

基于边缘特征融合的高分影像建筑物目标检测



赫晓慧¹ 邱芳冰² 程渐杰² 田智慧¹ 周广胜³

¹ 郑州大学地球科学与技术学院 郑州 450052

² 郑州大学信息工程学院 郑州 450001

³ 中国气象科学研究院郑州大学生态气象联合实验室 郑州 450052

摘要 高分辨率遥感图像建筑物目标检测在国土规划、地理监测、智慧城市等领域有着广泛的应用价值,但是由于遥感图像背景复杂,建筑物目标的部分细节特征与背景区分度较低,在进行检测任务时,容易出现建筑物轮廓失真、缺失等问题。针对这一问题,设计了自适应加权边缘特征融合网络(VAF-Net)。该方法针对遥感图像建筑物检测任务,对经典编解码器网络 U-Net 进行拓展,通过融合 RGB 特征图和边缘特征图,弥补了基础网络学习中的细节特征缺失;同时,借助网络的学习自动更新融合权重,实现自适应加权融合,充分利用不同特征的互补信息。该方法在 Massachusetts Buildings 数据集上进行了实验,其准确率、召回率和 F1-score 分别达到了 82.1%,82.5% 和 82.3%,综合指标 F1-score 相比于基础网络提升了约 6%。VAF-Net 有效提高了编解码器网络对于高分影像建筑物目标检测任务的表现性能,具有良好的实用价值。

关键词: 目标检测;特征融合;神经网络;边缘特征;U-Net

中图分类号 TP391.4

High-resolution Image Building Target Detection Based on Edge Feature Fusion

HE Xiao-hui¹, QIU Fang-bing², CHENG Xi-jie², TIAN Zhi-hui¹ and ZHOU Guang-sheng³

¹ School of Earth Science and Technology, Zhengzhou University, Zhengzhou 450052, China

² School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China

³ Joint Laboratory of Eco-Meteorology, Chinese Academy of Meteorological Sciences, Zhengzhou University, Zhengzhou 450052, China

Abstract High-resolution remote sensing image building target detection has a wide range of application value in territorial planning, geographic monitoring, smart city and other fields. However, due to the complex background of remote sensing images, some detailed features of building targets are less distinguishable from the background. During the task, it is prone to problems such as distortion and missing of the building outline. Aiming at this problem, an adaptive weighted edge feature fusion network (VAF-Net) is designed. This method is aimed at remote sensing image building detection tasks, expands the classic codec network U-Net network, and makes up for the lack of detailed features in basic network learning through the fusion of RGB feature maps and edge feature maps. At the same time, relying on network learning to automatically update the fusion weight, adaptive weighted fusion can be achieved, and the complementary information of different features can be full made use of. This method is tested on the Massachusetts Buildings data set, and its accuracy, recall and F1-score reach 82.1%, 82.5% and 82.3%, respectively. The comprehensive index F1-score increases by about 6% compared to the basic network. VAF-Net effectively improves the performance of the codec network for high-resolution image building target detection tasks, and has good practical value.

Keywords Target detection, Feature fusion, Neural network, Edge feature, U-Net

1 引言

近年来,随着遥感技术的不断发展,遥感图像的数量日益增长,空间分辨率不断提高,光谱信息更加丰富,为遥感图像各领域的研究提供了重要的分析条件和资源。目标检测作为遥感图像处理的一个重要部分,在军事领域和民用领域都具有重要的意义。从遥感影像中对建筑物进行目标检测在 GIS

数据库更新、军事侦察、土地利用分析、城市规划、灾害评估等领域有着广泛的应用^[1]。

传统的建筑物目标检测主要依靠底层视觉特征,如颜色、纹理、形状等。这类方法普适性较差,检测性能取决于人工筛选的低级特征,表达能力有限。近年来,随着计算机硬件的发展和超大规模学习样本的出现,以深度学习为代表的卷积神经网络在目标检测领域表现出了出色的检测性能,各类优秀

收稿日期:2020-08-01 返修日期:2020-09-10 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:第二次青藏高原综合科学考察研究项目(2019QZKK0106)

This work was supported by the Second Tibetan Plateau Scientific Expedition and Research(STEP) Program(2019QZKK0106).

通信作者:赫晓慧(hexh@zzu.edu.cn)

的神经网络模型极大地提高了建筑物目标检测领域的检测准确性。FCN^[2]网络的提出是目标检测领域的一个重大突破^[3],它摒弃了全连接层,并引入反卷积,解决了计算量大的缺点。但是,FCN忽略了像素与像素的关系,检测结果不够精细。U-Net^[4]在FCN的基础上对原网络结构进行了延伸,并增加了跳连结构,使其可以很好地捕捉上下文信息,保证了目标检测的精度,该网络也被相关学者应用于建筑物目标检测任务^[5]。Xu等^[6]利用改进的U-Net模型对建筑物区域进行像素级提取,在建筑物语义分割任务上获得了较高的精度。Ren等^[7]针对建筑物与周围环境信息混淆导致提取精度下降的问题,提出了一种低维特征信息增强型的U-Net网络,进一步优化了建筑物的提取效果。

然而,使用单一特征进行目标检测时,难以避免建筑物的细节信息在网络学习过程中的损失,使得网络无法准确地提取和表示相关特征,从而导致检测效果不理想。在这种情况下,相关学者尝试将特征融合^[8]的思想应用到建筑物目标检测任务中,利用纹理、颜色等传统特征的融合来弥补网络学习中细节信息的缺失。Sheng等^[9]将纹理特征应用于高分影像建筑区的提取任务中,提取精度提高了近5%。Feng等^[10]在单一输入的基础上,增加了灰度图像矩阵作为扩展输入,通过不同特征的融合进一步提高了检测精度。同时,为了充分发挥不同特征在目标检测中的优势,不少学者提出了加权融合的方法^[11]。Feng等^[12]提出了一种基于多输入多输出和多特征融合的全卷积网络。一方面将输入影像进行不同比例下采样,然后将结果和网络中对应编码段的特征进行融合;另一方面将不同解码段输出的特征图上采样后连接在一起实现跨尺度特征聚合。以上两种特征融合的方式构成了对称的多输入多输出结构,强化了网络对复杂地物的分析能力。Wang等^[13]针对网络中不同层提取出的特征图存在尺度、信息差异,给每个特征图设置可学习的融合权重,该方法在建筑物识别任务上取得了较好的效果。以上方法在高分辨率影像的建筑物识别任务上取得了一定的效果,但是由于光照、角度和阴影等因素影响,建筑物目标的部分细节特征与背景区分度较低^[14],所检测到的建筑物的轮廓经常存在一定程度的缺失或失真。研究表明,遥感图像中的边缘较好地保存了背景和目标之间的阶跃变化的细节信息,适当地融入边缘特征可以有效改

善检测结果中建筑物轮廓失真的情况^[15]。

因此,为了解决以上问题,本文基于特征融合的思想,结合边缘检测算法,提出了自适应加权边缘特征融合网络VAF-Net。首先,在原目标检测网络U-Net的基础上,通过增加卷积池化结构丰富网络的参数,弥补U-Net网络参数量过少的缺陷。在此基础上,构建边缘特征检测子网,利用轮廓细节特征与基础网络特征的融合,丰富特征图中的边缘细节信息。同时,考虑到不同特征对检测任务的贡献不同,引入自适应加权融合的方法,借助网络的学习自动更新融合权重,动态调节不同特征图的融合比例,进一步提高建筑物目标检测的效果。为验证本文方法的有效性,基于Massachusetts Buildings数据集进行实验,并将其与其他方法进行对比分析。

2 边缘特征融合网络的设计

2.1 基础网络UV-Net

与普通图像的目标检测相比,遥感图像背景复杂,建筑物形态各异、大小不一,因此一般的神经网络不能很好地完成建筑物检测任务^[16]。编解码网络架构具有较好的细节分辨率,更适用于遥感影像的目标提取任务。U-Net是由特征编码器和特征解码器构成的“U”形网络,网络前端为特征编码器,通过卷积和下采样操作将图像上下文信息转换为特征图;网络后端为特征解码器,将提取的特征图转换为二值化图像。此外,针对特征信息在传递过程中的损失问题,该网络将特征编码器中的特征信息与解码器中相对应的特征图进行叠加融合,从而增加了特征图像的细节信息。

然而由于U-Net网络参数偏少,在进行网络训练的过程中容易出现过拟合现象^[17],因此本文借鉴VGG模型的网络参数对U-Net网络进行改进,设计了建筑物目标检测基础网络UV-Net。如图1所示,在U-Net网络的基础上,新增了一组“上采样-卷积”块,并将有的跳连结构进行了相应的调整,形成了对称的“V”形结构。同时,为了增强网络的细节信息的学习能力,在编码器部分的“卷积-池化”块中分别增加一层卷积,在与之相对应的解码器部分也相应地加入一层卷积。卷积层均采用步长为1、尺寸为 3×3 的卷积核进行特征提取,池化层利用步长为2、尺寸为 2×2 的卷积核进行最大下采样操作。

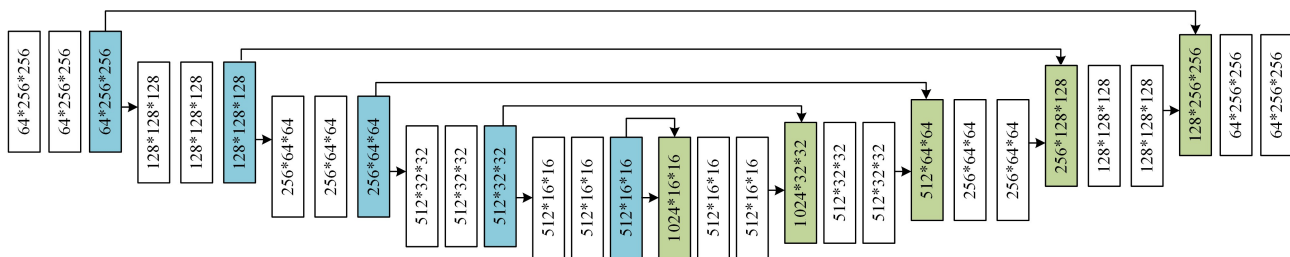


图1 UV-Net网络结构

Fig. 1 UV-Net network structure

2.2 基于UV-Net的特征融合网络VF-Net

由于遥感图像场景复杂,单一模型表征能力不足,仅仅使用UV-Net网络进行建筑物目标检测任务时,其检测结果无法令人满意。因此本文在基础网络UV-Net的基础上,增加

边缘特征检测子网,通过融合RGB特征图和边缘特征图,丰富特征图的细节信息,提高了建筑物目标检测精度。

为了得到干净有效的边缘信息,我们对RGB影像进行了边缘检测。Sobel算子^[18]是整像素边缘检测常用的算子之

一, 由于其在灰度渐变和噪声较多的图像的边缘检测任务中表现出色, 因此本文将其应用于遥感图像的边缘检测, 并以其检测结果作为边缘检测子网的输入图像。该算子以图像灰度值为基础, 利用图像边缘的一阶导数来检测目标的边缘。Sobel 算子的定义为:

$$S = (d_x^2 + d_y^2)^{\frac{1}{2}} \quad (1)$$

其中:

$$d_x = \Delta_x f(x, y) = \{f(x+1, y-1) + 2f(x+1, y) + f(x+1, y+1)\} - \{f(x-1, y-1) + 2f(x-1, y) + f(x-1, y+1)\} \quad (2)$$

$$d_y = \Delta_y f(x, y) = \{f(x-1, y+1) + 2f(x, y+1) + f(x+1, y+1)\} - \{f(x-1, y-1) + 2f(x, y-1) + f(x+1, y-1)\} \quad (3)$$

基于 Sobel 算子检测得到的边缘图像, 本文在 UV-Net 的基础上设计了边缘特征融合的目标检测网络 VF-Net。如图 2 所示, 该网络分为两大部分, 分部是目标检测基础网络和边缘特征学习子网, 网络中的 f 表示融合边缘特征图和 RGB 特征图的融合模块。

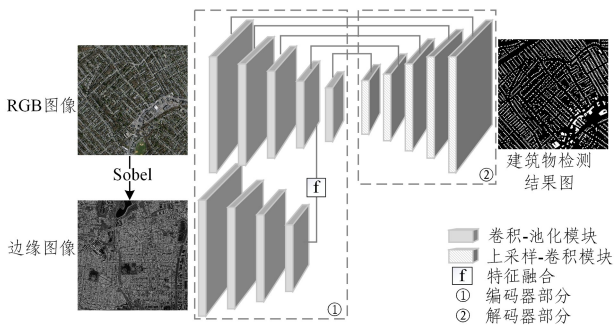


图 2 边缘特征图融合网络 VF-Net

Fig. 2 Edge feature fusion network VF-Net

UV-Net 的编码器部分由 5 组“卷积-池化”结构块组成, 作为基础检测网络, 其输入为原始的 RGB 图像。编码器在逐层降采样的过程中对 RGB 图像进行特征提取, 特征图中的全局上下文逐层增多, 但损失了目标的位置信息和边缘细节信息。因此, 本文在 UV-Net 的基础上增加边缘检测子网, 并通过特征图之间的融合丰富基础网络中的细节信息。

大多数融合网络进行融合的特征来自编码器的最后一层特征, 同时该类融合网络将融合后的特征图直接送入解码器进行解码操作, 这不利于后期解码器的学习和分类^[19]。因此在本文的网络设计中, 边缘特征学习子网的结构采用了基础网络编码器的前 4 层。边缘特征学习子网的特征图与 RGB 特征学习子网第 4 层的特征图进行融合, 并在第 5 组“卷积-池化”的学习之后, 再进入解码器部分, 这保证了不同特征的充分融合, 充分利用了边缘信息对目标检测的贡献。

通过引入边缘特征融合策略, 编码器在逐层降采样的过程中很好地保留了建筑物的边缘信息。为了更加充分地保留和利用边缘信息, 解码网络保留了基础网络的跨链接结构, 通过编码器特征图与解码器特征图的融合, 使网络可以更好地恢复部分细节信息。网络最终通过 Softmax 层输出逐像素分类的结果。

2.3 自适应加权融合模块

VF-Net 在基础检测网络 UV-Net 的基础上, 加入了边缘检测子网, 并进行了边缘特征图和基础网络特征图的融合。然而在 VF-Net 网络进行特征融合的过程中, 没有考虑到不同类型的特征对不同场景下目标检测的不同贡献, 直接进行等权重融合, 虽然在一定程度上弥补了基础网络中的特征损失, 但检测结果的精确度并不理想。因此, 为了充分利用两种图像特征的互补信息, 本文提出了一种自适应加权融合方法, 借助网络的学习, 自适应更新不同特征的融合权重, 保证所融合特征的贡献最大化。

在特征融合模块, 对进行融合的两特征图赋予不同的融合权重 w , 并对其赋予 $w=0.5$ 的初值进行融合。由于在网络训练的过程中, 网络可以利用反向传播^[20]将输出层的误差信息反向传播给所有神经元, 完成对网络中的权重的更新。借助这一过程, 网络可以在学习的过程中对融合权重 W 进行权重修正, 使其以适当的比例进行融合。自适应加权融合的流程如图 3 所示。

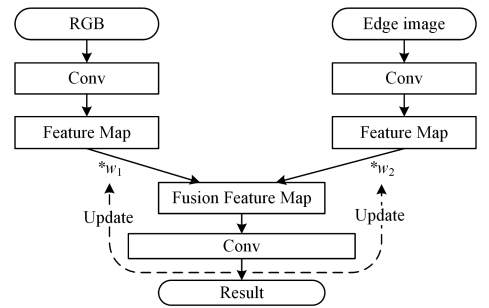


图 3 自适应加权融合流程

Fig. 3 Adaptive weighted fusion process

令 X^1 代表原始特征图, X^2 代表边缘特征图, 并分别赋予其融合权重 w_1, w_2 , 则融合后的特征图可以表示为:

$$Y = w_1 * X^1 \odot w_2 * X^2 \quad (4)$$

其中, w_1 和 w_2 是简单的标量变量, 在所有通道之间可共享; \odot 代表 Concat 操作。同时, 定义 $w_1, w_2 \in [0, 1]$ 且 $w_1 + w_2 = 1$ 。在网络学习的过程中, 将 w_1 和 w_2 作为网络中的参数进行学习, 利用神经网络反向传播算法判断识别误差、更新梯度。目标检测网络自适应地学习边缘特征和基础网络特征对目标检测任务的不同贡献, 在网络学习的过程中对融合权重进行更新, 从而达到自适应加权融合的效果。

$$w_1 = \frac{e^{\lambda_{w_1}^l}}{e^{\lambda_{w_1}^l} + e^{\lambda_{w_2}^l}}, w_2 = \frac{e^{\lambda_{w_2}^l}}{e^{\lambda_{w_1}^l} + e^{\lambda_{w_2}^l}} \quad (5)$$

其中, $\lambda_{w_1}^l, \lambda_{w_2}^l$ 通过 Softmax 函数进行定义, 作为控制参数对 w_1, w_2 进行更新。

3 实验及结果分析

为了验证本文提出的 VAF-Net 网络在遥感图像目标检测任务中的有效性, 基于 Massachusetts 建筑物数据集进行实验, 并选择 Precision, Recall, F1-score 作为网络性能的评价标准。

3.1 评价标准

本文采用精确率(Precision, P)、F1 值(F1-score)和召回率(Recall, R)3 个指标作为衡量标准。

精确率是针对预测结果而言的,它表示的是预测为正的样本中有多少是真正的正样本。因此在进行预测时,有正类预测为正类(TP)以及负类预测为正类(FP)两种可能性,则精确率 Precision 的计算如下:

$$Precision = TP / (TP + FP) \quad (6)$$

召回率是针对原来的样本而言的,它表示的是样本中的正例有多少被正确预测。那么其也有两种可能,一种是把原来的正类预测成正类(TP),另一种是把原来的正类预测为负类(FN)。

$$Recall = TP / (TP + FN) \quad (7)$$

F1-score 是基于召回率和精确率计算的,同样是针对某个样本而言的。一般 F1-score 用来综合 Precision 和 Recall。

$$F1-score = 2 * Precision * Recall / (Precision + Recall) \quad (8)$$

3.2 数据集介绍

本文所使用的实验数据来自 Massachusetts 建筑物数据集^[21]。该数据集由波士顿地区的 151 张航拍图像组成,并随机分为 137 个图像的训练集、10 个图像的测试集和 4 个图像的验证集。数据集中所有图像的大小均为 1500×1500 ,分辨率为 1m。每幅影像覆盖了 2.25 km^2 的区域,整个数据集大约覆盖 340 km^2 。建筑物标签来自 OpenStreetMaps。

3.3 参数设置

本文所提供的网络结构基于 TensorFlow 实现,网络训练过程基于单块 NVIDIA Tesla P100 进行。对于所有随机初始化的层,学习率 $\lambda = 0.01$,对于使用预训练模型初始化的层,学习率 $\lambda = 0.0001$,每 2 万次迭代将其缩小为原来的 10%。批处理的总迭代次数设置为 60000,批处理大小为 1。权重衰减 η 和动量因子 m 分别设置为 0.0005 和 0.9。

3.4 实验分析

为了验证本文提出的 VAF-Net 网络在建筑物目标检测任务上的有效性,基于 Massachusetts 建筑物数据集所提供的图像以及类别标签进行了实验,结果如图 4 所示。

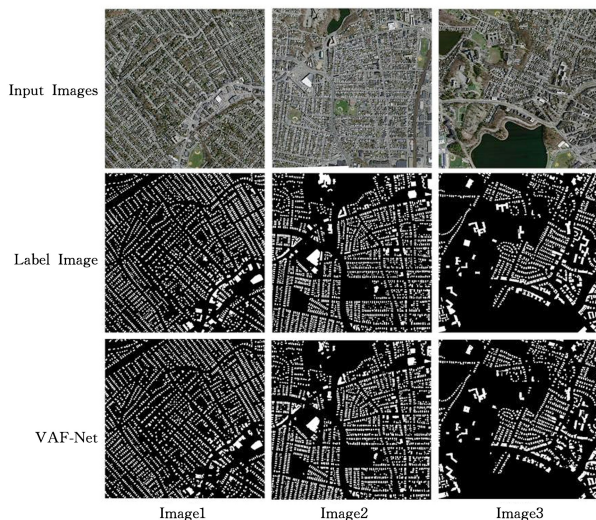


图 4 实验结果

Fig. 4 Experimental results

建筑物,检测结果中噪声点较少,对于颜色、形状、阴影等差异因素带来的不良影响具有很好的容错能力。同时可以看出,建筑物的轮廓清晰,边缘缺失或失真的现象也得到了很好的改善。此外,为了测试该模型的检测速度,在 Tesla P100, GeForce GTX 1080, Sugon GPU 3 种主流的深度学习显卡上进行了实验,结果如表 1 所列。可以看出,在计算资源充足的情况下,本文方法能够达到较好的检测速度。

表 1 在不同设备上的检测时间

Table 1 Test time on different devices

设备名称	Telsa P100	GeForce GTX 1080	Sugon GPU
整张影像处理时间 (1280×1280)/s	20.494	25.018	32.687
单张影像处理时间 (256×256)/s	0.819	1.000	1.307

将 VAF-Net 与其他几种具有代表性的方法进行对比,对比方法有 FCN, VGG, U-Net。实验结果如表 2 所列。FCN 网络是像素级别的目标检测网络,在进行建筑物目标检测过程中,忽略了像素与像素之间的关系,因此在面临背景复杂、建筑物密集的检测任务时,效果欠佳。VGG 网络作为 CNN 网络的一种,其卷积核大小单一,无法应对遥感影像中建筑物的大小各不相同的情况。U-Net 网络是针对医学影像分类问题所提出的,在医学影像的分类分割检测任务中有着出色的表现。然而医学影像的图像结构较为单一,结构比较固定,因此检测难度较低,而遥感图像中的建筑物大小不一、形状各不相同,因此在进行遥感图像目标检测时,其检测效果并不令人满意,并且常常因 U-Net 网络层数不足,造成特征表达不充分的现象。本文提出的 VAF-Net 网络,相较于这几种经典网络,在 Precision, Recall, F1-score 3 个指标上都有不同程度的提高,这说明引入自适应加权特征融合的方式,可以有效提高建筑物目标检测的精确度。此外,将基础网络 UV-Net 与 U-Net 进行对比,其 F1-score 提升了 6%,表明了改进网络 UV-Net 在目标检测任务上的有效性。同时,从整张影像的测试时间来看,FCN 网络由于网络层数较浅,网络参数量少,因此测试用时最少。本文提出的 VAF-Net,相比于基础网络 U-Net 以及改进网络 UV-Net,网络参数有一定的增多,因此其测试时间稍有延长,但相比 VGG 网络,其测试时间较短,且 F1-score 也有一定的提升。因此,在进行背景复杂的遥感图像建筑物目标检测时,网络中的单一特征的代表能力有限,适当地引入边缘特征,可以在检测时间较短的情况下,很好地提高建筑物目标检测的精确度。

表 2 不同方法对建筑物提取精度的对比

Table 2 Comparison of building extraction accuracy with different methods

实验名称	Precision	Recall	F1-score	整张影像测试时间/s
U-Net	0.638	0.872	0.734	19.043
FCN	0.611	0.667	0.632	12.195
VGG	0.804	0.801	0.802	25.364
UV-Net	0.808	0.788	0.798	19.578
VAF-Net	0.821	0.825	0.823	20.494

图 5 给出了不同网络检测到的建筑物图像。图 5(a)为

可以看出,VAF-Net 可以较为完整地检测到图像中的建

FCN 网络所检测到的建筑物,从图中可以看出,检测结果中建筑物边界失真严重且噪声较多。相比于 FCN, U-Net 网络和 VGG 网络的检测结果较为干净,一定程度上减少了检测结果中的噪声,但其建筑物边缘缺失的情况并没有得到很好的改善。图 5(d)为本文提出的 VAF-Net 所检测到的建筑物,相比于 FCN, VGG, U-Net, 可以明显看出,建筑物的边缘较为完整,缺失和模糊的现象得到了很好的改善。

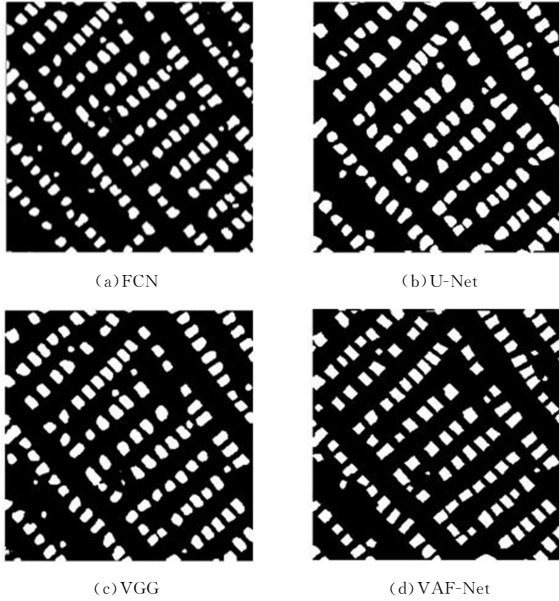


图 5 不同网络输出结果的对比

Fig. 5 Comparison of output results of different networks

为进一步验证本文方法的有效性,在基础网络 UV-Net 的基础上,将本文方法与目前流行的其他建筑物检测方法进行了对比,实验结果如表 3 所列。

表 3 不同融合方式对实验结果的影响

Table 3 Influence of different fusion methods on experimental results

实验名称	Precision	Recall	F1-score	整张影像 测试时间/s
实验 1 UV-Net	0.808	0.788	0.798	19.578
实验 2 ^[20] U-Net+CRF	0.780	0.857	0.790	26.604
实验 3 ^[21] U-Net+ 空洞卷积	0.602	0.798	0.682	24.168
实验 4 ^[9] 多源数据 融合+U-Net	0.756	0.781	0.767	20.481
实验 5 VAF-Net	0.821	0.825	0.823	20.494

实验 2 结合 U-Net 以及条件随机场 CRF 进行网络的设计^[20],其检测效果与本文所提的基础检测网络 UV-Net 相近,但是相比于 VAF-Net 仍有一定的差距。此外,由于 CRF 的引入,极大地增大了网络的参数量,因此其检测时间相比于 VAF-Net 较慢。实验 3 利用基于空洞卷积的级联来进行建筑物的检测^[21],然而遥感影像中的建筑物尺寸各异,同一张影像中通常同时包含较大尺寸的建筑物和众多小尺寸建筑物,空洞卷积在面对这样的情况时其检测效果和检测时间都不理想。实验 4 中,利用多源数据融合的方法进行目标检测^[9],由于在数据层面的融合可能会引入更多的噪声,因此其在建筑物目标检测任务中表现一般。本文提出的 VAF-Net

是在 U-Net 的基础上进行改进的,考虑到了不同特征对检测效果的重要程度,在建筑物目标检测任务中检测效果较好。图 6 给出了这 5 组实验的检测结果,从图中可以明显看出,本文方法不仅可以有效检测到影像中的目标建筑物,并且很好地解决了建筑物边缘失真的情况。

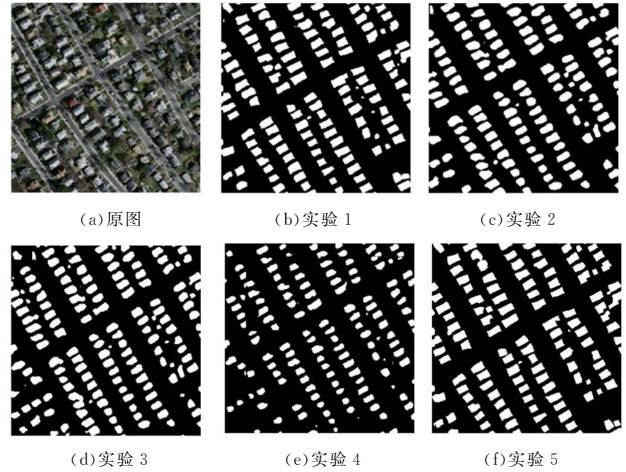


图 6 不同方法的实验对比图

Fig. 6 Experimental comparison diagram of different methods

为了探讨精度提升和测试用时的关系,以 U-Net 模型为基准,从 F1-score 和整张影像的检测用时方面进行了分析,结果如表 4 所列。可以看出,FCN 相比于其他模型的时间消耗最少,但其忽略了上下文的语义表征,故检测效果差强人意。VGG 模型在建筑物检测任务中的表现较好,但由于其网络层数过深,模型较大,因此检测速度相比于 U-Net 较为缓慢。此外,结合 CRF 进行建筑物检测,其 F1-score 值相比 U-Net 提高了 7.7%,但由于模型的参数量增加,其整张影像的检测时间相比于 U-Net 增加了将近 40%,时间成本消耗过大。相比 U-Net,本文方法(VAF-Net)在进行建筑物检测任务时,整张影像检测时间增加了 7.6%,但其 F1-score 值相比 U-Net 提高了 12.1%,在控制影像的检测时间损耗不大的情况下,提高了建筑物的检测效果,有效地改善了建筑物边缘失真的问题。

表 4 不同模型 F1-score 和检测时间的对比

Table 4 Comparison of F1-score and detection time of different models

模型名称	F1-score	整张影像 测试时间/s
U-Net	0.734	19.043
FCN	0.632 (-13.8%)	12.195 (-35.9%)
VGG	0.802 (+9.2%)	25.364 (+33.2%)
UV-Net	0.798 (+8.7%)	19.578 (+2.8%)
U-Net+CRF	0.790 (+7.7%)	26.604 (+39.7%)
多源融合+U-Net	0.767 (+4.5%)	20.481 (+7.5%)
U-Net+空洞卷积	0.682 (-7.1%)	24.168 (+26.9%)
VAF-Net	0.823 (+12.1%)	20.494 (+7.6%)

结束语 针对高分辨率影像中的目标检测任务中经常出现的检测结果边缘模糊、缺失的现象,本文进行了以下几个方面的工作:1)基于编码解码网络 U-Net 进行了目标检测基础网络 UV-Net 的构建,通过增加网络层数以及“卷积-池化”结构块适当地增加网络参数,解决了 U-Net 网络因参数过少引起的过拟合问题;2)构建了融合边缘特征的建筑物目标检测

网络 VF-Net,基于 Sobel 检测算子边缘检测的结果,在 UV-Net 网络的基础上,设计了并行的边缘特征检测子网,并在相应的层次融合两种不同的特征图,以此来丰富特征图中的边缘信息;3)针对特征融合部分设计了自适应加权特征模块,借助反向传播,自适应地更新融合权重,从而更好地利用不同的特征对网络的贡献。经过实验验证,该方法在 Precision, Recall 和 F1-score 3 个评价指标上都达到了较高的水平,最终的结果中检测出的建筑物较为完整,边缘模糊、缺失的现象有了很大的改善。然而,该方法针对较大型建筑物的检测效果欠佳,在后续的工作中,将会基于该问题进行研究,并尝试进行解决。

参 考 文 献

- [1] YANG Z, MU X D, WANG S Y. Scene classification of remote sensing images based on multiscale features fusion[J]. Optics and Precision Engineering, 2018, 26(12): 3099-3107.
- [2] LONG J, SHELHAMER E, DARRELL T. Fully Convolutional Networks for Semantic Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(4): 640-651.
- [3] XU Z J, YANG X B, HE L M, et al. Multiscale remote sensing semantic segmentation Network[J]. Computer Engineering and Applications, 2020, 56(21): 210-217.
- [4] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional Networks for Biomedical Image Segmentation[J]. arXiv: 1505. 04597, 2015.
- [5] LIU H, LUO J C, HUANG B, et al. DE-Net: Deep Encoding Network for Building Extraction from High-Resolution Remote Sensing Imagery[J]. Remote Sensing, 2019, 11(20): 2380.
- [6] XU Z H, LIU Y, QUAN J C, et al. Buildings egmentation of remote sensing images based on VGG16 pre-encoding[J]. Science Technology and Engineering, 2019, 19(17): 250-255.
- [7] REN X L, WANG Y P, YANG J Y. Building Detection form Remote Sensing Images Based on Improved U-net[J]. Laser and Optoelectronics Progress, 2019, 657(22): 195-202.
- [8] LI Z, ZHOU F. FSSD: Feature Fusion Single Shot Multibox Detector[J]. arXiv: 1712. 00960, 2018.
- [9] SHENG Y T, ZHAO Z, WANG T T. Building Areas Extraction in GF-3 Images Based on the Integration of Span Image and Texture Features[J]. Beijing Surveying and Mapping, 2020, 34(1): 73-78.
- [10] FENG F J, LI J P, DING Y Z. Target Detection from High Resolution Remote Sensing Images Based on Combination of Multi-scale Visual Features[J]. Journal of Applied Sciences, 2018, 36(3): 471-484.
- [11] LIU S, HUANG D, WANG Y. Learning Spatial Fusion for Single-Shot Object Detection[J]. arXiv: 1911. 09516, 2019.
- [12] FENG F, WANG S T. Building Extraction Based on Multi-input-multi-output and Multi-feature Fusion [J]. Journal of Zhengzhou Institute of Surveying and Mapping, 2020, 37(6): 575-580.
- [13] WANG Z H, LIU H Q. Building Recognition Based on Transfer Learning and Adaptive Feature Fusion[J]. Computer Technology and Development, 2019(12): 40-43.
- [14] LI X Y. Object Detection in Remote Sensing Images Based on Deep Learning[D]. Hefei: University of Science and Technology of China, 2019.
- [15] FENG J W, ZHANG L M, DENG X Y. Image segmentation based on multi-source fusion FCN[J]. Application Research of Computers, 2018, 35(9): 2877-2880.
- [16] ZHU G Y. Research on Building Extraction from Remote Sensing Images Based on Deep Learning[D]. Hangzhou: Zhejiang University, 2019.
- [17] JIN F, WANG L F, LIU Z, et al. Double U-Net Remote Sensing Image Road Extraction Method [J]. Journal of Geomatics Science and Technology, 2019, 36(4): 377-381, 387.
- [18] LIU R Y, SUN Q C, WANG C Y. Research on Edge Detection Algorithm in Digital Image[M]. Science Press, 2015.
- [19] PANG Y W, XIU Y X. Lane Semantic Segmentation Neural Network Based on Edge Feature Merging and Skip Connections [J]. Journal of Tianjin University (Science and Technology), 2019, 52(8): 779-787.
- [20] ZHANG H, ZHAO J H, ZHANG X G. High-resolution Image Building Extraction Using U-net Neural Network[J]. Remote Sensing Information, 2020, 35(3): 143-150.
- [21] XU Z H, LIU Y, QUAN J C, et al. Buildings Segmentation of Remote Sensing Images Based on U-Net Pre-encoding[J]. Science Technology and Engineering, 2019, 19(17): 250-255.



HE Xiao-hui, born in 1978, professor. Her main research interests include artificial intelligence, computer vision, remote sensing image processing, and data mining.