ICML | 2021
Thirty-eighth International Conference on Machine Learning

# Learning Generalized Intersection Over Union for Dense Pixelwise Prediction

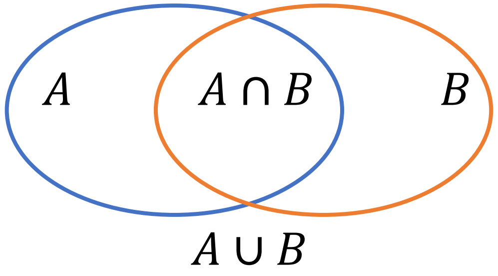
**Jiaqian Yu**[1], Jingtao Xu[1] , Yiwei Chen[1] , Weiming Li[1] , Qiang Wang[1] , Byung In Yoo[2] , Jae-Joon Han[2]
[1]Samsung Research China - Beijing (SRCB)
[2]Samsung Advanced Institute of Technology (SAIT)

Samsung Research

SAMSUNG ADVANCED INSTITUTE OF TECHNOLOGY

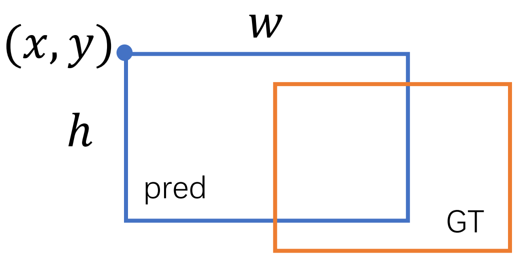
# Introduction

- Intersection over Union (IoU)



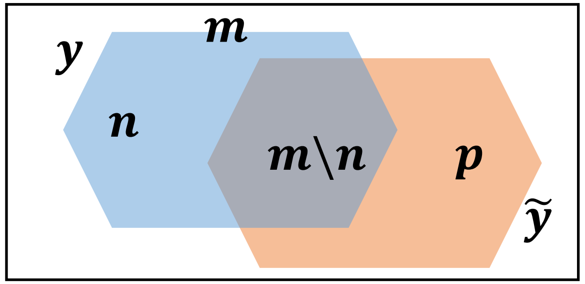

$$\mathrm{IoU} = \frac{A \cap B}{A \cup B}$$

- Bounding box regression




$$\mathrm{IoU} = \frac{\square \cap \square}{\square \cup \square} = \mathrm{IoU}(x, y, w, h)$$

- Dense pixelwise prediction

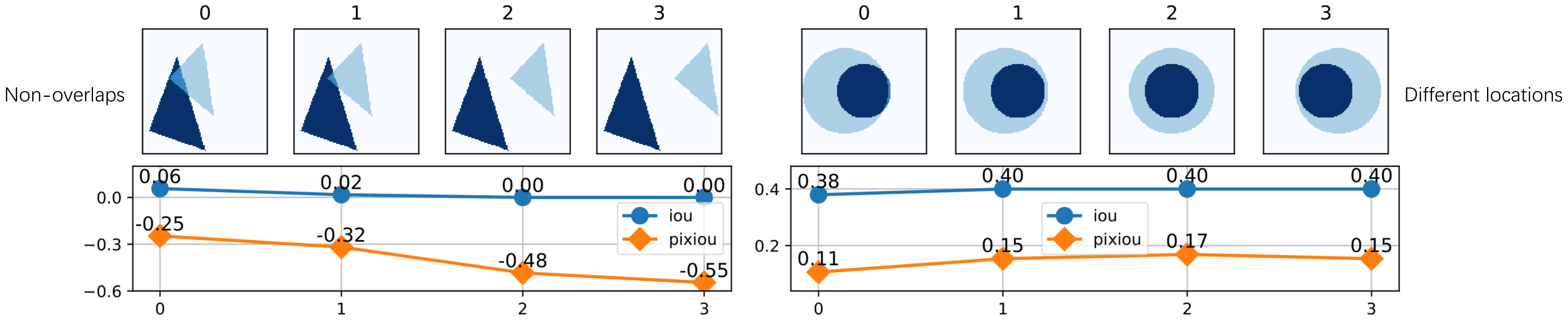



$$\mathrm{IoU} = \frac{\cap}{\cup} = \frac{|\boldsymbol{m} \backslash \boldsymbol{n}|}{|\boldsymbol{m} \cup \boldsymbol{p}|} = \frac{|\boldsymbol{m}| - |\boldsymbol{n}|}{|\boldsymbol{m}| + |\boldsymbol{p}|}$$

$$\boldsymbol{m} = \{\boldsymbol{y} = c\}$$
$$\boldsymbol{n} = \{\boldsymbol{y} = c, \tilde{\boldsymbol{y}} \neq c\}$$
$$\boldsymbol{p} = \{\boldsymbol{y} \neq c, \tilde{\boldsymbol{y}} = c\}$$

# Introduction

- Learning IoU: $L_{\text{iou}} = 1 - \text{IoU}$
- **Issue: zero-gradient for learning if (i) no-overlaps, (ii) only different locations**




- Human cognition can judge that the optimization should be further performed.
- Learning over IoU yields suboptimal performance and leads to slower convergence.
- **Our solution: PixIoU, steeper gradients for this cases.**

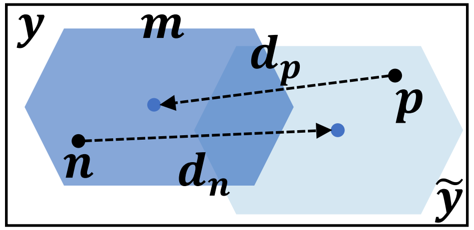
# Method

- Ingrate the coordinates of mispredicted pixels into the calculation:

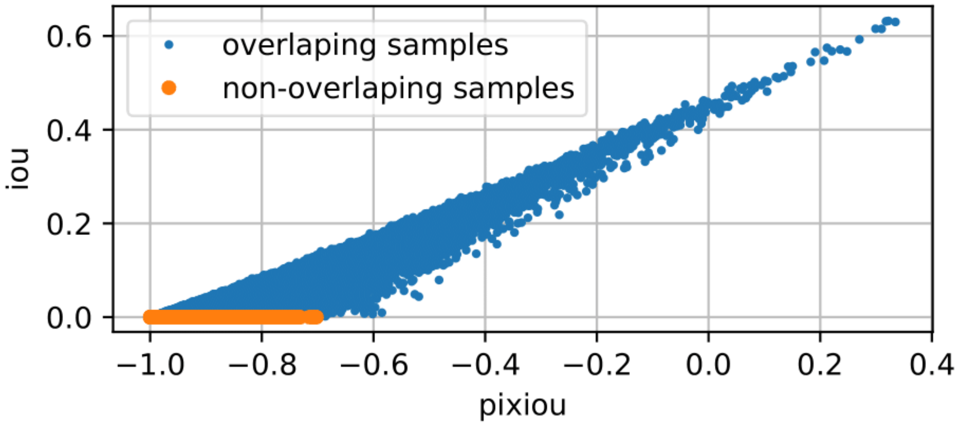
$$\mathrm{IoU} = \frac{|\boldsymbol{m}| - |\boldsymbol{n}|}{|\boldsymbol{m}| + |\boldsymbol{p}|}$$

**Definition**: $\mathrm{PixIoU} = \frac{|\boldsymbol{m}| - \langle \boldsymbol{d_n}, \boldsymbol{1_n} \rangle}{|\boldsymbol{m}| + \langle \boldsymbol{d_p}, \boldsymbol{1_p} \rangle} + \mathrm{IoU} - 1,$

$\boldsymbol{d_n}, \boldsymbol{d_p}$ is the normalized Euclidean distances to the centers.




- IoU: $O(N)$ for $N$ pixels.
- PixIoU: $O(kN)$ with $k$ additionally: a mean for the centers, an Euclidean, a dot product.




**Properties:**

1. PixIoU is invariant to the scale of the problem.
2. PixIoU is always a lower bound of IoU; it becomes tighter when the predictions get better.
3. PixIoU is well-bounded.

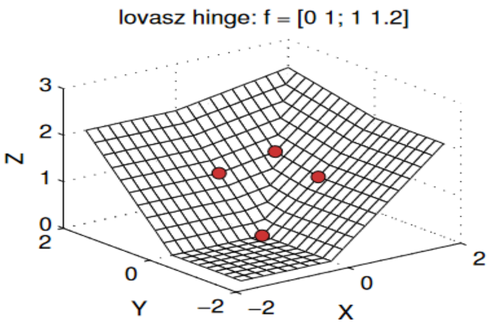
# Method

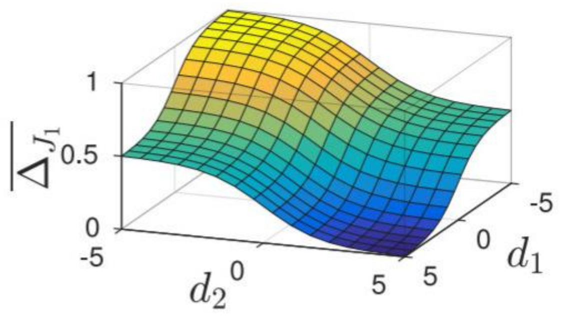
- Learning PixIoU: $L_{\text{pix}} = 1 - \text{PixIoU}$

> **Proposition**: Given a groundtruth, $L_{\text{pix}}(\cdot, \text{A})$, is **submodular** w.r.t. the set of mispredictions of A to the groundtruth.

- Submodular functions: diminishing returns property.
- Efficient surrogate function: the **Lovász surrogate**, yields a convex surface, provides a polynomial computation complexity.
- Lovász Softmax (Berman et al., 2018) is proposed for learning $L_{\text{iou}}$.

Yu et al., 2015, 2020



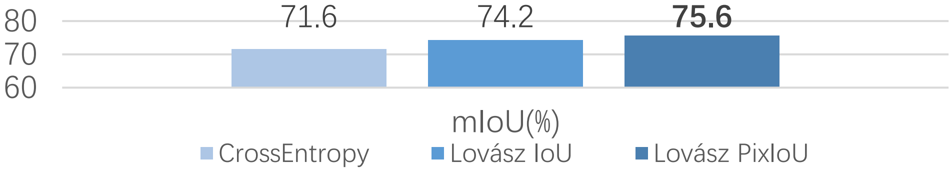



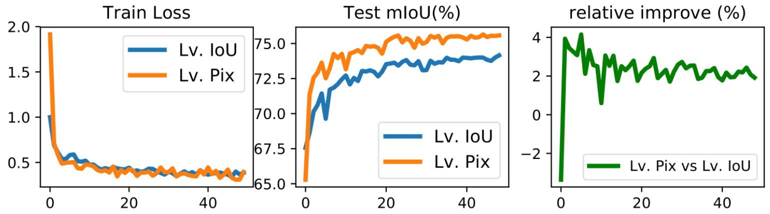

Berman et al., 2018

- **Lovász surrogate for learning $L_{\text{pix}}$: additionally $\langle d_n, 1_n \rangle$ and $\langle d_p, 1_p \rangle$.**

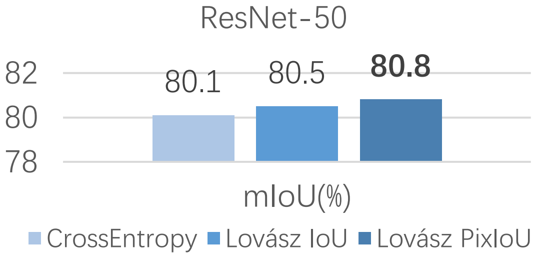
# Results

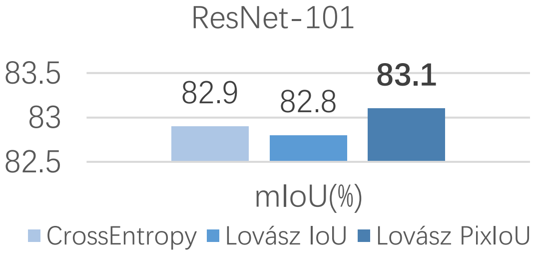
- Pixelwise Object Tracking on VOT2020

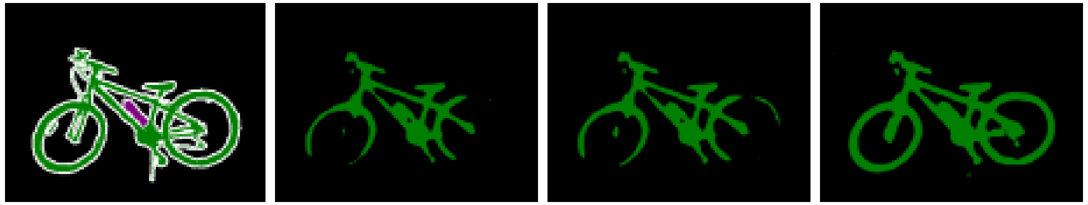



- Semantic segmentation on Pascal VOC

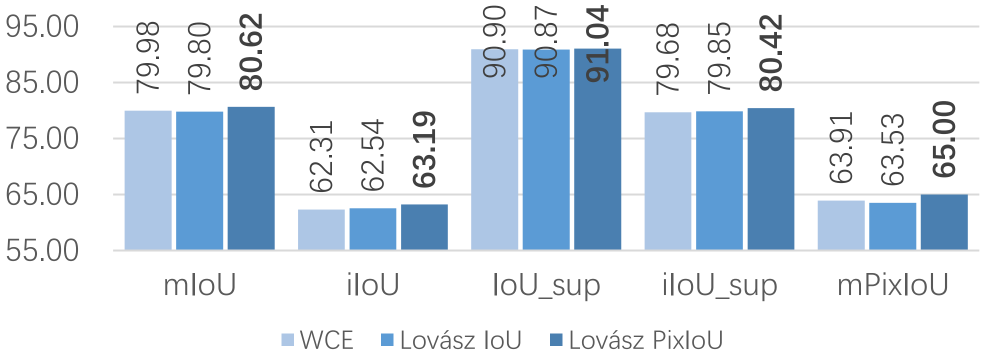


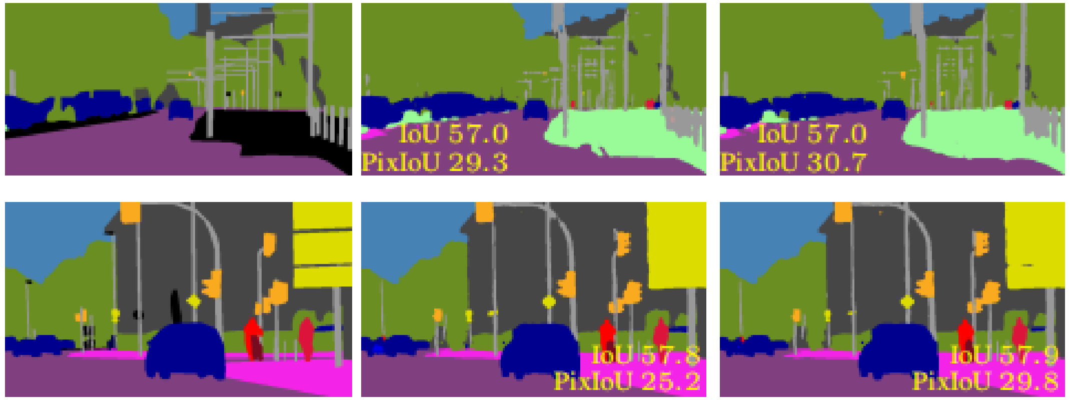



Groundtruth CrossEntropy Lovász IoU Lovász PixIoU

- Semantic segmentation on Cityscapes




PixIoU provides larger gradients than IoU.
Predictions with larger PixIoU provides better qualitative results.

More details please refer to our paper.

ICML | 2021

Thirty-eighth International Conference on Machine Learning

# Thank you!

## Learning Generalized Intersection Over Union for Dense Pixelwise Prediction

Paper: 2497

Samsung Research

SAMSUNG ADVANCED INSTITUTE OF TECHNOLOGY