

背景介绍

难点一：大中小目标分布不均

在遥感数据中

1. 小目标数量众多，排列密集或稀疏，中大目标的数量较少
2. 由于遥感图像需要进行切割，因此会加剧目标分布不均的情况。

在YOLOv5中，利用Mosaic数据增强方式可以有效解决这个问题。**Mosaic** 的拼接特性能过极大程度上改善 **DOTA** 数据集目标分布不均的情况，并且由于其拼接的随机性，随着训练时间的加长，改善效果会更明显。

难点二：目标尺度变化剧烈

大部分目标的尺度较小，提供的细节信息有限，且随着网络的加深，小目标的细节信息会逐渐丢失，因此我们需要网络结构对小目标有一定敏感性，即较强的多尺度特征提取能力。高空场景下，目标尺度时而密集，时而稀疏，模型需要满足单个图像中检测多个不同尺度多个对象的能力。

因此需要考虑选取合适的 featuremap 的结果（Backbone + Neck），既要保证丰富的浅层位置信息同时保证深层的语义信息。yolov5 的 SPP + PANet 对这个问题有很好的处理

难点三：目标长宽比大，排列密集

在遥感场景下，很多目标的长宽比较大，且排列密集，因此遥感目标更适合利用旋转框进行检测。

原始YOLOv5项目的应用场景为自然场景下的目标，目标检测边框为水平矩形框因此利用了改建后的YOLOv5，增加了一个预测角度的分支来进行旋转目标的检测。

损失计算：置信度损失、class分类损失、 θ 角度分类损失、bbox边框回归损失。

- class分类损失：无需更改，注意数据索引部分即可。
- θ 角度分类损失：由于我们添加的 θ 是分类任务，照葫芦画瓢，添加分类损失就可以了，
- bbox边框回归损失：边框回归损失部分依旧采用IOU/GIOU/CIU/DIOU损失函数。
- 置信度损失：置信度分支的权重系数依然选择水平边框的之间的IOU/GIOU/CIU/DIOU；

所做的工作

1. 对数据进行处理

1. 对标注格式的转换，将数据格式转换为YOLO的方式
 2. 对数据进行裁剪
 3. 对裁剪后的数据进行清洗，由于裁剪后的数据很多没有目标，因此遍历标签将不含有目标的数据进行剔除
2. 在进行训练的过程中，进行多尺度训练，按照两个不同的比例对图片进行裁剪，然后进行训练，这样的训练方式对最终的mAP提升效果不显著，只有零点几个点，且大幅增加了训练时间，因此在后期放弃了这个训练策略。
3. 进行多尺度训练检测，在进行检测的过程中，进行多尺度检测，将图片以两种不同的比率进行裁剪，进行检测，map 上升了 3 个百分点，其中小目标（车辆）的 AP 上升了 8 个百分点

最终的结果

达到了 68% 的 mAP，但是其中的大目标，如足球场，桥梁等较大的目标 AP 较低，只有 50% 左右，分析原因，一是在裁剪的过程中，将这些目标裁掉了，二是这些大目标在数据集中的数量占比较少，对 loss 的贡献较少。但是总的来说，这个结果还是较好的。

自我介绍

尊敬的面试官您好，我是胡森康，目前是北京理工大学电子信息工程的大四学生，已推荐免试至中国科学院大学空天信息研究院攻读博士研究生，从事计算机视觉和遥感信息智能处理方向的研究。下面我将从以下几个方面进行自我介绍。

首先是学习能力方面，在大学学习期间，我的 GPA 常年保持在 90 分以上，排名一直在 top 5%，并获得了教育部国家奖学金，CASC 奖学金，北京理工大学一等奖学金，北京理工大学“优秀学生标兵”等12项荣誉。

其次是实践能力和团队合作方面，在大学期间，我参与了很多的竞赛和项目，包括全国大学生数学建模竞赛，FPGA设计竞赛，电子设计竞赛，“挑战杯”等，并获得了10余项奖项。在参与这些竞赛的过程中，我一直作为队长带领队员进行实践，在团队合作中推进项目和竞赛的进展。

另外是科研能力方面，在本科期间我以第一作者发表两篇论文，以第二作者发表一篇论文，这在本科生中是较为突出的。

最后我介绍一下我为什么要申请这个岗位。人类文明发展到现在经历了三次工业革命，分别是以蒸汽机为代表的第一次工业革命，以内燃机、电气化为代表的第二次工业革命，以信息技术为代表的第三次工业革命，我相信在不久的将来人类会迈进以智能技术为代表的第四次工业革命，这会极大的推进人类文明的发展，因此我觉得我们一定要积极参与到这次浪潮中来，所以我在保研的时候选择了人工智能方向的导师，从去年保研之后，我在中科院实习了半年多，参与了一些实验室的工作。现在我想利用最近这段时间在工业界从事一些相关的算法工作，与优秀的人一起工作，提高自己的水平，开阔自己的视野。

最后，假如我能加入像旷视这样优秀的企业进行实习，我想这一定是一个很难忘的经历。

深度学习面试常见问题

bias/variance trade off ?

ReLU 在零点不可导，那么在反向传播中如何处理？

可以在 ReLU 的零点人为的给它赋值一个倒数，比如 0 或者 1

ReLU 的优缺点

所有的负值都变为 0，而正值不变，这种操作被称为**单侧抑制**。正因为有了单侧抑制，才使得神经网络中的神经元也有了稀疏激活性。

使用 ReLU 得到的 SGD 的收敛速度会比 sigmoid/tanh 快很多，计算复杂度低，不需要进行指数运算。

训练的时候很脆弱，很容易“die”。同时 ReLU 不会对数据做幅度压缩，所以数据的幅度会随着模型层数的增加不断扩张

四种归一化

- BN, batch normalization
- LN, layer normalization
- IN, Instance normalization, 实例归一化
- GN, group normalization

什么是梯度消失和梯度爆炸

关键在于激活函数。在反向传播中，我们使用的是梯度下降策略来优化参数，在链式法则求导的时候，涉及到了对激活函数求导，关键在于反向传播时激活函数的导数。如果导数大于 1，那么随着网络层数的增加，梯度更新将会朝着指数爆炸的方式增加，这就是梯度爆炸。同样如果导数小于 1，那么对着网络层数的增加梯度更新信息会朝着指数衰减的方式减少，这就是梯度消失。

什么是梯度弥散，，造成的问题

梯度弥散就是梯度消失，导数为0。梯度弥散造成靠近输出层的隐藏层梯度大，参数更新快，所以很快就会收敛。靠近输入层的隐藏层梯度小，参数更新慢，几乎和初始状态一样，随机分布。

如何解决梯度消失问题

- 使用 ReLU, Leaky ReLU 激活函数替代 Sigmoid
- BN, 通过每一层的输出规范为均值和方差一致的方法, 消除了权重参数放大缩小带来的影响

什么是端到端学习 (end-to-end)

端到端学习是一种解决问题的思路, 与之对应的是多步骤解决问题, 依旧是将一个问题拆分为多个步骤分布解决, 而端到端是由输入端的数据直接得到输出端的结果。

Softmax 的原理是什么, 有什么作用

softmax 用于多分类过程中, 它将多个神经元的输出, 映射到 $(0, 1)$ 区间内, 可以看成概率来理解, 从而进行多分类。

softmax 直白来讲就是将原来输出是 3, 1, -3 通过 softmax 函数作用, 就映射为 $(0, 1)$ 的值, 而且这些值类和为 1 (满足概率的性质), 那么我们就可以将它理解为概率, 在最后选取输出节点的时候, 我们就可以选取概率最大 (也就是值对应最大) 的节点, 作为我们的预测目标。

softmax 的好处:

1. 好求导
2. 使得好结果和坏结果之间的差异更加显著, 更有利于学习。SVM 只选自己喜欢的男神, softmax 把多有备胎全都拉出来评分, 最后还归一化一下。

dropout 是否用在测试集上?

不能, 只能用于训练集, 是训练过程中的一种正则化技术

列举几个梯度下降的方法并简介

常见降维方法, 并简介

Transformer 面试知识

写出 self attention 的公式表达式

$$\text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

为什么式 (1) 中要对 QK 进行 scaled

让 softmax 输入的数据分布变好，数值进入梯度敏感区间，能防止梯度小时，让模型好训练

self-attention 一定要这样表达吗

不需要，能刻画相关性，相似性等建模方式都可以。最好速度快，模型好学，表达能力够

有其他方法不用除以 $\sqrt{d_k}$ 吗？

有，同上，只要能做到每层参数等梯度保持在训练敏感范围内，不要太大，不要太小，那么这个网络就不叫好训练

为什么transformer用Layer Norm？有什么用？

让神经网络各层参数输入数据分布变好，数值进入梯度敏感区间，能防止梯度消失，让模型好训练

为什么不用BN？

- NLP不定长，好多位置填0，影响其他样本非0参数的计算
- transforme模型比较大，batch size拉不大，容易变得不稳定

transformer为什么要用三个不一样的QKV

增强网络的容量和表达能力

为什么要多头

增强网络的容量和表达能力