

# Regional feature learning using attribute structural analysis in bipartite attention framework for vehicle re-identification

Cynthia Sherin, Kayalvizhi Jayavel

Department of Networking and Communications, School of Computing, SRM Institute of Science and Technology, Chennai, India

---

## Article Info

### Article history:

Received Oct 18, 2022

Revised Feb 16, 2023

Accepted Mar 9, 2023

---

### Keywords:

Bipartite framework

Partition-alignment

Region detection

Self-attention

Vehicle re-ID

---

## ABSTRACT

Vehicle re-identification identifies target vehicles using images obtained by numerous non-overlapping real-time surveillance cameras. The effectiveness of re-identification is further challenging because of illumination changes, pose differences of captured images, and resolution. Fine-grained appearance changes in vehicles are recognized in addition to the coarse-grained characteristics like color of the vehicle along with model, and other custom features like logo stickers, annual service signs, and hangings to overcome these challenges. To prove the efficiency of our proposed bipartite attention framework, a novel dataset called Attributes27 which has 27 labelled attributes for each class are created. Our framework contains three major sections: The first section where the overall and semantic characteristics of every individual vehicle image are extracted by a double branch convolutional neural network (CNN) layer. Secondly, to identify the region of interests (ROIs) each branch has a self-attention block linked to it. Lastly to extract the regional features from the obtained ROIs, a partition-alignment block is deployed. The results of our proposed system's evaluation on the Attributes27 and VeRi-776 datasets has highlighted significant regional attributes of each vehicle and improved the accuracy. Attributes27 and VeRi-776 datasets exhibits 98.5% and 84.3% accuracy respectively which are comparatively higher than the existing methods with 78.6% accuracy.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



---

## Corresponding Author:

Cynthia Sherin

Department of Networking and Communications, School of Computing, SRM Institute of Science and Technology

Kattankulathur, Chennai-603203, India

Email: cb4881@srmist.edu.in

---

## 1. INTRODUCTION

Detection and re-identification of vehicles [1] have gained significant importance in computer vision because of its varied applications on video surveillance. In intelligent transportation, it is highly challenging to accurately perform vehicle re-identification (Re-ID) by retrieving images of vehicles captured by multiple non-overlapping cameras. This is mostly performed in illegal vehicle search systems and urban surveillance. Due to the similarity in vehicle appearance, person re-identification methods failed to provide adequate results for vehicle re-identification. They can be distinguished as similar vehicles in varied models (inter-class similarity) as in Figure 1(a) and same vehicle appearing different (intra-class differences) as in Figure 1(b) along with similarly looking various vehicles of the same model exhibited in Figure 1(c). More emphasis is given on the distinct regional features which differentiate every vehicle like the ocular lights, inspection markings of yearly maintenance, decorative mirror hangings and personalized designs [2]. These minute details of each vehicle are used for accurate re-identification. Thus, the key component of vehicle re-identification is to distinguish the fine-grained dissimilarities of each vehicle thereby highlighting the distinct regions. The

global features of the vehicle such as color, model and make details about the brand and the spatial-temporal information from varied camera angles provides the vehicle direction in the given space and time [3] along with its appearance. The license plate details distinguish each vehicle individually but are commonly not available because of privacy issues, possibility of faking it, distortions, occlusion, motion blur and illumination in the real-world scenario. Lastly the vehicle orientation and camera angles [4] are sometimes unreliable because of the low resolutions in capturing minute visual appearances.

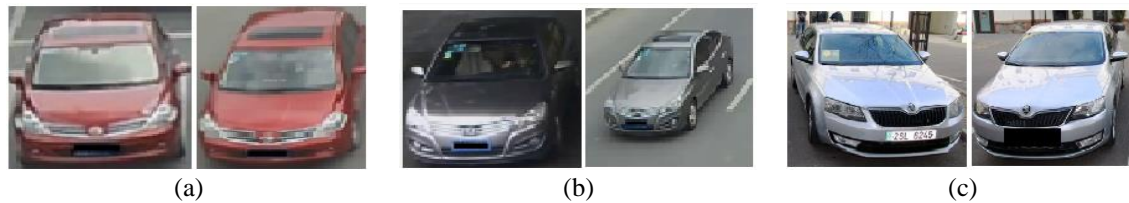


Figure 1. Changes in vehicle appearance (a) similar vehicles of varied models, (b) different appearances of same vehicle, and (c) different vehicles of same model

Personalized local features are considered in addition to the overall structural appearance to enhance the performance of re-identification. An ordered vehicle dataset with 27 labelled classes of attributes are collected and named as Attributes27 dataset in proper bounding boxes which includes various types of cars, windshield glass, wheels, tissue box and back mirror. These images were gathered from a variety of surveillance cameras in urban locations and features several vehicles from various angles (top, face, rear, along with left, and right sides of vehicle), both during the day and at night. Table 1 provides a comprehensive list of the vehicle attributes, when combined with the personalized attributes to improve the efficiency of vehicle re-identification. By using the characteristic attribute data gathered, Attributes27 dataset is leveraged for diverse purposes such as analysis of vehicles along with its tracking [5] and detection, highly defined model-based identification of vehicles [6], logo detection [7], detection of vehicle plate for improved re-identification [8]. Vehicle tracking, model detection, and vehicle re-ID requires varying degrees of structural data.

Table 1. Hierarchical list of labelled attributes (Attributes27)

No	Viewing angle	Dataset attributes						
1	Top	Sliding roof	Luggage	Antenna	Vehicle carrier			
2	Front and Rear	Front lights	New Signs	Back mirror	Safety belt			
		Entry License	Hanging	Light cover	Lay Ornament			
		Windshield glass	Logo	Tissue box	Number plate			
3	Sides (Left, Right)	Windows	Wheel	Annual service signs				
				Decorative markings				
4	Type of Vehicle	SUV	Sedan	Hatchback	Truck	Bus	Pick-up	Jeep

Numerous existing works successfully explains the retrieval mechanisms of positive images along with the loss relationship using multiple vehicle re-identification datasets. With the aid of 20 cameras Liu *et al.* [9] gathered 40,000 images of 619 vehicle identities from unobstructed traffic conditions and devised an re-identification method based on appearance to extract the global and semantic characteristics on the VeRi dataset. To calculate the relative distance between vehicles, Liu *et al.* [10] used a double branched deep convolutional neural network (CNN) on a large labeled re-ID dataset called VehicleID which contains 2,22,000 images of 26,267 vehicle identities from varied real world surveillance cameras. Distinct vehicle datasets based on regional features with proper labelled attributes [11] with the help of bounding boxes are utilized for license plate and logo identification [12]. Along with these information semantic data including vehicle color, surface texture, proportion, and geometric data also enhances the re-identification performance. Vehicle model, color, and unique identification features are taken into account by Liu *et al.* [13] in his multi-branch architecture to improve the performance of re-identification. The cross-sectional manipulation of vehicles is improved by generative adversarial networks (GAN) [14], by feeding the network with vehicle images from various angles. Zhouy and Shao [15] classification system divided the images based on viewpoints into various classes, which were combined together to generate the global features. He *et al.* [16] trained the model by merging the local model with the attention heatmaps to obtain the region of interest (ROI) recommendations. Teng *et al.* [17] obtained the final features of the image by making use of the spatial and channel attention data as a weight

map. Multi-attribute-based attention module was employed by Wang *et al.* [18] to enhance the reidentification task. Zhang *et al.* [19] presented a part-guided attention learning model to uniquely identify the fine-grained attributes such as distinct designs, yearly maintenance stickers, and tissue box to emphasize the various part regions. As a result, the regional features extracted by the self attention model are vital to show up the distinctive qualities of each vehicle image. The relationship between the correctly and wrongly identified images in comparison with the query is determined by the loss function in vehicle re-ID. To produce the loss function, Ding *et al.* [20] employed L2 norms to limit the loss function for the positive and negative images. Schroff *et al.* [21] developed a constant value from the triplet loss function by distinguishing the similar and dissimilar image pairs.

A bipartite framework divides the images into two non-coinciding parts using the numerous labelled attributes. The pattern branch and identity branch are the two main branches which make up the bipartite framework. The self attention model trains the labeled images effectively and the partition-alignment model enhances the accuracy of re-identification.

The significant contributions are in brief, as follows:

- To gather an annotated dataset of labelled vehicle images comprising of 27 attributes per class. For efficient vehicle re-identification, the selective features of individual vehicles are examined using the multi-class attributes.
- To construct a bipartite framework with multiple attributes and independent feature maps for the pattern branch as well as identity branch. The resultant framework produces an effective triplet loss function thereby diminishing the conflict between the multiple attributes.
- To design a self-attention model to produce centralized multi-attribute attention heatmaps in combination with partition-alignment model to precisely identify the windshield area and bumper sections of the vehicle to highlight the fine-grained information.

Our proposed research work has the following sections: section 2 elaborates the comprehensive mechanism of the proposed method along with its architecture. Section 3 displays the experimental results obtained for the proposed method on Attributes27 dataset and standard vehicle re-id dataset VeRi-776. Section 4 concludes the overall analysis.

## 2. METHOD

### 2.1. Characteristics of Attributes27 dataset

Our proposed Attributes27 dataset comprises of 8,271 images gathered from variety of datasets under various environments. The VAC21 vehicle dataset served as a model for Attributes27 dataset containing 21 classes of labeled attributes. To strengthen the regional feature learning capability and detection accuracy of smaller areas, our dataset has been updated with extra attributes. The labeled attributes for each image in the dataset are displayed in Table 1. The visual representation of the labeled attributes is given in Figure 2. This dataset captures minute information of the vehicle at various levels. In the entry level the type of the vehicle image captured is concluded by the body type such as bus, jeep, truck, sedan, sport utility vehicle (SUV), hatchback. In the second level, the vehicle make and model are determined. The final level records the unique vehicle characteristics like yearly maintenance stickers, newer signs, and decorative hangings.



Figure 2. Attributes distribution in a sample image of Attributes27 dataset

### 2.2. Overall architecture

Typically, multiple attributes are used in mono branched CNN architecture, by allocating a common feature space for vehicle re-identification techniques. To address this issue, a bipartite framework with dual

branches-the identity and pattern branch, as shown in Figure 3 is used to analyze multiple attributes with the help of many CNN layers. The input is delivered into the input distribution layer as m groups of n images. With ResNet50 as the foundation CNN, the initial 4 layers are kept the identical and the layer 5 embraces the bipartite network structure. A distinct self attention block creates the heatmaps for the various attributes for each image. The partition-alignment blocks are incorporated on the generated heatmaps which serves the input to create the ROI. To extract the regional features, the resulting ROIs are given as input to a small CNN, ResNet18. In order to train the model, the triplet loss for both the branches is obtained.

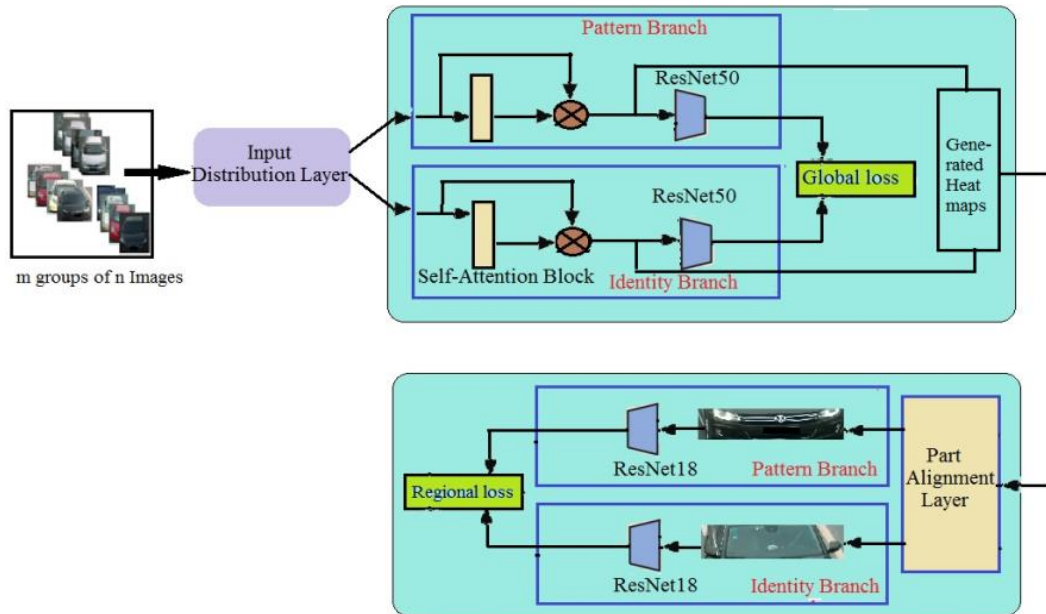


Figure 3. Architectural design of bipartite framework

**2.3. Loss functions**

As the classification loss ignores the fine-tuned appearance attributes acquired from semantic features, triplet loss tends to be comparatively more effective. The training phase is accompanied by both the triplet loss and label smoothing cross entropy loss. The cross-entropy loss with label smoothing is explained with (1).

$$Loss_{cross} = - \left( (1 - \epsilon) * t + \frac{\epsilon}{N} \right) * \sum_{i=1}^N \log \frac{\exp(p_i)}{\sum_{n=1}^N \exp(p_n)} \tag{1}$$

With the ground truth value  $t$  allocated as 1,  $\epsilon$  stands for the smoothing parameter,  $N$  denotes the number of classifications, and  $p_n$  stands for the un-normalized logits of the  $n^{th}$  classification. Pre-processing the input data is highly recommended for combining the cross entropy and triplet loss. The input images are thus divided into a number of groups  $G_1, G_2... G_k$  with a total of  $n$  similar images per group throughout the training phase. Each triplet unit for the given input consists of anchor, positive, negative images. The loss function is given by (2).

$$Loss_{Triplet} = \max(d(g_a, g_p) - d(g_a, g_n) + m, 0) \tag{2}$$

$g^a$  stands for the anchor image, the matching images with  $g^a$  are represented as  $g^p$  and the negative images which are not similar to  $g^a$  are represented as  $g^n$ .  $d$  denotes the distance between  $g^a, g^p, g^n$ .  $M$  stands for the margin value which separates the non-matching images from the matching ones. Thus, the overall loss is denoted as (3).

$$Loss = Loss_{cross} + \omega * Loss_{Triplet} \tag{3}$$

where  $\omega$  is the weights allocated. The dual branches namely the pattern branch and the identification branch are subjected to the triplet loss. Let us assume that there are sample images for the provided anchor image that

has different identities.  $L$  is the distance between the sample images and anchor image. When the identity branch is trained, the distance  $L$  increases as the images belong to various identities and when the images from the same model are trained,  $L$  becomes closer. With the advantage of dual branch in the proposed system it overcomes the conflicts of single branch framework, thereby improving the performance. The experimental analysis on VeRi-776 and Attributes27 dataset results in 84.3% and 98.5% of mAP respectively for the proposed bipartite structure with self attention. Figure 4 displays the comparative study of the framework with enhanced performance and accurate detection of windshield and bumper zones.

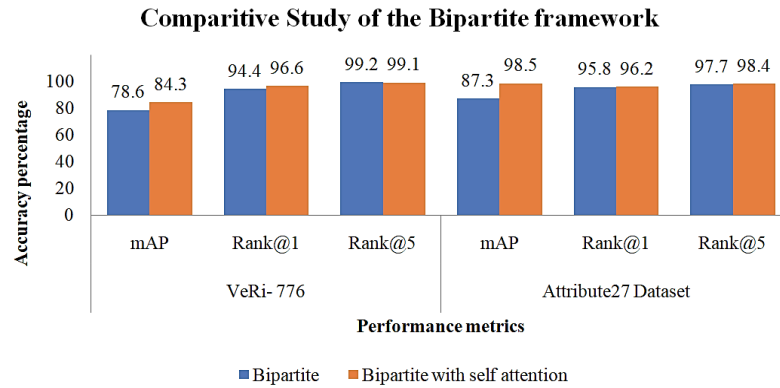


Figure 4. Comparative study of self-attention bipartite framework

**2.4. Self-attention block**

The bipartite structure has a separate feature space for each branch, and an attention model adds more emphasis to the specific areas. The various positions of the similar input image are focused by the self-attention model and the acquired output belongs to the same input series. Thus the independent features are brought together by the self-attention [22] mechanism incorporated in the bipartite network. Figure 5(a) gives the brief architecture of the self-attention framework. The three 1x1 convolution layers C1, C2, C3 are fed with the input  $X$  to generate the  $f$ ,  $g$  and  $h$  heatmaps. The correlation matrix is generated as a product of the heatmaps  $g$  and  $f$  yielding  $s$  which is the probability map. For the input  $X$  the final self-attention map  $SA$  is generated by the cross product of probability map  $s$  and feature map  $h$ . The self-attention map is given as (4).

$$Self - attention Map (SA) = h \otimes ( Softmax ( g^T \otimes f ) )^T \tag{4}$$

**2.5. Partition-alignment block**

With the generated attention heatmap of every attribute, the ROIs are constructed by the partition-alignment block for deriving the final regional features. Multi-level attention pooling is used in the partition-alignment block to produce the ROIs.  $SA_p$  and  $SA_i$  are the self attention heatmaps produced by the identity and pattern branches respectively which is indicated in the partition-alignment block sample structure as shown in Figure 5(b). The output feature maps highlight essential information with important features.

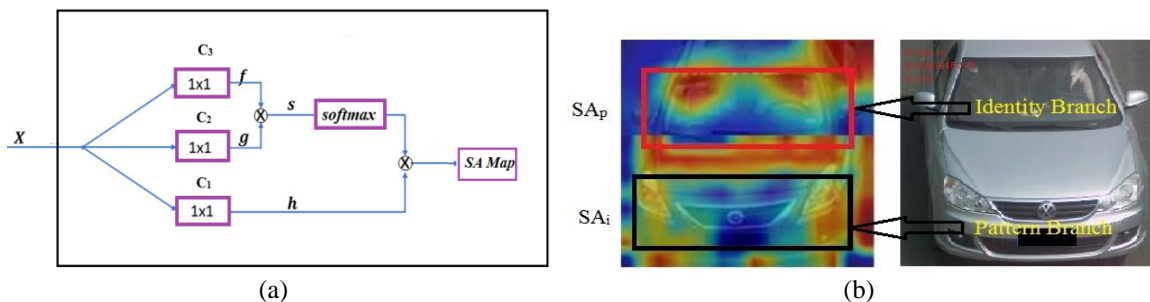


Figure 5. Bipartite network structure (a) self-attention block and (b) windshield and bumper regions highlighted by partition-alignment block

Various average pooling sizes are used since the size of the windshield area is less than the bumper area. After the detection of pattern branch, the bumper area is detected by applying  $4 \times 12$  average pooling on the heatmap  $SA_p$ . For detecting the smaller sized inspection stickers in the occluded windshield area  $SA_i$ , a  $3 \times 3$  pooling layer is utilized. The  $4 \times 10$  average pooling layer processes the pooled map which is assigned as the seed. The resulting heatmaps are separately linked with the input image to distinguish the bumper and windshield regions. The algorithmic explanations of both the self-attention block and partition-alignment block are listed in Table 2.

Table 2. Algorithmic explanation of self-attention block and partition-alignment block

Algorithm 1: Outline of self-attention block	Algorithm 2: Outline of partition-alignment block
Requirement: Let $X$ be the input image.	Requirement: Generated heatmaps for the pattern and identity branch, $SA_p$ and $SA_i$ respectively.
To Accomplish: Generation of the feature maps $f$ , $g$ , and $h$ along with resultant self-attention map $SA$ .	To Accomplish: Obtain bumper and windshield region coordinates $H_p$ and $H_i$ .
Step 1: Generate the feature maps $h$ , $f$ and $g$ by applying the input $X$ on three individual $1 \times 1$ convolution layers $C_1$ , $C_2$ , and $C_3$ respectively.	Step 1: Generation of new heatmap $SA'_p$ by executing $4 \times 12$ average pooling on $SA_p$ ;
Step 2: The probability map $s$ is produced by the product of $f$ and $g$ .	Step 2: The coordinates $H_p$ denotes the maximum value on $SA'_p$ ;
Step 3: The Self-Attention map $SA$ for each branch is generated by the product of $h$ and the probability map $s$ .	Step 3: Generate a new heatmap $SA'_i$ by performing a $3 \times 3$ average pooling on the initial heatmap $SA_i$ ;
	Step 4: Another heatmap $SA''_i$ is generated with $SA'_i$ by undergoing $4 \times 10$ average pooling;
	Step 5: In $SA''_i$ , the pixel with the highest value becomes the seed value;
	Step 6: The pixel in $SA''_i$ includes the region covered by $SA_i$ , adding them to the candidate areas;
	Step 7: The coordinates $H_i$ denotes the candidate region's maximum value;
	Step 8: Return the values $H_p$ and $H_i$ .

### 3. RESULTS AND DISCUSSION

To precisely accomplish the re-identification of vehicles, the entire experimental analysis is carried out on the Amazon elastic cloud compute (EC2) instance of amazon web services (AWS) console. Pytorch and Tensorflow are used in conjunction with a high frequency Intel Xeon Scalable Processor in the p2.xlarge instance for accelerated computing. The bipartite self attention framework is analyzed on both Attributes27 and VeRi-776 dataset. The VeRi-776 dataset consists of 50,000 images of 776 distinct vehicles taken by 20 cameras from various angles. A testing set contains 11,579 images of 200 vehicle identities, while the training set has 37,778 images of 576 vehicles. ResNet-50 and ResNet-18, allocated with ImageNet pretrained weights in the training phase are applied on the bipartite framework and subnet respectively. Through data augmentation techniques the input images are randomly cropped to  $224 \times 224 \times 3$  with a probability of 0.5. The Adam optimizer [23] is employed with a weight decay of  $5 \times 10^{-4}$  and momentum of 0.9. Each batch with 16 different identities of the augmented images are obtained with a size of 120 per batch. A random selection of instances with the same ID is performed, and the count is set to 10. The learning rate set to 0.001 for 100 epochs. The similarity rate between the query and the collection of galleries imaged are provided by Euclidean distance (L2), which is used to assess the final ranking outcome.

The re-identification performance is evaluated by mean average precision (mAP) and cumulative match curve (CMC). The area covered by the precision-recall curve is termed as average precision (AP) and mAP stands for the  $i^{\text{th}}$  query of mean average precision for the total of  $N$  gallery images. The collective percentage of the accurately matching images from top- $K$  images is given by CMC. Rank@1 and rank@5 scores are used to analyze the accuracy in re-identification. The formulae for mAP is given as (5).

$$mAP = \frac{1}{N} \sum_{i=1}^N AP(i) \quad (5)$$

The performance evaluation of the bipartite self-attention framework on VeRi-776 and Attributes27 dataset for both the self attention block and partition alignment block is given in the Table 3. The training accuracy and loss for an average of 10 epochs is tabulated for the trained ResNet-50 and ResNet-18 CNN models for both the blocks respectively to exhibit its robustness in the VeRi-776 and Attributes27 dataset. The overall performance analysis report of the bipartite framework is analyzed based on the metrics such as precision, accuracy and recall. The graph for these metrics plotted in Figure 6 clearly indicates that the accuracy, precision and recall of bipartite with self attention framework shows 86.4%, 86.1%, 86.3% in VeRi-776 and 98.5%, 97.1%, 98.3% in Attributes27 dataset respectively.

Table 3. Average values of accuracy and loss data for Attributes27 and VeRi-776 dataset

Epochs	VeRi-776 Dataset				Attributes27 Dataset			
	ResNet-50(Self-Attention block)		ResNet-18(Partition-Alignment block)		ResNet-50(Self-Attention block)		ResNet-18(Partition-Alignment block)	
	Loss	Accuracy	Loss	Accuracy	Loss	Accuracy	Loss	Accuracy
10	4.175	0.221	2.352	0.358	3.562	0.369	1.953	0.399
20	2.311	0.376	1.971	0.412	3.021	0.492	1.661	0.460
30	1.878	0.354	1.325	0.498	2.746	0.534	1.234	0.523
40	1.235	0.492	0.652	0.547	1.854	0.597	0.863	0.587
50	0.741	0.531	0.551	0.581	1.432	0.662	0.625	0.682
60	0.708	0.658	0.506	0.603	0.896	0.715	0.442	0.731
70	0.567	0.719	0.361	0.711	0.651	0.784	0.360	0.853
80	0.421	0.742	0.225	0.754	0.376	0.861	0.258	0.924
90	0.359	0.809	0.201	0.829	0.258	0.906	0.153	0.968
100	0.274	0.812	0.136	0.843	0.186	0.938	0.102	0.985

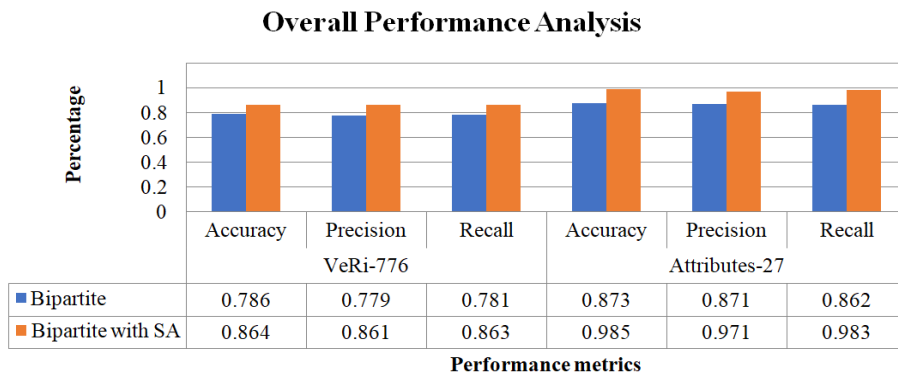


Figure 6. Overall performance analysis for the bipartite framework

**3.1. Comparison with other methods and results**

With the use of the multi-view images from VeRi-776 and Attributes27 datasets, feature maps are created to get the ROIs that provide the regional features needed to compare the performance of re-ID with other cutting-edge techniques. mAP and CMC are employed in VeRi-776 dataset in comparison with the state-of-art methods which is shown in Figure 7. The findings of the comparative analysis on the VeRi-776 dataset show that VAMI+STR [15] and FACT [9] obtain more information from multi-attributes. VAMI [15] and OIFE(4views+ST) [24] gets the multi-view information with the help of attention model. In the case of MSVF, SCAN the dual-branch architecture is utilized and PR+GLBW [16] makes use of LocalNet [25] to get the local information of each vehicle. Our bipartite framework when applied on Attributes27 and VeRi-776 dataset displayed a good improvement in performance of 98.5% and 84.3% respectively. Due to the varied image categories the percentage of accuracy in VeRi-779 dataset has decreased when compared with the Attributes27 dataset.

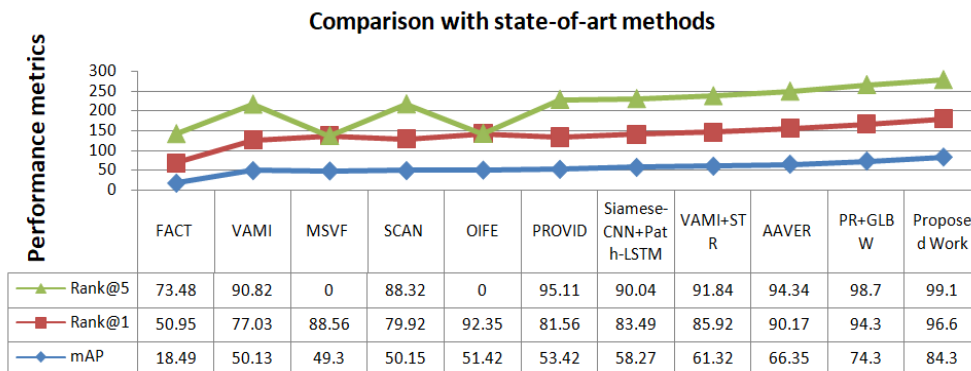


Figure 7. Comparison of VeRi-776 dataset with state-of-art methods

Figure 8(a) shows the outcome of partition alignment and the regions that were detected for the windshield and bumper. With the use of attention heatmaps, the windshield region provides data about the identity branch, and the bumper region provides details of the pattern branch. The back focused images are accurately found with the help of the local features which boosts re-identification. The first five ranking retrieval samples of VeRi-776 and Attributes27 dataset are displayed in Figure 8(b). The query image is present in the leftmost corner and the obtained results are from the gallery images. The correctly retrieved images to the anchor image are highlighted in green color and the wrong entries are displayed red. Similarity in appearance, local and global features, illumination changes can sometimes result in false retrievals in some cases. Thus, our proposed system has effectively overcome the defects of other methods and provided 98.5% re-identification accuracy.

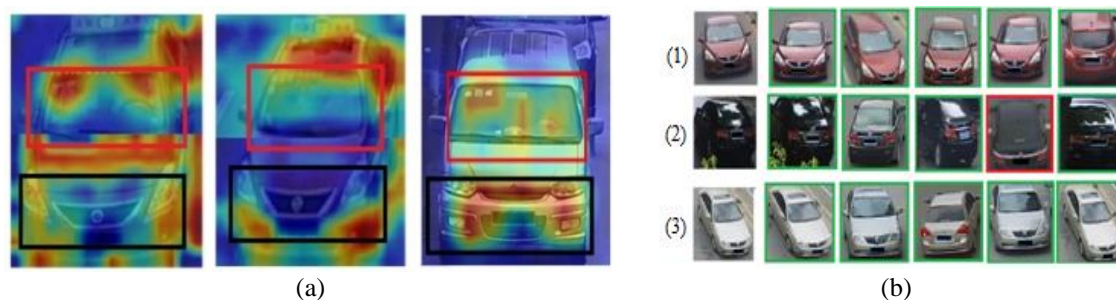


Figure 8. Results of experimental analysis (a) Heatmaps generated on VeRi-776 after partition-alignment and (b) Top 5 rank images on Attributes27(1) and VeRi-776 (2, 3)

#### 4. CONCLUSION

In this paper, we have effectively gathered the regional features with the help of 27 labelled classes of structured attributes in Attributes27 dataset. Both the Attributes27 and VeRi-776 dataset are tested on the proposed bipartite attention framework to achieve remarkable re-identification results. VeRi-776 dataset achieved 99.1% and Attributes27 dataset with 98.4% accuracy results with triplet loss incorporated in the bipartite framework when compared with the other techniques with only 65.7% and 78.6% accuracy. Bipartite architecture successfully overcomes the conflict between the triplet losses. The attention heatmaps generated by the self-attention block combines the individual features and with these heatmaps, the partition-alignment block detects the local regions. An excellent performance of 84.3% and 98.5% is achieved on VeRi-776 and Attributes27 datasets respectively with effective detection of windshield and bumper areas in the captured front angle images.

#### REFERENCES




- [1] W. Fang, J. Chen, C. Liang, Y. Wang, and R. Hu, "Vehicle re-identification collaborating visual and temporal-spatial network," in *Proceedings of the Fifth International Conference on Internet Multimedia Computing and Service*, Aug. 2013, pp. 121–125, doi: 10.1145/2499788.2499844.
- [2] L. Yang, P. Luo, C. C. Loy, and X. Tang, "A large-scale car dataset for fine-grained categorization and verification," *Prepr. arXiv.1506.08959*, Jun. 2015.
- [3] X. Liu, W. Liu, T. Mei, and H. Ma, "PROVID: Progressive and multimodal vehicle reidentification for large-scale urban surveillance," *IEEE Transactions on Multimedia*, vol. 20, no. 3, pp. 645–658, Mar. 2018, doi: 10.1109/TMM.2017.2751966.
- [4] Y. Lou, Y. Bai, J. Liu, S. Wang, and L.-Y. Duan, "Embedding adversarial learning for vehicle re-identification," *IEEE Transactions on Image Processing*, vol. 28, no. 8, pp. 3794–3807, Aug. 2019, doi: 10.1109/TIP.2019.2902112.
- [5] B. C. Matei, H. S. Sawhney, and S. Samarasekera, "Vehicle tracking across nonoverlapping cameras using joint kinematic and appearance features," in *CVPR 2011*, Jun. 2011, pp. 3465–3472, doi: 10.1109/CVPR.2011.5995575.
- [6] A. Psyllos, C. N. Anagnostopoulos, and E. Kayafas, "Vehicle model recognition from frontal view image measurements," *Computer Standards & Interfaces*, vol. 33, no. 2, pp. 142–151, Feb. 2011, doi: 10.1016/j.csi.2010.06.005.
- [7] Y. Huang, R. Wu, Y. Sun, W. Wang, and X. Ding, "Vehicle logo recognition system based on convolutional neural networks with a pretraining strategy," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 1951–1960, Aug. 2015, doi: 10.1109/TITS.2014.2387069.
- [8] D. Zapletal and A. Herout, "Vehicle re-identification for automatic video traffic surveillance," in *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2016, pp. 1568–1574, doi: 10.1109/CVPRW.2016.195.
- [9] X. Liu, W. Liu, H. Ma, and H. Fu, "Large-scale vehicle re-identification in urban surveillance videos," in *2016 IEEE International Conference on Multimedia and Expo (ICME)*, Jul. 2016, pp. 1–6, doi: 10.1109/ICME.2016.7553002.
- [10] H. Liu, Y. Tian, Y. Wang, L. Pang, and T. Huang, "Deep relative distance learning: Tell the difference between similar vehicles," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 2167–2175, doi: 10.1109/CVPR.2016.238.






- [11] Y. Zhao, C. Shen, H. Wang, and S. Chen, "Structural analysis of attributes for vehicle re-identification and retrieval," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 2, pp. 723–734, Feb. 2020, doi: 10.1109/TITS.2019.2896273.
- [12] D. F. Llorca, R. Arroyo, and M. A. Sotelo, "Vehicle logo recognition in traffic images using HOG features and SVM," in *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, Oct. 2013, pp. 2229–2234, doi: 10.1109/ITSC.2013.6728559.
- [13] X. Liu, S. Zhang, Q. Huang, and W. Gao, "RAM: a region-aware deep model for vehicle re-identification," *Prepr. arXiv.1806.09283*, Jun. 2018.
- [14] Y. Wang, Z. Liu, and F. Xiao, "A fast coarse-to-fine vehicle logo detection and recognition method," in *2007 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, Dec. 2007, pp. 691–696, doi: 10.1109/ROBIO.2007.4522246.
- [15] Y. Zhouy and L. Shao, "Viewpoint-aware attentive multi-view inference for vehicle re-identification," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2018, pp. 6489–6498, doi: 10.1109/CVPR.2018.00679.
- [16] B. He, J. Li, Y. Zhao, and Y. Tian, "Part-regularized near-duplicate vehicle re-identification," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2019, pp. 3992–4000, doi: 10.1109/CVPR.2019.00412.
- [17] S. Teng, X. Liu, S. Zhang, and Q. Huang, "SCAN: Spatial and channel attention network for vehicle re-identification," in *PCM 2018: Advances in Multimedia Information Processing – PCM 2018*, 2018, pp. 350–361.
- [18] H. Wang, J. Peng, D. Chen, G. Jiang, T. Zhao, and X. Fu, "Attribute-guided feature learning network for vehicle re-identification," *Prepr. arXiv.2001.03872*, Jan. 2020.
- [19] X. Zhang, R. Zhang, J. Cao, D. Gong, M. You, and C. Shen, "Part-guided attention learning for vehicle instance retrieval," *Prepr. arXiv.1909.06023*, Sep. 2019.
- [20] S. Ding, L. Lin, G. Wang, and H. Chao, "Deep feature learning with relative distance comparison for person re-identification," *Prepr. arXiv.1512.03622*, Dec. 2015.
- [21] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: a unified embedding for face recognition and clustering," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2015, pp. 815–823, doi: 10.1109/CVPR.2015.7298682.
- [22] A. Vaswani *et al.*, "Attention Is All You Need," *Prepr. arXiv:1706.03762*, Jun. 2017.
- [23] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *Prepr. arXiv.1412.6980*, Dec. 2014.
- [24] Z. Wang *et al.*, "Orientation invariant feature embedding and spatial temporal regularization for vehicle re-identification," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Oct. 2017, pp. 379–387, doi: 10.1109/ICCV.2017.49.
- [25] P. Khorramshahi, A. Kumar, N. Peri, S. S. Rambhatla, J.-C. Chen, and R. Chellappa, "A dual-path model with adaptive attention for vehicle re-identification," *Prepr. arXiv.1905.03397*, May 2019.

## BIOGRAPHIES OF AUTHORS



**Cynthia Sherin**    received the B.Tech. degree in Information Technology and M.E. degree in Computer Science Engineering from Anna University, Chennai, in 2012 and 2018. She is currently pursuing her Ph.D. degree in Networking and Communication engineering from SRM University, Chennai, India. She has a corporate experience of 2 years in networking and teaching experience of 2 years as Assistant Professor at the Department of Computer Science Engineering. Her research interests include machine learning and deep learning techniques in computer vision, image processing, ad hoc networks and internet of things. She can be contacted at email: cb4881@srmist.edu.in.



**Kayalvizhi Jayavel**    is founder of IoT Alliance club, a not to profit unit to train students in the field of IoT, www.iotalliance.in. She served as IT Coordinator for National Family Health Survey (NFHS5) for Tamilnadu-South under Ministry of Health and Family Welfare, Government of India. She has published papers in reputed journals with international collaborations with countries including Japan, Oman and Rwanda. She is Professor Mentor for few Samsung research worklets implemented at SRM. She is invited Professor and international research supervisor at African Centre of Excellence of Internet of Things (ACEIoT), University of Rwanda (2019 to till date). She is Faculty advisor for ACM SIGAI chapter at SRM, which received a fund of 2000 Euros from AI Journal towards the application submission "AAIW for under privileged" (2019). She has been a visiting researcher to NEC (Japan) for a collaborative research project in the domain of WSN (2011). She can be contacted at email: kayalvij@srmist.edu.in.