中图法分类号:TP391; U457　　**文献标识码**: A　　**文章编号**: 1006-8961(2025)02-0467-18

论文引用格式:Huang Z H, Luo H T and Guo B. 2025. Detecting the defects of bridge cables and tunnel lining via integrating attention and enhanced receptive field. Journal of Image and Graphics, 30(02):0467-0484(黄志海, 罗海涛, 郭波. 2025. 融合注意力及增强感受野的桥隧表面病害深度网络检测. 中国图象图形学报, 30(02):0467-0484)[DOI:10.11834/jig.240191]

融合注意力及增强感受野的桥隧表面病害深度网络检测

黄志海[1],罗海涛[2],郭波[1*]

1. 广东工业大学土木与交通工程学院,广州　510006; 2. 广州地铁设计研究院股份有限公司,广州　510030

摘　要: **目的**　在桥梁缆索和隧道环境中进行表面病害检测面临多重挑战:人工病害定位在复杂结构中的高投入成本;隧道内存在光照不足、病害区域的表面特征复杂、范围广阔和易遭遮挡等问题,这些因素使得传统图像处理技术在病害区域检测上表现出较低的抗干扰能力、识别精度不足和分割效果欠佳,鉴于此,提出一种基于融合注意力及增强感受野的深度网络模型。**方法**　该模型使用融合Transformer注意力机制的骨干网络提取目标特征信息,获得更为密切联系的全局特征表示,以解决光照不足导致局部特征缺少的问题;引入空间序列缩减法降低骨干网络的参数量;改进使用具有串并联关系的空洞卷积池化金字塔(series-parallel atrous convolutional pyramid, SACP)模块,使得卷积感受野进一步扩大并且彼此融合,更好地感知完整的病害范围;解码阶段融合卷积注意力模块(concentration-based attention module, CBAM),提高浅层特征的有效边界特征权重。改进损失函数加快模型收敛速度。**结果**　本文实地采集隧道内衬以及桥梁缆索病害部位图像构建数据集展开实验,结果表明:本文模型在隧道内衬病害提取上准确率(accuracy, Acc)达到94.4%,平均交并比(mean intersection over union, mIoU)达到78.14%,F1分数(F1-score)达到76.45%。在桥梁缆索病害提取上Acc达到97.15%,mIoU达到80.41%,F1分数达到77.92%。**结论**　相较于主流的分割网络,本文模型在桥隧表面病害提取的精度上均有提升,具有更优秀的提取效果和抗干扰能力,能更好地满足复杂环境下病害检测工程需求。

关键词: 深度学习;注意力机制;卷积感受野;图像分析;自动化病害检测;工程表面病害识别

Detecting the defects of bridge cables and tunnel lining via integrating attention and enhanced receptive field

Huang Zhihai[1], Luo Haitao[2], Guo Bo[1*]

1. School of Civil and Transportation Engineering, Guangdong University of Technology, Guangzhou　510006, China;

2. Guangzhou Metro Design & Research Institute Co., Ltd., Guangzhou　510030, China

Abstract: **Objective** Using computer image technology to identify defects in bridge cables and tunnel linings is an efficient and convenient method. Surface defect detection on bridge cables and tunnel linings faces multiple challenges: high costs of manual defect localization in complex structures. Insufficient tunnel lighting, complex defect features, and imbalanced target-background ratios often present apparent shortcomings in traditional image technology for defect area detection, such as poor anti-interference ability, low recognition rates, and subpar segmentation effects. Therefore, this study proposes a

收稿日期:2024-04-07;**修回日期**:2024-06-17;**预印本日期**:2024-06-23

*通信作者:郭波　guobo@gdut.edu.cn

基金项目:国家自然科学基金项目(U21A20139)

Supported by: National Natural Science Foundation of China(U21A20139)

deep network model based on the fusion attention and enhanced receptive field. **Method** The model employs a backbone network integrating a fused Transformer to extract target feature information and establish a more closely related global texture feature representation, which solves the problem of a lack of local features caused by insufficient lighting. Additionally, spatial reduction attention is introduced to reduce the parameter count of the backbone. The series-parallel atrous convolutional pyramid (SACP) module was introduced to further expand the convolutional receptive field and integrate it, thereby enhancing the perception of the complete defect scope with multiscale characteristics. The decoder uses a concentration-based attention module(CBAM) to strengthen the effective boundary feature weights of shallow features and suppress some occluded features, which sharpens the segmentation results. The model employs a composite loss function that combines cross-entropy loss and Dice loss to balance the contributions of positive and negative samples. Datasets for experimentation were constructed via the onsite collection of tunnel lining defect images and bridge cable climbing robot-acquired cable defect images. **Result** The results demonstrate the following: 1) the proposed multitransformer backbone reduces computational complexity while maintaining feature extraction capabilities. 2) The optimized SACP module improves the segmentation accuracy by 2%, and the use of depthwise separable convolution effectively reduces the complexity. 3) Owing to the use of an improved composite loss function, the convergence speed of the network has accelerated. 4) The proposed network model achieves an accuracy of 94.4%, a mean intersection over union of 78.14%, and an F1 score of 76.45% for the tunnel lining dataset. The generalization ability of the proposed model was also demonstrated in the bridge cable dataset. **Conclusion** Compared with mainstream segmentation networks, the proposed model improves the accuracy of bridge cable and tunnel lining defect detection, demonstrating enhanced extraction efficacy and anti-interference capability. Therefore, this model can better meet the engineering requirements for defect detection in complex environments.

Key words: deep learning; attention mechanism; convolution receptive field; image analysis; automated defect detection; structure surface defect recognition

0 引 言

随着我国交通事业的高速发展,桥隧交通工程作为交通基础设施的重要组成部分,在促进区域经济发展和社会互联互通方面发挥着不可替代的作用。然而,随着桥隧的不断投入使用和使用年限的增长,在使用过程中可能受到各种因素的影响,例如地质变化、交通负荷和自然灾害等,其病害问题逐渐显露出来。这些病害包括但不限于开裂、剥落、渗水甚至断裂等,尤其是在桥梁缆索以及隧道内衬这些受力关键部位,严重影响了桥隧的结构完整性和使用安全性(张劲泉 等,2023;洪江华,2023)。因此,及时准确地检测和定位桥隧表面病害,对维护基础交通设施的健康状态、延长使用寿命以及提高交通运输的安全性至关重要(彭孟开,2016)。

传统桥隧病害检测提取方法主要包括人工巡检和仪器检测。人工巡检易受人为主观因素的影响,仪器检测往往需要特种设备和专业技术人员进行操作,应用面不广(邓普,2015;李丹乐,2018)。近年来,图像设备已普遍用于快速、高精度的大场景数据采集,例如隧道图像采集系统和爬索机器人。利用采集的环境图像,基于边缘检测算法、形态学算法、区域生长算法等常规算法无法很好地应对光线、角度和拍摄距离等因素变化带来的挑战,泛化能力差,具有一定的盲目性(龚锐 等,2020;徐蔚波 等,2017;刁智华 等,2010)。同时,机器学习技术的快速发展,可以在大量数据的基础上进行自主学习和提取相关高阶特征,具有识别速度快、分割准确度高和鲁棒性高等优点,可以实现优秀的目标检测提取(孙刘杰等,2022)。早期机器学习算法如K均值聚类(K-means)、随机森林(random forest)以及支持向量机(support vector machine)等有效提高了图像自动病害检测的可靠程度(Shi等,2016;Li等,2017),但受限于算力,较低参数量使得模型鲁棒性不高,需要理想成像条件和较多预处理(许霄煜,2022)。

伴随计算硬件发展,同时受生物视觉系统的启发,Shelhamer等人(2017)将卷积神经网络(convolutional neural network, CNN)的全连接层替换为卷积层,首次提出全卷积网络(fully convolutional network, FCN),从而实现像素级的语义分割任务。随后卷积网络在医学及自然领域取得大量突破(Ronneberger等,2015;Badrinarayanan等,2017;Zhao等,2017)。

Bhowmick等人(2020)将U-Net改进为彩色图像

分割运用在桥梁裂缝检测并量化计算，但由于训练样本不足，分割效果也有待优化。Liang（2019）基于贝叶斯优化超参数，基于3层图像进行桥梁局部损伤级损伤定位及评估。佃松宜等人（2022）在U-Net基础上加入注意力补偿模块得到TCS-Net，可以更有效地提升目标关注权重，更好地探测壳上的细小裂缝，但对于细长特征外的检测效果无法得到保证。董绍江等人（2024）基于YOLO（you only look once）网络模型改进得到桥梁表观病害检测方法，引入轻量级注意力机制并进行多头预测，但是缺少像素级病害指示。王宝坤等人（2023）改进具有残差单元的注意力编码器，使用多达4 500幅影像构建盾构隧道病害数据集，使得识别精度相比传统模型有明显提高。这些卷积网络模型专注于局部特征信息获取，忽略了在全局下更加丰富的特征潜力，往往导致不一致的感知重点，出现零星分割问题。

注意力机制一直在推动各种识别任务的最新发展，其通过使用非归纳偏差和全局范围内对远程关系进行建模来增强特征的学习能力。继Vaswani等人（2023）提出的Transformer在自然语言处理（natural language processing，NLP）领域取得成功后，ViT（vision Transformer）（Dosovitskiy 等，2021）、Detr（Carion 等，2020）以及Swin Transformer（Liu 等，2021）网络模型相继提出，取得令人瞩目的目标检测精度表现，在病害检测方面同样受益。Fang等人（2022）基于Transformer构建编解码器框架TransU-Net，消除了阴影、噪声等负面因素的干扰。Zhou等人（2022）在骨干网络中加入挤压—激发网络（squeeze-and-excitation network，SENet）注意力模块，在复杂的隧道背景和光照条件下有效区分病害目标。周中等人（2024）使用Transformer以及卷积双分支进行特征提取，实现病害全局特征和局部细节特征互补，但同时带来巨大的计算量。

针对桥隧表面病害尺度跨度大、环境光线对比度高、可能存在复杂管线遮挡情况的问题，现存的神经网络还无法高效提取上下文多尺度信息，完成对目标病害的整体感知。本文在编码器—解码器网络框架上，设计了一个具有针对性的可靠桥隧表面病害区域检测的深度网络。

本文主要贡献如下：1）针对复杂度较高且注重局部特征感知的卷积骨干网络，融合使用无位置编码、具有多尺度输出骨干网络Multi-Transformer。通过提取目标上下文信息来提高在光线不足场景下的病害特征表示能力。其中空间序列缩减法降低Transformer编码器复杂度。在浅层特征输出中使用卷积注意力提升空间及通道重要特征权重，从而突出目标边界。2）针对空洞卷积金字塔结构中感受野相对固定的问题进行优化，增加部分分支，并融合各空洞卷积分支结果以进一步扩大感受野，改进得到具有串并联关系的空洞卷积池化金字塔（series-parallel atrous convolutional pyramid，SACP）结构，提高多尺度下目标区域完整的感知能力。3）针对以上改进方式开展实验，证明本文提出模型在面对病害区域检测任务的有效性。

1　基于融合注意力及增强感受野的语义分割网络

1.1　编码器—解码器结构分割网络

经典的编码器—解码器分段结构可以充分细化目标边缘信息，模型结构较适合于目标分割任务。DeepLabv3+在Pascal VOC（pattern analysis，statistical modeling and computational learning visual object classes）数据集上取得了良好的语义分割成绩（Chen等，2018），但较深架构堆叠许多局部卷积层容易引起归纳偏差以及模型退化，并且伴随网络的深度增加，计算量也急剧增大，对保留原始特征产生不利影响（He等，2016）。与此同时，网络架构存在着感受野相对固定、无法全面适应目标区域大小、边缘信息未充分利用以及容易出现零星特征分割块的问题。针对以上情况，本文选择基于编码器—解码器网络框架改进，得到更适用于复杂场景的分割模型。

1.2　融合注意力及增强感受野的网络模型

基于融合注意力及增强感受野的网络模型整体框架如图1所示。在网络编码阶段，基于全局的注意力机制可以很好地避免卷积出现的归纳偏差并且可以弥补全局缺失的上下文联系以加强信息交流（Xie等，2021），获得更为完整的全局信息，因此本文设计的网络融合Transformer构建骨干网络（Multi-Transformer），其具有多尺度不同精细的特征输出，详见1.2.1节。多层感知器（multi-layer perceptron，MLP）模块进行线性映射及拼接，以聚合多尺度信息，加强提取图像特征的能力，详见1.2.2节。同时，在空洞卷积金字塔结构中提出具有串并联的

SACP框架，添加额外空洞卷积分支并彼此连接，可以融合多尺度感受野提取的特征，提升病害区域完整感知能力，详见1.2.3节。

在网络解码阶段，卷积注意力提升输入解码器的有效目标特征的权重，抑制噪声，以应对存在遮挡的病害情况，详见1.2.4节。其输出与来自SACP的输出拼接进行融合，再经过3×3卷积以及4倍双线性上采样，最终得到分割输出。

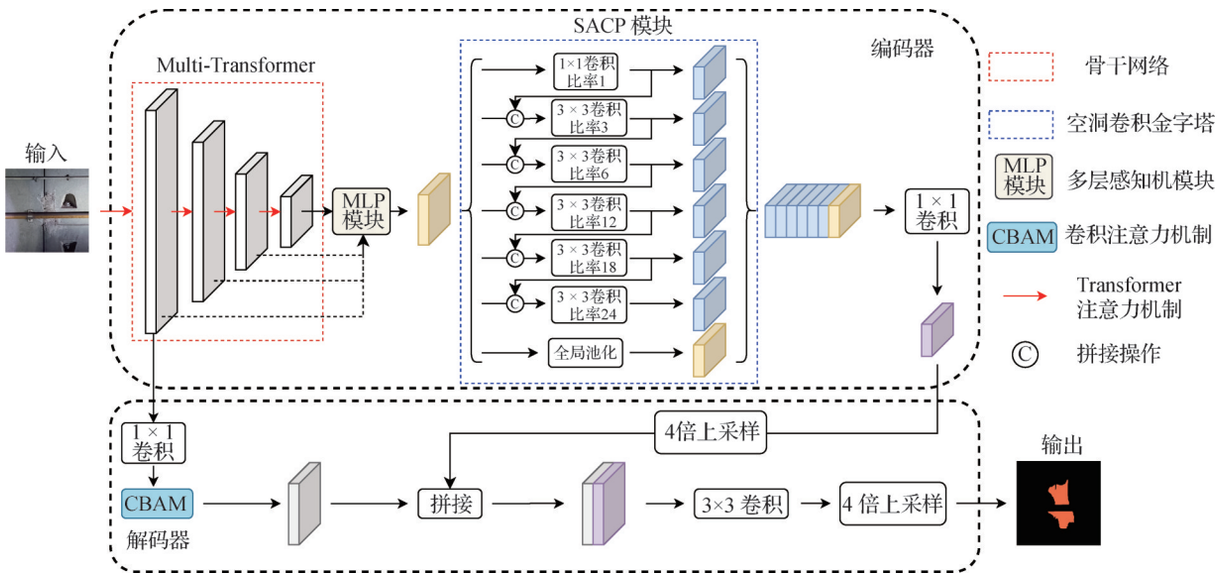


图1　基于融合注意力及增强感受野的网络模型

Fig. 1　The model based on attention and enhanced receptive field

1.2.1　基于注意力的特征提取骨干网络

由于CNN网络深且宽的架构堆叠，许多局部卷积层容易引起归纳偏差以及退化，自ViT(Dosovitskiy等，2021)模型诞生以来，基于全局的注意力机制Transformer的特征提取架构可以很好地弥补全局上下文联系。

由于图像Transformer计算保留序列块(patch)位置信息需要借助位置编码，PVT(pyramid vision Transformer)模型(Wang等，2021)采用位置嵌入编码来实现位置信息定位。然而面对不同的分辨率输入，涉及到位置编码插值问题(Xie等，2021)，因此本文使用的Multi-Transformer骨干网络没有在序列编码阶段加入位置编码，而在前馈神经网络模块中使用3×3深度卷积来添加填充信息，通过零填充来涉及位置信息，同时还可以降低自注意力编码的计算量(Islam等，2020)。在Multi-Transformer骨干网络中共分为4个阶段，每个阶段都包含一个Transformer编码模块、序列编码模块以及重塑层如图2所示。

在重叠patch嵌入及展平模块中，巧妙地利用了二维卷积遍历特征图以获取具有重叠关系的patch，这种重叠关系类似Swin-Transformer中保持特征块周围的局部连续性，并且同时对特征图完成下采样，具体参数见表1。在第1层中使用较大的patch以便快速获得广域信息，降低分辨率，随后层均使用小patch以掌握细腻的特征关系。

在完成patch嵌入后进行线性映射序列展平。其中，$p \times p$大小的patch组成序列块，每幅图像的序列长度为$N = \dfrac{H \times W}{P^2}$，再进行展平得到线序列$\left\{\boldsymbol{x}_p^i \in \mathbf{R}^{p^2 \times c} \mid i = 1, \cdots, N\right\}$，其中$i$代表图像的第$i$个patch，再将序列$\boldsymbol{x}_p^i$由$C$维映射到Transformer的$D$维空间(embed dimensions)中，具体为

$$\boldsymbol{z}_0 = [\boldsymbol{x}_p^1\boldsymbol{E}; \boldsymbol{x}_p^2\boldsymbol{E}; \cdots; \boldsymbol{x}_p^N\boldsymbol{E}] \tag{1}$$

式中，$\boldsymbol{x}_p^i \in \mathbf{R}^{p \times p}$为第$i$个patch，$\boldsymbol{E} \in \mathbf{R}^{(p^2 \times C) \times D}$为线性映射关系。

Transformer编码器以自注意力为核心模块，由L层多头自注意力(multi-head self-attention，MSA)、前馈神经网络(feedforward neural network，FNN)和层归一化(layer normalization，LN)组成，如图3(a)所示，L层Transformer堆叠方式可以表示为

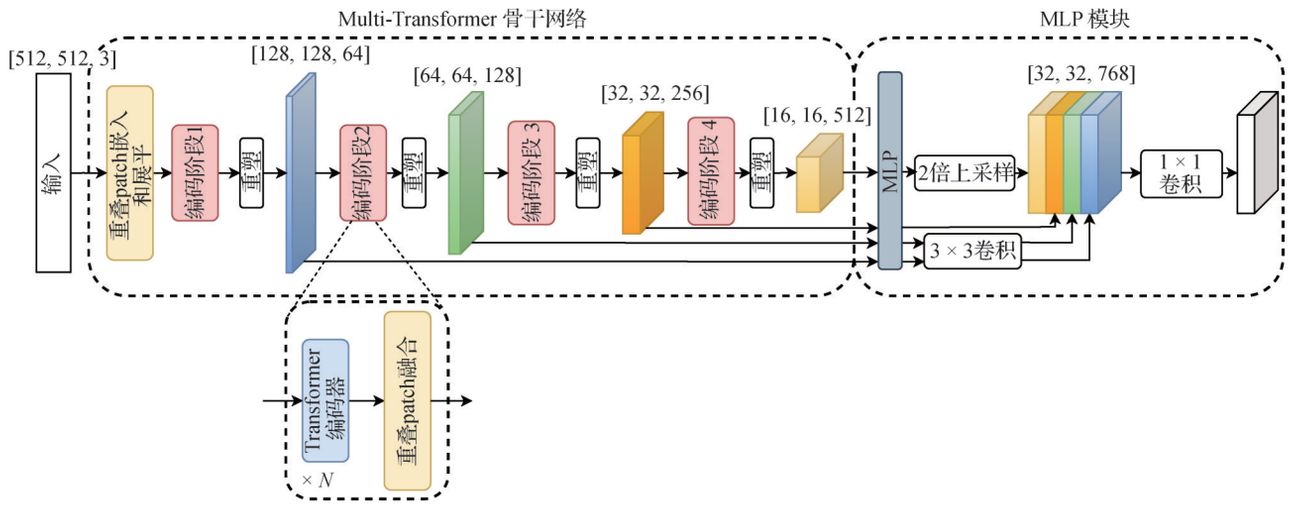


图 2　Multi-Transformer骨干网络以及MLP模块结构

Fig. 2　Multi-Transformer backbone and structure of MLP module

$$\begin{cases} z_l' = MSA\left(LN\left(z_{l-1}\right)\right) + z_{l-1} \\ z_l = FNN\left(LN\left(z_l'\right)\right) + z_l' \end{cases} \tag{2}$$

式中，z_l为第L层Transformer序列编码，$MSA(\cdot)$为多头注意力，$LN(\cdot)$为层归一化，$FNN(\cdot)$为前馈神经。

在一个自注意力结构中，针对输入序列$z_l \in \mathbf{R}^{N\times D}$，给定映射矩阵$\boldsymbol{W}_q, \boldsymbol{W}_k, \boldsymbol{W}_v \in \mathbf{R}^{D\times D_h}$，计算矩阵$\boldsymbol{query}, \boldsymbol{key}, \boldsymbol{value} \in \mathbf{R}^{N\times D_h}$，核心是求解每个元素的$\boldsymbol{v}$与其他元素的$\boldsymbol{k}$的相似性，经过缩放以及归一化求得加权和输出权重$\boldsymbol{A}_{ij}$，可表示为

$$\begin{aligned} &\left[\boldsymbol{q}, \boldsymbol{k}, \boldsymbol{v}\right] = z \times \boldsymbol{W}_{qkv}, \boldsymbol{W}_{qkv} \in \mathbf{R}^{D\times 3D_h} \\ &\boldsymbol{A} = f_{\text{softmax}}\left(\frac{\boldsymbol{q}\times \boldsymbol{k}^{\mathrm{T}}}{\sqrt{D_h}}\right), \boldsymbol{A} \in \mathbf{R}^{N\times N} \\ &SA(z) = \boldsymbol{A} \times \boldsymbol{v} \end{aligned} \tag{3}$$

式中，$\boldsymbol{W}_{qkv}$为映射矩阵，$\boldsymbol{q}$为$\boldsymbol{query}$矩阵，$\boldsymbol{k}$为$\boldsymbol{key}$矩阵，$\boldsymbol{v}$为$\boldsymbol{value}$矩阵，D_h为缩放因子，一般根据多头数量设置为$\frac{D}{h}$。

由于自注意力计算复杂度是$\mathrm{O}(N^2)$，这对于较高分辨率的图像分割是相当大的计算量（Dosovitskiy等，2021），因此引入渐进式空间序列缩减法（spatial-reduction attention，SRA）以降低计算量（Wang等，2021），保证网络模型的高效。如图3（c）所示，在这个过程中，只针对$\boldsymbol{key}$以及$\boldsymbol{value}$矩阵进行缩减，核心是通过将序列线性映射到较小尺度来完成自注意力计算。以$\boldsymbol{key}$为例可以表示为

$$\begin{cases} \boldsymbol{K}' = \mathrm{Reshape}\left(\frac{N}{R}, D_h \times R\right)(\boldsymbol{K}) \\ \hat{\boldsymbol{K}} = \mathrm{Linear}\left(D_h \times R, D_h\right)(\boldsymbol{K}') \end{cases} \tag{4}$$

式中，R为缩减度，$\boldsymbol{K} \in \mathbf{R}^{N\times D_h}$为原$\boldsymbol{key}$矩阵，$\hat{\boldsymbol{K}} \in \mathbf{R}^{\frac{N}{R}\times D_h}$为缩减矩阵，Reshape为重塑维度操作，Linear为维度映射操作。将$\boldsymbol{q}, \boldsymbol{k}, \boldsymbol{v}$代入式（3）仍然得到$SA(z) \in \mathbf{R}^{N\times D}$。由此，自注意力机制计算复杂度$\mathrm{O}(N^2)$降低到$\mathrm{O}\left(\frac{N^2}{R}\right)$。渐进式各阶段缩减度设置见表1。

此外，采用多头注意力机制可以输入特征映射到不同空间进行学习，增加网络理解能力，如图3（b）所示。在多头注意力中，一般是由h个头的自注意力构成。最终多个结果向量进行拼接后映射回原始尺度，表示为

$$MSA(z) = [SA_1(z); SA_2(z); \cdots; SA_h(z)] \times \boldsymbol{W}^0 \tag{5}$$

式中，$\boldsymbol{W}^0 \in \mathbf{R}^{h\times D_h\times D}$为级联映射矩阵，$SA(\cdot)$为第$i$个单头自注意力。

为保持全局上下文信息提取的非线性化，多头注意力机制输出需要经过FNN进行非线性激活，如图3（d）。FNN模块可以表示为

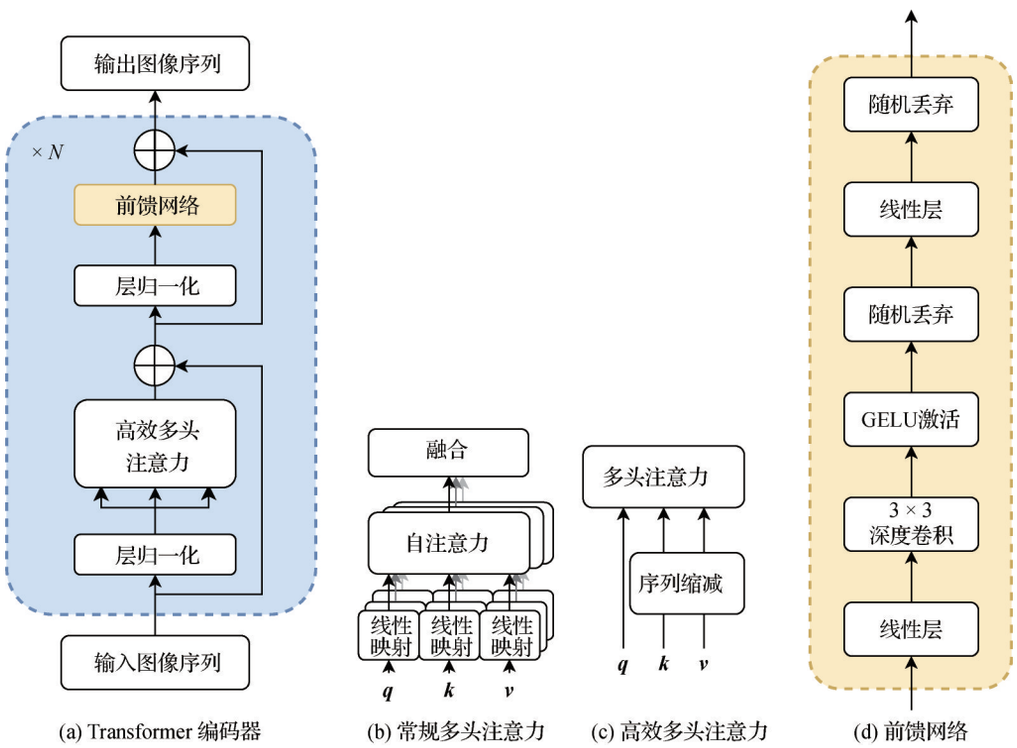
$$f_{\text{FNN}}(z) = f_{\text{MLP}}\left(f_{\text{GELU}}\left(f_{\text{DWConv}_{3\times 3}}\left(f_{\text{MLP}(z)}\right)\right)\right) + z \tag{6}$$

式中，MLP为多层感知机，GELU（Gaussian error linear units）为激活函数，DWConv为深度卷积，目的是填加零填充来弥补位置编码的缺失，还可以进一步

表1　Multi-Transformer网络参数

Table 1　Parameters of Multi-Transformer

阶段	重叠序列块嵌入(k, s, p)	输出步长(OS)	嵌入维度	多头数量	深度	缩减比率
1	(7, 4, 3)	4	64	1	3	8
2	(3, 2, 1)	8	128	2	4	4
3	(3, 2, 1)	16	320	5	6	2
4	(3, 2, 1)	32	512	8	3	1



(a) Transformer 编码器　(b) 常规多头注意力　(c) 高效多头注意力　(d) 前馈网络

图3　Transformer编码器结构

Fig. 3　Transformer encoder structure

((a)Transformer encoder; (b)general multi-head attention; (c)efficient multi-head attention; (d)feedforward neural network)

降低计算量。最终，编码结束后的序列$\frac{H \times W}{p^2}$重塑回特征层结构$\left[\frac{H}{OS}, \frac{W}{OS}\right]$。

1.2.2　MLP模块

在Multi-Transformer骨干网络中采用了分层的特征注意力提取结构，浅层特征关注于全局粗略关系，深层特征关注于局部细腻关系。为了充分利用多级信息，一个简单的MLP模块（如图1所示）加入到其后以进一步聚合各层的特征信息。在这个模块中，4个经过Transformer注意力计算得到的特征序列经过重塑分别得到相对于原尺寸下采样4、8、16以及32倍的特征图，随后通过MLP层映射到统一通道维度来平衡各特征图的贡献程度。由于在编码器阶段采用下采样16倍的输出提供给解码器生成掩膜，可以很好地权衡分割质量以及内存占用，因此将下采样32倍的特征进行上采样至16倍，下采样4和8倍的特征经过3×3卷积下采样到16倍。最终将四者拼接融合特征信息作为SACP模块的输入。

1.2.3　感受野增强的SACP模块

空洞卷积金字塔模块核心是通过构建不同的膨胀率的空洞卷积，从而得到不同感受野的卷积核，用来提取多尺度目标信息，但是每个分支的空洞卷积层相互独立，没有进行尺度特征信息的共享。

受到Xception骨干网络的启发，在改进模块中首先采用深度可分离卷积（depthwise separable convolution，DSConv）替代普通卷积以降低计算量

(Howard 等,2017)。深度可分离卷积将标准卷积分为两个部分:深度卷积(depthwise convolution)和点卷积(pointwise convolution)。在深度卷积中特征的每个通道都与唯一的卷积核对应,保证通道信息的独立性,随后在 1 × 1 的点卷积中将特征的不同通道信息融合到固定通道数,实现不同通道的信息交流。

此外,受到 DenseASPP 的启发(Yang 等,2018),为进一步加强金字塔结构的提取能力,本文将膨胀率为 3 以及 24 的空洞卷积分支加入到并行结构中,形成空洞卷积分支膨胀率 $\{x_i | i = 1, 3, 6, 12, 18, 24\}$ 的并行结构以进一步扩大感知范围,如图 1 中的 SACP 模块所示。得益于强大的骨干网络上下文特征提取能力,并没有采用高计算复杂度的稠密连接,而是改进为一种简单而高效的形式:将膨胀率为 1、3、6、12 以及 18 的空洞卷积块的输出结果通过添加拼接模块,与下一个卷积分支输入进行空间和通道尺度对齐堆叠,以此构建出联级空洞卷积关系,可以进一步融合感受野感知的范围,提高模块多尺度特征提取的能力,同时降低模块的计算量。提出的 SACP 具体结构如图 4 所示。

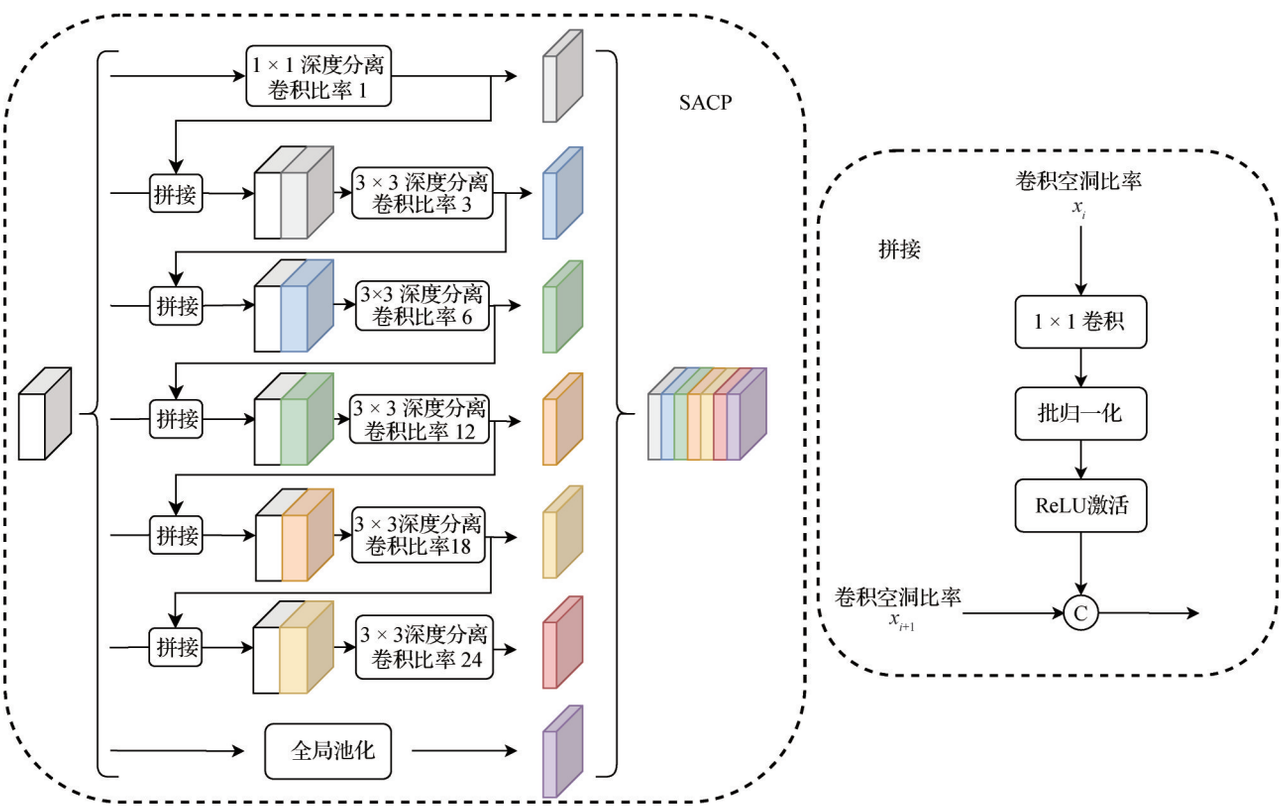


图 4　SACP 结构

Fig. 4　Series-parallel atrous convolutional pyramid structure

1.2.4　边界增强卷积注意力

为提高解码器能力,本文方法加入卷积注意力模块以加工来自浅层骨干网络的特征信息,提高边界信息的有效保留。卷积注意力模块(concentration-based attention module,CBAM)(Woo 等,2018)由空间注意力及通道注意力构成,卷积注意力机制可以提升浅层局部目标特征的权重,降低噪声带来的影响,因此可以高效地应对成像质量不高、存在管线遮挡的内衬病害提取问题。CBAM 主体结构如图 5 所示。

输入特征在 CBAM 模块中首先经过通道注意力模块完成空间向权重计算,随后与原始输入特征进行通道权重相乘融合,得到通道向有效特征信息增强。再将输出特征经过空间注意力模块完成通道向权重计算,随后同样与输入进行空间权重融合,得到空间向有效特征信息增强。CBAM 整体可以表示为

$$CBAM(F) = M_S \otimes F' = M_S \otimes (M_C \otimes F) \quad (7)$$

式中,F 为输入特征,M_C 为通道注意力权重,M_S 为空间注意力权重,⊗ 为逐像素相乘操作。

通道注意力模块中重点关注通道向信息,具体表示为

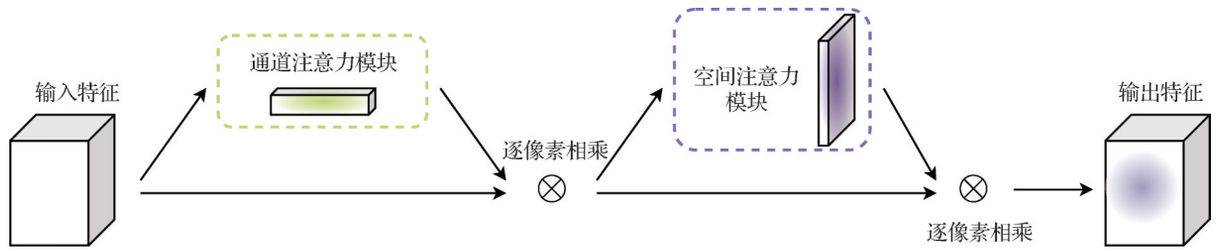


图5　CBAM模块

Fig. 5　Concentration-based attention module

$$M_C(F)=\sigma\left(f_{\mathrm{MLP}}\left(f_{\mathrm{AvgPool}}(F)\right)\right)+\sigma\left(f_{\mathrm{MLP}}\left(f_{\mathrm{MaxPool}}(F)\right)\right) \tag{8}$$

式中，σ为sigmoid激活函数，f_{MLP}为多层感知机，$f_{\mathrm{AvgPool}}(\cdot)$为全局平均池化，$f_{\mathrm{MaxPool}}(\cdot)$为全局最大池化。空间注意力模块将重点关注空间向信息，具体表示为

$$M_S(F)=\sigma\left(F^{7\times7}\left[\left(f_{\mathrm{AvgPool}}(F');f_{\mathrm{MaxPool}}(F')\right)\right]\right)=\sigma\left(F^{7\times7}\left(\left[F^S_{\mathrm{avg}};F^S_{\mathrm{max}}\right]\right)\right) \tag{9}$$

式中，σ为sigmoid激活函数，$F^{7\times7}$为7×7卷积操作，$f_{\mathrm{AvgPool}}(\cdot)$为全局平均池化，$f_{\mathrm{MaxPool}}(\cdot)$为全局最大池化。

2　实验内容

2.1　实验平台

实验所使用的硬件平台为CPU AMD 3960X，内存128 GB四通道，GPU RTX 3090×4。软件平台为操作系统Ubuntu 18.04，编程语言PyThon 3.8，深度学习库PyTorch 1.8，并行计算架构CUDA 11.1。

2.2　数据采集与处理

为验证本文模型的泛化能力，实验使用两大自定义病害数据集，包括盾构隧道内衬病害数据集以及桥梁缆索病害数据集。

2.2.1　隧道内衬病害数据集

实验采用广州市地铁提供的现场探测拍摄的48份RGB图像以及LiDAR扫描强度转化的61份灰度图像作为原始数据。统一裁剪到512×512像素大小后使用标注软件标注病害区域，重点关注病害出现的位置，如图6所示，其中各子图左边为病害、右边为标签。

为了弥补训练样本数量不足，网络模型易出现欠拟合情况，实验中对病害数据集进行数据增强操作，包括物理变化和色彩变化，具体见表2，处理效果如图7所示。经过增强后样本总量为1 000幅，依照7∶2∶1比例来划分训练集、验证集以及测试集，基于此数据集开展模型内的各模块有效性实验以及模型间对比实验。

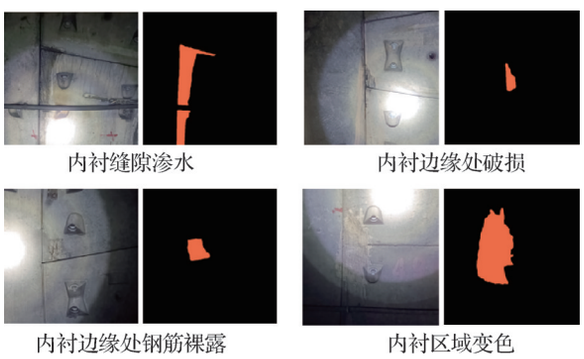


图6　隧道内衬常见病害及其标签

Fig. 6　Common defects of tunnel lining and their labels

表2　数据集增强类别及其参数

Table 2　Dataset enhancement categories and their parameters

类别		概率/%	范围
物理变化	旋转	80	±25°
	左右互换	50	–
	缩放	30	±15%
	小块变形	80	±10%
	剪切	100	10%
	裁剪	100	80%
	翻转	100	–
色彩变化	亮度	100	±50%
	颜色	100	100%
	对比度	100	±50%

注：“–”表示无需指定范围。

图 7　数据集增强操作

Fig. 7　Dataset enhancement operations

2.2.2　桥梁缆索病害数据集

实验采用由缆索机器人现场拍摄，位于山西某斜拉桥梁的缆索病害 RGB 图像作为原始数据。统一裁剪到 960 × 540 像素大小后进行标注，病害基本类型及其标签如图 8 所示，其中各子图左边为病害、右边为标签。样本总量为 2 304 幅，依照 7∶2∶1 比例来划分训练集、验证集以及测试集，基于此数据集开展模型间对比实验。

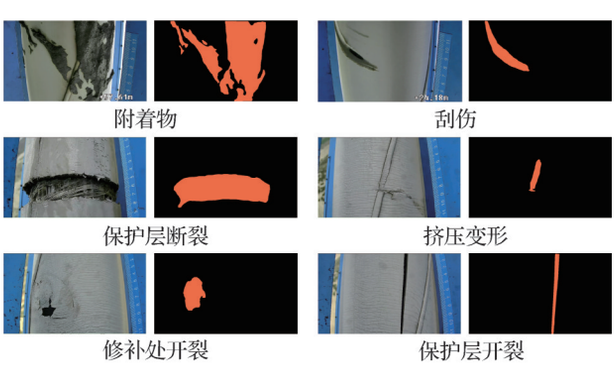


图 8　桥梁缆索常见病害及其标签

Fig. 8　Common defects of bridge cables and their labels

两个数据集存在病害类别不同以及应用场景的差异，但由于本文模型骨干部分使用 ImageNet-1K 大数据集进行预训练，随后整体在目标数据集中进行少量训练微调参数，以此将网络迁移到目标领域。经过迁移学习后，本文方法都能够达到较好的结果，证明该方法具有良好的泛化能力，不存在场景迁移的问题。

2.3　损失函数

在深度学习中，交叉熵损失函数（cross entropy loss，CE）经常作为多分类任务的损失函数，交叉熵损失可以表示为

$$L_{ce} = -\sum_{i=1}^{n} p(x_i) \times \ln\left(q(x_i)\right) \tag{10}$$

式中，x_i 为第 i 个样本，n 为总类别数，$p(\cdot)$ 为真实概率，$q(\cdot)$ 为预测概率。针对隧道内衬病害分布情况，负样本（即背景，类别为 0）占大部分，使得损失函数偏重于单个类别，不便于网络对病害特征的学习，因此改进采用复合损失函数。Dice 是一种集合相似度度量指标，通常用于计算两个样本的相似度，值域为 [0，1]，越接近 1 表示两个集合越相似。加权采用 Dice 损失函数使得网络更关注前景目标，加强学习效果。Dice 损失函数可以表示为

$$L_{Dice} = 1 - \frac{2\left|\boldsymbol{y} \cap \boldsymbol{y}'\right|}{\left|\boldsymbol{y}\right| + \left|\boldsymbol{y}'\right|} \tag{11}$$

式中，$\boldsymbol{y}$ 为真实值，$\boldsymbol{y}'$ 为预测值，$|\cdot|$ 为值的数量。最终，采用的复合损失函数表示为

$$L_{all} = \alpha \times L_{ce} + \beta \times L_{Dice} \tag{12}$$

式中，α 和 β 为平衡权重。

2.4　评价指标

语义分割是对图像的所有像素点进行分类，赋予不同的标签值。对于语义分割的评价指标，本文使用准确率、平均交并比以及 F1 分数作为评价指标。

准确率是指正确预测的样本数占总预测样本数的比值，可以表示为

$$I_{acc} = \frac{N_{TP} + N_{TN}}{N_{TP} + N_{FP} + N_{FN} + N_{TN}} \tag{13}$$

式中，N_{TP} 为真正例数量，N_{FP} 为假正例数量，N_{FN} 为真反例数量，N_{TN} 为假反例数量。

平均交并比（mean intersection over union，mIoU）表示各分类预测值与真实值之间交并比（IoU）的平均值，mIoU 的值域为 [0，1]，值越接近 1 表示分割精度越高，分割效果越好，计算式为

$$mIoU = \frac{1}{k+1} \sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} + p_{ii}} \tag{14}$$

式中，k 代表类别数，i 为真实值，j 为预测值，p_{ii} 为真

正例(true positive, TP),p_{ji}为假正例(false positive, FP),p_{ij}为假反例(false negative, FN)。

F1分数(F1-score)表示精确率和召回率的加权平均,F1-score的值域为[0,1],值越接近1表示分割精度越高,计算为

$$F1 = 2 \times \frac{P \times R}{P + R} \tag{15}$$

式中,P为精确率,R为召回率。具体为

$$\begin{aligned} P &= \frac{N_{TP}}{N_{TP} + N_{FP}} \\ R &= \frac{N_{TP}}{N_{TP} + N_{FN}} \end{aligned} \tag{16}$$

2.5 实验设置

网络训练时选用Adam优化器,其中动量设置为0.9,权重衰减设置为0.000 5;学习率衰减策略使用poly指数衰减,初始学习率设为0.007;默认采用随机种子;多卡训练使用SyncBatchNorm;批处理设置为16,迭代次数(epochs)为100。实验过程中均采用ImageNet-1K大规模数据集对骨干网络进行预训练,载入最佳预训练权重以加快网络迁移学习的收敛速度。

3 实验结果与分析

3.1 骨干网络的对比实验

为验证本文改进的骨干网络性能以及计算复杂度,将Multi-Transformer替换为主流骨干网络展开对比实验,实验结果如表3所示。其中ResNeXt-101在瓶颈结构卷积中使用32组,每组8个卷积核的深度可分离卷积,Swin-Transformer使用微型模型,ConvNeXt使用基础模型。

表3 基于不同骨干网络的参数量及精度对比

Table 3 Comparison of parameters and accuracy based on different backbone

骨干网络	参数量/M	浮点计算量/GFlops	精度	
			Acc/%	mIoU/%
Xception(Chen等,2018)	37.87	46.67	88.51	55.33
MobileNet-V2(Sandler等,2018)	15.40	10.69	92.66	73.74
ResNet-101(He等,2016)	42.50	178.20	94.25	77.69
ResNeXt-101(32×8 d)(Xie等,2017)	86.74	363.59	94.57	78.72
Swin Transformer-T(Liu等,2021)	27.52	24.76	93.62	74.70
ConvNeXtV2-B(Woo等,2023)	88.72	80.22	89.56	60.35
Multi-Transformer-原始	38.98	33.38	–	–
Multi-Transformer-序列缩减	24.20	28.64	**94.40**	**78.14**

注:加粗字体表示各列最优结果。"–"表示数据项无需评估。

由表3可以得出,所使用的Multi-Transformer骨干网络使用序列缩减法将参数量降低38%,浮点计算量也降低14%。相比于参数量最低的MobileNet-v2骨干网络mIoU精度提升4.4个点,同时略高于ResNet-101精度表现。得益于强大的多尺寸特征输出聚合设计,相较于Transformer骨干单输出的Swin-Transformer,在参数减少的同时精度提升3.44个点。距离结构复杂的ResNeXt-101仍有近0.6个点的mIoU精度差距,但参数量降低72%,浮点计算量降低92%。因此,Multi-Transformer在保持较低计算要求的前提下,表现出较为出色的分割能力,因此作为本文模型的最佳性价比选择。

3.2 SACP改进实验

为检验所提出的SACP模块改进方式的有效性,实验1定义原始空洞卷积金字塔模块为模型A,仅加入膨胀率为3、24空洞卷积分支为模型B,仅加入拼接模块为模型C,同步加入卷积分支和拼接模块的改进为模型D,结果如表4所示;实验2对比将SACP模块中的标准卷积改进为深度可分离卷积的参数量变化,结果如表5所示。

由表4得知,相比原始空洞卷积金字塔,改进后的模块mIoU精度提升2%,达到78.14%。而单独添加卷积分支以及加入拼接模块相比原版提升有限,均提升不到1%。由此得出结论,加强空洞卷积分支

表 4　基于不同改进方式的SACP模块精度对比

Table 4　Comparison of SACP module accuracy based on different improvement methods

SACP模块	精度	
	Acc/%	mIoU/%
模型 A	93.71	76.01
模型 B	94.09	76.86
模型 C	94.02	76.73
模型 D	**94.40**	**78.14**

注：加粗字体表示各列最优结果。

表 5　基于使用不同卷积方式的SACP模块参数量对比

Table 5　Comparison of SACP module parameter based on using different convolutional methods

SACP卷积方式	参数量/M
标准卷积	19.22
深度可分离卷积	7.46

并彼此融合各感受野可以更好地感知完整的目标区域，两者改进方式结合可以获得更好的分割结果。

由表 5 知，将模块中的标准卷积改进为深度可分离卷积可以将参数量降低 61. 18 %。综合以上，改进后SACP模块使用的深度可分离膨胀卷积融合了深度可分离卷积以及膨胀卷积的优势，获取多个较大感受野的同时，降低参数量和计算复杂度，提高模型的效率。

由图 9 的模块整体感受野热力图得知，相比模型 A 仅使用空洞卷积金字塔，改进后的模型 D 有更加均匀广泛的感受野分布，可以更加全面感知目标特征。此外，在中心关注区域感受野分布也更加均匀，因此改进后的方案更适合复杂分布的病害情况。

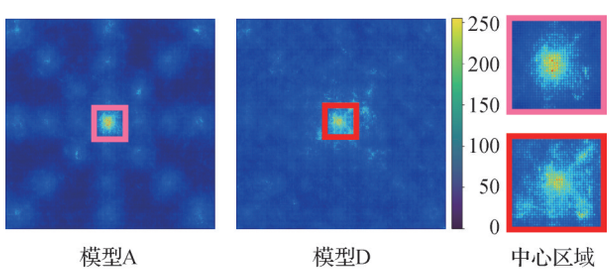


图 9　感受野热力图（数值表示关注度）

Fig. 9　Heat map of receptive field (the color bar indicates the level of attention)

3. 3　消融实验

由于使用多尺寸特征输出的骨干网络，为统一各尺寸的贡献程度，使用 MLP 模块进行尺度和通道对齐以融合多尺寸特征。在此实验中，MLP 模块与 3. 2 节中改进的SACP模块作为消融单位参与实验，以证明这两个模块的添加是否对网络模型带来分割精度改进，再加入 CBAM 模块进行验证实验，结果如表 6 所示。

表 6　网络的消融精度对比

Table 6　Comparison of ablation accuracy

实验	MLP	SACP	CBAM	精度/%	
				Acc	mIoU
1	–	–	–	93.32	75.16
2	√	–	–	93.91	76.42
3	–	√	–	93.95	76.14
4	–	–	√	93.67	75.82
5	√	√	–	94.15	77.37
6	√	√	√	**94.40**	**78.14**

注：加粗字体表示各列最优结果。“√”和“–”分别表示添加和不添加相应模块。

由表 6 可以得出，未经过 MLP 模块进行通道映射以及 SACP 加强的实验 1 模型分割精度最差，相较最优模型 mIoU 精度下降 3%。单独加入 MLP 模块的实验 2 可以提升 mIoU 精度 1. 3%，单独改进 SACP 模块的实验 3 可以提升 1%，综合两者改进的实验 5 对网络模型有分割精度提升贡献，单独加入 CBAM 模块的实验 4 模型精度也略有提升。综合 3 项改进的实验 6 整体模型 mIoU 精度为 78. 14%。

图 10 显示了模型特征关注区域的变化情况。在骨干网络提取后的注意分布中，存在着大量误关注区域，目标提取区域不明显，目的是广泛提取潜在的关注目标。经过 MLP 模块多层特征融合，目标区域基本聚焦，但依旧存在较多其他干扰区域。经过 SACP 模块多尺度感知，目标病害区域得到完备凸显，但由于较大感受野使得光线不均处存在部分噪声。最后经过 CBAM 模块注入注意力，目标区域权重提高，边界清晰，噪声得到有效抑制。

3. 4　损失函数对比实验

由于使用交叉熵损失函数以及 Dice 损失函数的复合损失函数来计算网络预测与实际的差距，本

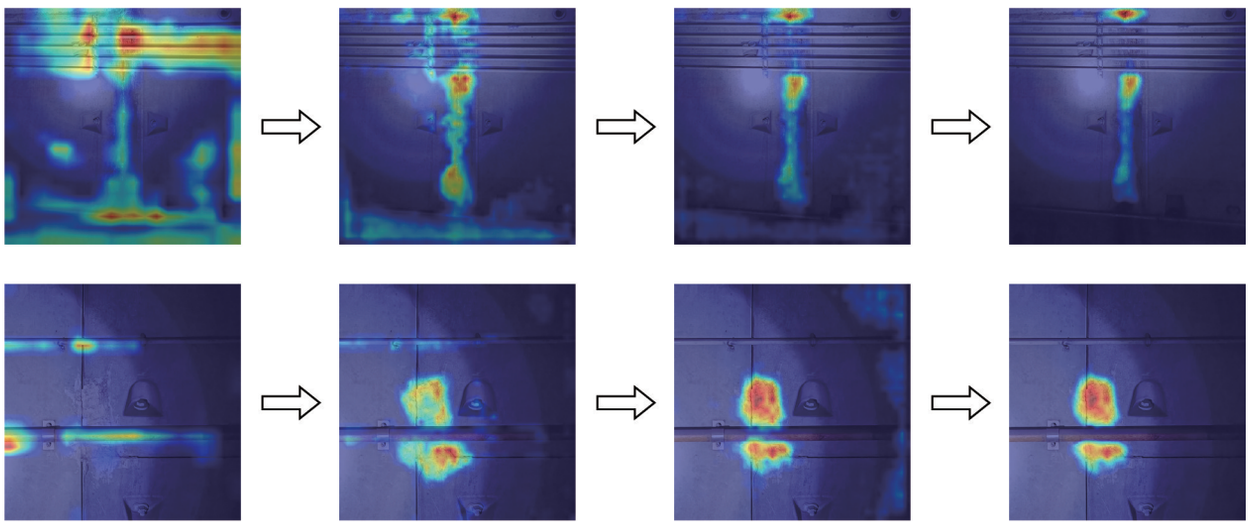


图10　特征注意力热力图变化(红色表示关注度高)

Fig. 10　Changes in feature attention heatmap (red indicates high level of attention)

实验针对复合函数的权重展开实验,实验结果如表7所示,训练中的损失值如图11所示。

表7　基于损失函数不同平衡权重训练精度对比

Table 7　Comparison of training accuracy based on different balance weights of loss functions

平衡权重		精度	
α	β	Acc/%	mIoU/%
1	0	93.85	75.94
0.8	0.2	93.74	76.39
0.5	0.5	**94.40**	**78.14**
0.2	0.8	94.06	77.74

注:加粗字体表示各列最优结果。

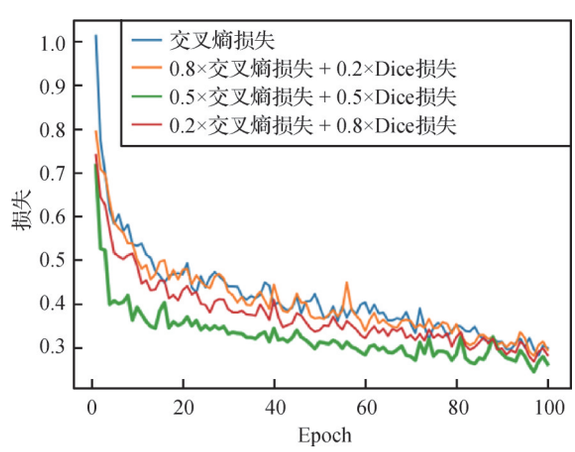


图11　复合损失函数损失值变化

Fig. 11　Change in loss value of composite loss function

由表7可以得到,加入Dice损失可以降低受到大量背景负样本的影响,使得整体模型分割mIoU精度提高2.43个百分点。由图11实验数据可知,0.5倍交叉熵损失搭配0.5倍Dice损失的复合函数初始损失较低并且收敛速度快于其他类型,因此作为后续实验的基本复合损失函数搭配。

3.5　模型间对比实验

为验证网络模型分割病害的有效性,本实验利用构建的隧道内衬数据集以及桥梁缆索数据集,将改进后的网络与现有主流网络展开训练验证。各网络模型分割评定指标如表8和表9所示。分别额外选取实地采集的病害图像进行分割预测。其中针对隧道内衬数据集,6幅图像代表管线遮挡、低光照以及大光比复杂的环境,分割结果如图12所示。针对桥梁缆索数据集,7幅图像代表细粒度损伤、轻度损伤、病害大跨度、阴影干扰以及多种病害情况,分割结果如图13所示。

由表8可知,FCN、SegNet、LR-ASPP以及Swin-Unet由于模型本身堆叠量的限制,无法有效提取病害信息,导致网络分割精度较低。SegNeXt使用多尺度卷积注意依然无法应对多尺度的病害特征,精度最低。LR-ASPP在实现最低计算需求下取得较为显著的精度提升,得益于精炼的骨干网络MobileNet-V3以及简化的ASPP结构。Bi-Former由于没有充分利用编码器的多层特征信息,以及缺少良好设计的解码器结构,尽管使用动态稀疏聚合全

表 8　各主流网络模型参数量及在隧道内衬数据集的精度对比

Table 8　Comparison of parameters and accuracy among mainstream networks based on tunnel lining dataset

网络模型	参数量/M	浮点计算量/GFlops	精度/%		
			Acc	mIoU	F1-score
FCN(VGG-16)(Shelhamer等,2017)	35.31	148.53	90.91	66.03	49.17
SegNet(VGG-16)(Badrinarayanan等,2017)	29.44	160.68	91.53	70.32	65.64
PSPNet(ResNet50)(Zhao等,2017)	46.71	69.46	93.86	72.06	68.35
DeepLabv3(ResNet50)(Chen等,2017)	41.99	173.79	91.10	66.10	50.38
DeepLabv3+(MobileNet-V2)(Chen等,2018)	41.83	33.75	92.66	73.74	71.14
DeepLabv3+(ResNet101)(Chen等,2018)	74.87	82.88	94.25	77.69	73.93
LR-ASPP(MobileNet-V3)(Howard等,2019)	3.22	2.07	93.31	74.05	71.29
SegNeXt(Base)(Guo等,2022)	29.91	41.55	90.31	64.13	48.89
Swin-Unet(Cao等,2022)	27.17	5.92	93.15	73.65	70.72
Bi-Former(Base)(Zhu等,2023)	58.40	91.10	92.61	71.17	66.95
本文	47.69	58.22	**94.40**	**78.14**	**76.45**

注:加粗字体表示最优精度结果。

表 9　各主流网络模型在桥梁缆索数据集的精度对比

Table 9　Comparison of accuracy among mainstream networks based on bridge cables dataset

网络模型	精度/%		
	Acc	mIoU	F1-score
DeepLabv3+ (MobileNet-V2)(Chen等,2018)	97.02	78.90	76.17
DeepLabv3+ (ResNet101)(Chen等,2018)	97.09	79.72	**78.32**
LR-ASPP (MobileNet-V3)(Howard等,2019)	96.54	76.39	72.12
SegNeXt (Base)(Guo等,2022)	94.66	65.34	56.13
Swin-Unet(Cao等,2022)	96.71	78.02	74.61
Bi-Former (Base)(Zhu等,2023)	95.12	68.02	58.25
本文	**97.15**	**80.41**	77.92

注:加粗字体表示最优精度结果。

局上下文特征关系,但计算量依然较高,有效信息得不到充分细化利用。PSPNet 由于采用 ResNet-50 骨干,分割精度显著上升,但参数量随之提高。DeepLabv3+模型采用 MobileNet 骨干网络的参数量基本与 DeepLabv3 相当,但 mIoU 精度提升 11.5%,F1 精度提升 20%,这表明 MobileNet 搭配解码器的使用可以在同等复杂度下提升分割精度。采用 ResNet-101 骨干网络的 DeepLabv3+模型达到 77.69% 的 mIoU 分割精度,高堆叠骨干网络使得特征信息利用率得到提高,代价是较大计算量。而本文提出的模块 SACP 弥补空洞卷积金字塔采样不够密集的缺点,使用 Transformer 骨干网络在保证特征提取效果的前提下,引入序列缩减机制使得参数量下降 12%,取得最高分割精度,mIoU 为 78.14%,F1 分数为 76.45。

由表 9 可知,本文模型在桥梁缆索数据集上保持较高的分割水平,其 mIoU 达到 80.41%,Acc 达到 97.15%。这一性能优势相对于隧道内衬数据集上的表现更为突出。这是因为缆索图像采集基本处于光线充足的环境,病害特征明显,拍摄距离相对固定,相对宽松的环境条件使得网络能够表现出更好的分割

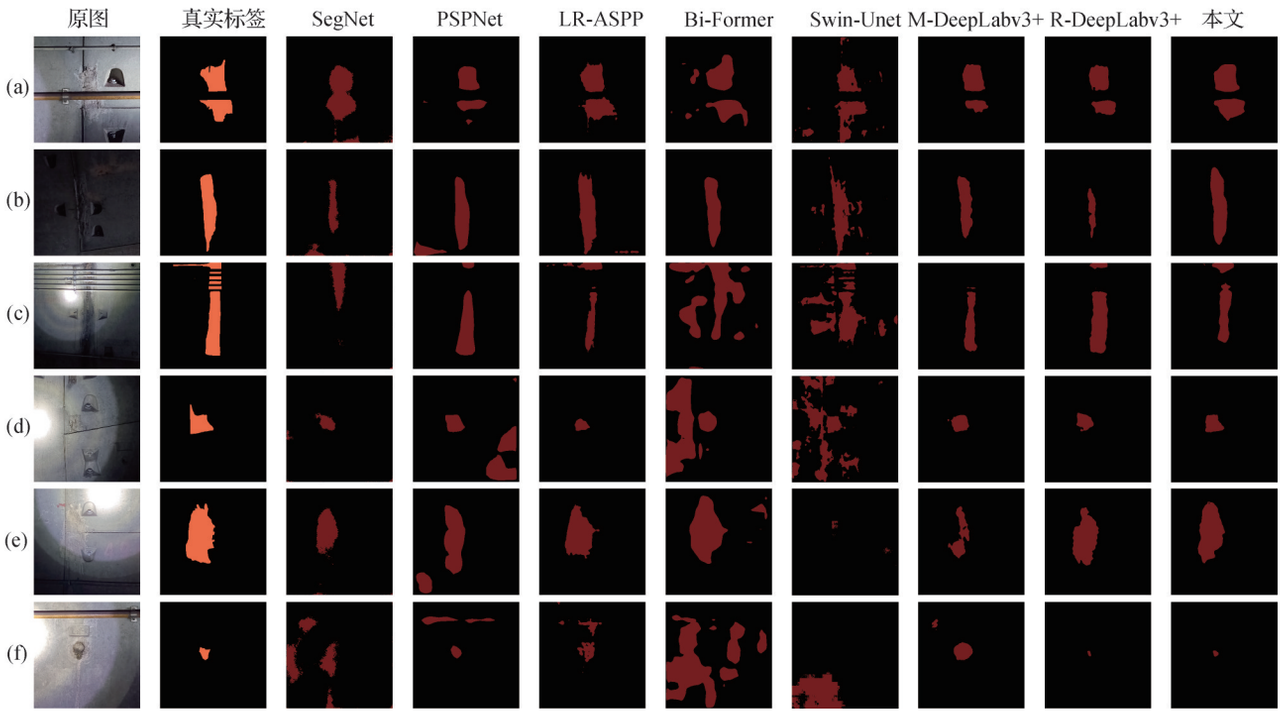


图12　隧道内衬图像分割结果

（M-DeepLabv3+为使用MobileNet骨干网络；R-DeepLabv3+为使用ResNet-101骨干网络）

Fig. 12　Image of tunnel lining segmentation results

（M-DeepLabv3+ represents the use of MobileNet backbone; R-DeepLabv3+ represents the use of ResNet-101 backbone）

精度。同时，明显的病害特征也使得本文模型与主流网络之间的差距减小。在F1分数上，使用ResNet101骨干网络的DeepLabv3+比本文模型精度高0.4%，这是由于缆索病害分布较为集中，面积相对固定，使得本文基于全局上下文关系的提取机制以及增强感受野的优势不能充分发挥。然而，本文模型在mIoU以及Acc精度上依然保持领先。

由图12得知，PSPNet与SegNet由于网络深度不足，无法有效应对复杂环境，在图12(b)(d)(f)中，SegNet边缘混乱，PSPNet出现大量的误判断区域。Swin-Unet以及Bi-Former在图12(c)(d)大光比环境中在同样出现较多误分割且零星区域，单纯提取上下文信息在复杂环境中无法很好地保留有效特征信息，表现较差。M-DeepLabv3+与R-DeepLabv3+情况较为相似，空洞卷积金字塔模块相对独立的感受野无法扩大对目标区域的感知，在对图12(b)(e)分割时都缺失较多的区域，R-DeepLabv3+对图12(c)的分割最准确，但依然无法明确区分管线遮挡的病害。LR-ASPP在图12(a)(c)中对病害整体分割较为精确，但感知面积偏小，同样出现零星误分割区域。而本文模型依靠强大的Transformer机制可以保证上下文信息的关系感知，SACP模块扩大目标的特征感知范围，解码器加入注意力机制提高边缘信息的保留能力，图12(a)(b)(d)(e)分割区域最为完整，边缘清晰，能够有效应对光线分布不均、存在线路遮挡的环境条件。图12(c)效果相对较差，但能够发现在管线遮挡位置尝试分割的痕迹，较符合真实标签。图12(f)针对细腻度病害检测，本文模型尽管分割面积较真实标签小，但依旧较其他模型准确，没有受到管线的信息干扰。综上，本文模型能更好地满足隧道内衬病害分割任务的需求。

由图13(a)得知，M-DeepLabv3+、LR-ASPP以及Bi-Former无法有效地对细粒度附着物进行提取，本文模型依靠多层特征利用机制，能够完整保留有效特征。在图13(b)中，对于轻度损伤，R-DeepLabv3+表现最佳，保留完整分段的病害，尽管本文模型出现对缆索外的区域误分割，但正确识别的病害区域与标签基本相同。在图13(c)中，本文模型依靠强大的上下文关系提取机制，面对大跨度病害区域提取表现最佳，基于Transformer机制的Bi-Former同样分

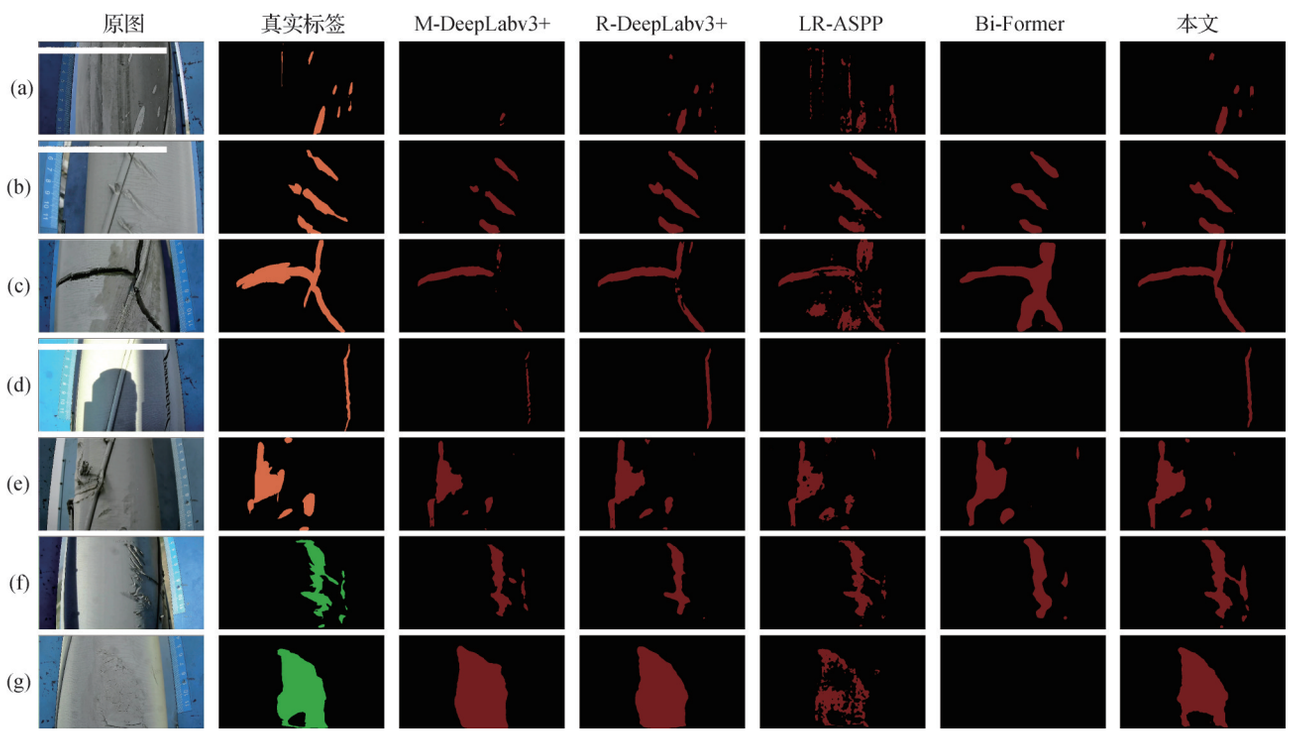


图 13　桥梁缆索图像分割结果

（M-DeepLabv3+为使用 MobileNet 骨干网络；R-DeepLabv3+为使用 ResNet-101 骨干网络）

Fig. 13　Image of bridge cables segmentation results

（M-DeepLabv3+ represents the use of MobileNet backbone; R-DeepLabv3+ represents the use of ResNet-101 backbone）

割区域完整，但提取面积过大，而其他基于卷积的模型均无法保留完整的开裂区域。在图 13(d)的阴影干扰环境中，本文模型与 R-DeepLabv3+和 LR-ASPP 均能保证病害提取完整。在图 13(e)(f)中，上述模型均能完整提取人工标签未发现的额外病害区域，LR-ASPP 由于相对固定的卷积金字塔导致零星空洞的分割区域。在图 13(g)中，面对轻度裂纹，M-DeepLabv3+和 R-DeepLabv3+都将部分正常区域界定为损伤，相比之下本文模型病害提取更为完整，且边界定位清晰。综上，本文模型同样能满足桥梁缆索病害分割任务的需求。

4　结　论

本文针对桥隧表面病害区域检测问题，提出一种融合注意力机制及增强感受野融合的深度网络模型。模型中采用具有强大全局信息关联能力的 Transformer 骨干网络，并且融合多层特征的机制可以更好地取得丰富的上下文特征信息。与此同时，通过改进密集空洞空间金字塔 SACP 以获得更大病害整体范围的感受野。此外，在解码器加入的卷积注意力也能更好地抑制噪声，提高边界细节的区分能力。最后表明使用深度可分离卷积可以有效降低模型的参数量，提高效率。本文实地采样隧道内衬病害以及桥梁缆索病害图像作为数据集，对本文模型展开实验。隧道内衬病害对比实验结果表明本文提出的模型相较于主流的 Unet、DeepLabv3、DeepLabv3+和 Swin-Transformer 模型有着更精确、更完整的分割效果。准确率能够达到 94.4%，平均交并比达到 78.14%，F1 分数达到 76.45%。此外，在桥梁缆索病害对比实验中同样展示出本文模型的较强泛化能力。因此，本文提出的深度网络模型在保证轻量化的同时实现较好的病害定位能力，可以较好地应用于实际工程中。由于数据集覆盖范围尚未全面、数据集不足以支撑多类别分割以及数据标注成本高等问题，在接下来的工作中将引入少样本学习策略，并且获取更多病害实例来确保更精准的病害预测。

参考文献（References）

Badrinarayanan V, Kendall A and Cipolla R. 2017. SegNet: a deep con-

volutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(12): 2481-2495 [DOI: 10.1109/TPAMI.2016.2644615]

Bhowmick S, Nagarajaiah S and Veeraraghavan A. 2020. Vision and deep learning-based algorithms to detect and quantify cracks on concrete surfaces from UAV videos. Sensors, 20(21): #6299 [DOI: 10.3390/s20216299]

Cao H, Wang Y Y, Chen J, Jiang D S, Zhang X P, Tian Q and Wang M N. 2022. Swin-Unet: unet-like pure transformer for medical image segmentation//Proceedings of 2022 European Conference on Computer Vision (ECCV). Tel Aviv, Israel: Springer: 205-218 [DOI: 10.1007/978-3-031-25066-8_9]

Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A and Zagoruyko S. 2020. End-to-end object detection with transformers//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer International Publishing: 213-229 [DOI: 10.1007/978-3-030-58452-8_13]

Chen L C, Papandreou G, Schroff F and Adam H. 2017. Rethinking atrous convolution for semantic image segmentation [EB/OL]. [2024-04-07]. https://arxiv.org/pdf/1706.05587.pdf

Chen L C, Zhu Y K, Papandreou G, Schroff F and Adam H. 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation//Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich, Germany: Springer: 833-851 [DOI: 10.1007/978-3-030-01234-2_49]

Deng P. 2015. Research on Detection of Railway Tunnel Seepage with Train-Mounted GPG. Chengdu: Southwest Jiaotong University (邓普. 2015. 铁路车载探地雷达检测隧道渗漏水的研究. 成都: 西南交通大学)

Dian S Y, Huang J J, Wu K J and Zhong Y Z. 2022. TCS-Net: a tiny crack segmentation network for nuclear containment vessel. Advanced Engineering Sciences, 54(5): 249-256 (佃松宜, 黄儆进, 吴克江, 钟羽中. 2022. TCS-Net:一种核电安全壳细小裂缝分割网络. 工程科学与技术, 54(5): 249-256) [DOI: 10.15961/j.jsuese.202100590]

Diao Z H, Zhao C J, Wu G and Guo X Y. 2010. Application research of mathematical morphology in image processing of crop disease. Journal of Image and Graphics, 15(2): 194-199 (刁智华, 赵春江, 吴刚, 郭新宇. 2010. 数学形态学在作物病害图像处理中的应用研究. 中国图象图形学报, 15(2): 194-199) [DOI: 10.11834/jig.20100202]

Dong S J, Tan H, Liu C and Hu X L. 2024. Apparent disease detection of bridges using improved YOLOv5s. Journal of Chongqing University, 47(9): 91-100 (董绍江, 谭浩, 刘超, 胡小林. 2024. 改进YOLOv5s的桥梁表观病害检测方法. 重庆大学学报, 47(9): 91-100)

Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X H, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J and Houlsby N. 2021. An image is worth 16 × 16 words: Transformers for image recognition at scale [EB/OL]. [2024-04-07]. https://arxiv.org/pdf/2010.11929.pdf

Fang J, Yang C, Shi Y T, Wang N and Zhao Y. 2022. External attention based TransUNet and label expansion strategy for crack detection. IEEE Transactions on Intelligent Transportation Systems, 23(10): 19054-19063 [DOI: 10.1109/TITS.2022.3154407]

Gong R, Ding S, Zhang C H and Shu H. 2020. Lightweight and multi-pose face recognition method based on deep learning. Journal of Computer Applications, 40(3): 704-709 (龚锐, 丁胜, 章超华, 苏浩. 2020. 基于深度学习的轻量级和多姿态人脸识别方法. 计算机应用, 40(3): 704-709) [DOI: 10.11772/j.issn.1001-9081.2019071272]

Guo M H, Lu C Z, Hou Q B, Liu Z N, Cheng M M and Hu S M. 2022. SegNeXt: rethinking convolutional attention design for semantic segmentation//Proceedings of the 36th International Conference on Neural Information Processing Systems. New Orleans, USA: Curran Associates Inc.: 1140-1156

He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]

Hong J H. 2023. Survey on visual inspection technique for surface damages of subway tunnel engineering. Railway Investigation and Surveying, 49(1): 18-22, 32 (洪江华. 2023. 地铁隧道表面病害视觉检测技术应用综述. 铁道勘察, 49(1): 18-22), 32) [DOI: 10.19630/j.cnki.tdkc.202201140001]

Howard A, Sandler M, Chen B, Wang W J, Chen L C, Tan M X, Chu G, Vasudevan V, Zhu Y K, Pang R M, Adam H and Le Q. 2019. Searching for MobileNetV3//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South): IEEE: 1314-1324 [DOI: 10.1109/ICCV.2019.00140]

Howard A G, Zhu M L, Chen B, Kalenichenko D, Wang W J, Weyand T, Andreetto M and Adam H. 2017. Mobilenets: efficient convolutional neural networks for mobile vision applications [EB/OL]. [2024-04-07]. https://arxiv.org/pdf/1704.04861.pdf

Islam M A, Jia S and Bruce N D B. 2020. How much position information do convolutional neural networks encode? [EB/OL]. [2024-04-07]. https://arxiv.org/pdf/2001.08248.pdf

Li D L. 2018. Structural Design and Simulation Analysis of the Upper-Car System of the Road-Railway Dual-Use Basket-Type Bridge Inspection Vehicle. Qingdao: Qingdao University of Technology (李丹乐. 2018. 公铁两用吊篮式桥梁检测车上装系统的结构开发与仿真研究. 青岛: 青岛理工大学)

Li G, Zhao X X, Du K, Ru F and Zhang Y B. 2017. Recognition and evaluation of bridge cracks with modified active contour model and greedy search-based support vector machine. Automation in Construction, 78: 51-61 [DOI: 10.1016/j.autcon.2017.01.019]

Liang X. 2019. Image-based post-disaster inspection of reinforced concrete bridge systems using deep learning with Bayesian optimiza-

tion. Computer-Aided Civil and Infrastructure Engineering, 34(5): 415-430 [DOI: 10.1111/mice.12425]

Liu Z, Lin Y T, Cao Y, Hu H, Wei Y X, Zhang Z, Lin S and Guo B N. 2021. Swin Transformer: hierarchical vision Transformer using shifted windows//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, Canada: IEEE: 9992-10002 [DOI: 10.1109/ICCV48922.2021.00986]

Peng M K. 2016. Discussion on improving the maintenance and management efficiency of electromechanical system equipmenturban subways. Architectural Engineering Technology and Design, 2016(23): 1588-1588 (彭孟开. 2016. 提高城市地铁机电系统设备的维护管理效率的探讨. 建筑工程技术与设计, 2016(23): 1588-1588) [DOI: 10.3969/j.issn.2095-6630.2016.23.556]

Ronneberger O, Fischer P and Brox T. 2015. U-net: convolutional networks for biomedical image segmentation//Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention——MICCAI 2015. Munich, Germany: Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]

Sandler M, Howard A, Zhu M L, Zhmoginov A and Chen L C. 2018. MobileNetV2: inverted residuals and linear bottlenecks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, USA: IEEE: 4510-4520 [DOI: 10.1109/CVPR.2018.00474]

Shelhamer E, Long J and Darrell T. 2017. Fully convolutional networks for semantic segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39 (4): 640-651 [DOI: 10.1109/TPAMI.2016.2572683]

Shi Y, Cui L M, Qi Z Q, Meng F and Chen Z S. 2016. Automatic road crack detection using random structured forests. IEEE Transactions on Intelligent Transportation Systems, 17(12): 3434-3445 [DOI: 10.1109/TITS.2016.2552248]

Sun L J, Zhang Y S, Wang W J and Zhao J. 2022. Lightweight semantic segmentation network for RGB-D image based on attention mechanism. Packaging Engineering, 43(3): 264-273 (孙刘杰, 张煜森, 王文举, 赵进. 2022. 基于注意力机制的轻量级RGB-D图像语义分割网络. 包装工程, 43(3): 264-273) [DOI: 10.19554/j.cnki.1001-3563.2022.03.033]

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser L and Polosukhin I. 2023. Attention is all you need [EB/OL]. [2024-04-07]. https://arxiv.org/pdf/1706.03762.pdf

Wang B K, Wang R L, Chen J J, Pan Y and Wang L J. 2023. Automatic detection method for surface diseases of shield tunnel based on deep learning. Journal of Shanghai Jiaotong University, 1-16 (王宝坤, 王如路, 陈锦剑, 潘越, 王鲁杰. 2023. 基于深度学习的盾构隧道表观病害自动检测方法. 上海交通大学学报 [DOI: 10.16183/j.cnki.jsjtu.2023.089]

Wang W H, Xie E Z, Li X, Fan D P, Song K T, Liang D, Lu T, Luo P and Shao L. 2021. Pyramid vision Transformer: a versatile backbone for dense prediction without convolutions//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, Canada: IEEE: 548-558 [DOI: 10.1109/ICCV48922.2021.00061]

Woo S, Debnath S, Hu R H, Chen X L, Liu Z, Kweon I S and Xie S N. 2023. Convnext v2: co-designing and scaling convnets with masked autoencoders//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE: 16133-16142 [DOI: 10.1109/CVPR52729.2023.01548]

Woo S, Park J, Lee J Y and Kweon I S. 2018. CBAM: convolutional block attention module//Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich, Germany: Springer: 3-19 [DOI: 10.1007/978-3-030-01234-2_1]

Xie E Z, Wang W H, Yu Z D, Anandkumar A, Alvarez J M and Luo P. 2021. SegFormer: simple and efficient design for semantic segmentation with transformers//Proceedings of the 35th International Conference on Neural Information Processing Systems. Virtual, Online: Curran Associates Inc.: 12077-12090

Xie S N, Girshick R, Dollár P, Tu Z W and He K M. 2017. Aggregated residual transformations for deep neural networks//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE: 5987-5995 [DOI: 10.1109/CVPR.2017.634]

Xu W B, Liu Y and Zhang H W. 2017. Research progress in image segmentation based on region growing. Beijing Biomedical Engineering, 36(3): 317-322 (徐蔚波, 刘颖, 章浩伟. 2017. 基于区域生长的图像分割研究进展. 北京生物医学工程, 36(3): 317-322) [DOI: 10.3969/j.issn.1002-3208.2017.03.16]

Xu X Y. 2022. Design of Tunnel Crack Detection and Image Processing System based on Support Vector Machine. Hangzhou: China Jiliang University (许霄煜. 2022. 基于支持向量机的隧洞裂缝检测与图像处理系统设计. 杭州: 中国计量大学) [DOI: 10.27819/d.cnki.gzgjl.2022.000164]

Yang M K, Yu K, Zhang C, Li Z W and Yang K Y. 2018. DenseASPP for semantic segmentation in street scenes//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, USA: IEEE: 3684-3692 [DOI: 10.1109/CVPR.2018.00388]

Zhang J Q, Jin J, Wang Y F, Zhou W and Liu Z C. 2023. Study progress of intelligent inspection technology and equipment for highway bridge. Journal of Highway and Transportation Research and Development, 40(1): 1-27, 58 (张劲泉, 晋杰, 汪云峰, 周炜, 刘智超. 2023. 公路桥梁智能检测技术与装备研究进展. 公路交通科技, 40(1): 1-27, 58) [DOI: 10.3969/j.issn.1002-0268.2023.01.001]

Zhao H S, Shi J P, Qi X J, Wang X G and Jia J Y. 2017. Pyramid scene parsing network//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE: 6230-6239 [DOI: 10.1109/CVPR.2017.660]

Zhou Z, Yan L B, Zhang J J, Zhang D M and Gong C J. 2024. Intelli-

gent detection of tunnel lining cracks based on self-attention mechanism and convolution neural network. Journal of the China Railway Society, 46(9): 182-192 (周中, 闫龙宾, 张俊杰, 张东明, 龚琛杰. 2024. 基于自注意力机制与卷积神经网络的隧道衬砌裂缝智能检测. 铁道学报, 46(9): 182-192) [DOI: 10.3969/j.issn.1001-8360.2024.09.021]

Zhou Z, Zhang J J and Gong C J. 2022. Automatic detection method of tunnel lining multi-defects via an enhanced you only look once network. Computer-Aided Civil and Infrastructure Engineering, 37(6): 762-780 [DOI: 10.1111/mice.12836]

Zhu L, Wang X J, Ke Z H, Zhang W and Lau R. 2023. BiFormer: vision transformer with bi-level routing attention//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE: 10323-10333 [DOI: 10.1109/CVPR52729.2023.00995]

作者简介

黄志海,男,硕士研究生,主要研究方向为基于深度学习的场景关键地物识别。E-mail:2112209044@mail2.gdut.edu.cn
郭波,通信作者,男,副教授,主要研究方向为三维激光扫描数据分类及地物建模。E-mail:guobo@gdut.edu.cn
罗海涛,男,高级工程师,主要研究方向为基于多信息融合的隧道病害提取分析。E-mail:luohaitao@dtsjy.com