doi:10.11887/j.issn.1001-2486.24080004 http://journal.nudt.edu.cn

三维Mesh建筑物立面半监督对比学习语义分割方法

杜 春,成浩维,资文杰*,陈 浩,李 军
(国防科技大学 电子科学学院,湖南 长沙 410073)

摘 要:从三维Mesh数据中分割建筑物立面以识别对象,是三维场景理解的关键,但现有方法多依赖高成本的精细标注数据。针对该问题,提出了一种半监督学习方法,引入一种基于对比学习和一致性正则化的半监督语义分割(semi-supervised semantic segmentation based on contrastive learning and consistency regularization,SS_CC)方法,用于分割三维Mesh数据的建筑物立面。在SS_CC方法中,改进后的对比学习模块利用正负样本之间的类可分性,能够更有效地利用类特征信息;提出的基于特征空间的一致性正则化损失函数,从挖掘全局特征的角度增强了对所提取建筑物立面特征的鉴别力。实验结果表明,所提出的SS_CC方法在F1分数、mIoU指标上优于当前一些主流方法,且在建筑物的墙面和窗户上的分割效果相对更好。

关键词:三维Mesh数据;建筑物立面;对比学习;语义分割
中图分类号:TP753 **文献标志码**:A **文章编号**:1001-2486(2025)06-235-10

论文拓展

Semi-supervised semantic segmentation method for 3D Mesh building facades based on contrastive learning

DU Chun, CHENG Haowei, ZI Wenjie, CHEN Hao, LI Jun*
(College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, China)

Abstract: Semantic segmentation of building facades from 3D mesh data is essential for scene understanding but often relies on costly fine-grained annotations. In response to this issue, a semi-supervised learning approach was proposed, introducing a semi-supervised semantic segmentation method based on contrastive learning SS_CC(semi-supervised semantic segmentation based on contrastive learning and consistency regularization) to segment building facades in 3D mesh data. In the SS_CC method, the enhanced contrastive learning module exploited the class separability between positive and negative samples to more effectively utilize class-specific feature information. Additionally, the proposed feature-space consistency regularization loss improved the discriminative capability of the extracted building facade features by leveraging global feature representations. Experimental results show that the proposed SS_CC method outperforms some mainstream methods in F1 score and mIoU, and has relatively better segmentation performance on building walls and windows.

Keywords: 3D Mesh data; building facades; contrastive learning; semantic segmentation

近年来,随着传感器、遥感测绘、地理信息处理等技术的发展,三维数据在智慧城市建模与分析[1]、数字孪生城市应用[2]、城市规划建设[3]等领域获得了广泛应用。在城市三维模型数据中,建筑物数据占有相当大的比重,如何实现对建筑物三维数据的处理分析和可视化具有非常重要的意义。从理解三维数据语义的角度出发,建筑物表面的墙、窗等是其最重要的几何特征,实现精确的建筑物立面语义分割,不仅是大规模城市三维数据分析的基础,也可为相关行业业务提供建筑物的拓扑和语义描述,有利于提升城市基础设施建设和城市治理水平。

点云和Mesh是城市建筑物三维数据的两种典型格式。建筑物的三维点云数据一般通过激光扫描、摄像机捕获实际物体表面点或通过机器视觉算法计算生成点云等方式获取,其能够表示建

收稿日期:2024-08-07
基金项目:国家自然科学基金重点资助项目(U19A2058)
第一作者:杜春(1983—),男,云南玉溪人,副教授,博士,E-mail:duchun@nudt.edu.cn
***通信作者**:资文杰(1997—),男,湖南耒阳人,博士研究生,E-mail:ziwenjie@nudt.edu.cn
引用格式:杜春,成浩维,资文杰,等. 三维Mesh建筑物立面半监督对比学习语义分割方法[J]. 国防科技大学学报,2025,47(6):235-244.
Citation: DU C, CHENG H W, ZI W J, et al. Semi-supervised semantic segmentation method for 3D Mesh building facades based on contrastive learning[J]. Journal of National University of Defense Technology, 2025, 47(6): 235-244.

筑物的位置、形状等信息,但由于缺乏表面拓扑结构信息,并不适宜对建筑物立面的墙、窗等精细结构进行语义分析。建筑物的三维 Mesh 数据由一系列的三角形面片构成,三角形面片的组合能够较好表示建筑物的具体形状和表面拓扑结构,适合进行可视化渲染、形状编辑等操作,相比三维点云数据更适合对建筑物立面进行语义分析。

考虑到三维 Mesh 数据本身具有结构复杂、不连续、不规则等特点,直接采用传统语义分割技术并不能有效实现对建筑物立面的语义分析任务。解决这一问题的一种直观思想是将三维 Mesh 数据转换到平面影像进行分析,以便通过影像数据较好地分割和提取建筑物立面上的门窗等结构及其相应的空间位置等信息[4]。

当前,语义分割工作主要面向二维影像,主要基于卷积神经网络[5](convolutional neural network,CNN)并提出了多种全监督类型的分割方案。这类方法大多采用编码器 - 解码器结构,在降低图像分辨率的同时能够提取丰富的语义特征图信息。UNet 方法[6]设计了独特的 U 型神经网络结构,可以为语义分割获取多层次的细节信息。特征金字塔[7]的提出解决了多尺度特征图分析的难题。金字塔场景解析网络(pyramid scene parsing network,PSPN)方法[8]在其基础上进一步改进,具备提取全局上下文信息的能力。ResNet 方法[9]构建了跳跃连接的残差网结构,一定程度上克服了网络层数加深导致的梯度消失问题。DeepLab 方法[10]采用空洞卷积扩大了特征感受野,能够获取多尺度的特征。最近的一些工作为了使模型更加关注有利于语义分割的信息,在模型架构上倾向于设计多种典型的通道和空间注意力机制,使得语义分割的性能不断得到提升。但是,上述语义分割工作主要面向二维影像,且基于全监督机器学习理论展开,其效果好坏依赖于是否有大量精细标注数据的支撑。

鉴于三维 Mesh 语义标注相比二维影像语义标注代价更高,且现有开源标注数据集非常稀缺,目前基于全监督学习进行三维 Mesh 建筑物立面分割尚存在较大困难。

半监督语义分割适合解决包含大量无标签数据但仅含有少量有标签数据的语义分割问题,尤其在数据获取容易但标注极为困难的场景中应用比较广泛。其核心在于如何通过有标签数据更好地挖掘无标签数据的分布信息,从而在没有显著增加人工标注负担的基础上达到提升语义分割性能的目的。半监督语义分割一般遵循自我训练和一致性正则化[11]两种典型的范式。这两种范式都依赖于网络自动生成的伪标签的质量,这使得它们容易受到确认偏见的影响,并导致出现训练过程中的错误累积。置信度阈值、样本在训练迭代中的组合、多视图增强[12]、信息转移[13]等在一定程度上解决了上述问题。

半监督对比分割[14](semi-supervised contrastive segmentation,SSC)是一种将对比学习和半监督学习结合在一起的语义分割的方法,其基于 mean-teacher 结构并构造用于语义分割的表征学习模块,在二维影像数据上取得了较好的语义分割效果。mean-teacher 的核心思想是通过在训练过程中,将未标记样本的预测结果整合到模型的训练中,从而强化模型的学习,其包含一个教师模型和一个学生模型。不过,由于该方法仅基于正值进行对比学习,且在每个样本处仅对类特征进行对比,使得其对全局信息和样本类别分布信息尚未充分利用,因此将其用于三维 Mesh 建筑物立面语义分割时的效果尚不够理想。

为解决上述问题,利用大量的无标签数据和少量有标签数据进行联合学习,提出了一种基于对比学习和一致性正则化的半监督语义分割(semi-supervised semantic segmentation based on contrastive learning and consistency regularization,SS_CC)方法,实现了三维 Mesh 建筑物立面语义分析任务。该方法一方面采用双分支结构,提出了用类特征对比学习模块来加强建筑物立面图像中不同类样本的可分性;另一方面,所提的特征一致性正则化损失函数可以从全局角度更加精确地把握建筑物立面的结构信息,一定程度上缓解了在少量数据标注条件下实现建筑物立面语义分割的困难。此外,鉴于目前尚未调研到公开的建筑物立面标注数据集,基于课题组构建的长沙市区三维 Mesh 数据集进行了多组比较实验。

1 面向三维建筑 Mesh 立面的半监督对比学习分割方法

为了更好地针对三维 Mesh 建筑物立面数据实施语义分割,提出了一种新的基于对比学习的半监督语义分割方法 SS_CC。与代表性的半监督对比分割方法 SSC 相比,该方法在三个方面进行了改进:①与 SSC 方法仅考虑正样本不同,SS_CC 方法在对比学习过程中同时考虑了正样本和负样本,使得各语义类的信息能被充分应用;②与 SSC 方法仅在每个样本处对类特征进行对比不同,SS_CC方法在特征空间中增加了一致性正则化损

失模块，从全局角度进一步提升了对 Mesh 数据整体特征信息的描述和利用；③与 SSC 方法考虑二维影像语义分割不同，本文方法为了更好地适应三维 Mesh 建筑物立面半监督语义分割的需求，在模型损失函数上综合考虑了伪标签、类特征和全局特征三个方面，能够更充分地提取无标签数据的分布信息，进一步提高了模型的准确性和鲁棒性。

1.1　SS_CC 模型的总体框架

在本文中，研究所涉及数据集为部分标注的稠密标记数据集，其同时包含少量有标签数据和大量无标签数据。具体地，定义三维 Mesh 建筑物立面数据集为 $X=\{X_1,X_u\}$，其中有标签样本为 $X_1=\{x_1,y_1\}$，x_1 表示原始三维 Mesh 立面图像，y_1 为其对应的标签数据；无标签样本为 $X_u=\{x_u\}$，x_u 为无标签样本影像。

SS_CC 方法基于 mean-teacher 结构进行训练，学生模型和教师模型在不同的训练阶段发挥不同的作用，并相互提供指导信息用于训练。为便于描述，定义学生模型和教师模型的参数分别为 f_θ 和 f_ξ，首先通过梯度对学生模型的权重进行更新，然后再用学生模型的权重来更新教师模型。

SS_CC 的模型总体框架如图 1 所示，其中 y_u 表示 x_u 对应的伪标签。为了使模型更好地挖掘 Mesh 立面影像的潜在信息，对无标签样本影像通过随机裁剪、大小调整、翻转和旋转等空间变换进行了样本增广，生成增广后的样本集 $\overline{X_u}$。

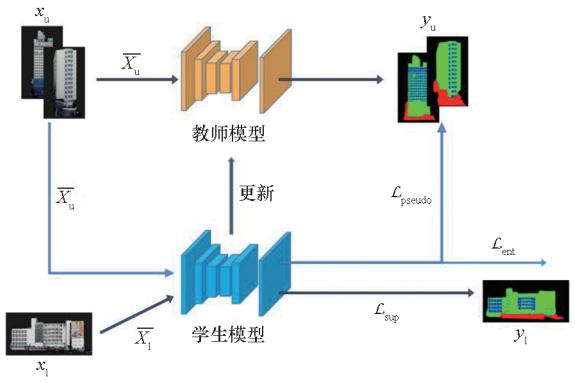


图 1　SS_CC 模型框架图

Fig. 1　Framework diagram of SS_CC model

在训练时，通过借助无标签的数据来提升有监督过程中的模型性能，以最小化加权交叉熵损失为优化目标。在基于全部已标注样本进行监督训练后，还要同时使用有标签数据和无标签数据对学生模型进行训练，其中无标签数据的训练由教师模型生成伪标签来进行：

$$y_u=f_\xi(\overline{X_u}) \tag{1}$$

$$H(x,y)=-\frac{1}{N}\sum\nolimits_N W_1 x W_2 \lg y \tag{2}$$

$$\mathcal{L}_{sup}=H(f_\theta(\overline{X_l}),y_l) \tag{3}$$

$$\mathcal{L}_{pseudo}=H(f_\theta(\overline{X_u}),y_u) \tag{4}$$

式中，H 表示标准的交叉熵损失，W_1、W_2 分别为两个输入的权重，N 为参与训练的数据长度，$\overline{X_l}$ 和 $\overline{X_u}$ 分别表示增广后的有标签样本和无标签样本，$f_\theta(\overline{X_l})$ 和 $f_\theta(\overline{X_u})$ 分别表示有标签样本和无标签样本的模型预测结果，$\mathcal{L}_{sup}$ 表示学生模型仅依靠有标签样本计算的全监督训练损失，$\mathcal{L}_{pseudo}$ 表示教师模型基于伪标签训练的损失函数。

在训练过程中，考虑到半监督学习过程中有标签样本数量相对较少，为进一步提升算法性能，本文将模型生成的高置信度的伪标签加入训练中，通过反复迭代更新模型，可以使决策边界更加趋于真实。具体地，决策边界一般不宜出现在稠密的数据点附近。因此，在训练时为了让模型预测输出具有更高的置信度，本文基于熵最小化理论增加了一项正则化损失 $\mathcal{L}_{ent}$，如下式所示：

$$\mathcal{L}_{ent}=-\sum f_\theta(x_u)\lg(f_\theta(x_u)) \tag{5}$$

分析可以发现，当上述模型预测的置信度比较低时，整体的熵值就会偏大，反之则越小。因此，通过对无标签数据预测的熵值最小化损失可以对决策界面的位置进行有效的优化。

当所有训练完成后，教师模型被赋予确定的网络权重。在模型测试阶段，输入待分析的建筑物立面数据进入教师模型，即可预测相应的语义分割结果。

1.2　对比学习过程

对比式方法是指通过将数据在特征空间分别与正例样本和负例样本进行对比，从而来学习样本特征表示的方法。由于对比式方法只在特征空间上学习类别区分性，其更侧重抽象的语义信息而不会过分关注像素细节，因此较为适用于语义分割。在本文中，SS_CC 方法首先通过全监督学习和伪标签训练，提取出置信度较高的特征。然后，专门设计对比学习模块并利用正负样本之间像素级特征的类可分性，以提升语义分割的性能。

图 2 给出了对比学习模块的具体构造。该模块以 DeepLabV2 作为基础模型，其中，W_θ 和 W_ξ 表示类特征经过学生模型和教师模型的自注意力模块生成的权重参数，$\mathcal{G}_\xi$ 和 $\mathcal{G}_\theta$ 代表教师模型和学生模型中附加的特征映射头，$\boldsymbol{Z}'$ 和 $\boldsymbol{Z}$ 为

模型提取到的特征经过这两个映射头后得到的低维度特征，q_θ 为学生模型附加的预测头，$\boldsymbol{P}$ 为经过预测头后的特征向量。这种不对称结构增加了学生模型和教师模型所提取到特征的不一致性，能有效地防止模型塌缩于零点，增强模型的鲁棒性。

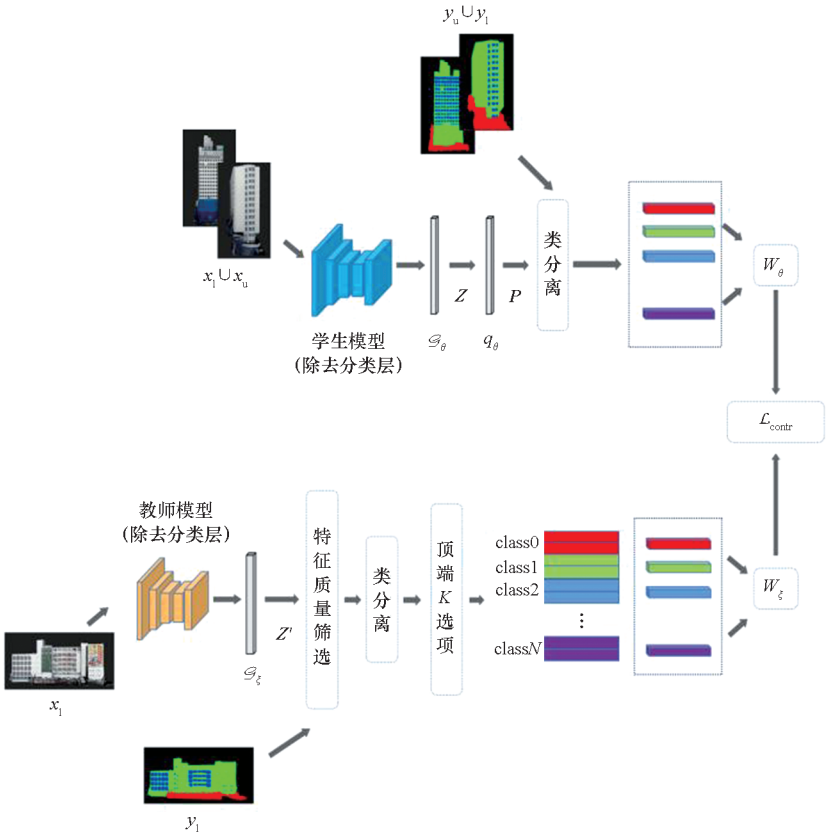


图 2 对比学习模块

Fig. 2 Contrastive learning module

在每次迭代中，通过教师模型 f_ξ 提取出有标签样本的特征值，这些特征经过一个特征映射头 $\mathcal{G}_\xi$，投影映射形成一个 n 维特征图 $\boldsymbol{Z}'$。定义教师模型和学生模型减去分类层后分别为 $f_{\xi\_}$ 和 $f_{\theta\_}$。提取的特征图 $\boldsymbol{Z}'$ 定义如下：

$$\boldsymbol{Z}' = \mathcal{G}_\xi(f_{\xi\_}(x_l)) \tag{6}$$

教师模型的整体输出 $f_\xi(x_u)$ 为每个像素点的类概率分布，分类结果 y 与置信度 T 定义如下：

$$y = \operatorname{argmax}(\operatorname{softmax}(f_\xi(x_l))) \tag{7}$$

$$T = \max(y) \tag{8}$$

上述模块提取的特征图中的每一个点可视为三维 Mesh 数据每个样本点对应的特征向量。由于标签与样本一一对应，根据标签可将特征图上的每个特征向量按类进行划分。随后，分别对每一类所提取的特征向量进行排序，挑选出置信度最高的 K 个特征向量存储到数据库中。

对比学习模块在具体工作时，首先向学生模型 f_θ 中输入有标签和无标签数据，并提取对应的特征图，经过特征映射头 $\mathcal{G}_\theta$ 投影映射成特征图 $\boldsymbol{Z}$。然后，为了增强学生模型和教师模型的不一致程度以提高模型的鲁棒性，加入一个预测头 q_θ，将 $\boldsymbol{Z}$ 映射为具有相同维度的特征图 $\boldsymbol{P}$，并根据对应的标签和伪标签进行特征向量的分类。最终得到的特征图 $\boldsymbol{P}$ 定义如下：

$$\boldsymbol{P} = q_\theta(\mathcal{G}_\theta(f_{\theta\_}(x_l))) \tag{9}$$

最后，对比学习模块将分离的类特征与数据库中的特征向量进行比对，相同类的特征作为正样本，不同类的特征作为负样本，强制每类的特征向量相似，同时尽可能增大不同类特征之间的差异性。

1.3 损失函数

SS_CC 模型的损失函数由三个部分组成，即全监督（伪标签）损失函数、对比损失函数和一致性正则化损失函数，具体如下：

$$\mathcal{L} = \mathcal{L}_{sup} + \boldsymbol{\lambda}_{pseudo}\mathcal{L}_{pseudo} + \boldsymbol{\lambda}_{ent}\mathcal{L}_{ent} + \boldsymbol{\lambda}_{consis}\mathcal{L}_{consis} + \boldsymbol{\lambda}_{contr}\mathcal{L}_{contr} \tag{10}$$

式中，$\mathcal{L}_{contr}$ 为对比损失，$\mathcal{L}_{consis}$ 为一致性正则化损失，$\boldsymbol{\lambda}_{pseudo}$、$\boldsymbol{\lambda}_{ent}$、$\boldsymbol{\lambda}_{consis}$ 和 $\boldsymbol{\lambda}_{contr}$ 为对应权重。其中，$\mathcal{L}_{sup}$、$\mathcal{L}_{pseudo}$、$\mathcal{L}_{ent}$ 已在 1.1 节进行定义，下面对后两项损失函数进行介绍。

1.3.1　对比损失 $\mathcal{L}_{\text{contr}}$

在对比学习中需要进行正负样本的构建,在训练过程中需要不断地减小正样本之间的距离,同时拉大正负样本之间的差距。因此正负样本选取的好坏与模型的精度高低息息相关。在 SS_CC 方法中,提出采用正负样本训练策略,将学生模型和教师模型所提取的相同类样本特征向量作为正样本,不同类样本的特征向量作为负样本。

定义 $\boldsymbol{P}_c=\{\boldsymbol{p}_c\}$ 为学生模型中每一类的特征预测值;$\boldsymbol{Z}'_c=\{\boldsymbol{z}'_c\}$ 为教师模型提取的经过映射头后的每一类特征预测,即 $\boldsymbol{Z}'_c=\mathcal{G}_\xi(f_{\xi\_}(x_u))$。

在对比学习中,要求预测的特征向量 $\boldsymbol{p}_c$ 与 $\boldsymbol{z}'_c$ 尽可能地相似,因此本文采用余弦相似度来表示两者之间的相似度,即

$$Cos(\boldsymbol{p}_c,\boldsymbol{z}'_c)=\frac{<\boldsymbol{p}_c,\boldsymbol{z}'_c>}{\|\boldsymbol{p}_c\|\|\boldsymbol{z}'_c\|_2} \tag{11}$$

式中:$\boldsymbol{p}_c$ 表示计算样本;$\boldsymbol{z}'_c$ 为其对应的正样本;Cos 表示余弦相似度,用于计算两个输入之间的相似性。

为了使对比学习更具有针对性,本文基于文献[15]中的方法,通过一个注意力模块为特征向量 $\boldsymbol{p}_c$ 与 $\boldsymbol{z}'_c$ 分配相应的权重因子 $w_{\boldsymbol{p}_c}$ 和 $w_{\boldsymbol{z}'_c}$,此时可定义两个特征向量的加权距离为

$$D_{\boldsymbol{p}_c,\boldsymbol{z}'_c}=-w_{\boldsymbol{p}_c}w_{\boldsymbol{z}'_c}Cos(\boldsymbol{p}_c,\boldsymbol{z}'_c) \tag{12}$$

由此,可定义 SS_CC 模型的对比损失函数为

$$\mathcal{L}_{\text{contr}}=-\lg\frac{D_{\boldsymbol{p}_c,\boldsymbol{z}'_c}/\delta}{D_{\boldsymbol{p}_c,\boldsymbol{z}'_c}/\delta+\Delta} \tag{13}$$

$$\Delta=\sum D_{\boldsymbol{p}_c,\boldsymbol{z}'_{c'}}/\delta \tag{14}$$

式中:$\boldsymbol{z}'_{c'}$ 为计算样本对应的负样本,即不同类的特征向量;δ 为温度系数,用于调节模型对困难样本的关注强度。较小的温度系数能够促使模型更加注重区分本样本与最相似样本之间的细微差异,从而提升样本间的可分性。

1.3.2　一致性正则化损失 $\mathcal{L}_{\text{consis}}$

一致性正则化的原理如图 3 所示。从三维 Mesh 数据语义分割的角度来说,其基本假设是:考虑将输入图像进行两种不同的数据增广(例如颜色扰动、拉伸、加噪等强增强方式,以及图像裁剪、缩放、翻转等弱增强方式),模型所提取到的特征向量应该是相似的。实质上,这一假设的成立意味着微小扰动前后训练样本输入模型后的预测输出应保持一致,其鼓励训练后的模型对样本的邻域具有光滑性,即从全局的角度对语义分割所需提取的特征进行了约束,有利于增强三维 Mesh 数据语义分割的鲁棒性。

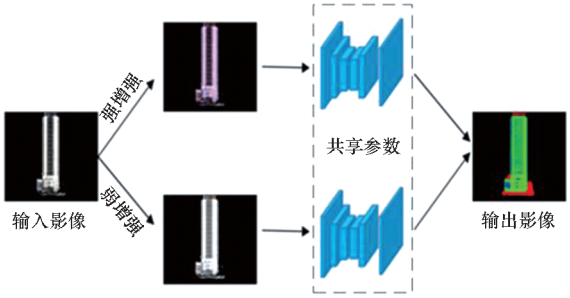


图 3　一致性正则化的原理

Fig. 3　Principle of consistency regularization

具体地,为了使模型更好地提取有利于三维 Mesh 数据语义分割的特征,SS_CC 方法提出一种改进的一致性正则化损失函数用于模型的训练,具体定义如下:

$$\mathcal{L}_{\text{consis}}=Cos(\mathcal{G}_\xi(f_{\theta\_}(\overline{X_1})),\mathcal{G}_\xi(f_{\theta\_}(\overline{X_2}))) \tag{15}$$

$$Cos(x,y)=\frac{<x,y>}{\|x\|_2\|y\|_2} \tag{16}$$

式中,$\overline{X_1}$、$\overline{X_2}$ 分别表示通过两种不同数据强增强的输入样本,由于在该任务中去除了模型的分类层,因此会得到一个维度非常高的特征向量,增加了模型的参数和计算的复杂度;且用过高维度表示数据会造成数据冗余,大量的维度无法保存有效的特征数据,容易造成模型的不稳定,甚至导致模型的性能下降。为了应对这个挑战,同时不增加模型的参数,SS_CC 方法将模型提取的特征通过特征映射头 $\mathcal{G}_\xi$,将高维度数据映射到较低的维度进行分析。

2　实验及结果分析

2.1　实验设置

本文所有实验均在 2 个 GeForce GTX 2080Ti GPU, 256 GB RAM 服务器上运行,依托于 PyTorch 深度学习框架实现。实验采用的基础网络框架为 DeepLabV2,特征提取层为标注的 ResNet101 模型的前 4 层。

实验一共训练 10 000 批次,采用动量为 0.9 的随机梯度下降(stochastic gradient descent,SGD)优化函数,初始学习率设置为 2×10^{-3},并在每个周期以 0.9 为指数进行衰减。损失函数的权值分别设置为:$\lambda_{\text{pseudo}}=1$,$\lambda_{\text{ent}}=1$,$\lambda_{\text{contr}}=0.01$。为了使模型预测的质量具有一定的可信度,前 4 000 批次设置 $\lambda_{\text{contr}}=0$,并提前 1 000 批次开始进行数据库的构建。在进行全监督和伪标签训练时,λ_{consis} 和 λ_{contr} 的值定义为 0。

2.2 数据集

采用2018年于长沙市高新产业园区拍摄的一组分辨率为0.15 m的倾斜摄影数据，提取出100张建筑物立面作为无标签数据样本。最终实现数据集中有标签样本、无标签样本和测试数据的比例为20∶60∶20。数据集命名为“CHANG_SHA_2022”。

建筑物立面提取及数据集构建流程如图4所示。第一步，将三维倾斜摄影数据转换为OBJ格式的文件；第二步，将对应区域的二维正射遥感影像进行语义分割，提取建筑物的角点信息；第三步，利用建筑物的角点信息对三维OBJ格式的数据进行建筑物单体化；第四步，调用Blender等开源软件进行数据读取和三维建模，得到三维建筑模型后通过不同的视角进行视场分析；第五步，通过人工筛选出具有清晰完整分割边缘的立面图像，根据需求进行标注并构建建筑物立面数据集。

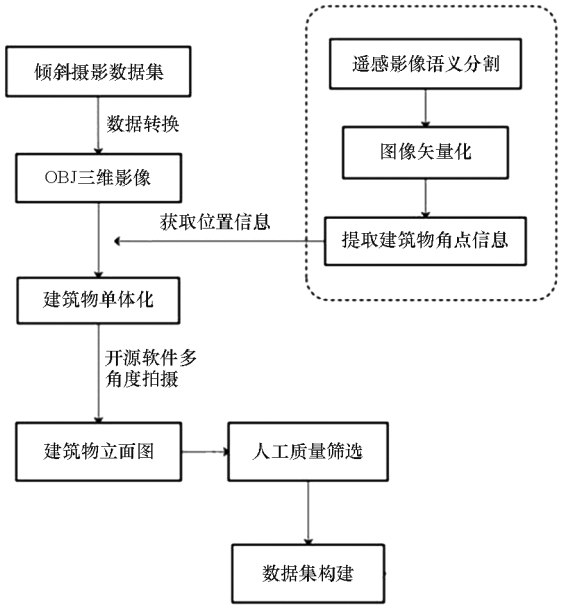


图4 建筑物立面提取及数据集构建流程

Fig. 4 Building facade extraction and dataset construction process

2.3 实验分析

为了验证SS_CC方法的有效性，本文挑选出几种代表性的模型进行对比试验。实验在1/16、1/30以及1/50量级的数据集场景中进行。本文采用F1分数（F1）和平均交并比（mean intersection over union，mIoU）对实验结果进行评估。下面对参与对比实验的半监督方法进行介绍。

1）Adversarial[16]：该模型基于生成对抗网络（generative adversarial network，GAN）进行语义分割，将生成对抗网络中的生成器替换为一个分割网络，模型的鉴别器采用全卷积网络，输入生成的概率图或真实标签，由鉴别器判断输入是否为模型生成。

2）s4GAN[17]：模型采用一个双分支网络来处理半监督语义分割问题。一条支路为标准的分割网络，用于生成输入图像中每个像素的类标签。另一条支路为mean-teacher结构的分类器。

3）ClassMix[18]：一种新的数据增强策略，从一幅图像中删除一半的预测类，并将它们粘贴到另一幅图像上，形成一个新的样本，同时不需要真正的注释，利用一致性正则化和伪标签对模型进行训练。

4）FixMatch[19]：是一致性正则化和伪标签训练的完整实现，将样本经过数据增强送入模型中，有标签数据和无标签数据分别通过标签和样本生成的伪标签进行训练。

表1展示了不同模型的预测结果。从表1可以看出，SS_CC方法在半监督情况下取得了最优的性能，在多项指标上都取得了良好的成绩。同时可以发现，随着可用标签数据的减少，不同模型的性能差异越来越大，这表明本文提出的模型在极少量的数据集的情况下仍能发挥比其他模型更加优秀的性能。

在1/16量级的情况下，SS_CC相较于SSC模型的整体性能提高不大，可能的原因是此时根据图像所能挖掘到的特征信息已经趋于饱和，尤其是在少量分类的任务中，基本满足根据特征对每一类进行判断的条件，难以有更大的性能提升。

当模型处在1/30和1/50量级的情况下，SS_CC模型的性能提升显著。这是由于此时数据量较少，SS_CC模型能够更好地挖掘出其他模型所不能理解的语义信息。在使用负样本和特征一致性正则化时，从更加全面的角度对每个样本点进行描述，提高了语义分割的性能。

同时需要说明，在表1给出实验结果中，SS_CC模型在分割墙面和杂质两类时具有优势，在分割窗户类时多数情况下也优于其他模型，但有时在mIoU指标上略低于SSC模型。这主要是由于在建筑物立面数据中墙面像素点一般较多，从中采集的正负样本自然较多，SS_CC模型相比其他模型能够更好地学习到墙面的空间特征。对于窗户而言，其所占像素点一般较少，模型采集的正负样本相对较少，且同一立面数据内不同窗户间纹理特征差异相比墙面也偏大，使得模型学习

表 1　不同模型的语义分割结果

Tab. 1　Semantic segmentation results of different models

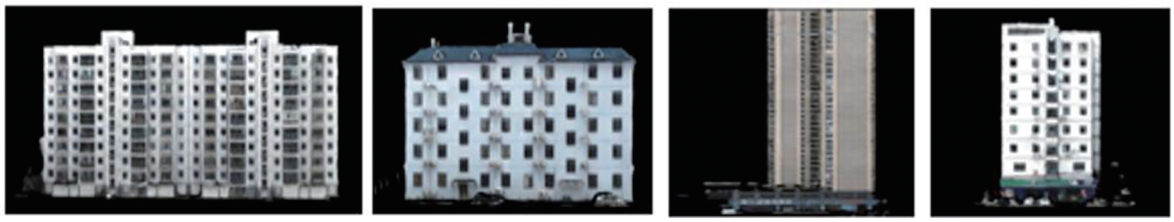
模型	1/16					1/30					1/50				
	F1	mIoU	墙面(mIoU)	窗户(mIoU)	杂质(mIoU)	F1	mIoU	墙面(mIoU)	窗户(mIoU)	杂质(mIoU)	F1	mIoU	墙面(mIoU)	窗户(mIoU)	杂质(mIoU)
Adversarial	0.780 0	0.653 5	0.841 2	0.467 4	0.652 1	0.624 6	0.501 1	0.724 7	0.151 0	0.627 8	0.564 3	0.437 5	0.807 5	0.194 4	0.310 7
s4GAN	0.730 0	0.589 3	0.740 7	0.388 3	0.639 1	0.554 2	0.452 0	0.709 7	0.040 5	0.605 9	0.628 0	0.486 1	0.765 0	0.258 6	0.435 0
ClassMix	0.549 9	0.407 2	0.694 4	0.217 0	0.310 3	0.569 9	0.428 4	0.697 3	0.188 7	0.399 2	0.309 6	0.241 7	0.653 5	0.039 7	0.032 1
FixMatch	0.814 6	0.696 4	0.782 7	0.573 9	0.732 8	0.798 2	0.668 1	0.764 3	0.567 3	0.672 9	0.753 5	0.612 0	0.721 1	0.460 1	0.656 2
SSC	0.824 9	0.706 2	0.790 8	0.597 1	0.731 9	0.808 8	0.681 9	0.768 2	0.593 2	0.684 5	0.753 4	0.612 3	0.736 4	0.464 7	0.636 0
SS_CC	0.825 7	0.707 4	0.841 4	0.594 5	0.736 4	0.813 5	0.690 4	0.782 4	0.570 9	0.718 0	0.763 1	0.624 8	0.800 2	0.475 0	0.659 2

到的窗户特征有时不够精确,影响了该类语义分割效果的进一步提升。

为了直观地展示模型的性能,本文将 1/30 数据量下的 4 个三维 Mesh 数据的实验结果进行可视化,结果如图 5 所示。通过对比观察可以看出:当数据量较少时,基于生成对抗网络的两种模型(Adversarial 和 s4GAN)难以达到良好的实验效果,它们对于杂质类和窗户类不能很好地进行区分。可能是由于这两种模型使用了 GAN 模型,生成器根据样本信息生成样本标签,再由判别器对每个类进行判定,当数据量较少时,模型强调数据的真实性,在墙面类上能够有很好的分辨能力。但当模型分辨较为相似的特征类时,如窗户和杂质,较难区分它们的真实性,导致模型误判,准确性降低。

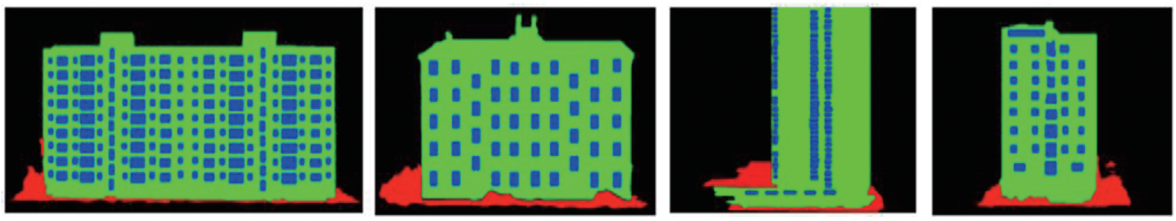
ClassMix 方法采用图像增强策略进行模型训练,当两类特征较为接近时(窗户和杂质),模型会比较敏感,从而对这两类数据产生误判,使模型的精度降低。

FixMatch 和 SSC 方法是两种代表性的半监督语义分割模型:FixMatch 方法采用数据增广策略能够有效地保留实验样本的原始特征,采用伪标



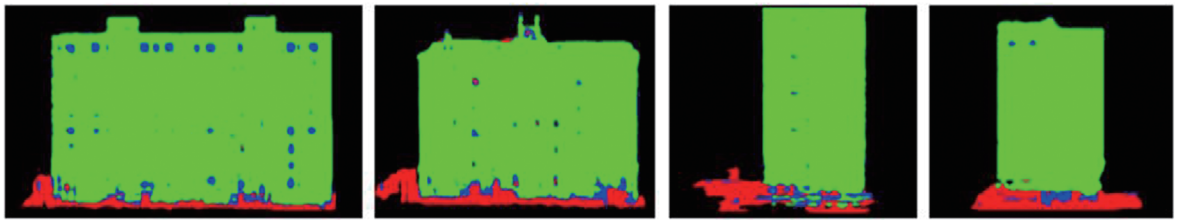
(a) 原始图像

(a) Original image

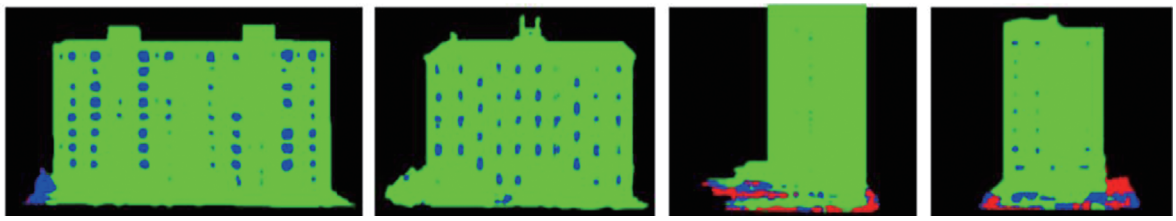


(b) 稠密标签数据

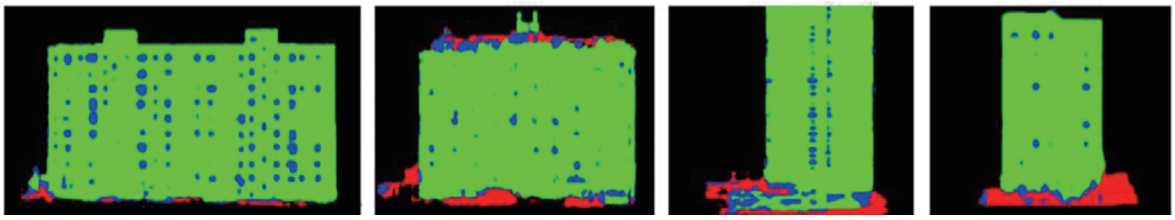
(b) Dense label data



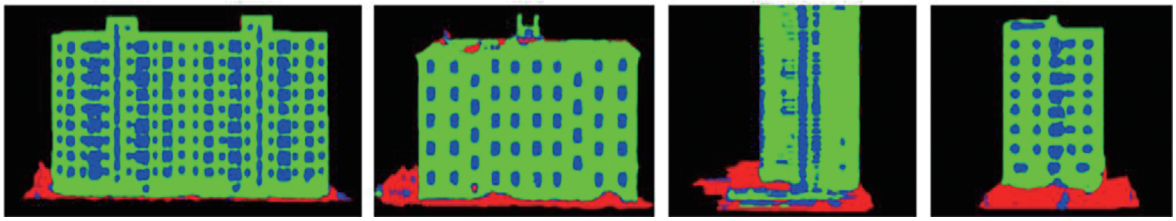
(c) s4GAN



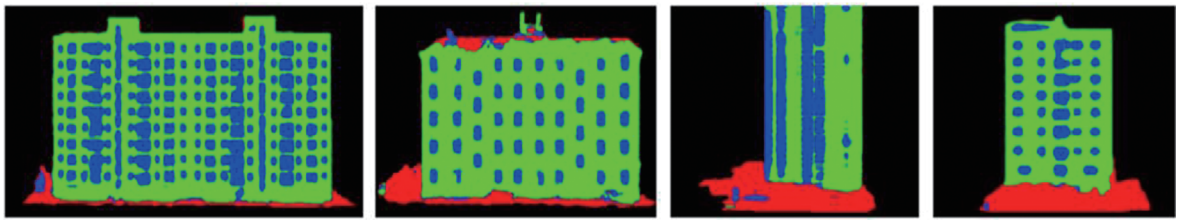
（d）ClassMix



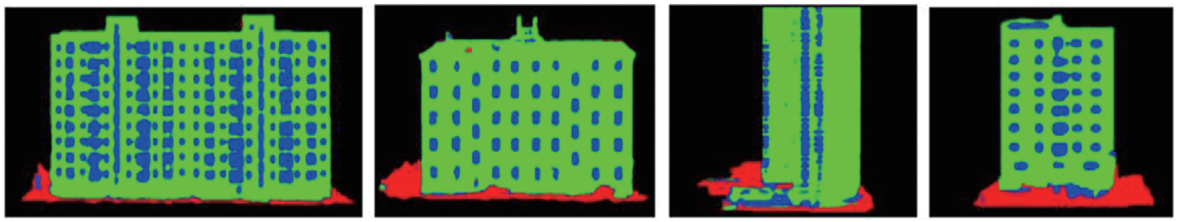
（e）Adversarial



（f）FixMatch



（g）SSC



（h）SS_CC

图 5　建筑物立面分割结果

Fig. 5　Facade segmentation results of buildings

签训练可以有效地挖掘出图像的隐含信息；SSC 方法引入对比学习策略，基于 mean-teacher 结构提出表征学习模块，并在不同样本之间强调像素级特征的类可分性。这两种半监督方法充分发挥了半监督学习的优势，在三维 Mesh 数据立面分割的效果方面优于前面三种方法。但是由于在对比学习中只考虑正值的对比学习，在类特征学习中缺乏对全局信息的描述，这两个方法在三维 Mesh 数据中对细节的分割效果还不理想。

SS_CC 在现有方法上做出改进，在对比学习中加入负样本，并通过一致性正则化损失来提升对特征的鉴别力，使其对墙面和窗户的识别效果得到较大的提升，有效提升了三维 Mesh 建筑物立面半监督语义分割方法的性能。

2.4　消融实验分析

为了更好地验证模型各个组成部分的有效性，在 1/30 的数据量上进行消融实验，将各个组成部分进行拆分，观察实验结果，以此来验证本文

提出的模型的有效性。为了更好地对消融实验结果进行展示,本文使用 DeepLabV2 作为基础模型,并根据不同的训练方式对模型进行命名。

1)SS_CC_sup:仅使用当前量级的标签数据,采用标准的交叉熵损失对模型进行全监督训练。

2)SS_CC_pse:为了验证加入 mean-teacher 结构和伪标签训练能够有效地提升模型性能,在 SS_CC_sup的基础上引入伪标签训练。

3)SS_CC_contr:为了验证对比模块的有效性,在 SS_CC_pse 中加入对比损失。

4)SS_CC_neg:为了验证不同类特征能够作为负样本加入模型训练并能取得良好的训练效果,引入负样本训练。

5)SS_CC:最后在模型的训练中加入一致性正则化损失来验证其在半监督学习中能够帮助模型性能的提升,同时能够很好地适应当前的训练任务。

表 2 列出了每种消融模型所采用的训练策略以及它们的 mIoU,展示了加入不同的训练模块对模型性能的影响。其中 $\mathcal{L}_{\text{contr}}^{+}$ 和 $\mathcal{L}_{\text{contr}}^{-}$ 分别代表仅采用正样本和添加负样本训练的对比损失。可以观察到,模块之间的作用是相辅相成的,从第二行到第五行不断地加入半监督学习模块,使得模型的性能稳步提升。也就是说,不同模块提取到的无标签数据的信息是不重复的,后续的模块对之前的模块有了进一步的补充。

表 2　消融不同模块的实验结果

Tab. 2　Experimental results of ablating different modules

模型	$\mathcal{L}_{\text{sup}}$	$\mathcal{L}_{\text{pseudo}}$	$\mathcal{L}_{\text{contr}}^{+}$	$\mathcal{L}_{\text{contr}}^{-}$	$\mathcal{L}_{\text{consis}}$	mIoU
SS_CC_sup	√	×	×	×	×	0.602 8
SS_CC_pse	√	√	×	×	×	0.668 1
SS_CC_contr	√	√	√	×	×	0.681 5
SS_CC_neg	√	√	√	√	×	0.688 3
SS_CC	√	√	√	√	√	0.690 4

通过对标签数据的训练,模型能够生成置信度较高的伪标签来拟合无标签数据,根据聚类假设,这些置信度较高的点,其伪标签的可信度是非常高的。通过伪标签训练可以最小化无标签数据的类概率条件熵,促进类之间的低密度分离,可以有效地利用无标签数据的分布信息。另外,采用 mean-teacher 结构来生成伪标签数据时,仅通过梯度对学生模型进行更新,不会引入额外的计算负担。加入对比损失则从类特征的角度对每个样本点进行分析。双分支网络提供了输入和输出特征的不一致性,但对于同一类物体而言,每个类特征是具有不变性的,这样就能很好地实现对于无标签数据的强制性类特征对比,使相同类的样本不断接近,不同类的样本相互远离。加入负样本进一步提高了模型的精度和鲁棒性,尤其是在数据量较少的情况下,能够提供更多的参考信息。加入一致性正则化损失函数,从挖掘全局特征的角度进一步增强了所提取特征的鉴别力。

3　结论

本文面向三维 Mesh 数据建筑物立面语义分析的应用场景,提出了一种基于对比学习的半监督语义分割方法 SS_CC。SS_CC 有效地利用无标签数据中的信息,缓解标签数据不足的问题。利用对比学习挖掘出正负样本之间的关联性,增强模型语义分割性能。

SS_CC 采用 mean-teacher 结构,使用对应的标签将特征按每个样本点的类进行划分,构建类特征的数据库;通过学生模型提取所有数据的类特征向量进行对比,相同类作为正样本,不同类作为负样本,减小正样本之间的距离,同时拉大负样本之间的差异,有效地加强建筑物立面图像中不同类样本点的可分性,增强模型的特征提取和识别能力。为了更好地利用无标签数据,SS_CC 加入了特征空间的一致性正则化损失,让模型对产生扰动的数据做出一致性的判断,从全局特征的角度对模型进行约束,使模型从更高的维度理解图像信息。

参考文献(References)

[1] ARCURI N, DE RUGGIERO M, SALVO F, et al. Automated valuation methods through the cost approach in a BIM and GIS integration framework for smart city appraisals [J]. Sustainability, 2020, 12(18): 7546.

[2] SHAHAT E, HYUN C T, YEOM C. City digital twin potentials: a review and research agenda[J]. Sustainability, 2021, 13(6): 3386.

[3] BUSH J, DOYON A. Building urban resilience with nature-based solutions: how can urban planning contribute? [J]. Cities, 2019, 95: 102483.

[4] YU B L, LIU H X, WU J P, et al. Automated derivation of urban building density information using airborne LiDAR data and object-based method [J]. Landscape and Urban Planning, 2010, 98(3/4): 210 - 219.

[5] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 3431 - 3440.

[6] YAN X Y, TANG H, SUN S L, et al. AFTer-UNet: axial fusion transformer UNet for medical image segmentation[C]// Proceedings of the IEEE/CVF Winter Conference on

Applications of Computer Vision (WACV), 2022: 3270 – 3280.

[7] GONG Y Q, YU X H, DING Y, et al. Effective fusion factor in FPN for tiny object detection[C]//Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), 2021: 1159 – 1167.

[8] ZHAO H S, SHI J P, QI X J, et al. Pyramid scene parsing network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2881 – 2890.

[9] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 770 – 778.

[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, 2018, 40(4): 834 – 848.

[11] YUAN X H, SHI J F, GU L C. A review of deep learning methods for semantic segmentation of remote sensing imagery[J]. Expert Systems with Applications, 2021, 169: 114417.

[12] HARTLEY R, ZISSERMAN A. Multiple view geometry in computer vision[M]. 2nd ed. Cambridge: Cambridge University Press, 2003.

[13] WU J W, FAN H Y, LI Z Y, et al. Information transfer in semi-supervised semantic segmentation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(2): 1174 – 1185.

[14] ALONSO I, SABATER A, FERSTL D, et al. Semi-supervised semantic segmentation with pixel-level contrastive learning from a class-wise memory bank[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021: 8199 – 8208.

[15] SUN S Y, CHEN L, SLABAUGH G, et al. Learning to sample the most useful training patches from images[EB/OL]. [2024 – 03 – 15]. https://arxiv.org/abs/2011.12097.

[16] HUNG W C, TSAI Y H, LIOU Y T, et al. Adversarial learning for semi-supervised semantic segmentation[EB/OL]. [2024 – 03 – 17]. https://arxiv.org/abs/1802.07934v2.

[17] MITTAL S, TATARCHENKO M, BROX T. Semi-supervised semantic segmentation with high- and low-level consistency[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(4): 1369 – 1379.

[18] OLSSON V, TRANHEDEN W, PINTO J, et al. ClassMix: segmentation-based data augmentation for semi-supervised learning[C]//Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), 2021: 1368 – 1377.

[19] SOHN K, BERTHELOT D, LI C L, et al. FixMatch: simplifying semi-supervised learning with consistency and confidence[C]//Proceedings of 34th Conference on Neural Information Processing Systems, 2020.