

基于复数域 Transformer-Unet 混合模型的 PolSAR 地物分类

谢雯, 张嘉鹏, 张哲哲, 闪晨超
(西安邮电大学通信与信息工程学院 西安 710121)

摘要: 传统的基于深度学习的极化合成孔径雷达 (Polarimetric Synthetic Aperture Radar, PolSAR) 地物分类方法, 通过堆叠卷积层提取图像局部特征, 难以建立长距离依赖关系。基于自注意力机制的深度学习模型 Transformer (变换) 在图像分类任务中取得了成功, 其自注意力机制能够捕获全局像素之间的关联性, 同时 PolSAR 地物分类任务被证实: 相比于实数域, 其在复数域上表现出更好的分类效果。因此, 本文将 Transformer 引入到复数域中, 提出了一种基于复数域的 Transformer 和 Unet (语义分割网络) 混合模型 (CT-Unet) 用于 PolSAR 地物分类, 将 Transformer 与 CNN 相结合, 对复数类型的 PolSAR 数据进行特征提取, 使用西安数据集和德国数据集进行 PolSAR 地物分类的实验结果表明: 提出的模型能够有效提高 PolSAR 地物分类的准确性, Transformer 有望在 PolSAR 地物分类任务中弥补卷积神经网络的不足。

关键词: 极化合成孔径雷达; 复数域; Transformer; Unet

中图分类号: TP75; TP183 文献标志码: A 文章编号: 2095-1000(2024)03-0035-08

DOI: 10.12347/j.ycyk.20240116002

引用格式: 谢雯, 张嘉鹏, 张哲哲, 等. 基于复数域 Transformer-Unet 混合模型的 PolSAR 地物分类[J]. 遥测遥控, 2024, 45(3): 35-42.

PolSAR Terrain Classification Based on Transformer-Unet Hybrid Model in the Complex Domain

XIE Wen, ZHANG Jiapeng, ZHANG Zhezhe, SHAN Chenchao

(School of Communications and Information Engineering, Xi'an University of Posts and Telecommunications, Xi'an 710121, China)

Abstract: The traditional deep learning-based Polarimetric Synthetic Aperture Radar (PolSAR) feature classification method extracts image local features by stacking convolutional layers, which makes it difficult to establish long-range dependencies. It is noted that Transformer, a deep learning model based on a self-attention mechanism that captures global pixel-to-pixel correlations, has achieved success in image classification tasks. Meanwhile, the PolSAR feature classification task has demonstrated better classification results in the complex domain compared to the real domain. Therefore, Transformer is introduced into the complex domain, and a hybrid model of Transformer and Unet based on the complex domain (CT-Unet) is proposed for PolSAR feature classification. This model combines Transformer with CNN for feature extraction on PolSAR data of complex type. The experimental results of PolSAR feature classification using the Xi'an dataset and German dataset show that the proposed model can effectively improve the accuracy of PolSAR feature classification. Transformer is expected to make up for the shortcomings of convolutional neural networks in the PolSAR feature classification task.

Keywords: Polarimetric synthetic aperture radar(PolSAR); Complex domain; Transformer; Unet

Citation: XIE Wen, ZHANG Jiapeng, ZHANG Zhezhe, et al. PolSAR Terrain Classification Based on Transformer-Unet Hybrid Model in the Complex Domain[J]. Journal of Telemetry, Tracking and Command, 2024, 45(3): 35-42.

0 引言

遥感卫星^[1]技术发展已久, 其中的极化合成孔径雷达(PolSAR)技术以其多极化、多频率和多角度的特点, 能够提供更丰富和详细的地物表面信息, 显著提高了地形分类的准确性, 对城市规划^[2]、农业监测^[3]、军事探测^[4]等相关领域具有重要意义和实际应用价值。通过深入研究 PolSAR 在地物分类中的应用, 可以为提高遥感数据解译精度提供技术支持^[5]。PolSAR 地物分类方法经过不断发展大致可分为三类: 第一类是基于 PolSAR 数据散射机制^[6]的算法, 通过分解 PolSAR 数据的相干矩阵、散射矩阵等得到有用的地物分类特征; 第二类是基于统计分布^[7]的算法, 通常根据不同的分布对 PolSAR 数据进行模型的搭建, 而传统的 PolSAR 地物分类方法通常需要从 PolSAR 数据中手工提取特征, 这导致了模型的泛化性能受限等问题。随着深度学习技术的发展, 第三类即基于深度学习的方法被应用在了 PolSAR 地物分类任务^[8], 卷积神经网络(Convolutional Neural Network, CNN)^[9]的应用减轻了对手工设计特征的依赖, 提高了在复杂场景中的地物分类性能。尽管该方法在学习 PolSAR 图像的空间和极化特征表示方面非常有效, 仍受限于相位信息和感受野的丧失。为了解决上述问题并考虑到 PolSAR 地物分类属于像素级别的分类任务, 全卷积神经网络(Fully Convolutional Network, FCN)^[10]被引入到 PolSAR 地物分类领域, Aduagna Mullissa 等人^[11]提出的模型 PolSARNet, 降低了计算复杂度的同时获得了令人满意的分类结果。尽管 FCN 可以实现对 PolSAR 图像的端到端像素级分类, 但现有的方法中没有考虑到不同特征图之间存在的潜在关系。其变体 U-Net^[12]通过连接编码器和解码器部分的特征图, 保留更多的低级语义信息并减少空间信息的丧失, 而这对于 PolSAR 地物分类任务来说是非常重要的, 但其中的卷积操作依旧具有局部性, 容易受到感受野大小的限制。而 Transformer^[13]模型中的自注意力机制能够有效捕捉输入序列中各个位置的全局依赖关系, 这意味着 Transformer 可以更好地保留远离中心像素的信息, 从而提高对全局图像结构的理解能力, Dosovitskiy 等^[14]提出了视觉转换器(ViT)模型, 凭借自注意力机制获取全局特征, 有效弥补了 CNN 的局部性, 可以很好地学习图像全局像素

的依赖关系。陈辉等^[15]提出了一种基于 Swin Transformer(采用平移窗口的分层视觉转换器)的多尺度混合光谱注意力网络(SMSaNet), 该网络实现了不同特征区域之间对全局信息建立长距离依赖关系, 有效提升了图像的分类准确率, 但上述方法无法充分利用 PolSAR 数据的相位信息。Zhang 等人^[16]首次将 PolSAR 数据的复数域信息作为输入数据提出了复值 CNN(Complex-valued CNN, CV-CNN)取得了较为满意的分类精度。Yu 等人^[17]在 CV-CNN 网络模型的基础上提出了 CVFCNN 网络模型, 该网络模型具有比 CV-CNN 网络模型更好的分类性能, 同时也验证了 PolSAR 数据的复数域中包含的相位信息的重要性。Xie 等人^[18]最近提出的 TF-CVFCN(复值深度全卷积神经网络)在 CV-FCN 的基础上引入迁移学习取得了令人满意的分类效果, 进一步说明了 PolSAR 地物分类在深度学习领域广阔的发展前景。

基于前述分析并受以前的工作启发, 本文提出了一种 PolSAR 地物分类模型 CT-Unet, 该模型在复数域中将 Transformer 模型与 Unet 模型进行结合, 从 PolSAR 图像中有效捕获全局、局部和多尺度的特征, 从而提高分类性能。实验采用西安地区数据集和德国 ESAR 数据集在多个不同的模型上进行验证, 所提出的模型为 PolSAR 地物分类任务提供了一种可行的方法。

1 Transformer-Unet 混合模型

本文提出的网络模型如图 1 所示。为了充分利用 PolSAR 数据的相位信息, 将复数域信息作为网络模型的输入数据, 在通道维度上将输入数据分割成实部与虚部, 接下来整体网络模型也将对输入数据的实部与虚部分别进行处理。首先, 对其进行多次 3D 卷积和 2D 卷积, 将提取的特征输入到不同尺度的 Transformer 中, 融合 Transformer 得到的特征图并重新转换作为解码器的输入, 解码器包括特征融合块、上采样卷积块以及用于整合实部和虚部的模块。最后, 通过重塑操作将输出的图像尺寸调整为与输入特征图的尺寸相同。

1.1 复数运算

在 PolSAR 地物分类任务当中, 网络模型利用相位信息能够提升其分类性能^[19], 复数域中则包含 PolSAR 数据的相位信息, 因此本文将在复数域上完成网络模型的搭建, 将卷积、池化及激活函数

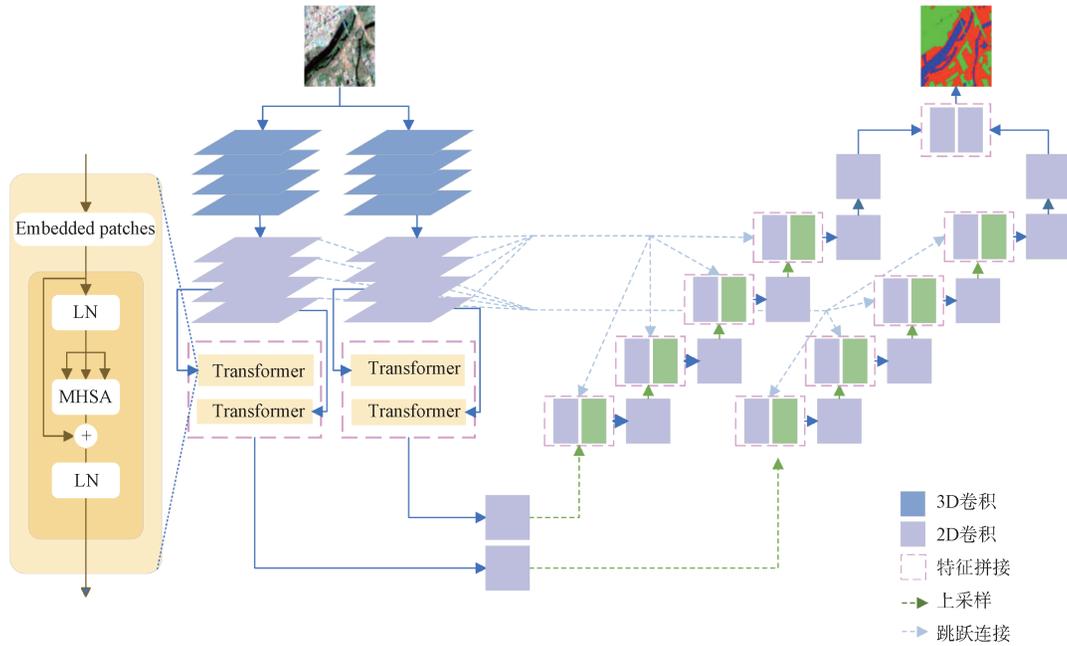


图1 CT-Unet网络模型结构图

Fig. 1 Network architecture of CT-Unet

从实数域扩展到复数域中。首先将输入的数据与复数卷积核进行卷积，以下公式所示：

$$w * h = (a + jb) * (x + jy) = (a * x - b * y) + j(b * x + a * y) = R(w * h) + jI(w * h) \quad (1)$$

其中， w 为复数卷积核， $w = a + jb$ ， h 为输入数据， $h = x + jy$ ， $R(w * h)$ 表示卷积结果的实部， $I(w * h)$ 表示卷积结果的虚部。通过提取并结合输入数据和复数卷积核的实部与虚部进行卷积操作，再结合得到复数卷积的结果。池化操作能够有效减小训练参数，增加模型泛化能力^[20]，扩展到复数域上的池化操作以平均池化为例，其池化过程如下式所示：

$$\text{AvePooling}(h) = \text{AvePooling}(x) + j\text{AvePooling}(y) = R(h) + jI(h) \quad (2)$$

其中， h 为输入数据，其定义为 $h = x + jy$ ， $R(h)$ 表示平均池化结果的实部， $I(h)$ 表示平均池化结果的虚部。通过对输入数据的实部和虚部分别进行平均池化操作，将结果进行组合得到复数平均池化的结果。激活函数使得神经网络能够学习和表示更为复杂的函数^[21]，本文将ReLU激活函数扩展到复数域，具体操作如下式所示：

$$\text{ReLU}(h) = \text{ReLU}(x) + j\text{ReLU}(y) = R(h) + jI(h) \quad (3)$$

其中， $R(h)$ 表示激活函数结果的实部， $I(h)$ 表示

激活函数结果的虚部。将输入数据的实部和虚部分别进行激活函数处理，将所得结果进行结合得到复数域激活函数的结果。

1.2 编码器

编码器部分位于图1的左侧，该部分包括卷积模块和Transformer模块，主要实现特征提取的功能，所提出的网络首先会将输入数据在通道维度分割成实部与虚部，此步骤旨在充分利用PolSAR数据中的相位信息。然后为了获取更加细致和精确的局部特征和全局特征，在Transformer模型之前利用卷积核大小不同的四层3D卷积和四层2D卷积提取输入数据的局部特征，该提取特征的方法能够在控制参数数量的前提下，充分提取数据的空间信息及通道信息^[22]，然后将提取到的特征输入到多尺度的Transformer模型以提取全局特征，提取全局特征的过程则依赖于多头自注意力机制^[23]，该机制的过程如图2所示。

实现过程如式(4)~式(6)所示，其中注意力函数计算公式如下

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \quad (4)$$

其中， \mathbf{Q} 、 \mathbf{K} 和 \mathbf{V} 分别是查询、键和值的矩阵， d_k 是键的维度。每个注意力头的计算公式如下

$$h_i = \text{Attention}(\mathbf{W}_i^Q \mathbf{Q}, \mathbf{W}_i^K \mathbf{K}, \mathbf{W}_i^V \mathbf{V}) \quad (5)$$

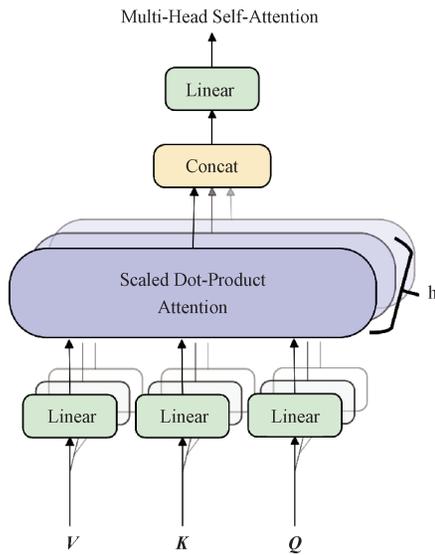


图 2 多头自注意力机制的过程

Fig. 2 The process of multi-head self-attention mechanism

其中, W_i^o 、 W_i^v 和 W_i^k 是参数矩阵。多头自注意力计算公式如下

$$\text{Multihead}(Q, K, V) = \text{Concat}(h_1, \dots, h_n) W^o \quad (6)$$

其中, W^o 是参数矩阵。多尺度 Transformer 部分一共由两个 Transformer 模块组成, 这里的多尺度指的是输入 Transformer 模块的特征尺寸不同, 第一个 Transformer 模块的输入来自第三层的 2D 卷积, 第二个 Transformer 模块的输入来自第四层的 2D 卷积, 其多头自注意力机制模块堆叠的次数不同, 其中的自注意力头数也不同。第一个 Transformer 模块中堆叠了八个多头自注意力机制模块, 自注意力头数为 6, 第二个 Transformer 模块中堆叠了 6 个多头自注意力机制模块, 自注意力头数为 4, 这样设计的编码器可在连续卷积的过程中尽可能地提取局部与全局信息。最终, 经过编码器获得的特征图进行融合与重构后输入到解码器中。

1.3 解码器

解码器部分位于图 1 的右侧, 该部分包括特征融合模块、上采样卷积模块、实部与虚部融合模块, 主要实现上采样并得到最后的分类结果图。在解码器中, 首先将输入的图像进行上采样, 将其结果与编码器部分经跳跃连接传递的特征进行拼接, 再进行卷积后得到的特征图继续上采样。该步骤通过获得编码阶段的浅层特征从而帮助模型更好地得到多尺度特征信息。重复上述步骤四次后将实部与虚部两个支路得到的特征图进行融

合, 并通过重构将其恢复到原始图像大小, 该解码器部分的设计在保留图像语义信息的同时有效地恢复了空间细节。

2 实验

2.1 数据集

本文采用西安数据集和德国数据集验证所提出模型的有效性。西安数据集是 2010 年由 RADARSAT-2(星载合成孔径雷达)卫星系统^[24]获取的中国西安西部渭河地区场景, 图像大小为 $8 \text{ m} \times 8 \text{ m}$, 选取其分辨率为 512×512 的子图像块进行实验, 该数据集共包括草地、城市和河流三种地形类别。德国数据集是由德国宇航中心 ESAR(电子扫描阵列雷达)系统获取的关于 Oberpfaffenhofen(奥伯法芬霍芬)地区的全极化雷达图像, 图像大小为 $3 \text{ m} \times 3 \text{ m}$, 分辨率为 1300×1200 , 该数据集共包括建筑区、开放区和林区三种地形类别。

2.2 实验结果

2.2.1 西安地区地物分类实验

将滑动窗口和步幅设置为 128×128 和 6 对西安地区图像进行处理, 训练集和测试集的样本数量分别为 125 和 516, 模型训练时的批量大小设置为 16, 并使用 $1\text{E}-3$ 作为学习率的 Adam(自适应矩估计)优化器来优化目标函数, 迭代次数设置为 300 次。所有实验均在配备 TensorFlow(机器学习服务平台)2.10.0 版本和 NVIDIA GTX 4070ti GPU(英伟达显卡)的 PC 上运行。为了说明本文所提出模型的有效性, 在相同的条件下分别测试了 FCN(全卷积网络), Unet(语义分割网络)和 TF-CVFCN 模型, 并进行了消融实验, 其中模型 1 为 Unet 与单尺度的 Transformer 进行结合, 模型 2 是在模型 1 的基础上加入了 3D 卷积, 以及 CT-Unet。在本文涉及的所有实验中, 均使用整体准确率(OA)、KAPPA(卡帕)系数和平均交并比(MIoU)作为其分类效果的评价指标, 对比实验的可视化结果如图 3 所示。每个类别具体的精度以及各个模型的数量和运算量如表 1 所示。图 3 展示了不同模型在西安数据集上的分类结果图, 绿色区域为草地, 红色区域为城市, 蓝色区域为河流。从图中可以看出, FCN 及 Unet 取得的分类结果图比较连贯平滑, 但还是存在较多的错误分类点, 在一些细节方面分类情况不是很理想。TF-CVFCN 的错误分类点较少, 整体效果更加接近本文提出的模型。从模型 1 开始加

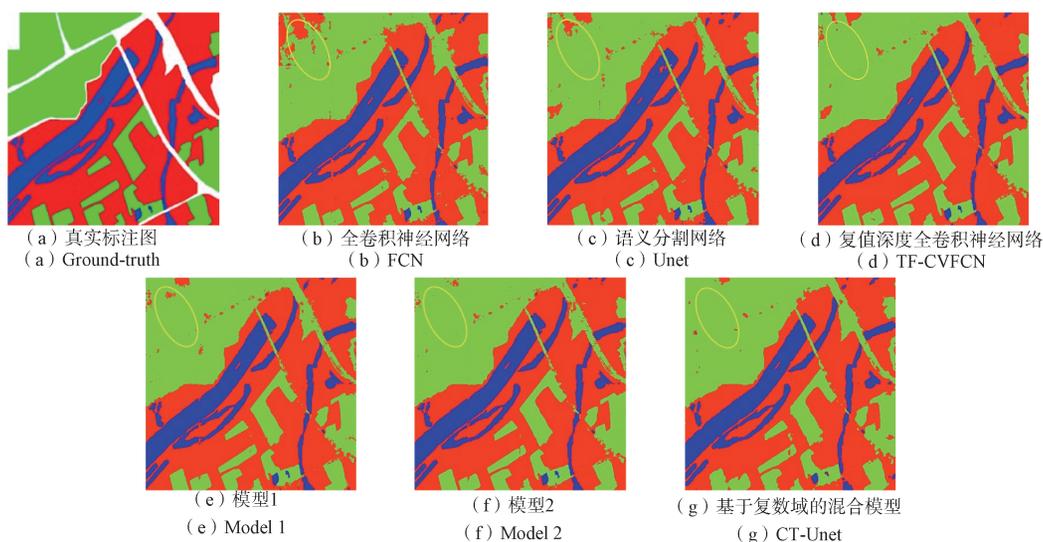


图3 西安地区数据集分类结果图

Fig. 3 The classification results of the Xi'an dataset

入Transformer后,能够较为有效地提取全局特征,加深了网络模型对于特征的提取能力,所得的分类结果也有所改善。而多尺度Transformer的采用更进一步加深了网络模型对于PolSAR数据的理解能力,特别是黄色圆圈圈出的区域图像更加清晰且错分现象大大减少,说明了通过多头自注意力机制能够更好地捕捉PolSAR的全局特征,更全面地理解地物的分布,从而提高地物分类的精度。表1中罗列了不同模型在西安数据集上的分类准确度。通过表1可以看到,提出的模型在OA值、

MIoU和Kappa系数分别为98.85%、97.89%和97.24%,相较于其他的模型都获得了最高的分数,其中OA值分别比FCN、Unet、TF-CVFCN、模型1及模型2提高了2.64%、2%、0.74%、1.77%、0.83%。MIoU分别提高了5.91%、4.03%、1.59%、3.19%、1.6%。Kappa分别提高了4.79%、3.23%、1.25%、2.51%、1.26%。在具体类别上包括草地、城市及河流也都取得了最高的准确度,所得到的结果图更加接近于真实标注图,同时由于Transformer模块的加入也伴随着运算量的升高。

表1 西安数据集分类准确度

Table 1 The classification accuracy of the Xi'an dataset

	FCN	Unet	TF-CVFCN	模型1	模型2	CT-Net
草地	0.971 5	0.984 7	0.989 9	0.969 1	0.984 9	0.991 7
城市	0.963 9	0.950 1	0.974 2	0.982 2	0.978 3	0.988 9
河流	0.926 8	0.958 5	0.968 9	0.956 2	0.969 4	0.977 3
OA	0.962 1	0.968 5	0.981 1	0.971 8	0.980 2	0.988 5
MIou	0.913 3	0.932 1	0.956 5	0.940 5	0.956 4	0.972 4
Kappa	0.931 0	0.946 6	0.966 4	0.953 8	0.966 3	0.978 9
GFLOPs	0.023 6	0.032 7	0.084 0	420.655 1	437.223 7	1 545.490 7
Params	11 849 740	10 898 480	42 153 928	15 400 480	15 505 746	27 813 130

2.2.2 德国ESAR地物分类实验

为了更好地验证CT-Net网络模型的性能,使用德国ESAR数据集进行测试,将滑动窗口和步幅设置为 128×128 和31对德国ESAR图像进行处理,训练集和测试集的样本数量分别为296和5 727,模型训练时的批量大小设置为110,并使用

$1E-3$ 作为学习率的Adam优化器来优化目标函数,迭代次数设置为200次。同样,分别测试了FCN,Unet和TF-CVFCN模型,并进行了消融实验,各个模型与西安地区地物分类实验相同。对比实验的可视化结果如图4所示,每个类别具体的精度及采用德国数据集后各个模型的运算量及参数量如

表 2 所示。图 4 展示了不同模型在德国 ESAR 数据集上的分类结果图, 建筑区为红色区域, 林区为绿色区域, 黄色为开放区区域^[25]。从图中可以看到 FCN 及 Unet 取得的分类结果图相较于真实标注图存在较多的错误分类点, 有部分将建筑区错分为林区的现象, TF-CVFCN 对于建筑区的分类较为准确, 但开放区错误分类的现象较多, 综合分类效果要优于模型 1 和模型 2。模型 1 加入了 Transformer 后分类结果图准确性较 FCN 和 Unet 有所提

升, 开放区的错分现象大幅减少, 也是所有本文提到的模型中在开放区类别上表现最好的模型。模型 2 在林区上的表现相较于模型 1 在准确率上进一步得到提升, 绿色的林区区域错分现象较少。最后, 本文提出的模型综合表现优于其他模型, 各个类别的准确率从结果图来看均达到较好的水平。特别是图像中圈出的建筑区区域更加平滑且几乎没有错分现象, 再次验证了多尺度 Transformer 能够有效捕捉 PolSAR 图像的全局特征。

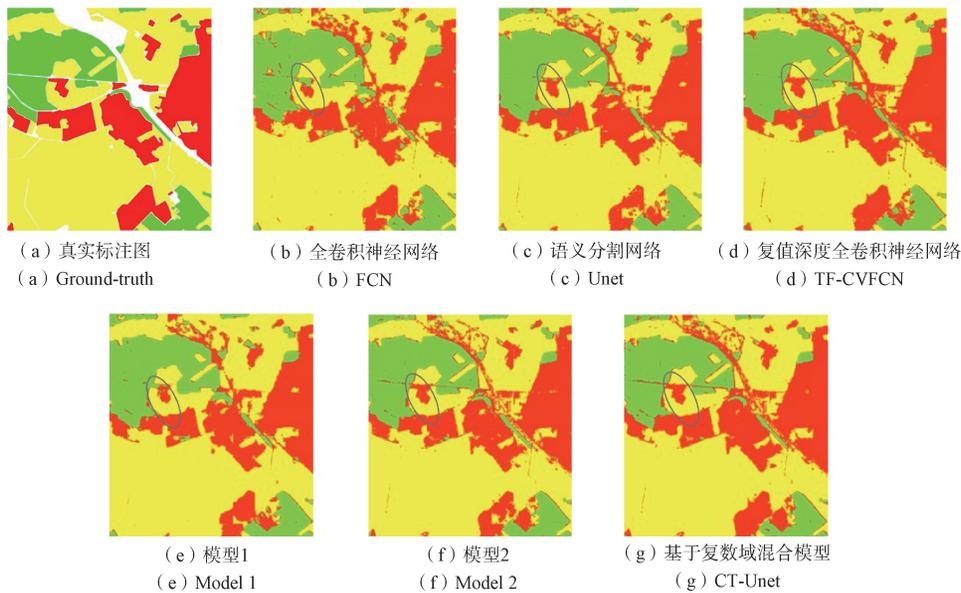


图 4 德国 ESAR 数据集分类结果图

Fig. 4 The classification results of the Germany dataset

表 2 中罗列了不同模型在德国数据集上的分类准确度, 通过表 2 可以看到所提出的模型在 OA 值、MIoU 和 Kappa 系数分别为 98.42%、96.01% 和 96.93%, 相较于其他的模型都获得了最高的分数, 其中 OA 值分别比 FCN、Unet、TF-CVFCN、模型 1 及模型 2 提高了 2.93%、1.46%、0.44%、1.08%

和 0.57%。MIoU 分别提高了 6.83%、3.33%、0.64%、2.73% 和 0.92%。Kappa 分别提高了 5.53%、2.65%、0.50%、2.15% 和 0.72%。而随着数据集的变化, 运算量也会随之变化, 在使用德国数据集后, 各个模型的运算量均有增长。表 2 中的数据再次证实了所提出的模型能够提升 PolSAR 地物分类

表 2 德国 ESAR 数据集分类准确度

Table 2 The classification accuracy of the Germany dataset.

	FCN	Unet	TF-CVFCN	模型 1	模型 2	CT-Unet
建筑区	0.906 8	0.923 0	0.975 0	0.932 8	0.955 2	0.984 4
林区	0.939 8	0.973 7	0.967 7	0.970 4	0.980 9	0.962 9
开放区	0.981 4	0.989 0	0.986 0	0.992 5	0.988 1	0.991 2
OA	0.954 9	0.969 6	0.979 8	0.973 4	0.978 5	0.984 2
MIou	0.891 8	0.926 8	0.953 7	0.932 8	0.950 9	0.960 1
Kappa	0.914 0	0.942 8	0.964 3	0.947 8	0.962 1	0.969 3
GFLOPs	0.033 4	0.045 7	0.090 1	2 891.865 6	3 005.774 4	25 403.761 3
Params	11 849 740	10 898 480	42 153 928	15 400 480	15 505 746	27 813 130

的精度。

3 结束语

本文提出了一种 PolSAR 地物分类模型 CT-Unet, 该模型实现了在复数域上将 Transformer 与 Unet 相结合进行 PolSAR 地物分类任务。为了充分提取 PolSAR 图像中的有效信息, 将 Transformer 用于编码器部分来提取特征信息, 可以在不丢失局部信息的同时, 更好地整合全局信息。Unet 则通过其编码器-解码器结构提供多尺度的特征, 通过将二者结合有效捕捉到 PolSAR 图像局部、全局、多尺度的特征, 从而提高 PolSAR 地物分类的精度。从实验结果也可以看出: 所提出的模型分类效果比其他方法更加平滑和准确。但由于多头自注意力机制带来的计算量, 本文提出的模型出现计算复杂及参数量膨胀的问题, 未来考虑将在不影响分类精度的情况下实现模型的轻量化。

参考文献

- [1] 叶晖, 李博, 王金华, 等. 光学遥感卫星地面图像数据测试的局限及解决方法初探[J]. 遥测遥控, 2015, 36(5): 62-67.
YE Hui, LI Bo, WANG Jinhua, et al. Preliminary exploration of a new method for image data testing of optical satellites[J]. Journal of Telemetry, Tracking and Command, 2015, 36(5): 62-67.
- [2] 肖雨彤, 张继贤, 黄国满, 等. FCN 与 CRF 结合的 PolSAR 影像建筑区域提取[J]. 遥感信息, 2020, 35(3): 44-49.
XIAO Yutong, ZHANG Jixian, HUANG Guoman, et al. Building area extraction of PolSAR images based on FCN and CRF[J]. Remote Sensing Information, 2020, 35(3): 44-49.
- [3] 郭交, 尉鹏亮, 周正舒, 等. 基于时变特征的多时相 PolSAR 农作物分类方法[J]. 农业机械学报, 2017, 48(12): 174-182.
GUO Jiao, WEI Pengliang, ZHOU Zhengshu, et al. Crop classification method with differential characteristics based on multi-temporal PolSAR images[J]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(12): 174-182.
- [4] OH Y, HONG S Y, KIM Y, et al. Polarimetric backscattering coefficients of flooded rice fields at L-and C-bands: Measurements, modeling, and data analysis[J]. IEEE Transactions on Geoscience and Remote Sensing, 2009, 47(8): 2714-2721.
- [5] SHARIF M, ATTARCHI S. Estimation of hyrcanian forests coverage using ALOS PALSAR and MODIS satellite data (2007—2017) [J]. Advances in Space Research, 2024, 73(1): 426-439.
- [6] LIANG D, TANG T, BAN Y, et al. Unsupervised polarimetric SAR urban area classification based on model-based decomposition with cross scattering[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2016, 116: 86-100.
- [7] 刘振. 小样本 PolSAR 图像地物分类技术研究[D]. 西安: 西安电子科技大学, 2018.
- [8] PARIKH H, PATEL S, PATEL V. Classification of SAR and PolSAR images using deep learning: A review[J]. International Journal of Image and Data Fusion, 2020, 11(1): 1-32.
- [9] CUI Y, LIU F, JIAO L, et al. Polarimetric multipath convolutional neural network for PolSAR image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 1-18.
- [10] XIE W, JIAO L, HUA W. Complex-valued multi-scale fully convolutional network with stacked-dilated convolution for PolSAR image classification[J]. Remote Sensing, 2022, 14(15): 3737.
- [11] MULLISSA A G, PERSELLO C, STEIN A. PolSAR-Net: A deep fully convolutional network for polarimetric SAR image classification[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2019, 12(12): 5300-5309.
- [12] HUANG H, LIN L, TONG R, et al. Unet 3+: A full-scale connected unet for medical image segmentation[C]//ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020: 1055-1059.
- [13] WANG W, CHEN W, QIU Q, et al. Crossformer++: A versatile vision transformer hinging on cross-scale attention[J]. arXiv preprint arXiv:2303.06908, 2023.
- [14] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[C]//International Conference on Learning Representations, 2021. arXiv: 2010.11929.
- [15] 陈辉, 张甜, 陈润斌. 基于轻量级卷积 Transformer 的图像分类方法及在遥感图像分类中的应用[J]. 电子与信息学报, 2022, 44: 1-9.
CHEN Hui, ZHANG Tian, CHEN Runbin. Image classification method based on lightweight convolutional transformer and its application in remote sensing[J].

- Journal of Electronics & Information Technology, 2022, 44: 1-9.
- [16] ZHANG Z, WANG H, XU F, et al. Complex-valued convolutional neural network and its application in polarimetric SAR image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(12): 7177-7188.
- [17] YU L, HU Y, XIE X, et al. Complex-valued full convolutional neural network for SAR target classification[J]. IEEE Geoscience and Remote Sensing Letters, 2019 (99): 1-5.
- [18] XIE W, SUN H, ZHANG Y, et al. Polsar image classification via transfer learning and fully convolutional network[C]//IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2023: 8026-8029.
- [19] 秦先祥, 余旺盛, 王鹏, 等. 基于复值卷积神经网络样本精选的极化 SAR 图像弱监督分类方法[J]. 雷达学报, 2020, 9(3): 525-538.
QIN Xianxiang, YU Wangsheng, WANG Peng, et al. Weakly supervised classification of PolSAR images based on sample refinement with complex-valued convolutional neural network[J]. Journal of Radars, 2020, 9(3): 525-538.
- [20] 江泽涛, 秦嘉奇, 张少钦. 参数池化卷积神经网络图像分类方法[J]. 电子学报, 2020, 48(9): 1729-1734.
JIANG Zetao, QIN Jiaqi, ZHANG Shaoqin. Parameterized pooling convolution neural network for image classification[J]. Acta Electronica Sinica, 2020, 48(9): 1729-1734.
- [21] 杜圣杰, 贾晓芬, 黄友锐, 等. 用于 CNN 模型图像分类任务的高效激活函数[J]. 红外与激光工程, 2022, 51(3): 483-491.
- DU Shengjie, JIA Xiaofen, HUANG Yourui, et al. High efficient activation function design for CNN model image classification task[J]. Infrared and Laser Engineering, 2022, 51(3): 483-491.
- [22] 张腊梅, 陈泽茜, 邹斌. 基于 3D 卷积神经网络的 PolSAR 图像精细分类[J]. 红外与激光工程, 2018, 47(7): 7-14.
ZHANG Lamei, CHEN Zexi, ZOU Bin. Fine classification of polarimetric SAR images based on 3D convolutional neural network[J]. Infrared and Laser Engineering, 2018, 47(7): 7-14.
- [23] HAN K, WANG Y, CHEN H, et al. A survey on vision transformer[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(1): 87-110.
- [24] LOU Y, KIM Y, ZYL J V, et al. The NASA/JPL airborne synthetic aperture radar's 1996 pacRim deployment[C]// International Geoscience and Remote Sensing Symposium. IEEE, 1997. DOI:10.1109/IGARSS.1997.606460.
- [25] VAN ZYL J, CARANDE R, LOU Y, et al. The NASA/JPL three-frequency polarimetric AIRSAR system[C]// IGARSS'92 International Geoscience and Remote Sensing Symposium. IEEE, 1992, 1: 649-651.
- [作者简介]
谢 雯 1989 年生, 副教授, 硕士生导师。
张嘉鹏 1998 年生, 硕士研究生。
张哲哲 1997 年生, 硕士研究生。
闪晨超 1999 年生, 硕士研究生。
- (本文编辑: 傅 杰)