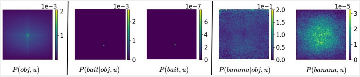
Long-tailed Instance Segmentation using Gumbel Optimized Loss

Konstantinos Panagiotis Alexandridis[1,2], Jiankang Deng[3], Anh Nguyen[2] and Shan Luo[1,2]
[1]King's College London, [2]University of Liverpool, [3]Imperial College London

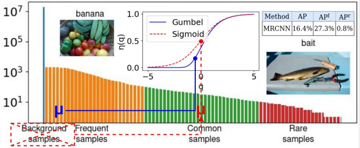
Summary

1. We identify the problem of activation functions in long-tailed instance segmentation for the first time, via extensive experiments.
2. We propose a new loss, i.e., Gumbel Optimized Loss (GOL), for long-tailed instance segmentation.
3. GOL surpasses the state-of-the-art in LVIS benchmark, by 1.1% AP.

Motivation



Let *obj* be object occurrence and *u* the location in the normalized grid. The long-tailed object distribution has low expected values for both *frequent* and *rare* objects. This is due to **class** and **location imbalance** problem.



Instance segmentation models, like Mask-RCNN (MRCNN), do not detect rare category objects having low average precision (AP^r) in long-tailed benchmarks like LVIS. They use Sigmoid activation that is not suitable for this task, in contrast our proposed method aligns better with the imbalanced distribution.

Hypothesis

Sigmoid/Softmax activation functions cannot effectively model long-tailed object distribution due to imbalance problem. Gumbel activation function is a better choice for modeling the extreme values of long-tailed object distribution.

Gumbel optimized Loss

We develop Gumbel activation η_γ as:

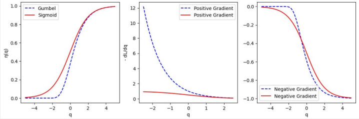
$$\eta_\gamma(q_i) = \exp(-\exp(-q_i)) \tag{1}$$

The loss using Gumbel Cross Entropy is:

$$L(\eta_\gamma(q_i), y_i) = \begin{cases} -\log(\eta_\gamma(q_i)), & if \;\; y_i = 1 \\ -\log(1-\eta_\gamma(q_i)), & if \;\; y_i = 0 \end{cases} \tag{2}$$

The gradient of Eq. 2 is:

$$\frac{dL(\eta_\gamma(q_i), y_i)}{dq_i} = \begin{cases} -\exp(-q_i), & if \;\; y_i = 1 \\ \frac{\exp(-q_i)}{\exp(\exp(-q_i))-1}, & if \;\; y_i = 0 \end{cases} \tag{3}$$

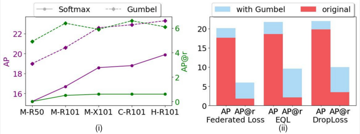


Using Gumbel activation, we develop GOL:

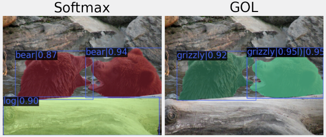
$$\mathcal{L}_{GOL} = -\sum_{j=1}^{C} \log(w_j^{Drop}\bar{p}_j), \;\; \bar{p}_j = \begin{cases} \eta_\gamma(q_i), & if \;\; y_j = 1 \\ 1-\eta_\gamma(q_i), & if \;\; not \end{cases} \tag{4}$$

where w_j^{Drop} are class specific weights proposed by DropLoss.

Results



(i) Gumbel has better performance than Softmax on MaskRCNN with ResNet-50, ResNet101, ResNeXt-101, Cascade-MaskRCNN with ResNet101 and Hybrid-Task-Cascade with ResNet101 using LVIS dataset. (ii) Gumbel activation boosts the performance of many SOTA models like Federated loss, EQL and DropLoss.



Method	AP	AP^r	AP^c	AP^f	AP^b
RFS[1]	23.7	13.3	23.0	29.0	24.7
Seesaw[2]	26.4	19.5	26.1	29.7	**27.6**
LOCE[3]	26.6	18.5	26.2	**30.7**	27.4
GOL (ours)	**27.7**	**21.4**	**27.7**	30.4	27.5

[1] Gupta et al.: *LVIS: A dataset for large vocabulary instance segmentation*, CVPR (2019)
[2] Wang et al.: *Seesaw loss for long-tailed instance segmentation*, CVPR (2021)
[3] Feng et al.: *Exploring classification equilibrium in long-tailed object detection*, CVPR (ICCV 2021)