A Novel Dataset for Detecting Pedestrian Heads in Crowds Using Deep Learning Algorithms

Mohammed Abubaker[2], *Zubayda* Alsadder[2], *Hamed* Abdelhaq[2,*], *Mohcine* Chraibi[1,**], *Maik* Boltes[1], and *Ahmed* Alia[1,2,***]

[1]Institute for Advanced Simulation, Forschungszentrum Jülich, 52425, Germany;
[2]Information Technology Department, Faculty of Engineering and Information Technology, An Najah National University, Nablus, Palestine

Abstract. This work introduces a diverse high-resolution dataset aimed at enhancing automatic pedestrian head detection in crowd videos at railway platforms and event entrances. It includes 109,914 human heads manually annotated from 64 videos. Experimental results demonstrate its significant impact on improving detection performance (Mean Average Precision) for deep learning algorithms in such environments.

1 Introduction

The automatic detection of pedestrian heads in crowded environments is crucial for various crowd analysis and management tasks, including crowd counting, density estimation, pedestrian trajectory extraction, and behavior detection. Despite advancements in deep learning algorithms for object detection, existing studies struggle with pedestrian head detection in crowded situations such as railway platforms and event entrances, where risks frequently arise. One main reason for the poor head detection performance is the underrepresentation of such scenarios in existing datasets. These scenarios are particularly challenging due to variations in lighting conditions, viewpoints, occlusions, scale changes, indoor/outdoor environments, and weather conditions.

To narrow this gap, we introduce a novel, diverse, and high-resolution dataset of human heads in crowds at Railway Platforms and Event Entrances, named the RPEE-Heads dataset.

2 Dataset Preparation

Here, we detail the steps required to generate the RPEE-Heads dataset, including the training, validation, and testing sets. First, a group of 64 video recordings was selected from closed and open data archives hosted by Forschungszentrum Jülich [1]. These videos encompass a wide range of real-life scenarios and experiments at railway platforms and event entrances, providing diversity in viewpoints, lighting conditions, weather, time of day, indoor/outdoor environments, head sizes, and frame resolutions. Figure 1 illustrates examples of scenes from

*e-mail: hamed@najah.edu
**e-mail: m.chraibi@fz-juelich.de
***e-mail: a.alia@fz-juelich.de

these data sources. In the second step, we extracted 1,888 frames from the selected videos and manually added accurate head-bounding box annotations. Eventually, the dataset was divided into 70% training, 15% validation, and 15% test sets. The generated RPEE-Heads dataset includes 106,207 head annotations: 75,963 for training, 15,412 for validation, and 14,822 for testing.

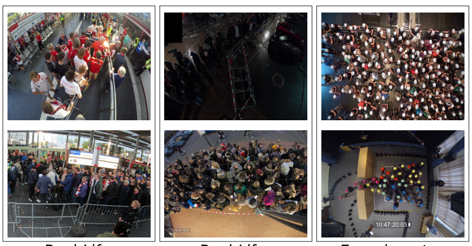

Figure 1. Illustrative examples from the selected data sources.

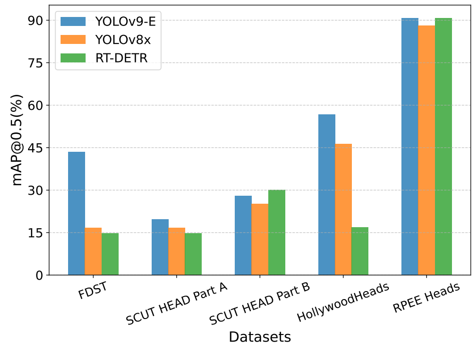

Figure 2. Mean Average Precision (mAP) across diverse datasets and models in the context of railway platforms and event entrances.

3 Evaluation and Results

This section aims to evaluate the effectiveness of the RPEE-Heads dataset in training reliable deep learning models for detecting pedestrian heads, particularly in crowded environments such as railway platforms and event entrances. To achieve this goal, we trained and evaluated three state-of-the-art deep learning-based object detection algorithms— YOLOv8x [2], YOLOv9-E [3], and Real-Time Detection Transformer (RT-DETR) [4]— using the proposed dataset. Additionally, we trained the same algorithms using four publicly available datasets with head annotations: Fudan-ShanghaiTech (FDST) [5], Scut Head Part A, Scut Head Part B, and HollywoodHeads. These trained models were then evaluated on the RPEE-Heads dataset. Results, as depicted in Figure 2, indicate that models trained on the PEE-Heads dataset achieved mAP exceeding 88%, whereas models trained on public datasets showed suboptimal performance, with the highest mAP being 57% for head detection at railway platforms and event entrances.

References

[1] Pedestrian Dynamics Data Archive 2008. *Data archive of experimental data from studies about pedestrian dynamics*. http://10.34735/ped.da.

[2] Jocher, Glenn and Chaurasia, Ayush and Qiu, Jing, 2023. *Ultralytics YOLOv8*. https://github.com/ultralytics/ultralytics.

[3] Wang, C.Y., Yeh, I.H. and Liao, H.Y.M., 2024. *YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information*. arXiv preprint. arXiv:2402.13616.wang2024yolov9.

[4] Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., Liu, Y. and Chen, J., 2023. *Detrs beat yolos on real-time object detection*. arXiv preprint arXiv:2304.08069.

[5] Fang, Y., Zhan, B., Cai, W., Gao, S. and Hu, B., 2019, July. Locality-constrained spatial transformer network for video crowd counting. In 2019 IEEE international conference on multimedia and expo (ICME) (pp. 814-819). IEEE.