

结合注意力机制的遥感图像场景分类模型构建

杨全海

陕西职业技术学院

DOI:10.32629/acair.v4i2.20398

[摘要] 遥感图像场景分类是地球观测数据分析的核心技术。本文以ResNet50为骨干架构,嵌入通道与空间双注意力模块构建分类模型,通过串行融合策略实现特征精细化筛选,并在AID和NWPU-RESISC45数据集上开展实验。结果表明,该模型分类精度分别达97.43%和95.38%,优于现有对比模型,有效提升遥感图像场景分类性能,为遥感图像智能解译提供技术支撑。

[关键词] 遥感图像; 场景分类; 注意力机制; 通道注意力; 空间注意力; 残差网络
中图分类号: P237 文献标识码: A

Construction of a Remote Sensing Image Scene Classification Model Integrating Attention Mechanism

Quanhai Yang

Shaanxi Vocational and Technical College

[Abstract] Remote sensing image scene classification is a core technology for Earth observation data analysis. This paper constructs a classification model using ResNet50 as the backbone architecture, embedding both channel and spatial attention modules to achieve refined feature selection through a serial fusion strategy. Experiments are conducted on the AID and NWPU-RESISC45 datasets. The results show that the model achieves classification accuracy of 97.43% and 95.38%, respectively, outperforming existing comparative models and effectively improving the performance of remote sensing image scene classification, thereby providing technical support for intelligent interpretation of remote sensing images.

[Key words] Remote sensing image; scene classification; attention mechanism; channel attention; spatial attention; residual network

遥感对地观测技术持续发展,海量高分辨率遥感图像数据的获取趋于常态化,从该类数据中自动且准确地识别场景语义类别,是遥感图像智能解译领域的核心研究内容^[1]。遥感图像场景分类存在诸多问题,遥感图像类内差异大、类间相似度高,关键目标与背景信息混杂情况突出,传统卷积神经网络受局部感受野限制,难以捕获全局上下文依赖关系,处理多尺度目标与大范围空间关系时存在不足^[2]。注意力机制可引导模型聚焦图像中具备判别力的区域,抑制无关背景噪声干扰,融合通道与空间注意力机制的遥感图像场景分类模型,以残差网络为骨干架构,嵌入双注意力模块完成特征提取与类别判定。

1 相关工作

1.1 遥感图像场景分类研究进展

遥感图像场景分类方法呈现出从传统特征工程到深度学习驱动的发展过程,早期方法依靠手工设计的尺度不变特征变换、方向梯度直方图、局部二值模式等特征描述子,搭配支持向量机分类器实现场景判别,此类方法的特征设计对领域经验依赖性

较强,泛化能力存在局限^[3]。AlexNet在ImageNet竞赛中取得突破后,基于卷积神经网络的深度学习方法成为遥感图像场景分类研究的主要方向,VGG、ResNet及其变体、DenseNet等架构应用于该任务,ResNet凭借跳跃连接结构缓解深层网络梯度消失问题,ResNet50在AID数据集上可实现约94.67%的分类精度。

1.2 注意力机制及其在遥感分类中的应用

深度学习中,注意力机制为不同特征分量赋予差异化权重系数,对特征信息进行处理。通道注意力机制针对特征通道重要性建模,SENet通过全局平均池化压缩空间信息,经全连接层学习通道依赖关系,空间注意力机制聚焦特征位置重要性,二者协同可从多维度完成特征筛选。遥感图像场景分类领域中,RCA网络结合残差结构与通道注意力,在RSSCN7、PatternNet数据集精度分别为97%、99%,GLDAN网络在RSSCN7数据集精度达98.75%,HLAE-Net较骨干网络精度最高提升1.04个百分点。

2 模型相关分析

2.1 整体架构

遥感图像场景分类模型整体架构遵循编码器—分类器的设计范式。以残差网络ResNet50作为特征编码骨干,在骨干网络的特征提取阶段并行嵌入通道注意力模块和空间注意力模块,形成双分支注意力增强的特征表示层。输入遥感图像首先经过骨干网络的卷积层和残差块进行初步特征提取,获得维度为 $C \times H \times W$ 的特征图 F ,其中 C 表示通道数, H 和 W 分别表示特征图的高度和宽度。随后,该特征图同时馈入通道注意力分支和空间注意力分支,分别生成通道注意力权重向量 $M_c \in \mathbb{R}^C$ 和空间注意力权重矩阵 $M_s \in \mathbb{R}^{H \times W}$ 。最终,特征图 F 经由两个注意力权重依次加权,得到增强后的特征表示 F' ,再经全局平均池化压缩空间维度,输入全连接层完成场景类别的概率预测^[4]。

2.2 通道注意力模块

通道注意力模块结构遵循SENet的通道压缩—激励范式,采用一维卷积替代全连接层以降低参数量。给定输入特征图 $F \in \mathbb{R}^{C \times H \times W}$,首先通过全局平均池化操作将每个通道的空间信息压缩为一个标量描述子:

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W F_c(i, j) \quad (1)$$

式中, z_c 表示第 c 个通道的全局统计量, $F_c(i, j)$ 表示第 c 个通道在空间位置 (i, j) 处的特征值, H 和 W 分别为特征图的空间高度和宽度。该操作将所有通道的空间分布压缩为维度为 C 的一维向量 z 。

随后,通过一维卷积层与Sigmoid激活函数生成通道注意力权重:

$$\alpha = \sigma(\text{Conv1D}_k(z)) \quad (2)$$

式中, $\alpha \in \mathbb{R}^C$ 为通道注意力权重向量, σ 表示Sigmoid函数, Conv1D_k 表示卷积核大小为 k 的一维卷积操作。一维卷积核的尺寸 k 根据通道数自适应确定,以实现跨通道的局部交互。通道注意力加权后的特征图 F_c 通过如下逐通道乘法获得:

$$F_c = \alpha \odot F \quad (3)$$

其中 $\odot$ 表示通道维度的逐元素乘法。

2.3 空间注意力模块

空间注意力模块对特征图中最具判别力的空间区域进行定位,增强模型对关键目标位置的响应强度^[5]。与通道注意力关注“哪些通道”不同,空间注意力关注“哪些位置”。

给定输入特征图 $F \in \mathbb{R}^{C \times H \times W}$,首先沿通道维度分别执行平均池化和最大池化操作,将多通道特征压缩为两个二维空间描述子:

$$F_{\text{avg}}(i, j) = \frac{1}{C} \sum_{c=1}^C F_c(i, j) \quad (4)$$

$$F_{\text{max}}(i, j) = \max_{c=1,2,\dots,C} F_c(i, j) \quad (5)$$

式中, $F_{\text{avg}}(i, j)$ 表示所有通道在空间位置 (i, j) 处的均值响应, $F_{\text{max}}(i, j)$ 表示该位置的最大响应值。平均池化捕捉全局空间分布特征,最大池化突出局部显著特征,两者互补以获取更丰富

的空间信息。

将上述两个二维描述子沿通道维度拼接,形成双通道的空间特征图,再经由一个卷积核大小为 7×7 的卷积层和Sigmoid激活函数生成空间注意力权重矩阵:

$$\beta = \sigma(\text{Conv}_{7 \times 7}([F_{\text{avg}}; F_{\text{max}}])) \quad (6)$$

式中, $\beta \in \mathbb{R}^{1 \times H \times W}$ 为空间注意力权重矩阵, $[F_{\text{avg}}; F_{\text{max}}]$ 表示沿通道维度的拼接操作, $\text{Conv}_{7 \times 7}$ 表示 7×7 的卷积运算。空间注意力加权后的特征图通过如下逐元素乘法获得:

$$F_s = \beta \odot F \quad (7)$$

2.4 注意力融合与分类输出

通道注意力模块和空间注意力模块分别从“通道重要性”和“空间重要性”两个维度对特征进行精细化筛选。采用串行融合策略:先通过通道注意力加权筛选出关键语义通道,再通过空间注意力加权突出关键空间位置。融合过程的数学表达为:

$$F' = \beta \odot (\alpha \odot F) \quad (8)$$

融合后的增强特征 F' 依次经过全局平均池化层和全连接分类层。全局平均池化将尺寸为 $C \times H \times W$ 的特征图压缩为 C 维特征向量。全连接层采用Softmax激活函数,输出图像属于各类别的后验概率:

$$P(y=j|x) = \frac{\exp(w_j^\top v + b_j)}{\sum_{k=1}^K \exp(w_k^\top v + b_k)} \quad (9)$$

式中, $v \in \mathbb{R}^C$ 为全局平均池化后的特征向量, w_j 和 b_j 分别为第 j 类的权重向量和偏置项, K 为总类别数。

3 实验相关分析

3.1 实验数据集

实验选用两类常用的遥感图像场景分类基准数据集。AID数据集由武汉大学和华中科技大学联合发布,包含30个场景类别,总计10000张航拍图像,每类图像数量在220至420张之间,图像尺寸为 600×600 像素,涵盖机场、海滩、森林、农田、体育场等多种场景。NWPU-RESISC45数据集由西北工业大学发布,包含45个场景类别,每类700张图像,总计31500张图像,图像尺寸为 256×256 像素,在地理覆盖、空间分辨率、拍摄角度、光照与季节变化等方面存在多样差异。

3.2 实验设置

实验按6:2:2的比例划分各数据集样本为训练集、验证集与测试集,AID数据集训练集约6000张图像,验证集与测试集各约2000张,NWPU-RESISC45数据集训练集约18900张,验证集与测试集各约6300张。输入图像统一缩放至 224×224 像素,通过随机水平翻转与随机裁剪完成数据增强。模型训练采用随机梯度下降优化器,动量因子0.9,权重衰减系数 $1e-4$,初始学习率0.01,以余弦退火策略衰减,训练90轮次,批次大小32。ResNet50采用ImageNet预训练权重初始化,注意力模块参数用Xavier初始化,损失函数为交叉熵损失,以总体分类准确率为评估指标,实验在NVIDIA RTX 3090 GPU上运行,框架为PyTorch 2.0。

3. 3实验结果

表1 各模型在AID数据集上的分类精度对比

模型	分类准确率 (%)	参数量 (M)
VGG16	91.23	138.36
ResNet34	93.16	21.80
ResNet50	94.67	25.56
ResNet101	94.89	44.55
ResNet50 + SE	96.14	26.08
本文模型	97.43	26.12

表2 各模型在NWPU-RESISC45数据集上的分类精度对比

模型	分类准确率 (%)
VGG16	88.36
ResNet34	90.47
ResNet50	93.19
ResNet101	93.62
ResNet50 + SE	94.21
本文模型	95.38

表3 注意力模块消融实验结果

通道注意力	空间注意力	分类准确率 (%)
×	×	94.67
√	×	96.14
×	√	95.83
√	√	97.43

3. 4消融实验

消融实验在AID数据集上开展, 分别评估仅使用通道注意力模块、仅使用空间注意力模块以及同时使用两个模块的模型性能。

4 结语

本文围绕遥感图像场景分类展开研究, 梳理了其从传统特征工程到深度学习驱动的发展历程, 以及注意力机制在该领域的应用现状。设计了以ResNet50为骨干, 嵌入通道与空间双分支注意力模块的分类模型, 通过串行融合策略实现特征精细化筛选。实验表明, 该模型在AID数据集上分类精度达97.43%, 在NWPU-RESISC45数据集上达95.38%, 均优于VGG16、ResNet系列及ResNet50+SE模型。消融实验证实, 双注意力模块协同作用可显著提升性能, 较单一模块效果更优。该模型兼顾分类精度与参数量, 为遥感图像场景分类提供了高效可行的解决方案。

[参考文献]

- [1]赵君君.基于多尺度注意力机制的遥感图像复杂场景分类方法研究[J].中国高新科技,2026,(03):145-146+149.
- [2]吴文彦.基于特征增强与多特征融合的遥感图像场景分类方法研究[D].安徽理工大学,2025.
- [3]袁昊.面向自然保护地的遥感图像场景分类研究与应用[D].济南大学,2025.
- [4]杨松,王晓晖,王晓燕,等.基于注意力机制与特征融合的遥感图像场景分类[J].传感技术学报,2024,37(12):2078-2083.
- [5]李智,高连如,郑珂,等.高分辨率遥感图像场景分类研究进展[J].遥感学报,2024,28(11):2739-2760.

作者简介:

杨全海(1974--),男,汉族,陕西澄城人,工学硕士,讲师,研究方向:图像处理。