结合注意力与上下文融合的遥感图像道路提取

李云红，王 梅，苏雪平，李丽敏，张富星，郝特吉

（西安工程大学 电子信息学院，西安 710048）

摘 要：针对遥感图像地物复杂，道路存在细长、连续分布且易受遮挡的问题，提出了一种结合注意力与上下文融合的遥感图像道路提取模型（ACFD-LinkNet）。该模型以D-LinkNet网络为基础，首先在D-LinkNet网络编码器最后卷积层后采用条带注意力模块增强不同尺度道路的特征提取能力，更好地捕捉道路的全局特征，捕获道路的长距离信息；其次，提出了一种上下文融合模块（CFM），并添加至网络编解码的特征传递部分预测相邻像素之间的道路连接，融合上下文不同层级之间的道路信息，解决障碍物遮挡干扰道路连接的问题；最后，对改进模型的交叉熵损失函数和Dice损失函数设置多损失函数超参数权重分配，解决数据集正负样本不均的问题，通过调整权重比值获取最佳分割精度。在DeepGlobe和CHN6-CUG数据集上进行实验，综合指标F_1值分别达到86.76%、92.12%，相比D-LinkNet模型分别提高了3.96%、1.13%。此外，相较于Unet、Deeplabv3+、A^2-FPN等网络，本文模型有最优的性能表现。

关键词：图像处理；遥感图像；道路提取；注意力机制；上下文特征融合；超参数权重分配

中图分类号：TP751 **文献标志码**：A **文章编号**：1671-5497(2025)12-4034-11

DOI：10.13229/j.cnki.jdxbgxb.20240442

Road extraction from remote sensing images combining attention and context fusion

LI Yun-hong，WANG Mei，SU Xue-ping，LI Li-min，ZHANG Fu-xing，HAO Te-ji

（*School of Electronics and Information*，*Xi'an Polytechnic University*，*Xi'an*710048，*China*）

Abstract：Aiming at the complexity of features in remote sensing images and the existence of an elongated and continuous distribution of roads that are easy to obscure，a Road Extraction Model for Remote Sensing Images Combining Attention and Context Fusion（ACFD-LinkNet）was proposed. The network is based on the D-LinkNet network. Firstly，a strip attention module was used in the codec part of the D-LinkNet network to enhance the feature extraction capability of roads at different scales，to better capture the global features of the roads，and to capture the long-distance information of the roads. Secondly，a Context Fusion Module（CFM）was proposed and added to the feature delivery part of the network codec to predict road connections between neighboring pixels，fusing road information between different layers of the context to solve the problem of obstacle obstruction interfering with road connections. Finally，the cross-

收稿日期：2024-04-24.

基金项目：国家自然科学基金项目（62203344）；陕西省自然科学基础研究计划重点项目（2022JZ-35）；西安市科技局“科学家+工程师”队伍建设项目（25KGYB00029）.

作者简介：李云红（1974-），女，教授，博士. 研究方向：图像处理，信号与信息处理技术.
E-mail：hitliyunhong@163.com

entropy loss function and Dice loss function of the improved model were set up with multiple loss function hyperparameter weight assignments to solve the dataset positive and negative sample inhomogeneity, and the optimal segmentation accuracy was obtained by adjusting the weight ratios. Experiments on the DeepGlobe and CHN6-CUG datasets resulted in F_1 values of 86.76% and 92.12% for the composite metrics, respectively, which is an improvement of 3.96% and 1.13% compared to the D-LinkNet model, in addition to optimal performance compared to semantic segmentation methods such as Unet, Deeplabv3+, A^2-FPN, etc.

Key words: image processing; remote sensing image; road extraction; attention mechanism; context feature fusion; hyperparameter weight allocation

0　引　言

遥感图像道路提取是遥感图像处理领域的关键任务，道路信息在城市规划[1]、灾害管理[2]和交通导航[3]等领域扮演着重要角色。然而，由于遥感图像通常包括多种地物和复杂的背景信息，道路跨度大且易被树木、建筑物等障碍物遮挡，因此，从高分辨率遥感图像中提取道路仍存在巨大挑战[4]。

近些年，随着计算机硬件和算力的不断提升，深度学习技术开始广泛应用于图像分割[5-7]任务。2015 年，全卷积神经网络[8](Fully convolutional networks，FCN)提出应用于处理像素级别的语义分割任务，极大地推动了语义分割技术的快速发展。尽管 FCN 使用卷积替代全连接，但是仍然独立处理像素，并未充分考虑像素间的关系。因此，提出了编-解码模型，该类模型有效解决了低分辨率图像处理中的空间信息丢失问题，恢复了像素空间信息。Wang 等[9]在 Unet[10]的基础上设计内部卷积网络，增强对道路拓扑结构和线性特征的学习，提出方向条件随机场，通过将道路方向添加到能量函数中以提高道路提取质量。为了保留更多的道路空间信息，Zhou 等[11]在 LinkNet[12]基础上提出了 D-LinkNet，通过在 LinkNet 模型中添加空洞卷积层扩大感受野，促进多尺度特征融合，但在障碍物遮挡下的道路提取效果不佳。针对这一问题，Wu 等[13]在 D-LinkNet 基础上，在中心部分引入坐标注意力模块来增强特征表示。同时，为了更好地整合不同分支的特征，使用注意力特征融合模块来取代线性特征融合操作，改善了细节提取。Kampffmeyer 等[14]提出了一种方向感知块，用于预测每个像素与相邻像素的连通性。Maji 等[15]提出了一种带引导解码器的深度学习生成器，通过加权引导损失增强解码层预测能力，提升输出精度。Dai 等[16]提出了一个道路增强可变形注意力网络 RADANet，用于学习特定道路像素的长程依赖关系。但在处理农村地区复杂狭窄的道路时，易出现细节丢失问题，进而影响提取性能。

针对以上算法在长跨度道路，以及由于障碍物遮挡导致道路提取不全的问题，本文提出了结合注意力与上下文融合的网络模型 ACFD-LinkNet。在特征提取阶段引入条带注意力模块(Strip attention module，SAM)[17]抑制非道路区域的干扰信息，减少噪声对道路提取结果的影响；同时，为了提高预测遮挡情况下道路的连通性，增强道路目标浅层和深层语义信息的传递，提出将上下文融合模块 CFM 添加至跳跃连接部分；再通过多损失函数加权调整交叉熵损失函数与 Dice 损失函数的权值比例调整道路目标像素较少的问题，实现最佳的分割性能。最后，在 DeepGlobe 和 CHN6-CUG 数据集进行实验，并开展对比实验，验证了本文算法的有效性和优越性。

1　模型概述

本文以 D-LinkNet 模型为主体框架构建道路提取模型 ACFD-LinkNet。网络整体结构由编码区、级联空洞卷积和解码区组成。首先，将 512×512 大小的图像作为网络的输入，编码区域由一个核大小为 7×7、步长为 2 的卷积和 4 个 Resnet 残差模块组成。预训练网络采用 Resnet34，经过 4 次下采样提取特征。中心部分采用级联的空洞卷积和跳跃连接，以增大感受野和实现细粒度特征融合。在网络编码部分和特征传递处分别嵌入条带注意力模块和上下文融合模块关注全局细节特征信息增强道路连接。解码区采用残差结构通

过1×1的卷积核降低计算复杂度。同时,利用转置卷积将图像恢复到原始尺寸,ACFD-LinkNet网络结构如图1所示。

1.1 条带注意力模块

道路通常呈现跨度大、条状连续分布,不同地区的道路网分布差异较大。传统注意力机制仅关注局部特征,无法适应道路跨度变化。因此,本文引入条带注意力模块关注水平和垂直方向的像素分布,捕捉细粒度的局部特征。同时,保持全局语义信息,提升道路分割的准确性。

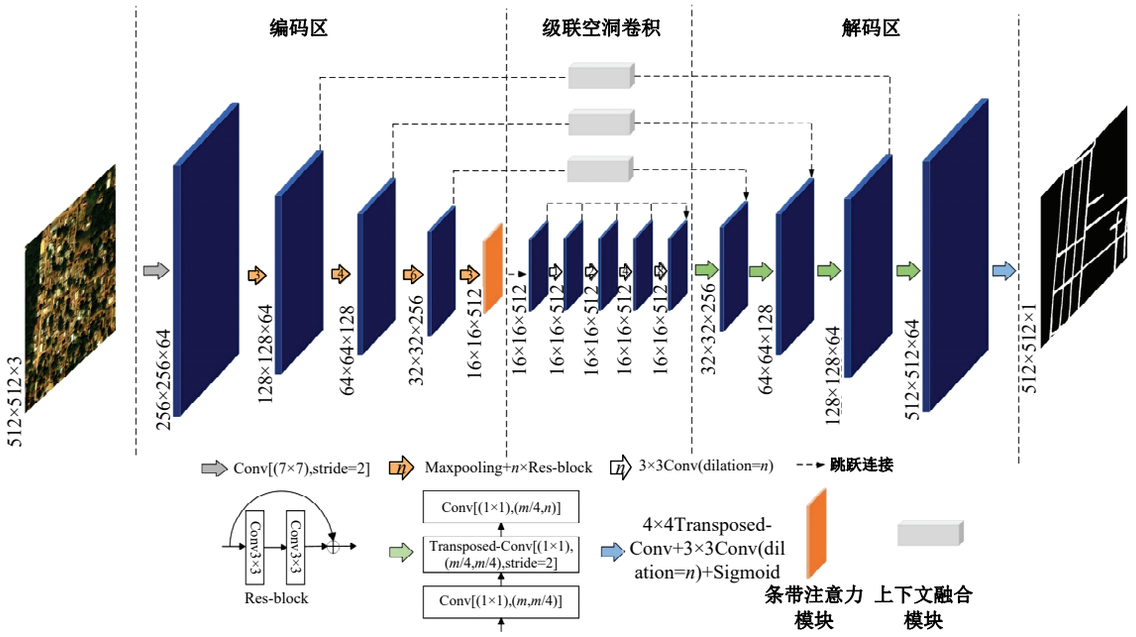

图1 ACFD-LinkNet网络结构

Fig. 1 ACFD-LinkNet network structure

条带注意力模块结构图如图2所示。通过在垂直方向部署长条形池化核编码全局信息,并在水平方向搜集像素与条带池化核的关联程度,每个像素可与不同列空间中的像素关联,获取更多上下文信息,有效解决空洞卷积的网格效应。给定特征图输入输出的大小分别为$F\in\mathbb{R}^{C\times H\times W}$和$F'\in\mathbb{R}^{C\times H\times W}$,特征图$F$分别经过1×1卷积后生成3个新的特征图$Q$、$K$、$V$。注意力图$A$是通过计算$Q^{\mathrm{T}}$和$K$之间的亲和性运算搜集水平方向上每个像素与其他像素之间的相关程度得到的,然后经过soft max函数归一化以获得注意力权重,特征图A定义如下所示。

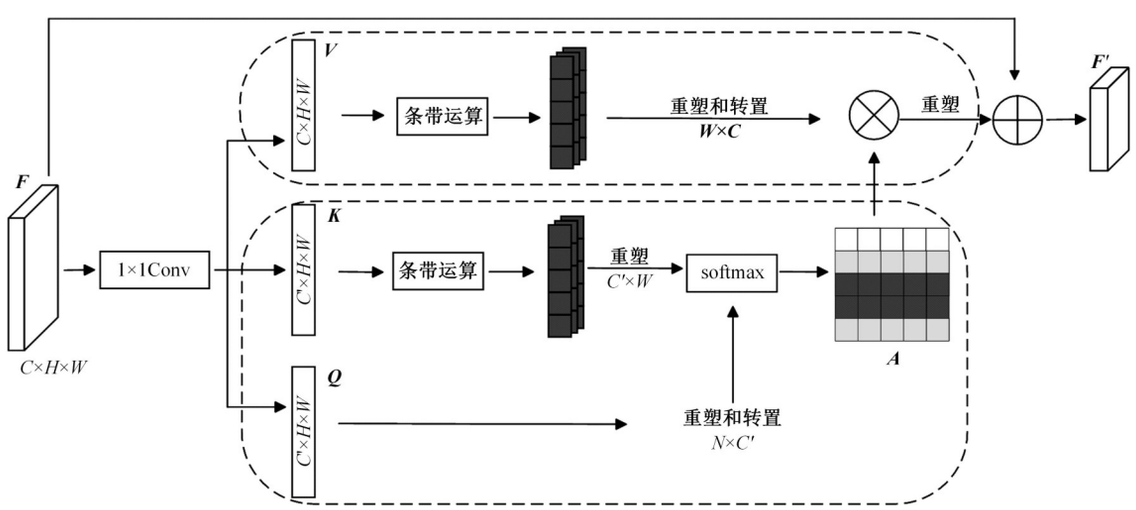

图2 条带注意力模块

Fig. 2 Strip attention module

$$A=s(R_Q(\mathrm{Conv}(F))^{\mathrm{T}}\times R_K(S_p(\mathrm{Conv}(F)))) \tag{1}$$

式中:$R_K(\cdot)$表示将特征图$\boldsymbol{K}$重塑为$\mathbb{R}^{C'\times W}$,$R_Q(\cdot)$表示将特征图$\boldsymbol{Q}$重塑为$\mathbb{R}^{C'\times N}$,$N=H\times W$;$S_p(\cdot)$表示条带运算;Conv(·)表示1×1卷积;T表示转置,$s(\cdot)$表示softmax。

另外,对特征图$\boldsymbol{V}$进行类似于$\boldsymbol{K}$的操作后与A加权融合,以获得更准确的特征表示。最后,对F进行逐元素求和得到最终的$F'\in\mathbb{R}^{C\times H\times W}$,$F'$定义如下所示:

$$F'=A\times R_V(R_K(S_p(\mathrm{Conv}(F))))^{\mathrm{T}}+F \tag{2}$$

式中:$R_V(\cdot)$表示将特征图$\boldsymbol{V}$重塑为$\mathbb{R}^{C\times H\times W}$。

将上下文信息添加到F中,以增强逐像素表示,从而捕捉条状结构的依赖程度。

1.2　上下文融合模块

遥感图像背景复杂,部分道路易被建筑物、树木阴影等遮挡,导致道路连通性断裂,给道路提取任务带来较大挑战。因此,本文提出了上下文融合模块,用于预测给定像素与相邻像素的连接性,模块结构如图3所示。该模块由卷积模块、高效多尺度注意力(Efficient multiscale attention, EMA)模块[18]及连接立方体构成。上下文融合模块最终输出的是$H\times W\times C$连接立方体来学习相邻像素间的连接关系。

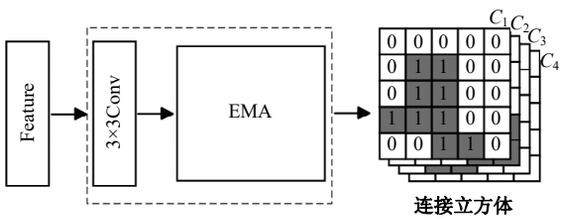

图3　上下文融合模块

Fig. 3　Context fusion module

EMA模块如图4所示。输入特征首先经过一个3×3卷积捕获像素周围的空间信息。随后进入EMA模块进行特征优化,具体流程如下:①将给定的特征映射沿通道维度方向划分为N个子特征映射,以学习不同的特征信息。②设计3条平行路径提取分组特征图的注意力权重,包含2条1×1卷积分支和1条3×3卷积分支。其中,1×1分支中利用二维全局平均池化对信道进行编码后,得到两个张量,再使用Sigmoid函数处理以调整编码权重获得精确的空间信息,输出像素属于道路或非道路的值。3×3分支使用3×3卷积捕获多尺度特征。③引入跨空间学习(Cross-spatial learning)机制,通过跨维交互间实现信息聚合,对1×1与3×3分支输出的全局空间特征依次进行二维全局平均池化、特征重构与Softmax归一化操作,并逐元素相乘,最终生成融合完整空间位置信息的输出特征图。

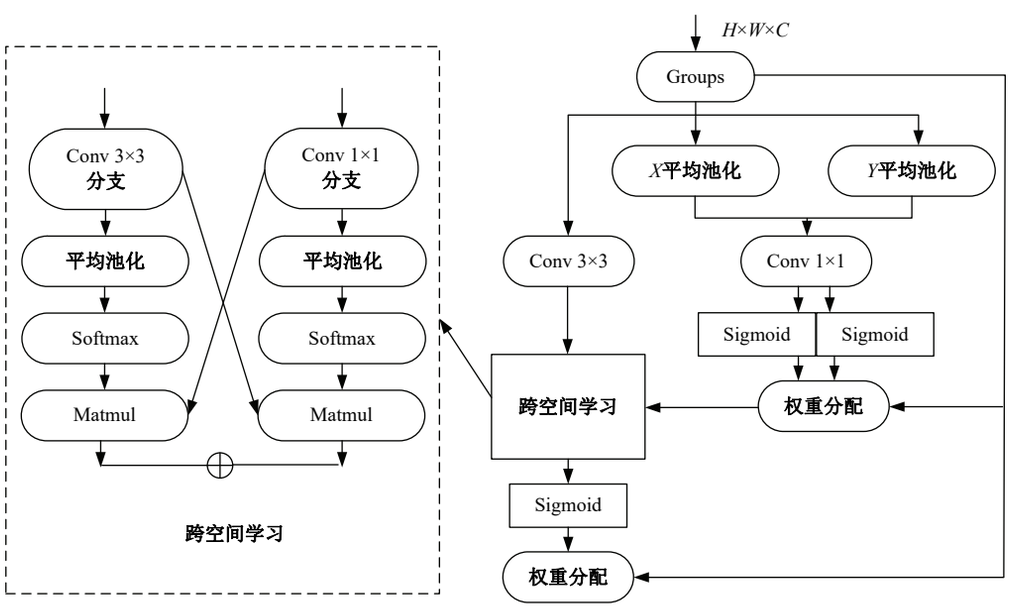

图4　高效多尺度注意力模块

Fig. 4　Efficient multi-scale attention

为了预测道路的连接关系,使用道路二值分割掩码生成连接立方体。如图5所示,给定像素P_i,采用4连通[14]计算P_i与上下左右4个方向相邻像素$c_1\sim c_4$的连接,P_{ijc_i}表示给定像素与相邻像素的连接性,i,j表示像素空间位置,c_i表示相邻像素。其中,每个像素位置上的数值表示该像素所

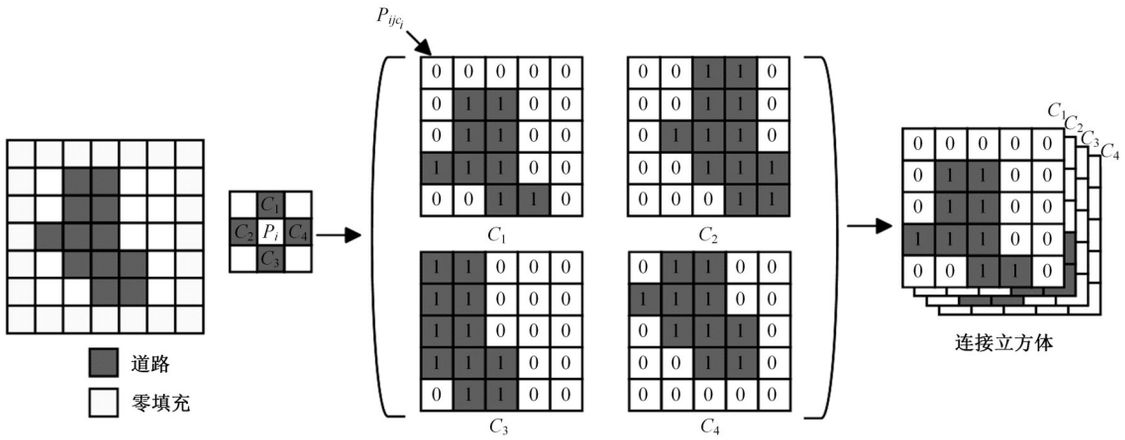

图5 连接立方体

Fig. 5 Connection cube

属区域,0表示不连接,1表示像素与相邻像素连接。通过堆叠c_1~c_4的二进制分割掩码生成$H\times W\times C$连接立方体P,其中,H和W分别为输入图像的高度和宽度,C为给定像素与相邻像素的数量,本文$C=4$。

1.3 多损失函数权重分配

交叉熵损失函数常用于分类任务,但在遥感图像中非道路像素远多于道路像素,导致样本不平衡和过度预测背景。Dice损失函数能捕捉道路细节和边界,但因对噪声敏感,故训练细粒度目标时不稳定。为解决以上问题,本文对交叉熵和Dice损失函数的比例设置了权重分配,使梯度大的Dice损失函数获得更高权重。同时,利用梯度较小的交叉熵损失保持模型稳定性,提高收敛速度和精度。交叉熵损失函数$\mathrm{BCE_{Loss}}$和Dice损失函数$\mathrm{Dice_{Loss}}$的计算公式分别如下所示:

$$\mathrm{BCE_{Loss}}=-\frac{1}{N}0(y_i\cdot\log(p_i)+(1-y_i)\cdot\log(1-p_i)) \tag{3}$$

$$\mathrm{Dice_{Loss}}=1-\frac{2\cdot\sum_{i=1}^{N}p_i\cdot y_i}{\sum_{i=1}^{N}{p_i}^2+\sum_{i=1}^{N}{y_i}^2} \tag{4}$$

$$\mathrm{Loss}=\alpha\mathrm{BCE_{Loss}}+\beta\mathrm{Dice_{Loss}} \tag{5}$$

式中:N为总像素数量;y_i为像素i的真实标签;p_i为像素i的预测概率值;α、β分别为交叉熵损失函数和Dice损失函数所占权重超参数,通过调节两者权重可获得最佳分割性能。

2 实验及结果分析

2.1 数据集介绍

本文选取DeepGlobe数据集和CHN6-CUG数据集对模型进行评估验证。

DeepGlobe数据集来自2018年"DeepGlobe道路提取挑战赛"。该数据集包括6 226张分辨率为0.5 m、大小为1 024×1 024像素的图像,1 243张验证图像,以及1 101张测试图像。图像覆盖了泰国、印度和印度尼西亚,包括水泥、沥青和山区的道路。本次实验将训练集图像剪裁成512×512大小并重新划分,选取其中5 000张、1 000张分别作为训练集和验证集。

CHN6-CUG数据集是首个中国代表性城市的新型大规模卫星影像数据集,选取了中国6个具有代表性的城市。数据集包含4 511张大小为512×512的像素的图像,分辨率为0.5 m,3 608张用于模型训练,903张用于测试以及结果评估。本次实验选取3 608张、750张分别作为训练集和验证集。

2.2 实验参数设置

2.2.1 实验环境

实验环境采用的GPU为NVIDIA GeForce RTX 3090,显存为24 GB,利用Python3.8、Pytorch1.9.0深度学习框架搭建网络模型。学习率设为2×10^{-4},批量大小设为16,epoch设为200,使用Adam优化器更新权重,训练过程采用学习率衰减方法优化模型性能。

2.2.2　评价指标

实验采用精确度（Precision）、召回率（Recall）、F_1分数（F_1-score）和平均交并比（Mean intersection over union，mIoU）4个指标进行模型有效性评估。精确度表示预测为道路的像素中真正的道路像素比例；召回率表示模型正确预测的真实道路像素比例；平均交并比衡量预测和真实道路区域的重叠程度；F_1分数是精确度和召回率的调和平均数，能反映模型的总体性能。

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

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

$$F_1\text{-score}=\frac{2\times \text{Precision}\times \text{Recall}}{\text{Precision}\times \text{Recall}} \tag{8}$$

$$\text{mIoU}=\frac{1}{k+1}\sum_{i=0}^{k}\frac{\text{TP}}{\text{FN}+\text{FP}+\text{TP}} \tag{9}$$

式中：$k+1$表示类别总数包括背景；TP（True positive）为模型正确预测为道路像素的样本数；FP（False positive）为模型错误预测为道路像素的样本数；FN（False negative）为实际为道路像素但被模型错误地预测为非道路像素的样本数；TN（True negative）为被正确预测为非道路的像素。

2.3　实验结果分析

2.3.1　权重分配

在改进模型上对交叉熵损失函数和Dice损失函数设置权重比，以优化分类和分割之间的权重平衡。将式（5）中的权重α设置先验值为1，通过调节β的取值分析网络性能，进而得到最优的提取精度。设置1∶1、1∶2、1∶3、1∶4、1∶5的比例进行实验，模型性能评价取mIoU指标。权重比性能对比量化结果如表1所示，当α∶β=1∶4时，两个数据集上的mIoU精度最高。图6为选取两种数据集的可视化提取结果对比，第1行图像选自DeepGlobe数据集，第2行图像选自CHN6-CUG数据集，当权重比在1∶4时道路提取结果最为完整。

表1　损失函数不同权重比性能对比

Table 1　Performance comparison of loss functions with different weight ratios　%

α∶β	mIoU	
	DeepGlobe数据集	CHN6-CUG数据集
1∶1	79.32	74.94
1∶2	79.38	75.40
1∶3	79.52	76.66
1∶4	79.68	77.90
1∶5	79.39	76.49

2.3.2　对比实验

为验证ACFD-LinkNet模型的有效性，将其与经典分割网络，如Unet[10]、FCN[8]、Deeplabv3+[19]、D-LinkNet[11]，以及最近的模型SGCN[20]、A^2-FPN[21]、EGE-UNet[22]和VM-UNet[23]进行了定性和定量对比分析。Unet和FCN是经典的分割模型代表。Deeplabv3+以Xception架构作为骨干网络，并引入特征金字塔网络。SGCN通过对道路特征的信道和空间特征进行提取，并结合图卷积网络捕获全局背景道路信息。A^2-FPN结合注意力聚合模块和特征金字塔网络进行道路分割。EGE-UNet和VM-UNet是用于医学图像分割的新方法，在边缘保留和细节恢复方面表现出色，适用于道路提取任务。图7、图8为选取的部分图像道路提取结果，虚线框为道路提取差异部分。

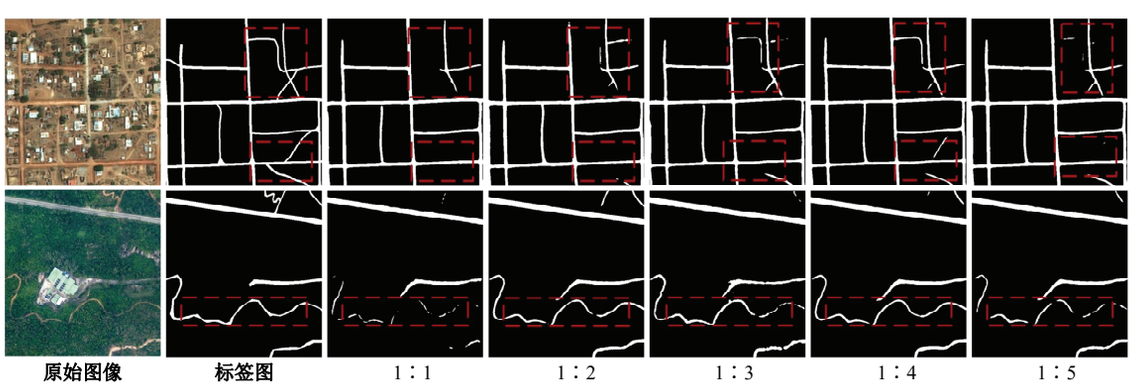

图6　损失函数不同权重比道路提取结果

Fig. 6　Loss function with different weight ratio road extraction results

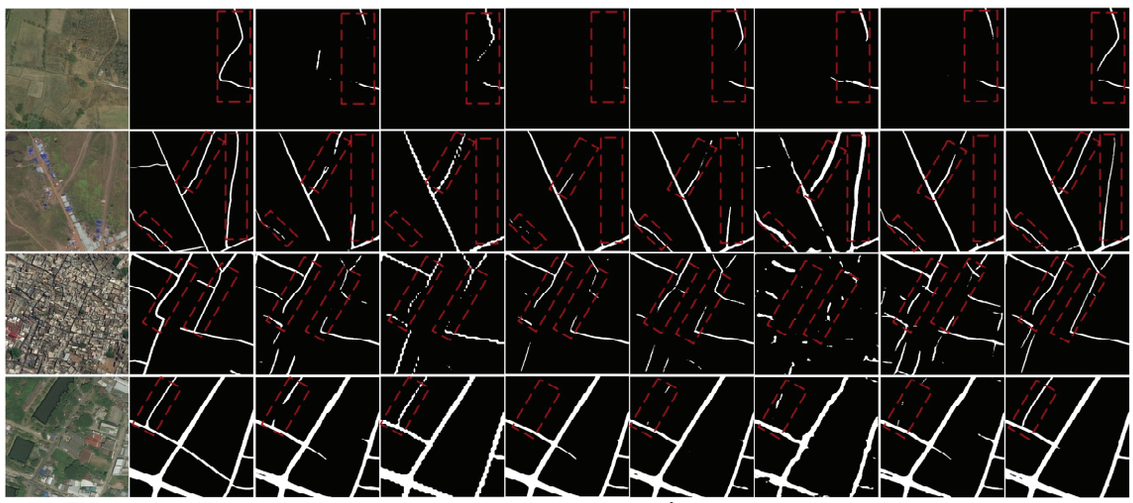

图7 ACFD-LinkNet与对比网络在DeepGlobe数据集提取结果

Fig. 7 ACFD-LinkNet with comparison network in DeepGlobe dataset extraction results

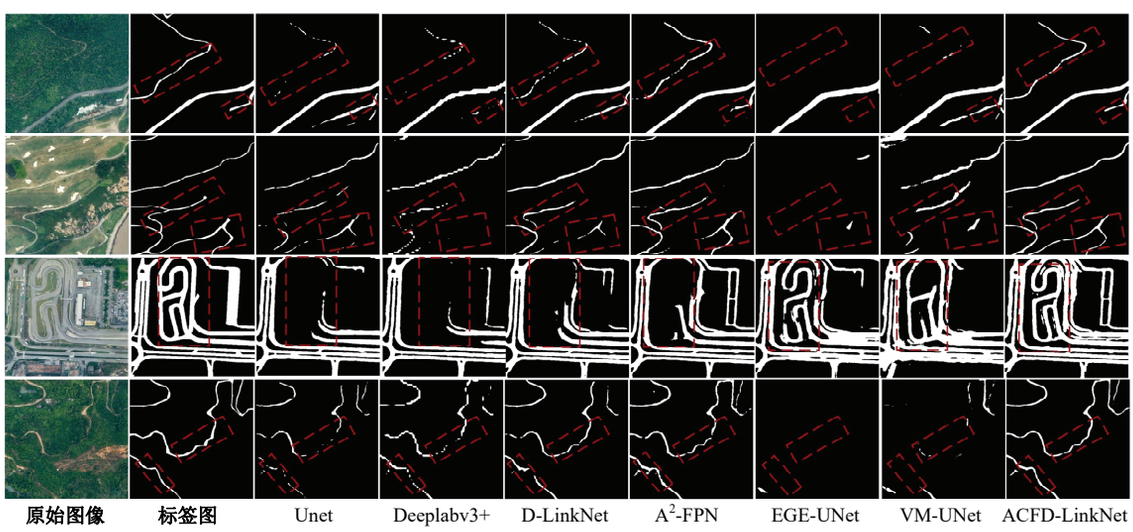

图8 ACFD-LinkNet与对比网络在CHN6-CUG数据集提取结果

Fig. 8 ACFD-LinkNet with comparison network in CHN6-CUG dataset extraction results

(1)DeepGlobe数据集定性定量评估

DeepGlobe数据集中裸地荒地较多,与道路颜色相似,并存在被建筑物植被遮挡的情况。从图7可以看出,在第1、2行原始图像中,道路与背景光谱特征较为相似,且整体场景以连续的长道路为主。相较于其他对比网络,ACFD-LinkNet通过条带注意力机制优化相关特征的提取,同时抑制无关特征的影响,能够在相同环境下较为完整地提取到道路目标。而Unet、Deeplabv3+和D-LinkNet漏提取较多,道路断裂现象明显,提取效果不佳。这是因为,U-Net的跳跃连接会引入过多局部细节,在与高层语义信息融合时导致道路边界模糊或误分割,从而影响道路提取的准确性和连续性。Deeplabv3+和D-LinkNet模型使用空洞卷积导致感受野内大量像素未被有效利用,进而引发"空洞效应",破坏数据的连续性特征。第3、4行图像为复杂城市遮挡场景,图中道路受建筑物和树荫遮挡较多,可以看出,EGE-UNet和VM-UNet提取的道路破碎且不连续,VM-UNet采用非对称的编-解码结构,会导致在解码过程中信息丢失或重建不准确。需要注意的是,由于医学图像通常关注细微的病理信息,而道路提取则更侧重于几何形状、纹理和上下文关系,二者任务特性存在明显差异,这也是EGE-UNet和VM-UNet在此类道路提取任务中表现不佳的重要原因。相比之下,ACFD-LinkNet提取结果与标签重合度最高,通过上下文融合模块能有效预测像素点相邻关系,很好地保持了道路的连通性。

表2为DeepGlobe数据集上各模型定量评估结果，其中，加粗数字表示该指标的最佳值，后同。从中可以看出，ACFD-LinkNet的Precision最高为95.79%，相较于Unet和SGCN，ACFD-LinkNet的Precision分别提升了1.16%和0.93%，说明ACFD-LinkNet预测的假阳性像素更少。从mIoU看，D-LinkNet和A²-FPN整体表现较好，分别达到了78.23%和78.21%。F_1能全面评价网络模型性能，综合结果显示，ACFD-LinkNet的F_1最高为86.76%，相较于次优的Unet提升了1.65%；SGCN的F_1最差，为68.31%，ACFD-LinkNet较之提升了18.45%。由此可见，ACFD-LinkNet各项指标均高于其他对比模型，能清晰识别道路，并区分各种干扰因素。

表2 在DeepGlobe数据集道路提取结果

Table 2　Road extraction results in DeepGlobe dataset

%

方法	DeepGlobe数据集			
	Precision	Recall	F_1-score	mIoU
Unet[8]	94.63	77.33	85.11	75.82
FCN[6]	82.44	77.87	80.09	75.11
Deeplabv3+[17]	90.94	76.90	83.33	75.84
D-LinkNet[9]	94.11	73.93	82.80	78.23
SGCN[18]	94.86	53.38	68.31	74.67
A²-FPN[19]	92.40	78.38	84.81	78.21
EGE-UNet[20]	76.93	69.44	72.99	65.59
VM-UNet[21]	78.10	71.55	74.68	75.28
ACFD-LinkNet	**95.79**	**79.29**	**86.76**	**79.68**

(2)CHN6-CUG数据集定性定量评估

为进一步验证ACFD-LinkNet模型的有效性，在CHN6-CUG数据集上进行验证。CHN6-CUG数据集包含铁路、公路、城市和农村道路等类型，道路周围树木高大密集遮挡严重，形状结构差异大，为提取完整性带来巨大挑战。从1、2、4行图像可以看出，尽管对比模型在一定程度上能够捕捉到道路信息，但与本文模型相比，其在道路提取过程中仍表现出较为明显的错误识别和遗漏现象，无法全面且准确提取出完整的道路信息。其中，EGE-UNet在树木遮挡区域漏提取最为严重，VM-UNet相比于EGE-UNet虽然能识别出部分细小道路，但是边缘部分的预测不够精细，其余对比网络提取的道路均存在不同程度的断裂和不连通；在第3行图像中，道路颜色与背景相似，边界模糊难以区分，A²-FPN提取结果仅次于ACFD-LinkNet，A²-FPN通过注意力引导的特征聚合增强多尺度特征学习，有助于提高对目标的准确性和鲁棒性。而EGE-UNet和VM-UNet在道路与背景有相似光谱特征的场景下提取结果相对较为完整，表明EGE-UNet和VM-UNet适合区分具有相似颜色或纹理结构的道路。经过综合分析，本文在解决图像中道路遮挡与连通性挑战时，通过融合注意力与上下文特征，显著提高模型在保留道路完整性和细节信息方面的能力，增强模型在处理复杂场景时的局限性，极大地提升了道路分割视觉效果，能更好地呈现道路分割结果。

表3为CHN6-CUG数据集上各模型定量评估结果。从表3中可以看出，ACFD-LinkNet模型性能最佳，各项指标分别达到了92.22%、92.02%、92.12%、77.90%，与Precision第二的Deeplabv3+相比，本文模型的Recall和F_1分数分别提升了1.67%和1.46%。F_1分数综合指标比相对较高的FCN和D-LinkNet模型分别提升了1.22%和1.13%，同时各项指标远超EGE-UNet，其中mIoU比其高出16.99%。实验结果表明，ACFD-LinkNet模型相比其他对比网络具有显著优势，能全面捕获道路长距离特征，准确提取道路边缘，并在处理纹理差异和多尺度道路方面表现出色，有效解决了障碍物遮挡带来的道路连接问题。

表3 在CHN6-CUG数据集道路提取结果

Table 3　Road extraction results in CHN6-CUG dataset

%

方法	CHN6-CUG数据集			
	Precision	Recall	F_1-score	mIoU
Unet[10]	78.56	88.30	83.15	73.56
FCN[8]	90.83	90.97	90.90	74.94
Deeplabv3+[19]	90.98	90.35	90.66	75.70
D-LinkNet[11]	90.97	91.01	90.99	76.66
SGCN[20]	88.97	52.64	66.14	67.10
A²-FPN[21]	90.19	78.42	83.89	75.89
EGE-UNet[22]	48.91	49.99	49.45	60.91
VM-UNet[23]	73.33	61.95	67.17	67.86
ACFD-LinkNet	**92.22**	**92.02**	**92.12**	**77.90**

2.3.3　消融实验

首先，通过在网络不同位置嵌入不同数量的条带注意力模块，全面验证网络整体性能表现，得到指标最优地嵌入位置，并在DeepGlobe和

CHN6-CUG数据集对不同模块性能进行消融实验。

表4为消融模型在两个数据集上的评价指标。在DeepGlobe数据集上,除Precision指标外,在空洞卷积左侧加入条带注意力模块的Recall、F_1-score和mIoU指标均最高,mIoU指标在两个数据集上比加入右侧和两侧分别高出了0.82%、1.22%,1.08%、1.65%。这说明在特征提取初期引入条带注意力模块,可缓解空洞卷积带来的网格效应,聚焦道路区域,提升道路提取精度。因此,本文网络结构基于在空洞卷积左侧引入条带注意力模块进行改进。

图9为消融模型在两个数据集上提取的可视化结果图,第1、2行选取自DeepGlobe数据集,第3、4行选取自CHN6-CUG数据集。只引入条带注意力模块后网络从水平和垂直方向上进行特征增强,因此识别到的道路整体较为连贯。由表4可知,在DeepGlobe数据集上Precision、Recall、F_1-score、mIoU精度较原模型分别提升了0.57%、3.52%、2.40%、1.02%,在CHN6-CUG数据集上分别提升了0.55%、0.92%、0.73%、0.39%。这表明在特征提取部分只嵌入条带注意力模块能够抑制不相关信息的干扰,提升模型的预测精度。同样,只添加上下文融合模块后,模型对建筑物和山体阴影遮挡有较强的鲁棒性,各项指标在DeepGlobe数据集上分别提升了0.59%、4.24%、2.84%、0.96%,在CHN6-CUG数据集上依次提升了0.63%、0.39%、0.51%、0.43%。这表明上下文融合模块能有效融合深浅层特征,增强相邻像素的连接从而提高分割精度。与D-LinkNet模型相比,引入条带注意力与上下文融合模块显著增强了特征识别与利用,能优化长距离道路特征的捕获与边界提取,有效解决了道路遮挡连接问题。

表4 消融模型道路提取结果

Table 4 Ablation model road extraction results %

方法	SAM数量	嵌入位置	DeepGlobe数据集				CHN6-CUG数据集			
			Precision	Recall	F_1-score	mIoU	Precision	Recall	F_1-score	mIoU
D-LinkNet	—	—	94.11	73.93	82.80	78.23	90.97	91.01	90.99	76.66
D-LinkNet+SAM	1	空洞卷积左侧	94.68	77.45	85.20	79.25	91.52	91.93	91.72	77.05
D-LinkNet+SAM	1	空洞卷积右侧	95.54	76.14	84.74	78.43	90.18	90.69	90.44	75.97
D-LinkNet+SAM	2	空洞卷积两侧	94.61	72.39	82.02	78.03	90.11	89.63	89.87	75.40
D-LinkNet+CFM	—	—	94.70	78.17	85.64	79.19	91.60	91.40	91.50	77.09
ACFD-LinkNet	—	—	**95.79**	**79.29**	**86.76**	**79.68**	**92.22**	**92.02**	**92.12**	**77.90**

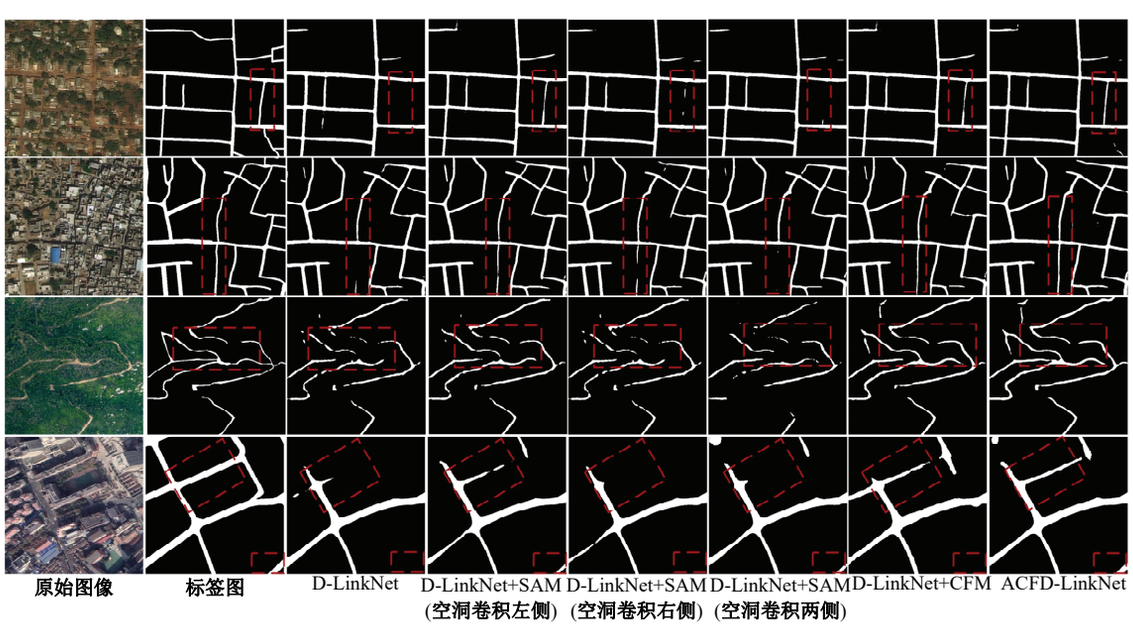

图9 消融模型道路提取结果

Fig. 9 Ablation model road extraction results

3　结束语

为了解决遥感图像中道路跨度大、障碍物遮挡道路连接及数据集正负样本占比不均的问题,本文提出了一种结合注意力与上下文融合的道路提取方法 ACFD-LinkNet。该方法基于D-LinkNet网络结构,首先引入条带注意力模块增强模型对道路的全局特征和局部连接信息的学习能力,使模型更加关注道路区域的重要特征。然后,提出上下文融合模块CFM预测像素之间的连接性,结合上下文信息推断相邻像素的道路属性,更好地融合不同尺度的语义信息。最后,在此基础上对交叉熵损失函数与Dice损失函数设置多损失函数加权解决正负样本失衡,以实现最优提取精度。综合实验结果分析,ACFD-LinkNet模型在两个数据集上的评价指标均取得了最高精度,mIoU分别达到了79.68%、77.90%,各项指标均优于8种对比网络。在主观视觉对比上,ACFD-LinkNet模型有效提高了道路的连续性和完整性,以及遮挡情况下导致的细小道路缺失的问题。

参考文献:

[1] Chen R, Li X, Hu Y, et al. Road extraction from remote sensing images in wildland-urban interface areas [J]. IEEE Geoscience and Remote Sensing Letters, 2020, 19: 1-5.

[2] Zhao K, Liu J, Wang Q, et al. Road damage detection from post-disaster high-resolution remote sensing images based on tld framework[J]. IEEE Access, 2022, 10: 43552-43561.

[3] Zhang X, Jiang Y, Wang L, et al. Complex mountain road extraction in high-resolution remote sensing images via a light roadformer and a new benchmark [J]. Remote Sensing, 2022, 14(19): No. 4729.

[4] 宦海, 盛宇, 顾晨曦. 基于遥感图像道路提取的全局指导多特征融合网络[J]. 浙江大学学报: 工学版, 2024, 58(4): 696-707.
Huan Hai, Sheng Yu, Gu Chen-xi. Global guidance multi-feature fusion networ kbased on remote sensing image road extraction[J]. Journal of Zhejiang University(Engineering Science Edition), 2024, 58(4): 696-707.

[5] 谭国金, 欧吉, 艾永明, 等. 基于改进DeepLabv3+模型的桥梁裂缝图像分割方法[J]. 吉林大学学报: 工学版, 2024, 54(1): 173-179.
Tan Guo-jin, Ou Ji, Ai Yong-ming, et al. Bridge crack image segmentation method based on improved DeepLabv3+ model[J]. Journal of Jilin University (Engineering and Technology Edition), 2024, 54(1): 173-179.

[6] 刘洋, 毛克明. 基于自适应反馈机制的小差异化图像纹理特征信息数据检索[J]. 江苏大学学报: 自然科学版, 2025, 46(1): 73-81.
Liu Yang, Mao Ke-ming. Retrieval of texture feature information data for small differentiated images based on adaptive feedback mechanism[J]. Journal of Jiangsu University(Natural Science Edition), 2025, 46(1): 73-81.

[7] 杨洋, 何童瑶, 詹永照, 等. 基于软聚类的深度图增强方法[J]. 江苏大学学报: 自然科学版, 2024, 45(2): 183-190.
Yang Yang, He Tong-yao, Zhan Yong-zhao, et al. Depth image enhancement method based on soft clustering[J]. Journal of Jiangsu University(Natural Science Edition), 2024, 45(2): 183-190.

[8] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015: 3431-3440.

[9] Wang S, Mu X, Yang D, et al. Road extraction from remote sensing images using the inner convolution integrated encoder-decoder network and directional conditional random fields[J]. Remote Sensing, 2021, 13(3): No. 465.

[10] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation [C]//18th International Conference on Medical Image Computing and Computer-assisted Intervention-MICCAI, Munich, Germany, 2015: 234-241.

[11] Zhou L, Zhang C, Wu M. D-LinkNet: LinkNet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, USA, 2018: 182-186.

[12] Chaurasia A, Culurciello E. Linknet: Exploiting encoder representations for efficient semantic segmentation[C]//2017 IEEE Visual Communications and Image Processing, St. Petersburg, USA, 2017: 1-4.

[13] Wu K, Cai F. Dual Attention D-LinkNet for road segmentation in remote sensing images[C]//2022 IEEE 14th International Conference on Advanced Infocomm Technology, Chongqing, China, 2022:

304-307.

[14] Kampffmeyer M, Dong N, Liang X, et al. ConnNet: a long-range relation-aware pixel-connectivity network for salient segmentation[J]. IEEE Transactions on Image Processing, 2018, 28(5): 2518-2529.

[15] Maji D, Sigedar P, Singh M. Attention Res-UNet with guided decoder for semantic segmentation of brain tumors[J]. Biomedical Signal Processing and Control, 2022, 71: No. 103077.

[16] Dai L, Zhang G, Zhang R. RADANet: road augmented deformable attention network for road extraction from complex high-resolution remote-sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-13.

[17] Song Q, Mei K, Huang R. AttaNet: Attention-augmented network for fast and accurate scene parsing [C]//Proceedings of the AAAI Conference on Artificial Intelligence, Menlo Park, USA, 2021, 35(3): 2567-2575.

[18] Ouyang D, He S, Zhang G, et al. Efficient multi-scale attention module with cross-spatial learning[C]// ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 2023: 1-5.

[19] 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, Munich, Germany, 2018: 801-818.

[20] Zhou G, Chen W, Gui Q, et al. Split depth-wise separable graph-convolution network for road extraction in complex environments from high-resolution remote-sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 1-15.

[21] Li R, Wang L, Zhang C, et al. A2-FPN for semantic segmentation of fine-resolution remotely sensed images[J]. International Journal of Remote Sensing, 2022, 43(3): 1131-1155.

[22] Ruan J, Xie M, Gao J, et al. Ege-unet: an efficient group enhanced unet for skin lesion segmentation[C]// International Conference on Medical Image Computing and Computer-Assisted Intervention, Vancouver, Canada, 2023: 481-490.

[23] Ruan J, Xiang S. Vm-unet: Vision mamba unet for medical image segmentation[J/OL]. [2024-03-25]. https://arxiv.org/abs/2402.02491