

融合多特征与超分辨的农田防护林带遥感语义分割模型

张琦，周云成^{*}，赵鸿舫，武文豪，黄跃坤

(沈阳农业大学信息与电气工程学院，沈阳 110866)

摘要：农田防护林带在遥感影像中呈现为典型的狭长线状地物，其精确分割因全局上下文依赖强、局部特征微弱而面临严峻挑战。为实现耕地与防护林带的高精度自动化提取，该研究提出一种名为 MFF-Net 的语义分割模型。首先设计了多特征融合模块，该模块通过空间门控融合机制，自适应整合类 Mamba 算子的长程依赖建模能力、卷积的局部细节感知能力及傅里叶变换的频域边缘增强能力。同时，引入一种任务导向的超分辨率预处理流程，以提升输入影像质量，辅助模型识别细微地物。在自建农田防护林数据集上的试验结果表明，MFF-Net 的耕地与防护林带分割精确率分别为 96.42% 与 82.83%，平均交并比达 83.45%，性能超越多种先进模型。消融试验证实，多特征融合策略贡献显著，其中频域特征使防护林带分割交并比提升 2.80%。超分预处理作为有效辅助，使防护林带分割交并比进一步大幅提升 6.61%。MFF-Net 语义分割模型通过融合全局、局部与频域多元特征，能有效应对狭长地物分割挑战，结合超分预处理可进一步挖掘模型潜力。该研究为农田林网的精准监测与动态评估提供了一种可选的技术路径。

关键词：遥感；语义分割；农田防护林带；多特征融合；超分辨率；深度学习

doi: 10.11975/j.issn.1002-6819.202509002

中图分类号: TP751

文献标志码: A

文章编号: 1002-6819(2026)-12-0268-11

张琦, 周云成, 赵鸿舫, 等. 融合多特征与超分辨的农田防护林带遥感语义分割模型[J]. 农业工程学报, 2026, 42(12): 268-278. doi: 10.11975/j.issn.1002-6819.202509002 <http://www.tcsae.org>

ZHANG Qi, ZHOU Yuncheng, ZHAO Hongge, et al. Semantic segmentation of farmland shelterbelts using multi-feature and super-resolution remote sensing[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2026, 42(12): 268-278. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202509002 <http://www.tcsae.org>

0 引言

农田防护林带是农田生态系统的重要组成部分^[1], 在防风固沙、水土保持、调节小气候等方面发挥着关键作用, 其构成的农田林网化程度已被纳入国家标准《耕地质量等级》(GB/T 33469—2016) 的核心评价指标之一^[2]。目前, 农田林网化程度调查主要依赖人工野外踏勘与遥感目视解译, 存在效率低、主观性强等问题。因此, 发展基于遥感影像的自动化、高精度防护林带提取方法, 对于提升耕地质量监测效率与科学性具有重要意义。

传统的遥感地物提取方法多依赖于人工设计特征(如光谱、纹理、形状等), 结合支持向量机^[3]、随机森林^[4]、决策树^[5]等机器学习分类器实现。例如, 高峰等^[6]基于 GF-2 影像提取多类特征, 构建 SVM 模型提取荒漠绿洲区防护林带; 李萌等^[7]融合多源遥感数据, 利用随机森林实现了林地类型的有效识别, 分类精度达 92.85%。然而, 这类方法在特征构造与选择上依赖专家经验, 且受限于中低分辨率影像的空间细节表达能力, 难以有效捕捉农田防护林带这类狭长、稀疏分布的线状地物。

近年来, 深度学习方法为遥感语义分割提供了新的路径。以 U-Net^[8] 为代表的编码器-解码器架构被广泛应用。陈妮等^[9] 基于国产高分系列与资源三号卫星数据, 应用 U-Net 实现了土地利用类型的提取, 准确率达 90.95%。XU 等^[10] 通过改进 U-Net 的跨越连接结构与损失函数, 增强了模型对耕地边缘特征的提取能力, 在 Landsat-5 影像上取得了 91.74% 的识别精度。朱昱等^[11] 在 U-Net 基础上引入 D-LinkNet 架构, 通过融合深层语义特征与边缘概率图, 显著提升了 GF-2 影像中耕地边界特征的完整性。为进一步增强多尺度特征提取能力, RU 等^[12] 将空间金字塔池化模块嵌入 U-Net 中, 构建了多尺度上下文特征融合网络。刘振国等^[13] 则以轻量化的 MobileNetV2 替换 DeepLabV3+^[14] 的 Xception 主干网络, 在保持特征提取效率的同时实现了荔枝种植区域的精准识别。然而, 当面对农田防护林带这一特殊地物时, 现有深度学习模型仍面临一个核心问题, 即模型特征表征能力与地物复杂形态需求之间的不匹配。首先是全局与局部特征的失衡, 基于 CNN 的模型(如 DeepLabV3+^[14]) 虽善于捕捉局部纹理, 但其固定感受野难以建模防护林带所需的长距离空间依赖性, 导致分割结果出现断裂; 而引入自注意力机制的 Transformer 类模型(如 Swin Transformer^[15]) 虽能捕获全局上下文, 却在局部细节与边缘锐度的保持上有所不足。其次是细微地物与有限分辨率之间的矛盾, 防护林带在影像中呈现为狭长条带状且像素占比极低, 受限于原始影像空间分辨率, 其纹理和边缘等高频信息微弱, 使得模型难以将其与光谱相似

收稿日期: 2025-09-01 修订日期: 2026-01-06

基金项目: 国家重点研发计划项目(2021YFD1500204; 2023YFD1501303)

作者简介: 张琦, 研究方向为深度学习与农业遥感。

Email: zq0106@stu.syau.edu.cn

※通信作者: 周云成, 博士, 教授, 研究方向为机器学习在农业信息处理中应用。Email: zhouyc2002@syau.edu.cn

的地物（如作物、草地）进行有效区分。

针对上述问题，本研究构建一种融合多特征与超分辨率的农田防护林带语义分割模型（multi-feature fusion network, MFF-Net）。本文的核心工作包括两个方面：其一，设计一个多特征融合模块，该模块通过空间门控融合机制，旨在整合类 Mamba 线性注意力的全局上下文建模、卷积网络的局部细节感知与傅里叶变换的频域边缘增强这三类能力，以探索其对改善狭长地物特征表征的作用。其二，在语义分割模型前端引入一个任务导向的超分辨率预处理流程，以期通过重建高分辨率影像来增强地物纹理与边界细节，从而在输入层面辅助后续的细微目标分割。在自建的农田防护林网数据集上开展试验，验证所提模型的有效性。

1 材料与方法

1.1 研究区与数据集构建

选取松嫩平原中南部部分区域（45°37′20.57″N~46°20′53.26″N, 125°26′46.64″E~126°07′29.10″E）作为研究区（图 1），面积约 4 100 km²。该区域位于黑龙江省农田防护林重点监测区内，中温带半湿润大陆性季风气候，无霜期 130~150 d，主要种植作物为玉米、水稻与大豆，一年一熟制，作物生长季内主导风向为偏南风，防护林带主要以杨、柳、松等树种为主。

研究选择吉林一号高分 03 星（JL1GF03）的遥感影像作为主要数据源，多光谱数据包含蓝（450~520 nm）、绿（520~600 nm）、红（630~690 nm）和近红外（760~900 nm）4 个波段，空间分辨率为 3 m，全色影像空间分辨率为 0.75 m。选用研究区 2020—2022 年 8—9 月期间，云量低于 3% 的 15 景遥感影像作为研究数据。数据经辐射定标、大气校正及几何精校正等预处理。选用多光谱数据的蓝（B）、绿（G）、红（R）波段与空间分辨率更高的全色影像，在 ENVI 软件中采用 Gram-Schmidt 算法进行数据融合以生成空间分辨率为 0.75m 的 RGB 影像。该处理旨在克服原始多光谱影像空间分辨率不足的局限，通过引入全色影像的精细细节，生成地物边界更清晰、纹理更丰富的高质量融合影像，为后续语义分割模型提供可靠的输入基础，这对于精确提取狭长的防护林带等

细微地物至关重要。最终影像存储为 GeoTIFF 格式以备后续使用。

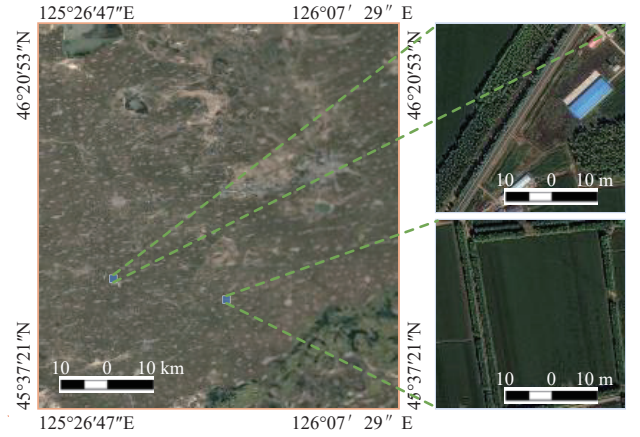


图 1 研究区域及影像数据示例

Fig.1 Study area and examples of images

为消除影像边缘的无数据区域（黑边），基于像素值阈值，识别并切除影像周边的无效黑边，提取出有效观测区域，进一步裁剪为大小 5 000×5 000 像素的系列图像，并剔除耕地与防护林带占比较少的图像，便于后续做语义标注，共得到 150 幅图像。针对语义分割模型的训练与评估需求，应用 LabelImg 标注工具对图像进行人工标注，标注出耕地田块和防护林带掩码，其他区域则被标注为背景，3 类像素占比分别为 65.52%、5.25% 和 29.93%。150 幅图像及对应的语义标注结果共同构成农田防护林网数据集。进一步按 4:1 的比例随机将数据集划分为训练集和测试集，用于模型的训练与评估。

1.2 总体技术框架

基于遥感影像实现农田防护林带的精确分割与空间分析，对耕地质量评价及生态效益评估具有重要意义。然而，防护林带在影像中通常呈现为狭长条带，其精确分割面临挑战，且基于像素级分割结果直接开展空间分布分析存在诸多不便。为此，本研究构建了一套集成遥感影像超分辨率预处理、语义分割与矢量化分析的技术框架（图 2），系统提升防护林带提取精度并拓展其应用。

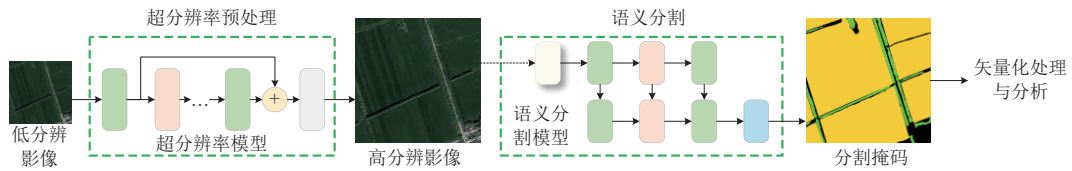


图 2 农田防护林带语义分割及矢量化应用技术框架

Fig.2 Technical framework for semantic segmentation and vectorization application of farmland shelterbelt

为提高防护林带的分割效果，首先引入超分辨率预处理作为前端增强手段。通过重建高分辨率影像，有效增加林带的像素宽度、锐化其边缘特征，从而为后续语义分割模型提供更优质的输入，缓解因目标狭窄、边界模糊导致的分割困难。语义分割是本文的核心环节，通过精心设计的语义分割模型生成像素级分类掩码 M 。

基于 M 可直接统计各地类像素数目，初步估算耕地与林带面积，为实现更复杂的空间分析，本研究进一步

对 M 进行矢量化处理，具体流程如图 3a。首先基于 M 生成耕地与防护林带的二值化图 B_c ，如式（1）所示：

$$B_c(i, j) = \begin{cases} 1 & \text{if } M_{i,j} = c \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

依次对 B_c 进行形态学闭运算与开运算，以填充孔洞、去除噪声、平滑边界；采用边界追踪算法^[16]提取区域轮廓像素坐标；利用 Douglas-Peucker 算法^[17]进行边界简化，

降低坐标冗余；基于遥感影像的地理坐标映射关系，通过地理空间数据抽象库（geospatial data abstraction library, GDAL）生成具有地理参考的面状矢量数据。为探讨矢量化结果工程化应用的可行性，以防护林带闭合度评估为例开展应用分析。首先获取单个田块多边形及其周边

缓冲区内的防护林带多边形；进而，将林带边界点与田块几何中心连线，计算其覆盖的最大几何角度，即保护角（图 3b）；最终，累计田块周围所有林带的保护角，其与 360° 的比值即为该田块的防护林带闭合度。用该指标评估农田林网化程度的达标情况。



图 3 分割掩码矢量化及应用示例

Fig.3 Vectorization and application examples of segmentation masks

1.3 基于多特征融合的语义分割模型设计

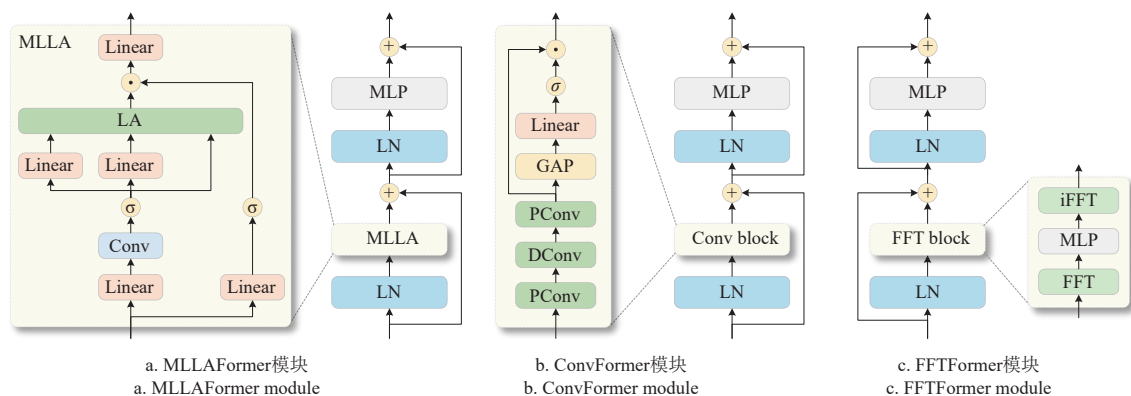
语义分割模型的分割性能直接影响矢量化结果及应用潜力。特征提取模块是语义分割模型的核心构造块，其特征提取能力决定模型性能。本研究根据田块和防护林带影像特点，优化设计特征提取模块，并基于该模块，应用编码器和解码器架构设计语义分割模型。

1.3.1 特征提取算子选择及优化设计

在遥感影像中，田块占图像大部分区域，属大尺度目标，防护林带则属于狭长目标，该特点要求特征提取模块能够在图像上下文中快速捕获全局信息，以建立同一目标各局部特征的空间相关性，使同一目标物被分割为连续区域，避免破碎化。作物、林木、草地等绿色植被反射光谱近似，区分难度大，需借助纹理特征的微弱差异，其要求特征提取模块能够有效捕获图像纹理的高频细节特征。因此，除基本的局部特征外，特征提取模

块也应具有全局上下文和频域特征的融合提取能力。

Transformer 的多头注意力（multi-head self-attention, MHSA）机制能够有效提取并融合全局上下文特征，但其计算复杂度为 $O(n^2)$ ， n 为特征数量，计算开销大。Mamba^[18] 是一种具有全局注意力的特征提取算子，该算子的核心是选择性状态空间模型（selective state space model, SSSM），其依赖并行扫描实现训练，但 GPU 对递归计算的优化不如矩阵乘法，会产生计算瓶颈。HAN 等^[19] 研究表明，SSSM 是线性注意力（linear attention, LA）^[20] 的一种特殊变种，该变种可表示为具有额外输入门和遗忘门的 LA，由此构建了类 Mamba 的线性注意力（Mamba-like linear attention, MLLA）算子。MLLA 在保持全局注意力的同时，将计算复杂度降为 $O(n)$ ，因此本研究用 MLLA 取代 Transformer 的 MHSA，构建 MLLAFormer（图 4a），用其提取并融合遥感影像的全局上下文特征。



注：Linear 为线性层； $\odot$ 和 $\oplus$ 分别代号为逐元素乘法和加法运算；LA 表示线性注意力； σ 表示 Sigmoid 激活；Conv 表示卷积； $\oplus$ 表示逐元素加法运算；MLP 表示多层感知机；LN 表示层归一化；PConv 表示点卷积；DConv 表示深度卷积；GAP 表示全局平均池化；FFT、iFFT 为快速傅里叶变换及其逆变换。下同。

Note: Linear refers to a linear layer; $\odot$ and $\oplus$ represent element-wise multiplication and addition, respectively; LA is linear attention; σ is Sigmoid activation; Conv means convolution; MLP is short for multi-layer perceptron; LN is layer normalization; PConv and DConv are pointwise and depthwise convolutional layers, respectively; GAP is global average pooling; FFT and iFFT are the fast Fourier transform and its inverse. The same below.

图 4 特征提取算子

Fig.4 Feature extraction operator

卷积模块在图像局部特征提取方面具有优势，其有效性已在多类计算机视觉任务中得到验证，为降低计算

复杂度，本研究用 PConv 和 DConv 组合作为局部特征提取算子。同时为增强局部特征表达能力，并降低潜在特

征冗余，在卷积算子后应用通道注意力算子 ϕ （式（2））。

$$\phi(X) = \sigma(W_l \omega(X)) \quad (2)$$

式中 X 为输入特征， ω 表示全局平均池化， W_l 为线性变换（Linear）的权重参数， σ 表示 Sigmoid 激活。应用算子 ϕ 的卷积块结构如图 4b。LIN 等^[21]研究表明 Transformer 的结构设计是其取得成功的关键因素之一，为利用该设计，并保持特征提取算子结构上的一致性，本研究用该卷积块取代 Transformer 的 MHSA，构建 ConvFormer（图 4b），用其提取图像局部特征。

Transformer 的多层感知机（multi-layer perceptron, MLP）在图像的空间域上更倾向于低频特征提取^[22]，对精细纹理和锐利边缘的捕捉能力有限，不利于保留分割结果中的精细化目标边缘。傅里叶特征变换能够捕获高频细节^[23]，因此本研究应用快速傅里叶变换（fast Fourier transform, FFT）将输入由空间域转换为频域，进一步由 MLP 直接在频域空间上提取图像特征，即在 FFT 基础上构建 FFTFormer（图 4c），用其提取影像的频域特征。

1.3.2 基于空间门控的多特征融合模块设计

MLLAFormer、ConvFormer 和 FFTFormer 分别在全局、局部和频域空间上提取遥感影像的多元特征。在语义分割网络的不同层次，3 种特征提取算子应具有不同的贡献度。本研究提出用空间门控融合机制（spatially gated fusion mechanism, SGFM）自适应地平衡不同算子的贡献度，该机制的计算过程如式（3）~（4）所示：

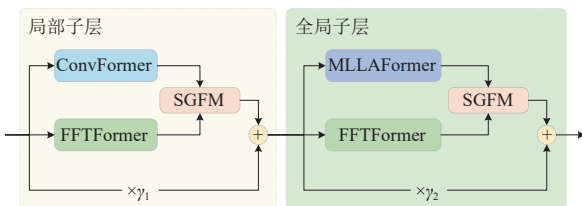
$$G = \sigma(\psi^{(1)}([X_1, X_2])) \quad (3)$$

$$Y = G \odot X_1 + (1 - G) \odot X_2 \quad (4)$$

式中 $X_1, X_2 \in \mathbb{R}^{c \times h \times w}$ 为不同特征提取算子提取的特征图， $[\cdot]$ 表示在通道维度上进行特征级联， $\psi^{(1)}$ 为输出通道数为 1 的点卷积， $G \in \mathbb{R}^{1 \times h \times w}$ 为空间维度上的门控平面， $Y \in \mathbb{R}^{c \times h \times w}$ 为 X_1, X_2 在空间维度上的融合结果， c, h, w 分别表示特征图 Y 的通道数、高度和宽度。

用 MLLAFormer、ConvFormer 和 FFTFormer 提取不同类型特征，用 SGFM 进行特征融合，构建多特征融合模块（multi-feature fusion block, MFFB）。MFFB 包含局部和全局两个子层（图 5），在局部子层中，输入特征图 $F_{in} \in \mathbb{R}^{c \times h \times w}$ 先经过 ConvFormer 和 FFTFormer 两个并行分支做局部和频域特征提取，获取 X_1, X_2 ，进一步用 SGFM 做特征融合，生成 Y 。为重用已有特征，为局部子层引入第 3 个并行分支，该分支通过可学习参数 $\gamma \in \mathbb{R}^c$ 对输入特征进行缩放，并与 Y 做进一步融合，如式（5）所示：

$$F_{out} = \gamma F_{in} + Y \quad (5)$$



注：SGFM 表示空间门控融合机制； $\times \gamma_1$ 等表示基于可学习参数 γ_1 的层缩放操作。下同。

Note: SGFM is short for spatially gated fusion mechanism; $\times \gamma_1$, etc., denote layer scaling operations based on the learnable parameter γ_1 . The same below.

图 5 多特征融合模块

Fig.5 Multi-feature fusion block (MFFB)

式中 $F_{out} \in \mathbb{R}^{c \times h \times w}$ 为局部子层输出。MFFB 的全局子层用 MLLAFormer 提取全局特征，其结构与局部子层类似。

1.3.3 编码器设计

以 MFFB 为基础，构建语义分割模型（图 6a），并命名为 MFF-Net。首先设计模型的编码器，该编码器包含 4 个阶段，各阶段先用由卷积层构成的模块（图 6b）进行基本特征提取，并将特征图通道数调整为 D_i ， $i \in \{1, 2, 3, 4\}$ 为阶段编号，同时用跨步卷积对输入特征图进行空间下采样，其后堆叠 L_i 个 MFFB 模块。 L_i 和 D_i 将作为编码器超参数，分别用于控制网络的深度（层数）和宽度（输出特征通道数）。编码器各阶段产生空间尺度不同的特征图 $F_i \in \mathbb{R}^{D_i \times H/2^{i+1} \times W/2^{i+1}}$ ， H, W 表示模型输入图像的高和宽。不同尺度的特征图 F_i 构成特征金字塔，用作解码器的输入。

1.3.4 解码器与多尺度环境特征提取模块设计

解码器首先对编码器输出的不同层特征图（ F_2, F_3 和 F_4 ）进行聚合。应用 3 个并行分支，将不同尺度的特征图上采样到相同的空间尺度。对于其中任一分支，首先用点卷积将特征图 D_i 的通道数由 D_i 统一调整为 C （用作解码器超参数），进而用双线性插值对特征图进行上采样，得到 $F_{upi} \in \mathbb{R}^{C \times H/8 \times W/8}$ 。对 3 个并行分支的上采样结果进行通道维度级联，并用点卷积对级联结果进行特征聚合和通道维度调整，如式（6）所示：

$$F = \psi^{(N)}([F_{up2}, F_{up3}, F_{up4}]) \quad (6)$$

式中 $F \in \mathbb{R}^{N \times H/8 \times W/8}$ 为聚合特征， N 表示 F 的通道数，也是解码器的另一超参数。

对输入影像的上下文环境进行多尺度建模，能够促进模型的跨尺度语义表征，对解决类内尺度变化、增强类别间区分度具有重要作用。进一步构建多尺度环境特征提取（multi-scale context extraction, MSCE）模块（图 6c），用于增强解码器的多尺度环境建模能力。MSCE 的输入为聚合特征 F ，该模块同样包含 3 个分支，各分支通过卷积和像素逆重排（pixel unshuffle, PU）操作实现多尺度上下文环境建模。其中分支 1 直接用 PU 将 F 在空间维上非重叠的 4×4 特征块重新排列在通道维上，产生该分支的输出 F_{s1} ，该操作在保持信息不丢失的前提下，将 4×4 区域的上下文环境信息直接融合为 1 个像素，该分支的等价感受野为 4×4 。分支 2 先采用步长为 2 的 3×3 卷积对 F 进行特征提取，进一步用 PU 将非重叠的 2×2 块排列在通道维度上，生成结果 F_{s2} ，该分支的等价感受野为 5×5 。分支 3 则用跨步为 4 的 5×5 卷积对 F 进行更大尺度的环境特征提取，生成 F_{s4} 。各分支的特征提取感受野存在差异，即分别提取了不同尺度的上下文环境信息。对 F_{s1}, F_{s2} 和 F_{s4} 进行融合，为降低计算复杂度同时平衡各尺度环境特征的贡献，融合前应用点卷积将三者的通道维度调整为 N 。MSCE 的 3 个分支分别提取了不同尺度的环境特征，但仍缺少全局性的环境特征。进一步应用 MLLAFormer，在级联后的多尺度环境特征上融合出全局特征，整个融合过程可表示为式（7）：

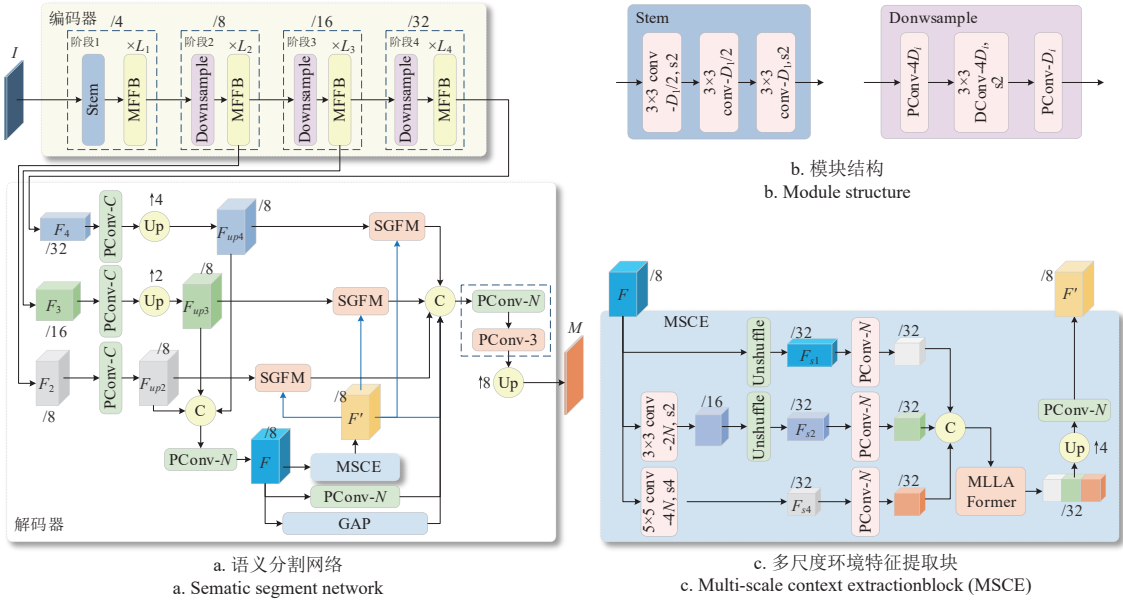
$$F_s = \text{MLLAFormer}([\psi^{(N)}(F_{s1}), \psi^{(N)}(F_{s2}), \psi^{(N)}(F_{s4})]) \quad (7)$$

式中 F_s 表示具有全局性的多尺度环境特征。进一步对 F_s

进行上采样, 使其与 F 的空间尺度相同, 并应用点卷积调整通道维度, 即有式 (8):

$$F' = \psi^{(N)}(U(F_s)) \quad (8)$$

式中 F' 为 MSCE 生成的多尺度环境特征, U 表示上采样。



注: I 、 M 分别表示输入图像与分割掩码; “/4” 等表示特征图空间维度为 I 的 $1/4$; $\times L_1$ 等表示模块重复 L_1 次; PCConv- N 等表示输出通道数为 N 的点卷积; F_4 、 F_{up4} 、 F 、 F' 、 F_{s1} 等表示不同尺度的特征图; Up 表示双线性插值上采样; $\uparrow 4$ 等表示空间上采样倍数为 4 倍; $3 \times 3 \text{ Conv-}D_1, s2$ 等表示卷积核大小为 3×3 , 通道数为 D_1 , 步长为 2 的卷积; Unshuffle 表示像素逆重排操作; C 表示在通道维度上的特征拼接操作。下同。
Note: I and M denote the input image and segmentation mask, respectively; “/4” indicates that the spatial resolution of the feature map is $1/4$ of I ; $\times L_1$ indicates that module is repeated L_1 times; PCConv- N denotes a pointwise convolution with N output channels; F_4 , F_{up4} , F , F' , F_{s1} represent feature maps at different scales; Up represents bilinear interpolation upsampling; $\uparrow 4$ indicates a spatial upsampling factor of 4; $3 \times 3 \text{ Conv-}D_1, s2$ denotes a convolution with a kernel size of 3×3 , D_1 output channels, and a stride of 2; Unshuffle denotes the pixel unshuffle operation; C denotes the feature concatenation operation along the channel dimension. Same below.

图 6 遥感影像语义分割模型

Fig.6 Remote sensing image semantic segmentation model

1.3.5 分割模型的损失函数定义

本研究构建的数据集中多种地类像素占比存在着明显的不平衡, 若直接采用常规的交叉熵损失进行优化, 分割模型会严重倾向于耕地主导类别, 而对占比极小的防护林带学习不足, 导致其分割性能下降。为解决该问题, 需设计对类别分布不敏感的损失函数。为此, 本研究采用了混合损失函数策略, 融合交叉熵损失与 Dice 损失的优点。交叉熵损失 L_{CE} 逐像素衡量预测与真实标签的概率分布差异, 有利于模型收敛。Dice 损失 L_{Dice} 则直接优化预测区域与真实区域之间的重叠度, 其对类别不平衡问题具有更强的鲁棒性, 能有效缓解训练过程中大类主导梯度的问题。最终, 语义分割模型的总损失函数 L 定义为二者的加权和, 如式 (9) ~ (11) 所示:

$$L_{CE} = -\frac{1}{HW} M_{gt} \log(M) \quad (9)$$

$$L_{Dice} = 1 - \frac{2M_{gt}M}{M_{gt}^2 + M^2} \quad (10)$$

$$L = \alpha L_{CE} + \beta L_{Dice} \quad (11)$$

式中 M_{gt} 为与输入图像对应的真实语义标注, α 、 β 为权重平衡系数。以 $\text{argmin } L$ 为目标, 对语义分割模型进行端到端优化, 直至损失函数收敛至稳定值。

1.4 面向分割任务的超分辨率预处理模型

为提升语义分割模型对防护林带等细微地物的识别能力, 本研究引入一个任务导向的超分辨率预处理流程。该流程旨在将输入的低分辨率遥感影像重建为高分辨率

影像, 以增强地物纹理、锐化目标边缘, 从而为后续分割模型提供更精细的输入。基于所设计的多特征融合模块, 构建超分辨率模型 (图 7)。该模型采用了当前研究^[24]中的通用架构: 首先通过卷积层提取浅层特征, 随后堆叠多个由 MFFB 模块构成的特征提取层 (feature extraction layer, FEL) 以挖掘深层特征, 最后通过一个专门设计的上采样器 (Upsampler) 重建高分辨率结果。本研究设计的上采样器集成了转置卷积与像素重排两种上采样路径, 并利用 SGFM 自适应地融合双路径输出, 以期在增大图像尺度的同时, 更好地恢复利于语义分割的高频结构信息。在训练阶段, 为促使重建影像在结构和细节上更贴合分割任务的需求, 采用结合 L_1 损失函数^[25]与结构相似性损失 L_{SSIM} ^[26]的混合损失函数进行超分辨率模型优化, 其定义如式 (13) 所示:

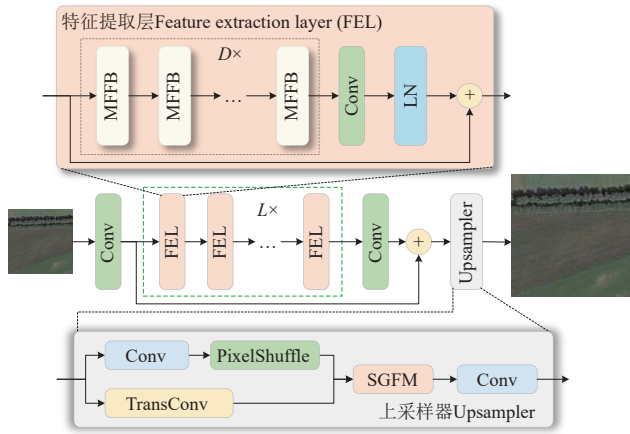
$$L_{sup} = \lambda_1 |HR_{GT} - HR_{SR}| + \lambda_2 L_{SSIM}(HR_{GT}, HR_{SR}) \quad (12)$$

式中 λ_1 、 λ_2 为平衡两项损失的权重系数, HR_{SR} 为模型输出的超分辨率重建图像, HR_{GT} 为高分辨率图像。最终, 训练完成的超分辨率模型被固定为前端处理器, 用于在语义分割前对输入影像进行分辨率提升。

1.5 模型的训练与评估

在配置有 Intel Xeon Silver 4314 处理器, 128 GB 内存, Nvidia L20 计算卡, Windows Server 2022 操作系统, CUDA 11.8 加速库的多台计算机上开展模型的训练及测试试验。基于 PyTorch 2.7, 应用 Python 实现遥感影像语义分割模型及超分辨率预处理模型, 同时实现图 2 技术框

架对应的计算机程序。用农田防护林网训练集和 AdamW 优化器训练分割模型，学习率设定为 10^{-5} ，输入图像分辨率设为 512×512 像素，训练过程中采用随机裁剪图像块，图像块随机翻转和颜色抖动等方法对数据进行增广。语义分割模型的性能评估采用类别精确率 (precision, P)、召回率 (recall, R)、交并比 (intersection over union, IoU)、像素精度 (pixel accuracy, PA) 和平均交并比 (mean IoU, mIoU) [26,34] 等指标。遥感影像超分辨率预处理模型的训练则采用与文献 [23] 相同的方法。



注：D 为 MFFB 的重复次数；L 为 FEL 的重复次数；PixelShuffle 表示像素重排操作；TransConv 表示反卷积。下同。
Note: D and L denote the number of repetitions for MFFB and FEL, respectively; PixelShuffle is pixel shuffling operation; TransConv is transposed convolution. Same below.

图 7 遥感影像超分辨率网络模型

Fig.7 Remote sensing image super-resolution network model

2 结果与分析

2.1 超分辨率预处理对语义分割效果的影响

为定量评估超分辨率预处理对农田林网语义分割的

有效性，采用本研究语义分割模型 MFF-Net，设置 3 组对照试验，试验 I 为对照组，直接输入原始分辨率影像，试验 II 的输入为经 2 倍超分辨率上采样的影像，试验 III 则输入经 4 倍超分辨率上采样的影像。在本研究中，MFF-Net 的超参数设置为 $L_1=2$ 、 $L_2=4$ 、 $L_3=8$ 、 $L_4=4$ 、 $D_1=64$ 、 $D_2=128$ 、 $D_3=256$ 、 $D_4=512$ 、 $C=320$ 、 $N=96$ 。分别计算 3 组试验的语义分割精度，结果如表 1 所示。

由表 1 可知，超分辨率预处理对分割性能具有明显的提升作用。与对照组相比，4 倍上采样的耕地和防护林带分割精确率分别提高 0.91 和 4.32 个百分点，IoU 分别提升 1.85% 和 6.61%，语义分割的 mIoU 提升 2.51%。超分辨率预处理对耕地和防护林带的分割均具有促进作用，特别是有效提高了防护林带的分割精度。超分辨率重建增强了图像的高频分量，可改善语义分割模型的纹理判别能力，同时超分辨率预处理提高了狭长防护林带的宽度和边缘细节信息，有效缓解了语义分割模型难以分割狭长目标的问题。本研究还同时开展了 8 倍上采样的分割试验，但出现了性能快速下降的问题，这是因为大幅上采样后的影像出现了伪影，影像风格与训练样本有很大差异，使分割模型性能大幅下降，故不再报告此数据。总体上，一定倍数的超分辨率预处理对分割精度具有显著的提升作用，特别是难以分割的狭长防护林带。

2.2 语义分割模型性能综合评估

2.2.1 不同模型的分割结果对比

为验证 MFF-Net 的有效性，将其与多种具有代表性的先进语义分割模型进行对比，包括基于 Mamba^[18] 的 RS³Mamba^[27]、基于 Swin Transformer^[15] 的 DC-Swin^[28]、在基础模型 SAM^[29] 微调的 SSRS^[30]、基于 SAM2^[31] 微调的 SAM2-CD^[32]，以及基于卷积的 DeepLabV3+^[14]、SegMAN^[33] 和 UNet^[8]。用农田林网数据集对各模型进行训练和测试，所有模型的输入均为 4 倍上采样超分辨率预处理后的影像，结果如表 2。

表 1 超分辨率对语义分割效果的影响

Table 1 Impact of super-resolution on semantic segmentation performance

编号 No.	上采样倍数 Upsampling factor	耕地 Cropland			防护林带 Shelterbelt			平均交并比 Mean IoU (mIoU)	像素精度 Pixel accuracy (PA)
		精确率 Precision P	召回率 Recall R	交并比 Intersection over union(IoU)	精确率 P	召回率 R	交并比 IoU		
I	-	95.51	97.70	93.40	78.51	85.75	69.44	81.41	93.95
II	2	96.36	98.17	94.87	79.70	87.65	71.88	82.69	94.20
III	4	96.42	98.40	95.13	82.83	87.99	74.03	83.45	94.89

表 2 不同语义分割模型的性能比较

Table 2 Performance comparison of semantic segmentation models

模型 Models	耕地 Cropland			防护林带 Shelterbelt			mIoU/%	PA/%	参数量 Parameters/M
	P /%	R /%	IoU/%	P /%	R /%	IoU/%			
MFF-Net	96.42	98.40	95.13	82.83	87.99	74.03	83.45	94.89	94.07
RS ³ Mamba ^[26]	93.09	95.81	92.08	82.53	85.05	73.99	81.69	93.20	45.63
DC-Swin ^[27]	94.10	96.41	93.07	82.55	85.06	74.01	82.36	94.55	96.17
SSRS ^[29]	90.21	95.69	91.74	79.91	82.82	70.68	80.14	91.01	48.72
SAM2-CD ^[31]	92.80	96.13	92.07	80.09	83.80	71.58	81.22	93.16	85.31
DeepLabV3+ ^[13]	94.53	96.85	93.90	80.47	84.80	73.62	82.51	93.57	43.32
SegMAN ^[32]	94.25	97.03	93.86	81.26	85.42	72.89	82.18	93.84	91.56
UNet ^[7]	65.12	80.76	67.73	54.66	72.65	57.05	61.59	78.37	18.97

由表 2 可知，本研究提出的 MFF-Net 模型在各项评估指标上均表现最佳。模型在耕地分割上的精确率与 IoU 分别达到 96.42% 与 95.13%，在更具挑战性的防护林带分割上，二者也分别达到 82.83% 与 74.03%。模型的

综合性能指标 mIoU 为 83.45%，像素精度 PA 达 94.89%。相较于同样引入状态空间模型的 RS³Mamba，MFF-Net 模型 mIoU 提升 2.15%，防护林带召回率提升 2.94 个百分点，表明其能更有效地捕捉狭长林带目标，减少漏检。

与基于 Swin Transformer 的 DC-Swin 相比, MFF-Net 模型 mIoU 提升 1.32%, 体现了多特征融合策略相对于单一 Transformer 架构的优越性。与基于 SAM 系列微调的 SSRS 和 SAM2-CD 相比, MFF-Net 也有明显的优势。与广泛应用的 DeepLab V3+相比, 其在耕地和防护林带分割 IoU 上分别提升 1.31% 与 0.56%, mIoU 提高 1.14%, PA 提高 1.32 个百分点, 凸显了本文模型架构在特定任务上的针对性提升。与同样注重多尺度上下文建模的 SegMAN 相比, MFF-Net 在防护林带分割的召回率和 IoU 上也均有明显提升, 表明其融合频域特征的策略在捕捉狭长目标的连续性和完整性方面更具优势。与 UNet 相比, MFF-Net 的分割精度远高于前者。

在参数量方面, MFF-Net 低于 DC-Swin, 与 SegMAN

接近, 高于常用模型 DeepLabV3+等, 但本研究模型在耕地和防护林带的分割性能上也高于对比模型, 特别是有效提高了狭长林带的召回率。考虑到农田防护林监测、高精度农业遥感制图等任务对完整性要求极高, 精度优先的设计是合理的。图 8 展示了各模型在典型区域的分割结果可视化对比, MFF-Net 生成的分割掩码最接近真实标注。特别是在防护林带区域, RS³ Mamba 与 DeepLabV3+ 均出现明显断裂, 而 MFF-Net 能有效保持林带的连续性与完整性。同时, 在耕地与林带边界处, MFF-Net 的边界更加平滑精确, 展现出良好的细节保持能力。综上, MFF-Net 通过融合全局、局部与频域特征, 能够有效缓解传统模型在狭长林带分割中出现的断裂与边界模糊问题, 有效提升了耕地与防护林带的分割精度。

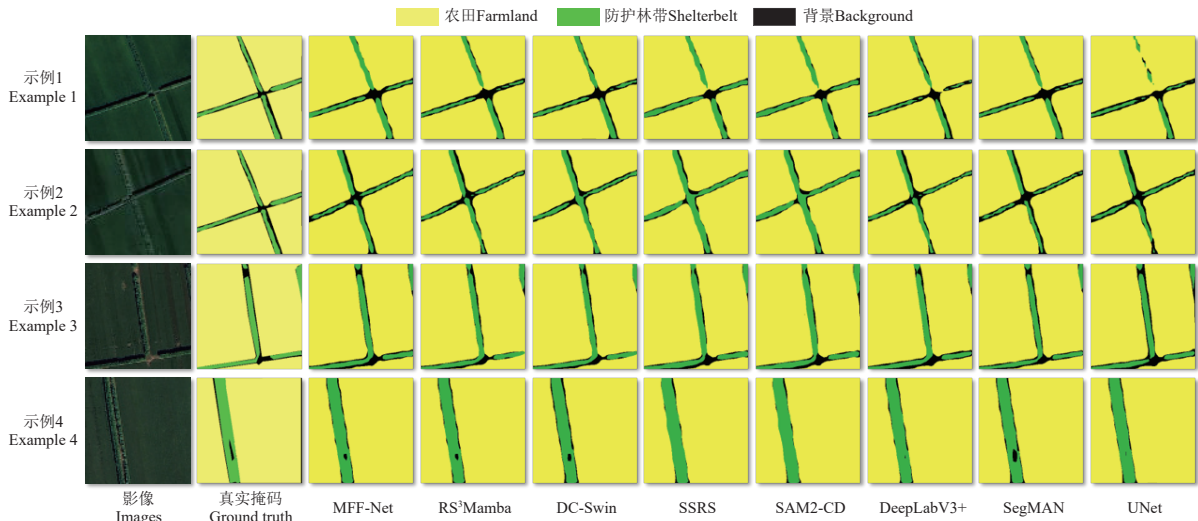


图 8 不同语义分割模型的分割结果
Fig.8 Segmentation results of different semantic segmentation models

为深入验证 MFF-Net 的泛化能力, 进一步在 LOVEDA 数据集^[35]上开展方法间比较试验 (表 3)。由表 3 可知, MFF-Net 在各类地物上取得了最高的分割精度。由于 LOVEDA 数据集场景复杂、类别多样且样本分布不均, 其中耕地更为小样本, 具有较高的分割难度。分割模型在耕地类别上得不到有效训练, 分割精度较低, 但本研究模型在该类地物的分割上仍优于其他方法。尤其在道路、河流等狭长地物上, MFF-Net 具有明显优势。上述结果说明, 即使在高难度数据集上, 本方法仍能保持最优性能, 表明方法是有效性。

表 3 不同语义分割模型在 LOVEDA 上的分割精度比较
Table 3 Performance comparison of different semantic segmentation models on the LoveDA dataset %

方法 Methods	IoU						mIoU
	建筑 Building	道路 Road	河流 Water	荒地 Barren	森林 Forest	耕地 Cropland	
MFF-Net	60.17	59.29	68.43	38.51	41.15	27.38	50.54
RS ³ Mamba ^[26]	58.75	57.92	61.00	37.24	39.67	25.64	46.93
DC-Swin ^[27]	58.91	57.57	67.69	25.93	37.10	20.85	49.22
SSRS ^[29]	57.83	56.18	66.37	26.39	38.94	23.52	47.99
SAM2-CD ^[31]	59.08	53.16	66.14	26.88	38.91	23.57	48.71
DeepLabV3+ ^[13]	52.78	51.89	63.47	29.84	34.23	18.44	44.12
SegMAN ^[32]	58.96	50.50	66.91	35.72	37.41	25.65	49.57
UNet ^[7]	52.35	48.17	54.26	20.19	32.47	16.32	42.61

2.2.2 基于分割结果的面积提取与应用分析
像素级的精度评估验证了 MFF-Net 模型的分割质量。然而, 在遥感宏观监测应用中, 地物面积的准确提取至关重要。为此, 本研究从农田防护林测试集中随机选择 4 个图像样本, 用 MFF-Net 生成分割掩码, 按图 3a 所示技术方法转换为矢量面数据 (Shape 格式), 并计算耕地与防护林带的面积提取精度, 进而从地理信息产品层面评估本方法的实用价值。其中样本图像覆盖区域的耕地与防护林带实际面积来自于第三次全国国土调查 (2021 年发布), 结果如表 4。

分析表 4 中的数据可知, 基于 MFF-Net 分割结果矢量化得到的面积具有较高的准确性。耕地面积反演的平均相对误差约为 7.50%, 而形态细长、破碎的防护林带, 其面积反演的平均相对误差甚至更低, 约为 6.76%。表明本研究方法不仅能在像素级别实现精确分割, 也具有产出较高精度地理信息产品的能力, 以满足实际业务中对面积统计的需求。防护林带面积反演精度高, 且因为 MFF-Net 对林带整体连续性和边界形态的良好保持, 部分抵消了因面积基数小而产生的绝对误差影响。

为进一步展示矢量化结果在空间分析中的应用潜力, 对其中一幅测试样本图像对应区域进行基于矢量数据的防护林带闭合度分析。根据国家标准《农田防护林工程设计规范》(GB/T 50817-2013), 田块周围防护林带的

闭合度在 0.75 以上为达标。基于矢量化结果，按图 3b 所示方法计算防护林带闭合度。结果表明（图 9），在影像覆盖范围内，按面积计算，91.95% 的田块周边防护林带闭合度达到标准要求。同时基于人工标注结果计算的闭合度达标情况为 89.37%，两者差异较小。该结果初步证明了本方法所提取的矢量数据能够有效支持复杂空间分析，为农田林网化程度的自动化、定量化评估提供了可能，展现了良好的工程应用前景。

表 4 耕地与防护林带面积提取精度评估
Table 4 Accuracy assessment of cultivated land and shelterbelt area extraction

图像样本 Image samples	耕地 Cropland				防护林带 Shelterbelt			
	实际面积 Actual area/hm ²	估计面积 Estimated area/hm ²	绝对误差 Absolute error/hm ²	相对误差 Relative error/%	实际面积 Actual area/hm ²	估计面积 Estimated area/hm ²	绝对误差 Absolute error/hm ²	相对误差 Relative error/%
1	1 134.64	1 062.85	71.79	6.33	82.43	86.85	4.42	5.36
2	1 233.59	1 108.49	125.10	10.14	81.26	86.19	4.93	6.07
3	1 211.37	1 141.47	69.90	5.77	60.97	67.09	6.12	10.04
4	1 202.05	1 108.66	93.39	7.77	89.77	94.77	5.00	5.57

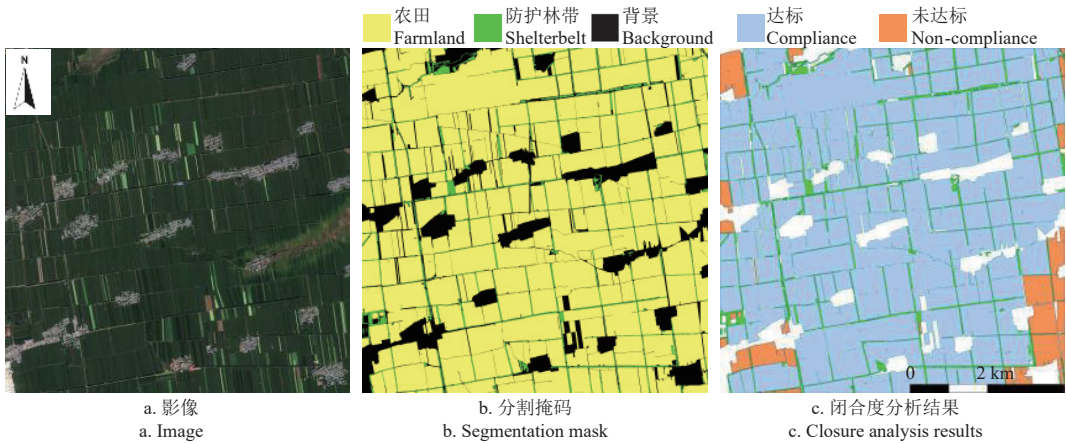


图 9 基于矢量化数据的防护林带闭合度分析结果
Fig.9 Analysis results of shelterbelt closure based on vectorized data

2.2.3 多特征融合模块的消融试验

为进一步分析 MFFB 核心子模块 MLLAFormer、FFTFormer 和 SGFM 对语义分割的作用，构建了 4 个变体模型。基线模型 A 不启用 MLLAFormer、FFTFormer 和 SGFM，即在 MFFB 中用 ConvFormer 取代 MLLA-

Former 和 FFTFormer，用逐元素加运算取代 SGFM。模型 B 在模型 A 基础上启用 MLLAFormer，模型 C 在模型 B 基础上启用 FFTFormer，模型 D 为本研究的最终模型，进一步启用了 SGFM。在农田林网数据集上，对各模型采用相同的训练和测试方法，结果如表 5。

表 5 不同变体模型的语义分割结果比较
Table 5 Comparison of semantic segmentation results across model variants %

模型编号 Model No.	处理 Treatments			耕地 Cropland			防护林带 Shelterbelt			mIoU	PA
	MLLAFormer	FFTFormer	SGFM	P	R	IoU	P	R	IoU		
A	×	×	×	94.86	98.01	93.38	79.19	86.76	71.13	79.37	93.28
B	√	×	×	95.76	98.12	94.03	79.97	86.93	71.87	80.63	93.86
C	√	√	×	96.29	98.38	94.45	81.58	87.48	73.88	82.05	94.05
D	√	√	√	96.42	98.40	95.13	82.83	87.99	74.03	83.45	94.89

由表 5 可知，随着 MFFB 中各子模块的逐步引入，模型的语义分割性能逐渐提高。在模型 A 的基础上，模型 B 引入 MLLAFormer，使得耕地分割 IoU 提高了 0.70%，精确率提高了 0.90 个百分点，防护林带分割 IoU 提高 1.04%，表明 MLLAFormer 的全局上下文建模能力，对大尺度和狭长目标的分割均具有促进作用。模型 C 在模型 B 的基础上，引入 FFTFormer，使得耕地和防护林带分割精确率分别提高 0.53 和 1.61 个百分点，IoU 分别提高 0.45% 和 2.80%，mIoU 提高 1.76%，表明 FFTFormer 的频域特征对狭长目标分割具有重要作用。频域特征提取能够增强狭长地物边界特征的清晰度，提高不同地物边界的可区分性。模型 D 在模型 C 的基础上，进一步引入 SGFM，使得防护林带分割精确率提高了 1.25 个百分点，mIoU 提升 1.71%，表明 SGFM 对语义

分割精度的提高具有一定贡献。SGFM 的空间门控加权能够协调局部、全局、频域空间特征的协同优化，对提升模型语义分割性能具有正向作用。相较于模型 A，模型 D 的耕地分割精确率和 IoU 分别提升 1.56 个百分点和 1.87%，防护林带的分割精确率和 IoU 分别提升 3.64 个百分点和 4.08%，mIoU 提高 5.14%，验证了 MFFB 模块设计的有效性。总体上，多特征融合的耦合设计，有效提升了农田林网语义分割精度。

3 结 论

本文提出了一种融合多特征与超分辨预处理的农田防护林带遥感语义分割模型，核心是设计了多特征融合模块 MFFB，其通过空间门控融合机制整合了类 Mamba 线性注意力的全局上下文建模能力、卷积的局部细节感

知能力以及傅里叶变换的频域边缘增强能力。在自建的农田防护林数据集上开展了试验验证, 主要结论如下:

1) 基于 MFFB 构建的语义分割模型 MFF-Net, 在农田防护林数据集上取得了耕地分割精确率 96.42%、防护林带分割精确率 82.83%、平均交并比 83.45% 的优异性能。

2) 通过消融试验验证了多特征融合与超分辨率预处理的有效性。频域特征的引入使防护林带分割交并比提升 2.80%, 空间门控融合机制进一步将平均交并比提升了 1.71%。引入超分辨率预处理流程, 通过 4 倍上采样重建影像细节, 使防护林带分割交并比大幅提升 6.61%, 平均交并比提升 2.51%, 有效缓解了狭长目标的分割难题。

3) 完成了从像素级分割掩码到矢量地理信息的转换, 并基于矢量化结果开展了面积提取与防护林带闭合度评估应用, 实现了完整技术闭环。面积提取结果表明, 耕地与防护林带的平均相对误差分别为 7.50% 与 6.76%, 为农田林网化程度的自动化、定量化评估提供了可靠的技术路径, 具备较强的工程应用前景。

[参 考 文 献]

- [1] 陈正发, 龚爱民, 张刘东, 等. 基于质量评价的省域尺度坡耕地质量调控体系构建[J]. 农业工程学报, 2021, 37(20): 136-145.
CHEN Zhengfa, GONG Aimin, ZHANG Liudong, et al. Construction of the quality regulation system for provincial scale slope farmland based on quality evaluation[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(20): 136-145. (in Chinese with English abstract)
- [2] 何新莹, 聂艳, 王朴, 等. 基于改进灰靶模型的耕地质量评价方法与实证[J]. 土壤学报, 2023, 60(4): 1007-1016.
HE Xinying, NIE Yan, WANG Pu, et al. Cultivated land quality evaluation method and demonstration based on improved grey target model[J]. Acta Pedologica Sinica, 2023, 60(4): 1007-1016. (in Chinese with English abstract)
- [3] 王春雷, 卢彩云, 李洪文, 等. 基于支持向量机的玉米根茬行图像分割[J]. 农业工程学报, 2021, 37(16): 117-126.
WANG Chunlei, LU Caiyun, LI Hongwen, et al. Image segmentation of maize stubble row based on SVM[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(16): 117-126. (in Chinese with English abstract)
- [4] 陈元鹏, 罗明, 彭军还, 等. 基于网格搜索随机森林算法的工矿复垦区土地利用分类[J]. 农业工程学报, 2017, 33(14): 250-257.
CHEN Yuanpeng, LUO Ming, PENG Junhuan, et al. Classification of land use in industrial and mining reclamation area based grid-search and random forest classifier[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2017, 33(14): 250-257. (in Chinese with English abstract)
- [5] CHANMEE S, KESORN K. Semantic decision trees: A new learning system for the ID3-based algorithm using a knowledge base[J]. Advanced Engineering Informatics, 2023, 58: 102156.
- [6] 高峰, 姜群鸥, 辛智鸣, 等. 荒漠绿洲区带状防护林遥感提取方法研究——以磴口为例[J]. 光谱学与光谱分析, 2022, 42(12): 3896-3905.
GAO Feng, JIANG Qunou, XIN Zhiming, et al. Extraction method of oasis shelterbelt systems based on remote-sensing images: A case study of Dengkou County[J]. Spectroscopy and Spectral Analysis, 2022, 42(12): 3896-3905. (in Chinese with English abstract)
- [7] 李萌, 年雁云, 边瑞, 等. 基于多源遥感影像的青海云杉和祁连圆柏分类[J]. 遥感技术与应用, 2020, 35(4): 855-863.
LI Meng, NIAN Yanyun, BIAN Rui, et al. Classification of Picea crassifolia and Sabina przewalskii based on multi-source remote sensing images[J]. Remote Sensing Technology and Application, 2020, 35(4): 855-863. (in Chinese with English abstract)
- [8] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C] // Medical Image Computing and Computer-Assisted Intervention — MICCAI 2015, Munich, Germany. Cham: Springer, 2015: 234-241.
- [9] 陈妮, 应丰, 王静, 等. 基于 U-Net 的高分辨率遥感图像土地利用信息提取[J]. 遥感技术与应用, 2021, 36(2): 285-292.
CHEN Ni, YING Feng, WANG Jing, et al. Research on land use information extraction based on U-Net[J]. Remote Sensing Technology and Application, 2021, 36(2): 285-292. (in Chinese with English abstract)
- [10] XU W, DENG X, GUO S, et al. High-resolution U-net: Preserving image details for cultivated land extraction[J]. Sensors, 2020, 20(15): 4064.
- [11] 朱昱, 潘耀忠, 张杜娟. 基于深度卷积网络和分水岭分割的耕地地块识别方法[J]. 地球信息科学学报, 2022, 24(12): 2389-2403.
ZHU Yu, PAN Yaozhong, ZHANG Dujuan. An agriculture parcel identification method based on convolutional neural network and watershed segmentation[J]. Journal of Geo-information Science, 2022, 24(12): 2389-2403. (in Chinese with English abstract)
- [12] RU F X, ZULKIFLEY M A, ABDANI S R, et al. Forest segmentation with spatial pyramid pooling modules: A surveillance system based on satellite images[J]. Forests, 2023, 14(2): 405.
- [13] 刘振国, 孙永旺, 张喜珍, 等. 基于改进 DeepLabV3+网络的荔枝种植面积提取方法[J]. 农业工程学报, 2025, 41(12): 191-197.
LIU Zhenguo, SUN Yongwang, ZHANG Xizhen, et al. Extracting lychee planting areas using improved DeepLabV3+ network[J]. Transactions of the Chinese Society of Agricultural Engineering (Transaction of the CSAE), 2025, 41(12): 191-197.

- (in Chinese with English abstract)
- [14] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C] //Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany. Cham: Springer, 2018: 801-818.
- [15] LIU Z, LIN Y T, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021, Montreal, QC, Canada. Piscataway, NJ: IEEE, 2021: 10012-10022.
- [16] SUZUKI K, ABE K. Topological structural analysis of digitized binary images by border following[J]. *Computer Vision, Graphics, and Image Processing*, 1985, 30(1): 32-46.
- [17] VISVALINGAM M, WHYATT J D. The Douglas-Peucker algorithm for line simplification: Re-evaluation through visualization[J]. *Computer Graphics Forum*, 1990, 9(3): 213-225.
- [18] GU A, DAO T. Mamba: Linear-time sequence modeling with selective state spaces[EB/OL]. (2023-12-01) [2024-07-20]. <https://arxiv.org/abs/2312.00752>.
- [19] HAN D, WANG Z, XIA Z, et al. Demystify Mamba in vision: A linear attention perspective[EB/OL]. (2024-05-26) [2024-07-20]. <https://arxiv.org/abs/2405.16605>.
- [20] KATHAROPOULOS A, VYAS A, PAPPAS N, et al. Transformers are RNNs: Fast autoregressive transformers with linear attention[C] //Proceedings of the 37th International Conference on Machine Learning (ICML), Virtual Only Conference, Online. New York: ACM, 2020: 5156-5165.
- [21] LIN X, YAN Z, DENG X, et al. ConvFormer: Plug-and-play CNN-style transformers for improving medical image segmentation[C] //The 26th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Vancouver, Canada. Cham: Springer, 2023: 642-651.
- [22] RAHAMAN N, BARATIN A, ARPIT D, et al. On the spectral bias of neural networks[C] //Proceedings of the 36th International Conference on Machine Learning (ICML), Long Beach, CA, USA. New York: ACM, 2019: 5301-5310.
- [23] FENG J X, LIU X Y. Frequency-quantized variational autoencoder based on 2D-FFT for enhanced image reconstruction and generation[J]. *Computers, Materials & Continua*, 2025, 83(2): 2087-2107.
- [24] QIAO J, LIAO J, LI W, et al. Hi-mamba: Hierarchical mamba for efficient image super-resolution[EB/OL]. (2024-10-16) [2024-07-20]. <https://arxiv.org/abs/2410.10140>.
- [25] TANCHENKO A. Visual-PSNR measure of image quality[J]. *Journal of Visual Communication and Image Representation*, 2014, 25(5): 874-878.
- [26] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: From error visibility to structural similarity[J]. *IEEE Transactions on Image Processing*, 2004, 13(4): 600-612.
- [27] MA X, ZHANG X, PUN M O. RS3Mamba: Visual state space model for remote sensing image semantic segmentation[J]. *IEEE Geoscience and Remote Sensing Letters*, 2024, 21: 1-5.
- [28] WANG L, LI R, DUAN C, et al. A novel transformer based semantic segmentation scheme for fine-resolution remote sensing images[J]. *IEEE Geoscience and Remote Sensing Letters*, 2022, 19: 1-5.
- [29] KIRILLOV A, MINTUN E, RAVI N, et al. Segment anything[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France. Piscataway, NJ: IEEE, 2023: 4015-4026.
- [30] MA X, WU Q, ZHAO X, et al. SAM-assisted remote sensing imagery semantic segmentation with object and boundary constraints[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 1-16.
- [31] RAVI N, GABEUR V, HU Y T, et al. SAM 2: segment anything in images and videos[EB/OL]. (2024-08-01) [2024-08-01]. <https://arxiv.org/abs/2408.00714>.
- [32] QIN Y, WANG C, FAN Y, et al. SAM2-CD: Remote sensing image change detection with SAM2[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2025.
- [33] FU Y, LOU M, YU Y Z. SegMAN: Omni-scale context modeling with state space models and local attention for semantic segmentation[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA. Piscataway, NJ: IEEE, 2025: 19077-19087.
- [34] 张羽丰, 杨景, 邓寒冰, 等. 基于 RGB 和深度双模态的温室番茄图像语义分割模型[J]. *农业工程学报*, 2024, 40(2): 295-306.
ZHANG Yufeng, YANG Jing, DENG Hanbing, et al. Semantic segmentation model for greenhouse tomato images using RGB and depth bimodal[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2024, 40(2): 295-306. (in Chinese with English abstract)
- [35] WANG J J, ZHENG Z, MA A L, et al. LoveDA: A remote sensing land-cover dataset for domain adaptive semantic segmentation[EB/OL]. (2021-10-17) [2024-07-20]. <https://arxiv.org/abs/2110.08733>.

Semantic segmentation of farmland shelterbelts using multi-feature and super-resolution remote sensing

ZHANG Qi , ZHOU Yuncheng^{*} , ZHAO Hongge , WU Wenhao , HUANG Yuekun

(College of Information and Electrical Engineering, Shenyang Agricultural University, Shenyang 110866, China)

Abstract: Farmland shelterbelts can be represented as typical narrow linear features in remote sensing imagery. However, the significant challenges remain in accurate segmentation, due to their strong global contextual dependencies and weak local features. In this study, a semantic segmentation model, named MFF-Net, was proposed to accurately and rapidly extract cultivated land and shelterbelts. Multi-feature fusion block (MFFB) and spatially gated fusion mechanism (SGFM) were also adaptively integrated with three complementary approaches: the long-range dependency modeling using a mamba-like linear attention (MLLA) operator, the local detail perception in convolutional networks, and the frequency-domain edge enhancement via fast Fourier transform (FFT). The feature representation was effectively balanced when segmenting elongated objects. Furthermore, a task-oriented super-resolution preprocessing was introduced into the network. This front-end step was designed to reconstruct high-resolution images. Thereby, the textual details were enhanced to sharpen the boundaries of subtle features, like shelterbelts. Superior input was provided for the subsequent segmentation model. A series of experiments was performed on the self-constructed dataset of the farmland shelterbelt. The results demonstrate that the superior performance of MFF-Net was achieved in the precision rates of 96.42% for cropland and 82.83% for shelterbelts, with a mean intersection over union (mIoU) of 83.45%, thus outperforming a range of advanced models, including RS³Mamba, DC-Swin, DeepLabV3+, and SegMAN. Ablation studies validated the effectiveness of each component within the multi-feature fusion. Frequency-domain features via FFTFormer contributed to a 2.80% increase in the shelterbelt IoU, indicating the enhanced boundary discrimination. The spatially gated fusion mechanism (SGFM) further boosted the overall mIoU by 1.71%, indicating the adaptive balance on the contribution rates from different feature domains. Compared with a baseline model, the full MFFB was improved mIoU of 5.14%. Super-resolution preprocessing was integrated with the highly effective auxiliary strategy. Taking 4x upsampled images as the input, there was a remarkable 6.61% increase in shelterbelt IoU and a 2.47% gain in overall mIoU. The difficulties with segmenting narrow targets were effectively mitigated to augment their pixel-width and edge clarity. A complete technical pipeline was established from pixel-level segmentation to vectorization and application. Accurate area estimation was achieved with average relative errors of 7.50% for cropland and 6.76% for shelterbelts, compared with the national survey data. An application analysis of shelterbelt closure degree was also made on individual farmland plots. The 91.95% of the plots met the national standard requirement (≥ 0.75). In conclusion, the MFF-Net model can be expected to effectively segment the narrow linear features in complex landscapes. Synergistic fusion of global, local, and frequency-domain features was combined with task-oriented super-resolution enhancement. This research can provide a robust technical pathway for the precise monitoring and dynamic evaluation of farmland shelterbelt networks in ecological conservation.

Keywords: remote sensing; semantic segmentation; farmland shelterbelt; multi-feature fusion; super-resolution; deep learning