

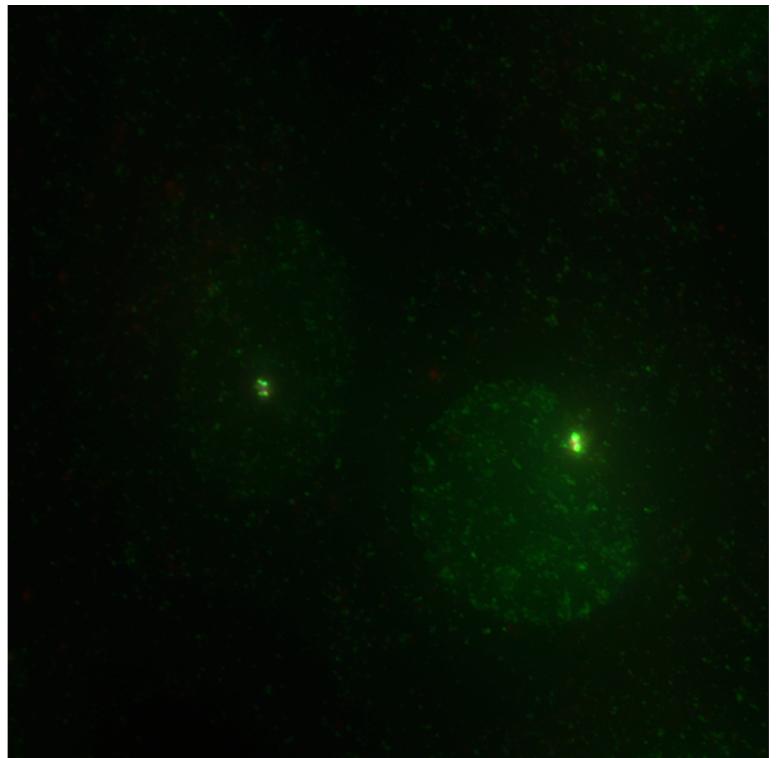
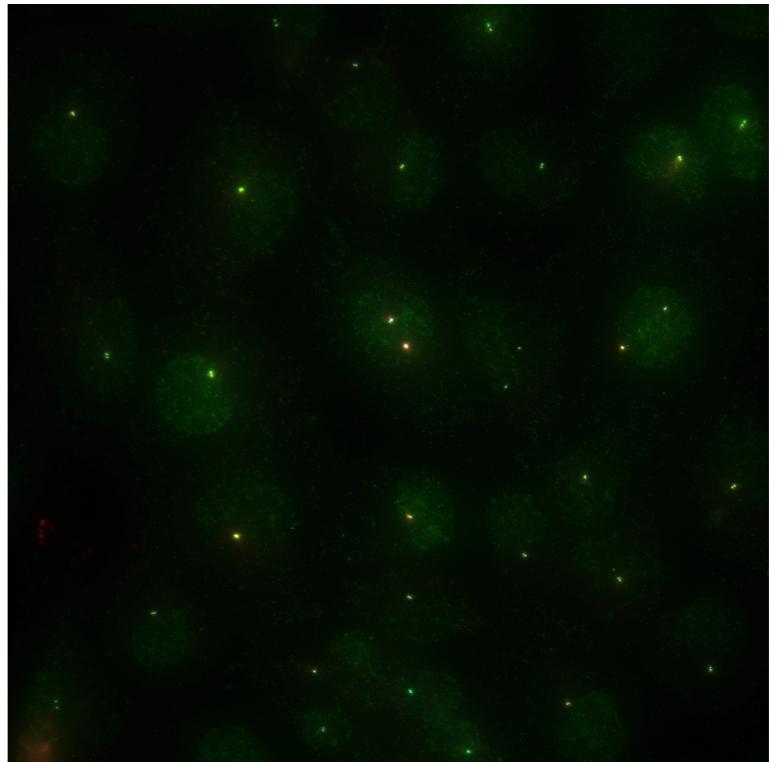
# Developing a ML pipeline to detect centrioles in human cells

Leo Burgy  
2021-11-08

# Aim

1. Detect centrioles
2. Assign centrioles to cell

=> Determine  $N_c$



# Rationale

- Cells have to **maintain** the centriole number ( $N_c$ ) during cell division.
- Loss of the control leads to **division problems**.
- Besides, centrioles, aka basal bodies, participate in the **formation of cilia**.
- The components of the centrioles have been **studied** using a **combination** of approaches such as proteomics, genetics, fluorescence microscopy and electron microscopy.
- The absence of a protein of interest may **modify**  $N_c$ .
- Yet, not all proteins involved have been characterised.
- Thus knowing the variation of  $N_c$  is a key step when analysing a protein of interest.

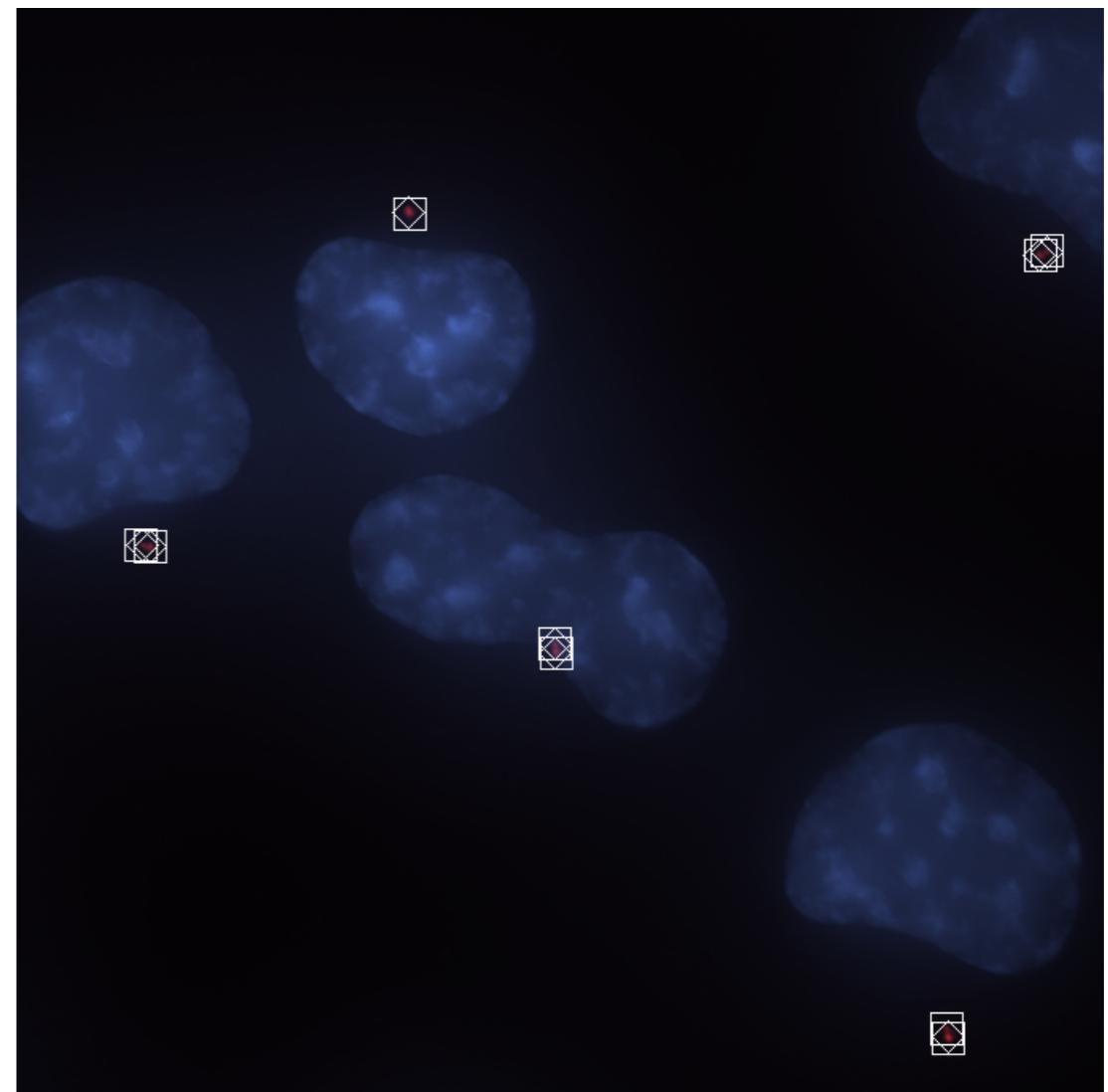
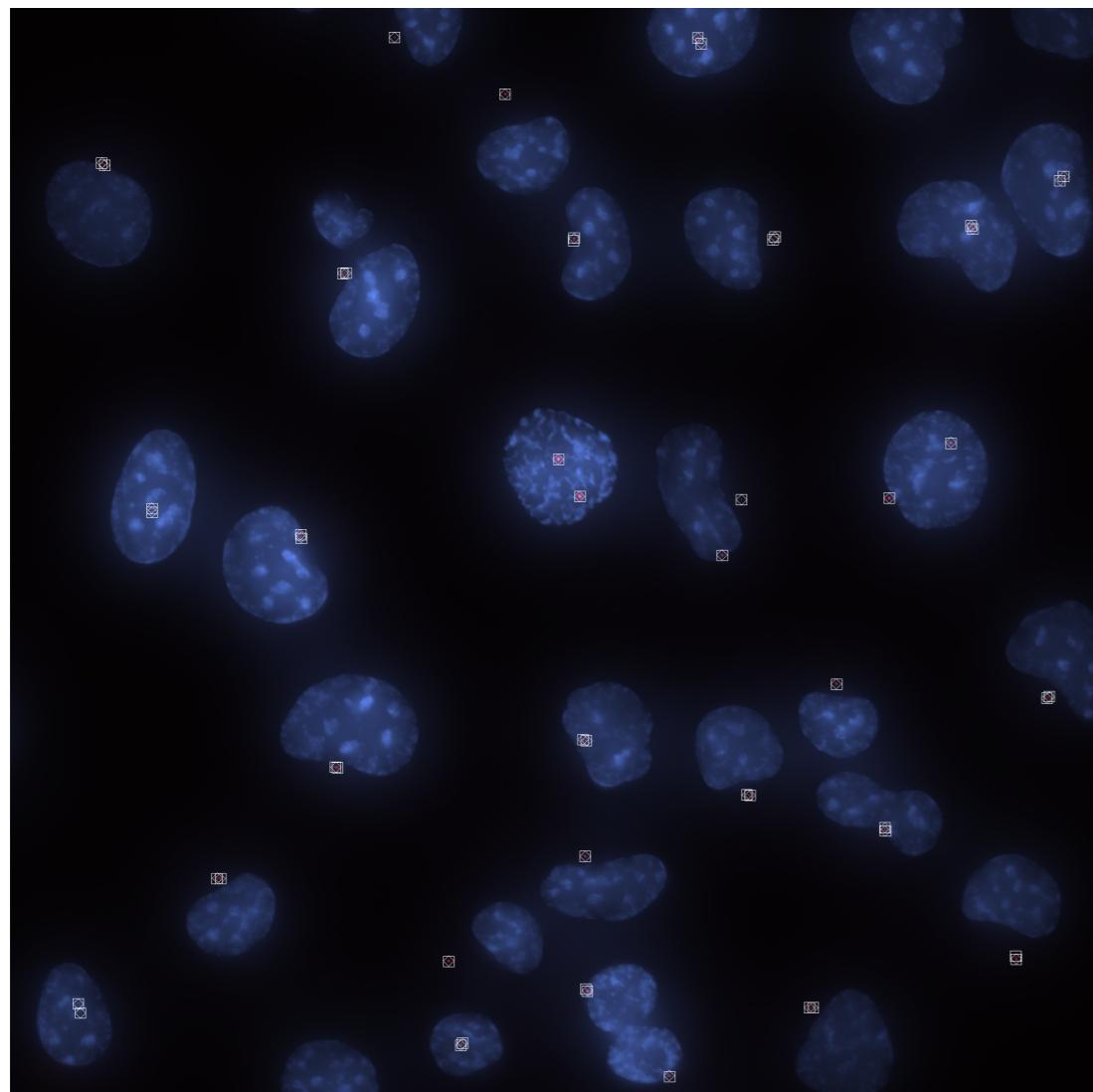
# Manual solution

- Score manually at the microscope
  - not reproducible
  - risk of scoring the same
  - Not easy to discern (too dim signal for the eye)
  - slow / low throughput

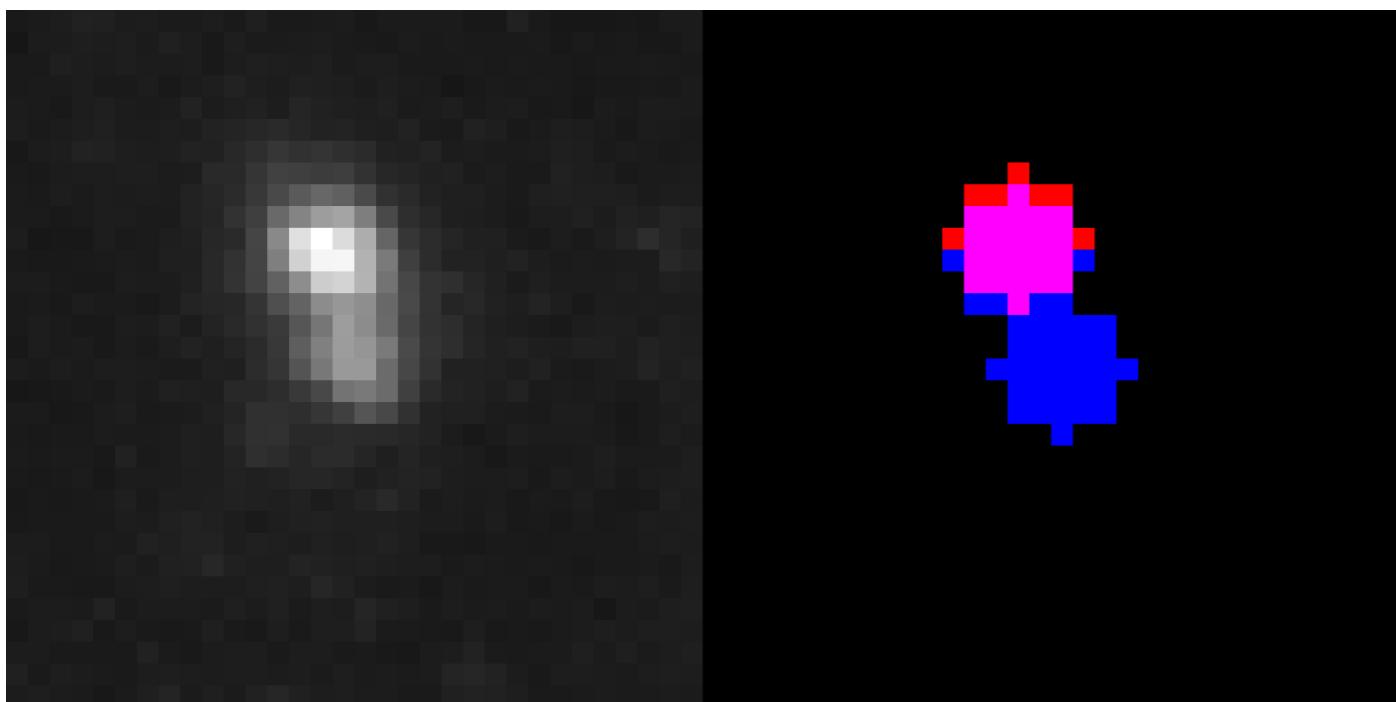
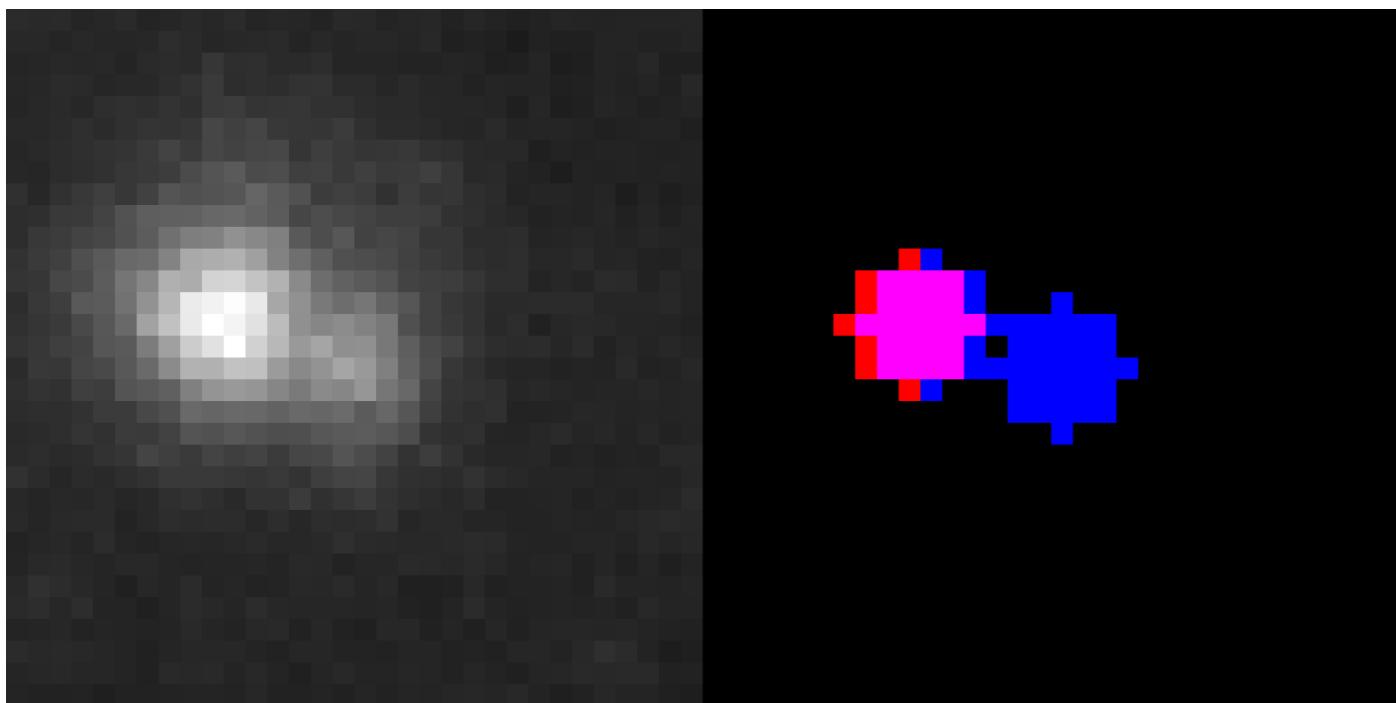
# Automating detection

- Telling apart the fore- and background: threshold
- Determining the scale-space representation: blurring
- Removing the spurious blobs: max-filter for extracting centrosomes
- Picking the foci: local peak
- Relaxing the coordinate space: up-sampling
- Determining the minimal inter-peak distance

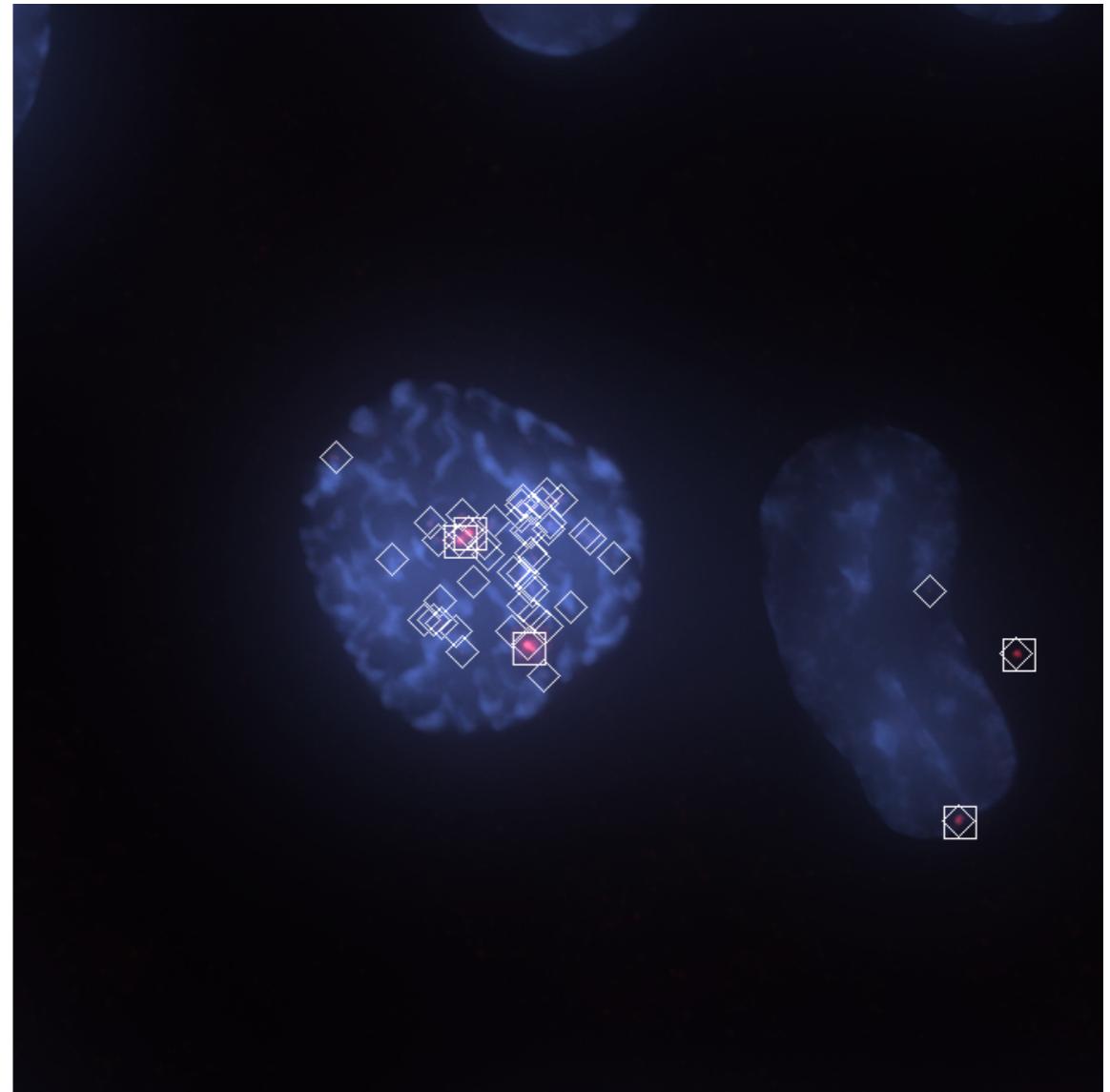
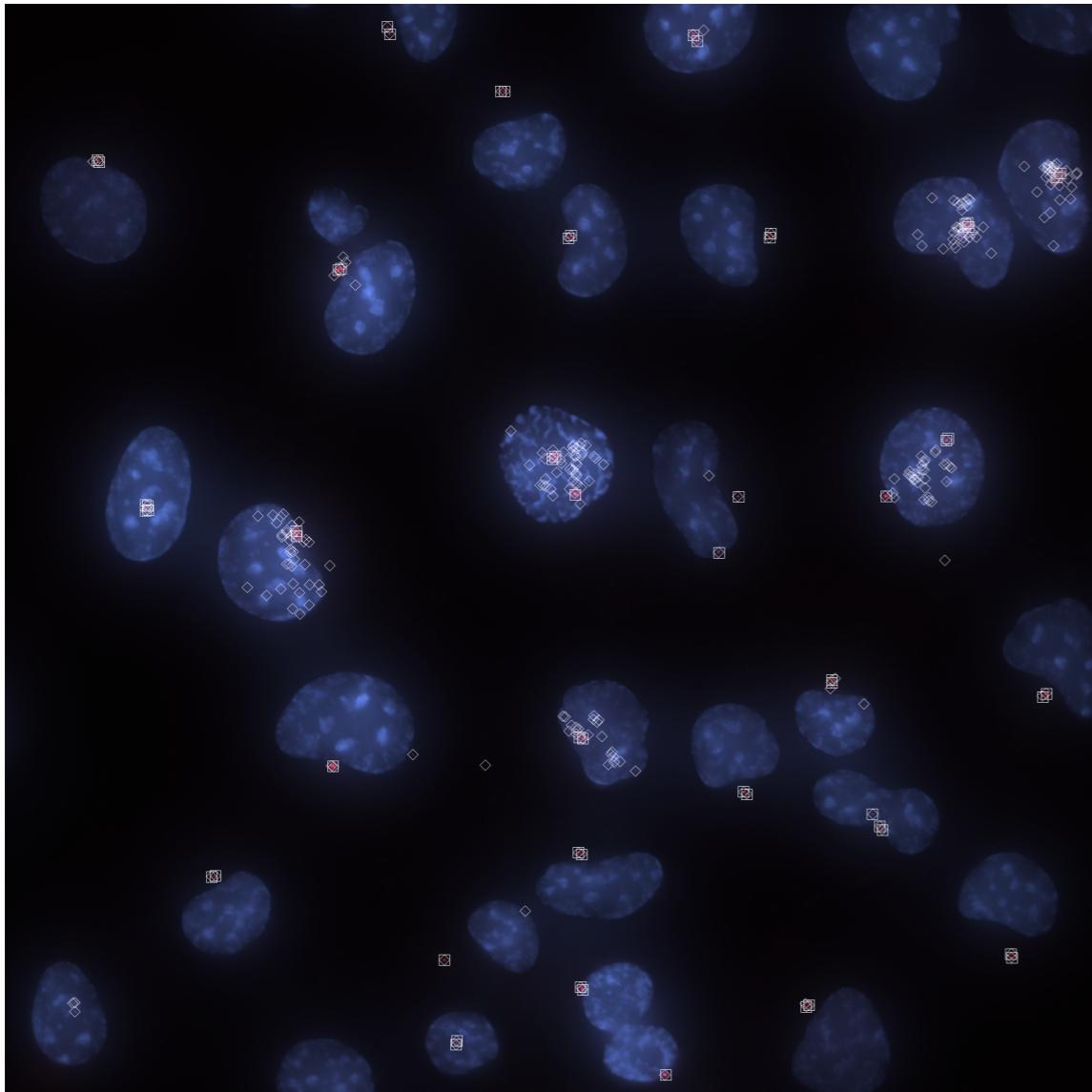
# Annotation (CEP63)



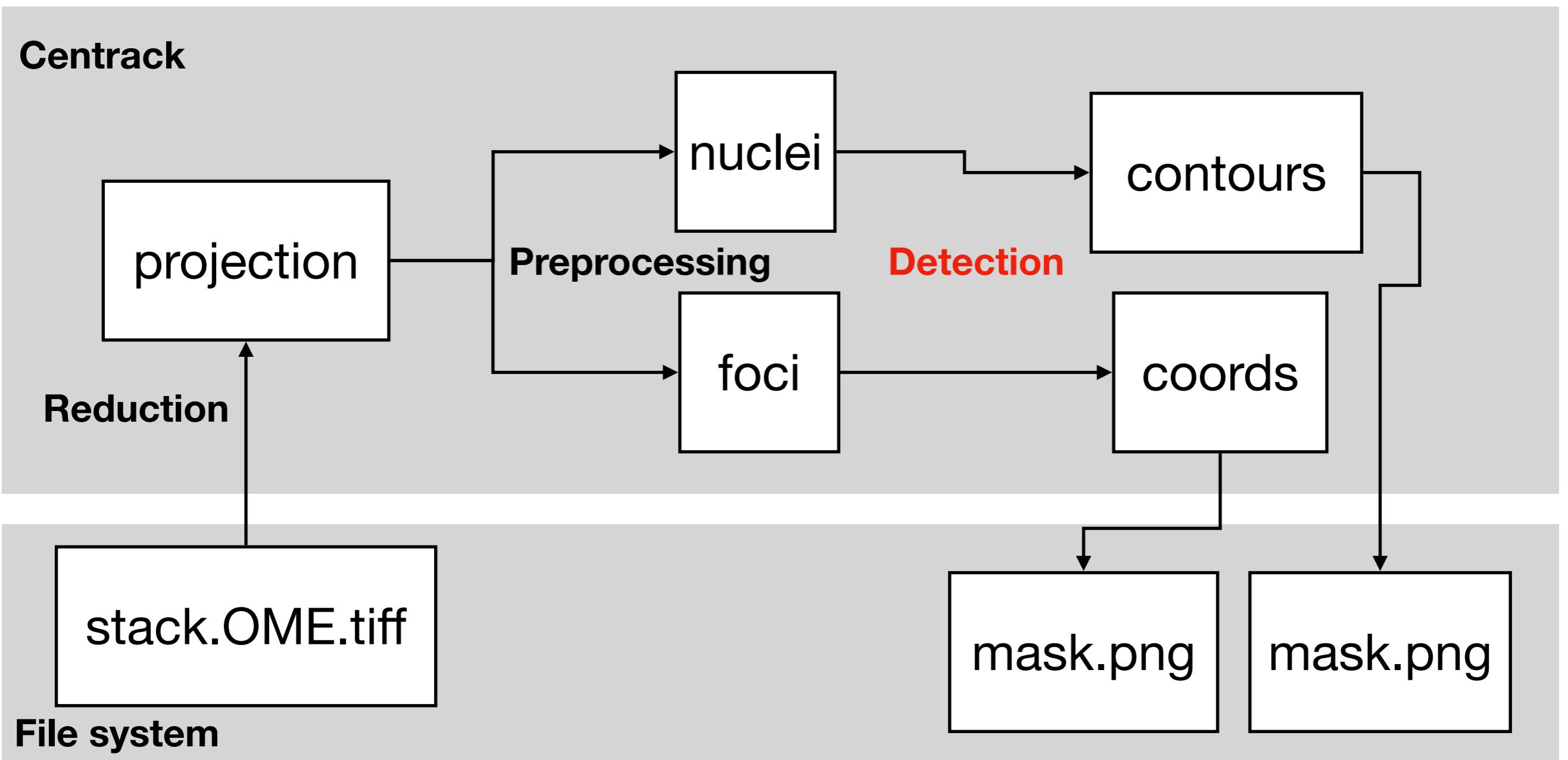
# False negative 0:1



# Annotation (CETN2)



# Pipeline structure



# The dataset

- Images of human cells, in which centrioles are labeled by tagging centrin (component present in both the centriole and the procentriole) with fluorescence; also contains additional centriole markers.
- Collection of positions of centrioles present in one image (by channel, max-z-projection)
- DAPI channel (nucleus marker).

# What you will be doing

In this project, you will

- learn to use the concepts we have seen in the lectures and practiced in the labs on a real-world dataset, start to finish.
- explore your dataset and your features,
- process and engineer your dataset and extract more meaningful information,
- implement and use machine learning methods on real data,
- analyse your model and **Object detection? Segmentation? Regression?**
- generate predictions and **Use one or more channels?**
- report your findings. **How to score a prediction?**

# Resources

- <https://ai.stackexchange.com/questions/31710/which-neural-network-architecture-to-use-to-detect-very-close-and-very-small-blob>
- <https://www.dropbox.com/sh/na34gq3ikfumgn0/AAAAaQiFkHEpsCxM8N5dBqnXca?dl=0>
- [labelbox.com](https://labelbox.com)
- More than one group? job splitting? in parallel?