第 43 卷 第 24 期/2023 年 12 月/光学学报 研究论文

光学学报

注意力引导与多特征融合的遥感影像分割

张印辉[1]，张枫[1]，何自芬[1*]，杨小冈[2]，卢瑞涛[2]，陈光晨[1]

[1]昆明理工大学机电工程学院，云南 昆明 650500；

[2]火箭军工程大学导弹工程学院，陕西 西安 710025

摘要 从遥感影像能够获取到精度高、范围广的地物信息，因而遥感影像在高空侦察和精确制导等领域得到广泛应用。针对遥感影像地物目标边缘模糊、尺度多变导致难以精准分割的问题，提出以深度残差网络为主干并结合注意力引导与多特征融合的分割方法，命名为AMSNet。首先，采用类别引导通道注意力模块提高模型对难分辨区域的敏感性；其次，嵌入特征复用模块减少遥感影像特征提取过程中边缘损失和小尺度目标丢失的问题；最后，设计跨区域特征融合模块以增强对多尺度特征信息的获取能力，并耦合多尺度损失融合模块对损失函数进行优化，综合提升模型对多尺度遥感影像目标的分割能力。选取3组遥感影像数据集进行对比实验，结果表明，AMSNet能够有效分割遥感影像地物目标边缘和多尺度目标。

关键词 遥感影像；语义分割；注意力机制；多尺度特征

中图分类号 P237 **文献标志码** A **DOI**：10.3788/AOS230631

1 引 言

遥感卫星成像技术凭借成像精度高、探测范围广、时效性好、不受地形限制等诸多优势[1]，成为高空侦察、精确制导及地形匹配等领域获取地物信息的重要方式。然而随着如今数据量的激增，当前智能化与自动化目标提取方法难以满足需求[2]。因此如何实现高效、精准的自动化目标提取，已成为遥感卫星图像领域研究的重点。

传统遥感图像提取包括边缘检测[3]、阈值分割[4]和区域分割[5]等方法，这些方法对具有显著轮廓边界的遥感目标分割效果较好，但面对复杂多变的遥感目标时缺乏自适应调节能力。相比传统遥感提取技术，卷积神经网络提取图像中多层级语义信息，具有表征能力强、扩展性能强及鲁棒性能好等优势。在卷积神经网络语义分割方面，文献[6-7]为还原更多目标细节信息，先采用主干网络提取图像主要目标特征，再利用上采样融合深层与浅层特征，以此构建编码-解码的网络模型。文献[8]和文献[9]在编码-解码结构的基础上分别结合压缩激励模块和自适应空间池化模块，提升网络在遥感影像中的分割效果。文献[10-11]为强化上下文信息，建立池化金字塔结构进行多尺度目标特征信息融合。文献[12]和文献[13]在池化金字塔结构基础上分别构建多尺度自适应模块和自适应融合模块，改善遥感影像中目标尺度差异过大引起的分割精度不足的问题。此外，文献[14-15]为提升网络对特定遥感目标的关注能力，从注意力机制方面着手，利用注意力模块提取遥感图像关键语义信息并对特定对象进行自适应特征优化。文献[16-17]分别加入级联注意力模块和位置注意力模块，以提高对遥感影像小尺度目标的检测能力。文献[18]加入压缩注意力模块，提升模型在复杂场景下的海陆边界分割精度。基于神经网络的分割方法在遥感影像中已取得长足进步。然而，遥感影像存在地物目标分布不均、边缘模糊、尺寸多变等问题，使分割网络在特征提取过程中容易丢失边缘信息和多尺度特征信息；此外复杂场景下云层遮挡遥感目标，更是加剧目标边缘和多尺度目标特征信息损失。

针对上述问题，本文设计了以深度残差网络为主干同时结合注意力引导与多特征融合的分割方法，命名为AMSNet。首先，在主干中加入类别引导通道注意力模块，引导网络关注与分割类别有关的重要通道信息，提升对难分辨目标区域信息的获取能力；其次，嵌入特征复用模块，关联浅层特征与深层特征，弥补深层特征中细节缺失的信息；最后，提出跨区域特征融合模块，通过融合多层特征信息提升对多尺度目标信息的获取能力并借助MLPSeg Head解码模块进行辅助损失计算，对所得结果进行加权融合，进一步提升网络

收稿日期：2023-03-06；**修回日期**：2023-03-29；**录用日期**：2023-04-24；**网络首发日期**：2023-05-08

基金项目：国家自然科学基金(62061022,62171206)

通信作者：*zyhhzf1998@168.com

对多尺度目标的分割效果。在3组数据集中进行实验，结果表明，相比其他主流分割网络，所提网络对遥感目标边缘分割更加清晰，对多尺度目标分割更加精准。

2 网络结构

所提结合注意力引导与多特征融合的遥感影像语义分割网络AMSNet的结构如图1所示。在编码结构部分(Encoder Section)，采用D_Resnet50作为主干网络提取遥感影像中的主要特征信息，并向其中插入类别引导通道注意模块(CGCAM)，加强网络对遥感影像中难分辨、不规则形状区域的分割能力，再加入特征复用模块(FRM)以解决网络在提取特征过程中边缘细节信息损失和零散小尺度目标消失的问题。在解码结构部分(Decoder Section)，利用跨区域特征融合模块(CRFFM)将多特征信息融合，并结合多尺度损失融合模块(MLFM)进一步提升网络对多尺度目标的分割效果。

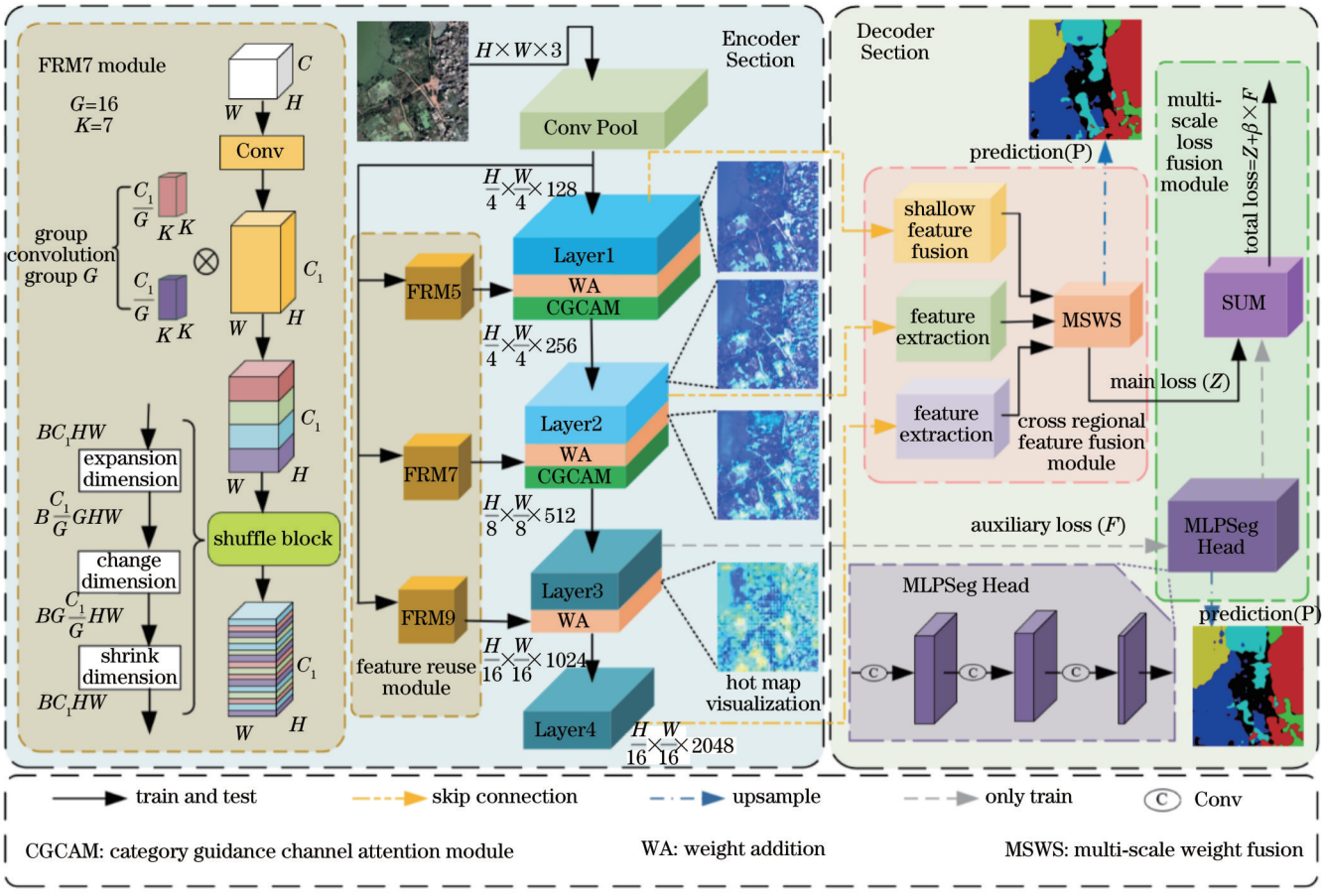

图1 AMSNet的结构

Fig. 1 Architecture of the AMSNet

2.1 主干网络

遥感影像中地物目标交错相融并且小尺度目标分布零散，导致主干网络在特征提取过程中会丢失大量的细节信息。而深度残差网络可以将浅层细节特征更可靠地抽象成高维语义信息，网络层数的增加有利于深层特征信息的提取，因此深度残差神经网络可以加深网络层数的同时降低梯度爆炸或者梯度消失的风险。在深度残差网络Resnet50[19]的基础上首先将1个7×7卷积替换为3个串联的3×3卷积并与最大池化组成卷积池化模块(Conv Pool)，其次降低模型下采样倍率以缓解特征提取过程中遥感目标边缘损失和零散小尺度目标消失的问题，最后在Layer3和Layer4中加入空洞率(Dilated Conv)分别为2和4的空洞卷积[20]，增大模型在深层特征上的感受野，获取更全面的遥感目标特征信息。最终提出的D_Resnet50主干网络能够提升对遥感影像地物目标边缘和小尺度目标等细节信息的获取能力。

2.2 类别引导通道注意力模块

遥感影像中地物目标相互交融导致边缘模糊，且受光线、云层及阴影等环境因素的干扰，增加了模型识别目标的难度。而注意力机制帮助网络聚焦目标区域内的重点信息，按照目标重要程度赋予不同权重，降低非目标特征干扰。本文设计了类别引导通道注意力模块，将通道信息与分割对象关联，引导网络关注与分割对象相关的通道信息，提升对应通道信息权重，增强网络对目标边缘和难分辨区域的识别能

力。图 2 所示的类别引导通道注意力模块首先利用平均池化(Avg Pool)操作将输入特征图的长宽压缩为 1×1 大小，其次利用全连接(liner)将通道维度上的特征信息转化为类别数量(N)。然后将转置(transpose)前后的特征图相乘，得到关于通道信息和类别数量的特征图，接着利用全连接对得到的特征图在通道维度上进行压缩(channel compression)和激励(channel expansion)，得到关于类别数量的通道权重(w)。最后将新得到的特征图与转置后的特征图相乘，并利用维度扩张(dimensional expansion)恢复至长宽为 1×1 大小的特征图，在此基础上利用 Sigmoid 函数对通道维度信息进行归一化，并与输入特征信息相乘，实现对图片通道信息的重新赋值。N 和 w 的表达式分别为

$$N = L_1\left[\mathrm{AP}(X)\right], \tag{1}$$

$$w = L_3 \times R\left[L_2(C \times N)\right], \tag{2}$$

式中：AP(·)为平均池化；R(·)为激活函数；L_1、L_2、L_3 为全连接层。

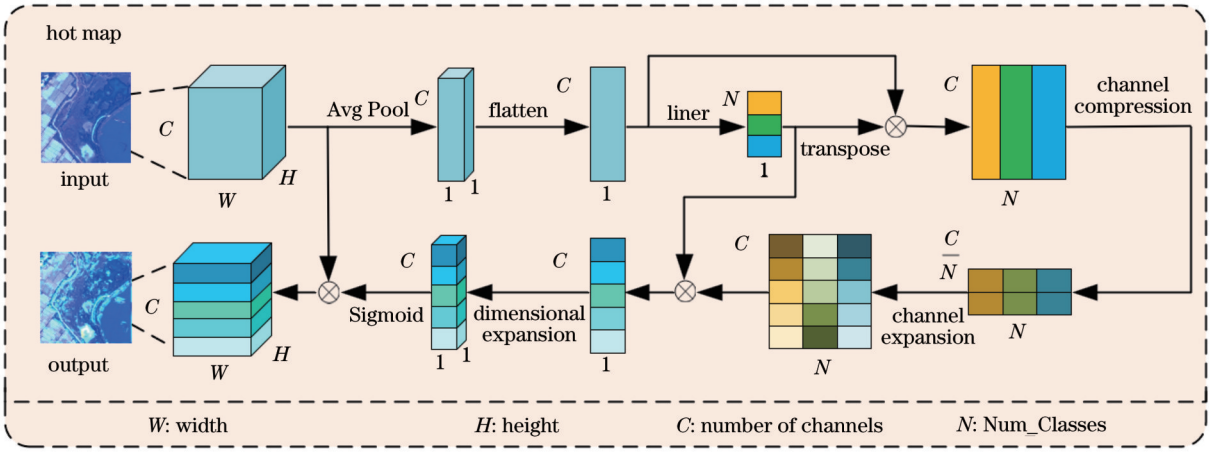

图 2　类别引导通道注意力模块

Fig. 2　Category guidance channel attention module

2.3　特征复用模块

主干网络在遥感影像特征提取过程中需要进行大量卷积操作，然而这会导致遥感影像中目标边缘信息损失和小尺度目标丢失。为弥补目标特征提取时的细节信息损失，设计了特征复用模块，该模块包含 FRM5、FRM7 和 FRM9 三个子模块(结构如图 1 所示)。先利用三个子模块分别对卷积池化模块产生的特征进行多尺度的特征信息提取，再利用跳跃连接对提取到的特征信息分别与 Layer1、Layer2 和 Layer3 产生的特征信息进行权重相加，实现多特征融合，通过向主干网络中融入复用特征信息，保留更多的遥感目标细节特征。三个子模块实现过程以 FRM7 模块为例，首先使用一个 1×1 卷积增加输入图片的通道数，其次为降低 7×7 卷积参数量和计算量，将通道划分为 16 个组后进行分组卷积。分组卷积的参数量(N_{params})和计算量(N_{FLOPs})为普通卷积参数量($N_{\mathrm{params\,p}}$)的 6.25% 和计算量($N_{\mathrm{FLOPs\,p}}$)的 6.251%。但分组卷积会导致通道间组别关联缺失，无法有效将各组特征信息分散到其他组中。因此加入混洗模块(shuffle block)[21]对输入特征在通道维度上进行扩张(expansion dimnension)，再交换通道维度(change dimension)信息，最后收缩维度(shrink dimension)还原到初始通道维度，建立组别信息关联，提升通道间信息的交互性。各参数的表达式分别为

$$N_{\mathrm{params\,p}} = K^2 \times C_{\mathrm{r}} \times C_{\mathrm{c}}, \tag{3}$$

$$N_{\mathrm{params}} = K^2 \times C_{\mathrm{r}}/g \times C_{\mathrm{c}}/g \times g, \tag{4}$$

$$N_{\mathrm{FLOPs\,p}} = \left(2 \times C_{\mathrm{r}} \times K^2\right) \times H \times W \times C_{\mathrm{c}}, \tag{5}$$

$$N_{\mathrm{FLOPs}} = \left[\left(2 \times K^2 \times C_{\mathrm{r}}/g + 1\right) \times H \times W \times C_{\mathrm{c}}/g\right] \times g, \tag{6}$$

式中：K 表示卷积核；g 表示分组数；C_{r} 表示输入通道数；C_{c} 表示输出通道数。

2.4　跨区域特征融合模块

解码过程一般采用类似特征金字塔网络(FPN)[22]的金字塔多层级解码结构或者 U 形结构，其功能都是将深层特征图上采样再融合原特征，以恢复图像在卷积过程中丢失的特征信息，但这些结构会导致模型参数量大幅增加。本文针对深层与浅层特征图所呈现的不同遥感目标特征及多尺度目标分割任务需求，设计了跨区域特征融合模块，如图 3 所示，该模块包含 4 个子模块，分别为浅层特征融合模块、特征提取模块、特征恢复模块及多尺度权重融合模块，通过这 4 个子模块对多层特征信息进行有效融合。浅层特征融合模块中，对 Layer1 上的特征图分别进行 3×3 卷积和最大池化操作，获取多尺度特征图，然后对获得的特征图在维度方向进行拼接，最后利用 1×1 卷积融合拼接后的特征信息。特征提取模块中，利用 3×3 卷积操作进一步提取 Layer2 上的特征信息。特征恢复模块中，

通过多层卷积感知机模块(MCP)将Layer4上的特征图的通道维度信息转换为新的通道维度信息并保留重要原始特征信息,然后与原特征进行跳跃连接并利用1×1卷积进行特征融合,最后对获取到的特征图进行2倍上采样,恢复下采样丢失的特征信息。多尺度权重融合模块中,对浅层特征融合模块与特征提取模块的输出沿通道维度进行拼接,再利用1×1卷积进行特征融合,并与特征恢复模块输出的特征信息进行权重相加实现多特征融合,提升模型对多尺度遥感目标的分割效果。

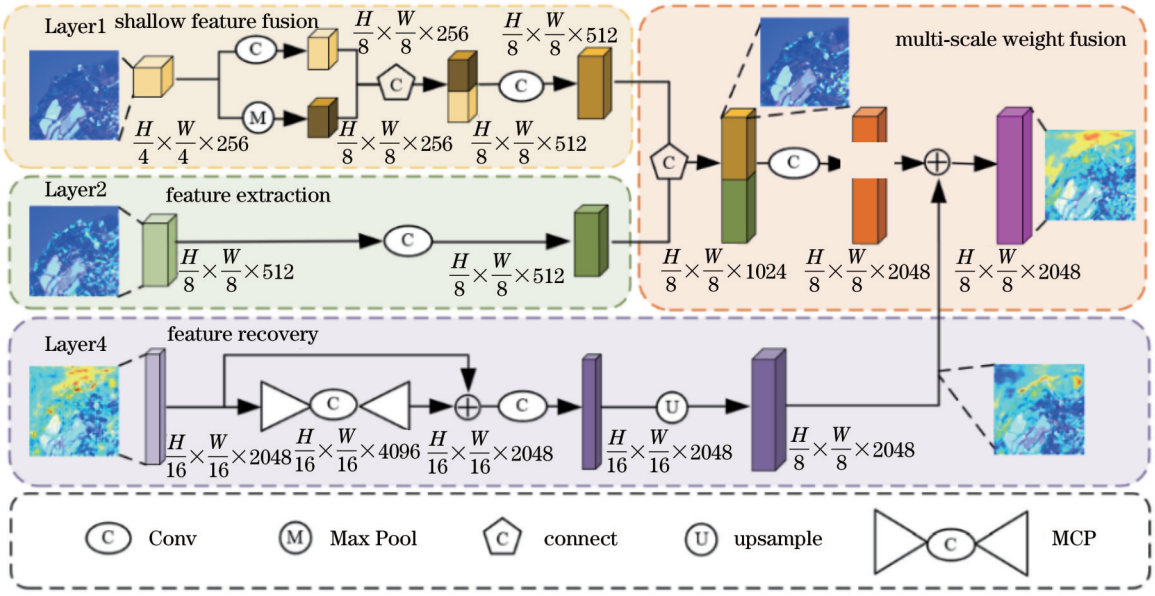

图3 跨区域特征融合模块

Fig. 3 Cross regional feature fusion module

2.5 多尺度损失融合模块

神经网络通过损失函数计算模型每次迭代后产生的结果与真实结果之间的差距,指导网络沿正确方向梯度下降,然而随着网络层数的增加,损失产生的梯度无法有效回传到网络各层。PspNet引入辅助损失函数优化学习过程。BiseNetv2[23]在多个网络层上使用多个损失函数,利用多个损失对梯度下降方向加以约束以提高分割精度。受此启发,本文设计了MLPSeg Head辅助解码模块,由于遥感目标尺度多变,对主解码模块和辅助解码模块分别在不同尺度特征图上进行交叉熵损失计算,并对两者损失进行加权融合,表达式为

$$L_z=-\sum_{c=1}^{N} y\log P_c+\left(-\beta\sum_{b=1}^{N} y'\log P_b\right),\quad (7)$$

式中:N表示类别数;y和y'分别表示主损失和辅助损失预测值,预测值与真实值相同为1,不同为0;P_c和P_b代表观测样本属于c和b类的预测概率。MLPSeg Head解码模块(如图1所示)首先利用1×1卷积上升维度,然后利用1×1卷积进行特征映射,最后利用1×1卷积降低通道数,进一步提取目标特征信息。

3 实验结果及分析

3.1 数据集建立

实验采用两个数据集:高原区域遥感影像数据集(数据集1);云层干扰下高原区域遥感影像数据集(数据集2)。遥感数据来源于WorldView-2卫星(光谱范围为450~800 mm,分辨率为0.46 m,全色)所拍摄的云南省昆明市滇池区域,并采用Labelme进行标注,标注对象为建筑区域(Building)、植被区域(Vegetation)、湖泊区域(Lake)、河流区域(River)、农田区域(Farmland)和背景(Background)共6类。对遥感卫星成像的图片进行裁剪,如图4所示,最终形成1600张含有标签信息的500×500像素大小的遥感影像数据集,按照4:1设置训练集与测试集。

图4 云层干扰遥感影像

Fig. 4 Cloud disturbance remote sensing image

云层干扰下高原区域遥感影像数据集是在高原区域遥感影像数据集的基础上考虑到在卫星遥感成像过程中存在云层干扰导致光线受到遮挡，成像效果受到干扰导致成像区域目标不清晰的情况采集的，如图 4 所示。为了更好地比较网络在不同气象条件下的鲁棒性，利用 Imagecorruptions 软件包进行云层干扰生成，如图 5 所示。考虑到云雾仅存在某一小片区域内，随机选取部分区域进行云雾干扰条件的生成，共生成 200 张带有云层干扰的图片。训练集和测试集与数据集 1 保持一致。

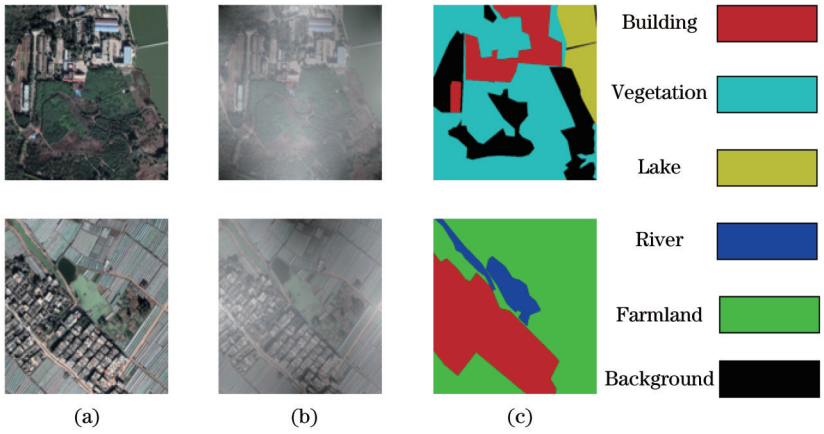

图 5 数据集展示。(a)原图；(b)气候变换图；(c)标签

Fig. 5 Show of dataset. (a) Original images; (b) after image corruption; (c) labels

3.2 实验环境

实验基于 Ubuntu18.04 操作系统，使用 CPU 为英特尔 i5-11400，内存为 40 GB，GPU 为 12G 版本 NVIDIA GeForce RTX3060 的硬件平台。深度学习框架为 PyTorch1.8.1，使用 CUDA11.1 和 cudnn8.0.4 加速模型训练，Python 版本为 3.7，图片输入尺寸为 500×500，图片批量数(batch size)为 4，学习率为 0.01，迭代步数为 50000，动量衰减为 0.0005，优化器使用随机梯度下降(SGD)。

3.3 评价指标

从总体精度(OA)、每类精度和预测速度 3 个角度评估网络在数据集上的表现，评价指标有平均交并比(mIoU)、平均像素精度(mPa)、平均 F1 分数(mF1)。计算公式分别为

$$R_{\mathrm{mIoU}}=\frac{1}{N+1}\sum_{n=0}^{N}\frac{p_{nm}}{\sum_{m=0}^{N}p_{nm}+\sum_{m=0}^{N}p_{mn}-p_{nn}},\tag{8}$$

$$R_{\mathrm{mPa}}=\frac{1}{N+1}\sum_{n=0}^{N}\frac{p_{nm}}{\sum_{m=0}^{N}p_{nm}},\tag{9}$$

$$R_{\mathrm{precision}}=\frac{p_{nn}}{\sum_{m=0}^{N}p_{nm}+p_{nn}},\tag{10}$$

$$R_{\mathrm{recall}}=\frac{p_{nn}}{\sum_{m=0}^{N}p_{mn}+p_{nn}},\tag{11}$$

$$R_{\mathrm{F1\text{-}scorce}}(n)=\frac{2\times R_{\mathrm{precision}}(n)\times R_{\mathrm{recall}}(n)}{R_{\mathrm{precision}}(n)+R_{\mathrm{recall}}(n)},\tag{12}$$

$$R_{\mathrm{mF1}}=\frac{1}{N+1}\sum_{n=0}^{N}R_{\mathrm{F1\text{-}scorce}}(n),\tag{13}$$

式中：N 为类别数量；n 代表第 n 类，m 代表第 m 类；P_{nm} 代表真实值是第 n 类，预测值是第 m 类的概率；$R_{\mathrm{precision}}$ 为精准率；R_{recall} 为召回率。

4 实验及结果分析

4.1 模块消融实验

为探究各模块对模型分割精度的影响，在 D_Resnet50 主干网络的基础上，依次加入多尺度损失融合模块(MLFM)、类别引导通道注意力模块(CGCAM)、跨区域特征融合模块(CRFFM)和特征复用模块(FRM) 4 个模块，分别组成 D_Resnet50、MLNet、MCNet、MCCNet 及 AMSNet 网络模型。各模块消融实验在高原区域遥感影像数据集上的结果如表 1 所示。

从表 1 可以看出，向主干网络中依次加入各模块后，模型的分割效果得到提升。首先，向主干网络(D_Resnet50)中加入多尺度损失融合模块组成的 MLNet 能够对遥感影像中尺度多变的地物目标进行有效目标分割，mIoU 与 mPa 相比模块加入前增长 1.27 个百分点与 0.97 个百分点。向 MLNet 中加入类别引导通道注意力模块组成的 MCNet 能够提高对遥感影像中难分辨区域内目标的分割能力，mIoU 为 75.59%，mPa 为 85.15%，mF1 为 85.80%。接着，向 MCNet 中加入跨区域特征融合模块组成的 MCCNet 能够对遥感影像中多层特征信息进行跨区域融合，提高网络对遥感影像中多尺度特征信息的提取能力，mIoU 为 76.56%，mPa 为 85.80%，mF1 为 86.48%。最后，向 MCCNet 中加入特征复用模块，缓解网络在特征提取过程中边缘特征和小尺度目标信息丢失的问题，组成的 AMSNet 相比初始的骨干网络模型 mIoU 增长 4.24 个百分点，mPa 增长 3.27 个百分点，mF1 增长 2.87 个百分点。

表 1 在 D_Resnet50 主干网络上各模块的消融实验结果

Table 1 Results of ablation test of each module on D_Resnet50 backbone

Network	MLFM	CGCAM	CRFFM	FRM	mIoU /%	mPa /%	mF1 /%
D_Resnet50					73.53	83.34	84.40
MLNet	√				74.80	84.31	85.26
MCNet	√	√			75.59	85.15	85.80
MCCNet	√	√	√		76.56	85.80	86.48
AMSNet	√	√	√	√	77.77	86.61	87.27

4.2 主干网络的消融实验

主干是网络进行特征提取的重要组成部分，为了探究不同主干对网络分割的影响程度，分别选取 Resnet50、D_Resnet50、Shufflenetv2[24] 及 ConvNeXt[25] 4 个主干网络，在多尺度损失融合模块、类别引导通道注意力模块、跨区域特征融合模块和特征复用模块这 4 个模块不变的情况下仅仅变换主干，在高原区域遥感影像数据集上的实验结果如表 2 所示。D_Resnet50 作为主干的网络的分割效果优于其他主干网络，相比 Resnet50 主干，有着更深的网络结构、更大的感受野，mIoU 增长 4.92 个百分点，mPa 增长 3.52 个百分点，mF1 增长 3.32 个百分点。Shufflenetv2 为轻量化主干，在所有对比主干中，所需的计算复杂度最小，浮点运算数（FLOPs）仅仅只有 49.89×10^9，相应的网络深度较浅，通道信息较少，分割精度较低，但推理速度更快，达 8.8 frame/s。在所有对比主干中，ConvNeXt 地物目标分割精度最低，mIoU 为 66.69%，mPa 为 78.94%，mF1 为 79.48%，原因在于遥感影像中存在零散分布的小尺度目标，目标与目标间交错相融分割边界不清晰，ConvNeXt 使用 7×7 的深度可分离卷积，一方面大卷积核容易导致小尺度目标消失和目标边缘损失，另一方面深度可分离卷积在减少计算参数量的同时破坏通道信息，降低通道信息的承载量。

表 2 不同主干的消融实验

Table 2 Ablation test of different backbones

Backbone	mIoU /%	mPa /%	mF1 /%	GLOPs /10^9	Speed /(frame·s^{-1})
Resnet50	72.85	83.09	83.95	256.81	7.1
D_Resnet50	77.77	86.61	87.27	546.98	6.4
Shufflenetv2	71.12	81.60	82.64	49.89	8.8
ConvNeXt	66.69	78.94	79.48	127.47	7.8

4.3 类别引导通道注意力模块消融实验

注意力模块可以加强网络对特定对象的关注能力，为了探究类别引导通道注意力模块在不同通道维度上给网络分割精度带来的影响，在主干网络 D_Resnet50 的 4 个不同通道上加入注意力模块，在高原区域遥感影像数据集上的实验结果如表 3 所示。当在单个位置上使用注意力模块时，浅层特征上使用注意力模块的分割效果总体要好于深层特征上使用注意力模块并且在 Layer2 上取得较好的分割效果，mIoU 为 76.44%，相比 Layer4 上的分割效果，mIoU 提高 1.13 个百分点；在多个位置上使用注意力模块时，在主干前两层加入注意力模块后分割效果达到最好，相比在主干后两层加入注意力模块，mIoU 提高 2.3 个百分点。原因在于浅层特征上使用注意力模块，注意力会在浅层特征通道维度上选择与类别对象有关的通道信息，能够为深层特征通道信息进行引导，而在深层特征上使用注意力模块，注意力更关注深层特征通道，缺乏对浅层通道的影响，因此在浅层特征上使用注意力模块取得的效果较好。借助图 2 的热力图（hot map）进行分析：添加注意力模块前，对于输入的特征图（input），主要关注农田和小部分河流区域；添加注意力模块后，对于输出的特征图（output），对农田区域和湖泊区域的关注范围变大同时对河流、农田及湖泊等边缘模糊及难分辨区域给予更多关注。

4.4 特征复用模块消融实验

网络在进行特征提取过程中会产生目标边缘损失

表 3 类别引导通道注意力模块的消融实验

Table 3 Ablation experiment of category guidance channel attention module

Place for attention	mIoU /%	mPa /%	mF1 /%
Layer1	76.36	85.97	86.34
Layer2	76.44	85.85	86.40
Layer3	76.32	85.77	86.32
Layer4	75.31	84.91	85.63
Layer1+Layer2	77.77	86.61	87.27
Layer3+Layer4	75.47	85.06	85.73
Layer1+Layer2+Layer3	76.50	86.09	86.43
Layer1+Layer2+Layer3+Layer4	75.70	85.21	85.89

和小尺度目标丢失的问题。为解决这个问题，利用特征复用模块(FRM)将卷积池化模块(Conv Pool)产生的特征图直接与主干网络中不同特征层相融合，形成融合多层特征信息的特征图，并进行消融实验，在高原区域遥感影像数据集上的实验结果如表4所示。

表4 特征复用模块消融试验

Table 4 Ablation experiment of feature reuse module

FRM	mIoU /%	mPa /%	mF1 /%
Origin	76.56	85.80	86.48
FRM5	76.61	85.89	86.51
FRM5+FRM7	77.35	86.38	87.00
FRM5+FRM7+FRM9	77.77	86.61	87.27
FRM5+FRM7+FRM9+FRM11	77.14	86.27	86.85

根据表4可知：在网络中加入FRM5模块，相比加入前，mIoU增长0.05个百分点，mPa增长0.09个百分点，mF1增长0.03个百分点，评估指标涨幅微小，原因在于浅层特征上使用特征复用模块后，由于浅层特征与输入图片相距较近，在特征提取过程中并不会丢失太多的目标边缘特征和小尺度目标等细节信息，因此在浅层特征图中加入特征复用模块并不会对精度有较大提升；随着向网络中继续依次加入FRM7和FRM9模块，网络的mIoU依次增长0.74个百分点和0.42个百分点，mPa依次增长0.49个百分点和0.23个百分点，原因在于深层特征与输入图片相距较远，经多层卷积获取到比浅层特征更丰富的语义信息，但在卷积过程中丢失大量细节特征和小尺度目标信息，使得分割精度下降；往深层特征图中加入特征复用模块，弥补特征提取过程中细节特征的损失，mIoU比加入前增长1.21个百分点。

4.5 多尺度损失融合模块消融实验

损失函数用来评估模型的预测值与真实值之间的误差，引入辅助损失函数有助于优化网络的学习过程，在实验中将MLPSeg Head解码模块的损失权重β值设置在0到1.4之间，在高原区域遥感影像数据集上的实验结果如表5所示。

表5 多尺度损失融合模块的消融实验

Table 5 Ablation experiment of multi-scale loss fusion module

Main loss weight	β	mIoU /%	mPa /%	mF1 /%
1	0	75.61	85.37	85.86
1	0.2	76.46	86.12	86.42
1	0.4	76.73	86.23	86.60
1	0.6	77.49	86.61	87.09
1	0.8	77.22	86.32	86.91
1	1	77.77	86.61	87.27
1	1.2	77.41	86.34	87.04
1	1.4	76.61	85.50	86.50

根据表5可知：在网络中除主损失(main loss)分支外另加入MLPSeg损失分支后，网络的分割效果总体上是随着MLPSeg损失权重β值增加而提高的，当MLPSeg损失权重值为1时，此时网络的分割效果最好，相比不加入MLPSeg损失分支，mIoU提高2.16个百分点，mPa提高1.24个百分点，mF1分数提高1.41个百分点。结果表明MLPSeg损失权重的加入可以更好地约束网络学习的方向，从而能更有效地学习特征信息，提高网络的分割准确度。

4.6 高原区域遥感影像数据集网络对比实验

为了验证所提网络的有效性，对所提网络与其他主流语义分割网络进行比较，引入多种评价指标，如mIoU、mPa、mF1、每类对象精度(accuracy of each class)、计算复杂度(FLOPs)和网络每秒处理图片的速度，对网络进行全面评估。在高原区域遥感影像数据集上的实验结果如表6所示，其中加粗字体为最优值。

从表6可知：BiseNetv2[23]设置两个分支，一个分支

表6 不同网络在高原区域遥感影像数据集上的对比实验

Table 6 Comparative experiment of different networks on remote sensing image dataset of plateau region

Parameter		BiseNetv2	PspNet	Deeplabv3+	OCNet	ISANet	SegNext	AMSNet
mIoU /%		57.80	73.22	74.39	74.93	73.07	73.73	**77.77**
mPa /%		70.69	83.32	83.83	84.68	83.50	84.17	**86.61**
mF1 /%		72.26	84.22	84.98	85.41	84.15	84.56	**87.27**
Accuracy of each class /%	River	38.28	60.56	58.25	64.60	59.61	59.38	**69.97**
	Lake	81.02	87.92	86.79	89.22	84.49	87.13	**91.90**
	Farmland	68.71	81.53	83.39	81.33	81.93	82.17	**84.37**
	Vegetation	49.29	67.99	69.49	70.41	67.38	69.65	**70.74**
	Building	49.52	67.47	72.15	68.49	69.83	68.87	**72.35**
	Background	59.96	73.85	76.29	75.51	75.14	75.20	**77.30**
FLOPs /10^9		**43**	763	708	658	539	188	546
Speed /(frame·s^{-1})		**9.1**	5.7	5.5	6.0	6.0	6.2	6.4

为细节分支，用于捕捉低层特征中的细节信息，另一分支为语义分支，用于提取高层特征中的语义信息，并设计引导聚合层对两分支上的特征信息进行融合，但由于其为轻量化网络，相比其他网络，通道信息承载量较小，导致分割精度较低，因此分割精度mIoU仅为57.80%，mPa仅为70.69%，mF1仅为72.26%；PspNet设计金字塔池化模块融合多尺度特征信息，以更好地分割多尺度遥感目标，所得分割精度mIoU为73.22%，mPa为83.32%，mF1为84.22%；Deeplabv3+使用空洞卷积扩大多尺度特征感受野范围，并采用具有编码-解码的网络结构恢复卷积中损失的细节信息，相比PspNet，分割精度mIoU提高1.17个百分点，mPa提高0.51个百分点，mF1提高0.76个百分点；OCNet[26]计算目标当前像素，并对其与其他像素的相似度进行分类，建立目标语义池化，分割精度mIoU为74.93%；ISANet[26]进行双阶段自注意力计算，减少参数计算量，分割精度mIoU为73.07%；SegNext[27]提出的多尺度卷积注意力相比自注意力能更为高效地获取上下文信息，相比ISANet，分割精度mIoU提高0.66个百分点，mPa提高0.67个百分点，mF1提高0.41个百分点；AMSNet将特征复用模块和跨区域特征融合模块等相结合，提高网络对目标边缘及多尺度目标的分割效果，同时结合注意力模块，重点关注与类别对象有关的通道信息，提高对难分辨区域的分割效果，mIoU、mPa以及mF1分别为77.77%、86.61%、87.27%。

对每类对象精度(accuracy of each class)进行分析：所提网络对每类分割对象都取得了较好的分割效果；在河流区域(River)，相比SegNext，分割精度提高10.59个百分点，相比OCNet，分割精度提高5.37个百分点；对湖泊区域(Lake)的分割精度达91.90%，对农田区域(Farmland)、建筑区域(Building)、植被区域(Vegetation)和背景区域(Background)的分割精度相比其他主流分割网络也都有一定提升。

根据模型计算复杂度进行分析：AMSNet的计算复杂度在所有对比网络中并不是最低的，与相似计算量大小的ISANet相比，分割精度mIoU提升4.70个百分点。根据推理速度进行分析：在训练过程中使用MLPSeg Head解码模块能提高模型的训练效果，推理过程中抛弃此模块可以加快网络的推理速度，虽然低于BiseNetv2轻量化网络推理速度，但除BiseNetv2，所提网络取得了最高的推理速度，为6.4 frame/s。

图6为不同网络在高原区域遥感影像数据集上的分割效果。选择7种网络模型在测试集上进行分割结果测试，并挑选其中3种遥感影像场景进行分析。从整体分割效果上分析，OCNet、SegNext及BiseNetv2在部分遥感场景下存在分割类别错误的情况；Deeplabv3+和PspNet采取多尺度空洞卷积扩大网络感受野范围以及融合多尺度特征信息，增强了网络对不同尺度目标的辨别能力，有效降低了网络对不同场景下不同尺度目标的错分概率；ISANet通过分段自注意力方法关联相似像素，但忽视了地物目标边缘模糊和尺度多变的问题，导致分割目标边缘出现断裂，小尺度目标分割不完整。在场景1中，植被区域和河流区域边缘模糊，所设计的类别引导通道注意力模块能提高网络对遥感目标边缘像素的辨别能力，相比其他网络，能有效降低植被区域与河流区域边缘像素类别错分的概率。场景2中植被区域与湖泊区域分布极不均匀，场景3中建筑区域尺度变化较大，导致对比网络在分割过程中出现像素漏分的情况，所设计的跨区域特征融合模块与特征复用模块对遥感影像中不同尺度特征信息进行多特征融合，并结合多尺度损失融合模块，提升了网络对多尺度遥感目标的分割效果，使得所提网络相比其他网络能够更加准确地分割出分布不均的植被区域和尺度多变的建筑区域。

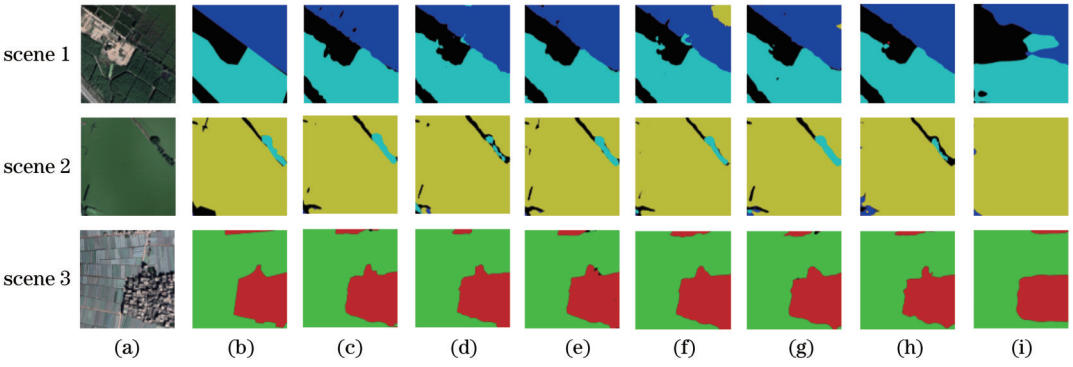

图6 不同网络在高原区域遥感影像数据集上的分割效果。(a)原图；(b)标签；(c)AMSNet；(d)SegNext；(e) ISANet；(f) OCNet；(g) Deeplabv3+；(h)PspNet；(i)BiseNetv2

Fig. 6 Segmentation rendering of different networks on remote sensing image dataset of plateau region. (a) Original image; (b) label; (c) AMSNet; (d) SegNext; (e) ISANet; (f) OCNet; (g) Deeplabv3+; (h) PspNet; (i) BiseNetv2

4.7 云层干扰下高原区域遥感影像数据集网络对比实验

在云层干扰高原区域遥感影像数据集上所提网络与其他分割网络的对比实验结果如表7所示。

在干扰气象条件下，遥感影像数据部分细节特征受到云层干扰，导致分割精度相比于无干扰条件下的均有不同程度的下滑。为了能够在云层干扰下获得更好的分割效果，将AMSNet使用的特征复用模块中分组卷积的分组数缩减到4，以减少分组过程中通道信息丢失的问题。在所有对比的模型中，AMSNet依然取得了较好的分割效果，mIoU、mPa和mF1分别为76.67%、86.03%和86.56%，相比分割效果第二好的Deeplabv3+网络，mIoU、mPa和mF1分别提高了3.85个百分点、3.37个百分点和2.64个百分点，相比OCNet，mIoU、mPa和mF1分别提高了9.63个百分点、6.54个百分点和6.87个百分点。在云层干扰下，OCNet中相似度计算受到云层的干扰，导致分割精度出现下降。

表7 不同网络在云层干扰下高原区域遥感影像数据集上的对比实验

Table 7 Comparative experiment of different networks on remote sensing image dataset of plateau area under cloud disturbance

Parameter		BiseNetv2	PspNet	Deeplabv3+	OCNet	ISANet	SegNext	AMSNet groups are 4
mIoU /%		55.44	68.00	72.82	67.04	71.03	71.93	**76.67**
mPa /%		67.75	79.06	82.66	79.49	82.34	82.84	**86.03**
mF1 /%		70.13	80.65	83.92	79.69	82.70	83.31	**86.56**
Accuracy of each class /%	River	35.24	56.63	57.04	53.76	55.50	57.31	**68.56**
	Lake	88.92	85.02	86.52	86.29	85.99	86.21	**91.12**
	Farmland	66.21	67.33	80.45	76.25	77.62	80.47	**82.96**
	Vegetation	47.68	65.48	68.57	63.16	67.47	68.03	**69.87**
	Building	43.83	63.33	68.77	54.04	67.63	65.76	**71.05**
	Background	58.77	70.23	75.53	68.73	71.95	73.80	**76.48**
FLOPs /10^9		**43**	763	708	658	539	188	669
Speed /(frame·s^{-1})		**9.1**	5.7	5.6	6.0	6.0	6.2	6.0

根据每类对象精度进行分析：AMSNet对河流区域(River)的分割精度相比BiseNetv2和Deeplabv3+分别提高33.32个百分点和11.52个百分点；对湖泊区域(Lake)的分割精度达到91.12%；对农田区域(Farmland)的分割精度为82.96%，相比SegNext，分割精度提高2.49个百分点。

从模型复杂度分析，由于减少了分组卷积的分组数，模型的FLOPs达669×10^9，相比分组卷积分组数减少前，提升22.5%。这导致模型的推理速度也出现一定程度的下降，推理速度为6.0 frame/s。

图7为不同网络在云层干扰高原区域遥感影像数据集上的分割效果，云层干扰对网络模型的预测效果均产生不同程度的影响。利用分割效果图进行定性分析，选取的3个场景均存在遥感目标分布不均的情况。在场景1中，由于云层干扰，植被区域边缘及小部分建筑物区域难以得到准确分辨，所设计的类别引导通道注意力模块对目标边缘及分辨难、尺寸小的建筑区域给予更多关注，相比其他网络，对植被区域边缘像素的分割错误率和建筑区域的漏检率更低。在场景2和场景3中，面对尺度多变的建筑区域和植被区域时，所设

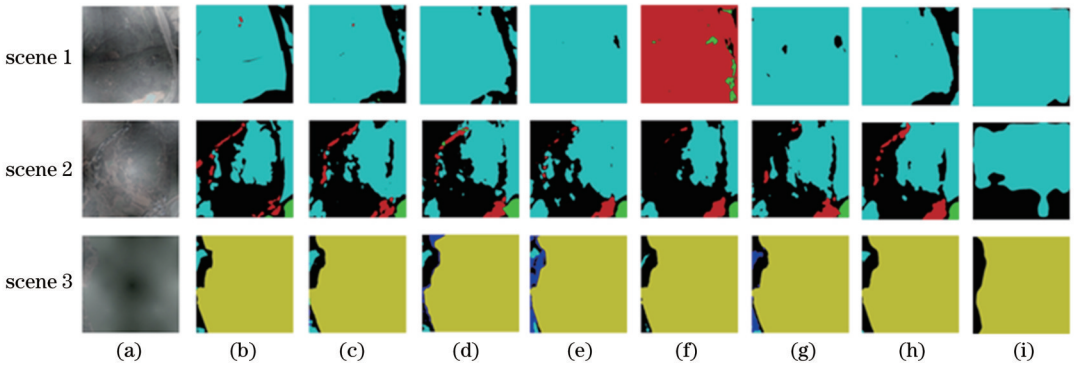

图7 不同网络在云层干扰下高原区域遥感影像数据集上的分割效果。(a)原图；(b)标签；(c)AMSNet；(d)SegNext；(e) ISANet；(f) OCNet；(g) Deeplabv3+；(h)PspNet；(i)BiseNetv2

Fig. 7 Segmentation rendering of different networks on remote sensing image dataset of plateau area under cloud disturbance. (a) Original image; (b) label; (c) AMSNet; (d) SegNext; (e) ISANet; (f) OCNet; (g) Deeplabv3+; (h) PspNet; (i) BiseNetv2

计的跨区域特征融合模块、特征复用模块及多尺度损失融合模块加强了网络对多尺度目标的特征提取能力，相比其他网络，对不同尺度的建筑区域和植被区域有着更高的分割准确度。

在高原区域遥感影像数据集和云层干扰下高原区域遥感影像数据集上，相比其他语义分割网络，所提网络无论在有无云层干扰条件下均有较好的分割效果，并且分割效果受云层干扰影响较小，在有云层干扰下对地物目标的mIoU、mPa、mF1仅比无云层干扰下的低1.10个百分点、0.58个百分点、0.71个百分点，低于其他语义分割网络在不同云层气象干扰条件下对分割效果的影响。

4.8 ISPRS Vaihingen数据集实验

为了验证AMSNet网络分割效果的泛化性，选用在德国Vaihingen地区采集的International Society for Photogrammetry and Remote Sensing(ISPRS)数据集，其包含树木(Tree)、低矮植被(Vegetationn)、建筑物(Building)、汽车(Car)、背景(Background)和不透水面(River)共6类。选取其中17张，裁剪成250×250像素大小共1048张，随机分成848张训练集和200张测试集。

为了在ISPRS Vaihingen数据集上更好地适应图片尺寸，将AMSNet中特征复用模块的分组卷积的分组数缩减到4。从表8可以看出：所提网络相比其他网络依旧有较好的分割表现；相比PspNet和OCNet，mIoU分别提高5.09个百分点和5.57个百分点；相比Deeplabv3+，mIoU提高3.47个百分点，mPa提高3.56个百分点，mF1提高2.78个百分点。图8为不同网络在ISPRS Vaihingen数据集上的分割效果，相比其他网络，所提网络对建筑物边缘和小尺度汽车有着更低的错分率、更少的漏分情况和更准确的分割边界。

表8 不同网络在ISPRS Vaihingen数据集上的对比实验

Table 8 Comparative experiment of different networks on ISPRS Vaihingen dataset

Parameter	BiseNetv2	PspNet	Deeplabv3+	OCNet	ISANet	SegNext	AMSNet groups are 4
mIoU /%	59.00	66.82	68.44	66.34	63.37	70.58	**71.91**
mPa /%	70.65	77.73	77.36	76.96	77.51	80.17	**80.92**
mF1 /%	71.45	79.06	80.08	78.65	76.97	81.61	**82.86**

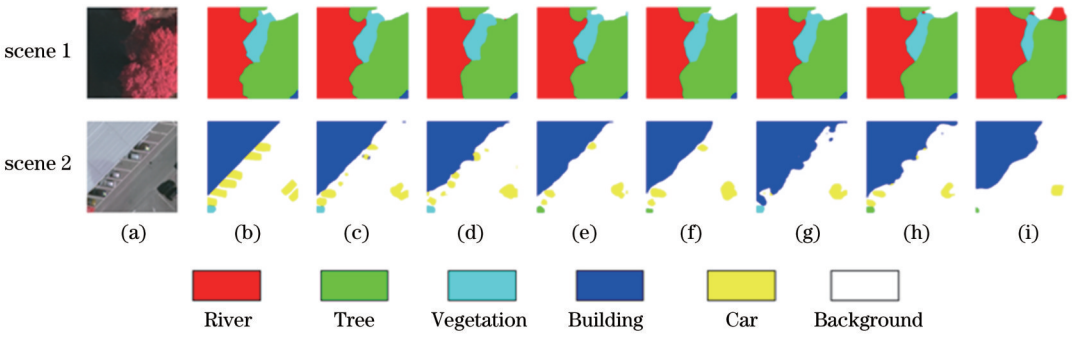

图8 不同网络在数据集ISPRS Vaihingen上的分割效果。(a)原图；(b)标签；(c)AMSNet；(d)SegNext；(e) ISANet；(f) OCNet；(g) Deeplabv3+；(h)PspNet；(i)BiseNetv2

Fig. 8 Segmentation rendering of different networks on ISPRS Vaihingen dataset. (a) Original image; (b) label; (c) AMSNet; (d) SegNext; (e) ISANet; (f) OCNet; (g) Deeplabv3+; (h) PspNet; (i) BiseNetv2

5 结 论

提出了一种基于编码-解码结构的网络模型，命名为AMSNet。在编码部分，采用D_Resnet50作为主干，提取遥感影像的主要特征信息，借助类别引导通道注意力模块降低通道噪声对分割对象的干扰，提高对难分辨区域内目标的分割效果，嵌入特征复用模块弥补特征提取过程中目标边缘损失和小尺度目标丢失的问题。在网络解码部分，设计跨区域特征融合模块融合多层特征并结合多尺度损失融合模块，在不同尺度上计算特征损失，提高网络对多尺度目标的分割效果。在高原区域遥感影像数据集、云层干扰下高原区域遥感影像数据集和公开数据集上进行实验，相比BiseNetv2、PspNet、Deeplabv3+等语义分割网络，所提网络在mIoU、mPa和mF1的评价指标上均取得较好的结果。可视化结果表明，所提网络能有效分割出遥感影像中交错相融难分辨区域的地物目标以及分布零散的多尺度目标，并且在云层干扰下依旧有较好分割效果，具备良好的鲁棒性。虽然所提网络有一定的鲁棒性，在两种气象条件下均能取得较好的分割效果，但推理速度较慢。因此下一步工作是在保证精度不变的前提下降低网络参数量并提高网络推理速度。

参 考 文 献

[1] 陶泽远. 大幅面遥感影像阵地目标检测与识别方法研究[D]. 武汉: 华中科技大学, 2021: 1-152.
Tao Z Y. Research on detection and recognition method of position target in large-format remote sensing image[D]. Wuhan: Huazhong University of Science and Technology, 2021: 1-152.

[2] 陈鑫. 基于可见光遥感图像的典型目标自动检测技术研究[D]. 长春: 中国科学院长春光学精密机械与物理研究所, 2022: 1-121.
Chen X. Research on automatic detection technology of typical targets based on visible light remote sensing images[D]. Changchun: Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, 2022: 1-121.

[3] 代沁伶, 罗斌, 郑晨, 等. 区域多尺度马尔可夫随机场的遥感影像分类[J]. 遥感学报, 2020, 24(3): 245-253.
Dai Q L, Luo B, Zheng C, et al. Regional multiscale Markov random field for remote sensing image classification[J]. Journal of Remote Sensing, 2020, 24(3): 245-253.

[4] 王小鹏, 文昊天, 王伟, 等. 形态学边缘检测和区域生长相结合的遥感图像水体分割[J]. 测绘科学技术学报, 2019, 36(2): 149-154, 160.
Wang X P, Wen H T, Wang W, et al. Water segmentation of remote sensing image using morphological edge detection and region growing[J]. Journal of Geomatics Science and Technology, 2019, 36(2): 149-154, 160.

[5] 杨蕴, 李玉, 赵泉华. 高分辨率全色遥感图像多级阈值分割[J]. 光学精密工程, 2020, 28(10): 2370-2383.
Yang Y, Zhao Q H. Multi-level threshold segmentation of high-resolution panchromatic remote sensing imagery[J]. Optics and Precision Engineering, 2020, 28(10): 2370-2383.

[6] Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.

[7] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation[M]//Navab N, Hornegger J, Wells W M, et al. Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015. Lecture notes in computer science. Cham: Springer, 2015, 9351: 234-241.

[8] 罗松强, 李浩, 陈仁喜. 多尺度特征增强的ResUNet+遥感影像建筑物提取[J]. 激光与光电子学进展, 2022, 59(8): 0828007.
Luo S Q, Li H, Chen R X. Building extraction of remote sensing images using ResUNet+ with enhanced multiscale features[J]. Laser & Optoelectronics Progress, 2022, 59(8): 0828007.

[9] Zhang J, Lin S F, Ding L, et al. Multi-scale context aggregation for semantic segmentation of remote sensing images [J]. Remote Sensing, 2020, 12(4): 701.

[10] Zhao H S, Shi J P, Qi X J, et al. Pyramid scene parsing network [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6230-6239.

[11] Chen L C, Zhu Y K, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision - ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 833-851.

[12] Shang R H, Zhang J Y, Jiao L C, et al. Multi-scale adaptive feature fusion network for semantic segmentation in remote sensing images[J]. Remote Sensing, 2020, 12(5): 872.

[13] Li G, Li L L, Zhu H, et al. Adaptive multiscale deep fusion residual network for remote sensing image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(11): 8506-8521.

[14] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 7132-7141.

[15] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision - ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 3-19.

[16] 张寅, 朱桂熠, 施天俊, 等. 基于特征融合与注意力的遥感图像小目标检测[J]. 光学学报, 2022, 42(24): 2415001.
Zhang Y, Zhu G Y, Shi T J, et al. Small object detection in remote sensing images based on feature fusion and attention[J]. Acta Optica Sinica, 2022, 42(24): 2415001.

[17] 汪亚妮, 汪西莉. 基于注意力和特征融合的遥感图像目标检测模型[J]. 激光与光电子学进展, 2021, 58(2): 0228003.
Wang Y N, Wang X L. Remote sensing image target detection model based on attention and feature fusion[J]. Laser & Optoelectronics Progress, 2021, 58(2): 0228003.

[18] 高慧, 阎晓东, 张衡, 等. 基于Res2Net的多尺度遥感影像海陆分割方法[J]. 光学学报, 2022, 42(18): 1828004.
Gao H, Yan X D, Zhang H, et al. Multi-scale sea-land segmentation method for remote sensing images based on Res2Net[J]. Acta Optica Sinica, 2022, 42(18): 1828004.

[19] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.

[20] Yu F, Koltun V, Funkhouser T. Dilated residual networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 636-644.

[21] Zhang X Y, Zhou X Y, Lin M X, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 6848-6856.

[22] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 936-944.

[23] Yu C Q, Gao C X, Wang J B, et al. BiSeNet V2: bilateral network with guided aggregation for real-time semantic segmentation[J]. International Journal of Computer Vision, 2021, 129(11): 3051-3068.

[24] Ma N N, Zhang X Y, Zheng H T, et al. ShuffleNet V2: practical guidelines for efficient CNN architecture design[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision - ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11218: 122-138.

[25] Liu Z, Mao H Z, Wu C Y, et al. A ConvNet for the 2020s[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 18-24, 2022, New Orleans, LA, USA. New York: IEEE Press, 2022: 11966-11976.

[26] Yuan Y H, Huang L, Guo J Y, et al. OCNet: object context for semantic segmentation[J]. International Journal of Computer Vision, 2021, 129(8): 2375-2398.

[27] Guo M H, Lu C Z, Hou Q B, et al. SegNeXt: rethinking convolutional attention design for semantic segmentation[EB/OL]. (2022-09-18)[2023-03-06]. https://arxiv.org/2209.08575.

Remote Sensing Image Segmentation Based on Attention Guidance and Multi-Feature Fusion

Zhang Yinhui[1], Zhang Feng[1], He Zifen[1*], Yang Xiaogang[2], Lu Ruitao[2], Chen Guangchen[1]

[1]*Faculty of Mechanical and Electrical Engineering, Kunming University of Science and Technology, Kunming* 650500, *Yunnan, China*;

[2]*College of Missile Engineering, Rocket Force Engineering University, Xi'an* 710025, *Shaanxi, China*

Abstract

Objective Remote sensing images have a large detection range, long dynamic monitoring time, and a large amount of carrying information, making the obtained ground feature information more comprehensive and rich. By extracting ground object targets from remote sensing images, more detailed and accurate ground object information in the imaging area can be obtained, providing data support for high-altitude reconnaissance, precision guidance, and terrain matching. However, with the rapid increase in data volume, the current low level of intelligent and automated target extraction methods is difficult to embrace the demand. Traditional image extraction techniques contain edge detection, threshold segmentation, and region segmentation. These methods have good segmentation performance for remote sensing targets with significant contour boundaries but lack the ability of adaptive adjustment while facing complex and ever-changing remote sensing targets. Convolutional neural networks have stronger representation ability, scalability, and robustness than traditional methods by providing multi-level semantic information in images. Due to the uneven distribution, blurred edges, and variable scales of ground objects in remote sensing images, convolutional neural networks are prone to losing edge information and multi-scale feature information during feature extraction. In addition, cloud cover of remote sensing targets in complex scenes exacerbates the loss of target edge and multi-scale information, making it more difficult for convolutional neural networks to accurately segment remote sensing ground objects. In order to solve the above problems, we propose a segmentation method that uses deep residual networks as the backbone and combines attention guidance and multi-feature fusion to enhance the network's ability to segment remote sensing image ground object edges and multi-scale objects.

Methods We propose a remote sensing image semantic segmentation network called AMSNet, which combines attention guidance and multi-feature fusion. In the Encoder Section, D_ Resnet50 is applied as the backbone network to extract the main feature information from remote sensing images, which can enhance the acquisition of detailed information such as edge and small-scale targets in remote sensing images. The category guidance channel attention module is inserted into the backbone to enhance the network's segmentation ability for difficult-to-distinguish and irregularly shaped areas in remote sensing images. A feature reuse module is added to the backbone network to solve the loss of edge detail information and the disappearance of scattered small-scale targets during feature extraction. In the Decoder Section, the cross-regional feature fusion module is applied to fuse the multi-feature information, improving the acquisition of multi-scale target information. Multi-scale loss fusion module is also joined to further enhance the segmentation performance of the network for multi-scale targets.

Results and Discussions From the analysis of experimental results on the remote sensing image dataset of the plateau region and the remote sensing image dataset of the plateau region under cloud interference, compared with other semantic segmentation networks, the proposed network has better segmentation performance (Table 6 and Table 7) regardless of cloud interference. In addition, the segmentation performance is less affected by cloud interference. Even under cloud interference, the segmentation accuracy of ground targets is only 1.10 percentage points lower than that without cloud interference in mIoU, 0.58 percentage points lower than that in mPa, and 0.71 percentage points lower than that in mF1, which is lower than the influence of other semantic segmentation networks on segmentation effect under different cloud meteorological interference conditions. In addition, in order to verify the generalization performance of the AMSNet network segmentation effect, the International Society for Photogrammetry and Remote Sensing (ISPRS) dataset in the Vaihingen region of Germany is selected. In order to better fit the picture size, number of grouping convolutions of feature multiplexing modules in the AMSNet network is reduced to four groups. From the experimental results in Table 8, the network still performs better than other networks. This network is compared with PspNet and OCNet, with mIoU increased by 5.09 percentage points and 5.57 percentage points, Deeplabv3+ network with mIoU by 3.47 percentage points, mPa by 3.56 percentage points, and mF1 by 2.78 percentage points. From the segmenting effect diagram of Fig. 8, this network has a lower error rate, fewer omission, and a more accurate segmenting boundary for building edges and

small-scale cars than other networks.

Conclusions We propose a network model based on encoding-decoding structure—AMSNet. In the encoding part, the D_Resnet50 network is applied as the backbone to extract the main feature information of remote sensing images. We also use a category-guided channel attention module to reduce the interference of channel noise on segmented objects and improve the segmentation effect of targets in difficult-to-distinguish areas. We embed a feature reuse module to compensate for the problem of target edge loss and small-scale target loss during the feature extraction process. In the decoding part, the cross-regional feature fusion module is designed to integrate multi-layer features and combine the multi-scale loss fusion module to calculate the feature loss at different scales to improve the segmentation effect of the network on multi-scale targets. This network conducts experiments on the remote sensing image dataset of the plateau region, remote sensing image dataset of the plateau region under cloud interference, and a public dataset. Compared with semantic segmentation networks such as BiseNetv2, PspNet, and Deeplabv3+, the proposed network achieves better results in the evaluation indicators of mIoU, mPa, and mF1. The visualization results show that the proposed network can effectively segment the ground object targets and scattered multi-scale targets in the interlaced and hard-to-distinguish areas in the remote sensing images, and it has good segmentation performance and good robustness in cloud interference.

Key words remote sensing image; semantic segmentation; attention mechanism; multi-scale feature