



DOI: 10.12086/oe.2020.200036

## 基于U型全卷积神经网络的路面裂缝检测

陈涵深<sup>1,2</sup>, 姚明海<sup>1\*</sup>, 瞿心昱<sup>2</sup><sup>1</sup>浙江工业大学信息工程学院, 浙江 杭州 310023;<sup>2</sup>浙江交通职业技术学院, 浙江 杭州 311112

U-shaped fully convolutional neural network



**摘要:** 路面裂缝检测是道路运营和维护的一项重要工作, 由于裂缝没有固定形状而且纹理特征受光照影响大, 基于图像的精确裂缝检测是一项巨大的挑战。本文针对裂缝图像的特点, 提出了一种U型结构的卷积神经网络UCrackNet。首先在跳跃连接中加入Dropout层来提高网络的泛化能力; 其次, 针对上采样中容易产生边缘轮廓失真的问题, 采用池化索引对图像边界特征进行高保真恢复; 最后, 为了更好地提取局部细节和全局上下文信息, 采用不同扩张系数的空洞卷积密集连接来实现感受野的均衡, 同时嵌入多层输出融合来进一步提升模型的检测精度。在公开的道路裂缝数据集CrackTree206和AIMCrack上测试表明, 该算法能有效地检测出路面裂缝, 并且具有一定的鲁棒性。

**关键词:** 裂缝检测; 卷积神经网络; UCrackNet; 感受野

**中图分类号:** TP391.4

**文献标志码:** A

**引用格式:** 陈涵深, 姚明海, 瞿心昱. 基于U型全卷积神经网络的路面裂缝检测[J]. 光电工程, 2020, 47(12): 200036

## Pavement crack detection based on the U-shaped fully convolutional neural network

Chen Hanshen<sup>1,2</sup>, Yao Minghai<sup>1\*</sup>, Qu Xinyu<sup>2</sup><sup>1</sup>College of Information Engineering, Zhejiang University of Technology, Hangzhou, Zhejiang 310023, China;<sup>2</sup>Zhejiang Institute of Communications, Hangzhou, Zhejiang 311112, China

**Abstract:** Crack detection is one of the most important works in the system of pavement management. Cracks do not have a certain shape and the appearance of cracks usually changes drastically in different lighting conditions, making it hard to be detected by the algorithm with imagery analytics. To address these issues, we propose an effective U-shaped fully convolutional neural network called UCrackNet. First, a dropout layer is added into the skip connection to achieve better generalization. Second, pooling indices is used to reduce the shift and distortion during the up-sampling process. Third, four atrous convolutions with different dilation rates are densely connected in the bridge block, so that the receptive field of the network could cover each pixel of the whole image. In addition, multi-level fusion is introduced in the output stage to achieve better performance. Evaluations on the two public CrackTree206 and AIMCrack datasets demonstrate that the proposed method achieves high accuracy results and good generalization ability.

**Keywords:** crack detection; convolutional neural network; UCrackNet; receptive field

收稿日期: 2020-01-20; 收到修改稿日期: 2020-04-10

基金项目: 国家自然科学基金资助项目(61871350); 浙江省自然科学基金资助(GG19E050005)

作者简介: 陈涵深(1983-), 男, 博士研究生, 讲师, 主要从事计算机视觉和机器学习的研究。E-mail: chs9811@163.com

通信作者: 姚明海(1963-), 男, 博士, 教授, 主要从事机器学习和模式识别的研究。E-mail: ymh@zjut.edu.cn

版权所有©2020 中国科学院光电技术研究所

**Citation:** Chen H S, Yao M H, Qu X Y. Pavement crack detection based on the U-shaped fully convolutional neural network[J]. *Opto-Electronic Engineering*, 2020, 47(12): 200036

## 1 引言

近年来,随着国民经济发展的日新月异,公路交通事业得到了快速的发展,截止2019年中旬,我国交通公路总里程数为484.65万公里<sup>[1]</sup>,已跃居世界第一。公路日常养护问题变得日益突出,及时发现和维护受损路面可以极大地节约公路养护成本,同时,随着我国全国机动车保有量增加,因交通道路维护不及时,交通路面的凹坑、坑井以及塌陷等异常情况会直接影响到交通效率和行车安全<sup>[2]</sup>。因此,为了更好地保障道路性能和安全,对路面缺陷进行快速、准确的分析和评估已经成为当务之急。

裂缝是一种最常见的路面缺陷类型<sup>[3]</sup>,基于视觉的自动检测裂缝已成为研究的热点。目前基于视觉的裂缝检测方法大致可分为传统图像处理方法和深度学习方法,其中传统图像处理方法主要通过分析裂缝的图像特征<sup>[4-5]</sup>,如纹理、边缘和对比度等,通过人为地设计分割和提取的依据,能够在特定的数据集上取得良好的检测效果。深度学习方法已经在多个视觉应用场合表现出卓越的性能,研究者对基于深度学习的图像裂缝检测也展开了广泛的研究,提出的方法根据任务类型不同大致可分为三类:1) 图像分类方法<sup>[6-7]</sup>,判断采集的图像中是否存在裂缝;2) 目标检测方法<sup>[8-9]</sup>,定位裂缝在图像的位置;3) 像素级(pixel-level)预测方法<sup>[10-11]</sup>,能够对图中的每一个像素都得到一个对应的分类结果。前两种算法能够在路面图像中定位裂缝,但无法逐像素检测裂缝,而像素级预测方法可以获得裂缝的几何特征,例如形状、方向、长度和宽度,这对于准确评估路面状况并做出路面维护决策至关重要。

在实际场景中,像素级的裂缝检测和自然图像的语义分割区别较大,主要体现在以下三点:1) 对比常见的语义分割图像,裂缝没有固定的形状并且常常存在极端的长宽比;2) 前景图像和背景图像的像素比极度不均衡;3) 目前公开的数据集规模均较小,可训练的标注图像较为有限。因此,尽管一些语义分割模型在自然图像的应用场景下具有出色的整体性能,但应用在裂缝检测时效果并不一定好。比如目前比较流行的语义分割模型 DeepLab v3+<sup>[12]</sup>,在编码器(encoder)

下采样4倍后,会有一个分支连接到解码器(decoder)部分,而解码器的输出前会有一个直接4倍双线性插值操作,即输入到输出有一个4倍下采样和4倍上采样操作。如果一个目标物的长度或者宽度小于4个像素,即便在特征提取的下采样过程中保留了完整的细节,直接4倍双线性插值也是不能恢复小于4个像素的空间信息,而裂纹在图像中的宽度有时会小于4个像素。因此,将 DeepLab v3+ 直接应用到裂缝检测,识别效果并不出色,但该网络的一些设计思想,如多尺度信息的捕获和融合,对裂缝检测的网络模型设计具有一定的参考价值。

本文主要关注基于U型卷积神经网络的模型设计,对提高像素级裂缝检测的性能展开研究。由于路面状况复杂,很难找到对所有不同路面均有效的特征提取方法,室外环境下的像素级裂缝检测仍然是一个挑战。为了使神经网络具备强大的特征表示能力,目前已经提出了几种多尺度和多层次特征提取的裂缝检测方法<sup>[13-14]</sup>,这些方法均使用特征金字塔模块作为多尺度特征提取器,在多种分辨率条件下捕获丰富的上下文信息。此外,Mei等人<sup>[15]</sup>使用密集连接(densely connected)的卷积神经网络来实现裂缝检测,并引入了像素之间连通性的损失函数,以克服反卷积层输出中存在的裂缝分散问题。Fei等人<sup>[16]</sup>提出了一种用于3D沥青路面裂缝检测的CrackNet-V高效深度神经网络,CrackNet-V建立在先前的CrackNet<sup>[17]</sup>工作基础上,使用多个小卷积核(3×3)的卷积层来增加网络结构的深度,以此在不增加额外参数的基础上提高准确性和计算效率。因上述研究使用不同的裂缝数据集或者同数据集不同评估方法进行实验,故无法直接对比其实际效果。

本文借鉴了最新语义分割模型的设计思路,针对裂缝图像的特点,提出一种有效的基于U型全卷积神经网络的裂缝检测模型,这里称为UCrackNet,用于挖掘更多有利于裂缝分割的特征信息,从而提高分割精度和泛化能力,所提出的方法具有以下优点:

- 1) 在模型中同时引入跳跃连接和池化索引,减少下采样和上采样过程中的特征丢失,从而改善边缘轮廓的检测效果;
- 2) 考虑裂缝没有固定的形状并且常常展现出极

端的长宽比,使用不同扩张系数的空洞卷积密集连接来改善网络的感受野,使其在检测中既关注细节又能充分利用上下文信息;

3) 为了进一步提升对不同尺度物体的分割鲁棒性,在解码器部分采用多层输出融合技术,该方法能更好的学习来自于不同层次的卷积特征信息。

## 2 裂缝检测算法

### 2.1 算法框架

基于深度学习的像素级裂缝检测通常分为三步:

1) 通过图像传感器获取道路图像,提取感兴趣区域(region of interest, ROI),减少不相关物体的干扰;2) 图像预处理,为了减轻检测模型对计算资源的要求,使用滑动窗口在 ROI 图像截取图像块,然后在输入分割模型之前,对其进行简单的归一化预处理,即把输入图像减去每个通道的均值,再除以 128。考虑卷积神经网络具有较强的拟合能力,本文未使用对比度均衡、伽马校正等图像预增强技术;3) 图像分割,使用分割模型对图像中的每一个像素都进行一个对应的分类输出,最终通过合并图像块的预测结果得到整个路面裂缝区域。由于前两步骤比较简单,本文着重介绍第三步骤。

### 2.2 模型总体结构

U-Net<sup>[18]</sup>是一种经典的U型结构全卷积神经网络,在小样本的眼底视网膜分割、细胞分割等应用场景中

获得了非常优秀的结果。本文借鉴了 U-Net 网络的对称式设计思路,提出一种 U 型结构的全卷积神经网络 UCrackNet,全卷积神经网络由于没有全连接层,能够保留一定的上下文特征和空间信息。UCrackNet 的模型结构如图 1 所示,它由编码器、桥接单元(bridge block)和解码器三部分组成。

考虑到 VGG-16<sup>[19]</sup>在很多项任务上均比 VGG-13 展现出更好的性能,同时 VGG-16 相对于 VGG-13 具有更深的网络深度和更强非线性拟合的能力,本文将 VGG-16 作为 UCrackNet 编码器的基础结构(backbone)。U-Net 桥接部分是由两个 1024 通道 3×3 卷积组成,而本文使用的是多个密集连接的空洞卷积,具体介绍详见 2.4 小节。模型的解码器部分采用编码器类似结构的对称设计,其中 VGG-16 中的池化层更换成上采样层,最终输出像素级的预测结果。

UCrackNet 也采用了 U-Net 结构中跳跃连接方式(skip connection),在最大池化和反卷积之间引入直连方式,能够把某些特征从编码器中直接传播到对应的解码器中,该方式不仅可保留来自编码部分的上下文信息,而且引入多个路径进行反向传播梯度,可减轻梯度消失的问题。不同之处是在每层跳跃连接中加入一个参数为 0.3 的 Dropout 层来缓解网络过拟合,达到提升网络泛化能力的效果。

### 2.3 池化索引

池化层可以实现平移的不变性但也导致细小的空间位置偏移。UCrackNet 模型中使用 4 个最大池化层,

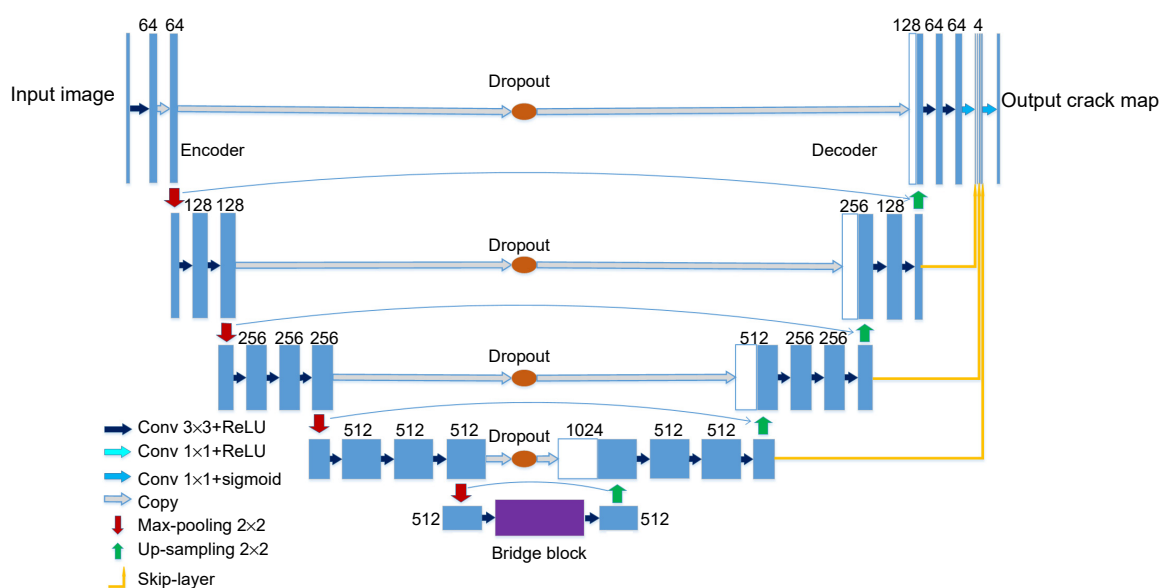


图 1 UCrackNet 模型结构图

Fig. 1 The structure of the proposed UCrackNet

这会加剧特征图中的空间位置信息丢失，导致在图像分割时产生边缘细节的损失。道路裂缝往往具有宽度窄、长度长等几何特点，边缘位置信息的正确定位显得更加重要。本文借鉴 SegNet 中提出的池化索引(pooling indices)思想<sup>[20]</sup>，在裂缝检测的下采样过程中记录每个池中最大特征值的位置，在上采样时使用该位置索引对图像特征进行恢复，得到一个稀疏特征图，最后通过卷积层生成稠密特征图。该方法具有以下优点：1) 能够改善边缘轮廓的检测效果；2) 只需要增加少量参数(存储池化索引)且该过程不需要训练。

在编码器的池化层中，把输入特征图划分为  $K$  个矩形区域  $P_m$ ，其中  $P_m$  为特征向量区域  $m$  的集合。最大池化操作如式(1)所示，该式是对每个子区域  $P_m$  求最大值。这里加入索引存储  $B_{m,j}$ ，用于记录每个子区域  $P_m$  中最大值数的相对位置，定义  $\max\text{Loc}$  为求区域最大值的索引函数，其计算为式(2)。在解码器的上采样层中，设  $\mathbf{y}_{m,j}$  为输入的一个特征向量，这时利用式(3)将该特征向量根据索引矩阵恢复到对应的空间位置。

$$\mathbf{h}_{m,j} = \max_{i \in P_m} \mathbf{a}_{i,j} \quad (j=1, \dots, K), \quad (1)$$

$$B_{m,j} = \max\text{Loc} a_{i,j} \quad (j=1, \dots, K), \quad (2)$$

其中：

$$B_{m,j} = \{b^{(i)}; b \in \{0,1\}\},$$

$$\mathbf{Y}_{m,j} = \mathbf{y}_{m,j} \times B_{m,j} \quad (j=1, \dots, K), \quad (3)$$

其中： $\mathbf{h}_{m,j}$ 是最大池化操作后的输出向量， $\mathbf{a}_{i,j}$ 为输入池化操作的一个特征向量， $\mathbf{Y}_{m,j}$ 是特征图上采样后的输出矩阵。相关的操作示例如图 2 所示。

### 2.4 感受野均衡

考虑到裂缝长宽不一，为了能更好地提取局部细节和全局上下文信息，要求模型同时具备大小不同的感受野<sup>[21]</sup>。使用池化(pooling)可以成倍增加特征点的

感受野，但在下采样过程中会降低中心特征图的分辨率而可能丢失空间信息。针对该问题，可以使用空洞卷积(atrous convolution)<sup>[22]</sup>，它能够在不损失分辨率和不增加计算量的情况下增大感受野，然而直接使用空洞卷积存在网格效应(griding effect)<sup>[23]</sup>，也会产生临近信息的丢失，影响信息的连续性。例如，使用大扩张系数(dilation rate)的空洞卷积在卷积过程中会插入多个空洞来获取大的感受野，这对于小的物体而言，可能会采集不到。因此如何均衡不同大小物体之间的关系，是设计空洞卷积网络的关键。

ASPP<sup>[22]</sup>在网络解码器中对于不同尺度使用不同大小的扩张系数来获取多尺度信息，每个尺度的特征提取是一个独立的分支，在网络最后把它们融合起来再接一个卷积层输出，这样的设计可以有效避免在编码器上冗余的信息获取，而直接关注物体内部之间的相关性。然而，ASPP 使用了多个大扩张系数(插入更多的零)的空洞卷积并联组成，对裂缝检测而言，ASPP 获得的多尺度视野仍然不够密集。受 DenseASPP<sup>[24]</sup>使用密集连接来覆盖大范围感受野的启发，在桥接单元加入不同扩张系数的空洞卷积，并设计成密集连接来减少此类问题的影响。从网络局部分解来看，整个网络由扩张系数分别为 1、2、3、4 的  $3 \times 3$  卷积层串联而成。其中扩张系数为 1 的空洞卷积就是普通卷积，扩张系数为 2、3、4 的空洞卷积会在相邻两个卷积核权值之间分别有 1、2、3 个空洞。最后在空洞卷积的末尾连接一个  $1 \times 3$  卷积和  $3 \times 1$  卷积的组合，来提高网络的非线性拟合能力和减少输出的通道数，最终得到网络结构如图 3 所示。

图中， $d$  表示空洞卷积的扩张系数， $c_{ch}$  表示网络通道数，当前  $i$  层的理论感受野(receptive field, RF，在公式中用  $R_i$  表示)的计算式为

$$R_i = (R_{i-1} - 1) \times S_i + d_i(K_i - 1) + 1, \quad (4)$$

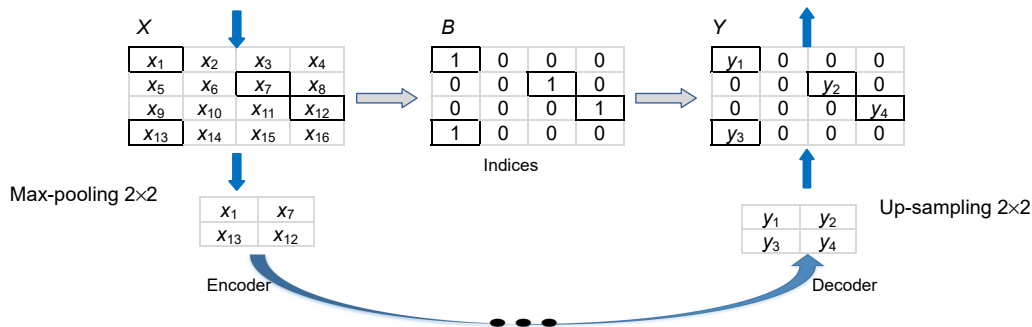


图 2 池化索引操作示意图。左侧是  $2 \times 2$  的最大池化操作，右侧是带位置索引的  $2 \times 2$  上采样操作  
Fig. 2 An illustration of pooling indices. The left and right sides show the operation of max-pooling and up-sampling, respectively

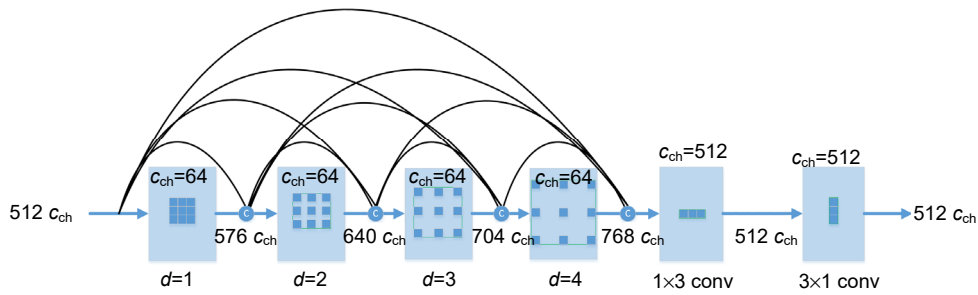


图 3 混合空洞卷积的桥接单元

Fig. 3 The bridge block with atrous convolution

其中： $R_{i-1}$  是前一层的感受野大小， $d_i$ 、 $K_i$ 、 $S_i$  分别为扩张系数、卷积核大小(kernel size)和步长(stride)。当  $d_i=1$  时，就是普通卷积，其等效的感受野计算公式可推算为

$$R_i = (R_{i-1} - 1) \times S_i + K_i \quad (5)$$

根据式(4)和式(5)，得到密集连接部分的感受野如表 1 所示，其中 RF current 和 RF stacking 分别代表当前层感受野和堆叠后的感受野。

表 1 各层的感受野

Table 1 The receptive field of different layers

Layer	Parameter	RF current	RF stacked
1	$d_1=1, K_1=3, S_1=1$	3	3
2	$d_2=2, K_2=3, S_2=1$	5	7
3	$d_3=3, K_3=3, S_3=1$	7	13
4	$d_4=4, K_4=3, S_4=1$	9	21

从表 1 可以看出，密集连接部分的理论感受野尺寸已经达到了  $21 \times 21$ 。扩张系数之所以选择(1,2,3,4)而不是(1,3,5,7)或者 DenseASPP 的(3,6,12,18,24)之类，是因为 AIMCrack<sup>[10]</sup> 的图像块分辨率为  $192 \times 192$ ，通过 UCrackNet 编码器处理后的特征图大小为  $12 \times 12$ 。文献[25]指出感受野的影响分布是近似于高斯分布，高斯分布从中间衰减得十分快，实际有效的感受野大小只是理论感受野的小部分，因此网络的理论感受野应大于特征图的尺寸。(1,2,3,4)的组合能够使空洞卷积之后的特征图像像素点很好地覆盖到整个桥接部分的输入特征图，而(1,3,5,7)或者(3,6,12,18,24)的组合会产生过大的感受野而引起特征细节的丢失。

### 2.5 多层输出融合

目前大多数 U 型结构的全卷积语义分割网络(例如 SegNet、DeepLab v3+)，仅利用卷积网络的最后一层作为输出，在深层次的特征信息中确定浅层的细节

信息，故容易造成高层特征信息的丢失。此外，裂缝图像存在严重前景和背景像素量的不平衡，易出现梯度爆炸或消失的情况，从而使得网络训练时收敛较慢甚至无法收敛。本文借鉴 HED 网络架构<sup>[26]</sup>思想，在解码器的每一个单元(block)中都计算其损失函数，该方式的优势是可以更好地学习来自不同层次的卷积特征信息。UCrackNet 还将解码部分所有层的特征信息进行融合得到最终的特征，其深层次的特征比较粗糙，对于较大的目标以及目标的部分边缘处可以得到较强的响应，而浅层的特征可以为深层特征补充充分的细节信息，同时每一个单元的感受野皆不相同，都被很好地集成到最终的特征表中。

多层输出融合使用图像金字塔策略如图 4 所示。在解码器每个 Block 的输出连接了一个卷积核大小为  $1 \times 1$  卷积来降低特征信息的通道数，然后将所得到的概率映射使用反卷积恢复到原始尺寸。最后将这些概率映射进行堆叠，通过  $1 \times 1$  卷积得到最终的预测结果。鉴于精度与速度间平衡性的考虑，该部分的所有卷积和反卷积的通道数都设置成为 1。

### 2.6 损失函数

裂缝检测属于像素级的二分类问题，二进制的交叉熵(cross-entropy)损失函数可以直接用于网络的训练。定义训练数据集为  $S = \{X_n, Y_n\}$ ，这里的  $X_n = \{x_i^{(n)}, i=1, \dots, M\}$  是输入的图像，而  $Y_n = \{y_i^{(n)}, i=1, \dots, M\}$ ， $y_i^{(n)} \in \{0, 1\}$  是裂缝标注图， $M$  表示每个图像中的像素数量，其损失函数定义如下：

$$L(W)_{\text{seg}} = - \sum_{i=1}^M y_i^{(n)} \log(F_d(x_i^{(n)}; W)) - (1 - y_i^{(n)}) \log(1 - F_d(x_i^{(n)}; W)) \quad (6)$$

其中： $W$  是神经网络模型的可训练参数， $F_d(x_i^{(n)}; W)$  为模型预测样本  $x_i^{(n)}$  为裂缝的概率。利用式(6)计算检测误差，再通过反向传播算法更新其模型参数，经过反复操作最终获得模型的最优参数。

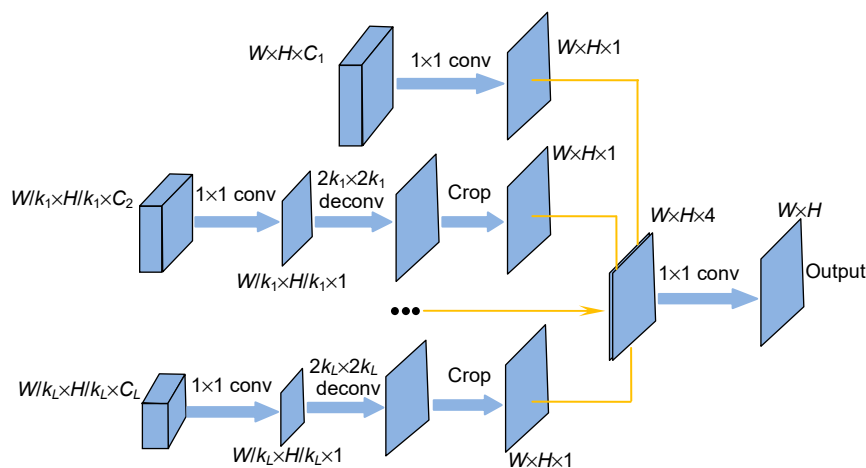


图 4 多层输出融合的网络结构图

Fig. 4 The output stage using multi-level fusion

### 3 实验结果

UCrackNet 模型使用 Python 语言在基于 Tensorflow 的 Keras 深度学习框架上搭建,编写的程序在 Ubuntu16.04 操作系统运行。硬件环境为 Intel i5 8500 处理器, 16 G DDR4 内存和一块 Nvidia GTX1080Ti 11 G 显卡。

#### 3.1 实验数据集

本文选取两个规模相对较大的道路裂缝数据集进行训练和测试。

1) CrackTree206<sup>[27]</sup>: 包含 206 张 800 pixels×600 pixels 的道路裂缝图片,数据集中的裂缝较细并且存在遮挡和阴影干扰。本文随机选取 126 张图片为训练样本, 余下 80 张图片为验证样本。

2) AIMCrack<sup>[10]</sup>: 由安装在车辆内部的行车记录仪以透视图方式(perspective view)拍摄的道路裂缝图像。数据集由韩国市内不同地点、不同时间段拍摄的 527 张 1920 pixels×1080 pixels 的彩色图片组成, 样本干扰物多, 多样性强, 光照影响大, 检测难度相对较高。由于拍摄相机视角的原因, 图像中包含路面和路面上方的信息, 而裂缝识别只需要关注路面信息, 所以本文选用固定参数的感兴趣区域(region of interest, ROI)方法对图像进行裁剪, 以减少非道路信息的干扰。裁

剪后的图像分辨率为 1920 × 384, 并随机划分成 327 张训练样本和 200 张验证样本。

#### 3.2 训练参数设定

由于硬件资源的限制, 将裂缝图像按比例切割成多个图像块作为模型的训练集, 同时为了增加训练样本, 采用间隔为 1/2 图像块宽度的交叠切割方法对训练集进行扩充, 最终得到的训练样本统计数据详见表 2。在训练时, 使用随机对比度(± 5%)、随机亮度(± 5%)和水平翻转的图像增强技术来避免网络的过拟合, 采用 Adam 优化器<sup>[28]</sup>, 设定小批尺寸(mini-batch size)为 12。在数据集 AIMCrack 中, 初始学习率(learning rate)设置为 0.001, 每 40 个周期(epoch)减少 1/3; 对数据集 CrackTree206, 设置初始学习率为 0.0005, 每 20 周期减少 1/3。

#### 3.3 评估指标

将基于像素点的精度(precision, P)、召回率(recall, R)和 F1 分数(F<sub>1</sub>), 三个检测领域常用度量方法作为裂缝预测结果的评价指标。另外, 裂缝检测也可以被看作是一种二进制的语义分割, 从而可以使用语义分割的重叠度(intersection-over-union, IoU)来评价裂缝的检测结果, 不同的是因为道路图像中 90%以上的像素都是背景(非裂缝), 所以这里只计算前景(裂缝)的 IoU。各项评价指标定义如下:

表 2 两个裂缝数据集的训练样本对比

Table 2 Comparison of the two different training sets

Dataset	Image resolution	Patch resolution	Patches
CrackTree206	800×600	160×160	7938
AIMCrack	1920×384	192×192	18639

$$P = \frac{1}{N} \sum_{i=1}^N \frac{T_p(i)}{T_p(i) + F_p(i)}, \quad (7)$$

$$R = \frac{1}{N} \sum_{i=1}^N \frac{T_p(i)}{T_p(i) + F_N(i)}, \quad (8)$$

$$F_1 = \frac{2P \times R}{P + R}, \quad (9)$$

$$IoU = \frac{1}{N} \sum_{i=1}^N \frac{T_p(i)}{T_p(i) + F_p(i) + F_N(i)}, \quad (10)$$

其中： $T_p$  为正确检测到裂缝的像素点个数， $F_p$  为非裂缝区域误检为裂缝的像素点个数， $F_N$  为未正确检测到裂缝的像素点个数， $N$  是验证集中的图像数量。鉴于裂缝的分割结果是作为后期道路质量定量评估的依据，在评估中不考虑像素距离差，也就是说在评估分割结果时，只有在预测边界与真实边界完全像素匹配的情况下才被认为是正确。实验中，设定输出阈值为 0.3，仅将大于该阈值的检测结果视为裂缝，另外，算法的运行时间是通过计算所有验证集样本的神经网络前向传播时间，并求其平均值而得到。

### 3.4 消融实验

为了分析各种改进措施对算法检测效果的影响，UCrackNet 在 AIMCrack 数据集上进行了消融实验，这里的基线是指基于 VGG-16、桥接单元通道数为 512 的 U-Net 模型。从表 3 可以看出，在跳跃连接中加入 Dropout 层对模型性能提升有帮助，同时增加池化素

引可以将算法在验证集的 IoU 再提高 1.7%。另外，使用感受野均衡比原算法在 IoU 上提升 3.6%，验证了合理设计网络感受野对模型设计的重要性，同时加入多层输出融合模块表现更好，对 IoU 和 F1 分别有 3.8%、1.8%的提升。

### 3.5 对比实验

为了验证本文方法的优越性，在 CrackTree206 和 AIMCrack 数据集上与几个主流的网络模型展开了对比实验。CrackForest<sup>[4]</sup>是采用传统图像处理方法来获取裂纹特征，再使用结构化随机森林来实现分类输出的一种裂缝检测算法。CrackForest 的实验结果是由 i5 8500 处理器，16 G 内存的硬件环境和 Matlab2016 软件环境中运行得到。LinkNet<sup>[29]</sup>、DeepCrack<sup>[14]</sup>、U-Net 和 ResUNet<sup>[30]</sup>都是 U 型结构的全卷积神经网络，其中 LinkNet 编码器部分是 ResNet-18<sup>[31]</sup>网络，解码器部分由轻量级的全卷积网络组成；ResUNet 是一种全残差网络设计的 U-Net 模型，为了速度和性能的均衡，该模型只使用了 6 个残差学习单元和 2 个卷积层；DeepCrack 是最近提出的基于 SegNet 基础结构的裂缝检测模型，该模型融合编码器和解码器中各层的多尺度特征，取得了不错的检测效果。

各种方法的定量对比结果见表 4，本文算法在 CrackTree206 数据集的测试 F1 分数达到 81.2%，相比

表 3 UCrackNet 消融实验的结果

Table 3 The result of ablation study for UCrackNet

Baseline	Dropout	Pooling indices	Larger receptive field	Multi-level fusion	IoU	R	P	F1
✓					0.295	0.392	0.484	0.433
✓	✓				0.299	0.423	0.466	0.443
✓	✓	✓			0.304	0.396	0.505	0.444
✓	✓	✓	✓		0.315	0.405	0.512	0.452
✓	✓	✓	✓	✓	0.327	0.408	0.527	0.460

表 4 不同算法在数据集 CrackTree206 和 AIMCrack 上的结果对比

Table 4 Comparison of performance of various methods on the CrackTree206 and AIMCrack datasets

Method	CrackTree206			AIMCrack		
	IoU	F1	Time/ms	IoU	F1	Time/ms
CrackForest	0.160	0.281	1545	0.140	0.271	2493
LinkNet	0.521	0.684	<b>155</b>	0.285	0.421	<b>156</b>
ResUNet	0.668	0.797	181	0.254	0.377	221
U-Net	0.676	0.804	213	0.309	0.446	242
DeepCrack	0.646	0.783	505	0.312	0.450	510
UCrackNet	<b>0.688</b>	<b>0.812</b>	421	<b>0.327</b>	<b>0.460</b>	453

于 DeepCrack 和 U-Net 分别有 3.7%和 1.0%的提升,而在 AIMCrack 数据集上比 DeepCrack 和 U-Net 也分别有 2.2%和 3.1%的提升。从模型的运行时间来看,本文提出的算法在单块 GTX1080Ti 显卡上能超过 2 fps,尽管相比 LinkNet、U-Net 和 ResUNet 显得劣势,但仍快于 DeepCrack。因此,综合考虑各类方法的检测效果与计算速度,UCrackNet 更能满足实际裂缝检测的需求。

根据各种算法在两种数据集上的检测结果,给出了对应的精度-召回率曲线(precision-recall curves)。通过选定不同阈值,以精度和召回率作为变量而绘制出的曲线如图 5。从图 5 可以看出,本文的方法在不同的阈值设定下,精度和召回率上均优于其他对比方法。

此外,分别给出了各种方法在两个数据集上的输出结果,详见图 6 和图 7,红色框表示模型预测中产生的噪声,绿色框表示一些细节的预测结果。可以看出,CrackForest 只能提取图像中的部分裂缝,不适用

于复杂场景下的裂缝图像分割,LinkNet 和 ResUNet 在有干扰物影响的部位准确率较低。DeepCrack 在噪声抑制上好于 U-Net,但 U-Net 在裂缝细节方面要优于 DeepCrack,而本文提出的算法(图 6 和图 7 绿色框部位)相比 LinkNet、ResUNet、U-Net 和 DeepCrack,在裂缝边缘处,完整度上与真实更加接近,而且输出噪声更低。

最后,本文还给出了复杂场景的预测结果对比,如图 8 所示,红色框表示误检区域,黄色框表示一些明显的漏检区域。样本图像选自 AIMCrack 测试集,存在路面灰度不均匀、裂缝深浅各异和双黄线标识干扰等情况。从图中可以看出,LinkNet 和 ResUNet 存在许多漏检现象,未能正确识别出一些明显裂缝;UNet 会发生一定的误检,如图中的红色框所示,将一些非裂缝的像素预测成了裂缝;而 DeepCrack 在双黄线部位不能识别出裂缝;相对而言,本文方法在检测上效果更好,在浅层的裂缝处和道路标识干扰下具有

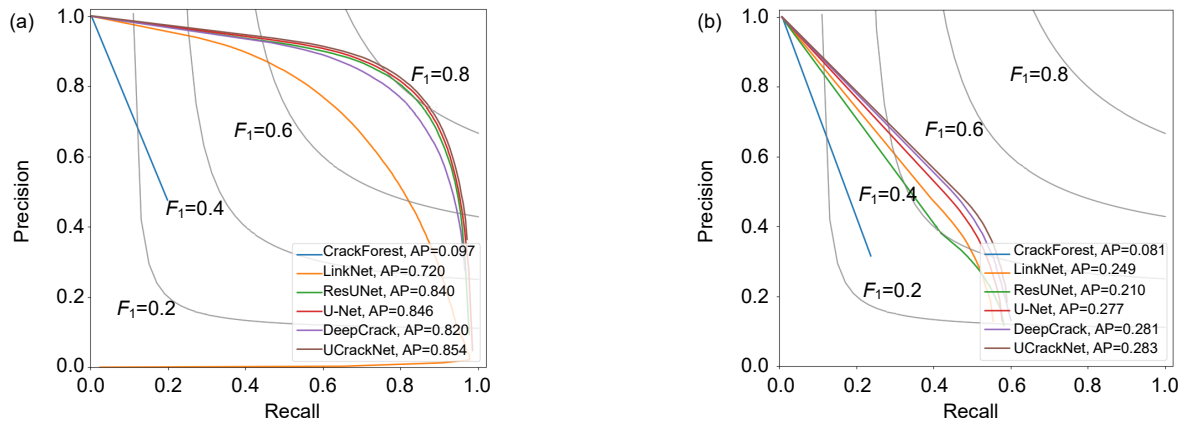


图 5 不同算法在两个数据集中的精度-召回率(P-R)曲线图。(a) CrackTree206; (b) AIMCrack  
Fig. 5 The precision-recall (PR) curves of various methods on the two datasets. (a) CrackTree206; (b) AIMCrack

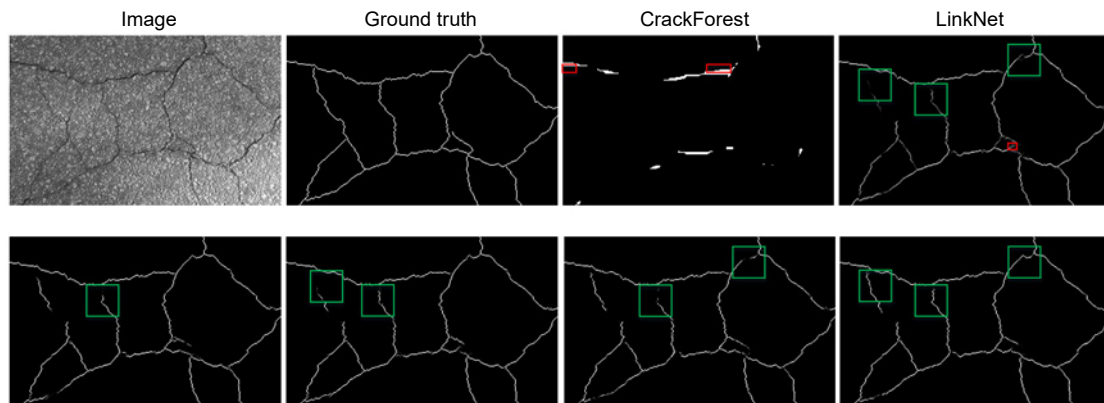


图 6 CrackTree206 数据集上的检测结果对比  
Fig. 6 The predicted results by different methods on the CrackTree206 dataset



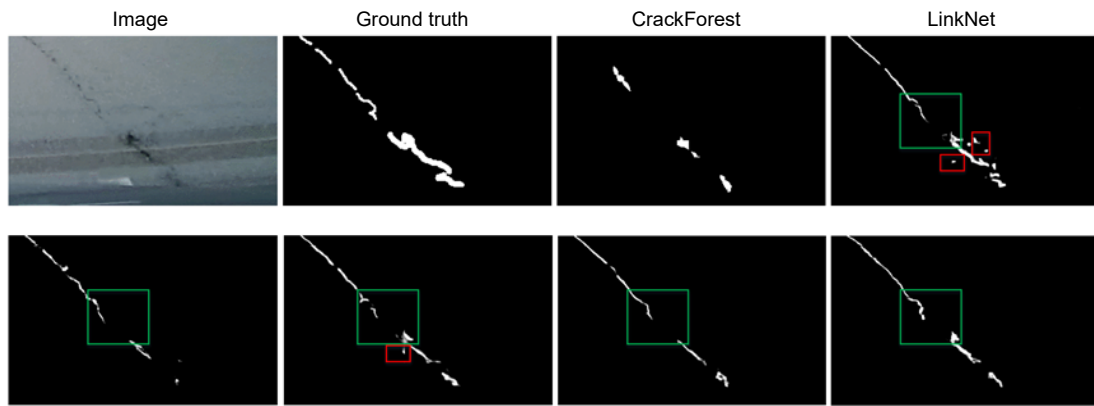


图 7 AIMCrack 数据集上的检测结果对比

Fig. 7 The predicted results by different methods on the AIMCrack dataset

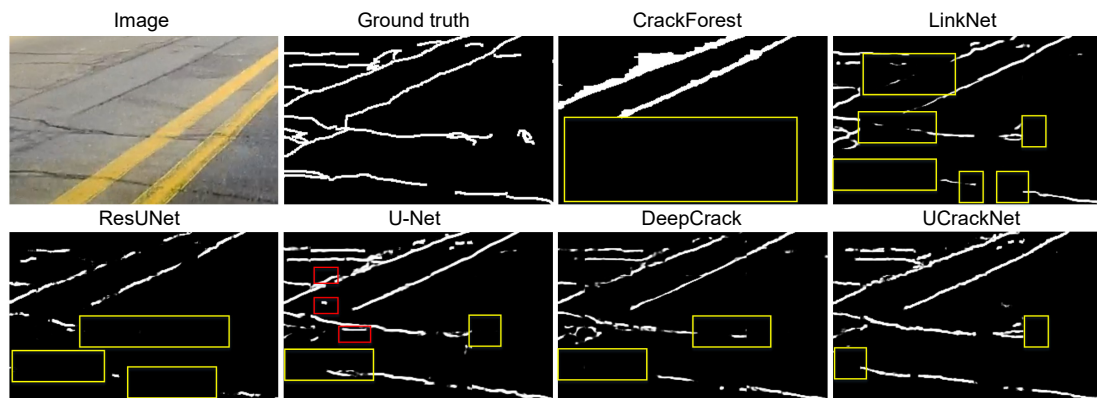


图 8 复杂场景下的检测结果对比

Fig. 8 The predicted results by different methods on the challenging scenario

一定的鲁棒性。当然，本文方法也存在一定的局限性，从图 8 的 UCrackNet 输出结果中可以看到，在某些裂缝特征微弱的地方，也出现一些漏检。在后续的工作中，可以通过扩大训练样本数量来进一步提升该方法的实际检测效果。

## 4 结 论

根据裂缝图像的特点，结合最新的分割模型设计思想，提出了一种基于 U 型全卷积神经网络的裂缝检测模型。通过在公开数据集 CrackTree206 和 AIMCrack 上的测试，获得了比目前主流模型更优的结果。该方法也可通过适当的修改应用于其他语义分割的场景，比如桥梁裂缝检测、物体表面缺陷检测。

现有的基于深度学习的裂缝检测算法本质都是通过神经网络的强大拟合能力来实现裂缝检测，但是这些方法并没有考虑到裂缝的几何约束，因此如何结合裂缝的几何特征和卷积神经网络的学习能力来实现更

加精准的裂缝检测，将作为今后的一个研究方向。

## 参考文献

- [1] [http://www.zgjt.com/2019-10/08/content\\_230254.htm](http://www.zgjt.com/2019-10/08/content_230254.htm).
- [2] Schnebele E, Tanyu B F, Cervone G, et al. Review of remote sensing methodologies for pavement management and assessment[J]. *European Transport Research Review*, 2015, 7(2): 7.
- [3] Zhang D J, Li Q Q. A review of pavement high speed detection technology[J]. *Journal of Geomatics*, 2015, 40(1): 1-8.  
张德津, 李清泉. 公路路面快速检测技术发展综述[J]. *测绘地理信息*, 2015, 40(1): 1-8.
- [4] Shi Y, Cui L M, Qi Z Q, et al. Automatic road crack detection using random structured forests[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2016, 17(12): 3434-3445.
- [5] Xu W, Tang Z M, Lv J Y. Pavement crack detection based on image saliency[J]. *Journal of Image and Graphics*, 2013, 18(1): 69-77.  
徐威, 唐振民, 吕建勇. 基于图像显著性的路面裂缝检测[J]. *中国图象图形学报*, 2013, 18(1): 69-77.
- [6] Zhang L, Yang F, Zhang Y D, et al. Road crack detection using deep convolutional neural network[C]//*International Conference on Image Processing (ICIP)*, Phoenix, AZ, USA, 2016: 3708-3712.

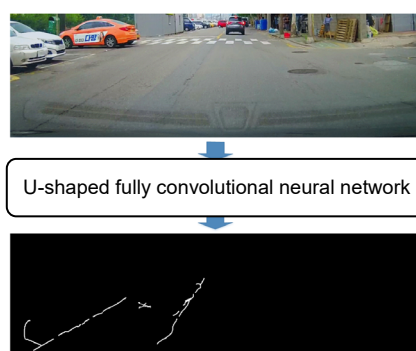
- [7] Cha Y J, Choi W, Büyüköztürk O. Deep learning-based crack damage detection using convolutional neural networks[J]. *Computer-Aided Civil and Infrastructure Engineering*, 2017, **32**(5): 361–378.
- [8] Maeda H, Sekimoto Y, Seto T, *et al.* Road damage detection and classification using deep neural networks with smartphone images[J]. *Computer-Aided Civil and Infrastructure Engineering*, 2018, **33**(12): 1127–1141.
- [9] Carr T A, Jenkins M D, Iglesias M I, *et al.* Road crack detection using a single stage detector based deep neural network[C]//2018 IEEE Workshop on Environmental, Energy and Structural Monitoring Systems, Salerno, Italy, 2018.
- [10] Bang S, Park S, Kim H, *et al.* Encoder–decoder network for pixel-level road crack detection in black-box images[J]. *Computer-Aided Civil and Infrastructure Engineering*, 2019, **34**(8): 713–727.
- [11] Yang X C, Li H, Yu Y T, *et al.* Automatic pixel-level crack detection and measurement using fully convolutional network[J]. *Computer-Aided Civil and Infrastructure Engineering*, 2018, **33**(12): 1090–1109.
- [12] Chen L C, Zhu Y K, Papandreou G, *et al.* Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//European Conference on Computer Vision, Glasgow, United Kingdom, 2018: 833–851.
- [13] Yang F, Zhang L, Yu S J, *et al.* Feature pyramid and hierarchical boosting network for pavement crack detection[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2020, **21**(4): 1525–1535.
- [14] Zou Q, Zhang Z, Li Q Q, *et al.* DeepCrack: learning hierarchical convolutional features for crack detection[J]. *IEEE Transactions on Image Processing*, 2019, **28**(3): 1498–1512.
- [15] Mei Q P, Gül M, Azim M R, *et al.* Densely connected deep neural network considering connectivity of pixels for automatic crack detection[J]. *Automation in Construction*, 2020, **110**: 103018.
- [16] Fei Y, Wang K C P, Zhang A, *et al.* Pixel-level cracking detection on 3D asphalt pavement images through deep-learning-based CrackNet-V[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2020, **21**(1): 273–284.
- [17] Zhang A, Wang K C P, Li B X, *et al.* Automated pixel-level pavement crack detection on 3D asphalt surfaces using a deep-learning network[J]. *Computer-Aided Civil and Infrastructure Engineering*, 2017, **32**(10): 805–819.
- [18] Ronneberger O, Fischer P, Brox T, *et al.* U-Net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention, Lima, Peru, 2015: 234–241.
- [19] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[C]//International Conference on Learning Representations, 2015.
- [20] Badrinarayanan V, Kendall A, Cipolla R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, **39**(12): 2481–2495.
- [21] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions[C]//International Conference on Learning Representations, 2016.
- [22] Chen L C, Papandreou G, Kokkinos I, *et al.* DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, **40**(4): 834–848.
- [23] Wang P Q, Chen P F, Yuan Y, *et al.* Understanding convolution for semantic segmentation[C]//2018 IEEE Winter Conference on Applications of Computer Vision, Lake Tahoe, NV, USA, 2018: 1451–1460.
- [24] Yang M K, Yu K, Zhang C, *et al.* DenseASPP for semantic segmentation in street scenes[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018: 3684–3692.
- [25] Luo W J, Li Y J, Urtasun R, *et al.* Understanding the effective receptive field in deep convolutional neural networks[C]//Proceedings of the 30th International Conference on Neural Information Processing Systems, 2016: 4905–4913.
- [26] Xie S, Tu Z. Holistically-Nested Edge Detection[J]. *International Journal of Computer Vision*, 2015, **125**(1-3): 3–18.
- [27] Zou Q, Cao Y, Li Q Q, *et al.* CrackTree: Automatic crack detection from pavement images[J]. *Pattern Recognition Letters*, 2012, **33**(3): 227–238.
- [28] Kingma D P, Ba L J. Adam: A method for stochastic optimization[C]//International Conference on Learning Representations, Ithaca, NY, 2015.
- [29] Chaurasia A, Culurciello E. LinkNet: Exploiting encoder representations for efficient semantic segmentation[C]//2007 IEEE Visual Communications and Image Processing (VCIP), St. Petersburg, FL, USA, 2017: 1–4.
- [30] Zhang Z X, Liu Q J, Wang Y H. Road extraction by deep residual U-Net[J]. *IEEE Geoscience and Remote Sensing Letters*, 2018, **15**(5): 749–753.
- [31] He K M, Zhang X Y, Ren S Q, *et al.* Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016: 770–778.

# Pavement crack detection based on the U-shaped fully convolutional neural network

Chen Hanshen<sup>1,2</sup>, Yao Minghai<sup>1\*</sup>, Qu Xinyu<sup>2</sup>

<sup>1</sup>College of Information Engineering, Zhejiang University of Technology, Hangzhou, Zhejiang 310023, China;

<sup>2</sup>Zhejiang Institute of Communications, Hangzhou, Zhejiang 311112, China



Processing flow of the proposed method

**Overview:** Cracks are one of the most common categories of pavement distress. Early locating and repairing the cracks can not only reduce the cost of pavement maintenance but also decrease the probability of road accidents happening. Precise measurement of the crack is an essential step toward identifying the road condition and determining rehabilitation strategies. Nevertheless, cracks do not have a certain shape and the appearance of cracks usually changes drastically in different lighting conditions, making it hard to be detected by the algorithm with imagery analytics. Therefore, fully automated and comprehensive crack detection is still challenging. In this paper, we focus on pixel-level crack detection in 2D vision and propose an effective U-shaped fully convolutional neural network called UCrackNet, which is the enhanced version of U-Net. It consists of three main components: an encoder, a bridge block, and a decoder. The backbone of the encoder is the pre-trained VGG-16 that extracts spatial features from the pavement image. The last convolutional layer at each scale in the encoder has a skip connection to connect the corresponding layer in the decoder to preserve and reuse feature maps at different pooling stages. To minimize the possibility of overfitting and achieve better generalization ability, we add a dropout layer into each skip connection. The bridge block is a bridge path between the encoder and the decoder. Motivated by DenseASPP, four densely connected atrous convolutional layers with different dilation rates are employed in the bridge block, so that it generates features with a larger receptive field to effectively capture multi-scale information. The decoder has four convolutional blocks, and in each block, up-sampling with indices is used to reduce the shift and distortion during the up-sampling operation. Furthermore, multi-level fusion is introduced in the output stage to utilize multiscale and multilevel information of objects. The idea of multi-level fusion is inspired by the success of the HED network architecture, which showed that it is capable of fully exploiting the rich feature hierarchies from convolutional neural network (CNNs). Specifically, the feature maps of each stage are first up-sampled to the size of the output image, then a  $1 \times 1$  convolutional is used to fuse these maps to get the final prediction map. Qualitative evaluations on the two public CrackTree206 and AIMCrack datasets demonstrate that the proposed method achieves superior performance compared with CrackForest, LinkNet, ResUNet, U-Net, and DeepCrack. The qualitative results show that our method produces high-quality crack maps, which are closer to the ground-truth and have lower noise compared with the other methods.

**Citation:** Chen H S, Yao M H, Qu X Y. Pavement crack detection based on the U-shaped fully convolutional neural network[J]. *Opto-Electronic Engineering*, 2020, 47(12): 200036

Supported by National Natural Science Foundation of China (61871350) and Zhejiang Provincial National Science Foundation of China (GG19E050005)

\* E-mail: ymh@zjut.edu.cn