

基于 Word Embedding 的遥感影像检测分割

尤洪峰¹, 田生伟², 禹 龙³, 吕亚龙¹

(1. 新疆大学信息科学与工程学院, 乌鲁木齐 830046; 2. 新疆大学软件学院, 乌鲁木齐 830046;
3. 新疆大学网络中心, 乌鲁木齐 830046)

摘 要: 遥感影像检测分割技术通常需提取影像特征并通过深度学习算法挖掘影像的深层特征来实现. 然而传统特征(如颜色特征、纹理特征、空间关系特征等)不能充分描述影像语义信息, 而单一结构或串联算法无法充分挖掘影像的深层特征和上下文语义信息. 针对上述问题, 本文通过词嵌入将空间关系特征映射成实数密集向量, 与颜色、纹理特征的结合. 其次, 本文构建基于注意力机制下图卷积网络和独立循环神经网络的遥感影像检测分割并联算法(Attention Graph Convolution Networks and Independently Recurrent Neural Network, ATGIR). 该算法首先通过注意力机制对结合后的特征进行概率权重分配; 然后利用图卷积网络(GCNs)算法对高权重的特征进一步挖掘并生成方向标签, 同时使用独立循环神经网络(IndRNN)算法挖掘影像特征中的上下文信息, 最后用 Sigmoid 分类器完成影像检测分割任务. 以胡杨林遥感影像检测分割任务为例, 我们验证了提出的特征提取方法和 ATGIR 算法能有效提升胡杨林检测分割任务的性能.

关键词: 注意力机制; 图卷积网络; 独立循环神经网络; 并联算法; 词嵌入

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2020)01-0075-09

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2020.01.009

Remote Sensing Image Detection and Segmentation Based on Word Embedding

YOU Hong-feng¹, TIAN Sheng-wei², YU Long³, LÜ Ya-long¹

(1. School of Information Science and Engineering, Xinjiang University, Urumqi, Xinjiang 830046, China;
2. Software College, Xinjiang University, Urumqi, Xinjiang 830046, China;
3. Network Center, Xinjiang University, Urumqi, Xinjiang 830046, China)

Abstract: Remote sensing image detection and segmentation technology usually needs to extract image features and mine the deep features of images through deep learning algorithm. However, traditional imaging features (e. g., color, texture, spatial relationship) cannot fully reflect the semantic information of the images, while single/sequential algorithm cannot fully exploit the deep features and the contextual semantic information of the images. Aiming at the above challenges, in this paper, the spatial relation features are mapped into real dense vectors by word embedding, which are combined with color and texture features. Further, we propose a new parallel algorithm referred to as attention graph convolution networks and independently recurrent neural network (ATGIR) based on graph convolution network and independent recurrent neural network under attention mechanism for remote sensing image detection and segmentation. Our algorithm first assigns probabilistic weights to the combined features based on attention mechanism; then extracts deep features based on the features with high weights to generate labels with directions by using graph convolution network (GCNs) algorithms, extracts contextual semantic information of the images by using the independently recurrent neural network (IndRNN) algorithm; finally, our algorithm realizes image detection and segmentation by using Sigmoid. For remote sensing image detection and segmentation of populous euphratica forest as an instance, we prove that our feature extraction method and proposed ATGIR algorithm can effectively improve the detection and segmentation tasks.

Key words: attention mechanism; GCNs; IndRNN; parallel algorithm; word embedding

1 引言

随着遥感技术的快速发展,遥感影像的光谱和空间分辨率不断提高,其应用领域越来越广泛,包括场景分类、目标识别和影像分割等领域。目前大量影像分割检测算法都是基于传统特征,难以描述特征之间的上下文语义信息。大量研究表明递归神经网络(RNN)^[1]、长短期记忆网络(LSTM)^[2]等算法通过存储整合特征之间的上下文语义信息进一步提高了影像分割的准确率,并展现出了广阔的前景;卷积神经网络(CNN)通过挖掘深层特征提高了影像分割的准确率。虽然上述方法在影像的识别结果上获得一些成果,但并不能很好的适用于具有丰富特征的遥感影像的分割任务。因此如何更好的挖掘深层特征并整合上下文语义信息^[3]变得尤为重要。

与此相比,本文提出的 ATGIR 算法结合词嵌入技术在保留高权重特征的同时更好的挖掘遥感影像中的深层特征和整合上下文语义信息。本文的主要贡献有以下五点:

(1)结合自然语言处理方法来生成实数密集向量,进一步建立空间关系特征的上下文语义信息;

(2)通过注意力机制加强检测目标的概率权重,以获取高权重特征;

(3)使用图卷积网络 GCNs^[4]算法不仅实现深层特征提取,同时还自动构建了深层特征之间的方向标签;

(4)利用独立循环神经网络 IndRNN^[5]算法记忆更多特征间的信息和捕捉上下文语义信息;

(5)本文提出一种并联算法取代串联算法,从而保留了更多的深层特征,最终在对胡杨林遥感影像检测分割任务上取得了有效的结果。

2 相关研究

由于遥感影像本身特征的丰富性,增强特征之间的关系可以更好的表达整个影像;慕彩虹^[6]等人将原始遥感图像划分成互不重叠的小块,并对小块进行主成分分析得到特征向量,最终通过特征向量提高了分类的准确率;宋婉莹^[7]等人考虑到遥感影像样本在特征空间上的距离,提出了一种加权合成核函数权重系数的自适应确定方法,该方法不仅提高了分类结果且更好地保留了图像边缘信息;李玉^[8]等人提出了一种可变性聚类算法,根据影像的相似图构建权值矩阵和标准 Laplacians 矩阵计算出特征向量,最后通过 FCM (Fuzzy C-Means)算法实现遥感影像分割。由上述研究可知,在特征之间建立关系可以更好的实现遥感影像的分割检测任务。

深度学习在深层特征的挖掘和特征语义信息的整合上占据着不可替代的位置;Xu Y^[9]提出一个关于 LSTM 的端对端模型,对高光谱影像进行分类,该模型充分利用相邻像元之间的语义信息,实验结果有明显提升;Han D^[10]等人通过单一结构 CNN 提取有效的影像特征来表示识别目标,继而进行分类,但是 CNN 仅限于提取影像的物理特征,难以整合影像内的上下文语义信息;LIM H 等人通过 CNN 来获取深层视觉特征,再通过双向长短期记忆网络(BiLSTM)^[11]学习时域中的深层特征,并对特征进行整合,克服了 CNN 难以整合上下文语义信息的问题。实验表明串联算法因网络层次更深,对全局特征信息捕捉更完善,获得良好效果。虽然串联算法^[12]在一定程度上实现了特征联合抽取,但在串联算法中,后算法是在前算法提取特征的基础上进行二次挖掘,不能充分挖掘深层特征。LSTM、BiLSTM 等算法对长序列数据不能够充分存储特征的上下文语义信息^[5]。

针对上述研究中出现的问题,本文通过词嵌入技术对空间关系特征进行扩充和编码生成实数密集向量更好地构建特征之间关系,从而加强了特征对影像的语义描述能力;利用注意力机制^[13]对混合特征进行概率权重分配,获取高权重的特征,从而最大限度的保留了深层特征;接着通过 GCNs 算法挖掘深层特征,生成深层特征之间的方向标签,增强深层特征之间的关系;同时利用 IndRNN 算法挖掘特征的上下文语义信息并整合,解决了对上下文语义信息保留不够充分的问题。实验结果表明,所提特征和算法在对胡杨林遥感影像检测分割任务上取得了有效的结果。

3 ATGIR 算法和特征提取

3.1 ATGIR 算法

为了更好地挖掘和处理深层特征之间关系,本文提出了一种结合词嵌入的 ATGIR 算法。由输入层、注意力机制、GCNs、IndRNN 和分类层五部分组成。首先,利用词嵌入技术将空间关系特征生成实数密集向量特征,并与纹理特征和颜色特征结合;接着利用注意力机制对混合特征进行概率权重分配,获取权重较高的特征;然后图卷积算法对注意力机制生成的高权重特征做进一步提取,生成深层特征,并自动构建深层特征之间的方向关系;同时利用独立循环神经网络算法充分学习特征的上下文语义信息并整合。最后融合 GCNs 和 IndRNN 的深层特征,通过 Sigmoid 进行分类。ATGIR 算法如图 1 所示。

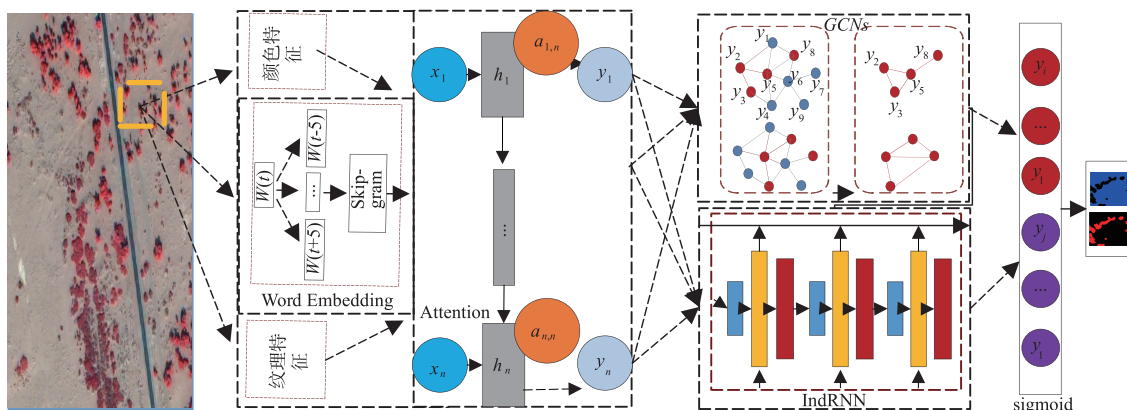


图1 ATGIR算法网络结构图

3.2 特征提取

3.2.1 词嵌入提取实数密集向量特征

Hinton^[14]提出 Word Embedding 将词映射到低维空间,解决向量稀疏问题和反映出它们在语义层面上的联系.本文中,为了充分挖掘影像像元之间的空间关系,利用词嵌入技术将影像中空间关系特征转换成实数密集向量,通过 Word2Vec 将影像中的当前像元同周围 5 个像元的空间关系特征做向量映射,以此来充分挖掘像元之间的上下文关系.如图 2 所示,其中 1 代表当前像元的空间关系特征,2 代表 1 的前后各 5 个像元的空间关系特征.

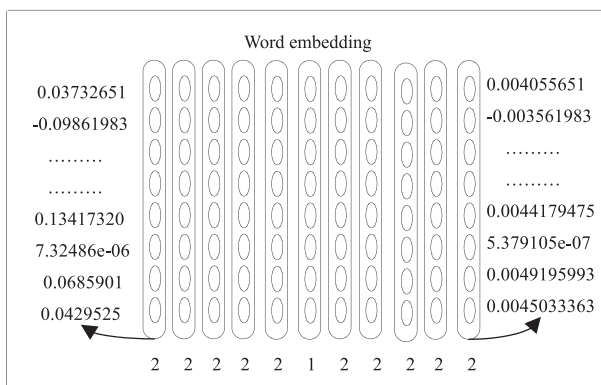


图2 Word Embedding映射结果图

左右两边的数字代表生成的实数密集向量特征.本文使用的是 Word2Vec 中的 Skip-gram^[15]模型,通过已知像元 w_t 的空间关系特征来预测上下像元 $w_{t-5}, \dots, w_{t-1}, w_t, w_{t+1}, \dots, w_{t+5}$ 空间关系特征.实数密集向量通过余弦公式(式(1))来表示空间关系相似度.

$$S(v, u) = \frac{\langle w_v \times w_u \rangle}{|w_v| \times |w_u|} \quad (1)$$

其中 w_v, w_u 代表像元的空间关系特征. $S(v, u)$ 表示余弦计算之后的相似度结果.

3.2.2 纹理特征

本文所提纹理特征分为三部分:

- ①将原图转为灰度图像;
- ②将灰度图切分成多种尺度 $N \times N$ 的矢量块;
- ③将矢量块的中心点与周围点进行比较,中心点大则返回 1,中心点小则返回 0.将 1,0 生成的特征序列作为纹理特征.

3.2.3 HSV 特征

由于 HSV 比 RGB 对颜色的敏感度要大,而且更易于进行图像数字化处理,因此本文提取了 HSV 特征.该特征提取分为两部分:

- ①将原图像切分成多种尺度 $N \times N$ 的矢量块;
- ②通过 HSV 颜色计算公式计算原 R, G, B 参数,生成 HSV 特征.

我们将原始影像切分成 $3 \times 3, 5 \times 5, 7 \times 7$ (实验部分用 9, 25, 49 表示)大小不同的图像块,并通过这些图像块生成实数密集向量特征,纹理特征、HSV 特征来表示图像块的中心像元,最终输入到算法模型中,实现对中心像元的检测.

3.3 算法意义及作用

3.3.1 注意力机制

注意力机制 (Attention Mechanism) 是一种概率权重分配机制,通过计算不同时刻下的遥感影像特征,使得检测目标的相关特征得到更多关注,从而提升隐含层影像特征的质量.注意力机制的计算公式见式(2)和式(3).具体网络构造在图 1 中的 Attention 部分给出.

$$a_{ij} = \frac{\exp(e_{ij})}{\sum_{j=1}^n \exp(e_{ik})} \quad (2)$$

$$c_i = \sum_{j=1}^n a_{ij} h_j \quad (3)$$

a_{ij} 表示隐含层各时刻权重系数, h_j 是隐含层的信息, e 代表第 i 时刻的能量值, c_i 代表 i 这个点生成的上下文关系.

3.3.2 图卷积网络

注意力机制输出特征作为图卷积网络的输入特

征. GCNs 自动挖掘深层特征,通过隐藏层单元数量来确定深层特征的数量,并在深层特征之间生成特征方向标签来建立特征关系,第一层图卷积公式(式(4)).

$$h_v^1 = \text{ReLU} \left(\sum_{u \in N(v)} (W x_u + b) \right) \quad (4)$$

h_v^1 表示第一层图卷积特征输出; x_u 表示 Attention Layer 的输出; W 表示这一层算法权重; b 表示偏置函数; ReLU 代表激活函数.

通过叠加图卷积网络的层数可以学习到更深层的特征,将上层输出 h_v^1 作为下一层的输入. 如有 k 层 GCNs, 第 k 层图卷积的公式见式(5).

$$h_v^{k+1} = \text{ReLU} \left(\sum_{u \in N(v)} (W^{(k)} h_u^{(k)} + b^{(k)}) \right) \quad (5)$$

当特征向量存在多个较高的权重特征时,设其两特征分别为 $A1, A2$. 则 $(A1, A2)$ 和 $(A2, A1)$ 是这两个特征之间建立的方向关系. 图卷积网络将它们自动生成的方向标签分别记作 $L(A1, A2)$ 和 $L(A2, A1)$, 同时特征本身在数据中都存在着自己固定的位置,由此会产生一个 Self 标签. 改进之后如式(6). Label 表示上面生成的 3 种标签. 如图 3 所示.

$$h_v^{k+1} = \text{ReLU} \left(\sum_{u \in N(v)} (W_{L(\text{label})}^{(k)} h_u^{(k)} + b_{L(\text{label})}^{(k)}) \right) \quad (6)$$

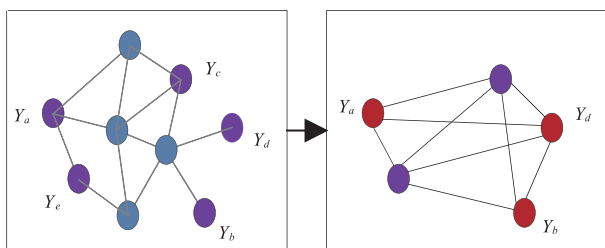


图3 图卷积网络特征选择图

图 3 左图中紫色圆 Y_a 等和蓝色圆表示注意力机制下概率权重较高的特征,特征关系两两相连,而 Y_a 等特征相比于蓝色圆概率权重较高,将作为右图的特征输入. 右图中 3 个红色圆概率权重较高,因此将作为下一层的特征输入,以此来挖掘更深层的影像特征.

3.3.3 独立循环神经网络

LSTM 和 GRU^[13] 在处理长文本数据时存在丢失序列前面特征的问题,且随着算法层数的堆叠, LSTM、GRU 中使用双曲正切和 Sigmoid 函数会使得 LSTM、GRU 的梯度衰减,同时 LSTM 和 GRU 算法中同层的所有神经元会相互影响,增加冗余特征,从而降低算法检测分割的准确率. 为了解决上述问题,本文使用 IndRNN. 式(7)如下.

$$h_t = \sigma(\omega x_t + \lambda \cdot h_{t-1} + b) \quad (7)$$

其中: λ 代表循环权重向量; \bullet 代表阿达马乘积; σ 代表了激活函数; ω 表示权重; h_{t-1} 表示 $t-1$ 时刻的隐藏层状态; x_t 为注意力机制输出的高权重特征.

IndRNN 算法中每一层的每个神经元各自独立,每个神经元仅接收前一时间步的输入信息和它自己当前时刻的隐藏状态信息,因此不会被同一时刻的特征间的关系所影响,消除了同一时刻的冗余特征. 同时 IndRNN 通过调节时间梯度反向传播解决该网络梯度衰减的问题,使得网络处理长序列数据效果明显. IndRNN 的激活函数采用非饱和激活函数 ReLU 以增强鲁棒性. 式(8)如下.

$$h_{t,n} = \sigma(\omega_n x_t + \lambda_n \cdot h_{t-1,n} + b_n) \quad (8)$$

ω_n, λ_n 分别代表第 n 行输入的当前权重. 基本框架如图 4 所示^[5], 通过 weight 和激活函数 ReLU 对每一步骤的特征进行运算和循环处理,通过 Batch Normalization (BN) 对每一步特征进行归一化处理.

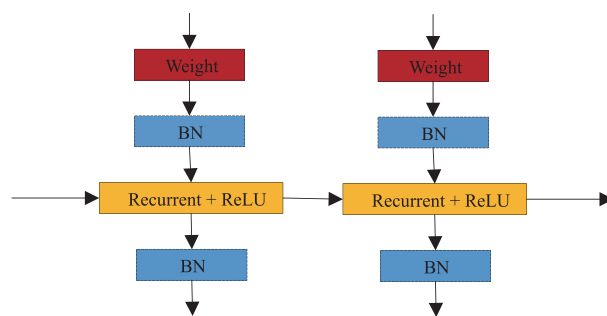


图4 独立循环神经网络

3.4 特征融合

提取 IndRNN 算法最终特征输出结果 y_i , 提取 GCNs 算法最终特征输出结果 y_j , 融合两算法特征, 最后通过 Sigmoid 进行分类. 式(9)如下.

$$\text{Merge}([y_i, y_j]) \quad (9)$$

4 实验结果与分析

4.1 实验数据

本文数据来源于两组 Quick Bird 卫星拍摄的遥感影像数据集: (1) 轮台县胡杨林群; (2) 塔里木河胡杨林群. 为了提高遥感影像的质量, 在实验前先对遥感影像进行大气校正、几何校正和锐化等^[16] 预处理. 实验数据中训练集和测试集的样本比例 4:1; 本文研究的是一个二分类问题: 正样本是胡杨林, 负样本是背景 (沙漠、树影、公路等等); 数据总量是 165336 个图像块.

4.2 实验评价标准

本实验使用多种评价指标对实验进行验证如: Accuracy (Acc) 和 Precision (P)、Recall (R)、F-Measure (F). Acc 是准确率; P 值是指精确率; R 值是指召回率; F 值是综合考虑正负样本的一个评价指标, 可以更好的反映算法检测效果. (TN 真负例, FN 假负例, FP 假正例, TP 真正例) 具体见式 (10) ~ (13).

$$\text{Acc} = (TP + TN) / (TP + FN + FP + TN) \quad (10)$$

$$P = TP / (TP + FP) \quad (11)$$

$$R = TP / (TP + FN) \quad (12)$$

$$F = (2P * R) / (P + R) \quad (13)$$

4.3 实验算法超参数设置

超参数的设置会影响算法对特征的挖掘和利用^[17], 实验结果表明当参数设置如表 1 时分割结果最佳.

表 1 超参数设置

类型	超参数
Adam	lr = 0.01 bata_1 = 0.9 bata_2 = 0.999 Epsilon = 1^{e-08} Decay = 3^{e-8}
Attention	bias = True
Parallelmodel-graphConv	filters = 128/128/128 neighbors_ix_mat = 8/4/2 num_neighbors = 25/16/4
Maxpooling	pool_size = 2/2
IndRNN	units = 64/64/64 recurrent_clip_min = -1/-1/-1 recurrent_clip_max = -1/-1/-1 dropout = 0.35
Dense 1 layer	Units = 64
Dense 2 layer	Units = 2

本文使用简单高效、对参数更新时不受梯度伸缩变化影响的 Adam 优化器,其参数:lr 表示学习率为 0.01;bata_1 表示一阶矩估计的指数衰减率为 0.9;bata_2 表示二阶矩估计的指数衰减率 0.999;Epsilon 为了防止分母为 0,将 1^{e-08} 来代替 0 值;Decay 表示每次参数更新后学习率衰减值为 3^{e-8} . Attention 算法中 bias 函数设置为 0 值. 本文使用了 3 层图卷积,其参数:neighbors_ix_mat 表示当前像元特征与周边 n 的平方个像元特征建立联系;num_neighbors 表示在 n 平方个像元特征中保留 n^2 个像元关键特征. 本文使用了 3 层的 IndRNN 的参数有:units 表示神经元个数;recurrent_clip_min = -1 和 recurrent_clip_max = -1 表示使用默认的时间步长.

4.4 数据特征对实验的影响

实验使用三种特征:颜色特征 HSV、纹理特征 T 和空间关系特征 S, 其中将空间关系特征 S 通过 Word2Vec 映射成向量用 S_Word2Vec 表示. 为了验证所提特征的有效性,将 HSV、S、T、S_Word2Vec 和全特征 All_Feature (S_Word2Vec + T + HSV) 作为模型输入. 以下实验都在两组数据集上利用多种评价指标进行验证. 轮台县胡杨林群数据设为 R1, 塔里木河胡杨林群数据设为 R2. 具体实验结果如表 2 所示.

表 2 不同特征的实验结果

模型	R1			
	Acc	P	R	F
HSV	96.43%	90.40%	83.29%	86.70%
S	88.0%	55.07%	51.67%	53.32%
T	96.53%	88.95%	81.64%	85.14%
S_Word2vec	90.42%	58.99%	56.44%	57.69%
All_Feature	97.43%	91.61%	84.77%	88.06%
模型	R2			
	Acc	P	R	F
HSV	97.73%	84.32%	83.52%	83.92%
S	92.72%	74.67%	82.5%	78.39%
T	96.15%	73.40%	71.15%	72.26%
S_Word2vec	94.58%	72.49%	91.03%	80.71%
All_Feature	98.23%	81.95%	89.83%	85.71%

ATGIR 算法根据不同的特征学习到不同的信息. 由表 2 可以看出,本文提出的 S_Word2vec 特征在实验结果上看明显优于 S 特征. All_Feature 特征可以更好的表达图像.

4.5 算法对实验的影响

为了验证 ATGIR 算法的有效性,本小段使用特征 All_Feature 进行实验. 其中注意力机制用 A 表示. 具体实验结果如表 3 所示.

表 3 不同算法的实验结果

模型	R1			
	Acc	P	R	F
CNN	96.09%	89.79%	81.27%	85.32%
LSTM	95.57%	93.61%	77.07%	84.54%
CNN-LSTM	96.31%	91.51%	81.60%	86.27%
P-CNN-LSTM	96.94%	93.48%	80.85%	86.71%
A_GCNs	96.43%	93.08%	74.85%	82.98%
A_IndRNN	96.58%	82.42%	84.31%	86.28%
A_GCNs-IndRNN	96.52%	88.34%	83.96%	85.94%
ATGIR	97.43%	91.61%	84.77%	88.06%
模型	R2			
	Acc	P	R	F
CNN	97.03%	89.34%	80.56%	84.72%
LSTM	96.85%	92.36%	71.89%	80.85%
CNN-LSTM	97.28%	86.49%	78.42%	82.26%
P-CNN-LSTM	97.8%	92.52%	76.08%	83.50%
A_GCNs	97.26%	92.36%	75.28%	82.95%
A_IndRNN	97.53%	88.00%	78.91%	83.21%
A_GCNs-IndRNN	97.86%	85.07%	84.25%	84.66%
ATGIR	98.23%	81.95%	89.83%	85.71%

通过表3中的 Acc 和 F 值评价指标可以看出,相比于 CNN、LSTM 这类单一结构算法,CNN-LSTM、A_GC-Ns、A_IndRNN、A_GCNs-IndRNN 的串联算法表现出更好的性能.由于并联算法保留了更多的深层特征,实验结果可以看出并联算法(P-CNN-LSTM、ATGIR)比串联算法结果更优,学习能力更强.本文提出的 ATGIR 算法在 Acc 和 F 值两项综合指标均优于其他算法,证明了本文所提算法的有效性.生成的效果如图5所示.

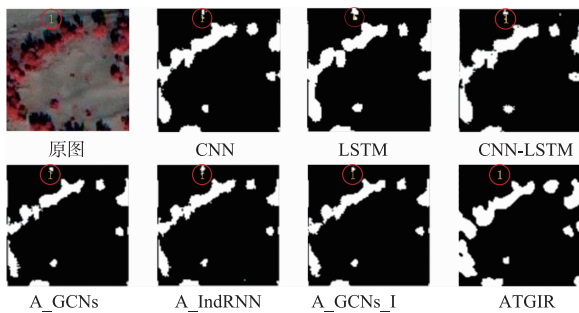


图5 不同算法的效果图

图5中不同算法生成的效果图可以明显看出:“1”标记的区域在原图中没有胡杨林,但由于单一结构算法和串联算法不足以充分挖掘深层特征,从而产生了噪声点,而本文所提算法效果更优.

本文所提 ATGIR 算法与各算法相比,由于 Acc 结果优势不明显,因此我们通过 McNemars 检测来进一步

评估各模型之间的优劣.具体评测结果如表4所示.

表4 McNemars 检测结果

模型	R1	R2
	卡方值	卡方值
CNN	86	94
LSTM	101	101
CNN-LSTM	73	75
PI-CNN-LSTM	55	65
A_GCNs	48	36
A_IndRNN	50	32
A_GCNs-IndRNN	46	24
ATGIR	35	15

所有算法的预测结果与影像标签结果做 McNemars 检测.针对二分类问题,我们选取错误假设率为 $\alpha = 0.05$,自由度取值为1,则 McNemars 卡方值越接近3.84说明算法效果越好.由表4可以非常直观看出我们所提的算法在胡杨林检测分割任务中性能最优.

4.6 数据量对实验结果的影响

图像块大小对实验结果起着一定的影响,本文将数据集切成9、25、49大小不同的图像块进行实验.注意力机制、GCNs 和 IndRNN 组成的串联算法简写成 A_G_I.其中注意力机制用 A 表示.具体实验结果如表5和图6所示.

表5 不同数据量的实验结果

模型	R1				R2			
	Acc	P	R	F	Acc	P	R	F
9_CNN	92.6%	77%	82.83%	79.81%	94.8%	95.25%	64.60%	76.99%
9_LSTM	93.01%	81.73%	77.63%	79.63%	94.38%	94.06%	64.76%	76.71%
9_CNN-LSTM	93.21%	80.27%	82.29%	81.27%	95.08%	91.73%	69.37%	79.0%
9_P-CNN-LSTM	93.83%	92.3%	72.26%	81.06%	96.61%	90.23%	73.02%	80.72%
9_A_GCNs	93.21%	79.72%	81.35%	80.53%	94.68%	93.16%	64.82%	76.45%
9_A_IndRNN	93.28%	82.59%	78.97%	80.74%	95.35%	93.07%	69.23%	79.40%
9_A_G_I	94.01%	86.09%	79.62%	82.73%	95.46%	90.26%	71.54%	79.82%
9_ATGIR	94.7%	92.53%	77.14%	84.14%	96.63%	82.95%	79.84%	81.37%
25_CNN	94.99%	93.13%	74.37%	82.70%	96.1%	94.49%	68.18%	79.21%
25_LSTM	94.87%	92.69%	73.85%	82.21%	95.65%	96.16%	66.33%	78.51%
25_CNN-LSTM	96.14%	90.46%	82.39%	86.24%	96.3%	92.01%	71.69%	80.59%
25_P-CNN-LSTM	96.35%	87.68%	82.89%	85.22%	97.16%	83.28%	81.59%	82.43%
25_A_GCNs	95.91%	89.37%	81.71%	85.37%	96.23%	91.82%	71.95%	80.68%
25_A_IndRNN	96.15%	88.05%	83.71%	85.83%	96.28%	92.02%	70.97%	80.14%
25_A_G_I	96.36%	87.27%	85.54%	86.40%	96.35%	93.13%	72.07%	81.26%
25_ATGIR	96.8%	92.85%	82.84%	87.56%	97.2%	85.81%	81.95%	83.84%

续表

模型	R1				R2			
	Acc	P	R	F	Acc	P	R	F
49_CNN	96.09%	89.79%	81.27%	85.32%	97.03%	89.34%	73.35%	80.56%
49_LSTM	95.57%	93.61%	77.07%	84.54%	96.85%	92.36%	71.89%	80.85%
49_CNN-LSTM	96.31%	91.51%	81.59%	86.27%	97.28%	86.49%	78.42%	82.26%
49_P-CNN-LSTM	96.94%	93.48%	80.85%	86.71%	97.8%	92.52%	76.08%	83.50%
49_A_GCNs	96.43%	93.08%	74.85%	82.98%	97.26%	92.36%	75.28%	82.95%
49_A_IndRNN	96.58%	88.34%	84.31%	86.28%	97.53%	88.00%	78.91%	83.21%
49_A_G_I	96.52%	88.01%	83.96%	85.94%	97.86%	85.07%	84.25%	84.66%
49_ATGIR	97.43%	91.61%	84.77%	88.06%	98.23%	81.95%	89.83%	85.71%

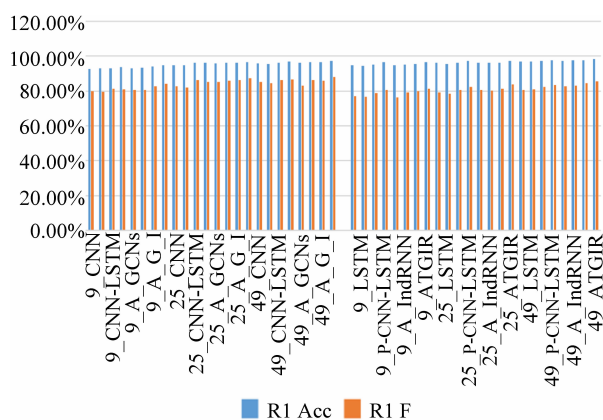


图6 数据量对实验的影响

由表 5 可知,将影像切成 49 的图像块时,Acc 和 F 值最高,表明图像块保留特征最多.当切成 9、25 的图像块时,图像块保留信息不完整,导致实验结果较差.由图 6 可以直观看出,当影像被切成 49 的图像块时,效果高于其他尺寸的图像块.生成的效果如图 7 所示.

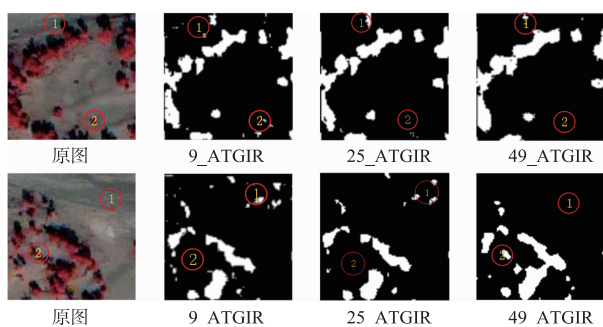


图7 不同矢量图像块的效果图

由图 7 中“1”标记的区域可以看出,当影像被切成 9 的图像块时,效果图中的噪声点较多,9 图像块保留的特征和语义信息较少,导致分割效果较差;当影像被切成 25 的图像块时,可以明显看出噪声点数量较少,可知随着图像块尺度的增大,特征和语义信息增多,噪声

点随之减少;当影像被切成 49 的图像块时,噪声点明显消除,证明了 49 的图像块可以更好的保留特征和语义信息. “2”标记的区域可以看出,由于 9 和 25 的图像块中保留的特征和语义信息较少,导致了最终保留的深层特征和语义信息较少,对小目标分割效果较差.而 49 图像块保留大量的特征,使得实验效果更好,从而验证了数据量对实验结果的影响.

4.7 网络层数对实验的影响

网络层数对实验结果有着至关重要影响.测试 GCNs 层数时保持 IndRNN 层数不变,测试 IndRNN 层数时保持 GCNs 层数不变.具体实验结果如表 6 所示.

表 6 不同层数的实验结果

模型	R1			
	Acc	P	R	F
GCNs_1layer	96.65%	87.86%	84.67%	86.24%
GCNs_2layers	96.8%	89.31%	84.57%	86.88%
GCNs_4layers	96.83%	88.99%	84.62%	86.75%
IndRNN_11ayer	96.45%	89.56%	80.77%	84.94%
IndRNN_2layers	96.71%	87.18%	87.30%	87.24%
IndRNN_4layers	96.95%	89.17%	84.23%	86.63%
ATGIR	97.43%	91.61%	84.77%	88.06%
模型	R2			
	Acc	P	R	F
GCNs_1layer	97.75%	83.32%	83.94%	83.63%
GCNs_2layers	97.85%	86.45%	82.14%	84.24%
GCNs_4layers	97.83%	86.85%	79.44%	82.98%
IndRNN_11ayer	97.6%	87.77%	78.23%	82.73%
IndRNN_2layers	97.78%	85.76%	80.92%	83.27%
IndRNN_4layers	97.98%	83.2%	85.18%	84.18%
ATGIR	98.23%	81.95%	89.83%	85.71%

由表 6 可知,当 GCNs 层数和 IndRNN 层数都为 3 层

时,实验效果最好.实验结果表明,当网络层数较少时,冗余特征使得实验效果变差;当网络层数较多时,丢失一些深层特征和语义信息,从而降低检测分割的准确率.

5 总结与展望

本文提出了一种结合词嵌入技术(Word Embedding)生成实数密集向量特征提取方法,增强了对遥感影像的语义描述能力.构建并联算法 ATGIR 加强对深层特征挖掘和上下文语义信息整合.利用注意力机制加强特征的概率权重,获取高权重特征,从而最大限度的保留了深层特征;随后通过 GCNs 算法进一步挖掘深层特征,生成深层特征之间的方向标签,自动构建各深层特征之间的关系;同时通过 IndRNN 算法挖掘特征的上下文语义信息并整合,解决了对上下文语义保留不够充分的问题,最终有效地提升了胡杨林检测分割任务的性能.

本文所提特征和方法为今后的遥感影像检测分割提出了新思路.我们下一步工作将对遥感小目标分割的测进行研究.

参考文献

- [1] WEI H, ZHANG H, GAO G. Representing word image spotting on historical document images[A]. IEEE International Conference on Multimedia and Expo [C]. US: IEEE, 2017. 1368 – 1373.
- [2] TANG P, WANG H, KWONG S. Deep sequential fusion LSTM network for image description [J]. Neurocomputing, 2018, 312: 54 – 164.
- [3] URICCHIO T, BALLAN L, SEIDENARI L, et al. Automatic image annotation via label transfer in the semantic space[J]. Pattern Recognition, 2017, 71: 144 – 157.
- [4] MARCHEGGIANI D, TITOV I. Encoding sentences with graph convolutional networks for semantic role labeling [A]. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing [C]. US: ACL, 2017. 1506 – 1515.
- [5] LI S, LI W, COOK C, et al. Independently recurrent neural network (indrnn): Building a longer and deeper RNN [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. US: IEEE, 2018. 5457 – 5466.
- [6] 慕彩红, 霍利利, 刘逸, 等. 基于小波融合和 PCA-核模糊聚类的遥感图像变化检测 [J]. 电子学报, 2014, 43(7): 1375 – 1381.
MU Cai-hong, HUO Li-li, LIU Yi, et al. Change detection for remote sensing images based on wavelet fusion and PCA-kernel fuzzy clustering [J]. Acta Electronica Sinica, 2014, 43(7): 1375 – 1381. (in Chinese)
- [7] 宋婉莹, 李明, 张鹏, 等. 基于加权合成核与三重 Markov 场的极化 SAR 图像分类方法 [J]. 电子学报, 2015, 44(3): 520 – 526.
SONG Wan-ying, LI Ming, ZHANG Peng, et al. A classification method of PolSAR image based on weighted composite kernel and triplet Markov field [J]. Acta Electronica Sinica, 2015, 44(3): 520 – 526. (in Chinese)
- [8] 李玉, 袁永华, 赵雪梅. 可变类谱聚类遥感影像分割 [J]. 电子学报, 2018, 46(12): 3021 – 3028.
LI Yu, YUAN Yong-hua, ZHAO Xue-mei. Spectral clustering of variable class for remote sensing image segmentation [J]. Acta Electronica Sinica, 2018, 46(12): 3021 – 3028. (in Chinese)
- [9] XU Y, DU B, ZHANG L, et al. A band grouping based LSTM algorithm for hyperspectral image classification [A]. Chinese Conference on Computer Vision [C]. China: CCF, 2017. 421 – 432.
- [10] HAN D, LIU Q, FAN W. A new image classification method using CNN transfer learning and web data augmentation [J]. Expert Systems with Applications, 2018, 95: 43 – 56.
- [11] LIM H, RHYU S, LEE K. Chord Generation from Symbolic Melody Using BLSTM Networks [OL]. <https://arxiv.org/abs/1712.01011v1>, 2017.
- [12] CLICHE M. BB_twtr at SemEval-2017 Task 4: Twitter sentiment analysis with CNNs and LSTMs [A]. Proceedings of the 11th International Workshop on Semantic Evaluation [C]. US: SemEval, 2017. 573 – 580.
- [13] ROZENTAL A, FLEISCHER D. AMOBEE at SemEval-2018 Task 1: GRU neural network with a CNN attention mechanism for sentiment classification [A]. Proceedings of the 12th International Workshop on Semantic Evaluation [C]. US: SemEval, 2018. 218 – 225.
- [14] MIWA M, BANSAL M. End-to-end relation extraction using LSTMs on sequences and tree structures [A]. Association for Computational Linguistics [C]. US: ACL, 2016. 1105 – 1116.
- [15] KIM A Y, HA J G, CHOI H, et al. Automated text analysis based on skip-gram model for food evaluation in predicting consumer acceptance [A]. Computational Intelligence and Neuroscience [C]. US: Hindawi, 2018. ID9293437.
- [16] ALEXANDER ARIZA. Empirical line model for the atmospheric correction of Sentinel-2A MSI images in the caribbean islands journal [J]. European Journal of Remote Sensing, 2018, 51(1): 765 – 776.
- [17] MA X, HOVY E. End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF [A]. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics [C]. US: ACL, 2016. 64 – 74.

作者简介



尤洪峰 男,1991 年 6 月生,江苏省南通人. 现为新疆大学硕士研究生,主要研究方向为人工智能和图像处理.

E-mail:1053109177@qq.com



田生伟(通信作者) 男,1973 年 10 月出生,新疆乌鲁木齐人. 教授、博士生导师. 现为新疆大学软件学院院长、新疆计算机学会、电子学会、软件学会理事,主要研究方向为人工智能、大数据分析、图像处理.

E-mail:tianshengwei@163.com



禹龙 女,1974 年 10 月出生,新疆乌鲁木齐人. 教授、博士生导师. 现为网络与信息技术中心网络部主任,主要研究方向为人工智能、网络空间安全、自然语言处理.