

基于图像拼接的无人机视角目标检测系统设计与实现

姜 飞, 朱月强, 刘永微, 马文斌, 陈启冠

(国营洛阳丹城无线电厂, 河南 洛阳 471000)

摘要: 目标检测技术在无人机遥感、工业缺陷检测及生物学分析等领域具有重要应用价值, 但传统方法受限于相机焦距与传感器性能, 难以实现大视野场景下的高效目标识别; 为解决这一问题, 研究提出了一种融合图像拼接与深度学习的目标检测方法, 通过将多幅具有相同特征点的局部图像拼接为全景图, 构建了宽视野检测的方法; 该方法采用基于深度学习的方法实现图像精准拼接, 结合基于深度学习的检测网络, 并创新性地引入自适应滑动窗口机制以优化检测精度; 实验结果表明, 该系统在无人机航拍数据集上实现了视野范围扩大 3 倍以上, 目标检测数量提升 50%, 检测速度相比于 YOLOv8-L 提升了 3.4 ms, 同时通过滑动窗口策略使检测准确率提高 12%; 实际应用表明, 该方法可有效满足大范围场景下的多目标检测需求, 配套开发的人机交互界面进一步提升了系统实用性, 为宽视野目标检测提供了完整的技术解决方案。

关键词: 滑动窗口检测; 无人机视角; 目标检测; 图像拼接; EDNet 算法

Design and Implementation of a Detection System for UAV Perspective Targets Based on Image Stitching

JIANG Fei, ZHU Yueqiang, LIU Yongwei, MA Wenbin, CHEN Qiguan

(State-Owned Luoyang Dancheng Radio Factory, Luoyang 471000, China)

Abstract: It is of great practical value for target detection in fields such as unmanned aerial vehicle (UAV) remote sensing, industrial defect inspection, and biomedical analysis. However, conventional methods are often limited to camera focal length and sensor performance, which are difficult to effectively identify targets in scenarios with a large field of view (FoV). To address this issue, a novel target detection framework that integrates image stitching with deep learning techniques is proposed. By aligning and stitching multiple local images that share same feature points to create a panoramic view, a detection method for targets with the large FoV is built. A deep learning-based image stitching algorithm is used to realize precise image stitching. Combined with a deep learning-based detection network, an adaptive sliding window mechanism is introduced to optimize the detection precision. Experimental evaluations on the UAV aerial datasets demonstrate that the proposed system increases the FoV by a factor of over 3, the number of target detection is increased by 50%, and compared with YOLOv8-L, this method improves the detection speed by 3.4 ms. Furthermore, the adaptive sliding window improves the detection accuracy by 12%. Practical applications show that the proposed approach can effectively meet the requirement of multi-target detection in scenarios with a large FoV. Developing a user-friendly human-computer interaction interface further improves the system's usability, offering a comprehensive and practical solution for wide FoV target detection.

Keywords: sliding window detection; UAV view; object detection; image stitching; EDNet algorithm

0 引言

目标检测 (OD, object detection) 是计算机视觉领域的重要研究方向^[1-2], 其核心任务是在图像或视频中自动识别并定位目标对象。近年来, 随着深度学习技术的

飞速发展, 基于卷积神经网络 (CNN, convolutional neural network) 的目标检测算法在精度和效率上取得了突破性进展^[3]。经典算法如 Faster R-CNN (Faster R-CNN, fast region-based convolutional network)^[4]、YOLO (YOLO, you only look once)^[5] 和 SSD (SSD, single shot multi-

收稿日期: 2025-05-09; 修回日期: 2025-06-05。

作者简介: 姜 飞 (1981-), 男, 大学本科, 高级工程师。

通讯作者: 朱月强 (1993-), 男, 博士, 工程师。

引用格式: 姜 飞, 朱月强, 刘永微, 等. 基于图像拼接的无人机视角目标检测系统设计与实现[J]. 计算机测量与控制, 2025, 33(10): 64-71.

box detector)^[6]已被广泛应用于多个实际场景, 包括人脸识别^[7]、交通监管^[8]、自动驾驶^[9]、无人机目标跟踪^[10]等, 在各领域发挥了重要作用。然而, 在实际应用中, 这些方法仍然面临诸多挑战, 尤其是当单张图像的视野范围较小时, 传统的目标检测方法的检测范围受到限制, 难以有效覆盖整个场景中的所有目标。在无人机执行大范围侦察与监控任务时尤为突出, 不仅影响目标检测的全面性, 还降低了系统的可靠性。

为了解决这一问题, 图像拼接技术应运而生^[11]。图像拼接是一种通过将多张具有共同特征点的图像融合为一幅具有更大视野、更完整信息的高分辨率或全景图像的技术。该技术主要包括特征点匹配、几何变换和图像融合等关键步骤, 使得拼接后的图像在扩展视野的同时, 实现无缝衔接与自然过渡。由于拼接后的图像不仅能提供更宽广的场景视图, 还能保留更多细节信息, 图像拼接技术已广泛应用于多个领域。例如, 在自动驾驶中, 它用于环境感知与障碍物检测, 提高行车安全性^[12]; 在虚拟现实, 它用于沉浸式场景生成, 增强用户体验^[13]; 在遥感领域, 它助力地形测绘与目标监测, 提升数据获取的精度与广度^[14]; 在地图服务中, 它用于街景图构建, 优化导航与地理信息服务^[15]。此外, 将图像拼接技术应用在智能监控系统中能显著提升系统的环境感知能力, 使监控设备能够全面捕捉区域动态信息, 从而为复杂场景下的目标检测提供更全面的数据支持。

然而, 据我们所知, 关于将图像拼接技术与目标检测方法相结合, 以实现大视野目标检测的研究仍较为有限。这可能是因为, 尽管图像拼接技术能够有效扩展视野范围, 但在拼接过程中可能引入视差失真、光照变化及遮挡等问题, 这些因素可能直接或间接影响目标检测算法的稳定性与精度。此外, 该领域仍面临诸多技术挑战。一方面, 如何高效融合来自不同视角的图像信息(或同一设备在不同时间拍摄的图像序列), 并确保目标对象的一致性和准确性, 仍然是一个亟待解决的难题。另一方面, 如何将图像拼接技术与目标检测算法有机结合, 使其在大范围场景下实现更精准、更鲁棒的目标检测, 也是当前研究的重要方向。因此, 探索图像拼接与目标检测的融合方法, 以提升大视野场景下的目标检测性能, 具有重要的研究价值与应用前景。

综上所述, 现有目标检测方法受限于相机镜头焦距和传感器性能, 单张图像的视野范围较窄, 难以覆盖大范围监控区域的所有目标, 尤其在大场景监控任务中表现出明显不足。这一局限性不仅削弱了检测系统对整体目标的全局认知能力, 还影响了检测的全面性和准确性。为解决这一问题, 将图像拼接技术与目标检测相结合, 利用拼接后的全景图像扩展检测范围, 成为提升检

测系统性能的有效途径。针对传统目标检测方法视野受限的问题, 提出了一种融合图像拼接技术的新型检测方法。该方法通过拓宽视野范围, 不仅有效增强了对大尺寸目标的检测能力, 同时提高了对多个小缺陷的识别精度, 尤其适用于需要覆盖大范围场景的应用, 为目标检测技术的发展提供了新的思路和解决方案。

1 系统结构及原理

图 1 展示了基于图像拼接技术的目标检测系统整体架构, 包括以下几个部分: 图 (a) 为基于无监督深度学习框架的图像拼接算法流程图; 图 (b) 利用扩散模型对拼接后图像进行处理的流程示意图; 图 (c) 滑动窗口对全景图进行分割的过程示意图, 其中虚线框表示滑动窗口的大小, 箭头表示其扫描路径; 图 (d) ED-Net 目标检测算法的检测头结构示意图, 用于检测 (c) 中所生成的全景图像。

图像拼接算法负责对待检测的图像进行全景图拼接, 全景图像滑动窗口则对全景图像进行滑动切分, 而目标检测算法则用于对全景图像滑动窗口中的图像进行目标检测。

首先, 本节将介绍目标检测方法, 第 2 节将进一步阐述图像拼接的相关技术。目标检测是计算机视觉领域的核心任务之一, 旨在通过特征提取自动识别图像或视频中的目标物体, 并精确确定其位置与类别。近年来, 随着深度学习的迅猛发展, 目标检测技术取得了显著进展。当前主流方法主要分为两大类: 两阶段 (Two-Stage) 方法与单阶段 (One-Stage) 方法。

以 R-CNN 系列为代表的两阶段方法通常包括两个步骤: 1) 生成候选区域; 2) 对候选区域进行分类并回归边界框。此类方法在检测精度方面表现优异, 但计算复杂度较高, 难以满足实时性要求。相比之下, 单阶段方法省略了候选区域生成步骤, 直接在整幅图像上同时进行目标分类与边界框回归, 具有更高的检测效率。YOLO 系列是单阶段方法的代表, 通过将图像划分为多个网格, 并由每个网格直接预测目标类别与边界框, 实现了高效、快速的目标检测。

然而, 由于 YOLO 在无人机视角下的小目标检测能力较弱, 本文选用了 EDNet (EDNet, EdgeDroneNet) 目标检测算法^[16]。该算法在 YOLOv10 的基础上对用于特征提取的主干网络 (Backbone) 进行了改进, 将原有的 C2f (CSPF, cross-stage partial fusion) 模块替换为 C2f-FCA (C2f-FCA, C2f-faster context attention) 模块。该模块引入了一种定制化的快速上下文注意力机制, 有效提升了特征提取能力的同时降低了计算复杂度。此外, EDNet 增加了超小检测头和交叉连接策略, 进一步优化了多尺度特征的融合效果。边界框回归方

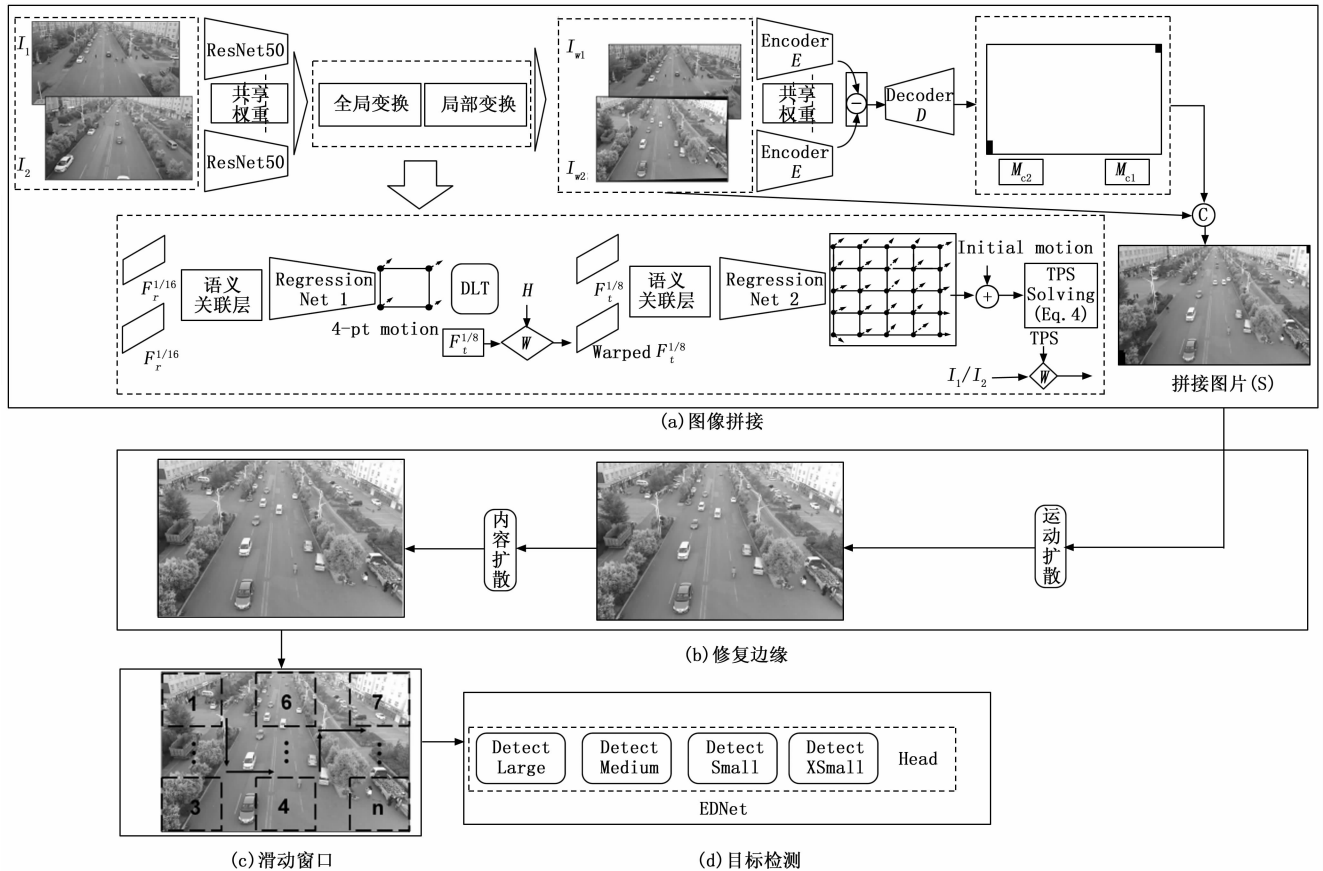


图 1 基于图像拼接技术的目标检测系统整体架构

面,采用基于 WIoUv3 (加权交并比) 的算法,以提升目标定位的精度。

如图 2 所示, FCA 模块通过 3×3 的空间混合操作整合选定输入通道的空间信息,相较于传统卷积方式,其计算复杂度显著降低。该模块还包含一个前馈网络,由两个 1×1 的点卷积 (PWConv, pointwise convolution) 层和上下文感知注意力 CAA (Context-Aware Attention) 机制组成,能够有效捕捉远距离像素之间的长程上下文依赖关系,特别适用于多同类目标混杂场景下的特征增强。FCA 模块的具体实现包括两个阶段:首先进行局部特征提取,通过平均池化和点卷积获取局部上下文信息;随后采用深度可分离带状卷积,以轻量化的方式近似大卷积核的感受野,从而在降低计算成本的同时增强特征表征能力。

2 图像拼接算法

其次是系统架构的图像拼接部分,传统的图像拼接算法通过引入更复杂的几何特征,以实现更精确的图像对齐,并尽可能保持图像的结构完整性。例如,最初 SIFT (SIFT, scale-invariant feature transform) 被广泛应用于图像拼接任务^[17],用于提取两幅或多幅图像中的关键点,并进行自适应图像变换,以确保拼接的准确

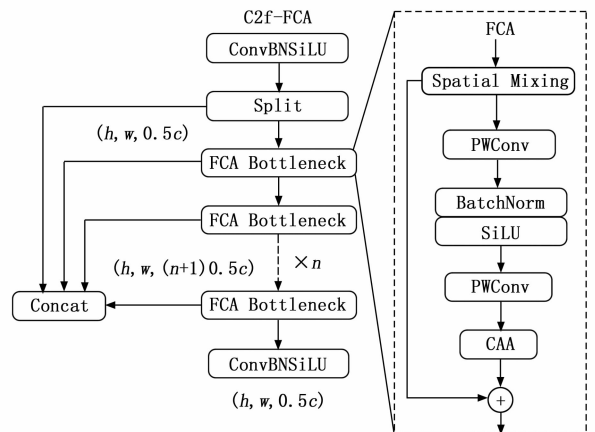


图 2 C2f-FCA 模块,该模块保持高效计算的同时优化特征提取能力

性和稳定性。随后,线结构 (LS, line structure) 作为另一种独特的特征^[18],被用于提升拼接质量,同时更好地保持场景中的线性结构^[19],使拼接后的图像更加自然流畅。此外,研究人员还探索了其他高级特征,以进一步优化拼接效果,例如深度图 (DM, depth map)^[20]通过引入场景的三维信息提升拼接的精度、语义平面结构 (SPS, semantic plane structure)^[21]则利用高层次语义信息增强拼接的稳定性和一致性。这些技术

的不断发展, 使图像拼接在复杂场景下的适应性和鲁棒性得到了显著提升。

在计算图像之间的几何变换后, 通常采用拼接缝消除方法来去除图像重叠区域产生的伪影。然而, 过度依赖几何特征的传统图像拼接算法存在两个主要问题: 1) 几何特征不足时, 图像拼接质量下降或失败——当图像中缺乏足够的几何特征时, 拼接结果可能出现模糊、失真, 甚至完全失败; 2) 几何结构复杂时, 计算成本急剧上升——在几何结构过于复杂的情况下, 传统方法的计算开销会显著增加, 影响拼接效率。相比之下, 基于深度学习的图像拼接技术不依赖复杂的几何特征, 而是能够从图像数据中捕提高级语义特征, 使其能够适应各种复杂的拼接场景。然而, 不管基于传统的图像拼接和基于深度学习的图像拼接存在边缘非矩形化的问题。基于此, 本文结合无监督深度图像拼接方案^[22], 通过引入运动扩散和内容扩散技术修补拼接后的图像边缘缺陷, 从而进一步提升拼接图像的质量和一致性。

整体无监督深度图像拼接方案如图 1 所示, 图像拼接框架由扭曲变换和组合两个模块组成。其中扭曲变换部分包含由具有重叠区域的两幅图像或几幅图像组成。通过扭曲变换将具有重叠区域的待拼接图像 I_1 和 I_2 作为输入, 并通过残差网络、全局变换和局部变换形成具有高鲁棒性扭曲变换后的图像。随后, 将扭曲变换后的图像 (I_{w1}, I_{w2}) 输入至第二个模块, 以预测合成掩码 (M_{c1}, M_{c2}), 最终拼接后的图像可以通过以下形式表示为:

$$S = M_{c1} \times I_{w1} + M_{c2} \times I_{w2} \quad (1)$$

针对第一部分实现扭曲变换后的图像, 传统的图像拼接算法通常使用基于网格的多重单应性变换实现图像的扭曲和变形, 无法实现高效的局部变形。本文通过采用薄板样条插值 (TPS, thin plate spline)^[23] 来实现高效的局部变换。TPS 变换是一种非线性、灵活的变换, 通常用于非刚性物体的变换 (具体细节见参考文献 [23], 这里我们取平衡因子为 0, 以严格约束控制点的运动)。TPS 由两组控制点决定, 在变换前的平面图像和变换后的图像之间具有一一对应关系。假设为平面图像上的 N 个控制点为 $\mathbf{P} = [P_1, \dots, P_N]^T$, 变换后的对应点为 $\mathbf{P}' = [P'_1, \dots, P'_N]^T$ (其中 $p_i, p'_i \in \mathbb{R}^{2 \times 1}$)。通过最小化一个包含数据项和失真项的能量函数, TPS 变换可以参数化为^[22]:

$$p' = \mathfrak{z}(p) = C + Mp + \sum_{i=1}^N \omega_i O(\|p - p_i\|_2) \quad (2)$$

其中: p 是平面图像上的任意一点, p' 是变换后图像上的对应点。 $C \in \mathbb{R}^{2 \times 1}$, $M \in \mathbb{R}^{2 \times 2}$, 和 $\omega_i \in \mathbb{R}^{2 \times 1}$ 是变换参数。 $O(r) = r^2 \log r^2$ 是一个径向基函数, 表示每个控制点对 p 的影响, 为了求解这些参数, 根据公式

(2) 使用 N 对控制点制定 N 个数据约束, 并施加额外的维度约束:

$$\sum_{i=1}^N \omega_i = 0 \text{ 和 } \sum_{i=1}^N p_i \omega_i^T = 0 \quad (3)$$

将公式 (3) 重新写成矩阵计算的形式, 参数可以通过以下方式求解:

$$\begin{bmatrix} \mathbf{C} \\ \mathbf{M} \\ \mathbf{W} \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{P} & \mathbf{K} \\ 0 & 0 & \mathbf{1}^T \\ 0 & 0 & \mathbf{P}^T \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{P}' \\ 0 \\ 0 \end{bmatrix} \quad (4)$$

其中: $\mathbf{1}$ 是一个 $N \times 1$ 的全 1 矩阵。矩阵 $\mathbf{K} \in \mathbb{R}^{N \times N}$ 中的每个元素 k_{ij} 由 $O(\|p - p_i\|_2)$ 决定, 且 $\mathbf{W} = [\omega_1, \dots, \omega_N]^T$ 。类似于, 单应性变换的四点参数化, TPS 也可以参数化为控制点的运动, 通过定义在目标图像上均匀分布的 $(U+1) \times (V+1)$ 个控制点, 然后预测每个控制点的运动。为了将全局单应性变换与局部 TPS 变换连接起来, 我们首先通过回归单应性变换, 以提供控制点的初始运动。然后, 我们可以预测残差运动, 以进一步实现灵活的图像变换。

如图 1 所示, 给定 I_1, I_2 , 通过采用带有预训练参数的 ResNet50^[21] 作为骨干网络用于提取语义特征。ResNet50 通过将一个三通道图像映射到高维语义特征, 并将分辨率缩小至原始图像的 $1/16$ 。然后, 通过使用上下文相关层将这些特征图 ($F_r^{1/16}, F_t^{1/16}$) 之间的相关性聚合成 2 通道的特征流。随后, 回归网络用于估计单应性变换的四点参数化。接下来, 通过将分辨率更高的特征图 $F_t^{1/8}$ 进行变换, 以将单应性先验嵌入到后续工作流程中。经过另一个上下文相关层和回归网络后, 预测控制点的残差运动, 从而实现一个鲁棒且灵活的 TPS 变换。随后经过优化图像变换, 无监督图像无缝合成以及自适应迭代变换便可实现两幅图像的无缝拼接。上述虽然能实现两幅图像的拼接, 但拼接后的图像存在非矩形边缘问题 (即拼接后的图像不是矩形的), 这会导致目标检测存在误差。

为了解决这个问题, 本文对拼接后的图像进行处理, 通过引入参考文献 [24] 中的运动扩散模型处理拼接后的图像, 能够有效的将拼接不规则边界过渡到几何校正的中间态。随后, 内容扩散模型用于图像细节的精细化处理。其中扩散模型简洁表示如下:

$$q(x_{1:T} | x_0) = \prod_{t=1}^T q(x_t | x_{t-1}) \quad (5)$$

其中: $\{x_1 \dots x_t\}$ 表示一系列逐渐增加的噪声图像, $x_0 \sim q(x)$ 表示将一系列高斯噪声添加进原始图像。在每个时间间隔中引入的噪声遵循一个由方差调度 $\{\beta_t \in (0, 1)\}_{t=1}^T$ 指定的高斯分布:

$$q(x_t | x_{t-1}) = N(x_t; \sqrt{1 - \beta_t} x_{t-1}, \beta_t S) \quad (6)$$

随后, 通过采用重参数化技术, 可以从任意中间分

布 x_t (对于任意的 $t \in (1, t)$) 进行采样:

$$q(x_t | x_0) = N[x_t; \sqrt{\bar{\alpha}_t} x_0, (1 - \bar{\alpha}_t) S] \quad (7)$$

其中: $\alpha_t = 1 - \beta_t$, $\bar{\alpha}_t = \prod_{i=1}^t \alpha_i$ 。通过引入优化后的去噪模型 θ , 以逆转扩散过程, 从而从各向同性的高斯噪声 $x_t \sim N(0, S)$:

$$p_\theta(x_{0:T}) = p(x_T) \prod_{t=1}^T p_\theta(x_{t-1} | x_t) \quad (8)$$

$$p_\theta(x_{t-1} | x_t) = N[x_{t-1}; \mu_\theta(x_t, t), \sigma_t^2 S] \quad (9)$$

通过执行这种逆向转换, 模型被赋予了将高斯分布还原为初始数据分布的能力。为了增强模型对生成过程的控制能力并提高生成图像的保真度, 还需要在模型框架中引入额外的条件变量 y 。这一方法遵循条件机制, 条件机制通过将这些变量与中间噪声数据合并, 从而实现更优的生成效果:

$$p_\theta(x_{t-1} | x_t, y) = N[x_{t-1}; \mu_\theta(x_t, t, y), \sigma_t^2 S] \quad (10)$$

基于公式 (10) 可解决拼接后的图像非矩形边缘问题, 最终通过内容填充的方式进行图像细节化处理, 得到最终拼接后的图像。此外, 对于多幅图像, 采用逐步进行拼接的方式。即首先对于图像 I_1, I_2 , 使用上述方法得到 $S_{1,2}$, 然后将 $S_{1,2}$ 与 I_3 进行拼接, 得到 $S_{1,2,3}$, 继续此过程, 将结果与接下来的图像 $I_4, I_5, \dots, I_n$ 逐步拼接, 最终得到多幅图像的拼接结果。

由于 EDNet 算法是基于较小像素的图像大小进行训练的, 而拼接后的图像通常较大, 直接对拼接后的图像进行目标检测会导致结果不准确。这是因为拼接后的图像具有更大的视野范围, 而算法的特征提取能力有限, 无法全面识别图像中的目标, 尤其是小目标检测能力会丧失。因此, 为了有效地对拼接后的全景图像进行目标检测, 需要采用 Z 字形扫描的方式。具体来说, 通过设定特定的分割核 (S_E) 和滑动模块 (S_t) 来实现全景图像的逐步检测, 以满足模型输入图像的要求。同时, 滑动模块的大小可控制全景图像的扫描速度和检测精度。具体而言, S_t 越小, 检测精度越高, 但扫描速度较慢; 反之, S_t 越大, 检测精度可能会降低, 但扫描速度较快。

3 系统软件设计

本节设计了如图 3 所示的软件系统模块, 该系统模块由多个部分组成, 包括初始化系统参数、配置系统参数、选择输入源、两种检测方式 (选择文件和选择摄像头)、数据处理、EDNet 算法预测、全景图像生成、滑动窗口、目标检测结果实时显示、交互界面、结束界面等。其中, 初始化系统用于硬件检测、加载系统固件以及设置系统环境; 配置系统参数用于设置软件运行所需的各类参数; 4 种检测方式用于加载待检测的目标图像, 包括支持图片、图片文件夹、视频和相机实时检测

等多种输入方式; 数据处理模块对上传的图像进行增强和滤波处理, 以消除环境因素对图像质量的影响。全景图像模块用于对两幅或多幅具有共同特征的图像进行拼接, 从而获得更广泛的检测视野; EDNet 算法加载预训练模型, 实现对拼接后的全景图像进行目标预测; 目标实时检测模块用于实时展示全景图像中的目标检测结果。

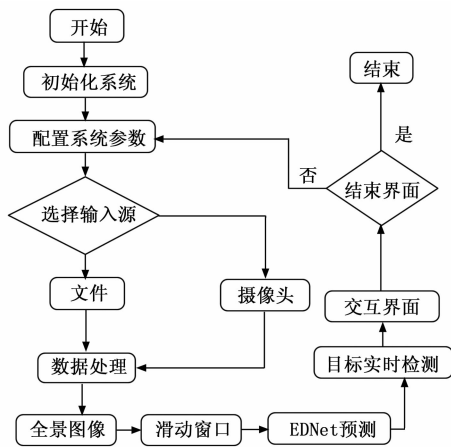


图 3 软件系统设计流程图

基于软件系统设计流程如图 3 所示, 并结合 PyQt5 和 QtDesigner, 设计了如图 4 所示的简洁 GUI 界面。该界面简化了操作步骤, 用户仅需要通过点击选择文件即可打开对应的文件, 随后点击开始运行按钮, 一方面便会将实时检测结果显示在界面上, 另一方面会根据时间戳自动的将检测结果存储在当前的输出目录下。此外, 点击开始运行按钮时, 会自动根据全景图像的边缘密度 (ED, edge density) 选择合适的滑动窗口进行检测 (程序中设定边缘密度小于 0.6 时滑动窗口设置为 800 像素, 边缘密度 (详见后续实验结果分析) 大于 0.6 时, 滑动窗口大小设置为 640 像素)。图 4 中的实时检测结果显示区域, 用来实时显示目标检测结果。操作提示按钮用于提示操作人员如何进行操作, 操作人员按钮用于显示操作的登录的操作人员。



图 4 基于图像拼接的目标检测系统软件系统

4 基于图像拼接的无人机视角下目标检测结果

在无人机视角场景下, 实现基于图像拼接技术的目

标检测的必要性从下述几个角度进行探讨: 1) 全域监控与覆盖范围扩展, 即无人机可以从空中对广阔的地面区域进行监控, 图像技术能够将无人机拍摄到的多个视角合并成一个连贯的具有大视野的图像, 从而提供更加全面的监控视野。这种全景视图可以克服传统地面监控设备因视角受限导致的盲区问题, 尤其是在开阔地或复杂地形中, 能够为安全人员提供更广泛的实时监控画面, 确保没有遗漏任何潜在的安全隐患; 2) 在无人机执行巡逻或监控任务时, 通过图像拼接技术, 可以将不同角度和时间的画面合成一个监控视频。这使得在发生突发事件或异常行为时, 监控系统能够迅速拼接多个视角的视频流, 重构现场的全过程, 为事件分析、应急响应和后续调查提供充分的数据支持; 3) 结合图像拼接与人群目标检测技术, 无人机能够对大范围区域进行高效的人群密集度检测与行为分析; 4) 智能化监控与决策支持, 通过无人机视角下的图像拼接和人群目标检测技术的结合, 可以实现更高效的视频分析与智能化监控。例如, 系统能够自动检测人群密集度过高、人员集中度异常等潜在风险, 并及时发出预警。

基于此, 本文的实验是在 Windows 系统上进行的, 利用 python 编程语言并结合深度学习来构建实验环境, 实验采用在单个 GPU 的 RTX4060Ti (16 GB) 上进行训练和测试, 在 PyTorch1.9.0、Python3.8 和 Cuda11.1 下进行实验。实验采用 VisDrone 数据集, 该数据集共 6 471 张训练数据, 验证数据共 548 张, 测试集 1 610 张, 其中检测类型包括共 10 种类型, 包括“pedestrian”“people”“bicycle”“car”“van”“truck”“tricycle”“awning-tricycle”“bus”“motor”。为评估基于图像拼接的目标检测方法整体性能, 本文使用检测目标数量 (N, number) 来评估未进行图像拼接和进行图像拼接后的检测结果性能差异, 使用 FPS 来评估检测的速度。如图 5 所示, 是基于图像拼接的目标检测系统分别对单幅图像和拼接后的图像进行检测的实验结果, 从图中可以看出单幅图像的检测视野明显小于图像拼接后检测视野 (图 5 (b) 中虚线框所示是单幅图像检测不到的范围)。其中, 图 5 (a) 中检测的目标数量 N 共 29 个, 图 5 (b) 中的目标检测数量 N 共 44 个。值得注意的是, 通过设置 $S_i=800$ 可以实现图 5 (a) 与图 5 (b) 具有相同的检测速度 $FPS=0.5$ ms。

如图 6 所示, 为了更清楚地展示无人机视角下的图像拼接检测结果, 我们选取了其中 9 幅图像对其进行拼接, 并以其中的 4 幅作为代表展示如图 6 (a) 所示, 从图中实线框可以看出单幅图像检测视野范围有限且对小目标检测能力较弱。图 6 (b) 为经图像拼接后的目标检测结果, 可以看出全景图像检测扩大了无人机检测的视野范围, 但同样对小目标的检测能力较弱, 如图中



(a) 单幅图像目标检测结果



(b) 图像拼接后的目标检测结果

图 5 基于图像拼接技术的无人机视角下目标检测对比结果

的虚线框所示, 对应的放大图像如图 6 (d) 所示。造成这种原因是由于在训练期间采用的较小的图像, 而拼接后的图像通常较大, 网络不能够有效地提取小目标特征图中的虚线框所示, 对应的放大图像如图 6 (d) 所示。造成这种原因是由于在训练期间采用的较小的图像, 而拼接后的图像通常较大, 网络不能够有效地提取小目标特征。因此, 在保证检测速度的同时, 在全景图像中设置合适的滑动窗口是非常必要的, 如图 6 (c) 所示, 将滑动窗口设置为训练图像的两倍进行的检测, 全景图像不仅增加了检测的视野范围, 还能保留小目标的检测能力, 从图中的虚线框可以看出, 对应的放大图如图 6 (e) 所示。

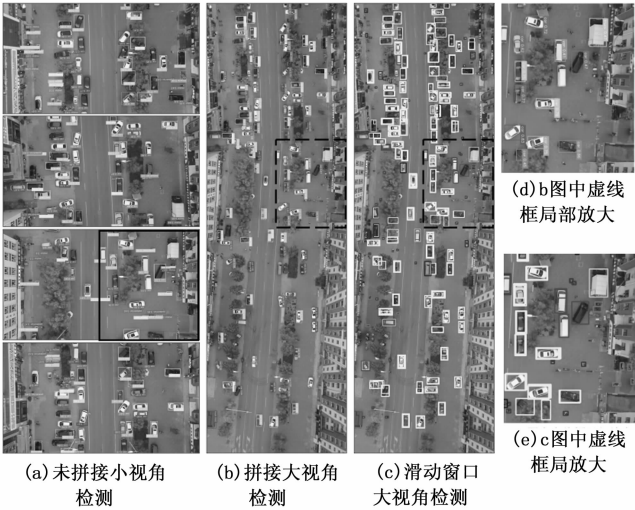


图 6 基于图像拼接技术的无人机视角下目标检测对比结果

为了全面验证系统在复杂环境下的鲁棒性，我们在多种典型场景下进行了测试。具体包括：图 7 (a) 显示了人员密集、遮挡较严重的拥挤环境（由虚线标注区域突出体现）；图 7 (b) 为光照条件剧烈变化的户外与室内过渡区域；图 7 (c) 则展示了包含不同尺度与姿态目标的多目标混合场景。实验结果表明，系统在应对复杂视觉条件时依然能够保持较高的检测稳定性和识别准确性，体现出良好的鲁棒性。

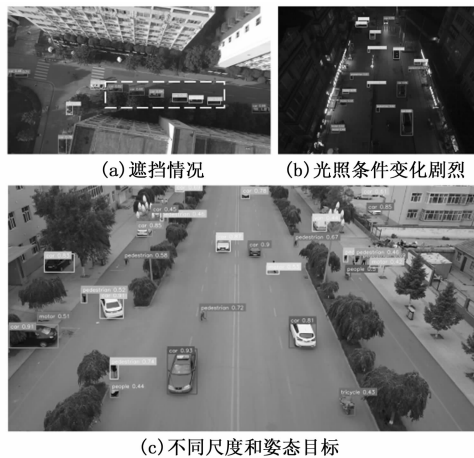


图 7 复杂场景下目标检测结果

如表 1 所示（各参数含义详见文献 [16]），我们选择了在全景图像检测下的 EDNet 模型与当前主流的目标检测方法进行了对比实验，并在相同数据集与测试条件下进行评估。结果显示，通过全景图拼接并设置滑动窗口，在提升目标检测平均精确率（50.1%）和检测速度（4.2 ms）方面具有明显优势，尤其在图像拼接处理后的大视角输入条件下，对小目标（图 6）与边缘目标的检测（图 9 (b)）表现更优。

表 1 在 VisDrone2019-DET-VAL 数据集与其他模型的对比

模型	参数/M	mAP50/%	mAP50~95/%	速度/ms
YOLOv5-L	53.1	42.9	26.2	6.3
YOLOv8-L	43.6	43.6	26.8	7.6
YOLOv10-L	25.7	44.1	27.1	5.9
RT-DETR-L	32	38.1	21.8	6.7
EDNet-B	25.5	50.1	31.4	4.2

此外，可通过设置 S_t 的大小来提升检测速度， S_t 和 FPS 之间的关系如图 8 所示，可以看出随着滑动窗口的增大，检测速度在增大；相反，随着滑动窗口的增大，检测数量却在降低。以图 5 所示的拼接图像为例，当滑动窗口的大小设置为 800 个像素时，检测速度仅仅为 0.5 ms，检测的目标数量可达 44 个。因此，在应用于基于图像拼接的目标检测系统时，可通过适当提高滑动窗口的大小提升目标检测的速度。

为了实现检测速度与精度的最佳平衡，我们通过引

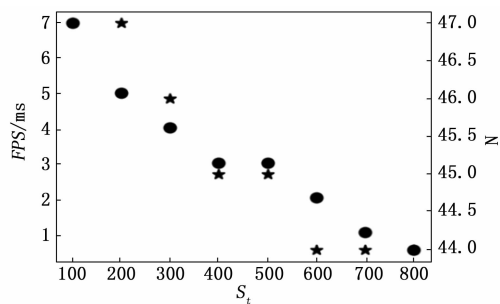


图 8 检测速度和检测数量随滑动模块之间的关系

入 ED 概念来粗略衡量待检测图像的复杂度与目标分布特征。图像中边缘密度越高，往往意味着包含更多细节和可能的目标信息，因此适合使用较小的滑动窗口以提高检测精度；而对于边缘密度较低的区域，可以适当加大窗口尺寸，从而提升整体处理速度。边缘密度计算公式如下：

$$ED = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W \frac{G(i,j)}{G_{\max}} \quad (11)$$

其中： H 和 W 分别为图像的高度和宽度， $G(i, j)$ 表示图像在像素位置 (i, j) 处的梯度幅值， T 表示梯度幅值的阈值，用于判定该像素是否为边缘。

当 ED 大于 0.6 时，取滑动窗口为 640 像素，当 ED 小于 0.6 时，取滑动窗口为 800 像素。如图 9 分别是不同 ED 下的目标检测结果，可以看出通过初步优化可以实现目标的快速且高效的检测，且设置滑动窗口的检测精度更高。



(a) $ED > 0.6, S_e = 640, S_t = 640$ (b) $ED < 0.6, S_e = 800, S_t = 800$

图 9 不同边缘密度下采用不同大小的滑动窗口检测结果

在未来的研究工作中，将进一步引入目标分布热图的自适应策略，动态调整滑动窗口大小和步长。通过构建时间—精度平衡的代价函数，结合不同任务场景（如实时监控、静态分析等）的需求，自动确定最优窗口策略。此外，本文所采用的运动扩散与内容扩散方法，在常规场景下能够有效缓解非矩形边缘对目标检测带来的干扰。在处理更复杂的边界形状或大规模图像拼接任务时，该方法可能面临性能下降或扩散效果不稳定等问题，未来将探索更具鲁棒性的扩散建模方法，如结构感知扩散网络或基于图结构的生成模型，以提升边界处理的结构保真度。

5 结束语

本研究提出了一种基于图像拼接技术的目标检测系统, 采用 EDNet 目标检测方法并引入全景图形和滑动窗口有效提升了对小目标的检测能力。系统通过深度学习驱动的图像拼接技术, 将运动扩散与内容扩散相结合, 将待检测图像拼接为视野更广的全景图像。实验结果表明, 拼接后的图像质量优良, 且不存在明显的非矩形边界问题。同时, 本文设计了针对无人机视角的多目标检测 GUI 界面, 显著提升了集成性与交互性。为解决全景图像可能带来的检测性能下降问题, 系统引入滑动模块 (S_i), 进一步优化了检测速度。实验结果显示, 在无人机视角下, 与传统的单幅图像检测方法相比, 该系统在保持相同检测速度的情况下, 不仅拥有更大的检测视野, 还能检测出更多目标。总体而言, 本文提出的基于图像拼接的无人机视角下的目标检测系统在宽视野需求的目标检测领域展现出广阔的应用前景。

参考文献:

- [1] ZOU X. A review of object detection techniques [C] // International Conference on Smart Grid and Electrical Automation, 2019: 251–254.
- [2] 高港耀. 基于深度学习的无人机视角小目标检测算法研究 [D]. 成都: 电子科技大学, 2023.
- [3] 曹家乐, 李亚利, 孙汉卿. 基于深度学习的视觉目标检测技术综述 [J]. 中国图象图形学报, 2022, 27 (6): 1697–1722.
- [4] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39 (6): 1137–1149.
- [5] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [J]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779–788.
- [6] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [J]. Computer Vision-ECCV, Cham: Springer International Publishing, 2016: 21–37.
- [7] LI L, MU X, LI S, et al. A review of face recognition technology [J]. IEEE Access, 2020, 8: 139110–139120.
- [8] LIN C J, JHANG J Y. Intelligent traffic-monitoring system based on YOLO and convolutional fuzzy neural networks [J]. IEEE Access, 2022, 10: 14120–14133.
- [9] KUKREJA R, RINCHEN S, VAIDYA B, et al. Evaluating traffic signs detection using faster R-CNN for autonomous driving [J]. IEEE 25th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks, 2020: 1–6.
- [10] WANG F, WANG H, QIN Z, et al. UAV target detection algorithm based on improved YOLOv8 [J]. IEEE Access, 2023, 11: 116534–116544.
- [11] WANG Z, YANG Z. Review on image-stitching techniques [J]. Multimedia Systems, 2020, 26 (4): 413–430.
- [12] KINZIG C, CORTES I, FERNANDEZ C, et al. Real-time seamless image stitching in autonomous driving [J]. International Con. on Information Fusion, 2022: 1–8.
- [13] PM K, NAIR A R, PHILIP J W, et al. Efficient computing for immersive VR: real-time panoramic video stitching on GPU platforms [J]. International Conference for Emerging Technology, 2024: 1–6.
- [14] PAN W, LI A, LIU X, et al. Unmanned aerial vehicle image stitching based on multi-region segmentation [J]. IET Image Processing, 2024, 18 (14): 4607–4622.
- [15] LIN Z, TANG X, WEI H. Non-ghosting stitching for street-view image [J]. Cross Strait Radio Science & Wireless Technology Conference, 2020: 1–3.
- [16] SONG Z, ZHANG Y, EBAYYEH A A R M A. EDNet: edge-optimized small target detection in UAV imagery-faster context attention, better feature fusion, and hardware acceleration [C] // IEEE Smart World Congress, 2024: 28–38.
- [17] LOWE D G. Distinctive image features from scale-invariant keypoints [J]. International Journal of Computer Vision, 2004, 60 (2): 91–110.
- [18] XIANG T, XIA G, BAI X, et al. Image stitching by line-guided local warping with global similarity constraint [J]. Pattern Recognition, 2018, 83 (C): 481–497.
- [19] JIA Q, LI Z, FAN X, et al. Leveraging line-point consistency to preserve structures for wide parallax image stitching [C] // IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 12181–12190.
- [20] LIN W, LIU S, MATSUSHITA Y, et al. Smoothly varying affine stitching [J]. IEEE Conference on Computer Vision and Pattern Recognition, 2011: 345–352.
- [21] LI A, GUO J, GUO Y. Image stitching based on semantic planar region consensus [J]. IEEE Transactions on Image Processing, 2021, 30: 5545–5558.
- [22] NIE L, LIN C, LIAO K, et al. Parallax-tolerant unsupervised deep image stitching [C] // IEEE/CVF International Conference on Computer Vision, 2023: 7365–7374.
- [23] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [J]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770–778.
- [24] ZHOU T, LI H, WANG Z, et al. RecDiffusion: rectangling for image stitching with diffusion models [C] // IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 2692–2701.