HA-VReID: An Effective Hard Attention Model with Deep Learning for Vehicle Re-Identification

Imen ZITOUNI *, Emna BEN BAOUES†, Taher SLIMI ‡, Ibtissem CHERNI §, Anouar BEN KHALIFA¶
* Université de Tunis El Manar, Ecole Nationale d'Ingénieurs de Tunis, 1002, Tunis, Tunisie;
*†¶Université de Sousse, Ecole Nationale d'Ingénieurs de Sousse,
LATIS- Laboratory of Advanced Technology and Intelligent Systems, 4023, Sousse, Tunisie;
† ‡ ¶ Université de Jendouba, Institut National des Technologies et des Sciences du Kef, 7100, Le Kef, Tunisie;
§ Université de Jendouba, Institut Supérieur d'Informatique du Kef, 7100, Le Kef, Tunisie;
Email: *imen.zitouni@enit.utm.tn, ¶anouar.benkhalifa@eniso.u-sousse.tn

***Abstract*—Vehicle Re-Identification involves identifying and matching a target vehicle with images captured from different views in a multi-camera network. This topic holds significant importance in various applications including intelligent transportation systems, video surveillance and smart city. However, Vehicle Re-Identification faces significant challenges in dynamic environments due to viewpoint variations, inter-vehicle appearance similarity, intra-class variability, illumination variation, occlusion and background clutter. To address these limitations, we propose HA-VReID, An Effective Hard Attention Model with Deep Learning for Vehicle Re-Identification that combines Hard attention mechanism for background removal and vehicle shape focus with EfficientNet-powered feature extraction for robust vehicle representation. Extensive experiments on the VeRi-776 and VRAI benchmarks demonstrate that our approach outperforms state-of-the-art methods in vehicle Re-Identification tasks.**

***Index Terms*—Vehicle Re-Identification, Hard Attention, EfficientNet, Background Removal, Intelligent Transportation Systems.**

I. Introduction

Global population growth has led to a surge in vehicle numbers, surpassing sustainable levels and making them a key research focus. Vehicle based applications including vehicle detection, vehicle tracking and vehicle type recognition, play important roles in our daily lives such as intelligent transportation systems, video-surveillance, border control and security [1].

Recently, with significant advancements in deep learning paradigms and their success in image analysis and pattern recognition systems, vehicle Re-Identification (Re-ID) has garnered the attention of many researchers. This topic aims to obtain additional information about a specific target in order to re-identify, track and match vehicles based on their similarities. However, vehicle Re-ID remains a challenging task due to multiple factors including viewpoint and illumination changes, resolution variations, spatial and temporal uncertainty and background clutter, especially when extracting relevant and robust features from vehicle image including type, color and the overall vehicle structure [1], [2].

Despite these difficulties, numerous approaches aim to develop solutions using deep learning-based methods and seek to extract characteristics of the input image. However, in order to obtain information at different degrees of granularity and to identify more discriminative information to characterize target vehicles, hierarchical feature learning is necessary. In reality, multi-granularity feature learning [3] aims to capture global information such as the overall vehicle shape, color and model of the vehicle, as well as various local properties that facilitate the recognition and extraction of specific information from defined areas of the image such as side mirrors, rear lights and wheels.

Recent advances in vehicle Re-ID have increasingly leveraged attention mechanism, inspired by human the visual perception systems [4]–[6]. These self-adaptive approaches enhance feature discrimination by selectively focusing on relevant regions while suppressing irrelevant information. Current implementations primarily utilize two complementary paradigms: (1) soft attention, which applies continuous weighting to image regions [5], and (2) hard attention that focuses discretely on the critical elements while aggressively filtering out background clutter and non-essential objects [4].

Building on this foundation, researchers have developed specialized attention variants for vehicle recognition. Notably, spatial attention mechanisms specifically target discriminative local features such as license plates and manufacturer logos [7]. Meanwhile, self-attention architectures dynamically compute inter-region dependencies across the entire input image, enabling richer and more coherent feature integration [6]. This evolution from basic attention frameworks to sophisticated, vehicle-specific implementations has significantly advanced the field's capability to handle real-world challenges including viewpoint variation, occlusion and background clutter.

Despite these significant advancements, a major challenge in vehicle Re-ID remains associated with scalability, particularly regarding the depth, width and resolution of neural networks. While increasing network depth enhances the ability to learn complex feature identification, it risks gradient vanishing issues. Similarly, expanding network width may improve learning capacity but incurs significant computational overhead, particularly for high-resolution vehicle images. Moreover, excessive resolution in complex environments can diminish performance by over-emphasizing irrelevant details.

To address these challenges, EfficientNet's compound scaling method provides an elegant solution [8]. Unlike conventional approaches that scale dimensions arbitrarily, EfficientNet optimizes depth, width and resolution simultaneously through a unified coefficient. This balanced scaling approach has established EfficientNet as a versatile architecture in computer vision, achieving superior performance with reduced computational demands. These scaling advantages directly support the attention mechanisms discussed previously. By efficiently processing high-resolution vehicle images while maintaining manageable computational costs, EfficientNet enables robust feature extraction for both spatial attention (local details like license plates) and self-attention (global and local vehicle representations). This synergy between optimized network architecture and adaptive attention mechanism explains their combined effectiveness in vehicle Re-ID systems.

Thus we present, in this study, a new vehicle Re-ID framework entitled An Effective Hard Attention Model with Deep Learning for Vehicle Re-Identification (HA-VReID), which efficiently removes noisy-background clutter by processing only the vehicle region, therefore reducing unnecessary computations. Moreover, our proposed approach excel in addressing traditional scaling problems using the EfficientNet model by focusing on relevant vehicle structure.

The main contributions of our paper are outlined below:

- We propose an innovative approach called HA-VReID that not only optimizes vehicle Re-ID performance in complex scenarios but also protects sensitive vehicle location data.
- To address traditional scaling challenges, we adapt EfficientNet by modifying its classifier component and replacing it with a specific linear layer optimized for vehicle Re-ID
- Rich experimentation on two public benchmark datasets, successfully confirm our HA-VReID's superiority over state-of-the-art methods in handling real-world vehicle Re-ID complexities.

The remaining part of our paper is organized as follows: Section 2 presents a literature review on vehicle Re-ID using various approaches. Section 3 provides a detailed explanation of our HA-VReID approach. Section 4 presents qualitative and quantitative evaluations. The conclusion is presented in Section 5.

II. Related Work on Vehicle Re-Identification

In recent years, research has given increased emphasis to person Re-ID by paying precious attention to the human body, facial features and clothing attributes [9]. However, vehicle Re-ID presents a complex set of challenges compared to person Re-ID. First, many vehicles from different brands share nearly identical designs, making them hard to distinguish. Second, the same vehicle can appear drastically different across camera viewpoints or lighting conditions. Additionally, color perception varies due to camera settings, illumination changes and environmental factors. Finally, cluttered or dynamic backgrounds further complicate recognition by introducing visual noise. These factors make vehicle Re-ID particularly complex in real-world surveillance scenarios [10].

An overview of relevant research in vehicle Re-ID reveals numerous proposed methods, which can be categorized into four main approaches: (1) traditional machine learning, (2) local and global feature extraction, (3) metric learning and (4) attention mechanisms, as summarized in Table I.

Traditional machine learning is a feature extraction method that focuses on certain aspects of the image including edge information (Histogram of Oriented Gradient(HOG)), local appearance (scale-Invariant Feature Transform) and image texture (Local Binary Patterns). For instance, D. Zapletal and A. Herout apply color histograms and HOG models associated with a linear regressor to solve vehicle detection problem via a 3D full-fledged bounding-box [11].

Furthermore, vehicle Re-ID based deep learning methods have seen real growth in recent years. They combine multiple layers to automatically capture high-level features from large training datasets. Therefore, their generalization ability and the integration of a large number of parameters ensure the best performance during target identification.

Due to the high similarity between vehicles, some research works focus on global and local features. In this case, Liu *et.al* proposed a model that takes into account regions (RAM) and extracts distinctive visual cues from local areas [12]. This method weights global vehicle parts according to the local region. Moreover, Ma *et.al* [10] proposed a refined-part approach to learn and extract feature embedding and capture fine details using a spatial-based network transformer.

Metric learning, also known as similarity learning, is an approach that learns an optimal distance metric to transform input features into an embedding space, where similar samples are clustered closely while dissimilar samples are separated. This approach reduces the distance between two vehicles sharing the same identifier such as the research work presented in [13] .

In the same context, Chu *et.al* [14] have proposed a novel viewpoint-aware based metric learning method, to tackle poses challenges for vehicle Re-ID. Moreover, Spagnolo et al. [15] proposed an integrated approach combining a CNN backbone with a convolutional baseline using triplet part loss to extract discriminative features. This method specifically addresses challenging scenarios where vehicles are captured from extreme viewing angles or under low-light conditions.

However, to overcome these limitations, recent work has increasingly focused on integrating attention mechanism with deep learning through the use of masks. For instance, Du *et.al* [7] proposed a multi-scale attention framework that effectively combines local and global features to extract discriminative characteristics from vehicle images. Similarly, Gao et al. [16] introduced a multi-dimensional attention network incorporating three distinct attention forms: channel, width and height attention using ResNet-50 backbone. Further advancing this direction, Masum *et.al* [17] developed a hybrid model that integrates self-attention with a global and local orthogonal network, demonstrating improved feature learning for vehicle re-identification.

Unlike most vehicle Re-ID methods that scale only a single dimension (typically depth or width), EfficientNet [8] has emerged as a powerful alternative by simultaneously optimizing depth, width and resolution through its compound scaling method. This comprehensive approach has made EfficientNet particularly valuable for vehicle classification and accurate matching tasks [15].

In our study, we advance this further by uniquely integrating EfficientNet with hard attention, combining the model's inherent architectural efficiency with targeted feature localization to effectively address critical vehicle Re-ID challenges, in particular occlusion and background clutter.

III. HA-VReID: An Effective Hard Attention Model with Deep Learning for Vehicle Re-Identification

We present HA-VReID, an EfficientNet-based vehicle Re-ID framework that employs hard attention to focus exclusively on vehicle regions while eliminating background interference. As shown in Figure 1, our architecture consists of two key components:

- Hard Attention and Data Reconstruction: Processing and enhancing only the vehicle region by removing background from the whole image.
- Vehicle Re-Identification and Image Classification: Using EfficientNet model for better feature extraction and accurate vehicle matching.

A. Hard Attention and Data Reconstruction

Hard attention is a dynamic mechanism that selectively focuses on the most discriminative features of an image while suppressing less relevant information. This technique enhances robustness against lighting variations, viewpoint changes and occlusions, while minimizing interference from complex backgrounds. By concentrating computational resources on key regions, hard attention significantly improves both efficiency and robustness in vehicle Re-ID.

In this section, we present a systematic framework for reconstructing vehicle datasets while preserving identity-critical features. Our goal is to enhance data quality by eliminating noise and irrelevant information, thereby optimizing the input for downstream classification tasks.

Leveraging a deep learning-based hard attention mechanism, our preprocessing pipeline consists of three key phases, as illustrated in Figure 2:

- Image Enhancement: Adjustments in contrast, brightness and sharpness are applied to improve visual clarity and highlight structural details of the vehicle. This step ensures consistent quality across diverse imaging conditions.
- Foreground Extraction: Advanced techniques including adaptive thresholding, semantic segmentation and edge detection are employed to precisely isolate the vehicle from its surroundings. This phase accurately identifies the region of interest, ensuring that only relevant features are retained.
- Background Removal: The final output is generated by masking out non-essential background pixels, which are replaced with a uniform black fill. This step eliminates distractions and standardizes the input format, focusing solely on the vehicle's structure.

These preprocessing stages are crucial for reducing intra-class variance caused by environmental factors particularly occlusion and background clutter while amplifying discriminative features essential for Re-ID. The obtained result is a refined dataset as depicted in Figure 3, that enhances our model generalizability and performance.

B. Vehicle Re-Identification and Image Classification

In visual recognition, image classification is a critical task that aims to assign an image to a specific label within a predefined set. The choice of algorithm plays a pivotal role in this process. Among CNN-based architectures, EfficientNet stands out as a highly efficient and effective scaling method, uniquely optimizing the balance between depth, width and resolution of the network.

Unlike conventional approaches that arbitrarily scale individual dimensions (depth, width or resolution), EfficientNet employs a compound scaling method to systematically adjust all three dimensions using a unified coefficient. This method is grounded in the principle that larger input images demand deeper networks (to expand receptive fields) and wider channels (to capture finer details). The architecture comprises hierarchical blocks, each containing convolutional layers followed by batch normalization and activation functions. Network scalability is achieved by uniformly increasing channel counts across all blocks, as depicted in Figure 1.

EfficientNet's efficiency reduces inference times, making it suitable for real-time applications including surveillance systems and vehicle Re-ID. By fine-tuning the architecture and pre-training on large vehicle Re-ID datasets, we leverage in this study, the EfficientNet effectiveness to extract discriminative features, enabling accurate cross-camera vehicle matching.

Finally, for vehicle classification, we adopt a cross-entropy loss function, also called softmax loss, to address classification problems and avoid overfitting. We also find an optimal solution that automatically adjusts the weights of a learning model.

The cross-entropy loss function is described by equation (1), where CN shows the whole number of classes, N the total batch size, p_i the prediction probability for sample i, $\mathrm{q}(p_i, j)$ is a binary value that returns 1 if $p_i = j$ and 0 otherwise. $\mathrm{S}(\mathrm{W}_i)_j$ returns the model's prediction probability for class j in sample i.

$$\text{Loss} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{CN} \mathrm{q}(p_i, j) \cdot \log(\mathrm{S}(\mathrm{W}_i)_j) \tag{1}$$

The integration of EfficientNet's compound scaling with cross-entropy optimization establishes a robust framework for

TABLE I
SUMMARY OF RELATED WORK IN VEHICLE RE-IDENTIFICATION

Methods		Ref	Year	Characteristics
	Traditional Machine Learning	[11]	2016	This study uses color histograms and HOG combined with linear regression.
Deep Learning based Methods	Global & Local Features	[12]	2018	This work proposes a Region Aware Deep Model (RAM) to incorporate detailed visual representations in local regions by extracting local characteristics for each region.
		[10]	2019	This study employs local feature extraction for each segment of vehicle images to capture fine details and enhance the model's ability to discriminate between similar vehicles.
	Metric Learning	[13]	2017	This work improves discrimination between similar vehicles and adapts a fusion of characteristics including vehicle model and color.
		[18]	2018	This approach uses a triplet loss function that optimizes feature learning.
		[14]	2019	This work uses metric learning to enhance vehicle model performance under a variety of viewpoints.
		[15]	2022	Spagnolo *et al.* propose a CNN and Part-Based Convolutional Baseline approach with triplet loss, evaluating backbones (ResNet, SENet, EfficientNet) to extract discriminative vehicle features.
	Attention Mechanism	[7]	2021	Du *et.al* propose a multi-scale attention mechanism that combines both local and global features to extract discriminative features from vehicle images, while accounting for scale variations and differences between similar vehicles.
		[19]	2021	TAihuan Yao *et.al* propose a network based on simple color annotations and attention masks from Unmanned Aerial Vehicles (UAV) images.
		[16]	2022	The author employs a multi-dimensional attention network that combines three different forms of attention: channel, width and height attention, to enhance the learning of discriminative features using Resnet-50.
		[17]	2023	Masum *et.al* propose a deep global and local orthogonal network for vehicle Re-ID, integrating self-attention.
		[6]	2024	This method fuses spatial-channel self-attention while reducing occlusion and background interference.

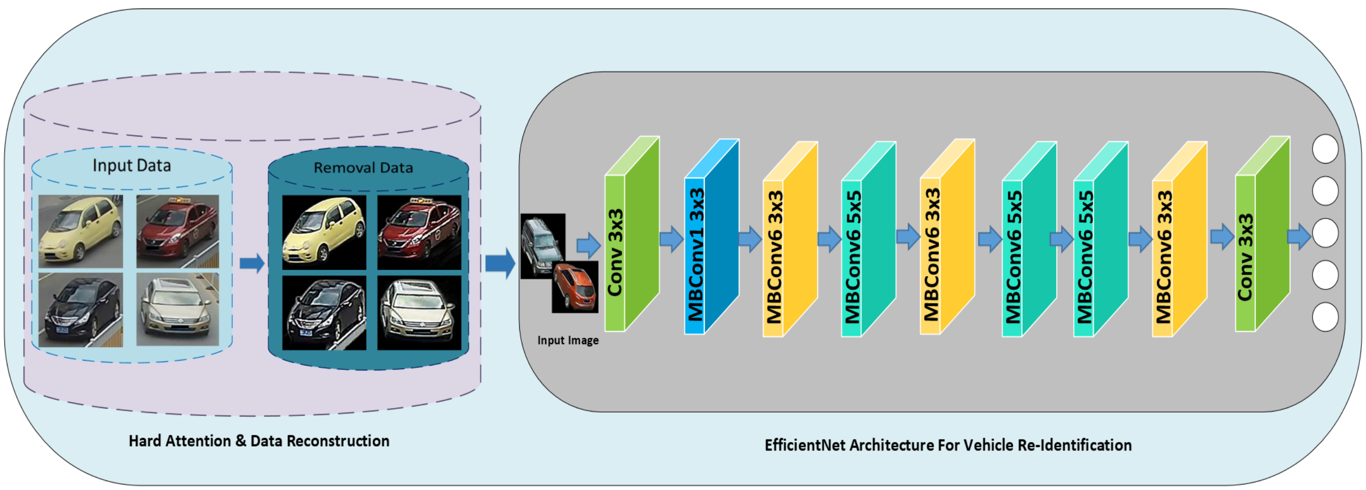

Fig. 1. The HA-VReID framework: An EfficientNet-based vehicle Re-ID system using hard attention for foreground focus and background removal.

vehicle Re-ID, systematically balancing model capacity and efficiency. The formulated loss function (Eq. 1) ensures discriminative feature learning, while the hierarchical architecture enables scalable feature extraction across varying resolutions.

IV. EXPERIMENTAL RESULTS

We evaluate our HA-VReID framework against state-of-the-art vehicle Re-ID methods using two large-scale benchmarks: VeRi-776 [20] and VRAI [21]. These datasets capture real-world surveillance scenarios from both Closed-Circuit Television (CCTV) systems and Unmanned Aerial Vehicles (UAV) platforms, providing comprehensive testing conditions under varying viewpoints, illumination changes and occlusions as depicted in Table II.

A. Experimental Setup

All experiments were conducted on a workstation equipped with an Intel ® Core™ i7 processor (64-bit architecture), 16GB DDR4 RAM and an NVIDIA GeForce RTX 3060 GPU with 8GB GDDR6 memory. The system ran on Ubuntu 20.04 LTS with PyTorch 1.9.0 and CUDA 11.1 for GPU acceleration.

B. Datasets and Experimental Evaluation Protocol

VeRi-776 [20] a widely-used dataset from Beijing University of Posts and Telecommunications, contains 49,360 images of 776 vehicle identities (576 IDs for the training set with 37,781 images and 200 IDs for the test set). VeRi-776 is divided into 1,678 query vehicle images and 11,579 gallery images. These images are captured by 20 surveillance cameras in a real scene and labeled by different attributes

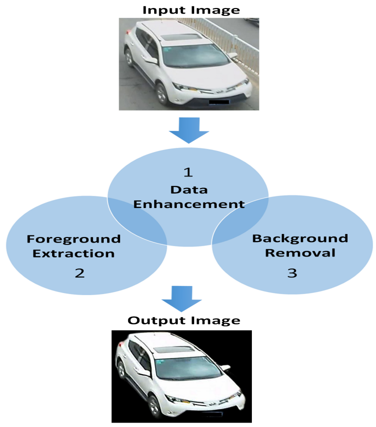

Fig. 2. Visual pipeline of our vehicle preprocessing framework showing: (1) Image enhancement through contrast/brightness adjustment, (2) Foreground extraction via semantic segmentation and (3) Background removal with uniform masking, producing clean vehicle-centric inputs for Re-ID

Fig. 3. Results of our background removal approach, demonstrating clean vehicle segmentation while maintaining discriminative Re-ID features.

including brands, models, colors and license plate according to different viewpoints, lighting conditions and occlusion.

VRAI (Vehicle Re-identification for Aerial Image) [21], a large-scale UAV dataset published by School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an, China. VRAI contains 137,613 images of 13,022 vehicles with 6,302 identities for the training with 66,113 images and 71,500 images for testing with 6,720 identities. Moreover, this dataset is further divided into a query set (25% images) and a gallery set (75% images). Thus we find 15,747 images for queries and 55,753 images for galleries, each class is captured by two drones in different locations with different viewing angles labeled by different attributes including vehicle type, color, binary attributes such as spare tire, skylight, bumper or luggage rack and discriminating parts (frontal windows, luggage racks).

C. Experimental results

Quantitative Results: We compare our HA-VReID method with state-of-the-art approaches in Tables III and IV. Our method achieves superior performance with mean Average Precision (mAP) scores of 82.99% on VeRi-776 and 83.2%

TABLE II
OVERVIEW AND STATISTICS OF VERI-776 AND VRAI DATASETS

Dataset	VeRi-776	VRAI
Year	2016	2019
Images	49,357	137,613
Identities	776	13,022
Train	37,781	66,113
Test	11,579	71,500
Cameras	20	2
Occlusion	✓	×
Complex Background	✓	×

on VRAI, outperforming existing deep learning-based approaches.

The EfficientNet backbone provides effective feature extraction and vehicle classification capabilities. By modifying its final layer, we optimize the model for the Re-ID task while maintaining its computational efficiency. Our attention mechanism successfully focuses on vehicle features while eliminating background clutter, leading to these competitive results.

TABLE III
PERFORMANCE OF HA-VREID APPROACHE IN VEHICLE RE-ID ON VERI-776

Model	mAP(%)
OIFE [22]	48
RAM [12]	61.5
VANet [14]	66.3
SPAN [23]	68.6
PVEN [24]	79.5
VGM [25]	79.71
ViT-ReID [26]	82.1
OURS	82.99

TABLE IV
PERFORMANCE OF HA-VREID APPROACHE IN VEHICLE RE-ID ON VRAI

Model	mAP(%)
RAM [25]	69.37
SPAN [25]	76.71
VGM [25]	79.71
SPCN [27]	80.12
OURS	83.2

Qualitative Results: We assess HA-VReID's retrieval performance through cosine similarity analysis between query and gallery features. Figure 4 demonstrates our model's top-5 matches from the VeRi-776 dataset, ranked by similarity scores. The results showcase the system's ability to: (1) accurately identify query vehicles across different camera views, (2) maintain discriminative power among visually similar instances, and (3) remain robust against viewpoint variations and background clutter. This visual analysis complements our quantitative metrics, confirming the effectiveness of our attention-based feature learning approach.

V. CONCLUSION

Vehicle Re-ID is a key component for a wide range of application areas such as border control, video surveillance and

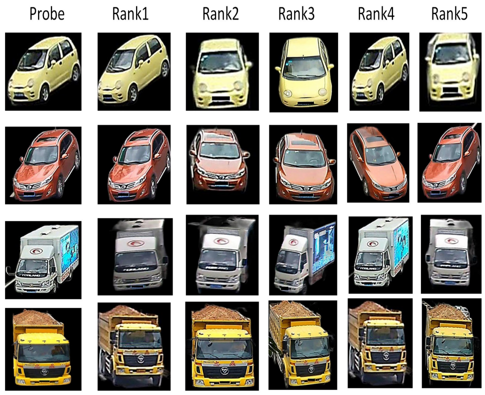

Fig. 4. Top-5 retrieval results of our HA-VReID approach on the VeRi-776 dataset, showing successful matches despite appearance variations.

security. However, several issues including viewpoint changes, illumination variation, dynamic and background clutter are present making this task more difficult. To address these challenges, we propose HA-VReID that leverage the potential of hard attention for background removal reducing unnecessary computations and EfficientNet model for effective feature extraction. Our method achieves state-of-the-art performance (82.99% mAP on VeRi-776, 83.2% on VRAI) by focusing computation on relevant vehicle regions while eliminating distracting elements. The results demonstrate superior mAP and matching accuracy compared to existing approaches, validating HA-VReID's effectiveness for real-world surveillance and intelligent transportation systems.

REFERENCES

[1] H. Wang, J. Hou, and N. Chen, "A survey of vehicle re-identification based on deep learning," *IEEE Access*, vol. 7, pp. 172 443–172 469, 2019.

[2] Zakria, J. Deng, Y. Hao, M. S. Khokhar, R. Kumar, J. Cai, J. Kumar, and M. U. Aftab, "Trends in vehicle re-identification past, present, and future: A comprehensive review," *Mathematics*, vol. 9, no. 24, p. 3162, 2021.

[3] X. Tian, X. Pang, G. Jiang, Q. Meng, and Y. Zheng, "Vehicle re-identification based on global relational attention and multi-granularity feature learning," *IEEE Access*, vol. 10, pp. 17 674–17 682, 2022.

[4] E. B. Baoues, I. Jegham, S. Ameur, and A. B. Khalifa, "Hard attention based efficientnet for person re-identification," in *2024 10th International Conference on Control, Decision and Information Technologies (CoDIT)*, 2024, pp. 976–981.

[5] E. Ben Baoues, I. Jegham, M. E. Yacoubi, and A. B. Khalifa, "Soft-attention based person re-identification in real-world settings using variational autoencoders," in *2024 16th International Conference on Human System Interaction (HSI)*, 2024, pp. 1–6.

[6] Y. Jiao, S. Qiu, L. Sun, D. Han, Q. Li, and M. Chen, "Dsa-scgc: A dual self-attention mechanism based on space-channel grouped compression for vehicle re-identification," in *2024 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2024, pp. 1–8.

[7] L. Du, C. Yu, C. Shuai, X. Liu, J. Yang, and Y. Zhang, "A multiscale attention mechanism based vehicle re-identification," in *2021 7th International Conference on Systems and Informatics (ICSAI)*. IEEE, 2021, pp. 1–6.

[8] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*. PMLR, 2019, pp. 6105–6114.

[9] E. B. Baoues, I. Jegham, S. Ameur, and A. B. Khalifa, "Gf2preid: A novel framework for person re-identification using generative networks," in *2023 International Conference on Cyberworlds (CW)*, 2023, pp. 102–109.

[10] X. Ma, K. Zhu, H. Guo, J. Wang, M. Huang, and Q. Miao, "Vehicle re-identification with refined part model," in *2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. IEEE, 2019, pp. 603–606.

[11] D. Zapletal and A. Herout, "Vehicle re-identification for automatic video traffic surveillance," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2016, pp. 25–31.

[12] X. Liu, S. Zhang, Q. Huang, and W. Gao, "Ram: a region-aware deep model for vehicle re-identification," in *2018 IEEE international conference on multimedia and expo (ICME)*. IEEE, 2018, pp. 1–6.

[13] C. Cui, N. Sang, C. Gao, and L. Zou, "Vehicle re-identification by fusing multiple deep neural networks," in *2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*. IEEE, 2017, pp. 1–6.

[14] R. Chu, Y. Sun, Y. Li, Z. Liu, C. Zhang, and Y. Wei, "Vehicle re-identification with viewpoint-aware metric learning," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 8282–8291.

[15] P. Spagnolo, P. L. Mazzeo, F. Buccoliero, P. L. Carcagnì, and C. Distante, "A deep learning approach for vehicle re-identification," in *2022 7th International Conference on Smart and Sustainable Technologies (SpliTech)*. IEEE, 2022, pp. 1–6.

[16] Y. Gao, L. Lu, B. Xu, and D. Chu, "Multi-dimensional attention network for vehicle re-identification," in *2022 6th CAA International Conference on Vehicular Control and Intelligence (CVCI)*. IEEE, 2022, pp. 1–5.

[17] S. I. H. Masum, M. N. Sheikh, M. Rahman, M. S. Islam, M. A. Islam, and S. E. Siddika, "Vehicle re-identification for intelligent transport system using dolgnet attention model," in *2023 8th International Conference on Image, Vision and Computing (ICIVC)*. IEEE, 2023, pp. 693–698.

[18] J. Zhu, H. Zeng, Z. Lei, S. Liao, L. Zheng, and C. Cai, "A shortly and densely connected convolutional neural network for vehicle re-identification," in *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE, 2018, pp. 3285–3290.

[19] A. Yao, M. Huang, J. Qi, and P. Zhong, "Attention mask-based network with simple color annotation for uav vehicle re-identification," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2021.

[20] X. Liu, W. Liu, T. Mei, and H. Ma, "A deep learning-based approach to progressive vehicle re-identification for urban surveillance," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*. Springer, 2016, pp. 869–884.

[21] P. Wang, B. Jiao, L. Yang, Y. Yang, S. Zhang, W. Wei, and Y. Zhang, "Vehicle re-identification in aerial imagery: Dataset and approach," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 460–469.

[22] Z. Wang, L. Tang, X. Liu, Z. Yao, S. Yi, J. Shao, J. Yan, S. Wang, H. Li, and X. Wang, "Orientation invariant feature embedding and spatial temporal regularization for vehicle re-identification," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 379–387.

[23] T.-S. Chen, C.-T. Liu, C.-W. Wu, and S.-Y. Chien, "Orientation-aware vehicle re-identification with semantics-guided part attention network," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*. Springer, 2020, pp. 330–346.

[24] D. Meng, L. Li, X. Liu, Y. Li, S. Yang, Z.-J. Zha, X. Gao, S. Wang, and Q. Huang, "Parsing-based view-aware embedding network for vehicle re-identification," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 7103–7112.

[25] C. Zhang, C. Yang, D. Wu, H. Dong, and B. Deng, "Cross-view vehicle re-identification based on graph matching," *Applied Intelligence*, vol. 52, no. 13, pp. 14 799–14 810, 2022.

[26] L. Du, K. Huang, and H. Yan, "Vit-reid: A vehicle re-identification method using visual transformer," in *2023 3rd International Conference on Neural Networks, Information and Communication Engineering (NNICE)*. IEEE, 2023, pp. 287–290.

[27] J. Li, Y. Cong, L. Zhou, Z. Tian, and J. Qiu, "Super-resolution-based part collaboration network for vehicle re-identification," *World Wide Web*, vol. 26, no. 2, pp. 519–538, 2023.