

中图法分类号: TP751 文献标识码: A 文章编号: 1006-8961(2023)02-0628-15

论文引用格式: Zhou F, Hang R L, Xu C, Liu Q S and Yang G W. 2023. Airborne image segmentation via progressive multi-scale causal intervention. Journal of Image and Graphics, 28(02): 0628-0642 (周峰, 杭仁龙, 徐超, 刘青山, 杨国为. 2023. 渐进式多尺度因果干预航拍图像分割. 中国图象图形学报, 28(02): 0628-0642) [DOI:10.11834/jig.211036]

渐进式多尺度因果干预航拍图像分割

周峰¹, 杭仁龙², 徐超¹, 刘青山^{2*}, 杨国为^{1,3}

1. 南京审计大学计算机学院, 南京 211815; 2. 南京信息工程大学计算机与软件学院, 南京 210044;
3. 青岛大学电子信息学院, 青岛 266071

摘要: 目的 航拍图像分割为遥感领域中许多实际应用提供支撑。与传统方法相比, 深度学习方法能够自适应地学习与任务相关的特征, 极大提升了分割精度, 但忽略了数据集中的偏置问题。由偏置引起的混杂因子干扰使分割方法容易获得模糊的物体边缘, 并且难以区分易混淆物体。针对这个问题, 提出了一种基于渐进式多尺度因果干预的模型。方法 首先, 使用深度卷积神经网络提取航拍图像的卷积特征。然后, 解混杂模块引入类别隐特征, 近似表示混杂因子特征。同时, 使用混杂因子特征以因果干预的方式将卷积特征分解成对应每一种混杂因子下的特征表示, 抑制特定混杂因子的干扰。最后, 由深层解混杂特征得到的分割结果, 经过融合模块指导浅层解混杂特征生成分割结果, 以此得到每个尺度的分割结果, 并以加权求和的方式得到最终分割结果。结果 实验在公开的航拍图像数据集 Potsdam 和 Vaihingen 上进行, 与 6 种先进的深度学习方法分割方法和 7 种公开的基准方法进行对比。本文方法在 Potsdam 和 Vaihingen 数据集中的总体准确率分别为 90.3% 和 90.8%, 相比性能第 2 的深度学习方法分别提高了 0.6% 和 0.8%。与性能第 2 的基准方法相比, 本文方法在 Potsdam 和 Vaihingen 数据集上的总体准确率分别提升了 1.3% 和 0.5%。结论 本文提出的分割模型能够有效缓解数据集中的偏置问题, 提升了航拍图像分割性能。

关键词: 航拍图像; 语义分割; 卷积神经网络 (CNN); 因果干预; 解混杂

Airborne image segmentation via progressive multi-scale causal intervention

Zhou Feng¹, Hang Renlong², Xu Chao¹, Liu Qingshan^{2*}, Yang Guowei^{1,3}

1. School of Computer Science, Nanjing Audit University, Nanjing 211815, China;
2. School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044, China;
3. School of Electronic Information, Qingdao University, Qingdao 266071, China

Abstract: Objective Airborne-relevant image segmentation is one of the essential tasks for remote sensing, which can assign a semantic label to each pixel in an image. Its applications have been developing in related to such research domain like land use, urban planning, and environmental surveillance. To analyze the segmentation results of airborne image, most of conventional methods are concerned about label-manual features like scale-invariant feature transform (SIFT) and histo-

收稿日期: 2021-11-05; 修回日期: 2022-02-22; 预印本日期: 2022-02-28

* 通信作者: 刘青山 qslu@nuist.edu.cn

基金项目: 国家杰出青年科学基金项目 (61825601); 国家自然科学基金项目 (71972102, 62172229, 61906096); 江苏省自然科学基金项目 (BK20211295)

Supported by: National Science Fund for Distinguished Young Scholars (61825601); National Natural Science Foundation of China (71972102, 62172229, 61906096); Natural Science Foundation of Jiangsu Province, China (BK20211295)

gram of oriented gradient (HOG) . Their performance is constrained of features-selected intensively. It is still challenged to deal with such complex scene. To optimize image classification tasks, the deep convolution neural network (DCNN) has been melted into pixel-wise classification issues like airborne image segmentation. The ability of DCNN is linked to auto task-adaptive features extraction for training to a certain extent. Fully convolutional network (FCN) can be used to improve the performance of airborne image segmentation. FCN-based UNet and SegNet are followed and developed further. A newly encoder-decoder design is involved in for airborne image segmentation. The fixed-size convolutional kernels are employed to capture contextual information for segmentation. Deep learning technique is beneficial for airborne image segmentation, but the output-learned is restricted of single-scale and local. In fact, it is required to handle the two challenging issues in airborne image segmentation as mentioned below: 1) remote sensing based multiple objects, and 2) multi-source images based heterogeneity. The first task is focused on multi-scale contexts for segmentation. The second one is developed to get discriminative information more in terms of global extraction. To alleviate the limitations and improve the performance, these two kinds of methods are compared to FCN-based methods. However, the mutual benefits are not included and the interference of confounders is leaked out. So, we develop a causal and effect-intervened segmentation method to suppress the interference of confounders. **Method** In this study, a progressive multi-scale cause and effect intervention model (PM-SCIM) is built up. First, the PM-SCIM takes ResNet18 as backbone network to extract convolutional features of airborne images. Then, a de-confounded module is designed to measure the average cause and effect of confounders on the convolutional feature through stratifying the confounders into different cases. In this way, to suppress the interference of a specific confounder, it is possible to collect objects in any context confounders indirectly. Next, the de-confounded feature generated is used to analyze the segmentation result from the deepest layer. This overall scale segmentation result can be obtained while a fusion module is fed into the segmentation results are guided in terms of de-confounded features from shallow layers. Finally, all segmentation results are fused via sum-weighted. The PM-SCIM is trained on two datasets of those are Potsdam and Vaihingen. For Potsdam, we choose 24 images for training and the remaining 14 images for testing. For Vaihingen, we select 16 images for training and the remaining 17 images for testing. To make full use of computing resources, a 256×256 sliding window is used to crop the input images for generating training samples. At inference phase, the same sliding method is used to crop input tiles from the original testing image and they are processed gradually. For training, the momentum parameter is set to 0.9, the learning rate is kept to 0.01, and the weight decay is configured at 0.000 01. The SGD (stochastic gradient descent) learning procedure is accelerated using a NVIDIA GTX TITAN X GPU device. A poly learning rate pathway is employed to update each iteration-after learning rate as well. **Result** Our demonstration is compared to 4 popular state-of-the-art deep methods and 7 public benchmark data sets. The quantitative evaluation metrics are composed of overall accuracy (OA) and F1 score, and we offer several segmentation maps of benched results for comparison. Specifically, the OA is increased by 0.6% and 0.8% each (higher is better), and mean F1 increased by 0.7% and 1% of each as well (higher is better) compared to DANet on Potsdam and Vaihingen. The OA is increased by 1.3%, and the mean F1 is increased by 0.3% in comparison with CVEO2 on Potsdam. The OA is increased by 0.5% and the mean F1 is increased by 0.5% in terms of the comparative analysis with DLR_10 on Vaihingen. The segmentation maps showed that our method has its potentials for small objects (e.g., car) and ambiguous objects (e.g., tree and lawn). Additionally, to clarify the effectiveness of multiple modules in PM-SCIM, a series of ablation studies on Potsdam and Vaihingen are carried out. **Conclusion** To suppress the interference of confounders using cause and effect intervention, a novel segmentation method is proposed and developed through melting de-confounded module and fusion module into ResNet18.

Key words: airborne image; semantic segmentation; convolutional neural network; causal intervention; de-confound

0 引言

随着传感器技术的发展,航拍图像的空间分辨率可以达到厘米级(Zhou等,2021)。得益于超高分

辨率的空间信息,航拍图像成功应用于城市规划(余帅和汪西莉,2021)、土地利用、环境监测、水体检测(李鑫伟等,2021)和军事等领域。航拍图像分割任务作为实现上述应用的重要步骤,一直在遥感领域受到广泛关注。航拍图像分割旨在为航拍图像

中每个像素点分配对应的语义标签。然而在复杂的遥感场景中,物体外观呈现的多样性为航拍图像分割带来挑战。

近些年,深度学习尤其是深度卷积神经网络(deep convolutional neural network, DCNN)在遥感图像领域取得了突破性进展(Ding 等,2021;Long 等,2017;Mou 等,2017;Zheng 等,2021)。DCNN 通过多层非线性映射,能够自动地从输入图像中学习具有判别性的特征。Long 等人(2015)将 DCNN 中的全连接层替换成卷积层,提出一个端到端的全卷积网络(fully convolutional network, FCN),极大提升了航拍图像的分割精度(Sun 和 Wang,2018;Volpi 和 Tui-ia,2017)。FCN 的分割结果是通过卷积特征进行 32 倍率的空域上采样得到的,但高倍率的上采样极易丢失空间细节信息,因此仅基于 FCN 无法取得更好的分割结果。Ronneberger 等人(2015)提出对称的编码—解码网络结构 U-Net,将输入图像逐尺度地下采样编码,然后逐尺度地上采样解码得到最终的分割结果。同时,U-Net 使用跨层连接将编码器中每个尺度的特征与解码器中对应尺度的特征融合,达到捕获空间位置信息的目的。Badrinarayanan 等人(2017)提出的编码—解码网络 SegNet 将编码器中池化索引值用于解码器中的上采样,以保留边缘信息。这 3 个经典的分割模型通过固定大小的卷积核学习上下文信息,成功地将深度学习应用到分割领域,但受限于卷积核感受野的局限性,仅能捕获单一尺度且为局部的上下文信息,对航拍图像分割任务中的两个问题无法应对。第 1 个问题是遥感场景中物体(例如车)尺度多变。第 2 个问题是不同类物体(例如道路和楼顶)之间呈现较小的差异性。

研究表明,提取多尺度上下文能够很好地解决第 1 个问题。基于此,金字塔场景解析网络(pyramid scene parsing network, PSPNet)(Zhao 等,2017)和 DeepLabv3+(Chen 等,2018b)相继提出。PSPNet 中的金字塔池化模块(pyramid pooling module, PPM)使用多个尺度(1×1 、 2×2 、 3×3 、 6×6)的池化核提取多尺度上下文,并进行融合。DeepLabv3+ 引入由多个空洞卷积构成的空洞空间金字塔池化(atrous spatial pyramid pooling, ASPP)模块,通过不同空洞率的空洞卷积大幅增加感受野范围,有效获取不同尺度下的上下文。与 DeepLabv3+ 类似,韩彬彬等人(2020)使用 ASPP 提取不同尺度目

标及其边缘。除此之外,Audebert 等人(2016)将不同尺度下的航拍图像输入到编码—解码结构的 DCNN 中,得到对应的分割结果。为了融合多尺度上下文信息,将不同尺度的分割结果相加得到最终结果。Liu 等人(2018)提出基于 U-Net 的自级联网络结构,使用不同空洞率的空洞卷积提取多尺度上下文,并以自级联方式进行融合。

针对第 2 个问题,局部上下文信息不足以帮助区分易混淆物体,如果能够引入更多场景信息(即全局上下文信息),就可以抑制与场景不相容的物体。Zhang 等人(2018)提出一个上下文编码模块,旨在捕获场景的语义上下文信息,用于增强相关类别物体的特征,抑制其他物体特征。随着自注意力方法的发展,研究者尝试用其提取全局上下文信息。Fu 等人(2019)提出双注意力网络(dual attention network, DANet),通过加权对应域内所有像素点特征,构成目标像素点的上下文表示。其中,每个像素点对应的权重是与目标像素点的特征相似度。与 DANet 类似,Mou 等人(2019)提出空间和通道自注意力模块,用于捕获对应维度的全局上下文信息。然而,自注意力需要考虑所有像素点对目标像素点的影响,非常消耗计算资源。Huang 等人(2020)提出十字交叉网络(criss-cross network, CCNet),仅使用目标像素点的垂直和水平方向内像素点计算上下文,大幅降低了计算资源的消耗。

自注意力方法能够自适应地集成局部特征与全局依赖性,而基于多尺度上下文信息的方法忽略了上下文信息中特征之间的依赖关系。因此,基于自注意力机制的全局上下文信息能够帮助模型获得更好的分割性能。但上述方法都忽视了数据集中固有的偏置问题。首先,遥感场景中任何物体都不是单一存在的,有极大概率与特定物体同时出现。例如,大多数航拍图像中,树和草地经常相邻出现,分割模型会误以为树和草地必须作为整体出现,导致树和草地之间的边界模糊,如图 1 第 1 行所示。其次,由于遥感场景中车辆外观的多样性,容易出现车身部分外观(颜色与形状)与周围建筑物楼顶相似的情况,如图 1 第 2 行所示。这种情况下,分割模型只能将受环境影响较小,即车辆最具有判别性的部位(车头)准确识别。上述现象中,树和车的先验分布分别引入草地和建筑物两种混杂因子,一般的分割模型只捕获到它们之间的伪相关性,忽略了因果关

系。具体地说,对树而言,草地与它并无因果关系,但由于它们在数据集中的分布呈现正相关(大概率同时出现),导致分割模型将二者建立了联系。而对车而言,具有因果关系的车头和车身受建筑物干扰被分割模型分离。为了解决数据集中的偏置问题,本文提出使用因果干预的方式抑制由混杂因子带来的伪相关性,提升目标物体的分割精度。同时,使用深层特征得到的分割结果指导浅层特征,以渐进式分割的方式更有效地利用多尺度上下文信息。基于这两个思想,本文构建了渐进式多尺度因果干预模型。

本文主要贡献如下:1)提出渐进式多尺度因果干预模型(progressive multi-scale causal intervention model, PM-SCIM),通过对特征进行逐尺度的因果干

预,解决数据集的偏置问题,提高航拍图像分割性能。2)为了能够抑制混杂因子对目标物体的干扰,提出一个解混杂模块(de-confounded module, DCM),通过因果干预的方式建模目标物体在所有混杂因子下的分布情况,消除特定混杂因子的干扰。3)提出一个融合模块(fusion module, FM),能够使用深层特征得到的分割结果,指导浅层分割结果的预测,从而更好地利用多尺度上下文信息。4)在ISPRS(International Society for Photogrammetry and Remote Sensing)提供的Potsdam和Vaihingen数据集上与相关的深度学习方法 and 公开的基准方法进行对比。实验结果表明,PM-SCIM能够获得具有竞争力的分割性能。

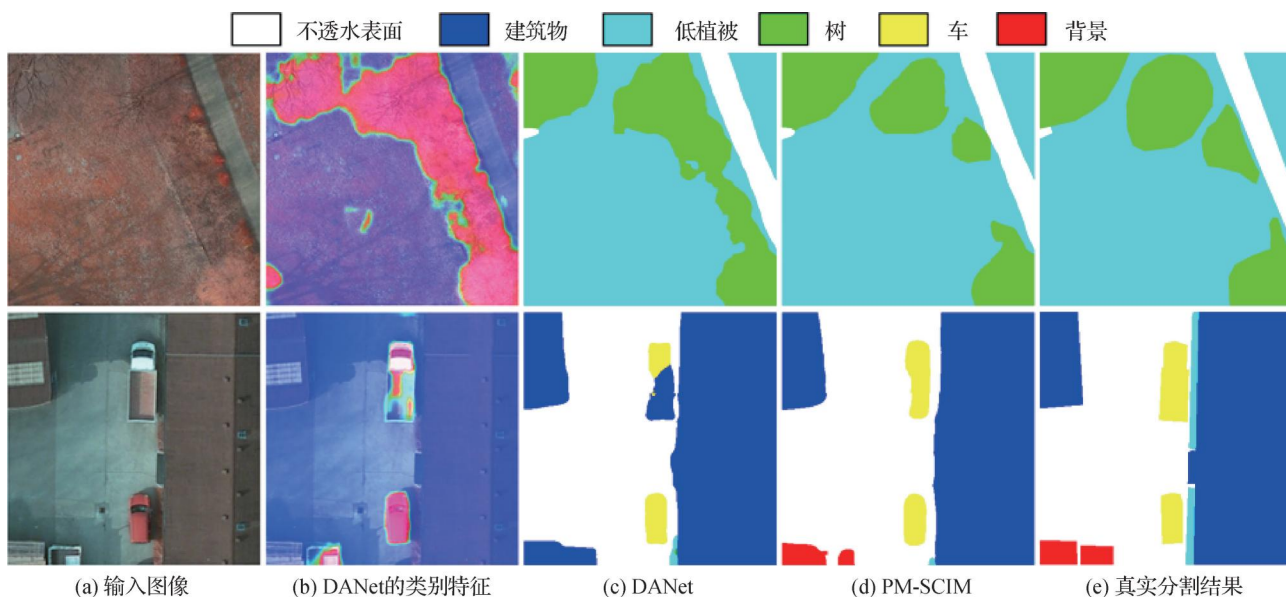


图1 分割结果对比

Fig. 1 Comparison of segmentation results

((a) input images; (b) class-wise feature of DANet; (c) DANet; (d) PM-SCIM; (e) ground truth)

1 渐进式多尺度因果干预模型

渐进式多尺度因果干预模型的网络结构如图2所示。该网络主要由骨干网络、解混杂模块和融合模块构成。首先,使用ResNet18(He等,2016)作为骨干网络,提取输入航拍图像的卷积特征。然后,将每个尺度的卷积特征都输入解混杂模块,通过因果干预方式将卷积特征分解,以消除某些混杂因子对目标物体的干扰。为了能够有效利用所有尺度下的信息,将深层解混杂特征通过 1×1 卷积核得到分

割结果,再经过融合模块,引导浅层解混杂特征得到分割结果。然后,逐尺度地进行该过程,直至得到所有尺度的分割结果。最后,使用加权求和的方式自适应地融合所有尺度的结果,得到最终分割结果。

1.1 解混杂模块

航拍图像中的物体通常出现在特定场景下,这些场景对目标物体识别的影响具有两面性。例如树木总是和草地相邻出现,造成树木和草地的正相关性,分割模型错误地将二者建立联系。但是,Zhou等人(2021)提出的类别引导特征解耦网络(class-guided feature decoupling network, CGFDN)利用车辆

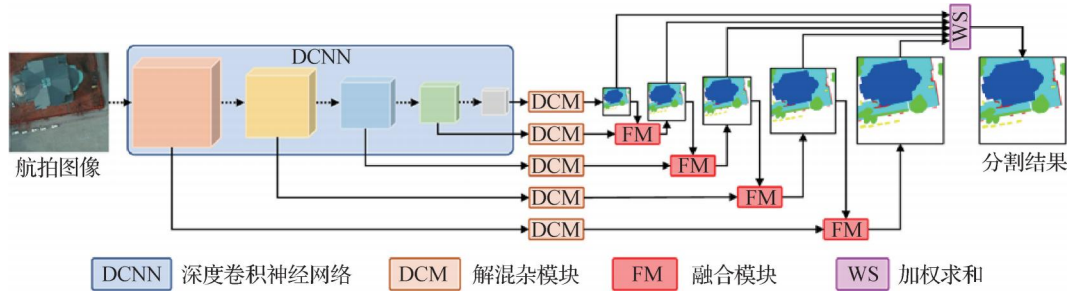


图2 PM-SCIM 网络结构

Fig. 2 The network structure of PM-SCIM

和道路的共生关系,帮助区分与车辆具有高度相似性的建筑物。先前的深度学习方法并没有考虑这个问题,都是将航拍图像的卷积特征 X 输入模块 $P(Y|X)$,得到包含丰富上下文信息的特征 Y ,直接用于预测分割结果。整个过程忽略了场景中混杂因子对目标物体的干扰。为了解决这个问题,本文提出解混杂模块,使用 $P(Y|do(X))$ 替换 $P(Y|X)$ 。 do 操作旨在排除混杂因子带来的干扰,找到 X 与 Y 之间的因果关系。

首先,构建如图3所示的结构因果模型,定义 X 和 Y 之间的因果关系,并找出混杂因子干扰的原因。图中的箭头代表由起因指向结果的因果关系。假设图中的 B 代表的是常识(数据集中大多数图像体现的规律),如树木和草地一般相邻出现。 $B \rightarrow X$ 表示在该常识下,模型提取的关于树木的图像特征; $X \rightarrow M$ 表示 X 受到偏置 B 影响之后变换为特征 M 。即使树木与草地本身没有关系, B 也会将它们关联起来,形成特征 M 。 $M \rightarrow Y$ 表示将掺杂草地信息的树木特征用于分割,导致分割结果中树木和草地的边缘难以区分。常规的分割模型只关注了 $X \rightarrow Y$,而忽视了后门通道 $X \leftarrow B \rightarrow M \rightarrow Y$ 对最终结果 Y 的影响。通过对图3分析,发现 B 通过后门通道是可以影响树木和草地(混杂因子)的关系,导致最终学习到的特征中草地和树木难以区分。本文的目的就是通过因果干预来消除混杂因子的影响。

基于因果干预的思想,本文提出解混杂模块,使用 $P(Y|do(X))$ 建模 X 和 Y 之间的因果关系。得到 $P(Y|do(X))$ 最理想的方法应该是物理干预,即收集目标物体在所有场景下的图像,以此捕获它与所有混杂因子的关系。但是,实现物理干预是不现实的,所以解混杂模块通过后门调整(Wang等, 2020)的方式虚拟地得到干预结果。后门调整的主

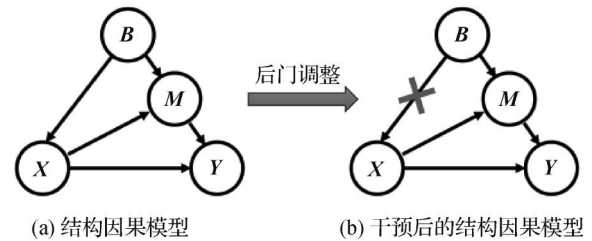


图3 因果关系示意图

Fig. 3 The causality diagram ((a) structural causal model; (b) intervened structural causal model)

要过程包括两步,首先切断 X 和 B 之间的联系,然后将 B 拆分成集合 $\{b\}$,每个元素代表一种混杂因子。基于此, B 和 X 不再有直接联系,可以间接衡量 X 在不同混杂因子 $\{b\}$ 干扰下,对 Y 的平均因果影响。具体方式为

$$P(Y|do(X)) = \sum_b P(Y|X, B=b)P(B=b) \quad (1)$$

为了使用深度网络实现式(1),本文使用归一化加权几何平均(normalized weighted geometric mean, NWGM)(Yang等, 2021)的方法将这个过程近似为

$$P(Y|do(X)) \approx \sum_b P(f(b)|g(X))l(b) \quad (2)$$

式中, $f(\cdot)$, $g(\cdot)$, $l(\cdot)$ 是参数化网络。如图4所示,先将卷积特征 $X \in \mathbf{R}^{H \times W \times C}$ 和混杂因子特征 $B \in \mathbf{R}^{K \times C}$ 分别输入到两个网络中,得到对应的新表征 $\hat{X} \in \mathbf{R}^{H \times W \times C/2}$ 和 $\hat{B} \in \mathbf{R}^{K \times C/2}$,其中 H 和 W 分别表示特征图的宽和高, C 表示特征的通道个数, K 代表数据集中类别的个数。然后将 $\hat{X}$ 重塑,通过矩阵乘法计算出 $A = \text{softmax}(\hat{B}\hat{X}^T)$ 。最后,与经过网络变化的 $\hat{B}$ 相乘得到最终的结果 $A^T\hat{B}$ 。其中, A 可以近似地衡量 $P(f(b)|g(X))$ 。在实现过程中,混杂因子特

征通过使用航拍图像数据集对 ResNet18 进行预训练得到。具体地说,解混杂模块引入类别隐特征 $B \in \mathbf{R}^{K \times C}$,与 ResNet18 中最后一层卷积层的输出特征相乘,得到维度为 $H \times W \times K$ 的分割结果,与真实分割结果进行损失函数计算并优化。同时,假设对目标物体会产生干扰的混杂因子来自于已标注的物

体类别中,所以解混杂模块将每一类物体隐特征视做对应的混杂因子特征。

解混杂模块使用后门调整,抑制了卷积特征 X 中混杂因子的影响,得到解混杂特征 Y 。接下来,需要借助融合模块将解混杂特征进行渐进式的多尺度分割预测。

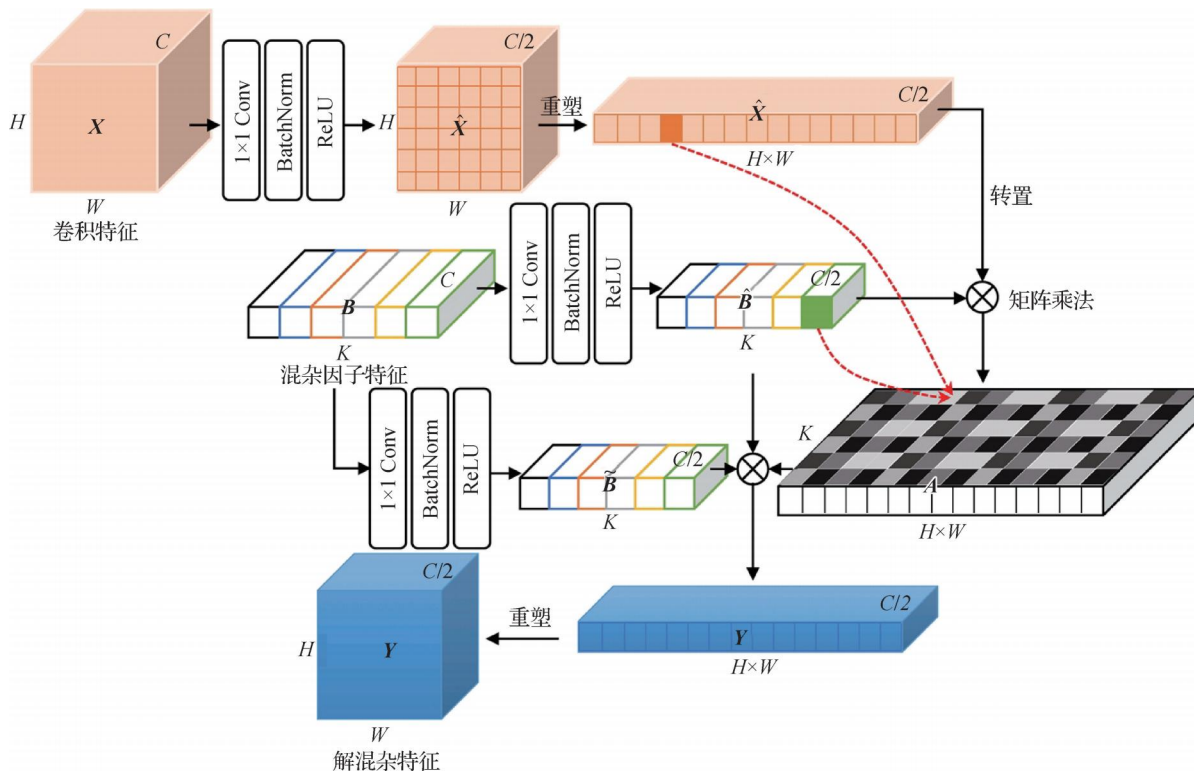


图4 解混杂模块

Fig. 4 De-confounded module

1.2 融合模块

航拍图像的高空间分辨率带来的另一个挑战是场景内物体尺度不一的问题。尤其是车辆这种小尺度目标会增加分割难度。为解决这个问题,提出了融合模块,进行渐进式的多尺度分割结果预测。深层的特征包含更加丰富精准的语义信息,即场景中出现物体的类别信息,而浅层的特征则包含更多的空间位置信息。利用深层的语义信息确定场景中包含的物体类别,通过渐进式地指导浅层特征,确定不同类别物体出现的空间位置。

为了能够在多个尺度下进行预测,首先将 ResNet18 每一个尺度的卷积特征输入解混杂模块得到对应的解混杂特征。然后,将最后一个尺度的解混杂特征经过一层 1×1 卷积,得到分割结果。同时,将其输入如图5所示的融合模块中,进行2倍的上

采样操作,与上一尺度的解混杂特征进行通道维度的串联。最后,将串联特征输入到一个4层的网络中,得到上一尺度下的分割结果。通过循环上述过程,可以依次得到所有尺度下的分割结果。

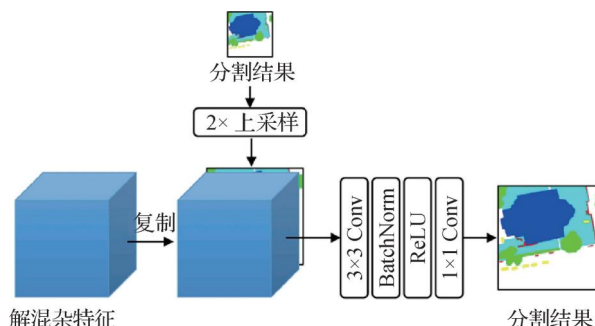


图5 融合模块

Fig. 5 Fusion module

1.3 损失函数

在渐进式多尺度分割结果预测过程中,每个尺度都需要一个相对精确的分割结果作为引导。因此使用真实分割结果对每个尺度的预测结果进行监督训练。同时,为了能够进一步利用不同尺度下的信息,PM-SCIM 将每个尺度的分割结果进行加权求和,得到最终的分割结果。假设 $\hat{Y}_i$ 代表第 i 个尺度预测的分割结果, T_i 表示第 i 个尺度的真实分割结果。最终的分割结果 $\hat{Y}$ 为

$$\hat{Y} = \sum_i \alpha_i \hat{Y}_i \quad (3)$$

式中,可学习参数 α_i 为第 i 个尺度预测分割结果的权重。最终,PM-SCIM 的损失函数定义为

$$L = - (T\hat{Y} + \sum_i T_i \log \hat{Y}_i) \quad (4)$$

式中, T 代表与最终预测结果 $\hat{Y}$ 尺度对应的真实分割结果。

2 实验结果分析

2.1 数据集及指标

2.1.1 数据集

为了验证 PM-SCIM 的有效性,与相关深度学习方法及公开基准方法在 ISPRS Potsdam 和 Vaihingen 数据集上进行对比实验。

Potsdam 数据集包含 38 幅空间分辨率为 5 cm 的图像。所有图像的空间尺寸都为 6 000 × 6 000 像素。每幅图像由 4 个通道组成,分别是近红外、红、绿和蓝。数据集还提供了对应的数字地面模型 (digital surface model, DSM) 结果,包含地表物体的高度信息。这些图像中所有像素点都标注了对应的语义标签,包含不透水表面 (impervious surface)、建筑物 (building)、低植被 (low vegetation)、树 (tree)、车 (car) 和背景 (background) 等 6 类物体。根据官网给出的实验设置,实验时,使用 24 幅图像用于训练,剩下的 14 幅图像用于测试。PM-SCIM 仅使用由近红外、红和绿 3 个通道构成的图像作为输入,没有使用任何额外数据。

Vaihingen 数据集包含 33 幅空间分辨率为 9 cm 的图像。图像的平均空间尺寸为 2 494 × 2 064 像素。每幅图像由 3 个通道构成,分别是近红外、红和绿。与 Potsdam 数据集类似,该数据集也包含不透水表面、建筑物、低植被、树、车和背景等 6 种地物。

按照官网划分数据集的设置,将 16 幅图像用于训练,剩下的 17 幅用于测试。针对该数据集,PM-SCIM 仅使用上述三通道的图像作为输入,没有使用 DSM 等额外数据。

图 6 是两种数据集的图像示例。

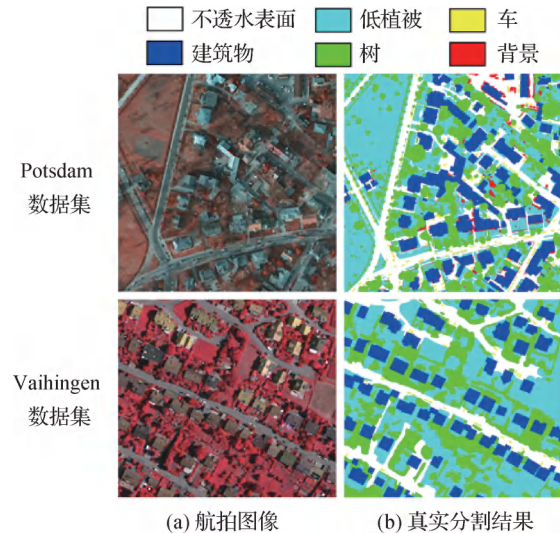


图 6 Potsdam 和 Vaihingen 数据集示例

Fig. 6 Examples of Potsdam and Vaihingen datasets

((a) airborne images; (b) ground truth)

2.1.2 评价指标

本文使用总体准确率 (overall accuracy, OA) 和 $F1$ 分数评估分割结果的精度。OA 衡量正确分类的像素点个数占总像素点个数的比例。 $F1$ 分数定义为精确率 (precision, P) 和召回率 (recall, R) 的调和平均数。具体为

$$\begin{cases} F1 = 2 \frac{PR}{P + R} \\ P = \frac{TP}{TP + FP} \\ R = \frac{TP}{TP + FN} \end{cases} \quad (5)$$

式中, TP 代表模型分类正确的正样本, FP 代表模型分类错误的负样本, FN 代表模型分类错误的正样本。

2.2 实验细节

由于目前的显卡无法直接处理高分辨率的航拍图像,使用一个 256 × 256 像素的滑动窗口以步长为 64 的跨度从原始航拍图像中裁剪训练图像块。在测试阶段,同样使用滑动窗口的方式从原始图像裁剪出测试图像依次输入到模型中,然后按照输入

的顺序拼接成原图大小。在拼接过程中,重叠的部分使用平均值作为预测结果。

本文使用 PyTorch 在一台 64 位的计算机上实现 PM-SCIM。该计算机配备 i7-4790 CPU、32 GB 运行内存和 1 块 GTX TITAN X 显卡。同时,使用随机梯度下降 (stochastic gradient descent, SGD) 优化 PM-SCIM。基础学习率设置为 0.01,动量设置为 0.9,权重衰减设置为 0.000 01,并采用 poly 学习率衰减策略 (Liu 等,2015)。每次迭代后,初始的学习率乘以 $(1 - i/t)^{0.9}$ 进行衰减。其中, i 为当前迭代次数, t 为总迭代次数。训练周期设置为 200。

2.3 消融实验

2.3.1 定量分析

PM-SCIM 包含解混杂模块 (DCM) 和融合模块 (FM)。为了验证这两个模块的有效性,首先测试基础网络 ResNet18 在 ISPRS Potsdam 和 Vaihingen 数据集上的分割性能。然后,在 ResNet18 上分别单独附加两个模块测试其性能,如表 1 所示。可以看出,加入解混杂模块后,PM-SCIM 在 Potsdam 和 Vaihingen 数据集上的 OA 分别提升了 2% 和 4.8%。同时,使用融合模块进行渐进式的多尺度因果干预,ResNet18 在 Potsdam 和 Vaihingen 数据集上的 OA 分别提升了 1.6% 和 4.7%。将两个模块同时连接到 ResNet18,PM-SCIM 在 Potsdam 和 Vaihingen 数据集上的 OA 分别提升到 89.7% 和 90.3%。最后,测试了加权求和是否会对最终的分割结果有影响。通过加权求和操作能够自适应地融合多尺度分割结果,模型的分割性能提升了 0.6% 和 0.5%。

表 1 不同设置下 PM-SCIM 的性能对比

Table 1 Performance comparison of different setups of PM-SCIM

骨干网络	DCM	FM	WS	OA/%	
				Potsdam	Vaihingen
ResNet18	—	—	—	87.2	84.6
ResNet18	✓	—	—	89.2	89.4
ResNet18	—	✓	—	88.8	89.3
ResNet18	✓	✓	—	89.7	90.3
ResNet18	✓	✓	✓	90.3	90.8

注:加粗字体表示各列最优结果,“✓”表示采用,“—”表示未采用。

2.3.2 定性分析

首先,为了更好地理解解混杂模块如何帮助模型提升分割性能,使用网络的中间特征进行可视化分析。图 7 展示了树和车两组个例。baseline 表示仅使用 ResNet18 作为分割模型,PM-SCIM 表示在 ResNet18 基础上附加解混杂模块。图中的类别特征是通过将 ResNet18 输出的卷积特征与用于预测分割结果的 1×1 卷积参数相乘得到。从图 7 第 1 行可以看出,虽然树木和草地外观差异明显,但 baseline 仍然难以区分它们,导致 baseline 得到的树木特征边缘模糊。造成这种现象的原因是数据集中存在偏置问题,即树和草地同时相邻出现在大量样本中。通过因果干预,解混杂模块成功抑制了草地和树的伪相关性,使 PM-SCIM 能够获得清晰的边缘。从图 7 第 2 行可以看出,由于车身与周围建筑物非常相似,导致 baseline 只能识别出受环境影响较小的车头和车尾。而 PM-SCIM 通过因果干预保留了车头、车身和车尾之间的因果关系,抑制了建筑物的干扰信息,将车辆完整地识别并定位。融合模块的作用如图 8 所示,该模块能够有效识别并定位场景中的小目标,例如车辆。

2.4 性能对比实验

2.4.1 与相关深度学习方法对比

表 2 和表 3 分别为 PM-SCIM 与目前流行的基于多尺度上下文信息和全局上下文信息的深度学习方法的对比结果。具体地说,选取 PSPNet (Zhao 等,2017)、DeepLabv3+ (Chen 等,2018b)、S-RA-FCN (serial relation augmented fully convolutional network) (Mou 等,2019)、DANet (Fu 等,2019)、CCNet (Huang 等,2019) 和 CGFDN (Zhou 等,2021) 作为对比方法。为了实验的公平性,对比方法都采用 ResNet18 作为骨干网络。可以看出,DANet 在 Potsdam 和 Vaihingen 数据集上的 OA 和 mean F1 分数优于 DeepLabv3+ 和 PSPNet。这是因为 DANet 考虑了上下文内部特征之间的依赖关系。CCNet 由于简化了 DANet 中的自注意力操作,丢失了空间上部分像素点的信息,导致最终的分割结果略低于 DANet。而 PM-SCIM 通过因果干预缓解了数据集中的偏置问题,与 DANet 相比,在 Potsdam 和 Vaihingen 数据集上的 OA 分别提高了 0.6% 和 0.8%、mean F1 分数

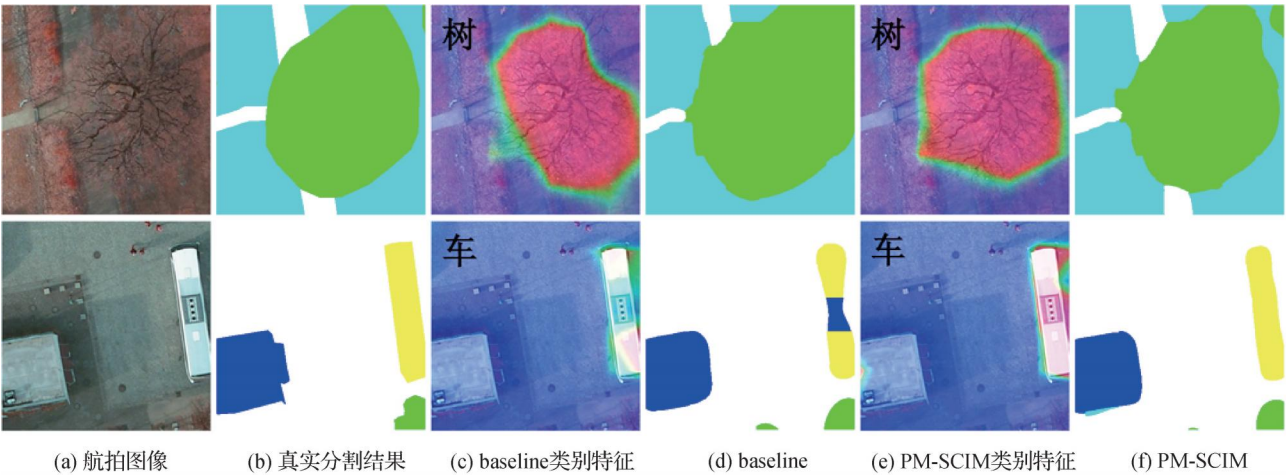


图 7 解混杂模块(DCM)的影响

Fig. 7 The effect of DCM ((a) airborne images;(b) ground truth;(c) class-wise features of baseline;(d) baseline;(e) class-wise features of PM-SCIM;(f) PM-SCIM)

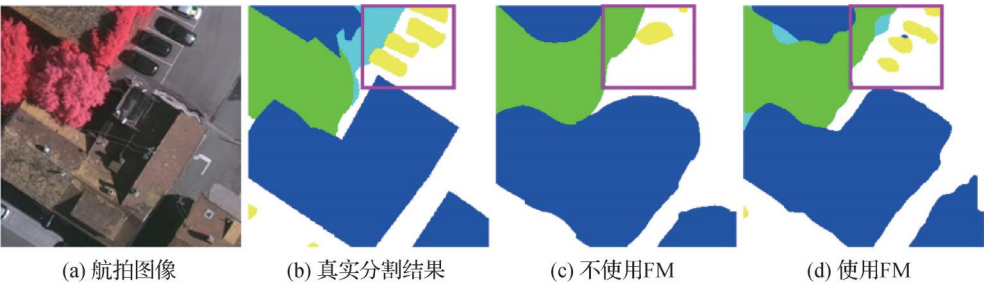


图 8 融合模块(FM)的影响

Fig. 8 The effect of FM ((a) airborne image;(b) ground truth;(c) without FM;(d) with FM)

表 2 相关深度学习方法在 Potsdam 数据集上的分割结果

Table 2 Segmentation results of related deep learning methods on Potsdam dataset

模型	地物类型					OA	mean F1	交并比 IOU
	不透水表面	建筑物	低植被	树	车			
PSPNet(Zhao 等,2017)	90. 7	94. 8	84. 1	85. 9	90. 5	88. 7	89. 2	71. 7
DeepLabv3 + (Chen 等,2018b)	89. 8	92. 6	82. 6	83. 3	93. 2	87. 1	88. 3	70. 0
S-RA-FCN(Mou 等,2019)	90. 7	94. 2	83. 8	85. 8	93. 6	88. 5	89. 6	72. 5
DANet(Fu 等,2019)	91. 6	95. 4	85. 1	86. 8	94. 2	89. 7	90. 6	74. 8
CCNet(Huang 等,2019)	90. 3	94. 1	84. 3	86. 0	90. 1	88. 3	89. 0	70. 9
CGFDN(Zhou 等,2021)	92. 1	95. 6	86. 3	87. 9	94. 9	90. 3	91. 4	75. 6
PM-SCIM	92. 1	95. 9	86. 0	87. 9	94. 5	90. 3	91. 3	75. 4

注:加粗字体表示各列最优结果。

分别提高了 0.7% 和 1%。值得注意的是,CGFDN 与 PM-SCIM 在 Potsdam 数据集上取得了几乎相同的 OA 和 Mean F1 分数。其原因是 Potsdam 数据集中存在的偏置问题,主要体现为车辆和建筑物部分

外观相似,导致车辆难以精确识别。而该问题通过引入车辆和道路的共生关系能够很好地缓解。因此,基于共生关系的 CGFDN 能与 PM-SCIM 达到相同的效果。而 Vaihingen 数据集中,主要由另一种现

表 3 相关深度学习方法在 Vaihingen 数据集上的分割结果
Table 3 Segmentation results of related deep learning methods on Vaihingen dataset

模型	地物类型					OA	mean F1	交并比 IOU
	不透水表面	建筑物	低植被	树	车			
PSPNet(Zhao 等,2017)	90.6	94.3	79.0	87.0	70.7	89.1	84.3	74.1
DeepLabv3 +(Chen 等,2018b)	91.0	94.2	79.9	87.8	85.0	89.7	87.6	78.5
S-RA-FCN(Mou 等,2019)	90.5	93.8	79.6	87.5	82.6	89.2	86.8	77.3
DANet(Fu 等,2019)	91.0	94.5	80.9	88.1	83.9	90.0	87.7	78.7
CCNet(Huang 等,2019)	90.8	94.3	80.5	88.1	72.1	89.7	85.2	75.2
CGFDN(Zhou 等,2021)	91.9	95.0	81.5	88.7	85.0	90.6	88.4	79.8
PM-SCIM	92.0	95.3	81.8	88.7	85.5	90.8	88.7	80.2

注:加粗字体表示各列最优结果。

象带来偏置问题,即树与草地相邻出现,导致边界模糊。而共生关系无法应对这类问题。借助因果干预,本文方法能够成功抑制偏置问题。所以,PM-SCIM 的 OA 和 Mean F1 分数分别比 CGFDN 高 0.2% 和 0.3%。

深度学习方法对比实验的可视化结果如图 9 和图 10 所示。可以看出,本文提出的 PM-SCIM 能够精确地识别整个车辆,保证车辆内部的语义一致性;同时,PM-SCIM 的分割结果在易混淆邻近物体的边缘细节上优于其他深度学习方法。

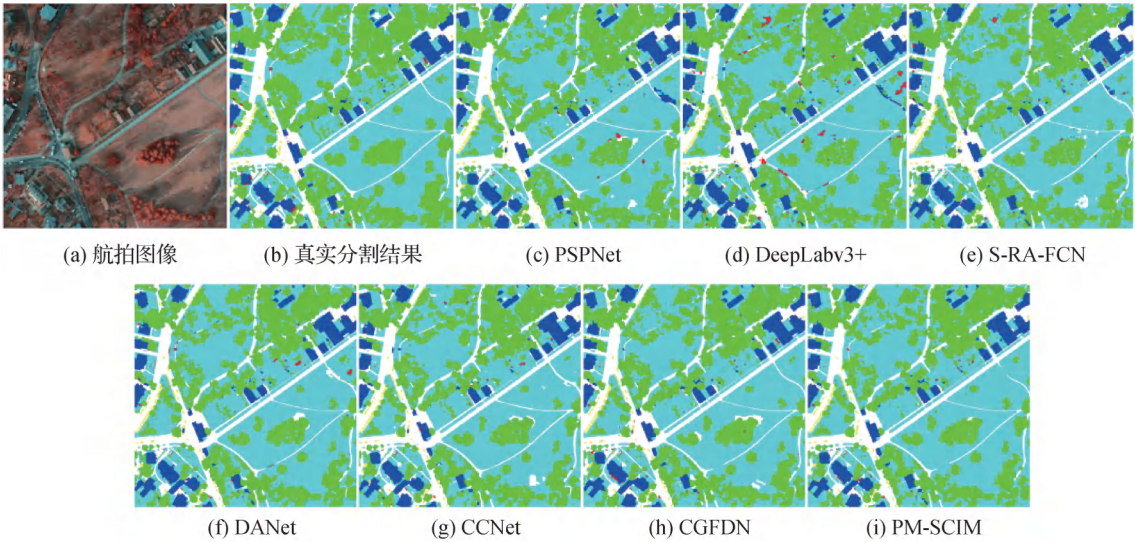


图 9 不同深度学习方法在 Potsdam 数据集上的可视化对比结果

Fig. 9 The visualization comparison between different deep learning methods on Potsdam dataset ((a) airborne image; (b) ground truth;(c) PSPNet;(d) DeepLabv3 + ;(e) S-RA-FCN;(f) DANet;(g) CCNet;(h) CGFDN;(i) PM-SCIM)

表 4 列出了相关深度学习方法的空间和时间复杂度,即模型参数量和运行时间。表中运行时间为对比方法使用 GTX TITAN X 显卡测试 100 个图像块(尺寸为 256 × 256)的时间。与 PSPNet、S-RA-FCN、DANet 和 CCNet 相比,PM-SCIM 仅增加了至多 10 MB 左右参数量和 6.5 s 运行时间,将总体准确率提升了至少 0.6%。PM-SCIM 虽然运行时间比 DeepLabv3 + 多了 1.5 s,但参数量实际低了近 32 M,且

总体准确率提升至少 1.1%。同样,相对于 CGFDN,PM-SCIM 虽然延长了运行时间,但减少了近 7 M 参数量,并能更全面地考虑偏置问题,在 Vaihingen 数据集上取得了更好的分割性能。

2.4.2 与基准方法对比

为了进一步验证 PM-SCIM 的有效性,与 7 个公开的基准方法进行对比,包括 CVEO* (Chen 等, 2018a)、UFMG_4(Nogueira 等,2019)、RIT_L* (Liu

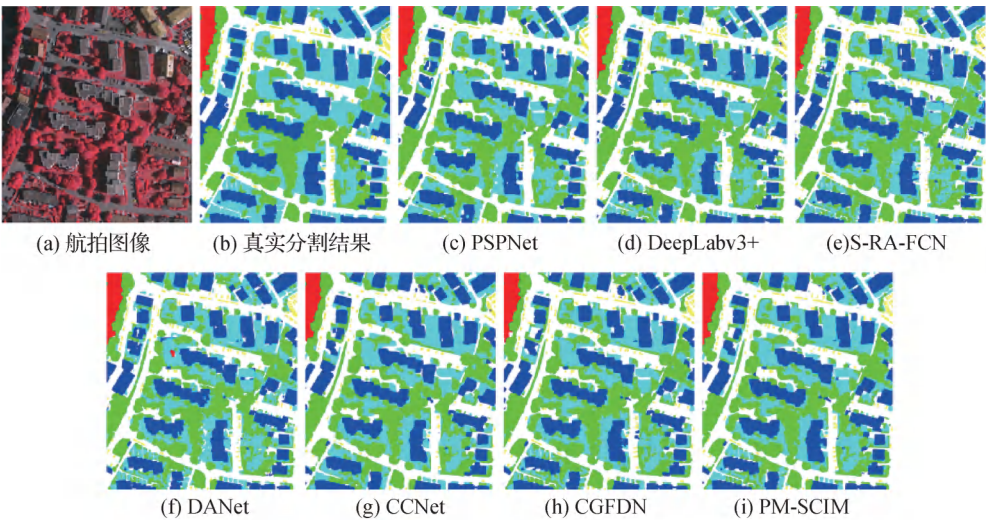


图 10 不同深度学习方法在 Vaihingen 数据集上的可视化对比结果

Fig. 10 The visualization comparison between different deep learning methods on Vaihingen dataset ((a) airborne image; (b) ground truth; (c) PSPNet; (d) DeepLabv3+; (e) S-RA-FCN; (f) DANet; (g) CCNet; (h) CGFDN; (i) PM-SCIM)

表 4 相关深度学习方法的时空复杂度对比

Table 4 Comparison of time and space complexity with related deep methods

模型	参数量/MB	时间/s
PSPNet(Zhao 等,2017)	16.8	1.0
DeepLabv3+(Chen 等,2018b)	54.9	5.9
S-RA-FCN(Mou 等,2019)	13.2	1.1
DANet(Fu 等,2019)	12.6	2.6
CCNet(Huang 等,2019)	13.3	5.4
CGFDN(Zhou 等,2021)	29.5	2.3
PM-SCIM	23.0	7.5

注:加粗字体表示各列最优结果。

等,2017)、UZ_1(Volpi 和 Tuia,2017)、SVL_*(Gerke,2015)、KLab_2(Kemker 等,2018)和 DLR_10(Marmanis 等,2018)。所有方法在 Potsdam 数据集的分割结果如表 5 所示。可以看出,本文提出的 PM-SCIM 的分割结果在 OA 和 mean F1 两项指标上均超过了其他公开基准方法。具体地说,PM-SCIM 将次好的模型 CVEO2 的 OA 值从 89.0% 提升至 90.3%,mean F1 从 90.1% 提升至 91.3%。主要是因为这些基准方法只关注如何提取多尺度上下文信息,忽略了数据集中的偏置问题。而 PM-SCIM 通过因果干预的方式有效缓解了数据集中偏置带来的混杂因子问题。图 11 展示了所有方法在 Potsdam 测

表 5 本文方法与基准方法在 Potsdam 数据集上的对比结果

Table 5 Comparison of results between the benchmark methods and ours on Potsdam dataset

模型	地物类型					OA	mean F1
	不透水表面	建筑物	低植被	树	车		
CVEO2(Chen 等,2018a)	91.3	94.5	86.5	87.7	95.1	89.0	91.0
UFMG_4(Nogueira 等,2019)	90.8	95.6	84.4	84.3	92.4	87.9	89.5
RIT_L7(Liu 等,2017)	91.2	94.6	85.1	85.1	92.8	88.4	89.8
UZ_1(Volpi 和 Tuia,2017)	89.3	95.4	81.8	80.5	86.5	85.8	86.7
SVL_1(Gerke,2015)	83.5	91.7	72.2	63.2	62.2	77.8	74.6
KLab_2(Kemker 等,2018)	89.7	92.7	83.7	84.0	92.1	86.7	88.4
PM-SCIM	92.1	95.9	86.0	87.9	94.5	90.3	91.3

注:加粗字体表示各列最优结果。

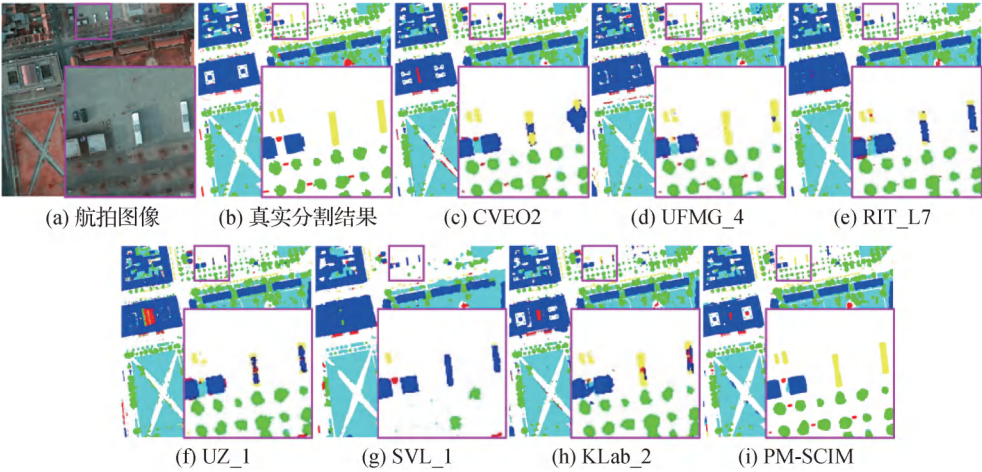


图 11 本文方法与基准方法在 Potsdam 的可视化对比结果

Fig. 11 Comparison of visualization between different benchmark methods and ours on Potsdam dataset ((a) airborne image; (b) ground truth; (c) CVEO2; (d) UFMG_4; (e) RIT_L7; (f) UZ_1; (g) SVL_1; (h) KLab_2; (i) PM-SCIM)

试图像上的分割结果,几乎所有基准方法得到的车辆分割结果都含有噪声。它们将部分车身错分成建筑物,破坏了车身之间的因果关系。PM-SCIM 通过因果干预的方式,成功保留了整车与部分车身之间的因果关系,抑制了对车具有干扰性的建筑物信息。表 6 为本文方法与基准方法在 Vaihingen 数据集

上的对比结果。可以看出,PM-SCIM 取得了最好的 OA 和 mean $F1$ 分数。与次好的模型 DLR_10 相比,OA 从 90.3% 提升到 90.8%,mean $F1$ 从 88.3% 提升到 88.7%。图 12 为本文方法与基准方法在 Vaihingen 数据集上的可视化对比结果。可以看出,PM-SCIM 能够准确识别出其他基准方法漏分的建筑物。

表 6 本文方法与基准方法在 Vaihingen 数据集上的对比结果

Table 6 Comparison of results between the benchmark methods and ours on Vaihingen dataset

模型	地物类型					OA	mean $F1$
	不透水表面	建筑物	低植被	树	车		
CVEO3(Chen 等,2018a)	91.0	93.0	81.3	88.3	83.0	88.6	87.3
UFMG_4(Nogueira 等,2019)	91.1	94.5	82.9	88.8	81.3	89.4	87.7
RIT_L8(Liu 等,2017)	89.6	92.2	81.6	88.6	76.0	87.8	85.6
UZ_1(Volpi 和 Tuia,2017)	89.2	92.5	81.6	86.9	57.3	87.3	81.5
SVL_3(Gerke,2015)	86.6	91.0	77.0	85.0	55.6	84.8	79.0
DLR_10(Marmanis 等,2018)	92.3	95.2	84.1	90.0	79.3	90.3	88.2
PM-SCIM	92.0	95.3	81.8	88.7	85.5	90.8	88.7

注:加粗字体表示各列最优结果。

图 13 展示了本文提出的 PM-SCIM 和基准方法在 Potsdam 和 Vaihingen 数据集上的分割精度和模型的参数量。可以看出,PM-SCIM 能够在很小的参数量情况下,获得比对比基准方法更好的分割性能。表明 PM-SCIM 能够在分割精度和复杂度两者之间取得很好的平衡。

3 结 论

针对航拍图像数据集中存在的偏置问题,即场景中的上下文先验对识别物体的干扰,本文提出一种基于渐进式多尺度因果干预模型的分割方法,在

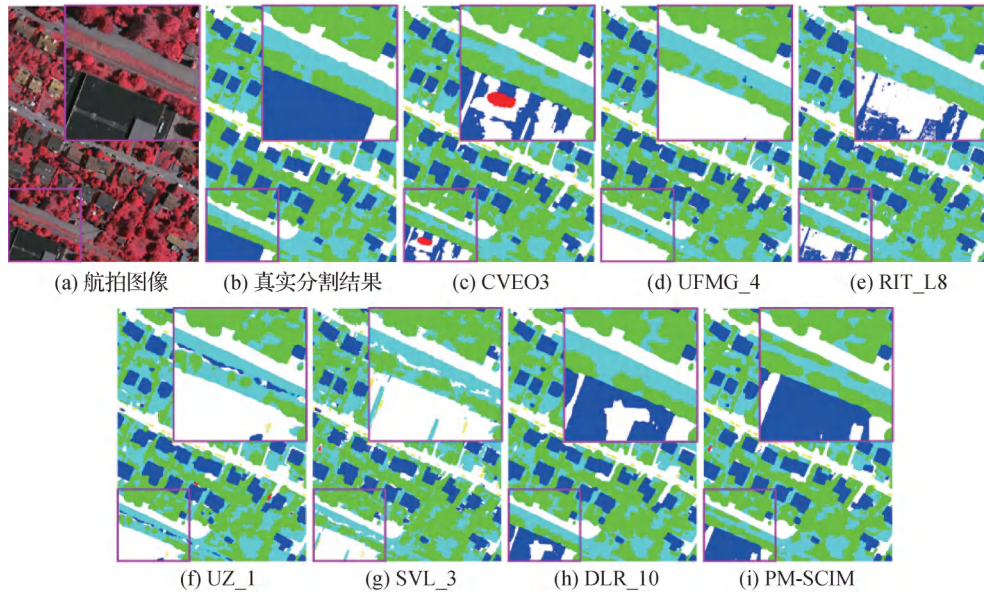


图 12 本文方法与基准方法在 Vaihingen 数据集上的可视化对比结果

Fig. 12 Comparison of visualization between different benchmark methods and ours on Vaihingen dataset ((a) airborne image; (b) ground truth; (c) CVEO3; (d) UFMG_4; (e) RIT_L8; (f) UZ_1; (g) SVL_3; (h) DLR_10; (i) PM-SCIM)

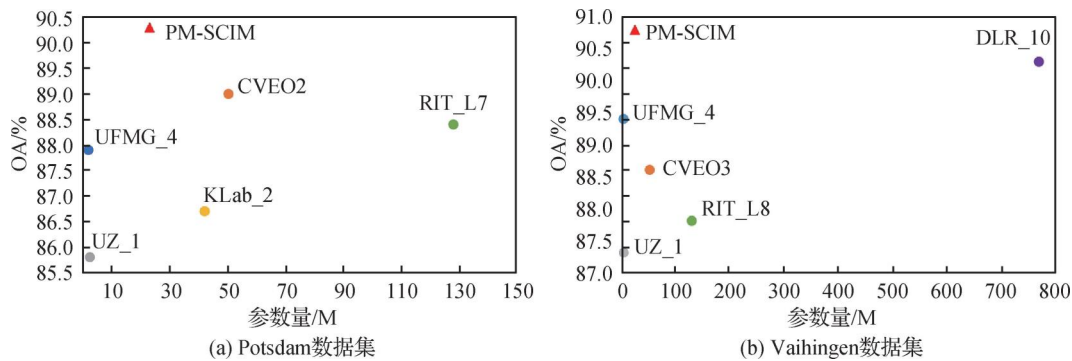


图 13 PM-SCIM 与基准方法的参数量

Fig. 13 The number of parameters of PM-SCIM and benchmark methods ((a) Potsdam dataset; (b) Vaihingen dataset)

ResNet18 结构上添加解混杂模块和融合模块。通过解混杂模块中的后门调整操作,将卷积特征分解成多个与特定的混杂因子相关的子特征。然后将所有混杂因子的子特征融合,以此抑制与目标物体伪相关但无因果关系的干扰物体。同时为了能够充分利用不同尺度的上下文信息,本文首先将 ResNet18 每个尺度的卷积特征经过解混杂模块得到解混杂特征,然后利用最小尺度的解混杂特征得到分割结果,并通过融合模块指导预测上一尺度分割结果。逐尺度循环该过程得到每个尺度对应的分割结果。最后将所有尺度的预测结果进行加权求和得到最终的分割结果。实验结果表明,与相关的深度学习方法相比,渐进式多尺度因果干预模型能够有效提升分割

性能,解决类似树和草地边缘模糊、车和建筑物难以区分的问题。同时,提出的模型能够得到优于大量基准方法的分割结果,进一步验证了在航拍图像分割领域的有效性。

本文方法能够缓解数据集偏置带来干扰信息的问题,但无法处理一些特殊场景。例如混杂因子是场景中已标注类别以外的物体。未来的工作主要是考虑通过无监督学习的方式,自适应地从数据集中提取存在的所有混杂因子,并使用本文方法进行分割结果预测。

参考文献 (References)

Audebert N, Le Saux B and Lefèvre S. 2016. Semantic segmentation of

- earth observation data using multimodal and multi-scale deep networks//Proceedings of the 13th Asian Conference on Computer Vision. Taipei, China: Springer; 180-196 [DOI: 10.1007/978-3-319-54181-5_12]
- Badrinarayanan V, Kendall A and Cipolla R. 2017. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12): 2481-2495 [DOI: 10.1109/TPAMI.2016.2644615]
- Chen G Z, Zhang X D, Wang Q, Dai F, Gong Y F and Zhu K. 2018a. Symmetrical dense-shortcut deep fully convolutional networks for semantic segmentation of very-high-resolution remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(5): 1633-1644 [DOI: 10.1109/JSTARS.2018.2810320]
- Chen L C, Zhu Y K, Papandreou G, Schroff F and Adam H. 2018b. Encoder-decoder with atrous separable convolution for semantic image segmentation//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer; 833-851 [DOI: 10.1007/978-3-030-01234-2_49]
- Ding L, Tang H and Bruzzone L. 2021. LANet: local attention embedding to improve the semantic segmentation of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 59(1): 426-435 [DOI: 10.1109/TGRS.2020.2994150]
- Fu J, Liu J, Tian H J, Li Y, Bao Y J, Fang Z W and Lu H Q. 2019. Dual attention network for scene segmentation//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE; 3141-3149 [DOI: 10.1109/CVPR.2019.00326]
- Gerke M. 2014. Use of the stair vision library within the ISPRS 2D semantic labeling benchmark (Vaihingen) [EB/OL]. [2021-10-20]. <https://research.utuente.cl/en/publications/use-of-the-stair-vision-library-within-the-isprs-2d-semantic-labe>
- Han B B, Zhang Y T, Pan Z X, Tai X Q and Li F F. 2020. Residual dense spatial pyramid network for urban remote sensing image segmentation. *Journal of Image and Graphics*, 25(12): 2656-2664 (韩彬彬, 张月婷, 潘宗序, 台宪青, 李芳芳. 2020. 残差密集空间金字塔网络的遥感图像分割. *中国图象图形学报*, 25(12): 2656-2664) [DOI: 10.11834/jig.190557]
- 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. Las Vegas, USA: IEEE; 770-778 [DOI: 10.1109/CVPR.2016.90]
- Huang Z L, Wang X G, Huang L C, Huang C, Wei Y C and Liu W Y. 2019. CCNet: Criss-cross attention for semantic segmentation//Proceedings of 2019 IEEE/CVF Conference on International Conference on Computer Vision. Seoul, Korea(South): IEEE; 603-612 [DOI: 10.1109/iccv.2019.00069]
- Kemker R, Salvaggio C and Kanan C. 2018. Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 145: 60-77 [DOI: 10.1016/j.isprsjprs.2018.04.014]
- Li X W, Li Y S and Zhang Y J. 2021. Weakly supervised deep semantic segmentation network for water body extraction based on multi-source remote sensing imagery. *Journal of Image and Graphics*, 26(12): 3015-3026 (李鑫伟, 李彦胜, 张永军. 2021. 弱监督深度语义分割网络的多源遥感影像水体检测. *中国图象图形学报*, 26(12): 3015-3026) [DOI: 10.11834/jig.200192]
- Liu W, Rabinovich A and Berg A C. 2015. ParseNet: looking wider to see better [EB/OL]. [2015-11-19]. <https://arxiv.org/pdf/1506.04579v2.pdf>
- Liu Y C, Fan B, Wang L F, Bai J, Xiang S M and Pan C H. 2018. Semantic labeling in very high resolution images via a self-cascaded convolutional neural network. *ISPRS Journal of Photogrammetry and Remote Sensing*, 145: 78-95 [DOI: 10.1016/j.isprsjprs.2017.12.007]
- Liu Y S, Piramanayagam S, Monteiro S T and Saber E. 2017. Dense semantic labeling of very-high-resolution aerial imagery and lidar with fully-convolutional neural networks and higher-order CRFs//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Hawaii, USA: IEEE; 1561-1570 [DOI: 10.1109/CVPRW.2017.200]
- Long J, Shelhamer E and Darrell T. 2015. Fully convolutional networks for semantic segmentation//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE; 3431-3440 [DOI: 10.1109/CVPR.2015.7298965]
- Long Y, Gong Y P, Xiao Z F and Liu Q. 2017. Accurate object localization in remote sensing images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 55(5): 2486-2498 [DOI: 10.1109/TGRS.2016.2645610]
- Marmanis D, Schindler K, Wegner J D, Galliani S, Datcu M and Stilla U. 2018. Classification with an edge: improving semantic image segmentation with boundary detection. *ISPRS Journal of Photogrammetry and Remote Sensing*, 135: 158-172 [DOI: 10.1016/j.isprsjprs.2017.11.009]
- Mou L C, Ghamisi P and Zhu X X. 2017. Deep recurrent neural networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(7): 3639-3655 [DOI: 10.1109/TGRS.2016.2636241]
- Mou L C, Hua Y S and Zhu X X. 2019. A relation-augmented fully convolutional network for semantic segmentation in aerial scenes//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE; 12408-12417 [DOI: 10.1109/CVPR.2019.01270]
- Nogueira K, Mura M D, Chanussot J, Schwartz W R and Dos Santos J A. 2019. Dynamic multicontext segmentation of remote sensing images based on convolutional networks. *IEEE Transactions on Geoscience and Remote Sensing*, 57(10): 7503-7520 [DOI: 10.1109/TGRS.2019.2913861]

- 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. Munich, Germany: Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]
- Sun W W and Wang R S. 2018. Fully convolutional networks for semantic segmentation of very high resolution remotely sensed images combined with DSM. IEEE Geoscience and Remote Sensing Letters, 15(3): 474-478 [DOI: 10.1109/LGRS.2018.2795531]
- Volpi M and Tuia D. 2017. Dense semantic labeling of subdecimeter resolution images with convolutional neural networks. IEEE Transactions on Geoscience and Remote Sensing, 55(2): 881-893 [DOI: 10.1109/TGRS.2016.2616585]
- Wang T, Huang J Q, Zhang H W and Sun Q R. 2020. Visual common-sense R-CNN//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 10757-10767 [DOI: 10.1109/CVPR42600.2020.01077]
- Yang X, Zhang H W, Qi G J and Cai J F. 2021. Causal attention for vision-language tasks//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 9842-9852 [DOI: 10.1109/CVPR46437.2021.00972]
- Yu S and Wang X L. 2021. Remote Sensing building segmentation by CGAN with multilevel channel attention mechanism. Journal of Image and Graphics, 26(3): 686-699 (余帅, 汪西莉. 2021. 含多级通道注意力机制的 CGAN 遥感图像建筑物分割. 中国图象图形学报, 26(3): 686-699) [DOI: 10.11834/jig.200059]
- Zhang H, Dana K, Shi J P, Zhang Z Y, Wang X G, Tyagi A and Agrawal A. 2018. Context encoding for semantic segmentation//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 7151-7160 [DOI: 10.1109/CVPR.2018.00747]
- 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. Hawaii, USA: IEEE: 2881-2890 [DOI: 10.1109/CVPR.2017.660]
- Zheng J H, Liu X Y and Wang X D. 2021. Single image cloud removal using u-net and generative adversarial networks. IEEE Transactions on Geoscience and Remote Sensing, 59(8): 6371-6385 [DOI: 10.1109/TGRS.2020.3027819]
- Zhou F, Hang R L and Liu Q S. 2021. Class-guided feature decoupling network for airborne image segmentation. IEEE Transactions on Geoscience and Remote Sensing, 59(3): 2245-2255 [DOI: 10.1109/TGRS.2020.3006872]

作者简介

周峰,男,讲师,主要研究方向为遥感图像分割。

E-mail: 13057588879@163.com

刘青山,通信作者,男,教授,主要研究方向为模式识别和图像处理。E-mail: qslu@nuist.edu.cn

杭仁龙,男,副研究员,主要研究方向为遥感图像处理。

E-mail: renlong_hang@163.com

徐超,男,教授,主要研究方向为大数据处理。

E-mail: 270174@nau.edu.cn

杨国为,男,教授,主要研究方向为大数据处理。

E-mail: 270178@nau.edu.cn