

基于改进长短焦图像融合技术的轨道交通障碍物检测系统研究

徐振森, 吕宇, 田野, 袁小军, 李晨, 袁希文

株洲中车时代电气股份有限公司数据与智能技术中心, 湖南 长沙

收稿日期: 2025年2月19日; 录用日期: 2025年3月18日; 发布日期: 2025年3月26日

摘要

近年来, 随着轨道交通网络的快速扩展和列车速度的不断提升, 传统轨交障碍物检测系统在视觉感知能力上面临目标模糊、范围受限等挑战, 难以满足日益严格的铁路运输安全需求。针对这一问题, 本文提出了一种基于改进长短焦图像融合技术的轨道交通障碍物检测系统。该系统通过结合长焦相机的高分辨率细节捕捉能力与短焦相机的广角场景感知优势, 生成清晰且信息丰富的融合图像用于障碍物检测。同时, 设计了一种结合SIFT特征提取与相位相关法的创新配准算法, 将边缘对齐误差从2.16像素显著降低至0.27像素, 显著提升了铁轨区域的对齐精度和全局图像一致性。在此基础上, 利用YOLO11目标检测模型的多尺度特征提取能力与高效推理性能开发障碍物检测系统。检测结果表明, 相比单一长焦或短焦图像, 融合图像可有效探测近距离障碍物并兼顾远距离细节, 目标检测mAP50-95指标由64.19%提升至81.32%, 增幅达26.69%。这展现了融合图像在远近目标识别、铁轨细节捕捉和全局视野感知方面的显著优势。本研究为轨道交通障碍物检测系统的视觉感知能力提升提供了重要的技术参考。

关键词

轨道交通, 长短焦图像融合, 障碍物检测, SIFT特征, 相位相关法, YOLO11

Research on Railway Transportation Obstacle Detection System Based on Improved Long and Short Focal Image Fusion Technology

Zhensen Xu, Yu Lyu, Ye Tian, Xiaojun Yuan, Chen Li, Xiwen Yuan

Data and Intelligence R&D Center, Zhuzhou CRRC Times Electric Co., Ltd., Changsha Hunan

Received: Feb. 19th, 2025; accepted: Mar. 18th, 2025; published: Mar. 26th, 2025

文章引用: 徐振森, 吕宇, 田野, 袁小军, 李晨, 袁希文. 基于改进长短焦图像融合技术的轨道交通障碍物检测系统研究[J]. 人工智能与机器人研究, 2025, 14(2): 414-426. DOI: 10.12677/airr.2025.142041

Abstract

In recent years, with the rapid expansion of rail transit networks and the continuous increase in train speeds, traditional rail transit obstacle detection systems face challenges such as target blurriness and limited detection range, making it difficult to meet the increasingly stringent requirements of railway transportation safety. To address these issues, this paper proposes a rail transit obstacle detection system based on improved long-focal and short-focal image fusion technology. By combining the high-resolution detail capture capability of long-focal cameras with the wide-angle scene perception advantage of short-focal cameras, the system generates clear and information-rich fused images for obstacle detection. An innovative registration algorithm that integrates SIFT feature extraction and phase correlation is designed, reducing edge alignment error from 2.16 pixels to 0.27 pixels, significantly enhancing alignment accuracy in rail regions and improving overall image consistency. On this basis, the YOLO11 object detection model's multi-scale feature extraction capability and efficient inference performance are leveraged to develop the obstacle detection system. Experimental results show that, compared to individual long-focal or short-focal images, the fused image effectively detects nearby obstacles while preserving distant details, with the mAP50-95 metric for target detection improving from 64.19% to 81.32%, an increase of 26.69%. This demonstrates the significant advantages of fused images in recognizing both near and distant targets, capturing rail track details, and perceiving the global scene. This study provides a valuable technical reference for enhancing the visual perception capabilities of rail transit obstacle detection systems.

Keywords

Railway Transportation, Long-Focal and Short-Focal Image Fusion, Obstacle Detection, SIFT Features, Phase Correlation Method, YOLO11

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. 引言

轨道交通在当今社会中占据着至关重要的地位，随着铁路网络的扩展和列车速度的提升，铁路安全成为提高运输效能与保障公共安全的核心目标[1]。近年来，伴随着智能交通技术的蓬勃发展，障碍物检测系统作为一种先进的安全防护手段，已逐步在铁路运营中得到广泛应用[2]。这些系统通过摄像设备获取轨道周围环境的视觉信息，并利用计算机视觉技术对潜在障碍物进行检测[3][4]。然而，现有的单一焦距成像技术在远近目标检测中存在一定局限性，例如远处目标的细节捕捉不清晰，或近处场景的信息缺失，影响了障碍物检测的全面性和准确性。

针对上述问题，长短焦图像融合技术通过结合长焦相机的细节捕捉优势与短焦相机的广角覆盖能力，生成清晰且信息丰富的融合图像，为障碍物检测提供了有效解决方案[5]。近年来，图像融合技术受到了广泛的研究，其典型流程包括图像采集、图像预处理、图像配准和图像融合等，而图像配准作为核心步骤，是决定融合质量的关键[6]。无法有效配准会导致明显的对不齐问题，影响后续检测任务的准确性[7]。

图像配准方法主要分为两类：基于空间域的图像配准方法[8]-[10]和基于变换域的图像配准方法[11]-[13]。基于空间域的图像配准方法主要通过提取多焦点图像中的特征并进行特征匹配来实现配准。在[8]中，通过提取长短焦相机图像的 SIFT 特征，在空间域内进行特征匹配，利用单应性矩阵进行全局配准。

在[9]中,提出了一种基于密集 SIFT 特征的多焦点图像配准方法,通过活动水平图和特征匹配生成精细的配准图像。在[10]中,提出了一种自动多焦点图像配准算法,通过 BRISK 和 SURF 特征匹配,并结合改进的 RANSAC 实现图像配准。

在变换域内,图像配准方法主要依赖于频域的多尺度、多方向特征分解,通过对图像在不同频段的系数进行处理和选择,实现精确配准。在[11]中,提出了一种基于非下采样轮廓波变换(NSCT)的多焦点图像配准方法,通过选择低频和带通方向子带系数,实现高精度的变换域图像配准。在[12]中,提出了一种基于 Daubechies 复小波变换(DCWT)的图像配准方法,通过多分辨率分解和最大选择融合波特系数,实现无冗余、对噪声鲁棒的变换域配准。在[13]中,提出了一种基于图像周期分解和傅里叶变换的变换域图像配准方法,通过消除图像边界效应,提高了相位相关配准的准确性和成功率。

在现有的长短焦图像融合技术中,尽管已取得一些重要进展,但在公开文献中,尚未有专门针对轨道交通系统的图像融合方法,尤其是在轨道交通中的特定应用场景下,如铁路轨道交界处的处理仍面临诸多挑战。图像在这些交界处的错位和不对齐现象会严重影响融合图像的整体质量,以及系统在后续障碍物检测中的精度。为此,本文将改进的长短焦图像融合技术引入轨道交通障碍物检测系统的应用中,提出了一种基于 SIFT 特征提取和相位相关法的创新图像配准算法。该方法既考虑了 SIFT 特征等空间域信息,又引入了变换域的轨道相位信息进行细化对齐。相比单一的空间域方法,本文方法能够更好地应对轨道场景下的复杂融合需求,实现对图像细节与全局信息的良好兼顾,解决了传统方法难以应对的铁轨错位和割裂问题。此外,试验结果表明,基于所提出的长短焦图像融合技术的轨道交通障碍物检测系统能够充分利用长焦图像的高分辨率细节与短焦图像的广角视野,实现早期预警和大范围目标检测,显著提升了系统的整体性能和安全性。

2. 轨交障碍物检测系统方案

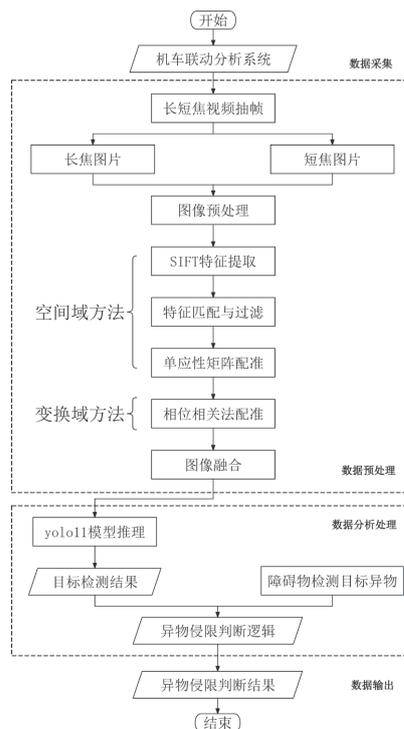


Figure 1. Process of the obstacle detection system for rail transit
图 1. 轨道交通障碍物检测系统的流程

本文提出的轨道交通障碍物检测系统整体方案如图 1 所示, 涵盖数据采集、数据预处理、数据分析处理及数据输出四个核心模块。首先, 系统从机车联动分析系统获取分辨率为 1920×1080 、帧率为每秒 25 帧的长焦(焦距 75 mm)和短焦(焦距 22.8 mm)轨道场景视频流, 并按固定间隔(每 0.1 秒抽一帧)同步提取长焦和短焦图片。图像预处理阶段基于相机标定参数对相机进行畸变校正和立体校正, 生成无畸变且成像平面一致的图像对。随后, 在空间域中, 基于 SIFT 算法对长短焦图像构建高斯金字塔, 通过差分高斯极值检测与 Hessian 矩阵边缘过滤提取鲁棒特征点, 结合暴力匹配与 Lowe's 比例测试筛选高质量匹配对, 利用 RANSAC 算法优化单应性矩阵完成全局粗配准; 接下来, 针对铁轨区域局部错位问题, 引入变换域的轨道相位信息进行傅里叶频谱分析, 通过相位差峰值检测实现亚像素级平移修正, 最终采用加权平均法生成细节清晰、全局一致的融合图像。然后, 将处理后的图像输入 YOLO11 目标检测模型进行模型推理, 输出边界框与类别概率。最终, 结合 YOLO11 的目标检测结果, 基于轨交系统特定的目标侵限标准输出检测报告和预警信号。

3. 改进的长短焦图像融合技术

长短焦图像融合技术旨在综合利用长焦相机在细节捕捉方面的优势与短焦相机在全局场景感知上的广角能力, 以生成既具高分辨率又包含完整场景信息的融合图像。为实现这一目标, 本章结合轨道交通场景的实际需求, 提出了一种基于空间域和变换域相结合的融合策略。具体而言, 本章首先对长短焦图像进行预处理, 包括畸变校正和立体校正, 以消除成像畸变并对齐两幅图像的成像平面; 随后, 通过空间域的 SIFT 特征提取、特征匹配与过滤以及单应性矩阵对齐, 完成长短焦图像的初步配准; 最后, 结合变换域的相位相关法对铁轨等关键区域进行精细配准, 以进一步提高融合图像的边缘对齐精度。

3.1. 长短焦图像的预处理

长短焦图像的预处理包括畸变校正和立体校正, 旨在消除相机镜头畸变以及对齐长短焦相机的成像平面。

(1) 畸变校正用于矫正相机镜头的径向和切向畸变, 使图像变形恢复为真实场景的直观表示。基于现场机车相机的内参矩阵 \mathbf{K} 和畸变系数 D , 可以使用以下公式进行校正:

$$x' = \mathbf{K} \cdot (x + \Delta x) \quad (1)$$

其中, x 为原始像素坐标, Δx 为畸变校正项(包含径向和切向畸变的修正值)。

(2) 立体校正通过对两台相机的旋转和平移进行校正, 确保两幅图像的成像平面平行化, 从而方便特征匹配与深度估计。校正过程的关键在于计算两个相机的相对旋转矩阵 \mathbf{R} 和相对平移向量 \mathbf{T} :

$$\mathbf{R} = \mathbf{R}_2 \times \mathbf{R}_1^T, \mathbf{T} = \mathbf{T}_2 - \mathbf{R} \times \mathbf{T}_1 \quad (2)$$

其中, \mathbf{R}_1 和 \mathbf{R}_2 为两相机的旋转矩阵, \mathbf{T}_1 和 \mathbf{T}_2 为两相机的平移向量。在得到后 \mathbf{R} 和 \mathbf{T} , 使用 OpenCV 的 stereoRectify 方法生成校正变换矩阵和重映射参数。

3.2. SIFT 特征提取

尺度不变特征变换(Scale-Invariant Feature Transform, SIFT)是一种广泛应用于图像特征提取的算法 [14] [15], 旨在从输入图像中提取具有尺度和旋转不变性的关键点及其描述符。SIFT 算法的特征提取过程主要包括尺度空间构建、关键点检测、关键点精确定位、方向分配以及描述符, 步骤如下:

(1) SIFT 算法首先通过高斯模糊和金字塔下采样构建尺度空间, 以检测不同尺度下的特征。高斯金字塔的构建公式为:

$$G(x, y, \sigma) = I(x, y) * G(x, y, \sigma) \quad (3)$$

其中, $G(x, y, \sigma)$ 表示尺度 σ 为的高斯模糊核, $I(x, y)$ 为输入图像, $*$ 表示卷积操作, σ 为尺度参数。

(2) 在尺度空间构建完成后, SIFT 算法通过计算差分高斯(Difference of Gaussian, DoG)来检测潜在的特征点。差分高斯的计算公式为:

$$D(x, y, \sigma) = G(x, y, k\sigma) - G(x, y, \sigma) \quad (4)$$

其中, $D(x, y, \sigma)$ 是尺度 σ 下的 DoG 图像, k 为尺度空间的倍增因子, 通常 $k = \sqrt{2}$ 。通过在 DoG 空间中寻找局部极值点, 初步确定图像中的潜在特征点。

(3) 初步检测到的关键点需要进一步精确定位, 以提高其稳定性和准确性。通过拟合二次曲面, 并根据对比度阈值和 Hessian 矩阵比值 R 过滤掉低对比度或位于强边缘的特征点:

$$R = \frac{Tr(H)^2}{\det(H)} \quad (5)$$

其中, $Tr(H)$ 是 Hessian 矩阵的迹, $\det(H)$ 是其行列式。当比值 R 大于预设阈值时, 认为该特征点位于边缘, 需予以剔除。

(4) 最后是方向分配, 在以特征点为中心的邻域采样, 根据图像小区域计算的梯度方向直方图, 分配不同的关键点一个单一或不同的方向。

3.3. 特征匹配与过滤

在完成 SIFT 特征提取后, 下一步是进行特征匹配, 以找到长短焦图像之间对应的特征点对。然而, 直接进行特征匹配往往会引入误匹配点, 因此需要进一步过滤以提高匹配的准确性。在本算法中使用暴力匹配(Brute Force Matching, BFMatcher)搜索对图像中的特征点进行匹配, 使用 Lowe's 比例测试过滤掉不良匹配[16]。

(1) BFMatcher 通过逐一比较两个图像中所有特征点的描述符, 并计算它们的欧几里得距离。对于两个描述符向量 d_1 和 d_2 , 其欧几里得距离定义为:

$$dist(d_1, d_2) = \sqrt{\sum_{i=1}^n (d_{1i} - d_{2i})^2} \quad (6)$$

其中, n 是描述符的维度。

(2) Lowe's 比例测试则通过将每个特征点的最近邻和次近邻的距离进行比较, 计算它们的距离比值:

$$Ratio = \frac{d_1}{d_2} \quad (7)$$

其中, d_1 和 d_2 分别是最近邻和次近邻的欧几里得距离。若比值低于预设阈值(通常为 0.75), 则认为该匹配点对为有效匹配; 否则, 将其视为误匹配点予以舍弃。

3.4. 单应性矩阵对齐

在获得高质量的特征匹配点对后, 算法利用单应性矩阵(Homography Matrix)实现长短焦图像的初步配准。

(1) 单应性矩阵 H 是一个描述从一个平面投影到另一个平面的变换矩阵[17], 为了估计单应性矩阵 H , 采用最小二乘法, 通过最小化以下误差函数来优化匹配点对的拟合精度:

$$E(H) = \sum_{i=1}^n \|q_i - Hp_i\|^2 \quad (8)$$

其中, n 为匹配点的数量, p_i 和 q_i 分别为长焦图像 I_1 和短焦图像 I_2 中的匹配点坐标。即对于长焦图像 I_1 和短焦图像 I_2 中的每一对匹配点, 通过 H 可以将 I_1 中的点 p_i 映射到 I_2 中的对应点 q_i 。

(2) 一旦获得了单应性矩阵 H , 就可以使用该矩阵对长焦图像进行透视变换。透视变换的公式如下:

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = H \cdot \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (9)$$

其中, (x, y) 是长焦图像中的像素坐标, (x', y') 是透视变换后的图像坐标。

3.5. 图像融合

在第 3.2 至 3.4 节中, 所描述的方法是基于空间域的特征配准算法, 包括 SIFT 特征提取、特征匹配与过滤以及单应性矩阵对齐。这些方法与文献[8]-[10]中的经典空间域特征配准方法类似, 均通过提取图像中的局部特征并进行特征匹配以完成图像配准。

现在本研究将这一方法应用于轨道机车长短焦场景中, 以解决实际轨交环境中因焦距差异导致的图像融合问题。本算法采用的图像融合算法是加权融合, 将配准后的长焦图像 I_1 和短焦图像 I_2 在重叠区域使用加权平均公式:

$$I_F(x, y) = \alpha \cdot I_1(x, y) + (1 - \alpha) \cdot I_2(x, y) \quad (10)$$

其中, $\alpha \in [0, 1]$ 是加权系数, 用于调整长焦图像和短焦图像在融合结果中的贡献权重。



Figure 2. Long and short focal length images collected on-site by the locomotive
图 2. 机车现场采集的长短焦图像



Figure 3. Fused image after spatial domain registration
图 3. 空间域配准后的融合图像

如图 2 所示, 为机车现场采集的原始长短焦图像, 可以发现, 短焦图像具有较大的视野范围, 能够涵盖近处的信号灯、接触网立柱及监控设备等场景; 而长焦图像则凭借其优越的高分辨率, 能够清晰捕捉远处的目标, 例如图中的土挡和信号灯。根据相机的内参参数计算, 长焦图像与短焦图像的焦距比约

为 3.28, 长焦图像的内容是短焦图像中心区域。通过将第 3.2 至 3.4 节中描述的空间域特征配准算法应用于图 2 中的长短焦图像, 生成了图 3 所示的融合图像。结果显示, 长焦图像被配准到短焦图像的中心区域, 二者在整体结构上实现了较为精确的对齐。然而, 在铁轨等边缘区域, 仍然存在一定程度的错位, 即图像配准的局部误差尚未完全消除。这种误差的出现主要归因于基于空间域特征配准算法的固有局限性: 该方法依赖于局部特征点的提取和匹配, 而铁轨等边缘区域缺乏明显的拐角和纹理变化, 导致其特征点稀疏, 因而配准的鲁棒性和精度会显著下降。这表明, 仅依靠空间域特征配准难以解决铁轨边缘的配准问题, 需进一步结合其他的配准方法予以优化。

3.6. 相位相关法精细配准

在 3.5 节中, 基于空间域的特征配准算法已实现了长短焦图像的初步整体对齐, 但在铁轨等边缘区域仍存在一定的错位现象。这种错位问题主要归因于局部特征点匹配的不足, 空间域方法对特定场景复杂性的适应能力有限。为了进一步提高配准精度, 本文引入了基于变换域的相位相关法(Phase Correlation Method), 用于对铁轨等关键区域进行精细配准。

相位相关法的核心思想是利用图像频域中的相位信息来计算图像之间的平移量, 避免了空间域计算中的模糊误差[18]。特别是在轨道场景中, 铁轨在频域中的表现尤为显著。

相位相关法基于傅里叶变换和移位定理。移位定理表明, 若两幅图像之间存在平移变换, 那么它们的频谱幅值是相同的, 仅在相位上存在差异:

$$F_1(u, v) = F_2(u, v) \times e^{j2\pi(u \times d_x + v \times d_y)} \quad (11)$$

其中, $F_1(u, v)$ 和 $F_2(u, v)$ 分别为长焦图像 I_1 和短焦图像 I_2 的傅里叶变换, d_x, d_y 为图像在空间域中的平移量, u, v 为图像在频域中的坐标。

(1) 为了估算图像的平移量, 首先计算两幅图像铁轨区域之间的相位差, 公式如下:

$$\Delta\theta(u, v) = \arg(F_1(u, v)) - \arg(F_2(u, v)) \quad (12)$$

(2) 接着通过逆傅里叶变换将相位差转换回空间域, 理想情况下, 逆变换后的结果会在某个位置形成一个尖锐的峰值, 峰值位置即为两幅图像的平移量(d_x, d_y)。最终公式如下:

$$(dx, dy) = \arg \max [F^{-1}(\Delta\theta(u, v))] \quad (13)$$

其中, F^{-1} 表示傅里叶逆变换, $\arg \max$ 表示找到的峰值位置。

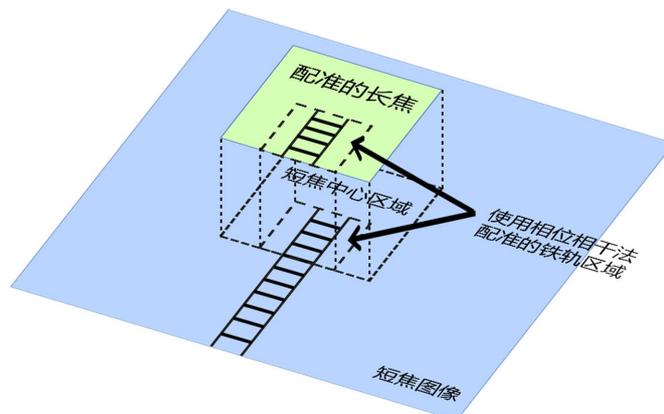


Figure 4. Schematic diagram of the registration area for the phase correlation method
图 4. 相位相关法配准区域的示意图

(3) 基于长短焦图像在轨交场景应用的先验知识,如图4所示,铁轨主要集中在长焦图像中间1/3区域的下半部分。根据此特点,可将相位相关法的计算范围限定于该特定区域,从而有效提高配准效率,减少非目标区域的干扰,实现长焦图像与短焦图像在铁轨区域的进一步精细配准。

3.7. YOLO11 算法概述

YOLO 算法作为首个实现单阶段目标检测的框架,将目标检测问题转化为回归问题,通过直接回归目标包络框的类别和位置,展现了卓越的实时性和工程应用价值[19]。YOLO11 是 Ultralytics 团队于 2024 年 10 月发布的最新模型,通过重新设计网络架构和优化训练流程,进一步提升了模型的精度与效率。如图 5 所示,YOLO11 的网络结构主要引入了 C3k2 (Cross Stage Partial with kernel size 2)和 C2PSA (Convolutional block with Parallel Spatial Attention)等创新组件,增强了特征提取和处理能力。其中, C3k2 模块结合 C2f 和 C3 模块的优势,通过将 Bottleneck 的卷积核转变为两个核大小为 3 的卷积层提升了特征提取能力,并且 C3k2 模块存在两种结构,支持灵活切换标准 Bottleneck 或 C3k 结构,以适应多尺度目标检测任务。

C2PSA 模块在 C2f 的基础上进一步增强了特征的表达能力,主要通过引入位置自注意力(PSA)机制来提升模型性能。它使用改进的多头注意力机制(PSABlock)替代了传统的 Bottleneck 层,增强了对目标局部特征和位置关系的捕捉能力。引入无激活卷积层替代 LayerNorm 层,并用两个卷积层代替 MLP 层,提高了特征提取的效率和适应性。这些改进显著提升了 YOLO11 的检测性能。相比 YOLOv8m, YOLO11m 在 COCO 数据集上的 mAP 实现提升,同时参数量减少了 22%,实现了更高的检测精度和更低的计算成本。

YOLO11 凭借其高效的特征提取能力和优化的网络结构,能够在轨道交通环境中实时检测潜在障碍物提供预警,显著提升行车安全性。其多尺度特征提取机制能够有效检测不同距离和尺寸的目标,它优化的训练方法和高效推理能力,使其能够适应轨交场景中高实时性的要求,为动态监控系统提供高效可靠的检测结果。结合 YOLO11 的创新,我们将在第 4 章展示其在轨道交通障碍物检测系统中的应用效果和性能表现。

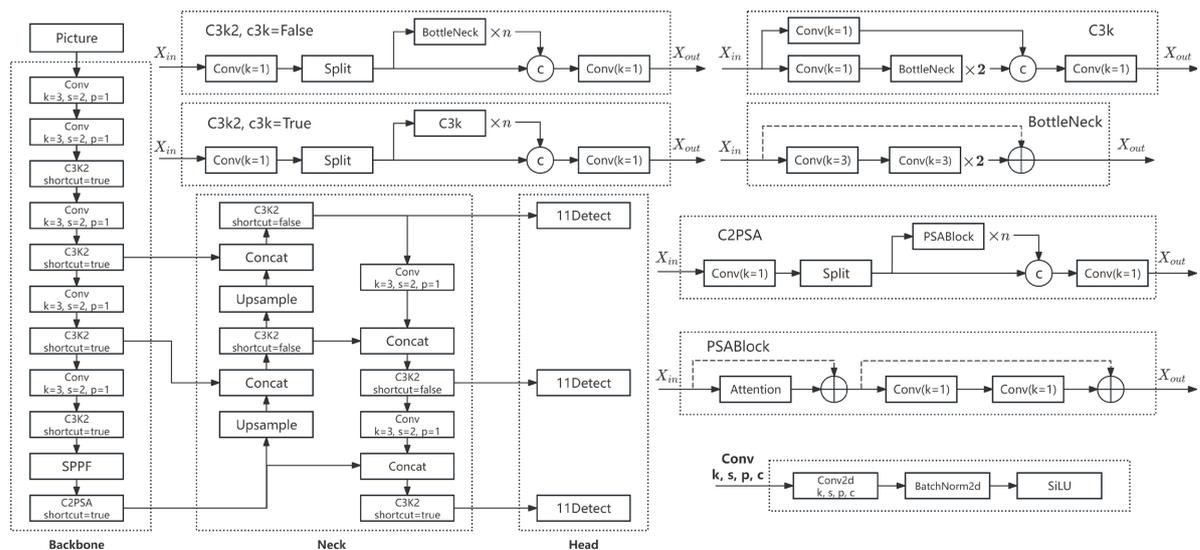


Figure 5. Network structure diagram of YOLO11

图 5. YOLO11 网络结构图

4. 改进的长短焦图像融合技术结果对比分析

图 6 展示了在空间域配准的基础上引入变换域相位相关法精细配准后的融合图像。从图中可以看出,

融合图像的边界对齐效果相比图 3 有了明显提升，尤其是在铁轨附近区域，由原来的明显错位实现了精准对齐。这一改进得益于长短焦图像中铁轨几何结构在中间区域的高相似性，以及铁轨在变换域中显著的相位信息特征，为相位相关法准确计算精细配准平移量提供了可靠依据。

表 1 显示，相比于原始短焦图像，融合图像在清晰度上得到了显著提升。通过 Brenner 梯度法[20]评估图像清晰度，融合图像的 Brenner 值从 2.43×10^7 提高至 2.61×10^7 ；Laplacian 算子[20]的评估结果显示，融合图像的清晰度值从 53.04 提高至 66.08。此外，使用 Tenengrad 算法[20]评估图像清晰度时，融合图像的 Tenengrad 值从 3.75×10^9 提高至 4.17×10^9 。这些结果表明，融合图像在保留短焦图像大视野优势的同时，显著增强了细节的清晰度。

为进一步验证本算法在边缘对齐精度和时间效率方面的表现，表 2 给出了与空间域特征配准算法的对比结果。在 1000 组图片的对比实验中，边缘对齐误差(边缘错位的像素值)评估表明，本算法将误差从 2.16 像素显著降低至 0.27 像素，减少幅度高达 87.5%。在时间效率方面，本算法的整体耗时为 86 ms，相较于空间域特征配准算法的 78 ms 略有增加，但这一时间成本在大幅提升对齐精度的情况下完全可以接受。

通过结合空间域特征配准与变换域精细配准的优势，第二章提出的改进长短焦图像融合策略有效解决了铁轨区域的错位问题。特别是，相位相关法的引入显著提升了融合图像在局部细节处理和全局一致性上的表现，为轨道交通场景中的智能视觉感知提供了高质量的数据基础，第 4 章将进一步验证其在实际任务中的性能优势。



Figure 6. Fused image after fine registration using the phase correlation method
图 6. 相位相干法精细配准后的融合图像

Table 1. Clarity metrics for fused images

表 1. 融合图像清晰度指标

指标	Brenner	Laplacian	Tenengrad
原始短焦	2.43×10^7	53.04	3.75×10^9
融合图像	2.61×10^7	66.08	4.17×10^9

Table 2. Comparison of algorithm metrics

表 2. 算法指标对比

指标	边缘对齐误差	时间效率
空间域特征配准	2.16 px	78 ms
本算法	0.27 px	86 ms

5. 基于 YOLO11 的障碍物检测系统应用测试

在第 3 章中, 我们详细介绍了改进的长短焦图像融合技术, 并于第 3.7 节概述了 YOLO11 算法的网络结构创新及其在目标检测中的优势。这些技术为轨道交通障碍物检测提供了高质量的图像输入和高效率的检测能力。本节将重点探讨上述技术在轨道交通场景中的实际应用测试。我们在实验室专用视频分析服务器平台上部署并测试了该系统, 验证了融合图像在提升障碍物检测性能方面的效果, 并与原始短焦图像进行了对比分析。

5.1. 试验环境

本实验的障碍物检测模型训练并部署于实验室专用视频分析服务器平台, 硬件配置包括 512 GB RAM、8 块 NVIDIA A30 GPU (单卡显存 24 GB) 和 2 颗 Intel Xeon Gold 6342 处理器(2.80 GHz, 24 核心 48 线程)。软件环境为 Ubuntu 20.04 (64 位) 操作系统, CUDA 12.4 和 cuDNN 9.1 提供计算加速, 利用 PyTorch 深度学习框架搭建深度神经网络模型, 构建了高性能实验平台以支持障碍物检测模型的开发与测试。

长短焦视频数据来源于国铁南宁局机务段现场部署的机车联动分析系统, 视频分辨率为 1920×1080 , 每段视频时长约为 15 分钟。实验对长焦和短焦视频分别进行抽帧, 并根据障碍物检测任务中目标异物的需求制作图像训练和验证数据集。数据集涵盖了土挡、信号灯、行人、体积大于 $30 \times 30 \times 30 \text{ cm}^3$ 的障碍物以及车辆等检测目标。长焦和短焦图片各标注了 1000 张, 并按 4:1 的比例划分为训练集和测试集。

在模型训练过程中, 本研究设置输入图片的批量大小(batch size)为 64, 图像尺寸(imgsz)为 640, 以确保模型能够有效处理高分辨率图像数据。训练中采用了改进的自适应动量估计算法 AdamW (Adaptive Moment Estimation with Weight Decay), 初始学习率(learning rate)设置为 0.000714, 动量(momentum)为 0.9, 权值衰减(weight decay)为 0.0005。此配置在确保快速收敛的同时, 能够有效提升优化过程中的性能表现。训练过程利用 8 块 NVIDIA A30 GPU 进行多卡并行计算, 显著提高了训练效率。训练总迭代次数(epochs)设置为 500, 以充分学习复杂场景特征, 提升模型的检测精度和鲁棒性。

5.2. 障碍物检测试验分析

为了说明融合图像相较于长焦图像和短焦图像的优势, 本研究将训练完成的障碍物检测模型分别应用于同一时刻的长焦图像、短焦图像以及融合图像进行检测, 检测结果如图 7 所示。场景 1 模拟了对近处信号灯以及远处土挡和信号灯的检测。从短焦图像的检测结果可以看出, 近处的信号灯被准确识别, 置信度为 0.94, 但由于远处土挡和信号灯较为模糊, 未能被模型检测到。而在长焦图像的检测结果中, 近处信号灯未包含在图像范围内, 但远处的土挡和信号灯均被检测到, 置信度分别为 0.89 和 0.85。对于融合图像的检测结果, 近处的信号灯被检测到, 置信度为 0.94, 远处的土挡和信号灯也被识别, 置信度分别为 0.77 和 0.76。

场景 2 模拟了对近处摩托车以及远处行人的检测。从短焦图像的检测结果中可以看到, 近处的摩托车被成功检测, 置信度为 0.93, 但远处行人由于图像模糊仅被低置信度检测, 置信度为 0.31。长焦图像的检测结果表明, 近处摩托车未包含在图像范围内, 但远处的行人被准确识别, 置信度为 0.92。融合图像的检测结果显示, 近处的摩托车被成功检测, 置信度为 0.93, 远处的行人也被高置信度检测到, 置信度提升至 0.81。

试验结果表明, 融合图像有效结合了长焦和短焦图像的优势, 不仅在远处的细节捕捉方面展现了长焦图像的高分辨率特性, 还在全局场景感知上保留了短焦图像的广角能力, 从而显著增强了远处目标的细节呈现与整体检测性能。

为了进一步证明融合图像在障碍物检测中的整体优势，表 3 展示了短焦图像和融合图像在障碍物检测模型中的性能对比。由于长焦图像在同一场景下仅能够捕捉部分目标物，未能涵盖全部检测目标，因此无法在同一数据集上进行全面评估。通过在同一时刻场景的短焦图像与融合图像数据集上进行评估，结果表明，融合图像在所有关键指标上均明显优于短焦图像。具体来说，mAP50 (平均准确率, IoU = 0.5) 从短焦图像的 90.95% 提升至融合图像的 98.87%，提升幅度达 8.71%；mAP50-95 (平均准确率, IoU = 0.5~0.95) 从 64.19% 提升至 81.32%，提升幅度高达 26.69%。在 Precision、Recall 和 F1-Score 方面，融合图像相较短焦图像分别提升了 0.83%、17.57% 和 9.16%。这一结果进一步验证了融合图像在障碍物检测任务中的优势，特别是在同时具备长焦图像的高分辨率和短焦图像的广角特性时，能够有效提升模型对不同距离目标的检测能力。

以上结果表明，改进的长短焦图像融合技术不仅可以克服单一视角图像所存在的局限性，还能够通过集成不同焦段的优点，进一步提升障碍物检测系统的环境感知能力，为轨道交通等智能驾驶系统的安全性和可靠性提供有力支持。

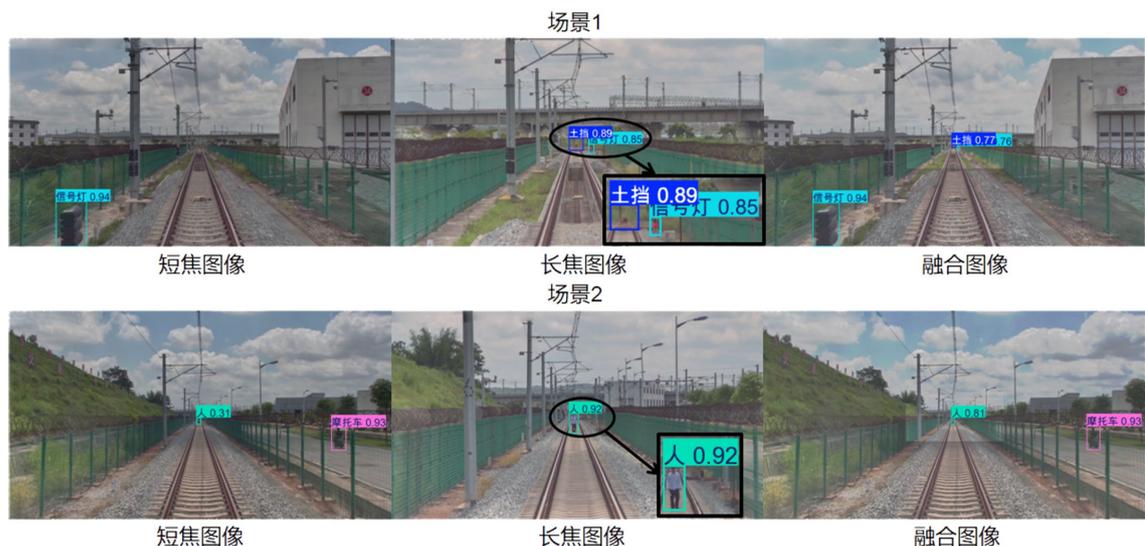


Figure 7. Detection results of the obstacle detection model

图 7. 障碍物检测模型检测结果

Table 3. Comparison of obstacle detection performance between short focal length images and fused images

表 3. 短焦与融合图像在障碍物检测中的性能对比

指标	mAP50	mAP50-95	Precision	Recall	F1-Score
原始短焦	90.95%	64.19%	94.97%	82.30%	88.18%
融合图像	98.87%	81.32%	95.76%	96.76%	96.26%

6. 结论

本文将改进的长短焦图像融合技术应用于轨道交通障碍物检测系统，提出了一种基于 SIFT 特征提取与相位相关法相结合的创新图像配准方法。该方法不仅利用了 SIFT 在空间域中对图像特征的高效提取与匹配能力，还引入了变换域中的相位相关技术，针对轨道交通场景中的铁轨区域进行了精细对齐。相比传统的空间域特征配准方法，本文的方法在处理轨道场景下复杂的图像融合需求方面表现出更高的鲁棒性和准确性，将边缘对齐误差从 2.16 像素降低至 0.27 像素，解决了传统方法在铁轨错位和图像割裂上的不足。

基于该图像融合技术, 本文进一步结合 YOLO11 目标检测模型的高效特征提取与实时推理能力, 构建了轨道交通障碍物检测系统。试验结果表明, 该系统不仅保留了长焦相机在捕捉远处高分辨率细节方面的优势, 还结合了短焦相机在全局场景感知上的广角能力, 实现了对远近目标的精准检测和早期预警。其目标检测性能的 mAP50-95 指标从短焦图像的 64.19% 提升至 81.32%, 显著提升了系统的整体性能和安全性。研究成果为轨道交通障碍物检测系统的视觉感知能力提升提供了重要参考, 展现出良好的应用价值与推广潜力。

未来的研究可围绕长短焦图像融合技术的工程化应用。一方面, 针对嵌入式硬件平台优化算法设计, 提高实时性与计算效率; 另一方面, 探索更高效的配准与融合算法, 增强对复杂轨道场景的适应性。同时, 开发面向不同轨道交通场景的定制化解决方案, 推进技术从实验室向实际工程的转化, 进一步提升系统的稳定性和实用性, 为轨道交通安全保障提供更加全面的技术支持。

基金项目

本文由国家重点研发计划项目(2022YFB4300602)资助。

参考文献

- [1] Chellaswamy, C., Dhanalakshmi, A., Chinnammal, V. and Malarvizhi, C. (2017) An IoT-Based Frontal Collision Avoidance System for Railways. 2017 *IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI)*, Chennai, 21-22 September 2017, 1082-1087. <https://doi.org/10.1109/icpsi.2017.8391877>
- [2] Bhavsar, S.S. and Kulkarni, A.N. (2016) Train Collision Avoidance System by Using RFID. 2016 *International Conference on Computing, Analytics and Security Trends (CAST)*, Pune, 19-21 December 2016, 30-34. <https://doi.org/10.1109/cast.2016.7914935>
- [3] 姚巍巍, 田野, 李晨. 基于改进 SOLO 的列车主动避障视觉算法研究[J]. 电子测量技术, 2022, 45(9): 133-139.
- [4] Sahba, F. and Sahba, R. (2018) Prevention of Metro Rail Accidents and Incidents in Stations Using RFID Technology. 2018 *World Automation Congress (WAC)*, Stevenson, 3-6 Jun 2018, 1-5. <https://doi.org/10.23919/wac.2018.8430408>
- [5] Li, X., Li, X., Cheng, X., Wang, M. and Tan, H. (2023) MCDFD: Multifocus Image Fusion Based on Multiscale Cross-Difference and Focus Detection. *IEEE Sensors Journal*, **23**, 30913-30926. <https://doi.org/10.1109/jsen.2023.3330871>
- [6] 唐文彦. 图像拼接关键技术研究[学位论文]. 上海: 上海交通大学, 2014.
- [7] 杜俊康. 场景自适应的图像拼接[D]. [硕士学位论文]. 南京: 东南大学, 2020.
- [8] 关腾腾, 刘兵, 曾建波, 等. 一种长短焦相机图像融合的方法[P]. 中国, CN116740522A. 2023-09-12.
- [9] Liu, Y., Liu, S. and Wang, Z. (2015) Multi-Focus Image Fusion with Dense Sift. *Information Fusion*, **23**, 139-155. <https://doi.org/10.1016/j.inffus.2014.05.004>
- [10] Liu, Y. and Yu, F. (2015) An Automatic Image Fusion Algorithm for Unregistered Multiply Multi-Focus Images. *Optics Communications*, **341**, 101-113. <https://doi.org/10.1016/j.optcom.2014.12.015>
- [11] Zhang, Q. and Guo, B. (2009) Multifocus Image Fusion Using the Nonsubsampled Contourlet Transform. *Signal Processing*, **89**, 1334-1346. <https://doi.org/10.1016/j.sigpro.2009.01.012>
- [12] Miao, Q., Shi, C., Xu, P., Yang, M. and Shi, Y. (2011) A Novel Algorithm of Image Fusion Using Shearlets. *Optics Communications*, **284**, 1540-1547. <https://doi.org/10.1016/j.optcom.2010.11.048>
- [13] Singh, R. and Khare, A. (2014) Fusion of Multimodal Medical Images Using Daubechies Complex Wavelet Transform—A Multiresolution Approach. *Information Fusion*, **19**, 49-60. <https://doi.org/10.1016/j.inffus.2012.09.005>
- [14] Tang, G., Wei, Z. and Zhuang, L. (2024) SAR Image Registration with Sift Features and Edge Points. *IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium*, Athens, 7-12 July 2024, 9711-9715. <https://doi.org/10.1109/igarss53475.2024.10642824>
- [15] Song, K., Zhu, F. and Song, L. (2022) Moving Target Detection Algorithm Based on SIFT Feature Matching. 2022 *International Conference on Frontiers of Artificial Intelligence and Machine Learning (FAIML)*, Hangzhou, 19-21 June 2022, 196-199. <https://doi.org/10.1109/faiml57028.2022.00045>
- [16] Renjith, R., Reshma, R. and Arun, K.V. (2017) Design and Implementation of Traffic Sign and Obstacle Detection in a Self-Driving Car Using SURF Detector and Brute Force Matcher. 2017 *IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI)*, Chennai, 21-22 September 2017, 1985-1989.

- <https://doi.org/10.1109/icpsi.2017.8392062>
- [17] Luo, Y., Li, Z. and Luo, S. (2023) Research on Camera Calibration Method Based on Homography Matrix. 2023 *IEEE 6th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, Chongqing, 24-26 February 2023, 464-468. <https://doi.org/10.1109/itnec56291.2023.10082094>
- [18] Gerganov, G., Papucharov, A., Kawrkow, I. and Mitev, K. (2013) Portal Image Registration Using the Phase Correlation Method. *Proceedings of the 2013 IEEE Nuclear Science Symposium and Medical Imaging Conference (2013 NSS/MIC)*, Seoul, 26 October-2 November 2013, 1-3.
- [19] Liu, A., Liu, Y. and Kifah, S. (2024) Deep Convolutional Neural Network for Enhancing Traffic Sign Recognition Developed on Yolo V5. 2024 *International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC)*, Bhubaneswar, 27-29 January 2024, 1-6. <https://doi.org/10.1109/assic60049.2024.10508025>
- [20] Her, L. and Yang, X. (2019) Research of Image Sharpness Assessment Algorithm for Autofocus. 2019 *IEEE 4th International Conference on Image, Vision and Computing (ICIVC)*, Xiamen, 5-7 July 2019, 93-98. <https://doi.org/10.1109/icivc47709.2019.8980980>