Optimizing Deep Learning Models for Cell Recognition in Fluorescence Microscopy: the Impact of Loss Functions on Performance and Generalization

Luca Clissa, Antonio Macaluso, Antonio Zoccoli

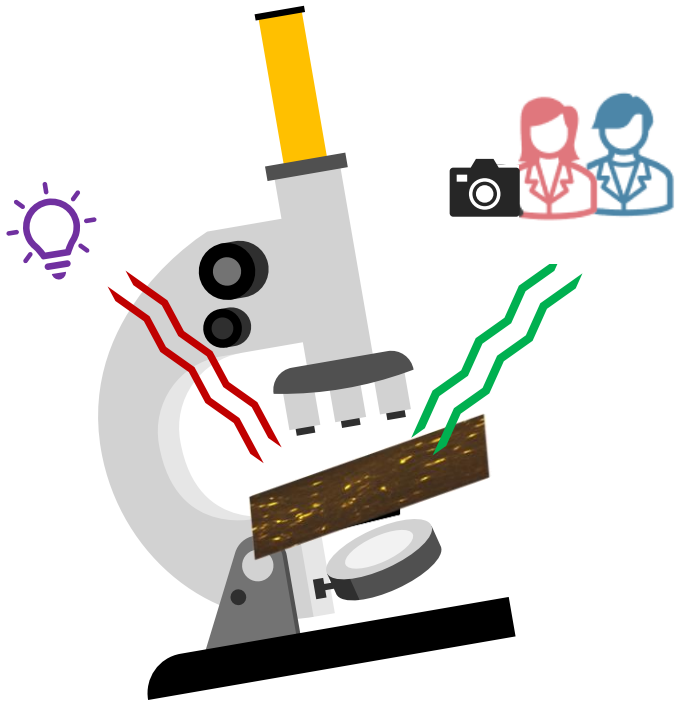
luca.clissa2@unibo.it

Outline

- Fluorescence Microscopy
- Challenges
- Experimental setup
- Results
- Conclusions

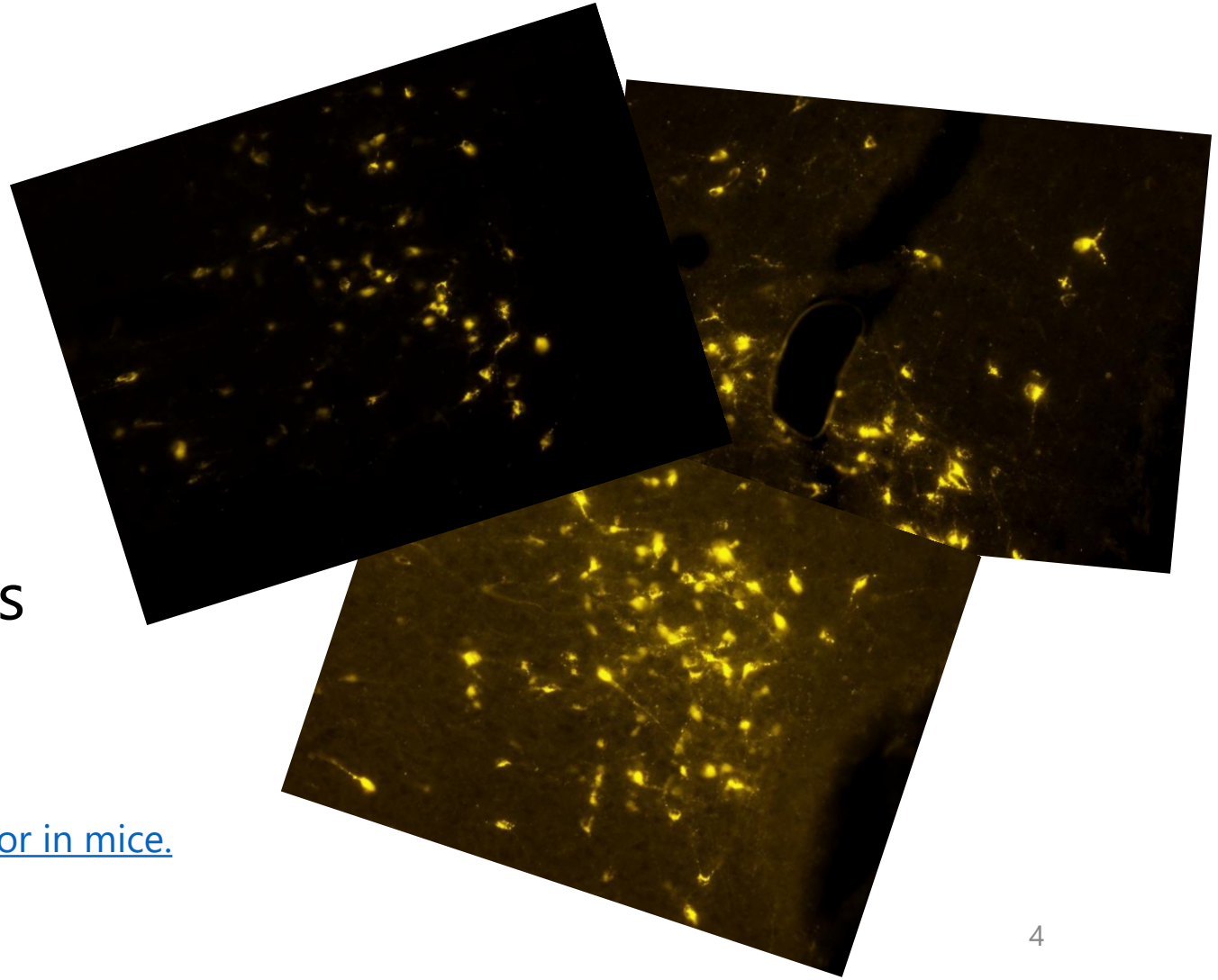
Fluorescence microscopy

- Physics-based imaging technique
- Exploits light absorption/emission properties
- Used to mark/tag/stain biological compounds



Applications

- Very popular in life science
  - Torpor onset [1]
- Cytoplasmatic neuronal structures
- Variability in shape, size and color hue
- **Goal:** count stained structures



[1] Hitrec, T., et al.: Neural control of fasting-induced torpor in mice. Scientific Reports 9(1) (oct 2019).

Problem

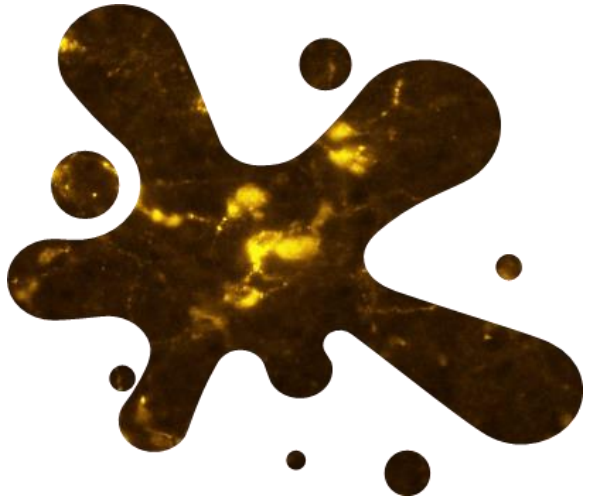
- Manual processing
  - Time-consuming
  - Error-prone
  - Subjectivity of borderline cases
- Hard to adapt Deep Learning solutions
  - Domain shift
  - Few in-domain annotated datasets
  - How to train? How to evaluate?

Contributions

- Semantic segmentation using c-ResUnet [2] and Fluorescent Neuronal Cells v2 dataset (FNC v2) [3]
- Show the impact of loss functions on model performance
  - 60 ablation studies
  - 6 loss functions
- Inspect pros and cons of several evaluation metrics
- Discuss characteristics affecting out-of-sample generalization

[2] Morelli, R., et al.: Automating cell counting in fluorescent microscopy through deep learning with c-ResUnet. Scientific Reports 11(1), 22920 (2021).
[3] Clissa, L., et al.: Fluorescent neuronal cells v2: Multi-task, multi-format annotations for deep learning in microscopy. arXiv preprint (submitted to Scientific Data) (2023)

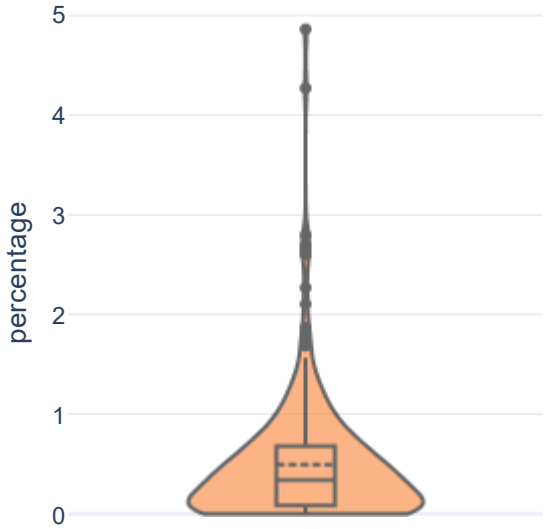
Challenges

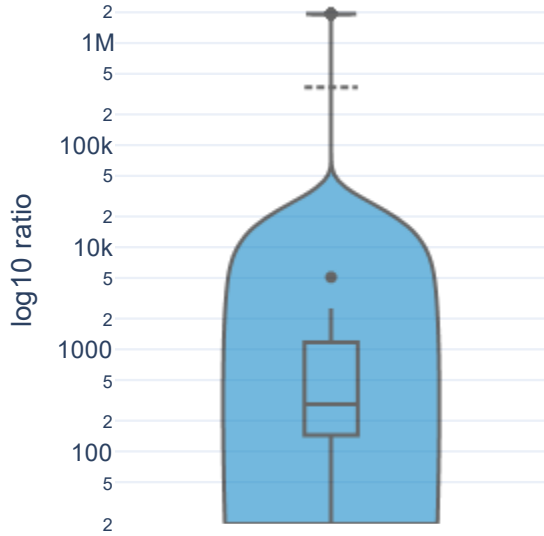


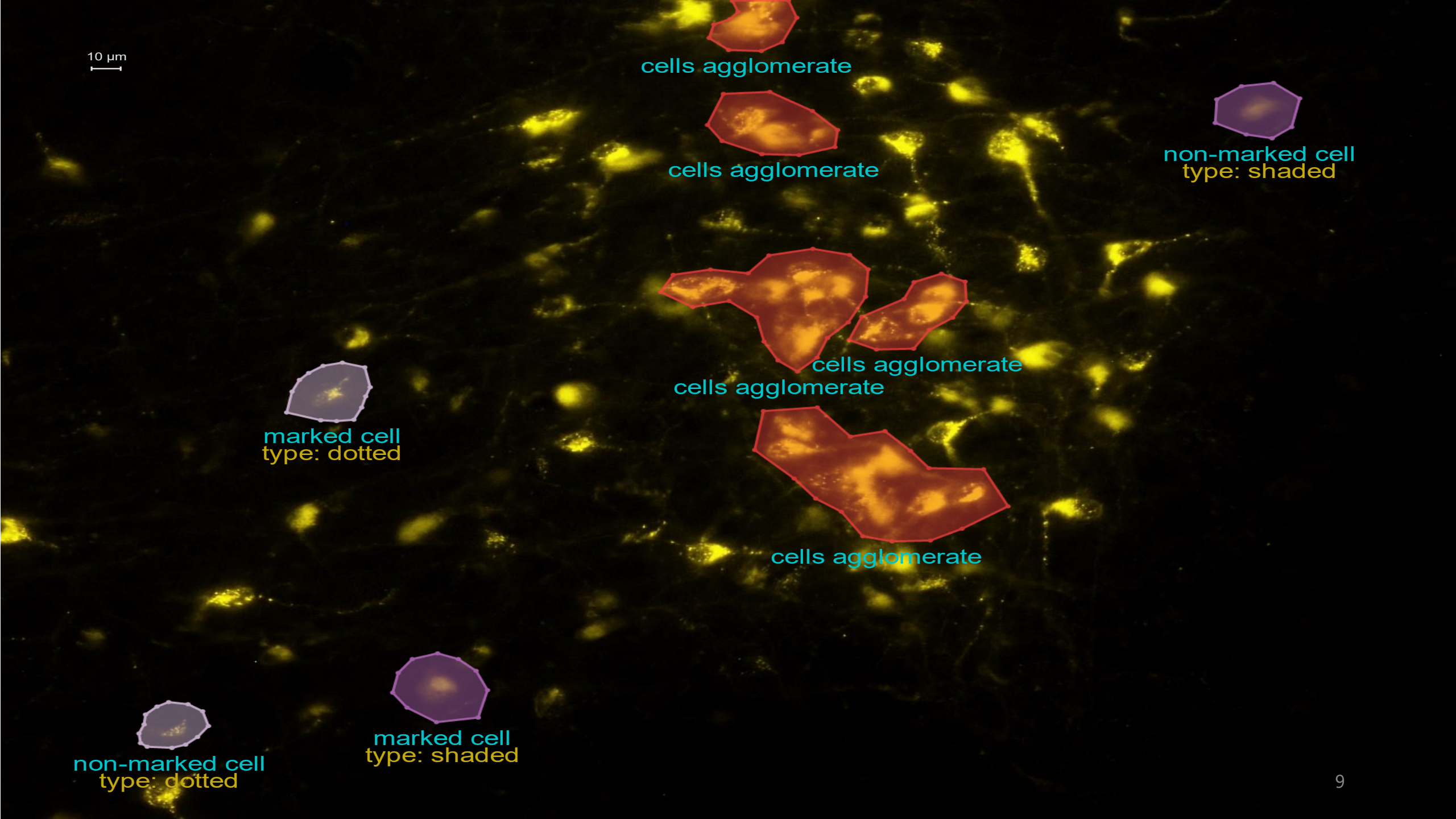
- Class imbalance
- Overcrowding
- Noisy labels
- Artefacts

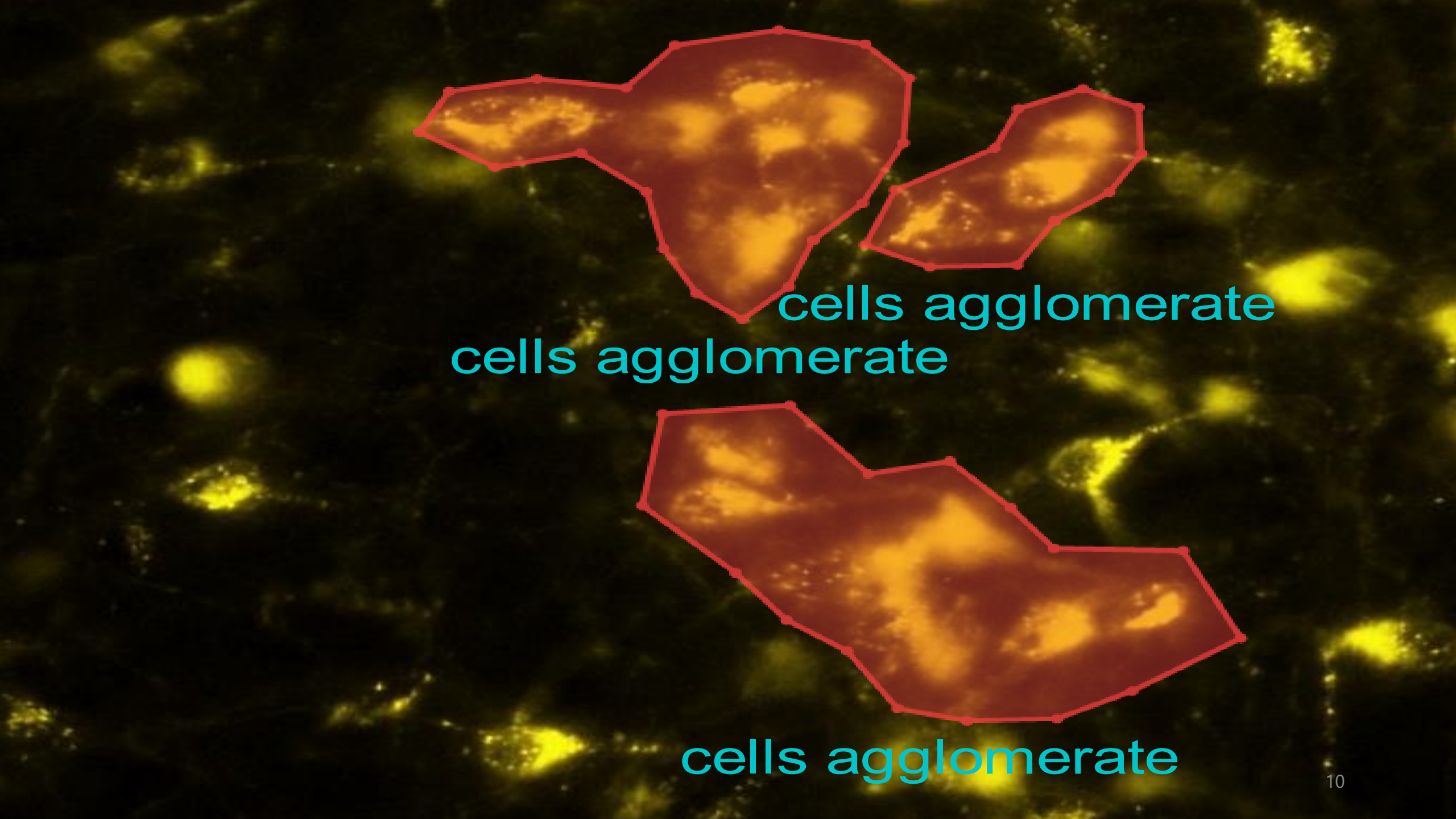
Class imbalance

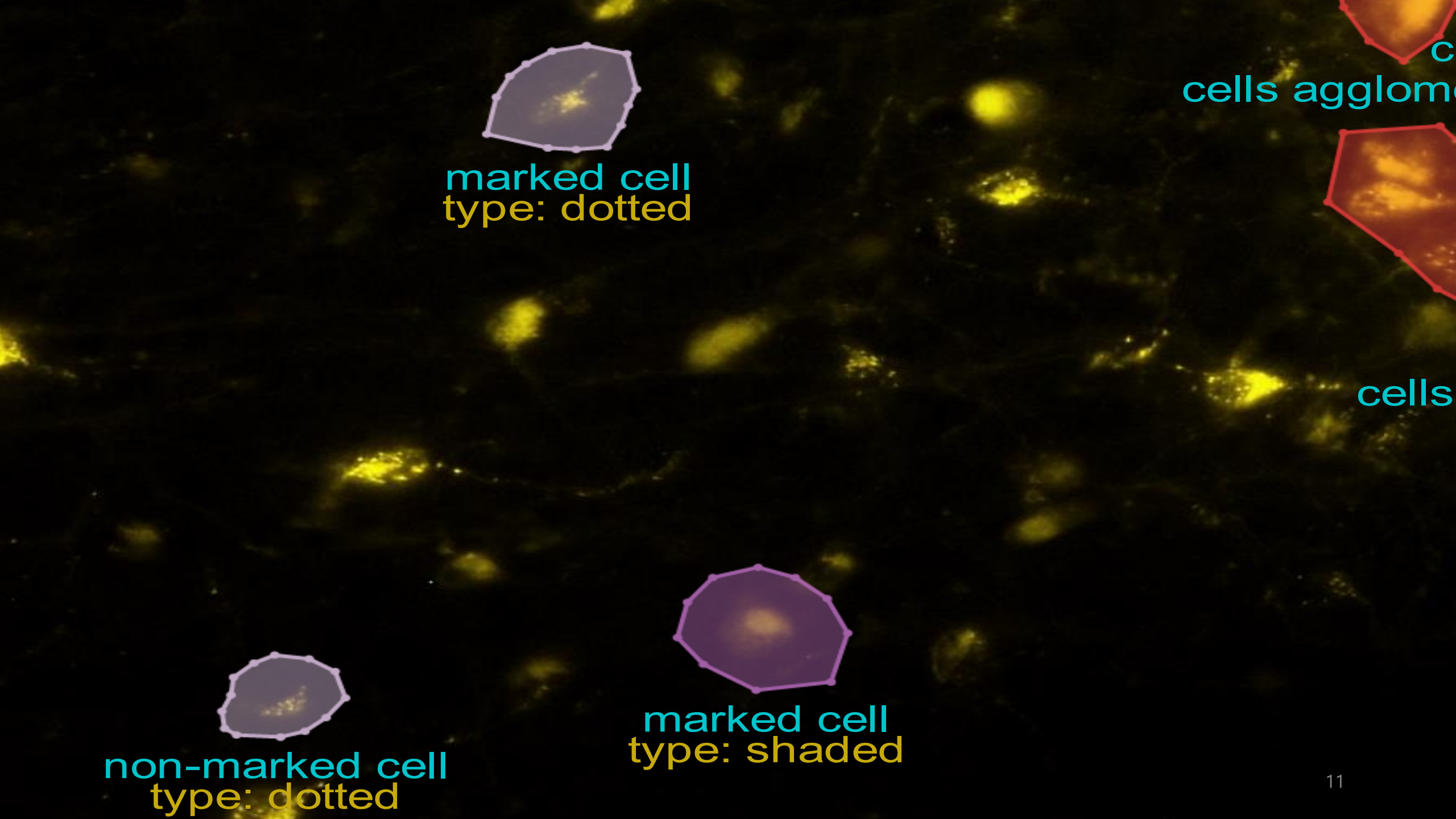
quantile	signal (%)	signal ratio
mean	0.50	367k
s.d.	0.61	756k
min	0	19.57
10%	0	92.39
25%	0.09	145.35
50%	0.34	291.10
75%	0.68	1k
90%	1.07	1.9M
max	4.86	1.9M

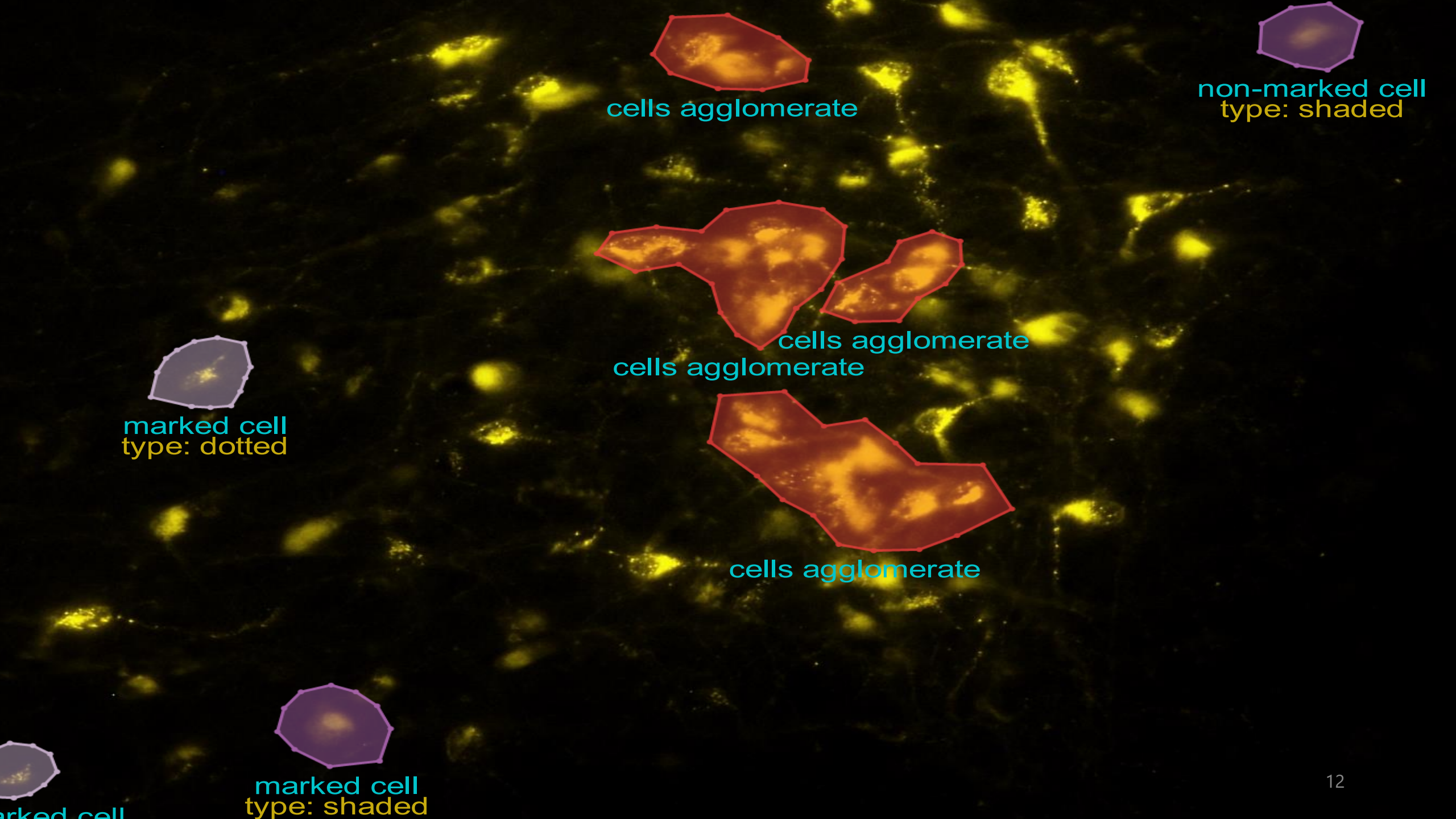


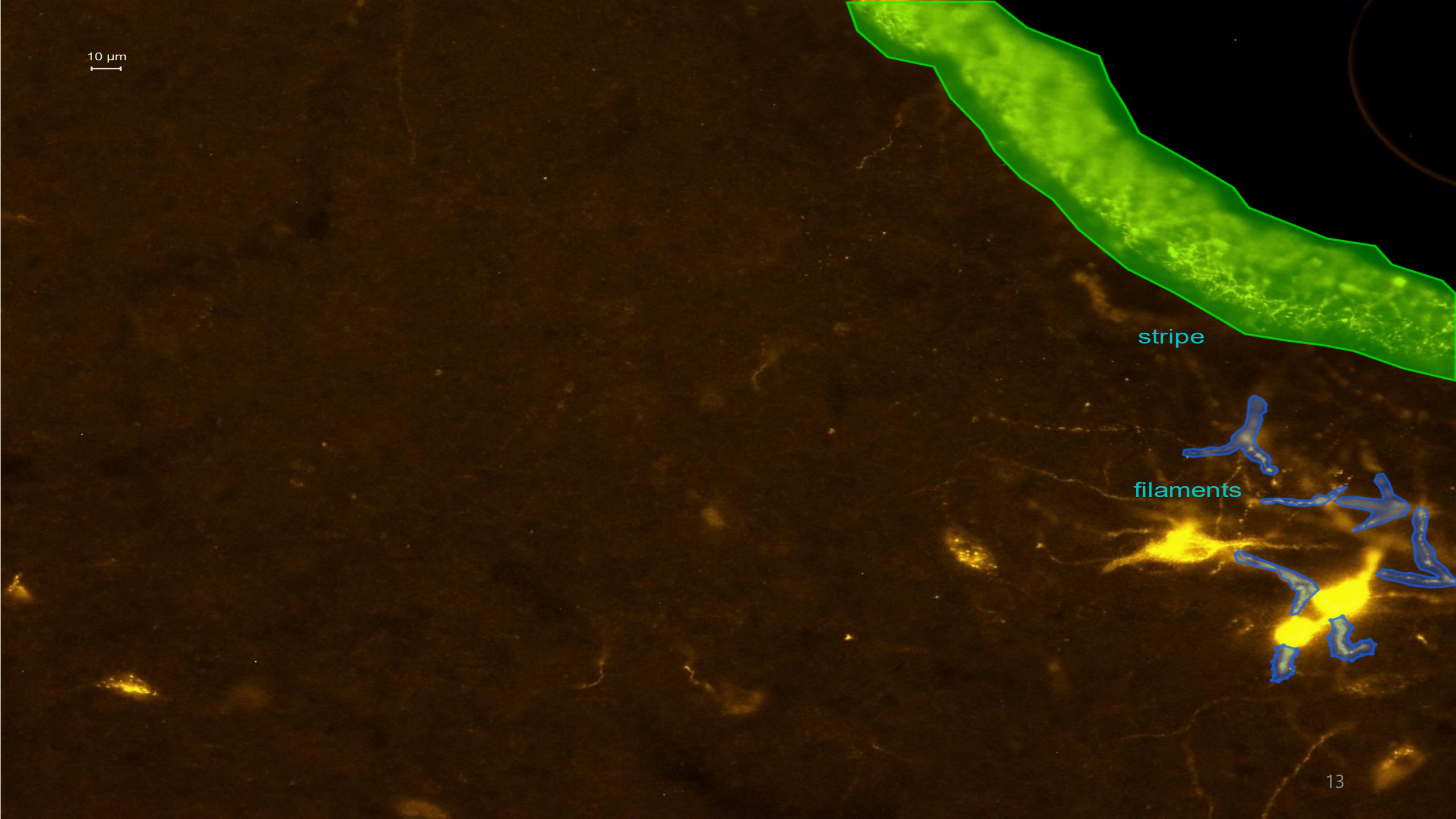



10 µm
cells agglomerate
cells agglomerate
non-marked cell
type: shaded
cells agglomerate
cells agglomerate
marked cell
type: dotted
cells agglomerate
marked cell
type: shaded
non-marked cell
type: dotted


cells agglomerate
cells agglomerate
cells agglomerate


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

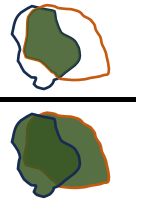
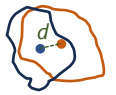

cells agglomerate
non-marked cell
type: shaded
cells agglomerate
cells agglomerate
marked cell
type: dotted
cells agglomerate
marked cell
type: shaded


10 µm
stripe
filaments

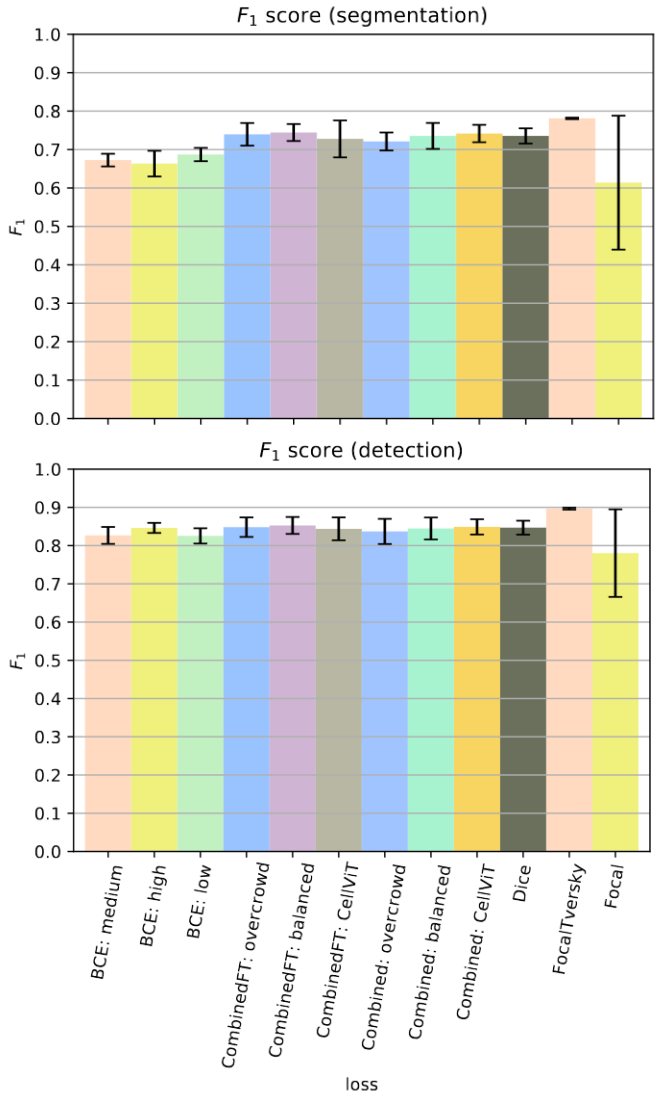
Ablation studies

- 4 alternative losses
  - Weighted Binary Cross Entropy (BCE): $w_{cell} = 50, 100, 200; w_{bkgd} = 1$
  - Dice Loss
  - Focal Loss
  - Focal Tversky Loss
- 2 combined losses
  - CombinedLoss $= \lambda_1 BCE + \lambda_2$Dice + λ_3Focal
  - CombinedFTLoss $= \lambda_1 BCE + \lambda_2$Dice + λ_3Focal Tversky
    - Balanced: $\lambda_1 = 0.3, \lambda_2 = 0.3, \lambda_3 = 0.4$
    - Overcrowd: $\lambda_1 = 0.2, \lambda_2 = 0.5, \lambda_3 = 0.3$
    - CellViT: $\lambda_1 = 0.5, \lambda_2 = 0.3, \lambda_3 = 0.5$

Evaluation Metrics

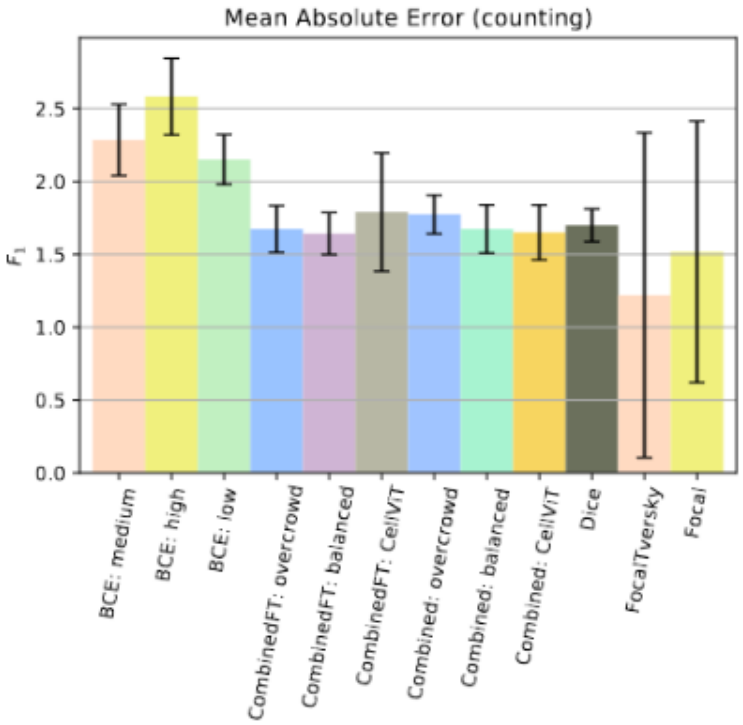
- Segmentation
  - Mean Intersection over Union (mIoU) = $\frac{Intersection}{Union}$ 
  - threshold: 0.4
- Detection
  - Centers distance
  - threshold: 40 pixels (mean cell diameter) 
- Counting
  - Mean Absolute Error
  - Median Absolute Error
  - Mean Percentage Error: $\frac{(n_t - n_p)}{\max(n_t, 1)} * 100$

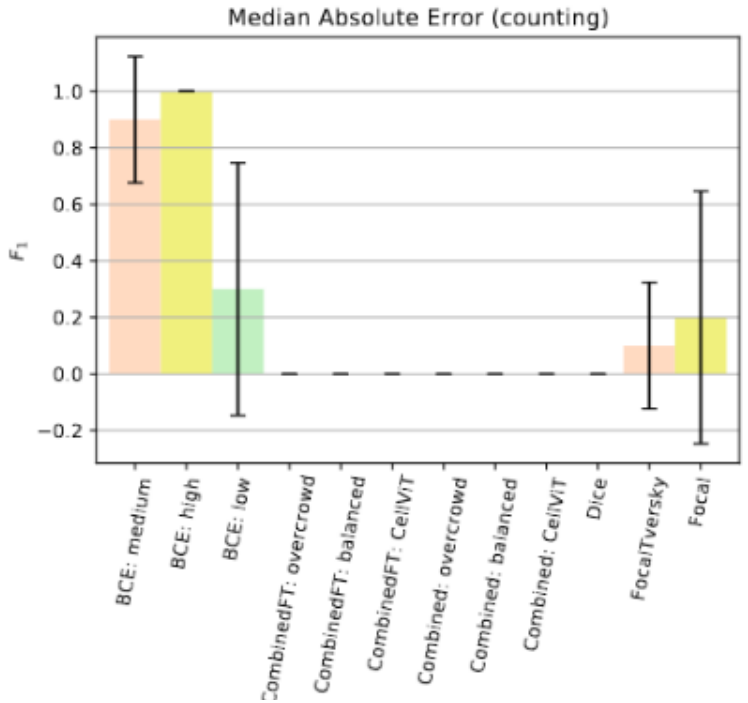
Segmentation & Detection

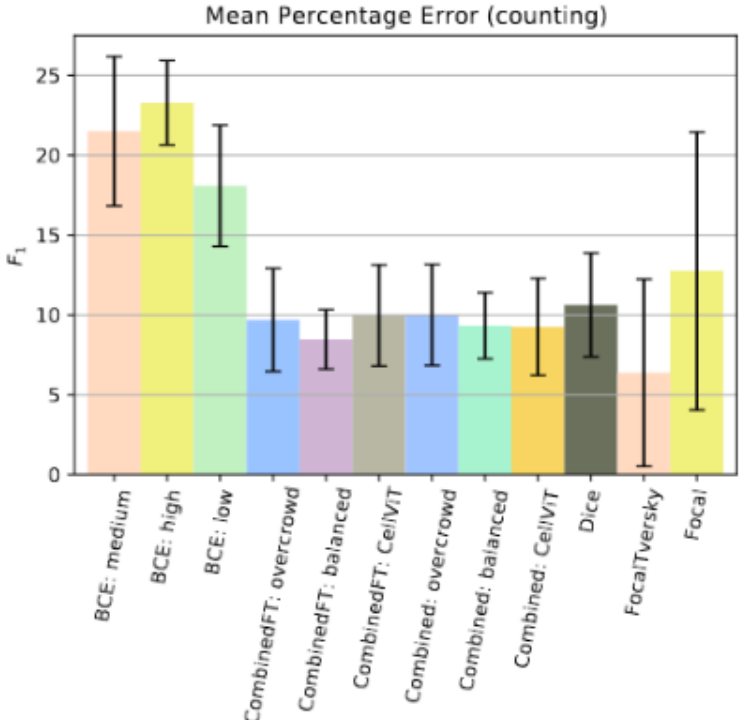


Loss	F1 score (IoU)	F1 score (distance)
BCE: medium	0.673±0.017	0.827±0.022
BCE: high	0.663±0.033	0.846±0.013
BCE: low	0.687±0.017	0.825±0.020
CombinedFT: overcrowd	0.740±0.029	0.848±0.026
CombinedFT: balanced	0.744±0.022	0.853±0.022
CombinedFT: CellViT	0.728±0.048	0.844±0.030
Combined: overcrowd	0.721±0.023	0.837±0.033
Combined: balanced	0.735±0.034	0.845±0.029
Combined: CellViT	0.742±0.023	0.849±0.020
Dice	0.735±0.020	0.847±0.018
Focal Tversky	**0.781±0.002**	**0.897±0.003**
Focal	0.614±0.027	0.780±0.034

Counting



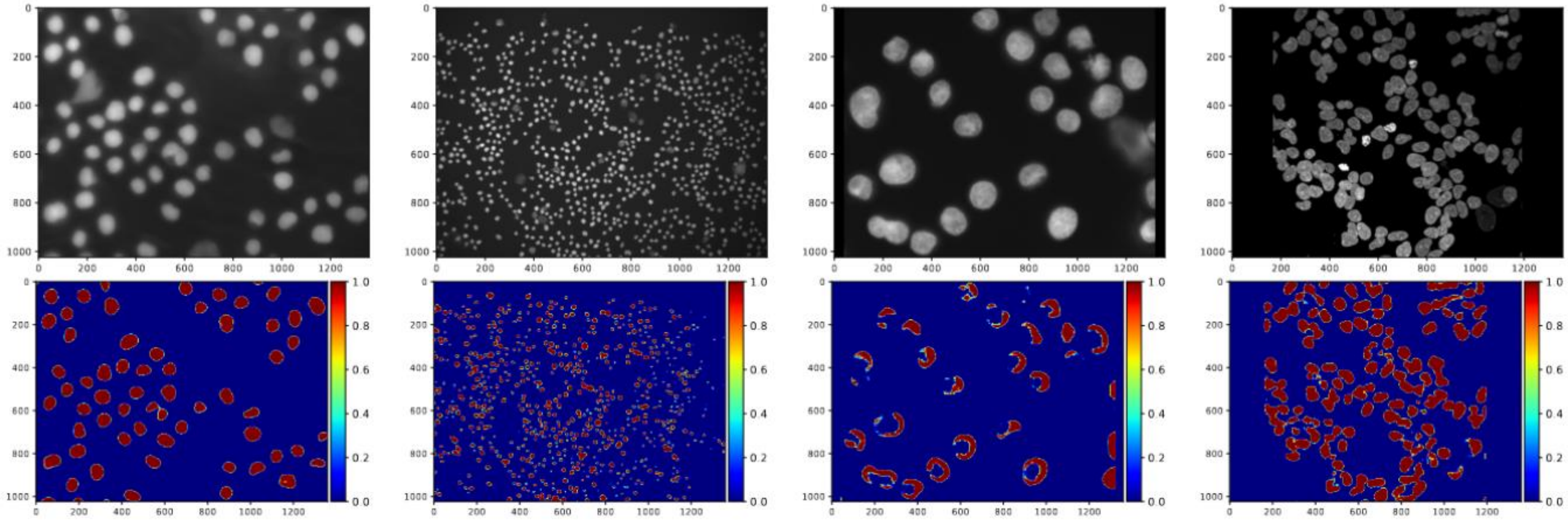




	BCE: medium	*BCE: high*	*BCE: low*	*CombinedFT: overcrowd*	*CombinedFT: balanced*	*CombinedFT: CellViT*	*Combined: overcrowd*	*Combined: balanced*	*Combined: CellViT*	*Dice*	*Focal Tversky*	*Focal*
MAE	2.286±0.245	2.583±0.263	2.151±0.171	1.674±0.161	1.643±0.144	1.791±0.406	1.774±0.132	1.674±0.166	1.651±0.188	1.700±0.113	1.220±1.115	1.517±0.365
MedAE	0.9±0.224	1±0	0.3±0.447	0±0	0±0	0±0	0±0	0±0	0±0	0±0	0.1±0.224	0.2±0.4
MPE	21.498±4.679	23.281±2.637	18.079±3.791	9.682±3.221	8.465±1.859	9.968±3.163	10±3.160	9.322±2.074	9.249±3.039	10.621±3.247	**6.373±1.607**	12.747±3.525

Out-of-sample generalization

[4] S-BSST265 dataset



(a) Medium-sized, sharp objects

(b) Small objects

(c) Uneven texture and filling

(d) Overcrodwing

Conclusions

- Focal Tversky loss overperforms other losses
- Still some troubles separating crowded objects
  → careful post-processing needed (hole filling, small objects, watershed)
- Combined losses are competitive
  - Better tuning of lambda weights
- Generalization
  - High variability
  - Dedicated augmentation may help
  - Panoptic loss
- Integrating more metrics enables more comprehensive assessment

Thanks for your attention!

Questions?

References

[1] Hitrec, T., et al.: Neural control of fasting-induced torpor in mice. Scientific Reports 9(1) (oct 2019).

[2] Morelli, R., et al.: Automating cell counting in fluorescent microscopy through deep learning with c-ResUnet. Scientific Reports 11(1), 22920 (2021).

[3] Clissa, L., et al.: Fluorescent neuronal cells v2: Multi-task, multi-format annotations for deep learning in microscopy. arXiv preprint (submitted to Scientific Data) (2023)

[4] Kromp, F., et al.: An annotated fluorescence image dataset for training nuclear segmentation methods. Scientific Data 7(1), 262 (2020)

[5] Jadon, S.: A survey of loss functions for semantic segmentation. In: 2020 IEEE conference on computational intelligence in bioinformatics and computational biology (CIBCB). pp. 1–7 (2020)

Backup

Ablation studies configurations

	BCE	Dice	Focal	Focal Tversky	Combined	Combined FT
Hyperparameters	w_cell	smooth	gamma	gamma	λ_1, λ_2, λ_3	λ_1, λ_2, λ_3
Values	[50, 100, 200]	1×10^{-6}	2	2	balanced: [0.3, 0.3, 0.4] overcrowd: [0.2, 0.5, 0.3] CellViT: [0.5, 0.3, 0.5]	balanced: [0.3, 0.3, 0.4] overcrowd: [0.2, 0.5, 0.3] CellViT: [0.5, 0.3, 0.5]

Training setup

- Adam optimizer
- Learning rate test for initial LR
- 200 epochs
- Cyclical lerarning rates
- Best model on validation dice coefficient

Loss functions

- Weighted Binary Cross Entropy: higher weight to underrepresented class

$$L_{W-BCE}(y,\hat{y}) = -(\beta * ylog(\hat{y}) + (1-y)log(1-\hat{y}))$$

- Dice Loss: targets segmentation performance directly, low impact of small objects

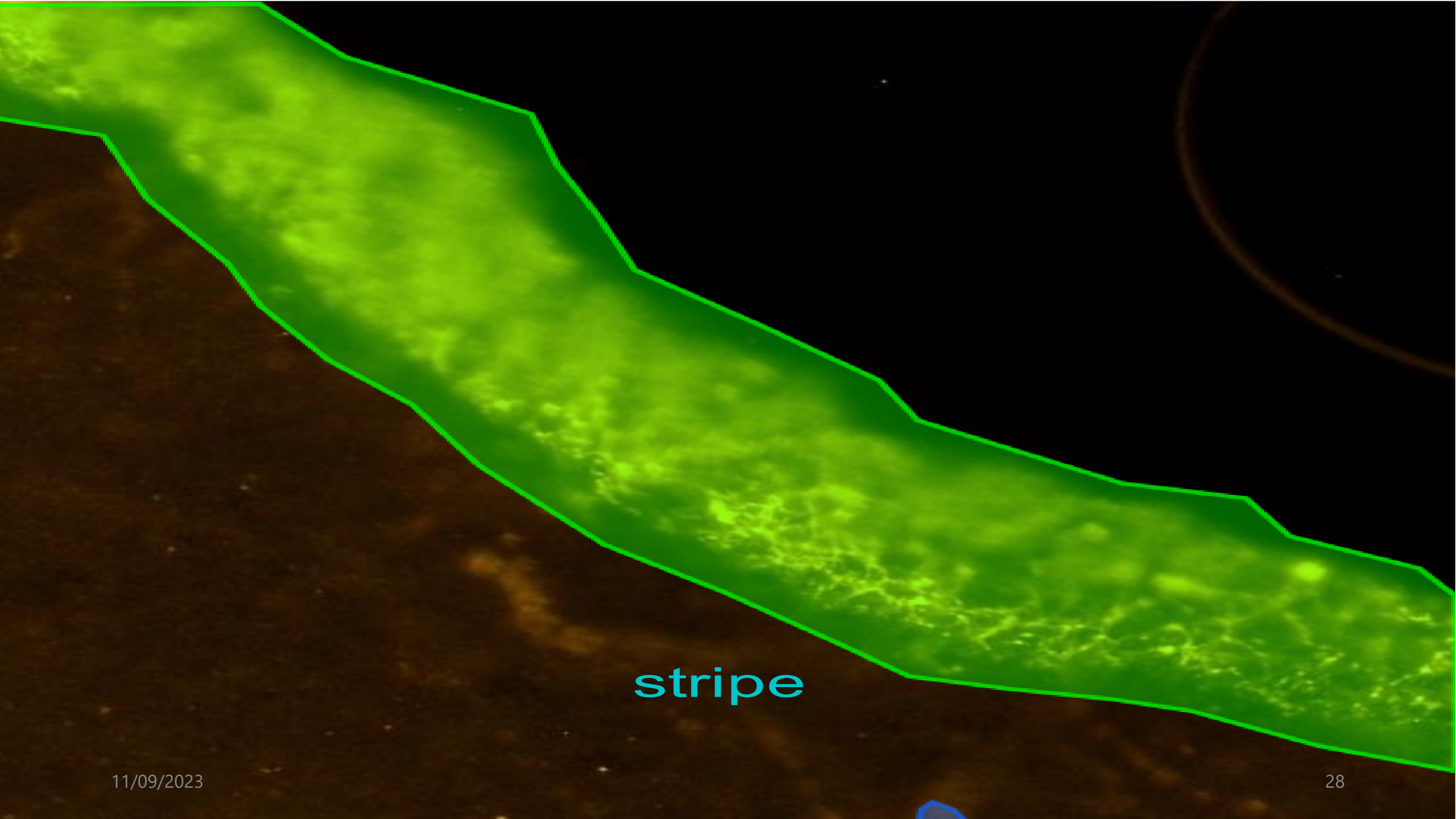
$$DL(y,\hat{p}) = 1 - \frac{2y\hat{p}+1}{y+\hat{p}+1}$$

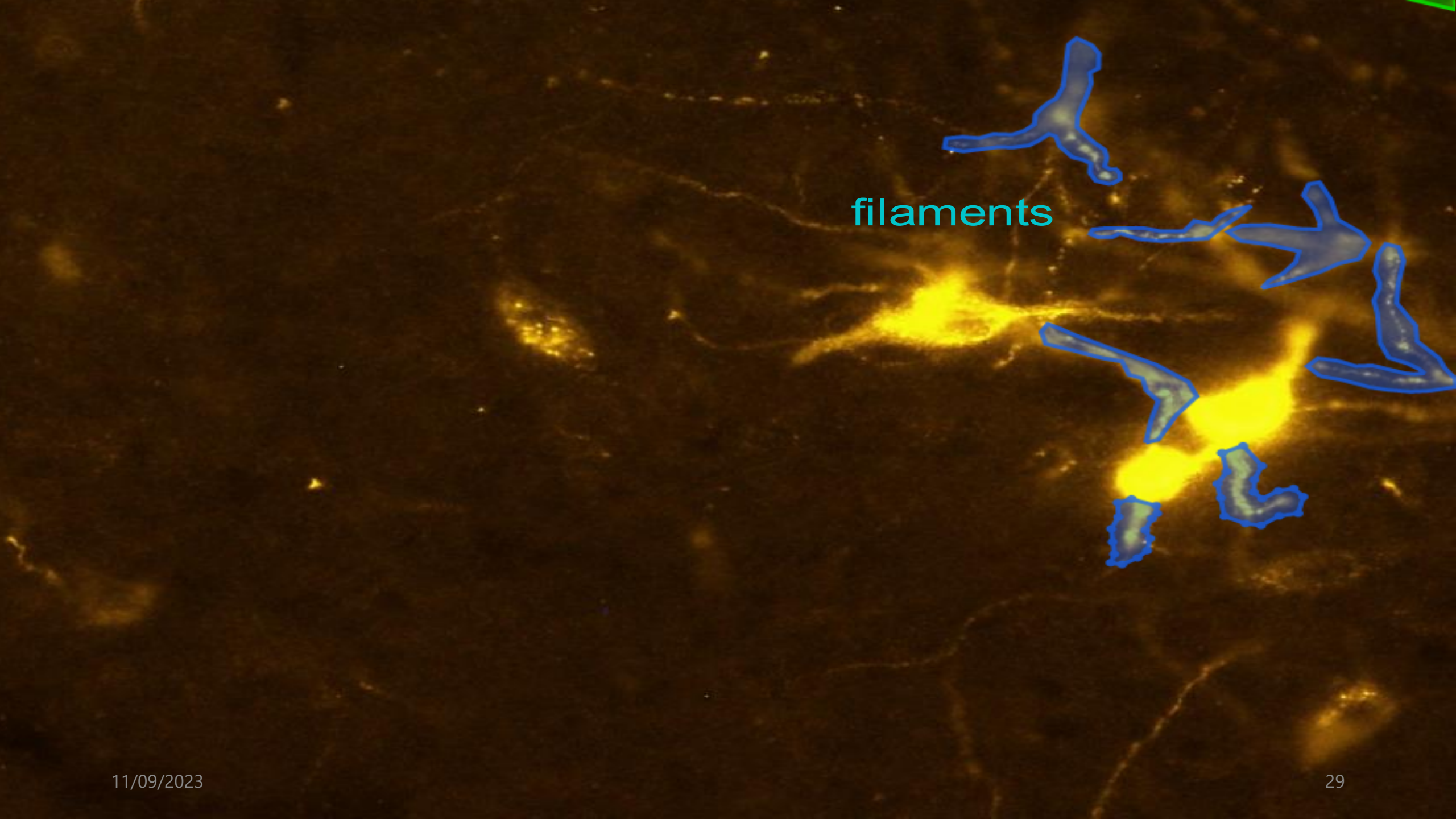
- Focal Loss: oversample hard examples

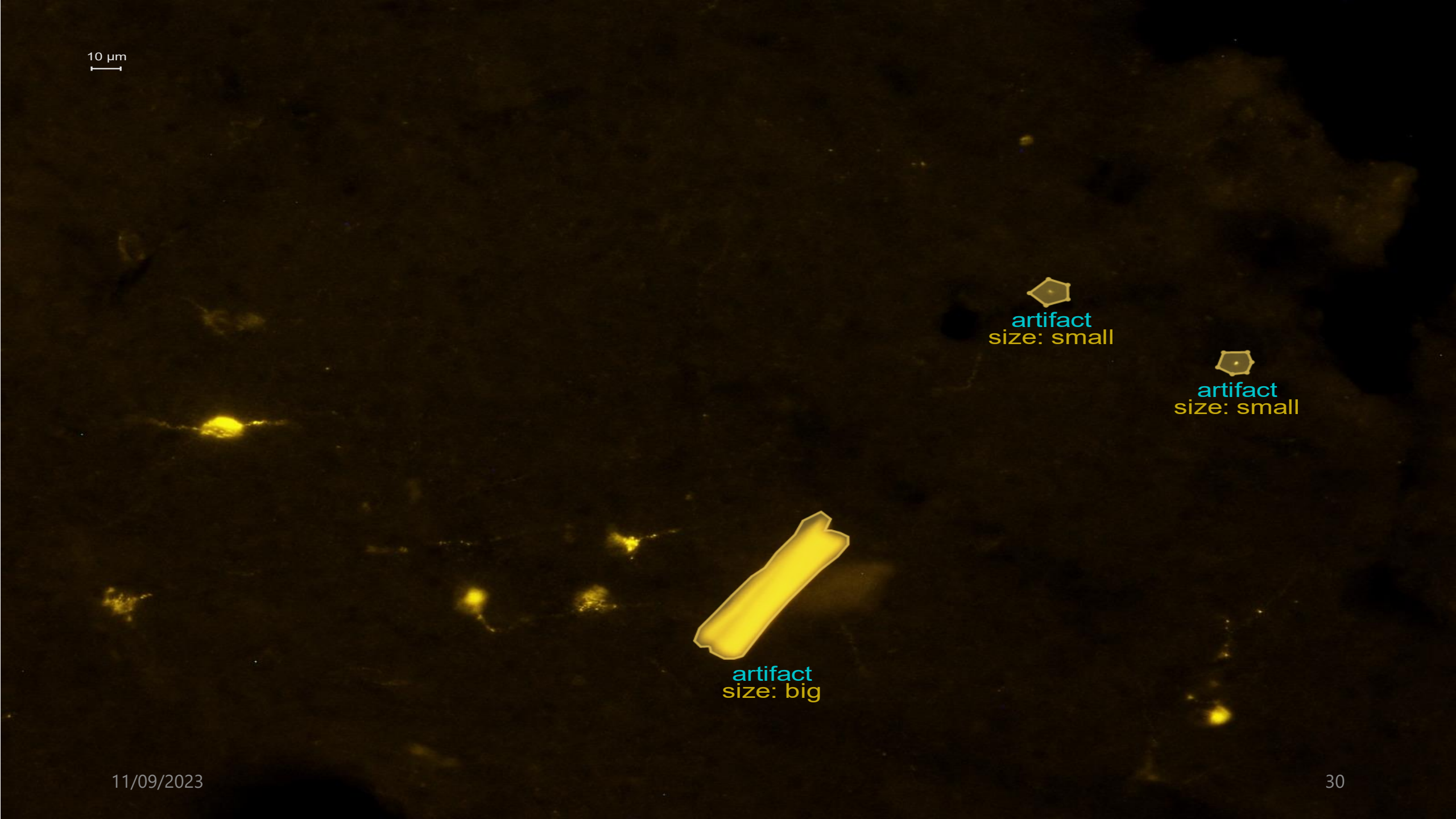
$$FL(p_t) = -\alpha_t(1-p_t)^{\gamma}log(p_t), \qquad p_t = \begin{cases} p, & \text{if } y = 1 \\ 1-p, & \text{otherwise} \end{cases}$$

- Focal Tversky Loss: bring together advantages of Dice and Focal losses

$$FTL = \sum_c (1-TI_c)^{\gamma}, \qquad TI(p,\hat{p}) = \frac{p\hat{p}}{p\hat{p} + \beta(1-p)\hat{p} + (1-\beta)p(1-\hat{p})}$$


stripe


filaments


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