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

Luca Clissa[1,2], Antonio Macaluso[3], Antonio Zoccoli[1,2]

luca.clissa2@unibo.it

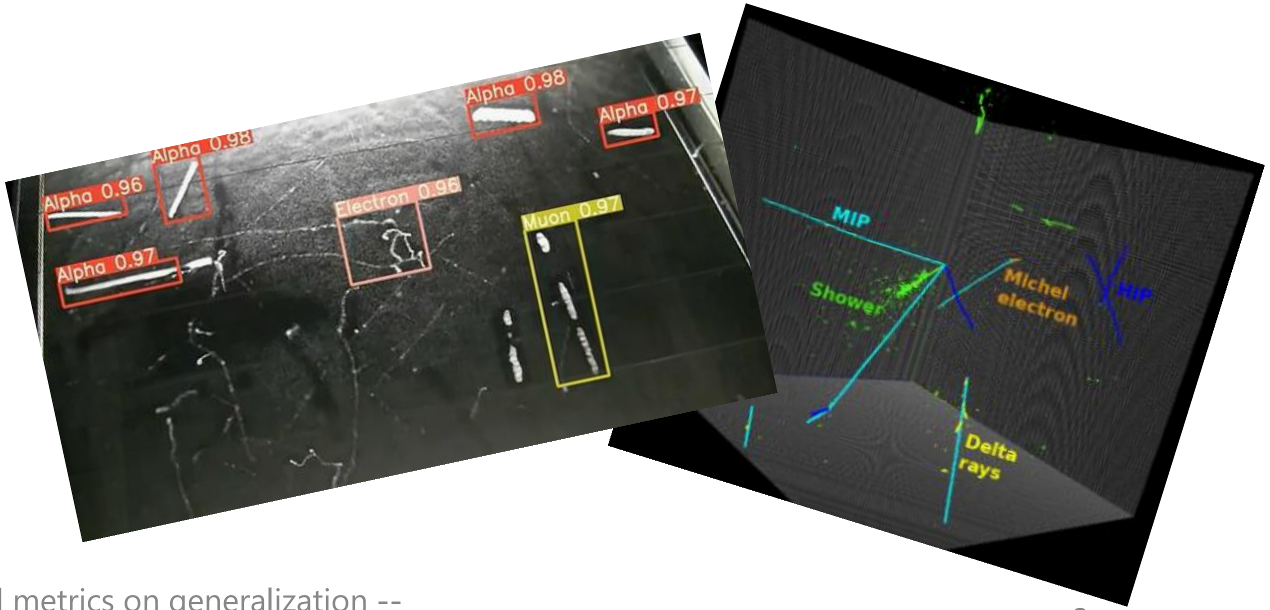
[1] University of Bologna, Department of Physics and Astronomy
[2] National Institute for Nuclear Physics, Bologna
[3] German Research Center for Artificial Intelligence (DFKI), Agents and Simulated Reality Department, Saarbruecken

Outline

- Why is this relevant?
- Background & Challenges
- Loss functions
- Performance metrics
- Results & Conclusions

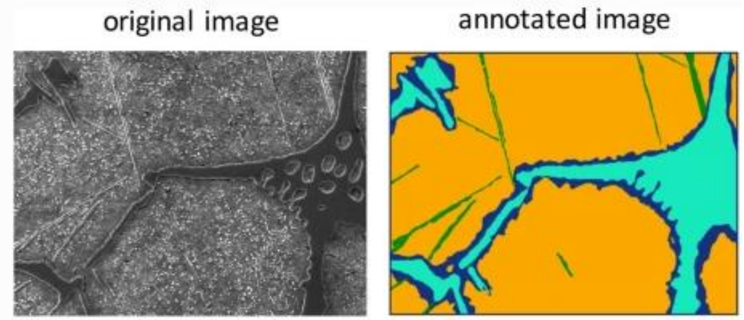
Why is this relevant?

- Object «recognition» is a popular learning task
  - Segmentation, detection and counting objectives
- Applications also in physics
  - 3D semantic segmentation task on simulated LArTPC samples [1]
  - Particle identification in cloud chamber



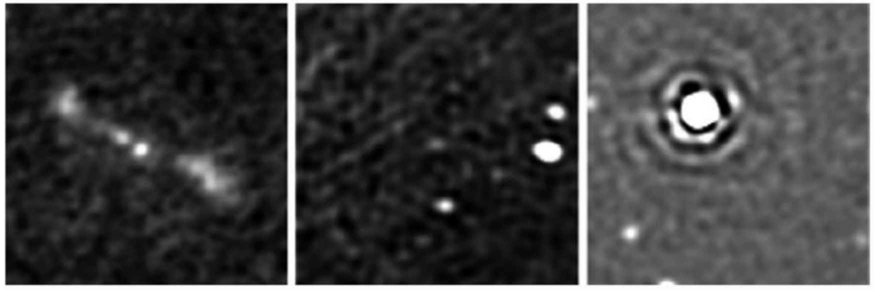
Why is this relevant?

- Object «recognition» is a popular learning task
  - Segmentation, detection and counting objectives
- Applications also in physics
  - 3D semantic segmentation task on simulated LArTPC samples [1]
  - Particle identification in cloud chambers
  - Semantic segmentation of microstructural constituents in SEM micrograph of ultrahigh carbon steel [2]



Why is this relevant?

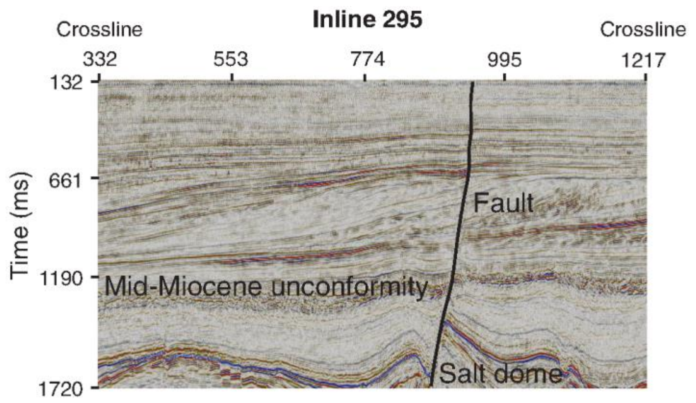
- Object «recognition» is a popular learning task
  - Segmentation, detection and counting objectives
- Applications also in physics
  - 3D semantic segmentation task on simulated LArTPC samples [1]
  - Particle identification in cloud chambers
  - Semantic segmentation of microstructural constituents in SEM micrograph of ultrahigh carbon steel [2]
  - Semantic Segmentation of Radio-Astronomical Images [3]



Examples of (left) galaxies, (center) sources and (right) sidelobes.

Why is this relevant?

- Object «recognition» is a popular learning task
  - Segmentation, detection and counting objectives
- Applications also in physics
  - 3D semantic segmentation task on simulated LArTPC samples [1]
  - Particle identification in cloud chambers
  - Semantic segmentation of microstructural constituents in SEM micrograph of ultrahigh carbon steel [2]
  - Semantic Segmentation of Radio-Astronomical Images [3]
  - Seismic facies interpretation [4]



Why is this relevant?

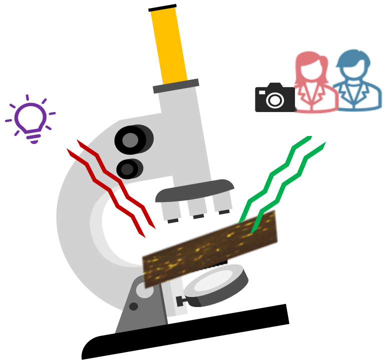
- Object «recognition» is a popular learning task
  - Segmentation, detection and counting objectives
- Applications also in physics
- Different applications typically share **similar challenges**
- **Aim**
  - Investigate loss effectiveness
  - Explore and compare several evaluation strategies

Background

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

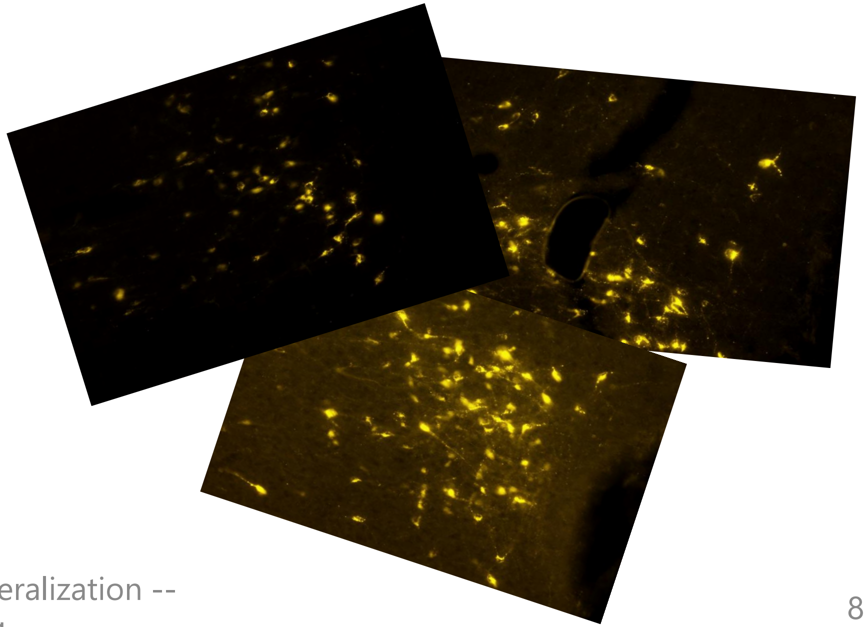
Fluorescent Microscopy

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



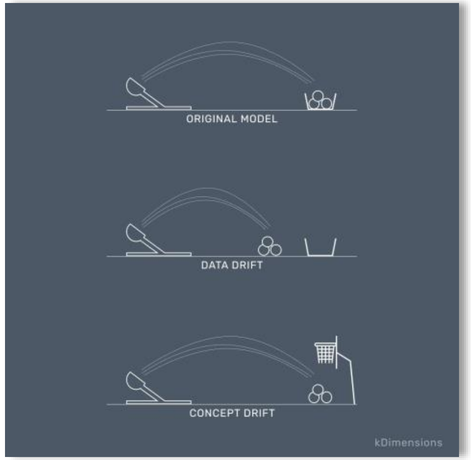
Applications

- Very popular in life science
- Torpor onset [5, 6, 7]
- Cytoplasmatic neuronal structures
- Variability in shape, size and color hue
- **Goal:** count stained structures



Problem

Visual by kDimensions



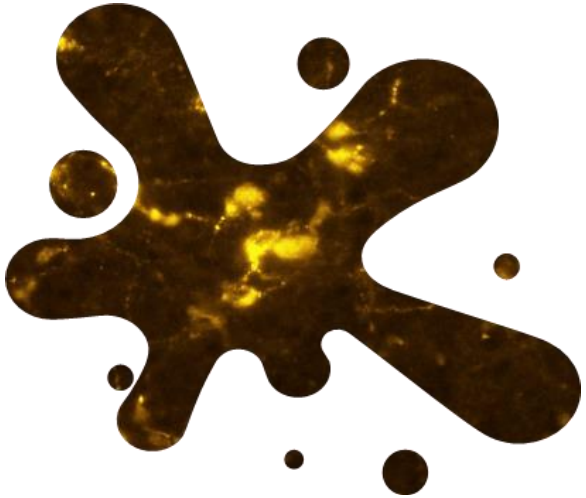
Manual processing

Time-consuming
Error-prone
Subjectivity of borderline cases

Hard to adapt Deep Learning solutions

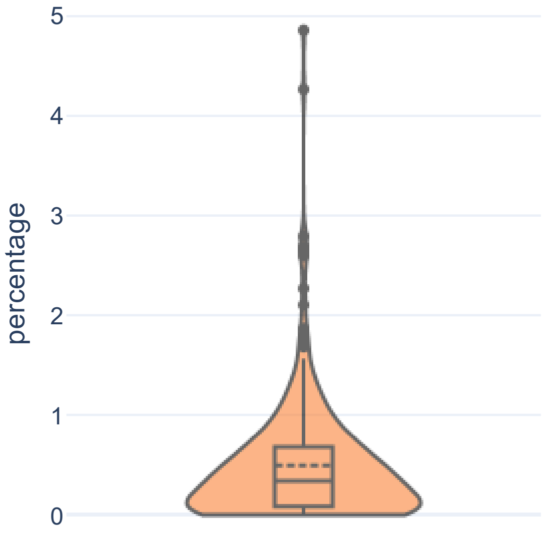
Domain shift
Few in-domain annotated datasets
How to train? How to evaluate?

Challenges

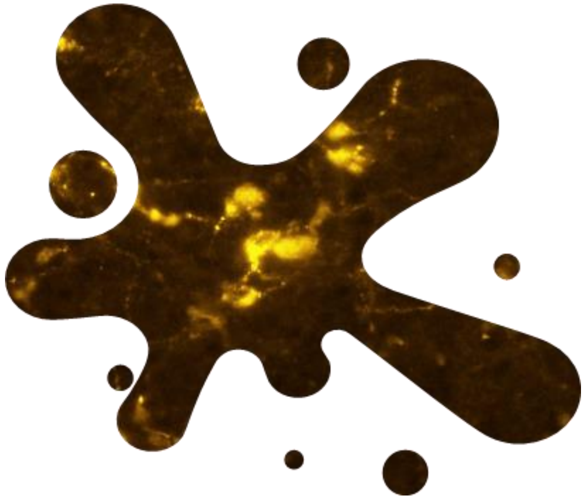


Class imbalance

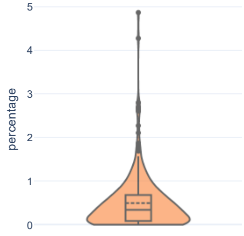
LESS THAN 5% SIGNAL PIXELS!



Challenges

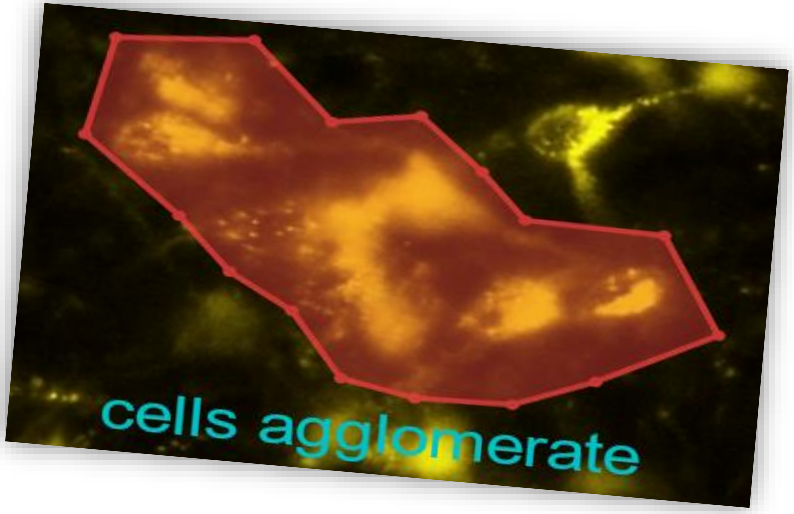


Class imbalance

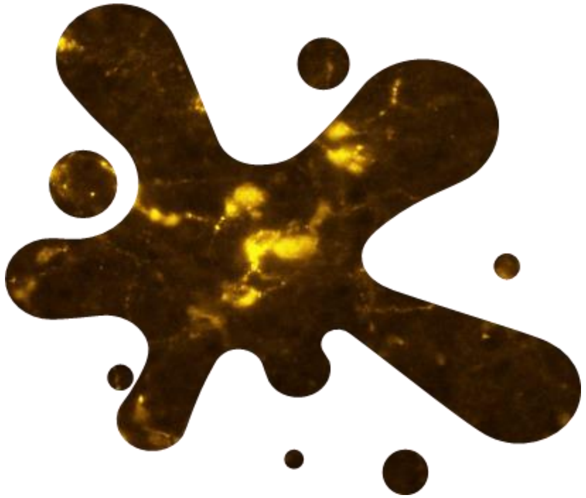


LESS THAN 5% SIGNAL PIXELS!

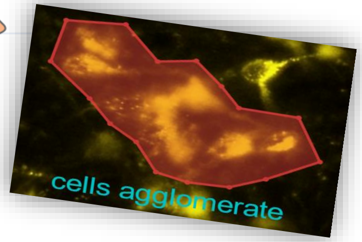
Overcrowding

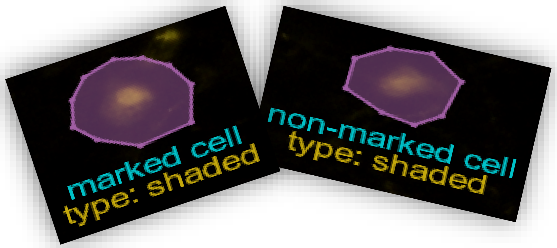


Challenges

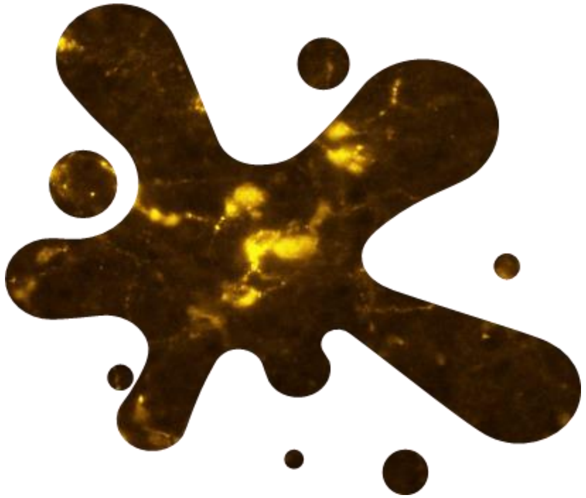


- Class imbalance
- Overcrowding
- Noisy labels

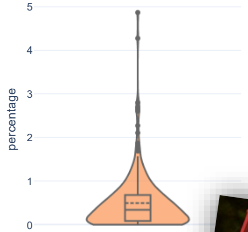




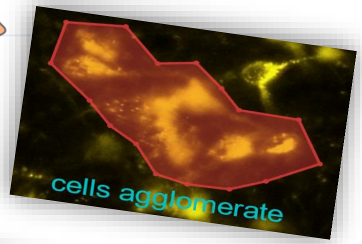
Challenges

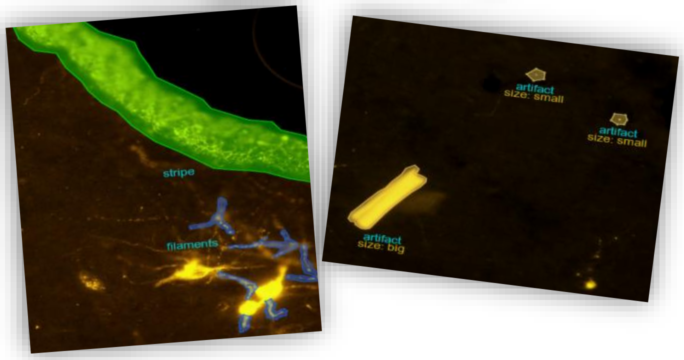


- Class imbalance
- Overcrowding
- Noisy labels
- Artefacts



LESS THAN 5% SIGNAL PIXELS!





Loss functions

Which loss should we choose to address or mitigate these challenges?

Weighted Binary Cross Entropy [8]

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

- de facto standard for classification
  → suitable for binary segmentation
- weighted version mitigates **class imbalance**
  → assign higher weights to underrepresented class
- No explicit **segmentation focus**
- No explicit **object-level error focus**
- No explicit **noise and systematics focus**

Focal Loss [9]

$$FL(p_t) = -\alpha_t(1 - p_t)^{\gamma} log(p_t)$$

$$p_t = \begin{cases} p, & \text{if } y = 1 \\ 1 - p, & \text{otherwise} \end{cases}$$

- Oversample wrong predictions, focus on hard examples
  → indirectly helpful for any challenge
- No explicit **segmentation focus**
- No explicit **object-level error focus**

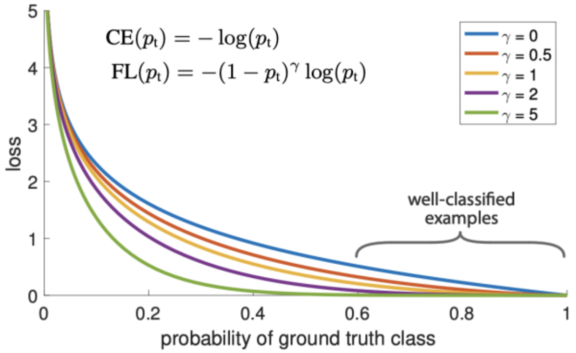


Figure 1. We propose a novel loss we term the *Focal Loss* that adds a factor $(1 - p_t)^{\gamma}$ to the standard cross entropy criterion. Setting $\gamma > 0$ reduces the relative loss for well-classified examples ($p_t > .5$), putting more focus on hard, misclassified examples. As our experiments will demonstrate, the proposed focal loss enables training highly accurate dense object detectors in the presence of vast numbers of easy background examples.

Dice Loss [10]

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

- 👍 targets **segmentation performance** directly
  → optimization aligned with learning goal
- 👎 No explicit **object-level error focus**
- 👎 No explicit **noise and systematics focus**
- 👎 Low impact of **small objects**
  → poor generalization WRT object size

Focal-Tversky Loss [11]

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

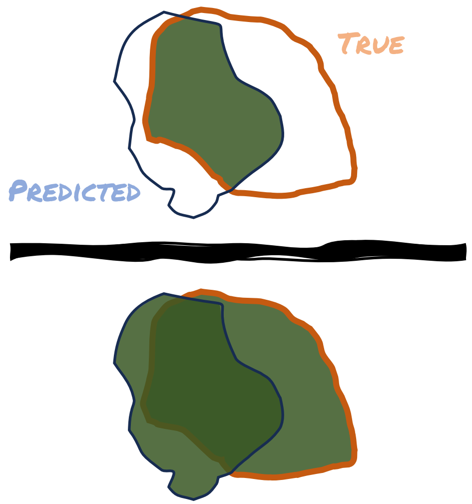
- 👍 Brings together advantages of Dice and Focal losses
  - → direct focus on segmentation
  - → indirectly helpful for any challenge
- 👎 No explicit **object-level error focus**

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: $\lambda_1 = 0.5, \lambda_2 = 0.3, \lambda_3 = 0.5$

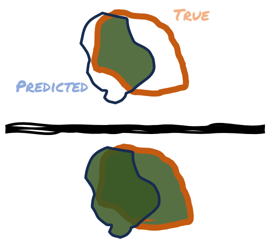
Evaluation Metrics

- Segmentation
  - Mean Intersection over Union (mIoU) = $\frac{Intersection}{Union}$
  - threshold: 0.4



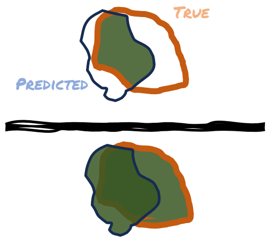
Evaluation Metrics

- Segmentation
  - Mean Intersection over Union (mIoU) = $\frac{Intersection}{Union}$
  - threshold: 0.4

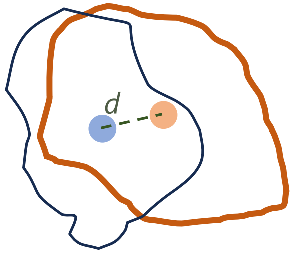


Evaluation Metrics

- Segmentation
  - Mean Intersection over Union (mIoU) = $\frac{Intersection}{Union}$
  - threshold: 0.4

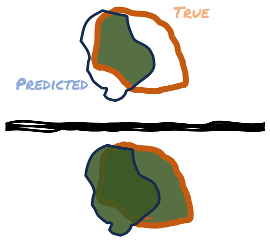


- Detection
  - Centers distance
  - threshold: 40 pixels (mean cell diameter)

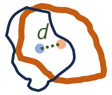


Evaluation Metrics

- Segmentation
  - Mean Intersection over Union (mIoU) = $\frac{Intersection}{Union}$
  - threshold: 0.4

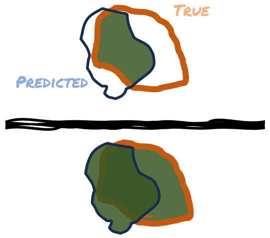


- Detection
  - Centers distance
  - threshold: 40 pixels (mean cell diameter)

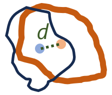


Evaluation Metrics

- Segmentation
  - Mean Intersection over Union (mIoU) = $\frac{Intersection}{Union}$
  - threshold: 0.4

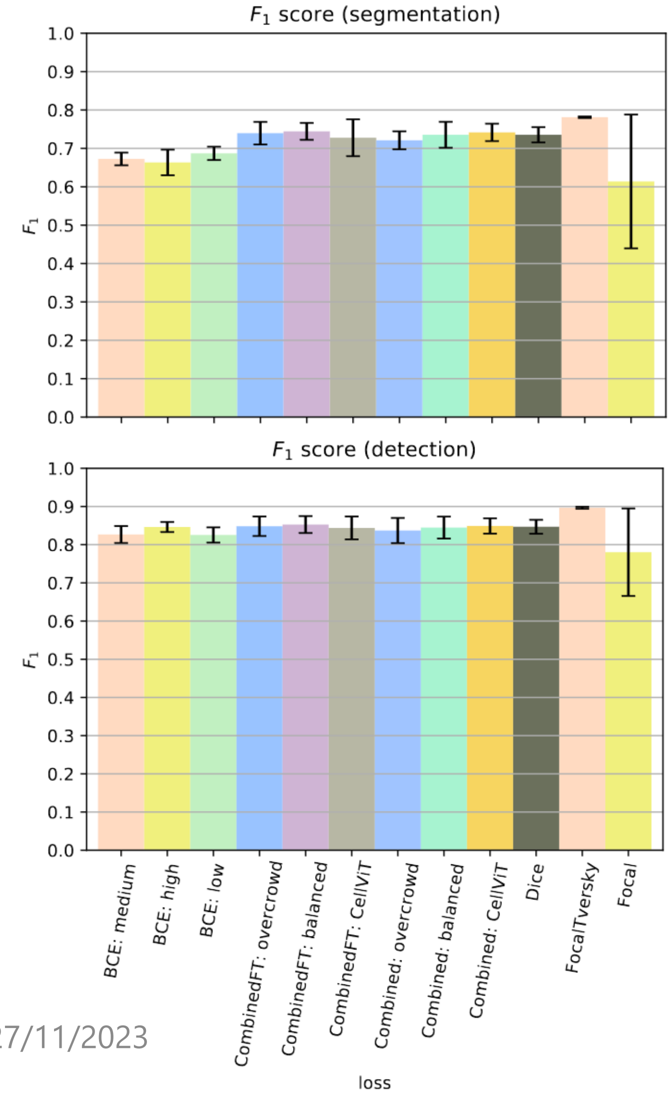


- Detection
  - Centers distance
  - threshold: 40 pixels (mean cell diameter)



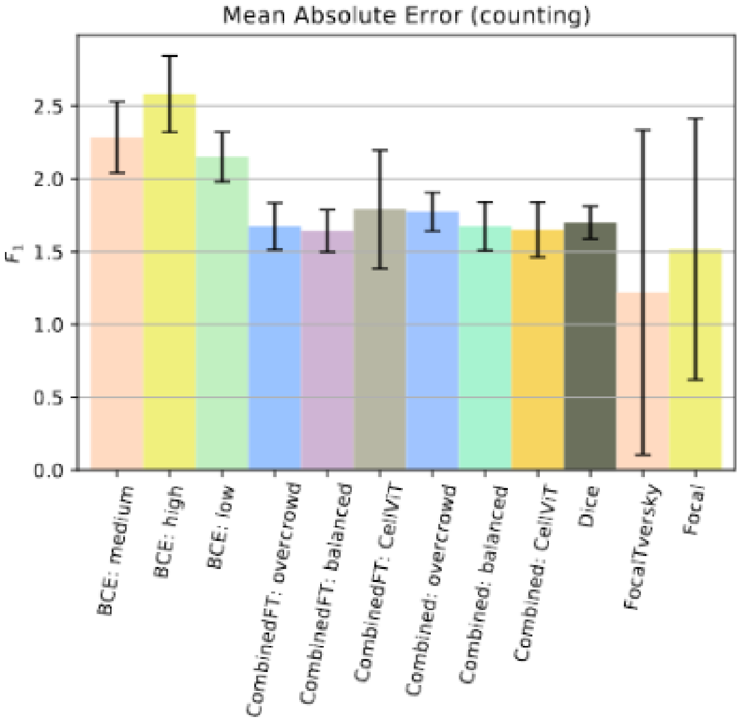
- Counting
  - Mean Absolute Error
  - Median Absolute Error
  - Mean Percentage Error: $\frac{(n_t - n_p)}{\max(n_t, 1)} * 100$

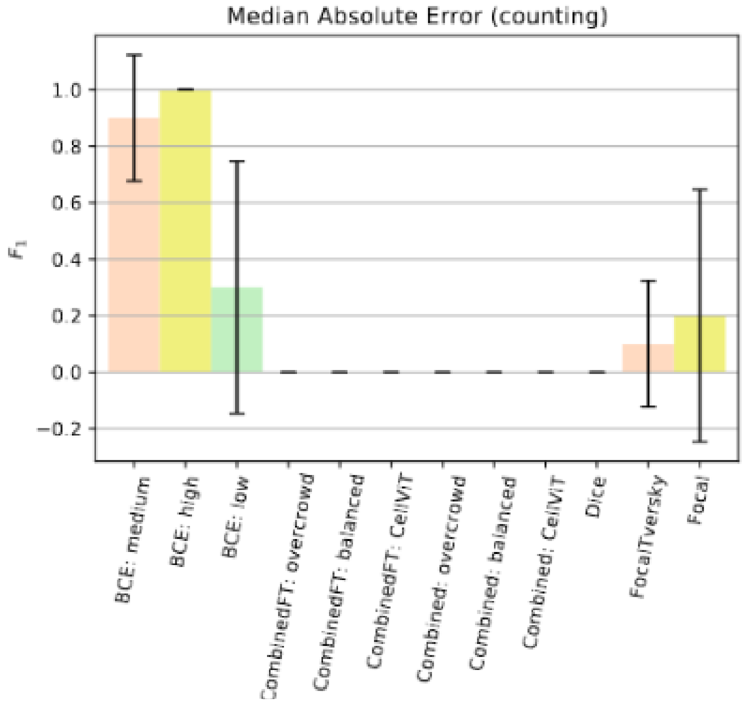
Segmentation & Detection

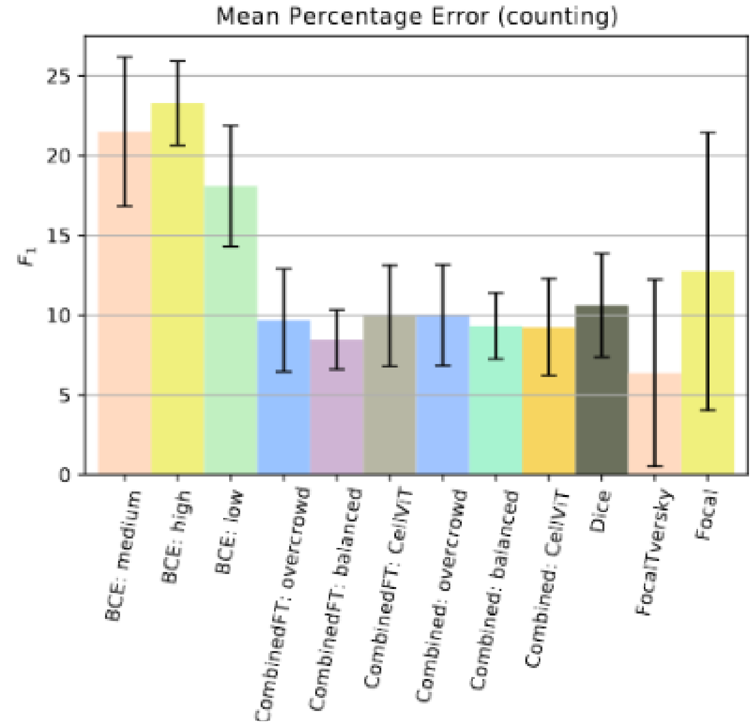


Loss	F1 score (IoU)	F1 score (distance)
BCE: medium	0.673±0.017	0.827±0.022
BCE: high	0.663±0.033	0.846±0.013
BCE: low	0.687±0.017	0.825±0.020
CombinedFT: overcrowd	0.740±0.029	0.848±0.026
CombinedFT: balanced	0.744±0.022	0.853±0.022
CombinedFT: CellViT	0.728±0.048	0.844±0.030
Combined: overcrowd	0.721±0.023	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



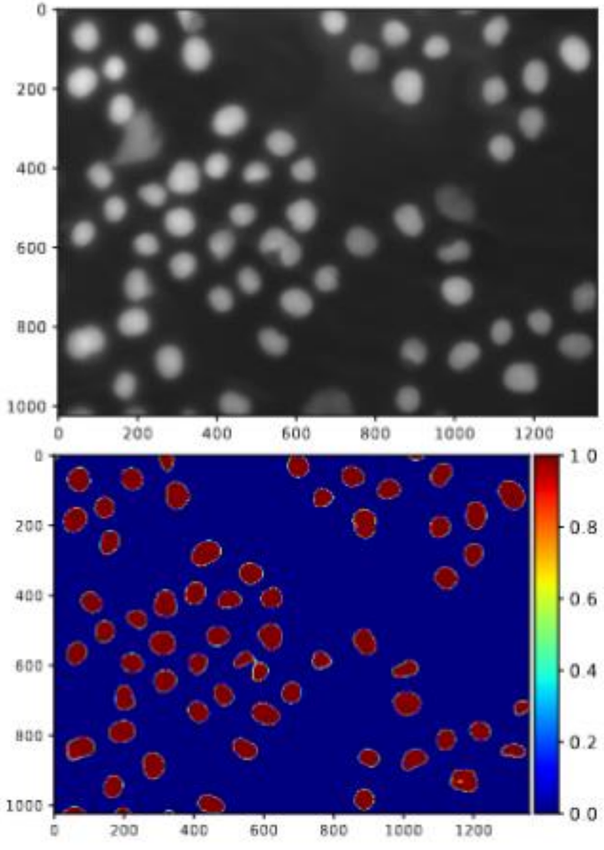




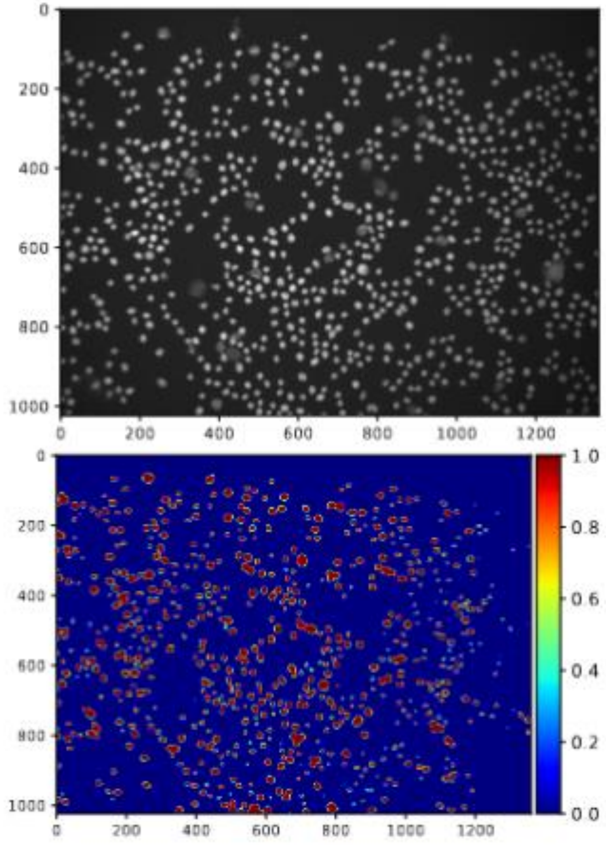
	BCE: medium	*BCE: high*	*BCE: low*	*CombinedFT: overcrowd*	*CombinedFT: balanced*	*CombinedFT: CellViT*	*Combined: overcrowd*	*Combined: balanced*	*Combined: CellViT*	*Dice*	*Focal Tversky*	*Focal*
MAE	2.286±0.245	2.583±0.263	2.151±0.171	1.674±0.161	1.643±0.144	1.791±0.406	1.774±0.132	1.674±0.166	1.651±0.188	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

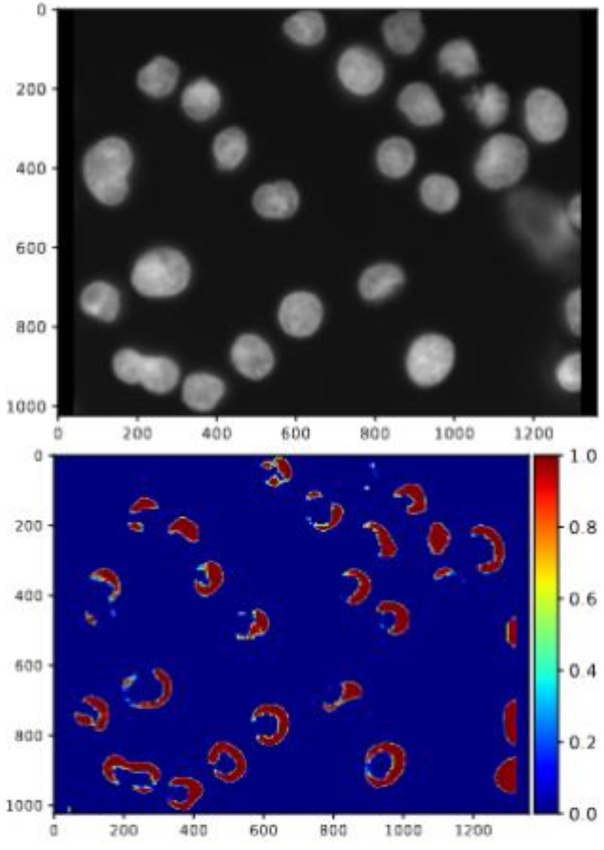
[12] S-BSST265 dataset



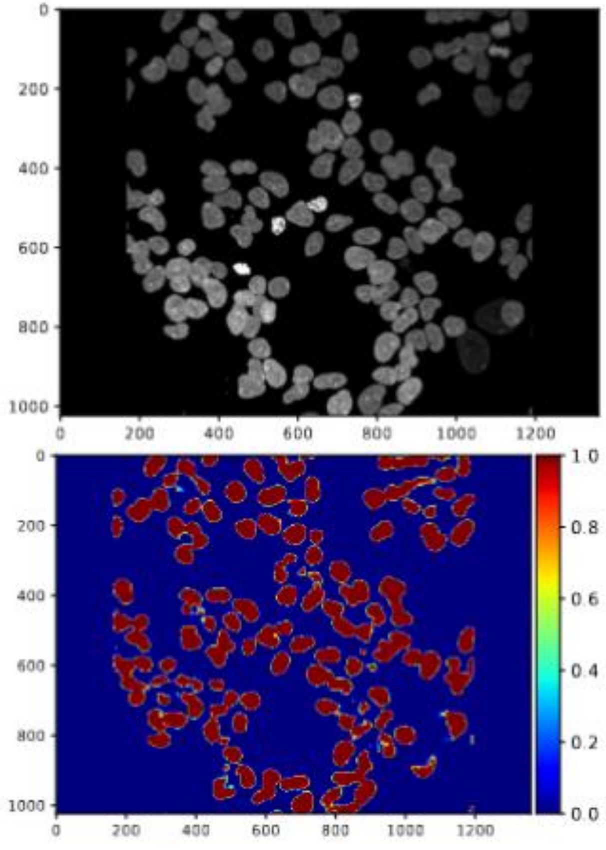
(a) Medium-sized, sharp objects



(b) Small objects



(c) Uneven texture and filling



(d) Overcrodwing

Conclusions

- **Focal Tversky** loss overperforms other losses
- Still some trouble 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 → single object errors**
- Multiple metrics enable comprehensive assessment

References

[1] Dominé, L. and Terao, K: Scalable deep convolutional neural networks for sparse, locally dense liquid argon time projection chamber data. Phys. Rev. D 102 (2020).

[2] Holm, E.A., *et al.* Overview: Computer Vision and Machine Learning for Microstructural Characterization and Analysis. *Metall Mater Trans A* **51** (2020).

[3] Pino, C., et al.: Semantic Segmentation of Radio-Astronomical Images. In Progress in Artificial Intelligence and Pattern Recognition. IWAIPR 2021. Lecture Notes in Computer Science (2021).

[4] Zhang, H. et al.: Automatic seismic facies interpretation using supervised deep learning. *Geophysics* (2021)

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

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

[7] 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)

[8] Pihur, V., Datta, S., Datta, S.: Weighted rank aggregation of cluster validation measures: a monte carlo cross-entropy approach. Bioinformatics 23(13), (2007).

[9] Lin, T. Y., et al.: Focal loss for dense object detection. ICCV Proceedings (2017).

[10] Sudre, C.H., et al.: Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. (2017).

[11] Nabila, A., and Khan, N. M.: "A Novel Focal Tversky Loss Function With Improved Attention U-Net for Lesion Segmentation." 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019) (2018).

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

[13] 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)

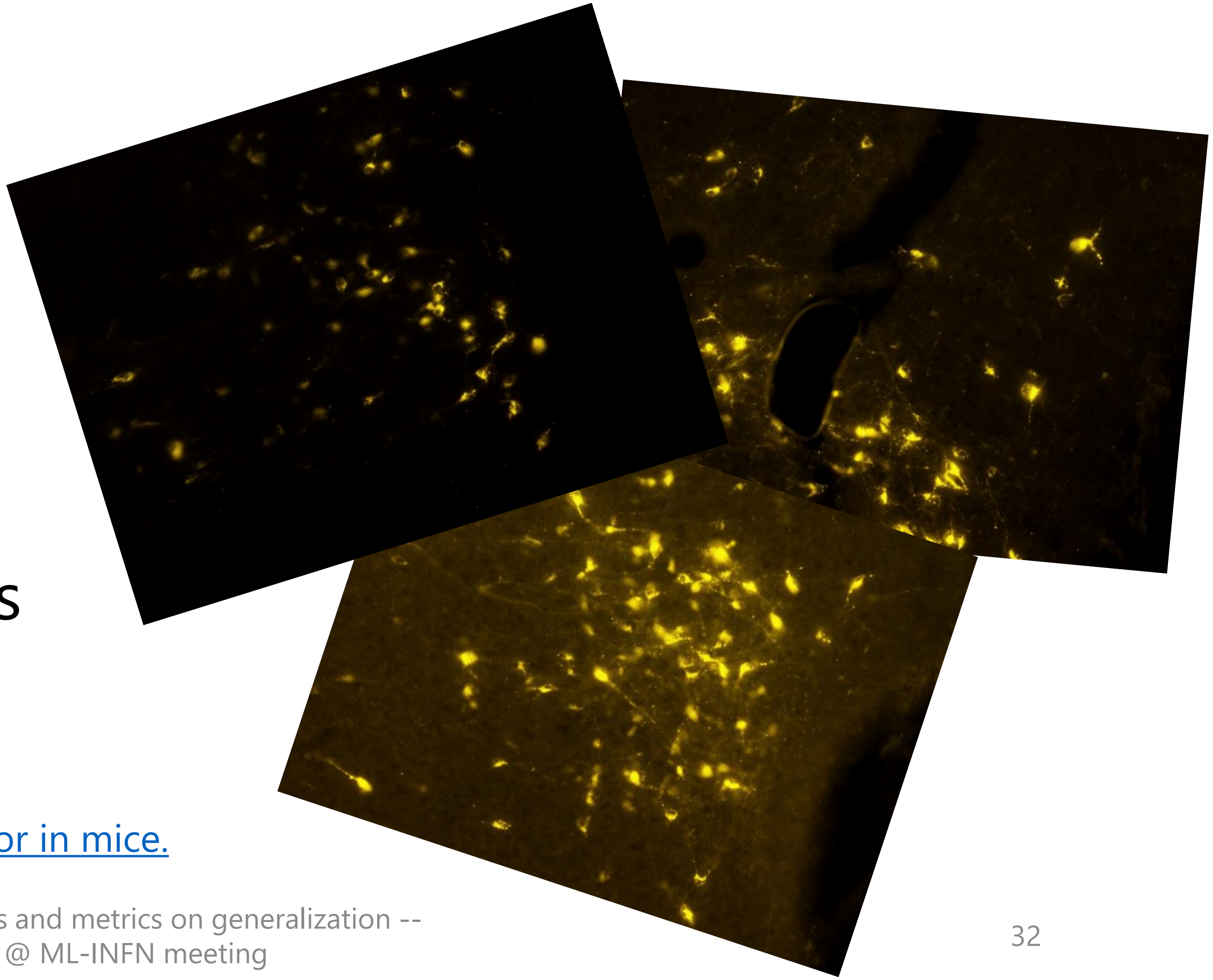
Thanks for your attention!

Questions?

Backup

Torpor onset

- 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).

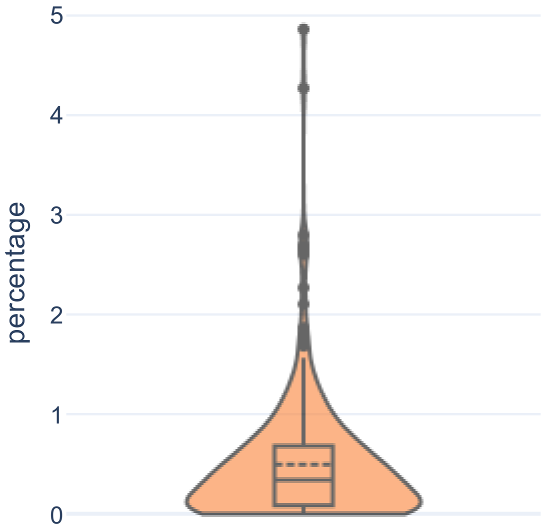
Contributions

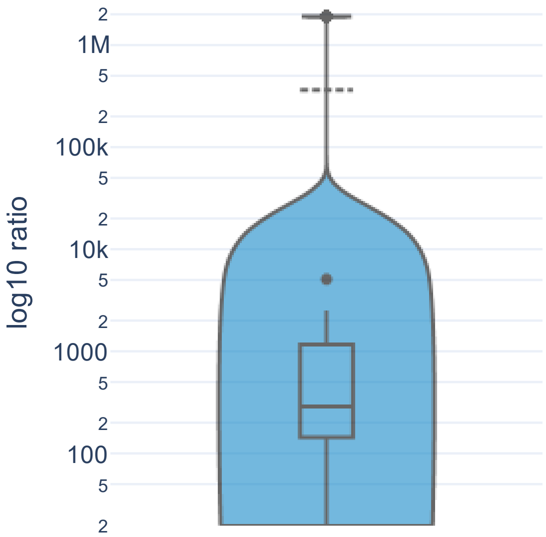
- Semantic segmentation using c-ResUnet [2] and Fluorescent Neuronal Cells v2 dataset (FNC v2) [3]
- Show the impact of loss functions on model performance
  - 60 ablation studies
  - 6 loss functions
- Inspect pros and cons of several evaluation metrics
- Discuss characteristics affecting out-of-sample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

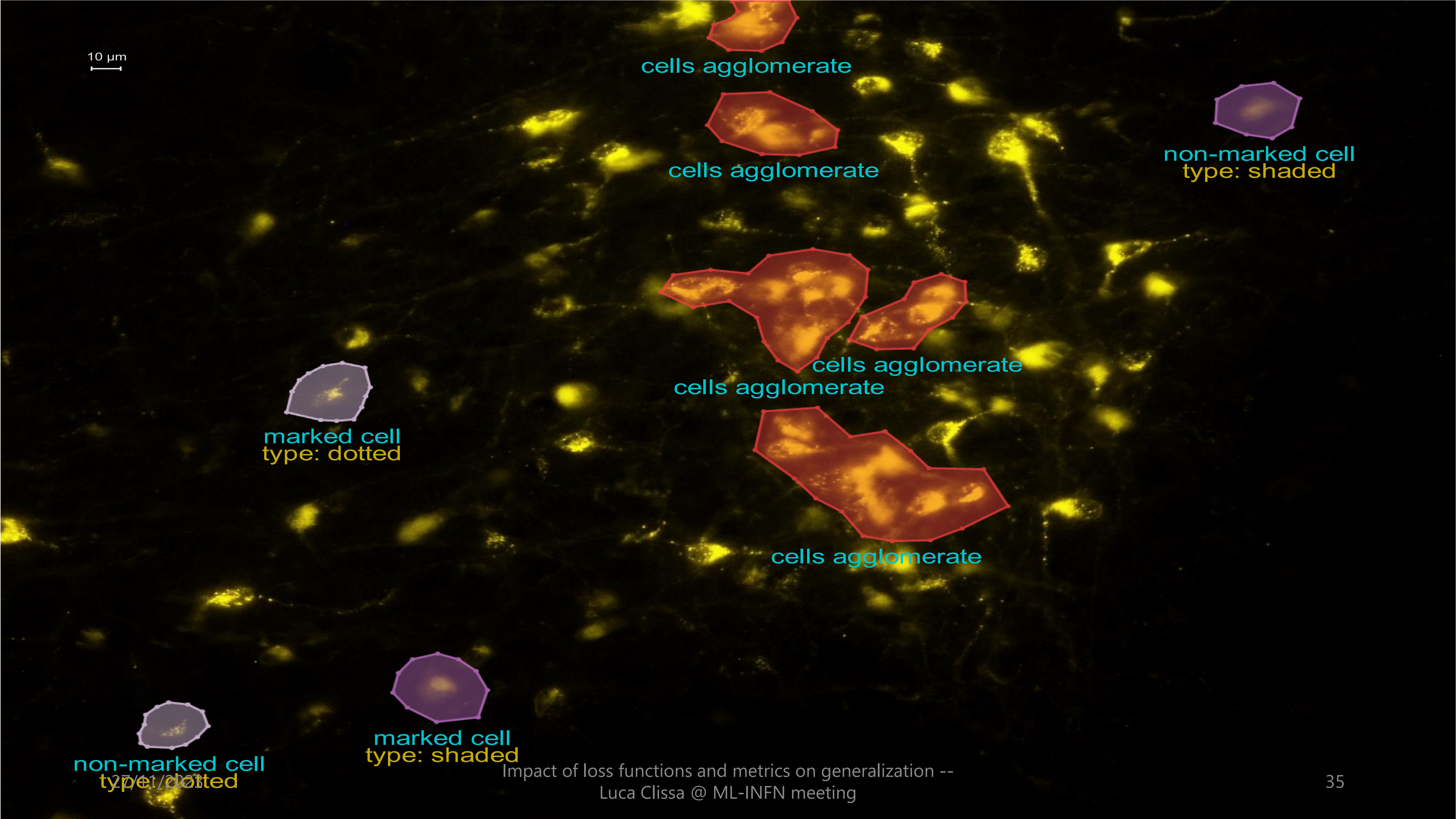
[2] Morelli, R., et al.: Automating cell counting in fluorescent microscopy through deep learning with c-ResUnet. Scientific Reports 11(1), 22920 (2021).
[3] Clissa, L., et al.: Fluorescent neuronal cells v2: Multi-task, multi-format annotations for deep learning in microscopy. arXiv preprint (submitted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

Class imbalance

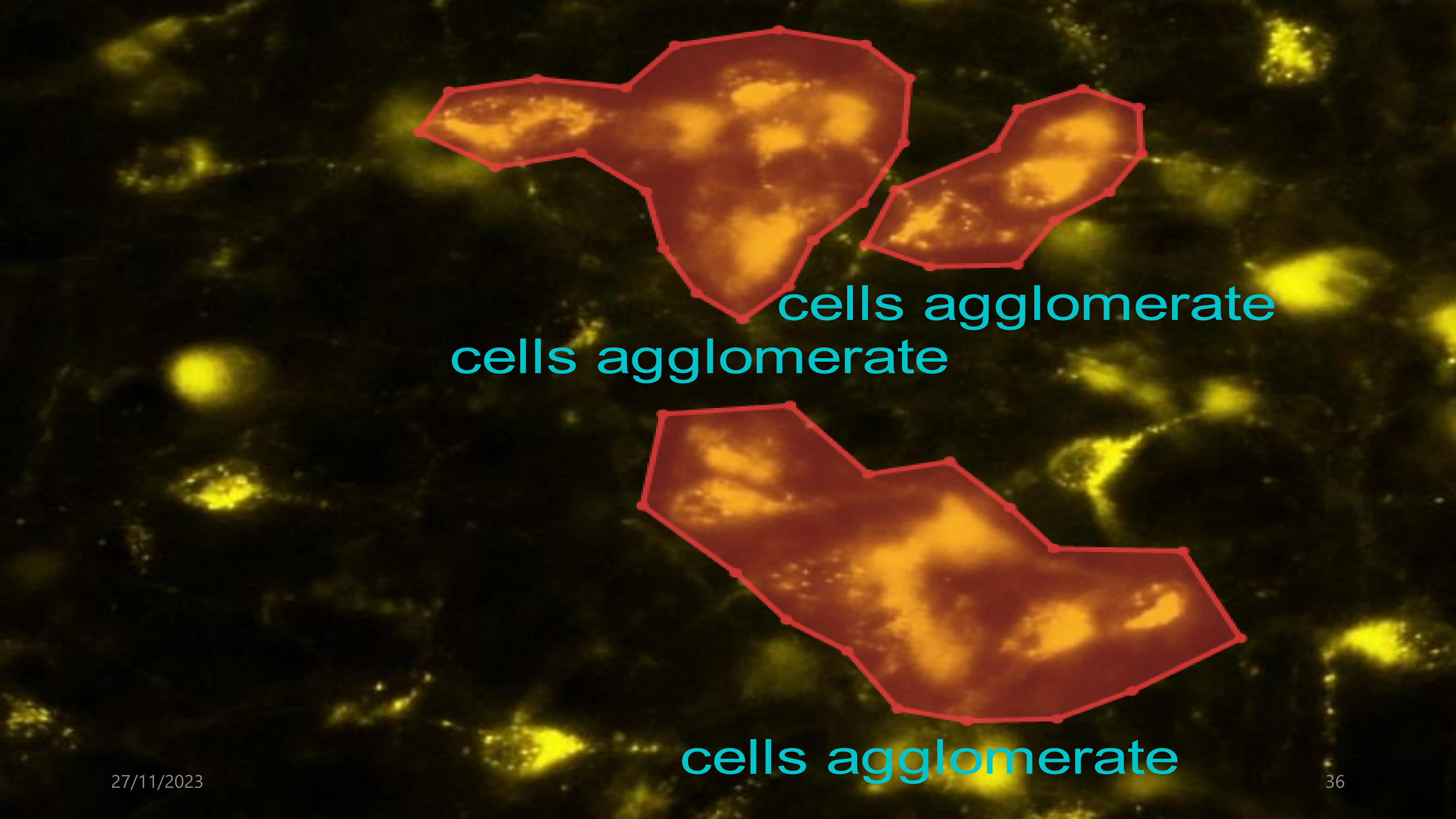
quantile	signal (%)	signal ratio
mean	0.50	367k
s.d.	0.61	756k
min	0	19.57
10%	0	92.39
25%	0.09	145.35
50%	0.34	291.10
75%	0.68	1k
90%	1.07	1.9M
max	4.86	1.9M

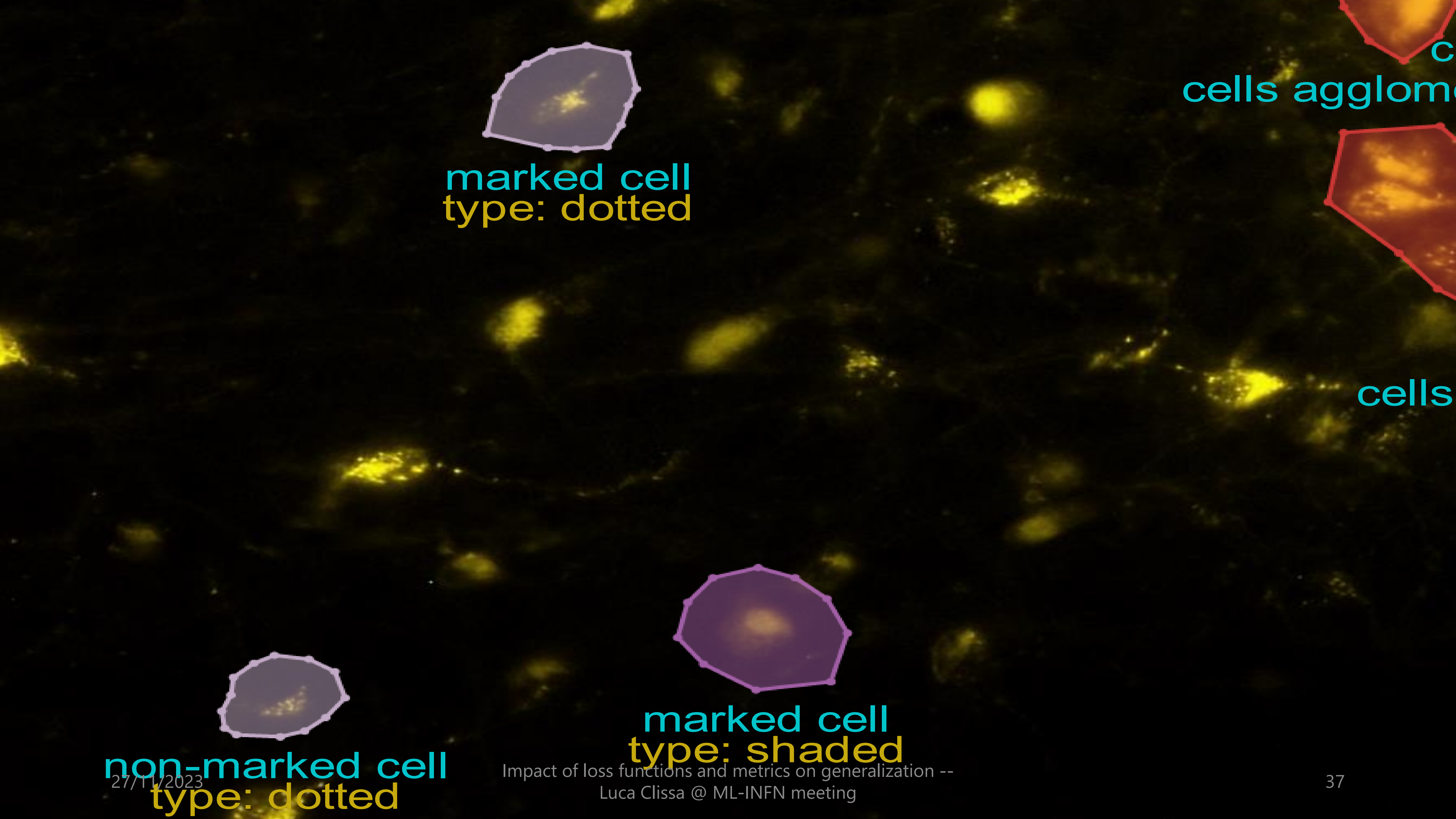


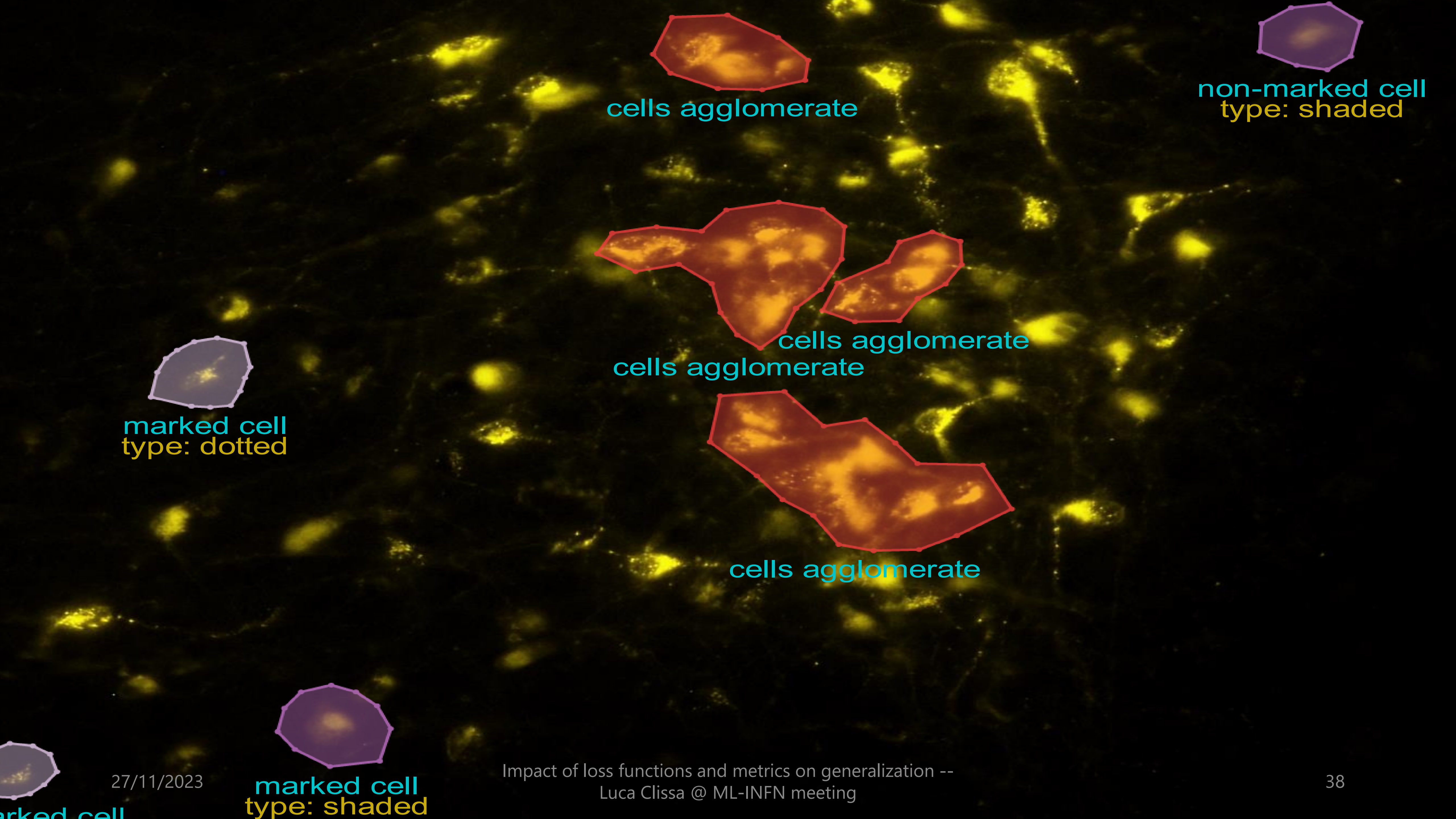




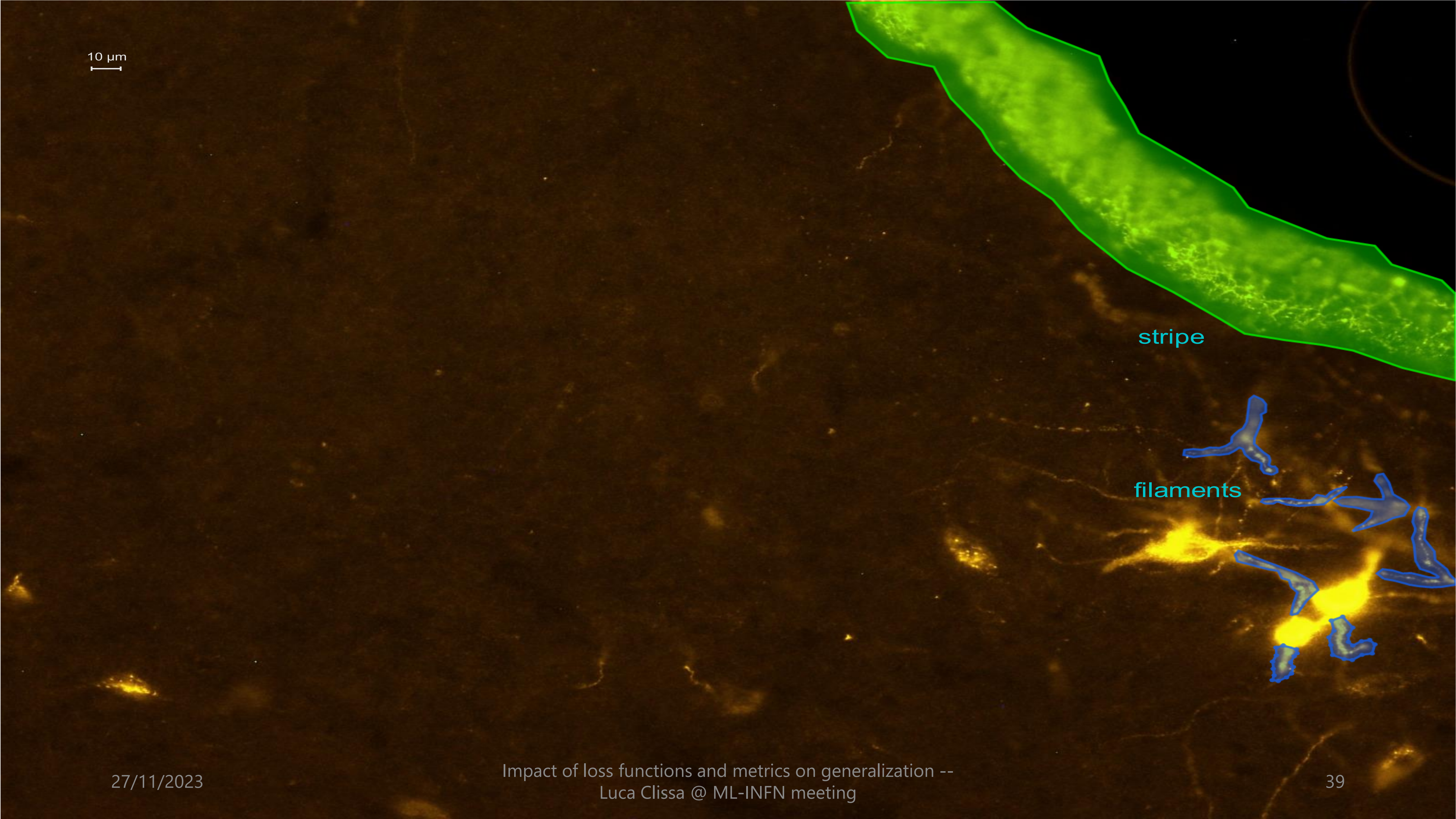
10 µm
cells agglomerate
cells agglomerate
non-marked cell
type: shaded
cells agglomerate
cells agglomerate
marked cell
type: dotted
cells agglomerate
marked cell
type: shaded
non-marked cell
type: dotted


cells agglomerate
cells agglomerate
cells agglomerate


marked cell
type: dotted
cells agglom
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 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

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

- 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

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