Applications of Deep Learning for High-Throughput Imaging

Gianluca Pegoraro
CCR High-Throughput Imaging Facility

Jun 15 2021

HiTIF

NIH NATIONAL CANCER INSTITUTE

High-Throughput Imaging (HTI)

PI

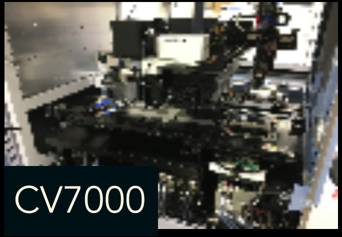
Experimental perturbation → Imaging-based cellular assay → Phenotypic change

HiTIF

Automated liquid handling

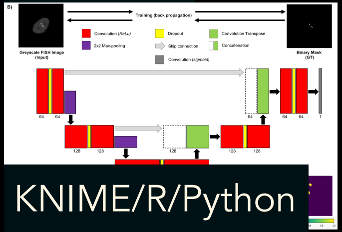
→

High-throughput microscopy



→

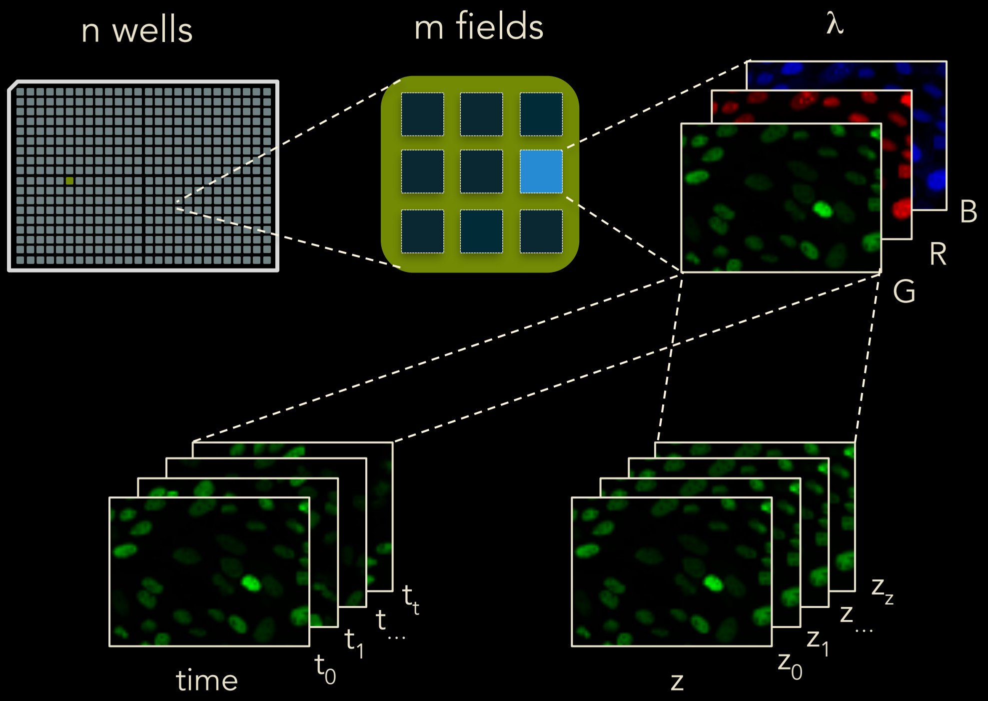
High-content image analysis

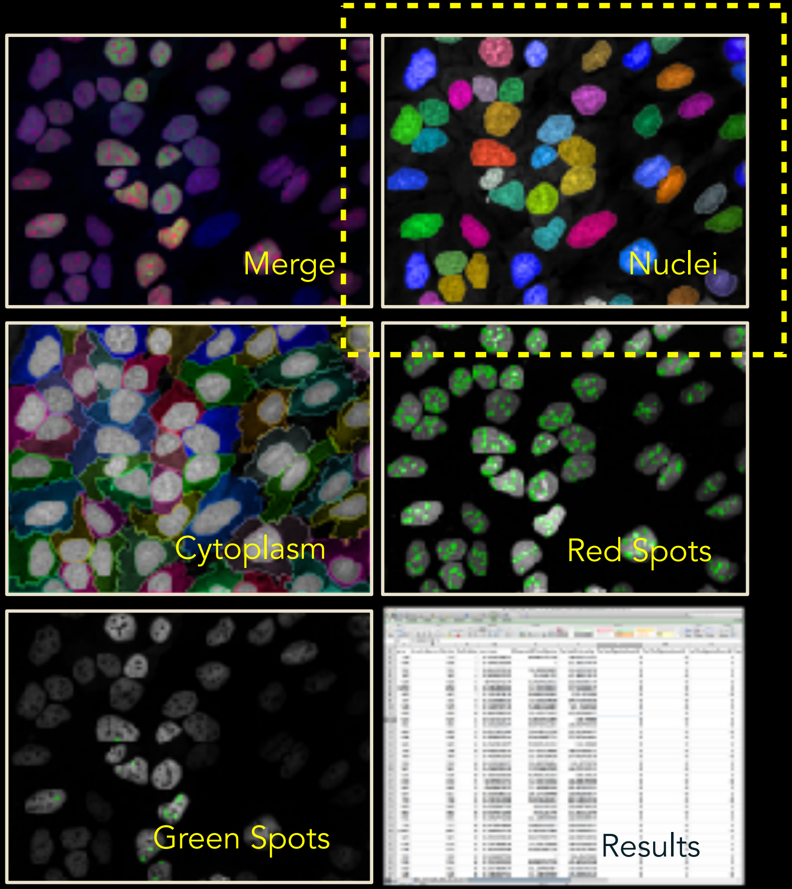


Up to:

- 10^4 Wells
- 10^4 Cells/Well
- 10^2 Feat./Cell

High-Throughput Acquisition and Analysis



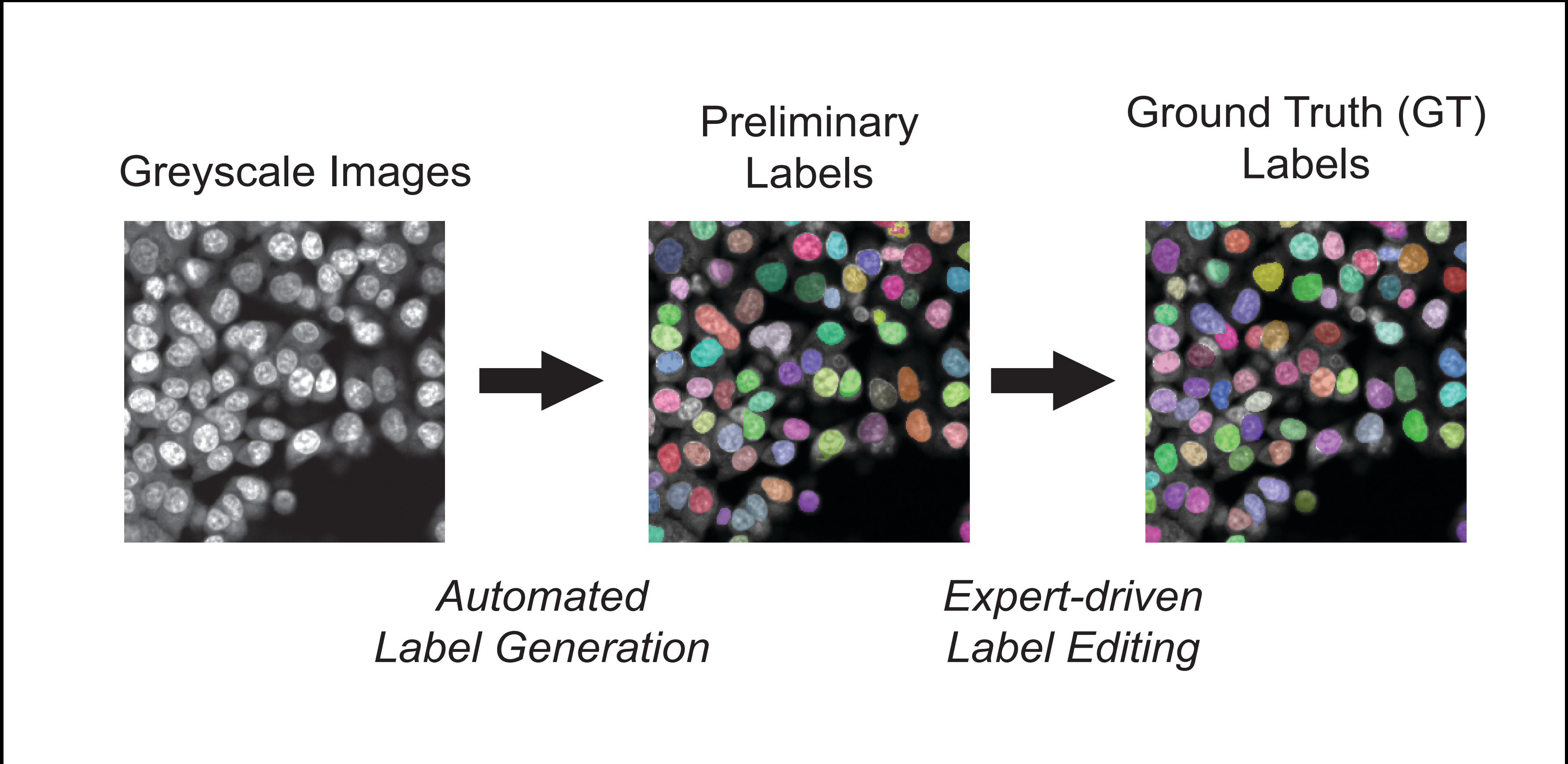


2D images/day = $n * m * \lambda * z * t \approx$ *up to* $2*10^5$

Deep Learning for Nucleus Segmentation

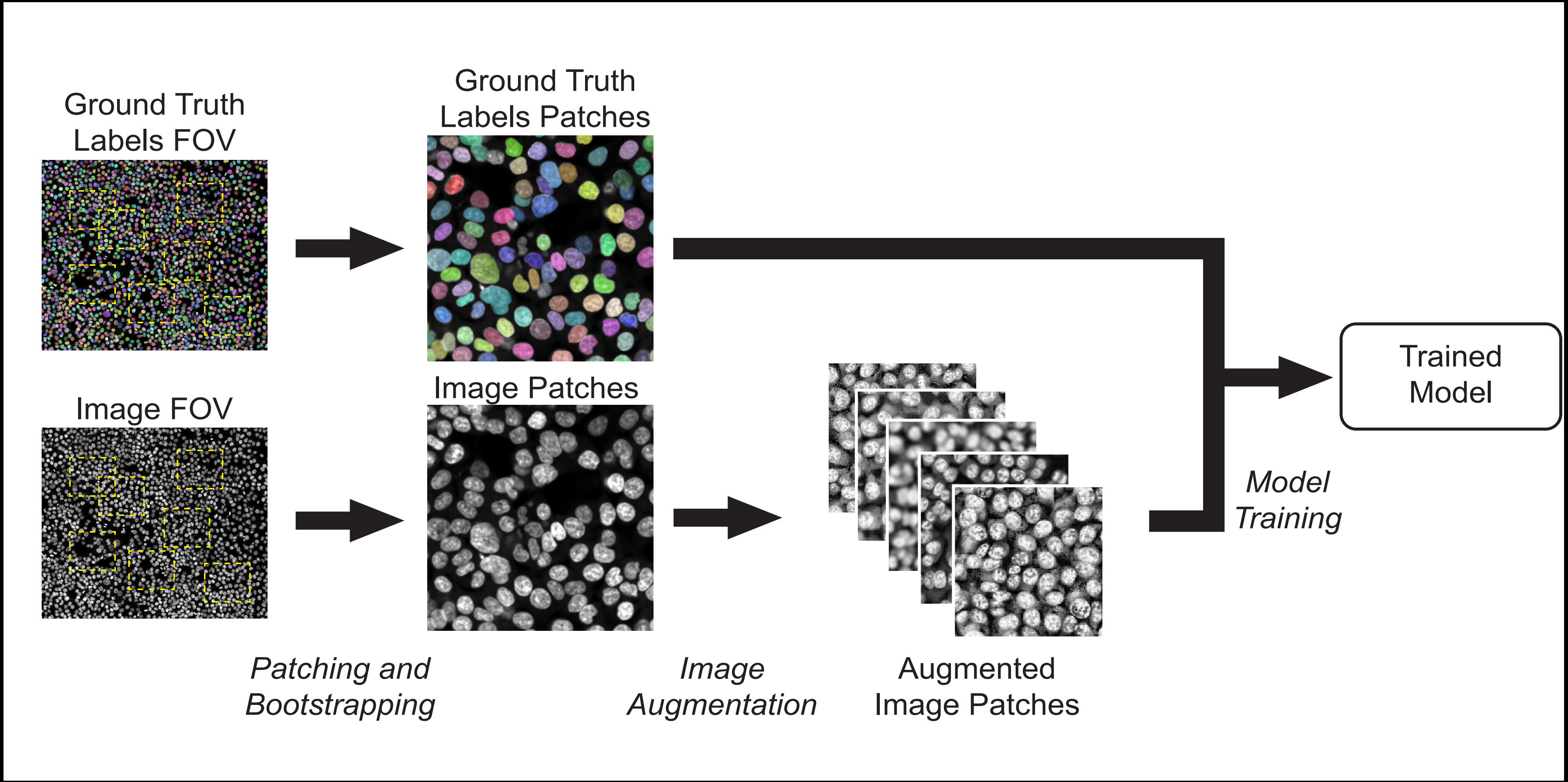
- Accurate Detection:
    - 90%-95% accuracy
- Practical:
    - Trainable with ~ 10 FOVs (~500 - 1,000 objects)
    - Fast inference (~ 1s/FOV)
- Robust and Generic:
    - Different cell types
    - Different magnifications
    - Different confluency

Semi-automated GT Label Generation

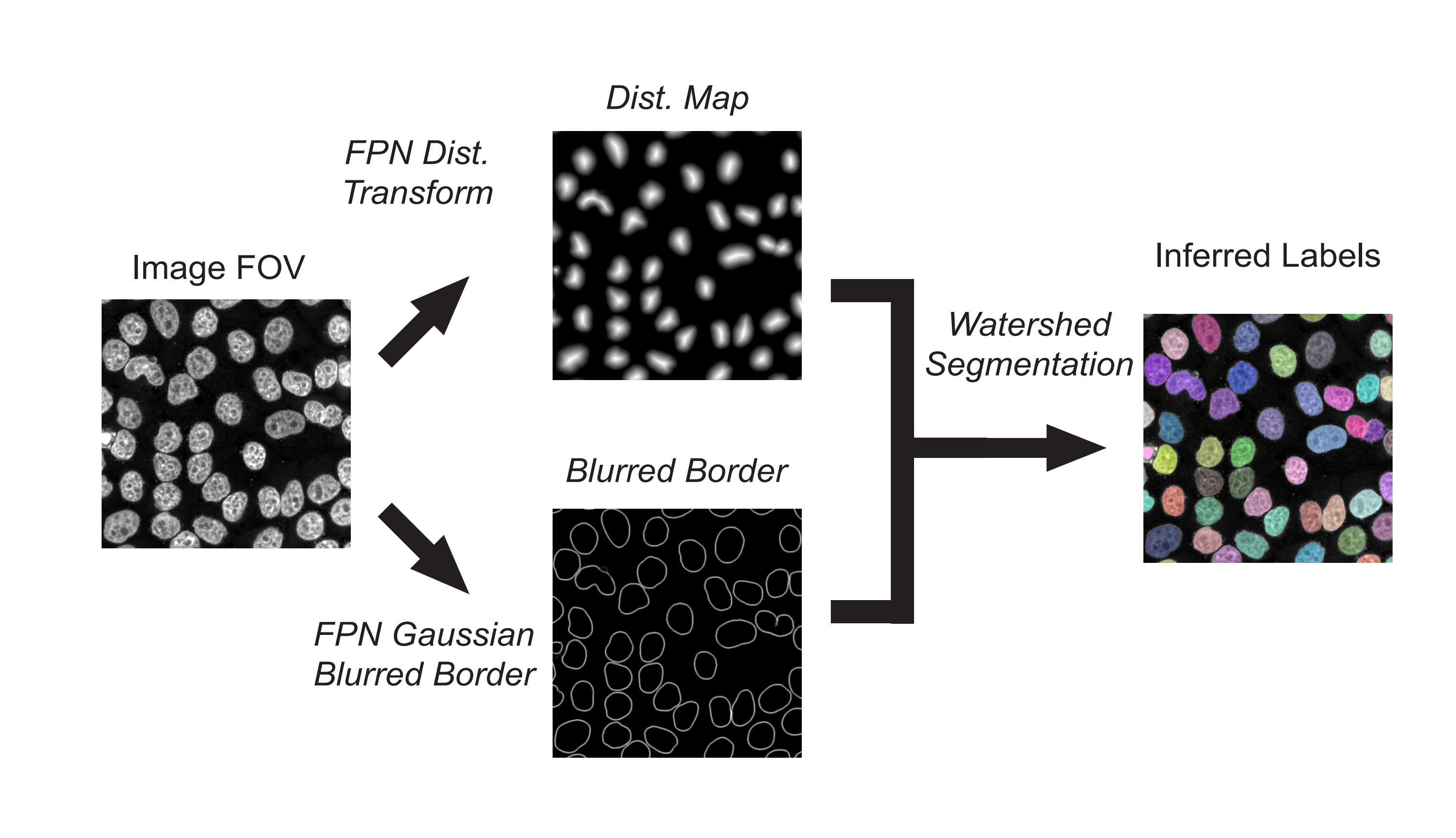


G. Zaki, Gudla P., et al., Cytometry A, 2020

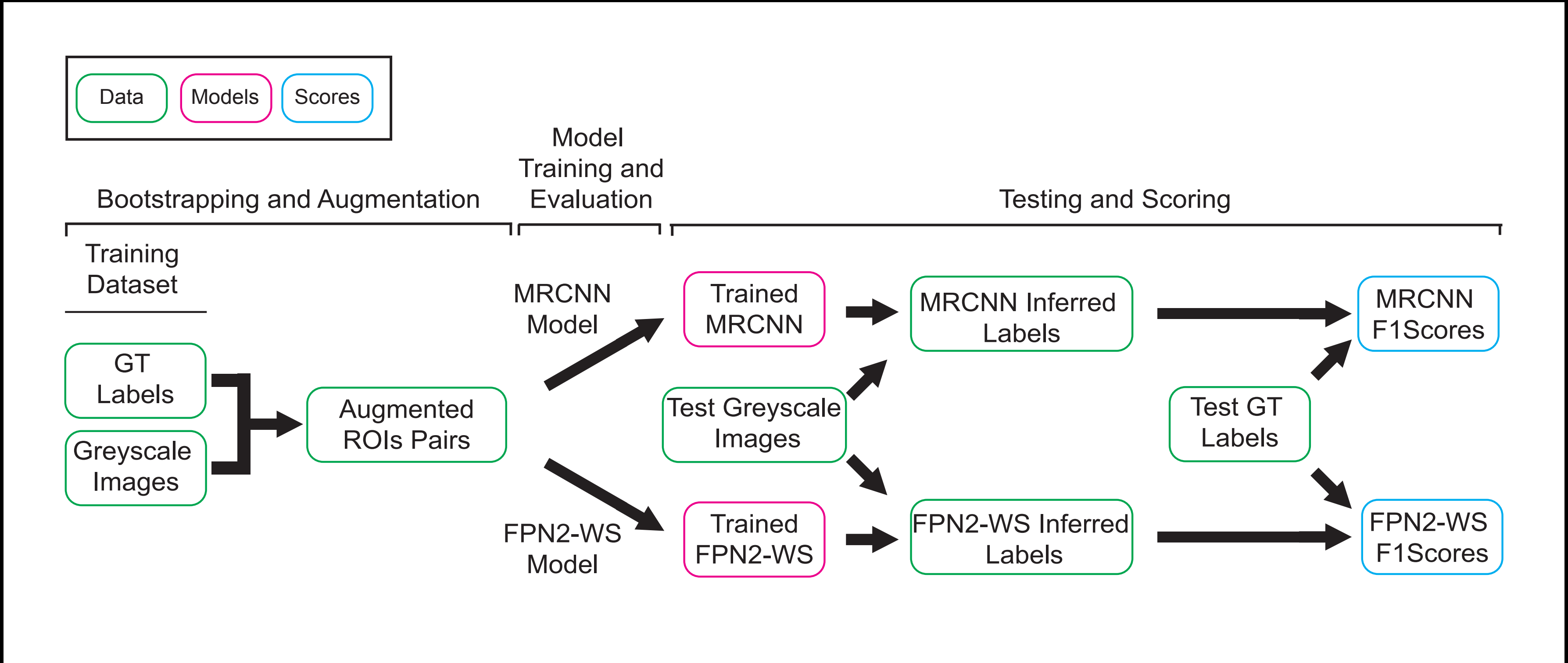
Image Augmentation and Bootstrapping



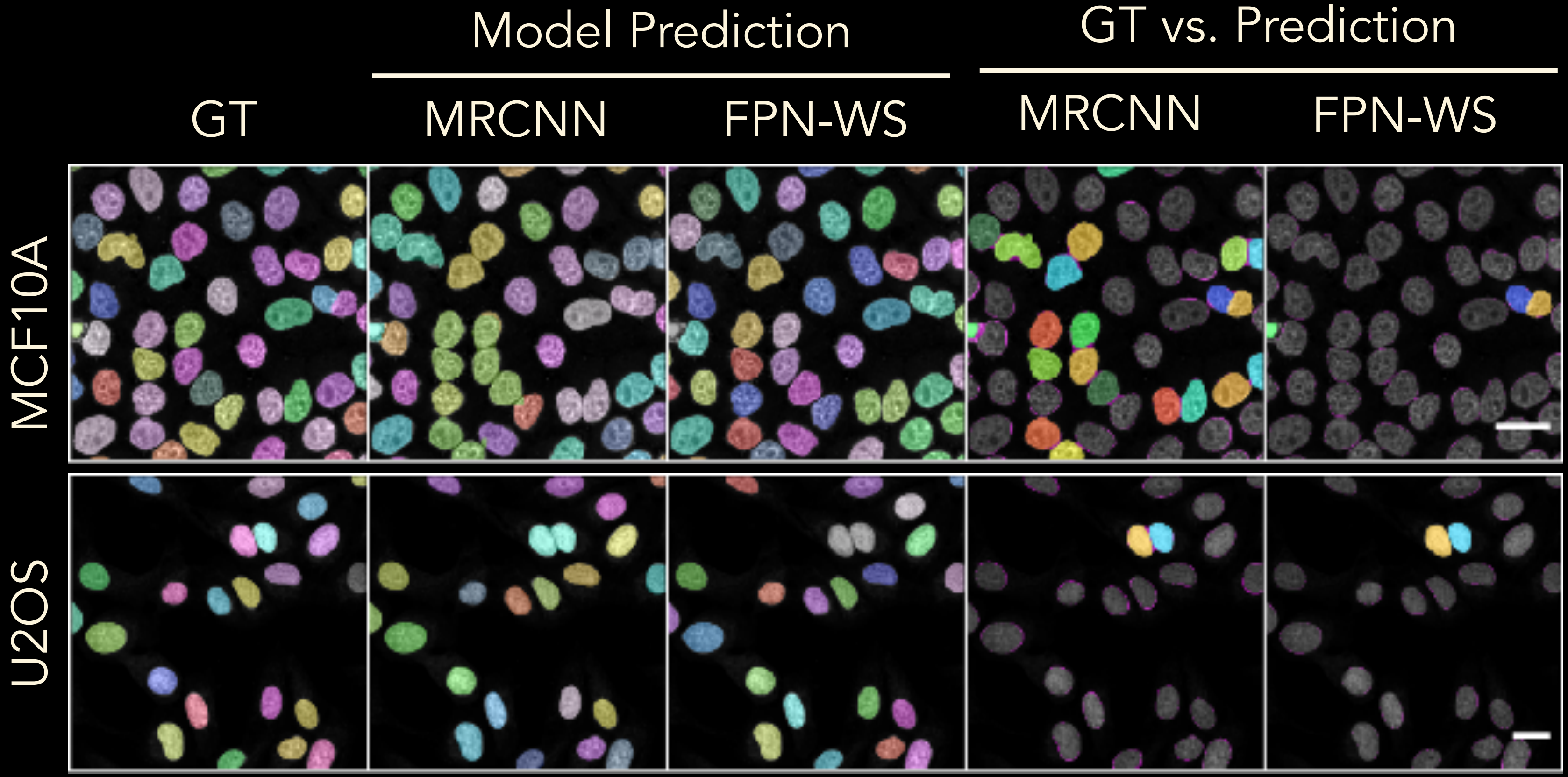
Feature Pyramid Networks (FPN)-Watershed (WS)



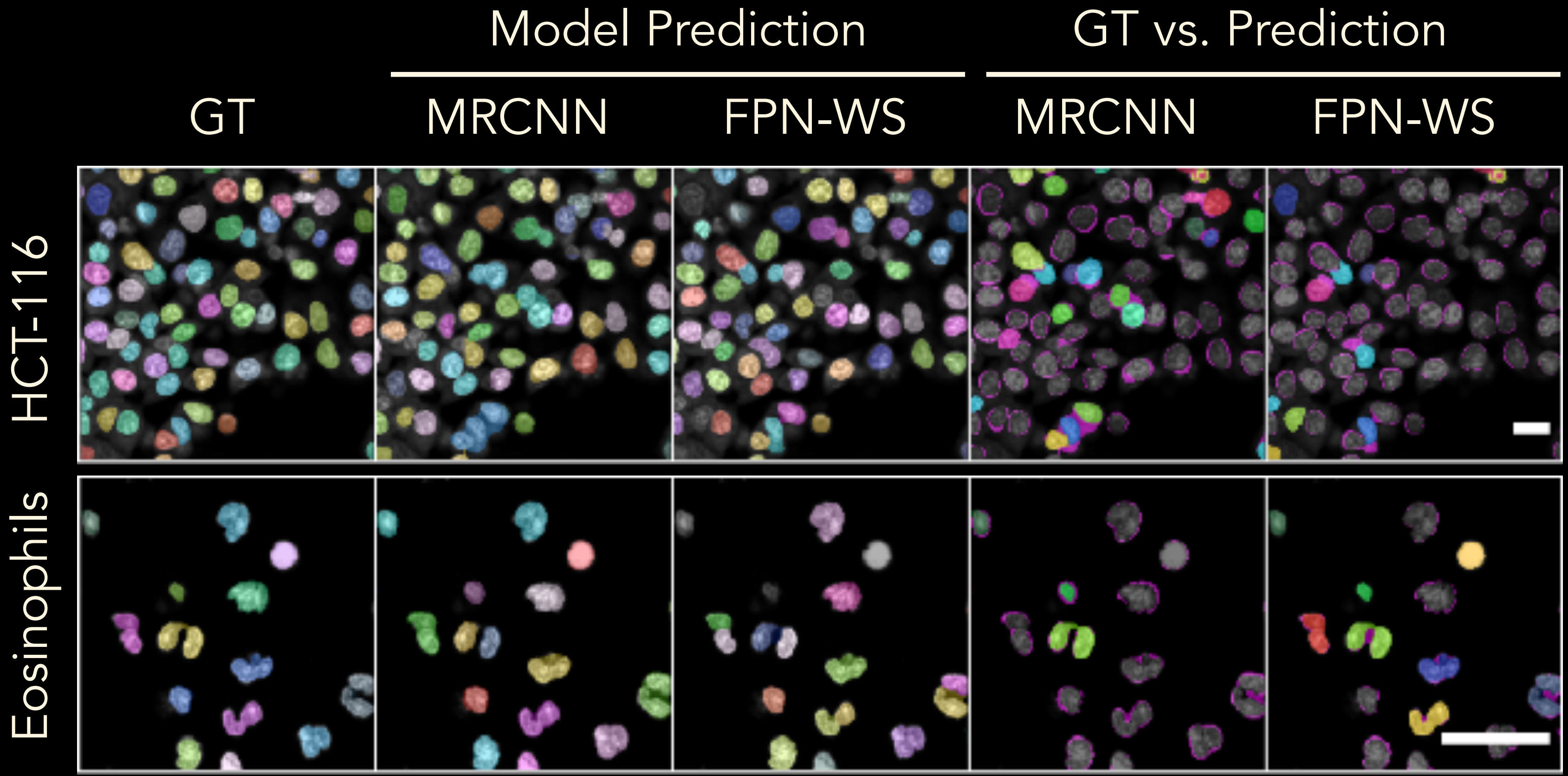
Pipeline for Training and Testing DL Models



DL Models Trained on MCF10A Images (1)



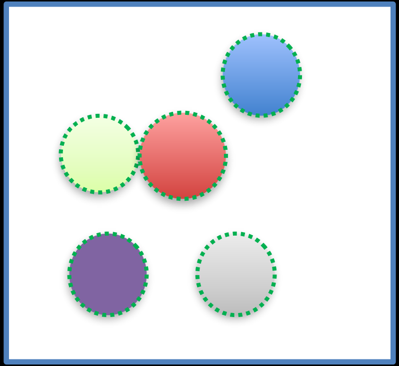
DL Models Trained on MCF10A Images (2)



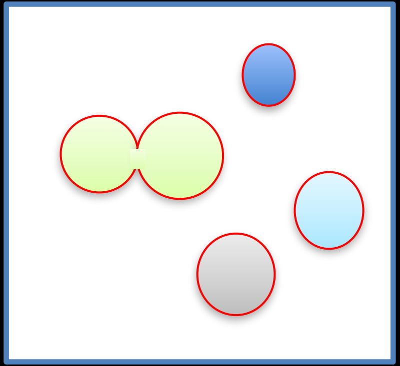
F1 Score to Measure Inference Performance

$$F1(t) = TP(t)/(TP(t) + (FP(t) + FN(t))/2)$$

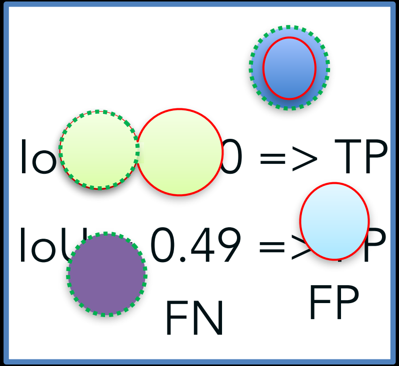
IoU(Threshold) = 0.50

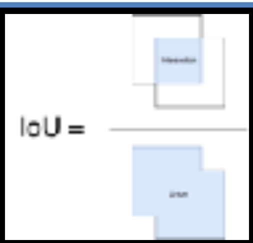


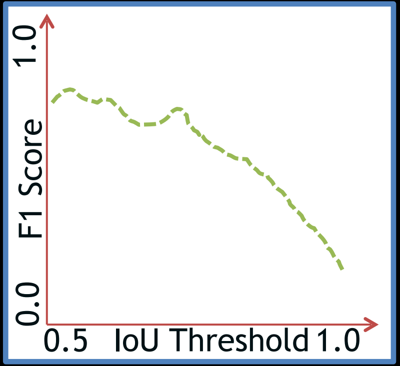
Ground Truth



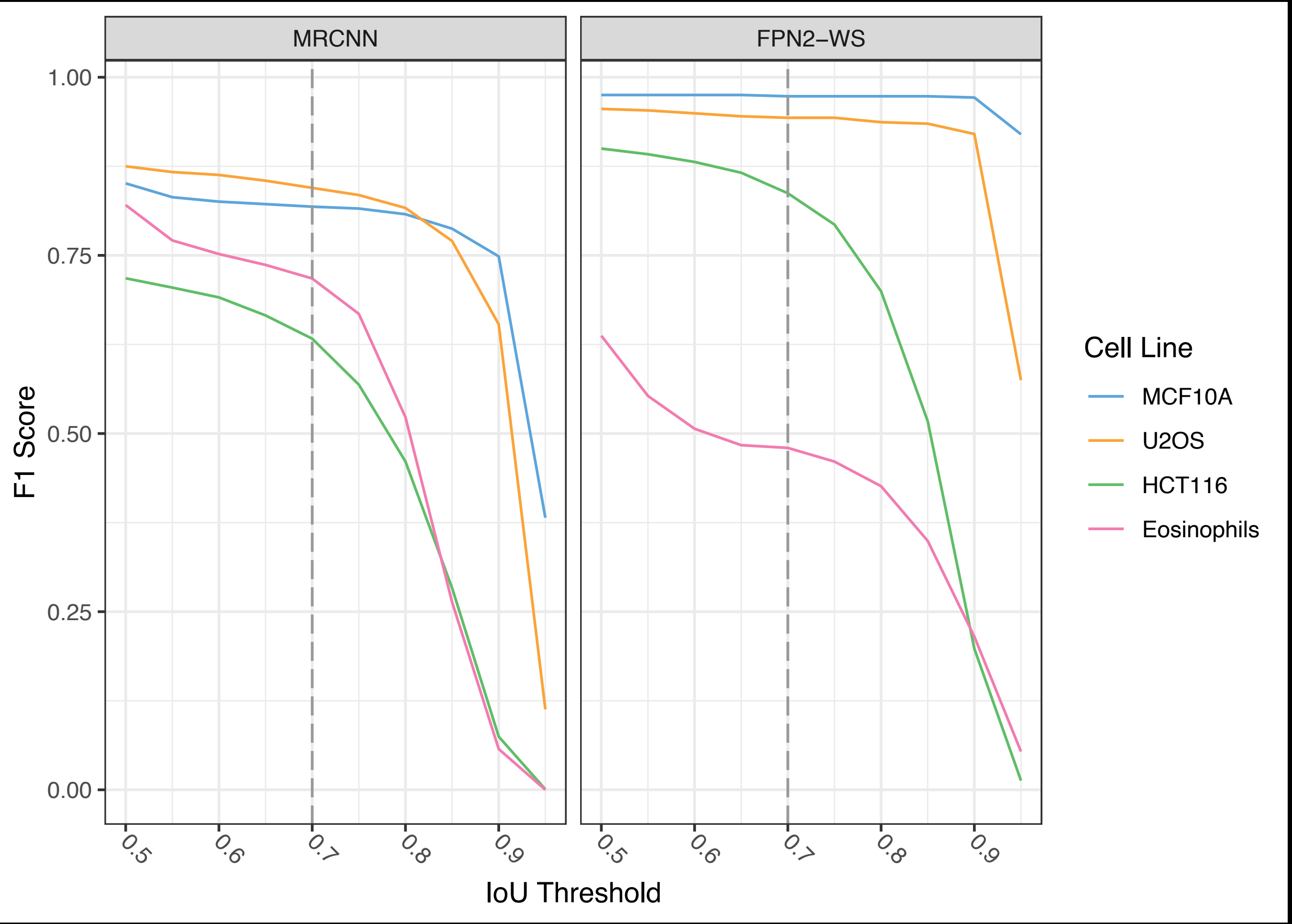
Prediction



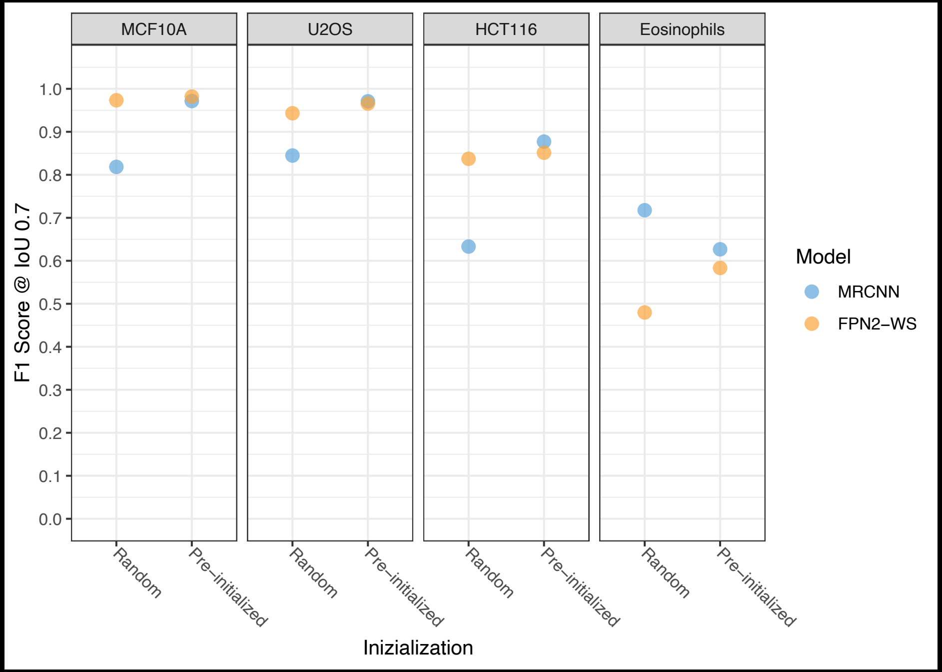




Inference Performance of Baseline DL Models



Transfer Learning Improves MRCNN Performance



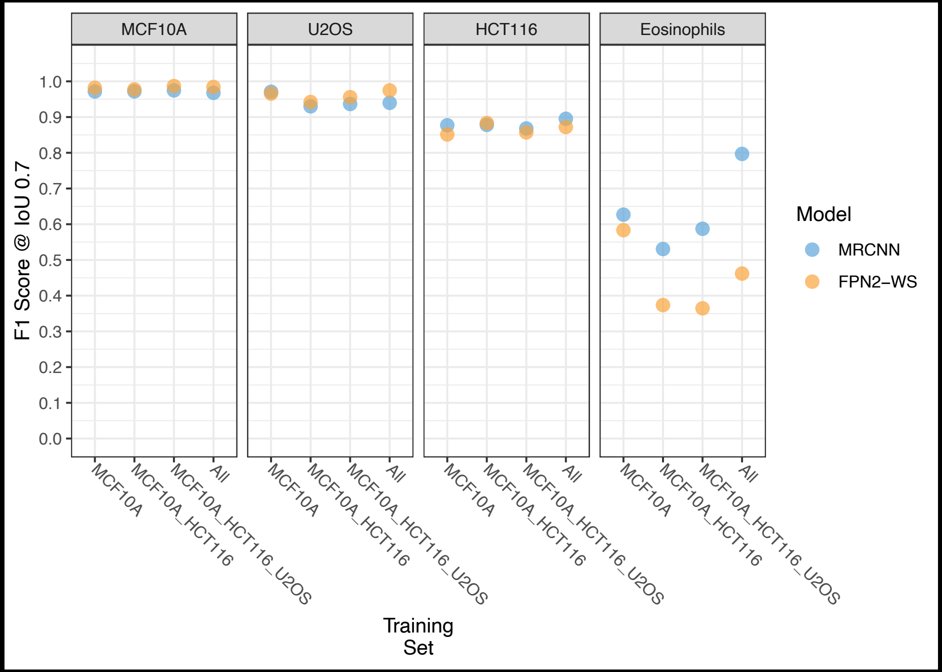
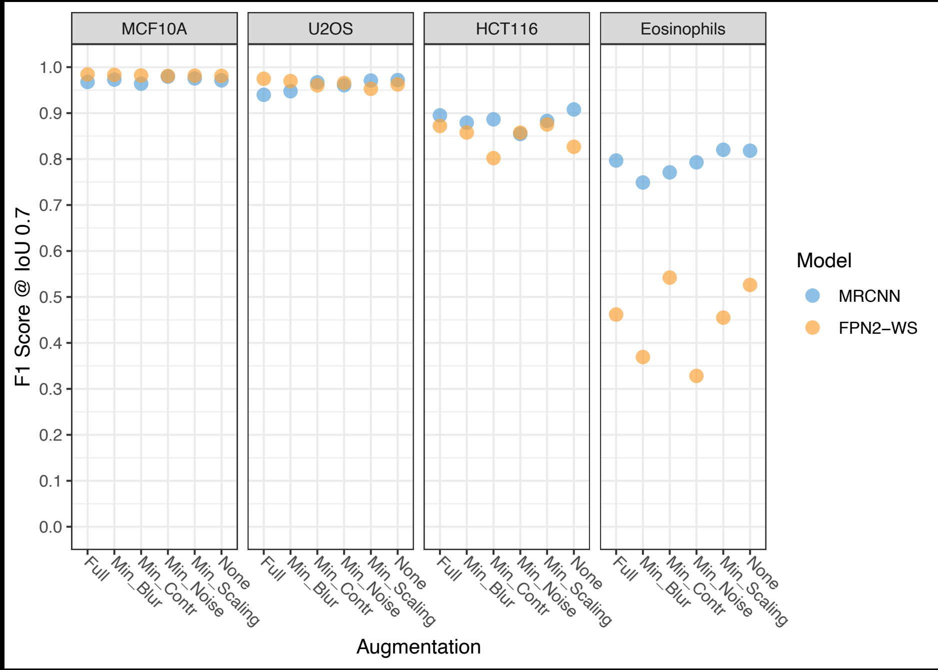
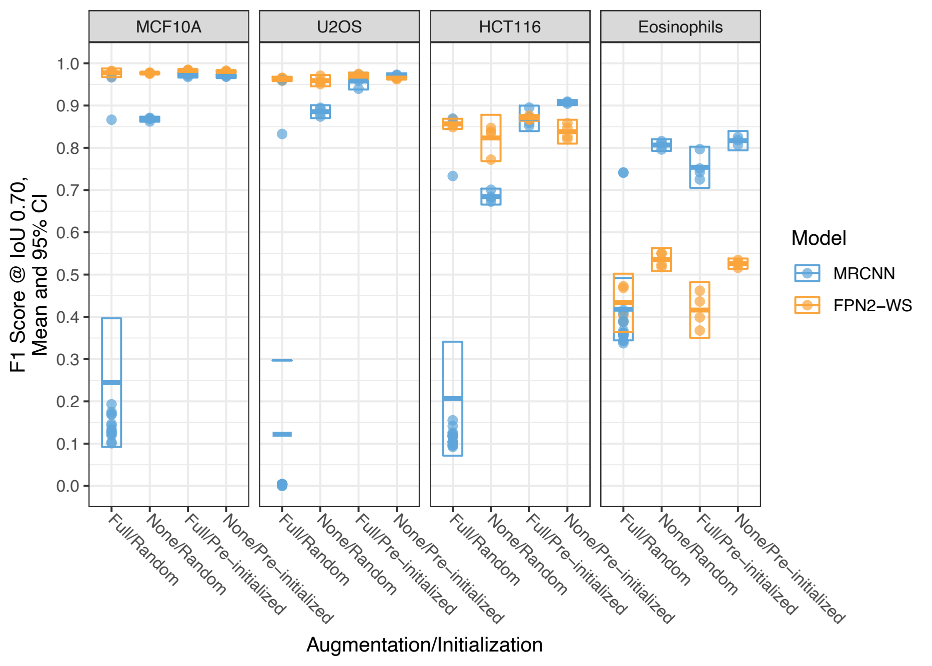
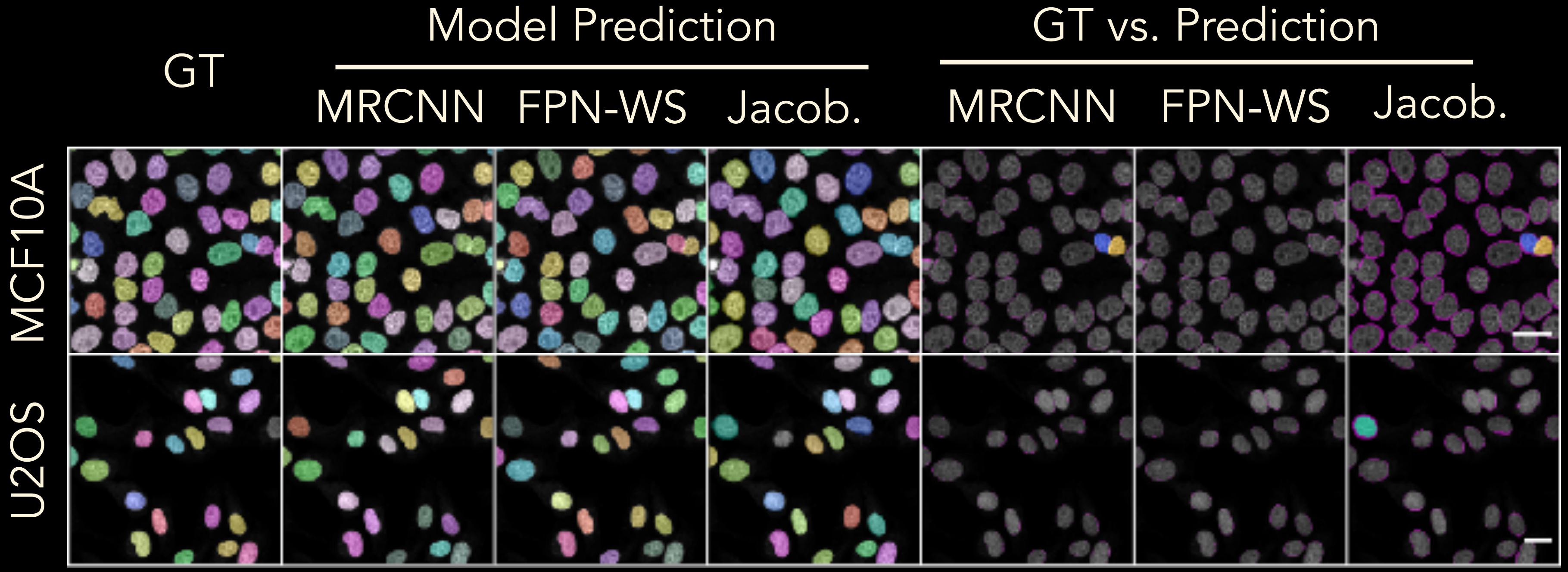


Image Augmentation is not Required

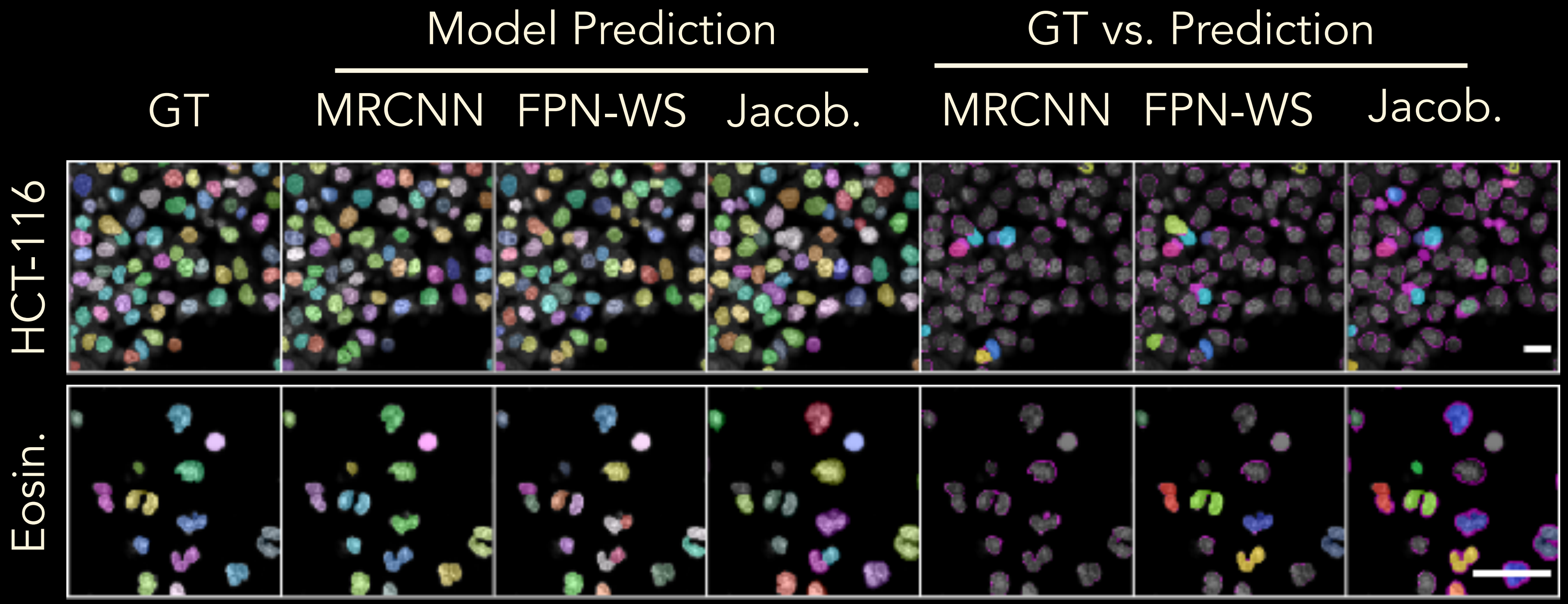




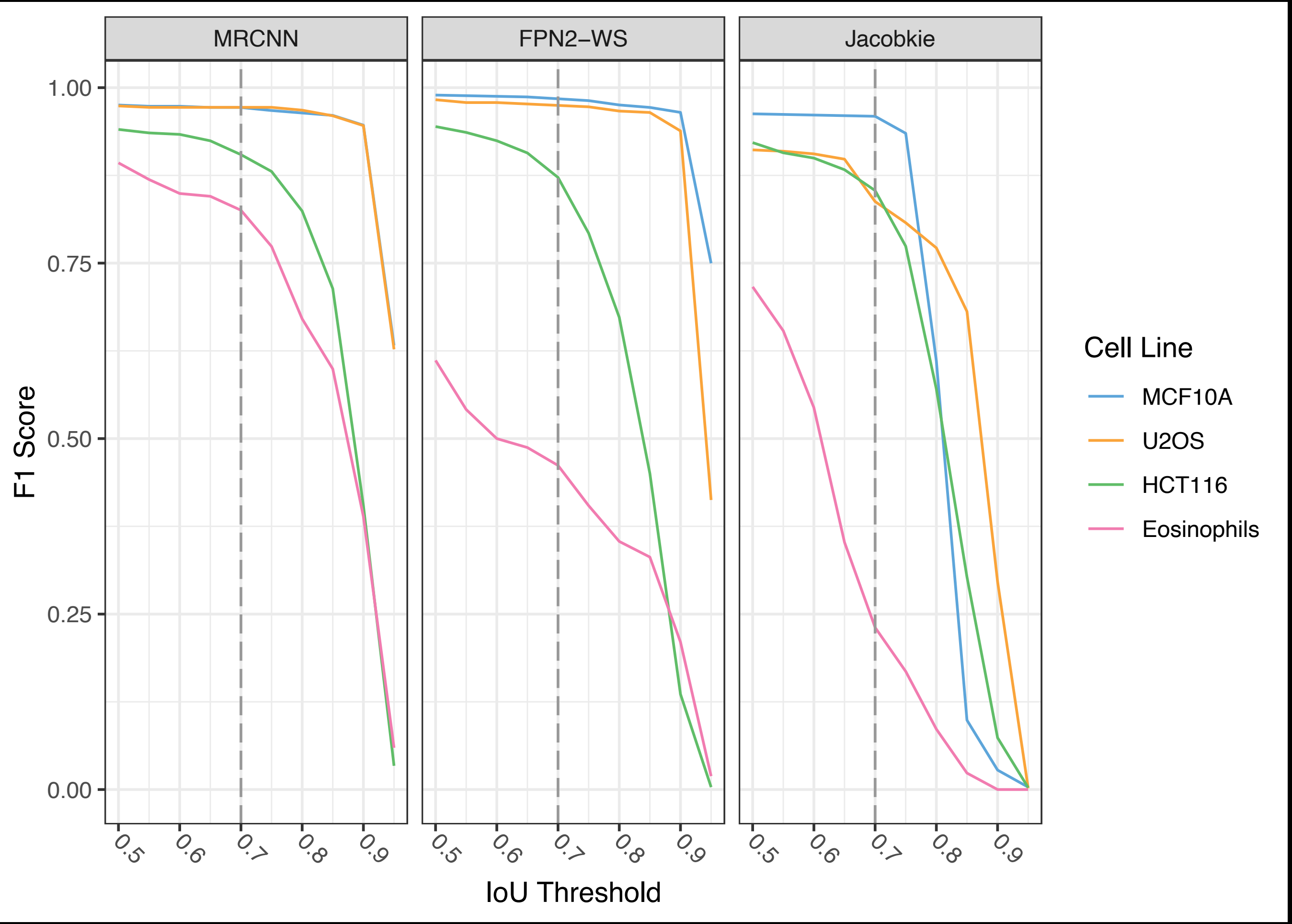
Final DL Models (1)



Final DL Models (2)



Final Models Performance



Summary 1)

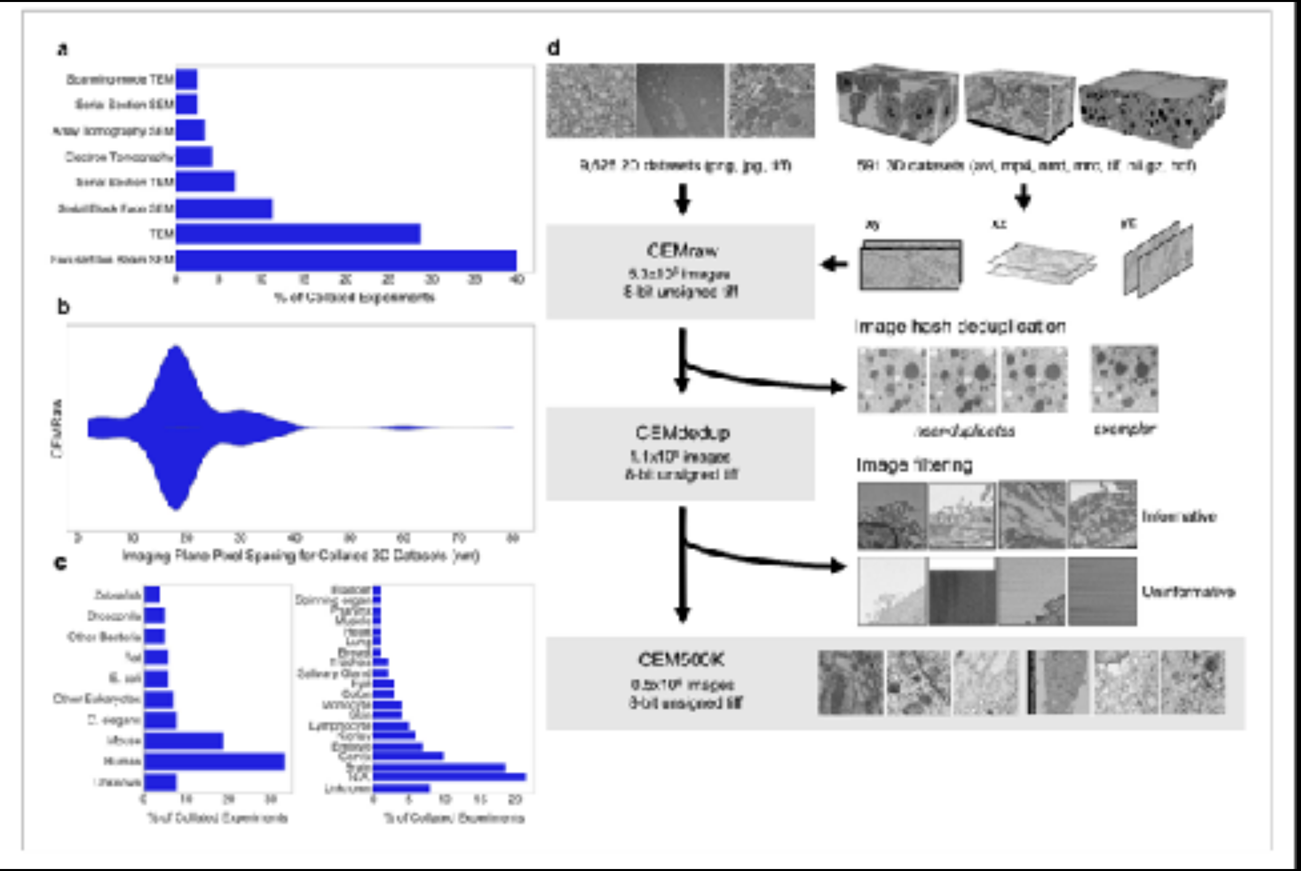
- Semi-automated computational pipeline for DL models training/testing
- Transfer learning can improve performance by using networks weights obtained from training on everyday objects
- Training vs. out of the box: it depends…
- Other DL applications: classification, denoising, inpainting

Future Areas of Improvement for DL in Bioimaging

Datasets → GT Annotation → Model Architecture → Inference

Datasets	GT Annotation	Model Architecture	Inference
Size	Interactive	Accurate	Scalable
Variety	Easy to use	Fast	Cost effective
Quality	Train. Integrated	Generic	Easy to use

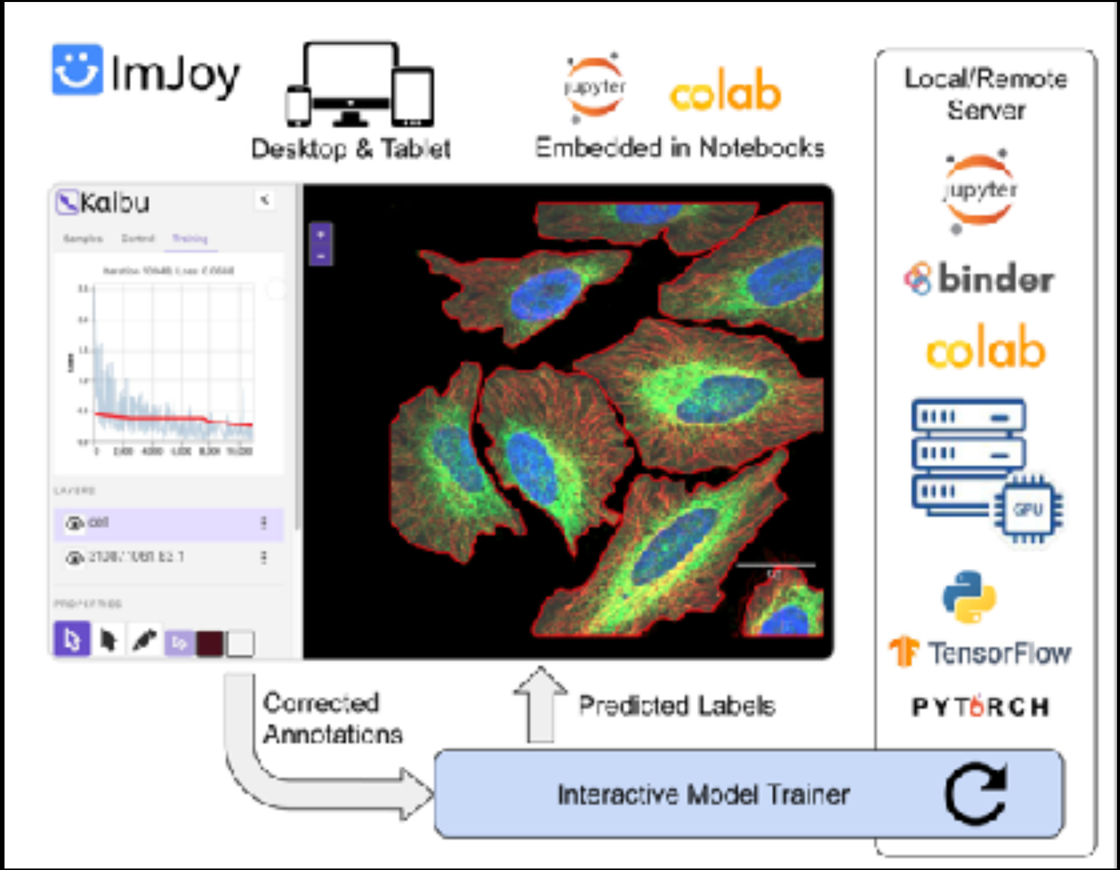
CEM500K

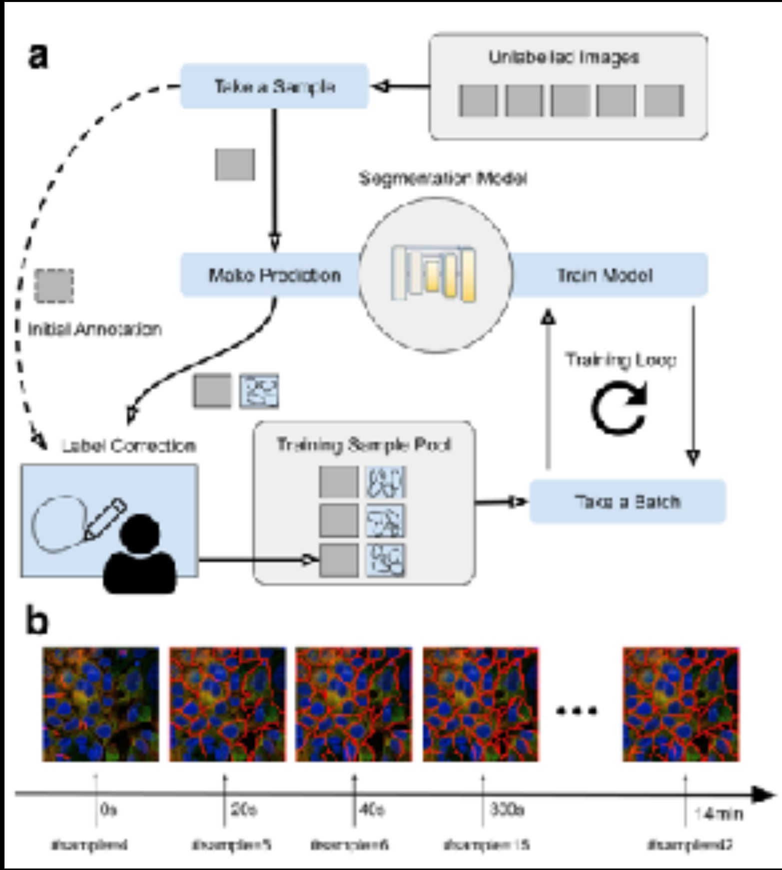


Conrad R. and Narayan K., eLife, 2021

Datasets

Imjoy

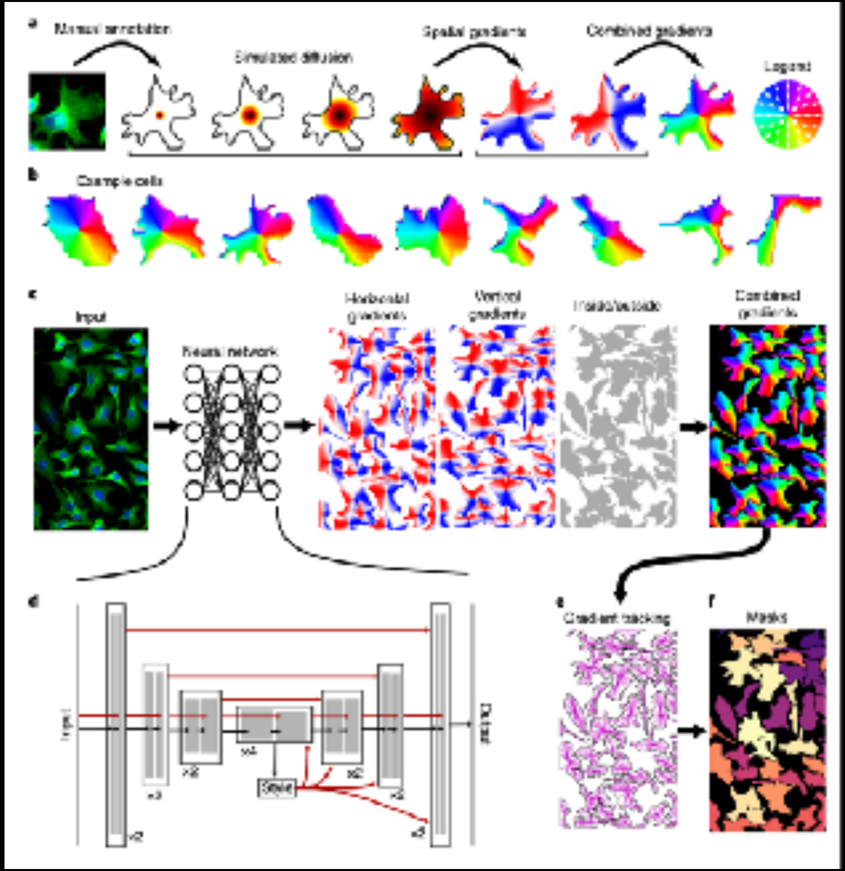


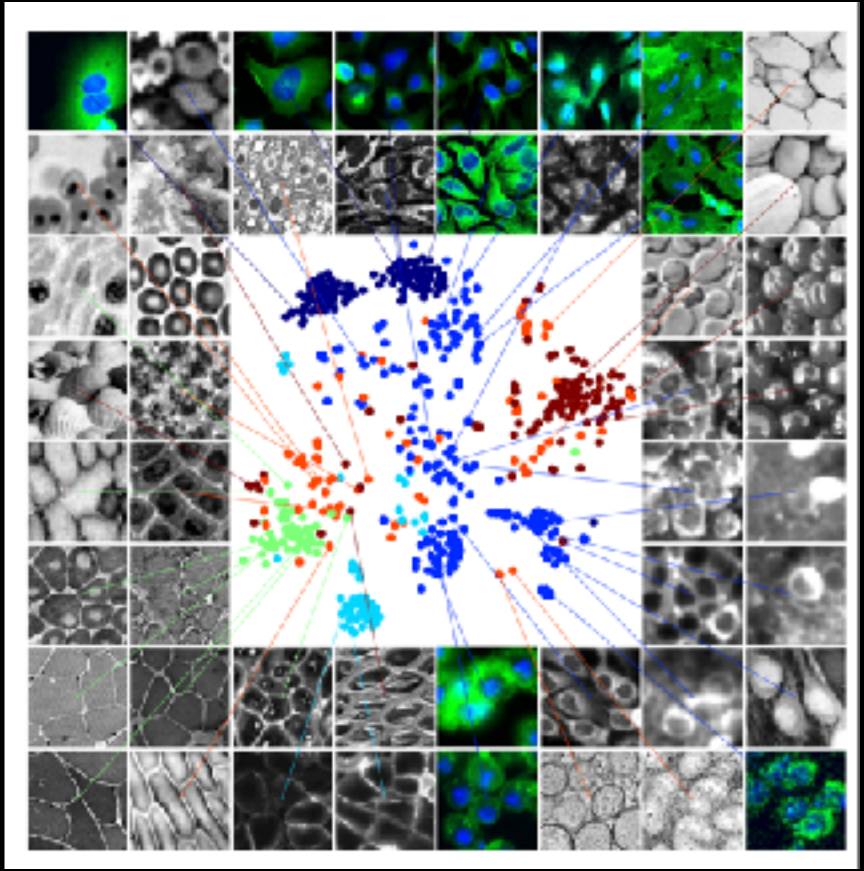


Ouyang W., Nat. Meth., 2019; Ouyang W., F1000Res, 2021

Annotation

CellPose: 2D Segmentation



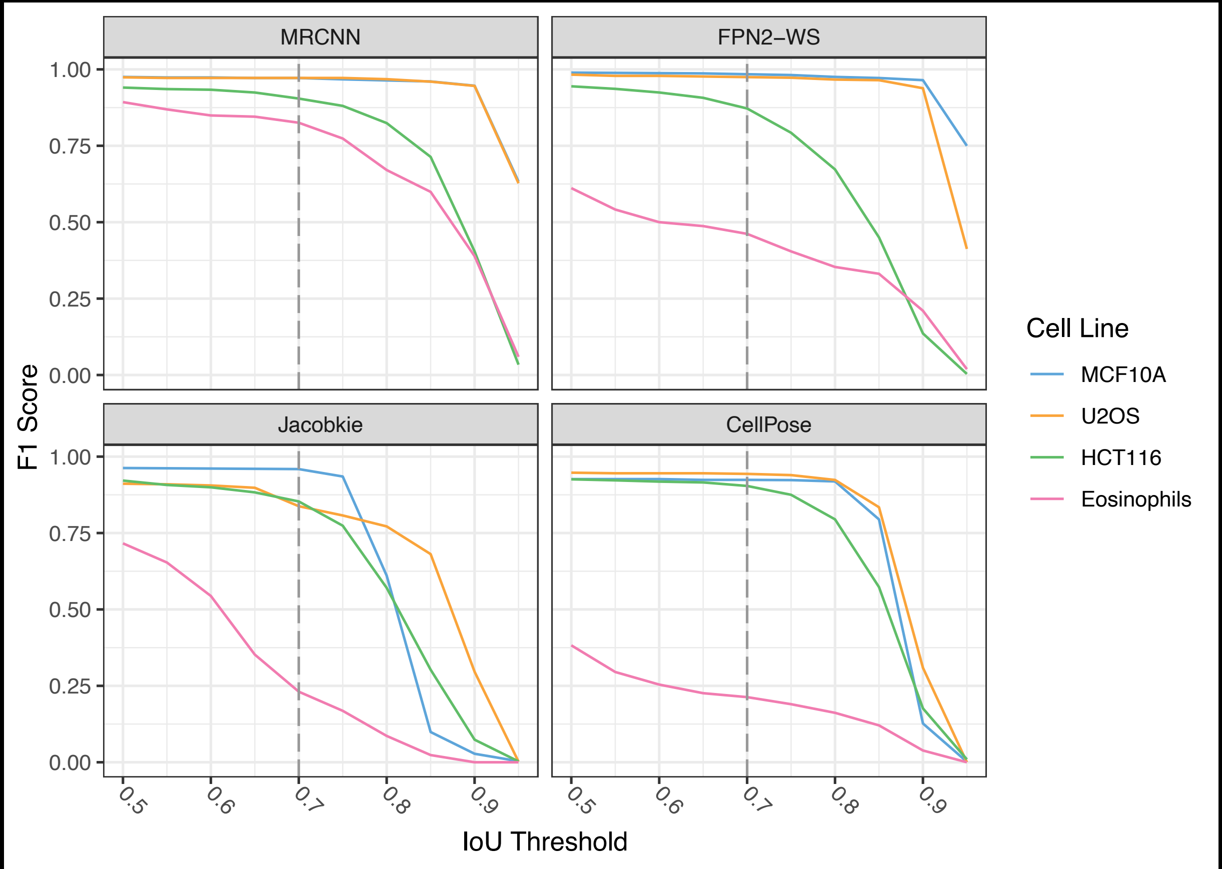


Stringer C., Nat. Meth., 2021

Model

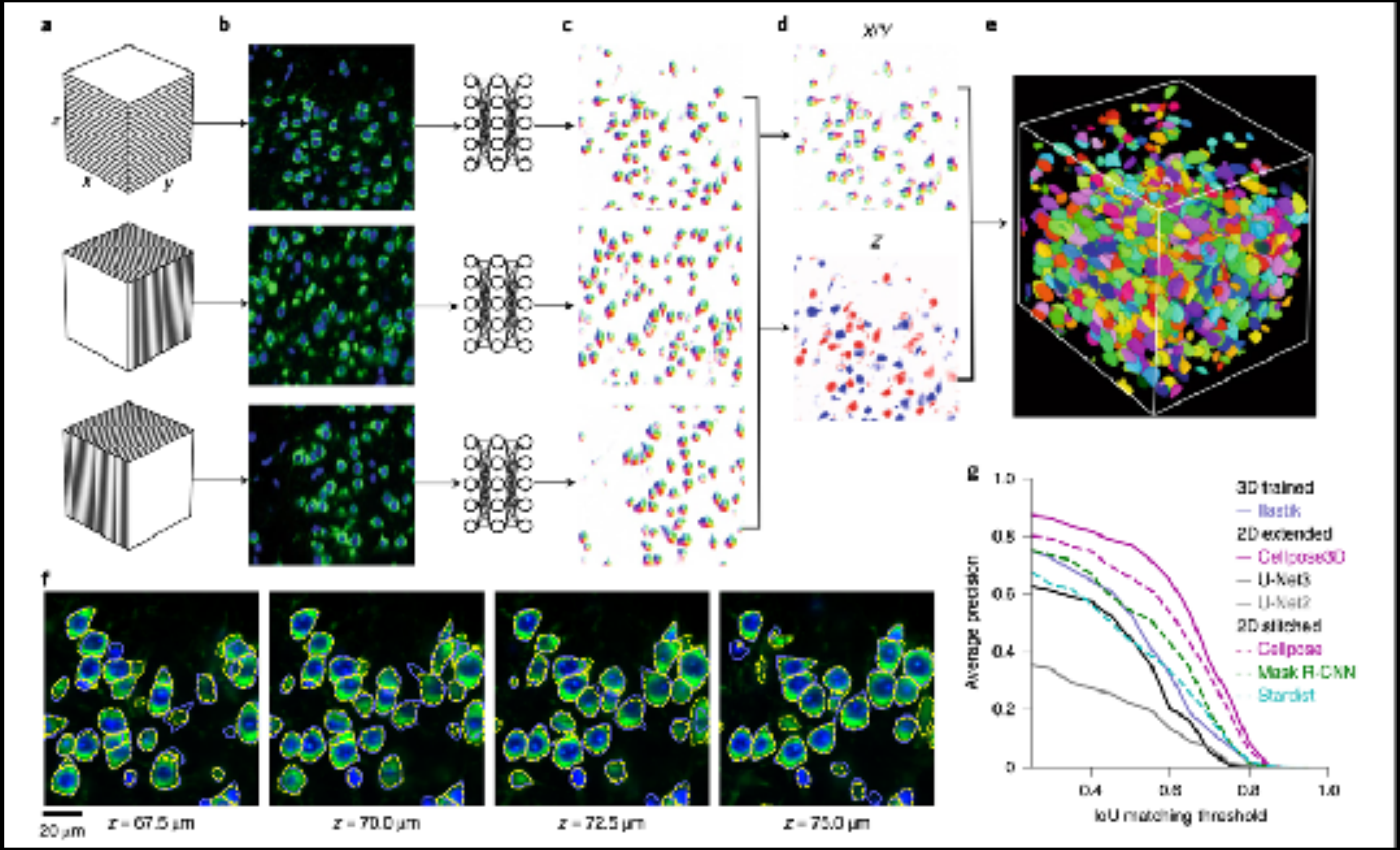
Datasets

CellPose Works “Out of the box”



G. Zaki and A. Keikhosravi

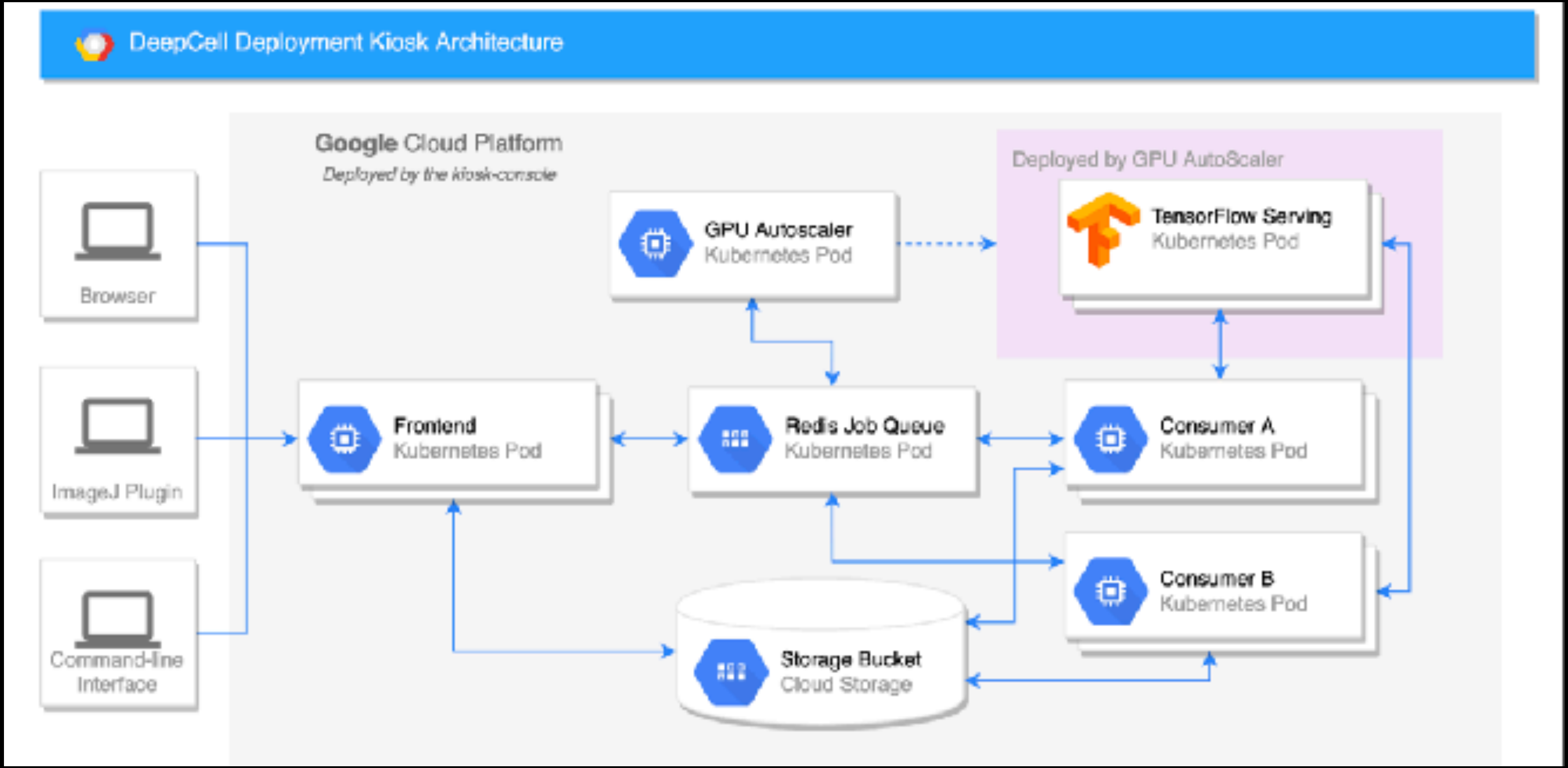
CellPose: 3D Segmentation



Stringer C., Nat. Meth., 2021

Model

Better Tools to Serve Models: Deep Cell Kiosk



Bannon D., Nat. Meth., 2021

Inference

Summary 2)

- Rapid improvements in making DL more accessible for biologists, larger curated datasets, better model architectures, higher-throughput at inference
- Biologists should pair up with ML/DL experts

Acknowledgements

HiTIF

P. Gudla (f)
L. Ozbun
K. Lee (f)
A. Keikhosravi

FNL

G. Zaki
J. Kim

Collaborators

S. Shachar (f)
T. Misteli (CCR)
M. Gadkari
L. Franco (NIAMS)
J. Sung
I. Fraser (NIAD)

NIH HPC Team

NIH NATIONAL CANCER INSTITUTE

Contact Information

HiTIF

Laboratory of Receptor Biology and Gene Expression

Center for Cancer Research/NCI

Gianluca Pegoraro

gianluca.pegoraro@nih.gov

NIH NATIONAL CANCER INSTITUTE