张进, 龚声蓉, 周立凡, 等. 一种基于改进 YOLOv7 的地表形变和大气相位检测方法[J]. 智能计算机与应用,2025,15(6):114-121. DOI:10.20169/j.issn.2095-2163.250617

一种基于改进 YOLOv7 的地表形变和大气相位检测方法

张 进[1], 龚声蓉[2], 周立凡[2], 凤黄浩[2], 刘 浩[2]
(1 湖州师范学院 信息工程学院, 浙江 湖州 313000; 2 常熟理工学院 计算机科学与工程学院, 江苏 常熟 215500)

摘 要: 地表形变的检测效果对于地质灾害的形变探测极其重要,针对 InSAR 成像极其容易受到大气相位的影响,使得 InSAR 成像精度低、导致难以判断准确的地质灾害形变的问题,本文提出了一种基于改进 YOLOv7 的地表形变和大气相位目标检测方法。首先将 Efficient MOdel(EMO)轻量化网络和 CoTAttention 注意力模块嵌入到 YOLOv7 的骨干网络中,然后将 CFPNet 中的 EVC-Block 嵌入到颈部,获得更多特征信息,采用 *MPDIoU* 损失函数提高模型对大气相位和实际地表形变目标的定位能力。实验结果表明,所提算法的均值平均精度 *mAP* 达到 74.9%,与 YOLOv7 基线算法相比分别提高 2.2%。

关键词: 大气相位; 地表形变; 目标检测; YOLOv7; 损失函数
中图分类号: TP391 **文献标志码**: A **文章编号**: 2095-2163(2025)06-0114-08

A surface deformation and atmospheric phase detection method based on improved YOLOv7

ZHANG Jin[1], GONG Shengrong[2], ZHOU Lifan[2], FENG Huanghao[2], LIU Hao[2]
(1 School of Information Engineering, Huzhou University, Huzhou 313000, Zhejiang, China;
2 School of Computer Science and Engineering, Changshu Institute of Technology, Changshu 215500, Jiangsu, China)

Abstract: The detection effect of surface deformation is extremely important for the detection of geological disasters. InSAR imaging is extremely susceptible to the influence of atmospheric phase, resulting in low accuracy of InSAR imaging and difficulty in accurately determining geological disaster deformation. Therefore, this paper proposes a surface deformation and atmospheric phase target detection method based on improved YOLOv7. Firstly, the Efficient MOdel (EMO) lightweight network and CoTAttention attention module are embedded into the backbone network of YOLOv7. Then, the EVC-Block in CFPNet is embedded into the neck to obtain more feature information. The *MPDIoU* loss function is used to improve the model's positioning ability for atmospheric phase and actual surface deformation targets. The experimental results show that the average accuracy *mAP* of the proposed algorithm reaches 74.9%, which is 2.2% higher than the accuracy of the YOLOv7 baseline algorithm.

Key words: atmospheric phase; surface deformation; target detection; YOLOv7; loss function

0 引 言

地表形变是在自然或人类因素作用下形成的对广大人民的生产生活秩序和社会的稳步发展造成破坏的地质灾害之一[1],不仅会对人们的生命财产造成破坏,还会对社会的和谐稳定带来威胁。在地质灾害发生之前,通常伴有明显的地表形变变化,尽可能早地发现地表形变,对于预防地质灾害、减少损失具有重要作用。因此,开展以形变为主的地质灾害隐患检测可以为地质灾害的及时防范提供科学依据。

地质灾害的地表形变检测主要包括地面监测技

基金项目: 国家自然科学基金(61972059, 62376041, 42071438, 62102347); 中国博士后科学基金项目(2021M69236); 教育部符号与知识工程重要实验室项目(吉林大学)(93K172021K01)。

作者简介: 张 进(1999—),男,硕士研究生,主要研究方向:图像处理,目标检测; 周立凡(1984—),男,博士,副教授,主要研究方向:人工智能与地球科学交叉研究; 凤黄浩(1984—),男,博士,讲师,主要研究方向:社交机器人,人机交互,情感计算,音乐科学; 刘 浩(1994—),男,博士,讲师,主要研究方向:深度学习,激光雷达点云,多源多尺度遥感,森林物候研究等。
通信作者: 龚声蓉(1966—),男,博士,教授,博士生导师,主要研究方向:视频图像智能分析,数字媒体安全与取证分析,遥感智能信息处理等。Email:shrgong@cslg.edu.cn。

收稿日期: 2023-11-03

术和遥感监测技术等，但是地面监测技术有着无法进行大规模形变监测的缺点，而遥感技术以其宏观、快速等优势弥补了地面监测技术的缺点，被广泛应用地质灾害形变探测中，特别是干涉合成孔径雷达（Interferometry Synthetic Aperture Radar，InSAR）已经被证明是检测地表形变的最独特、有效的手段方法。而现有的地质灾害的地表形变检测识别有赖于专家经验识别形变区域，具有一定的局限性且需要大量的人力资源。此外，InSAR 成像易受大气相位影响，导致 InSAR 影像中地表形变目标不准确。

随着深度学习不断发展，深度学习目标检测方法不断应用在遥感图像领域，通过深度卷积网络来进行自动特征学习，基于深度学习的目标检测算法分为 2 类。一类是以 R-CNN[2]、Fast R-CNN[3]、FasterR-CNN[4]、Mask R-CNN[5]等为代表的双阶段目标检测算法，另一类是以 SSD[6]、RetinaNet[7]、YOLOv4[8]、YOLOv6[9]等为代表的单阶段目标检测算法。目前，基于深度学习的图像处理方法是遥感图像目标检测领域的研究热点，赵珊等学者[10]通过双注意力机制来构建深度神经网络实现神经网络自动学习特征，通过添加细节提取模块以及通道注意力特征融合模块来提取更多的细节特征，解决高分辨率特征经过深度 CNN 后导致的信息丢失问题，通过结合 KL 散度优化损失函数解决了神经网络直接用于目标检测存在的问题。杨晨等学者[11]通过引入特征复用以增加特征图中的小目标特征信息，以解决下采样导致的特征图中包含的小目标信息较少或消失的问题，使用 EMFFN（Efficient Multi-scale Feature Fusion Network）的特征融合网络代替原有的 PANet，通过添加跳跃连接以及跨层连接高效融合不同尺度的特征图信息，同时提出一种包含通道与像素的双向特征注意力机制，以提高模型在复杂背景下的检测效果，但是改进措施极大地提高了模型的参数量，模型的训练速度与推理速度存在一定程度的下降。刘涛等学者[12]为了强化对遥感图像的多尺度特征表达能力，通过增加一个融合浅层语义信息的细粒度检测层来提高对小目标的检测效果，但是模型对全局语义信息的感知能力以及预测框的定位能力有待提升。余俊宇等学者[13]利用集中特征金字塔（Centralized Feature Pyramid，CFP）[14]解决遥感图像因目标分布密集以及检测背景复杂导致检测效率较低的问题，引入混合注意力模块 ACmix 加强网络对于小目标检测的敏感度，优化损失函数提升模型性能。

综上所述，本文基于 InSAR 技术获得地表形变的相位数据，提出了一种改进 YOLOv7 的 InSAR 遥感图像检测方法，自动精确检测出影像 InSAR 成像精度的大气相位和实际地表形变目标，以便预防地质灾害。首先替换骨干网络为 EMO[15]轻量化网络，降低网络计算量和参数复杂度，其次在骨干网络中引入 CoTAttention[16]注意力机制加强骨干网络对图像的特征提取能力，抑制背景信息的干扰，使模型更加关注目标特征，获得更多的大气相位和地表形变特征信息，然后在颈部引入集中金字塔 CFPNet 中的 EVC-Block 模块，获得特征图的更多局部信息，提高检测精度。此外，通过将损失函数替换成 *MPDIoU*[17] 进一步提升预测框预测的准确性。

1　模型及改进

1.1　YOLOv7 网络模型及其改进

YOLOv7 是目前 YOLO 算法系列杰出代表之一，其性能超越以往多数的目标检测算法。YOLOv7 网络主要由输入端、骨干网络、检测头网络三部分组成。输入端输入 $S \times S \times 3$ 尺寸的图像，经过预处理后进入骨干网络，骨干网络由多个 CBS、ELAN 和 MP 模块组成进行特征提取，输出 3 个不同尺度特征图，检测头网络由 SPPCSPC 结构、引入 ELAN-H 和 MP2 结构的特征提取网络以及 RepConv 结构组成，接收骨干网络输出 3 个特征图，然后经过整合输出预测结果。

本文以 YOLOv7 为基础框架，设计了一种可以快速自动检测出大气相位和地表形变的网络模型，改进的 YOLOv7 网络整体架构如图 1 所示。本文首先将 YOLOv7 的骨干网络替换成 EMO 轻量网络来降低模型的计算量和参数复杂度，然后在骨干网络末尾集成 CoTAttention 注意力机制来提取地表形变和大气相位的关键特征信息，并使用集中金字塔 CFPNet 中的 EVC-Block 嵌入到网络颈部，使网络充分提取检测目标的特征信息，提高对目标的检测精度。

1.2　EMO

YOLOv7 原本的骨干网络是由多个卷积组成，使得网络计算量大且参数复杂。为了降低网络的计算量和参数复杂度，在保证精度的情况下，通过引入高效、轻量级的模型 EMO 轻量化模块来降低模型的计算量和参数复杂度，实现在参数、计算量和性能之间的平衡。EMO 由多个反向残差移动块（iRMB）堆叠而成，通过将卷积神经网络（CNN）和 Transformer

的有效组件统一起来,吸收了类似卷积神经网络的效率来模拟短距离依赖和类似 Transformer 的动态建模能力来学习长距离交互,实现了高效的模型性能。

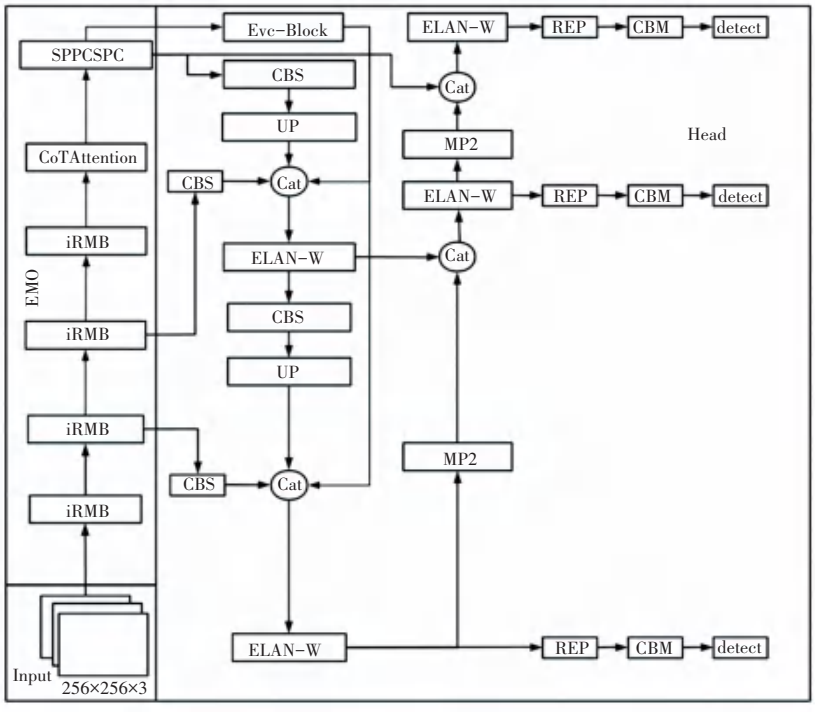


图 1 改进的 YOLOv7 网络整体架构

Fig. 1 Overall architecture of improved YOLOv7 network

由于 Transformer 中的多头注意力机制(MHSA)和前馈神经网络(FFN)模块以及 MobileNet-v2 中的倒残差模块(Inverted Residual Block, IRB)具有相似的结构,可以通过抽象出一种 Meta Mobile Block 结构来表示,即采用不同的参数扩展率 λ 和高效算子 F 来实例化不同的模块,其网络结构如图 2 所示。

iRMB 以 Meta Mobile Block 结构为基础,仅由标准卷积和多头自注意力组成,没有其他复杂的运算符。其网络结构如图 3 所示。

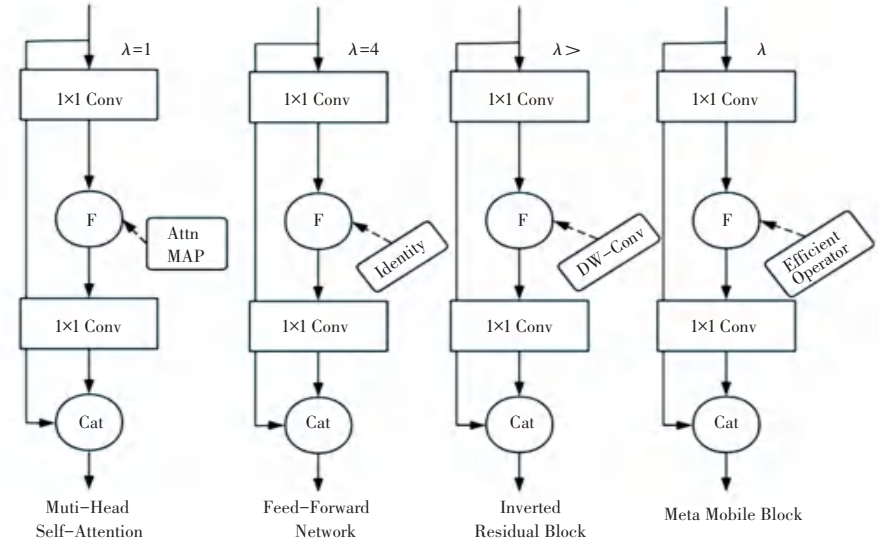


图 2 Meta Mobile Block 结构

Fig. 2 Meta Mobile Block structure

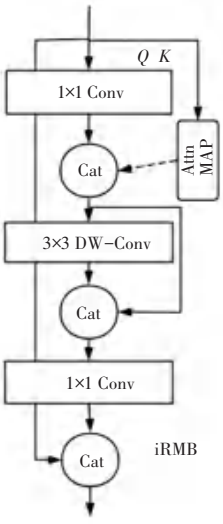


图 3　iRMB 网络结构

Fig. 3　iRMB network structure

iRMB 中的 F 被建模为级联的 MHSA 和卷积运算，公式可以抽象为：

$$F(\cdot) = Conv(MHSA(\cdot)) \tag{1}$$

此外，iRMB 通过使用 DW-Conv 来实现卷积，借助 Window-Based MHSA（W-MHSA）使得计算复杂度变为线性，从而节约计算开销，并受益于 DW-Conv，iRMB 还可以通过步长适应下采样操作，并且不需要任何位置嵌入来向 MHSA 引入位置偏差。

EMO 模型的详细架构如图 4 所示。将其替换成 YOLOv7 的骨干网络，图像在输入到网络中，图像分别经过 4 倍、8 倍、16 倍、32 倍的下采样，其中在 4 倍和 8 倍下采样的 iRMB 模块中采用 BN+SiLU，在 16 倍和 32 倍下采样的 iRMB 模块中采用 LN+ReLU，通过不同的归一化和激活函数来提升网络的稳定性和效率，根据多头注意力机制更适合为更深层次的语义特征建模的特性，将其嵌入到后面 2 个 iRMB 模块来进一步提升网络的性能，有助于学习对象间的关系，使模型能够从大邻域中收集和关联特征信息。

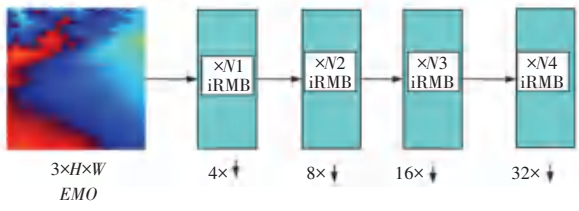


图 4　EMO 模型的详细架构

Fig. 4　Detailed architecture of EMO model

1.3　注意力机制

为了有效降低地表形变图像复杂背景的干扰，提高对地表形变目标和大气相位目标的识别能力，在骨干网络中嵌入 CoTAttention，实现更好的信息交换和整合，获得更多大气相位和地表形变目标特征信息。CoTAttention 结合 Transformer 与卷积的优点，不仅可以加强网络对地表形变图像中目标的敏感度，而且还能降低由背景带来的噪声影响。CoTAttention 的结构如图 5 所示。

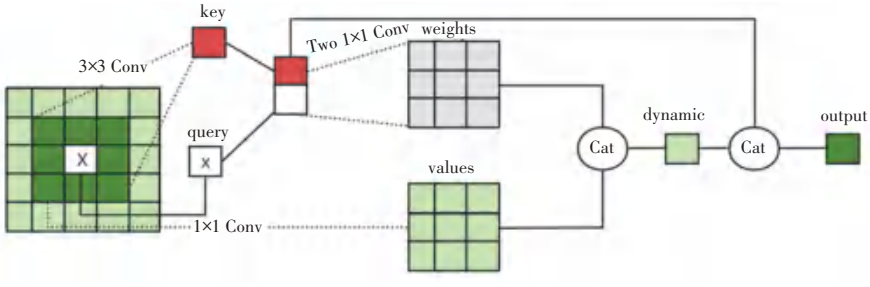


图 5　CoTAttention 结构图

Fig. 5　Structure diagram of CoTAttention

与传统的自注意仅利用孤立的查询键对来测量注意矩阵相比，CoTAttention 充分利用的键之间留下了丰富的上下文。CoTAttention 首先通过卷积对输入键进行上下文编码，从而得到输入的静态上下文表示，然后进一步将编码的键与输入查询连接起来，通过 2 个连续的 1×1 卷积来学习动态多头注意矩阵。将学习到的注意矩阵乘以输入值，实现输入的动态上下文表示。最后，将静态和动态上下文表示的融合作为输出[18]。

考虑到 CoTAttention 更适合深层卷积且过早使用 CoTAttention 可能会导致网络相对较浅而特征映射相对较大，丢失重要的上下文信息，因此，本文将 CoTAttention 模块应用于骨干网络的末尾以提取更多的差异化特征，让网络提升对重要区域内与非重要区域的目标特征的提取能力。

1.4　CFPNet 中的 EVC-Block

对于地表形变遥感图像而言，因其背景范围较大且存在大气相位影响的原因，YOLOv7 网络对于

地表形变目标和大气相位特征的提取不够充分,造成最终的地表形变目标检测结果不佳。因此,本文通过在 YOLOv7 网络颈部引入 CFPNet 中的 EVC-Block,即一种集中特征金字塔,来加强网络对大气相位和地表形变目标的特征提取能力。其输入是骨干网络最高层提取的特征图。该特征图会经过 CoTAttention 注意力模块的处理,加强特征图的特征,获取更多细节信息,然后再输入到 EVC-Block 中,EVC-Block 是由轻量级别 MLP 和视觉中心 LVC 并行连接的模块组成的,其结构如图 6 所示。

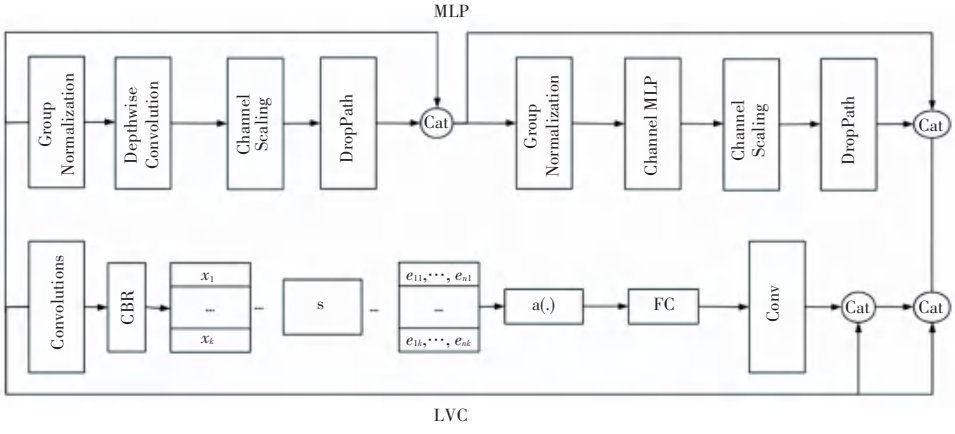


图 6 EVC-Block 结构图

Fig. 6 EVC-Block structure diagram

通过 EVC-Block 中的 MLP 可以获得更全面、更准确的地表形变遥感图像中尺度差异较大的目标特征,从而提高对大气相位和地表形变目标的识别、分类和定位的准确性。EVC-Block 中的 LVC 主要由卷积层、全连接层以及字典编码器组成,通过卷积层对输入的地表形变特征和大气相位特征进行编码,并使用具有归一化的卷积和 Relu 激活函数组成的 CBR 模块对编码进行处理后再输送到字典编码器当中,使用编码器能够获得关于编码的整个图像的完整信息,之后通过将编码器的输出馈送到全连接层和卷积层以预测突出关键类的特征。通过 LVC 模块可以获取图像局部角落细节信息,可以更好地分辨出地表形变遥感图像中的地表形变目标和大气相位。此外,该模块还可以通过调整网络关注的区域来避免漏检现象出现,提高检测的准确性。

总地来说,通过将 CFPNet 中的 EVC-Block 嵌入到 YOLOv7 的颈部不仅可以获得顶层特征图目标特征,而且可以获得局部信息,通过局部信息使网络可以充分提取大气相位和地表形变的特征信息,提高检测精度。

1.5 *MPDIoU* 损失函数

在 YOLOv7 的原网络框架中采用的是 *CIoU*[19] 损失函数来计算预测框的坐标,其公式如下所示:

$$CIoU = IoU - \frac{\rho^2(b,b^{gt})}{c^2} + \alpha v \tag{2}$$

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right) \tag{3}$$

$$\alpha = \frac{v}{1 - IoU + v} \tag{4}$$

其中,w^{gt}、h^{gt}、w、h 分别表示地面真值框与预测框的宽度与高;*IoU* 表示真实框与预测框交并比;$\rho^2(b,b^{gt})$ 表示求预测框与地面真值框中心点之间的欧式距离;c 表示预测框与真值框的最小包围框的最短对角线长度;α表示用于平衡参数;v 表示用于衡量长宽比是否一致[20]。

虽然 *CIoU* 相对于大多损失函数而言,已经考虑到了预测框的重叠面积、中心距离、高宽比,但是当预测框与真实框重合,长宽比的惩罚项没有起到任何作用,并且在预测框的回归中,高质量的预测框一般而言要比低质量的预测框少得多,将影响网络的训练。此外,*CIoU* 是边界框回归损失函数的一种,而现有的边界框回归的损失函数在不同的预测结果下具有相同的值,当预测的边界框与真值边界框具有相同的长宽比,但宽度和高度的值完全不同时,无法进行有效优化,这降低了边界框回归的收敛速度和精度。

MPDIoU 以提高边界框回归的收敛速度和精度为目标,直接最小化预测边界框与真实边界框之间的左上和右下点距离,简化了 2 个边界框之间的相

似性比较。*MPDIoU* 不仅包含了现有损失函数中考虑的所有相关因素，即重叠或不重叠区域、中心点距离、宽度和高度偏差等，而且简化了计算过程，数学公式定义如下：

$$MPDIoU = IoU - \frac{d_1^2}{h^2 + w^2} - \frac{d_2^2}{h^2 + w^2} \tag{5}$$

其中，*IoU* 表示预测边界框与真实边界框的相交面积和并集面积之比；d_1，d_2 分别表示预测边界框与真实边界框之间的左上点和右上点的最小距离；h 和 w 分别表示图像预测框的高度和宽度。

2　实验与结果分析

2.1　数据集

本实验所采用的数据集是通过 SyInterferoPy 模拟生成地表形变 InSAR 图像数据集，该数据集主要包含 2 种类别，分别是大气相位和地表形变，同时也包含了部分背景图片以提升训练后模型的鲁棒性，数据集总共包含 2 000 张图片，为满足实验需求，将数据集按照 8∶1∶1 的比例划分为训练集、验证集和测试集。样本图片如图 7 所示。左侧蓝色框标记的是地表形变，右侧橙色框标记的大气相位。

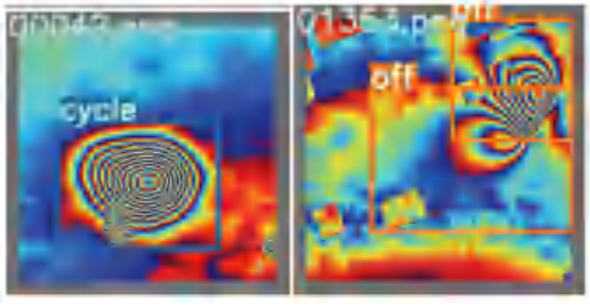


图 7　样本图片

Fig. 7　Sample image

2.2　实验环境

本实验的运行环境是 Win11 操作系统，处理器为 Intel(R) Core(TM) i7-12700H 2.30 GHz，内存为 16 GB，GPU 是 NVIDIA GeForce RTX 3070，采用 python 版本是 3.8.16，配置深度学习环境为 Pytorch1.12.1，Torchvision 为 0.13.1，CUDA 版本 11.3。在各模型输入的图像大小统一设为 256×256，训练代数 *epoch* 设为 400，优化器选择 SGD，权重衰减为 5e-4，初始学习率为 1e-2，并采用余弦退火算法调整学习率。

2.3　评价指标

实验采用精度（*P*）、召回率（*R*）、*F*1 分数、平均精度均值（*mAP*）以及浮点运算（*FLOPs*）作为评价指标，对改进后的 YOLOv7 模型的性能进行综合评价。相关指标计算公式具体如下：

$$P = \frac{TP}{TP + FP} \tag{6}$$

$$R = \frac{TP}{TP + FN} \tag{7}$$

$$F1 = \frac{2PR}{P + R} \tag{8}$$

$$AP = \int_0^1 P(R)\,\mathrm{d}R \tag{9}$$

$$mAP = \frac{\sum_{j=1}^{c} AP(j)}{C} \tag{10}$$

其中，*TP* 表示真正例；*FP* 表示假正例；*FN* 表示假负例；*P* 表示精度；*R* 表示召回率；*C* 表示目标检测的总类别数[21]。在模型性能的评价中，精度高或者召回率高并不一定意味着模型是准确的，通常使用精度和召回率的调和平均值，即 *F*1 分数作为模型的综合评价指标。而 *mAP* 是用来评估模型检测所有类别综合能力的指标，通过取所有类平均精度（*AP*）的平均值来度量算法的精度[22]。此外，为了比较不同模型的计算复杂度，使用 *Flops* 指标来表示不同算法之间的差异[23]。*fps* 性能指标可以用来评价一个模型的检测速度，其数值越大，表明模型的检测速度越快。

2.4　消融实验

为了进一步验证本文改进 YOLOv7 模型中各改进方法的有效性，在同一配置和相同的地表形变数据集下通过消融实验验证各个方法有效性，改进 YOLOv7 模型的各个模块消融实验结果见表 1。将 YOLOv7 原来的骨干网络替换为 EMO 轻量化网络，其精度提升了 1.4%，网络整体计算量下降了很多；将 CoTAttention 嵌入骨干网络末尾，其精度提升了 1.3%；将 CFPNet 中的 EVC-Block 模块嵌入到网络颈部，其精度提升了 0.8%；将原损失函数替换成 *MPDIoU* 损失函数，其精度提升了 1.1%。此外，改进的各个模块在 *fps* 指标上都优于 YOLOv7，在计算量上除了 CoTAttention 注意力机制增加了网络的计算量。其余均未增加计算量，在参数复杂度上，除了 CoTAttention 注意力机制增加了网络的参数复杂度，其余均低于原网络的参数复杂度。因此，实验结果证明了本文所改进的方法在地表形变 InSAR 遥感图像目标检测上具有很好的效果。

表 1 改进 YOLOv7 各模块消融实验

Table 1 Improved YOLOv7 ablation experiments for each module

Model	*mAP*@0.5/%	*fps*/FPS	*Param*/M	*Flops*/G
YOLOv7	72.7	108.70	36.9	103.2
+EMO	74.1	128.20	23.1	40.0
+CoTAttention	74.0	144.92	46.1	110.5
+EVC-Block	73.5	185.20	7.1	16.3
+MPDIoU	73.8	138.90	36.5	103.2

2.5 对比实验

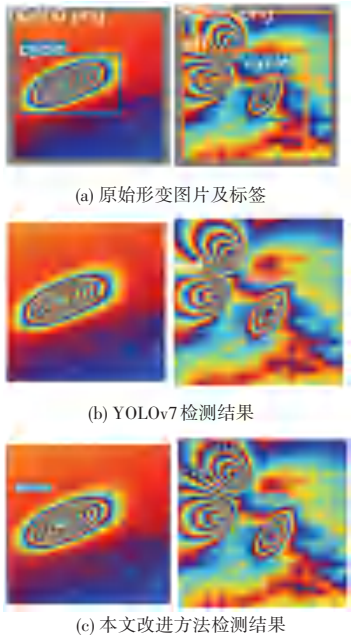
为了验证本文改进算法的检测性能,将改进的网络和原网络与具有代表性的网络如 Faster R-CNN、DCN、Deformable_Detr、YOLOX 等其它网络以 *mAP*@0.5、*Flops*(*G*) 和 *F*1、*fps* 指标和 *Param* 进行对比,其实验结果见表 2。从表 2 中可以看出,本文的方法的精度值、*fps* 和 *F*1 比其它网络都高,*Flops* 虽然高于 YOLOX,但是相较于其它网络,计算量是远低于其它网络的,参数量虽然比 Ld 和 YOLOX 高,但是相较于其它网络,参数量是比较低的。本文方法的 *mAP*@0.5 比 Faster R-CNN、DCN、Deformable_Detr、Lab、Ld、Libra R-CNN、RetinaNet、Sparse R-CNN、VFNet、YOLOX、YOLOv7 分别提高了 4.2、4.6、12.7、11.6、10.7、7.0、10.3、25.1、9.3、7.5、2.2 个百分点。总地来说,改进 YOLOv7 网络模型在地表形变检测方面更有优势。

表 2 不同网络的对比

Table 2 Comparison of different networks

Model	*mAP*@0.5/%	*Flops*/G	*F*1/%	*fps*/FPS	*Param*/M
Faster R-CNN	70.7	206.67	63.57	55.2	41.13
DCN	69.6	213.16	53.10	50.3	97.11
Deformable_Detr	62.2	195.23	3.90	57.5	39.82
Lab	66.3	201.46	2.76	87.7	31.89
Ld	64.2	157.91	21.69	36.8	19.08
Libra R-CNN	67.9	157.08	65.90	43.7	40.80
RetinaNet	64.6	204.80	16.28	48.8	36.13
Sparse R-CNN	49.8	149.90	2.75	64.5	105.95
VFNet	65.6	189.02	60.08	48.3	32.49
YOLOX	67.4	33.30	16.31	27.8	8.94
YOLOv7	72.7	103.20	72.40	104.2	36.49
本文	74.9	53.30	75.40	144.9	27.70

为了更直观地体现实验效果,分别使用 YOLOv7 和本文改进的方法对地表形变图像进行可视化实验展示,其检测结果如图 8 所示。图 8(a)为原始形变图片以及标签,图 8(b)为 YOLOv7 检测出的结果,图 8(c)为本文改进方法检测出来的结果。可以看出在地表形变 InSAR 图像检测任务中,相较于 YOLOv7 而言,本文的方法利用 EMO 和 CoTAttention 在轻量化网络的同时,通过注意力机制获取语义信息,来学习形变和大气相位对象间的关系,在颈部通过 EVC-Block 加强对顶层特征局部信息获取,可以更准确地检测到大气相位和地表形变,降低漏检率,进一步提升了检测效果,保证了地表形变检测的有效性。



(a) 原始形变图片及标签

(b) YOLOv7 检测结果

(c) 本文改进方法检测结果

图 8 检测结果

Fig. 8 Test results

3 结束语

为了解决当前 InSAR 地表形变检测方法中存在的检测时间长、精度低、人工成本高且由于大气相位使得地表形变判断准确性不高的问题,本文以 YOLOv7 网络框架为基础,通过引入 EMO 轻量化网络、CoTAttention、集中特征金字塔 CFPNet 中的 EVC-Block 模块以及将损失函数替换成 *MPDIoU* 来解决对 InSAR 图像地表形变识别不高且漏检的情况。通过大量实验证明改进后的网络模型在推测速度和检测的评价精度上都有所提升。该模型实现了对 InSAR 图像中大气相位和地表形变目标的快速、精确检测,有利于预防和有效监督地质灾害发生。由于使用的地表形变数据集较小,本文模型在准确呈现上还存在一定的优化空间。此外,YOLOv7 检测网络中采用 PANET,旨在特征融合过程中能够均

衡发挥不同尺度输入特征的作用，但实际上，不同尺度输入对最终输出的贡献是不同的，地表形变图像具有多尺度的目标，对检测的精度会有一定影响，可以引入不同的特征金字塔网络结构进行改进，适应不同的输入特征，并进一步挖掘信息，获得更高水平的特征融合，提升其性能。

参考文献

[1] 周吕. 雷达干涉测量地表沉降和建筑物变形监测及分析[D]. 武汉：武汉大学，2018.

[2] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2014: 580-587.

[3] GIRSHICK R. Fast R-CNN [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2015: 1440-1448.

[4] REN Shaoqing, HE Kaiming, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[5] HE Kaiming, GKIOXARI G, DOLLAR P, et al. Mask RCNN [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 2961-2969.

[6] LIU Wei, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector [C]// European Conference on Computer Vision. Cham: Springer, 2016: 21-37.

[7] 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. Piscataway, NJ: IEEE, 2017: 2980-2988.

[8] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv, 2004.10934, 2020.

[9] LI Chuyi, LI Lulu, JIANG Hongliang, et al. YOLOv6: A single-stage object detection framework for industrial applications [J]. arXiv preprint arXiv, 2209.02976, 2022.

[10] 赵珊，郑爱玲，刘子路，等. 通道分离双注意力机制的目标检测算法[J]. 计算机科学与探索，2023，17(5)：1112-1125.

[11] 杨晨，佘璐，杨璐，等. 改进 YOLOv5 的遥感影像目标检测算法[J]. 计算机工程与应用，2023，59(15)：76-86.

[12] 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.

[13] 余俊宇，刘孙俊，许桃. 融合注意力机制的 YOLOv7 遥感小目标检测算法研究[J]. 计算机工程与应用，2023，59(20)：167-175.

[14] QUAN Yu, ZHANG Dong, ZHANG Liyan, et al. Centralized feature pyramid for object detection [J]. arXiv preprint arXiv, 2210.02093, 2022.

[15] ZHANG Jiangning, LI Xiangtai, LI Jian, et al. Rethinking mobile block for efficient attention-based models [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway, NJ: IEEE, 2023: 1389-1400.

[16] LI Yehao, YAO Ting, PAN Yingwei, et al. Contextual Transformer networks for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(2): 1489-1500.

[17] MA Siliang, XU Yong. MPDIoU: A loss for efficient and accurate bounding box regression[J]. arXiv preprint arXiv, 2307.07662, 2023.

[18] 赵浩然. 基于卷积神经网络的塑封芯片内部缺陷智能识别方法[D]. 绵阳：西南科技大学，2023.

[19] ZHENG Zhaohui, WANG Ping, REN Dongwei, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation [J]. IEEE Transactions on Cybernetics, 2021, 52(8): 8574-8586.

[20] 李小军，邓月明，陈正浩，等. 改进 YOLOv5 的机场跑道异物目标检测算法[J]. 计算机工程与应用，2023，59(2)：202-211.

[21] JIANG K, XIE T, YAN R, et al. An attention mechanism-improved YOLOv7 object detection algorithm for hemp duck count estimation[J]. Agriculture, 2022, 12(10): 1659.

[22] NEPAL U, ESLAMIAT H. Comparing YOLOv3, YOLOv4 and YOLOv5 for autonomous landing spot detection in faulty UAVs [J]. Sensors, 2022, 22(2): 464.

[23] LIU M, XIE J, HAO J, et al. Alightweight and accurate recognition framework for signs of Xray weld images [J]. Computers in Industry, 2022, 135: 103559.