多注意力机制在遥感图像语义分割中的应用

张垚杰 [1]
ZHANG Yaojie

摘 要 为解决遥感图像语义分割中传统注意力机制资源消耗过大的问题，优化分割效果与资源利用效率。文章提出一种创新的多注意力模块，将线性空间注意力与通道注意力相结合，利用多注意力模块构建解码器来完成分割任务。实验表明，所构建的多注意力模块有效提升了遥感图像语义分割的精度，实现了资源效率与分割效果的双重优化。多注意力机制能够在保证资源高效利用的前提下提升分割效果，为后续研究提供了一种可行的注意力机制应用方案。

关键词 深度学习；语义分割；注意力机制；遥感影像

doi：10.3969/j.issn.1672-9528.2025.03.015

0 引言

遥感图像语义分割任务是利用遥感技术获取地表信息，并通过图像处理和深度学习等技术对遥感图像进行分析，目的是将图像中的每个像素分配到特定的类别中。与传统的图像分类不同，语义分割不仅关注整体图像的类别，而且关注图像中每个像素所对应的具体类别，从而实现更为精细的分析。通过对遥感图像的语义分割，可以有效识别城市、森林、农业用地等不同土地利用类型，帮助监测土地利用变化，支持可持续发展，在土地资源管理、参量估算和经济评估等应用中起着至关重要的作用。

近年来，注意力机制逐渐融入到遥感图像的分割网络中，但普遍存在资源消耗大的问题。为此，本文设计使用线性空间注意力和通道注意力结合的多注意力模块来保持资源利用和图像分割的高效性。具体来说，利用 ResNet101 作为骨干网络来获取上下文信息，然后使用多注意力模块来高效地提取特征。最后在 GID 数据集上进行实验验证，实验结果表明，与遥感图像分割任务中常用的方法相比，分割效果有明显提升，取得了 90.95% 的像素准确率，83.67% 的平均像素准确率，76.72% 的平均交并比。

1 相关工作

遥感图像语义分割任务主要存在以下挑战：首先，遥感图像通常具有高分辨率和大范围的特点，处理和分析这些大规模数据需要强大的计算能力和存储资源。其次，遥感图像中的物体类别往往多样且复杂，如城市建筑、道路、农田、水体等，导致分割任务中的类间相似性增加。最后，在实际场景中，物体可能被遮挡，或像素可能包含多个类别（混合像素），这使得准确分割更加困难。这些挑战使得遥感图像语义分割任务在技术和应用层面都具有很高的复杂性和研究价值。

为了解决上述问题，研究者进行了多项工作。早期的遥感图像分割方法主要基于像素的光谱特征和空间特征。常见的技术包括基于阈值的方法、区域生长、分水岭算法等。这些方法通常依赖于人工设计的特征，处理速度较快，但在复杂场景下的表现较差。随着机器学习的兴起，基于特征提取和分类器的分割方法逐渐被提出。例如，支持向量机（SVM）、随机森林等方法被广泛应用于遥感图像的分类任务。这些方法在一定程度上提高了分割的精度，但仍然受到特征选择和模型泛化能力的限制。深度学习的兴起为遥感图像的语义分割提供了新的解决方案 [1-2]。其中卷积神经网络（CNN）通过自动学习图像特征，在语义分割中表现出色。近年来，许多基于深度学习的分割网络被提出，其中最具代表性的是全卷积网络（FCN）、U-Net、DeepLab 等 [3-6]。全卷积网络（FCN）通过将全连接层替换为卷积层，使得能够进行端到端的像素级预测。U-Net 最初用于医学图像分割，但其对称的编码 - 解码结构也适用于遥感图像分割，通过跳跃连接将高分辨率特征与低分辨率特征结合，提高了分割精度。DeepLab 系列模型采用空洞卷积（dilated convolution）和条件随机场（CRF）后处理，有效捕捉多尺度信息，进一步提升了分割效果。最近，注意力机制逐渐被引入遥感图像分割中 [7]。注意力机制能够帮助模型聚焦于重要特征，提高分割精度，但缺点是资源消耗巨大。

1. 长治学院计算机系 山西长治 046000
[基金项目] 长治学院校级科研项目（020/XN0580）

2 多注意力分割网络

2.1 线性空间注意力

空间注意力机制（spatial attention mechanism）是深度学习中用于计算机视觉任务的一种技术，其核心思想是让模型能够聚焦于输入图像或特征图中的关键空间区域。空间注意力机制通过学习图像空间上的重要区域，增强了模型在捕捉关键特征方面的能力，通过增强与选定位置或物体相关的感知处理并抑制与非选定刺激相关的处理来促进环境中的自适应交互。常见的空间注意力机制基于点积注意力，如图 1 所示。尽管点积注意力有效，但它具有巨大的计算和内存成本，尤其是在大规模输入的情况下。点积注意力的内存和计算成本随输入大小呈二次方增长，限制了其在高分辨率图像和长序列等大规模场景中的应用。下面介绍由点积注意力推广得到的线性空间注意力（linear spatial attention mechanism, LSAM）。

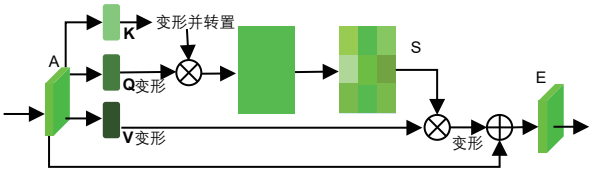

图 1 点积注意力

假设 N 和 C 分别表示输入序列的长度和输入通道的数量，其中 $N=H\times W$，H 和 W 分别表示输入的高度和宽度。给定一个特征 $\boldsymbol{X}=[x_1,\cdots,x_n]\in\mathbf{R}^{N\times C}$，点积注意力使用 3 个投影矩阵 $\boldsymbol{W}_q\in\mathbf{R}^{D_x\times D_k}$、$\boldsymbol{W}_k\in\mathbf{R}^{D_x\times D_k}$ 和 $\boldsymbol{W}_v\in\mathbf{R}^{D_x\times D_v}$ 来生成相应的查询矩阵 $\boldsymbol{Q}$、键矩阵 $\boldsymbol{K}$ 和值矩阵 $\boldsymbol{V}$。需要注意的是 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 的维度应该是相同的。因此，用相同的符号来表示它们的形状。通过一个标准化函数 $\rho\left(q_i{}^Tk_j\right)$ 来评估第 i 个查询特征 $q_i{}^T$ 和第 j 个键特征 k_j 之间的相似性。通常，由于查询特征和键特征是由不同层生成的，$\rho\left(q_i{}^Tk_j\right)$ 和 $\rho\left(q_j{}^Tk_i\right)$ 之间的相似性不是对称的。通过计算所有位置对之间的相似性并将其作为权重，点积注意力模块根据加权求和计算位置 i 的值：

$$D(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\rho(\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}})\boldsymbol{V} \tag{1}$$

使用 softmax 来作为标准化函数：

$$\rho(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{K})=\mathrm{softmax}(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{K}) \tag{2}$$

式中：softmax 表示沿着矩阵 $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{K}$ 的每一行应用 softmax 函数。$\rho(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{K})$ 建模了所有位置对之间的相似性。但是，由于 $\boldsymbol{Q}\in\mathbf{R}^{N\times D_k}$ 和 $\boldsymbol{K}^{\mathrm{T}}\in\mathbf{R}^{D_k\times N}$，$\boldsymbol{Q}$ 和 $\boldsymbol{K}^{\mathrm{T}}$ 的乘积属于 $\mathbf{R}^{N\times N}$，导致 $O(N^2)$ 的内存复杂度和 $O(N^2)$ 的计算复杂度。因此，点积注意力的高资源需求严重限制了其在大规模输入中的应用。

在不改变 softmax 函数的前提下，点积注意力模块公式（1）生成的结果矩阵的第 i 行可以写成：

$$D(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})_i=\frac{\sum_{j=1}^{N}\mathrm{sim}(\boldsymbol{q}_i,\boldsymbol{k}_j)\boldsymbol{v}_j}{\sum_{j=1}^{N}\mathrm{sim}(\boldsymbol{q}_i,\boldsymbol{k}_j)} \tag{3}$$

式中：$\mathrm{sim}(\boldsymbol{q}_i,\boldsymbol{k}_j)$ 表示 $\boldsymbol{q}_i$ 和 $\boldsymbol{k}_j$ 之间相似度的函数。并且 $\mathrm{sim}(\boldsymbol{q}_i,\boldsymbol{k}_j)$ 可以进一步展开为 $\mathrm{sim}(\boldsymbol{q}_i,\boldsymbol{k}_j)=\phi(\boldsymbol{q}_i)^{\mathrm{T}}\varphi(\boldsymbol{k}_j)$，其中 $\phi(\cdot)$ 和 $\varphi(\cdot)$ 可以被视为核平滑器，如果 $\phi=\varphi$。因此相应的内积空间可定义为 $\langle\phi(\boldsymbol{q}_i),\varphi(\boldsymbol{k}_j)\rangle$。式（3）可以进一步被重写为：

$$D(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})_i=\frac{\phi(\boldsymbol{q}_i)^{\mathrm{T}}\sum_{j=1}^{N}\varphi(\boldsymbol{k}_j)\boldsymbol{v}_j{}^{\mathrm{T}}}{\phi(\boldsymbol{q}_i)^{\mathrm{T}}\sum_{j=1}^{N}\varphi(\boldsymbol{k}_j)} \tag{4}$$

取 $\phi(\cdot)=\varphi(\cdot)=\mathrm{softplus}(\cdot)$，其中：

$$\mathrm{softplus}(x)=\log(1+\mathrm{e}^x) \tag{5}$$

选择 softplus(·)，而不是 ReLU(·) 的原因是 softplus 的非零属性使得注意力在输入为负时能够避免零梯度。然后，相似度函数可以转换为：

$$\mathrm{sim}(\boldsymbol{q}_i,\boldsymbol{k}_j)=\mathrm{softplus}(\boldsymbol{q}_i)^{\mathrm{T}}\mathrm{softplus}(\boldsymbol{k}_j) \tag{6}$$

从而将公式（4）重写为：

$$D(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\frac{\mathrm{softplus}(\boldsymbol{Q})\mathrm{softplus}(\boldsymbol{K})^{\mathrm{T}}\boldsymbol{V}}{\mathrm{softplus}(\boldsymbol{Q})\sum_j\mathrm{softplus}(\boldsymbol{K})_{ij}{}^{\mathrm{T}}} \tag{7}$$

由于 $\sum_{j=1}^{N}\mathrm{softplus}(\boldsymbol{k}_j)\boldsymbol{v}_j{}^{\mathrm{T}}$ 和 $\sum_{j=1}^{N}\mathrm{softplus}(\boldsymbol{k}_j)$ 可以预先计算并重用，基于式（7）的注意力机制的时间和空间复杂度仅为 $O(N)$。

2.2 整体架构

除了上述空间注意力之外，通道注意力机制是另外一种常用的注意力机制。通道注意力机制（channel attention mechanism, CAM）是一种在深度学习中，特别是在卷积神经网络中使用的技术，它通过关注特征图的不同通道来提升模型性能。这种机制的核心思想是，在一个特征图中，不同的通道可能捕捉到的信息重要性是不同的，因此，通过学习每个通道的重要性，可以增强模型对关键信息的捕捉能力。

基于线性空间注意力和通道注意力的多注意力模块如图 2 所示。

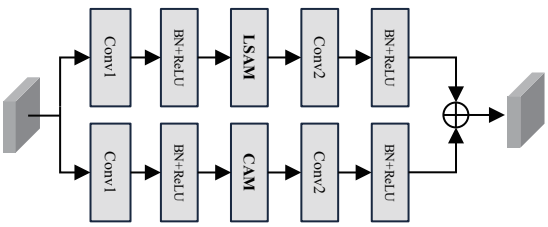

图 2 多注意力模块

对于空间维度，由于点积注意力机制的计算复杂度与输入大小呈二次方关系，使用线性空间注意力机制，如图中 LSAM 所示。对于通道维度，输入通道数 C 通常远小于特征图中包含的像素数（即 $C\ll N$）。因此，根据式（3），通道的 softmax 函数的复杂度（即 $O(C^2)$）并不大。因此，使用基于点积的通道注意力机制，如图 2 中 CAM 所示。然后将这两种注意力机制结合，设计了一个多注意力模块，以增强每

一层提取的特征图的区分能力。

方法整体架构如图 3 所示，利用在 ImageNet 上预训练的 ResNet-101 来提取特征图。具体来说，使用从 ResNet 四个残差块的输出中获取的 4 个不同尺度的特征图，然后被送入一个多注意力模块，最后使用双线性插值方法上采样得到与输入相同分辨率的分割图。

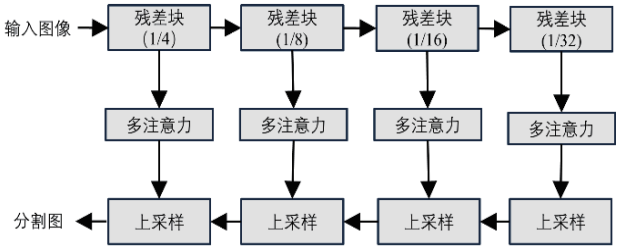

图 3 整体架构图

3 实验结果和分析

3.1 实验设置

实验用的数据集是高分辨率的 GID 数据集（gaofen image dataset）。GID 包含 10 幅 RGB 图像，大小为 7 200 px×6 800 px，由中国 Gaofen-2 卫星拍摄。每幅图像覆盖 506 平方公里的地理区域，并被标记为 15 类土地利用类别。实验将每幅图像分别划分为大小为 256 px×256 px 的不重叠块，并丢弃不能被 256 整除的边缘像素。因此，获得了 7 280 个子图。然后随机选择 60% 进行训练，20% 用于验证，其余 20% 用于测试。

实验采用语义分割常用的评估指标：像素精度（pixel accuracy, PA）、平均像素精度（mean pixel accuracy, MPA）和平均交并比（mean intersection over union, mIoU）来评估方法的性能。其中 PA 对少数类别不敏感，而 mIoU 对不平衡数据集中的少数类别过于敏感。

3.2 对比实验

为了全面评估多注意力网络的性能，与经典网络进行了对比，包括 U-Net、DeepLab V3、DeepLab V3+ 和 PSPNet[8]。所有模型都是用 PyTorch 实现的，使用 Adam 优化器，学习率为 0.000 3，批量大小为 16。所有实验都在配备有 24 GB RAM 的单个 NVIDIA GeForce RTX 3090 GPU 上进行。实验结果如表 1 所示。

表 1 对比实验结果

方法	PA	MPA	mIoU
U-Net	86.38	74.53	64.52
Deeplab V3	89.39	80.91	71.81
Deeplab V3+	90.13	81.48	72.67
PSPNet	90.57	82.21	74.80
本文	90.95	83.67	76.72

实验结果表明，所提出的方法在所有定量评估指标上的表现均超过了其他算法。对实验结果进行分析，由于 U-Net 结构简单且浅层，其特征提取能力有限，因此 U-Net 的性能不足。通过使用扩张空间金字塔池化（DeepLab V3 和 DeepLab V3+）、金字塔池化模块（PSPNet）聚合的多尺度上下文信息，有助于提高最终分割图的准确性。

与通过扩张卷积、池化操作或单独的注意力捕获上下文信息不同，多注意力网络直接在多层上对输入的每对像素建模全局依赖性。由于线性空间注意力显著降低了内存和计算需求，因此可以不考虑输入特征图的大小，利用多层全局上下文信息。因此，所提出的多注意力网络设计，它在四个尺度上对长距离依赖性进行建模，有助于捕获输入的精细特征。

3.3 消融实验

在所提出的多注意力分割网络中，使用 ResNet-101 作为骨干网络来获取全局上下文表示，然后利用注意力机制增强特征提取能力。为了进一步评估注意力块和骨干网络的贡献，在 GID 数据集上使用不同的设置进行了消融实验。结果如表 2 所示。

表 2 消融实验结果

骨干网络	注意力	PA	MPA	mIoU
ResNet-50		90.38	83.09	73.86
ResNet-50	√	90.53	83.94	75.27
ResNet-101		90.44	81.94	74.49
ResNet-101	√	90.95	83.67	76.72

ResNet-101 比 ResNet-50 产生了更高的准确率，因为更深的骨干网络使得网络能够捕获更精细的特征。这种深度结构的优势在于能够逐层构建更加抽象且具有辨识能力的特征表示，进而提高模型在复杂任务中的表现。对于不同的骨干网络，注意力机制平均提高了 1% 的 mIoU。这一结果表明，注意力机制通过加强对关键信息的关注，能够有效增强网络在特征选择和信息聚焦方面的能力，从而提高整体性能。注意力机制受生物机制启发，已经成为一种计算效率高且具有可配置灵活性的结构。通过将线性空间注意力和通道注意力的结合，构成的一个复杂度为 $O(N)$ 的多注意力模块，在实验中验证了其有效性和效率。

4 结语

本文针对注意力机制在遥感图像语义分割任务中存在的资源消耗巨大问题，设计了一种多注意力模块，将线性空间注意力与通道注意力两种注意力相结合。其中线性空间注意力将点积注意力机制的复杂度降低至 $O(n)$。通过整合该多注意力模块和骨干网络 ResNet-101，设计了一个多注意力网络，用于在不同层级整合语义信息，利用局部特征的上下文依赖性，提高了准确性和计算效率。

（下转第 74 页）

5 总结

本文将无线数据传输技术与Flash快速烧写技术相结合，创新性的实现了非接触式 FPGA 快速烧写方法，可在不脱去外部罩子、不增加裸露的硬件连线的情况下实现 FPGA 软件的升级，未来可以扩展到更多类型器件中，具有较好的应用意义。

后续还可考虑使用 FPGA Multi-boot 技术 [10]，通过将 RS 管脚连接到 Flash 地址线的高两位，使配置 Flash 划分为多块不同分区，防止意外烧写错误而导致 FPGA 无法正常加载现象。

参考文献：

[1] 戴政 , 陈小敏 , 廖志忠 , 等 . 基于 FPGA 的随机衰落及硬件实时统计实现 [J]. 航空兵器 , 2020, 27(2):71-76.

[2]Micron Technology. PC28F00AP30TF datasheet[EB/OL]. [2024-06-21]. https://www.alldatasheet.com/datasheet-pdf/pdf/558688/MICRON/PC28F00AP30TF.html.

[3] AMD. 7 series FPGAs configuration user guide (UG470)[EB/OL].(2023-12-05)[2024-06-25]. https://docs.amd.com/v/u/en-US/ug470_7Series_Config.

[4] NAIK K C, BABU J C, PADMAPRIYA K, et al. Implementing NAND flash controller using product reed solomon code on FPGA chip[J]. International journal ofengineering research and applications , 2013, 3(2): 224-229.

[5] 李鹏 , 兰巨龙 . 用 CPLD 和 Flash 实现 FPGA 配置 [J]. 电子技术应用 , 2006(6):101-103.

[6] 关珊珊 , 周洁敏 . 基于 Xilinx FPGA 的 SPI Flash 控制器设计与验证 [J]. 电子器件 , 2012, 35(2): 216-220.

[7] 康嘉 . 基于 FPGA 配置的电路系统设计 [D]. 西安 : 西安电子科技大学 , 2014.

[8] 郑吉华 . 一种 FPGA 芯片在线更新配置电路及方法 : CN202011134086.9[P]. 2021-02-05.

[9] 刘宇波 , 施文韬 . 基于 FPGA 和 PowerPC 的 FPGA 启动加载 Flash 升级系统及方法 [P]. 2016-12-7.

[10] 王群泽 , 胡方明 , 林汉成 , 等 . 基于 Flash 和 JTAG 接口的 FPGA 多配置系统 [J]. 单片机与嵌入式系统应用 , 2011, 11(4): 37-40.

【作者简介】

王博（1993—），男，河南鲁山人，硕士，工程师，研究方向：嵌入式系统与雷达信号处理。

（收稿日期：2024-11-14）

（上接第 70 页）

实验结果表明，在高分辨率遥感图像的语义分割这一复杂任务中，本文方法在实验指标上展现出良好的性能。在 GID 数据集上的实验，所提出的方法的性能在所有准确性指标上超过了对比的基线方法。通过消融研究，评估多注意力模块的影响，结果表明多注意力模块能够明显提升分割效率。另外，多注意力模块所展示的资源效率使其能更普遍地应用在多种网络中。

参考文献：

[1] 马震环 . 深度学习技术在图像语义分割中的应用研究[D]. 成都 : 中国科学院大学 ,2020.

[2] 张琛亮 . 几类基于 Transformer 的遥感图像语义分割方法研究 [D]. 南京 : 南京信息工程大学 ,2023.

[3] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation[DB/OL]. (2017-12-05)[2024-06-16].https://doi.org/10.48550/arXiv.1706.05587.

[4] CHEN L C, ZHU Y K, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[J].Computer vision ,2018,11211:833-851.

[5] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[J].IEEE transactions on pattern analysis and machine intelligence, 2017, 39(4):640-651.

[6] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation[J]. Medical image computing and computer-assisted intervention, 2015, 9351: 234-241.

[7] LI R, ZHENG S Y, ZHANG C, et al. Multiattention network for semantic segmentation of fine-resolution remote sensing images [J]. IEEE transactions on geoscience and remote sensing, 2021, 60: 1-13.

[8] ZHAO H S, SHI J P, QI X J, et al. Pyramid scene parsing network[DB/OL]. (2017-04-27)[2024-05-11]. https://doi.org/10.48550/arXiv.1612.01105.

【作者简介】

张垚杰（1993—），男，山西长治人，硕士研究生，助教，研究方向：深度学习、算法设计与分析。

（收稿日期：2024-11-19）