文章编号：2095-2163(2021)05-0218-05 中图分类号：TP24 文献标志码：A

基于 CBAM-EfficientNet 的垃圾图像分类算法研究

叶 冲，杨晶东
（上海理工大学 光电信息与计算机工程学院，上海 200093）

摘 要：针对垃圾分类数据集，本文采用基于 Imagenet 数据集的迁移参数初始化 Efficient-net 模型，与经典的 VGG 和 ResNet50 模型对比，得到了较高的泛化性能和准确率。为了降低源领域数据集的特征参数对于目标领域数据集特征参数产生负迁移的影响，本文加入了 CBAM 注意力机制增强重要特征并忽视无效特征，同时使用批归一化和随机失活模块加速网络的训练并减轻过拟合程度，从而得到高性能、高效率的 CBAM-EfficientNet 垃圾分类模型。实验结果表明，基于 Efficient-net 模型的垃圾分类的准确率高于经典的 VGG 和 ResNet50 模型 5%以上，而本文所提出的 CBAM-EfficientNet 进一步提高了 2.5%。

关键词：CBAM；Efficient-net；泛化性能；分类模型

Algorithm research on garbage image classification based on CBAM-EfficientNet

YE Chong, YANG Jingdong
(School of Optoelectronic Information and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

【Abstract】 For the garbage classification data set, this paper uses the migration parameters of Imagenet data set to initialize the efficient net model. Compared with the classical VGG and resnet50 models, it gets higher generalization performance and accuracy. In order to reduce the negative effect of the feature parameters of the source domain data set on the feature parameters of the target domain data set, this paper adds the CBAM attention mechanism to enhance the important features and ignore the invalid features, and uses the batch normalization and random deactivation module to accelerate the network training and reduce the degree of over fitting, so as to obtain a high-performance and efficient CBAM efficientnet Garbage classification model. The experimental results show that the garbage classification accuracy based on efficient net model is more than 5% higher than the classical VGG and resnet50 models, and the CBAM-EfficientNet proposed in this paper is further improved by 2.5%.

【Key words】 CBAM; Efficient-net; generalization performance; classification model

0 引 言

如今，由人工进行垃圾分拣不仅对人体健康有伤害，而且垃圾分拣效率低。在日常生活中，每天会产生大量不同种类的垃圾，人工分拣只能解决其中的小部分，大多数都会进行填埋，从而对环境有很大的污染。随着计算机视觉的发展和图像数据的日益增多，深度学习方法在垃圾分类领域被广泛的应用。通过对不同种类的垃圾图像进行检测，让机器自动进行识别和分拣，从而提高资源利用率，减少环境污染。

传统的图像分类方法主要分为两部分，一是通过人为设置特征提取器提取图像特征，例如 HOG 特征[1]、SIFT 特征[2]、LBP 特征[3]等；二是通过设计更好的分类器算法来提高分类结果的准确率。然而随着移动互联网的普及以及各类互联网产品的推出，每时每刻都有海量的图像数据产生，这些图像数据复杂而且多样化，传统的图像分类算法并不足以支撑这些数据的分类。近年来，由于计算机技术的快速迭代和计算能力的日异丰富，给基于神经网络学习的图像分类算法提供了可能性。因此，越来越多的学者将目光投向基于神经网络的图像分类算法研究，以便找到更加快速高效的图像分类技术。深度学习通过多层非线性层的叠加[4-5]，使得模型可以拟合更加复杂的非线性映射。2014 年，来自牛津大学的 Simonyan K[6]等人提出了 VGG 网络，该网络常用的有 VGG16 和 VGG19 两种结构，除了在网络深度上的不同，二者在本质上并无区别。相较于 AlexNe[7]直接使用具有较大感受野的 11 * 11、7 * 7、5 * 5 大小的卷积核，VGG 用 3 * 3 的卷积核堆叠的方法来代替。通过这种方法不仅大大减少了网络的参数量，并且通过堆叠方式增加了网络的非线性程度，可以使网络拟合更加复杂的分布。同年，来自 Google 公司的 Szegedy[8]等人提出了 GoogleNet，相

作者简介：叶 冲(1994-)男，硕士研究生，主要研究方向：深度学习、计算机视觉；杨晶东(1973-)男，博士，副教授，硕士生导师，主要研究方向：智能机器人、计算机视觉。

收稿日期：2021-03-04

比于 VGG 网络通过增加网络深度来提高模型的拟合能力，GoogleNet 使用了 Inception 模块化结构。该结构通过不同大小的卷积核获得不同大小的感受野，然后进行拼接，将不同尺度的特征进行融合，从而提升网络的表达能力。2015 年，Szegedy 等[9]提出了 Inception V2 结构，该结构首先吸收了 VGG 网络的优点，将多个大卷积替换为小卷积叠加，节省了大量的计算量；二是提出了批归一化，降低了网络对初始化权重的敏感性，并且加快了网络的收敛速度。随后，Szegedy 等[10]对 Inception 结构进行了进一步的挖掘和改进，形成了 Inception V3 结构。该结构相比于 Inception V2 主要有 3 点改进：一是将 $n*n$ 的卷积结构分解为先进行 $1*n$ 卷积，然后再进行 $n*1$ 卷积的结构，大大降低了网络的参数量，从而可以堆叠更多的模块来提升网络的拟合能力；二是通过并行结构来优化池化部分，实现不同特征尺度的融合；三是使用标签平滑对网络输出进行正则化。通过这些改进，InceptionV3 结构相比于 GoogleNet 在 ImageNet 数据集上降低了约 8% 的错误率。同年，KaiMing He[11]等人提出了 ResNet 神经网络，相比于 VGG 网络，Resnet 多了一条残差回流通道，直接绕道将输入信息传给输出，保护了信息的完整性。Resnet 的残差结构使得网络只需学习输入和输出的差值，简化了训练目标，降低了训练难度，一定程度上解决了梯度消失和梯度爆炸的问题。2017 年，为了解决过多超参数给网络设计和计算带来额外难度的问题，Saining Xie 等[12]提出了 ResNext 网络。该网络使用了一种平行堆叠的结构来代替 ResNet 的三层卷积结构，使得网络在不明显增加参数的同时，明显提高了准确率。另外由于基础拓扑结构相同，减少了大量的超参数，便于训练和移植。为了解决单一增加网络宽度或单一增加网络深度导致的性能瓶颈和参数过剩问题，Tan M 等[13]提出了一种搜索网络架构 Efficient-net，该网络相比之前的卷积神经网络，主要有 Efficient-net-b0 到 Efficient-net-b7 共 8 种结构分别对应于不同分辨率的图像。

上述方法主要是基于大型数据集进行训练，在算法上实现了较大的创新。然而现实生活中常常面对小批量的数据分类任务，由于网络参数过多、太少的数据集，使以上网络模型难以得到充分的训练或者容易出现过拟合现象。虽然迁移学习可以解决这一问题，但面对不同的数据集任务，容易忽略数据之间的差异性。

因此，本文提出了一种自适应注意力机制和数据增强方法。通过数据增强，提高数据集的多样性，降低过拟合效应；在原模型中添加自适应注意力机制，通过自适应注意力机制，提取目标数据集较重要的特征，最终获得准确率的提高。

本文的主要贡献如下：

(1)提出一种数据增强的方法，帮助扩展训练集的多样性；

(2)将 EfficientNet、VGG 和 ResNet 进行对比，说明了 EfficientNet 的高效性和将其作为垃圾分类迁移学习主干网络的正确性；

(3)将 CBAM 注意力机制用于 EfficientNet 迁移学习网络中，并通过 Grad-CAM[14]对原图信息的重要性进行可视化，说明了本文所提模型能够加强特征提取功能，从而得到较优的分类准确率。

1　模型架构

1.1　Efficient-net 网络模型

经典的神经网络一般有以下 3 个特点：一是利用残差神经网络来增加网络的深度，通过更深的神经网络层数来提取更深的特征，并获取一定的性能提升；二是通过改变每一层提取的特征层数，实现更多的特征提取，得到更多的特征，以此来增加网络的多尺度表达特性；三是通过增大输入图像的分辨率，来帮助网络学到更多的图像细节。Efficient-net 是将这 3 个特点结合起来，通过调整输入图像的分辨率、深度、宽度 3 个维度，来得到更好的网络结构。其基线结构如图 1 所示。

在该网络的基础上，作者通过复合缩放的方法对网络的深度、宽度以及输入图像的分辨率 3 个维度进行优化，以获得在一定资源条件限制下的准确率最高的模型。其对缩放的关系如式(1)所示：

$$\begin{aligned} &depth: d = a^{\emptyset}, \\ &width: w = \beta^{\emptyset}, \\ &resolution: r = \gamma^{\emptyset}, \\ &\text{s.t. } a \cdot \beta^2 \cdot \gamma^2 \approx \gamma^{\emptyset}, \\ &a \geqslant 1, \beta \geqslant 1, \gamma \geqslant 1. \end{aligned} \tag{1}$$

其中，a、β、γ 是通过神经网络结构搜索得到的常数；$\emptyset$ 是根据计算资源大小设置的常数；d、w、r 分别代表网络的深度、宽度和分辨率的缩放系数。Efficient-net-b1~b7 是通过确定 a、β、γ 的最优取值后，调整 $\emptyset$ 所得到的。因此 Efficient-net 在有限的资源环境下，可以获得良好的性能提高。

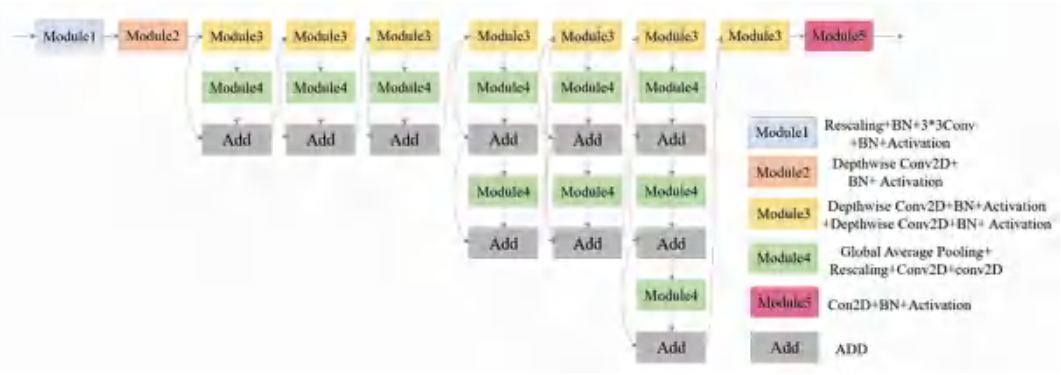


图 1 EfficienNet B0 结构图

Fig. 1 Structure of EfficienNet B0

1.2 CBAM 注意力机制

CBAM 注意力机制[15]是一种简单而高效的注意力模块。其将给定的中间特征图沿空间和通道 2 个独立的维度依次判断特征注意力图,并且与原始特征图相乘进行自适应优化,其结构如图 2 所示。

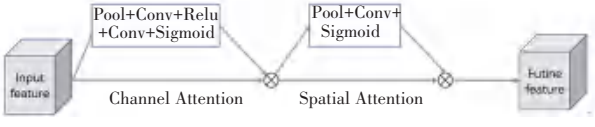


图 2 CBAM 模型结构图

Fig. 2 Structure of CBAM

1.3 整体模型结构

由于目标数据集与源数据集之间的差异,仅通过迁移学习提取特征容易造成特征提取不充分,从而导致最终精度的损失。一般做法是,放开所有预训练网络参数进行微调,但由于目标数据集数量的不足,该方法较容易得到过拟合的模型,不具有很强的泛化性能。因此,本文在预训练网络中的每一次下采样前加入 CBAM 注意力机制,以此来提升网络对于某些重要特征图和重要空间的注意力,从而提高模型准确率。结构如图 3 所示。

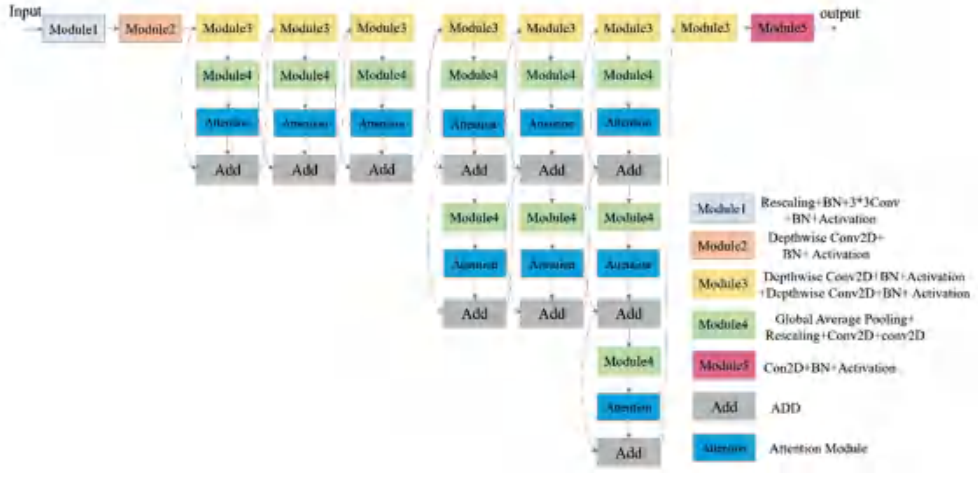


图 3 基于 CBAM 注意力机制的 EfficientNet 结构图

Fig. 3 Eficientnet structure based on CBAM attention mechanism

1.4 损失函数

由于垃圾图像数据集中,存在类别不平衡以及少部分图像未正确标注的问题,本文采用了标签平滑(Label Smooth)[16]和 Focal loss[17]相结合的损失函数。标签平滑是假定标签并不是 100%正确,将预定的类别设置为一个较大的概率,其它类别分配相应较小的概率。在存在较多分类类别、标注异常时,会起到较大的改善作用。Focal loss 是通过动态增加难分类样本的损失函数权重,降低易分类样本的损失函数权重,使得模型着重于难训练样本的训练,从而缓解样本不平衡的问题。如式(2)所示:

$$FL(p_t) = -(1-p_t)^{\gamma}\log p_t. \tag{2}$$

1.5 训练策略

首先,将 EfficientNet 网络去除最后一层全连接层,然后使用 Google 公司在 ImageNet 上训练完成的 EfficientNet B0 权重将其初始化;其次,加入自适应注意力机制重构网络,并将其余权重采用 He_norma 的方式进行初始化;固定前 3 层网络,采用本文所提

的损失函数进行训练；最后在验证集 loss 最低处保存模型。具体训练步骤如下：

输入：形状为[N,H,W,C]的图像和标签，

Step1：构建 EfficientNet 网络；

Step2：加载 ImageNet 预训练参数；

Step3：加入自适应注意力机制重构网络；

Step4：固定网络前 3 层进行训练；

Step5：在验证集 loss 最低处保存网络模型。

输出：图像的分类结果。

2　实验结果分析

本实验平台为两块 GPU 显卡，GPU 型号为 GeForce RTX 2080 Ti，该 GPU 显存为 11.07GB，显卡频率 1.112(GHZ)。实验基于深度学习框架为 keras2.2.4 和 tensorflow1.14，使用 python3.5 语言编程。

2.1　数据集描述

本文所采用的数据集来自于 2019 年华为杯垃圾分类挑战赛，该数据集共有 14 802 张图片，分为 40 个类别。图 4 为该数据集的统计结果。由图中可以发现，该数据集主要有以下特点：一是各类样本数量不均衡，其中第 4 类样本最少，仅有 100 张图片不到；二是图片分辨率大小不一致，包含多种分辨率的图像。由于本文采取 Efficient-net b0 提取图像特征，因此本文经过预处理将图片统一为 300 ´ 300 分辨率大小的图片。并将数据集划分为训练集、验证集和测试集，其划分比例为 7∶2∶1。

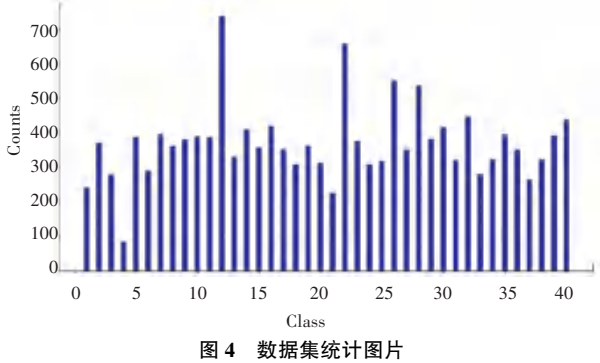


图 4　数据集统计图片

Fig. 4　Statistical picture of data set

2.2　数据集增强

本文首先按照图像最大边长与输入图像分辨率(300)，进行等比拉伸，然后将图像进行随机裁剪、颜色扭曲、图像旋转等进行图像增强。最后，将图像边界用 0 填充至输入图像分辨率大小(300 * 300)，如图 5 所示。左边为数据集部分原图，中间为本文所提数据增强方法增强后结果图，右边为直接 resize 后的结果图。从图中可以发现，经本文所提方法增强后，其图像比例与原始图像保持一致，而直接 resize 后其图像有所失真，不利于模型识别。

图 5　数据增强结果图

Fig. 5　Data enhancement results

2.3　实验结果与分析

利用 Efficient-net 网络与 VGG 和 Resnet 在该数据集上进行了实验。为验证本文所采用迁移学习网络在垃圾分类数据集上的高效性和准确性，本文将训练阶段验证集准确率和验证集 loss 的变化进行了可视化，如图 6、图 7 所示。由图可知，Efficient-net 采用了更高效网络特征提取模块，对未来实现垃圾分类自动化分拣有一定现实意义。其网络相比于 VGG 和 Resnet 收敛较快，并且有较高的准确率。在验证集上相比于 Resnet 提高了 5.1%，相比于 VGG 其准确率提高了 5.8%，说明了 Efficient-net 网络在垃圾分类数据集上的有效性和高效性。

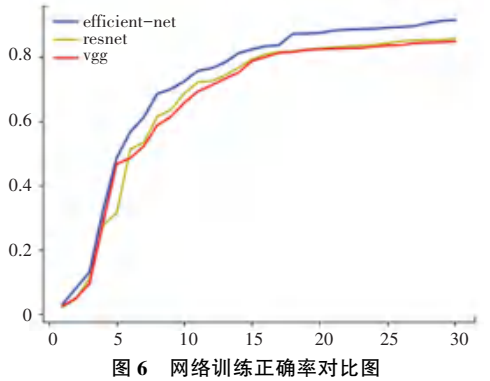


图 6　网络训练正确率对比图

Fig. 6　Comparison of network training accuracy

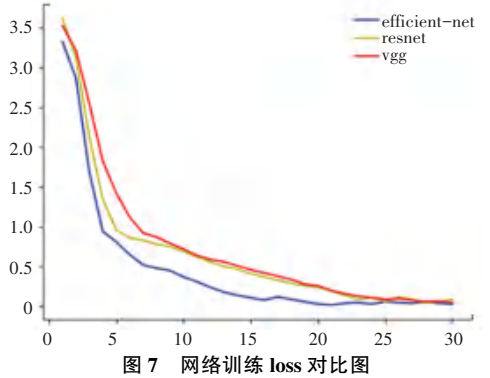
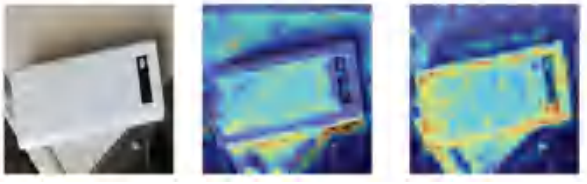


图 7　网络训练 loss 对比图

Fig. 7　Comparison of network training loss

为进一步说明自适应注意力机制的有效性,本文采用了 Grad-CAM[14](Grad-CAM 是最近提出的一种可视化方法,其使用梯度来计算卷积层中空间位置的重要性)对结果图进行可视化,如图 8 所示。其中,图 8(a)为原图,图 8(b)为 EfficientNet 最后一层特征层可视化结果,图 8(c)为在预训练网络中加入自适应注意力后,最后一层特征可视化结果。由图 8 可以发现,加入注意力机制后,其特征提取更为精确,且该图像所属类别的置信度更高,进一步说明了基于自适应注意力机制在 EfficientNet 网络的有效性。



(a) 输入 (b) 高效网络 (P = 0.12) (c) 时空注意力高效网络 (P = 0.46)

(a) Input (b) EfficientNet (P = 0.12) (c) CBAM-EfficientNet (P = 0.46)

图 8 图像空间重要性可视化结果图

Fig. 8 Visualization results of image spatial importance

最后,本文对比了 Efficient-net 和 CBAM-EfficientNet 在测试集上的准确率和模型大小,结果,见表 1。由表 1 可以发现,加入 CBAM 注意力机制后,虽然模型参数量有所增加,但是计算速度却降低了近 1%,并且准确率提升了 2.5%,证明了 CBAM-EfficientNet 在垃圾图像分类数据集上的有效性。

表 1 模型结果对比

Tab. 1 Comparison of model results

模型	模型大小	计算速度	准确率
Efficient-net b0	42.1 M	9.98 fps	89.6%
CBAM-EfficientNet	45.6 M	9.92 FPS	92.1%

3 结束语

本文将 EfficientNet 网络应用于垃圾图像分类数据集中,并与 VGG 和 Resnet 网络进行了对比。实验结果显示,EfficientNet 相较于 VGG 和 Resnet 网络具有较高的准确率和高效性,适合轻量化部署。针对迁移学习特征提取不充分问题,本文在预训练网络中加入了 CBAM 注意力机制,增强了预训练网络中重要的特征层权重,同时抑制无效特征层的影响,并与 EfficientNet 进行对比,实验结果证明 CBAM-EfficientNet 相比于 EfficientNet 提高了 2.5%的准确率,且计算速度未明显降低。进一步的工作将集中于更加适合迁移学习的注意力机制模型研究,目的是进一步增强网络的特征提取能力,从而得到较好的性能指标。

参考文献

[1] GRABNER M, GRABNER H, BISCHOF H. Fast approximated SIFT [C]//Asian conference on computer vision. Springer, Berlin, Heidelberg, 2006: 918-927.

[2] HE L, ZOU C, ZHAO L, et al. An enhanced LBP feature based on facial expression recognition [C]//2005 IEEE Engineering in Medicine and Biology 27th Annual Conference. IEEE, 2006: 3300-3303.

[3] JULINA J K J, SHARMILA T S. Facial recognition using histogram of gradients and support vector machines [C]//2017 International Conference on Computer, Communication and Signal Processing (ICCCSP). IEEE, 2017: 1-5.

[4] BENGIO Y. Learning Deep Architectures for AI[J]. Foundations & Trends® in Machine Learning, 2009, 2(1):1-127.

[5] BENGIO Y, LAMBLIN P, DAN P, et al. Greedy layer-wise training of deep networks [J]. Advances in Neural Information Processing Systems, 2007, 19:153-160.

[6] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. arXiv preprint arXiv:1409.1556, 2014.

[7] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks [C]// Advances in neural information processing systems. 2012: 1097-1105.

[8] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.

[9] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift [C]// International conference on machine learning. PMLR, 2015: 448-456.

[10] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2818-2826.

[11] He, Kaiming, et al. Deep residual learning for image recognition [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.

[12] XIE S, GIRSHICK R, DOLLÁR P, et al. Aggregated residual transformations for deep neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1492-1500.

[13] TAN M, LE Q V. Efficientnet: Rethinking model scaling for convolutional neural networks [J]. arXiv preprint arXiv: 1905.11946, 2019.

[14] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-cam: Visual explanations from deep networks via gradient-based localization[C]//Proceedings of the IEEE international conference on computer vision. 2017: 618-626.

[15] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 3-19.

[16] MÜLLER R, KORNBLITH S, HINTON G. When does label smoothing help[J]. arXiv preprint arXiv:1906.02629, 2019.

[17] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C]//Proceedings of the IEEE international conference on computer vision. 2017: 2980-2988.