

GCAT-U-Net 嵌入全局坐标注意力机制的 遥感地块分割网络

苏 耀¹, 于 濂¹, 周 伟²

(1. 北京师范大学 数学科学学院, 北京 100875;

2. 北京师范大学 互联网教育智能技术及应用国家工程实验室, 北京 100875)

摘要: 遥感影像的地块背景特征复杂, 当前地块分割方法不能较好地处理模糊的边缘信息, 导致分割精度不理想; 文章利用注意力机制处理地块特征, 提出了一种基于全局坐标注意力机制的遥感地块分割网络: GCAT-U-Net; 该方法在 U-Net 网络基础上嵌入了全局坐标注意力机制, 加强了深度神经网络对于遥感影像数据中重要特征的关注度; 在公开的 GID 数据集上的实验结果表明, 文章提出的模型将准确率从 0.904 1 提升到了 0.922 7, 比传统 U-Net 网络提高了 2 个百分点; 结合特征自身重要性和特征位置信息的全局坐标注意力机制有助于更精确的目标定位, 其输出相较于嵌入单一注意力机制, 地块边界更为清晰, 提升效果更为显著。

关键词: 地块语义分割; 注意力机制; 模式识别; U-Net 网络; 卷积神经网络

GCAT-U-Net: Remote Sensing Plot Segmentation Network with Global Coordinate Attention Mechanism

SU Yao¹, YU Lian¹, ZHOU Wei²

(1. School of Mathematical Sciences, Beijing Normal University, Beijing 100875, China;

2. National Engineering Laboratory for Cyberlearning and Intelligent Technology,
Beijing Normal University, Beijing 100875, China)

Abstract: The plots background features in remote sensing images are complex. The current plot segmentation methods cannot preferably handle the fuzzy edge information, so the segmentation accuracy is not ideal. The attention mechanism is used to process the land parcel features, and a remote sensing land parcel segmentation network based on the global coordinate attention mechanism is proposed. This method embeds the global coordinate attention mechanism on the basis of U-Net network, which strengthens the deep neural network's attention to important features in remote sensing image data. The experimental results on the public GID data set show that the model proposed in the article increases the accuracy from 0.904 1 to 0.922 7, which is 2% higher than the traditional U-Net network. The global coordinate attention mechanism that combines the feature importance itself and the feature location information is helpful for more accurate target positioning. Compared with the embedded single attention mechanism, the parcel boundary is more clearer and the improvement effect is more significant.

Keywords: parcel semantic segmentation; attention mechanism; pattern recognition; U-Net network; convolution neural network

收稿日期: 2021-09-30; 修回日期: 2021-10-26。

作者简介: 苏 耀(1995-), 男, 山西运城人, 硕士研究生, 主要从事模糊数学与人工智能方向的研究。

通讯作者: 周 伟(1989-), 男, 湖南益阳人, 博士研究生, 主要从事模式识别方向的研究。

引用格式: 苏 耀, 于 濂, 周 伟. GCAT-U-Net 嵌入全局坐标注意力机制的遥感地块分割网络[J]. 计算机测量与控制, 2022, 30(2): 222-228, 236.

0 引言

耕地的数量和质量是保持农业可持续发展的关键, 利用高分辨率的卫星遥感影像^[1]可以识别并获取耕地区域, 准确的耕地分布能够为国家决策部门提供重要支撑。随着遥感技术的迅速发展, 现如今遥感图像以及相关技术的应用越来越广泛^[2-4], 耕地面积的统计提取、农作物识别以及地块识别等图像语义分割技术, 对促进农业发展有重要的科研意义和经济价值。

学者们针对遥感地块语义分割这一实际问题进行了一系列的研究, 提出了很多种方法。有基于边缘、形态学、区域、随机场等传统图像分割特征的方法, 例如: 使用形态学中自适应全局阈值配合方法^[5]、高斯马尔可夫随机场结合支持向量机算法^[6]等。很显然, 上述传统语义分割方法, 在特定规模的数据上有着良好的效果, 但其依赖于专家知识、人工参数选取以及大量重复性实验等先验知识, 对于大规模的数据来说, 随着特征复杂程度的提高, 会有一些的限制, 相较于深度学习方法, 其在准确率、精度、以及效率上尚存在着差距。

在深度学习领域, 卷积神经网络凭借其独特的自主学习能力以及处理大数据集的优势, 在语义分割中展现了强大的潜力。当前广泛使用的语义分割网络主要有: Long 等人提出来的全连接层替代卷积层, 保留图像上下文空间特征, 并且是图像像素端到端预测的 FCN 网络模型^[7]; Badrinarayanan 等人在继承了 FCN 网络模型的思想的同时, 进一步提出了 SegNet 网络模型, 该模型去掉了全连接层并且在编码器 (Encoder) 信息和解码器 (Decoder) 信息之间采用直接连接的方式, 保留了数据结构中大量有用的特征信息, 使得网络在训练过程以及精准度上都有了提升^[8]; Chaurasia 等人同样在基于编码-解码结构网络上, 通过改进特征融合方式, 在解码阶段融合编码阶段的数据特征信息, 提出了 LinkNet 网络模型, 获得了更为精细的实验效果^[9]; Ronneberger 等人同样在 FCN 的基础上使用对称的编码器和解码器以及添加跳跃连接提出了 U-Net 网络模型, 其主要意义在于解码阶段重复利用了编码阶段的数据的特征信息, 因此对于数据特征的纹理信息能够更好的利用^[10]; Oktay 等人将注意力机制引入到了 U-Net 网

络中, 其在编码器和解码器特征进行跳跃连接的地方, 引入了一个门控制信号, 旨在控制不同特征的重要性, 对于 U-Net 网络的检测精度进行了提升^[11]; Zhou 等人在 U-Net 网络中使用了嵌套密集跳跃连接来替代原来的跳跃连接, 提出了 U-Net++ 网络模型, 在减小了编码阶段特征图和解码阶段特征图的语义鸿沟的同时, 增加了 U-Net 的网络性能^[12]。

相对于一般场景特征来说, 遥感影像具有数据量大、范围广的特点^[13], 其综合反映了某一时段内某一地区的各种地物的形态和分布, 包括: 地质、地貌、土壤、植被、水文、人工建筑物等, 因此遥感影像存在纹理特征复杂、边缘信息繁琐、背景特征丰富的特点。随着遥感影像空间分辨率的提高, 在信息更加丰富的同时, 带来的还有更多的数据干扰的问题, 例如: 房屋、树木等, 其几何形状以及结构内容具有同一性或结构性, 给土地的细化分类, 带来了更多的挑战。为了更好地处理遥感影像数据, 本文在 U-Net 网络结构中引入注意力机制。实验结果表明注意力机制的加入, 能够有效提升语义分割的精确度以及地块边缘的清晰度。

本文的主要创新点如下:

1) 针对地块语义分割这一现实问题, 结合网络结构中通道以及数据位置等特征的研究, 本文在 U-Net 的网络架构基础上嵌入通道注意力机制和坐标注意力机制, 用于提高地块分割网络性能, 证明了 U-Net 网络在遥感地块分割中的实用性, 以及注意力机制的加入, 能够有效提升语义分割的精确度以及地块边缘的清晰度;

2) 关注网络特征提取中对于特征自身重要性以及特征位置信息的提取, 既考虑全局信息, 又考虑坐标信息, 创新的提出综合考虑全局信息及其位置信息的全局坐标注意力机制, 进一步提高了模型性能, 相较于同类注意力机制的嵌入, 提高了分割准确性, 并且对于边界的分割也更为清晰。

1 U-Net 网络体系结构及注意力机制

1.1 U-Net 网络体系结构

U-Net 网络是 2015 年 Ronneberger 等人提出来的用于医学图像处理的网络, 是图像分割领域著名的深度学习网络模型^[10]。该模型主要使用卷积压缩和扩展的编码-解码结构和跳级连接, 实现了对语义信

息以及图像信息进行融合的特征提取。由于其对称的两个路径形似 U 型结构而命名为“U-Net”。

U-Net 网络模型结构如图 1 所示，该模型是一端到端的网络模型，由左半边的捕获上下文信息的压缩通道 (Encoder) 和右半边的扩展通道 (Decoder) 以及连接压缩通道和扩展通道的跳跃连接组成。从模型结构来分析，主要分为 3 个部分：下采样，上采样和跳跃连接。左半部分即 Encoder，每个下采样结构由两个 3×3 的卷积层，卷积层后使用 ReLU 函数激活，然后使用 2×2 的步长为 2 的最大池化操作组成，每经过一次下采样，通道数翻倍；右半部分 Decoder，每个上采样结构由一个 2×2 的上采样卷积层，每个卷积层后使用 ReLU 函数激活后与对应的 Encoder 层的输出特征图结果相加之后进行 2 个 3×3 的卷积层，最后进行 ReLU 函数激活组成，其中上采样将特征通道的数量减半。在最后一层，使用 1×1 的卷积，将输出映射到所需的类别数。网络中的卷积皆采用 valid 的填充方式，主要是用来保证结果都是在没有缺失的上下文特征中得到的，且网络中通过中间的跳跃连接，将在网络压缩阶段所产生的浅层次特征图与扩张阶段所产生的深层次特征图相结合，根据所获取的特征图进行地块语义的预测分割。

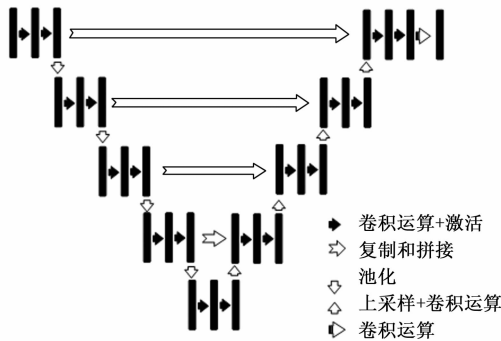


图 1 U-Net 网络体系结构

1.2 注意力机制

注意力模型在当下已经成为神经网络中优化网络的一个重要研究方向，很多研究者都颇为关注^[14-16]。在非大幅度增加模型复杂程度的基础上，注意力机制通过参数调整，加强网络对于数据中重要特征的关注度，抑制数据中的背景特征，以提高模型预测结果的分割精度，尤其对细节处的提升，有较为明显的优势。

在遥感图像中，地块的分布是错综复杂的，不仅要考虑部分的遮挡，例如大树、房屋等，还要考虑田

间道路等细微处的分割，因此引入注意力机制，能够有利于细化分割。本文提出了一种全局坐标注意力机制，对 U-Net 网络进行了优化改进。

1.2.1 通道注意力机制 (SE, squeeze-and-excitation channel attention mechanism)

遥感图像在 U-Net 网络的卷积操作的过程中，不同的特征通道对于最后的图像分割也会存在不同的影响，原有的 U-Net 网络，是直接采用相加拼接的方式进行特征的融合，忽略了不同特征通道的权重，不利于在最后的分割结果中某些重要特征通道的信息增强。因此考虑特征通道之间的关系，提出了通道注意力机制^[17]，该机制是通过特征图中各通道间对于输出结果的影响的要性来进行的分析计算，这有利于增强特征图中重要通道的影响，提高 U-Net 网络对最终图像分割的精准度。

如图 2 所示，通道注意力机制是一种先压缩再扩张的模型。具体来说，给定输入特征 X ，假设原始特征图的维度为 $H \times W \times C$ ，其中 H 是高度 (Height)， W 是宽度 (Width)， C 是通道数 (Channel)。压缩部分是将 $H \times W \times C$ 压缩为 $1 \times 1 \times C$ ，相当于把 $H \times W$ 压缩成一维，实际中一般使用全局平均池化来实现，并不使用其他的池化方式，因为每个通道的特征图最后会池化为一个值，在这种情况下，全局平均池化更能代表其整体的特征。 $H \times W$ 压缩成一维后，相当于这一维参数获得了之前 $H \times W$ 全局的视野，感受区域更广，将特征通道数据进行压缩变成 $1 \times 1 \times C$ 后，加入一个 FC 全连接层 (Fully Connected)，进行降维操作，其目的在于扩大感受野的同时，降低计算量。之后通过 ReLU 激活，再接一个 FC 全连接层完成升维操作，升维后的向量通过 Sigmoid 激活，变成 C 维向量，所代表的是每个通道的重要性。在得到不同通道的重要性大小后再乘 (激励) 到之前的特征图对应通道上。通道注意力机制有着复杂度低、新增参数和计算量小的优势。

1.2.2 坐标注意力机制 (CAT, coordinate attention mechanism)

通道注意力机制仅仅考虑了卷积之后不同特征通道之间的权重分配，虽然它能够有效反映通道间相关性，但却忽视了位置信息的问题，因此应考虑在不同的通道特征图中位置对于输出结果的影响，坐标注意力机制可以将位置信息嵌入到通道注意力中，通过融

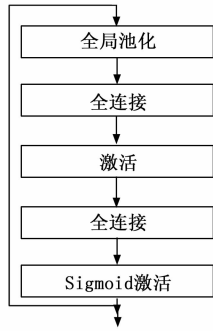


图 2 通道注意力机制结构

入水平与垂直不同方向的注意力机制, 以捕获特征位置的依赖关系^[18]。

所提位置注意力机制模块的结构如图 3 所示, 相较于通道注意力机制, 其将全局池化拆分成了两个特征编码操作, 具体来说, 给定输入特征 X , 假设其维度为 $H \times W \times C$, 采用两个池化核 $(H, 1)$ 、 $(1, W)$ 对每个通道沿着水平与垂直坐标进行编码, 描述如下:

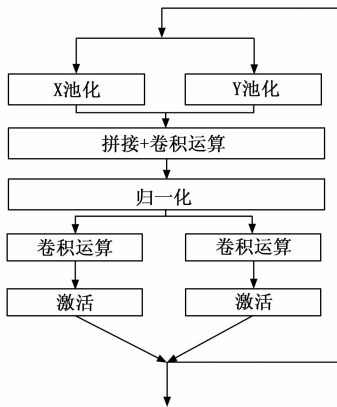


图 3 坐标注意力机制结构

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i \leq W} x_c(h, i) \quad (1)$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq i \leq H} x_c(j, w) \quad (2)$$

通过这两个变换可以沿着两个空间方向集成特征, 并生成与该方向相关的特征图。这两种变换使得所提注意力模块可以更好地捕获位置依赖性, 有助于进行更精确的目标定位。结合上一步所提取的特征, 首先对其进行拼接, 这一步是为了保障两个方向的输出拥有一致的量纲, 将拼接之后的特征数据送入到共享 1×1 卷积 F_1 并得到:

$$f = \delta(F_1([z^h, z^w])) \quad (3)$$

最后将 f 沿着空间维度拆分为两个独立张量 $f^h \in R^{R/r \times H}$, $f^w \in R^{C/r \times W}$, 采并用两个额外的 1×1 卷积 F_h, F_w 对其进行分离为和 X 相同维度的张量。

$$g^h = \sigma(F_h(f^h)) \quad (4)$$

$$g^w = \sigma(F_w(f^w)) \quad (5)$$

接下来, 对输出进行扩展用以作为注意力权重, 并作用于输入得到该注意力模块的最后输出:

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (6)$$

位置注意力机制本质是能够在卷积后的通道图上, 结合其所处的位置信息, 提取所有像素的显著特征。对于输出的特征图来说, 其建立了特征像素和位置之间的关联信息, 在对于边界点以及细节处等微小的地方, 网络的分割能力有显著提高。

1.2.3 全局坐标注意力机制 (GCAT, global coordinate attention mechanism)

上述两个注意力机制, 分别考虑了卷积之后的特征图不同通道之间以及各个通道特征图中不同位置之间的关系之后提出的注意力机制。在此基础之上, 本文不仅考虑不同位置之间的影响, 同时还考虑全局信息对于输出特征图的影响, 提出了全局坐标注意力机制, 将进一步关注特征图中每一特征自身对于整体特征的重要性及其所处位置信息对于输出结果的重要性, 这对于地块语义分割来说能够提高其特征选取效率, 细化选取有效的地块语义分割特征。

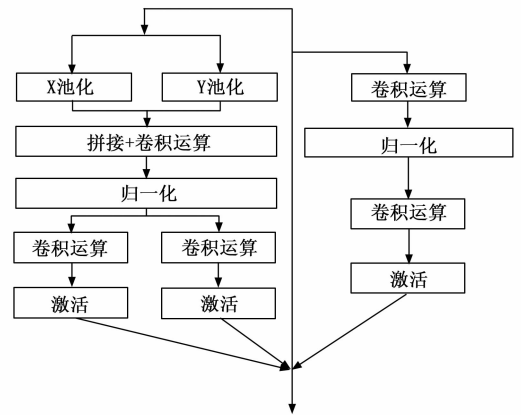


图 4 全局坐标注意力机制结构

所提出的注意力机制模块如图 4 所示, 主要分为两个部分: 第一部分同于上述坐标注意力机制, 通过变换沿着水平与垂直两个空间方向集成特征, 生成方向相关特征图。

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (7)$$

第二部分融入特征图自身对于输出的影响，将输入特征图送入到共享 1×1 卷积，之后进行标准化操作，最后采用额外的 1×1 卷积以及 Sigmoid 激活，输出结果为与 X 相同维度的张量。

$$y_n = \sigma(F_1(\sigma(F_1(x_c(i,j)))))) \quad (8)$$

该注意力机制模块最终输出为与输入得到与 X 相同维度的张量 Y 。

$$Y = y_n + y_c \quad (9)$$

经过这两种变换使得所提全局注意力机制模块不仅可以捕获到位置依赖性，而且考虑其自身对于输出的作用，在有助于进行更精确的目标定位同时，细化输出，加强了重要特征的提取。

2 模型构建

2.1 模型结构

本文对 U-Net 网络进行了改进，改进后的网络设计如图 5 所示。

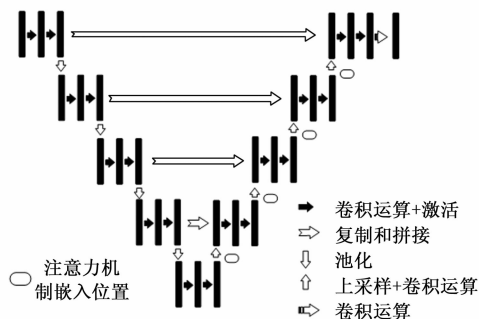


图 5 注意力机制嵌入位置

本文在一个标准的 U-Net 架构上，嵌入不同的注意力机制。综合考量 U-Net 的网络结构，为充分利用其对于图像的多个尺度的特征提取，因此考虑跳跃连接之后对结合了网络压缩阶段所产生的浅层次特征图与扩张阶段所产生的深层次特征图的特征进行注意力机制改进。这样做，不仅有利于改进模型对于粗略的特征图的上下文信息提取，有利于突出显示最终需要的特征类别和位置，并且对于感受野的捕捉有促进作用。从结果输出来分析，这 4 个位置进行注意力机制嵌入，也有利于其突出跳跃连接传递的显著特征，相较于压缩阶段嵌入来说，更加贴近输出特征图，对于结果的改进更为直接。此外，相关注意力机制的嵌入可以让网络更加关注其卷积层间的信息、特征图的位置信息以及自身信息。

结合图 5 的注意力机制嵌入方式以及 SE 通道注

意力机制、CAT 坐标注意力机制、GCAT 全局坐标注意力机制，我们设计了 SE-U-Net 通道注意力机制嵌入网络、CAT-U-Net 坐标注意力机制嵌入网以及 GCAT-U-Net 全局坐标注意力机制嵌入网络。

2.2 参数设置

在训练过程中，每次实验参数设置相同，batch size 每批次输入大小为 8，训练迭代次数为 100 次，训练图像每张大小为 480 像素 * 480 像素，实验使用 Adam 优化算法，使用 Adam 中默认参数^[19]，其中学习率设为 0.001，指数衰减率 0.9，调整学习率为间隔调整学习率 (Step LR)。

2.3 损失函数

将损失函数设计为 CELoss 和 DiceLoss 二者的加权和，考虑 Dice 系数为评价指标引导模型进行更好的训练。

损失函数：

$$loss = loss(x_i, y_i) + s \quad (10)$$

其中： x_i 表示样本 i 的预测概率； y_i 表示样本 i 的真实标签，取值为 0 或 1； w_i 表示样本 i 的权重：

$$loss(x_i, y_i) = -w_i[y_i \log x_i + (1 - y_i) \log(1 - x_i)] \quad (11)$$

Dice 系数差异函数 (Dice loss)，是一种集合相似度度量函数，通常用于计算两个样本的相似度 (值范围为 $[0, 1]$)，公式如下：

$$s = 1 - \frac{2|X \cap Y|}{|X| + |Y|} \quad (12)$$

$|X|$ 和 $|Y|$ 分别表示 X 图片真实标注的标签和 Y 网络预测的结果标签的元素个数。其中，分子中的系数 2，是因为分母存在重复计算 X 和 Y 之间的共同元素的原因。

2.4 评价指标

本文主要研究的是遥感地块语义分割，因此使用语义分割中普遍使用的准确率度量：像素准确率 (PA, pixel accuracy)、平均像素准确率 (MPA, mean pixel accuracy)、平均 IOU (MIOU, mean intersection over union) 和加权平均 IOU (FWIoU, frequency weighted intersection over union)。

3 实验结果与分析

3.1 数据获取与处理

GID 数据集^[20]是武汉大学的开源高分影像数据

集,它是利用高分 2 号 (GF-2) 卫星影像构建的大规模土地覆盖数据集,该数据集覆盖范围大,分布范围广,空间分辨率高,优于先前的土地覆盖数据集。因为本实验专注于地块提取,所以利用其精细分类集,精细分类集包含有:工业用地、城市住宅、农村住宅、交通用地、灌溉地、水田、旱地、河、湖、池塘等 15 个类别。

在精细分类遥感数据集中,挑选了一副土地占有率为 56.7% 的 $7\,200$ 像素 \times $6\,800$ 像素的高分辨率遥感图像。通过对标注数据的二值化,得到训练所需的地面真值标签,其中 1 为地块类别,0 为其他类别。随着所设计的网络的深度、宽度的不断增加,网络模型的参数量都是数以百计的,因此需要大量的图像数据进行训练以得到更好的模型,而实际情况中数据并没有那么多,为了更好的提取图像特征,使用以下方法对图像训练集进行丰富。在获得 batch 数据之后,对这个 batch 的数据进行了数据增强,主要方法手段有:旋转 90° 、旋转 180° 、旋转 270° 、翻转、光照调整、模糊操作、增加噪声等方式,将训练数据集扩充到 3 000 张,其中训练集与测试集比例为 7:3。一方面,它增加了训练的数据量,提高了模型的泛化能力;另一方面,它增加噪声数据,提升了模型的鲁棒性。

3.2 结果与讨论

本节讨论 U-Net 网络在嵌入不同注意力机制,在 GID 实验数据集上的实验结果和分析。

实验所有模型的输入图像尺寸为 480 像素 \times 480 像素,输出图像为 480 像素 \times 480 像素大小的预测标签图。部分分割结果对比如图 6,依次为:原图、Ground Truth 以及 U-Net、SE-U-Net、CAT-U-Net 和 GCAT-U-Net 的预测结果。

其中,SE-U-Net、CAT-U-Net 分别为:单独嵌入通道注意力机制、坐标注意机制的 U-Net 网络,GCAT-U-Net 为本文设计的针对问题研究提出来的全局坐标注意力机制改进网络。所有对比网络在相同环境下运行,且损失函数、数据预处理以及参数设置等方面均保持一致。

从预测图中可以看出,U-Net 模型的预测结果与真实标注 label 相差较大,而将通道注意力机制嵌入之后的 SE-U-Net 模型预测的计算结果则有了

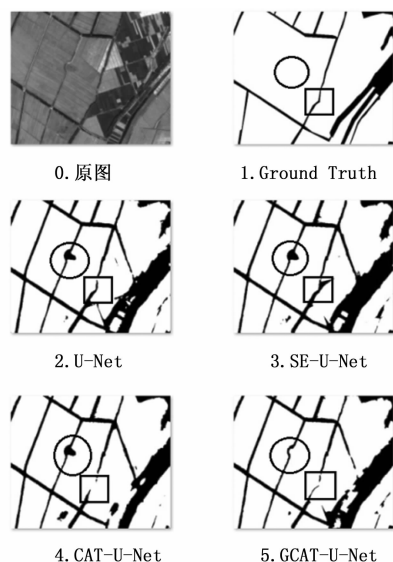


图 6 网络在 GID 测试集上分割效果图

明显提升,进一步嵌入坐标注意力机制的 CAT-U-Net 模型预测结果在 SE-U-Net 的基础上有了更进一步的提高。相较于前两种注意力机制来说,本文提出的全局坐标注意力机制嵌入的 GCAT-U-Net 模型则对于预测结果的精确度提升效果最好。如图 8,在每个预测图的圆框内,只有本文设计的 GCAT-U-Net 模型的预测结果分类较为准确,其余模型,有明显的将地块类别错分为其余类别的问题。另外,从图 8 中的每个预测图的方框中可以看出,GCAT-U-Net 模型的预测结果相较于 CAT-U-Net 模型的预测结果边界分割更清晰,分类更为准确;而 SE-U-Net 模型的预测结果,有明显的将地块类别错分为其余类别的问题。由此说明注意力机制可以明显提高遥感地块语义分割与提取的精度,所得到的分割边界也更为清晰。

不同注意力机制嵌入的 U-Net 语义分割网络模型在验证数据集上的基于分割准确率评价指标的结果分析如表 1 所示。通过对表进行分析,基于语义分割评价指标 PA,与 U-Net 进行对比,SE-U-Net 提升了 1.43%,CAT-U-Net 提升了 1.83%,GCAT-U-Net 提升了 2.05%;而在评价指标 MIoU 上,SE-U-Net 提升了 2.58%,CAT-U-Net 提升了 3.35%,GCAT-U-Net 提升了 3.72%,可以发现,本文所提出的全局坐标注意力机制,在各个评价指标上,相较于其他注意力机制来说

取得了显著的提升,对于图像的分割,也取得了较高的分割精度。

表 1 模型在 GID 测试集上的比较

方法	PA	MPA	MIoU	FWIoU
U-Net	0.904 1	0.921 1	0.819 3	0.828 2
SE-U-Net	0.917 0	0.928 8	0.840 4	0.849 3
CAT-U-Net	0.920 7	0.931 8	0.846 7	0.855 4
GCAT-U-Net	0.922 7	0.931 9	0.849 8	0.858 6

4 结束语

本文是关注遥感地块语义分割中对于 U-NET 网络的注意力机制改进,旨在提高细节处遥感图像的分割,提高分割的准确率。针对 U-Net 网络的特性,在前人研究的通道注意力机制以及坐标注意力机制的基础上,创新性的提出了全局坐标注意力机制,提高了遥感地块语义分割的准确度。全局坐标注意力机制改进的 U-Net 网络模型,相比较于单一嵌入通道注意力机制、坐标注意力机制的 U-Net 网络,在武汉大学 GID 数据集上,从不同的指标来看,都有了很大的改进,尤其对于一些原始网络中的错分、漏分和边缘粗糙等问题来说,输出结果均有所改善。但还有很多可提升的空间,例如:增加网络的泛化性以及鲁棒性;降低模型复杂程度、参数数量、运行时间等。在接下来的研究中,可以考虑多尺度模型设计,以及其他改进方式,例如 transform 等,在利用好深度学习网络模型的同时,尝试融入当下理论体系相对较完善的传统语义分割理论,增加模型的可解释性,更深层次的增进遥感地块语义分割的探索与研究。

参考文献:

[1] BRUZZONE L, CARLIN L. A multilevel context-based system for classification of very high spatial resolution images [J]. IEEE Transactions on Geoscience and Remote Sensing, 2006, 44 (9): 2587-2600.

[2] KAMPPFFMEYER M, SALBERG A B, JENSSEN R. Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks [C] // The IEEE Conference on Computer Vision and Pattern Recognition

(CVPR), IEEE, 2016.

- [3] SUN W, WANG R. Fully Convolutional networks for semantic segmentation of very high resolution remotely sensed images combined with DSM [J]. IEEE Geoscience & Remote Sensing Letters, 2018; 1-5.
- [4] MOU L, HUA Y, ZHU X X. A relation-augmented fully convolutional network for semantic segmentation in aerial scenes [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2019.
- [5] SINGH PANKAJ P, GARG R D. Automatic road extraction from high resolution satellite image using adaptive global thresholding and morphological operations [J]. Journal of the Indian Society of Remote Sensing, 2013, 41 (3): 631-640.
- [6] SHEN Z F. Extraction of road network from high resolution remote sensed imagery with the combination of Gaussian Markov random field texture model and support vector machine [J]. Journal of Remote Sensing, 2005 (3): 271-176.
- [7] EVAN S, JONATHAN L, TREVOR D. Fully convolutional networks for semantic segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39 (4): 640-651.
- [8] VIJAY B, ALEX K, ROBERTO C. SegNet: a deep convolutional encoder-decoder architecture for image segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39 (12): 2481-2495.
- [9] CHAURASIA A, CULURCIELLO E. LinkNet: exploiting encoder representations for efficient semantic segmentation [C] // 2017 IEEE Visual Communications and Image Processing (VCIP). IEEE, 2018.
- [10] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation [C] // International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015: 234-241.
- [11] OKTAY O, SCHLEMPER J, FOLGOC L L, et al. Attention u-net: Learning where to look for the pancreas [C] // arXiv Preprint arXiv: 1804.03999, 2018.

(下转第 236 页)