

# 优化特征融合的多尺度遥感图像目标检测方法

张昊<sup>1,2</sup> 刘凤<sup>1,2</sup> 谭富祥<sup>1,2</sup> 钱育蓉<sup>1,2,3\*</sup>

<sup>1</sup>(新疆大学软件学院 新疆 乌鲁木齐 830008)

<sup>2</sup>(新疆维吾尔自治区信号检测与处理重点实验室 新疆 乌鲁木齐 830008)

<sup>3</sup>(新疆大学软件工程重点实验室 新疆 乌鲁木齐 830046)

**摘要** 为解决遥感图像场景下多尺度目标检测准确率低的问题,提出 DAFFNet 遥感图像目标检测算法。该算法基于 SSD 进行了三方面的改进:为增强多尺度特征信息的获取能力,设计一种基于分组的特征融合方法;引入基于注意力机制的多维度特征优化方法,来解决复杂背景下目标分类困难的问题;将 Focal loss 作为新的边界框置信度损失函数,令模型聚焦于难分类的正样本,以改善正负样本不平衡对目标分类所造成的干扰。在遥感公共数据集 NWPU VHR-10 上进行模型评估,实验结果表明,该算法相较于原算法均值平均精度提高 5.1 百分点,能有效地提高遥感图像目标检测准确率。

**关键词** 遥感图像 目标检测 分组特征融合 多维度特征优化 注意力机制

中图分类号 TP3

文献标志码 A

DOI:10.3969/j.issn.1000-386x.2024.08.022

## MULTI-SCALE REMOTE SENSING IMAGE TARGET DETECTION METHOD BASED ON OPTIMIZED FEATURE FUSION

Zhang Hao<sup>1,2</sup> Liu Feng<sup>1,2</sup> Tan Fuxiang<sup>1,2</sup> Qian Yurong<sup>1,2,3\*</sup>

<sup>1</sup>(College of Software, Xinjiang University, Urumqi 830008, Xinjiang, China)

<sup>2</sup>(Key Laboratory of Signal Detection and Processing in Xinjiang Uygur Autonomous Region, Urumqi 830008, Xinjiang, China)

<sup>3</sup>(Key Laboratory of Software Engineering, Xinjiang University, Urumqi 830046, Xinjiang, China)

**Abstract** A DAFFNet remote sensing image object detection algorithm is proposed to solve the problem of low accuracy of multi-scale object detection in the remote sensing image scene. Based on SSD, the algorithm was improved in three aspects. We designed a group-based feature fusion method to enhance the ability to acquire multi-scale feature information. A multi-dimensional feature optimized method based on the attention mechanism was introduced to solve the difficulty of target classification in a complex background. The focal loss was used as a new bounding box confidential loss function to make the model focus on the positive samples that were difficult to classify, so as to improve the interference caused by the imbalance of positive and negative samples to target classification. The model was evaluated on the remote sensing public dataset NWPU VHR-10. The experimental result shows that the proposed algorithm improves the mean average precision by 5.1 percentage points compared with the original algorithm, which can effectively increase the object detection accuracy of remote sensing image.

**Keywords** Remote sensing images Object detection Group-based feature fusion Multi-dimensional feature optimization Attention mechanism

## 0 引言

近年来,目标检测与识别技术一直是学术界与工业界的一大研究热点。随着科学技术的迅猛发展,遥感技术也愈发成熟,遥感图像目标检测更是一跃成为计算机视觉领域的重要研究方向<sup>[1]</sup>。遥感图像相较于常规图像具有更为广阔的拍摄视野和更加丰富的正负样本量,遥感目标检测技术在军事和民用领域均具有广泛的应用<sup>[2]</sup>,占据着不可或缺的一席之地。

近年来基于深度学习的目标检测算法在图像处理领域中异彩纷呈,其可分为两阶段算法和单阶段算法两种类型。Faster R-CNN<sup>[3]</sup>、Cascade R-CNN<sup>[4]</sup>、Trident-Net<sup>[5]</sup>均是两阶段算法的典型代表,它们的检测准确率较高但检测速度较慢。YOLO(You Only Look Once)<sup>[6]</sup>、SSD(Single Shot MultiBox Detector)<sup>[7]</sup>、RetinaNet<sup>[8]</sup>、RefineDet<sup>[9]</sup>均是单阶段算法的典型代表,它们提供端到端的方式直接进行预测,检测速度快,精度略低于两阶段算法。

目前基于深度学习的目标检测算法已经被广泛应用到遥感图像目标检测与识别领域中。Audebert 等<sup>[10]</sup>将语义分割和目标检测进行结合应用于遥感图像领域。Yang 等<sup>[11]</sup>提出双路径特征注意力网络,利用目标真实标注框中的掩码信息来引导网络在训练过程中对目标物体的关注度,抑制无关背景信息带来的干扰。Fang 等<sup>[12]</sup>提出了一种半锚式检测器,利用旋转无锚分支精确回归 OBB(Oriented Bounding Box)以增强对前景特征的学习,从而优化对目标的定位能力。Tayara 等<sup>[13]</sup>提出了一种密集连接的特征金字塔网络,通过获得融合多尺度的高质量信息特征图来解决高分辨率遥感图像目标尺度差异大的问题。Wang 等<sup>[14]</sup>提出 Inception 横向连接来提高特征金字塔网络对多尺度特征提取的能力,但增加了计算成本。刘凤等<sup>[15]</sup>将更深的密集卷积网络作为新的骨干网络,充分优化了模型的特征提取能力。Liu 等<sup>[16]</sup>将轻量级通道注意力与残差网络相结合,使网络更加关注特征层中更有效的通道,实现不降维的局部跨通道交互,增强特征提取能力。总体来看,基于深度学习的方法取得了显著成果,但是仍面临以下挑战:

(1) 遥感图像具有较宽的拍摄广角,其中包含复杂背景以及多类目标。图像中正负样本不平衡,无关背景干扰较大,均导致目标分类难度增加。

(2) 遥感图像具有较高的拍摄高度,因此各类目标尺度差异较大,会导致特征提取能力受到影响。

针对上述问题,研究者们基于 SSD 算法在特征融合方面做出了相应的改进: Fu 等<sup>[17]</sup>针对小目标物体检测鲁棒性差的问题,通过反卷积操作和跳跃连接进行特征融合构造出 DSSD 算法。Li 等<sup>[18]</sup>将 SSD 与 FPN 相结合,并采用 Concat 融合操作构造出 FSSD 算法。

由于 DSSD 中的跳跃连接在传递信息时所产生的大量冗余信息会限制检测精度的提升, FSSD 中的 Concat 操作仅顾及通道维度上的特征融合。因此本文提出了基于分组的特征融合方法对多尺度特征进行融合,同时利用注意力在空间维度和通道维度上对融合后的特征进行多维度特征优化,最终构造出了一种适用于遥感图像的目标检测方法(Dual Attention and Feature Fusion Network, DAFFNet)。将其在 NWPU VHR-10<sup>[19]</sup>数据集上进行对比实验,实验表明本文所提 DAFFNet 算法具有更高的目标检测精度。

## 1 相关模型介绍

### 1.1 SSD 算法

SSD 目标检测算法是由刘伟等于 2016 年提出的一种基于端到端训练学习的单阶段目标检测算法。它通过结合 YOLO 的回归思想以及 Faster R-CNN 中的 Anchor 机制直接预测目标类别和位置,提高检测精度。SSD 的网络结构如图 1 所示。

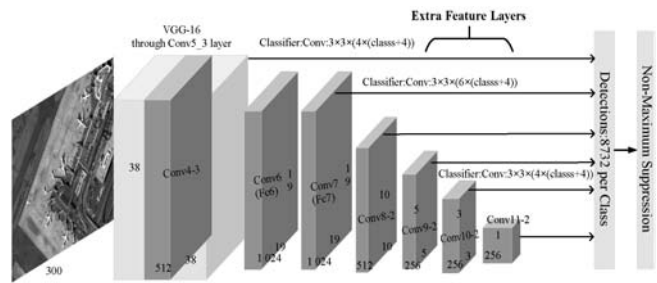


图1 SSD的结构

该算法借鉴了特征图像金字塔的思想,利用 Conv4\_3、Conv7、Conv8\_2、Conv9\_2、Conv10\_2 以及 Conv11\_2 生成 6 种不同大小的特征图直接进行目标检测和分类,经过非极大值抑制(Non-Maximum Suppression, NMS)方法来去除掉多余候选框,从而获得正确的检测结果。

SSD 算法选择从偏高的 Conv4\_3 层开始进行金字塔的构建,这在很大程度上无法充分利用低层特征图的空间信息,缺乏小目标的特征信息,不利于遥感图像小目标的检测与识别。虽然利用了多个尺度大小的特征图进行预测,但忽略了特征间的上下文联系,未能将特征信息进行共享,无法精确地对遥感图像中多尺度

目标物体进行检测。

### 1.2 FPN 算法

特征金字塔 FPN<sup>[20]</sup>算法 (Feature Pyramid Network) 由 Lin 等根据深度卷积神经网络固有的多尺度、多层级的金字塔结构来进行构建的。该金字塔进行自上而下的横向连接,其结构由自底向上、自顶向下以及横向连接路径组合而成。FPN 算法结构如图 2 所示。

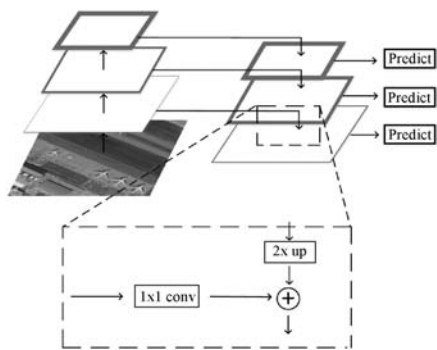


图 2 FPN 结构

该算法将语义信息更丰富的高层特征图做 2 倍上采样,然后各级较低层特征通过  $1 \times 1$  卷积调整通道数为 256,并将所得的上采样结果与经过  $1 \times 1$  卷积的结果对应元素依次相加,在所有尺度上将具有高分辨率的低层特征图和具有丰富语义特征的高层特征图进行融合,从而丰富多尺度特征,有效提升检测精度。本算法中每层都是独立预测的,以确保能够充分提取多尺度的信息特征,提升检测精度,降低漏检率。

### 1.3 CBAM 模型

在人类的感知功能中,注意力是非常重要的存在,其发挥的作用是显而易见的。人类视觉系统不会尝试一次性地处理整个场景,而是有选择性地关注突出部分以更好地捕捉视觉结构。注意力的作用不仅表现在寻找重点关注的位置,更表现在可以优秀地表示关注点。因此研究者们开始尝试在卷积神经网络 (Convolutional Neural Networks, CNN) 中加入注意力,用于提高其大规模分类任务的性能。

双重注意力机制 (Convolutional Block Attention Module, CBAM)<sup>[21]</sup>具有良好的通用性以及轻量级模型的优势,因此能够轻松地将它与卷积神经网络进行结合,进行端到端的训练,最终达到提升网络模型特征提取能力的效果。该 CBAM 模型将输入特征依次通过通道注意力模块和空间注意力模块输出,以获得注意力权重,然后对特征进行自适应调整。通过将整个注意力过程分为两个独立的部分,有效地获得通道维度和空间维度之间特征的相关性和重要性。

## 2 DAFFNet 遥感图像目标检测方法

### 2.1 网络模型设计

SSD 算法在 MSCOCO、PASCAL VOC 等数据集上具有优异的检测性能。但由于遥感图像与常规图像的场景不同,具有目标尺度差异大、背景复杂度高、正负样本不平衡的特点,这使得 SSD 算法在遥感图像领域中面临较大的挑战。

因此本文将在 SSD 算法的基础上,根据遥感图像的特点来构建适合的目标检测算法,提出了基于分组的特征融合方法和基于注意力机制的多维度特征优化方法,并将 Focal loss<sup>[22]</sup>损失函数作为新的类别损失函数。DAFFNet 算法的网络结构如图 3 所示。算法首先将遥感图像输入骨干网络进行特征提取,获得多尺度特征图。再通过基于分组的特征融合对多尺度特征图进行处理,将融合后的结果通过基于注意力的多维度特征优化方法,最终生成 6 种不同大小的特征图用于目标检测和分类。

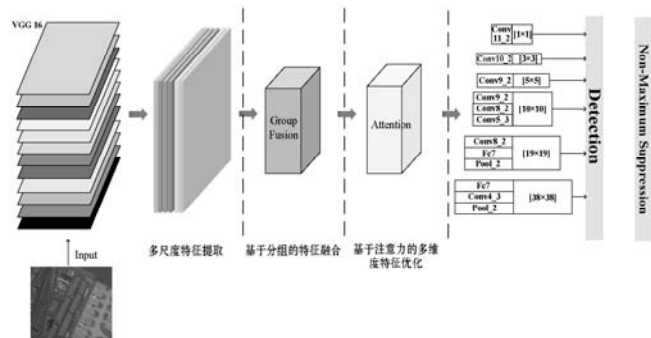


图 3 DAFFNet 的结构

### 2.2 基于分组的特征融合

由于 SSD 算法不仅未能充分利用低层特征图的空间信息,而且未对多尺度特征信息进行共享,因此本文提出了一种基于分组的特征融合方法 (Group Fusion) 来进行改进,其网络结构如图 4 所示。

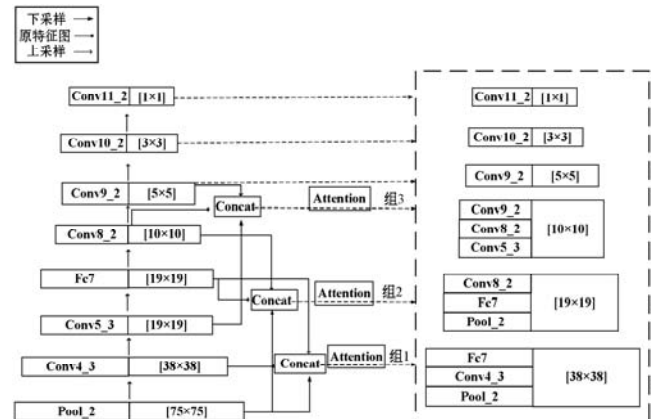


图 4 基于分组的特征融合

该方法将从 VGG-16 网络中的 Pool2 层特征图开始构建特征金字塔,该层相较于 Conv4\_3 层分辨率更高且能够充分利用低层特征图的空间信息,具有更为详细的小目标特征信息与位置信息,有利于小目标的检测与识别。因此其自底向上地依次抽取了 Pool2、Conv4\_3、Conv5\_3、Fc7、Conv8\_2、Conv9\_2、Conv10\_2、Conv11\_2 共 8 个特征层进行特征提取,并将所获特征图划分为三个待融合的特征层组,分别进行多尺度特征融合,分组细节如表 1 所示。

表 1 待融合特征层组的分组细节

组序	待融合特征层组	融合特征图的大小
组 1	Pool_2、Conv4_3、Fc7	38 × 38
组 2	Pool_2、Fc7、Conv8_2	19 × 19
组 3	Conv5_3、Conv8_2、Conv9_2	10 × 10

由于遥感图像内小尺度目标物体的外观和结构质量较差、所包含的特征信息较少,因此本文方法将在前两个待融合特征层组内利用具有详细小目标信息的 Pool2 层与不同深度的浅层特征进行特征融合,从而兼顾了中小尺度目标的检测。第三个待融合特征层组则对尺寸为 19 × 19、10 × 10、5 × 5 的特征层进行特征融合,以兼顾大、中尺度目标的检测。在第一个特征层组中,使用空洞卷积操作和反卷积操作分别对 Pool2 层的 75 × 75 大小的特征图、Fc7 层的 19 × 19 大小的特征图进行下采样和上采样操作,所得结果与 Conv4\_3 层输出的特征图尺寸一致,大小为 38 × 38,然后对该特征层组内的三个特征层进行 Concat 特征融合操作。同理,依次将第二个和第三个特征层组内的特征图尺寸统一为 19 × 19 和 10 × 10 的大小,并进行 Concat 特征融合操作。最终将低分辨率、高语义信息的高层特征和高分辨率、低语义信息的低层特征进行充分融合,从而丰富多尺度特征的语义信息,以强化多尺度特征提取的能力。

### 2.3 基于注意力机制的多维度特征优化

由于遥感图像背景复杂会导致目标分类难度增加而且进行特征融合所采用的 Concat 方法仅能够在通道维度上进行特征图之间的融合,无法很好地反映出不同维度间特征的相关性及重要性。因此受 CBAM 模型的影响,将通道注意力模块和空间注意力模块加入特征融合过程当中,提出了基于注意力机制的多维度特征优化方法。图 5 为多维度特征优化过程。

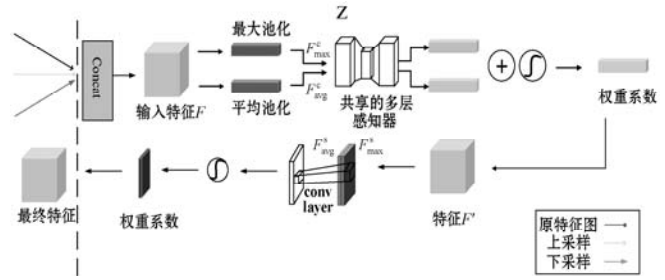


图 5 多维度特征优化过程

首先对经过尺度统一的多尺度特征进行 Concat 特征融合操作,再将融合后的特征  $F$  同时通过最大池化和平均池化以生成通道最大池特征  $F_{\max}^c$  和通道平均池特征  $F_{\text{avg}}^c$ ,两种池化同时使用能够更大程度提高网络的代表能力。以上两特征分别经过共享的多层感知器处理而获得两个新的特征,对两个新特征依次进行加和操作以及 Sigmoid 激活操作从而获得权重系数,利用该权重系数对原特征进行自适应调整而获得新的特征  $F'$ 。使得原特征能够在通道维度上更好地聚焦于具有重要意义部分,并抑制其中的冗余部分,有效地削弱了无关背景的干扰。

接下来将特征  $F'$  依次经过最大池化和平均池化以生成空间最大池特征  $F_{\max}^s$  和空间平均池特征  $F_{\text{avg}}^s$ ,并对二者进行拼接,然后经过卷积操作和 Sigmoid 激活操作从而获得权重系数,令该权重系数与输入特征  $F'$  相乘以生成最终的特征。这样,可以有效地在空间维度上对各尺度特征所占的权重重新进行分配,聚焦于特征中的重要位置,将特征融合后的结果更大程度地聚焦于目标物体当中。

多维度特征优化方法能够有效地增强特征融合结果中感兴趣区域的影响,以提高网络对目标的显著性,同时弥补 Concat 操作的缺点。

### 2.4 Loss 函数

为了进一步提高该算法在遥感图像领域中的目标检测精度,本文算法将采用 Focal Loss 函数作为新的分类损失函数来解决遥感图像中正负样本不平衡所导致目标分类难度增加的问题。

由于数据集中大量易分类的负样本会在损失函数中起主导作用,Focal Loss 中的调制系数  $(1 - p_i)^\gamma$  能够加强难分类样本在训练中所占的权重,削减易分类样本在训练中所占的权重,使得网络侧重于对难分类样本的学习,实现对难分类样本的有效训练。Focal Loss 函数公式如式(1)所示。

$$FL(p_i) = -\alpha(1 - p_i)^\gamma \log(p_i) \quad (1)$$

式中: $\gamma$  为可调整聚焦参数; $1 - p_i$  为调制因子; $\alpha$  为加权因子。

### 3 实验验证与结果分析

#### 3.1 数据集与评价指标

为了验证本文算法的有效性,完成遥感图像目标检测及分类工作,在 TitanV GPU 服务器上搭建基于 PyTorch 的目标检测框架。实验操作系统为 Windows 10,编程语言为 Python 3.7,主要函数库包括 OpenCV、Visdom 等。

本实验采用 NWPU VHR-10 数据集,该数据集分别从 Google Earth 和 Vaihingen 数据集裁剪而来,并由西北工业大学专家团队手动标注。数据集中共包含 800 幅高分辨率卫星图像,其中 150 幅图像无目标,另外 650 幅图像包含各种多尺度大小的目标,数据集的详细信息如表 2 所示。我们将数据集转换为 PASCAL VOC 格式,并按照 8:2 划分为训练集和测试集两部分。在实验中,本文将基础学习率、动量、权重衰减系数和网络单次输入数据的数量分别设置为 0.001、0.9、0.0005 和 4,利用预训练模型 VGG16 进行权重参数的初始化,迭代次数共 5 万次。

表 2 NWPU VHR-10 数据集的详细信息

目标类别	目标数量	目标尺寸 (Pixel)
飞机	757	50 × 77 - 104 × 117
船舶	302	20 × 40 - 30 × 52
油桶	655	27 × 22 - 61 × 51
棒球场	390	66 × 70 - 109 × 129
网球场	524	45 × 54 - 122 × 127
篮球场	159	52 × 52 - 179 × 179
操场	163	195 × 152 - 344 × 307
港口	224	95 × 32 - 175 × 50
桥梁	124	88 × 90 - 370 × 401
汽车	477	20 × 41 - 45 × 80

本文采用平均精度 (Average Precision, AP) 和均值平均精度 (mean Average Precision, mAP) 两个目标检测领域的标准评价指标来评估图像目标检测网络的性能。AP 代表某一类别目标的检测平均精度, mAP 代表所有目标类别 AP 的平均值,数值大小在 0 到 1 之间,其值越大说明检测精度越高,目标检测效果越好。mAP 的计算公式如式 (2) 所示,式中  $n$  表示目标类别数。

$$mAP = \frac{\sum_{i=1}^n AP_i}{n} \quad i = 1, 2, \dots, n \quad (2)$$

#### 3.2 遥感图像目标检测实验结果与分析

##### 3.2.1 消融实验对比分析

为了更好地理解 DAFFNet 算法中各个改进模块对检测效果的影响,在 NWPU VHR-10 数据集上进行了消融实验,分析结果如表 3 所示。

表 3 在 NWPU VHR-10 数据集上的消融学习

SSD Baseline	Group Fusion	Attention	Focal loss	mAP/%
✓				85.5
✓	✓			86.9
✓	✓	✓		88.4
✓	✓	✓	✓	90.6

由表 3 可知,将 Group Fusion 模块嵌入基线 SSD 算法中,能够有效结合多尺度特征的上下文信息,提高网络对不同尺度目标的检测与识别能力,与 SSD 算法相比, mAP 提高了 1.4 百分点。

将 Attention 与 Group Fusion 模块同时嵌入基线 SSD 中,可以使得网络自主学习多维度特征间的相关性和重要性,提高网络对目标的显著性。通过这两个模块的结合学习,相较于原始基线, mAP 从 85.5% 上升到 88.4%。

Focal loss 函数的引入使得网络能够有效训练难分类样本,降低复杂背景所带来的干扰。将 Attention 与 Group Fusion 模块以及 Focal loss 函数同时嵌入基线 SSD 中,结果表明通过这三个模块的结合学习,相较于原始基线, mAP 从 85.5% 上升到 90.6%。

##### 3.2.2 与不同算法的对比

将本文算法与 SSD 系列算法进行对比实验,实验结果如表 4 所示。在 NWPU VHR-10 数据集中, DAFFNet 算法相较于 SSD、DSSD、FSSD 在均值平均精度上分别提升了 5.1 百分点、4.7 百分点和 8.3 百分点。

表 4 DAFFNet 与 SSD 系列算法的实验对比 (%)

模型	AP										mAP
	飞机	棒球场	篮球场	操场	网球场	桥梁	港口	船舶	油罐	汽车	
SSD	99.5	97.9	85.9	99.5	84.1	65.2	90.9	98.9	85.8	47.2	85.5
DSSD	96.7	97.1	79.3	98.9	85.6	64.7	88.8	94.7	87.2	66.5	85.9
FSSD	96.5	99.2	78.1	94.5	76.8	68.1	86.8	83.8	76.9	62.6	82.3
DAFFNet	99.8	91.4	96.5	99.9	87.0	82.0	98.9	97.2	86.2	66.6	90.6

本文所提出的 Group Fusion 模块增强了网络对多尺度特征的学习能力, Attention 模块提升了网络对有效目标的关注度,与 SSD 算法相比,对其中 6 个类别的检测精度均有所提升。在小目标中, DAFFNet 算法在油罐和汽车这两类目标中相较于原始 SSD 算法分别

提升了 0.4 百分点、19.4 百分点。在大目标中,篮球场、操场、桥、港口这四类目标相较于原始 SSD 算法分别提升了 10.6 百分点、0.4 百分点、16.8 百分点、8.0 百分点。与 DSSD 算法相比,DAFFNet 算法对其中 8 个类别的检测精度均有所提升。尤其是篮球场、桥梁、港口这三类目标效果最为明显,分别提升了 17.2 百分点、17.3 百分点、10.1 百分点。与 FSSD 算法相比,DAFFNet 算法除棒球场这类目标以外,对其余所有目标的检测精度均有所提升。

为了综合评估模型的检测与识别能力,本文将 DAFFNet 算法与目前较为流行的 Faster R-CNN、CascadeR-CNN、RetinaNet 算法进行对比,骨干网络分别为 VGG16、ResNet-50、ResNet-50,得到的结果见表 5。在针对小目标与大目标的检测中,本文算法在除了棒球场和汽车这两类目标均达到了单类别检测准确率的最高值。相较于 Faster R-CNN、CascadeR-CNN、RetinaNet,DAFFNet 算法在船舶类别中分别提升了 15.1 百分点、7 百分点、18.3 百分点,在油罐类别中分别提升了 11.8 百分点、9.4 百分点、3.4 百分点。这说明 DAFFNet 算法能够有效地对多尺度目标进行检测与识别。

表 5 DAFFNet 与其他算法的实验对比 (%)

目标	Faster R-CNN	Cascade R-CNN	RetinaNet	DAFFNet
飞机	98.5	98.9	99.5	<b>99.8</b>
棒球场	96.6	<b>98.3</b>	96.3	91.4
篮球场	91.7	92.5	83.4	<b>96.5</b>
操场	99.8	<b>99.9</b>	99.4	<b>99.9</b>
网球场	80.4	73.4	73.8	<b>87.0</b>
桥梁	81.2	61.6	81.7	<b>82.0</b>
港口	93.7	86.4	78.1	<b>98.9</b>
船舶	82.1	90.2	78.9	<b>97.2</b>
油罐	74.4	76.8	82.8	<b>86.2</b>
汽车	75.6	63.5	79.3	66.6

总体而言,从均值平均精度的角度来对比,Faster R-CNN、Cascade R-CNN、RetinaNet 的 mAP 分别为 87.4%、84.2%、85.3%,而本文算法的 mAP 最高,达到了 90.6%。从各类别检测准确率的角度来对比,采用本文提出的算法对遥感图像各个目标进行检测的准确率基本高于采用 Faster R-CNN、Cascade R-CNN 与 RetinaNet 对目标的检测准确率。从模型大小角度来对比,Faster R-CNN、Cascade R-CNN、RetinaNet 大小分别为 1 034 MB、1 016 MB、636 MB,而本文算法的模型

大小最小,约为 130 MB。

### 3.2.3 不同迭代次数的 mAP 值对比

图 6 为在 NWPU VHR-10 数据集上训练的不同迭代次数的 mAP 变化情况。实线代表本文改进方法折线,虚线代表原 SSD 算法折线。横坐标表示训练次数,纵坐标表示训练过程中 mAP 值。可以看出,在不同的迭代次数下,本文方法的 mAP 值基本高于原始 SSD 算法。由整体变化趋势来看,DAFFNet 算法的 mAP 值在前期有少许波动,而后期呈现稳步上升趋势。该算法在第 5 万次迭代时所取得的 mAP 值最高,同时与 SSD 算法的 mAP 值拉开的差距最为明显。

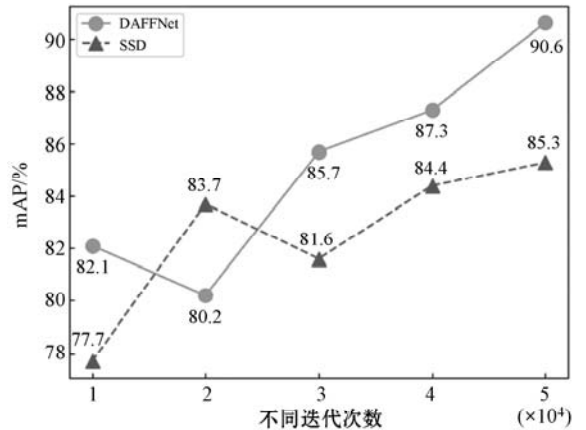


图 6 各迭代次数 mAP 值对比

### 3.2.4 遥感图像目标检测效果对比

为了验证本文算法性能,选用 SSD 算法与本文算法在 NWPU VHR-10 数据集做对比实验。图 7 为两算法实验结果对比图。在第一列中,对于具有不同姿态、尺寸的飞机,两算法均可对其进行准确的检测识别,但由于油罐顶部呈圆形,并且较浅的灰白色与周围明亮的背景相近,因此 SSD 算法出现了误检的现象。在第二列中,由于篮球场的颜色与周围草木颜色接近,因此导致目标物与背景区分度较小。而篮球场中的地面划线原本能够与周围明显区分,却因拍摄距离较远,而不够清晰,因此该篮球场被漏检,同时误将矩形状灰白相间的背景识别为车辆。同理,如第三列所示,整幅图中超过 3/4 的部分均被茂密的绿色背景所占据,其中呈现浅绿色不规则椭圆形的封闭区域被 SSD 算法误检为棒球场。由此可见 DAFFNet 算法能够针对正负样本不平衡以及复杂无关背景的干扰进行有效的改进,从而提升了此类情况目标的检测与识别效果。



(a) SSD 算法



(b) MFA-SSD 算法

图 7 复杂背景下的目标检测结果对比

针对遥感图像中尺度差异大的目标进行了对比分析,图 8 为两算法实验结果对比图。在第一列中,检测油罐这类小而密集的目标时,本文所提出的 DAFFNet 算法能够有效改善小目标漏检的现象,尽管仍有漏检现象发生,总体而言具有良好的检测与识别性能。在第二列中操场这类尺度大的目标,相较于原 SSD 算法本文所提出算法的识别框能够更精准且全面地对操场四周进行覆盖,同时提升了检测精度。由此可见,DAFFNet 算法在目标检测与识别任务中能够有效兼顾大尺度目标与小尺度目标。



(a) SSD 算法



(b) DAFFNet 算法

图 8 大尺度差异下的目标检测结果对比

## 4 结 语

提出了一种优化特征融合的 DAFFNet 算法,旨在提高遥感图像场景下多尺度目标物体的检测精度。通过结合基于分组的特征融合方法和基于注意力机制的多维度特征优化方法充分地融合了多尺度特征信息,并使网络能够自主学习各特征维度间的相关性和重要性,同时利用 Focal Loss 来重新构造损失函数,以提高目标检测准确率。该算法的均值平均精度达到了 90.6%,相较于原算法提高了 5.1 百分点,多项对比实验充分展现出 DAFFNet 拥有较好的检测性能。在未来的工作中,将进一步对网络结构进行优化,考虑加入密集卷积网络提高遥感图像目标检测精度。

## 参 考 文 献

- [ 1 ] Zhong Y F, Han X B, Zhang L P. Multi-class geospatial object detection based on a position-sensitive balancing framework for high spatial resolution remote sensing imagery[J]. ISPRS Journal of Photogrammetry and Remote Sensing,2018, 138:281 - 294.
- [ 2 ] Jia J L, Lai Z Y, Qian Y R, et al. Aerial video trackers review[J]. Entropy,2020,22(12):1358.
- [ 3 ] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2015,39(6):1137 - 1149.
- [ 4 ] Cai Z W, Vasconcelos N. Cascade R-CNN: Delving into high quality object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition,2018:6154 - 6162.
- [ 5 ] Li Y H, Chen Y T, Wang N Y, et al. Scale-aware trident networks for object detection[C]//IEEE International Conference on Computer Vision,2020:6053 - 6062.
- [ 6 ] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition,2016: 779 - 778.
- [ 7 ] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector[C]//European Conference on Computer Vision,2016:21 - 37.
- [ 8 ] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//IEEE International Conference on Computer Vision,2017:2999 - 3007.
- [ 9 ] Zhang S F, Wen L Y, Bian X, et al. Single-shot refinement neural network for object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition,2018:4203 - 4212.
- [ 10 ] Audebert N, Saux B L, Lefevre S, et al. Segment-before-detect: Vehicle detection and classification through semantic segmentation of aerial images[J]. Remote Sensing,2017,9 (4):368 - 386.
- [ 11 ] Yang F, Li W T, Hu H W, et al. Multi-scale feature integrated attention-based rotation network for object detection in VHR aerial images[J]. Sensors,2020,20(6):1686.
- [ 12 ] Fang Z Y, Ren J C, Sun H, et al. SAFDet: A semi-anchor-free detector for effective detection of oriented objects in aerial images[J]. Remote Sensing,2020,12(19):3225.
- [ 13 ] Tayara H, Chong K T. Object detection in very high-resolution aerial images using one-stage densely connected feature pyramid network[J]. Sensors,2018,18(10):3341.

续表 2

识别方法	JAFFE	CK +	Fer2013
深度 ResNet	91.32	92.14	72.21
低功耗 ResNet	88.75	91.51	68.11
本文方法	<b>95.61</b>	<b>97.75</b>	<b>73.6</b>

## 4 结 语

本文在考虑到关键点与表情识别的关联性后,提出了一种基于注意力改进残差网络的表情识别方法,通过双流网络提取关键点并进行粗特征表情识别,利用关键点生成注意力图增大关键点周边特征的权重、减小干扰区域的特征响应值。以残差网络为基础模型,改进了残差块之间的跳跃连接方式,并使用分组卷积来替代残差块中的普通卷积,以更好地提取关键点特征。最后完成特征融合和表情识别。在公开数据集上,将该方法与其他主流方法进行对比研究,结果表明其准确率有较大提高。

## 参 考 文 献

- [1] Yang B, Cao J M, Ni R, et al. Facial expression recognition using weighted mixture deep neural network based on double-channel facial images[J]. IEEE Access, 2018, 6: 4630 - 4640.
- [2] 李勇, 林小竹, 蒋梦莹. 基于跨连接 LeNet-5 网络的面部表情识别[J]. 自动化学报, 2018, 44(1): 176 - 182.
- [3] 王琳琳, 刘敬浩, 付晓梅. 融合局部特征与深度置信网络的人脸表情识别[J]. 激光与光电子学进展, 2018, 55(1): 196 - 204.
- [4] Ozbey N, Topal C. Expression recognition with appearance-based features of facial landmarks[C]//26th Signal Processing and Communications Applications Conference, 2018: 1 - 4.
- [5] Munasinghe M I. Facial expression recognition using facial landmarks and random forest classifier[C]//17th International Conference on Computer and Information Science, 2018: 423 - 427.
- [6] 孔英会, 陈咨彤, 车轺麟. 基于关键子区域及特征提取的表情识别[J]. 科学技术与工程, 2017, 17(34): 257 - 262.
- [7] 杜进, 陈云华, 张灵, 等. 基于改进深度残差网络的低功耗表情识别[J]. 计算机科学, 2018, 45(9): 303 - 307.
- [8] 钱勇生, 邵洁, 季欣欣, 等. 基于改进卷积神经网络的多视角人脸表情识别[J]. 计算机工程与应用, 2018, 54(24): 12 - 19.
- [9] Bilen H, Fernando B, Gavves E, et al. Action recognition with dynamic image networks[J]. IEEE Transactions on

Pattern Analysis & Machine Intelligence, 2016, 40(12): 2799 - 2813.

- [10] Xu Q, Zhang M, Gu Z H, et al. Overfitting remedy by sparsifying regularization on fully-connected layers of CNNs[J]. Neurocomputing, 2019, 328(7): 69 - 74.
- [11] Wang X J, Kan M N, Shan S G, et al. Fully learnable group convolution for acceleration of deep neural networks[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2020: 9041 - 9050.
- [12] 龚跃, 张真真, 黄小珂, 等. 基于动态双向优先级的任务分配与调度算法[J]. 计算机应用, 2009(4): 1131 - 1134.
- [13] Wang X, Liu X G. Learning the discriminate patches from the key landmarks for facial expression recognition[C]//IEEE International Conference on Smart City/SocialCom/SustainCom, 2015: 345 - 348.
- [14] 卢官明, 朱海锐, 郝强, 等. 基于深度残差网络的人脸表情识别[J]. 数据采集与处理, 2019, 34(1): 50 - 57.

## (上接第 161 页)

- [14] Wang J W, Ding J, Guo H W, et al. Mask OBB: A semantic attention-based mask oriented bounding box representation for multi-category object detection in aerial images[J]. Remote Sensing, 2019, 11(24): 2930 - 2951.
- [15] 刘凤, 李华, 南方哲, 等. 优化特征提取的多目标交通标志检测方法[J]. 计算机工程与设计, 2021, 42(5): 425 - 432.
- [16] Liu F, Qian Y R, Li H, et al. CAFFNet: Channel attention and feature fusion network for multi-target traffic sign detection[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2021, 35(7): 2152008.
- [17] Fu C Y, Liu W, Ranga A, et al. DSSD: Deconvolutional single shot detector[EB]. arXiv:1701.06659, 2017.
- [18] Li Z X, Yang L, Zhou F Q. FSSD: Feature fusion single shot multibox detector[EB]. arXiv:1712.00960, 2017.
- [19] Cheng G, Zhou P C, Han J W. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(12): 7405 - 7415.
- [20] Lin T Y, Dollar P, Girshick R, et al. Feature pyramid networks for object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117 - 2125.
- [21] Woo S, Park J C, Lee J Y, et al. CBAM: Convolutional block attention module[C]//European Conference on Computer Vision, 2018: 3 - 19.
- [22] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//IEEE International Conference on Computer Vision, 2017: 2999 - 3007.