激光与光电子学进展 先进成像

基于改进YOLOv8算法的遥感图像目标检测

张秀再[1,2*]，沈涛[1]，许岱[1]

[1]南京信息工程大学电子与信息工程学院，江苏 南京 210044；

[2]南京信息工程大学江苏省大气环境与装备技术协同创新中心，江苏 南京 210044

摘要 针对遥感图像目标检测算法漏检和误检率高、目标定位不精确、无法准确识别目标类别等问题，提出一种基于改进YOLOv8的目标检测算法。为提高模型的损失函数对梯度分配的灵活性，适应各种形状和尺寸的物体，设计了非单调聚焦机制与边界框几何因素相结合的边界框回归损失函数；为扩大模型的感受野并削弱遥感图像背景对检测目标的影响，采用全局注意力机制与残差块结合的方式，设计了残差全局注意力机制；为使模型适应遥感图像中目标物体的形变与不规则排列，对YOLOv8模型中的C2f模块进行改进，融入可变形卷积与可变形感兴趣区域池化层。实验结果表明，在DOTA数据集和RSOD数据集上，所提算法的平均精度均值(mAP@0.5)达到72.1%和94.6%，优于对比算法，提高了遥感图像目标检测精度，为遥感图像识别提供了新的手段。

关键词 目标检测；YOLOv8；WIoU；全局注意力机制；可变形卷积

中图分类号 TP393 **文献标志码** A **DOI**：10.3788/LOP231803

Remote-Sensing Image Object Detection Based on Improved YOLOv8 Algorithm

Zhang Xiuzai[1,2*], Shen Tao[1], Xu Dai[1]

[1]*School of Electronic and Information Engineering, Nanjing University of Information Science & Technology, Nanjing* 210044, *Jiangsu*, *China*;

[2]*Jiangsu Province Atmospheric Environment and Equipment Technology Collaborative Innovation Center, Nanjing University of Information Science & Technology, Nanjing* 210044, *Jiangsu*, *China*

Abstract A target detection algorithm based on improved YOLOv8 is proposed to address the issues of high-missed and false-detection rates, inaccurate target positioning, and inability to accurately identify target categories in remote-sensing image target detection algorithms. To improve the flexibility of the loss function of the model in gradient allocation and adapt to various object shapes and sizes, a boundary box regression loss function is designed, which combines a nonmonotonic focusing mechanism with geometric factors of the boundary box. To expand the receptive field of the model and weaken the influence of the remote-sensing image background on the detection target, a residual global attention mechanism is designed by combining global attention mechanism and residual blocks. To adapt the model to the deformation and irregular arrangement of target objects in remote-sensing images, the C2f module in the YOLOv8 model is improved by incorporating deformable convolution and deformable region-of-interest pooling layers. Experimental results show that on DOTA and RSOD datasets, mean average precision (mAP@0.5) of the improved YOLOv8 algorithm reaches 72.1% and 94.6%, which are better than other mainstream algorithms. It improves the accuracy of remote sensing image target detection and provides a new means for remote sensing image target detection.

Key words target detection; YOLOv8; WIoU; global attention mechanism; deformable convolution

1 引 言

目标检测[1]技术是计算机视觉领域的一个重要研究方向，近年来受到了广泛的关注。目标检测技术已经被应用于各种领域，包括自动驾驶、安防监控、无人机、医疗影像等。在这些领域中，目标检测可以自动地从图像或视频中识别和跟踪目标，为人们的生产和生活带来便利。

目标检测是指在图像或视频中自动识别并定位出

收稿日期：2023-07-31；**修回日期**：2023-10-04；**录用日期**：2023-10-13；**网络首发日期**：2023-11-07

基金项目：国家社会科学基金一般项目(22BZZ080)

通信作者：*zxzhering@163.com

目标物体的位置和类别。根据检测方法的不同，目标检测算法可以分为一阶段算法和二阶段算法。一阶段算法是指直接从图像中提取特征，然后使用分类器或回归器来预测物体的位置和类别。常见的一阶段算法包括YOLO[2]、SSD[3]等。二阶段算法则将目标检测问题分为两个子任务：目标提议和目标分类。首先，需要提取出所有可能包含目标的区域，即目标提议。然后，针对每个提议区域进行分类，确定其是否包含目标以及目标的类别。常见的二阶段算法包括Faster R-CNN[4]等。相比而言：一阶段算法通常具有较快的检测速度和较低的计算复杂度，但检测精度略低；而二阶段算法则具有更高的检测精度，但速度较慢、计算复杂度较高。

目标检测技术在遥感图像中的应用广泛。通过目标检测技术能够自动识别和定位遥感图像中的不同地物，如建筑物、道路、水体和森林，为城市规划、环境监测和资源管理提供重要支持。此外，目标检测还能用于监测和响应自然灾害对地表的影响，包括洪水、火灾和地震等灾害事件。通过目标检测技术，可以实现对遥感图像的高效分析和地物提取，为各种应用领域提供有价值的信息和决策支持。

现阶段的目标检测算法在遥感图像中的应用仍存在部分缺陷：遥感图像包含众多小目标，如车辆或设备，当前主流算法对这些目标漏检率与误检率较高；遥感图像中的目标有方向性多、纵横比大等特点，对于模型适应目标的旋转与目标的特殊尺寸要求高；由于光照条件和拍摄视角的变化，遥感图像中的目标可能呈现出不同的外观，当前的目标检测算法对于光照和视角变化的适应性有限，容易产生误检或漏检。

为解决目标检测算法在遥感图像中的缺陷：刘涛等[5]提出一种改进的YOLOv5算法，引入通道注意力机制以提升特征捕获与融合能力，增加融合浅层语义信息的细粒度检测层提高对小目标的检测效果，并用Copy-Paste进行数据增强；张正等[6]提出一种改进SFP-DETR算法，结合膨胀卷积设计单级特征金字塔结构，构造新的边界框回归损失，实现了无锚框旋转目标检测，并在解码器的交叉注意力上加入权重约束，将全局注意力计算限制在局部范围内；原瑜蔓等[7]提出一种改进的FCOS模型，通过将多尺度特征缩放进行特征融合和优化，并利用适当的感受野提取高层特征的上下文信息。这些算法一定程度上提高了各类目标检测器的性能，但检测精度仍有提升空间。

针对遥感图像检测中漏检或误检率高、目标定位不精确、无法准确识别目标等问题，本文提出一种基于改进YOLOv8[8-9]的目标检测算法。改进边界框回归损失函数算法，将非单调聚焦机制与边界框几何因素相结合，提高边界框的质量；添加注意力机制并融入残差网络，扩大模型感受野，提高网络学习能力；对C2f模块进行改进，融入可变形卷积与可变形池化，使模型的采样更贴近目标的形状和尺寸。

2　YOLOv8网络模型及算法原理

YOLOv8算法模型主要由主干特征提取模块（Backbone）、特征加强模块（Neck）、检测模块（Detect）等3部分构成，如图1所示。

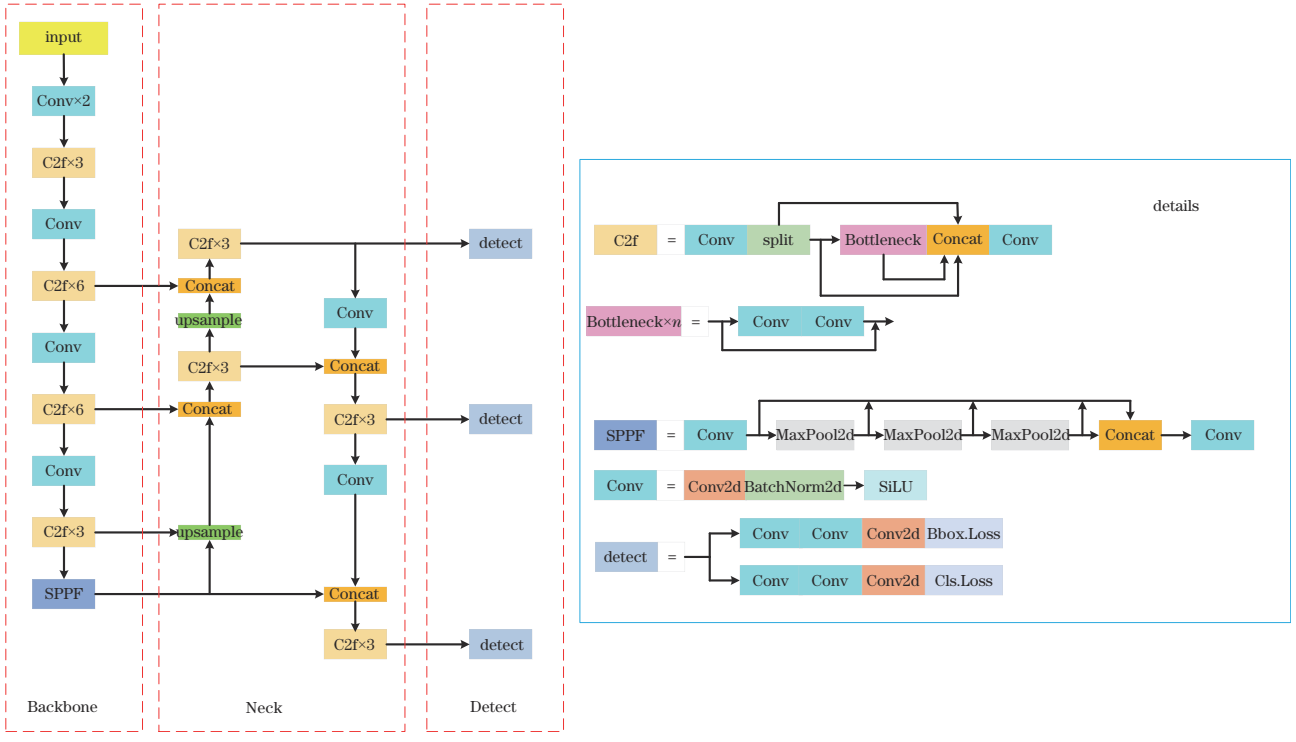

图1　YOLOv8算法模型结构

Fig. 1　YOLOv8 algorithm model structure

YOLOv8的主干特征提取模块沿用了CSPDarkNet结构，并采用梯度流更加丰富的C2f模块代替C3模块，并调整了不同尺度模型的通道数，降低了模型计算量，提升了收敛速度与收敛效果。特征加强模块采用Path Aggregation Network(PANet)结构，通过上采样、通道融合，最终将PANet的3个输出分支送入检测模块。检测模块采用解耦头结构进行检测，将回归分支和预测分支分离，加速模型收敛。

YOLOv8的主要思想是将目标检测问题转化为一个回归问题。YOLOv8将输入的图像分成$S\times S$大小的网格。对于每个网格，YOLO预测其中是否存在物体以及物体的位置和大小，并预测一组边界框，每个边界框由4个值确定，分别是中心点的坐标和边界框的宽度和高度。此外，每个边界框会分配一个表示该边界框中是否包含物体的置信度分数。由于同一个物体可能会被多个边界框检测，为了去除重复的检测结果，YOLOv8采用非极大值抑制算法[10]来筛选最终的检测结果并预测类别。

3　改进YOLOv8算法

基于YOLOv8模型对遥感图像目标检测漏检率高等问题进行改进，具体工作如下：1)由于遥感图像中目标的形状和尺寸多种多样，边界框的长宽比通常也具有较大的差异，为使得模型更好地适应各种形状和尺寸的物体，减少遥感图像中特殊尺寸目标的漏检率，通过考虑静态聚焦机制与目标的几何因素，设计DWIoU边界框损失函数；2)遥感图像背景复杂且检测目标占图像面积小，导致图像迭代冗余特征占比大，为使模型选择性增强包含目标信息量最大的特征通道，将注意力机制与残差结构结合，设计残差注意力模块；3)考虑到遥感图像中目标排列不规则、部分目标被遮挡，设计C2f-DCN模块，采用可变形卷积替代部分常规卷积，并融入可变性池化层，以适应遥感图像中目标物体的形变和旋转。改进YOLOv8结构如图2所示。

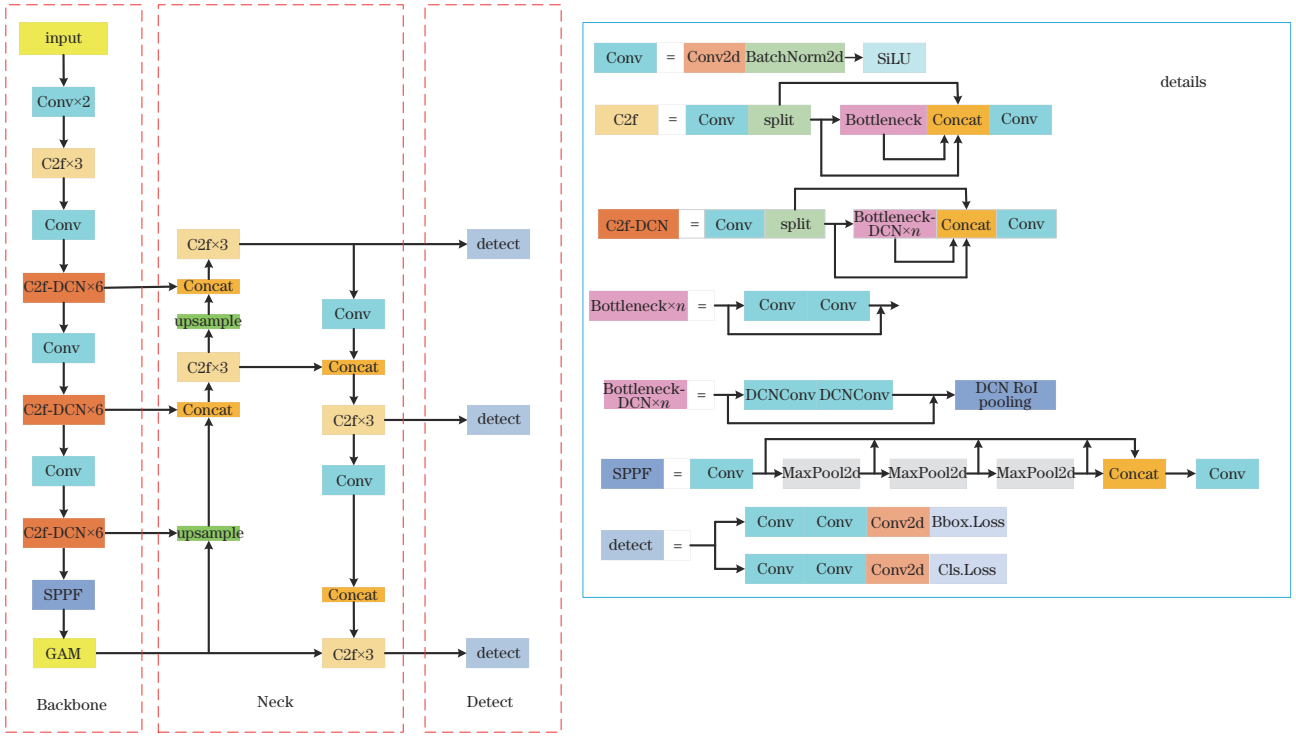

图2　改进YOLOv8结构

Fig. 2　Improved YOLOv8 structure

3.1　DWIoU

边界框回归损失函数对目标检测至关重要，模型通过学习预测边界框的位置，使其尽可能接近真实的边界框。

YOLOv8采用CIoU[11]作为边界框回归损失函数，如式(1)～(3)所示：

$$L_{\mathrm{CIoU}}=1-R_{\mathrm{IoU}}+\frac{(x-x_{\mathrm{gt}})^2+(y-y_{\mathrm{gt}})^2}{W_{\mathrm{g}}^2+H_{\mathrm{g}}^2}+\alpha\upsilon,\tag{1}$$

$$\alpha=\frac{\upsilon}{1-R_{\mathrm{IoU}}+\upsilon},\tag{2}$$

$$\upsilon=\frac{4}{\pi^2}\left(\arctan\frac{w}{h}-\arctan\frac{w_{\mathrm{gt}}}{h_{\mathrm{gt}}}\right)^2,\tag{3}$$

式中：IoU[12]表示预测框与真实框的交并比；x、y为预测框的中心点坐标；x_{gt}、y_{gt}为真实框的中心点坐标；W_{g}和H_{g}是预测框与真实框构成的最小矩形框的宽高值；α为超参数；w、h分别为预测框的宽高值；w_{gt}、h_{gt}分别为真实框的宽高值。

在CIoU中引入预测框与真实框的中心点距离与长宽比等因素，并对反向传播的梯度计算进行优化，但未

考虑难易样本的平衡问题，导致网络收敛速度慢且效率低。因此，引入WIoU[13]作为模型的边界框损失函数。

考虑到训练数据不可避免地包含低质量的例子，盲目地在低质量样本上加强边界框回归会降低模型的泛化性能。当预选框与目标框重合度较高时，良好的损失函数应该削弱几何因素的惩罚，使模型获得更好的泛化能力。WIoU采用非单调聚焦机制，通过构造动态的梯度增益系数，用离群度代替IoU来评价锚框的质量，提供了一种明智的梯度增益分配策略，离群度小意味着预选框的质量高，只需分配较低的梯度增益，以便将损失回归集中在普通质量的预选框上。WIoU如式(4)～(7)所示：

$$L_{\mathrm{WIoU}}=\gamma R_{\mathrm{WIoU}} L_{\mathrm{IoU}}, \tag{4}$$

$$\gamma=\frac{\beta}{\delta\alpha^{\beta-\delta}}, \tag{5}$$

$$\beta=\frac{L_{\mathrm{IoU}}^{*}}{\overline{L_{\mathrm{IoU}}}}, \tag{6}$$

$$R_{\mathrm{WIoU}}=\exp\left[\frac{(x-x_{\mathrm{gt}})^2-(y-y_{\mathrm{gt}})^2}{(W_{\mathrm{g}}^2+H_{\mathrm{g}}^2)^{*}}\right], \tag{7}$$

式中：δ、α为超参数(训练中δ取3，α取2)；γ为梯度增益；β为离群度；$\overline{L_{\mathrm{IoU}}}$是具有动量$m$的动态平均IoU值；为防止产生阻碍收敛的梯度，将$\beta$中的IoU、WIoU中的$W_{\mathrm{g}}$和$H_{\mathrm{g}}$从梯度计算中分离出来(上标*表示此操作)。由于$L_{\mathrm{IoU}}$是动态的，预测框的质量划分标准也是动态的，这使得WIoU能够在每一个时刻作出最符合当前情况的梯度增益分配策略。

在遥感图像的检测中，由于目标的形状和尺寸多种多样，边界框的长宽比通常也具有较大的差异。如果模型只关注中心坐标的回归，而忽略了边界框长宽比的影响，可能导致细长或压缩的物体难以准确检测。因此，为进一步达到合理的梯度分配，基于WIoU设计一种DWIoU算法，在尽可能减少几何因素对梯度分配的影响下，在WIoU中加入长宽比的惩罚项，使模型在适当关注长宽比的情况下进行准确回归。DWIoU的表达式如式(8)～(11)所示：

$$L_{\mathrm{DWIoU}}=\gamma R_{\mathrm{WIoU}} L_{\mathrm{IoU}}, \tag{8}$$

$$\gamma=\frac{\beta}{\delta\alpha^{\beta-\delta}}, \tag{9}$$

$$\beta=\frac{L_{\mathrm{IoU}}^{*}}{\overline{L_{\mathrm{IoU}}}}, \tag{10}$$

$$R_{\mathrm{WIoU}}=\exp\left[\frac{(x-x_{\mathrm{gt}})^2-(y-y_{\mathrm{gt}})^2}{(W_{\mathrm{g}}^2+H_{\mathrm{g}}^2)^{*}}-\left(1-\frac{w}{w_{\mathrm{gt}}}\right)\times\left(1-\frac{h}{h_{\mathrm{gt}}}\right)\right]。 \tag{11}$$

长宽比的引入使得DWIoU在聚焦普通质量预选框、削弱低质量预选框梯度分配的同时，在一定程度上考虑边界框的几何因素对边界框回归的影响，帮助模型更好地适应各种形状和尺寸的物体，减少遥感图像中特殊尺寸目标的漏检率，使模型具有更好的泛化性能。

3.2　残差注意力网络

注意力机制可以通过给予特定区域更大的权重来强调图像中的重要信息，促进不同层次或尺度之间的信息融合与交互，选择性增强包含目标信息量最大的特征通道，使模型更加聚焦于重要的区域以减小背景信息对模型的影响。常见的注意力机制有Squeeze-and-Excitation(SE)[14]、Convolutional Block Attention Module(CBAM)[15]等。SE是第一个使用通道注意力和通道特征融合来抑制不重要通道的网络，但在抑制不重要的像素方面效率较低；CBAM同时考虑空间维度和通道维度，但忽略了空间和通道的相互作用。因此，所提算法引入Global Attention Mechanism(GAM)[16]注意力机制，通过减少信息丢失和放大全局交互表示来提高模型的性能。

GAM沿用了CBAM的结构并改进通道、空间两个子模块，利用通道、空间宽度和空间高度等3个维度之间的注意力权重来提高效率，放大跨维度的相互作用。GAM注意力结构、通道注意力子模块和空间注意力子模块结构图分别如图3～5所示。

GAM注意力机制可由式(12)～(14)表示：

$$\boldsymbol{F}_2=M_{\mathrm{C}}(\boldsymbol{F}_1)\otimes\boldsymbol{F}_1=\mathrm{Sigmoid}\left[\boldsymbol{F}_1\cdot\mathrm{ReLU}(\boldsymbol{w}_2\boldsymbol{y}+\boldsymbol{b}_2)^{\mathrm{T}}\right], \tag{12}$$

$$\boldsymbol{y}=\boldsymbol{w}_1\boldsymbol{F}_1^{\mathrm{T}}+\boldsymbol{b}_1, \tag{13}$$

$$\boldsymbol{F}_3=M_{\mathrm{S}}(\boldsymbol{F}_2)\otimes\boldsymbol{F}_2=\mathrm{Sigmoid}\left\{\mathrm{ConvBN}\left[\mathrm{ConvBNReLU}(\boldsymbol{F}_2)\right]\right\}, \tag{14}$$

式中：$\boldsymbol{F}_1$为输入特征图；$\boldsymbol{F}_2$为通道注意力子模块输出特征图；$\boldsymbol{w}_1$、$\boldsymbol{w}_2$和$\boldsymbol{b}_1$、$\boldsymbol{b}_2$分别为多层感知机(MLP)[17]的初始权值和偏执项；M_{C}为通道注意力函数；$\boldsymbol{F}_3$为GAM注意力输出特征图；M_{S}为空间注意力函数。

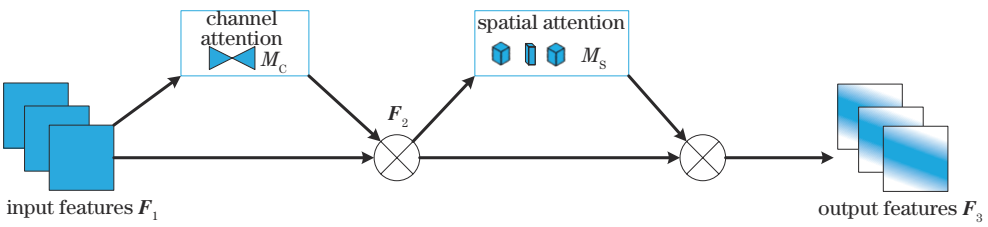

图3　GAM注意力机制结构

Fig. 3　Structure of GAM attention mechanism

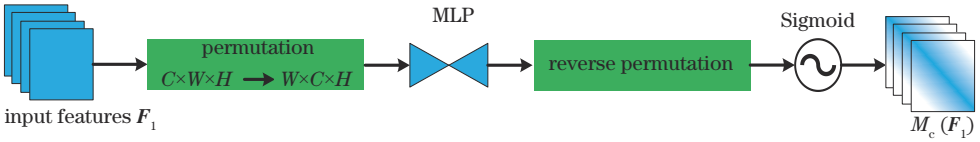

图 4　通道注意力子模块

Fig. 4　Channel attention submodule

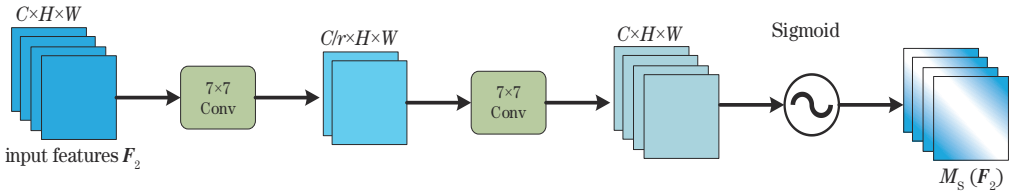

图 5　空间注意力子模块

Fig. 5　Spatial attention submodule

对于给定维度顺序为 $C\times W\times H$ 的输入特征图 $\boldsymbol{F}_1$，通道注意力模块首先对特征图变换维度排列顺序以保留通道信息，变换后的维度顺序为 $W\times H\times C$。然后通过两层多层感知机引入非线性能力，使模型学习自适应的权重和非线性变换来捕捉特征之间的复杂关系，增强维度通道之间的空间依赖性并在层间使用 ReLU 激活函数防止梯度弥散与梯度爆炸。最后将三维排列恢复为 $C\times W\times H$ 后与原输入特征图 $\boldsymbol{F}_1$ 相乘得到通道注意力输出特征图 $\boldsymbol{F}_2$。空间注意力模块接收特征图 $\boldsymbol{F}_2$ 后，使用两个卷积核大小为 7×7、填充为 3 的卷积进行空间维度语义信息的提取和融合，增强模型对空间信息的关注度，为了进一步保留空间信息，不再采用池化处理，直接通过 Sigmoid 激活函数后与注意力输出特征图 $\boldsymbol{F}_2$ 相乘得到最终输出特征图 $\boldsymbol{F}_3$。

考虑随着网络深度的增加，训练误差也随之增加，会产生网络退化等问题，在 GAM 注意力机制中融入残差结构[18]，设计残差注意力模块（ReGAM）以缓解这一问题。通过将输入特征图分别与通道注意力输出特征图和 GAM 注意力输入特征图融合，得到新的输出特征图。引入残差结构，使得模型能够提取不同层次的信息，增强特征融合能力，缓解网络退化问题，使模型更容易学习输入和输出之间的映射关系，并且防止了梯度消失与弥散。残差注意力模块可描述为

$$\boldsymbol{F}_4=\boldsymbol{F}_1+M_S(\boldsymbol{F}_2)\otimes(\boldsymbol{F}_1+\boldsymbol{F}_2),\tag{15}$$

式中：$\boldsymbol{F}_4$ 为残差注意力的输出特征图。残差注意力模块的结构如图 6 所示。

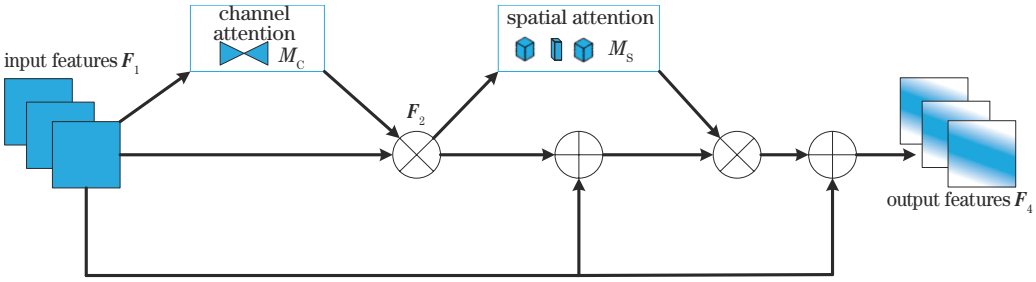

图 6　残差注意力模块

Fig. 6　Residual attention module

遥感图像背景复杂，且检测目标占图像面积小，导致模型迭代过程中冗余特征占比大。残差注意力机制的引入使得模型削弱了背景信息的权重，增强了包含特征信息量大的通道，使模型拟合重要的语义信息，更加聚焦目标区域，并且防止了梯度爆炸与弥散，保留了原始输入特征图中的有效信息，允许模型捕获更复杂的特征。

3.3　C2f-DCN

3.3.1　可变形卷积

传统的卷积操作中，卷积核具有固定像素点的位置，对输入图像的每个位置应用相同的卷积核。而在实际上，不同位置的图像可能具有不同的形变，常规卷积无法适应遥感图像的不规则布局与非刚性形变，因此可能导致提取的遥感目标特征不准确。

以 3×3 卷积为例，普通卷积对于每一个输出特征图 $\boldsymbol{Y}$，都要从输入特征图进行规则采样，再经过加权计算。其中，采样是指以中心位置向四周扩散得到 9 个点，所得到的网格定义为 $\boldsymbol{R}$，如式（16）所示：

$$\boldsymbol{R}=\{(-1,-1),(-1,0),\cdots,(0,1),(1,1)\}。\tag{16}$$

对于任意位置 $\boldsymbol{p}_0$，传统卷积输出特征图 $\boldsymbol{Y}$ 如式（17）所示：

$$\boldsymbol{Y}(\boldsymbol{p}_0)=\sum_{\boldsymbol{p}_n\in\boldsymbol{R}}W(\boldsymbol{p}_n)\cdot\boldsymbol{X}(\boldsymbol{p}_0+\boldsymbol{p}_n),\tag{17}$$

式中：$\boldsymbol{p}_n$表示枚举网格$\boldsymbol{R}$中的像素位置；$W(\cdot)$是加权计算函数；$\boldsymbol{X}(\cdot)$是采样后的输入特征图。

在 Deformable ConvNets(DCN)[19-20]中，卷积核每个像素点的位置通过一个偏移量确定，使得采样网格不再受限于规则分布，能够适应不同形状的物体。同样以3×3卷积为例，对规则采样网格$\boldsymbol{R}$增加偏移量进行扩充，如式(18)所示：

$$\boldsymbol{R}=\{\Delta\boldsymbol{p}_n|n=1,\cdots,n\}。\tag{18}$$

为了使得采样点的偏移精确覆盖检测目标，引入调制变量，增强可变形卷积操纵空间的能力，让模型学习采样点的偏移，学习每个采样点的权重，减轻了无关因素的干扰。则对任意位置$\boldsymbol{p}$，输出特征图$\boldsymbol{Y}$可表示为

$$\boldsymbol{Y}(\boldsymbol{p})=\sum_{k=1}^{K}w_k\cdot\boldsymbol{x}(\boldsymbol{p}+\boldsymbol{p}_k+\Delta\boldsymbol{p}_k)\cdot\Delta m_k,\tag{19}$$

式中：K表示卷积核的个数；w_k和$\boldsymbol{p}_k$分别表示第k个位置的权重和偏移量；$\Delta\boldsymbol{p}_k$和Δm_k分别表示第k个位置的可学习偏移量和调制标量。Δm_k对卷积在输入特征图上采样点的偏移量计算权重，去除无关的上下文信息，对于无关的采样点权重直接设置为0。

由于偏移量的引入，采样的9个位置不再规则，偏移量$\Delta\boldsymbol{p}_n$通常为分数，所以采用双线性插值实现，可由式(20)～(22)表示：

$$X(p)=\sum_{q}G(q,p)\cdot g(q_y,p_y),\tag{20}$$

$$G(q,p)=g(q_x,p_x)\cdot g(q_y,p_y),\tag{21}$$

$$g\left(q_y,p_y\right)=\max\left(0,1-|a-b|\right),\tag{22}$$

式中：p表示任意(分数)位置；q枚举特征图$\boldsymbol{X}$中的所有积分空间位置；$G(\cdot,\cdot)$是双线性插值核。

对于输入特征图，可变形卷积首先进行预处理，生成偏移量与调制量，规则分布的像素点通过不同的偏移方向得到不规则分布的像素点，再根据生成的新的像素点对输入图像进行采样，得到采样后的特征图。最后将采样后的特征图与卷积核进行逐元素相乘，并进行求和操作，得到最终卷积结果。可变形卷积过程如图7所示。

3.3.2　可变形感兴趣区域池化

池化层是卷积神经网络中常用的一种层级结构，用于降低输入数据的维度，从而减小模型参数和计算量，同时也能够提取输入数据中的关键特征。在传统的池化方法中，通常是对局部区域内的特征值进行统计汇总，从而得到该区域的特征表示。可变形感兴趣区域(RoI)池化[19-20]引入偏移量来增强特征表示的表达能力。RoI池化定义经过最后一次卷积后特征图上的一个矩形框为RoI。对于给定输入特征图，RoI池化层将RoI分为$k\times k$个直方图，并输出$k\times k$个输出特征图。定义第(i,j)个直方图为$\mathrm{bin}(i,j)(0\leqslant i,j\leqslant k)$，则

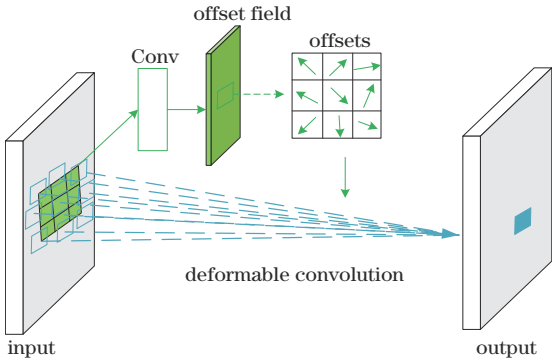

图7　可变形卷积过程

Fig. 7　Variable convolution process

对应的输出特征图$\boldsymbol{y}(i,j)$为

$$\boldsymbol{y}(i,j)=\sum_{\boldsymbol{p}\in\mathrm{bin}(i,j)}\boldsymbol{x}(\boldsymbol{p}_0+\boldsymbol{p})/n_{ij},\tag{23}$$

式中：$\boldsymbol{x}(\cdot)$为输入特征图；$\boldsymbol{p}_0$为感兴趣的左上角坐标；$\boldsymbol{p}$为枚举直方图$\mathrm{bin}(i,j)$中的像素点；n_{ij}为直方图$\mathrm{bin}(i,j)$的像素点个数。

与可变形卷积类似，在可变形RoI池化中添加偏移量$\Delta\boldsymbol{p}_{ij}$，产生新的像素点位置，则对应的输出特征图$\boldsymbol{y}(i,j)$为

$$\boldsymbol{y}(i,j)=\sum_{\boldsymbol{p}\in\mathrm{bin}(i,j)}\boldsymbol{x}(\boldsymbol{p}_0+\boldsymbol{p}+\Delta\boldsymbol{p}_{ij})/n_{ij}。\tag{24}$$

首先，对于输入特征图，通过可变形RoI池化生成对应的RoI。其次，在特征图上，针对每个RoI的位置，使用一个额外可变形卷积层预测RoI中各个位置的偏移量，并将其用于调整采样位置。再用RoI的坐标和预测的偏移量来划分RoI为一系列子区域。对于每个子区域，根据偏移量调整采样位置，从特征图中取得对应位置的特征值。然后对于每个子区域，对其采样得到的特征值进行自适应最大池化操作，得到固定大小的输出。最后将所有子区域的池化结果按照一定的顺序拼接在一起，形成最终的可变形RoI池化结果。可变形RoI池化过程如图8所示。

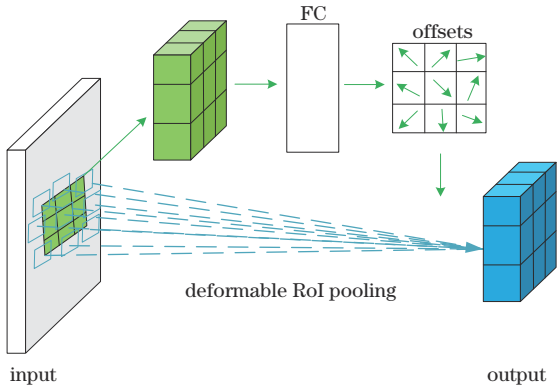

图8　可变形RoI池化过程

Fig. 8　Deformable RoI pooling process

3.3.3　C2f-DCN 结构

为使模型适应遥感图像中目标物体的形变和旋转，对 Backbone 模块中的第 4、6、8 层的 C2f 层进行改进，将 Bottleneck 结构中的常规卷积改为可变形卷积并融入可变形 RoI 池化层。改进后的 C2f-DCN 结构如图 9 所示。

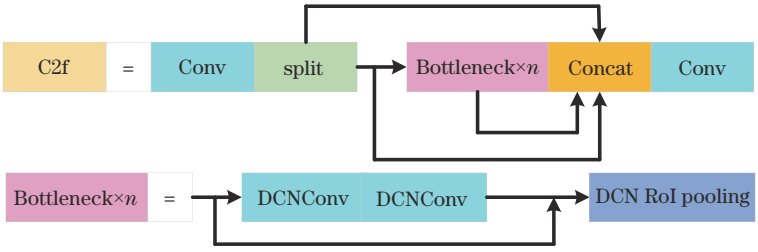

图 9　C2f-DCN 结构

Fig. 9　Structure of C2f-DCN

输入特征图经过两层可变形卷积，模型允许采样网格的自由变形，通过额外的卷积层从特征映射中学习到的偏移量，充分学习目标物体的尺度、姿态、视点和形变。经历卷积层后通过残差连接与输入特征图进行特征融合，最后通过可变形 RoI 池化，同样从前面的特征映射和 RoI 池化中学习偏移量，从而扩大模型感受野，提取更复杂的特征，实现不同形状物体自适应学习以适应遥感图像中不规则排列与布局的目标，进一步提高模型的鲁棒性和泛化能力。

4　实验数据与分析

4.1　实验环境

模型训练环境为 Windows 10 操作系统、Intel Core i5 12490F、32 GB 内存和 RTX 2060s(8 GB)。

训练参数设置如下：训练周期为 500，批处理量为 32，进程数为 8，输入图像尺寸为 640×640。模型使用 Adam 优化器对学习率进行优化，最大学习率设为 1×10^{-3}，最小学习率为 1×10^{-5}。模型使用权重衰减策略防止过拟合，权重衰减值设为 5×10^{-4}。模型训练使用 Early-Stopping，模型检测出验证损失趋于平稳则自动终止训练，模型基本收敛。

4.2　实验数据集

实验采用 DOTA 数据集[21]与 RSOD 数据集[22]。DOTA 数据集是用于航拍图像中目标检测的大型图像数据集，可用于发现和评估航拍图像中的物体。无论从数量还是质量上来说，DOTA 在同类型数据集中都具有优势。DOTA 数据集有 2806 张图片，188282 个实例，共 15 个类别。采用切割的方式将数据集切割为 21046 张分辨率为 1024×1024 的图片，并按 7∶2∶1 的比例分为训练集、验证集与测试集。

RSOD 数据集是一个开放的目标检测数据集，用于遥感图像中的目标检测。数据集包含飞机、油箱、运动场和立交桥等 4 个类别。数据集共计 976 张图，6950 个目标，按 7∶2∶1 的比例分为训练集、验证集与测试集。

4.3　实验评估指标

实验采用精确率(P)、召回率(R)和平均精度均值(mAP)作为实验评估指标。

精确率以预测结果为判断依据，是预测为正例的样本中预测正确的比例。预测为正例的结果包括正例(TP)和负例(FP)，通常把样本 IoU 大于等于置信度阈值分为正例，样本 IoU 小于置信度阈值分为负例。则精确率的表达式为

$$P=\frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}}+N_{\mathrm{FP}}}。\tag{25}$$

召回率以实际样本为判断依据，是被预测正确的正例占总实际正例样本的比例。实际为正例的样本中，预测正确定义为 TP，预测错误定义为 FN。则召回率的表达式为

$$R=\frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}}+N_{\mathrm{FN}}}。\tag{26}$$

以精确率作为纵轴，召回率为横轴，可以得到模型的性能 P-R 曲线。P-R 曲线下的面积为平均精度(AP)：

$$P_{\mathrm{AP}}=\int_0^1 P(t)\mathrm{d}t。\tag{27}$$

对于 N 个类别的目标检测，其平均精度均值为

$$P_{\mathrm{mAP}}=\frac{\sum_{n=1}^{N}P_{\mathrm{AP}\,n}}{N}。\tag{28}$$

目标检测实际应用中，网络模型不仅需要较高的准确性，其检测速度也要满足实时性。网络的检测速度由 FPS 来衡量，FPS 为模型每秒检测图片数量。

4.4　数据分析

为验证各个模块的性能，以 YOLOv8 原始模型为基准方法，采用 DOTA 数据集设计一系列消融实验进行对比验证，使用 P、R 和 mAP@0.5(0.5 为 IoU 阈值)为定量评价指标，实验结果如表 1 所示。

选取每组消融实验最优结果进行分析，表 1 中：对比方法 2 与方法 5 可知，模型引入残差注意力网络后，精确率基本保持不变，召回率提升约 0.7 百分点，平均

表 1 消融实验

Table 1 Ablation experiments

Method	Type	P /%	R /%	mAP@0.5 /%	FPS /(frame/s)
1	YOLOv8	75.5	66.1	69.8	**98**
2	YOLOv8+DWIoU	75.8	65.9	70.2	95
3	YOLOv8+ReGAM	74.9	66.7	70.3	92
4	YOLOv8+C2f-DCN	76.1	66.5	71.0	81
5	YOLOv8+DWIoU+ReGAM	75.8	66.6	70.9	89
6	YOLOv8+DWIoU+C2f-DCN	76.1	66.7	71.5	80
7	YOLOv8+ReGAM+C2f-DCN	75.9	66.6	71.3	78
8	YOLOV8+DWIoU+ReGAM+C2f-DCN	**76.2**	**66.8**	**72.1**	77

精度均值提升约0.7百分点，FPS降低约6 frame/s；对比方法2与方法6可知，模型引入C2f-DCN模块后，精确率提升约0.3百分点，召回率提升约0.8百分点，平均精度均值提升约1.3百分点，FPS约下降15 frame/s；对比方法3与方法5可知，模型引入DWIoU后，精确率提升约0.9百分点，召回率基本不变，平均精度均值提升约0.6百分点，FPS约下降3 frame/s。由数据可知，所提出模块对网络检测性能均有所提升，但FPS有所下降，原因为加入改进模块后，网络模型复杂度上升，计算量增大，一定程度影响了网络检测速度。

表1实验结果表明，所提改进算法优于消融实验其他模型，说明了所提算法用于遥感图像目标检测的可行性与有效性。

为验证改进YOLOv8算法的检测性能，使用FPS和mAP@0.5为定量指标，将其与主流算法Faster R-CNN、SSD、YOLOv3、YOLOv4、YOLOv5、YOLOv8在DOTA数据集上的目标检测结果进行定量分析，结果如表2所示。

表 2 不同网络模型在DOTA数据集上的性能对比

Table 2 Performance comparison of different network models on DOTA dataset

Method	FPS /(frame/s)	mAP@0.5 /%
Faster R-CNN	8	42
SSD	38	59.6
YOLOv3	18	68.2
YOLOv4	17	68.4
YOLOv5	38	69.4
YOLOv8	**98**	69.8
Improved YOLOv8	77	**72.1**

分析表2可得，改进YOLOv8相较于Faster R-CNN、SSD算法在PFS与平均精度均值上均有明显提升，原因在于R-CNN算法相对其他算法在特征提取网络中没有对高层特征进行多尺度融合，对不同尺度的目标检测鲁棒性较差。SSD算法采用多个先验框来检测不同大小的目标，但是当物体的长宽比不一致时，检测效果较差。改进YOLOv8相较于YOLOv3、YOLOV4、YOLOv5、YOLOv8检测效果有一定提升，在平均精度均值上分别提升3.9百分点、3.7百分点、2.7百分点和2.3百分点。原因在于YOLOv3使用的DarkNet架构相对较旧，与最新的深度学习架构相比，其可扩展性和灵活性较低。YOLOv4采用空间金字塔池化(SPP)来加强目标特征提取，一定程度上提高了精确率，但面对小目标的漏检率较高。由于YOLOV5在训练过程中使用Anchor-Free的方式，因此在小目标检测时需要进一步优化。YOLOv8采用了梯度流更加丰富的结构，性能明显改善，FPS达到98 frame/s，相较其他模型有显著优势，但缺乏对形变目标的建模与泛化能力。改进YOLOv8算法的平均精度均值达到85.3%，优于其他算法，但FPS较YOLOv8有明显降低，原因在于引入了更复杂的卷积方式与主干结构，但综合性能优于对比算法。

为进一步验证改进YOLOv8算法的鲁棒性，使用FPS和mAP@0.5为定量指标，将其与主流算法Faster R-CNN、SSD、YOLOv3、YOLOv4、YOLOv5、YOLOv8在RSOD数据集上进一步进行对比实验，其结果如表3所示

表 3 不同网络模型在RSOD数据集上的性能对比

Table 3 Performance comparison of different network models on RSOD dataset

Method	FPS /(frame/s)	mAP@0.5 /%
Faster R-CNN	12	90.3
SSD	46	89.2
YOLOv3	33	84.7
YOLOv4	57	87.8
YOLOv5	62	91.8
YOLOv8	**137**	93.4
Improved YOLOv8	123	**94.6**

由表3可知：在RSOD数据集上，YOLOv3算法平均精度均值表现最差，为83.7%；Faster R-CNN算法平均精度均值达到90.3%，但FPS表现最差，为12 frame/s；YOLOv8算法FPS表现最优，达到137 frame/s，但平均精度均值低于改进YOLOv8算法；

改进YOLOv8算法的FPS略低于YOLOv8算法，但平均精度均值达到94.6%，优于其他主流算法，表明改进YOLOv8算法在遥感图像目标检测中具有鲁棒性。

此外，将改进YOLOv8算法与YOLOv3、YOLOv5、YOLOv8算法在DOTA数据集的部分图像检测效果进行直观对比，结果如图10所示。由图10(d)～(f)可知，YOLOv3对于小目标与排列密集目标漏检率高，原因在于YOLOv3的检测网络只在3个不同尺度的特征图上预测目标，导致无法精确地处理图像语义信息。由图10(g)～(i)可知，YOLOv5的置信度明显上升，但

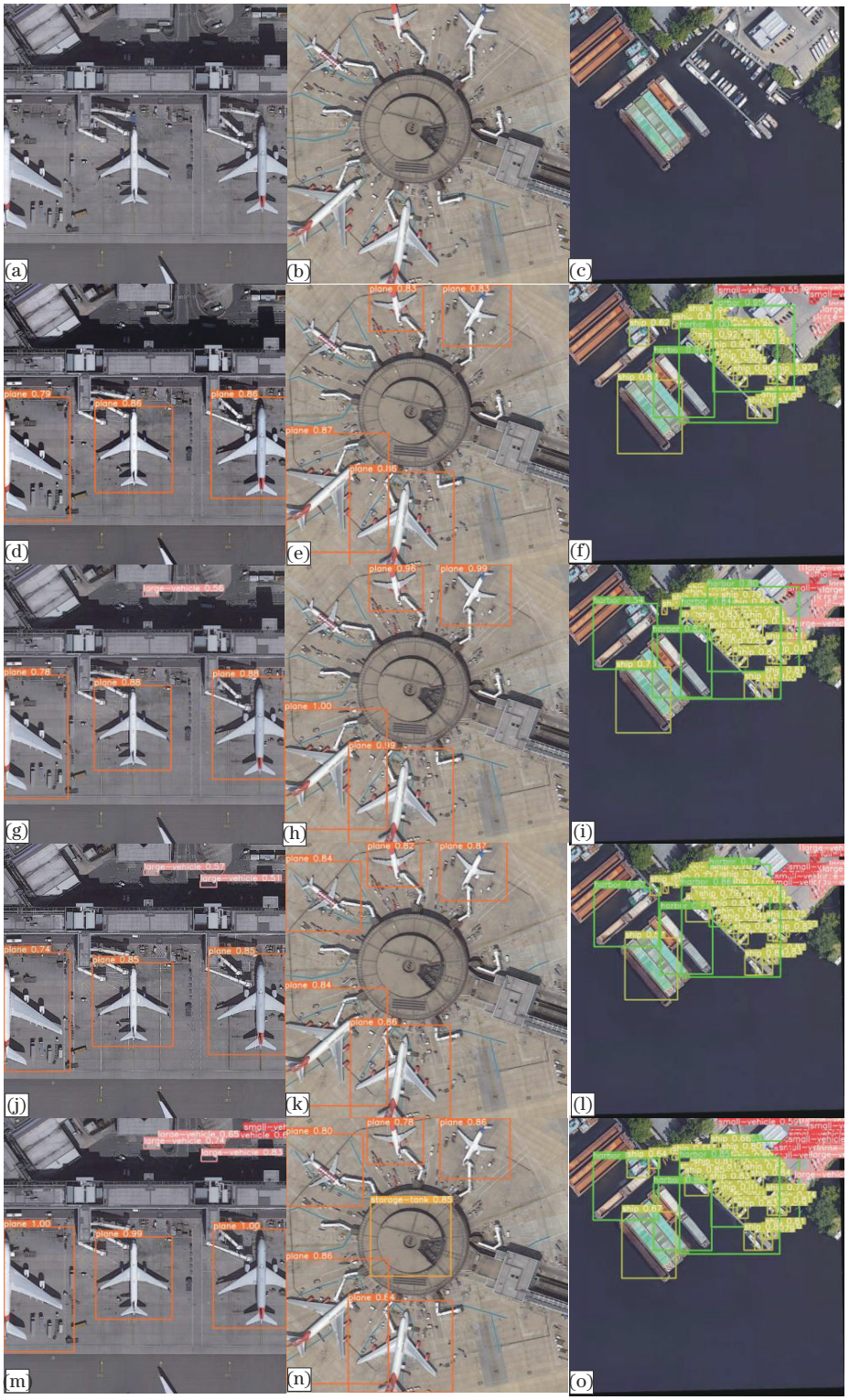

图10 不同模型检测效果。(a)～(c)原始图像；(d)～(f) YOLOv3检测效果；(g)～(i) YOLOv5检测效果；(j)～(i) YOLOv8检测效果；(m)～(o)改进YOLOv8检测效果

Fig. 10 Detection effects of different models. (a)–(c) Original images; (d)–(f) detection effects of improved YOLOv3; (g)–(i) detection effects of YOLOv5; (j)–(i) detection effects of YOLOv8; (m)–(o) detection effects of improved YOLOv8

漏检率仍然较高，原因在于尽管YOLOv5采用了多尺度检测头与自适应训练策略，但在处理极端情况下的目标（如极小目标或极大目标）时存在一定的限制。由图10(j)～(l)可知，YOLOv8算法对小目标与排列密集目标的漏检率明显降低，原因为YOLOv8采用了梯度流更加丰富的C2f模块代替C3模块，并且采用解耦头代替耦合头，对小目标的检测能力明显改善。由图10(m)～(o)可知，改进YOLOv8算法相较于其他算法总体置信度有所提升，漏检率明显降低，尤其体现在小目标与不规则目标，总体性能优于对比检测算法。

5 结　论

为提高遥感图像目标检测算法性能，提出改进YOLOv8算法。为使模型更好地适应各种形状和尺寸的物体，降低对遥感图像中特殊尺寸目标的漏检率，通过考虑静态聚焦机制与目标的几何因素，设计DWIoU边界框损失函数；为使模型选择性增强包含目标信息量最大的特征通道，将GAM注意力机制与残差结构相结合，设计残差注意力模块，有效减少遥感图像背景对检测目标的影响；采用可变形卷积替代部分常规卷积，并融入可变形感兴趣区域池化层，设计C2f-DCN模块，以适应遥感图像中目标物体的形变和旋转，缓解遥感图像中目标排列不规则、部分目标被遮挡等问题。在DOTA数据集和RSOD数据集上，改进YOLOv8算法的平均精度均值相对于YOLOv8算法分别提升2.3百分点和1.2百分点。但在实际测试中，模型对遮挡严重的小目标检测效果仍有待提升，且FPS较原模型有所降低。未来主要工作仍聚焦于进一步优化网络模型以提高对遥感图像的目标检测性能。

参 考 文 献

[1] Zou Z X, Chen K Y, Shi Z W, et al. Object detection in 20 years: a survey[EB/OL]. (2019-05-13)[2021-05-06]. https://arxiv.org/abs/1905.05055.

[2] 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 (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.

[3] 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.

[4] Girshick R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2016: 1440-1448.

[5] 刘涛, 丁雪妍, 张冰冰, 等. 改进YOLOv5的遥感图像检测方法[J]. 计算机工程与应用, 2023, 59(10): 253-261.
Liu T, Ding X Y, Zhang B B, et al. Improved YOLOv5 for remote sensing image detection[J]. Computer Engineering and Applications, 2023, 59(10): 253-261.

[6] 张正, 白佳华, 田青. 基于单级特征金字塔的图像旋转目标检测[J]. 计算机工程与应用, 2023, 59(15): 235-242.
Zhang Z, Bai J H, Tian Q. Image rotating objects detection based on single level feature pyramid[J]. Computer Engineering and Applications, 2023, 59(15): 235-242.

[7] 原瑜蔓, 白宏阳, 郭宏伟, 等. HourglassNet: 一种用于遥感目标检测的改进FCOS算法[J]. 南京理工大学学报, 2022, 46(6): 719-727, 741.
Yuan Y M, Bai H Y, Guo H W, et al. HourglassNet: an improved FCOS algorithm for remote sensing target detection[J]. Journal of Nanjing University of Science and Technology, 2022, 46(6): 719-727, 741.

[8] Redmon J, Farhadi A. YOLOV3: an incremental improvement[EB/OL]. (2018-04-08)[2021-05-06]. https://arxiv.org/abs/1804.02767.

[9] Lou H T, Duan X H, Guo J M, et al. DC-YOLOv8: small-size object detection algorithm based on camera sensor[J]. Electronics, 2023, 12(10): 2323.

[10] Neubeck A, Van Gool L. Efficient non-maximum suppression[C]//18th International Conference on Pattern Recognition (ICPR'06), August 20-24, 2006, Hong Kong, China. New York: IEEE Press, 2006: 850-855.

[11] Zheng Z H, Wang P, Liu W, et al. Distance-IoU loss: faster and better learning for bounding box regression[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.

[12] Rezatofighi H, Tsoi N, Gwak J, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2020: 658-666.

[13] Tong Z J, Chen Y H, Xu Z W, et al. Wise-IoU: bounding box regression loss with dynamic focusing mechanism[EB/OL]. (2023-01-24)[2023-05-06]. https://arxiv.org/abs/2301.10051.

[14] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 7132-7141.

[15] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 3-19.

[16] Liu Y C, Shao Z R, Hoffmann N. Global attention mechanism: retain information to enhance channel-spatial interactions[EB/OL]. (2021-12-10)[2022-05-06]. https://arxiv.org/abs/2112.05561.

[17] Tolstikhin I, Houlsby N, Kolesnikov A, et al. MLP-mixer: an all-MLP architecture for vision[EB/OL]. (2021-05-04)[2021-05-06]. https://arxiv.org/abs/2105.01601.

[18] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),

June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.

[19] Dai J F, Qi H Z, Xiong Y W, et al. Deformable convolutional networks[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 764-773.

[20] Zhu X Z, Hu H, Lin S, et al. Deformable ConvNets V2: more deformable, better results[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 9300-9308.

[21] Xia G S, Bai X, Ding J, et al. DOTA: a large-scale dataset for object detection in aerial images[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 3974-3983.

[22] Xiao Z F, Liu Q, Tang G F, et al. Elliptic Fourier transformation-based histograms of oriented gradients for rotationally invariant object detection in remote-sensing images[J]. International Journal of Remote Sensing, 2015, 36(2): 618-644.