Social Distancing Monitoring System using Deep Learning

Abhishek Gagneja, Department of Electrical Engineering, Indian Institute of Technology, New Delhi
Dr. Amit Kumar Gupta, DVIEW AI, Sydney, NSW, Australia
Prof. Brejesh Lall, Department of Electrical Engineering, Indian Institute of Technology, New Delhi

Abstract:

We also developed a socially relevant use case application of pedestrian detection with a Social Distancing Monitoring System. We used a pre-trained YOLO V3 and SORT tracker to generate detection and assign them unique IDs for the duration of their visibility. We then use a combination of homographic projection and Euclidean distance measurement to record whether a given pair of IDs are violating the social distancing norms of distance and duration of violation, hence incorporating both guidelines issued by the WHO regarding social distancing.

Keywords: *Social Distancing Monitoring, Deep Learning, Object Tracking, Homographic Projection*

Introduction

As a social application of pedestrian detection, the use case of Social Distancing Monitoring System presented itself to be promising challenge. During the pandemic lockdown, WHO issued strict guidelines regarding social distancing norms to be maintained for public health and safety. These guidelines stated that the distance between two individuals must not be less than 6ft for more than 3sec. Since this is a surveillance problem, the feed to be used must be of a wall mounted camera. The view of such a camera is tilted on an angle, rendering any absolute calculation of distance between two individuals extremely difficult. The distance measurement is then dependent upon the intrinsic parameters of the camera and the area for which it is capturing the feed. These parameters need to be defined beforehand. Also, to incorporate the time aspect of the guidelines, the violations need to be tracked across multiple frames, hence making the need of a fast tracker also apparent. We tasked ourselves with creating a simple yet powerful solution which is lightweight, easy to scale and flexible to incorporate better models as and wherever required.

For the overhead camera feed, we used the Oxford Town Centre Dataset. This dataset was one of the most used ones with respect to the social distancing detection applications. A sample image from the dataset is shown in Figure 1.

Figure 1. Sample Image from Oxford Town Centre Dataset

Importance of Multi Object Tracking and Behaviour Estimation strategies

Multi-object tracking (MOT) finds extensive applications across diverse domains, contributing significantly to fields such as surveillance, autonomous vehicles, sports analysis, and human-computer interaction. In surveillance, MOT plays a pivotal role in monitoring crowded scenes, detecting suspicious activities, and ensuring public safety. Moreover, MOT facilitates real-time object tracking in autonomous vehicles, enabling obstacle avoidance and enhancing navigation capabilities. In sports analysis, MOT aids in player tracking, performance evaluation, and tactical assessment, providing valuable insights to coaches and analysts. Furthermore, it assists in human-computer interaction scenarios, enabling gesture recognition, augmented reality applications, and immersive experiences.

Several tracking techniques have emerged to address the complexities of MOT, leveraging a combination of algorithms and methodologies. Data association methods,

including Kalman Filters, Particle Filters, and Hungarian Algorithms, are widely used to associate object detections across frames, managing occlusions and identity switches effectively. Additionally, deep learning-based approaches utilizing Convolutional Neural Networks (CNNs) and Recurrent Neural Networks have gained prominence due to their ability to learn complex motion patterns and appearance features, contributing to enhanced tracking accuracy and robustness. Hybrid methods combining feature-based tracking, motion models, and appearance-based descriptors have also shown promising results by exploiting both spatial and temporal information for object tracking.

Among simple yet effective tracking algorithms, Simple Online and Realtime Tracking (SORT}) algorithm is a popular multi-object tracker. SORT operates in real-time and exhibits high-speed performance, making it suitable for applications requiring low-latency tracking capabilities. Its simplicity lies in its straightforward design, leveraging a combination of motion prediction, bounding box association, and Kalman filtering for reliable object tracking. Moreover, SORT's modular architecture allows for easy integration with various detection algorithms and association methods, ensuring adaptability and flexibility across different tracking scenarios. Additionally, SORT achieves competitive tracking accuracy while maintaining computational efficiency, rendering it an appealing choice for real-world applications demanding real-time multi-object tracking capabilities.

Behavior monitoring, a subset of computer vision and surveillance, focuses on understanding and analyzing patterns, interactions, and activities of individuals or groups within visual scenes. This field plays a critical role in various applications, including security, crowd management, social science research, and human-computer interaction. Research in behavior monitoring encompasses trajectory analysis, crowd dynamics, anomaly detection, and social interaction modeling, aiming to derive insights into human behavior within diverse environments.

Trajectory analysis forms a fundamental aspect of behavior monitoring, involving the tracking and analysis of individuals' movement patterns. Studies such as the Social LSTM model and attentive GAN for trajectory prediction, exemplify advancements in predicting future trajectories within crowded spaces. These methods leverage deep learning architectures to model temporal dependencies and interactions, enabling more accurate trajectory forecasting and behavior prediction.

Crowd dynamics analysis aims to comprehend collective behavior within groups or crowds. Mehran et al. introduced a Social Force Model for abnormal crowd behavior detection, emphasizing the influence of social forces in simulating crowd movement. Additionally, works by Helbing and Molnár and Moussaid et al. shed light on collective behavior models and crowd simulation, elucidating emergent patterns and behaviors within large groups.

Anomaly detection within visual scenes is crucial for identifying irregular or suspicious activities. Research by Cong et. al. introduced anomaly detection frameworks based on motion patterns, scene context, and abnormal behavior recognition. These methods utilize statistical models and machine learning techniques to identify deviations from normal behavior, aiding in security and surveillance applications.

Models focusing on social interactions and group behavior, such as social force models, graph-based representations, and trajectory clustering techniques, have been explored in works by Pellegrini et. al. and Leal-Taixe et. al. These approaches analyze spatio-temporal relationships, social cues, and interaction patterns, facilitating the understanding of group dynamics and social behavior within visual scenes.

The interdisciplinary nature of behavior monitoring integrates computer vision, machine learning, and social science concepts, fostering advancements in understanding human behavior, social interactions, and collective dynamics within various environments.

Distance Estimation: Homography

To solve the challenge of distance estimation between two pedestrian in a skewed view, we make use of homographic transformation. The model asks the user to mark 4 points on the first frame of the video in order to demarcate the area with reference to which the distance measurement is to take place. If camera intrinsic parameters and particulars of the area of camera deployment are known then this step can be automated. We generate the homographic projection matrix for the selected area using the OpenCV library. Then, using homographic transform, the perspective of the view of the street is mapped on to bird's eye view, represented as a blank image. The blank image is then marked with a grid which is mapped back on to the original image for reference. For distance measurement, each grid cell is estimated to represent a 2m by 2m area. The choice made for the sample image as shown in Figure 2 is of an estimated area of 20m by 16m.

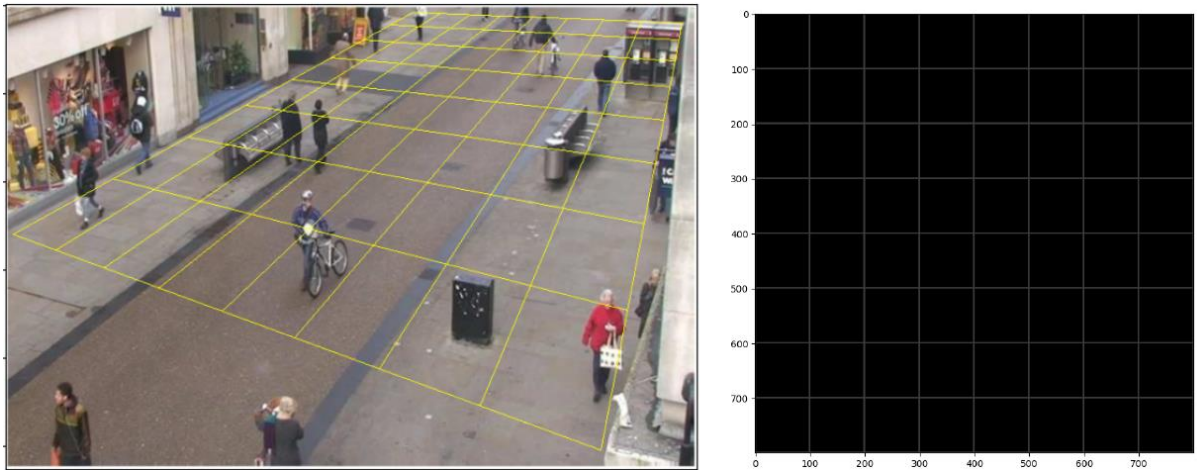

Figure 2. Grid formation using homographic transformation

Now, with the perspective shifted on to a bird's eye view, the distance between any pair of detections mapped on to the projected grid can be evaluated using simple Euclidean distance between them. Since we have assumed the required distance of 6ft (or 2m) be represented by 100 px, therefore any pair of detection at a distance less than 100 px would be considered violators.

To find the detection, we pass the video frames through a YOLO V3 detector. The advantage of using this detector is that it has very little computational cost and still maintains decent accuracy of detection. This is also the model of choice for many other versions of social distancing monitoring systems developed in the past few years. The detections generated are in the conventional bounding box format. However, to map the detection on to the projected grid image, we require detections to be represented as singular points on the grid made on the ground in the original image. For this, we consider only the mean point of the bottom coordinates of the bounding boxes generated by the detector. Now these detections are presentable as singular points and the same is achieved using same the homographic projection matrix which was generated to map the grid.

Once mapped to the grid, the Euclidean distance is measured between all possible pairs of detection and the pairs with less than 100 px separation between them are identified. These detected pedestrians would be called the distance norm violators. In order to indicate their movement through the video, the non-violating detections are represented by a green dot and the violators are initially marked with a red dot. This however, brings us to the aspect of identifying if the violation is happening for more or less than the defined threshold of 3sec.

Timing estimation: SORT tracker

In order to estimate the time for which violation is occurring, we need to identify that the violation is happening between the same two detections across multiple frames. For this, each detection needs to be assigned unique IDs by a tracking system which remains assigned to them till their presence in the field of view. We chose to use SORT tracker for this study, since it is lightweight and works in real-time.

Now, since we have access to unique detections, we keep track of the violations across 90 frames, i.e., 3sec} (the dataset is recorded at 30fps). If the same pair of IDs are violating the distance norms in multiple frames, the color of their representation keeps switching from yellow to red and back, every 15 frames (or 0.5sec). If the violations stop, the color marking the two detections is switched back to green. In case the violations continue for more than the threshold 90 frames, the color of the detections is switched to red for the remainder of their visibility in the video stream. A sample image of detections is shown in Figure 3.

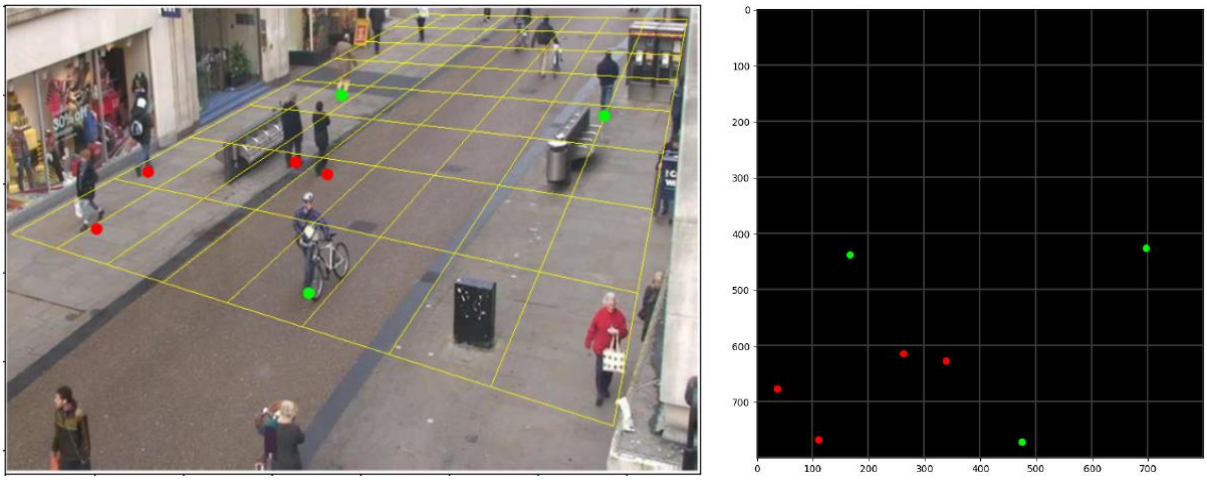

Figure 3. Marking the safe and unsafe detections

Results

The system was used to create a demonstration on the Oxford dataset for DView AI. For the demo, we ran the model on different snippets extracted from the dataset, which is lengthy surveillance camera feed. The idea was to highlight the different challenges faced in the task of social distancing and how the model would tackle them.

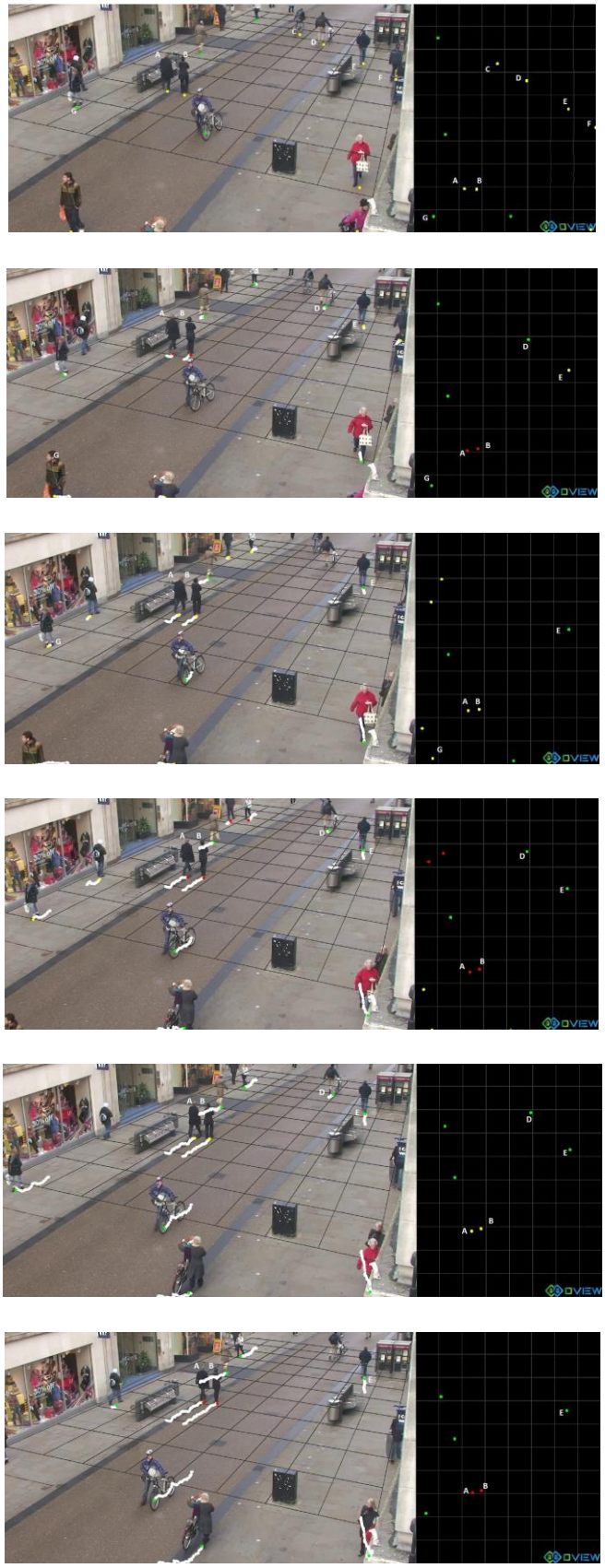

Figure 4. Social Distancing Demo snapshots

The process is better explained with an example as shown in Figure 4. Here we demonstrate the behaviour of the model at 2 fps, i.e. we display every 15th frame of a snippet of the demo. In the first frame, 7 detections are identified and marked with the characters A, B, C, D, E, F and G.

As is visible, detection A and B are walking together violating social distancing norms. Therefore, they are initially marked with yellow dot and the color keeps switching between yellow to red for 6 frames, i.e. 3 seconds. In the 7th frame, and all subsequent frames, these detections stay red having been marked unsafe because of violating social distancing norms.

Detection C, D, E and F are all initially walking at relatively low distance from each other, hence initially all being marked with yellow. Subsequently, the detection C and F walk out of the frame of reference and changing the color of detection to green. We note that in frame 3, the detector fails to identify detection D but recovers the detection in frame 4. After frame 4, the distance between detection 4 and 5 is enough to make them both safe, hence maintaining the green color.

Detection G can be seen in frame 1, to be blocking the visibility of another pedestrian. From the angle of perception, the model fails to identify that pedestrian and hence, the model is unable to identify that detection G is at an unsafe distance. This is identifiable in frame 3, when both the pedestrian and detection G are marked yellow. Since the two are walking away from each other, the model changes the color back to green in subsequent frame.

As is evident, the pedestrian detection accuracy is limited due to use of smaller models. However, the concept used is easily applicable with any detection and tracking system.

Figure 5. A group of more than 2 violators

Figure 5 shows the case where more than 2 detections are found to be violating social distancing at the same time. Since distance between violating instances is measured in pairs, the IDs of

the violating instances are entered in relevant sets (red or yellow), hence creating sets of unique IDs which can then be assigned the respective colors they should.

Upon analysis on the Oxford dataset, we find that our model presents the following metric values.

Metric	**Value**
Detection Precision	68.27%
Detection Recall	73.19%
IDF1-score	69.34%
MOTA	62.48%
Frames per Second (FPS)	25
System Latency	0.28 s
Violation Detection Rate	37.63%
Violation Detection Precision	84.24%
Fraction of violators at risk	57.36%

Here, IDF1 score indicates the retention of the IDs of the tracked objects and MOTA is a metric which evaluates overall accuracy of the tracked individuals. We also present the findings of social distancing violators on the dataset. We observe that there is a 37.63 % social distancing violation rate determined at an 84.24 % precision. Out of all the violations, 57.36 % were found to be for over 3 seconds, hence marking the violators to be at risk of exposure.

Conclusion

The IDs marked with red indefinitely are considered to be at risk of exposure due to violation of Social distancing and those marked green are considered safe.

Therefore, this demo serves as a good proof of concept as an efficient Social Distancing Monitoring System. The solution was used by DView AI, an Australia based startup that deals with AI} based solutions, as a part of their social distancing measurement solution developed during the COVID-19 pandemic.

In the demo solution created for DView AI, the individual tracks of the moving pedestrians across past frames was indicated using white dots. Detectors generate detections based on the entire field of view. However, for this application the detections for which the bottom of the bounding box coordinates lie outside the marked grid area, are ignored from analysis.

The model created here is very lightweight and can be run using extremely simple machines. With marginal latency, the model can also be run on a simple CPU machine with sufficient

memory. The model is flexible enough to incorporate detections from any other advanced systems. Use of better versions of SORT like is also a possibility.

References:

1. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 779–788.
2. J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 7263–7271.
3. J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," arXiv preprint arXiv:1804.02767, 2018.
4. A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," arXiv preprint arXiv:2004.10934, 2020.
5. G. Jocher, YOLOv5 by Ultralytics. 2020. doi: 10.5281/zenodo.3908559.
6. B. Benfold and I. Reid, "Stable multi-target tracking in real-time surveillance video," in CVPR 2011, 2011, pp. 3457–3464.
7. W. Nie et al., "Single/cross-camera multiple-person tracking by graph matching," Neurocomputing, vol. 139, pp. 220–232, 2014.
8. D. Yang, E. Yurtsever, V. Renganathan, K. A. Redmill, and Ü. Özgüner, "A vision-based social distancing and critical density detection system for COVID-19," Sensors, vol. 21, no. 13, p. 4608, 2021.
9. R. Magoo, H. Singh, N. Jindal, N. Hooda, and P. S. Rana, "Deep learning-based bird eye view social distancing monitoring using surveillance video for curbing the COVID-19 spread," Neural Computing and Applications, vol. 33, no. 22, pp. 15807–15814, 2021.
10. Y. C. Hou, M. Z. Baharuddin, S. Yussof, and S. Dzulkifly, "Social distancing detection with deep learning model," in 2020 8th International conference on information technology and multimedia (ICIMU), 2020, pp. 334–338.
11. A. Shukla, I. Garkoti, A. Choudhary, and P. Dhaka, "Social distancing detection using open CV and yolo object detector [J]," International Journal for Modern Trends in Science and Technology, vol. 7, no. 1, pp. 93–95, 2021.
12. K. Bhambani, T. Jain, and K. A. Sultanpure, "Real-time face mask and social distancing violation detection system using yolo," in 2020 IEEE Bangalore Humanitarian Technology Conference (B-HTC), 2020, pp. 1–6.

13. N. S. Punn, S. K. Sonbhadra, S. Agarwal, and G. Rai, “Monitoring COVID-19 social distancing with person detection and tracking via fine-tuned YOLO v3 and Deepsort techniques,” arXiv preprint arXiv:2005.01385, 2020.
14. S. R. C. De Guzman, L. C. Tan, and J. F. Villaverde, “Social Distancing Violation Monitoring Using YOLO for Human Detection,” in 2021 IEEE 7th International Conference on Control Science and Systems Engineering (ICCSSE), 2021, pp. 216–222.
15. A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, “Simple online and realtime tracking,” in 2016 IEEE international conference on image processing (ICIP), 2016, pp. 3464–3468.
16. N. Wojke, A. Bewley, and D. Paulus, “Simple online and realtime tracking with a deep association metric,” in 2017 IEEE international conference on image processing (ICIP), 2017, pp. 3645–3649.
17. N. Aharon, R. Orfaig, and B.-Z. Bobrovsky, “BoT-SORT: Robust associations multi-pedestrian tracking,” arXiv preprint arXiv:2206.14651, 2022.
18. J. C. Gower, “Euclidean distance geometry,” Math. Sci, vol. 7, no. 1, pp. 1–14, 1982.
19. G. Ciaparrone, F. L. Sánchez, S. Tabik, L. Troiano, R. Tagliaferri, and F. Herrera, “Deep learning in video multi-object tracking: A survey,” Neurocomputing, vol. 381, pp. 61–88, 2020.
20. A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, “Social LSTM: Human Trajectory Prediction in Crowded Spaces,” Jun. 2016.
21. A. Sadeghian, V. Kosaraju, A. Sadeghian, N. Hirose, H. Rezatofighi, and S. Savarese, “SoPhie: An Attentive GAN for Predicting Paths Compliant to Social and Physical Constraints,” in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 1349–1358. doi: 10.1109/CVPR.2019.00144.
22. R. Mehran, A. Oyama, and M. Shah, “Abnormal crowd behavior detection using social force model,” in 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 935–942. doi: 10.1109/CVPR.2009.5206641.
23. D. Helbing and P. Molnar, “Social force model for pedestrian dynamics,” Physical review E, vol. 51, no. 5, p. 4282, 1995.
24. M. Moussaïd, D. Helbing, S. Garnier, A. Johansson, M. Combe, and G. Theraulaz, “Experimental study of the behavioural mechanisms underlying self-organization in human crowds,” Proceedings of the Royal Society B: Biological Sciences, vol. 276, no. 1668, pp. 2755–2762, 2009.
25. Y. Cong, J. Yuan, and J. Liu, “Sparse reconstruction cost for abnormal event detection,” in CVPR 2011, 2011, pp. 3449–3456.

26. S. Pellegrini, A. Ess, K. Schindler, and L. Van Gool, “You’ll never walk alone: Modeling social behavior for multi-target tracking,” in 2009 IEEE 12th international conference on computer vision, 2009, pp. 261–268.
27. E. Ristani, F. Solera, R. Zou, R. Cucchiara, and C. Tomasi, “Performance measures and a data set for multi-target, multi-camera tracking,” in European conference on computer vision, 2016, pp. 17–35.
28. K. Bernardin and R. Stiefelhagen, “Evaluating multiple object tracking performance: the clear mot metrics,” EURASIP Journal on Image and Video Processing, vol. 2008, pp. 1–10, 2008.
29. A. El-Nouby, Y. Baydanov, I. Nour Eldin, and A. El-Sallam, “A Review of Multi-Object Tracking,” IEEE Transactions on Intelligent Vehicles, vol. 1, no. 2, pp. 187–197, 2016.
30. A. W. Khan, A. Yilmaz, and V. T. Hopper, “Visual Object Tracking: A Survey,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 10, pp. 1944–1961, 2012.
31. S. Chen, Y. Xu, X. Zhou, and F. Li, “Deep Learning for Multiple Object Tracking: A Survey,” IET Computer Vision, vol. 13, Jan. 2019, doi: 10.1049/iet-cvi.2018.5598.