

# 基于模拟数据的古建筑火灾检测改进算法研究

江金松<sup>1,2</sup>, 史东辉<sup>1,2</sup>, 王鹏林<sup>1,2</sup>, 周思恒<sup>3</sup>, 王飞龙<sup>1,2</sup>

(1. 安徽建筑大学电子与信息工程学院, 安徽合肥 230601; 2. 安徽建筑大学智能建筑与建筑节能安徽省重点实验室, 安徽合肥 230022; 3. 安庆师范大学计算机与信息学院, 安徽安庆 246133)

**摘要:** 在古建筑火灾预防算法研究中, 数据集稀缺成为主要挑战。针对此问题, 提出了一种新的古建筑火灾检测与预防方法。构建一个虚拟平台以生成模拟火灾数据, 该平台能够模拟火焰和烟雾在古建筑中的扩散情况。在算法层面, 提出一种改进后的算法模型YOLOv5s-STI, 该模型引入了Involution卷积计算方法, 能够更加灵活地适应输入数据的特性; 模型主干部分集成了Swin Transformer, 能够有效捕捉火焰和烟雾的空间分布关系, 增强了模型对于火灾动态的识别能力; 在C3模块融入SimAM自注意力机制, 通过对特征的空间相关性进行自适应调整, 提升了模型对图像中火灾区域的识别精度。通过对仿真数据的评估实验发现, 改进模型在模拟数据集上的平均精度mAP@0.5提升了1.2个百分点, 且在实景火焰和烟雾检测中展现了良好性能, 证明了该方法在古建筑火灾预防中的有效性。

**关键词:** 火灾检测; YOLOv5s; 虚拟仿真; 注意力机制; 卷积算子; Transformer

DOI: 10.11907/rjdk.241581

开放科学(资源服务)标识码(OSID):

中图分类号: TP391.41

文献标识码: A

文章编号: 1672-7800(2025)008-0027-11



## Research on Improved Fire Detection Algorithms for Ancient Buildings Based on Simulated Data

JIANG Jinsong<sup>1,2</sup>, SHI Donghui<sup>1,2</sup>, WANG Penglin<sup>1,2</sup>, ZHOU Siheng<sup>3</sup>, WANG Feilong<sup>1,2</sup>

(1. School of Electronic and Information Engineering, Anhui Jianzhu University, Hefei 230601, China;

2. Anhui Province Key Laboratory of Intelligent Building & Building Energy Saving, Anhui Jianzhu University, Hefei 230022, China;

3. College of Computer and Information, Anqing Normal University, Anqing 246133, China)

**Abstract:** In the research on fire prevention algorithms for ancient buildings, the scarcity of datasets has become a major challenge. To address this issue, this study proposes a novel method for fire detection and prevention in ancient buildings. First, a virtual platform was developed to generate simulated fire data, which can model the spread of flames and smoke in ancient structures. On the algorithmic level, this study introduces an improved algorithm model, YOLOv5s-STI, which incorporates the Involution convolution computation method, allowing the model to more flexibly adapt to the characteristics of the input data. The backbone of the model integrates the Swin Transformer, which effectively captures the spatial distribution of flames and smoke, enhancing the model's ability to recognize fire dynamics. The C3 module is enhanced with the SimAM self-attention mechanism, which adaptively adjusts the spatial correlation of features, improving the model's accuracy in identifying fire regions within images. Evaluation experiments on the simulated dataset show that the improved model achieved a 1.2 percentage increase in mean Average Precision (mAP@0.5) and demonstrated strong performance in detecting real-world flames and smoke. These results validate the effectiveness of this method in the fire prevention of ancient buildings.

**Key Words:** fire detection; YOLOv5s; virtual simulation; attention mechanism; convolution operator; Transformer

收稿日期: 2024-09-10

扫描二维码阅读全文:



**基金项目:** 安徽省高校人文社会科学研究重点项目(2022AH050224); 智能建筑与建筑节能安徽省重点实验室主任基金项目(IBES2024ZR01); 中国建设教育协会教育教学科研课题重点项目(2023003); 安徽省研究生教育质量工程项目(2023cx-cysj129, 2024zyxwjsxalk126)

**作者简介:** 江金松(1999-), 男, CCF 学生会会员, 安徽建筑大学电子与信息工程学院硕士研究生, 研究方向为图像识别与目标检测; 史东辉(1966-), 男, 博士, CCF 会员, 安徽建筑大学电子与信息工程学院教授、硕士生导师, 研究方向为机器学习、人工智能; 王鹏林(1997-), 男, 安徽建筑大学电子与信息工程学院硕士研究生, 研究方向为人工智能、虚拟现实; 周思恒(2003-), 男, 安庆师范大学计算机与信息学院学生, 研究方向为三维建模; 王飞龙(2000-), 男, 安徽建筑大学电子与信息工程学院硕士研究生, 研究方向为大语言模型、机器学习。本文通讯作者: 史东辉。

## 0 引言

材料老化和自然灾害等因素使得古建筑存在火灾隐患,社会对古建筑文化遗产重视程度不断提高。由于古建筑火灾数据集缺乏,视觉相关检测研究进展缓慢。目前,大多数古建筑使用传统的火灾检测装置,如烟雾探测器和温度传感器,这些装置通常存在反应迟缓和误报率高等问题。基于视频的目标检测技术正成为一种更为有效的建筑火灾检测手段,该技术可以通过监控摄像头获取古建筑内部和外部的实时视频图像数据,并对其中的对象进行检测和识别。

本文提出一种新的预防古建筑火灾的方法,通过构建虚拟平台模拟古建筑发生火灾时的真实场景,获得虚拟古建筑火灾数据集。基于此数据集,提出一种新的检测算法YOLOv5s-STI,即通过自通道注意力机制SimAM<sup>[1]</sup>、使用新的卷积算子<sup>[2]</sup>替换传统的卷积计算以及在YOLOv5s主干部分增加一层Swin Transformer<sup>[3]</sup>以提高对火焰烟雾的检测效果。优化的网络架构有效地契合了古建筑等特定场景的需求,并通过实验手段证实了虚拟数据集的有效性以及模型在火灾检测精度方面的优越性。

本文主要贡献如下:

(1)虚拟仿真平台与数据获取。鉴于当前古建筑火灾数据集的不足,本研究开发了一个虚拟仿真系统,以模拟古建筑火灾实际情景。该系统不仅能够生成和采集高度逼真的火焰与烟雾数据,还提供了对火焰和烟雾过程的精确控制,进而生成多样化的数据样本,这一过程对于训练出更高精度的火灾检测模型至关重要。

(2)模型定制化改进。针对古建筑在复杂环境下的火灾检测需求,对主流的YOLOv5s模型进行了专门优化。通过集成SimAM自注意力机制,模型对火焰和烟雾特征的关注能力得到增强,从而通过聚焦图像关键区域以提升检测精确度。在模型的主干网络中引入了Swin Transformer层,采用自适应窗口技术以优化对场景空间层次结构的处理,进一步提高了模型对古建筑中火焰和烟雾独特空间分布的识别能力。此外,采用新型Involution卷积算子替代传统卷积,以更精细地处理图像局部细节,同时保持对全局信息的敏感性。该算子以其独特的结构特性,能够自适应地响应图像不同区域的需求,从而增强模型对复杂场景动态变化的捕捉能力。这3种模块的融合不仅克服了原模型在复杂环境下的局限性,而且为古建筑火灾预防和检测提供了一种创新性的解决方案。

## 1 相关工作

古建筑是珍贵的历史文化遗产,我国许多古建筑采用砖木结构,因而面临较高的火灾风险。一旦遭受火灾,不

仅会造成无法估量的文化损失,还可能对人民群众的生命财产安全构成严重威胁,因而保护古建筑免受火灾损害具有极其重要的意义。故申俊云<sup>[4]</sup>总结了古建筑火灾发生的因素和预防,为非建筑文物保护单位加入耐火极限较高的现代建筑材料、对可燃组件进行阻燃处理、增加消防设施、引入先进的烟雾报警设备等。陆万里等<sup>[5]</sup>研究了古建筑的火灾风险和蔓延规律,并进行火灾数值模拟研究,最后提出了相应的消防对策。尽管这些工作为古建筑的火灾防范提供了有价值的信息,但火灾数值模拟在模拟古建筑火灾时可能受到建筑结构和材料等多方面因素的影响,导致模型精确性受到限制。尤其是古建筑的独特结构和复杂性,可能难以完全还原,影响模拟结果准确性。刘慧等<sup>[6]</sup>基于BIM及仿真技术,通过计算机模拟火灾场景,对古建筑火灾防控技术开展研究,对古建筑的安全保护起到一定作用。但是,由于一些古建筑的特殊性或保护性质,实际数据采集非常困难,甚至不可能进行。在这种情况下,传统的建筑研究和保护工作受到限制,难以有效保护古建筑。

目前,广泛应用的古建筑火灾检测技术设备主要依赖于传感器,如温度和烟雾传感器,往往在阈值达到后才激活报警,有时这会延误扑救火灾的最佳时机。随着计算机视觉领域深度学习技术的快速进步和视频监控设备的广泛部署,基于视频的火灾监测方法开始涌现。这些方法提供了更高的识别精度,并能抵御环境变化的干扰。

传统火焰与烟雾识别主要基于特征提取,如Chen等<sup>[7]</sup>采用基于RGB色彩模型的无序度分析进行火焰检测。这类方法对简单场景下的火焰识别效果明显,但面对复杂环境时,误判和遗漏情况仍较为常见。相较之下,深度学习技术展现出更强的表征学习能力,能够捕捉火焰图像的多层次特征<sup>[8]</sup>。Muhammadk等<sup>[9]</sup>将卷积神经网络(Convolutional Neural Network, CNN)运用于视频监控系统中,实现火焰检测监控视频架构,并使用CNN进行火焰检测。石磊等<sup>[10]</sup>提出利用深度学习以提升对微小火焰的识别率,同时解决了样本不均衡问题。Bochkovskiy等<sup>[11]</sup>则通过使用优化的YOLOv4-tiny算法,有效提升了火焰烟雾检测准确性和速度,增强了模型泛化和鲁棒性。然而,深度学习技术在火焰识别领域虽有突破,但仍受限于某些因素。例如,这些技术依赖于大规模标注数据集进行模型训练,这类数据集的获取在特定场合颇具挑战。进一步地,深度学习模型对运算要求较高,可能超出了一些计算资源有限场景的可行性范围。尤其是,在不同环境和光照条件下的泛化能力有所欠缺,对复杂场景中的火焰识别准确率有待提升等。

火灾监测对于模型实时性和检测准确性提出了高要求,而YOLOv5S模型展现出了实时性与准确性之间的良好平衡,引起了研究者的关注。郑远攀等<sup>[12]</sup>对YOLOv5S进行了优化,增强了其在烟雾识别方面的能力。王一旭等<sup>[13]</sup>

通过改进YOLOv5S模型,显著提升了对火焰和烟雾的检测效率。Zhang等<sup>[14]</sup>和Wang等<sup>[15]</sup>、Ji等<sup>[16]</sup>对YOLOv5S的C3模块加入了SimAM注意力机制,一定程度上提高了精度和参数量,对火灾检测研究具有一定意义。杨国为等<sup>[29]</sup>提出了一种基于YOLOv8的多尺度特征融合轻量级火灾检测算法,该算法通过设计EDBAN模块、优化BiFPN结构以及引入LOTT检测模块,显著提升了模型检测精度、泛化能力和实时性,同时减少了模型参数量和计算负担。尽管YOLOv8模型具有较高的检测精度,但其对GPU设备的性能有一定依赖性。相比之下,YOLOv5系列模型更为成熟,对运行设备的要求较为宽松。目前,针对YOLO系列的研究主要聚焦于单一火焰或烟雾检测,且鲜有研究考虑到模型在特定应用场景中的性能表现。这表明,尽管现有研究取得了一定进展,但在火灾检测的深入应用和全面考量方面,仍存在显著提升空间。

针对现有研究中的不足,本文构建了一个模拟平台,以采集在现实环境中难以获取的古建筑火灾数据集,并基于YOLO系统框架对网络结构进行了优化,提出了基于YOLOv5s-STI的火焰检测方法。为了使模型更准确地捕捉输入序列中的关键信息,在YOLOv5s的C3模块中融入了SimAM自注意力机制<sup>[17]</sup>。传统YOLOv5s模型主要依赖于卷积层以提取特征,这种局部感受野的设计可能限制了网络对长距离依赖关系的建模能力,因而在主干网络中添加一层Swin Transformer,以更好地捕捉全局上下文信息<sup>[18-19]</sup>。为了克服固定的感受野大小、计算成本较高以及对不同尺度信息处理能力有限等问题,引入了一种新的卷积计算方法Involution<sup>[20]</sup>。总体而言,本文方法有效提升了古建筑火灾检测精度和速度,为古建筑文物保护提供了一种可行的技术途径。

## 2 古建筑火灾仿真数据集构建

古建筑火灾仿真数据集构建框架总体上划分为3个主要环节:古建筑场景设计、古建筑火灾场景生成,以及数据采集与处理。系统框架如图1所示。

具体步骤如下:首先,进行古建筑场景设计;其次,采用仿真软件对古建筑火灾场景进行模拟,以实现火灾发生动态过程的再现,涵盖火势蔓延与烟雾扩散等现象;最后,在数据采集阶段,运用深度学习模型YOLO(You Only Look Once)框架,实现实时火焰与烟雾检测。

### 2.1 古建筑场景设计

在开展古建筑建模工作之初,本文进行了广泛的文献回顾与资料搜集,整理了众多古建筑的参考图像和详尽的平面图,确保了建筑历史原貌的精确重现。利用3ds Max软件,采用基础几何体构建了古建筑的主要结构,严格依照现存古建筑的尺寸比例和结构细节进行设计。在此基础上,进一步添加了窗户、门扇和精细雕刻等细节,以增强

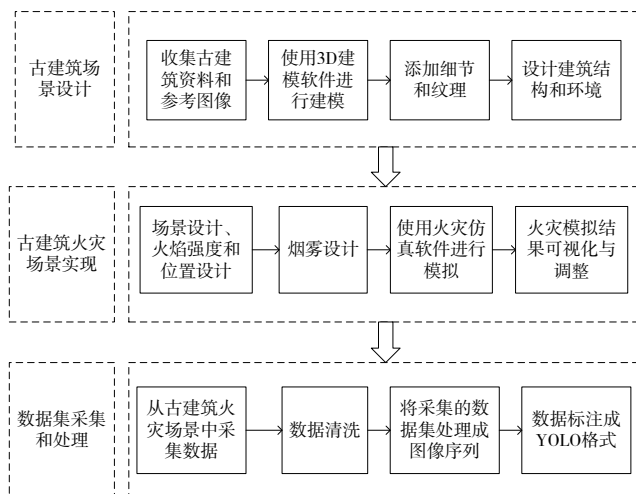


Fig. 1 System framework for the construction of an ancient building fire simulation dataset

图1 古建筑火灾仿真数据集构建系统框架

模型的视觉真实感和艺术表现。为了提高模型的逼真度,对建筑表面实施了精细的纹理映射,应用了与原始建筑材料相匹配的贴图技术。此外,模型还模拟了建筑对光照的真实反应,从而加深了视觉模拟的深度和层次感。

在场景布局方面,本文重建古建筑的空间布局并设计了与古建筑相匹配的环境设置,包括周围的山水景观和植被分布。在设计火焰和烟雾效果时,选用了高对比的明亮黄色和深黑色,以更逼真地模拟古建筑火灾的真实性。

通过上述建模流程,成功地重现了古建筑模型,并验证了建模工具的精确度。通过与实际古建筑的对比分析,本文设计的模型在形态和纹理细节上展现出了极高的相似性,为后续火灾模拟研究奠定了基础,保障了实验数据质量和研究成果可靠性。图2展示了模型结构设计细节,图3呈现了建模完成后的视觉效果与实际古建筑的对比。



Fig. 2 The ancient building modeling process

图2 古建筑建模过程

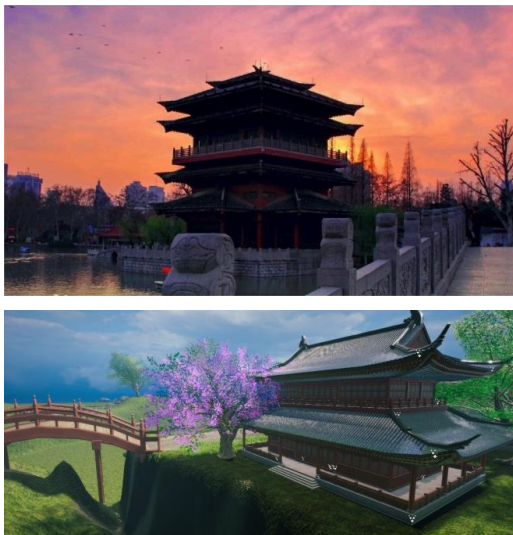


Fig. 3 Comparison of ancient building modeling with real building

图3 古建筑建模与现实建筑比较

### 2.2 古建筑火灾场景实现

采用3ds Max软件对古建筑场景进行了精细的建模工作,旨在构建一个高度真实的背景环境。借助先进的建模工具、纹理编辑和UV映射技术,确保了古建筑的结构细节和装饰特性得以精确复现,从而为模型赋予了浓郁的历史氛围和高度的逼真性。经过优化处理,模型被导入Unity引擎,其细节和质感得到保留,确保了在模拟平台中的展示效果与真实感。

此外,为增强模拟平台的紧迫性和真实性,在Unity中利用粒子系统设计了逼真的火焰和烟雾效果,如图4所示。通过调整发射器的属性和粒子的颜色、大小、透明度及生命周期,成功模拟了火焰升腾、烟雾消散的动态过程,提升了古建筑火灾情景的视觉真实感。

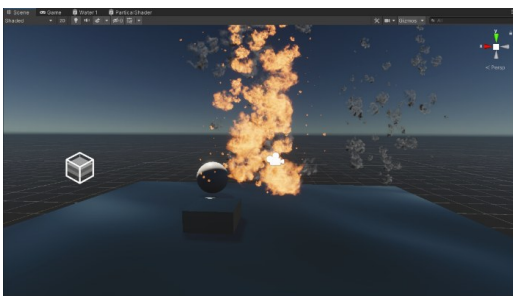


Fig. 4 Design effect of flames and smoke

图4 火焰与烟雾设计效果

在完成3ds Max的建模过程后,将古建筑模型以FBX格式从3ds Max导出并导入Unity,以保证几何结构、材质、纹理及动画信息的一致性。导入Unity后,对模型构造和纹理显示进行了细致检查,并通过定位、旋转和缩放操作对模型位置进行了精确调整。此外,为增强视觉效果,对场景光照进行了优化。为了提升运行效率,还对网络进行了简化并合并了纹理,从而成功地在Unity中实现了模型的完整集成。

### 2.3 数据集采集与处理

本文旨在通过高质量数据集的采集,为模拟平台提供真实场景的输入数据,从而提升火灾检测与预防算法的性能。在构建高效且精确的仿真平台以服务于古建筑火灾预防研究的过程中,数据采集环节发挥了至关重要的作用。为此,利用Unity引擎制作并部署了一套高分辨率摄像头漫游系统,以全方位记录建筑模型以及火焰和烟雾的动态变化。

通过该漫游系统,实现了对整个古建筑的全方位环绕拍摄,捕捉到火焰和烟雾在建筑不同视角下的动态特性。这些视频数据经过进一步处理,转化为细致的图像序列,以便输入到算法模型中,图5展示了处理后的序列图片。在代码转换过程中,注重保持数据质量,确保图像序列的清晰性和连贯性。对采集到的数据进行了彻底清理,剔除可能影响模型训练和场景渲染质量的瑕疵,从而确保数据集的准确性和可靠性。

此外,本文还对数据进行了精细的标注工作,特别关注火焰和烟雾的关键位置。这些标注包括但不限于火焰中心点、烟雾密度分布以及火焰和烟雾的动态变化。这些标注信息对于后续模型分析和效果模拟至关重要,为训练算法提供了必要参考。火焰和烟雾的关键位置标注不仅使得数据集的应用更加灵活和高效,还提高了模型在模拟古建筑场景中火焰和烟雾效果时的准确性和真实感。结合大量历史火灾影像资料可知,本文设计的仿真场景与实际情况高度一致,为后续算法开发和评估提供了坚实的数据支持。

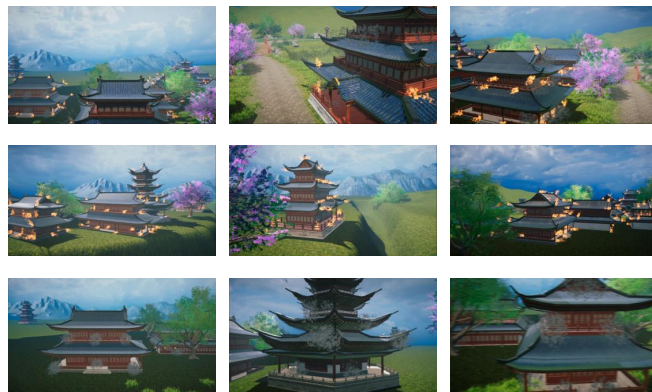


Fig. 5 Results after processing video into image sequences

图5 处理视频为图像序列后的结果

## 3 YOLOv5s-STI算法

YOLOv5s输入部分是640×640大小的图片,骨干部分主要使用CSP(Cross Stage Partial networks)模块和CSP-Darknet进行特征提取,颈部部分使用FPN(Feature Pyramid Network)和PANet(Path Aggregation Network)以聚合图像特征,预测部分用于输出检测结果<sup>[12-15]</sup>。

YOLOv5s-STI网络结构如图6所示,其在YOLOv5s的

C3 模块中融入了 SimAM 自注意力机制, 目的是综合考虑空间和通道维度, 这与模拟火灾数据的复杂特性紧密对应。本文构建的模拟数据集包含了多样化的火灾场景, 其中火焰和烟雾在空间和通道维度上展现出了复杂的相互作用, SimAM 机制能够更精确地提取这些关键特征。此外, 在模型的主干网络中引入了 Swin Transformer 层, 增强了模型对多尺度特征的识别能力。同时, 将传统的卷积算子替换为 Involution 超算子, 以更高效地处理这些多样化的

特征。选择将这些方法相结合, 旨在应对火灾场景中火焰与烟雾的高度动态性和复杂性。SimAM 机制首先确保模型能够聚焦于火灾场景中的关键特征, Swin Transformer 则有效地捕捉了全局和局部的空间分布特征, 而 Involution 超算子则为模型提供了灵活的特征变换能力, 使其能够自适应地处理多样化的输入数据。这种组合方式不仅提升了模型鲁棒性和精度, 还增强了模型在应对复杂火灾场景中的适应性和灵活性。

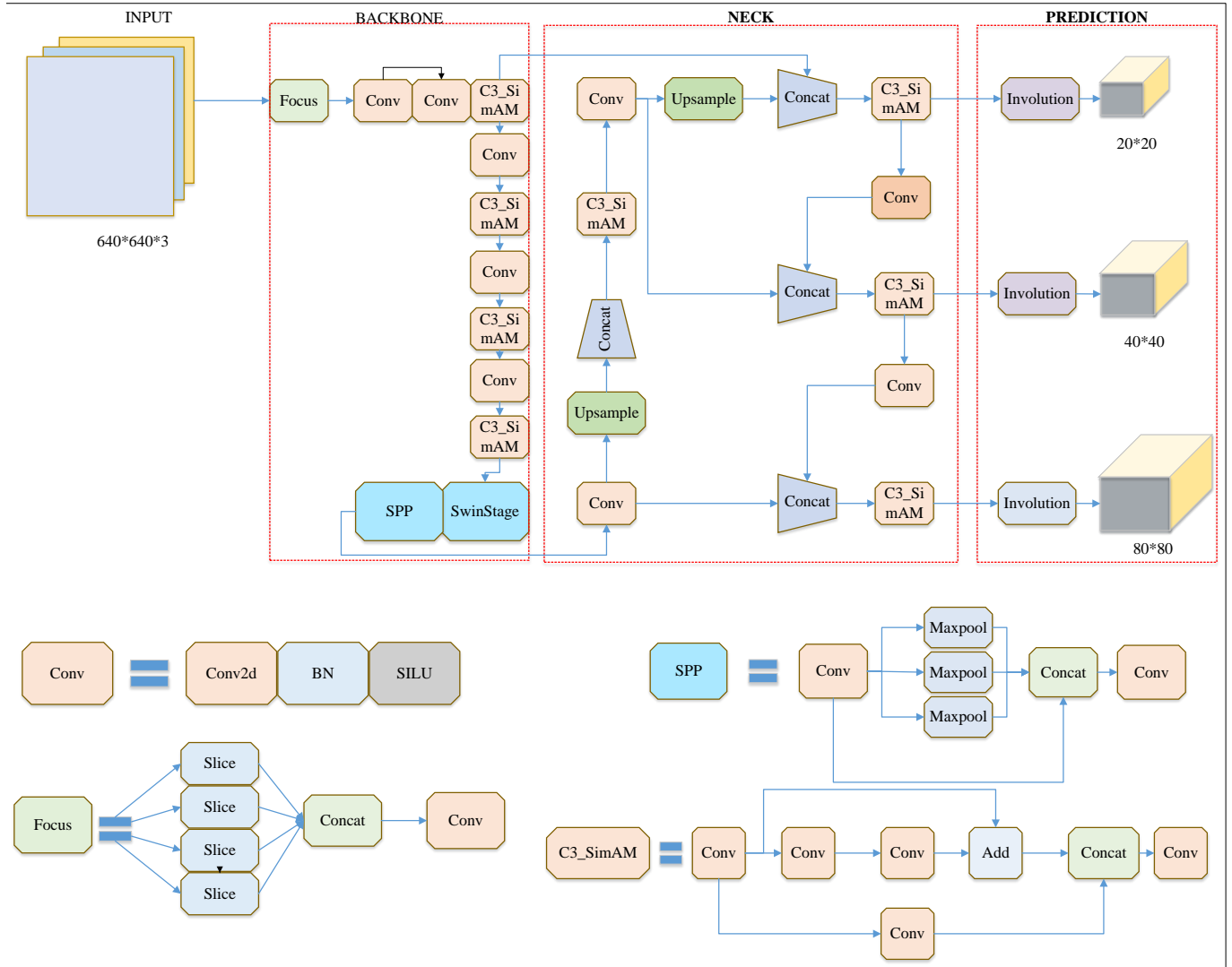


Fig. 6 YOLOv5s-STI network structure

图6 YOLOv5s-STI网络结构

### 3.1 SimAM注意力机制

现有的注意力模块通常集成在每个网络块中, 用于改进前一层的输出。这些改进可以在通道或空间维度上进行, 生成一维或二维的权重, 均等处理各通道或空间位置的神经元。传统的注意力机制分为两种: 一维通道注意力 (1D Attention) 和二维空间注意力 (2D Attention)。一维通道注意力区别对待各通道, 但对所有空间位置的神经元处理一致; 二维空间注意力则区别对待各空间位置, 但对所有通道中的神经元处理一致。

传统的注意力机制可能限制了模型对更多区分性线索的学习能力。相比之下, 三维权重注意力机制表现更为卓越, 这是因为该机制能够同时考虑通道和空间两个维度, 从而提高了学习的全面性和准确性。通过引入三维权重的 SimAM 注意力机制 (见图 7), 模型可以更准确地调整输入特征的权重, 突出重要的通道和空间位置, 增强对关键特征的感知和识别能力。普通注意力模块如 BAM 和 CBAM 尽管融合了空间与通道注意力, 但未能精确模拟其相互协作效应。神经科学研究表明, 信息量大的神经元会

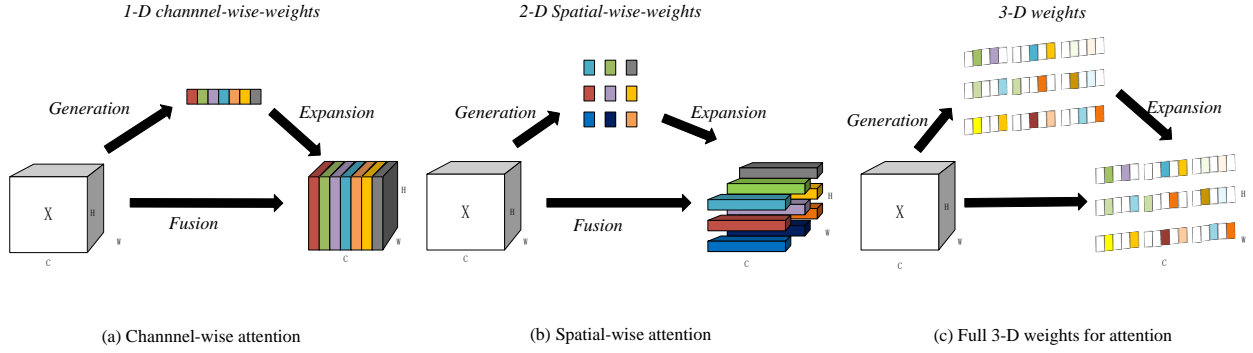


Fig. 7 SimAM attention mechanism structure

图7 SimAM注意力机制结构

显示出空间抑制效应,应获得更高权重。SimAM通过能量评价函数量化神经元的重要性,实现通道与空间信息的优化整合,增强模型对关键特征的聚焦与识别。引入这种统一权值的注意力机制后,本期望提升模型在各种任务中的性能,使其更贴近人脑的工作方式。能量评价函数如式(1)所示。

$$e_i(\omega_i, b_i, y, x_i) = (y_i - t)^2 + \frac{1}{M-1} \sum_{i=1}^{M-1} (y_0 - \hat{x}_i)^2 \quad (1)$$

这里的  $\hat{t} = w_t t + b_t$  和  $\hat{x}_i = w_i x_i + b_i$  是  $t$  和  $x_i$  的线性变换,其中  $t$  和  $x_i$  是输入特征  $X \in \mathbb{R}^{C \times H \times W}$  中的目标神经元和其他神经元,  $i$  是空间维度的索引,  $M = H \times W$  是该通道上的神经元数量,  $\omega_i$  和  $b_i$  分别表示模型的权重和偏置参数。当  $t$  等于  $y_i$  且所有其他  $\hat{x}$  等于  $y_0$  时,式(1)得到最小值,其中  $y_i$  和  $y_0$  是两个不同的值。最小化式(1)相当于增强同一通道内神经元与其他神经元的线性可分性。为了简化过程,使用二值标签并引入正则化项,详细的能量评价函数定义如式(2)所示。

$$e_i(\omega_i, b_i, y, x_i) = \frac{1}{M-1} \sum_{i=1}^{M-1} \left( -1 - (\omega_i x_i + b_i)^2 \right) + \left( 1 - (\omega_i x_i + b_i)^2 \right) + \lambda \omega_i^2 \quad (2)$$

式(2)中  $\omega_i$  和  $b_i$  的计算如式(3)和式(4)所示。

$$\omega_i = -\frac{2(t - \mu_i)}{(t - \mu_i)^2 + 2\sigma_i^2 + 2\lambda} \quad (3)$$

$$b_i = -\frac{1}{2}(t + \mu_i)\omega_i \quad (4)$$

其中,  $\lambda$  表示正则化参数,用于平衡预测误差和正则化项的影响。这里的  $\hat{\mu} = \frac{1}{M} \sum_{i=1}^M x_i$  和  $\hat{\sigma}^2 = \frac{1}{M} \sum_{i=1}^M (x_i - \hat{\mu})^2$  分别代表神经元  $t$  所在通道中除去神经元  $t$  后所有其他神经元的均值和方差。此计算是在单个通道内进行,即假设每个通道内的神经元激活值服从相同分布。

由于每个通道上的所有神经元都遵循相同分布,可以对所有神经元计算平均值和方差,并为该通道上的所有神经元重用,避免为每个位置反复计算  $\hat{\mu}$  和  $\hat{\sigma}^2$ 。最小能量  $e_i^*$  计算如式(5)所示,其表明能量越低,神经元  $t$  与周围神经元的区别越大,重要性也越高。

$$e_i^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \hat{\mu})^2 + 2\hat{\sigma}^2 + 2\lambda} \quad (5)$$

基于能量函数确定每个神经元的重要性,这种做法受到哺乳动物大脑中注意力调节机制启发,通常表现为神经反应的增益(或称为放大)效果。因此,与常见的添加操作不同,选择使用缩放操作以进行特征细化。注意力机制特征细化的定义可以表示为如式(6)所示。

$$\tilde{X} = \text{sigmoid}\left(\frac{1}{E}\right) \odot X \quad (6)$$

式(6)用于描述整个特征细化过程,其中  $E$  包含了所有通道和空间维度的  $e_i$  (从能量函数得到的值)。使用 Sigmoid 函数以限制  $E$  中的值过大,保证了处理过程中每个神经元的相对重要性不会受到影响,因为 Sigmoid 是单调函数。

### 3.2 Swin Transformer

Swin Transformer (Shifted Window Transformer) 是一种专为计算机视觉任务而设计的深度学习模型,其关键创新在于引入了移动窗口机制,使模型能够在局部区域内捕捉细节信息,同时保持对全局上下文的理解。传统的 YOLOv5s 主要依靠卷积层提取特征,局部感受野的设计限制了网络对长距离依赖的建模能力。通过引入自注意力机制, Swin Transformer 可以捕捉全局上下文信息,更好地处理长距离依赖性,提升对目标间关系的建模能力。YOLOv5s 中的特征融合依赖多尺度特征金字塔和跨层连接,而 Swin Transformer 的多层结构能够自适应地捕捉不同层次的特征,通过跨层自注意力机制实现高质量的跨尺度特征融合,增强模型对目标的多尺度感知能力。

如图8所展示, Swin Transformer 的微型版本(Swin-T)起始于一个补丁分割模块,该模块将输入的 RGB 图像分割为非重叠的补丁。这些补丁被当作单独的“Token”,其特征定义为原始 RGB 像素值的串联。具体而言,每个补丁采用  $4 \times 4$  的大小,导致每个补丁的特征维数为  $4 \times 4 \times 3 = 48$ 。随后,这些补丁特征通过一个线性嵌入层被映射到指定的维数  $C$ 。接下来,在这些标记上施加若干个改进自注意力机制的 Transformer 块(Swin Transformer 块),这些块保持了标记总数不变,定义了“Stage 1”。为了创建层次化的表示,

网络逐渐减少标记数量。第一个补丁合并层通过将每个  $2 \times 2$  的相邻补丁的特征进行串联, 并通过一个维度为  $4C$  的线性层进行处理, 从而将标记数量减少了  $1/4$  (分辨率减半至  $2C$ )。此后, Swin Transformer 块继续对特征进行变换, 保持  $H \times 8 \times W \times 8$  的分辨率, 这定义了“Stage 2”。通过重复此过程, 形成了“Stage 3”和“Stage 4”, 输出分辨率分别为

$H \times 16 \times W \times 16$  和  $H \times 32 \times W \times 32$ 。通过这种层次化和渐进式的补丁合并与特征转换过程, Swin Transformer 架构有效地实现了图像的综合建模与特征提取, 提供了不同分辨率的特征表征。Swin Transformer 架构通过这种分层结构可以对图像进行更全面的建模和特征提取。

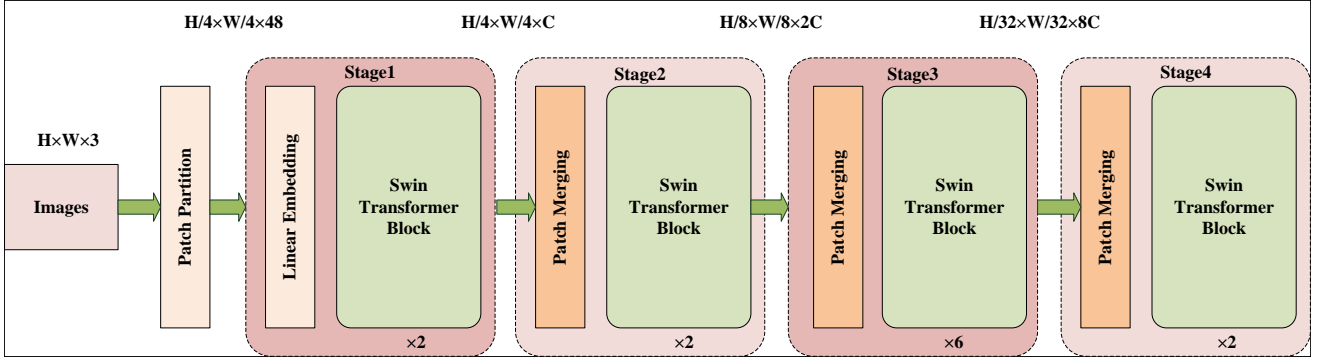


Fig. 8 Swin Transformer structure

图 8 Swin Transformer 的结构

### 3.3 Involution 超算子

传统的卷积操作在处理图像时存在一些限制, 如固定的感受野大小、计算开销较大以及对不同尺度的信息处理能力有限等问题<sup>[27-28]</sup>。为了解决这些问题, 采用了 Involution 超算子, 其核心思想是通过学习动态的卷积核以适应不同尺度的特征。

为了确保内卷定义的自足性, 首先明确标准卷积操作的概念。将输入特征表示为  $X$ , 其高度、宽度和通道数分别用  $H$ 、 $W$  和  $C$  表示。在特征张量  $X$  中, 位于图像单元的每个特征向量  $X_{i,j}$  可以被理解为代表一种复杂语义模式的像素。在这些图像点阵单元中, 每个特征向量都承载着高级语义信息。采用一个具有固定核大小的卷积滤波器  $\Phi$ , 每个滤波器包括  $C_i$  个卷积核, 这些核通过滑动窗口技术对输入特征进行乘加操作, 从而产生输出特征映射  $Y$ , 具体映射过程定义如式(7)所示。

$$Y_{i,j,k} = \sum_{c=1}^{C_i} \sum_{(u,v) \in \Delta_k} \mathcal{F}_{k,c,u+\lfloor \frac{K}{2} \rfloor, v+\lfloor \frac{K}{2} \rfloor} X_{i+u, j+v, c} \quad (7)$$

其中,  $Y_{i,j,k}$  表示第  $k$  个特征图在位置  $(i, j)$  处的输出值。 $\sum_{c=1}^{C_i} \sum_{(u,v) \in \Delta_k}$  表示对通道维度进行求和,  $c$  的取值范围是从 1 到输入通道的总数  $C_i$ , 同时在滤波器的空间范围内进行求和,  $u$  和  $v$  的取值范围在  $\Delta_k$  集合中。该集合通常表示滤波器核的索引, 通常是一个大小为  $K \times K$  的方阵。

$\mathcal{F}_{k,c,u+\lfloor \frac{K}{2} \rfloor, v+\lfloor \frac{K}{2} \rfloor}$  表示在位置  $(k, c, u + \lfloor \frac{K}{2} \rfloor, v + \lfloor \frac{K}{2} \rfloor)$  处的权重 (或滤波器系数), 该权重用于与输入数据进行卷积运算。 $X_{i+u, j+v, c}$  表示第  $c$  个通道在位置  $(i+u, j+v)$  处的输入值。

其中,  $\Delta_k$  为考虑对中心像素进行卷积的邻域偏移量集合, 计算原理如式(8)所示 (这里的  $\times$  表示笛卡尔积)。

$$\Delta_k = \left[ -\left\lfloor \frac{K}{2} \right\rfloor, \dots, \left\lfloor \frac{K}{2} \right\rfloor \right] \times \left[ -\left\lfloor \frac{K}{2} \right\rfloor, \dots, \left\lfloor \frac{K}{2} \right\rfloor \right] \quad (8)$$

与传统的深度卷积相比, 卷积核之所以称为 Involution, 是因为它包括在空间和通道域中处理逆特性的变换<sup>[29]</sup>。在 Involution 的结构设计中, Involution 核特别为每个特定坐标的像素量身定制, 而在通道维度上实现共享。此外, 计算过程包括确定每个组中共享相同 Involution 核的数量。通过对输入使用 Involution 核进行乘加计算, 得到的输出特征图如式(9)所示。

$$Y_{i,j,k} = \sum_{(u,v) \in \Delta_k} \mathcal{H}_{i,j,u+\lfloor \frac{K}{2} \rfloor, v+\lfloor \frac{K}{2} \rfloor, \lfloor \frac{KC}{C} \rfloor} X_{i+u, j+v, k} \quad (9)$$

不同于传统的卷积核, Involution 核  $H$  的构形是由输入特征映射  $X$  的维度所决定。其设计目的是在原始输入张量的条件下生成 Involution 核, 以确保输出核与输入核相匹配。核生成过程中, 将生成操作符表示为  $\phi$ , 并将位于每个位置  $(i, j)$  的函数映射抽象为如式(10)所示。

$$\mathcal{H}_{i,j} = \phi(X_{\Psi_{i,j}}) \quad (10)$$

其中,  $\Psi_{i,j}$  是以  $\mathcal{H}_{i,j}$  为条件的像素集。具体的 Involution 示意图如图 9 所示。

图 9 中的 Involution 核  $\mathcal{H}_{i,j} \in R_V$  (为了便于演示, 本示例中  $G=1$ ) 由函数  $\phi$  产生, 该函数以  $(i, j)$  处的单个像素为条件, 然后进行通道到空间的重新排列。对核的乘加运算分解为两个步骤: 乘积运算为将  $C$  个通道的张量分别与内卷核  $H$  相乘, 相加运算则是将内卷核范围内的元素相加于内卷核核心。

## 4 实验与结果分析

实验对 4 种目标检测模型 YOLOv4-tiny、YOLOv5s、YOLOv8s 和 YOLOv5s-STI 进行了全面的性能评估。评估

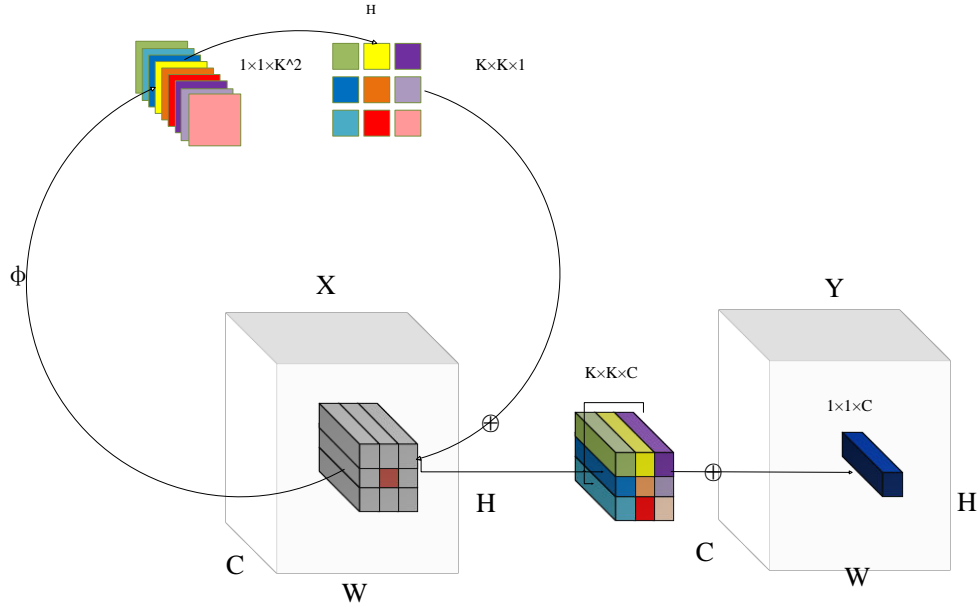


Fig. 9 Specific example of Involution  
图9 Involution 具体示例

工作在模拟数据集上展开,依据精确率、召回率、平均精度均值(mAP)和处理速度等指标对模型性能进行了量化。在消融实验中,逐步向YOLOv5s模型引入SimAM、Swin Transformer和Involution模块,以探究这些模块对模型性能的具体影响。进一步地,通过在3个公开的现实场景火焰和烟雾数据集上进行实验,对比了不同模型在真实火焰和烟雾数据上的表现,从而验证了它们在实际应用中的可行性。

#### 4.1 环境与数据集

在训练过程中,模型超参数设置如表1所示。实验使用的设备环境为Linux系统,CPU为12th Gen Intel(R) Core(TM) i9-13900F@2.50 GHz,GPU为RTX3090,24 G显存。

鉴于古建筑火灾数据集的稀缺,本文采用了通过模拟数据平台采集的火焰和烟雾数据集。为了尽可能地模拟现实场景,数据集涵盖了3种不同状态下的古建筑火灾场景:①仅有火焰的场景;②仅有烟雾的场景;③火焰与烟雾并存的场景。数据集标注工作是通过Labelimg软件完成。考虑到YOLOv5s网络结构对输入图像尺寸的固定要求,所有输入图像均被统一调整至640×640像素。实验中,数据图片采用人工标注的方法,分别设置了火焰和烟雾两个类别标签。

该数据集包含总计1500张图片,其中1200张用作训练集,150张用作测试集,另外150张用于验证集。在第

Table 1 Model hyperparameter configuration

表1 模型超参数设置

Parameters	Value
Learning rate	1.00E-05
Weight decay	1.00E-04
Total epochs	100
Btatch size	16

4.3.3节的实验中,所使用的OpenData1、OpenData2、OpenData3数据集均来源于百度公开标注的火焰烟雾数据集。具体数据集链接分别为<https://aistudio.baidu.com/datasetdetail/118365>、<https://aistudio.baidu.com/datasetdetail/105341>以及<https://aistudio.baidu.com/datasetdetail/120467>。OpenData1数据集包含了6908张不同场景下的火焰和烟雾图像,OpenData2数据集则包含了492张火焰和烟雾图像,而OpenData3数据集完全由烟雾图像组成,共计1021张,所有这些数据均经过了精确标注。为了确保实验一致性和结果可靠性,在进行实景火焰与烟雾检测的实验验证时,采用了与模拟数据集相同的划分方法。

#### 4.2 评价指标

在YOLOv5S的预测过程中,采用了广义交并比(Generalized Intersection over Union,GIoU)作为边界框的损失函数,以优化预测框的定位准确性。GIoU损失的计算公式如式(11)和式(12)所示。

$$L_{GIoU} = 1 - IoU + \frac{|C - (B \cup B^{gt})|}{|C|} \quad (11)$$

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (12)$$

其中, $IoU = \frac{|A \cap B|}{|A \cup B|}$ 表示预测框 $B = (x, y, w, h)$ 是预测框与真实框 $B^{gt} = (x^{gt}, y^{gt}, w^{gt}, h^{gt})$ 的交并比。 $C$ 是同时覆盖 $B$ 和 $B^{gt}$ 的最小闭合框。

为了进一步提高检测性能,YOLOv5s还采用了非极大值抑制(NMS)以过滤重叠的预测框,确保每个对象只被检测一次。本文通过使用召回率(Recall)、精准率(Precision)、平均精度AP(Average Precision)和平均精度均值MAP(Mean Average Precision)以量化模型准确性。评价指

标计算公式如式(13)—式(16)所示。

$$Precision = \frac{TP}{TP + FP} \quad (13)$$

$$Recall = \frac{TP}{TP + FN} \quad (14)$$

$$AP = \int_0^1 p(r) dr \quad (15)$$

$$mAP = \frac{1}{m} \sum_{i=1}^m AP_i \quad (16)$$

其中,  $m$  表示样本类别数;  $p(r)$  表示 Precision 以 recall 为参数的一个函数; TP (True Positives) 表示被正确识别的正样本; TN (True Negatives) 表示未被正确识别的负样本; FP (False Positives) 表示负样本被错误识别为正样本; FN (False Negatives) 表示正样本被错误识别为负样本。AP 值是精准率-召回率曲线下面积的近似值。

### 4.3 实验结果分析

#### 4.3.1 对比实验

为了验证本文方法在模拟古建筑火焰和烟雾检测任务中的优越性, 在模拟数据集上对该方法进行了详尽的性能评估, 并将其与流行的检测模型 YOLOv5s、YOLOv8s 和 YOLOv4-tiny 进行了比较。评估过程中, 采用了多个关键性能指标, 包括模型参数量、平均精度 (mAP) 和每秒处理帧数 (FPS), 以全面评估各模型检测性能和运行效率。图 10 展示了改进模型在训练过程中精度与损失随迭代次数的变化趋势, 表 2 则详细列出了不同模型在模拟数据集上的性能对比结果。

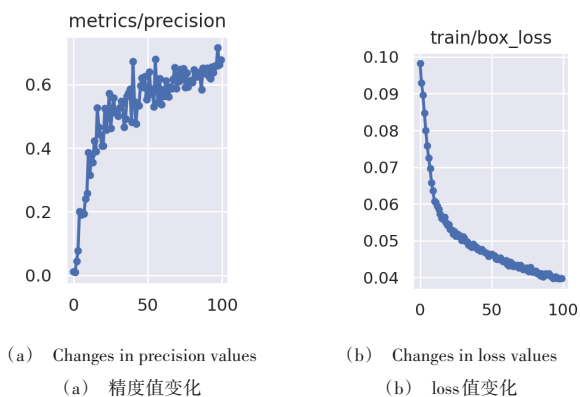


Fig. 10 Changes in data during model training

图 10 模型训练过程数据变化

由图 10 可知, 从训练开始 (Epochs=0), 精度迅速从接近 0 开始上升, 而训练损失则快速下降, 表明模型逐渐开始有效地学习图像特征。在训练达到约 50 次迭代时, 精度曲线逐渐趋于稳定, 波动显著减少。此时, 损失值已降至约 0.04, 下降速度有所减缓。到训练次数达到 100 左右时, 精度曲线稳定在 0.66 左右, 而训练损失则稳定在约 0.04。整体来看, 改进后的模型在训练过程中表现稳定, 具备良好的学习能力和较高的性能。

表 2 数据显示了在模拟数据集上对火焰和烟雾检测

Table 2 Comparison of test results of algorithms on simulated datasets

表 2 模拟数据集检测算法测试结果比较

Dataset	Models	$P$	$R$	$mAP$	速度/帧· $S^{-1}$	模型参 数/M
Our simulated datasets	YOLOv4-tiny	0.61	0.574	0.692	38	6.4
	YOLOv5S	0.693	0.514	0.696	98	7.01
	YOLOv8s	0.688	0.585	0.705	86	11.2
	YOLOv5s-STI	0.668	0.59	0.708	78	13.6

性能的评估结果, 4 种算法的实验结果揭示了它们在不同性能指标上的表现差异。YOLOv4-tiny 模型在维持较高精确度的同时, 提高了召回率, 显示出了较好的性能平衡, 特别是在小目标和快速移动目标的检测上具有优势, 尽管其精确度尚有提升空间。YOLOv5s 模型虽然在精确度上表现最佳, 但召回率最低, 意味着该模型在准确识别火灾事件的同时, 遗漏了一定数量的火灾案例, 反映出其在优化精确度时牺牲了一定的召回率。YOLOv5s-STI 模型在保持高精确度的同时, 显著提升了召回率和 mAP, 与 YOLOv8s 模型相当, 且改进后的模型在检测速度上稍胜一筹。图 11 展示了 YOLOv5s-STI 模型在自建模拟数据集上对火焰和烟雾的检测效果。总体而言, 所提出的 YOLOv5s-STI 模型在火焰和烟雾检测任务中展现了卓越性能, 有效提升了模型在复杂场景下的适应性和识别能力。



Fig. 11 Detection performance of YOLOv5s-STI on simulated data for flames and smoke

图 11 YOLOv5s-STI 在自建模拟数据集上的火焰和烟雾检测效果

#### 4.3.2 消融实验

在相同的实验环境下, 将本文提出的 YOLOv5s-STI 算法进行消融实验对比分析, 算法实验对比结果如表 3 所示。

Table 3 Comparison of ablation experiments on simulated datasets

表 3 模拟数据集消融实验对比

Dataset	Models	$P$	$R$	$mAP$	速度/帧· $S^{-1}$
Our Simu- lated Data	YOLOv5s	0.693	0.514	0.696	98
	YOLOv5s-SimAM	0.672	0.525	0.679	83
	YOLOv5s-SimAM-Swin	0.643	0.562	0.699	79
	Transformer	0.668	0.59	0.708	78
	YOLOv5s-SimAM-Swin Transformer-Involution	0.668	0.59	0.708	78

此次消融实验逐步将 SimAM 自注意力机制、Swin Transformer 和 Involution 方法集成至 YOLOv5s 模型,以评估这 3 个模块对模型性能的具体影响。由实验结果可知,集成 SimAM 自注意力机制后,模型精确度略有下降,而召回率从 0.514 提升至 0.525, mAP 从 0.696 降至 0.679。这表明 SimAM 在增强召回率的同时,以一定程度的精确度下降为代价。SimAM 的加入有助于检测到更多目标,但也可能引入额外的误检。引入 Swin Transformer 后,模型召回率显著提高,而精确度和 mAP 略有下降,表明 Swin Transformer 增强了模型对复杂目标的泛化能力和识别能力,这可能是由于 Swin Transformer 在处理全局特征时,对目标边界的识别能力得到了加强。最终,结合 Involution 方法,模型在召回率(0.590)和 mAP(0.708)上均实现了显著提升,同时精确度回升至 0.668,检测速度略下降至 78 帧/s。Involution 的集成显著增强了模型在复杂场景下的检测能力,尤其是在提高召回率和整体检测准确性方面的效果尤为突出。实验结果证实,通过逐步集成 SimAM、Swin Transformer 和 Involution 模块,模型整体性能得到了显著增强,尤其是在召回率和 mAP 指标上有所改善。这些结构性的改进有效提升了 YOLOv5s 模型在目标检测任务中的性能,实验结果验证了改进方法的有效性和适用性。

#### 4.3.3 实景火焰与烟雾检测实验验证

鉴于目前尚未有公开的古建筑火灾数据集,为了验证改进算法在实际火焰和烟雾检测中的有效性,采用了普通的公开火焰和烟雾数据集 OpenData1、OpenData2、OpenData3 进行方法可行性验证。使用改进后的模型对实际公开的火烟数据集进行检测,其验证精度结果如表 4 所示。

**Table 4 The results of different models on public fire and smoke datasets**

表 4 不同模型在公开火焰和烟雾数据集上的结果

Dataset	Models	P	R	mAP	速度/帧·S <sup>-1</sup>
Opendata1	YOLOv4-tiny	0.74	0.735	0.721	38
	YOLOv5S	0.804	0.792	0.789	98
	YOLOv8s	0.828	0.796	0.814	89
	YOLOv5s-STI	0.825	0.811	0.809	68
Opendata2	YOLOv4-tiny	0.69	0.684	0.669	59
	YOLOv5S	0.767	0.752	0.748	78
	YOLOv8s	0.795	0.773	0.771	80
	YOLOv5s-STI	0.787	0.771	0.768	73
Opendata3	YOLOv4-tiny	0.715	0.709	0.694	53
	YOLOv5S	0.783	0.769	0.765	82
	YOLOv8s	0.812	0.791	0.783	88
	YOLOv5s-STI	0.803	0.788	0.784	86

本文将提出的新模型在 3 个公开的火焰和烟雾数据集 (OpenData1、OpenData2、OpenData3) 上进行了性能评估。表 4 结果显示, YOLOv5s-STI 模型在 OpenData1 数据集上的整体表现最为突出, 精确度为 0.825, 召回率为 0.811, 平均精确度 (mAP) 为 0.809, 处理速度为 68 帧/秒, 表明该模型在精确性和处理速度方面实现了良好的平衡, 适

合于需要快速响应的火灾检测等场景。在 OpenData2 数据集上, 模型精确度为 0.787, 召回率为 0.771, mAP 为 0.768, 处理速度为 73 帧/s, 显示出在真实环境下的高性能。在 OpenData3 数据集上, 模型精确度为 0.803, 召回率为 0.788, mAP 为 0.784, 处理速度为 86 帧/秒, 在保持高精确度的同时提供了快速响应。尽管 YOLOv5s-STI 模型精度略低于当前主流的最新模型, 但其 mAP 和推理速度的表现却略优于 YOLOv8s。综上所述, 改进模型在不同数据集上均展现出了高效的性能和良好的适应性, 满足实时火灾监测的严格要求。通过对比实验, 也证明了本文模型在真实火焰和烟雾检测任务中的有效性。

## 5 结语

本文通过构建基于模拟数据的古建筑火灾图像数据集, 提出了一种新的 YOLOv5s-STI 模型并进行了实验对比, 旨在提升古建筑火灾检测准确性与效率。实验结果表明, 改进后的 YOLOv5s-STI 模型在古建筑火灾场景中展现出了较高的识别精度和良好的检测速度, 证明了其在复杂环境下的应用潜力。

尽管本研究在古建筑火灾检测领域取得了一定进展, 但在实际应用中仍存在一些局限性。首先, 尽管现有的模拟平台和数据集涵盖了多种火灾场景, 但对真实环境的复杂性和多样性的模拟仍有待加强; 其次, 模型泛化能力和精确度在很大程度上依赖于数据集的真实性和多样性, 而当前数据集在这些方面仍有不足。面对这些挑战, 将对模拟平台和检测算法进行深入优化, 以更好地反映真实环境。同时, 致力于扩充和提高数据集的真实度和范围, 确保其能更全面地覆盖各类古建筑场景。

### 参考文献:

- [1] YANG L, ZHANG R Y, LI L, et al. Simam: a simple, parameter-free attention module for convolutional neural networks [C]//Proceedings of the International Conference on Machine Learning, 2021: 11863-11874.
- [2] LI D, HU J, WANG C, et al. Involution: inverting the inheritance of convolution for visual recognition [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 12321-12330.
- [3] LIU Z, LIN Y, CAO Y, et al. Swin transformer: hierarchical vision transformer using shifted windows [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 10012-10022.
- [4] SHEN J Y. Discussion on the characteristics of ancient architectural fires and protective measures [J]. Journal of Safety Science and Technology, 2007, 3(6): 134-137.  
申俊云. 浅谈古建筑火灾特点和防护对策[J]. 中国安全生产科学技术, 2007, 3(6): 134-137.
- [5] LU W L, CHENG J X. Numerical simulation analysis of the fire at puning temple in ancient buildings [J]. Fire Science and Technology, 2011, 30(4): 290-293.  
陆万里, 程金新. 古建筑普宁寺火灾数值模拟分析[J]. 消防科学与技术, 2011, 30(4): 290-293.
- [6] LIU H, LI T, ZHONG C. Research on fire prevention and control technolo-

- gy for ancient buildings based on BIM and simulation technology [J]. Fire Science and Technology, 2023, 42(9): 1228–1232.
- 刘慧, 李亭, 钟晨. 基于 BIM 及仿真技术的古建筑火灾防控技术研究 [J]. 消防科学与技术, 2023, 42(9): 1228–1232.
- [7] CHEN T H, WU P H, CHIOU Y C. An early fire-detection method based on image processing [C]//Proceedings of the 2004 International Conference on Image Processing, 2004: 1707–1710.
- [8] WANG H, SHAN W J, FANG B F. Object detection algorithm based on multi-layer context convolutional neural networks [J]. Pattern Recognition and Artificial Intelligence, 2020, 33(2): 113–120.
- 王浩, 单文静, 方宝富. 基于多层上下文卷积神经网络的目标检测算法 [J]. 模式识别与人工智能, 2020, 33(2): 113–120.
- [9] MUHAMMAD K, AHMAD J, MEHMOOD I, et al. Convolutional neural networks based fire detection in surveillance videos [J]. IEEE Access, 2018, 6: 18174–18183.
- [10] SHI L, ZHANG H G, YANG J F. Video smoke and fire detection algorithm based on an improved SSD [J]. Computer Applications and Software, 2021, 38(12): 161–167.
- 石磊, 张海刚, 杨金锋. 基于改进型 SSD 的视频烟火检测算法 [J]. 计算机应用与软件, 2021, 38(12): 161–167.
- [11] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection [DB/OL]. <https://arxiv.org/abs/2004.10934>.
- [12] ZHENG Y P, XU B Y, WANG Z Y. Improved YOLOv5 smoke detection model [J]. Journal of Computer Engineering & Applications, 2023, 59(7): 214–221.
- 郑远攀, 许博阳, 王振宇. 改进的 YOLOv5 烟雾检测模型 [J]. 计算机工程与应用, 2023, 59(7): 214–221.
- [13] WANG Y X, XIAO X L, WANG P F, et al. Improved YOLOv5s algorithm for small object smoke and flame detection [J]. Computer Engineering and Applications, 2023, 59(1): 72–81.
- 王一旭, 肖小玲, 王鹏飞, 等. 改进 YOLOv5s 的小目标烟雾火焰检测算法 [J]. 计算机工程与应用, 2023, 59(1): 72–81.
- [14] ZHANG L, LI J, ZHANG F. An efficient forest fire target detection model based on improved YOLOv5 [J]. Fire, 2023, 6(8): 291.
- [15] WANG Y, HU Y. Smoke and fire detection algorithm based on improved YOLOv5 [C]//Proceedings of the 2023 6th International Conference on Intelligent Autonomous Systems (ICoIAS), 2023: 82–87.
- [16] JI Z, PAN W, LIU S. Detecting foreign objects on transmission lines via improved YOLOv5 optimized by attention mechanism [C]//Proceedings of the 2024 IEEE 4th International Conference on Power, Electronics and Computer Applications (ICPECA), 2024: 173–178.
- [17] WANG X, CAI L, ZHOU S, et al. Fire safety detection based on CAGSA-YOLO network [J]. Fire, 2023, 6(8): 297.
- [18] ZHANG S G, ZHANG F, DING Y, et al. Swin-YOLOv5: research and application of fire and smoke detection algorithm based on YOLOv5 [J]. Computational Intelligence and Neuroscience, 2022, 2022(1): 6081680.
- [19] LI A, ZHAO Y, ZHENG Z. Novel recursive BiFPN combining with Swin Transformer for wildland fire smoke detection [J]. Forests, 2022, 13(12): 2032.
- [20] WANG G, CHEN S, HU G, et al. Detection algorithm of abnormal flow state fluid on closed vibrating screen based on improved YOLOv5 [J]. Engineering Applications of Artificial Intelligence, 2023, 123: 106272.
- [21] LI X, GENG S. Improved traffic sign detection algorithm for YOLOv5S [C]//2023 4th International Conference on Computer Engineering and Application (ICCEA), 2023: 696–699.
- [22] WANG H, BAI Y, LI H. YOLOv5S-fire: a lightweight model for flame detection [C]//Proceedings of the 2023 6th International Conference on Image and Graphics Processing, 2023: 51–57.
- [23] YANG J, LIU Y, FANG W, et al. Real-time fire detection model based on attention mechanism and multi-scale features map [C]//Matsuyama: 2022 9th International Conference on Behavioural and Social Computing (BESC), 2022.
- [24] ZHANG L Q, LIU Z T, JIANG C S. An improved SimAM Based CNN for facial expression recognition [C]//2022 41st Chinese Control Conference (CCC), 2022: 6510–6514.
- [25] NABIJIANG M, WAN X, HUANG S, et al. BAM: block attention mechanism for OCT image classification [J]. IET Image Processing, 2022, 16(5): 1376–1388.
- [26] YANG N, HE C. Malaria detection based on ResNet+ CBAM attention mechanism [C]//2022 3rd International Conference on Information Science, Parallel and Distributed Systems (ISPDS), 2022: 271–275.
- [27] FU T, QIU S, WANG Y, et al. Correction of ring artifacts with Swin-Conv-U-Net for X-ray computed tomography [J]. Applied Physics Letters, 2023, 122(12): 121107.
- [28] ZHANG J, ZHAO J. Dual-stream architecture and improved action recognition based on 3D convolutional neural network fusion [C]//2023 4th International Conference on Computer Engineering and Application (ICCEA), 2023: 181–185.
- [29] YANG G W, LIU X, GAO M, et al. Lightweight fire detection algorithm based on multi-scale feature fusion [J]. Computer Engineering and Applications, 2024, 60(23): 229–237.
- 杨国为, 刘璇, 郜敏, 等. 基于多尺度特征融合的轻量级火灾检测算法 [J]. 计算机工程与应用, 2024, 60(23): 229–237.

(责任编辑: 孙娟)