E-mail: jig@aircas.ac.cn
Website: www.cjig.cn
Tel: 010-58887035
中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS

中图法分类号:TP391　　**文献标识码**: A　　**文章编号**: 1006-8961(2025)03-0710-14

论文引用格式:Xu H P, Liu G, Xi J T and Tong J. 2025. Infrared aircraft detection algorithm based on adaptive weighted fusion of global-local context. Journal of Image and Graphics, 30(3):0710-0723(徐红鹏, 刘刚, 习江涛, 童军. 2025. 基于全局—局部上下文自适应加权融合的红外飞机检测算法. 中国图象图形学报, 30(3):0710-0723)[DOI:10.11834/jig.240271]

基于全局—局部上下文自适应加权融合的红外飞机检测算法

徐红鹏[1],刘刚[1*],习江涛[2],童军[2]

1. 河南科技大学信息工程学院,洛阳　471023; 2. 伍伦贡大学电气计算机与通信工程学院,澳大利亚　伍伦贡　NSW　2522

摘　要:**目的**　针对远距离红外飞机目标检测中存在的由于成像面积小、辐射强度较弱造成无法充分提取目标特征进而影响检测性能的问题,提出一种基于全局—局部上下文自适应加权融合(adaptive weighted fusion of global-local context,AWFGLC)机制的红外飞机目标检测算法。**方法**　基于全局—局部上下文自适应加权融合机制,沿着通道维度随机进行划分与重组,将输入特征图切分为两个特征图。一个特征图使用自注意力进行全局上下文建模,建立目标特征与背景特征之间的相关性,突出目标较显著的特征,使得检测算法更好地感知目标的全局特征。对另一特征图进行窗口划分并在每个窗口内进行最大池化和平均池化以突出目标局部特征,随后使用自注意力对池化特征图进行局部上下文建模,建立目标与其周围邻域的相关性,进一步增强目标特征较弱的部分,使得检测算法更好地感知目标的局部特征。根据目标特点,利用可学习参数的自适应加权融合策略将全局上下文和局部上下文特征图进行聚合,得到包含较完整目标信息的特征图,增强检测算法对目标与背景的判别能力。**结果**　将全局—局部上下文自适应加权融合机制引入YOLOv7(you only look once version 7)并对红外飞机目标进行检测,实验结果表明,提出算法在自制和公开红外飞机数据集的mAP50(mean average precision 50)分别达到97.8%、88.7%,mAP50:95分别达到65.7%、61.2%。**结论**　本文所提出的红外飞机检测算法,优于经典的目标检测算法,能够有效实现红外飞机目标检测。

关键词:红外飞机;目标检测;全局上下文;局部上下文;自适应加权

Infrared aircraft detection algorithm based on adaptive weighted fusion of global-local context

Xu Hongpeng[1], Liu Gang[1*], Xi Jiangtao[2], Tong Jun[2]

1. College of Information Engineering, Henan University of Science and Technology, Luoyang　471023, China; 2. School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, Wollongong, New South Walse　2522, Australia

Abstract: **Objective** An infrared aircraft target detection algorithm based on adaptive weighted fusion of global-local context (AWFGLC) is proposed to address the challenge of insufficient target feature extraction due to the small imaging area and weak radiation intensity in long-range infrared aircraft target detection. The global context focuses on the overall distribution of the target, providing the detection algorithm with global information regarding targets in images with strong radia-

收稿日期:2024-06-03;**修回日期**:2024-08-28;**预印本日期**:2024-09-04

*通信作者:刘刚　lg19741011@163.com

基金项目:国家留学基金资助([2022]20);河南省高等学校重点科研项目(21A520012)

Supported by: National Scholarship Fund for Overseas Study([2022]20); Key Research Projects of Higher Education Institutions in Henan Province(21A520012)

tion intensity and clear contours. In contrast, the local context emphasizes the local details of the target and the surrounding background information, offering the detection algorithm local information regarding targets with weak radiation intensity and blurred contours. Therefore, the global context and local context should be combined in accordance with the target characteristics in practical applications. **Method** Based on the global-local context adaptive weighted fusion mechanism, the input feature map is randomly divided and reorganized along the channel dimensions, resulting in two separate feature maps. Compared with global and local context modeling for input feature maps based on a specific arrangement or simply dividing the input feature map into two feature maps, the arrangement of the input feature map during iterative training can be diversely changed by randomly reorganizing it with different random numbers in each training round. Global and local context modeling can help the detection algorithm learn additional diversified and comprehensive features by combining feature maps with different arrangements. The global and local context modeling of different combinations of feature maps can enable learning of additional diverse and comprehensive features through the detection algorithm. Moreover, the complexity of contextual modeling is reduced by half by dividing the input feature maps equally in the channel dimension and performing global and local contextual modeling to reduce the complexity of contextual modeling of input feature maps. A feature map is subjected to global context modeling using self-attention to realize the following: establish the dependencies between each pixel in the feature map and other pixels, demonstrate the correlation between the target and background features, emphasize the more salient features of the target, and enable the detection algorithm to perceive the global features of the target effectively. The other feature map is divided into windows, and maximum and average pooling are performed within each window to highlight the local features of the target. The pooled feature map is then subjected to local contextual modeling using self-attention to establish the correlation between the target and its surrounding neighborhoods, which further enhances the weaker parts of the target features and allows the detection algorithm to perceive the local features of the target effectively. According to the target characteristics, the global and local context feature maps are adaptively weighted and channel spliced using an adaptive weighted fusion strategy of learnable parameters, updating the learnable parameters along with the optimizer under the guidance of loss function reduction of the target detection algorithm. Subsequently, feature maps containing more complete target information are obtained using convolution, batch normalization, and activation functions to enhance the detection the capability of the algorithm to discriminate between target and background. The mechanism of global-local context-adaptive weighted fusion is incorporated into the YOLOv7 feature extraction network, and the context modeling is performed on the downsampled 4-fold feature maps and the downsampled 32-fold feature maps to maximize the physical information in the shallow feature maps and the semantic information in the deeper feature maps, improving the capability of the model to extract infrared aircraft target features. **Result** Experimental results show that the proposed AWFGLC mechanism outperforms the contextual mechanisms such as global context, position attention, and local Transformer in terms of detection accuracy under the condition of increasing the number of parameters and computation in the homemade infrared aircraft dataset. The proposed AWFGLC mechanism tends to learn the global features of the target. Compared with the YOLOv7 detection algorithm, which has the second-best performance, the proposed AWFGLC-YOLO algorithm improves the mAP50 and mAP50:95 by 1.4% and 4.4%, respectively. In the publicly available dataset of weak aircraft target detection and tracking in infrared images in the ground/air context, the proposed AWFGLC mechanism learns the local features of the target. Compared to the YOLOv8 detection algorithm, which exhibits the second-best performance, the proposed AWFGLC-YOLO algorithm improves mAP50 and mAP50:95 by 3.0% and 4.0%, respectively. **Conclusion** The infrared aircraft detection algorithm proposed in this paper is superior to classical target detection algorithms, effectively achieving infrared aircraft target detection.

Key words: infrared aircraft; target detection; global context; local context; adaptive weighted

0 引　言

红外成像制导空空导弹是近距空战的主战武器，是决定近距空战胜负的关键。红外空空导弹的攻击目标主要是战斗机、直升机和无人机等，当红外成像制导空空导弹距离目标较远时，目标在成像平面内面积小、辐射特征弱，使得现有的目标检测算法

无法充分地提取红外飞机目标特征，导致红外飞机目标检测性能较差。目前，红外飞机目标检测主要依赖对目标特征精确建模的传统方法，但精确建模依赖人工先验知识，作战环境一旦超出预设的规则，则传统方法将无能为力。深度学习网络模拟人类大脑的神经连接结构，具备强大的目标特征自主学习和表示能力，是解决红外目标检测问题的有效途径。深度学习目标检测方法主要分为基于候选区域的双阶段检测算法如R-CNN（region-based convolutional neural network）系列及基于回归的单阶段检测算法，如YOLO（you only look once）系列、SSD（single shot multibox detector）系列等。相比双阶段算法，单阶段检测算法检测速度更快，能满足检测实时性的要求（程旭 等，2021）。

YOLO系列算法（Redmon和Farhadi，2018；Wang等，2021a；Wang等，2023；Ge等，2021）直接对目标的坐标和类别进行回归，这种端到端的检测方式使得检测精度高且检测速度快。相比其他YOLO系列检测算法，YOLOv5和YOLOv7在检测效率与精度之间取得较好的平衡，因此许多研究人员使用YOLOv5和YOLOv7并与具体任务相结合来解决实际问题（肖锋 等，2024；薛珊 等，2024）。Zhou等人（2022）以超分辨率重构图像作为检测算法的输入，并在YOLOv5的特征融合网络中结合自适应挤压激励模块、空间金字塔池化模块和多级感受野结构，使得检测算法能够保留红外飞机原始信息并提高目标特征利用效率。Wang等人（2019）在YOLOv5的特征提取网络中使用3次下采样操作来调整特征图的大小，并利用密集连接将YOLOv5各层的输出进行矩阵相加，更好地整合浅层网络和深层网络中红外飞机的位置信息，提高网络的定位能力。周葳楠等人（2024）在YOLOv5的特征提取网络嵌入挤压激励模块和Inception结构，提高检测算法在复杂背景下对目标特征的提取能力。同时，在特征融合网络中采用新的特征融合模块来融合不同层级，从而提高检测算法对目标特征的利用率。Zhao等人（2023）使用MobileViT（mobile vision Transformer）网络作为YOLOv7的特征提取网络来降低检测算法复杂度，并在特征融合网络中使用CARAFE（content-aware reassembly of features）作为上采样方法来更多地保留目标特征信息。Hu等人（2023）在YOLOv7中根据真值框的长宽比分配锚框，为检测算法提供目标形状的先验信息。同时，使用难样本挖掘损失函数来指导网络增强对硬样本的学习，增强检测算法对目标的分类能力。张光华等人（2024）将YOLOv7的激活函数LeakyReLU（leaky rectified linear unit）替换为SiLU（sigmoid liner unit），从而提升模型训练时的收敛速度与模型泛化性。此外，其设计了小目标检测层，进而有效地检测小目标。

上述方法从超分辨与感受野、密集连接、注意力与特征融合、轻量化与上采样、锚框先验与分类损失以及激活函数的角度改进检测算法，并没有利用目标的全局上下文信息。然而，在目标检测任务中融入上下文信息有利于区分目标与背景，从而提升检测算法对目标的检测效果。谭建豪等人（2021）引入全局上下文模块对输入特征图进行全局上下文建模，从而使得检测算法能够更好地区分目标与背景。Hou等人（2022）提出全局上下文信息聚合模块来提高检测算法的全局感知能力，增强检测算法对在夜间和遮挡下红外目标与背景特征的感知能力。王振学等人（2023）利用多尺度下采样的方式获得其对应的多尺度特征并选择性地集成多级特征的空间上下文信息，提升模型对目标空间位置的感知能力。Zhang等人（2021）利用不同大小的特征图来聚合上下文信息，从而增强背景和前景的空间关联性。以上研究利用全局上下文机制挖掘目标与场景中其他对象之间的关系，使得模型能够学习目标的上下文信息，增强检测算法对目标特征与背景特征的判别能力。然而，全局上下文机制大多是基于图像全局信息进行建模，没有突出目标局部信息，对目标及其周围邻域的建模能力较弱。基于局部上下文建模（Yang等，2024；Jamali等，2023）可以建立目标与其周围邻域的相关性，增强目标特征较弱的部分，使得检测算法更专注于目标周围的局部信息，从而增强对目标局部特征的提取能力。在实际应用中应将全局和局部上下文结合起来并根据目标特点有所侧重。

基于上述分析，本文提出一种基于全局—局部上下文自适应加权融合（adaptive weighted fusion of global-local context，AWFGLC）机制的YOLOv7红外飞机目标检测算法。本文主要贡献如下：1）针对红外飞机目标特点，提出了一种全局—局部上下文自适应加权融合机制，为检测算法提供较为完整的目标特征信息；2）将全局—局部上下文自适应加权融

合机制融入 YOLOv7 特征提取网络不同层级，充分利用浅层特征图中的物理信息和深层特征图中的语义信息，提出 AWFGLC-YOLO 算法，并应用于红外飞机目标检测。

1　全局—局部上下文自适应加权融合机制

全局—局部上下文自适应加权融合能够突出目标特征，提高检测算法对目标特征的感知能力，并增强对目标与背景的判别能力。全局上下文和局部上下文在融合特征图中的占比根据目标的不同特性存在差异，目标特征较强时，融合特征图应更侧重全局上下文信息，反之则应偏重于局部上下文。提出的机制将输入特征图在通道维度进行随机重组与划分得到两个特征图。一个特征图进行全局上下文建模；另一个特征图进行窗口划分并进行池化，随后在池化窗口内部进行局部上下文建模并通过空间重组得到局部上下文特征图。使用可学习参数的自适应加权策略聚合全局上下文信息和局部上下文信息，丰富目标的特征表达，提高检测算法特征提取网络对红外飞机目标的特征提取能力。全局—局部上下文自适应加权融合机制的整体流程如图 1 所示。

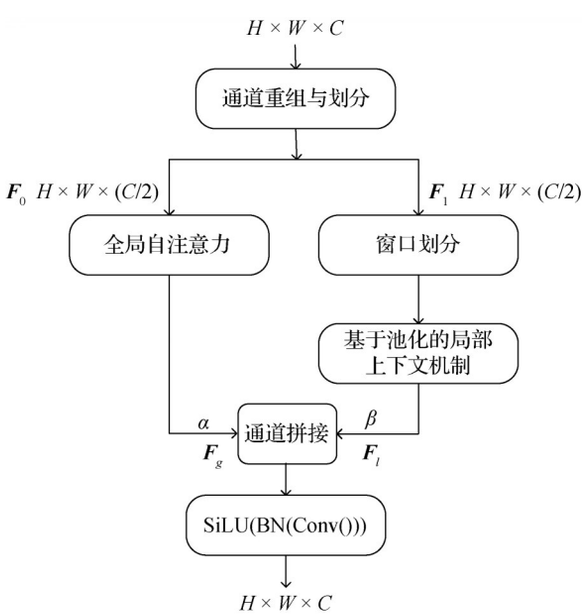


图 1　全局—局部上下文自适应加权融合机制

Fig. 1　Adaptive weighted fusion mechanism of global-local context

流程包括通道重组与划分、窗口划分、全局上下文建模、局部上下文建模和加权融合 5 个功能单元，具体如下：

1）相比对基于特定排列方式的输入特征图或简单地将输入特征图划分为两个特征图进行全局上下文建模和局部上下文建模，在迭代训练时，每个训练轮次通过设定不同的随机数进行随机重组，可以多样性地改变输入特征图的排列方式，对不同排列组合的特征图进行全局上下文建模和局部上下文建模可以帮助检测算法学习到更多样化、更全面的特征。此外，为减小对输入特征图上下文建模的复杂度，对输入特征图在通道维度进行均分并分别进行全局上下文建模和局部上下文建模。假设输入特征图为 $\boldsymbol{F} \in \mathbf{R}^{H\times W\times C}$，先将 $\boldsymbol{F}$ 在通道维度进行随机重组得到中间特征图 $\boldsymbol{F}' \in \mathbf{R}^{H\times W\times C}$。随后，在通道维度对 $\boldsymbol{F}'$ 进行划分生成 $\boldsymbol{F}_0 \in \mathbf{R}^{H\times W\times (C/2)}$ 和 $\boldsymbol{F}_1 \in \mathbf{R}^{H\times W\times (C/2)}$，如图 2 所示。

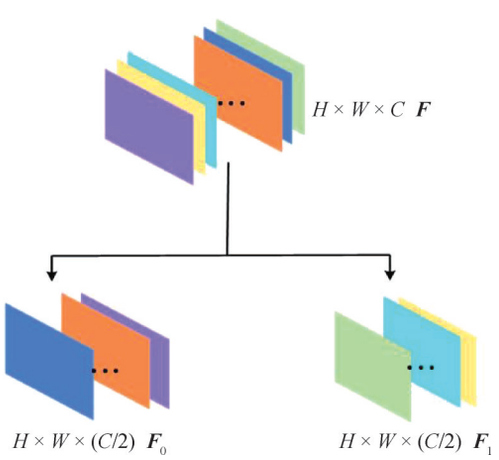


图 2　通道重组与划分

Fig. 2　Channel reorganisation and segmentation

随机重组是取出特征图 $\boldsymbol{F}$ 中每个通道的位置索引，将位置索引打乱并按照打乱后的位置索引进行通道拼接；划分是将通道拼接特征图在通道维度进行切片，得到两个相同维度的特征图。具体为

$$\boldsymbol{F}'[\boldsymbol{x}_3,\boldsymbol{x}_1,\cdots,\boldsymbol{x}_c,\cdots,\boldsymbol{x}_i]=f_r^c(\boldsymbol{F}[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_i,\cdots,\boldsymbol{x}_c]) \tag{1}$$

$$\boldsymbol{F}_0,\boldsymbol{F}_1=f_d^2(\boldsymbol{F}'[\boldsymbol{x}_3,\boldsymbol{x}_1,\cdots,\boldsymbol{x}_c,\cdots,\boldsymbol{x}_i]) \tag{2}$$

式中，f_r^c 表示对输入特征图在通道维度上进行重组；f_d^2 表示对中间特征图进行划分。$\boldsymbol{x}_i$ 表示输入特征图的第 i 个通道。

2）使用自注意力建立特征图 $\boldsymbol{F}_0$ 中每个像素点

与其他像素点之间的依赖关系，获得全局上下文特征图 $\boldsymbol{F}_g$，使用自注意力对 $\boldsymbol{F}_0$ 建模流程如图3所示。具体而言，将 $\boldsymbol{F}_0$ 的维度变换为 $(C/2)\times(H\times W)$ 并通过线性变换映射到 $\boldsymbol{F}_0^q$、$\boldsymbol{F}_0^k$、$\boldsymbol{F}_0^v$。为避免softmax函数的输入绝对值过大，导致反向传播时函数的梯度过小，从而影响线性变换中参数的更新，将 $\boldsymbol{F}_0^q$ 除以 $\sqrt{d}$ 进行缩放，d 是输入向量维度。将缩放后的 $\boldsymbol{F}_0^q$ 与 $\boldsymbol{F}_0^k$ 的转置 $\boldsymbol{F}_0^{k^{\mathrm{T}}}$ 相乘，然后进行softmax操作，结果与 $\boldsymbol{F}_0^v$ 相乘并将维度转为与 $\boldsymbol{F}_0$ 相同的形式，得到自注意力结果。整个计算过程为

$$\boldsymbol{F}_0^q,\boldsymbol{F}_0^k,\boldsymbol{F}_0^v=f_{\mathrm{Linear}}^q(\boldsymbol{F}_0),f_{\mathrm{Linear}}^k(\boldsymbol{F}_0),f_{\mathrm{Linear}}^v(\boldsymbol{F}_0) \tag{3}$$

$$\boldsymbol{F}_g=\mathrm{softmax}\left(\frac{\boldsymbol{F}_0^q\boldsymbol{F}_0^{k^{\mathrm{T}}}}{\sqrt{d}}\right)\boldsymbol{F}_0^v \tag{4}$$

式中，f_{Linear} 表示线性变换；softmax表示激活函数。

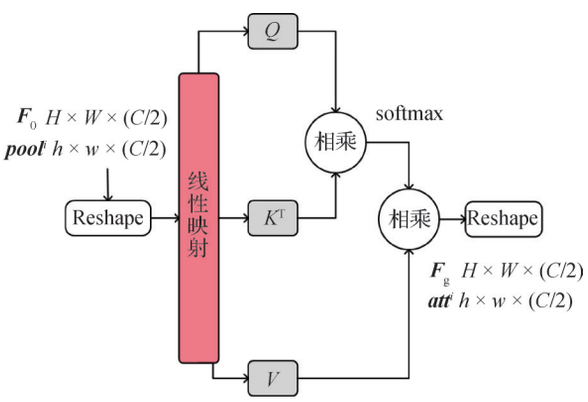


图3　全局(局部)自注意力建模

Fig. 3　Global (local) self-attention modelling

3)为增强目标局部特征，以大小、步长均为 $h\times w\times(C/2)$ 的窗口在 $\boldsymbol{F}_1$ 滑动，将其划分为 i 个维度为 $h\times w\times(C/2)$ 的 $\boldsymbol{patch}$，如图4所示。在每个 $\boldsymbol{patch}$ 中进行最大池化和平均池化来突出目标区域的显著特征和平均特征，并得到最大池化特征图 $\boldsymbol{pool}_{\max}^i$、平均池化特征图 $\boldsymbol{pool}_{\mathrm{avg}}^i$。将 $\boldsymbol{pool}_{\max}^i$ 与 $\boldsymbol{pool}_{\mathrm{avg}}^i$ 进行矩阵相加得到池化特征图 $\boldsymbol{pool}^i$。计算过程为

$$\boldsymbol{pool}_{\max}^i=f_{\max}^{k\times k\times s}(\boldsymbol{patch}^i) \tag{5}$$

$$\boldsymbol{pool}_{\mathrm{avg}}^i=f_{\mathrm{avg}}^{k\times k\times s}(\boldsymbol{patch}^i) \tag{6}$$

$$\boldsymbol{pool}^i=\boldsymbol{pool}_{\max}^i\oplus\boldsymbol{pool}_{\mathrm{avg}}^i \tag{7}$$

式中，$f_{\mathrm{avg}}^{k\times k\times s}$ 表示使用大小、步长分别为 $k\times k$、s 的卷积进行平均池化；$f_{\max}^{k\times k\times s}$ 表示使用大小、步长均为 $k\times k$、s 的卷积进行最大池化；$\oplus$ 表示矩阵相加；$\boldsymbol{pool}^i$ 表示第 i 个局部池化特征图。

在第 i 个局部池化特征图 $\boldsymbol{pool}^i$ 上进行自注意力

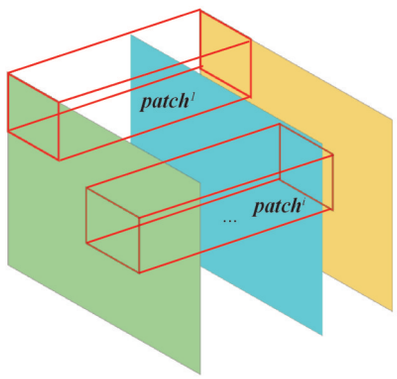


图4　窗口划分

Fig. 4　Window segmentation

建模来调整特征图中每个像素点与其他像素点之间的关联性，得到基于池化的局部上下文特征图 $\boldsymbol{att}^i$，从而将特征学习的焦点更多地放在目标及其邻域。将 $\boldsymbol{att}^i$ 重组获得局部上下文特征图 $\boldsymbol{F}_l\in\mathbf{R}^{H\times W\times(C/2)}$，如图5所示。每个窗口的自注意力建模方法与 $\boldsymbol{F}_0$ 相同，具体为

$$\boldsymbol{pool}_i^q,\boldsymbol{pool}_i^k,\boldsymbol{pool}_i^v=f_{\mathrm{Linear}}^q(\boldsymbol{pool}^i),f_{\mathrm{Linear}}^k(\boldsymbol{pool}^i),f_{\mathrm{Linear}}^v(\boldsymbol{pool}^i) \tag{8}$$

$$\boldsymbol{att}^i=\mathrm{softmax}\left(\frac{\boldsymbol{pool}_i^q\boldsymbol{pool}_i^{k^{\mathrm{T}}}}{\sqrt{d}}\right)\boldsymbol{pool}_i^v \tag{9}$$

$$\boldsymbol{F}_l=f_r^s([\boldsymbol{att}^1,\boldsymbol{att}^2,\cdots,\boldsymbol{att}^i]) \tag{10}$$

式中，f_r^s 表示对 $\boldsymbol{att}^i$ 在空间维度上进行重组。

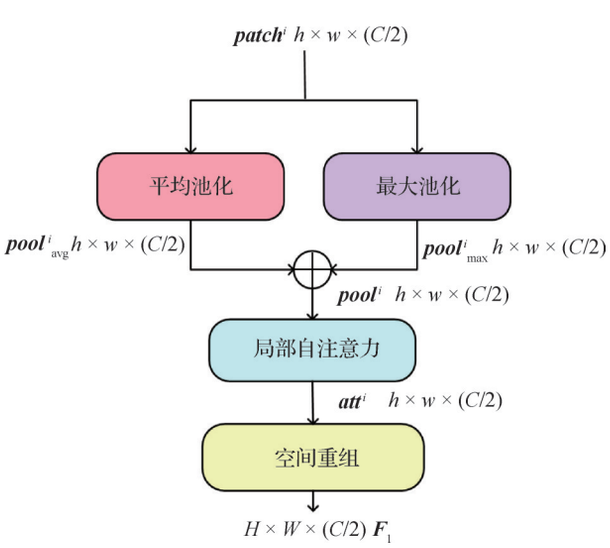


图5　基于池化的局部上下文机制

Fig. 5　Pooling-based local context mechanism

4)使用可学习参数的自适应加权策略对将全局上下文特征图 $\boldsymbol{F}_g$ 与局部上下文 $\boldsymbol{F}_l$ 进行自适应加权并进行通道拼接操作，随后，使用卷积、批次归一化

和激活函数获得包含全局和局部的上下文信息的特征图$\boldsymbol{F}_{gl}$，如图6所示。

通过自适应加权策略，检测算法可以根据红外飞机目标的特点来动态调整$\boldsymbol{F}_{gl}$中目标的全局和局部信息的权重，当目标特征较明显时应更多关注全局上下文，反之则应更侧重于局部上下文，从而使得检测算法更好地感知目标所处环境并突出目标特征，提高对目标和背景的判别能力。具体而言，将全局上下文和局部上下文的可学习加权参数分别设定为α和β，在目标检测算法损失函数最小化的指导下，随着优化器进行参数更新。假设卷积层中的卷积为$\boldsymbol{W}=[w_1,w_2,\cdots,w_i,w_c]$，对加权后的全局上下文特征图和局部上下文特征图进行特征聚合得到聚合特征图$\boldsymbol{F}_{agg}$，具体为

$$\boldsymbol{F}_{agg}([\boldsymbol{x}'_3,\boldsymbol{x}'_1,\cdots,\boldsymbol{x}'_i,\cdots,\boldsymbol{x}'_c])=\boldsymbol{W}\times[\alpha\times\boldsymbol{x}_3,\alpha\times\boldsymbol{x}_1,\cdots,\beta\times\boldsymbol{x}_i,\cdots,\beta\times\boldsymbol{x}_c] \tag{11}$$

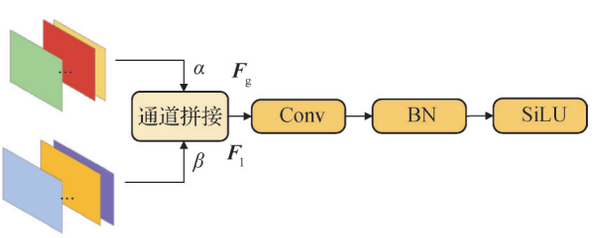


图6　自适应加权融合

Fig. 6　Adaptive weighted fusion

聚合特征图中，每个通道特征图的上下文特征的聚合方式为

$$\boldsymbol{x}'_i=w_i\times\alpha\times\boldsymbol{x}_3+w_i\times\alpha\times\boldsymbol{x}_1+\cdots+w_i\times\beta\times\boldsymbol{x}_i+\cdots+w_i\times\beta\times\boldsymbol{x}_c=w_i\times[\alpha\times\boldsymbol{x}_3,\alpha\times\boldsymbol{x}_1,\cdots,\beta\times\boldsymbol{x}_i,\cdots,\beta\times\boldsymbol{x}_c] \tag{12}$$

批归一化（batch normalization，BN）层通过对聚合特征图的归一化，使得聚合特征图的数据分布更加稳定，从而帮助检测算法更快地收敛，具体为

$$\boldsymbol{F}_{agg}^{bn}=\gamma\times\frac{\boldsymbol{F}_{agg}-\mu}{\sqrt{\sigma^2+\varepsilon}}+\delta \tag{13}$$

同时，通过激活函数使得检测算法能够捕获和学习数据中的非线性表达，从而学习到更高级的语义信息，具体为

$$\boldsymbol{F}_{gl}=\frac{\boldsymbol{F}_{agg}^{bn}}{1+e^{-\boldsymbol{F}_{agg}^{bn}}} \tag{14}$$

全局—局部上下文自适应加权融合的完整计算过程具体为

$$\boldsymbol{F}_{gl}=SiLU(BN(Conv(f_c(\alpha\times\boldsymbol{F}_g,\beta\times\boldsymbol{F}_l)))) \tag{15}$$

式中，f_c表示通道拼接；*Conv*表示卷积核大小、步长均为1×1的卷积层；*BN*表示批次归一化；*SiLU*表示激活函数；μ表示聚合特征图的均值；σ表示聚合特征图的方差；γ、δ表示*BN*层中的可学习参数；ε表示*BN*层中的正则化参数。

2　基于全局—局部上下文自适应加权融合机制的红外飞机检测算法

本文以YOLOv7为基础框架，应用提出的全局—局部上下文自适应加权融合机制，设计了红外飞机检测算法AWFGLC-YOLO，整体结构如图7所示。基础的YOLOv7目标检测模型架构主要由特征提取网络、特征融合网络和检测头3部分组成。特征提取网络由基本卷积模块（conv batch normalization SiLU，CBS）、扩展高效层聚合网络-1（extended efficient layer aggregation networks-1，ELAN-1）和最大池化卷积模块（maxpooling and conv batch normalization SiLU，MP-C3）3部分组成。特征融合采用路径聚集网络结构，通过自顶向下与自底向上的方式将不同尺度的特征信息进行融合，其中。检测头使用RepConv在下采样为8倍、16倍、32倍的特征图上进行结果预测，输出不同尺度上的预测结果。

本文将全局—局部上下文自适应加权融合机制融入到YOLOv7特征提取网络中，对下采样4倍的特征图和下采样32倍的特征图进行上下文建模，充分利用浅层特征图中的物理信息和深层特征图中的语义信息，提高检测算法对红外飞机目标特征的提取能力，从而更好地区分目标和背景。

为验证基于自适应加权的全局—局部上下文机制AWFGLC能够提高目标与背景的判别能力并突出目标特征，绘制特征提取网络的4倍下采样特征图，如图8所示。图8(a)表示原始图像，图8(b)表示输入到全局—局部上下文自适应加权融合机制的输入特征图，图8(c)表示全局上下文建模，图8(d)和图8(e)分别表示窗口划分与池化特征图和局部上下文建模特征图，图8(f)表示全局—局部上下文自适应加权融合机制的特征图。从图8(b)可以看出，经过特征提取网络前4层的特征提取与下采样后，输入到全局—局部上下文自适应加权融合机制AWFGLC特征图中目标特征微弱且面积较小。

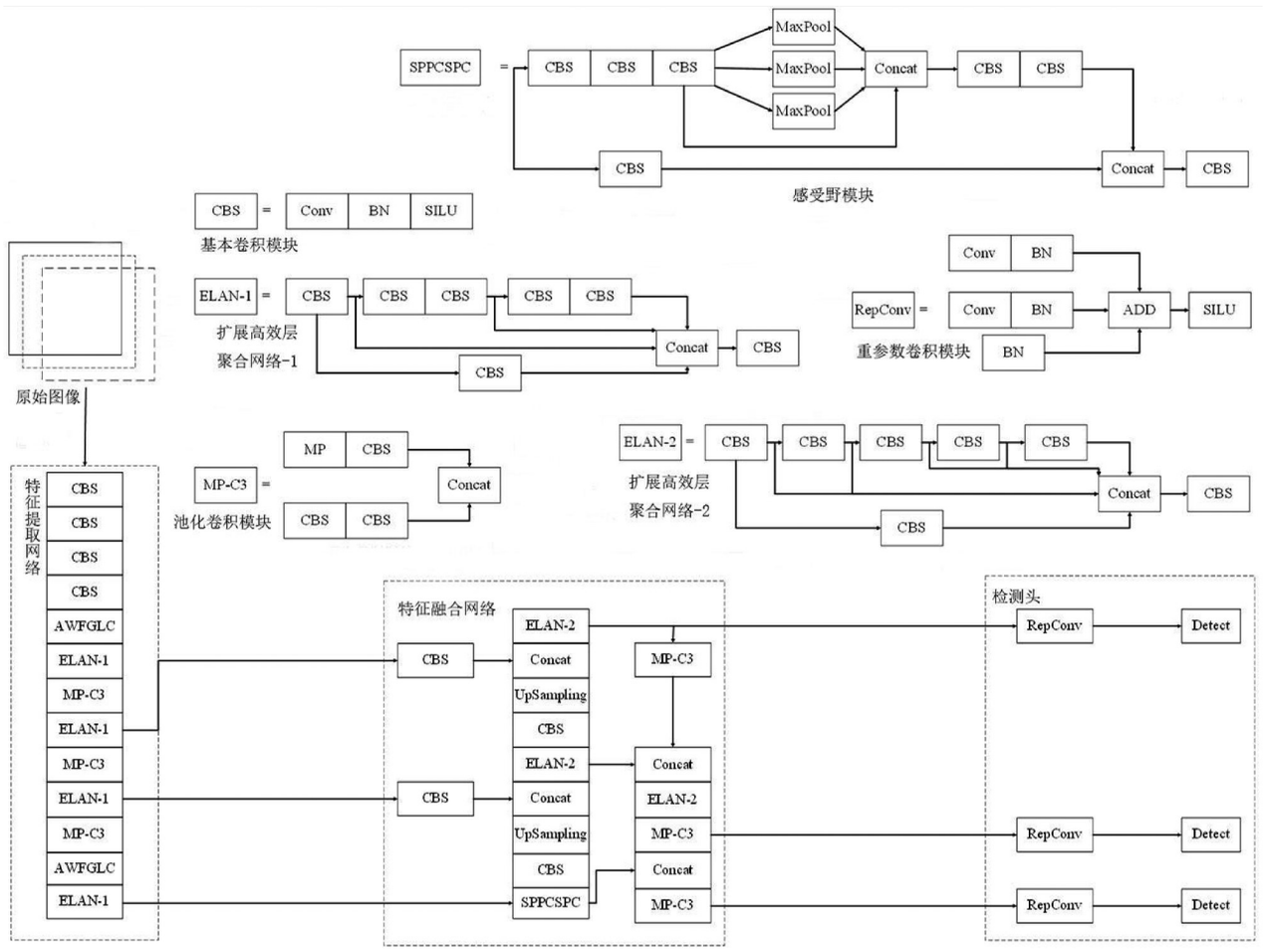


图7　AWFGLC-YOLO网络结构图

Fig. 7　AWFGLC-YOLO network structure diagram

AWFGLC对输入特征图进行通道重组与划分得到两个特征图，并对两个特征图进行不同的特征提取建模。从图8(c)可以看出，由于全局上下文特征图能够反映目标与全局图像中其他背景物像素点的关联性，突出的信息为目标中较亮的尾焰(黄色部分)，有利于区分目标与背景。但全局上下文特征图对目标与其周围邻域像素点之间的关联性建模能力较弱，没有突出较暗的飞机机身。从图8(d)和图8(e)可以看出，由于基于池化的局部上下文建模特征图能够反映目标特征与周围邻域特征的相关性，因此飞机机身特征得到了增强，这有助于检测算法更好地感知目标局部的特征。从图8(f)可以看出，自适应加权融合策略能够将包含较亮尾焰信息的全局上下文特征图和增强飞机机身特征的局部上下文特征图进行有效聚合，从而获得包含飞机整体轮廓信息的特征图。

3　实验与分析

3.1　实验环境

本文实验采用的硬件环境：CPU型号为Intel(R)Core(TM)i7-8700KCPU@3.70 GHz，运行内存为32 GB。GPU型号为NVIDIA 3090 Ti，显存大小为24 GB。软件环境：操作系统为Windows10，基于PyTorch版本为1.13深度学习框架，编程语言为Python3.7，使用CUDA11.1对GPU进行加速。

3.2　自制红外飞机数据集实验

自制红外飞机数据集共5 831帧图像，分为背向姿态(back-attitude)、侧向姿态(lateral-attitude)和后向姿态(backward-attitude)3种姿态。检测时区分机身(fuselage)和尾焰(tailflame)。人工标注的目标类别为6类：背向机身(back-attitude-fuselage，BAF)、背向尾焰(back-attitude-tailflame，BAT)、侧向机身

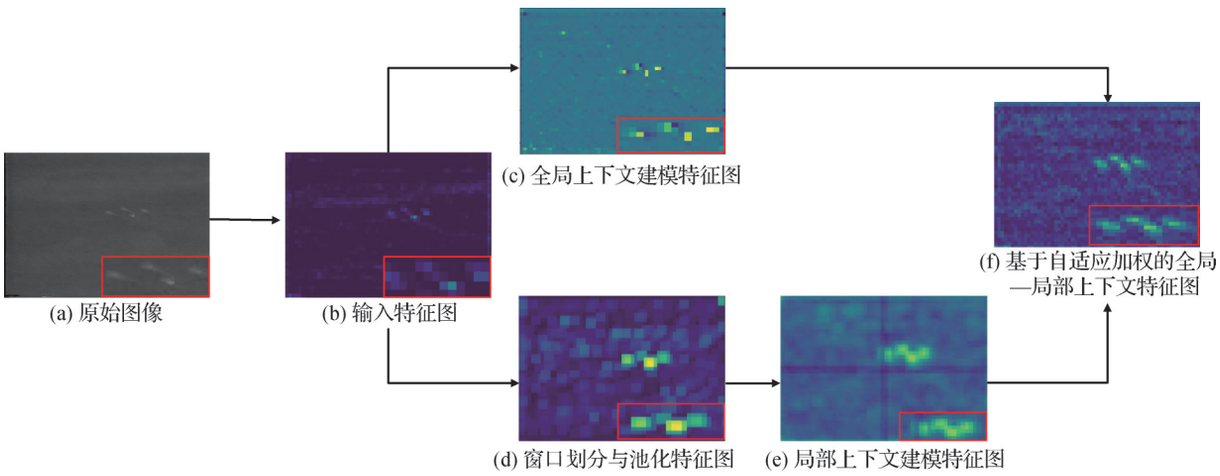


图 8　增强目标特征的效果图

Fig. 8　Enhanced effectiveness of targeted features ((a) original image; (b) input feature map; (c) global context modeling feature map; (d) window segmentation and pooled feature maps; (e) local context modeling feature map; (f) global-local contextual feature map based on adaptive weighting)

(lateral-attitude-fuselage, LAF)、侧向尾焰(lateral-attitude-tailflame, LAT)、后向机身(backward-attitude-fuselage, BWF)和后向尾焰(backward-attitude-tailflame, BWT)。将红外飞机数据集按照 10% 的比例随机划分出的数据作为测试集,余下的数据集利用十字交叉法,按照 8:2 的比例划分出训练集与验证集。

3.2.1　实验参数

在模型训练时,选用 Adam 随机梯度下降算法为优化器,初始学习率设为 0.001,动量因子设置为 0.937,权重衰减设置为 0.000 5。可学习参数 α 与 β 的初始值均为 0.5。

绘制训练过程中可学习参数 α 与 β 的变化曲线,如图 9 所示。从图 9 可以看出,可学习参数 α、β 的值随着训练逐渐下降且相较于作用在局部上下文机制上的 β,作用在全局上下文机制上的 α 的值更大一些。最优权重的选择是以 mAP50:95(mean average precision 50:95)的最高值为依据进行判定,在最优权重中 α 和 β 的值分别为 0.168 和 0.151。这表明在自制红外飞机数据集上基于全局—局部上下文自适应加权融合的机制更加倾向于学习目标的全局特征。

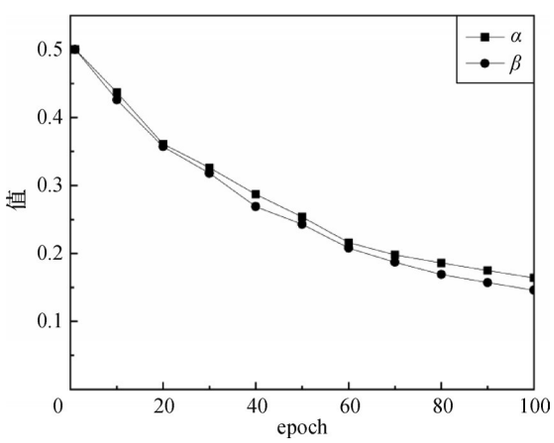


图 9　在自制红外飞机数据集上的可学习参数变化曲线

Fig. 9　Variation curves of learnable parameters in homemade infrared aircraft dataset

3.2.2　不同上下文机制的对比

以 YOLOv7 为基准检测算法,在其特征提取网络 4 倍下采样特征图和 32 倍下采样特征图的位置分别添加全局上下文(global context, GC)(Cao 等, 2019)、基于位置注意力的上下文机制(position attention, PA)(Fu 等, 2019)、基于局部 Transformer 的上下文机制(local Transformer, LT)(Liu 等, 2021)和基于自适应加权的全局—局部上下文机制 AWFGLC,从而对比不同上下文机制在红外飞机目标检测上的表现,实验结果如表 1 所示。

从表 1 可以观察到,相对于 YOLOv7, YOLOv7 + GC 和 YOLOv7 + LT 的 mAP50 分别下降 2.1% 和 0.9%, mAP50:95 分别下降 1.2% 和 0.6%, YOLOv7 + PA 的 mAP50 提高 0.3%, mAP50:95 提高 1.2%。在 YOLOv7 基础上添加全局—局部上下文自适应加权融合机制 AWFGLC 后,在参数量和浮点计算量分别增加 3.6% 和 1.7% 的条件下,相较于 YOLOv7, mAP50 和 mAP50:95 指标分别提高 1.4% 和 4.3%。这表明 AWFGLC 能够有效挖掘输入特征图中全局

表1　不同上下文机制的对比实验

Table 1　Comparative experiments with different contextual mechanisms

方法	mAP50/%	mAP50:95/%	Param./M	FLOPs/G
YOLOv7	96.4	61.4	**35.6**	**105.4**
YOLOv7+GC	94.3	60.2	36.2	105.9
YOLOv7+PA	96.7	62.6	37.1	106.1
YOLOv7+LT	95.5	60.8	37.9	106.7
AWFGLC-YOLO	**97.8**	**65.7**	36.9	107.2

注：加粗字体表示各列最优结果。

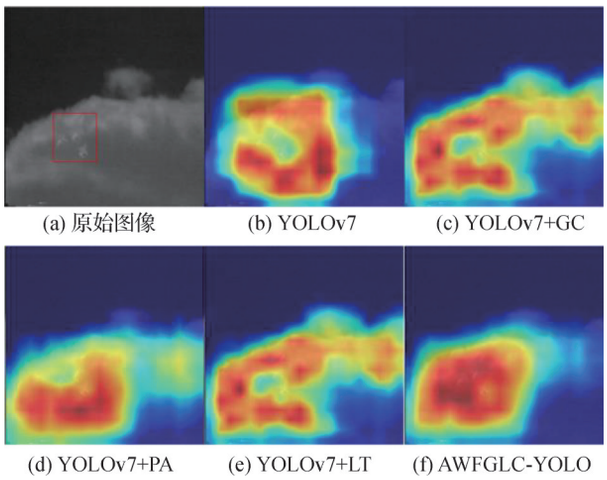
上下文和局部上下文信息，突出目标特征并增强算法对目标与背景的判别能力。

为了更加直观地观察使用基于自适应加权的全局—局部上下文机制的YOLOv7目标检测算法是否能够更好地提取目标特征，本文通过Eigen-CAM（Muhammad和Yeasin，2020）来可视化卷积层学习到的特征，如图10所示。Eigen-CAM通过计算卷积神经网络的特征图与预测层权重的线性组合，生成一个突出重要区域的热力图。热力图中的颜色通常表示模型对图像中不同区域的关注程度。具体而言，红色区域表示模型最关注的区域，这些区域对模型预测的贡献最大；黄色区域表示模型较为关注的区域，这些区域对模型预测有较大贡献；绿色区域表示模型稍微关注的区域，这些区域对模型预测的贡献较小；蓝色区域表示模型几乎不关注的区域，这些区域对模型预测的贡献最小。

在图10(a)中，机身辐射强度与周围邻域的辐射强度相近、可利用信息少且部分机身被云层遮挡。在图10(c)—图10(e)中，在YOLOv7的特征提取网络中分别使用GC、PA、LT后，相比原YOLOv7模型，可以关注到部分目标（黄色区域），但更加关注的是背景区域（红色区域）。从图10(f)可以看出，当在YOLOv7特征提取网络中融入基于自适应加权全局—局部上下文机制后，模型更加关注的是目标及其周围邻域（红色区域）。

3.2.3　不同检测算法的对比实验

将本文算法与基于关键点检测的检测算法CenterNet（Duan等，2019）、基于稀疏特征的双阶段检测算法Sparse R-CNN（Sun等，2021）、高效检测网络Efficientdet（Tan等，2020）、基于标签自动分配的检测算法Autoassign（Zhu等，2020）、YOLOF（Chen等，2021）、基于Transformer的检测算法Deformable DETR（deformable detection Transformer）（Zhu等，2021）、基于密集可区分查询的检测算法DDQ（dense distinct query）（Zhang等，2023）、YOLOv5、YOLOv7（Wang等，2023）和YOLOv8（Reis等，2024）等经典算法进行对比实验，如表2所示。



(a) 原始图像　(b) YOLOv7　(c) YOLOv7+GC

(d) YOLOv7+PA　(e) YOLOv7+LT　(f) AWFGLC-YOLO

图10　基于Eigen-CAM可视化热力图

Fig. 10　Eigen-CAM based visualization of thermal maps ((a)original image; (b)YOLOv7; (c)YOLOv7+GC; (d)YOLOv7+PA; (e)YOLOv7+LT; (f)AWFGLC-YOLO)

从表2可以看出，Efficientdet和YOLOF的

表2　自制红外飞机数据集上的不同算法比较

Table 2　Comparison of different algorithms in homemade infrared aircraft dataset

方法	mAP50/%	mAP50:95/%	Param./M	FLOPs/G
CenterNet	92.4	59.1	20.4	**14.2**
Autoassign	94.1	60.5	79.1	36.0
Efficientdet	73.7	45.9	18.3	46.4
Sparse R-CNN	90.8	59.7	105.9	64.6
YOLOF	80.6	49.3	42.3	41.2
Deformable DETR	93.8	59.2	40.1	27.4
DDQ	97.2	65.3	47.2	203.9
YOLOv5	94.4	63.7	**13.9**	16.4
YOLOv7	96.4	61.4	35.6	105.4
YOLOv8	96.0	62.3	25.0	79.1
AWFGLC-YOLO	**97.8**	**65.7**	36.9	107.2

注：加粗字体表示各列最优结果。

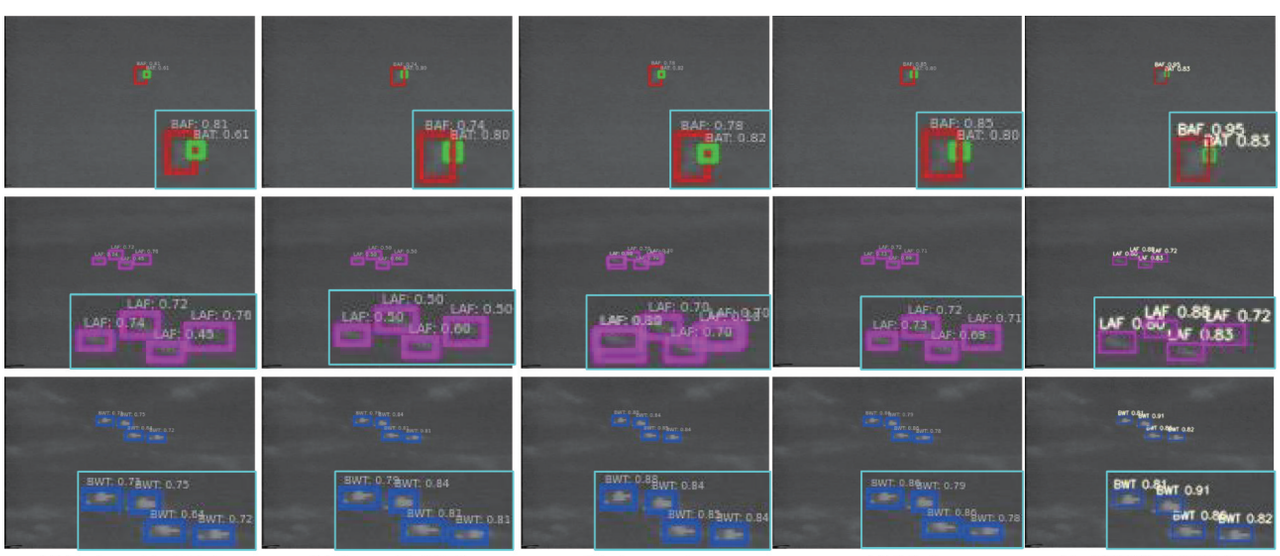
mAP50 和 mAP50:95 较低。Sparse R-CNN 有着较好的检测性能，mAP50 和 mAP50:95 达到 90.8% 和 59.7%，但其模型参数量与浮点计算量较大，分别达到 105.9 M 和 64.6 G。DDQ 的 mAP50 和 mAP50:95 分别为 97.2% 和 65.3%，但计算量较大。YOLOv5 与 CenterNet、Autoassign 和 Deformable DETR 的 mAP50 较为接近，但 YOLOv5 的 mAP50:95 比第 4 名的 YOLOv8 高 1.4%，且模型参数量与浮点计算量较小，为 13.9 M 和 16.4 G。与 YOLOv5 和 YOLOv8 相比，YOLOv7 的 mAP50 较高，达到 96.4%，但 YOLOv7 的 mAP50:95 比 YOLOv5 低 2.3%，且 YOLOv7 的模型参数量与计算量较大，分别为 35.6 M 和 105.4 G。与其他检测算法相比，本文算法在 mAP50 和 mAP50:95 均达到最高，分别为 97.8%、65.7%。因为引入基于自适应加权的全局—局部上下文机制，所以本文算法的参数量和浮点计算量相比 YOLOv7 分别增加 3.7%、1.7%。

不同算法的部分可视化检测结果如图 11 所示。从第 1 行图像可以看出，在图 11(a)中，Deformerble DETR 检测出背向机身(BAF)和背向尾焰(BAT)，置信度分别为 0.81 和 0.61。在图 11(b)中，Sparse R-CNN 检测出背向机身(BAF)和背向尾焰(BAT)，置信度分别为 0.74 和 0.80。在图 11(c)中，Autoassign 检测出背向机身(BAF)和背向尾焰(BAT)，置信度分别为 0.78 和 0.82。在图 11(d)中，DDQ 检测出背向机身(BAF)和背向尾焰(BAT)，置信度分别为 0.85 和 0.80。从图 11(e)可以看出，AWFGLC-YOLO 算法检测到的背向机身(BAF)、背向尾焰(BAT)置信度最高，分别为 0.95 和 0.83。

从第 2 行图像可以看出，Deformerble DETR、Sparse R-CNN 和 DDQ 均能够正确检测出侧向机身(LAF)的个数和位置，每个检测算法中目标的平均置信度分别为 0.67、0.53 和 0.71。Autoassign 误检出两个侧向机身(LAF)。AWFGLC-YOLO 正确检测出目标，平均检测置信度为 0.81。

从第 3 行图像可以看出，Deformerble DETR、Sparse R-CNN、Autoassign 和 DDQ 能够正确检测出后向尾焰(BWT)且平均置信度分别为 0.71、0.81、0.85 和 0.82；本文算法正确检测出红外飞机的后向尾焰(BWT)的数量、位置且平均置信度最高为 0.85。

图 12 为不同算法在自制红外飞机数据集上的训练损失曲线及 80～100 训练轮次的局部放大图。在 0～20 轮次的初始训练阶段中，所有算法的初始损失都较高并在这 20 个轮次内迅速下降。在 20～100 轮次的训练阶段，各个算法的损失逐渐趋于稳定。相较参与比较的其他算法，AWFGLC-YOLO 损失值最低，表明预测结果与真实标签之间的误差最小。



(a) Deformable DETR　(b) Sparse R-CNN　(c) Autoassign　(d) DDQ　(e) AWFGLC-YOLO

图 11　不同算法在自制红外飞机数据集上的检测结果

Fig. 11　Detection results of different algorithms in homemade infrared aircraft dataset

((a) Deformable DETR; (b) Sparse R-CNN; (c) Autoassign; (d) DDQ; (e) AWFGLC-YOLO)

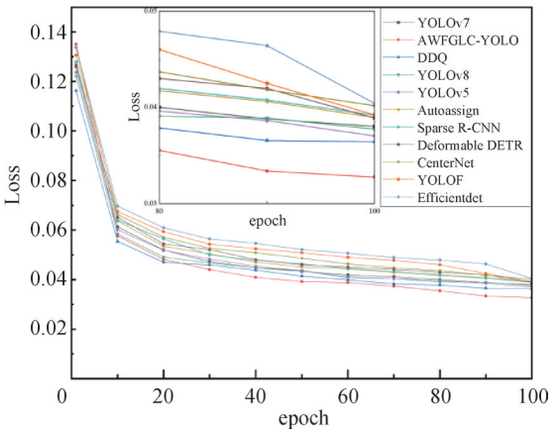


图12　不同算法在自制红外飞机数据集的训练损失曲线

Fig. 12　Training loss curves for homemade infrared aircraft dataset with different algorithms

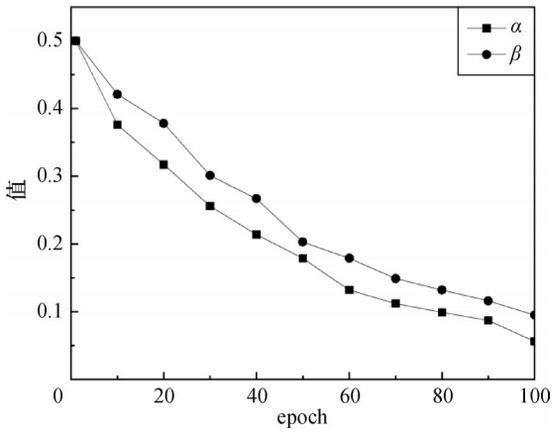


图13　在公开红外小目标数据集上的可学习参数变化曲线

Fig. 13　Variation curves of learnable parameters in public infrared small target dataset

3.3　红外小目标数据集实验

采用地/空背景下红外图像弱小飞机目标检测跟踪数据集(回丙伟 等,2020),该数据集共有22种典型场景,涵盖了单目标天空背景、两目标交叉飞行天空背景、单目标地面复杂背景、目标由远及近和目标由近到远等情形。在数据集中共有16 177帧图像、16 944个目标及标注,目标尺寸大多为3×3、5×5像素且目标为无人机。小目标尺寸较小,携带的特征信息较少,缺乏将其与背景或类似物体区分开的有效外观信息,在自然场景下的目标检测任务中融入目标周围的上下文信息能够丰富特征信息的表达,有利于区分小目标,从而提升模型对小目标的检测效果。红外小目标数据的训练集、验证集和测试集与红外飞机数据集采用相同的划分策略。同时,该数据集的实验参数与红外飞机数据集相同。

图13为训练过程中可学习参数α和β的变化曲线。从图13可以看出,可学习参数α和β的值随着训练迭代轮次的增加逐渐下降。相比于作用在全局上下文机制上的α,作用在局部上下文机制上的β的值更大一些。在最优权重中α和β的值分别为0.065和0.102。这是因为在公开数据集中,目标特征的尺寸较小、轮廓模糊,这使得AWFGLC更加倾向于学习目标的局部特征。

不同检测算法在地/空背景下红外小目标的检测结果如表3所示。从表3可以看出,本文算法的mAP50和mAP50:95分别达到88.7%和61.2%,均高于其他检测算法。

图14为不同算法在公开红外小目标数据集上的训练损失曲线及80～100训练轮次的局部放大图。在0～20轮次的初始训练阶段中,所有算法的初始损失较高并在20个轮次内迅速下降。在20～90轮次的训练阶段,各个算法的损失下降趋势变缓。相较参与比较的其他算法,AWFGLC-YOLO损失值最低,表明预测结果与真实标签之间的误差最小。在90～100轮次,DDQ算法的损失达到最低。

表3　公开红外小目标数据集上的不同算法比较

Table 3　Comparison of different algorithms in public infrared small target dataset

方法	mAP50/%	mAP50:95/%	Param./M	FLOPs/G
CenterNet	79.6	50.8	20.4	**14.2**
Autoassign	80.0	51.7	79.1	36.0
Efficientdet	64.2	48.07	18.3	46.4
Sparse R-CNN	76.4	50.2	105.9	64.6
YOLOF	54.2	20.4	42.3	41.2
Deformable DETR	81.5	52.2	40.1	27.4
DDQ	87.3	60.4	47.2	203.9
YOLOv5	84.7	57.4	**13.9**	16.4
YOLOv7	85.3	56.7	35.6	105.4
YOLOv8	85.7	57.2	25.0	79.1
AWFGLC-YOLO	**88.7**	**61.2**	36.9	107.2

注:加粗字体表示各列最优结果。

不同算法的部分可视化检测结果如图15所示。从图15(a)的第1幅图像可以看出,无人机处于地面树林上空,目标辐射强度比周围背景辐射强度大。

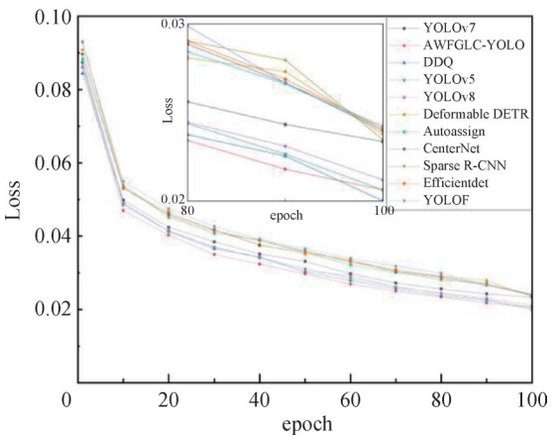
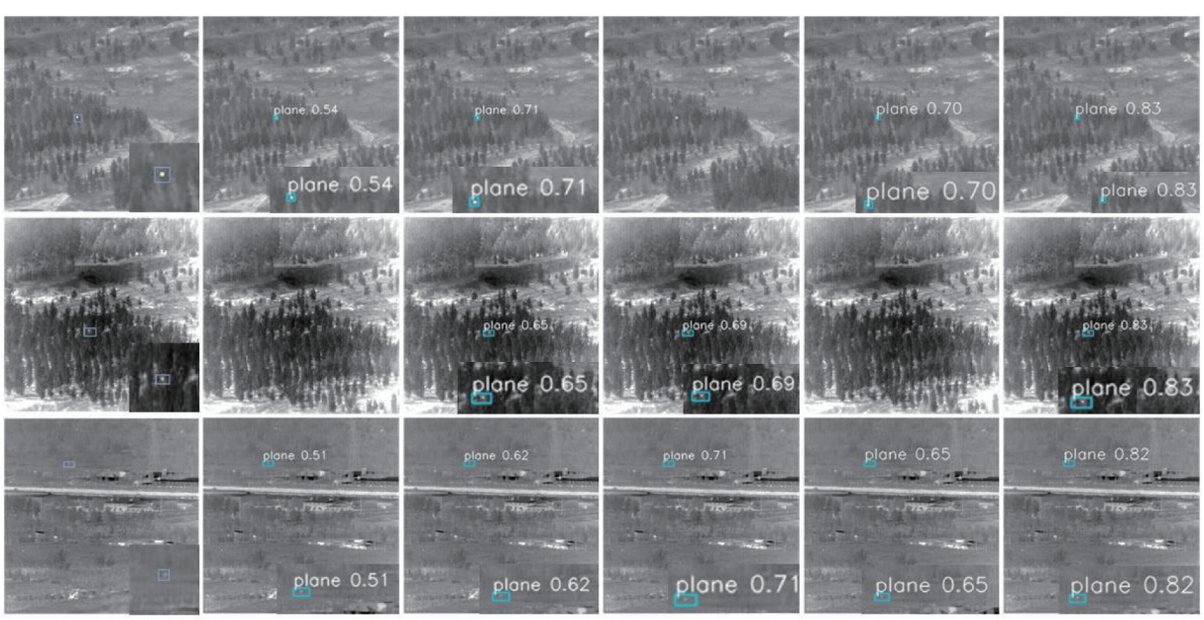


图 14　不同算法在公开数据集上的训练损失曲线

Fig. 14　Training loss curves for the public dataset with different algorithms

在图 15(a)的第 2 幅图像中，无人机处于地面树林上空，目标辐射强度与地面树林中其他背景的辐射强度相似。在图 15(a)的第 3 幅图像中，无人机处于包含道路、建筑等复杂地面背景的上空，目标辐射强度较弱。从图 15(b)—图 15(f)的第 1 行图像可以看出，除了 Autoassign 算法以外，其他检测算法均能检测出目标且本文算法检测结果的置信度最高，达到 0. 83。从第 2 行图像可以看出，Deformable DETR 算法没有检测出目标；DDQ 算法检测出两个目标，真实目标的置信度分别为 0. 79，误检目标的置信度为 0. 68；参与对比的其他检测算法能检测出目标且本文算法检测结果的置信度最高为 0. 83。从第 3 行图像可以看出，参与对比的检测算法均能检测出目标，本文算法检测结果的置信度为 0. 82，高于其他检测算法结果的置信度。



(a) 原始图像　(b) Deformable DETR　(c) Sparse R-CNN　(d) Autoassign　(e) DDQ　(f) AWFGLC-YOLO

图 15　不同算法在公开红外小目标数据集上的检测结果

Fig. 15　Detection results of different algorithms in public infrared small target dataset

((a) original images; (b) Deformable DETR; (c) Sparse R-CNN; (d) Autoassign; (e) DDQ; (f) AWFGLC-YOLO)

4　结　论

针对远距离红外飞机目标检测中存在的由于成像面积小、辐射强度较弱造成无法充分提取目标特征进而影响检测性能的问题，本文设计了基于全局—局部上下文自适应加权融合机制的红外飞机检测算法。该机制应用自注意力分别对飞机图像进行全局上下文和局部上下文建模，并根据目标特点利用自适应加权策略对全局和局部上下文信息进行融合，增强对红外飞机目标的特征表达，提高检测算法对目标与背景的判别能力。理论分析与实验结果表明，本文提出算法 AWFGLC-YOLO 能够对红外飞机目标进行有效检测。

深度卷积神经网络具有优异的性能，但较高的复杂度和非线性导致其透明度低、可解释性差，利用知识指导深度模型的推理过程使得建模过程具备解释性是未来的一个研究方向(杨朋波 等，2023)。因

此，后续研究工作可以考虑对红外飞机辐射、运动等知识进行建模并将其融入深度学习目标检测算法中，进一步提升检测算法的性能及其可解释性。

参考文献（References）

Cao Y, Xu J R, Lin S, Wei F Y and Hu H. 2019. GCNet: non-local networks meet squeeze-excitation networks and beyond//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision Workshop. Seoul, Korea (South): IEEE: 1971-1980 [DOI: 10.1109/ICCVW.2019.00246]

Chen Q, Wang Y M, Yang T, Zhang X Y, Cheng J and Sun J. 2021. You only look one-level feature//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 13034-13043 [DOI: 10.1109/CVPR46437.2021.01284]

Cheng X, Song C, Shi J G, Zhou L, Zhang Y F and Zheng Y H. 2021. A survey of generic object detection methods based on deep learning. Acta Electronica Sinica, 49(7): 1428-1438 (程旭, 宋晨, 史金钢, 周琳, 张毅锋, 郑钰辉. 2021. 基于深度学习的通用目标检测研究综述. 电子学报, 49(7): 1428-1438) [DOI: 10.12263/DZXB.20200570]

Duan K W, Bai S, Xie L X, Qi H G, Huang Q M and Tian Q. 2019. CenterNet: keypoint triplets for object detection//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 6568-6577 [DOI: 10.1109/ICCV.2019.00667]

Fu J, Liu J, Tian H J, Li Y, Bao Y J, Fang Z W and Lu H Q. 2019. Dual attention network for scene segmentation//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 3141-3149 [DOI: 10.1109/CVPR.2019.00326]

Ge Z, Liu S T, Wang F, Li Z M and Sun J. 2021. YOLOX: exceeding YOLO series in 2021 [EB/OL]. [2024-06-03]. https://arxiv.org/pdf/2107.08430.pdf

Hou Z Q, Sun Y, Guo H, Li J J, Ma S G and Fan J L. 2022. M-YOLO: an object detector based on global context information for infrared images. Journal of Real-Time Image Processing, 19(6): 1009-1022 [DOI: 10.1007/s11554-022-01242-y]

Hu S M, Zhao F, Lu H Z, Deng Y J, Du J M and Shen X L. 2023. Improving YOLOv7-tiny for infrared and visible light image object detection on drones. Remote Sensing, 15(13): #3214 [DOI: 10.3390/rs15133214]

Hui B W, Song Z Y, Fan H Q, Zhong P, Hu W D, Zhang X F, Lin J G, Su H Y, Jin W, Zhang Y J and Bai Y X. 2020. A dataset for infrared detection and tracking of dim-small aircraft targets under ground/air background. China Scientific Data, 5(3): 291-302 (回丙伟, 宋志勇, 范红旗, 钟平, 胡卫东, 张晓峰, 凌建国, 苏宏艳, 金威, 张永杰, 白亚茜. 2020. 地/空背景下红外图像弱小飞机目标检测跟踪数据集. 中国科学数据, 5(3): 291-302) [DOI: 10.11922/csdata.2019.0074.zh]

Jamali A, Roy S K, Bhattacharya A and Ghamisi P. 2023. Local window attention transformer for polarimetric SAR image classification. IEEE Geoscience and Remote Sensing Letters, 20: #4004205 [DOI: 10.1109/LGRS.2023.3239263]

Liu Z, Lin Y T, Cao Y, Hu H, Wei Y X, Zhang Z, Lin S and Guo B N. 2021. Swin transformer: hierarchical vision transformer using shifted windows//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 9992-10002 [DOI: 10.1109/ICCV48922.2021.00986]

Muhammad M B and Yeasin M. 2020. Eigen-CAM: class activation map using principal components//Proceedings of 2020 International Joint Conference on Neural Networks. Glasgow, UK: IEEE: 1-7 [DOI: 10.1109/IJCNN48605.2020.9206626]

Redmon J and Farhadi A. 2018. YOLOV3: an incremental improvement [EB/OL]. [2024-06-03]. https://arxiv.org/pdf/1804.02767.pdf

Reis D, Kupec J, Hong J and Daoudi A. 2024. Real-time flying object detection with YOLOv8 [EB/OL]. [2024-06-03]. https://arxiv.org/pdf/1804.02767.pdf

Sun P Z, Zhang R F, Jiang Y, Kong T, Xu C F, Zhan W, Tomizuka M, Li L, Yuan Z H, Wang C H and Luo P. 2021. Sparse R-CNN: end-to-end object detection with learnable proposals//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 14449-14458 [DOI: 10.1109/CVPR46437.2021.01422]

Tan J H, Yin W, Liu L M and Wang Y N. 2021. DenseNet-siamese network with global context feature module for object tracking. Journal of Electronics and Information Technology, 43(1): 179-186 (谭建豪, 殷旺, 刘力铭, 王耀南. 2021. 引入全局上下文特征模块的DenseNet孪生网络目标跟踪. 电子与信息学报, 43(1): 179-186) [DOI: 10.11999/JEIT190788]

Tan M X, Pang R M and Le Q V. 2020. EfficientDet: scalable and efficient object detection//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 10778-10787 [DOI: 10.1109/CVPR42600.2020.01079]

Wang C Y, Bochkovskiy A and Liao H Y M. 2021a. Scaled-YOLOv4: scaling cross stage partial network//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 13024-13033 [DOI: 10.1109/CVPR46437.2021.01283]

Wang C Y, Bochkovskiy A and Liao H Y M. 2023. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 7464-7475 [DOI: 10.1109/CVPR52729.2023.00721]

Wang K D, Li S Y, Niu S S and Zhang K. 2019. Detection of infrared small targets using feature fusion convolutional network. IEEE

Access, 7: 146081-146092 [DOI: 10.1109/ACCESS. 2019. 2944661]

Wang Z X, Xu Z M, Xue Y Y, Lang C Y, Li Z and Wei L L. 2023. Global and spatial multi-scale contexts fusion for vehicle re-identification. Journal of Image and Graphics, 28(2): 471-482 (王振学, 许喆铭, 雪洋洋, 郎丛妍, 李尊, 魏莉莉. 2023. 融合全局与空间多尺度上下文信息的车辆重识别. 中国图象图形学报, 28(2): 471-482) [DOI: 10.11834/jig.210849]

Xiao F, Lu H, Zhang W J, Huang S J, Jiao Y L, Lu Z T and Li Z S. 2024. Aerial infrared image target recognition algorithm based on rotation equivariant convolution. Acta Armamentarii, 45(8): 2817-2827 (肖锋, 卢浩, 张文娟, 黄姝娟, 焦雨林, 卢昭廷, 李照山. 2024. 基于旋转等变卷积的航拍红外图像目标识别算法. 兵工学报, 45(8): 2817-2827) [DOI: 10.12382/bgxb.2023.0503]

Xue S, An H Y, Lyu Q Y and Cao G H. 2024. Image target detection algorithm based on YOLOv7-tiny in complex background. Infrared and Laser Engineering, 53(1): #20230472 (薛珊, 安宏宇, 吕琼莹, 曹国华. 2024. 复杂背景下基于YOLOv7-tiny的图像目标检测算法. 红外与激光工程, 53(1): #20230472) [DOI: 10.3788/IRLA20230472]

Yang P B, Sang J T, Zhang B, Feng Y G and Yu J. 2023. Survey on interpretability of deep models for image classification. Journal of Software, 34(1): 230-254 (杨朋波, 桑基韬, 张彪, 冯耀功, 于剑. 2023. 面向图像分类的深度模型可解释性研究综述. 软件学报, 34(1): 230-254) [DOI: 10.13328/j.cnki.jos.006415]

Yang Z G, Xia X Y, Liu Y M, Wen G W, Zhang W E and Guo L M. 2024. LPST-Det: local-perception-enhanced swin transformer for SAR ship detection. Remote Sensing, 16(3): #483 [DOI: 10.3390/rs16030483]

Zhang G H, Li C F, Li G Y and Lu W D. 2024. Small target detection algorithm for UAV aerial images based on improved YOLOv7-tiny. Engineering Science and Technology (张光华, 李聪发, 李钢硬, 卢为党. 2024. 基于改进YOLOv7-tiny的无人机航拍图像小目标检测算法. 工程科学与技术 [DOI: 10.15961/j. jsuese. 202300593]

Zhang S L, Wang X J, Wang J Q, Pang J M, Lyu C Q, Zhang W W, Luo P and Chen K. 2023. Dense distinct query for end-to-end object detection//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 7329-7338 [DOI: 10.1109/CVPR52729.2023.00708]

Zhang W C, Fu C, Xie H Y, Zhu M, Tie M and Chen J X. 2021. Global context aware RCNN for object detection. Neural Computing and Applications, 33(18): 11627-11639 [DOI: 10.1007/s00521-021-05867-1]

Zhao X F, Xia Y T, Zhang W W, Zheng C and Zhang Z L. 2023. YOLO-ViT-based method for unmanned aerial vehicle infrared vehicle target detection. Remote Sensing, 15(15): #3778 [DOI: 10.3390/rs15153778]

Zhou W N, Wu Z H, Zhang Z D, Peng L and Xie L B. 2024. Lightweight small target detection method based on weak feature enhancement. Control and Decision, 39(2): 381-390 (周葳楠, 吴治海, 张正道, 彭力, 谢林柏. 2024. 基于弱特征增强的轻量化小目标检测方法. 控制与决策, 39(2): 381-390) [DOI: 10.13195/j.kzyjc.2022.1432]

Zhou X, Jiang L, Hu C X, Lei S, Zhang T T and Mou X A. 2022. YOLO-SASE: an improved YOLO algorithm for the small targets detection in complex backgrounds. Sensors, 22(12): #4600 [DOI: 10.3390/s22124600]

Zhu B J, Wang J F, Jiang Z K, Zong F H, Liu S T, Li Z M and Sun J. 2020. AutoAssign: differentiable label assignment for dense object detection [EB/OL]. [2024-06-03]. https://arxiv.org/pdf/2007.03496.pdf

Zhu X Z, Su W J, Lu L W, Li B, Wang X G and Dai J F. 2021. Deformable DETR: deformable transformers for end-to-end object detection [EB/OL]. [2024-06-03]. https://arxiv.org/pdf/2010.04159.pdf

作者简介

徐红鹏，男，硕士研究生，主要研究方向为计算机视觉和机器学习。E-mail：1317736871@qq.com

刘刚，通信作者，男，教授，主要研究方向为计算机视觉和机器学习。E-mail：lg19741011@163.com

习江涛，男，教授，主要研究方向为信号处理和机器学习。E-mail：jiangtao@uow.edu.au

童军，男，副教授，主要研究方向为信号处理和机器学习。E-mail：jtong@uow.edu.au