

基于深度学习的自然灾害遥感影像语义分割^①



王关茗, 胡乃平

(青岛科技大学 信息科学技术学院, 青岛 266061)

通信作者: 王关茗, E-mail: 348102401@qq.com

摘要: 自然灾害种类繁多, 通过遥感影像语义分割相对比较困难. 为了能够更好地实现遥感影像分割, 本文提出一种基于生成对抗网络的 3 层遥感影像语义分割模型, 针对不同场景的解析, 基于全卷积神经网络 FCN, 设计一种多层次的遥感语义分割框架. 有效对遥感图像语义分割进行处理, 从而提高了模型的分割精度. 实验表明利用这种模型是有效的, 特别是受损建筑的分割结果, $mIoU$ 为 82.28%, 通过该模型与其他网络模型进行对比, 其性能评价指标明显优于其他网络模型. 最后, 通过对自然灾害各种场景影像进行分析, 为应急管理部门提供一份可靠的数据报告.

关键词: 自然灾害; 遥感影像; 深度学习; 语义分割

引用格式: 王关茗, 胡乃平. 基于深度学习的自然灾害遥感影像语义分割. 计算机系统应用, 2023, 32(2): 322-328. <http://www.c-s-a.org.cn/1003-3254/8994.html>

Semantic Segmentation of Natural Disaster Remote Sensing Image Based on Deep Learning

WANG Guan-Ming, HU Nai-Ping

(College of Information Science and Technology, Qingdao University of Science and Technology, Qingdao 266061, China)

Abstract: There are many kinds of natural disasters, and it is relatively difficult to semantically segment remote sensing images. In order to better realize remote sensing image segmentation, this study proposes a three-layer semantic segmentation model for remote sensing images based on a generative adversarial network. For the analysis of different scenes, a multi-level remote-sensing semantic segmentation framework is designed based on a fully convolutional network (FCN). The semantic segmentation of remote sensing images is effectively performed, and thus the segmentation accuracy of the model is enhanced. Experiments show that this model is effective, which can be directly observed from the segmentation results of damaged buildings, with $mIoU$ being 82.28%. In addition, this model is compared with other network models, and its performance evaluation index is significantly better than that of other network models. Finally, a reliable data report is provided to emergency management departments by analyzing various scene images of natural disasters.

Key words: natural disaster; remote sensing image; deep learning; semantic segmentation

近年来, 地震、洪灾、火灾等自然灾害频繁发生, 给国家和人民带来难以磨灭的伤害. 如何挽救宝贵生命和减少经济损失是我们面临的重大课题, 其中最重要的步骤是准确评估这些事件造成的损失^[1]. 对损失的正确估计可以有效地掌握灾情, 并使救援小组能够合

理地安排紧急援助^[2]. 当前随着社会的高速发展, 越来越多的高精尖技术应用在防范自然灾害领域, 其中, 计算机视觉是该领域的核心研究课题, 计算机视觉主要分为图像分类、目标检测和语义分割, 本文主要采用图像语义分割对自然灾害场景分类进行研究^[3]. 图像的

① 收稿时间: 2022-07-12; 修改时间: 2022-09-07, 2022-09-15; 采用时间: 2022-09-27; csa 在线出版时间: 2022-12-02

CNKI 网络首发时间: 2022-12-05

语义分割由 Shelhamer 等人^[4]首次提出,其主要的内容是对分割后的图像加上语义特征分类,就是给分割后图像中的相同类的物体加上标签,输入一般是彩色深度图像.与传统的图像分割相比,图像语义分割为图像中的目标添加了一定的语义信息,可以根据图像本身的场景、纹理等语义特征识别图像本身所表达的信息.语义分割是将图像按照不同属性划分为不同的区域,用相同的颜色代表同一个区域;图像分割后不同区域有明显的界线.随着计算机视觉的快速发展,尤其是以全卷积神经网络 (FCN) 为基础框架的图像语义分割等深度学习的技术.鉴于目前国内还没有全面论述基于深度学习的自然灾害遥感图像语义分割的综述文献,该论文主要基于遥感图像,应用于防范自然灾害,提高快速救援自然灾害的能力,为科学救灾提供可靠的依据.针对上述问题,本文提出一种基于生成对抗网络的3层遥感影像语义分割模型,针对不同场景的解析,基于全卷积神经网络 FCN,设计一种多层次的遥感语义分割框架,有效对遥感图像语义分割进行处理,从而提高了模型的分割精度.针对当前自然灾害的复杂性,本文主要采用公开的 xBD 数据集,该数据集自然灾害种类多、规模大、分辨率高.

1 相关工作

本文在研究了基于深度学习的图像语义分割基础上,结合遥感影像中地物目标具有多尺度的特性,探索

高分辨率遥感图像的多尺度语义分割,具有重要的研究和应用价值.自然灾害数据集最初可分为两类:成像数据集和非成像数据集.根据图像捕获位置,现有的成像自然灾害数据集可进一步分为3类:地面图像、卫星图像和航空图像.最近,研究人员介绍了几种用于自然灾害损失评估的数据集.Nguyen 等人^[5]提出了 AIDR 系统的扩展,主要是从中国的社交媒体收集数据中提出了 AIST 建筑物变化检测 (ABCD) 数据集,其中包括海啸后的航空图像,以确定建筑物是否已被冲毁.Chen 等人^[6]在中提出了一种融合不同数据资源的方法,用于飓风后建筑物损伤的自动检测.Christie 等人^[7]介绍了一组名为“世界功能地图”的建筑物和土地图像.Rudner 等人^[8]提出了一种从 Sentinel-1 和 Sentinel-2 卫星收集的卫星图像,用于洪水建筑物的语义分割.Gupta 等人^[9]提出了 XBD,其中包括事件前后的卫星图像,以评估建筑物损坏.Zhu et 等人^[10]提出了一个名为 MSNet 的空中视频实例分段,以评估建筑物损坏.针对以上研究现状,结合自然灾害的多变性和复杂性,本文提出一种基于生成对抗网络的3层遥感影像语义分割模型,通过改进 PSPNet 网络模型与 PSPNet^[11]、DeepLab^[12] 等网络进行对比,在多层网络模型使用空洞卷积^[13]扩大感受野,通常采用编码-解码模型,该模型广泛应用于计算机视觉领域,其中包括位置评估^[14]、目标检测^[15]和语义分割^[16].本文的框架如图1所示.

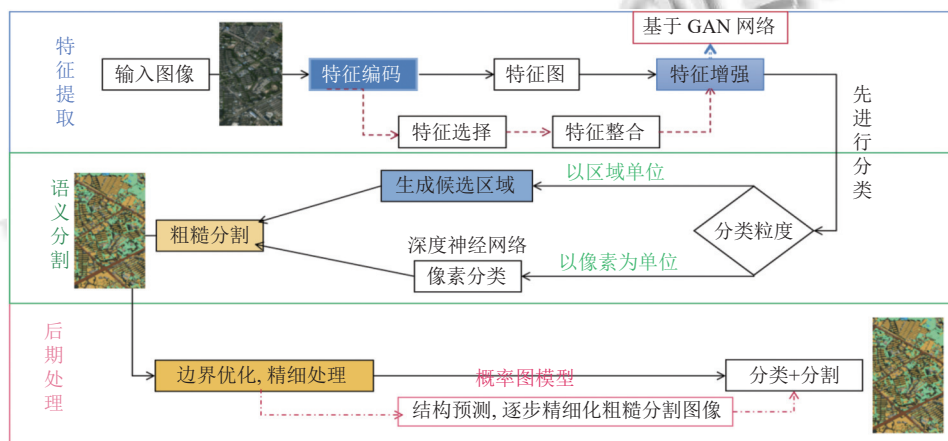


图1 遥感影像语义分割框架

2 融合特征增强与通道域注意力分割模型

本文通过对 PSPNet 网络模型增加通道域注意力模块,将输入影像中的关键特征提取并标注,经过不断

的学习和训练,让网络模型更加准确获得影像的特征信息^[17].同时,在每个通道设置不同的权重,并不断增强特征,进而生成通道域注意力模型.xBD 数据集标签

和灾害类型分布高度不均衡.此外,区分不同程度损伤的视觉线索可能非常小.不同类型灾害的负面图像的不同可用性也可能是定位模型的一个障碍. ResNet50 是在 ImageNet^[18] 上预先训练的,而较小的边缘网络是用随机权重初始化的.使用 ReLU^[19] 激活所有卷积层.首先输入高精度遥感影像,使用数据增强对抗神经网络模型,增加像雾霾、多云、雷雨等实际场景,更好的提取特征信息;然后不同尺寸进行池化操作得到多尺

度特征图,再经过通道注意力机制卷积神经网络来减少通道数量,最后采用强化模块的上采样方法得到相同尺寸的特征图,并在通道上拼接起来,它包含了不同尺度之间的变化信息.由于xBD数据集自然灾害种类较多,通过先分类再分割的方法,将相同灾害类型放在同一个模块,这样更好的减少卷积网络的任务,能够实现更高的预测结果和更多的应用场景,网络模型如图2所示.

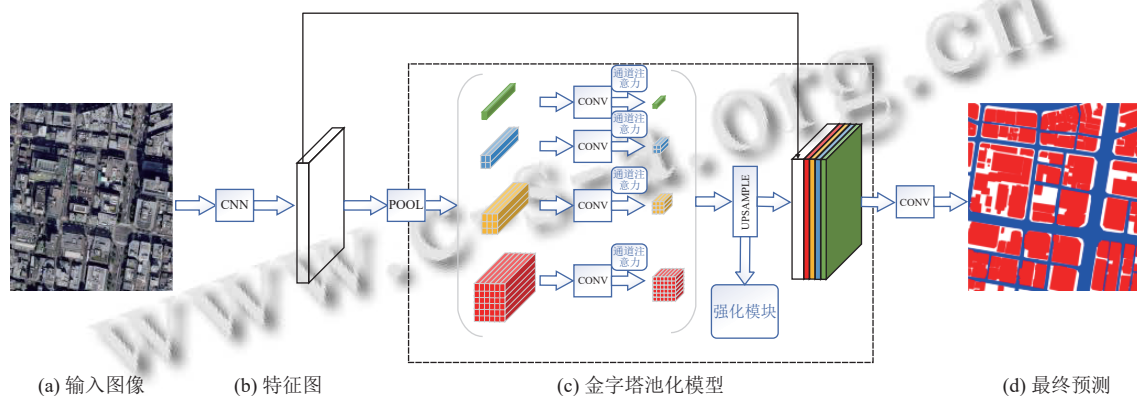


图2 改进 PSPNet 网络结构

2.1 全卷积神经网络 (FCN)

目前,大多数语义分割方法依赖于深度卷积神经网络 (CNN).但是 CNN 对局部接收的卷积会影响语境空间关系的建模.通过基于全卷积神经网络 (FCN) 的图像语义分割方法,改善了 CNN 对卷积的影响.目前,大多数较好的语义切分网络都是基于 FCN 进行改进的.图像的语义分割是通过图像中的每个像素进行分类,主要是为了解决语义图像分割的问题,FCN 能够输入不同大小的图像.通过反卷积方法对最后卷积的特征图进行上取样,得到各像素的预测,同时保持其原始图像的空间信息.然后将上取样的特征图像素进行归类. FCN 网络主要通过卷积神经网络对图像的强大学习能力,采用全卷积的方法,利用卷积层代替现有的语义分割深度网络模型.因为卷积层后面有一个连接池层,而连接池层又叫下采样层,所以会影响图片的分辨率. FCN 采用反卷积方法实现上采样,以确保图像的清晰度. FCN 和 CNN 区别就是将 CNN 的全连接层替换为全卷积,并将其作为一个标志. FCN 作为图像语义分割具有里程碑的意义,是后来很多图像语义分割网络模型的研究的理论基础. FCN 能接受任意大小图片输入、运算高效等特点. Max-Pool、Convolution 和 Dense

的输出被输入通道和空间关系模块,用于生成关系增强特征.这些特征随后通过 1×1 滤波器反射各自的卷积层,以将通道数压缩为类别数.最后,将卷积特征映射放大到所需的全分辨率,并按元素添加,以生成最终的分割映射.

2.2 空洞卷积

首先,空洞卷积提出的是一种用于处理图像的方法,常用的图像分割方法一般采用池化或卷积来增大感知范围,并将其特征图尺寸减小,最后采用上采样进行恢复;由于特征图的缩小和放大处理会导致测量的准确性下降,所以必须采用一种既能提高感知范围又能保证特征图大小不改变的运算,取代下、上采样,既能保证图像的几何特性,又能保证图像的信息不会丢失.在网络层要求较大的感知范围时,使用空洞卷积效果最好. r 设置为“膨胀率”,即孔卷积的步长,代表输入信号的采样间隔.当 $r = 1$ 时,空卷积是标准卷积.在孔卷积中,卷积核以“膨胀率”进行扩展,在空间维度在相邻的权系数之间插入 $r-1$ 个零,创建稀疏滤波器.空洞卷积的感受野的大小可以用式 (1) 表示:

$$rf((k+1)(r-1)+k)^2 \quad (1)$$

如图 3 所示, 通过扩大滤波器, 网络扩大了其感受野 (因为权重将以更稀疏的形状排列), 但保留了分辨率, 并在数据中执行下采样. 因此, 支持感受野的扩展,

而不增加每层可训练参数的数量, 这减少了计算负担; 保留了特征图分辨率, 这可能有助于网络从数据中提取更多有用的信息, 主要是小对象.

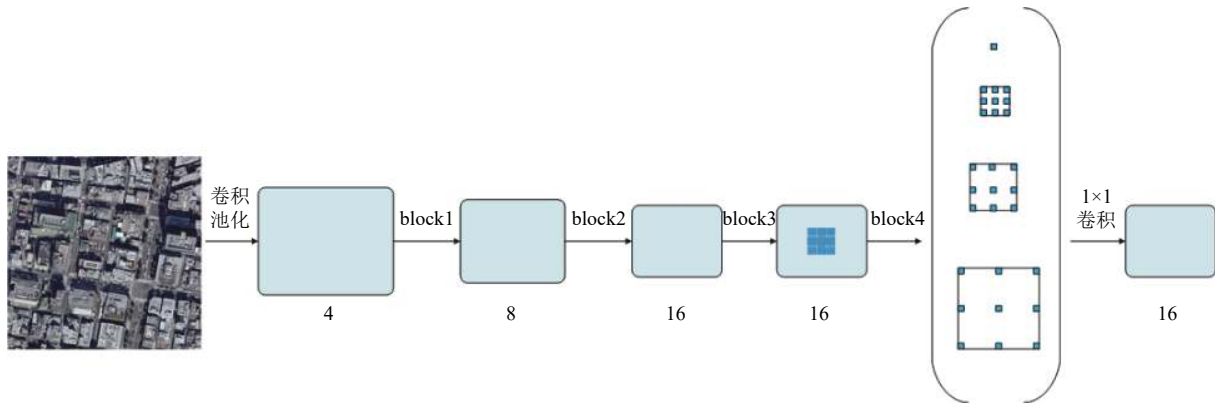


图 3 空洞卷积

2.3 数据增强对抗神经网络

在改进的 PSPNet 中, 为了实现数据增强现实场景, 通过上采用结合数据增强网络结构 (DAGAN) 模型, DAGAN 如图 4 所示. 生成网络是一个编码器集, 用于采集 c 级的输入影像并将其投射到一个低维矢量上. 将所述随机矢量 (z_i) 变换并且与所述映射矢量相关联; 所有的信息都传送给解码器网络, 并由解码器产生一个加强的图像.

同一类别产生的其他实际图片) 和伪分布 (由发生器网络产生的图片). 对抗培训通过从老图片中产生新图片 (无论哪个类别), 看上去都属于相同类别, 但看上去完全不一样. 因为判别网络有 $2k$ 个输出, 判别网络主要有两部分组成, 即实际数据和生成数据. 几种 GAN 网络的对比如表 1 所示.

$$(L_c)' = L_{data} + L_{gen} \tag{3}$$

$$L_{data} = -E_{x,y \sim P_{data}(x,y)} \log[p(y|x, y < k + 1) + p(y + k|x, y < k + 1)] \tag{4}$$

$$L_{gen} = -E_{x,y \sim p_g(x,y)} \log[p(y|x, y < y < 2k + 1) + p(y - k|x, k < y < 2k + 1)] \tag{5}$$

表 1 几种 GAN 网络的对比

模型	类别数量	训练数据
GAN	2	unlabeled
Conditional GAN	2	labeled
Semi-supervised	$k+1$	labeled+unlabeled
DAGAN	$2k$	labeled

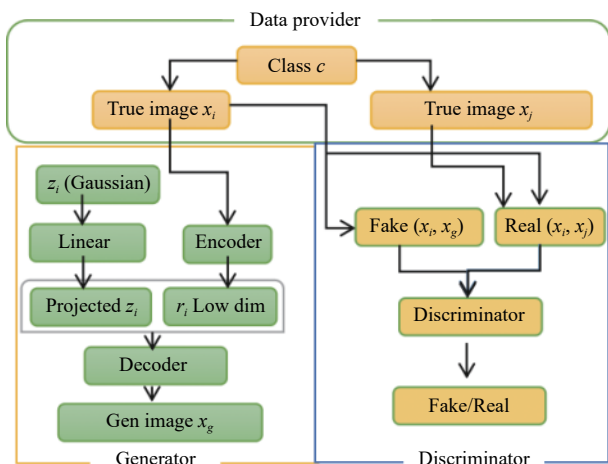


图 4 DAGAN 网络结构

对于判别网络, 其损失函数如下:

$$L_c = -E_{x,y \sim P_{data}(x,y)} \log[p(y|x, y < k + 1)] - E_{x,y \sim P_g(x,y)} \log[p(y|x, y < y < 2k + 1)] \tag{2}$$

对抗识别网络进行培训以识别从实际的分布 (从

2.4 全连接条件随机场

本文采用先分类再分割的方法进行遥感影像语义分割, 在图像的语义划分中, 首先要对图像进行分级, 然后进行图像的图像分割, 而对图像进行的非线性卷积上抽样会造成图像的丢失, 从而造成图像的分割边缘不清; 为改善边缘划分的最佳化, 获得更为准确的分割效果, 本文提出采用概率图法进行求解. 利用 RGB 彩色特性与空间方位之间的相关性, 对神经网络的后

验分配进行了预报,以改善分割准确率.条件随机场将单个像素的一元势能与邻近像素的二元势能相结合,实现了在不同的位置上相邻像素都是同一标记,因此对目标的边界进行了光滑的预测.而在遥感图像中,由于采用了网络结构,其边界仍存在着一定的模糊性,因此,必须对图像的具体细节进行更准确的处理.将基于图像的完全连通状态随机场与深度学习方法进行图像的语义划分,利用图像中各像素的预测和真实的空间间距来判定各像素的相似度,进而形成点对势能,达到最大限度地分割图像的边界.全连接条件随机场结合了整体的数据,并结合了图像级别的特点,能够较好地实现高分辨率的遥感影像边缘的分割,实验进行对比如表2所示.

表2 实验 CRF 和未使用 CRF

算法	Kappa	正确率 (%)	漏检率 (%)
未使用CRF	0.823	91.34	8.66
使用CRF	0.834	93.04	6.96

3 实验分析

3.1 数据集

本文使用 xBD 数据集,其中标记 19 种自然灾害的 22 068 张图像.该数据集包含 22 068 张图像,标记有 19 种不同的事件,包括地震、洪水、野火、火山爆发和车祸等.这些图像包括了灾前、灾后图像,图像可用于构建定位和损伤评估这两项任务.自然灾害救援行动需要准确评估建筑物的损坏情况,以便有效分配援助、协助和资源.目前,这是一项危险且耗时的人工任务,必须由人类应答者来执行.通过认识到病变中有明显的视觉模式来自动完成这项任务,xBD 提供了大量的图像和注释多边形,以及注释比例.这将使创建一个可视化模型成为可能,该模型可以自动评估建筑物的损害并远程执行.如图5所示,xBD 包含了世界不同可视区域的各种类型的灾害.这满足了许多救灾机构提出的许多实际需求.

3.2 数据集评审指标及模型参数

本文评价指标采用的是平均交并比 (mIoU) 交并比是预测值和真实值的交集和并集的比.其计算公式为:

$$mIoU = \frac{1}{k} \sum_{i=1}^k \frac{P \cap G}{P \cup G} \quad (6)$$

其中, P 表示预测值, G 表示真实值, k 表示总类型的个数,公式表示每个类型的预测值和真实值的交集除以并集,然后取平均.

3.3 实验结果分析

如图6所示,在 xDB 数据集上,当训练轮数达到 50 epoch 时,模型平均损失稳定在 0.1 附近,之后平均损失函数值基本无降低,表明训练过程收敛,模型平均损失在 0.1 附近,训练收敛.实验环节是在 PyTorch 中实现本文的网络结构,实验室的 GPU 型号为 GTX1080Ti.实验学习率为 0.1, epoch 设置为 50, β_1 为 0.9, β_2 为 0.999 的 Adam 优化器^[20]用于训练.同时,采用了加权 F1 分数是评价数据集,加权 F1 平衡了查准率和查全率,非常适合 xBD 数据集.准确度本身就是一个有缺陷的指标,因为预测所有图像都是“无损”的,分类器可以保持 75% 准确度.对于基线模型,本文加权 F1 总分是 0.265 4.



图5 xBD 数据灾害类型

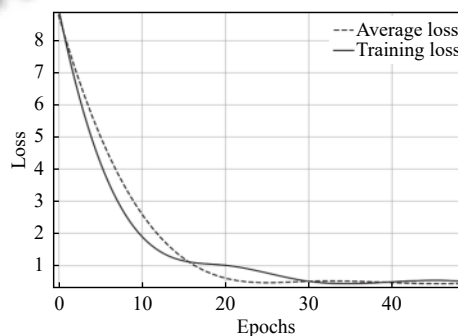


图6 训练损失折线图

3.4 对比验证

为了只检测建筑物和道路,我们将所有类型的建筑物合并到一个单一的建筑类别中,根据 4 个损坏级别分别进行注释.所有网络模型在建筑物和道路的分

段上都表现出显著的性能.这4种方法都能很好地分割建筑物和道路.与其他两个实验相比,在本实验中,ENet^[21]在DeepLabV3+^[22]中表现出更好的性能.但PSPNet的性能最好,分割结果近乎完美,*mIoU*为96.11.从实验结果可以明显看出,与基于编码器-解码器的方法ENet甚至DeepLabV3+相比,作为全局上下文的金字塔池模型能够更精确地分割建筑物和道路.尽管DeepLabV3+在分割模型的设计中也实现了全局上下文优先,但与DeepLabV3+中使用的空洞卷积相比,改进PSPNet中的金字塔池显示出了优越的性能.

本文分析了4种不同类型建筑物的分割性能:建筑物未受损、建筑物中度受损、建筑物严重受损和建筑物完全受损.网络模型显示,尽管完全损坏和未损坏的建筑物可以以良好的准确度被分割,但是很难区分中度损坏的建筑物和严重损坏的建筑物.改进PSPNet在所有3种模型中表现最好,它们根据各自的破坏程度对不同类型的建筑物进行分类.特别是未受损建筑和完全受损建筑的分割结果最为完美,*mIoU*分别为

95.24和96.34,如表3所示.这4种性能的一个共同点是,这两种网络在检测中度和重度受损建筑物时性能最差.从这两个模型产生的结果来看,改进PSPNet在xBD数据集上的表现显然最好.

表3 xBD测试数据集的结果

算法	无破坏	中等破坏	重大破坏	全部破坏	<i>mIoU</i>
ENet	62.47	42.10	34.50	88.96	57.01
DeepLabV3+	71.91	61.12	58.39	60.63	62.98
PSPNet	91.34	93.23	89.45	92.34	91.59
改进PSPNet	95.24	96.34	96.28	96.55	96.11

在图7曲线图中,PSPNet比DeepLabV3+更好地学习数据集.性能随着PSPNet和DeepLabV3+的基本学习率的变化而显著变化.总体而言,PSPNet在xBD数据集的动态特性方面显示了最佳结果.PSPNet中的金字塔池模块能够,比DeepLabV3+中的空洞卷积更好地收集所有类的全局上下文.PSPNet比ENet的优越性能证明,基于金字塔池的分割模型在xBD数据集的语义分割上优于基于编码器-解码器的模型.

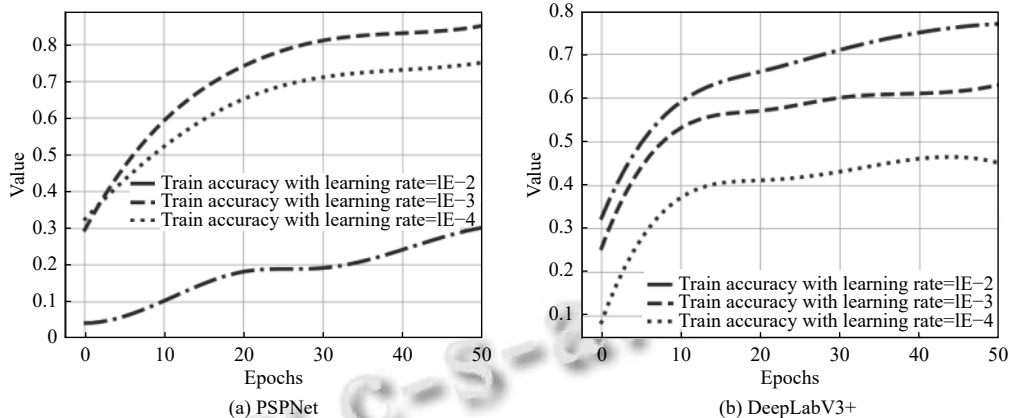


图7 PSPNet和DeepLabV3+实验结果

4 结论与展望

本文提出一个基于深度学习的自然灾害遥感语义分割,通过优化PSPNet语义分割网络结构模型,并将DAGAN网络训练,对图像进行数据增强.通过金字塔池生成的不同层次的特征图最终被展平并连接到一个完全连接的层中进行分类.自然灾害救援行动需要准确评估建筑物受损情况,以便有效分配援助、援助和资源.xBD包含世界上不同视觉区域的各种灾难类型.这满足了多个灾难响应机构规定的许多现实需求.我们讨论了在该数据集上进行语义分割所面临的挑战,

以及所提出的模型在该数据集上的优异性能.我们希望这篇论文将有助于未来的自然灾害损失评估研究,并有助于通过有效的自然灾害管理减少人员和经济损失.

参考文献

- 史培军,王季薇,张钢锋,等.透视中国自然灾害区域分异规律与区划研究.地理研究,2017,36(8):1401-1414.
- 徐鹏杰,邓磊.遥感技术在减灾救灾中的应用.遥感技术与应用,2011,26(4):512-519.[doi:10.11873/j.issn.1004-0323.2011.4.512]

- 3 Chowdhury T, Rahnemounfar M. Attention based semantic segmentation on UAV dataset for natural disaster damage assessment. Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Brussels: IEEE, 2021. 2325–2328.
- 4 Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640–651. [doi: 10.1109/TPAMI.2016.2572683]
- 5 Nguyen DT, Alam F, Ofli F, *et al.* Automatic image filtering on social networks using deep learning and perceptual hashing during crises. Proceedings of the 14th International Conference on Information Systems for Crisis Response and Management. Albi: ISCRAM Association, 2017.
- 6 Chen SA, Escay A, Haberland C, *et al.* Benchmark dataset for automatic damaged building detection from post-hurricane remotely sensed imagery. arXiv:1812.05581, 2018.
- 7 Christie G, Fendley N, Wilson J, *et al.* Functional map of the world. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 6172–6180.
- 8 Rudner TGJ, Rußwurm M, Fil J, *et al.* Multi3Net: Segmenting flooded buildings via fusion of multiresolution, multisensor, and multitemporal satellite imagery. Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Honolulu: AAAI Press, 2019. 702–709.
- 9 Gupta R, Hosfelt R, Sajeev S, *et al.* Creating xBD: A dataset for assessing building damage from satellite imagery. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Long Beach: IEEE, 2019. 10–17.
- 10 Zhu XY, Liang JW, Hauptmann A. MSNet: A multilevel instance segmentation network for natural disaster damage assessment in aerial videos. Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision (WACV). Waikoloa: IEEE, 2020. 2022–2031.
- 11 Zhao HS, Shi JP, Qi XJ, *et al.* Pyramid scene parsing network. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017. 6230–6239.
- 12 Chen LC, Papandreou G, Kokkinos I, *et al.* Semantic image segmentation with deep convolutional nets and fully connected CRFs. Proceedings of the 3rd International Conference on Learning Representations. San Diego: ICLR, 2015.
- 13 Liu W, Rabinovich A, Berg AC. ParseNet: Looking wider to see better. arXiv:1506.04579, 2015.
- 14 Bao WX, Yang YP, Liang D, *et al.* Multi-residual module stacked hourglass networks for human pose estimation. Journal of Beijing Institute of Technology, 2020, 29(1): 110–119.
- 15 Chen LC, Papandreou G, Schroff F, *et al.* Rethinking atrous convolution for semantic image segmentation. arXiv:1706.05587, 2017.
- 16 Lin TY, Dollár P, Girshick R, *et al.* Feature pyramid networks for object detection. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017. 936–944.
- 17 Jia D, Wei D, Socher R, *et al.* ImageNet: A large-scale hierarchical image database. Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami: IEEE, 2009. 248–255.
- 18 Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe: Curran Associates Inc., 2012. 1097–1105.
- 19 Chen MZ, Challita U, Saad W, *et al.* Artificial neural networks-based machine learning for wireless networks: A tutorial. IEEE Communications Surveys & Tutorials, 2019, 21(4): 3039–3071.
- 20 Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv:1412.6980, 2014.
- 21 Paszke A, Chaurasia A, Kim S, *et al.* ENet: A deep neural network architecture for real-time semantic segmentation. arXiv:1606.02147, 2016.
- 22 Nie Z, Xu JC, Zhang SC. Analysis on DeepLabV3+ performance for automatic steel defects detection. arXiv:2004.04822v2, 2020.

(校对责编: 孙君艳)