中图法分类号:TP391.41　　**文献标识码**:A　　**文章编号**:1006-8961(XXXX)XX-0001-20

论文引用格式:Hu Jing, Lv Xiaopeng, Wang Fang, Zhang Rui. Spatial-frequency consistency-driven semi-supervised semantic segmentation: SF-Match[J/OL]. Journal of Image and Graphics, XXXX:1-20. DOI: 10.11834/jig.260246. (胡静, 吕晓鹏, 王芳, 张睿. 空频域一致性驱动的半监督语义分割 SF-Match[J/OL]. 中国图象图形学报, XXXX:1-20. DOI: 10.11834/jig.260246.)[DOI:10.11834/jig.260246]

空频域一致性驱动的半监督语义分割 SF-Match

胡静,吕晓鹏*,王芳,张睿

太原科技大学计算机科学与技术学院,太原　030024

摘　要:**目的**　针对现有半监督语义分割方法特征表达不足,以及全图统一一致性约束易造成语义边界模糊和目标内部预测不连贯的问题,提出空频一致性驱动的半监督语义分割框架SF-Match(Spatial-Frequency Match)。**方法**　以自监督预训练DINOv2为编码器、DPT(Dense Prediction Transformer)为解码器,构建兼顾全局语义与局部细节的密集预测网络;在教师—学生框架中,对CutMix后的强增强图像构造高频和低频视图,并依据教师伪标签的形态学梯度动态生成边界与内部掩膜;分别在边界区域施加高频一致性约束、在内部区域施加低频一致性约束,与RGB全局一致性和监督损失联合优化,实现频域信息与空间语义区域的解耦对齐。**结果**　在Pascal VOC 2012和Cityscapes数据集上开展多标注比例对比、消融、困难类别、复杂度及定性实验。Pascal VOC 2012上,SF-Match的平均交并比(mean Intersection over Union,mIoU)为86.4%～90.6%,较semiVL提高1.9～3.6个百分点,较UniMatch提高9.4～11.2个百分点;Cityscapes上,mIoU为83.6%～85.0%,较各划分下最佳外部对比方法提高3.7～5.1个百分点。消融实验中,Pascal VOC 2012的1/4标注划分下,mIoU由传统基线的77.8%提高至89.6%,表明网络重构与区域化空频对齐具有协同作用。**结论**　目前的实验结果支持SF-Match用于少量像素级标注、较多同域未标注图像且重视边界细节的自然场景和城市街景分割。其区域化空频对齐思想具有扩展至其他教师—学生式分割框架的潜力,但尚需进一步验证;对于高度遮挡、远距离小目标、复杂高频背景和跨领域数据,其有效性仍有限或未得到充分验证,且多视角训练会增加计算量与显存开销。

关键词:半监督学习;语义分割;空域-频域一致性;边界感知;视觉Transformer

Spatial-frequency consistency-driven semi-supervised semantic segmentation: SF-Match

Hu Jing, Lv Xiaopeng*, Wang Fang, Zhang Rui

School of Computer Science and Technology, Taiyuan University of Science and Technology, Taiyuan 030000, China

Abstract: **Objective** Semi-supervised semantic segmentation aims to learn robust pixel-level representations by leveraging a small amount of labeled data and a large amount of unlabeled data, thereby reducing the cost of acquiring high-quality pixel-level annotations. Although existing methods have achieved significant progress under the framework of consistency regularization, most of them still rely on traditional ResNet-based encoders, which limits their feature representation capa-

收稿日期:2026-04-29;**修回日期**:2026-08-02

*通信作者:吕晓鹏 lvxiaopeng@tyust.edu.cn

基金项目:山西省自然科学基金项目(项目编号:202203021211189);企业委托技术开发项目(项目编号:2025556);山西省研究生教育创新项目(项目编号:2025SZ13)

Supported by: Shanxi Provincial Natural Science Foundation (202203021211189); Enterprise-commissioned Technology Development Project (2025556); Shanxi Provincial Postgraduate Education Innovation Project (2025SZ13)

bility. Moreover, existing methods usually impose uniform consistency constraints on the entire image, ignoring the spatial and frequency-domain differences between semantic boundary regions and target interior regions. This may lead to blurred boundaries and inconsistent predictions within object regions. To address these issues, this paper proposes SF-Match, a semi-supervised semantic segmentation framework driven by spatial-frequency consistency. **Method** First, the traditional ResNet encoder is replaced with DINOv2, a Vision Transformer-based self-supervised pre-trained model. A DPT decoder is further introduced to reconstruct spatial feature maps from sequential representations, forming a dense prediction architecture that is more suitable for semi-supervised semantic segmentation. This architecture enhances the model's ability to jointly capture global semantic information and local structural details. Based on this architecture, a spatial-frequency consistency mechanism is designed. Specifically, strongly augmented views of unlabeled images are explicitly decomposed into high-frequency and low-frequency views. Meanwhile, morphological gradients are computed from the pseudo-labels generated by the teacher model to dynamically construct boundary-region masks and interior-region masks. High-frequency consistency constraints are applied to boundary regions to strengthen the modeling of object contours and fine-grained details, whereas low-frequency consistency constraints are applied to interior regions to improve structural coherence and class-level consistency. In this way, frequency-domain information and spatial regions are decoupled and aligned in a region-adaptive manner. **Result** Experimental results on two benchmark datasets, PASCAL VOC 2012 and Cityscapes, demonstrate the effectiveness of the proposed method. On the PASCAL VOC 2012 dataset, compared with the second-best method, semiVL, SF-Match improves the mean Intersection over Union (mIoU) by 1.9 to 3.6 percentage points under different labeled-data settings. Compared with the strong baseline UniMatch, SF-Match achieves consistent mIoU improvements across all partition settings, with a maximum gain of 11.2 percentage points. On the Cityscapes dataset, compared with the second-best method MPMC, SF-Match improves the mIoU by 5.1 and 3.7 percentage points under the 1/16 and 1/4 labeled-data settings, respectively. In particular, under the 1/16 labeled-data setting, SF-Match achieves an mIoU of 83.6%. Ablation experiments and qualitative analyses further demonstrate that the proposed method effectively improves complex boundary delineation and enhances prediction consistency within object regions. **Conclusion** The proposed SF-Match framework constructs a DINOv2-DPT dense prediction network and integrates a spatial-frequency consistency alignment strategy. By jointly enhancing the modeling of global semantic information and fine-grained structural details, SF-Match alleviates the limitations of insufficient feature representation and coarse consistency constraints in existing semi-supervised semantic segmentation methods. Experimental results show that the proposed method achieves superior segmentation performance on multiple benchmark datasets.

Key words: semi-supervised learning; semantic segmentation; spatial-frequency consistency; boundary awareness; Vision Transformer

论文引用格式:【DOI:10.11834/jig.260246】

0 引 言

半监督语义分割利用少量标记数据和大量未标记图像来学习鲁棒的像素级分类器,已成为解决密集预测任务中数据匮乏问题的关键方案。当前以UniMatch(YANG L 等,2023)及其变体为代表的最先进范式,主要依赖于一致性正则化(Consistency Regularization)。其核心思想直观且有效:强制模型在未标记图像的"强增强(Strongly Augmented)"视图上的预测结果,与"弱增强(Weakly Augmented)"视图生成的伪标签保持一致。这种"弱到强"的监督机制推动了该领域近年来的显著进步。国内相关综述也系统梳理了弱标注形式、代表性方法及大模型辅助等研究方向(项伟康 等,2024)。

尽管半监督语义分割的训练策略不断迭代,但其基础特征网络的设计仍相对滞后。现有方法大多沿用基于ImageNet-1K预训练的残差网络(residual network,ResNet)编码器(He 等,2016)。这类卷积骨干在多阶段下采样过程中容易削弱边界细节、小目标结构和局部纹理信息,从而限制一致性学习对细粒度特征的利用。随着视觉基础模型的发展,基于大规模自监督预训练的Vision Transformer为半监督语义分割提供了新的特征基础。因此,本文基于DINOv2(Oquab M 等,2024)的视觉表征能力,重新

设计基础特征网络，以获得更适合密集预测任务的细粒度视觉表征。初步实验结果（图1）所示，在仅使用有限标注数据的情况下，该设计已能够带来明显性能提升。此外，近期研究表明，轻量级CNN与Transformer的协同建模有助于兼顾局部细节提取与全局上下文建模（侯志强 等，2025）。

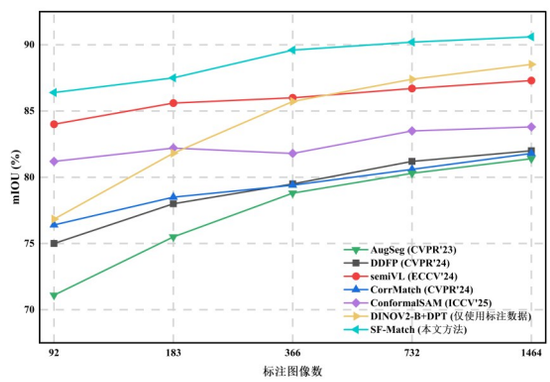


图1　在Pascal VOC 2012数据集上，SF-Match与其他SOTA方法的对比

Fig. 1　Comparison of SF-Match with other SOTA methods on the Pascal VOC 2012 dataset

然而，更强的特征表征也进一步凸显了现有一致性学习机制与精细特征之间的不匹配。DINOv2能够同时捕获目标主体的连贯语义结构和边界区域的局部细节，但传统“全局一致性”约束通常对整幅图像施加统一监督，将边界像素、目标内部像素和平坦背景区域等同处理。这种空间均匀的约束方式难以区分不同区域的特征需求，也无法充分利用DINOv2所提供的细粒度表征。例如，过强的统一约束可能削弱边界处本应保留的高频细节，也可能在较稳定的低频语义区域中引入不必要的扰动。因此，要进一步发挥基础视觉模型在半监督语义分割中的作用，亟需一种与特征相匹配、且能充分挖掘图像内在空频差异的精细化一致性策略。

为此，本文提出了SF-Match（Spatial-Frequency Match），旨在实现空域-频域一致性（Spatial-Frequency Consistency）。语义区域天然地与频率分量相对应：边界对应高频信号，而物体内部对应低频信号。这一特性为设计新方法提供了理论基础。

本文的主要贡献总结如下：

网络重构：基于DINOv2的自监督视觉表征，结合DPT解码器重构特征提取网络，构建了适应空频一致性约束的密集预测网络。该架构提升了全局语义建模能力，并通过高分辨率特征图保留了更多局部细节，增强了边界细节和复杂结构的捕捉能力，提供了更强的特征提取基础。

方法创新：提出了SF-Match框架，结合了空频一致性策略，采用联合解耦频域和空域的方法。通过将高频（细节增强）和低频（结构增强）特征与边界/内部区域精准对齐，解决了传统全局一致性方法中引起的模糊性问题。

实验验证：在Pascal VOC 2012和Cityscapes数据集上进行广泛实验，结果表明SF-Match在多个任务上取得了更优结果，有效缓解了边界混淆并显著提升了区域内部的一致性。

1　相关工作

1.1　半监督学习

半监督学习旨在通过利用少量标记数据和大量未标记数据来提升模型性能。早期的半监督学习方法主要基于熵最小化（Entropy Minimization）（GRANDVALET Y 等，2004）和伪标签（Proxy-label）方法（LEE D H 等，2013），通过鼓励模型对未标记数据做出低熵预测来隐式地推开决策边界。

近年来，一致性正则化已成为半监督学习的主流范式。其核心假设是模型对输入图像的微小扰动应具有鲁棒性。Mean Teacher（TARVAINEN A 等，2017）引入了教师-学生架构，通过对学生模型权重的指数移动平均（exponential moving average，EMA）来构建教师模型，并强制两者预测一致。MixMatch（BERTHELOT D 等，2019）和ReMixMatch（BERTHELOT D 等，2020）引入了MixUp（ZHANG 等，2018）数据增强策略，在图像层面混合标记和未标记样本以平滑标签分布。

而在当前半监督学习领域最具影响力的当属FixMatch（SOHN K 等，2020）。它提出了一种“弱到强（Weak-to-Strong）”的一致性机制：利用弱增强视图生成的伪标签来监督强增强视图的预测。基于FixMatch，后续工作进一步优化了伪标签的质量和利用率。例如，FlexMatch（ZHANG 等，2021）通过课程学习（Curriculum Learning）动态调整每个类别的置信度阈值；FreeMatch（WANG 等，2023）则提出了一种自适应阈值策略，根据模型的训练状态自动调

整置信度。此外，SimMatch（ZHENG 等，2022）和CoMatch（LI J 等，2021）将对比学习（Contrastive Learning）引入半监督学习，通过特征空间的对齐进一步增强了特征表示。尽管上述方法在图像分类任务上取得了巨大成功，但直接将其迁移到像素级预测任务（如语义分割）仍面临空间结构对齐的挑战。

1.2 半监督语义分割

半监督语义分割将半监督学习的思想扩展到了密集预测任务。现有的方法大致可以分为基于生成对抗网络（generative adversarial network，GAN）的方法和基于一致性正则化的方法。国内研究还在生成对抗框架中引入自注意力与谱归一化，以增强长程依赖建模并提升半监督训练稳定性（李志欣 等，2022）。

1.2.1 基于扰动与一致性的方法

目前半监督语义分割的最具代表性的先进（state of the art，SOTA）方法主要由一致性正则化主导。早期工作如交叉一致性训练（cross-consistency training，CCT）（OUALI Y 等，2020）通过对编码器输出添加扰动来强制一致性。交叉伪监督（cross pseudo supervision，CPS）（CHEN 等，2021）提出了交叉伪监督策略，使用两个具有不同初始化的网络互相监督。Adversarial Learning（HUNG W C 等，2018）则利用判别器来区分预测掩膜和真实标签，促进预测分布的对齐。

随着数据增强技术的发展，区域级混合（Region Mixing）被证明对半监督语义分割至关重要。CutMix（YUN 等，2019）被广泛用于在图像间剪切和粘贴区域以生成强增强样本。ClassMix（OLSSON V 等，2021）通过利用预测的语义掩膜来混合属于特定类别的像素，从而保留了物体的语义完整性。基于这些增强策略，ST++（YANG 等，2022）在伪标签生成的稳定性上做了深入探索。U2PL（WANG 等，2022）采用了直接丢弃低置信度像素的做法，利用负样本学习策略从未标记数据中挖掘更多信息。在此基础上，AugSeg（ZHAO 等，2023）指出单纯在未标记样本间进行混合可能因伪标签错误而引入确认偏差。为此，该方法提出了自适应标签辅助的CutMix，即根据未标记样本的置信度，自适应地将带有真实标签的样本混合到低置信度的未标记样本中，从而利用可靠的监督信息来稳定未标记数据的训练。UniMatch是当前的基准方法之一，它统一了图像级和特征级的一致性，利用双流扰动显著提升了性能。然而，包括UniMatch在内的大多数方法主要关注空间域的全局对齐，往往对整张图像施加统一的一致性约束，忽略了不同语义区域在特征分布上的内在差异。

1.2.2 频域感知与边界学习

图像的频域特征包含丰富的语义线索：低频分量通常反映图像的整体结构、主体形状和风格信息，而高频分量更多对应边缘、纹理及局部变化。在域适应领域，傅里叶域适应（Fourier domain adaptation，FDA）（Yang 等，2020）通过交换频谱低频分量缩小源域和目标域之间的外观差异。Zhang等（2023）针对少样本无监督域适应问题，提出基于频谱敏感性分析的谱对抗混合方法SAMix，通过生成目标域风格样本降低模型对域偏移的敏感性。Liu等（2023）则面向跨模态医学图像无监督域适应，通过频域与空间域知识迁移和多教师蒸馏学习结构相关表征。上述研究主要解决跨域分布差异问题，并非全监督条件下的边界细化方法。

在弱监督语义分割中，FFR（Yang 等，2025）指出 Vision Transformers（ViT）（DOSOVITSKIY A 等，2021）特征中的高频信息衰减可能造成分割结果过度平滑，并通过频率特征校正增强解码器对高频结构的学习能力。在半监督语义分割中，CFCG（Li 等，2023）利用交叉融合监督提高伪标签质量，并通过自适应轮廓引导模块识别不可靠空间区域，从而加强轮廓附近的监督。XNet v2（Zhou 等，2024）利用离散小波变换构建互补的高低频信息流，以增强不同频率特征之间的融合。

CorrMatch（Sun 等，2024）和ScaleMatch（Lv 等，2025）分别从相关性传播和多尺度一致性角度增强未标记数据监督；CFCG虽然显式引入了轮廓引导，但并未从频域角度对边界区域和目标内部施加差异化的一致性约束。本文提出的SF-Match进一步将高频视角与语义边界、低频视角与目标内部进行区域化对应，以实现空域区域和频域信息的解耦对齐。

1.2.3 基础模型在分割中的应用

长期以来，半监督语义分割领域一直依赖ResNet系列作为标准编码器。尽管ViT和SegFormer（XIE 等，2021）在全监督任务中已展现出超越卷积神经网络（convolutional neural networks，CNN）的性能，但在半监督领域，针对ViT的探索相

对滞后，仅有少量工作尝试将其引入半监督医学图像分割(He 等，2023)。

近年来，随着基础模型的爆发，研究者开始尝试引入更强大的先验知识来提升半监督分割性能。例如，SemiVL(HOYER L 等，2024)利用视觉-语言模型(如对比语言-图像预训练(contrastive language-image pre-training，CLIP))的强语义先验来指导半监督训练，有效缓解了标记数据受限时的类别混淆问题。AllSpark(WANG 等，2024)依托视觉 Transformer 架构，借助无标注视觉特征重构标注特征，依托视觉先验缓解标注数据主导训练的缺陷，提升半监督分割精度；而 ConformalSAM(CHEN 等，2025)则进一步探索了将“分割一切”模型(segment anything model，SAM)应用于半监督语义分割，通过共形预测校准 SAM 的不确定性，从而生成高质量的伪标签监督。

然而，尽管这些利用多模态或特定分割大模型的方法表现优异，纯视觉自监督基础模型在半监督语义分割中的潜力仍有进一步挖掘空间。DINOv2 通过大规模自监督学习获得了较强的视觉表征能力，能够同时提取语义信息与局部细节，并在密集预测任务中表现出良好的迁移潜力。但现有的粗糙一致性损失往往无法充分利用 DINOv2 提供的精细特征。因此，本文以 DINOv2 为特征基础，结合密集预测解码结构，并设计空频一致性策略，进一步释放其在半监督语义分割中的潜力。

2 研究方法

2.1 总体框架

半监督语义分割的核心目标在于通过挖掘海量未标记数据的内在结构信息，来弥补标记数据匮乏带来的监督信号缺失。主流方法主要遵循“弱增强到强增强”的一致性正则化范式，即利用弱增强视图生成的伪标签来监督强增强视图的预测。然而，SOTA 方法在面对复杂场景时仍面临两个核心瓶颈：

1. 传统网络的特征表达局限：大多数方法仍沿用过时的 ResNet 作为编码器，难以捕捉长距离语义依赖，且缺乏大规模预训练带来的通用视觉特征。

2. 空间-频域的均一化处理：传统的一致性约束对整张图像的所有像素和所有频段施加等强度约束。然而，图像的高频分量(对应边缘、纹理)和低频分量(对应结构、主体)在语义理解中扮演着截然不同的角色，且容易受到不同类型的噪声干扰。

针对上述问题，本文提出了一种改进的半监督分割框架——SF-Match。该框架在 FixMatch 的强弱一致性策略基础上进行了两项关键改进：首先，本文以 DINOv2 为特征基础，重构密集预测网络架构，增强模型对全局语义与细粒度结构的表达能力；其次，设计了空频联合解耦模块(Spatial-Frequency Decoupling Module)，通过显式的频域变换和形态学空间划分，实现了“高频关注边界、低频关注内部”的分治对齐策略。SF-Match 模型整体网络架构，如图 2 所示。

2.2 问题定义与基线回顾

2.2.1 半监督分割设定

假设给定一个包含 N_l 个样本的标记数据集 $\boldsymbol{D}_l=\{(\mathbf{x}_i,\mathbf{y}_i)\}_{i=1}^{N_l}$，其中 $x_i\in \mathrm{R}^{H\times W\times 3}$ 表示输入图像，$y_i\in\{0,1\}^{H\times W\times C}$ 表示对应的像素级真实标签(One-hot 编码)，C 为语义类别数；一个规模远大于标记集的未标记数据集 $\boldsymbol{D}_u=\{\mathbf{u}_i\}_{i=1}^{N_u}$，通常 $N_u\gg N_l$。

模型的训练目标是学习一个映射函数 f_θ: $\mathbf{X}\rightarrow\mathbf{Y}$，使得在测试集上的分割性能最优。总的优化目标函数 L 通常由有监督损失 L_s 和无监督损失 L_u 组成：

$$L=L_s+\lambda_u L_u \tag{1}$$

式中，L 为总损失；L_s 为有监督损失；L_u 为无监督损失；λ_u 为无监督损失权重。

2.2.2 基础一致性范式

SF-Match 以 Teacher-Student 一致性学习框架为基础，并在此范式上构建后续的空频一致性约束。具体而言，学生模型(Student)负责通过反向传播进行参数更新，教师模型(Teacher)则由学生模型参数的指数移动平均生成，以提供更稳定的监督信号。学生模型(参数为 $\boldsymbol{\theta}_s$)通过反向传播更新，教师模型(参数为 $\boldsymbol{\theta}_t$)则是学生模型的指数移动平均(EMA)：

$$\boldsymbol{\theta}_t\leftarrow\alpha\boldsymbol{\theta}_t+(1-\alpha)\boldsymbol{\theta}_s \tag{2}$$

式中，$\boldsymbol{\theta}_t$ 和 $\boldsymbol{\theta}_s$ 分别为教师模型参数向量和学生模型参数向量；α 为指数移动平均的动量系数；箭头表示参数赋值更新，这种更新策略使得教师模型能够提供更稳定、更平滑的预测结果。

对于未标记图像 u_l，分别应用弱增强 A^w(如随机翻转、裁剪)和强增强 A^s(如色彩抖动、CutMix、Class-

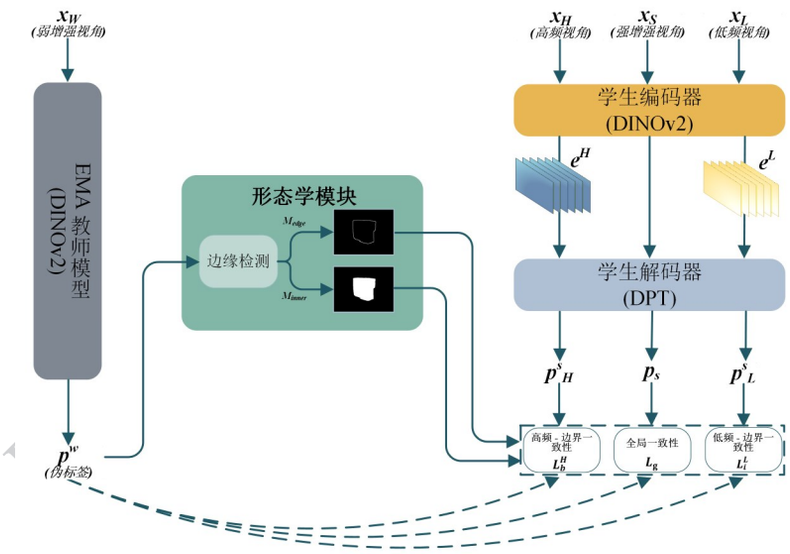


图2　SF-Match模型整体网络架构

Fig. 2　The overall network architecture of the SF-Match model

Mix)。教师模型对弱增强图像生成预测概率分布，并利用置信度阈值筛选出高质量的伪标签。学生模型则需要对强增强图像做出预测，并最小化其与伪标签之间的交叉熵损失：

$$L_{\mathrm{u}}^{base}=|\boldsymbol{\Omega}|^{-1}\sum_{j\in\boldsymbol{\Omega}}q_j\,l_{\mathrm{c}}\left(\mathbf{p}_j^{\mathrm{s}},\hat{\mathbf{y}}_j\right),\quad q_j=I\left(\max_{1\leqslant r\leqslant C}\mathbf{p}_{j,r}^{\mathrm{w}}>\tau\right)\tag{3}$$

式中，L_{u}^{base}为基线无监督损失；$\boldsymbol{\Omega}$为全部像素位置构成的集合；$|\boldsymbol{\Omega}|$为像素位置总数；j为像素位置索引；$q_j\in\{0,1\}$为第j个像素的置信度筛选量；I为示性函数；max为最大值函数；r为类别索引；C为语义类别数；$\mathbf{p}_{j,r}^{\mathrm{w}}$为教师模型在弱增强视图第$j$个像素处对第$r$类的预测概率；$\tau$为置信度阈值；$\mathbf{p}_j^{\mathrm{s}}$为学生模型在强增强视图第$j$个像素处的预测概率向量；$\hat{\mathbf{y}}_j$为第$j$个像素的伪标签；$l_{\mathrm{c}}$为像素级交叉熵损失函数；上标w和s分别表示弱增强视图和强增强视图。

这一基线构成了SF-Match的核心基础，但其仍面临两大显著局限：一是未能充分挖掘图像内在的频域结构信息；二是受限于传统ResNet编码器有限的感受野与特征建模能力。

2.3　面向空频一致性对齐的特征网络重构

针对上述底层特征提取的瓶颈，首先对基础网络进行了重构。现有的半监督分割框架普遍受限于传统特征网络在下采样过程中的信息衰减，难以支撑细粒度的空频特征提取。为此，SF-Match构建了一种面向空频一致性对齐的密集预测网络架构。

该架构基于预训练的DINOv2模型。相比于传统网络，它能够提供全局感受野并避免深层特征的过度平滑。然而，DINOv2原始输出的离散序列嵌入无法直接应用于像素级的频域解耦。为了弥合这一表征缺陷，SF-Match在解码侧引入了DPT解码器，利用其独特的序列到空间(Sequence-to-Space)重塑能力，通过渐进式的特征融合与上采样，重构出极具判别性的高分辨率空间特征图。

DINOv2特征编码与DPT空间重建的结合，是SF-Match实现空频一致性对齐的核心前提。它不仅完整地保留了用于指导边界对齐的高频线索，还稳固了用于维持主体连贯性的低频语义结构。

此外，在训练策略上，该架构支持对DINOv2骨干网络进行冻结(Lock Backbone)，从而在显著降低显存开销的条件下，仍能保持优异的分割表现。

2.4　空频一致性框架

基于上述特征网络，针对未标记数据中边界细节与主体语义利用不足的问题，SF-Match创新性地提出了一种空频联合解耦机制。

2.4.1　频域解耦视角生成

对于经过CutMix处理后的强增强图像u_s，SF-Match构造两个频域增强视角：低频视角u^L和高频视角u^H。

1 低频分量提取

低频分量$\mathbf{I}_{\mathrm{L}}$代表图像的整体结构、光照分布和主体形状。为有效提取这一基础特征，本文采用核大小为s_{g}、标准差为σ的高斯低通滤波器对图像进行平滑处理：

$$\boldsymbol{I}_L = \boldsymbol{u}^s * \mathbf{G}\left(s_{\mathrm{g}}, \sigma\right) \tag{4}$$

式中，$\mathbf{G}\left(x, y\right) = \dfrac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$，$\boldsymbol{u}^s$为强增强图像。

低频增强视角$\mathbf{u}^{\mathrm{L}}$的构造方式为直接使用提取的低频分量（模拟失焦、运动模糊场景），迫使模型不依赖纹理细节，而是通过上下文结构来识别$\mathbf{I}_{\mathrm{L}} = \mathbf{u}^{\mathrm{s}} * \mathbf{G}\left(s_{\mathrm{g}}, \sigma\right)$物体：

$$\boldsymbol{u}^L = \alpha_L \boldsymbol{I}_L \tag{5}$$

式中，$\boldsymbol{u}^L$为低频增强图像；α_L为低频视角的强度系数；$\boldsymbol{I}_L$为式（4）得到的低频图像。

2 高频分量提取

高频分量$\boldsymbol{I}_H$对应图像中的剧烈变化区域，如物体轮廓、纹理纹路和噪点。本文采用拉普拉斯算子（Laplacian Operator）进行高效提取：

$$\boldsymbol{K}_L = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} \tag{6}$$

式中，$\boldsymbol{K}_L$为3 × 3拉普拉斯卷积核矩阵；L表示“拉普拉斯”。

$$\boldsymbol{I}_H = \boldsymbol{u}^s * \boldsymbol{K}_L \tag{7}$$

考虑到纯高频响应通常只包含稀疏的边缘与纹理变化，直接作为输入可能导致语义信息缺失，因此将高频分量作为残差扰动叠加到原始强增强视角上，得到高频增强视角$\boldsymbol{u}^H$：

$$\boldsymbol{u}^H = \boldsymbol{u}^s + \alpha_H \boldsymbol{I}_H \tag{8}$$

式中，$\boldsymbol{u}^H$为高频增强图像，α_{H}为高频视角的强度系数。

通过以上方式，得到了两个互补的增强视角：u^L强调“结构一致性”，$\boldsymbol{u}^H$强调“细节鲁棒性”。

需要指出的是，高频视角与低频视角均是在输入空间中构造的确定性增强视图，而不是额外引入的独立网络分支。三类输入视角$\boldsymbol{u}^s$、$\boldsymbol{u}^H$和$\boldsymbol{u}^L$均送入同一个学生模型，共享同一套编码器-解码器参数。因此，高频与低频信息对模型的影响并非停留在输入层面，而是通过后续区域化一致性损失产生梯度，并共同回传至共享特征空间。这使得模型能够在统一表征中同时学习基础语义、边界细节和区域结构，而不是将频率信息割裂为互不相关的分支。

2.4 2基于形态学的空间区域解耦

仅构造高频和低频视角仍不足以保证有效监督。如果在平坦的内部区域强制模型拟合高频噪声，会导致预测出现孔洞；如果在边缘区域强制模型拟合模糊的低频信息，会导致边界平滑化。因此，需要将频域特征与空间区域进行对齐，使不同频率视角只在适合的区域内产生有效梯度。

SF-Match利用教师模型生成的伪标签$\hat{y}$对空间区域进行动态划分。由于$\hat{y}$是离散的类别索引图，利用形态学梯度（Morphological Gradient）来定位语义边界。

定义形态学膨胀（Dilation，⊕）和腐蚀（Erosion，⊖）操作。通过最大池化（Max Pooling）高效实现：

$$\text{膨胀：}\boldsymbol{D} = \hat{\boldsymbol{y}} \oplus \boldsymbol{B}_k \tag{9}$$

$$\text{腐蚀：}\boldsymbol{E} = \hat{\boldsymbol{y}} \ominus \boldsymbol{B}_k \tag{10}$$

式中，D为伪标签数组的膨胀结果；$\boldsymbol{E}$为伪标签数组的腐蚀结果；$\hat{\boldsymbol{y}}$为伪标签数组；$\boldsymbol{B}_k$为大小为$k \times k$的结构元素数组；k为形态学操作核大小，即局部邻域范围。对于一个像素点，若其膨胀结果不等于腐蚀结果，说明该像素的邻域内存在不同的类别标签，即该像素位于语义边界上。

根据上述原理，生成二值掩膜：

边界掩膜$\boldsymbol{M}_b$定义为膨胀结果与腐蚀结果不一致的像素集合。该掩膜$\boldsymbol{M}_b \in \{0,1\}^{H \times W}$覆盖了所有语义类别之间的交界区域：

$$\boldsymbol{M}_b = I\left(\boldsymbol{D} \neq \boldsymbol{E}\right) \tag{11}$$

式中，$\boldsymbol{M}_b \in \{0,1\}^{H \times W}$为边界二值掩膜数组；$I$为示性函数；$H$和$W$分别为图像高度和图像宽度；其边界宽度$boundary_width$由形态学核大小$k$控制：

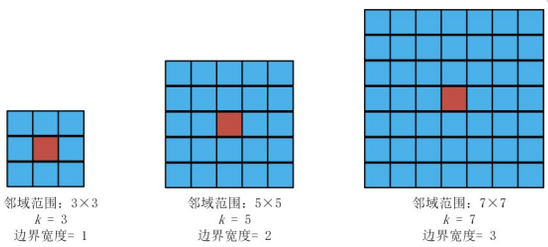
内部掩膜$\boldsymbol{M}_i$定义为边界区域之外的像素集合，即对边界掩膜取反得到。该掩膜$\boldsymbol{M}_i \in \{0,1\}^{H \times W}$覆

图3　参数关系示意图

Fig. 3　Schematic diagram of parameter relationships

$$boundary\_width = \frac{k-1}{2} \tag{12}$$

参数关系示意图，如图3所示。



盖了所有语义类别内部相对平坦且远离类别交界的区域，可用于约束模型更加关注非边界区域的稳定语义表征：

$$\boldsymbol{M}_i = 1 - \boldsymbol{M}_b \tag{13}$$

2.4　3区域感知的一致性对齐与梯度约束

在得到三类输入视角和两类空间掩膜后，SF-Match构建区域感知的空频一致性损失。该损失由RGB全局一致性损失、高频—边界一致性损失和低频—内部一致性损失三部分组成：

$$L_u = L_g + \lambda_b L_b^H + \lambda_i L_i^L \tag{14}$$

式中，L_g、L_b^H和L_i^L分别为全局一致性损失、高频—边界一致性损失和低频—内部一致性损失；λ_b和λ_i分别为边界损失与内部损失的权重；

RGB全局一致性损失作用于原始强增强视角，用于保持学生模型与教师伪标签之间的基础语义一致性：

$$\begin{aligned} L_g &= Q_g^{-1}\sum_{j\in\boldsymbol{\Omega}} q_j\, l_c\left(\boldsymbol{p}_j^s, \hat{\boldsymbol{y}}_j\right), \\ Q_g &= \sum_{j\in\boldsymbol{\Omega}} q_j \end{aligned} \tag{15}$$

式中，Q_g为参与全局一致性损失计算的有效像素数；$\boldsymbol{\Omega}$为全部像素位置集合；j为像素位置索引；q_j为置信度筛选量；l_c为交叉熵损失函数；$\mathbf{p}_j^s$为强增强视图第j个像素处的预测概率向量；$\hat{\mathbf{y}}_j$为伪标签。

高频-边界对齐损失L_b^H用于约束高频视角$\boldsymbol{u}^H$下的学生预测p^H。仅在边界掩膜$\boldsymbol{M}_b$覆盖的区域内计算一致性损失：

$$\begin{aligned} L_b^H &= Q_b^{-1}\sum_{j\in\boldsymbol{\Omega}} q_j M_{b,j}\, l_c\left(\boldsymbol{p}_j^H, \hat{\boldsymbol{y}}_j\right), \\ Q_b &= \sum_{j\in\boldsymbol{\Omega}} q_j M_{b,j} \end{aligned} \tag{16}$$

式中，L_b^H为高频—边界一致性损失；Q_b为参与计算的有效边界像素数；$\boldsymbol{\Omega}$为全部像素位置集合；q_j为置信度筛选量；$M_{b,j}$为边界掩膜在第j个像素位置处的取值；$\mathbf{p}_j^{\mathrm{H}}$为高频视角第$j$个像素处的预测概率向量。

该项损失使高频视角产生的梯度主要集中在语义边界附近，从而强化共享特征空间中对轮廓变化、类别交界和局部细节的响应。由于p^H与RGB视角预测共享同一网络参数，该梯度不仅更新高频输入下的预测能力，也会反向影响编码器和解码器中的通用边界表征。

低频-内部对齐损失L_i^L用于约束低频视角u^L下的学生预测p^L。仅在内部掩膜M_i覆盖的区域内计算一致性损失：

$$\begin{aligned} L_i^L &= Q_i^{-1}\sum_{j\in\boldsymbol{\Omega}} q_j M_{i,j}\, l_c\left(\boldsymbol{p}_j^L, \hat{\boldsymbol{y}}_j\right), \\ Q_i &= \sum_{j\in\boldsymbol{\Omega}} q_j M_{i,j} \end{aligned} \tag{17}$$

式中，Q_i为参与计算的有效内部像素数；$M_{i,j}$为内部掩膜在第j个像素位置处的取值；$\mathbf{p}_j^{\mathrm{L}}$为低频视角第$j$个像素处的预测概率向量。

该项损失使低频视角产生的梯度主要作用于远离类别交界的区域，促使模型在纹理弱化的情况下仍能保持类内预测一致性和主体结构连贯性。该梯度同样回传至共享参数，模型可以在同一特征空间中增强对低频结构和上下文语义的建模能力。

SF-Match的最终优化目标定义为：

$$L = L_l + \lambda_u L_u \tag{18}$$

式中，L为总损失；L_l为有监督损失；λ_u为无监督一致性损失的总权重；L_u为无监督损失。

通过上述设计，SF-Match构建了完整的空频一致性框架，将频域视角解耦、空间区域划分和共享特征梯度更新统一起来，实现了“高频约束边界、低频稳定内部、RGB保持语义”的区域感知空频一致性学习。空频一致性框架图，如图4所示。

2.5　总体优化与算法描述

考虑到语义分割任务中常存在类别分布不均衡、易分类像素占比较高等问题，本文引入在线困难样本挖掘策略，仅选择损失值较高的困难像素参与反向传播，从而提升模型对少数类、边界区域和复杂区域的学习能力。

具体而言，对于有标签图像中的每个有效像素，首先计算其像素级交叉熵损失，并忽略标注为ignore的无效像素。随后，根据模型对真实类别的

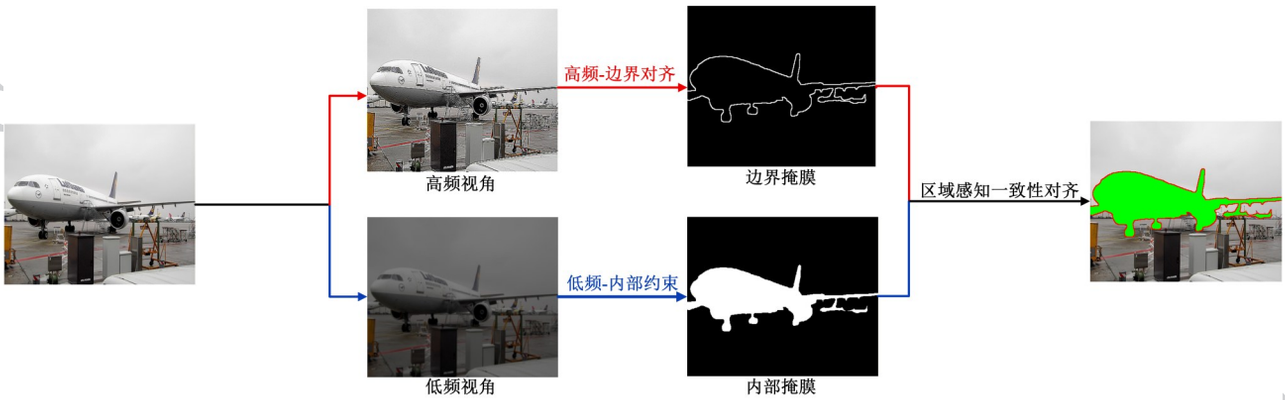


图4　空频一致性框架图

Fig. 4　Spatial-frequency decoupling mechanism diagram

预测置信度，或等价地根据像素级交叉熵损失，对所有有效像素进行排序。预测置信度较低、交叉熵损失较高的像素被视为困难像素（Hard Pixels）。训练时，仅保留满足困难样本阈值条件的像素参与监督损失计算；当满足阈值条件的像素数量不足时，进一步选取损失值最高的前若干个像素，以保证每次迭代中具有足够的有效监督信号。其监督损失定义为：

$$L_l = N_h^{-1}\sum_{i=1}^{N_l}\sum_{j\in\Omega_h} l_c\left(\boldsymbol{p}_{i,j},\boldsymbol{y}_{i,j}\right),$$
$$\boldsymbol{p}_{i,j} = f_{\boldsymbol{\theta}_s}\left(\boldsymbol{x}_i\right)_j \tag{19}$$

式中，N_h为参与损失计算的困难像素总数；N_l为标记样本数；i为样本索引；$\boldsymbol{\Omega}_h$为困难像素位置集合；j为像素位置索引；l_c为交叉熵损失函数；$\mathbf{p}_{i,j}$为学生模型对第i幅标记图像在第j个像素处的预测概率向量；$\mathbf{y}_{i,j}$为对应的真实标签向量；f为分割映射函数；$\boldsymbol{\theta}_s$为学生模型参数向量；$\mathbf{x}_i$为第i幅标记图像；下标h表示“困难像素”。

为保证CutMix操作后输入图像与伪标签等监督信号在像素空间严格对齐，本文对强增强图像、伪标签、置信度图及忽略掩膜执行同步CutMix处理。

对批内第i个未标注样本，先由教师模型在弱增强视图上生成伪标签与对应置信度图；再生成二值CutMix掩膜，以掩膜取值控制像素来源：取值为1的位置由批内配对样本替换，取值为0的位置保留当前样本像素。由于采用二值掩膜进行像素级区域替换（而非标签插值或概率加权），混合后的图像、伪标签、置信度图与忽略掩膜在空间位置上完全一致，确保监督信号与输入图像逐像素对应。

后续基于混合后的置信度图与忽略掩膜筛选有效监督像素，并以混合伪标签为基础生成边界掩膜与内部掩膜。

SF-Match的总体训练流程如算法1所示。每次迭代中，模型同时利用有标签数据和无标签数据进行联合训练。对于无标签数据，教师模型首先生成高置信度伪标签，并结合CutMix增强、形态学区域掩膜以及高低频视角解耦，计算区域感知的空频一致性损失。最终，模型联合优化有监督损失与无监督一致性损失，并通过EMA策略更新教师模型。

至此，SF-Match形成了完整的空频一致性训练框架。以DINOv2的全局语义表征和DPT的空间重建能力为基础，构建了更适合密集预测任务的特征网络，并结合高低频视角解耦与边界内部掩膜划分机制，实现了高频关注边界、低频连贯内部的分治学习，克服了传统一致性算法全局约束的局限，显著提升模型对复杂语义区域的精细化感知能力。

3　实验与分析

本章将对SF-Match进行实验验证与深入分析。首先，介绍实验所使用的数据集、评估指标以及具体的网络训练和实施细节；其次，在两个主流的半监督语义分割基准数据集（Pascal VOC 2012和Cityscapes）上，将SF-Match与SOTA方法进行性能对比；随后，通过一系列消融实验，剥离并验证SF-Match中各个核心创新组件（如空域解耦、形态学空间掩膜等）的有效性及超参数的敏感性，以及评估预训练DINOv2编码器在微调（默认）与冻结机制下的性能差异与模型复杂度分析；最后，通过定性可视化

算法1.SF-Match总体训练算法

输入:有标签数据集D_l,无标签数据集D_u,最大训练轮次max_epochs,置信度阈值τ
输出:训练好的语义分割模型f

1) 初始化语义分割模型参数
2) $epochs \leftarrow 0$
3) **WHILE** $epochs < max\_epochs$ **DO**
(1) **FOR** $(X_l,Y_l),(X_w,X_s)$ in $zip(D_l,D_u)$ **DO**
① 计算有标签数据(X_l,Y_l)的监督损失
② 教师模型预测X_w,生成伪标签$\hat{Y}^w$与高置信度掩膜M_c
③ 生成二值CutMix掩膜M^{cm},对强增强图像X_s、伪标签$\hat{Y}^w$、高置信度掩膜M_c和忽略掩膜I执行同步像素级区域替换,得到$\tilde{X}_s$、$\tilde{Y}$、$\tilde{C}$和$\tilde{I}$
④ 根据$\tilde{C} \geqslant \tau$且$\tilde{I} \neq 255$构造有效监督掩膜$\tilde{M}_c$
⑤ 基于混合后的伪标签$\tilde{Y}$提取形态学边界掩膜M_{edge}与内部掩膜M_{inner}
⑥ 对混合后的强增强图像$\tilde{X}_s$解耦出高频视角X_H(锐化)与低频视角X_L(失焦)
⑦ 学生模型预测各视角,结合掩膜计算空频一致性损失L_u
⑧ 计算总损失$L = L_s + \lambda_u L_u$
⑨ 反向传播最小化L,更新学生模型f
⑩ 通过EMA策略更新教师模型f_{ema}
(2) **END FOR**
(3) $epochs \leftarrow epochs + 1$
4) **END WHILE**

分析,直观展示本方法在改善边界模糊和提升内部连贯性方面的表现。

3.1 实验设置

3.1.1 数据集

Pascal VOC 2012:该数据集包含20个前景类别和1个背景类别。在标准的半监督设定中,该数据集由两部分组成:经典的高质量标注训练集(包含1,464张图像)和验证集(包含1,449张图像)。

Cityscapes:该数据集专注于城市街道场景的语义理解,包含19个用于评估的语义类别。相比于Pascal VOC,Cityscapes具有更高的图像分辨率(通常为1024×2048),且场景中包含大量密集分布的小目标(如行人、交通标志)以及复杂的精细边缘(如电线杆、自行车)。其标准划分包含2,975张训练图像和500张验证图像。

3.1.2 评估指标

语义分割的本质是像素级的多分类任务。本文采用该领域最主流的评价指标——平均交并比(mean intersection over union,mIoU)来量化模型的分割精度。mIoU越高,表明模型在各个类别上的综合分割性能越好。

3.1.3 实验环境与参数

实验均基于PyTorch深度学习框架实现,并采用重构的DINOv2-B与DPT结合的密集预测网络作为统一分割架构。模型优化采用AdamW优化器(权重衰减设为0.01)。为防止端到端微调破坏基础模型的先验特征,引入分层学习率策略,将骨干网络的学习率设定为解码器头部的0.1倍,全局基础学习率依多项式衰减策略(Poly LR)进行退火:

$$\eta_t = \eta_0\left(1 - \frac{t}{T}\right)^{0.9} \tag{20}$$

式中,η_t为第t次迭代的学习率;η_0为初始学习率;t为当前迭代次数;T为总迭代次数;指数0.9为多项式衰减指数。

所有实验均在4张32GB vGPU上通过分布式数据并行(distributed data parallel,DDP)与同步批量归一化(synchronized batch normalization,SyncBN)完成,全局批次大小设为4。Pascal VOC 2012与Cityscapes数据集的训练周期分别设定为60和180个Epoch。

在SF-Match的具体实现中,为促使模型学习对强增强伪影的空间不变性,频域视角的解耦操作被严格置于CutMix混合之后。频域提取方面,低频分支采用核大小为$s_g = 5$、标准差为$\sigma = 1.0$的高斯低通滤波器,高频分支采用3×3的拉普拉斯卷积核,两者的残差混合系数α均设定为0.5。空域解耦方面,形态学梯度计算的领域范围设为3×3($edge\_width=1$),以精细提取边界并防止掩膜过度侵占主体内部。此外,伪标签置信度阈值τ设为0.95,无监督总损失权重λ_u设定为1.0,其中高频和低频分支的损失系数λ_H和λ_L默认设置为0.5。

3.2 对比实验

为了全面验证SF-Match的有效性,将其与当前主流的半监督分割算法进行了对比,包括ST++、AugSeg、iMAS、DAW以及强基线方法UniMatch等。实验分别在Pascal VOC 2012和Cityscapes两个公开数据集上进行。为分析模型在小目标、细长结构和复杂边界类别上的表现,本文还在Pascal VOC 2012数据集上开展困难类别性能对比。

3.2.1 主流数据集整体性能对比

首先,在Pascal VOC 2012数据集上,对比了SF-

表1　不同含标签比例的分区协议下Pascal VOC 2012数据集的性能对比(%)

Table 1　Comparison with state-of-the-art methods on Pascal(%)

方法	发表会议/期刊	编码器	有标注图像比例及对应数量				
			1/16(92)	1/8(183)	1/4(366)	1/2(732)	全标注(1464)
Labeled Only	–	RN-101	45.1	55.3	64.8	69.7	73.5
ST++	CVPR′22	RN-101	65.2	71.0	74.6	77.3	79.1
PS-MT	CVPR′22	RN-101	65.8	69.6	76.6	78.4	80.0
GTA-Seg	NeurIPS′22	RN-101	70.0	73.2	75.6	78.4	80.5
PCR	NeurIPS′22	RN-101	70.1	74.7	77.2	78.5	80.7
iMAS	CVPR′23	RN-101	68.8	74.4	78.5	79.5	81.2
UniMatch	CVPR′23	RN-101	75.2	77.2	78.8	79.9	81.2
AugSeg	CVPR′23	RN-101	71.1	75.5	78.8	80.3	81.4
CCVC	CVPR′23	RN-101	70.2	74.4	77.4	79.1	80.5
DAW	NeurIPS′23	RN-101	74.8	77.4	79.5	80.6	81.5
CorrMatch	CVPR′24	RN-101	76.4	78.5	79.4	80.6	81.8
DDFP	CVPR′24	RN-101	75.0	78.0	79.5	81.2	82.0
MPMC	ECCV′24	RN-101	77.3	78.6	79.8	80.8	81.7
semiVL	ECCV′24	CLIP-B	84.0	85.6	86.0	86.7	87.3
SWSEG	CVPR′25	RN-101	79.0	79.3	79.9	–	–
ConformalSAM	ICCV′25	SegFormer-B5	81.2	82.2	81.8	83.5	83.8
Labeled Only	–	DINOv2-B	76.8	81.8	85.7	87.4	88.5
SF-Match(本文)	–		**86.4**	**87.5**	**89.6**	**90.2**	**90.6**

注：性能最优的结果以粗体标注，次优结果以下划线标注，数据缺失以 – 标注。

Match与其他先进半监督语义分割方法的性能表现，具体定量结果如表1所示。与传统的强基线方法UniMatch相比，SF-Match在多个分区协议下实现了显著的性能提升：在1/16、1/8、1/4、1/2和Full分区下，分别领先11.2%、10.3%、10.8%、10.3%和9.4%。尽管与当前引入了多模态或强大先验的最新SOTA方法(如基于CLIP-B架构的semiVL)相比，SF-Match依然保持了明显的优势：在资源最少的1/16分区，SF-Match比semiVL高出2.4%；在数据量逐渐增加的1/4、1/2和Full分区下，领先幅度进一步扩大至3.6%、3.5%和3.3%。这一实验结果有力地验证了：通过重构DINOv2+DPT密集预测网络，并与空频解耦一致性机制相结合，SF-Match能在标签数据稀缺的情况下有效提取并对齐关键的边缘和结构信息，从而缓解半监督学习中的冷启动问题；也成功突破了传统网络架构的性能瓶颈，在标签数据充足时持续释放出卓越的特征表征潜力。

表2对比了SF-Match与现有主流模型在Cityscapes数据集上的性能差异。数据显示，SF-Match在各项分区协议下均实现了对现有方法的超越。在仅含186个标记样本的极低数据域(1/16分区)下，SF-Match取得了83.6%的优异表现，较当前SOTA方法MPMC提升5.1%。充分印证了空频联合解耦机制在标签稀缺场景下对复杂语义边界的精准定位能力。此外，在标签规模相对充足的1/4分区协议中，SF-Match依然保持显著优势，其mIoU分别领先次优模型MPMC 3.7%与经典强基线UniMatch 5.4%，展现出极具竞争力的表征鲁棒性。

3.2.2　困难类别性能对比

仅使用mIoU作为整体平均指标，难以充分反映模型在小目标、细长结构和边界复杂类别上的表现。为进一步验证SF-Match对困难类别的分割能力，本文在Pascal VOC 2012数据集1/4标注划分下选取bicycle、bottle、chair、dining table、motorbike和potted

表2　Cityscapes数据集不同含标签比例的分区协议下分割性能对比(%)

Table 2　Comparison with state-of-the-art methods on Cityscapes(%)

方法	发表会议/期刊	编码器	有标注图像比例及对应数量			
			1/16(186)	1/8(372)	1/4(744)	1/2(1488)
Labeled Only	–	RN-101	66.3	72.8	75.0	78.0
U^{2}PL	CVPR′22	RN-101	74.9	76.5	78.5	79.1
PS-MT	CVPR′22	RN-101	–	76.9	77.6	79.1
GTA-Seg	NeurIPS′22	RN-101	69.4	72.0	76.1	–
PCR	NeurIPS′22	RN-101	73.4	76.3	78.4	79.1
iMAS	CVPR′23	RN-101	74.3	77.4	78.1	79.3
UniMatch	CVPR′23	RN-101	76.6	77.9	79.2	79.5
AugSeg	CVPR′23	RN-101	75.2	77.8	79.6	80.4
CCVC	CVPR′23	RN-101	74.9	76.4	77.3	–
DAW	NeurIPS′23	RN-101	76.6	78.4	79.8	80.6
CorrMatch	CVPR′24	RN-101	77.3	78.5	79.4	80.4
DDFP	CVPR′24	RN-101	77.1	78.2	79.9	80.8
MPMC	ECCV′24	RN-101	<u>78.5</u>	79.2	<u>80.9</u>	<u>81.3</u>
semiVL	ECCV′24	CLIP-B	77.9	<u>79.4</u>	80.3	80.6
SWSEG	CVPR′25	RN-101	76.8	78.6	79.5	–
CSL	ICCV′25	RN-101	78.2	78.8	80.0	81.1
Labeled Only	–	DINOv2-B	81.0	82.5	83.8	84.5
SF-Match(本文)	–		**83.6**	**84.4**	**84.6**	**85.0**

注:性能最优的结果以粗体标注,次优结果以下划线标注,数据缺失以 – 标注。

plant等具有明显细粒度结构或复杂边界的类别进行单独统计,并计算这些类别的平均IoU,记为Hard-Avg。实验结果如表3所示。

表3　困难类别分割性能对比(%)

Table 3　Comparison of segmentation performance on difficult classes (%)

方法	自行车	瓶子	椅子	餐桌	摩托车	盆栽植物	困难类别平均IoU	mIoU
UniMatch	70.63	79.64	26.31	67.62	83.63	57.39	64.20	78.8
Augseg	72.07	76.92	33.75	66.37	84.98	57.14	65.21	78.8
DDFP	71.12	77.42	36.01	64.97	79.85	55.31	64.11	79.5
semiVL	78.30	85.97	44.23	74.63	88.95	**74.70**	74.46	86.0
SF-Match(本文)	**84.30**	**88.31**	**69.62**	**77.07**	**94.22**	<u>73.29</u>	**81.14**	**89.6**

注:性能最优的结果以粗体标注,次优结果以下划线标注。

由表3可见,SF-Match不仅取得了89.6%的整体mIoU,在困难类别上也表现出较强的分割能力。其中,bicycle和motorbike包含车轮、车架等细长结构,bottle和potted plant往往具有较小目标尺度或细碎边界,chair与dining table则容易受到遮挡和室内复杂背景干扰。SF-Match在这些类别上的Hard-Avg达到81.14%,说明该方法并非仅依赖大目标或易分类类别提升整体mIoU,而是在小目标、细粒度结构

和复杂边界类别上同样具有较好的鲁棒性。这主要得益于DINOv2+DPT对全局语义和空间细节的联合建模能力，以及空频一致性策略对区域结构信息的约束，使模型能够更有效地保持目标内部一致性并增强边界区域的判别能力。

3.3 消融实验

在本节中，将分析SF-Match中不同设计模块的有效性以及多分支网络的性能变化。消融实验均在经典的Pascal VOC 2012数据集上进行。

3.3.1 关键模块有效性分析

本小节探究了网络重构与空频模块对模型性能的影响，分析结果展示在表4中。“高频视角”是指提取图像纹理和边缘信息的增强分支，“低频视角”是指保留图像主体与结构信息的增强分支，而“掩膜对齐”是指利用形态学梯度动态将频域特征与空间区域进行分治匹配的模块。此外，在表5中还探究了高频分量提取采用不同滤波策略以及邻域范围对模型性能的影响。

在表4的成分消融中，方案A采用传统ResNet-101网络，且不引入高低频视角和空间掩膜对齐模块，仅使用传统全局一致性正则化。从结果可以看出，将传统网络重构为DINOv2+DPT密集预测网络后（方案B），模型在1/4和1/2分区下的性能由77.8%和79.2%提升至88.3%和89.4%，说明重构网络显著增强了模型的基础特征表达能力。在此基础上，单独引入高频视角（方案C）或低频视角（方案D）时，模型性能进一步提升至89.3%和89.4%，表明高频视角有助于强化边界细节，低频视角则有利于维持目标内部语义连贯性。然而，直接引入双频域分支但不进行空间掩膜对齐（方案E）时，性能相比单低频分支略有下降，说明全局统一的高低频约束可能产生冲突，即高频扰动影响平滑区域，低频约束削弱边界细节。加入空间掩膜对齐后（方案F），模型在1/4和1/2分区下达到最优性能89.6%和90.2%。这表明空间掩膜对齐能够协调高低频分支的作用范围，使高频约束聚焦边界区域，低频约束作用于目标内部，从而验证了网络重构、空频解耦与空间对齐协同设计的有效性。

如表5所示，本实验测试了1/4标签比例的分区协议下Pascal VOC 2012数据集在不同高频分量提取策略和邻域范围（k）大小下的性能变化。“拉普拉斯算子”提取策略采用空域内的二阶微分边缘检测算子进行高频提取。“FFT”提取策略则在频域利用快速傅里叶变换配合高通掩膜进行提取。从结果中可以看出，当选择邻域范围为3×3（k = 3）的“拉普拉斯算子”策略时，模型表现出最佳性能（89.6%）。相比之下，融合更广邻域信息的5×5（k = 5）和7×7（k = 7）范围反而导致模型性能逐步下降。其原因在于，较大的邻域范围引入了过多的上下文空间信息，导致提取出的高频边缘特征过厚，增加了非边界像素点混入掩膜区域的可能性，从而破坏了高频特征与真实语义边界对齐的精准度。此外，空域的拉普拉斯算子相比快速傅里叶变换（fast Fourier transform，FFT）策略，不仅避免了频域截断带来的潜在振铃效应，还能更精细地贴合局部梯度变化，因而在各项设定下均取得了更优的分割结果。

表4 网络重构与空频模块消融实验结果（%）

Table 4 Ablation results of network reconstruction and spatial-frequency modules（%）

方案	传统网络	重构网络	高频视角	低频视角	掩膜对齐	1/4 (366)	1/2 (732)
A	✓					77.8	79.2
B		✓				88.3	89.4
C		✓	✓			89.3	89.9
D		✓		✓		<u>89.4</u>	<u>90.1</u>
E		✓	✓	✓		89.2	89.7
F		✓	✓	✓	✓	**89.6**	**90.2**

注：性能最优的结果以粗体标注，次优结果以下划线标注，“✓”表示对应实验方案启用该模块/设置。

表5 高频分量提取策略和领域范围的消融实验结果（%）

Table 5 Ablation results of high-frequency component extraction strategies and neighborhood ranges（%）

高频分量提取方法	邻域范围(k)		
	3×3	5×5	7×7
拉普拉斯算子	**89.6**	<u>89.4</u>	88.8
FFT	89.3	89.0	88.3

注：性能最优的结果以粗体标注，次优结果以下划线标注。

3.3.2 微调与冻结DINOv2编码器

本文中的主要实验均对整个模型（编码器+解码器）进行了全微调。然而，受限于GPU显存大小，部分硬件资源有限的研究团队往往难以采用端到端的全量训练策略。针对这一问题，本文尝试在训练阶

段冻结预训练的DINOv2-B编码器权重，仅对随机初始化的解码器进行优化。该做法能够大幅降低训练的硬件门槛：在Pascal VOC 2012数据集上使用DINOv2-B架构时，GPU显存占用从28G减少至约12G，且单轮训练时间减少近50%。

表6　SF-Match框架下对预训练DINOv2编码器进行微调（默认）与冻结的消融实验（%）

Table 6　Ablation study on fine-tuning (default) vs. freezing pre-trained DINOv2 encoder in SF-Match (%)

冻结	有标注图像比例及对应数量				
	1/16(92)	1/8(183)	1/4(366)	1/2(732)	全标注(1464)
	86.4	**87.5**	**89.6**	**90.2**	**90.6**
✓	85.0	86.7	88.0	89.5	89.8
	−1.4	−1.0	−1.6	−0.7	−0.8

注：性能最优的结果以粗体标注，次优结果以下划线标注，“✓”表示该组实验启用编码器冻结策略。

表6列出了在不同标签比例下，编码器微调与冻结两种策略对SF-Match的性能影响。从结果可以看到，在涵盖从极低资源（1/16）到全量数据的所有实验设置下，微调策略的性能均实现了对冻结策略的显著超越，提升幅度介于0.7%-1.6%之间。这一实验现象表明：尽管DINOv2提取的“冻结特征”本身已经具备了强大的零样本表征能力，但允许基础模型在下游特定的语义分割数据上进行端到端的微调更新，能够更彻底地与SF-Match提出的空频解耦机制相契合，从而在特征对齐阶段释放出更大的性能潜力。

3.3.3　超参数敏感性分析

为验证SF-Match对核心超参数的鲁棒性，本文在Pascal VOC 2012数据集的1/4标注划分协议下，对伪标签置信度阈值τ、空频一致性损失权重λ_H,λ_L以及频域残差混合系数α进行了敏感性分析，定量结果如表7所示。

由表7可见，当伪标签置信度阈值τ设置为0.95、空频一致性损失权重λ_H和λ_L均设置为0.5、频域残差混合系数α设置为0.5时，模型取得最佳性能。过低的置信度阈值容易引入噪声伪标签，过高的阈值则会减少有效监督像素；过小的空频损失权重难以充分发挥高低频分支的作用，而过大的权重可能削弱原始RGB一致性分支的稳定性；频域残差混合系数过小时边界增强不足，过大时则可能引入过强高频噪声。总体而言，SF-Match在不同超参数设置下性能波动较小，表明该方法具有较好的超参数鲁棒性。

表7　不同超参数设置下的性能对比（%）

Table 7　Performance comparison under different hyper-parameter settings (%)

超参数变量	设定值	mIoU
置信度阈值τ	0.90	89.2
	0.95	**89.6**
	0.99	88.4
一致性权重λ_H,λ_L	0.25,0.25	88.9
	0.50,0.50	**89.6**
	1.00,1.00	88.8
残差混合系数α	0.25	89.0
	0.50	**89.6**
	1.00	88.6

注：性能最优的结果以粗体标注，次优结果以下划线标注。

3.4　复杂度分析

为进一步评估SF-Match的实际运行效率，本文在Pascal VOC 2012数据集的1/4标注划分协议下，从参数量、运算量、推理速度和单轮训练时间四个方面分析模型复杂度，并结合分割性能与代表性半监督语义分割方法UniMatch、DDFP和semiVL进行综合对比。所有方法均在相同硬件环境、相同输入分辨率和相同数据划分条件下进行测试。其中，参数量表示模型所包含的参数总数，单位为百万；运算量表示单张输入图像完成一次前向推理所需的浮点运算次数，单位为十亿次；推理速度表示批大小为1时模型平均每秒处理的图像数量；单轮训练时间表示在相同训练协议下完成一轮训练所需的平均时间，单位为分钟。为减小偶然波动对测试结果的影响，推理速度和单轮训练时间均在模型充分预热后进行多次测试，并取平均值。实验结果如表8所示。

由表8可见，SF-Match在精度与复杂度之间取得了较好的折中。相比传统ResNet-101编码器的UniMatch与DDFP，SF-Match由于采用DINOv2-B与DPT解码器，参数量、运算量和推理速度均有所逊色，说明更强的Transformer表征和空间重建结构会带来一定推理开销。然而，SF-Match的mIoU达到

表8　不同方法的复杂度与性能对比

Table 8　Complexity and performance comparison of different methods

方法	编码器	参数量	运算量	推理速度	单轮训练时间(分钟)	mIoU (%)
UniMatch	RN-101	59.5	100.8	74.8	11.0	78.8
DDFP	RN-101	59.5	103.6	28.3	10.5	79.5
semiVL	CLIP-B	89.0	159.2	46.5	17.0	<u>86.0</u>
our	DINOv2-B	97.5	192.3	51.3	13.5	**89.6**

注:性能最优的结果以粗体标注,次优结果以下划线标注。

89.6%,较UniMatch提升10.8个百分点,表明额外计算量能够带来明显的精度收益。与多模态模型semiVL相比,SF-Match参数量虽有所上升但推理速度更快,单轮训练耗时更短,同时mIoU提升3.6个百分点。这表明SF-Match在不引入大规模视觉-语言先验的情况下,仍能获得更优的精度和较好的运行效率。

这一结果说明,在半监督语义分割这类密集预测任务中,强语言先验并不一定能够完全转化为像素级分割优势。semiVL等视觉-语言模型主要依赖语言语义先验来缓解类别识别与语义混淆问题,但其优势更多体现在图像级或区域级语义理解上,对于边界细节、局部纹理以及类内结构一致性等像素级问题仍存在一定局限。相比之下,DINOv2通过大规模自监督学习获得了更适合视觉结构建模的表征能力,结合DPT解码器后能够更好地恢复空间细节。在此基础上,SF-Match进一步利用空频一致性策略,将高频信息显式约束于语义边界,将低频信息约束于目标内部区域,使模型在训练过程中同时强化边界判别能力和区域一致性。因此,尽管SF-Match不引入语言先验,但其视觉表征与空频约束更直接地作用于语义分割所需的像素级结构建模,从而在本实验设置下取得了优于semiVL的性能。

3.5　定性结果与局限性分析

3.5.1　定性结果分析

图5展示SF-Match在1/4划分下Pascal VOC 2012数据集上的部分定性结果。与全监督基线、经典半监督方法FixMatch相比,该方法在小目标精准分割、复杂背景鲁棒性上展现出显著优势。

针对小目标分割,SF-Match有效解决了漏检、错分与轮廓失真问题:图(a)中小孩手持的物品为典型小目标,全监督基线完全漏检该目标,FixMatch出现类别错分与形状残缺,SF-Match精准还原目标轮廓与类别,与真实标注高度一致;图(b)中客厅场景的小型座椅,FixMatch将其错分为沙发类,全监督分割轮廓存在明显偏差,SF-Match准确完成类别区分与完整轮廓分割,保障了小目标的分割精度。

针对复杂背景场景,SF-Match具备更强的抗干扰能力与类别区分度:图(c)为多类别密集排布的复杂室内场景,全监督与FixMatch均出现船模目标完全漏检、沙发类别混淆的问题,SF-Match在密集背景干扰下精准识别船模,同时修正背景区域的错分,实现多类别精准区分;图(d)为多行人密集场景,背景与目标特征相似、目标间存在遮挡,对比方法均出现背包分割轮廓失真、类别混淆,SF-Match有效克服背景干扰与遮挡影响,精准完成目标分割。

这证明了SF-Match的有效性:通过空频解耦与空间对齐机制,模型不仅能维持主体结构的连贯性,还极大增强了对高频边缘和小目标的感知能力,在一定程度上改善了传统一致性方法在复杂场景下的边界模糊与类别混淆问题。

3.5.2　失败案例与局限性分析

为进一步分析SF-Match在复杂场景下的性能局限,本文从验证集中选取两个具有代表性的低精度样本,如图6所示。图中从左至右依次为原始图像、真实标注、预测结果、错误区域图、形态学边界掩膜和置信度图。

图6(a)中,在婴儿被座椅和包裹物大面积遮挡的场景中,SF-Match能够识别出主要的person类别区域,但在身体、腿部与座椅/包裹物交界处仍出现明显误分。该场景中真实人体区域呈离散分布,且座椅纹理、衣物褶皱和遮挡边缘会产生较强高频响应,容易干扰模型对真实语义边界的判断。同时,预测结果中的形态学边界掩膜围绕错误扩张的人体轮廓生成,使得边界区域的高频一致性约束可能进一步保留局部错误边界。该现象表明,在高度遮挡和

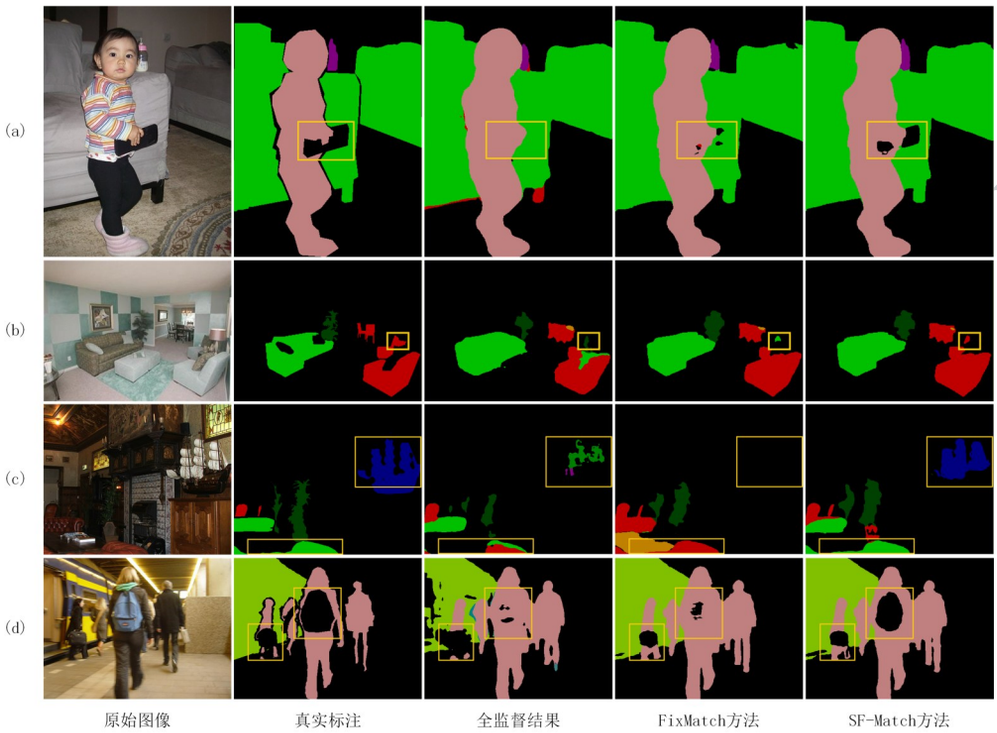


图 5　1/4 划分下 SF-Match 在 Pascal VOC2012 数据集上的定性结果；(a)手持小目标分割；(b)小型座椅分割；(c)复杂室内多类别分割；(d)密集行人遮挡分割。

Fig. 5　Qualitative results of SF-Match on the Pascal VOC2012 dataset under the 1/4 split. Subfigures: (a) egmentation of a small handheld object; (b)segmentation of a small chair; (c)multi-class segmentation in a complex indoor scene; (d)segmentation in a crowded pedestrian scene with occlusion.

纹理密集场景中，基于伪标签的形态学掩膜仍可能受到错误轮廓影响。

图 6(b)中，真实目标为远距离牛群，目标尺度小、边界细碎，并受到栏杆和栈桥结构干扰。模型虽然在牛群附近产生一定响应，但也将部分栈桥或平台区域误预测为前景类别，说明复杂结构化背景中的非语义边缘仍可能被高频增强放大。进一步观察形态学边界掩膜可以发现，错误预测区域同样生成了明显边界，这说明当预测结果或伪标签中存在背景误激活时，基于形态学梯度的空间掩膜可能无法区分真实语义边界与非语义纹理边界，从而导致局部错误被保留。

上述失败案例表明，SF-Match 虽然能够有效提升多数场景下的边界刻画能力和区域一致性，但其性能仍受到教师伪标签质量、形态学边界掩膜准确性以及复杂高频背景干扰的影响。特别是在高度遮挡、目标与背景粘连、远距离小目标和结构化背景较强的情况下，错误预测可能导致边界掩膜偏移，高频一致性约束也可能同时增强真实语义边界和非语义纹理边缘。

4　结　论

本文提出一种空频一致性驱动的半监督语义分割网络(SF-Match)。首先，基于 DINOv2 与 DPT 重构密集预测网络，使模型能够更充分地利用自监督视觉表征，并提升其在标签受限场景下的特征表达与泛化能力。其次，针对全局一致性策略易引发的语义冲突问题，设计了空频解耦与对齐机制。通过将图像显式分离为高低频视角，并结合动态生成的形态学掩膜，实现了“高频对齐边界、低频对齐内部”的分治监督。该设计有效避免了高频噪声对平滑主体的干扰及低频模糊对精细边缘的破坏。实验表明，SF-Match 在 Pascal VOC 2012 和 Cityscapes 数据集上均取得了更优结果。

尽管取得了突破性进展，SF-Match 仍存在一定局限性：一是基于大参数量 Transformer 的多视角前向传播，导致训练阶段显存占用大、耗时较长；二是

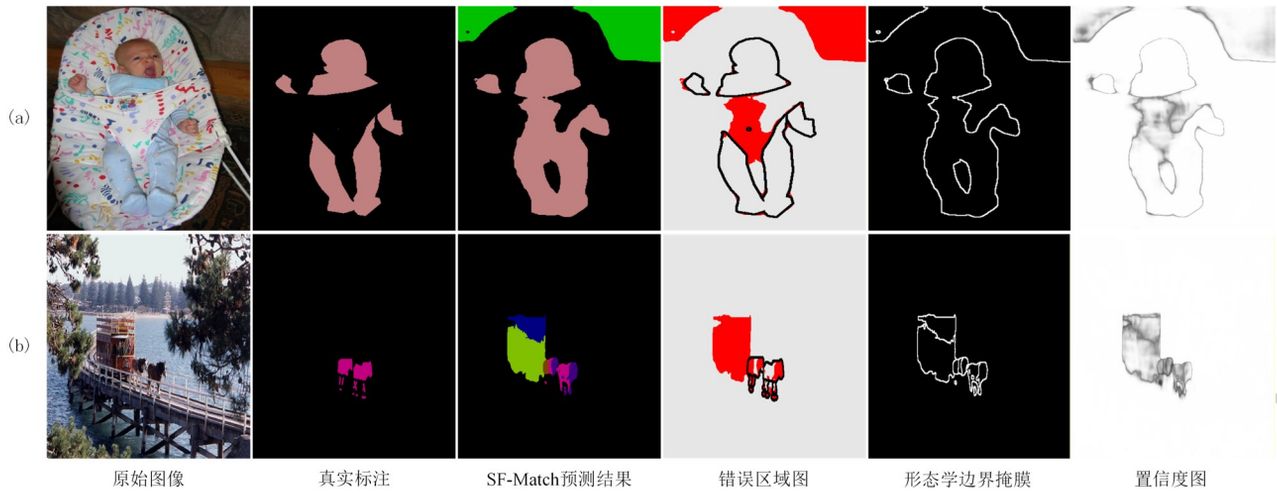


图 6　SF-Match在遮挡与小目标场景中的典型失败案例；(a)高度遮挡婴儿场景；(b)远距离小目标牛群场景。

Fig. 6　Typical failure cases of SF-Match under occlusion and small-object scenarios. Subfigures: (a) highly occluded infant scene; (b) distant cattle scene with small targets.

在高度遮挡、小目标以及复杂高频背景场景中，形态学掩膜仍可能受到错误伪标签或预测边界偏移的影响，导致局部误检或边界错分。

针对上述局限，未来的研究工作计划从以下两个方向展开优化：

降低多视角训练的计算开销：探索从特征层替代图像层实现空频解耦。例如，在网络特征提取阶段引入轻量级频域建模模块，使模型在单次前向传播中获得不同频率信息，从而减少多视角输入带来的显存占用和训练时间开销。

提升复杂场景下区域一致性监督的鲁棒性：针对遮挡、小目标和结构化高频背景中伪标签边界易偏移、非语义纹理易被强化的问题，后续将引入伪标签可靠性评估、边界不确定性建模和自适应区域划分机制，动态调整不同区域的一致性监督权重；或尝试结合无需额外训练的分割大模型，如SAM，提供更稳定的零样本边界先验，以降低错误掩膜对特征对齐训练的干扰。

参考文献(References)

Berthelot D, Carlini N, Goodfellow I, Papernot N, Oliver A and Raffel C A. 2019. MixMatch: a holistic approach to semi-supervised learning//Advances in Neural Information Processing Systems. Vancouver, Canada: Curran Associates, Inc.: 5050-5060

Berthelot D, Carlini N, Cubuk E D, Kurakin A, Sohn K, Zhang H and Raffel C A. 2020. ReMixMatch: semi-supervised learning with distribution matching and augmentation anchoring//Proceedings of the International Conference on Learning Representations. Addis Ababa, Ethiopia: ICLR

Chen D H, Liu Z Q, Yang C X, Wang D, Yan Y, Xu Y and Ji X Y. 2025. ConformalSAM: unlocking the potential of foundational segmentation models in semi-supervised semantic segmentation with conformal prediction//Proceedings of the IEEE/CVF International Conference on Computer Vision. Honolulu, USA: IEEE: 24045-24055

Chen X K, Yuan Y H, Zeng G and Wang J D. 2021. Semi-supervised semantic segmentation with cross pseudo supervision//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 2613-2622

Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J and Houlsby N. 2021. An image is worth 16×16 words: transformers for image recognition at scale//Proceedings of the International Conference on Learning Representations. Virtual: ICLR

Grandvalet Y and Bengio Y. 2004. Semi-supervised learning by entropy minimization//Advances in Neural Information Processing Systems. Vancouver, Canada: MIT Press: 529-536

He K, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 770-778

He R, Yang J and Qi X J. 2023. Rethinking semi-supervised medical image segmentation with transformers//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 23862-23871

Hou Z Q, Qu M J, Li J G, Ma S G, Wang Y C and Yang X B. 2025. Lightweight CNN-Transformer combined network for real-time semantic segmentation. Journal of Image and Graphics, 30(7):

2437-2450 (侯志强, 屈敏杰, 李俊歌, 马素刚, 王昀琛, 杨小宝. 2025. 轻量级 CNN-Transformer 相结合的实时语义分割网络. 中国图象图形学报, 30(7): 2437-2450) [DOI:10.11834/jig.240527]

Howlader P, Das S, Le H and Samaras D. 2024. Beyond pixels: semi-supervised semantic segmentation with a multi-scale patch-based multi-label classifier//Proceedings of the European Conference on Computer Vision. Milan, Italy: Springer: 342-360

Hoyer L, Tan D J, Naeem M F, Van Gool L and Tombari F. 2024. SemiVL: semi-supervised semantic segmentation with vision-language guidance//Proceedings of the European Conference on Computer Vision. Milan, Italy: Springer: 257-275

Hung W C, Tsai Y H, Liou Y T, Lin Y Y and Yang M H. 2018. Adversarial learning for semi-supervised semantic segmentation//Proceedings of the British Machine Vision Conference. Newcastle, UK: BMVA Press: 65

Jin Y, Wang J Q and Lin D H. 2022. Semi-supervised semantic segmentation via gentle teaching assistant//Advances in Neural Information Processing Systems. New Orleans, USA: Curran Associates, Inc.: 2803-2816

Lee D H. 2013. Pseudo-label: the simple and efficient semi-supervised learning method for deep neural networks//Workshop on Challenges in Representation Learning, ICML. Atlanta, USA: ICML: 1-6

Li J N, Xiong C M and Hoi S C H. 2021. CoMatch: semi-supervised learning with contrastive graph regularization//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 9475-9484

Li L J, He Y, Xie G, Zhang H X and Bai Y H. 2024. Cross-layer detail perception and group attention-guided semantic segmentation network for remote sensing images. Journal of Image and Graphics, 29(5): 1277-1290 (李林娟, 贺赟, 谢刚, 张浩雪, 柏艳红. 2024. 跨层细节感知和分组注意力引导的遥感图像语义分割. 中国图象图形学报, 29(5): 1277-1290) [DOI:10.11834/jig.230653]

Li S, He Y, Zhang W M, Zhang W, Tan X, Han J Y, Ding E R and Wang J D. 2023. CFCG: semi-supervised semantic segmentation via cross-fusion and contour guidance supervision//Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 16348-16358

Li Z X, Zhang J, Wu J L and Ma H F. 2022. Semi-supervised adversarial learning based semantic image segmentation. Journal of Image and Graphics, 27(7): 2157-2170 (李志欣, 张佳, 吴璟莉, 马慧芳. 2022. 基于半监督对抗学习的图像语义分割. 中国图象图形学报, 27(7): 2157-2170) [DOI:10.11834/jig.200600]

Liu P and Liu J S. 2025. When confidence fails: revisiting pseudo-label selection in semi-supervised semantic segmentation//Proceedings of the IEEE/CVF International Conference on Computer Vision. Honolulu, USA: IEEE: 21874-21884

Liu S L, Yin S Q, Qu L H, Wang M and Song Z J. 2023. A structure-aware framework of unsupervised cross-modality domain adaptation via frequency and spatial knowledge distillation. IEEE Transactions on Medical Imaging, 42(12): 3919-3931 [DOI: 10.1109/TMI.2023.3318006]

Liu Y Y, Tian Y, Chen Y H, Liu F B, Belagiannis V and Carneiro G. 2022. Perturbed and strict mean teachers for semi-supervised semantic segmentation//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 4258-4267

Lu C Y, Derakhshandeh K and Chaterji S. 2025. Improving semi-supervised semantic segmentation with sliced-Wasserstein feature alignment and uniformity//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 20233-20243

Lv L and Zhang L F. 2025. ScaleMatch: multi-scale consistency enhancement for semi-supervised semantic segmentation//Proceedings of the AAAI Conference on Artificial Intelligence. Philadelphia, USA: AAAI Press: 5910-5918 [DOI: 10.1609/aaai.v39i6.32631]

Olsson V, Tranheden W, Pinto J and Svensson L. 2021. ClassMix: segmentation-based data augmentation for semi-supervised learning//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa, USA: IEEE: 1369-1378

Oquab M, Darcet T, Moutakanni T, Vo H V, Szafraniec M, Khalidov V, Fernandez P, Haziza D, Massa F, El-Nouby A, Assran M, Ballas N, Galuba W, Howes R, Huang P Y, Li S W, Misra I, Rabbat M, Sharma V, Synnaeve G, Xu H, Jégou H, Mairal J, Labatut P, Joulin A and Bojanowski P. 2024. DINOv2: learning robust visual features without supervision. Transactions on Machine Learning Research

Ouali Y, Hudelot C and Tami M. 2020. Semi-supervised semantic segmentation with cross-consistency training//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 12674-12684

Sohn K, Berthelot D, Carlini N, Zhang Z Z, Zhang H, Raffel C A, Cubuk E D, Kurakin A and Li C L. 2020. FixMatch: simplifying semi-supervised learning with consistency and confidence//Advances in Neural Information Processing Systems. Vancouver, Canada: Curran Associates, Inc.: 596-608

Sun B Y, Yang Y Q, Zhang L, Cheng M M and Hou Q B. 2024. CorrMatch: label propagation via correlation matching for semi-supervised semantic segmentation//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 3097-3107

Sun R, Mai H Y, Zhang T Z and Wu F. 2023. DAW: exploring the better weighting function for semi-supervised semantic segmentation//Advances in Neural Information Processing Systems. New Orleans, USA: Curran Associates, Inc.: 61792-61805

Tarvainen A and Valpola H. 2017. Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised

deep learning results//Advances in Neural Information Processing Systems. Long Beach, USA: Curran Associates, Inc.: 1195-1204

Wang H N, Zhang Q X, Li Y and Li X M. 2024. AllSpark: reborn labeled features from unlabeled in transformer for semi-supervised semantic segmentation//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 3627-3636

Wang X Y, Bai H H, Yu L M, Zhao Y and Xiao J M. 2024. Towards the uncharted: density-descending feature perturbation for semi-supervised semantic segmentation//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 3303-3312

Wang Y C, Wang H C, Shen Y J, Fei J J, Li W, Jin G Q, Wu L W, Zhao R and Le X Y. 2022. Semi-supervised semantic segmentation using unreliable pseudo-labels//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 4248-4257

Wang Y D, Chen H, Heng Q, Hou W X, Fan Y, Wu Z, Wang J D, Savvides M, Shinozaki T, Raj B, Schiele B and Xie X. 2023. FreeMatch: self-adaptive thresholding for semi-supervised learning//Proceedings of the International Conference on Learning Representations. Kigali, Rwanda: ICLR

Wang Z C, Zhao Z, Xing X X, Xu D, Kong X Y and Zhou L P. 2023. Conflict-based cross-view consistency for semi-supervised semantic segmentation//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 19585-19595

Xiang W K, Zhou Q, Cui J C, Mo Z Y, Wu X F, Ou W H, Wang J D and Liu W Y. 2024. Weakly supervised semantic segmentation based on deep learning. Journal of Image and Graphics, 29(5): 1146-1168 (项伟康, 周全, 崔景程, 莫智懿, 吴晓富, 欧卫华, 王井东, 刘文予. 2024. 基于深度学习的弱监督语义分割方法综述. 中国图象图形学报, 29(5): 1146-1168) [DOI:10.11834/jig.230628]

Xie E Z, Wang W H, Yu Z D, Anandkumar A, Alvarez J M and Luo P. 2021. SegFormer: simple and efficient design for semantic segmentation with transformers//Advances in Neural Information Processing Systems. Virtual: Curran Associates, Inc.: 12077-12090

Xu H M, Liu L Q, Bian Q C and Yang Z. 2022. Semi-supervised semantic segmentation with prototype-based consistency regularization//Advances in Neural Information Processing Systems. New Orleans, USA: Curran Associates, Inc.: 26007-26020

Yang L H, Zhuo W, Qi L, Shi Y H and Gao Y. 2022. ST++: make self-training work better for semi-supervised semantic segmentation//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 4268-4277

Yang L H, Qi L, Feng L T, Zhang W and Shi Y H. 2023. Revisiting weak-to-strong consistency in semi-supervised semantic segmentation//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 7236-7246

Yang Y and Soatto S. 2020. FDA: Fourier domain adaptation for semantic segmentation//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 4085-4095

Yang Z Q, Zhao X Q, Wang X L, Zhang Q and Xiao J M. 2025. FFR: frequency feature rectification for weakly supervised semantic segmentation//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 30261-30270

Yun S, Han D Y, Oh S J, Chun S, Choe J and Yoo Y J. 2019. CutMix: regularization strategy to train strong classifiers with localizable features//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, South Korea: IEEE: 6023-6032

Zhang B W, Wang Y D, Hou W X, Wu H, Wang J D, Okumura M and Shinozaki T. 2021. FlexMatch: boosting semi-supervised learning with curriculum pseudo labeling//Advances in Neural Information Processing Systems. Virtual: Curran Associates, Inc.: 18408-18419

Zhang H Y, Cissé M, Dauphin Y N and Lopez-Paz D. 2018. mixup: beyond empirical risk minimization//Proceedings of the International Conference on Learning Representations. Vancouver, Canada: ICLR

Zhang J J, Chao H Q, Dhurandhar A, Chen P Y, Tajer A, Xu Y Y and Yan P K. 2023. Spectral adversarial mixup for few-shot unsupervised domain adaptation//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland: 728-738

Zhao Z, Long S F, Pi J M, Wang J D and Zhou L P. 2023. Instance-specific and model-adaptive supervision for semi-supervised semantic segmentation//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 23705-23714

Zhao Z, Yang L H, Long S F, Pi J M, Zhou L P and Wang J D. 2023. Augmentation matters: a simple-yet-effective approach to semi-supervised semantic segmentation//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 11350-11359

Zheng M K, You S, Huang L, Wang F, Qian C and Xu C. 2022. SimMatch: semi-supervised learning with similarity matching//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 14471-14481

Zhou Y F, Li L R, Wang Z C, Liu G L, Liu Z W and Yang G. 2024. XNet v2: fewer limitations, better results and greater universality//Proceedings of the 2024 IEEE International Conference on Bioinformatics and Biomedicine. Lisbon, Portugal: IEEE: 4070-4075 [DOI:10.1109/BIBM62325.2024.10822699]

Zhu Z Q, Sun M W, Qi G Q, Li Y Y, Yang M J, Cheng J and Liu Y. 2026. Frequency adaptation and feature transformation network for

medical image segmentation. Journal of Image and Graphics, 31(1): 303-319 (朱智勤, 孙梦薇, 齐观秋, 李嫄源, 杨梦杰, 程俊, 刘羽. 2026. 融合频率自适应和特征变换的医学图像分割. 中国图象图形学报, 31(1): 303-319) [DOI: 10.11834/jig.250100]

作者简介

胡静,女,教授,主要研究方向为图像处理与深度学习。E-mail:279641292@qq.com

吕晓鹏,通信作者,男,硕士研究生,主要研究方向为计算机视觉、深度学习和图像分割。E-mail:lvxiaopeng@tyust.edu.cn

王芳,女,讲师,主要研究方向为图像分割与计算机视觉。E-mail:wangfang6@tyust.edu.cn

张睿,男,教授,CCF高级会员,主要研究方向为计算机视觉、智能信息处理、自动机器学习。E-mail:zhangrui@tyust.edu.cn