Generalized Focal Loss: Learning Qualified and Distributed Bounding Boxes for Dense Object Detection

Xiang Li[1,2], Wenhai Wang[3,2], Lijun Wu[4], Shuo Chen[5,1], Xiaolin Hu[6], Jun Li[1], Jinhui Tang[1], Jian Yang[1*]

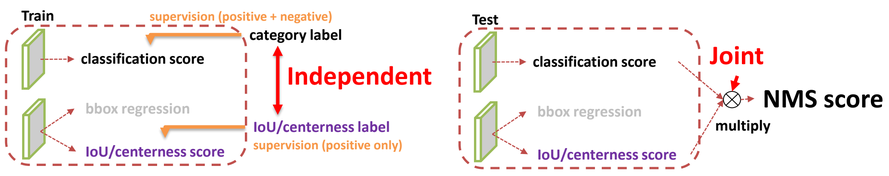
1. Nanjing University of Science and Technology 2. Momenta 3. Nanjing University
4. Microsoft Research 5. RIKEN Center for Advanced Intelligence Project 6. Tsinghua University

Abstract

One-stage detector basically formulates object detection as dense classification and localization. The classification is usually optimized by Focal Loss and the box location is commonly learned under Dirac delta distribution. A recent trend for one-stage detectors is to introduce an individual prediction branch to estimate the quality of localization, where the predicted quality facilitates the classification to improve detection performance. This paper delves into the representations of the above three fundamental elements: quality estimation, classification and localization. Two problems are discovered in existing practices, including (1) the inconsistent usage of the quality estimation and classification between training and inference, and (2) the inflexible Dirac delta distribution for localization. To address the problems, we design new representations for these elements. Specifically, we merge the quality estimation into the class prediction vector to form a joint representation, and use a vector to represent arbitrary distribution of box locations. The improved representations eliminate the inconsistency risk and accurately depict the flexible distribution in real data, but contain continuous labels, which is beyond the scope of Focal Loss. We then propose Generalized Focal Loss (GFL) that generalizes Focal Loss from its discrete form to the continuous version for successful optimization. On COCO test-dev, GFL achieves 45.0% AP using ResNet-101 backbone, surpassing state-of-the-art SAPD (43.5\%) and ATSS (43.6\%) with higher or comparable inference speed.

Motivation

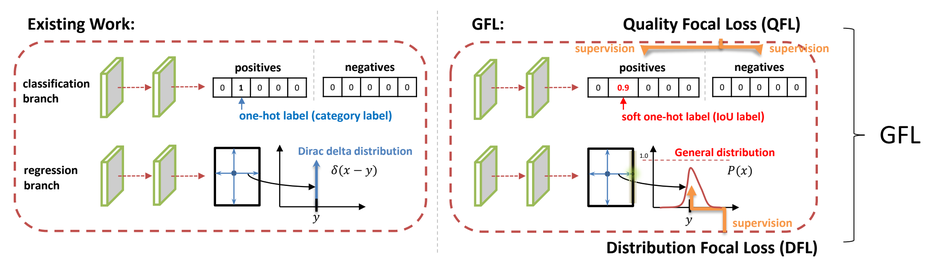
Inconsistent usage of box quality and cls score between train and test



Inflexible representation of bounding box

Improved Representation

Improved Representation for cls, box quality and box regression

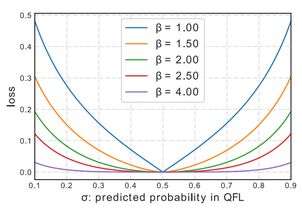


Generalized Focal Loss

Quality Focal Loss (QFL)

$$\mathbf{FL}(p) = -(1-p_t)^{\gamma}\log(p_t), p_t = \begin{cases} p, & \text{when } y=1 \\ 1-p, & \text{when } y=0 \end{cases}$$

$$\mathbf{QFL}(\sigma) = \mathbf{GFL}(1-\sigma,\sigma) = -\left|y-\sigma\right|^{\beta}\left((1-y)\log(1-\sigma)+y\log(\sigma)\right)$$

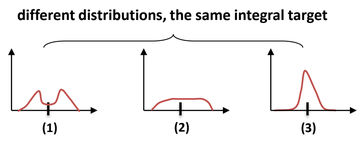


Distribution Focal Loss (DFL)

$$y = \int_{-\infty}^{+\infty}\delta(x-y)x\,\mathrm{d}x \Rightarrow \hat{y} = \int_{-\infty}^{+\infty}P(x)x\,\mathrm{d}x = \int_{y_0}^{y_n}P(x)x\,\mathrm{d}x \Rightarrow \hat{y} = \sum_{i=0}^{n}P(y_i)y_i \leftarrow \text{softmax } \mathcal{S}(\cdot) \text{ layer}$$

discretizing the range $[y_0, y_n]$ into a set of $\{y_0, y_1, ..., y_i, y_{i+1}, ..., y_{n-1}, y_n\}$

$$\mathbf{DFL}(\mathcal{S}) = \mathbf{GFL}(\mathcal{S}_i,\mathcal{S}_{i+1}) = -\left((y_{i+1}-y)\log(\mathcal{S}_i)+(y-y_i)\log(\mathcal{S}_{i+1})\right)$$



Generalized Focal Loss (GFL)

$$\mathbf{GFL}(p_{y_l},p_{y_r}) = -\left|y-(y_l p_{y_l}+y_r p_{y_r})\right|^{\beta}\left((y_r-y)\log(p_{y_l})+(y-y_l)\log(p_{y_r})\right), \quad \text{given } p_{y_l}+p_{y_r}=1$$

C FL, QFL and DFL are special cases of GFL

In this section, we show how GFL can be specialized into the form of FL, QFL and DFL, respectively.

FL: Letting $\beta=\gamma, y_l=0, y_r=1, p_{y_r}=p, p_{y_l}=1-p$ and $y\in\{1,0\}$ in GFL, we can obtain FL:

$$\begin{aligned}\mathbf{FL}(p) &= \mathbf{GFL}(1-p,p) = -\left|y-p\right|^{\gamma}\left((1-y)\log(1-p)+y\log(p)\right), y\in\{1,0\} \\ &= -(1-p_t)^{\gamma}\log(p_t), p_t = \begin{cases} p, & \text{when } y=1 \\ 1-p, & \text{when } y=0 \end{cases}\end{aligned} \tag{9}$$

QFL: Having $y_l=0, y_r=1, p_{y_r}=\sigma$ and $p_{y_l}=1-\sigma$ in GFL, the form of QFL can be written as:

$$\mathbf{QFL}(\sigma) = \mathbf{GFL}(1-\sigma,\sigma) = -\left|y-\sigma\right|^{\beta}\left((1-y)\log(1-\sigma)+y\log(\sigma)\right). \tag{10}$$

DFL: By substituting $\beta=0, y_l=y_i, y_r=y_{i+1}, p_{y_l}=P(y_l)=P(y_i)=\mathcal{S}_i, p_{y_r}=P(y_r)=P(y_{i+1})=\mathcal{S}_{i+1}$ in GFL, we can have DFL:

$$\mathbf{DFL}(\mathcal{S}_i,\mathcal{S}_{i+1}) = \mathbf{GFL}(\mathcal{S}_i,\mathcal{S}_{i+1}) = -\left((y_{i+1}-y)\log(\mathcal{S}_i)+(y-y_i)\log(\mathcal{S}_{i+1})\right). \tag{11}$$

Experimental Results and Analyses

The Effect of QFL and DFL / Speed-Accuracy Trade-off

QFL	DFL	FPS	AP	AP_{50}	AP_{75}
		19.4	39.2	57.4	42.2
✓		19.4	39.9	58.5	43.0
	✓	19.4	39.5	57.3	42.8
✓	✓	19.4	**40.2**	**58.6**	**43.4**

Table 3: **The effect of QFL and DFL on ATSS**: The effects of QFL and DFL are orthogonal, whilst utilizing both can boost 1% AP over the strong ATSS baseline, without introducing additional overhead practically.

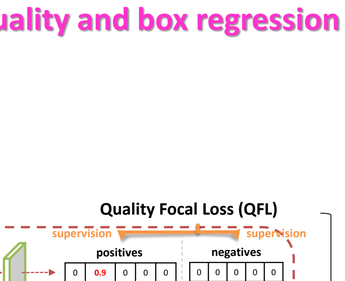
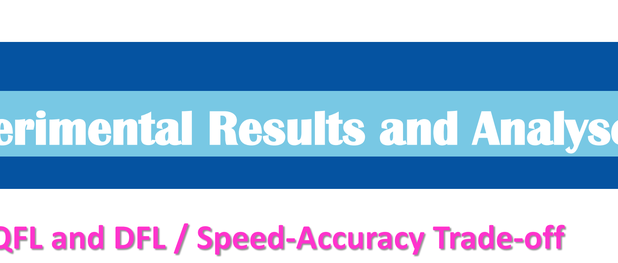


Figure 8: Single-model single-scale speed (ms) vs. accuracy (AP) on COCO test-dev among state-of-the-art approaches. GFL achieves better speed-accuracy trade-off than many competitive counterparts.

Analyses

QFL:

DFL:



Acknowledgement

This work was supported by Postdoctoral Innovative Talent Support Program of China under Grant BX20200168, NSFC 62072242, 61836014, U19B2034 and U1713208, Program for Changjiang Scholars.