RailRelax
* Enhancing train travel comfort

Prof. Lifna C.S
Assistant Professor, Department of Computer Engineering
Vivekanand Education Society's Institute of Technology
Chembur Mumbai 400074
lifna.cs@ves.ac.in

Wafiya Shaikh
Student of Third Year Computer Engineering
Vivekanand Education Society's Institute of Technology
Chembur Mumbai 400074
2022.wafiya.shaikh@ves.ac.in

Anjali Thakrani
Student of Third Year Computer Engineering
Vivekanand Education Society's Institute of Technology
Chembur Mumbai 400074
2022.anjali.thakrani@ves.ac.in

Anisha Shankar
Student of Third Year Computer Engineering
Vivekanand Education Society's Institute of Technology
Chembur Mumbai 400074
2022.anisha.shankar@ves.ac.in

Himaja Pannati
Student of Third Year Computer Engineering
Vivekanand Education Society's Institute of Technology
Chembur Mumbai 400074
2022.himaja.pannati@ves.ac.in

***Abstract*—In highly populated urban areas such as Mumbai, the issue of overcrowding in local trains poses a significant challenge and inconvenience for the millions of passengers who travel. This paper introduces a sophisticated solution that employs the YOLO object detection algorithm to accurately tell the head count of passengers in each train compartment. It utilizes real-time video data from CCTV systems to provide passengers with information about compartment occupancy prior to arrival, enabling them to make informed decisions that optimize boarding and space usage. This also reduces the need for manual video monitoring and increases the efficiency of passenger flow, resulting in a more organized and comfortable journey. The paper signifies a significant advancement in managing public transport systems by using the latest computer vision approach to address the persistent issue of overcrowding in Mumbai's local train network.**

***Index Terms*—[TODO]**

I. Introduction

Mumbai City has over 20 million people within its limits and relies on the suburban railway system to transport 7.5 million passengers to their destinations daily. Local trains, on the other hand, endure the most severe overcrowding, with compartments frequently filled to six times their authorized capacity. Tragic instances have occurred, indicating the dangers that commuters face. According to a 2018 estimate, overloaded trains killed 711 individuals last year. This problem requires immediate, meaningful action. The inefficiency of current systems is also contributing to overcrowding. The Mumbai Urban Transport Project (MUTP) was designed to decongest the city, but it has seen lower-than-expected passenger growth despite spiking demand. The projected growth by 2030 of around 6-8 million again raises other questions and contradictions in terms of preservation of existing transportation infrastructures. The introduction of double-deckers besides the extension of the metro system are suggested solutions that require tremendous investments and consume much time. Therefore, they can not guarantee safety and comfort for the people who commute in this city.

Current methods of passenger-flow management depend on traditional practices and do not have a dynamic method for dealing with overcrowding. This paper presents a novel approach based on advancing computer vision technology, which focuses on the YOLO V4-tiny algorithm to count real-time head counts for the number of passengers in each compartment of the train. Our solution, analyzing live video data coming from CCTV systems, sheds light on compartment occupancy and helps the passenger make rational decisions about where to board. Thus, it optimizes space usage and, hence, enhances passengers' safety.

Moreover, such an approach is much cheaper than that of double-deckers and can be introduced much faster than the latter. Additionally, onboard behavior can be checked with the help of this solution, which deters theft and other types of bad behavior inside the train.

The incorporation of real-time data analysis into the system would help the passengers deal with overcrowding. This technology, when incorporated, will enhance the safety of the passengers while also leaving an open avenue to more efficient management of suburban railway networks which makes the commute of millions of Mumbaikars more organized and comfortable.

II. Literature Survey

The table provides a comparative overview of various research studies on crowd counting and passenger detection in public transport systems, focusing on models like YOLOv3, YOLOv4, YOLOv5, Faster R-CNN, and EspiNet. It highlights the use of diverse datasets such as PAMELA-UANDES, COCO, Hollywood Heads, and Scuthead to evaluate these models. The studies examine parameters like camera setup, environmental adaptability, and object detection accuracy, with performance metrics including precision, recall, F1-scores, and mAP. Key findings show that YOLO-based models, especially YOLOv5, significantly improve detection rates, particularly for small heads in crowded scenes, with detection rates improving from 66% to 78%. Overall, the research underscores the potential of deep learning models to enhance real-time crowd counting and passenger flow management, while also acknowledging challenges related to complex environments and varying object sizes.

TABLE I: Comparison of People Detection and Counting Models

Title	Parameters	Model Name	Evaluation	Inferences
Detecting, Tracking and Counting People Getting On/Off a Metropolitan Train Using a Standard Video Camera	**Dataset:** PAMELA-UANDES dataset, consisting of 348 video sequences of passengers boarding and alighting. **Variables Studied:** Fare collection method (card-based vs. none), Vertical gap between platform and vehicle, Door widths (800 mm and 1600 mm), Passenger density inside vehicles and on platforms. **Camera Setup:** Multiple cameras capturing video from various angles, though not synchronized or calibrated.	EspiNet, Faster R-CNN, YOLOv3	**Performance Metrics:** 1. F1 Score: Achieved above 95% for counting people across 322 video sequences. 2. Detection performance evaluated using the Multi Object Tracking (MOT) Challenge Development Kit. 3. Comparison of detection capabilities among the three models, with the best achieving an F1 score close to 90%. 4. Multi Object Tracking Accuracy (MOTA) around 80% for the benchmark trackers evaluated.	The research underscores the critical role of accurate people detection and tracking in enhancing public transport efficiency and safety. It suggests that advanced computer vision techniques can improve passenger flow management, potentially increasing public transport usage. Ongoing research is essential to address challenges related to varying environmental conditions and passenger behaviors.
Passenger Detection and Counting during Getting on and off from Public Transport Systems.[1].	**Detection Techniques:** Utilizes histograms of oriented gradients and supports vector machines. **Environmental Adaptability:** Designed for both indoor and outdoor scenarios. **Datasets Used:** Experiments conducted using PAMELA metropolitan train and bus datasets.	Computer Vision-Based Passenger Detection and Counting System	**Accuracy:** Achieved detection accuracies between 91.2% and 86.24%. Counting Errors: Relative counting errors ranged from 10% to 13%.	Performance: The system demonstrated robust performance under various conditions, including occlusion and illumination challenges. Automation: Successfully automates the counting process, reducing reliance on manual methods.

Continued on next page

Title	Parameters	Model Name	Evaluation	Inferences
An Improved Architecture for Automatic People Counting in Public Transport using Deep Learning [3].	The datasets used were obtained from the transport system of the city of Pereira.	Yolov3-tiny, Yolov4, Yolov4-tiny, Yolov8	Yolov3-tiny: mAP - 33.1, Frames per second - 220 Yolov4: mAP - 43.5, Frames per second - 65 Yolov4-tiny: mAP - 22, Frames per second - 443 Yolov8: mAP - 50.2, Frames per second - 500 Input image size (pixels) - 640 Precision - 96.8%, Recall - 92.0%, F1-score - 94.4%, Confidence - 95%	The proposed architecture represents a significant advancement in automatic passenger counting systems for public transport, leveraging deep learning and computer vision technologies. Its high performance metrics suggest it could greatly enhance the management and planning of transportation resources, although practical implementation considerations such as camera placement and computational demands must be addressed.
Human Detection And Crowd Counting Using YOLO	**Datasets used:** COCO (Common Objects in Context) dataset, VOC (Visual Object Classes) dataset, ImageNet dataset.	Two models have been used for human detection and crowd counting: Faster R-CNN and SSD, Object detection: YOLOv8 model. Model was trained on a NVIDIA GTX 1650 Ti GPU with 4 GB memory.	-	YOLO is highly effective for human detection and crowd counting, delivering speed, efficiency, and state-of-the-art performance. It faces challenges in complex scenes, generalization to new domains, and detecting small or distant objects. These limitations can be mitigated with preprocessing or post-processing techniques and careful application in specific contexts. YOLO remains a reliable method for real-time detection, especially in low-power applications, offering significant advantages in these areas.

Continued on next page

Title	Parameters	Model Name	Evaluation	Inferences
Crowd Counting Using Deep Learning Based Head Detection	**Datasets:** The datasets used are Casablanca, Hollywood heads dataset, Scuthead dataset, Merge dataset. Backbone of YOLO v5 uses CSPNet to extract rich and meaningful features from an input image. The FPN and PAN module is used as a neck in YOLO v5 to generate feature pyramids. It uses anchor boxes to construct the final vectors consisting of the output class along with class probabilities, objectiveness scores, and parameters defining the boundary box coordinates.	YOLOv5	The mAP of Casablanca is high i.e. 83%. Hollywood heads mAP is 70.9% and the Scuthead dataset contains all small heads so its mAP is very low i.e. 6%. The mean average precision is 83% when the step size is 100 and the precision obtained is 95%.	The use of YOLOv5 improves accuracy in crowd counting via head detection, particularly in overcrowded scenes. YOLOv5 achieved a detection rate for small heads of 78%, up from 66% using Fast RCNN. Accuracy on the Hollywood Heads dataset increased from 56% to 75%, and the mAP for the Scuthead dataset improved from 66% to 78%. The research resulted in a stabilized pre-trained model capable of detecting heads of various sizes across multiple datasets.

III. Proposed System

To address the issue of train passengers moving easily within a congested train compartment, we have proposed a formal system based on live video analytics and machine learning. Project initiation is done through live video capture from the cameras set inside the compartments to monitor the passengers. In the detection phase, every video frame is forwarded with the applied model YOLOv4-tiny (You Only Look Once) to detect people and surround them with a bounding box. This finally enables the estimation of the count of passengers in specific segments.

The system classifies people density into three categories: High, Medium, and Low-by the count of people it detected within a frame. The process is embedded in a mobile application which is also equipped with train maps, communication features, and alarms for updates. Lastly, this intends to inform the passengers of the level of crowding within the compartments at any given time so that seated passengers might opt to move to less occupied areas and make their trips more enjoyable and productive.

Fig. 1 represents the abstract pipeline structure.

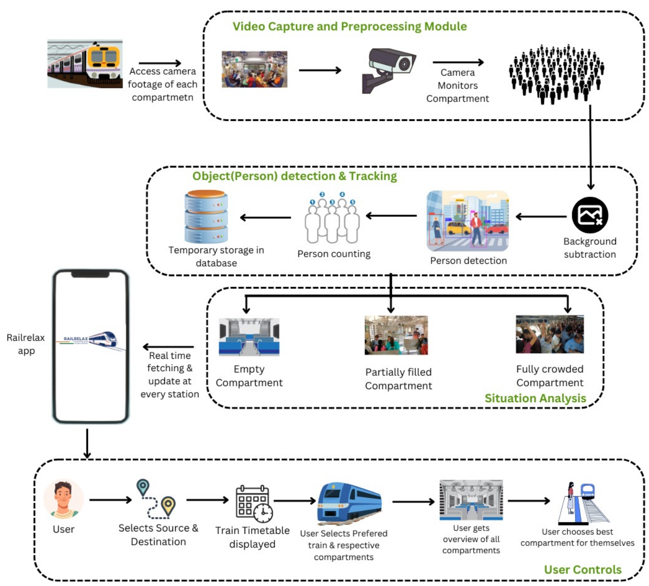

Fig. 1. Flow of the data.

A. Collection of Dataset

For training and testing the YOLOv4-tiny model,we collected annotated images through the Roboflow platform. A total of 1057 images were gathered for the training process. These images were sourced from a combination of pre-existing datasets on Roboflow. Since the annotation labels varied across datasets (e.g., some used ”person,” while others used ”people”),for accurate results the labels are consolidated into a single class labeled ”person” to maintain consistency in the model.

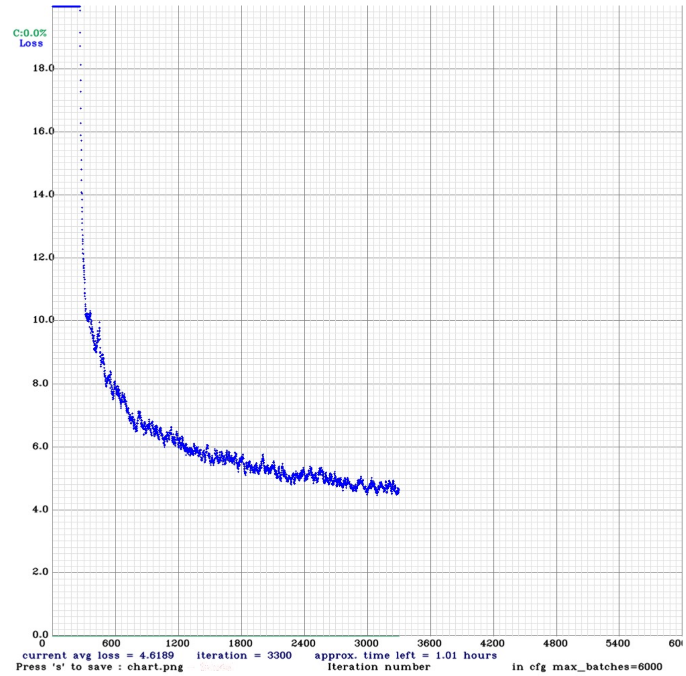

Fig. 2. Model Training Results : Loss vs Batch Size Graph

B. Data Preprocessing

The data preprocessing phase is designed to transform raw video data into a format suitable for efficient and accurate model training. This involves several key steps that ensure both the quality and manageability of the input data.

- Frame Extraction: Each video stream is broken down into individual frames for easier processing.
- Normalization Resizing: The frames are resized and normalized to reduce computational complexity without compromising accuracy.
- Background Subtraction: Techniques like Gaussian Mixture Models (GMM) or deep learning-based methods are used to separate passengers from the background, detecting moving objects.

C. Training

For the proposed system, several variations of the YOLO model were tested to determine the most effective approach for real-time crowd density classification in train compartments. As shown in Table 1, we experimented with YOLOv4-tiny, YOLOv8, YOLOv11, and YOLO-NAS models, evaluating their performance based on Precision, Recall, and mAP@0.50 metrics. These models were trained on annotated frames extracted from video streams to detect passengers and classify the density levels within each compartment.

Each model brings distinct advantages and trade-offs. For instance, YOLOv4-tiny performed with high precision but lower recall, making it suitable for scenarios requiring fewer false positives but not ideal for situations demanding high sensitivity. YOLOv8, on the other hand, provided a balanced performance, achieving both high precision and recall, making it an optimal choice for real-time applications where both detection accuracy and responsiveness are crucial. YOLOv11 also showed competitive results, particularly in terms of precision, while YOLO-NAS demonstrated high recall, excelling in identifying most objects but with reduced precision.

The results obtained through these experiments allowed us to fine-tune and select the most suitable model for integration into the final system.

TABLE II
COMPARISON OF VARIOUS MODELS TESTED USING VARIOUS EVALUATION MEASURES

Model	mAP@0.50	Precision	Recall
YOLOv4-tiny	0.622	0.84	0.51
YOLOv8	0.782	0.838	0.681
YOLOv11	0.776	0.817	0.683
YOLO-NAS	0.754	0.261	0.844

D. Evaluation

For the crop recommendation model, precision, recall and F1-score were used as evaluation measures.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

For the evaluation of the yield prediction model as well as the price forecasting model, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) were calculated.

$$MAE = \sum_{i=1}^{D} |x_i - y_i| \tag{4}$$

$$RMSE = \sum_{i=1}^{D} (x_i - y_i)^2 \tag{5}$$

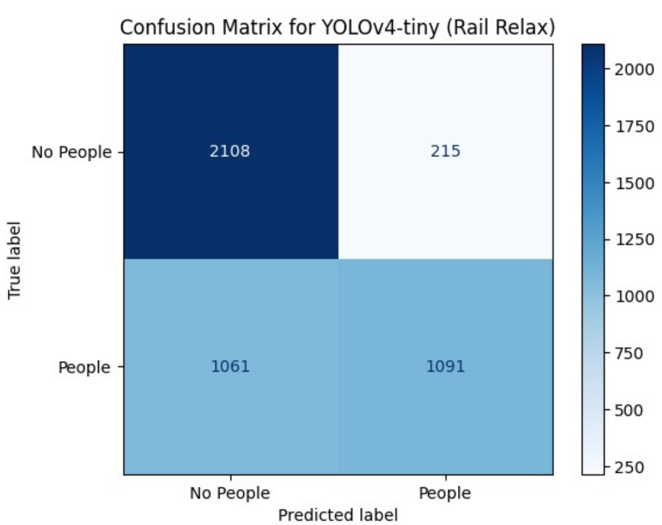

Fig. 3. Confusion Matrix for Yolov4 tiny model

The confusion matrix indicates that the YOLOv4-tiny model performs well with 1091 true positives in detecting individuals, but it struggles with 1061 false negatives, often missing people in images. Additionally, 215 false positives reveal some misclassifications, while 2108 true negatives show a reasonable ability to identify scenes without people.

E. Results

After comparing the performance of various YOLO models for crowd density classification, YOLOv4-tiny was selected as the most suitable model for our project due to its higher precision. As shown in Table 1, YOLOv4-tiny achieved the highest precision of 0.84, making it particularly effective in detecting passengers with a minimal rate of false positives. While YOLOv8 showed a slightly higher mAP@0.50 score, YOLOv4-tiny was preferred because its precision ensures that the detection of passengers is highly accurate, especially in critical real-time applications. YOLOv4-tiny, a lightweight and fast version of YOLOv4, was chosen for its balance between accuracy and computational efficiency. Its ability to process high-resolution video frames quickly and accurately makes it ideal for real-time video-based applications, such as monitoring passenger density in train compartments.

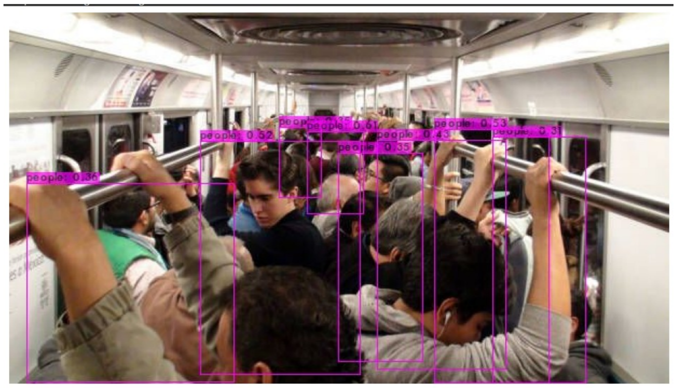

Fig. 4. Model Testing on real crowded Image

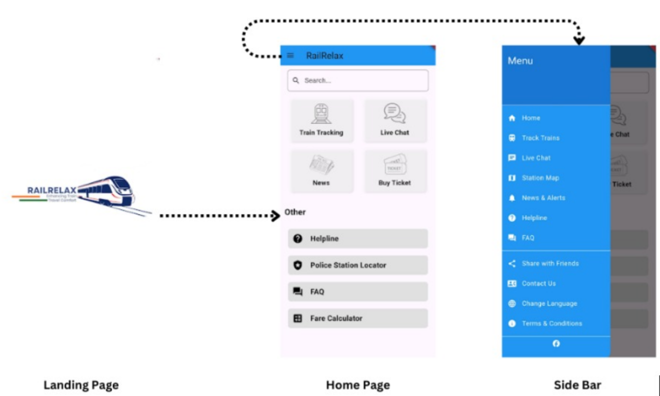

Fig. 5. Initial GUI of Rail Relax App

Fig. 6. Side Bar Page : Features Page of Rail Relax App

Fig. 6 is one of the feature , Helpline Page which shows the Emergency Contacts in case of emergency,

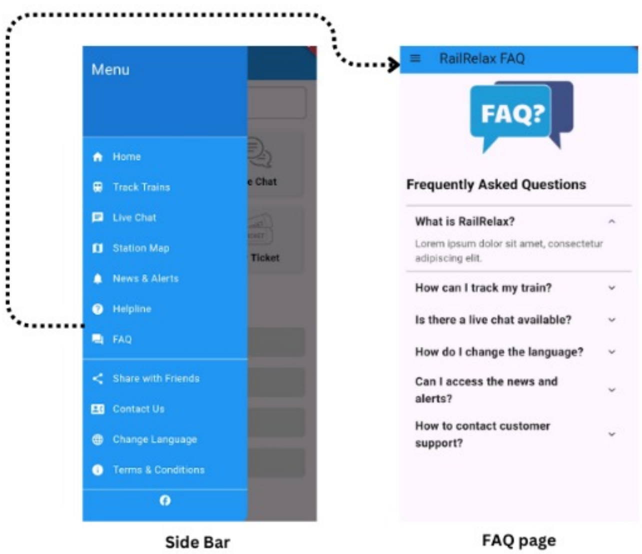

Fig. 7. FAQ Page : Features Page of Rail Relax App

Fig. 7 shows the FAQ page, which will answer queries regarding the navigation of the app, and feature accessing.

IV. Conclusion and Future work

The real-time train occupancy monitoring system presents an innovative approach to improving the efficiency and comfort of public transportation. By leveraging deep learning-based object detection and tracking algorithms, the system accurately counts and monitors passengers in real time. This not only provides passengers with up-to-date information on train occupancy but also enables train operators to optimize crowd management and ensure safety, especially during peak travel times. The system's architecture ensures scalability, real-time performance, and high accuracy, making it a valuable tool in modernizing public transit systems. Through the integration of video-based passenger detection, the system provides a non-invasive solution to track real-time occupancy without requiring any physical interaction from passengers. The benefits of this system include enhanced passenger experience, reduced congestion, better-informed travelers, and improved operational decision-making for transit authorities. While the proposed system demonstrates robust functionality, several areas can be enhanced and expanded in future iterations:

1) **Enhanced Passenger Detection:**
   - Further refine the detection model to better handle edge cases such as detecting passengers in crowded situations, low-light conditions, or partial occlusions.
   - Implement advanced AI models to differentiate between standing and seated passengers and identify luggage or other large objects that may distort passenger counts.
2) **Multimodal Data Integration:**
   - Integrate additional sensor data, such as thermal imaging or RFID, to complement visual-based passenger detection, improving accuracy in extreme weather conditions or during camera malfunctions.
3) **Real-Time Predictive Analytics:**
   - Develop predictive algorithms that forecast train occupancy at upcoming stations, allowing passengers and operators to anticipate crowd levels and make proactive decisions.
4) **Extended Application Scope:**
   - Expand the system's application to monitor station platforms, bus systems, and other public transport modes, creating a unified transportation occupancy management solution.
   - Explore the use of this technology in intercity trains, metro systems, and long-distance buses, where crowd management is equally important.
5) **Edge Computing Deployment:**
   - Explore edge computing to process video streams locally at train compartments, reducing bandwidth usage and enhancing real-time performance by minimizing latency.
6) **Integration with Smart City Ecosystem:**
   - Integrate the system with a broader smart city framework, where real-time occupancy data can inform city planners, optimize traffic flow, and enhance transportation infrastructure.
7) **Passenger Comfort Analytics:**
   - Analyze data on seating availability, overcrowding trends, and passenger movement patterns to develop insights into overall passenger comfort, which can be used to enhance future train designs and service schedules.

References

[1] S. H. Khan, M. H. Yousaf, F. Murtaza, and S. A. Velastin, "Passenger Detection and Counting during Getting on and off from Public Transport Systems," NED University Journal of Research, vol. 2, no. 17, pp. 35-46, 2019. DOI: 10.35453/NEDJR-ASCN-2019-0016

[2] U. Bhangale et al., ”Near Real-time Crowd Counting using Deep Learning Approach,” Procedia Computer Science, vol. 171, pp. 770–779, 2020. DOI: 10.1016/j.procs.2020.04.084
[3] G. Curiel et al., ”An Improved Architecture for Automatic People Counting in Public Transport using Deep Learning,” in 2023 IEEE Colombian Caribbean Conference (C3), Barranquilla, Colombia, pp. 1-8. DOI: 10.1109/C358072.2023.10436152
[4] C. Wiboonsiriruk et al., ”Efficient Passenger Counting in Public Transport Based on Machine Learning,” in TENCON 2023 - IEEE Region 10 Conference, Chiang Mai, Thailand, pp. 214-217. DOI: 10.1109/TENCON58879.2023.10322516
[5] Velastin, S., Fernandez, R., Espinosa Oviedo, J. E. (2020). Detecting, Tracking and Counting People Getting On/Off a Metropolitan Train Using a Standard Video Camera.
[6] Pankaj, A. (2022). A Comprehensive Review on Real-Time Object Detection using Deep Learning Models.
[7] https://researchdata.edu.au/crowd-counting-database/448416
[8] https://www.kaggle.com/datasets/fmena14/crowd-counting
[9] https://mivia.unisa.it/people-detection-dataset/