

在目标检测被“遗忘”领域进行探索后，百度开源最新力作UMOP：即插即用、无痛涨点

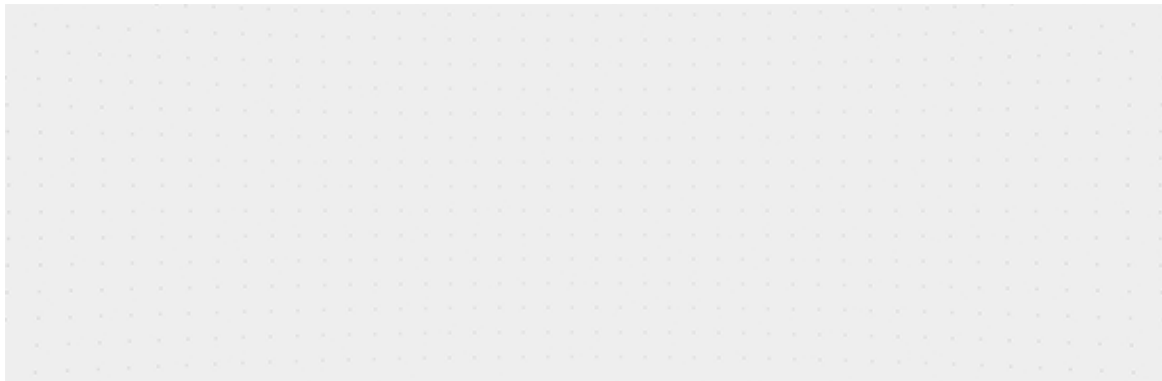
原创 CV开发者都爱看的 极市平台 2021-09-17 22:00:28 手机阅读 𠄎

收录于话题

#目标检测

31个

↑ 点击蓝字 关注极市平台



作者 | happy

编辑 | 极市平台

极市导读

百度开源了一项即插即用的一阶段检测器模块，主要是对金字塔层级目标分布不平衡进行的探索，所提模型在COCO test-dev数据集上取得了截至目前最高的55.1AP指标(一阶段检测器)。>>加入极市CV技术交流群，走在计算机视觉的最前沿

Progressive Hard-case Mining across Pyramid Levels in Object Detection

**Binghong Wu, Yehui Yang*, Dalu Yang,
Junde Wu, Haifeng Huang, Lei Wang, Junwei Liu, Yanwu Xu**

Artificial Intelligence Group, Baidu Inc.
No.10 Xibeiwang East Road, Baidu Technology Park Building No.2
Haidian District, Beijing, China, 100193

文章链接：<https://arxiv.org/pdf/2109.07217.pdf>

代码链接：<https://github.com/zimoqingfeng/UMOP>

本文是百度的研究员在目标检测领域的最新力作，对一个被“遗忘”的地带(金字塔层级目标分布不平衡)进行了探索。首先，作者基于FPN架构对不同检测器的不同层级目标分布进行了分析并得出：不同层级目标分布确实存在Level Imbalance问题；基于所发现问题，从现有固定参数FocalLoss出发设计了一种新的PFL损失；与此同时，针对训练过程中难易样例的角色转换现象提出了渐进式动态调整超参的机制进行更进一步的改善。值得一提的是，所提UMOP一种“即插即用”、“无损涨点”的模块，它可以轻易的嵌入到现有一阶段检测器中并大幅提升检测性能，性能提升高达~1.5AP。不得不说，百度最近两年最detection方面的工作还是非常值得称道的，顶！

Abstract

在目标检测领域，多级预测(比如FPN、YOLO)与重采样技术(如Focal Loss、ATSS)极大的提升了一阶段检测器性能。然而，如何通过逐级优化特征金字塔提升性能尚未进行探索。我们发现：在训练过程中，不同金字塔层级的正负样例比例是不同的，即存在Level Imbalance，而这尚未被一阶段检测器解决。

为缓解Level Imbalance影响，本文提出一种UMOP(Unified Multi-level Optimization Paradigm)，它包含两个成分：

- 基于不同层级目标重采样考量，对每个层级特征添加独立分类损失进行监督
- 提出一种跨金字塔层级的难例挖掘损失，且不会引入额外的层级相关设置

基于所提即插即用UMOP机制，现有一阶段检测可以用更少的训练迭代次数取得~1.5AP指标提升，且不会导致额外的计算负担。所得最佳模型在COCO test-dev数据集上取得了55.1AP指标。

本文主要由以下三点贡献：

- 本文首次证实：FPN性能在某种程度上会受到Level Imbalanced影响；
- 集成与所提UMOP，现有一阶段检测器可以用更少的训练迭代次数取得~1.5AP指标提升，且不会导致额外的计算负担；
- 在COCO test-dev数据集上，最佳模型取得了截至目前最高的55.1AP指标(一阶段检测器)。

The Statistical on Level Imbalance

在目标检测领域，多级架构在一阶段检测器中被广泛应用并极大提升了检测性能。在SSD与YOLO系列中，多级预测框架通过提供更多的稠密候选显著提升了检测器性能。与SSD和YOLO不同，FP

N还引入了自顶而下的路径进行特征聚合。接下来，我们将从更广义视角对anchor-based与anchor-free检测的Level Imbalance问题进行确认。

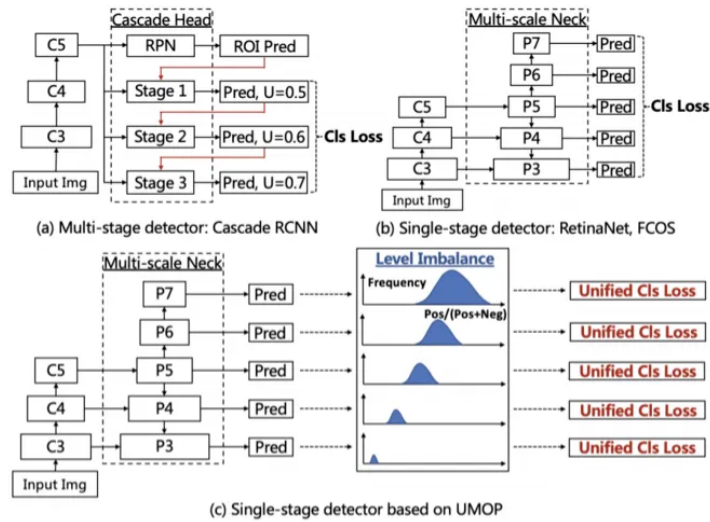


Figure 1: Differences and similarities. (a) Multi-stage detectors resample proposals with different IOU thresholds, shown as U in heads. (b) One-stage detectors utilize various pyramid levels for dense prediction, with a single reweighting loss form for all potentials. (c) UMOP optimizes all samples with a dynamic hyperparameter adjusting strategy, based on the convergence situation at which level they are.

The experiment settings on level imbalance 不失一般性，我们分别以两种检测器(Anchor-based: RetineNet, ATSS, Anchor-free: FCOS, VFNet)进行分析，我们以MSCOCO作为基准数据，对C3-C5添加FPN生成P3-P7五种不同分辨率特征，见上图b和c。与此同时，我们固定所有其他设置以确保公平对比。

The statistical analysis on level imbalance 在分析过程中，我们精心的对不同层级的目标分布影响进行分析，同时确保图像分辨率与模型复杂度固定。

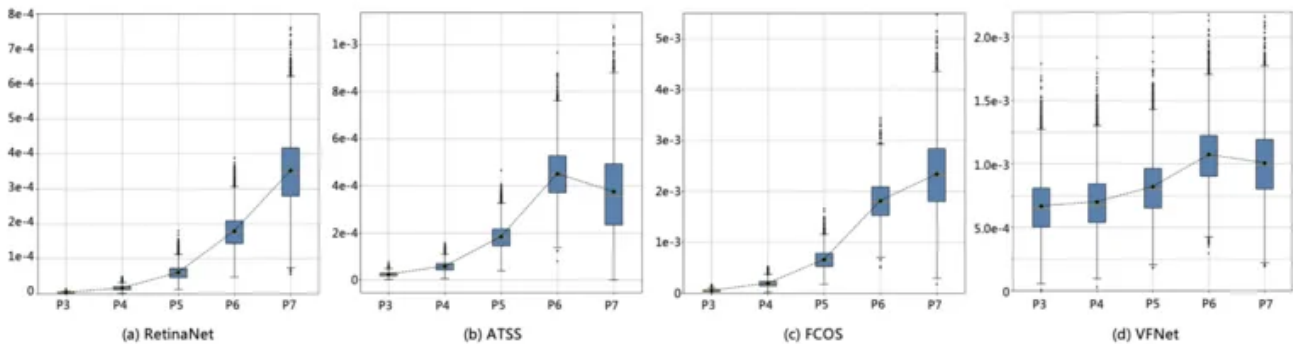


Figure 2: The statistical analysis on level imbalance. P3 to P5 is defined as the level index from FPN, indicating the results from different pyramid level predictions. In each training iteration, the proportions of positive samples to total samples in each pyramid level are recorded and summarized as box plots. The height of each box surround by upper quartile and lower quartile indicates the variance of the recorded proportions per pyramid level, and the mean values are drawn as dots and connected with dotted lines.

上图对比了不同检测器不同层级特征的目标分布，可以看到：

- 不同金字塔层级的目标分布完全不同；
- 训练过程中P7的每个候选具有更高概率匹配正样例。

上述结果清晰的表明了：**多阶段检测器存在Level Imbalance现象**。因此，我们怀疑：**带固定超参的FocalLoss不能同时对所有金字塔层级做好最好均衡**。

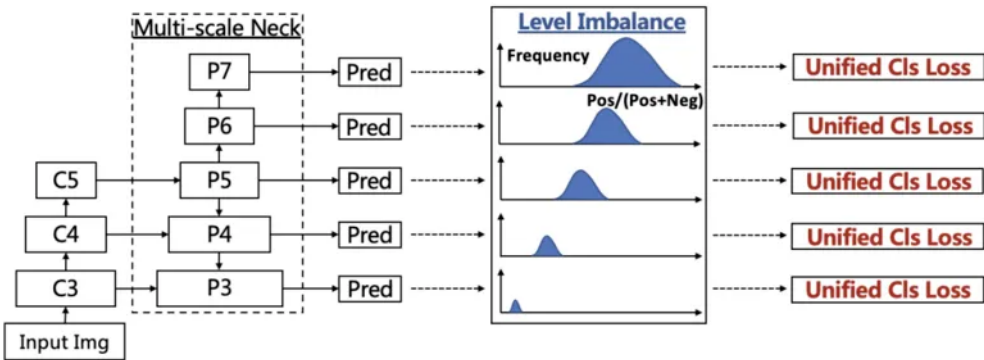
Unified Multi-level Optimization Paradigm

本文提出了UMOP以缓解一阶段检测器中的Level Imbalance问题。所提方法包含两部分：

- Level-wise Resampling Paradigm(LRP)，基于独立重采样考量，对每个金字塔层级设置不同的分类损失；
- Progressive Focal Loss(PFL)，它基于每个金字塔层级的正样例预测渐进的调整难例挖掘度。

Level-wise Resampling Paradigm

多阶段检测器通过迭代采样机制缓解不平衡现象，这种机制可以排查简单的低质采样，防止无匹配问题，进而提升预测质量。我们提出一种LRP，一种类似的框架，它基于灵活的难例度量提升一阶段检测器的高质量预测。



LRP是一种解决跨不同金字塔层级采样不平衡问题的多级优化框架，可见上图。在训练过程中，我们对每个金字塔层级独立的计算分类损失，定义如下：

$$Loss_{cls} = \frac{1}{L} \sum_{l=1}^L Loss_l(P_l, Y_l)$$

该损失赋予了每个金字塔层级足够灵活性以适应特定的采样所面临的不平衡性。

Progressive Focal Loss

与此同时，我们还提出了PFL以自动调节每个金字塔层级的难例挖掘自由度。对于二分类来说，Sigmoid Focal Loss定义如下：

$$FL(p_i, y_i) = \begin{cases} -\alpha(1 - p_i)^\gamma \log(p_i), & y_i = 1 \\ -(1 - \alpha)p_i^\gamma \log(1 - p_i), & y_i = 0 \end{cases}$$

在原始损失中， α 用于保持正负样例的梯度梯度， γ 则用于缓解正负样例不平衡。

从分而治之的角度出发，我们提出了一种新的动态超参调整策略，它可以按照每个金字塔层级的收敛位置动态调整上述超参。对于不同层级的不平衡性，真正需要的超参数 γ_i 完全不同的。

在训练阶段，早期的样例通常比较难训练，而在后期则相对容易。所提策略可以按照各自的收敛位置动态调整难例挖掘强度。所提PFL则满足上述特性，定义如下：

$$PFL(p_i, y_i) = \begin{cases} -\alpha_{ad}(1 - p_i)^{\gamma_{ad}} \log(p_i), & y_i = 1 \\ -(1 - \alpha_{ad})p_i^{\gamma_{ad}} \log(1 - p_i), & y_i = 0 \end{cases}$$

其中，作为难例挖掘的超参 γ_{ad} 与 α_{ad} 定义如下：

$$\gamma_{ad} = -\log\left(\frac{1}{n_{pos}} \sum_{i=1}^n y_i \cdot p_i\right) \alpha_{ad} = w/\gamma_{ad}$$

注： γ_{ad} 类似于交叉熵损失，它能够自然的反映不同金字塔层级的收敛位置。为确保训练稳定性，超参数被截断在有效区间 $\gamma_{ad} \in [\gamma - \delta, \gamma + \delta]$ ，其中 δ 为常熟。

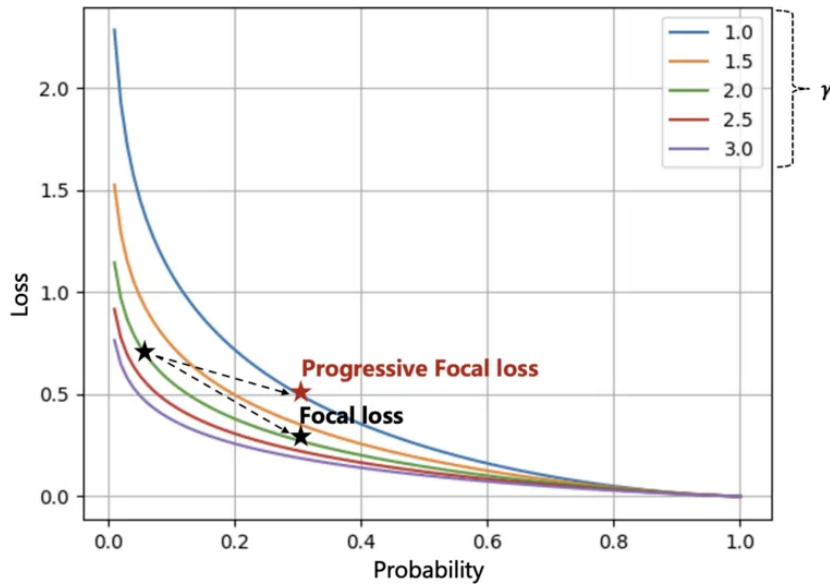


Figure 3: The comparison between Progressive Focal Loss and Focal Loss. During training, the cases are generally harder in the early training stage while relatively easier in the later. Progressive Focal Loss could strengthen the hard-case mining degree progressively.

如上图所示，PFL中的动态调整机制使得模型可以在早期聚焦难度，而当难例与简单样例无法判别时提升模型的判别能力。下图给出了UMOP的整个处理步骤，**UMOP**可以作为一种即插即用模块嵌入到带多级结构的任意一阶段检测器中。

Algorithm 1: Unified Multi-level Optimization Paradigm

Input:
P is a set of prediction results
Y is a set of ground truth corresponding to prediction
Output: $Loss_{cls}$ is the total classification loss

1: split P into $S_p = [P_1, P_2, \dots]$ by each pyramid level.
2: split Y into $S_y = [Y_1, Y_2, \dots]$ by each pyramid level.
3: **for** predictions $P_1 \in S_p$ and ground truths $Y_1 \in S_y$ **do**
4: calculate γ_{ad} by $y_i \in Y_1$ and $p_i \in P_1$ according to Eq. (4).
5: calculate α_{ad} by γ_{ad} according to Eq. (5).
6: calculate PFL_L according to Eq. (3) as the l^{th} level loss.
7: **end for**
8: calculate the $Loss_{cls}$ according to the Eq. (1)
9: **return** $Loss_{cls}$

Experiments

Method	Backbone	Size	Epoch	AP	AP ₅₀	AP ₇₅	AP _S	AP _M	AP _L
<i>multi-stage:</i>									
Faster R-CNN w/ FPN (Lin et al. 2017)	R-101	800*	24	36.2	59.1	39.0	18.2	39.0	48.2
Cascade R-CNN (Cai and Vasconcelos 2018)	R-101	800*	18	42.8	62.1	46.3	23.7	45.5	55.2
CenterNet2 (Zhou, Koltun, and Krähenbühl 2021)	X-101-DCN	800*	24	50.2	68.0	55.0	31.2	53.5	63.6
<i>one-stage:</i>									
CornerNet (Law and Deng 2018)	Hg-104	512	200	40.6	56.4	43.2	19.1	42.8	54.3
CenterNet (Zhou, Wang, and Krähenbühl 2019)	Hg-104	512	190	44.9	62.4	48.1	25.6	47.4	57.4
FASF (Zhu, He, and Savvides 2019)	X-101	800*	18	42.9	63.8	46.3	26.6	46.2	52.7
FCOS (Tian et al. 2019)	X-101	800*	24	44.7	64.1	48.4	27.6	47.5	55.6
ATSS (Zhang et al. 2020)	X-101-DCN	800*	24	47.7	66.5	51.9	29.7	50.8	59.4
EfficientDet (Tan, Pang, and Le 2020)	EffNet-B7	1536	450	55.1	74.3	59.9	37.2	57.9	68.0
<i>ours:</i>									
ATSS w/ UMOP	R2-101-DCN	960*	24	50.3	69.5	54.9	31.8	53.8	63.1
ATSS w/ UMOP	Swin-S-1K	960*	24	50.3	70.0	54.9	32.0	53.6	63.1
ATSS w/ UMOP	Swin-B-22K	960*	24	51.9	71.6	56.6	33.4	55.4	65.0
ATSS w/ UMOP	Swin-L-22K	960*	24	53.1	72.7	58.0	34.9	56.5	66.4
ATSS w/ UMOP (multi-scale testing)	Swin-L-22K	960*	24	55.1	74.2	60.7	38.1	58.6	66.8

Table 3: Single-model performance comparison with state-of-the-art detectors on COCO test-dev, 'R': ResNet, 'X': ResNeXt-64x4d, 'R2': Res2Net, 'Hg': Hourglass, 'EffNet': EfficientNet 'DCN': Deformable convolution network v2, 960*: resize the shorter side to 960 and the longer side less or equal to 1333 with the aspect ratio kept, MS_{train}: training image scale range 1333×[640:800] for 800*, and 1333×[480:960] for 960*.

上图给出ATSS+UMOP与其他检测器在COCO test-dev数据集上的性能对比，从中可以看拿到：相比其他需要长训练周期、大图像分辨率的高性能检测器，基于常规实验设置+常见图像分辨率，所提**ATSS+UMOP**即可取得了**55.1AP**指标，大幅优于其他检测器。

Ablation Study

Method	Backbone	w/ PFL	w/ LRP	AP
RetinaNet	R-50			35.7
RetinaNet	R-50	✓		36.7
RetinaNet	R-50	✓	✓	36.9
RetinaNet w / imprv.	R-101			38.9
RetinaNet w / imprv	R-101	✓		39.7

RetinaNet w / imprv.	R-101	✓	✓	40.5
ATSS	R-50			39.3
ATSS	R-50	✓		40.1
ATSS	R-50	✓	✓	40.4
ATSS w / imprv.	R-101			46.1
ATSS w / imprv.	R-101	✓		46.7
ATSS w / imprv.	R-101	✓	✓	47.6
ATSS w / imprv.	X-101			47.7
ATSS w / imprv.	X-101	✓		48.4
ATSS w / imprv.	X-101	✓	✓	48.8

Table 1: Average precision (AP) improvements on COCO *minival*. 'R': ResNet, 'X': ResNeXt-64x4d. We show the performance improving when only applying PFL and the whole method. For a stronger baseline, we optionally applied multi-scale training and deformable convolutional layers on both detectors.

上图说明所提方案的泛化性：可以作为即插即用模块轻易嵌入到几乎所有一阶段检测中并提升性能。

- 对于RetinaNet-R50，UMOP可以将其性能从35.7提升到36.9；
- 对于RetinaNet-R101，UMLP可以将其性能从38.9提升到40.5；
- 对于ATSS-R50，UMOP可以将其性能从39.3提升到40.4；
- 对于ATSS-R101，UMOP可以将其性能从46.1提升到47.6；



Figure 4: Some detection results on COCO *minival*. ResNet-50 is used as the backbone and the score threshold for visualization is 0.3. As shown in the figure, UMOP works well with a wide range of objects including crowded, highly overlapped, and extremely small objects.

上图给出了基于UMOP训练的ATSS与常规训练的检测效果对比，所提方案的泛化性极好，可以解决模糊目标检测、遮挡目标、小目标、极限比例目标等常见难例问题。

Method	Backbone	Epoch	AP	AP ₅₀	AP ₇₅	AP _s	AP _m	AP _l
ATSS	ResNet-101-DCN	24	46.3	64.7	50.4	27.7	49.8	58.4
ATSS w/ UMOP		18	47.7	66.9	52.1	29.1	51.0	59.7
ATSS	ResNeXt-101-64x4d-DCN	24	47.7	66.5	51.9	29.7	50.8	59.4
ATSS w/ UMOP		18	49.1	68.5	53.6	30.8	52.6	61.1

Table 2: Convergence speed analysis on COCO *test-dev*, our method achieves a better performance with fewer iterations

上表对所提方案的收敛性进行了分析对比，从中可以看到：原始ATSS需要24epoch达到收敛，而装备上UMOP后仅需18epoch即可取得更高的指标，性能提升1.4dB。

本文亮点总结

1.为缓解Level Imbalance影响，本文提出一种UMOP(Unified Multi-level Optimization Paradigm)，它包含两个成分：

- 基于不同层级目标重采样考量，对每个层级特征添加独立分类损失进行监督
- 提出一种跨金字塔层级的难例挖掘损失，且不会引入额外的层级相关设置

2.本文主要由以下三点贡献：

- 本文首次证实：FPN性能在某种程度上会受到Level Imbalanced影响；
- 集成与所提UMOP，现有一阶段检测器可以用更少的训练迭代次数取得~1.5AP指标提升，且不会导致额外的计算负担；
- 在COCO test-dev数据集上，最佳模型取得了截至目前最高的55.1AP指标(一阶段检测器)。

如果觉得有用，就请分享到朋友圈吧！



极市平台

专注计算机视觉前沿资讯和技术干货，官网：www.cvmart.net
624篇原创内容

公众号

△点击卡片关注极市平台，获取最新CV干货

公众号后台回复“CVPR21检测”获取CVPR2021目标检测论文下载~

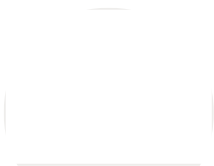
极市干货

深度学习环境搭建：如何配置一台深度学习工作站？

实操教程：OpenVINO2021.4+YOLOX目标检测模型测试部署 | 为什么你的显卡利用率总是0%？

算法技巧 (trick)：图像分类算法优化技巧 | 21个深度学习调参的实用技巧

极市平台签约作者



happy
知乎：AIWalker

AIWalker运营、CV技术深度Follower、爱造各种轮子
研究领域：专注low-level，对CNN、Transformer、MLP等前沿网络架构保持学习心态，倾心于AI技术产品化。
公众号：AIWalker

作品精选

- 吊打一切现有版本的YOLO！旷视重磅开源YOLOX：新一代目标检测性能速度担当！
- YOLOv4团队开源最新力作！1774fps、COCO最高精度，分别适合高低端GPU的YOLO
- 图像增强领域大突破！以1.66ms的速度处理4K图像，港理工提出图像自适应的3DLUT



投稿方式：

添加小编微信Fengcall（微信号：fengcall19），备注：姓名-投稿



△长按添加极市平台小编

觉得有用麻烦给个在看啦~

阅读原文

喜欢此内容的人还喜欢

VideoCLIP-Facebook&CMU开源视频文本理解的对比学习预训练，性能SOTA！

我爱计算机视觉

这个开源项目用Pytorch实现了17种强化学习算法

开源前线

阿里巴巴开源“KNAS”：一种无需训练即可评估给定架构的自动化机器学习算法

相约机器人
