Object counting with Deep Learning

Luca Clissa
INFN Bologna
clissa@bo.infn.it

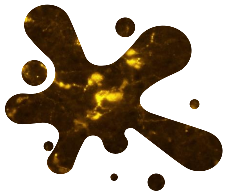
About me

Luca
Clissa

PhD in Data Science & Computation

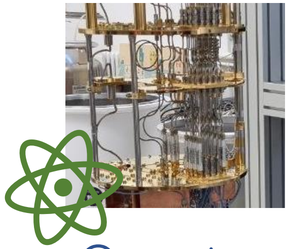
Collaborazioni

Aree di ricerca



Computer Vision
[1, 2, 3]

Text Processing
[3, 4, 5, 6]



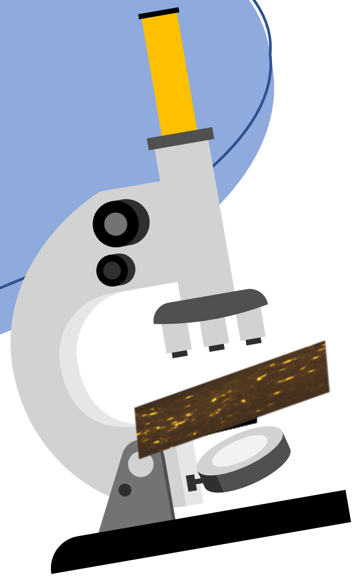
Quantum Computing
[7, 8, 9]

Time Series, Survival Analysis

Outline

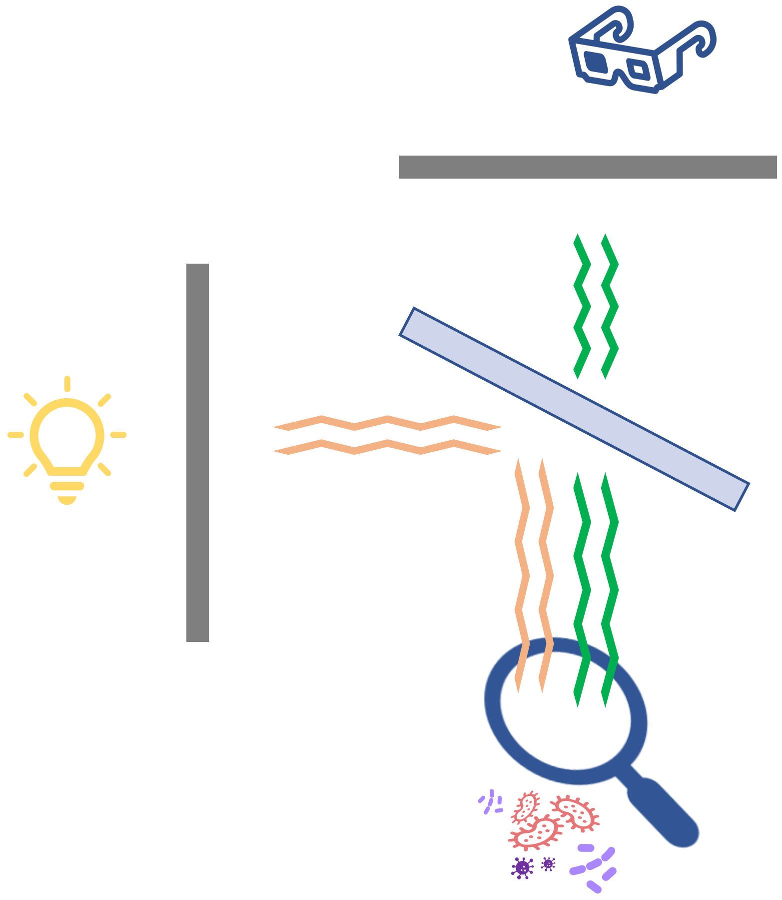
- Microscopia in fluorescenza
- Fluorescent Neuronal Cells dataset
- Cell ResUnet (c-ResUnet)
- Ispezione risultati
- Possibili sviluppi

Microscopia in Fluorescenza



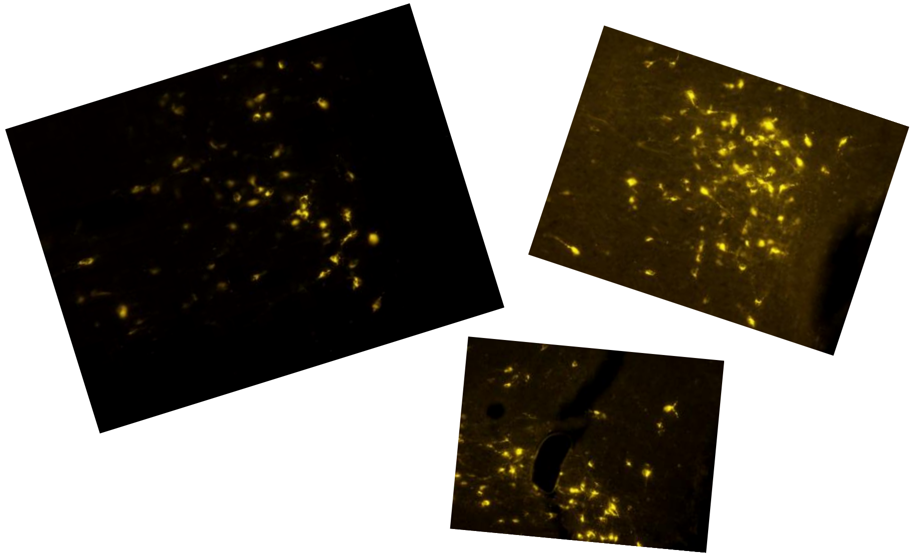
Cos'è?

- Tecnica di imaging
- Assorbimento/emissione luce
- Spesso utilizzata in life science



Applicazioni

- Studio meccanismi responsabili del torpore nei roditori [10]
  - Importanti ricadute umane
- Neuroni marcati con fluoroforo giallo
- Varie tonalità, forme e grandezze
- Obiettivo: conteggio neuroni marcati



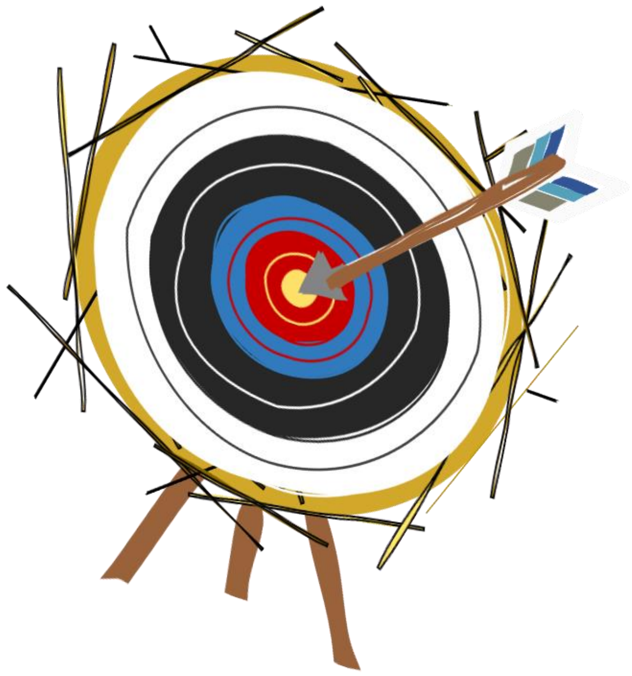
Problema

- Manuale
  - Time-consuming
  - Error-prone
- Interpretazione soggettiva casi limite

Obiettivo

Automazione

Interpretabilità



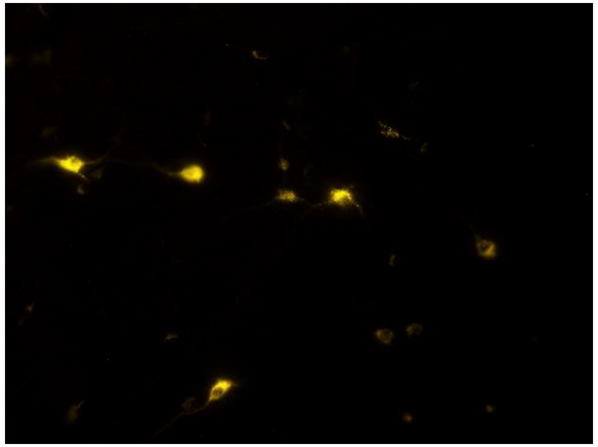
Contributo

- Approccio Deep Learning per riconoscimento e conteggio
  - cell-ResUnet (c-ResUnet)
  - test against SOTA
- Ablation studies
  - Weight Map (WM)
  - Artifacts Oversampling (AO)
- Fluorescent Neuronal Cells dataset

[1] Morelli, R. et al. 2021
[2] Clissa, L. et al. 2021
[3] Clissa, L. et al. 2022

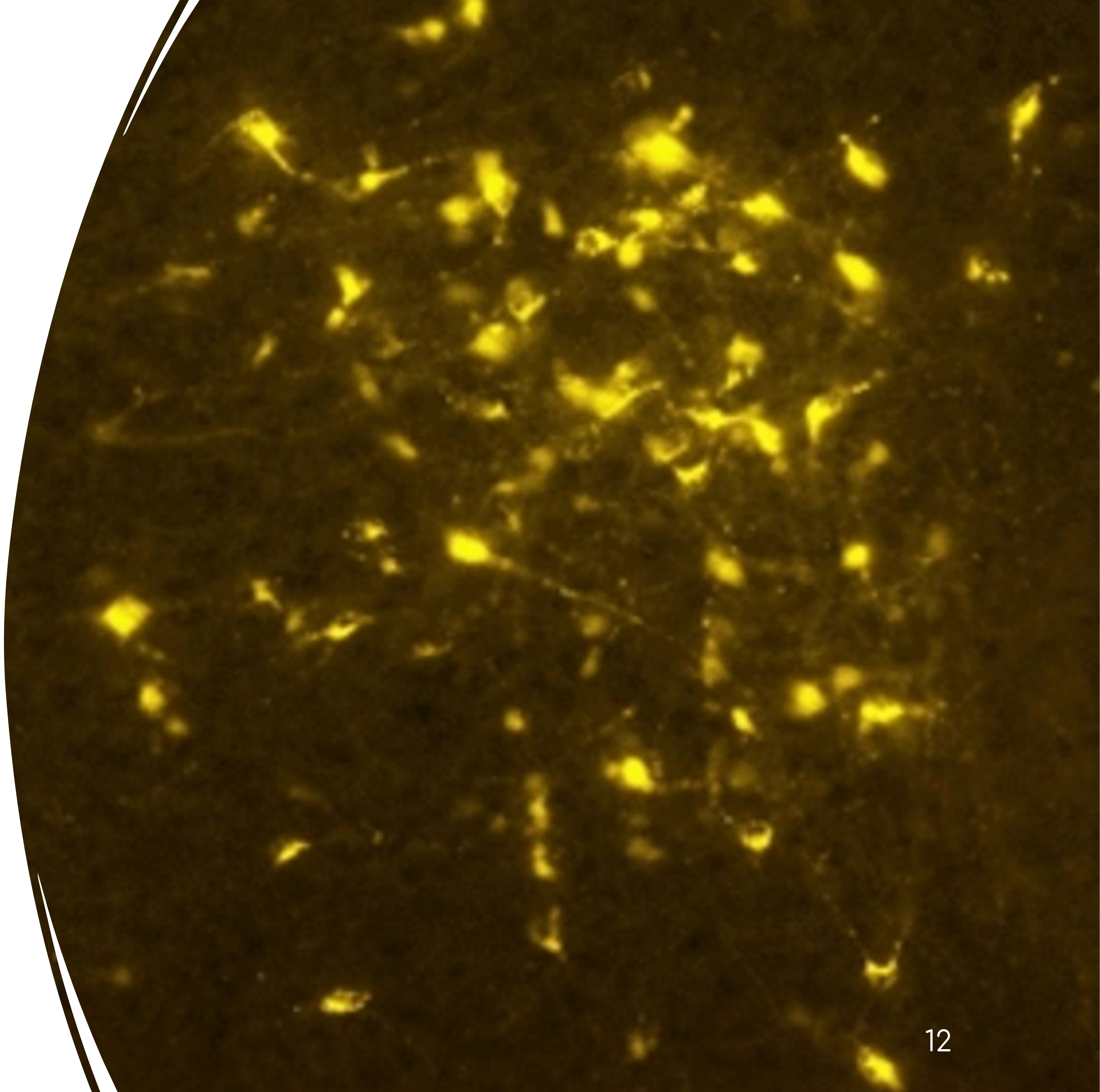
Strategia

- Counting by segmentation
  - Semantic segmentation
  - Convolutional neural network

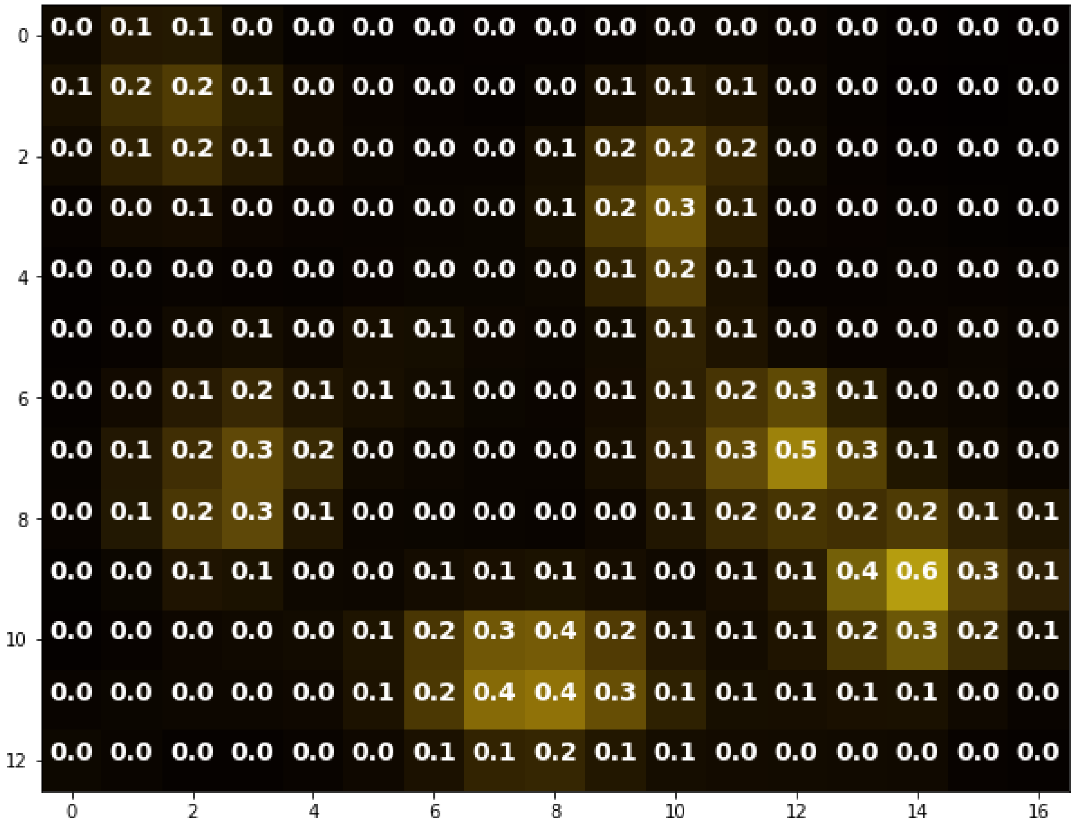


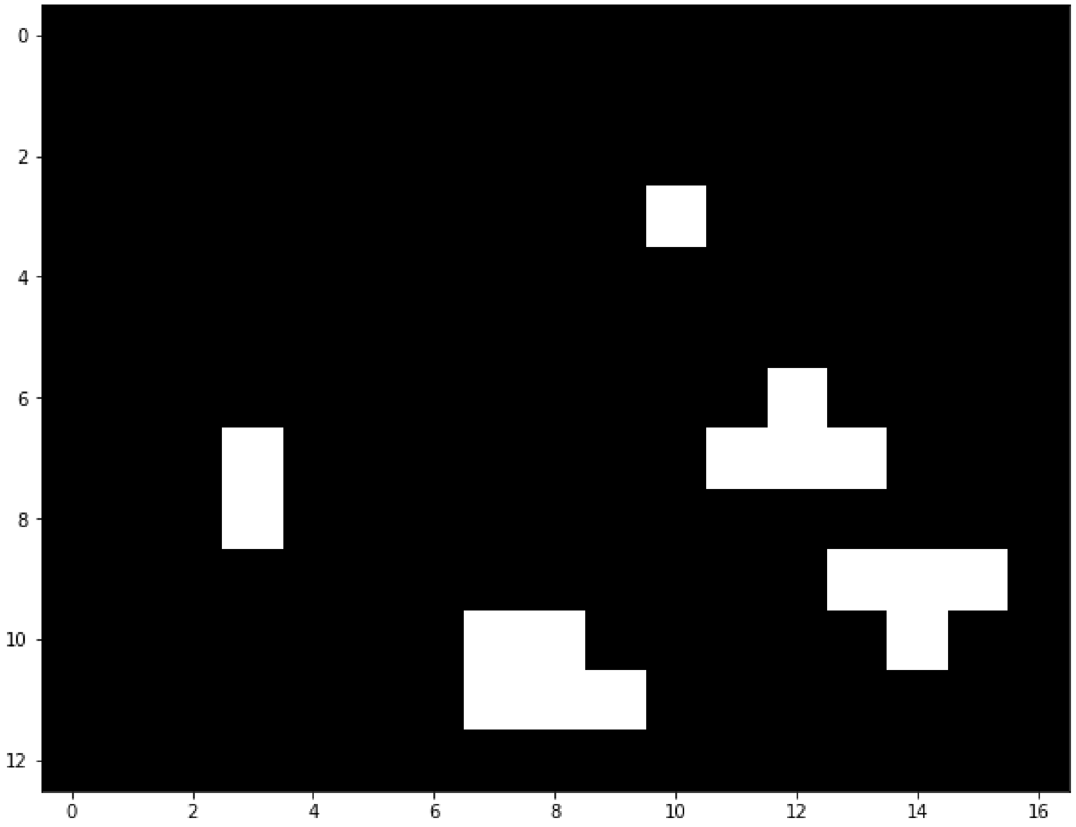
Fluorescent Neuronal Cells dataset

- 283 immagini ad alta risoluzione (1200x1600 px)
- Ottenere annotazioni è molto dispendioso
- Approccio semi-automatico
  - 252 bozze basate su filtro luminosità, rifinite poi da esperti
  - 31 immagini segmentate manualmente



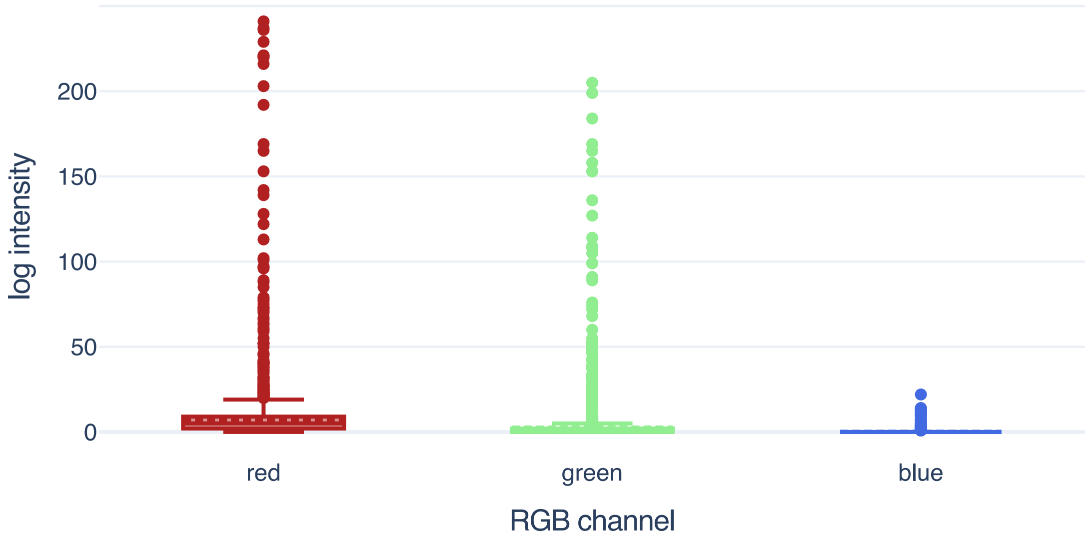
Filtro semi-automatico





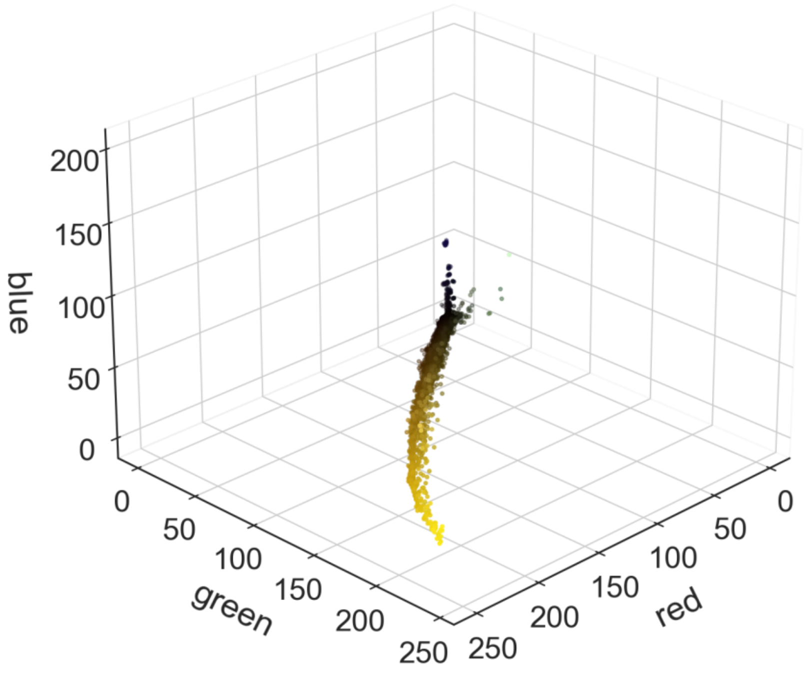
Canali RGB

quantile	red	green	blue
mean	7.32	2.83	0.20
s.d.	16.81	13.30	1.43
min	0	0	0
10%	0	0	0
25%	2	0	0
50%	5	2	0
75%	9	3	0
90%	12	4	0
max	252	251	87

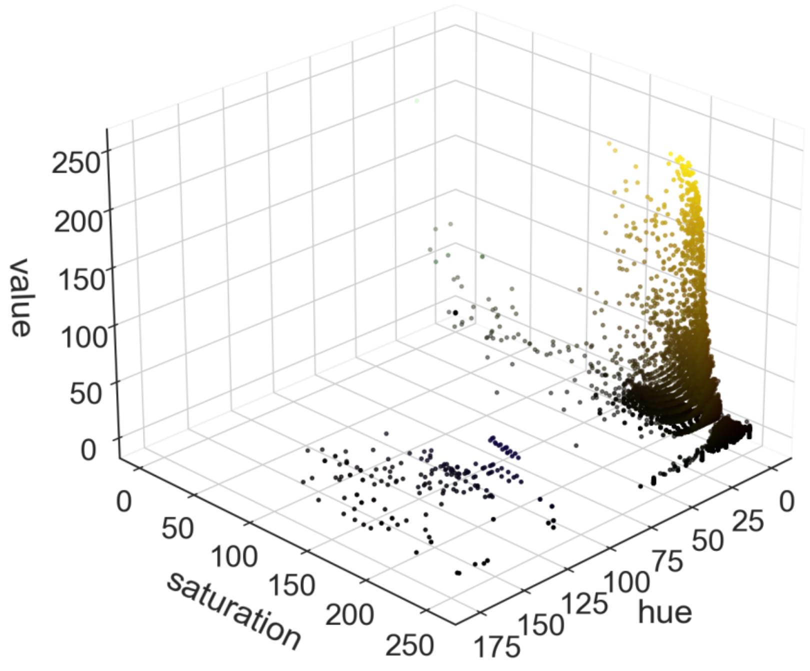


Encoding colore

RGB

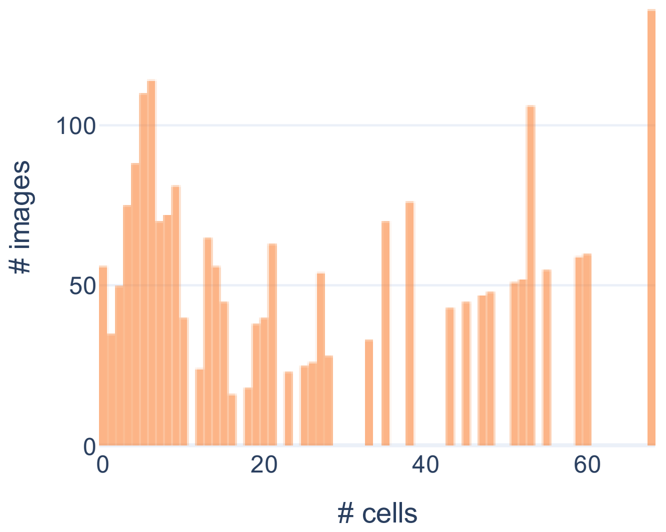


HSV



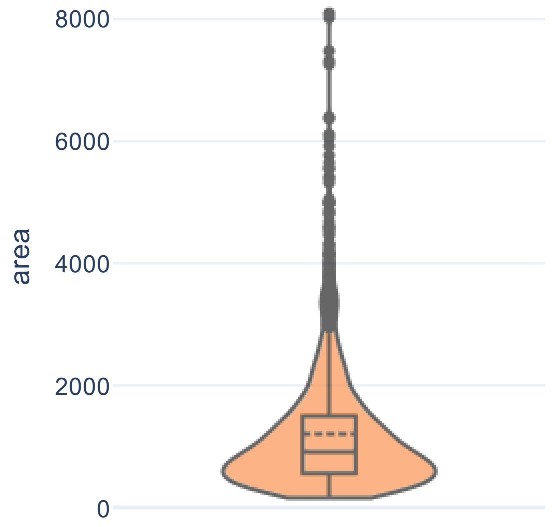
Distribuzione conteggi

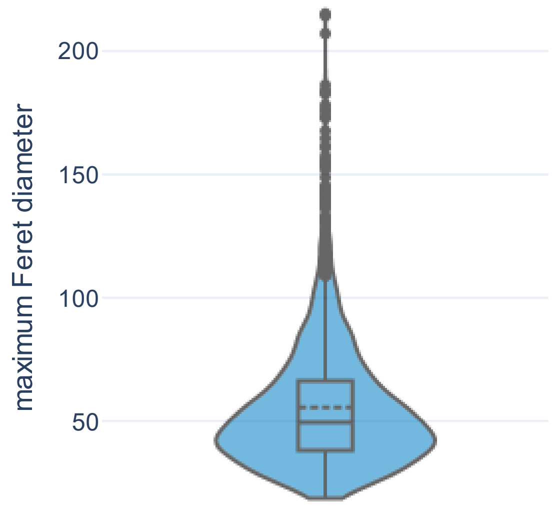
quantile	# cells/image
mean	27.05
s.d.	21.75
min	0
10%	4
25%	7
50%	21
75%	48
90%	59
max	68



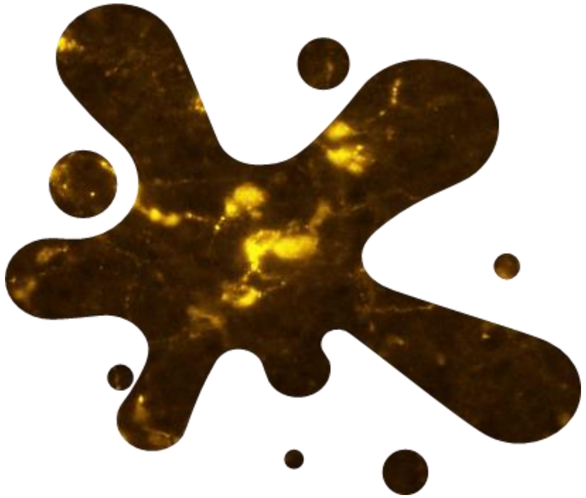
Area e Feret diameter

quantile	area (µm^2)	Feret Diameter (µm)
mean	119.30	17.48
s.d.	97.96	8.20
min	15.94	5.86
10%	35.23	9.42
25%	55.51	11.95
50%	89.86	15.53
75%	148.02	20.86
90%	237.09	27.61
max	796.39	67.48





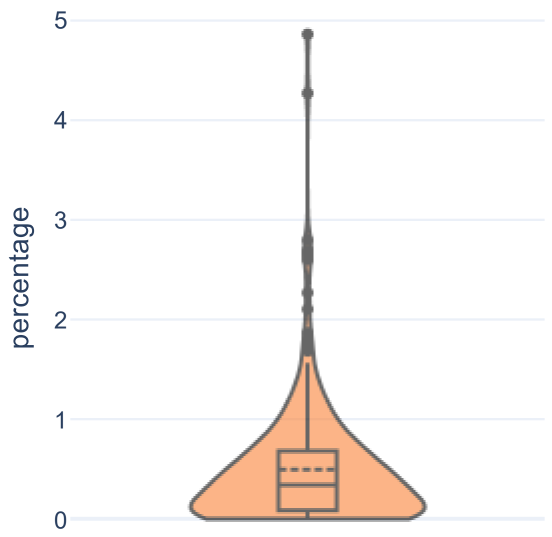
Sfide

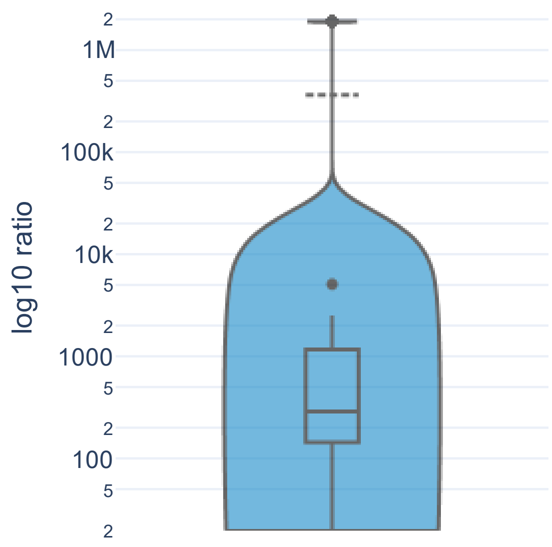


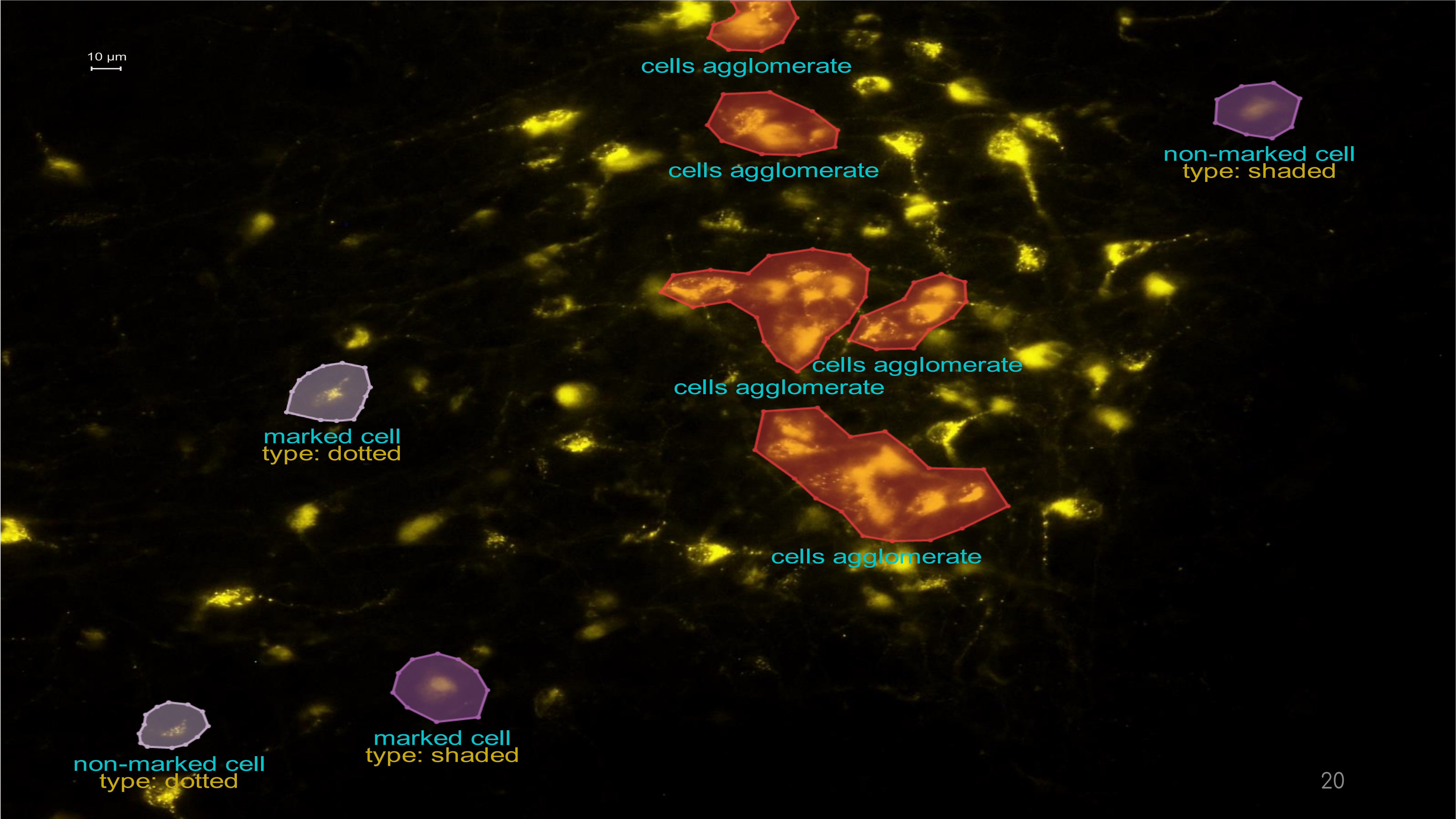
- Sbilanciamento classi
- Overcrowding
- Rumore etichette
- Artefatti

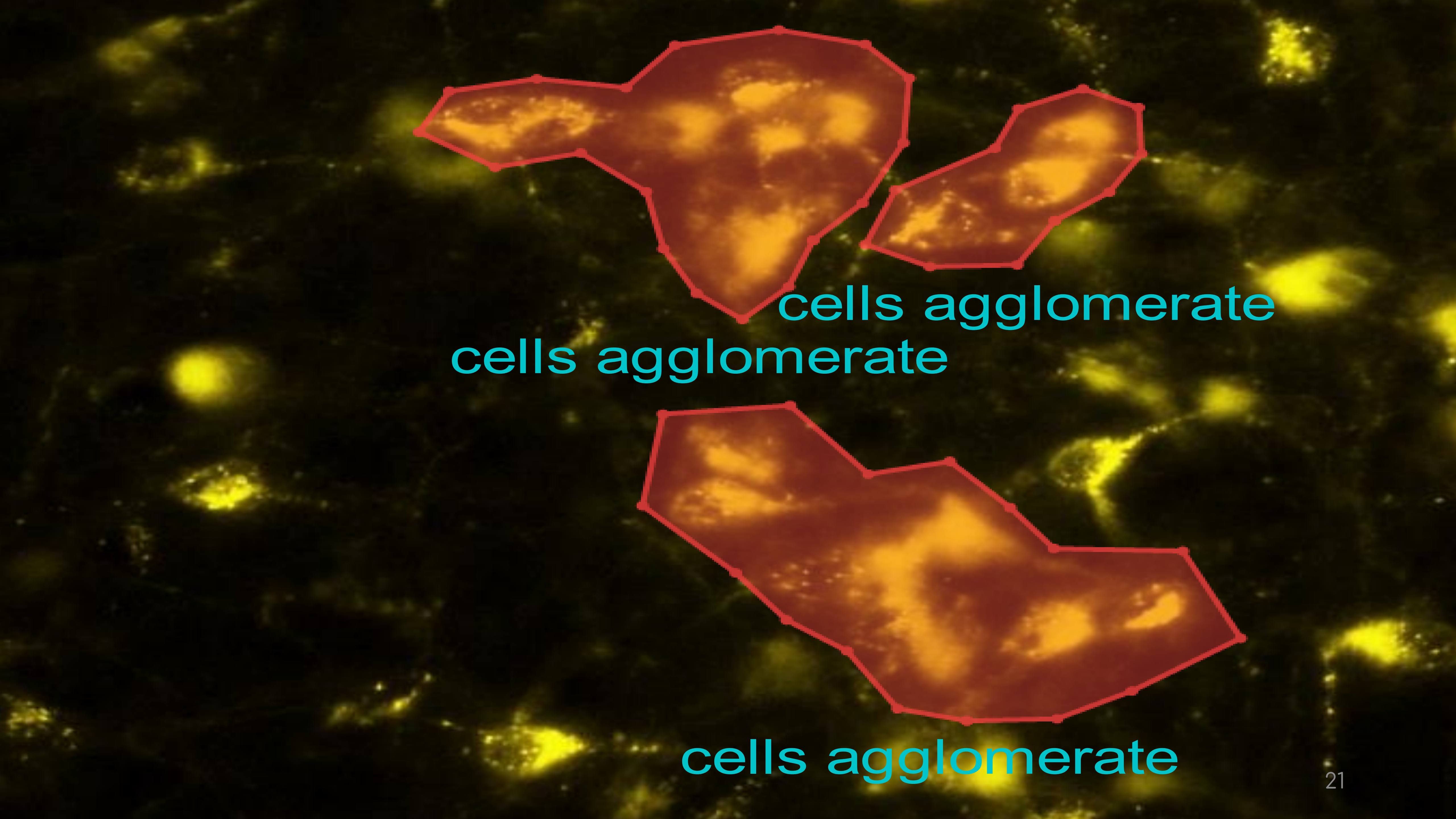
Class imbalance

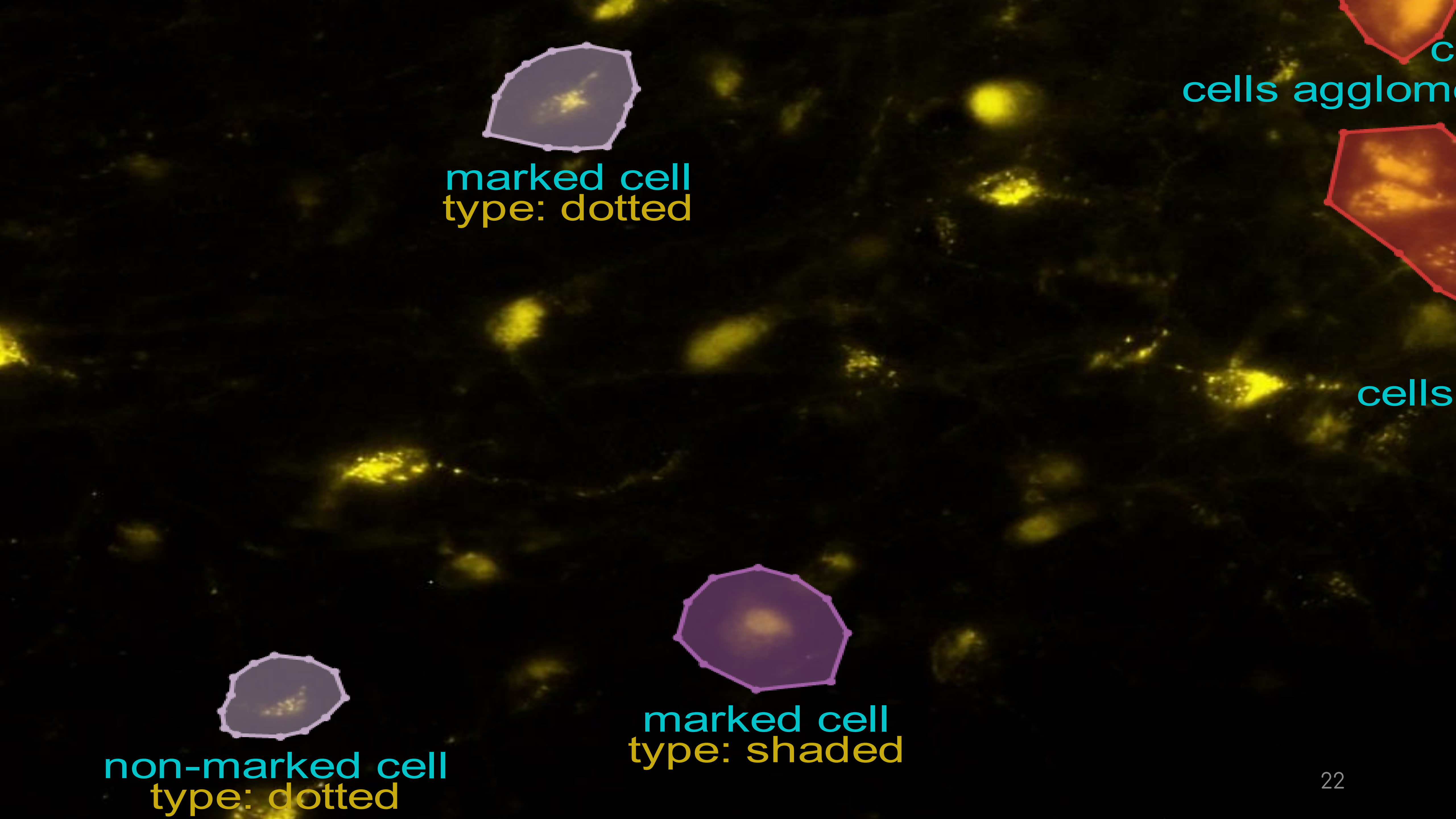
quantile	signal (%)	signal ratio
mean	0.50	367k
s.d.	0.61	756k
min	0	19.57
10%	0	92.39
25%	0.09	145.35
50%	0.34	291.10
75%	0.68	1k
90%	1.07	1.9M
max	4.86	1.9M

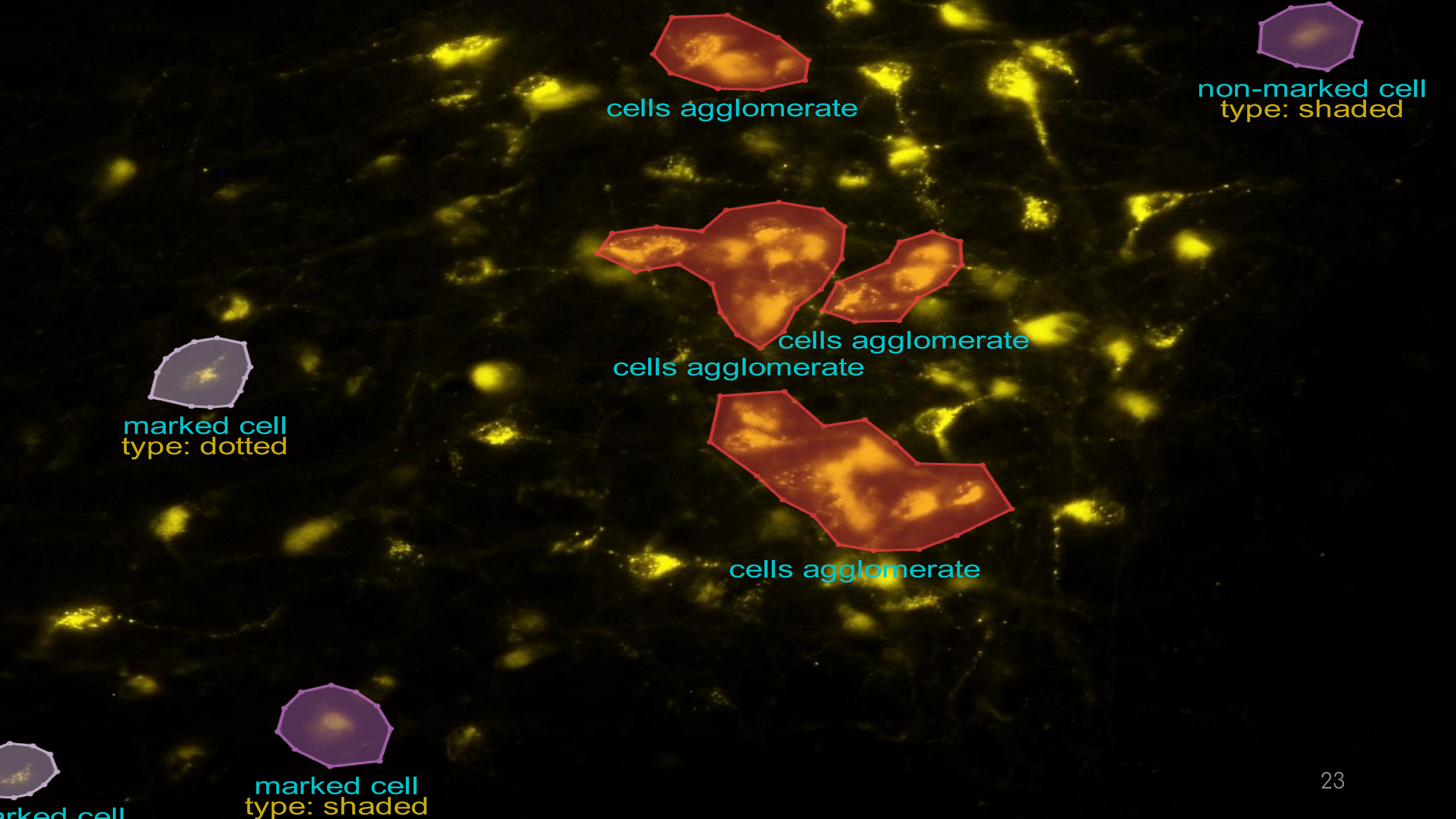


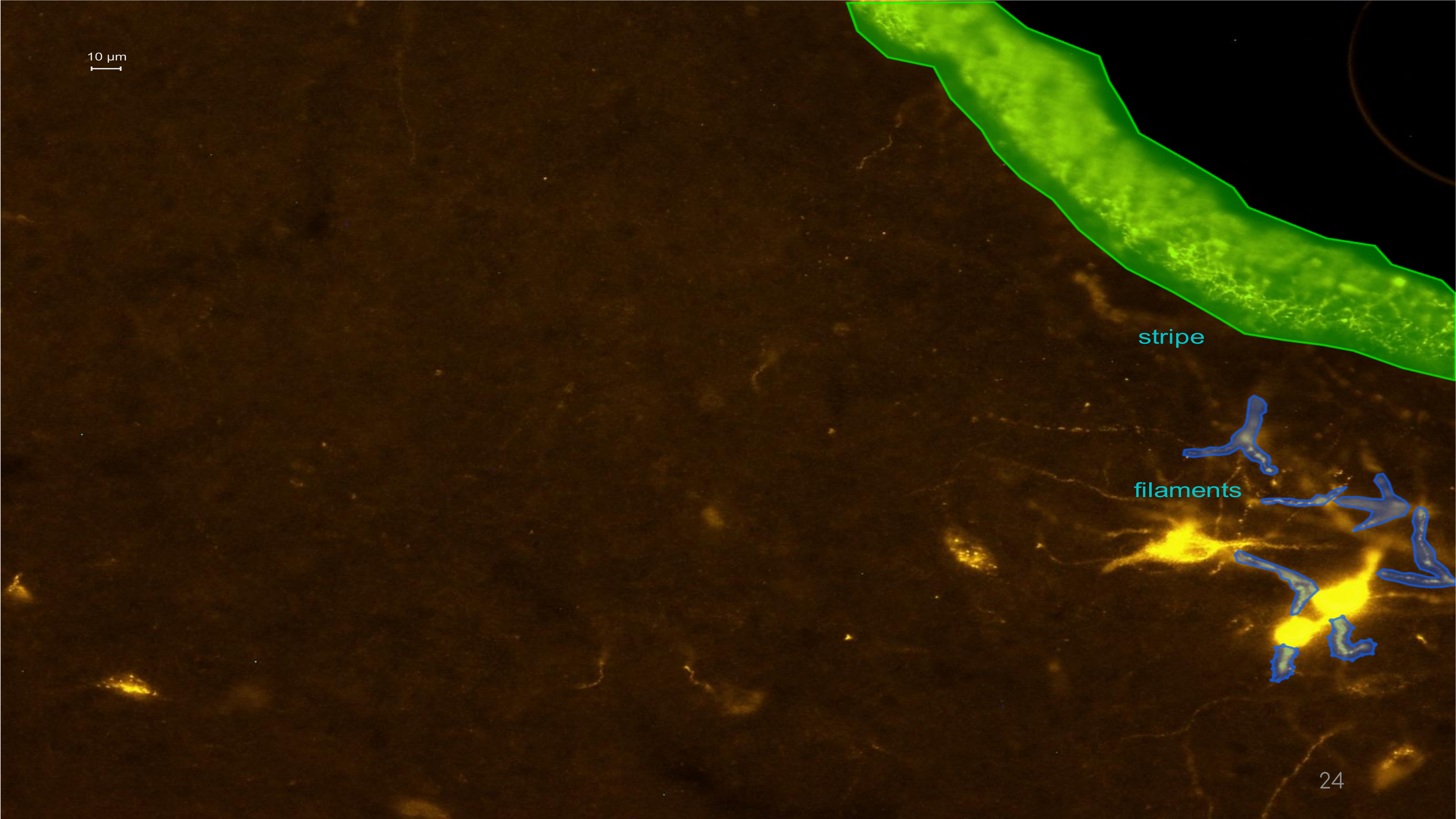


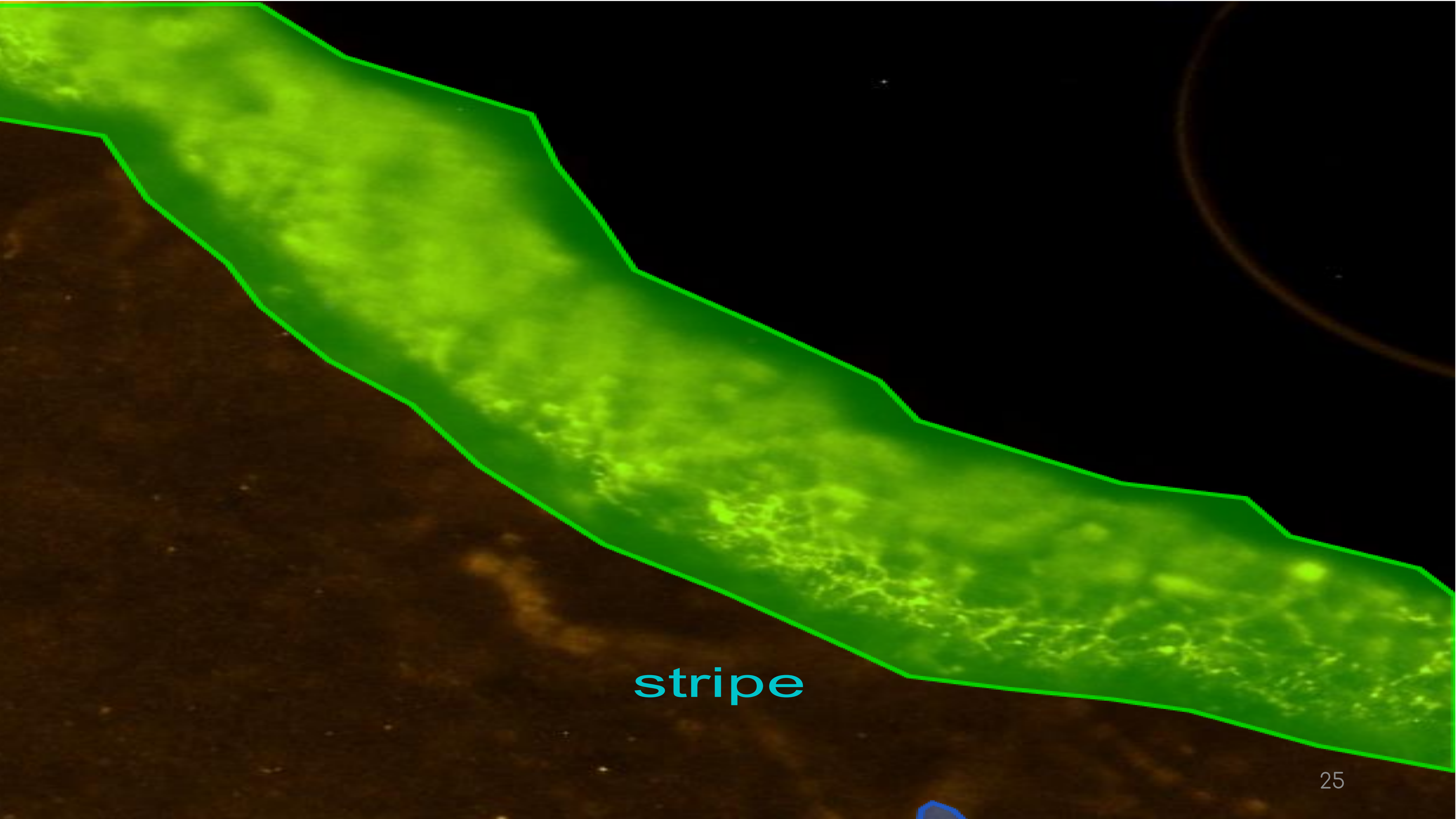

10 µm
cells agglomerate
cells agglomerate
non-marked cell
type: shaded
cells agglomerate
cells agglomerate
marked cell
type: dotted
cells agglomerate
non-marked cell
type: dotted
marked cell
type: shaded

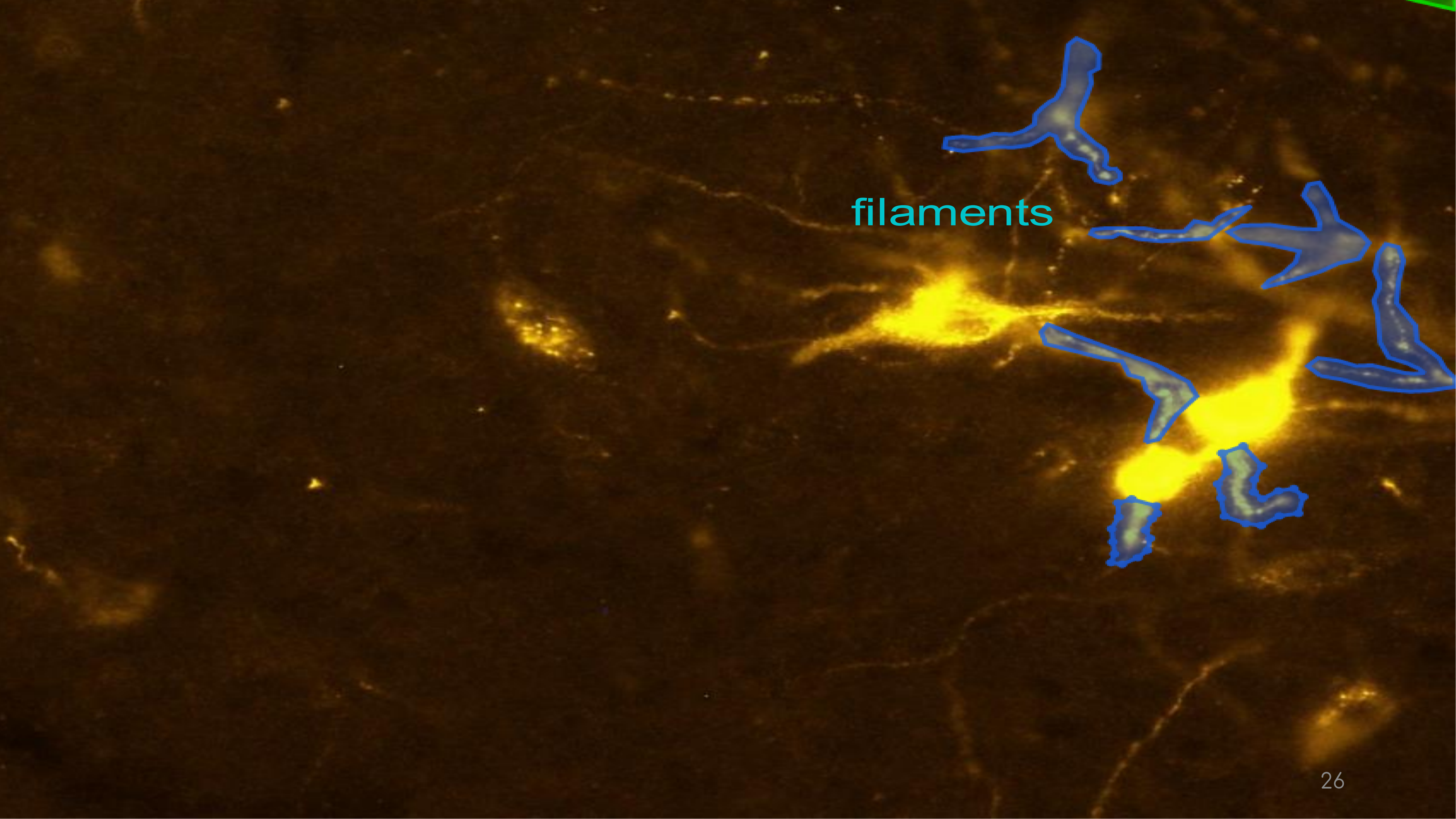

cells agglomerate
cells agglomerate
cells agglomerate

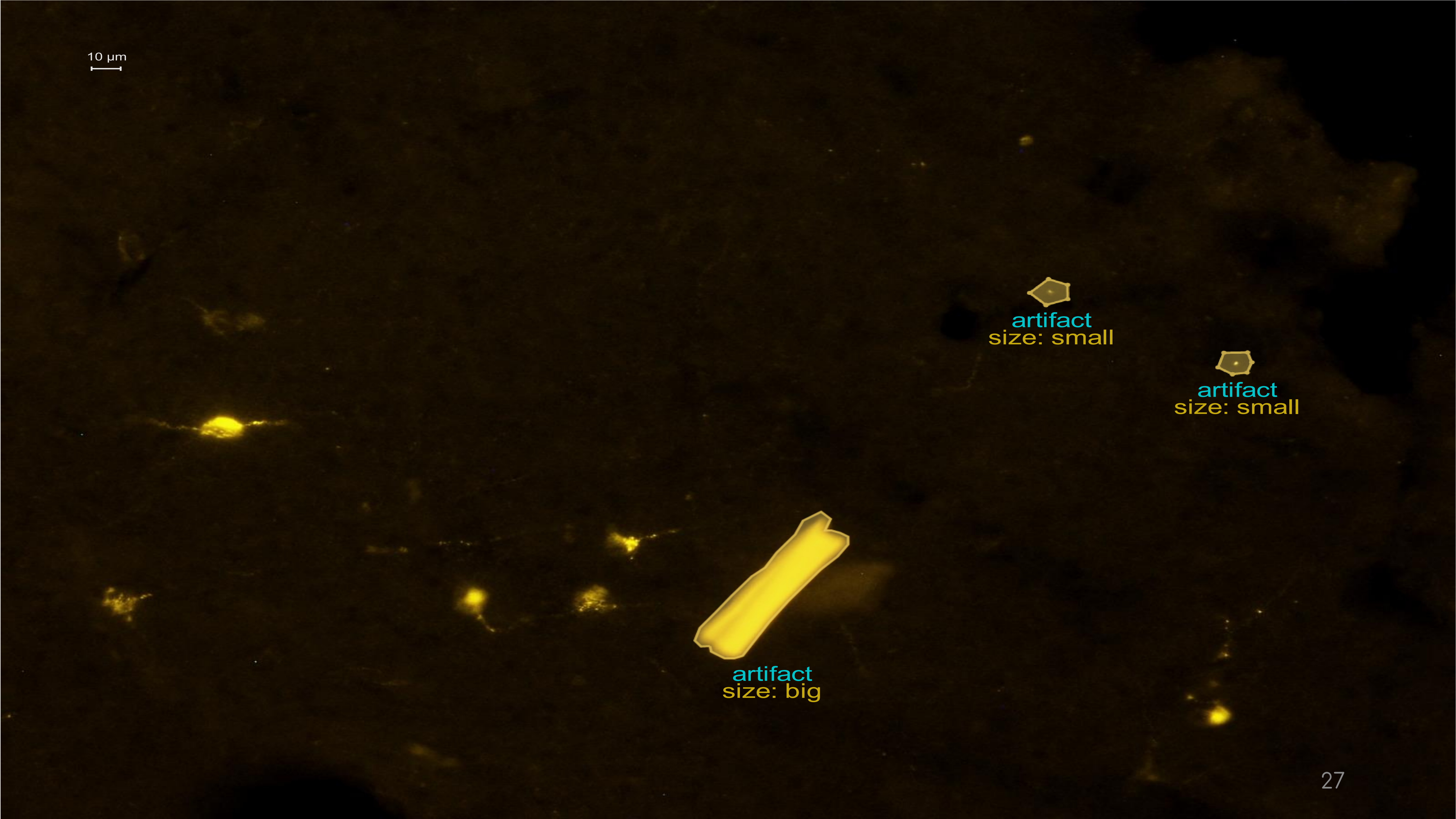

cells agglom
marked cell
type: dotted
cells
marked cell
type: shaded
non-marked cell
type: dotted


cells agglomerate
non-marked cell
type: shaded
cells agglomerate
cells agglomerate
marked cell
type: dotted
cells agglomerate
marked cell
type: shaded

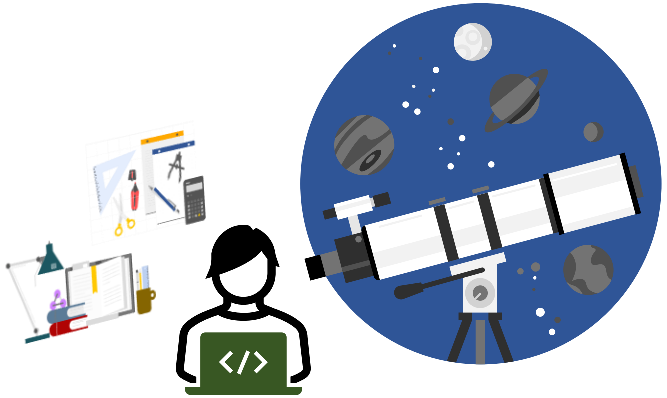

10 µm
stripe
filaments


stripe


filaments


10 µm
artifact
size: small
artifact
size: small
artifact
size: big

Metodi



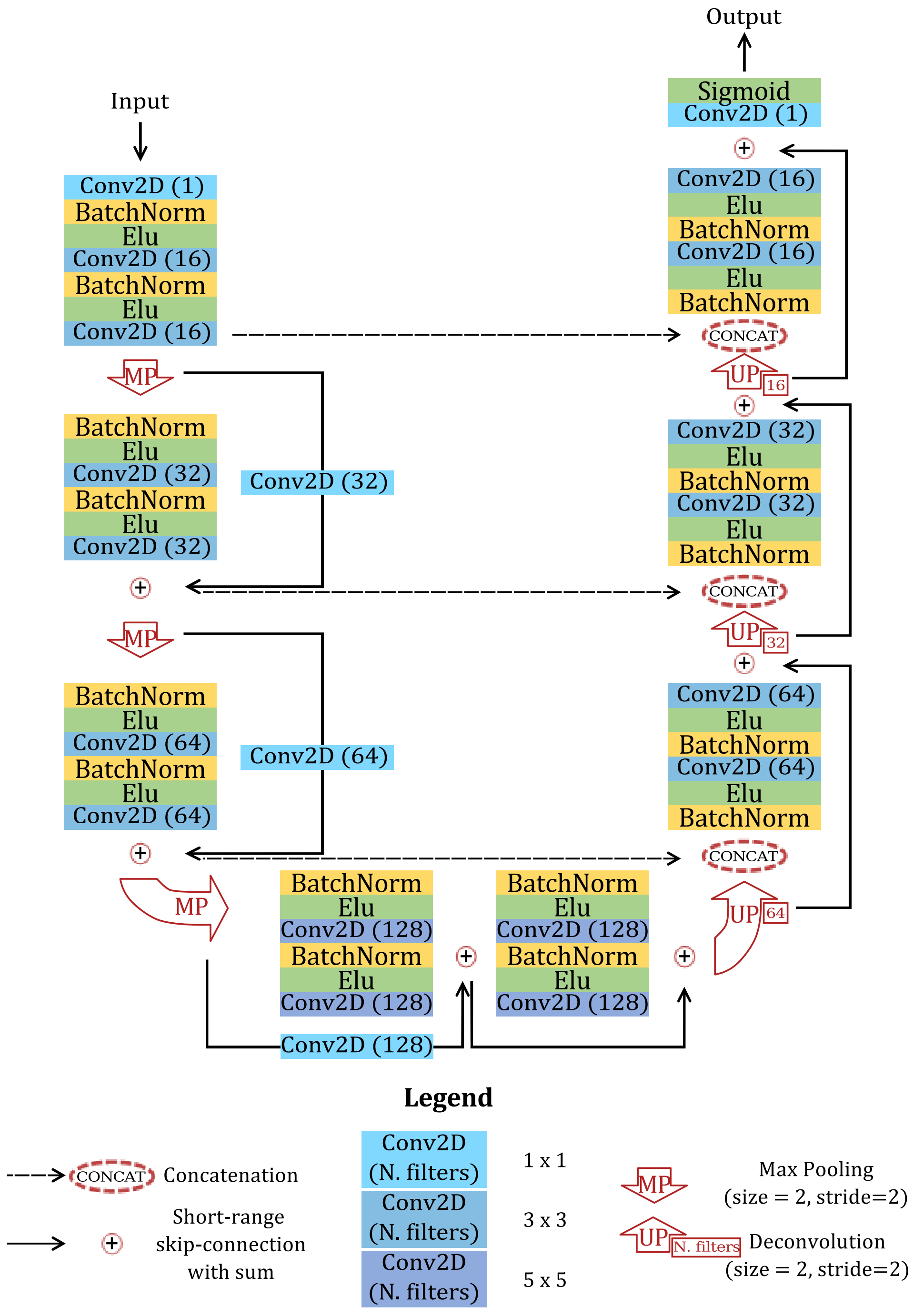
- c-ResUnet
- Ablation studies
  - Weight maps (WM)
  - Artifacts oversampling (AO)

Cell ResUnet

- Convoluzione 1x1 iniziale
  - Conversione grayscale

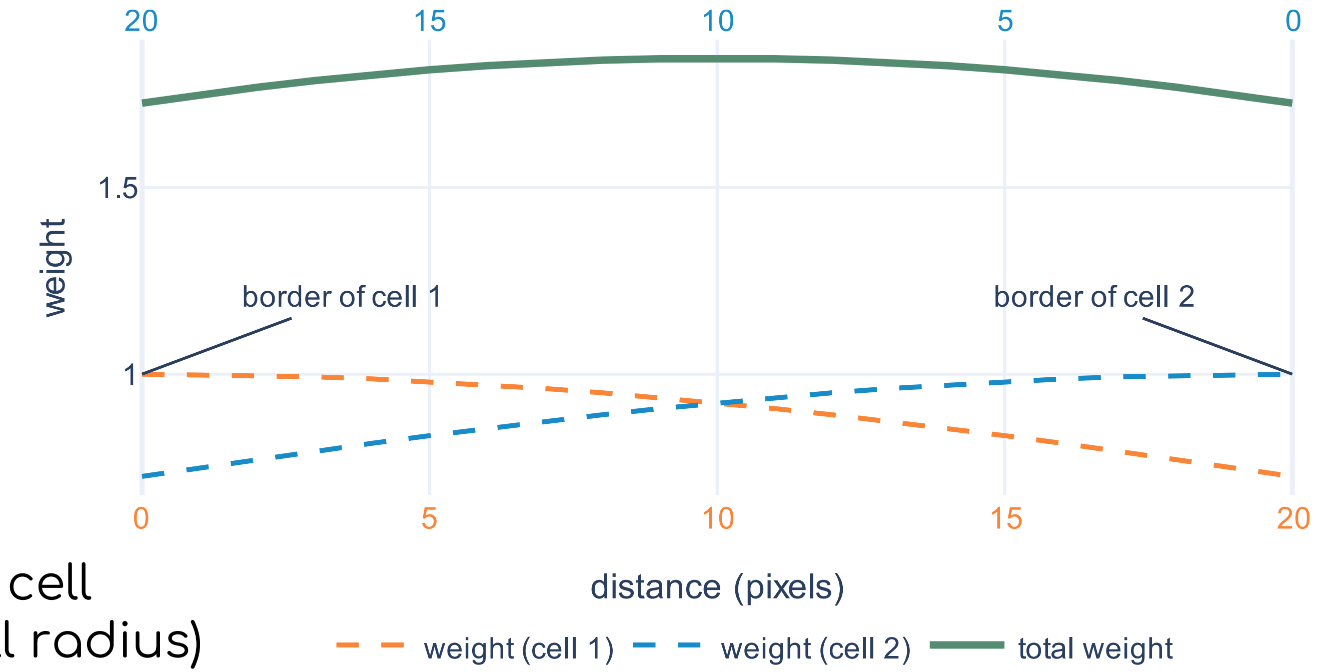
    → Colourspace appreso
- Bottleneck
  - Blocco residuo addizionale
  - Filtri 5x5 (anzichè 3x3)

    → Maggiore field-of-view



Weight maps

- Penalizzano errore in aree sovrafollate
- Effetto additivo

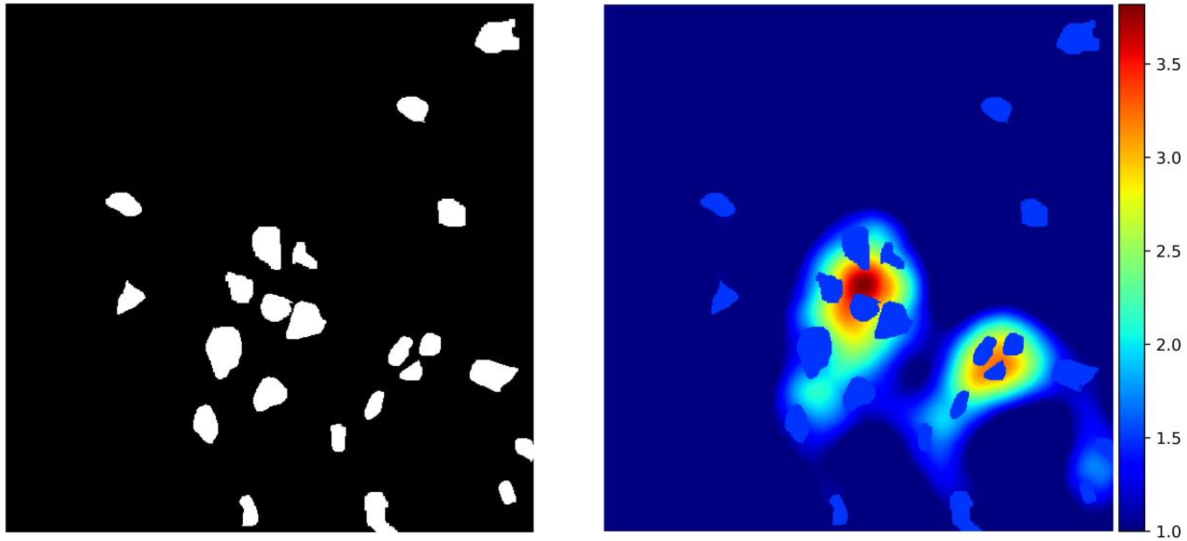


$$w = \exp\left\{-\frac{d^2}{2\sigma^2}\right\},$$

where
d: distance from a cell
σ = 25px (mean cell radius)

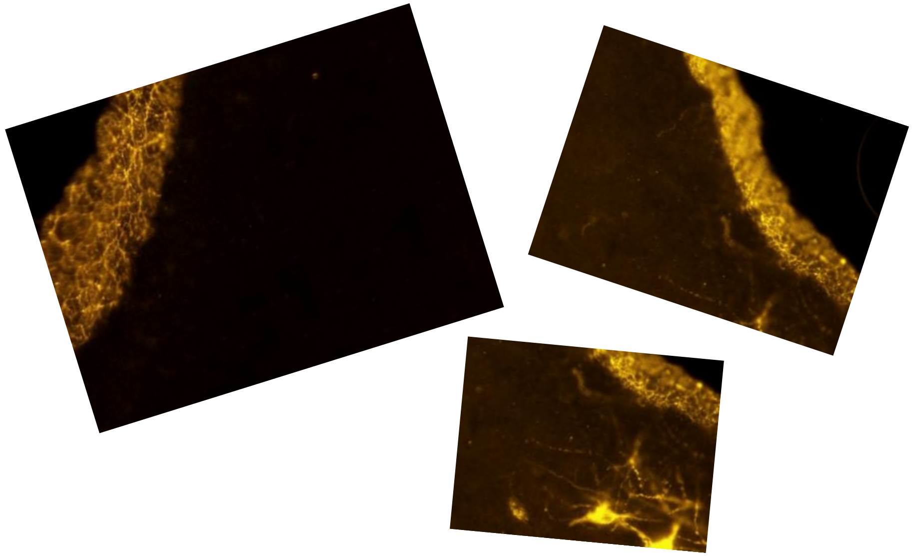
Weight maps

- Penalizzano errore in aree sovrafollate
- Effetto additivo



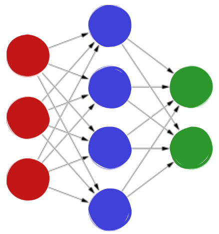
Artifacts oversampling

- Difficile da gestire
- Sottorappresentati
- Ricampionati con fattore 6



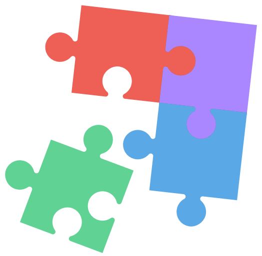
Experiments

Benchmark against SOTA architectures



- c-ResUnet (1.7M params)
- ResUnet (887k params)
- Unet (14M params)
- small Unet (876k params)

Ablation studies



- Full design (WM + AO)
- Weight maps only (no AO)
- Artifacts oversampling only (no WM)

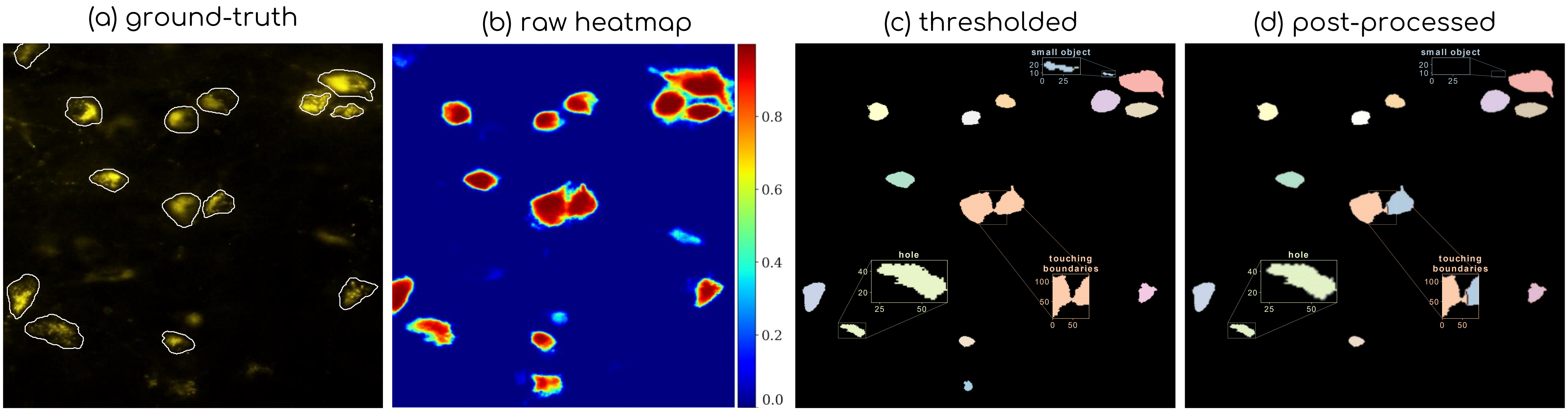
Metrics of performance

- Detection: F_1 score, AUC
- Counting : R^2, MAE

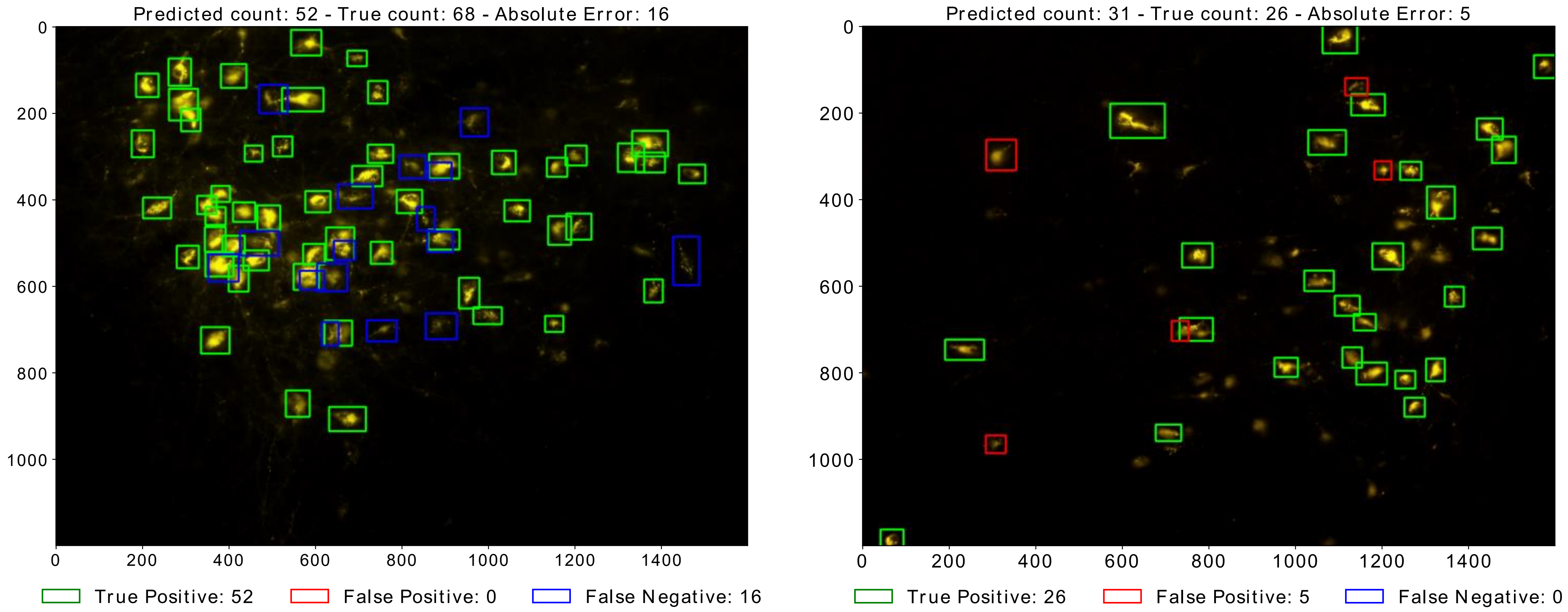
Training pipeline

- 150 training; 63 validation; 70 test
- 512x512 crop
- Augmentation
  - Rotazioni, rumore Gaussiano, variazione luminosità, deformazioni elastiche
  - Fattore re-sampling
    - 4 per immagini segmentate semi-automatically
    - 10 per immagini segmentate manualmente
    - 25 per artefatti

Post-processing

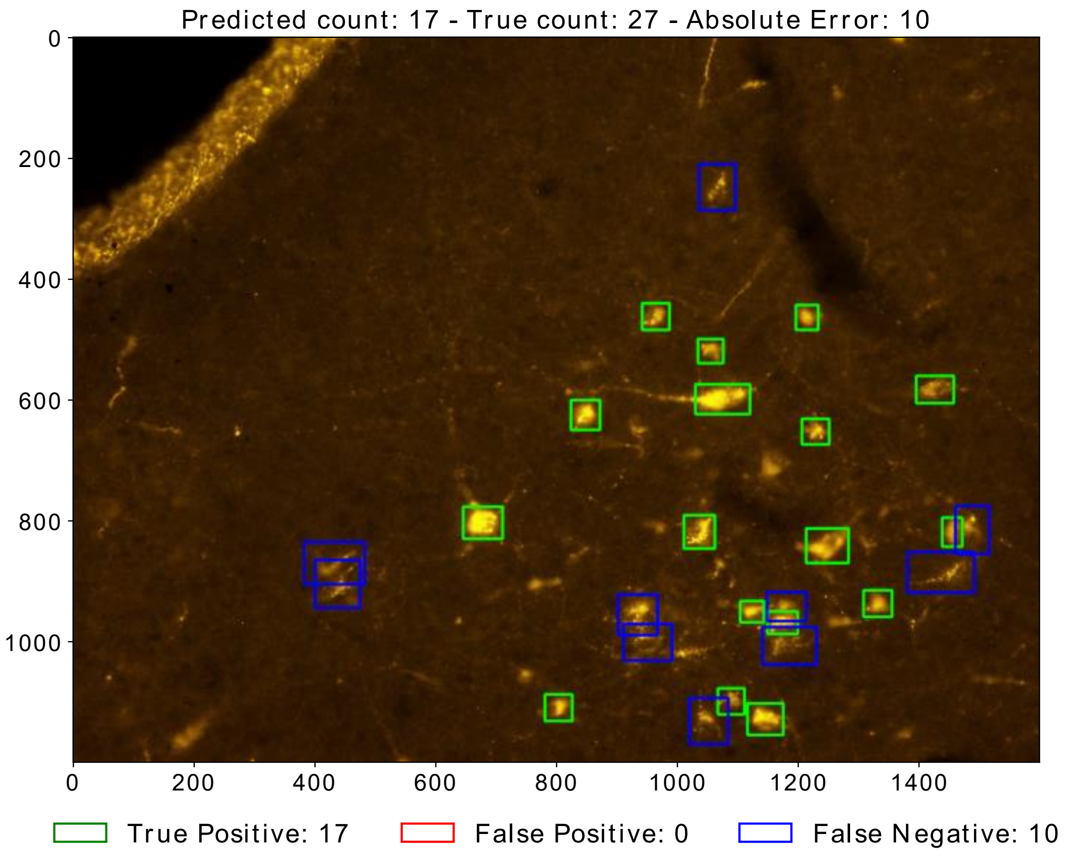


Qualitative evaluation: prediction

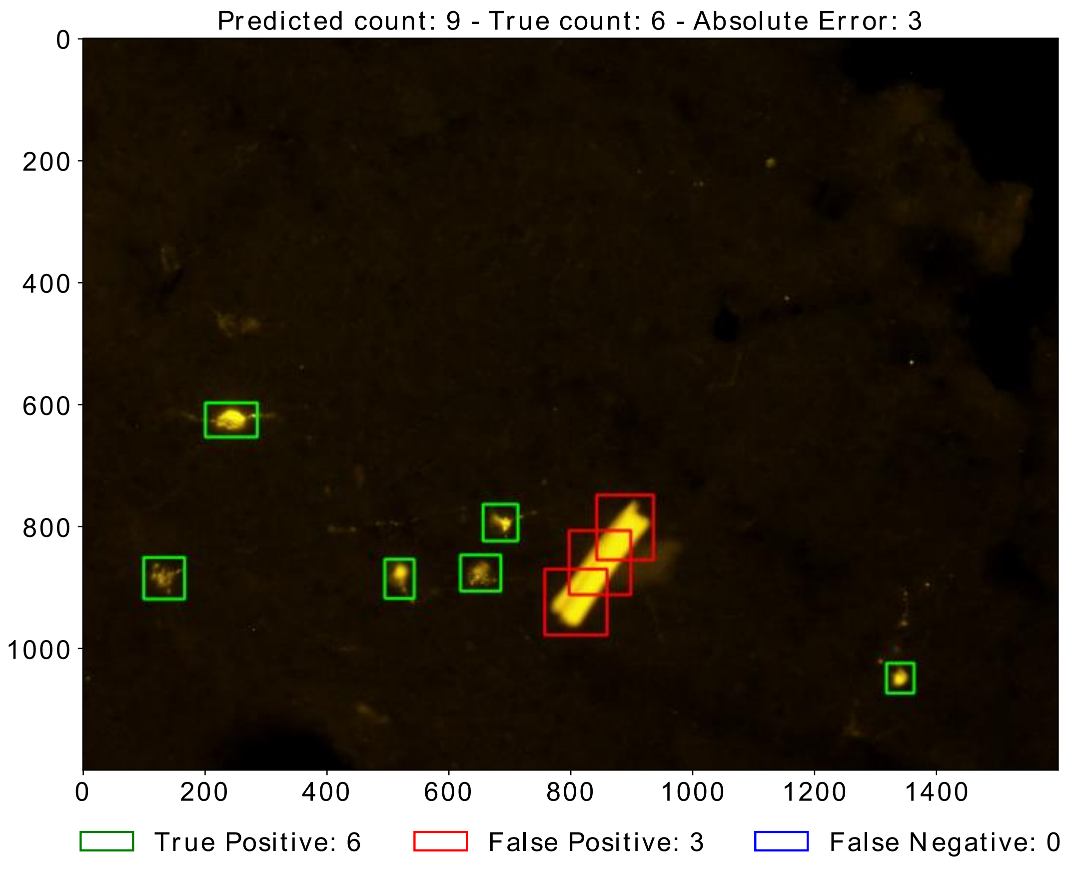


Qualitative evaluation: artifacts

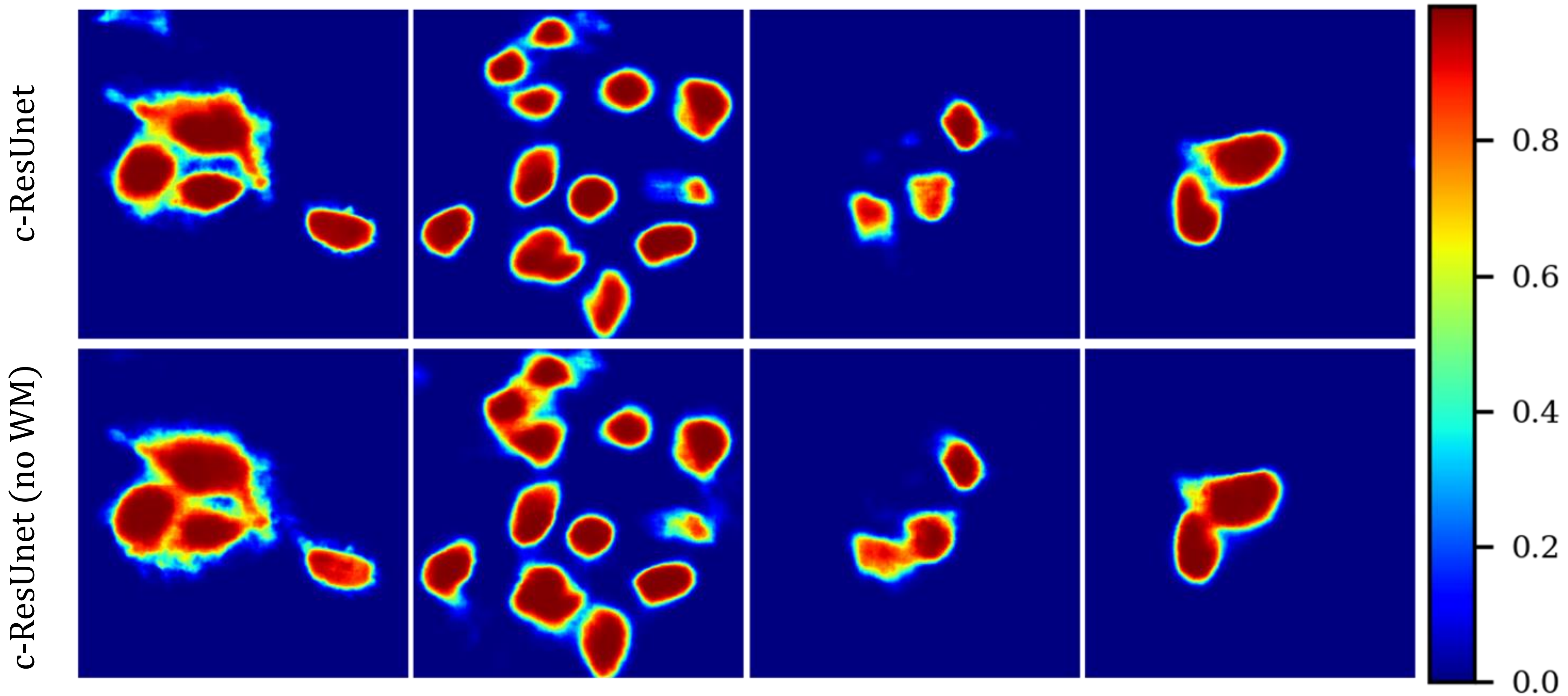
c-ResUnet (no AO)



c-ResUnet



Qualitative evaluation: weight maps



Quantitative evaluation

Model	Threshold	F_1	AUC	Accuracy	Precision	Recall	R^2	MAE	MedAE	MPE (%)
c-ResUnet	**0.875**	**0.8149**	**0.8705**	**0.6877**	0.9081	**0.7391**	**0.8215**	**3.0857**	**1.0**	-5.13
c-ResUnet (no AO)	0.875	0.8047	0.8741	0.6732	0.9019	0.7264	0.8077	3.0857	1.5	-6.24
c-ResUnet (no WM)	0.875	0.7613	0.8594	0.6147	0.9418	0.6389	0.7048	3.6857	**1.0**	-19.14
ResUnet	0.850	0.7855	0.8579	0.6468	0.8865	0.7052	0.7831	3.3286	**1.0**	**-4.84**
ResUnet (no WM)	0.850	0.7513	0.8643	0.6016	0.9387	0.6262	0.6955	4.0571	2.0	-24.12
Unet	0.875	0.7724	0.8609	0.6291	0.9117	0.6700	0.7560	3.5143	1.5	-14.36
Unet (no WM)	0.850	0.7886	0.8461	0.6510	0.8989	0.7024	0.8069	3.1571	2.0	-9.23
small Unet	0.875	0.7563	0.8691	0.6081	0.9264	0.6389	0.7682	3.5714	2.0	-21.37
small Unet (no WM)	0.825	0.6697	0.8326	0.5034	**0.9483**	0.5176	0.5723	4.7714	2.0	-32.01
Adaptive Threshold	0.994	0.6106	0.0865	0.4394	0.5680	0.6601	0.3565	8.0143	6.0	78.26

Table 5.1: **Performance metrics.** Test set performance using the optimal *kneed* threshold. The first five columns report the detection metrics, while the latter ones evaluate counting performance.

Quantitative evaluation

Model	Threshold	F_1	AUC	Accuracy	Precision	Recall	R^2	MAE	MedAE	MPE (%)
c-ResUnet	**0.875**	**0.8149**	**0.8705**	**0.6877**	0.9081	**0.7391**	**0.8215**	**3.0857**	**1.0**	-5.13
c-ResUnet (no AO)	0.875	0.8047	0.8741	0.6732	0.9019	0.7264	0.8077	3.0857	1.5	-6.24
c-ResUnet (no WM)	0.875	0.7613	0.8594	0.6147	0.9418	0.6389	0.7048	3.6857	**1.0**	-19.14
ResUnet	0.850	0.7855	0.8579	0.6468	0.8865	0.7052	0.7831	3.3286	**1.0**	**-4.84**
ResUnet (no WM)	0.850	0.7513	0.8643	0.6016	0.9387	0.6262	0.6955	4.0571	2.0	-24.12
Unet	0.875	0.7724	0.8609	0.6291	0.9117	0.6700	0.7560	3.5143	1.5	-14.36
Unet (no WM)	0.850	0.7886	0.8461	0.6510	0.8989	0.7024	0.8069	3.1571	2.0	-9.23
small Unet	0.875	0.7563	0.8691	0.6081	0.9264	0.6389	0.7682	3.5714	2.0	-21.37
small Unet (no WM)	0.825	0.6697	0.8326	0.5034	**0.9483**	0.5176	0.5723	4.7714	2.0	-32.01
Adaptive Threshold	0.994	0.6106	0.0865	0.4394	0.5680	0.6601	0.3565	8.0143	6.0	78.26

Table 5.1: **Performance metrics.** Test set performance using the optimal *kneed* threshold. The first five columns report the detection metrics, while the latter ones evaluate counting performance.

Quantitative evaluation

Model	Threshold	F_1	AUC	Accuracy	Precision	Recall	R^2	MAE	MedAE	MPE (%)
c-ResUnet	**0.875**	**0.8149**	**0.8705**	**0.6877**	0.9081	**0.7391**	**0.8215**	**3.0857**	**1.0**	-5.13
c-ResUnet (no AO)	0.875	0.8047	0.8741	0.6732	0.9019	0.7264	0.8077	3.0857	1.5	-6.24
c-ResUnet (no WM)	0.875	0.7613	0.8594	0.6147	0.9418	0.6389	0.7048	3.6857	**1.0**	-19.14
ResUnet	0.850	0.7855	0.8579	0.6468	0.8965	0.7052	0.7831	3.3286	**1.0**	**-4.84**
ResUnet (no WM)	0.850	0.7513	0.8643	0.6016	0.9387	0.6262	0.6955	4.0571	2.0	-24.12
Unet	0.875	0.7724	0.8609	0.6291	0.9117	0.6700	0.7560	3.5143	1.5	-14.36
Unet (no WM)	0.850	0.7886	0.8461	0.6510	0.8989	0.7024	0.8069	3.1571	2.0	-9.23
small Unet	0.875	0.7563	0.8691	0.6081	0.9264	0.6389	0.7682	3.5714	2.0	-21.37
small Unet (no WM)	0.825	0.6697	0.8326	0.5034	**0.9483**	0.5176	0.5723	4.7714	2.0	-32.01
Adaptive Threshold	0.994	0.6106	0.0865	0.4394	0.5680	0.6601	0.3565	8.0143	6.0	78.26

Table 5.1: **Performance metrics.** Test set performance using the optimal *kneed* threshold. The first five columns report the detection metrics, while the latter ones evaluate counting performance.

Quantitative evaluation

Model	Threshold	F_1
c-ResUnet	**0.875**	0.8149
c-ResUnet (no AO)	0.875	0.8047
c-ResUnet (no WM)	0.875	0.7613
ResUnet	0.850	0.7855
ResUnet (no WM)	0.850	0.7513
Unet	0.875	0.7724
Unet (no WM)	0.850	0.7886
small Unet	0.875	0.7563
small Unet (no WM)	0.825	0.6697

Quantitative evaluation

Model	Threshold	F_1	AUC	Accuracy	Precision	Recall	R^2	MAE	MedAE	MPE (%)
c-ResUnet	**0.875**	[illegible]	<u>0.8705</u>	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
c-ResUnet (no AO)	0.875	[illegible]	0.8741	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
c-ResUnet (no WM)	0.875	[illegible]	0.8594	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ResUnet	0.850	[illegible]	0.8579	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ResUnet (no WM)	0.850	[illegible]	0.8643	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Unet	0.875	[illegible]	0.8609	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Unet (no WM)	0.850	[illegible]	0.8461	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
small Unet	0.875	[illegible]	0.8691	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
small Unet (no WM)	0.825	[illegible]	0.8326	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Adaptive Threshold	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Table 5.1: **Performance metrics.** Test set performance using the optimal *kneed* threshold. The first five columns report the detection metrics, while the latter ...es evaluate counting performance.

Quantitative evaluation

Model	Threshold	F_1	AUC	Accuracy	Precision	Recall	R^2	MAE	MedAE	MPE (%)
c-ResUnet	**0.875**	**0.8149**	**0.8705**	**0.6877**	0.9081	**0.7391**	**0.8215**	**3.0857**	**1.0**	-5.13
c-ResUnet (no AO)	0.875	0.8047	0.8741	0.6732	0.9019	0.7264	0.8077	3.0857	1.5	-6.24
c-ResUnet (no WM)	0.875	0.7613	0.8594	0.6147	0.9418	0.6389	0.7048	3.6857	1.0	-19.14
ResUnet	0.850	0.7855	0.8579	0.6468	0.8865	0.7052	0.7831	3.3286	1.0	**-4.84**
ResUnet (no WM)	0.850	0.7513	0.8643	0.6016	0.9387	0.6262	0.6955	4.0571	2.0	-24.12
Unet	0.875	0.7724	0.8609	0.6291	0.9117	0.6700	0.7560	3.5143	1.5	-14.36
Unet (no WM)	0.850	0.7886	0.8461	0.6510	0.8989	0.7024	0.8069	3.1571	2.0	-9.23
small Unet	0.875	0.7563	0.8691	0.6081	0.9264	0.6389	0.7682	3.5714	2.0	-21.37
small Unet (no WM)	0.825	0.6697	0.8326	0.5034	**0.9483**	0.5176	0.5723	4.7714	2.0	-32.01
Adaptive Threshold	0.994	0.6106	0.0865	0.4394	0.5680	0.6601	0.3565	8.0143	6.0	78.26

Table 5.1: **Performance metrics.** Test set performance using the optimal *kneed* threshold. The first five columns report the detection metrics, while the latter …es evaluate counting performance.

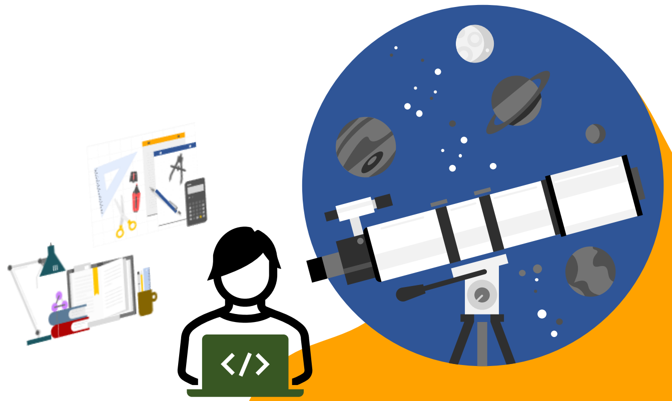
Quantitative evaluation

Model	Threshold	F_1	AUC	Accuracy	Precision	Recall	R^2	MAE	MedAE	MPE (%)
c-ResUnet	**0.875**	**0.8149**	**0.8705**	[illegible]	[illegible]	[illegible]	**0.8215**	**3.0857**	[illegible]	[illegible]
c-ResUnet (no AO)	0.875	0.8047	0.8741	[illegible]	[illegible]	[illegible]	0.8077	3.0857	[illegible]	[illegible]
c-ResUnet (no WM)	0.875	0.7613	0.8594	[illegible]	[illegible]	[illegible]	0.7048	3.6857	[illegible]	[illegible]
ResUnet	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ResUnet (no WM)	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Unet	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Unet (no WM)	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
small Unet	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
small Unet (no WM)	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Adaptive Threshold	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Table 5.1: **Performance metrics.** Test set performance using the optimal *kneed* threshold. The first five columns report the detection metrics, while the latter ...es evaluate counting performance.

Recap

- 👍 c-ResUnet performa meglio dei competitor
- 👍 Prestazioni simili ad operatore umano
- 👍 "Effetto operatore" sistematico
- 👍 Weight map utili
- 👍 Pubblicazione dataset e modello allenato
- 👎 Artefatti



Punti aperti

Lifecycle completo:

- Messa in produzione
- Labeling

Punti aperti: produzione

Poco fruibile senza un punto d'accesso per utilizzatori finali (e.g. servizio web)

- Messa in produzione modello

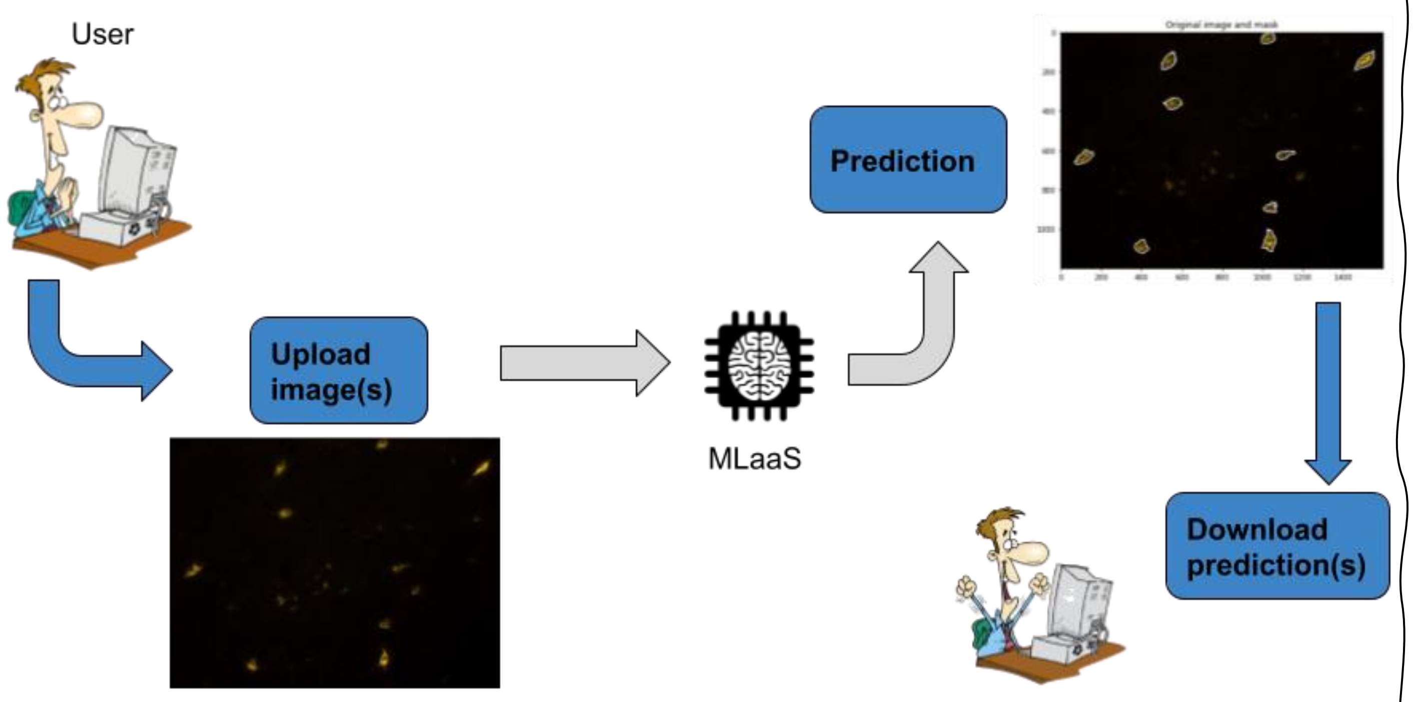


Figura 1: minimal workflow

Punti aperti: produzione

Poco fruibile senza un punto d'accesso per utilizzatori finali (e.g. servizio web)

- Messa in produzione modello
- Data access

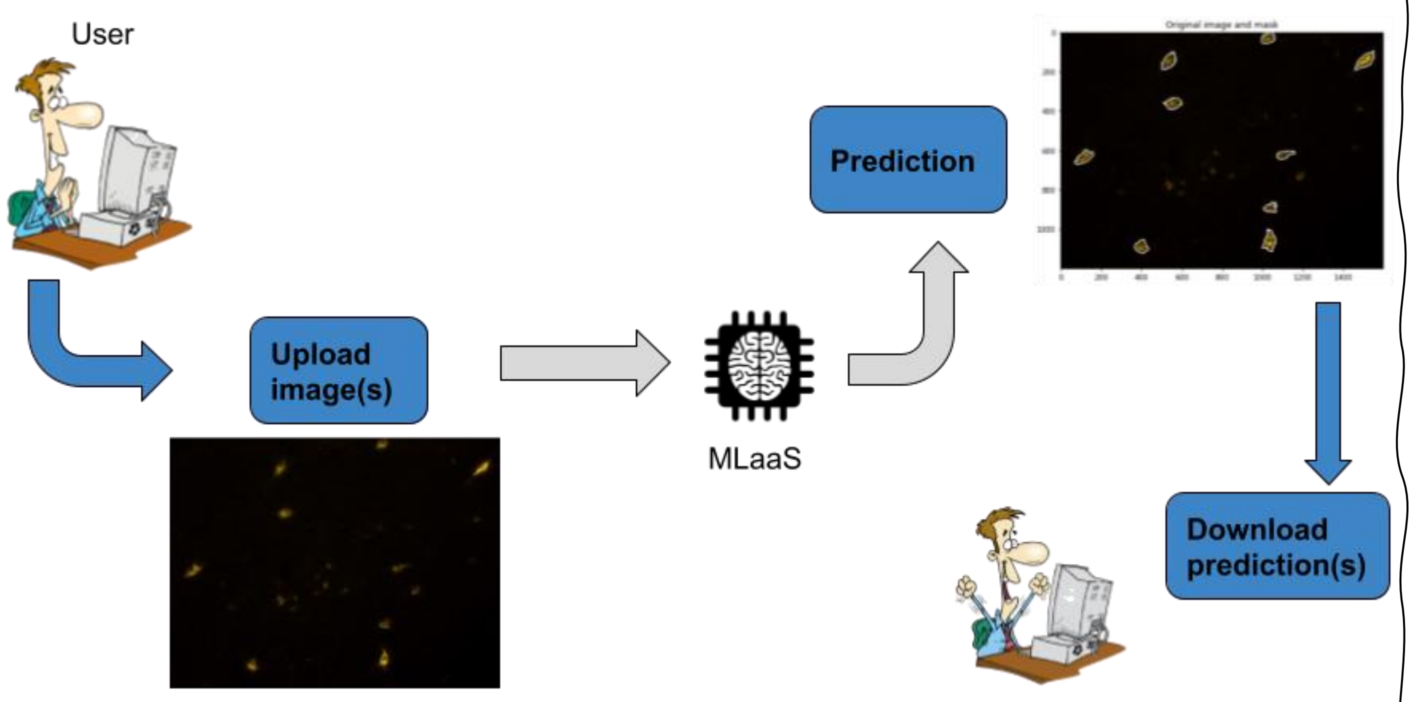


Figura 1: minimal workflow

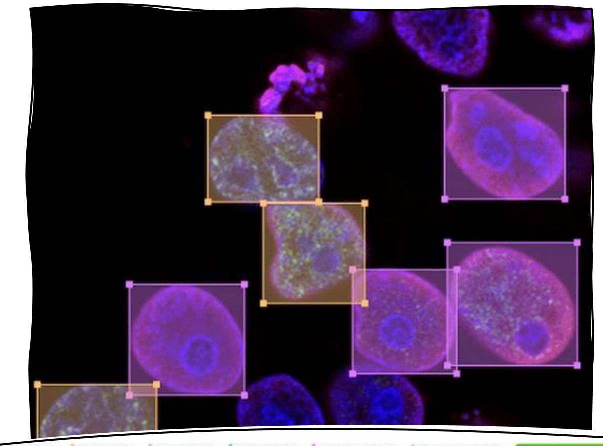
Punti aperti: labeling

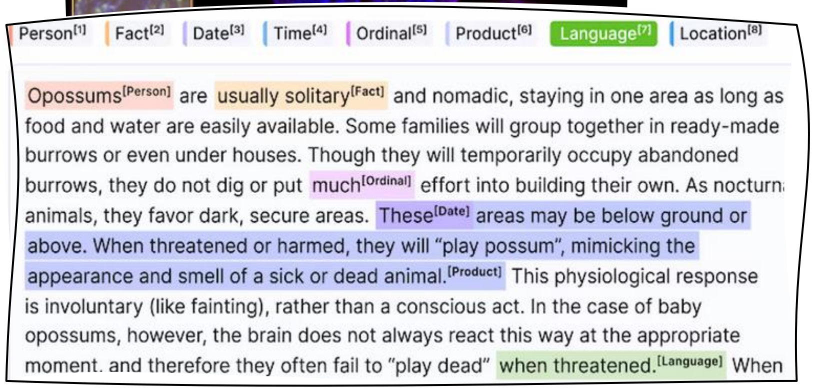
- Annotazioni fondamentali per supervised learning
- Labeling richiede più iterazioni
  - Rifinitura maschere
  - Annotazioni multiple
- In generale, molto dispendioso

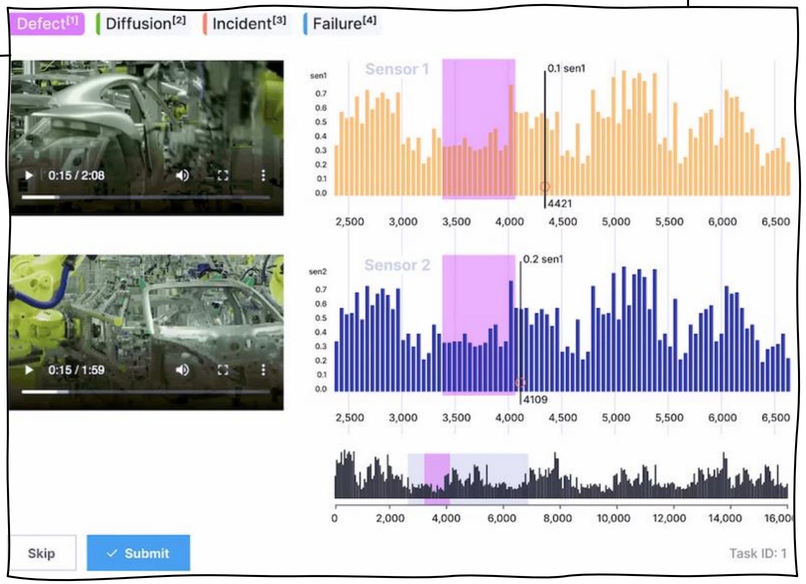
→ Tool per labeling

Punti aperti: labeling

- Esistono molti strumenti commerciali (e.g. Label Studio)







Punti aperti: labeling

- Esistono molti strumenti commerciali (e.g. Label Studio)
- Principali necessità
  - Collaborativo e accessibile
  - Authentication & Authorization
  - Configurabile
- Nice-to-have
  - CI/CD (e.g. ML-assisted labeling, support active/online learning)

End-to-end lifecycle

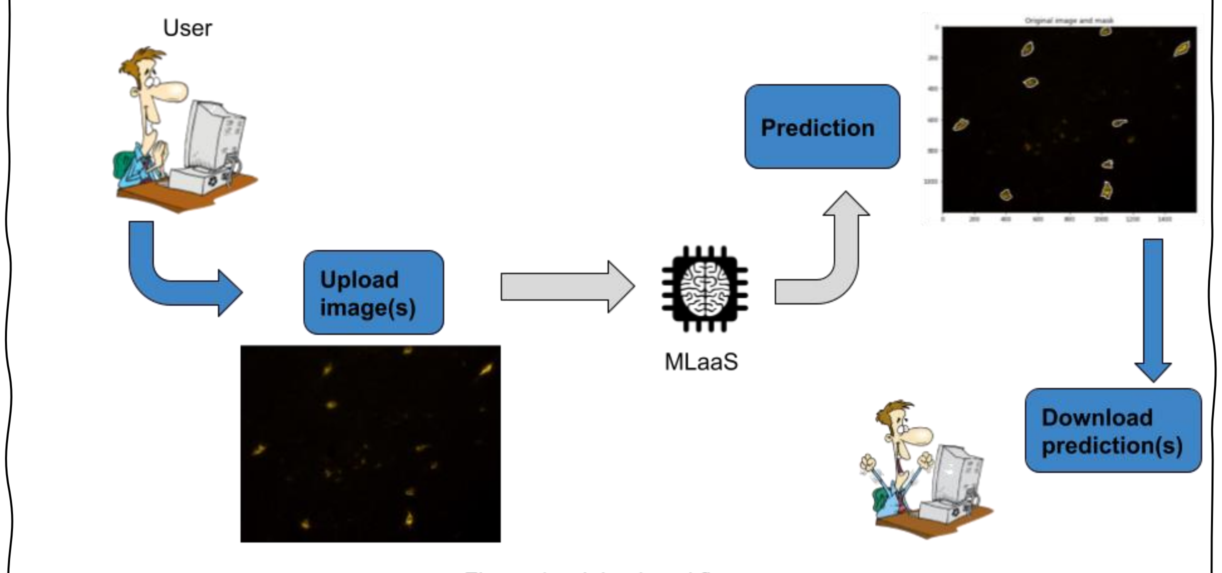


Figura 1: minimal workflow

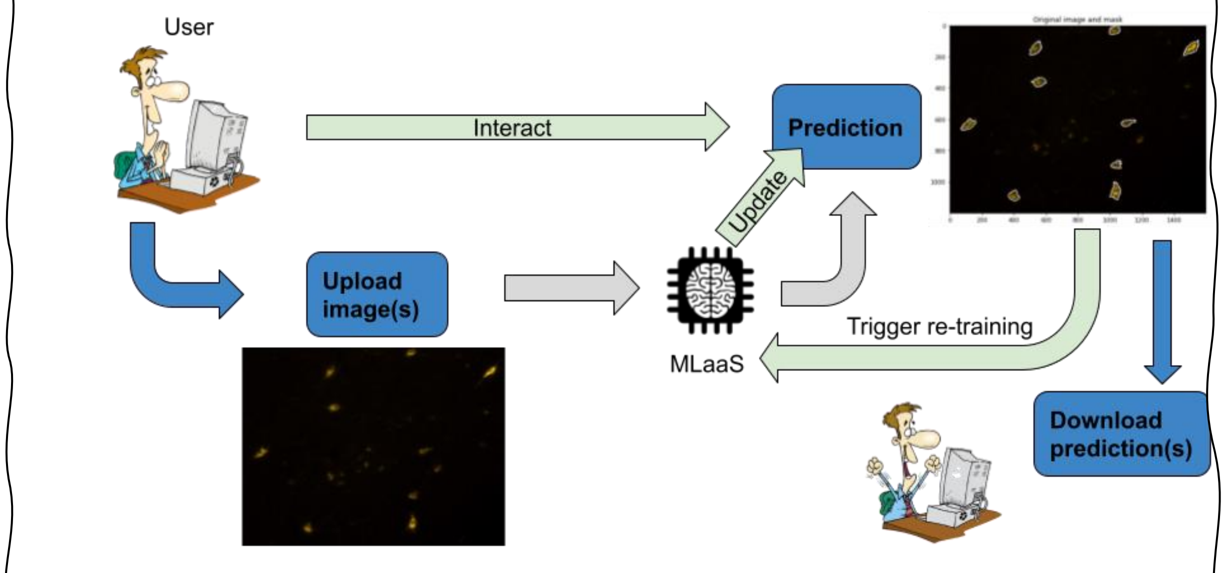


Figura 2: complete workflow

Punti aperti: labeling

- Esistono molti strumenti commerciali (e.g. Label Studio)
- Principali necessità
  - Collaborativo e accessibile
  - Authentication & Authorization
  - Configurabile
- Nice-to-have
  - CI/CD (e.g. ML-assisted labeling, support active/online learning)
- Principali difficoltà
  - Self-hosting (in particolare storage/calcolo)
  - Authentication & Authorization

References

[1] Morelli R., et al., (2021), Automating cell counting in fluorescent microscopy through deep learning with c-ResUnet, *Scientific Reports*

[2] Clissa L., et al., (2021), Fluorescent Neuronal Cells dataset, *AMS Acta*

[3] Clissa L., (2022), Supporting Scientific Research Through Machine and Deep Learning: Fluorescence Microscopy and Operational Intelligence Use Cases, *AMS Dottorato*

[4] Clissa L., et al., (2022), Analyzing WLCG File Transfer Errors Through Machine Learning, *Computing and Software for Big Science*

[5] Di Girolamo A., et al., (2022), Preparing Distributed Computing Operations for the HL-LHC Era With Operational Intelligence, *Frontiers in Big Data*

[6] Di Girolamo A., et al., (2020), for Distributed Computing Systems for Exascale Science, CHEP

[7] Macaluso A., et al., (2020), A Variational Algorithm for Quantum Neural Networks, *International Conference on Computational Science*

[8] Macaluso A., et al., (2020), Quantum Ensemble for Classification, *arXiv*

[9] Macaluso A., et al., (2020), Quantum splines for non-linear approximations, *International Conference on Computing Frontiers*

[10] Hitrec T., et al., (2019), Neural control of fasting-induced torpor in mice, *Scientific Reports*

[11] Clissa L., (2023), Survey of Big Data sizes in 2021, *arXiv*

GitHub: https://github.com/robomorelli/cell_counting_yellow

Backup

Training pipeline

- Crop images in 512x512 partially overlapping crops
- Augmentation
  - rotations, Gaussian noise, brightness variation, elastic transformation
  - Augmentation factors set to:
    - 4 for semi-automatically segmented images
    - 10 for manually segmented images
    - 25 for crops in AO
- Train from scratch

Training settings

- Adam optimizer
- Learning rate (*lr*) starts at 0.06
- Decrease *lr* by 30% if no improvement in 4 epochs
- Weighted Binary Cross Entropy loss (1.5 weight for cells, 1 for background)
- Train until no improvement in 20 epochs
- TensorFlow implementation through Keras
- 4 NVIDIA V100 GPUs hosted at CNAF

Post-processing

- Output grezzo: heatmap
- Thresholding
  - Introduce iper-parametro
  - Ottimizzazione threshold
- Pulizia output

Metriche performance

- True positives (TP), False Positives (FP) e False Negatives (FN)

```
for center in predicted_centers:
        compute distance (px) from all true_centers
        if minimum_distance < 50:
                increase TP count
                remove corresponding true center
FN = n_true – TP
FP = n_pred - TP
```

- Detection: F_1 score, AUC
- Conteggio: R^2, MAE

Metriche detection: F_1 score e AUC

- F_1 score come metrica principale:

$$accuracy = \frac{TP}{TP + FP + FN},$$

$$precision = \frac{TP}{TP + FP},$$

$$recall = \frac{TP}{TP + FN},$$

$$F_1 = \frac{2 \cdot precision \cdot recall}{precision + recall},$$

- Area Under precision/recall Curve (AUC)

Counting metrics: MAE and R^2

Errore assoluto (AE) e percentuale (PE):

$$AE = \left| n_{true} - n_{pred} \right|,$$

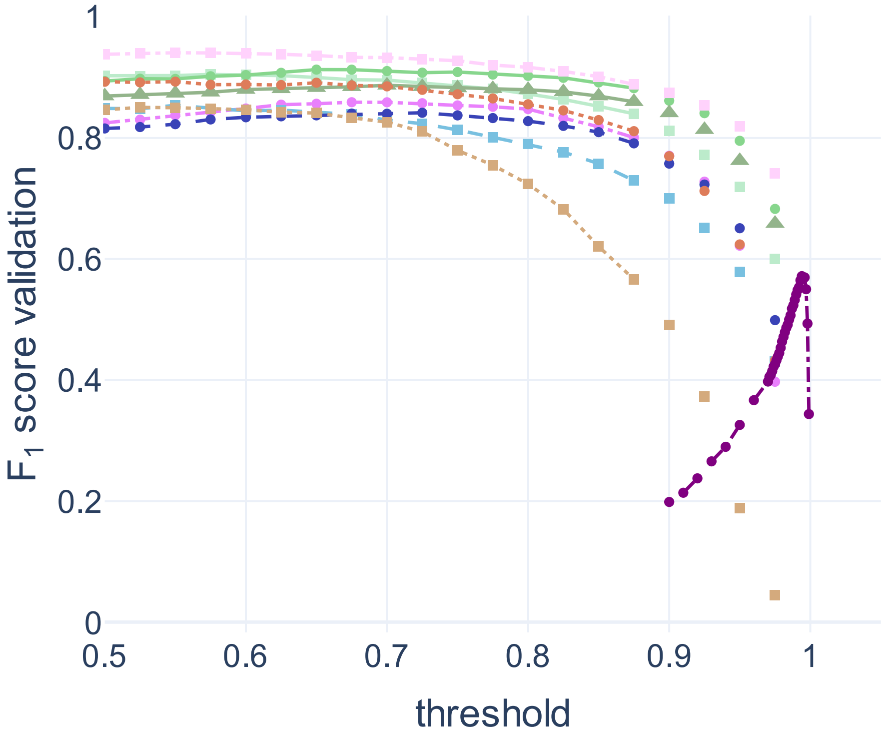
$$PE = \frac{n_{true} - n_{pred}}{n_{true}}$$

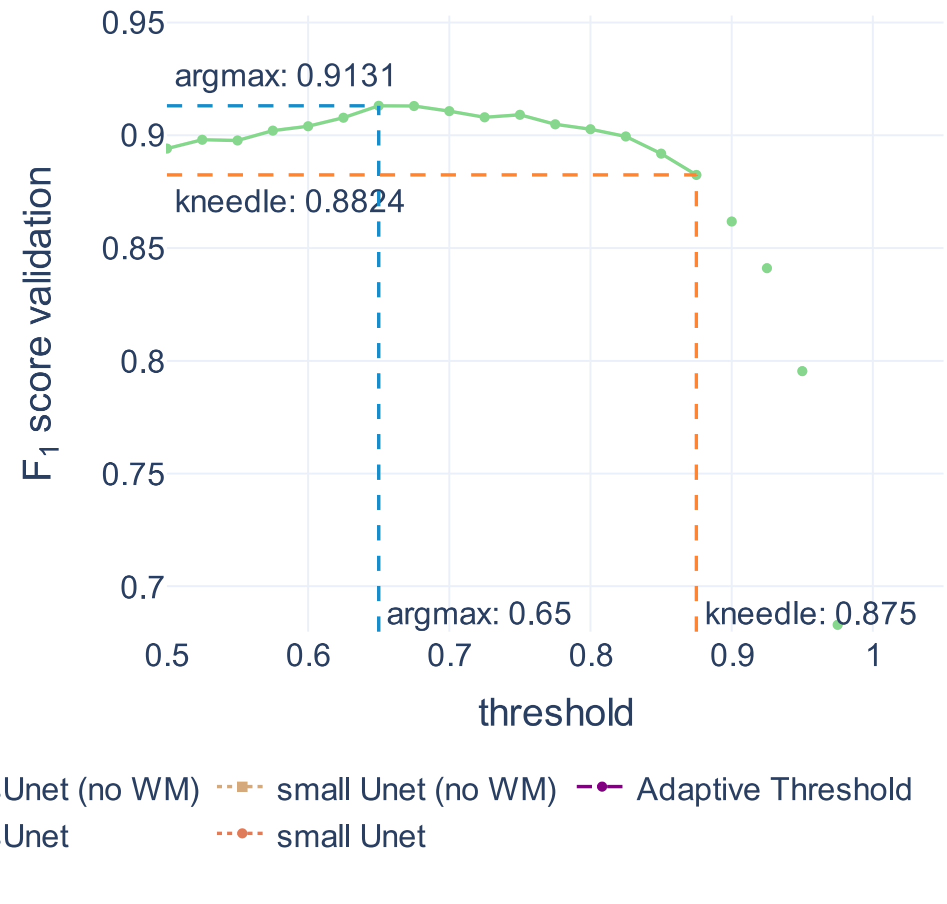
Coefficiente di determinazione, R^2:

$$R^2 = 1 - \frac{\sum_i \left(n_{true}^i - n_{pred}^i\right)^2}{\sum_i \left(n_{true}^i - \bar{n}\right)^2},$$

Dove n_{true}^i, n_{pred}^i sono il numero di cellule vere e predette per l'immagine i,
$\bar{n}$ è il numero medio di cellule per immagine

Threshold optimization





ROC curve

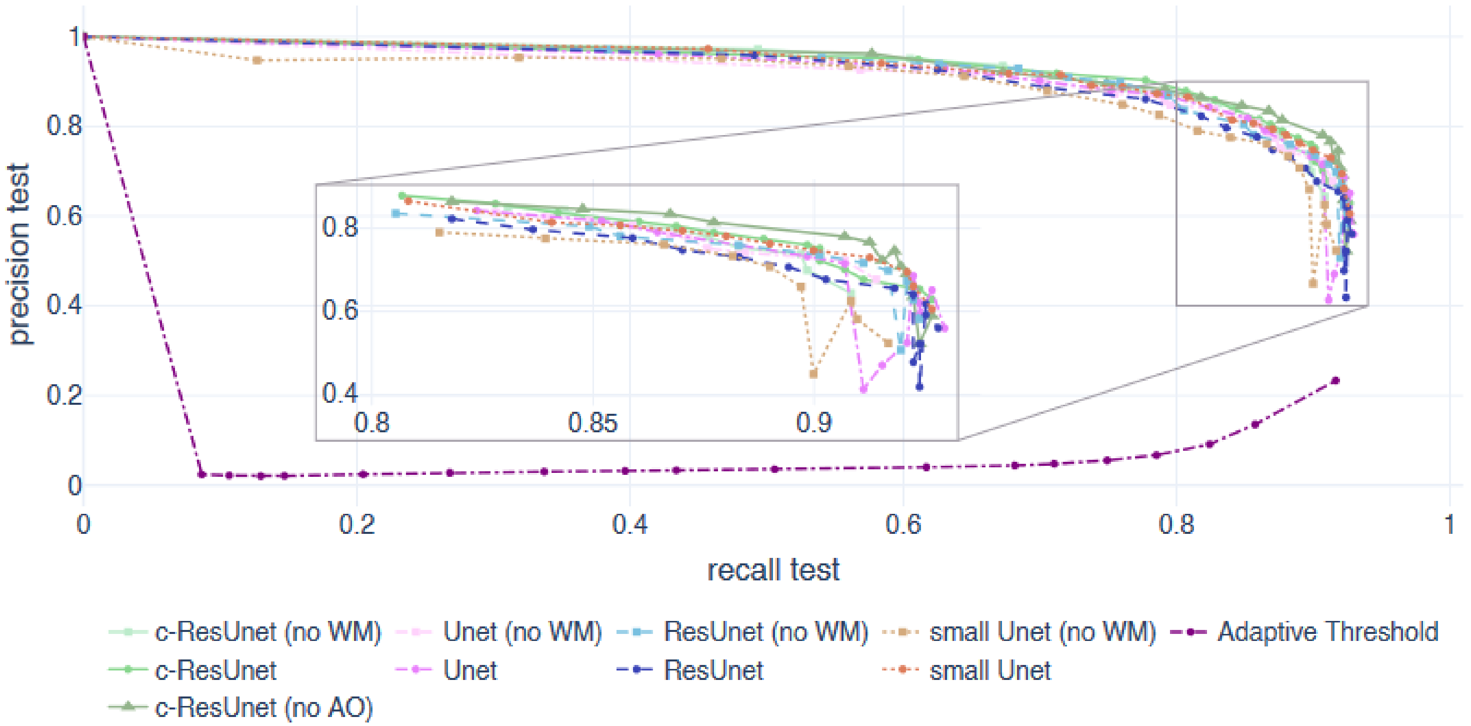


Figure 5.1: **Precision/Recall plot.** Test set precision/recall curves varying the threshold for predicted heatmap binarization. The inset plot reports a zoom of the top right corner to highlight differences of the various curves.

Quantitative evaluation

Model	Threshold	F_1	AUC	Accuracy	Precision	Recall	R^2	MAE	MedAE	MPE (%)
c-ResUnet	**0.875**	**0.8149**	**0.8705**	[illegible]	[illegible]	[illegible]	**0.8215**	**3.0857**	[illegible]	[illegible]
c-ResUnet (no AO)	0.875	0.8047	0.8741	[illegible]	[illegible]	[illegible]	0.8077	3.0857	[illegible]	[illegible]
c-ResUnet (no WM)	0.875	0.7613	0.8594	[illegible]	[illegible]	[illegible]	0.7048	3.6857	[illegible]	[illegible]
ResUnet	0.850	0.7855	0.8579	[illegible]	[illegible]	[illegible]	0.7831	3.3286	[illegible]	[illegible]
ResUnet (no WM)	0.850	0.7513	0.8643	[illegible]	[illegible]	[illegible]	0.6955	4.0571	[illegible]	[illegible]
Unet	0.875	0.7724	0.8609	[illegible]	[illegible]	[illegible]	0.7560	3.5143	[illegible]	[illegible]
Unet (no WM)	0.850	0.7886	0.8461	[illegible]	[illegible]	[illegible]	0.8069	3.1571	[illegible]	[illegible]
small Unet	0.875	0.7563	0.8691	[illegible]	[illegible]	[illegible]	0.7682	3.5714	[illegible]	[illegible]
small Unet (no WM)	0.825	0.6697	0.8326	[illegible]	[illegible]	[illegible]	0.5723	4.7714	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Table 5.1: **Performance metrics.** Test set performance using the optimal *kneed* threshold. The first five columns report the detection metrics, while the latter ...es evaluate counting performance.