软件学报 ISSN 1000-9825, CODEN RUXUEW
[doi: 10.13328/j.cnki.jos.007412] [CSTR: 32375.14.jos.007412]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

# 基于差异化特征提取的交叉半监督语义分割网络*

陈亚当[1], 李家戚[1], 车 洵[2], 吴恩华[3,4]

[1](南京信息工程大学 计算机学院, 江苏 南京 210044)

[2](南京理工大学 计算机科学与工程学院, 江苏 南京 210094)

[3](基础软件与系统重点实验室 (中国科学院 软件研究所), 北京 100190)

[4](计算机科学国家重点实验室 (中国科学院 软件研究所), 北京 100190)

通信作者: 车洵, E-mail: chexun@njust.edu.cn

**摘 要**: 半监督语义分割方法通常采用不同数据增强方案来确保多分支网络输入信息的差异化, 以实现分支之间相互监督. 虽然该方法取了一定成效, 但其存在以下问题: 1) 特征提取差异不足, 造成推理特征信息同化; 2) 监督信号差异不足, 造成末端损失学习同化. 以上两个问题都会促使网络中不同分支收敛到相似的解决方案, 导致多分支网络功能退化, 出现多个分支对错误保持相似置信度的问题, 错误引导网络分支收敛. 针对上述问题, 提出了一种基于差异化特征提取的交叉半监督语义分割网络. 首先, 采用差异化特征提取策略, 通过让网络分支分别关注纹理、语义和形状等不同信息, 从特征提取角度使特征提取信息始终存在差异性, 减少网络对数据增强的依赖; 其次, 提出一种交叉融合伪标签方法, 使网络分支交替生成邻域像素融合伪标签, 以此增强网络末端监督信号的差异性, 最终促使网络分支收敛向不同的解决方案. 实验结果证明, 方法在 Pascal VOC 2012 和 Cityscapes 验证集上分别达到了 80.2% 和 76.8% 的优异性能, 领先于最新方法 0.3% 和 1.3%.

**关键词**: 计算机视觉; 语义分割; 半监督学习; 协同训练; 伪标签

**中图法分类号**: TP391

中文引用格式: 陈亚当, 李家戚, 车洵, 吴恩华. 基于差异化特征提取的交叉半监督语义分割网络. 软件学报. http://www.jos.org.cn/1000-9825/7412.htm

英文引用格式: Chen YD, Li JQ, Che X, Wu EH. Cross Semi-supervised Semantic Segmentation Network Based on Differential Feature Extraction. Ruan Jian Xue Bao/Journal of Software (in Chinese). http://www.jos.org.cn/1000-9825/7412.htm

## Cross Semi-supervised Semantic Segmentation Network Based on Differential Feature Extraction

CHEN Ya-Dang[1], LI Jia-Qi[1], CHE Xun[2], WU En-Hua[3,4]

[1](School of Computer Science, Nanjing University of Information Science and Technology, Nanjing 210044, China)

[2](School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China)

[3](State Laboratory of System Software (Institute of Software, Chinese Academy of Sciences), Beijing 100190, China)

[4](State Key Laboratory of Computer Science (Institute of Software, Chinese Academy of Sciences), Beijing 100190, China)

**Abstract**: Semi-supervised semantic segmentation methods typically employ various data augmentation schemes to ensure differentiation in the input of network branches, enabling mutual self-supervision. While successful, this approach faces several issues: 1) insufficient diversity in feature extraction leads to feature signal assimilation during inference; 2) inadequate diversity in supervision signals results in the assimilation of loss learning. These issues cause network branches to converge on similar solutions, degrading the functionality of multi-branch networks. To address these issues, a cross semi-supervised semantic segmentation method based on differential feature extraction is proposed. First, a differential feature extraction strategy is employed, ensuring that branches focus on distinct information, such as texture, semantics, and shapes, thus reducing reliance on data augmentation. Second, a cross-fusion pseudo-labeling method is

* 基金项目: 国家自然科学基金 (62473201, 62477026, 62332015, 62072449); 无锡市产业创新研究院先导技术预研项目
收稿时间: 2024-10-23; 修改时间: 2024-12-09, 2025-01-11; 采用时间: 2025-02-11; jos 在线出版时间: 2025-08-27

introduced, where branches alternately generate neighboring pixel fusion pseudo-labels, enhancing the diversity of supervision signals and guiding branches toward different solutions. Experimental results demonstrate this method achieves excellent performance on the Pascal VOC 2012 and Cityscapes validation datasets, with scores of 80.2% and 76.8%, outperforming the latest methods by 0.3% and 1.3%, respectively.

**Key words**: computer vision; semantic segmentation; semi-supervised learning; co-training; pseudo-label

基于深度神经网络的语义分割任务已取得显著成功, 但这在很大程度上依赖于大量标注数据集[1−4]. 由于语义分割需要精确到像素级的标注, 标注人员必须手动标记每张图像中的数十万个像素. 这使得收集完全准确的标注数据用于训练深度神经网络的成本极为高昂[5−7].

为减少模型对标注数据的依赖, 半监督语义分割提出利用大量未标注样本来增强网络在少量标注样本上的学习能力, 提升模型的泛化性和通用性[8−12]. 这一方法在标注成本高、数据量庞大且需要精细像素级标注的领域, 如自动驾驶和医疗影像分析中, 具有广阔的应用前景[13,14].

显然, 半监督语义分割中有标签数据的数量远远少于无标签数据, 因此可用的标注信息非常有限. 如何充分利用无标签数据来辅助有标签数据进行模型训练, 成为了一个关键问题. 一种直观的解决方案是为无标签数据生成标签, 即通过模型推理生成伪标签, 以监督无标签数据的训练, 进而增加数据量[15−17]. 然而, 在单分支网络中, 伪标签是由模型自身的预测生成的, 如果伪标签出现错误, 模型会倾向于重复错误的预测, 导致错误信息在网络中不断积累, 进而引发确认偏差[18].

为了解决单分支网络在处理无标签数据时的准确性问题, 半监督语义分割引入了多分支网络架构. 通过多个独立或部分共享权重的分支, 学习不同的特征表达, 以此增强网络对无标签数据的利用[19−21]. 比如, 图 1 所示传统半监督语义分割方法通过不同的数据增强方案进行训练, 具体来说, 将经过不同程度增强的图像输入到网络分支中进行推理和分割. 其中, 弱数据增强仅包括图像的裁剪和随机翻转, 而强数据增强则在此基础上增加了色彩变换和灰度化等高级操作, 以进一步丰富训练数据的多样性[22−25].

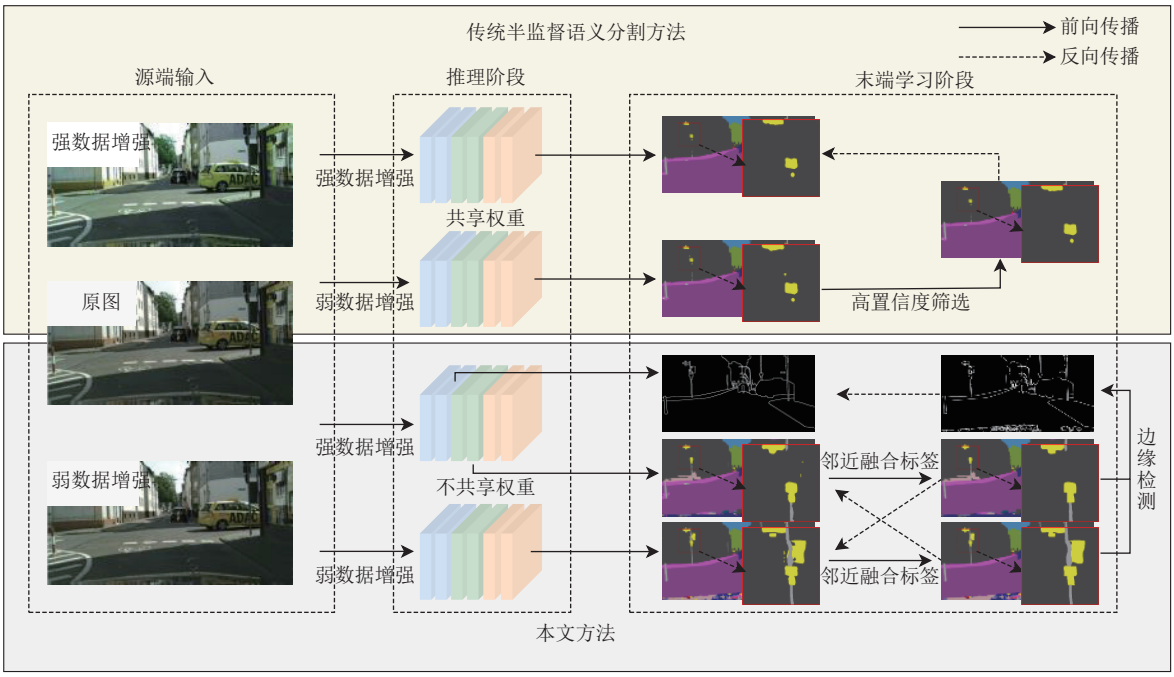



图 1 现有的半监督语义分割方法与本文框架对比

神经网络可分为源端输入、推理阶段和末端学习阶段. 现有的强弱对比数据增强方案仅在源端输入阶段引入差异, 期望通过这种方式在推理阶段促使各分支收敛到不同的解决方案, 以应对同化问题. 然而, 在推理阶段和末

端损失学习阶段未能有效考虑差异化. 相同的分支推理架构和末端监督信号都会促进各分支的学习路径趋于一致, 导致分支网络同化.

针对上述阶段差异化不足导致的网络同化问题[26−29], 本文结合强弱对比增强方案创造源端差异的基础上, 进一步引入差异化特征提取策略和交叉融合伪标签方法, 从推理阶段和末端损失学习阶段有效解决了网络分支同化的问题, 由此提出基于差异化特征提取的交叉半监督语义分割网络, 如图 1(b) 所示.

本文采用差异化特征提取策略, 使各分支分别提取细节纹理、语义上下文和边缘形状信息. 细节纹理集中于网络的低层特征, 而语义上下文则位于高层特征, 从而在网络的深度分布上保证了分支差异化. 细节纹理和语义上下文分支执行语义分割任务, 而边缘形状分支生成二值图, 仅划分边缘和非边缘区域, 任务的不同使边缘形状分支信息自然与其他分支产生差异. 综上, 通过在网络层次与任务导向上的差异化设计, 确保了分支信号多样性, 有效避免了推理阶段信息差异不足的问题.

本文设计了一种交叉融合伪标签方法, 通过邻域像素的置信度优化伪标签生成. 具体而言, 各网络分支融合邻域像素的置信度, 并生成用于其他分支的监督信号. 在网络末端的监督过程中, 不同分支接受不同的监督信号, 分支之间的学习路径被分化, 避免了所有分支共享相同的监督信号而导致的错误同化积累. 同时, 邻域像素的置信度融合进一步提高了伪标签的质量, 减少了高置信度错误预测对其他分支的负面影响. 这种方法从根本上打破了传统伪标签方案中各分支在监督信号上的一致性, 从而解决了末端损失学习同化的问题.

本文在多个半监督语义分割基准数据集上验证了所提出方法的性能. 相比于最新方法, 本文方法在 Classic Pascal VOC 2012 (732 张标签训练图像)[30]上拥有了 0.3% 的 mIoU 提升, 在 Cityscapes (186 张标签训练图像)[31]上也有了 1.3% 的 mIoU 提升. 上述实验均在相同骨干网络下进行, 结果充分证明了所提出方法的合理性和有效性.

## 1 相关工作

为减少对大规模全标签数据集的依赖, 半监督学习方法提出了一种利用少量标注数据和大量未标注数据相结合的方式来训练网络模型[16,32−34].

一种直观的解决方案是为无标签数据生成相应的监督信号, 即伪标签方法[15−17]. 例如在 PseudoSeg[17]中, 无标签数据的监督信号来源于模型自身预测的高置信度结果. 然而高置信度伪标签出现错误时, 会干扰网络的收敛过程, 进而导致确认偏差问题.

因此, 半监督方法提出了一种基于多分支网络架构的策略, 即使用弱数据增强处理的输入预测生成伪标签, 然后将这些伪标签作为监督信号, 用于训练经过强数据增强处理的输入预测[35−38]. 例如, 在半监督分类领域中, FixMatch[35]通过强弱对比的数据增强策略对网络进行训练. 尽管这种方法避免了模型自身生成监督信号的局限, 但由于两个网络分支共享权重, 导致它们的收敛方向趋于一致, 未能彻底消除确认偏差问题. 此外, 该方法依赖于手动设计的数据增强策略, 难以找到最优的增强方案.

为了避免网络同化导致的确认偏差, 需要确保不同分支能够提取差异化且互补的特征信息. 因此, 部分半监督语义分割研究从网络源端输入入手, 通过设计使各分支在特征提取过程中具备足够的差异性, 确保当某个分支出现错误预测时, 其他分支能够进行纠正. 例如, CCT (cross-consistency training)[20]方法引入了特征扰动, 通过对不同网络分支编码后的特征进行 Dropout 扰动, 保证解码器输入的信息存在差异性. UniMatch[27]结合了 FixMatch[35]和 CCT[20]的思想, 通过多个图像增强分支和特征扰动来增加分支信息的多样性. 然而, 这些方法依赖于人工设计的扰动方案, 难以确保不同分支能够提取出足够差异化的特征, 也难以找到最优的方案.

因此, 部分研究从推理架构层面入手, 尝试更根本地解决这一问题. 例如 CPS (cross pseudo supervision)[19]通过对不同的网络分支进行初始化, 期望它们收敛到不同的解决方案, 然而该方法依赖于多分支网络的数量, 不仅增加了模型的参数量, 还导致性能提升与资源消耗不成正比. 此外, 相同的监督信号可能会引导各分支收敛至相似的解决方案. CCVC (conflict-based cross-view consistency)[39]引入特征差异损失, 强制不同分支提取差异化的特征, 以避免网络同化的问题, 但其特征差异损失依赖线性映射, 难以确保特征的充分差异化. PCR (prototype-based

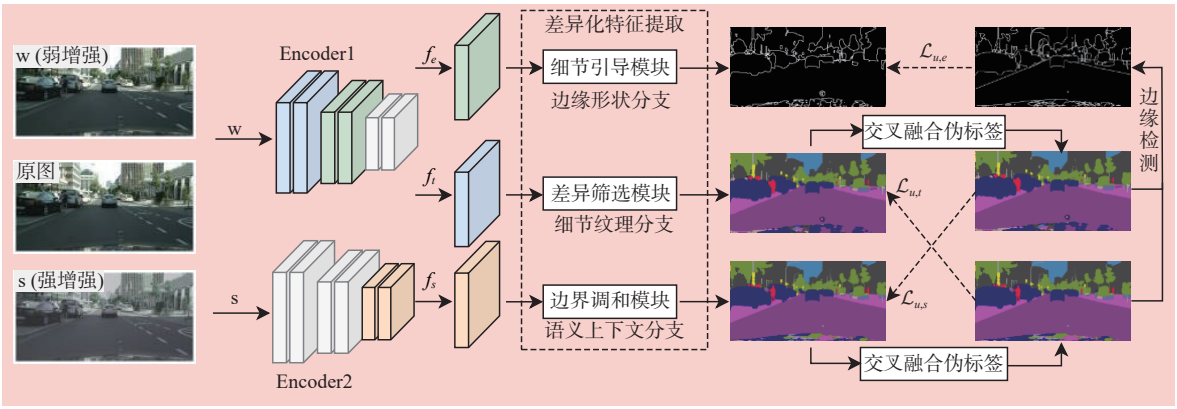
consistency regularization)[40]在网络中引入了原型网络, 通过探索强扰动分支的多样性提升性能, 但受限于原型网络的聚类机制, 其在细粒度分割任务中的效果不佳.

另有部分方法从网络末端入手, 例如 N-CPS[19], 通过不同分支间的伪标签相互监督, 确保各分支在学习过程中保持足够的差异性, 并通过一致性正则化引导它们朝正确的方向收敛. 然而, 伪标签本身可能存在错误, 反而可能误导其他分支学习错误信息, 影响整体模型性能. U2PL (using unreliable pseudo-labels)[41]通过将低置信度的预测用于负样本生成, 以此平衡类别训练. 但该方案仍未解决高置信度错误预测导致的确认偏差问题, 限制了模型的进一步提升. CPSR (class probability space regularization)[42]旨在综合利用所有未标记数据的有效性息, 将不确定像素的信息进行处理并转化为有效信息, 但是依然只从一个阶段解决同化问题. IPixMatch[43]同样从网络末端的损失函数入手, 引入相关性一致性损失来结合像素间的上下文信息. VC3 (view-coherent correlation consistency)[44]综合考虑了网络源端和末端, 引入了新的数据增强策略和损失计算, 但是缺乏对于网络推理阶段的差异化处理.

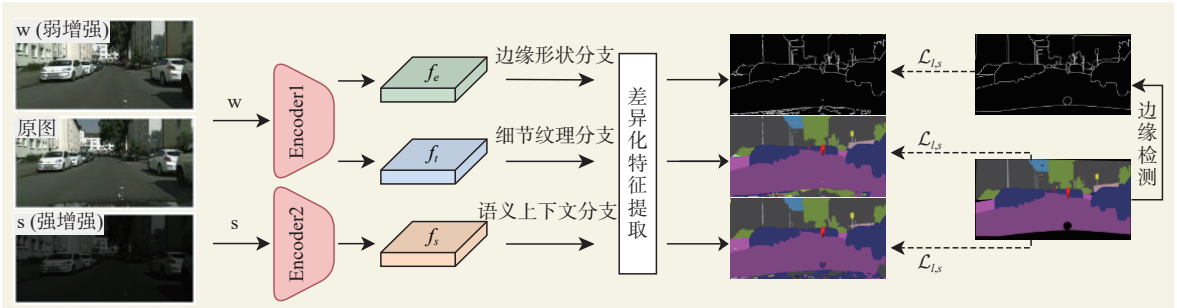
与上述方法相比, 本文通过在神经网络的不同阶段均引入差异化, 在源端增强策略差异性的基础上, 进一步引入了推理阶段的特征提取差异性和末端损失学习阶段的监督信号差异性, 全阶段差异性最大程度上保证了分支收敛的异向性. 通过上述方法, 本文有效地解决了多分支网络中功能退化的问题, 确保了各分支在特征提取和推理过程中的差异化.

## 2 基于差异化特征提取的交叉半监督语义分割网络

本节详细介绍本文所提出的基于差异化特征提取的交叉半监督语义分割网络. 如图 2 所示, 每次迭代均进行一次标签图像训练和无标签图像训练.




(a) 无标签图像训练过程




(b) 标签图像训练过程

图 2 本文方法总体训练流程

多分支网络通过两个不共享权重的 ResNet 网络提取基础特征映射, 在标签训练过程中细节纹理分支 $f_t$ 与语义上下文分支 $f_s$ 的预测由真实标签直接监督训练, 边缘形状分支 $f_e$ 则由真实标签生成的边缘检测图进行监督, 无标签训练过程与标签训练不同的是, 细节纹理分支 $f_t$ 与语义上下文分支 $f_s$ 融合生成对方的监督信号进行训练, 边缘形状分支 $f_e$ 则由 $f_t$ 和 $f_s$ 融合生成的边缘检测结果作为监督信号. 图中灰色特征层表示与后续特征提取无关的内容. 本文方法框架包括差异化特征提取 (第 2.1 节) 和交叉融合伪标签(第 2.2 节) 两个部分. 在 2.3 节将结合算法 1 详细介绍本文方法的总体训练迭代流程和损失计算过程.

### 2.1 差异化特征提取

本文采用多分支网络架构、差异化特征提取分别聚焦于细节纹理信息、语义上下文信息和边缘形状信息, 从不同的特征信息角度优化分割任务. 这 3 个差异化特征提取分支分别标记为细节纹理分支 $f_t$、语义上下文分支 $f_s$、边缘形状分支 $f_e$.

细节纹理分支和语义上下文分支提取的特征信息具有差异性, 这是因为在神经网络信息分布中, 细节纹理信息和语义上下文信息分别集中在神经网络的浅层和深层特征中. 这种特征信息的层次分布引导了细节纹理分支 $f_t$ 和语义上下文分支 $f_s$ 进行不同的特征提取, 确保了它们在提取过程中各自关注的信息得以充分表达.

尽管细节纹理分支 $f_t$ 和语义上下文分支 $f_s$ 分别提取了不同的特征信息, 但 $f_t$ 和 $f_s$ 各自应用在语义分割任务中效果并不理想[45]. 细节纹理分支 $f_t$ 虽然提取了大量细节和空间位置信息, 但缺乏高层次语义抽象能力, 难以准确分割, 易产生过拟合现象. 为解决这一问题, 本文提出了差异筛选模块 (第 2.1.1 节), 有选择性地从语义上下文分支 $f_s$ 中获取部分高级抽象信息, 既融合了高级语义信息, 又避免了全盘接收语义信息导致的分支同化.

语义上下文分支 $f_s$ 提取的深层特征信息虽然具备高级别的抽象和语义信息, 但丢失了过多的细节, 导致分割边缘不够精细. 为解决这一缺陷, 本文提出了边界调和模块 (第 2.1.2 节), 在边界区域使用细节纹理信息进行填充. 这一策略不仅改善了高倍率下采样的深层特征提取导致的边界分割效果不佳的问题, 还保留了主体区域的高级语义信息, 确保与细节纹理分支 $f_t$ 提取的特征信息保持差异化.

在边界调和模块中, 本文提出在边界区域上使用纹理信息对语义上下文的特征信息进行填充, 因此在多分支网络中引入边缘形状分支. 通过优化任务引导, 使得监督信号仅包含图像的边缘区域, 不包含细节等其他分割信息[46]. 这样, 边缘形状分支可以专注于提取边界信息, 忽略其他细节, 确保边界信息的准确性, 同时保持与其他分支特征的差异性.

神经网络在处理边界信息时, 卷积操作的平滑效应会导致边界信息不够精确, 特别是在细微和复杂的边界区域. 为了解决这一问题, 本文提出了细节引导模块 (第 2.1.3 节), 该模块在边界空间注意力的引导下, 有选择地整合细节纹理分支的信息, 生成更具代表性的边界特征, 从而显著提升边界特征提取的准确性和鲁棒性. 该模块通过关注局部细节, 使得边界特征在处理复杂背景时也能保持较高的辨识度.

#### 2.1.1 差异筛选模块

神经网络的浅层特征保留了更多的细节信息, 因此, 细节纹理分支 $f_t$ 从 1/4 下采样的特征图中提取和解码细节信息. 然而, 由于浅层特征缺乏高级语义表示, 直接使用低级特征进行解码和分类容易导致过拟合现象.

为解决这一问题, 如图 3 所示, 本方法将语义上下文分支 $f_s$ 作为细节纹理分支 $f_t$ 的语义信息补充, 利用基于余弦相似度的融合算法, 使得细节纹理分支有选择性地获取语义信息, 既不会因为过于关注特定局部信息发生过拟合, 也不会与语义上下文分支过于相似导致网络同化. 将特征映射中细节纹理分支和语义上下文分支对应像素的向量分别定义为 $\overrightarrow{v_t}$ 和 $\overrightarrow{v_s}$, 那么该算法用公式可以表述为:

$$Out = \sigma \cdot \overrightarrow{v_t} + (1-\sigma) \cdot \overrightarrow{v_s} \tag{1}$$

其中, $\sigma = \dfrac{0.5(\overrightarrow{v_t} \cdot \overrightarrow{v_s} + 1)}{|\overrightarrow{v_t}||\overrightarrow{v_s}|} \in [0,1]$ 表示这两个向量的相似度, 如果 $\sigma$ 接近 1, 则更加信赖 $\overrightarrow{v_s}$, 因为语义上下文分支提供了更准确丰富的语义信息; 反之, 则更加信赖具有丰富解析能力的纹理分支向量 $\overrightarrow{v_t}$, 以保留更多细节信息.

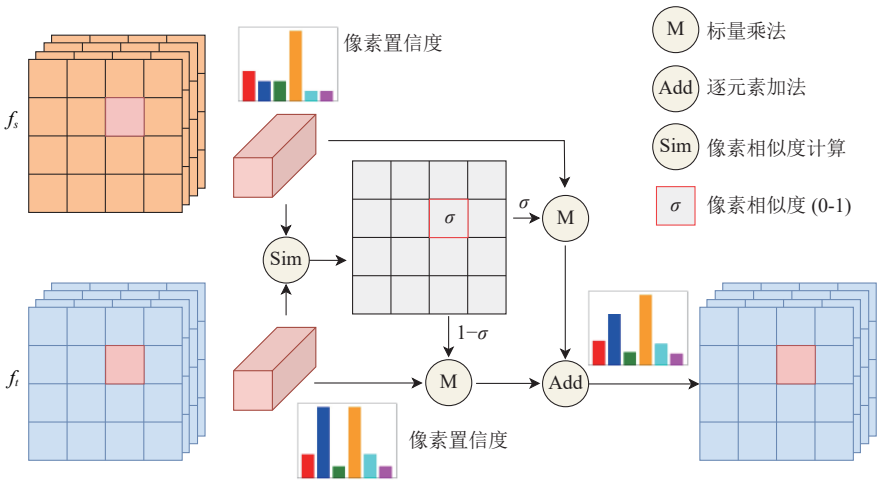



图 3　差异筛选模块示意图

2.1.2　边界调和模块

深层网络中承载着丰富的高级语义信息, 因此, 本文中的语义上下文分支从 1/32 下采样的特征图中提取和解码图像信息. 然而, 过高的下采样倍率使得语义上下文分支缺乏了细节纹理和边界形状信息, 这影响了其对图像的准确理解. 特别是在物体的边界区域, 分割效果受到了较大的影响, 表现出不足之处.

相反, 细节纹理分支以及边缘形状分支则分别保留了空间细节信息和边界形状信息. 因此, 本文提出了借用边界预测调和语义信息和细节纹理信息的合成, 如图 4 所示, 在边缘形状预测 $\tilde{y}_e$ 引导下使用浅层纹理细节信息 $f_t$ 填补语义分支 $f_s$ 的边界区域信息, 在边界区域上填补更多细节纹理信息, 在其他主体区域上更加信任语义上下文分支. 这一机制不仅丰富了语义分支的特征信息, 还避免了语义分支过度融合细节信息导致的过拟合现象. 该调和模块可以表示为:

$$Out_{\text{merge}} = f_{\text{out}}((1-v)\overrightarrow{v_s} + v\cdot\overrightarrow{v_t}) \tag{2}$$

其中, $\overrightarrow{v_t}$、$\overrightarrow{v_s}$ 分别表示细节纹理分支和上下文分支的对应像素特征信息, 边缘预测对应像素表示为 $v\in[0, 1]$ (越接近 1 表示越可能是边缘类别), $f_{\text{out}}$ 表示卷积、归一化和 ReLU 的结合.

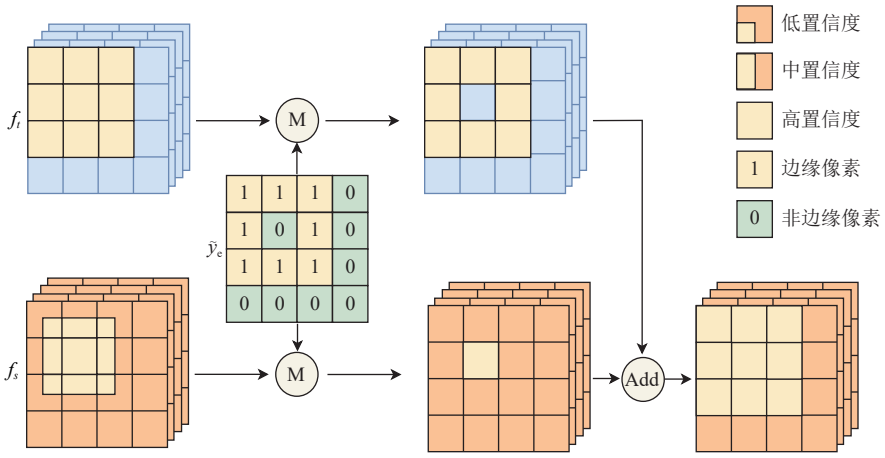



图 4　边界调和模块

2.1.3 细节引导模块

在边界调和模块中, 边界信息被用于调和语义信息和细节纹理信息和合成. 因此本文引入了边缘形状分支来提取自然图像的边界信息. 神经网络的中低层拥有丰富的边缘信息, 因此边缘形状分支提取来自 1/8 下采样的特征图, 通过端到端的学习从图像数据中学习到最优的边缘特征. 本文设计了通过对真实标签和伪标签进行 Canny 边缘检测得到边缘标签作为监督信号, 促使边缘形状分支仅提取物体的边缘信息, 而不对物体内部的其他细节边缘信息关注.

由于神经网络中卷积操作的平滑效应可能导致边界信息的精度下降, 而细节纹理信息通常包含更为准确的边界信息. 因此, 本文引入了边缘信息的空间注意力机制[47], 以将细节纹理信息 $\vec{v}_t$ 逐像素融合到边缘形状信息 $\vec{v}_e$ 中, 从而生成更具有空间区域代表性的边缘特征 (如图 5 所示). 这种融合方式可以用以下方式表示:

$$\overrightarrow{v_{\text{out}}} = \text{Sigmoid}(Att)(\vec{v}_e + \vec{v}_t) \tag{3}$$

其中, $\vec{v}_e$ 表示边缘形状分支向量, $\vec{v}_t$ 表示细节纹理分支向量, $Att$ 表示空间注意力图对应像素权重, Sigmoid 函数保证了注意力权重在 0–1 之间.

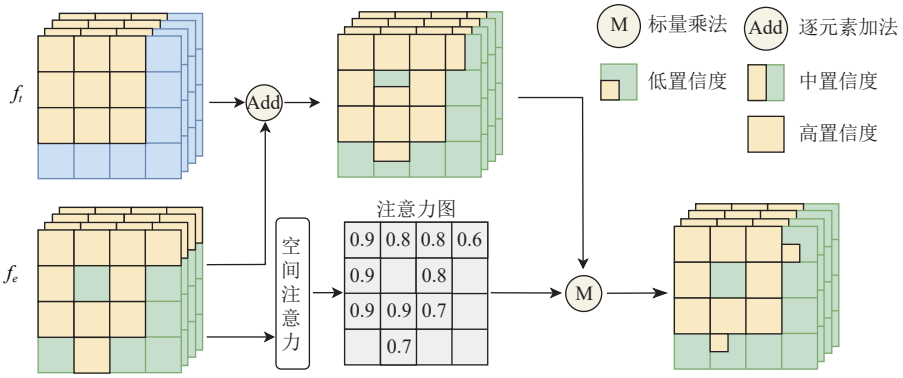



图 5 细节引导模块

## 2.2 交叉融合伪标签

现实图像往往具有很强的空间相关性, 相似的语义类别通常会在空间上聚集, 图像中的邻近像素之间的语义类别一般也是相似的. 这意味着邻近像素表征之间存在较高的正相关性, 且这种正相关性会随着像素之间的距离变近而增强.

借用自然图像的这种性质, 通过空间上聚集上邻近像素来改善伪标签的生成. 定义给定像素向量为 $\vec{v}_{i,j}$, 类别预测为 $p_c(\vec{v}_{i,j})$, 邻域像素向量及类别预测被表示为 $\vec{v}_{k,l}$ 和 $p_c(\vec{v}_{k,l})$, 计算两个类别同属于一个类别 $c$ 的联合修正概率为:

$$\tilde{p}_c(\vec{v}_{i,j} \cup \vec{v}_{k,l}) = p_c(\vec{v}_{i,j}) + p_c(\vec{v}_{k,l}) - p_c(\vec{v}_{i,j} \cap \vec{v}_{k,l}) \tag{4}$$

在像素 $\vec{v}_{i,j}$ 和 $\vec{v}_{k,l}$ 相互独立时, 联合概率 $p_c(\vec{v}_{i,j} \cap \vec{v}_{k,l})$ 可以被表述为公式 (5), 但是图像具有强相关性, 且这种相关性随着像素的距离变近而增强, 因此简单的假定所有像素独立并不能有理想效果.

$$p_c(\vec{v}_{i,j} \cap \vec{v}_{k,l}) = p_c(\vec{v}_{i,j}) \cdot p_c(\vec{v}_{k,l}) \tag{5}$$

$$p_c(\vec{v}_{i,j} \cap \vec{v}_{k,l}) = p_c(\vec{v}_{i,j}) \cdot p_c(\vec{v}_{k,l} | \vec{v}_{i,j}) \tag{6}$$

在公式 (6) 中, 本文计算了在像素存在正相关时的真实联合概率, 由于邻近像素表现出正相关, $\vec{v}_{i,j}$ 属于类别 $c$ 会增加 $\vec{v}_{k,l}$ 也属于类别 $c$ 的概率, 这时 $p_c(\vec{v}_{k,l} | \vec{v}_{i,j}) > p_c(\vec{v}_{k,l})$, 因此真实修正联合概率公式 (4) 存在一个修正概率上限:

$$p_c(\vec{v}_{i,j} \cap \vec{v}_{k,l}) \geqslant p_c(\vec{v}_{i,j}) \cdot p_c(\vec{v}_{k,l}) \tag{7}$$

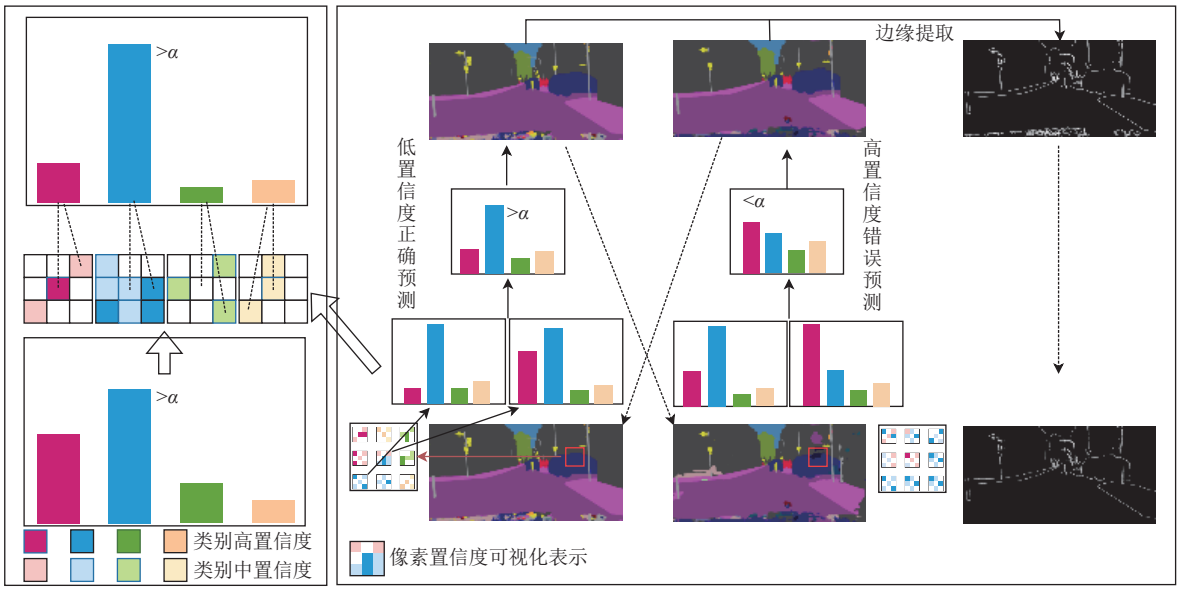
$$\tilde{p}_c(\vec{v}_{i,j} \cup \vec{v}_{k,l}) \leqslant p_c(\vec{v}_{i,j}) + p_c(\vec{v}_{k,l}) - p_c(\vec{v}_{i,j} \cap \vec{v}_{k,l}) \tag{8}$$

虽然得到了联合修改概率上限, 但是邻近像素的正相关性随着距离变近而增强, 简单的假设所有邻近像素相互独立, 距离越近计算出的联合修正概率偏差越大. 因此, 在假定像素相互独立时, 本文要求邻近像素与给定像素的距离尽可能相近, 比如在一个 3×3 的邻域中, 所有像素到中心像素的距离是相同的, 这样可以保证相关性的一致性, 同时所有邻域像素联合修正概率偏差较为一致, 在筛选最大熵的类别联合修正概率较为准确.

具体而言, 如图 6(a) 所示, 考虑每个像素的周围 3×3 像素邻域内的预测来改进每个像素的伪标签, 通过计算每个类别的邻近像素最大熵作为类别预测, 选择熵最大的类别作为伪标签类别. 给定像素向量 $\vec{v}_{i,j}$ 及其对应的每个类别预测 $p_c(\vec{v}_{i,j})$, 考虑每个像素周围 3×3 像素的邻域像素 $\vec{v}_{k,l}$, 计算两个的像素中至少有一个属于类别 $c$ 的联合概率修正:

$$\tilde{p}_c(\vec{v}_{i,j} \cup \vec{v}_{k,l}) \leqslant p_c(\vec{v}_{i,j}) + p_c(\vec{v}_{k,l}) - p_c(\vec{v}_{i,j}, \vec{v}_{k,l}) \tag{9}$$

其中, $p_c(\vec{v}_{i,j}, \vec{v}_{k,l}) = p_c(\vec{v}_{i,j}) \cdot p_c(\vec{v}_{k,l})$ 为像素相互独立的联合概率.




(a) 邻近像素融合标签 (b) 交叉伪标签

图 6 交叉融合伪标签

最大增益的邻近预测能够提供周围像素的主要特征信息, 确保置信度更高的信息优先被整合, 因此类别 $c$ 的联合修正概率需要筛选出具有最大信息熵的邻域修正概率:

$$\tilde{p}_c^M(\vec{v}_{i,j}) = \max_{k,l}(\tilde{p}_c(\vec{v}_{i,j} \cup \vec{v}_{k,l})) \tag{10}$$

通过计算所有类别的联合修正概率, 可以得到每个类别对应的最大信息熵. 本文选择具有最大信息熵的类别作为该像素的伪标签. 虽然在物体边缘区域, 像素容易受到物体外部类别信息的干扰, 从而导致错误类别的信息熵增大, 但由于空间相关性, 正确类别的邻域像素会产生更大的信息熵, 这有效避免了其他类别的干扰. 此外, 这种方法还解决了低置信度下正确预测无法被传播的问题.

尽管邻近像素融合优化了分支生成的伪标签, 但这些伪标签仍可能存在某些错误预测, 进而引入不正确的监督信号, 误导模型训练. 因此本文采用交叉伪标签监督训练无标签图像, 如图 6(b) 所示. 具体而言, 细节纹理分支 $f_t$ 和上下文分支 $f_s$ 生成的伪标签被用作对方的监督信号, 以避免分支生成自己的监督信号, 从而无法检测自身的

错误预测产生的确认偏差.

考虑到边缘形状分支 $f_e$ 需要特殊的伪标签, 本文采用对两个 $f_t$ 和 $f_s$ 生成的伪标签像素置信度更高的像素合成边缘伪标签的前置伪标签, 即:

$$\vec{v}_{e,(i,j)}) = \max(\vec{v}_{t,(i,j)}, \vec{v}_{s,(i,j)}) \tag{11}$$

通过对前置伪标签应用边缘检测算法[46]提取边缘信息, 从而获得最终的边缘伪标签, 表示为:

$$\hat{y}_e^u = \mathbb{E}\sum_{j=0}^{W}\sum_{i=0}^{H}(\vec{v}_{e,(i,j)}) \tag{12}$$

其中, $\mathbb{E}$ 表示边缘伪标签检测算法.

### 2.3 总体损失与算法描述

经过解码的分支特征预测记为 $\tilde{y}_i^\alpha$, 其中 $\alpha \in \{l,u\}$ 标签图像和无标签图像, $i \in \{t,s,e\}$ 表示细节纹理、语义上下文和边缘形状分支.

如图 2 所示, 在无标签图像训练中, 纹理分支和语义上下文分支之间交叉监督, 因此细节纹理分支预测 $\tilde{y}_t^u$ 由语义分支预测生成的伪标签 $\hat{y}_s^u$ 进行监督. 因此细节纹理分支的交叉熵损失被表述为:

$$\mathcal{L}_{u,t} = \frac{1}{N}\sum\nolimits_{m=1}^{N}\frac{1}{W\times H}\sum\nolimits_{n=0}^{W\times H}\ell_{\text{ce}}(\tilde{y}_{mn,t}^u, \hat{y}_{mn,s}^u) \tag{13}$$

其中, $\tilde{y}_{mn,t}^u$ 表示细节纹理分支对第 $m$ 张图像的第 $n$ 个像素的预测值, 而 $\hat{y}_{mn,s}^u$ 表示语义分支生成的交叉融合伪标签的置信度, 同样针对第 $m$ 张图像的第 $n$ 个像素.

同样的, 语义上下文分支的损失被表示为 $\mathcal{L}_{u,s}$, 边缘形状分支预测 $\tilde{y}_e^u$ 受到其他分支边缘提取的伪标签 $\hat{y}_e^u$ 的监督, 相应的损失表述为 $\mathcal{L}_{u,e}$. 因此, 无监督损失可以被表述为:

$$\mathcal{L}_u = \lambda_t\mathcal{L}_{u,t} + \lambda_s\mathcal{L}_{u,s} + \lambda_e\mathcal{L}_{u,e} \tag{14}$$

根据实验效果变化, 分别将 $\lambda_t$、$\lambda_s$、$\lambda_e$ 设置为 0.6, 1.8, 10.

如图 2 所示, 标签图像与无标签图像在损失计算的差别仅在与标签图像使用真实标签 $y_i^l$ 作为监督信号训练网络. 因此监督损失记为:

$$\mathcal{L}_l = \lambda_t\mathcal{L}_{l,t} + \lambda_s\mathcal{L}_{l,s} + \lambda_e\mathcal{L}_{l,e} \tag{15}$$

根据实验效果变化, 分别将 $\lambda_t$、$\lambda_s$、$\lambda_e$ 设置为 0.6, 1.8, 20.

综上, 训练阶段损失被分为监督损失 $\mathcal{L}_l$ 和无监督损失 $\mathcal{L}_u$, 总损失计算如下所示:

$$\mathcal{L} = \lambda_u\mathcal{L}_u + \lambda_l\mathcal{L}_l \tag{16}$$

其中, $\lambda_u$、$\lambda_l$ 是无监督损失和监督损失的平衡参数.

总体训练流程如算法 1 所示, 每次迭代均会对标签图像和无标签图像进行训练, 无标签图像训练并不额外生成伪标签加入标签数据中进行训练, 仅使用其他分支预测生成的融合伪标签作为每一次迭代的监督信号. 本文计算监督损失和无监督损失作为总训练损失来更新模型参数.

**算法 1**. 半监督语义分割总体训练算法.

输入: 标签图像数据集 $(X_l, \mathrm{Y}_l) \in D_l$, 无标签图像数据集 $X_u \in D_u$, 最大训练轮次 *max_epochs*;
输出: 训练好的语义分割模型.

1. 初始化语义分割模型参数
2. *epochs* ← 0
3. **WHILE** *epochs* < *max_epochs* **DO**
4. **FOR** ($X_l$, $Y_l$, $X_u$) in *zip*($D_l$, $D_u$) **DO**
5. 使用 ($X_l$, $Y_l$) 训练语义分割模型, 计算监督损失 $\mathcal{L}_l$
6. 使用 $X_u$ 训练语义分割模型, 计算无监督损失 $\mathcal{L}_u$

7. 计算总损失 $\mathcal{L}=\lambda_u\mathcal{L}_u+\lambda_l\mathcal{L}_l$
8. 更新语义分割模型参数以最小化损失 $\mathcal{L}$
9. **END FOR**
10. $epochs \leftarrow epochs+1$
11. **END WHILE**

## 3 实验结果及分析

### 3.1 数据集

Pascal VOC 2012 数据集[30]是由来自 21 个类的超过 13 000 张图像组成的半监督语义分割 (SSS) 基准数据集. 它包含 1 464 张用于训练的全注释图像, 1 449 张用于验证的图像和 1 456 张用于测试的图像. 之后又采用 Blender Pascal VOC 2012[48]的渲染标记图像, 并将标记数据的数量扩展到 10 582 个. 渲染的标签图像质量较低, 其中一些伴有噪声.

Cityscapes 是 SSS 的难度较大的来自 50 个不同城市的 30 个类别的一个基准数据集[31], 它专注于城市场景, 由来自 19 个类的 2 975 张带注释的训练图像, 500 张验证图像和 1 525 张测试图像组成.

### 3.2 实验细节

为了验证方法有效性的公平性, 本研究遵循先前方法所采用的 ResNet 骨干网络方案. 表 1 详细列出了主要网络结构及关键超参数, 包括卷积核大小、步长、激活函数等, 以便于复现. 纹理分支 Layer3 输入来自边缘分支 Layer1 输出, 网络结构*2 指的是结构循环 2 次, 如 Conv*2 为 Conv(3, 64, 3, 2, 1)+Conv(64, 64, 3, 1, 1). 本文使用 SGD 优化器对 Pascal VOC 2012 数据集和 Cityscapes 数据集上分别设置初始学习率为 0.001 和 0.005 进行实验. Pascal 和 Cityscapes 的 *Epochs*, *CropSize* 和 *BatchSize* 分别设置为 [80, 512, 24], [250, 712, 8]. 在每个 batch 中标记数据和无标记数据的数量相等, 使用 mIoU 作为语义分割评估指标. 本文在超参数 $\lambda_u$、$\lambda_l$ 的设定上沿用以往研究方法[28,39]的数值, 在 Pascal VOC 2012 数据集上设置为 5.0 和 2.0, 在 Cityscapes 数据集上设置为 1.0 和 1.0. 针对置信区间细化的引入, 本文依据实验效果将伪标签阈值设置为 $\alpha_0=0.4$, 并随着训练进度调整 $\alpha=\alpha_0(1+epochs/epochs)$. 为了进一步提升模型的鲁棒性, 我们在训练过程中使用了 CutMix 数据增强, 并优化了学习率、损失函数等超参数设置, 具体 Cityscape 数据集超参数设置见表 2, 由于部分超参数的最佳设置只能通过实验验证其性能优化效果, 无法直接验证本文方法的有效性, 因此本文仅给出了推荐的参数设置, 而未提供不同配置下的消融实验结果.

表 1 网络结构及关键超参数

| 提取分支 | 网络层次 | 网络结构 | 输入通道 | 输出通道 | 卷积核 | 步长 | 填充 |
|---|---|---|---|---|---|---|---|
| 语义分支 | Conv1 | Conv*2 | 3 | 64 | 3 | 2 | 1 |
| | Layer1 | BasicBlock*2 | 64 | 64 | 3 | 1 | 1 |
| | Layer2 | | 64 | 128 | 3 | 2 | 1 |
| | Layer3 | | 128 | 256 | | | |
| | Layer4 | | 256 | 512 | | | |
| 边缘分支 | Conv1 | Conv*2 | 3 | 64 | 3 | 2 | 1 |
| | Layer1 | BasicBlock*2 | 64 | 64 | 3 | 1 | 1 |
| | Layer2 | | 64 | 128 | 3 | 2 | 1 |
| | Layer3 | | 128 | 64 | 3 | 1 | 1 |
| | Layer4 | | 64 | 128 | | | |
| | Layer5 | | 128 | 256 | | | |
| 纹理分支 | Layer3 | BasicBlock*2 | 64 | 128 | 3 | 1 | 1 |
| | Layer4 | | 128 | 128 | | | |
| | Layer5 | | 128 | 256 | | | |

表 2 Cityscape 数据集上的部分超参数设置

| 参数名称 | 数值 | 描述 |
|---|---|---|
| *BackBone* | ResNet-50 | 使用ResNet公共网络进行基础特征提取 |
| 初始学习率 | 0.005 | 使用SGD进行优化 |
| *BatchSize* | 8 | 每个GPU分配的*BatchSize* |
| *Epochs* | 250 | 总训练轮次 |
| *CropSize* | 712 | 随机裁剪尺寸 |
| 权重衰减 | 0.000 5 | 防止过拟合 |
| 数据增强 | 旋转、翻转、裁剪、CutMix等 | 提高泛化能力 |
| 边缘检测 | Canny(0.1, 0.2) | 算法提取边缘标签 |
| 损失函数 | 交叉熵+一致性损失 | $\lambda_u$、$\lambda_l$ 均为1.0 |
| 伪标签策略 | 初始阈值0.4 | 阈值随训练进度调整 |
| 优化策略 | 学习率衰减 | 采用 Poly Learning Rate |

### 3.3 对比实验

为了验证本文方法有效性, 在不同基准数据集下对比了最新的半监督语义分割方法, 包括 PCR[40], CCVC[39] 等. 此外还展示了仅使用标记数据的监督训练的结果 (记为 Baseline). 由于半监督语义分割现有方案对比均基于相同 *BackBone* 下的 mIoU 对比, 因此本文也遵循现有的对比原则, 并未对参数量和推理速度等指标进行对比.

首先在原始 Pascal VOC 2012 数据集[30]上选用 ResNet-101 作为骨干网络验证了本文方法与其他方法的性能表现, 结果如表 3 所示, 红色最优, 蓝色次之. 可以看到本文方法在不同有标签比例的分区协议 (分区协议 1/*n* 表示将 1/*n* 图像作为标签数据集, 其余图像作为无标签数据集) 下都拥有较强的竞争力. 本文方法在 1/2 分区协议下领先 UniMatch[27]0.3%, 在 1/4 和 full 分区协议下仅分别落后 UniMatch[27]0.2% 和 0.1%. 然而在 1/16 分区下, 本文方法性能相较 UniMatch 落后最多, 差距为 2.6%. 这一结果侧面印证了尽管本文方法在标签数据较少的情况下仍表现出较强的性能, 但在标签数据极为稀缺的情况下仍存在进一步优化的空间.

表 3 不同含标签比例的分区协议下 Classic Pascal VOC 2012[30]数据集和 Blender Pascal VOC 2012[48]数据集的性能对比 (%)

| 方法 | Classic Pascal VOC 2012 | | | | | Blender Pascal VOC 2012 | | |
|---|---|---|---|---|---|---|---|---|
| | 1/16 (92) | 1/8 (183) | 1/4 (366) | 1/2 (732) | Full (1 464) | 1/16 (662) | 1/8 (1 323) | 1/4 (2 646) |
| Baseline | 45.1 | 55.3 | 64.8 | 69.7 | 73.5 | 67.5 | 71.1 | 74.2 |
| Pseudo[17] | 57.6 | 65.5 | 69.1 | 72.4 | 73.2 | — | — | — |
| CPS[19] | 64.1 | 67.4 | 71.7 | 75.9 | — | 74.5 | 76.4 | 77.7 |
| RC2L[49] | 65.3 | 68.9 | 72.2 | 77.1 | 79.3 | — | — | — |
| PCR[40] | 70.1 | 74.7 | 77.2 | 78.5 | 80.7 | 78.6 | **80.7** | **80.8** |
| CCVC[39] | 70.2 | 74.4 | 77.4 | 79.1 | 80.5 | 77.2 | 78.4 | 79.0 |
| S4MC[50] | 71.0 | 71.7 | 75.4 | 77.7 | 80.6 | 78.5 | 79.7 | 79.9 |
| UniMatch[27] | **75.2** | **77.2** | **78.8** | **79.9** | **81.2** | 78.1 | 78.4 | 80.4 |
| MLLC[51] | 70.6 | 74.3 | 77.4 | 79.3 | — | **78.9** | 80.3 | **80.8** |
| IPixMatch[43] | **73.9** | 74.6 | 77.1 | 78.9 | 79.4 | 77.2 | 78.2 | 78.8 |
| Ours | 72.6 | **76.3** | **78.6** | **80.2** | **81.1** | **80.7** | **82.1** | **81.8** |

此外还进一步验证了本文方法在 Blender Pascal VOC 2012 数据集[48]上的性能表现, 仅使用 ResNet-101 作为骨干网络的结果展示在表 3 中. 本文方法在 1/4 的划分比例下超过 PCR[40]和 MLLC[51]方法 1.0%, 在标签数据更少的 1/16 分区协议下也拥有相较于 MLLC[51]2.1% 的性能领先, 上述实验结果证实了本文方法的有效性.

本文对不同方法性能差异的原因进行了深入分析. PseudoSeg[17]方法仅对无标签图像生成伪标签作为监督信号, 网络生成的错误预测可能导致伪标签存在错误, 从而使网络受到错误的监督信号影响, 导致性能下降.

CCVC[39]虽然也提出了分支网络提取差异信息, 但其网络架构完全相同, 差异化依赖于线性特征映射层, 线性映射前不同分支的特征提取差异化无法得到保证, 网络分支同化的问题仍然存在. PCR[40]将原型网络引入半监督语义分割, 与全卷积神经网络构成多分支网络, 但其受限于一致正则化, 完全相同的监督信号在出现错误时无法准确检测并排除, 网络末端的错误检测能力得不到保证. IPixMatch[43]引入了像素上下文损失, 在极少标签的情况下表现较好, 也侧面印证了本文交叉融合伪标签的合理性. UniMatch[27]将强弱对比策略进行了扩展, 不仅在数据增强阶段应用强弱对比方案, 还在解码阶段通过使用 dropout 对编码进行扰动生成新的分支. 四分支网络在标签数据极少的情况下能够获取更多的可用信息. 此外, UniMatch[27]在无标签端的损失计算采用了熵最小化的一致性损失, 而不是采用伪标签方法, 这使得其在处理错误预测的收敛更慢, 有效避免错误干扰. 尽管在 Classic Pascal VOC 2012[30]的 1/16 分区协议下, UniMatch 性能领先本工作 2.6%, 但其训练时间较长, 训练成本相对较高. 相对而言, Blender Pascal VOC 2012 数据集[48]的数据量较大, 因此在不同分区协议下本方法能够获得更多准确监督, 这也导致在各分区协议下本文方法的性能普遍优于 MLLC[51]等方法. 不同数据集下的性能差异表明, 本文方法在数据量较大的情况下表现更为出色, 而在标签数据极少的情况下, 仍会受到较多错误信息的干扰, 从而影响性能.

表 4 对比了本文方法以及现有方法在 Cityscapes 数据集上的性能差异 (CPS 结果转载自 U2PL[40]结果), 红色最优, 蓝色次之. 由于 Cityscapes 数据集训练规模更大、耗时更长, 为了加快实验进程, 本文仅采用 ResNet-50 进行训练. 从结果可以看到本文方法性能相较于其他方法均有不同幅度领先, 特别在只有 186 个标签数据的 1/16 分区协议下, 本文方法更是超过 CPSR[42]方法 1.3%, 即使在标签数据较多的 1/4 分区协议时也与 CPSR[42]并列最优, 并领先 UniMatch 和 IPixMatch 方法 0.5%, 取得了有竞争力的结果.

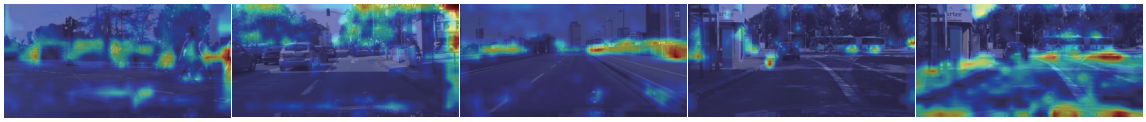
表 4　Cityscapes 数据集不同含标签比例的分区协议下分割性能对比 (%)

| Method | 1/16 (186) | 1/8 (372) | 1/4 (744) |
|---|---|---|---|
| Baseline | 63.3 | 65.8 | 68.4 |
| CCT[20] | 66.4 | 72.5 | 75.7 |
| GCT[52] | 65.8 | 71.3 | 75.3 |
| CPS[19] | 69.8 | 74.3 | 74.6 |
| ELN[15] | — | 70.3 | 73.5 |
| U2PL[41] | 69.0 | 73.0 | 76.3 |
| USRN[53] | 71.2 | 75.0 | — |
| CCVC[39] | 74.9 | 76.4 | 77.3 |
| UniMatch[27] | 75.0 | 76.8 | **77.5** |
| CPSR[42] | **75.5** | **77.3** | **78.0** |
| VC3[44] | 74.8 | 76.8 | 77.2 |
| IPixMatch[43] | 74.1 | 76.0 | **77.5** |
| Ours | **76.8** | **77.5** | **78.0** |

从图 7 中可以看到本文交叉监督的差异化特征提取方法的合理性, 语义分支对于主体区域的识别更加准确, 细节纹理分支可以保留关注更多细节信息, 而边缘分支更多的关注图像中的边界区域. 语义分支和纹理分支提取不同的信息, 保障了不同分支之间的差异性, 解决了现有半监督语义分割方法网络分支同化的缺陷.

在图 8 中, 展示了本文方法在 1/4 分区下的 Cityscapes 数据集上的分割效果, 可以看出, 本文方法在各种难以识别分割的区域都有明显的改善. 本文的方法在识别和分割形状边缘更加明显的信号灯、信号牌等小目标物体时表现出更强的能力. 相比于其他方法, 第 2 行分割效果图中细粒度更高的 bus 分割任务中, 本文的方法能够更准确地区分对应像素类别, 而其他方法会将部分区域错误标记为 car. 此外, 在第 3 行具有长连续区域分割效果展示的 sidewalk 分割任务中, 本文的方法在处理边缘区域时表现更出色, 具有更好的连续性.


(a) 边界调和模块后的语义分支热力图可视化



(b) 边界调和模块前的语义分支热力图可视化

(c) 差异筛选模块后的细节纹理分支热力图可视化

(d) 差异筛选模块前的细节纹理分支热力图可视化

(e) 细节引导模块后的边缘形状分支热力图可视化

图 7　单类别热力激活可视化结果

(a) 传统一致正则化方法

(b) PCR方法

(c) UniMatch方法

(d) 本文方法

图 8　来自于 Cityscapes 数据集的 1/4 有标签数据的分区协议下推理对比

(e) 标签图像, 黑色区域为无需分割区域

图 8 来自于 Cityscapes 数据集的 1/4 有标签数据的分区协议下推理对比 (续)

### 3.4 消融实验

在本节中, 将分析基于差异化特征提取的交叉半监督语义分割网络中不同设计模块的有效性以及多分支网络的性能变化[54,55]. 本文选用基于 ResNet-101 作为 *BackBone* 的网络架构, 在 Classic Pascal VOC 2012 数据集上进行所有的消融实验, 分区协议均为 1/4 含标签数据.

#### 3.4.1 成分的有效性

本文探究了差异化特征提取策略以及交叉融合伪标签方法对模型性能的影响, 分析结果展示在表 5 中. “纹理”指的是差异筛选模块, “边缘” 指的是细节引导模块, 而“上下文”指的是边界调和模块. 红色最优, 蓝色次之. 此外, 在表 6 本文还探索了交叉融合伪标签采用不同融合策略以及邻域范围对网络性能的影响.

表 5 差异化特征提取方法在 ResNet-101 上的消融实验结果 (%)

| 差异化特征提取 | | | 交叉伪标签 | mIoU |
|---|---|---|---|---|
| 纹理 | 边缘 | 上下文 | | |
| — | √ | √ | √ | **78.1** |
| √ | — | √ | √ | 77.8 |
| √ | √ | — | √ | 76.1 |
| — | — | — | √ | 75.7 |
| √ | √ | √ | — | 77.4 |
| √ | √ | √ | √ | **78.6** |

表 6 交叉融合伪标签邻域像素融合策略和邻域范围的消融实验 (%)

| 融合策略 | 邻域范围 | | | |
|---|---|---|---|---|
| | None | 3×3 | 5×5 | 7×7 |
| 联合概率修正 | 77.4 | **78.6** | **78.3** | 77.5 |
| 余弦相似度 | | 77.8 | 77.4 | 76.7 |
| 欧氏距离 | | 77.6 | 76.9 | 76.2 |

在表 5 的成分消融中, 不采用差异筛选模块指的是使用特征相加来融合其他分支信息, 其他模块同理, 不使用交叉伪标签则是采用传统一致正则化方法. 从结果中可以看出, 模型性能在失去了差异化特征提取的任意分支模块后均有不同程度的降低. 在语义上下文分支中的边界调和模块最为显著, 性能降低了 2.5%. 在失去了完整差异化特征提取策略后, 纹理分支和语义上下文分支完全一致, 导致性能降低 2.9%. 交叉伪标签保障了分支监督信号的准确性和差异性, 对比一致性方法提升了 1.2%.

如表 6 所示, 本文测试了不同邻域像素融合策略和邻域范围大小的性能变化, “余弦相似度”融合策略采用邻域内和给定像素余弦相似度最高的邻域像素进行平均融合. “欧式距离”融合策略将邻域内和给定像素欧式距离最近的邻域像素进行平均融合. “None”表示不采用像素融合策略, 仅依靠不同网络分支一致性损失来训练无标签图像. 从结果中可以看出, 当选择邻域范围为 3×3 的“联合概率修正策略”时, 模型表现出最佳性能. 而融合过多邻域信息的 5×5 和 7×7 反而导致模型性能下降. 其原因在于, 较大的领域范围引入了更多的特征信息, 增加了出现非中心像素点类别信息的可能性, 从而影响置信度的融合效果.

#### 3.4.2 方法有效性

如图 9 所示, 细节纹理分支和语义上下文分支保持近似的分割性能, 纹理分支在 mIoU 表现上稍逊于语义上

下文分支. 同时, 本文仅对置信度高于伪标签阈值的像素进行传播[18,56−60], 这并不会在损失传播阶段过度干扰语义上下文分支的训练, 证明了本文方法多分支网络相互监督能力的有效性.

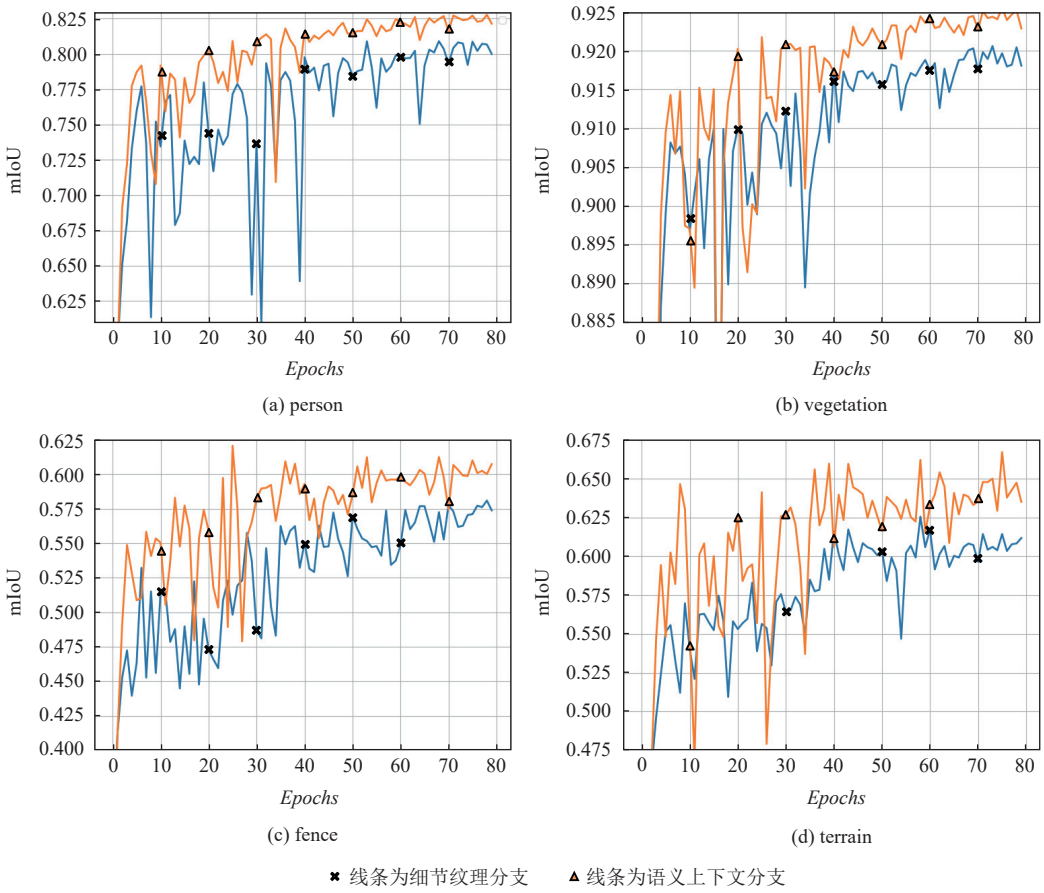



图 9 Cityscapes 数据集中的 4 个类别的 mIoU 变化曲线

### 3.5 缺陷与分析

尽管本文方法在不同基准数据集上的性能均优于现有方法, 但仍有两方面需要进一步优化: (1) 虽然推理阶段仅使用语义上下文分支进行预测, 但在训练阶段, 差异化特征提取过程中涉及的边缘监督、信号合成以及邻近融合伪标签的计算耗费了更多的时间. (2) 如图 9 所示, 纹理分支的性能在大多数情况下略低于语义上下文分支. 图 10 中展示了分支分割性能差异更加明显的可视化分割效果, 细节纹理分支的分割效果相对较差, 尤其在语义信息相似的细粒度分类中表现较为不足. 图 10 第 1 行为语义上下文分支, 第 3 行为细节纹理分支, 第 2 行为细节纹理分支和语义上下文分支的局部放大图. 尽管本文采用了伪标签阈值来过滤低置信度的伪标签, 但仍有可能出现高置信度的错误伪标签干扰语义上下文分支的收敛. 伪标签阈值筛选的局限在于, 即便是正确预测的低置信度伪标签, 也可能因为低于阈值而被过滤掉. 虽然阈值筛选有效避免了错误标签的传播, 但未充分训练低阈值伪标签对应的像素, 这会在一定程度上影响网络的收敛速度. 此外, 本工作在 1/16 标签比例的 Classic Pascal VOC 2012 数据集上, 相较于 UniMatch 方法的性能落后了 2.6%, 这表明在标签数据极少的情况下, 本方法仍存在进一步优化的空间. 如何在监督信息极为有限的条件下有效避免错误信息的干扰, 将是本工作未来研究的一个重要方向. 总体而言, 未来的研究方向将聚焦于如何进一步优化网络训练时间, 并平衡多分支网络的性能表现, 并优化极少标签下的性能表现.

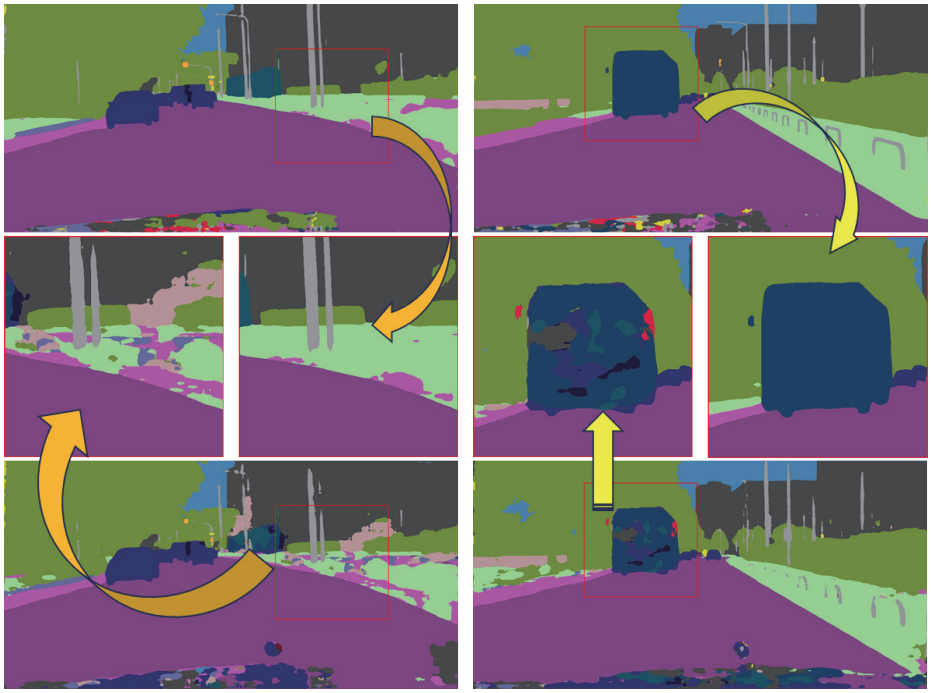


图 10　在 Cityscapes 数据集下的细节纹理分支和语义上下文分支的分割可视化

## 4 结　论

本文提出了一种基于差异化特征提取的交叉半监督语义分割网络. 首先, 差异化特征提取策略增强了纹理和边缘信息的关注程度, 并利用边界预测来平衡纹理和语义信息的合成, 从特征提取角度使源端信号始终存在差异性, 更有效地利用无标记数据来优化网络性能. 其次, 交叉融合伪标签策略通过邻近预测和交叉监督提高了网络末端监督信号的准确性和差异性, 避免了因网络同化而出现纠错能力失效的问题. 本文方法在不同的基准数据集下均取得了有竞争力的结果.

虽然本文的方法取得了较好的性能表现, 但仍存在一些缺陷, 例如多分支网络训练时间较长以及性能表现不均衡的问题.

在未来的研究工作中, 本文计划从以下几个角度优化现有的问题: (1) 减少多分支网络的数量: 将致力于从一个网络分支中提取不同的特征信息, 例如同时提取细节纹理信息和语义上下文信息, 使它们均能参与分割任务. 此外, 本工作计划优化边缘形状分支, 采用更加精准高效的边缘检测算法, 从而减轻网络的训练时间和压力. (2) 提高多分支网络性能的均衡性和极少标签下的性能表现: 针对多分支网络性能不均衡的问题, 计划引入高置信度不更新策略. 当分支网络对像素的预测保持高置信度时, 不对该分支的像素进行监督和反向传播, 以此避免性能较差的分支生成高置信度的错误伪标签, 干扰高性能分支的训练.

**References:**

[1] Cai DG, Zhao LC, Zhang J, Sheng L, Xu D. 3DJCG: A unified framework for joint dense captioning and visual grounding on 3D point clouds. In: Proc. of the 2022 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022. 16443–16452. [doi: 10.1109/CVPR52688.2022.01597]

[2] Fu J, Liu J, Tian HJ, Li Y, Bao YJ, Fang ZW, Lu HQ. Dual attention network for scene segmentation. In: Proc. of the 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 3141–3149. [doi: 10.1109/CVPR.2019.00326]

[3] Li XT, Zhao HL, Han L, Tong YH, Tan SH, Yang KY. Gated fully fusion for semantic segmentation. In: Proc. of the 34th AAAI Conf.

on Artificial Intelligence. New York: AAAI, 2020. 11418–11425. [doi: 10.1609/aaai.v34i07.6805]

[4] Tian X, Wang L, Ding Q. Review of image semantic segmentation based on deep learning. Ruan Jian Xue Bao/Journal of Software, 2019, 30(2): 440–468 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/5659.htm [doi: 10.13328/j.cnki.jos.005659]

[5] Ahn J, Kwak S. Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation. In: Proc. of the 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 4981–4990. [doi: 10.1109/CVPR.2018.00523]

[6] Khoreva A, Benenson R, Hosang J, Hein M, Schiele B. Simple does it: Weakly supervised instance and semantic segmentation. In: Proc. of the 2017 IEEE Conf. on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 1665–1674. [doi: 10.1109/CVPR.2017.181]

[7] Lee J, Choi J, Mok J, Yoon S. Reducing information bottleneck for weakly supervised semantic segmentation. In: Proc. of the 35th Int'l Conf. on Neural Information Processing Systems. Virtual Event: Curran Associates Inc., 2021. 27408–27421.

[8] Yun F, Yin YJ, Zhang WX, Zhi M. Adversarial semi-supervised semantic segmentation with attention mechanism. Computer Engineering & Applications, 2023, 59(8): 254–262 (in Chinese with English abstract). [doi: 10.3778/j.issn.1002-8331.2112-0484]

[9] Feng X, Yang J, Zhou T, Gong C. Weakly supervised object localization based on attention mechanism and categorical hierarchy. Ruan Jian Xue Bao/Journal of Software, 2023, 34(10): 4916–4929 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/6675.htm [doi: 10.13328/j.cnki.jos.006675]

[10] Alonso I, Sabater A, Ferstl D, Montesano L, Murillo AC. Semi-supervised semantic segmentation with pixel-level contrastive learning from a class-wise memory bank. In: Proc. of the 2021 IEEE/CVF Int' Conf. on Computer Vision. Montreal: IEEE, 2021. 8199–8208. [doi: 10.1109/ICCV48922.2021.00811]

[11] Chen HA, Jin Y, Jin GQ, Zhu CA, Chen EH. Semisupervised semantic segmentation by improving prediction confidence. IEEE Trans. on Neural Networks and Learning Systems, 2022, 33(9): 4991–5003. [doi: 10.1109/TNNLS.2021.3066850]

[12] French G, Laine S, Aila T, Mackiewicz M, Finlayson G. Semi-supervised semantic segmentation needs strong, varied perturbations. arXiv:1906.01916, 2020.

[13] Bachute MR, Subhedar JM. Autonomous driving architectures: Insights of machine learning and deep learning algorithms. Machine Learning with Applications, 2021, 6: 100164. [doi: 10.1016/j.mlwa.2021.100164]

[14] Jiao RS, Zhang YC, Ding L, Xue BS, Zhang JC, Cai R, Jin C. Learning with limited annotations: A survey on deep semi-supervised learning for medical image segmentation. Computers in Biology and Medicine, 2024, 169: 107840. [doi: 10.1016/j.compbiomed.2023.107840]

[15] Kwon D, Kwak S. Semi-supervised semantic segmentation with error localization network. In: Proc. of the 2022 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022. 9947–9957. [doi: 10.1109/CVPR52688.2022.00972]

[16] Mendel R, De Souza LA Jr, Rauber D, Papa JP, Palm C. Semi-supervised segmentation based on error-correcting supervision. In: Proc. of the 16th European Conf. Computer Vision (ECCV 2020). Glasgow: Springer Int'l Publishing, 2020. 141–157. [doi: 10.1007/978-3-030-58526-6_9]

[17] Zou YL, Zhang ZZ, Zhang H, Li CL, Bian X, Huang JB, Pfister T. PseudoSeg: Designing pseudo labels for semantic segmentation. In: Proc. of the 9th Int'l Conf. on Learning Representations. Virtual Conf. 2021.

[18] Yang LH, Zhuo W, Qi L, Shi YH, Gao Y. ST++: Make self-training work better for semi-supervised semantic segmentation. In: Proc. of the 2022 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022. 4258–4267. [doi: 10.1109/CVPR52688.2022.00423]

[19] Chen XK, Yuan YH, Zeng G, Wang JD. Semi-supervised semantic segmentation with cross pseudo supervision. In: Proc. of the 2021 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 2613–2622. [doi: 10.1109/CVPR46437.2021.00264]

[20] Ouali Y, Hudelot C, Tami M. Semi-supervised semantic segmentation with cross-consistency training. In: Proc. of the 2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 12671–12681. [doi: 10.1109/CVPR42600.2020.01269]

[21] Na J, Ha JW, Chang HJ, Han D, Hwang W. Switching temporary teachers for semi-supervised semantic segmentation. In: Proc. of the 37th Int'l Conf. on Neural Information Processing Systems. New Orleans: Curran Associates Inc., 2023. 40367–40380.

[22] Zhao Z, Yang LH, Long SF, Pi JM, Zhou LP, Wang JD. Augmentation matters: A simple-yet-effective approach to semi-supervised semantic segmentation. In: Proc. of the 2023 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023. 11350–11359. [doi: 10.1109/CVPR52729.2023.01092]

[23] Xu HJ, Xiao YF. Semi-supervised semantic segmentation method based on multiple teacher network model. Computer Science, 2023, 50(12): 279–284 (in Chinese with English abstract). [doi: 10.11896/jsjkx.221000245]

[24] Kalluri T, Varma G, Chandraker M, Jawahar CV. Universal semi-supervised semantic segmentation. In: Proc. of the 2019 IEEE/CVF Int'l

Conf. on Computer vision. Seoul: IEEE, 2019. 5258–5269. [doi: 10.1109/ICCV.2019.00536]

[25] Berthelot D, Carlini N, Goodfellow I, Oliver A, Papernot N, Raffel C. MixMatch: A holistic approach to semi-supervised learning. In: Proc. of the 33rd Int'l Conf. on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2019. 5049–5059.

[26] Arazo E, Ortego D, Albert P, O'Connor NE, McGuinness K. Pseudo-labeling and confirmation bias in deep semi-supervised learning. In: Proc. of the 2020 Int'l Joint Conf. on Neural Networks (IJCNN). Glasgow: IEEE, 2020. 1–8. [doi: 10.1109/IJCNN48605.2020.9207304]

[27] Yang LH, Qi L, Feng LT, Zhang W, Shi YH. Revisiting weak-to-strong consistency in semi-supervised semantic segmentation. In: Proc. of the 2023 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023. 7236–7246. [doi: 10.1109/CVPR52729.2023.00699]

[28] Liu LM, Zong JX, Xiao ZJ, Lan H, Qu HC. Cross-consistent semantic segmentation algorithm based on manifold regularization. Journal of Image and Graphics, 2022, 27(12): 3542–3552 (in Chinese with English abstract). [doi: 10.11834/jig.210571]

[29] Li YT, Zhang CL. Algorithm research of deep learning in semi-supervised semantic segmentation. Artificial Intelligence and Robotics Research, 2023, 12(4): 328–339 (in Chinese with English abstract). [doi: 10.12677/AIRR.2023.124036]

[30] Everingham M, Van Gool L, Williams CKI, Winn J, Zisserman A. The pascal visual object classes (voc) challenge. Int'l Journal of Computer Vision, 2010, 88(2): 303–338. [doi: 10.1007/s11263-009-0275-4]

[31] Cordts M, Omran M, Ramos S, Rehfeld T, Enzweiler M, Benenson R, Franke U, Roth S, Schiele B. The cityscapes dataset for semantic urban scene understanding. In: Proc. of the 2016 IEEE Conf. on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 3213–3223. [doi: 10.1109/CVPR.2016.350]

[32] Zhao Z, Long SF, Pi JM, Wang JD, Zhou LP. Instance-specific and model-adaptive supervision for semi-supervised semantic segmentation. In: Proc. of the 2023 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023. 23705–23714. [doi: 10.1109/CVPR52729.2023.02270]

[33] Chen SH, Chen YD, Zheng YH, Yang ZX, Wu EH. A transformer-based adaptive prototype matching network for few-shot semantic segmentation. In: Proc. of the 33rd Int'l Joint Conf. on Artificial Intelligence (IJCAI), Jeju, 2024. 659–667.

[34] Kurakin A, Raffel C, Berthelot D. RemixMatch: Semi-supervised learning with distribution matching and augmentation anchoring. 2020.

[35] Sohn K, Berthelot D, Li CL, Zhang ZZ, Carlini N, Cubuk ED, Kurakin A, Zhang H, Raffel C. FixMatch: Simplifying semi-supervised learning with consistency and confidence. In: Proc. of the 34th Int'l Conf. on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2020. 596–608.

[36] Liu YY, Tian Y, Chen YH, Liu FB, Belagiannis V, Carneiro G. Perturbed and strict mean teachers for semi-supervised semantic segmentation. In: Proc. of the 2022 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022. 4248–4257. [doi: 10.1109/CVPR52688.2022.00422]

[37] Olsson V, Tranheden W, Pinto J, Svensson L. ClassMix: Segmentation-based data augmentation for semi-supervised learning. In: Proc. of the 2021 IEEE Winter Conf. on Applications of Computer Vision. Waikoloa: IEEE, 2021. 1368–1377. [doi: 10.1109/WACV48630.2021.00141]

[38] Jiang F, Gu Q, Hao HZ, Li N, Guo YW, Chen DX. Survey on content-based image segmentation methods. Ruan Jian Xue Bao/Journal of Software, 2017, 28(1): 160–183 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/5136.htm [doi: 10.13328/j.cnki.jos.005136]

[39] Wang ZC, Zhao Z, Xing XX, Xu D, Kong XY, Zhou LP. Conflict-based cross-view consistency for semi-supervised semantic segmentation. In: Proc. of the 2023 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023. 19585–19595. [doi: 10.1109/CVPR52729.2023.01876]

[40] Xu HM, Liu LQ, Bian QC, Yang Z. Semi-supervised semantic segmentation with prototype-based consistency regularization. In: Proc. of the 36th Int'l Conf. on Neural Information Processing Systems. New Orleans: Curran Associates Inc., 2022. 26007–26020.

[41] Wang YC, Wang HC, Shen YJ, Fei JJ, Li W, Jin GQ, Wu LW, Zhao R, Le XY. Semi-supervised semantic segmentation using unreliable pseudo-labels. In: Proc. of the 2022 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022. 4238–4247. [doi: 10.1109/CVPR52688.2022.00421]

[42] Yin JJ, Yan S, Chen T, Chen Y, Yao YZ. Class probability space regularization for semi-supervised semantic segmentation. Computer Vision and Image Understanding, 2024, 249: 104146. [doi: 10.1016/j.cviu.2024.104146]

[43] Wu KB, Li WB, Xiao XF. IPixMatch: Boost semi-supervised semantic segmentation with inter-pixel relation. arXiv:2404.18891, 2024.

[44] Hou YZ, Gould S, Zheng L. View-coherent correlation consistency for semi-supervised semantic segmentation. Pattern Recognition, 2024, 147: 110089. [doi: 10.1016/j.patcog.2023.110089]

[45] Li YL, Gao Y, Yan JL, Zou BH, Wang JM. Image inpainting methods based on deep neural networks: A review. Chinese Journal of Computes, 2021, 44(11): 2295–2316 (in Chinese with English abstract). [doi: 10.11897/SP.J.1016.2021.02295]

[46] Rong WB, Li ZJ, Zhang W, Sun LN. An improved CANNY edge detection algorithm. In: Proc. of the 2014 IEEE Int'l Conf. on Mechatronics and Automation. Tianjin: IEEE, 2014. 577–582. [doi: 10.1109/ICMA.2014.6885761]

[47] Zhu XZ, Cheng DZ, Zhang Z, Lin S, Dai JF. An empirical study of spatial attention mechanisms in deep networks. In: Proc. of the 2019 IEEE/CVF Int'l Conf. on Computer Vision. Seoul: IEEE, 2019. 6687–6696. [doi: 10.1109/ICCV.2019.00679]

[48] Hariharan B, Arbeláez P, Bourdev L, Maji S, Malik J. Semantic contours from inverse detectors. In: Proc. of the 2011 Int'l Conf. on Computer Vision. Barcelona: IEEE, 2011. 991–998. [doi: 10.1109/ICCV.2011.6126343]

[49] Zhang JR, Wu TY, Ding CH, Zhao HW, Guo GD. Region-level contrastive and consistency learning for semi-supervised semantic segmentation. arXiv:2204.13314, 2022.

[50] Kimhi M, Kimhi S, Zheltonozhskii E, Litany O, Baskin C. Semi-supervised semantic segmentation via marginal contextual information. Trans. on Machine Learning Research, 2024.

[51] Xiao H, Hong YT, Dong L, Yan DQ, Xiong JJ, Zhuang JY, Liang DT, Peng CB. Multi-level label correction by distilling proximate patterns for semi-supervised semantic segmentation. IEEE Trans. on Multimedia, 2024, 26: 8077–8087. [doi: 10.1109/TMM.2024.3374594]

[52] Ke ZH, Qiu D, Li KC, Yang Q, Lau RWH. Guided collaborative training for pixel-wise semi-supervised learning. In: Proc. of the 16th European Conf. on Computer Vision (ECCV 2020). Glasgow: Springer, 2020. 429–445. [doi: 10.1007/978-3-030-58601-0_26]

[53] Guan DY, Huang JX, Xiao AR, Lu SJ. Unbiased subclass regularization for semi-supervised semantic segmentation. In: Proc. of the 2022 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022. 9958–9968. [doi: 10.1109/CVPR52688.2022.00973]

[54] Chen YD, Zhao YB, Wu EH. Robust semi-supervised video object segmentation with dynamic embedding. Journal of Beijing University of Aeronautics and Astronautics, 2025, 51(7): 2253–2261. (in Chinese with English abstract). [doi: 10.13700/j.bh.1001-5965.2023.0354]

[55] You CY, Dai WC, Min YF, Staib L, Duncan JS. Bootstrapping semi-supervised medical image segmentation with anatomical-aware contrastive distillation. In: Proc. of the 28th Int'l Conf. on Information Processing in Medical Imaging. San Carlos de Bariloche: Springer, 2023. 641–653. [doi: 10.1007/978-3-031-34048-2_49]

[56] Dai ZH, Liu HX, Le QV, Tan MX. CoAtNet: Marrying convolution and attention for all data sizes. In: Proc. of the 35th Int'l Conf. on Neural Information Processing Systems. Virtual Event: Curran Associates Inc., 2021. 3965–3977.

[57] Xu Y, Shang L, Ye JX, Qian Q, Li YF, Sun BG, Li H, Jin R. Dash: Semi-supervised learning with dynamic thresholding. In: Proc. of the 38th Int'l Conf. on Machine Learning. 2021. 11525–11536.

[58] Zuo SM, Yu Y, Liang C, Jiang HM, Er S, Zhang C, Zhao T, Zha HY. Self-training with differentiable teacher. In: Proc. of the Findings of the Association for Computational Linguistics: NAACL 2022. Seattle: Association for Computational Linguistics, 2022. 933–949. [doi: 10.18653/v1/2022.findings-naacl.70]

[59] Bartolomei L, Teixeira L, Chli M. Perception-aware path planning for UAVs using semantic segmentation. In: Proc. of the 2020 IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems (IROS). Las Vegas: IEEE, 2020. 5808–5815. [doi: 10.1109/IROS45743.2020.9341347]

[60] Chen LC, Zhu YK, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proc. of the 15th European Conf. on Computer Vision. Munich: Springer, 2018. 833–851. [doi: 10.1007/978-3-030-01234-2_49]

**附中文参考文献**:

[4] 田萱, 王亮, 丁琪. 基于深度学习的图像语义分割方法综述. 软件学报, 2019, 30(2): 440–468. http://www.jos.org.cn/1000-9825/5659.htm [doi: 10.13328/j.cnki.jos.005659]

[8] 云飞, 殷雁君, 张文轩, 智敏. 融合注意力机制的对抗式半监督语义分割. 计算机工程与应用, 2023, 59(8): 254–262. [doi: 10.3778/j.issn.1002-8331.2112-0484]

[9] 冯迅, 杨健, 周涛, 宫辰. 基于注意力机制及类别层次结构的弱监督目标定位. 软件学报, 2023, 34(10): 4916–4929. http://www.jos.org.cn/1000-9825/6675.htm [doi: 10.13328/j.cnki.jos.006675]

[23] 许华杰, 肖毅烽. 基于多教师网络模型的半监督语义分割方法. 计算机科学, 2023, 50(12): 279–284. [doi: 10.11896/jsjkx.221000245]

[28] 刘腊梅, 宗佳旭, 肖振久, 兰海, 曲海成. 流形正则化的交叉一致性语义分割算法. 中国图象图形学报, 2022, 27(12): 3542–3552. [doi: 10.11834/jig.210571]

[29] 李一彤, 张长伦. 基于深度学习的半监督语义分割算法研究. 人工智能与机器人研究, 2023, 12(4): 328–339. [doi: 10.12677/AIRR.2023.124036]

[38] 姜枫, 顾庆, 郝慧珍, 李娜, 郭延文, 陈道蓄. 基于内容的图像分割方法综述. 软件学报, 2017, 28(1): 160–183. http://www.jos.org.cn/1000-9825/5136.htm [doi: 10.13328/j.cnki.jos.005136]

[45] 李月龙, 高云, 闫家良, 邹佰翰, 汪剑鸣. 基于深度神经网络的图像缺损修复方法综述. 计算机学报, 2021, 44(11): 2295–2316. [doi: 10.11897/SP.J.1016.2021.02295]

[54] 陈亚当, 赵翊冰, 吴恩华. 基于动态嵌入特征的鲁棒半监督视频目标分割. 北京航空航天大学学报, 2025, 51(7): 2253–2261. [doi: 10.13700/j.bh.1001-5965.2023.0354]

**陈亚当**(1985－), 男, 博士, 副教授, 主要研究领域为基于机器学习, 深度学习的计算机视觉技术.

**车洵**(1985－), 男, 博士生, 教授, CCF 高级会员, 主要研究领域为模型鲁棒性与安全性研究.

**李家威**(1999－), 男, 硕士生, 主要研究领域为计算机视觉, 半监督语义分割.

**吴恩华**(1947－), 男, 博士, 研究员, 博士生导师, CCF 会士, 主要研究领域为计算机图形学, 计算机软件.