



基于图像识别技术的龙溪口航电工程 智能视频监控系统设计

姜旭东¹, 宋本扬², 杜泽永¹, 邢小军¹, 王佳俊², 郑春³

(1. 四川岷江港航电开发有限责任公司, 四川乐山 614000;

2. 天津大学, 水利工程仿真与安全国家重点实验室, 天津 300350;

3. 四川港航建设工程有限公司, 四川成都 610023)

摘要: 为加强施工现场的安全管理, 对工人的安全帽佩戴进行智能化检测, 提出一种基于图像识别技术的航电工程智能视频监控系统, 由目标检测模块和网络通信模块构成, 按照 C/S 架构部署, 并成功应用于岷江龙溪口航电枢纽施工现场。其中目标检测模块采用先进的目标检测算法, 通过卷积神经网络对图像中的安全带和安全帽进行特征提取、分类和定位, 完成识别; 网络通信模块通过网络连接视频采集终端、服务器和视频播放终端完成图像采集、传输和显示。对系统各组分的测试表明, 系统各项功能完备, 各模块间连接完好, 能够完成工程建设过程各场景的实时图像采集、传输、识别和显示功能, 为龙溪口航电工程建设安全管理提供了坚实的基础。

关键词: 航电工程; 图像识别; 视频监控; 安全帽识别; 目标检测

中图分类号: U61

文献标志码: A

文章编号: 1002-4972(2023)10-0183-05

Design of intelligent video monitoring system of Longxikou Navigation-power project based on image recognition technology

JIANG Xudong¹, SONG Benyang², DU Zeyong¹, XING Xiaojun¹, WANG Jiajun², ZHENG Chun³

(1. Sichuan Minjiang Port & Shipping & Electricity Power Development Co., Ltd., Leshan 614000, China;

2. State Key Laboratory of Hydraulic Engineering Simulation and Safety, Tianjin University, Tianjin 300350, China;

3. Sichuan Port and Channel Construction Engineering Co., Ltd., Chengdu 610023, China)

Abstract: To strengthen the safety management of construction site and intelligently detect the helmets worn by workers, this paper proposes an intelligent video monitoring system for navigation and power project based on object detection, which consists of object detection module and network communication module, deployed according to the C/S architecture, and successfully applied to the construction site of navigation and power hub project on Minjiang River. The object detection module uses advanced object detection algorithm to extract, classify and locate the features of safety belts and helmets in the image through convolutional neural network, and thus complete the recognition of safety belts and helmets. The network communication module connects the video capture terminal, server and video player terminal through the network to complete the image capture, transmission and display functions. The test of each component of the system shows that the system has complete functions and the modules are well connected. It can complete the real-time image acquisition, transmission, detection and display functions of each scene in the process of project construction, providing a solid foundation for the safety management of Longxikou navigation and power project construction.

Keywords: navigation and power project; image recognition; video monitoring; helmet detection; object detection

收稿日期: 2023-06-07

作者简介: 姜旭东 (1984—), 男, 工程师, 从事港口航道及水利水电工程建设管理。

近年来，施工现场人员安全事故频发。据统计，建筑工程安全生产事故中高处坠落和由物体打击引起的事故位列前两名，分别占 53.69% 和 15.91%^[1]，逐渐发展为威胁施工现场人员安全的重大隐患。施工用安全帽是保护施工人员免受上述伤害的有效措施^[2]，但由于施工人员安全意识缺乏、设备舒适度低等问题，施工现场安全保障设备的实际佩戴情况并不乐观^[3]。随着施工现场视频监控的普及和机器学习智能算法的兴起，借助基于图像识别技术的智能化手段对监控视频中的人员安全保障设备进行自动化检测成为可能。基于图像识别的方法可以有效避免人为疏忽和不必要的人力投入、节约成本，同时还能高效地完成检测工作，保证现场施工工人安全，对提升施工现场的管理水平有指导意义。

本文提出一种基于图像识别技术的智能视频监控系统，能够通过施工现场布置的视频监控系统实时采集视频资料，采用基于深度学习的图像识别算法^[4]对视频中采集到的人员安全带、安全帽佩戴情况进行实时检测，并将检测结果在视频中以醒目的形式标注，最终反馈到大屏系统中。经试验分析，系统能够完成施工现场的图像采集、传输和识别，为一线施工现场安全管理提供数据支持和指导，并已成功应用于四川省岷江龙溪口航电枢纽工程。

1 目标识别

目标检测技术近年发展迅速，按照检测方式主要可分为基于区域建议的两阶段模型和基于回归的单阶段模型。其中，两阶段模型通过区域建议和目标分类 2 个阶段完成目标检测任务，精度较高但检测速度较慢，如 R-CNN^[5]、Fast R-CNN^[6]、Faster-R-CNN^[7] 及 Mask R-CNN^[8] 等模型。单阶段模型通过回归直接进行目标检测任务，检测速度快但检测精度较低，容易造成漏检、误检等问题，如 YOLO^[9]、YOLOv2^[10]、YOLOv3^[11]、YOLOv4^[12]、SSD^[13]、RetinaNet^[14] 等模型。

考虑到施工现场安全帽检测的准确度要求高

于速度要求，本系统的识别模块主要采用基于深度学习的计算机视觉手段，采用先进的 Faster R-CNN 目标检测方法对施工现场实时采集到的视频监控画面进行识别与分析。其主要结构可分为 2 个阶段，第 1 阶段生成目标可能出现的位置，由负责特征提取的卷积神经网络 CNN 和区域建议网络 RPN 组成，第 2 阶段对目标进行分类与定位，由一系列堆叠的神经网络层组成。

1) 卷积神经网络。卷积神经网络是一种基于卷积数学运算的前馈神经网络，相较于传统的全连接前馈神经网络，卷积神经网络具有参数量少、运算速度快、适用范围广泛等特点。此外，由于二维卷积本身对输入的平移变换并不敏感，卷积神经网络能够高效地完成图像识别、分类等工作。在以图像识别和分类为基础的目标检测模型中，常使用卷积神经网络作为主干，自动提取图像中各种物体的特征，避免传统图像识别技术中需要人工设定物体特征的缺点。卷积神经网络数学运算式：

$$f(x)g(x) = \int_{-\infty}^{\infty} f(\tau)g(x - \tau) d\tau \quad (1)$$

$$f(x)g(x) = \sum_{\tau=-\infty}^{\infty} f(\tau)g(x - \tau) \quad (2)$$

式中： f 、 g 为函数； x 、 τ 为自变量。

2) 区域建议网络。区域建议网络的主要作用是在通过卷积神经网络进行特征提取得到的特征图中生成目标候选框。RPN 采用滑动窗口方式，在特征图上滑动 1 个 $n \times n$ 的窗口，通过卷积神经网络提取窗口中的特征，送入 2 个不同的全连接层分别作边框回归和分类，得到框体坐标和分类置信度，其中滑动窗口使用 1 个尺寸为 $n \times n$ 的卷积核在特征图上执行卷积运算实现。因此，RPN 相较于 CNN 实际上并无网络层的增加，存储与计算开销与经典 CNN 网络基本一致。

此外，在每一个滑动窗口上，RPN 采用以窗口中心为锚点的 3 种边长(8, 16, 32)和 3 种宽高比(0.5, 1, 2)共计 9 种组合的锚盒(anchor box)。锚盒尺寸和宽高比例的多种组合初步解决了图像中不同目标尺寸、比例不尽相同的问题。

2 网络通信

2.1 基本架构

网络通信采用 RTSP (real-time stream protocol) 实时流媒体数据网络传输协议, 采用 C/S 架构实现视频监控终端-服务器-网页端的通信和视频流传输。如图 1 所示, C/S 架构由若干视频监控终端、1 台网络硬盘录像机、1 台服务器及若干视频播放终端构成。

其中视频监控终端是系统的图像采集终端, 负责施工现场各处监控画面的实时采集工作; 网络硬盘录像机位于监控终端与服务器之间, 负责对现场采集到的实时视频数据进行整合并接入网络; 服务器接收网络硬盘录像机发送视频流并负责存储、处理、转发至播放终端; 视频播放终端则是系统的显示终端, 负责将服务器发送到网络端的视频流解码并显示。

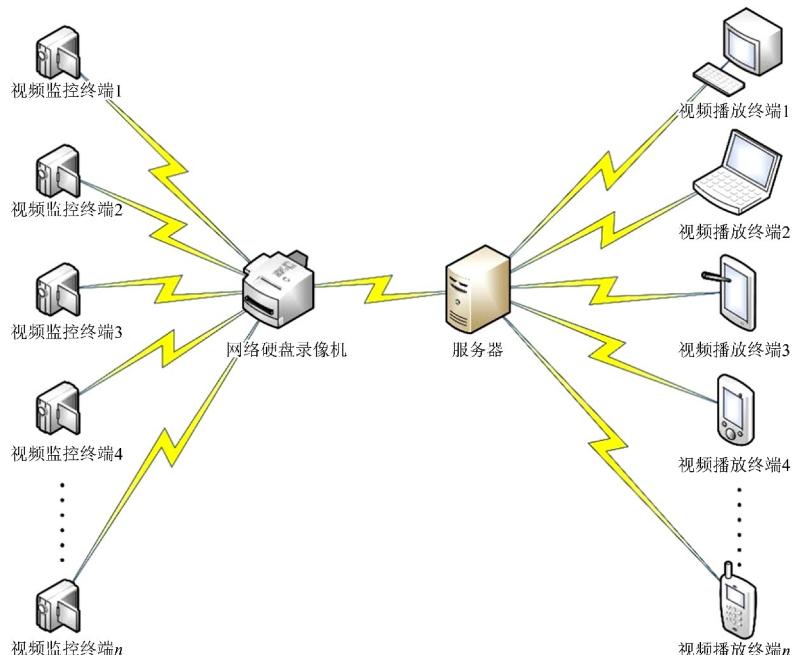


图 1 系统 C/S 架构

2.2 服务器

服务器是系统的核心组成部分, 负责通过 RTSP 协议拉取网络硬盘录像机中的视频流, 在本地存储视频并完成安全带及安全帽的检测, 并将检测后已标注的视频流重新按照 RTSP 协议格式封装并发送到网络。实际运行过程中, 服务器与网络硬盘录像机以 C/S 架构按照 TCP 协议建立通信。数据发送前, 在客户端与服务器之间进行 3 次握手建立连接; 之后网络

硬盘录像机按照 RTSP 协议对各摄像机的实时采集视频分通道进行封装编码, 并通过 TCP 协议网络发送到服务器; 服务器接收到网络中的视频流后, 通过解码器对视频流进行解码, 送入目标检测模块进行安全带与安全帽的识别; 识别结果以在原视频逐帧进行带框标注的形式呈现, 并继续按照 RTSP 协议封装编码后发送至网络端; 最后服务器与客户端进行 4 次握手断开连接。服务器工作流程见图 2。

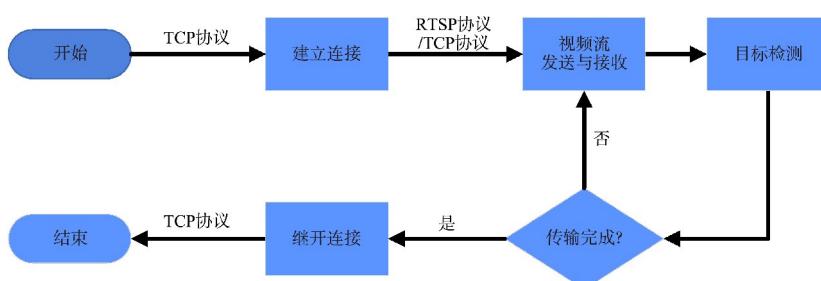


图 2 服务器工作流程

3 系统测试

3.1 目标检测模块

测试使用设备型号为 Intel Core i7 10700 CPU 和 NVIDIA GeForce GTX 1660Super GPU (6 GB)，采用 Ubuntu 20.04 操作系统、CUDA 11.2 科学计算库以及基于 Tensorflow1.4 版本的深度学习框架完成 Faster R-CNN 网络搭建、训练和测试。其中 CNN 使用在 ImageNet 数据集上预先训练的 ResNet101^[15]模型参数作为初始化权重参数，RPN 与 CNN 共享卷积层和池化层参数，其他主要初始化参数见表 1。

目标检测模块设计要求为：截取实时采集到的监控画面进行安全带和安全帽检测，并将检测结果标注于原画面上。本次试验使用公开数据集和现场数据集共同进行检测，检测平均精确度为 80.8%，检测速度为 0.440 s/张。如图 3a)、3c) 所示，在一般场景 1、3 下检测精度较高，能够检测出全部安全帽；而在场景 2 的复杂情况下，则可能出现错检，如图 3b)。针对现场情况，如图 4 所示，无论是目标距离较远如图 4a)、4b) 或较近如图 4c)，系统均能够准确识别安全帽佩戴情况，模块表现满足要求。

表 1 模型主要初始化参数

学习率	学习率衰减步长	迭代次数	优化器	锚盒边长	锚盒宽高比	RPN 批次大小(batch size)
0.001, 0.000 1	30 000	80 000	SGD	8, 16, 32	0.5, 1.0, 2.0	256



图 3 公开数据集安全帽检测结果

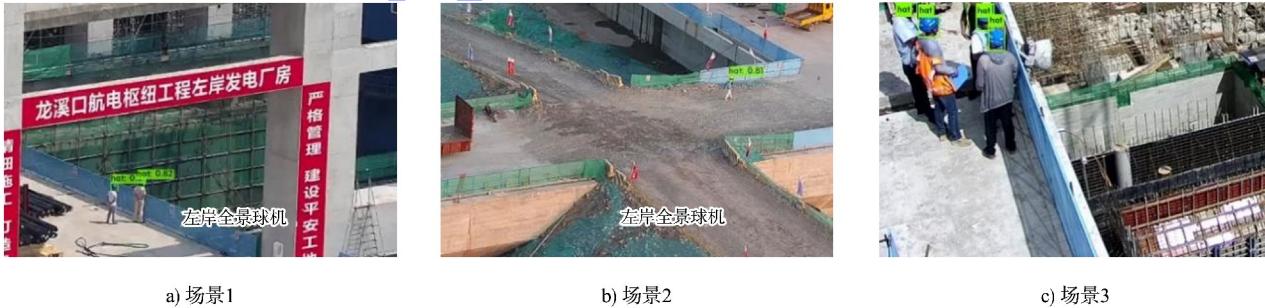


图 4 龙溪口航电枢纽现场安全帽检测结果

3.2 网络通信模块

网络通信模块测试过程使用的设备型号同目标检测模块相同。测试分为 6 个部分，测试结果见

表 2。经测试，服务器能够成功连接至采集终端并任意选取终端将画面传输、处理、存储并发送至播放端。因此，网络通信模块功能完备，部署情况良好。

表 2 模型测试结果

测试内容	测试方法	测试结果
服务器与网络硬盘录像机是否能够通过 TCP 协议建立连接	Ping 指令	连接良好
是否能够任意选取连接到网络硬盘录像机的视频监控终端	随机选取若干终端	选取功能完善
服务器是否能够向终端发出视频流请求信号	根据上一步发布对应的视频请求指令	发送功能良好
服务器是否能够接收到所请求的视频流	根据上一步发布的视频请求指令接收视频流	接收功能良好
服务器是否能够存储、处理、转发至播放终端	对接收到的视频流进行存储、目标检测并发送至播放端	存储、处理、转发功能良好
视频播放终端是否能够正常将视频流解码并显示	采用播放器对视频流进行播放	视频解码播放功能良好

4 结语

1) 基于 Faster R-CNN 图像识别技术和 C/S 架构的网络通信方法, 搭建航电工程智能视频监控系统, 该系统通过高准确率的目标检测模块和基于 C/S 架构的网络通信模块的相互耦合, 完成龙溪口航电工程施工现场的监控, 包括①各复杂场景的图像实时采集; ②低延迟图像数据传输; ③图像内容高准确率识别; ④检测结果实时播放。

2) 本文提出的方法为龙溪口航电工程建设过程的安全监管提供智能化方法, 降低人工监管成本, 有效地提高了施工现场安全管理水平。

参考文献:

- [1] 住房和城乡建设部办公厅关于 2020 年房屋市政工程生产安全事故情况的通报. 中华人民共和国住房和城乡建设部 [EB/OL]. (2021-05-21) [2023-06-05]. https://www.mohurd.gov.cn/gongkai/zhengce/zhengcefillelib/202210/20221026_768565.html.
- [2] 吕谋贵. 分析施工现场安全帽佩戴情况监控技术 [J]. 低碳世界, 2018(7): 205-206.
- [3] DONG X S, SANG D C, BORCHARDT J G, et al. Fatal falls from roofs among U. S. construction workers [J]. Journal of safety research, 2013, 44(2): 17-24.
- [4] HINTON G E, OSINDERO S, TEH Y W. A fast learning algorithm for deep belief nets [J]. Neural computation, 2006, 18(7): 1527-1554.
- [5] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//2014 IEEE conference on computer vision and pattern recognition, Columbus: Institute of electrical and electronics engineers, 2014: 580-587.
- [6] GIRSHICK R. Fast R-CNN [C]//2015 IEEE international conference on computer vision (ICCV), Santiago: IEEE Computer Society, 2015: 1440-1448.
- [7] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE transactions on pattern analysis & machine intelligence, 2017, 39(6): 1137-1149.
- [8] HE K, GKIOXARI G, DOLLAR P, et al. Mask R-CNN [C]//2017 IEEE international conference on computer vision (ICCV), Venice: IEEE Computer Society, 2017: 2980-2988.
- [9] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]//2016 IEEE conference on computer vision and pattern recognition (CVPR), Las Vegas: IEEE, 2016: 779-788.
- [10] REDMON J, FARHADI A. YOLO9000: Better, Faster, Stronger [C]//2017 IEEE conference on computer vision and pattern recognition (CVPR), Honolulu: IEEE, 2017: 6517-6525.
- [11] JIANG X, GAO T, ZHU Z, et al. Real-time face mask detection method based on YOLOv3 [J]. Electronics, 2021, 10(7): 837.
- [12] JIANG P, ERGU D, LIU F, et al. A review of Yolo algorithm developments [J]. Procedia computer science, 2022, 199: 1066-1073.
- [13] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]//Computer vision-ECCV 2016: 14th European conference, Amsterdam: Springer International Publishing, 2016: 21-37.
- [14] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal Loss for Dense Object Detection [C]//2017 IEEE international conference on computer vision (ICCV), Venice: IEEE Computer Society, 2017: 2999-3007.
- [15] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]//2016 IEEE conference on computer vision and pattern recognition (CVPR), Las Vegas: IEEE Computer Society, 2016: 770-778.