

StRDAN: Synthetic-to-Real Domain Adaptation Network for Vehicle Re-identification

Sangrok Lee¹, Eunsoo Park¹, Hongsuk Yi², and Sang Hun Lee^{*3}

¹MODULABS

srl@modulabs.ai, es.park@modulabs.co.kr

²Korea Institute of Science and Technology Information

hsyi@kisti.re.kr

³Kookmin University

shlee@kookmin.ac.kr

Abstract

Vehicle re-identification aims to obtain the same vehicles from vehicle images. This is challenging but essential for analyzing and predicting traffic flow in the city. Although deep learning methods have achieved enormous progress for this task, their large data requirement is a critical shortcoming. Therefore, we propose a synthetic-to-real domain adaptation network (StRDAN) framework, which can be trained with inexpensive large-scale synthetic and real data to improve performance. The StRDAN training method combines domain adaptation and semi-supervised learning methods and their associated losses. StRDAN offers significant improvement over the baseline model, which can only be trained using real data, for VeRi and CityFlow-ReID datasets, achieving 3.1% and 12.9% improved mean average precision, respectively.

1. Introduction

Vehicle re-identification (Re-ID) aims to identify the same vehicles that are captured by various cameras. It is an essential technology for analyzing and predicting traffic flow in smart cities and uses visual appearance based Re-ID methods in general. However, vehicle Re-ID is challenging for two reasons.

- Different lighting and complex environments create difficulties with appearance based vehicle Re-ID, and large apparent variations can be generated using different cameras.

*Corresponding author

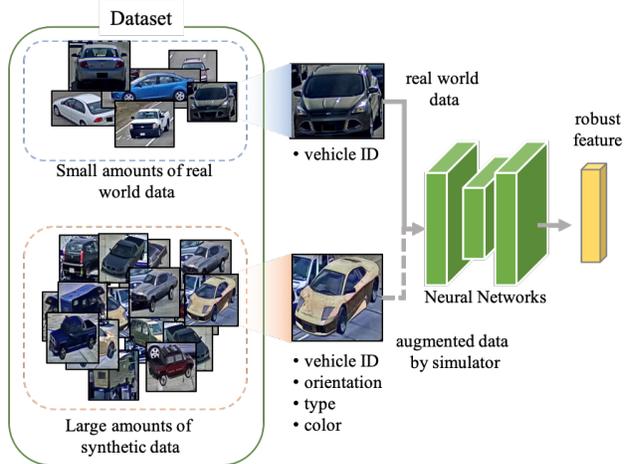


Figure 1. Proposed synthetic-to-real domain adaptation method to improve vehicle re-identification performance. It can be difficult to obtain meaningful labels for real data, but it is relatively simple for synthetic data.

- Different vehicles can be visually very similar when they are in the same type category.

Deep learning methods [23, 10, 17] are commonly employed to tackle this complex vehicle Re-ID task with significant progress. These models extract features using deep learning networks and distinguish vehicles by comparing feature distances. However, they require large datasets for training and improved performance, which rapidly becomes a drawback. Many studies [30] have confirmed that more training data provides better model performance. Therefore, data from real environments require considerable an-

notation workload. On the other hand, domain adaptation approaches employ inexpensive synthetic data to replace real data.

This paper explores how to improve model performance using inexpensive synthetic data (see Fig. 1). We adopted an adversarial domain adaptation approach [4] where an artificial neural network (ANN) learns the best discriminating features for classification using real data, while simultaneously learning indistinguishable features between real and synthetic data [1] [5]. To implement this concept, we introduce a domain discrimination layer and associated cross-entropy loss to train the whole network to be indiscriminative for both domains. We also adopted a semi-supervised learning method to better exploit specific synthetic data labels, such as color, type, and orientation. Since these labels only exist for synthetic data, a semi-supervised learning approach that can handle unlabeled data is appropriate to improve performance. In training, classification losses for the exclusive labels are selectively applied depending on the data domain [30]. The proposed model trained on real and synthetic data from the AI City Challenge using domain adaptation and semi-supervised learning approaches achieved 12.9% improvement over the baseline model, which was trained with only real data.

This work proposes a novel synthetic-to-real domain adaptation network StRDAN framework, with major contributions as follows.

- StRDAN can be successfully trained with inexpensive large-scale synthetic as well as real data to improve performance.
- We propose a new training approach for StRDAN, combining domain adaptation and semi-supervised learning methods and corresponding losses.
- StRDAN shows significant improvement over the baseline model for two significant data sets: VeRi [15] and CityFlow-ReID [25].

2. Related Work

This section reviews relevant prior studies regarding vehicle Re-ID and domain adaptation methods with synthetic data.

Vehicle Re-ID: Vehicle Re-ID methods generally incorporate contrastive loss and spatio-temporal features. Previous studies [14, 15, 16] have proposed several contrastive loss based methods, such as siamese networks, triplet loss, and metric learning. Liu *et al.* [15] introduced the VeRi dataset, being the first large scale vehicle Re-ID benchmark. Spatio-temporal features are critical for performance improvement and have helped vehicle Re-ID studies to achieve huge progress. Tan *et al.* [23] used spatial-temporal features for multi-camera vehicle tracking and vehicle Re-ID to win

the AI City Challenge in 2019 [18]. Shen *et al.* [22] proposed a two-stage framework to match visual appearance based on an long-short term memory (LSTM) based path inference mechanism.

Domain Adaptation with Synthetic Data: To overcome the lack of data, Zhou *et al.* [34, 35] proposed a method to improve Re-ID performance by augmenting various viewpoint vehicle images with generative adversarial networks (GANs). Performance significantly reduces when deploying trained models onto new datasets due to differences among the datasets, commonly called domain bias. Peng *et al.* [19] proposed a domain adaptation framework to address this problem, incorporating an image-to-image translation network and an attention based feature learning network. The VehicleX [29] simulator also leverages synthetic data and domain randomization to overcome the reality gap [26, 27]. Although Liu *et al.* [12] proposed a domain adaptation method, they only considered real-to-real domain adaptation. The recently proposed PAMTRI approach [24] uses synthetic data to improve model performance and has similar architecture to the proposed StRDAN framework. However, PAMTRI requires considerable effort to obtain vehicle pose and labels for real data, whereas StRDAN uses domain adaptation to utilize synthetic data and semi-supervised learning does not require additional annotation workload. Thus, StRDAN is somewhat simpler and easier to train.

3. Proposed Synthetic-to-Real Domain Adaptation Network

3.1. Datasets

We developed an ANN using real and synthetic vehicle datasets provided for Track 2 of the 2020 AI City Challenge. The real dataset was the CityFlow-reID dataset, a subset of CityFlow made available for the Track 2 challenge comprising 56,277 images for 666 unique vehicles collected from 40 cameras, with 36,935 images from 333 vehicle identities for training, and 18,290 images from the other 333 identities for testing. The remaining 1052 images in the test set were provided as query data.

The synthetic vehicle dataset comprised 192,150 images from 1,362 distinct vehicles created using the VehicleX [29] synthetic dataset generator, forming an augmented training set. The synthetic dataset includes vehicle ID, color, type, and object orientation; whereas the real dataset includes only vehicle ID. Vehicles were distinguished into 12 colors and 11 types, and orientation was represented by rotation angle $[0, 360)$ on the horizontal plane.

We trained and evaluated the proposed StRDAN model using the VeRi real dataset [8] and City Challenge synthetic data to examine validity and robustness for the approach. The Veri dataset contains over 50,000 images for 776 vehi-

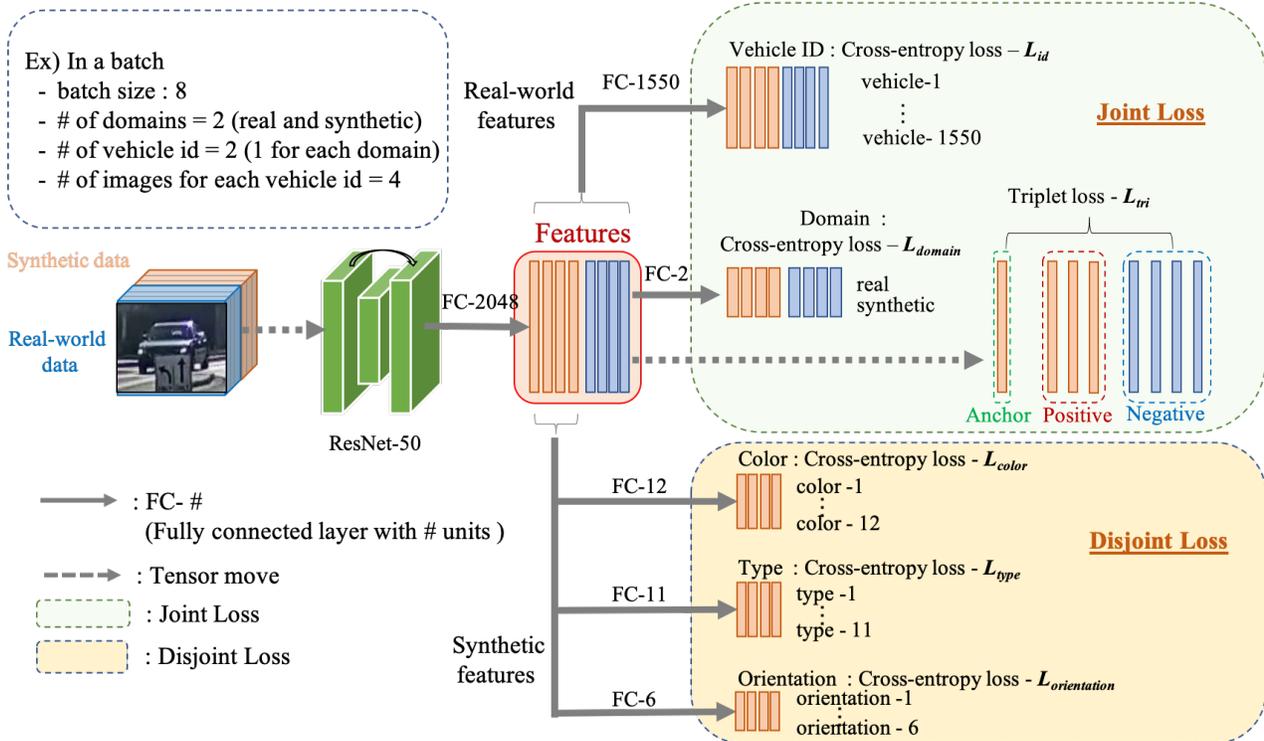


Figure 2. Proposed synthetic-to-real domain adaptation network architecture comprising a ResNet-50 backbone for feature extraction and five fully connected softmax layers for classification, trained using joint and disjoint losses between synthetic and real data.

cles captured by 20 cameras, with the training set containing 37,781 images for 576 vehicles, and the testing set containing 11,579 images for 200 vehicles. The VeRi dataset includes labels for vehicle color and type.

3.2. Overall Architecture

Figure 2 shows the proposed overall StRDAN architecture. The model comprises a backbone network for feature extraction and multiple fully connected (FC) softmax layers for classification. Input images are batch sampled in equal numbers from the real and synthetic datasets. For a mini-batch, n different vehicle identities are chosen from the real and synthetic datasets, respectively, then m samples are randomly selected from these chosen images. Therefore, each batch contains $2 \times n \times m$ images.

The backbone network extracts a highly abstracted feature vector ($dim = 2048$) from the input image. In principle, any convolution neural network (CNN) designed for image classification can be used as the backbone network, and a variety of CNNs have been employed in previous studies, including VGG-CNN-M1024 [3], MobileNet [9], and ResNet [7], as vehicle Re-ID model backbone. We selected ResNet-50 as the backbone network for StRDAN. Feature maps extracted by the backbone network are flattened and fed into various FC softmax layers to classify ve-

hicle id, real or synthetic, color, type, and orientation. Outputs are then fed into five cross-entropy loss functions and one triplet loss function. StRDAN was end-to-end trained by updating the network parameters to reduce total loss, combining cross-entropy and triplet losses.

3.3. Key Features

Adversarial Domain Adaptation. An annotated dataset is essential for deep neural network supervised learning. However, collecting and manually annotating large datasets is time consuming and expensive. Therefore, the VehicleX approach was introduced in the AI City Challenge to generate automatically labeled data using a graphic simulator and hence overcome the dearth of real data. However, synthetic data has similar but different distributions than real data. Therefore, it is necessary to train the ANN to predict the classification, regardless of the input domain.

We adopted the adversarial domain adaptation approach, where the ANN learns features that are most discriminative for classification on the real domain and simultaneously as indistinguishable as possible between the real and synthetic domains [1] [5]. To implement this approach, we introduced a domain discrimination layer and its associated cross-entropy loss train the network to be indiscriminate to the domains. We also introduced a vehicle-id classification

layer and its associated cross-entropy loss along with triplet loss to train the network to better discriminate vehicle identities and shape signatures.

Semi-supervised Learning. In contrast with the real data, synthetic data includes vehicle type, color, and orientation labels, and we use these labels under multitask learning to improve generalization performance for all tasks [31]. Many semi-supervised learning approaches improve learning accuracy by combining a small amount of labeled data with a large amount of unlabeled data during training. Zhai *et al.*'s work [30] created artificial labels for unlabeled and labeled data and utilize them in training, and this approach inspired us to use joint and disjoint labels between real and synthetic data to improving performance. Joint labels attached to real and synthetic data are vehicle ID and domain (real or synthetic), whereas disjoint labels were attached to only synthetic data and includes vehicle type, color, and orientation. Losses were classified as joint or disjoint losses, associated with joint and disjoint labels, respectively, as shown in Fig. 2. Triplet loss is classified as a joint loss because the vehicle id contributes to distinguishing batch images into anchor, positive, or negative images.

The semi-supervised learning approach considered here has learning objective

$$\min_{\theta} \mathcal{L}_{joint}(\theta) + w\mathcal{L}_{disjoint}(\theta), \quad (1)$$

where \mathcal{L}_{joint} is the joint loss defined in both real and synthetic domains and $\mathcal{L}_{disjoint}$ is the disjoint loss defined in the synthetic domain. θ is parameters of the network. In the next section, we will describe the losses in more detail.

4. Loss Function

4.1. Joint Losses

Vehicle ID. Cross-entropy following the softmax function is the most commonly employed loss for image classification, and can be represented for vehicle ID classifier, Lid, as follows

$$L_{id} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C y_{ij} \log(\hat{y}_{ij}) \quad (2)$$

where N denotes the number of images in a mini-batch, C is the number of classes, y_{ij} is the j^{th} element of an one hot encoded vector for the ground-truth of the i^{th} sample in a mini-batch, and \hat{y}_{ij} is the j^{th} element of the softmax FC layer for the i^{th} image.

Domain. We adopted the adversarial domain adaptation approach, with real and synthetic domains. A softmax FC layer was added to the backbone network for domain discrimination, with loss function to train the network to be

indiscriminate to the domains,

$$L_{domain} = \frac{1}{N} \sum_{i=1}^N y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i). \quad (3)$$

Domain discrimination loss is defined as the negative of binary cross-entropy loss. Since cross-entropy loss trains the network to discriminative between the domains, the negative loss trains the model to be less discriminating. Thus, if a vehicle captured by a camera is drawn by a graphic simulator in the same orientation, features extracted from a synthetic image would be similar to that from a real image since domain dependent features are suppressed. The negative cross-entropy loss function was implemented by a gradient reversal layer [4].

Triplet Loss. In a mini-batch that contains P identities and Q images for each identity, each image (anchor) has $Q - 1$ images of the same identity (positives) and $(P - 1) \times Q$ images of different identities (negatives). Triplet loss pulls the positive pair (a, p) together while pushing the negative pair (a, n) away by some margin. Thus, triplet loss trains the network to minimize the distance between features from the same image classes and simultaneously maximizes distance between features from different image classes. Triplet loss is defined as [8]

$$L_{tri} = \sum_{i=1}^P \sum_{a=1}^Q \left[m + \max_{\substack{p=1 \dots Q \\ p \neq a}} D(v_{a,i}, v_{p,i}) - \min_{\substack{j=1 \dots P \\ n=1 \dots Q \\ j \neq i}} D(v_{a,i}, v_{n,j}) \right]_+ \quad (4)$$

where $v_{a,i}$ is the prediction vector for the a^{th} image of the i^{th} identity group, and m is the margin to control the difference between positive and negative pair distances and helps cluster the distribution more densely.

4.2. Disjoint Losses

Color, Type, and Orientation. Softmax cross-entropy loss was applied for these three targets. Orientation is continuous and numerical, whereas color and type are categorical and nominal. Therefore, it seems reasonable to use regression to predict orientation, but orientation is a difficult problem for regression due to the wide range of the regression target. Indeed, experiments showed that optimization would not converge for regression. Therefore, we predicted orientation as direct classification into n discrete bins, with softmax cross-entropy loss [33] or [6], dividing the 360 degrees of orientation space into six bins of 60 degrees. Cross-entropy losses for color, type, and orientation were only applied to synthetic images and set to zero for real images.

The loss function can be expressed as

$$L_x = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C \delta_i y_{ij} \log(\hat{y}_{ij}), \quad \delta_i \in \{1, 0\}, \quad (5)$$

where x is one of color, type, and orientation, and δ_i is a mask value that is set to 1 if the i^{th} data in a mini-batch has x , and 0 otherwise.

5. Experiments

5.1. Evaluation Metric

We used rank-K mean Average Precision (mAP), the official AI City Challenge evaluation metric, to evaluate model performances. mAP measures the mean average precision for each query considering only the top K matches, where we chose $K = 100$. Average precision was computed for each query image from the area under the precision-recall curve, and then the mean of the average precision over all queries was computed.

5.2. Implementation

The chosen backbone network, ResNet-50, was initialized with weights pre-trained on ImageNet [21] to accelerate training. We trained the model end-to-end with an AMSGrad optimizer [20] for 60 epochs. Initial learning rate = 0.0003, reduced by 0.1 after 20 and 40 epochs. Weight decay factor for L2 regulation was set = 0.0005, and batch size = 64. For each mini-batch, two different vehicle-IDs were selected from each of the real and synthetic datasets, and four images with the same ID were sampled. Therefore, 16 different images with four different IDs from real and synthetic datasets were sampled. Input images were resized to 128 x 256 pixel, and we employed horizontal flip and random erasure augmentations. Postprocessing used the re-ranking algorithm [32], to order the distance matrix between features with Jaccard and original output distance.

5.3. Results and Discussion

We trained and evaluated our models using the CityFlow-reID and VeRi real datasets along with synthetic data generated by VehicleX, with selected disjoint losses as shown in Table 1 and Table 2.

Performance on AI City Dataset. The baseline (Case 1) model comprised the backbone network and vehicle ID classifier. The baseline was trained with the real dataset using vehicle-ID cross-entropy and triplet losses. Table 1 shows that the proposed domain adaptation and semi-supervised learning approaches significantly improved model performance compared with the baseline by at least 8.5% (Case 8) up to 12.9% (Case 4). The proposed model exhibited best performance for Case 4, where only

Table 1. Evaluation results of the StRDAN trained with CityFlow-reID and VehicleX datasets for the 2020 AI City Challenge, Track 2. The results are from the official evaluation leaderboard of the challenge.

Case	O	C	T	V	D	Dataset	mAP
1				✓		R	25.5
2	✓			✓	✓	R+S	No Conv.
3		✓		✓	✓	R+S	35.2
4			✓	✓	✓	R+S	38.4
5	✓	✓		✓	✓	R+S	34.1
6	✓		✓	✓	✓	R+S	37.5
7		✓	✓	✓	✓	R+S	35.3
8	✓	✓	✓	✓	✓	R+S	34.0

Table 2. Evaluation results of the StRDAN trained with VeRi and VehicleX datasets.

Case	O	C	T	V	D	Dataset	mAP
1				✓		R	73.0
2	✓			✓	✓	R+S	76.1
3		✓		✓	✓	R+S	74.2
4			✓	✓	✓	R+S	74.9
5	✓	✓		✓	✓	R+S	74.7
6	✓		✓	✓	✓	R+S	75.3
7		✓	✓	✓	✓	R+S	74.8
8	✓	✓	✓	✓	✓	R+S	74.6

Notes: O, C, T, V, D = orientation, color, type, vehicle ID, and domain, respectively.

R, S = real and synthetic data, respectively.

No Conv. = no convergence

mAP = mean average precision

Boxes are checked if target loss was included.

Best result is shown in bold.

vehicle type was considered, whereas Case 8 considered all three labels and exhibited the worst performance.

Performance on VeRi Dataset. Table 2 also shows that domain adaptation and semi-supervised learning approaches with synthetic dataset and additional losses helped improve performance by at least 1.2% (Case 3) up to 3.1% (Case 2). In contrast to the AI City dataset, the best performance was when considering only orientation (Case 2). However, the model could not converge with the AI City dataset. Veri data model performances were much superior than those for AI City data. Table 3 compares the proposed StRDAN approach with other methods. All models were trained using only the VeRi dataset, aside from PAMTRI and StRDAN (R+S). The proposed StRDAN (R+S) model outperformed all other considered methods.

Domain Adaptation and Semi-supervised Learning. The experimental results verified that domain adaptation

Table 3. Comparing Deep learning methods on the VeRi dataset.

Method	mAP
FACT[13]	18.7
ABLN[35]	24.9
OIFE[28]	48.0
PROVID[16]	48.5
PathLSTM[22]	58.3
GSTE[2]	59.5
VAMI[36]	61.3
BA[11]	66.9
BS[11]	67.6
PAMTRI[24]	71.9
StRDAN (R, baseline)	73.0
StRDAN (R+S, best)	76.1

and semi-supervised learning approaches help extract more important semantic features for vehicle Re-ID. However, further research is required regarding unexpected phenomena as follows.

- The best model was trained with only one loss of three disjoint losses.
- Performance degraded with including more disjoint losses.
- Best performance depended on the real dataset.

6. Conclusions

This paper proposes using domain adaptation and semi-supervised learning to fully utilize synthetic data. Experiment results confirmed that increasing training data via with domain adaptation improved performance. We also showed that using semi-supervised learning with labels only available for synthetic data helped the model extract more semantic features.

Future work will investigate the following issues.

- Synergy between disjoint losses and real-world data dependency on disjoint losses, as discussed above.
- Reality effects on synthetic data. Image data synthesized by VehicleX was far from realistic and hence easily distinguishable from real image data. More realistic synthetic data from more sophisticated simulation software could further improve performance.
- Orientation prediction. We converted orientation regression to six bin classification, but we have not optimized bin count. Since orientation is a key feature to identify vehicles captured at various camera angles, proper orientation representation will also help to improve performance.

References

- [1] Hana Ajakan, Pascal Germain, Hugo Larochelle, Francois Laviolette, and Mario Marchand. Domain-adversarial neural networks. *ArXiv*, abs/1412.4446, 2014.
- [2] Yan Bai, Yihang Lou, Feng Gao, Shiqi Wang, Yuwei Wu, and Lingyu Duan. Group-sensitive triplet embedding for vehicle reidentification. *IEEE Transactions on Multimedia*, 20:2385–2399, 2018.
- [3] Ken Chatfield, Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Return of the devil in the details: Delving deep into convolutional nets. *ArXiv*, abs/1405.3531, 2014.
- [4] Yaroslav Ganin and Victor S. Lempitsky. Unsupervised domain adaptation by backpropagation. *ArXiv*, abs/1409.7495, 2014.
- [5] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, Francois Laviolette, Mario Marchand, and Victor S. Lempitsky. Domain-adversarial training of neural networks. *ArXiv*, abs/1505.07818, 2016.
- [6] Spyros Gidaris, Praveer Singh, and Nikos Komodakis. Unsupervised representation learning by predicting image rotations. *ArXiv*, abs/1803.07728, 2018.
- [7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. *ArXiv*, abs/1603.05027, 2016.
- [8] Alexander Hermans, Lucas Beyer, and Bastian Leibe. In defense of the triplet loss for person re-identification. *ArXiv*, abs/1703.07737, 2017.
- [9] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *ArXiv*, abs/1704.04861, 2017.
- [10] Tsung-Wei Huang, Jiarui Cai, Hao Yang, Hung-Min Hsu, and Jenq-Neng Hwang. Multi-view vehicle re-identification using temporal attention model and metadata re-ranking. In *CVPR Workshops*, 2019.
- [11] Ratnesh Kumar, Edwin Weill, Farzin Aghdasi, and P. Sriram. Vehicle re-identification: an efficient baseline using triplet embedding. *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–9, 2019.
- [12] Chih-Ting Liu, Man-Yu Lee, Chih-Wei Wu, Bo-Ying Chen, Tsai-Shien Chen, Yao-Ting Hsu, Shao-Yi Chien, and NTU IoX Center. Supervised joint domain learning for vehicle re-identification. In *Proc. CVPR Workshops*, pages 45–52, 2019.
- [13] Hongye Liu, Yonghong Tian, Yaowei Wang, Lu Pang, and Tiejun Huang. Deep relative distance learning: Tell the difference between similar vehicles. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2167–2175, 2016.
- [14] Xinchun Liu, Wu Liu, Huadong Ma, and Huiyuan Fu. Large-scale vehicle re-identification in urban surveillance videos. *2016 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6, 2016.
- [15] Xinchun Liu, Wu Liu, Tao Mei, and Huadong Ma. A deep learning-based approach to progressive vehicle re-

- identification for urban surveillance. In *European conference on computer vision*, pages 869–884. Springer, 2016.
- [16] Xinchun Liu, Wu Liu, Tao Mei, and Huadong Ma. Provid: Progressive and multimodal vehicle reidentification for large-scale urban surveillance. *IEEE Transactions on Multimedia*, 20(3):645–658, 2018.
- [17] Kai Lv, Heming Du, Yunzhong Hou, Weijian Deng, Hao Sheng, Jianbin Jiao, and Liang Zheng. Vehicle re-identification with location and time stamps. In *CVPR Workshops*, 2019.
- [18] Milind Naphade, Zheng Tang, Ming-Ching Chang, David C Anastasiu, Anuj Sharma, Rama Chellappa, Shuo Wang, Pranamesh Chakraborty, Tingting Huang, Jenq-Neng Hwang, et al. The 2019 ai city challenge. In *CVPR Workshops*, 2019.
- [19] Jinjia Peng, Huibing Wang, Tongtong Zhao, and Xianping Fu. Cross domain knowledge transfer for unsupervised vehicle re-identification. In *2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pages 453–458. IEEE, 2019.
- [20] Sashank J Reddi, Satyen Kale, and Sanjiv Kumar. On the convergence of adam and beyond. *arXiv preprint arXiv:1904.09237*, 2019.
- [21] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael S. Bernstein, Alexander C. Berg, and Fei-Fei Li. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115:211–252, 2015.
- [22] Yantao Shen, Tong Xiao, Hongsheng Li, Shuai Yi, and Xiaogang Wang. Learning deep neural networks for vehicle re-id with visual-spatio-temporal path proposals. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1900–1909, 2017.
- [23] Xiao Tan, Zhigang Wang, Minyue Jiang, Xipeng Yang, Jian Wang, Yuan Gao, Xiangbo Su, Xiaoqing Ye, Yuchen Yuan, Dongliang He, Shilei Wen, and Errui Ding. Multi-camera vehicle tracking and re-identification based on visual and spatial-temporal features. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 275–284, 2019.
- [24] Zheng Tang, Milind Naphade, Stan Birchfield, Jonathan Tremblay, William G. Hodge, Ratnesh Kumar, Shuo Wang, and Xiaodong Yang. Pamtri: Pose-aware multi-task learning for vehicle re-identification using highly randomized synthetic data. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 211–220, 2019.
- [25] Zheng Tang, Milind Naphade, Ming-Yu Liu, Xiaodong Yang, Stanley T. Birchfield, Shuo Wang, Ratnesh Kumar, David C. Anastasiu, and Jenq-Neng Hwang. 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)*, pages 8789–8798, 2019.
- [26] Josh Tobin, Rachel Fong, Alex Ray, Jonas Schneider, Wojciech Zaremba, and Pieter Abbeel. Domain randomization for transferring deep neural networks from simulation to the real world. In *2017 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 23–30. IEEE, 2017.
- [27] Jonathan Tremblay, Aayush Prakash, David Acuna, Mark Brophy, Varun Jampani, Cem Anil, Thang To, Eric Cameracci, Shaad Boochoon, and Stan Birchfield. Training deep networks with synthetic data: Bridging the reality gap by domain randomization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 969–977, 2018.
- [28] Zhongdao Wang, Luming Tang, Xihui Liu, Zhuliang Yao, Shuai Yi, Jing Shao, Junjie Yan, Shengjin Wang, Hongsheng Li, and Xiaogang Wang. Orientation invariant feature embedding and spatial temporal regularization for vehicle re-identification. *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 379–387, 2017.
- [29] Yue Yao, Liang Zheng, Xiaodong Yang, Milind Naphade, and Tom Gedeon. Simulating content consistent vehicle datasets with attribute descent. *arXiv:1912.08855*, 2019.
- [30] Xiaohua Zhai, Avital Oliver, Alexander Kolesnikov, and Lucas Beyer. S4l: Self-supervised semi-supervised learning. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1476–1485, 2019.
- [31] Yu Lin Zhang and Qiang Yang. A survey on multi-task learning. *ArXiv*, abs/1707.08114, 2017.
- [32] Zhun Zhong, Liang Zheng, Donglin Cao, and Shaozi Li. Re-ranking person re-identification with k-reciprocal encoding. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3652–3661, 2017.
- [33] Yi Zhou, Connelly Barnes, Jingwan Lu, Jimei Yang, and Hao Li. On the continuity of rotation representations in neural networks. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5738–5746, 2018.
- [34] Yi Zhou and Ling Shao. Aware attentive multi-view inference for vehicle re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6489–6498, 2018.
- [35] Yi Zhou and Ling Shao. Vehicle re-identification by adversarial bi-directional lstm network. *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 653–662, 2018.
- [36] Yi Zhou and Ling Shao. Viewpoint-aware attentive multi-view inference for vehicle re-identification. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6489–6498, 2018.