

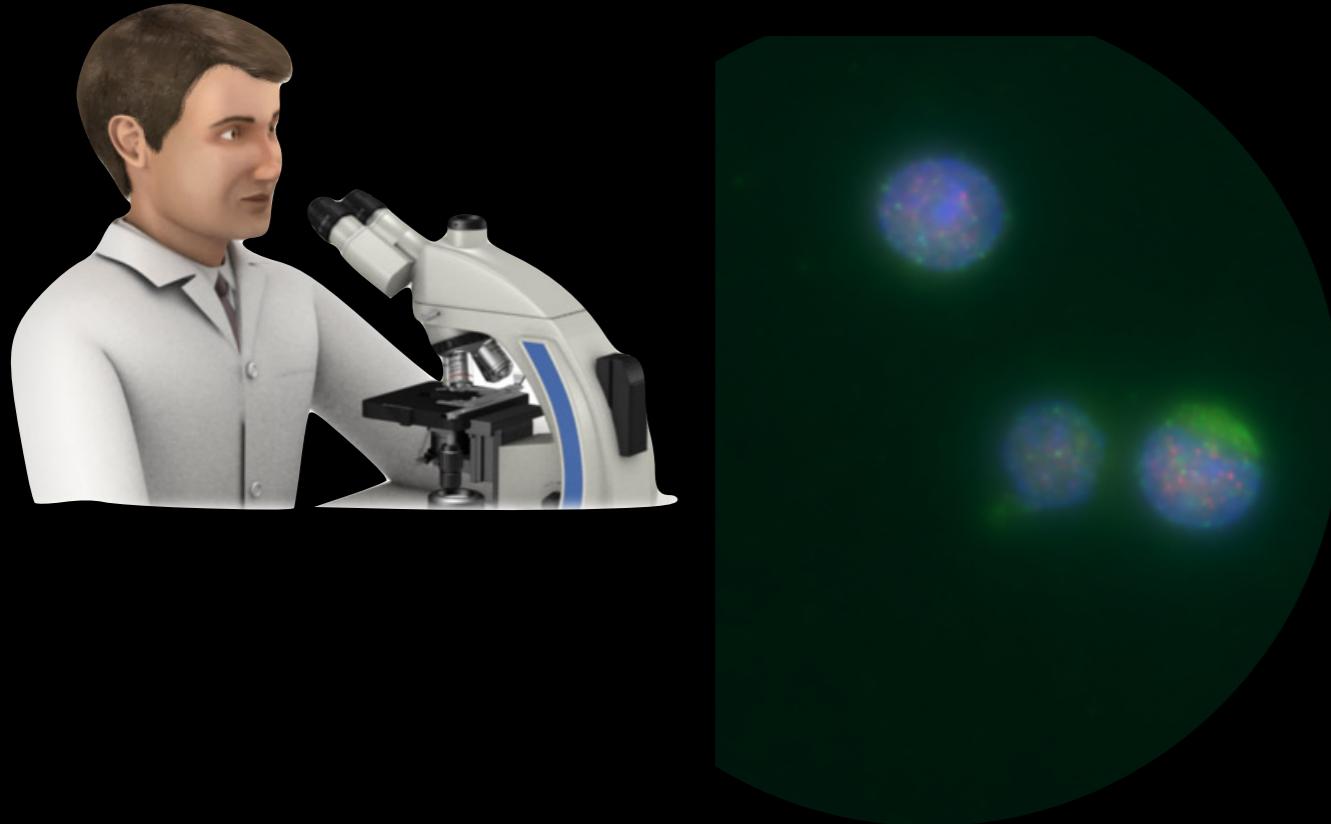
# Automating microscopy acquisition with deep learning and augmented reality.

By Dr. Dominic Waite UKRI Innovation Fellow  
Weatherall Institute of Molecular Medicine, University of Oxford.

Bioimage Analysis in Python Course Dec 2019  
Cambridge



# Microscopy acquisition in life sciences



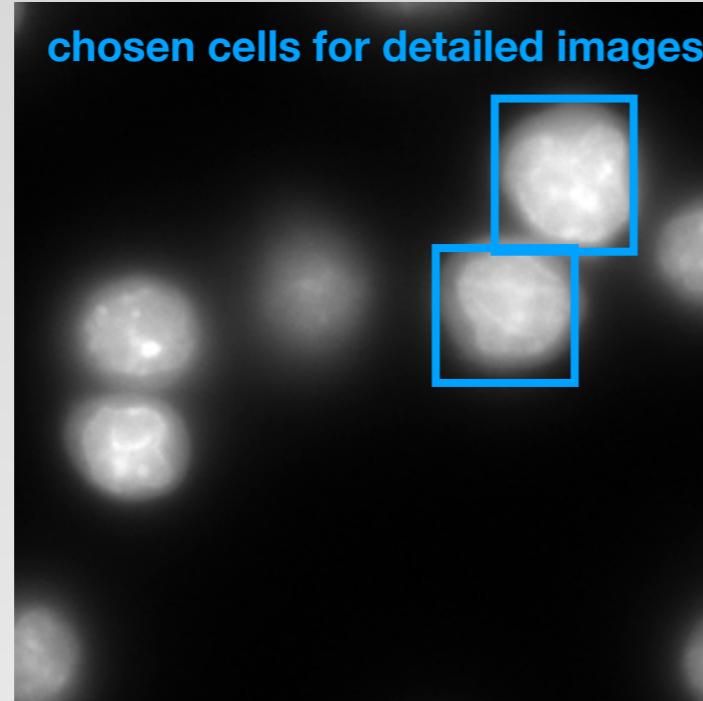
During an experiment we make choices often based on limited information. In microscopy we often don't have time to image all the cells and so we acquire only a sample of the cells present. The choice of cells is often a human decision.

# Motivation for project

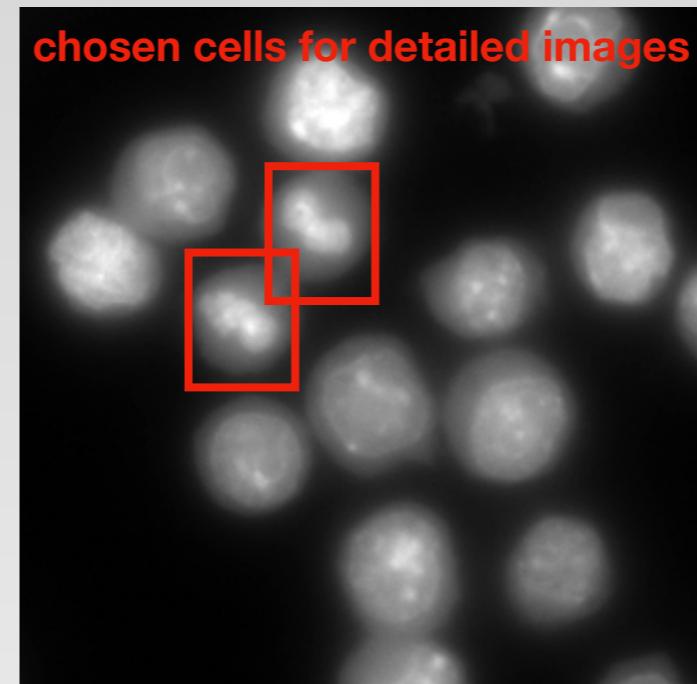
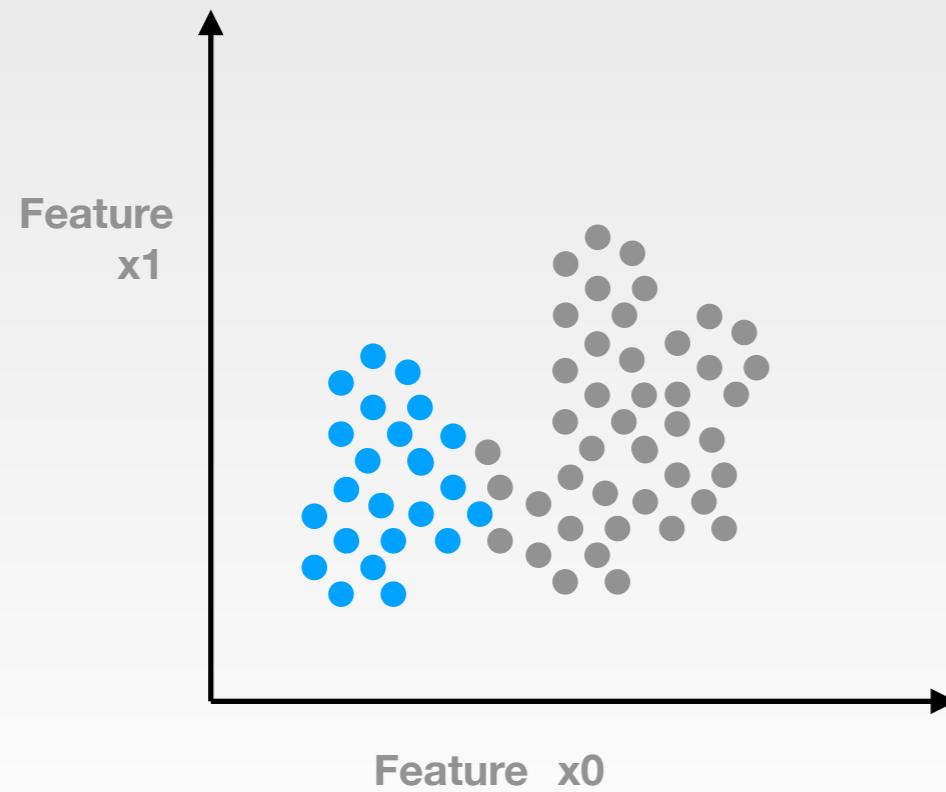
**"Contrary to the rules of philosophers of science, who advise testing hypotheses by trying to refute them, people (and scientists, quite often) seek data that are likely to be compatible with the beliefs they currently hold."**

Therefore we need to keep working on the methodology of science to make this harder to do.

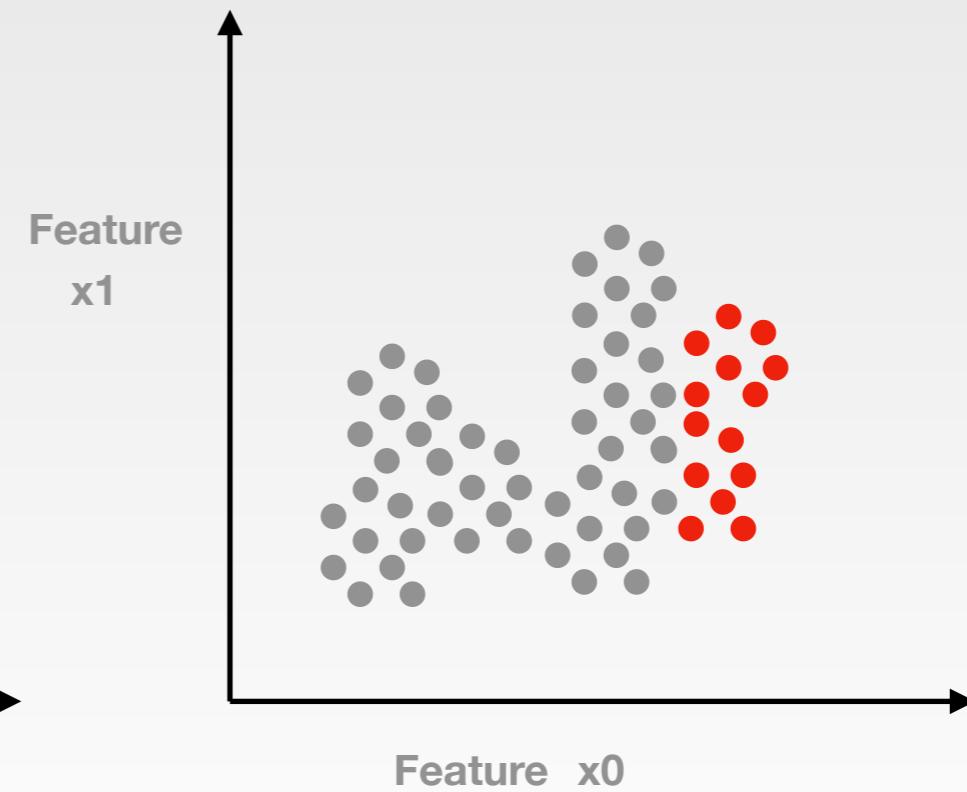
# Humans unconscious are drawn towards differences



Condition A (e.g drug treatment)

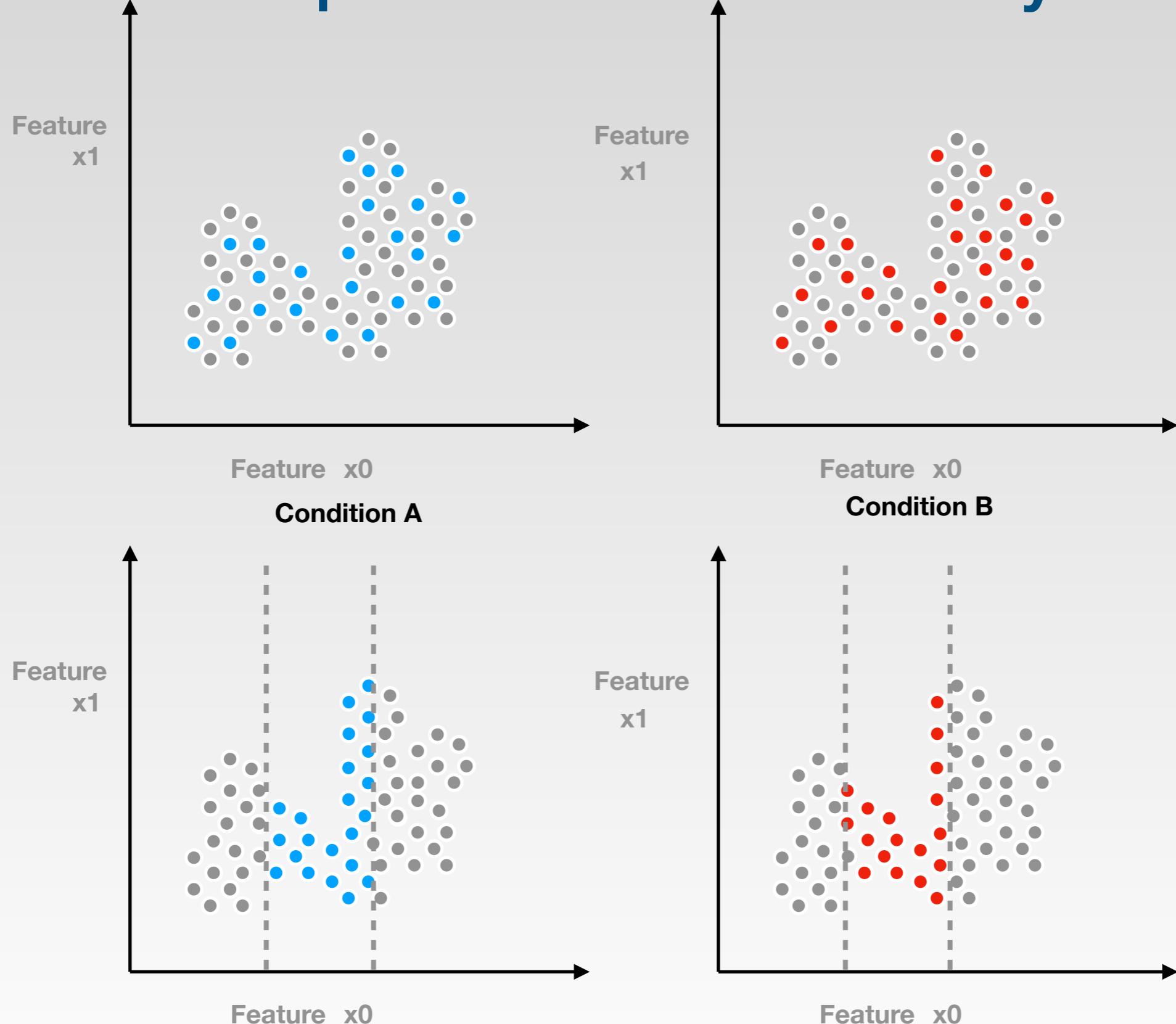


Condition B (e.g. placebo)



Our measurements will appear different but the conditions are the same.

# If we can sample and measure everything.



Now we have an idea of the whole distribution we can sample randomly or specifically according to some feature.

# We have the hardware to do this. This is my microscope



# Potential cell localiser algorithm: Object detection algorithms

arXiv:1506.01497v3 [cs.CV] 6 Jan 2016

73.2 mAP 7 FPS 2015

## Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks

Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun

**Abstract**—State-of-the-art object detection networks depend on region proposal algorithms to hypothesize object locations. Advances like SPPNet [1] and Fast R-CNN [2] have reduced the running time of these detection networks, exposing region proposal computation as a bottleneck. In this work, we introduce a *Region Proposal Network* (RPN) that shares full-image convolutional features with the detection network, thus enabling nearly cost-free region proposals. An RPN is a fully convolutional network that simultaneously predicts object bounds and objectness scores at each position. The RPN is trained end-to-end to generate high-quality region proposals, which are used by Fast R-CNN for detection. We further merge RPN and Fast R-CNN into a single network by sharing their convolutional features—using the recently popular terminology of neural networks with “attention” mechanisms, the RPN component tells the unified network where to look. For the very deep VGG-16 model [3], our detection system has a frame rate of 5fps (including all steps) on a GPU, while achieving state-of-the-art object detection accuracy on PASCAL VOC 2007, 2012, and MS COCO datasets with only 300 proposals per image. In ILSVRC and COCO 2015 competitions, Faster R-CNN and RPN are the foundations of the 1st-place winning entries in several tracks. Code has been made publicly available.

**Index Terms**—Object Detection, Region Proposal, Convolutional Neural Network.

### 1 INTRODUCTION

Recent advances in object detection are driven by the success of region proposal methods (e.g., [4]) and region-based convolutional neural networks (R-CNNs) [5]. Although region-based CNNs were computationally expensive as originally developed in [5], their cost has been drastically reduced thanks to sharing convolutions across proposals [1], [2]. The latest incarnation, Fast R-CNN [2], achieves near real-time rates using very deep networks [3], when ignoring the time spent on region proposals. Now, proposals are the test-time computational bottleneck in state-of-the-art detection systems.

Region proposal methods typically rely on inexpensive features and economical inference schemes. Selective Search [4], one of the most popular methods, greedily merges superpixels based on engineered low-level features. Yet when compared to efficient detection networks [2], Selective Search is an order of magnitude slower, at 2 seconds per image in a CPU implementation. EdgeBoxes [6] currently provides the best tradeoff between proposal quality and speed, at 0.2 seconds per image. Nevertheless, the region proposal step still consumes as much running time as the detection network.

• S. Ren is with University of Science and Technology of China, Hefei, China. This work was done when S. Ren was an intern at Microsoft Research. Email: sren@mail.ustc.edu.cn  
 • K. He and J. Sun are with Visual Computing Group, Microsoft Research. E-mail: {kaihe,jiansun}@microsoft.com  
 • R. Girshick is with Facebook AI Research. The majority of this work was done when R. Girshick was with Microsoft Research. E-mail: rbg@fb.com

One may note that fast region-based CNNs take advantage of GPUs, while the region proposal methods used in research are implemented on the CPU, making such runtime comparisons inequitable. An obvious way to accelerate proposal computation is to re-implement it for the GPU. This may be an effective engineering solution, but re-implementation ignores the down-stream detection network and therefore misses important opportunities for sharing computation.

In this paper, we show that an algorithmic change—computing proposals with a deep convolutional neural network—leads to an elegant and effective solution where proposal computation is nearly cost-free given the detection network’s computation. To this end, we introduce novel *Region Proposal Networks* (RPNs) that share convolutional layers with state-of-the-art object detection networks [1], [2]. By sharing convolutions at test-time, the marginal cost for computing proposals is small (e.g., 10ms per image).

Our observation is that the convolutional feature maps used by region-based detectors, like Fast R-CNN, can also be used for generating region proposals. On top of these convolutional features, we construct an RPN by adding a few additional convolutional layers that simultaneously regress region bounds and objectness scores at each location on a regular grid. The RPN is thus a kind of fully convolutional network (FCN) [7] and can be trained end-to-end specifically for the task for generating detection proposals.

RPNs are designed to efficiently predict region proposals with a wide range of scales and aspect ratios. In contrast to prevalent methods [8], [9], [1], [2] that use

1

arXiv:1612.08242v1 [cs.CV] 25 Dec 2016

76.8 mAP 67 FPS 2016

## YOLO9000: Better, Faster, Stronger

Joseph Redmon<sup>\*†</sup>, Ali Farhadi<sup>\*†</sup>

University of Washington\*, Allen Institute for AI†

<http://pjreddie.com/yolo9000/>

### Abstract

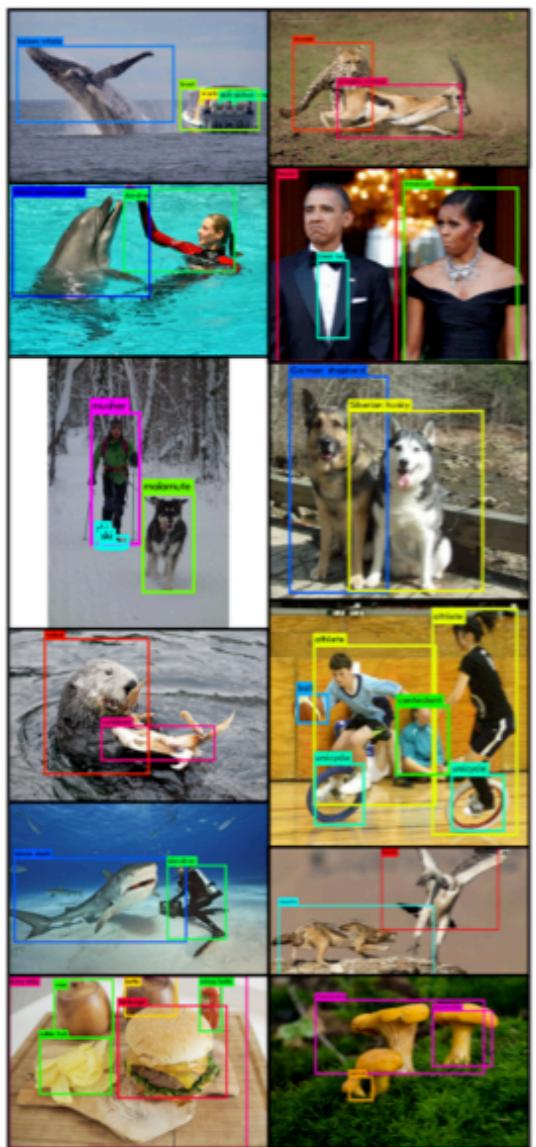
We introduce YOLO9000, a state-of-the-art, real-time object detection system that can detect over 9000 object categories. First we propose various improvements to the YOLO detection method, both novel and drawn from prior work. The improved model, YOLOv2, is state-of-the-art on standard detection tasks like PASCAL VOC and COCO. Using a novel, multi-scale training method the same YOLOv2 model can run at varying sizes, offering an easy tradeoff between speed and accuracy. At 67 FPS, YOLOv2 gets 76.8 mAP on VOC 2007. At 40 FPS, YOLOv2 gets 78.6 mAP, outperforming state-of-the-art methods like Faster R-CNN with ResNet and SSD while still running significantly faster. Finally we propose a method to jointly train on object detection and classification. Using this method we train YOLO9000 simultaneously on the COCO detection dataset and the ImageNet classification dataset. Our joint training allows YOLO9000 to predict detections for object classes that don’t have labelled detection data. We validate our approach on the ImageNet detection task. YOLO9000 gets 19.7 mAP on the ImageNet detection validation set despite only having detection data for 44 of the 200 classes. On the 156 classes not in COCO, YOLO9000 gets 16.0 mAP. But YOLO can detect more than just 200 classes; it predicts detections for more than 9000 different object categories. And it still runs in real-time.

### 1. Introduction

General purpose object detection should be fast, accurate, and able to recognize a wide variety of objects. Since the introduction of neural networks, detection frameworks have become increasingly fast and accurate. However, most detection methods are still constrained to a small set of objects.

Current object detection datasets are limited compared to datasets for other tasks like classification and tagging. The most common detection datasets contain thousands to hundreds of thousands of images with dozens to hundreds of tags [3] [10] [2]. Classification datasets have millions of images with tens or hundreds of thousands of categories [20] [2].

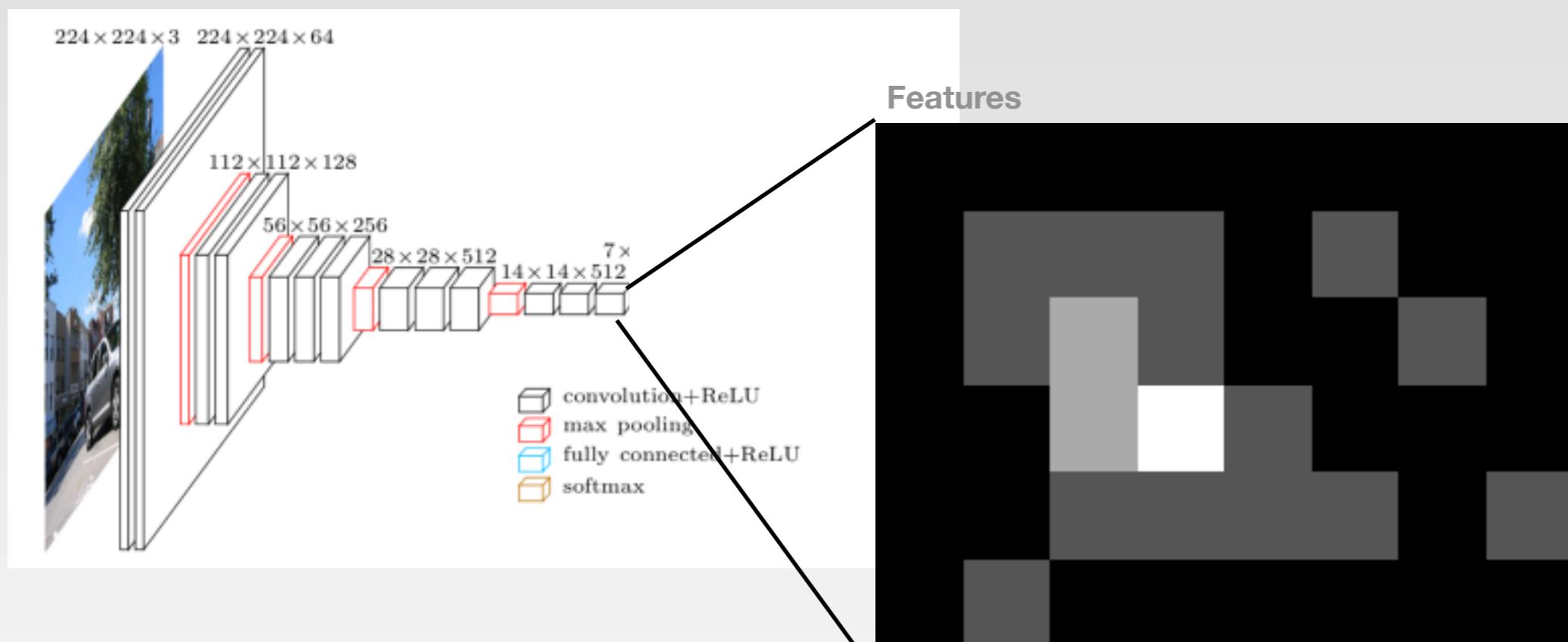
We would like detection to scale to level of object classification. However, labelling images for detection is far more expensive than labelling for classification or tagging (tags are often user-supplied for free). Thus we are unlikely



**Figure 1: YOLO9000.** YOLO9000 can detect a wide variety of object classes in real-time.

# Feature representation

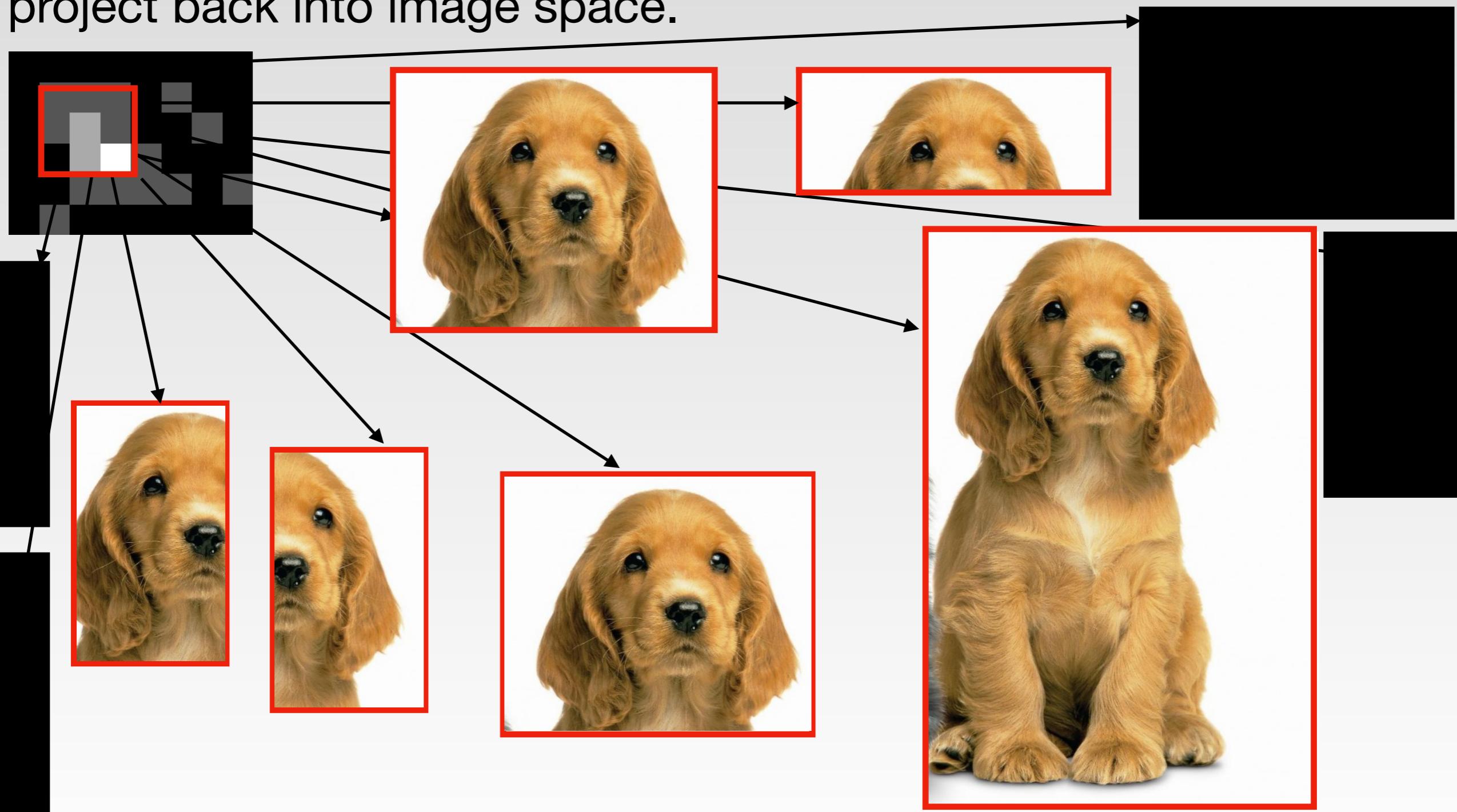
- Using a deep convolutional neural network we calculate high-level features of our image



- Max pooling means our feature map is much smaller spatially compared to our input image. Each pixel corresponds to a square region of input image.

# For each position we generate a number of bounding boxes

- For each spatial position in the CNN feature map we have a generate a large number of bounding boxes (n anchors) and project back into image space.

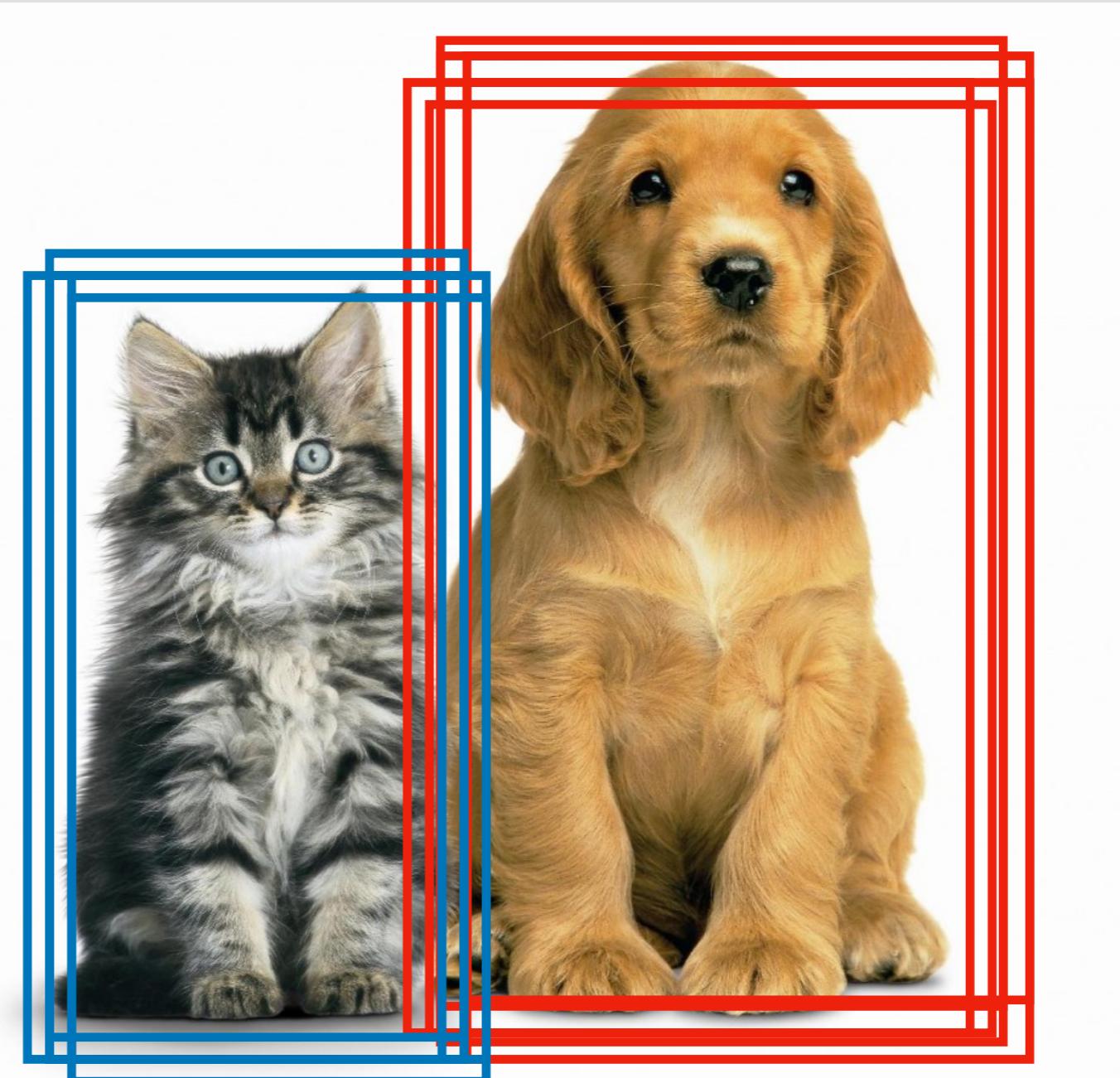
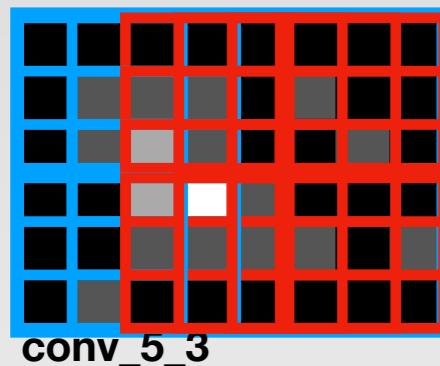


- However it is unlikely that any one of these regions will give us the perfect bounding box

WaiteD Source:

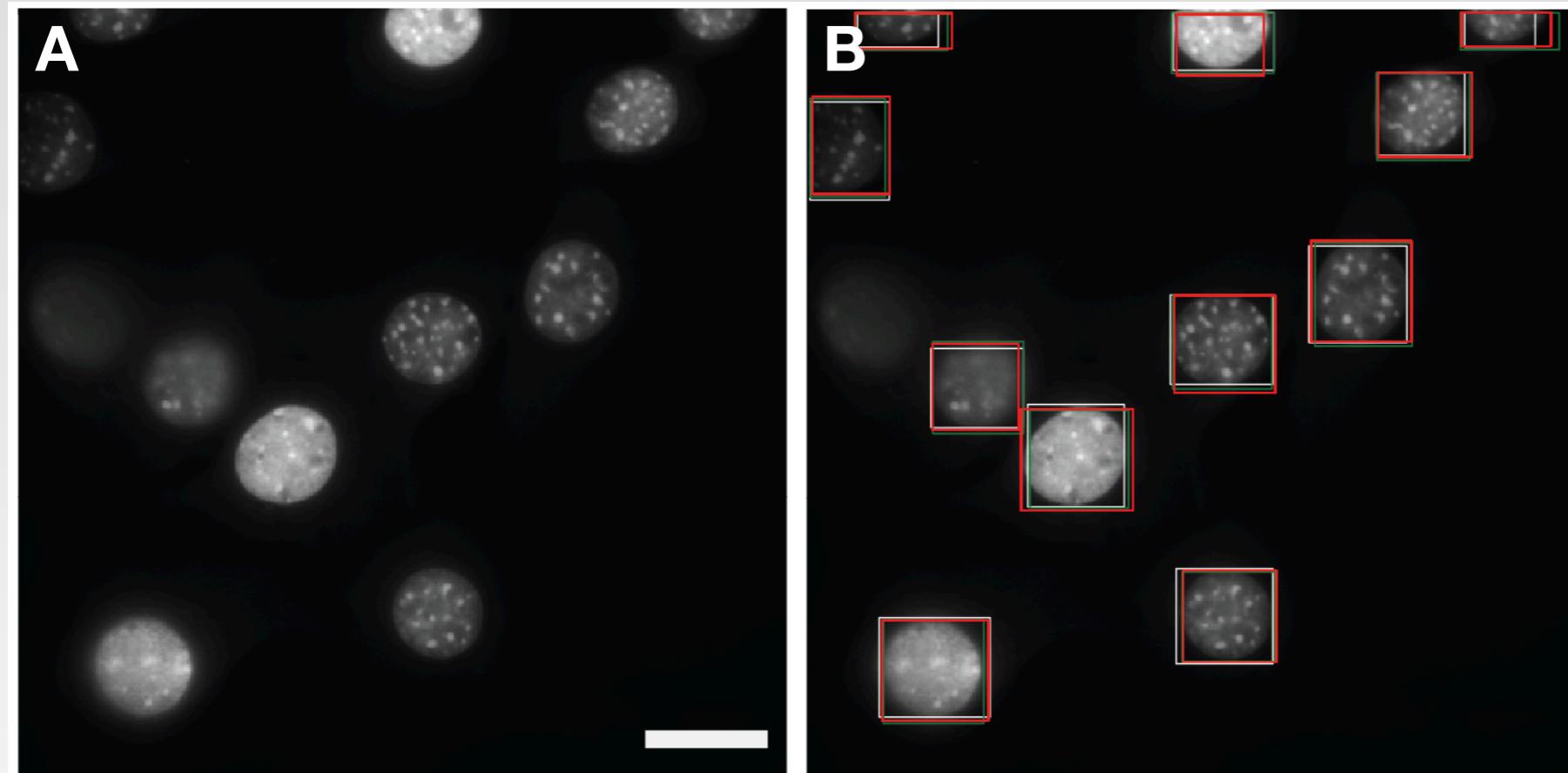
# Then we learn a correction also.

- Based on our feature space we learn a correction for each of our anchors. to make it fit the objects



# Training images for microscopy

C127 cells with dapi staining.

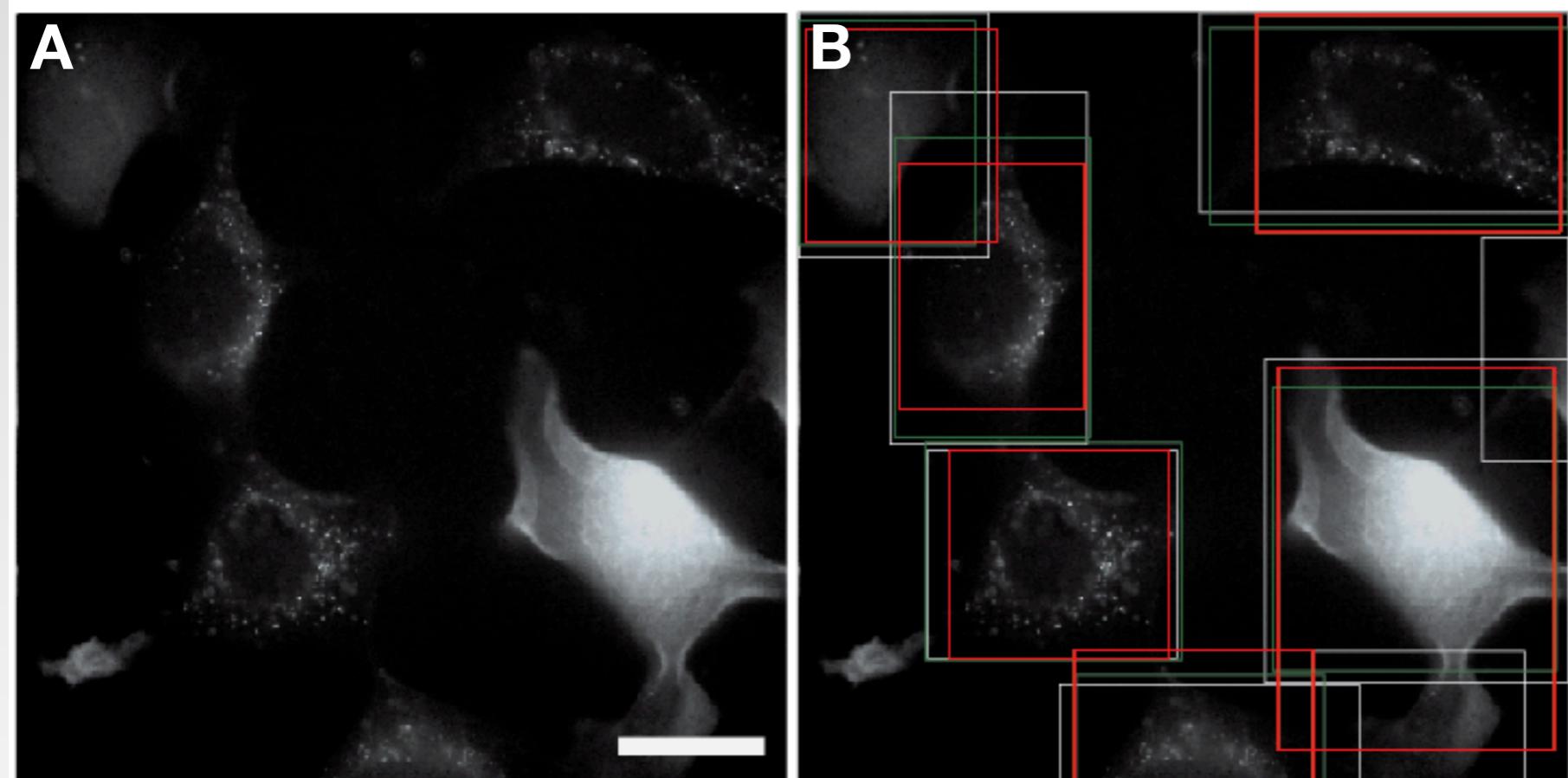


min = 0, max = 255.0, scale bar (25  $\mu$ m).  
30 training images and 30 test images.  
white boxes (ground-truth),  
green boxes Faster-RCNN prediction  
red boxes YOLO v2. prediction

Source: Description: C127 cells stained with dapi. 108651.jpg 108639.jpg  
TADA DIG+ TADB 594. 2 x2 bin 10 ms exp

# Additional Datasets: Hela Peroxisomes.

Hela Peroxisome (GFP-SCP2) stained. Very challenging.



min = 0, max = 33.0, scale bar (25 µm).

55 training images and 55 test images.

white boxes (ground-truth),

green boxes Faster-RCNN prediction

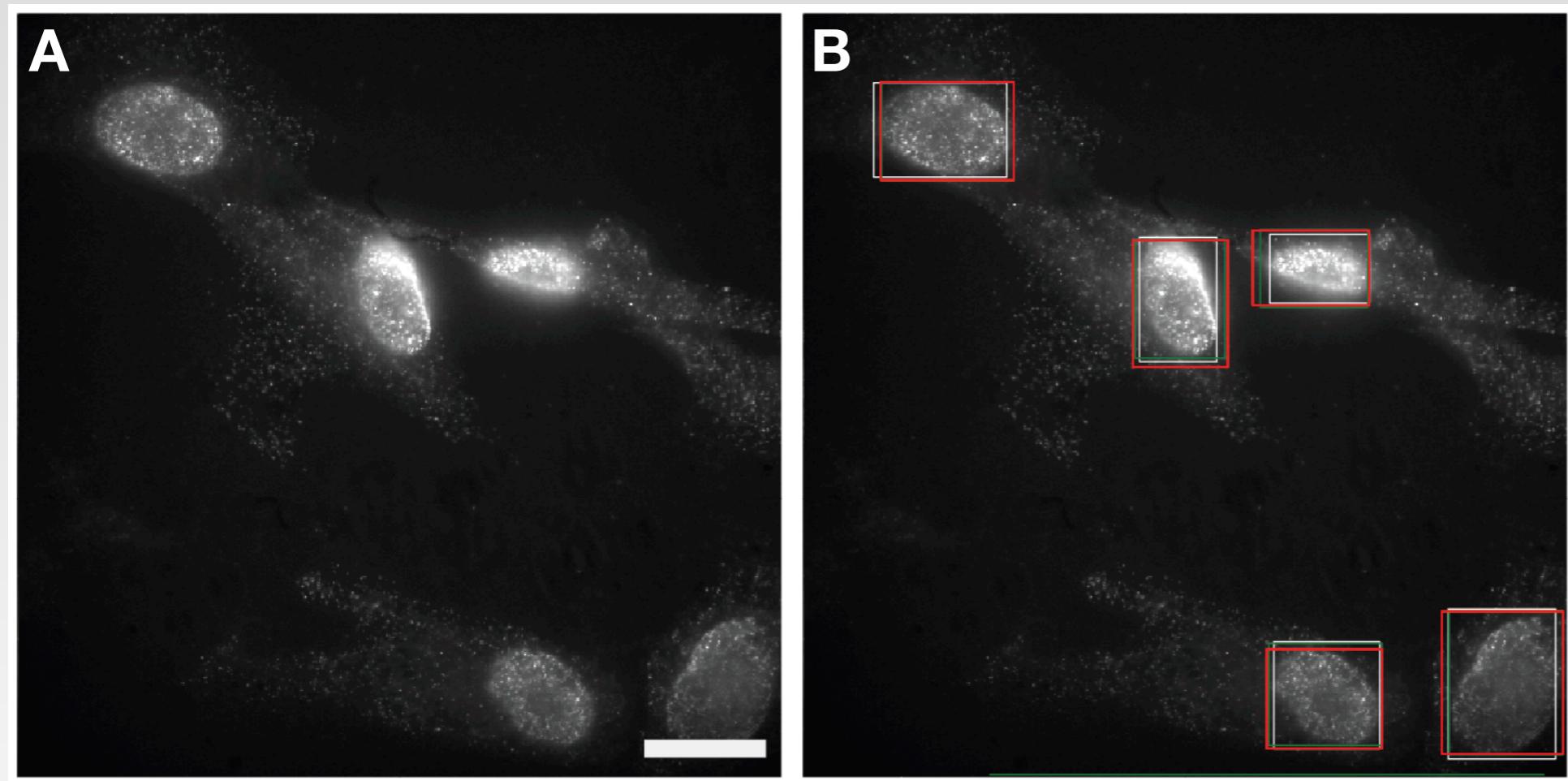
red boxes YOLO v2. prediction

Source: Slides prepared by Katharina imaged on wide-field PicoQuant system. 55 training and 55 test images.

illumination: protein GFP-SCP2, 2x2 bin, 10 ms exposure resized to 250 nm/pixel. NAME = 107592.JPG, 107637.JPG. Fixed cells.

# Additional Datasets: Fibroblast Nucleopore

Anti-Nup153 mouse antibody (Abcam) and counterstained using Alex Fluor 488 (anti-mouse). Small dataset.



min = 0, max = 255.0, scale bar (25 µm).

26 training images and 20 test images.

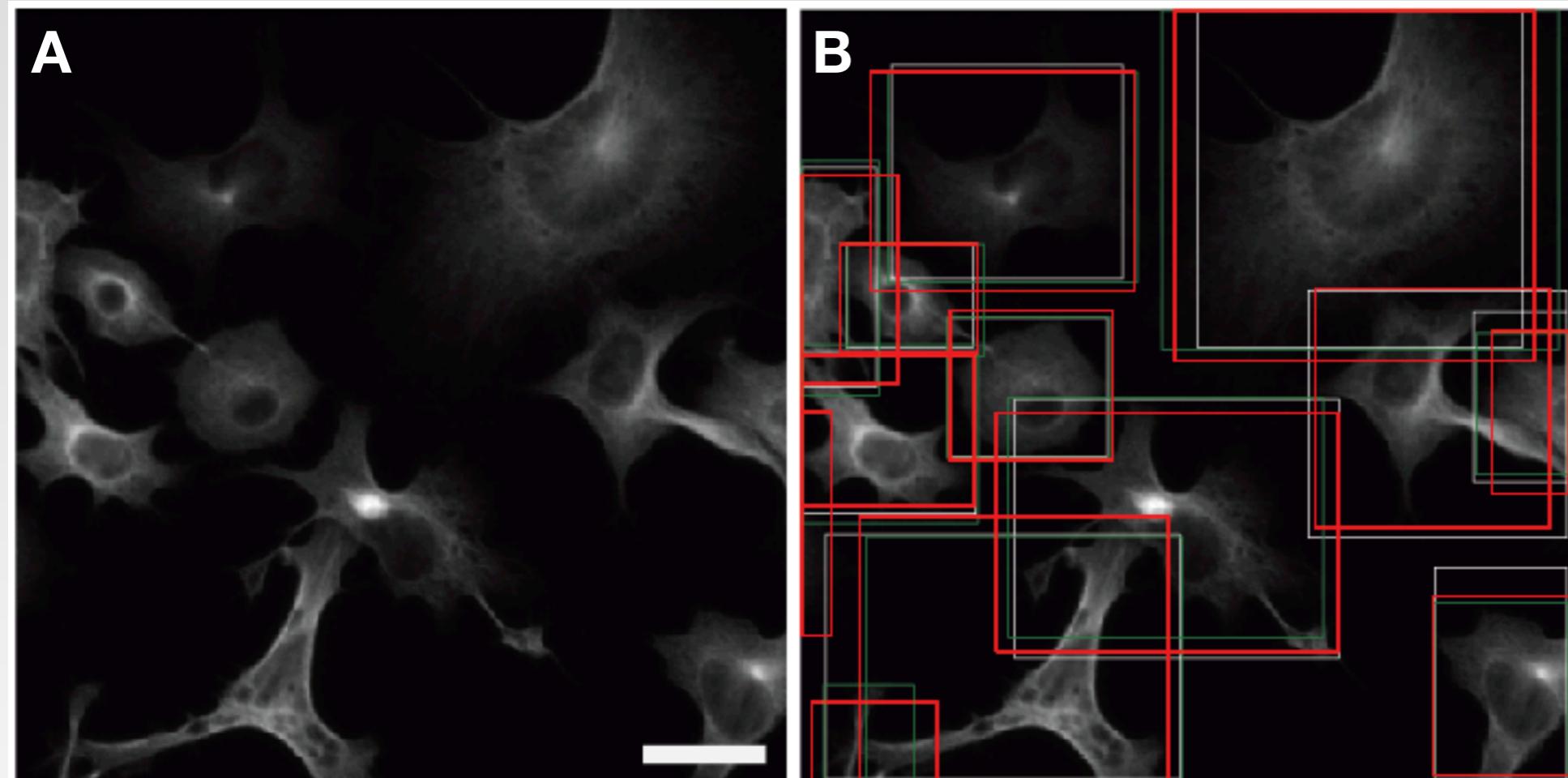
white boxes (ground-truth),

green boxes Faster-RCNN prediction

red boxes YOLO v2. prediction

# Additional Datasets: Neuroblastoma phalloidin

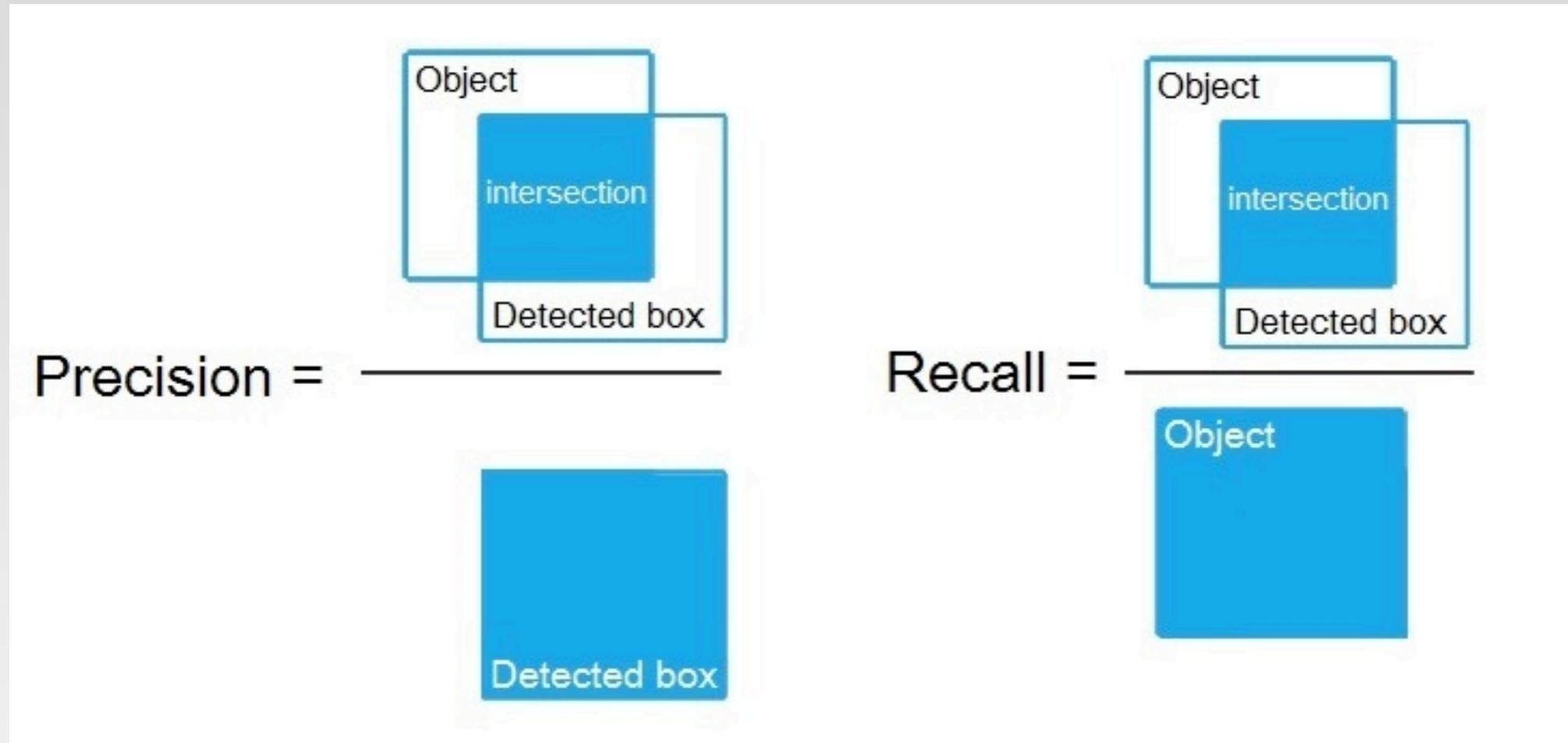
Neuroblastoma phalloidin stained. Challenging dataset.



**min = 1, max = 111.0, scale bar (25 µm)**  
**180 training images and 180 test images.**  
**white boxes (ground-truth),**  
**green boxes Faster-RCNN prediction**  
**red boxes YOLO v2. prediction**

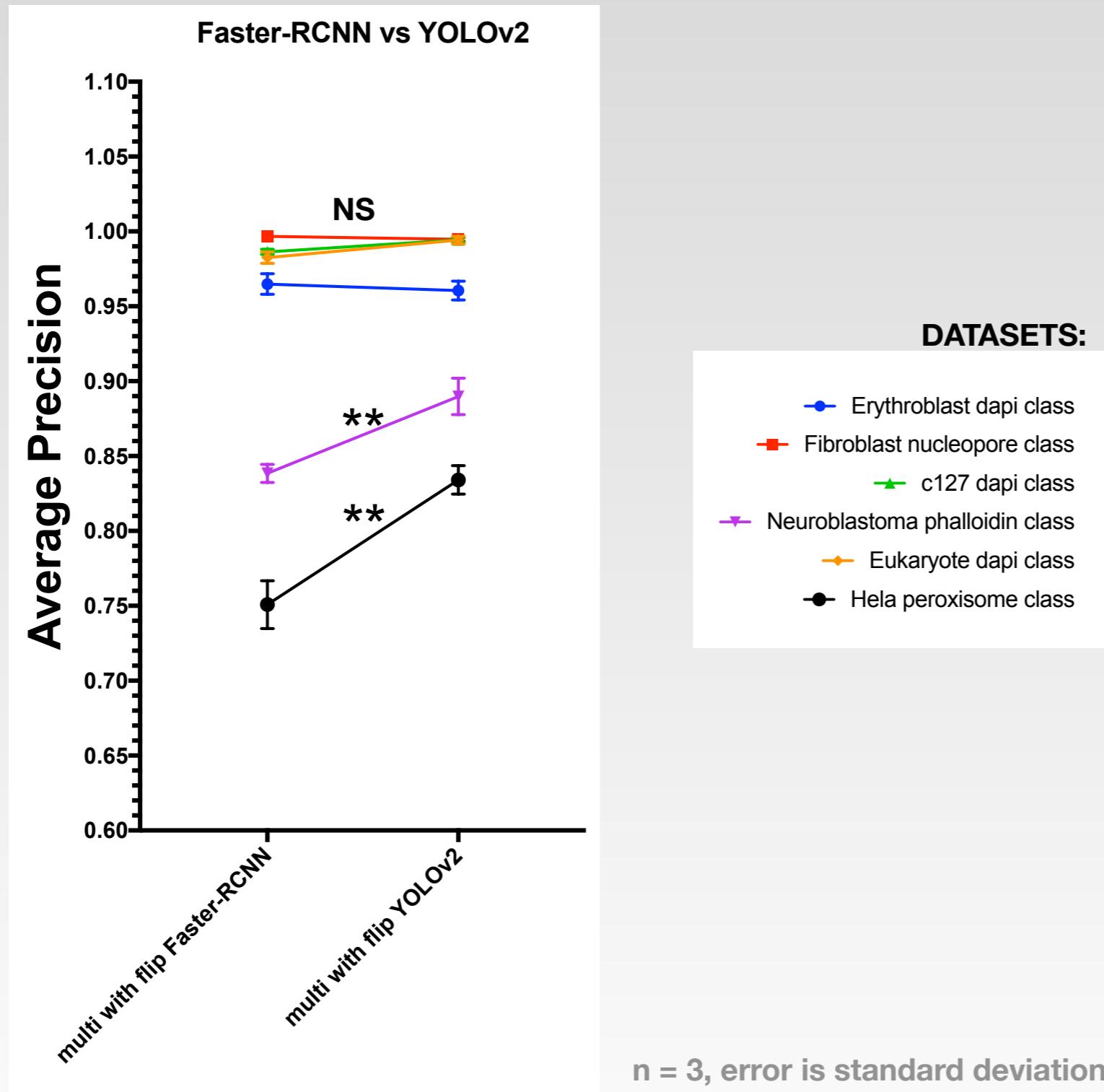
Source: Description: To develop automated algorithms for analysis of neurite outgrowth in high content screens of cultured neurons. This data set consists of wide field epifluorescent images of cultured neurons with both cytoplasmic (phalloidin) and nuclear stains (DAPI) and a set of manual segmentations of neuronal and nuclear boundaries that can be used as benchmarking data sets for the development of segmentation algorithms. 110098 110095

# I use the industry standard metric for localisation.



- Average precision is average value of 11 points on PR-curve for each possible threshold (each probability of detection) for the same class (Precision-Recall in terms of PascalVOC, where Precision=TP/(TP+FP) and Recall=TP/(TP+FN) ), page-11.

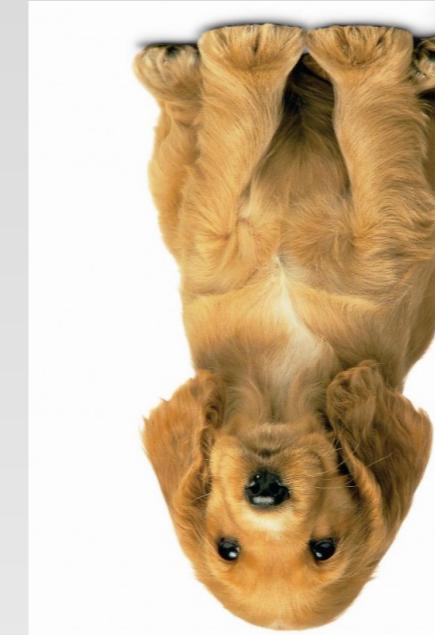
# YOLOv2 outperforms Faster-RCNN on challenging data.



- I tried both SSD network and YOLOv3, in my hands the underperformed compared these two methods

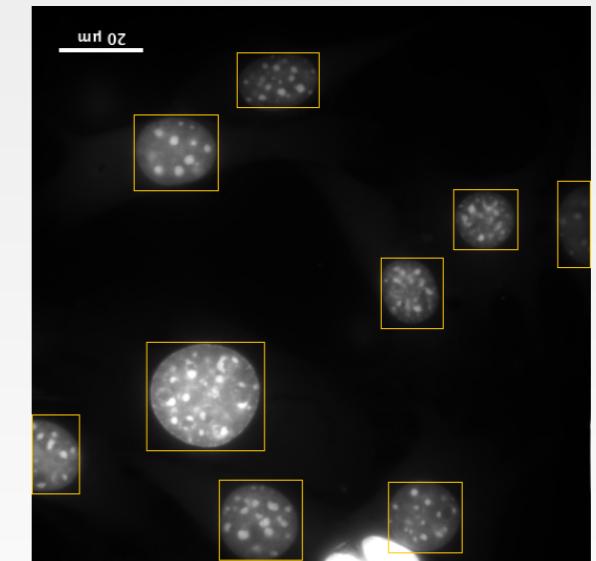
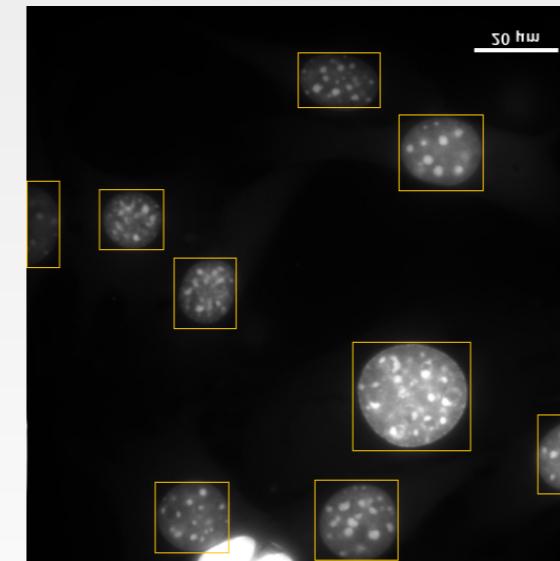
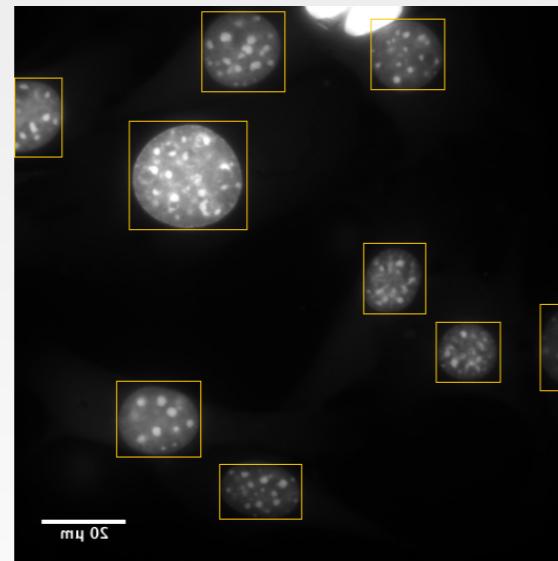
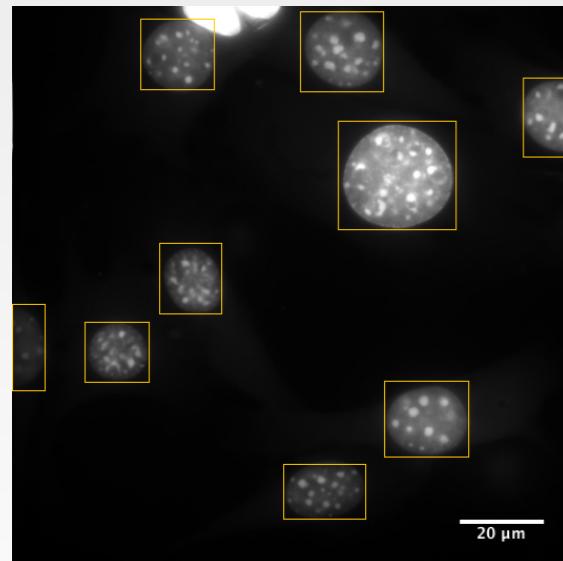
# Performance

- Training images are flipped horizontally to increase training data.



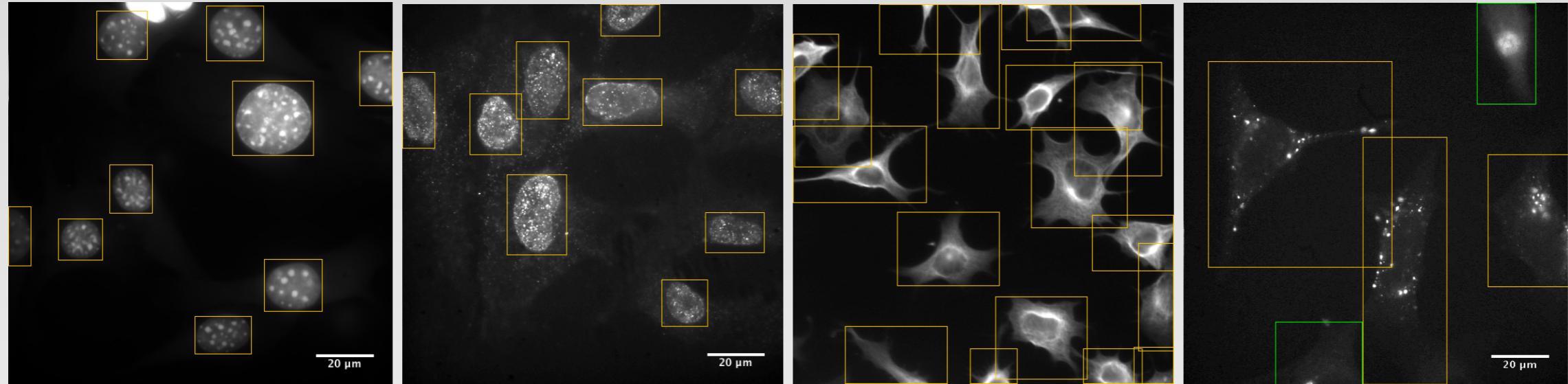
- With photography it doesn't make sense to flip vertically. Because you won't see in nature.

In fluorescence microscopy however we can flip both ways and so double our effective training data.

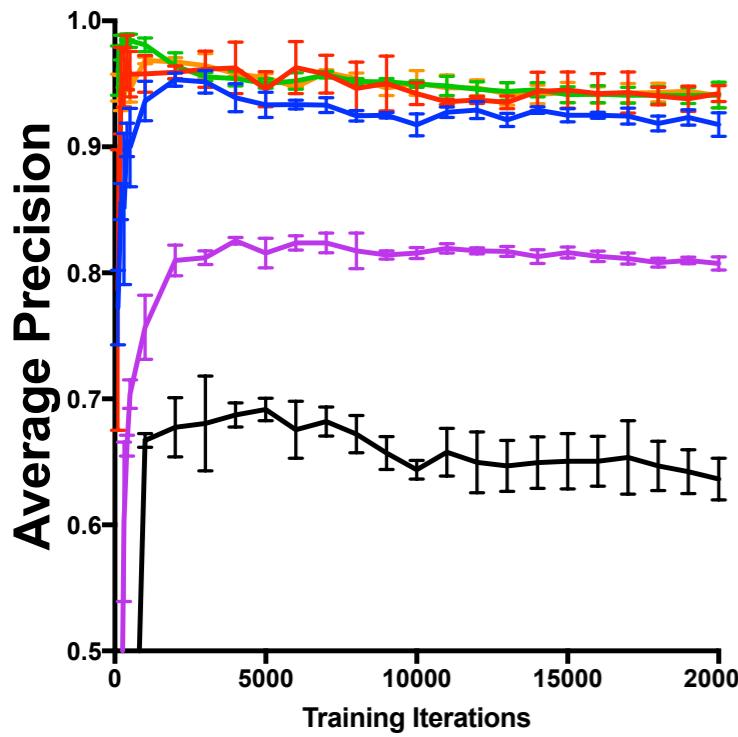


# Performance

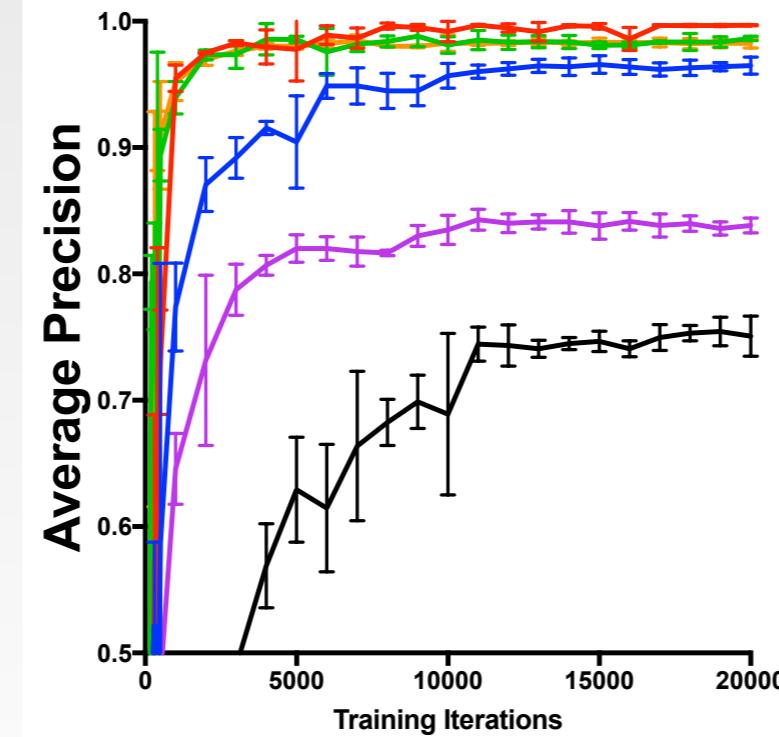
A dataset combining all the images.



Faster-RCNN trained on individual datasets with no additional data augmentation



Faster-RCNN trained on multi-class dataset with additional vertically flipped data augmentation

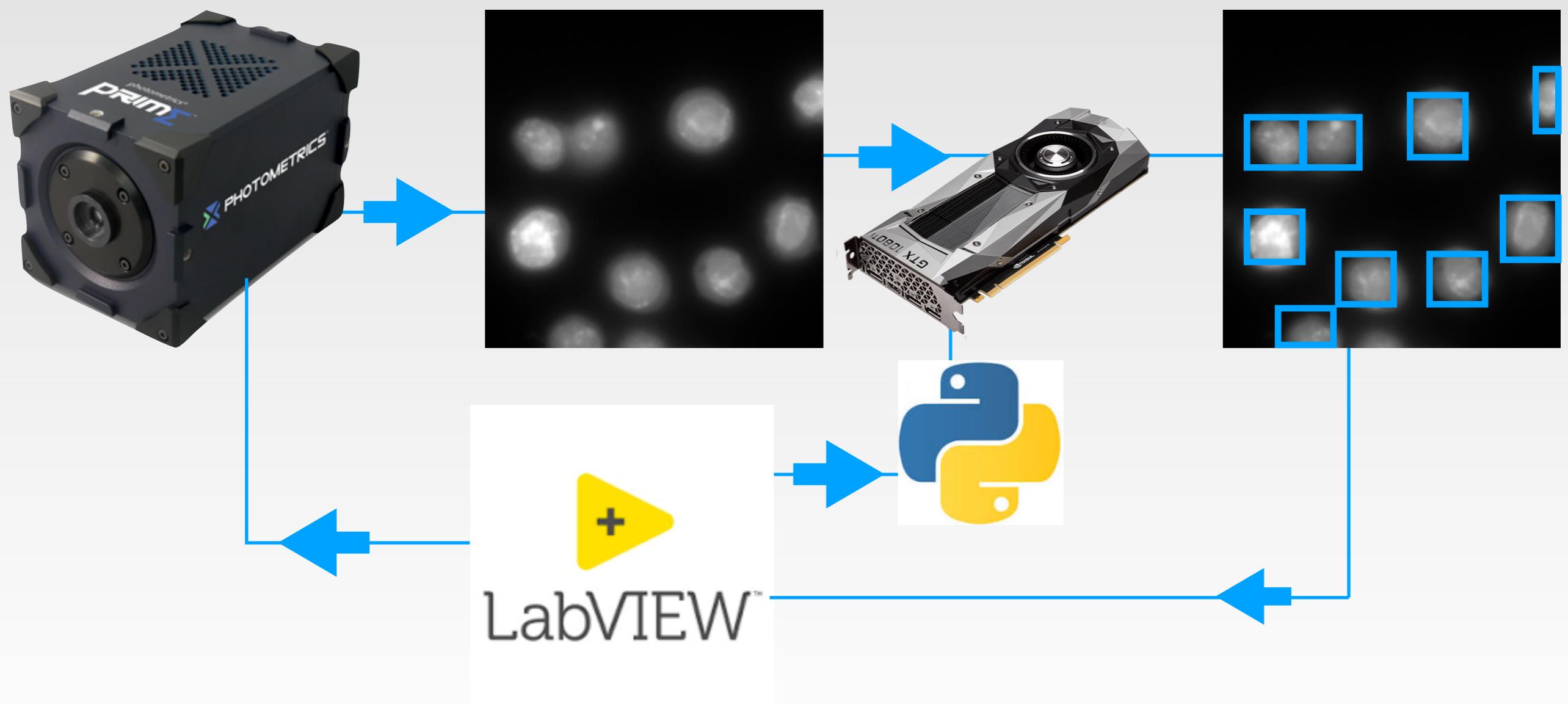


## DATASETS:

- Erythroblast dapi class
- Fibroblast nucleopore class
- ▲ c127 dapi class
- ▼ Neuroblastoma phalloidin class
- ◆ Eukaryote dapi class
- Hela peroxisome class

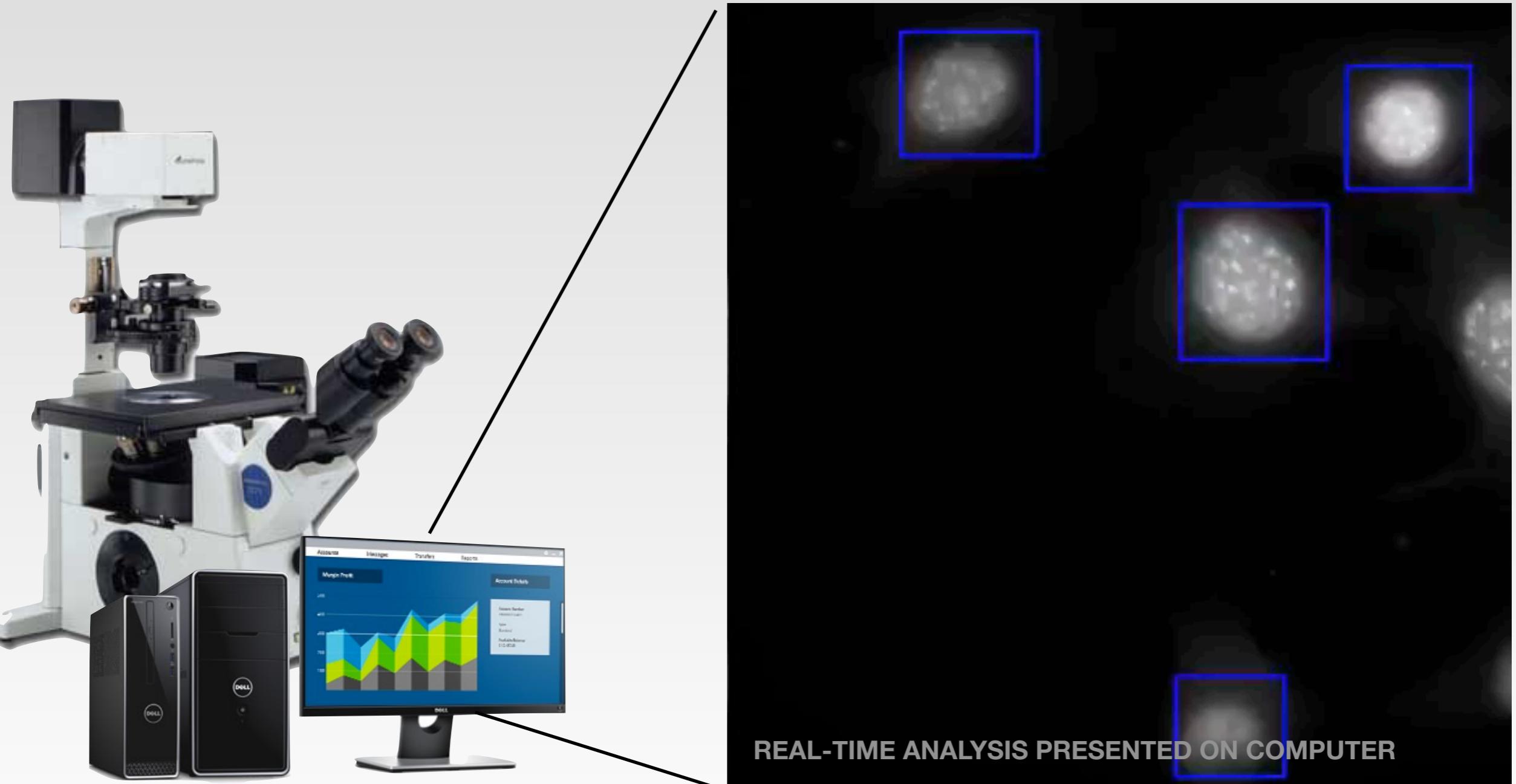
# Acquisition and detection

- Hook up image feed to cell localiser algorithm to find and record cells in real-time in 3-D.



# Trained object detection algorithm to recognise the cells.

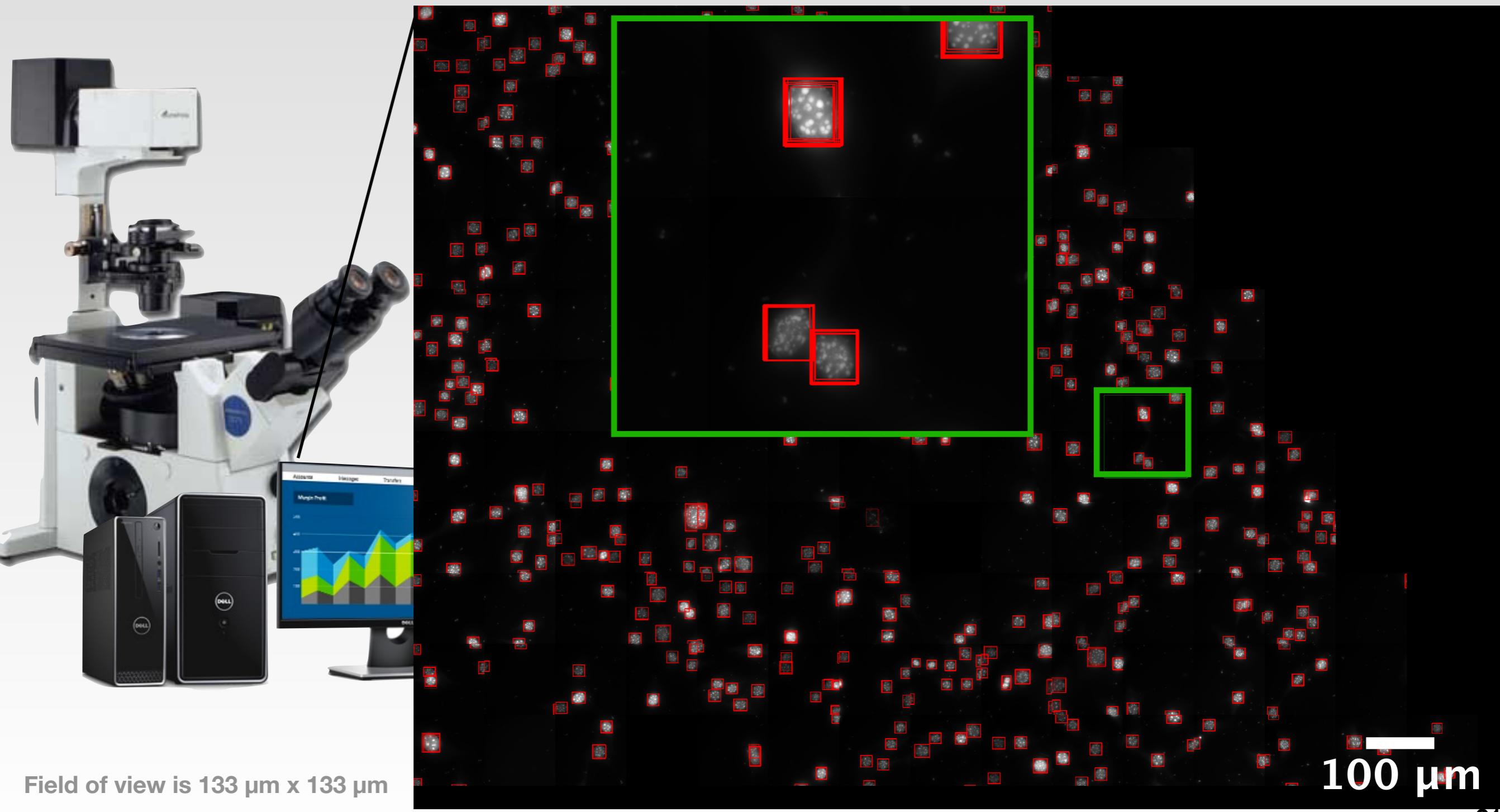
- Faster-RCNN output after training with 20 images.
- Applied to live-data, moving stage manually.



Field of view is 133 µm x 133 µm

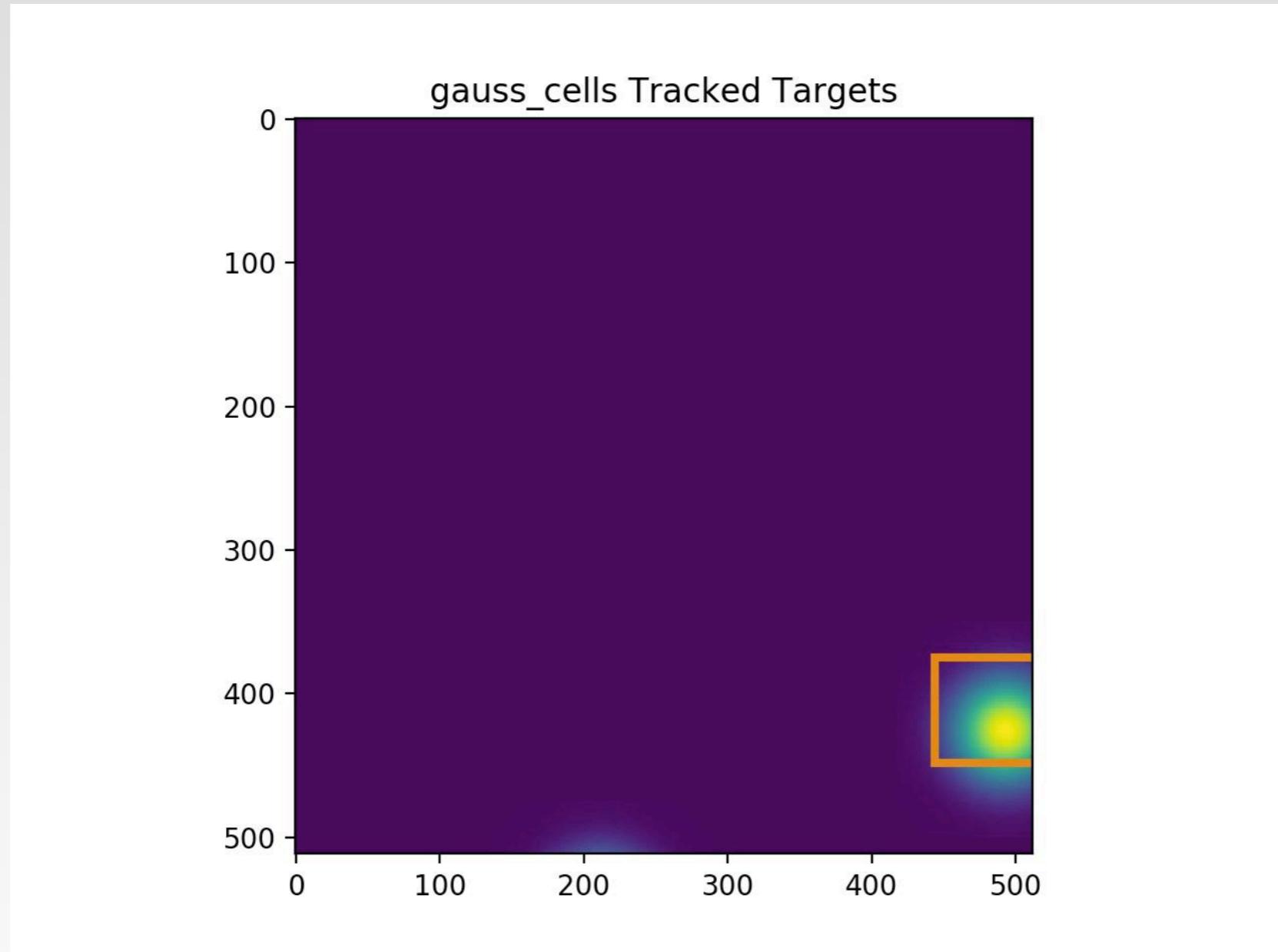
# scanning large regions of slide.

- System systematically scans slide but only acquires images where cells are present.
- Much faster than exhaustive sampling of same area.



# Linking detections between images

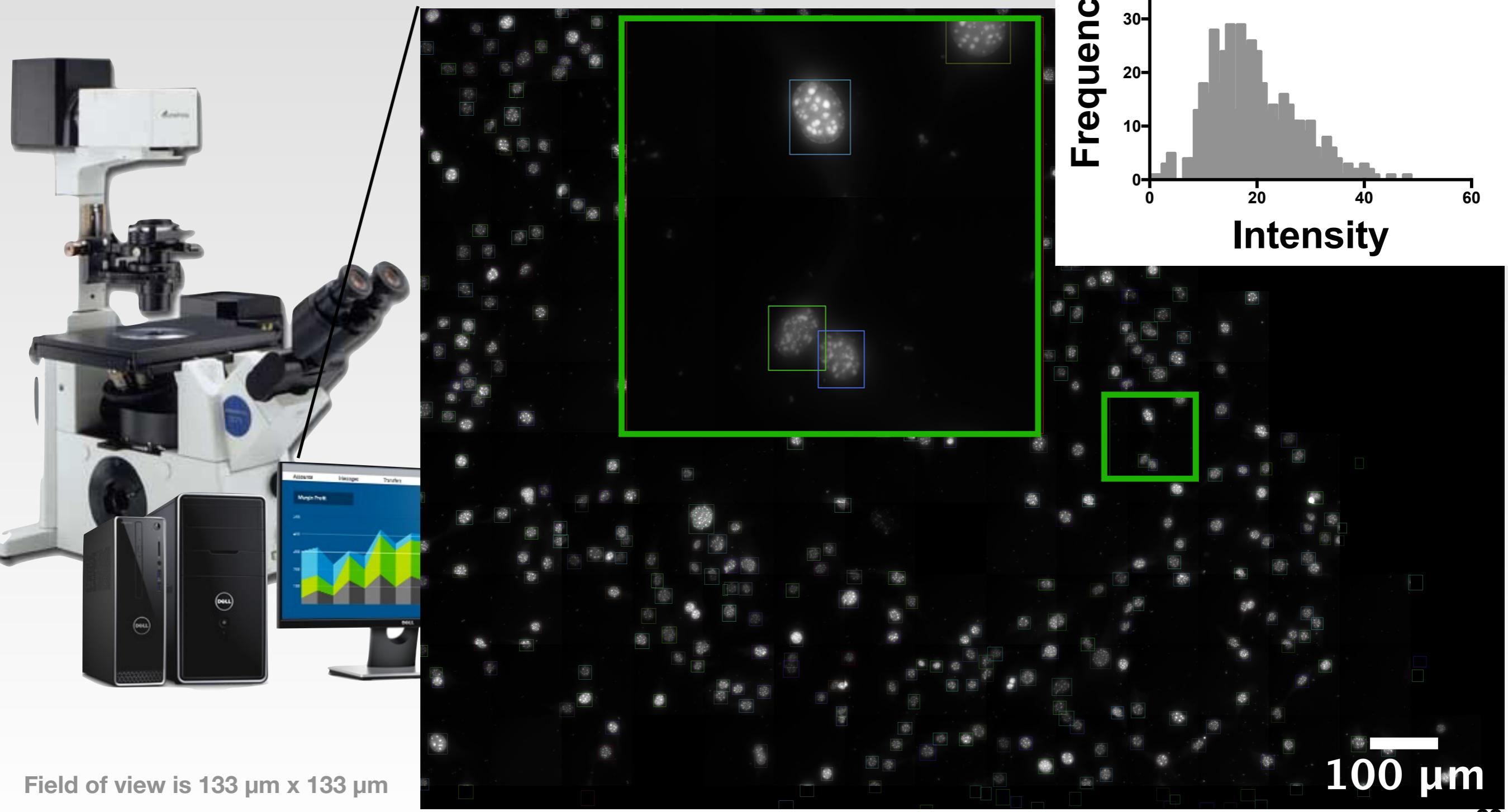
- Detecting cells is one thing, but you need to be able to link the observations between frames to ensure that they are identified correctly. Not a new problem.



- This is shown above by the detections being joined together with the same colour

# scanning large regions of slide.

- Merging of regions across z and measurement.
- Distributions are now possible.

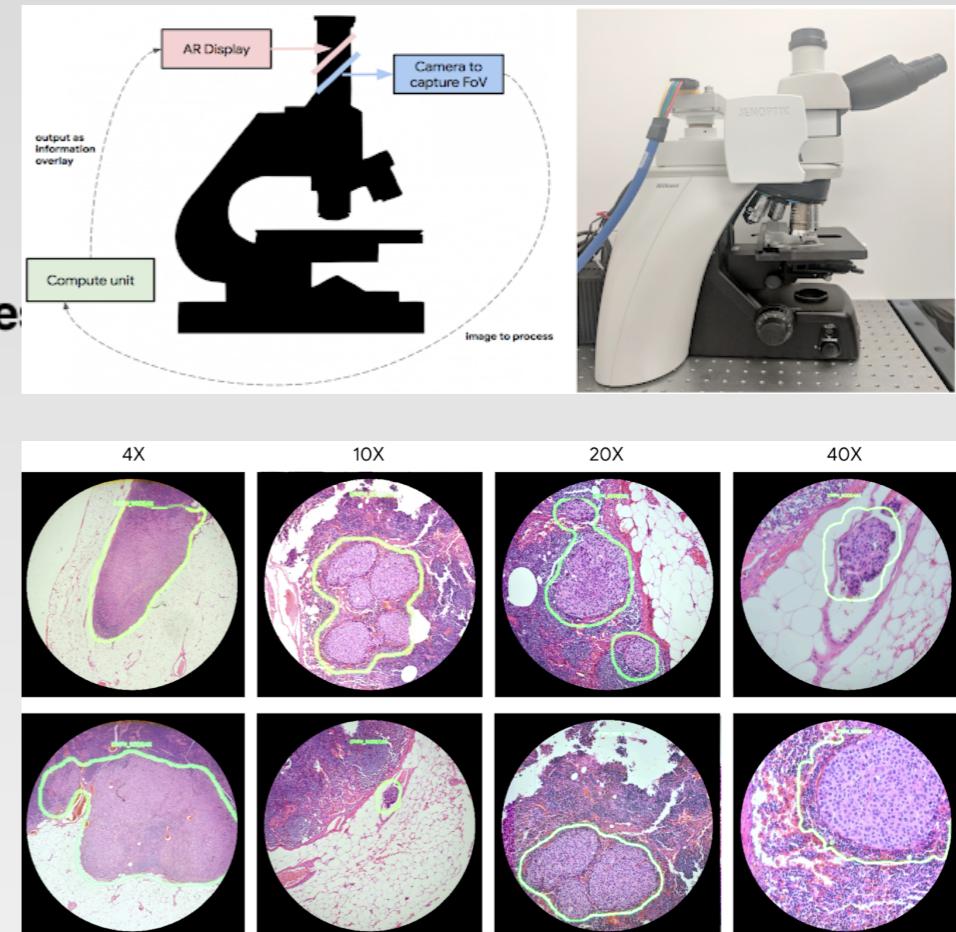
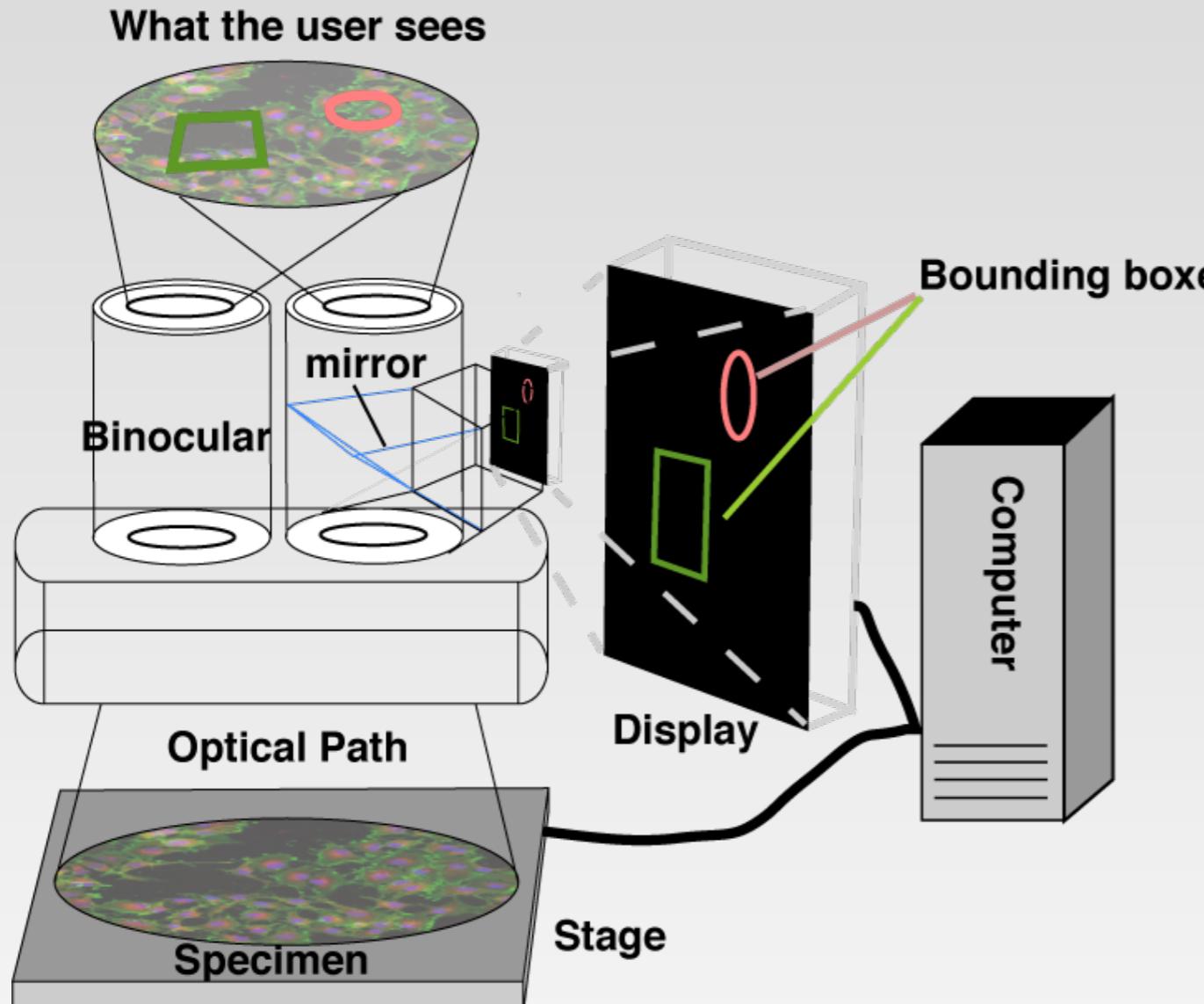


# But!!! It is important to keep the human the picture



Source: <https://www.avanade.com/en/blogs/avanade-insights/artificial-intelligence/how-humans-machines-will-work-together>

# Augmented reality display for fluorescence microscopy can help bridge the gap.



Google have developed their own augmented reality microscope.

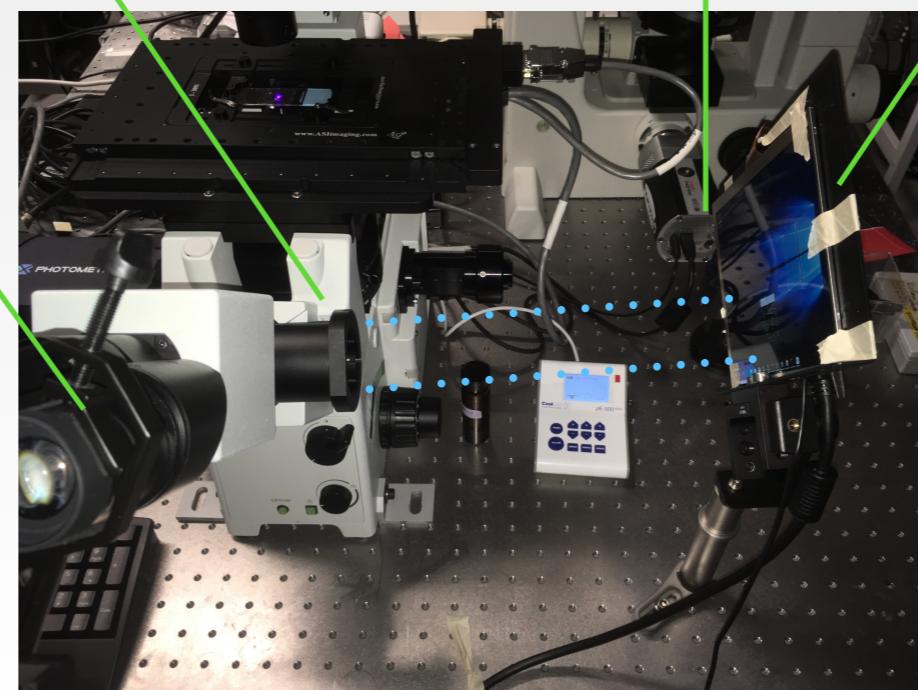
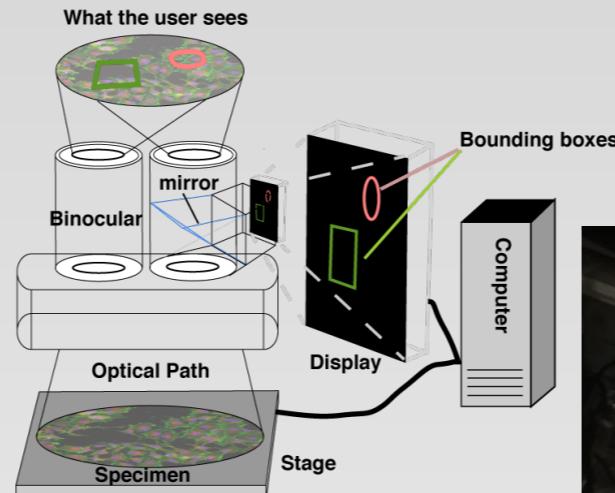
## AXRIV: Microscope 2.0: An Augmented Reality Microscope with Real-time Artificial Intelligence Integration

[Po-Hsuan Cameron Chen](#), [Krishna Gadepalli](#), [Robert MacDonald](#), [Yun Liu](#), [Kunal Nagpal](#), [Timo Kohlberger](#), [Jeffrey Dean](#), [Greg S. Corrado](#), [Jason D. Hipp](#), [Martin C. Stumpe](#)

# Augmented reality hardware.

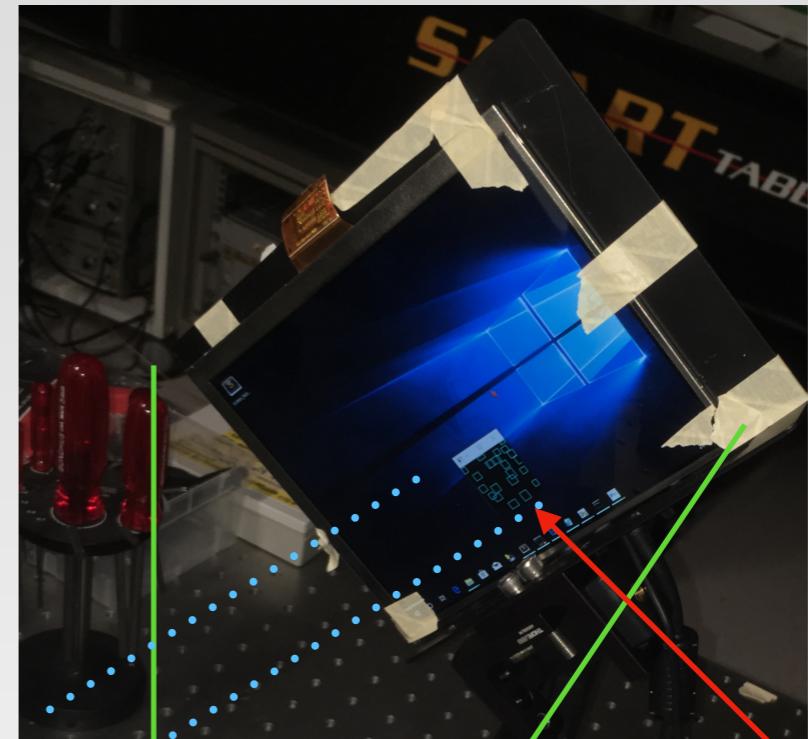


Mightex Dichroic/filter cube for coupling Polygon DSI into an Olympus BX upright.  
 + We engineering coupling.  
 + We insert 50:50 beamsplitter



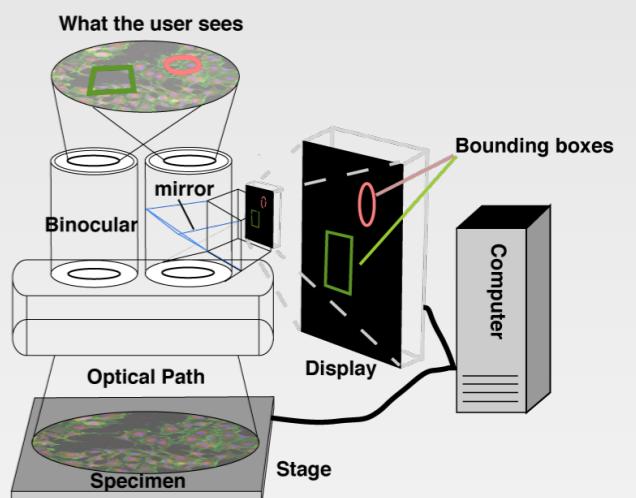
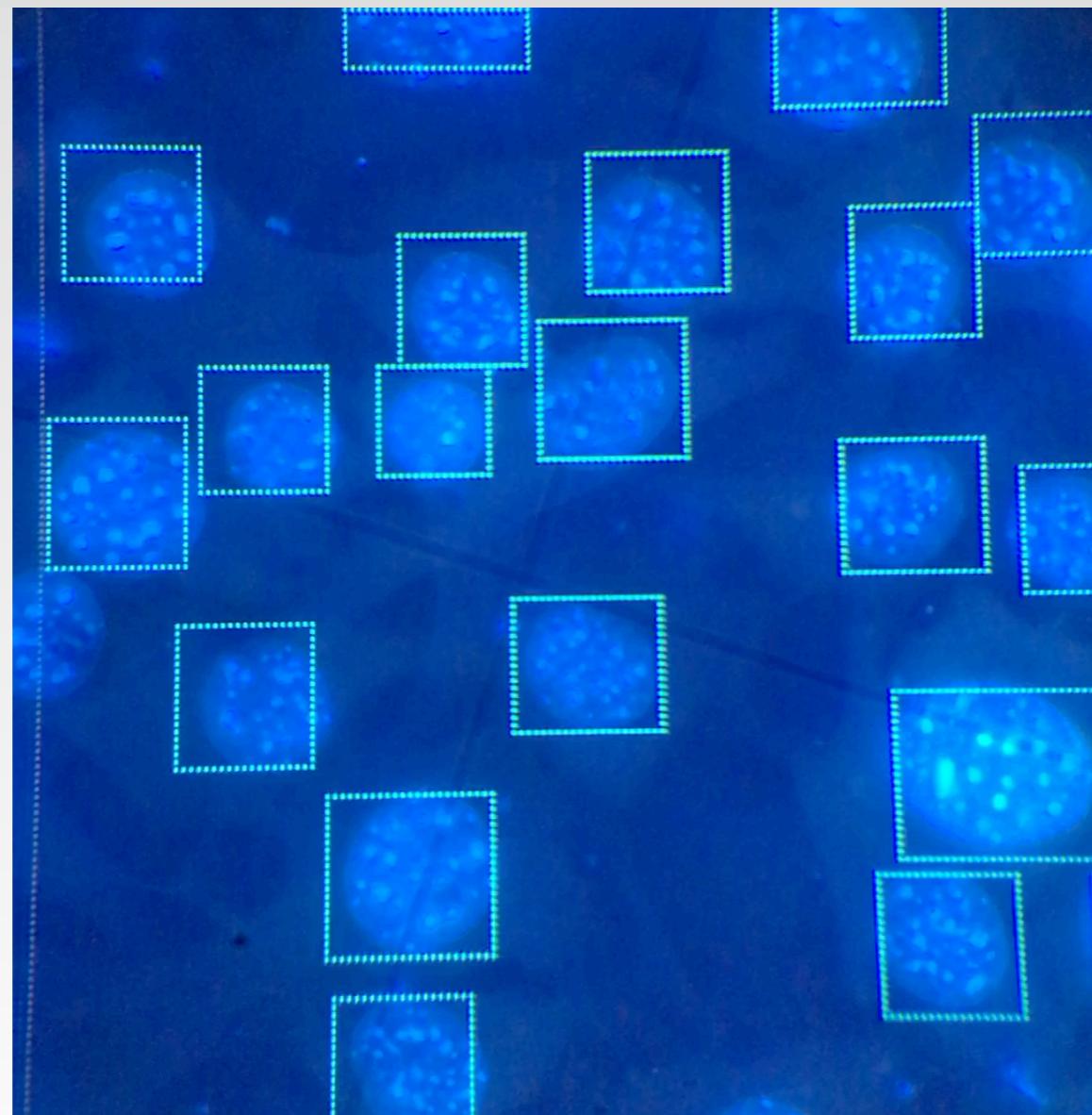
This is a Pimoroni HDMI 8" IPS LCD Screen Kit

regions



# Augmented reality display.

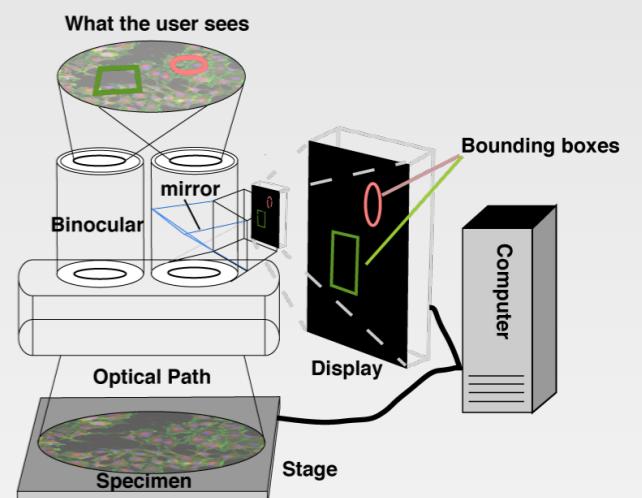
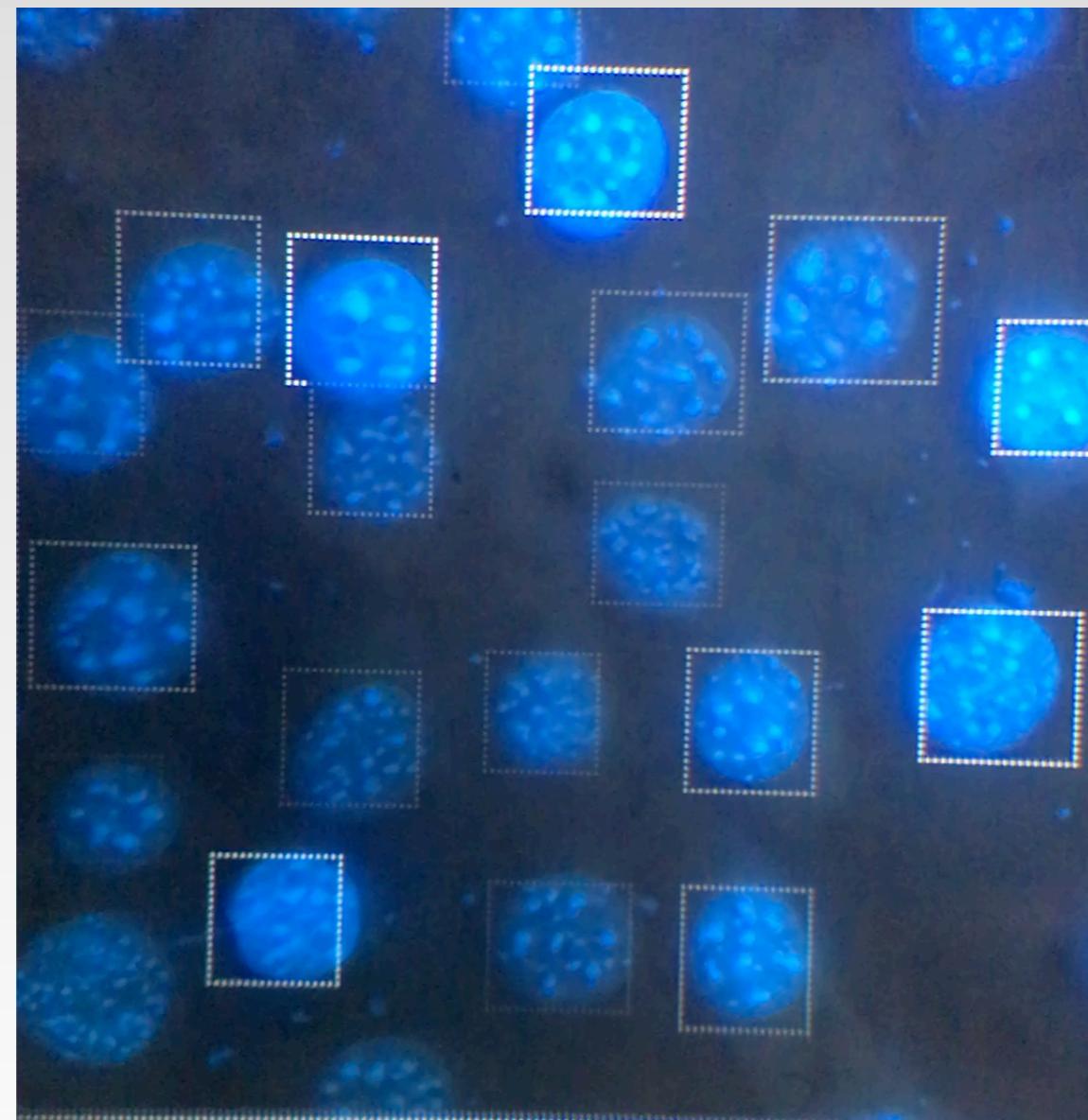
- Real-time (sub-second update).



Source: Looking down the binocular.

# Instantaneous measurements

- Visualising the intensity of the cells in real-time. (brighter square, brighter cell)



Source: Looking down the binocular.

# Summary

- Have found two excellent object detection algorithm for cellular identification.
- Have benchmarked it and other algorithms on actual data.
- Surprise is that it works well without too much data.
- Have enhanced the training and justified why we need a corpus of data.
- Can distinguish cells in images and can be used to acquire 3-D volumes of cells
- Have developed a real-time acquisition system for previewing and screening.

## Next steps

- Want to start screens with biological researchers.
- Expand modalities of fine-grain analysis.

# Thanks for your attention

Thanks to Katharina Reglinski for Peroxisome samples.

Thanks to Isabel Diez-Sevilla for glycophorinA samples.

Thanks to Jill Brown for erythroid examples

Thanks to Christian Eggeling

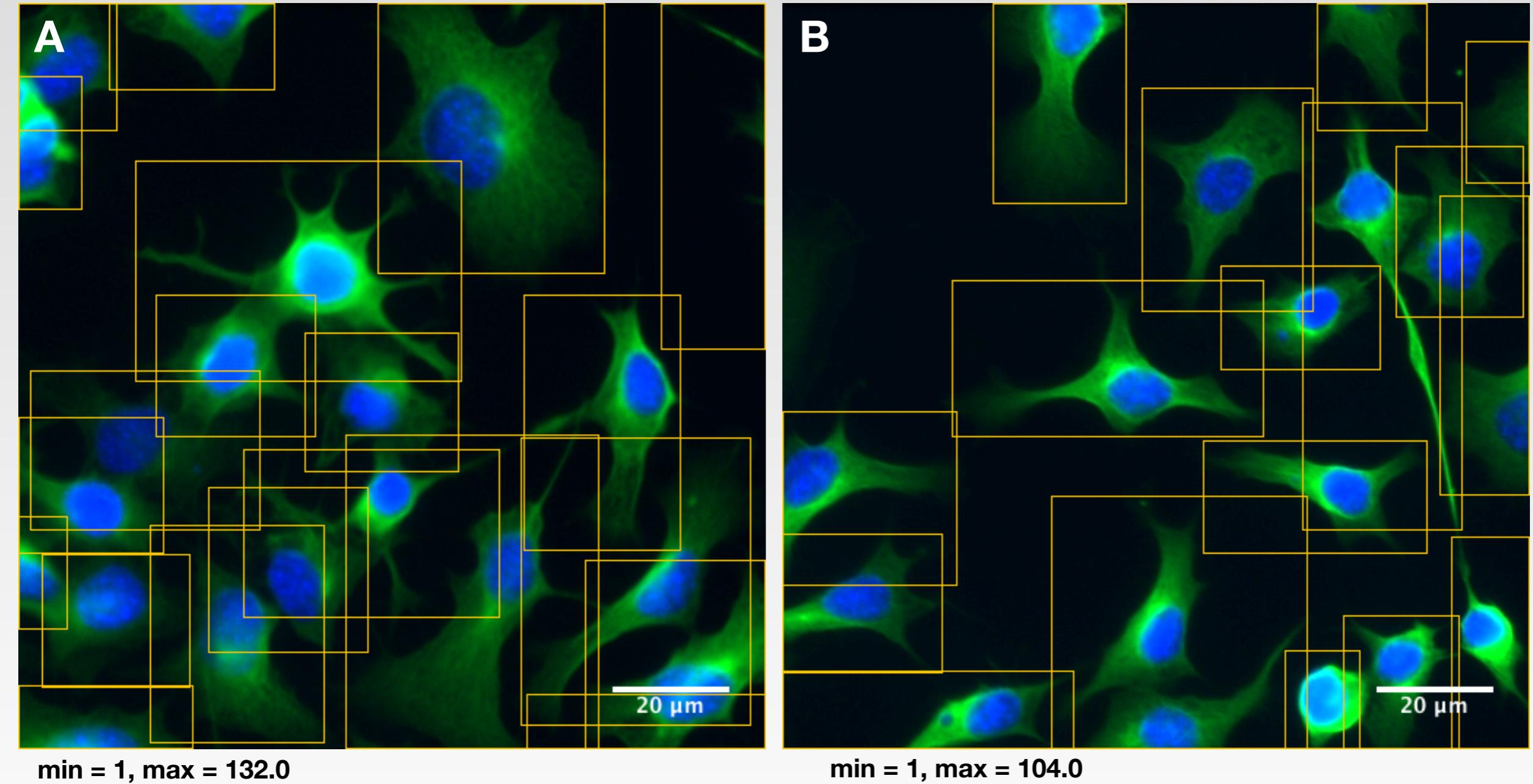
Thanks to Eggeling Group and Wolfson Imaging Centre

Please follow me on Twitter.

[dwaithe@twitter.com](mailto:dwaithe@twitter.com).

# Additional Datasets: Neuroblastoma + dapi

- Neuroblastoma (phalloidin and actin). Challenging dataset (less than -dapi).



Source: Description: To develop automated algorithms for analysis of neurite outgrowth in high content screens of cultured neurons. This data set consists of wide field epifluorescent images of cultured neurons with both cytoplasmic (phalloidin) and nuclear stains (DAPI) and a set of manual segmentations of neuronal and nuclear boundaries that can be used as benchmarking data sets for the development of segmentation algorithms. 110098 110095

