计算机工程与科学
Computer Engineering & Science
ISSN 1007-130X,CN 43-1258/TP

《计算机工程与科学》网络首发论文

题目：一种基于注意力机制的小目标检测深度学习模型
作者：吴湘宁，贺鹏，邓中港，李佳琪，王稳，陈苗
收稿日期：2020-03-13
网络首发日期：2020-10-14
引用格式：吴湘宁，贺鹏，邓中港，李佳琪，王稳，陈苗．一种基于注意力机制的小目标检测深度学习模型[J/OL]．计算机工程与科学.
https://kns.cnki.net/kcms/detail/43.1258.tp.20201012.1130.002.html

CN 43-1258/TP
ISSN 1007-130X

计算机工程与科学
Computer Engineering & Science

一种基于注意力机制的小目标检测深度学习模型*

吴湘宁，贺鹏，邓中港，李佳琪，王 稳，陈 苗

（中国地质大学（武汉）计算机学院， 湖北 武汉 430078）

摘 要：小目标检测用来识别图像中小像素尺寸目标。传统目标识别算法泛化性差，而通用的深度卷积神经网络算法容易丢失小目标的特征，对小目标识别的效果不甚理想。针对以上问题，提出了一种基于注意力机制的小目标检测深度学习模型，模型在 ResNet101 主干网络和候选区域生成网络中使用了通道注意力和空间注意力，通道注意力模块实现了通道维度上的特征加权标定，空间注意力模块实现了空间维度上的特征聚焦，从而提升了小目标的捕获效果。此外，模型使用数据增强技术和多尺度特征融合技术，保证了小目标特征提取的有效性。通过在遥感影像数据集中识别船只的实验表明，注意力模块可带来小目标检测的性能提升。

关键词：小目标检测；深度学习；遥感图像；注意力机制

中图分类号：TP753 文献标志码：A

A deep learning model of small object detection based on attention mechanism

WU Xiang-ning，HE Peng，DENG Zhong-gang，LI Jia-qi，WANG Wen，CHEN Miao

（School of Computer Science, China University of Geosciences(Wuhan), Wuhan 430078, China）

Abstract: Small target detection is used to identify small pixel size targets in image. The generalization of traditional target recognition algorithms are poor, and general depth convolution neural network algorithms are easy to lose the characteristics of small target, so these algorithms are not ideal for small target recognition. To solve the above problems, a small target detection deep learning model based on attention mechanism is proposed. The model uses channel attention and spatial attention in ResNet101 backbone network and region proposal network. The channel attention module implements feature weighting calibration in channel dimension, and the spatial attention module realizes feature focusing in spatial dimension, thus improving capture effect of small targets. In addition, the model uses data enhancement technology and multi-scale feature fusion technology to ensure the effectiveness of small target feature extraction. The experiment of ship recognition in remote sensing image data set shows that the attention module can improve the performance of small target detection.

Key words: small object detection; deep learning; remote sensing image; attention mechanism

*收稿日期：2020－03－13；修回日期：2020－04－20
基金项目：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项目(CUG2019ZR11)；国家自然科学基金项目(U1711266)
通讯地址：430078 湖北省武汉市东湖新技术开发区锦程街 68 号中国地质大学计算机学院
Address: School of Computer Science, China University of Geosciences, No.68, Jincheng Street, East Lake High-tech Development Zone, Wuhan 430078, Hubei , P.R.China

1 引言

在目标检测领域，小目标通常指尺寸小于 32×32 像素的目标[1]。小目标检测的难点在于图像中目标的分辨率低、图像模糊、信息量少，能提取到的特征信息少。传统目标检测常用的特征提取算法有 SIFT[2]、HOG[3]、图像金字塔[4]等，这类算法难以从海量的数据集中学习出一个有效的分类器来充分挖掘数据之间的关联，不适合解决图像的小目标检测问题。

近年来，深度卷积神经网络(ConvNet)[5]用于目标检测并取得了很大的进展，ConvNet 实现了特征、候选区域、边界框的提取以及对象类别的判别，然而，ConvNet 检测器不太适合检测小目标，这是因为在卷积神经网络中，特征图的分辨率比原始输入图像的分辨率要低得多，分辨率甚至会降低16倍之多，这使得分类和边界框回归非常困难[6]。因此，不论是一段式的 YOLO[7]和 SSD[8]，还是两段式的 Faster-RCNN[9]，对小目标检测效果都不理想。此后深度学习领域出现了专门针对小目标检测的改进方法，如多尺度融合、尺度不变性等方法。金字塔网络 FPN[10]将低层位置信息与高层语义信息融合，解决了以往的深度学习目标检测算法只采用顶层特征映射进行分类预测，而忽略了低层特征的位置信息的问题。尺度不变性目标检测模型 SNIP[11] 实现了多尺寸图像输入，提高了预选框精度，对小目标检测效果有一定提升。

此外，在计算视觉领域出现了注意力思想，注意力机制可以有效捕捉图片中有用的区域，通过给关键特征标识权重，使模型能学到需要关注的区域。视觉注意力分为软注意力(soft attention)和强注意力(hard attention)。软注意力关注的是区域和通道，是一个完全可微的确定性机制，可以通过网络模型算出梯度并反向转播学习到注意力的权重。而强注意力则更关注图像中的点，具有不确定性。软注意力可分为空间域、通道域、混合域的注意力。空间域注意力的代表有空间转换网络模型 STN [12]。通道域注意力的代表有挤压及激励网络模型 SENet[13]，通过学习每个通道的权重，从而产生通道域的注意力。混合域注意力的代表有剩余注意网络模型模型[14]，通过注意力掩码，不只是对空间域或通道域，同时也对每个特征元素找出对应的注意力权重。这些算法被证明可有效捕捉关键区域。

本文提出了一种基于注意力的掩码区域卷积神经网络 AM-R-CNN 模型(Attention based Mask R-CNN)，模型使用数据增强技术和多尺度特征融合技术，保证小目标的特征得到加强且不易流失，并在卷积神经网络中引入注意力感知机制，使得不同模块的特征会随着网络的加深产生适应性改变。实验结果表明，模型提高了遥感影像中船只目标识别的准确性。

2 模型实现及在遥感影像小目标识别中的应用

2.1 模型结构

AM-R-CNN 模型采用了 Mask R-CNN[15]作为基本框架，这是因为 Mask R-CNN 是一个灵活的目标检测框架，它在 Faster RCNN 的基础上扩展，在预测分支和边界框分支上添加了一个用于预测目标掩码的分支，适用于目标检测、语义分割、人体姿态识别等领域。

AM-R-CNN 模型整体结构如图 1 所示，模型分为以下几个主要部分：

（1）数据预处理模块：对原始图像进行预处理和数据增强；

（2）ResNet101 骨干(backbone)网络：将 ResNet 101 和特征金字塔结合后构成 ResNet101 FPN。负责从输入的数据中提取特征，输出为特征图集合；

（3）候选区域生成网络(RPN)：从特征图中提取候选区域；

（4）混合注意力模块：为 ResNet101 及 RPN 提供通道注意力及空间注意力机制。

（5）头部网络：对 Faster RCNN 进行改进和扩展。含 ROI (Region Of Interest，感兴趣区域) Align 及三个分支网络。ROI Align 使用双线性插值来更精确地找到每个块对应的特征，它从特征图中提取出固定长度的特征向量和并列的 Mask。每个特征向量都会被输送到全连接层(FC 层)序列中，FC 层又分支成两个同级输出分支，一个是分类分支，用来产生 softmax 概率分布来对目标分类，输出每一个 ROI 中的目标关于 K 个分类(包括背景类)的概率分布。另一个分支是边框回归分支，可输出 K 个类的精确边界框位置(四个实数编码值)。而掩码分支采用掩码编码来识别目标的空间布局。每一个 ROI 定义了多任务损失：$L = L_{class}+L_{boxes}+L_{mask}$，三个损失参数分别对应三个分支。

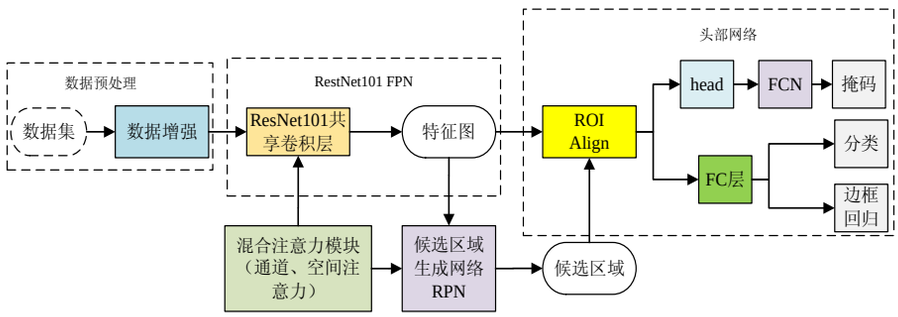

Figure 1 Architecture of the AM-R-CNN Model

图 1 AM-R-CNN 模型的体系结构

2.2 特征金字塔网络 ResNet101 FPN 的实现

模型将特征金字塔与 ResNet101 有效地结合，构成了一个特征金字塔网络 ResNet101 FPN。用于提取 8×8 至 64×64 像素这种小目标的语义特征。

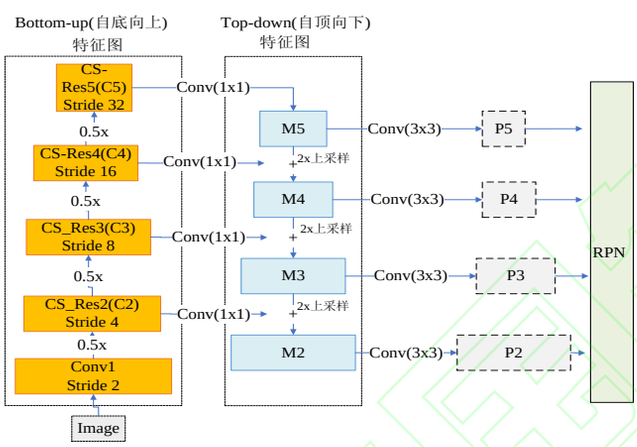

Figure 2 Structrue of ResNet101 FPN

图 2 ResNet101 FPN 的结构

FPN 实现过程如图 2 所示，左边“自底向上路径”是卷积网络的前馈计算，计算由不同比例的特征映射组成的特征图，其缩放步长为 2，ResNet 块 Conv1，CS-Res2，CS-Res3，CS-Res4，CS-Res5 的 stride(步距)分别设置为 2，4，8，16，32，输出的特征图为{C2，C3，C4，C5}。而右边的“自顶向下路径”，通过对更抽象但语义更强的高层特征图进行上采样，来幻化高分辨率的特征。对于上层邻近的特征空间做 2 倍最近邻上采样，然后将“自底向上路径”中的特征图经过 1*1 的卷积变换后的结果与上层的上采样结果相加合并，再对合并结果{ M2，M3，M4，M5}使用 3*3 的卷积，以减少上采样带来的混叠效应。最终得到多尺度特征图集合{P2，P3，P4，P5}，这些特征图与{C2，C3，C4，C5}中对应的特征图有相同的尺寸。

2.3 注意力机制的实现

在模型中，在 ResNet101 中和 RPN 两处使用了混合注意力模式，包含通道注意力和空间注意力。其中通道注意力思想借鉴了 Inception[16] 和 MobileNet[17] 网络。

通道域注意力的原理是首先将一个通道上整个空间特征编码聚合为一个全局特征，再通过另一种运算提取通道之间的关系。假设 $V=[v_1,v_2,\ldots,v_c]$表示学习到的卷积核的集合，其中的v_c表示第c个卷积核的参数，输出为 $U=[u_1,u_2,\ldots,u_c]$，计算公式如下：

$$u_c = v_c * x = \sum_{x=1}^{C} v_c^s * x^s \qquad (1)$$

公式中的*表示卷积运算，由于输出是通过通道求和产生，通道间的关系隐式地嵌入在v_c中。卷积运算建模的通道关系本质上是局部相关的，因此，需要通过建模通道相互依赖关系来实现全局通道信息的相关。通道注意力模块的计算公式如下：

$$\begin{aligned} M_c(F) &= \sigma\Big(MLP\big(AvgPool(F)\big) \\ &\quad + MLP\big(MaxPool(F)\big)\Big) \\ &= \sigma\Big(W_1\Big(W_0\big(F_{avg}^c\big)\Big) \\ &\quad + W_1\big(W_0(F_{max}^c)\big)\Big) \end{aligned} \qquad (2)$$

公式中，σ表示 sigmoid 函数，*MLP* 表示多层感知器，用于共享参数。F_{avg}^c和F_{max}^c分别代表全局平均池化和全局最大值池化输出的特征。输入的特征图 *F* 分别通过基于宽和高的全局最大值池化和全局平均值池化，然后分别通过多层感知器，将感知器输出的特征进行基于逐元素的相加操作，最后经过 sigmoid 激活函数，生成最终的通道注意力特征 *Mc*。

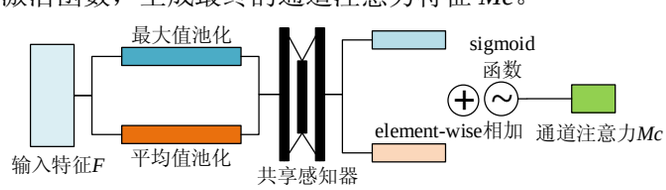

Figure 3 Channel Attention Module

图 3 通道注意力模块

通道注意力模块的实现过程如图 3 所示。算法借鉴了 SENet 模型[13]的方法，选用全局平均池化来实现简单的聚合运算，单一的全局平均池化聚合的特征信

息对特征图中每个像素都有反馈，引入全局最大值进行梯度反向传播，计算特征图中响应最大的区域，并将两个池化得到的一维矢量相加，丰富全局平均池化提取的信息。

除了通道间存在相关注意力信息，在空间层面上也存在着注意力信息。空间域的注意力用于准确定位空间中的目标特征，在目标检测数据集中，小目标像素的占比很小，添加空间域注意力能准确定位小目标，提高检测的准确率。空间注意力模块的计算公式如下：

$$M_s(F)=\sigma\left(f^{7*7}([AvgPool(F);MaxPool(F)])\right)=\sigma\left(f^{7*7}\left([F^s_{avg};F^s_{max}]\right)\right) \quad (3)$$

公式中，σ 表示 sigmoid 函数：$F^s_{avg}\epsilon\mathbb{R}^{1*H*W}$ $F^s_{max}\epsilon\mathbb{R}^{1*H*W}$，卷积层使用 7*7 的卷积核。

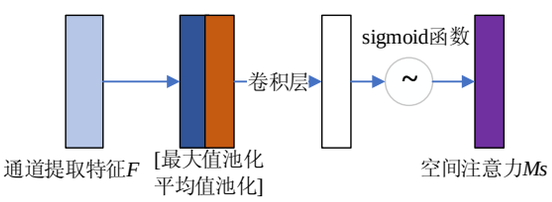

Figure 4 Spatial Attention Module

图 4 空间注意力模块

空间注意力模块的实现如图 4 所示。首先，使用全局平均池化和全局最大值池化对输入的特征图 F 进行压缩操作，对输入特征分别在通道维度上做 mean 和 max 操作。然后将得到的两个特征图按通道维度拼接，再经过一个卷积操作，降维为 1 个通道，保证得到的特征图在空间维度上与输入的特征图一致，最后，经过 sigmoid 函数生成空间注意力特征 M_S。

在 ResNet101 网络的 ResNet 块中添加混合注意力的方法如图 5 所示。将上一层产生的特征图做一次卷积计算产生输入特征图 F，F 经过通道注意力模块后得到的通道注意力特征 Mc，将 F 与 M_C 进行逐元素乘法操作得到新的特征图F'，然后将F'输入到空间注意力模块得到空间注意力特征 M_S，再将 M_S 与F'进行逐元素乘法操作，得到的混合注意力特征图F''与 F 进行相加操作，保留 ResNet 的残差模块，最后生成的特征图F'''作为下一个模块的输入。

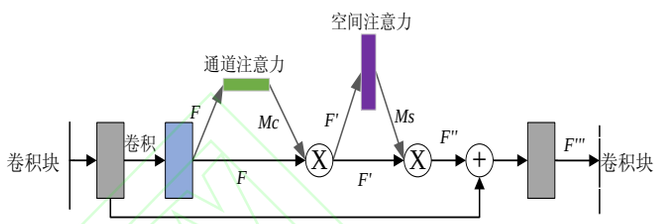

Figure5 ResNet Module with Attention Added

图 5 添加了注意力的 ResNet 模块

在 RPN 中添加混合注意力机制的方法如图 6 所示，F 表示从 FPN 输出的多尺度特征图，F 通过一层卷积计算和 Relu 激活函数后得到特征图 $F1$，将 $F1$ 输入通道注意力模块后得到通道注意力特征 M_C，$F1$ 与 M_C 进行逐元素乘法操作后得到F'，将F'输入到空间注意力模块后得到空间注意力特征 M_S，再将 M_S 与F'进行逐元素乘法操作，得到的混合注意力特征图F''，然后再通过一批归一化和池化层后得到最终的混合注意力特征图F'''。

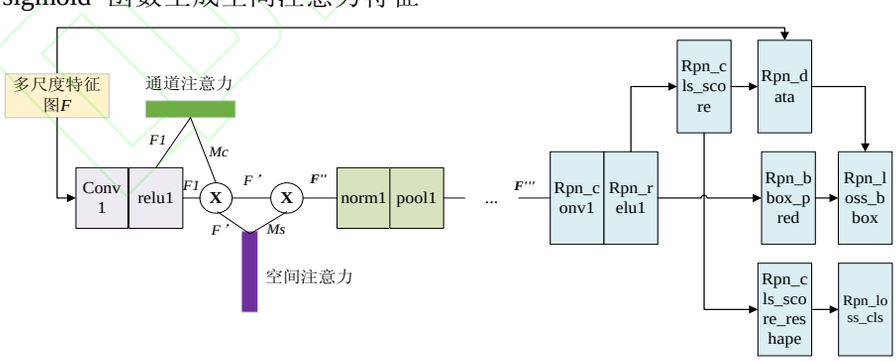

Figure 6 Attention Module in RPN

图 6 RPN 的注意力结构示意图

RPN 实现了一个滑动窗口式的目标检测器，RPN 将输入由单一尺寸的特征图替换成了 FPN 生成的具有多尺度的金字塔特征图，在特征金字塔的每一层添加与 RPN 相同的 head 结构(3*3 卷积和两个并列的 1*1 卷积)。由于 head 会在特征金字塔中所有层级的所有位置上滑动，因此不需要在每个层级上使用多尺度的锚框。RPN 依靠一个在共享特征图上滑动的窗口，为每个位置生成 9 种预先设置好长宽比与面积的锚框。这 9 种初始锚框包含三种面积(128×128，256×256，512×512)，每种面积又包含三种长宽比(1:1，

1:2，2:1)。

2.4 数据增强

本文使用遥感影像海洋船只检测数据集，遥感影像已做了去噪、平滑、滤波等预处理，数据集包含训练集和测试集，训练集有 1925526 张图片，测试集有 15606 张图片。csv 文件为训练图像提供行程长度编码，用来定位船只，并生成图像的 mask 及边界框。

数据集根据不同的 IoU(Intersection over Union,交并比)阈值下的 F2 分值进行评估，IoU 用于预测区域目标像素和真实的目标像素的重叠度，计算公式为：

$$IoU(A,B)=\frac{A\cap B}{A\cup B} \quad (4)$$

IoU 阈值范围 0.5~0.95，步长为 0.05，在预测阈值为 0.5 时，预测目标与真实目标的 IoU 大于 0.5 时表示“命中”。根据预测目标与所有真实目标比较得到的真阳性(*TP*)、假阴性(*FN*)和假阳性(*FP*)三个指标来计算 F2 分值。F2 分值的计算公式如下：

$$F_\beta(t)=\frac{(1+\beta^2)TP(t)}{(1+\beta^2)TP(t)+\beta^2FN(t)+FP(t)} \quad if\ \beta=2 \quad (5)$$

公式中$F_\beta(t)$得分表示精确率和召回率的调和值，β表示召回率的重要程度相对于精准率的倍数，当β为 1 时，精确率和召回率都很重要，权重相同，得数被称为 F1 分值。在本模型中，认为召回率更重要些，因此将β设为 2，因而得数被称为 F2 分值。式中的 t 表示数据样本，TP 指单个预测目标与真值目标匹配时的 IoU 高于阈值的样本个数，FP 表示有预测目标却没有相关真实目标的样本个数，FN 表示存在真实目标却没有相关联预测目标的样本个数。

Table 1 The encoded information of the training set image

表 1 训练集图像的编码信息

	ImageId	EncodedPixels	Ship_count
0	00003e153.jpg	NaN	0
1	0001124c7.jpg	NaN	0
2	000155de5.jpg	264661 17 265429 33 266197 33 266965 33…	1
3	000194a2d.jpg	360468 1 361252 4 362019 5 362785 8 …	5
4	000194a2d.jpg	51834 9 33602 9 53370 9 54138 9 54906 9…	5
…	…	…	…
231719	fffedbb6b.jpg	NaN	0
231720	ffff2aa57.jpg	NaN	0

训练集图像的编码信息及船只统计结果如表 1 所示。表中编码表示一些矩形框，用来框定图像中的船只，若编码为 NaN 表示图中没有船只。编码的字符串格式为：起点，长度，起点，长度，…，其中每对(起点，长度) 表示从位置开始绘制一定长度的像素线。起始位置不是二维(x,y)坐标，而是一维数组的索引，从而将二维图像压缩为一维像素序列。读取行程长度编码，解码后数组中的 1 表示 mask、0 表示背景，将得到的 mask 信息覆盖到对应的图像中，并使用透明的颜色实现可视化。处理后的结果如图 7 所示。

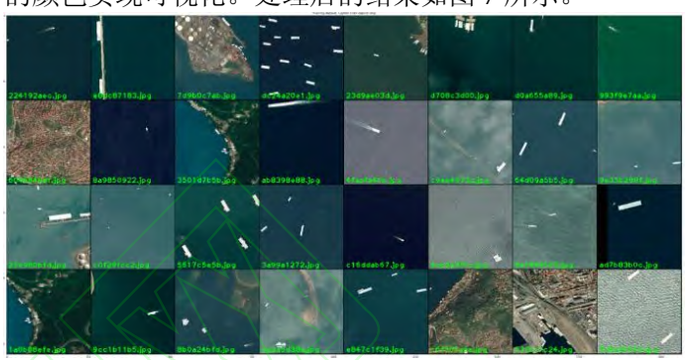

Figure 7 Parts of ships in training sets

图 7 训练集中的船只示意图

图 8 是训练集中图片的统计图，可以看出有船只的图片占比为 78%，所有图片中包含的船只数量为 81723。由于样本存在着类别不平衡的问题，因此，对没有船只的图片进行了下采样处理，以防止模型训练时引入过多噪声。在有船只的图片中，出现 1~2 只船的图片占绝大多数，小目标的数量太少意味着小目标信息量很少，可能导致训练出的模型更加关注其他信息特征，因此，针对包含某些数量船只的样本进行过采样，来保证样本种类相对平衡。

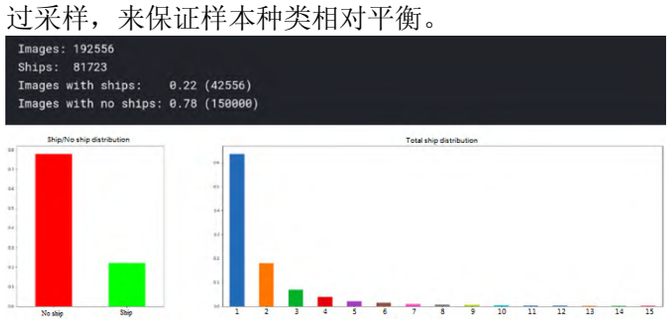

Figure8 Statistics of ships in training sets

图 8 训练集船只个数统计示意图

图 9 是从训练集的掩码中提取的船只边界框(bbox)。bbox 用于后续模型的训练，格式为(min_row, min_col, max_row, max_col)。

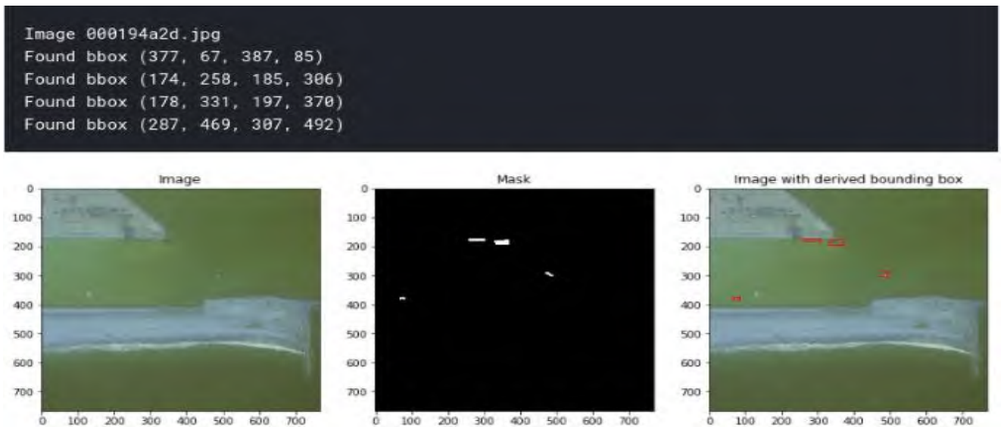

Figure9 Bounding boxes extracted from masks of ships in training set

图 9 从训练集中船只的掩码提取的边界框

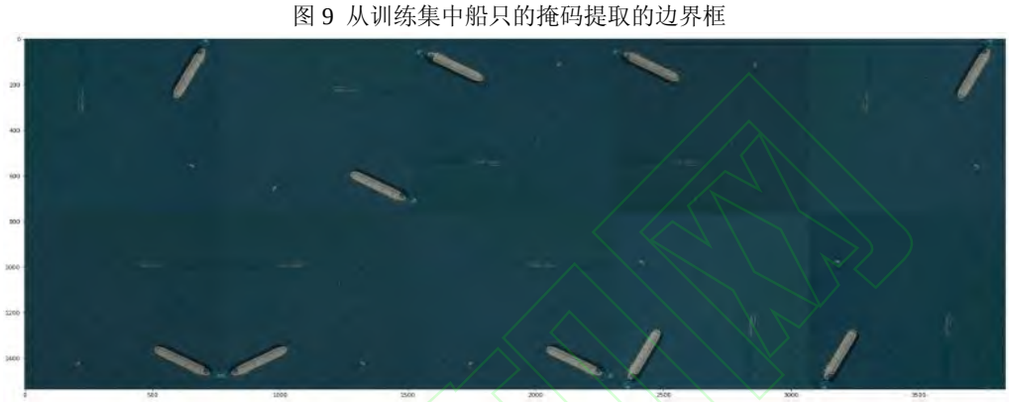

Figure 10 Positive samples after data augmentation

图 10 增强后的正样本数据

遥感图像中被检测目标存在角度的多样性，例如遥感图像中船、汽车的方向都可能与常规目标检测算法存在较大的差别[18]，这会增加检测的难度，因此需要对遥感图像做尺度变换、旋转等数据增强操作。同时，使用过采样来解决包含小目标的图像较少的问题，方法是将小目标从原始位置复制后粘贴到不同的位置，通过人为增加小目标的数量，使匹配的锚数增加，从而有效提高对小目标检测的性能。图 10 显示如何对有船只的图片进行随机 90° 倍数旋转，以增加样本的数量和方位多样性。

2.5 模型训练及评估

遥感图像数据经过数据增强处理后，划分为训练数据集及验证数据集，用训练数据集训练后得到 AM-R-CNN 模型，再将验证数据输入模型作验证和评估。

模型采用添加注意力模块后的 ResNet101 作为骨干网络，训练过程使用随机梯度下降进行优化。RPN 边框的 scales 分别为 4,8,16,32,64。Batch-size 设置为 8，Mask 的尺寸为 28×28，类别个数为 2，权重衰减系数为 0.0001，权重损失初始化为[‘rpn class loss’: 30.0, ‘rpn bbox loss’: 0.8, ‘mrcnn class loss’: 6.0, ‘mrcnn bbox loss’: 1.0, ‘mrcnn mask loss’: 1.2]，输入图片大小为 800×800，学习率为 0.001，图片的通道数为 3，每张图片的 ROI 数为 200，验证步数为 50，Top-Down 金字塔尺寸为 256。代码采用 Python 实现，实验平台为 Nvidia Telsa K40C，Xeon E5，32G RAM。

训练模型时，先加载 COCO 的预训练权重，使用预训练模型会使模型训练收敛更快，效果更好，训练更少的轮数，而且有可能获得低误差模型，避免陷入局部最优点。

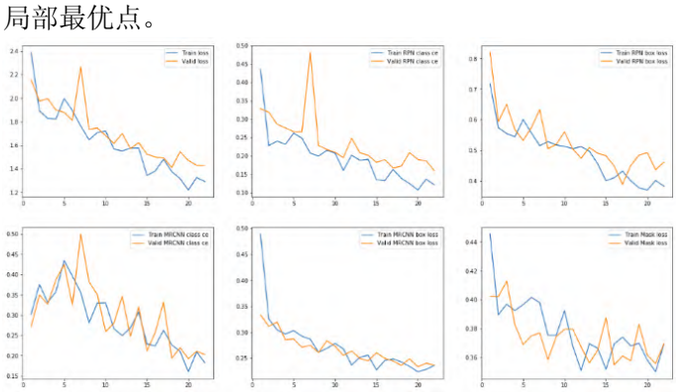

Figure 11 The Records of Training Loss

图 11 训练损失的记录图

随着训练的深入，模型的各项损失逐渐降低并最终收敛。模型的损失曲线如图 11 所示。蓝色的线表示训练集上的损失，黄色的线表示测试集上的损失。图中记录的损失函数包括整体损失、rpn 分类损失、rpn 边界框回归损失、mrcnn 分类损失、mrcnn 边界框回归损失、mrcnn 的掩码损失。由图中可以看出，所有损失呈整体下降的趋势，训练过程中，最好的轮数为 18，验证集的整体损失达到最小值，值为 1.4102766604423522。

图 12 显示了不同卷积层的图像特征表示，从左到右分别为原始图像、图像掩码和 ResNet 的 layer25、layer35、layer45 卷积层的输出特征图。可以看出，随着模型越来越深，图像的空间信息逐渐模糊，而背景的位置信息逐渐明显。

图 13 是 FPN 到 RPN 的训练过程，将 FPN 输出的特征图 P2,P3,P4,P5 输入到 RPN 中，对应锚框大小分别为 322,642,1282,2562，每一层锚框宽高比例为 {1:2,1:1,2:1}，共有 12 个锚框，锚框通过 ROI Align 层后分别得到 Mask 分支和特征向量，特征向量再分支为分类和边界框回归。训练时根据 IoU 的大小，为每个锚框贴上正负样本的浅标签。Heads 的参数在特征金字塔各层级中共享，可以使用一个通用的 head 分类在任意图片尺寸中进行预测[10]。

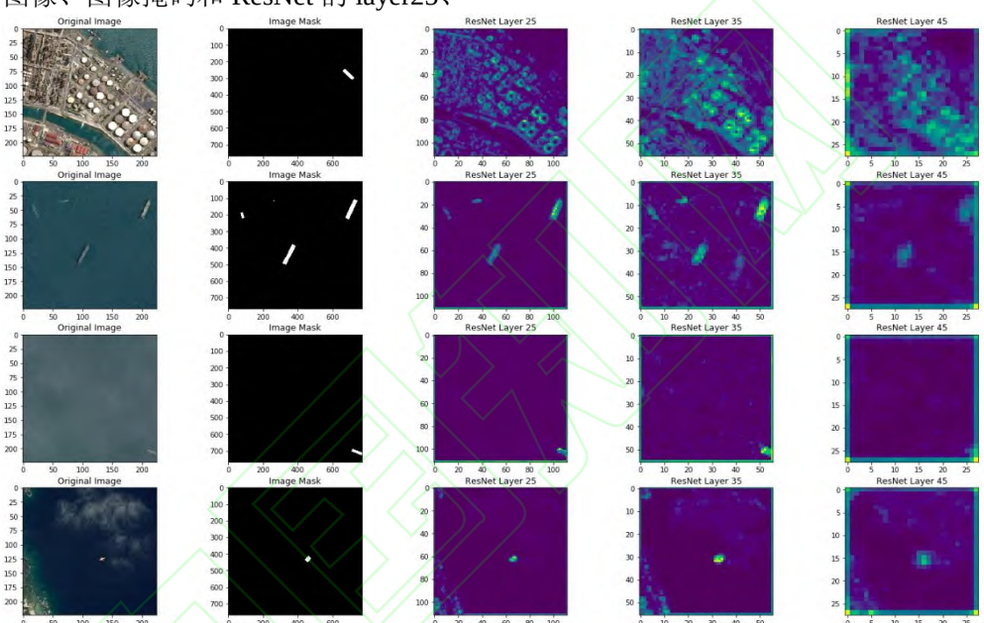

Figure 12 Convolutional Feature Maps of each layer in ResNet

图 12 ResNet 各层的卷积特征图

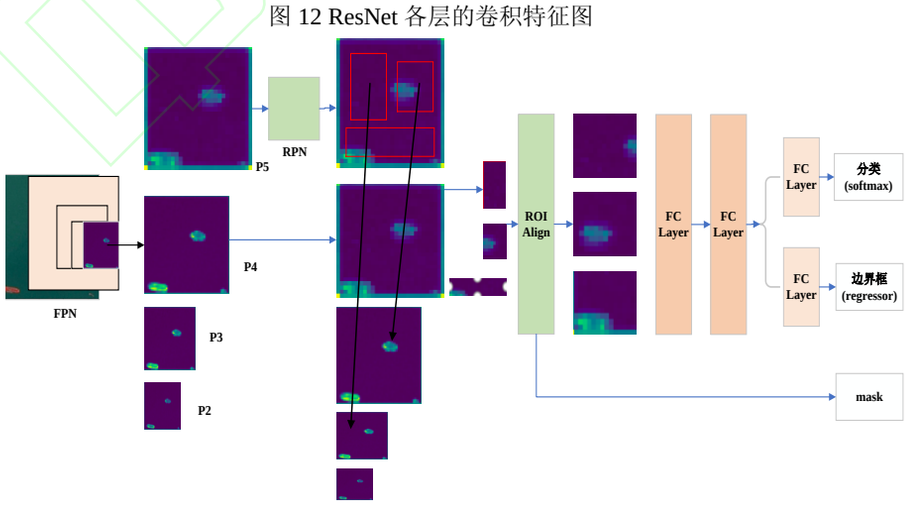

Figure13 Training process from FPN to RPN

图 13 FPN 到 RPN 的训练过程

在图 14 中，从左到右特征图的输出尺度分别为 391×391、548×548、768×768。小尺度的特征图像素

信息更少，位置信息更明显，而大尺度的特征图像素信息更多，图像相对模糊。在训练阶段，如果 ROI 与真实框的 IoU 大于 0.5 则被认为是正样本，否则为负样本，通过计算 L_{mask} 得到正样本上的掩码损失。在预测阶段，RPN 在每个样本中提取的 ROI 数量为 300，然后进行边框预测，并使用 Soft-NMS 算法[19]消除多余(交叉重复)的边框，找出目标最佳检测位置，在这里我们选择得分最高的 100 个边框，再对这些边框应用掩码分支，每个 ROI 预测出 K 个掩码，并将其缩放到 ROI 的尺寸大小，然后依据阈值 0.5 对掩码的像素值进行二值化操作，将 ROI 分为前景和背景。

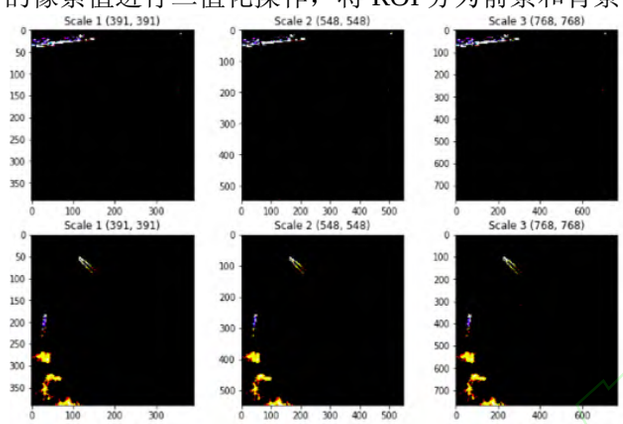

Figure 14 Multi-scale Feature Map

图 14 多尺度特征图

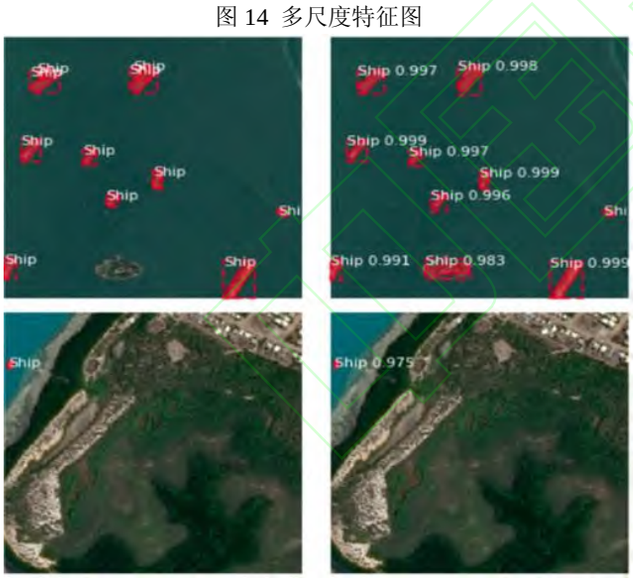

Figure 15 Comparison of real data and verification results

图 15 真实数据和验证结果的对比图

在验证阶段，加载训练第 18 轮所得到的权重来初始化注意力 ResNet101-FPN，使用模型对验证集进行验证，样本通过整个模型后，输出回归边界框、类别概率、掩码。图 15 是验证数据和预测结果的对比，左边的图表示原始图像及标注的船舶掩码前景及其边界框，而右边的图表示模型识别出的船舶掩码前景、边界框以及作为“ship”类的概率。可以看出在验证数据集上的检测精度还是比较高的。

最后，将测试数据集加载到经过验证的模型中，用于预测真实遥感图像中的船舶目标，设置最大值抑制的阈值为 0.45，最后得到的 F2 分值为 0.817，预测结果如图 16 所示，左图是识别出的船只边框，右图是船只的灰度图掩码，掩码的可能性为 0.985。

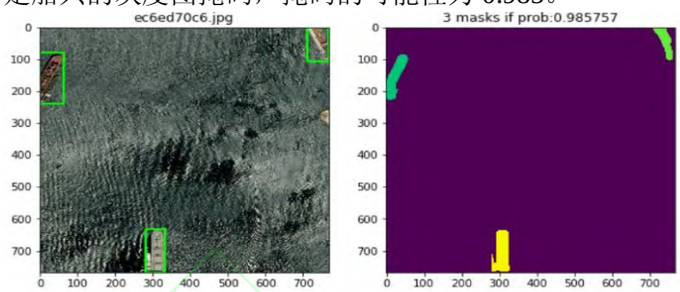

Figure 16 The result of predicting ship using test sets

图 16 测试集船只预测示意图

为了验证注意力机制在小目标检测的有效性，将 AM-R-CNN 模型与 Mask R-CNN、U-Net[20]、RetinaNet[21]、YOLO V3[22]这 4 个模型进行了比较。这些模型所采用的主干网络均为 ResNet101。各个模型采用相同的航拍遥感图像作为检测数据集。

不同模型的处理过程保持相同，在取不同阈值的情况下训练并测试后，记录其 F2 分值，最后将 F2 分值的平均值作为该模型的最后 F2 分值。

Table 2 Baseline model evaluation form

表 2 基准模型评估表

Model	Recall(%)	Precision(%)	F2(%)	Time Costs
U-Net	71.5	74.8	75.3	0.13
YOLO V3	69.4	73.1	71.6	0.15
RetinaNet	68.0	71.5	70.1	0.15
Mask R-CNN	74.7	75.3	76.1	0.18
AM-R-CNN	77.1	82.4	79.8	0.25

表 2 是几种模型的评估指标。指标包括召回率、精确率、F2 分值、推断时间。召回率是表示实际的正样本中，预测为正样本的比例，定义为$Recall = TP/(TP + FN)$，精确率表示被预测为正样本的实例中实际为正样本的比例，定义为$Precision = TP/(TP + FP)$。YOLO V3 和 RetinaNet 精确率和召回率不如其他两个模型，是因为此类模型属一段式检测模型，本身更注重算法的实时性，模型的准确率比不上 Mask R-CNN 等模型。U-Net 模型结构虽然简单，但船只检测数据集本身属于像素级特征，而模型的编码器-解码器结构更方便得到较高的准确率，实际上，U-Net 与 Mask R-CNN 有融合的可能。Mask

R-CNN 有效利用了数据集中的 Mask 信息，因此其各项指标较其它三个模型都要好，但时间的消耗略高。AM-R-CNN 模型在多个阶段添加了注意力机制，得到的区域提案锚框更准确，因此各项指标明显优于其他模型，但是推断时间也略长一点。实验结果表明，将注意力机制与 RPN 相结合，可有效提高小目标检测效果。

3 结束语

基于注意力机制的小目标检测深度学习模型 AM-R-CNN，在目标检测框架 Mask R-CNN 的基础上，引入了混合注意力机制。在 ResNet 网络，以及 RPN 中添加注意力模块，通过注意力掩码将图片中小目标的关键特征标识出来，从而帮助模型学习到需要关注的小目标区域。同时，模型设计了针对小目标的、可实现多尺度特征融合的 FPN，可以更好地提取8×8 至 64×64像素的小目标的特征。通过实验对比发现，AM-R-CNN 在经过数据增强的遥感图像数据集上，对船只的检测识别具有更好的表现。

参考文献：

[1] Bosquet B, Mucientes M, Brea V M. STDnet: A ConvNet for Small Target Detection[C]//BMVC.2018:253.

[2] D.G. Lowe Dept. of Comput. Sci., British Columbia Univ., Vancouver, BC, Canada Object recognition from local scale-invariant feat ures[C]. In IEEE,1999.

[3] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05).IEEE,2005,1:886-893.

[4] Adelson E H, Anderson C H, Bergen J R, et al. Pyramid methods in image processing[J]. RCA engineer,1984,29(6):33-41.

[5] Krizhevsky A,Sutskever I,Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems.2012:1097-1105.

[6] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE transactions on pattern analysis and machine intelligence,2015,37(9):1904-1916.

[7] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition.2016:779-788.

[8] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//European conference on computer vision. Springer, Cham,2016:21-37.

[9] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[C]//Advances in neural information processing systems.2015:91-99.

[10] Kirillov A, Girshick R, He K, et al. Panoptic feature pyramid networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.2019:6399-6408.

[11] Singh B, Davis L S. An analysis of scale invariance in object detection snip[C]//Proceedings of the IEEE conference on computer vision and pattern recognition.2018:3578-3587.

[12] Jaderberg M, Simonyan K, Zisserman A. Spatial transformer networks[C]//Advances in neural information processing systems.2015:2017-2025.

[13] J. Hu, L. Shen, G. Sun. Squeeze-and-excitation networks[C].Proceedings of the IEEE conference on computer vision and pattern recognition.2018:7132–7141.

[14] F. Wang, M. Jiang, C. Qian, et al. Residual attention network for image classification[C].Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.2017:3156–3164.

[15] He K, Gkioxari G, Dollár P, et al. Mask r-cnn[C]//Proceedings of the IEEE international conference on computer vision.2017:2961-2969.

[16] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions [C]//Proceedings of the IEEE conference on computer vision and pattern recognition.2015:1-9.

[17] Howard A G, Zhu M, Chen B, et al. Efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint ArXiv: 1704.0486,2017.

[18] Huang Jie, Jiang Zhiguo, Zhang Haopeng, et al. Remote sensing image ship target detection based on convolutional neural network[J]. Journal of Beijing university of aeronautics and astronautics,2017,43(9):1841-1848.

[19] Bodla N, Singh B, Chellappa R, et al. Soft-NMS--improving object detection with one line of code[C]//Proceedings of the IEEE inte rnational conference on computer vision. 2017: 5561-5569.

[20] 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. Springer, Cham,2015:234-241.

[21] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE international conference on computer vision.2017:2980-2988.

[22] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767,2018.

附中文参考文献：

[18]黄洁,姜志国,张浩鹏,等.基于卷积神经网络的遥感图像舰船目标检测[J].北京航空航天大学学报,2017, 43(9):1841-1848.

作者简介：

吴湘宁（1972-），男，湖南衡阳人，副教授，研究方向为机器学习及大数据技术。Email: wxning@cug.edu.cn

WU Xiang-ning, born in 1972, associate professor, his research interest includes machine learning and big data technology.

贺鹏（1995-），男，湖南双峰人，硕士生，研究方向为机器学习。Email: penghe6666@gmail.com

HE Peng, born in 1995, MS candidate, his research interest includes machine learning.

邓中港（1995-），男，湖北黄冈人，硕士生，研究方向为机器学习。Email:dzg337120@163.com

DENG Zhong-gang, born in 1995, MS candidate, his research interest includes machine learning.

李佳琪（1995-），女，河南安阳人，硕士生，研究方向为机器学习。Email:2957208668@qq.com

LI Jia-qi, born in 1995, MS candidate, her research interest includes machine learning.

王稳（1996-），男，湖北天门人，硕士生，研究方向为机器学习。Email: sillywa@foxmail.com

WANG Wen, born in 1996, MS candidate, his research interest includes machine learning.

陈苗（1997-），男，湖南常宁人，硕士生，研究方向为机器学习。Email: 2273259101@qq.com

CHEN Miao, born in 1997, MS candidate, his research interest includes machine learning.