YOLOX:在2021超越yolo系列

1. 要解决什么问题

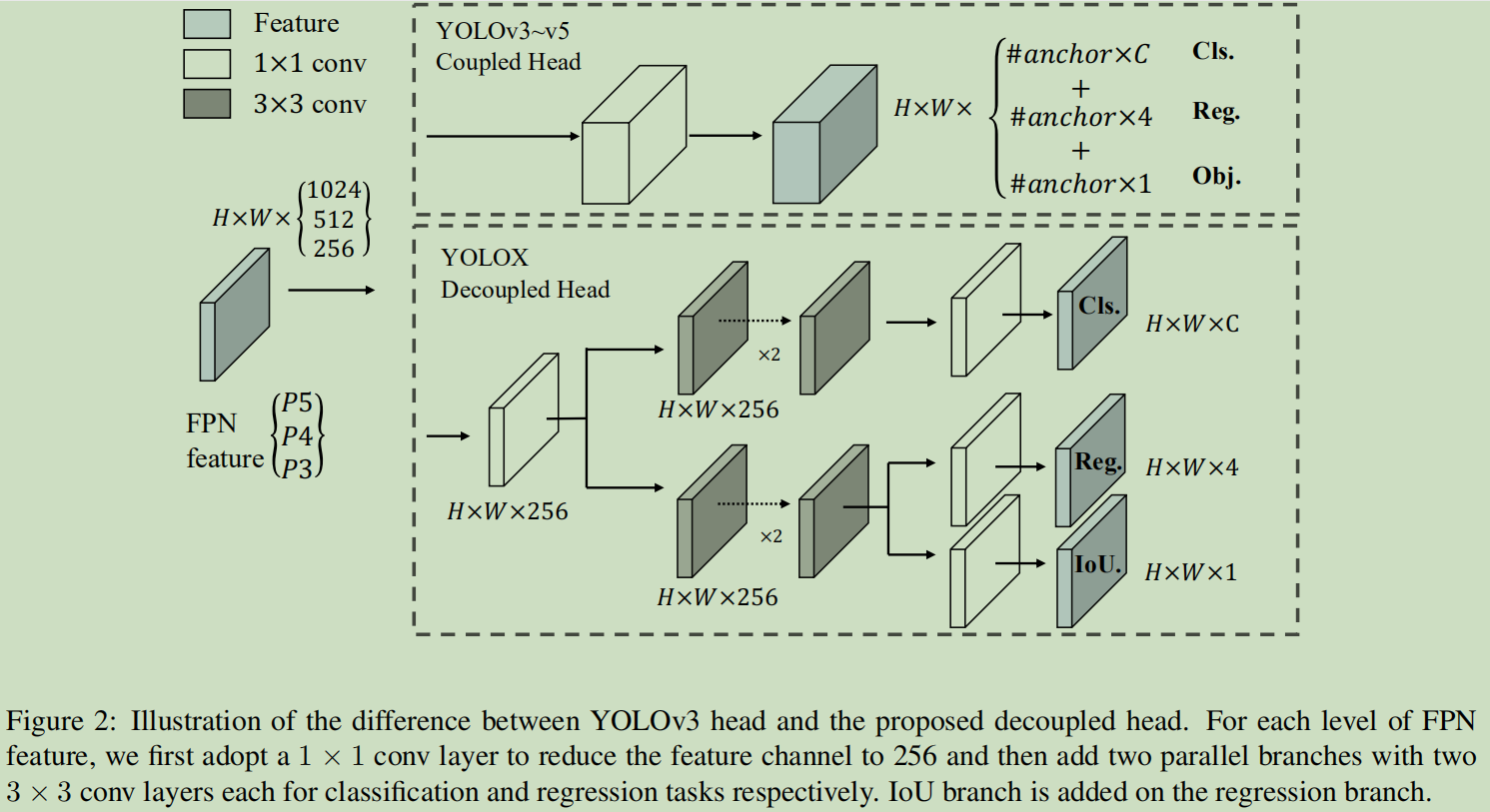
将yolo系列算法从anchor-base转为anchor-free，做法参考的是FCOS。

增加了一系列技术优化，将模型提升到state-of-the-art。

1. 用了什么方法

使用yolov3的主干。对训练策略进行了微调：添加了EMA权值更新，cosine学习率机制，IoU损失，IoU感知分支。采用BCE损失训练cls和obj分支，IoU损失训练reg分支。此外，添加了RandomHorizontalFlip、ColorJitter以及多尺度训练，不使用RandomResizedCrop。

Decoupled head。classification和 regression 是有冲突的，但是yolo系列工作还是将两者放在一个head里面。Yolox使用解耦的检测头，性能得到提升，速度略微下降。



强大的数据增强。使用Mosaic 和 MixUp来提升yolox的性能,但是在最后的15个epoch中关闭。由于加入了这么强的两个数据增强后，YOLOX从这里开始就不再需要pre-trained mode了，而是直接train from scratch。

Anchor-free。Anchor机制存在诸多问题：（1）为了获得最优检测性能，需要在训练之前进行聚类分析以确定最佳anchor集合，这些anchor集合存在数据相关性，泛化性能较差；（2）anchor机制提升了检测头的复杂度和每张图片预测结果的数量，这在边缘设备上可能存在瓶颈。

Yolox去掉了anchor-box，每个grid只预测一个目标，同时参考FCOS，预先设定一个尺寸范围，根绝每个gt的size来判断应该分配到哪个尺度上去。

Multi positives。为了确保与yolov3的一致性，前述的anchor-free版本仅仅对每个目标赋予一个正样本，而忽略了其他高质量预测，于是参考FCOS，简单的赋予中心区域为正样本，增加正样本数量。

SimOTA。为什么 Anchor Free 现在可以上 YOLO ，并且性能不降反升，这与样本匹配有密不可分的联系。与 Anchor Free 比起来，样本匹配在业界似乎没有什么关注度。但是一个好的样本匹配算法可以天然缓解拥挤场景的检测问题（ LLA、OTA 里使用动态样本匹配可以在 CrowdHuman 上提升 FCOS 将近 10 个点），缓解极端长宽比的物体的检测效果差的问题，以及极端大小目标正样本不均衡的问题。甚至可能可以缓解旋转物体检测效果不好的问题，这些问题本质上都是样本匹配的问题。在我们的认知中，样本匹配有 4 个因素十分重要：

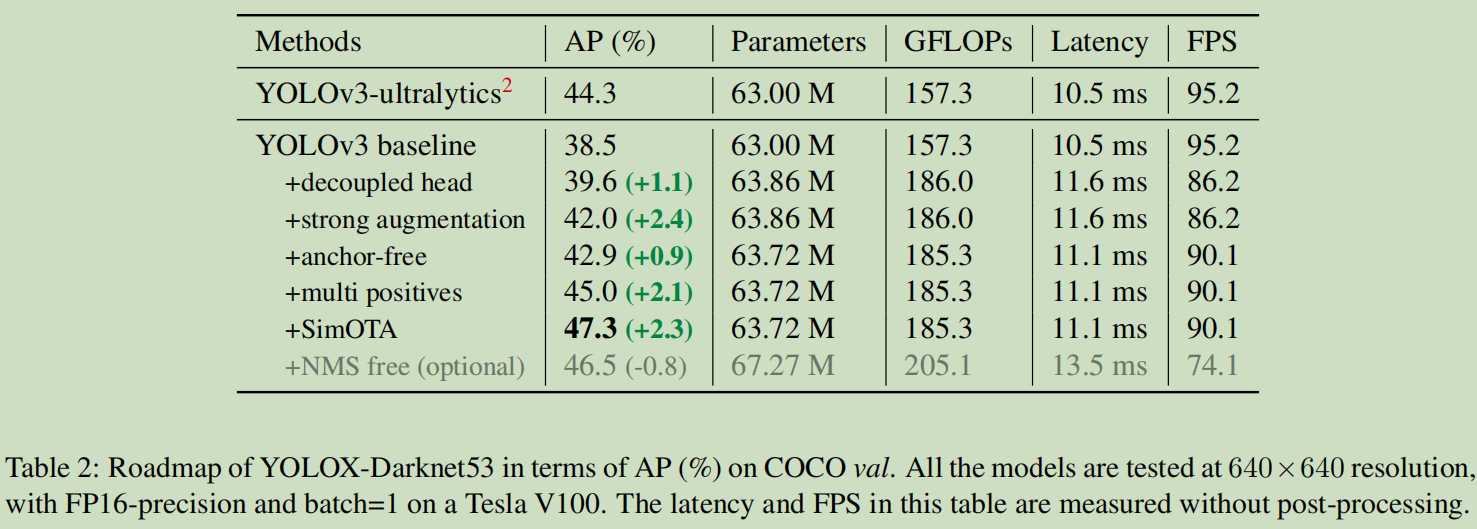
1) loss/quality/prediction aware ：基于网络自身的预测来计算 anchor box 或者 anchor point 与 gt 的匹配关系，充分考虑到了不同结构/复杂度的模型可能会有不同行为，是一种真正的 dynamic 样本匹配。而 loss aware 后续也被发现对于 DeTR 和 DeFCN 这类端到端检测器至关重要。与之相对的，基于 IoU 阈值 /in Grid(YOLOv1)/in Box or Center(FCOS) 都属于依赖人为定义的几何先验做样本匹配，目前来看都属于次优方案。

2) center prior : 考虑到感受野的问题，以及大部分场景下，目标的质心都与目标的几何中心有一定的联系，将正样本限定在目标中心的一定区域内做 loss/quality aware 样本匹配能很好地解决收敛不稳定的问题。

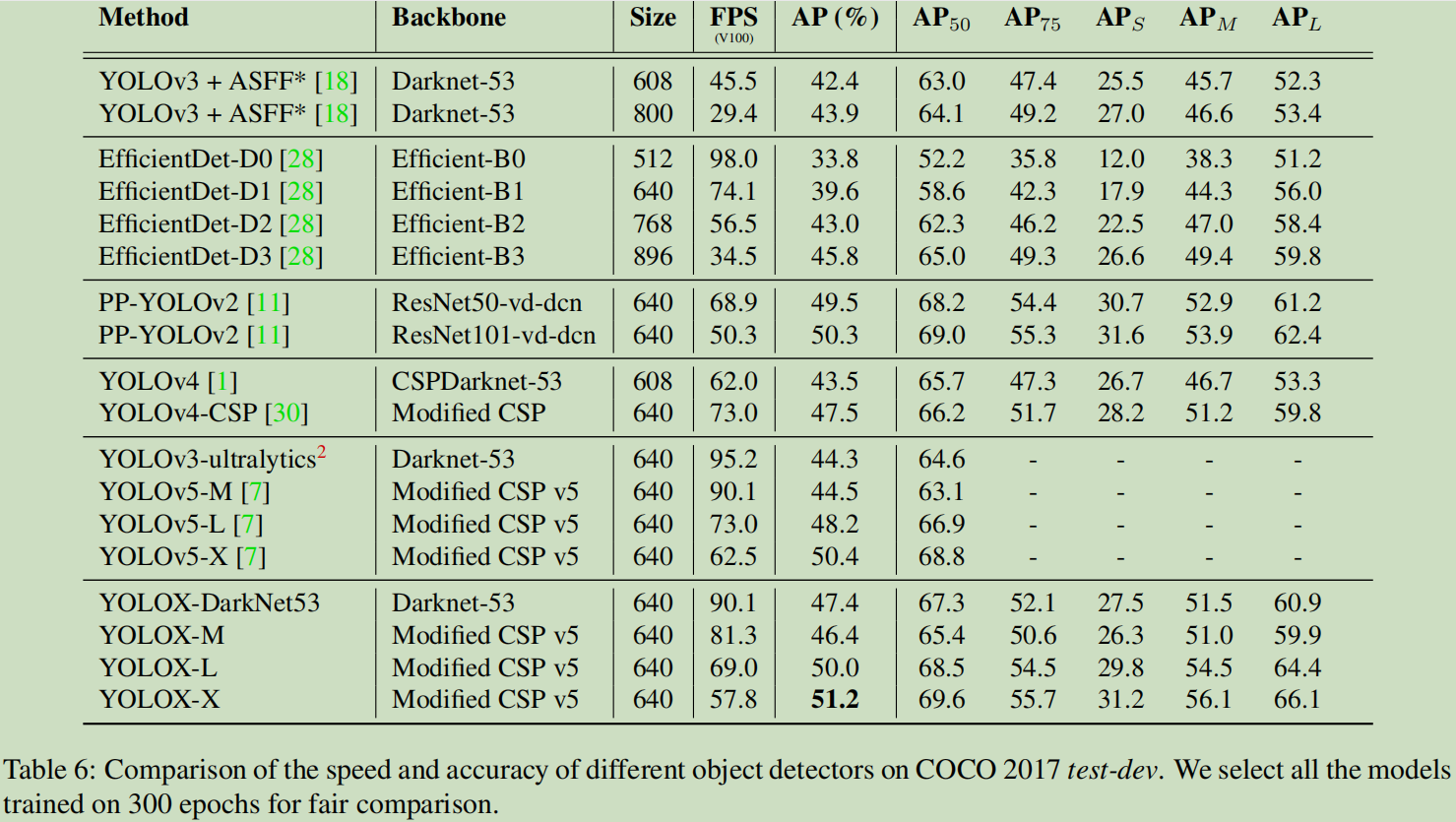
3) 不同目标设定不同的正样本数量( dynamic k )：我们不可能为同一场景下的西瓜和蚂蚁分配同样的正样本数，如果真是那样，那要么蚂蚁有很多低质量的正样本，要么西瓜仅仅只有一两个正样本。Dynamic k 的关键在于如何确定k，有些方法通过其他方式间接实现了动态 k ，比如 ATSS、PAA ，甚至 RetinaNet ，同时，k的估计依然可以是 prediction aware 的，我们具体的做法是首先计算每个目标最接近的10个预测，然后把这个 10 个预测与 gt 的 iou 加起来求得最终的k，很简单有效，对 10 这个数字也不是很敏感，在 5~15 调整几乎没有影响。

4) 全局信息：有些 anchor box/point 处于正样本之间的交界处、或者正负样本之间的交界处，这类 anchor box/point 的正负划分，甚至若为正，该是谁的正样本，都应充分考虑全局信息。

OTA 充分考虑到了以上 4 点，通过把样本匹配建模成最优传输问题，求得了全局信息下的最优样本匹配方案。但是 OTA 最大的问题是会增加约 20~25 %的额外训练时间，对于动辄 300epoch 的 COCO 训练来说是有些吃不消的，此外 Sinkhorn-Iter 也会占用大量的显存，所以在 YOLOX 上，我们去掉了 OTA 里的最优方案求解过程，保留上面 4 点的前 3 点，简而言之： loss aware dynamic top k。由于相对 OTA 去掉了Sinkhorn-Iter 求最优解的过程，我们把 YOLOX 采用的样本匹配方案称为 SimOTA ( Simplified OTA )。没有复杂的数学公式和原理，不增加额外的计算时间，但是有效。使用SimOTA之后，FCOS样本匹配阶段的FPN分层就被取消了，匹配(包括分层)由SimOTA自动完成。



1. 效果如何



1. 还存在什么问题和可借鉴的地方

解耦head的精度提升也有可能是解耦后又添加了几个卷积有关，毕竟计算量增加了。

使用数据增强后，在最后的几个epoch时关闭。可以想象，Mosaic+Mixup 生成的训练图片，远远脱离自然图片的真实分布，并且 Mosaic 大量的 crop 操作会带来很多不准确的标注框。因此该操作有助于增强模型性能。

参考FASF，分支添加anchor-base部分进行联合训练精度是否会提升一些，哪怕inference的时候把anchor-base部分去掉，相当于anchor-base部分当成一个辅助loss部分，这样精度因该也会提升一些，至少误检会少一些。