基于对比学习的改进 SSD 目标检测算法

胡 焱，原子昊，涂晓光，刘建华，雷 霞，王文敬
（中国民用航空飞行学院 航空电子电气学院，四川 广汉 618307）

摘要：现有基于深度学习的目标检测算法在图像的目标检测过程中存在物体视角的多样性、目标本身形变、检测物体受遮挡、光照性以及小目标检测等问题。为了解决这些问题，本文将对比学习思想引入到 SSD（Single Shot MutiBox Detectior）目标检测网络中，对原有的 SSD 算法进行改进。首先，通过采用图像截块的方式随机截取样本图片中的目标图片与背景图片，将目标图像块与背景图像块输入到对比学习网络中提取图片特征进行对比损失计算。随后，使用监督学习的方法对 SSD 网络进行训练，将对比损失传入到 SSD 网络中与 SSD 损失值加权求和反馈给 SSD 网络，进行网络参数的优化。由于在目标检测网络中加入了对比学习的思想，提高了背景和目标在特征空间中的区分度。因此所提出的算法能显著提高 SSD 网络对于目标检测的精度，并在可见光和热红外图像中均取得了令人满意的检测效果。在 PASCAL VOC2012 数据集实验中，AP50 值提升了 0.3%，在 LLVIP 数据集实验中，AP50 值提升了 0.2%。

关键词：深度学习；SSD；目标检测；对比学习；红外热成像；图像截块
中图分类号：TP391.41 **文献标识码**：A **文章编号**：1001-8891(2024)05-0548-08

Improved SSD Object Detection Algorithm Based on Contrastive Learning

HU Yan，YUAN Zihao，TU Xiaoguang，LIU Jianhua，LEI Xia，WANG Wenjing
(*Institute of Electronic and Electrical Engineering, Civil Aviation Flight University of China, Guanghan* 618307, *China*)

Abstract：The existing deep learning-based object detection algorithms encounter various issues during the object detection process in images, such as object viewpoint diversity, object deformation, detection occlusion, illumination variations, and detection of small objects. To address these issues, this paper introduces the concept of contrastive learning into the SSD object detection network and improves the original SSD algorithm. First, by randomly cropping object images and background images from sample images using the method of image cropping, the object image blocks and background image blocks are input into the contrastive learning network for feature extraction and contrastive loss calculation. The supervised learning method is then used to train the SSD network, and the contrastive loss is fed into the SSD network and weighted and summed with the SSD loss value for feedback to optimize the network parameters. Because the contrastive learning concept is introduced into the object detection network, the distinction between the background and object in the feature space is improved. Therefore, the proposed algorithm significantly improves the accuracy of the SSD network for object detection, and obtains satisfactory detection results in both visible and thermal infrared images. In the experiment on the PASCAL VOC2012 dataset, the proposed algorithm shows an increase in the AP50 value by 0.3%, whereas in the case of the LLVIP dataset, the corresponding increase in AP50 value is 0.2%.

Key words：deep learning, SSD, object detection, contrastive learning, infrared thermal, image cropping

0 引言

目标检测是计算机视觉领域的一项重要研究课题[1]，其目的是要从一幅场景（图片）中找出目标并

收稿日期：2023-05-18；**修订日期**：2023-07-11.
作者简介：胡焱（1973-），男，四川大英人，教授，硕士生导师，研究方向：航空电子设备维修、测控。E-mail：huyan@cafuc.edu.cn。
通信作者：原子昊（1999-），男，河南焦作人，硕士研究生，主要从事计算机视觉、深度学习目标检测的研究。E-mail：769606514@qq.com。
基金项目：中国博士后科学基金（2022M722248）；四川省无人系统智能采集控制技术工程实验室开放课题（WRXT2021-001）；民航飞行技术飞行与安全重点实验室开放项目资助（FZ2022KF06）；中国民用航空飞行学院面上项目（J2023-026）；中央高校基本科研业务费（ZHMH2022-004, J2022-025）。

对其进行定位，包括检测和识别两个过程[2]。现如今，目标检测已经广泛地应用于我们的日常生活当中，如自动驾驶、机器人视觉、视频监控等领域[3]。

现阶段深度学习的科研进展突飞猛进，卷积神经网络在目标检测这一领域也被充分应用并获得了不小的成就。目前主流的目标检测算法划分为两大类：基于边界框回归的一阶段目标检测算法和基于候选区域的两阶段目标检测算法。一阶段目标检测算法在目标图片上做边界框且实施分类回归，如 SSD（Single Shot MutiBox Detectior）算法[4]、YOLO（You only look once）系列算法[5-8]等。两阶段目标检测在图像上生成候选区的基础上，在图片候选区域中进行特征提取后进行目标分类回归，如 R-CNN（Region-based Convolutional Neural Network）算法[9]、Fast R-CNN（Fast Region-based Convolutional Neural Network）算法[10]、FasterR-CNN（Faster Region-based Convolutional Neural Network）算法[11]等。目前这些经典的有监督学习目标检测算法对于目标的检测已经取得了不错的效果。

然而现如今有监督学习目标检测算法依然存在着种种挑战，例如检测物体视角的多样性问题、目标本身形变问题、检测物体受遮挡问题、光照性问题以及小目标检测问题等都会对感受野提取图片特征造成一定的麻烦。以自监督学习为主的对比学习算法（Contrastive Learning, CL）根据图像本身通过对比图片之间的相似度来进行网络训练，这和目标检测过程中判断目标和背景的差异性思想是一致的，因此，可以设想在目标检测算法中引入对比学习的思想，提取和背景差异更加明显的目标特征，将会有效地提升模型的检测效果[12-13]。图 1 为将对比学习思想引入目标检测网络中的图示。

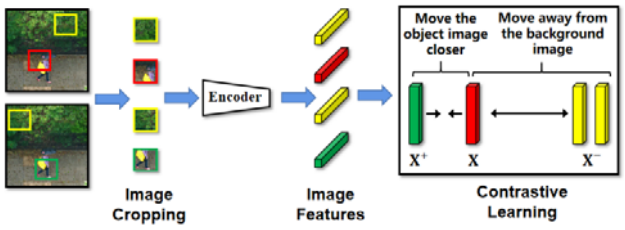

图 1 将对比学习思想引入目标检测网络中的图示

Fig.1 An illustration of integrating the concept of contrastive learning into object detection networks

SSD 是一种基于深度神经网络的目标检测算法，它在许多实际应用中取得了成功。SSD 是一种单阶段的检测器，不需要使用候选区域来生成可能包含物体的区域。这个特点使得 SSD 的速度非常快，可以实时地处理大量图像和视频数据[14]。SSD 算法使用了多层的卷积神经网络来提取图像特征，这些特征能够准确地描述不同物体的形状和纹理。此外，SSD 还采用了多个尺度的特征图来检测不同大小的物体，这也有助于提高检测精度。SSD 算法使用了多尺度的卷积特征图，这使得它能够有效地检测小物体。与其他目标检测算法相比，SSD 在检测小物体方面有很大的优势。另外，SSD 可以通过使用多个尺度的特征图来检测不同大小和长宽比的物体，这使得它能够在不同场景下具有更好的适应性和鲁棒性。综上所述，SSD 是一种速度快、精度高、对小目标检测效果好、对小物体和不同长宽比的物体具有良好检测效果的目标检测算法。因此，本文选择 SSD 目标检测算法作为改进算法的主体算法，用来和对比学习思想进行融合。

1 相关工作

1.1 目标检测算法

计算机视觉中的一个基本任务是目标检测，它涉及在图像或视频中识别对象的存在和位置。目标检测算法经过多年发展，分为两类：两阶段算法，如 R-CNN[9]、Fast R-CNN[10]和 Faster R-CNN[11]，以及一阶段算法，如 SSD[4]和 YOLO 系列[5-8]。虽然 R-CNN 相对于传统算法有了显著的改进，但在 CNN 中进行候选区域框计算导致计算量增加，严重影响了测试速度。Fast R-CNN 减少了计算量，但仍无法实现真正的实时性或端到端的训练和测试。因此，提出了 Faster R-CNN，将特征提取、候选框选择、分类和边界框回归集成到一个框架中，提高了准确性和速度，并实现了端到端的目标检测。然而，实时目标检测与 Faster R-CNN 之间仍存在差距，导致出现了 SSD 和 YOLO 等一阶段算法。虽然 YOLO 系列将目标检测作为回归问题解决，但与 Faster R-CNN 相比存在定位误差。YOLOv2 在保持速度优势的同时改进了原始算法，而 YOLOv3 使用深度残差网络提取图像特征。

许多学者针对 SSD 算法进行了改进，以提升其检测能力。例如，Fu 等人提出了 DSSD（Deconvolutional Single Shot Multibox Detector）模型[15]，通过增加反卷积模块来融合上下文信息，但由于网络结构较复杂，计算量大，导致检测速度显著降低。Jeong 等人提出了 RSSD 模型[16]，改进了特征融合方式，增加了有效的特征图用于检测，充分利用了模型特征，检测性能与 DSSD 持平。Li 等人在 SSD 中引入 FocalLoss 损失函数[17]，降低了易分类样本的权重，使模型更关注难分类样本，提高了训练效率，但计算速度较慢。Li 等提出了 FSSD（Feature Fusion Single Shot Multibox Detector）模型[18]，将 SSD 和 FPN（Feature Pyramid Network）思想结合起来，构建了一组新的有效特征

层，加快了速度，但速度提高有限。

本文提出的算法相较于传统以监督学习为主的目标检测算法，引入了对比学习的思想。对比学习算法以自监督学习为主，根据图像本身通过对比图片之间的相似度来进行网络训练，这和目标检测过程中判断目标和背景的差异性思想是一致的。因此，在目标检测算法中引入对比学习的思想，提取和背景差异更加明显的目标特征，将会有效地提升模型的检测效果。

1.2 对比学习

对比学习是一种无监督学习方法，通过对比相似和不相似的示例来学习有用的特征表示。在机器学习领域，对比学习引起了广泛关注。目前，经典的对比学习算法包括 SimCLR（A Simple Framework for Contrastive Learning of Visual Representations）算法[19]、MoCo（Momentum Contrast）算法[20]、BYOL（Bootstrap Your Own Latent）算法[21]、SwAV（Swapped-Asymmetric learning for Visual recognition）算法[22]和 SimSiam（Exploring Simple Siamese Representation Learning）算法[23]。对于对比学习，训练数据是未标记的图像，网络模型可以从中学习图像的特征表示，以实现图像分类、检索等任务。

Chen 等人提出了 SimCLR 算法[19]，这是一种经典的同步对称网络架构的对比学习框架。它使网络能够学习图像特征，并根据不同的数据增强模式和图像的多维特征将这些特征传输到不同的下游任务。在 ImageNet 数据集中，SimCLR 的对比学习实验结果类似于 ResNet50，并实现了 top-1 的无监督分类准确率。然而，在网络训练过程中可能会出现模型退化的问题。为了解决这些问题，He 等人提出了 MoCo 算法[20]，这是一种异步对称网络架构的对比学习网络框架。它通过在内存中维护队列的方式，并使用动量更新方法来训练模型，在 ImageNet 数据集中，MoCo 的检测性能超过了 SimCLR。BYOL 算法[21]和 SimSiam 算法[23]是异步不对称网络结构的例子。这两种方法都采用异步不对称网络结构，其中 BYOL 首次仅使用正样本计算对比学习损失。虽然 BYOL 的骨干网络架构是 MoCo，但它在训练中不使用负样本。然而，如果上下网络分支相同，则可能会在训练过程中出现网络崩溃的问题。为了解决这个问题，SimSiam 提出了交叉梯度更新方法，防止网络崩溃。与 BYOL 类似，SimSiam 不需要对网络进行负样本训练。

1.3 SSD 改进算法

针对目前有监督目标检测算法中，检测物体视角的多样性问题、目标本身形变问题、检测物体受遮挡问题、光照性问题以及小目标检测精度不高等问题，本文改进 SSD 目标检测算法引入对比学习的思想，提取和背景差异更加明显的目标特征，以提高背景和目标在特征空间中的区分度，并使用监督学习的方法对网络进行训练，提高 SSD 网络目标检测精度，从而有效地提升其网络模型的检测效果。

1.4 图像预处理

本文的重点内容在于使用对比学习思想来提高图片背景和目标在特征空间中的区分度。在获取正负样本的流程中，采用图片随机截块的方式分别截取样本图片中的目标图像与背景图像，其中目标图像为正样本，背景图像为负样本。首先找到图像中的目标中心点坐标，然后以目标 ground truth 框为基准随机截取图像块，图像块大小为 64×64。当截取的图像块与 ground truth 框之间的 IOU（Intersection over Union）大于截取图像块一半以上的目标区域，即认为截取到图像为正样本图像，如图 2 中的绿色方框所示；当截取的图像包含一半以下的目标区域，即认为截图到的图像为背景图像，如图 2 中的黄色方框所示。

截取到正负样本后，将样本图片进行图像增强，图像增强即从同一张图像衍生出不同的图像，衍生出的图像和原图像内容相同，只是大小、尺度、亮度、颜色等会发生一定的变化，衍生图像和原图像认为其正负类别相同。本文采用的图像增强操作方法包括裁剪、旋转、颜色调整、尺度调整、以及光照亮度调整等操作，并对每种操作之间进行随机组合。值得注意的是，算法模型测试阶段的图像截块方式与训练阶段图像截块方式相同。

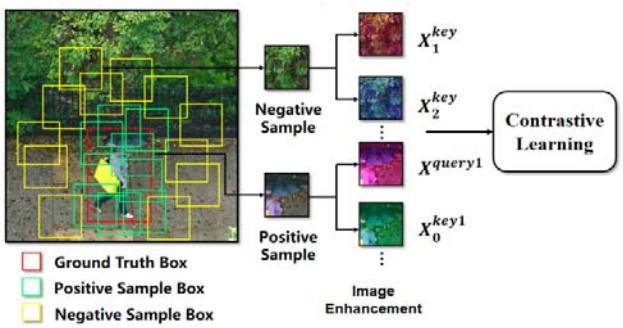

图 2 正负样本截取及图像增强示意图

Fig.2 Diagram illustrating positive and negative sample extraction and image Enhancement

1.5 整体网络结构

本文整体网络结构一共包含两种网络结构：SSD 网络结构与对比学习网络结构，SSD 与对比学习网络结构编码器部分都是由若干个卷积层、池化层以及全

连接层组成，整体网络结构如图 3 所示。其中，SSD 网络结构主要分为输入层、特征提取层、检测层、输出层、非极大值抑制层，特征提取层通常由一系列卷积层和池化层组成，用于从输入图像中提取特征，以便于后续的目标检测。

本文采用 MoCo 对比学习网络结构，作为改进 SSD 算法的对比学习方法。MoCo 网络结构由图像增强、特征提取和计算对比损失 3 个部分组成。图像增强部分采用随机图像增强，包括裁剪、旋转、颜色调整、尺度调整、以及光照亮度调整等操作，并执行多种增强操作的随机组合。特征提取部分采用 ResNet 残差网络，由多个残差块组成。

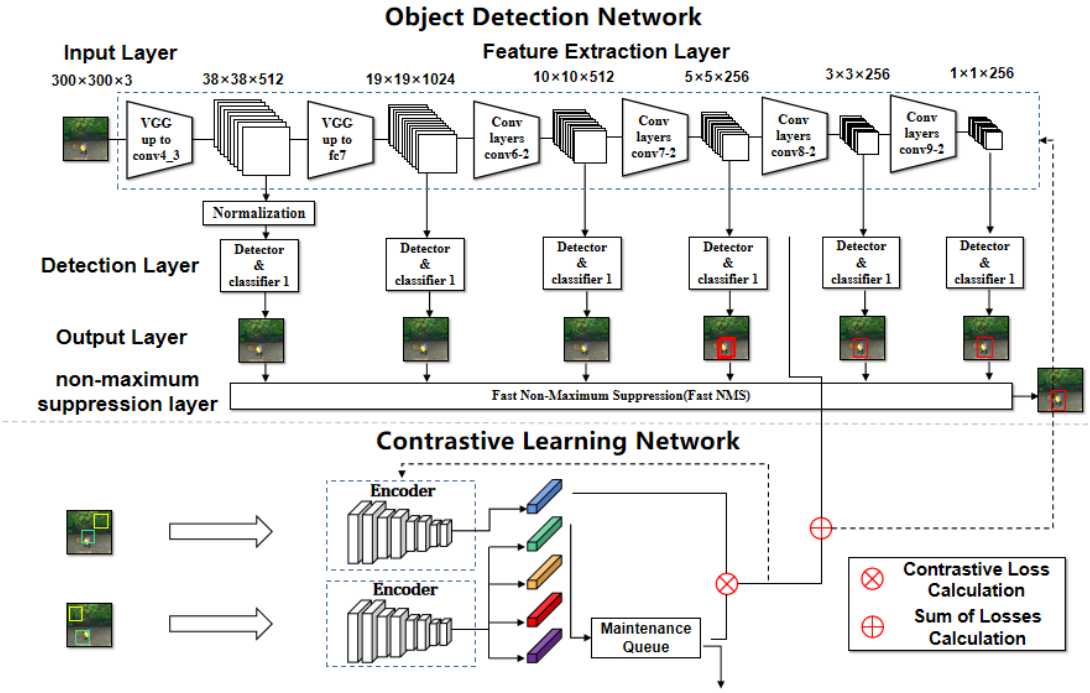

图 3 改进 SSD 算法的整体网络结构图

Fig.3 Improved overall network architecture diagram of SSD algorithm

截取到图片的正负样本经过图像增强之后正样本图像生成锚点图像 X^{query1} 与正键图像 X_0^{key}，负样本图像生成若干负键图像 X_1^{key}、X_2^{key}……，最终将锚点图像 X^{query1} 输入到查询编码器中提取图像的特征，将正键图像 X_0^{key} 与若干负键图像 X_1^{key}、X_2^{key}……输入到动量编码器中提取图像的特征。查询编码器提取的是正样本图像增强后的其中一个作为原始对比图像的特征。而动量编码器需要提取的是正样本图像增强后的另一个图像与负样本图像增强后的图像特征，从而计算正负样本的对比损失。MoCo 模型使用到了维护队列来存储动量编码器提取到的图像特征，然后输出图像特征与查询编码器提取到的特征做对比的损失计算结果。维护队列的大小通常是一个超参数，它可以影响模型的性能和训练时间，队列长度小，模型的性能差，队列长度大，往往模型的性能也不会有显著的提升，模型训练时间也会变长且会占用过多的内存空间。本文 MoCo 模型存储样本特征所采用的维护队列大小设置为 4096，即队列可存储的样本数最大为 4096，在这个长度下的维护队列模型性能较高且训练时间不长、内存空间占用也不大[20]。

在网络训练过程中，计算得到对比损失后，将该值反向传播到两个编码器中，以更新优化网络参数。同时，对比学习损失和 SSD 损失进行加权求和，得到总损失值，并将其反向传播到 SSD 网络中，以更新优化 SSD 的网络参数。

1.6 损失函数

SSD 目标检测算法的损失函数通常由分类损失 L_{conf} 和回归损失 L_{loc} 两部分组成，用于训练模型并提高目标检测的准确性。分类损失函数通常用于训练模型的目标分类任务，该损失函数计算模型预测每个预测框的类别概率与其真实标签之间的交叉熵损失。具体地，对于每个预测框，分类损失函数计算其类别预测概率与其真实类别标签的交叉熵损失，并对所有预测框的损失进行加权平均。回归损失函数通常用于训练模型的目标定位任务，该损失函数计算模型预测每个预测框的位置偏移量与其真实标签之间的损失。具

体地，对于每个预测框，回归损失函数计算其位置偏移量与其真实位置标签之间的 Smooth L1 损失，并对所有预测框的损失进行加权平均。最终，SSD 模型的总损失函数为分类损失 L_{conf} 和回归损失 L_{loc} 的加权和：

$$\text{Loss}_{\text{SSD}}=\frac{1}{N}\left(L_{\text{conf}}+\alpha L_{\text{loc}}\right) \tag{1}$$

式中：α表示分类损失和回归损失之间的权重系数。在训练过程中，可以调整α的值以平衡分类损失和回归损失的权重，提高模型的准确性和检测性能。

对比损失公式定义为：

$$L_i=-\log\frac{\exp\left(qk_+/\tau\right)}{\sum_{i=1}^{k}\exp\left(qk_i/\tau\right)} \tag{2}$$

式中：q 表示查询编码器从目标图像中提取的特征；k_i 表示动量编码器提取的特征；k_+表示正样本的特征（假设有且只有一个），使用τ作为超参数来调整上述对比损失。

在完成了正负样本的对比损失计算后，接下来需要计算交叉熵损失函数。为了计算交叉熵损失函数，需要将正负样本的对比损失作为损失样本。交叉熵损失函数的计算公式如下：

$$\text{Loss}_{\text{CL}}=\sum_{i=1}^{n}L_i\log\hat{L}_i \tag{3}$$

式中：n 表示正样本和负样本之间的样本数；L_i 表示第 i 次预期对比损失；$\hat{L}_i$ 表示网络模型计算的第 i 次对比损失。$\hat{L}_1$ 为正样本与样本图像之间的对比损失，而 $\hat{L}_2$、$\hat{L}_3$…是负样本与样本图像之间的对比损失。然后，通过一步一步的迭代运算，$\hat{L}_i$ 逐渐逼近 L_i。在每个 epoch 中，计算损失函数，使样本能够满足拉进正样本和拉出负样本的目标。从原始图像中裁剪正负区域图像，然后增强图像。得到的增强对象和背景图像被输入编码器以提取它们的特征。得到的对比损耗用于更新对比学习编码器的网络参数，也添加到 SSD 损耗中进行整体训练。最终整体损失定义为：

$$\text{Loss}=\beta\text{Loss}_{\text{SSD}}+\gamma\text{Loss}_{\text{CL}} \tag{4}$$

上式为 SSD 整体优化目标函数，它由两部分组成：Loss_{SSD} 表示 SSD 目标检测训练损失；Loss_{CL} 表示对比学习损失；β与γ为超参数用于调节两者的权重。

2 实验与结果分析

实验采用 Pytorch 深度学习框架，CUDA 为 11.3 版本，Pytorch 为 1.7.1 版本。网络权值的初始学习效率为 0.01，衰减系数设置为 0.0005。batch-size＝16，共 50 批次。SSD 损失函数中α的值设为 1，整体损失函数中 SSD 损失函数权重β设为 1，对比学习损失函数权重γ设为 0.01。优化器使用 optimizer 优化器。

2.1 数据集介绍

实验使用 PASCAL VOC2012 数据集与 LLVIP 数据集作为实验数据集进行改进 SSD 网络结构的检测精度实验。

PASCAL VOC2012 是一个常用的计算机视觉数据集，用于目标检测、图像分割、物体分类等任务。该数据集包含 20 个物体类别，涵盖了常见的物体类别，如人、汽车、飞机、动物等。该数据集共包含 17125 张图像，其中 12281 张图像是训练集，3388 张图像是验证集，1456 张图像是测试集。PASCAL VOC2012 数据集是计算机视觉领域中使用最广泛的数据集之一，常用于评估目标检测、图像分割等算法的性能和表现。

LLVIP 数据集包含了 15488 对红外图像，涵盖了 26 个实时场景，其中大部分使用 8～14 μm 波段在低光条件下拍摄。在实验中，我们选择 19 个实时场景作为训练与验证集，其中 6009 张图像作为训练集，6012 张图像作为验证集，共计 12021 张图片，3467 张图片作为测试集。

2.2 对比实验结果

本文采用 AP、AP_{50}、AP_{75}、AP_S、AP_M、AP_L 指标对改进 SSD 网络模型进行评估，AP 的计算方法是在不同的置信度（confidence）阈值下计算模型的准确率（Precision）和召回率（Recall），并对准确率-召回率曲线下面积进行积分；AP_{50} 为当 IOU 阈值为 0.5 时的平均精度；AP_{75} 为当 IOU 阈值为 0.75 时的平均精度；AP_S 为对小物体（面积小于 32×32 像素）检测的平均精度；AP_M 为对中等大小物体（32×32 像素到 96×96 像素）检测的平均精度；AP_L 为对大物体（面积大于 96×96 像素）检测的平均精度。

将 PASCAL VOC2012 数据集分别输入进 SSD 网络与改进 SSD 网络中训练达 50 迭代次数时，损失函数便开始收敛。通过对改进 SSD 算法与原 SSD 算法性能指标结果比较可以看出改进 SSD 算法比原 SSD 算法有所提升，其结果如表 1 所示。从表 1 可以看出，改进 SSD 算法与原 SSD 算法相比较，AP 值提升了 0.1%，AP_{50} 值提升 0.3%，AP_S 值提升了 0.3%，AP_M 值提升了 1.1%。从表 1 可以看出，本文改进算法在 PASCAL VOC2012 数据集中表现最好。改进 SSD 与原 SSD 算法的检测结果如图 4 所示。

基于表 1 中 AP_{75} 与 AP_L 值并未高出原算法的问题，我们讨论了改进 SSD 算法精度是否与截取图像块

尺寸大小有关，于是本文分别采用了截取尺寸为 32×32 的图像块和截取尺寸为96×96的图像块进行实验，结果如表 2 所示。可以看出，当截取图像块尺寸变大时其 AP_L 值也会相应变大，说明 AP_S、AP_M、AP_L 指标是会受截取图像块大小的影响。但图像尺寸过小所截取图像中的特征信息偏少，会影响到对比学习的精度，而图像块尺寸过大，正样本中所含概的背景元素太多也会影响到对比学习的精度，因此截取图像块的尺寸大小本文使用 64×64。

将 LLVIP 数据集分别输入 SSD 网络与改进 SSD 网络中训练达 50 迭代次数时，损失函数便开始收敛。改进 SSD 算法与原 SSD 算法性能指标结果比较可以看出改进 SSD 算法比原 SSD 算法有所提升，其结果如表 3 所示。从表 3 可以看出，改进 SSD 算法与原 SSD 算法相比较，AP 值提升了 0.2%，AP_{50} 值提升了 0.2%，AP_{75} 值提升了 0.3%，AP_S 值提升了 0.2%，AP_L 值提升了 0.2%。从表 3 可以看出，本文改进算法在 LLVIP 数据集中表现最好。改进 SSD 与原 SSD 算法的检测结果如图 5 所示。

表 1 PASCAL VOC2012 数据集上改进 SSD 算法与原 SSD 算法结果比较

Table 1 Comparison of the results between the improved SSD algorithm and the original SSD algorithm on the PASCAL VOC2012 dataset

Models	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Improved SSD algorithm	0.452	0.718	0.487	0.091	0.281	0.522
Original SSD algorithm	0.451	0.715	0.491	0.088	0.270	0.524

图 4 改进 SSD 与原 SSD 算法在 PASCAL VOC2012 数据集上的检测结果（上排为改进 SSD 算法的检测效果图，下排为原 SSD 算法检测效果图）

Fig.4 The detection results of the improved SSD and the original SSD algorithms on the PASCAL VOC2012 dataset (The top row shows the detection results of the improved SSD algorithm, while the bottom row shows the detection results of the original SSD algorithm)

表 2 在 PASCAL VOC2012 数据集中不同图像块截取尺寸下的算法结果比较

Table 2 Comparison of algorithm results under different sizes of image cropping on the PASCAL VOC2012 dataset

Image cropping size (Pixels)	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Improved SSD Algorithm（32×32）	0.450	0.717	0.485	0.091	0.281	0.521
Improved SSD Algorithm（64×64）	0.452	0.718	0.487	0.091	0.281	0.522
Improved SSD Algorithm（96×96）	0.449	0.713	0.485	0.085	0.279	0.524

表 3 LLVIP 数据集上改进 SSD 算法与原 SSD 算法结果比较

Table 3 Comparison of results between the improved SSD algorithm and the original SSD algorithm on the LLVIP dataset

Models	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Improved SSD algorithm	0.524	0.928	0.539	0.013	0.272	0.539
Original SSD algorithm	0.522	0.926	0.536	0.011	0.275	0.537

图 5 改进 SSD 与原 SSD 算法在 LLVIP 数据集上的检测结果（上排为改进 SSD 算法的检测效果图，下排为原 SSD 算法检测效果图）

Fig.5 The detection results of the improved SSD and the original SSD algorithms on the LLVIP dataset (The top row shows the detection results of the improved SSD algorithm, while the bottom row shows the detection results of the original SSD algorithm)

2.3 其他目标检测算法精度对比

为了进一步证明改进 SSD 算法的有效性，本文对几种当前主流的目标检测算法进行了实验比较，包括 YOLO-v2、SSD、Fast R-CNN、Faster R-CNN、ION[24] 等。在所有实验中，我们统一了配置环境和初始训练参数，数据集使用 MS COCO2017 数据集，数据集包括 118287 个训练集，5000 个验证集和 40670 个测试集。实验数据结果见表 4。

在表 4 中，我们评估了改进算法的准确度，可以看出和其他主流目标检测算法相比，改进后的 SSD 算法的各种 AP 值都有所提高。验证了改进后的 SSD 算法在目标检测任务中的有效性。

3 结语

本文所提出的改进算法对于其他目标检测算法的优势在于，引用了对比学习的思想进行目标和背景特征的提取，通过同时约束目标检测损失和对比损失，可使得同类目标在特征空间中分布更紧凑，目标和背景在特征空间中分布距离更远，从而提升目标和背景之间的区分度，提升 SSD 目标检测算法的检测精度。本文改进算法为通用型的算法，该对比学习机制不仅可以适用于 SSD 目标检测模型，还可适用于其他基于深度学习的目标检测方法，如 RCNN 系列、YOLO 系列、SPP-Net、R-FCN 等。

表 4 MS COCO2017 数据集上改进 SSD 算法与其他目标检测算法结果比较

Table 4 Comparison of results between the improved SSD algorithm and other object detection algorithms on the MS COCO 2017 dataset

Models	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
YOLOv2 [6]	21.6	44.0	19.2	5.0	22.4	35.5
YOLOv3	33.0	57.9	34.4	18.3	35.4	41.9
YOLOv5	36.9	58.4	-	-	-	-
SSD [4]	23.2	41.2	23.4	5.3	23.2	39.6
Fast R-CNN [10]	20.5	39.9	19.4	4.1	20.0	35.8
Faster R-CNN [11]	21.9	42.7	-	-	-	-
ION [24]	23.6	43.2	23.6	6.4	24.1	38.3
Improved SSD Algorithm	28.9	47.5	30.7	5.5	26.5	43.5

参考文献:

[1] XIA G S, BAI X, DING J, et al. DOTA: a large scale dataset for object detection in aerial images[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 3974-3983.

[2] ZHANG J, LIANG X, WANG M, et al. Coarse-to-fine object detection in unmanned aerial vehicle imagery using lightweight convolutional neural network and deep motion saliency[J]. *Neurocomputing*, 2020, **398**: 555-565.

[3] Pathak A, Pandey M, Rautaray S. Application of deep learning for object detection[J]. *Procedia Computer Science*, 2018, **132**: 1706-1717.

[4] LIU W, Anguelov D, Erhan D, et al. SSD: Single shot MultiBox detector[C]//*Proceedings of the* 14th 284 *European Conference on Computer Vision*, 2016: 21-37.

[5] LIU G, NOUAZE J C, TOUKO P L, et al. YOLO-tomato: A robust algorithm for tomato detection based on YOLOv3[J]. *Sensors*, 2020, **20**(7): 2145.1-2145.20.

[6] Redmon J, Farhadi A. Yolo9000: Better, faster, stronger[C]// *Computer Vision and Pattern Recognition* (CVPR), 2017: 6517-6525.

[7] Sruthi M S, Poovathingal M J, Nandana V N, et al. YOLOv5 based open-source UAV for human detection during search and rescue (SAR) [C]//

10th *International Conference on* 13 *Advances in Computing and Communications*, 2021: 1-6.

[8] ZHU X K, LYU S C, WANG X, et al. TPH-YOLOv5: improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C]//*Proceedings of the IEEE International Conference on Computer Vision*, 2021: 2778-2788.

[9] CHEN C, LIU M Y, Tuzel O, et al. R-CNN for small object detection[C]//*Asian Conference on Computer Vision*, 2016: 214-230.

[10] Girshick R. Fast R-CNN[C]//*IEEE International Conference on Computer Vision*, 2015: 1440-1448.

[11] REN S Q, HE K M, Girshick R, et al. Faster CNN: Towards real-time object detection with region proposal networks[C]//*Proceedings of the 28th International Conference on Neural Information Processing Systems*, 2015: 91-99.

[12] WANG Longguang, WANG Yingqian, DONG Xiaoyu, et al. Unsupervised degradation representation learning for blind super-resolution[C]//*CVPR*, 2021: 10581-10590.

[13] HUANG Y, TU X, FU G, et al. Low-Light image enhancement by learning contrastive representations in spatial and frequency domains[J]. arXiv preprint arXiv: 2303.13412, 2023.

[14] SUN X H, GU J N, HUANG R. A modified SSD method for electronic computer fast recognition[J]. *Optik*, 2020, **205**: 163767.

[15] FU C Y, LIU W, Ranga A, et al. Dssd: DeConvolutional single shot detector[J]. arXiv preprint arXiv: 1701. 06659, 2017.

[16] Jeong J, Park H, Kwak N. Enhancement of SSD by con-catenating feature maps for object detection[J]. arXiv preprint arXiv: 1705. 09587, 2017.

[17] 李文涛, 彭力. 多尺度通道注意力融合网络的小目标检测算法[J]. **计算机科学与探索**, 2021, **15**(12): 2390-2400.
LI Wentao, PENG Li. Small objects detection algorithm with multi-scale channel attention fusion network[J]. *Journal of Frontiers of Computer Science & Technology*, 2021, **15**(12): 2390-2400.

[18] LI Z, ZHOU F. FSSD: feature fusion single shot multibox detector[J]. arXiv preprint arXiv: 1712. 00960, 2017.

[19] CHEN T, Kornblith S, Norouzi M, et al. A simple framework for contrastive learning of visual representations[C]//*Proceedings of the* 37*th International Conference on Machine Learning*, 2020: 1597-1607.

[20] HE K M, FAN H Q, WU Y X, et al. Momentum contrast for unsupervised visual representation learning[C]//*Proceedings of* 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 9726-9735.

[21] Grill J B, Strub F, Altche F, et al. Bootstrap your own latent a new approach to self-supervised learning[C]//*Proceedings of the* 34th *International Conference on Neural Information Processing Systems* (NIPS), 2020: 2127121284.

[22] Caron M, Misra I, Mairal J, et al. Unsupervised learning of visual features by contrasting cluster assignments[C]//*Proceedings of the* 34th *International Conference on Neural Information Processing Systems*, 2020: 99129924.

[23] CHEN X L, HE K M. Exploring simple Siamese representation learning[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (CVPR), 2021: 1574515753.

[24] Bell S, Zitnick CL, Bala K, Girshick R. Inside-outside net: Detecting objects in context with skip pooling and recurrent neural networks[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 2874-2883.