

Open Access

<https://doi.org/10.48130/dts-0024-0019>
Digital Transportation and Safety 2024, 3(4): 210–219

A privacy-compliant approach to responsible dataset utilisation for vehicle re-identification

Yan Qian¹, Johan Barthélemy², Bo Du^{3*} and Jun Shen¹

¹ School of Computing and Information Technology, University of Wollongong, Wollongong, NSW 2522, Australia

² NVIDIA, Santa Clara, CA 95051, USA

³ Department of Business Strategy and Innovation, Griffith University, Brisbane, QLD 4111, Australia

* Corresponding author, E-mail: bo.du@griffith.edu.au

Abstract

Modern surveillance systems increasingly adopt artificial intelligence (AI) for their automated reasoning capacities. While AI can save manual labor and improve efficiency, addressing the ethical concerns of such technologies is often overlooked. One of these AI application technologies is vehicle re-identification - the process of identifying vehicles through multiple cameras. If vehicle re-identification is going to be used on and with humans, we need to ensure the ethical and trusted operations of these systems. Creating reliable re-identification models relies on large volumes of training datasets. This paper identifies, for the first time, limitations in a commonly used training dataset that impacts the research in vehicle re-identification. The limitations include noises due to writing on images and, most importantly, visible faces of drivers or passengers. There is an issue if facial recognition is indirectly performed by these black box models as a by-product. To this end, an approach using an image-to-image translation model to generate less noisy training data that can guarantee the privacy and anonymity of people for vehicle re-identification is proposed.

Keywords: Vehicle re-identification; Image-to-image translation; Responsible data generation; Artificial intelligence ethics

Citation: Qian Y, Barthélemy J, Du B, Shen J. 2024. A privacy-compliant approach to responsible dataset utilisation for vehicle re-identification. *Digital Transportation and Safety* 3(4): 210–219 <https://doi.org/10.48130/dts-0024-0019>

Introduction

Cameras are ubiquitous and deployed in public areas, producing large quantities of video data. To get a better understanding of the dynamic traffic systems, vehicle re-identification (i.e., vehicle reID or v-reID) can be performed. Each camera in the road network captures an image of every passing vehicle and stores them in a gallery. The cameras can have overlapping or non-overlapping views. Given a query image, i.e., an image taken from a vehicle, v-reID aims to re-identify the vehicle by finding its occurrences from the gallery. Like any deep learning model, a lot of training data is required for v-reID models. Two of the widely used and publicly available datasets are VeRi-776^[1] and VehicleID^[2]. Although they are commonly used, to the best of our knowledge, there is no existing research that specifically comments on the content and limitations of such datasets.

Artificial intelligence (AI) ethics refers to a framework of ethical principles and methods aimed at guiding the responsible creation and use of AI technologies^[3]. There is no prior research on ethics regarding v-reID. We assume that the reason behind is that no humans are directly involved in the re-identification process. However, what about humans being indirectly involved? One major issue is that some faces of drivers and passengers in the dataset are visible. Because existing reID models are black boxes, it is unknown what happens behind the scenes. In addition, we notice other observations in the datasets that may decrease the performance of v-reID models. This paper uses CycleGAN, an image-to-image translator, to tackle these limitations.

While the datasets in this paper have been extremely valuable for the v-reID research, they are not without their limitations. The primary purpose of this paper is to highlight the limitations within commonly used training datasets for v-reID and propose solutions to address these concerns. By identifying and addressing these limitations, we aim to promote safer research practices and raise awareness about the ethical implications of AI models involving humans. In this way, we can promote safer research, educate our readers, and build a more human-centered AI. This paper proposes the use of existing image-to-image translation models, such as CycleGAN, to mitigate the identified limitations. By employing CycleGAN, we aim to train it using two datasets: one containing the limitations present in current datasets and another without these limitations. Through this approach, we can generate images absent of the identified limitations. In general, our main contribution is threefold: (1) we discuss for the first time about the ethics involved in vehicle re-identification; (2) different limitations of existing and widely used datasets that can impede the performance of reID models are discussed, and (3) a method is developed to generate data that is anonymous and respects the privacy of drivers and passengers.

Literature review

AI ethics principles

To ensure that AI is reliable throughout its lifecycle, the Australian Government has proposed eight voluntary AI ethics principles^[4]. If V-reID were to be deployed in public and used in

Privacy-compliant vehicle re-identification

multi-camera vehicle tracking and traffic surveillance^[5], the following considerations, among others, would need to be prioritised:

- **Model:** the deep learning models used for v-reID should be transparent for users and those that are impacted. Users should comprehend the hows and whys of the models, while the general public should receive a more accessible education on the AI systems (principle: transparency and explainability).

- **Data:** the data collection and usage should guarantee the anonymity of drivers and passengers, i.e., their privacy should be protected and respected (principle: privacy protection and security). Furthermore, the v-reID model should enable diversity and should not perform any sort of unjustified surveillance (principle: human-centered values).

To the best of our knowledge, limited studies have been done on re-identification in consideration of ethical issues. Dietlmeier et al.^[6] anonymized datasets by blurring faces on person reID benchmarks and demonstrated that in doing so did not compromise the performance of person reID. To introduce privacy and security in person reID, Ahmad et al.^[7] solved person reID using event cameras. The latter captures dynamic scenes by responding to brightness changes only without providing any RGB image content. Richardwebster et al.^[8] used saliency maps to help differentiate between individuals that are visually similar. This helps to better understand the person reID model's decisions and helps to reduce false matches in high-stake reID, such as autonomous driving, criminal justice, and healthcare^[9]. There is a need to set boundaries or regulations to ensure a safer practice of v-reID research.

Limitations of deep learning models

Methods using Deep Neural Networks (DNNs), such as Convolutional Neural Networks (CNNs)^[10–16], Recurrent Neural Network (RNNs)^[17] and Transformers^[18–22], have been extensively used in v-reID. Although v-reID has resulted in improvements in performance, these deep learning models still remain black boxes. This means that the internal inference processes are either unknown or non-interpretable to us^[23].

On one hand, some research primarily aims to improve the performance of v-reID models using esoteric algorithms. However, the outcomes are not useful for real-world applications (or called open-world re-identification^[24]). Some works have tackled the issue of rendering re-identification more explainable in the context of persons^[25] and vehicles^[19]. Chen et al.^[25] proposed an Attribute-guided Metric Distillation (AMD) method that learns an interpreter that uses semantic attributes to explain the results of person reID methods. The interpreter is capable of quantifying the contributions of attributes so that users can know what attributes differentiate two people. Moreover, it can visualize attention maps of attributes to show what the most significant attributes are. Our previous work^[19] aimed to render vehicle reID research more digestible. A step-by-step guide was proposed on how to train a reID model, as existing research papers are more complex.

Conversely, people cannot entirely trust the results produced by these black-box models. A small alteration of an image that is undetectable for the human eye, can lead to a DNN being confused, thinking it is something completely different^[26]. Moreover, even though v-reID does not aim to re-identify people directly, facial recognition can happen as a by-product if the training data involves images of humans. This means that data should guarantee the anonymity of people or that deep

learning models should not perform any unknown behind-the-scenes facial recognition.

Limitations of vehicle re-identification datasets

With the growing introduction of publicly available vehicle large-scale datasets, v-reID has increased in popularity. Some publicly available v-reID datasets include: Comprehensive Cars (CompCars)^[27], PASCAL VOC^[28], PKU-VD1 and PKU-VD2^[29], Vehicle-1M^[30], CityFlow^[31], VERI-Wild^[32], PKU VehicleID^[2], and VeRi-776^[1]. In this paper, the spotlight is on the VehicleID and VeRi-776 datasets, due to their relevance as widely recognized benchmarks in the field of v-reID. Meanwhile, VehicleID is a large-scale dataset with controlled views, VeRi-776 provides real-world diversity.

The VehicleID^[2] dataset contains 221,763 images of 26,267 vehicle images captured by multiple real-world surveillance cameras in a small city in China, offering one of the largest available collections for v-reID. This dataset focuses primarily on front and rear viewpoints. VeRi-776^[1] is a vehicle re-identification dataset which contains 49,357 images of 776 vehicles from 20 cameras. The dataset is collected in the real traffic scenario, also in China. In contrast to VehicleID, VeRi-776 presents a more complex and realistic dataset by including varying environmental conditions, such as different times of day and diverse camera angles. Samples of both datasets are shown in Fig. 1.

VehicleID and VeRi-776 are compared in terms of four categories: resolution, privacy, noise, and views.

Resolution

The images captured by the different cameras are in higher resolution for VehicleID compared to VeRi-776. This is important for v-reID as every local feature is taken into consideration when performing the re-identification task. The smallest detail, such as the size, shape, or color of the windshield sticker is crucial.

Privacy

Unfortunately, the high resolution has its flaws. In the case of VehicleID, the faces of drivers and passengers are visible. This is not privacy-compliant. Fig. 2 shows examples of VehicleID data where faces are clearly visible when zooming in. This is an issue if data were leaked and distributed, and the faces of people, including children, could be accessed. Furthermore, the AI could learn the faces instead of the vehicle features by performing facial recognition indirectly. In the worst scenario, the AI could be performing unjustified surveillance without our awareness.

Noise

Some images in VehicleID have red writing on them, see samples in Fig. 2. The writing indicates the timestamp and trip-related information. This adds additional noise to the reID model which can therefore impact its performance either by thinking that the red writings are essential to the reID process, or by being confused and deducing that it has nothing to do with vehicles anymore.

Views

Finally, VehicleID is captured by an unknown number of cameras from two views (e.g. front and back), while the images in VeRi-776 with the same identity have eight different appearances captured from 20 different camera views. A reID dataset should preferably contain vehicles from various perspectives.



Fig. 1 Sample images of different vehicle models from (a) the VehicleID and (b) VeRI-776 datasets.

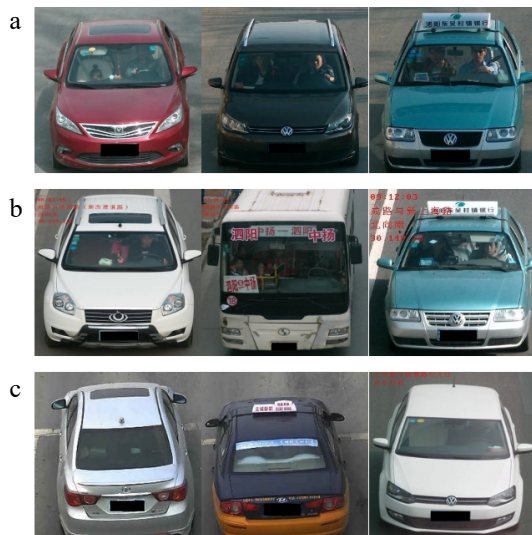


Fig. 2 The pros and cons of the VehicleID dataset: (a) visible faces, (b) noisy red time stamp, (c) only two directions. (The images were captured in high quality).

Image-to-image translation for vehicle re-identification

Image-to-image translation^[33] involves learning a mapping function between two domains X and Y , to generate an image from one domain to another domain. Examples include converting images from greyscale to color, from synthetic objects to real objects, or photos to paintings. In v-reID, image-to-image translation becomes useful in two cases: (1) when training data is scarce and needs to be augmented, or (2) implementing a model that can generalize, i.e., a model that is trained on a source dataset X and then applied on a target dataset Y , where X and Y are of different domains. This is also called domain adaptation (DA)^[34]. The differences in domains can be due to illumination, view, or environment changes, or different camera networks, types, settings and resolutions.

Let X be the source domain and Y the target domain. The goal is to construct a mapping function from X to Y , i.e., given an image $x \in X$, we want to function a mapping function that transforms x into $y \in Y$. Traditionally, image-to-image translation techniques needed paired data. This means that for the same examples in X , the same examples in Y with the required modifications are provided, e.g. a dog from a certain position and a drawing of the same dog in the same position. An example of paired image-to-image translation is Pix2Pix^[33]. While paired image-to-image translation produces great results, the requirement for training data is very limited and expensive to prepare, or even impossible depending on the field, e.g., v-reID. As a consequence, Zhu et al.^[35] introduced CycleGAN which is an approach to unpaired image-to-image translation built on the Generative Adversarial Network (GAN)^[36].

Image-to-image translation techniques have been widely used in vehicle reID^[37–40]. Given an input view of a vehicle, Zhou & Shao^[37] generated cross-view vehicle images. Wang et al.^[38] and Luo et al.^[40] augmented their data using image-to-image translation techniques, such as SPGAN^[41] and CycleGAN^[35], while Zhou et al.^[39] employed a GAN-Siamese network to transform images from day-time domain to night-time domain, and vice versa. However, to the best of our knowledge, no prior work has tackled the safety issue of the existing datasets using image-to-image translation thus far.

Existing works transfer image styles to augment their data or to adapt to the different domains. The above-mentioned shortcomings in terms of noise, views, and privacy for VehicleID can limit the performance of reID models. To this end, we propose a method to circumvent these issues related to the dataset. This is the very first attempt to tackle this topic.

Proposed method

CycleGAN is an image-to-image translation technique that does not require paired examples. This means that the training data does not need to be related, i.e., CycleGAN requires a source set $\{x_i\}_{i=1}^N$ and a target set $\{y_j\}_{j=1}^M$ with no information on which $x_i \in X$ matches to which $y_j \in Y$. This flexibility makes

Privacy-compliant vehicle re-identification

CycleGAN particularly well-suited for our task that works with v-reID datasets that do not have a direct one-to-one relationship. Additionally, CycleGAN has been widely adopted in research due to its versatility in applications such as style transfer, domain adaptation, and data augmentation. Its open-source code is well-documented and easy to understand, making it accessible for researchers and developers to implement and customise.

CycleGAN is constructed using GANs, and GAN is an unsupervised technique that is built on two models: a generator model and a discriminator model. As the name suggests, the generator generates outputs from the domain, and the discriminator receives synthetic data from the generator and the real dataset. The discriminator then determines whether its input is real or fake (generated). Both models are trained until the generator has learned to fool the discriminator and the discriminator is not able to distinguish real from generated data.

CycleGAN involves training two GANs, i.e., training two generators and two discriminators simultaneously. Let X and Y be two domains with training samples $\{x_i\}_{i=1}^N$ ($x_i \in X$) and $\{y_j\}_{j=1}^M$ ($y_j \in Y$). The two generators are $G: X \rightarrow Y$ and $F: Y \rightarrow X$. Generator G tries to generate images $G(x)$ that look similar to images from domain Y . Likewise, generator F aims to generate images $F(y)$ that look most similar to images from domain X . The two discriminators are D_X and D_Y , where D_Y aims to distinguish between images $\{x\}$ in X and the generated images $\{F(y)\}$, and D_X aims to discern between $\{y\}$ and $\{G(x)\}$.

The overall objective function involved in training consists of two adversarial losses^[36] and a cycle consistency loss^[42]. The adversarial losses are applied to both mapping functions G and F , and match the distribution of the generated images to the data distribution in the target domain:

$$\mathcal{L}_{GAN}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{data(y)}} [\log(D_Y(y))] + \mathbb{E}_{x \sim p_{data(x)}} [\log(1 - D_Y(G(x)))] \quad (1)$$

$$\mathcal{L}_{GAN}(F, D_X, Y, X) = \mathbb{E}_{x \sim p_{data(x)}} [\log(D_X(x))] + \mathbb{E}_{y \sim p_{data(y)}} [\log(1 - D_X(G(y)))] \quad (2)$$

where, $x \sim p_{data(x)}$ and $y \sim p_{data(y)}$ denote the data distributions of X and Y , respectively. Generator G tries to minimize Eqn (1), while adversary D_Y aims to maximize it, i.e., $\min_G \max_{D_Y} \mathcal{L}_{GAN}(G, D_Y, X, Y)$. Similarly, generator F tries to

minimize Eqn (2), while the adversary D_X aims to maximize it, i.e., $\min_F \max_{D_X} \mathcal{L}_{GAN}(F, D_X, Y, X)$.

The cycle consistency loss aims to reduce the space of possible mapping functions, and therefore to encourage G and F to be bijections. Zhu et al.^[35] illustrated this property with an example in translation: if a sentence is translated from English to French, and the output is translated back to English from French, the result should be the original sentence^[43]. It indicates that both mapping functions should be cycle-consistent, i.e., $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$ (forward cycle consistency) and $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$ (backward cycle consistency). The cycle consistency loss is then formulated as Eqn (3):

$$\mathcal{L}_{cyc}(G, F) = \mathbb{E}_{x \sim p_{data(x)}} [\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_{data(y)}} [\|G(F(y)) - y\|_1] \quad (3)$$

The full objective function is:

$$\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{GAN}(G, D_Y, X, Y) + \mathcal{L}_{GAN}(F, D_X, Y, X) + \lambda \mathcal{L}_{cyc}(G, F) \quad (4)$$

where, λ controls the relative importance of the two objectives. Finally, the aim is to solve for Eqn (4):

$$G^*, F^* = \operatorname{argmin}_{G, F} \max_{D_X, D_Y} \mathcal{L}(G, F, D_X, D_Y) \quad (5)$$

Details of CycleGAN are available in Zhu et al.^[35].

In the context of our work and for the sake of visualization, Fig. 3 shows how CycleGAN can be employed to translate images from the synthetic domain to the real domain, and vice-versa. The two chosen datasets are unpaired, i.e., the images are captured at different locations and times. This means that we do not have the exact same correspondence in both datasets. CycleGAN consists of GAN1, which transfers photos from the real domain to the synthetic domain, and GAN2, transferring images from the synthetic domain to generate images from the real domain, whilst solving Eqn (5). We use CycleGAN to transfer images from VehicleID to VeRi-776 to tackle the issues related to privacy, noise, and views mentioned previously.

Experiments

CycleGAN is trained on VeRi-776 and VehicleID, and employed to transfer images from VehicleID to VeRi-776 style.

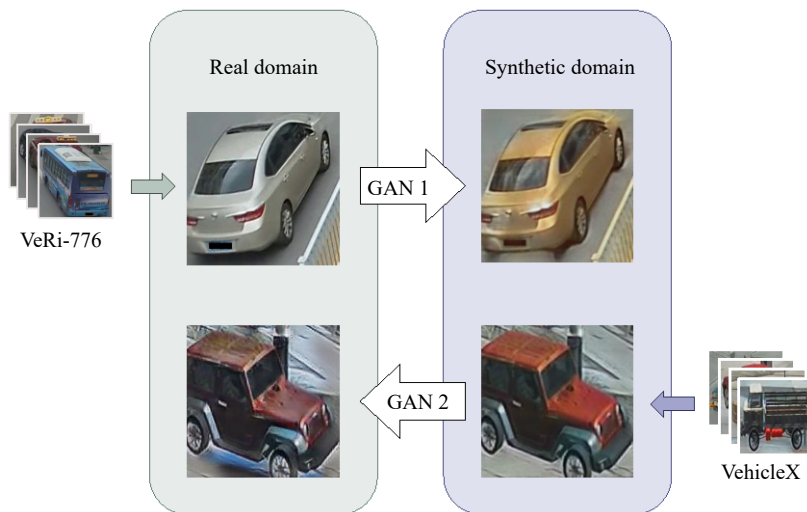


Fig. 3 CycleGAN model architecture transferring between VeRi-776 (real domain) and VehicleX (synthetic domain).

This section presents the implementation details and the results generated from experiments VCGAN-A and VCGAN-B.

Implementation details

Two experiments have been run using the default settings of CycleGAN. Both were trained for 100 epochs with an initial learning rate of 0.0002, followed by 200 epochs with a linear learning rate decay. The default optimiser is Adam^[44]. The input images were scaled to 256 × 256 pixels and each experiment was trained on three NVIDIA A100 80GB GPUs.

The difference between the two experiments lies in the training size, as shown in Table 1. VCGAN-A is trained using the same number of data in both datasets, while VCGAN-B is trained using the entirety of both datasets. While a testing set is not required for the training of CycleGAN, it is beneficial for evaluating the model's performance on unseen data. In the next paragraphs, examples from the VehicleID test set will be used to assess the performance of the trained models. In the following paragraphs, VCGAN-A and VCGAN-B will also be referred to as 'A' and 'B' respectively for simplicity.

Observations

The properties of datasets mentioned previously and how CycleGAN addresses them are described in this subsection, including four categories: resolution, privacy, noise, and views.

Resolution

The resolution is crucial in the v-reID process. The higher the visibility of the details, the more discriminant features can be extracted and can therefore aid the v-reID decision. As we train CycleGAN on transforming images from VehicleID to VeRi-776,

Table 1. Training data sizes for VCGAN-A and VCGAN-B.

	VehicleID	VeRi-776
VCGAN-A	37,778	37,778
VCGAN-B	113,346	37,778

which has a decreased resolution compared to VehicleID, it is only logical that the images output by CycleGAN also present a lower resolution. This could be problematic depending on what features get lost. If we do not want to perform any license plate or facial recognition, losing features related to the license plates, faces, or camera descriptions, becomes an advantage. However, if the lost features include logos or windshield stickers, then this can influence the performance of the v-reID model.

Privacy

Privacy is an important factor that needs to be addressed. Even though the reID model is a black box, we need to ensure that no facial recognition is unintentionally conducted. As shown in Fig. 4 (left), we notice that in experiments A and B, the faces are blurred out, such that they are not visible anymore. While the outputs of A are satisfactory, the transformed images of B appear much smoother (look at the windshield, where there are fewer 'white strokes'). CycleGAN can erase the faces from the windshields.

Noise

The model might incorporate the red timestamps in its re-identification decision. After all, these timestamps are more noise than anything else. One could circumvent this by cropping the image, such that the timestamps are not visible anymore. While this would be possible for some samples, e.g., Fig. 4a (right), this wouldn't work for others, e.g., Fig. 4b–d (right), as a chunk of the vehicle would then be cut out. The results look promising when transforming the images using the trained CycleGAN. The outputs in Fig. 4 (right) for A and B show that the red writing is completely (Fig. 4a–c right), or partially (Fig. 4d right) erased.

Views

Finally, the cars in VehicleID are captured from two views (back and front), while VeRi-776 has more. This characteristic is visible in Fig. 5, where given a vehicle input from the front, CycleGAN generates vehicles from different views: side or back.

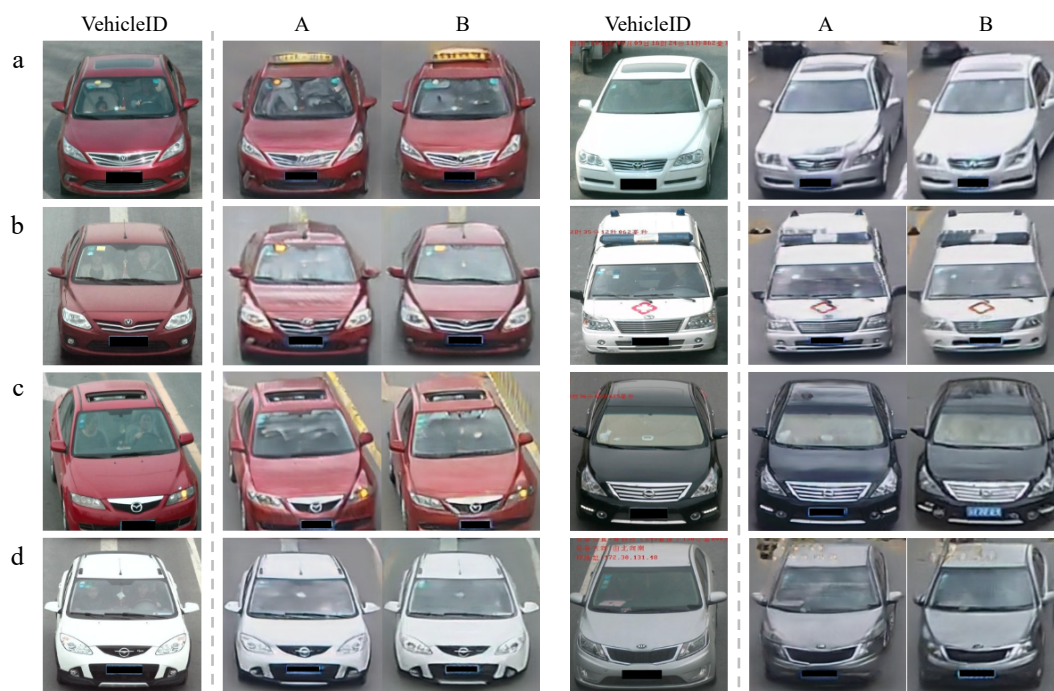


Fig. 4 Left: faces are blurred out, such that the privacy of the driver is ensured; Right: the timestamps in red are (partially) erased.

Privacy-compliant vehicle re-identification

CycleGAN managed to erase or blur out the faces of people as well as to remove the noisy red timestamps from the images. This is beneficial as this renders the dataset anonymous as well as less distractive. On the other hand, the resulting images are of lower resolution, and potential discriminative features get lost.

Unexpected outcomes

Besides our observations, we summarize additional outcomes that can benefit the research. These outcomes were unexpected and are worthy of mention. It is worth mentioning that this is not an exhaustive list, and there might be other elements we failed to notice.

Background

One of the unexpected outcomes is that the background, such as the lane or the pedestrian crossing markings, were removed, as shown in Fig. 6 (left). This can aid the v-reID model in focusing only on the vehicle rather than the background, as the latter is useless for reID.

Partial vehicles

Surprisingly, CycleGAN handled cropped vehicles well. As shown in Fig. 6 (right), it can be observed that even though the vehicles were partially captured due to their position or camera glitch, CycleGAN managed to transform the vehicle only.

I spy with my little eye

CycleGAN saw certain objects in specific types of images that we did not perceive with our eyes, as shown in Fig. 7. Some vehicles in red that were captured from a specific frontal view were transformed into a taxi (Fig. 7a), highlighting one potential bias of the dataset. Fig. 8a shows four images taken of different taxis. We don't know the ratio of taxis and red vehicles. However, given the confusion by CycleGAN, the ratio must be on the higher end. VeRi-776 doesn't have many images captured during the nighttime. Hence, CycleGAN misinterprets images that are taken at night. Depending on the image, two scenarios result: a bush or an abstract image of the vehicle. On the one hand, when there are some variations of yellow or green on the bottom, CycleGAN detects bushes (Fig. 7b). As we



Fig. 5 Generated views by CycleGAN.

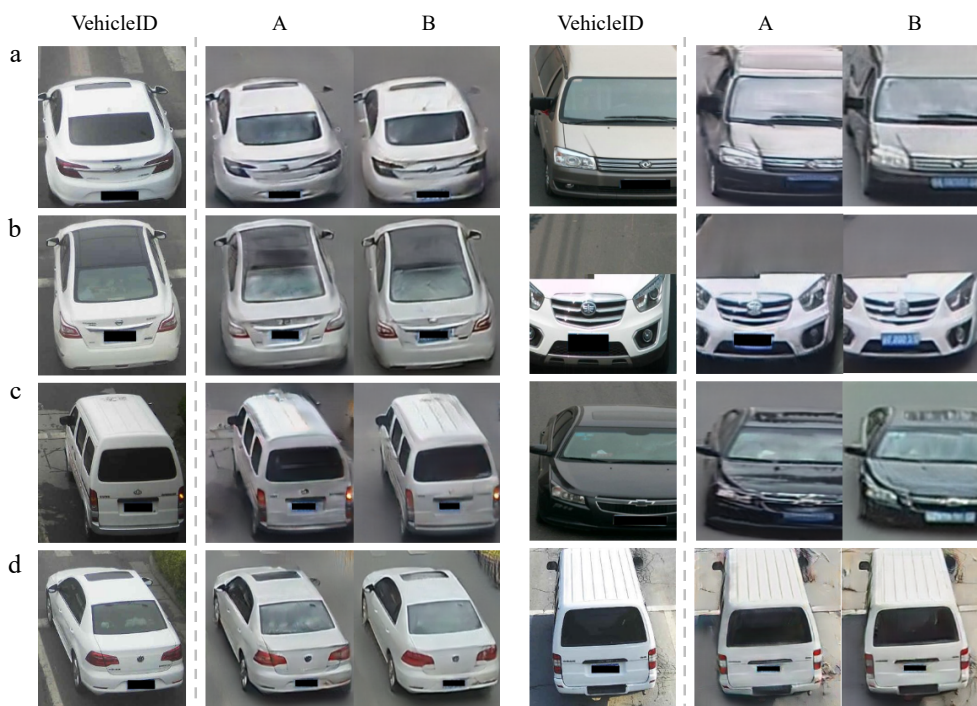


Fig. 6 Left: background removal; Right: cropped vehicles still look like vehicles.

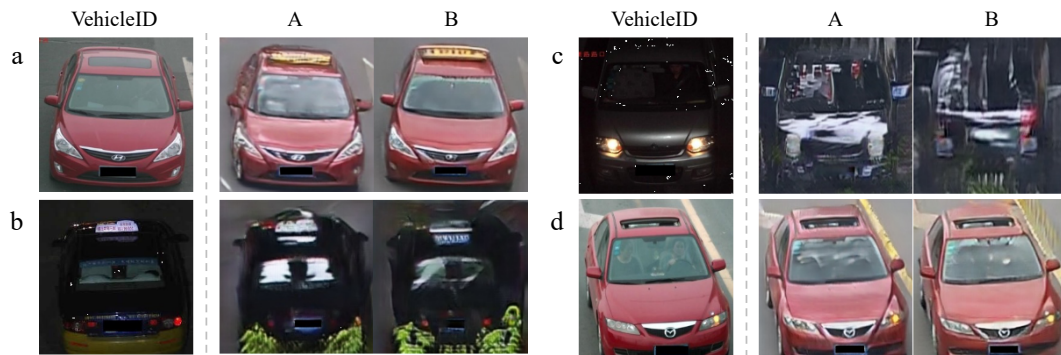


Fig. 7 I spy with my convoluted eyes: (a) a taxi, (b) a bush, (c) an abstract self portrait, (d) a fence.



Fig. 8 VeRi-776: samples of (a) taxis and yellow fence, and (b) vehicles captured behind bushes and non-blackened out license plates.

notice, both VCGAN-A and VCGAN-B transformed the rear bumper into a bush. This transformation is due to the amount of certain camera views in VeRi-776 where the vehicles are hidden behind bushes, as shown in Fig. 8b, where we notice the bushes on the bottom of the images. On the other hand, when these yellow patches are absent, CycleGAN struggles and produces an output that does not make much sense (Fig. 7c). Finally, yellow street markings are transformed by CycleGAN into fences. Referring to Fig. 8a, we can deduce that this behavior is due to some camera views that are positioned such that these fences are visible. Interestingly enough, CycleGAN added something we did not expect: a license plate, see Fig. 4 (right (c)), Fig. 5, and Fig. 6. This was unexpected, but could mean that there is a large amount of data in VeRi-776 where the license plates have not been blackened out completely. Upon checking, we can verify this observation. Fig. 8b clearly shows that some vehicles in VeRi-776 do not have their license plates blackened out.

CycleGAN transformed yellow-green patches into bushes, yellow markings into fences, added license plates, and transformed red vehicles into taxis. When the images are taken at night, CycleGAN produces outputs that do not make much sense. Rather than being useful for v-reID, these observations show that the source dataset has its drawbacks, notably a lack of images taken during the night, occluded vehicles by bushes, and most importantly, license plates that were not removed or hidden. This also means that we could go the opposite way, and generate vehicle images without license plates or images in a night setting.

Multiple potential solutions can be explored to address these unexpected outcomes that could negatively impact the quality of the generated dataset. One potential solution is to ensure

that the datasets are more diverse; however, this approach goes back to data collection, which can be more challenging and time-consuming. Another potential solution is to apply preprocessing steps to remove or obscure license plates before training, which would prevent the model from adding them during transformations. Finally, domain adaptation techniques, where transformations are more closely aligned with the target domain could be employed to better control the outputs generated by CycleGAN.

Qualitative and quantitative comparison

In this subsection, we compare VCGAN-A and VCGAN-B quantitatively and qualitatively.

Qualitative comparison

There is a trade-off between smoothness and details. Images generated by VCGAN-A fits its purpose, however VCGAN-B does a better job overall. Based on all the previous figures, we deduce that VCGAN-B does a cleaner job in the sense that the outputs turn out to be smoother and resemble less to synthetic images than of VCGAN-A. Furthermore, referring to Fig. 9,

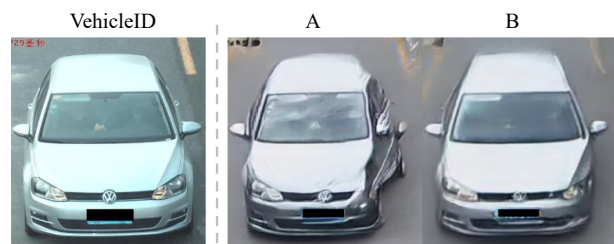


Fig. 9 Comparing the two experiments: the outcomes of B are qualitatively more reliable than A.

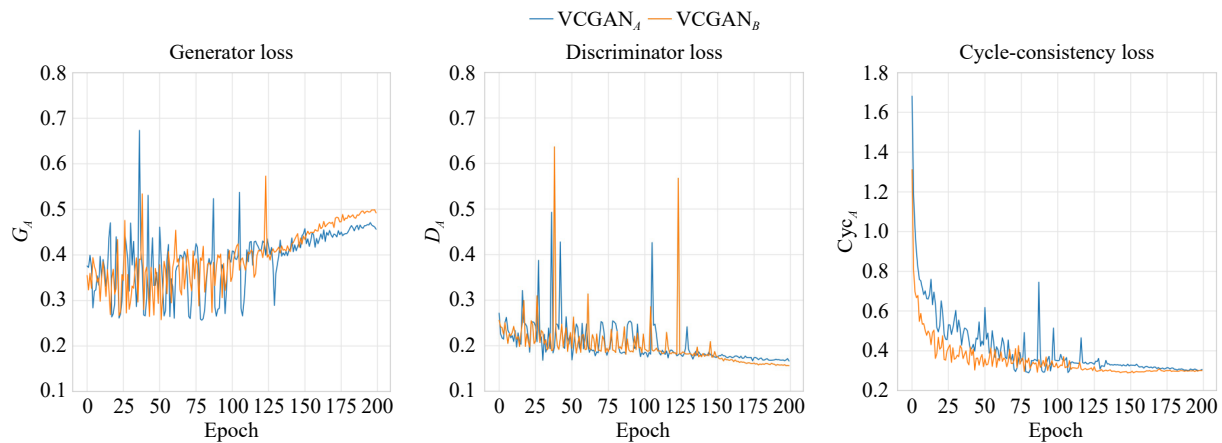


Fig. 10 Generator loss, discriminator loss and forward cycle consistency loss curves per epoch for VCGAN-A and VCGAN-B.

VCGAN-B can produce 'whole' vehicles, while VCGAN-A outputs a deformed vehicle.

Quantitative comparison

To compare both experiments numerically, we plot the generator, discriminator, and forward cycle consistency losses (Fig. 10). While the discriminator and cycle-consistency losses are decreasing by each epoch, the generator loss increases from epoch 120 on. This indicates that the training of the generator and discriminator is imbalanced, with the discriminator becoming too strong. Compared to VCGAN-A, the discriminator loss and forward cycle consistency loss are lower for VCGAN-B. However, the generator loss shows the opposite behavior.

Discussion and conclusions

This paper is the first-of-its-kind to analyze the ethical aspects and limitations of v-reID models as well as widely used vehicle re-identification datasets. It is found that there is a lack of explainable v-reID works and that existing datasets are not privacy compliant or contain biases that can confuse v-reID models. Due to the black-box properties of deep neural networks, researchers do not understand the internal procedures of how v-reID models perform their reasoning. Hence, we need to ensure the following: training data should not contain: (1) any face of drivers and passengers, to respect people's privacy and to ensure that no indirect facial recognition is performed, and (2) any artifact in images that do not originate from the source scene, to avoid confusion of the models. It is found that VehicleID did not respect these characteristics since some samples included faces that were visible and recognizable, and others carried red writing.

To this end, CycleGAN was proposed to transfer images from VehicleID to VeRi-776. With the proposed method, we managed to generate samples of vehicle images where faces were blurred out and the red timestamps were removed. The experiments produced images that did not resemble vehicles anymore or added objects that were not present in the source image (e.g., bushes, fences, or license plates). These limitations, however, point out further data bias of VehicleID and VeRi-776 datasets, such as the lack of variety in images or non-hidden license plates that should be further investigated.

Using GAN-based blurring offers advantages over simple resolution reduction by preserving image quality and targeting

specific areas. Unlike resolution reduction, which can degrade the entire image, GAN-based methods maintain the overall details of the image. Additionally, GANs can precisely target the face area, leaving other parts of the image, such as the vehicle itself, untouched, which is not possible with simple resolution reduction.

With the introduction of generative models such as Stable Diffusion1 or Midjourney AI2, vehicle re-identification data generation could also be improved, producing better results than the proposed method. However, our proposed approach is more manageable and smaller in terms of parameters and resources that are needed. Furthermore, it is also more affordable compared to existing generative models as no external server is needed for training. It should be noted that both datasets are built from images collected in China. This makes these datasets a bit biased towards China in terms of people, car models, backgrounds, and direction of traffic. Additionally, this work focuses mainly on two datasets: VehicleID and VeRi-776. While we have not specifically examined other v-ReID datasets for similar problems, these limitations likely exist in them as well. High-resolution images could potentially expose faces, raising privacy issues if not carefully managed.

In future work, additional datasets will be inspected to identify any further limitations or challenges they may present, and more datasets from different countries and areas will be used to further validate the proposed method regarding efficiency and effectiveness. Furthermore, it is necessary to develop proper regulations and laws to ensure the legal use of v-ReID technologies^[45].

Author contributions

The authors confirm contribution to the paper as follows: study conception and design: Qian Y; data collection: Qian Y; analysis and interpretation of results: Qian Y; draft manuscript preparation: Qian Y, Barthélemy J, Du B, Shen J. All authors reviewed the results and approved the final version of the manuscript.

Data availability

The data used in this study are from the following resources available in the public domain: VeRi-776 (<https://vehiclereid>).

github.io/VeRi/) and VehicleID (www.pkuml.org/resources/pku-vehicleid.html).

Conflict of interest

The authors declare that they have no conflict of interest. Bo Du and Jun Shen are the Editorial Board members of *Digital Transportation and Safety* who were blinded from reviewing or making decisions on the manuscript. The article was subject to the journal's standard procedures, with peer-review handled independently of these Editorial Board members and their research groups.

Dates

Received 19 August 2024; Revised 9 September 2024; Accepted 14 September 2024; Published online 27 December 2024

References

- Liu X, Liu W, Mei T, Ma H. 2016. A deep learning-based approach to progressive vehicle re-identification for urban surveillance. *Computer Vision – ECCV 2016: 14th European Conference Part II, Amsterdam, The Netherlands, October 11–14, 2016*. Amsterdam, The Netherlands: Springer. pp. 869–84. doi: [10.1007/978-3-319-46475-6_53](https://doi.org/10.1007/978-3-319-46475-6_53)
- Liu H, Tian Y, Wang Y, Pang L, Huang T. 2016. Deep relative distance learning: Tell the difference between similar vehicles. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016*. USA: IEEE. pp. 2167–75. doi: [10.1109/CVPR.2016.238](https://doi.org/10.1109/CVPR.2016.238)
- Kazim E, Koshiyama AS. 2021. A high-level overview of AI ethics. *Patterns* 2(9):100314
- Department of Industry, Science and Resources. 2019. *Australia's AI Ethics Principles*. www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles (Accessed on March 4, 2023)
- Yang HF, Cai J, Liu C, Ke R, Wang Y. 2023. Cooperative multi-camera vehicle tracking and traffic surveillance with edge artificial intelligence and representation learning. *Transportation Research Part C: Emerging Technologies* 148:103982
- Dietmeier J, Antony J, McGuinness K, O'Connor NE. 2021. How important are faces for person re-identification? *The 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021*. USA: IEEE. pp. 6912–19. doi: [10.1109/ICPR48806.2021.9412340](https://doi.org/10.1109/ICPR48806.2021.9412340)
- Ahmad S, Scarpellini G, Morerio P, Bue AD. 2022. Event-driven re-id: a new benchmark and method towards privacy-preserving person re-identification. *2022 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW), Waikoloa, HI, USA, 4–8 January 2022*. USA: IEEE. pp. 459–68. doi: [10.1109/WACVW54805.2022.00052](https://doi.org/10.1109/WACVW54805.2022.00052)
- RichardWebster B, Hu B, Fieldhouse K, Hoogs A. 2022. Doppelgänger saliency: towards more ethical person re-identification. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), New Orleans, LA, USA, 19–20 June 2022*. USA: IEEE. pp. 2847–57. doi: [10.1109/CVPRW56347.2022.00322](https://doi.org/10.1109/CVPRW56347.2022.00322)
- Rudin C. 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence* 1:206–15
- Wang Z, Tang L, Liu X, Yao Z, Yi S, et al. 2017. Orientation invariant feature embedding and spatial temporal regularization for vehicle re-identification. *2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017*. Venice, Italy: IEEE. pp. 379–87. doi: [10.1109/ICCV.2017.49](https://doi.org/10.1109/ICCV.2017.49)
- Zhang Y, Liu D, Zha ZJ. 2017. Improving triplet-wise training of convolutional neural network for vehicle re-identification. *2017 IEEE International Conference on Multimedia and Expo (ICME), Hong Kong, China, 10–14 July 2017*. USA: IEEE. pp. 1386–91. doi: [10.1109/ICME.2017.8019491](https://doi.org/10.1109/ICME.2017.8019491)
- Hou J, Zeng H, Zhu J, Chen J, et al. 2019. Deep quadruplet appearance learning for vehicle re-identification. *IEEE Transactions on Vehicular Technology* 68(9):8512–22
- Qian J, Pan M, Tong W, Law R, Wu EQ. 2023. URRNet: A unified relational reasoning network for vehicle re-identification. *IEEE Transactions on Vehicular Technology* 72(9):11156–68
- He Z, Zhao H, Wang J, Feng W. 2023. Multi-level progressive learning for unsupervised vehicle re-identification. *IEEE Transactions on Vehicular Technology* 72(4):4357–71
- Zhu W, Peng B. 2022. Manifold-based aggregation clustering for unsupervised vehicle re-identification. *Knowledge-Based Systems* 235:107624
- Zhang H, Kuang Z, Cheng L, Liu Y, Ding X, et al. 2024. AIVR-Net: attribute-based invariant visual representation learning for vehicle re-identification. *Knowledge-Based Systems* 289:111455
- Wei XS, Zhang CL, Liu L, Shen C, Wu J. 2019. Coarse-to-fine: a RNN-based hierarchical attention model for vehicle re-identification. *Computer Vision – ACCV 2018: 14th Asian Conference on Computer Vision Part II, Perth, Australia, 2–6 December 2018*. Perth, Australia: Springer. pp. 575–91. doi: [10.1007/978-3-030-20890-5_37](https://doi.org/10.1007/978-3-030-20890-5_37)
- He ST, Luo H, Wang PC, Wang F, Li H, et al. 2021. TransReID: Transformer-based object re-identification. *2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021*. USA: IEEE. pp. 14993–5022. doi: [10.1109/ICCV48922.2021.01474](https://doi.org/10.1109/ICCV48922.2021.01474)
- Qian Y, Barthelemy J, Iqbal U, Perez P. 2022. V²ReID: Vision-outlooker-based vehicle re-identification. *Sensors* 22(22):8651
- Li H, Li C, Zheng A, Tang J, Luo B. 2022. MsKAT: Multi-scale knowledge-aware transformer for vehicle re-identification. *IEEE Transactions on Intelligent Transportation Systems* 23(10):19557–68
- Lu Z, Lin R, Hu H. 2023. MART: Mask-aware reasoning transformer for vehicle re-identification. *IEEE Transactions on Intelligent Transportation Systems* 24(2):1994–2009
- Huang F, Lv X, Zhang L. 2023. Coarse-to-fine sparse self-attention for vehicle re-identification. *Knowledge-Based Systems* 270:110526
- Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, et al. 2019. A survey of methods for explaining black box models. *ACM Computing Surveys* 51(5):1–42
- Ye M, Shen J, Lin G, Xiang T, Shao L, et al. 2022. Deep learning for person re-identification: a survey and outlook. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44(6):2872–93
- Chen X, Liu X, Liu W, Zhang XP, Zhang Y, et al. 2021. Explainable person re-identification with attribute-guided metric distillation. *2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021*. USA: IEEE. pp. 11813–22. doi: [10.1109/ICCV48922.2021.01160](https://doi.org/10.1109/ICCV48922.2021.01160)
- Szegedy C, Zaremba W, Sutskever I, Bruna J, Erhan D, et al. 2013. Intriguing properties of neural networks. *arXiv preprint*
- Yang L, Luo P, Loy CC, Tang X. 2015. A large-scale car dataset for fine-grained categorization and verification. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015*. USA: IEEE. pp. 3973–81. doi: [10.1109/CVPR.2015.7299023](https://doi.org/10.1109/CVPR.2015.7299023)
- Everingham M, Van Gool L, Williams CKI, Winn J, Zisserman A. 2010. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision* 88(2):303–38
- Yan K, Tian Y, Wang Y, Zeng W, Huang T. 2017. Exploiting multi-grain ranking constraints for precisely searching visually-similar vehicles. *2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017*. USA: IEEE. pp. 562–70. doi: [10.1109/ICCV.2017.68](https://doi.org/10.1109/ICCV.2017.68)

Privacy-compliant vehicle re-identification

30. Guo H, Zhao C, Liu Z, Wang J, Lu H. 2018. Learning coarse-to-fine structured feature embedding for vehicle re-identification. *Proceedings of the AAAI Conference on Artificial Intelligence* 32(1):6853–60
31. Tang Z, Naphade M, Liu MY, Yang X, Birchfield S, et. al. 2019. CityFlow: A city-scale benchmark for multi-target multi-camera vehicle tracking and re-identification. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019*. USA: IEEE. pp. 8797–806. doi: [10.1109/CVPR.2019.00900](https://doi.org/10.1109/CVPR.2019.00900)
32. Lou Y, Bai Y, Liu J, Wang S, Duan L. 2019. VERI-Wild: A large dataset and a new method for vehicle re-identification in the wild. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019*. USA: IEEE. pp. 3235–43. doi: [10.1109/CVPR.2019.00335](https://doi.org/10.1109/CVPR.2019.00335)
33. Isola P, Zhu JY, Zhou T, Efros AA. 2017. Image-to-image translation with conditional adversarial networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017*. USA: IEEE. pp. 1125–34. <https://doi.org/10.1109/CVPR.2017.632>
34. Wang M, Deng W. 2018. Deep visual domain adaptation: A survey. *Neurocomputing* 312:135–53
35. Zhu JY, Park T, Isola P, Efros AA. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. *2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017*. USA: IEEE. pp. 2242–51. doi: [10.1109/ICCV.2017.244](https://doi.org/10.1109/ICCV.2017.244)
36. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, et. al. 2020. Generative adversarial networks. *Communications of the ACM* 63(11):139–144
37. Zhou Y, Shao L. 2017. Cross-view gan based vehicle generation for re-identification. *Proc. British Machine Vision Conference, 2017*. London UK. pp. 1–12.
38. Wang Q, Min W, Han Q, Liu Q, Zha C, et al. 2022. Inter-domain adaptation label for data augmentation in vehicle re-identification. *IEEE Transactions on Multimedia* 24:1031–41
39. Zhou Z, Li Y, Li J, Yu K, Kou G, et al. 2023. GAN-Siamese network for cross-domain vehicle re-identification in intelligent transport systems. *IEEE Transactions on Network Science and Engineering* 10(5):2779–90
40. Luo H, Chen W, Xu X, Gu J, Zhang Y, et al. 2021. An empirical study of vehicle re-identification on the AI city challenge. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Nashville, TN, USA, 19–25 June 2021*. USA: IEEE. pp. 4095–102. doi: [10.1109/CVPRW53098.2021.00462](https://doi.org/10.1109/CVPRW53098.2021.00462)
41. Deng W, Zheng L, Ye Q, Kang G, Yang Y. 2018. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018*. USA: IEEE. pp. 994–1003. doi: [10.1109/CVPR.2018.00110](https://doi.org/10.1109/CVPR.2018.00110)
42. Zhou T, Krähenbühl P, Aubry M, Huang Q, Efros AA. 2016. Learning dense correspondence via 3D-guided cycle consistency. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016*. USA: IEEE. pp. 117–26. doi: [10.1109/CVPR.2016.20](https://doi.org/10.1109/CVPR.2016.20)
43. Brislin RW. 1970. Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology* 1(3):185–216
44. Kingma DP, Ba J. 2014. Adam: A method for stochastic optimization. *arXiv Preprint*
45. Qian Y, Barthélemy J, Du B, Shen J. 2024. Paying attention to vehicles: a systematic review on transformer-based vehicle re-identification. *ACM Transactions on Multimedia Computing, Communications and Applications*



Copyright: © 2024 by the author(s). Published by Maximum Academic Press, Fayetteville, GA. This article is an open access article distributed under Creative Commons Attribution License (CC BY 4.0), visit <https://creativecommons.org/licenses/by/4.0/>.