Gradient Harmonized Single-Stage Detector

AAAI 2019

Paper Reading by Yunyan Yan

- 作者介绍
- 研究背景
- 本文方法
- 实验效果
- 总结反思

作者介绍

Biography

Buyu Li is a PhD candidate at Multimedia Lab (MMLab), The Chinese University of Hong Kong, supervised by Prof. Xiaogang Wang. He also has a close research collaboration with Yu Liu, Quanquan Li, Junjie Yan and Prof. Wanli Ouyang.

He used to work as a computer vision researcher in Sensetime Research (2016-2017). During this period, he was a member of the CUImage team that won the first place in ILSVRC2016 DET (ImageNet).

His research interests now include but not limit to object detection (both 2D and 3D), and 3D animation.

Buyu Li

PhD Candidate

MMLab, The Chinese University of Hong Kong

Yu Liu

General Manager & Senior Director, SenseTime
PI, Shanghai AI Lab
Industrial doctoral supervisor, SJTU
~~PhD @ Multimedia Lab, CUHK~~

liuyuisanai@gmail.com

Google Scholar
Resume
GitHub

News

- **Job positions and PhD program are open for talents interested in deep RL, super HUGE neural architecture, AutoML and Game AI.**
- [2022] My team have 9 papers published on ECCV/CoRL/NeurIPS/AAAI in 2022
- [2022] We release DI-Star, an implementation of AlphaStar in pyTorch, beating pro players with 6000+ MMR.
- [2021] My team won the best paper of ICCV21 MFR workshop
- [2021] My team won 3 championships of ICCV21 The Masked Face Recognition Challenge
- [2021] We release OpenDILab, an open source decision intelligence platform
- [2020] 5 papers with 1 oral got published on CVPR/ECCV 2020

王晓刚
港中文副教授
商汤研究院院长

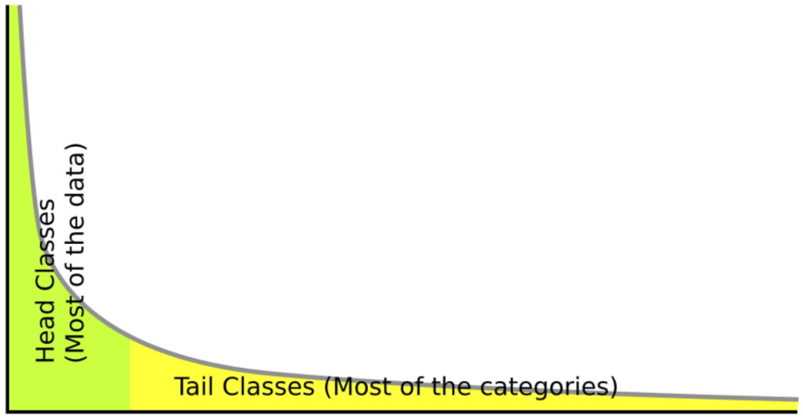
- 作者介绍
- 研究背景
- 本文方法
- 实验效果
- 总结反思

Long-Tailed Distribution

任务：分类、
目标检测、
实例分割

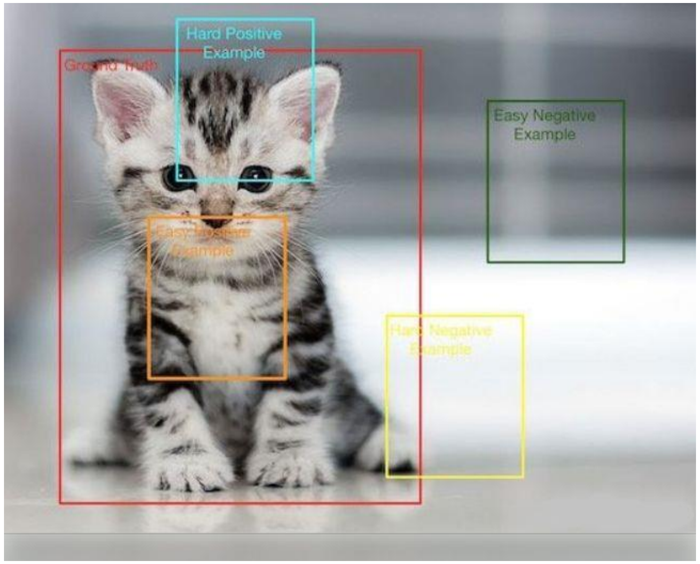
特点：少数类别样本数量庞大，
多数类别样本数量稀少

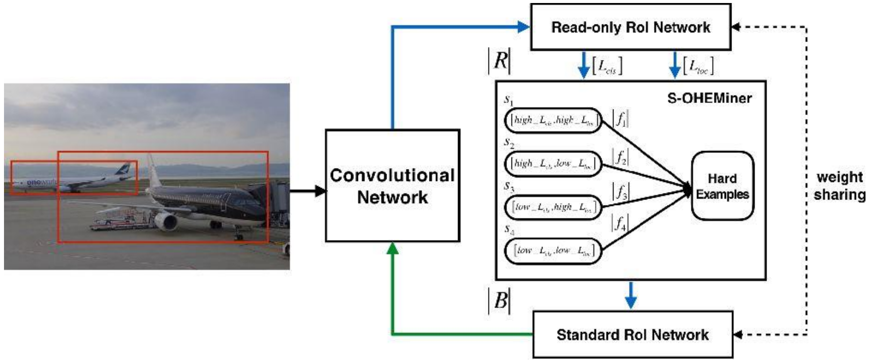
问题：数据分布先验影响分类精度
偏向头部类别的误分类
尾部类别样本量过少，难以建模

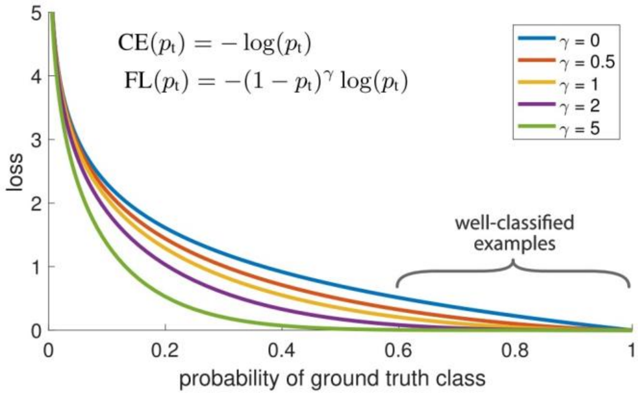


Data Imbalance in OD

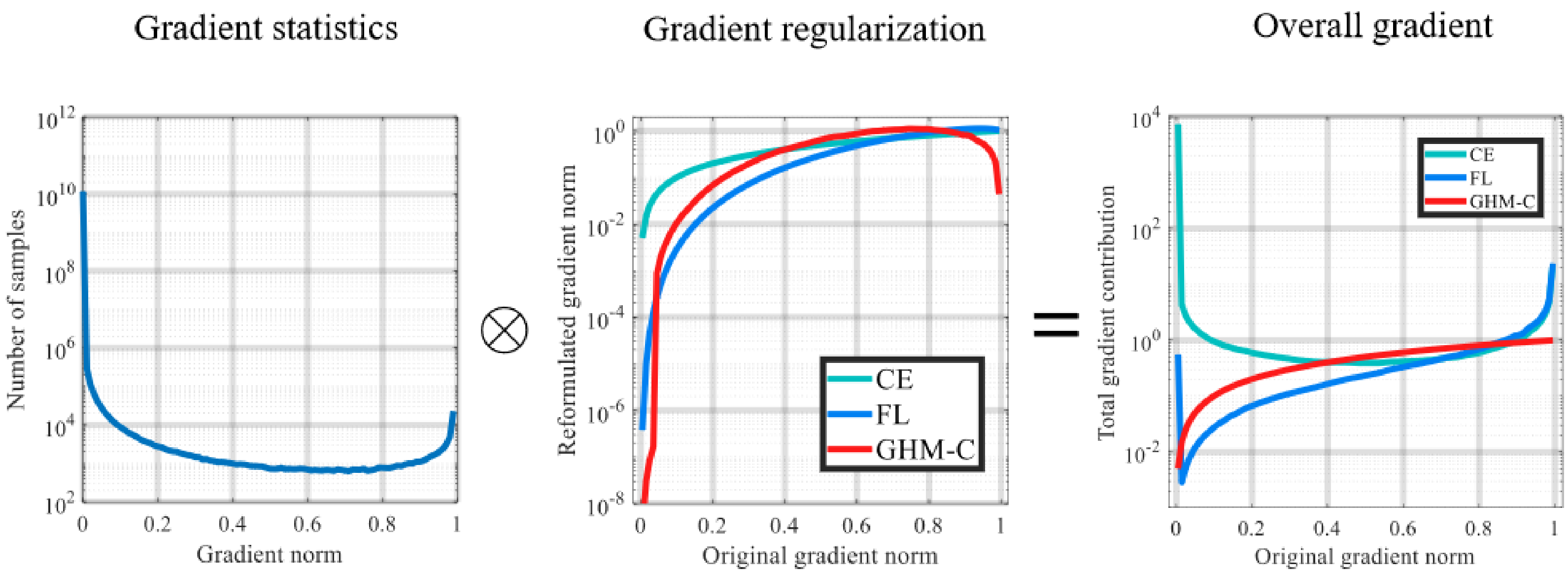
- Data Imbalance in OD
  - OHEM/S-OHEM
  - FL







Motivation



梯度->样本对网络的更新作用
梯度分布极为不均衡

简单样本的梯度数值小但数量庞大
总梯度累积量，CE和FL表现均欠佳

- 作者介绍
- 研究背景
- 本文方法
- 实验效果
- 总结反思

Gradient Harmonizing Mechanism

Overview

简单样本，没啥大用
累积效应还会影响分类结果

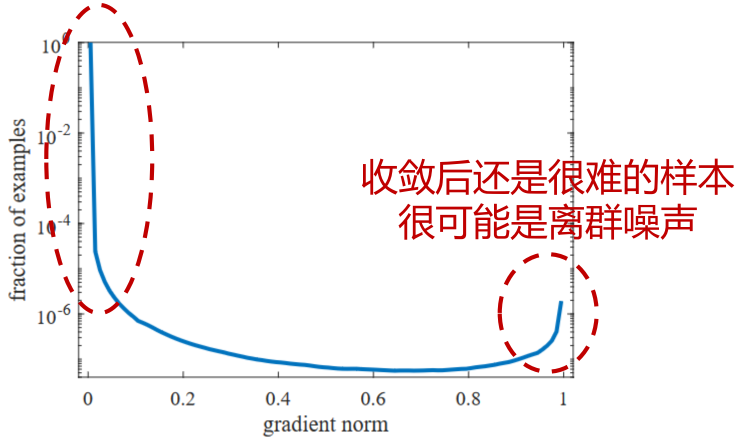


Figure 2: The distribution of the gradient norm g from a converged one-stage detection model. Note that the y-axis uses log scale since the number of examples with different gradient norm can differ by orders of magnitude.

按梯度分布统计 -> 重加权

$$GD(g) = \frac{1}{l_\epsilon(g)} \sum_{k=1}^{N} \delta_\epsilon(g_k, g)$$

小邻域指示函数 $\delta_\epsilon(x,y) = \begin{cases} 1 & \text{if } y - \frac{\epsilon}{2} <= x < y + \frac{\epsilon}{2} \\ 0 & \text{otherwise} \end{cases}$

小邻域长度 $l_\epsilon(g) = \min(g + \frac{\epsilon}{2}, 1) - \max(g - \frac{\epsilon}{2}, 0)$

GHM权重 $$\beta_i = \frac{N}{GD(g_i)} = \frac{1}{GD(g_i)/N}$$

Gradient Harmonizing Mechanism

GHM-C Loss

$$L_{CE}(p,p^*) = \begin{cases} -\log(p) & \text{if } p^*=1 \\ -\log(1-p) & \text{if } p^*=0 \end{cases}$$

$$\frac{\partial L_{CE}}{\partial x} = \begin{cases} p-1 & \text{if } p^*=1 \\ p & \text{if } p^*=0 \end{cases} = p - p^*$$

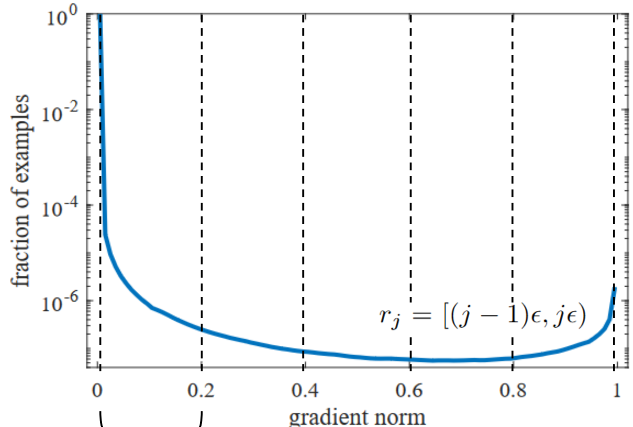
$$g = |p-p^*| = \begin{cases} 1-p & \text{if } p^*=1 \\ p & \text{if } p^*=0 \end{cases}$$

$$\hat{GD}(g) = \frac{R_{ind(g)}}{\epsilon} = R_{ind(g)}M$$

$$\hat{\beta}_i = \frac{N}{\hat{GD}(g_i)}$$

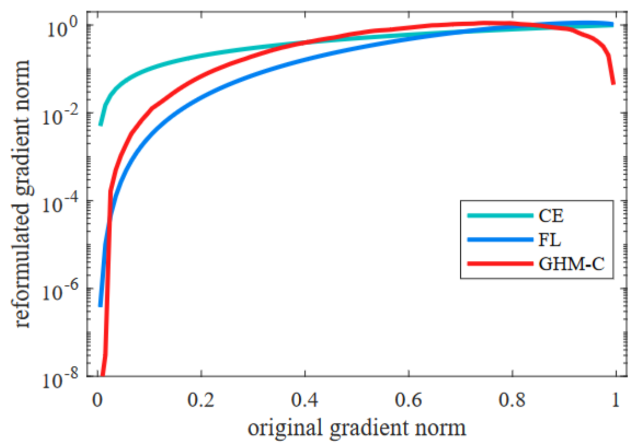
$$\hat{L}_{GHM-C} = \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}_i L_{CE}(p_i,p_i^*)$$

$$= \sum_{i=1}^{N}\frac{L_{CE}(p_i,p_i^*)}{\hat{GD}(g_i)}$$



ϵ

$$S_j^{(t)} = \alpha S_j^{(t-1)} + (1-\alpha)R_j^{(t)}$$



Gradient Harmonizing Mechanism

GHM-L Loss

$$L_{reg} = \sum_{i \in \{x,y,w,h\}} SL_1(t_i - t_i^*)$$

$$SL_1(d) = \begin{cases} \frac{d^2}{2\delta} & \text{if } |d| <= \delta \\ |d| - \frac{\delta}{2} & \text{otherwise} \end{cases}$$

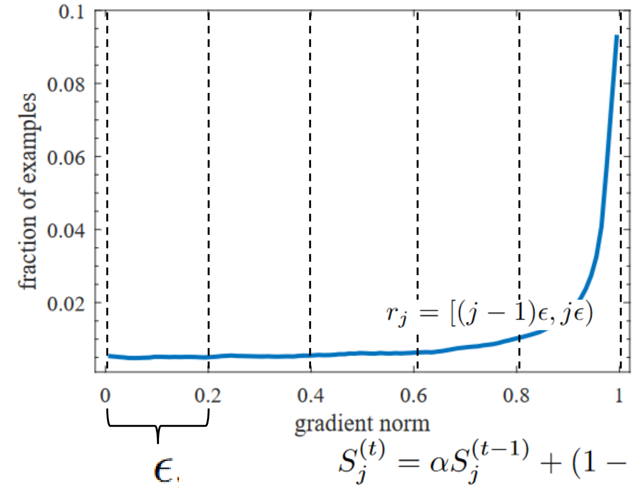
$$\frac{\partial SL_1}{\partial t_i} = \frac{\partial SL_1}{\partial d} = \begin{cases} \frac{d}{\delta} & \text{if } |d| <= \delta \\ sgn(d) & \text{otherwise} \end{cases}$$

$$ASL_1(d) = \sqrt{d^2 + \mu^2} - \mu$$

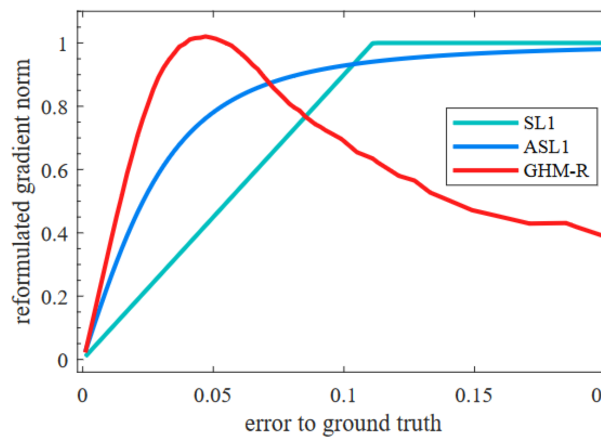
$$L_{GHM-R} = \frac{1}{N} \sum_{i=1}^{N} \beta_i ASL_1(d_i)$$

$$= \sum_{i=1}^{N} \frac{ASL_1(d_i)}{GD(gr_i)}$$

回归只在正样本上进行



$$S_j^{(t)} = \alpha S_j^{(t-1)} + (1-\alpha) R_j^{(t)}$$



- 作者介绍
- 研究背景
- 本文方法
- 实验效果
- 总结反思

SOTA Comparison

- COCO

method	network	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Faster RCNN (Ren et al. 2015)	FPN-ResNet-101	36.2	59.1	39.0	18.2	39.0	48.2
Mask RCNN (He et al. 2017)	FPN-ResNet-101	38.2	60.3	41.7	20.1	41.1	50.2
Mask RCNN (He et al. 2017)	FPN-ResNeXt-101	39.8	62.3	43.4	22.1	43.2	51.2
YOLOv3 (Redmon and Farhadi 2018)	DarkNet-53	33.0	57.9	34.4	18.3	35.4	41.9
DSSD513 (Fu et al. 2017)	DSSD-ResNet-101	33.2	53.3	35.2	13.0	35.4	51.1
Focal Loss (Lin et al. 2017b)	RetinaNet-FPN-ResNet-101	39.1	59.1	42.3	21.8	42.7	50.2
Focal Loss (Lin et al. 2017b)	RetinaNet-FPN-ResNeXt-101	40.8	61.1	44.1	**24.1**	44.2	51.2
GHM-C + GHM-R (ours)	RetinaNet-FPN-ResNet-101	39.9	60.8	42.5	20.3	43.6	54.1
GHM-C + GHM-R (ours)	RetinaNet-FPN-ResNeXt-101	**41.6**	**62.8**	**44.2**	22.3	**45.1**	**55.3**

Table 3: Comparison with state-of-the-art methods (single model) on COCO *test-dev* set.

Ablation Study

- 单位梯度区间 M

M	AP	$AP_{.5}$	$AP_{.75}$	AP_S	AP_M	AP_L
5	33.4	51.7	35.6	18.6	36.8	45.7
10	34.6	53.9	36.5	19.5	37.1	46.1
20	35.2	54.4	36.9	19.4	38.4	46.3
30	**35.8**	**55.5**	**38.1**	**19.6**	**39.6**	**46.7**
40	35.4	54.8	36.3	19.5	38.5	46.3

Table 1: Results of varying number of unit regions for GHM-C loss.

- GHM-C

method	AP	$AP_{.5}$	$AP_{.75}$	AP_S	AP_M	AP_L
CE	28.6	43.3	30.7	11.4	30.7	40.7
OHEM	31.1	47.2	33.2	-	-	-
FL	35.6	**55.6**	**38.2**	19.1	39.2	46.3
GHM-C	**35.8**	55.5	38.1	**19.6**	**39.6**	**46.7**

Table 4: Comparison of other loss functions. Note that the 'OHEM' is trained with ResNet-101 while others are trained with ResNet-50.

Ablation Study

- GHM-L

method	AP	$AP_{.5}$	$AP_{.75}$	AP_S	AP_M	AP_L
SL_1	35.8	**55.5**	38.1	19.6	39.6	46.7
ASL_1	35.7	55.0	38.1	19.7	39.7	45.9
GHM-R	**36.4**	54.6	**38.7**	**20.5**	**40.6**	**47.8**

Table 5: Comparison of different loss functions for regression.

method	AP	$AP_{.5}$	$AP_{.6}$	$AP_{.7}$	$AP_{.8}$	$AP_{.9}$
SL_1	35.8	**55.5**	51.2	43.4	31.4	11.9
ASL_1	35.7	55.0	51.1	43.5	31.5	12.1
GHM-R	**36.4**	54.6	**51.4**	**44.0**	**32.2**	**13.1**

Table 6: Comparison of AP at different IoU thresholds.

- 两阶段

method	AP	$AP_{.5}$	$AP_{.75}$	AP_S	AP_M	AP_L
SL_1	36.4	58.7	38.8	21.1	39.6	47.0
GHM-R	**37.4**	**58.9**	**39.9**	**21.8**	**40.8**	**48.8**

Table 7: Comparison of regression loss functions on two-stage detector.

- 作者介绍
- 研究背景
- 本文方法
- 实验效果
- 总结反思

总结反思

- 总结
  - 同时解决不均衡问题和噪声样本的问题
  - 从梯度分布的统计结果进行重加权

- 思考
  - 回归与分类问题，梯度分布有较大差别
  - 将简单类舍弃，相当于改变了输入数据的分布，是否会影响网络对简单类的建模，降低多数类的精度

Thanks for Attention!