IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 14 June 2024, accepted 10 July 2024, date of publication 15 July 2024, date of current version 30 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3428438

RESEARCH ARTICLE

Pedestrian Group Re-Identification and Trajectory Prediction Through Zone-Based Clustering

MINGZUOYANG CHEN, SHADI BANITAAN, (Member, IEEE), AND MINA MALEKI

Department of Electrical and Computer Engineering and Computer Science, University of Detroit Mercy, Detroit, MI 48221, USA

Corresponding author: Mingzuoyang Chen (chenmi7@udmercy.edu)

ABSTRACT Pedestrian trajectory prediction is a critical aspect of computer vision, aimed at predicting a pedestrian's future locations by analyzing their past movements. Traditional trajectory prediction models primarily focus on individuals, which can be challenging in densely populated areas due to occlusions. These occlusions not only complicate the re-identification of pedestrians once they reappear but also increase processing times due to the more complex procedures and the greater number of objects involved. However, a common observation is that pedestrians often travel in groups. This insight led us to propose a novel approach, predicting the future trajectories of pedestrian groups instead of individuals. This strategy effectively addresses the complexities of predicting movements in crowded environments and the issues related to pedestrian occlusion. In this work, we introduced a distinctive methodology for identifying pedestrian groups, re-identifying them, and predicting their future trajectories. Our approach, unlike traditional state-of-the-art re-identification and trajectory prediction methods of individual pedestrians, focuses on re-identifying pedestrian groups and predicting their future trajectories while emphasizing processing time reduction with great accuracy. The process started with object detection to ascertain pedestrian coordinates. Subsequently, a zone-based clustering method was employed to form groups. Following this, a specific group re-identification was utilized to construct continuous trajectories for these groups, rather than for individual pedestrians. Finally, the group trajectory prediction technique was applied to estimate the future movements of these groups. Both the object detection and group detection methods were applied every five frames to generate these trajectories. The effectiveness of our approach has been validated using several evaluation metrics, including Average Displacement Error (ADE), Final Displacement Error (FDE), Cumulative Matching Characteristics (CMC) scores, IDF1 scores, and IDs, all assessed using the MOT17 dataset. These evaluations not only confirm the practicality and accuracy of our method in predicting pedestrian trajectories but also highlight its efficiency, with a reduction in processing time by 7.6% compared to individual trajectory prediction. This efficiency demonstrates the potential of our method for real-time applications and underscores its capacity to prevent accidents.

INDEX TERMS Hungarian algorithm, human view camera, LSTM, re-identification, trajectory prediction, zone-based group detection.

I. INTRODUCTION

In recent times, automated vehicles have gained significant importance due to their potential in reducing vehicle accidents and protecting human lives [1]. Beyond enhancing safety, these vehicles also promise to elevate the efficiency and convenience of daily life by increasing driver comfort and decreasing workload [2], [3]. The integration of various automotive technologies into automated vehicle systems necessitates effective communication not only with drivers

The associate editor coordinating the review of this manuscript and approving it for publication was Mohammad J. Abdel-Rahman.

but also with other road users. In this context, the recognition of pedestrian behavior emerges as a crucial component of these technologies, playing a pivotal role in augmenting system efficiency [4], [5].

Anticipating pedestrian behavior presents a significant challenge within the automotive driving domain. Modern methodologies for autonomous driving break down the intricate process into four sequential stages: object detection, object tracking, trajectory prediction, and path planning [6], [7], [8]. Initially, autonomous vehicles employ object detection and tracking techniques to understand the behavior of pedestrians. Following this, trajectory prediction models are used to forecast their future movements. At last, vehicles utilize the path planning module to formulate driving decisions. While recognizing and monitoring road objects is pivotal, discerning and prognosticating their imminent actions, such as trajectories, is equally crucial.

Pedestrian trajectory prediction holds significant importance in computer vision, with the objective of forecasting a pedestrian's future locations based on their historical path. The trajectory in these methods represents the continuous positioning of a pedestrian across video frames [9]. Current trajectory prediction models primarily use the past positions of pedestrians as a key feature, yielding high accuracy in cases of linear motion patterns. Nevertheless, to accurately forecast future pedestrian behavior, especially from vehicle-mounted cameras, additional features are often necessary [10], [11]. In response, two main types of trajectory prediction methods are evolving: model-based methods and LSTM-based methods [12], [13], [14]. Model-based approaches depend on handcrafted parameters, like pedestrian interactions, for trajectory forecasting [15], while LSTM-based approaches focus on analyzing pedestrian movement patterns or include data from nearby entities during their training [16], [17], [18]. The choice between LSTM-based and model-based techniques generally hinges on the specific data available. Some strategies are adopting a hybrid approach, combining both methods to improve the accuracy of predictions.

A key issue arises in linking object detection results across continuous frames to create object trajectories. The object detection method may not consistently assign the same ID to an object across successive frames [19]. This challenge introduces the need for re-identification (Re-ID). Re-ID technology is tasked with recognizing the same individuals across images captured at different times [20], [21]. The methodologies of Re-ID can vary, typically relying on either the similarity of image data within bounding boxes or spatial distances to establish connections. This divergence gives rise to two main categories of Re-ID: supervised and unsupervised learning [22], [23]. The former concentrates on precise re-identification across various cameras [24], [25], while the latter is more commonly employed in tracking scenarios. Once we successfully identify the same object over a defined time span, we can construct continuous trajectories, setting the stage for accurate trajectory prediction.

However, given the multifaceted behaviors of pedestrians, this remains a formidable challenge. Many external factors, such as obstacles on the road and the presence of other individuals, play a critical role in influencing a pedestrian's trajectory. Additionally, predicting the movements of pedestrians in crowded settings further complicates the task [26], [27], [28].

Since pedestrians often walk in groups, leveraging group information instead of focusing solely on individuals can be advantageous in these scenarios [29], [30], [31]. However, most existing research, such as Mei et al. [24] and Chen et al. [25], seeks to use group information to aid in the analysis of individual pedestrians. Our approach differs from these studies by working entirely with groups. In this work, we introduced a concept named pedestrian group trajectory prediction. We first use the object detection method to find pedestrians and forward detected pedestrian bounding box coordinates to the zone-based group detection method. Then, the grouped coordinates were linked with prior grouped results by our group Re-ID method to form continuous trajectories for pedestrian groups. The next step involved applying the group trajectory prediction algorithm to predict the movements of these groups. Lastly, we utilized the Cumulative Matching Characteristics (CMC), IDF1 scores, and IDs for performance evaluation on the group Re-ID method, while the Average displacement Error (ADE) and Final Displacement Error (FDE) metrics for performance evaluation on group trajectory prediction.

The main contributions of this work are as follows:

- We proposed a comprehensive zone-based pedestrian group trajectory prediction method that included object detection, group detection, group re-identification, and group trajectory prediction. This method addressed the challenge of predicting the future behavior of occluded pedestrians, leading to more accurate trajectory prediction for these pedestrians.
- We demonstrated that predicting pedestrian groups' future trajectories was more time-efficient than predicting individual pedestrians, reducing the prediction processing time by 7.6%. This improvement could enable autonomous vehicles to react more quickly and help prevent accidents.
- We introduced a zone-based pedestrian group re-identification method to consistently identify the same groups over continuous frames. This approach achieved a CMC rank-1 score of 86.51%, an IDF1 score of 95.82, and 58 IDs on the MOT17 dataset, resulting in more accurate trajectory data for future predictions.

The rest of the paper is structured as follows: Section II describes the background of pedestrian re-identification and trajectory prediction. The method details are described in Section III, followed by the result based on the evaluation metrics in Section IV. Finally, Section V discusses the conclusions and future work.

II. RELATED WORK

Trajectory prediction plays a crucial role in the realm of autonomous driving, where it is employed to forecast the movements of various objects. However, the trajectory prediction method typically lacks the capability to independently detect and re-identify objects. Consequently, it is essential to integrate supplementary processes like object detection and re-identification to effectively generate continuous and reliable trajectories.

A. RE-IDENTIFICATION

Re-identification is necessary and irreplaceable once we wanna keep the focus on the same object. This process could help us identify the same object throughout the entire video sequence. The goal of Re-ID is to determine whether the specific object appeared in another camera or even the same camera at different frames. This methodology is normally needed due to the occlusion, low-image resolutions, detection method accuracy, and so on [32].

Pedestrian Re-ID normally focuses on two different situations, one for finding the same pedestrian within different cameras, and the other is to find the same pedestrian in a video or image sequence. This causes the Re-ID methods to be divided into two different types, supervised learning and unsupervised learning. The supervised Re-ID learning methods need the detected object information as the input and make the connection. However, these methods will take a lot of memory once the object number is huge, and the training dataset is also required for more accuracy. Unsupervised Re-ID learning set up the assignment based on the observation between objects.

The field of Re-ID has garnered considerable attention from researchers. For instance, Ning et al. [33] developed a feature refinement and filter network. This network primarily focuses on complete features rather than solely on highly valuable ones, emphasizing person-centric information over background details. It also incorporates a multi-branch attention network to enhance its effectiveness. Similarly, Yan et al. [34] introduced another notable approach with their multi-attention context graph model, specifically designed for group-based re-identification. By utilizing both intra-group and inter-group information, they demonstrated considerable success. Zhu et al. [35] crafted a group context graph neural network designed for graph representation learning. This network acquires contextual data from each member's k nearest neighbors, significantly enhancing the CMC scores of their method. Park and Ham [36] focused on part-level features to discern relationships between different body parts, applying this to a global contrastive pooling method to enhance re-identification, particularly for people with similar attributes. Mei et al. [24] introduced the Siamese Verification-Identification-based Group Retrieval method, which employs minimum distance matching for feature extraction and group retrieval. This approach utilizes the correlation from group retrieval to refine the re-identification process. Similarly, Chen et al. [25] proposed a two-stream attentive network comprising four distinct sub-networks to extract features for both individuals and groups. This method integrates a novel re-ranking algorithm that applies a cosine similarity calculation formula to obtain similarity scores, which are subsequently adjusted by manually defined weights to achieve final matching. In the realm of transformer-based Re-ID, TransReID [37] stands out by encoding images into a sequence of patches and employing a jigsaw patch module for robust feature generation. They also introduced side information embeddings to fortify their framework. Furthermore, Zheng et al. [38] proposed the Group-aware Label Transfer (GLT) strategy, which innovatively uses labels for both training and clustering, thereby boosting the method's effectiveness. Dai et al. [39] developed Cluster Contrast, an innovative approach that processes features at the cluster level, supplemented by a momentum update strategy to maintain feature consistency across continuous frames. This combination significantly improved the performance of their method. In a similar vein, AAformer [40] automated the identification of human parts and obstacles, extracting their features using learnable vectors. The integration of an auto-alignment mechanism then clusters these vectors, further amplifying the efficiency of their methodology. For a comparative perspective, Table. 1 presents CMC rank-1 scores across various Re-ID methods. Here, it's notable that the feature refinement and filter network achieved the highest score on the Market-1501 dataset, while TransReID excelled on the DukeMTMC-reID dataset.

B. TRAJECTORY PREDICTION

Deep learning-based methods have been designed to handle time series data based on their exceptional performance in addressing computer vision challenges. Recurrent Neural Networks (RNN) and its derivatives, such as Long Short-Term Memory (LSTM), are widely used across diverse tasks, with pedestrian trajectory prediction being a prime example. Contrasted with model-based techniques, employing LSTM for trajectory prediction stands out as a more universal, data-centric approach.

Numerous existing studies have explored the application of LSTM for trajectory prediction. As detailed in [41], a model was presented that leverages environmental data for trajectory prediction, utilizing unlabeled image data to adapt unsupervised learning to new environments and tasks. The integration of environmental context with agent motion in this model leads to more accurate predictions. The combination between the given environment and the agent motions resulted in a more reliable prediction. Xue et al. [42] introduced the Local-Velocity-Temporal Attention LSTM model, designed to predict trajectories solely based on observed

TABLE 1. Summary of Re-identification methods.

Algorithm	Dataset	CMC Rank-1	Year	Paper
Feature refinement and filter network	Market-1501, DukeMTMC-reID	98.3/90.3	2020	[33]
Multi-attention context graph model	CUHK-SYSU-Group, DukeMTMC Group	63.2/57.4	2020	[34]
Group context graph neural network	Road Group, DukeMTMC Group	81.7/53.6	2020	[35]
Relation network	Market-1501, DukeMTMC-reID	95.2/89.7	2020	[36]
SVIGR	SYSU-Group, i-LIDS	94.5/46.2	2020	[24]
Two-Stream Attentive network	DukeGroupVid, i-LIDS-VID	82.2/85.3	2021	[25]
TransReID	Market-1501, DukeMTMC-reID	95.2/90.7	2021	[37]
Group-aware label transfer algorithm	Market-1501, DukeMTMC-reID	92.2/82.0	2021	[38]
Cluster Contrast	Market-1501, MSMT17	92.9/62.0	2022	[39]
AAformer	Market-1501, DukeMTMC-reID	95.4/90.1	2023	[40]

TABLE 2. Summary of LSTM-based trajectory prediction methods.

Algorithm	Dataset	ADE	FDE	Year	Paper
Stochastic LSTM	ETH, UCY	0.07	0.127	2020	[41]
LVTA	ETH, UCY	0.46	0.92	2020	[42]
PoPPL	NYGC, Edinburgh	0.396/1.304	0.617/2.314	2020	[14]
CG-LSTM	ETH, UCY	0.43	0.63	2020	[43]
Deep convolutional LSTM network	ETH, UCY	0.61	0.96	2020	[44]
GAN-Tri model	UCY, store	1.895/5.375	3.738/9.350	2021	[45]
SRAI-LSTM	ETH, UCY	0.26	0.53	2022	[46]
STS LSTM	ETH, UCY	0.316	0.568	2023	[47]

data. This model employs dual mechanisms to extract location and velocity features, using a location-velocity attention layer to refine predicted positions and speeds, thereby achieving competitive results. The PoPPL model [14] initially categorizes pedestrian trajectories into various paths, predicts destination regions through an LSTM classification network, and then applies an LSTM-based method for future destination prediction. In contrast, CF-LSTM [43] captures dynamic human-human interactions by extracting features from two consecutive time steps, effectively utilizing past location and velocity data without needing information about other pedestrians. Song et al. [44] developed a model that stores pedestrian features in tensors and employs a convolutional LSTM for trajectory prediction. This combination of tensor storage and convolution enhances learning about pedestrian interactions, significantly improving prediction accuracy. In [45], an innovative deep learning architecture was introduced that prioritizes increased diversity while minimizing linearity. This innovative model integrates LSTM with the Unimodal Generative Adversarial Network for singular trajectory predictions. Moreover, for scenarios requiring multimodal forecasts, it harnesses the potential of the Trimodal Generative Adversarial Network. Peng et al. [46] proposed the Social Relation Attention-based Interaction-aware LSTM. This model features a social relation encoder to decipher relationships and attention between pedestrians based on their past positions, substantially boosting accuracy on prominent pedestrian trajectory datasets. Similarly, the STS LSTM [47] focuses on aggregating spatial, temporal, and spectral data to inform its predictive capabilities. By integrating LSTM, CNNs, and Transformers, this model achieves remarkable accuracy in forecasting pedestrian movements.

Several studies have explored the use of group information as a parameter for predicting future pedestrian trajectories. For instance, Group LSTM [48] developed a coherent filtering approach to assess the similarity among detected pedestrian trajectories, enabling efficient trajectory clustering. It introduced a social pooling layer combined with LSTM to enhance path prediction efficacy and applicability. Zhou et al. [49] advanced this field with the development of a social graph convolutional LSTM neural network. This network is adept at deciphering the intricate relationships between pedestrians and their immediate neighbors. By incorporating an emotion gate, their model effectively filters out extraneous data, enhancing overall efficiency. In another innovative approach, AG-GAN [50] amalgamated coherent group clustering with a global attention mechanism within an LSTM-based Generative Adversarial Network. This integration enables the model to more accurately represent neighbor relationships and focus on subtle, hidden information, leading to a marked increase in method accuracy. Similarly, Bae et al. [51] took a different tack by initially identifying pedestrian groups and then exploring both inter-group and intra-group interactions. This analysis

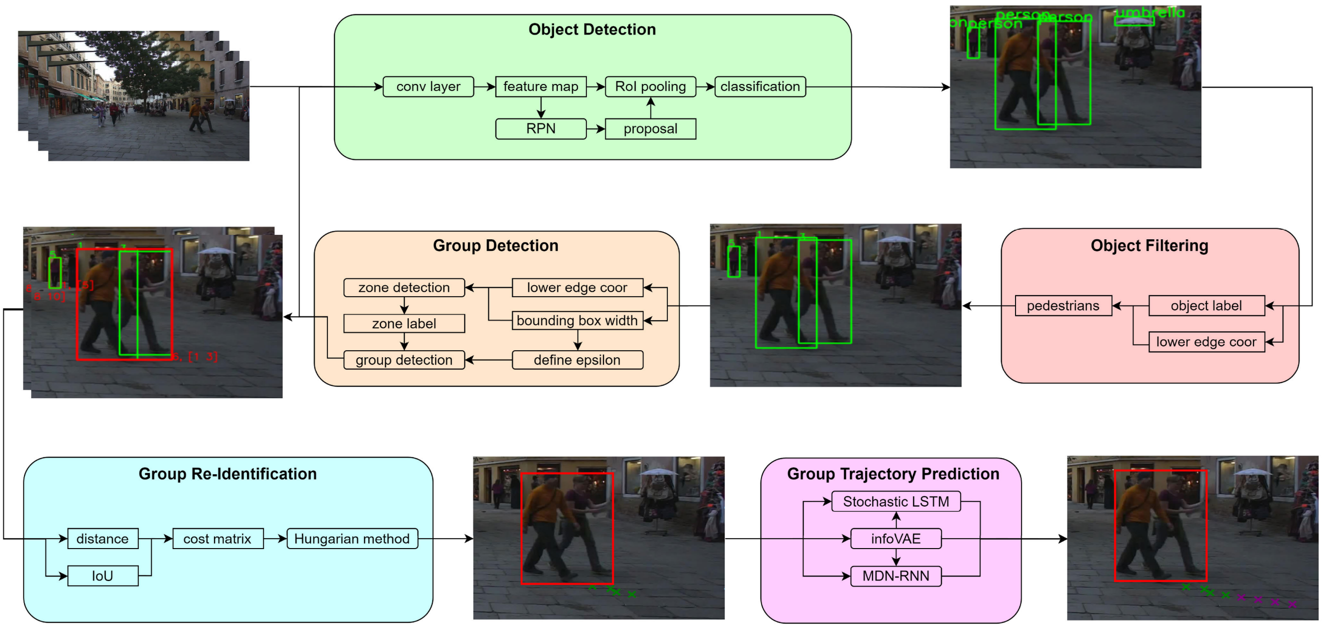

FIGURE 1. **Overall structure of the pedestrian group trajectory prediction.**

significantly aids in understanding the relationships between pedestrian trajectories, thereby boosting model performance. Furthermore, the SG-LSTM [52] model builds on the foundations of Social-LSTM by integrating social awareness into its framework. This feature is particularly effective in facilitating robot navigation in environments populated by groups, ensuring smooth and unobstructed movement. Table. 2 and Table. 3 display the ADE and FDE across various trajectory prediction models, providing a clear comparative analysis of their performance.

C. LIMITATIONS OF EXISTING TECHNIQUES

Previous research on pedestrian trajectory prediction has primarily depended on ground truth coordinates supplied by dataset generators, which often led researchers to overlook the challenge of predicting occluded pedestrians. In methods that integrate re-identification with object detection results, the distance between pedestrians is frequently used as a key parameter, enhancing the accuracy of individual trajectory predictions. However, many trajectory prediction methods do not account for processing time as an evaluative measure and typically utilize overhead cameras, potentially introducing a linear bias into the predictions. Moreover, methods focusing on re-identification generally depend on pre-trained deep learning models to consistently re-identify the same pedestrians across various camera views. Even studies like [24] and [25] have incorporated group information into deep learning networks to address pedestrian occlusion issues and improve the accuracy of individual re-identification methods. Nevertheless, these approaches often fail to leverage the potential of group detection for predicting the trajectories of obscured pedestrians. In contrast, our research underscores the importance of group detection as an essential tool to tackle the challenges posed by non-visible pedestrians. Instead of merely re-identifying and predicting the paths of individually detected pedestrians, our study focused on re-identifying and forecasting the future trajectories of pedestrian groups, which aided in preventing accidents by facilitating quicker responses.

TABLE 3. **Summary of trajectory prediction methods that used group information.**

Algorithm	Dataset	ADE/FDE	Year	Paper
Group LSTM	ETH, UCY	0.34/1.18	2018	[48]
SGC-LSTM	ETH, UCY	0.47/1.01	2021	[49]
AG-GAN	ETH, UCY	0.46/0.90	2021	[50]
GP-Graph	ETH, UCY	0.23/0.39	2022	[51]
SG-LSTM	ETH, Hotel	0.35/0.68	2023	[52]

III. METHODOLOGY

As pedestrians navigate through streets, occlusions, where one pedestrian blocks the view of another, are a common occurrence, as illustrated in Fig. 2. These occlusions present significant challenges in object detection, tracking, and trajectory prediction. Our methodology is designed to tackle this issue by employing pedestrian group detection, which requires object detection to accurately simulate occlusions and demonstrate the effectiveness of our approach. The workflow of our approach is outlined in Fig. 1 and further explicated in the subsequent subsection. Our process begins with object detection, followed by the application of a group detection method to identify pedestrian clusters. Once a sufficient number of sequence frames of object

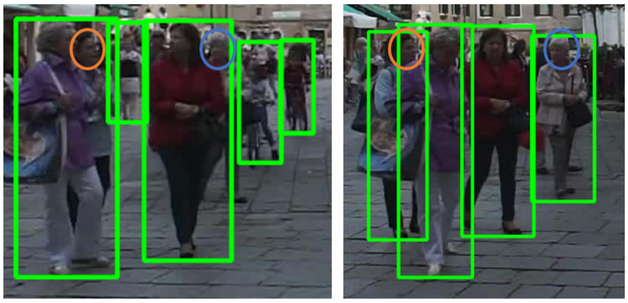

FIGURE 2. Example on a pedestrian being blocked.

coordinates have been accumulated, the group Re-ID method is implemented to construct continuous group trajectories. These group trajectories' coordinates are then input into a pedestrian group trajectory prediction method. The accuracy of our predictions is gauged using metrics like the average displacement error and the final displacement error. In addition, we use the Cumulative Matching Characteristics metric, IDF1 scores, and IDs to evaluate the efficacy of our Re-ID methods. To further demonstrate the time efficiency of our methodology, we also record the processing time.

A. OBJECT DETECTION

In our study, we employed the Faster R-CNN model implemented in PyTorch as our primary object detection method [53]. R-CNN, or region-based convolutional neural network, revolutionized object detection by incorporating selective search to identify regions of interest, followed by the use of CNNs for feature extraction and SVMs for classification. The integration of a Region Proposal Network (RPN) enabled Faster R-CNN to surpass many state-of-the-art methods in terms of accuracy and processing efficiency. Deviating from the traditional selective search, Faster R-CNN segments the network into distinct components, each learning region proposals autonomously. This design allows for real-time performance.

The comprehensive procedure is detailed in Fig. 1. As depicted, the initial step involves forwarding the images to a convolutional layer, which produces a feature map. This feature map is then employed by the RPN to create proposals. These proposals are subsequently processed by the Region of Interest (RoI) pooling layer, which generates feature vectors essential for the classification phase.

In our approach, images are directly input into the Faster R-CNN model, bypassing any form of compression. This model then analyzes the images, subsequently generating outputs that include labels of detected objects as well as the coordinates of their bounding boxes, which are identified by the top-left and lower-right points of each box. This methodology guarantees a high degree of accuracy in identifying pedestrians, setting a robust foundation for the subsequent stages of our process.

B. GROUP DETECTION

The initial step of our group detection involved filtering out non-pedestrian objects using their labels. Following this, we addressed the challenge posed by the object detection method's limited ability to detect distant pedestrians, as objects appear smaller with increased distance. To maintain a focus on generating continuous trajectories, it was necessary to exclude pedestrians that the algorithm consistently failed to detect. For this purpose, we set a threshold at half the height of the image. Any detected pedestrians whose lower edge exceeded this threshold were consequently removed from consideration.

In our previous work, we observed that traditional clustering methods were not fully effective with images taken from a human-view camera. Frequently, individual pedestrians, even if they weren't part of any group, were erroneously clustered together. We attributed this challenge to the depth perception associated with human vision. The perceived distance between pedestrians can vary depending on their proximity to the observer, with the space between them seeming larger as they move closer.

To overcome this, we developed the Z-DBSCAN method. This technique involves dividing the image into several zones and assigning a variable epsilon value to each zone, allowing for more accurate clustering of pedestrians by their apparent size. This approach ensures that pedestrians of similar magnification are grouped together. Implementing zone detection requires a sorted list of the lower-bound coordinates of each bounding box in every image. Given that pedestrians in a group are usually close to each other, their y-coordinates are also closely aligned. Therefore, for effective zone detection, we utilized the widths of the bounding boxes as a key parameter.

The detailed structure and functioning of the Z-DBSCAN method are illustrated in Fig. 1. Initially, we employ the object labels and the coordinates of the lower edge for object filtering. Subsequently, the filtered pedestrians' lower edge coordinates and bounding box widths are used to divide the image into distinct zones. Once the epsilon is determined based on the bounding box width, we proceed to execute the group detection method, ultimately outputting the coordinates of each identified group.

C. PEDESTRIAN GROUP RE-IDENTIFICATION

To construct continuous trajectories for the pedestrian groups we have identified, it is imperative to implement group re-identification. Group Re-ID, distinct from individual Re-ID, seeks to link each individual or group in the current frame with their corresponding counterparts from previous frames. This approach is essential for overcoming the challenges inherent in the traditional pedestrian Re-ID method, especially when it involves the dispersion and later regrouping of pedestrian clusters. In our study, we utilized the Hungarian algorithm for initial matching between consecutive frames. This method requires the construction of a cost matrix, where

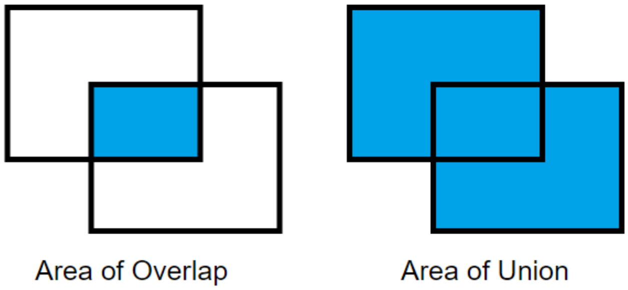

FIGURE 3. **Interaction over Union.**

weights are assigned based on the likelihood of a pedestrian in the current frame correlating with one in the previous frame. To derive these weights, we integrated both spatial distance and Intersection over Union (IoU) metrics into the cost matrix.

Fig. 1 elaborates on the process following the acquisition of continuous frame coordinates for pedestrian groups. Utilizing both distance metrics and IoU values, we generate a cost matrix tailored for the Hungarian method. This method then facilitates the assignment process, which is crucial for identifying the same objects across frames, thereby enabling the formation of continuous trajectories.

Considering our dataset was captured using a human-view camera, the field of view within each image varies significantly. Employing just Euclidean Distance would inaccurately represent the distances between pedestrian trajectory sample points, especially for those closer to the camera. To address this, we differentiated the weights assigned to vertical and horizontal distances, thereby enhancing the precision of our group Re-ID method.

IoU, also known as the Jaccard Index, is widely used in computer vision, especially for object detection tasks. It calculates the ratio of the overlapping area between the selected bounding box and the compared bounding box. This calculation, detailed in Equation 1 and Fig. 3, provides a numerical assessment of how closely the individual or group in the current frame corresponds to their counterparts in the previous frame, crucial for accurate Re-ID and, consequently, for accurate trajectory prediction. By combining both distance and IoU for generating the cost matrix, our group Re-ID process is significantly enhanced in terms of reliability and accuracy.

$$IoU = \frac{AreaofOverlap}{AreaofUnion} \tag{1}$$

However, one inherent limitation of the Hungarian method is that it assigns each object in the subsequent frame to a specific object in the previous frame, based solely on the criteria set by the cost matrix. This can sometimes result in incorrect pairings, where there is no real connection between the pedestrian and the object to which they are assigned. To address this, we established a threshold to exclude these incorrect assignments, especially in cases where the pedestrians are not actually connected. Following this filtration, we focus on those objects that have not been assigned. The next phase of our process is to rearrange these unassigned objects, taking into account their specific weights. This step is crucial as it ensures that any split objects are accurately connected to their appropriate groups from the previous frame.

Additionally, we encounter the issue of short-term occlusion, where pedestrians are temporarily undetectable in certain frames due to obstructions. When these pedestrians reappear and are detected, the absence of their data in the previous frames prompts the traditional system to generate a new trajectory matrix. This not only slows down the process but also impacts the accuracy of trajectory predictions. To overcome this issue, we employed a global coordinate system. This system is designed to keep track of pedestrians who remain unassigned by the Hungarian algorithm or those who are excluded based on the criteria defined in the cost matrix. By effectively associating these unassigned pedestrians with others that reappear in subsequent frames, or by integrating them into nearby groups, we facilitate continuous tracking. This approach is particularly effective in managing short-term occlusion challenges, ensuring a more robust and accurate trajectory prediction process.

Another critical challenge we face is the dynamic nature of pedestrian groupings. Groups may occasionally disperse into smaller clusters or individuals, only to reassemble into a larger group later. Addressing this requires an effective method to continuously generate pedestrian trajectories through these phases of splitting and rejoining. Our strategy begins with assigning a single trajectory to all members of a group. Upon detecting their division into multiple individuals or smaller groups, we replicate the group's historical trajectory for each separate entity and proceed with individual trajectory predictions for each. This approach allows us to maintain trajectory continuity while accommodating the fluidity of pedestrian group dynamics.

The diagrams in Fig. 4 and Fig. 5 illustrate the complex process involved in our pedestrian group Re-ID method. Specifically, Fig.4 details the detection results for each frame corresponding to the trajectory points in Fig.5. The images labeled *a, b, c* and *d* in Fig. 4 depict a scenario where a pedestrian group splits. Initially, these pedestrians are part of the same group but begin to diverge from the third sample frame onwards. As evident from *a* and *b* in Fig.5, during trajectory generation, these pedestrians initially share a common path while they are grouped together. Once they split, we replicate their shared trajectory history and assign it individually to each separated entity, after which trajectory predictions are run independently for each. Conversely, images *e, f, g* and *h* in Fig. 4 demonstrate a contrasting scenario where pedestrian groups or individuals initially separate and merge into a single group over time. In such cases, the combined trajectories from the point of merging, as depicted by the 3rd and 4th trajectory points in images *c* and *d* in Fig.5, are shared among all members of the

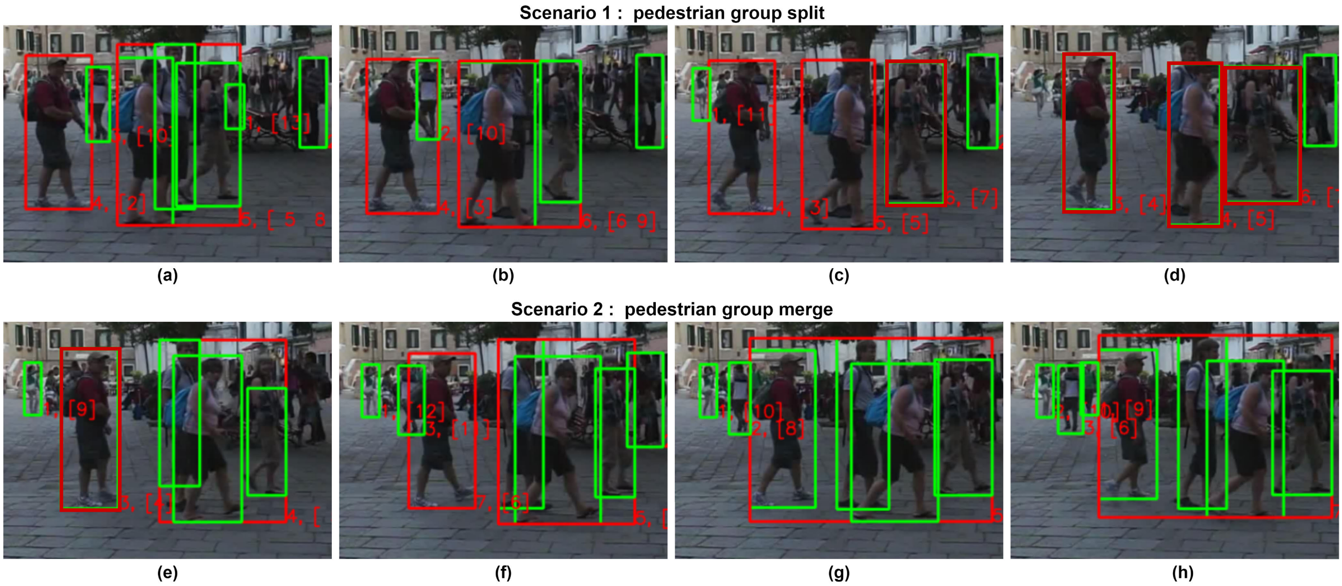

FIGURE 4. **The pedestrian group detection results for two distinct scenarios: 1) In the images *a, b, c* and *d*, which are four consecutive sample frames, we observe that the pedestrians are initially part of the same group but begin to split starting from image *c*. 2) Conversely, in the images *e, f, g* and *h*, also four continuous sample frames, we see that the pedestrians, initially in separate groups, merge into a single group beginning with image *g*.**

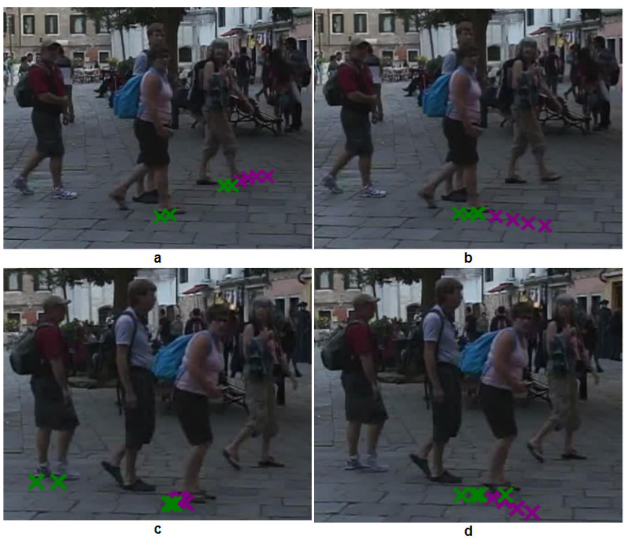

FIGURE 5. **Pedestrian group split and rejoin on two sample frames of the MOT17 dataset.**

newly formed group. This approach ensures continuity and coherence in the trajectory prediction as the group dynamics change.

D. PEDESTRIAN GROUP TRAJECTORY PREDICTION

Long Short-Term Memory (LSTM) is a specialized variant of Recurrent Neural Networks (RNNs), which are commonly employed for processing sequential data [54], [55]. Traditional RNNs often struggle with learning and effectively adjusting weights in their earlier layers. Additionally, as the network depth increases, stability issues can arise due to gradient-related challenges. LSTMs are engineered to overcome these limitations [43], [56]. Their unique architecture enables them to retain and learn from early data across extended sequences, making them particularly suited for tasks like trajectory analysis.

In light of these advantages, we utilized the LSTM network for trajectory prediction in our study. Our methodology aligns closely with the model outlined in [41], characterized by three integral components: an unsupervised model, a supervised model, and a global dynamics model that effectively unifies the first two. This integrated model structure is depicted in Fig. 1. Such an architecture is specifically designed for model-based trajectory prediction, harnessing the synergy of all three components. The process begins with the encoding of image data, followed by connecting information from successive frames. This compiled data is then processed through a conventional stochastic LSTM model, facilitating accurate and effective trajectory forecasting.

The model was initially pre-trained on the ETH and UCY datasets, which enabled it with the capability to analyze a specific trajectory length before predicting subsequent behaviors. However, in our current study, we have shifted our focus from predicting individual pedestrian trajectories to emphasizing the prediction of group behaviors.

Our research is centered on pedestrian group trajectory prediction. Consequently, all the trajectories we generated for each pedestrian were at the group level. After obtaining the coordinates of pedestrian groups for specific frames, we connect these coordinates to those of their preceding frame using our group Re-ID process. A critical aspect to consider is the dynamic nature of pedestrian groups, which may split or merge over time. This could result in instances where some group members exhibit overlapping trajectories in consecutive frames once they are split, particularly when groups divide.

IV. RESULT AND DISCUSSION

In this study, we introduced a framework for forecasting the future trajectories of pedestrian groups. This framework improves the efficiency of re-identifying pedestrian groups, thereby increasing prediction speed by minimizing the number of objects that need to be processed. Furthermore, it enables the prediction of future movements for closely interacting pedestrians by utilizing insights from group dynamics. In this section, we will discuss the parameter settings, followed by a description of the datasets used and the evaluation metrics, including the generation of their ground truth data. We then present experimental results for each step of our methodology, including group detection, group re-identification, and group trajectory prediction, along with an analysis of the associated hyperparameters. Additionally, a comparison with existing work is provided to emphasize the advancements achieved through our approach.

A. PARAMETER SETTINGS

Our methodology begins with the detection of pedestrians in selected video frames from the MOT17Det dataset using the fasterrcnn_resnet50_fpn model provided by PyTorch. This model can be easily executed with the examples provided in its documentation, and various parameters can be adjusted, such as weights, progress, and num_classes. However, these parameters are optional, and for our testing, we used the default values without any specific customizations, except for the weights. We employed the pre-trained weights from COCO_v1.

Following this, pedestrian groups are delineated using a zone-based clustering method known as Z-DBSCAN. This method receives the top-left and lower-right coordinates from the object detection results and outputs group coordinates in the same format. The process begins by generating a sorted list of pedestrian bounding boxes based on their y-axis values, along with their corresponding widths. If the distance between adjacent bounding boxes exceeds the average bounding box width, the average y-axis value is selected as the threshold to divide the image. Subsequently, the parameter ϵ is calculated based on the average width of bounding boxes within each zone, and the default value for *MinPts* is set to 1, assuming a single pedestrian as a group.

Subsequently, group re-identification is utilized to establish continuous trajectories for the identified groups. In this stage, we compare the current group coordinates with the previous ones and employ the Hungarian algorithm to generate continuous trajectory points. To improve the accuracy of group re-identification, we use both spatial and IoU metrics with weights of 0.3 and 0.7, respectively, to construct the cost matrix.

The final step involves utilizing a Stochastic LSTM model to forecast the future movements of these pedestrian groups, which requires a sequence of past trajectories as input to generate future trajectories. The entire model is built upon the BasicLSTMCell from TensorFlow without any activation function and is guided by the maximum mean discrepancy loss function. This method requires an observed length and a predicted length as initial inputs for prediction. Here, we set both values to 4, meaning that 4 continuous trajectory points are observed to predict the next 4 future trajectory points. Once 4 trajectory points are collected, they are fed one by one into the LSTM network, which then outputs the next 4 continuous future trajectory points.

B. DATASET

In this study, we concentrated on using the bounding box coordinates provided by the object detection method to predict the future trajectories of pedestrians. For every individual pedestrian, we constructed a matrix capturing the xy-coordinates across a specific sequence duration. As our objective was to apply this process on streets using a camera mounted on a vehicle, the MOT17Det dataset [57] was utilized in our test. This dataset contains a range of videos, each providing coordinates for all objects detected within them.

The MOT17Det dataset includes images with a resolution of 1920×1080 pixels in JPG format, spanning across multiple videos. Specifically, MOT17-02 features a video comprising 600 frames spread over a duration of 20 seconds. This specific video captures pedestrian movements on the street from the perspective of a stationary, human-eye-level camera. This vantage point closely resembles what would be seen from a vehicle-mounted camera, making it an excellent fit for our study.

C. EVALUATION MATRICS

In order to evaluate our pedestrian group trajectory prediction result, we used Average Displacement Error and Final Displacement Error as the measurements for trajectory prediction [58], [59], [60]. While ADE represents the mean squared error between predicted trajectories and the ground truth, FDE quantifies the discrepancy between the final predicted and actual positions. For both ADE and FDE, lower values are indicative of better performance. Hence, a method that yields low ADE and FDE values is typically regarded as showing a more precise trajectory prediction.

They can be defined in Equations 2 and 3, in which $\mathbf{p}_t$ represents the actual position at time t, $\hat{\mathbf{p}}_t$ represents the predicted position at time t, and T is the total number of trajectory points.

$$\mathrm{ADE} = \frac{1}{T}\sum_{t=1}^{T}\|\mathbf{p}_t - \hat{\mathbf{p}}_t\|_2 \tag{2}$$

$$\mathrm{FDE} = \|\mathbf{p}_T - \hat{\mathbf{p}}_T\|_2 \tag{3}$$

We also employed the Cumulative Matching Characteristic (CMC) curve, an evaluation metric prevalent in Re-ID methods. This curve quantifies the likelihood that objects in the current frame are accurately identified among their top matches, making it a rank-based metric. The process involves generating a list of potential matches for each object, ordered by their similarity scores derived from the similarity matrix. In the context of a CMC curve, the 'rank' refers to the specific position of a gallery item in this list, sorted according to its resemblance to the query object. The effectiveness of the Re-ID method is primarily gauged by the probability of correctly identifying the target gallery item within the top 'k' matches.

To further assess the effectiveness of our Re-ID method, we employed the IDF1 score and IDs, both integral metrics in the MOT (Multiple Object Tracking) evaluation suite, commonly used for evaluating multi-object tracking algorithms. The IDF1 score primarily gauges the accuracy of consistently identifying the same object across different frames. This score is derived using Equations 4. In this equation, 'IDTP' (ID True Positives) represents the count of objects whose IDs match the ground truth. 'IDFP' (ID False Positives) quantifies the objects incorrectly identified, not matching the ground truth. 'IDFN' (ID False Negatives) denotes the number of actual objects either missed entirely or identified with an incorrect ID.

In contrast, the IDs score offers a more straightforward metric than IDF1. It tallies the frequency of ID changes for an object as it moves across various frames. This score is particularly useful for understanding the consistency of ID assignments in tracking scenarios.

$$IDF1 = \frac{2 * IDTP}{2 * IDTP + IDFP + IDFN} \tag{4}$$

However, the IDF1, IDs, and CMC scores, which are metrics essential for accurate re-identification analysis, require ground truth data. The generation of this ground truth data entails identifying the same pedestrian or pedestrian group across consecutive frames within detection results. Due to the comparative simplicity of tracking the same pedestrians or groups consistently, we limited the task to a single annotator. We supplied detection results for 121 consecutive sample frames and instructed the annotator to create a list that links each pedestrian in the current frame to the corresponding individual or group in the previous frame.

D. GROUP DETECTION

In our prior research, we noted that some traditional clustering algorithms struggled to cluster points across varying fields of view effectively. To address this limitation, we introduced advanced zone-based clustering methods which segmented the image into distinct zones. Unlike grid clustering, our approach divided the image only along the horizontal axis, using non-uniform intervals. In this work, we used the Z-DBSCAN as the group detection method. We separate the pedestrians based on the lower coordinates of the pedestrians' bounding boxes. Then, we deployed DBSCAN to cluster pedestrians within each zone. As illustrated in our previous work and Fig. 6, our Z-DBSCAN method for pedestrian clustering showed a marked improvement over traditional DBSCAN. As shown in the figure, the pedestrians were described by a green rectangle bounding box with the detection ID on the top left, and the pedestrian groups were described by a red rectangle bounding box with the group ID on the lower right. The red horizontal lines denoted the zone defined by our Z-DBSCAN method.

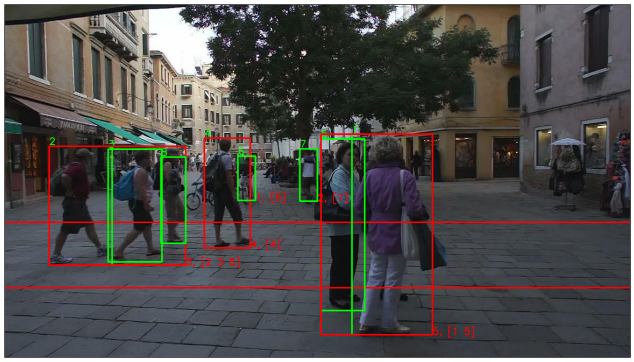

FIGURE 6. Group detection with our Z-DBSCAN method on a sample frame of the MOT17 dataset.

E. PEDESTRIAN GROUP RE-IDENTIFICATION

The primary challenge addressed in this paper is the re-identification of pedestrian groups. To tackle this, we utilize the Hungarian algorithm, a well-known combinatorial optimization algorithm renowned for resolving assignment problems. This method typically requires identifying the most efficient assignment pairing between two sets of entities. The Hungarian method executes assignments based on the aggregate cost, specifically linking each object to its most correlated counterpart. The process begins with the creation of a cost matrix that calculates the costs involved in pairing objects from the previous frame with those in the current frame. Subsequently, we employ the linear_sum_assignment function in Python to establish the assignments. This function is designed to identify the most cost-effective complete assignment based on the provided cost matrix, significantly enhancing the accuracy of the Re-ID method across frames.

In order to make the best match between the two frame detection results, we utilized a combination of IoU and distance measurements, assigning them weights of 0.3 and 0.7 respectively, to construct the cost matrix for the Hungarian algorithm. The IoU would provide a simple and efficient way to help us measure the overlap between pedestrians across different frames. By complementing this with distance measurements, we can achieve more precise Re-ID, particularly addressing scenarios where pedestrians move quickly and have no overlapping areas (IoU) between frames.

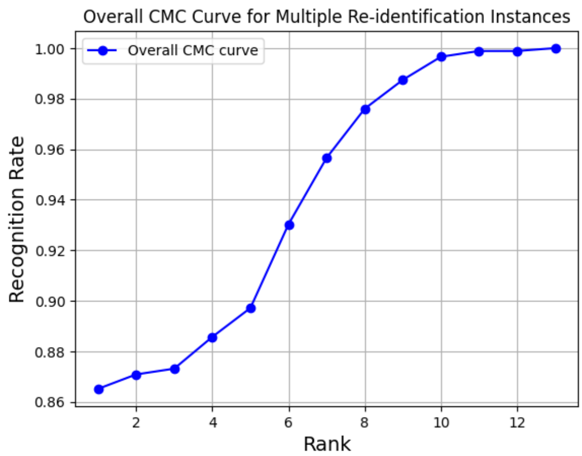

FIGURE 7. Overall CMC Curve on the MOT17 dataset.

TABLE 4. Group re-identification evaluation on MOT17 dataset.

Methods	IDF1	IDs
SAMOT [61]	60.9	1,064
FairMOT [19]	72.3	3,303
ByteTrack [62]	77.3	2,196
Deep OC-SORT [63]	80.6	1,023
FRoG-MOT [64]	77.8	2,244
Ours	95.82	58

While our focus was on leveraging pedestrian group detection and trajectory prediction for forecasting future group trajectories and reducing processing time, we must also address the complexities arising when pedestrian groups split and recombine. To tackle this, we have enhanced the assignment process within the Hungarian algorithm. Initially, after setting up the primary assignment based on the cost matrix, we filtered out assignments with excessively high costs, specifically exceeding 0.35 in our tests. Then, we established connections between unassigned objects and those from the previous frame. This step was crucial for maintaining the continuity of group members who had temporarily split from their original group. To further assess the efficiency of our group Re-ID, we conducted evaluations using CMC scores, IDF1 scores, and IDs, and the results are shown in Fig. 7 and Table 4. Notably, our method demonstrated exceptional results in IDF1 scores and ID switch comparisons. This happened because of the utilization of pedestrian group detection, which limited the number of objects considered throughout the process. Additionally, our method focused on creating continuous trajectories over a sequence of 5 frames, thereby analyzing 120 frames instead of the entire 600-frame dataset. Despite these constraints, the CMC curve effectively illustrated the efficiency of our group Re-ID method and achieved an 86.51% rank-1 score.

FIGURE 8. Pedestrian group trajectory prediction on a sample frame of the MOT17 dataset.

F. PEDESTRIAN GROUP TRAJECTORY PREDICTION

To validate the feasibility and effectiveness of pedestrian group trajectory prediction, it is essential to apply the trajectory prediction method to our pedestrian group detection results. After the pedestrian group Re-ID, we were able to generate the continuous trajectories for the pedestrian groups, whether they were split or combined. If the pedestrian group is not split into several subgroups, we will generate a single trajectory for all of the members, or we will duplicate the trajectory and assign it to all of the subgroups. In our methodology, we chose the Stochastic LSTM for its superior accuracy and applied it to the outputs of the group detection method. Crucially, our processing focused on pedestrian groups rather than individual pedestrians. As depicted in Fig. 8 *a* to *c*. In this figure, we can observe that the trajectories of the pedestrian group in the red bounding box are in green color and the predicted trajectories in purple. This showed that our method proficiently forecasts the future movement patterns of pedestrian groups. We were taking four sample points to predict three future trajectories of the selected object.

The comparison between the individual trajectory prediction and our proposed group trajectory prediction is illustrated in Fig. 9. According to the evaluation metrics we utilized, we could observe that our results show a slight decline in the accuracy of trajectory prediction models when group detection is incorporated. This supports our initial hypothesis that it is feasible to execute pedestrian group trajectory prediction.

Our proposed framework for pedestrian group trajectory prediction also has the potential to reduce the processing time. This is achieved by minimizing the number of trajectories that need to be analyzed. To validate this, we conducted a comparative analysis of the total processing time with and without the integration of group detection and group Re-ID. The comparison result showed our pedestrian group trajectory prediction achieved a 7.6% reduction in processing time, decreasing from 7.17s to 6.62s. This marked decrease underscores the superior time efficiency of our proposed method compared to traditional approaches. While the time saving of 0.55 seconds might seem minimal, it can

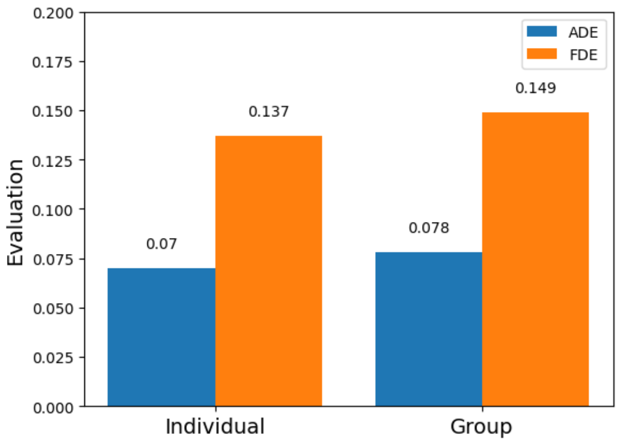

FIGURE 9. **ADE and FDE on our pedestrian group trajectory prediction.**

significantly enhance the safety of autonomous vehicles. For instance, at a speed of 30 MPH, a time difference of 0.55 seconds allows a vehicle to react faster over an additional distance of approximately 7.38 meters. An earlier response in steering could also result in greater horizontal movement, thereby reducing the likelihood of an accident. Importantly, this reduction in time persists even with the use of more powerful computational resources. This approach could be seamlessly integrated into existing methods, maintaining similar accuracy while reducing processing time. However, in [24] and [25], they were more focused on accurately finding the same pedestrians even when they were being blocked in some of the camera views. This required their method to contain a pre-trained deep learning network, which might take a longer time than unsupervised learning-based re-identification methods.

G. STUDY LIMITATION

However, the accuracy of our group re-identification method and group trajectory prediction depends on the quality of object detection and trajectory prediction. If the object detection methods fail to deliver accurate results, we cannot accurately identify the groups. Moreover, if the trajectory prediction method employed after group detection is imprecise, the resulting forecasts will be unreliable. Furthermore, achieving high accuracy in scenarios involving dense crowds of pedestrians can be challenging.

V. CONCLUSION

Our research highlights the importance of changing how we predict pedestrian paths. Instead of focusing only on individuals, we looked at how people often move in groups. This new approach helps us deal with crowded areas where it is hard to see everyone clearly, enabling us to make faster predictions with satisfactory accuracy. Through the integration of object detection, zone-based clustering, and group re-identification techniques, we have established a robust framework for identifying and predicting the future trajectories of pedestrian groups.

Our approach, validated using rigorous evaluation metrics on the MOT17 dataset, has demonstrated superior performance in ADE, FDE, IDF1, IDs, and CMC compared to traditional individual re-identification and trajectory prediction methods. Our methodology achieved a commendable CMC rank-1 accuracy of 86.51%. Moreover, our method exhibits a notable improvement in processing time efficiency, with a reduction of 7.6% compared to conventional approaches. This efficiency enhancement is pivotal for real-time applications, where timely and accurate trajectory prediction is paramount for ensuring pedestrian safety and optimizing traffic flow.

In future work, we aim to further refine our methodology, particularly in pedestrian recognition, and explore its applicability in diverse real-world scenarios to enhance pedestrian safety and urban mobility management.

REFERENCES

[1] Q. Zhang, X. J. Yang, and L. P. Robert, "Individual differences and expectations of automated vehicles," *Int. J. Hum.-Comput. Interact.*, vol. 38, no. 9, pp. 825–836, May 2022.

[2] Q. Zhang, X. J. Yang, and L. P. Robert Jr., "Finding the right voice: Exploring the impact of gender similarity and gender-role congruity on the efficacy of automated vehicle explanations," in *Proc. AAAI Symp. Ser.*, 2023, pp. 219–223.

[3] M. Marcano, S. Díaz, J. Pérez, and E. Irigoyen, "A review of shared control for automated vehicles: Theory and applications," *IEEE Trans. Human-Mach. Syst.*, vol. 50, no. 6, pp. 475–491, Dec. 2020.

[4] H. Detjen, S. Faltaous, B. Pfleging, S. Geisler, and S. Schneegass, "How to increase automated vehicles' acceptance through in-vehicle interaction design: A review," *Int. J. Hum.-Comput. Interact.*, vol. 37, no. 4, pp. 308–330, Feb. 2021.

[5] Q. Zhang, S. Hu, J. Sun, Q. A. Chen, and Z. M. Mao, "On adversarial robustness of trajectory prediction for autonomous vehicles," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 15138–15147.

[6] A. Singh, "Trajectory-prediction with vision: A survey," 2023, *arXiv:2303.13354*.

[7] S. Fadadu, S. Pandey, D. Hegde, Y. Shi, F.-C. Chou, N. Djuric, and C. Vallespi-Gonzalez, "Multi-view fusion of sensor data for improved perception and prediction in autonomous driving," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2022, pp. 2349–2357.

[8] Z. Zhang, J. Gao, J. Mao, Y. Liu, D. Anguelov, and C. Li, "STINet: Spatio-temporal-interactive network for pedestrian detection and trajectory prediction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11343–11352.

[9] A. Rudenko, L. Palmieri, M. Herman, K. M. Kitani, D. M. Gavrila, and K. O. Arras, "Human motion trajectory prediction: A survey," *Int. J. Robot. Res.*, vol. 39, no. 8, pp. 895–935, Jul. 2020.

[10] P. Czech, M. Braun, U. Kreßel, and B. Yang, "Behavior-aware pedestrian trajectory prediction in ego-centric camera views with spatio-temporal ego-motion estimation," *Mach. Learn. Knowl. Extraction*, vol. 5, no. 3, pp. 957–978, Aug. 2023.

[11] L. Neumann and A. Vedaldi, "Pedestrian and ego-vehicle trajectory prediction from monocular camera," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10204–10212.

[12] P. Zhang, W. Ouyang, P. Zhang, J. Xue, and N. Zheng, "SR-LSTM: State refinement for LSTM towards pedestrian trajectory prediction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 12077–12086.

[13] H. Xue, D. Q. Huynh, and M. Reynolds, "SS-LSTM: A hierarchical LSTM model for pedestrian trajectory prediction," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2018, pp. 1186–1194.

[14] H. Xue, D. Q. Huynh, and M. Reynolds, "PoPPL: Pedestrian trajectory prediction by LSTM with automatic route class clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 77–90, Jan. 2021.

[15] Y. Xu, J. Xie, T. Zhao, C. Baker, Y. Zhao, and Y. Nian Wu, "Energy-based continuous inverse optimal control," 2019, *arXiv:1904.05453*.

[16] P. Jia, H. Chen, L. Zhang, and D. Han, "Attention-LSTM based prediction model for aircraft 4-D trajectory," *Sci. Rep.*, vol. 12, no. 1, p. 15533, Sep. 2022.

[17] F. Hui, C. Wei, W. ShangGuan, R. Ando, and S. Fang, "Deep encoder–decoder-NN: A deep learning-based autonomous vehicle trajectory prediction and correction model," *Phys. A, Stat. Mech. Appl.*, vol. 593, May 2022, Art. no. 126869.

[18] A. Zaroug, D. T. H. Lai, K. Mudie, and R. Begg, "Lower limb kinematics trajectory prediction using long short-term memory neural networks," *Frontiers Bioeng. Biotechnol.*, vol. 8, p. 362, May 2020.

[19] Y. Zhang, C. Wang, X. Wang, W. Zeng, and W. Liu, "FairMOT: On the fairness of detection and re-identification in multiple object tracking," *Int. J. Comput. Vis.*, vol. 129, no. 11, pp. 3069–3087, Nov. 2021.

[20] N. Peri, P. Khorramshahi, S. S. Rambhatla, V. Shenoy, S. Rawat, J.-C. Chen, and R. Chellappa, "Towards real-time systems for vehicle re-identification, multi-camera tracking, and anomaly detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 2648–2657.

[21] P. Xu and X. Zhu, "DeepChange: A long-term person re-identification benchmark with clothes change," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 11196–11205.

[22] G. Wu, X. Zhu, and S. Gong, "Tracklet self-supervised learning for unsupervised person re-identification," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 7, pp. 12362–12369.

[23] X. Lin, P. Ren, C.-H. Yeh, L. Yao, A. Song, and X. Chang, "Unsupervised person re-identification: A systematic survey of challenges and solutions," 2021, *arXiv:2109.06057*.

[24] L. Mei, J. Lai, Z. Feng, and X. Xie, "From pedestrian to group retrieval via Siamese network and correlation," *Neurocomputing*, vol. 412, pp. 447–460, Oct. 2020.

[25] L. Chen, H. Yang, Q. Xu, and Z. Gao, "Harmonious attention network for person re-identification via complementarity between groups and individuals," *Neurocomputing*, vol. 453, pp. 766–776, Sep. 2021.

[26] K. Li, M. Shan, K. Narula, S. Worrall, and E. Nebot, "Socially aware crowd navigation with multimodal pedestrian trajectory prediction for autonomous vehicles," in *Proc. IEEE 23rd Int. Conf. Intell. Transp. Syst. (ITSC)*, Sep. 2020, pp. 1–8.

[27] C. Yu, X. Ma, J. Ren, H. Zhao, and S. Yi, "Spatio-temporal graph transformer networks for pedestrian trajectory prediction," in *Computer Vision—ECCV*, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds. Cham, Switzerland: Springer, 2020, pp. 507–523.

[28] X. Shi, X. Shao, Z. Fan, R. Jiang, H. Zhang, Z. Guo, G. Wu, W. Yuan, and R. Shibasaki, "Multimodal interaction-aware trajectory prediction in crowded space," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 7, pp. 11982–11989.

[29] M. Chen, S. Banitaan, M. Maleki, and Y. Li, "Pedestrian group detection with K-means and DBSCAN clustering methods," in *Proc. IEEE Int. Conf. Electro Inf. Technol. (eIT)*, May 2022, pp. 1–6.

[30] M. Chen, S. Banitaan, M. Maleki, and Y. Li, "GDSCAN: Pedestrian group detection using dynamic epsilon," in *Proc. 21st IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2022, pp. 1748–1753.

[31] M. Chen, S. Banitaan, and M. Maleki, "Enhancing pedestrian group detection and tracking through zone-based clustering," *IEEE Access*, vol. 11, pp. 132162–132179, 2023.

[32] M. Ye, J. Shen, G. Lin, T. Xiang, L. Shao, and S. C. H. Hoi, "Deep learning for person re-identification: A survey and outlook," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 6, pp. 2872–2893, Jun. 2022.

[33] X. Ning, K. Gong, W. Li, L. Zhang, X. Bai, and S. Tian, "Feature refinement and filter network for person re-identification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 9, pp. 3391–3402, Sep. 2021.

[34] Y. Yan, J. Qin, B. Ni, J. Chen, L. Liu, F. Zhu, W.-S. Zheng, X. Yang, and L. Shao, "Learning multi-attention context graph for group-based re-identification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 6, pp. 7001–7018, Jun. 2023.

[35] J. Zhu, H. Yang, W. Lin, N. Liu, J. Wang, and W. Zhang, "Group re-identification with group context graph neural networks," *IEEE Trans. Multimedia*, vol. 23, pp. 2614–2626, 2021.

[36] H. Park and B. Ham, "Relation network for person re-identification," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 7, 2020, pp. 11839–11847.

[37] S. He, H. Luo, P. Wang, F. Wang, H. Li, and W. Jiang, "TransReID: Transformer-based object re-identification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 14993–15002.

[38] K. Zheng, W. Liu, L. He, T. Mei, J. Luo, and Z.-J. Zha, "Group-aware label transfer for domain adaptive person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 5306–5315.

[39] Z. Dai, G. Wang, W. Yuan, S. Zhu, and P. Tan, "Cluster contrast for unsupervised person re-identification," in *Proc. Asian Conf. Comput. Vis. (ACCV)*, 2022, pp. 1142–1160.

[40] K. Zhu, H. Guo, S. Zhang, Y. Wang, J. Liu, J. Wang, and M. Tang, "AAformer: Auto-aligned transformer for person re-identification," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Aug. 25, 2024, doi: 10.1109/TNNLS.2023.3301856.

[41] T. Davchev, M. Burke, and S. Ramamoorthy, "Learning structured representations of spatial and interactive dynamics for trajectory prediction in crowded scenes," *IEEE Robot. Autom. Lett.*, vol. 6, no. 2, pp. 707–714, Apr. 2021.

[42] H. Xue, D. Q. Huynh, and M. Reynolds, "A location-velocity-temporal attention LSTM model for pedestrian trajectory prediction," *IEEE Access*, vol. 8, pp. 44576–44589, 2020.

[43] Y. Xu, J. Yang, and S. Du, "CF-LSTM: Cascaded feature-based long short-term networks for predicting pedestrian trajectory," in *Proc. AAAI Conf. Art. Intel.*, 2020, vol. 34, no. 7, pp. 12541–12548.

[44] X. Song, K. Chen, X. Li, J. Sun, B. Hou, Y. Cui, B. Zhang, G. Xiong, and Z. Wang, "Pedestrian trajectory prediction based on deep convolutional LSTM network," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 6, pp. 3285–3302, Jun. 2021.

[45] L. Rossi, M. Paolanti, R. Pierdicca, and E. Frontoni, "Human trajectory prediction and generation using LSTM models and GANs," *Pattern Recognit.*, vol. 120, Dec. 2021, Art. no. 108136.

[46] Y. Peng, G. Zhang, J. Shi, B. Xu, and L. Zheng, "SRAI-LSTM: A social relation attention-based interaction-aware LSTM for human trajectory prediction," *Neurocomputing*, vol. 490, pp. 258–268, Jun. 2022.

[47] C. Zhang, Z. Ni, and C. Berger, "Spatial–temporal–spectral LSTM: A transferable model for pedestrian trajectory prediction," *IEEE Trans. Intell. Vehicles*, vol. 9, no. 1, pp. 2836–2849, Jan. 2023.

[48] N. Bisagno, B. Zhang, and N. Conci, "Group LSTM: Group trajectory prediction in crowded scenarios," in *Proc. Eur. Conf. Comput. Vis. (ECCV) Workshops*, 2018.

[49] Y. Zhou, H. Wu, H. Cheng, K. Qi, K. Hu, C. Kang, and J. Zheng, "Social graph convolutional LSTM for pedestrian trajectory prediction," *IET Intell. Transp. Syst.*, vol. 15, no. 3, pp. 396–405, Mar. 2021.

[50] Y. Song, N. Bisagno, S. Z. Hassan, and N. Conci, "AG-GAN: An attentive group-aware GAN for pedestrian trajectory prediction," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 8703–8710.

[51] I. Bae, J.-H. Park, and H.-G. Jeon, "Learning pedestrian group representations for multi-modal trajectory prediction," in *Computer Vision—ECCV*, S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, Eds. Cham, Switzerland: Springer, 2022, pp. 270–289.

[52] R. Bhaskara, M. Chiu, and A. Bera, "SG-LSTM: Social group LSTM for robot navigation through dense crowds," 2023, *arXiv:2303.04320*.

[53] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32. Red Hook, NY, USA: Curran Associates, 2019, pp. 8024–8035. [Online]. Available: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

[54] G. Xie, A. Shangguan, R. Fei, W. Ji, W. Ma, and X. Hei, "Motion trajectory prediction based on a CNN-LSTM sequential model," *Sci. China Inf. Sci.*, vol. 63, no. 11, pp. 1–21, Nov. 2020.

[55] R. Jiang, H. Xu, G. Gong, Y. Kuang, and Z. Liu, "Spatial-temporal attentive LSTM for vehicle-trajectory prediction," *ISPRS Int. J. Geo-Inf.*, vol. 11, no. 7, p. 354, Jun. 2022.

[56] S. Mikhailov and A. Kashevnik, "Car tourist trajectory prediction based on bidirectional LSTM neural network," *Electronics*, vol. 10, no. 12, p. 1390, Jun. 2021.

[57] A. Milan, L. Leal-Taixé, I. D. Reid, S. Roth, and K. Schindler, "MOT16: A benchmark for multi-object tracking," 2016, *arXiv:1603.00831*.

[58] S. Pellegrini, A. Ess, K. Schindler, and L. van Gool, "You'll never walk alone: Modeling social behavior for multi-target tracking," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 261–268.

[59] A. Mohamed, K. Qian, M. Elhoseiny, and C. Claudel, ''Social-STGCNN: A social spatio-temporal graph convolutional neural network for human trajectory prediction,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 14412–14420.
[60] T. Gu, G. Chen, J. Li, C. Lin, Y. Rao, J. Zhou, and J. Lu, ''Stochastic trajectory prediction via motion indeterminacy diffusion,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 17113–17122.
[61] W. Feng, Z. Hu, B. Li, W. Gan, W. Wu, and W. Ouyang, ''SAMOT: Switcher-aware multi-object tracking and still another MOT measure,'' 2020, *arXiv:2009.10338*.
[62] Y. Zhang et al., ''ByteTrack: Multi-object tracking by associating every detection box,'' in *Computer Vision—ECCV*, S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, Eds. Cham, Switzerland: Springer, 2022, pp. 1–21.
[63] G. Maggiolino, A. Ahmad, J. Cao, and K. Kitani, ''Deep OC-SORT: Multi-pedestrian tracking by adaptive re-identification,'' 2023, *arXiv:2302.11813*.
[64] T. Ogawa, T. Shibata, and T. Hosoi, ''FRoG-MOT: Fast and robust generic multiple-object tracking by IoU and motion-state associations,'' in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2024, pp. 6563–6572.

MINGZUOYANG CHEN received the B.S. and M.S. degrees in electrical and computer engineering from the University of Detroit Mercy, MI, USA, in 2018 and 2019, respectively, where he is currently pursuing the Ph.D. degree in electrical and computer engineering. From 2020 to 2024, he was a Research Assistant in electrical and computer engineering at the University of Detroit Mercy. His research interests include machine learning and image processing.

SHADI BANITAAN (Member, IEEE) received the B.S. degree in computer science from Yarmouk University, the M.S. degree in computer and information sciences, and the Ph.D. degree in computer science from North Dakota State University. Currently, he is working as the Director of computer science and software engineering and he joined as an Associate Professor with the University of Detroit Mercy, in 2013. Before this, he worked as an Instructor at the University of Nizwa, from 2004 to 2009. His research interests include artificial intelligence, machine learning, and software engineering. He is a member of the Association for Computing Machinery (ACM) and the IEEE Computer Society.

MINA MALEKI received the master's degree in software engineering from Tehran Polytechnic University, in 2007, and the Ph.D. degree in computer science from the University of Windsor, in 2014. She is currently an Assistant Professor of computer science and software engineering with the University of Detroit Mercy, MI, USA. Prior to this role, she worked as a Sessional Instructor at the University of Windsor, Canada, and as a SOSCIP TalentEdge Postdoctoral Research Fellow at the Cross-Border Institute (CBI), University of Windsor. Her research interests include machine learning, deep learning, pattern recognition, and data analysis.