

基于注意力机制的遥感目标检测方法 研究与实现

李聪慧

西京学院计算机学院, 陕西 西安

收稿日期: 2025年10月17日; 录用日期: 2025年11月17日; 发布日期: 2025年11月26日

摘 要

遥感目标检测在环境监测、城市规划等领域具有重要应用价值,但由于遥感图像中目标尺度变化大、背景复杂以及特征提取不够鲁棒等问题,传统的检测方法难以满足实际需求。本文旨在研究一种基于注意力机制的遥感目标检测方法,以提升检测精度和鲁棒性,为遥感图像分析提供更高效的技术支持。本文提出了一种基于YOLOv8改进的遥感目标检测模型YOLO-Shuffle-MSDA,通过引入ShuffleAttention模块和多尺度特征融合模块(MSDA)来增强特征表达能力。ShuffleAttention模块通过注意力机制优化特征提取,有效区分目标与复杂背景;MSDA模块则通过多尺度特征融合,提升模型对多尺度目标的检测能力。通过多组对比实验和可视化检测结果可以看出,本文方法有效提升了检测精度,减少了误检和漏检。此外,本文设计并实现了完整的遥感目标检测系统,包括模型训练、验证和测试模块,并开发了用户友好的演示验证系统,支持单张图像、批量图像、视频以及实时摄像头检测。

关键词

目标检测, 注意力机制, YOLOv8, 遥感图像

Research and Implementation of Remote Sensing Object Detection Method Based on Attention Mechanism

Conghui Li

School of Computer Science, Xijing University, Xi'an Shaanxi

Received: October 17, 2025; accepted: November 17, 2025; published: November 26, 2025

文章引用: 李聪慧. 基于注意力机制的遥感目标检测方法研究与实现[J]. 计算机科学与应用, 2025, 15(11): 234-246.
DOI: 10.12677/csa.2025.1511300

Abstract

Remote sensing target detection has important application values in monitoring, urban planning and other fields, but due to the large change in target scale, complex background and insufficiently robust feature extraction in remote sensing images, detection is difficult to actually require. This paper aims to study a remote sensing object detection based on attention mechanism to improve detection accuracy and robustness and provide more efficient technical support for remote sensing images. This paper proposes a remote sensing object detection model YOLO-Shuffle-MSDA based on YOLOv8, which enhances feature expression capabilities by introducing the ShuffleAttention module and the multi-scale feature fusion module (MSDA). The ShuffleAttention module optimizes feature extraction through attention mechanism to effectively distinguish targets from complex backgrounds; the MSDA module improves the model's detection ability of multi-scale targets through multi-scale feature fusion. Through multiple sets of comparison experiments and visual inspection results, it can be seen that this paper effectively improves detection accuracy, false detection and missed detection. In addition, this paper designs and implements a complete remote sensing object detection system, including model training, verification and testing modules, and a user-friendly demonstration verification system, supporting single images, batch images, videos and real-time camera detection.

Keywords

Target Detection, Attention Mechanism, YOLOv8, Remote Sensing Imagery

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 绪论

1.1. 研究背景与意义

随着遥感技术的快速发展, 遥感图像目标检测在国防、农业、城市规划等领域具有重要应用价值。然而, 遥感图像中目标通常具有尺度多变、背景复杂、分布稠密等特点, 传统算法在特征提取与定位精度上存在局限。深度学习尤其是注意力机制的引入, 为解决这些问题提供了新思路。本文基于 YOLOv8 框架提出 YOLO-Shuffle-MSDA 模型, 通过引入 ShuffleAttention 模块与多尺度特征融合模块(MSDA), 提升模型对小目标及复杂背景的检测性能。

1.2. 国内外研究现状

目前的目标检测算法大致分为单阶段检测算法和两阶段检测算法两大类型策略。

两阶段检测算法采用分阶段的方法对目标进行定位与分类, 检测精度高。其中, Faster R-CNN [1]、Mask R-CNN、R-CNN [2]、Fast R-CNN [3]、FSOD [1]、Cascade R-CNN [4]等是典型的代表。

单阶段检测算法检测速度快, 适用于实时应用。YOLO [5]系列单次检测器和 SSD [6]系列单网多尺度检测器是这一类算法的代表。它们能够直接在输入的图像中预测目标的位置和种类, 而不需要创建候选区域, 这使得它们在预测速度上具有明显的优越性。

1.2.1. 两阶段目标检测算法研究现状

R-CNN [2]这一经典的两阶段目标检测技术, 巧妙地结合了选择性搜索、迁移学习方法以及高效的分

类器，从而达到了目标检测的高效率和准确性。

Faster R-CNN [1]采纳了空间金字塔池化(SPP)的核心思想，成功地将各种大小的感兴趣区域(ROI-Pooling)有效地映射到一个统一的特征图中，这样可以省去不必要的计算步骤并加快特征提取速度。Faster R-CNN [1]摒弃了传统的选择性搜索方式，而是采用区域提议网络(RPN)自动生成候选区域，从而有效地解决了多目标检测中的计算复杂性问题。

在 Mask R-CNN 中引入掩模划分任务，通过引入附加分支产生像素级目标掩模以提供更精细目标信息。这种方法在小物体与大物体总体召回率上都有所提高，平均精度也得到增强。

Cascade R-CNN [4]指出，由于目标检测任务的高度复杂性和变化性，单次预测很难准确地锁定目标的具体位置信息。使单次预测精度无法充分生成精确的位置信息，易受目标大小，形状，姿态和光照等诸多因素干扰。

1.2.2. 单阶段目标检测算法研究现状

对于遥感图像中的目标快速检测，很多研究者对 YOLO [5]算法进行了大量有益的改进与尝试。例如，Gong 及其团队[7]考虑到卫星图像中目标和方向的多样性，巧妙地采用了 Swin Transformer [8]作为核心网络，多尺度特征被有效地提取出来，算法对于目标及方向改变的敏感性通过注意力机制得到显著提高。Wu 及其团队[9]提出了一种结合 YOLOv5 和全卷积神经网络(FCN)优点的创新快速目标检测系统，另外，李坤亚和他的团队[10]进一步提高该算法提取不同尺度物体特征的能力，使遥感图像检测得到明显发展。

国内外有很多研究人员在遥感影像检测领域采用了基于深度学习目标检测的方法。在 YOLOv3 骨干网络重建和 RFB 结构介绍下，王建军等人[11]增强了遥感影像小目标检测精度。但这种方法会加大模型推理时间，不能满足算力较低平台下实时检测需求。Wang 等人[12]用 C3CMix 模块降低参数量和加快检测速度。Shi 等人[13]在中提出了对滑动切片和边缘丢弃，有效地解决了多次检测和定位模糊的问题。

1.2.3. 注意力机制的研究现状

该机制使模型能在计算资源受限的情况下快速地从全图中抽取更有区分度的数据，在视觉目标检测用得较多。Zhang [14]等在注意力信息的指导下，设计出一种新的检测方法，有效地降低无关信息对模型的干扰。Chen 和他的团队[15]提出了一种全新的三流注意力显著性目标检测技术，以注意力机制为基础，通过对特征提取使目标更加清晰。

1.3. 应用前景

基于注意力机制遥感目标检测方法引入注意力机制增强了遥感图像目标检测精度与效率。它不仅促进遥感图像处理技术进步，而且在自然资源管理，城市规划和建设以及灾害监测和评估方面提供强大技术支持。伴随着深度学习技术不断进步以及遥感数据不断丰富，该方法已经在更加广泛的领域中扮演了重要的角色，给人类社会提供了更加便捷与价值所在。

2. 系统分析与设计

2.1. 系统需求分析

基于注意力机制进行遥感目标检测的目的是设计和实现一个基于注意力的遥感目标检测模型能有效应对遥感图像目标尺度多变和背景复杂的特点；该系统需要支持数据集导入，模型训练，验证与测试等功能，使用户能够自行定义数据集并优化模型参数；同时开发了用户友好演示与验证系统，该系统支持单张图像，批量图像，视频及实时摄像头检测功能等，且可加载与运行以注意力机制为核心的优化模型，本发明实现了检测结果实时显示与可视化，达到了实际工作中遥感目标检测高精度高效能的要求。

功能需求

操作人员用例图如图 1 所示。

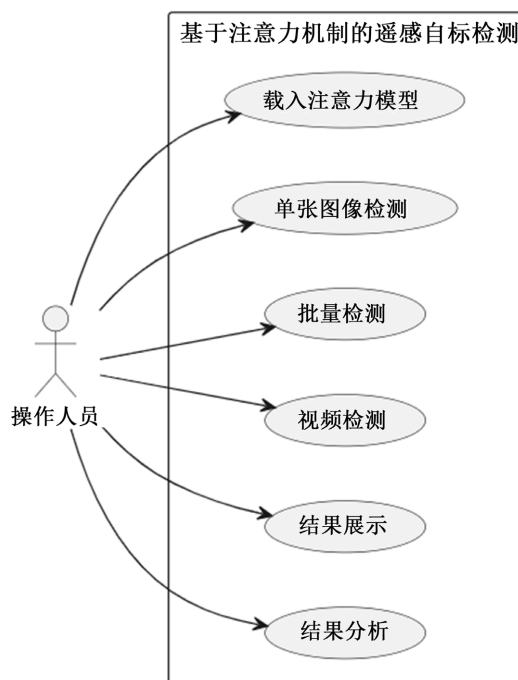


Figure 1. Operator use case diagram

图 1. 操作人员用例图

2.2. 系统设计

注意力机制的遥感目标检测系统功能整体分为两个模块：模型设计和系统设计。系统的重难点集中在模型设计模块，系统按注意力机制的遥感目标检测算法框架将模型设计划分为三个结果模块，从数据准备开始，经过模型训练到模型测试，形成一个完整的模型开发流程，确保模型的准确性和鲁棒性；在系统设计中，通过系统功能设计和测试用例设计，保证系统功能的完善性和测试的全面性。系统总体结构组成图 2 所示。

2.2.1. 系统功能设计

注意力机制的遥感目标检测系统主要由载入注意力模型，单张图像检测，批量检测，结果展示，结果分析七个功能组成。具体实现如下：

(1) 载入注意力模型功能：系统需要加载一个预先训练好的注意力机制模型，这个模型能够对遥感图像中的目标进行有效的识别和定位。

(2) 单张图像检测功能：操作员需选择适合自己的模型及图片，点击即可启动检测。检测结束后系统将检测结果展示于界面。

(3) 批量检测功能：操作员选择文件夹开启，可一次录入多幅图片，系统会并行地处理这些图片，并迅速给出每一幅图片的检测结果。

(4) 视频检测功能：操作员选择开启视频，既可视频文件，也可实时视频流。该系统会处理视频内容，并将这些检测数据整合成视频，或者以帧的方式呈现在用户界面中。

(5) 摄像头检测功能：操作员开启摄像头使系统与摄像头设备连接。实时采集摄像头图像帧并检测采

集各帧图像中目标。该系统将检测结果实时地显示于界面上，并对视频流进行检测目标标注。

(6) 结果展示：在检测结束时，该系统会显示每一幅图片的结果。一般都是将检测目标直接标记于图像中，并对目标的类别、位置等有关属性信息进行标记。

(7) 结果分析：该遥感目标检测系统能在保留图片检测与视频检测结果的前提下，充分发挥注意力机制优点，提高遥感影像目标检测准确性与效率。

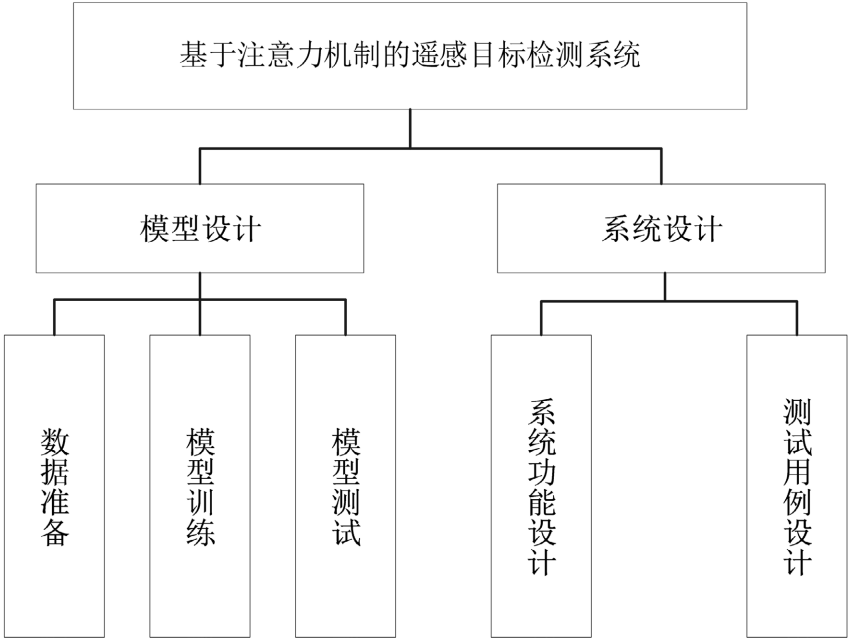


Figure 2. System overall structure design diagram
图 2. 系统总体结构设计图

系统功能模块如图 3 所示：

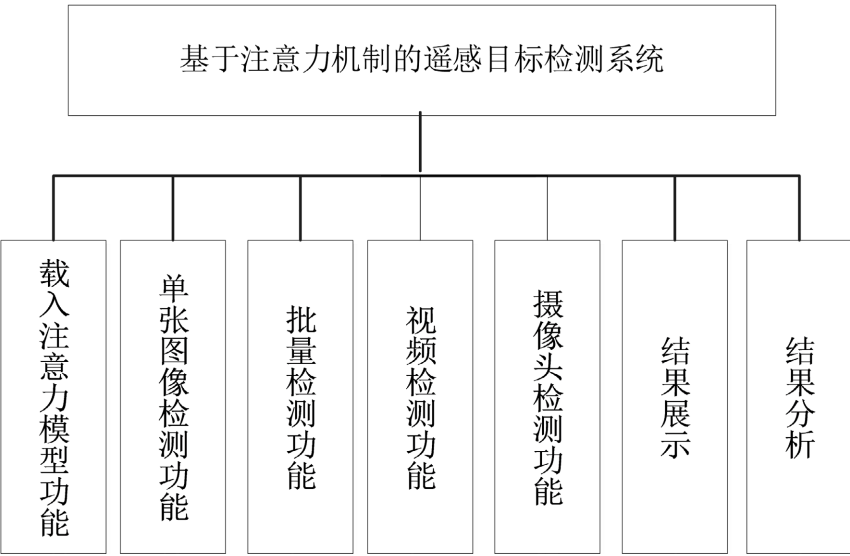


Figure 3. Functional module diagram of remote sensing target detection system with attention mechanism
图 3. 注意力机制的遥感目标检测系统功能模块图

2.2.2. 原型设计

界面上方显示系统名称、左侧划分为显示画面及位置信息区域以实时显示检测内容及目标的位置；右上包含检测参数设置及检测结果区域便于使用者对参数的调节及对检测结果的观察；底层提供了开启图片，文件夹，视频，摄像头和注意力模型的系列功能按钮，以支持使用者灵活地上传资料，发起检测，保存结果和退出系统。整个接口布局合理、操作方便，能满足用户检测遥感目标的现实需要。

3. 基于注意力机制的遥感目标检测算法

3.1. 数据集

本实验所用数据集为武汉大学于 2017 年所制作的 RSOD 遥感数据集。对于遥感数据集而言，由于其拍摄获取的方式都是由空中对地进行数据的获取，拍摄的角度、方式将使拍摄的目标特征与真实的目标特征不一致，从数据集中可以看出，这些检测的目标通常存在分布密集、大小不一，并且由于光照等因素，将导致数据的检测背景十分复杂，通过考虑数据集的大小以及设备因素。

RSOD 以 PASCALVOC 数据集的格式标注，RSOD 数据集是用于遥感图像中物体检测的数据集，包括飞机(aircraft)、油箱(oiltank)、桥(overpass)和操场(playground)。其中 655 选择张图片作为训练集，188 张图片作为验证集。

3.2. YOLO-Shuffle-MSDA 模型

在遥感图像中，目标的大小因尺度而异，尤其是在处理航拍或卫星图像时，目标的尺寸差异可能非常显著，从远距离的大型结构到小型细节目标都存在，这使得传统的检测模型很难处理这种尺度差异。而遥感图像中一般都会含有复杂而凌乱的背景如植被，建筑以及道路。该模型在此条件下易使目标和背景相混淆，针对遥感目标检测中的困难，该文提出基于注意力机制 YOLO-Shuffle-MSDA 的遥感目标检测模型，通过集成多尺度特征融合模块(C2f_MSDA)和特征增强注意力(ShuffleAttention)，成功地解决了一系列相关问题。在 C2f_MSDA 模块中集成了不同尺度下的特性，提高了模型在多尺度下的目标探测能力，降低了漏检与误检现象；ShuffleAttention 模块对特征的表达进行了加强，将目标从复杂的背景中分辨出来，加强了特征的鲁棒性。此外，该模型使用了分层清晰的架构设计方法(Backbone, Neck, Head)，通过逐步提取和整合特征，从而进一步提升了检测的准确性。通过上述创新，该模型对复杂遥感场景下的目标检测更加精确和稳健。

3.2.1. 模型整体结构

同主流目标检测网络相似，YOLO-Shuffle-MSDA 由 Backbone, Neck, Head 三部分构成。图 4 展示的那样。红色框所显示的是本论文中的改进模块。

Backbone 从卷积层(Conv)开始，并通过一系列的专用层(例如 C2f, SPPF)来提取多尺度特征图。在主干网络上增加 Shuffle Attention 帮助模型更加高效地集中在图像的关键区域以克服复杂背景带来的困难。它动态调整注意力权重，增加重要区域注意力，改善背景杂乱现象。

Neck 将特征拼接与 MSDA 模块相结合，对多尺度特征进行融合以加强模型对不同特征尺度及上下文信息的聚合。MSDA 将多尺度膨胀注意力(Multi-Scale Dilated Attention: MSDA)引入 C2f 中，用 MSDA 块替换 C2f 中的 Bottleneck。该模块的核心是采用 CSP 思想，分块处理特征，兼顾特征保留和上下文增强。使用膨胀卷积和拼接操作，增强小目标的语义信息捕获。灵活调整的 CSP 结构，既轻量化又高效。这种设计能够显著提升遥感目标检测的性能，尤其在复杂背景、目标尺度多样的小目标检测任务中表现出色。显著提升了数据集的检测精度。

Head 层接收来自 Neck 的特征，并执行目标检测任务，包括预测目标的位置和类别。

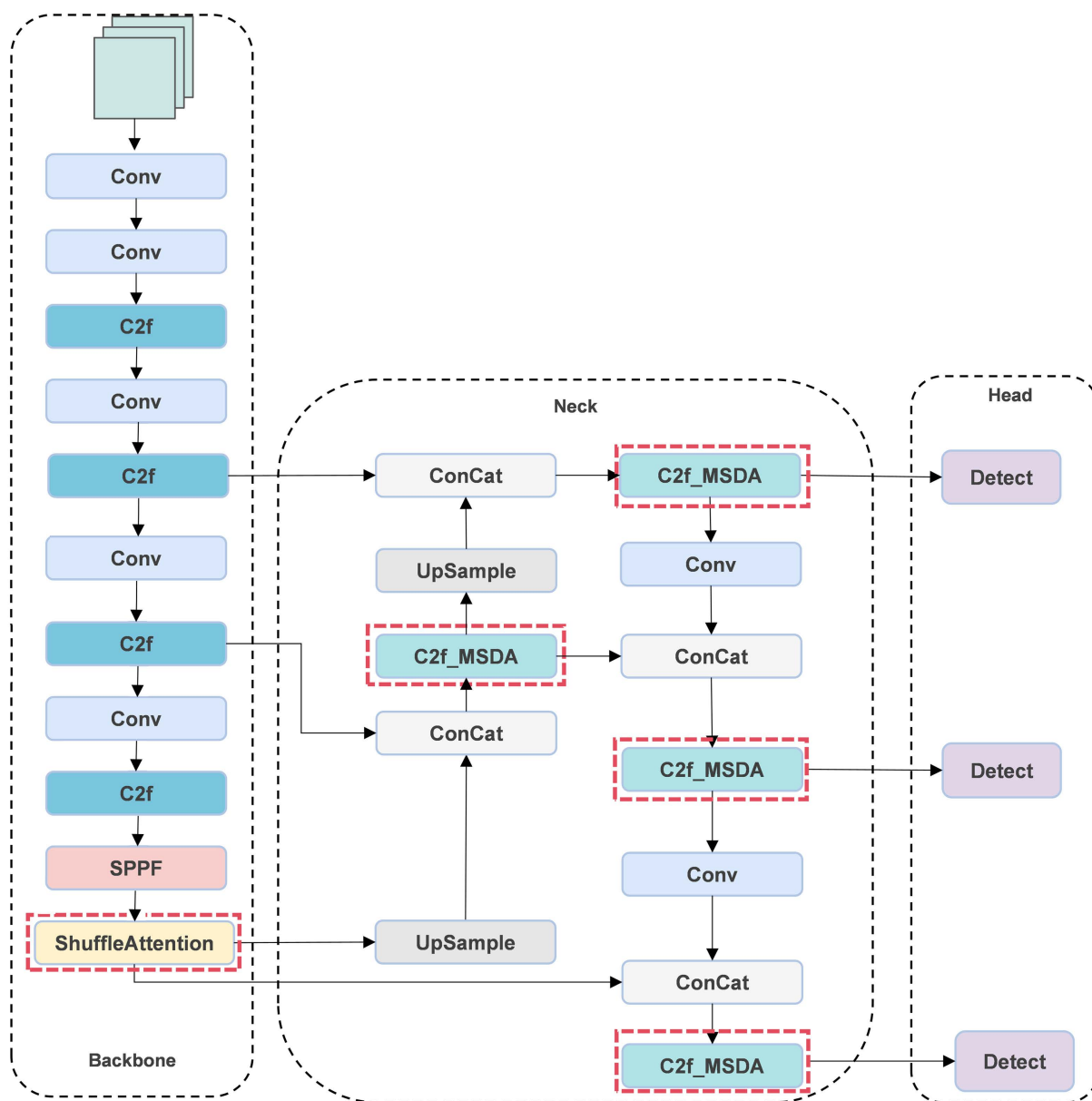


Figure 4. Overall structure diagram of YOLO-Shuffle-MSDA

图 4. YOLO-Shuffle-MSDA 整体结构图

3.2.2. 模型介绍

(1) C2f 网络结构模块:

首先通过第一个卷积层 $cv1$ 进行特征提取, 随后将得到的特征图分为两支。一支特征图直接传递到后续的拼接阶段, 而另一支则经过一系列 Bottleneck 模块的逐层处理, 以增强特征的表达能力。最终, 这两个特征图在通道维度上进行拼接, 并通过第二个卷积层 $cv2$ 进行融合与调整, 从而生成模块的最终输出。图 5 所示。

该结构旨在保持计算效率前提下加强网络的特征提取与融合能力, 在目标检测任务中显得尤为重要。该网络通过分割与拼接操作可以在不同规模下同时保存特征信息, 对不同规模目标的探测十分有用。同时采用 Bottleneck 模块有利于降低计算量并使网络更轻量化。

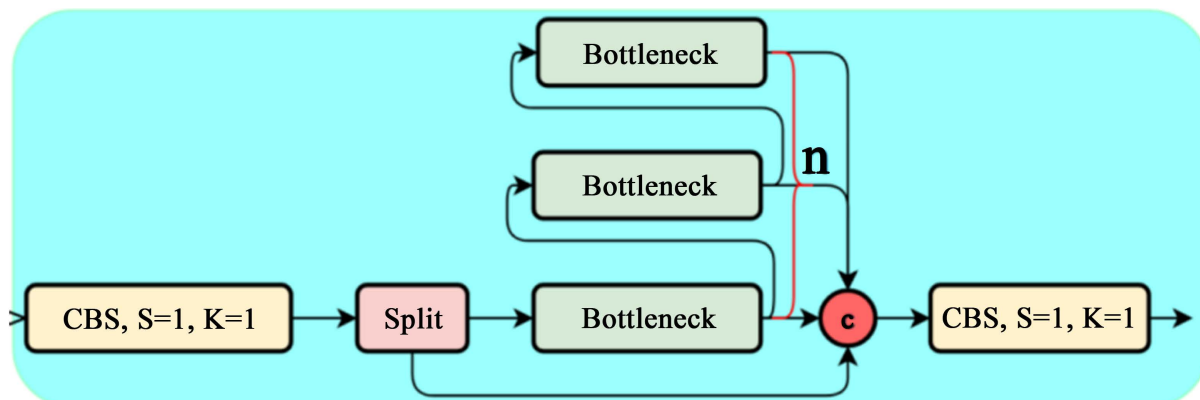


Figure 5. C2f structure diagram

图 5. C2f 结构图

(2) MSDA 结构模块:

MSDA (Multi-Scale Dilated Attention)的结构被视为一种集中注意力的机制，它利用膨胀卷积(dilated convolution)在多个尺度上捕捉特征信息。该构造尤其适合处理背景复杂，目标多尺度的遥感图像。现将 MSDA 的构造简介如下：

首先将特征贴图中通道划分为不同头。接着对不同头周边窗口彩色斑块采用不同膨胀率自注意操作。另外将不同头的特性联系起来再输入线性层中。在默认设置中，我们采用了 3×3 的核尺寸，其扩张率分别是 $r = 1, 2$ 和 3 ，而不同头部的感受野大小则是 $3 \times 3, 5 \times 5$ 和 7×7 。MSDA 结构图图 6 所示：

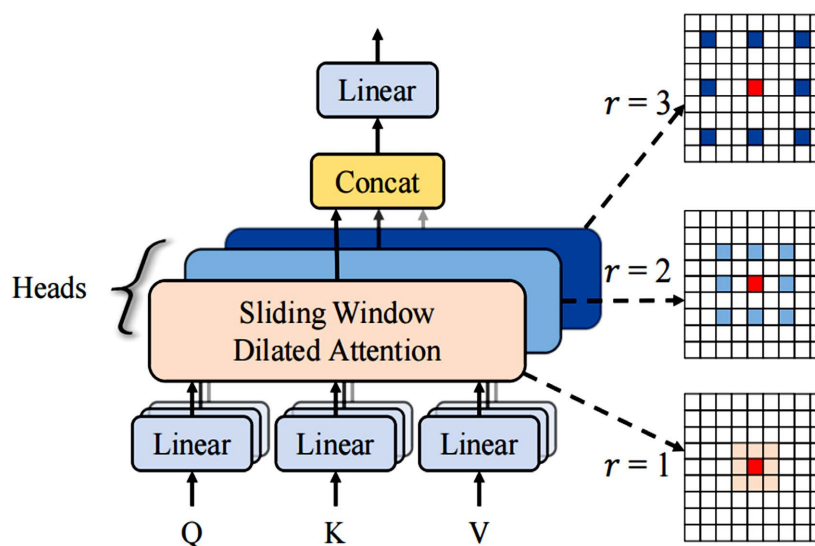


Figure 6. Structure of MSDA

图 6. MSDA 结构

在实现层面，滑动窗口膨胀注意力由深度膨胀卷积近似实现，图 7 所示。该方法以局部卷积的形式高效实现多尺度注意力计算，兼顾全局感受野扩展与计算轻量化。

(3) ShuffleAttention 模块

ShuffleAttention 模块作为注意力机制可以将通道注意力与空间注意力进行有效融合来提高深度神经网络性能。模块先对输入的特征图进行分组，采用 Shuffle 单元实现通道注意与空间注意的融合。该方法

对输入特征进行分组，注意力计算和应用，有利于模型对重要特征进行较好的关注和对不相关信息进行压制。该注意力机制加入 YOLOv8 系列，提高了模型的目标识别能力。

```
class C2f_MSDA(nn.Module):
    def __init__(self, channels, heads=3, dilation_rates=(1,2,3), se_ratio=8):
        super().__init__()
        assert channels % heads == 0
        self.heads = heads
        self.Cm = channels // heads
        self.dilations = dilation_rates
        # depthwise dilated convs
        self.dw = nn.ModuleList([
            nn.Conv2d(self.Cm, self.Cm, kernel_size=3, padding=d, dilation=d, groups=self.Cm, bias=False)
            for d in dilation_rates
        ])
        # pointwise to mix channel per head
        self.pw = nn.ModuleList([nn.Conv2d(self.Cm, self.Cm, kernel_size=1, bias=False) for _ in range(heads)])
        self.bn_pw = nn.ModuleList([nn.BatchNorm2d(self.Cm) for _ in range(heads)])
        # SE (channel attention) per head
        self.se_fc1 = nn.ModuleList([nn.Linear(self.Cm, self.Cm//se_ratio) for _ in range(heads)])
        self.se_fc2 = nn.ModuleList([nn.Linear(self.Cm//se_ratio, self.Cm) for _ in range(heads)])
        # spatial attention conv per head (keeps single-channel map)
        self.sa_conv = nn.ModuleList([nn.Conv2d(2, 1, kernel_size=3, padding=1, bias=False) for _ in range(heads)])
        # fusion 1x1
        self.fuse = nn.Conv2d(channels, channels, kernel_size=1, bias=False)
        self.out_bn = nn.BatchNorm2d(channels)
```

Figure 7. Code

图 7. 代码

(4) SPPF 模块

其被广泛地应用于目标检测任务，尤其适用于 YOLOv8 多尺度特征信息的有效抽取。SPPF 模块设计使网络能够对不同规模的特性进行处理，对于目标检测任务来说尤其有用，因为图像中的目标可能会呈现为不同规模。模块采用多尺度融合的方法捕获不同尺寸目标以提高特征提取能力。SPPF 模块采用简化版空间金字塔池化方法，实现了不同规模特征图的拼接，增强了不同规模目标检测能力。

3.3. 实验

3.3.1. 模型训练

(1) 数据集参数配置

训练数据集：data.yaml 这是 RSOD 数据集的配置文件，指定了训练和验证路径、测试集数据集的路径、数据集中类别有 4 类。类别 0 被命名为“aircraft”（飞机），类别 1 被命名为“oiltank”（油罐），类别 2 被命名为“overpass”（立交桥）；类别 3 被命名为“playground”（操场）。

(2) 模型训练

模型训练代码用于配置和启动 YOLO-Shuffle-MSDA 模型的训练过程，图 8 所示，主要功能包括模型初始化、设置训练参数以及启动训练任务。首先，通过 YOLO ('ultralytics/cfg/models/v8/yolvos-att2-MSDA.yaml') 初始化模型，并指定了模型配置文件；接着，通过 model.train() 方法配置训练参数，包括数据集路径、图像尺寸、训练轮数、批次大小、设备和优化器等。此外，代码还支持设置多线程工作、项目保存路径和实验名称等。通过这段代码，用户可以灵活调整训练配置，并在指定设备上启动 YOLO-Shuffle-MSDA 模型的训练任务。

```

if __name__ == '__main__':
    model = YOLO('ultralytics/cfg/models/v8/yolov8-att2-MSDA.yaml')
    #model.load('yolov8n.pt') # loading pretrain weights
    model.train(data='/storage/home/018023/dataset/RSOD/RSOD/data.yaml',
                cache=False,
                imgsz=640,
                epochs=300,
                batch=32,
                close_mosaic=10,
                workers=12,
                device='1',
                optimizer='SGD', # using SGD
                # resume='', # last.pt path
                # amp=False, # close amp
                # fraction=0.2,
                project='runs/train/',
                name='exp',
                )

```

Figure 8. Training code

图 8. 训练代码

(3) 训练结果

模型的训练成果展示了在模型的训练和验证阶段，多个损失函数和评估指标的变化趋势，这包括了训练和验证的 box loss、cls loss、dfl loss，以及精度(precision)、召回率(recall)、mAP50 和 mAP50-95 等指标。从图中可以看出，训练初期各项损失函数下降慢慢稳定，表明模型在快速学习并逐步收敛；同时，评估指标(如 precision、recall、mAP50 和 mAP50-95)在训练过程中稳步上升并最终趋于平稳，说明模型的性能不断提升并达到较优状态。蓝色实线代表实际结果，橙色虚线为平滑后的结果，两者对比清晰，便于观察变化趋势。整体来看，模型在训练过程中表现出良好的优化效果，各项指标的变化趋势符合预期，验证了训练方法的有效性。

3.3.2. 验证与测试

验证代码如图 9 所示：

```

import warnings
warnings.filterwarnings('ignore')
from ultralytics import YOLO

if __name__ == '__main__':
    model = YOLO('runs/train/exp7/weights/best.pt')
    model.val(data='/storage/home/018023/dataset/RSOD/RSOD/data.yaml',
             split='val',
             imgsz=640,
             batch=4,
             # iou=0.7,
             # rect=False,
             # save_json=True, # if you need to cal coco metrice
             project='runs/val',
             name='exp',
             )

```

Figure 9. Verification code

图 9. 验证代码

4. 算法评估与演示验证系统

实验结果

为综合评价本论文算法对遥感目标检测时模型的性能，对几种目标检测算法进行了比较，并在相同数据集上对各种算法进行了性能差异实验，并用表格，曲线图及检测效果的比较展示了实验的结果。

表 1 中为了比较不同 YOLOv8 变体模型在实验中的表现，我们考察了参数数量(Parameters)、每秒浮点运算次数(GFLOPs)、精度(P)、召回率(R)、mAP50 以及 mAP50-95 等多个关键指标。本文提出的 YOLO-Shuffle-MSDA 模型在多个关键指标上表现优异，尤其是在精度(P)和 mAP50-95 方面，分别达到 0.939 和 0.958，显著优于其他模型。这一成果的取得归功于该模型将 ShuffleAttention 模块与多尺度的特征融合(MSDA)技术相结合，ShuffleAttention 利用增强的特征表达，有效地分辨出目标和复杂的背景，并且 MSDA 通过多尺度特征融合，增强模型在多尺度下的目标检测。此外，该模型在参数的数量和计算的复杂性(GFLOPs)方面都维持了一个相对较低的标准，确保了性能和效率之间的均衡。所以 YOLO-Shuffle-MSDA 无论是检测精度还是整体性能都有明显的提高。

Table 1. Comparison of experimental results of different network models

表 1. 不同网络模型实验结果对比

No.	Method	Parameters	GFLOPs	P	R	mAP50	mAP50-95
1	YOLOv8	3006038	8.1	0.755	0.9	0.914	0.636
2	YOLOv8-MHSA	3203804	8.2	0.867	0.89	0.922	0.641
3	YOLOv8-ShuffleAttention	3006524	8.1	0.901	0.907	0.948	0.635
4	YOLOv8-SimAM-1	3006428	8.1	0.888	0.912	0.93	0.625
5	YOLOv8-SimAM-2	3006428	8.1	0.908	0.918	0.95	0.635
6	YOLOv8-C2f_Attention	3047892	8.3	0.938	0.909	0.944	0.645
7	YOLOv8-SEAttention	3014620	8.1	0.858	0.933	0.931	0.63
8	YOLOv8-C2f-MSDA	2649564	7.4	0.902	0.904	0.942	0.616
9	YOLO-Shuffle-MSDA (Ours)	2649468	7.4	0.933	0.939	0.958	0.628

图 10 所显示的是文中方法的检测效果图，从中可看出该模型对不同种类，尺寸及方向遥感目标特别是复杂场景下目标的探测性能，及不同光照，遮挡，背景干扰等情况下，文中方法鲁棒性较强。

5. 总结与展望

5.1. 总结

遥感目标检测是计算机视觉及遥感图像处理中的一个重点研究领域，这一技术主旨是在遥感影像上对具体目标类别进行精准定位和空间分布及几何参数信息同步进行获取。近年来，随着深度神经网络架构不断演化，这一领域的技术效能取得了实质性的突破，但是仍然需要解决许多技术瓶颈问题：目标几何尺度微小化问题、空间分布的密集型特点，目标之间的互相遮挡，多尺度的目标混杂，复杂的地表背景的干扰是难以解决的。



Figure 10. Detection results of the method in this paper on the RSOD dataset

图 10. 本文方法在 RSOD 数据集上的检测结果

5.2. 展望

本研究通过引入注意力机制对模型进行优化改进，在增强系统抗干扰能力和实时响应效率的基础上，有效提升了目标检测性能，重点攻克了遥感影像中背景干扰复杂、目标尺度动态变化及目标遮挡等技术瓶颈。

参考文献

- [1] Girshick, R., Donahue, J., Darrell, T. and Malik, J. (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014 *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 24-27 June 2014, 580-587. <https://doi.org/10.1109/cvpr.2014.81>
- [2] Cao, Y., Wang, J., Jin, Y., Wu, T., Chen, K., Liu, Z. and Lin, D. (2021) Few-Shot Object Detection via Association and Discrimination. *NeurIPS 2021*, 6-14 December 2021, 16570-16581.
- [3] Girshick, R. (2015) Fast R-CNN. 2015 *IEEE International Conference on Computer Vision (ICCV)*, Santiago, 7-13 December 2015, 1440-1448. <https://doi.org/10.1109/iccv.2015.169>
- [4] Cai, Z. and Vasconcelos, N. (2018) Cascade R-CNN: Delving into High Quality Object Detection. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 6154-6162. <https://doi.org/10.1109/cvpr.2018.00644>
- [5] 胡惠娟, 秦一锋, 徐鹤, 等. 面向无人机航拍图像的 YOLOv8 目标检测改进算法[J]. 计算机科学, 2025, 52(4): 202-211. <http://kns.cnki.net/kcms/detail/50.1075.TP.20240925.1400.013.html>, 2024-09-28.
- [6] Liu, W., Anguelov, D., Erhan, D., et al. (2015) SSD: Single Shot MultiBox Detector.

-
- [7] Gong, H., Mu, T., Li, Q., Dai, H., Li, C., He, Z., *et al.* (2022) Swin-Transformer-Enabled Yolov5 with Attention Mechanism for Small Object Detection on Satellite Images. *Remote Sensing*, **14**, Article No. 2861. <https://doi.org/10.3390/rs14122861>
 - [8] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., *et al.* (2021) Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 9992-10002. <https://doi.org/10.1109/iccv48922.2021.00986>
 - [9] Wu, W., Liu, H., Li, L., Long, Y., Wang, X., Wang, Z., *et al.* (2021) Application of Local Fully Convolutional Neural Network Combined with YOLOv5 Algorithm in Small Target Detection of Remote Sensing Image. *PLOS ONE*, **16**, e0259283. <https://doi.org/10.1371/journal.pone.0259283>
 - [10] 李坤亚, 欧鸥, 刘广滨, 等. 改进 YOLOv5 的遥感图像目标检测算法[J]. 计算机工程与应用, 2023, 59(9): 207-214.
 - [11] 王建军, 魏江, 梅少辉, 王健. 面向遥感图像小目标检测的改进 YOLOv3 算法[J]. 计算机工程与应用, 2021, 57(20): 133-141.
 - [12] Wang, H. and Han, J. (2023) Improved Military Equipment Identification Algorithm Based on YOLOv5 Framework. 2023 *IEEE 6th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, Chongqing, 24-26 February 2023, 1195-1199. <https://doi.org/10.1109/itmec56291.2023.10081984>
 - [13] Shi, P., Jiang, Q., Shi, C., Xi, J., Tao, G., Zhang, S., *et al.* (2021) Oil Well Detection via Large-Scale and High-Resolution Remote Sensing Images Based on Improved YOLOv4. *Remote Sensing*, **13**, Article No. 3243. <https://doi.org/10.3390/rs13163243>
 - [14] Zhang, X., Wang, T., Qi, J., Lu, H. and Wang, G. (2018) Progressive Attention Guided Recurrent Network for Salient Object Detection. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 714-722. <https://doi.org/10.1109/cvpr.2018.00081>
 - [15] Chen, H. and Li, Y. (2019) Three-Stream Attention-Aware Network for RGB-D Salient Object Detection. *IEEE Transactions on Image Processing*, **28**, 2825-2835. <https://doi.org/10.1109/tip.2019.2891104>