

## 遥感目标空间关系检测方法

农元君, 王俊杰\*, 赵雪冰, 张军航, 耿慧, 徐晓东

中国海洋大学工程学院, 山东 青岛 266100

**摘要** 当前基于深度学习的遥感目标检测方法只能识别出遥感目标的类别及位置, 无法检测出遥感目标间的空间关系。针对该问题, 提出一种遥感目标空间关系检测方法。首先采用卷积神经网络构建视觉模块, 以提取出遥感图像中的视觉特征; 其次构建语义模块将提取出的视觉特征映射至语义嵌入空间中, 以实现遥感目标视觉特征和语义特征的深度融合; 最后将 Softmax 函数和视觉一致性损失函数引入传统的三元组损失函数中, 设计一种改进的三元组损失函数。在自制的遥感目标空间关系检测的数据集中采用所提方法进行实验。实验结果表明, 在前 20、50 和 100 个的预测结果中, 所提方法的召回率分别为 76.32%、78.54% 和 81.47%, 说明所提方法具有良好的空间关系检测性能, 能够精准检测出遥感图像中的遥感目标及其空间关系。

**关键词** 遥感; 遥感图像; 空间关系检测; 卷积神经网络; 视觉模块; 语义模块

中图分类号 TP751; TP183

文献标志码 A

doi: 10.3788/AOS202141.1628001

## Spatial Relationship Detection Method of Remote Sensing Objects

Nong Yuanjun, Wang Junjie\*, Zhao Xuebing, Zhang Junhang, Geng Hui, Xu Xiaodong

*School of Engineering, Ocean University of China, Qingdao, Shandong 266100, China*

**Abstract** Current remote sensing target detection methods based on deep learning can only identify the type and location of remote sensing targets, but cannot detect the spatial relationship between remote sensing targets. Aiming at this problem, a method for detecting the spatial relationship of remote sensing targets is proposed. First, a convolutional neural network is used to construct a vision module to extract the visual features in the remote sensing image. Second, a semantic module is constructed to map the extracted visual features to the semantic embedding space to achieve the deep fusion of the visual features and semantic features of the remote sensing target. Finally, the Softmax function and the visual consistency loss function are introduced into the traditional triplet loss function, and an improved triplet loss function is designed. The proposed method is used to conduct experiments on the self-made remote sensing target spatial relationship detection dataset. The experimental results show that among the top 20, 50 and 100 prediction results, the recall rates of the proposed method are 76.32%, 78.54% and 81.47%, respectively, indicating that the proposed method has good spatial relationship detection performance and can accurately detect remote sensing objects and their spatial relationships in remote sensing images.

**Key words** remote sensing; remote sensing images; spatial relationship detection; convolution neural network; visual module; semantic module

**OCIS codes** 280.4788; 100.3008; 110.2960

## 1 引言

近年来, 随着遥感技术的快速发展, 高分辨率遥感图像数据日益丰富, 极大地推动了遥感领域的应

用研究。遥感图像中存在飞机、船舶和桥梁等重要目标, 对遥感图像中的目标进行识别并检测出遥感目标间的空间关系, 在国防安全、港口管理、交通疏导、海上救援等军用和民用领域具有广泛的应用价

收稿日期: 2021-02-22; 修回日期: 2021-03-02; 录用日期: 2021-03-18

基金项目: 山东省重点研发计划 (2019GHY112081)

通信作者: \*wj@ouc.edu.cn

值。如在海上救援中,通过识别遇难船只及其位置并检测出遇难船只与地面救援中心间的方位和距离等空间关系,可以辅助地面救援人员制定救援计划和展开救援行动,因此实现遥感目标的空间关系检测具有重要的意义。

目前关于遥感目标的研究侧重于采用基于深度学习的目标检测方法对遥感目标进行识别和检测,如戴媛等<sup>[1]</sup>采用改进的 Faster R-CNN 检测方法<sup>[2]</sup>,实现了多场景下的遥感目标检测任务。韩永赛等<sup>[3]</sup>提出了一种优化的区域卷积神经网络检测方法,实现了对机场区域目标的自动检测。王俊强等<sup>[4]</sup>采用改进的 SSD 方法<sup>[5]</sup>对遥感小目标进行检测,取得了较好的检测效果。杜泽星等<sup>[6]</sup>采用密集连接网络对遥感图像目标进行检测,该网络对多尺度遥感目标具有良好的检测性能。鞠默然等<sup>[7]</sup>采用改进的 YOLOv3 方法<sup>[8]</sup>对遥感小目标进行检测,取得了较高的检测精度。尽管采用基于深度学习的目标检测方法可实现对遥感目标的自动识别和检测,但只能识别出单一的遥感目标,无法检测出不同遥感目标间的空间关系,不能充分理解遥感图像中更深层次的空间语义信息,故具有一定的局限性。

与目标检测不同,空间关系检测<sup>[9-10]</sup>通过对遥感图像的视觉特征和语义特征进行建模,采用基于卷积神经网络的视觉模块对遥感图像进行特征提取,以获取丰富的遥感目标区域特征、位置特征和目标类别等视觉特征。将提取出的视觉特征输入语义模块中以映射至更高层次的语义嵌入空间,经过视觉特征和语义特征的深度融合,最终不仅能识别出遥感目标的类别及位置,还能检测出遥感目标间的空间关系,实现了对遥感图像中视觉信息与空间信息的深度感知和理解。

当前关于遥感目标空间关系检测的研究较为匮乏,无法满足遥感目标空间关系检测的需求。鉴于此,本文提出一种基于深度学习的遥感目标空间关系检测方法。该方法采用基于卷积神经网络的视觉模块对遥感图像进行特征提取,并构建语义模块将提取出的视觉特征映射至语义嵌入空间中,同时采用 Softmax 函数和视觉一致性损失函数对传统三元组损失函数进行改进。实验结果表明,所提方法可以取得较高的准确率,具有良好的视觉和空间理解性能,能够精准地检测出遥感目标及其空间关系。

## 2 遥感目标空间关系检测模型

受到文献[11]的启发,提出一种基于深度学习的遥感目标空间关系检测(Remote Sensing Spatial Relationship Detection, RSSRD)模型,其网络结构如图 1 所示,该模型由视觉模块和语义模块组成,其中  $z^s$ 、 $z^p$  和  $z^o$  分别表示卷积神经网络(CNN)所提取出的主体、客体和关系特征,  $w$  表示全连接层,  $h$  表示隐含特征,  $x$  表示嵌入特征,  $L$  表示损失函数。首先采用视觉模块对输入的遥感图像进行特征提取,以获取遥感图像中丰富的遥感目标区域特征、位置特征和目标类别等视觉特征;其次将提取的视觉特征输入语义模块中以映射至语义嵌入空间,经过视觉特征和语义特征的深度融合,最终输出遥感目标空间关系的检测结果。在训练阶段,将 Softmax 函数和视觉一致性损失函数引入传统三元组损失函数中,为此设计一种改进的三元组损失函数,采用该函数对视觉和语义嵌入信息进行匹配,可以获得更佳的模型训练效果。在测试阶段,将遥感目标及其关系词向量输入模型中,并采用最近邻搜索策略进行搜索,以获得遥感目标空间关系的检测结果。

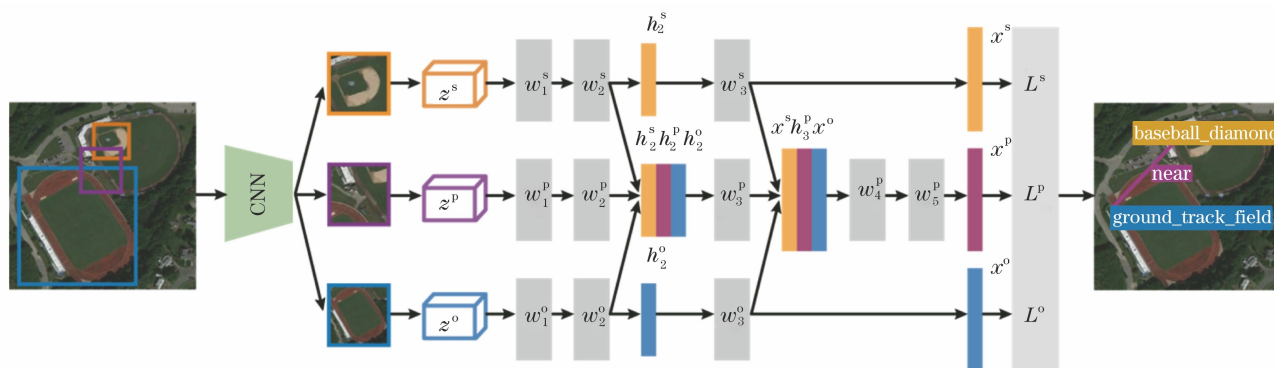


图 1 遥感目标空间关系检测模型的网络结构

Fig. 1 Network structure of remote sensing spatial relationship detection model

## 2.1 视觉模块

精准地识别出遥感图像中的目标是正确检测不同遥感目标空间关系的基础,因此 RSSRD 模型中的视觉模块采用卷积神经网络并对输入的遥感图像进行特征提取,可以获取遥感图像中丰富的目标区域特征、位置特征和目标类别等视觉特征,并将视觉特征中的局部信息和全局信息进行深层次传递,使网络充分地学习和理解遥感目标的视觉特征。图 1 中的视觉模块采用卷积神经网络并对输入的遥感图像进行特征提取,其次将提取的主体、客体 and 关系特征  $z^s$ 、 $z^o$  和  $z^p$  与对应的目标区域进行 ROI (Region of Interest) 池化,并分别输入两个全连接层中进行推理,可以获得三个中间隐含特征  $h_2^s$ 、 $h_2^o$  和  $h_2^p$ 。为了进一步增强所提取特征的显著性,对于主体和客体分支,增加全连接层  $w_3^s$  和  $w_3^o$  并对中间隐含特征  $h_2^s$  和  $h_2^o$  进行进一步处理,可以获取主客体嵌入特征  $x^s$  和  $x^o$ 。对于关系分支,采用两级特征融合策略对关系特征进行增强。首先将三个隐含特征  $h_2^s$ 、 $h_2^o$  和  $h_2^p$  进行第一次融合并输入全连接层  $w_3^p$  中,可以获得更高层次的隐含特征  $h_3^p$ ,其次将主客体嵌入特征  $x^s$  和  $x^o$  与隐含特征  $h_3^p$  进行第二次融合并输入两个全连接层  $w_4^p$  和  $w_5^p$  中,从而得到关系嵌入特征  $x^p$ 。

## 2.2 语义模块

在语义特征方面,语义模块通过将视觉特征映射至比原始词向量空间更具判别性却保持了语义相似性的语义嵌入空间中,使网络学习到更高层次的语义特征信息,实现视觉特征和语义特征的深度融合。与视觉模块类似,语义模块中的主客体分支共享权重,而关系分支则保持独立,使网络能充分地学习视觉特征和关系特征,以提高模型的鲁棒性和泛化性。在训练阶段,将每个主客体关系的真实标签以及负类标签输入至语义模块中进行训练;在测试阶段,将所有主客体关系标签输入语义模块中并采用最近邻搜索策略进行搜索,以获得遥感目标空间关系的检测结果。

在遥感目标空间关系的语义空间较大而训练数据量有限的情况下,提升语言先验能力有利于提升模型的空间关系检测效果。因此,引入预训练好的 word2vec 词嵌入向量<sup>[12]</sup>作为基准,并与关系级共嵌入向量和 node2vec 嵌入向量<sup>[13]</sup>相结合,以提升遥感目标空间关系的检测准确率,并丰富预测关系的多样性,提高模型的泛化性和少样本学习能力。

## 2.3 损失函数

为了使网络充分学习到视觉和语义的交叉特征,提升模型的训练效果,设计了一种改进的三元组损失函数。传统的三元组损失函数<sup>[14]</sup>可使具有相同标签的样本在嵌入空间中的距离较近,不同标签的样本在嵌入空间中的距离较远,而改进的三元组损失函数对相同标签的样本具有更强的嵌入能力。传统的三元组损失函数  $L^{\text{Tr}}$  为  $L_x^{\text{Tr}}$  分支与  $L_y^{\text{Tr}}$  分支的代数和,表达式分别为

$$L_x^{\text{Tr}} = \frac{1}{NK} \sum_i^N \sum_j^K \max[0, m + s(y_i, \bar{x}_{ij}) - s(y_i, x_i)], \quad (1)$$

$$L_y^{\text{Tr}} = \frac{1}{NK} \sum_i^N \sum_j^K \max[0, m + s(x_i, \bar{y}_{ij}) - s(x_i, y_i)], \quad (2)$$

$$L^{\text{Tr}} = L_x^{\text{Tr}} + L_y^{\text{Tr}}, \quad (3)$$

式中:  $N$  为正向 ROI 的数量;  $K$  为每个正向 ROI 采样的负样本数量;  $m$  为正负对之间的边缘距离;  $s(\cdot)$  为相似度函数;  $y$  为从训练数据集中随机选取的一个样本;  $\bar{x}$  为与  $y$  同一类别的样本,即正样本;  $x$  为与  $y$  不同类别的样本,即负样本。

传统的三元组损失函数易于训练,但存在一定的局限性。在网络训练过程中,若正负样本相似度的差距超过了边缘距离,则三元组损失函数将变为零,导致模型无法从损失函数中学习新的特征信息,网络的权重参数无法更新,训练效果不佳。为了解决该问题,将 Softmax 函数引入传统的三元组损失函数中,分别表示为

$$L_x^{\text{TrSm}} = \frac{1}{N} \sum_i^N \ln \left\{ \frac{\exp[s(y_i, x_i)]}{\exp[s(y_i, x_i)] + \sum_j^K \exp[s(y_i, \bar{x}_{ij})]} \right\}, \quad (4)$$

$$L_y^{\text{TrSm}} = \frac{1}{N} \sum_i^N \ln \left\{ \frac{\exp[s(x_i, y_i)]}{\exp[s(x_i, y_i)] + \sum_j^K \exp[s(x_i, \bar{y}_{ij})]} \right\}, \quad (5)$$

$$L^{\text{TrSm}} = L_x^{\text{TrSm}} + L_y^{\text{TrSm}}. \quad (6)$$

通过将正负样本间的相似度输入 Softmax 层中再进行逻辑回归分类,可以使正样本之间的相似度朝 1 训练,而负样本之间的相似度则朝 0 训练,有效解决了当正负样本相似度的差距超过边缘距离时所引起的三元组损失函数变为零的问题。

为了进一步使嵌入向量更具区分性,还引入了一个视觉一致性损失函数。该函数通过控制同类样本间的最小相似度大于不同类样本间的相似度,使同类样本在嵌入空间中的距离更近,不同类样本在嵌入空间中的距离更远,从而提高了嵌入向量的区分度。视觉一致性损失函数可表示为

$$L_c = \frac{1}{NK} \cdot \sum_i^N \sum_j^K \max\{0, m + s(x_i, \bar{x}_{ij}) - \min_{l \in C(i)} [s(x_i, x_l)]\}, \quad (7)$$

式中: $C(i)$ 为同类样本中正向 ROI 的集合。

为了使模型保持一定的松弛性,避免模型在训练过程中出现梯度爆炸,改进三元组损失函数的  $L_y^{\text{Tr}}$  分支引入 Softmax 函数,  $L_x^{\text{Tr}}$  分支则保持不变。因此,最终改进的三元组损失函数为

$$L = L_y^{\text{TrSm}} + \alpha L_x^{\text{Tr}} + \beta L_c, \quad (8)$$

式中: $\alpha$  和  $\beta$  为比例参数,通常  $\alpha = \beta = 1$ 。

### 3 实验与结果

#### 3.1 实验数据与平台

遥感目标空间关系的检测任务需要检测出遥感图像中的目标及其空间关系,因此用于遥感目标空间关系检测任务的数据集不仅需要标注出遥感图像中相关遥感目标的类别及坐标位置信息,还需要确定不同遥感目标间的空间关系,以使网络模型能充分提取和学习遥感图像中的视觉特征,以及理解和感知不同遥感目标间的空间关系语义特征。当前公开的遥感目标检测数据集只标注出了遥感目标的类别和坐标位置信息,只能用于遥感目标的检测任务,无法满足遥感目标空间关系检测任务的需求。

鉴于目前尚无用于遥感目标空间关系检测任务的数据集,本文构建了一个用于遥感目标空间关系检测的数据集。首先通过网上搜集以及从遥感目标检测公共数据集中选取的方式收集遥感图像,其次采用 VIA 标注软件<sup>[15]</sup>对遥感图像中的遥感目标进行标注,以获取遥感目标的类别和坐标位置信息,最后仿照视觉关系检测公共数据集 (Visual Genome<sup>[16]</sup>) 的数据格式对标注出的遥感目标进行整理,以确定不同遥感目标间的空间关系。数据集中共标注了飞机、船只、桥梁、棒球场、网球场、篮球场和田径场 7 种遥感目标,并根据遥感图像为俯视视角的特性,制定了 6 种不同遥感目标间的空间关系,分别为左右关系 (left/right)、上下关系 (up/down) 和远近关系 (far/near)。数据集共包含 300 张遥感图像,随机将其中的 60% 作

为训练集,剩余的 40% 作为测试集。

实验选择在搭载有 GPU 图形处理器的工作站中进行,工作站的配置: Ubuntu16.04 操作系统, GPU 为 GeForce GTX 1060, 内存为 16 G, CUDA 9.0, CUDNN 7.6.2。

#### 3.2 评价指标

为了评价不同模型的空间关系检测结果,采用视觉关系检测任务中常用的  $R@K'$  召回率作为评价指标。 $R@K'$  召回率表示前  $K'$  个预测结果中正确结果所占的比例,其中预测结果正确的标准为空间关系中的两顶点类别与真实值中的两顶点相同,且 IoU (Intersection-over-Union) 值大于或等于 0.5。平均准确率均值为遥感目标检测任务中常用的评价指标,但其在空间关系检测任务中会惩罚那些被模型正确预测但数据集中并未标注出的空间关系,这会导致检测结果的准确率降低,无法客观准确地评估模型的检测性能,因此平均准确率均值通常不作为空间关系检测任务的评价指标。关于  $K$  的取值,选取常用的  $R@20$ 、 $R@50$  和  $R@100$  召回率作为最终的评价指标。

#### 3.3 训练过程可视化

实验基于 PyTorch 深度学习框架进行,将提出的 RSSRD 空间关系检测模型在构建的遥感目标空间关系检测的训练集上进行训练,训练过程中网络的损失值变化曲线如图 2 所示。从图 2 可以看到,网络的损失值在训练前期较大,随着迭代次数的增加,损失值不断减小;当迭代至 1 万次时,损失值已降至 1 左右并继续保持小幅度下降;迭代至 2 万次后,损失值基本稳定在 0.8 左右,表明网络的训练效果理想,模型具有收敛性。

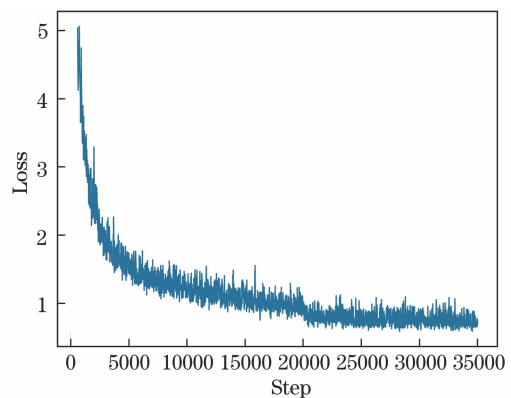


图 2 训练过程中损失值的变化曲线

Fig. 2 Change curve of loss value during training

#### 3.4 实验结果

为了验证本文所提出模型的有效性,以  $R@20$ 、

R@50 和 R@100 召回率作为评价指标,将训练好的 RSSRD 模型在构建的遥感目标空间关系检测的测试集中进行测试,并与当前关系检测性能较好的 TDE(Total Direct Effect)模型<sup>[17]</sup>进行比较,结果如表 1 所示。从表 1 可以看到,提出的 RSSRD 模型在测试集中取得了较高的精度,R@20、R@50 和 R@100 召回率分别为 76.32%、78.54% 和 81.47%,较 CTDE 模型分别提高了 3.13、2.71、2.93 个百分点,充分验证了本文模型的有效性。究其原因,所提出的 RSSRD 模型构建了视觉模块和语义模块,其中视觉模块采用具有强大空间感知能力的卷积神经网络对输入的遥感图像进行特征提取,使模型获得了丰富的遥感目标区域特征、位置特征和目标类别等视觉特征;同时,构建的语义模块将视觉特征映射至比原始词向量空间更具判别性却保持了语义相似性的语义嵌入空间中,使网络学习到更高层次的语义特征信息;此外,本文采用 Softmax 函数和视觉一致性损失函数对传统的三元组损失函数进行改进,有效解决了当正负样本相似度的差距超过边缘距离

时所引起传统三元组损失函数变为零的问题,提高了网络的学习能力。

表 1 不同方法的空间关系检测结果

Table 1 Spatial relationship detection results of different methods

Method	R@20/%	R@50/%	R@100/%
TDE	73.19	75.83	78.54
RSSRD	76.32	78.54	81.47

### 3.5 空间关系检测结果可视化

图 3 为提出的 RSSRD 模型在遥感目标空间关系检测的测试集中的部分检测效果。从图 3 可以看到,RSSRD 模型具有良好的空间关系检测性能,不仅精准地识别出了遥感目标及其位置,还准确地检测出了遥感目标间的空间关系。尤为重要的是,对于多个遥感目标的空间关系,如田径场与网球场和棒球场的空间关系,模型也能精准地进行检测和识别。检测结果表明,提出的 RSSRD 模型具有良好的视觉和空间理解性能,能充分理解遥感图像中的视觉信息和语义信息,精准地检测出遥感目标及其空间关系。

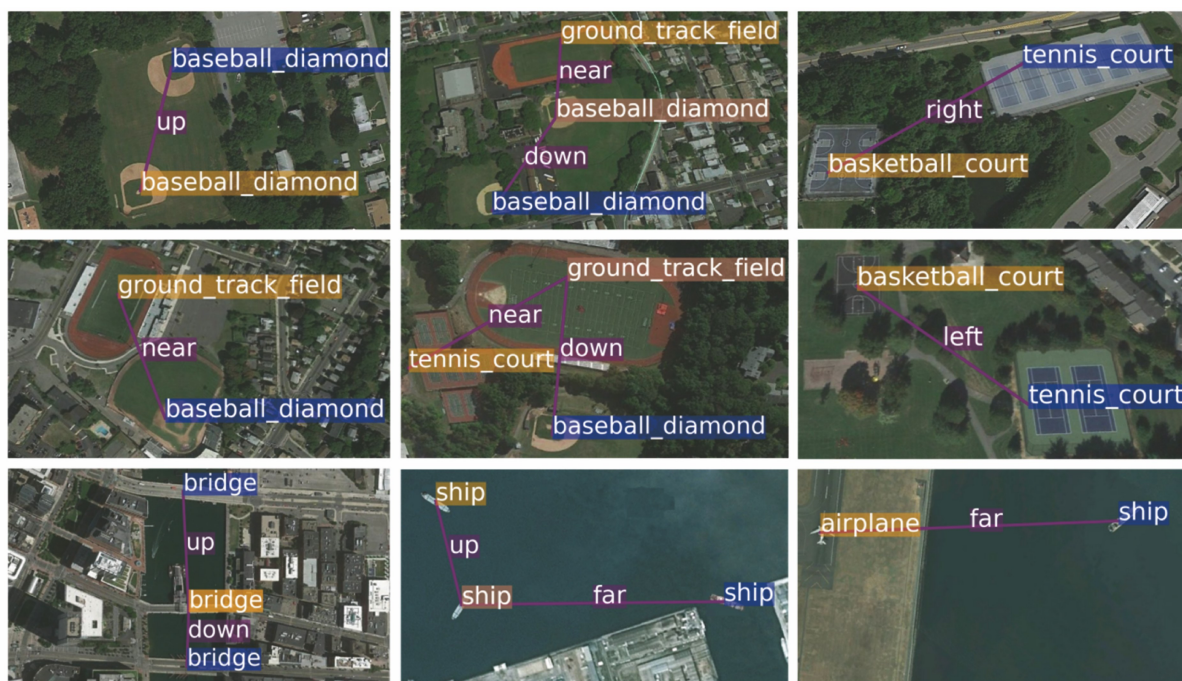


图 3 RSSRD 模型的检测结果

Fig. 3 Detection results of RSSRD model

## 4 结 论

为了实现遥感目标空间关系的检测,通过卷积神经网络来构建视觉模块和语义模块,并设计一种改进的三元组损失函数,提出了一种基于深度学习的遥感目标空间关系检测方法。实验结果表明,所

提方法在自制的遥感目标空间关系检测的数据集中取得了较高的准确率,具有良好的视觉和空间理解性能,能够准确识别和检测出遥感图像中的目标及其空间关系。下一步将对数据集进行扩充和完善,以提升模型在复杂场景下遥感目标空间关系的检测性能。

## 参 考 文 献

- [1] Dai Y, Yi B S, Xiao J S, et al. Object detection of remote sensing image based on improved rotation region proposal network [J]. *Acta Optica Sinica*, 2020, 40(1): 0111020.  
戴媛, 易本顺, 肖进胜, 等. 基于改进旋转区域生成网络的遥感图像目标检测[J]. *光学学报*, 2020, 40(1): 0111020.
- [2] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137-1149.
- [3] Han Y S, Ma S P, Li S, et al. Airport area detection based on optimized regional convolutional neural network [J]. *Laser & Optoelectronics Progress*, 2020, 57(10): 101021.  
韩永赛, 马时平, 李帅, 等. 基于优化区域卷积神经网络的机场区域检测[J]. *激光与光电子学进展*, 2020, 57(10): 101021.
- [4] Wang J Q, Li J S, Zhou X W, et al. Improved SSD algorithm and its performance analysis of small target detection in remote sensing images [J]. *Acta Optica Sinica*, 2019, 39(6): 0628005.  
王俊强, 李建胜, 周学文, 等. 改进的 SSD 算法及其对遥感影像小目标检测性能的分析[J]. *光学学报*, 2019, 39(6): 0628005.
- [5] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M]//Leibe B, Matas J, Sebe N, et al. *Computer vision-ECCV 2016. Lecture notes in computer science*. Cham: Springer, 2016, 9905: 21-37.
- [6] Du Z X, Yin J Y, Yang J. Remote sensing image detection based on dense connected networks [J]. *Laser & Optoelectronics Progress*, 2019, 56(22): 222803.  
杜泽星, 殷进勇, 杨建. 基于密集连接网络的遥感图像检测方法[J]. *激光与光电子学进展*, 2019, 56(22): 222803.
- [7] Ju M R, Luo H B, Wang Z B, et al. Improved YOLO V3 algorithm and its application in small target detection [J]. *Acta Optica Sinica*, 2019, 39(7): 0715004.  
鞠默然, 罗海波, 王仲博, 等. 改进的 YOLO V3 算法及其在小目标检测中的应用[J]. *光学学报*, 2019, 39(7): 0715004.
- [8] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08)[2021-02-21]. <https://arxiv.org/abs/1804.02767>.
- [9] Lu C W, Krishna R, Bernstein M, et al. Visual relationship detection with language priors[M]//Leibe B, Matas J, Sebe N, et al. *Computer vision-ECCV 2016. Lecture notes in computer science*. Cham: Springer, 2016, 9905: 852-869.
- [10] Zhang H W, Kyaw Z, Chang S F, et al. Visual translation embedding network for visual relation detection [C] //2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 3107-3115.
- [11] Zhang J, Kalantidis Y, Rohrbach M, et al. Large-scale visual relationship understanding [EB/OL]. (2019-08-16)[2021-02-21]. <https://arxiv.org/abs/1804.10660>.
- [12] Mikolov T, Sutskever I, Chen K, et al. Constructing distributed representations using additive clustering [EB/OL]. (2002-04-01)[2021-02-21]. [https://www.researchgate.net/publication/2538177\\_Constructing\\_Distributed\\_Representations\\_Using\\_Additive\\_Clustering](https://www.researchgate.net/publication/2538177_Constructing_Distributed_Representations_Using_Additive_Clustering).
- [13] Grover A, Leskovec J. node2vec: scalable feature learning for networks [C] //Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 13-17, 2016, San Francisco, California, USA. New York: ACM, 2016: 855-864.
- [14] Kiros R, Salakhutdinov R, Zemel R. Multimodal neural language models [EB/OL]. (2014-11-10)[2021-02-21]. <https://arxiv.org/abs/1411.2539>.
- [15] Dutta A, Zisserman A. The VIA annotation software for images, audio and video [C] //Proceedings of the 27th ACM International Conference on Multimedia, October 15, 2019, Nice, France. New York: ACM, 2019: 2276-2279.
- [16] Krishna R, Zhu Y K, Groth O, et al. Visual genome: connecting language and vision using crowdsourced dense image annotations [EB/OL]. (2016-02-23)[2021-02-21], <https://arxiv.org/abs/1602.07332>.
- [17] Tang K H, Niu Y L, Huang J Q, et al. Unbiased scene graph generation from biased training [C] //2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 3713-3722.