

注意力特征融合 SSD 算法对遥感图像的目标检测*

尹法林, 王天一

(贵州大学 大数据与信息工程学院, 贵州 贵阳 550025)

摘要: 针对多尺度单发射击检测(Single Shot MultiBox Detector, SSD)算法对小目标物体检测效果不佳的问题, 提出注意力特征融合 SSD(Attention Feature Fusion SSD, AFF-SSD)算法。首先, 为了提升网络对小目标物体的检测性能, 使用注意力特征融合模块对浅层特征图中的特征信息融合, 在降低噪声的同时增强特征图中远距离像素的相关性; 其次, 针对训练过程中正负样本失衡导致的模型退化问题, 结合聚焦分类损失函数对 SSD 算法中的损失函数优化; 最后, 引入迁移学习解决因训练数据较少导致的过拟合问题。实验结果表明, 与 SSD 算法相比, AFF-SSD 算法平均准确率均值提高 8.09%, 经过迁移学习后, AFF-SSD 算法平均准确率均值提高 3.47%。

关键词: 遥感图像; 目标识别; 注意力特征融合; 损失函数; 迁移学习

中图分类号: TP389

文献标识码: A

DOI: 10.20044/j.csdg.2017-1788.2022.03.011

引用格式: 尹法林, 王天一. 注意力特征融合 SSD 算法对遥感图像的目标检测[J]. 网络安全与数据治理, 2022, 41(3): 67-73.

Target detection of remote sensing image based on attention feature fusion SSD algorithm

Yin Falin, Wang Tianyi

(College of Big Data and Information Engineering, Guizhou University, Guiyang 550025, China)

Abstract: Aiming at the problem that the single shot multibox detector(SSD) algorithm is not effective for small target object detection, an attention feature fusion SSD(AFF-SSD) algorithm is proposed. Firstly, in order to improve the detection performance of the network for small target objects, the attention feature fusion module is used to fuse the feature information in the shallow feature map, which reduces the noise and enhances the correlation of distant pixels in the feature map. Secondly, for the model degradation caused by the imbalance between positive and negative samples in the training process, combined with the focus classification loss function, the loss function in the SSD algorithm is optimized. Finally, transfer learning is introduced to solve the problem of overfitting caused by less training data. The experimental results show that the average accuracy of the AFF-SSD algorithm is increased by 8.09% compared with the SSD algorithm. After the migration, the average accuracy of the AFF-SSD algorithm is increased by 3.47%.

Key words: remote sensing image; target recognition; attention feature fusion; loss function; transfer learning

0 引言

近年来卫星遥感技术在自然灾害救助和高空目标侦察等方面得到广泛的应用, 成为军事侦察、海洋勘测等领域不可缺少的工具^[1-3]。气候、光照等自然条件的影响, 使得识别遥感图像中的目标有很多困难。因此, 对复杂场景下的遥感图像目标检测

识别的研究具有重要的价值^[4-5]。

随着深度学习^[6]在计算机视觉领域的发展, 使用卷积神经网络^[7-8]对图像中的目标进行识别已经成为研究的热门课题。2012 年 AlexNet^[8]网络在 ImageNet 图像分类比赛中成绩突出, 从此出现了以深度卷积神经网络为基础的目标检测算法。之后出现的双阶段目标检测算法 Fast R-CNN^[9]、Faster R-CNN^[10]等进一步提高了检测的精度, 但是检测

* 基金项目: 贵州省科技支撑计划(黔科合支撑[2021]一般 176)

速度很慢。2016 年 Redmon 等在 CVPR 会议上提出统一实时目标检测算法 YOLO^[11],该算法利用回归得到边界框和类别概率,在检测速度上有明显的提升,但是检测精度偏低。同年,Liu 等在 ECCV 会议上提出了多尺度单发射击检测算法 SSD^[12],该算法通过用不同尺度的特征图来提取特征,在满足实时性的同时,提高了检测精度。

针对小目标检测的准确率不高,文献[13]提出了自注意力特征融合模块。遥感数据集中基本都是小目标物体,而小目标物体主要以浅层特征图来检测^[14],因此对 SSD 算法网络中的浅层特征图进行融合,可提高目标检测的准确性;针对训练过程中正负样本失衡导致的模型退化问题,采用聚焦分类损失函数(focal classification loss)^[15-16]对原始的损失函数进行优化。本文在原始 SSD 算法的基础上,提出了注意力特征融合 SSD(Attention Feature Fusion SSD, AFF-SSD)算法,以提升对遥感图像目标检测的平均准确率。

1 AFF-SSD 算法

如图 1 所示,AFF-SSD 算法网络结构是基于 VGG16^[17]网络修改而来的,将 VGG16 网络的 FC6、FC7 全连接层改为 Conv6、FC7 卷积层,并在后面添加了 4 个卷积层(Conv8、Conv9、Conv10、Conv11),然后将 Conv4_3 层和 FC7 层特征融合组成新的 Conv4_3 层,将 FC7 层和 Conv6_2 层特征融合组成新的 FC7 层。整个网络结构有 6 个特征图,这 6 个特征图有不同的感受野,会产生不同尺寸的候选框,预测不同大小的目标物体。在网络结构中,由于浅层特征图感受野小、分辨率高,用来预测小目标物体;深层特征图感受野较大,用来预测较大的物体目标。

网络以 300×300 图片作为输入,通过池化下采样使得图片尺寸逐渐减小,6 个特征图尺寸分别为 38×38,19×19,10×10,5×5,3×3,1×1,每个特征图生成候选框的数量分别为 4,6,6,6,4,4,整个网络会产生 8 732 个候选框来预测目标物体。最后通过非极大值抑制(Non-Maximum Suppression, NMS)算法^[18]过滤掉多余边界框并产生最终检测结果。

1.1 注意力模块

注意力模块本质上是通过对矩阵转置相乘进行相关性计算,增加依赖性较强的特征权重,降低噪声干扰,提高对有效信息的利用率。在卷积神经网络模型中,由于卷积核选择一般都比较小,导致每次卷积操作只能提取部分的特征信息,对于距离卷积核较远的特征信息不容易被提取。为了捕捉数据和特征的内部相关性,使得远距离的特征存在依赖关系,本文提出了通道注意力模块和空间注意力模块,结构如图 2、图 3 所示。

通道注意力模块如图 2 所示,其作用为建立通道之间的依赖关系,增强语义信息的表达。输入特征图 A 维度为 $W \times H \times C$,通过对特征图 A 卷积得到特征图 B 、 C ,维度均为 $W \times H \times C_1$,将特征图 B 进行维度重排和矩阵转置得到矩阵 D ,维度为 $C_1 \times (W \times H)$,将特征图 C 进行维度重排得到矩阵 E ,维度为 $(W \times H) \times C_1$ 。然后矩阵 D 和矩阵 E 相乘,得到特征图矩阵 H ,维度为 $C_1 \times C_1$,将矩阵 H 在列方向上使用 softmax 进行归一化,得到通道注意力矩阵,具体操作如下:

$$B_{ji} = \frac{\exp(H_{ij})}{\sum_{i=1}^{C_1} \exp(H_{ij})} \quad (1)$$

式中 B_{ji} 为第 i 个像素和第 j 个像素之间的相关性,

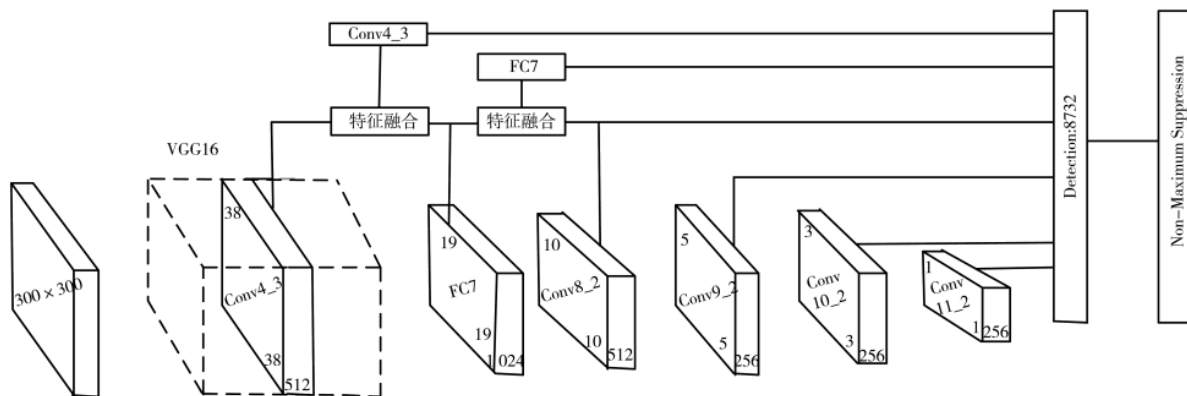


图 1 AFF-SSD 算法网络结构

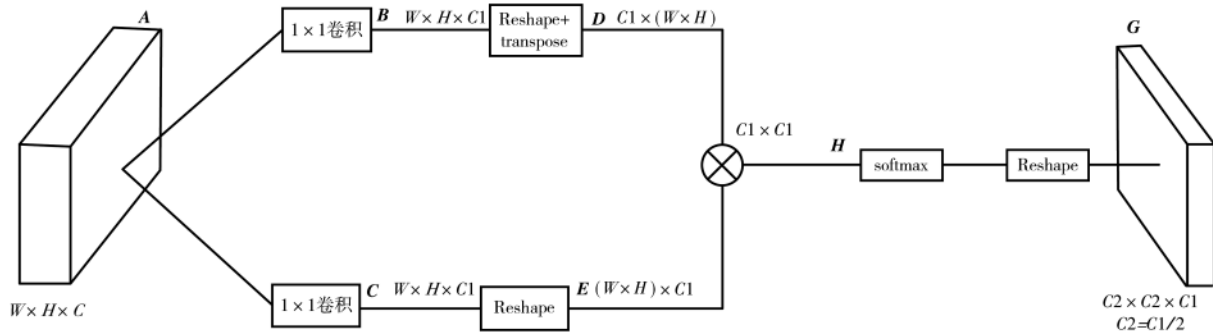


图2 通道注意力模块

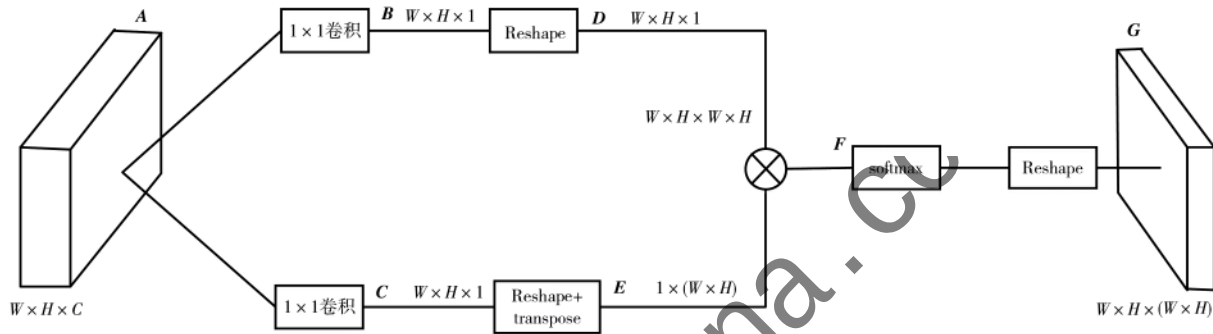


图3 空间注意力模块

二者特征越相似,则转置相乘得到的相关性越大,在注意力特征图像上具有更高的特征值。 H_{ij} 为矩阵 H 中的各元素。最后将矩阵 H 维度重排得到通道注意力特征图 G ,维度为 $C2 \times C2 \times C1$,其中 $C2=C1/2$ 。

空间注意力模块如图3所示,其作用为降低噪声,加强对重要特征信息的提取。输入特征图 A 维度为 $W \times H \times C$,通过对特征图 A 卷积得到特征图 B 、 C ,维度为 $W \times H \times 1$,将特征图 B 进行维度重排得到矩阵 D ,维度为 $W \times H \times 1$,将特征图 C 进行维度重排和矩阵转置得到矩阵 E ,维度为 $1 \times (W \times H)$,然后矩阵 D 和矩阵 E 相乘,得到特征图矩阵 F ,维度为 $W \times H \times W \times H$,将矩阵 F 在列方向上使用softmax进行归一化,得到空间注意力矩阵。最后将矩阵 F 维度重排得到空间注意力特征图 G ,维度为 $W \times H \times (W \times H)$ 。

1.2 注意力特征融合模块

在SSD算法中,不同特征图的特征信息没有融合,使得用于检测小目标物体的浅层特征图缺乏足够的语义信息,因此对小目标物体的检测效果不好。传统的特征融合方法都要进行上采样操作,而上采样会引入大量噪声,对小物体目标识别会有影响。同时在卷积神经网络中,由于卷积核尺寸一般设置都比较小,导致每次卷积操作只能覆盖卷积核

附近的像素点区域。对于图片上距离较远的相关特征信息不容易被卷积核捕获。针对以上问题,结合通道注意力模块和空间注意力模块,本文提出了注意力特征融合模块,对SSD算法卷积神经网络浅层特征图特征融合。

图4所示为注意力特征融合模块,在模块中有两个输入:第 n 个特征图和第 $n+1$ 个特征图。

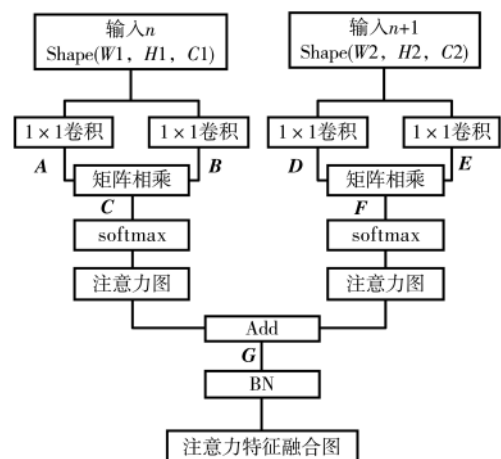


图4 注意力特征融合模块

第 n 个特征图,输入维度为 $W1 \times H1 \times C1$,其中 $W1$ 为输入宽, $H1$ 为输入高, $C1$ 为输入通道数。首

先,通过 2 个 1×1 卷积降低通道数,使得通道数变为 1,维度变为 $W1 \times H1 \times 1$ 。然后,合并宽和高得到矩阵 A 和矩阵 B ,矩阵 A 维度为 $W1 \times H1$;矩阵 B 维度为 $W1 \times H1$ 。将矩阵 A 与矩阵 B 相乘,得到矩阵 C ,维度 $W=W1 \times H1$,将矩阵 C 在列方向上使用 softmax 归一化,得到特征图上所有特征点与某一特征点的相关性。将矩阵 C 的维度展开得到第 n 个特征图的空间注意力图。

第 $n+1$ 个特征图,输入维度为 $W2 \times H2 \times C2$,其中 $W2$ 为输入宽, $H2$ 为输入高, $C2$ 为输入通道数。首先,通过 2 个 1×1 卷积改变通道数,使得通道数 $C2=W1 \times H1$ 。然后合并宽和高,得到矩阵 D 和 E 。矩阵 D 维度为 $(W, N; W=W1 \times H1, N=W2 \times H2)$, 矩阵 E 维度为 $(N, W; N=W2 \times H2, W=W1 \times H1)$, 将矩阵 D 与矩阵 E 相乘,得到矩阵 F ,其维度 $W=W1 \times H1$,将矩阵 F 在列方向上使用 softmax 归一化,得到特征图上所有特征点与某一特征点的相关性。将矩阵 F 的维度展开得到第 $n+1$ 个特征图的通道注意力图。将第 n 个特征图的空间注意力图与第 $n+1$ 个特征图的通道注意力图使用特征增强方法特征融合。最后,为了增加特征多样性,采用通道拼接的特征融合方法与第 n 个特征图特征融合,得到注意力特征融合图。

在 SSD 网络中,主要由 Conv4_3 层、FC7 层、Conv8_2 层来检测小目标物体,相比于 Conv4_3 层,FC7 层拥有更多的语义信息,而 Conv4_3 层拥有更多的特征信息;与 FC7 层相比,Conv8_2 层拥有更多的语义信息,而 FC7 层拥有更多的特征信息。因此,在 FC7 层增加通道注意力模块,可以提取更多的语义信息,在 Conv4_3 层增加空间注意力模块,可以加强对重要信息的提取,降低噪声的影响,增强对小物体目标的检测能力。

1.3 损失函数优化

为避免因正负样本分布不均导致模型训练退化的问题,结合聚焦分类损失函数,对 SSD 算法的损失函数进行优化,优化后的损失函数由聚焦分类损失和位置回归损失加权求和获得,公式如下:

$$L(x, c, p, l, g) = \frac{1}{N} (L_{cl}(x, c, p) + \alpha L_{lrf}(x, l, g)) \quad (2)$$

聚焦分类损失公式如下:

$$L_{cl}(x, c, p) = - \sum_{i \in \text{pos}} x_{ij}^k p \log(\hat{C}_i^k) - \sum_{i \in \text{Neg}} (1-p) \log(\hat{C}_i^0) \quad (3)$$

其中, x_{ij}^k 中 i 为候选框的序号, j 为真实框序号, k 为类别序号; p 为预测类别概率; N 是匹配到真实框的默认框的数量。公式前半部分 $-\sum_{i \in \text{pos}} x_{ij}^k p \log(\hat{C}_i^k)$

是正样本的损失,后半部分 $-\sum_{i \in \text{Neg}} (1-p) \log(\hat{C}_i^0)$ 是负样本的损失, $\hat{C}_i^k$ 是 softmax 函数。

位置回归损失公式如下:

$$L_{lrf}(x, l, g) = \sum_{i \in \text{pos}} \sum_{m \in \{cx, cy, w, h\}} x_{ij}^k \text{smooth}_{L1}(L_i^m - \hat{g}_j^m) \quad (4)$$

其中, cx, cy 是中心点坐标; w, h 分别为默认框的宽和高; L_i^m 为候选框的值, $\hat{g}_j^m$ 表示真实框相对于候选框编码。

1.4 迁移学习

迁移学习是解决小样本数据集相关问题的一个很好方法,其将一个任务上训练好的卷积神经网络模型迁移到另一个任务中,然后通过简单的调整使其适用于一个新的任务。通过模型参数共享,将在大规模数据集中预训练好的网络模型运用到其他任务中,作为其他任务的特征提取器。图 5 为 AFF-SSD 算法迁移过程,该算法网络中包含有大量要训练的参数,而训练这些参数需要大量的训练数据,由于本文使用的遥感数据集较少,因此在训练网络时容易过拟合,为解决该问题,本文采用迁移学习方法训练网络。

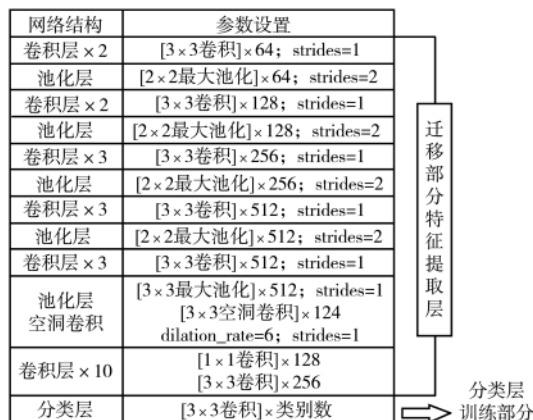


图 5 迁移学习过程

2 数据集

本文使用的遥感数据集较小,训练时比较困难,

因此在训练时进行数据增强。图像变换方式包括随机亮度变化($-30\% \sim 30\%$)、随机平移($-20\% \sim 20\%$)、随机旋转($30^\circ \sim 180^\circ$)、随机翻转(水平或垂直翻转)、尺寸变化($-20\% \sim 100\%$)、随机裁剪等。

首先,在 VOC2012 数据集进行预训练,将预训练模型的特征提取层和相应权重迁移到本文的训练任务;然后,根据本文训练任务重新设计一个分类层,和迁移过来的特征提取层重新组建一个新的卷积神经网络,用于本文数据集的训练。实验中采用两种方法进行训练:(1)冻结特征提取层,只训练分类层;(2)所有层都训练。

3 实验结果与分析

3.1 实验环境及训练方式

在 Python3.7、TensorFlow2.0 环境下进行实验,应用 labeling 标注数据集;用 RTX 2080Ti 加速训练,CPU 为 i7-9700 3.6k@3.60 GHz*8。实验使用 Adam 优化器,初始学习率设为 0.000 4, batch_size 设置为 8, Dropout 为 0。

3.2 实验结果与分析

表 1 为不同融合方式在测试集上的实验结果。本文采用平均精度(Average Precision, AP)、平均检测精度(mean Average Precision, mAP)、平均检测时间三个评价指标对不同融合方式进行检验。AP 是评价目标检测精度的最重要的指标之一,其大小可是以查全率为横轴、召回率为纵轴所围成的面积大小,所围成面积越大,AP 值越高,对目标检测越准确。mAP 用来判别对多目标物体的检测精度,对每个类别求出 AP 再求和,然后除以类别数便得到

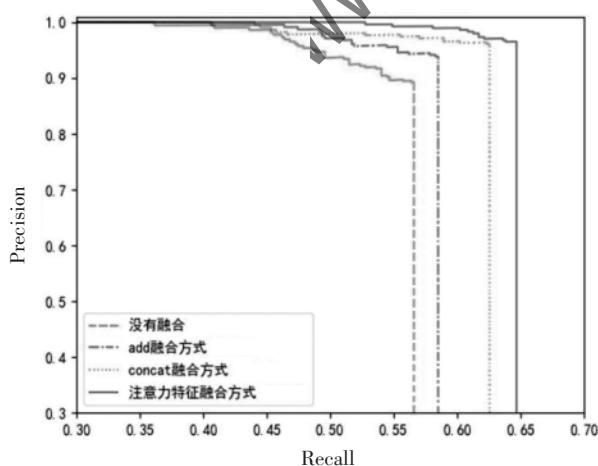
表 1 不同融合方式在测试集上的实验结果

实验 算法	融合 方式	类别	AP/%	mAP/%	平均检测 时间/ms
SSD	-	Aircraft	55.78	63.63	31.33
		Oiltank	71.48		
SSD	add (特征增强)	Aircraf	58.05	65.95	43.68
		Oiltank	73.86		
SSD	concat (通道拼接)	Aircraf	60.12	67.23	45.03
		Oiltank	74.35		
SSD	本文方法	Aircraf	64.51	71.72	38.75
		Oiltank	78.93		

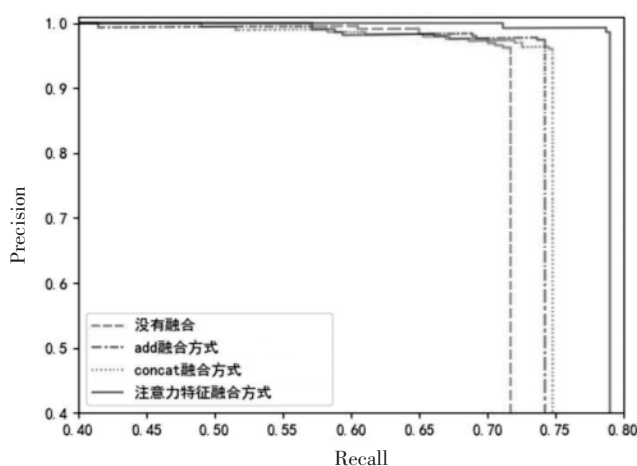
mAP 的值,mAP 的取值范围是 $[0, 1]$,其值越高检测精度越好。除了检测精度,目标检测的另一个检测指标是平均检测时间,只有平均检测时间足够短,才能实现实时检测。

为了检测注意力特征融合方法的效果,实验中与传统的融合方法做了比较。由实验结果可知,与传统 SSD 算法相比,经过注意力特征融合后的 SSD 算法,虽然平均检测时间增加了 7.42 ms,但是平均准确率均值提高了 8.09%;与特征增强、通道拼接等传统融合方法相比,无论平均检测时间还是平均准确率均值,都有大幅度的提高。

图 6 为不同融合方式下的 PR 曲线。在图 6 中,垂直直线部分不在 PR 曲线内,只是一条辅助线,为了显示出不同 PR 曲线与坐标轴围成面积的差异,从而体现不同融合方式下 AP 值的差异。以垂直直线为区分条件,从左向右第一条为没有融合的 PR 曲线,第二条为 add(特征增强)融合方式的



(a) 飞机类别在不同融合方式下的 PR 曲线



(b) 油桶类别在不同融合方式下的 PR 曲线

图 6 不同类别在不同融合方式下的 PR 曲线

PR 曲线,第三条为 concat(通道拼接)特征融合方式的 PR 曲线,第四条为本文注意力特征融合方式的 PR 曲线。由图 6 可知注意力特征融合方式的 PR 曲线所围成的面积最大,即该方式的 AP 值最高。

在图 7 中,灰色框(上面一层)为 SSD 算法的检测效果,白色框(下面一层)为 AFF-SSD 算法的检测效果,可以看出,AFF-SSD 算法对小目标物体的检测效果好于 SSD 算法。

表 2 为不同训练方式在测试集上的实验结果,由实验结果可知,平均准确率均值最低为迁移学习只训练分类层的训练方式,平均准确率均值为 57.27%。原因是本文使用的数据集与 VOC2012 数据集存在差异,只训练分类层,特征提取层参数无法得到更新。而迁移学习训练全部层的平均准确率均值最高,达到 75.19%,与迁移学习只训练分类层相比,平均准确率均值提高 17.92%,与全新训练方式相比,平均准确率均值提高 3.47%,因此,本文使用

表 2 不同训练方式在测试集上的实验结果

实验算法	训练方式	类别	AP/%	mAP/%
AFF-SSD	全新训练	Aircraft	64.51	71.72
		Oiltank	78.93	
AFF-SSD	训练分类层	Aircraft	48.70	57.27
		Oiltank	65.84	
AFF-SSD	训练全部层	Aircraft	67.39	75.19
		Oiltank	82.99	

迁移学习训练全部层作为迁移学习实验结果。

图 8 为 AFF-SSD 算法迁移学习后的 PR 曲线。在图 8 中,垂直直线部分不在 PR 曲线内,只是一条辅助线,为了显示出 PR 曲线与坐标轴围成面积,从而体现出 AP 值。由图 8 可知经过迁移学习后,飞机类别和油桶类别 PR 曲线所围成的面积比未经过迁移学习所围成的面积大,即经过迁移学习后,对目标检测的 AP 有了进一步的提高。

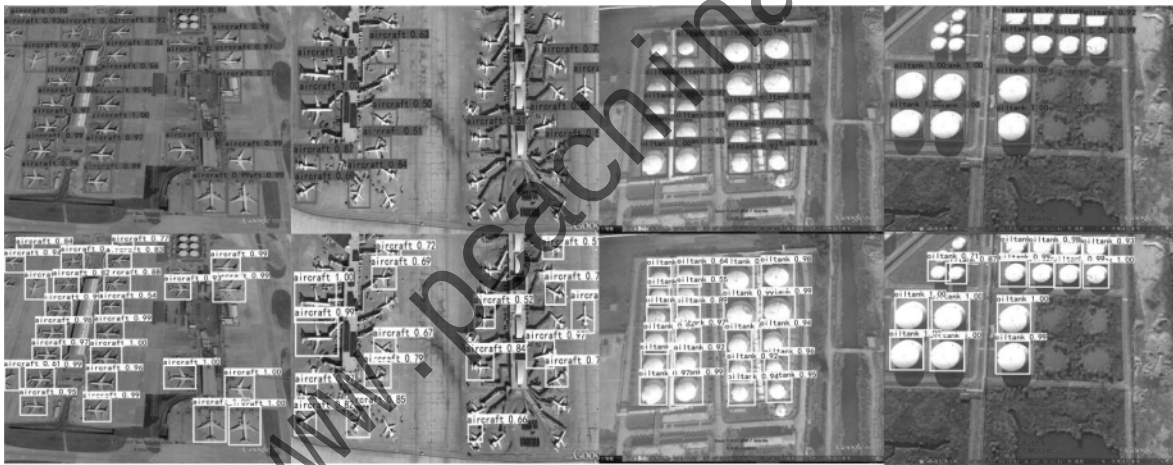


图 7 SSD 算法与 AFF-SSD 算法检测效果对比

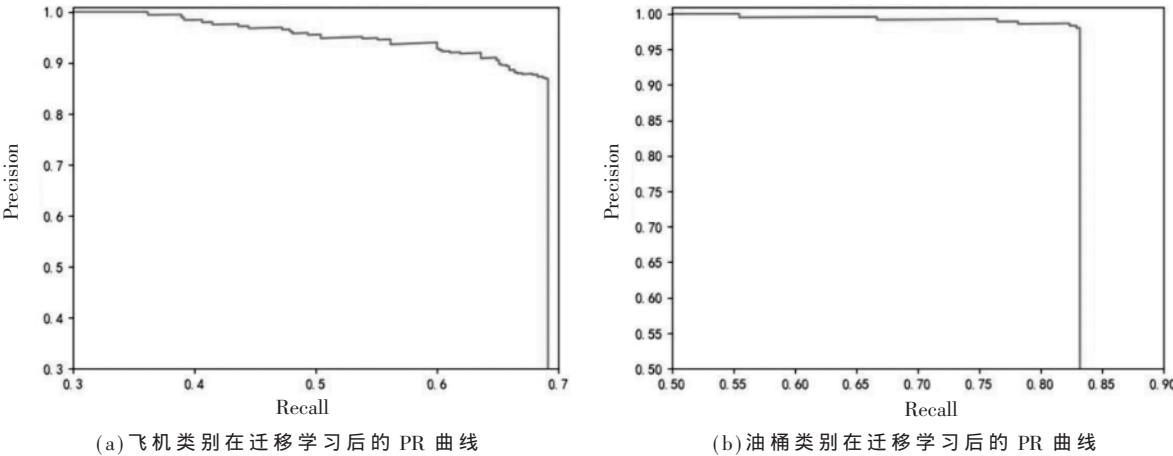


图 8 不同类别在迁移学习后的 PR 曲线

4 结论

针对 SSD 算法对遥感图像小目标物体检测效果不佳的缺陷,本文提出了一种特征融合增强的 SSD 小目标检测算法——AFF-SSD 算法。该算法的核心思想是加强 SSD 算法浅层特征图之间的信息交流,使用注意力特征融合模块对 SSD 算法网络结构的浅层特征图进行特征融合,增强浅层特征图的语义信息,从而提高小目标的检测准确率。此外,针对训练过程中正负样本失衡导致模型退化的问题,对原始的损失函数进行优化。最后,使用迁移学习的方法对网络进行训练,进一步提升检测精度。由实验结果可知,该算法对小目标具有良好的检测效果。

参考文献

- [1] XU H J, YANG Z W, CHEN G Z, et al. A ground moving target detection approach based on shadow feature with multi-channel high-resolution synthetic aperture radar[J]. IEEE Geoscience and Remote Sensing Letters, 2016, 13(10): 1572–1576.
- [2] KIM S, SONG W J, KIM S H. Robust ground target detection by SAR and IR sensor fusion using ada-boost-based feature selection[J]. Sensors, 2016, 16(7): 1117–1134.
- [3] CHENG G, HAN J W. A survey on object detection in optical remote sensing images[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2016, 117: 11–28.
- [4] YI W, ZENG Y, YUAN Z. Fusion of SAR and optical images based on the nonsubsapled contourlet transform[J]. Acta Optica Sinica, 2018, 38(11): 1110002.
- [5] BALL J E, ANDERSON D T, CHAN C S. A comprehensive survey of deep learning in remote sensing: theories, tools and challenges for the community[J]. Journal of Applied Remote Sensing, 2017, 11(4): 042609.
- [6] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521: 436–444.
- [7] XIE S, GIRSHICK R, DOLLAR P, et al. Aggregated residual transformations for deep neural networks[C]//Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 1492–1500.
- [8] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[C]//Proceedings of Neural Information Processing Systems, 2012: 1097–1105.
- [9] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision. Santiago: IEEE, 2015: 1440–1448.
- [10] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]//Advances in Neural Information Processing Systems, 2015: 91–99.
- [11] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779–788.
- [12] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//European Conference on Computer Vision, 2016: 21–37.
- [13] 王冬丽, 廖春江, 牟金震, 等. 基于特征融合的 SSD 视觉小目标检测[J]. 计算机工程与应用, 2020, 56(16): 31–36.
- [14] 储岳中, 黄勇, 张学锋, 等. 基于自注意力的 SSD 图像目标检测算法[J]. 华中科技大学学报(自然科学版), 2020, 48(9): 70–75.
- [15] 史文旭, 谭代伦, 鲍胜利. 特征增强 SSD 算法及其在遥感目标检测中的应用[J]. 光子学报, 2020, 49(1): 154–163.
- [16] 沈新峰, 姜平, 周根荣. 改进 SSD 算法在零部件检测中的应用研究[J]. 计算机工程与应用, 2021, 57(7): 257–262.
- [17] ZHU L, CHEN Y, GHAMISI P, et al. Generative adversarial networks for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(9): 5046–5063.
- [18] NEUBECK A, VAN G L. Efficient non-maximum suppression[C]//18th International Conference on Pattern Recognition(ICPR'06). Hong Kong, China: IEEE, 2006: 850–855.

(收稿日期: 2022-06-07)

作者简介:

尹法林(1996–), 男, 硕士研究生, 主要研究方向: 深度学习、图像处理。

王天一(1989–), 通信作者, 男, 博士, 副教授, 主要研究方向: 量子通信、图像处理。E-mail: tywang@gzu.edu.cn。

版权声明

凡《网络安全与数据治理》录用的文章，如作者没有关于汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权等版权的特殊声明，即视作该文章署名作者同意将该文章的汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权授予本刊，本刊有权授权本刊合作数据库、合作媒体等合作伙伴使用。同时，本刊支付的稿酬已包含上述使用的费用，特此声明。

《网络安全与数据治理》编辑部

www.pcachina.cn