Zero Shot Vehicle Re-Identification

Author

Muhammad Zohaib Nasir

MS-17 (CSE) 00000204890

Supervisor

Dr. Ali Hassan

DEPARTMENT OF COMPUTER ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD

MAY 2021

Zero Shot Vehicle Re-Identification

Author

Muhammad Zohaib Nasir

MS-17 (CSE) 00000204890

A thesis submitted in fulfillment of the requirements for the degree of

MS Software Engineering

Thesis Supervisor

Dr. Ali Hassan

Thesis Supervisor's Signature:

_____

DEPARTMENT OF COMPUTER ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY,

ISLAMABAD

MAY 2021

Declaration

I certify that this research work titled "*Zero Shot Vehicle Re-Identification*" is my own work. The work has not been presented elsewhere for assessment. The material that has been used from other sources it has been properly acknowledged / referred.

Signature of Student

Muhammad Zohaib Nasir

MS-17 (CSE) 00000204890

Language Correctness Certificate

This thesis has been read by an English expert and is free of typing, syntax, semantic, grammatical and spelling mistakes. Thesis is also according to the format given by the university.

Signature of Student

Muhammad Zohaib Nasir

MS-17 (CSE) 00000204890

Signature of Supervisor

Dr. Ali Hassan

Copyright Statement

Acknowledgements

All praise and glory to Almighty **Allah** (the most glorified, the highest) who gave me the courage, patience, knowledge and ability to carry out this work and to persevere and complete it satisfactorily. Undoubtedly, HE eased my way and without HIS blessings I can achieve nothing.

I would like to express my sincere gratitude to my advisor **Dr. Ali Hassan** for boosting my morale and for his continual assistance, motivation, dedication and invaluable guidance in my quest for knowledge. I am blessed to have such a co-operative advisor and kind mentor for my research.

Along with my advisor, I am profusely thankful to **Dr. Arslan Shaukat** for an excellent guidance throughout this journey and for being part of my evaluation committee.

It is indeed a privilege to thank my father, my mother, my wife, and my brothers for their constant encouragement throughout my degree and research period. The sense of belief that they instilled in me has helped me sail through this journey. I would like to thank my family & friends, who have rendered valuable assistance to my study.

Finally, I would like to express my gratitude to all my friends and the individuals who have encouraged and supported me through this entire period.

Dedicated to my exceptional parents, wife, and brothers whose tremendous support and cooperation led me to this accomplishment. At the end this thesis is dedicated to all those who believe in the richness of learning

Abstract

Vehicle re-identification is very useful in intelligent traffic monitoring systems. Its application is not just limited to vehicle monitoring or surveillance but having an efficient vehicle re-identification procedure allows a system to accurately and timely detect/track a vehicle, which can play an important part in doing forensic analysis as well. The procedure that will be based on deep neural network where a random camera input of vehicle ID will be given to the system and the system will learn to distinguish between different vehicles. Most of the current algorithms solve this problem in fully-supervised manner that require large number of labeled training data. However, it is almost impossible to get large labeled dataset due to high cost. Besides this, in practical scenarios, testing data contains unseen vehicle images on which model is not trained. So, a more robust model is required to handle unseen data. Zero Shot Vehicle Re-Identification, an unsupervised model is proposed to handle unseen data to handle real time data. Two consistencies are proposed to work the model on unseen data, cross view support consistency (CVSC) and cross view projection consistency (CVPC). Let's suppose we have vehicle images of two cameras C_a and C_b. In spite of images important viewpoints distortion and object occlusion, it can be said that visual appearance of images from C_a to C_b will face same illumination changes and blur variation. So, vehicle image from camera C_a can be denoted with images from C_b and vehicle image in camera C_b can be denoted with images in camera C_a. Cross view support consistency says that one image can be represented by other images by sparse coding. So, representation of probe and gallery images are selected and those gallery images are selected whose representatives have maximum overlap with gallery images representatives. The idea behind cross view projection consistency is that probe and gallery image of same vehicle should have more common neighborhoods than probe and gallery image of different vehicles. The neighborhood of the vehicle images are identified by

calculating Euclidean distance between gallery and probe image. KNN of gallery and probe images are selected and the gallery images who have more overlapping neighborhoods with neighborhoods of probe image have stronger projection consistency with the probe image. The neighborhoods of image of camera C_a is directly calculated by taking Euclidean distance, but for neighborhoods of images of C_b, first images of C_b and basic reference subset in C_a is projected to virtual camera C_v then distance is calculated by the learnt metric.

Table of Contents

DECLARATION I

LANGUAGE CORRECTNESS CERTIFICATE II

COPYRIGHT STATEMENT III

ACKNOWLEDGEMENTS IV

ABSTRACT VI

TABLE OF CONTENTS VIII

LIST OF FIGURES IX

LIST OF TABLES X

CHAPTER 1: INTRODUCTION **11**

1.1 Background, Scope and Motivation 12

1.2 Aims and Objectives 13

1.3 Structure of Thesis 13

CHAPTER 2: RELATED WORK **14**

2.1 Part-regularized Near-duplicate Vehicle Re-identification 14

2.2 Joint Semi-supervised Learning and Re-ranking for Vehicle Re-identification 18

2.3 Vehicle Re-identification using PROgressive Unsupervised Deep architecture 21

2.4 Viewpoint-aware Attentive Multi-view Inference for Vehicle Re-identification 23

CHAPTER 3: METHODOLOGY **25**

3.1 Cross View Support Consistency 26

3.2 Cross View Projection Consistency 29

3.3 Data Driven Distance Metric 31

3.3.1 Cross View Support Factor *32*

3.3.2 Cross View Projection Factor *34*

CHAPTER 4: EXPERIMENTATION **36**

4.1 Images Classification 36

4.2 Features Extraction 37

4.3 Implementation 37

4.3.1 Dataset *38*

4.3.2 Results *38*

CHAPTER 5: CONCLUSIONS & FUTURE WORK **48**

REFERENCES **50**

List of Figures

Figure 1-Parts of Vehicles with bounding box 14
Figure 2-Discriminative parts of vehicles 15
Figure 3-Pipeline of part regularize framework 16
Figure 4-Part definition of the model 16
Figure 5- Training stage of proposed model 18
Figure 6- Testing stage of proposed model 19
Figure 7- Architecture of VR-PROUD 21
Figure 8- VR-PROUD cluster convergence 22
Figure 9- VAMI proposed model 23
Figure 10- Cross view support consistency 28
Figure 11- Results of Cross view support consistency 28
Figure 12- Cross view projection consistency 30
Figure 13- Results of cross view projection consistency 31
Figure 14- illustration of exploiting support and projection consistency 33
Figure 15- Front and back vehicle images before classification 36
Figure 16- Front and rear vehicle images after classification 37
Figure 17- Results on VehicleID dataset at 25K feature vector 39
Figure 18- Results on VehicleID dataset at 50K feature vector 39
Figure 19- Results on VehicleID dataset at 75K feature vector 40
Figure 20- Results on VehicleID dataset at 100K feature vector 41
Figure 21- Results on VeRi-776 dataset at 25K feature vector 42
Figure 22- Results on VeRi-776 dataset at 50K feature vector 43
Figure 23- Results on VeRi-776 dataset at 75K feature vector 44
Figure 24- Results on VeRi-776 dataset at 100K feature vector 45

List of Tables

Table 1- Accuracy table of part regularize model 17
Table 2- Accuracy table of joint semi-supervised and re-ranking model 20
Table 3- Accuracy table of VR-PROUD model 22
Table 4- Accuracy table of VAMI model 24
Table 5- Accuracy table of our proposed model 45
Table 6- Comparison of all the models 47

CHAPTER 1: INTRODUCTION

Due to increasing no. of vehicles and expanding traffic, traffic flow network as well as security is main concern, Manual surveillance and traffic monitoring is almost impossible. So vehicle re-identification is step toward the automated surveillance and traffic monitoring system. Its main objective is to identify vehicle in other cameras. Moreover, build a vehicle Re ID system that has the ability to accurately and timely detect/track a vehicle for surveillance and traffic monitoring. Vehicle re-identification is very useful in intelligent traffic monitoring systems. Its main application is vehicle monitoring and surveillance. Beyond this, Having an efficient vehicle re-identification procedure allows a system to accurately and timely detect/track a vehicle, which can play an important part in doing forensic analysis as well as surveillance or traffic monitoring. Most of the current algorithms solve this problem fully-supervised manner that require large number of labeled training data. However, it is almost impossible to get large labeled dataset due to high cost. Besides this, in practical scenarios, testing data contains unseen vehicle images on which model is not trained. So, a more robust model is required to handle unseen data. Zero Shot Vehicle Re-Identification, an unsupervised model is proposed to handle unseen data for real time data. Two consistencies are proposed to work the model on unseen data, CVSC and CVPC. Let's suppose we have vehicle images of two cameras C_a and C_b. In spite of images important viewpoints distortion and object occlusion, it can be said that visual appearance of images from C_a to C_b will face same illumination changes and blur variation. So, vehicle image from camera C_a can be denoted with images from C_b and vehicle image in camera C_b can be denoted with images in camera C_a. Cross view support consistency says that one image can be represented by other images by sparse coding. So, representation of probe and gallery images are selected and those gallery images are selected whose

representatives have maximum overlap with gallery images representatives. The idea behind cross view projection consistency is that probe and gallery image of same vehicle should have more common neighborhoods than probe and gallery image of different vehicles. The neighborhood of the vehicle images are identified by calculating Euclidean distance between probe and gallery image. K nearest neighbors of probe and gallery images are selected and the gallery images who have more overlapping neighborhoods with neighborhoods of probe image have stronger projection consistency with the probe image. The neighborhoods of image of camera C_a is directly calculated by taking Euclidean distance, but for neighborhoods of image of C_b, first images of C_b and basic reference subset in C_a is projected to virtual camera C_v then distance is calculated by the learnt metric.

1.1 Background, Scope and Motivation

Due to increasing no. of vehicles and expanding traffic, traffic flow network as well as security is main concern, Manual surveillance and traffic monitoring is almost impossible. So vehicle re-identification is step toward the automated surveillance and traffic monitoring system. Its main objective is to identify vehicle in other cameras. Moreover, build a vehicle Re ID system that has the ability to accurately and timely detect/track a vehicle for surveillance and traffic monitoring. Vehicle re-identification is very useful in intelligent traffic monitoring systems. Its main application is vehicle monitoring and surveillance. Beyond this, Having an efficient vehicle re-identification procedure allows a system to accurately and timely detect/track a vehicle, which can play an important part in doing forensic analysis as well as surveillance or traffic monitoring. Most of the current algorithms solve this problem fully-supervised manner that require large number of labeled training data. However, it is almost impossible to get large labeled dataset due to high cost. Besides this, in practical scenarios, testing data contains unseen vehicle images on which

model is not trained. So, a more robust model is required to handle unseen data. That is why zero shot vehicle re-identification is proposed.

1.2 Aims and Objectives

Vehicle re-identification is very useful in intelligent traffic monitoring systems. Its main application is vehicle monitoring and surveillance. Beyond this, Having an efficient vehicle re-identification procedure allows a system to accurately and timely detect/track a vehicle, which can play an important part in doing forensic analysis as well as surveillance or traffic monitoring. In current security situation of Pakistan, surveillance and traffic monitoring is national need to detect, track and prevent terrorist activities. It is also beneficial from Counter Intelligence (CI) and Counter Terrorism (CT) angle.

1.3 Structure of Thesis

The thesis is structured as follows:

Chapter 2 gives review of the literature and the significant work done by researchers in past few years on the Vehicle Re-Identification.
Chapter 3 discusses proposed methodology in detail.
Chapter 4 discusses the experimental results in detail with all desired figures and tables.
Chapter 5 concludes the thesis and reveals future scope of this research.

CHAPTER 2: RELATED WORK

This chapter discusses the related work done in vehicle re-identification problem. Here is the latest work done in this problem domain.

2.1 Part-regularized Near-duplicate Vehicle Re-identification

Bing et al. [1] proposed the method of vehicle re-identification using local parts of the vehicles. There are thousands type of vehicles exist in the market. It is really difficult to distinguish these vehicles. So, in this paper [1], local discriminative parts of the vehicles are used to better distinguish the vehicles.

Figure 1-Parts of Vehicles with bounding box [1]

Figure 2-Discriminative parts of vehicles [1]

For example, in figure-1, in the first row, vehicles are difficult to distinguish despite that vehicles belong to different identities of same vehicle model. In figure-2, vehicles can be easily distinguished by using discriminative parts of the vehicles.

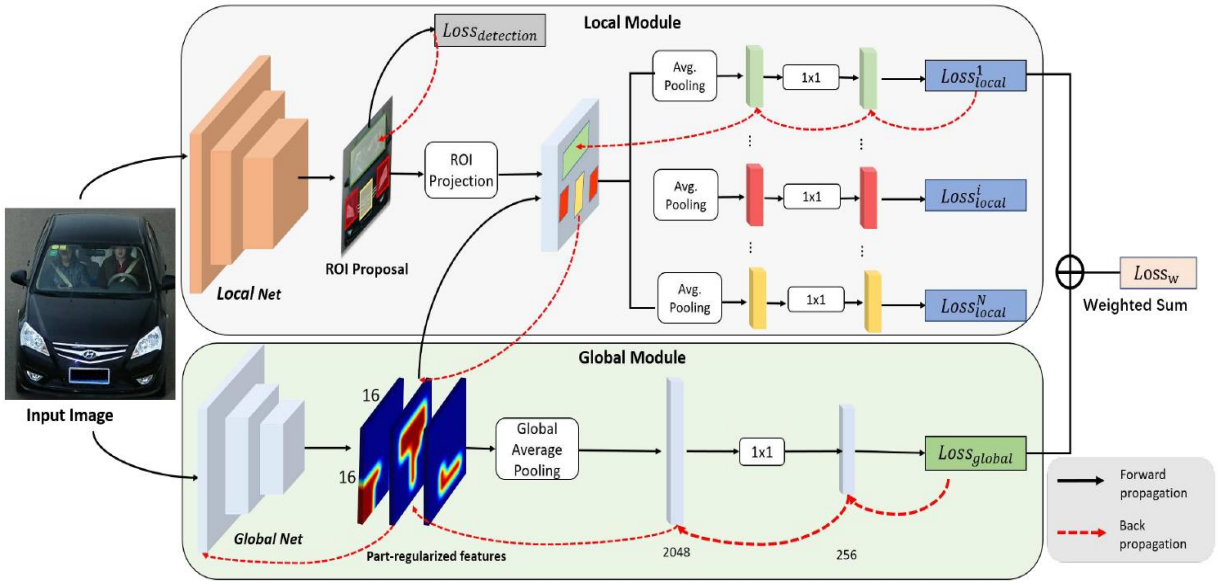

Figure 3-Pipeline of part regularize framework [1]

Figure 4-Part definition of the model [1]

The framework is divided into 2 parts, local module to correct classification and global module to Re-Id categorization. Local module does the vehicle part regularization. In part regularization process, three parts of the vehicles are selected, front and back lights, vehicle brand, and front and back window. As vehicle's front and back lights, and brand are the most discriminant part of the vehicles. A tight bounding box of the light is defined and extended it to bottom of the vehicle as sown in figure-4. In next step, location of parts of the vehicles are identified. Input image is fed to Local Net (YOLO in experiments) and part detection results as output are received. If occlusion is presence in the image, then next image of the same vehicle is given as input to Local Net. In global module, features extracted through ResNet-50. Experiments were conducted on datasets VeRi-776 and VehicleID. Here are the accuracy table of both datasets.

	CMC@Rank 1 in %	CMC@Rank 5 in %
VehicleID	78.4	92.3
VeRi-776	87.8	95.2

Table 1- Accuracy table of part regularize model [1]

2.2 Joint Semi-supervised Learning and Re-ranking for Vehicle Re-identification

This research paper focuses on a method of semi-supervised learning. The idea behind semi supervised learning is that in real world problems, dataset finding is the major problem. It is difficult to obtain data for training process. Even if small dataset is available, the dataset is not sufficient enough to train the algorithm. So, this research paper proposed a joint semi-supervised learning and re-ranking method.

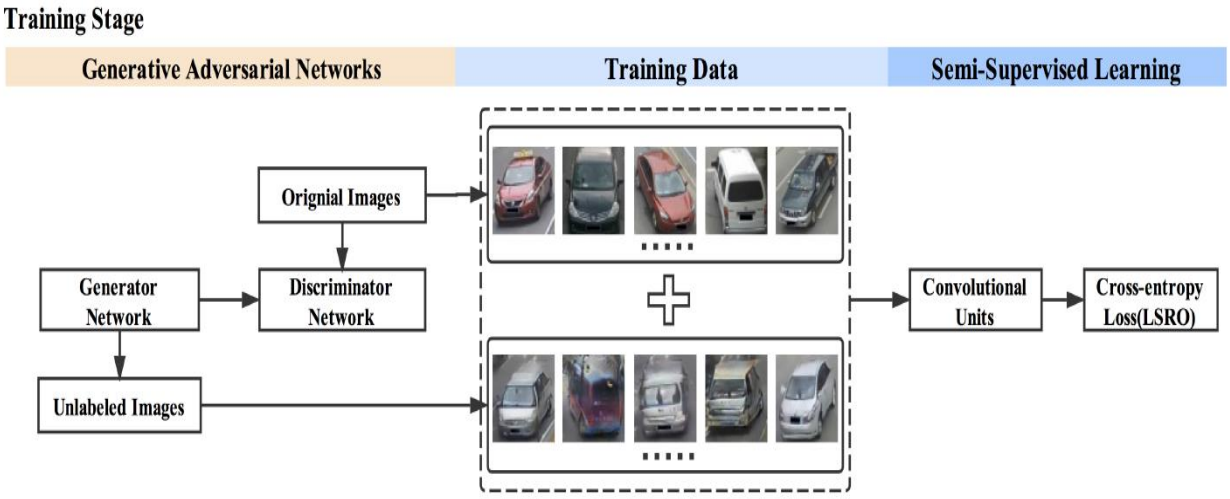

Figure 5- Training stage of proposed model [102]

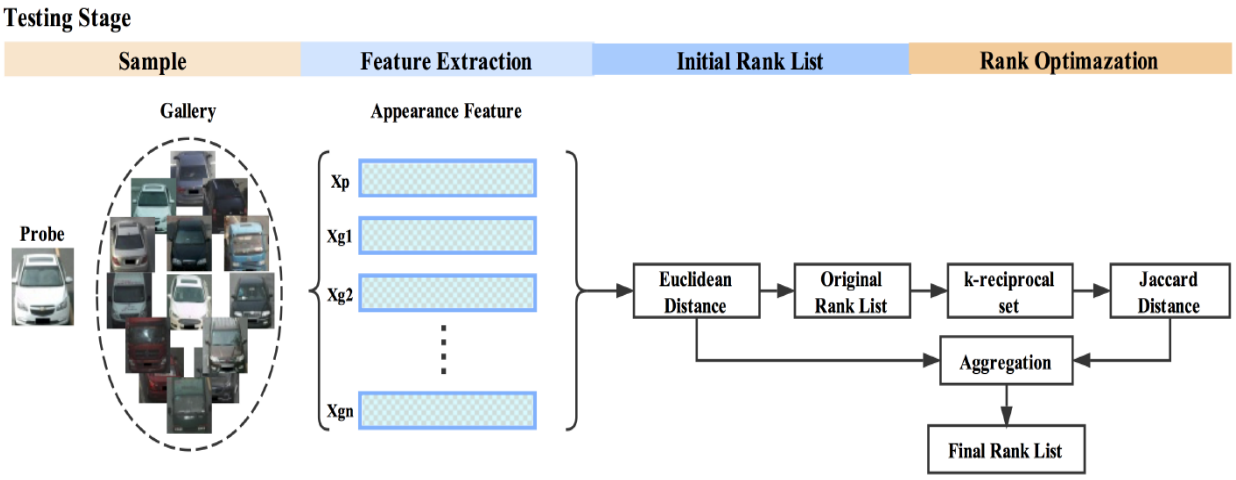

Figure 6- Testing stage of proposed model [102]

In training phase, Features are extracted from training images and using these features new images are generated using generative adversarial network. Existing and new generated images are combined. Then LSRO is applied to incorporate the unlabeled images in the network. Using LSRO, outliers are dealt efficiently.

In re-ranking method, initial rank list $L(p, G) = \{g1, g2, \ldots, gN\}$ is obtained by taking Euclidean distance of probe image with gallery images g_i. Using K-NN, top K vehicles images are selected $N(p, k) = \{g1, g2, \ldots, gN\}$. The goal is to re-rank the initial rank list to get more positive images in the ranking list to improve performance of the model. For this purpose, K-reciprocal nearest neighbors are defined $R(p,k)=\{g_i|(g_i) \in N(p,k) \Lambda (p \in N(g_i,k))\}$. Due to occlusions, illuminations and variations in views, the k-nearest neighbors and the k-reciprocal nearest neighbors may not include the positive images. To cater this problem, 1/4 k-reciprocal nearest neighbors of each vehicle in R(p, k) added incrementally into a more robust set $R^{*}\leftarrow(p, k)= \{g_i|(g_i) \in N(p,k) \Lambda (p \in N(g_i,k))\}$. $R\leftarrow(p, k)$ as

contextual knowledge is considered to re-calculate the distance between the appearance features of the probe and the gallery. The pairwise distance between the probe p and the gallery gi will be re-calculated by comparing their k-reciprocal nearest neighbor sets. we believe that if two images are similar, their k-reciprocal nearest neighbor sets overlap, i.e., there are some duplicate samples in the sets. The more duplicate samples, the more similar the two images are. Model was tested on VeRi-776 and VehicleID datasets. Here is the accuracy graph on these datasets.

	CMC@Rank 1 in %	CMC@Rank 5 in %
VehicleID	83.3	85.9
VeRi-776	87.8	94.2

Table 2- Accuracy table of joint semi-supervised and re-ranking model [102]

2.3 Vehicle Re-identification using PROgressive Unsupervised Deep architecture

This research paper proposed progressive unsupervised deep architecture for vehicle re-identification. Dataset labelling is a major issue in training process of machine learning models. This framework uses clustering method to train the model.

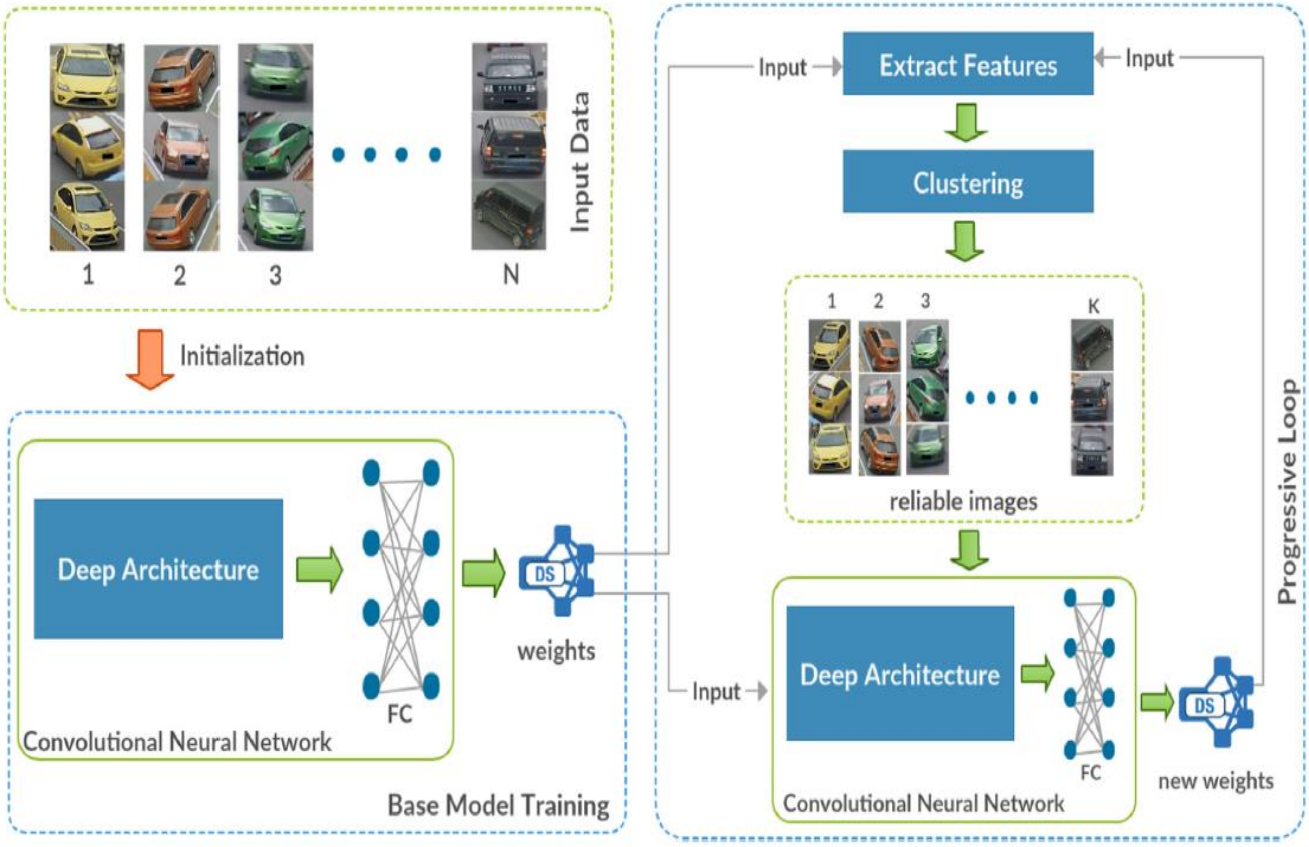

Figure 7- Architecture of VR-PROUD [103]

In the first step, image features are extracted using base deep pre-trained CNN model which is used in next subsequent step. ResNet-50 (with last layer replaced) trained on ImageNet is used to extract features. Then in the next step, clustering is done on extracted features. Using K-means, features are clustered and cluster ids are assigned as pseudo labels to the clustered vehicles for next iteration. Clustering results are

further refined to obtain stable and accurate clusters by enforcing certain heuristic constraint. These robust clusters (representing vehicles) are then utilized to fine-tune another CNN network having the same architecture as the base CNN. The process is iteratively performed where the training sample set grows in every iteration with increasingly robust clusters to enable unsupervised self-progressive learning till convergence.

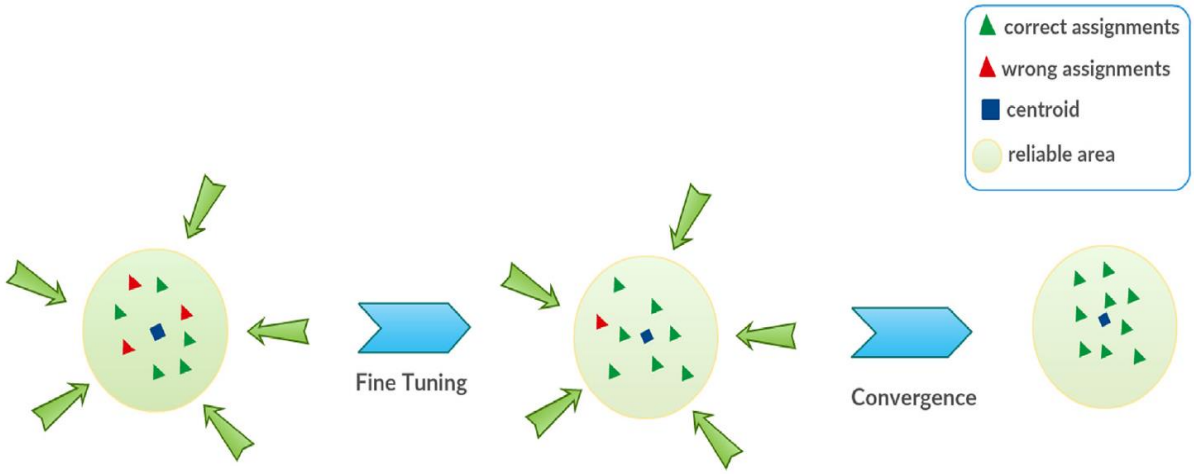

Figure 8- VR-PROUD cluster convergence [103]

Here is the accuracy of the above model on VehicleID and VeRi datasets.

	CMC@Rank 1 in %	CMC@Rank 5 in %
VehicleID	71.4	81.5
VeRi-776	82.8	90.4

Table 3- Accuracy table of VR-PROUD model [103]

2.4 Viewpoint-aware Attentive Multi-view Inference for Vehicle Re-identification

This paper proposed view point aware attentive multi view inference model that needs just visual information of the vehicles to solve multi view vehicle re-id problem. As, previous models need information of spatial temporal information to get fine results. For a vehicle image of arbitrary viewpoints, the model extracts the single view feature for each vehicle image and aims to transform the features into a global multi view feature representation so that pairwise distance metric learning can be better optimized in such a viewpoint invariant feature space. The VAMI adopts a viewpoint aware attention model to select core regions at different viewpoints and implement effective multi-view feature inference by an adversarial training architecture as shown in Figure-9.

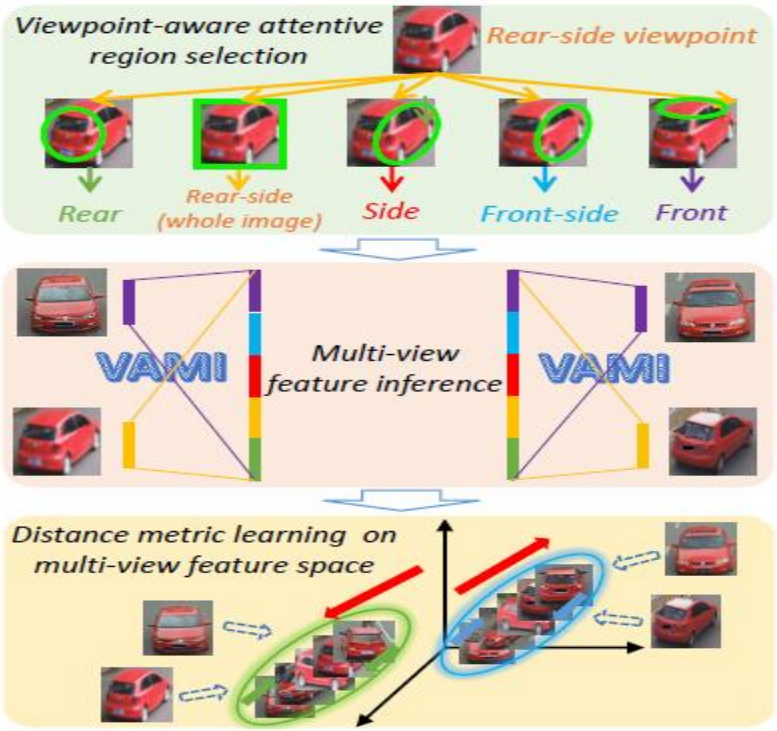

Figure 9- VAMI proposed model [104]

Here is the accuracy of the above model on VehicleID and VeRi datasets.

	CMC@Rank 1 in %	CMC@Rank 5 in %
VehicleID	52.1	77.4
VeRi-776	77.6	91.2

Table 4- Accuracy table of VAMI model [104]

CHAPTER 3: METHODOLOGY

This chapter discusses the methodology used in the vehicle re-identification using zero shot learning. Consider a pair of Cameras C_a and C_b with different views. A set of labeled Vehicles $V=\{v_1,v_2, v_3,\ldots\ldots v_m\}$ are associated with the pair of camera C_a and C_b where m is number of vehicles. Let's represent image of vehicle v_i captured by C_a by x_a^i and vehicle v_i captured by C_b by x_b^i, $x_a^i, x_b^i \in R^d$. Let's denote set of vehicle images captured by camera C_a with $X_{a,L}=\{x_a^1,\ldots..x_a^i,\ldots..x_a^m\}$ and camera C_b with $X_{b,L}=\{x_b^1,\ldots..x_b^j,\ldots..x_b^m\}$ where $1 \leq i,j \leq m$ and i=j means the same vehicle captured by C_a and C_b.

Let x_a^p represent testing data captured by C_a and $X_{b,U}=\{x_b^{m+1},\ldots..,x_b^q,\ldots..x_b^{m+n}\}$ where $m+1 \leq q \leq m+n$ represent the gallery of testing data captured by C_b and n is number of testing vehicle images captured by C_b. So, training and testing data is totally different. Our goal is to rank testing data captured by C_a denoted by $X_{a.U}$ in $X_{b,U}$ using zero shot learning algorithm.

A traditional supervised learning methods learnt metric learning methods using pair of vehicle images $X_{a,L}$ and $X_{b,L}$. Given a pair of images x_a^i and x_b^j, their Mahalanobis distance can be calculated by below equation

$$d_M(x_a^i, x_b^j) = \left(x_a^i - x_b^j\right)^T M(x_a^i - x_b^j)$$

where M is a positive semi definite matrix for metric validity. After metric learning between training images of $X_{a,L}$ and $X_{b,U}$, testing data of $X_{a,L}$ can be ranked in $X_{b,L}$ using following metric learning method

$$d_M(x_a^i, x_b^j) = \left(x_a^i - x_b^j\right)^T M(x_a^i - x_b^j) \qquad (1)$$

Matrix M is further decomposed with $M=L^TL$. Then equation (1) can be re-written as

$$d_M(x_a^i, x_b^j) = (x_a^i - x_b^j)^T L^T L(x_a^i - x_b^j) = ||L.x_a^p - L.x_b^p||^2.$$ By this definition, metric gives projection matrix. This matrix is used to obtain new feature space from original image features. Dimensions are not correlated in new feature space. This is equal to changing images of C_a/C_b to virtual camera C_v and computing distances with Euclidean distance. So, Mahalanobis distance is applied to vehicle images of different cameras and Euclidean distance is applied to vehicle images of same camera. So, traditional metric learning method is used to rank the matched images in image gallery but this metric learning methods has certain limitations. Metric learning models overfit on training data and can only rank seen data and cannot rank unseen data. So, a more robust method is required to deal with unseen data.

To deal with the above mentioned problem, a new method is proposed to rank unseen images. Cross view projection consistency and cross view support consistency is proposed to generalize the model on seen and unseen data. Images of same vehicles have stronger cross view projection and support consistencies than images with different vehicle.

3.1 Cross View Support Consistency

In spite of images important viewpoints distortion and object occlusion, it can be said that visual appearance of images from C_a to C_b will face identical illumination and blur variation. For example, vehicle image in camera C_a is represented by I_a and

vehicle image in camera C_b is represented by I_b, and illumination changes V and blur change B exist between images of same vehicle from C_a to C_b. So, images of same vehicle from C_a and C_b can be represented as

$$I_b = I_a * V * B.$$

I_a can further be represented by three other images from C_a like

$$I_a = w_1 {I_a}^1 + w_2 {I_a}^2 + w_3 {I_a}^3$$

Then I_b will be

$$I_b = w_1 {I_a}^1.V.B + w_2 {I_a}^2.V.B + w_3 {I_a}^3.V.B$$

Here ${I_a}^1$.V.B, ${I_a}^2$.V.B, ${I_a}^3$.V.B are the transformation of selected images ${I_b}^1$, ${I_b}^2$, ${I_b}^3$ in C_b. So, from above we can say that vehicle image from camera C_a can be denoted by images from C_b and vehicle image in camera C_b can be denoted by images from camera C_a.

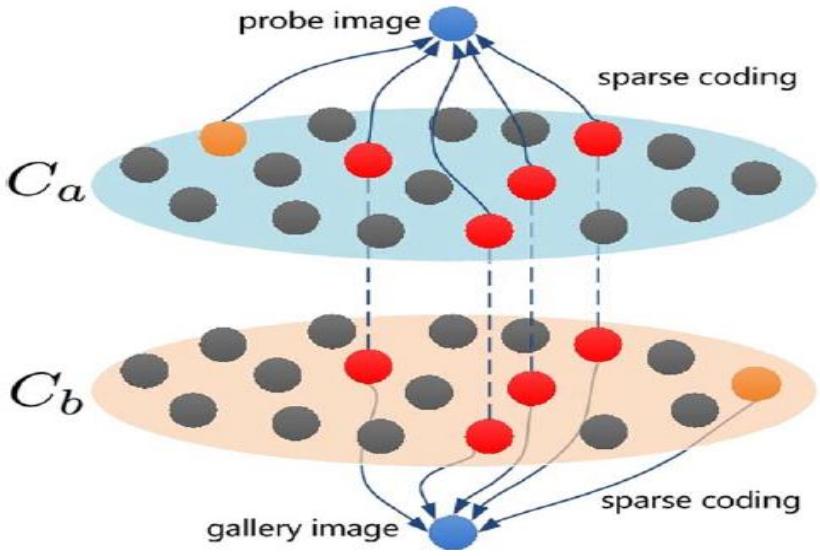

Figure 10- Cross view support consistency [105]

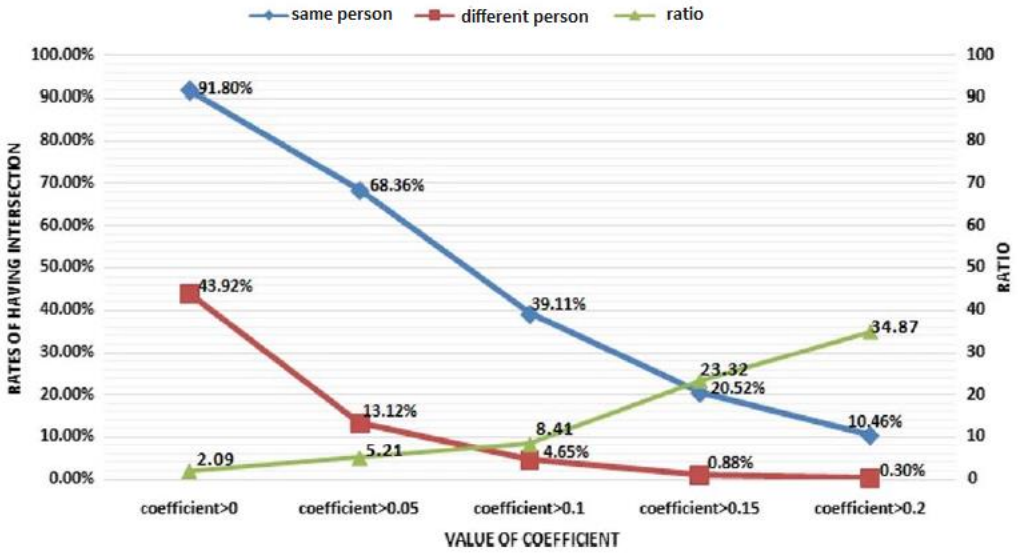

Figure 11- Results of Cross view support consistency

Support set of image are selected as sparsely selected images which together represent the image as shown in Fig (10). The idea behind is that two images with same vehicles should have more support sets in common than two images with different vehicles. So, support sets of probe image and gallery images are identified and intersection is taken of the support sets of probe and gallery images. In Fig (10), dashed lines that connect the red balls are showing the similar support sets of probe and gallery image. The gallery images whose support sets have maximum intersection with support set of probe image are closer to probe image. In order to achieve this, two dictionaries are maintained for images of camera C_a and C_b. Sparse coding method [2] is used to learn representation coefficients. The non-zero entries are the vehicle images selected that compose support set. It was seen in (fig 11) that the ratio of non-zero entries are much higher for the same vehicles than with the non-zero entries for the different vehicles. Ratio of value of same vehicles to that of different vehicles are much greater at higher sparse representation coefficients which represent more strong support set as shown in Fig (12). So, same vehicles has strong cross view support consistency.

3.2 Cross View Projection Consistency

The idea behind CVPC is that probe and gallery image of same vehicle should have more common neighborhoods than probe and gallery image of different vehicles. The neighborhood of the vehicle images are identified by calculating Euclidean distance between gallery and probe image. K nearest neighbors of gallery and probe images are selected and the gallery images who have more overlapping

neighborhoods with neighborhoods of probe image have stronger projection consistency with the probe image.

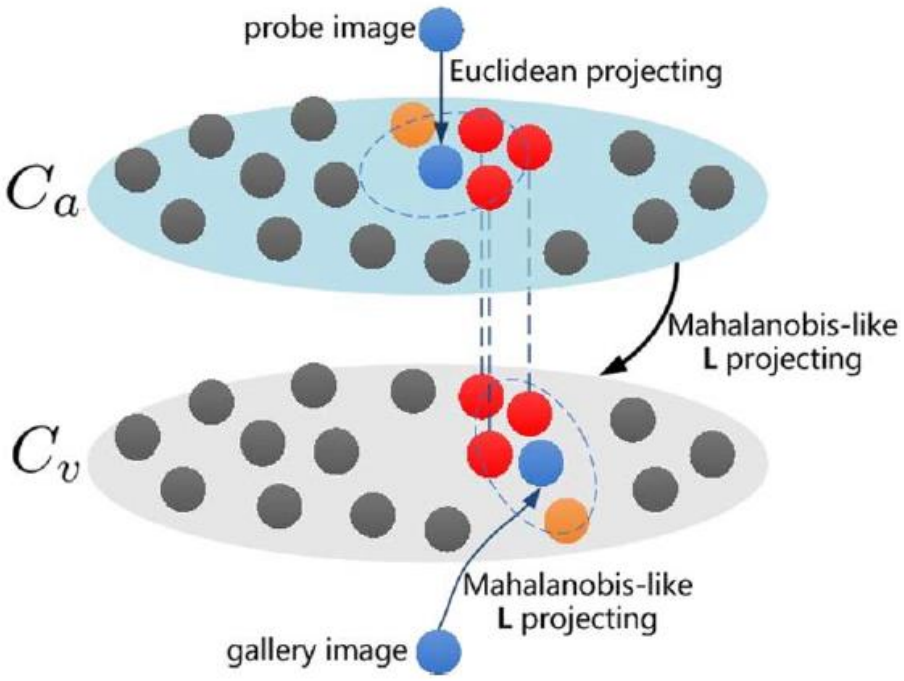

Figure 12- Cross view projection consistency [105]

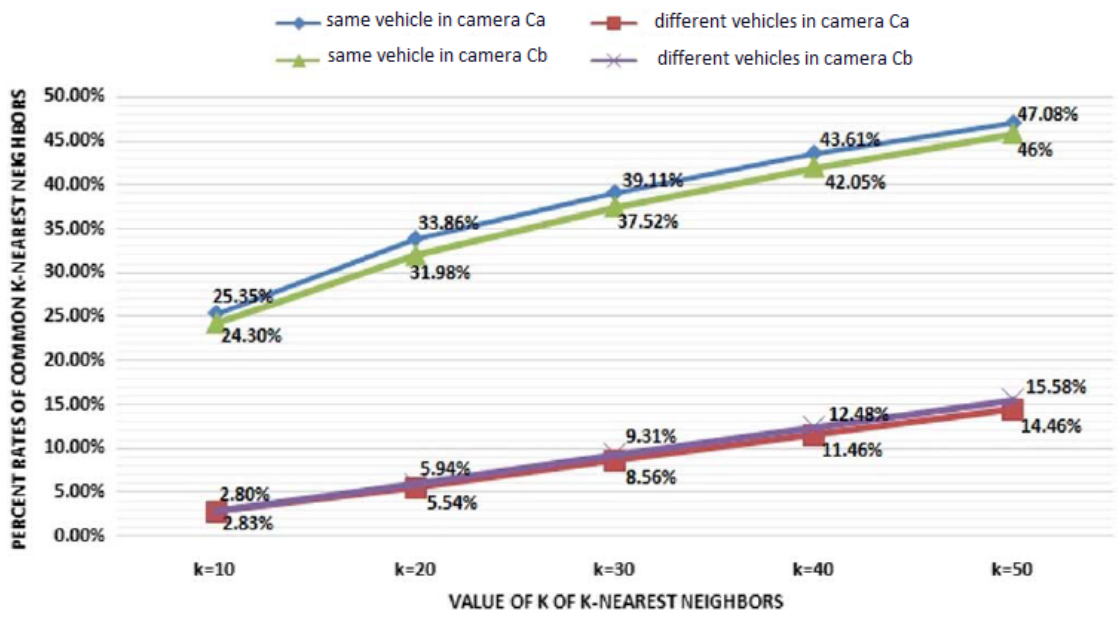

Figure 13- Results of cross view projection consistency

The neighborhoods of image of camera C_a is directly calculated by taking Euclidean distance, but for neighborhoods of image of C_b, first images of C_b and basic reference subset in C_a is projected to virtual camera C_v then distance is calculated by the learnt metric (KSS metric learning[3]) as shown in Figure-12. It was seen that images of same vehicles have more similar neighborhoods than images of different vehicles as shown in Figure-13. So, it can be said that cross view projection consistency is stronger for same vehicle images.

3.3 Data Driven Distance Metric

A data driven distance metric is proposed to improve original uniform metric method. By metric learning method, to adjust the uniform metric M, adaptive factors are learnt. For vehicle image pair x_a^p and x_b^q, our aim is to obtain M_{pq} to obtain the data specific distance.

$$d_{M_{pq}}(x_a{}^p, x_b{}^q) = (x_a{}^p - x_b{}^q)^T M_{pq}(x_a{}^p - x_b{}^q) \qquad (2)$$

Here the data specific distance is obtained by calculating adaptive factors. There are two adaptive factors, CVSAF f_s and CVPAF f_p. So, adaptive metric is obtained by equ (3).

$$M_{pq} = f_s(x_a{}^p, x_b{}^q).f_p(x_a{}^p, x_b{}^q).M \qquad (3)$$

3.3.1 Cross View Support Factor

In CVSC, for probe image $x_a{}^p$ in C_a, we select the images in $X_{a,L}$ by sparse coding to represent $x_a{}^p$. Similarly for gallery image $x_b{}^q$, we select the images in $X_{b,L}$ by sparse coding to represent $x_b{}^q$. We generate the dictionary $D_a=[x_a{}^1,.....x_a{}^i,.....x_a{}^m]$, $D_a \in R^{d*m}$ and then from dictionary, select sparsely to represent $x_a{}^p=D_a w_a{}^p*$ where (.)* stands for solution and $w_a{}^p* \in R^{m*1}$ shows the selected images with the coefficients. If image coefficient is greater than zero, it shows that image is chosen as support data for $x_a{}^p$. Sparse representation $w_a{}^p*$ is computed by the equ (4)

$$w_a{}^{p*} = \arg\min_{w_a{}^p \geq 0} \frac{1}{2}||D_a w_a{}^p - x_a{}^p||_2^2 + \frac{\rho}{2}||w_a{}^p||_2^2 + \lambda||w_a{}^p||1 \qquad (4)$$

Similarly, we generate dictionary

$$D_b = [x_b{}^1, \ldots . x_b{}^i, \ldots \ldots . x_b{}^m],\ D_b \in R^{d*m}$$

And then from dictionary, select sparsely to depict $x_b{}^q=D_b w_b{}^q*$ where (.)* stands for solution and $w_b{}^q* \in R^{m*1}$ shows the selected images with the coefficients. If image coefficient is greater than zero, it indicates that image is selected as support set for $x_b{}^q$. Sparse representation $w_b{}^q*$ is computed by the equ (5)

$$w_b{}^{q*} = \arg\min_{w_b{}^q \geq 0} \frac{1}{2}||D_b w_b{}^q - x_b{}^q||_2^2 + \frac{\rho}{2}||w_b{}^q||_2^2 + \lambda||w_b{}^q||1 \qquad (5)$$

Then cross view consistency is defined by equ (6)

$$sc(x_a{}^p, x_b{}^q) = size(w_a{}^{p*} .* w_b{}^{q*}) \qquad (6)$$

where $.*$ indicates element wise multiplication. And *size* gives count of non-zero elements in the vector.

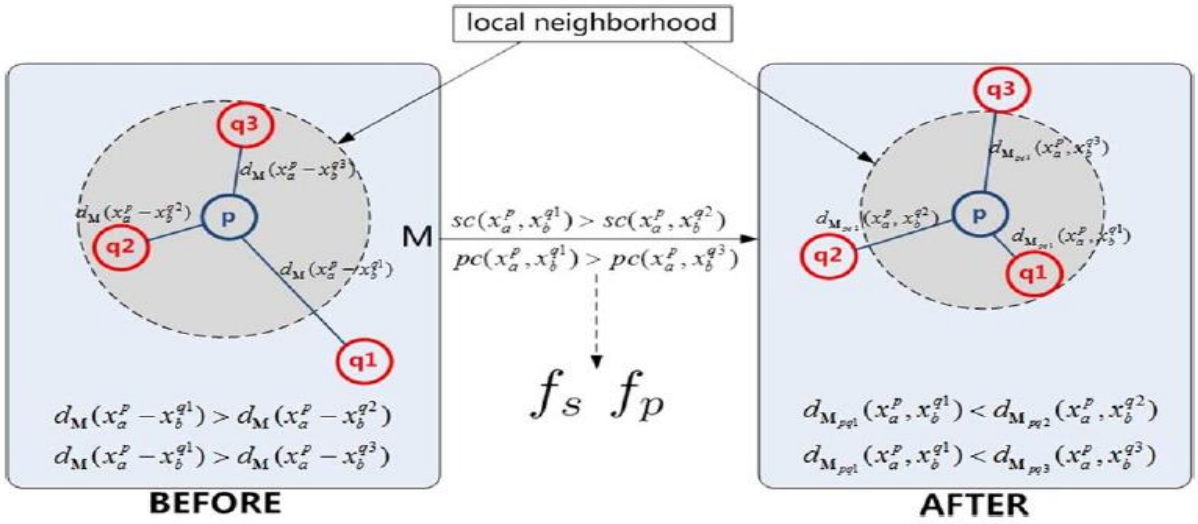

Figure 14- illustration of exploiting support and projection consistency [105]

Support adaptive factor is constructed with cross view support consistency. It is shown in Figure-14, cross view support consistency is stronger between p and q1 as compared to q2 and q3 and after metric adaption, q1 is closer to p as compared to q2 and q3. So, if CVSC is stronger, value of cross view support factor f_s will be smaller as shown in Equ (7).

$$f_s(x_a{}^p, x_b{}^q) = \left[\frac{1}{1+sc(x_a{}^p, x_b{}^q)}\right]^{\alpha}$$

$$= \left[\frac{1}{1+size(w_a{}^{p*} .* w_b{}^{q*})}\right]^{\alpha} , \quad \alpha > 0 \qquad (7)$$

Here α indicates contribution of cross view support consistency, greater the value of α, greater the impact of metric adaption.

3.3.2 Cross View Projection Factor

As already discussed above, the idea behind cross view projection consistency is that probe and gallery image will have stronger cross view projection consistency if they have similar context. The context of images means images have similar neighborhoods across the cameras. For the probe image $x_a{}^p$, the neighborhoods in C_a are calculated directly by taking Euclidean distance while the neighborhoods in C_b is calculated by first projecting images of C_b in virtual camera C_v and then Euclidean distance is calculated. The distance between $x_a{}^p$ and $x_b{}^j$ is calculated by $d_M(x_a{}^p,x_b{}^j)=(x_a{}^p-x_b{}^j)^T M(x_b{}^q-x_b{}^j)$. Then distances are ranked and K nearest neighbors of $x_a{}^p$ and $x_b{}^q$ are selected. The set of KNN of $x_a{}^p$ in $X_{b,L}$ is $knn(x_b{}^q|X_{b,L})$. Then projection consistency in C_b is defined as

$$pc_b(x_a{}^p, x_b{}^q) = |knn(x_a{}^p|X_{b,L}) \cap knn(x_b{}^q|X_{b,L})| \qquad (8)$$

Where $|knn(x_a{}^p|X_{b,L}) \cap knn(x_b{}^q|X_{b,L})|$ represents count of common k nearest neighbors of $x_a{}^p$ and $x_b{}^q$ in $X_{b,L}$.

Similarly, The set of KNN of $x_a{}^p$ in $X_{a,L}$ is $knn(x_b{}^q|X_{a,L})$. Then the projection consistency in C_a is defined as

$$pc_a(x_a^p, x_b^q) = |knn(x_a^p|X_{a,L}) \cap knn(x_b^q|X_{a,L})| \quad (9)$$

where $|knn(x_a^p|X_{a,L}) \cap knn(x_b^q|X_{a,L})|$ represents the number of common k nearest neighbors of x_a^p and x_b^q in $X_{a,L}$. Then the total projection consistency is calculated by following Equ (10).

$$pc(x_a^p, x_b^q) = pc_a(x_a^p, x_b^q) + pc_b(x_a^p, x_b^q) \quad (10)$$

As shown in figure. 5, cross view projection consistency is stronger between p and q1 as compared to q2 and q3 and after metric adaption, q1 is closer to p as compared to q2 and q3. So, stronger the CVPC, smaller the value of cross view support factor f_p as shown in Equ (11).

$$f_p(x_a^p, x_b^q) = \left[\frac{1}{1 + pc(x_a^p, x_b^q)}\right]^{\beta}$$

$$= \left[\frac{1}{1+pc_a(x_a^p, x_b^q)+pc_b(x_a^p, x_b^q)}\right]^{\beta}, \beta > 0 \quad (11)$$

Here β indicates contribution of CVPC, greater the value of β, greater the impact of metric adaption will be.

After learning CVSC and CVPC, new metric specific will be obtained for each x_a^p and x_b^q by Equ (3).

CHAPTER 4: EXPERIMENTATION

This chapter discusses the experiment performed in the thesis implementation. Experiments are performed on two datasets VehicleID and Veri-776. Both are widely renowned datasets.

4.1 Images Classification

Both the dataset contained front, back, and side view images. So, front view vehicle images are classified and extracted. Python code in Pycharm IDE with Python 2.7 is run on Linux environment. Here is the example of images before detection of vehicle view.

Figure 15- Front and back vehicle images before classification

Below is the image after detection of vehicle view. As shown in below image, first 3 images are front view vehicles. The last 2 images are rear view vehicles.

Figure 16- Front and rear vehicle images after classification

So, in that way, front view images are classified and placed separately into new directory. Then images are further classified into directories on the basis of camera ids and vehicle ids. For that, all the vehicle Ids of camera 1 are placed into c001 directory and all the vehicle Ids of camera 2 are placed into c002 directory. As each vehicle Id has multiple images. Images of same vehicle Id are placed in one directory. So, finally we have, inside camera Id directories, there are vehicle Id directories with multiple images in it.

4.2 Features Extraction

In the next step, features are extracted for each vehicle Id, as we have multiple images for each vehicle Id, that's why mean of multiple image features of same vehicle Id are taken. Pre trained Resnet-50 is used for features extraction. Feature vectors of size 25K, 50K, 75K, and 100K are extracted. ".m file" is generated and all the extracted features are stored in ".m file". That ".m file" will be used by the application.

4.3 Implementation

This section discusses in detail the standard datasets used in our research and also elaborate implementation details.

4.3.1 Dataset

For the implementation, we used following standard datasets in our research.

4.3.1.1 VeRi-776

VeRi dataset contains 50,000 images of 776 different vehicles captured in one kilometer area by 2-18 cameras, each placed on different locations in a time span of 24 hours.

4.3.1.2 VehicleID

VehicleID dataset contains 3 category images (small, medium, large) vehicles. Contains 2,22,628 images of 26,328 vehicles captured by multiple cameras in real-world traffic surveillance environment.

4.3.2 Results

For the implementation of our research, I used Matlab as IDE. Experiment is divided into 2 parts. Training phase and Testing Phase. In training phase, Training Matrix 'M' is obtained by metric learning of camera 1 and camera 2 images. KSS metric learning method is used for training. Testing were performed on the remaining data. Results are obtained on different features vector of 25K, 50K, 75K, and 100K. On features vector of 100K, model has the highest accuracy. Here are the results of different features vector on VehicleID dataset.

4.3.2.1 With 25K Features

With 25K features vector on VehicleID dataset, model has accuracy 53 and 70 at Rank 1 and Rank 5 respectively.

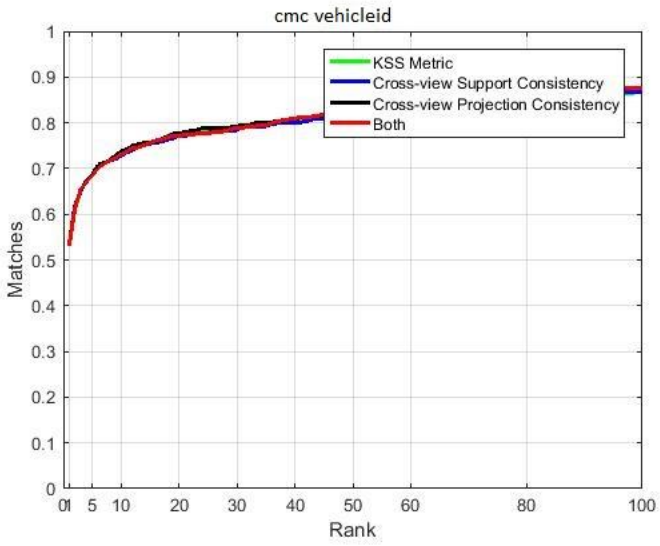

Figure 17- Results on VehicleID dataset at 25K feature vector

4.3.2.2 With 50K Features

With 50K features vector on VehicleID dataset, model has accuracy 70 and 80 at Rank1 and Rank 5 respectively.

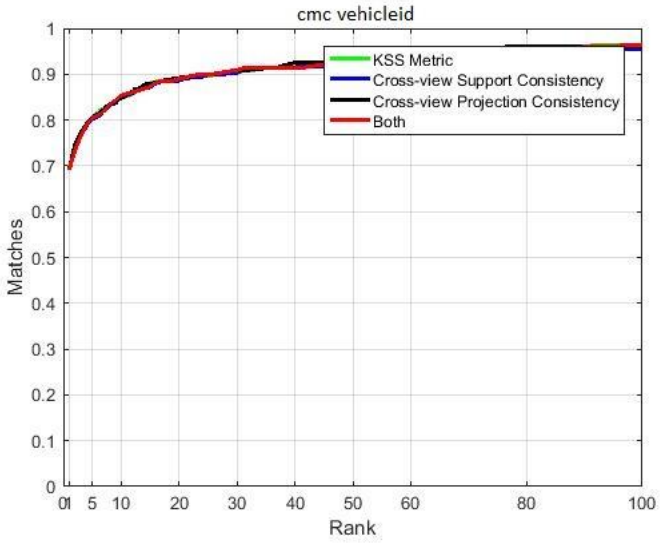

Figure 18- Results on VehicleID dataset at 50K feature vector

4.3.2.3 With 75K Features

With 75K features vector on VehicleID dataset, model has accuracy 80 and 90 at Rank1 and Rank 5 respectively.

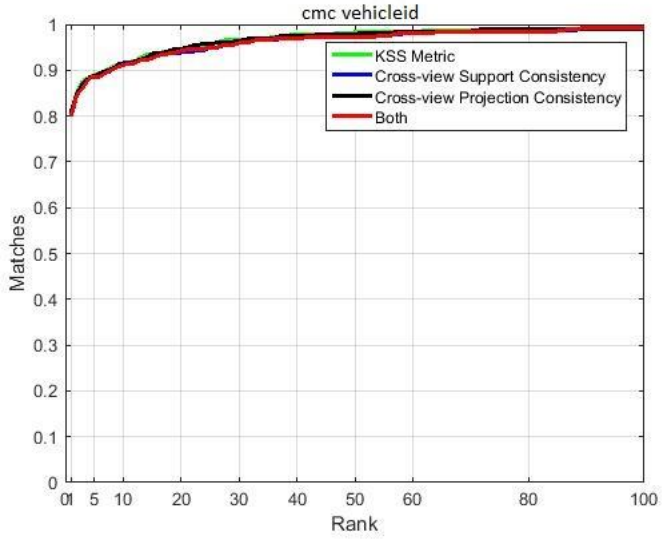

Figure 19- Results on VehicleID dataset at 75K feature vector

4.3.2.4 With 100K Features

With 100K features vector on VehicleID dataset, model has accuracy 84 and 93 at Rank1 and Rank 5 respectively.

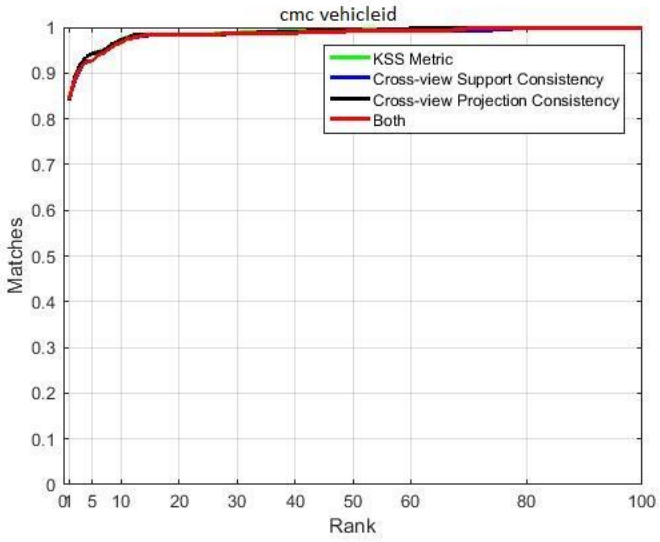

Figure 20- Results on VehicleID dataset at 100K feature vector

Here are the results of different features vector on VeRi-776 dataset.

4.3.2.5 With 25K Features

With 25K features vector on VeRi-776 dataset, model has accuracy 49 and 63 at Rank1 and Rank 5 respectively.

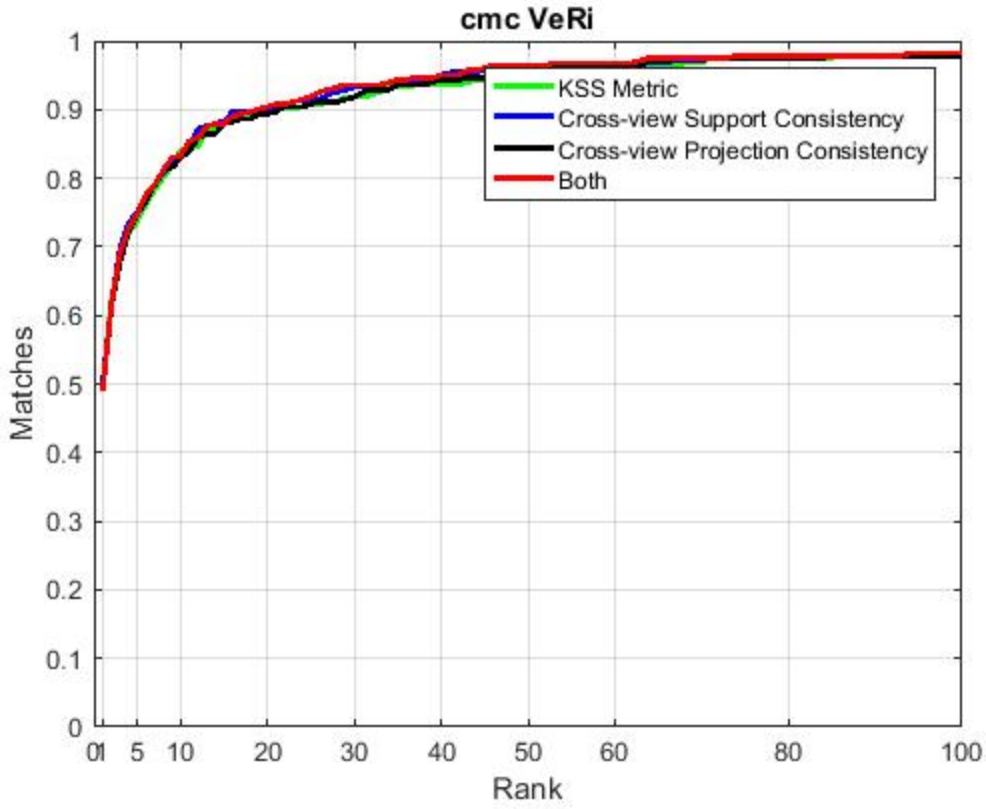

Figure 21- Results on VeRi-776 dataset at 25K feature vector

4.3.2.6 With 50K Features

With 50K features vector on VeRi-776 dataset, model has accuracy 71 and 88 at Rank1 and Rank 5 respectively.

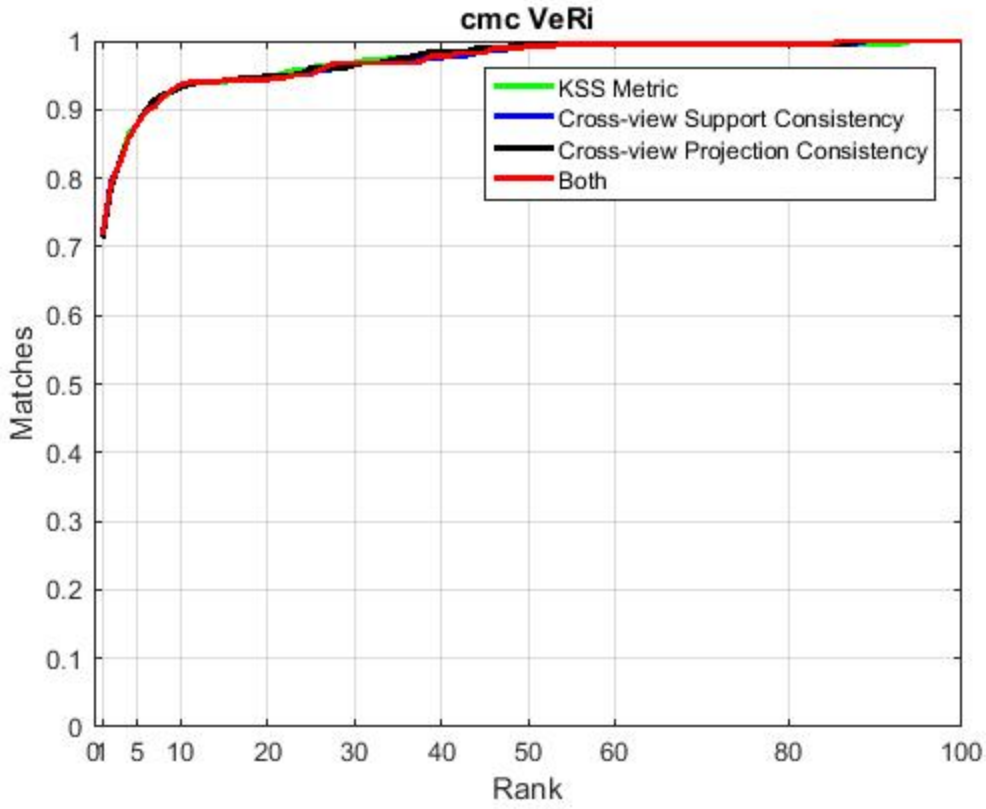

Figure 22- Results on VeRi-776 dataset at 50K feature vector

4.3.2.7 With 75K Features

With 75K features vector on VeRi-776 dataset, model has accuracy 78 and 97 at Rank1 and Rank 5 respectively.

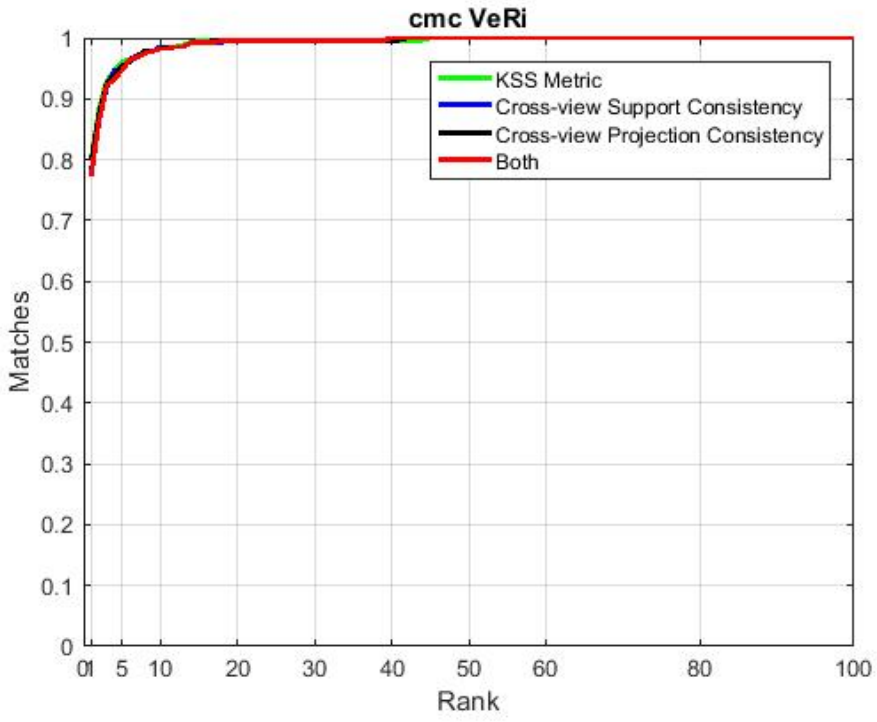

Figure 23- Results on VeRi-776 dataset at 75K feature vector

4.3.2.8 With 100K Features

With 100K features vector on VeRi-776 dataset, model has accuracy 83 and 98 at Rank1 and Rank 5 respectively.

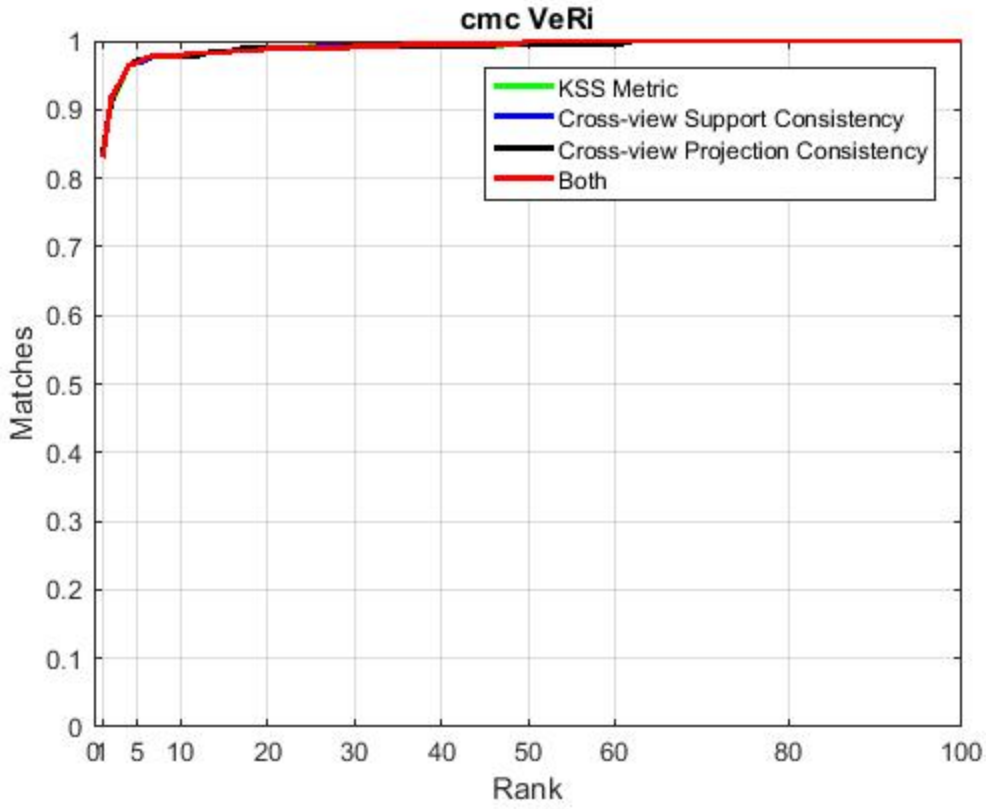

Figure 24- Results on VeRi-776 dataset at 100K feature vector

So, the model gives highest accuracy at 100K features vector. Here is the accuracy table on Veri-776 and VehicleID datasets.

	CMC@Rank 1 in %	CMC@Rank 5 in %
VehicleID	84.2	93.3
VeRi-776	83.2	98.4

Table 5- Accuracy table of our proposed model

In past few years, the solution of Vehicle Re-Identification problem has been proposed by many scientists in terms of supervised, semi-supervised and unsupervised manner. Usually, supervised models have more accuracy than un-supervised models. We proposed zero shot learning solution for Vehicle Re-Identification problem. Our model recognize vehicles on unseen data. Our model has 84% accuracy at Rank-1 and 93% accuracy at Rank-5 on VehicleID dataset and 83% accuracy at Rank-1 and 98% accuracy at Rank-5 on VeRi-776 dataset. Here is the comparison graph with previous models at different ranks.

	VehicleID CMC@Rank 1 in %	VehicleID CMC@Rank 5 in %	VeRi-776 CMC@Rank 1 in %	VeRi-776 CMC@Rank 5 in %
Part Regularize model	78.4	92.3	87.8	95.9
Joint semi supervised and re-ranking model	83.3	85.9	87.8	94.2
VR-PROUD Model	71.4	81.5	82.8	90.4

VAMI Model	52.1	77.4	77.6	91.2
Zero Shot Method (Our Model)	**84.2**	**93.3**	**83.2**	**98.4**

Table 6- Comparison of all the models

CHAPTER 5: CONCLUSIONS & FUTURE WORK

This chapter discusses the conclusions derived in this thesis and new ideas related to this problem for future work. In the thesis, data driven distance metric is calculated using cross view support consistency and cross view projection consistency. The idea behind the model is that learnt metric is refined by re-exploiting the training data in distance calculation stage. Cross view support consistency and cross view projection consistency is applied to calculate data specific adaptive metric that improves accuracy of the model. Zero shot learning solution is applied on datasets VehicleID and VeRi-776. Our zero shot learning model outperforms the existing state of the art models. Our zero shot method's accuracy on dataset VehicleID is 84 at Rank-1 and 93 at Rank-5 and on dataset VeRi-776 is 83 at Rank-1 and 98 at Rank-5. None of the model has that much accuracy on unsupervised and zero shot learning model.

Here are the few ideas for future work where accuracy can be enhanced.

- In this thesis two consistencies, cross view support consistency and cross view projection consistency is considered. We believe that there exists more consistencies/associations between multiple cameras. So, consistencies among more than two cameras can be analyzed.
- The cross view projection factor f_p is calculated by projecting images to virtual camera using metric M. However, new metric M is calculated in later step. So, f_p can be improved using new metric M. So, for future work, f_p can be calculated using new metric M.
- During cross view support factor f_s calculation, a dictionary is maintained of all the training images of individual camera. This dictionary can cause noise for

sparse coding because each vehicle image looks unique in appearance change across camera. So, a more effective dictionary is required after applying constraints.

References

[1] B. He, J. Li, Y. Zhao and Y. Tian, "Part-Regularized Near-Duplicate Vehicle Re-Identification," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[2] J. Liu and J. Ye, “Efficient Euclidean projections in linear time,” in Proc. Int. Conf. Mach. Learn., 2009, pp. 657–664.

[3] M. Kostinger, M. Hirzer, P. Wohlhart, P. Roth, and H. Bischof, “Large scale metric learning from equivalence constraints,” in Proc. IEEE Conf. Comput. Vis. Pattern Recog., Jun. 2012, pp. 2288–2295.

[4] Y. Bai, Y. Lou, F. Gao, S.Wang, Y.Wu, and L. Duan. Group sensitive triplet embedding for vehicle re-identification. *TM- M*, 2018.

[5] S. Ding, L. Lin, G. Wang, and H. Chao. Deep feature learning with relative distance comparison for person reidentification. *PR*, 48(10), 2015.

[6] J. Fu, H. Zheng, and T. Mei. Look closer to see better: Recurrent attention convolutional neural network for fine-grained image recognition. In *CVPR*, volume 2, 2017.

[7] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *CVPR*, 2016.

[8] S. Huang, Z. Xu, D. Tao, and Y. Zhang. Part-stacked cnn for fine-grained visual categorization. In *CVPR*, 2016.

[9] M. M. Kalayeh, E. Basaran, M. G¨okmen, M. E. Kamasak, and M. Shah. Human semantic parsing for person reidentification. In *CVPR*, 2018.

[10] D. Li, X. Chen, Z. Zhang, and K. Huang. Learning deep context-aware features over body and latent parts for person re-identification. In *CVPR*, 2017.

[11] D. Li, X. Chen, Z. Zhang, and K. Huang. Learning deep context-aware features over body and latent parts for person re-identification. In *CVPR*, 2017.

[12] S. Liao, Y. Hu, X. Zhu, and S. Z. Li. Person re-identification by local maximal occurrence representation and metric learning. In *CVPR*, 2015.

[13] D. Lin, X. Shen, C. Lu, and J. Jia. Deep lac: Deep localization, alignment and classification for fine-grained recognition. In *CVPR*, 2015.

[14] M. Lin, Q. Chen, and S. Yan. Network in network. *arXiv preprint arXiv:1312.4400*, 2013.

[15] H. Liu, Y. Tian, Y. Yang, L. Pang, and T. Huang. Deep relative distance learning: Tell the difference between similar vehicles. In *CVPR*, 2016.

[16] X. Liu, W. Liu, H. Ma, and H. Fu. Large-scale vehicle reidentification in urban surveillance videos. In *Multimedia and Expo (ICME), 2016 IEEE International Conference on*. IEEE, 2016.

[17] X. Liu, W. Liu, T. Mei, and H. Ma. A deep learning-based approach to progressive vehicle re-identification for urban surveillance. In *ECCV*. Springer, 2016.

[18] Y. Lou, Y. Bai, J. Liu, S. Wang, and L.-Y. Duan. Embedding adversarial learning for vehicle re-identification. *IEEE Transactions on Image Processing*, 2019.

[19] P. Luo, X. Wang, and X. Tang. Hierarchical face parsing via deep learning. In *CVPR*. IEEE, 2012.

[20] J.-J. Lv, X. Shao, J. Xing, C. Cheng, X. Zhou, et al. A deep regression architecture with two-stage re-initialization for high performance facial landmark detection. In *CVPR*, volume 1, 2017.

[21] O. M. Parkhi, A. Vedaldi, C. Jawahar, and A. Zisserman. The truth about cats and dogs. In *ICCV*. IEEE, 2011.

[22] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection. In *Proceed-ings of the IEEE conference on computer vision and patternrecognition*, pages 779–788, 2016.

[23] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In *CVPR*, 2015.

[24] Y. Shen, T. Xiao, H. Li, S. Yi, and X. Wang. Learning deep neural networks for vehicle re-id with visual-spatio-temporal path proposals. In *ICCV*. IEEE, 2017.

[25] C. Su, J. Li, S. Zhang, J. Xing, W. Gao, and Q. Tian. Posedriven deep convolutional model for person re-identification. In *ICCV*. IEEE, 2017.

[26] Y. Sun, X. Wang, and X. Tang. Deep convolutional network cascade for facial point detection. In *CVPR*, 2013.

[27] 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 reidentification. In *CVPR*, 2017.

[25] K. Q. Weinberger and L. K. Saul. Distance metric learning for large margin nearest neighbor classification. *Journal of Machine Learning Research*, 10(Feb), 2009.

[26] T. Xiao, H. Li,W. Ouyang, and X.Wang. Learning deep feature representations with domain guided dropout for person re-identification. In *CVPR*, 2016.

[27] K. Yan, Y. Tian, Y. Wang, W. Zeng, and T. Huang. Exploiting multi-grain ranking constraints for precisely searching visually-similar vehicles. In *ICCV*, 2017.

[28] L. Yang, P. Luo, C. Change Loy, and X. Tang. A large-scale car dataset for fine-grained categorization and verification. In *CVPR*, 2015.

[29] H. Zhang, T. Xu, M. Elhoseiny, X. Huang, S. Zhang, A. Elgammal, and D. Metaxas. Spda-cnn: Unifying semantic part detection and abstraction for fine-grained recognition. In *CVPR*, 2016.

[30] N. Zhang, J. Donahue, R. Girshick, and T. Darrell. Partbased r-cnns for fine-grained category detection. In *ECCV*. Springer, 2014.

[31] X. Zhang, H. Luo, X. Fan, W. Xiang, Y. Sun, Q. Xiao, W. Jiang, C. Zhang, and J. Sun. Alignedreid: Surpassing human-level performance in person re identification. *arXiv preprint arXiv:1711.08184*, 2017.

[32] X. Zhang, H. Xiong, W. Zhou, W. Lin, and Q. Tian. Picking deep filter responses for fine-grained image recognition. In *CVPR*, 2016.

[33] Z. Zhang, P. Luo, C. C. Loy, and X. Tang. Facial landmark detection by deep multi-task learning. In *ECCV*. Springer, 2014.

[34] H. Zhao, M. Tian, S. Sun, J. Shao, J. Yan, S. Yi, X. Wang, and X. Tang. Spindle net: Person re-identification with human body region guided feature decomposition and fusion. In *CVPR*, 2017.

[35] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian. Scalable person re-identification: A benchmark. In *CVPR*, 2015.

[36] Y. Zhou and L. Shao. Aware attentive multi-view inference for vehicle re-identification. In *CVPR*, 2018.

[37] S. Zhu, C. Li, C. Change Loy, and X. Tang. Face alignment by coarse-to-fine shape searching. In *CVPR*, 2015.

[38] H. Liu, Y. Tian, Y. Yang, L. Pang, and T. Huang, "Deep relative distance learning: Tell the difference between similar vehicles," in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016.

[39] S. Mahendran and R. Vidal, “Car segmentation and pose estimation using 3d object models,” arXiv preprint arXiv:1512.06790, 2015.

[40] B. C. Matei, H. S. Sawhney, and S. Samarasekera, “Vehicle tracking across nonoverlapping cameras using joint kinematic and appearance features,” in Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011, pp. 3465–3472.

[41] M. Valera and S. A. Velastin, “Intelligent distributed surveillance systems: a review,” IEE Proceedings-Vision, Image and Signal Processing, vol. 152, no. 2, pp. 192–204, 2005.

[42] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, “Datadriven intelligent transportation systems: A survey,” IEEE Transactions on Intelligent Transportation Systems, vol. 12, no. 4, pp. 1624–1639, 2011.

[43] M. A. Saghafi, A. Hussain, M. H. M. Saad, N. M. Tahir, H. B. Zaman, and M. Hannan, “Appearance-based methods in re-identification: a brief review,” in Signal Processing and its Applications (CSPA), 2012 IEEE 8th International Colloquium on. IEEE, 2012, pp. 404–408.

[44] 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.

[45] X. Liu, W. Liu, H. Ma, and H. Fu, “Large-scale vehicle re-identification in urban surveillance videos,” in Multimedia and Expo (ICME), 2016 IEEE International Conference on. IEEE, 2016, pp. 1–6.

[46] Z. Zhong, L. Zheng, D. Cao, and S. Li, “Re-ranking person re-identification with k-reciprocal encoding,” arXiv preprint arXiv:1701.08398, 2017.

[47] M. Ye, C. Liang, Y. Yu, Z. Wang, Q. Leng, C. Xiao, J. Chen, and R. Hu, “Person reidentification via ranking aggregation of similarity pulling and dissimilarity pushing,” IEEE Transactions on Multimedia, vol. 18, no. 12, pp. 2553–2566, 2016.

[48] A. J. Ma and P. Li, "Query based adaptive re-ranking for person reidentification," in Asian Conference on Computer Vision. Springer, 2014, pp. 397–412.

[49] E. Ahmed, M. Jones, and T. K. Marks, "An improved deep learning architecture for person re-identification," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3908–3916.

[50] T. Xiao, H. Li, W. Ouyang, and X. Wang, "Learning deep feature representations with domain guided dropout for person re-identification," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1249–1258.

[51] H. Liu, Y. Tian, Y. Yang, L. Pang, and T. Huang, "Deep relative distance learning: Tell the difference between similar vehicles," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2167–2175.

[52] D.-H. Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in Workshop on Challenges in Representation Learning, ICML, vol. 3, 2013, p. 2.

[53] A. Odena, "Semi-supervised learning with generative adversarial networks," arXiv preprint arXiv:1606.01583, 2016.

[54] Z. Zheng, L. Zheng, and Y. Yang, "Unlabeled samples generated by gan improve the person re-identification baseline in vitro," arXiv preprint arXiv:1701.07717, 2017.

[55] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," arXiv preprint arXiv:1511.06434, 2015.

[56] Y. Shen, T. Xiao, H. Li, S. Yi, and X. Wang, "Learning deep neural networks for vehicle re-id with visual-spatio-temporal path proposals," arXiv preprint arXiv:1708.03918, 2017.

[57] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[58] X. Liu, W. Liu, T. Mei, and H. Ma, "A deep learning-based approach to progressive vehicle re-identification for urban surveillance," in European Conference on Computer Vision. Springer, 2016, pp. 869–884.

[59] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818–2826.

[60] D. Qin, S. Gammeter, L. Bossard, T. Quack, and L. Van Gool, "Hello neighbor: Accurate object retrieval with k-reciprocal nearest neighbors," in Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011, pp. 777–784.

[61] S. Bai and X. Bai, "Sparse contextual activation for efficient visual reranking," IEEE Transactions on Image Processing, vol. 25, no. 3, pp. 1056–1069, 2016.

[62] A. Vedaldi and K. Lenc, "Matconvnet: Convolutional neural networks for matlab," in Proceedings of the 23rd ACM international conference on Multimedia. ACM, 2015, pp. 689–692.

[63] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard et al., "Tensorflow: A system for large-scale machine learning." in OSDI, vol. 16, 2016, pp. 265–283.

[64] S. Ding, L. Lin, G. Wang, and H. Chao, "Deep feature learning with relative distance comparison for person re-identification," Pattern Recognition, vol. 48, no. 10, pp. 2993–3003, 2015.

[65] W.-S. Zheng, S. Gong, and T. Xiang, "Person re-identification by probabilistic relative distance comparison," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2011, pp. 649–656.

[66] N. O'Hare and A. F. Smeaton, "Context-aware person identification in personal photo collections," *IEEE Trans. Multimedia,* vol. 11, no. 2, pp. 220–228, Feb. 2009.

[67] S. Gong, M. Cristani, S. Yan, and C. Loy, *Person Re-identification*. New York, NY, USA: Springer, 2014.

[68] X. Wang, T. Zhang, Tretter, and Q. Lin, "Personal clothing retrieval on photo collections by color and attributes," *IEEE Trans. Multimedia*, vol. 15, no. 8, pp. 2035–2045, Dec. 2013.

[69] K. W. Chen, C. C. Lai, P. J. Lee, C. S. Chen, and Y. P. Hung, "Adaptive learning for target tracking and true linking discovering across multiple non-overlapping cameras," *IEEE Trans.Multimedia,* vol. 13, no. 4, pp. 625–638, Aug. 2011.

[70] H. Ben Shitrit, J. Berclaz, F. Fleuret, and P. Fua, "Tracking multiple people under global appearance constraints," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 137–144.

[71] L. L. Presti, S. Sclaroff, and M. L. Cascia, "Path modeling and retrieval in distributed video surveillance databases," *IEEE Trans. Multimedia*, vol. 14, no. 2, pp. 346–360, Apr. 2012.

[72] J. W. Hsieh, Y. T. Hsu, H. Y. Liao, and C. C. Chen, "Video-based human movement analysis and its application to surveillance systems," *IEEE Trans. Multimedia*, vol. 10, no. 3, pp. 372–384, Apr. 2008.

[73] F. Chen, C. De Vleeschouwer, and A. Cavallaro, "Resource allocation for personalized video summarization," *IEEE Trans. Multimedia,* vol. 16, no. 2, pp. 455–469, Feb. 2014.

[74] R. Vezzani, D. Baltieri, and R. Cucchiara, "People reidentification in surveillance and forensics: A survey," *ACM Comput. Surveys*, vol. 46, no. 2, p. 29, 2013.

[75] N. A. Fox, R. Gross, J. F. Cohn, and R. B. Reilly, "Robust biometric person identification using automatic classifier fusion of speech, mouth, and face experts," *IEEE Trans. Multimedia*, vol. 9, no. 4, pp. 701–714, Jun. 2007.

[76] W. Yin, J. Luo, and C. W. Chen, "Event-based semantic image adaptation for user-centric mobile display devices," *IEEE Trans. Multimedia*, vol. 13, no. 3, pp. 432–442, Jun. 2011.

[77] X. Li, D. Tao, L. Jin, Y. Wang, and Y. Yuan, "Person re-identification by regularized smoothing kiss metric learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 10, pp. 1675–1685, Oct. 2013.

[78] Y. Wang, R. HU, C. Liang, C. Zhang, and Q. Leng, "Camera compensation using feature projection matrix for person re-identification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 8, pp. 1350–1361, Aug. 2014.

[79] L. Ma, X. Yang, and D. Tao, "Person re-identification over camera networks using multi-task distance metric learning," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3656–3670, Aug. 2014.

[80] Q. Leng, R. Hu, and C. Liang, "Bi-directional ranking for person re-identification," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2013, pp. 1–6.

[81] S. Ali, O. Javed, N. Haering, and T. Kanade, "Interactive retrieval of targets for wide area surveillance," in *Proc. ACM Int. Conf. Multimedia*, 2010.

[82] C. Liu, C. Loy, and S. Gong, "POP: Person re-identification post-rank optimisation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 441–448.

[83] Z. Wang, R. Hu, C. Liang, Q. Leng, and K. Sun, "Region-based interactive ranking optimization for person re-identification," in *Proc. Pacific- Rim Conf. Multimedia*, 2014, pp. 1–10.

[84] N. Gheissari, T. B. Sebastian, and R. Hartley, "Person reidentification using spatiotemporal appearance," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2006, vol. 2, pp. 1528–1535.

[85] X. Wang, G. Doretto, T. Sebastian, J. Rittscher, and P. Tu, "Shape and appearance context modeling," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–8.

[86] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, "Person re-identification by symmetry-driven accumulation of local features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2010, pp. 2360–2367.

[87] B. Ma, Y. Su, and F. Jurie, "Bicov: A novel image representation for person re-identification and face verification," in *Proc. Brit. Mach. Vis. Conf.*, 2012, p. 11.

[88] B. Layne, T. Hospedales, and S. Gong, "Person re-identification by attributes," in *Proc. Brit. Mach. Vis. Conf.*, 2012, p. 8.

[89] I. Kviatkovsky, A. Adam, and E. Rivlin, "Color invariants for person reidentification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 7, pp. 1622–1634, Jul. 2013.

[90] Z. Rui, O. Wanli, and W. Xiaogang, "Unsupervised salience learning for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2013, pp. 3586–3593.

[91] W. Li, R. Zhao, T. Xiao, and X. Wang, "DeepReID: Deep filter pairing neural network for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2014, pp. 152–159.

[92] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 262–275.

[93] B. Prosser, W.-S. Zheng, S. Gong, and T. Xiang, "Person reidentification by support vector ranking," in *Proc. Brit.Mach. Vis. Conf.*, 2010, p. 6.

[94] K. Q. Weinberger, J. Blitzer, and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *J. Mach. Learn. Res.*, vol. 10, pp. 207–244, 2009.

[95] M. Hirzer, C. Beleznai, M. Kstinger, P. M. Roth, and H. Bischof, “Dense appearance modeling and efficient learning of camera transitions for person re-identification,” in *Proc. IEEE Int. Conf. Image Process.*, Sep.–Oct. 2012, pp. 1617–1620.

[96] M. Dikmen, E. Akbas, T. Huang, and N. Ahuja, “Pedestrian recognition with a learned metric,” in *Proc. Asian Conf. Comput. Vis.*, 2010, pp. 501–512.

[97] M. Kostinger, M. Hirzer, P. Wohlhart, P. Roth, and H. Bischof, “Large scale metric learning from equivalence constraints,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2012, pp. 2288–2295.

[98] A. Mignon and F. Jurie, “PCCA: A new approach for distance learning from sparse pairwise constraints,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2012, pp. 2666–2672.

[99] S. Pedagadi, J. Orwell, S. Velastin, and B. Boghossian, “Local fisher discriminant analysis for pedestrian re-identification,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2013, pp. 3318–3325.

[100] Z. Li, S. Chang, F. Liang, T. Huang, L. Cao, and J. Smith, “Learning locally-adaptive decision functions for person verification,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2013, pp. 3610–3617.

[101] J. Jiang, R. Hu, Z. Wang, and Z. Han, “Noise robust face hallucination via locality-constrained representation,” *IEEE Trans. Multimedia*, vol. 16, no. 5, pp. 1268–1281, Aug. 2014.

[102] F. Wu, S. Yan, J. S. Smith and B. Zhang, "Joint Semi-supervised Learning and Re-ranking for Vehicle Re-identification," 2018 24th International Conference on Pattern Recognition (ICPR), 2018.

[103] R.M.S. Bashir, M. Shahzad, M.M. Fraz,VR-PROUD: Vehicle Re-identification using PROgressive Unsupervised Deep architecture, Pattern Recognition, Volume 90, 2019.

[104] Y. Zhouy and L. Shao, "Viewpoint-Aware Attentive Multi-view Inference for Vehicle Re-identification," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018.

[105] Z. Wang et al., "Zero-Shot Person Re-identification via Cross-View Consistency," in IEEE Transactions on Multimedia, vol. 18, no. 2, pp. 260-272, Feb. 2016