

基于频域筛选的遥感影像地物分割知识蒸馏方法

邓兴升, 李谦

长沙理工大学 航空工程学院, 长沙 410114

摘要: 随着遥感影像语义分割技术在边缘设备中的广泛应用, 如何在有限计算资源下实现精度与效率的平衡成为关键科学问题。本文提出了一种基于频域筛选的知识蒸馏方法 FDF (frequency domain filtering distillation)。该方法利用中心方形掩膜在傅里叶域中对师生特征进行滤波, 选择性传递教师模型中的结构性低频知识, 同时抑制无关高频成分, 从而有效克服了编码器前端特征蒸馏中的高频噪声干扰问题。为进一步探索高效蒸馏架构, 引入了轻量网络 U-Lite 作为学生模型, 其在参数量与推理速度方面具有显著优势。在 LoveDA (land-cover domain adaptation)、Vaihingen 与 Potsdam 等 3 个遥感数据集上的综合实验表明, FDF 的平均全局性能提升和平均区域重叠度提升均位于前列, 且其性能衰退风险显著低于主流频域方法, 其平均交并比差值 $\Delta mIoU$ (mean intersection over union) 标准差 (1.03) 优于通道蒸馏方法 CWD (channel-wise knowledge distillation) (1.25) 和傅里叶 (Fourier) 蒸馏方法 (1.34), 且避免了其他方法最高 2.37% 的性能下降。在计算效率方面, FDF 推理速度达 16.63 ms/帧, 比类内特征变化蒸馏 IFVD (intra-class feature variation distillation) 和傅里叶蒸馏方法快约 2.5 倍。这些优势使 FDF 在实际边缘计算场景中展现出独特的应用价值, 为资源受限环境提供了高可靠性的分割解决方案。

关键词: 深度学习网络, 知识蒸馏, 语义分割, 特征提取, 快速傅里叶变换

中图分类号: TP751.1/P2

引用格式: 邓兴升, 李谦. 2026. 基于频域筛选的遥感影像地物分割知识蒸馏方法. 遥感学报, 30(6): 1841–1855

Deng X S and Li Q. 2026. Distillation method for remote sensing image semantic segmentation via frequency-domain filtering. National Remote Sensing Bulletin, 30(6): 1841–1855 [DOI:10.11834/jrs.20265293]

1 引言

遥感图像分割是精准解译遥感信息、实现地物识别与变化监测等应用的关键环节 (马妍和古丽米拉·克孜尔别克, 2023)。近年来, 深度卷积神经网络 CNN (convolutional neural network) 凭借其强大的特征学习能力, 已成为高分辨率遥感影像地物分割的主流方法。以 FCN (fully convolutional network)、U-Net、SegNet、DeepLab、PSPNet (pyramid scene parsing network) (Long 等, 2015; Ronneberger 等, 2015; Badrinarayanan 等, 2017; Chen 等, 2018; Zhao 等, 2017) 及其衍生模型为代表的分割架构奠定了该领域的技术基础。然而, 深度学习网络 (尤其是大参数模型) 的训练严重依赖图形处理单元 GPU (graphics processing unit) 等高性能硬件资源, 对应用设备的算力提出了较

高要求, 这在相当程度上限制了人工智能技术在资源受限场景 (如边缘设备、嵌入式系统) 的广泛应用与发展 (黄震华 等, 2022)。为了使高性能的深度学习模型能够在低资源设备上实现低功耗和实时性等要求, 同时尽量不降低模型性能, 研究人员开始致力于深度学习模型的压缩研究。目前, 知识蒸馏 (Gou 等, 2021) 是模型压缩领域的热点方向。

Hinton 等 (2015) 提出将教师模型和学生模型共同的输出作为目标函数, 将复杂的教师模型中有用的信息迁移至简单的学生网络上进行训练, 其中知识被定义为教师模型输出的软化概率分布。这种将复杂教师模型的表征能力凝练并注入轻量学生模型的过程, 被称为知识蒸馏。后来, Gotmare 等 (2018) 发现, 教师模型的软标签主要在网络深层对学生模型进行训练指导, 而在特征

收稿日期: 2025-08-07; 预印本: 2026-04-13

基金项目: 湖南省地理空间信息工程技术研究中心开放课题 (编号: HNGIET2025001); 湖南省自然科学基金项目 (编号: 2024JJ8335)

第一作者简介: 邓兴升, 研究方向为遥感影像智能解译。E-mail: dengxs@csust.edu.cn

通信作者简介: 李谦, 研究方向为遥感影像语义分割。E-mail: 2648296612@qq.com

提取层的引导作用则相对较弱。Romero等(2015)提出,借助教师中间层的特征作为“提示”,增强对学生网络中间层的监督,从而解决了特征提取层引导作用弱的问题。虽然中间层蒸馏显著改善了学生网络的特征提取能力,但这类方法最初主要聚焦于图像分类场景,而遥感影像地物分割本质上属于语义分割任务。Li等(2017)和Liu等(2023)的研究发现,直接利用基于分类的知识蒸馏KD(knowledge distillation)方法来处理语义分割任务可能无法达到理想的性能。这是因为严格对齐教师和学生网络之间的粗糙特征图可能导致负约束,并忽略像素之间的结构化上下文信息(Yang等,2022)。

近年来,语义分割任务中的KD方法不断演化。例如,Wang等(2020)提出的类内特征变化蒸馏IFVD(intra-class feature variation distillation),通过特征分布对齐,实现了教师与学生模型之间的知识迁移;Yang等(2022)的跨图像关系知识蒸馏CIRKD(cross-image relational knowledge distillation)则利用全局像素—区域关联,进一步挖掘结构化知识;Liu等(2024)的边界敏感知识蒸馏BPKD(boundary Privileged knowledge distillation)首次将边缘与主体区域解耦,显著提升了边界精度并强化了内部一致性。此外,Zhang等(2024)提出的频域知识蒸馏FreeKD(frequency knowledge distillation)首次将语义频域提示引入蒸馏流程,通过像素级频率掩膜精准定位高低频关键信息,在COCO(common objects in context)目标检测与Cityscapes语义分割任务中均显著优于现有空间方法。同时,傅里叶(Fourier)蒸馏方法(Shi等,2025)通过在编码器后端的频域蒸馏中引入可学习参数,自适应融合教师与学生的振幅与相位信息,在密集预测任务中取得了良好效果,进一步验证了傅里叶变换在知识蒸馏中的有效性。

然而,这些现有方法普遍存在以下两个关键局限。(1)现有方法普遍聚焦于编码器深层或输出端的高级语义特征蒸馏,如BPKD(Liu等,2024)和Fourier蒸馏方法(Shi等,2025)均主要蒸馏深层通道。这种“末端导向”策略虽然有效,却严重低估了编码器前端提取的低级视觉特征(如边缘、纹理、颜色)在构建精确目标边界、区分细粒度类别以及提升模型空间细节感知能力中的基础性作用。赵雅等(2022)设计的上下文关

联知识蒸馏C2KD(context correlation knowledge distillation)方法通过利用教师模型编码器前端的多尺度上下文相关性知识进行蒸馏,显著提升了学生模型对全局上下文信息的捕捉能力,其实验结果甚至表明编码器前端蒸馏的贡献大于解码器端。(2)在遥感影像语义分割任务中,编码器前端特征往往包含丰富的空间细节与频域信息,这些低级特征虽有助于保留地物边缘与纹理,但同时也引入了大量与任务无关的背景噪声(如云层、阴影、传感器扰动等)。这类噪声在频域中通常表现为高频分量,若在未加筛选的情况下直接迁移,极易干扰学生网络对结构性语义的学习,导致知识蒸馏效率下降,甚至引发性能退化。现有主流蒸馏方法[如FitNets蒸馏方法(Romero等,2015)的特征模仿、通道蒸馏方法CWD(channel-wise knowledge distillation)(Shu等,2021)的通道关系迁移、相位幅蒸馏方法(Phase)(Yao等,2025)的傅里叶变换相位幅直接对齐]多针对自然图像分类或中间层特征设计,其假设特征已具备较高的语义抽象性,故在未充分考量遥感前端特征特性的前提下,直接迁移至编码器前端,往往表现出较差的场景适配性。本实验发现,这类方法在遥感分割任务中蒸馏效果不稳定。

针对上述问题,本文提出一种面向遥感前端特征的频域筛选知识蒸馏方法FDF(frequency domain filtering distillation),创新性地引入正方形中心掩膜滤波机制,在傅里叶空间中选择性迁移教师模型的结构性低频知识,同时抑制由卷积操作与背景噪声引入的高频干扰。与现有频域方法相比,本方法的创新在于:(1)直接在编码器前端进行高频噪声的筛选,避免了噪声在后续蒸馏过程中的传播和积累;(2)不引入任何额外的可学习参数,显著降低了计算复杂度和过拟合风险。通过这种机制,本方法在提升蒸馏效果的同时,显著增强了模型的稳定性和泛化能力。

2 方法与原理

教师模型与学生模型均以U-Net的编码器—解码器结构为基本骨架。本文选用U-Net及其3种变体,如DCSAU-Net(deeper and more compact split-attention u-shape network)(Xu等,2023)、ACC-UNet(a completely convolutional UNet)(Ibtehaz和Kihara,2023)、U-Lite(lightweight U-Net)(Dinh

等, 2023) 作为教师模型与学生模型的实现实例。这些变体在原始 U-Net 基础上引入了注意力、残差或多尺度跳跃连接等机制。

实验分两个阶段：(1) 在遥感影像分割数据集上独立训练教师模型并保存权重；(2) 冻结教

师模型，提取其编码器第一次下采样之后、即将送入对应跳跃连接的特征图，经频域筛选后，作为蒸馏监督信号（蒸馏损失见 2.1 节），整体流程如图 1 所示。

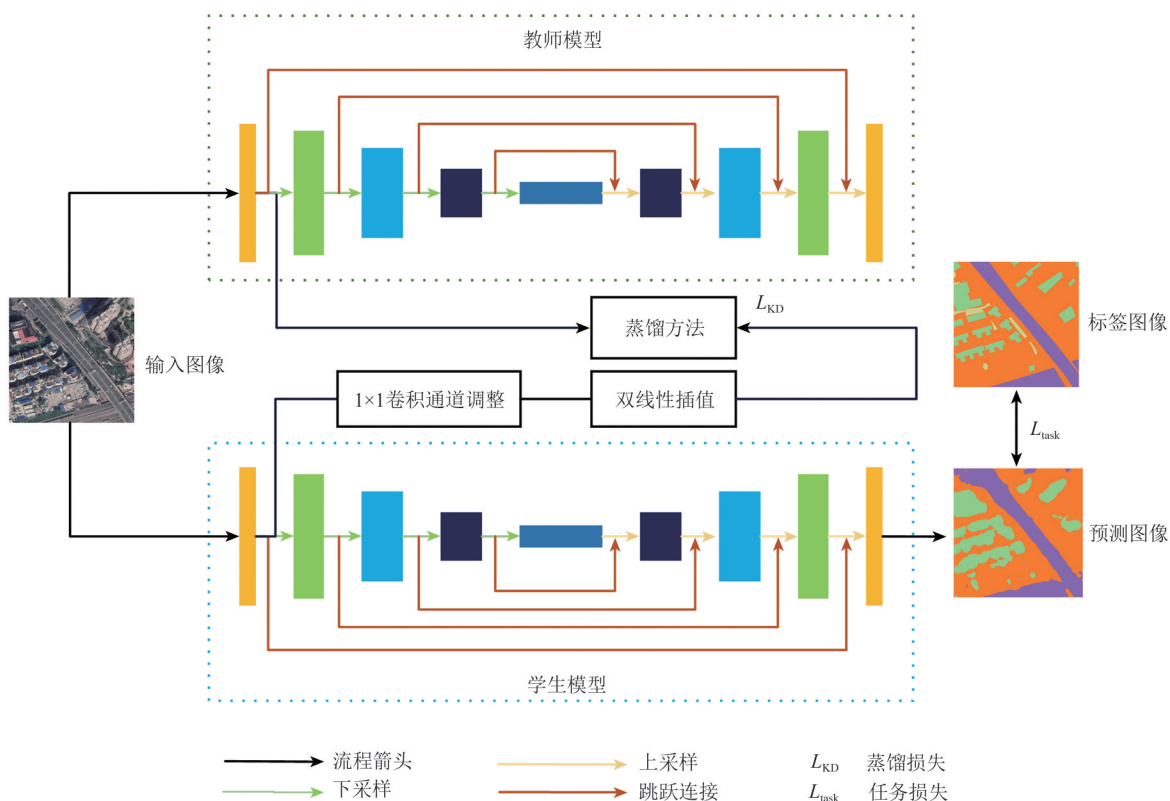


图 1 中间层蒸馏实验流程

Fig. 1 Flowchart of middle-layer distillation experiment

学生模型在蒸馏阶段同时最小化该频域蒸馏损失和真值标签的交叉熵损失（总体损失见 2.2 节）。训练完成后，仅需部署学生模型，即可完成遥感影像的语义分割预测。

2.1 蒸馏位置选择与蒸馏函数理论

2.1.1 前端蒸馏位置选择

教师—学生模型蒸馏的点位仅选择在 U-Net 结构模型的第一个跳跃连接处，该处特征将沿模型的整体处理路线传递至整个解码路径，使所学知识能够被学生模型更充分地吸收。Romero 等 (2015) 研究表明，仅利用教师模型中间层特征作为监督信号，不引入输出蒸馏 (logit distillation) (Hinton 等, 2015)，即可有效提升学生模型表征能力。

2.1.2 频域筛选原理

编码器前端提取的低级视觉特征虽然包含目标轮廓的基础信息，但也混杂了对语义分割无益的细节和噪声。Zhang 等 (2022) 研究表明，小波域中的高频子带虽然能够传递边缘和纹理等细粒度信息，但若不加筛选地迁移全部高频成分，容易将教师模型中的噪声引入学生模型，导致性能下降。例如，FitNets 和 Phase 方法在直接作用于前端特征时表现不佳：FitNets 在 DCSAU-Net 教师模型下平均交并比 mIoU (mean intersection over union) 降至 51.62%，低于基线模型的 52.13% (表 1)；Phase 在 UNET 教师模型下 mIoU 仅为 52.08% (表 2)。FitNets 方法蒸馏全局特征，Phase 方法蒸馏频域相位谱特征 (原始高频分量)，这些结果表明，这两种方法未能有效处理高频分量，可能是其性能不佳的潜在原因之一。

表1 FitNets方法蒸馏效果
Table 1 FitNets distillation results

数据集	教师模型	学生模型	PA/%	mIoU/%
LoveDA	DCSAU-Net	U-Lite	64.43	51.62
	ACC-UNet	U-Lite	66.32	53.94
	UNET	U-Lite	64.88	53.21
	基线模型	U-Lite	64.46	52.13

注：加粗的数值为低于基线模型指标的蒸馏结果；PA为像素准确率。

表2 Phase方法蒸馏效果
Table 2 Phase distillation results

数据集	教师模型	学生模型	PA/%	mIoU/%
LoveDA	DCSAU-Net	U-Lite	64.77	52.62
	ACC-UNet	U-Lite	64.80	52.26
	UNET	U-Lite	64.39	52.08
	基线模型	U-Lite	64.46	52.13

注：加粗的数值为低于基线模型指标的蒸馏结果；PA为像素准确率。

为解决上述问题，本文提出一种基于频域筛选的知识蒸馏方法。该方法利用快速傅里叶变换（Nussbaumer, 1981）将特征映射至频域，通过分析频谱能量分布，构建正方形掩膜以分离有效结构与高频噪声。具体而言，频谱中心区域对应于图像的低频成分，承载着建筑物轮廓与道路走向等主体结构信息；而外围区域则包含砖瓦纹理、边缘细节及卷积操作引入的高频噪声。通过保留中心低频区域并滤除外围高频成分，本方法能够在无需引入额外参数的前提下，重建出纯净的结构特征作为蒸馏监督信号。

根据 Rahaman 等（2019）提出的频谱偏置理论，神经网络在学习过程中倾向于优先捕获低频成分，且低频特征对参数扰动具有更强的鲁棒性。因此，通过聚焦于教师与学生模型中具鲁棒性的低频结构信息进行知识迁移，可有效提升蒸馏过程的稳定性与最终性能。

2.1.3 频域筛选算法流程

本文将模型中需要对齐中间层的特征图简单表示为 $\mathbf{F} \in \mathbb{R}^{C \times H \times W}$ ，其中 C 及 H 、 W 分别为特征通道数及特征空间尺寸的高和宽，学生特征记为 $\mathbf{S}$ ，教师特征记为 $\mathbf{T}$ 。

因教师特征与学生特征的通道与尺寸通常不一致，本文先通过 $g(\cdot)$ 即 1×1 卷积（式（1））

与双线性插值（式（2））完成学生特征对教师特征的通道与尺寸对齐， $\mathbf{F}'$ 为通道对齐结果， $\mathbf{F}''$ 为尺寸对齐结果。

$$\mathbf{F}' = g(\mathbf{S}) \quad (1)$$

$$\mathbf{F}'' = \text{Resize}(\mathbf{F}', \text{size}(\mathbf{T})) \quad (2)$$

$$\mathbf{S} = \mathbf{F}'' \quad (3)$$

频域筛选损失函数流程见图2。对 $\mathbf{T}$ 和 $\mathbf{S}$ 做快速傅里叶变换 FFT（fast Fourier transform），得到教师特征和学生特征的复数频谱 $\mathbf{T}^{\text{FFT}}$ （式（4））和 $\mathbf{S}^{\text{FFT}}$ （式（5））。

$$\mathbf{T}^{\text{FFT}} = \text{FFT}(\mathbf{T}) \quad (4)$$

$$\mathbf{S}^{\text{FFT}} = \text{FFT}(\mathbf{S}) \quad (5)$$

频域筛选模块设计，首先构造中心正方形低通掩膜 $\mathbf{M}_L$ （式（6）），其中掩膜尺寸 R 控制边长 L （式（7））， R 越小，通过的高频分量越少。随后逐元素乘掩膜得到筛选结果 $\mathbf{S}^{\text{LT}}$ （式（8））和 $\mathbf{T}^{\text{LT}}$ （式（9））。

$$\mathbf{M}_L = \begin{cases} 1, & \text{在以特征图中心点为中心,} \\ & \text{边长为 } L \text{ 的正方形区域内} \\ 0, & \text{特征图中的其他区域} \end{cases} \quad (6)$$

$$L = R \times \min(W, H) \quad (7)$$

$$\mathbf{S}^{\text{LT}} = \mathbf{S}^{\text{FFT}} \otimes \mathbf{M} \quad (8)$$

$$\mathbf{T}^{\text{LT}} = \mathbf{T}^{\text{FFT}} \otimes \mathbf{M} \quad (9)$$

执行快速傅里叶逆变换 IFFT（inverse fast Fourier transform）重建空间域特征图得到 $\mathbf{S}^{\text{IFFT}}$ （式（10））和 $\mathbf{T}^{\text{IFFT}}$ （式（11）），再将特征图归一化至 $[0, 1]$ ，得到 $\mathbf{S}^{\text{norm}}$ （式（12））与 $\mathbf{T}^{\text{norm}}$ （式（13）），计算公式如下：

$$\mathbf{S}^{\text{IFFT}} = \text{IFFT}(\mathbf{S}^{\text{LT}}) \quad (10)$$

$$\mathbf{T}^{\text{IFFT}} = \text{IFFT}(\mathbf{T}^{\text{LT}}) \quad (11)$$

$$\mathbf{S}^{\text{norm}} = \frac{\mathbf{S}^{\text{IFFT}} - \min(\mathbf{S}^{\text{IFFT}})}{\max(\mathbf{S}^{\text{IFFT}}) - \min(\mathbf{S}^{\text{IFFT}})} \quad (12)$$

$$\mathbf{T}^{\text{norm}} = \frac{\mathbf{T}^{\text{IFFT}} - \min(\mathbf{T}^{\text{IFFT}})}{\max(\mathbf{T}^{\text{IFFT}}) - \min(\mathbf{T}^{\text{IFFT}})} \quad (13)$$

最后计算特征图均方误差 MSE（mean square error）蒸馏损失。设 L_{KD} 为频域筛选损失函数， N 为特征图总元素数量，计算公式如下：

$$N = C \times H \times W \quad (14)$$

$$L_{\text{KD}} = \frac{1}{N} \sum_{c=1}^C \sum_{w=1}^W \sum_{h=1}^H (\mathbf{T}_{c,w,h}^{\text{norm}} - \mathbf{S}_{c,w,h}^{\text{norm}})^2 \quad (15)$$

2.1.4 频域掩膜 R 值选取

在知识蒸馏中，FDF 方法的目标是将教师模型习得的、对下游任务至关重要的特征表征传递

给学生模型。编码器前端（浅层网络）负责提取基础的、通用的特征，如边缘与纹理。这些特征与原始图像的像素级信息高度相关，但其表征形式更为抽象。由此提出假设：编码器前端特征图

所承载的关键结构性信息，其在频域中的分布，与驱动原始图像视觉内容的主要频带具有同源性，即构建高质量浅层特征所依赖的频域成分，主要由原始图像的频谱特性决定。

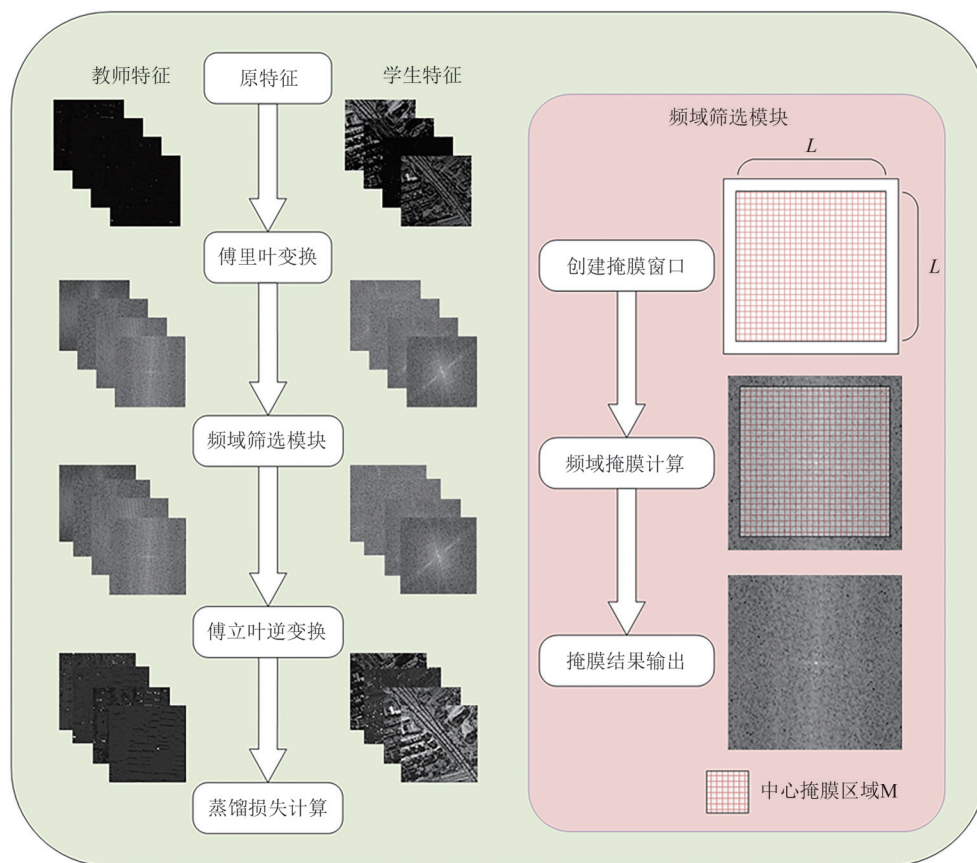


图2 频域筛选损失函数流程

Fig. 2 Frequency-domain filtered loss function flowchart

为验证这一假设并实现对特征图中频域成分的精准筛选，本文设计了一种基于原始图像频域先验的掩膜方法。该方法的基本思想是：利用原始图像的频域能量分布作为向导，识别并保留特征图中与图像内容相关的频带，同时滤除那些在原始图像层面能量微弱、可能对应噪声的高频成分。

在具体实现中，首先分析遥感图像的频域能量分布特性。给定图像 $I(x, y)$ ，其频域表示为 $F(u, v) = \mathcal{F}\{I(x, y)\}$ ， $\mathcal{F}\{\cdot\}$ 表示二维离散傅里叶变换。功率谱密度定义为 $P(u, v) = |F(u, v)|^2$ 。根据帕什瓦（Parseval）定理（Proakis 和 Manolakis, 2007），空域与频域满足能量守恒，即

$$\sum_{x,y} |I(x,y)|^2 = \frac{1}{H \times W} \sum_{u,v} P(u,v) \quad (16)$$

式中， $H \times W$ 为图片像素总数， H 、 W 分别为图像的高度与宽度。

本文定义一个中心方形掩膜函数 $M_L(u, v)$ ，其尺寸参数 $R \in [0, 1]$ 表示掩膜边长相对于图像尺寸的比例（取值范围为 0—1，对应 0%—100% 的边长占比）。掩膜后保留的能量比例为

$$E(R) = \frac{\sum P(u,v) \times M_L(u,v)}{\sum P(u,v)} \quad (17)$$

在 3 个数据集上分别随机抽取 100 张原始图像进行频域统计分析。能量累积曲线（图 3—5）表明，遥感图像的频域能量集中于低频区域：保留约 95% 的能量仅需覆盖不足全频谱 1/4 的窗口面积。这一现象揭示了高频成分的普遍低能量特性，为通过滤除高频成分以抑制噪声提供了理论依据。

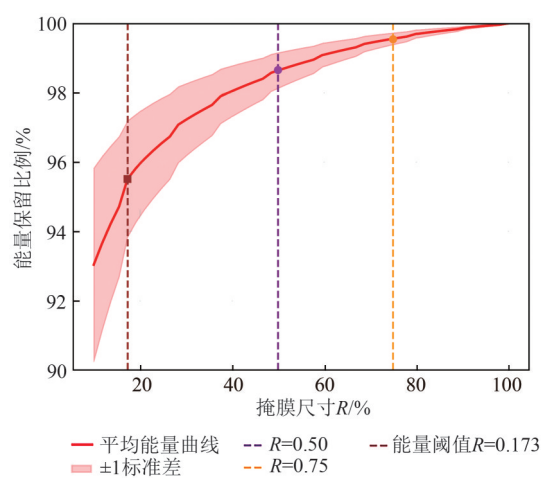


图3 LoveDA数据集能量掩膜分布曲线
Fig. 3 LoveDA Dataset energy mask distribution curve

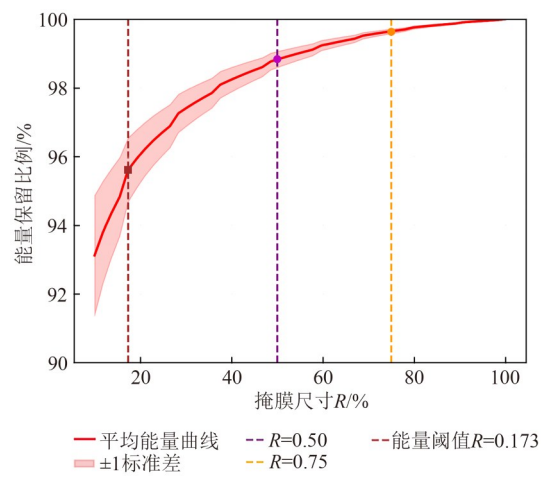


图4 Vaihingen数据集能量掩膜分布曲线
Fig. 4 Vaihingen Dataset energy mask distribution curve

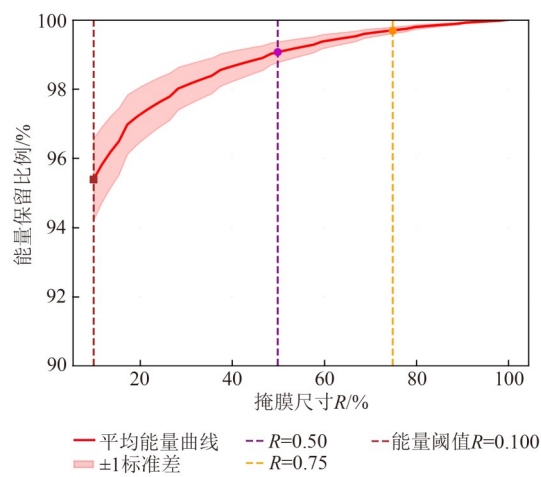


图5 Potsdam数据集能量掩膜分布曲线
Fig. 5 Potsdam Dataset energy mask distribution curve

基于上述分析，本文设置了4个能量保留级别，以系统评估频谱保留程度对蒸馏性能的影响。

(1) FDF (低)：保留约95%的能量。此级别旨在捕捉最主要的低频结构信息，对应 R 为10%—17.3% (依据数据集调整)。(2) FDF (中)：保留约98.86%的能量，统一设置 R 为50%。此级别在信息保留与噪声滤除之间取得平衡。(3) FDF (高)：保留约99.65%的能量，统一设置 R 为75%。此级别倾向于保留更多高频细节。(4) FDF (完整)：保留100%的能量，即 R 为100%。此级别作为基线对照，等同于未施加频域滤波。该方法结果与FitNets方法相当。各能量级别与掩膜尺寸的对应关系如表3所示。这种设计可以跨数据集以一致的能量保留水平进行比较，同时兼顾了数据集间的固有特性差异。

表3 能量级别与掩膜尺寸对应关系			
Table 3 Correspondence between energy levels and mask sizes			
能量级别	保留能量比例/%	掩膜尺寸(R)/%	描述
完整	100	100	完整频谱保留
高	99.65	75	接近完整频谱保留
中	98.86	50	平衡频谱保留
低	95	10—17.3	主要频谱保留

2.2 损失函数

2.2.1 任务损失函数

语义分割的目标是实现逐像素分类。为此，基线模型在训练过程中需要最小化其预测结果与真实像素标签之间的差异，该差异通过标准的像素级交叉熵损失函数 (L_{task}) (Mao 等, 2023) 进行度量。

$$L_{\text{task}} = \frac{1}{H \times W} \sum_{h=1}^H \sum_{w=1}^W CE(y_{h,w,c}, \sigma(Z_{h,w,c})) \quad (18)$$

式中， $CE(\cdot)$ 表示交叉熵损失函数， H 和 W 分别为图像的高度和宽度， c 为类别数， $y_{h,w,c}$ 是位置 (h, w) 处类别 c 的真实独热 (one-hot) 标签， $Z_{h,w,c}$ 是模型预测的逻辑值 (logits)， $\sigma(\cdot)$ 表示 softmax 函数。

2.2.2 知识蒸馏整体损失函数

在应用本文提出的前端频域筛选知识蒸馏方法时，学生模型的优化目标需要同时兼顾对真实标签的拟合以及从教师模型迁移的知识。因此，定义整体损失函数 (L_{total}) 为任务损失与蒸馏损失

的加权和：

$$L_{\text{total}} = (1 - \lambda)L_{\text{task}} + \lambda L_{\text{KD}} \quad (19)$$

式中， L_{task} 是式(18)定义的像素级交叉熵损失，用于确保学生模型学习基本的语义分割能力； L_{KD} 是频域筛选损失函数，基于教师模型和学生模型经过频域筛选后的前端特征计算而得，具体见式(15)，旨在引导学生模型学习教师模型所提取的关键结构信息； λ 是蒸馏损失权重系数，用于调节教师模型知识迁移对学生模型训练的贡献程度。 λ 值越大，学生模型越倾向于模仿教师特征； λ 值越小，则越依赖任务损失进行学习。

3 实验设置

3.1 实验数据集

3.1.1 LoveDA

LoveDA数据集由季顺平(2018)构建，旨在支持城—乡复杂场景下的高分辨率地表覆盖分类与域自适应研究。数据采自南京、重庆、哈尔滨，经下采样共获得5987张0.9 m分辨率的航空正射影像，单幅1024像素×1024像素，对应166768个精细勾绘的多边形对象。语义定义含8类：背景、建筑、道路、水体、荒地、森林、农业用地及无资料区，见图6(a)。为降低显存占用并统一输入尺寸，实验前将影像双线性插值至256×256，并按1000张、400张、400张划分为训练集、验证

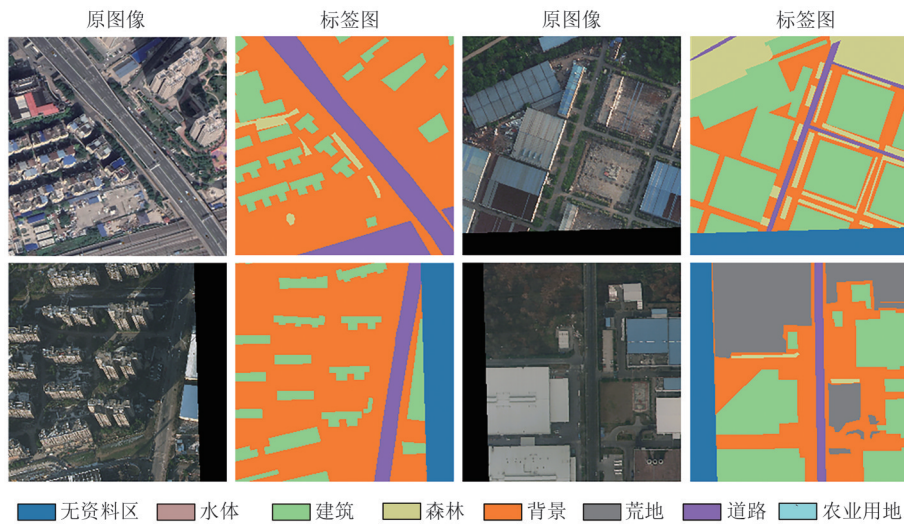
集、测试集。

3.1.2 Potsdam

Potsdam数据集由国际摄影测量与遥感学会ISPRS (International Society for Photogrammetry and Remote Sensing) 竞赛发布，共包含38张5 cm分辨率的航空影像及配套的数字表面模型DSM (digital surface model)、归一化数字表面模型(normalized digital surface model)，光谱通道为红、绿、蓝、近红外四波段，覆盖德国波茨坦市区域。官方标注6类：不透光表面、建筑、低矮植被、树木、汽车、背景，见图6(b)。经相同滑动窗口裁剪与降采样后，获得608张256像素×256像素的影像块，按400张、104张、104张分配训练集、验证集、测试集，用以验证算法在大范围城区影像上的蒸馏效果。

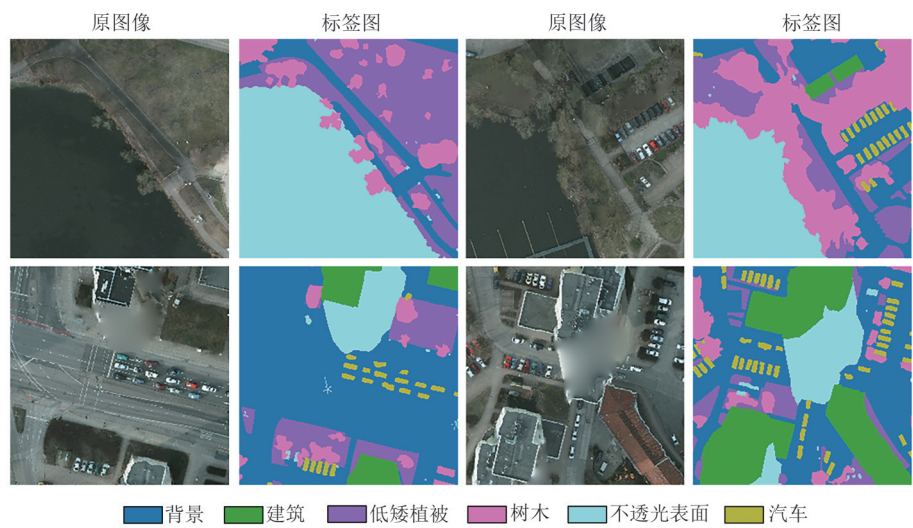
3.1.3 Vaihingen

Vaihingen数据集源自国际摄影测量与遥感学会二维语义标注竞赛 (ISPRS 2D Semantic Labeling Contest)，包含33张9 cm分辨率的真正射影像TOP (true orthophoto) 及同期获取的DSM、nDSM，光谱通道为红、绿、近红外三波段，测区位于德国斯图加特市郊法伊英根小镇。官方标注6类：道路、建筑、低矮植被、树木、汽车、背景，见图6(c)。本文采用滑动窗口将图像裁剪后降采样至256像素×256像素，共获得864张影像块，其中638张用于训练，113张用于验证，113张用于测试。



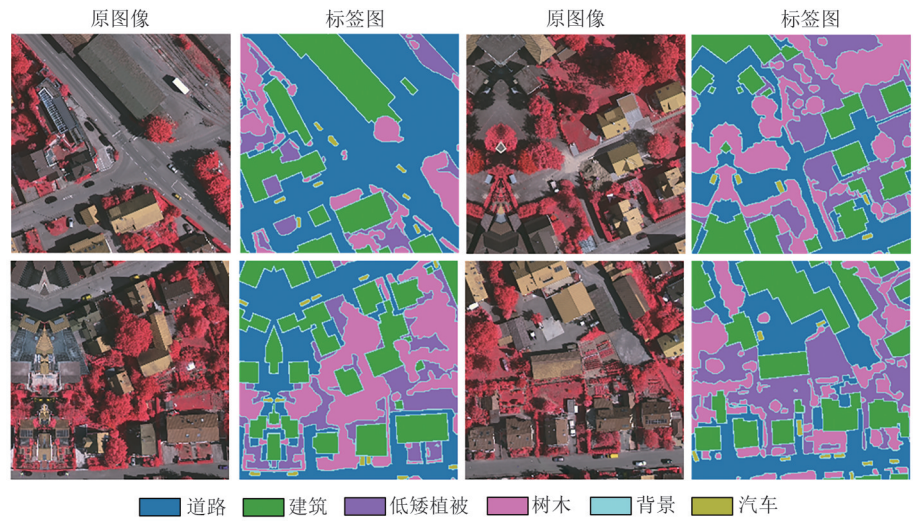
(a) LoveDA 数据集

(a) LoveDA Dataset



(b) Potsdam数据集

(b) Potsdam Dataset



(c) Vaihingen数据集

(c) Vaihingen Dataset

图6 实验数据集

Fig. 6 Experimental Dataset

3.2 实验参数设置

实验在以下软硬件环境下进行：操作系统为 Windows 11、Python 3.10.15、PyTorch 2.5.1、CUDA 12.4；硬件配置为 AMD Ryzen 5 5500 CPU 和 NVIDIA GeForce RTX 4070 Super GPU。

关键训练超参数设置见表4。所有实验批处理大小（Batch Size）设置为4。最大训练周期（Epochs）设置为100，此设置考虑了数据集规模及轻量化学生模型（U-Lite）较快的收敛特性。模型优化均使用Adam优化器（Kingma和Ba，2017）（参数为 $\beta_1=0.9$ 、 $\beta_2=0.999$ 、 $\varepsilon=1\times10^{-8}$ ），并采用10个周期的学习率预热（Warm-up）。

表4 关键训练超参数设置

Table 4 Key training hyperparameter settings		
参数	基线训练	知识蒸馏训练
批处理大小	4	4
学习率	{0.0003, 0.001, 0.003}	{0.0003, 0.001, 0.003}
最大训练周期	100	100
优化器	Adam	Adam
学习率预热周期	10	10
蒸馏损失权重 λ	无	{0.1, 0.2, 0.3}

基线训练使用交叉熵损失进行训练，如式（18）所示。学习率在 {0.0003, 0.001, 0.003} 范围内进行网格搜索以确定最优值。

知识蒸馏训练学生模型的总损失函数定义见式 (19)。为获得稳定且最优的蒸馏效果，本实验对学习率（范围同基线，为 $\{0.0003, 0.001, 0.003\}$ ）和蒸馏损失权重 λ （在 $\{0.1, 0.2, 0.3\}$ 范围内）进行了详尽的网格搜索（Grid Search）。

3.3 模型评价指标

在语义分割任务中，准确地评估模型性能是衡量其实际应用价值的关键环节。为了全面且客观地衡量模型的分类与分割能力，本文采用了像素准确率 PA（pixel accuracy）、平均交并比 mIoU（mean intersection over union）这两种经典的评估函数。

PA 为被正确分类的像素数量占所有像素数量的比例，其从像素层面直接反映分类正确率，数值越高意味着分割效果越好，公式为

$$PA = \frac{pix_{cn}}{pix_m} \quad (20)$$

式中， pix_{cn} 为正确分类的像素数， pix_m 为所有像素数。

mIoU 作为语义分割领域的标准度量指标，通过综合评估模型对各目标区域的分割效果，能够精准地衡量模型在复杂场景下的分割能力，公式为

$$mIoU = \left(\frac{1}{k} \right) \times \sum_{i=1}^k \left(\frac{TP_i}{TP_i + FP_i + FN_i} \right) \quad (21)$$

式中， k 为类别总数， TP_i 是第 i 类的真正例（True Positives）， FN_i 是第 i 类的假反例（False Negatives）， FP_i 是第 i 类的假正例（False Positives）。

3.4 实验模型

鉴于医学图像处理与遥感图像分析在特征提取和计算效率需求上的高度相似性，为探索高效遥感分割模型并服务于后续知识蒸馏研究，本文尝试将医学影像领域的先进轻量化模型应用于遥感分割任务中。这些模型基于 U-Net 架构进行优化，在保持较低计算复杂度的同时，展现了处理分割任务的潜力。具体模型参数如表 5 所示。

表 6 展示了 4 个网络模型在不同数据集上的 PA 及 mIoU。其中，U-Lite 模型凭借极低的参数量（0.88 M）和浮点运算量（3.30 GFLOPs），成为理想的轻量化学生模型候选，并且其基线性能与参数量更大的模型存在显著差距。这种性能差距为知识蒸馏技术提供了明确的应用空间——通过迁移大模型（教师模型）的知识来提升轻量模型

（学生模型）U-Lite 的性能。本文 ACC-UNet 模型的整体卷积通道数被调整为原 ACC-UNet 模型默认通道数的 1/4，参数量减少至 1.1 M。

表 5 模型参数

网络模型	参数量/M	浮点运算量/GFLOPs
U-Lite	0.88	3.30
DCSAU-Net	2.60	26.32
ACC-UNet	1.10	15.48
UNET	33.05	220.77

表 6 模型分割精度

数据集	网络模型	PA/%	mIoU/%
LoveDA	U-Lite	64.46	52.13
	DCSAU-Net	67.94	56.25
	ACC-UNet	68.65	57.26
	UNET	70.35	57.90
Vaihingen	U-Lite	76.01	53.78
	DCSAU-Net	78.61	58.34
	ACC-UNet	76.80	56.10
	UNET	79.07	60.88
Potsdam	U-Lite	76.77	57.25
	DCSAU-Net	77.21	58.96
	ACC-UNet	74.95	56.29
	UNET	79.05	61.36

4 实验结果分析

为了全面评估频域筛选知识蒸馏方法在遥感影像地物分割任务中的性能及其优势，本文选择了性能更优的 DCSAU-Net、ACC-UNet 和 UNET 作为教师模型，U-Lite 作为轻量级学生模型，构建了 3 组典型的教师—学生模型组合，并且在 3 个不同的遥感数据集上进行了广泛的测试，以验证所提出方法的有效性。

4.1 频域筛选方法消融实验

4.1.1 能量消融实验

为探究中心正方形掩膜边长尺寸 R 对学生模型的影响，本实验在不同能量级别下进行蒸馏实验，得出网格搜索后的最佳值。图 7 给出了不同 R 值下教师特征的可视化，可见 R 越小，高频细节丢失越严重。

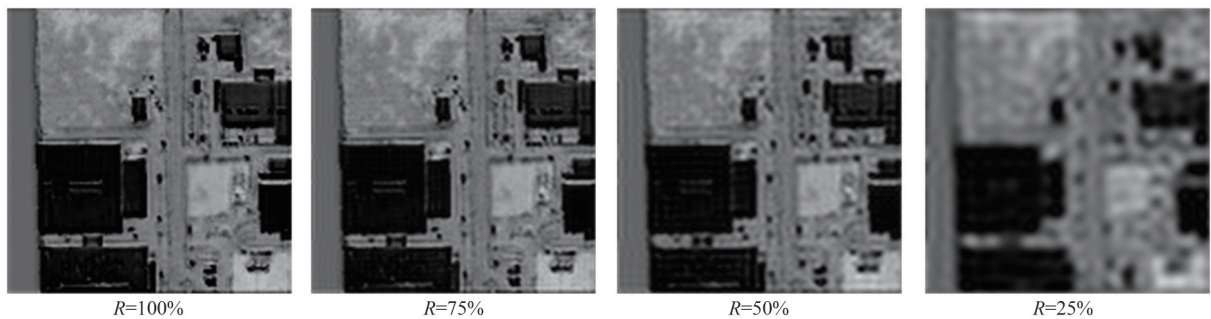


图7 频域筛选的特征可视化效果
Fig. 7 Visualization of frequency-domain filtered features

本次实验在3个数据集和3组教师—学生模型组合中进行，每个能量的取值都会生成9组数据。消融实验的详细结果记录于表7，其中 ΔPA 和 ΔmIoU 为实验结果减去基线模型精度的差值。最优指标的统计包括PA和mIoU两项，因此9组实验中每组都会产生2个最佳值，共计18个。低于基线的总数，则取决于每种方法下的9组数据：由于

统计指标也是PA和mIoU 2项，因此需要分别统计这2项指标低于基线指标的次数。总体而言，不同的能量级别对模型性能产生了显著影响，其中FDF（高）取得了最佳的综合性能。通过消融实验逐步移除高频分量（能量从完整降至低能级）发现，该方法在能量级别高时达到性能峰值，表明适度抑制高频分量是关键。

表7 不同能量级别下FDF方法的消融实验结果与精度统计

Table 7 Ablation experiment results and accuracy statistics of the FDF method under different energy levels						
方法	平均 $\Delta\text{PA}/\%$	ΔPA 标准差	平均 $\Delta\text{mIoU}/\%$	ΔmIoU 标准差	最优指标数/总数	指标低于基线数/总数
FDF(完整)	0.38	1.08	0.36	1.46	5/18	4/18
FDF(高)	0.55	0.69	0.66	1.19	7/18	4/18
FDF(中)	0.45	0.40	0.77	0.52	4/18	5/18
FDF(低)	0.00	0.86	-0.30	1.67	2/18	6/18

从性能增益分析来看，FDF（高）的设置在所有配置中脱颖而出，其平均性能提升最为显著。其在PA上的平均提升达到0.55%，且最佳指标数和指标数低于基线数也均为最佳，表明在保留大部分原始特征的同时引入少量扰动，最有利于模型学习到更具判别性的特征。相比之下，当能量下降至低时，平均 ΔmIoU 出现了负值（-0.30%），这意味着过于激进的特征丢弃策略反而损害了模型的整体分割精度。

从稳定性与鲁棒性分析来看，性能的稳定性是衡量方法可靠性的关键，可通过观察 ΔPA 和 ΔmIoU 的标准差来进行评估。如表7所示，FDF（高）的设置不仅在性能提升上领先，其 ΔPA 的标准差和 ΔmIoU 的标准差仅次于FDF（中），表明其在不同场景下的性能波动小，具有较好的鲁棒性。

4.1.2 λ 参数实验对比

本次实验在3个数据集和3组教师—学生模型

组合中进行，统计了所有方法（共8种，包括能量消融实验中的方法）在不同 λ 取值下的最佳PA数据。

在整体分析中，如图8所示， λ 取0.3时获得了最多的最优精度计数（32次），表明该参数设置在本研究任务中具有最佳的综合性能。 λ 取0.1时，最优精度计数为28次，与 λ 取0.3时接近，但其分布模式存在显著差异，即在特定模型中取得的指标较多。 λ 取0.2时表现相对较弱（12次），可能表明此参数区间存在性能敏感区域。

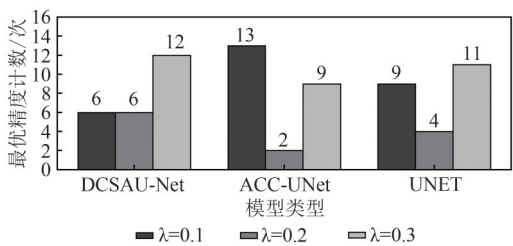


图8 模型对蒸馏权重系数 λ 的敏感性分析
Fig. 8 Sensitivity analysis of models to the distillation weight coefficient λ

在模型层面，如图8所示，DCSAU-Net模型呈现渐进式改进：随着 λ 值从0.1增加到0.3，性能从6次提升至12次，表明该模型能够从更大的正则化强度中获益。ACC-UNet模型对 λ 值最为敏感：在 λ 取0.1时表现最佳（13次），但在 λ 取0.2时急剧下降至2次，显示该模型对参数选择具有高度依赖性。UNET模型表现与ACC-UNet模型类似。

方法性能如图9所示，Phase方法在 λ 取0.3时表现突出（7次），但在取其他 λ 值时表现较差，特别是 λ 取0.2时最优精度计数为0次。IFVD方法展现出独特的稳定性，其在所有 λ 值下均保持3次最优精度计数，表明其对参数变化不敏感。FDF（低、中、高、完整）表现相似，在 λ 取0.1和 λ 取0.3时最优精度计数较多。

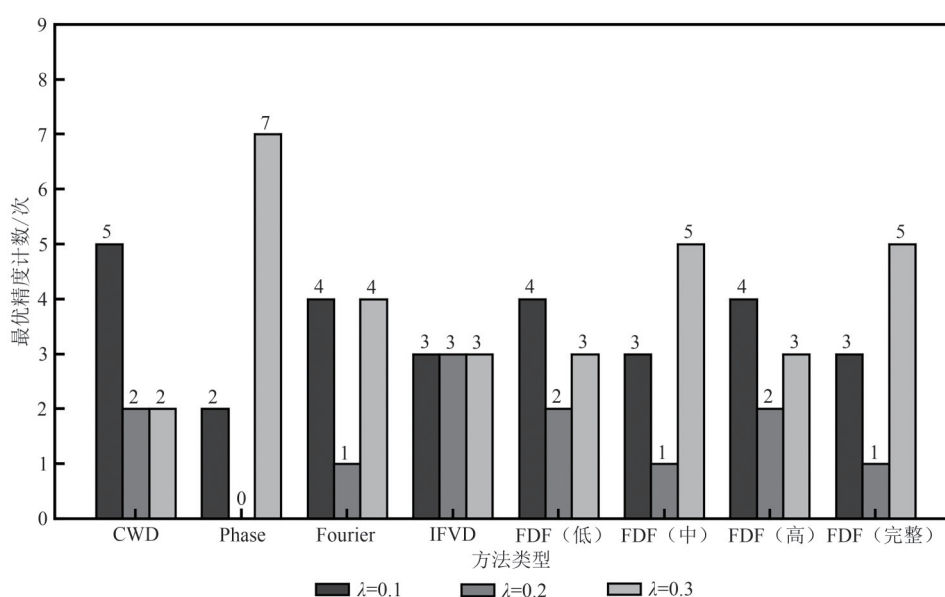


图9 方法层面对蒸馏权重系数 λ 的敏感性对比

Fig. 9 Sensitivity comparison of methods to the distillation weight coefficient λ

通过对不同 λ 参数设置的系统性评估，本研究发现模型和方法对正则化强度的敏感性存在显著差异。总体而言， λ 取0.3时在多数情况下提供了最优的性能平衡，共获得32次最优精度计数。在模型层面，ACC-UNet对 λ 参数的敏感性最高，性能随参数变化波动剧烈；DCSAU-Net的参数鲁棒性最优，在不同 λ 取值下性能稳定性最佳；UNET则表现出中等水平的参数敏感性与鲁棒性。在方法层面，Phase在较高正则化强度下表现优异；IFVD则显示出对参数变化的稳定性；FDF方法表现相似，在 λ 取0.1和 λ 取0.3时最优精度计数较多，表明性能的挖掘可能需要更多的取值尝试。这些发现为特定模型和方法选择适当的蒸馏损失权重参数提供了实证依据。

4.2 方法对比

将FDF方法与当前主流的先进知识蒸馏技术，包括基于像素关系蒸馏的方法IFVD（Wang等，

2020）、通道蒸馏方法CWD和两个基于频域处理的蒸馏方法Phase、Fourier，在PA、mIoU以及推理效率（平均单次耗时）3个维度上进行严格对比。所有方法均基于相同的实验设置，且FDF方法统一采用其FDF（高）（ $R=75\%$ ）配置，以确保对比的公平性与可复现性。

基于LoveDA数据集，U-Lite模型在FDF（高）方法蒸馏前后的建筑物注意力图与预测结果图（图10）对比表明，频域筛选方法蒸馏显著提升了学生模型的整体感知能力，其中教师模型是DCSAU-Net。在基线模型的预测中，学生模型的建筑物注意力热力图呈现碎片化的高响应，对应的预测掩膜边界残缺、内部空洞多；经全局蒸馏后，建筑物注意力热力图的能量集中于目标主体且平滑连贯，模型对建筑物的感知更加集中，对其边界的感知更加的完整，同时对误分类区域的注意力也更少。该结果验证了该方法能够将教师模型中的低频形状先验有效迁移至学生模型。

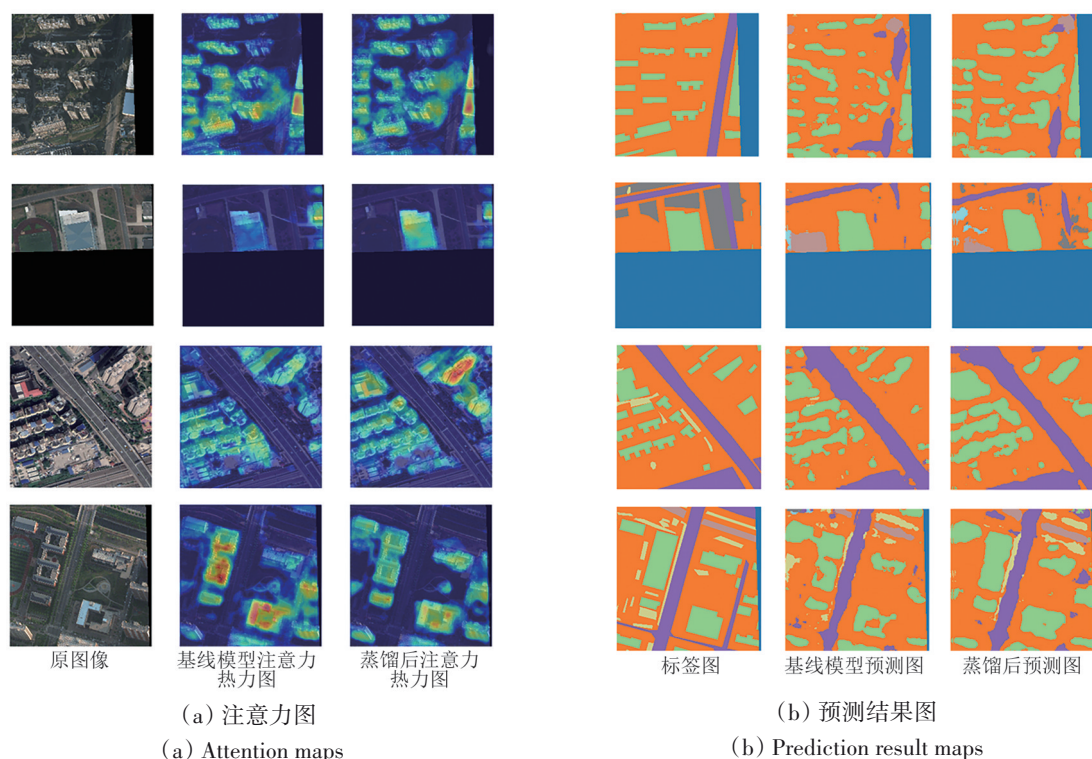


图 10 U-Lite模型在 LoveDA 数据集上的建筑物注意力图与预测结果图

Fig. 10 Attention maps and prediction result maps of the U-Lite model based on the LoveDA Dataset

如表 8 所示, 在 LoveDA 数据集上, FDF 方法表现出良好的竞争力。当以 DCSAU-Net 为教师模型时, FDF 的 PA 和 mIoU 指标分别达到 65.84% 和 53.81%, 均优于其他对比方法。在以 UNET 为教师模型时, FDF 也取得了具有竞争力的性能表现, 其 mIoU 达 53.70%, 显著高于基线学生模型。

在 Vaihingen 数据集上, FDF 方法保持了稳定的性能提升。与基线学生模型相比, FDF 在不同教师模型下均实现了稳定的性能增益, 其性能表现与 IFVD、Phase 等方法处于同一水平。

在 Potsdam 数据集上, FDF 方法的性能出现一定波动。这一现象可能与 Potsdam 数据集的图像特性有关: 在该数据中, 教师模型的表现较差, 甚至出精度指标现低于学生模型的现象。同时, 在该数据集的蒸馏实验中, 除 Phase 方法外, 其他方法的 PA 或 mIoU 均产生了负提升。此外, 经过观察, 该数据集具有较多的模糊边界与区域, 表明其数据质量相对于 LoveDA 和 Vaihingen 较差, 导致 FDF (高) 配置相对保守的滤波策略未充分发挥作用。

通过对全部 18 个实验设置 (3 个数据集×3 个教师模型×2 个指标) 进行统计分析 (表 9), 研究发现 FDF 方法在以下方面展现出独特优势。

在可靠性方面, FDF 方法的“指标低于基线数/总数”为 4/18, 低于 IFVD 和 Fourier 方法的 5/18。这一结果表明, FDF 方法具有较低的性能衰退风险。表 8 中 potsdam 数据集 DCSAU-Net 模型蒸馏结果显示, 当 Fourier 方法出现性能衰退时, 其 mIoU 下降幅度可达 2.37%, 而 FDF 方法的性能波动范围明显更小, 且平均 Δ PA 提升和平均 Δ mIoU 提升均位列前 3。

在稳定性方面, FDF 方法的 Δ mIoU 标准差为 1.03, 显著低于 CWD 方法 (1.25) 和 Fourier 方法 (1.34)。这表明 FDF 方法在不同实验设置下具有更稳定的性能表现, 其输出结果具有更好的可预测性。

在推理速度方面, FDF 方法平均单次推理耗时 16.63 ms, 仅慢于 CWD 方法 (2.99 ms) 与 Phase 方法 (6.51 ms), 但仅为 IFVD 方法 (46.58 ms) 与 Fourier 方法 (42.91 ms) 的 35.7% 与 38.8%。FDF 方法的额外延迟来源于一次 FFT、逐元素掩膜乘法与一次 IFFT, 其中 FFT 与 IFFT 的算法复杂度为 $O(N \log_2 N)$, 掩膜乘法为 $O(N)$; 并且掩膜可离线生成, 无需引入可学习参数, 从而以最小的计算代价实现了分割稳定性提升。

表 8 跨数据集蒸馏方法性能对比

Table 8 Cross-dataset performance comparison of knowledge distillation methods

方法	LoveDA 数据集		Vaihingen 数据集		Potsdam 数据集	
	PA/%	mIoU/%	PA/%	mIoU/%	PA/%	mIoU/%
T:DCSAU-Net	67.94	56.25	78.61	58.34	77.21	58.96
S:U-Lite	64.46	52.13	76.01	53.78	76.77	57.25
+IFVD	65.27	53.10	76.60	55.11	76.72	57.11
+CWD	65.12	51.43	76.49	55.34	78.17	58.58
+Phase	64.77	52.62	76.21	54.82	77.87	58.71
+Fourier	65.67	53.79	76.66	55.09	75.55	54.88
+FDF(ours)	65.84	53.81	76.35	54.44	76.25	55.62
T:ACC-UNet	68.65	57.27	76.80	56.10	74.95	56.29
S:U-Lite	64.46	52.13	76.01	53.78	76.77	57.25
+IFVD	65.30	53.32	76.19	54.52	77.91	58.21
+CWD	66.66	54.03	76.15	54.63	76.62	56.63
+Phase	64.80	52.26	76.23	55.19	77.53	58.47
+Fourier	65.56	53.26	76.38	54.27	76.61	56.37
+FDF(ours)	65.31	53.27	76.03	54.27	77.76	58.00
T:UNET	70.35	57.90	79.07	60.88	79.05	61.36
S:U-Lite	64.46	52.13	76.01	53.78	76.77	57.25
+IFVD	65.01	52.90	75.90	54.19	75.90	56.78
+CWD	66.28	54.54	76.63	55.09	76.84	56.22
+Phase	64.39	52.08	76.08	54.26	77.64	57.68
+Fourier	65.07	53.00	76.70	55.31	77.00	56.90
+FDF(ours)	65.90	53.70	76.61	55.17	76.58	57.14

注：加粗数字为评估指标的最佳结果,T代表教师模型,S代表学生模型,ours代表本文所提方法,采用FDF(高)配置。

表 9 蒸馏方法相对基线的精度提升与稳定性统计

Table 9 Accuracy gain and stability statistics of different distillation methods relative to baseline

方法	平均 Δ PA/%	Δ PA 标准差	平均 Δ mIoU/%	Δ mIoU 标准差	最优指标数/总数	指标低于基线数/总数
IFVD	0.34	0.62	0.62	0.61	1/18	5/18
CWD	0.80	0.63	0.82	1.25	6/18	4/18
Phase	0.42	0.40	0.40	0.56	5/18	2/18
Fourier	0.39	0.73	0.73	1.34	4/18	5/18
FDF(ours)	0.55	0.69	0.66	1.03	2/18	4/18

注：加粗字体为评估指标的最佳结果

表 10 蒸馏方法单次推理耗时对比

Table 10 Inference latency comparison of distillation methods

损失函数	平均时间/ms
Fourier	42.91
IFVD	46.58
CWD	2.99
Phase	6.51
FDF(ours)	16.63

综合分析表明，FDF 方法的优势在于其良好的稳定性与可靠性。虽然在部分实验设置下该方

法未能取得最优性能，但其较低的性能衰退风险和稳定的性能输出，使其在实际应用中具有独特价值。这一特性对于将知识蒸馏技术应用于实际遥感场景尤为重要，因为在实际部署中，方法的稳定性和可靠性往往与峰值性能同等重要。

5 结 论

本文提出的频域筛选蒸馏方法以极简的频域滤波机制，在资源受限场景下实现了对教师模型低频形状先验的高效迁移。实验表明，FDF 方法

在保持极低计算开销的同时,显著提升了学生模型的全局一致性,并稳健维持了区域结构完整性。尤其在遥感建筑物分割任务中,蒸馏后的注意力图从碎片化高响应转变为边界完整、主体集中的平滑分布,验证了本文方法在抑制高频噪声、强化几何感知方面的有效性。然而,该方法在极端复杂场景下的细节还原能力仍逊于部分高复杂度蒸馏策略,提示其更适用于对全局一致性要求高、对边缘细节容忍度适中的实际部署场景。未来可通过动态掩膜尺度调整或与其他蒸馏策略互补融合,进一步拓展其性能边界。

参考文献(References)

- Badrinarayanan V, Kendall A and Cipolla R. 2017. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12): 2481-2495 [DOI: 10.1109/TPAMI.2016.2644615]
- Chen L C, Papandreou G, Kokkinos I, Murphy K and Yuille A L. 2018. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4): 834-848 [DOI: 10.1109/TPAMI.2017.2699184]
- Dinh B D, Nguyen T T, Tran T T and Pham V T. 2023. 1M parameters are enough? A lightweight CNN-based model for medical image segmentation. *arXiv preprint arXiv:2306.16103* [DOI: 10.48550/arXiv.2306.16103]
- Gotmare A, Keskar N S, Xiong C M and Socher R. 2018. A closer look at deep learning heuristics: learning rate restarts, warmup and distillation. *arXiv preprint arXiv:1810.13243* [DOI: 10.48550/arXiv.1810.13243]
- Gou J P, Yu B S, Maybank S J and Tao D C. 2021. Knowledge distillation: a survey. *International Journal of Computer Vision*, 129(6): 1789-1819 [DOI: 10.1007/s11263-021-01453-z]
- Hinton G, Vinyals O and Dean J. 2015. Distilling the knowledge in a neural network. *arXiv preprint arXiv: 1503.02531* [DOI: 10.48550/arXiv.1503.02531]
- Huang Z H, Yang S Z, Lin W, Ni J, Sun S L, Chen Y W and Tang Y. 2022. Knowledge distillation: a survey. *Chinese Journal of Computers*, 45(3): 624-653 (黄震华, 杨顺志, 林威, 倪娟, 孙圣力, 陈运文, 汤庸. 2022. 知识蒸馏研究综述. *计算机学报*, 45(3): 624-653) [DOI: 10.11897/SP.J.1016.2022.00624]
- Ibtehaz N and Kihara D. 2023. ACC-UNet: a completely convolutional UNet model for the 2020s. *arXiv preprint arXiv: 2308.13680* [DOI: 10.48550/arXiv.2308.13680]
- Ji S P. 2018. *An Introduction to Intelligent Photogrammetry Introduction to Intelligent Photogrammetry*. Beijing: Science Press (季顺平. 2018. *智能摄影测量学导论*. 北京: 科学出版社)
- Kingma D P and Ba J L. 2017. Adam: a method for stochastic optimization. *arXiv preprint arXiv: 1412.6980* [DOI: 10.48550/arXiv.1412.6980]
- Li Q Q, Jin S Y and Yan J J. 2017. Mimicking very efficient network for object detection//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE: 7341-7349 [DOI: 10.1109/CVPR.2017.776]
- Liu L Y, Wang Z H, Phan M H, Zhang B W, Ge J C and Liu Y F. 2024. BPKD: boundary privileged knowledge distillation for semantic segmentation//2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa, HI, USA: IEEE: 1051-1061 [DOI: 10.1109/WACV57701.2024.00110]
- Liu Y F, Shu C Y, Wang J D and Shen C H. 2023. Structured knowledge distillation for dense prediction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(6): 7035-7049 [DOI: 10.1109/TPAMI.2020.3001940]
- Long J, Shelhamer E and Darrell T. 2015. Fully convolutional networks for semantic segmentation//2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE: 3431-3440 [DOI: 10.1109/CVPR.2015.7298965]
- Ma Y and Kelibieke G. 2023. Research review of image semantic segmentation method in high-resolution remote sensing image interpretation. *Journal of Frontiers of Computer Science and Technology*, 17(7): 1526-1548 (马妍, 古丽米拉·克孜尔别克. 2023. 图像语义分割方法在高分辨率遥感影像解译中的研究综述. *计算机科学与探索*, 17(7): 1526-1548) [DOI: 10.3778/j.issn.1673-9418.2211015]
- Mao A Q, Mohri M and Zhong Y T. 2023. Cross-entropy loss functions: theoretical analysis and applications. *arXiv preprint arXiv: 2304.07288* [DOI: 10.48550/arXiv.2304.07288]
- Nussbaumer H J. 1981. The fast fourier transform//Nussbaumer H J, ed. *Fast Fourier Transform and Convolution Algorithms*. Berlin: Springer: 80-111 [DOI: 10.1007/978-3-662-00551-4_4]
- Proakis J G, Manolakis D G. 2007. *Digital Signal Processing: Principles, Algorithms and Applications*. 4th ed. Upper Saddle River: Prentice Hall.
- Rahaman N, Baratin A, Arpit D, Draxler F, Lin M, Hamprecht F, Bengio Y and Courville A. 2019. On the spectral bias of neural networks. *arXiv preprint arXiv: 1806.08734* [DOI: 10.48550/arXiv.1806.08734]
- Romero A, Ballas N, Kahou S E, Chassang A, Gatta C and Bengio Y. 2015. FitNets: hints for thin deep nets. *arXiv preprint arXiv: 1412.6550* [DOI: 10.48550/arXiv.1412.6550]
- Ronneberger O, Fischer P and Brox T. 2015. U-Net: convolutional networks for biomedical image segmentation//18th International Conference on Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015. Munich: Springer International Publishing: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]
- Shi M, Zheng C K, Yi Q M, Weng J and Luo A W. 2025. Knowledge distillation in Fourier frequency domain for dense prediction. *IEEE Signal Processing Letters*, 32: 296-300 [DOI: 10.1109/LSP.2024.3515795]
- Shu C Y, Liu Y F, Gao J F, Yan Z and Shen C H. 2021. Channel-wise

- knowledge distillation for dense prediction//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE: 5291-5300 [DOI: 10.1109/ICCV48922.2021.00526]
- Wang Y K, Zhou W, Jiang T, Bai X and Xu Y C. 2020. Intra-class feature variation distillation for semantic segmentation//16th European Conference on Computer Vision – ECCV 2020. Glasgow: Springer: 346-362 [DOI: 10.1007/978-3-030-58571-6_21]
- Xu Q, Ma Z C, He N and Duan W T. 2023. DCSAU-Net: a deeper and more compact split-attention U-Net for medical image segmentation. *Computers in Biology and Medicine*, 154: 106626 [DOI: 10.1016/j.compbiomed.2023.106626]
- Yang C G, Zhou H L, An Z L, Jiang X, Xu Y J and Zhang Q. 2022. Cross-image relational knowledge distillation for semantic segmentation//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE: 12309-12318 [DOI: 10.1109/CVPR52688.2022.01200]
- Yao L, Liu F, Zhang C Y, Ou Z Q and Wu T. 2025. Domain-invariant progressive knowledge distillation for UAV-based object detection. *IEEE Geoscience and Remote Sensing Letters*, 22: 7500705 [DOI: 10.1109/LGRS.2024.3492187]
- Zhang L F, Chen X, Tu X B, Wan P F, Xu N, and Ma K S. 2022. Wavelet knowledge distillation: towards efficient image-to-image translation//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE: 12454-12464 [DOI: 10.1109/CVPR52688.2022.01214]
- Zhang Y, Huang T, Liu J M, Jiang T, Cheng K and Zhang S H. 2024. FreeKD: knowledge distillation via semantic frequency prompt//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE: 15931-15940 [DOI: 10.1109/CVPR52733.2024.01508]
- Zhao H S, Shi J P, Qi X J, Wang X G and Jia J Y. 2017. Pyramid scene parsing network//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE: 6230-6239 [DOI: 10.1109/CVPR.2017.660]
- Zhao Y, Feng Z L, Wang H Q and Song M L. 2022. Context correlation distillation for lip reading. *Journal of Computer-Aided Design and Computer Graphics*, 34(10): 1559-1566 (赵雅, 冯尊磊, 王慧琼, 宋明黎. 2022. 针对唇语识别的上下文相关性蒸馏方法. *计算机辅助设计与图形学学报*, 34(10): 1559-1566) [DOI: 10.3724/sp.j.1089.2022.19723]

Distillation method for remote sensing image semantic segmentation via frequency–domain filtering

DENG Xingsheng, LI Qian

School of Aeronautical Engineering, Changsha University of Science and Technology, Changsha 410114, China

Abstract: With the increasing deployment of remote sensing semantic segmentation models on edge devices, achieving an appropriate balance between segmentation accuracy and inference efficiency under limited computational resources has become a critical challenge. This study addresses high-frequency noise interference in front-end encoder features during knowledge distillation and aims to design a reliable and efficient distillation framework for remote sensing semantic segmentation in resource-constrained edge computing environments.

This study proposes a Frequency Domain Filtering Distillation (FDF) framework. In the Fourier domain, a central square mask is applied to the teacher and student features, enabling selective transfer of structurally meaningful low-frequency semantic knowledge while suppressing irrelevant high-frequency noise components. For the distillation architecture, U-Lite, a lightweight network with compact parameters and fast inference that has been validated in medical imaging tasks, is used as the student model to ensure edge-device efficiency. Comprehensive experiments on three widely used remote sensing semantic segmentation datasets (LoveDA, Vaihingen, Potsdam) are conducted to evaluate the proposed method.

Experimental results indicate that FDF provides competitive performance gains and stronger training and inference stability for remote sensing semantic segmentation. Specifically, FDF substantially reduces the risk of performance degradation: the standard deviation of $\Delta mIoU$ is only 1.03, which is lower than that of CWD (1.25) and Fourier (1.34), and it avoids the maximum 2.37% performance drop observed in other distillation methods. In terms of efficiency, the FDF-based model requires only 3.3G FLOPs of computational overhead and reaches an inference speed of 16.63 ms/frame, which is about 2.5 times faster than the IFVD and Fourier methods.

The proposed FDF framework effectively mitigates high-frequency noise interference during feature distillation for remote sensing semantic segmentation models, and its integration with the lightweight U-Lite model further ensures model efficiency. By balancing segmentation accuracy, model stability, and inference efficiency, FDF provides an effective, reliable, and efficient technical solution for deploying remote sensing semantic segmentation in resource-constrained edge computing environments.

Key words: deep learning network, distillation of knowledge, semantic segmentation, feature extraction, fast Fourier transform

Supported by Open Topic of Hunan Geospatial Information Engineering and Technology Research Center (No. HNGIET2025001); Natural Science Foundation of Hunan Province, China (No. 2024JJ8335)