引 文 格 式 ： 谢勇. 基于 GQA-AssemFormer 的遥感影像道路提取方法[J]. 地理空间信息, 2026, 24(7): 26-32. doi:10. 3969/j.issn. 1672-4623.2026.07.006

基于 GQA-AssemFormer 的遥感影像道路提取方法

谢　勇
（四川省交通勘察设计研究院有限公司，四川 成都 610000）

摘　要：针对当前模型提取道路结构时，受复杂地物影响存在的道路结构断裂消失、背景误提取等问题，提出了一种多尺度局部细节特征同步编码的遥感影像道路提取方法。通过分组查询—内嵌卷积视觉转换器（GQA-AssemFormer）搭建编码器网络，同步编码各尺度特征图中的细节与全局上下文特征；利用高效上采样（EUB）单元对深层特征图精细上采样，并通过混排注意力（SA）单元对采样后特征图赋权过滤；以广义骰子系数损失与自适应阈值焦点损失组合为损失函数，通过偏执学习道路样本，优化模型对复杂道路边缘的分割精度。结果表明，所构建模型在 DeepGlobe Road、Massachusetts Road 数据集上的 F_1-score 分别达到 90.27%与 90.85%，较主流模型视觉窗口转换 U 型网络提高了 7.14%与 7.38%，能在各种复杂场景下准确提取各类道路结构，具有良好的提取精度与泛化能力。

关键词：遥感影像；道路提取；GQA-AssemFormer；EUB；SA；优化损失函数
中图分类号：P237　　**文献标志码：**B　　**文章编号：**1672-4623（2026）07-0026-07

Road Extraction Method from Remote Sensing Images Based on GQA-AssemFormer

XIE Yong
(Sichuan Transportation Survey and Design Institute Co., Ltd., Chengdu 610000, China)

Abstract: In view of the problems of road structure breakage and disappearance, background mis-extraction caused by complex ground objects when extracting road structure by the current models, we proposed a remote sensing image road extraction method with synchronous encoding of multi-scale local detail features. The encoder network was built with grouped query attention-assembly-based convolutional vision transformer (GQA-AssemFormer) to synchronously encode the details and global context features in the feature maps of each scale. The efficient upsample block (EUB) unit was used to finely upsample the deep feature map, and the shuffle attention (SA) unit was used to weight and filter the sampled feature map. Taking the combination of generalized DiceLoss and adaptive threshold FocalLoss as the loss function, the model is optimized for the segmentation accuracy of complex road edges through paranoid learning of road samples. Experimental results show that the constructed model has an F_1-score of 90.27% and 90.85% on DeepGlobe Road and Massachusetts Road datasets, which are 7.14% and 7.38% higher than the mainstream model shift window transformer U-shape network. This method can accurately extract road structures of various complexities and has good extraction accuracy and generalization ability.

Key words: remote sensing image, road extraction, GQA-AssemFormer, EUB, SA, optimized loss function

随着遥感技术的快速发展，高分辨率光学影像已成为城市交通规划、灾害应急响应、智慧农业等领域的重要数据源[1-2]。基于遥感影像的道路提取，对于城市规划、导航服务和灾害救援路径优化具有重要意义[3-4]。光学遥感影像中的道路形态复杂多样，如何在大尺度、背景复杂的遥感影像中，完整提取道路结构，具有重要的研究价值。早期研究多采用边缘、纹理等模板匹配方法[5]或基于经典机器学习模型的像素分类方法[6]实现道路结构与背景的分割，在简单场景中表现尚可，但其特征表达能力有限，难以应对复杂背景与形态多变的道路结构[7]。近年来深度学习技术凭借其强大的特征自适应性，逐渐成为道路提取的主流方法[8]。以卷积神经网络（CNN）为基础的模型通过局部卷积操作提取道路边缘与纹理特征，而基于视觉变换器（ViT）的模型则通过全局自注意力机制提升长距离建模能力，进一步提升复杂道路网络的完整性[9-11]。目前基于遥感影像的道路提取结果主要存在断裂、毛刺、拓扑错误等问题，其主要原因可归纳为数据和模型两个层面。在数据层面，随着遥感影像空间分辨率的不断提高，其内部地物目标细节特征愈发丰富，提取任务易受空间层面的阴影遮挡、语义层面的类道路结构以及同物异谱、同谱异物等问题干扰，导

项目来源：四川省交通勘察设计研究院有限公司科研资助项目（242023009）。
作者简介：谢勇（1985—），本科，高级工程师，主要从事公路、水路等交通勘察测绘工作，E-mail：YYXz235@126.com。

致模型难以光滑连续的提取完整道路结构；此外，遥感影像内道路结构的像素占比远低于无关背景样本，导致现有方法难以充分学习道路结构特征。在模型层面，现有模型主要基于 CNN 与 ViT 单元构建，标准 CNN 模型虽能有效捕获局部特征，但其上下文感受野有限，对狭长形态的道路结构上下文建模不足，导致所提取道路的连通性较差；ViT 模型通过全局注意力提升长距离建模能力，但参数量庞大，且对图像预处理的分块机制还会导致细节特征难以被连贯学习。针对上述问题，本文提出了一种基于局部—全局特征联合编解码的遥感影像道路提取模型。基于内嵌卷积视觉转换器（AssemFormer）搭建特征编码器网络，利用卷积核的局部特征提取能力与 ViT 单元的全局上下文建模能力同步捕获道路细节纹理与全局上下文关联特征。在 ViT 建模阶段，以分组查询注意力（GQA）为注意力计算头，减少键值存储带来的额外计算开销。在特征解码阶段，利用高效上采样模块（EUB）对各级特征图进行精细上采样，同时结合通道—空间双域的混排注意力（SA）赋予特征图注意力权重，抑制无关背景对提取过程的干扰。在训练阶段，采用广义骰子系数损失（GDL）与自适应阈值焦点损失（ATFL）组合作为模型损失函数，通过类别平衡权重与动态聚焦难分类样本策略，优化模糊道路边缘的定位精度。

1　模型结构概述

1.1　GQA–AssemFormer 编码器

为提取不同尺度下的道路细节与全局上下文特征，通过 AssemFormer 搭建特征编码网络。AssemFormer 单元封装了尺寸为 3×3 、1×1 的标准卷积核与 ViT 单元，并在内部通过打包（Pack）、解包（Unpack）操作实施特征图与特征序列之间的双向转换。Pack 操作层通过补丁拆分与线性投影嵌入的方式将特征图拉伸为特征序列；Unpack 操作层通过逆向映射操作，将赋予全局上下文注意力的特征序列还原为图像。通过 Pack–Unpack 操作实现特征图与序列的转换，卷积核与 ViT 单元的协同编码。为控制计算效率，在 AssemFormer 的 ViT 单元采用 GQA 计算全局注意力。作为融合多查询注意力与多头注意力优点的多头注意力机制，GQA 将全部查询头分为 G 组，每组包含 H/G 个查询头（H 为注意力头的数量），各组内采用共享键（K）和值（V）的策略独立计算注意力，最后拼接各组所得注意力，作为输出结果。

$$\begin{cases} Q_g = X \times \boldsymbol{W}_g^Q \\ K_g = X \times \boldsymbol{W}_g^K \\ V_g = X \times \boldsymbol{W}_g^V \end{cases} \tag{1}$$

式中，X 为输入特征，Q_g 为第 $g \in [1, G]$ 组对应的查询；K_g 与 V_g 为组内共享键值；$\boldsymbol{W}_g^Q$ 、$\boldsymbol{W}_g^Q$ 、$\boldsymbol{W}_g^Q$ 为查询、键、值的投影矩阵。

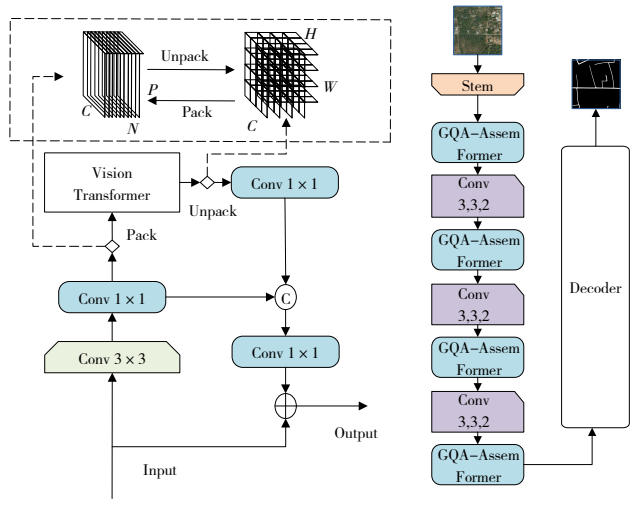
在组内计算注意力，再以通道拼接的形式合并各组结果，即

$$\mathrm{Att}(Q_g, K_g, V_g) = \mathrm{Softmax}\left(\frac{Q_g \boldsymbol{K}_g^{\mathrm{T}}}{\sqrt{d}}\right) V_g \tag{2}$$

$$\mathrm{Att} = \mathrm{Concat}(\mathrm{Att}_1, \mathrm{Att}_2, \cdots, \mathrm{Att}_n)\boldsymbol{W}^o \tag{3}$$

式中，$\mathrm{Att}(Q_g, K_g, V_g)$ 为第 g 组内所得注意力；d 为键的维度；$\boldsymbol{K}_g^{\mathrm{T}}$ 为 $\boldsymbol{K}^{\mathrm{T}}$ 的转置；Softmax 为归一化操作，Concat 为通道拼接；$\boldsymbol{W}^o$ 为对应的权重矩阵。

基于 GQA 构建的 AssemFormer 结构见图 1a。现有 ViT 架构的编码器通常在网络前端采用结合补丁拆分与线性投影嵌入的方式完成图像的序列化处理，而 GQA–AssemFormer 则通过内嵌的 Pack–Unpack 机制逐层实现特征图与序列的相互转换，因此编码网络的头部改为无损下采样的干细胞层，层间以 2 倍步长卷积的方式进行下采样，编码器完整结构见图 1b。



a GQA–AssemFormer 单元结构图　　b 编码器结构图

图 1　重构编码器结构示意图

1.2　多尺度特征过滤解码网络

为还原特征图平面尺寸，目前解码器多以双线性插值为基本上采样路径并增加多尺度特征图跨层拼接的方式，引导特征图重建道路的细节与全局特征；但简单的图像上采样对内部细节特征的恢复能力有限，导致特征图中仍存在各类背景噪声，不利于道路结构

的准确分割。针对该问题，本文构建了带有特征精细上采样与负样本过滤机制的特征解码器网络。解码器网络通过EUB机制构建特征多级上采样路径，同时融合编码器输出的多层特征图，提高图内特征信息的丰富程度。EUB单元由双线性插值上采样、标准卷积核、批量归一化—线性整流函数模块组成，在上采样的基础上，通过卷积与非线性激活操作，对特征进行强化处理。对于各解码器层得到的融合特征图，通过SA单元二次赋权，过滤特征图内的无关背景特征。SA单元同时在平面、通道两个维度计算注意力权重。SA单元沿通道切分将特征分为 G 组，对于任意一组子特征 $X_k, k\in(1,G)$，沿通道二次分割获得 X_{k1} 与 X_{k2}。在 X_{k1} 的基础上计算通道注意力，首先通过全局平均池化（GAP）将特征拉平；再通过由Sigmoid激活函数与全连接层组成的门控机制，在通道注意力的基础上生成通道注意力图，即

$$F_{gp}(X_{k1})=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}X_{k1}(i,j) \tag{4}$$

$$X_{k1}^{`}=\sigma(F_c(s))X_{k1} \tag{5}$$

式中，F_{gp} 为GAP计算；s 为通道注意力；$C\times H\times W$ 为特征图尺寸；σ 为Sigmoid激活函数；F_c 为全连接层组成的门控机制；$X_{k1}^{`}$ 为通道注意力图。

对于 X_{k2}，采用组归一化计算空间注意力，再通过门控机制得到空间注意力图，即

$$X_{k2}^{`}=\sigma(F_c(GN(X_{k2})))X_{k2} \tag{6}$$

式中，$GN()$ 为组归一化计算；$X_{k2}^{`}$ 为空间注意力图。

在每组赋权特征末端，采用1×1卷积核完成通道归一化，再将其与更浅层编码器输出的特征图进行拼接。

1.3 损失函数

对于大尺度遥感影像背景中的道路结构，其像素占比与背景像素差异显著，在无任何偏置引导的情况下，易使模型更偏向学习背景像素特征，影响最终的道路提取精度。针对上述问题，本文采用GDL与AT-FL的组合作为损失函数。

$$L_{\mathrm{GDL}}=1-\frac{2\times\sum_{m=1}^{M}W_l\sum_{n=1}^{N}r_n p_n}{\sum_{m=1}^{M}W_c(\sum_{n=1}^{N}r_n+\sum_{n=1}^{N}p_n)+\varpi} \tag{7}$$

$$L_{\mathrm{A}}=\begin{cases}-\alpha_t(\lambda-p_t)^{\mu}\log(p_t), & p_t\leqslant\theta\\ -\alpha_t(\lambda-p_t)^{v}\log(p_t), & p_t>\theta\end{cases} \tag{8}$$

式中，p_n、r_n 为各像素的真实、预测标签；$W_l=\frac{1}{\sum_{n=1}^{N}r_n^2}$ 为类别均衡权重参数；$M=2$ 为样本数量（包含背景像素）；N 为像素总数量；ϖ 为确保分母不为零的平衡参数；p_t 为模型预测正样本的概率；α_t 为平衡正负样本贡献的权重系数；θ 为用于区分难/易样本的阈值，$p_t<\theta$ 表示样本难以被分类；λ 为对简单样本的惩罚因子；μ、v 分别为控制易分类样本和难分类样本的调节因子。

GDL通过引入 W_l 对道路与背景样本加权，引导道路、背景像素在损失计算过程中的权重，使模型在训练阶段更加关注道路像素的预测准确性。L_{A} 通过自适应阈值和动态权重调整，为不同分类难度的像素分配不同的权重，增加模型对阴影遮挡、相似线状结构等区域像素的关注程度，提高模型对复杂场景中道路结构的提取性能。

2 实验结果与分析

2.1 实验设计

1）实验数据集。为验证本文模型的有效性，在遥感影像道路提取数据集DeepGlobe Road[10]和Massachusetts Road[11]上开展模型训练与性能验证实验。DeepGlobe Road数据集包含东南亚地区城市、乡村等环境的8 970张尺寸为1 024像素×1 024像素，空间分辨率为0.5 m的卫星遥感影像；Massachusetts Road数据集涵盖了美国马萨诸塞州不同区块的卫星遥感影像1 171张，尺寸为1 500像素×1 500像素。

2）数据集预处理。为完成模型的训练测试，按照8∶1∶1的比例将两组公开数据集分为相互独立的训练集、验证集、测试集。对于训练数据集内的样本，采用随机水平翻转、随机垂直翻转、随机角度旋转等影像增强措施提高影像样本的多样性，使训练后的模型能在不同道路结构和环境条件下均有效。

3）模型训练。实验均采用搭载2块NVIDIA GeForce RTX 3090显卡的服务器完成，设备操作系统为Ubuntu 20.04，中央处理器为Intel Xeon Platinum 8255C，服务器内存为52 GB，硬盘容量为1 TB。模型训练过程中，采用权重自适应矩估计算法优化模型参数，设置初始动量参数为0.935，权重衰减系数为0.001，单批次送入模型的训练样本数量为20个，模型总迭代次数为150次，初始学习率为0.02。

4）评价指标。本文选取精确率（P）、召回率（R）、总体精度（OA）、F_1-score、平均交并比（mIoU）作为实验结果的评价指标。

2.2　实验结果分析

2.2.1　多模型对比实验

为验证本文模型的实际表现，以UNet[12]、第三代U型网络（UNet3+）[13]、稠密连接网络（D-LinkNet）[14]、单分支卷积—转换器网络（SCTNet）[15]、Swin-UNet[16]、多尺度交换网络（MINet）[17]作为对照组，分别在两组数据集上进行模型训练和对比实验。实验结果见表1，可以看出，本文模型在两组数据集上的F_1-score、*OA*、mIoU分别达到90.27%、99.35%、85.24%以及90.85%、99.39%、85.65%，表现趋于一致，说明在不同环境下本文模型均能保持精度稳定。相较于对照模型，本文模型在DeepGlobe Road数据集上的F_1-score分别提高19.75%、15.86%、11.62%、8.43%、7.14%、4.92%；在Massachusetts Road数据集上的F_1-score分别提高19.21%、15.09%、11.03%、8.51%、7.38%、4.59%；本文模型的*OA*与mIoU也较对照模型有不同程度的提高，说明本文模型在遥感影像道路结构提取任务上的精度表现优于当前主流模型。本文模型与对照模型提取结果见图2，可以看出，对照模型提取的道路结构均出现了不同程度的道路结构断裂、部分消失、非道路结构错误划分等问题，而本文模型提取的道路结构较为完整，在不同模型反复出现提取错误的区域也能较完整、正确地分割出道路结构，进一步证明了本文模型在道路提取精度方面的优势。

表1　不同模型道路提取结果对比

数据集	模型	*P*	*R*	F_1-score	*OA*	mIoU
DeepGlobe Road	UNet	72.24	68.88	70.52	97.23	59.74
	UNet3+	76.18	72.72	74.41	97.76	64.23
	D-LinkNet	79.44	77.88	78.65	98.35	66.85
	SCTNet	81.15	81.93	81.54	98.83	72.29
	Swin-UNet	82.61	83.66	83.13	98.94	73.75
	MINet	84.34	86.38	85.35	99.12	78.86
	本文模型	89.23	91.33	90.27	99.35	85.24
Massachusetts Road	UNet	74.51	68.98	71.64	97.45	60.35
	UNet3+	77.37	74.22	75.76	97.89	64.87
	D-LinkNet	80.34	79.31	79.82	98.47	67.32
	SCTNet	82.61	82.07	82.34	98.94	73.85
	Swin-UNet	83.34	82.92	83.47	98.97	74.24
	MINet	85.75	86.78	86.26	99.18	79.23
	本文模型	89.78	91.95	90.85	99.39	85.65

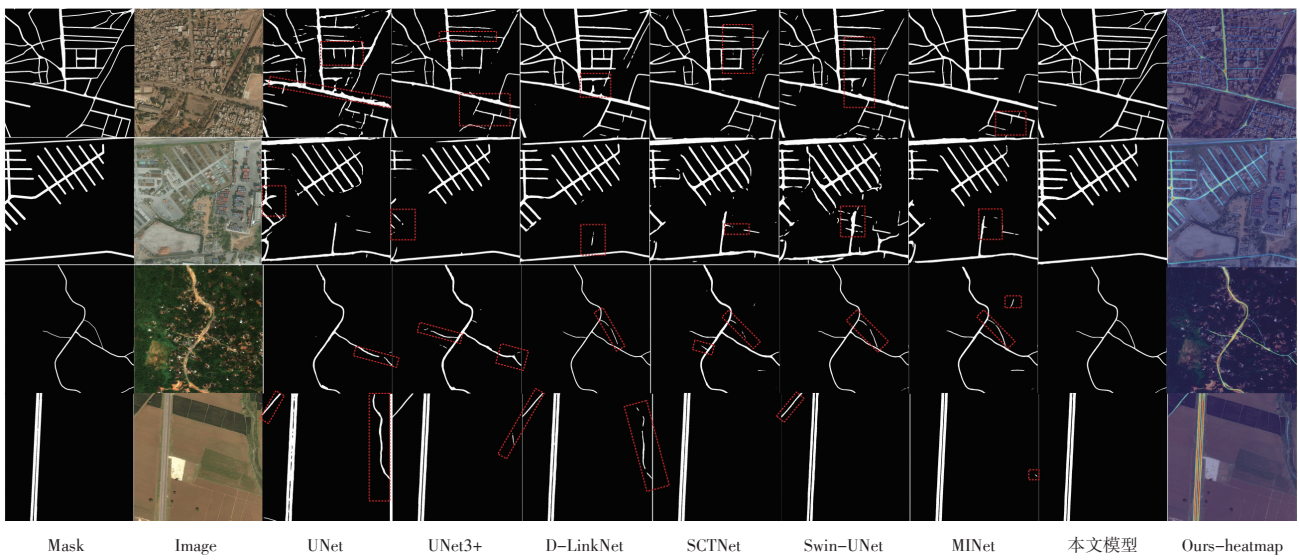


图2　不同模型道路提取结果对比图

2.2.2　模型结构验证

为验证以GQA-AssemFormer单元搭建编码器的实际性能表现，分别将语义分割转换器（SETR）[18]的编码器部分、Swin Transformer所搭建的编码器、GQA-AssemFormer搭建的编码器应用于本文模型中，在DeepGlobe Road数据集上进行编码器性能对比实验。结果表明（图3），SETR-Encoder、Swin Transformer、GQA-AssemFormer的F_1-score分别为85.42%、87.81%、90.27%，mIoU分别为78.93%、81.38%、85.24%；说明以GQA-AssemFormer搭建编码器能通过捕获不同尺度下的目标细节与全局特征，提高模型对道路结构的提取精度。

为验证编码器网络中EUB与SA注意力权重特征过滤组合对特征图内目标信息的还原效果，分别将UNet网络解码器、仅有EUB模块的解码器、仅有SA注意力单元的解码器、EUB-SA解码器应用于模型中开展对比实验。结果表明（图4），UNet、EUB-Only、SA-Only、EUB-SA解码器的F_1-score分别为73.79%、83.74%、81.89%、90.27%，mIoU分别为59.74%、74.26%、72.85%、85.24%；在构建解码器阶段仅使用

EUB或SA单元所得模型的F_1-score略低于二者组合效果，但仍优于UNet网络的解码器，说明在特征解码阶段引入精细化上采样与特征二次过滤，能有效提高模型的提取精度。

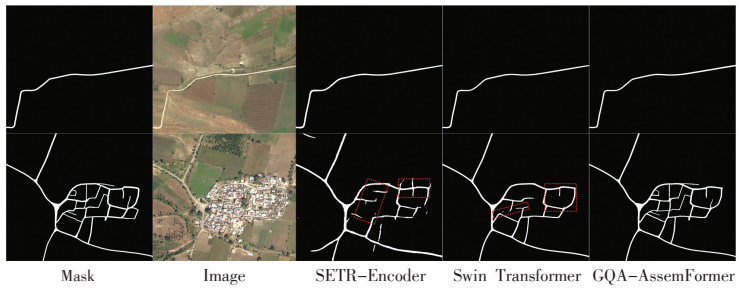


图3 不同编码网络提取结果对比

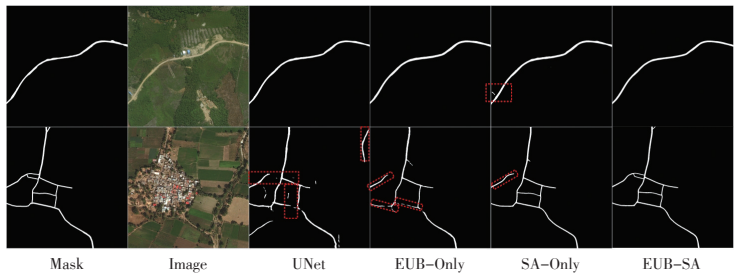


图4 不同解码网络提取结果对比

损失函数在训练过程中逐步计算各训练轮次的提取结果与真实掩膜之间的差异，其在训练阶段的表现对于训练效果与模型性能表现具有直接影响。为验证本文模型中的损失函数组合的性能，分别以Dice-BCE、Dice-ATFL、GDL-BCE、GDL-ATFL为组合，训练本文模型，开展对比实验。结果表明（图5），Dice-BCE、Dice-ATFL、GDL-BCE、GDL-ATFL的F_1-score分别为86.74%、88.29%、89.25%、90.27%，mIoU分别为81.13%、82.75%、83.87%、85.24%；验证了GDL、ATFL两组损失函数，在提升道路提取准确性方面的有效性。

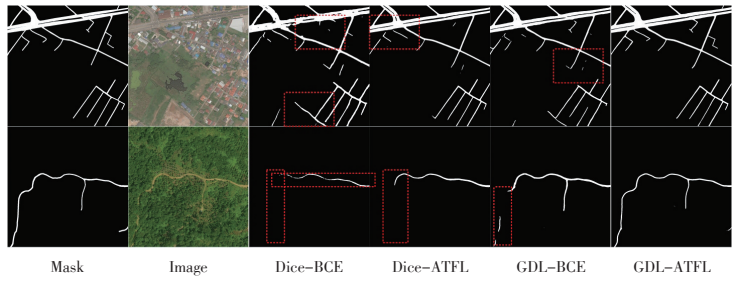


图5 不同损失函数模型提取结果对比

为确定GQA中不同共享计算头分组数以及SA单元中不同通道分组数的最优设置，参考已有文献[19-20]，调整GQA计算头分组数$G_1 \in [8,16,32,64]$，与SA单元的通道分组数$G_2 \in [8,16,32,64]$，基于DeepGlobe Road数据集开展梯度对比实验。实验结果见表2，随着G_1的增加，对应模型的F_1-score同步上涨，$G_1=32$时模型的F_1-score达到最优；随着G_2的增加，模型的F_1-score相应提高，$G_2=64$时模型的F_1-score达到最高。通过上述实验可确定GQA与SA单元中分组数的最优配置。

表2 GQA和SA不同分组参数设置对比实验

方法	分组数	F_1-score/%	mIoU%
GQA(G_1)	8	88.21	82.35
	16	88.56	82.87
	32	90.27	85.24
	64	89.74	84.39
SA(G_2)	8	88.62	82.73
	16	89.17	83.58
	32	89.53	84.16
	64	90.27	85.24

2.2.3 模型结构消融实验

为验证本文模型中不同结构对于提升道路提取精度的具体贡献，以Swin-UNet为基准模型，分别在两组数据集上开展消融实验。将GQA-AssemFormer编码器网络、EUB-SA组合解码器网络、GDL-ATFL损失函

数3组结构单独应用于基准模型上，构建消融模型1~3，将3组优化结构以不同组合形式应用于基准模型上，构建消融模型4~6，实验结果见表3。基准模型和各消融模型的检测结果见图6。

表3　消融实验结果

模型	Ablations			DeepGlobe Road		Massachusetts Road	
	GQA-AssemFormer	EUB-SA	GDL-ATFL	F_1-score/%	mIoU/%	F_1-score/%	mIoU/%
基准	–	–	–	83.13	73.75	83.47	74.24
1	√	–	–	87.41	81.25	87.53	81.87
2	–	√	–	85.35	78.86	85.64	79.14
3	–	–	√	84.63	76.32	84.97	76.73
4	√	√	–	89.56	84.65	89.79	85.14
5	–	√	√	86.45	79.12	86.52	79.48
6	√	–	√	88.36	83.65	88.93	84.27

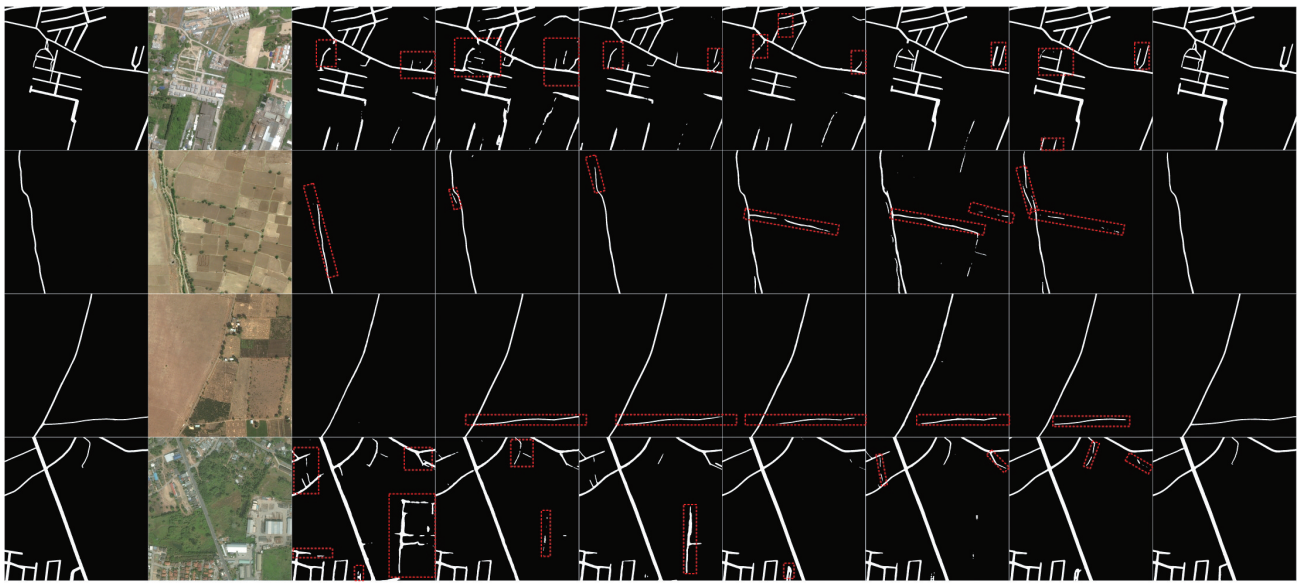


图6　基准模型与消融模型检测结果

基准模型在两组数据集上的F_1-score达到了83.13%与83.47%；替换编码器网络后，消融模型1的F_1-score分别提升至87.41%与87.53%。观察模型结构可知，以GQA-AssemFormer搭建的编码器网络，在各尺度特征图上先后进行卷积与全局自注意力计算，卷积层直接捕捉像素级道路边缘、材质、纹理等局部特征，自注意力机制则建模道路像素间跨越大区域、长距离的空间依赖关系，理解道路网络的拓扑连通性，同时借助单元内部特征图向卷积、GQA-AssemFormer的顺序传导路径与通道拼接机制，实现了细节纹理与全局上文的关联，能使模型更好地建立像素间跨区域的全局关联，从而更好地预判断裂道路的合理连接方式，使得消融模型1在提取精度上优于基准模型。采用EUB-SA组合解码器后，消融模型2的F_1-score提升至85.35%与85.64%。EUB-SA一方面通过精准上采样，充分恢复了特征图内的道路细节信息，避免传统插值上采样过程引入无关背景噪声；另一方面基于SA机制进行注意力建模，使模型更侧重学习道路结构特征并抑制无关背景噪声，能更准确地捕捉道路边界细节，减少对非道路结构的错误划分。以GDL-ATFL替换现有损失函数，消融模型3的F_1-score达到84.63%与84.97%。GDL通过像素级类别加权，确保模型关注正确预测的道路像素，有效抑制模型学习无关背景的倾向，而ATFL在此基础上引入自适应阈值动态区分难易样本，进一步优化模型对模糊或缺失道路的预测效果。GDL-ATFL组合在像素级分类与对象级相似度任务上实施双重保障，使得消融模型3在提取细小、复杂或被干扰的复杂道路时，提升了对道路结构和边缘提取的准确性。组合使用上述改进措施后，对应消融模型的提取精度均较基准模型有一定的提升。

3 结 语

为准确提取遥感影像中的道路结构，构建全局—局部特征联合编解码的遥感影像道路提取模型。以GQA-AssemFormer单元搭建编码器网络，在各尺度特征图上同步编码全局与细节特征；在解码器网络中利用EUB单元对特征图精细上采样，并采用SA单元通过空间与通道维度注意力进一步过滤特征图内无关背景特征；使用GDL-ATFL组合损失函数引导模型在训练期间关注学习道路样本和难分类区域的特征分布。实验结果表明，本文模型在DeepGlobe Road上的F_1-score达到90.27%，较对照模型分别提升了19.75%、15.86%、11.62%、8.43%、7.14%、4.92%；在Massachusetts Road数据集上的F_1-score达到90.85%，较对照模型分别提高了19.21%、15.09%、11.03%、8.51%、7.38%、4.59%。在不同场景下，本文模型均能更完整地提取复杂地表环境下的道路结构，结果与真实掩膜更接近，表现出良好的提取精度与泛化水平。后续研究将进一步优化模型结构，在增强提取精度的同时，压缩网络参数和训练拟合所用的时间。

参考文献

[1] 陶从辉,高青山,赵梦琳. 改进U-Net网络的高分辨率遥感影像建筑物提取方法[J]. 地理空间信息,2025,23(1):9-12

[2] 张京奎,卢桂洋,陈曙光,等. 结合超分辨率重建技术的遥感语义分割方法[J]. 地理空间信息,2025,23(1):29-33

[3] Mo S,Shi Y,Yuan Q,et al. A Survey of Deep Learning Road Extraction Algorithms Using High-resolution Remote Sensing Images[J]. Sensors,2024,24(5):1708

[4] Sharma P,Kumar R,Gupta M,et al. A Critical Analysis of Road Network Extraction Using Remote Sensing Images with Deep Learning[J]. Spatial Information Research,2024,32(4):485-495

[5] 严炳杨,冯志荣,鲍灵辉,等. DAR-Net:双空间注意力与边界细化的遥感影像滑坡提取网络[J]. 地理空间信息,2025,23(5):1-4

[6] Li J,Cai Y,Li Q,et al. A Review of Remote Sensing Image Segmentation by Deep Learning Methods[J]. International Journal of Digital Earth,2024,17(1):2328827

[7] 赵斌,王乾. 基于深度特征聚合网络的高分辨率遥感图像语义分割[J]. 地理空间信息,2025,23(1):80-84

[8] 王子明. 基于注意力机制的遥感影像道路提取方法[J]. 地理空间信息,2025,23(1):13-16

[9] 陈涛,葛小三. 融合道路特征的动态卷积遥感影像道路提取网络模型[J]. 地理与地理信息科学,2025,41(2):23-30

[10] Zhou H,He H,Xu L,et al. A Comparative Study of Deep Learning Methods for Automated Road Network Extraction from High-spatial-resolution Remotely Sensed Imagery[J]. Photogrammetric Engineering and Remote Sensing,2025,91(3):163-174

[11] Sharma P,Kumar R,Gupta M,et al. A Critical Analysis of Road Network Extraction Using Remote Sensing Images with Deep Learning[J]. Spatial Information Research,2024,32(4):485-495

[12] Liu X,Gao P,Yu T,et al. CSWin-UNet: Transformer UNet with Cross-shaped Windows for Medical Image Segmentation[J]. Information Fusion,2025,113:102634

[13] Ramya D,Lakshmi C. Segnet with UNet3+ and EfficientNet: A Novel Framework of Brain Tumour Segmentation and Classification Model by Multiscale Attention-based Deep Learning Techniques with Hybrid Heuristic Improvement Using 3D MRI Brain Images[J]. The Imaging Science Journal,2025,73(1):46-69

[14] Wu H,Bai J,Li J,et al. The Relationships Between Vegetation Changes and Groundwater Table Depths for Woody Plants in the Sangong River Basin, Northwest China[J]. Remote Sensing,2025,17(5):937

[15] Xu Z,Wu D,Yu C,et al. SCTNet: Single-branch CNN with Transformer Semantic Information for Real-time Segmentation[C]//Proceedings of the AAAI Conference on Artificial Intelligence,2024:6378-6386

[16] Cao H,Wang Y,Chen J,et al. Swin-UNet: UNet-like Pure Transformer for Medical Image Segmentation[C]. European Conference on Computer Vision, Cham, Springer Nature Switzerland,2022:205-218

[17] Li X,Song J,Jiao W,et al. MINet: Multi-scale Input Network for Fundus Microvascular Segmentation[J]. Computers in Biology and Medicine,2023,154:106608

[18] Zheng S,Lu J,Zhao H,et al. Rethinking Semantic Segmentation from a Sequence-to-sequence Perspective with Transformers[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2021:6881-6890

[19] Ainslie J,Lee-thorp J,de Jong M,et al. GQA: Training Generalized Multi-query Transformer Models from Multi-head Checkpoints[C]. The Conference on Empirical Methods in Natural Language Processing,2023

[20] Zhang Q L,Yang Y B. SA-Net: Shuffle Attention for Deep Convolutional Neural Networks[C]//ICASSP IEEE International Conference on Acoustics,Speech and Signal Processing,2021:2235-2239