

基于注意力及生成对抗网络的遥感影像目标检测^①



李佳琪, 邓玉娇, 吴湘宁, 代刚, 陈苗, 王稳, 方恒, 涂雨, 张锋

(中国地质大学(武汉)计算机学院, 武汉 430078)
通信作者: 吴湘宁, E-mail: wxning@cug.edu.cn

摘要: 针对传统目标检测算法在环境多变、背景复杂、目标聚集、小目标过多的航空遥感影像目标检测上效果不理想的问题, 本文提出了一种基于注意力机制及生成对抗网络的遥感影像目标检测模型 Attention-GAN-Mask R-CNN. 该模型将注意力、生成对抗网络和 Mask R-CNN 结合起来, 用以解决遥感影像目标检测中存在的问题. 实验结果表明, 在复杂的遥感影像数据集中, 该方法提升了目标检测的效率和准确率.

关键词: 遥感影像; 目标检测; 注意力机制; 生成对抗网络; 深度学习

引用格式: 李佳琪, 邓玉娇, 吴湘宁, 代刚, 陈苗, 王稳, 方恒, 涂雨, 张锋. 基于注意力及生成对抗网络的遥感影像目标检测. 计算机系统应用, 2022, 31(6): 182-191. <http://www.c-s-a.org.cn/1003-3254/8490.html>

Object Detection in Remote Sensing Image Based on Attention Mechanism and GAN

LI Jia-Qi, DENG Yu-Jiao, WU Xiang-Ning, DAI Gang, CHEN Miao, WANG Wen, FANG Heng, TU Yu, ZHANG Feng
(School of Computer Science, China University of Geosciences (Wuhan), Wuhan 430078, China)

Abstract: The traditional object detection algorithm is subject to changeable environment, complex background, target aggregation and too many small targets, showing disadvantages in the detection of aerial remote sensing images. This study presents the Attention-GAN-Mask R-CNN model for the object detection in remote sensing image based on attention mechanism and generative adversarial network (GAN). This model combines attention, generative adversarial network and Mask R-CNN to solve the above problems. The experimental results show that this method can improve the efficiency and accuracy of target detection in the complex remote sensing images.

Key words: remote sensing image; object detection; attention mechanism; generative adversarial network (GAN); deep learning

1 引言

遥感影像中飞机、船只以及近海海滩和岛屿的检测不仅在海洋军事信息研究领域是一个重要的应用方向, 在民用工业领域也同样发挥着重要的作用. 对于这类目标, 主要的任务需求是在海面、湖面或地面背景下对图像中的船只、海滩、岛屿以及机场中的飞机进行位置标注和掩膜分割. 而飞机和船只图像在尺度、分辨率上与自然图像存在差异, 海滩和岛屿也会因为崎岖的轮廓使得识别效果不明显.

针对高分辨率遥感图像中目标太小、识别掩码不

清晰、输入影像模糊等问题, 本文提出了一种基于注意力机制和生成对抗网络的模型 Attention-GAN-Mask R-CNN 简称 AG-Mask R-CNN, 将注意力、生成对抗网络^[1]和 Mask R-CNN^[2]结合起来. 本文的主要研究工作为: (1) 针对数据中目标尺度不同, 开展多尺度特征融合研究. 针对多尺度的输入, 在主干网络中添加特征金字塔网络 (FPN)^[3]. (2) 针对遥感图像中对象占比较小, 容易被漏检的问题, 开展注意力机制算法研究. 引入注意力机制, 嵌入在卷积层网络中, 产生注意力感知的特征, 使得网络关注更重要的区域, 以提升网络检测

① 基金项目: 国家自然科学基金 (U1711266); 中国地质大学地质探测与评估教育部 (B 类) 重点实验室主任基金 (CUG2019ZR11)

收稿时间: 2021-08-17; 修改时间: 2021-09-13; 采用时间: 2021-09-22; csa 在线出版时间: 2022-03-11

的精度。(3) 在 Mask (掩码) 分支中引入生成对抗网络。由于 Mask 分支生成器与生成对抗网络中的生成器定义相同, 因此利用单独的生成对抗网络对 Mask 分支的 Mask 生成网络进行预训练, 从而提升原始 Mask 分支中生成器的精度。(4) 模型的训练与评估。针对模型的训练结果进行验证分析, 并与初始 Mask R-CNN 网络等其他主流网络进行对比来说明模型的可行性和高效性。

2 AG-Mask R-CNN 模型

2.1 模型结构

AG-Mask R-CNN 模型的网络结构如图 1 所示。

- (1) 对遥感数据进行预处理作为输入。
- (2) 骨干网络采用 ResNet101^[4], 并且在 ResNet 卷

积层之间嵌入注意力模块以及特征金字塔网络, 以此来增强对小目标的语义提取和检测精度。注意力模块如图 2 所示, 采用先加入通道域注意力再加入空间域注意力的方式, 在卷积层的每一层网络之间都嵌入一个封装好的注意力模块。

(3) 沿用 Mask R-CNN 的模型架构, 针对 Mask 分支进行改进, 对 Mask 生成器进行生成对抗的生成器预训练。初始操作是对输入图像中物体的信息进行表达和融合, 这个过程中 CNN^[5] 模块充当特征提取网络, 对输入图像进行特征提取, 生成特征图。FPN 主要是减少尺度变换误差。之后候选区域生成网络 (RPN) 根据预设框进行候选区域的推荐, 输出经过筛选的正样本回归框和特征图。

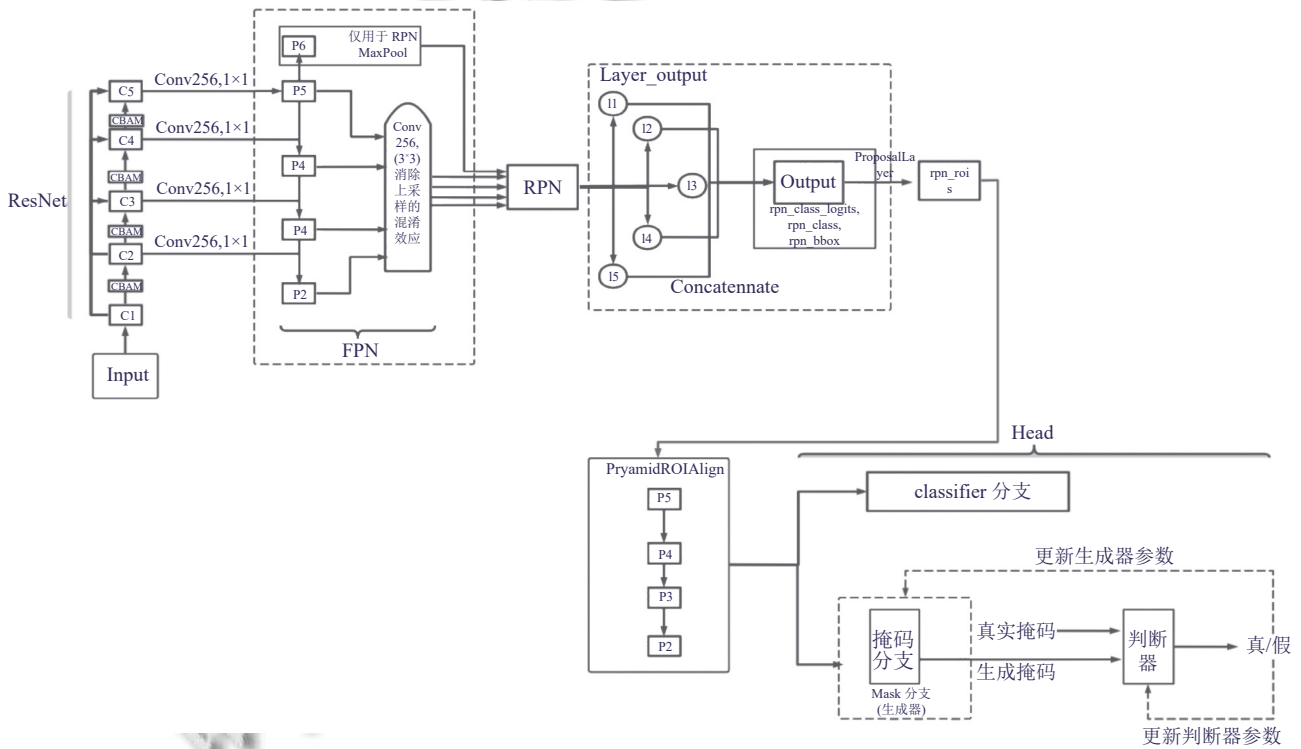


图 1 AG-Mask R-CNN 网络模型架构图

2.2 注意力机制的实现

AG-Mask R-CNN 的注意力模块如图 2, 采用先加入通道域注意力再加入空间域注意力的方式, 并且将 CBAM^[6] 嵌入到 ResBlock。这样连接的方式不仅保证了通道注意力和空间注意力的有效应用, 同时保留残差连接, 有效加速模型的收敛, 保证训练的效率。注意力模块详细架构图^[7] 如图 3, 在注意力模块中, 感知分

支通过接收特征提取器的注意图和特征图, 输出每个类的最终概率。VGGNet、ResNet 等图像分类模型的传统顶层感知分支结构相同, 如图 3(c)。使用式 (1) 和式 (2) 两种注意力机制中的一种将注意图应用到特征图中。 $g_c(X_i)$ 是特征提取器上的特征映射, $M(X_i)$ 是一个注意映射, $g'_c(X_i)$ 是注意感知分支的输出。多任务学习的 ABN 机理如图 3(a), $\{C_1, \dots, C\}$ 是通道的索引。

$$g'_c(X_i) = M(X_i) * g_c(X_i) \quad (1)$$

$$g'_c(X_i) = (1 + M(X_i)) * g_c(X_i) \quad (2)$$

式(1)是一个简单的点积之间的关注和特征图谱在特定频道的乘积. 相比之下, 式(2)可以在峰值突出

特征映射的注意力图, 而防止注意力图上特征权值较低的区域退化为零. 可以通过在两个分支机构使用损失进行端到端的训练, 训练损失函数 $L(X_i) = L_{att}(X_i) + L_{per}(X_i)$, 即注意分支和感知分支损失的简单总和.

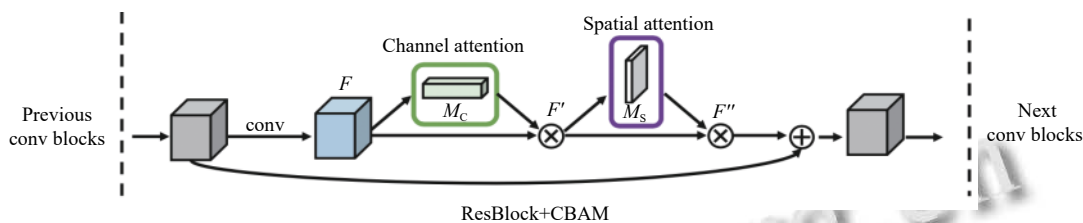
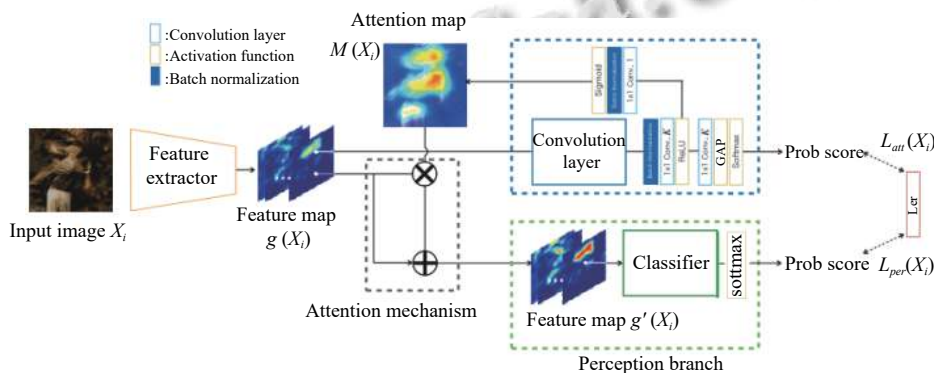


图2 注意力模块示意图^[6]



(a) 注意力模块

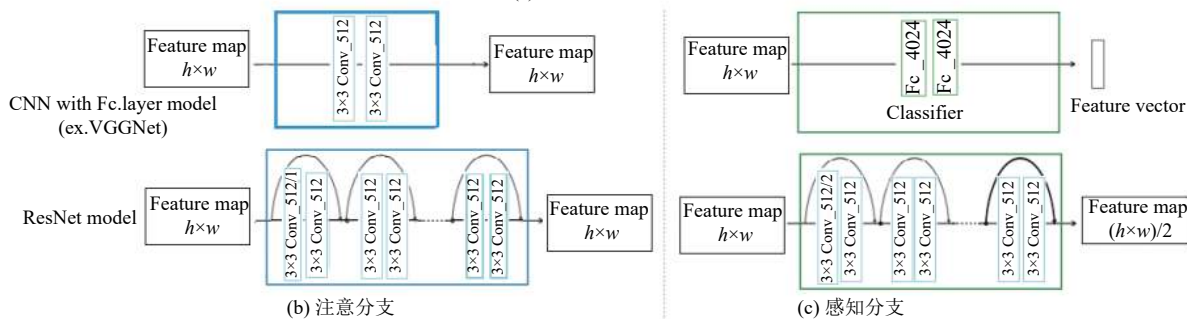


图3 注意力模块详细架构图^[7]

通道域注意力公式如式(3), *AvgPool* 和 *MaxPool* 分别表示平均池化和最大池化, *F* 表示输入的特征图, *W₀* 和 *W₁* 均表示通道注意力模型的权重矩阵, *M_c(F)* 表示生成最终的通道注意力特征. 在通道注意力模块中, 每个通道的特征代表一个特殊的检测器, 它会更关注那些对图像判别有意义的区域, 并帮助网络学习这些特征.

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \\ = \sigma(W_1(W_0(F_{avg}^c) + W_1(F_{max}^c))) \quad (3)$$

空间域注意力公式如式(4), *f^{7x7}* 表示在拼接后的特

征图上做 7x7 的卷积, *M_s(F)* 表示最终生成的空间注意力特征. 图4为通道域注意力与空间域注意力模块结构图.

$$M_s(F) = \sigma(f^{7 \times 7}([AvgPool(F); MaxPool(F)])) \\ = \sigma(f^{7 \times 7}([F_{avg}^s; F_{max}^s])) \quad (4)$$

2.3 特征金字塔网络的实现

目标检测在处理多尺度变化问题时存在一些不足之处, 如小物体本身包含的像素信息较少, 在下采样的过程中容易丢失, 为了处理这种物体大小差异十分明显的检测问题, 经典方法是利用图像金字塔的方式进行多尺

度变化增强,但这样会带来极大的计算量. Karras 等人^[8]在现有网络基础上进行层之间连接的修改,提出 FPN,在不增加原始模型计算量的情况下提高小目标检测的性能.

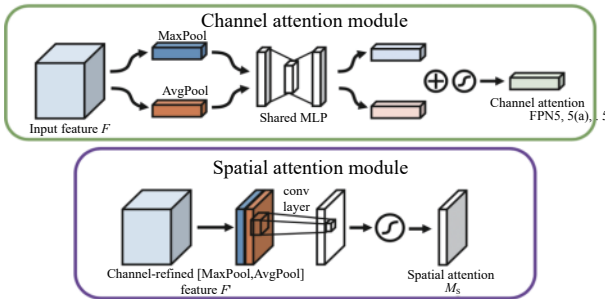


图4 通道域注意力与空间域注意力

FPN 的各种连接方式如图 5, 图 5(a) 的连接方式会把每一个尺度的特征图都抽取出来构建金字塔网络,检测速度慢. 图 5(b) 为了保证检测的速度,牺牲一定的精度,只对单一尺度的图像进行金字塔构建. 图 5(c) 通过卷积层特征提取后快速建立特征金字塔结构. 图 5(d) FPN 网络结合这些构建方式的优点,兼顾尺度变换与特征学习的速度. 特征金字塔网络由自底向上和自顶向下两个操作组成. 自底向上的操作是前向传播卷积神经网络的主要部分,计算一个缩放步长为 2 的特征层,得到每个阶段的最后一层特征. 自顶向下加入了横向连接,对图像进行上采样的操作,得到语义特征更强的图像. FPN 在对图像特征构建时的两个处理过程如图 6.

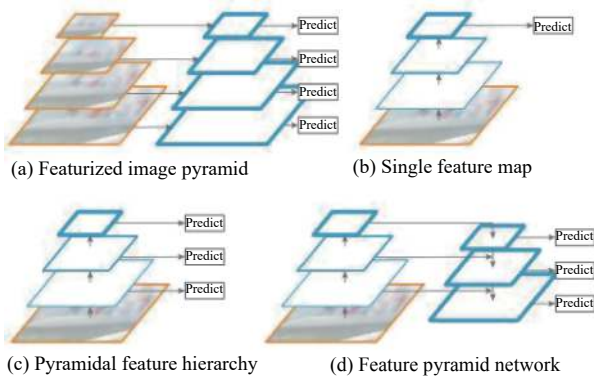


图5 FPN 各种连接方法对比图^[3]

2.4 生成对抗网络

生成对抗网络是一种使用对抗竞争思想的神经网络,生成器^[9]就是在多轮次(epoch)的训练中学习真实的文本向量或图像的分布规律.网络将这种规律和一个随机噪声进行特征映射,根据学习到的特征生成一

个新的图像或文本向量.而判别器主要是判断输入的向量是属于真实的数据集还是生成器生成的噪声向量.训练的过程中,两个模型都会进行学习,生成器会生成更真实的噪声混合在真实的数据中送入判别器.而竞争的过程中,判别器鉴别真假的能力也逐步提升. AG-Mask R-CNN 应用 Mask 分支网络为生成器,加入一个判别器,来判断 Mask 是否为真,进行对抗训练,以提高 Mask 分支所生成掩码的精准度.

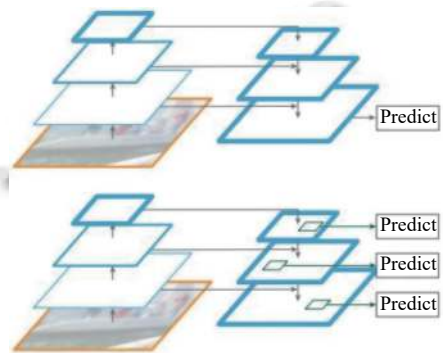


图6 FPN 网络处理过程图

如图 7 所示,首先需要有一个凭空捏造图片的生成器,还有一个判断青蛙图片是否为真的判别器.生成器接受一个 1 000 维的随机生成的数组,输出一个 64×64×3 通道的图片数据.判别器输入 64×64×3 的图片,输出数 1 或者 0,代表图片是否是青蛙.随后真图与假图进行拼接,打上标签,真图标签是 1,假图标签是 0,送入训练的网络.所有生成对抗网络的核心内容就是将生成器与判别器进行连接,而本文用于生成器强化训练,因此要设置判别器参数不进行修改,只在训练时对生成器网络的参数进行调整.

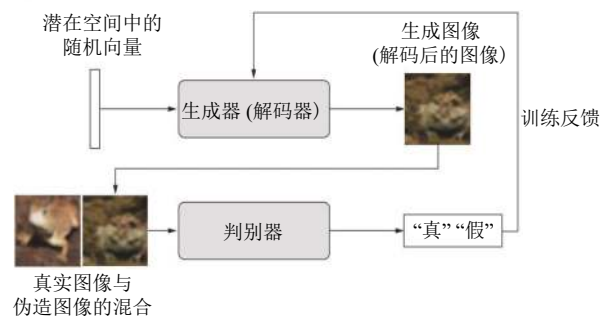


图7 生成对抗网络生成假图

3 数据集

所用数据集为从 DIOR, DOTA, NWPU 数据集中挑选出的飞机、船只、岛屿、海岸线图片作为自己的

原始数据集,其中船舶 7 000 张,飞机 7 000 张,岛屿 4 000 张、海滩 3 700 张.均以 256×256 的像素大小作为标准的网络输入.

图 8 为用于网络训练的数据集样本的部分图片,不

包含标注文件,直接选取图片用 labelme 进行标注.而且原图存在像素不一、目标区域所占比例不同等问题,需要进行初始裁剪去除多余的背景区域,并将含有目标区域切割成符合网络输入的 256×256 大小的图片.

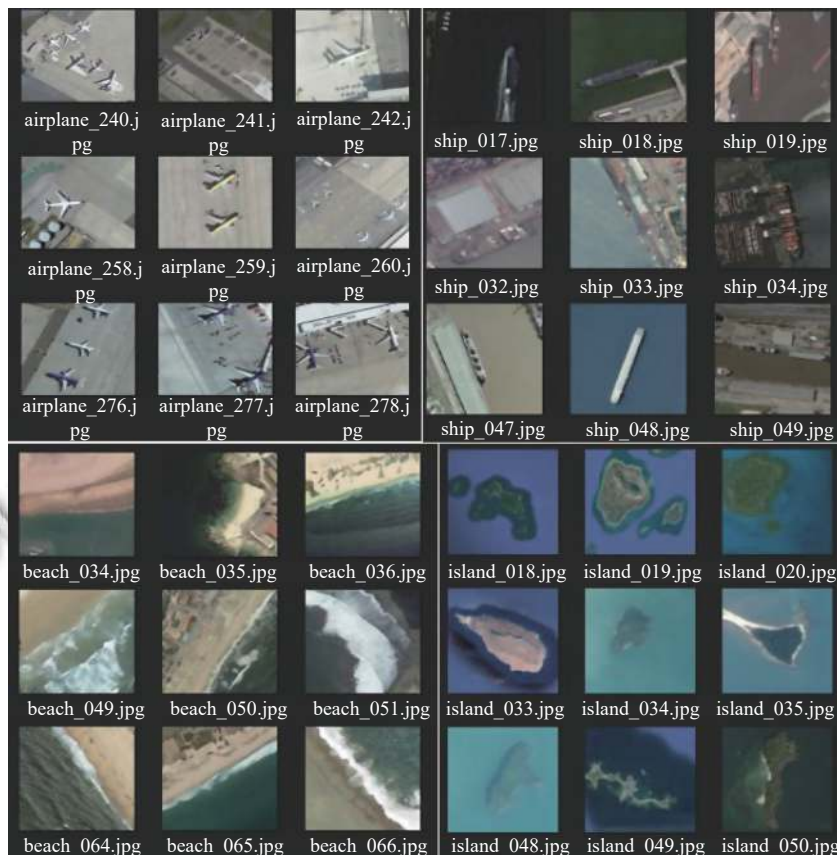


图 8 用于网络训练的数据集样本

遥感图像背景环境复杂,加上图像本身存在的一些无意义噪声,斑点白化等瑕疵,直接进行预测效果势必很差.因此需要对已经标注的数据集进行平滑、去噪、滤波等降噪处理.除此以外,遥感图像中的目标与一般目标检测中的汽车、船的方向和尺度有很大差别,因此需要对遥感图像做拉伸,旋转,高斯变换,亮度提升,去云等数据增强的操作,一方面扩大数据集,另一方面丰富多角度目标信息.由于输入数据中有相当一部分数据存在尺度分配不平衡的情况,需要对图像上的大小目标进行分析,对小目标进行过采样和复制等操作,提高其检测效率.

考虑到数据增强的工作量和效率问题,选择了 imageaug 数据增强库,为了更符合实际的训练场景,只对数据集中的部分数据进行随机的角度变换和仿射变换,并随机组合高斯模糊、均值模糊、中值模糊.其中随机的比例维持在 45%,可以确保数据集中超过 90%

的数据被再处理.图 9 为数据集中数据增强的结果.

4 模型训练和结果分析

图 10 为 AG-Mask R-CNN 训练流程图.模型训练流程如下.

(1) 输入训练样本数据集,对数据集进行预处理与数据增强操作.

(2) 将增强后的数据集送入骨干网络进行训练,调整每一层神经元节点对应的权重参数,以及获得最终输出的特征图.

(3) 对步骤(2)中获得的特征图做区域推荐操作,将结果传入区域推荐网络中进行正负样本的分类以及边界框的回归.过滤掉一些不符合条件的候选框.

(4) 对步骤(3)中得到的候选区域执行双线性插值,即 RoI Align 操作.

(5) 对步骤(4)中得到的候选区域做目标分类、边界框回归以及遥感目标的掩码生成,其中 Mask 分支添加生成对抗网络结构,在 Mask 生成器中添加一层判别器以提升 Mask 生成的精确度。

模型参数设置如表 1 所示: STEPS_PER_EPOCH 表示 AG-Mask R-CNN 每一轮迭代的步数. 更大的 VALIDATION_STEPS 可以提高验证统计数据的准确性. BACKBONE 是骨干网体系结构,支持的值为 ResNet50、ResNet101. RPN_ANCHOR_SCALES 表示正方形锚边的长度. RPN_ANCHOR_RATIOS 表示的是每个单元的锚固比,值为 1 表示方形锚定,值为 0.5 表示宽锚定. RPN_NMS_THRESHOLD 是过滤 RPN 建议的非最大抑制阈值. RPN_TRAIN_ANCHORS_PER_IMAGE 表示每个图像有多少锚用于 RPN 训练.

DETECTION_MIN_CONFIDENCE 是检测置信度. LEARNING_RATE 表示学习率.

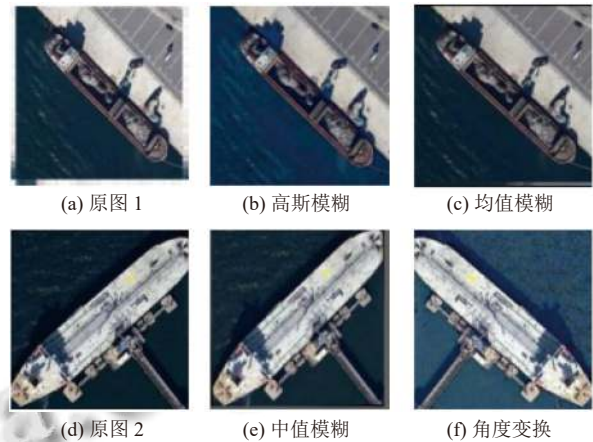


图 9 训练数据集增强后的图片

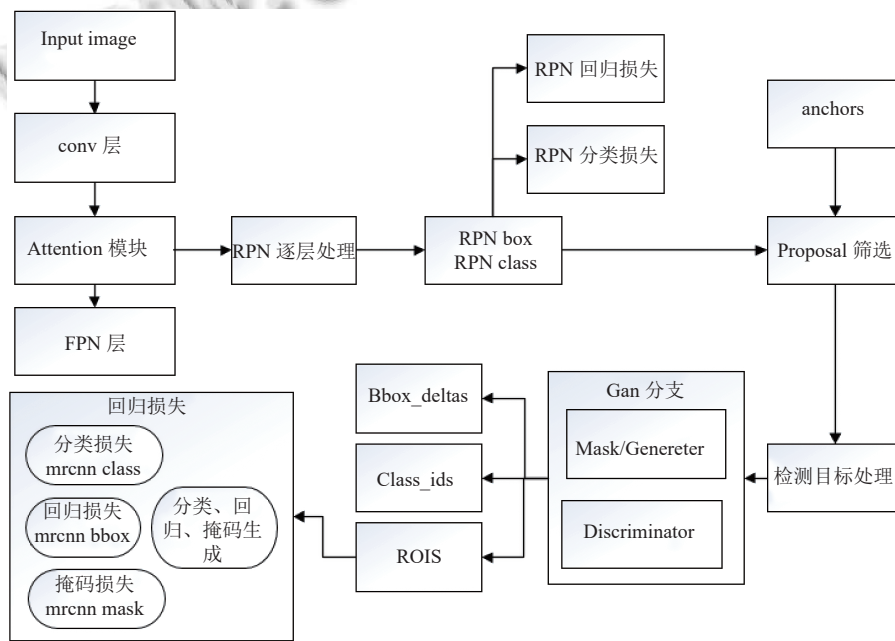


图 10 AG-Mask R-CNN 模型训练流程图

表 1 AG-Mask R-CNN 模型参数表

参数名称	参数值
STEPS_PER_EPOCH	1 000
VALIDATION_STEPS	50
BACKBONE	ResNet101
RPN_ANCHOR_SCALES	(32, 64, 128, 256, 512)
RPN_ANCHOR_RATIOS	[0.5, 1, 2]
RPN_TRAIN_ANCHORS_PER_IMAGE	256
RPN_NMS_THRESHOLD	0.7
DETECTION_MIN_CONFIDENCE	0.7
LEARNING_RATE	0.001

AG-Mask R-CNN 在训练过程中会存在一定的波动性,其结果不会因为使用相同的训练参数而相同,因此为了使实验结果更具说服力,以及考虑到整个实验过程对 GPU 算力资源的要求,实验方法中会用到迁移学习来训练深度神经网络.实现步骤主要是在训练遥感检测数据集之前先加载预训练的 COCO 权重,以此来加速模型的收敛,同时也可以一定程度上避免过拟合。

AG-Mask R-CNN 的头部网络训练需要 30 个轮

次, 训练初始阶段采用随机初始化, 其优势在于可以打破一些预训练的特征, 因此在头部网络训练时, 需要冻结网络其他的层. 头部网络训练完成后, 需要微调除了头部网络之外其他网络层的参数. 遥感检测数据集在输入网络中进行训练时, 刚开始需要将学习率设置得尽量小, 以免冻结层的卷积核发生较大的变化. 通过实验发现, 将学习率设置成 0.001 时损失函数曲线收敛更快.

全连接层训练完成后, 接下来再做微调训练, 规则是每过 10 个轮次, 学习率会减少一半. 通过观察, 模型会在第 35 个轮次过拟合, 而第 32 个轮次损失还在下

降, 因此训练轮次被设定为 30 轮. AG-Mask R-CNN 在训练过程中会冻结一部分层及其对应的参数, 防止网络出现严重的过拟合线性.

图 11 为 AG-Mask R-CNN 损失函数曲线图, 图中共有 5 种损失, 分别是 `rpn_bbox_loss`、`rpn_class_loss`、`mrcnn_class_loss`、`mrcnn_bbox_loss`、`mrcnn_mask_loss`. `loss` 是总的损失函数曲线, 横坐标为训练的轮次. 从图中可以看出, 损失函数最终是收敛状态, 说明模型训练是成功的. 但训练过程中会出现一些波动的情况, 原因可能在于处理时批次大小设置过小, 但是不影响模型整体收敛.

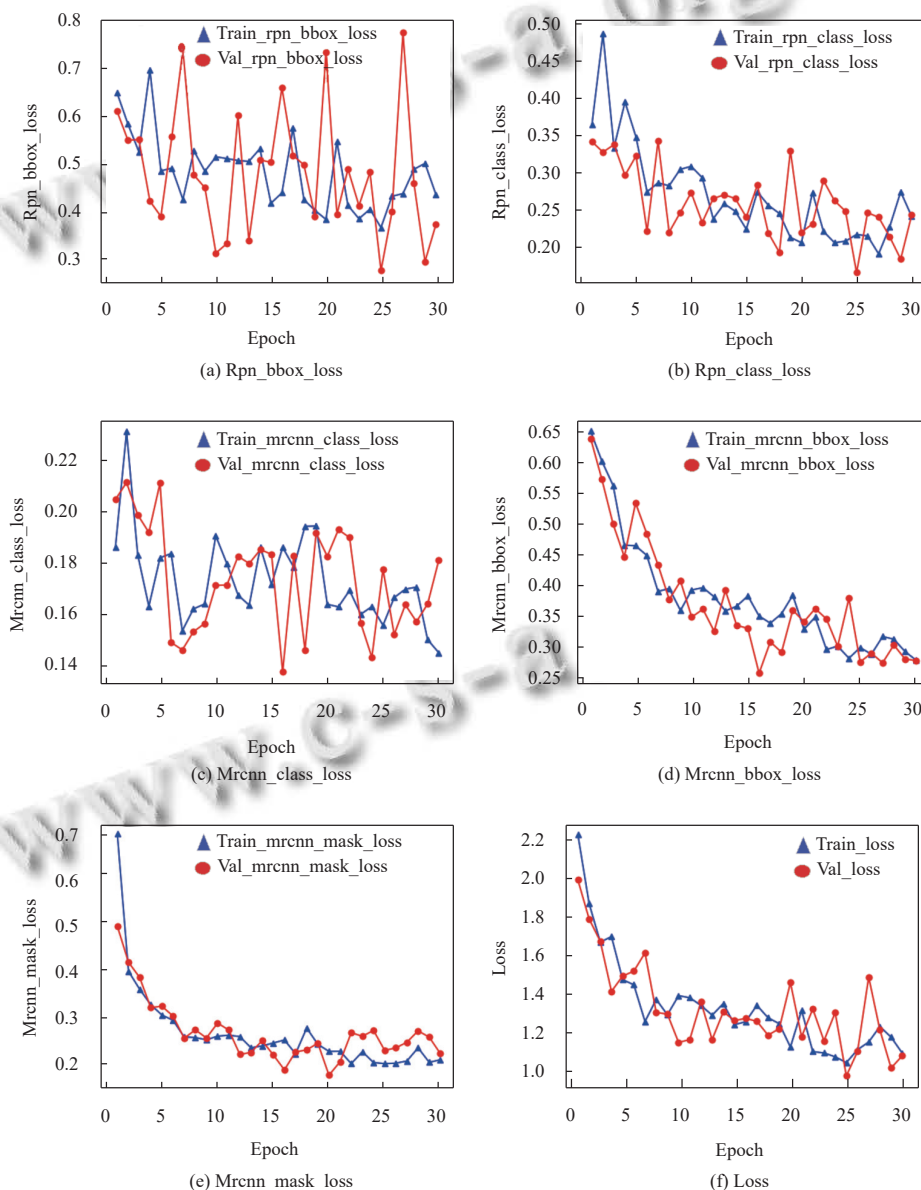


图 11 AG-Mask R-CNN 网络损失图

图 12 为添加了注意力机制后注意力在图片上的可视化热力图, 颜色越深表示该区域在网络中占的权重更大, 也就意味着对该区域的学习超过其他区域. 从图中可看出, 注意力模块对于数据集中目标的关注多于背景, 一定程度上对局部区域的关注可以增加网络的学习效率, 进而对最后的检测结果有一定的提升.

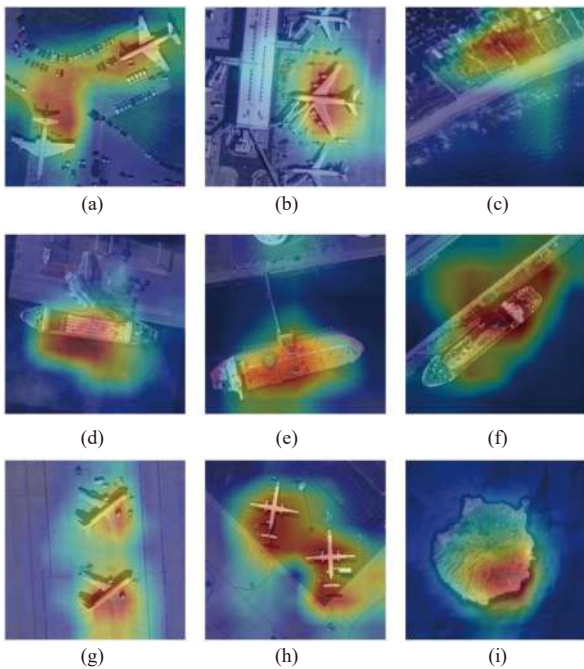


图 12 AG-Mask R-CNN 注意力通道可视化图层

图 13 为 AG-Mask R-CNN 网络中 64 个通道的可视化图, 第 1 行从左到右分别是原图, 第一通道 Tensor (张量) 转换的可视化图以及第二通道 Tensor 转换的可视化图, 在卷积神经网络中, 均以张量为计算输入的标准; 第 2 行为最大池化层的可视化图, 池层保留了样本的主要特征, 可以减少参数的个数和过拟合; 第 3 行是卷积层的可视化, 卷积过程中所有的参数是共享的. 从图中可以清楚地看出, 卷积核能够在不同位置保留相同的特征.

图 14 为部分生成对抗网络的训练集图片, 图 15 为对 Mask 分支网络作为生成器单独训练时, 生成器生成的飞机掩码, 原始的训练集就是标注过的 Mask 数据. 这里以飞机掩码生成为例, 共训练 10 000 个轮次, 直到判别器将生成器生成的掩码判别为真. 由图可得, 生成对抗网络结构中空间收敛具有强烈的不稳定性, 初始的生成几乎看不出掩码的轮廓形状, 经过大规模的训练后, 基本形状可以生成, 但是空间的不平衡很容

易导致生成对抗网络崩溃. 在实验过程中分别对生成器和判别器的学习参数进行了 True 和 False 的设置, 有两次训练过程中, 网络崩溃, 特征未收敛, 这有可能是导致后续将 Mask 分支加入主干网络后, 对掩码的生成准确度并没有质的提升的原因.

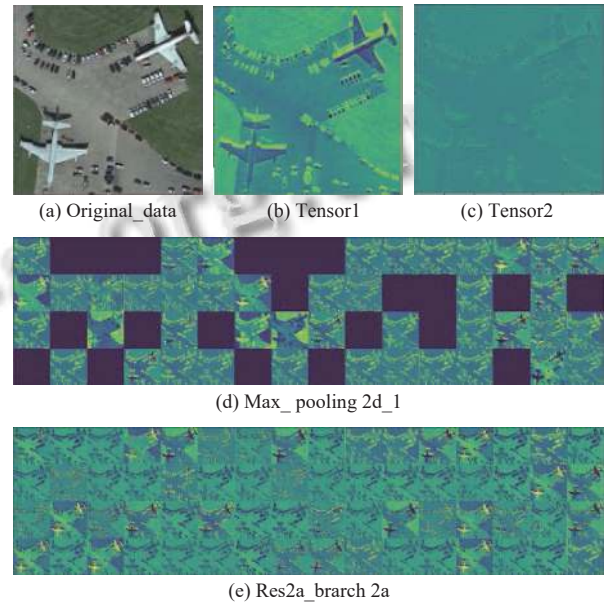


图 13 AG-Mask R-CNN 卷积通道可视化图



图 14 AG-Mask R-CNN 生成对抗网络训练 Mask 数据

图 16 为测试集部分图片的检测结果, 从检测结果来看, 船只的检测效果最好, 锚框大小很符合船只目标, 掩码形状也基本与目标重合. 对于岛屿和海滩这类目标而言, 由于本身在样本中占图片比例很大, 边缘特征也更易于学习, 检测效果也相对较好, 以上均为置信度为 0.9 以上的检测框. 而飞机图片目标的检测效果并不理想, 检测框有的时候会覆盖整张图片, 并且由于加入了负样本, 除了有一定的边角目标漏检, 还会在不含有目标的负样本中进行错误识别. 图中的结果显示掩码

在飞机目标图片的生成上也不能完全覆盖目标形状. 推测是目标轮廓过于复杂, 且飞机图片中目标所占比例很小, 这使得尺度小并且轮廓复杂目标的特征不易被网络学习, 从而也会影响到后续 Mask 的生成. 为了解决飞机目标检测效果差的问题, 重新标注了一部分飞机数据集 (300 张), 这次的方式是对飞机进行矩形框标注, 而不是采用一开始的精确轮廓标注. 在将数据集进行 10 倍增强后放入网络, 测试效果提升不明显, 虽然在检测框上精度有所提高, 但 Mask 的生成依然存在和之前一样的问题.

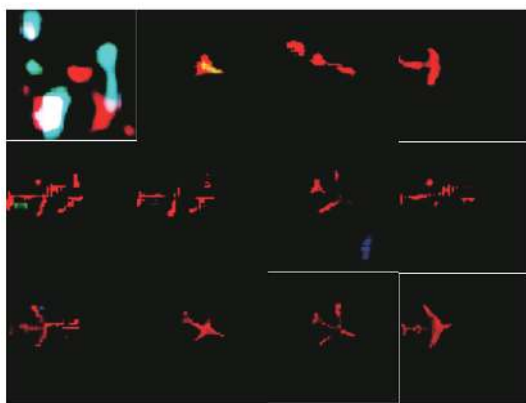


图 15 AG-Mask R-CNN 生成对抗网络结构生成掩码可视化

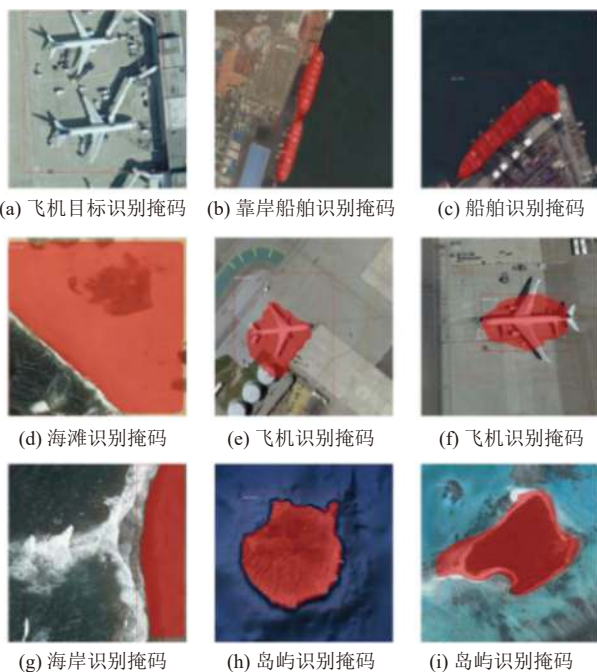


图 16 AG-Mask R-CNN 部分检测效果图

图 17 为初始 Mask R-CNN 模型在用同样的数据集训练后在测试集上的检测效果图, 可以直观地看出

初始模型在特定任务上的检测效果并不理想, 对于一些比较明显的目标不仅没有办法正确检测, 检测出来的目标框也并不接近目标大小, 掩码方面也不能完整地勾画. 而 AG-Mask R-CNN 无论在目标检测的位置、目标框的大小, 还是在掩码的精度上都明显强于 Mask R-CNN.

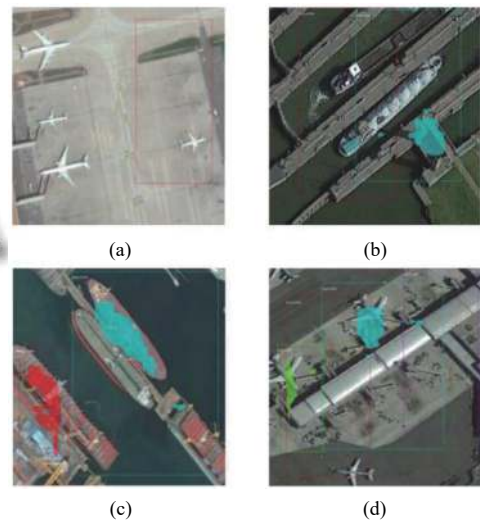


图 17 Mask R-CNN 部分检测效果图

表 2 是模型在准确率、召回率等方面与 Mask R-CNN、Faster R-CNN 等基准模型的对比. Precision 代表准确率, Recall 代表召回率, $F2$ 指数为 Recall 占比 2, mAP 为所有类别 AP 的平均值 (以 VOC mAP 计算方式为准). 通过表 2 可知, 基准模型 YOLOv4 与 SSD 在该任务上表现略差, 原因在于两者都是一阶段检测器, 更加注重检测的性能, 所以在查全率和查准率上表现略微差于二阶段检测器. 传统的二阶段检测器如 Faster R-CNN 在该任务上表现一般, 略差于 Mask R-CNN. Mask R-CNN 在使用 ResNet101 作为骨干网络的前提下, 无论在 Precision 还是 Recall 上都优于使用 ResNet50, 模型检测的 Precision 与 Recall 分别为 90.34 与 87.88, 原因在于 ResNet101 的网络结构更加复杂, 在该类任务上对于特征的提取能力强于 ResNet50.

为了分析注意力模块与生成对抗网络对模型产生的影响, 将这两种机制分别加入到基准模型并进行了对比实验. 表 3 的实验结果表明, 在骨干网络相同的情况下, 模型在单独加入注意力模块与单独加入生成对抗网络时, 检测精度均存在一定的提升. 前者的加入给模型带来了 1% 左右的 mAP 提升, 后者的加入给模型

带来了小幅度的 AP 提升. 最后一组是 AG-Mask R-CNN 实验结果, 由于使用 ResNet101 作为骨干网络, 并同时嵌入注意力机制与生成对抗网络, 因此在同类任务上表现最优, 达到了 92.22% 的检测精度.

表 2 基准模型对比 (%)

模型	Precision	Recall	F2	mAP
Faster R-CNN	88.27	86.32	86.63	89.40
Mask R-CNN (ResNet50+FPN)	89.12	87.44	87.77	90.87
Mask R-CNN (ResNet101+FPN)	90.34	87.88	88.36	91.80
YOLOv4	88.76	86.52	86.85	89.11
SSD	87.88	84.44	85.02	88.29

表 3 AG-Mask R-CNN 与基准模型加不同模块对比 (%)

模块	Precision	Recall	F2	mAP
Mask R-CNN (ResNet50+FPN)	89.12	87.44	87.77	90.87
Mask R-CNN (ResNet101+FPN)	90.34	87.88	88.36	91.80
Mask R-CNN (ResNet50+FPN) +Attention	89.23	87.55	87.88	90.92
Mask R-CNN (ResNet101+FPN) +Attention	91.29	88.65	89.09	92.44
Mask R-CNN (ResNet50+FPN) +GAN	89.22	87.66	87.96	90.88
AG-Mask R-CNN	92.22	89.33	89.89	93.48

通过对比实验可见, 本模型在加入生成对抗网络结构后, 效果并没有非常显著的提升, 可能原因是, 生成对抗网络结构想生成一个判别器判断为真的目标掩码图片时, 可能需要上千轮次训练, 而本文因设备限制只训练了 100 轮, 也没有在空间上进行收敛, 还有就是, 在 RoI 阶段的锚框定位也会影响后续 mask 的生成, 因此 mask 分支的改进对模型精度的提高效果并不明显, 这也是模型今后可以改进的方向.

5 结束语

基于注意力及生成对抗网络的遥感影像目标检测模型 AG-Mask R-CNN, 在目标检测框架 Mask R-CNN 的基础上引入注意力机制和生成对抗网络. 提出用于检测小目标的注意力机制方案, 将注意力模块添加到骨干网络 ResNet101 中, 给整个算法模型带来一部分性能上的提升. 采用特征金字塔来对各个尺寸大小的目标进行特征提取, 在模型中加入了生成对抗网络结

构预训练 Mask 分支. 实验结果表明, 生成对抗网络的加入对模型检测精度的提升带来正面的提升. 在经过数据增强的遥感图像数据集上的对比实验表明, AG-Mask R-CNN 模型对航空遥感目标影像的检测识别具有更好的表现.

参考文献

- Goodfellow I, Pouget-Abadie J, Mirza M, *et al.* Generative adversarial nets. Proceedings of the Advances in Neural Information Processing Systems. Montrel: NIPS, 2014. 2672–2680.
- He KM, Gkioxari G, Dollár P, *et al.* Mask R-CNN. Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 2980–2988.
- Lin TY, Dollár P, Girshick R, *et al.* Feature pyramid networks for object detection. Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 936–944.
- He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 770–778.
- LeCun Y, Boser B, Denker JS, *et al.* Backpropagation applied to handwritten zip code recognition. Neural Computation, 1989, 1(4): 541–551.
- Woo S, Park J, Lee J Y, *et al.* CBAM: Convolutional block attention module. Proceedings of the European Conference on Computer Vision (ECCV). Munich: Springer, 2018. 3–19.
- Fukui H, Hirakawa T, Yamashita T, *et al.* Attention branch network: Learning of attention mechanism for visual explanation. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 10697–10706.
- Karras T, Aila T, Laine S, *et al.* Progressive growing of GANs for improved quality, stability, and variation. arXiv: 1710.10196v3, 2017.
- Ledig C, Theis L, Huszár F, *et al.* Photo-realistic single image super-resolution using a generative adversarial network. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 105–114.