

局部和非局部特征的深度融合 滑坡精确识别

朱清, 陈林, 胡涵, 徐斌志, 张业婷, 李海峰

摘要- 滑坡库存量的精确测绘对于减灾至关重要。大多数滑坡通常与其他令人困惑的地质特征共存, 只有在大范围内才能明确推断出这些地区的存在。此外, 局部信息对于保存物体边界也很重要。针对这一问题, 本文提出了一种融合局部和非局部特征来克服上下文问题的有效方法。在遥感界广泛采用的 U-Net 架构的基础上, 我们利用了两个额外的模块。第一个使用扩展卷积和相应的空间金字塔池, 在不牺牲空间分辨率或增加内存使用的情况下扩大了接受野。第二个使用尺度注意机制, 通过学习的权重图来引导特征从粗层上采样。在实现中, 对原始 U-Net 的计算开销只有几个卷积层。实验评估显示, 所提出的方法优于最先进的通用语义分割方法。此外, 消融研究表明, 这两种模型在滑坡识别性能方面提供了广泛的增强。

索引术语: 滑坡制图, U-Net, 注意力, 扩展卷积

我的介绍。

L 滑坡是最具破坏性的自然灾害之一, 也可能引起一系列次生灾害,

比如堰塞湖溢流和大坝溃决引发的洪水

[1]。由于引起滑坡的因素复杂, 且滑坡易发地区在连续降雨期间或之后会突然发生滑坡, 滑坡易感性的灾前测绘不足以进行灾害管理

[2]。有效和精确的灾后滑坡区域测绘对于成功的应急响应以及预测由于下垫面不稳定而可能发生的二次滑坡也至关重要。

收稿日期: 2020 年 2 月 21 日。本文得到国家自然科学基金面上部分支持 (项目 No. 5); 项目资助: 41941019, 41871291, 41871314, 国家重点研发计划项目 (项目编号: 41871314); 2018 yfb0505404)。 (互通讯作者: 胡涵、李海峰)

朱青、陈林、徐斌志就职于中国成都西南交通大学地球科学与环境工程学院。(电子邮件: zhuq66@263.net; chenlin@my.swjtu.edu.cn; 徐 (binzhi@my.swjtu.edu.cn)

胡涵就职于中国铁路第一勘察设计院西安轨道交通工程信息化国家重点实验室和西南交通大学成都地球科学与环境工程学院。(电子邮件: han.hu@swjtu.edu.cn)

张业婷, 武汉大学测绘与遥感信息工程国家重点实验室, 中国武汉 (e-mail: zhangyeting@whu.edu.cn)

李海峰, 中南大学地球科学与信息物理学院, 中国长沙 (e-mail: lihaifeng@csu.edu.cn)

目前, 许多滑坡测绘系统需要交互式解译, 这显然依赖于人类专家的经验, 因此不足以实现快速响应。经典的方法倾向于使用光学或雷达图像的光谱特征的特定领域知识来描绘和分割滑坡的图像, 例如纹理模式[3]、地形特征[3]、[4]和植被指数[5]。当使用高分辨率数据时, 基于对象的策略[6]也被广泛用于提高可靠性[4]、[5]、[7]、[8]、[2]。

随着深度学习范式的出现, 基于卷积神经网络(CNN)的方法在许多图像处理目标中产生了令人印象深刻的结果。具体来说, 全卷积网络(fcn)[9]使用反卷积模块实现了端到端分割。fcn 及其后继算法[10]、[11]极大地促进了图像语义分割的发展。然而, 在将此类技术应用于滑坡测绘之前, 仍有一些挑战有待克服。(1) **局部接受场**。基于 cnn 的特征只使用局部区域的信息, 而滑坡往往具有由背景特征(如道路和居民区)产生的混淆光谱信息, 在分割过程中需要大背景来消除这些模糊性。(2) **边界保存**。FCN 本质上是使用上采样步骤将低分辨率特征图恢复到原始分辨率, 这导致大量边界信息的丢失。虽然金字塔结构如 Deeplab 和 U-Net[10]、[12]可以传播和聚合来自不同尺度的信息, 但这个问题仍然存在。为了更好地保持边界, 也考虑了变化检测策略[13], 但在不存在预灾害清单的情况下, 该策略并不适用。

在本研究中, 我们提出并开发了一种基于 U-Net 结构的方法, 通过融合本地和非本地信息来缓解上述问题[12]。在下采样期间, 我们使用不同膨胀大小的亚特拉斯卷积[14], [11]来模拟多尺度特征, 然后通过最大池化操作进行下采样[12]。此外, 受 Deeplab 的启发, 我们在 U-Net 的瓶颈处增加了一个融合步骤来聚合多尺度特征[10]。在上采样步骤中, 受尺度注意模块[15], [16], 特别是对齐模型[17], [18]的启发, 我们用注意模块增强 U-Net, 通过学习权重图来抑制粗尺度中不相关和混淆的特征。这两种增量修正共同提高了仅使用灾后无人机(UAV)图像绘制滑坡区域的精度和鲁棒性。

2 方法

A. 问题设置和方法概述

我们将 UAV 图像中的灾后滑坡映射制定为语义分割问题[9], 这在计算机视觉界得到了广泛的研究。更具体地说, 给定一张图像 \mathbf{X} , 目的是为每个像素 i 分配一个二进制标签 l_i ($l \in \{0, 1\}$), 它组成了图像对背景和滑坡区域的二进制分割 \mathbf{b} 。目标是以端到端方式学习 $f \rightarrow \mathbf{b}$ 映射[9], 使用训练二值分割样本和相应的 UAV 图像。

受先前遥感图像语义分割工作的启发[19], 我们基于著名的 U-Net 结构[12]构建了系统, 该结构以多尺度方式融合了 CNN 地图。U-Net 还具有低内存配置文件, 这对于大规模遥感应用至关重要。此外, U-Net 可以使用编码器-解码器策略进行密集的语义预测。在编码器中, U-Net 既掌握了从更细的空间分辨率(较浅的通道)到更粗的空间分辨率(分别为更深的通道)的低级灰度和梯度特征, 又掌握了高级上下文特征。在解码器中, U-Net 将左侧部分的编码器特征连接到反卷积特征(图 1)。它使用上述跳过层实现了多尺度信息的融合, 从而大大提高了映射分辨率。

然而, 在我们对灾后滑坡区域的测绘中, 我们经常会遇到结构非常小和非常大的区域。这就提出了一个问题, 因为由于计算效率的原因, 不可能不断增加层数来扩大接受场以涵盖更大的对象。此外, 小特征或局部特征的光谱特征可能会被具有相似光谱信息的区域(如裸土或道路)所遮挡。

为了克服这一障碍, 本研究在 U-Net 中增加了两个模块:(1)ASPP(空间金字塔池)模块, 用于扩大感受野, 而不会在编码器中太深而失去空间分辨率;(2)注意模块, 通过在解码器的粗层次上利用非局部上下文信息来抑制不相关或混淆的特征。该架构的概述如图 1 所示。下面的小节将详细介绍这两种修改。

B. 心房空间金字塔池扩大的感受野

1)心房卷积:经典卷积本质上是一种局部方法, 它只能解释一个固定的区域, 依靠池化操作(如 max-pooling)来扩大接受域, 代价是更粗的空间分辨率。心房卷积[10], [14]已经被用来克服这个问题, 因为它在卷积核中增加了孔, 这可以有效地扩大接受野, 而不牺牲空间分辨率或效率, 而不是扩大卷积核的大小。孔的数量由膨胀率 d 控制, 带孔的核的大小 k_{sh} 也与有效核大小 k 相关, $k_{sh} = k + (k - 1) \times (d - 1)$, 这是亚特罗斯卷积方法的另一个优点

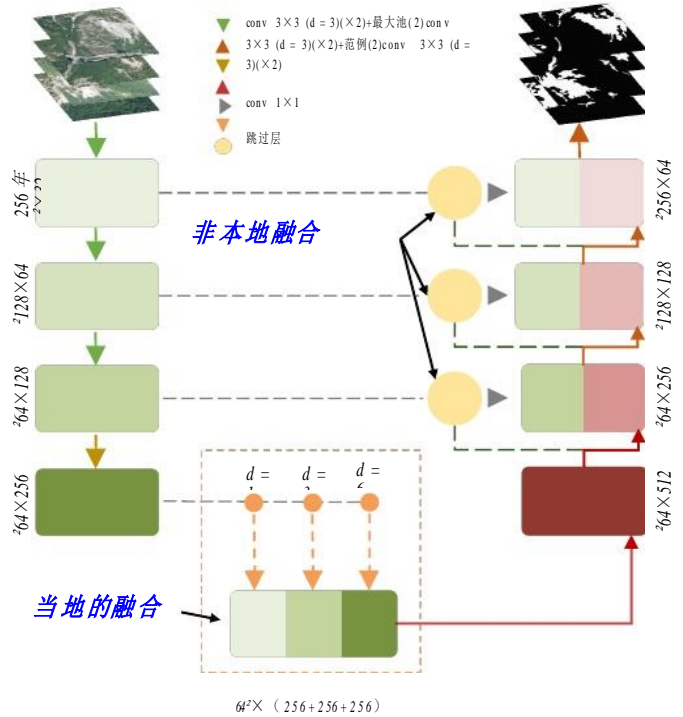


图 1: 滑坡制图的架构概述。

是通过在整个图像上滑动窗口来生成密集的特征地图。这一特性极大地简化了多尺度信息的融合。

2)亚特罗斯空间金字塔池化(Atrous spatial pyramid pooling):虽然亚特罗斯卷积可以解决感受场问题, 但如果扩张率在几个相关层上依次增加, 信息将不可避免地变得过于稀疏[14]。因此, 我们采用 ASPP 策略[10]来替代 U-Net 的瓶颈部分, 如图 1 所示。具体来说, 将融合记为

$$y = D_{3,1}(x) \oplus D_{3,3}(x) \oplus D_{3,6}(x), \quad (1)$$

其中 d, k , 表示膨胀率为 d 、核大小为 k 的属性卷积, \oplus 表示特征的拼接操作。通过与 1×1 核的卷积来调整融合特征的尺寸。通过使用 ASPP 融合, 接受野可以占整个磁片的四分之一以上。

C. 注意力增强上采样

注意机制起源于自然语言处理[20], 后来扩展到图像分类[17]和语义分割[15]、[21]、[16]、[22]。注意机制的前提是, 特征图的绝对值也揭示了它们的重要性;因此, 我们可以学习一个与特征图具有相同空间分辨率的权重图 $\alpha \in (0, 1)$, 以自适应地抑制不相关的特征。通常, 这样的权重图会考虑全局信息[15]、[22], 并且有两种模型(也称为兼容函数[17])可用, 即加性模型和乘法模型。双重注意网络[22]在通用语义分割中的成功激励了我们

将可加性模型扩展到 U-Net 结构中，即使用前人介绍的尺度注意模块

[15], [16].

具体来说，我们没有直接将来自编码器(即图 1 的左侧部分)的特征映射与来自解码器(即图 1 的右侧部分)的粗层上采样特征映射连接起来。相反，我们首先使用来自解码器的粗层特征来增强编码特征。由于粗糙层的特征包含更多的上下文信息，非局部信息有助于确定权重图。

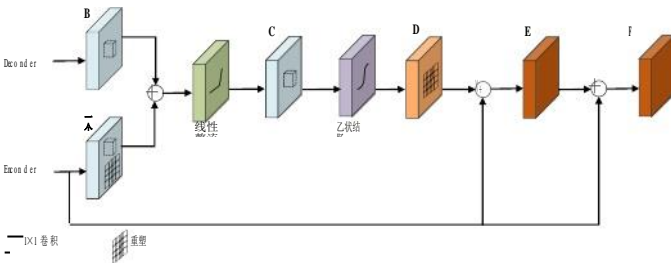


图 2:通过非局部融合来自较粗层次的特征来增强注意力的上采样。

图 2 为注意模块的放大表示，即图 1 中的圆圈。首先将输入编码器和解码器特征(即分别为 A 和 B)与 1×1 滤波器进行卷积，以使通道兼容;此外，还对编码器特征 A 进行了重塑，使其尺寸兼容。然后通过元素求和将两个地图连接起来，然后通过整流线性激活(ReLU)函数进行处理。然后，使用另一个只有 1 个通道的卷积来创建兼容性映射 C[17]。

有两个选项可用于创建注意图 α : $\text{softmax} \theta (x_i) = \text{Pe}^{\frac{x_i \exp \theta}{\sum_j x_j \exp \theta}}$ 和 $\text{sigmoid} \theta (x_i) = \frac{\exp \theta}{1 + \exp \theta}$

函数。虽然大多数方法都使用了 softmax

函数[15], [17], [22], 我们发现 softmax 函数中的归一化部分, 例如 $\sum_i \alpha_i = 1$, 会使 I 为 1

激活过于稀疏, 这对于多次应用的过程来说是不好的。因此, 使用 sigmoid 函数生成特征 D, 并重新采样到编码器特征 A 的空间分辨率, 类似于之前的工作[16]。注意增强特征 E 包括与权重 α 的元素明智乘法。最后, 在与上采样特征连接之前, 对地图 F 使用跳过连接[23], 如下所示:

$$F_i = A_i + \alpha_i A_i. \tag{2}$$

D. 实施细节

在 ArcMap 中以交互方式从原始 UAV 正射影片中获取标签, 并平铺成大小为 512×512 的片段。对于训练和测试, 图像被归一化为 $[-0.5, 0.5]$ 。由于滑坡主要位于森林中, 我们通过故意采样更令人困惑的区域(如裸露的土地、大岩石、道路和河流)来进行硬采矿。我们使用了随机翻转、旋转和缩放

对于数据增强过程。使用 Tensorflow 1.9 在一台带有 4 个 NVIDIA RTX Titan 图形计算单元(gpu)的机器上实现该框架。对于训练的超参数, 每个 GPU 的批大小为 12, 动量为 0.9, 学习率为 $1^{-3} e$, 正则化为 $5e^{-4}$ 。训练持续了 100000 次迭代, 我们每 20000 次迭代记录一次模型。选择测试性能最好的模型。最后, 结合二值分割损失函数, 使用 Softmax。在测试过程中, 对原始正射影像进行动态裁剪、加载和拼接。

3 实验评估

A. 实验设置和整体性能

2017 年 8 月 8 日, 中国九寨沟地震引发的山体滑坡获得了覆盖六个县的 UAV 图像。我们交互式地选择了 10^4 个瓦片进行训练, 并使用了 70 - 30% 的溢出。四架完整的 UAV 正射影像机用于测试, 其中包括大约 3000 块瓦片。使用三个常见指标来评估逐像素的结果, 即精度、召回率和 F1 score。

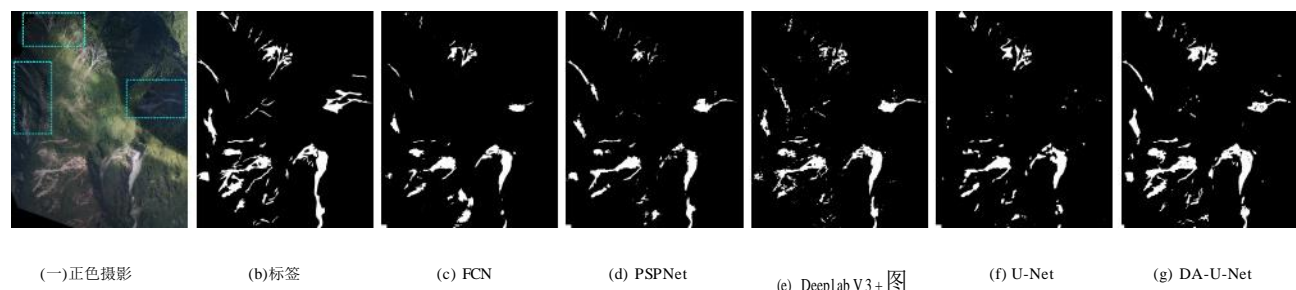
在下面, 我们使用前缀 “D” 来表示扩展卷积, “A” 来表示注意模块, 例如仅扩展卷积的 D-U-Net 和两个修改模块的 DA- U-Net。为了比较, 我们还重新实现了几种公开可用的方法, 例如以 VGG-16[24]为骨干的 FCN[9], 以 ResNet-101[23]为骨干的 PSPNet[25], 以 ResNet-101 为骨干的最新 DeepLabV3+[11]和 vanilla U-Net[12]。

图 3 比较了上述方法在评估一个典型场景时的性能。值得注意的是, 在阴影(日食区域)中, 几乎所有方法都无法识别滑坡区域;这种情况只能通过使用注意力模块, 例如 DA-U-Net 来改善。另一个有趣的发现是, 只有采用金字塔策略的建筑才能令人满意地识别出由滑坡和裸土(如矩形所示)组成的混乱区域中的滑坡区域。综上所述, 所提出的方法给出了最好的整体分割结果。

转向定量评估, 表 1 展示了所有方法的联合(IoU), 精度, 召回率和 f 分数的交集。所提出的 DA-U-Net 在最受关注的 IoU 度量以及召回率和 f 分方面的性能超过了第二好的方法, 即 DeeplabV3+[11]。考虑到目前最先进的 DeeplabV3+的性能也与 PSPNet 相当, 因此对 U-Net 的修改是有效的。

表 1:不同方法的定量比较。加粗的单元格表示性能最好的方法。

方法	精度	精度	回忆	f 值
FCN	48.15	75.96	56.81	65.00
U-Net	48.18	75.42	57.15	65.03
PSPNet	52.63	77.56	62.08	68.97
DeeplabV3+	57.25	79.6	67.1	72.81
DA-U-Net	59.41	70.06	79.62	74.54



3:与其他方法的定性比较。

B. 混淆区域的研究

为了有效地减轻灾害和迅速作出应急响应，需要精确探测的最关键的滑坡区域是那些发生在混乱地区的区域。由于训练中的大多数滑坡区域位于非混淆区域，因此我们通过选择偏好混淆区域的不平衡样本来优先考虑混淆区域。图 4 显示了两个典型的混淆区域，即道路(上)和裸露的土地(下)，每行中有趣的区域用青色多边形突出显示。由于如果不从大背景中进行推断，这些区域很难区分，因此如果不融合多尺度特征，几乎不可能识别出这些令人困惑的区域。不幸的是，fcn 没有处理多尺度特征的机制，因此预计性能会较差。U-Net 通过下采样传播和聚合有限的多尺度特征，这意味着它不能在不损失空间分辨率的情况下深入了解上下文信息。PSPNet 构建空间金字塔并融合不同层次，DeeplabV3 嵌入多尺度特征的 ASPP 模块;这两种方法都可以在不损失空间分辨率的情况下利用更大的上下文信息，但不能充分保留细粒度结构。由于采用了 ASPP 模块和注意引导上采样，所提出的 DA-U-Net 具有最好的性能。

C. 烧蚀研究

DA-U-Net 在传统 U-Net 的基础上增加了两个模块:1)瓶颈中的扩展卷积和 ASPP，用于开发更大的上下文信息;以及 2)用于引导上采样的注意力模块。表二比较了不同的 U-Net 变体，即 D-U-Net、A-U-Net 和 DA-U-Net。值得注意的是，可以看到，无论是扩张模块还是注意力模块，都清晰而广泛地提升了整体性能。

图 5 比较了三个混淆区域的不同 U-Net 架构。D-U-Net 和 A-U-Net 对受道路干扰的假滑坡区域的检测都有所提高，并且注意引导采样的效率高于膨胀模块和 ASPP 瓶颈。这也与表二所示的定量评价相一致。因此，结合起来，这两个模块展示了最佳的综合性能。

四. 结论

在本研究中，我们提出并开发了一种改进的滑坡库存测绘方法，该方法基于

表 2:基于 U- Net 的不同模块的烧蚀研究。加粗的单元格表示最佳性能。

方法	注意模块	扩展卷积+ASPP	借据
U-Net	×	×	48.18
D-U-Net	×	X	54.61
A-U-Net	X	×	52.44
DA-U-Net	X	X	59.41

增强 U-Net 结构，即扩张卷积[10]和注意引导的上采样[15]，[16]。由于滑坡或可能的滑坡区域通常与与滑坡具有相似光谱的区域共存，因此应考虑上下文信息以消除 cnn 产生的特征模糊性。这两个模块旨在融合局部和非局部信息，以缓解这一问题。这两个模块被用来同时扩大局部卷积的接受场，并保留密集的高分辨率特征图。未来的工作可能会致力于将正射影像和数字高程模型结合使用，以实现更精确和更稳健的滑坡测绘。此外，滑坡易发地区的灾害前易感性测绘[26]对于减轻灾害也至关重要。与本文对应的代码已公开发布¹。

参考文献

[1]李建军, 李建军, 李建军, “基于图像分析的滑坡检测方法研究”, 《地球科学与遥感学报》, vol. 49, no. 1. 12, pp. 4928 - 4943, 2011.

[2]窦俊杰. 张, 陈士生, A. Yunus, j - k. 刘海燕, “基于面向对象图像分析和遗传算法的滑坡自动检测方法”, 《遥感》, 第 7 卷, 第 1 期. 第 4 期, pp. 4318-4 342, 2015.

[3]李建军, “基于随机森林和面向对象分析的高分辨率滑坡制图”, 《环境科学》, 第 3 卷, pp 123-129, 2011.

[4] J.-Y. 劳.j. Jhan 和 r - j. Rau, “基于多传感器光学图像和 dem 的半自动面向对象滑坡识别方案”, 《IEEE 地球科学与遥感学报》, 第 52 卷, 第 2 期. 2, pp. 1336-1349, 2013.

[5]王晓明, “基于 gis 的滑坡特征提取方法研究”, 《遥感技术》第 5 卷, 第 1 期. 5, pp. 2389-2410, 2013.

[6]王志强, “基于地物的遥感图像分析”, 《遥感学报》, 第 6 卷, 第 6 期. 1, pp. 2-16, 2010.

[7]王 晓明, 王 晓明, 王 晓明, “基于高分辨率图像的滑坡图像检测 方法研究”, 《地球物理学报》, 第 4 卷, 第 1-11 页, 2018.

¹ <https://github.com/saedrna/DA-U-Net>

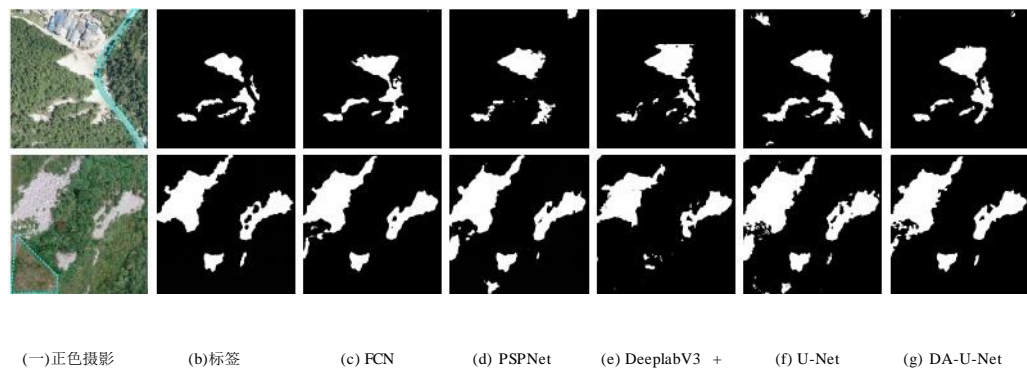


图 4:不同方法在典型混淆区域上的表现。突出显示的是常见的混淆区域。

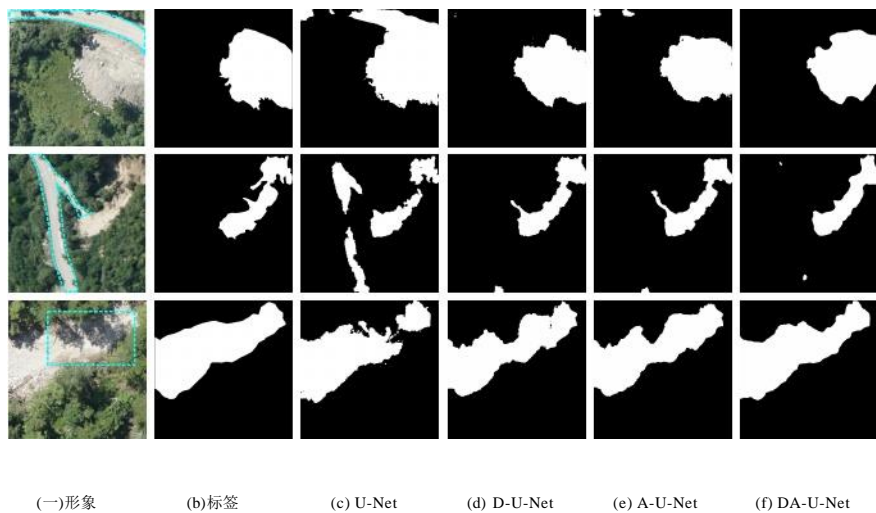


图 5:DA-U-Net 结构的消融研究。突出显示的是容易出错的区域。

[8]陈建军, 陈建军, 陈建军, 陈建军, 陈建军, 陈建军 等, “基于雷达和光学遥感技术的滑坡遥感监测研究——以长江三峡库区为例”, 《遥感技术与应用》, 2016, 第 4 卷, 第 1 - 8 页。

[9]王晓明, 王晓明, “基于卷积神经网络的语义分割”, 《IEEE 计算机视觉与模式识别会议论文集》, 2015, pp. 391 - 391。

[10]李志强. 陈, G. Papandreou, I. Kokkinos, K. Murphy 和 A. L. Yuille, “深度卷积:基于深度卷积神经网络的语义图像分割”, 《IEEE 模式分析与机器学习学报》, vol. 40, no. 1. 第 4 期, 第 834-848 页, 2017。

[11]李志强. Chen, Y. Zhu, G. Papandreou, F. Schroff 和 H. Adam, “用于语义图像分割的可分离卷积编码器-解码器”, 《欧洲计算机视觉会议论文集 (ECCV)》, 2018, 第 801-818 页。

[12]王晓明, 王晓明, “基于卷积神经网络的生物医学图像分割”, 《中国医学信息学报》, 2011. Springer, 2015, pp. 234-241。

[13]雷涛, 张震, 吕振杰, 李世生, 刘强, “基于深度卷积神经网络的滑坡清单映射”, 《IEEE 地球科学与遥感快报》, 第 16 卷, 第 1 期. 6, pp. 982-986, 2019。

[14]李志强. 陈, G. Papandreou, F. Schroff, H. Adam, “对语义图像分割的再思考”, *arXiv 预印本 arXiv:1706.05587*, 2017。

[15]李志强. 陈毅, 杨毅, 王勇, 徐伟, A. L. Yuille, “关注尺度:尺度感知语义图像分割”, 《IEEE 计算机视觉与模式识别会议论文集》, 2016, pp. 3640-3649。

[16]李晓明, 李晓明, 李晓明, 李晓明, 李晓明, 李晓明, 李晓明 等, “神经网络对胰腺的影响:基于神经网络的研究”, *arXiv 预印本 arXiv*, 2018。

[17]李晓明, 李晓明, 李晓明, 李晓明. “注意力的学习”, *arXiv, 预印本 arXiv: 1844 - 1844*, 2018。

[18]徐凯, 陈建军, 陈建军, 陈建军, 陈建军, 陈建军, 陈建军, 陈建军, “基于视觉注意力的神经图像标题生成:基于视觉注意力的神经图像标题生成”, 《中国计算机工程学报》, 2015, pp. 2048-2057。

[19]张震, 刘强, 王勇, “基于深度残差网络的道路提取”, 《IEEE 地球科学与遥感快报》, 第 15 卷, 第 15 期. 5, pp. 749-753, 2018。

[20]张晓明, 张晓明, 张晓明, 等. Kaiser, and I. Polosukhin, “Attention is all you need”, 《神经信息处理系统进展》, 2017, pp. 5998-6008。

[21]王晓, 王晓明, “非局部神经网络”, 《计算机视觉与模式识别会议》, 2018, pp. 794 - 7803。

[22]傅建军, 刘军, 田红红, 李艳, 包阳, 方志强, 卢辉, “基于双注意力网络的场景分割”, 《计算机视觉与模式识别会议论文集》, 2019, pp. 3146-3154。

[23]何开, 张晓霞, 任淑娟, 孙军, “基于深度残差学习的图像识别方法”, 《IEEE 计算机视觉与模式识别》, 2016, pp. 770-778。

[24]张晓明, 张晓明, 张晓明. 基于深度卷积神经网络的图像识别 [j]. 中文信息学报, 中文信息学报, 第 4 卷第 1 期, 2014。

[25]赵辉, 石静, 齐晓霞, 王晓, 贾军, “金字塔场景分析网络”, 《IEEE 视觉与模式识别会议论文集》, 2017, pp. 2881-2890。

[26]刘建军, 刘建军, “基于模糊神经网络的滑坡易感度反演方法研究”, 《中国地质大学学报(自然科学版)》, 第 4 卷, 第 1 期. 12, pp. 4164-4177, 2010。