

CV_10th homework

一、YOLOv4 测试结果如图 1 所示



(a) 源图像



(b) 检测结果

图 1: YOLOv4 目标检测测试结果

二、YOLOv4 的改进之处

YOLOv3 和 YOLOv4 的网络结构分别如图 2 和图 3 所示:

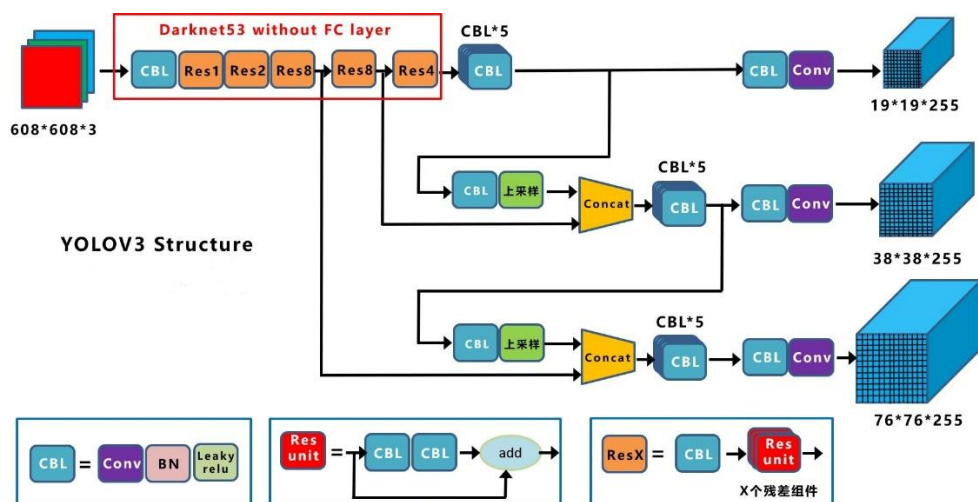


图 2: YOLOv3 网络结构

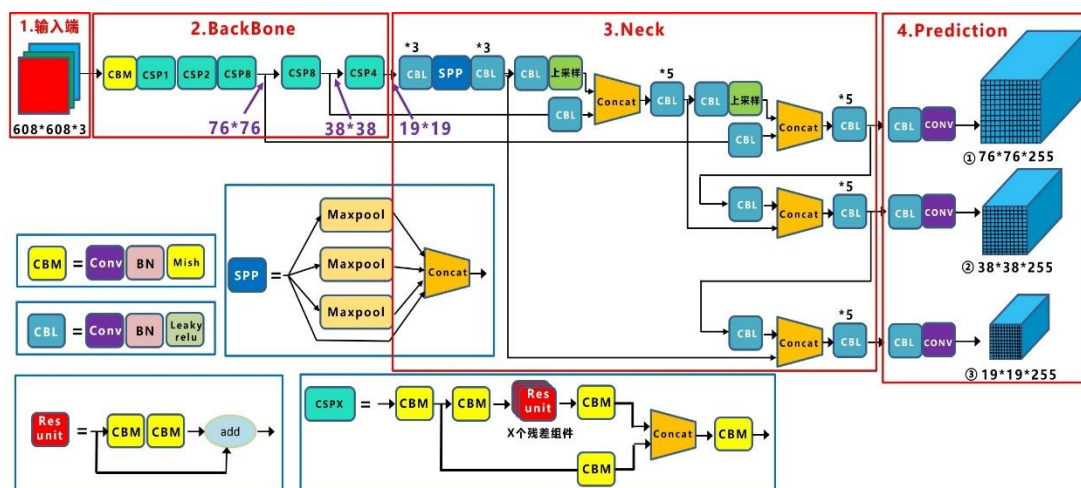


图 3：YOLOv4 网络结构

YOLOv3 是目标检测中 one-stage 中的经典算法，其网络包括 Darknet-53(主干网络)、anchor、FPN 等结构，如图 2-3 所示，具体组成上，YOLOv3 由 CBL(Conv+Bn+Leaky_relu)、残差结构 Res unit 以及 ResX 组成；与 YOLOv3 相比，YOLOv4 多了 CSP 结构和 PAN 结构，此外，YOLOv4 的基本组件中还包含 CBM(Conv+Bn+Mish)、CBL、Res unit 以及 SPP 结构。

YOLOv4 的创新之处主要为：

1. 输入端的创新：

YOLO v4 中使用的 Mosaic 基于 CutMix 数据增强的方式，二者相比，CutMix 只使用了两张图片进行拼接，Mosaic 则采用了对 4 张图片随即缩放、裁剪、排列的拼接方式进行数据增强，这种处理方式一方面丰富了数据集，其中随机缩放操作增加了很多小目标，使得网络的鲁棒性变得更好；另一方面，由于 Mosaic 增强训练时，可直接计算 4 张图片的数据，使得 Mini-batch 不需要很大，较少了 GPU 的使用。

2. Backbone 的创新：

1) CSPDarknet53 结构

YOLOv4 采用的 CSPDarknet53 主干网络结合了 YOLOv3 的 Darknet53 和 CSPNet 结构，一方面可以增强卷积网络的学习能力，使得网络具备轻量化和准确性较高的优点，另一方面，CSP 模块的引入，降低了计算瓶颈和内存成本。

2) Mish 激活函数的使用

YOLOv4 的 Backbone 中均使用了 Mish 激活函数，而在 Imagenet 数据集上使用 CSPDarknet53 的实验表明使用 Mish 激活函数可以提高分类精度。

3) Dropblock

代替 Cutout, YOLOv4 采用了 Dropblock 的方式缓解过拟合问题。与 cutout 直接将输入图像的部分区域清零相比, Dropblock 将 Cutout 应用到每一个特征图, 且在训练时以小归零比为初始值, 随训练过程线性增加该比率。Dropblock 可以在训练的不同阶段修改删减概率, 从空间层面和时间层面, 对网络的正则化过程都有更精细的改进。

3. Neck 处的创新:

1) SPP 模块的增加

YOLOv4 对目标检测采用了 SPP 模块, 实验对比表明, 与单纯使用 $k*k$ 最大池化的方式相比, YOLOv4 能够更有效的增加主干特征的接收范围, 显著的分隔了最重要的上下文特征。

2) FPN+PAN

区别于 YOLOv3, YOLOv4 在 FPN 层后加了一个自底向上的特征金字塔, 其中包含两个 PAN 结构, 这样的结合使得 FPN 层自顶向下传达强语义特征, 而特征金字塔则自底向上传达强定位特征, 二者从不同的主干层对不同的检测层进行参数聚合。

4. Prediction 处的创新:

目标检测任务的损失函数一般由分类损失函数和回归损失函数两部分构成。回归损失函数的选择上, YOLOv4 采用了 CIOU_LOSS 的回归方式, 使得预测框回归的速度和精度更高。在预测框的筛选上, 代替普通的 nms 方式, YOLOv4 将其中计算 IOU 的部分替换为 DIOU 的方式, 提高了筛选精度。