DSFC-Net：用于乡村道路提取的双编码器空间与频率协同感知网络

Zhengbo Zhang、Yihe Tian、Wanke Xia、陈林、Yue Sun、Kun Ding、Ying Wang、Bing Xu 和向世明

arXiv:2602.01278v1 [cs.CV] 1 Feb 2026

摘要——**从高分辨率遥感影像中准确提取乡村道路，对于基础设施规划和可持续发展至关重要。然而，由于多种因素，这项任务在乡村场景中面临独特挑战，包括地表材料多样导致的类内差异大、类间可分性低，频繁的植被遮挡破坏空间连续性，以及狭窄的道路宽度加剧检测难度。现有方法主要针对结构化的城市环境进行优化，往往忽视这些独特特征，因此在此类场景中表现不佳。为应对这些挑战，我们提出 DSFC-Net，这是一种协同融合空间信息与频域信息的双编码器框架。具体而言，我们采用卷积神经网络分支捕捉细粒度的局部道路边界和短程连续性，同时引入一种新颖的空间-频率混合 Transformer（SFT），以稳健地建模植被遮挡情况下的全局拓扑依赖关系。不同于易受频率偏差影响的标准注意力机制，SFT 引入了跨频率交互注意力（CFIA）模块，通过拉普拉斯金字塔策略显式分离高频和低频信息。这种设计能够动态交互空间细节与频率感知全局上下文，从而有效保持狭窄道路的连通性。此外，我们提出通道特征融合模块（CFFM），通过自适应地重新校准通道特征响应来连接两个分支，将局部纹理与全局语义无缝融合，实现精准分割。在 WHU-RuR+、DeepGlobe 和马萨诸塞州数据集上的全面实验验证了 DSFC-Net 优于当前最先进方法。值得注意的是，在具有挑战性的 WHU-RuR+ 数据集上，我们的方法取得了 69.93% 的 F1 分数和 53.77% 的交并比，彰显了其在道路狭窄、稀疏且受遮挡的农村场景中的稳健性。代码可在 https://github.com/ZhengboZhang/DSFC-Net 获取。**

索引词—**道路提取、遥感、语义分割、农村基础设施、VisionTransformer、卷积神经网络**

Zhengbo Zhang、陈林和向世明就职于中国科学院大学人工智能学院（中国北京，邮编100049），同时就职于中国科学院自动化研究所多模态人工智能系统全国重点实验室（中国北京，邮编100190）（电子邮箱：zhangzhengbo2025@ia.ac.cn；chenlin2024@ia.ac.cn；smxiang@nlpr.ia.ac.cn）。

Yihe Tian和Bing Xu就职于清华大学地球系统科学系（中国北京，邮编100084）（电子邮箱：tianyh25@mails.tsinghua.edu.cn；bingxu@tsinghua.edu.cn）

Wanke Xia就职于清华大学深圳国际研究生院（中国深圳，邮编518055）（电子邮箱：xwk25@mails.tsinghua.edu.cn）。

Yue Sun 任职于中国农业大学信息与电气工程学院，中国北京 100083（电子邮箱：ndsy@cau.edu.cn）。

Kun Ding 和 Ying Wang 任职于中国科学院自动化研究所多模态人工智能系统全国重点实验室，中国北京 100190（电子邮箱：kun.ding@ia.ac.cn；ywang@nlpr.ia.ac.cn）。

通讯作者：*Kun Ding* 和 *Bing Xu*；这些作者贡献均等：*Zhengbo Zhang* 和 *Yihe Tian*

I. 引言

乡村道路构成农业基础设施的骨干，连接偏远产区与外部市场。促进这种连通性是将农村人口融入现代经济的基本要求 [1]。这些道路网络的战略意义还体现在：农业食品系统支撑着全球近一半人口的生计，而这些人口主要居住在农村地区 [2], [3]。除了作为单纯的物理通道外，这些道路还是信息、技术和资本交换的重要动脉。因此，持续监测乡村道路网络对于实现可持续发展目标（可持续发展目标）至关重要 [4],并可作为消除贫困（可持续发展目标1）、终结饥饿（可持续发展目标2）以及建设韧性基础设施（可持续发展目标9）的关键指标 [5]。

准确的道路提取是有效开展路网监测的基础。传统方法主要依赖手工特征和几何建模。数学形态学 [6], [7]和基于知识的推理 [8], [9]等技术被广泛用于根据线性和宽度一致性识别类道路结构。然而，这些方法往往需要手动进行参数调优，且缺乏适应复杂环境所需的泛化能力 [10]。相比之下，在过去十年中，深度学习已成为该任务的主流范式。早期方法主要依靠卷积神经网络（CNN）提取图像特征 [11]–[13]。然而，由于局部感受野的限制，卷积神经网络往往难以捕获长程依赖。为解决这一问题，近期研究已转向 Transformer。这类架构利用自注意力机制有效建模全局上下文 [14]–[16]。最近，研究重点转向利用专门设计的模块，应对遮挡和拓扑连通性等复杂几何挑战 [17]–[19]。

然而，现有的大多数方法都是为城市场景设计的，在乡村场景中往往表现不佳。这种性能差距源于乡村道路网络与城市道路网络之间的显著差异。具体而言，如图所示，乡村道路面临三项独特挑战。 1。首先，它们具有较大的类内差异和较低的类间可分性（图 1（a））：多样的地表材料导致道路类别内部的光谱特征各异，而未铺装道路很容易与裸土、田埂等背景要素混淆。其次，乡村道路经常被树木和作物等植被遮挡，导致其空间连续性遭到破坏，也增加了准确提取的难度（图



图1. 从遥感影像中提取乡村道路面临的三项主要挑战示意图。(a) 类内差异大且类间可分性低，不同的道路表面与周围裸土或田野融为一体。(b) 植被遮挡：树木和作物会破坏道路连续性。(c) 道路狭窄，导致检测困难并造成前景-背景不平衡。

间连续性，并使准确提取更加困难（图 1（b））。最后，乡村道路通常较为狭窄，因此更难检测；此外，前景与背景要素之间的不平衡也会妨碍遥感影像中的道路检测（图 1（c））[20]。

为应对这些挑战，我们提出了 DSFC-Net，这是一种双编码器框架，可协同发挥空间域处理和频域处理各自的互补优势。具体而言，CNN分支能够有效提取局部道路特征，以减轻由地表材料多样性引起的分类误差；而配备跨频率交互注意力（CFIA）的空间-频率混合 Transformer（SFT）分支，则显式解耦频率分量，以建模全局拓扑依赖关系，并从复杂的背景噪声中分离道路边界。为连接这两个分支，我们引入通道特征融合模块（CFFM），自适应地重新校准通道特征响应，从而无缝对齐局部细节与全局语义。在 WHU-RuR+、DeepGlobe 和马萨诸塞州数据集上开展的全面实验验证了 DSFC-Net 达到最先进性能，展现出解析狭窄、稀疏和被遮挡乡村道路的卓越鲁棒性。

本文的主要贡献总结如下：

1) 我们提出了 DSFC-Net，这是一种专为从遥感图像中提取乡村道路而设计的新型双编码器网络。该网络对空间域和频域信息进行建模，以应对乡村道路提取中的独特挑战。

2）我们提出了一种空间-频率混合 Transformer（SFT），并引入跨频率交互注意力（CFIA）模块。通过分解高频和低频特征，该模块能够捕获长程依赖，并解决狭窄道路和空间不连续性带来的挑战。此外，通道特征融合模块（CFFM）通过自适应通道权重重新校准，连接卷积神经网络与 Transformer 分支。

3) 在三个具有不同乡村环境复杂性和稀疏道路密度的公共数据集上开展的大量实验表明，DSFC-Net 达到了最先进的性能。值得注意的是，我们的方法在 WHU-RuR+ 数据集上取得了 69.93% 的 F1分数和 53.77% 的交并比，彰显了其在具有挑战性的乡村场景中的卓越准确性和鲁棒性。

II. 相关工作

A. 道路提取方法

深度学习从根本上改变了道路提取的范式，其发展历程从基础的卷积神经网络（CNN）演进到融合注意力机制和拓扑感知模块的复杂架构。早期方法主要致力于优化编码器-解码器结构，以增强特征表征。例如，ResU-Net [12] 将残差块集成到 U-Net主干网络中，以促进深层特征学习；而具有开创性的 D-LinkNet [11]则采用空洞卷积，在不降低空间分辨率的情况下显著扩大感受野。然而，这些基于卷积神经网络的方法受限于其局部感受野。在类内差异较大等乡村场景中，它们往往难以从复杂背景中区分细微的道路特征，导致提取结果碎片化。

为克服卷积神经网络在捕获长程依赖方面的固有限制，后续研究引入了注意力机制和全局感知模块。GAM-SNet [21]利用多尺度残差学习捕获复杂的空间与通道依赖。Swin-Unet [15]标志着向基于Transformer的架构转变，并利用移位窗口自注意力有效地建模全局上下文。最近，研究人员将重点转向高效的上下文建模。CGCNet [22]引入了紧凑全局上下文感知模块，以较低的计算复杂度捕获长程依赖。CFRNet [23] 通过级联架构融合多个尺度的丰富特征，提升了分割精度；而 FRCFNet [24] 提出了一种分组多维特征重组模块，利用条带卷积融合不同方向上的上下文信息。尽管标准注意力机制能够有效进行全局建模，但它们会优先关注低频信号，因而存在频率偏差。这会导致边界过度平滑，并使模型难以在复杂的农村场景中保留细粒度道路细节。

研究重点进一步转向细化边界和保持拓扑连通性，尤其是为了解决道路断裂和遮挡问题。BT-RoadNet [25] 采用由粗到细策略来改善边界质量，而 Sat2Graph [17] 则直接提取道路图，以确保几何正确性。为应对复杂的遮挡，OARENet [26]设计了一种专门的遮挡感知

解码器，采用条带卷积来恢复密集城市环境中的道路连通性。此外，DENet[18]通过将道路边缘检测器与方向空间注意力相结合，强调方向特征的重要性，以追踪狭窄且细长的道路。最后，SegRoadv2[19]创新性地将可变形自注意力与可分组可变形卷积相结合，以自适应地追踪道路轨迹。然而，大多数拓扑感知方法依赖于针对结构化城市环境定制的几何先验。因此，它们通常难以泛化到乡村道路蜿蜒且不规则的几何形态，并且难以在无结构的冠层遮挡下恢复连续性。

尽管这些方法在城市数据集上占据主导地位，但应用于乡村影像时，其性能会急剧下降。面对复杂的乡村现实，这些方法难以实现良好的泛化；极强的纹理异质性和稀疏的道路连通性，使现有的、以城市为中心的先验知识失效。为应对这些挑战，我们提出的 DSFC-Net 采用双编码器策略，协同融合卷积神经网络和 Transformer 分支。通过空间-频率混合Transformer明确分离高频边界细节与低频拓扑依赖关系，我们的方法有效减轻了背景干扰，并保持了狭窄、蜿蜒农村道路的连续性。

B. 乡村道路提取

与结构规整的城市道路网络相比，乡村道路面临独特且严峻的挑战：它们通常窄得多，由于路面材料多样而具有较高的类内方差，并且常被植被严重遮挡。高质量数据集是数据驱动方法的基石，但经典数据集主要聚焦于城市区域，导致模型难以泛化到复杂的乡村环境。为弥补这一数据缺口，近期研究致力于构建乡村专用数据集。值得注意的是，WHU-RuR+ [20] 提供了一个大规模数据集，覆盖八个国家、总面积达 6,866 km2，专门用于呈现全球农村地区异质的背景。同样，WHU-CR数据集 [27] 旨在捕捉中国乡村地区特有的多样地理与环境情境。

尽管这些数据集相继问世，专用于乡村道路提取的算法仍然十分稀缺。在少数开创性工作中，R2-Net [28]提出了一种逆向细化机制，通过增强边界可分性来提取狭窄道路。为进一步缓解狭窄目标在深层编码过程中的特征损失，NANet [27] 采用了由粗到细框架，并融入前景-背景协同机制。

然而，农村场景的研究仍处于起步阶段。现有研究数量有限，往往难以同时应对农村遥感图像中固有的极端尺度变化和高频噪声。因此，迫切需要更稳健、专用的架构来应对这些复杂场景，而本研究旨在填补这一空白。

III. 方法

A. 总体架构

DSFC-Net的总体架构如图2所示，其中我们同时使用CNN分支和Transformer 分支，分别有效提取空间特征和频率特征。给定输入图像 $x \in \mathbb{R}^{H\times W\times 3}$，DSFC-Net首先应用主干块提取低层特征图$F_0 \in \mathbb{R}^{H\times W\times C}$，其中 $H\times W$表示空间维度， C 为通道数。接下来，特征图 F_0 依次通过四阶段编码器结构。在各阶段之间，CNN和Transformer分支共享一个下采样模块（层归一化以及一个 2×2、步长为2的卷积层）。在每个阶段内部，输入特征分别通过 N_i 个卷积神经网络块和 L_i 个SFT层提取特征图 F_C^i和 F_T^i。这两种特征由通道特征融合模块（CFFM）融合，得到低层特征图 F_l^i，其中 $i \in \{1,2,3,4\}$。随后，在解码器中应用转置卷积进行上采样，并通过跳跃连接将低层特征图 F_l^i 与高层特征图 F_h^i 拼接。最后，通过转置卷积将特征图放大至原始尺寸的4倍，并由分割头得到最终分割掩膜。

DSFC-Net包含两个关键组件：空间-频率混合Transformer（SFT）作为Transformer分支，用于感知全局特征；通道特征融合模块（CFFM）则用于融合两个分支的特征图。我们将在以下小节中介绍这两个组件。

B. 空间-频率混合*Transformer*（*SFT*）

尽管CNN分支擅长提取局部细节特征，乡村道路提取任务仍常面临道路结构连续且狭长等挑战，因此需要强大的全局上下文建模能力。为此，我们提出了空间-频率混合 Transformer（SFT），用于捕获乡村道路场景中的复杂空间结构和频率特征。该模块由三个组件构成：空间上下文聚合器（SCA）、跨频率交互注意力（CFIA）机制以及多尺度前馈网络（MFFN）。这些组件将局部空间感知能力与全局频率表示特性相结合，增强了对道路边缘、纹理变化和远距离上下文信息的表达能力。

如图所示2，SFT 采用并行架构，同时处理空间信息和频率信息。给定输入特征 $F_l^i \in \mathbb{R}^{H\times W\times C}$，首先对其进行层归一化（LN），然后分别送入空间上下文聚合器（SCA）和跨频率交互注意力（CFIA）分支。将这两个分支的输出聚合后，经过逐点卷积处理，并通过残差连接与输入相加。该过程可表示为：

$$\tilde{F_l^i} = F_l^i + \mathrm{PW}_{conv}(\mathrm{SCA}(\mathrm{LN}(F_l^i)) + \mathrm{CFIA}(\mathrm{LN}(F_l^i))) \quad (1)$$

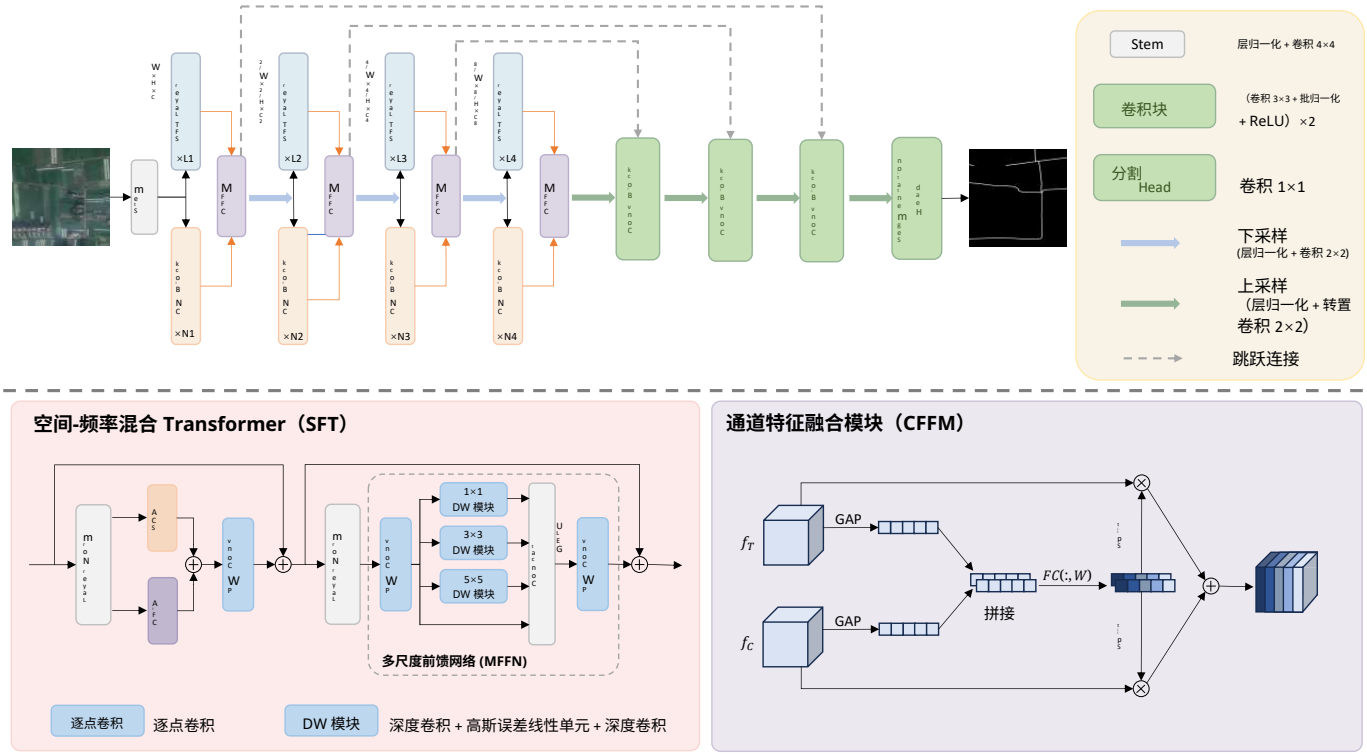

图 2 2 . 所提出的 DSFC-Net 总体架构。DSFC-Net 由两个并行编码器组成：一个采用 ConvNeXt-v 的 CNN 分支和一个采用我们提出的 SFT 的 Transformer 分支。CFFM 用于融合两个编码器输出的特征图。[29]

然后，将聚合后的特征送入多尺度前馈网络（MFFN），以进一步提取多尺度语义信息。随后，通过残差连接处理 MFFN 的输出，得到 SFT 的最终输出。

1）空间上下文聚合器（*SCA*）：道路目标的尺度变化极大，但固定大小的感受野不足以同时捕获这些多样的几何结构，因此难以维持长距离道路的连通性。为弥补这一不足，我们引入了空间上下文聚合器（SCA）模块。该模块采用具有不同膨胀率的并行深度空洞卷积，以捕获并聚合多尺度上下文线索。

第一个分支采用膨胀率为 $d = 1$ 的标准 3×3 深度卷积。该分支着重提取细粒度局部细节，并在紧凑的邻域内保持道路表面的连续性。第二个分支采用膨胀率为 $d = 2$ 的 3×3 深度卷积。该分支在不降低空间分辨率的情况下扩大感受野，从而捕获更广泛的上下文信息，这对于识别道路拓扑结构和连接不连续的道路路段至关重要。随后，拼接后的特征经由融合卷积模块处理，以促进跨通道交互并降低维度。

2）跨频率交互注意力（*CFIA*）：在道路分割中，语义信息存在于正交频域中：拓扑结构和连通的道路区域呈现平滑的强度变化，主要由低频分量表征，而道路边界、车道标线和路面裂缝涉及强烈的强度变化，因此对应高频分量。Transformer中的标准自注意力机制作用于混合信号，往往会导致频率偏差，即高频细节被占主导地位的低频模式所淹没。这种偏差会造成分割边界模糊和道路拓扑结构断裂。为解决这一局限，我们提出了跨频率交互注意力（CFIA）模块。如图3所示，我们的CFIA模块将输入特征$\mathbf{X} \in \mathbb{R}^{H \times W \times C}$显式解耦为两个不同的频率分量。

我们不再依赖计算开销较大的傅里叶变换，而是采用一种更高效、无需参数的拉普拉斯金字塔分解策略，以便有针对性地处理全局上下文。低频特征 $\mathbf{X}_L$ 通过步幅为 s 的最大池化操作获得，该操作充当结构低通滤波器，在保留主要道路布局的同时有效抑制噪声，并降低空间分辨率以实现高效的全局建模。高频分量$\mathbf{X}_H$包含边缘和纹理信息，由上采样后的低频特征与原始输入之间的残差差值得到。该分解可表示为：

$$\mathbf{X}_L = MaxPool_s(\mathbf{X}), \tag{2}$$

$$\mathbf{X}_H = UpSample(\mathbf{X}_L) - \mathbf{X} \tag{3}$$

CFIA 的核心创新在于其交互策略。我们不在各自独立的

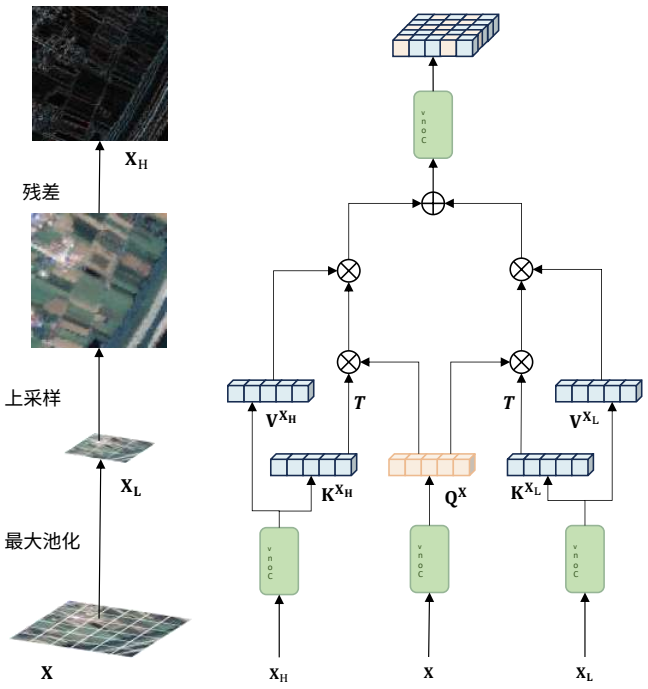

图 3. 所提出的跨频率交互注意力（CFIA）结构。左侧：拉普拉斯金字塔分解策略示例。右侧：CFIA 的详细结构，其中 $\mathbf{X}_L$ 和 $\mathbf{X}_H$ 由拉普拉斯金字塔分解策略获得。

频域内计算自注意力，而是将卷积投影后的原始特征$\mathbf{X}$用作语义查询锚点。此设计确保检索到的频率特定信息始终与原始语义上下文保持一致。我们从$\mathbf{X}$生成查询向量$\mathbf{Q}^{\mathbf{X}}$，同时从分解后的分量中获取键向量和值向量：高频分支的$\mathbf{K}^{\mathbf{X}_H}, \mathbf{V}^{\mathbf{X}_H}$，以及低频分支的$\mathbf{K}^{\mathbf{X}_L}, \mathbf{V}^{\mathbf{X}_L}$。这种双分支结构有助于实现互补的特征细化。高频分支侧重于局部细节恢复。通过查询高频键值对，模型能够关注突变区域，从而细化道路边界。低频分支利用$\mathbf{X}_L$的较低分辨率来捕获长程依赖关系和全局拓扑结构。接下来，我们分别计算两个分支的多头注意力：

$$\text{Attn}_i^{high} = Softmax\left(\frac{\mathbf{Q}_i^{\mathbf{X}} \cdot (\mathbf{K}_i^{\mathbf{X}_H})^T}{\sqrt{d_k}}\right) \cdot \mathbf{V}_i^{\mathbf{X}_H}, \quad (4)$$

$$\text{Attn}_i^{low} = Softmax\left(\frac{\mathbf{Q}_i^{\mathbf{X}} \cdot (\mathbf{K}_i^{\mathbf{X}_L})^T}{\sqrt{d_k}}\right) \cdot \mathbf{V}_i^{\mathbf{X}_L} \quad (5)$$

其中 i 表示头的索引，d_k 表示每个头的维度。随后，所有头的输出被拼接。最后，高频特征与低频特征被聚合，并投影回原始维度，从而得到全面的频率感知表示。这一全面的频率感知表示有效地融合了全局拓扑一致性来自低频特征，以及来自高频特征的精确边界细节。

3）多尺度前馈网络（*MFFN*）：对于道路分割任务，区分道路表面与背景在很大程度上依赖于局部纹理连续性和不同的结构尺度，因此我们提出多尺度前馈网络（MFFN），以从各种感受野中提取高维特征的局部关系。

给定前一模块输出的输入特征，我们首先使用逐点卷积降低通道维度，生成中间特征图。为捕获多样的空间模式，随后将中间特征图输入具有不同卷积核大小的并行深度卷积块进行处理。小卷积核分支保留细粒度的像素级信息，而大卷积核分支则感知更大范围的结构模式，例如宽阔的车道标线。此外，为促进高效的梯度流和特征复用，我们加入一条恒等路径，直接传递扩展后的特征。

C. 通道特征融合模块（*CFFM*）

在我们的双分支架构中，采用简单的逐元素求和或拼接等直接融合策略，会将来自不同领域的特征视为同等重要，从而忽略其信息量的动态变化。为此，我们提出了通道特征融合模块（CFFM），其结构如图2所示。受压缩与激励（SE）机制的启发 [30]，CFFM旨在显式建模两个分支之间的相互依赖关系，并自适应地重新校准通道级特征响应。

给定来自两个并行分支的输入特征图 $F_C \in \mathbb{R}^{H\times W\times C}$ 和 $F_T \in \mathbb{R}^{H\times W\times C}$，融合过程首先通过联合压缩操作聚合全局空间信息。我们分别对两个输入应用全局平均池化（GAP），并拼接其通道描述符，形成整体上下文向量$\mathbf{z}_{joint} \in \mathbb{R}^{2C}$。该联合向量概括了整个双分支系统的全局分布。为捕获两个分支间通道的非线性交互，$\mathbf{z}^{joint}$ 被输入多层感知机瓶颈结构。随后，通过 Sigmoid 门控从拼接后的表示中得到注意力权重$\mathbf{s} = \sigma(MLP(\mathbf{z}^{joint}))$，从而有效反映跨分支信息的重要性分数。

最后，生成的注意力向量 $\mathbf{s}$ 被拆分为两个子向量 $\mathbf{s}_C \in \mathbb{R}^C$ 和 $\mathbf{s}_T \in \mathbb{R}^C$，分别对应两个输入分支。这些权重作为软门控，对原始输入特征进行重新校准。最终融合输出 F_l 通过对 The final fused output F_l 重新校准后的特征进行加权求和得到。

IV. 实验与分析

A. 数据集

在我们的实验中，我们使用三个数据集来评估所提出的方法：WHU-RuR+ [20]、DeepGlobe [34],和马萨诸塞州 [35]。这些数据集非常适合模拟农村

表 I 在 WHU-RUR+ 数据集上，将所提出的 DSFC-Net 与最先进的道路提取方法进行定量比较，其中粗体表示最佳结果，下划线表示次佳结果。

方法	来源	精确率 (%)	召回率（%）	F1分数（%）	交并比（%）
U-Net [31]	$MICCAI_{2015}$	76.18	58.87	66.42	49.72
D-LinkNet [11]	$CVPR_{2018}$	72.46	65.88	69.01	52.68
SegFormer [32]	$NIPS_{2021}$	75.91	60.54	67.36	50.78
CFRNet [23]	$GRSL_{2024}$	79.01	57.15	66.33	49.62
FRCFNet [24]	$GRSL_{2024}$	81.18	48.83	60.98	43.86
OARENet [26]	$TGRS_{2024}$	77.09	59.18	66.96	50.33
C^2Net [33]	$TGRS_{2024}$	80.03	58.45	67.56	51.01
R2-Net [28]	$arXiv_{2024}$	65.49	73.79	<u>69.40</u>	<u>53.14</u>
DENet [18]	$TITS_{2025}$	71.89	56.90	63.52	46.54
SegRoadv2 [19]	$IJDE_{2025}$	74.48	61.54	67.39	50.82
CGCNet [22]	$TGRS_{2025}$	77.19	60.84	68.04	51.57
DSFC-Net（我们的方法）		75.77	64.93	**69.93**	**53.77**

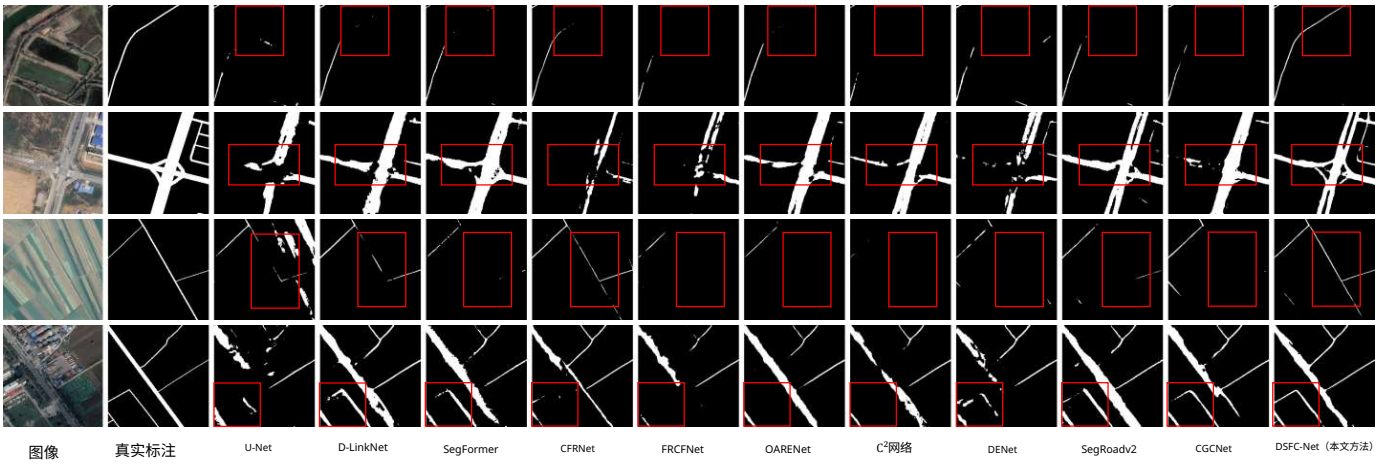

图 4. 不同方法在 WHU-RuR+ 数据集上的视觉比较。

道路场景。它们与标准城市数据集存在显著差异，因为其中农村背景占比很高，而道路像素密度非常低。具体而言，WHU-RuR+、DeepGlobe 和马萨诸塞州的农村区域覆盖率分别为 98.4%、74.31% 和 49.65%。与此同时，三个数据集中的道路像素比例均低于 5%。这种数据分布真实地反映了农村地区所面临的挑战：在广阔而复杂的背景中识别狭窄且稀疏道路十分困难。

1）WHU-RuR+：该数据集是一个专为乡村道路提取而打造的大规模数据集，包含 36,098 对图像（1024×1024），空间分辨率为 0.3–0.8 米。数据集覆盖八个国家的 6,866.35 平方千米^2，呈现了农村场景的独特复杂性，包括宽度各异的道路和未铺设路面的道路。我们遵循标准划分，其中 18,103 张图像用于训练，17,995 张用于测试。

2）DeepGlobe：该数据集源自 DeepGlobe 2018 挑战赛，提供覆盖泰国、印度尼西亚和印度的亚米级（0.5 m）卫星影像（1024×1024）。其中包含多种道路类型和光照条件覆盖总面积达 2,220 km^2。本研究使用官方训练集中的 4,980 幅图像进行训练。由于无法获取官方测试集的真实标签，我们采用包含 1,236 幅图像的完整官方验证集作为测试集进行评估。

3）马萨诸塞州：该数据集包含 1,171 张航空图像（1500×1500），空间分辨率为 1 米，覆盖了超过 2,600 平方千米^2 的多样化城乡地形。按照标准协议，数据集严格划分为 1,108 张训练图像和 63 张测试图像。

B. 实现细节

在 DSFC-Net 中，我们在 CNN分支采用 ConvNeXt-v 2 [29] 块，并在 Transformer 分支采用所提出的空间-频率混合 Transformer（SFT）层。实验在 4 块配备 24GB 显存的 NVIDIA RTX 3090 图形处理器上进行。对于 WHU-RuR+ 数据集，原始图像被调整为 768×768 像素。对于 DeepGlobe 数据集，原始图像被划分为四个 512×512 图块。对于马萨诸塞州数据集，原始图像被划分为九个

表 II 所提出的 DSFC-Net 与 DeepGlobe 数据集上最先进的道路提取方法的定量比较，其中粗体表示最佳结果，下划线表示次佳结果。

方法	来源	精确率（%）	召回率 (%)	F1分数（%）	交并比（%）
U-Net [31]	$MICCAI_{2015}$	84.59	73.81	78.83	65.06
D-LinkNet [11]	$CVPR_{2018}$	82.65	76.60	79.54	65.99
SegFormer [32]	$NIPS_{2021}$	82.10	76.91	79.42	65.86
CFRNet [23]	$GRSL_{2024}$	82.28	78.46	<u>80.32</u>	<u>67.12</u>
FRCFNet [24]	$GRSL_{2024}$	83.33	72.71	77.66	63.48
OARENet [26]	$TGRS_{2024}$	83.59	76.65	79.97	66.63
C^2Net [33]	$TGRS_{2024}$	82.94	73.44	77.90	63.80
DENet [18]	$TITS_{2025}$	73.92	84.25	78.75	64.95
SegRoadv2 [19]	$IJDE_{2025}$	83.55	76.55	79.90	66.52
CGCNet [22]	$TGRS_{2025}$	82.89	76.87	79.77	66.35
DSFC-Net（我们的方法）		82.68	80.23	**81.44**	**68.68**

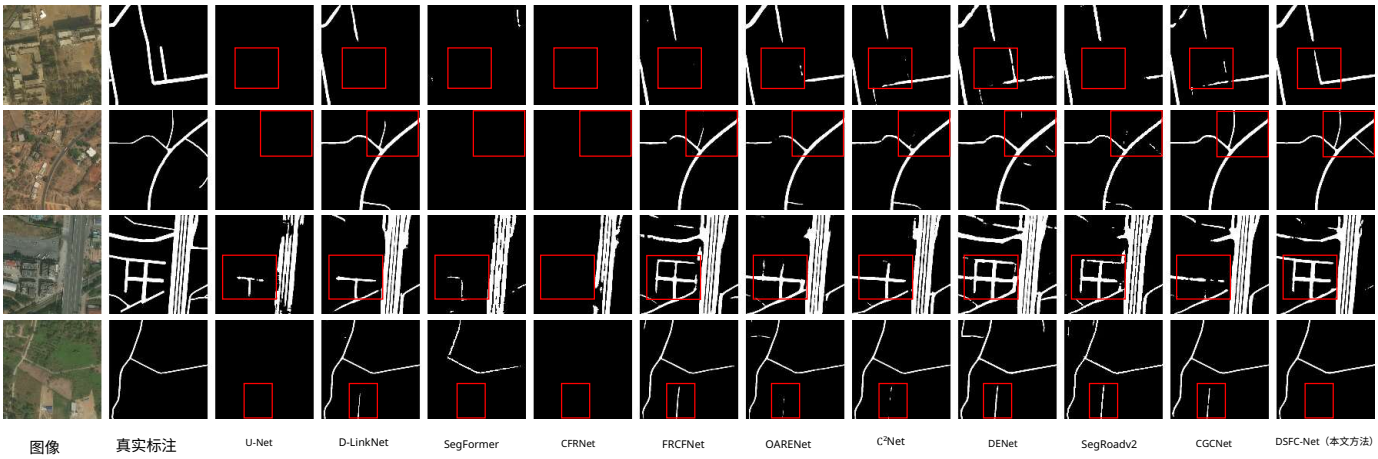

图 5. DeepGlobe 数据集上不同方法的可视化比较。

500×500 图块，并调整为 512×512 像素。在训练阶段，我们采用一些数据增强方法以避免过拟合，包括水平翻转、垂直翻转和对角翻转。所有模型均使用 AdamW 优化器进行训练，批大小为 4，共训练 100 个轮次，学习率初始设置为 1e-4。损失函数为二元交叉熵（BCE）损失与 Dice 损失之和。

C. 评估指标

所提出模型的性能采用精确率、召回率、F1 分数和交并比（Intersection over Union）进行评估。精确率是指正确预测的道路像素占全部预测像素的百分比，召回率是指正确预测的道路像素占全部实际道路像素的百分比。F1 分数是一种综合指标，用于平衡精确率与召回率。交并比衡量预测道路区域与真实标注道路区域之间的重叠程度。定义如下：

$$Precision = \frac{TP}{TP + FP}, \tag{6}$$

$$Recall = \frac{TP}{TP + FN}, \tag{7}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}, \tag{8}$$

$$IoU = \frac{TP}{TP + FP + FN}, \tag{9}$$

其中，TP、FP和 FN分别表示真正例、假正例和假负例。

D. 实验结果

为全面评估所提出的 DSFC-Net 在乡村道路分割任务中的性能，我们在三个数据集上进行了一系列实验，并将其与近年来几种具有代表性的方法进行比较，包括 U-Net [31],D-LinkNet [11],SegFormer [32], 以及 2024年–2025年新提出的方法，如 CFRNet[23], FRCFNet [24], OARENet [26],C^2CNet [33], DENet [18],SegRoadv2 [19], 和 CGCNet [22]。所有实验均采用相同的数据划分和训练策略，以确保公平比较。

表 III 马萨诸塞州数据集上本文提出的 DSFC-Net 与当前最先进道路提取方法的定量比较，其中粗体表示最佳结果，下划线表示次佳结果。

方法	来源	精确率（%）	召回率 (%)	F1分数 (%)	交并比 (%)
U-Net [31]	$MICCAI_{2015}$	85.66	72.39	78.47	64.56
D-LinkNet [11]	$CVPR_{2018}$	82.81	74.38	78.37	64.43
SegFormer [32]	$NIPS_{2021}$	83.51	75.02	79.04	65.34
CFRNet [23]	$GRSL_{2024}$	85.16	75.72	80.16	66.89
FRCFNet [24]	$GRSL_{2024}$	85.23	75.30	79.96	66.61
OARENet [26]	$TGRS_{2024}$	84.92	75.08	79.70	66.25
C^2Net [33]	$TGRS_{2024}$	85.55	74.22	79.48	65.95
DENet [18]	$TITS_{2025}$	80.30	80.56	<u>80.43</u>	<u>67.26</u>
SegRoadv2 [19]	$IJDE_{2025}$	85.41	74.27	79.45	65.91
CGCNet [22]	$TGRS_{2025}$	85.35	74.64	79.63	66.16
DSFC-Net（本文方法）		85.18	76.49	**80.60**	**67.50**

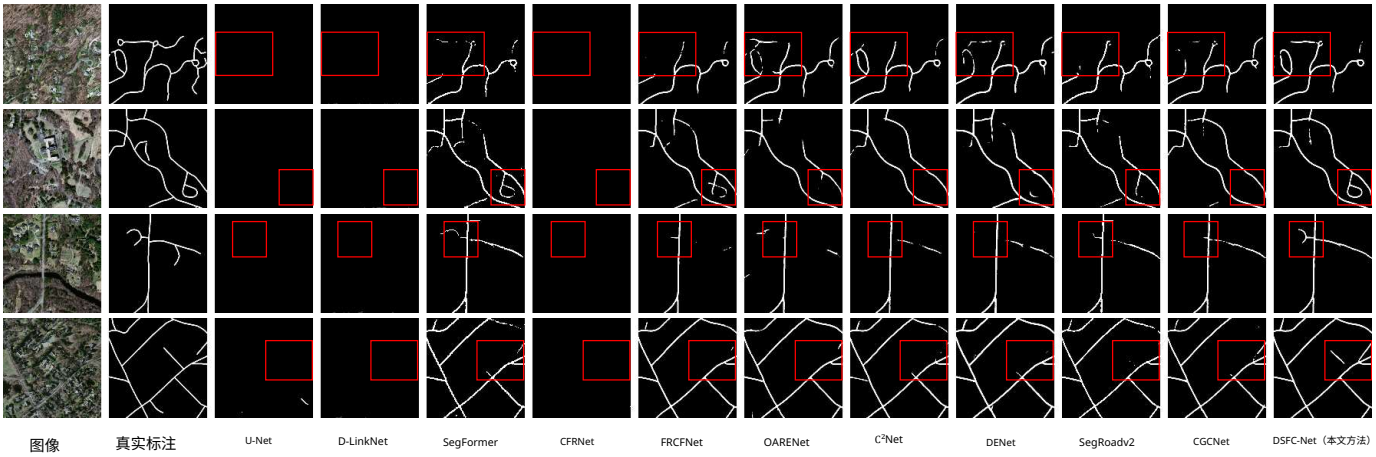

图 6. 马萨诸塞州数据集上不同方法的视觉对比。

1）*WHU-RuR* 的结果+： WHU-RuR+数据集包含大量复杂的乡村道路场景，背景干扰强烈且道路特征模糊，极具挑战性。定量结果见表 I。DSFC-Net 以 69.93% 的 F1 分数和 53.77% 的交并比取得最先进性能，优于所有对比方法。

在交并比指标上，我们的方法比次优的 D-LinkNet 高出 1.09%。在精确率和召回率指标上，近期方法（如 C^2Net [33] 和 CFRNet）取得了出色的精确率。然而，这是以较低的召回率为代价的，表明这些模型无法检测出假阴性目标，例如细微或被遮挡的乡村道路。相反，D-LinkNet 的召回率最高，但精确率较低，这说明它容易将背景纹理误分类为道路。

DSFC-Net 有效地应对了精确率与召回率之间的权衡。与高召回率的 D-LinkNet 相比，我们的精确率显著提高了 3.31%，展现了模型抑制噪声的能力。同时，我们的召回率较高精确率的 C^2Net 大幅提升了 6.48%，证明了我们恢复断裂或微弱道路特征的能力。这种鲁棒性证实，所提出的频率感知注意力和多尺度融合机制在复杂乡村环境中提取完整道路拓扑结构方面尤其有效。

为进一步验证 DSFC-Net 的有效性，我们在图+4 中展示了其在 WHU-RuR数据集上的可视化对比。可视化结果表明，与当前最先进的基线方法相比，我们的方法始终能够生成更完整、更准确的道路分割掩码。如第二行所示，该行包含复杂的交叉口和分支结构，DSFC-Net 保持了拓扑完整性，并维持了道路路段之间平滑的连通性。相比之下，SegFormer 和 CFRNet 等模型在交叉口区域出现了过度平滑或假阳性，这可能是由于它们依赖全局注意力，却缺乏显式的频率感知边界细化。

2）*DeepGlobe* 的结果：DeepGlobe 数据集的定量结果如表 II 所示。总体而言，所提出的 DSFC-Net 展现出最佳的综合性能。具体来说，DSFC-Net 的 F1分数达到 81.44%，交并比达到 68.68%，均显著优于其他对比方法。值得注意的是，我们的方法

在 F1 分数上比排名第二的 CFRNet 高出 1.12%，在交并比上高出 1.56%。尽管 DeepGlobe 数据集中农村场景的比例有所下降，我们提出的模型仍能在各项关键指标上保持稳健的性能，展现出其在城市和乡村场景中的强大泛化能力。

图 5 展示了不同方法在 DeepGlobe 数据集上的道路提取可视化结果。值得注意的是，在第四行的示例中，几乎所有方法都错误地检测出了图像下方的道路，但 DSFC Net 在复杂场景中仍保持鲁棒性。

3）马萨诸塞州结果：作为混合场景复杂度的数据集，马萨诸塞州数据集的特点是城市道路网络占比较高（约 50% 为非乡村背景）。尽管这些道路不像纯乡村道路数据集中的道路那样狭窄和稀疏，但它们带来了独特的挑战，尤其是树木和建筑物造成的严重遮挡，以及相似人工纹理的干扰。因此，在该数据集上评估我们的模型，对于验证其在乡村特定领域之外的泛化能力至关重要。

如表 III所示，DSFC-Net 在所有评估方法中取得了最高的整体性能，F1 分数为 80.60%，交并比为 67.50%。相比之下，DENet 采用激进的预测策略，将边缘图像作为分割的辅助特征。虽然该策略提高了召回率，但由于复杂的城市背景，也引入了大量误检噪声。关键在于，尽管 F1 分数的数值领先幅度有限，DSFC-Net 仍以 85.18% 的高精确率展现出更优的可靠性，能够有效区分道路与城市噪声，并确保较低的虚警率。

图 6 展示了体现这种鲁棒性的可视化结果。第一行中，场景包含蜿蜒的道路，道路穿行于茂密的植被和阴影区域之中。由于上下文感知能力有限，或受到建筑物阴影的干扰，大多数基线模型都无法检测到狭窄的道路路段。DSFC-Net 利用空间-频率混合 Transformer（SFT）的长程依赖建模能力，成功保留了道路的连续性和拓扑结构。相比之下，SegFormer 等基于 Transformer 的方法往往会产生断裂或过度平滑的输出，尤其是在急弯和交叉路口处，这表明它们在这些复杂环境中的边界细化能力不足。

E. 消融研究与讨论

为探究所提出的 DSFC-Net 的有效性，我们在具有挑战性的 WHU-RuR+数据集上开展了一系列消融研究。我们分析了双分支架构、SFT 模块内部组件的贡献，以及不同 CNN 骨干网络的影响。

1）模型复杂度与计算成本比较：为评估性能与效率之间的权衡，我们比较了模型参数（参数量）和浮点运算次数（FLOPs）。尤为重要的是，FLOPs 的计算采用了与我们的准确率基准测试完全相同的 768×768输入分辨率，以确保严格一致。如表IV所示，DSFC-Net 在实现更高准确率（平均 F1：77.32%，平均 IoU：63.32%）的同时，相较于最先进的方法仍保持了有竞争力的计算效率。

与 U-Net、DENet、C^{2}Net 和 SegRoadv2 等计算开销较大的方法相比，DSFC-Net 的效率更高。例如，与排名第二的 SegRoadv2 相比，其平均 F1 高出 1.74%，同时浮点运算次数减少了 56.5%。DSFC-Net 的高效性主要归功于其双编码器设计和自适应融合机制，两者协同作用，减少了冗余计算。

相比之下，尽管 FRCFNet 和 CGCNet 等轻量级架构在减少参数量和浮点运算次数方面表现出色，但我们的实证结果表明，适当增加计算资源能够显著提升性能。例如，与高效的 CGCNet 相比，DSFC-Net 利用额外的模型容量，使平均 F1 提高了 1.51%，平均 IoU 提高了 1.96%。这表明，额外的参数量和计算成本得到了有效利用，有助于解决复杂的分割难题，例如狭窄且被遮挡的乡村道路。

最后，与基于 Transformer 的基线模型 SegFormer 相比，DSFC-Net 在计算成本略低的同时实现了更高的精度。总体而言，DSFC-Net 在精度和效率之间实现了理想的平衡，适用于大规模乡村道路提取。

2）双分支架构与融合的影响：我们首先验证双分支设计和通道特征融合模块（CFFM）的必要性。如表V所示，我们构建了三个变体：移除 Transformer 分支（无 Transformer）、移除 CNN分支（无 CNN），以及用简单的求和拼接替代 CFFM（无 CFFM）。

移除 Transformer 分支会导致性能下降最为显著，F1分数降低 1.83%，交并比降低 2.14%。这证实了 Transformer 所建模的长距离语义依赖对于在复杂环境中识别连续道路结构至关重要。移除 CNN分支会使 F1分数和交并比略有下降，这表明尽管 Transformer 功能强大，但 CNN 提取的局部纹理特征对于细化边界仍然不可或缺。与完整模型相比，以其他模块替换 CFFM 会导致结果欠佳，这验证了我们的跨分支联合重校准策略优于简单的特征聚合。

3）*SFT* 组件分析：为评估空间-频率混合 Transformer（SFT）的内部机制，我们对其三个核心组件进行消融：空间上下文聚合器（SCA）、跨频率交互注意力（CFIA）和多尺度前馈网络（MFFN）。对于 SCA 和 CFIA，我们直接将其移除；对于 MFFN，则将其替换为简单的前馈网络（FFN）。结果汇总于表VI。

单独部署时，SCA模块的表现优于CFIA模块，这表明空间上下文是道路提取的基础。结合CFIA和SCA可显著提升性能，证明频域和空间域具有互补性。CFIA能够有效捕获SCA可能遗漏的高频边缘细节。无论采用哪种注意力机制，加入MFFN都能进一步提升性能。完整配置达到

表 IV 三个数据集上的平均性能、模型复杂度与计算成本比较，采用平均 F1 分数（平均 F1）、平均交并比（平均 IoU）、参数量（参数量）和浮点运算次数。

方法	平均 F1（%）	平均 IoU（%）	浮点运算次数（G）	参数量（M）
U-Net [31]	74.57	59.78	456.87	28.99
D-LinkNet [11]	75.64	61.03	80.35	44.81
SegFormer [32]	75.27	60.66	127.61	27.34
CFRNet [23]	75.60	61.21	85.23	33.64
FRCFNet [24]	72.87	57.92	85.14	**7.89**
OARENet [26]	75.54	61.07	78.85	33.48
C^2Net [33]	74.98	60.25	278.44	59.61
DENet [18]	74.23	59.58	357.92	25.41
SegRoadv2 [19]	75.58	61.08	282.01	33.55
CGCNet [22]	75.81	61.36	**62.63**	8.45
DSFC-Net（本文方法）	**77.32**	**63.32**	122.57	58.06

表 V WHU-RUR+ 数据集上双分支架构与融合机制的消融研究。

方法	F1分数（%）	交并比（%）
不使用 Transformer	68.10	51.63
不使用卷积神经网络	69.56	53.33
不使用 CFFM	69.63	53.41
DSFC-Net（本文方法）	**69.93**	**53.77**

最高F1分数为69.93%，证实在前馈阶段增强局部特征变换有助于提升性能。

表 VI 空间-频率混合 Transformer（SFT）各组件在 WHU-RUR+ 数据集上的消融研究。

模块			F1分数（%）	交并比（%）
CFIA	SCA	MFFN		
✓	✓	-	69.23	52.94
-	✓	✓	68.95	52.61
✓	-	✓	68.26	51.81
✓	-	-	68.05	51.57
-	✓	-	68.76	52.39
✓	✓	✓	**69.93**	**53.77**

*4) CNN*骨干网络的选择：我们研究了不同 CNN 骨干网络对整体性能的影响，如表 VII所示。我们选取了四种常用的 CNN 骨干网络：ResNet [36],MobileNet-v2 [37], ConvNeXt [38]以及 ConvNeXt-v2 [29]。

实验结果表明，由于特征表征能力有限，MobileNetV2 的表现最差。ResNet 提供了稳定的基线性能。ConvNeXt-v2 采用先进的全局响应归一化设计，并具备更强的特征提取能力，性能超过了原始 ConvNeXt，取得了最佳的分割精度。因此，我们选择 ConvNeXt-v2 作为 CNN分支的骨干网络。

表VII WHU-RUR+ 数据集上CNN骨干网络选择的消融研究。

CNN骨干网络	F1分数 (%)	交并比(%)
ResNet	69.08	52.76
MobileNet-v2	68.46	52.04
ConvNeXt	69.65	53.43
ConvNeXt-v2（我们的方法）	**69.93**	**53.77**

5）热力图分析：为了更全面地验证模型的有效性，我们将模型的热力图可视化，以分析关键组件的影响。热力图展示了编码器、解码器和各个 CFFM 层的输出。我们采用双线性插值将热力图转换为相同尺寸。如图7所示，这些热力图揭示了不同组件如何在多个阶段逐步优化道路相关特征。在早期阶段，ConvNext-v2 和 SFT 对田野、建筑物等大尺度结构都表现出强烈响应，但两者存在显著差异：ConvNext-v2 侧重局部纹理对比，而 SFT 分支则已开始通过频率注意力突出狭窄的线性模式。这表明，SFT 中的 CFIA模块即使在浅层也能增强对细粒度结构的敏感性。在更深层中，CNN分支逐渐捕获更多关于道路连通性和交叉路口的上下文信息，而 SFT 模块则持续抑制背景噪声，并通过关注与线性目标相关的主要频谱频率来锐化道路边界。CFFM 的输出凸显了这两个分支的互补性。经过自适应融合后，特征图在高分辨率细节与低分辨率上下文之间呈现出更强的一致性，道路路段上的激活更加清晰，非道路区域的误激活也有所减少。在解码器中，融合后的特征得到逐步优化，从而产生能够准确

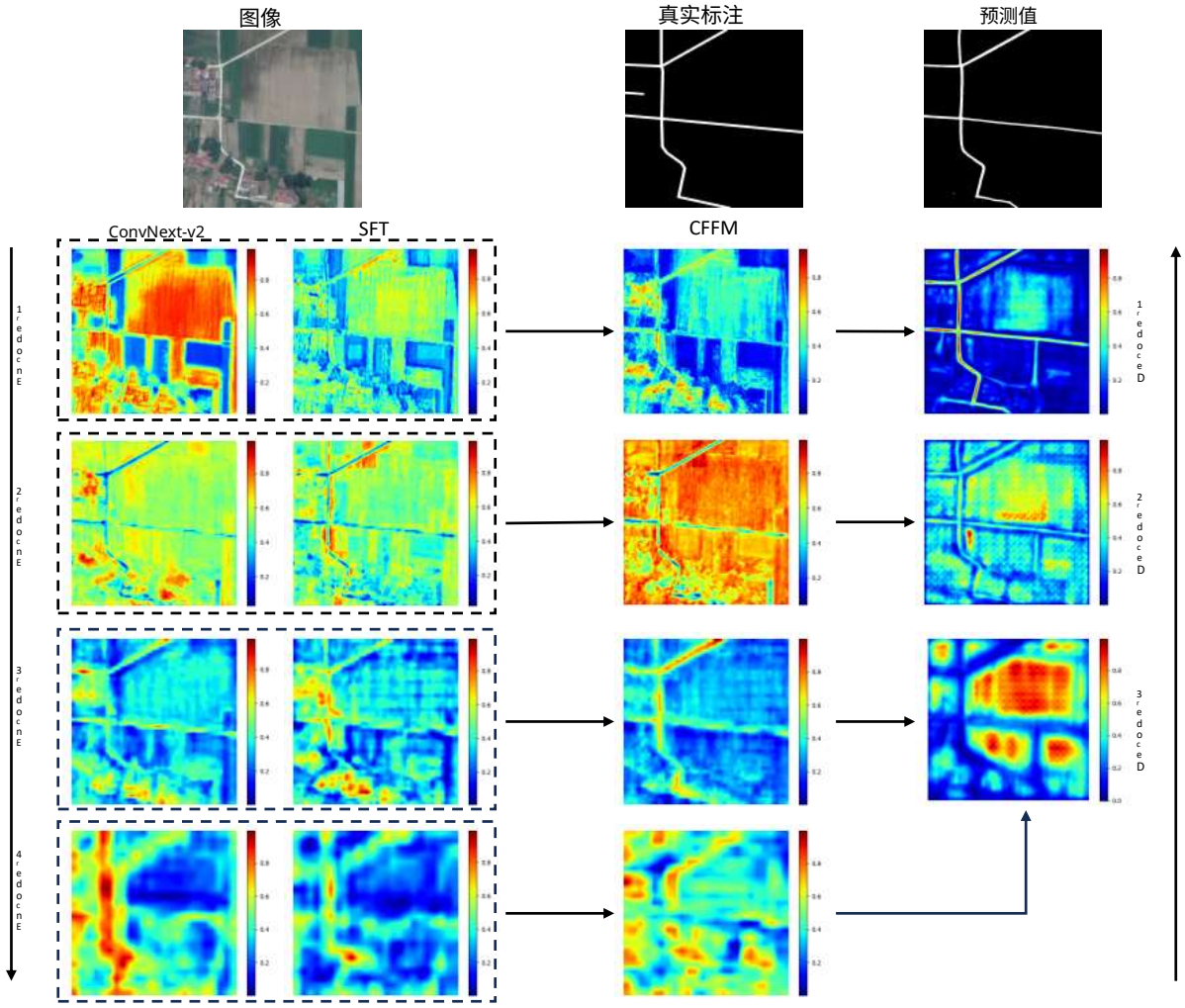

图7. DSFC-Net 的热力图，包括双编码器、CFFM 和解码器的输出。

保留乡村道路连续性和拓扑结构的最终预测。值得注意的是，与早期阶段相比，解码器的热力图在道路边缘表现出更强的激活，说明判别信号通过融合机制从双分支编码器中得到了有效传播。

这些可视化结果证实，通过 SFT 进行频域推理的集成显著提升了模型区分狭窄道路与复杂背景的能力。卷积分支能够有效捕捉乡村道路的局部纹理细节。此外，CFFM 成功平衡了多尺度信息，确保模型在整个过程中保留局部纹理线索和全局结构上下文。这些关键组件协同工作，使模型能够更好地还原道路结构。

五、结论

从遥感影像中提取乡村道路主要面临三个独特挑战，即：1）类内变异性高且类间可分性低，2）遮挡导致的道路不连续性，以及3）狭窄道路形状。为应对这些独特挑战，我们提出了DSFC-Net，一种新颖的双编码器架构协同融合空间与频域信息。一项关键创新是空间-频率混合 Transformer（SFT）层，专门用于建模蜿蜒农村道路的全局上下文依赖关系。SFT配备空间上下文聚合器（SCA）和跨频率交互注意力（CFIA），通过SCA聚合局部空间上下文，同时CFIA借助拉普拉斯金字塔分解分析跨频率特征，从而准确感知狭窄道路并保持其空间连续性。此外，我们提出的通道特征融合模块（CFFM）提供了一种自适应机制，用于融合来自卷积神经网络与Transformer分支的特征，确保表征稳健而全面。在三个公开且具有挑战性的数据集（WHU-RuR+、DeepGlobe和马萨诸塞）上的全面实验验证了DSFC-Net的优越性。它取得了最先进性能，并展现出强大的跨数据集泛化能力，能够有效提取完整且准确的乡村路网。这项工作为农村基础设施监测提供了一种有效的新工具，这对于推进关键可持续发展目标至关重要。尽管DSFC-Net展现出良好的性能，但仍有一些方面

有待进一步优化。当前的双编码器设计虽然能有效捕获多域特征，但与轻量级基于卷积神经网络的模型相比，计算需求更高。这可能给资源受限的边缘设备或大规模运行系统中的部署带来实际考量。未来研究将探索动态频率选择或稀疏注意力机制，以平衡计算效率与分割精度。

参考文献

[1] D. J. Weiss, A. Nelson, H. Gibson, W. Temperley, S. Peedell, A. Lieber, M. Hancher, E. Poyart, S. Belchior, N. Fullman 等，“2015年城市出行时间全球地图：评估可达性不平等”，《自然》，第553卷，第7688期，第333–336页，2018年。[2] J. Rockstr¨om, O. Edenhofer, J. Gaertner, 和 F. DeClerck，“为全球粮食系统保驾护航”，《自然·食品》，第1卷，第5期，2020年，第1、3页。[3] J. Verschuur, Y. Vittis, M. Obersteiner, 和 J. W. Hall，“贸易谷物和油料作物到岸成本的差异导致粮食获取机会不平等”，《自然·食品》，第6卷，第1期，第36–46页，2025年。[4] M. Burke, A. Driscoll, D. B. Lobell, 和 S. Ermon，“利用卫星影像理解并促进可持续发展”，《科学》，第371卷，第6535期，文章编号eabe8628，2021年。[5] E. E. Koks, J. Rozenberg, C. Zorn, M. Tariverdi, M. Vousdoukas, S. A. Fraser, J. Hall, 和 S. Hallegatte，“道路与铁路基础设施资产的全球多灾害风险分析”，《自然·通讯》，第10卷，第1期，文章编号2677，2019年。[6] P. Soille 和 M. Pesaresi，“应用于地球科学与遥感的数学形态学进展”，*IEEE*地球科学与遥感汇刊，第40卷，第9期，第2042–2055页，2002年。[7] S. Valero, J. Chanussot, J. A. Benediktsson, H. Talbot, 和 B. Waske，“用于检测超高分辨率遥感图像中道路网络的先进方向数学形态学”，《模式识别快报》，第31卷，第10期，第1120–1127页，2010年。[8] S. Hinz 和 A. Baumgartner，“从多视角航空影像中自动提取城市道路网络”，《国际摄影测量与遥感学会摄影测量与遥感杂志》，第58卷，第1-2期，第83–98页，2003年。[9] J. Liu, Q. Qin, J. Li, 和 Y. Li，“基于几何特征推断从高分辨率遥感图像中提取乡村道路”，《国际摄影测量与遥感学会国际地理信息杂志》，第6卷，文章编号314，2017年，第10期。[10] J. B. Mena，“自动道路提取技术现状：一种用于地理信息系统更新的新分类”，《模式识别快报》，第24卷，第16期，第3037–3058页，2003年。[11] L. Zhou, C. Zhang, 和 M. Wu，“D-LinkNet：采用预训练编码器和空洞卷积的LinkNet，用于高分辨率卫星影像道路提取”，载于*IEEE*计算机视觉与模式识别会议研讨会论文集，2018年，第182–186页。[12] Z. Zhang, Q. Liu, 和 Y. Wang，“基于深度残差U-Net的道路提取”，*IEEE*地球科学与遥感快报，第15卷，第5期，第749–753页，2018年。[13] R. Chen, X. Li, Y. Hu, C. Wen, 和 L. Peng，“野地—城市交界区域的遥感图像道路提取”，*IEEE*地球科学与遥感快报，第19卷，第1–5页，2022年。[14] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, 和 B. Guo，“Swin Transformer：采用移位窗口的分层视觉变换器”，载于*IEEE/CVF*国际计算机视觉大会论文集，2021年，第10 012–10 022页。[15] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, 和 M. Wang，“Swin-Unet：类U-Net的纯Transformer医学图像分割模型”，载于欧洲计算机视觉会议。Springer，2022年，第205–218页。[16] L. Gao, Y. Zhou, J. Tian, 和 W. Cai，“DdctNet：一种用于从高分辨率遥感图像中提取道路的可变形动态交叉变换器网络”，*IEEE*地球科学与遥感汇刊，第62卷，第1–19页，2024年。[17] S. He, F. Bastani, S. Jagwani, M. Alizadeh, H. Balakrishnan, S. Chawla, M. M. Elshrif, S. Madden, 和 M. A. Sadeghi，“Sat2Graph：通过图张量编码提取道路图”，载于欧洲计算机视觉会议。Springer，2020年，第51–67页。[18] T. Guo, Y. Gao, F. Luo, L. Zhang, B. Du, 和 X. Gao，“DENet：用于从高分辨率遥感影像中提取道路的方向与边缘协同感知网络”，《*IEEE*智能交通系统汇刊》，2025年。

[19] Z. Yu、Z. Chen、K. Xiao、X. Lei、R. Tang、Q. He、Z. Sun 和 H. Guo，《Segroadv2：一种结合可变形自注意力与卷积、用于具有连通性结构的道路提取的混合网络》，《国际数字地球学报》，第18卷，第1期，第2480267页，2025年。[20] N. Wang、X. Wang、Y. Pan、W. Yao 和 Y. Zhong，《Whu-rur+：全球高分辨率乡村道路提取基准数据集》，《国际地球观测与地理信息应用学报》，第139卷，第104518页，2025年。[21] X. Lu、Y. Zhong、Z. Zheng 和 L. Zhang，《Gamsnet：具有多尺度残差学习的全局感知道路检测网络》，《ISPRS摄影测量与遥感学报》，第175卷，352，2021 340pp. – 。[22] P. Liu、X. Gao、C. Shi、Y. Lu、L. Bai、Y. Fan、Y. Xing 和 Y. Qian，《Cgcnet：利用紧凑全局上下文感知从遥感图像中提取道路》，《IEEE地球科学与遥感汇刊》，2025年。[23] Y. Xiong、L. Li、D. Yuan、H. Wang、T. Ma、Z. Wang 和 Y. Yang，《Cfrnet：基于级联融合网络的遥感图像道路提取》，《IEEE地球科学与遥感快报》，第21卷，第1–5页，2024年。[24] H. Wang、L. Bai、D. Xue、M. C. Momi、Z. Ye 和 S. Quan，《Frcfnet：用于道路提取的特征重组与上下文信息融合网络》，《IEEE地球科学与遥感快报》，第21卷，第1–5页，2024年。[25] M. Zhou、H. Sui、S. Chen、J. Wang 和 X. Chen，《Bt-roadnet：一种用于从高分辨率遥感影像中提取道路的边界与拓扑感知神经网络》，《ISPRS摄影测量与遥感学报》，第168卷，第288–306页，2020年。[26] R. Yang、Y. Zhong、Y. Liu、X. Lu 和 L. Zhang，《用于高分辨率遥感影像的遮挡感知道路提取网络》，《IEEE地球科学与遥感汇刊》，第62卷，16，2024 1pp. – 。[27] N. Wang、X. Wang、Y. Pan、L. Lei、W. Yao 和 Y. Zhong，《在遥感影像中识别乡村道路：从基准数据集到由粗到精的提取网络——以中国为例》，《ISPRS摄影测量与遥感学报》，第231卷，第487–506页，2026年。[28] N. Wang、X. Wang、Y. Pan、W. Yao 和 Y. Zhong，《用于高分辨率卫星影像中狭窄乡村道路检测的反向细化网络》，arXiv预印本 arXiv:2410.10389，2024年。[29] S. Woo、S. Debnath、R. Hu、X. Chen、Z. Liu、I. S. Kweon 和 S. Xie，《ConvNeXt V2：通过掩码自编码器协同设计并扩展卷积网络》，载于《IEEE/CVF计算机视觉与模式识别会议论文集》，2023年，第16 133–16 142页。[30] J. Hu、L. Shen 和 G. Sun，《挤压激励网络》，载于《IEEE计算机视觉与模式识别会议论文集》，2018年，第7132–7141页。[31] O. Ronneberger、P. Fischer 和 T. Brox，《U-Net：用于生物医学图像分割的卷积网络》，载于《医学图像计算与计算机辅助干预国际会议》。Springer，2015年，第234–241页。[32] E. Xie、W. Wang、Z. Yu、A. Anandkumar、J. M. Alvarez 和 P. Luo，《SegFormer：利用Transformer 实现语义分割的简单高效设计》，《神经信息处理系统进展》，第34卷，第12 077–12 090页，2021年。[33] Z. Yang、W. Zhang、Q. Li、W. Ni、J. Wu 和 Q. Wang，《C2net：通过上下文感知与跨空间尺度特征交互进行道路提取》，《IEEE地球科学与遥感汇刊》，第62卷，第1–11页，2024年。[34] I. Demir、K. Koperski、D. Lindenbaum、G. Pang、J. Huang、S. Basu、F. Hughes、D. Tuia 和 R. Raskar，《DeepGlobe 2018：一项通过卫星图像解析地球的挑战》，载于《IEEE计算机视觉与模式识别会议研讨会论文集》，2018年，第172–181页。[35] V. Mnih，《航空图像标注的机器学习》。多伦多大学（加拿大），2013年。[36] K. He、X. Zhang、S. Ren 和 J. Sun，《用于图像识别的深度残差学习》，载于《IEEE计算机视觉与模式识别会议论文集》，2016年，第770–778页。[37] M. Sandler、A. Howard、M. Zhu、A. Zhmoginov 和 L.-C. Chen，《MobileNetV2：倒置残差与线性瓶颈》，载于《IEEE计算机视觉与模式识别会议论文集》，2018年，第4510–4520页。[38] Z. Liu、H. Mao、C.-Y. Wu、C. Feichtenhofer、T. Darrell 和 S. Xie，《面向2020年代的卷积网络》，载于《IEEE/CVF计算机视觉与模式识别会议论文集》，2022年，第11 976–11 986页。