计算机科学
COMPUTER SCIENCE

基于注意力机制和可变形卷积的路面裂缝检测

隆涛, 董安国, 刘来君

引用本文

隆涛, 董安国, 刘来君.基于注意力机制和可变形卷积的路面裂缝检测[J]. 计算机科学, 2023, 50(6A): 220300214-6.

LONG Tao, DONG Anguo, LIU Laijun. Pavement Crack Detection Based on Attention Mechanism and Deformable Convolution [J]. Computer Science, 2023, 50(6A): 220300214-6.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.220300214

基于注意力机制和可变形卷积的路面裂缝检测

隆　涛[1]　董安国[1]　刘来君[2]
1 长安大学理学院　西安 710064
2 长安大学公路学院　西安 710064

摘　要　针对较复杂背景下路面裂缝检测问题，由于基于深度学习的图像分割算法检测效果不甚理想，以及裂缝图像自身像素类别不平衡，提出了一种基于注意力机制和可变形卷积的路面裂缝检测网络，该网络基于编码-解码结构进行构建。为了解决较为复杂背景裂缝检测困难的问题，首先，由可变形卷积提升网络对不同形状裂缝线性特征的学习能力；其次，使用密集连接机制强化特征信息；然后，在解码阶段采用转置卷积和桥接方式与编码阶段特征逐步融合，并结合多级特征融合的思想，提高网络的检测精度；最后，引入注意力模块(SimAM)，在不增加网络参数的前提下，更加关注目标特征的提取，抑制背景特征。在两个公开裂缝数据集上进行实验来验证该算法的有效性，实验结果表明，该算法的各项性能评价指标均优于对比算法，BCrack数据集的平均像素精度、平均交并比分别达到92.12%和84.79%，CFD数据集的平均像素精度、平均交并比分别达到91.02%和74.75%，在复杂背景裂缝检测下表现良好，可应用于路面维修工程。

关键词：裂缝检测；编码-解码结构；可变形卷积；密集连接机制；注意力模块
中图法分类号　TP391

Pavement Crack Detection Based on Attention Mechanism and Deformable Convolution

LONG Tao[1], DONG Anguo[1] and LIU Laijun[2]
1 School of Science, Chang'an University, Xi'an 710064, China
2 School of Highway, Chang'an University, Xi'an 710064, China

Abstract　Aiming at the pavement crack detection problem under complex background, due to the unsatisfactory detection effect of image segmentation algorithm based on deep learning, and the imbalance of pixel categories in the crack image itself, this paper proposes a pavement crack detection network based on attention mechanism and deformable convolution, which is constructed based on encoder-decoder structure. In order to solve the problem of difficult crack detection in complex background, firstly, deformable convolutional is used to improve the learning ability of linear features of cracks with different shapes. Secondly, the dense connection mechanism is used to strengthen the feature information. Then, in the decoder stage, the feature fusion of transpose convolution and bridge are adopted, and the multi-stage feature fusion is combined to improve the detection accuracy of the network. Finally, the attention module(SimAM) is introduced to pay more attention to the extraction of target features and suppress background features without increasing network parameters. Experiments are carried out on two open crack datasets to verify the effectiveness of the algorithm. The experimental results show that the performance evaluation criteria of the algorithm are better than the comparison algorithms. The mean pixel accuracy and mean intersection over union of the BCrack dataset reached 92.12% and 84.79%, respectively. The mean pixel accuracy and mean intersection over union of the CFD dataset reached 91.02% and 74.75%, respectively. The average accuracy and average intersection ratio of CFD data set is 91.02% and 74.75%, respectively. The algorithm performs well in crack detection under complex background, and can be applied to pavement maintenance engineering.

Keywords　Crack detection, Encoder-decoder structure, Deformable convolutional, Dense connection mechanism, Attention module

1　引言

道路是交通系统的重要组成部分，为提高服务水平和使用寿命，需要对其进行损伤检测。及时采用有效方法进行裂缝检测并修补损伤，有助于保持交通安全和正常运行[1]。目前裂缝检测方法仍采用人工检测为主，该方法耗时低效、存在主观性，还容易引发安全问题。因此，利用图像处理技术进行道路裂缝检测成为了近年学者研究的一个热门问题。

近年来，随着图像处理技术的研究，基于图像处理的裂缝检测算法大量涌现，如基于阈值[2]、区域[3]和边缘[4]检测是早期常用的分割算法。虽然传统的图像处理方法可以实现裂缝分割，但图像处理过程复杂，且易受光照、阴影、污渍的影响，对于复杂背景裂缝图像检测，传统方法会出现检测的裂缝不完整、错误检测或遗漏检测等问题。

基金项目：陕西省重点产业创新链项目(2020ZDLGY09-09)；国家自然科学基金青年项目(12001057)
This work was supported by the Key Industry Innovation Chain Project of Shaanxi, China(2020ZDLGY09-09) and Youth Project of National Natural Science Foundation of China(12001057).
通信作者：隆涛(longtao1220@163.com)

随着深度学习技术在图像分类上取得突破，基于深度学习的裂缝检测算法的研究发展迅速。目前基于深度学习的裂缝检测方法主要分为3类：基于图像分类的方法、基于目标检测的方法和像素级预测方法。基于图像分类的方法，Cha等[5]首次结合CNN模型和滑动窗口法来检测混凝土表面裂缝，结果表明CNN模型优于传统裂缝检测方法，具有更高的精度和鲁棒性。Li等[6]利用滑动窗口法将桥梁裂缝图像切分为较小面元图像，利用CNN对面元进行分类识别，从而完成裂缝提取。上述基于图像分类的方法由于全连接层的存在，且基于滑动窗口法感受野受限、效率低、计算量较大，速度有待进一步提升。基于目标检测的方法，Zhang等[7]提出了一种基于YOLO3的裂缝检测算法，实现对裂缝的检测。这类方法虽然可以定位裂缝位置，但由于裂缝是不规则的，导致检测精度低，精准定位裂缝困难。上述两种方法虽然能够定位裂缝，但无法逐像素检测裂缝，而像素级预测方法可以获得裂缝几何特征，是目前流行的研究方向。Lecun等[8]为弥补卷积神经网络在像素级分割识别方面的不足，提出了全卷积神经网络(FCN)。自FCN被提出后图像分割领域迅速发展，常用的图像分割方法包括FCN，SegNet[9]，PSPNet[10]，DeepLab[11]，U-Net[12]等，这些普遍使用的分割方法在裂缝检测方面也得到了发展，Yang等[13]引入FCN来同时解决检测和测量裂缝问题，但是测量误差较大；Ren等[14]提出了基于SegNet的裂缝检测网络CrackSegNet，但是处理效率仍不够理想。由于像素级的裂缝检测和一般图像分割有较大差别，因此一般的图像分割模型直接应用在裂缝检测时，很难达到理想的效果。随着自然语言处理技术的发展，注意力机制逐渐被应用于裂缝检测来提升网络的抗噪能力。Zhou等[15]利用空间和通道注意力机制来捕获裂缝的空间结构信息和高级上下文特征，提升了检测精度。Qiao等[16]引入scSE注意力机制模块来增强重要信息特征，同时抑制空间和通道中不重要的信息特征，有效地提高了模型的检测效率，降低了计算成本。有研究将U-Net应用于裂缝分割[17]，并取得了良好的检测效果，U-Net是编码器-解码器结构的全卷积神经网络模型，其网络结构能适用于细微裂缝检测，能实现任意分辨率的检测。以往研究中的模型也存在一些问题[18]，自制裂缝分割数据集困难，一般采用少量图像训练，泛化能力弱；在更复杂背景下的裂缝检测有待进一步提升。此外，对于形态多样且具有线性特征的裂缝图像，普通卷积由于尺寸形态固定，其激活单元的感受野也相对固定，减弱了裂缝特征的表征能力。因此，如何根据检测任务的难度来选择适当深度的模型是另一个挑战。

因此，本文提出了一种基于注意力机制和可变形卷积的路面裂缝检测网络，基于编码-解码结构进行网络构建。编码阶段通过可变形卷积改进网络VGG16作为特征提取网络；使用密集连接机制强化网络；在解码阶段使用多级特征融合模块融合语义信息和细节信息；最后结合注意力模块，提升检测精度。本文提出的网络，更加具有泛化性，并能实现对较复杂背景下的路面裂缝的有效语义分割。

2 相关知识

2.1 转置卷积

转置卷积(Transposed Convolution)，又称反卷积[19]，与卷积操作类似。通过在输入特征边缘填充大量零来实现输出高宽大于输入高宽的效果，从而实现向上采样的目的。在解码阶段需要将特征图逐步还原到原图的分辨率，这时需要对其进行上采样，转置卷积可以引入参数让网络自动学习卷积核的权重以更好地恢复空间分辨率，因此，我们用转置卷积来替代常规的上采样操作(最近领插值、双线性插值等)。

2.2 Dropout层

深层神经网络训练出来的模型很容易产生过拟合的现象，为了防止过拟合发生，我们在网络中添加Dropout层[20]，随机断开一定概率的参数，提升模型的泛化能力，如图1所示。其中图1(a)为标准全连接层，图1(b)为添加Dropout的全连接网络。

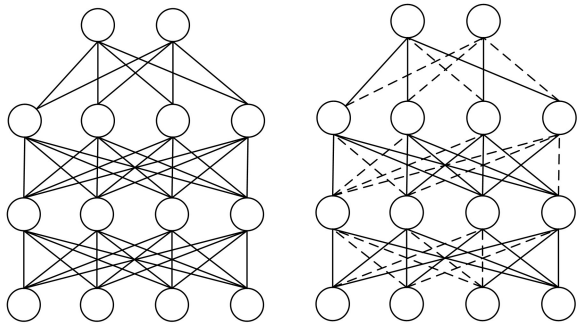

图1 Dropout示意图

Fig. 1 Schematic diagram of Dropout

2.3 Dense Block结构

DenseNet提出了一种密集连接机制构成的Dense Block结构，其作用是增强特征传播，特征能够重用，强化高层特征对细节特征的获取。图2为Dens Block结构图，其中密集连接机制为相互连接所有层，每一层的输入来自前面所有层的输出。本文受Dense Block结构中的密集连接机制启发，设计了特征强化网络，进一步加强编码阶段各阶段特征信息的利用率。

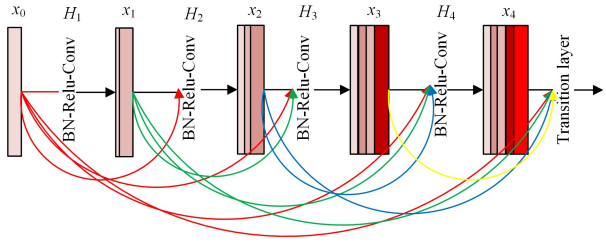

图2 Dens Block结构图

Fig. 2 Diagram of Dens Block structure

本文用x_0表示输入层，x_l表示其他层的输出，其中l是层数，则各层的输出可定义为：

$$x_l = H_l([x_0, x_1, \cdots, x_{l-1}]) \tag{1}$$

其中，$H_l(\cdot)$代表非线性转化函数，它是一个组合操作，其包括一系列的BN，ReLU及Conv操作，这里l层与$l-1$层之间实际上包含多个卷积层。

3 基于注意力机制和可变形卷积的裂缝检测模型

本文提出的网络模型包含了5个模块，分别是特征编码器模块(Feature Encoder Module)、特征强化模块(Feature Enhancement Module)、特征解码器模块(Feature Decoder Module)、特征融合模块(Feature Fusion Module)和注意力模块(Attention Module)，其详细网络结构如图3所示。所提出

的网络采用编码-解码结构。首先，将大小为 256×256 的原始彩色图像输入进特征编码器模块，由可变形卷积学习不同形状的裂缝特征；其次，特征编码器模块各阶段提取到的特征送入特征强化模块，加强各阶段特征信息的利用率；然后，将加强后的特征经过 Dropout 层以桥接形式和上采样得到的特征进行拼接，对解码器模块各阶段的特征进行融合，细化裂缝分割结果；最后，将融合的特征输入到注意力模块，抑制背景特征，提高检测精度。

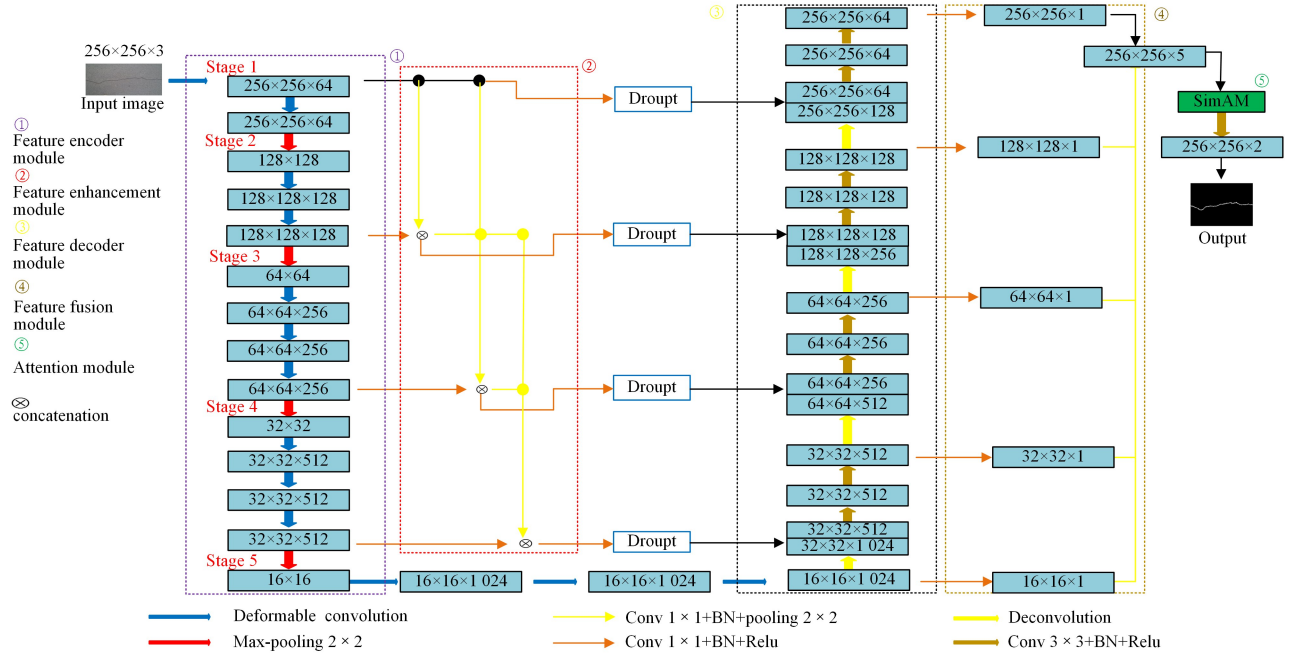

图 3　网络模型结构图

Fig. 3　Diagram of network model structure

3.1　编码器模块

为了更好地学习不同形状裂缝特征信息，本文基于 VGG16 结构作为特征提取网络，将 VGG16 所有卷积层替换为可变形卷积，为使网络得到充分卷积，增加可变形卷积层数，最终构成编码器模块。

裂缝检测的难点在于裂缝具有形状多样的几何线性结构特征，现有的卷积神经网络模型对于物体几何形变的适应能力几乎完全来自数据本身所具有的多样性，其模型内部并不具有适应几何形变的机制。因为卷积操作本身具有固定的几何结构，而由其层叠搭建而成的卷积网络的几何结构也是固定的，所以不具有对于几何形变建模的能力。针对此问题，本文引入了具有空间几何形变能力的可变形卷积[21]。图 4 给出了标准卷积与可形变卷积提取裂缝特征的方式。

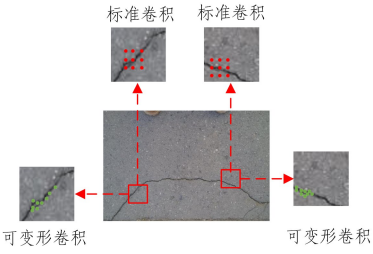

图 4　3×3 标准卷积和可形变卷积示意图

Fig. 4　Schematic diagram of 3×3 standard convolution and deformable convolution

可变形卷积是利用额外的卷积层去学习相应的偏移量，将得到的偏移量叠加到输入特征图中相应位置的像素点中，让卷积核在输入特征图能够发散采样，从而使网络能够聚焦目标函数。

卷积核 R 定义了感受野的大小和膨胀率。此处定义了一个 3×3 大小、膨胀率为 1 的卷积核 R，$R=\{(-1,-1),(-1,0),\cdots,(0,1),(1,1)\}$。对于标准卷积，输入特征图 x 上 p_0 的输出特征映射 y 可以定义为：

$$y(p_0)=\sum_{p_n\in R}\omega(p_n)\cdot x(p_0+p_n) \tag{2}$$

其中，p_n 为 R 中的点，$\omega(*)$为采样点权重。

可变形卷积是在标准卷积采样点位置 R 上增加偏移量 Δp_n，$\{\Delta p_n \mid n=1,2,\cdots,N\}$，$N=|G|$。因此，可变形卷积定义为：

$$y(p_0)=\sum_{p_n\in R}\omega(p_n)\cdot x(p_0+p_n+\Delta p_n) \tag{3}$$

可变形卷积提取裂缝特征的过程如图 5 所示，为学习偏移量 Δp_n，对特征图施加一个卷积层。在训练期间，用于生成输出特征的卷积核和生成偏移量的卷积核是同步学习的。其中偏移量的学习是利用插值算法，通过反向传播进行学习。

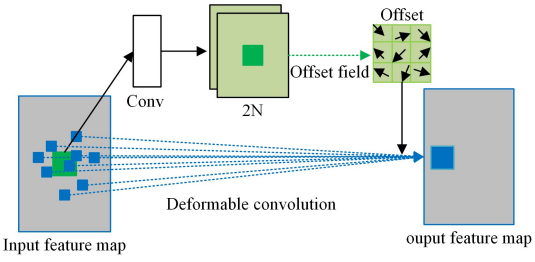

图 5　可变形卷积特征的提取过程

Fig. 5　Extraction process of deformable convolution feature

3.2　特征强化模块

为了提高编码器模块各阶段特征信息的利用率，受密集连接机制的启发设计特征强化模块，即图 3 中的模块 2，详细结构如图 6 所示。将编码器模块 stage1-stage4 提取到的特征图输入强化网络，浅层特征经过 1×1 Conv 以及下采样

和深层特征经过 1×1 Conv 进行特征融合，即强化网络各阶段输出为：融合自身阶段特征层与前面阶段所有特征层的输出。因此，编码器模块的特征经过强化网络后特征信息能够有效利用。

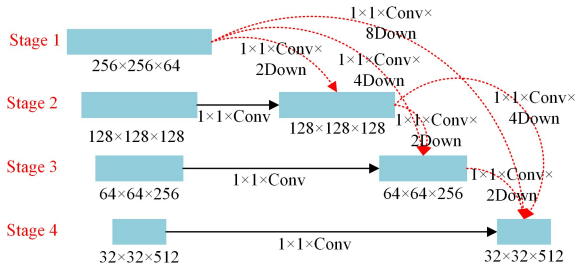

图 6 特征强化模块的结构图

Fig. 6 Structure diagram of feature enhancement module

3.3 特征融合模块

一般语义分割算法使用网络最后一层的特征图进行结果预测，可能导致语义信息不完全利用。在还原图像分辨率的过程中，不同特征层的语义信息是不一样的。为使不同层的信息进一步得到有效利用，设计多级特征融合模块，如图 3 所示。对于解码器模块得到的 16×16，32×32，64×64，128×128，256×256 大小的 5 个特征图分支，先分别进行 1×1 卷积压缩特征，再分别通过转置卷积变为 256×256 大小的特征图，进一步对这 5 个特征图进行堆叠。

3.4 注意力模块

复杂背景裂缝检测，容易出现遗漏检测（细节裂缝）、错误检测（背景噪声）。为了突出目标特征，抑制背景噪声，本文加入了最新注意力模块 SimAM[22]。与现有通道、空间注意力模块不同，它是一种无参数注意力机制，添加到网络中不增加网络参数，能为特征图推导 3D 注意力权重。图 7 给出了注意力模块 SimAM 的原理。

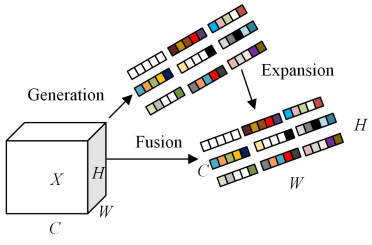

图 7 注意力模块 SimAM 的原理图

Fig. 7 Schematic diagram of attention module SimAM

通过定义一个简单的能量函数，度量神经元之间的线性可分性，计算最小能量的计算式如式(4)所示：

$$e(t)=\frac{4(\sigma^2+\lambda)}{(t-\mu)^2+2\sigma^2+2\lambda} \tag{4}$$

其中，$\mu=\frac{1}{M}\sum_{i=1}^{M}x_i$，$\sigma^2=\frac{1}{M}\sum_{i=1}^{M}(x_i-\mu)^2$。

上述公式表明，能量 $e(t)$ 越小，取值为 t 的神经元与周围神经元的区别越大，对视觉处理更重要。因此，每个神经元的重要性可以通过 $\frac{1}{e(t)}$ 来衡量。

按照注意力机制的定义，对特征进行增强处理，得到输出与输入特征的关系，如式(5)所示：

$$\tilde{x}_i=S\left(\frac{1}{e(x_i)}\right)x_i,\ i=1,2,\cdots,M \tag{5}$$

其中，$S(x)$ 表示 Sigmoid 激活函数，$X=\{x_1,x_2,\cdots,x_M\}$ 为输入特征，$\tilde{X}=\{\tilde{x}_1,\tilde{x}_2,\cdots,\tilde{x}_M\}$ 为输出特征。

3.5 损失函数

损失函数设计的好坏对网络性能的影响很大，以往大多数分割任务使用交叉熵作为损失函数[23]，式(6)为二分类问题的交叉熵损失函数。

$$L_{CE}=-\frac{1}{N}\sum_{i=1}^{N}[y_i\log\hat{y}_i+(1-y_i)\log(1-\hat{y}_i)] \tag{6}$$

其中，N 为像素数，y_i 为第 i 个像素点的标记值，$\hat{y}_i$ 为第 i 个像素点的预测结果。

交叉熵损失函数侧重于确定每个像素的类别，且对每个类别的关注度相同，因此容易受到类别不平衡的影响。对于裂缝图像来说，目标裂缝占比小，使用交叉熵损失函数，网络更偏向于背景预测。为解决类别不平衡带来的问题，我们引入骰子损失函数（Dice Loss）[24]，一种用于评估两个样本之间相似性度量的函数，取值范围为 0～1，值越大表示两个样本的相似度越高，骰子损失函数（二分类）的表达式如下：

$$L_{DL}=1-\frac{2\sum_{i=1}^{N}y_i\cdot\hat{y}_i+\varepsilon}{\sum_{i=1}^{N}y_i^2+\sum_{i=1}^{N}\hat{y}_i^2+\varepsilon} \tag{7}$$

其中，参数 ε 设置为 1，防止分母为 0 情况，减少过拟合，避免更快收敛。

骰子损失函数有利于正负样本不平衡的情况，侧重于对前景的挖掘，在训练过程中，小目标较多的情况下波动大，极端情况下会出现梯度饱和的情况，训练困难。

综合考虑交叉熵损失函数与骰子损失函数的特点，针对样本不平衡问题，对两种损失函数进行加权求和，得到新的损失函数，这样新的损失函数既能关注像素级别和图像级别的显著性，又能解决正负样本不平衡问题，如式(8)所示：

$$L=\alpha L_{CE}+(1-\alpha)L_{DL} \tag{8}$$

其中，L 为本文损失函数，$\alpha\in[0,1]$ 为权重系数，本文中 α 设为 0.15。

4 实验结果与分析

本文基于 Python 下的 Tensorflow2 深度学习框架设计实验。模型训练与测试配置了 Windows 系统、Intel Xeon CPU E5-2680 v2 2.8GHz 8 核处理器、GeForce RTX 2070 SUPER 显卡。

4.1 数据集

(1)DBCCrack。DBCCrack 数据集是由 Li 等[6]收集的公共裂缝数据集，包含 2068 张 1024×1024 尺寸的 RGB 图像，裂缝特征比较明显。用 labelme 软件标注得到相应的标签图像，以 256 为步长将其切割为 256×256 尺寸的图像，共 33088 张，从中随机筛选出 11000 张含裂缝的图像，按 8∶1∶1 划分为训练集、验证集、测试集，最终获得 8800 张训练集图像和1100 张验证集图像以及 1100 张测试集图像。

(2)CFD。CFD 数据集是 Shi 等[25]提出的一个混凝土路面裂缝数据集，包含 118 张大小为 480×320 尺寸的 RGB 图像，图像裂缝较细小，包含背景噪声。用 labelme 软件标注得到相应的标签图像，由于数据集太小，因此进行数据增强，包括对图像做旋转、镜像操作，最终获得 1000 张图像，按 8∶1∶1 划分为训练集、验证集、测试集，最终获得 800 张训练集图像，100 张验证集图像，100 张测试集图像。由于图像在输入网络前会被统一调整为 256×256 的尺寸，因此不需要对 CFD

数据集的尺寸进行裁剪。

4.2　性能评价指标

为评估本文模型的性能，本文选择了 3 个评价指标[26]，即像素精度（PA）、平均像素精度（MPA）、平均交并比（MIoU），其中 PA 表示预测正确的像素点占总像素的比例，如式(9)所示：

$$PA=\frac{\sum_{i=0}^{m}p_{ii}}{\sum_{i=0}^{m}\sum_{j=0}^{m}p_{ij}} \tag{9}$$

MPA 表示计算每个类内被预测正确分类像素数的比例，然后对所有类求平均，如式(10)所示：

$$MPA=\frac{1}{m+1}\sum_{i=0}^{m}\frac{p_{ii}}{\sum_{j=0}^{m}p_{ij}} \tag{10}$$

MIoU 是图像分割问题的标准评估度量，表示真实值与预测值两个集和的交集和并集的重叠比，如式(11)所示：

$$MIoU=\frac{1}{m+1}\sum_{i=0}^{m}\frac{p_{ii}}{\sum_{j=0}^{m}p_{ij}+\sum_{j=0}^{m}p_{ji}-p_{ii}} \tag{11}$$

式(9)－式(11)中，m 表示类别，背景为附加类别，p_{ii} 表示预测正确分类的像素数量，p_{ij} 表示属于 i 类但预测为 j 类的像素数数量。

4.3　实验参数设置

本文算法在训练过程中采用 eager 模式进行训练，冻结训练，总共迭代 200 个轮回，前 100 个轮回采用冻结训练加快训练速度，也防止在训练初期权值被破坏，Batch size 设置为 16，初始学习率设为 1×10^{-4}；后 100 个轮回不采用冻结训练，Batch size 设置为 8，初始学习率设为 1×10^{-5}。两阶段都采用 Adam 优化器优化网络模型，网络激活函数采用 ReLu。

4.4　实验结果与分析

为验证本文算法的优越性，将本文算法同 U-Net，SegNet，Deeplabv3，RCF 算法进行性能比较，最终实验结果为多次实验的平均结果。

(1)在 DBCCrack 数据集上的实验。本文算法和 4 种对比算法经过训练，从测试集中挑选不同类型的裂缝图像进行识别效果比较，如图 8 所示。SegNet，Deeplabv3，RCF，U-Net 算法的预测结果是不连续的，会出现断点情况，本文算法能较好地去除噪声并较完整且精确地识别出裂缝。

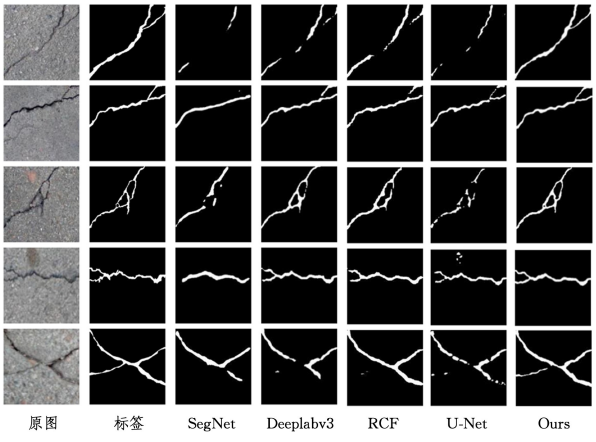

图 8　DBCCrack 数据集上的定性比较结果

Fig. 8　Qualitative comparison results on DBCCrack dataset

表 1 列出了这些算法的 PA，MPA，MIoU。由表 1 可知，本文算法的各项指标都高于对比算法，其 MPA 达到了 92.12%。因此，本文算法具有一定的优越性。

表 1　不同算法在 DBCCrack 数据集上的性能对比

Table 1　Performance comparison of different algorithms on DBCCrack dataset

（单位：%）

算法	PA	MPA	MIoU
SegNet	69.58	84.24	75.51
Deeplabv3	78.92	89.20	81.29
RCF	81.33	90.19	81.31
U-Net	72.30	85.85	80.11
Ours	85.01	92.12	84.79

(2)在 CFD 数据集上的实验。在 DBCCrack 数据集上已验证本文算法的优越性，为进一步验证本文算法的优越性以及稳定性，我们在相对复杂的 CFD 数据集上进行性能对比实验。图 9 给出了不同类型的裂缝图像测试结果，可以看出本文算法在 CFD 数据集上的识别效果优于其他方法。当裂缝图像背景相对简单、裂缝与背景区域有显著差异时，本文算法与对比算法都能较好地提取裂缝，SegNet 算法的识别效果稍差（见图 9 中的第 1 行与第 2 行）。当裂缝图像受到背景因素影响时，对比算法预测结果出现多处误检与漏检情况（见图 9 中的第 3 行与第 4 行）。当裂缝图像为块状裂缝时，对比算法预测结果是不完整的（见图 9 中的第 5 行与第 6 行）。由本文算法在不同类型裂缝图像的检测结果可以看出，本文算法能完整并较好地识别出图像中的裂缝。

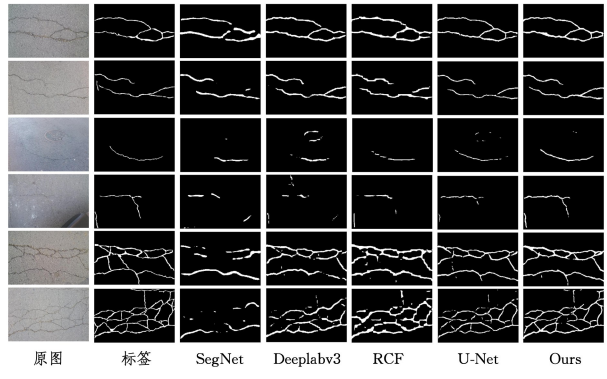

图 9　CFD 数据集上的定性比较结果

Fig. 9　Qualitative comparison results on CFD dataset

表 2 列出了各算法的评价指标结果。由表 2 可知，本文算法各项指标依旧高于其他算法，且具有更高的检测精度，在两个数据集上 PA 和 MPA 的实验结果接近，数据集的复杂程度对本文算法的影响较小，证明了本文算法的优越性和稳定性。

表 2　不同算法在 CFD 数据集上的性能对比

Table 2　Performance comparison of different algorithm on CFD dataset

（单位：%）

算法	PA	MPA	MIoU
SegNet	47.67	73.32	63.32
Deeplabv3	60.88	79.99	68.49
RCF	64.95	81.96	68.75
U-Net	63.77	81.67	74.53
Ours	83.02	91.02	74.75

结束语　针对复杂背景裂缝检测困难的问题，本文提出了一种基于注意力机制和可变形卷积的路面裂缝检测算法，以编码-解码结构进行构建。用可变形卷积学习形态多样的

裂缝几何线性特征；使用密集连接机制强化特征信息；采用转置卷积和桥接方式与编码阶段特征逐步融合，并结合多级特征融合的思想，融合裂缝的语义信息和细节信息；引入注意力模块，抑制背景特征，提高检测精度。此外，结合交叉熵损失和骰子损失函数解决裂缝像素类别不平衡的问题。实验结果表明，本文算法具有较强的稳定性、鲁棒性和优越性，可以解决较复杂背景干扰的裂缝检测问题，满足一定的工程检测需求。但由于实际场景中裂缝类型、粗细、光照及背景的复杂程度具有很强的不确定性，要完整、无误地提取出裂缝，还需要做深入地研究。在后续工作中可以考虑结合目标检测算法构建一个端到端的裂缝检测，在简化分割网络的参数量的同时也避免了背景复杂度的影响。

参 考 文 献

[1] XU H, SU X, WANG Y, et al. Automatic bridge crack detection using a convolutional neural network[J]. Applied Sciences, 2019, 9(14): 2867.

[2] CHENG Y, TIAN L, YIN C, et al. A magnetic domain spots filtering method with self-adapting threshold value selecting for crack detection based on the MOI[J]. Nonlinear Dynamics, 2016, 86(2): 741-750.

[3] XU L L, ZHU X P, HOU Y X, et al. Culvert crack defect segmentation algorithm based on enhanced tone features[J]. Progress in Laser and Optoelectronics, 2020, 57(8): 081016.

[4] HOANG N D, NGUYEN Q L, TRAN V D. Automatic recognition of asphalt pavement cracks usingmetaheuristic optimized edge detection algorithms and convolution neural network[J]. Automation in Construction, 2018, 94: 203-213.

[5] CHA Y J, CHOI W, BUYUKOZTURK O. Deep learning-based crack damage detection using convolutional neural networks[J]. Computer-Aided Civil and Infrastructure Engineering, 2017, 32(5): 361-378.

[6] LI F L, Ma W F, LI L, et al. Research on Bridge Crack Detection Algorithm Based on Deep Learning[J]. Acta Automatica Sinica, 2019, 45(9): 1727-1742.

[7] ZHANG Y, HUANG J, CAI F. On Bridge Surface Crack Detection Based on an Improved YOLO v3Algorithm[J]. IFAC-PapersOnLine, 2020, 53(2): 8205-8210.

[8] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. nature, 2015, 521(7553): 436-444.

[9] 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.

[10] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 6230-6239.

[11] 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, 2017, 40(4): 834-848.

[12] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]// International Conference on Medical Image Computing and Computer-assisted Intervention. 2015: 234-241.

[13] YANG X, LI H, YU Y, 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.

[14] REN Y, HUANG J, HONG Z, et al. Image-based concrete crack detection in tunnels using deep fully convolutional networks[J]. Construction and Building Materials, 2020, 234: 117367.

[15] ZHOU Q, QU Z, CAO C. Mixed pooling and richer attention feature fusion for crack detection[J]. Pattern Recognition Letters, 2021, 145: 96-102.

[16] QIAO W, LIU Q, WU X, et al. Automatic pixel-level pavement crack recognition using a deep feature aggregation segmentation network with a scSE attention mechanism module[J]. Sensors, 2021, 21(9): 2902.

[17] LIU J, YANG X, LAU S, et al. Automated pavement crack detection and segmentation based on two-step convolutional neural network[J]. Computer-Aided Civil and Infrastructure Engineering, 2020, 35(11): 1291-1305.

[18] ZHANG L X, SHEN J K, ZHU B J. A research on an improved Unet-based concrete crack detection algorithm[J]. Structural Health Monitoring, 2021, 20(4): 1864-1879.

[19] GAO H, YUAN H, WANG Z, et al. Pixel transposed convolutional networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 42(5): 1218-1227.

[20] CAI S, SHU Y, CHEN G, et al. Effective and efficient dropout for deep convolutional neural networks[J]. arXiv, 2019, 1904.03392.

[21] LIU Y, WANG W, Li Q, et al. DCNet: A deformable convolutional cloud detection network for remote sensing imagery[J]. IEEE Geoscience and Remote Sensing Letters, 2021, 19(1): 1-5.

[22] YANG L, ZHANG R Y, LI L, et al. Simam: A simple, parameter-free attention module for convolutional neural networks[C]// Proceedings of Machine Learning Research. 2021: 11863-11874.

[23] LI G, MA B, HE S, et al. Automatic tunnel crack detection based on u-net and a convolutional neural network with alternately updated clique[J]. Sensors, 2020, 20(3): 717.

[24] EELBODE T, BERTELS J, BERMAN M, et al. Optimization for medical image segmentation: theory and practice when evaluating with Dice score or Jaccard index[J]. IEEE Transactions on Medical Imaging, 2020, 39(11): 3679-3690.

[25] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2020: 318-327.

[26] FU H X, MENG D, LI W H, et al. Bridge Crack Semantic Segmentation Based on Improved Deeplabv3+[J]. Journal of Marine Science and Engineering, 2021, 9(6): 671.

LONG Tao, born in 1997, postgraduate. His main research interests include computer vision and crack detection based on deep learning.