2020 IEEE 2nd International Conference on Civil Aviation Safety and Information Technology (ICCASIT) | 978-1-7281-9948-1/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICCASIT50869.2020.9368533

Research on Vehicle Re-Identition based on Image Feature fusion

Yanyang Liu
AIPARK
Zhangjiakou City, Hebei Province, China
e-mail: liuyanyang @aipark.com

Jun Yan
AIPARK
Zhangjiakou City, Hebei Province, China
e-mail: yanjun @aipark.com

***Abstract*—In order to improve the accuracy of vehicle Re-Identition, a method combining global features and local features was proposed to solve the problem. Vehicle appearance blur, pedestrian and motor vehicle block, and vegetation block could effect the accuracy of the algorithm. This study proposed a new method. First of all, the siamese network was used to match the shape of the vehicle's head, vehicle's tail and the whole vehicle. Then, a license plate recognition network based on YOLOv3 was used to identify the characters. The experimental results showed that the algorithm could achieve better vehicle Re-Identition effect in the complex road environment, and the accuracy of Re-Identition reached 95.82%, which was 7.89, 4.02 and 1.81 percentage points higher than the three classic Re-Identition models of DRDL, OIFE and RAM, respectively.**

***Keywords*—image processing, target recognition, deep learning, vehicle re-identification, artificial intelligence**

I. INTRODUCTION

As an important object in urban monitoring, vehicle involved a lot of tasks such as detection, tracking and dispatching. Vehicle Re-Identification is to find out the same vehicle photographed by different cameras, or the same vehicle photographed by the same camera under different lights and different perspectives. Through vehicle Re-Identification technology, cross-lens automatic identification and lock of the same vehicle could be realized, which played a very important role in urban traffic scheduling, illegal vehicle tracking and other specific tasks, and was beneficial to the planning and development of intelligent traffic and smart city.

Image fusion was to make redundant complementary of two or more images with different focuses in the same scene, and finally obtained the image with complete information. In recent years, with the development of deep learning, inspired by pedestrian Re-Identition, deep learning technology had been widely applied in vehicle recognition tasks. Different from traditional methods, the vehicle re-recognition method based on deep learning did not need manual design features. It can automatically learn various features of vehicle images through neural network, and can process a large number of data. License plate recognition is the main method of vehicle re-recognition in the early stage and has been widely used. Although the license plate is the unique identification of the vehicle, in some special circumstances, the license plate information cannot be accurately recognized for some reasons. For example, in the process of driving, the resolution of the image is not enough and the image is not clear because of the environment and the camera. In addition, in actual road traffic scenes, license plates are often obscured, removed, or even forged. Therefore, it is very necessary to combine license plate recognition and vehicle appearance recognition to carry out re-recognition research.

Yan[1] proposed a new triad loss function to improve the accuracy of vehicle re-Identition. This function utilized both in-class error and inter-class error in vehicle Re-Identition process, but this method only utilized the appearance characteristics of vehicles, so the identition accuracy was still not high in complex environment. Zhou[2] proposed a two-way Long-Short Time Memory (LSTM) to solve vehicle Re-Identition in complex scenes. Aiming at the problem of vehicle Re-Identition of different vehicles of the same model, Li[3] adopted the partial instead of the whole method to obtain Windows and vehicle face areas with large differences of different vehicles, extracted and fused the vehicle features of the detected vehicle window and vehicle face areas, and generate new fusion features. Classification and recognition are carried out by calculating the distance between image features.The test results on the public data set VRIde-1 of Sun Yat-sen University showed that the Rank1 matching rate of the algorithm reached 66.67%, which verified the feasibility and effectiveness of the algorithm.De[4] put forward a kind of double layer cascade vehicles to identify the network, the input image into two different size of the image, respectively by two cascade twin network detection vehicle image appearance characteristics of different sizes, improves the vehicle identification accuracy, two cameras in the self-built dataset of recognition accuracy rate reached 92.6% and 98.7% respectively. Zhou[5] used convolutional neural network and LSTM network to jointly learn image features from different perspectives in vehicle images, which improved the accuracy of vehicle Re-Identition. However, it still only considered the appearance features of vehicle images, and feature extraction had certain limitations.

Most of the above algorithms only consider unilateral factors and the accuracy of vehicle Re-Identition is limited. The vehicle Re-Identition method based on image feature fusion could extract the local and global features of the vehicle and had great advantages in feature extraction.In the task of vehicle Re-Identition, the image resolution affected the accuracy of image matching. In the process of training vehicle Re-Identition model, the neural network designed in the past needs to change the resolution of the original image to adapt to the training network. The change of image resolution may reduce the accuracy of license plate number recognition.To sum up, this paper proposes a method combining global features and local features to solve the problems of vehicle license plate blur and vehicle appearance

978-1-7281-9948-1/20/$31.00 ©2020 IEEE

blur due to poor picture quality, pedestrian and motor vehicle occlusion, and vegetation occlusion resulting in low license plate recognition accuracy in vehicle re-recognition. First, the twin network is used to match the appearance and shape of the detected vehicle images. Then, the license plate recognition network based on YOLOv3 is used to combine the local shape, overall shape and license plate recognition for vehicle re-recognition. The experimental results show that the algorithm can realize vehicle re-recognition in changeable road environment and at night.

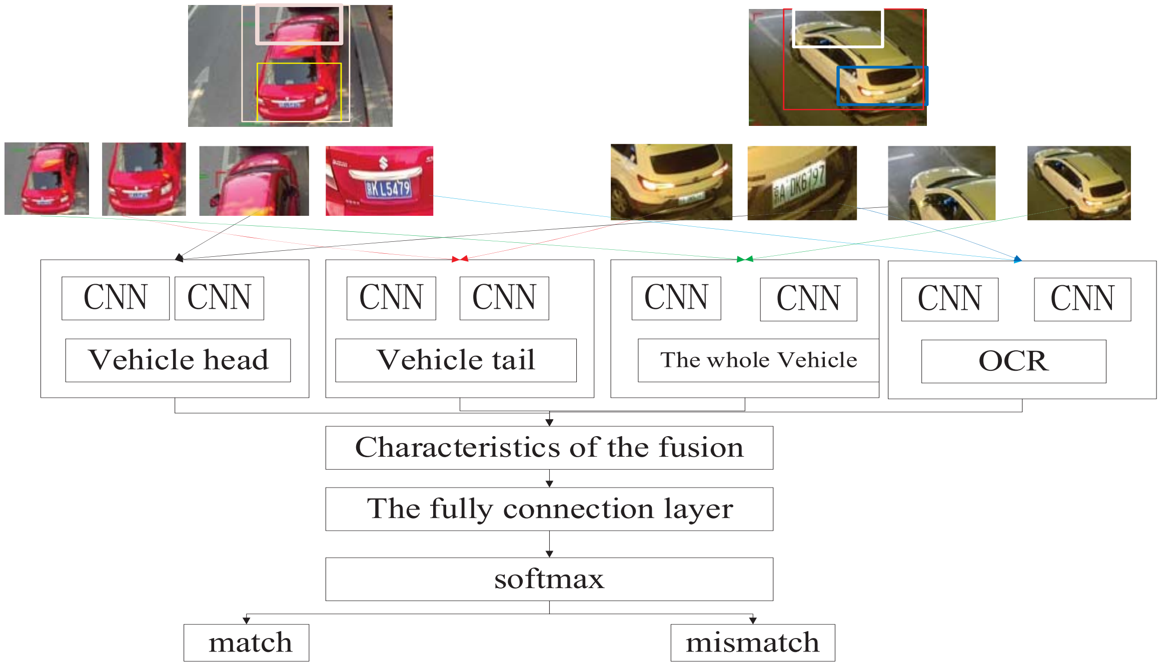

Fig. 1. Overall flow chart

II. Vehicle re-identification network

In this study, vehicle appearance area and license plate character area are taken as two independent local features, and two network models are selected to process the two areas respectively. The twin neural network is used to identify the shape similarity of vehicle appearance. The network compares two images input at the same time by sharing weight to obtain the similarity of vehicle appearance. YOLOv3 network is used as the license plate character recognition network to recognize license plate number information in two vehicle images entered at the same time. The accuracy of license plate recognition is determined by the character similarity of license plate number. A full connection layer is adopted to combine the similarity of shape recognition and license plate number recognition, and the recognition results after 5 detection modules are combined with the joint probability matching through the full connection layer to obtain the final probability model of re-recognition.Fig.1 is the overall flow chart of vehicle re-identification. As can be seen from Fig. 1, a multi-branch detection module is designed for image decomposition operation. The detection module includes the vehicle global appearance detection module, the front appearance detection module, the rear appearance detection module, and the license plate character detection module. By matching the vehicle images taken by different cameras, the re-recognition method proposed in this study, on the one hand, improves the accuracy of the re-recognition model, and on the other hand, improves the interpretability of the re-recognition algorithm, making it easy to realize.

A. Siamese Network

Siamese network in 2005 by Chopra[6]improved, since then, the original siamese network was divided into two parts, the first part was used for feature extraction, the second half with a character vector distance measurement, enter into the network with two pictures, two characteristic vector and the output calculation of the input image similarity to identify the similarity of two images.

This method is a similarity measurement method, which can be applied to the identification and classification of categories when the number of samples for each category is small.In this paper, the twin network is adopted. The two parts on the left and right sides are identical in structure and share the same parameters, namely weight and bias.The siamese network is composed of two convolutional layers, two pooling layers, a residual block and a full connection layer.

The number of channels in both the convolutional layer and the pooling layer is 16, the size of the convolution kernel is 3 × 3, the size of the pooling window is 2 × 2, the convolution mode is zero filling, and the activation function used Relu.

The loss function adopted by the twin network is the contrast loss function, and the contrast loss function is formula 1:

$$L = \frac{1}{2N}\sum_{n=1}^{N} yd^2 + (1-y)\max(m\arg in - d, 0)^2 \quad (1)$$

Where, D is the Euclidian distance between the two sample features, Y is the similarity or matching of the two samples, and margin is the threshold set.As shown in Fig.2, is the Siamese network structure diagram.

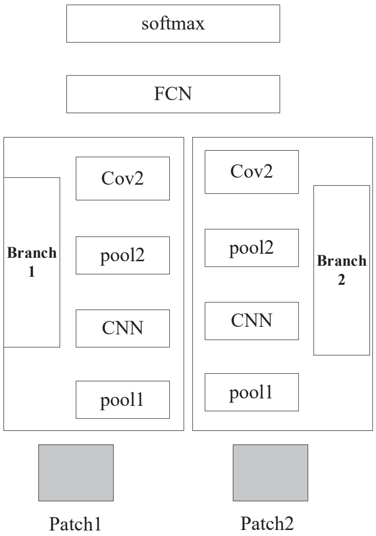

Fig.2. Siamese network structure diagram

B. YOLOv3 network

YOLO v3[7] is an end-to-end deep neural network system, whose loss function is composed of bounding box coordinate regression loss, classification loss and confidence loss.Assuming that the target box position information of license plate characters predicted by YOLO V3 is (X1, Y1,W1,H1), and the real license plate character position information is (X2,Y2,W2,H2), then the regression loss of boundary box coordinates is formula 2:

$$loss_{coo} = (2 - W1 \times H1) \cdot \| [X1, Y1, W1, H1] - [W2, Y2, W2, H2] \|^2 \quad (2)$$

The loss function is mainly used for the regression of license plate character position.Confidence is used to determine whether there is a license plate character target in the grid of the decision diagram. The corresponding confidence loss is shown in Equation 3:

$$loss_{conf} = (conf_e - conf_r)^2 \quad (3)$$

Respectively represent the confidence predicted by license plate characters and the true confidence.The classification loss can be expressed as a typical form of cross entropy. The loss function is shown in Equation 4:

$$loss_{cls} = -\sum_{i=0}^{h \times w} [c_i^e \cdot \log(c_i^r) + (1 - c_i^e) \log(1 - c_i^r)] \quad (4)$$

Where c_i^e, c_i^r,represents the prediction probability and true probability of the ith grid point belonging to the license plate character class, in which the true probability is based on the IOU of the actual position target box of the grid point and the license plate character.The total loss function is the weighted sum of the above losses:

$$loss = \lambda_{coo} \cdot loss_{coo} + \lambda_{conf} \cdot loss_{conf} + \lambda_{cls} \cdot loss_{cls} \quad (5)$$

In the above equation, $\lambda_{coo}, \lambda_{cls}$ is a positive weight parameter, and the network parameters are adjusted by means of error reverse transfer for the above losses.

Compared with YOLOv1 and YOLOv2, the YOLOv3 network model is faster and more accurate for target detection and recognition in pictures. The detection process is as follows:

(1) First, the input vehicle image should be processed, and the image size should be converted to 416×416, which is an integer multiple of 32.

(2) Next, the input image after scaling is segmented into grids. After extracting features through Darknet-53 feature extraction network, 3 feature maps with different scales are extracted, with the size of 13, 26 and 52 respectively.

(3) Darknet-53 can reduce the extracted feature map to one third of the original input image, and each scale feature map will predict the license plate characters in the image.

(4) In the YOLOv3 network structure, a variety of scale feature maps appear, thus generating different sizes of predicted anchor points.

For the detection and classification of final license plate characters, YOLOv3 network USES binary cross entropy and logistic regression methods to complete end-to-end training and realize license plate character recognition. The structure of YOLOv3 is shown in Fig 3.

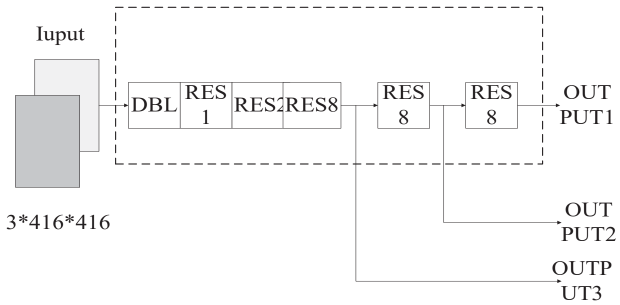

Fig.3. Yolov3 network structure diagram

III. Analysis of Test Results

A. The data set

In order to verify the effectiveness of the vehicle re-identification network proposed in this paper, it was verified in the scene of day and night respectively, in which the self-built data set was used for training and verification. All the image data were taken by Intelligent Communication Technology Co., Ltd. in Changping District, Beijing from June to November, 2019.The data set covers images of different vehicles taken by two cameras: images of vehicles with the same ID taken by different cameras, and images of vehicles with different ids taken by the same camera. Ten videos from two different cameras were used. Each video was about 20 minutes long and the resolution was 1920× 1080 and 35 frames per second. In addition, an XML file was created for each video, in which each category corresponds to a vehicle with a different ID, containing the license plate number and the corresponding identifier converted to ASCII, the location of the shot, and the vehicle model, color, year, and brand. The video was recorded at different times of the day so that the video images taken at different times of the day had different brightness. Table I shows the statistics of the number of vehicle pictures and license plate number taken by different cameras.

TABLE I. NUMBER OF IMAGES IN THE DATA SET

Period of time	Camera 1		Camera 2	
	number of vehicles	*the license plate number*	*number of vehicles*	*the license plate number*
1	385	342	277	245
2	350	301	244	225
3	340	312	273	252
4	280	258	230	196
5	345	299	242	205
total number	1700	1512	1266	1123

Camera 1 and camera 2 captured the different period of vehicle image, to create images, the acquired images are mostly taken from vehicle video frames in a row, because the appearance of the vehicles in a sequence of consecutive frame images usually have small change, acquisition of consecutive frame images can avoid vehicle movement caused by the fuzzy effects, vision brightness, background, etc. Therefore, 5 sets of vehicle ID matching/mismatching pairs were generated between the two cameras, as shown in TABLE II.

TABLE II. THE NUMBER OF ID MATCHING/MISMATCHED VEHICLE IMAGES GENERATED PER TIME PERIOD

Period of time	Mismatched object	matched objec
1	83255	19560
2	66732	17370
3	76681	19520
4	49750	14177
5	60321	16030
total number	336739	86657

As shown in Table II, in the surveillance images taken by the two cameras, the number of matches with the same car with the same ID appears to be 86,657, while the remaining 336,739 cars captured by the two cameras cannot be matched.

B. Data characteristics

The data set contains the vehicle image, including the appearance of the front, the appearance of the rear, the overall appearance of the vehicle, the appearance of the license plate area and the license plate characters. In the front and rear area, different vehicles may have different decorations, annual inspection marks, etc. These significant features can be used to distinguish different vehicles and effectively reduce inter-class error and intra-class error in the process of vehicle re-identification. At the same time, the license plate is the unique identification mark of the vehicle, and integrating it into the vehicle Re-Identition network can effectively improve the accuracy of the re-recognition model. Table III describes the characteristics of vehicle appearance and license plate area.

TABLE III. DATA CHARACTERISTICS

Characteristics of the area	Character description
Head of vehicle	There are different decorations, annual inspection marks, the location and appearance of decorations, as well as different degrees of scratches, antennas, headlights, front Windows, models, colors, etc.
Tail of vehicle	There are different decorations, the location and appearance of decorations, as well as different degrees of scratches, antennas, rear lights, rear Windows, models, colors, etc.
The whole vehicle	Various perspectives, overall appearance features, decorations, window features and other local features occupy the proportion of the whole car.
The license plate	Contains a license plate number representing the unique ID of the vehicle.

C. The evaluation index

Precision (P) and recall (R) are usually used to evaluate the performance effect of vehicle re-recognition model. The function of accuracy and recall rate is shown in Equation 6:

$$P = \frac{|tp|}{|tp| + |fp|}$$
$$R = \frac{|tp|}{|tp| + |fn|} \tag{6}$$

Where | tp | represents the true number of vehicle ID matches between camera 1 and camera 2, b| fp | is the number of false matches, and | fn | is the number of missing matches.In order to more accurately evaluate the performance of the re-recognition network, the harmonic average F of accuracy and recall rate is adopted, as shown in Formula 7:

$$F = \frac{2}{1/P + 1/R} \tag{7}$$

D. Vehicle re-identification results

The experiment was carried out on a PC. The hardware configuration of the PC was as follows: CPU core-i7, GeForce RTX2080 and 16GB of memory. The environment configuration is: Tensorflow1.10.0, Python 3.6, Linux Ubuntu 18.04.

To expand the data set, the brightness of the original image data was changed to 80% and 120% of the original brightness, respectively. The source image and the image after the brightness transformation are rotated left and right and front and back. The enhanced vehicle images with different brightness and different directions were obtained, and the data set was expanded.

In order to verify the effectiveness of the license plate recognition network proposed in this paper and the twin network for shape feature matching, four networks of VGG16, ResNe8, GoogleNet and MatchNet were used for feature extraction, and the accuracy was compared and evaluated. The shape features include the shape features of the caboome, the shape features of the head and the overall shape features of the vehicle. The detection effect of vehicle re-identification is shown in Table IV.

TABLE IV. THE RE-RECOGNITION ACCURACY OF DIFFERENT BACKBONES

backbone	P	R	F
VGG16	93.43%	95.54%	94.47%
ResNe8	75.73%	86.59%	80.80%
GoogleNet	80.87%	89.33%	84.89%
MatchNet	90.25%	93.11%	91.66%

As shown in Table IV, the re-recognition accuracy of the twin network with VGG16 as the feature extraction network is the highest, reaching 93.43%, and the Re-Identition recall rate of the twin network with MatchNet as the feature extraction network is the highest, reaching 90.25%. Considering the accuracy rate of re-recognition and recall rate comprehensively, the twin network with VGG16 as the feature extraction network had the highest average value of F, and the value of F reached 94.47%.

TABLE V. STRUCTURAL PARAMETERS OF SIAMESE NETWORKS BASED ON VGG-16

layer	filters	size	input	output
Conv	64	3*3/1	64*64*3	64*64*64
max		2*2/2	64*64*64	32*32*64
Conv	128	3*3/1	32*32*64	32*32*128
max		2*2/2	32*32*128	16*16*128
Conv	128	3*3/1	16*16*128	16*16*128
max		2*2/2	16*16*128	8*8*128
Conv	256	3*3/1	8*8*128	8*8*256
max		2*2/2	8*8*256	4*4*256
Conv	512	3*3/1	4*4*256	4*4*512
max		2*2/2	4*4*512	2*2*512

In order to compare the performance of the neural network described in this paper, the method of contrast experiment is adopted. The vehicle re-recognition neural network was set up, and the data set collected in this paper was used for network training and testing. The results are shown in Table VI.As can be seen from Table VI, the content of license plate area is added on the basis of the original shape and appearance, which increases the diversity of input parameters. Compared with other neural network models that input only shape features, the recognition accuracy of the proposed algorithm is improved.

TABLE VI. ACCURACY COMPARISON

Features	P	R	F
Head + tail + whole vehicle	90.50%	88.40%	89.44%
Head + tail	85.25%	84.65%	84.95%
Tail+OCR	95.85%	93.40%	94.61%
Head + OCR	92.00%	91.20%	91.60%
Head + tail + whole vehicle+ OCR	96.15%	95.50%	95.82%

E. Algorithm performance comparison

To verify the effectiveness of the method proposed in this paper, evaluating the performance of the vehicle identification model again, in the same input characteristic parameters, feature extraction under the condition of network selection for VGG16, this method respectively with DRDL[8], OIFE[9], RAM [10] the vehicle recognition method again, can be seen from table VII, the method proposed in this paper the vehicle identification accuracy is superior to other two methods, to identify the comprehensive accuracy reached 95.82%,Compared with DRDL, OIFE and RAM, the accuracy of the three classic re-recognition models is 7.89, 4.02 and 1.81 percentage points higher respectively.

TABLE VII. ACCURACY COMPARISON

methods	P	R	F
DRDL	86.35%	88.62%	87.93%
OIFE	90.55%	92.78%	91.80%
RAM	92.88%	94.67%	94.01%
My method	96.15%	95.50%	95.82%

IV. CONCLUSION AND PROSPECT

In this paper, a method combining global feature and local feature is proposed to solve the problems of vehicle license plate blurring, vehicle appearance blurring, pedestrian and motor vehicle occlusion, and green vegetation occlusion in vehicle re-recognition, which exist in the process of vehicle recognition. First, the twin network was used to match the shape of the detected vehicle images. Then, a license plate recognition network based on YOLOv3 was used for license plate recognition. The method proposed by this algorithm can achieve better vehicle Re-Identition effect in the complex road environment and the accuracy of Re-Identition reached 95.82%, which is better than the three classic re-Identition models of DRDL, OIFE and RAM.

The next research direction is how to better extract plate in the video and how to select more appropriate vehicle features for vehicle Re-Identition. At present, the vehicle Re-Identition algorithm proposed in this paper fails to meet the real-time requirements. Therefore, it is necessary to continue to explore how to improve the speed of vehicle Re-Identition in the next step, which will be an urgent problem to be solved in the future and also one of the hotspots of future research.

REFERENCES

[1] Y. Bai, Y. Lou, F. Gao, S. Wang, Y. Wu, and L. Duan, "Group sensitive triplet embedding for vehicle re-identification," IEEE Transactions on Multimedia, pp. 2385 – 2399, 2018.

[2] Zhou Y , Shao L . Vehicle Re-Identification by Adversarial Bi-Directional LSTM Network[C]// 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2018.

[3] Xiying L I , Zhihao Z , Mingkai Q . Vehicle Re-identification Algorithm Based on Component Fusion Feature[J]. Computer Engineering, 2019.

[4] De Oliveira I O , Fonseca K V O , Minetto R . A Two-Stream Siamese Neural Network for Vehicle Re-Identification by Using Non-Overlapping Cameras[J]. 2019.

[5] Zhou Y , Liu L , Shao L . Vehicle Re-Identification by Deep Hidden Multi-View Inference[J]. IEEE Transactions on Image Processing, 2018, PP(7):1-1.

[6] S. Chopra, R. Hadsell, and Y. LeCun. Learning a similarity metric discriminatively, with application to face verification[C]. In Computer Vision and Pattern Recognition, CVPR, 2005: 539 – 546.

[7] Redmon J, Farhadi A. Yolov3:An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.

[8] LIU Hongye, TIAN Yonghong, YANG Yaowei, et al. Deep relative distance learning:Tell the difference between similar vehicles[C]//IEEE Conference on Computer Vision and Pattern Recognition. 2016: 2167-2175.

[9] WANG Zhongdao, TANG Luming,LIU Xihui,et al.Orientation invariant feature embedding and spatial temporal regularization for vehicle re-identification[C]//IEEE International Conference on Computer Vision.2017: 379-387.

[10] WANG Zhongdao, TANG Luming, LIU Xihui, et al. Orientation invariant feature embedding and spatial temporal regularization for vehicle re-identification[C]// IEEE International Conference on Computer Vision. 2017: 379-387.

[11] M. Saravanan and A. Priya (2019). An Algorithm for Security Enhancement in Image Transmission Using Steganography. Journal of the Institute of Electronics and Computer, 1, 1-8.