激光与光电子学进展 先进成像

基于改进YOLOv4的遥感图像目标检测方法

肖振久，杨玥莹*，孔祥旭
辽宁工程技术大学软件学院，辽宁 葫芦岛 125105

摘要 遥感图像存在背景复杂、目标小且密集排列等问题，基于深度学习的目标检测方法可以提高目标检测的准确率，但是普遍存在模型参数量较多、检测速度一般的问题。针对上述问题，提出一种基于改进YOLOv4的遥感图像目标检测方法。首先采用轻量化网络Mobile NetV3代替YOLOv4的原特征提取网络，提高检测速度；其次在预测层中串联自注意力机制，使用改进非极大值抑制算法进行后处理；最后，在图像预处理中通过Mosaic方法进行数据增强，使用K-means方法获得更匹配遥感目标的候选框参数，在预测层中使用Complete Intersection Over Union（CIoU）损失函数进行坐标框定位。实验数据集由NWPU VHR-10和DOTA两个经典遥感数据集共同组成，包含船、车辆、港口等10个类别。结果表明，当遥感图像输入尺寸为608×608时，检测速度为54 frame/s，是YOLOv4检测速度的1.6倍，平均精度均值达到85.60%，所提方法在保持较高检测精度的同时，减小了参数量、提高了检测速度。

关键词 遥感；目标检测；遥感图像；YOLOv4；轻量化网络
中图分类号 TP391.4 **文献标志码** A **DOI**：10.3788/LOP213399

Object Detection Method Based on Improved YOLOv4 Network for Remote Sensing Images

Xiao Zhenjiu, Yang Yueying*, Kong Xiangxu
College of Software, Liaoning Technical University, Huludao 125105, *Liaoning, China*

Abstract Remote sensing images have many problems, such as complex background, small targets, and dense arrangement. The target detection method based on depth learning can improve the accuracy of target detection, but there are many problems, such as more model parameters and general detection speed. Aiming at the above problems, a remote sensing image target detection method based on improved YOLOv4 is proposed. First, the lightweight network Mobile NetV3 is used to replace the original feature extraction network of YOLOv4 to improve the detection speed; second, the self-attention mechanism is concatenated in the prediction layer, and the improved non maximum suppression algorithm is used for post-processing; finally, in the image preprocessing, Mosaic method is used to enhance the data, K-means method is used to obtain the candidate frame parameters that better match the remote sensing target, and Complete Intersection Over Union (CIoU) loss function is used in the prediction layer to locate the coordinate frame. The experimental data set consists of two classical remote sensing datasets, NWPUVHR-10 and DOTA, including 10 categories of ships, vehicles, and ports. The results show that when the input size of remote sensing image is 608×608, the detection speed is 54 frame/s, 1.6 times that of YOLOv4, and the average accuracy is 85.60%. The proposed method reduces the parameter amount and improves the detection speed while maintaining a high detection accuracy.

Key words remote sensing; object detection; remote sensing images; YOLOv4; lightweight network

1 引 言

目标检测是我国计算机视觉技术领域中应用研究的热门方向。随着航天遥感技术的发展，对遥感图像进行目标检测在空中侦察、无人机、军事航海、安全防护等人工智能技术领域有着广泛的应用[1-2]。因此，遥感图像目标检测方法及其应用有着重大的研究意义和价值[3-4]。

对自然场景图像进行多分类的目标检测已经有很多优秀的检测方法，如单阶段目标检测方法SSD[5]、YOLO系列[6-9]等，两阶段目标检测方法Faster R-

收稿日期：2021-12-30；**修回日期**：2022-01-17；**录用日期**：2022-01-21；**网络首发日期**：2022-01-30
基金项目：辽宁省教育厅科学技术研究项目（LJ2020JCL023）
通信作者：*719633801@qq.com

CNN[10]、Mask R-CNN[11]等，还有新兴目标检测方法FCOS[12]、CenterNet[13]等，这些方法在实际应用中都有良好的发展前景。对于遥感图像来说，在实际工程应用中不仅需要考虑检测的准确性，模型大小、参数量以及检测速度也尤为重要。就像在多数应用场景中，比如空中侦察、实时导弹预警、灾难检测救助等，不仅要保证准确率，还要对场景进行快速检测，这样才能更好地应用于实际工程中。使用基于深度学习的目标检测方法、选择更为深层次的卷积神经网络，模型复杂度便会提高，同时使得检测准确率也提高[14]。但是多数复杂模型都存在参数量多、模型较大的问题，同时对部署平台的计算性能要求高、存储设备需求大等，小型设备根本不能部署这些模型。能满足快速处理目标检测的应用设备普遍使用强大的独立显卡，如 Titan RTX、RTX2080Ti 等，显存一般都超过 10 GB，很难部署在无人机或卫星等内存、算力和质量受限的平台。因此，需要对算法进行压缩和加速。现今，常用的加速方法有两类：一类是通过模型剪枝[15]或者知识蒸馏来压缩模型[16]；另一类是使用更为轻量化的网络模型[17-18]。其中，压缩模型会导致卷积层提取的特征被破坏，使得检测精度下降，而轻量化的网络模型通过更为高效的卷积方式，如可分离卷积、深度可分离卷积等，尽可能地减少模型的参数并保证模型的性能。

目前也有很多专门针对遥感图像的目标检测网络。Yang 等[19]在 2018 年提出一种可对飞机目标进行检测的高效方法，该方法将残差网络和超矢量编码相结合以提高检测效果。同期 Xu 等[20]将特征融合技术应用到全卷积网络中，提高飞机目标定位精度。Liu 等[21]在 YOLOv3 特征提取网络中增加 CBL［由 Conv＋批归一化（BN）＋Leaky_ReLU 激活函数三者组成的模块］，更有效进行特征提取，用于航拍汽车检测。Xu 等[22]将密集连接网络加入 YOLOv3 中，代替部分残差连接，与原 YOLOv3 相比，对遥感目标的检测能力有所提高。

针对上述问题，本文提出一种基于改进 YOLOv4 的遥感图像目标检测方法。首先，针对遥感图像目标检测轻量快速的要求，采用轻量化网络 Mobile NetV3 进行特征提取，引入深度可分离卷积，降低参数量，减小模型大小，提升检测速度；其次，在预测层中串联自注意力机制，扩大卷积神经网络感受野，使用改进非极大值抑制（NMS）算法——Pyramid Shifted MaxpoolNMS with Relationship Recovery（PSRR-MaxpoolNMS）替换原非极大值抑制算法进行后处理，保证检测精度的同时，进一步提高检测速度；最后，针对遥感图像背景复杂、小目标多且密集的特征，在图像预处理中通过 Mosaic 方法进行数据增强，丰富遥感图像数据集，预先使用 K-means 方法获得更匹配遥感目标的候选框参数。在预测层中使用 Complete Intersection Over Union（CIoU）损失函数进行坐标框定位，进一步提升检测精度。

2　改进 YOLOv4 的遥感图像目标检测

2.1　改进 YOLOv4 的网络结构

改进 YOLOv4 的目标检测网络结构如图 1 所示。

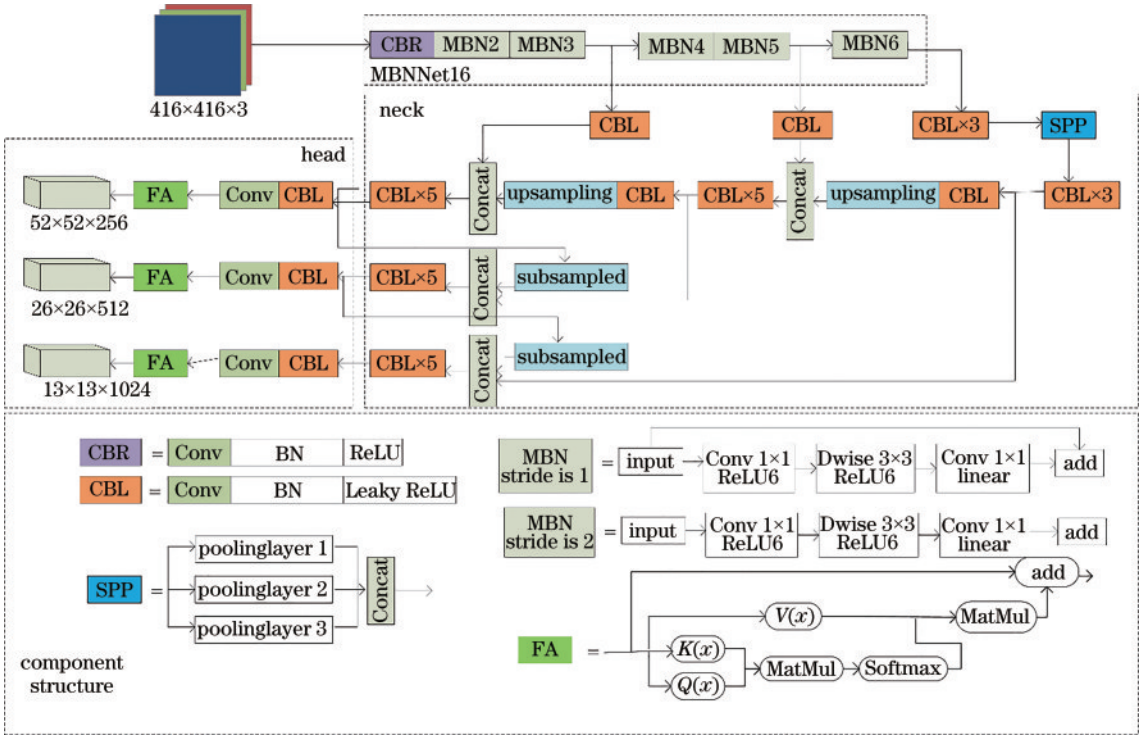

图 1　改进 YOLOv4 的目标检测网络结构

Fig. 1　Target detection network structure of improved YOLOv4

使用 Mobile NetV3 代替 CSPDarkNet53 进行特征提取，使用深度可分离卷积替换普通卷积，使 YOLOv4 更轻量化，在 head 部分添加自注意力机制。本实验采用多尺度检测，最后生成 3 种尺寸的特征图，分别为 13×13、26×26、52×52，将图片分成与其尺寸对应的网格，在每个网格中心建立多个先验框，预测判断框是否包含物体和物体种类。

2.1.1 特征提取网络

在 YOLO 系列中，YOLOv4 的准确率和检测效率都较好，原因是其选择以特征金字塔为基础的 CSPDarkNet53 作为特征提取网络。该网络通过特征金字塔结构预测不同尺寸的目标，参考 ResNet 残差结构，在某些层之间设置跳转连接，保持了较好的检测精度，但是其参数量和计算量都较大。所提网络使用 Mobile NetV3 代替 CSPDarkNet53 进行特征提取，在参数量和检测精度等方面保持了良好的平衡。保留 Mobile NetV2 中的线性瓶颈逆残差结构，引入 3×3 的深度可分离卷积，引入轻量级注意力模型 Squeeze-and-Excitation(SE)，调整每个通道的权重，使用 Swish 激活函数。在实际检测层中，选择 bottleneck 中的第 7、13 和 17 层进行提取，丢弃 17 层之后模块，并对这 3 层进行 pointwise(PW)卷积操作，变换维度使其与检测层连接，最后得到的这 3 层用来替换原来 YOLOv4 中的初步有效特征层。

加强特征提取网络采用空间金字塔池化，将卷积层通过 3 个不同尺寸的最大池化操作，将其拼接成 1 个一维向量，以适应不同尺寸图片的输入。采用路径聚合网络(PANet)中的特征图金字塔网络(FPN)进行特征融合，提取出更高质量的特征，将之前的初步有效特征层变为更高质量的用于预测的特征层。在所采用的卷积操作中，使用 3×3 的深度可分离卷积替代 3×3 的普通卷积，进一步实现模型轻量化。

2.1.2 自注意力机制

自注意力机制擅于发现数据或特征的内部相关性，减少对外部信息的依赖，用于模拟局部的长距离依赖关系。将目标检测问题重新定位为一个类似查询的问题，通过自注意力模块从输入的特征图中估计相关信息，构建特征图内部所有特征像素之间的全局依赖关系，针对不同区域进行更有效的目标检测。自注意力模块结构如图 1 所示。

设由特征层输出的预测特征图 $\boldsymbol{x}^t \in \mathbb{R}^{C^t \times N^t}$，$t \in \{1, \cdots, t\}$，其中 C 和 N 分别为特征映射中的通道数和总空间位置，t 为比例尺。首先将特征图 $\boldsymbol{x}^t$ 经过线性变换得到 3 个不同的特征空间 Q、K 和 V，矩阵 $\boldsymbol{W}_Q$、$\boldsymbol{W}_K$、$\boldsymbol{W}_V$ 是需要学习的矩阵，其公式分别为

$$Q(\boldsymbol{x}^t) = \boldsymbol{W}_Q^{t\top} \boldsymbol{x}^t, \boldsymbol{W}_Q^t \in \mathbb{R}^{C^t \times C'}, C' = C^t/8, \tag{1}$$

$$K(\boldsymbol{x}^t) = \boldsymbol{W}_K^{t\top} \boldsymbol{x}^t, \boldsymbol{W}_K^t \in \mathbb{R}^{C^t \times C'}, \tag{2}$$

$$V(\boldsymbol{x}^t) = \boldsymbol{W}_V^{t\top} \boldsymbol{x}^t, \boldsymbol{W}_V^t \in \mathbb{R}^{C^t \times C'}。 \tag{3}$$

接下来 $Q(\boldsymbol{x}^t)$ 和 $K(\boldsymbol{x}^t)$ 通过矩阵相乘(MatMul)得到注意力权重矩阵 $\boldsymbol{F}$，然后通过 Softmax 进行归一化操作，公式为

$$\boldsymbol{F}^t = Q(\boldsymbol{x}^t)^\top K(\boldsymbol{x}^t), \tag{4}$$

$$\overline{\boldsymbol{F}_{ij}}^t = \frac{\exp(\boldsymbol{F}_{ij}^t)}{\sum_j^{N^t} \exp(\boldsymbol{F}_{ij}^t)}, i, j = 1, 2, \cdots, N^t, \tag{5}$$

式中：$\overline{\boldsymbol{F}_{ij}}^t$ 为归一化后的预测特征图中坐标为 (i, j) 的像素点的注意力权重。通过得到的 $N \times N$ 的注意力权重矩阵，得出预测特征图中像素间的相关关系。

然后将 $V(\boldsymbol{x}^t)$ 和注意力权重 $\overline{\boldsymbol{F}}^t$ 进行矩阵相乘，得到预测特征图中每个位置的权重矩阵并加权求和。最后，将矩阵相乘结果添加到原预测特征图得到新的预测特征图 $\boldsymbol{x}^{t'}$，其公式为

$$\boldsymbol{x}^{t'} = \boldsymbol{x}^t + \left[\overline{\boldsymbol{F}}^t V(\boldsymbol{x}^t)^\top\right]^\top。 \tag{6}$$

2.2 改进非极大值抑制

NMS 在目标检测中用于提取置信度高的目标检测边界框，抑制置信度低的误检框。在预测阶段，会以输入图像的每个像素为中心生成多个大小和比例不同的边界框，NMS 就是用于去除这些多余的边界框，以便获得最终的预测边界框的算法。GreedyNMS 即传统的 NMS 算法是现今使用最广泛的算法，其先根据置信度分数降序对这些锚框进行排序，消除与最终预测有较大重叠的其他框，最后从剩余的框中选出最终的预测边界框。MaxpoolNMS 将 NMS 重新定义，通过最大池化操作删除冗余框，实现 NMS 的加速。但是 MaxpoolNMS 只能应用于两阶段目标检测方法中的第 1 阶段，不能使用于单阶段目标检测方法的后处理阶段。所以使用 PSRR-MaxpoolNMS 替换原 NMS，引入简单的关系恢复模块和 Pyramid Shifted MaxpoolNMS 模块，PSRR-MaxpoolNMS 比 MaxpoolNMS 更贴近 NMS，可以抑制跨通道的重复边界框，减少网络参数量，加速第 1 阶段目标检测方法的后处理过程。

关系恢复模块将经微调边界框得到的回归框映射到得分图上，回归框在空间位置、大小和形状方面通常更准确地包围预测目标。该模块具体由空间恢复、通道恢复和分数分配等 3 部分组成。空间恢复对于给定输入图像中的 1 个回归框的中心位置 $[x_c, y_c]$，使得映射到得分图上的空间索引为 $[X, Y]$，公式为

$$\begin{cases} X = \left[\dfrac{x_c}{\alpha}\right] \\ Y = \left[\dfrac{y_c}{\alpha}\right] \end{cases}, \tag{7}$$

式中：α 为得分图的下采样率。通常根据回归框的默认比例 s_0 和比率 r_0，将边界框映射到得分图的 1 个通道 $C(s_0, r_0)$，但是如果回归框的比例和比率出现巨大变化，会造成通道映射错误。通道恢复是为了避免这种情况，将一个长宽为 W、H 的回归框，通过欧氏距离计

算最接近的比例 $s=W\times H$、$r=\frac{H}{W}$，并选择 $C(s,r)$ 作为映射通道。在确定所有回归框的空间位置和通道后，得分图的每个单元格都不止有 1 个回归框，分数分配将通过 1×1 最大池化操作保留每个单元格分数最高的回归框。这 3 部分操作简单，具有高度的并行性，只需要回归框的位置和大小，是无需先验框的方法。

Pyramid Shifted MaxpoolNMS 模块包含 Pyramid MaxpoolNMS 和 Shifted MaxpoolNMS 两部分操作。Pyramid MaxpoolNMS 通过在不同信道组合的置信得分图上一个接一个进行最大池化操作，逐步诱导得分图变得稀疏，如图 2 所示。从图 2 可以清楚看到，随着 Pyramid MaxpoolNMS 操作：先经过单通道最大池，单通道在单分数图上独立运行，通道内核大小和步长均为 1；然后是交叉比率和交叉比例最大池，通过相邻比率或者比例串联通道在多分数图上运行；最后是交叉所有通道最大池，在所有通道上运行。跨多通道操作最大池时内核大小和步长为连接通道的内核大小和步长的最小值。通过这样的方式，将感受野从局部的单分数图增加到全局所有分数图。Shifted MaxpoolNMS 通过引入具有空间移位的额外最大池，给定内核大小为 1，Shifted MaxpoolNMS 将在得分图边缘用 0 填充 1/2 的范围，可以进一步增加分数图的稀疏性，有效消除重叠框。

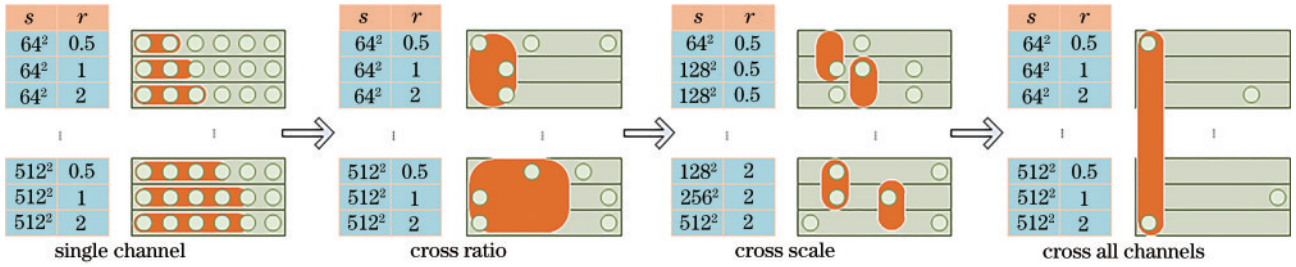

图 2　Pyramid MaxpoolNMS 操作流程

Fig. 2　Operating process of Pyramid MaxpoolNMS

2.3　损失函数

在训练过程中，损失函数由定位损失函数、分类损失函数和置信度损失函数组成，公式为

$$L_{\text{yolo}}=L_{\text{loc}}+L_{\text{cla}}+L_{\text{conf}}。\tag{8}$$

只有出现所需检测的目标时，才会有定位和分类损失，增加惩罚项。而置信度损失则完全不一样，当没有出现所需检测的目标时希望置信度为 0，当存在所需检测的目标时，希望达到较高的置信度，其意味着一个边界框位置的准确性和某个目标的可能大小。置信度采用交叉熵损失函数，计算公式为

$$\begin{aligned}L_{\text{conf}}=L_{\text{conf+obj}}+L_{\text{conf+noobj}}=&\sum_{i=0}^{G\times G}\sum_{j=0}^{B}I_{ij}^{\text{obj}}\left[C_i\log(\hat{C}_i)+\right.\\&\left.(1-C_i)\log(1-\hat{C}_i)\right]-\sum_{i=0}^{G\times G}\sum_{j=0}^{B}I_{ij}^{\text{noobj}}\left[C_i\log(\hat{C}_i)+\right.\\&\left.(1-C_i)\log(1-\hat{C}_i)\right],\end{aligned}\tag{9}$$

式中：obj 和 noobj 分别表示是否出现所需检测的目标；$G\times G$ 为检测层的网格数；B 为边界框；C 为置信度真实值；$\hat{C}_i$ 为置信度预测值；当第 i 个网格的第 j 个边界框不存在所需检测的目标时，I_{ij}^{noobj} 等于 1，若存在，则等于 0；而 I_{ij}^{obj} 则刚好与之相反。

分类损失函数公式为

$$\begin{aligned}L_{\text{cla}}=&\sum_{i=0}^{G\times G}\sum_{J=0}^{B}\sum_{c=0}^{M}I_{ij}^{\text{obj}}\left[\hat{p}_{ij}^{c}\log(p_{ij}^{c})+(1-\hat{p}_{ij}^{c})\times\right.\\&\left.\log(1-p_{ij}^{c})\right],\end{aligned}\tag{10}$$

式中：M 为遥感图像数据集的类别个数，即为 10；$\hat{p}_{ij}^{c}$ 为其中某一类的真实概率值；p_{ij}^{c} 为某一类的预测概率值，且只计算包含所需检测的目标的分类损失。

定位损失函数使用 CIoU 损失函数。首先介绍交并比（IoU），IoU 一般用来衡量目标检测算法所得到的边界框和实际边界框的相似程度，先设定一个阈值，若 IoU 大于这个阈值，则将边界框认定预测较为准确，否则认定预测错误。IoU 的计算公式为

$$R_{\text{IoU}}=\frac{\text{area}(d)\cap\text{area}(g)}{\text{area}(d)\cup\text{area}(g)},\tag{11}$$

式中：area 表示边界框的面积；d 代表预测边界框；g 代表真实边界框。若两个边界框完全重合，则该值为 1。CIoU 损失函数公式为

$$R_{\text{CIoU}}=1-R_{\text{IoU}}+\frac{\rho^2\left[\text{area}(d),\text{area}(g)\right]}{c^2}+\alpha v,\tag{12}$$

$$\alpha=\frac{v}{(1-R_{\text{IoU}})+v},\tag{13}$$

$$v=\frac{4}{\pi^2}\left(\arctan\frac{w^{\text{gt}}}{h^{\text{gt}}}-\arctan\frac{w}{h}\right)^2,\tag{14}$$

式中：ρ^2 为预测边界框和真实框中心点距离的平方；c^2 为刚好可以包括预测边界框和真实框的最小对角线的长度。最终定位损失函数公式为

$$L_{\text{loc}}=\lambda\sum_{i=0}^{G\times G}\sum_{j=0}^{B}I_{ij}^{\text{obj}}(1-R_{\text{CIoU}_{ij}}),\tag{15}$$

式中：λ 为位置定位损失函数的权重。

2.4　遥感图像预处理

遥感图像每张图片中样本数量不同，且分布不均

匀，所以对遥感图像训练集先进行Mosaic图像增强操作。Mosaic图像增强方法是从遥感图像训练集中取出一个批处理数据，随机使用4张图片进行不同比例的缩放或者扩大，按照矩形4个角的方向放置，将多余部分进行裁剪，最后拼接成1张固定大小的图片进行训练，如图3所示。采用该方法进行图像预处理有3个优点：1）可以丰富遥感图像数据集，当1张图片目标过少时，拼接后的图片就会包含多个目标，随机缩放也使得目标尺寸更为全面；2）对于遥感图像中小目标的检测性能有所帮助，使得小目标检测更为准确，增加了小目标的目标个数；3）一般情况下为了更好地训练网络，通常将批处理的尺寸尽可能地设置大一些，当4张图片拼为1张进行训练时，变相增加了批处理的图像个数，当设备GPU显存受限的时候，可以更快速进行训练，使得模型有更好的鲁棒性。

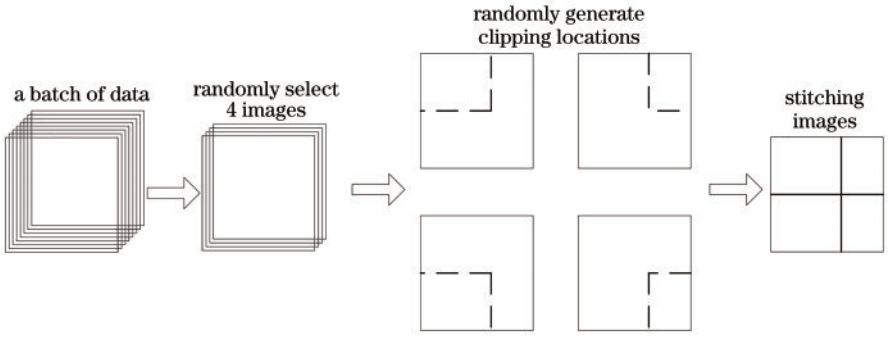

图3 Mosaic数据增强原理

Fig. 3 Principle of Mosaic data enhancement

YOLOv4使用的先验框机制是针对自然场景下的目标检测，为了使得模型能够更适合遥感图像的目标检测，采用K-means算法得到适合遥感图像的先验框参数大小，以便能够更好地预测遥感图像中的目标，提高检测精度。计算流程如下：

1）得到遥感图像训练集中所有目标的先验信息；

2）由于有3个检测层，所以随机初始化9个聚类，作为先验框的宽和高；

3）将目标和9个聚类的中心进行比较，分配给距离最短的类别，形成新的9个簇，并更新聚类中心；

4）重复执行步骤3），直到类别的中心值不变。

2.5 遥感图像目标检测方法

遥感图像目标检测方法包括两个阶段。在训练阶段：首先准备遥感图像训练数据集，使用Mosaic数据增强进行图像预处理，设置训练参数；然后加载预训练的改进YOLOv4模型权重，并使用K-means算法得到先验框参数大小；最后开始训练模型，直到损失收敛，最终得到改进YOLOv4的权重文件。在测试阶段：首先输入待检测图片，并加载已训练的改进YOLOv4模型，通过特征提取得到3种尺寸的特征图，尺寸分别为13×13、26×26、52×52；然后按照特征图的尺寸分别进行检测，预测目标分类和边界框位置；最后对预测结果进行得分排序，使用改进非极大值抑制过滤冗余的边界框，得到最终预测框信息，并在图像上框出目标位置、目标类别和预测得分。

3 实验结果及分析

3.1 数据集

本实验所使用的遥感数据集来自NWPU VHR-10数据集和DOTA数据集。NWPU VHR-10数据集是10类地理空间物体检测数据集，由西北工业大学收集并公开。该数据集共有800张高分辨率的VHR光学遥感图像。DOTA数据集是航拍图像中最大的目标检测遥感数据集，一共有15类目标，其中包括NWPU VHR-10数据集中10个类别的目标，选取其中620张图像，图像中的目标类别与NWPU VHR-10数据集的10个类别一致，将选出的图像进行水平标注目标。最终数据集一共包含1270张图像，具体类别如表1所示。

表1 数据集类别

Table 1 Dataset category

Class	plane	ship	tank	tennis court	basketball court	baseball court	track-and-field ground	bridge	port	vehicle
Training set	1290	661	1283	1076	204	370	165	142	302	1020
Test set	555	284	550	462	89	159	74	62	131	437
Total	1845	945	1833	1538	293	529	239	204	433	1457

本实验选择RSOD数据集来验证模型的迁移泛化能力。RSOD数据集由武汉大学标注提供，包含飞机、田径场、立交桥和油罐等4种类别，使用VOC数据集格式进行标注。其中，有446张图像包含4993个架飞机，

189张图像包含191个田径场，176张图像包含180座立交桥，164张图像包含1586个油罐，共975张图像。

3.2　评价指标

对于检测效果的评价，通常选用平均精度（AP）衡量单目标检测效果，选用平均精度均值（mAP）衡量多目标检测效果。对于模型大小的评价，通常使用参数量（parameter）和计算量来衡量，参数量和计算量越小，模型越轻量化。对于检测速度的评价，通常使用每秒钟检测图像的帧数（FPS）来衡量，FPS越大算法的检测速度越快。本实验是对10种类别的遥感图像数据集进行目标检测，所以选用AP、mAP、FPS、参数量和模型大小作为模型的评价指标。其中，AP为*P*-*R*曲线的面积，AP越高代表检测性能越好，mAP为所有类别AP的平均值，计算公式为

$$P=\frac{N_{TP}}{N_{TP}+N_{FP}},\tag{16}$$

$$R=\frac{N_{TP}}{N_{TP}+N_{FN}},\tag{17}$$

$$P_{AP}=\int_0^1 P(R)\mathrm{d}R,\tag{18}$$

$$P_{mAP}=\sum_{i=1}^{N}\frac{P_{AP_i}}{N},\tag{19}$$

式中：N_{TP}为真样本，表示检测到的目标类别与真实类别一致的样本；N_{FP}为假样本，表示检测到的目标类别与真实类别不一致的样本；N_{FN}为假负样本，表示实际存在的目标，但未被检测出的样本。图4为改进YOLOv4算法中类别油桶的*P*-*R*曲线，*P*-*R*曲线所围成的面积即为该类别的AP值。

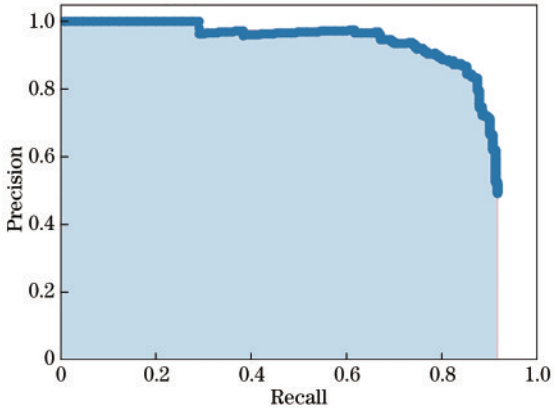

图4　改进YOLOv4算法类别中油罐的*P*-*R*曲线

Fig. 4　*P*-*R* curve of oil tank in improved YOLOv4 algorithm category

3.3　实验结果与分析

实验使用的深度学习框架为Pytorch 1.2.0，显卡为NVIDIA GeForce GTX 1080Ti，CPU为i7-8700k，内存为16 GB，硬盘为512 GB，操作系统为64位的Windows 10。实验将输入图像尺寸设置为608×608，batch size设置为4，epoch设置为1000，学习率设置为0.0001，采用余弦退火策略调整学习率，采用Mosaic数据增强方法进行数据增强。由于原始anchor大小对于遥感目标不匹配，采用K-means算法得到的先验框参数取值为（13，22）、（19，24）、（20，34）、（23，52）、（30，31）、（35，52）、（49，79）、（111，192）、（232，245），并将其应用到改进前后的网络中。图5为改进后的YOLOv4模型训练时CIoU损失曲线图，在约100个epoch时，模型开始逐渐收敛，震荡越来越微弱，最终稳定在2.6左右。

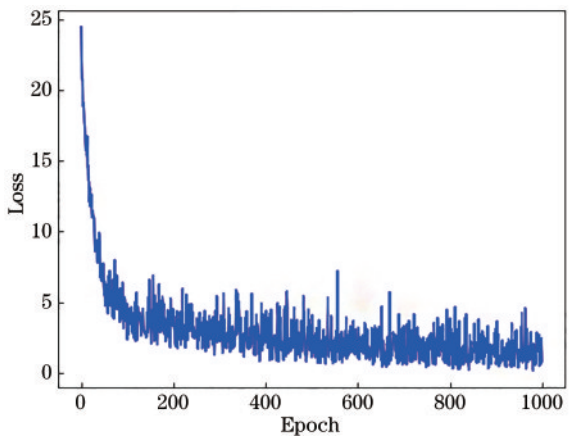

图5　模型训练时损失变化趋势

Fig. 5　Loss trend during model training

为验证改进模块的有效性，进行消融实验，结果如表2所示。添加自注意力机制后，检测精度有了一定的提升，检测速度小幅度降低。使用轻量化网络进行特征提取，并使用改进非极大值抑制进行后处理，检测精度小幅度降低，为85.62%，但是检测速度有了较大的提升，FPS为59，是原YOLOv4检测速度的1.7倍。在原YOLOv4基础上引入自注意力机制、改进NMS，提高了检测精度，检测精度达到了87.23%，FPS为35，与原YOLOv4相比检测精度和检测速度均有所提升。所提算法同时使用3种改进方法，检测精度达到86.50%，相比原YOLOv4虽然减少，但是FPS达到了54，提高了21。从表2可以清楚看到，每个改进方法均在遥感目标检测方法中起到重要的作用，相辅相成，缺一不可。

使用上述遥感数据集进行训练并测试，对Faster R-CNN、SSD、YOLOv3、YOLOv4、YOLOv5s算法、所提算法在参数量、模型大小和FPS方面进行对比，数据如表3所示。所提算法参数量为1.17×10^7，模型大小为44.7 MB，FPS达到了54。虽然在这3个指标中，所提算法与YOLOv5s相比，均没有达到最好，但是相比Faster R-CNN、SSD和其他YOLO系列，参数量和模型大小都有很大的减少。模型大小相比YOLOv3和YOLOv4下降了近80%。所提算法有效降低了模型大小和参数量，且保持了较高的检测速度。

为验证改进YOLOv4算法的有效性，使用本实验

表 2 不同改进方法的性能比较

Table 2 Performance comparison of different improvement methods

Method	MobileNetV3	Self-attention	PSRR-MaxpoolNMS	mAP /%	FPS
YOLOv4				87.17	33
1	√			85.16	57
2		√		87.67	29
3			√	86.19	36
4	√	√		86.81	49
5	√		√	85.62	59
6		√	√	87.23	35
Proposed method	√	√	√	86.50	54

表 3 不同方法的参数量、模型大小和 FPS 对比

Table 3 Comparison of parameters, model size, and FPS of different methods

Method	Parameter / 10^6	Model size / MB	FPS
Faster R-CNN	76.81		7
SSD512	62.54	102.50	18
YOLOv3	63.95	246.50	19
YOLOv4	61.13	244.30	33
YOLOv5s	7.20	35.70	63
Propoed Method	11.70	44.70	54

遥感图像数据集进行对比实验，对比 Faster R-CNN、SSD、YOLOv3、YOLOv4、YOLOv5s 算法在遥感图像测试数据集上的 AP 值和 mAP 值。Faster R-CNN、SSD、YOLOv3、YOLOv4、YOLOv5s 算法在测试遥感数据集上的 AP 值和 mAP 值如表 4 所示。所提算法的 mAP 值为 86.50%，和 Faster R-CNN、SSD、YOLOv3 相比，分别提升了 5.54 个百分点、10.7 个百分点、1.6 个百分点，相比 YOLOv4 降低了 0.67 个百分点，相比 YOLOv5s 轻量级算法提升了 4.6 个百分点。虽然相比 YOLOv4 检测精度稍微降低，但是在多数类别的 AP 值和其他算法比较均有不同程度的提升。

结合表 2 和表 3 的参数量、模型大小、检测速度和 mAP 值对比结果可得，YOLOv4 相比 YOLOv3 参数量和模型大小下降不多，但是检测速度得到了提升，而 YOLOv5s 与 YOLOv4 相比，虽然模型达到最小，检测速度也最快，但是检测精度却下降了 4.6 个百分比。所提算法在 YOLOv4 的基础上进行改进，虽然没有达到最好的检测效果但是保持检测性能稳定，检测速度高于 YOLOv4。

表 4 不同方法在各个类别上的检测结果对比

Table 4 Comparison of different detection methods in each category unit: %

Method	Faster R-CNN	SSD	YOLOv3	YOLOv4	YOLOv5s	Propoed method
plane	0.94	0.84	0.99	0.99	0.95	0.99
ship	0.82	0.62	0.84	0.82	0.76	0.87
tank	0.65	0.78	0.89	0.88	0.86	0.87
tennis court	0.82	0.79	0.99	0.95	0.88	0.99
basketball court	0.89	0.88	0.60	0.81	0.70	0.75
baseball court	0.95	0.89	0.96	0.95	0.90	0.97
track-and-field ground	0.92	0.80	0.98	0.94	0.92	0.97
bridge	0.58	0.65	0.62	0.66	0.71	0.64
port	0.72	0.71	0.91	0.87	0.84	0.89
vehicle	0.77	0.62	0.70	0.81	0.67	0.71
mAP	80.96	75.80	84.90	87.17	81.90	86.50

图 6 为改进前后部分遥感图像目标检测结果的对比图：原 YOLOv4 算法对于如飞机、船和油罐这种目标分布无规律且目标较小的类别，会出现定位不准确、漏检的情况，如图 6(a1)、(a2)、(a3)所示；但是使用改进后的 YOLOv4 算法进行检测时，定位更加准确，且没有发生漏检的情况，如图 6(b1)、(b2)、(b3)所示。YOLOv4 算法对于与背景颜色相似的车辆，因为区别太小，检测结果很不理想，出现很多漏检情况，如图 6(a4)所示。而改进后的 YOLOv4 算法在这种情况下，也能检测出多数车辆，如图 6(b4)所示。

为了验证改进 YOLOv4 遥感图像目标检测方法的普适性，将该算法在 RSOD 数据集上进行实验，RSOD 数据集是由武汉大学收集整理的遥感数据集，主要包括飞机、油桶等 4 类目标，其中飞机数量最多。

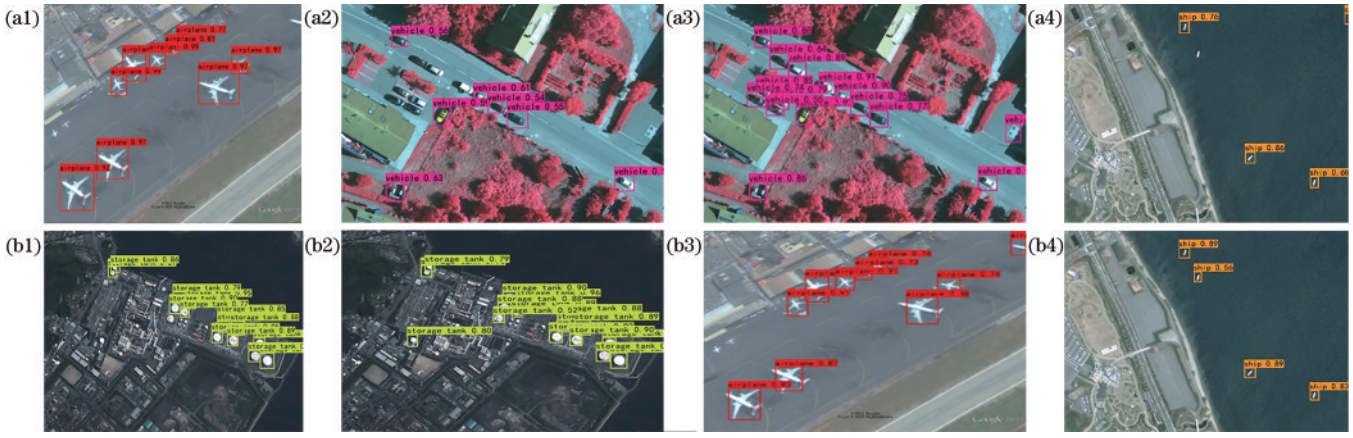

图 6　改进前后部分检测结果对比图。(a) YOLOv4 算法；(b)改进后 YOLOv4 算法

Fig. 6　Comparison of test results before and after improvement. (a) YOLOv4 algorithm; (b) improved YOLOv4 algorithm

实验参数与之前保持一致，结果如图 7 所示。从图 7 可以看出，改进 YOLOv4 算法的 mAP 值达到了 89.33%，达到了较好的检测效果，其中立交桥的 AP 值最低，分析发现可能由于立交桥目标数量过小，样本不均衡。

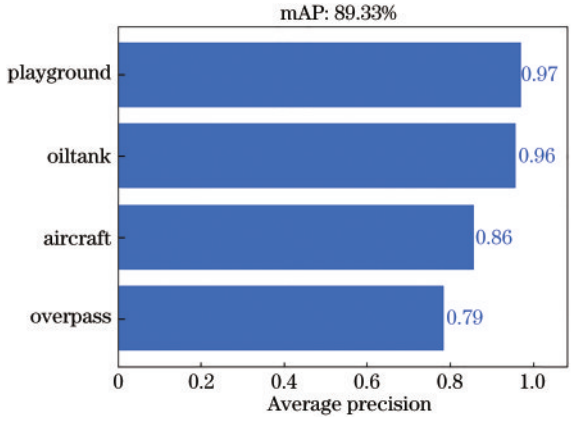

图 7　改进 YOLOv4 算法在 RSOD 上的 AP 和 mAP

Fig. 7　AP and mAP of improved YOLOv4 algorithm on RSOD dataset

4　结　　论

本实验提出一种基于改进 YOLOv4 的遥感图像目标检测方法。首先针对遥感图像目标检测轻量快速的要求，采用轻量化网络 Mobile NetV3 代替 YOLOv4 的原特征提取网络，检测速度为原来的 1.7 倍；其次在预测层中串联自注意力机制，针对遥感图像背景复杂、小目标多且密集的特征，在图像预处理中通过 Mosaic 方法进行数据增强，预先使用 K-means 方法获得更匹配遥感目标的候选框参数，提升检测精度。实验结果表明，所提方法在 NWPU VHR-10 和 DOTA 数据集共同组成的遥感图像数据集上取得良好的表现，与原 YOLOv4 方法相比在检测速度方面有明显的提升。在接下来的工作中，还需在嵌入式平台进行遥感目标检测的仿真实验，进一步研究如何更好地提高所提方法的检测效果。

参　考　文　献

[1] Ševo I, Avramović A. Convolutional neural network based automatic object detection on aerial images[J]. IEEE Geoscience and Remote Sensing Letters, 2016, 13(5): 740-744.

[2] 汪鹏, 刘瑞, 辛雪静, 等. 基于残差网络的光学遥感图像场景分类算法[J]. 激光与光电子学进展, 2021, 58(2): 0210001.
Wang P, Liu R, Xin X J, et al. Scene classification of optical remote sensing images based on residual networks [J]. Laser & Optoelectronics Progress, 2021, 58(2): 0210001.

[3] 王彦情, 马雷, 田原. 光学遥感图像舰船目标检测与识别综述[J]. 自动化学报, 2011, 37(9): 1029-1039.
Wang Y Q, Ma L, Tian Y. State-of-the-art of ship detection and recognition in optical remotely sensed imagery [J]. Acta Automatica Sinica, 2011, 37(9): 1029-1039.

[4] 汪亚妮, 汪西莉. 基于注意力和特征融合的遥感图像目标检测模型[J]. 激光与光电子学进展, 2021, 58(2): 0228003.
Wang Y N, Wang X L. Remote sensing image target detection model based on attention and feature fusion[J]. Laser & Optoelectronics Progress, 2021, 58(2): 0228003.

[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] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.

[7] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6517-6525.

[8] Redmon J, Farhadi A. Yolov3: an incremental

improvement[EB/OL]. (2018-04-08)[2021-02-05]. https://arxiv.org/abs/1804.02767.

[9] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2021-02-05]. https://arxiv.org/abs/2004.10934.

[10] 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.

[11] He K M, Gkioxari G, Dollar P, et al. Mask R-CNN [C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice. New York: IEEE Press, 2017: 386-397.

[12] Tian Z, Shen C H, Chen H, et al. FCOS: fully convolutional one-stage object detection[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 9626-9635.

[13] Duan K W, Bai S, Xie L X, et al. CenterNet: keypoint triplets for object detection[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 6568-6577.

[14] Zhang W H, Jiao L C, Liu X, et al. Multi-scale feature fusion network for object detection in VHR optical remote sensing images[C]//IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, July 28-August 2, 2019, Yokohama, Japan. New York: IEEE Press, 2019: 330-333.

[15] Han S, Pool J, Tran J, et al. Learning both weights and connections for efficient neural networks[C]//NIPS'15: Proceedings of the 28th International Conference on Neural Information Processing Systems-Volume 1, December 7, 2015, Cambridge, MA, United States. New York: ACM Press, 2015: 1135-1143.

[16] Cheng Y, Wang D, Zhou P, et al. A survey of model compression and acceleration for deep neural networks [EB/OL]. (2017-10-23)[2021-02-05]. https://arxiv.org/abs/1710.09282.

[17] Howard A G, Zhu M, Chen B, et al. Mobilenets: efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-17)[2021-02-05]. https://arxiv.org/abs/1704.04861.

[18] Ma N N, Zhang X Y, Zheng H T, et al. ShuffleNet V2: practical guidelines for efficient CNN architecture design [M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11218: 122-138.

[19] Yang J C, Zhu Y H, Jiang B, et al. Aircraft detection in remote sensing images based on a deep residual network and Super-Vector coding[J]. Remote Sensing Letters, 2018, 9(3): 228-236.

[20] Xu Y L, Zhu M M, Xin P, et al. Rapid airplane detection in remote sensing images based on multilayer feature fusion in fully convolutional neural networks[J]. Sensors, 2018, 18(7): 2335.

[21] Liu M J, Wang X H, Zhou A J, et al. UAV-YOLO: small object detection on unmanned aerial vehicle perspective[J]. Sensors, 2020, 20(8): 2238.

[22] Xu D Q, Wu Y Q. Improved YOLO-V3 with DenseNet for multi-scale remote sensing target detection[J]. Sensors, 2020, 20(15): 4276.