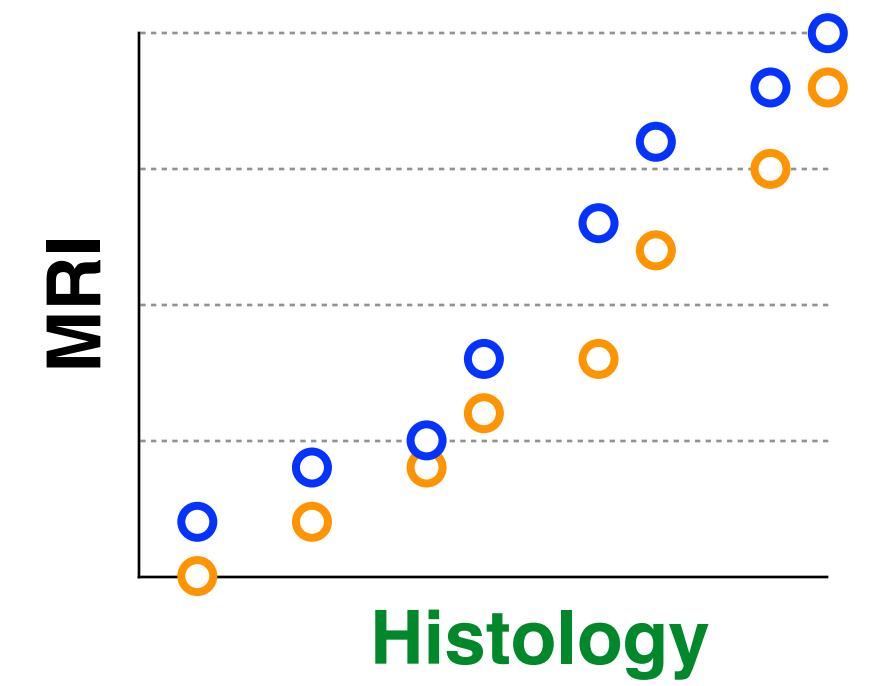


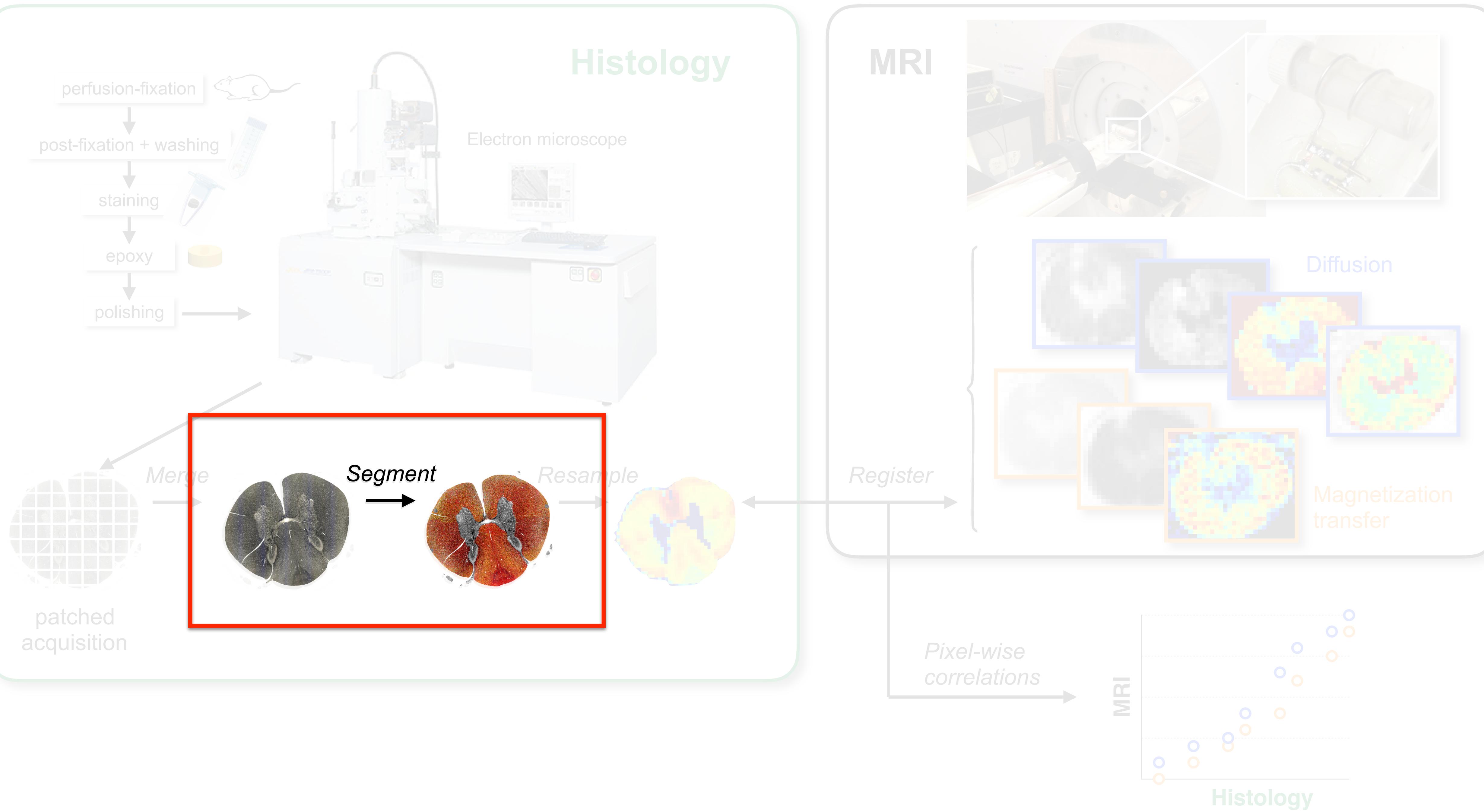
Fancy qMRI
techniques

Register

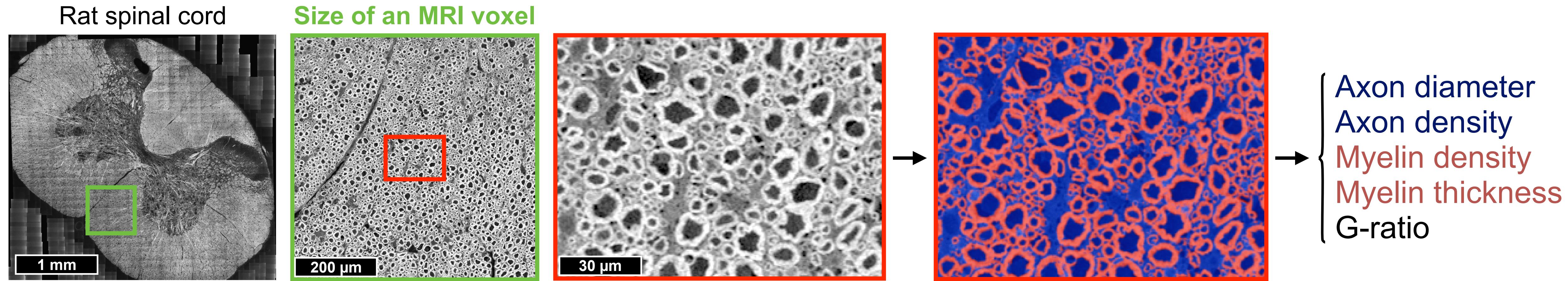


Pixel-wise
correlations





Large-scale histology



Manual segmentation of each individual axon and myelin sheath:

- Density: ~100,000 axons/mm²
- White matter cross-sectional area: ~10mm²
- Time to segment one axon + myelin: ~1min

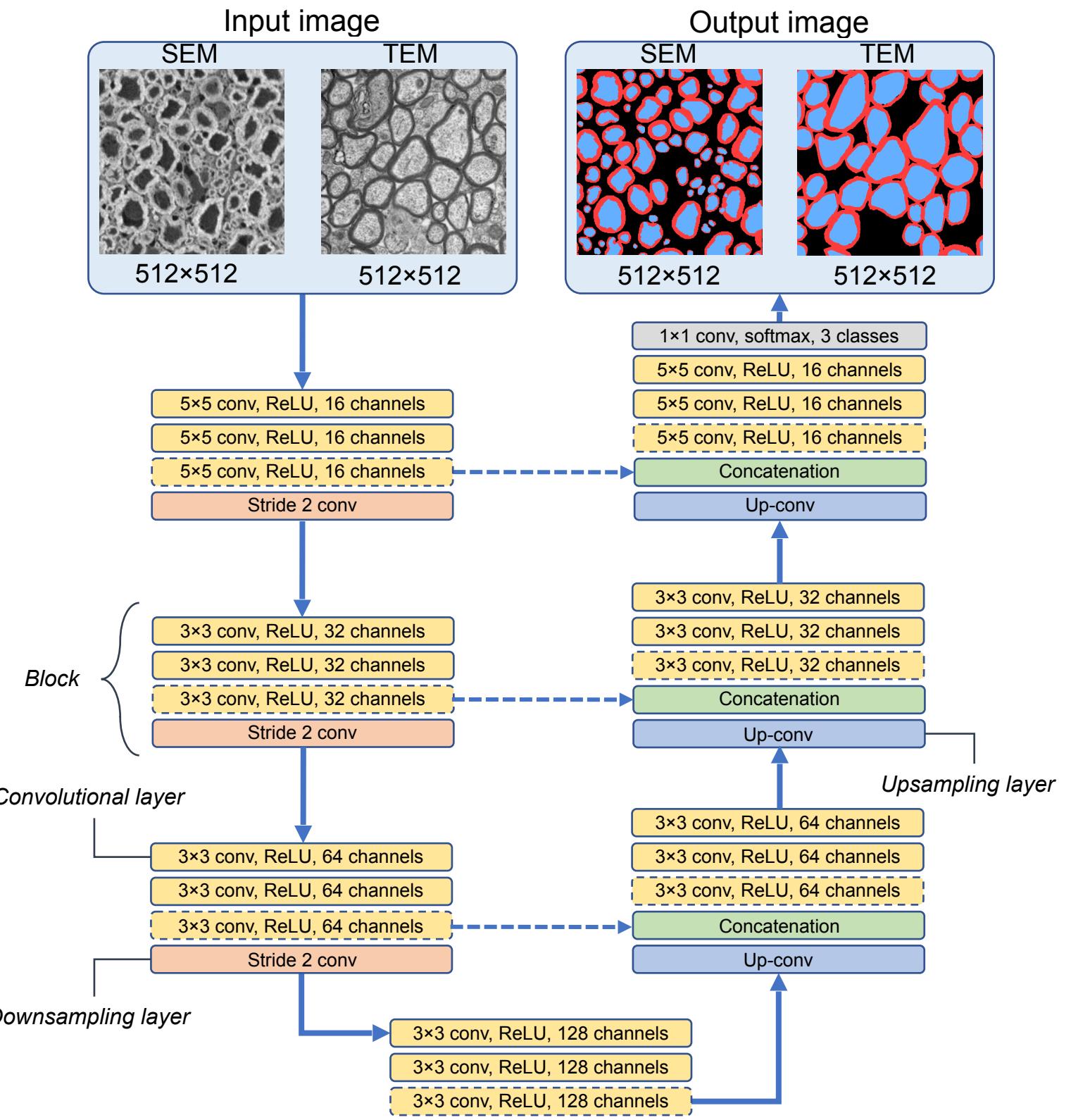
→ Time to segment one slice: **~2 years** (without eating, sleeping, watching Netflix, etc.)



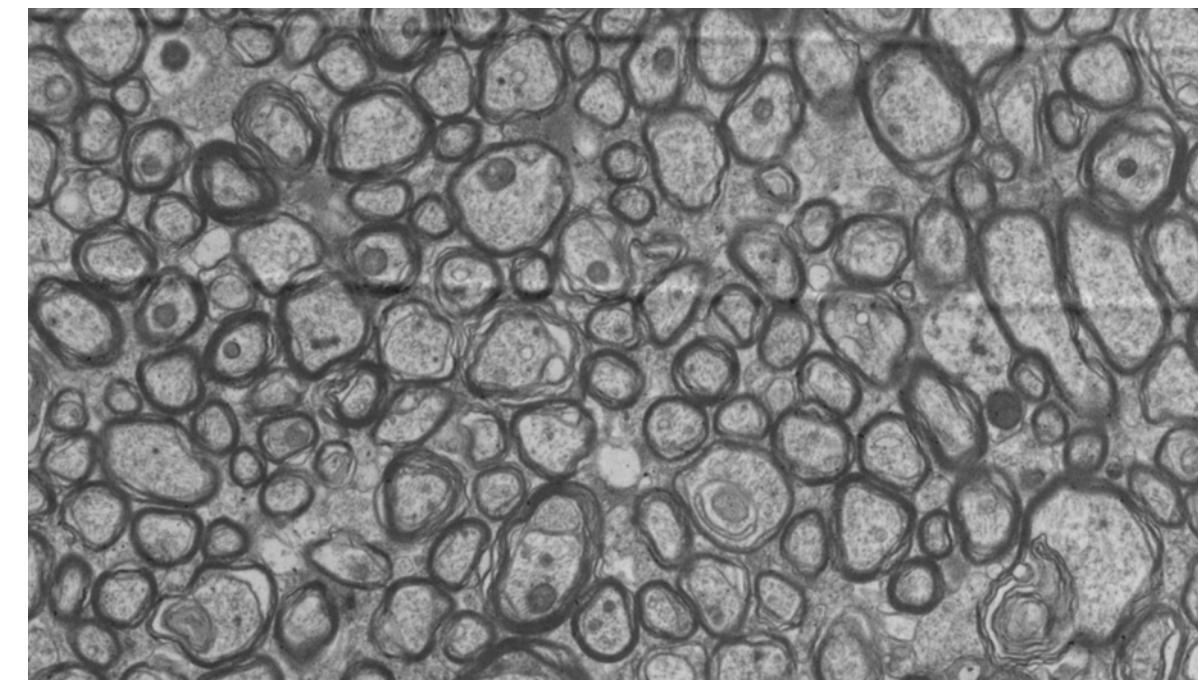


AxonDeepSeg Toolbox

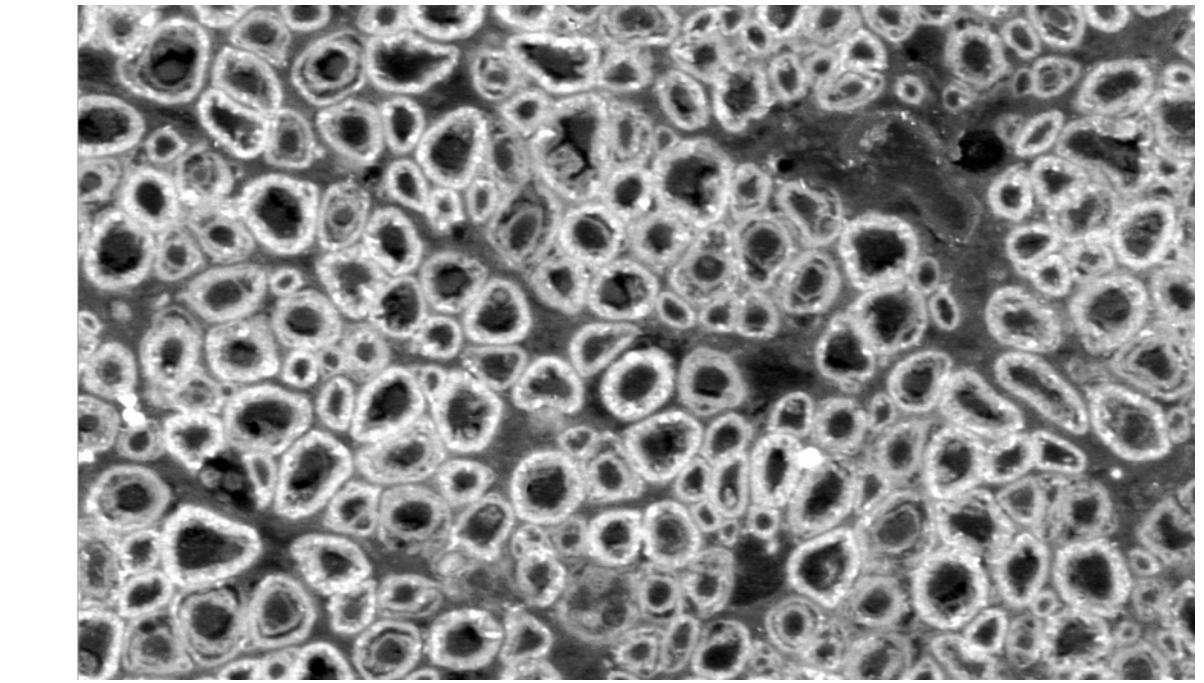
- Network:** U-Net [Ronneberger et al.] with TensorFlow
- Two models:** SEM & TEM
- Training data:** 500 patches (512x512)
- Accuracy: 93%, Precision: 87%, Sensitivity: 96%**



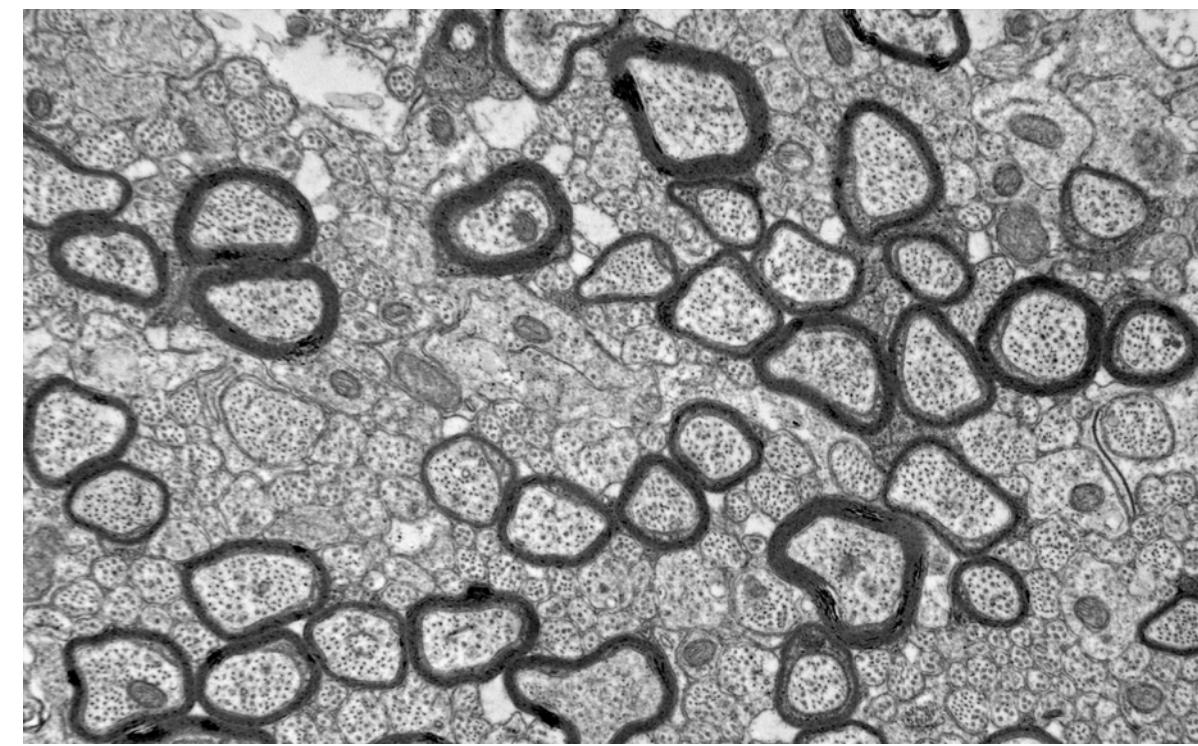
Macaque TEM



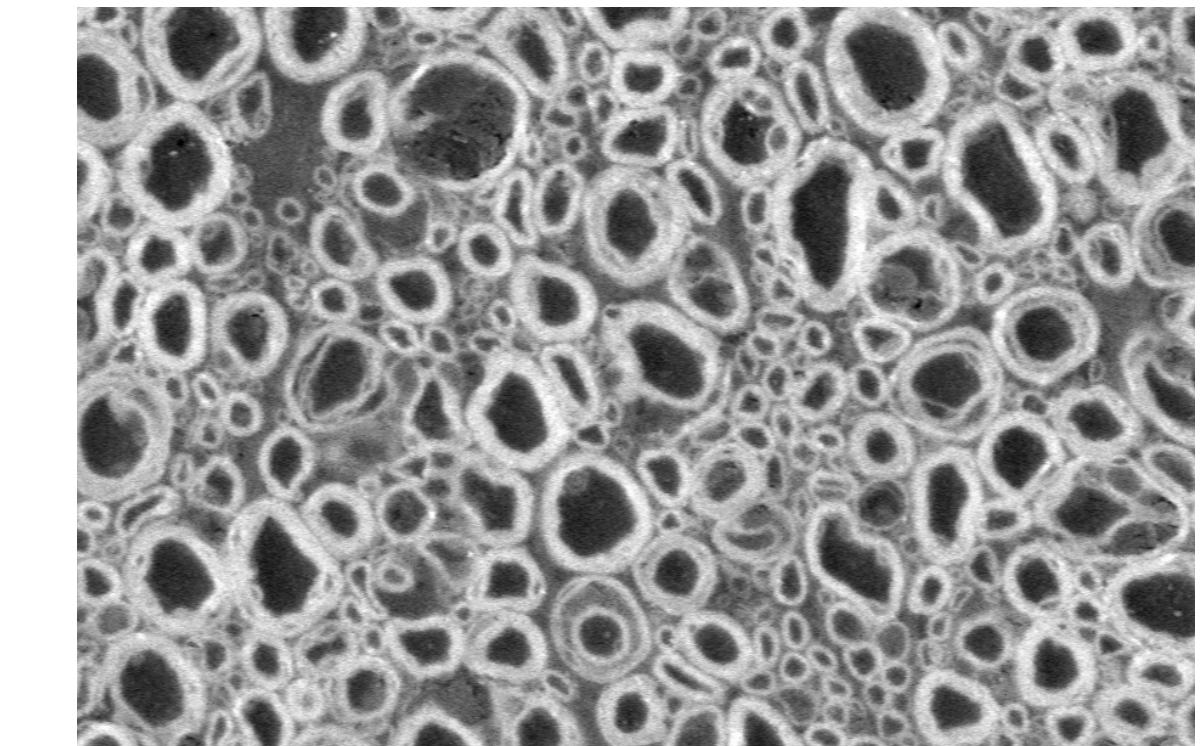
Human SEM

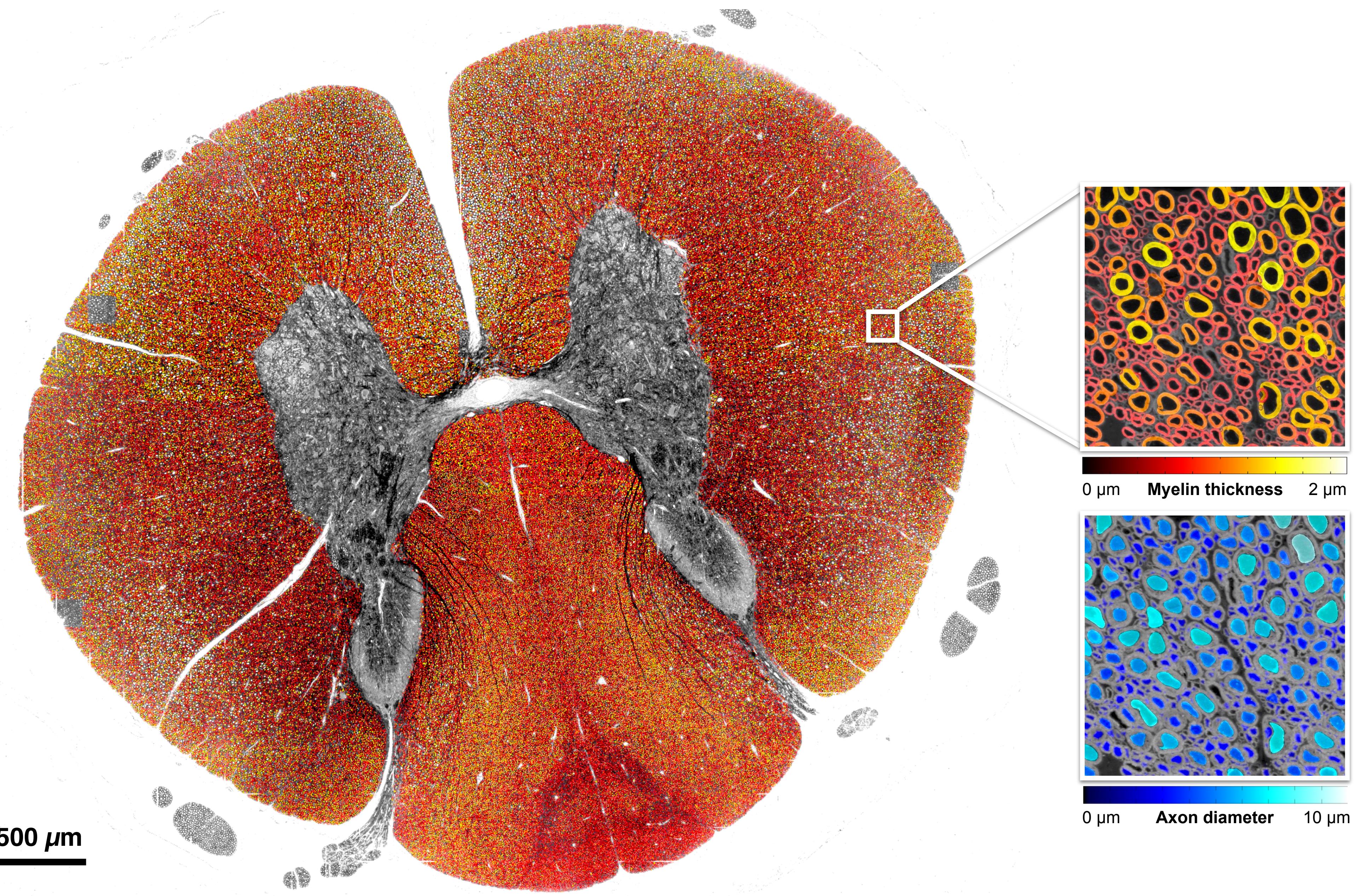


Mouse TEM

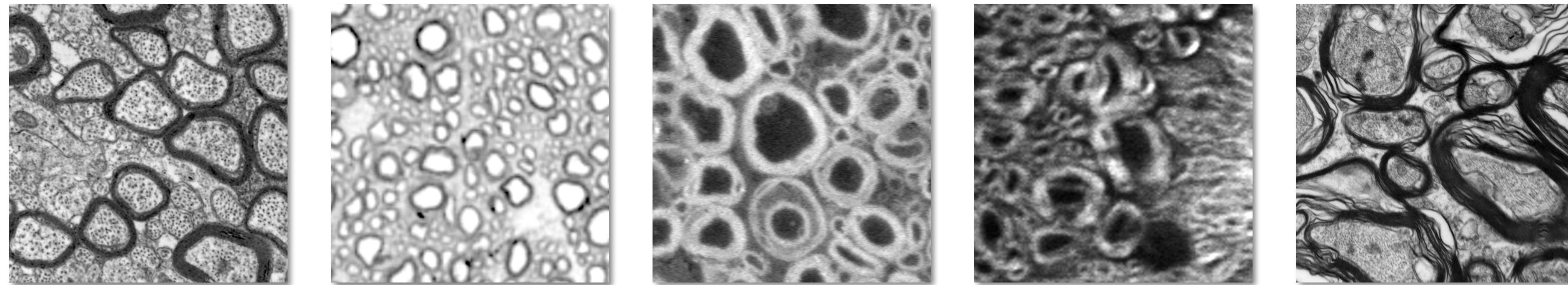


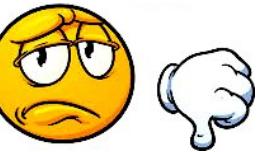
Rat SEM





Problem: Heterogeneity across acquisition parameters



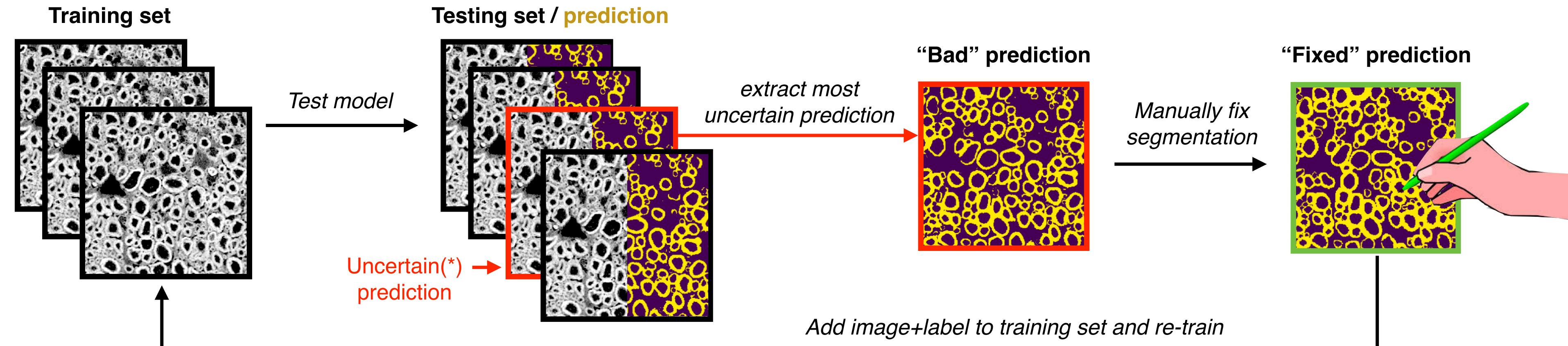
- Users have very specific needs:
 - Intra-modality —> good homogeneity of datasets 
 - Across modalities, projects and acquisition parameters —> poor homogeneity 
- Our strategy: Provide a framework to train a model for each application
- How to minimize users efforts in creating labels?



Uncertainty & Active learning

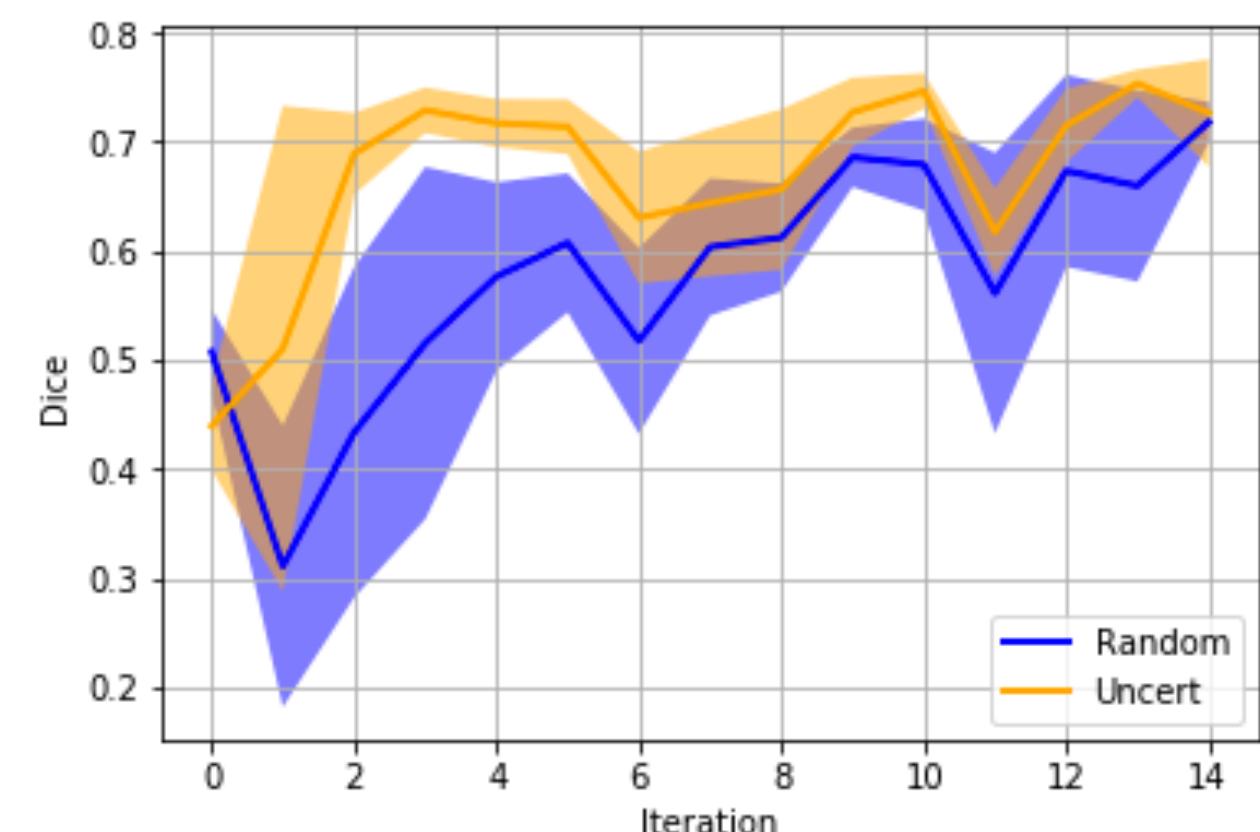
Mélanie
Lubrano

- **Active learning [1]**: Select patches that are **most uncertain** at prediction, ask user to fix them, then re-train.



(*) “**Uncertainty**” obtained from Monte-Carlo samples of the models with dropout at test-time.

- **Results**: Encouraging, but room for improvement. Maybe combine with Mean Teacher approach.



Conclusion

- Large databases are exciting... but small datasets with few labels is the reality for most medical applications
- Quest for generalizability: better to provide several application-specific models or one-that-fits-all?
- **Real impact** can be made if (i) good solutions are provided for such scarce data and (ii) clean code is shared.
- **Innovative solutions** fostered by bringing communities together



"Still nothing. You sure this is the
best computer we have?"

Acknowledgements



Nikola Stikov

Agah Karakuzu

Aldo Zaimi

Alexandru Foias

Ariane Saliani

Atef Badji

Benjamin De Leener

Charley Gros

Christian Perone

Dominique Eden

Gabriel Mangeat

Grégoire Germain

Harris Nami

Ilana Leppert

Jennifer Campbell

Jérôme Carretero

Lucas Rouhier

Matthieu Parizet

Mathieu Boudreau

Mélanie Lubrano

Nibardo Lopez-Rios

Nicolas Pinon

Oumayma Bounou

Ryan Topfer

Tanguy Duval

Tommy Boshkovski



Michael Fehlings

Allan Martin



Claudia Gandini Wheeler-Kingshott

Ferran Prados

Becky Samson



Yoshua Bengio

Christopher Pal

Adriana Romero Soriano



Shigeki Aoki

Masaaki Hori

Akifumi Hagiwara

Koji Kamagata

Kouhei Kamiya



Pierre Bellec

Carollyn Hurst

André Cyr



Gerald Moran

