Hindawi
Computational Intelligence and Neuroscience
Volume 2023, Article ID 3677387, 10 pages
https://doi.org/10.1155/2023/3677387

Research Article

HAZMAT Vehicle Reidentification in Road Tunnels Based on the Fusion of Appearance and Spatiotemporal Information

Lei Jia,[1,2] Xiaobao Li,[1,2] Wen Wang,[1,2] Jianzhu Wang,[1,2] Haomin Yu,[1,2] Tianyuan Wang,[3] and Qingyong Li[1,2]

[1]*Beijing Key Lab of Traffic Data Analysis and Mining, Beijing Jiaotong University, Beijing 100044, China*
[2]*Frontiers Science Center for Smart High-Speed Railway System, Beijing Jiaotong University, Beijing 100044, China*
[3]*Shenzhen Urban Transport Planning Center Co. Ltd., Shenzhen 518000, China*

Correspondence should be addressed to Qingyong Li; liqy@bjtu.edu.cn

Received 12 June 2022; Revised 9 October 2022; Accepted 11 October 2022; Published 14 February 2023

Academic Editor: Nian Zhang

Vehicles transporting hazardous material (HAZMAT) pose a severe threat to highway safety, especially in road tunnels. Vehicle reidentification is essential for identifying and warning abnormal states of HAZMAT vehicles in road tunnels. However, there is still no public dataset for benchmarking this task. To this end, this work releases a real-world tunnel HAZMAT vehicle reidentification dataset, VisInt-THV-ReID, including 10,048 images with 865 HAZMAT vehicles and their spatiotemporal information. A method based on multimodal information fusion is proposed to realize vehicle reidentification by fusing vehicle appearance and spatiotemporal information. We design a spatiotemporal similarity determination method for vehicles based on the spatiotemporal law of vehicles in tunnels. Compared with other reidentification methods based on multimodal information fusion, i.e., PROVID, Visual + ST, and Siamese-CNN, experimental results show that our approach significantly improves the vehicle reidentification recognition precision.

1. Introduction

Hazardous materials (HAZMAT) could endanger the health and safety of people, environment, and property. With the increasing demand of HAZMAT, traffic accidents occurred frequently during HAZMAT transportation, and especially, a risk increase is generally observed in the presence of tunnels [1–3], which makes it of great importance to tighten regulation for vehicles transporting HAZMAT in tunnels.

HAZMAT vehicle reidentification (ReID) methods face the following challenges in tunnel scenes: (1) the strong reflection of the tank of a HAZMAT vehicle can cause large differences in its appearance under the uneven lighting conditions of a tunnel; (2) it is difficult to distinguish the HAZMAT vehicles with the same vehicle type effectively, due to their close appearance. However, there still remains a research gap both in HAZMAT vehicle data and in specialized algorithms. This motivates us to focus on the study of HAZMAT vehicle reidentification in tunnels.

Vehicle ReID aims to determine whether a vehicle image captured in nonoverlapping cameras belongs to the same vehicle in traffic monitoring scenarios. Existing methods mainly perform research on vehicle ReID based on the vehicle appearance [4]. However, due to the special and complex tunnel environment containing dim illumination and limited viewing field, it is more challenging for the tunnel vehicle ReID problem than that in open road scenes [5, 6]. Thus, large fluctuation can be seen by merely conducting tunnel vehicle ReID based on the appearance information. As shown in Figure 1, the red, green, and blue lines in each subfigure are RGB channel color histograms for each image. Vehicles for the second and third images may have similar appearance features, whereas they are actually two different IDs. From such instance, we can see that in real-world applications, it is extremely sensitive to environmental changes to merely perform vehicle ReID via appearance information.

To address the above problem, except for appearance information, the spatiotemporal information is further leveraged

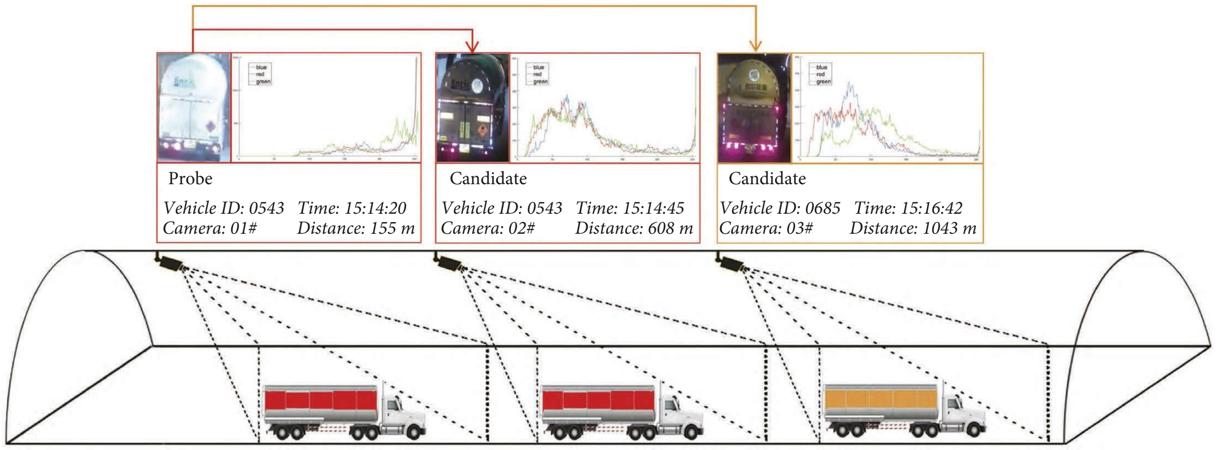

blue
red
green
Probe
Vehicle ID: 0543 Time: 15:14:20
Camera: 01# Distance: 155 m
Candidate
Vehicle ID: 0543 Time: 15:14:45
Camera: 02# Distance: 608 m
Candidate
Vehicle ID: 0685 Time: 15:16:42
Camera: 03# Distance: 1043 m

Figure 1: The HAZMAT vehicles are difficult to distinguish due to their close appearance. The reflection of the tank causes significant differences in its appearance under the variable lighting conditions in the tunnel.

to improve vehicle ReID performance in recent works [7–9]. This is inspired by the fact that the vehicle movements follow some implicit motion pattern according to the traffic rules. However, due to the randomness of vehicle motion, it is difficult to accurately model the spatiotemporal motion laws of vehicles in the open road. But the traffic rules of vehicles in tunnels are more distinct than in the open road, such as vehicles are expected to move in one fixed direction within limited speed, and no U-turns. It leads to the urgent need for a special spatiotemporal model tailored to the tunnel scene.

Therefore, to realize HAZMAT vehicle ReID in tunnel scenes, this work proposes a vehicle ReID method based on the fusion of vehicle appearance and tunnel spatiotemporal information. For vehicle appearance modeling, a deep residual network (i.e., Resnet50 [10]) is chosen as a feature extractor to model the complex appearance variation of tunnel vehicle. Meanwhile, to capture the spatiotemporal cues between cameras and vehicles, we develop a novel spatiotemporal similarity metric to model the between-vehicle structure correlation as well as the camera-vehicle topological relationship.

Furthermore, the extracted appearance representation and the spatiotemporal model are combined to efficiently encode the appearance variation and movement pattern for the tunnel vehicles. Moreover, to evaluate the HAZMAT vehicle ReID problem in the tunnel scenes, we construct and release a real-world HAZMAT Vehicle ReID dataset, named by VisInt-THV-ReID, containing 10,048 images of 865 HAZMAT vehicles collected from four high-resolution cameras. These images were captured by 4 cameras in the tunnel. Each camera monitors a space with a range of 150 meters and takes around 3 pictures of vehicles with far, middle, and near distances, respectively. Each vehicle is attached by the camera mileage and the picture shooting time. According to the spatial coordinate transformation method [11], we infer the spatial positions of vehicles in tunnel from the perspective of camera monitoring and obtain their temporal information by comparing timestamps of monitoring cameras. We use the vehicle ReID to determine whether the HAZMAT vehicles are exiting the tunnel within a normal time. If one vehicle passes the tunnel more than once, we identify the HAZMAT vehicle with a different vehicle ID for each time in the dataset. More attention is paid to the driving condition of the HAZMAT vehicle each time when it passes through the tunnel. The proposed method is evaluated to be effective through exhaustive experiments on the VisInt-THV-ReID dataset.

The main contributions of this work are summarized as follows:

(i) We extend the scenarios of vehicle ReID task to the challenging problem of HAZMAT vehicle ReID in tunnel scenes and propose a method that fuses both appearance modeling and spatiotemporal mining for more precise vehicle ReID.

(ii) We design a spatiotemporal metric approach based on the movement law of vehicles in road tunnels which brings in the description of between-vehicle structure correlation as well as the camera-vehicle topological relationship.

(iii) We build a real-world tunnel HAZMAT vehicle ReID dataset, named as VisInt-THV-ReID. As far as we know, the released VisInt-THV-ReID is the first HAZMAT vehicle ReID dataset captured in the tunnel scenes, which is crucial for the promotion of automatic regulation of HAZMAT transportation. Exhaustive experiments demonstrate that the proposed method can generate a state-of-the-art performance.

The rest of this work is organized as follows: The review related works are presented in Section 2. Section 3 details the proposed HAZMAT vehicle ReID method. In Section 4, we execute experiments for the evaluation of the proposed approach on VisInt-THV-ReID. Finally, we conclude this work in Section 5.

2. Related Work

Vehicle ReID in traffic monitoring scenarios can be seen as a part of multicamera tracking. Given an image of a vehicle in a specific area, the task is to find its image as captured under other cameras. This work studies vehicle ReID with spatiotemporal information fusion in tunnel scenes. We introduce related work from the aspects of vehicle ReID in tunnel scenes and multimodal information fusion.

2.1. Vehicle ReID Methods in Tunnels. Vehicle ReID in tunnel scenes is challenging due to low resolution, dim light, and dramatic changes in vehicle appearance. A vehicle is detected and tracked by each camera in road tunnels, and a detected vehicle is matched with the previous camera.

Frías-Velázquez et al. [6] proposed a probabilistic framework based on a two-step strategy that reidentifies vehicles in road tunnels. They built a Bayesian model that finds the optimal assignment between vehicles of connected groups based on descriptors such as trace transform signatures, lane changes, and motion discrepancies. Rios-Cabrera et al. [12] presented an integrated solution to detect, track, and identify vehicles in a tunnel surveillance application, taking into account practical constraints, such as real-time operation, imaging conditions, and decentralized architecture. AdaBoost [13] cascade is used for vehicle detection, and a comprehensive confidence score integrates the information of all stages of the cascade. Jelača et al. [14] proposed a real-time tracking method of multiple non-overlapping cameras in a road tunnel monitoring scene, using AdaBoost for vehicle detection. The vehicle detector and a Kalman filter of average optical flow are used for tracking. The ReID algorithm applies the projection feature similarity of a radon transform between vehicle images. Chen et al. [15] proposed a spatiotemporal successive dynamic programming algorithm to identify vehicles between pairs of cameras. They extracted features based on Harris corner detection and OpponentSIFT descriptors, considering color information [16]. Zhu et al. [5] proposed a synergistically cascaded forest model to gradually construct the linking relationships between vehicle samples with increasing alternative random forest and extremely randomized forest layers.

The abovementioned methods generally focus on the extraction of hand-designed features of vehicle images, which can only show good performance in specific scenes. These manual features are susceptible to the interference of a complex tunnel environment, and they are difficult to improve the precision of ReID.

2.2. Methods Using Multimodal Information. As a vehicle is far from cameras and the illumination is insufficient, the image resolution is low. Due to their similarity, it is impractical to effectively identify HAZMAT vehicles without special markings only by appearance. Recent work on vehicle ReID has improved the model by combining multi-dimensional information of vehicle attributes such as type, color, time, and space information with appearance features.

To reidentify vehicles based on fusion different appearance information, Liu et al. [17] designed a network using BOW-SIFT [18], BOW-CN [19], and GoogLeNet [20] to extract texture, color, and semantic features, respectively. Handmade features are fused with the vehicle type and color features obtained through deep learning. Liu et al. [21] proposed PROVID, which makes full use of appearance features, license plates, camera locations, and semantic information to carry out a progressive search from coarse to fine in the feature domain and from near to far in physical space.

To reidentify vehicles based on spatiotemporal information, Zhong et al. [7] proposed a vehicle pose guide model using a spatiotemporal probability model based on the Gaussian distribution to predict the spatiotemporal motion of vehicles. A convolution neural network (CNN) was used to predict the driving direction of a vehicle and the results of visual appearance, and then, the driving direction and spatiotemporal models were fused. Shen et al. [8] proposed a two-stage framework incorporating complex spatiotemporal information to effectively regularize ReID results. A candidate visual-spatiotemporal path was generated by a chain Markov random field model with a deeply learned potential function. A Siamese-CNN + Path-LSTM model takes the candidate path and pairwise queries to generate a similarity score. Jiang et al. [9] proposed an approach with a multi-branch architecture and a reranking strategy using the spatiotemporal relationship among vehicles from multiple cameras.

3. Method

3.1. Overview. Typically, a tunnel surveillance system consists of a series of cameras $C = \{C_0, C_1, C_2, \ldots, C_M\}$, with nonoverlapping visual receptive fields. $\overrightarrow{A_i}$ denotes the 2048-dimensional appearance feature vector obtained from the i -th vehicle image through the image appearance feature extraction network, and $\overrightarrow{S_i}$ denotes the spatiotemporal feature vector of the i-th vehicle collected by the camera. The spatiotemporal features involved are velocity v_i, timestamp t_i , and space position l_i of the tunnel.

We use $P_a(i, j)$ to represent the similarity of the appearance feature vectors of vehicles i and j from upstream and downstream cameras and $P_{st}(i, j)$ to represent the similarity of the spatiotemporal features of the vehicle pairs. $P(i, j)$ is the probability that vehicle pairs are identical after fusing multimodal information. The inputs of the proposed model are vehicle image pairs (i, j) and their spatiotemporal features $(\overrightarrow{S_i}, \overrightarrow{S_j})$ involved velocity, timestamp, and space position in the tunnel. The output is the probability $P(i, j)$ of whether the pair of vehicle images is the same vehicle.

The framework of the proposed method has three parts, as shown in Figure 2.

(1) Similarity calculation of vehicle appearance features. Resnet50 [10] is used as the feature extractor to obtain a 2048-dimensional appearance feature vector of a vehicle.

(2) Based on the spatiotemporal movement law of HAZMAT vehicles, we calculate the theoretical distance and the actual distance of the vehicle pairs. The tunnel spatial discrepancy ε_{ij} is used to evaluate

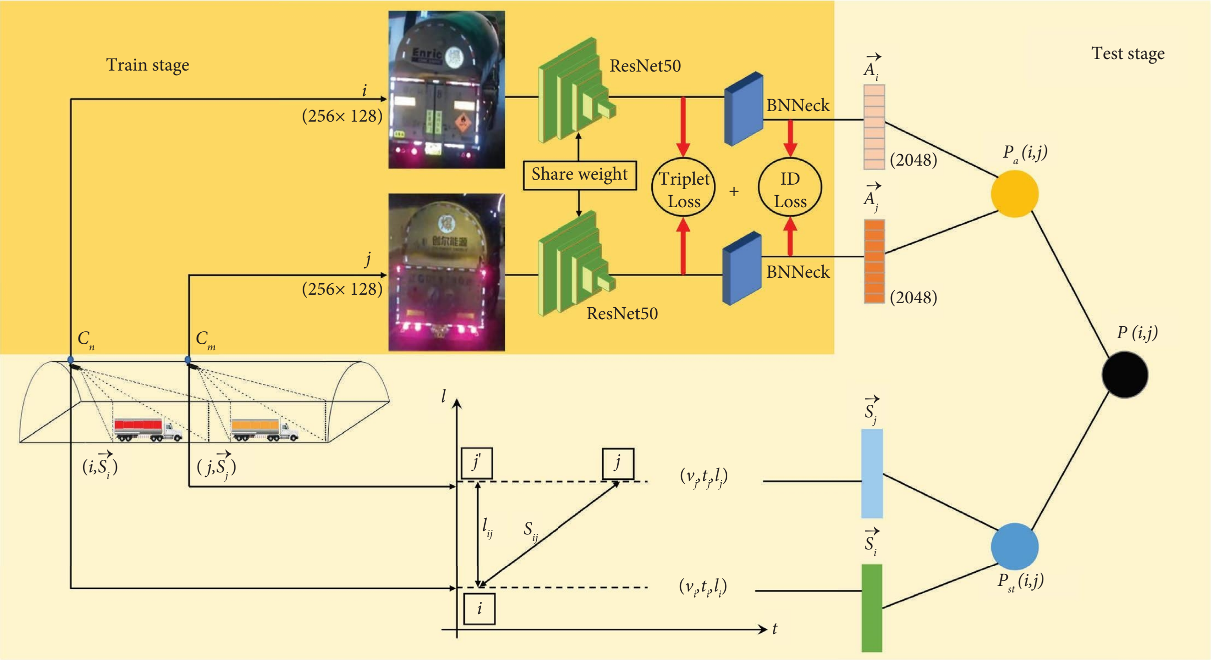


Figure 2: Vehicle ReID pipeline based on the fusion of appearance and spatiotemporal information.

the diversity between the theoretical distance and the actual distance.

(3) Similarity calculation of multimodal information fusion. Based on parts 1 and 2, the spatiotemporal and appearance similarity of the input vehicle image pairs are summed with a weight. We rerank the vehicle similarity of fusion information.

3.2. Appearance Features of Vehicle ReID. The vehicle appearance feature extraction network is shown in Figure 3. We use Resnet50 as the feature extraction backbone network and adjust each image to 256 × 128 pixels. Given an input image x_i with label y_i, the predicted probability of x_i being recognized as class y_i is encoded with a softmax function, represented by $p(y_i \mid x_i)$. ID prediction $p(y_i \mid x_i)$ is used to calculate ID loss [22]. The model outputs ReID feature $\overrightarrow{A_i}$ which is used to calculate triplet loss [23]. The output dimension of the full connection layer is changed to the number of vehicle IDs in the training dataset.

The ID loss treats the training process of vehicle ReID as an image classification problem [24], i.e., each identity is a distinct class. In the testing phase, the output of the pooling layer or embedding layer is adopted as the feature extractor. The identity loss is then computed by the cross-entropy.

$$L_{\mathrm{ID}} = -\frac{1}{N}\sum_{i=1}^{N}\log\left(p\left(y_i \mid x_i\right)\right), \tag{1}$$

where N represents the number of training samples within each batch.

The triple loss for feature extraction can reduce the intraclass distance of positive pairs and increase the interclass distance of negative pairs. Given a triplet (x^a, x^p, x^n), including an anchor image x^a, a positive x^p, and negative x^n, the triplet loss is formulated as follows:

$$L_{\mathrm{Tri}} = \sum_{i=1}^{N}\left[\left\|f\left(x_i^a\right) - f\left(x_i^p\right)\right\|_2^2 - \left\|f\left(x_i^a\right) - f\left(x_i^n\right)\right\|_2^2 + \alpha\right], \tag{2}$$

where α is a margin and usually set to 0.3. N is the number of training samples within each batch. $f(\bullet)$ stands for the appearance feature extractor.

In this work, we use ID loss and triplet loss together for optimizing the model. For image pairs in the embedding space, ID loss mainly optimizes the cosine distances while triplet loss focuses on the Euclidean distances. The feature vectors of the two losses are inconsistent in the embedding space. To address this problem, the BNNeck [22] is applied for more effective loss computation. BNNeck adds a batch normalization (BN) layer before the classifier FC layers of the model. The feature before the BN layer is denoted as $\overrightarrow{A_i}$. We let $\overrightarrow{A_i}$ pass through the BN layer to acquire a normalized feature $\vec{a_i}$. In the training stage, the feature $\overrightarrow{A_i}$ is used to compute the triplet loss. The feature $\vec{a_i}$ is used to compute the ID loss. Finally, the triplet loss and ID loss are combined to optimize the model. To train the ReID model, we combine ID loss and triple loss as follows:

$$L = L_{\mathrm{ID}} + L_{\mathrm{Tri}}. \tag{3}$$

In the test stage, the appearance features $(\overrightarrow{A_i}, \overrightarrow{A_j})$ for input image pairs (i, j) are generated using the vehicle

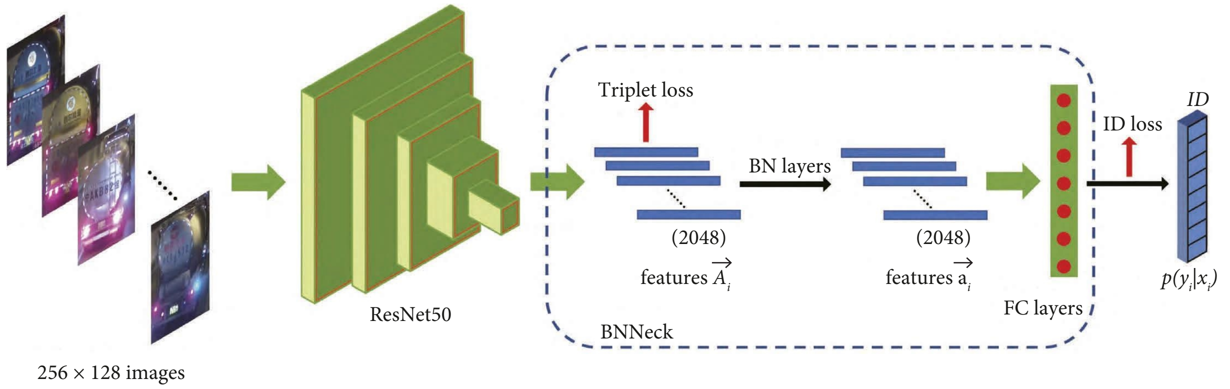


Figure 3: The framework of vehicle appearance modeling.

appearance feature extraction network. We use the cosine distance to measure the similarity between features and is expressed as follows:

$$P_a(i,j) = \frac{\overrightarrow{A_i}\cdot\overrightarrow{A_j}}{\left\|\overrightarrow{A_i}\right\|\left\|\overrightarrow{A_j}\right\|}. \tag{4}$$

3.3. Vehicle Spatiotemporal Features. The motion of the vehicle is limited by its speed and spatiotemporal motion. The time that the vehicle travels through a pair of cameras should be within a reasonable range. In a highway tunnel monitoring system, the driving speed of a vehicle is within the range of 10–80 km/h. The time interval of vehicle movement is affected by the camera installation position and the topological relationship of the tunnel and cameras. We analyze the motion law of the vehicle time interval between cameras in the VisInt-THV-ReID dataset. For each pair of cameras, the vehicle space interval can be modeled as a random variable that follows a certain distribution [6, 7].

In order to derive the spatiotemporal similarity probability distribution of the vehicle, we propose a feature called spatial discrepancy. We introduce the spatial discrepancy by considering Figure 4(a). This figure shows the spatiotemporal graph that relates vehicle i observed in upstream camera with another vehicle j observed in downstream camera. The motion variables involved are velocity v_i of vehicle i, timestamp t_i, and space position l_i of the tunnel. The state vector $\overrightarrow{S_i}$ expresses the spatiotemporal state of vehicle i.

To construct the spatiotemporal similarity relationship between the vehicle pairs, we calculate the theoretical distance and the actual distance of the vehicle pairs and define the indicator ε_{ij} to calculate the diversity of the distances. According to the constant acceleration model, the theoretical distance of the vehicle is calculated as follows according to the upstream and downstream cameras of the tunnel:

$$s_{ij} = \frac{v_i + v_j}{2}\cdot\left(t_j - t_i\right). \tag{5}$$

The actual distance between the current position of the vehicle collected by the upstream and downstream cameras is expressed as follows:

$$l_{ij} = \left|l_j - l_i\right|. \tag{6}$$

The spatial discrepancy ε_{ij} evaluates the fitness between the displacement estimate s_{ij} and the actual distance l_{ij} as stated in Figure 4(a). The tunnel spatial discrepancy is expressed as follows:

$$\varepsilon_{ij} = \frac{\left(s_{ij} - l_{ij}\right)}{\left|s_{ij}\right| + \left|l_{ij}\right|} \in (-1, 1), \tag{7}$$

which is used to evaluate the diversity between the theoretical distance and the actual distance. The spatial discrepancy ε_{ij} is evaluated by the vehicle spatiotemporal features involving velocity, timestamp, and space position.

To maintain the consistency of the data structure of the multimodal data fusion, we maintain the consistency of the spatiotemporal similarity discriminant method with the appearance feature discriminant method and use the chord function to represent the spatiotemporal similarity probability distribution of the vehicle. The $P_{\text{st}}(i,j)$ is defined as follows:

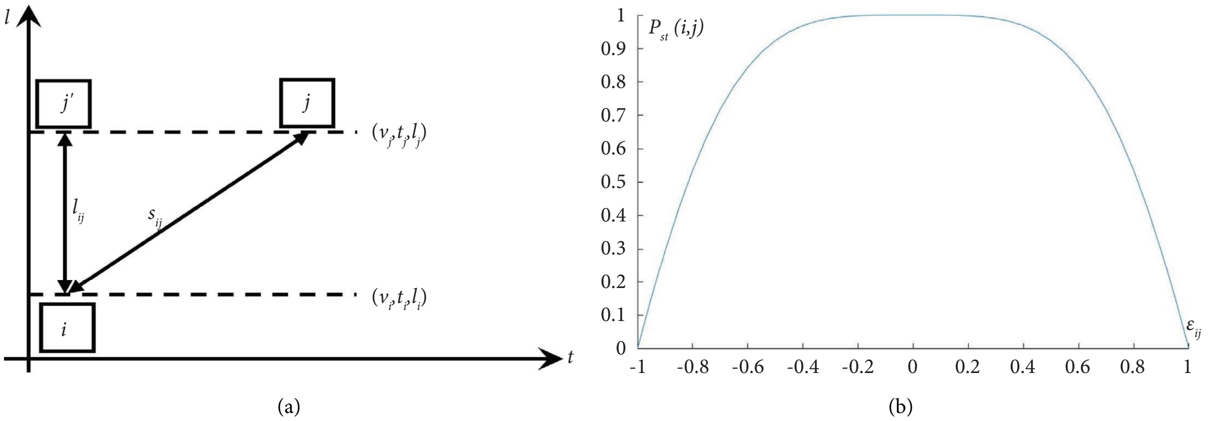
$$P_{\text{st}}(i,j) = \cos\left(\varepsilon_{ij}^2\cdot\frac{\pi}{2}\right). \tag{8}$$

As shown in Figure 4(b), $P_{\text{st}}(i,j)$ increases as ε_{ij} tends to 0. Based on $P_{\text{st}}(i,j)$, we can determine candidate matching vehicles according to the spatiotemporal similarity in tunnels.

3.4. Vehicle ReID by Fusing Image and Tunnel Spatiotemporal Information. To make full use of the vehicle appearance and spatiotemporal information, we establish a multimodal information strategy. The vehicle ReID probability is defined as follows:

$$P(i,j) = \lambda P_a(i,j) + (1-\lambda)P_{st}(i,j), \tag{9}$$

where the weight coefficient, $\lambda \in (0, 1)$, is used to fuse the spatiotemporal and appearance similarity.



(a)

(b)

Figure 4: (a) Motion states of vehicles i and j. (b) Spatiotemporal similarity distribution in tunnels.

4. Experiments

4.1. VisInt-THV-ReID Dataset. We verified the effectiveness of the proposed method on the VisInt-THV-ReID (The dataset is open-sourced at the following website: https://github.com/jialei-bjtu/VisInt-THV-ReID) dataset, which is collected from four cameras deployed in Taijia Expressway Linxian No. 3 tunnel in Shanxi province, China, providing high-definition video data of 6 million pixels and spaced at 300 meters. We collected video data for 10 hours daily over 3 days, from November 26 to 28, 2020, from 10:00 to 20:00. We annotated 10,048 pictures of 865 HAZMAT vehicles with their spatial position, speed, and timestamp information. To the best of our knowledge, this is the first open-source HAZMAT vehicle ReID dataset. The sample dataset is shown in Figure 5.

To mark the spatiotemporal and speed information of a vehicle, we must transform its spatial coordinates. Perspective transformation is used to transform the vehicle driving area under the camera vision to a fixed-size rectangle [11], as shown in Figure 6.

The position (x_i, y_i) of a vehicle in the camera field of view in the tunnel is calculated as follows:

$$\begin{cases} \left[x', y', \omega'\right] = \left[x^o, y^o, 1\right] \cdot T, \\ T = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \\ \left[x_i, y_i\right] = \left[\dfrac{x'}{\omega'}, \dfrac{y'}{\omega'}\right], \end{cases} \tag{10}$$

where x_i is the lateral distance of the vehicle from the left wall of the tunnel, y_i is its longitudinal distance from the current camera installation position, (x^o, y^o) is the lower midpoint of the vehicle object detection box in the image, and T is the transformation matrix defining the mapping between the original region and the transformation region. Using the image sequence taken by the surveillance camera, the speed of vehicle i in the tunnel can be obtained as follows:

$$v_i = \left(\sqrt{x_i^2 + y_i^2} - \sqrt{x_{i-1}^2 + y_{i-1}^2}\right) \cdot f, \tag{11}$$

where f is the frame rate of the monitoring camera, the spatial position vector l_i obtained by the camera at time t_i is (x_i, y_i), and the spatiotemporal vector of vehicle i is $\overrightarrow{S_i}\,(v_i, t_i, l_i)$.

We trained and tested the model on the VisInt-THV-ReID dataset, whose 10,048 images of 865 HAZMAT vehicles were divided into training, query, and test sets at a 10 : 1 : 9 ratio. The training set had 433 HAZMAT vehicles and 4980 images. There were 432 HAZMAT vehicles in the query and test sets, with 432 vehicle images in the query set and 4636 in the test set.

4.2. Experimental Settings. The mAP [21] and cumulative matching characteristic (CMC) curve [25] were used to evaluate the performance of the proposed method on the VisInt-THV-ReID dataset. The average precision for a query image is calculated as follows:

$$AP = \frac{\sum_{k=1}^{n} P(k) \cdot \mathrm{gt}(k)}{N_{\mathrm{gt}}}, \tag{12}$$

where n is the number of images in the test set, N_{gt} is the number of ground truths, $P(k)$ is the current precision result of the k-th query image, and $\mathrm{gt}(k)$ is an indicator function. When the matching result of the k-th query image is correct, $\mathrm{gt}(k) = 1$, and $\mathrm{gt}(k) = 0$ when it is incorrect.

The mAP is calculated as follows:

$$\mathrm{mAP} = \frac{\sum_{q=1}^{Q} AP(q)}{Q}, \tag{13}$$

where Q is the number of pictures in the query dataset. The CMC curve shows the probability that the correct matching image of the vehicle appears in the candidate lists. The CMC of the k-th position is as follows:

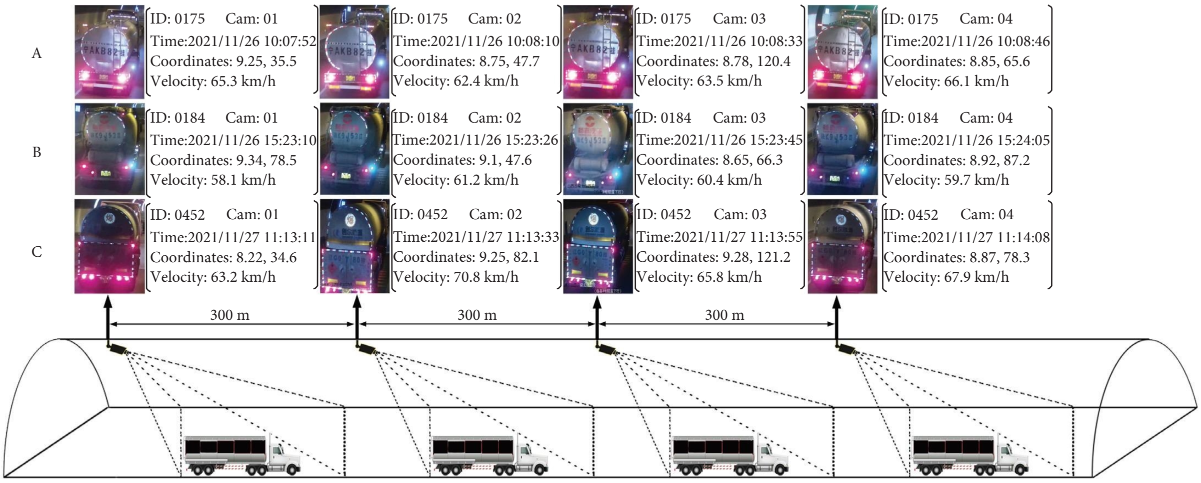


Figure 5: VisInt-THV-ReID dataset.

$$\mathrm{CMC}(k) = \frac{\sum_{q=1}^{Q} \mathrm{gt}(q,k)}{Q}, \tag{14}$$

where $\mathrm{gt}(q,k)$ is an indicator function, which equals 1 when the ground truth of the q query image appears before the k position. We also used Rank-1, Rank-5, Rank-10, and Rank-20 in the field of ReID to evaluate the model.

4.3. Ablation Study. Table 1 compares the experimental results of the multimodal fusion ReID method with those of Visual and ST-COS, which are appearance-based and spatiotemporal-based, respectively.

The method of Visual achieved 89.7% mAP and 96.3% Rank-1. The method of ST-COS achieved 85.5% mAP and 71.3% Rank-1. The fusion method Visual + ST-COS achieved 99.7% mAP and 99.8% Rank-1. The mAP of the fusion method increases by 142% and 10% compared to Visual and ST-COS and the Rank-1 rises by 3.5% and 28.5%.

The above results show that the multimodal information fusion method is superior to the use of appearance or spatiotemporal information alone and verify the effectiveness of the proposed multimodal information fusion method.

4.4. Comparison with Baselines. Table 2 shows the recognition precision of three baseline methods, PROVID [21], Visual + ST [7], and Siamese-CNN [8], comparing to that of Visual + ST-COS on the VisInt-THV-ReID dataset.

4.4.1. Appearance Feature Extraction and STR Spatiotemporal Fusion (PROVID). The method of PROVID extracts the appearance features of HAZMAT vehicles by the Resnet50 network and uses the STR method to measure the spatiotemporal relationship [21]. The STR is defined as follows:

$$\mathrm{STR}(i,j) = \frac{T_i - T_j}{T_{\max}} \cdot \frac{\delta\left(C_i, C_j\right)}{D_{\max}}, \tag{15}$$

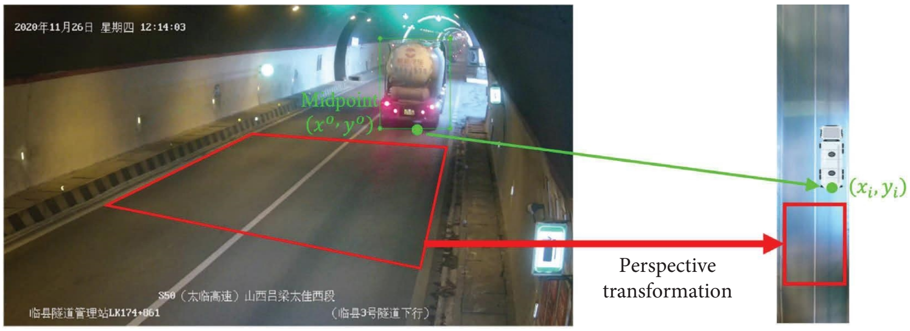
where T_i and T_j are the timestamps for the vehicles i and j captured by the cameras. $T_{\max}$ is the maximum time interval of vehicles passing through the tunnel. $\delta(C_i, C_j)$ is the actual distance between the current position of the vehicles collected by the upstream and downstream cameras, and $D_{\max}$ is the global maximum distance between any vehicles. We set $D_{\max}$ as the length of the tunnel.

4.4.2. Visual + ST. The method of Visual + ST extracts the appearance features of HAZMAT vehicles with the Resnet50 network and uses a spatiotemporal model based on the Gaussian distribution to predict the probability of vehicles [7]. $P_{\mathrm{stG}}(i,j)$ presents the similarity of the spatiotemporal features of vehicle pairs, and it is defined as follows:

$$P_{\mathrm{stG}}(i,j) = e^{\left(-10\cdot\varepsilon_{ij}^2\right)}, \tag{16}$$

where ε_{ij} is the tunnel spatial discrepancy as defined in equation (7).

4.4.3. Siamese-CNN. The method of Siamese-CNN uses a Resnet50 network to extract the appearance features of HAZMAT vehicles, and a multilayer perception network is applied to obtain their spatial and temporal relationships [8]. The spatiotemporal branch computes the spatiotemporal compatibility. Given the timestamps (t^i, t^j) and the positions (l^i, l^j) of vehilces, the input features of the branch are calculated as their time difference $\Delta \mathrm{t}(t^i, t^j)$ and spatial difference $\Delta \mathrm{d}(l^i, l^j)$. The scalar spatiotemporal compatibility is obtained by feeding the concatenated features, $[\Delta t(t^i, t^j), \Delta d(l^i, l^j)]^T$, into a multilayer perception with two fully connected layers. The outputs of the two branches are concatenated and input into a 2×1 fully connected layer with a sigmoid function to obtain the final compatibility


2020年11月26日 星期四 12:14:03
Midpoint
(x^o, y^o)
S50（太临高速）山西吕梁太佳西段
临县隧道管理站LK174+861
（临县3号隧道下行）
(x_i, y_i)
Perspective transformation

Original region transformed region

Figure 6: Coordinate transformation of vehicle position in tunnel space based on surveillance video.

Table 1: Results of ablation experiment.

Methods	mAP (%)	Rank-1 (%)	Rank-5 (%)	Rank-10 (%)	Rank-20 (%)
Visual	89.7	96.3	99.5	99.5	99.8
ST-COS	85.5	71.3	85.9	98.8	100
Visual + ST-COS	99.7	**99.8**	**100**	**100**	**100**

The bold values in Table 1 are the best values from the same column of data.

Table 2: Results of comparative experiments.

Methods	mAP (%)	Rank-1 (%)	Rank-5 (%)	Rank-10 (%)	Rank-20 (%)
PROVID	90.0	95.6	99.5	99.8	99.8
Visual + ST	90.8	96.1	99.5	99.8	99.8
Siamese-CNN	82.2	96.8	98.4	99.1	99.3
Our method	**99.7**	**99.8**	**100**	**100**	**100**

The bold values in Table 2 are the best values from the same column of data.

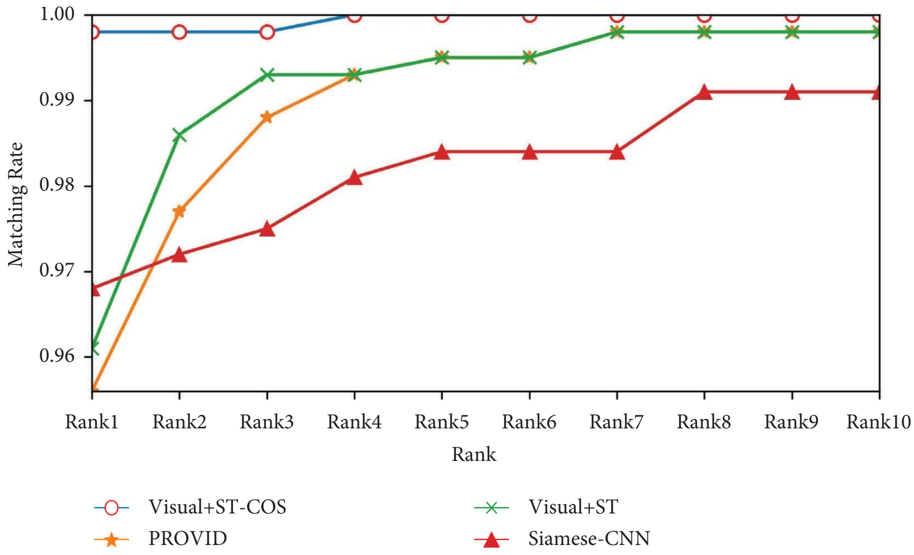


Figure 7: CMC curves on VisInt-THV-ReID dataset.

between the two states. Siamese-CNN takes all visual, spatial, and temporal information into consideration.

The results show that the proposed method achieves the best performance. It improves *mAP* and Rank-1 by 9.7% and 4.2%, respectively, compared with PROVID. This indicates that the STR spatiotemporal measurement method is not accurate enough to express the spatiotemporal information of vehicles in road tunnels. Compared with Siamese-CNN,

Table 3: Experimental results of coefficient λ under different values.

Results	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda =$ **0.35**	$\lambda = 0.4$	$\lambda = 0.5$	$\lambda = 0.6$	$\lambda = 0.7$	$\lambda = 0.8$	$\lambda = 0.9$
mAP	88.2	97.3	99.3	**99.7**	99.7	99.7	99.7	99.5	98.6	96.2
Rank-1	81.0	98.8	99.8	**99.8**	99.8	99.7	99.8	99.8	99.5	99.1
Rank-5	94.7	100	100	**100**	99.8	99.8	99.8	99.8	99.8	99.8
Rank-10	99.8	100	100	**100**	100	100	99.8	99.8	99.8	99.8
Rank-20	100	100	100	**100**	100	100	100	99.8	99.8	99.8

The bold values in Table 3 are the best values from the same column of data.

the proposed method improves mAP and Rank-1 by 17.5% and 3.0%. Since Siamese-CNN uses a multilayer perception network to train the spatial and temporal information of vehicles, the difficulty of model training is decreased and the precision is not ideal. Compared with Visual + ST, the proposed method improves mAP and Rank-1 by 8.9% and 3.7%, respectively. This shows that the proposed cosine spatiotemporal model can more accurately express the spatiotemporal state of a tunnel compared with Gaussian distribution. The CMC curves of all methods are shown in Figure 7.

4.5. Parameter Analysis. We experimented with the parameters of λ in the interval of 0.1–0.9. The best fusion result is achieved when λ equals 0.35. The comparison results of the parametric experiments are shown in Table 3. It can be observed from the table that a larger λ would cause appearance features to dominate vehicle identification, while a smaller λ causes spatiotemporal information to dominate. Table 3 shows that λ can have an important effect on the fusion results, and λ is relatively insensitive to the results in the interval 0.3–0.7.

5. Conclusion and Future Work

In this study, we presented a vehicle ReID method based on the fusion of vehicle appearance and tunnel spatiotemporal information for the task of HAZMAT vehicle ReID in road tunnels. The proposed method was evaluated on the VisInt-THV-ReID dataset. This study could play a role in promoting HAZMAT vehicle monitoring and traffic safety management in road tunnels.

Our future work has two aspects. Based on vehicle ReID research, we will study multicamera vehicle tracking technology to collect vehicle trajectories. In addition, we will use the time-to-collision (TTC) to indirectly evaluate safety and study a tunnel accident risk prediction model based on the traffic flow state.

Data Availability

The data that support the findings of this study are openly available in GitHub at https://github.com/jialei-bjtu/VisInt-THV-ReID.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Jia L. and Li X. conceived and designed the experiments; Li X. and Wang J. performed the experiments; Tianyuan W. and Haomin Y. analyzed the data; Jia L. and Wang W. wrote the paper; Li Q. reviewed and edited the paper. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

This work was supported in part by the Fundamental Research Funds for the Central Universities (Science and Technology Leading Talent Team project, grant no. 2022JBQY007), NIM R&D Project under grant no. 35-AKYZD2116-1, and China Academy of Railway Sciences Co., Ltd R&D Project under grant no. 2021IMXM04.

References

[1] R. Bubbico, S. Di Cave, B. Mazzarotta, and B. Silvetti, “Preliminary study on the transport of hazardous materials through tunnels,” *Accident Analysis & Prevention*, vol. 41, no. 6, pp. 1199–1205, 2009.

[2] B. Fabiano and E. Palazzi, “HazMat transportation by heavy vehicles and road tunnels: a simplified modelling procedure to risk assessment and mitigation applied to an Italian case study,” *International Journal of Heavy Vehicle Systems*, vol. 17, no. 3/4, p. 216, 2010.

[3] L. Jia, J. Wang, T. Wang, X. Li, H. Yu, and Q. H. M. D.-N. Li, “HMD-net: a vehicle hazmat marker detection benchmark,” *Entropy*, vol. 24, no. 4, p. 466, 2022.

[4] J. Deng, Y. Hao, M. S. Khokhar et al., “Trends in vehicle Re-identification past, present, and future: a comprehensive review,” *Mathematics*, vol. 9, no. 24, p. 3162, 2021.

[5] R. Zhu, J. Fang, S. Li et al., “Vehicle re-identification in tunnel scenes via synergistically cascade forests,” *Neurocomputing*, vol. 381, pp. 227–239, 2020.

[6] A. Frías-Velázquez, P. Van Hese, A. Pižurica, and W. Philips, “Split-and-match: a Bayesian framework for vehicle re-identification in road tunnels,” *Engineering Applications of Artificial Intelligence*, vol. 45, pp. 220–233, 2015.

[7] X. Zhong, M. Feng, W. Huang, Z. Wang, and S. Satoh, “Poses guide spatiotemporal model for vehicle Re-identification,” in *MultiMedia Modeling*Springer International Publishing, Berlin, Germany, 2018.

[8] Y. Shen, T. Xiao, H. Li, S. Yi, and X. Wang, “Learning deep neural networks for vehicle Re-ID with visual-spatio-temporal path proposals,” in *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*, IEEE, Venice, Italy, October 2017.

[9] N. Jiang, Y. Xu, Z. Zhou, and W. Wu, "Multi-attribute driven vehicle Re-identification with spatial-temporal Re-ranking," in *Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP)*, IEEE, Athens, Greece, October 2018.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Las Vegas, NV, USA, June 2016.

[11] Z. Xiong, M. Li, Y. Ma, and X. Wu, "Vehicle Re-identification with image processing and car-following model using multiple surveillance cameras from urban arterials," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 12, pp. 7619–7630, 2021.

[12] R. Rios-Cabrera, T. Tuytelaars, and L. Van Gool, "Efficient multi-camera vehicle detection, tracking, and identification in a tunnel surveillance application," *Computer Vision and Image Understanding*, vol. 116, no. 6, pp. 742–753, 2012.

[13] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, 1997.

[14] V. Jelača, J. O. N. Castañeda, A. Frías-Velázquez, A. Pižurica, and W. Philips, "Real-time Vehicle Matching for Multi-Camera Tunnel Surveillance," *Real-Time Image and Video Processing 2011*, pp. 232–239, Society of Photographic Instrumentation Engineers, Cergy-Pontoise, France, 2011.

[15] H. T. Chen, M. C. Chu, C. L. Chou, S. Y. Lee, and B. S. Lin, "Multi-camera vehicle identification in tunnel surveillance system," in *Proceedings of the 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, IEEE, Turin, Italy, June 2015.

[16] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek, "Evaluating color descriptors for object and scene recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1582–1596, 2010.

[17] X. Liu, W. Liu, H. Ma, and H. Fu, "Large-scale vehicle re-identification in urban surveillance videos," in *Proceedings of the 2016 IEEE International Conference on Multimedia and Expo (ICME)*, IEEE, Seattle, WA, USA, July 2016.

[18] L. Zheng, S. Wang, W. Zhou, and Q. Tian, "Bayes merging of multiple vocabularies for scalable image retrieval," in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Columbus, OH, USA, June 2014.

[19] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person Re-identification: a benchmark," in *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, IEEE, Santiago, Chile, December 2015.

[20] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," in *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Boston, MA, USA, June 2015.

[21] X. Liu, W. Liu, T. Mei, and H. Ma, "PROVID: progressive and multimodal vehicle reidentification for large-scale urban surveillance," *IEEE Transactions on Multimedia*, vol. 20, no. 3, pp. 645–658, 2018.

[22] H. Luo, Y. Gu, X. Liao, S. Lai, and W. Jiang, "Bag of tricks and a strong baseline for deep person Re-identification," in *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, IEEE, Long Beach, USA, March 2019.

[23] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: a unified embedding for face recognition and clustering," in *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Boston, MA, USA, June 2015.

[24] L. Zheng, H. Zhang, S. Sun, M. Chandraker, Y. Yang, and Q. Tian, "Person Re-identification in the wild," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3346–3355, Honolulu, HI, USA, July 2017.

[25] H. Liu, Y. Tian, Y. Wang, L. Pang, and T. Huang, "Deep relative distance learning: tell the difference between similar vehicles," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Las Vegas, NV, USA, June 2016.