****

**模式识别与机器学习实验报告**

**题目名称: 目标检测**

**作业内容： VOC数据集目标检测与分类**

**姓名与学号: 马彦祥 2018300150**

1.简介

目标检测在及你几年发展势头很猛，很多算法都是3年内提出的。我本次实验主要使用了faster rcnn、cascade rcnn[1]和可变卷积（dcn）版的casacde rcnn。经过查阅资料，我发现使用resnet骨干的cascade rcnn网络，在近几年的历年的CVPR UG2+比赛和上都取得了不错的成绩，在小的干扰下也具有较高的鲁棒性，因此我主要针对cascade rcnn做了一些研究和变种，其中效果最好的就是使用可变卷积层dcn来代替骨干中的卷积层，还有一些其他变种如更换骨干，骨干网和neck中更换更适合于多分类的损失函数，头部加载预训练模型进行权重和偏置的微调并更换损失函数。另外，我还加入了其他技巧，如金字塔特征、多尺度训练、学习率下降等等。本次实验主要以mAP为指标，其意义是平均精确率，受到每一个类别的精确率影响。

2.数据集

本次实验使用的数据集是Pascal VOC 2007数据集，其中共包含5011张大小和比例不等的图片。其中，包含5011张图片用于训练，4952张图片用于验证和测试共9963张图片。样本共包含20个不同种类，除了。本次实验规定的训练集包含2501张图片，验证集包含2510张图片，测试集同验证集。

3.相关工作

Paperwithcode查论文，cascade，faster，dcn

经过上网查找，我发现近几年在目标检测领域，主要的方法都是两阶段的，但大多数方法都是以R-CNN为基础的

4.方法

4.1设计思路

第一步：分析数据集；第二步：选择网络模型；第三步：实验；第四步：分析结果；第五步：优化模型，再实验，直到模型效果较好。

实验过程中主要优化方法是是根据实验输出调整参数或加入技巧。

4.1分析数据集：

先将数据集划分训练集和验证集。根据要求，训练集共有2501张图片，验证集共有2510张图片。仔细观察数据及后我发现，图像的大小不等，比例不一，而且部分图像存在旋转和遮挡，部分目标很小或黑暗，还有部分目标易误判。

(a) (b)

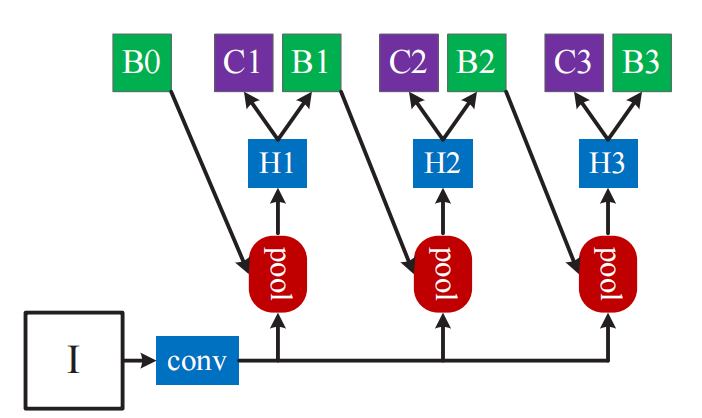
(c) (d)

图1，(a)为遮挡和小目标，图中狗狗尺寸很小；(b)为一误判目标，图中灌木和“盆栽”类很像；(c)为黑暗目标，“人”类别的样本与其他同类别样本相比颜色更黑暗，且色差较大；(d)为旋转目标

经过上网查找对于小目标，可以将cascade中最小的roihead中的IOU阈值调低来适应；对于黑暗目标

本次实验使用的是rcnn的变种网络：高速rcnn（faster\_rcnn）和级联rcnn（cascade\_rcnn）以及级联rcnn的变种。

rcnn是...组成成分包含骨干网络、颈部网络和头部网络，其中骨干网络的作用是，颈部网络的作用是，头部网络的作用是。



本次试验中用到的骨干网是残差网络Resnet，它的前44层...后面每3层组成1级，每一级包含一个卷积层，一个池化层，一个BN层，每一级将BN层的输出传送给下一级的卷积层，每级之间使用的是串联连接。

图片输入大小：500x400，450x350，400x300（受限于显存），不补全不拉伸

Faster使用的是...这种方法可以加快rcnn的训练和推理速度，但是并没有很好地克服rcnn的锚框过多的问题。

Cascade使用的是将上一层的BN输出结果直接输入给下一级的池化层，在传递语义信息的同时可以将深层信息与浅层信息混合。这些级联可以将语义信息层层传递，最大限度的保留各层语义信息，在后面聚类的时候就会获得更多有用的信息，保留浅层语义信息可以更好地提取小的特征，使聚类结果更倾向于真实情况，因此在处理一些较小的目标时，这种方法可以使性能有很大的提升。

Dcn可变卷积，这是一种可变形的卷积层，在目标不是正向的时候，可以通过变形适应目标的位置

【插图：rcnn、faster、cascade、dcn】

5.实验&结果分析

5.1.faster\_rcnn

faster\_rcnn实验结果非常差，初步试验中mAP仅有0.4左右，相比于使用同样技巧的cascade\_rcnn模型相差了10个点以上，因此我将faster\_rcnn作为一个参照，通过比较其结果的差异来分析哪些问题是由于数据集带来的，哪些是由于模型带来的。

cascade rcnn中，我对ROI进行了微调，同时修改了ROI头部的损失函数，从原始的smooth L1loss修改成了GIoUloss，经过六轮epoch的训练，我发现在epoch达到二的时候mAP就已经到达了峰值，但仅为0.56，而且loss开始波动，呈现出过拟合的状态。

在对cascade rcnn dcn进行训练的时候，我发现跟上一个实验出现了同样的问题，收敛速度很快，但马上就过拟合，而且mAP也不高，仅为0.6。因此我分析是学习率过高引起的。

观察每个类别的AP，可以很清楚地发现有一些样本数量少的类别，其AP值较低。结合数据及分析中发现的旋转、遮挡、黑暗、不清晰等问题，我认为是该类别样本中有效样本数量过少导致的，经过上网查找论文和博客，我发现解决这类问题最好的办法就是多尺度训练。加入多尺度训练，可以在不改变原数据集的前提下，有效的扩充有效样本的数量，同时适应各种不同状态的样本，在解决旋转样本、小样本和模糊样本上都有一定的帮助。另外我还注意到盆栽类别的AP值很低，和其他类别保持了10点以上的差距，同时AR值也和别的类别有一定差距。结合数据集分析可以发现，在一些图像背景中有和盆栽类别很相似的物体，如灌木、树冠等。图2画出热力图可以发现这些背景中的物品也被注意到了。这说明存在一部分误判的情况。我曾经尝试将验证集中的大部分盆栽样本转移到训练集中，但对盆栽类别的AP没有太大影响，这说明这类误判并不能通过直接增加训练集得到很有效的解除。经过询问学长和上网学习，我学会了在线难目标挖掘（ohem）并加入了模型，

实验介绍：目标检测，共有20个类别（基于mmdetection框架，进行了数据增强，加入了一些技巧，修改了cascadercnn的部分参数和结构）使用了一种基于锚框的目标分类与检测深度学习算法

指标：

mAP最高时为0.804，网络是cascade\_rcnn\_fpn\_dcn\_c3-c5\_1x网络，直接以voc格式读取数据集。20类数据的最高AP，除了pottedplant和chair外，其余的都可以达到0.7以上；但pottedplant（盆栽）和chair（座椅）的AP最低，仅为0.5~0.65，经观察数据集发现共有两个原因：第一个原因是样本太少了，训练集内pottedplant样本仅有200多个，chair也仅有300多个；第二个原因是大部分样本都不能让提供有效的特征，存在遮挡和旋转，难以得到准确的anchor，更难以提取特征，例如，大部分的chair样本上都有人或狗狗等其他样本覆盖。

模型：级联rcnn（cascade\_rcnn）/级联变形卷积rcnn（cascade\_rcnn\_dcn），backbone：resnet101，neck：1x，head：roi，epoch：4~8

实验1：cascade\_rcnn\_resnet101\_dcn\_x3-x5\_fpn\_giou\_1x\_voc

实验2：cascade\_rcnn\_resnet101\_fpn\_giou\_1x\_voc

实验3：cascade\_rcnn\_resnet101\_dcn\_x3-x5\_fpn\_giou\_1x\_voc\_lrdecrease

实验4：cascade\_rcnn\_resnet101\_fpn\_giou\_1x\_voc\_lrdecrease

① faster\_rcnn与另外两个模型比较，其效率太低，AP永远拉开超过0.1，因此，实验过程中，我决定淘汰faster\_rcnn

② 对于骨干网络的选择，resnet50与101相比，50训练和推理速度更快，但mAP和同等条件的101相比低了3~5个点，因此，在不需要考虑推理时间的情况下，我决定淘汰resnet50。

技巧：

1. fpn：金字塔特征，可以有效对各个尺度的特征进行提取

2.多尺度训练：图像多尺度输入，随机从多尺度中选取一个尺度进行输入

3. coco-finetune：使用coco的预训练进行微调，更新初始权重和偏置

初步结果：这4个模型中，最优的是cascade\_rcnn\_resnet101\_dcn

其结果为：mAP = 0.78

实验分析：

（1）多次实验可以看出训练中的bus的AP开始很低，然后慢慢增加经观察与查证发现，该网络对过大的目标检测能力比较差；

思路：缩小输入图像尺度，并进行多尺度训练；

（2）在经过（1）的调整后，发现训练遇到了瓶颈，map无法突破5.7，最高到达5.67，且在波动，怀疑是lr太高导致；

思路：降低lr，将lr变成原来的1/4，并加入lr下降机制，在第3和第5个epoch上进行lr下降；

（3）在经过（2）调整后，cascade\_rcnn和cascade\_rcnn\_dcn的mAP最高达到了0.76和0.78，但“盆栽”的AP一直很低，最高达到过0.5，经观察输出，发现是样本太少了，仅有200~300个样本；

思路：增加epoch；

下一步实验（6.4~6.15）

1.换模型，使用yolo\_v3实验

2.加入数据增强：

2.1.旋转：扩展数据集。经观察发现，大部分的car等类别，每张图像仅存在一个样本，但很少有旋转的情况，为了应对一部分旋转的样本，我将图片旋转一些角度（±30°，±60°），多余的地方留白，并更新对应的标注xml

2.2.增亮：部分图片存在太黑的问题，rgb三通道的值之和太低，这将影响特征提取（我是用的resnet是将3通道的值分别进入数个卷积核扩展通道）。我将平均rgb值求出，小于一定阈值认为是黑暗，然后将所有值放大1. 5~3倍（直到rgb值达到或超过阈值）

2.3调整数据集：针对较少样本的类别，将一部分处于val的图片调整至train内，或重复train内的图片，以达到扩展数据集的效果

8.参考文献

1. Cai, Z., & Vasconcelos, N. (2018). Cascade R-CNN: Delving Into High Quality Object Detection. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6154-6162.
2. Huang, Z., Chen, Z., Li, Q., Zhang, H., & Wang, N. (2020). 1st Place Solutions of Waymo Open Dataset Challenge 2020 - 2D Object Detection Track. *ArXiv, abs/2008.01365*.

|  |
| --- |
| 0.784 |
| 0.750 |
| 0.634 |
| 0.668 |
| 0.589 |
| 0.720 |
| 0.685 |
| 0.788 |
| 0.642 |
| 0.807 |
| 0.710 |
| 0.785 |
| 0.810 |
| 0.775 |
| 0.809 |
| 0.577 |
| 0.704 |
| 0.725 |
| 0.820 |
| 0.742 |

|  |
| --- |
| 0.757 |
| 0.833 |
| 0.681 |
| 0.779 |
| 0.677 |
| 0.784 |
| 0.815 |
| 0.834 |
| 0.704 |
| 0.767 |
| 0.739 |
| 0.848 |
| 0.784 |
| 0.869 |
| 0.856 |
| 0.660 |
| 0.774 |
| 0.779 |
| 0.833 |
| 0.803 |