CN 51-1346/O4 ISSN 1003-501X (印刷版) ISSN 2094-4019 (网络版)

光电工程
Opto-Electronic Engineering

中文核心期刊 中国科技核心期刊
Scopus CSCD

融合ResNeSt和多尺度特征融合的遥感影像道路提取

郝明，白鹤，徐婷婷

引用本文:

郝明，白鹤，徐婷婷. 融合ResNeSt和多尺度特征融合的遥感影像道路提取[J]. 光电工程，2025， **52**(1): 240236.

Hao M, Bai H, Xu T T. Remote sensing image road extraction by integrating ResNeSt and multi-scale feature fusion[J]. *Opto-Electron Eng*, 2025, **52**(1): 240236.

https://doi.org/10.12086/oee.2025.240236

收稿日期: 2024-10-09; 修改日期: 2024-12-16; 录用日期: 2024-12-16

相关论文

融合元素乘法和细节优化的道路提取算法
张进，吕明海，冯永安，张莹
光电工程 2024, **51**(12): 240210 doi: 10.12086/oee.2024.240210

无人机视角下的道路损伤检测算法MAS-YOLOv8n
王晓燕，王禧钰，李杰，梁文辉，牟建宏，毕楚然
光电工程 2024, **51**(10): 240170 doi: 10.12086/oee.2024.240170

面向道路场景语义分割的移动窗口变换神经网络设计
杭昊，黄影平，张栩瑞，罗鑫
光电工程 2024, **51**(1): 230304 doi: 10.12086/oee.2024.230304

更多相关论文见光电期刊集群网站

OEE 光电工程 Opto-Electronic Engineering
http://cn.oejournal.org/oee
OE_Journal Website

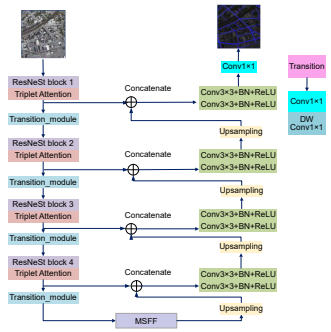
Opto-Electronic Engineering
光 电 工 程

Article

2025 年, 第 52 卷, 第 1 期

DOI: 10.12086/oee.2025.240236 CSTR: 32245.14.oee.2025.240236

融合 ResNeSt 和多尺度特征融合的遥感影像道路提取



郝 明*，白 鹤，徐婷婷

辽宁理工学院信息工程学院，辽宁 锦州 121000

摘要：针对高分辨率遥感影像的道路提取存在道路边缘分割不连续、小目标道路分割精度不高和目标道路误分的问题，本文提出了结合 ResNeSt 和多尺度特征融合的遥感影像道路提取方法用于遥感影像道路提取 (ResT-UNet)。参考 ResNeSt 网络模块构造 U 型网编码器，使前期编码器可以更完整的提取信息，分割目标边缘更加连续；首先在编码器部分引入 Triplet Attention 注意力机制，抑制无用的特征信息；其次使用卷积块代替最大池化操作，增加特征维度和网络深度，减少道路信息丢失；最后在编码器网络和解码器网络的桥连接部分使用多尺度特征融合模块 (multi-acale feature fusion, MSFF)，以捕获区域间的远程依赖关系，提高道路的分割效果。实验在 Massachusetts 道路数据集和 DeepGlobe 数据集上进行实验，实验结果表明，该方法分别在数据集上 IoU 达到了 64.76% 和 64.45%，相比于近几年网络 MINet 模型提高了 1.42% 和 1.74%，表明 ResT-UNet 网络有效提高遥感影像道路的提取精度，为解译遥感图像语义信息提供一种新思路。

关键词：遥感影像；道路提取；ResNeSt 网络；多尺度特征融合；注意力机制

中图分类号：TP391 **文献标志码：**A

郝明，白鹤，徐婷婷. 融合 ResNeSt 和多尺度特征融合的遥感影像道路提取 [J]. 光电工程，2025，**52**(1): 240236

Hao M, Bai H, Xu T T. Remote sensing image road extraction by integrating ResNeSt and multi-scale feature fusion[J]. *Opto-Electron Eng*, 2025, **52**(1): 240236

Remote sensing image road extraction by integrating ResNeSt and multi-scale feature fusion

Hao Ming*, Bai He, Xu Tingting

School of Information Engineering, Liaoning Institute of Science and Technology, Jinzhou, Liaoning 121000, China

Abstract: Aiming at the issues of discontinuous road edge segmentation, low accuracy in segmenting small-scale roads, and misclassification of target roads in high-resolution remote sensing imagery, this paper proposes a road extraction method that integrates ResNeSt and multi-scale feature fusion for road extraction from remote sensing imagery. Referencing the ResNeSt network module, a U-shaped network encoder is constructed to enable the initial encoder to extract information more entirely and ensure more continuous segmentation of target edges.

收稿日期：2024-10-09；修回日期：2024-12-16；录用日期：2024-12-16

基金项目：辽宁省教育厅基本科研项目 (JYTMS20230965)

*通信作者：郝明，haoming1232023@163.com。

Firstly, Triplet Attention is introduced into the encoder to suppress useless feature information. Secondly, convolutional blocks replace max pooling operations, increasing feature dimensionality and network depth while reducing the loss of road information. Finally, a multi-scale feature fusion (MSFF) module is utilized at the bridge connection between the encoder and decoder networks to capture long-range dependencies between regions and improve road segmentation performance. The experiments were conducted on the Massachusetts Roads dataset and the DeepGlobe dataset. The experimental results demonstrate that our proposed method achieved Intersection over Union scores of 65.39% and 65.45% , respectively, on these datasets, representing improvements of 1.42% and 1.74% compared to the original MINet model. These findings indicate that the ResT-UNet network effectively enhances the extraction accuracy of road features in remote sensing imagery, providing a novel approach for interpreting semantic information in remote sensing images.

Keywords: remote sensing imagery; road extraction; ResNeSt network; multi-scale feature fusion; attention mechanism

1 引　言

从遥感影像中提取道路信息可以应用于城市规划[1]、自动驾驶[2]和道路信息更新等诸多领域。深度学习中的语义分割技术通过对图片中的每个像素进行分类，将图像分为目标和背景，通过语义分割技术提取道路信息已经成为遥感影像任务的主流[3-5]。但遥感影像中的道路信息往往是复杂的、不规则的，要想准确、完整地提取道路目标仍是一项挑战。

随着深度学习的发展，利用卷积神经网络进行城乡道路提取已成为研究热点。随着空洞卷积[6]、注意力机制[7]等方法的逐渐使用，Lin 等[8]提出基于空洞卷积 U-Net 的遥感影像道路提取方法，学习更多的语义信息来改善提取结果出现的模糊问题。Yang 等[9]提出使用结合了上下文信息和注意力机制的 U-Net 网络模型对道路进行提取，利用上下文信息提取模块对道路的上下文信息进行整合，以达到提取道路的最优结果。受深度残差学习[10]和 U-Net[11]的启发，Zhang 等[12]提出深度残差 U-Net 用于道路提取，通过残差结构简化深层网络的训练，减少参数量，同时网络中大量跳跃连接促进了信息传播，实现了更好的道路提取效果。为进一步获取详细的空间信息，Zhou 等[13]提出 D-LinkNet，用于高分辨率卫星图像的语义分割，在 Linknet34 中心区域增加了串并联的扩张卷积，解决了 Linknet34 中心感受野不足的问题，提高了提取道路的精度。这一改进不仅增强了网络对道路特征的捕捉能力，还为后续的道路提取研究提供了新的思路。在此基础上，Zhang 等[14]进一步提出了 F-LinkNet，F-LinkNet 在 LinkNet 的基础上加入了多尺度特征的聚合模块，这一模块能够扩大特征感受野范围，从而改善网络提取结果的连通性问题。此外，研究者们还尝试将其他网络架构与 D-LinkNet 相结合，以进一步提升道路提取的性能。例如，Gao 等[15]提出了 DAD-LinkNet 网络架构，该架构引入了一种双重注意力机制来捕获多样的道路特征。同时，还通过引入一个权重参数来改进损失函数，以提高道路提取结果的准确性。

上述网络方法虽然一定程度上提高了网络的分割精度，但基于编码器和解码器与空洞卷积结合的方法会忽略信息边缘的分割，导致边缘分割不连续。其次，基于空洞卷积方法虽然在增大感受野的同时保持特征图的分辨率不变，但当多个空洞卷积进行叠加时，会产生网格效应，损失信息的连续性和相关性。因此本文提出了一种结合 ResNeSt 和多尺度特征融合的遥感影像道路提取方法 (ResT-UNet)。针对 U-Net 网络进行遥感图像特征提取时，出现梯度消失、模型过拟合的现象，提出用 ResNeSt block 和 Triplet Attention 模块来代替 U-Net 网络特征提取部分的卷积，增强网络特征提取能力；针对上下文信息考虑不完整问题，在 U-Net 的编码器和解码器连接部分引入多尺度特征融合 (multi-scale feature fusion, MSFF) 模块，旨在捕捉道路的多维度、多尺度特性，从而实现对道路细节和全局结构的精准预测。

2 相关理论

2.1 ResNeSt 网络

ResNeSt 网络[16]结构如图 1 所示，其是对 ResNet 的改进，其核心模块将输入划分为 K 个基数组，每个基数组再进一步划分为 R 个分散单元。这些分散单元

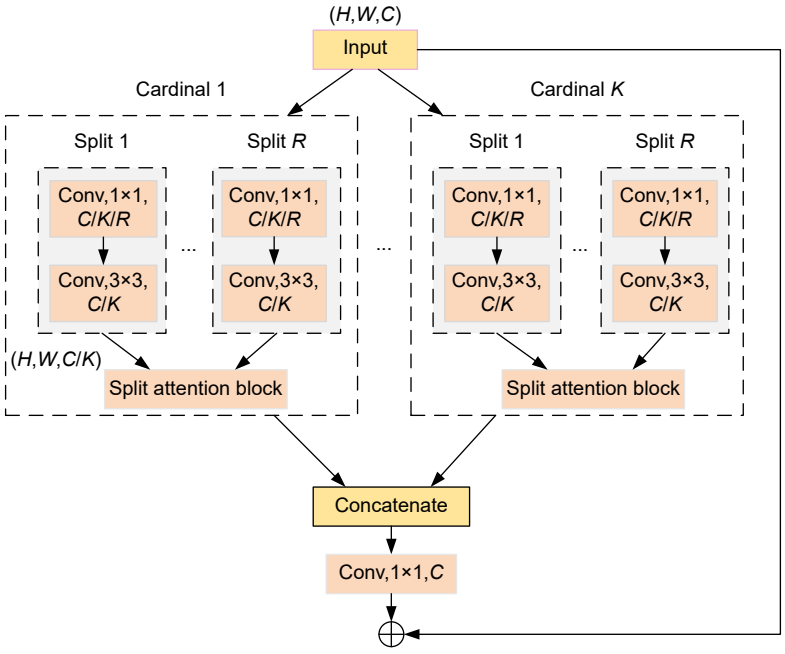


图 1 ResNeSt 网络总体结构图

Fig. 1 Overall structure diagram of ResNeSt network

通过分散注意力模块处理，该模块为各个通道分配权重，并将所有处理结果拼接起来。这种设计旨在有效减少梯度消失和网络退化问题。每个基数组均包含一个分散注意力模块 (split-attention block)，具体结构如图 2 所示，其中 U 表示由不同输入组成的特征向量的集合，j 为索引，用于标记不同分支的特征。

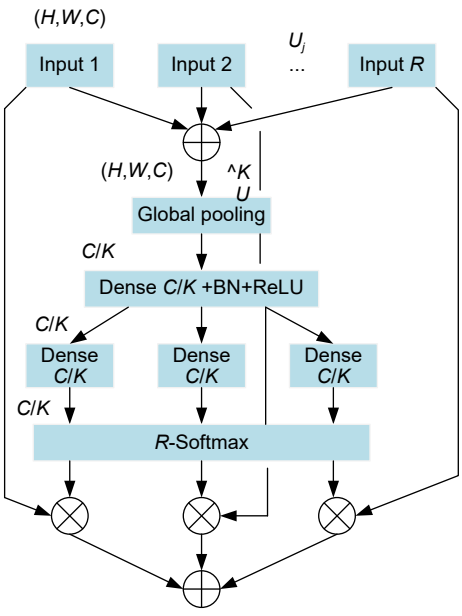


图 2 Split-attention block 结构图

Fig. 2 Structure diagram of split-attention block

在分散注意力模块中，输入的特征图首先被划分为 K 个基数组，每个基数组再进一步被划分为 R 个分散单元。这些分散单元随后被送入模块内部进行处理。模块的核心操作是为每个分散单元分配权重，模块通过对每个特征图小组施加全局平均池化，得到表示各个通道权重的特征向量。然后，经过一系列变换 (BN+ReLU+Softmax) 对权重向量进行修正，最终得到每个特征图小组的加权组合。这种加权组合的方式实现了对特征图跨组的注意力机制，使得网络能够更加关注重要的特征，同时抑制不重要的背景噪声。

2.2 三重注意力机制

为了增强 U-Net 编码器对通道和空间位置间相互依赖性的捕捉能力，提升信息输出效率和边缘细节分割精度，引入了轻量化的 Triplet Attention 注意力机制[17]，融合了不同维度的信息交互与增强。Triplet Attention 注意力机制结构图如图 3 所示，前两个分支并行工作，主要目标是探索并增强特征图中通道维度 (C) 与空间维度 (高度 H 和宽度 W) 之间的复杂关系。各自采用不同的策略来捕捉这种跨维度的相互作用。这样，每个分支都能从对方维度中提取关键信息，实现通道与空间信息的互补与增强。

第三个分支：该分支专注于构建空间注意力图，该图能够高亮特征图中对于当前任务最为关键的空间位置。通过分析空间维度上的特征分布，该分支能够识别出哪些区域或像素点对于整体特征表示最为重要。

通过生成空间注意力权重，该分支能够增强这些关键区域的信息，同时抑制不重要的背景噪声，从而提高模型对空间信息的敏感度和利用效率。

3 结合 ResNeSt 和多尺度特征融合的遥感影像道路提取

本文 ResT-UNet 网络架构以 U-Net 网络为基准模型进行改进，ResT-UNet 网络结构由编码器、解码器、中间层和跳跃连接组成，如图 4 所示。

3.1 ResT-UNet 编码器网络结构

ResT-UNet 网络的编码器由 U-Net 编码器基于 ResNeSt 网络重构而成。首先将 U-Net 的第一层第一个 Conv 替换为 ResNeSt block，其次在第一层第二个 Conv 替换为 Triplet Attention 注意力机制；最后将每

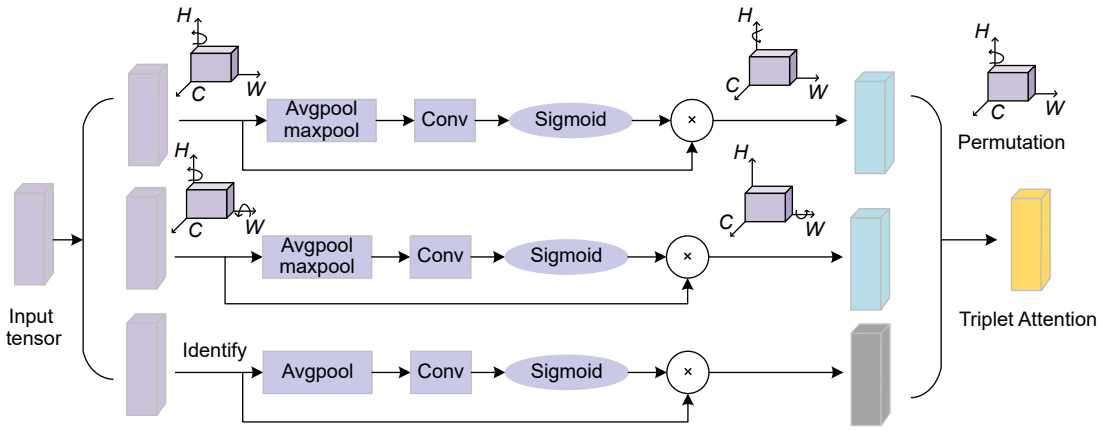


图 3 三重注意力机制原理图

Fig. 3 Schematic diagram of Triplet Attention mechanism

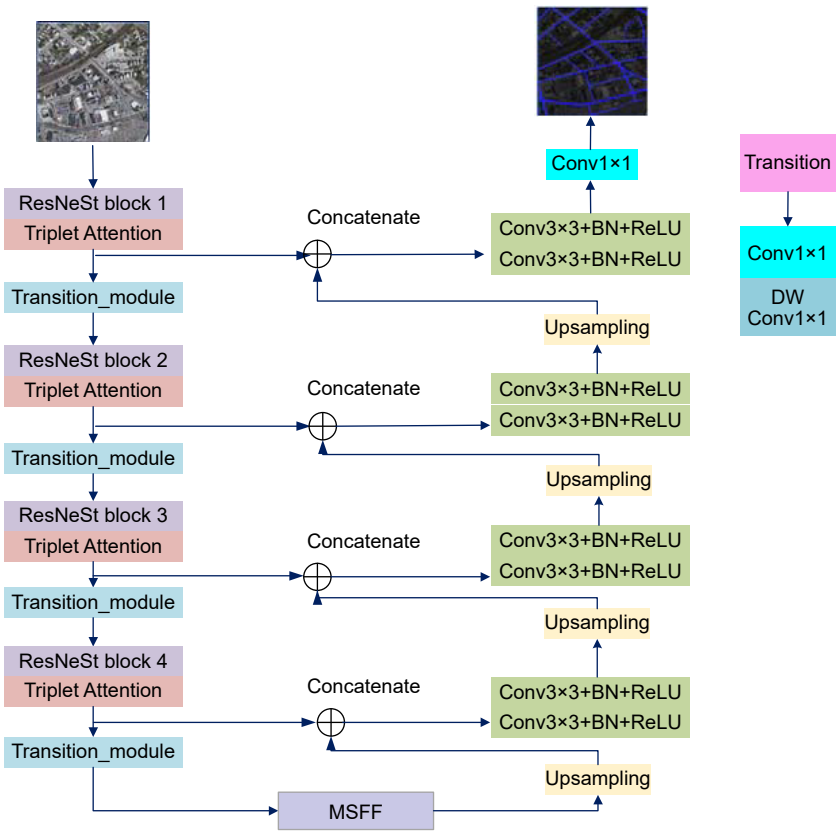


图 4 ResT-UNet 网络总体结构图

Fig. 4 The overall structure of ResT-UNet network

一层的 Max pooling 层替换为 Transition。其中：

1) ResNeSt block 网络结构如图 1 所示。在 ResNeSt 网络中提出了一个分散注意力块 (split-attention block)，通过对分散注意力块的叠加，使网络能够跨特征提取语义信息，提高分辨率遥感图像道路特征。

2) 在降采样操作时引入 Triplet Attention 注意力机制，使得网络对关键信息进行提取。

3) 将 Max pooling 层替换成 Transition，Transition 对比于最大池化模块，Transition 可以通过 1×1 Conv 增加特征维度和网络深度，同时 3×3 DWconv 进行 2 倍下采样。

3.2 ResT-UNet 中间层

在 ResT-UNet 架构中，多尺度特征融合 (multi-scale feature fusion, MSFF) 模块作为连接编码器和解码器的关键组件，运用了空洞卷积技术，以优化特征提取与融合过程。这一设计策略不仅有效降低了模型参数的数量，还通过扩大卷积操作的感受野，使得每个卷积层能够捕获并输出更为丰富和全面的信息。如图 5 所示，MSFF 模块的核心为多路径并行处理机制，能够将编码器深层输出的高维特征图作为输入，随后将这些特征图分配到三条独立的处理路径上。每条路径均配置有空洞卷积层，但各自采用不同的空洞率，以此实现对输入特征图在不同尺度上的特征提取。这种并行处理的方式确保了模型能够同时关注到道路图像的局部细节以及全局结构，从而增强了模型对道路形状和宽度多样性的适应能力。通过应用不同空洞率的空洞卷积，MSFF 模块能够捕获到不同感受野下的道路特征信息。这些特征信息在各自路径上被独立处理后，再经过特征融合步骤被整合成一个统一的特征表示。这一过程不仅丰富了特征图的语义内容，还提高了模型对复杂道路场景的理解能力。计算过程可表示为

$$F_{\mathrm{a}} = f_1^{\mathrm{a}}(F_{ia}) + f_2^{\mathrm{a}}[f_1^{\mathrm{a}}(F_{ia})] + f_4^{\mathrm{a}}\{f_2^{\mathrm{a}}[f_1^{\mathrm{a}}(F_{ia})]\} + f_8^{\mathrm{a}}\{f_4^{\mathrm{a}}\{f_2^{\mathrm{a}}[f_1^{\mathrm{a}}(F_{ia})]\}\}, \tag{1}$$

$$F_{\mathrm{b}} = f_1^{\mathrm{b}}(F_{ia}) + f_2^{\mathrm{b}}\left[f_1^{\mathrm{b}}(F_{ia})\right] + f_4^{\mathrm{b}}\left\{f_2^{\mathrm{b}}\left[f_1^{\mathrm{b}}(F_{ia})\right]\right\} + f_8^{\mathrm{b}}\left\{f_4^{\mathrm{b}}\left\{f_2^{\mathrm{b}}\left[f_1^{\mathrm{b}}(F_{ia})\right]\right\}\right\}, \tag{2}$$

$$F_{\mathrm{c}} = f_1^{\mathrm{c}}(F_{ia}) + f_2^{\mathrm{c}}[f_1^{\mathrm{c}}(F_{ia})] + f_4^{\mathrm{c}}\{f_2^{\mathrm{c}}[f_1^{\mathrm{c}}(F_{ia})]\} + f_8^{\mathrm{c}}\{f_4^{\mathrm{c}}\{f_2^{\mathrm{c}}[f_1^{\mathrm{c}}(F_{ia})]\}\}, \tag{3}$$

式中：F_{a}、F_{b}和F_{c}分别表示三条并联的空洞卷积支路各自负责提取输入特征图在不同感受野下的特征信息，其中f_i^{a}采用空洞率为 i 的普通空洞卷积，关注一般的局部特征；f_i^{b}采用水平空洞卷积，其设计可能侧重于捕捉水平方向上的特征；f_i^{c}采用垂直空洞卷积，更侧重于垂直方向上的特征提取；上、下标 a 表示普通空洞卷积分支的特征图，负责提取全局特征；上、下标 b 表示水平空洞卷积分支的特征图，专注于捕获水平方向的上下文信息；上、下标 c 表示垂直空洞卷积分支的特征图，专注于捕获垂直方向的上下文信息。除了这三条并行的空洞卷积支路外，MSFF 模块还包含两条全局池化支路，分别用于进行水平池化和垂直池化。这两条支路的作用是从输入特征图中提取全局性的水平特征和垂直特征。通过全局池化操作，能够捕捉到整个特征图在水平或垂直方向上的统计信息，过程可表示为

$$F_{\mathrm{bp}} = deconv\{CBR[AvgPool_{H,1}(F_{ia})]\}, \tag{4}$$

$$F_{\mathrm{cp}} = deconv\{CBR[AvgPool_{1,W}(F_{ia})]\}, \tag{5}$$

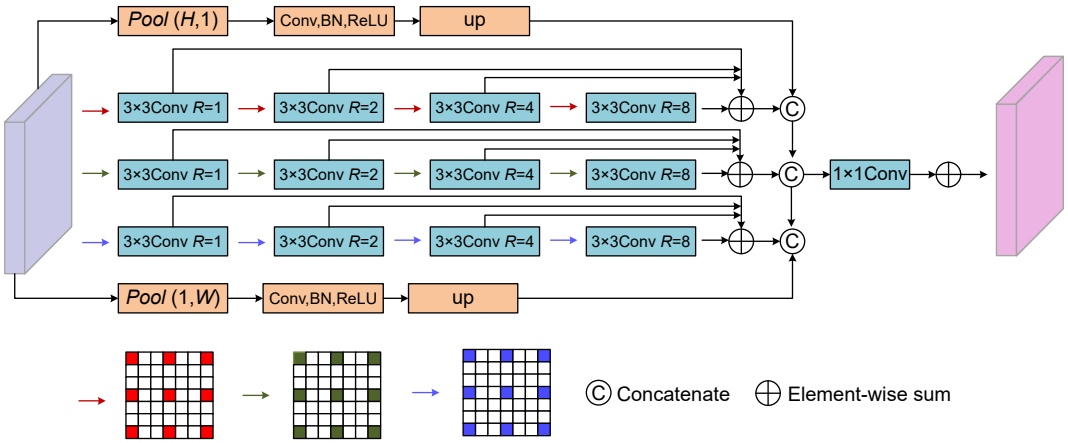


图 5 多尺度特征融合模块

Fig. 5 Multi-scale feature fusion module

式中：$deconv(\cdot)$表示反卷积；$CBR(\cdot)$包含卷积 Conv、批归一化和激活函数 ReLU；$AvgPool_{H,1}(\cdot)$和$AvgPool_{1,W}(\cdot)$分别表示池化核大小为$H\times1$和$1\times W$的平均池化；下标 bp 表示水平方向的局部上下文特征；下标 cp 表示垂直方向的局部上下文特征。最后将各个支路获得的特征图融合，此外，采用带权重的残差结构抑制模块产生的冗余信息，实现更有效的上下文信息提取。

$$F_{\text{out}}=p\cdot f^{1\times1}[concat(F_{\text{a}};F_{\text{b}};F_{\text{c}};F_{\text{bp}};F_{\text{cp}})]+(1-p)\cdot F_{ia}\quad,\tag{6}$$

式中：$concat(\cdot)$表示拼接操作；$f^{1\times1}$为1×1卷积，用于通道调整；p为设置的可训练参数，用于自学习地为残差连接分配权重。

3.3 ResT-UNet 解码器网络结构

ResT-UNet 网络结构的解码器由 U-Net 的解码器构成，利用跳跃连接，能够使网络在每一次上采样过程中，将编码器和解码器对应层的特征拼接起来，即考虑了高层特征也考虑了低层特征；其次由上采样和 3×3 Conv 进行 4 次序列运算，最后利用 1×1 Conv 进行降维处理，得到二分类的特征图。

1) 3×3 Conv：减少网络的参数和复杂度，加快网络的训练速度，以捕获更小的特征。

2) 1×1 Conv：对每个像素点在不同的通道上进行线性组合，实现二分类。

3) Upsampling：使用转置 Conv 进行上采样，扩大特征图的大小，合并特征信息，使特征图恢复成原始输入图像大小。

4 实验结果与分析

4.1 实验环境

实验环境设定于 64 位 Windows 10 操作系统之上，采用 PyTorch 深度学习框架作为模型构建的基础工具。为了加速计算过程与提升训练效率，实验配置了高性能的硬件支持，具体为 NVIDIA GeForce RTX 3090 图形处理单元的计算平台。

4.2 数据集

实验采用了马萨诸塞道路数据集与 DeepGlobe 道路数据集来验证所提出的方法。以下是对这两个数据集的具体阐述：

1) 马萨诸塞道路数据集 (Massachusetts)，该数据集收录了 1171 张遥感图像，每张图像的尺寸为 1500 pixel×1500 pixel，地面分辨率高达 1.2 m/pixel。数据集被精心划分为 1108 张训练样本、49 张测试样本以及 14 张验证样本。如图 6(a) 所示，前景道路像素以白色呈现，而背景像素则以黑色表示。

2) DeepGlobe 道路数据集源自 2018 年的 DeepGlobe 道路提取挑战赛。该数据集总共包含了 6226 张图像，每张图像的分辨率为 1024 pixel×1024 pixel，并具有三个颜色通道，其地面分辨率为 0.5 m。这些图像广泛覆盖了泰国、印度和印度尼西亚等多个地区的山路、混凝土路以及沥青路。如图 6(b) 所示，前景道路同样以白色显示，背景则以黑色标示。

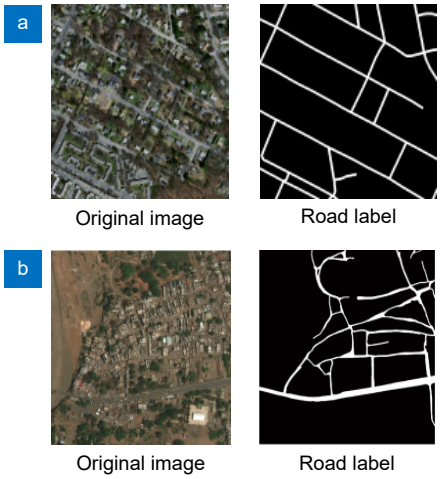


图 6 部分道路标签。(a) Massachusetts; (b) DeepGlobe

Fig. 6 Partial road labe. (a) Massachusetts; (b) DeepGlobe

4.3 数据增强

图像数据增强的目的是通过各种技术和方法来改善图像的视觉效果、提高图像清晰度、突出机器学习模型关注的特定信息，并减少模型对训练数据的过度依赖以防止过拟合。由于上述两个道路数据集的图像尺寸较大，而设备和计算能力有限，因此将原始 1024×1024 或 1500×1500 的图像裁剪成 512×512 的图像。为了更高效、全面地利用有限的训练数据，采用了一系列几何和度量的方法进行图像数据增强。几何增强方法包括随机裁剪、水平和垂直翻转等，而度量增强方法包括随机亮度变换、随机对比度变换、饱和度和色调变换等。这些增强方法主要用于训练集，而测试集和验证集的数据增强则相对简单，使用水平和垂直翻转等变换已足够。在最终模型推理阶段，通过将各种增强方法处理过的图像输入模型，并对获得的推理结果取平均值，从而得到预测的道路特征图。

4.4 网络参数设置

训练过程中实验参数设置：每批次输入大小为 8；总迭代次数为 200 次；优化器使用随机优化算法更新网络参数 Adam，学习率设置为 0.01。

4.5 评价指标

在评估道路提取模型的性能时，采用了三个关键指标：召回率 (Recall)、F_1-Score、交并比 (intersection over union, IoU) 以及平均交并比 (mean intersection over union, mIoU)，这些指标共同构成了对模型性能全面而细致的衡量体系。由于道路提取本质上被视为一个二分类问题，模型的预测结果自然地被划分为四个类别：真正例 (true positives, TP)：这些像素点被模型正确地识别为道路部分。既在真实标签中为道路，也在模型的预测结果中被判定为道路。假正例 (false positives, FP)：这些像素点被模型错误地识别为道路，而实际上属于背景区域。这种错误可能导致道路边界的模糊或背景噪声的误识别。假负例 (false negatives, FN)：这些像素点实际上属于道路区域，但模型却错误地将它们预测为背景。这种错误可能导致道路信息的遗漏，影响道路网络的完整性。真负例 (true negatives, TN)：这些像素点被模型正确地识别为背景区域。既在真实标签中为背景，也在模型的预测结果中被判定为背景。基于这四个类别，计算出以下三个评估指标：

1) 召回率：衡量的是模型识别出所有实际正样本 (道路像素) 的能力，即在所有真实的道路像素中，模型成功识别出的比例。

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

2) F1 分数 (F_1)：F1 分数是准确率 (Precision) 和召回率 (Recall) 的调和平均，综合考虑了模型的准确性和召回能力。

$$F_1 = 2\times\frac{Precision\times Recall}{Precision+Recall}, \tag{8}$$

式中：$$Precision = \frac{TP}{TP+FP}. \tag{9}$$

3) 交并比：交并比衡量的是预测区域与实际区域的重叠程度，是预测和真实区域的交集与并集的比率。反映了模型在空间上对道路区域的准确度。

$$IoU = \frac{TP}{TP+FN+FP}. \tag{10}$$

4) 平均交并比：衡量的是预测区域和实际区域交集除以预测区域和实际区域的并集，即模型对每一类预测的结果和真实值的交集与并集的比值，之后求和再计算平均。

$$mIoU = \frac{1}{K+1}\sum_{i=0}^{K}\frac{TP}{TP+FP+FN}. \tag{11}$$

4.6 损失函数

实验的标签只有道路和背景，因此本文采用 Focal 损失函数与 Dice 损失函数相结合的方法。其中 Dice 损失函数对当前的损失除了对当前的像素的预测值有关，还与其他点的值也相关，因此并不受大量背景像素的影像，更倾向于挖掘前景区域。但针对较小道路目标分割情况下，Dice 损失函数训练容易不稳定；Focal 损失函数则基于预测准确度对样本权重进行调整，使得模型更加关注分类错误和分类困难的样本：

$$L_{\text{Focal}} = -\alpha(1-p)^{\lambda}\log p, \tag{12}$$

$$L_{\text{Dice}} = 1-\frac{\sum_{m=1}^{N}p_m g_m}{\sum_{m=1}^{N}p_m^2+\sum_{m=1}^{N}g_m^2}, \tag{13}$$

式中：α为类别权重，用来衡量正负样本不均衡问题；λ为难以判断的样本权重，用来衡量难分样本和易分样本；N 为像素总数；g_m为像素 m 的真实标签值；p_m表示像素 m 的预测值。

为了更有效地应对道路提取任务中的挑战，特别是在处理类别不平衡问题时，考虑设计一种复合型损失函数融合了 Focal 损失函数与 Dice 损失函数的优点：

$$L_{\text{Loss}} = L_{\text{Focal}} + L_{\text{Dice}}. \tag{14}$$

4.7 ResT-UNet 网络实验结果和分析

在本节内容中，通过一系列设计的消融实验来评估和分析所提出的 ResT-UNet 网络在针对 Massachusetts 道路数据集和 DeepGlobe 数据集进行道路分割任务上的性能。这一网络是在经典 U-Net 架构的基础上，通过集成一种增强的残差网络块 (ResNeSt block)、Triplet Attention 注意力机制、Transition 层以及 MSFF 模块等元素进行显著改进而构建的。我们的目标是通过实验验证这些改进如何协同工作，以提升网络在遥感图像道路分割任务中的精度和效率。不同改进实验结果对比见表 1，表 1 详细列出了在逐步引入不同模块至 U-Net 基础架构后，模型在 Massachusetts 道路数据集和 DeepGlobe 数据集上分割性能的对比结果。

表 1 和表 2 分别展示了 8 种不同模块的模型在

Massachusetts 道路数据集和 DeepGlobe 数据集上的性能评估结果，这些模型均基于 U-Net 架构进行了不同程度的改进：ResNeSt block+U-Net 通过强化特征提取提升了分割基础；Att+U-Net 引入三重注意力机制，增强了关键特征聚焦并减少了背景噪声；Transition+U-Net 优化了特征传递过程，保留了更多细节；U-Net+MSFF 则利用多尺度特征提高了复杂道路结构的识别能力。而 ResNeSt block+Att+U-Net 在特征提取和特征关注上双重优化，虽整体准确率 (OA) 略降，但 F_1 和 IoU 显著提升，表明这些改进对复杂场景分割至关重要。ResNeSt block+Att+Transition+U-Net+MSFF (即 ResT-UNet) 展现了最优分割性能，证明了这些措施协同作用下的强大效果。为验证这一方法的有效性，以 Massachusetts 道路数据集为例选取了郊区和城市道路进行了直观的视觉对比，如图 7 和图 8 所示。

面对郊区道路的复杂场景，U-Net 网络虽然能够大致捕捉到道路的整体轮廓，但在区分道路与周围树木、植被等遮挡物时显得力不从心。这导致提取结果中，道路边界模糊，且易受背景噪声干扰。以 U-Net 为基础框架，在其编码器部分深度融合了 ResNeSt 网络。ResNeSt 以其强大的特征提取能力，显著提升了模型对复杂场景的解析度，使得道路与遮挡物之间的界限变得更加清晰，从而提高了道路分割的整体精度。进一步地，在 ResNeSt Block 之后加入了 Triplet Attention，使得模型能够更加精准地聚焦于道路区域，有效抑制了非道路信息的干扰。通过增强对关键道路特征的识别能力，模型在复杂环境中也能准确“看到”并提取出道路，提升了分割的准确性和鲁棒性。当 ResNeSt block、Triplet Attention 与 Transition 层以及 MSFF 模块相结合时，模型的性能达到了新的高度。Transition 层的引入优化了特征图的传递过程，保留了更多对分割有用的细节信息；而 MSFF 则充分利用了不同尺度特征的优势，进一步增强了模型对道路边界和细节的捕捉能力。因此，在 ResNeSt block+Att+Transition+U-Net+MSFF 模型下，道路边界的像素点提取变得更加精准且连续，道路分割的完整性得到了显著提升，同时道路的细节信息也得到了更好的

表 1　不同改进实验结果对比 (Massachusetts)

Table 1　Comparison of experimental results of different improvements (Massachusetts)

Method	OA/%	F_1/%	IoU/%	mIoU/%
U-Net	97.71	87.46	62.37	80.01
ResNeSt block+U-Net	97.79	87.51	62.79	80.36
Att+U-Net	97.76	87.49	62.41	80.15
Transition+U-Net	97.73	87.48	62.39	80.10
U-Net+MSFF	97.85	87.67	62.76	80.59
ResNeSt block+Att+U-Net	97.89	87.83	63.11	80.68
ResNeSt block+Att+Transition+U-Net	97.91	87.91	63.57	80.74
ResNeSt block+Att+Transition+U-Net+MSFF	**98.09**	**88.83**	**64.76**	**81.94**

表 2　不同改进实验结果对比 (DeepGlobe)

Table 2　Comparison of experimental results of different improvements (DeepGlobe)

Method	OA/%	F_1/%	IoU/%	mIoU/%
U-Net	97.07	86.12	61.98	79.82
ResNeSt block+U-Net	97.13	86.27	62.08	80.01
Att+U-Net	97.32	86.41	62.34	80.13
Transition+U-Net	97.11	86.03	61.13	79.69
U-Net+MSFF	97.73	86.91	62.60	80.61
ResNeSt block+Att+U-Net	97.81	87.29	63.07	80.94
ResNeSt block+Att+Transition+U-Net	97.93	87.55	63.41	81.01
ResNeSt block+Att+Transition+U-Net+MSFF	**98.05**	**88.01**	**64.45**	**81.35**

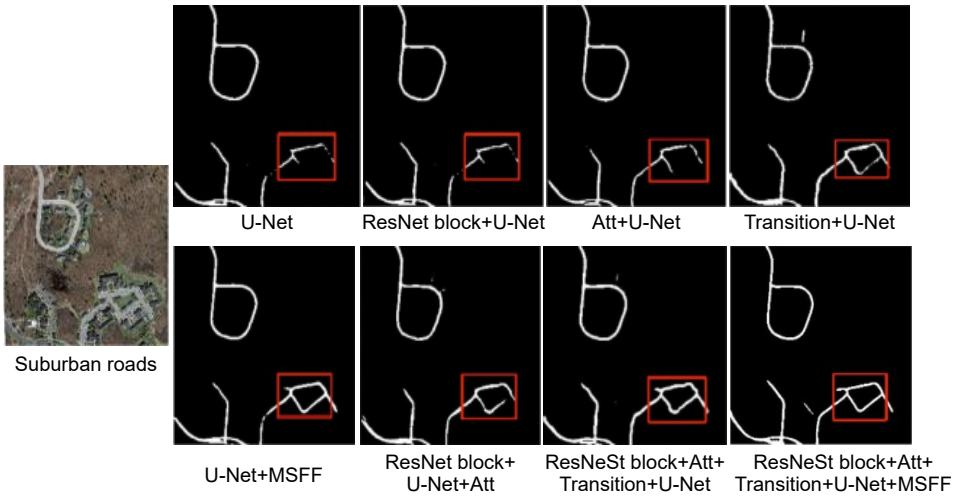


图 7　不同模块对郊区道路提取效果

Fig. 7　Extraction effect of different modules on suburban roads

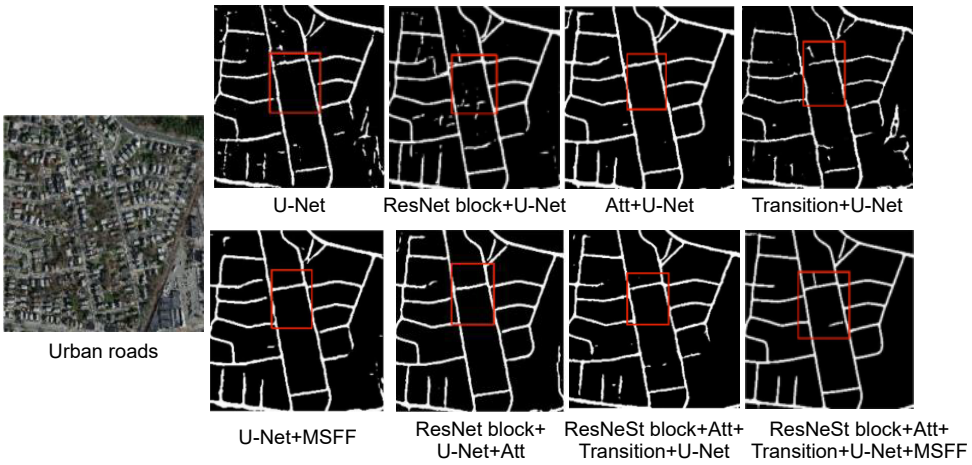


图 8　不同模块对城市道路提取效果

Fig. 8　The extraction effect of different modules on urban roads

保留和优化。

在处理城市道路图像时，U-Net 对于细小的道路目标存在分割不明显的问题，容易遗漏道路像素，导致分割结果不够完整。通过在编码器部分融合 ResNeSt 网络，整体网络的分割精度得到了显著提升。ResNeSt 的强大特征提取能力使得模型能够更好地识别并区分出不同尺寸的道路目标，为后续的精确分割打下了坚实基础。进一步地，ResNeSt block 与 Triplet Attention 融合不仅增强了对大道路目标的细节捕捉能力，还显著提升了对小道路目标的识别精度。这种双重优化使得模型在处理城市道路图像时，无论目标大小，都能实现更加精确的整体和细节分割。最后，通过添加 Transition 层和 MSFF 模块，模型在处理小道路目标时的整体分割效果得到了明显改善。Transition层优化了特征图的传递过程，减少了信息丢失；而 MSFF 则通过多尺度特征融合，进一步增强了模型对道路边界的识别能力。这两者的结合使得分割结果的边界信息更加平滑、自然，提高了道路分割的视觉效果和实用性。

为了全面评估本文提出的 ResT-UNet 网络在道路提取中的性能优势，设计了一系列对比实验，将 ResT-UNet 与当前流行的几种网络架构进行对比，包括经典的 U-Net[11] 和 DeepLabV3[18]，以及近几年的网络 ResUNet[19]、DDUNet[20]、D-LinkNet[13]、U^2-Net[21] 和 MINet[22]。通过对比分析这些网络在相同数据集上的表现，可以直观地验证 ResT-UNet 网络的有效性。实验结果见表 3 和表 4。

由表 3 和表 4 可以看出，本文提出的 ResT-UNet 网络在 OA、F_1 和 IoU 评价指标上较其他方法都获得了相对令人满意的效果。在 Massachusetts 数据集上，

不同网络道路提取的结果呈现出一定的差异。从总体准确率来看，所有网络的准确率都较高，其中对比网络的准确率超过了 97%，表现尤为出色。在 F_1 方面，MINet 和 ResT-UNet 的 F_1 较高，分别为 88.10% 和 88.83%，这表明在精确度和召回率之间取得了较好的平衡。在交并比方面，ResT-UNet 的 loU 最高，为 64.76%，显示出其在道路提取方面的优势。此外，从平均交并比来看，ResT-UNet 的 mloU 也相对较高，为 81.94%，进一步验证了其良好的性能。在 DeepGlobe 数据集上，ResT-UNet 同样展现出了出色的性能。与其他方法相比，ResT-UNet 在 OA、F_1、loU 以及 mloU 等指标上均取得了较高的得分。这进一步验证了 ResT-UNet 在道路提取任务中的有效性和优越性。

为了直观地对比不同模型的道路提取效果，本文选取了部分 DeepLabV3+、U-Net、ResUNet、DDUNet、D-LinkNet、U^2-Net 和 MINet 网络模型的实验结果图，如图 9、图 10 所示。

由图 9 和图 10 中标注的红色框区域可以看出，ResT-UNet 网络和其他网络分割的结果相比，ResT-UNet 网络较为平整，而 DeepLabV3+、U-Net、ResUNet、DDUNet、D-LinkNet、U^2-Net 和 MINet 网络的分割结果较为粗糙。由两个数据集的结果图中可以看出，存在树木遮挡时，本文提出的方法 ResT-UNet 网络较其他网络分割精度明显得到了一定的提升。其中 ResUNet 网络在提取道路方面 F_1 和 IoU 表现最差，丢失程度最高。对比于本文提出的 ResT-UNet 网络模型对道路提取没有出现道路边缘分割不连续、道路断割的情况；以及再树木遮挡的情况下，ResT-UNet 网络较为完整的提取道路信息。

表 5 统计了不同网络模型的训练时间，从表 5 可以看出：不同深度学习网络算法在计算资源 (FLOPs)

表 3　不同网络道路提取结果比较 (Massachusetts)

Table 3　Comparison of road extraction results from different networks (Massachusetts)

Method	OA/%	F_1/%	IoU/%	mIoU/%
U-Net	97.71	87.46	62.37	80.01
DDUNet	97.62	87.13	61.67	79.63
ResUNet	97.72	87.88	62.43	80.16
DeepLabV3+	97.62	87.48	61.62	79.71
D-LinkNet	97.83	87.91	63.11	80.83
U^2-Net	97.65	87.16	61.85	79.65
MINet	97.88	88.10	63.34	80.37
ResT-UNet	**98.09**	**88.83**	**64.76**	**81.94**

表 4　不同网络道路提取结果比较 (DeepGlobe)

Table 4　Comparison of road extraction results from different networks (DeepGlobe)

Method	OA/%	F_1/%	IoU/%	mIoU/%
U-Net	97.07	86.12	61.98	79.82
DDUNet	97.23	86.87	62.18	80.12
ResUNet	97.11	86.28	62.10	79.97
DeepLabV3+	96.13	85.84	61.02	78.94
D-LinkNet	97.52	86.57	63.07	80.76
U^2-Net	97.43	86.18	62.84	80.56
MINet	97.48	86.88	62.71	80.41
ResT-UNet	**98.05**	**88.01**	**64.45**	**81.35**

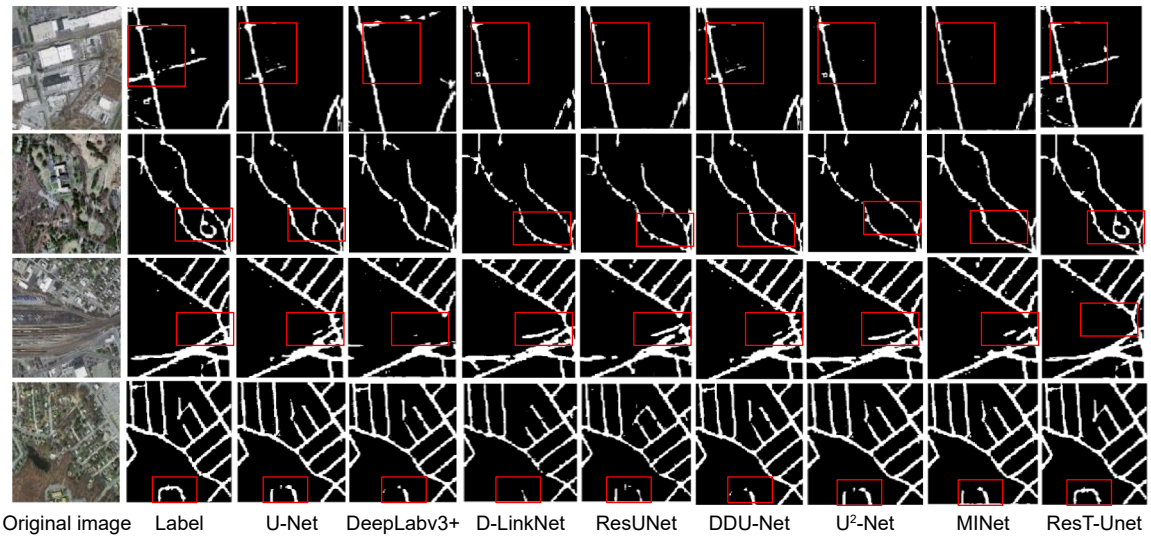


图 9　不同模型的道路提取效果图 (Massachusetts)

Fig. 9　Road extraction results of different models (Massachusetts)

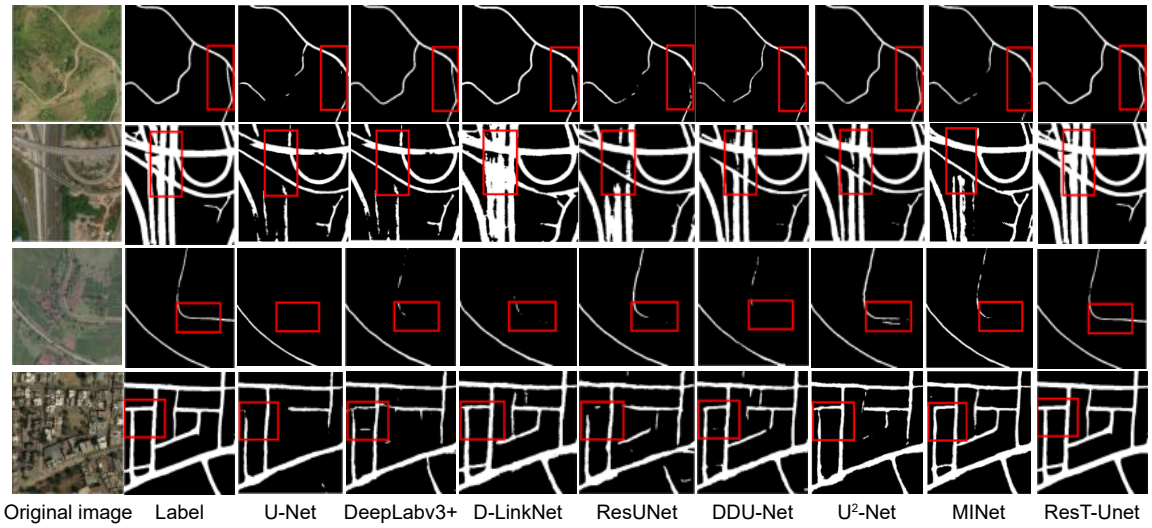


图 10 不同模型的道路提取效果图 (DeepGlobe)

Fig. 10 Road extraction results of different models (DeepGlobe)

表 5 不同算法的复杂度

Table 5 Complexity of different algorithms

Algorithm	U-Net	DeepLabv3+	D-LinkNet	ResUNet	U^2-Net	DDUNet	MINet	ResT-UNet
FLOPs/G	25947.00	26474.89	120408.56	41837.93	157679.07	77489.09	384409.60	54994.39
Params/M	24.73	5.83	213.87	22.60	44.17	65.84	47.56	51.13
Latency/s	3.78	3.49	4.79	3.71	5.00	4.13	6.45	3.57

上的需求存在差异。FLOPs 值反映了算法在执行过程中所需的计算能力，数值越高表示算法计算复杂度越大。在这些算法中，MINet 的 FLOPs 值最高，达到了 384409.6 MFLOPS，说明其计算复杂度相对较高。而 U^2-Net 和 D-LinkNett 的 FLOPs 值分别为 157679.07 MFLOPS 和 120408.56 MFLOPS，也位于较高水平。相比之下，U-Net、DeepLabv3+和 ResUNet 的 FLOPs 值较低，说明计算复杂度相对较低。除了计算复杂度外，表格还列出了各算法的参数数量 (Params) 和延迟时间 (Latency)。参数数量反映了算法模型的规模，而延迟时间则衡量了算法的执行速度。从参数数量上看，D-LinkNet 的参数数量最多，达到了 213.87 M，而 DeepLabv3+的参数数量最少，仅有 5.83 M。在延迟时间方面，各算法的延迟时间相差不大，其中 MINet 的延迟时间最长，为 6.45 s，而 DeepLabv3+和 ResT-UNet 的延迟时间相对较短，分别为 3.49 s 和 3.57 s。ResT-UNet 作为本文所提出的方法，在计算复杂度上表现良好，FLOPs 值位于中等水平，同时参数数量和延迟时间也相对较优。这表明 ResT-UNet 在保持较高计算效率的同时，还能够有效地控制模型规模和执行速度，具有较好的实用性和性能表现。

5 结 论

本文在 U-Net 架构的基础上，提出了结合 ResNeSt 和多尺度特征融合的遥感影像道路提取方法。该网络通过设计的组件——ResNeSt block、Triplet Attention 机制以及 Transition 层，显著优化了编码器的性能。具体而言，ResNeSt block 利用分组加权融合的策略，有效提升了跨通道语义信息的提取能力，使模型能够深入理解图像中的道路特征。Triplet Attention 机制则通过精准聚焦关键信息并忽略无关细节，进一步提高了道路分割的精确度。此外，Transition 层以卷积操作替代传统的最大池化，有效减少了在特征降维过程中道路信息的损失。在编码器与解码器之间的桥接部分，本文引入了多尺度特征融合模块，该模块能够捕获并整合来自不同尺度感受野的信息，从而增强网络对道路细节及边界的捕捉能力。这一设计使得 ResT-UNet 在提取道路信息时更加全面且细致。实验结果表明，本文提出的 ResT-UNet 网络结构具有更好的分割效果，提取的道路信息更加完整。未来，计划进一步拓展和优化道路提取的数据集，以涵盖更多样化的场景和更复杂的道路结构。同时，也将不断探索和尝试新的神经网络架构，以期在道路提取领域实现更高的性能突破。

参考文献

[1] Huang B, Zhao B, Song Y M. Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery[J]. *Remote Sens Environ*, 2018, **214**: 73−86.

[2] Xu Y X, Chen H, Du C, et al. MSACon: mining spatial attention-based contextual information for road extraction[J]. *IEEE Trans Geosci Remote Sens*, 2022, **60**: 5604317.

[3] Xiao Z J, Zhang J H, Lin B H. Feature coordination and fine-grained perception of small targets in remote sensing images[J]. *Opto-Electron Eng*, 2024, **51**(6): 240066.
肖振久, 张杰浩, 林渤翰. 特征协同与细粒度感知的遥感图像小目标检测[J]. 光电工程, 2024, **51**(6): 240066.

[4] Liang L M, Chen K Q, Wang C B, et al. Remote sensing image detection algorithm integrating visual center mechanism and parallel patch perception[J]. *Opto-Electron Eng*, 2024, **51**(7): 240099.
梁礼明, 陈康泉, 王成斌, 等. 融合视觉中心机制和并行补丁感知的遥感图像检测算法[J]. 光电工程, 2024, **51**(7): 240099.

[5] Yuan Q Q, Shen H F, Li T W, et al. Deep learning in environmental remote sensing: achievements and challenges[J]. *Remote Sens Environ*, 2020, **241**: 111716.

[6] Zhu Q Q, Zhang Y A, Wang L Z, et al. A global context-aware and batch-independent network for road extraction from VHR satellite imagery[J]. *ISPRS J Photogramm Remote Sens*, 2021, **175**: 353−365

[7] He D, Shi Q, Liu X P, et al. Generating 2 m fine-scale urban tree cover product over 34 metropolises in China based on deep context-aware sub-pixel mapping network[J]. *Int J Appl Earth Obs Geoinf*, 2022, **106**: 102667.

[8] Lin N, Zhang X Q, Wang L, et al. Road extraction from remote sensing images based on dilated convolutions U-Net[J]. *Sci Surv Mapp*, 2021, **46**(9): 109−114,156.
林娜, 张小青, 王岚, 等. 空洞卷积 U-Net 的遥感影像道路提取方法[J]. 测绘科学, 2021, **46**(9): 109−114,156.

[9] Yang J L, Guo X J, Chen Z H. Road extraction method from remote sensing images based on improved U-Net network[J]. *J Image Graph*, 2021, **26**(12): 3005−3014.
杨佳林, 郭学俊, 陈泽华. 改进 U-Net 型网络的遥感图像道路提取[J]. 中国图象图形学报, 2021, **26**(12): 3005−3014.

[10] Shafiq M, Gu Z Q. Deep residual learning for image recognition: a survey[J]. *Appl Sci*, 2022, **12**(18): 8972.

[11] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation[C]//*Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention*, Munich, Germany, 2015: 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.

[12] Zhang Z X, Liu Q J, Wang Y H. Road extraction by deep residual U-Net[J]. *IEEE Geosci Remote Sens Lett*, 2018, **15**(5): 749−753.

[13] Zhou L C, Zhang C, Wu M. D-LinkNet: LinkNet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction[C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, Salt Lake City, UT, USA, 2018: 182–186. https://doi.org/10.1109/CVPRW.2018.00034.

[14] Zhang Z, Chen Z, Liu C. Road extraction technology from remote sensing images based on LinkNet and feature aggregation module[J]. *China Science and Technology Information*, 2022, **672**(7): 116−119
张正, 陈仲柱, 柳长安. 基于 LinkNet 和特征聚合模块的遥感图像中道路提取技术[J]. 中国科技信息, 2022, **672**(7): 116−119

[15] Gao L P, Wang J Y, Wang Q X, et al. Road extraction using a dual attention dilated-LinkNet based on satellite images and floating vehicle trajectory data[J]. *IEEE J Sel Top Appl Earth Obs Remote Sens*, 2021, **14**: 10428−10438.

[16] Zhang H, Wu C R, Zhang Z Y, et al. ResNeSt: split-attention networks[C]//*Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, New Orleans, LA, USA, 2022: 2735–2745. https://doi.org/10.1109/CVPRW56347.2022.00309.

[17] Cui Y, Yu Z K, Han J C, et al. Dual-triple attention network for hyperspectral image classification using limited training samples[J]. *IEEE Geosci Remote Sens Lett*, 2022, **19**: 5504705.

[18] Chen L C, Zhu Y K, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//*Proceedings of the 15th European Conference on Computer Vision*, Munich, Germany, 2018: 833–851. https://doi.org/10.1007/978-3-030-01234-2_49.

[19] Diakogiannis F I, Waldner F, Caccetta P, et al. ResUNet-a: a deep learning framework for semantic segmentation of remotely sensed data[J]. *ISPRS J Photogramm Remote Sens*, 2020, **162**: 94−114.

[20] Wang Y, Peng Y X, Li W, et al. DDU-Net: dual-decoder-U-Net for road extraction using high-resolution remote sensing images[J]. *IEEE Trans Geosci Remote Sens*, 2022, **60**: 4412612.

[21] Qin X B, Zhang Z C, Huang C Y, et al. U^2-Net: going deeper with nested U-structure for salient object detection[J]. *Pattern Recognit*, 2020, **106**: 107404.

[22] Pang Y W, Zhao X Q, Zhang L H, et al. Multi-scale interactive network for salient object detection[C]//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, 2020: 9410–9419. https://doi.org/10.1109/CVPR42600.2020.00943.

作者简介

【通信作者】郝明 (2000-)，女，硕士研究生，主要研究方向为图像与视觉信息计算。

E-mail：haoming1232023@163.com

徐婷婷 (1997-)，女，硕士，主要研究方向为知识发现与智能信息处理。

E-mail：2842312826@qq.com

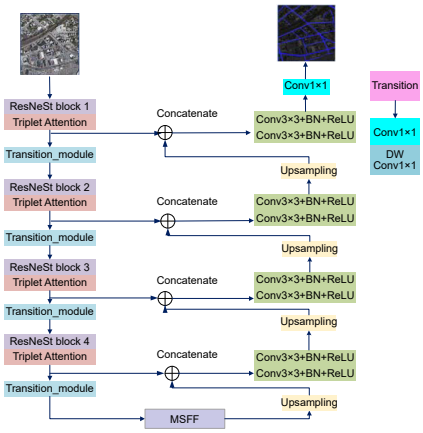
白鹤 (1980-)，女，硕士，副教授，主要研究方向为大数据技术驱动应急管理需求。

E-mail：29262488@qq.com

扫描二维码，获取PDF全文

Remote sensing image road extraction by integrating ResNeSt and multi-scale feature fusion

Hao Ming*, Bai He, Xu Tingting



Remote sensing image road extraction by integrating resnest and multi-scale feature fusion

Overview: Road extraction from high-resolution remote sensing imagery is critical for applications like urban planning, autonomous driving, and road network updates. However, challenges such as discontinuous road edges, low accuracy in small road feature segmentation, and misclassification remain. This paper proposes ResT-UNet, a novel method that integrates the ResNeSt network and multi-scale feature fusion to address these challenges and improve road extraction accuracy. The main objective of this study is to enhance road extraction performance by improving feature extraction and preserving road details. The ResT-UNet architecture builds upon the U-Net model, which is widely used in semantic segmentation. The first modification replaces U-Net's initial convolution layer with a ResNeSt block, which enhances feature extraction and ensures smoother road edge segmentation. Additionally, a triplet attention mechanism is introduced in the encoder to suppress irrelevant features and focus on key road-related information, improving the capture of fine road details by strengthening spatial and channel relationships. Furthermore, ResT-UNet replaces max pooling with convolutional blocks to retain more spatial information, reducing road feature loss. A multi-scale feature fusion (MSFF) module is added between the encoder and decoder, enabling the network to capture long-range dependencies and multi-scale features. This fusion of features from different scales improves road detection in complex environments. The method was evaluated on the Massachusetts Roads and DeepGlobe datasets. Experimental results showed that ResT-UNet outperformed the MINet model, achieving intersection over union (IoU) scores of 64.76% and 64.45%, respectively, representing improvements of 1.42% and 1.74%. These results confirm that ResT-UNet significantly enhances road extraction accuracy, especially in handling complex road boundaries and small-scale features. In conclusion, ResT-UNet offers an effective solution for road extraction from remote sensing imagery, with improved segmentation accuracy. The integration of the ResNeSt block, triplet attention, and multi-scale feature fusion significantly enhances road detection, making the model suitable for applications in autonomous driving, urban planning, and geographic information systems. Future work will focus on further optimization and application to more complex datasets.

Hao M, Bai H, Xu T T. Remote sensing image road extraction by integrating ResNeSt and multi-scale feature fusion[J]. *Opto-Electron Eng*, 2025, **52**(1): 240236; DOI: 10.12086/oee.2025.240236

Foundation item: Basic Scientific Research Project of the Education Department of Liaoning Province (JYTMS20230965)
School of Information Engineering, Liaoning Institute of Science and Technology, Jinzhou, Liaoning 121000, China
* E-mail: haoming1232023@163.com