基于深度学习的目标检测鲁棒性研究

**摘 要**

【缺一小段介绍目标检测的意义、现有方法的问题】

为了使目标检测技术能够应对多种干扰，从而更加适用于复杂的实际应用场景，本文针对基于深度学习的目标检测鲁棒性进行研究，主要工作包括：1）以目标检测工具箱MMDetection作为基线模型，融合了一系列数据增强和算法改进方法，包括Cascade R-CNN、ResNeXt101 64x4d、Cutout、Mixup、Soft NMS、Albumentations和标签平滑等；2）提出了适用于目标检测的CopyPaste增强策略、改进的WBF【第一次出现加全称】和带处理伪标签等方法，使之更适用于应对复杂干扰场景的目标检测任务。3）实验结果表明，模型经过优化后应对复杂干扰的性能得到了显著提升，目标检测CopyPaste、改进的WBF和带处理伪标签可以为目标检测算法的实际应用提供有益参考。

**关键词** 目标检测 深度学习 鲁棒性

**Study on the robustness of object detection based on deep learning**

**ABSTRACT**

【参考中文】

In order to make the object detection technology comprehensive to deal with a variety of interference, so as to be more suitable for complex real application scenarios, this paper reads and practices several literatures. I study the basic concepts and current applications of computer vision object detection, understands the development history of object detection from traditional learning to deep learning, and studies the latest robustness-related research on object detection, and proposes a study on the robustness of object detection based on deep learning. Using the benchmark dataset and six interference datasets provided by the competition, experiments are conducted using MMDetection, an open source object detection toolbox based on the PyTorch framework, and the mean average accuracy is used as the final evaluation metric. A series of data enhancement and algorithm improvement methods are used in the experiments, including Cascade R-CNN, ResNeXt101 64x4d, Cutout, Mixup, Soft NMS,, Albumentations, and label smoothing, the basic theories and specific algorithms of these methods were briefly introduced in sequence. And the object detection CopyPaste, improved WBF and pseudo-labeling with processing are proposed, adding new mechanisms to the traditional methods and making the methods more suitable for object detection tasks dealing with complex interference scenes. The experimental results are then compared and analyzed in detail in terms of data, and the reasons for their ability to improve robustness are further elaborated by introducing the features and advantages of each method. The experimental results show that the performance of the model in dealing with complex interference has been significantly improved after optimization, and that the object detection CopyPaste, improved WBF and pseudo-labeling with processing can provide useful references for the practical application of object detection algorithms. Finally, this study is summarized and reflected, according to the existing problems of object detection, the future development trend of object detection robustness research is prospected.

**KEY WORDS** object detection deep learning robustness

**第一章 绪论**

**1.1 研究背景**

**1.1.1 目标检测简介【标题前后应该有0.5倍行距】**

计算机视觉是计算机科学的一个领域，其重点是使计算机能够识别和理解图像和视频中的物体和人，试图复制人类的视觉方式以及人类对所见事物的理解方式。在计算机视觉领域中有几个基本的任务：图像分类、目标检测、实例分割及语义分割[1]，如图1-1所示。其中目标检测作为计算机视觉的重要问题，是其他计算机视觉任务的基础。目标检测的主要目标是解决计算机视觉应用中两个最基本的问题，包括分类和定位两方面，即该物体是什么？该物体在哪里？

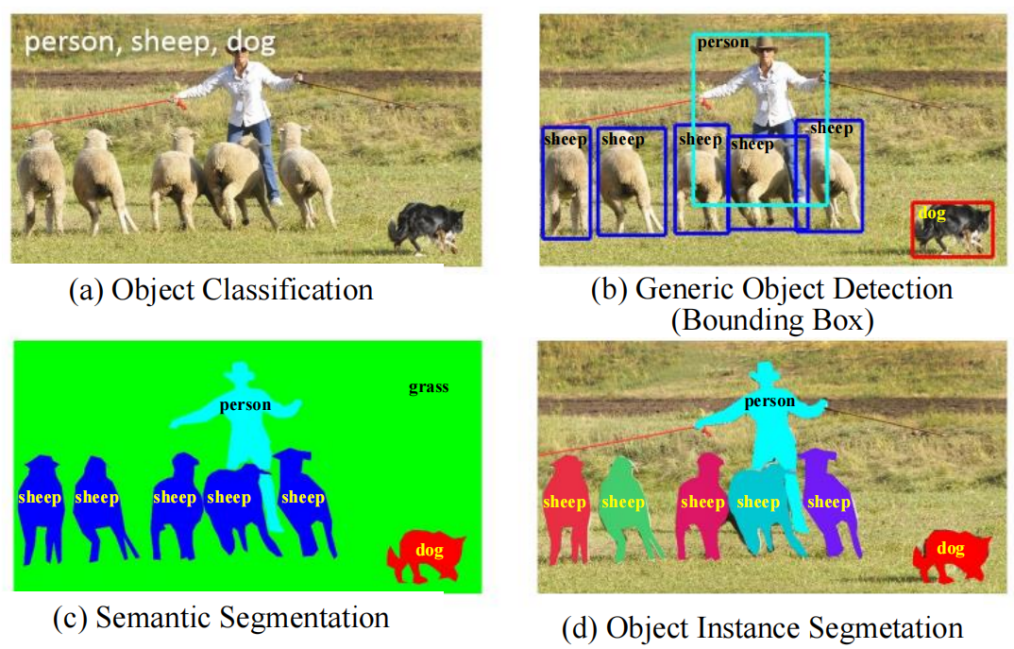


图1-1 计算机视觉重要任务[1]**【图标题后应该有0.5倍行距】**

目标检测已渗透于日常生活中，在多个领域都有广泛的实际应用，包括行人检测、人脸检测、文本检测、车辆检测等。行人检测作为一种重要的目标检测应用，在车辆辅助驾驶系统、智能视频监控、刑事侦查等领域得到了广泛的关注，发挥重大作用。人脸检测作为最古老的计算机视觉应用之一，现在已经应用于各行各业，例如手机照相中常用的笑容识别功能、在超市等结账场景下的刷脸支付、视频软件中的人脸化妆等。

目标检测的研究在学术界和产业界都有着广泛且深刻的影响，存在着大量的相关文献研究和讨论，近年来，随着深度学习技术的快速发展，极大地推动了目标检测的研究，目标检测技术的不断提升和完善，将对计算机视觉其他任务产生积极影响，并且朝着更加适应现实生活的需要的方向发展，不仅将在商业领域产生更大的经济价值，更会给人们的日常生活带来方便。

**1.1.2 目标检测发展历程**

如图1-2所示，以2014年为分界点，目标检测的发展大致经历了14年之前的“传统目标检测期”和14年之后的“基于深度学习的检测期”两个历史时期[2]。

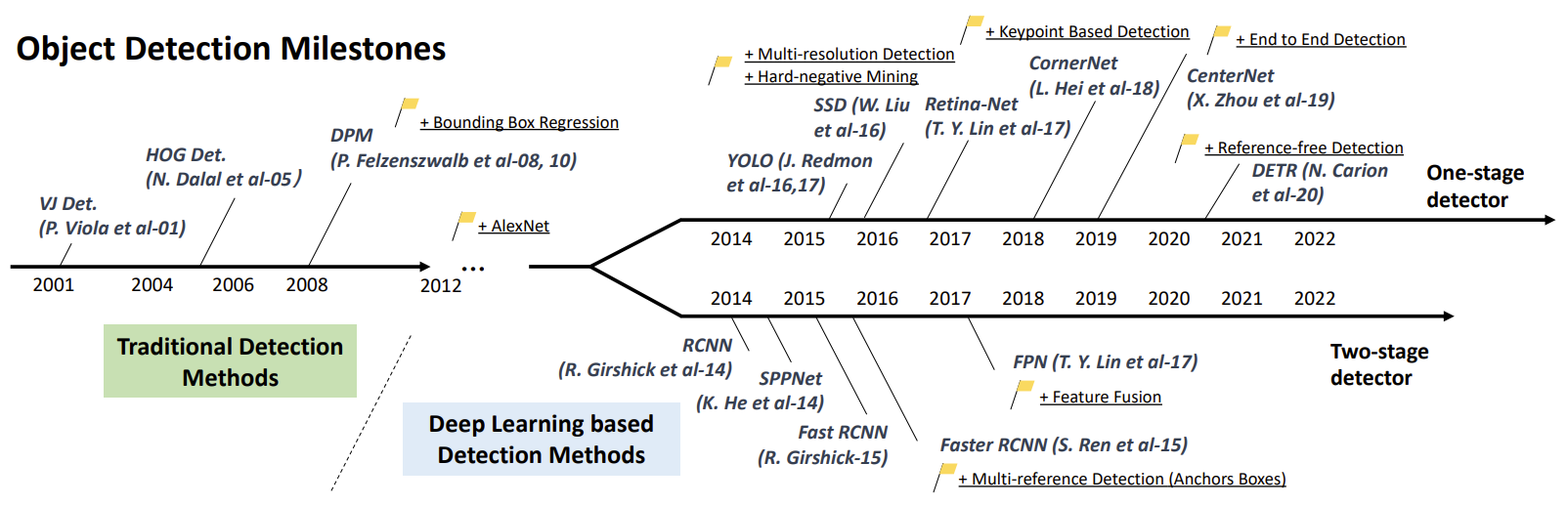


图1-2 目标检测里程碑[2]

传统的目标检测方法主要包括区域选择、特征提取和分类三个阶段。一般采用滑动窗口算法进行候选区域选择，通过不同大小的滑动窗口在输入图像上划定目标可能出现的区域，从而初步定位目标。采用局部二值模式、方向梯度直方图等算法提取候选区域的特征。采用支持向量机、自适应增强等算法对提取的图像特征进行分类。

传统的目标检测算法效率低、检测精度低，对于不同形状的不规则物体的鲁棒性较差。随着深度学习技术的发展，利用卷积神经网络（Convolutional Neural Networks, CNN）提取图像特征来实现目标检测已成为机器视觉领域的热点之一。目前，基于深度学习的目标检测方法主要有两个分支，即基于区域提取的两阶段目标检测模型和基于直接位置回归的一阶段目标检测模型。[3]

两阶段目标检测区域提取算法的核心是卷积神经网络CNN，具有局部连接和权值共享的优点，在目标分类应用中具有良好的鲁棒性。区域提取操作采用候选框来提取候选目标，从特征图中找出可能的候选区域，最后对候选区域进行滑动窗口操作，进一步判断目标的位置和类别信息，能够更好地处理目标的遮挡、旋转等复杂场景，适应不同尺度、形态的目标。还可以进行多种筛选和后处理方法来进一步提高检测结果的准确率，如非极大值抑制、筛选低置信度区域等，这些后处理方法能够极大地提高两阶段目标检测的鲁棒性和实用性。

随着两级检测方法的不断发展和优化，一级检测网络从实时检测的角度对检测方法进行了优化，适用于不同场景下的任务，YOLO（You Only Look Once）系列是一阶段中的经典方法。一阶段目标检测算法不通过中间层提取候选区域，而是在整个卷积网络中进行特征提取、目标分类和位置回归。通过逆向计算得到目标位置和类别。在识别精度略弱于两级目标检测算法的前提下，速度有了很大的提高，这使得基于深度学习的目标检测算法适用于许多需要较高推理速度的任务。

**1.1.3 鲁棒性研究现状**

鲁棒性是指系统在面对扰动、噪声或异常情况时，能够保持其良好的表现和性能，维持其稳定性和可靠性的能力，是图像识别、语音识别、自然语言处理等领域的重要概念。目标检测中的鲁棒性是指目标检测模型在遭遇图像质量变差、环境变化、特征干扰等情况下，仍然能够准确地检测出目标，对于实际应用具有非常重要的意义，主要通过采用多种方法改进算法、引入异常数据提高数据集的质量来提高目标检测中的鲁棒性。

CVPR2023也出现了有关目标检测鲁棒性的研究代表性的算法有。

3D目标检测在常见极端场景下的鲁棒性基准[4]研究了自动驾驶中的3D目标检测，指出现有3D检测器缺乏对真实世界中恶劣天气和传感器噪声等情况下的鲁棒性，引发了人们对自动驾驶系统安全性和可靠性的担忧。为了全面而严格地评估3D检测器的损坏鲁棒性，该研究考虑到真实世界的驾驶场景，设计了共27种3D目标检测中的常见损坏，分为天气、传感器、运动、对象和对齐过程五种级别，涵盖了现实世界中的大多数损坏情况。使用自动驾驶中最常使用的数据集KITTI、nuScenes和Waymo，建立了三个损坏鲁棒性基准，并在公共数据集上测试了24个不同的3D对象检测模型的鲁棒性。最终得出了以下结论：在所有损坏类型中，运动级破坏是导致所有模型性能显著下降的最主要因素，在正常天气下训练的模型容易受到天气级破坏的影响；单独使用相机模型鲁棒性较差，而激光雷达相机融合模型能有效缓解图像损坏的影响。该研究为3D目标检测的开发和提高自动驾驶系统的安全性和可靠性提供重要参考。

通过抽象增强鲁棒图像分类的验证训练[5]介绍了一种基于抽象的、可验证的训练方法，用于增强深度神经网络的鲁棒性，提高其对抗性攻击的能力。该方法通过将扰动图像映射到区间中进行训练，为所有映射到同一区间的图像赋予相同标签，抽象化处理能够减少训练集的方差，使损失函数更加平滑，显著提高训练模型的鲁棒性。并提出了黑盒验证方法用于验证训练后的神经网络的鲁棒性，在有限的输入空间内进行验证，在保证准确性的前提下，提高了验证效率，使用基于梯度下降的算法调整抽象颗粒度，以获得更细的训练间隔，从而提高神经网络的鲁棒性。在各种基准测试中对该方法进行了评估，并与现有的方法进行了比较，在训练和验证中都具有更低的验证错误和更少的计算开销。该研究可以有效提高深度神经网络对抗攻击的鲁棒性，进而提高模型的稳定性。

**1.2 研究内容**

**1.2.1 研究目的**

虽然目标检测算法发展迅速，但仍存在很多未解决的问题，在特殊场景的性能上仍有很大的优化空间。比如：遮挡导致特征表达不完整，增加了分类和定位的难度，一般情况下，在检测车辆和人群时容易发生遮挡现象；在恶劣天气如雪、雨、雾、强光等条件下，目标检测性能大大降低，在自动驾驶应用场景中是一个十分重要的安全问题；除此之外，还有目标的3D、姿势、形状、纹理等干扰因素，对目标检测的鲁棒性提出了挑战，是为在各类实际应用场景中获得良好效果而必须要解决的难题。

针对上述问题，本文提出了基于深度学习的目标检测鲁棒性研究，通过数据增强和改进算法的方式，使其更好面对遮挡、形态差异和天气等干扰问题，进行全面而精准的识别，从而有效应对复杂多样的实际情况。

**1.2.2 干扰数据**

数据集在目标检测研究的历史上起着重要的作用，不仅是比较竞争算法性能的共同基础，也促使该领域变得更加复杂和具有挑战性。通过互联网能够获得数百万图像的海量数据集，帮助目标识别领域建立更全面的图像和分类数据，从而实现更丰富和多样化的目标。

在过去的10年里，已经发布了许多著名的数据集，如PASCAL VOC【】、ImageNet【】、MS-COCO【】等。

本文的干扰数据集选用了ECCV 2022中OOD-CV(Out-Of-Distribution Generalization in Computer Vision)比赛提供的基准数据集，包含形状、姿势、纹理、上下文、天气和遮挡六种干扰，如图1-3所示，主要来自PASCAL VOC和ImageNet数据集，该比赛旨在提高计算机视觉模型在未知领域的泛化能力，改善计算机视觉技术在实际应用中的可靠性和鲁棒性。

(a)形状 (b)姿势 (c)纹理

(d)上下文 (e)天气 (f)遮挡

图1-3 干扰数据类型

PASCAL VOC数据集，涵盖日常生活中常见的20个类别，大量训练图像，接近真实世界的应用，一个图像中的多个对象，包含许多困难的样本。ImageNet包括大量的对象类别，每个图像有更多的对象实例和类别，比PASCAL VOC更具挑战性。

六种干扰中的上下文类型干扰主要包括室内、室外和水下场景，水下场景[6]相对于室内外场景更加复杂，由于光线迅速衰减，导致图像模糊不清，并产生噪音和失真，而形状、姿势、纹理、天气和遮挡较为容易理解。

**1.2.3 框架工具**

基于PyTorch框架的开源目标检测工具箱MMDetection[7]，为目标检测领域研究提供了一个强大、高效、易用的框架，本文将其作为为基线模型。

MMDetection适用于多种的应用场景，如目标检测、实例分割等领域，提供良好的API文档，可以减少用户对物体检测算法的研究和实现时间，使用户能够专注于模型优化、数据集自定义和应用实现上。

该工具箱提供了丰富的目标检测算法的实现，支持多种Backbone网络，内置多种数据增强方法，包括图片旋转、裁剪、缩放等操作，可以有效地扩充数据集并提高模型性能。包含多种目标检测算法的预训练模型，也可以使用自己的数据集对这些模型进行微调。

并且提供了高效的数据加载器，可以在数据预处理时提高代码执行速度。支持分布式训练，可以在多个机器上进行平行训练，加快模型训练速度和提升模型的性能。提供了简单易用的预测接口，用户可以输入一张图片，立即得到其物体检测结果。具备多GPU训练和推理等功能。

MMDetection具有速度快、性能高、易于阅读和修改的优点，采用模块化设计，可以灵活地组合和调整各个模块，用户可以根据自己的需求和领域特点，定制化自己的物体检测算法，从而得到需要的检测算法，可以为用户提供高效的算法实现和快速的实验验证。因此被广泛应用于各种目标检测任务和研究领域。

使用MMDetection进行数据预处理方法、模型结构优化和训练方案设置的设计，进行训练并保存训练日志以及模型。通过针对干扰数据集的类别预测代码，使用训练好的模型，预测干扰图片类别并将各干扰类型的检测结果分别保存。使用数据集的评估代码，根据预测结果进行评估，将各干扰类型的平均精度以及平均精度均值输出。

**1.2.4 评价指标**

评价检测算法性能的标准有：检测速度、准确率(accuracy)、精确率(precision)、召回率(recall)等。

准确率表示分类对的样本占所有样本的比例，是正确预测类别的样本数量与样本总数的比值。召回率是指在测试集中所有的正样本中被正确识别的正样本所占的比例。

最常用的目标检测评价方法是平均精度（Average Precision, AP），通过精度和召回得出。该方法首次在VOC2007中提出，被定义为在召回情况下的平均检测精度，通常以特定于类别的方式进行评估，需要每个目标类别分别计算。

在一般的目标检测中，检测器通常通过检测一些目标类别来进行测试。为了比较所有目标类别的性能，使用所有目标类别的AP平均值：平均精度均值（Mean Average Precision, mAP）作为最终的性能度量。

模型评估使用AP50，即交并比IoU（Intersection over Union）为0.5时的平均精度均值，交并比IoU公式如下：

式（1-1）

其中，A和B分别为预测边界框和真实边界框。通过判断检测物体与真实物体的重叠程度是否达到了阈值，评估模型检测物体的性能。

**第二章 目标检测模型相关方法**

**2.1 深度检测模型**

**2.1.1 Cascade R-CNN**

Cascade R-CNN[8]（Cascade Region-based Convolutional Neural Network）是目标检测任务中一种高效的多级级联神经网络，其主要思想是在训练过程中通过使用多级级联的检测器，根据IoU进行逐步筛选，在不牺牲检测速度的情况下提高目标检测的性能。

Cascade R-CNN通常使用三级的检测器来具体实现，包括：级联1、级联2和级联3，如图2-1所示。这三个检测器网络的结构是完全相同的，其功能是计算物体类别的概率，同时预测物体的边界框。但三个检测器网络的训练阈值不同，每个级联分别使用逐步增加的三个IoU交叉验证阈值进行训练，以便检测出在前一个级联中被忽略或错误检测的物体，从而逐步提高检测精度。

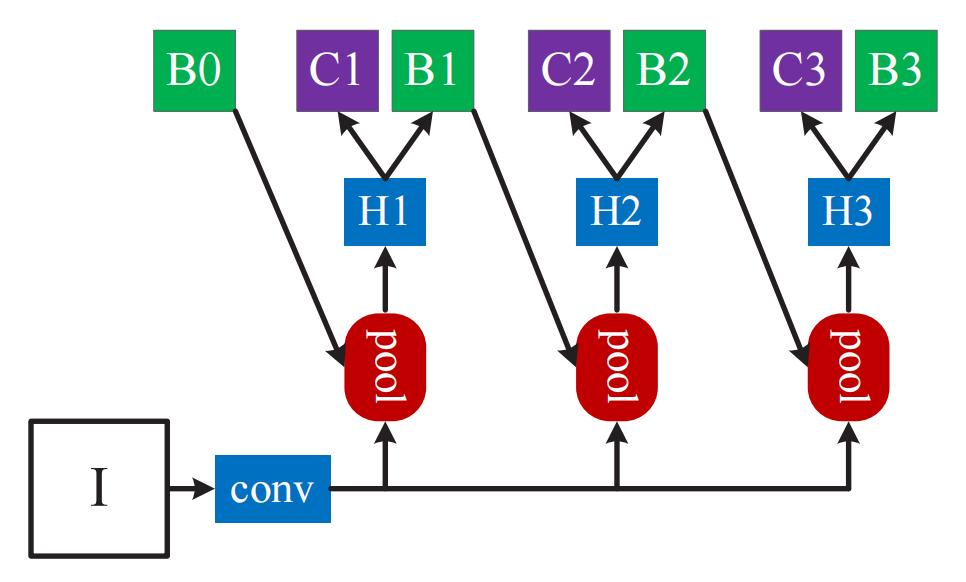


图2-1 Cascade R-CNN框架结构[8]

在训练过程中，级联1首先在较低的IoU阈值下识别出物体，在过滤掉大量的背景负样本的同时，保留一些负样本并分配较高的分数，从而减少漏检保证召回率。级联2接着使用较高的IoU阈值进行进一步的筛选，此时级联1产生的假阳性会被识别出来，同时未被正确识别的对象也被保留，由级联2进行进一步筛选。级联3最后则对检测结果实行更为严格的判断，使用最高的IoU阈值对检测结果进行验证，在保持高效率的同时进一步提高准确性。

**2.1.2 ResNeXt101 64x4d**

ResNet[9]（Residual Network）在经典卷积神经网络的基础上加入了残差块，即在每个卷积层之间引入了一个跨层连接使得输入信息能够沿着这些连接直接传递到后面的层。将输入特征映射与输出特征映射相加，得到残差特征映射：

式（2-1）

其中可以是一个任意的函数。然后再通过非线性激活函数将残差特征映射输入到下一个残差块中。这种跨层连接的设计有效地减小了网络的参数量，同时提高了模型的准确性，从而避免了网络过深时梯度消失和梯度爆炸等问题。

ResNet主要有五个常见的版本：ResNet-18、ResNet-34、ResNet-50、ResNet-101、ResNet-152，其中数字表示网络深度，即卷积层的数量。

ResNeXt[9]和ResNet一样都是基于残差块的思想深度卷积神经网络模型，通过残差连接将输入特征映射和输出特征映射相加，从而得到最终的残差特征映射，可以有效地解决深度神经网络中梯度消失和梯度爆炸的问题，使得训练更加稳定和容易。

ResNeXt在ResNet的基础上引入了分组卷积的结构，其输入通道被分成多个组，将多个相似的卷积操作组合，并在每个分组中使用相同的卷积核，最后将各个组的输出通道拼接起来，作为卷积层的输出特征映射。

在保持卷积核数量不变的情况下，增加了网络的宽度，可以大幅度提高模型的表达能力和性能，还可以有效地减少计算量和参数数量，同时不会影响模型的输入输出大小。

ResNeXt101 64x4d是一个基于 ResNeXt架构的卷积神经网络模型，是该系列模型中的一种，深度为101层且使用了64个cardinality和4d的bottleneck width，进一步提升模型的表达能力和性能。

**2.1.3 Soft NMS**

非极大值抑制[12]（Non-Maximum Suppression, NMS）是一个在目标检测领域中被广泛使用的方法。在目标检测中，通常会使用卷积神经网络等深度学习模型来检测出图片中存在的物体，但是在同一目标的多次检测结果中，会出现多个重叠的框，此时需要使用非极大值抑制将重叠的框进行抑制，在所有的框中寻找得分最高的那个框，并将与其重叠度较大的其他框剔除，最终得到唯一可靠的框，达到去除重复检测结果的目的。

NMS的基本原理是对训练集中的每张图片进行标注，得到真实边界框，并随机生成一定数量的预测框，计算它们与真实边界框的IoU值，并根据IoU值从大到小进行排序。选取当前得分最高的框作为参考框，并将其与其它框之间的IoU值进行调整，如果IoU值大于给定阈值，则将这个框删除：

式（2-2）

其中表示第i个目标的分数，表示第i个目标与分数最大的边界框的交并比，表示交并比的阈值。

当多个候选框与目标框的重叠面积大于一定阈值时，NMS只保留置信度最高的目标框，而将其它冗余框剔除，在目标框重叠较多时容易误判，删除掉与目标重叠面积更大、但置信度较低的框，从而使得目标检测结果不准确。

软性非极大值抑制[12]（Soft Non-Maximum Suppression, Soft NMS）改进了传统 NMS的算法流程，通过采用高斯函数对冗余框的权重进行衰减来消除对最终权重计算的干扰，使得重叠相对较少的框保持其权重不变，重叠较多的框则逐渐降低其权重，在不增加其复杂度的情况下，更加有效地应对目标重叠的情况，从而提高了目标检测的准确性和泛化能力：

式（2-3）

**2.2** **数据增强**

**2.2.1 Cutout**

Cutout[10]是一种常用的数据增强方法，其基本实现原理是通过随机遮盖图像的一部分区域来生成新的训练样本，即从输入图片中随机选择n个矩形区域R1、R2、…、Rn，将矩形区域R1、R2、…、Rn内的所有像素点替换为0或者其他预定义的数值，从而得到新的训练图像，如图2-2所示。



图2-2 应用Cutout的CIFAR-10数据集[10]

可以将这些被遮盖的区域看作是删除了一些有关物体的信息，从而迫使模型关注其他区域，因此能够更好地学习到不同位置的图像特征。避免模型出现过拟合，提高模型的预测精度，增强深度学习模型的泛化能力和鲁棒性。

**2.2.2 Mixup**

Mixup[11]是一种简单而有效的数据增强方法，通过将两张不同图像按照一定比例混合产生一个新的图像，同时对应的标签也按照这个比例进行融合，可以有效减少过拟合现象，提高训练的鲁棒性。

Mixup的基本思想是在训练集中随机取两张图片，并按照一定的比例混合它们的像素值和标签，生成一个新的训练样本：

式（2-5）

式（2-4）

其中为混合系数，是一个在范围内均匀分布的随机数，和是从训练数据中随机抽取的两个特征目标向量。越小，混合后的样本就越接近于原样本，但也容易造成过拟合；越大，混合后的样本就更具有随机性，可以有效地扩充数据分布，但也可能会带来一定的噪声。

**2.2.3 Albumentations**

Albumentations[14]是一个用于图像增强的Python库，与深度学习框架PyTorch、TensorFlow一起使用，该库旨在提供一个灵活且高效的框架，被广泛用于图形分类、目标检测、语义分割、实例分割等计算机视觉任务中的数据增强，从现有数据中创建新的训练样本，从而提高训练模型的质量。

该库提供了简单的API，可以处理RBG图像、灰度图像、多光谱图像、分割掩码、边界框和关键点数据类型，包含了超过70种图像变换方式。

像素级变换只改变输入图像，不改变分割掩码、边界框和关键点，例如：RGB shift对图像的RGB通道进行随机偏移，Hue Saturation Value对图像的色调、饱和度和亮度进行随机调整，Channel Shuffle对图像的RGB通道进行随机交换，JPEG Compression对JPEG格式的图像进行随机压缩，Blur对图像进行均值模糊处理，Median Blur对图像进行中值滤波处理，To Gray将彩色图像转换为灰度图像，Random Brightness Contrast对图像进行随机的亮度和对比度调整。



图2-3 Albumentations像素级变换生成新图像[14]

空间级变换将同时改变输入图像并对分割掩码、边界框和关键点进行更改，例如：Vertical Flip沿着垂直轴对图像进行翻转，Horizontal Flip沿着水平轴对图像进行翻转，RandomRotate90对图像进行90度的随机旋转，Shift Scale Rotate对图像进行平移、缩放和旋转等随机变换。



图2-4 Albumentations空间级变换生成新图像[14]

以上是Albumentations库中部分常见的图像变换方式，涵盖了旋转、缩放、裁剪、翻转、仿射变换、变形、颜色调整等多种操作，它们可以根据不同场景和需求灵活组合使用，并且支持随机应用多个变换，让每个样本都可以得到不同的处理结果，进一步扩充了数据集的多样性。

**2.3 正则化**

**2.3.1 标签平滑**

交叉熵（Cross Entropy）是用来衡量两个概率分布之间的差异的度量方法，设P与Q是两个概率分布，表示中某一事件发生的概率，表示中该事件发生的概率，则交叉熵定义为：

式（2-6）

交叉熵损失函数[15]（Cross Entropy Loss）是深度学习分类任务中的一种常用的损失函数，可以用于计算模型预测的类别标签与正确的类别标签之间的差异，来评估模型的性能，并用于参数优化和反向传播，从而得到更好的分类结果。其中，多分类的交叉熵损失函数可以被归纳为：

式（2-7）

其中，表示样本总数，表示类别数量，是一个one-hot向量，表示第i个样本的真实标签，如果该样本属于第j个类别，则，否则，是模型输出的预测结果，表示预测第i个样本属于第j个类别的概率。

在某些情况下，交叉熵损失函数可能会导致过度拟合、过于自信等问题。当真实标签为1时，传统的交叉熵损失函数会最大化预测结果中1所对应的概率，并尽可能抑制其他标签的概率。这样，在训练过程中往往会导致模型对于真实标签非常自信，而忽略了其他标签的概率，从而影响了模型的泛化性能。

而标签平滑[15]（Label Smoothing）则是一种在交叉熵损失函数基础上加入一个正则化项，不再将真实标签视为一个one-hot向量，而是采用平滑后的概率分布作为真实标签，强制模型更加关注不同类别之间的区别，避免对某个特定类别过分依赖，有助于防止模型对训练数据中噪声标签的过拟合，提高了模型的泛化能力和鲁棒性。标签平滑可以通过以下公式具体实现：

式（2-8）

其中，表示类别数量，为原始标签概率分布，为修正后标签概率分布，是一个平滑参数。该公式包含两项损失，第一项是常规的交叉熵损失，第二项则是考虑标签的真实分布，对非真实标签的预测概率进行惩罚，从而抑制模型的过拟合，使得模型变得更加稳定。

【缺本章小节】

**第三章 基于半监督学习的目标检测优化算法**

**3.1 数据增强**

**3.1.1 目标检测CopyPaste**

CopyPaste[13]是一种基于图像语义分割任务的数据增强方法，通常用于解决在语义分割任务中的稀有类别和小目标等问题，基本思想是通过复制粘贴的方式，从一张图像中复制粘贴特定的目标物体到另一张图像中，快速、简单地生成大量的、更复杂的新训练样本，可以有效地扩充图像训练集，可以在不增加显著计算资源的情况下提高训练效果，提高模型的性能，并且能够有效模拟遮挡干扰，使模型更好地学习到不同位置的图像特征。

本实验引入了CopyPaste的思想来解决最为困难的遮挡干扰识别问题。对于语义分割任务，CopyPaste需要使用标注好的像素级别的分割掩码，以便将物体从原始图像中裁剪出来，并精确地贴到新的背景图片上。为了适应目标检测任务的需要，实验中对CopyPaste的代码进行改进，提出基于目标检测的CopyPaste，只需要使用标注好的物体检测框就能够完成目标检测数据增强，生成具有不同背景、物体尺寸、位置和姿态等属性的新训练样本，帮助模型更好地学习图像特征，从而提高模型的泛化能力和鲁棒性，如图3-1所示。



图3-1 基于目标检测的CopyPaste

基于目标检测的CopyPaste基本流程如下：

1. 输入一组图像数据集及其相应的检测框信息。
2. 通过概率参数prob决定是否使用该方法进行数据增强，并通过参数num控制每个图像增强的次数，可以根据需要灵活的选择。
3. 从输入的数据集中随机选择一个具有检测框标注的源图像，提取该源图像中的所有检测框和标签，并选择其中某个检测框进行处理。
4. 从目标图像中随机选择一个目标位置，将源图像中所选对象直接粘贴到目标图像的目标位置中，并根据源图像所选对象的大小进行调整，从而生成新的图像。
5. 添加粘贴图像后产生的检测框和标签，更新目标图像的检测框标注信息。
6. 重复该过程，返回增强后的num个图像结果。

改进后的基于目标检测的CopyPaste算法仅需输入检测框信息即可确定所需复制对象的位置和大小，适应目标检测任务的实际应用，同时通过随机参数与控制循环次数的方式提高算法效率，避免出现过拟合，这一优化使得该算法更容易使用且具有实际应用意义，能够在突破遮挡干扰提升瓶颈方面产生重要作用。

**3.2 模型集成**

**3.2.1 改进的WBF**

模型集成是机器学习领域中的一种技术，其基本思想是将多个模型的预测结果进行融合，可以通过多种方式实现，例如投票法、加权投票法、平均法、加权平均法、堆叠泛化法等。其优势在于可以从不同角度对同一个问题进行建模，并从各模型之间互补的不足中得到更好的表现，借助多个模型之间的交互和协作来提高整体性能。

具有较高偏差的单个模型不能有效地拟合训练数据，通过采用将多个模型进行集成的方式，可以得到一个更加稳定的集成模型，提高整体预测的准确性，从而提高整个模型的性能。具有较高方差的单个模型可能在测试数据上表现不佳，通过集成多个模型，可以使其更具有鲁棒性。单个模型可能会过度拟合训练数据，因此在新数据上的泛化性能会下降，将多个模型的预测结果进行组合，可以缓解这种过拟合现象，提高整体模型的泛化能力。

加权边界框融合[16]（Weighted Boxes Fusion, WBF）是一种基于目标检测模型的预测结果进行加权融合的方法，能够在图像中提出一组候选框，每个候选框都包含一个边界框和相应的置信度得分，对每个检测器和输入图像预测一组边界框和其置信度得分，将每个检测器的边界框集合合并成一个总的边界框集合，可以更充分地利用多个模型的预测信息，有效减少虚假检测和漏检，从而提高模型的检测性能。



图3-2 改进的加权边界框融合WBF

为了有效利用各模型对某种干扰的优势，实验中引入了加权边界框融合WBF的后处理方法，并对传统的WBF进行了改进，如图3-2所示，改进的WBF算法具有以下功能：

引入了对检测结果的预处理，去除置信度低于指定阈值的预测框，并对框的坐标进行规范化。对每个模型的检测结果，遍历每个边界框，如果该框的置信度低于给定的阈值，则将该框剔除；如果该框的坐标值不符合规范，例如x2<x1或者y2<y1，则对其进行修正；如果该框的面积为0，则将其剔除。引入对检测结果的预处理可以有效地过滤掉一些低置信度的检测结果和错误的坐标值，从而提高后面加权融合的准确性。

可以为每个模型赋予不同的权重因子，表示每个模型对于融合结果的贡献权重，通过调整权重因子，更好地平衡不同模型之间的贡献和影响，使得最终融合结果更加准确和可靠。由于不同模型在不同数据集上的训练精度不同，或者对于某些类别的检测性能表现更好等原因，需要对权重进行调整，可以将较高的权重值分配给表现更好的模型，从而更加突显其在融合结果中的作用，例如图3-2所示蓝色检测框。

支持多种加权融合方式，包括平均值法、最大值法和加权融合法，提高了算法的灵活性和实用性，可以根据实际情况选择不同的方式进行融合。最简单的平均值法直接将所有预测框求平均作为融合结果；最大值法直接使用置信度值最大的预测框作为融合结果；加权融合法基于平均值法，额外考虑了不同模型预测的差异，为每个模型赋予相应的权重。

此外，通过使用numpy数组计算交并比，取代了原来的循环操作，从而极大地提高了算法的效率，在内存占用方面也有一定的优势，特别是对于大规模的数据集，在计算速度上获得了明显的提升。

综上所述，改进的WBF算法可以提供更具有可调节性和适应性的融合方案，提高算法对于不同检测模型和不同干扰数据的适应能力，还可以大幅提高算法的速度和效率，同时降低内存占用，使得算法更加实用和高效，有利于综合应对复杂多样的干扰。

**3.3 半监督学习**

**3.3.1 带处理伪标签**

半监督学习是一种介于有监督学习和无监督学习之间的机器学习方法，通过结合有限的带标签数据和大量未标记的无监督数据来进行模型的训练，可以更好地利用数据资源，提高模型的泛化能力和鲁棒性。有监督学习需要使用大量标注数据来训练模型，但获取困难会导致训练数据不足，从而影响模型的性能，而无监督学习则利用无标签的数据来训练模型，但其效果往往比较有限。

本实验引入了一种简单有效的深度神经网络半监督学习方法——伪标签。通过使用已有模型对未标记的数据进行推断，得到每个物体的类别和位置，将这些类别和位置信息作为伪标签，与已有标注数据一同送入模型进行训练，在训练过程中，模型会根据这些伪标签调整自身的参数，使得模型具备更强的特征提取能力，可以解决训练数据数量和多样性不足的问题，提高模型的泛化性能，进而提高模型的检测效果，增加模型的鲁棒性。

为了缓解伪标签可能存在的问题，比如输入太多置信度较低的预测结果干扰模型性能的提升，以及直接导入伪标签数据可能会导致过拟合现象，在实验中对伪标签进行了筛选，并进行了数据增强的处理，使用带处理伪标签与原始数据集构成新训练数据集，如图3-3所示。

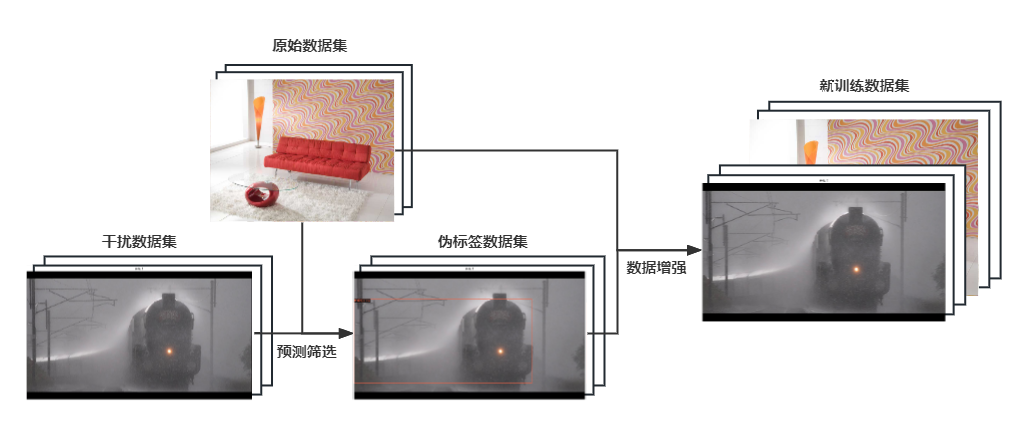


图3-3 带处理伪标签

带处理伪标签同样引入了对预测结果的预处理，去除置信度低于指定阈值的预测框，并对框的坐标进行规范化。对每个经过预测得到伪标签的图像，遍历每个边界框，如果该框的置信度低于给定的阈值，则将该框剔除；如果该框的坐标值不符合规范，例如x2<x1或者y2<y1，则对其进行修正；如果该框的面积为0，则将其剔除。引入对伪标签的预处理能够有效减少干扰信息进入训练过程，提高模型的训练效果。

同时利用数据增强避免过拟合，通过改变现有数据增强方法的参数，对伪标签数据集进行更大程度的变换，形成差异较大的新数据集。例如：增加Cutout的矩形框数量，删除更多区域；将CopyPaste参数设定为较高概率使用该方法进行遮挡；调整Albumentations中Shift Scale Rotate的参数，进行幅度更大的平移、缩放和旋转等随机变换。通过数据增强破坏伪标签数据集原始形态，一定程度上能够有效抑制过拟合现象的发生。

带处理的伪标签不但可以像传统伪标签一样，解决训练数据数量和多样性不足的问题，提高模型的泛化性能和检测效果，增加模型的鲁棒性。还可以保证数据集的纯净性和准确性，进而提高模型的训练效果，并且不是仅仅记忆了伪标签数据集中的样本信息，进一步避免过拟合现象的发生，可以使得模型在真实场景下的表现更加鲁棒，在实验过程都采用了带处理伪标签的方法。

【缺本章小节】

**第四章 实验结果**

**4.1 实验过程**

**4.1.1 Cascade R-CNN**

如表4-1所示，首先对Cascade R-CNN、DyHead、TOOD、VFNet、DDOD进行了测试，所用的Backbone为ResNet-50，Neck为FPN，迭代次数为12 epochs。

表4-1 各Head的平均精度均值及运行时间

|  |  |  |
| --- | --- | --- |
| 方法 | 平均精度均值 | 运行时间 |
| Cascade R-CNN | 58.16% | 3h |
| DyHead | 56.98% | 5.5h |
| TOOD | 53.40% | 2.5h |
| VFNet | 49.90% | 3h |
| DDOD | 35.40% | 2.5h |

在面对多种干扰时，Cascade R-CNN平均精度均值最高为58.16%，并且运行时间相对较短，因此选择针对干扰数据预测最优的模型Cascade R-CNN进行实验。

Cascade R-CNN通过逐级提高阈值来筛选出真正的物体，逐步提高检测器的准确性，有效消除背景的干扰，更好地聚焦于物体。并且不同级联之间的训练数据共享，不但在维护模型的原有速度和精度上做出了平衡，而且可以大大减少训练所需的训练样本，在较低的计算复杂度下实现高质量的目标检测。除了提高检测性能以外，还有助于减轻过拟合的风险，增加模型的泛化能力，提高在新环境下的稳定性。

**4.1.2 ResNet-101 64x4d**

如表4-2所示，列出了ResNet-50、ResNet-101、ResNet-101 64x4d三种[Backbone](http://www.baidu.com/link?url=o4C7GKSAHAKV4S82lE3YmFY9P5M5EdDFxHm4pXqHvOS3wD646rlHbgx1fdi6PZbb2FKbafiJhfplEbJH_TaG3G7v2MIwRLq4mkj28KWVByW" \t "https://www.baidu.com/_blank)的实验结果。

表4-2 各[Backbone](http://www.baidu.com/link?url=o4C7GKSAHAKV4S82lE3YmFY9P5M5EdDFxHm4pXqHvOS3wD646rlHbgx1fdi6PZbb2FKbafiJhfplEbJH_TaG3G7v2MIwRLq4mkj28KWVByW" \t "https://www.baidu.com/_blank)的平均精度均值及运行时间

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 骨干 | 平均精度均值 |  | 学习率 | 运行时间 |
| ResNet-50 | 58.16% |  | 0.02 | 3h |
| ResNet-101 | 59.29% |  | 0.02 | 3.5h |
| ResNet-101 64x4d | 61.33% |  | 0.01 | 8h |

ResNet-50只具有50个卷积层，是一种较浅的网络，因此平均精度均值最低；ResNet-101拥有101个卷积层，相对提高了约1.14%的平均精度均值。而ResNet-101 64x4d则在ResNet-101的基础上，进一步进行了优化，增加了通道数并减少了卷积核的大小，这种方法进一步提高了约2.03%的平均精度均值。由于ResNet-101 64x4d经过了更多次的迭代，并且运行空间的限制迫使ResNet-101 64x4d采用了更小的学习率0.01，所需时间更长。

ResNet-101 64x4d在平均精度均值上明显优于其他方法，通道数的增加使得模型特征更加丰富，卷积核大小的减少使得模型更加精细，这些优化措施可以使得模型在训练过程中更好地捕获信息，从而提高准确性。

将使用了ResNet-101 64x4d的模型作为Baseline，对Cascade R-CNN的IoU阈值进行调整，对于形状、姿势、纹理、上下文、天气和遮挡的平均精度如表4-3所示。

表4-3 各干扰类型平均精度

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 方法 | 形状 | 姿势 | 纹理 | 上下文 | 天气 | 遮挡 |  | mAP |
| Baseline | 65.7% | 62.1% | 66.6% | 67.0% | 70.3% | 36.1% |  | 61.33% |
| +IoU调参 | 67.9% | 65.3% | 69.2% | 68.4% | 71.8% | 38.6% |  | 63.55% |

可以发现在六种干扰中，模型对于遮挡的平均精度远远低于其他类型干扰，明显拉低了平均精度均值，因此推测改善模型对于遮挡的平均精度，能够有效提升平均精度均值。

**4.1.3 Cutout**

将对Cascade R-CNN的IoU阈值进行调整后的平均精度结果作为新的Baseline，对比加入Cutout后对遮挡和平均精度均值的影响，如表4-4所示。

表4-4 Cutout对遮挡和平均精度均值的影响

|  |  |  |  |
| --- | --- | --- | --- |
| 方法 | 遮挡 |  | 平均精度均值 |
| Baseline | 38.6% |  | 63.55% |
| +Cutout | 48.1% |  | 66.31% |

从表格数据中可以看出，Cutout针对遮挡的平均精度提高了9.5%，并且对六种干扰的平均精度均有提升，使得平均精度均值提高了2.76%。

Cutout让模型不仅关注到样本的主要特征，同时也学习到背景和噪声等特征信息，当图像中的对象具有不同的尺寸和形状时，可以学习到更多的位置和形态信息。在识别复杂物体或场景时，可以提高模型对于复杂因素的鲁棒性，帮助模型更好地学习到对象的关键特征，让模型更加具有泛化能力，增加数据多样性，避免模型过拟合，并提高预测精度。

**4.1.4 Mixup**

将加入Cutout后的平均精度结果作为新的Baseline，对比加入Mixup后对遮挡和平均精度均值的影响，如表4-5所示。

表4-5 Mixup对遮挡和平均精度均值的影响

|  |  |  |  |
| --- | --- | --- | --- |
| 方法 | 遮挡 |  | 平均精度均值 |
| Baseline | 48.1% |  | 66.31% |
| +Mixup | 61.9% |  | 71.21% |

从表格数据中可以看出，Mixup针对遮挡的平均精度提高了13.8%，并且对六种干扰的平均精度均有提升，使得平均精度均值提高了4.9%。

Mixup能够通过对训练数据进行随机混合，产生新的样本来增加数据集的大小，从而最大化地利用数据集中的信息，使得模型在学习到标准的训练数据的同时，学习到更广泛的数据分布，从而更好地适应新的输入数据，提高模型的鲁棒性和泛化能力。并且由于每个样本都是通过线性插值进行组合，特征会更加平滑，因此可以减少对标签噪声的敏感度。

**4.1.5 Soft NMS**

对比非极大值抑制NMS和软性非极大值抑制Soft NMS对平均精度均值的影响，以及参数对平均精度均值的影响，如表4-6所示。

表4-6 Soft NMS以及参数对平均精度均值的影响

|  |  |  |
| --- | --- | --- |
| 方法 | 参数 | 平均精度均值 |
| NMS | 0.05 | 71.21% |
| Soft NMS | 0.05 | 71.15% |
| NMS | 0.005 | 71.79% |
| Soft NMS | 0.005 | 71.82% |

当阈值取0.05时，NMS和Soft NMS的平均精度均值都比较低，只有71.21%和71.15%。而当阈值减小至0.005时，NMS和Soft NMS的平均精度均值明显提高，可以达到71.79%和71.82%。这是因为过高的阈值会导致过多的框被舍弃，而较低的阈值则会使得较多的框被保留，从而影响检测准确性。

综合来看，Soft NMS相比于传统的NMS方法在一定程度上能够降低检测中的错误框，所以在阈值较小的情况下表现更好。并且Soft NM方法非常直观简单、容易实现，有效解决了传统NMS方法中存在漏检、重复检测等问题。

**4.1.6 目标检测CopyPaste**

将参数为0.005的Soft NMS的平均精度结果作为新的Baseline，并对比加入基于目标检测的CopyPaste对遮挡和平均精度均值的影响，如表4-7所示。

表4-7 CopyPaste 对遮挡和平均精度均值的影响

|  |  |  |  |
| --- | --- | --- | --- |
| 方法 | 遮挡 |  | 平均精度均值 |
| Baseline | 60.7% |  | 71.82% |
| +目标检测CopyPaste | 74.6% |  | 71.96% |

从表格数据中可以看出，基于目标检测的CopyPaste针对遮挡的平均精度提高了13.9%，对其他五种干扰的平均精度稍有下降，平均精度均值提高了0.14%。

改进后的目标检测CopyPaste通过模拟遮挡干扰增加复杂度，可以有效地保留图像的特征信息，能够。使得训练集中的图像更加丰富，提高了模型的泛化能力，从而提升模型性能。并且只需要对已有数据进行复制粘贴操作，不需要增加过多计算资源，相对于使用传统数据增强方法数据增长的比例更高。同时非常容易实现和操作。

**4.1.7 Albumentations**

将加入CopyPaste后的平均精度结果作为新的Baseline，对比加入Albumentations后对于六种干扰的平均精度和平均精度均值的影响，如表4-8所示。

表4-8 Albumentations对各干扰类型平均精度的影响

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 方法 | 形状 | 姿势 | 纹理 | 上下文 | 天气 | 遮挡 |  | mAP |
| Baseline | 69.5% | 66.6% | 73.5% | 69.5% | 78.0% | 74.6% |  | 71.96% |
| +Albumentations | 68.9% | 66.8% | 72.9% | 70.2% | 78.8% | 75.2% |  | 72.14% |

使用Albumentations的平均精度均值有一定程度的提高，约0.18%，说明加入增强数据对模型的性能有积极的影响。其中，对于姿势、上下文、天气和遮挡等因素，Albumentations的平均精度有不到1%的提升，相较于Baseline有所提高。

Albumentations可以高效地实现丰富多样的图像转换操作，并提供了简洁而强大的图像增强接口，经过了针对性的优化以提高目标检测的性能，具有高效的计算速度，以进一步提高图像增强的效率，通过增加样本数量和数据多样性来提高模型的鲁棒性，从而达到更好的精度和泛化性能。

**4.1.8 标签平滑**

将加入Albumentations后的平均精度结果作为新的Baseline，对比加入标签平滑后对于六种干扰平均精度和平均精度均值的影响，如表4-9所示。

表4-9 标签平滑对各干扰类型平均精度的影响

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 方法 | 形状 | 姿势 | 纹理 | 上下文 | 天气 | 遮挡 |  | mAP |
| Baseline | 68.9% | 66.8% | 72.9% | 70.2% | 78.8% | 75.2% |  | 72.14% |
| +标签平滑 | 69.0% | 66.9% | 73.1% | 71.3% | 78.6% | 75.0% |  | 72.32% |

使用标签平滑的平均精度在形状、姿势、纹理和上下文方面都有所提高，其中在上下文方面提高最明显，提高了1.1%，总的平均精度均值增加了0.18%。

采用标签平滑技术会减少真实样本标签的权重，同时将一部分权重分配给其他标签，使得模型不会在训练时过度自信地预测标签，让模型对于噪声和干扰更具有鲁棒性，减缓模型对于少数样本过拟合的风险，从而增强模型的泛化能力，提高目标检测的准确性。

将加入标签平滑后的平均精度结果作为新的Baseline，对比改进Albumentations中旋转、平移、缩放（Shift Scale Rotate）操作后对于六种干扰平均精度和平均精度均值的影响，如表4-10所示。

表4-10 Shift Scale Rotate对各干扰类型平均精度的影响

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 方法 | 形状 | 姿势 | 纹理 | 上下文 | 天气 | 遮挡 |  | mAP |
| Baseline | 69.0% | 66.9% | 73.1% | 71.3% | 78.6% | 75.0% |  | 72.32% |
| +Shift Scale Rotate | 69.7% | 67.0% | 73.3% | 71.4% | 78.3% | 75.9% |  | 72.61% |

Shift Scale Rotate的平均精度均值提高了0.29%，对形状和遮挡两种干扰的平均精度提升相对比较明显，其中针对形状的平均精度提高了0.7%，针对遮挡的平均精度提高了0.9%，说明平移、缩放和旋转有助于应对形状和遮挡干扰。

通过采用旋转、平移、缩放等方式对图像进行随机变换，模拟实际情况下的不确定性和复杂性，可以有效地增加数据集的多样性，减轻过拟合问题，使得模型具有更强的鲁棒性和泛化能力，更好地适应各种场景，从而提高目标检测的准确率和稳定性。

**4.1.10 改进的WBF**

将加入Shift Scale Rotate后的平均精度结果作为新的Baseline，对比使用了改进的加权边界框融合WBF后对于六种干扰的平均精度和平均精度均值的影响，如表4-11所示。

表4-11 WBF对各干扰类型平均精度的影响

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 方法 | 形状 | 姿势 | 纹理 | 上下文 | 天气 | 遮挡 |  | mAP |
| Baseline | 69.7% | 67.0% | 73.3% | 71.4% | 78.3% | 75.9% |  | 72.61% |
| +改进的WBF | 75.3% | 75.6% | 76.4% | 76.3% | 81.7% | 79.0% |  | 77.40% |

改进的加权边界框融合WBF对六种干扰的平均精度均有明显提升，平均精度均值提高了4.79%。其中针对姿势的平均精度提高了8.6%，原先姿势作为平均精度最低的一项，与第二低的形状相差2.7%，与最高的天气相差11.3%，加入改进的WBF后，姿势比形状高出0.3%，与最高的天气差距缩小至6.1%。

在单个模型的后期优化中，提升某类干扰的平均精度通常会降低其他干扰的平均精度，导致整体平均精度均值的下降，改进的WBF可以有效利用各模型对不同干扰的优势，显著提高目标检测算法应对复杂多样干扰的性能，获得更准确的检测结果，并且具有高度灵活性和适应性，可以通过改变权重方式来适应不同的目标检测任务，适用于各种不同的目标检测场景，在实际应用中能够发挥巨大作用。

**4.2 实验总结**

**4.2.1 提升效果**

总结模型训练中使用的方法对平均精度均值的影响，如表4-12所示。

表4-12 各方法对平均精度均值的影响

|  |  |
| --- | --- |
| 方法 | 平均精度均值 |
| Cascade R-CNN | 58.16% |
| +ResNet-101 64x4d | 61.33% |
| +IoU调参 | 63.55% |
| +Cutout | 66.31% |
| +Mixup | 71.21% |
| +NMS调参0.005 | 71.79% |
| +Soft NMS | 71.82% |
| +目标检测CopyPaste | 71.96% |
| +Albumentations | 72.14% |
| +标签平滑 | 72.32% |
| +Shift Scale Rotate | 72.61% |
| +改进的WBF | 77.40% |

在以选用的Cascade R-CNN作为基准的情况下，使用ResNet-101 64x4d相较于原骨干ResNet-50提高了3.17%的准确率，继续使用IoU调参进一步提高了2.23%。加入Cutout方法使得平均精度均值提高了2.76%，再加上Mixup方法让性能提高达到了71.21%，大幅度提升了4.9%，随后平均精度均值的增加情况变得较为缓慢。通过对NMS调参为0.005接着改用Soft NMS分别提高了0.19%和0.03%，使用基于目标检测的CopyPaste方法增强模型应对遮挡的能力，同时让性能再次提高0.14%，Albumentations和标签平滑都进一步提高了0.18%的准确率，Shift Scale Rotate方法提高了0.29%，最后采用改进的WBF进行后处理，综合各模型对不同干扰的优势，平均精度均值大幅度提升了4.79%。

以Cascade R-CNN作为基准，平均精度均值累计提高了19.25%，在面临存在多种干扰的环境和场景时，能够保持较好的表现和准确率，有效地提高目标检测模型的鲁棒性。

【缺本章小节】

**第五章 总结与展望**

**5.1 本文工作总结**

**5.1.1 研究及结果**

本次研究主要是针对深度学习中的目标检测鲁棒性问题进行优化，重点关注了针对形状、姿势、纹理、上下文、天气和遮挡六种干扰的鲁棒性提升问题，旨在提高目标检测算法的鲁棒性，减少漏检和误检情况。基于Cascade R-CNN、ResNet-101 64x4d网络，并在改进的半监督学习带处理伪标签下进行优化，通过多个优化方法来提高模型的性能，采用数据增强、模型优化、加权融合等技术手段，有效提高目标检测算法在复杂场景下的鲁棒性。

采用Cutout、Mixup、Albumentations图像增强库等方法对数据进行增强，并且使用了基于目标检测的CopyPaste，这些方法可以有效地扩充数据集，增加训练数据的多样性，降低过拟合风险，提高模型的性能。

对IOU参数、NMS参数和Soft NMS参数进行了调优，主要通过调整检测框合并算法的参数，以优化模型在不同尺度、密度场景下的检测效果，提高模型识别的准确率，从而提高模型的准确度和鲁棒性。

通过标签平滑方法对标签进行平滑处理，缓解模型对训练集中噪声标签的过拟合现象，提高模型的泛化性能，有助于鲁棒性的提升。

使用改进的加权框融合WBF解决单个模型容易过拟合的问题，将多个模型的输出结果进行融合，结合各模型对于不同干扰的优势，提高模型的准确度和鲁棒性，从而提升整体的检测效果，可以有效提高模型的鲁棒性。

图4-1 平均精度均值提升折线图

如图4-1所示，实验结果表明，经过一系列的优化操作后，模型的平均精度均值逐步提升，以Cascade R-CNN作为基准总共提高了19.25%，对于包含六种干扰的复杂场景的综合识别有显著改善，鲁棒性得到了显著提高。实验中提出了目标检测CopyPaste、改进的WBF、带处理伪标签，这些优化可以为目标检测算法的实际应用提供有益参考。

**5.1.2 研究感悟**

在实验设计的过程中，由于相关知识储备和实践经验的不足，遇到了很多困难和挑战。算法本身具有一定的的复杂性，除了需要花费时间理解其原理，更多的精力消耗在代码实现和修改过程中，并且因为每一次模型训练的运行时间都比较长，试错成本较高，通常需要在每一次训练、预测和评价。

这些挑战也激发了丰富知识和提高能力的动力，在实验的过程中继续学习相关资料，包括但不限于官方文档、博客、教程等，并且充分研读实验代码，探究参数对算法性能的影响，寻找更加合适的超参数组合，提高模型的性能表现。有效利用实验数据和日志，对比原始算法和优化算法在精确度和运行速度方面的表现，通过实验结果的分析，评估优化算法的效果和可行性，解决遇到的问题和困难，积极开展研究工作，争取更好的优化效果。

在此过程中我也发现，目标检测领域研究的发展进步不仅注重算法本身的改进，还注重算法的简化和优化，使其易于理解和实现，越来越多的学习资源和平台，降低了学习门槛，使得初学者能够快速地掌握相关知识和技能，为目标检测领域的发展注入新的活力和动力，从而推动相关领域的发展。

相信随着技术的不断发展，目标检测领域的研究将会越来越成熟和完善，在未来的学习和工作中还要继续努力和创新，相信通过不断地学习和实践，能够更好地理解并掌握更多先进的技术手段和方法，不断探索本领域的前沿技术和思想，为解决实际问题提供更多的道路和方法，为社会的进步和发展作出自己的贡献。

**5.2 未来工作展望**

**5.2.1 问题挑战**

目标检测作为计算机视觉领域的重要研究任务之一，在实际应用中仍存在一些鲁棒性问题。不同目标光照、角度、大小等因素的变化，图像中噪声、模糊等因素，以及目标与背景之间的相似性、遮挡情况等因素，都会对目标检测造成影响，干扰目标检测的准确性。目标检测鲁棒性研究的难点和挑战主要表现在以下几个方面。

为了更好的性能，目标检测算法的鲁棒性研究需要丰富的数据集支持，但获取大量带有标签的数据样本需要做很多的工作，很多应用场景根本无法获取，小样本检测精度不理想。而且获取的数据样本通常存在不均衡的情况，在现实生活中某些物体可能比其他物体更容易被遮挡或者受到干扰。数据集的不充分性和不均衡性会导致模型不能学习到足够的特征，具有较弱的鲁棒性。

除此之外，小目标的检测[17]效果总是很差，主要原因在于小目标图像分辨率低，导致特征表达能力弱。并且在训练数据集中，只有少数图像包含小物体，容易导致检测器偏向于关注中型和大型物体。而且小物体覆盖区域小，位置缺乏多样性。在自动驾驶等领域，缺乏深度信息的二维图像无法有效避免碰撞。目标探测器对密集分布的物体检测[17]效果较差，主要是因为物体通常在各个方向上排列成不同的形状。很难确定相邻相似对象的预测框的位置。密集分布的对象在NMS处理后很容易被视为单个对象。

综上所述，目标检测鲁棒性研究中存在很多的难点和挑战，在未来的研究中，需要从实际应用场景出发综合考虑这些方面，提升目标检测鲁棒性，从而增强准确性和稳定性。

**5.2.2 未来趋势**

目标检测鲁棒性研究是计算机视觉领域中极具挑战和发展前景的一个方向，目前有几种常见的解决方法，使用图像增强技术对原始图像进行变换和扰动来生成具有更高鲁棒性的图像，开发和改进新型的网络结构来提高模型在复杂环境下的鲁棒性，改善数据采集和标注环节数据集的质量和多样性，这些在未来仍然是重要的研究方向

未来将会探索更多的新思路和方法提高鲁棒性，从而实现更为准确可靠的目标检测，有效的应用在实际场景中。目标检测的数据来源包含了传感器、图像、视频等多种方式，每种数据都有各自的优缺点，由于单一模态的数据通常不能全面的表达目标信息，可以采用多模态融合的方式提高目标检测的鲁棒性。弱监督学习[18]可以利用相对较少的监督信息来训练模型，降低了标注数据的成本，扩大了数据集规模，增加了训练数据的多样性，对CAM方法的改进不需要实时准确地标注每个目标的具体位置，降低了噪声和标注错误的影响，从而提高了模型的鲁棒性。过去的鲁棒性测试往往针对特定干扰设计，为目标检测鲁棒性构建全面综合性的基准[19]，提供更具挑战的测试环境、更加标准化的评估方法，可以推动目标检测算法的改进和发展。此外，还需要更多的技术手段来提高模型的可解释性，帮助研究人员更好地理解模型的运行过程，从而更好地提高算法的鲁棒性。甚至可以模型的视觉鲁棒性[20]，通过分析人类视觉系统的特点和优势，提出更接近人类视觉机制的机器学习模型和算法，寻找提升目标检测等计算机视觉领域鲁棒性的新方法，

相信随着计算机硬件和算法的不断进步，目标检测技术将在更多应用场景中得到应用，并发挥其更大的社会价值。同时在鲁棒性问题上取得更大的突破，更加准确、快速地判断目标信息，实现复杂场景下全天候、多角度、多目标的定位和分类，有效应用于智慧城市、医疗、安防等，带来更加智能、便捷且安全的生活。