

基于改进 U-Net 网络的半监督裂缝分割方法

罗 杨¹, 万黎明¹, 李 理¹, 刘知贵^{1,2}

(1. 西南科技大学 计算机科学与技术学院, 四川 绵阳 621000;

2. 西南科技大学 信息工程学院, 四川 绵阳 621000)

摘 要: 裂缝反映了结构的受力状态, 是结构健康检测的重要关注对象之一。基于数字图像利用深度学习方法进行结构表面裂缝自动识别具有速度快、精度高等优势, 不过深度学习方法严重依赖像素级标注信息, 为此, 提出一种基于半监督学习的改进 U-Net 方法。使用特征提取能力更佳的残差网络作为主干特征提取网络代替 U-Net 中由卷积层和池化层进行简单堆叠而成的下采样部分; 在主干网络中插入池化窗口长且窄的条带池化注意力辅助下采样进行特征的细化, 增强特征提取能力; 针对裂缝图像中裂缝区域的亮度普遍暗于背景区域的情况, 网络中的池化操作均采用平均池化使网络能更好地处理裂缝图像; 利用半监督学习, 在训练时同时训练两个网络并利用其分割结果相互监督从而使深度学习分割方法降低对标签数据的依赖度。改进的 U-Net 分割方法在自建裂缝数据集上进行了对比实验, 结果表明, 相较于原始 U-Net 网络, 改进方法具有更高的分割精度, 训练时可使用更少的标签数据。

关键词: 裂缝分割; 半监督; 注意力机制; 深度学习; U-Net

中图分类号: TP391

文献标识码: A

文章编号: 1673-629X(2022)12-0179-06

doi:10.3969/j.issn.1673-629X.2022.12.027

A Semi-supervised Crack Segmentation Method Based on Improved U-Net Network

LUO Yang¹, WAN Li-ming¹, LI Li¹, LIU Zhi-gui^{1,2}

(1. School of Computer Science and Technology, Southwest University of Science and Technology, Mianyang 621000, China;

2. School of Information Engineering, Southwest University of Science and Technology, Mianyang 621000, China)

Abstract: Cracks reflect the stress state of a structure and are one of the important objects of concern for structural health inspection. The automatic recognition of cracks based on digital images using deep learning methods has the advantages of high speed and accuracy, but the deep learning methods rely heavily on pixel-level annotation, so an improved U-Net method based on semi-supervised learning is proposed. A residual network with better feature extraction capability is used as the backbone feature extraction network instead of the down-sampling part of U-Net, which consists of simple stacking of convolutional and pooling layers. A strip pooling attention with long and narrow pooling window is inserted into the backbone to assist down-sampling for feature refinement and enhance feature extraction capability. For the situation that the brightness of crack region in a crack image is generally darker than that of the background region, all of the pooling operation in the network is average pooling so that the network can better handle the crack images. Using semi-supervised learning, two networks are trained at the same time and their segmentation results are used to supervise each other so that the deep learning segmentation method can reduce the dependence of labeled data. The improved U-Net segmentation method is tested on a self-built crack dataset, the experimental results show that the proposed method has higher accuracy than the original U-Net network and fewer label data needed for training.

Key words: crack segmentation; semi-supervised; attention mechanism; deep learning; U-Net

0 引 言

裂缝反映了混凝土结构的受力状态、耐久性与安

全性等问题, 是一种十分常见的结构损伤。受恒活荷载、温度、沉降不均、过度使用、自然灾害等影响均可能

收稿日期: 2022-01-12

修回日期: 2022-05-13

基金项目: 国家自然科学基金(U21A20157)

作者简介: 罗 杨(1993-), 男, 硕士研究生, CCF 会员(J6616G), 研究方向为深度学习与目标检测、图像分割; 通信作者: 刘知贵(1966-), 男, 教授, 硕、博士生导师, 研究方向为计算机应用技术、控制理论应用及自动化装置。

导致裂缝的产生^[1]。而裂缝的产生将导致保护层对内部钢筋的保护失效,引起钢筋锈蚀等问题加剧结构耐久性的降低,即使在没有钢筋的路面结构,也会因裂缝的产生而对路基的稳定性产生影响,严重时可能引发路面沉降、塌陷等问题,因此对于裂缝的准确检测将有助于结构健康的监测与维护。伴随着机器视觉技术及相关硬件的发展,利用数字图像进行裂缝检测成为了一种有效的结构表面损伤无损检测方法,得到了广泛的应用,并逐渐代替人工实现自动化检测。

在裂缝检测任务中,传统的检测方法主要利用阈值分割、边缘检测以及基于手工特征的机器学习方法,虽然取得了较大的进步,但仍然存在精度不高、算法泛用性差等问题,在应用中具有一定的局限性。得益于 GPU (Graphics Processing Unit, 图像处理器) 的发展, Alex Krizhevsky 等人^[2]利用 GPU 高效的并行运算能力,训练了 1 个包含 5 个卷积层和 3 个全连接层的神经网络,使深度学习技术在图像领域取得了突破。这是一种无需人工进行特征设计的方式,深度学习的主干特征提取网络可以自动进行特征提取,根据样本数据训练深度学习网络不断地更新权重,直至最优。利用深度学习的方式进行裂缝检测可以有效地提升任务效率,并且具有较高的可拓展性,受到了工业界的关注并进行了大量研究。刘鑫等人^[3]利用改进的 U-Net

网络实现岩层的裂缝图像分割;余加勇等人^[4]利用无人机与 Mask R-CNN 实现桥梁的裂缝识别;瞿中等人^[5]利用空洞卷积神经网络和多特征融合的方式实现混凝土路面的裂缝图像分割;刘凡等人^[6]利用注意力机制与 U-Net 网络实现路面与大坝的裂缝图像分割。深度学习是以数据为驱动的一种机器学习方式,需要大量的样本数据及对应标签,经学习得到较为理想的结果,不过相较于图像的获取,数据的标注是非常费时费力的。为此,该文提出利用 CPS (Cross Pseudo Supervision, 交叉伪监督) 的训练方式训练 U-Net 网络进行裂缝检测,并对 U-Net 网络进行改进,实现利用较少的标签数据达到较为理想的检测效果。

1 相关工作

1.1 U-Net 网络

卷积神经网络是一种具有权值共享和局部连接等特点的前馈网络,利用卷积层和池化层的堆叠来提取图像特征,实现输入图像特征映射到输出空间。U-Net 网络^[7]则是利用卷积神经网络,在 FCN (Fully Convolutional Network)^[8]的基础上进行改进和优化,用于医学图像分割,整体结构如图 1 所示,由编码器和解码器两个部分组成。

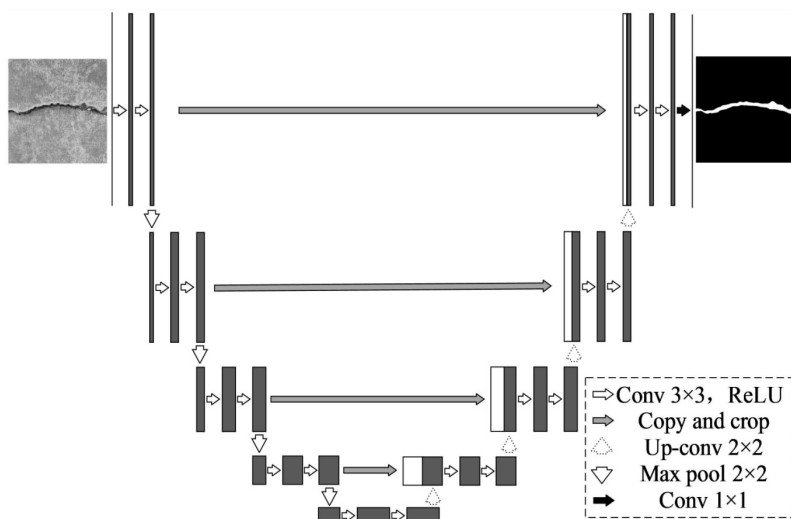


图 1 U-Net 网络基础框架

编码器部分可分为 4 个子模块,由卷积层和池化层组成,用于图像特征提取与下采样,而解码器部分的 4 个子模块与编码器一一对应并跳跃连接,同时利用反卷积操作使图像尺寸逐渐变大,并与跳跃连接的编码器输出进行融合。虽然 U-Net 网络是针对医学图像分割提出的,不过其良好的表现使得很多领域的分割任务也都采用该网络完成,如杨佳林等人^[9]利用改进 U-Net 方法进行遥感图像道路提取,李大海等人^[10]利用改进 U-Net 进行遥感图像云分割。

1.2 残差网络

研究发现,通过堆叠卷积层加深网络可以带来更好的效果,但同样也面临着梯度消失、梯度爆炸以及网络退化等难题。为此,He 等人^[11]提出了 ResNet 网络,对于一个堆叠块(如图 2 所示),采用跳跃连接的方式进行残差学习,使得底层映射为 $H(x)$,通过调整使学习到残差 $F(x) = H(x) - x$,再将原始映射重铸为 $F(x) + x$,从而令整个网络最坏的结果也仅仅是等效于浅层网络,保证了网络的特征提取能力。

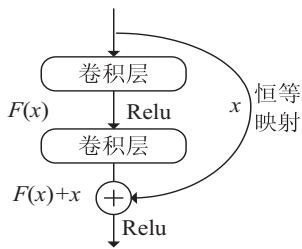


图 2 残差块结构

1.3 注意力机制

将注意力集中至图像中的重要区域并忽略不相关信息,是人类视觉系统中常用的帮助高效理解和分析复杂场景的方法,即为注意力机制。研究人员将这一特性引入到计算机视觉任务中,设计了一个动态选择过程,根据输入信息中的重要性来自适应地加权特征,从而实现重点关注某一区域,在神经网络中,其核心思想是通过这一过程来获取到每张特征图重要性的差异,将神经网络的计算资源集中^[12]。因此注意力模块的使用可以使图像分割网络获得更加密集的上下文信息,从而辅助分割任务进行更精确的像素分类。

1.4 半监督图像分割

图像分割的标签数据相比于原始图像的获取,其成本较为高昂,因此,半监督图像分割的思想便是利用大量的无标签数据来提升算法的性能。常见的半监督学习可分为两类:一致性正则化(Consistency Regularization)和自训练(Self-training)。一致性正则化通过对图像增加各种噪声扰动,使算法不同变换的同样样本有着相似的输出,例如文献[13]对编码器加入特征扰动。自训练^[14]则是通过少量有标签的样本训练网络,利用训练好的网络对无标签数据生成分割标签,称之为伪标签,最后将原标签数据与伪标签数据混合再次对网络进行训练。

2 网络结构

2.1 池化层

池化操作就是利用固定大小的窗口对输入图像以固定步长进行遍历,并对窗口区域内的像素值按一定的规则输出为一个像素值,用于降低计算量,并在一定程度上防止过拟合。常用的有最大池化和均值池化,

即选择区域内的最大值或者平均值作为输出,两者的输出效果也不尽相同,其对比结果如图 3 所示。最大池化有助于保留较暗背景中的高亮目标,而平均池化有助于保留高亮背景中的较暗目标^[15]。在实际使用中普遍情况是目标亮度大于背景亮度,故而更多地使用了最大池化,而在裂缝图像中,往往是裂缝区域的亮度低于其他非裂缝区域(与图 3 中的白底黑线类似),在进行下采样的过程中需要尽可能保留这一部分信息。因此,该文采用的是平均池化,代替原来的最大池化。

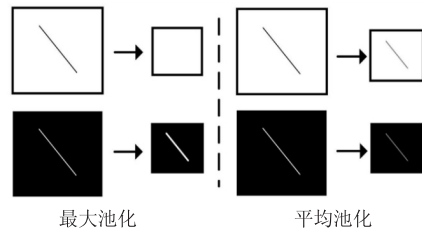


图 3 池化操作对比情况

2.2 Strip Pooling 注意力

对于常规的池化操作,通常是在一个范围大小为 $h \times w$ 的区域内进行,这种方式在处理不规则物体时常会不可避免地合并一些不相关的区域。为了缓解这一情况,SP(Strip Pooling)注意力^[16]使用一个条带状的池化窗口沿着水平或垂直方向进行池化操作,其结构如图 4 所示。对于输入尺寸为 $H \times W$ 的特征图,分别采用尺寸大小为 $H \times 1$ 和 $1 \times W$ 的池化窗口进行池化遍历,对窗口内元素求平均值:

$$y_i^h = \frac{1}{W} \sum_{0 \leq j \leq W} x_{i,j} \quad (1)$$

$$y_j^w = \frac{1}{H} \sum_{0 \leq i \leq H} x_{i,j} \quad (2)$$

其中, y_i^h 和 y_j^w 分别表示水平和垂直方向上的输出,分别得到 $1 \times W$ 和 $H \times 1$ 的特征图。对池化输出的特征图进行一维卷积再扩张至 $H \times W$ 大小,随后将两个维度方向上的特征图融合,经过 1×1 卷积和 Sigmoid 函数后得到注意力图,将其作用于原特征图进行信息权重的调整。利用这种条带状的池化窗口进行池化可以很容易在离散区域内建立长距离依赖关系,并且捕获细节信息。

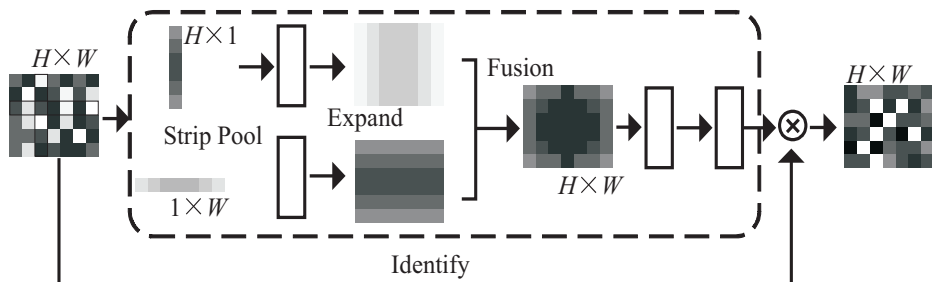


图 4 SP 注意力结构

2.3 网络总体框架

整个网络以具有 34 层深度的残差网络 ResNet-34^[11] 为基础,第一卷积阶段对输入图像以大小为 7×7 、通道数 64 的卷积核进行卷积处理,第二卷积阶段首先进行 3×3 的平均池化操作,随后的卷积操作由残差块构成,每个块分别由两次卷积、批正则化、ReLU 组成,并在 Cov2 到 Cov5 的四个卷积阶段中分别进行 3、4、6、3 次的残差块堆叠,并且在每个残差块中,加入了 SP 注意力模块,用于提升网络特征提取能力。主干网络采样过程中,采用扩边 padding=1 的方式使 3×3 卷

积前后图尺寸大小保持一致,而 7×7 的卷积层则填充为 3。

改进后整个 U-Net 的网络结构如图 5 所示,SP 模块与卷积模块处理同一输入,随后融合。编码器的输出与解码器对应跳跃连接,用于辅助解码器恢复下采样过程中所丢失的特征信息,从而更好地恢复图像。跳跃连接首先对特征图经过 1×1 卷积,再与解码器中上采样后的特征图进行拼接,随后进行 3×3 卷积,反复 4 次后再次经过 1×1 卷积调整通道数,经 Softmax 激活后得到与原图尺寸相同的概率图。

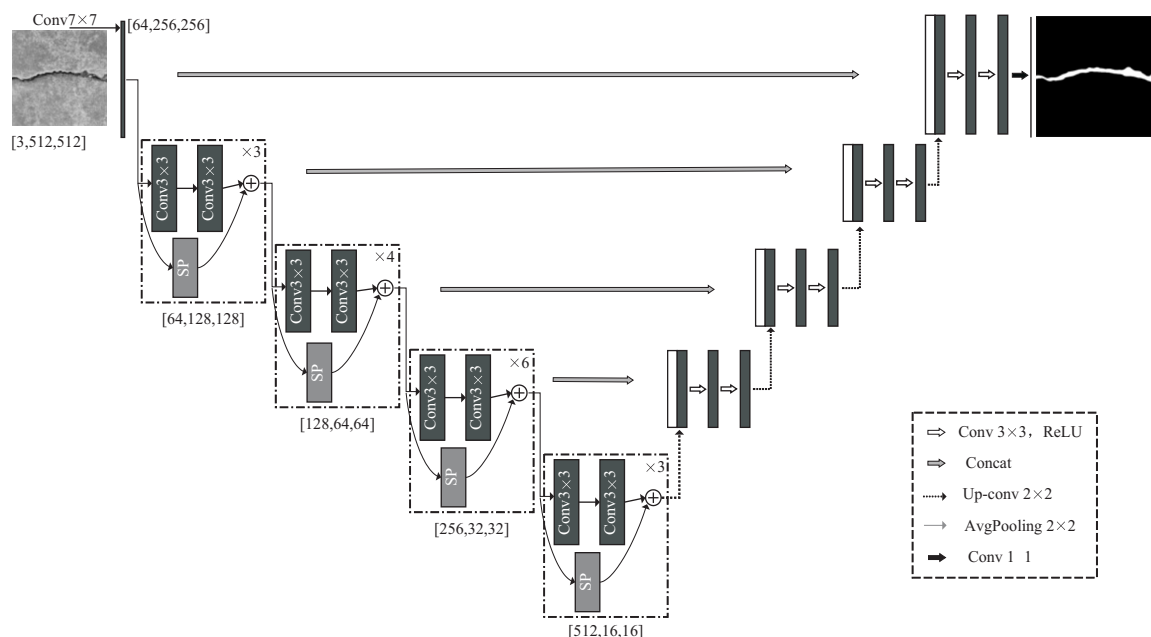


图 5 SP-Unet 网络结构

2.4 交叉伪监督训练

数据集的标签制作是一个非常耗时且费力的过程,对于标签数据有限的数据集而言,利用半监督学习训练网络是一种行之有效的办法。自训练可以利用伪标签扩充标签数据,而一致性学习则通过约束条件使网络学习到更加紧密的特征编码,CPS^[17] (Cross Pseudo Supervision,交叉伪监督)则是将两者相结合,利用两种方法的优势,提高网络精度。CPS 结构如图 6 所示。训练时,随机初始化两个网络 $f(\theta_1)$ 和 $f(\theta_2)$ 处理用于同一输入 X ,得到两个不同的 softmax 输出 Out_1 和 Out_2 ,以及对应的伪分割图 Y_1 和 Y_2 ,利用伪分割图对 Out_1 和 Out_2 进行损失监督。

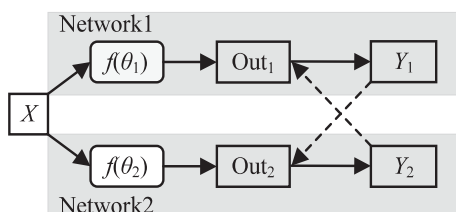


图 6 CPS 结构(---►表示损失监督)

3 实验分析

3.1 实验设置

该文所使用的混凝土裂缝数据集为自建数据集,共包含尺寸为 1024×1024 的图像数据 241 张并全部由专业工程师利用 Labelme 工具进行标注,将其按 7:1:2 的比例分别设置为训练集、验证集和测试集。训练时使用 Adam 算法^[18] 作为优化器,学习率初始值设定为 0.0001,因同时训练两个网络,需要更大的显存空间,受制于硬件设备,参数 batch_size 设定为 2,并且为了防止数据量过大,训练时将输入图像尺寸重塑为 512×512 大小。全部实验均基于 pyTorch 深度学习框架,在显存为 8 GB 的 NVIDIA RTX3070 Ti GPU 上进行,编程语言为 Python3.8,操作系统为 Windows 10。

3.2 评价指标

图像分割可视为一个对像素点进行前景与背景分类的二分类任务,正样本被正确分类时记为 TP (True Positive),正样本被错误分类时记为 FN (False

Negative), 负样本被正确分类时记为 TN (True Negative), 负样本被错误分类时即为 FP (False Positive), 则常用的评价指标精度 (precision) 和 IOU (Intersection Over Union) 为:

$$\text{precision} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{IOU} = \frac{TP}{TP + FP + FN} \quad (4)$$

为了更加准确地评估算法性能, 还选取了 DICE 系数、MIOU (Mean IOU) 和 FPS (Frames Per Second, 每秒传输帧数) 作为补充评价指标, 其中 DICE 系数是一种用于度量相似度的函数, 当达到完美分割时其值为 1, 因此也常用于分割网络的损失函数, 其公式为:

$$\text{DICE} = \frac{2|X \cap Y|}{|X| + |Y|} \quad (5)$$

MIOU 为真实值和预测值两个集合间的交并比, 其公式为:

$$\text{mIOU} = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - p_{ii}} \quad (6)$$

式中, p_{ij} 表示真实值为 i , 被预测为 j 的数量, $k+1$ 为类别个数, p_{ii} 为真值的数量, p_{ij} 和 p_{ji} 分别表示假正例和假负例。mIOU 的计算是基于全局进行的, 因此是一个全局性的评价。

FPS 用于分割速度的评价, 值越大速度越快, 其计算公式为:

$$\text{FPS} = \frac{S_{\text{num}}}{T} \quad (7)$$

式中, S_{num} 表示测试集样本数量, T 表示测试所用时长。

3.3 实验结果分析

为了评估方法的有效性, 该文对比了 U-Net、Res-UNet、PSPNet^[19] 和 DeepLabv3+^[20] 等经典的图像分割方法, 结果如表 1 所示。

表 1 分割效果对比实验

方法	精度/%	IOU/%	mIOU/%	DICE/%	FPS
Res-UNet	92.2	85.3	92.2	92.1	11.2
U-Net	89.4	83.4	91.1	90.9	11.9
PSPNet	92.1	84.0	91.5	91.3	13.7
DeepLabv3+	73.7	72.3	85.3	83.9	11.8
文中方法	94.6	85.6	92.3	92.2	10.6

可以看出, 该方法的分割精度最佳, 并且其他各项指标也均取得最佳结果, 这是由于在加入了 SP 注意力模块与平均池化后, 网络特征提取效果得到了提高, 更有利于裂缝的特征提取, 并且交叉伪监督的训练方式变相扩充了数据集, 对网络训练更有利。在方法的速度方面, 由于使用了注意力机制, 在图像处理过程中, 插入的注意力模块会参与到特征提取过程中进行特征的细化, 这部分的计算带来了些许时间成本, 导致了速度的下降。虽然改进后的方法对 FPS 指标略有影响, 但分割精度有较大提升, 整体上依然具有一定的优势。

为了评估交叉伪监督训练的有效性, 该文将训练集减少至 50% 再次进行训练, 并利用相同测试集进行测试, 结果如表 2 所示。

可以看出, 几乎所有的方法在分割效果上都出现了不同程度的降低, 这是训练数据不足所引起的, 网络没有得到足够多的训练样本从而在特征提取方面表现不佳。相比于其他方法, 文中方法影响最小, 精度降低 1.0%, IOU 降低 0.1%, 这两项指标略有降低, 而 mIOU 和 DICE 系数与原来持平, 总体上该方法还是保持在了一个较高的水平。

表 2 训练结果对比 (使用 50% 数据集) %

方法	精度	IOU	mIOU	DICE
Res-UNet	89.6	84.4	91.7	91.5
U-Net	84.6	79	88.8	88.3
PSPNet	88.8	78	88.2	87.6
DeepLabv3+	58.4	57.7	77.6	75.2
文中方法	93.6	85.5	92.3	92.2

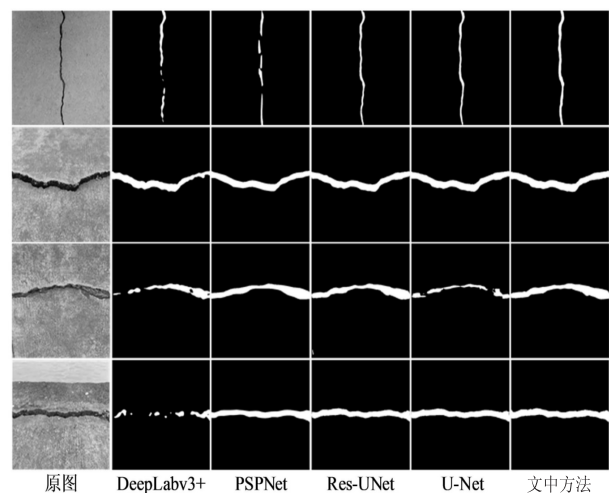


图 7 不同分割方法对比

图 7 展示了不同算法的分割结果,第一列为原图,从无标签数据中随机选取的图像,后面 5 列分别为 5 种算法分割后的二值化图像。可以看出,DeepLab3+、PSPNet 和 U-Net 分割方法出现了不同程度的不连续分割,没能完全分割出裂缝的整体轮廓,而 Res-UNet 方法则在第三张的左下角处出现了错误分割。相较于其他方法,文中分割方法整体上表现较好,整体分割更为细致,更少的像素被错误分类。

4 结束语

以 U-Net 为基础网络进行改进,提出了一种嵌套条带注意力机制的分割方法 (SP-Unet),将网络主干下采样部分替换为具有更强特征提取能力的 ResNet-34,并在其中插入 SP 注意力模块,同时将全部的池化层替换为平均池化,强化下采样部分对裂缝区域的特征提取能力。另外,为了应对数据集不足的问题,引入了交叉伪监督训练方法,同时训练两个网络并利用输出约束彼此,从而使网络在具有较少标签数据时依然能够保持一个较高的精度。

参考文献:

- [1] 刘宇飞,樊健生,聂建国,等. 结构表面裂缝数字图像法识别研究综述与前景展望[J]. 土木工程学报,2021,54(6): 79-98.
- [2] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. Advances in Neural Information Processing Systems, 2012, 25: 1097-1105.
- [3] 刘鑫,赵慧杰,武苏雯. 基于 U-Net 的岩层图像裂缝检测算法应用[J]. 计算机技术与发展,2021,31(5): 73-78.
- [4] 余加勇,李锋,薛现凯,等. 基于无人机及 Mask R-CNN 的桥梁结构裂缝智能识别[J]. 中国公路学报,2021,34(12): 80-90.
- [5] 瞿中,陈雯. 基于空洞卷积和多特征融合的混凝土路面裂缝检测[J]. 计算机科学,2022,49(3): 192-196.
- [6] 刘凡,王君锋,陈峙宇,等. 基于并行注意力 UNet 的裂缝检测方法[J]. 计算机研究与发展,2021,58(8): 1718-1726.
- [7] RÖNNBERGER O, FISCHER P, BROX T. U-net; convolutional networks for biomedical image segmentation[C]//International conference on medical image computing and computer-assisted intervention. [s. l.]: Springer, 2015: 234-241.
- [8] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Boston: IEEE, 2015: 3431-3440.
- [9] 杨佳林,郭学俊,陈泽华. 改进 U-Net 型网络的遥感图像道路提取[J]. 中国图象图形学报,2021,26(12): 3005-3014.
- [10] 李大海,王榆峰,王振东. 面向遥感图像云分割问题的新型 U-Net 模型[J]. 计算机应用研究,2021,38(11): 3506-3509.
- [11] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas: IEEE, 2016: 770-778.
- [12] 张宸嘉,朱磊,俞璐. 卷积神经网络中的注意力机制综述[J]. 计算机工程与应用,2021,57(20): 64-72.
- [13] OUALI Y, HUDELOT C, TAMI M. Semi-supervised semantic segmentation with cross-consistency training[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Seattle: IEEE, 2020: 12674-12684.
- [14] 刘培,黄雅平. 基于图像语义分割的半监督裂纹检测方法[J]. 激光与光电子学进展,2021,58(6): 317-328.
- [15] YU D, WANG H, CHEN P, et al. Mixed pooling for convolutional neural networks[C]//International conference on rough sets and knowledge technology. [s. l.]: Springer, 2014: 364-375.
- [16] HOU Q, ZHANG L, CHENG M M, et al. Strip pooling: rethinking spatial pooling for scene parsing[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Seattle: IEEE, 2020: 4003-4012.
- [17] CHEN X, YUAN Y, ZENG G, et al. Semi-supervised semantic segmentation with cross pseudo supervision[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. [s. l.]: IEEE, 2021: 2613-2622.
- [18] KINGMA D, BA J. Adam: a method for stochastic optimization[C]//Proceedings of the international conference on learning representations. San Diego: ICLR, 2015: 1-15.
- [19] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Honolulu: IEEE, 2017: 2881-2890.
- [20] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Proceedings of the European conference on computer vision (ECCV). [s. l.]: Springer, 2018: 801-818.