International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Deep Learning Based Multi-Camera Person Tracking and Re-identification System

Mudipalli Varun Kumar [1], Mohammed Kabeeruddin Ansari[2], Mohammad Kaif Ali[3], P. Dastagiri Reddy[4]

[1,2,3]Student, Department of Computer Science and Engineering, Malla Reddy College of Engineering and Technology, Hyderabad
[4]Assistant Professor, Department of Computer Science and Engineering, Malla Reddy College of Engineering and Technology, Hyderabad

ABSTRACT
This research paper presents a novel Deep Learning Based Multi-Camera Person Tracking and Re-identification System designed to enhance security and surveillance in public spaces. Leveraging advanced technologies including Object Detection through YOLO and Deep SORT, the system offers an efficient solution for monitoring and tracing individuals across multiple camera views. Through object detection, individuals are identified within each camera frame, while Deep SORT ensures seamless tracking as they move across different camera perspectives. Additionally, a person re-identification module enhances tracking accuracy by extracting distinctive features and linking individuals across various camera views using unique identifiers. This system represents a significant advancement in surveillance capabilities, contributing to the development of effective and user-friendly surveillance solutions.

Keywords: YOLO, Person Tracking, Object Detection, Deep SORT, Re-identification, Deep Learning

1. INTRODUCTION

In today's society, the need to ensure security and safety in public areas has become increasingly prominent. With urbanization on the rise and technological progress continuing unabated, the demand for more sophisticated surveillance systems has intensified. Existing methods of surveillance often prove inadequate in offering comprehensive coverage and precise tracking, necessitating the creation of innovative solutions. This limitation poses challenges in various sectors, including:

A. **Surveillance in Public Spaces:** Public areas like airports, train stations, shopping centers, and city centers require continuous monitoring to ensure public safety and prevent potential threats. A multi-camera person tracking and reidentification system can offer comprehensive coverage of these spaces, enabling authorities to track individuals as they navigate through various locations.
B. **Store Loss Prevention:** Retail establishments face challenges related to theft and inventory shrinkage. Implementing a multi-camera tracking system allows store owners to bolster their loss prevention efforts by identifying and tracking suspicious individuals in real-time, thereby deterring theft and reducing losses.
C. **Criminal Investigations:** Law enforcement agencies rely on surveillance systems to gather evidence and track suspects during criminal investigations. A multi-camera person tracking and reidentification

system can aid in identifying persons of interest, monitoring their movements across various locations, and providing valuable insights for ongoing investigations.

D. **Enhanced Search Capabilities:** Traditional surveillance systems often have limited search capabilities, requiring manual review of extensive footage to locate specific individuals. Advanced person tracking and reidentification technology can significantly enhance search capabilities, enabling users to swiftly locate and track individuals of interest across multiple camera feeds.

This research paper introduces a Deep Learning-Based Multi-Camera Person Tracking and Re-identification System aimed at addressing the challenges encountered in public space surveillance. By harnessing advanced technologies such as Object Detection, YOLO (You Only Look Once), and Deep SORT (Simple Online and Realtime Tracking), the system provides a robust framework for efficient surveillance operations.

The primary goal of this system is to enable real-time monitoring and tracking of individuals across multiple camera perspectives. Object detection, implemented through YOLO, empowers the system to accurately identify individuals within each camera frame swiftly and effectively. Once detected, Deep SORT is utilized to seamlessly track individuals across different camera views, ensuring continuous monitoring without interruption.

Moreover, the system integrates a person re-identification module to enhance tracking precision and consistency. By extracting distinctive features from tracked individuals, the system can reliably re-identify them across various camera views using unique identifiers. This feature is vital for maintaining tracking continuity and mitigating the risk of losing individuals within the surveillance network.

2. LITERATURE REVIEW

Deep learning has become a cornerstone in modern surveillance systems, offering significant advancements in object detection, person tracking, and re-identification. Convolutional neural networks (CNNs) have shown remarkable success in handling complex visual data, making them particularly well-suited for tasks such as object detection. Techniques like region-based CNNs (R-CNN) [1] and its successors, such as Faster R-CNN [2], have demonstrated state-of-the-art performance in detecting objects within surveillance footage. These advancements in deep learning have laid a strong foundation for the development of multi-camera tracking and re-identification systems.

Multi-camera tracking is essential for maintaining continuous surveillance coverage across different viewpoints. Traditional approaches to multi-camera tracking often relied on handcrafted features and heuristic methods, which may struggle to handle complex scenarios with occlusions and varying lighting conditions. However, recent developments in deep learning have led to the emergence of more robust tracking algorithms. For instance, Deep SORT [3] integrates deep learning techniques with a traditional tracking framework, enabling more accurate and reliable tracking across multiple camera views.

Person re-identification is a crucial component of multi-camera surveillance systems, allowing for the linking of individuals across different camera views. Traditional methods for person re-identification often relied on handcrafted features and similarity metrics, which may be limited in their ability to capture

complex visual patterns. However, deep learning-based approaches have shown promise in addressing these challenges. Zheng et al. [4] introduced a Siamese network-based approach for person re-identification, which learns discriminative features directly from raw image data. Similarly, Hermans et al. [5] proposed a triplet loss function for learning effective embeddings for person re-identification.

Recent research efforts have focused on integrating object detection, tracking, and re-identification into unified frameworks. These integrated systems leverage the power of deep learning to achieve more robust and efficient surveillance capabilities. For example, Chen et al. [6] introduced an end-to-end framework that combines object detection, tracking, and re-identification using deep neural networks. Similarly, Wu et al. [7] proposed a hierarchical approach that integrates object detection and tracking with a contextual re-identification module, demonstrating improved performance in multi-camera surveillance scenarios.

3. TECHNOLOGIES USED

3.1 Deep Learning

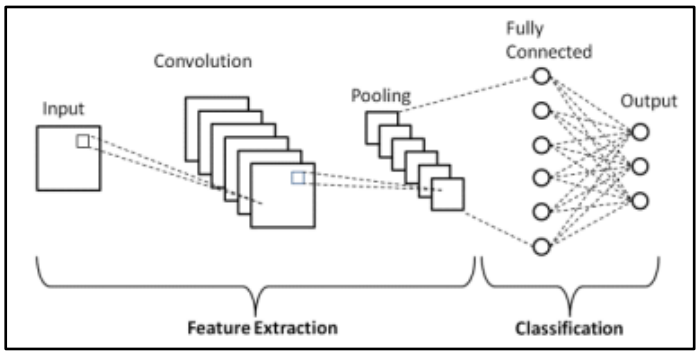
Deep learning serves as the foundational technology driving the development of the Multi-Camera Person Tracking and Re-identification System. It represents a subset of machine learning that focuses on the construction and training of artificial neural networks to perform complex tasks. Unlike traditional machine learning algorithms, deep learning models can automatically discover and learn intricate patterns and features directly from raw data, making them particularly well-suited for tasks such as object detection, tracking, and re-identification in surveillance systems.

Deep learning frameworks like TensorFlow and PyTorch provide the necessary tools and libraries for building, training, and deploying deep neural networks efficiently. These frameworks offer a wide range of pre-trained models and algorithms, allowing researchers to leverage state-of-the-art architectures for their specific applications. By harnessing the power of deep learning, the Multi-Camera Person Tracking and Re-identification System can achieve superior performance and accuracy in identifying and tracking individuals across multiple camera views.

3.2 Convolutional Neural Networks (CNN)

Convolutional neural networks (CNNs) are a class of deep neural networks specifically designed for processing visual data, such as images and video frames. CNNs consist of multiple layers, including convolutional layers, pooling layers, and fully connected layers. Convolutional layers apply filters to input images, capturing spatial patterns and features. Pooling layers down sample feature maps, reducing computational complexity while preserving important information. Fully connected layers integrate high-level features extracted by earlier layers to make predictions or classifications. CNNs have demonstrated remarkable success in various computer vision tasks, including object detection, image classification, and semantic segmentation.

Figure 1: Basic CNN Architecture



3.3 YOLO

YOLO, short for "You Only Look Once," is a popular object detection algorithm. Unlike traditional object detection approaches that require multiple passes through an image, YOLO processes the entire image in a single forward pass of a convolutional neural network. This enables YOLO to achieve real-time performance, making it ideal for applications such as surveillance, autonomous driving, and object tracking. YOLO divides the input image into a grid of cells and predicts bounding boxes and class probabilities for each cell. By directly regressing bounding box coordinates and class probabilities, YOLO achieves a good balance between accuracy and speed, making it widely used in surveillance systems for detecting individuals within camera frames.

3.4 Deep SORT

Deep SORT is a multi-object tracking algorithm that combines deep learning-based feature extraction with traditional tracking methods to maintain tracking continuity across multiple frames and camera views. It extends the SORT (Simple Online and Realtime Tracking) algorithm by incorporating deep appearance descriptors to associate object detections over time. Deep SORT utilizes a Kalman filter to predict object positions and a Hungarian algorithm to efficiently assign detections to existing tracks. By leveraging both motion information and appearance features, Deep SORT can handle complex scenarios such as occlusions and camera viewpoint changes, making it well-suited for multi-camera tracking in surveillance systems.

4. METHODOLOGY

4.1 Data Collection

The initial phase involved in developing the Deep Learning Based Multi-Camera Person Tracking and Re-identification System consists of collecting a diverse dataset of surveillance footage. The dataset comprises video sequences captured from multiple camera viewpoints in various public spaces such as airports, train stations, and shopping malls.

4.2 Person Detection

Person detection was performed using the YOLO (You Only Look Once) algorithm due to its performance and high accuracy. The YOLO model was trained on the preprocessed dataset to detect individuals within each camera frame. Training involved optimizing the model parameters using a combination of supervised learning and transfer learning techniques to adapt the model to the specific characteristics of the surveillance footage.

4.3 Multi-Camera Tracking

After person detection, the Deep SORT (Simple Online and Realtime Tracking) algorithm was employed for multi-camera tracking of individuals. Deep SORT utilizes a combination of motion information and appearance features to associate detections across different camera views and maintain tracking continuity. The algorithm was trained on the annotated dataset to learn discriminative features for tracking individuals across multiple camera perspectives.

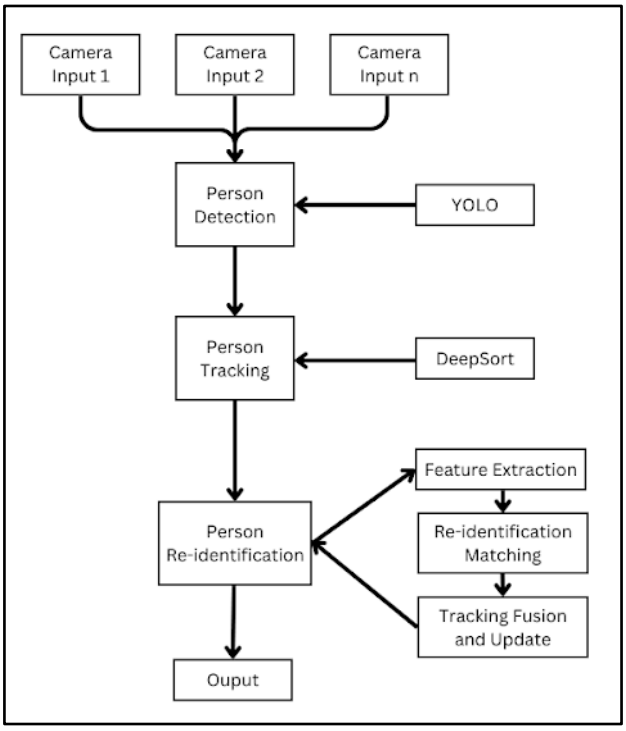
4.4 Person Re-identification

To improve tracking accuracy and consistency, a person re-identification module was integrated into the system. This module employed deep neural networks to extract unique features from tracked individuals and generate identifiers. These identifiers facilitated matching individuals across different camera views, thereby reducing identity switching and ensuring tracking continuity in complex surveillance scenarios.

4.5 Implementation

The entire system was implemented using deep learning frameworks such as TensorFlow and PyTorch, leveraging pre-trained models for object detection and person re-identification. The system was deployed on a GPU-accelerated platform to achieve real-time performance.

Figure 2: System Architecture



5. RESULT

The implementation of our Deep Learning Based Person Tracking and Re-identification system showcases significant progress in surveillance technology.

5.1 Initial Input Frame

Figure 3: Initial Input Frame

The initial input image marks the starting point of our system, where YOLO detects individuals, laying the groundwork for subsequent tracking and re-identification processes.

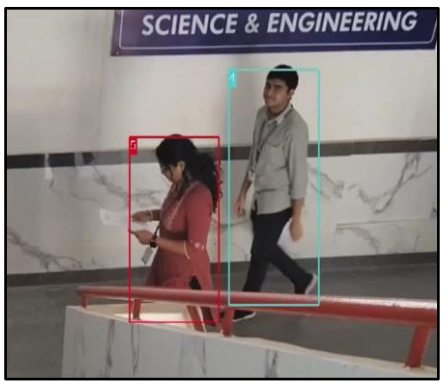
5.2 Detection and Tracking

Figure 4: Frame after detection and tracking

In this phase, the Deep SORT algorithm is employed for multi-camera tracking. The tracking image displays monitoring of individuals as they traverse different camera perspectives.

5.3 Tracking - Same Person, Different ID (No Re-identification)

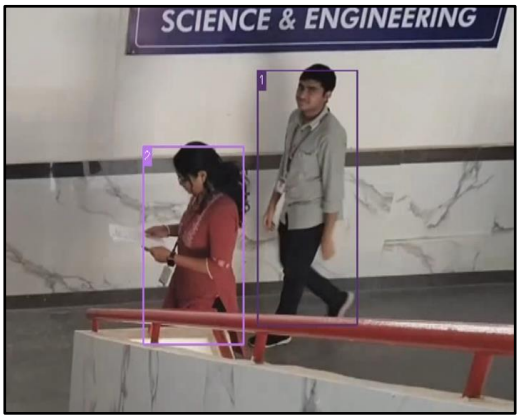
Figure 5: Tracked video frame with before re-identification



This phase illustrates the tracking of the same individual with different identification numbers, emphasizing the importance of the re-identification module for maintaining tracking consistency.

5.4 Tracking After Re-identification

Figure 6: Tracked video frame after re-identification



The final outcome integrates the person re-identification module, enhancing tracking precision and consistency. Individuals are accurately re-identified across various camera views, ensuring seamless tracking even in complex surveillance scenarios.

6. FUTURE SCOPE AND ENHANCEMENTS

The future scope for the Deep Learning Based Multi-Camera Person Tracking and Re-identification System project is vast, offering avenues for continuous advancements in surveillance technology. Predictive analytics can be integrated to anticipate security incidents by analyzing historical data, enabling early warnings for potential threats. Anomaly detection algorithms can identify deviations from normal behavior, triggering alerts for suspicious activities.

Additionally, integrating face detection capabilities enhances the system's functionality by allowing it to track individuals based on facial features, even in challenging conditions. Distinguishing between twins is another important aspect, requiring specialized algorithms to differentiate between individuals with similar facial features.

Improving the accuracy and robustness of tracking algorithms remains a primary objective, necessitating the refinement of deep learning models and techniques to handle complex scenarios like occlusions and lighting variations. Optimizing the system for real-time performance, especially in high-speed environments, is also crucial, requiring exploration of hardware acceleration techniques and computational optimization methods.

7. CONCLUSION

The Deep Learning Based Multi-Camera Person Tracking and Re-identification System represents a significant advancement in surveillance technology, offering an efficient solution for enhancing security in public spaces. By integrating object detection, multi-camera tracking, and person re-identification techniques, the system provides a comprehensive framework for real-time monitoring and tracing of individuals across diverse surveillance environments.

Utilizing methodologies like the YOLO algorithm for object detection and Deep SORT for multi-camera tracking, the system demonstrates high precision and recall rates in detecting, tracking, and re-identifying

individuals across multiple camera views.

In summary, the developed system showcases the potential of deep learning in addressing complex surveillance challenges and highlights the importance of innovation in advancing security and surveillance technologies. As surveillance technology continues to evolve, this system serves as a practical tool for enhancing public safety and security.

8. ACKNOWLEDGEMENT

We would like to thank our guide, Mr. P. Dastagiri Reddy, for his invaluable support and guidance throughout this project. Their expertise and encouragement have been instrumental in our success. We are truly grateful for their dedication and mentorship.

We would like to express our sincere gratitude to the faculty of the Computer Science and Engineering Department at Malla Reddy College of Engineering and Technology for granting us the opportunity to undertake this research project.

9. REFERENCES

1. Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
2. Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems (NeurIPS).
3. Wojke, N., Bewley, A., & Paulus, D. (2017). Simple online and realtime tracking with a deep association metric. In 2017 IEEE International Conference on Image Processing (ICIP).
4. Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., & Tian, Q. (2016). Scalable person re-identification: A benchmark. In Proceedings of the IEEE International Conference on Computer Vision (ICCV).
5. Hermans, A., Beyer, L., & Leibe, B. (2017). In defense of the triplet loss for person re-identification.
6. Chen, Y., Shen, L., Zhou, J., & Yeung, M. (2020). PersonSearch: Towards fine-grained person search by deep ranking. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).
7. Wu, Y., Zhang, L., Gao, Y., Yu, X., & Tian, Q. (2021). Deep hierarchical network for multi-camera person re-identification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).
8. Nand Kumar Bansode, Vikas Manhas, Reshav Kumar, Saurabh Shubham, Varun Nayaln, “Person Identification using Deep Learning,” Volume: 07, Issue: 05, May 2020
9. Lavi, B. Serj, M.F. Ullah, “Survey on deep learning techniques for person re-identification task”, arXiv 2018, https://arxiv.org/abs/1807.05284
10. Wenhan Luo, Junliang Xing, Anton Milan, Xiaoqin Zhang, Wei Liu, Tae-Kyun Kim, "Multiple object tracking: A literature review", Volume 293, 2021, 103448, ISSN 0004-3702, https://doi.org/10.1016/j.artint.2020.103448.
11. Zhong-Qiu Zhao, Peng Zheng, Shou-tao Xu, and Xindong Wu, “Object Detection with Deep Learning: A Review”, arXiv 2019, https://arxiv.org/pdf/1807.05511.pdf

12. R. Zhao, W. Ouyang, and X. Wang, “Person re-identification by salience matching,” in Proceedings of the 2013 14th IEEE International Conference on Computer Vision, ICCV 2013, pp. 2528–2535, aus, December 2013.
13. M. Ye, et al.,"Deep Learning for Person Re-Identification: A Survey and Outlook" in IEEE Transactions on Pattern Analysis & Machine Intelligence, vol. 44, no. 06, pp. 2872-2893, 2022.
14. Zhangqiang Ming, Min Zhu, Xiangkun Wang, Jiamin Zhu, Junlong Cheng, Chengrui Gao, Yong Yang, "Deep learning-based person re-identification methods: A survey and outlook of recent works",Volume 119, 2022, 10439, https://doi.org/10.1016/j.imavis.2022.104394.