Inverse Image Frequency for Long-tailed Image Recognition

Konstantinos Panagiotis Alexandridis[1,2], Shan Luo[1,2], Anh Nguyen[2], Jiankang Deng[3] and Stefanos Zafeiriou[3]

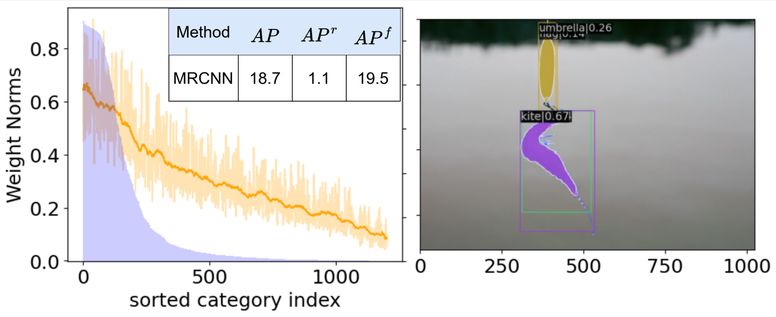
[1]King's College London, [2]University of Liverpool, [3]Imperial College London

Summary

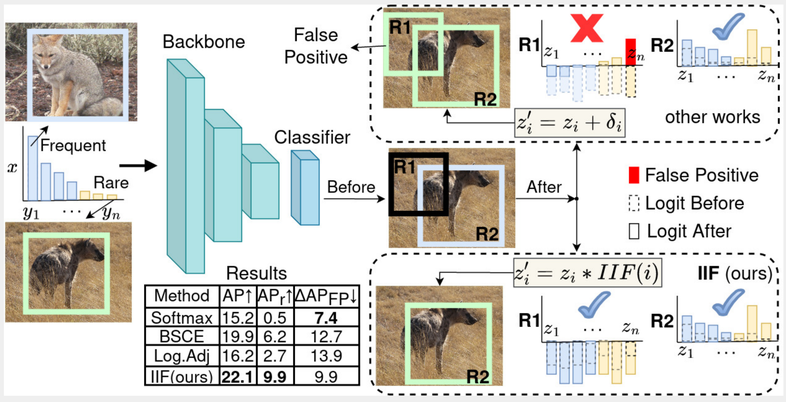
1. Models trained with class-imbalanced data, cannot recognise the rare classes of the dataset.
2. We propose Inverse Image Frequency IIF, for imbalanced object recognition and detection, IIF debiases the classifier and reduces false positive detections.
3. IIF surpasses the SOTA on both imbalanced image classification and instance segmentation benchmarks.

Motivation

Models trained with imbalanced data have biased classification norms that hinder their recognition ability for rare classes like *bait* and lower rare-class Average Precision AP^r.



Previous classifier debiasing methods, produce false positives, when transferred to imbalanced object detection, as they might confuse the background for a rare class.



Inverse Image Frequency

Image Frequency $IF(y, D)$ of a class $y \in \mathbb{N}$ is computed as the number of images, inside a trainset D, where an object o_y appears:

$$IF(y, D) = |\{image \in D : o_y \in image\}| \tag{1}$$

The class prior probability $p_D(y)$ of class y is defined as:

$$p_D(y) = \frac{IF(y, D)}{\sum_{y=1}^{y=C} IF(y, D)} \tag{2}$$

where C is the total number of classes in D and IIF is measured by taking the logarithm of the inverse of $p_D(y)$:

$$IIF(y, D) = -\log(p_D(y)) \tag{3}$$

IIF reweigths the model's logit z_y using the class prior probability $p_D(y)$, and during object detection, it is applied only to the foreground class predictions:

$$z_{IIF,y} = \begin{cases} -z_y \log(p_D(y)) & \text{if } y = \text{foreground} \\ z_y & \text{if } y = \text{background} \end{cases} \tag{4}$$

The calibrated IIF-prediction $\bar{q}_y$ is taken using Softmax:

$$\bar{q}_y = \frac{e^{-z_y \log(p_D(y))}}{e^{z_{bg}} + \sum_{j=1}^{C} e^{-z_j \log(p_D(j))}} = \frac{p_D(y)^{-z_y}}{e^{z_{bg}} + \sum_{j=1}^{C} p_D(j)^{-z_j}} \tag{5}$$

Eq. 5 makes Softmax more sensitive to the rare classes, because as $p_D(y) \to 1$, $p_D(y)^{-z_y} \to 1, \forall z_y \in \mathbb{R}$.
The gradient of Cross-Entropy (CE) with respect to z_i is:

$$\frac{\partial CE_{IIF}}{\partial z_i} = \begin{cases} -\log(p_D(i))(\bar{q}_i - 1) & \text{if } i = y \\ -\log(p_D(i))\bar{q}_i & \text{otherwise} \end{cases} \tag{6}$$

The positive IIF gradient, i.e. when $i = y$, is magnified when the class prior probability $p_D(i)$ is low, encouraging the model to learn more from rare classes.

Contact: konstantinos.alexandridis@kcl.ac.uk

Results

Results on long-tailed image classification, using Squeeze and Excite-ResNets, our method outperforms the SOTA.

Method	Acc
LADE	45.4
Log. Adj	43.0
RIDE	47.0
IIF (ours)	48.8

(a)CIFAR100-LT

Method	Acc
LADE	53.0
DisAlign	53.4
RIDE	55.9
IIF (ours)	**56.2**

(b)ImageNet-LT

Method	Acc
BSCE	38.7
DisAlign	39.3
LADE	38.8
IIF (ours)	**40.2**

(c)Places-LT

Results on LVISv1 using Mask-RCNN-ResNet50-RSB, our IIF achieves the best rare class performance AP^r.

Method	AP	AP^r	AP^c	AP^f	AP^b
RFS	25.4	13.0	25.5	30.9	24.9
DropLoss	25.7	14.4	26.6	29.7	25.1
NorCal	27.1	18.4	26.6	31.5	26.8
IIF (ours)	**27.4**	**19.4**	**26.8**	**31.5**	**27.4**

IIF detects the rare classes *parrot, owl, horse-carriage, giant panda* in contrast to Softmax, however, it fails to detect the *eagle* in the last image.

(a) IIF has better AP, using various models, (b) it produces more balanced classifier norms compared to Softmax.