

# 基于深度学习的无人机遥感影像灾损评估

韩东哲<sup>1</sup> 鲁王泽<sup>1</sup>

HAN Dongzhe LU Wangze

## 摘要

频发自然灾害威胁着人类的生命和财产安全,无人机机动灵活,采集的遥感图像分辨率高,结合深度学习图像分割技术可以高效地识别受灾目标、评估灾害损失。受灾后目标外观会发生很大改变,尤其是建筑物,受灾后整体结构被破坏,与周围的碎片、树木、道路等混淆在一起,难以有效区分各类目标。为此,提出了一个编解码结构模型 UDANet 旨在增强目标间的长距离依赖关系,扩大感受野以提高目标识别精度。首先,UDANet 模型编码端采用 ResNet50 作为特征提取器,在最深层特征处连接空洞卷积金字塔模块 ASPP,在浅层特征处使用空间注意力机制 PAM。然后,解码端对特定的浅层和深层进行融合后解码得到最终预测结果。最后,使用 RescueNet 数据集进行实验,实验结果 Pix\_Acuray 达到 83.6%。

## 关键词

无人机遥感图像;图像分割;灾损评估;深度学习;注意力机制

doi: 10.3969/j.issn.1672-9528.2024.01.004

## 0 引言

近年来世界各地自然灾害频发,给许多国家和人民带来了经济损失和人身伤害。如何高效识别受灾目标、评估自然灾害带来的损失,是制定应急救援预案的关键一环。受地形限制和灾害连续性的影响,救援人员不能及时进入受灾区域,但可借助遥感图像技术获取灾害信息。卫星、航空飞机和无人机都是有效获取受灾区域影像的可行手段。由于无人机遥感图像获取方式灵活、不受地形限制、速度快、周期短、可用成像手段多等特点,无人机遥感影像技术在灾害损失评估问题上得到广泛的应用。

随着无人机遥感设备的改进和摄影测量技术的成熟,获取到的无人机光学遥感图像有着超高分辨率,目标细节丰富且清晰。遥感图像拍摄幅面较广使得图像背景复杂,包含目标类别众多且尺寸多变,差异很大。以往基于人工设计特征规则提取目标的方法并不适用于超高分辨率无人机遥感图像。近年来随着计算机硬件性能的提升和计算能力突破,深度学习方法得到蓬勃发展,卷积神经网络模型可以自动学习目标特征,更准确识别多种目标、更易泛化。因此可以借助深度学习技术处理高分辨率无人机遥感图像。

目前,基于深度学习的图像处理技术主要有三大方向:分别是图像分类、目标检测、图像分割。其中,图像分割按最小分割单元可分为:像素级语义分割、对象级实例分割、

场景级全景分割<sup>[1]</sup>。图像语义分割实质是逐像素点进行分割,常用的图像语义分割方法大致分为三种:基于 FCN 全卷积神经网络的方法、基于 UNet 编解码结构的方法和基于 DeepLab 空洞卷积金字塔的方法<sup>[2]</sup>。

FCN 模型<sup>[3]</sup>是基于 CNN 图像分割时存在的固定尺度问题改进的网络,去掉了所有的全连接层,使得模型可以接受任意尺寸的图像输入,大大降低了语义分割网络的使用限制。初始时,模型直接对最深层特征图进行 32 倍上采样,恢复图像原始分辨率。但经过不断的池化下采样之后,最后一层特征图有效信息太少,导致模型分割精度不高。FCN 首先采用逐点相加的方式,融合来自第四层输出的特征图进行 16 倍上采样,然后又融合来自第三层输出的特征图进行 8 倍上采样,最终使模型达到较好的分割效果。

U-Net 模型<sup>[4]</sup>是基于 FCN 简单上采样过程出现的细节丢失问题而改进的网络,是对称式编解码器结构。编码端使用 VGG 网络提取图像特征,解码器端采用反卷积恢复图像尺寸,并通过跳跃连接将编码端提取到的各层特征图与高层特征进行通道拼接,这一策略可以使各层特征信息完整传递进来,抵消编码端池化操作引起的细节损失。后来研究者们对 U-Net 模型改进核心在于跳跃连接机制。U-Net++ 模型<sup>[5]</sup>提出在每层之间加满节点,复合长短跳连接,借助深监督自由选择下采样层数,以满足不同分辨率图像对网络深浅层次的不同要求。U-Net3+ 模型<sup>[6]</sup>采用全尺度的跳跃连接,通过密集叠加上下文信息增加空间位置感知以增强边缘信息,但是密集连接的思想会使得模型参数量增大较多,在实际训练

1. 防灾科技学院 河北廊坊 065201

[基金项目] 国家自然科学基金项目“遥感图像中基于深度学习网络的自然灾害破坏程度评估”(42007422)

时耗时且占用内存也比较多。U-Net 另一个演化思路是基于编码端特征提取结构的改进。ResUNet 模型<sup>[7]</sup>提出卷积模块使用残差结构，明显增加了网络的深度。ResUNet++ 模型<sup>[8]</sup>提出将金字塔模块 ASPP 融合到编码器末端，增加多尺度的上下文信息。U-Net 系列的模型，在结构较均匀、语义信息简单的图像上有着较高的分割精度。

DeepLab 系列模型是基于 FCN 提取特征不充分和边缘轮廓等空间位置信息细节丢失问题改进的网络，以期扩大感受野同时不降低图像分辨率。DeepLabv1 模型<sup>[9]</sup>提出用空洞卷积替换掉传统卷积来扩大感受野。但是，空洞卷积本身存在理论缺陷的，即棋盘问题。使用空洞卷积使邻近像素之间缺乏依赖，容易造成局部信息丢失。DeepLabv2 模型<sup>[10]</sup>设计了空洞卷积金字塔模块(ASPP)，并行设置不同扩张率空洞卷积，得到不同大小的感受野，使模型获得多尺度的特征，同时大的扩张率丢失的局部信息，可由小扩张率空洞卷积来补充，克服了棋盘问题。然而，过大的扩张率会使一次空洞卷积的有效覆盖像素减小到 1，这是与 ASPP 设计的初衷相背离的。DeepLabv3 模型<sup>[11]</sup>给出的改进措施，去掉过大的扩张率，并行增加一个 1\*1 卷积模块，减少参数个数，同时，并行增加一个全局平均池化模块 image pooling，补充全局特征信息。DeepLabv3+ 模型<sup>[12]</sup>增加了一个由浅层到上采样层的 skip 跳跃连接，以补充特征图细节信息，由于这种空洞卷积特征金字塔模块的设计，DeepLab 系列的模型在背景复杂、目标尺寸差异比较大的图像中，有着较高的分割效果。

综合分析，DeepLab 系列核心是在下采样阶段借助空洞卷积扩大感受野，并使用 ASPP 提取多个尺度的特征，且使用线性插值法上采样，同时融合特定的浅层和深层的特征。U-Net 系列则是使用反卷积法进行上采样和逐层的上下文信息融合，且融合了所有分支的特征图。现有研究证明，这两个系列的分割模型在大多数场景中都是很有效的。因此，本文尝试将 U-Net 系列的模型和 DeepLab 系列的模型的优势融合，设计了 UDANet 模型，使其可以更好地应用在无人机遥感图像分割任务中。

首先，通常情况下浅层特征包含较多的空间细节信息，深层特征包括更多的语义信息，不同层次间特征图语义信息差异较大。无人机遥感图像分辨率高，受灾后的细节信息比较丰富，浅层包含不仅是目标信息，还有许多杂乱物体

的干扰信息。因此，本文先探究了深层特征和浅层特征的融合比例，实验确定去掉第 1 分支浅层特征，仅融合第 2、3、4 分支的输出特征。

其次，理论上深层的感受野相比浅层要大，看到的空间位置信息更多，但实际上由深度增加的感受野是远远不够的。而且灾后无人机遥感图像部分目标结构被破坏，混淆在一起，难以有效区分开。本文希望模型关注到更多的不同尺度的像素依赖关系以强化目标相关特征。因此，本文在最深层特征（第四分支输出特征）处连接空洞卷积金字塔模块 ASPP，扩大了感受野，同时提升模型对多尺寸目标的适应能力。

最后，为适当减轻浅层杂乱细节的干扰作用，本文在第二分支特征处使用空间自注意力机制 PAM 来关注更多与目标相关的重要信息。

综上所述，本文提出了一个编码解码结构模型 UDANet，编码端采用预训练的 ResNet50 提取特征，首先在最深层特征处连接一个空洞卷积金字塔模块 ASPP，在浅层特征处使用空间注意力机制 PAM。然后解码端使用线性插值法进行上采样以及使用通道拼接法逐层融合了三个尺度的特征图，得到最终输出结果。最后，模型使用 RescueNet 数据集进行实验，实验结果 Pix\_Acuray 达到 83.6%，MIou 值达到 64.3%。

## 1 方法

### 1.1 模型结构

UDANet 模型整体结构如图 1 所示。编码端使用预训练的 ResNet50 提取特征，在第四输出分支处连接空洞卷积金字塔结构的模块 ASPP；在第三输出分支处直接使用两个堆叠的 3\*3 卷积压缩通道；在第二输出分支处连接空间注意力机制模块 PAM。解码端，对高层特征进行线性插值上采样后使用通道拼接的方式融合相邻低层的特征图，再使用两个堆叠的 3\*3 卷积进一步增强特征融合效果。解码端逐层融入 2、3、4 分支的特征图后，使用 1\*1 卷积直接压缩通道数到目标类别数，再一步上采样到原始图像尺寸，输入分类器 Softmax 进行最终的类别预测。

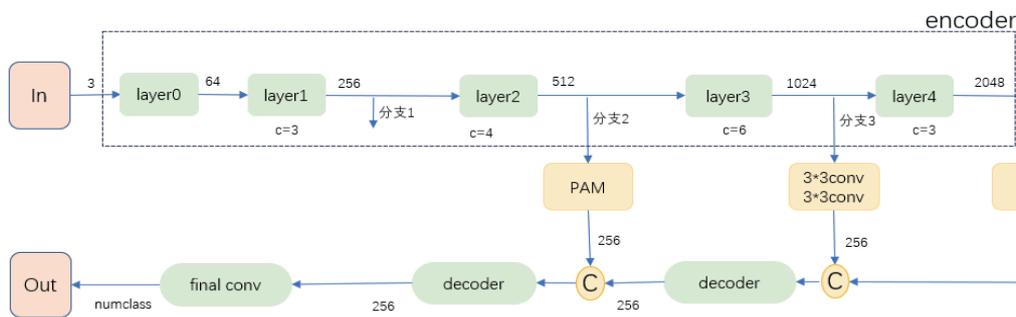


图 1 UDANet 结构示意图 (C 表示通道拼接)

1.2 空洞空间金字塔模块 ASPP

由于标准卷积的感受野是固定的，仅能看到局限的临近像素依赖关系，这使得模型难以建模长距离的依赖关系，而对于相似易混淆的目标来说，增加全局上下文信息是很有帮助的。因此，在第四分支处引入了一个能关注全局信息的多尺度的特征提取模块 ASPP，如图 2 所示。ASPP 是空洞卷积金字塔结构，包含 5 个分支。其中 1\*1 卷积分支保留模型原有的感受野。空洞卷积分支的膨胀率设置为 6、12、18，获得多个尺度的大感受野。全局平均池化分支旨在获得全局特征。将所有分支进行通道拼接后使用 1\*1 卷积压缩通道数，整合重要特征信息。

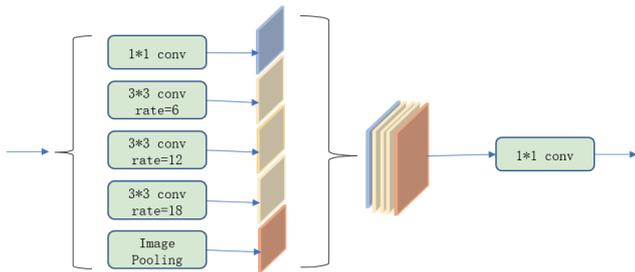


图 2 ASPP 结构示意图

1.3 空间自注意力机制模块 PAM

空间自注意力机制 position attention module(PAM) 整体结构如图 3 所示，计算公式可表示为：

$$E_j = \alpha \sum_{i=1}^N (s_{ij} D_i) + A_j \quad (1)$$

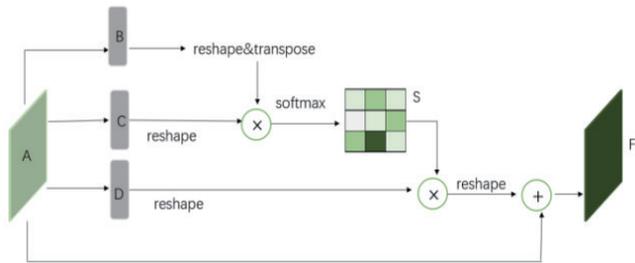


图 3 PAM 结构示意图

特征图 A 首先经过 3 个卷积得到三个特征图 B、C、D，再将 B、C、D 转换为二维向量。其次将 B 转置后与 C 相乘，送入 Softmax 计算得到空间位置注意力图 S，图中每个点表示的是 A 中两个像素点的空间位置关系强度。然后将 S 与 D 相乘后再转换为原来的形状，又乘了一个尺度系数，在训练中从 0 开始学习。最后将得到的矩阵图与原特征图 A 相加后输出最终特征图 F。

2 实验

2.1 实验环境

硬件环境：CPU 为 Intel(R) Xeon(R) Gold 6230R CPU @ 2.10 GHz，内存为 252 GB，GPU 为 RTX3090。软件环境：操作系统为 Ubuntu20.04，64 位，深度学习框架使用的是 PyTorch1.8，cuda 版本为 11.4，Python 环境为 3.7。训练参数：图像输入尺寸 713×713 像素，batchsize=4，采用预训练 ResNet50 权重文件初始化，使用交叉熵损失函数，使用随机梯度下降 SGD 优化器，lr=0.001，momentum=0.9，学习率调度函数 poly，epoch=100。

2.2 数据集

本文实验选用公开的无人机遥感数据 RescueNet<sup>[13]</sup>。该数据集由马里兰大学发布，使用小型无人机平台 DJI Mavic Pro 四轴飞行器收集的，拍摄于飓风灾后，有部分地区同时受到洪灾，分辨率为 4000 像素×3000 像素，共有 4494 张图像。数据集人工标注了 12 个类别，语义分割标签包括：1 背景、2 碎片、3 水、4 未受损建筑、5 中等受损建筑物、6 主体受损建筑物、7 完全损坏建筑物、8 车辆、9 道路、10 树、11 池塘、12 沙子。

2.3 评价指标

实验中，选用  $M_{IoU}$  作为结果评价指标。 $M_{IoU}$  (mean intersection over union) 是衡量语义分割任务中模型性能的一种指标，它是预测像素标签和真实像素标签的交集面积与它们的并集面积的比值，即：

$$M_{IoU} = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - p_{ii}} \quad (2)$$

式中： $p_{ij}$  表示真实值为  $i$ ，被预测为  $j$  的数量， $K+1$  是类别个数（包含空类）。 $M_{IoU}$  表示模型在所有像素上的平均像素级别准确度，它的取值范围在 [0,1] 之间，值越高表示模型的性能越好。

2.4 实验结果及分析

UNet 解码融合第 1、2、3、4 层四个分支的特征图，DeepLab 只融合了第 1、4 层 2 个分支特征图，从表 1 对比可知，使用 50 层的 ResNet 提取特征时，UNet 因融合特征层次较多，大部分类别的识别精度都比 DeepLab 较高。

表 1 常见模型预测结果

常见模型	结果		class IoU											
	Pixel Accuracy	MIou	0	1	2	3	4	5	6	7	8	9	10	11
			背景	碎片	水体	未受损	中等受损	严重受损	完全受损	汽车	道路	树	池塘	沙子
resnet50_unet	81.6	58.99	46.1	62.9	80.7	61.7	48.7	45.3	57.1	51.1	72.4	79.3	28.1	75.2
resnet101_unet	82.0	57.70	45.0	63.4	81.5	63.1	50.1	46.2	58.2	53.3	72.9	80.0	3.7	74.8
resnet50_deeplab	81.4	57.49	52.5	64.2	77.2	55.4	45.4	42.3	52.4	45.7	69.6	79.4	32	73.6
resnet101_deeplab	81.6	58.60	53.6	63.6	76.6	59.0	48.7	45.0	55.4	46.6	71.2	79.1	29.5	74.5

但是, 101 层的 ResNet50\_UNet 明显丢掉了池塘这类目标的信息, 这类目标特点是尺寸小, 目标总数少, 而 DeepLab101 对该类别的识别率比较稳定。为探究这一问题, 实验中首先对主干网络进行了一些细节的调整, 消融实验中没有使用 ASPP 等强化模块, 仅保留编码器和解码器, 调整结果见表 2。从表 2 分析, 线性插值上采样效果要优于转置卷积上采样, 直接对 ResUNet 编码器施加空洞卷积效果会变差, 但仍优于 ResDeepLab 模型。因此可知每个层次的特征图语义信息是有差别的, 互为辅助, 进行多层次的特征融合是必须的。但是需要对融合哪几个层次的特征进行微调, 调整结果见表 3。

从表 3 可知, 在保留第 1 分支特征图前提下, 增加第 2 或者第 3 层的特征信息, 分割结果变化不大。经可视化分割结果分析, 第 1 浅层特征图存在较多的与目标无关的细节, 冗余特征过多, 影响了模型的训练。因此, 实验中又去掉将第 1 浅层特征图, 仅融合第 2、3、4 分支的特征图, 识别效果明显提高。这是针对该数据集分割任务的最优特征融合方式。

综上, 本文的 UDANet 模型在融合了第 2、3、4 分支的特征后, 对第 4 分支使用 ASPP 增加多尺度特征信息, 对第 2 分支使用了空间注意力机制 PAM, 使模型尽可能

表 2 消融实验 1

常见模型	结果		class Iou											
	Pixel_Accuracy	MIou	0	1	2	3	4	5	6	7	8	9	10	11
			背景	碎片	水体	未受损	中等受损	严重受损	完全受损	汽车	道路	树	池塘	沙子
resnet50_unet(1234 分支)	81.6	58.99	46.1	62.9	80.7	61.7	48.7	45.3	57.1	51.1	72.4	79.3	28.1	75.2
resnet50_unet (反卷积) layer4: dilation=2	80.8	56.4	52.7	63.1	74.1	57.7	44.8	42.4	55	49.9	71.2	77.7	15.4	73.2
resnet50_unet (线性插值) layer4: dilation=2	82.0	57.70	45.0	63.4	81.5	63.1	50.1	46.2	58.2	53.3	72.9	72.9	34.8	74.8
resnet50_deeplab (14 分支)	80.8	56.80	50.4	62.1	77.5	56.3	45.1	41.1	53.1	44.3	70.7	78.8	28.3	74.1
resnet50_deeplab (线性插值) layer4: dilation=0 保留第一分支 48 通道特征图	81.0	59.6	34.9	61.8	80.4	62.1	50.5	47.1	58.9	48.2	71.6	79.5	46.2	74.6
resnet50_deeplab (线性插值) layer4: dilation=0 保留第一分支 256 通道特征图	81.2	61.1	42.2	62.1	80.5	61.0	51.0	47.8	59.3	48.5	72.3	78.9	55.7	74.6

表 3 消融实验 2

常见模型	结果		class Iou											
	Pixel_Accuracy	MIou	0	1	2	3	4	5	6	7	8	9	10	11
			背景	碎片	水体	未受损	中等受损	严重受损	完全受损	汽车	道路	树	池塘	沙子
分支 1,4	81.2	61.1	42.2	62.1	80.5	61.0	51	47.8	59.3	48.5	72.3	78.9	55.7	74.6
分支 1,2,4	81.4	61.9	46.8	62.1	80.0	62.9	49.8	48.0	59.2	52.9	72.3	79.2	55.1	74.0
分支 1,3,4	81.0	60.1	47.0	62.3	78.4	60.8	47.5	45.1	56.3	50.8	72.1	78.6	55.8	73.7
分支 1,2,3,4	81.5	61.9	42.8	62.8	79.9	63.1	50.8	48.1	57.3	55.1	72.8	79.7	56	74.2
分支 2,3,4	82.4	63.2	46.2	64.0	80.7	63.0	52.2	49.8	60	54.5	73.5	80.1	58	75.9
分支 2,4	81.3	62.2	42.9	61.9	80.1	63.3	50.4	48.5	59.7	53.3	72.3	78.5	59.9	75.3
分支 2,3	80.5	60.6	52.2	62.7	79	60.1	48.7	45.5	54.6	50	71.6	78.9	51.2	72.2
分支 2,3,4+ASPP	82.7	64.0	52.4	64.7	80.9	63.4	53.1	51.3	60.7	56.3	73.3	80.0	58.7	76.2
分支 2+PAM,3,4+ASPP (ours)	83.6	64.3	54.8	67.1	80.3	64.2	53.6	51.6	58.6	51.1	73.4	81.1	59.5	76.4

从表 2 可知, ResDeepLab 模型在编码端提取特征时去掉空洞卷积后, 识别效果明显提升。同时解码端保留第一分支全部的 256 个通道的特征图, 这个微调结果是最优的, 进一步验证了模型预测需要用到浅层特征图补充细节信息。因此, 之后的所有实验统一设置为去掉空洞卷积, 使用线性插值上采样, 并通过两个堆叠的 3\*3 卷积块, 保留各个特征分支的 256 个通道。

关注到与目标相关的信息, 而不被无关信息干扰。最终, UDANet 预测结果 Pix\_Accuracy 达到 83.6%, MIou 值达到 64.3%。分割结果如图 4 所示。从图 4 可视化结果看, UDANet 对受损的建筑物识别效果有所提升。道路的边缘也明显变得清晰。同时, UDANet 对于受损的位置也能较清晰地展示出来。整体分割结果可以作为灾损程度评估的一个依据。

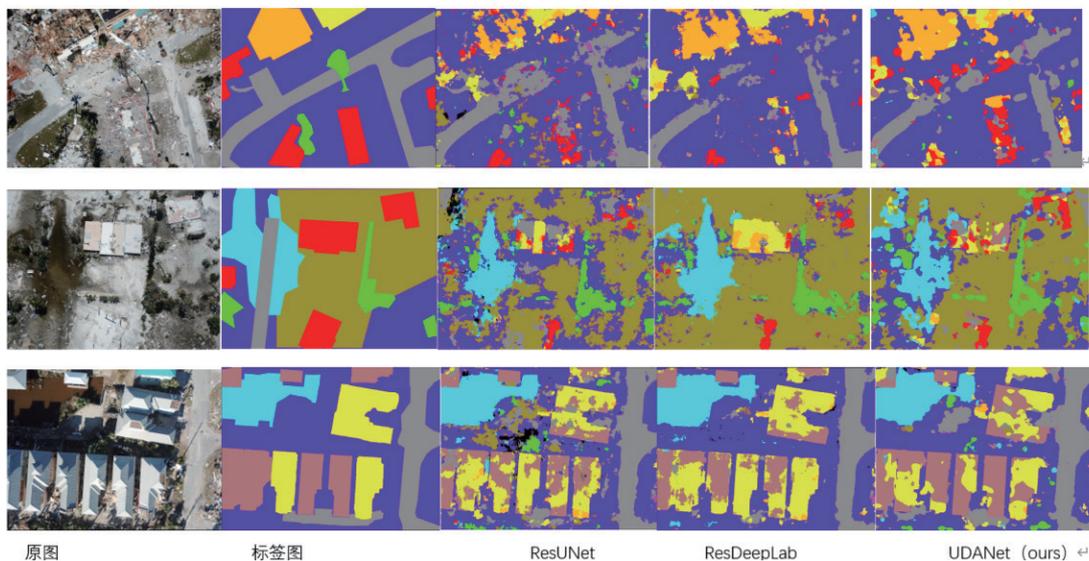


图4 分割结果对比

#### 4 结语

本文提出的UDANet模型,是希望能融合UNet和DeepLab的优势,对主干网络的一些细节进行了微调,重点放在对每层输出的特征进行强化处理和融合上。但是,浅层特征与深层特征包含的语义信息差异很大,在融合特征时如何借助可视化特征图手段,调整特征融合方式以减少误分割现象,还有待探究。

#### 参考文献:

[1] 刘小波. 基于深度学习的光学遥感图像目标检测研究进展[J]. 自动化学报, 2021,47(9):2078-2089.

[2] 丁婷婷, 于晓鹏, 李紫薇. 基于神经网络的遥感图像分割方法研究[J]. 智能计算机与应用, 2020,58(5):2095-2163.

[3] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation[J]. IEEE transactions on pattern analysis and machine intelligence, 2015,39(4): 640-651.

[4] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation[C]//MICCAI 2015-Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2015: 234-241.

[5] ZHOU Z. UNet++: a nested U-Net architecture for medical image segmentation[C]//Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Cham: Springer, 2018: 3-11.

[6] HUANG H, LIN L, TONG R, et al. UNet 3+: a full-scale connected UNet for medical image segmentation[C]//IEEE International Conference on Acoustics, Speech and Signal

Processing. Piscataway: IEEE, 2020: 1055-1059.

[7] ZHANG Z, LIU Q, WANG Y. Road extraction by deep residual U-Net[J]. IEEE geoscience and remote sensing letters, 2018,15(5): 749-753.

[8] JHA D, SMEDS-RUDPH, RIEGLER M A,

et al. ResUNet++: an advanced architecture for medical image segmentation[C]//2019 IEEE International Symposium on Multimedia (ISM). Piscataway: IEEE, 2019: 225-255.

[9] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Semantic image segmentation with deep convolutional nets and fully connected CRFs[C]//International Conference on Learning Representations - ICLR 2015. Piscataway: IEEE, 2015: 215-220.

[10] 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.

[11] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Rethinking atrous convolution for semantic image segmentation[J]. Computer vision and pattern recognition, 2017(17): 22-26.

[12] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Encoder-decoder with atrous separable convolution for semantic image segmentations[C]//European Conference on Computer Vision. Cham: Springer, 2018: 833-851.

[13] CHOWDHURY T, MURPHY R, TAHNEMOONFAR M. RescueNet: a high resolution UAV semantic segmentation benchmark dataset for natural disaster damage assessment[J]. Journal of latex class files, 2021, 14(8): 125-131.

#### 【作者简介】

鲁东哲(1991—),女,河北石家庄人,硕士,研究方向:灾害信息处理技术。

鲁王泽(2001—),男,安徽铜陵人,硕士,研究方向:灾害信息处理技术。

(收稿日期: 2023-09-21)