



# 阿里天池商标检测技术报告

队名 人菜瘾大

学校 南京理工大学

队长 蔡鑫浩

2022 年 6 月 11 日

# 目录

一、 赛题分析	II
二、 比赛策略	II
2.1 数据增强 . . . . .	II
2.2 模型选择 . . . . .	II
2.3 其他技巧 . . . . .	III
三、 消融实验	IV
参考文献	VI

## 一、赛题分析

初赛训练集 50 类，每类 50 个样本。复赛训练集 50 类，每类 20 个样本。首先就是样本少，其次对训练集的标注进行了仔细观察后，发现每个商标类标注中会有多个完全不同的特征（例如耐克的“勾子”商标，“NIKE”英文商标，“耐克”中文商标），而这些不同特征之间分布也是不均衡的，有些中文特征的标注甚至只有一两个。这两个问题在初赛和复赛中均有出现，如何解决这两个问题成为了比赛的关键。

## 二、比赛策略

### 2.1 数据增强

我们为了克服训练集训练样本少，测试集图片很多的情况，对训练集进行了在线/离线数据增强，并通过 Repeat Dataset 的方法，希望可以模拟测试集的分布。具体来说，对于离线数据增强，我们使用了 Copy Paste[1] 的方法将标注扣下来，并随机贴到其他的训练集中。为了解决均衡类内样本特征不均衡的问题，我们对那些特征少的样本贴的相对多一些，反之亦然。接着我们对扩充的训练集做了在线的数据增强，如位移、旋转、翻转、颜色通道变换等。我们通过对测试集的前期检测结果进行了可视化，发现对于测试集中出现的大量商标旋转、颜色变换、遮挡、空间形变等情况，模型不能很好地检测到。于是，我们增加了在线数据增强中旋转和颜色通道变化的比例。后期我们加入了 mixup[2] 数据增强，扩大训练样本量的同时增加了模型的鲁棒性。为了解决遮挡问题，我们加入了 cutout[3]。这些数据增强的效果将在消融实验部分进行详细介绍。

### 2.2 模型选择

我们选择了 CBNetV2[4] 作为我们的骨干网络框架。CBNetV2 融合了多个骨干网络的高层和底层的特征，并逐渐扩大感受野，以更有效地进行目标检测，如图2.1所示。

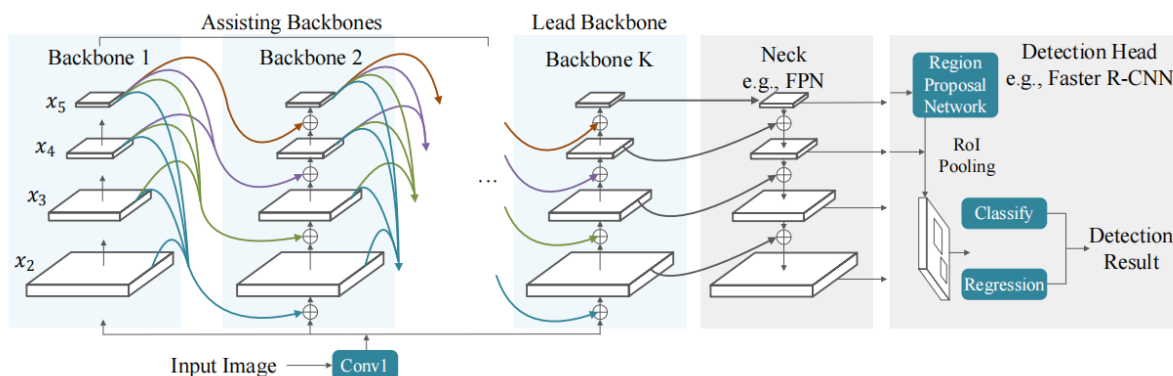


图 2.1 CBNetV2 网络结构

由于本次训练数据量少，训练时间较短，所以可以使用训练代价较昂贵的 backone。我们选择了目前性能 SOTA 的 Swin-Transformer[5] 作为我们的 backone。前期我们使用 Swin-Small，后期替换为 Swin-Base。

我们使用了 Cascade-Mask-Rcnn 作为检测框架。HTC 框架 [6] 由于时间和资源关系没能尝试，这是本次比赛略有遗憾的地方。

## 2.3 其他技巧

- **损失函数：** loss function 方面，我们使用交叉熵作为分类损失，GIOU 作为检测框损失；
- **训练和测试尺度：** 使用 [(2048,800),(2048,1400)] 作为训练多尺度。测试使用多尺度和翻转作为 TTA 的策略；
- **训练策略：** 我们使用 AdamW 作为优化策略，并使用学习率递减的方法，初始学习率设为  $1e-4$ 。共训练 30 个 epoch，在第 21、25、27 个 epoch 后分别衰减学习率。
- **预训练权重：** 初赛使用 CBNetV2 在 ImageNet1K 上的结果作为预训练权重。由于复赛可以使用初赛的测试集，故使用加上初赛的伪标签重新训练后得到的模型作为复赛的预训练权重。
- **训练集候选框聚类：** 我们将初赛和复赛的训练集的标注框进行了聚类，一共聚类为 6 类设置为候选框的比例。这样可以加速模型的收敛。
- **类内样本重新聚类：** 我们对复赛的训练集尝试过类内聚类，将类内不同的特征划为新的类重新进行训练，最终原本 50 个类被重新划分为 131 个类。这样会造成类间样本不平衡的情况，我们采用了 ClassBalanced 进行样本的重采样和 Seasawloss 作为分类损失函数来克服类间样本不平衡。但是实验结果表明，这种方法对此任务无效。

### 三、 消融实验

我们的实验都是在四张 Tesla V100-32G 上进行的。初赛和复赛阶段，我们都对训练集进行了 8: 2 的划分，来进行消融实验。由于算力的关系，我们前期使用 resnet101 作为 backbone 在划分数据集上进行实验，后期主要使用 CBNetV2 在所有训练集上进行训练，直接提交到网站上得到结果。

首先是我们在划分地验证集上对在线和离线的数据增强策略进行的消融实验，如表3.1所示：

**表 3.1 消融实验：离线、在线数据增强策略**

type	backbone	CopyPaste	repeat_dataset	Albu	划分验证集上的结果 (map)
cascade rcnn	resnet50	×	*1	×	0.466
cascade rcnn	resnet50	×	*3	×	0.494
cascade rcnn	resnet50	×	*3	√	0.503
cascade rcnn	resnet50	√	*3	×	0.501
cascade rcnn	resnet50	√	*3	√	0.522

接着我们进行了高级的在线数据增强策略的消融实验，包括 mixup 和 cutout，如表3.2所示：

**表 3.2 消融实验：mixup、coutout 在线数据增强策略**

type	backbone	CopyPaste	Albu	mixup	cutout	划分验证集上的结果 (map)
cascade rcnn	resnet50	√	√	×	×	0.522
cascade rcnn	resnet50	√	√	√	×	0.530
cascade rcnn	resnet50	√	√	√	√	0.536

接着我们进行了检测框架和 backbone 选择的消融实验，以下部分的实验都使用了上述的数据增强。如表3.3所示：

**表 3.3 消融实验：检测框架和 backbone**

type	backbone	batchsize	划分验证集上的结果 (map)
cascade rcnn	resnet50	8	0.536
cascade rcnn	resnet101	8	0.555
cascade mask rcnn	resnet101	8	0.558
cascade mask rcnn	swin-small	8	0.605
cascade mask rcnn	swin-base	8	0.620
cascade mask rcnn	CBNet-swin-small	8	0.644
cascade mask rcnn	CBNet-swin-base	4	0.667

## 参考文献

- [1] GHIASI G, CUI Y, SRINIVAS A, et al. Simple copy-paste is a strong data augmentation method for instance segmentation[J], 2021: 2918–2928.
- [2] ZHANG H, CISSE M, DAUPHIN Y N, et al. mixup: Beyond empirical risk minimization[J]. arXiv preprint arXiv:1710.09412, 2017.
- [3] DEVRIES T, TAYLOR G W. Improved regularization of convolutional neural networks with cutout[J]. arXiv preprint arXiv:1708.04552, 2017.
- [4] LIANG T, CHU X, LIU Y, et al. Cbnetv2: A composite backbone network architecture for object detection[J]. arXiv preprint arXiv:2107.00420, 2021.
- [5] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C] // ANON. Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 10012–10022.
- [6] CHEN K, PANG J, WANG J, et al. Hybrid task cascade for instance segmentation[C] // ANON. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 4974–4983.