electronics

MDPI

Article

Visual Features with Spatio-Temporal-Based Fusion Model for Cross-Dataset Vehicle Re-Identification

Zakria [1,*], Jianhua Deng [1], Jingye Cai [1,*], Muhammad Umar Aftab [1], Muhammad Saddam Khokhar [2] and Rajesh Kumar [3]

1 School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China; jianhua.deng@uestc.edu.cn (J.D.); ms.umaraftab@yahoo.com (M.U.A.)
2 School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang 212013, China; saddam_khokhar@hotmail.com
3 School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China; rajakumarlohano@gmail.com
* Correspondence: zakria.uestc@hotmail.com (Z.); jycai@uestc.edu.cn (J.C.)

Received: 31 May 2020; Accepted: 29 June 2020; Published: 1 July 2020

Abstract: Vehicle re-identification (Re-Id) is the key module in an intelligent transportation system (ITS). Due to its versatile applicability in metropolitan cities, this task has received increasing attention these days. It aims to identify whether the specific vehicle has already appeared over the surveillance network or not. Mostly, the vehicle Re-Id method are evaluated on a single dataset, in which training and testing of the model is performed on the same dataset. However in practice, this negatively effects model generalization ability due to biased datasets along with the significant difference between training and testing data; hence, the model becomes weak in a practical environment. To demonstrate this issue, we have empirically shown that the current vehicle Re-Id datasets are usually strongly biased. In this regard, we also conduct an extensive study on the cross and the same dataset to examine the impact on the performance of the vehicle Re-Id system, considering existing methods. To address the problem, in this paper, we have proposed an approach with augmentation of the training dataset to reduce the influence of pose, angle, camera color response, and background information in vehicle images; whereas, spatio-temporal patterns of unlabelled target datasets are learned by transferring siamese neural network classifiers trained on a source-labelled dataset. We finally calculate the composite similarity score of spatio-temporal patterns with siamese neural-network-based classifier visual features. Extensive experiments on multiple datasets are examined and results suggest that the proposed approach has the ability to generalize adequately.

Keywords: intelligent transportation system; fusion model; cross-dataset; vehicle re-identification; surveillance camera; siamese neural network; dataset bias

1. Introduction

In metropolitan cities, cameras are widely deployed in numerous areas for activity monitoring, from home surveillance systems to small and large business applications. Especially, a significant number of cameras are used for security purposes in public places, like parking lots, downtown, airports, and other sensitive areas. Globally, camera surveillance is one of the core modules in public transportation systems and has a large capability to contribute to the planning and controlling the traffic networks. Surveillance cameras cover large overlapping and nonoverlapping views; however, recorded frames of surveillance cameras are often manually analyzed in different organizations, which is a time and resource-consuming activity with hectic and tiresome tasks. Furthermore, the main goal of the surveillance system application is to develop such type of intelligent system that automates the human

decision-taking mechanism. Like a human operator, the automatic system should respond according to the scenario while making the system operator-independent. Meanwhile, the advances in the area of artificial intelligence, image processing, computer vision, and pattern recognition enable the systems to be more reliable, efficient and robust.

Similar to the person re-identification (Re-Id) task, the vehicle Re-Id is also a demanding task in camera surveillance. Vehicle Re-Id aims at matching a specific vehicle in another camera of surveillance system [1]. It is considered as the main module in intelligent transportation systems. However, with the installation of surveillance cameras on the roads for smart cities and traffic management, the demand to perform vehicle search from the gallery set is increased. Vehicle Re-Id is similar to several other applications, such as person Re-Id [2], behavior analysis [3], cross-camera tracking [4], vehicle classification [5], object retrieval [6], object recognition [7], and so on. However, it is significantly challenging for vehicle Re-Id algorithms to extract robust visual features from images to re-identify a vehicle. Intra-class variability, inter-class similarity and pose are major issues in which two different vehicle images look the same and the same vehicle looks different because of varied vehicle angles, illumination, resolution, viewpoint shift and camera color response [8], as shown in Figure 1.

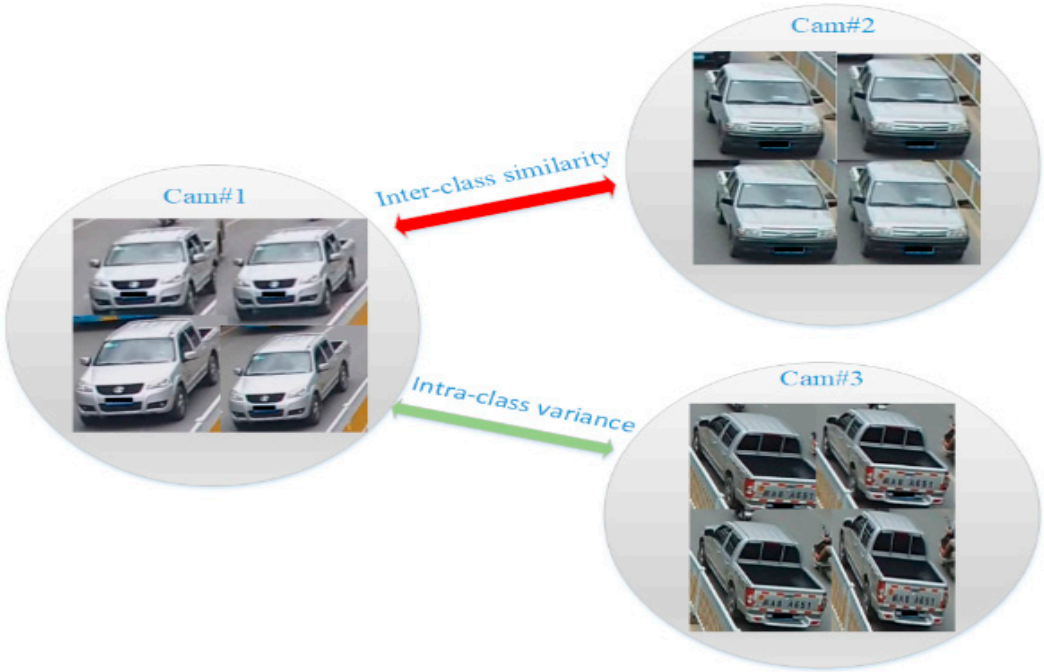

Figure 1. Illustrates two major challenges, intra-class variability and inter-class similarity, in vehicle re-identification (Re-Id).

In the last few years, research performed on vehicle Re-Id has successfully overcome many challenges. However, there is an ample gap in current approaches pertaining to cross-dataset vehicle Re-Id yet to be covered. It is also worth mentioning that this problem is more practical in real scenarios. In this context, the common approach to evaluate the model is dividing dataset into two different parts; one part is used for the training and the second part is used for testing the model. However, for the evaluation of proposed vehicle Re-Id algorithms researchers mostly use same dataset for testing and training purpose, which limits performance in practical scenario as the mostly used datasets contain biased and correlated data. Moreover, overfitting may also occur due to the overly tuned system parameters on certain features. Aim of the cross-dataset vehicle Re-Id is to develop such a model that performs with the same accuracy in practical scenario.

Moreover, it is noticed that vehicle Re-Id dataset images are taken by limited cameras, and consequently, the vehicle images have same angle, pose, background, and camera color response.

Furthermore, treating all pixels in image equally may bias the training algorithm and get the similarities between images with similar pose, angle, color, and backgrounds. Note that the pixels of foreground and background in each image have the same effect on the learning algorithm [9], and it degrades the system performance. The influence of background on the performance of vehicle Re-Id systems is not investigated before.

It is very difficult in a practical environment to reidentify a vehicle accurately on the basis of appearance information only. In a surveillance scenario, each vehicle image provides us visual appearance information as well as spatio-temporal information. Therefore, it may help if we exploit the appearance information as well as spatio-temporal clues for cross-dataset vehicle Re-Id, and examine the spatio-temporal relation between every pair of vehicle image. Practically, when a security officer needs to search a suspicious vehicle in large-scale surveillance system, the officer tries to reach the exact vehicle by utilizing appearance information such as model, color, type, etc., and, at the same time, searches in close to far positioned cameras with respect to time and space. Moreover, recently vehicle Re-Id gains more attention in the research community because of various significant practical applications where cross-dataset vehicle Re-Id can be deployed.

However, deploying directly person Re-Id models on vehicle Re-Id may not perform well. It can be observed from Figure 2 that the difference (based on visual features) between two viewpoints of a vehicle is much higher than the two viewpoints of a person. Whereas it is worth noticing that visual patterns of the vehicle are non-overlapping for the front, rear, and side viewpoint. In contrast, the visual appearance of a person is shared over different viewpoints. In fact, the color and texture of human body clothes normally do not change as much by changing the viewpoint, because of upright body posture. Initial models for person Re-Id did not consider different viewpoints of a person while processing the data. Moreover, such person Re-Id approaches split the human body vertically for feature extraction [10,11]. However, horizontal changes are not largely considered. Therefore, conventional methods for person Re-Id are not explicitly appropriate for vehicle Re-Id. Furthermore, some methods adopt subspace learning with distance matrix [12,13] and neural networks [14,15] in brute-force manner where the distance between images of same person is low and difference between images of different person is high. In this case, we cannot apply brute-force for vehicle Re-Id because images of a vehicle captured from different viewpoints look like different vehicles. On the other hand, the images of two different vehicles captured from same view angle look more alike.

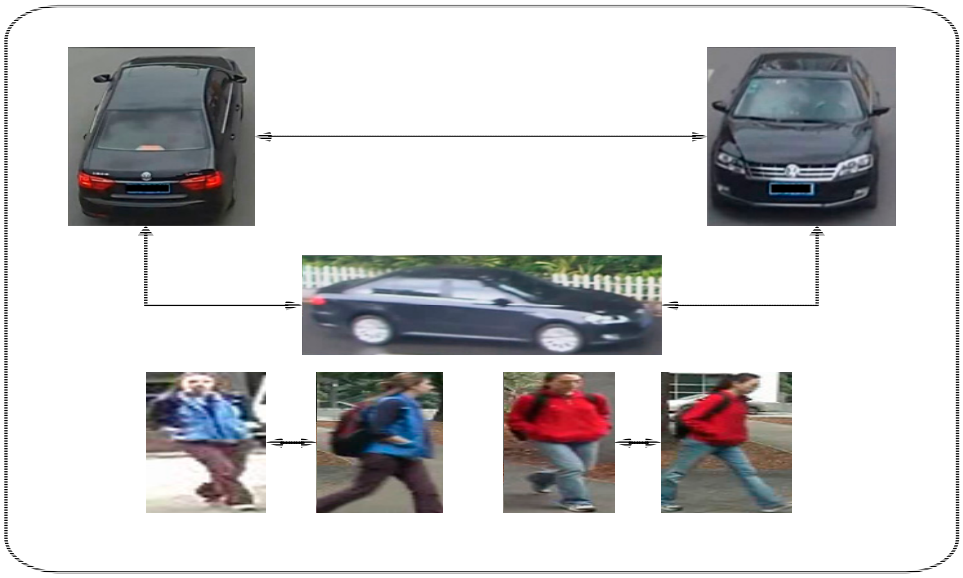

Figure 2. Different viewpoints of a vehicle and a person.

Significant Real-World Applications of Vehicle Re-Id

Some of the significant applications that satisfy the needs of real life where vehicle Re-Id system can be adopted are discussed as follows:

- Suspicious vehicle search: Most of the time, terrorists use a vehicle for their criminal activities and soon leave their location. It is very difficult to fast search suspicious vehicles manually from the surveillance camera;
- Cross-camera vehicle tracking: In vehicle race sports, some of the viewers on television want to watch only one targeted vehicle. When the vehicle comes in the field of view of camera network at different time and place, the broadcaster only focuses on that vehicle, and also the vehicle Re-Id system is also helpful to track the vehicle path;
- Automatic toll collection: This system can be used at toll gates to identify vehicle types, like small, medium and large, and charge the toll rate accordingly. Automatic toll collection reduces delay and improves toll collection performance by saving travelers time and fuel consumption;
- Road access restriction management: In big cities, heavy vehicles like trucks are not permitted in the daytime, or some of the vehicles with specific license plate numbers are permitted on specific days to avoid congestion in the city, or officially authorized vehicles can enter in city at any time;
- Parking lot access: This system can be deployed at the parking lot of different places (i.e., head offices, residential societies, etc.) where there is lack of parking areas; so that only the authorized vehicles are allowed to park;
- Traffic behavior analysis: Vehicle Re-Id can be used to examine the traffic pressure on different roads at a different time such as peak hours calculation or particular vehicle type behavior;
- Vehicle counting: The system can be useful to count certain type of vehicle;
- Speed restriction management system: Vehicle Re-Id system can be utilized to calculate the average velocity of a vehicle when it is crossing from two subsequent surveillance camera positions;
- Travel time estimation: Travel time information is important for a person who is travelling on road, it can be calculated when a vehicle is passing in between consecutive surveillance cameras;
- Traffic congestion estimation: By knowing the number of vehicles flowing from one point to another point within a specific time period using vehicle Re-Id system, we can estimate traffic congestion at the common spot from where all vehicles may cross;
- Delay estimation: Specific commercial vehicle delay can be estimated after predicting traffic congestion on the route that the vehicle follows;
- Highway data collection: Highway data can be collected through surveillance cameras that are installed on roadsides and it can be used for any purpose after processing and analyzing at the traffic control center;
- Traffic management systems (TMS): Vehicle Re-Id is an integral part of TMS; it helps to increase transportation performance, for instance, safe movement, flow, and economic productivity. TMS gathers the real-time data from the surveillance cameras network and streams into the transportation management center (TMC) for data processing and analyzing;
- Weather precautionary measures: When a specific vehicle is identified that may be affected by weather, then TMS notifies that vehicle about weather conditions like wind velocity, severe weather, etc.;
- Emergency vehicle pre-emption: If any suspicious vehicle is identified at any event or road then vehicle pre-emption system passes message towards lifesaving agencies such as security, firefighters, ambulance, and traffic police, etc. to reach in time and stabilize the scene. With this system, we can maximize safety and minimize response time;
- Access control: The vehicle Re-Id system can be implemented for providing safety and security, logging and event management. With the implementation of the system only authorized members can get an automatic door opening facility, which is helpful for guards on duty;

- Border control: The vehicle Re-Id system can be adopted at different check posts to minimize illegal vehicle border crossing. The system can provide vehicle and owner information as it approaches a security officer after identifying the vehicle. Normally these illegal vehicles are involved in cargo smuggling;
- Traffic signal light: When the traffic light is red and any vehicle crosses stop line, then the vehicle Re-Id system can be implemented there to identify that vehicle to charge the fine.

Our key contributions in this paper are summarized as follows:

- We attempt to investigate the vehicle Re-Id dataset bias problem using deep CNN models (dataset classification) to show the significance of cross-dataset vehicle Re-Id study;
- We conduct deep empirical study on cross- and single-dataset vehicle Re-Id to further examine the impact on the performance of existing methods;
- We propose the data augmentation method to reduce the influence of pose, angle, camera color response, and background on a model in vehicle images;
- We present a novel model that combines captured spatio-temporal patterns and siamese neural network classifiers features to achieve significant improvement in cross-dataset vehicle Re-Id.

The rest of the paper is organized as follows. The previous related research works are discussed in Section 2. Section 3 discusses the proposed approach for cross-dataset vehicle Re-Id. The experimentation settings, results and analysis on three publicly available datasets are described in Section 4. Section 5 discusses the overall results. Last, we conclude the paper in Section 6, including the contributions and provision of the potential future directions.

2. Related Work

Mainly object Re-Id tasks have been studied in person Re-Id and vehicle Re-Id domains. Initially, Re-Id problems are examined and applied on person. Approaches for person Re-Id are based on convolutional neural networks and are commonly divided into two parts: the first is enhancing the robustness and discriminative features [16,17] and the second is to learn the distance metric [12]. Additionally, deep learning-based methods are broadly investigated in the last five years for discriminative features and the distance metric at the same time. For example, deep-filter deep neural networks are generally designed to model geometric transformation and complex photometric [14]. Ahmed et al. [15] proposed a method to learn local correlation using a cross-input neighborhood, whereas Wang et al. [18] examined the combined learning of cross- and single-image representation for Re-Id. However, these approaches are not directly applicable for vehicle Re-Id problems.

In the initial work, vehicle Re-Id is done using license plate recognition and multiple sensors. Both techniques are accurate enough but require a constraint environment. If we consider license plate recognition based vehicle Re-Id then it would be difficult to capture vehicle images in which the vehicle license plate is clear enough and with good resolution. On the other hand, license plates are only visible from the front and rear side of vehicle. Typically, it is infeasible because of top view camera installations. With all these limitations, we can only capture images at toll stations or parking areas. Furthermore, it is also possible that vehicles may appear with a fake license plate. However, sensor-based vehicle Re-Id approaches are accurate, but most of the sensor-based techniques require vehicle owner's cooperation to install hardware/device in vehicle that is not suitable for criminal investigations as well as for the security. It is also costly, which is why researchers recommend vision-based vehicle Re-Id.

Before the advancement in deep learning, researchers used traditional computer vision techniques for Re-Id tasks such as local maximal occurrence (LOMO) [19], scale invariant feature transform (SIFT) [20], and bag-of-words (BOW) [21] for low-level feature extraction. Nowadays, because of the fast development and performance of deep neural network, various state-of-the-art appearance-based approaches using neural nets are proposed. Liu et al. [22] used the fusion of attributes and color features (FACT) model in which they combine low-level features (color and texture) and high-level

deep features. Zhu et al. [23] then proposed a siamese neural network for vehicle Re-Id and trained the network using deep features and similarity. As a further study, Tang et al. [24] employed deep features and handcrafted features for a multimodal metric learning network. To overcome intraclass variance and interclass similarity, Bai et al. [25] proposed deep metric and loss function by examining the vehicle shape. Liu et al. [26] proposed a model called region aware deep model (RAM) that utilizes the vehicle's local and global features for Re-Id.

The contextual cues, such as the spatio-temporal information, camera topology and object locations, have also been widely examined in nonoverlapping multicamera systems [27,28]. For instance, Javed et al. [27] studied camera correspondence using spatio-temporal relation for person tracking over surveillance cameras. Xu et al. [28] proposed graph-based object retrieval architecture to search cyclists and people on the campus. Ellis et al. [29] proposed a method to learn spatio-temporal transitions from path data that are acquired from a multi-camera surveillance network. Loy et al. [30] designed an approach for multiple-camera activity correlation analysis which is used to estimate the spatio-temporal topology of the surveillance cameras network [31]. In metropolitan areas with heavy traffic situations, for example, roads, crossroads, and camera topology are often in an unconstrained environment with uncertainty because of the different road topology and complicated environment. We still utilize and gain some benefit by exploiting the contextual information for vehicle Re-Id.

In person Re-Id, some of the researchers have exploited the cross-dataset technique and proposed different approaches to enhance the cross-dataset Re-Id performance. Lv et al. [32] proposed an unsupervised algorithm using an incremental learning technique, T Fusion, that learns the spatio-temporal patterns of pedestrians' unlabelled test dataset by utilizing a labelled training set. Fan et al. [33] used pretrained models with CNN and k-means clustering for the target domain. The study utilized labelled person Re-Id data for the baseline model, and, for feature extraction from the target, domain baseline was adopted then k-means clustering was applied. Lv et al. [34] proposed a method for unlabelled data in the target domain to fit in the model by using a generative adversarial network (GAN). GAN with cycle consistency constraint converts the unlabelled dataset of the target domain in the style of the input domain. Unfortunately, till today, to the best of authors' knowledge, no one has worked on cross-dataset vehicle Re-Id using spatio-temporal data.

3. Overview of Proposed Approach

We present the design and implementation procedure of the proposed approach for cross-dataset vehicle Re-Id in Figure 3. The data augmentation is adapted here to maximize the variability of training dataset and also counter the issue of partial occlusion, lighting condition, pose and consistent image background, along with reduces the dataset bias. During the training process, we augment the dataset with color, crop, rotation, affine, and random background substitution. In this way, the system learns to focus on the parts of the images belonging to the subjects, and image features that are strongly related to identity, while learning to ignore irrelevant sources of variability such as background. We then apply the siamese neural network (SNN) based classifier for visual feature extraction of source dataset. Moreover, we adapt the transfer learning technique for the spatio-temporal patterns of the unlabelled target dataset using labelled source dataset, and, lastly, we calculate together the similarity score of visual features and spatio-temporal patterns. Next, we discuss in detail the working procedure of the proposed method.

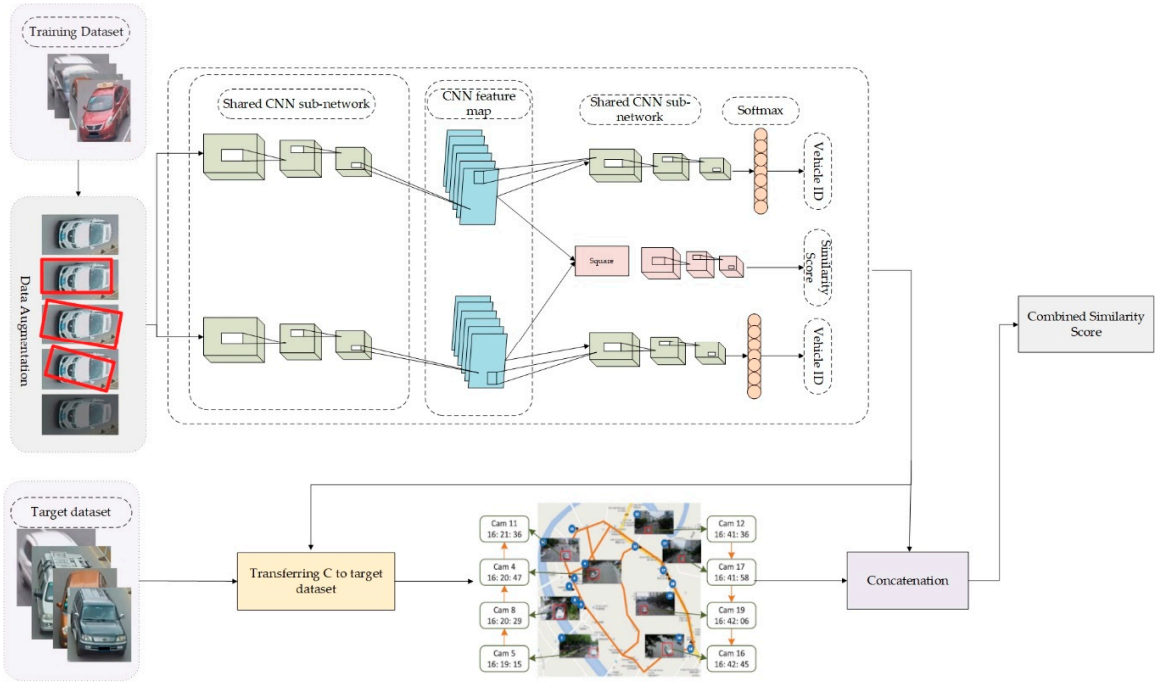

Figure 3. The proposed framework for cross-dataset vehicle Re-Id.

This approach mainly consists of the following steps:

- Data augmentation
- Siamese neural network based classifier
- Spatio-temporal pattern generation of the target dataset
- Calculation of composite similarity score

3.1. Data Augmentation

In a real-world vehicle Re-Id system situation, usually, the images are captured from several angles, pose and images illumination, but the class of the vehicle does not change by changing these characteristics. However, due to a small number of training images, the model may not represent each characteristic of the real scenario, and, as a result, overfitting takes place. Therefore, to increase the variability of images, we adopt data augmentation in the training set by introducing random transformation, for instance, rotation, mirroring, affine and changing color, and also changing image background, as illustrated in Figure 4. During training, we use the color casting technique by adjusting rgb channel of each image. To add further variety in the training set we cropped the training image 224×224 pixel and make sure that the cropped pixels are less informative. Generally, middle pixels of image are found to be informative. Horizontal mirroring, flipping, affine, and rotation are also utilized to add more variability in the data sample.

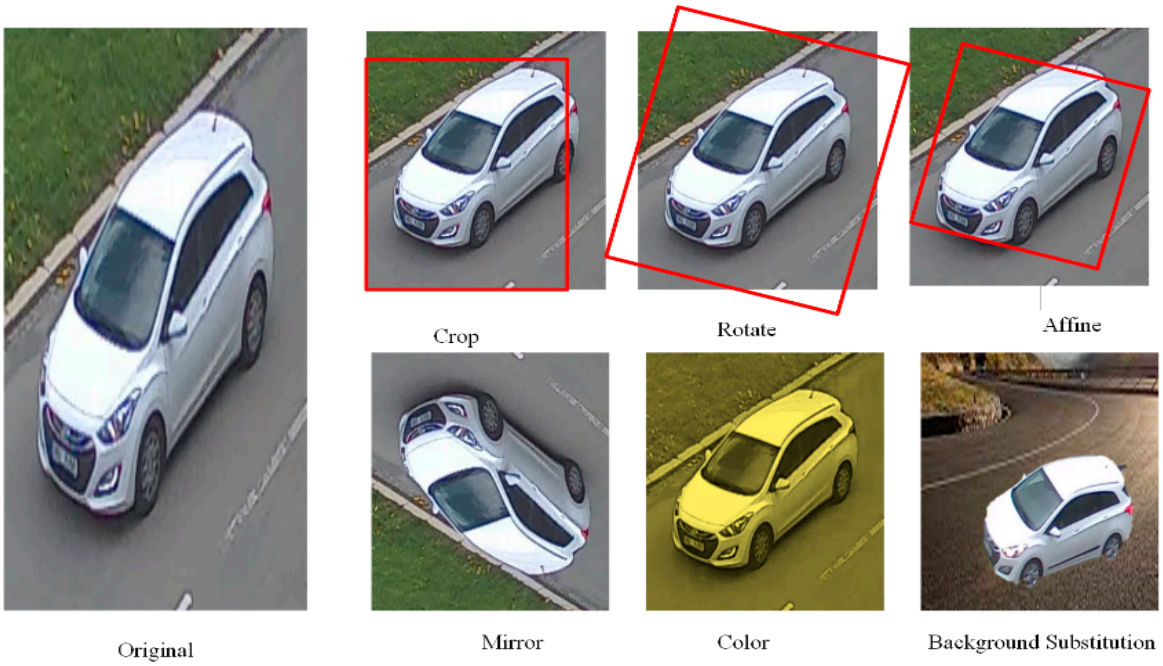

Figure 4. Illustrates of data augmentation process to add variability in training set images.

For robust vehicle Re-Id, the model should only focus on the vehicle and ignore irrelevant information which is not useful. Practically, the captured images have a random background. The system does not perform well if background of training set images are correlated with the foreground. To minimize the effect of image background on our training model, first, it is necessary that the vehicle image should be separated from the image background. Once a vehicle is separated from an image, we substitute another background in an image from a real-world background substitution gallery set. Modifying the training images in this way helps the Re-Id system to discriminate between the different sources of image variability. By doing data augmentation before training, we can encourage our model to learn and concentrate only on the foreground region of vehicle image, and the network learns to discriminate the vehicle well, which helps in generalizing on another dataset. Furthermore, the data augmentation technique is computationally inexpensive.

3.2. Siamese Neural Network-Based Classifier

A convolutional neural network (CNN) mainly consists of two structures: the identification model and the verification model. These are differentiable in terms of input images, feature vector extraction and loss function. We input a pair of images in the verification model and the model determines whether these two vehicle images are the same or not. In the verification network, a pair of images of the same vehicle are mapped close in feature space. On the other hand, the identification model treats vehicle Re-Id as a task of multiclass recognition from input vehicle images to vehicle ID model that directly learns nonlinear function, and the final layer is used for cross-entropy loss at the testing time. However, the fully connected layer is used for feature extraction and then normalized to measure the similarity between two vehicles images by using Euclidean distance between normalized CNN embedding.

As presented in Figure 5, we have utilized a recently proposed SNN-based classifier [35] for vehicle Re-Id. The SNN-based classifier simultaneously calculates the verification and identification loss on the basis of a pair of vehicle training images. The classifier learns discriminative features and measures similarity score to predict whether the pair of input vehicles contain same vehicle or not. The network consists of two ImageNet pretrained CNN modules that have a shared parameters with the same weights to extract features from a pair of input vehicle images. CNN modules consist of

ResNet-50 by removing the final layer to achieve higher accuracy. Furthermore, several researchers have used deep architecture, but it creates a vanishing gradient problem. To prevent this, ResNet-50 introduces residual blocks with reference to block input, and an intermediate layer of a block learns a residual function. We use here residual function as a refinement step in which we learn how to adjust the input feature map for higher quality features. This compares with a "plain" network in which each layer is expected to achieve new and distinct feature maps using ResNet-50 for preventing it. Just like traditional multiclass recognition methods for identity prediction, we adopt cross-entropy loss, which is expressed as Equations (1) and (2):

$$\hat{p} = softmax(\theta_I \, o \, x) \tag{1}$$

$$\text{Identify}(x, \, c, \, \theta_I) = \sum_{i=1}^{n} -p_i \, \log(\hat{p}_i) \tag{2}$$

where, o denotes convolution operation, x is $1 \times 1 \times 4096$ tensor, whereas c, θ_I and $\hat{p}$ represent target class, indicate the parameters of a convolutional layer and predicted probability respectively. p_i denotes the target probability.

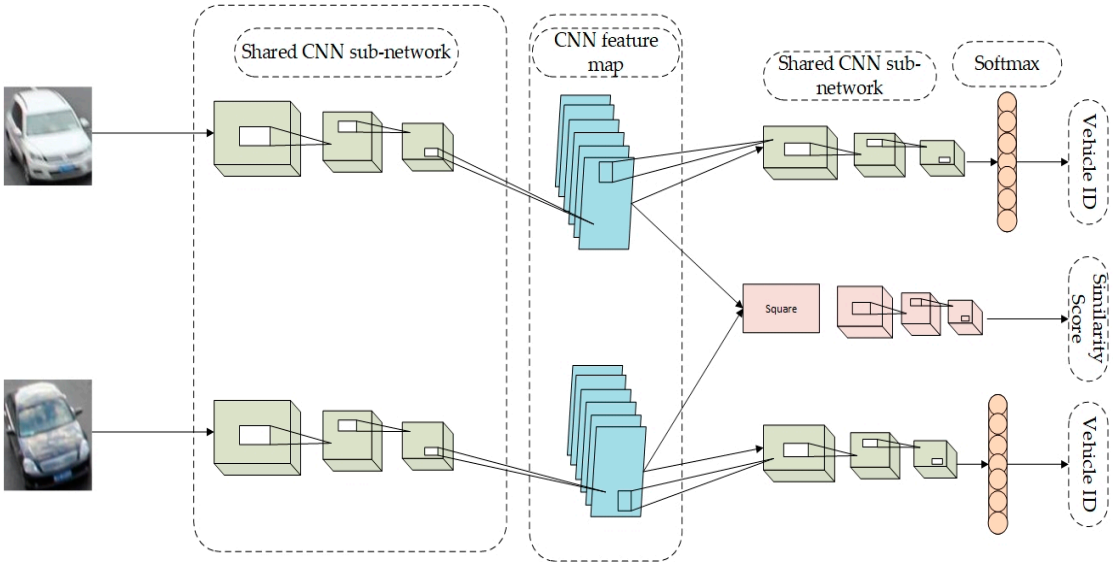

Figure 5. Siamese neural network (SNN-) based classifier.

For similarity estimation, the verification loss compares the high-level features x_1, x_2 and also directly supervises the vehicle descriptor x_1 and x_2 in the identification model. In this regard, for high-level feature comparison, we introduce a nonparametric layer which is known as a square layer. Nonparametric layers are represented as $x_r = (x_1 - x_2)^2$, where x_1, x_2 are the 4096-dim embedding and x_r is the output tensor of the nonparametric layer. To predict the probability of two query images that belong to the same class or not ($\hat{q}_1 + \hat{q}_2 = 1$), the convolutional layers and softmax function are embedded on the output tensor x_r to a 2D vector ($\hat{q}_1$, $\hat{q}_2$). x_r is used as input in the convolutional layer and filters it with a two-kernel of size $1 \times 1 \times 4096$. In this, we consider vehicle verification as a binary class task and utilize cross-entropy, which is similar to identification loss, as expressed in Equations (3) and (4):

$$\hat{q} = softmax(\theta_r \, o \, x_r) \tag{3}$$

$$\text{verify}(x_1, \, x_2, \, r, \, \theta_r) = \sum_{i=1}^{2} -q_i \, \log(\hat{q}_i) \tag{4}$$

where, r, θ_r and $\hat{q}$ represent target class (different/same), parameters of a convolutional layer and predicted probability respectively. If a pair of vehicle images is the same, then, $\hat{q}_1 = 1$, $\hat{q}_2 = 0$; otherwise, $\hat{q}_1 = 0$, $\hat{q}_2 = 1$.

The SNN-based classifier is used to extract the vehicle visual features. Input vehicles' image feature vectors are denoted by $\vec{v_n^i}$ and gallery set vehicles' feature vectors are denoted by $\vec{v_m^j}$ and these feature vectors are extracted by embedding ResNet 50 module without a final fully-connected layer. Formally, we denote a vehicle by O and $Y(O)$ represents its ID. The ith vehicle at camera n (C_n) is represented by O_n^i. The matching probability of a vehicle on the basis of its appearance can be calculated as follows:

$$P_r\left(Y\left(O_n^i\right) = Y\left(O_m^j\right) \middle| \vec{v_n^i}, \vec{v_m^j}\right) = \frac{\vec{v_n^i} \cdot \vec{v_m^j}}{\|\vec{v_n^i}\| \|\vec{v_m^j}\|} \tag{5}$$

3.3. Spatio-Temporal Pattern

In real-world monitoring scenarios, appearance information may not be enough to differentiate vehicles, especially in cases where vehicles are of the same type and model without any decoration. Here it is worth mentioning that each independent vehicle image not only provides appearance information but also provides spatio-temporal information, because, in camera surveillance environments, space and time information of any vehicle can be obtained easily. It is also possible to refine vehicle search results with the help of such spatio-temporal information. Moreover, it has been often observed that the vehicles normally follow the same roads for daily routine, which can be used as an important cue with visual features. For instance, we have two cameras A and B installed alongside the road in a series and, due to environmental constraints like road condition and destination distance, etc., the vehicle follows a specific path more often. Thus, we can get easily the entrance and exit location with the time of vehicle between two cameras A and B, as shown in Figure 6.

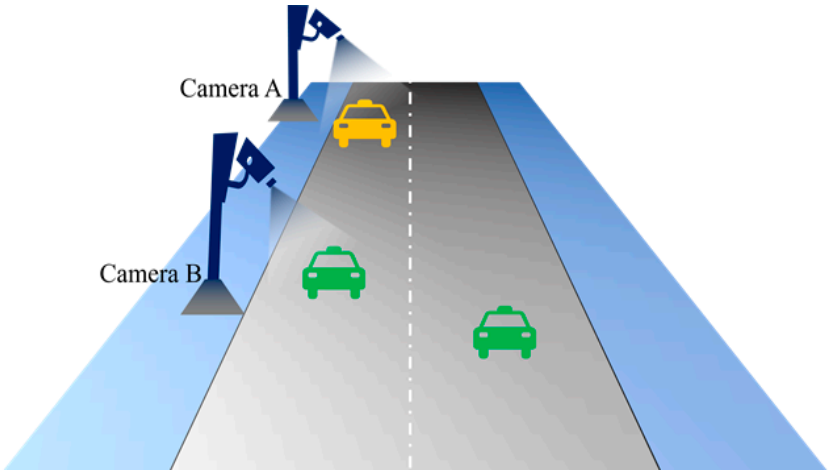

Figure 6. Illustrates the path of the vehicle between camera A and B.

The time between two consecutive cameras can be calculated as $\Delta t_{n \to m}^{i,j} = t_n^i - t_m^j$ when vehicles pass through them; here, t_n^i represents time of vehicle O_n^i when the vehicle is captured at camera C_n, t_m^j represents time of vehicle O_m^j when the vehicle is captured at camera C_m. For vehicle Re-Id, our task to find exact same vehicle as query image from different surveillance camera network and it is defined as $Pr\left(Y\left(O_n^i\right) = Y\left(O_m^j\right), n \neq m\right)$.

To calculate the spatio-temporal pattern of images between two consecutive cameras, it is necessary to know whether the vehicle is the same or not. However, it is hard to calculate in the unlabelled vehicle Re-Id dataset. To tackle this issue, we adopt the transfer learning technique to calculate the

spatio-temporal pattern by transferring an SNN-based classifier trained on a source-labelled dataset to each image pair of an unlabelled targeted dataset. After applying an SNN-based classifier to each pair of target dataset vehicle images, we are able get a rough identification result that O_n^i, the vehicle images captured at camera C_n are same as O_m^j, the vehicle images captured at camera C_m, or not. Furthermore, we can get information that demonstrates the probability of time interval and camera number of the same pair of vehicle images as well as a different pair of vehicle images. This helps us to get a rough result of a pair of images in the targeted dataset. The vehicles' spatio-temporal pattern between different surveillance cameras can be measured as:

$$Pr\left(\Delta t_{n\to m}^{i,j}, c_n, c_m | Y\left(O_n^i\right) = Y\left(O_m^j\right)\right) \tag{6}$$

3.4. Calculation of Composite Similarity Score

Lastly, to obtain significant performance with generalization capability of vehicle Re-Id, in our proposed approach, we attempt to incorporate the spatio-temporal pattern $Pr\left(\Delta t_{n\to m}^{i,j}, c_n, c_m | Y\left(O_n^i\right) = Y\left(O_m^j\right)\right)$ of the target dataset and discriminative appearance features vector. The spatio-temporal pattern is calculated using an SNN-based visual classifier of training dataset. By combining the SNN-based visual feature with spatio-temporal patterns, we calculate the composite similarity score as per Bayesian conditional probability, defined as:

$$Pr\left(Y\left(O_n^i\right) = Y\left(O_m^j\right) | \vec{v_n^i}, \vec{v_m^j}, \Delta t_{n\to m}^{i,j}, C_n, C_m\right) \tag{7}$$

Here, we have utilized visual features probability and spatio-temporal probability. O_n^i is the vehicle image of the target dataset captured by camera C_n at the time t_n^i and O_m^j is vehicle image of the same target dataset captured by camera C_m at the time t_m^j. Moreover, the visual feature vectors of these images are denoted as $\vec{v_n^i}$ and $\vec{v_m^j}$. The time between these two consecutive surveillance cameras is calculated as $\Delta t_{n\to m}^{i,j} = t_n^i - t_m^j$.

4. Experiment and Analysis

In this section, first we have given an introduction to the datasets and settings used for the evaluation of the proposed method. Then, we show the experimental outcome and analysis. In this section, we attempt to investigate the vehicle Re-Id dataset bias problem using deep CNN models (dataset classification) to demonstrate the importance of cross-dataset vehicle Re-Id research. Afterward, we quantitatively compare the state-of-the-art methods. Subsequently, qualitative analyses are performed to know the impact on the performance of vehicle Re-Id systems due to the same training and testing dataset as well as different training and testing datasets. Finally, we adopt our proposed approach to analyze significant improvement on the performance of the cross-dataset vehicle Re-Id system.

4.1. Vehicle Re-Id Benchmark Datasets

There are various datasets proposed in the literature for vehicle Re-Id problems. However, the majority of researchers for vehicle Re-Id commonly use these datasets [1,25,36]. Table 1 is a summary of well-known publicly available datasets: VeRi-776 [22], BoxCars116k [36], VehicleID [37]. A brief description of the datasets is given below:

- VeRi-776: VeRi-776 [22] is a publically available vehicle Re-Id dataset, and often adopted by the computer vision research community. Dataset images are gathered in real scenarios using surveillance cameras and the total images in the dataset are 50,000 of 776 different vehicles. Each captured vehicle images have a 2–18 viewpoint with different resolution, occlusion and illumination. Furthermore, spatio-temporal relations and license plates are annotated for all

vehicles. To make the dataset more robust, images are labelled with color, type and vehicle model. In Figure 7, various types of vehicles from VeRi-776 dataset are shown.

- BoxCars116k: The BoxCar116k [36] dataset was developed using 37 surveillance cameras, and this dataset consists of 116,286 images of 27,496 vehicles. For the preparation of dataset, 45 brands of vehicle were used. Moreover, captured images of the vehicle in the dataset are in an arbitrary viewpoint, i.e., side, back, front, and roof. All vehicle images in the dataset are annotated with a 3D bounding box, model, make and type. Images from BoxCars116k are shown in Figure 8.
- VehicleID: VehicleID [37] dataset was developed by Peking University with the funding of the Chinese national natural science foundation and national basic research program of China in the national engineering laboratory for video technology (NELVT). The vehicle dataset consists of 221,763 images of 26,267 vehicles and all the images were captured during daytime in a small town of China using multiple surveillance cameras with 10,319 vehicles' model information i.e., "Audi A6L", "MINI-cooper" and "BMW 1 Series" labelled manually. Figure 9 presents the sample images of the VehicleID dataset.

Figure 7. Depicts sample images of the VeRi-776 dataset.

Figure 8. Depicts sample images of the BoxCar116k dataset.

Figure 9. Depicts sample images of the VehicleID dataset.

Table 1. Summary of publicly available dataset.

S.No	Dataset	Year	Total Number of Images	Number of Vehicles	Images Per Vehicle	Number of Viewpoints
1	VeRi-776 [22]	2016	50,000	776	64.43	~20
2	BoxCars116k [36]	2017	116,286	27,496	4.22	~4
3	VehicleID [37]	2016	221,763	26,267	8.44	~2

4.2. Implementation Details

The approach was executed on a NVIDIA GeForce GTX 1080 GPU, 8 GB RAM (Nvidia, Santa Clara, CA, USA) memory, Ubuntu 16.04 LTS operating system with Core i7 Intel CPU (Intel, Santa Clara, CA, USA). A graphics processing unit (GPU) was used to speed up the experiment. All experiments were conducted using Pytorch. We developed the whole framework in Python which is widely used in deep learning and, mostly, the libraries used are scikit-learn, SciPy, matplotlib, and NumPy.

4.3. Evaluation Measures

To evaluate the performance of the approach, we used HIT@1 (precision at rank-1) and HIT@5 (precision at rank-5). Rank was utilized to measure the matching score of a test image to its own class, and higher value of rank indicates the improved performance of the system. Furthermore, the input has more than one ground truth, so recall and precision are examined in experiments. Therefore, we used mean average precision (*mAP*) to evaluate performance. The *mAP* measures the overall performance for vehicle Re-Id. The average precision (AP) was computed for each query as Equation (8).

$$AP = \sum_{k=1}^{n} \frac{p(k) \times f(k)}{n_{total}} \tag{8}$$

where n represents the number of vehicles in the result list, k represents the rank of retrieved vehicles, $p(k)$ represents the precision at kth position of the result, $f(k)$ equal to 1 when kth result is correctly

matched otherwise 0. The n_{total} is the number of relevant vehicle images. The mAP is mean of all the AP, defined as Equation (9).

$$mAP = \sum_{t=1}^{T} \frac{AP(t)}{T} \tag{9}$$

where T is the total number of the query images.

Classification results were generally measured in true positive (TP) and false negative (FN). TP represents the percentage of samples of a class correctly predicted by the classifier and is defined as Equation (10).

$$TP = \frac{n}{N} \times 100 \tag{10}$$

where n is the number of correctly predicted samples of a class and N denotes the total number of samples related to that class in the dataset.

FN represents the percentage of samples that are misclassified and is defined as Equation (11).

$$FN = \frac{v}{N} \times 100 \tag{11}$$

where v denotes the number of samples that are not predicted correctly of a class.

4.4. Vehicle Re-Id Dataset Classification

The state-of-the-art models assume that vehicle Re-Id datasets are not biased and all vehicle Re-Id datasets consist of a fair random sampling of vehicle images from the real-world environment. Supposing these assumptions are true and existing datasets are fulfilling gold-standard dataset criteria, then it is hard for a classifier (i.e., CNN) to distinguish between the images of each dataset if we combine all the datasets to build a single (aggregated) dataset. In other words, there exists no correlation between images of a dataset by which it can be differentiated. However, if the data in each dataset is correlated or biased then obviously the model trained on the correlated (biased) dataset may drop its performance when the dataset is changed.

To investigate the existence of a bias problem in the existing vehicle Re-Id datasets, we conducted an empirical study to prove the correlation between images of each dataset. For this purpose, we followed the proposed work of Torralba et al. [38]. We combined all the images from the VeRi-776 [22], BoxCars116k [36], and VehicleID [37] datasets to construct an aggregate dataset. The aggregate dataset consisted of equal numbers of images from each individual dataset. We removed existing labels from images and assign label to each image of aggregate dataset with its parent (original dataset) name. In our case, we had three classes of images in the aggregate dataset. For the classification task, the aggregated dataset was divided into 70% for training, 15% for validation, and 15% for testing. The training, validation and testing samples contained equal number of instances from each dataset. We trained fine-tuned Inception-v3 and VGG-16 CNN models for dataset classification.

The overall classification results in the form of a confusion matrix are shown in Table 2. Interestingly, it can be observed that the classifiers have achieved very high accuracy by correctly identifying images of their parent dataset. For example, only 2% of VeRi-776 are false classified by exploiting Inception-v3 model. This shows that the vehicle Re-Id datasets are strongly biased and there exists very low variance between images of each dataset, otherwise, it is not possible to train the network to classify the aggregated dataset into three different parent datasets (VeRi-776, BoxCars116 and VehicleID). Therefore, we conclude that these datasets are biased just because each dataset's images are captured from specific height, angle, background, resolution, viewpoint and some focused on entire scenes and others on single vehicles.

Table 2. Confusion matrix for a classifier trained to predict the parent dataset of a given re-identification (Re-Id) image.

		Predicted		
		VeRi-776	**BoxCars116k**	**VehicleID**
Test	Inception-v3			
	VeRi-776	98%	1.6%	0.4%
	BoxCars116k	1.1%	97.3%	1.6%
	VehicleID	1.3%	0.9%	97.8%
	VGG-16			
	VeRi-776	98.2%	1%	0.8%
	BoxCars116k	1.7%	96.9%	1.4%
	VehicleID	1.3%	1.6%	97.1%
	Inception-v3 (After applying data augmentation)			
	VeRi-776	67.1%	13.5%	19.4%
	BoxCars116k	19.3%	67.9%	12.8%
	VehicleID	12.9%	18.8%	68.3%

With augmentation of data, we have added more variability in aggregated dataset that simulates practical environment and tried to overcome the dataset biased problem. After applying a data augmentation technique on the aggregated dataset we again used Inception-v3 to classify aggregated dataset into three different parent datasets (VeRi-776, BoxCars116 and VehicleID). Classification results in Table 2 after applying data augmentation show increase in misclassification and this reflects that data augmentation have the ability to reduce the vehicle Re-Id dataset bias.

4.5. Single Dataset Vehicle Reidentification

In the above discussion, it is experimentally demonstrated that the current vehicle Re-Id datasets are highly biased. In this section, we report extensive experiments on different existing state-of-the-art approaches. All the conducted experiments in this setting were on the same training and testing dataset to compare the results with our proposed approach. However, our primary focus is not a single dataset vehicle Re-Id, but it is necessary to prove experimentally that the model trained on any specific dataset did not necessarily generalize well on other training datasets, which means that the training and testing datasets are different. Therefore, here we are presenting a single dataset vehicle Re-Id scenario and then in the subsequent section the cross-dataset vehicle Re-Id. For this purpose, we have compared several state-of-the-art approaches for vehicle Re-Id such as, VGG [37], GoogLeNet [39], FACT [22], ABLN-Ft-16 [40], SCCN-Ft + CLBL-8-Ft [41], SCNN [42], and DA + A + LPV [8]. This comparative study is presented in Table 3, which gives the performance comparison in terms of *mAP* on VeRi-776 dataset. The results highlight the superiority of proposed approach on many previous approaches.

Table 3. Comparison of our approach with other approaches on VeRi-776.

Method	***mAP***	**HIT@1**	**HIT@5**
VGG [37]	12.76	44.10	62.63
GoogLeNet [39]	17.89	52.32	72.17
FACT [22]	18.75	52.21	72.88
ABLN-Ft-16 [40]	24.92	60.49	77.33
SCCN-Ft + CLBL-8-Ft [41]	25.12	60.83	78.55
SCNN [42]	13.21	40.32	67.01
DA + A + LPV [8]	62.80	85.07	95.23
Ours	48.39	73.56	84.52

4.6. Cross-Dataset Vehicle Reidentification

In this section, we report a more complex experiment that coincides with practical applications. In practical cases, we normally first collect a large-scale vehicle Re-Id dataset and train a system on it. The trained system is then extended to apply on other videos or datasets to Re-Id vehicle. A realistic vehicle Re-Id system should have a promising performance on cross-datasets. Previous models mainly focus on viewpoint changes; however, the cross-database vehicle Re-Id issue was less studied.

For fair comparison, we categorized the proposed approaches to know whether it performed an optimization on the target dataset or not. We selected two datasets, VeRi-776 [22] and BoxCars116k [36], for the training, and evaluated the learned model on all datasets. Under this setting, the target dataset was unlabelled. To compare our approach with other methods, we employed the SNN-based Re-Id system. The model was trained as part of a SNN architecture [43], the same as the approach adopted in [42]. Furthermore, it was validated on the recently proposed approach [8], in which the authors reidentify vehicles based on vehicle appearance and license plate. However, they did not consider the spatio-temporal patterns to reduce the negative effect of visual features. In contrast, our proposed approach generalizes well on target datasets despite different training dataset with very large variations in terms of pose, background, illumination and viewpoint because of spatio-temporal cues. Yet, the results are significantly impressive on target datasets VehicleID [37], VeRi-776 [22] and BoxCars116k [36], due to data augmentation and spatio-temporal cues. This confirms the importance of the proposed approach that preserves the identity of vehicle. The effect of spatio-temporal cues can be seen in Table 4 for all three datasets in terms of HIT @1 and HIT @5.

Table 4. Cross-dataset vehicle Re-Id results on different settings.

Source Dataset	Target Dataset	SNN-Based Classifier		Our Complete Model	
		HIT@1	HIT@5	HIT@1	HIT@5
VeRi-776	VehicleID	30.19	37.9	54.19	69.40
BoxCars116k	VeRi-776	35.25	41.82	56.01	70.40
VeRi-776	BoxCar116k	28.19	32.17	51.91	66.50

Tables 3 and 5 show the performance comparison with baseline approaches and it can be observed that the proposed model trained on images of source dataset can also reidentify the images of the target dataset. It can be analyzed that the performance of previous models is quite weak on cross-dataset vehicle Re-Id results compared to the single dataset results. This is due to the different correlated sources and target datasets as well as lack of ability to generalize model on the distinct datasets. As our approach transfers the visual features from BoxCars116 to VeRi-776 and doesn't even use the labelled data of VeRi-776, we demonstrate the results in Figure 10 where our trained model outperforms other methods. In Figure 10, all baseline approaches use a single dataset for training and testing (i.e., VeRi-776), but, despite that, our approach with a cross-dataset shows robustness.

Table 5. Cross-dataset vehicle Re-Id results on different datasets and comparison with other approaches.

Method	Source Dataset	Target Dataset	*mAP*	HIT@1	HIT@5
SCNN [42]	BoxCars116k	VeRi-776	07.98	24.01	40.56
SCNN [42]	VeRi-776	VehicleID	09.88	28.74	48.92
DA + A + LPV [8]	VehicleID	VeRi-776	20.36	43.62	59.83
DA + A + LPV [8]	VeRi-776	BoxCar116k	23.53	47.24	64.79
Ours	VeRi-776	VehicleID	42.46	54.19	69.40
Ours	BoxCars116k	VeRi-776	44.53	56.01	70.40
Ours	VeRi-776	BoxCar116k	41.27	51.91	66.50

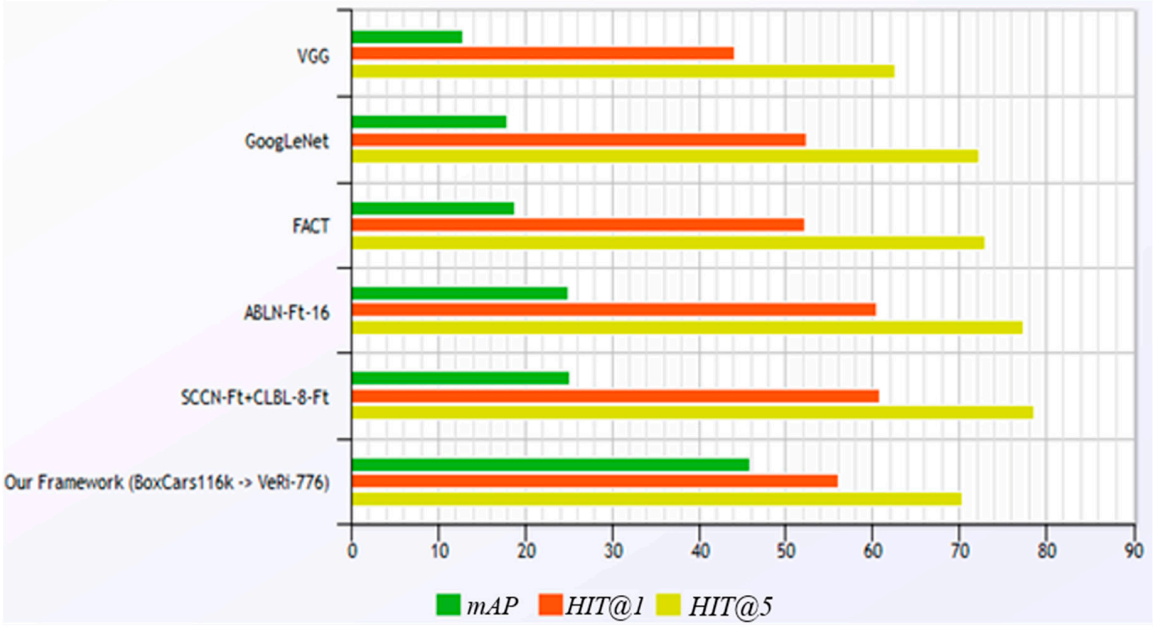

Figure 10. Shows the performance (%) comparison of our approach with other approaches on VeRi-776.

5. Discussion

It is often considered that the drop in performance of cross-dataset vehicle Re-Id systems is because of the bad dataset. However, this claim is unfair, because it is possible that the vehicle representation and different vehicle Re-Id algorithms used are not as efficient at learning the visual data that is related to a task, not to the dataset. On the other hand, in human-based scenarios, the Re-Id of objects, despite so many local visual biases, is performed effectively. So, we can conclude that both (algorithm and datasets) are equally responsible for the drop in performance of cross-dataset vehicle Re-Id. For example, when the dataset defines a specific vehicle with the front view, then there remains no reason to criticize the system for ineffectively reidentifying the target vehicle from rear view, as proved by the results (Section 4.4) on the current vehicle Re-Id dataset, which showed that datasets are extremely biased. Therefore, it is needed that the dataset should represent a real-world environment with enough images of different viewpoints, resolution, etc. per class. To overcome this issue, we have applied the data augmentation technique on the training set. As far a visual data learning problem in Re-Id algorithms is concerned, we adopt spatio-temporal cues. By using spatio-temporal information, the camera specifications have become similar to each other by reducing the negative effect of visual features like viewpoint resolution, etc.

6. Conclusions and Future Directions

The cross-dataset vehicle Re-Id problem has not received much attention before. However, it is crucial for real-world applications. In this paper, for the first time, we identified and investigated bias problems in vehicle Re-Id datasets. The defined bias problems are empirically proved on three vehicle Re-Id datasets using Inception-v3 and VGG-16 classifiers. In addition, we analyzed the cross dataset vehicle Re-Id problem on existing approaches. In this connection, we examined the performance of existing state-of-the-art methods on single and cross-datasets. For the cross-dataset vehicle Re-Id, we proposed an approach with augmentation of the training dataset to reduce the influence of pose, angle, camera color response and background information in vehicle images, and also calculated the composite similarity score of spatio-temporal pattern with SNN-based visual classifier features. Our approach generalized well when tested on three benchmark dataset VeRi-776 [22], BoxCars116k [36], and VehicleID [37]. The extensive experiments validated the effectiveness of our proposed approach on reducing dataset bias problems and robustness on the cross-datasets vehicle Re-Id.

The major reason in the progress of vehicle Re-Id is the development of large-scale real-world datasets. However, existing datasets often provide specific range of images with correlated data which cause over-fitting because parameters are over tuned on specific data. Hence, the system may not generalize effectively on other data. Therefore, the conducted empirical study leads us towards required quantitative and qualitative characteristics for next generation vehicle Re-Id dataset to fairly evaluate the algorithms. The qualitative and quantitative characteristics include:

(1) Unconstrained environment: it ensures a real traffic scenario with different weather conditions, time (day and night), traffic congestion, occluded vehicle and background clutter.
(2) Heterogeneous cameras: each vehicle should be captured by multiple surveillance cameras from different angles, resolution, viewpoint, background, lightening condition and camera settings.
(3) Image metadata: each image in dataset should be labelled with attributes like camera ID, timestamp, location, vehicle color, model, view (front, side, aerial, and rear), brand and distance between cameras of captured vehicle.
(4) License-plate information: each vehicle image in the dataset should be labelled with license number if the license plate is visible.

Including the need for a next-generation dataset, there are many aspects of the vehicle Re-Id task that can be improved. Specifically, the CNN works on edges, shapes, and original vehicle features. However, the relationship between all these features are not considered; hence the performance of the model is often found unsatisfactory when vehicle images are rotated or captured with a different rotation. In contrast, a recently introduced capsule network [44] to handle different poses, orientation, and occluded objects is yet to be explored to enhance the model performance.

Author Contributions: Conceptualization, Z.; Formal analysis, M.U.A.; Methodology, Z. and M.S.K.; Supervision, J.C.; Writing—original draft, Z.; Writing—review and editing, J.D.; Software, R.K. All authors have read and agreed to the published version of the manuscript.

Funding: This paper has been supported The National Key Research and Development Program of China (2017YFC0821505), Funding of Zhongyanggaoxiao ZYGX2018J075 and also supported by Sichuan Science and Technology Program 2019YFS0487.

Acknowledgments: I am grateful to my worthy supervisor as well as all the lab mates for their endless support.

Conflicts of Interest: There is no conflict between authors.

References

1. Lou, Y.; Bai, Y.; Liu, J.; Wang, S.; Duan, L.Y. Embedding adversarial learning for vehicle re-identification. *IEEE Trans. Image Process.* **2019**, *28*, 3794–3807. [CrossRef] [PubMed]
2. Wu, Y.; Lin, Y.; Dong, X.; Yan, Y.; Bian, W.; Yang, Y. Progressive learning for person re-identification with one example. *IEEE Trans. Image Process.* **2019**, *28*, 2872–2881. [CrossRef] [PubMed]
3. Rasouli, A.; Kotseruba, I.; Tsotsos, J.K. Understanding pedestrian behavior in complex traffic scenes. *IEEE Trans. Intell. Veh.* **2018**, *3*, 61–70. [CrossRef]
4. Chen, Z.; Liao, W.; Xu, B.; Liu, H.; Li, Q.; Li, H.; Xiao, C.; Zhang, H.; Li, Y.; Bao, W.; et al. Object tracking over a multiple-camera network. In Proceedings of the IEEE International Conference on Multimedia Big Data, Beijing, China, 20–22 April 2015; pp. 276–279.
5. Zakria; Cai, J.; Deng, J.; Khokhar, M.S.; Umar, A.M. Vehicle classification based on deep convolutional neural networks model for traffic surveillance systems. In Proceedings of the 15th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), Chengdu, China, 14–16 December 2018; pp. 224–227.
6. Arandjelovic, R.; Zisserman, A. Three things everyone should know to improve object retrieval. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 2911–2918.
7. Yan, L.; Wang, Y.; Song, T.; Yin, Z. An incremental intelligent object recognition system based on deep learning. In Proceedings of the Chinese Automation Congress (CAC), Jinan, China, 20–22 October 2017; pp. 7135–7138.

8. Zakria; Cai, J.; Deng, J.; Aftab, M.; Khokhar, M.; Kumar, R. Efficient and deep vehicle re-identification using multi-level feature extraction. *Appl. Sci.* **2019**, *9*, 1291. [CrossRef]
9. Tian, M.; Yi, S.; Li, H.; Li, S.; Zhang, X.; Shi, J.; Yan, J.; Wang, X. Eliminating background-bias for robust person re-identification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 5794–5803.
10. Farenzena, M.; Bazzani, L.; Perina, A.; Murino, V.; Cristani, M. Person re-identification by symmetry-driven accumulation of local features. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 2360–2367.
11. Zhao, R.; Ouyang, W.; Wang, X. Learning mid-level filters for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 24–27 June 2014; pp. 144–151.
12. Yi, D.; Lei, Z.; Li, S.Z. Deep metric learning for practical person re-Identification. In Proceedings of the 2014 22nd International Conference on Pattern Recognition, Stockholm, Sweden, 24–28 August 2014; pp. 207–244.
13. Xiong, F.; Gou, M.; Camps, O.; Sznaier, M. Person re-identification using kernel-based metric learning methods. In Proceedings of the European Conference on Computer Vision—ECCV, Zurich, Switzerland, 6–12 September 2014; pp. 1–16.
14. Li, W.; Zhao, R.; Xiao, T.; Wang, X. Deepreid: Deep filter pairing neural network for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 24–27 June 2014; pp. 152–159.
15. Ahmed, E.; Jones, M.; Marks, T.K. An improved deep learning architecture for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 8–10 June 2015; pp. 3908–3916.
16. Gheissari, N.; Sebastian, T.; Hartley, R. Person Reidentification Using Spatiotemporal Appearance. In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), New York, NY, USA, 17–22 June 2006; Volume 2, pp. 1528–1535.
17. Wu, Z.; Li, Y.; Radke, R.J. Viewpoint invariant human re-identification in camera networks using pose priors and subject-discriminative features. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1095–1108. [CrossRef] [PubMed]
18. Wang, F.; Zuo, W.; Lin, L.; Zhang, D.; Zhang, L. Joint Learning of Single-Image and Cross-Image Representations for Person Re-identification. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 1288–1296.
19. Liao, S.; Hu, Y.; Zhu, X.; Li, S.Z. Person re-identification by Local Maximal Occurrence representation and metric learning. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 2197–2206.
20. Zhao, R.; Ouyang, W.; Wang, X. Unsupervised Salience Learning for Person Re-identification. In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 3586–3593.
21. Zheng, L.; Shen, L.; Tian, L.; Wang, S.; Wang, J.; Tian, Q. Scalable Person Re-identification: A Benchmark. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 11–18 December 2015; pp. 1116–1124.
22. Liu, X.; Liu, W.; Ma, H.; Fu, H. Large-scale vehicle re-identification in urban surveillance videos. In Proceedings of the 2016 IEEE International Conference on Multimedia and Expo (ICME), Seattle, WA, USA, 11–15 July 2016; pp. 1–6.
23. Zhu, J.; Zeng, H.; Du, Y.; Lei, Z.; Zheng, L.; Cai, C. Joint feature and similarity deep learning for vehicle re-identification. *IEEE Access* **2018**, *6*, 43724–43731. [CrossRef]
24. Tang, Y.; Wu, D.; Jin, Z.; Zou, W.; Li, X. Multi-modal metric learning for vehicle re-identification in traffic surveillance environment. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 2254–2258.
25. Bai, Y.; Lou, Y.; Gao, F.; Wang, S.; Wu, Y.; Duan, L.-Y. Group-sensitive triplet embedding for vehicle reidentification. *IEEE Trans. Multimed.* **2018**, *20*, 2385–2399. [CrossRef]
26. Liu, X.; Zhang, S.; Huang, Q.; Gao, W. RAM: A region-aware deep model for vehicle re-identification. In Proceedings of the 2018 IEEE International Conference on Multimedia and Expo (ICME), San Diego, CA, USA, 23–27 July 2018; pp. 1–6.

27. Javed, O.; Shafique, K.; Rasheed, Z.; Shah, M. Modeling inter-camera space–time and appearance relationships for tracking across non-overlapping views. *Comput. Vis. Image Underst.* **2008**, *109*, 146–162. [CrossRef]
28. Xu, J.; Jagadeesh, V.; Ni, Z.; Sunderrajan, S.; Manjunath, B.S. Graph-Based Topic-Focused Retrieval in Distributed Camera Network. *IEEE Trans. Multimed.* **2013**, *15*, 2046–2057. [CrossRef]
29. Ellis, T.; Makris, D.; Black, J.; Engineers, E. Learning a multi-camera topology. In Proceedings of the In Joint IEEE Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, Lausanne, Switzerland, 12–13 October 2003; pp. 165–171.
30. Loy, C.C.; Xiang, T.; Gong, S. Multi-camera activity correlation analysis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 1988–1995.
31. Liu, X.; Liu, W.; Mei, T.; Ma, H. PROVID: Progressive and multimodal vehicle reidentification for large-scale urban surveillance. *IEEE Trans. Multimed.* **2018**, *20*, 645–658. [CrossRef]
32. Lv, J.; Chen, W.; Li, Q.; Yang, C. Unsupervised Cross-Dataset Person Re-identification by Transfer Learning of Spatial-Temporal Patterns. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7948–7956.
33. Fan, H.; Zheng, L.; Yang, Y. Unsupervised Person Re-identification: Clustering and Fine-tuning. *ACM Trans. Multimed. Comput. Commun. Appl.* **2017**, *14*, 1–18. [CrossRef]
34. Lv, J.; Wang, X. Cross-Dataset Person Re-identification Using Similarity Preserved Generative Adversarial Networks. In Proceedings of the International Conference on Knowledge Science, Engineering and Management, Changchun, China, 17–19 August 2018; pp. 171–183.
35. Zheng, Z.; Zheng, L.; Yang, Y. A discriminatively learned CNN embedding for person reidentification. *ACM Trans. Multimed. Comput. Commun. Appl.* **2017**, *14*, 1–20. [CrossRef]
36. Sochor, J.; Spanhel, J.; Herout, A. Boxcars: Improving fine-grained recognition of vehicles using 3-D bounding boxes in traffic surveillance. *IEEE Trans. Intell. Trans. Syst.* **2019**, *20*, 97–108. [CrossRef]
37. Liu, H.; Tian, Y.; Wang, Y.; Pang, L.; Huang, T. Deep relative distance learning: Tell the difference between similar vehicles. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 2167–2175.
38. Hou, J.; Zeng, H.; Cai, L.; Zhu, J.; Chen, J.; Ma, K.-K. Multi-label learning with multi-label smoothing regularization for vehicle re-identification. *Neurocomputing* **2019**, *345*, 15–22. [CrossRef]
39. Yang, L.; Luo, P.; Loy, C.C.; Tang, X. A large-scale car dataset for fine-grained categorization and verification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3973–3981.
40. Zhou, Y.; Shao, L. Vehicle re-identification by adversarial bi-directional LSTM network. In Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 12–15 March 2018; pp. 653–662.
41. Zhou, Y.; Liu, L.; Shao, L. Vehicle re-identification by deep hidden multi-view inference. *IEEE Trans. Image Process.* **2018**, *27*, 3275–3287. [CrossRef] [PubMed]
42. Yi, D.; Lei, Z.; Liao, S.; Li, S.Z. Deep metric learning for person re-identification. In Proceedings of the 22nd International Conference on Pattern Recognition, Stockholm, Sweden, 24–28 August 2014; pp. 34–39.
43. Hadsell, R.; Chopra, S.; LeCun, Y. Dimensionality reduction by learning an invariant mapping. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), New York, NY, USA, 17–22 June 2006; pp. 1735–1742.
44. Sabour, S.; Frosst, N.; Hinton, G.E. Dynamic routing between capsules. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 7–9 December 2017; pp. 3859–3869.