Camera Model Identification with Convolution Neural Network

Harsh Kaushik[1], Harshita Chadha[2], Neelam Sharma[3], Harsh Nigam[4], Eklavya Sharma[5], Komal Bhagat[6]

[1, 2, 3]Department of Computer Science & Engineering, Maharaja Agrasen Institute of Technology, New Delhi, India
[4, 5, 6]Department of Electronics & Communication Engineering, Bhagwan Parshuram Institute of Technology Delhi, India
[4]harsh.nigm@gmail.com, [5]ekyusharma@gmail.com, [6]bhagat.k15@gmail.com

Abstract—With the vast improvements in multimedia, image forgery has become a growing threat to forensics. In this research, we present a deep learning-based strategy for tracing an image back to its original camera model. We categorized a set of seven camera models; however the concepts presented in this research can be easily expanded to classification of almost any number of models based on the training data provided to the system. A comparison of dense neural networks and convolutional neural networks was conducted, and both of these cutting-edge techniques produced a robust solution.

Keywords—Dense Neural Networks, Convolutional Neural Networks. Digital Images, Camera model identification.

I. INTRODUCTION

With the advent of technology and the recent influx of gadgets, cyber forensics has become a hotbed of computer related research. In today's technology oriented world, there is no shortage of digital footprints and these can inadvertently help in solving criminal cases, finding individuals, carrying out relevant research, etc. [1]. For more than a decade, identifying the brand and model of the camera that recorded a picture has been a significant study subject in information forensics [2]. Determining the type of camera used to acquire an image can aid in determining its source. Although metadata may convey information about the origin of a picture, it is simple to change [3]. As a result, signal processing is developed to determine the inherent traces left on a picture after the digital cycle. This has applications in proof verification because it substantially reduces the time required by using digital technology; an algorithm to fix this can be placed on smartphones, enabling real-time verification while eliminating the logistics of transporting image/video proof to a forensic lab [4]. It will be far more crucial in a country like India due to the absence of judges and an increase in case time. Up until now, hand-crafted filtering and feature extraction techniques have largely been used in camera source identification and other image forensics [5]. However, following the success of picture recognition reported in a number of deep learning based studies have recently appeared. The camera model problem illustration is given in Fig. 1. To create an algorithm, we must first comprehend the process of capturing a digital image. Fig. 2 depicts the pipeline and several steps of the picture capture process. In reality, current forensic research has proven that each physical and algorithmic component creates an inherent trail that is individually identifiable to the camera model, and these remnants may be employed to connect the image to the model of its manufacturer [6].

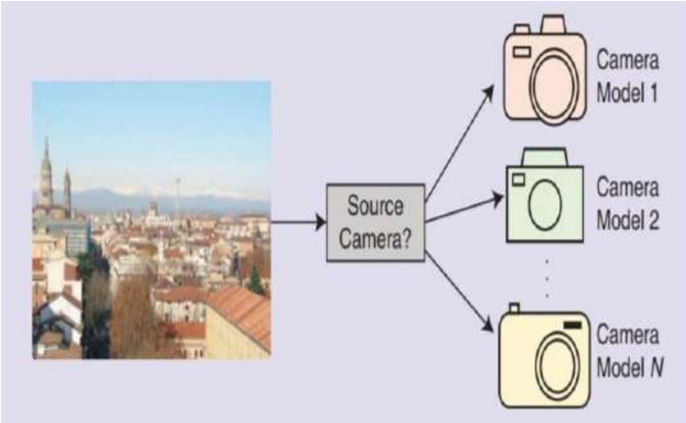

Fig.1. Camera Model Problem Illustration

Various camera types, for instance, employ different lenses, which might inject distinct noise into the image. The camera gathering process pipeline is given in Fig. 2. According to recent forensic research, each physical and algorithmic component generates an inherent trace that is uniquely traceable to the camera model [7]; those remnants may be utilized to link the image to the model of its manufacturer. For instance, various camera types employ different approximation demo-spacing methods.

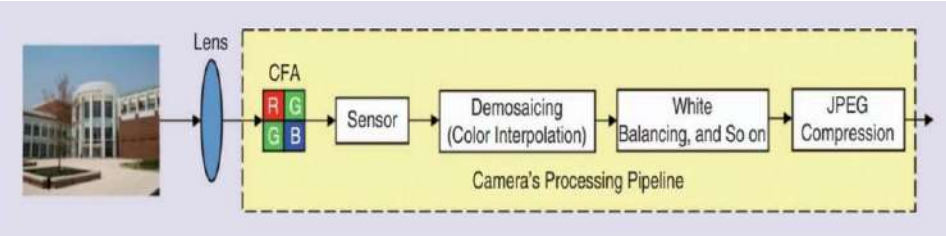

Fig.2. Camera gathering process pipeline

In this research article, we present a convolution neural network based solution for the posited problem. We first articulate the problem in section 1. In section 2, discuss notable work done in the recent past that has been conducted in this field. Sections 3-5 are dedicated towards a description of the presented proposed approach. Section 6, presents the results followed by Conclusion and Future research directions.

II. LITERATURE REVIEW

In previous works, rule based approaches have been explored. Feature engineering is done by domain experts and despite the fact that these methods have produced very good results but they are limited to their respective models and data, their extension is not guaranteed to work on unknown camera models as they don't have the ability to adapt to new data sources [8]. Statistical models have been developed to model the sensor noise pattern to classify the model of the camera used [7], [8]. Features based on JPEG compression have been built to solve this challenge [9].

For a brief overview of the research in this field, survey papers can be referred to [10]–[13]. Hu et al. [14] suggested three types of local characteristics for pedestrian portrayal: Hierarchical Weighted Histograms (HWH), Gabor Ternary Pattern HSV (GTP-HSV), and Maximally Stable Color Regions (MSCR), which could impose limiting factors on the part-level, pixel-level, and blob-level, including both. Based on the symmetry and asymmetry perceptual concepts, Bazzani et al. [15] proposed the Symmetry-Driven Accumulation of Local Characteristics (SDALFs) to identify three complementary characteristics for pedestrian photos. Yang et al. [16] created the Salient Color Names-based Color Descriptor (SCNCD) to compute colour name patterns across multiple colour models in order to manage lighting fluctuations. The Local Maximal Occurrence (LOMO) [17] technique seeks to overcome viewpoint variations by maximizing the frequency of extracted features of horizontal areas. There was a time when hand-crafted features were popular. [18]–[19] Prior to the widespread adoption of deep learning technologies in humans Gheissari et al. [20] used normalized colour and salient edge histograms to depict pedestrian pictures that are resistant to changes in surroundings and aspects.

III. ARTIFICIAL NEURAL NETWORKS

Artificial neural networks, the architecture of which is influenced by the biological brain, [21] have been known for a long time and have historically been able to tackle certain types of issues. However, more recently, neural network architectures that contained layers of hidden neurons (in addition to the input and output layers) were developed, and it is this added degree of complexity that enables deep learning and gives a more powerful set of problem-solving capabilities [22-23]. The general model of the deep neural network is given in Fig3.

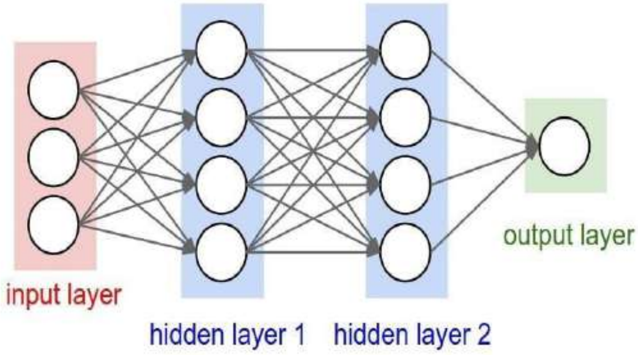

Fig.3. Deep Neural Network model

Since ANNs have such a wide range of architectures, there is no universally accepted neural network description. The ability to approximate nonlinear functions and the possession of adaptive weight sets are two common aspects of all ANNs [24].

IV. CONVOLUTION NEURAL NETWORKS

Convolution Neural Networks are a class of deep learning algorithms that are based on the mathematical operation of convolution [25]. These powerful neural networks that derive inspiration from the natural world form the backbone of the computer vision industry. A layered neuron architecture is employed here and the spatial characteristics of the images remain intact which is one of the main driving forces behind their large-scale success [26].The general model of a convolution neural network is given in Fig. 4.

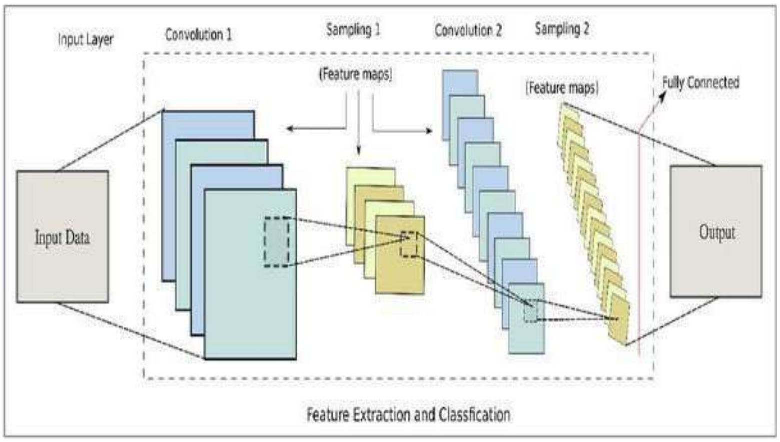

Fig.4. A general Model of a Convolution Neural Network

In keeping with the general rule of thumb when it comes to convolutional neural networks, the first few layers are charged with detecting low-level features and the deeper the network is the more details are detected. The matrices are called as kernels are convolved, i.e., a weighted sum of products is taken, with the image [27]. This exercise is repeated multiple times often coupled with pooling operations to reduce image dimensionality. Some combination of dense/ fully-connected layers is also often used to derive meaningful insights [28].

V. EXPERIMENTAL SETUP

To verify our model’s efficiency and generalization capacity, we will partition the dataset into two sets: training and validation. The training set is used to train our model so that the CNN may learn high-level characteristics from it. The test set and train set do not intersect since doing so would result in the leaking of validation set information into training, rendering the split ineffective. Our issue is primarily one of categorization. We will utilize the Dresden database, which is used to create benchmarks for numerous image related challenges in the field of forensics, making comparison simple.

To stay below our computational budget, we will create our model for 7 distinct camera models, making it a 7 multi classification task. The resolution of each image was modified to 100*100*3. In training and testing, around 13000 photos were used.

A. DENSE NEURAL NETWORKS (DNN)

The first approach employed is a basic dense connected neural network with a softmax function as the last layer. Cross entropy loss is the definition of loss, while accuracy is the assessment metric. We also look at the confusion matrix to observe how the DNN performs with various models. The model is based on local Fourier characteristics and a local binary pattern.

These characteristics capture the local variance in the image. We acquire roughly 250 features for each image once we extract these features; hence our training set has the shape (10000*250).

The network’s weights are randomly initialized, and RMS prop is used to update gradients. At highly linked layers, activation ReLU is utilized, whereas in the softmax layer, sigmoid is used. It took a day for the training to come together. The accuracy of 0.82 has been achieved on the test set. The model’s parameters considered are as in Table1.

Table.1. Summary of the model generated

LAYERS	SHAPE
input 1 (input layer)	(None, 250)
Dense 1 (dense)	(None, 1000)
Dense 2 (dense)	(None, 512)
Dense 3 (dense)	(None, 512)
Dense 4 (dense)	(None, 256)
Dense 5 (dense)	(None, 128)
softmax$_1$(*dense*)	(None, 7)

B. CONVOLUTION NEURAL NETWORK (CNN)

The second strategy is based on CNN, a specific neural network design for image-based challenges. Convnets, which have demonstrated effectiveness in recognition and classification, will be used in this design. Back-propagation is used to update the weights, the Adam optimizer to update the gradients, and the leaky-ReLU activation function in convolution layers and dense layers. Convolution layers are designed to extract features, and as we go deeper, more advanced features can be extracted, and finally convolution layer, the output is flattened and fed to a dense layer, after that we have another 2 dense layers and then in the final layers which is the softmax layer, we make calculate the probabilities associated with each class for a particular instance.

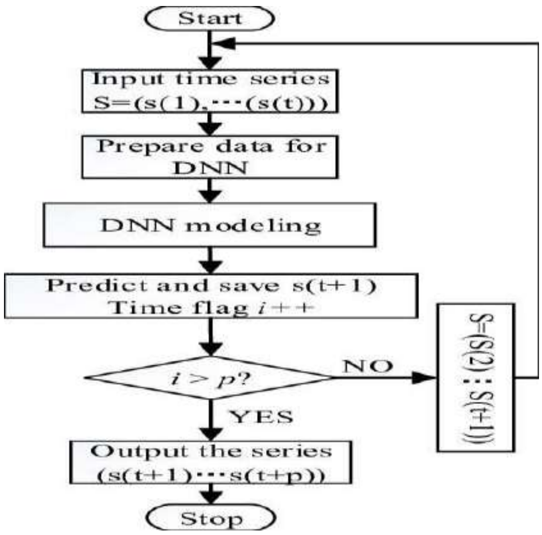

Fig.5. Flowchart of DNN training

Training took the same time as above; we made sure the amount of parameters in this is the same as the previous approach to see which architecture makes better use of the parameters.

The flowchart of the architecture of DNN and CNN training model are as in Fig.5 and Fig.6. An Accuracy of 0.91 has been achieved for the test set.

Table2. Confusion Matrix for DNN

	0	1	2	3	4	5	6
0	0.82	0.07	0.05	0.01	0.04	0.01	0
1	0.07	0.83	0.02	0.03	0.01	0	0.04
2	0.05	0.02	0.84	0.08	0.01	0	0
3	0.01	0.03	0.08	0.79	0	0.08	0.02
4	0.04	0.01	0.01	0	0.84	0.07	0.03
5	0.01	0	0	0.08	0.07	0.789	0.054
6	0	0.04	0	0.02	0.03	0.054	0.832

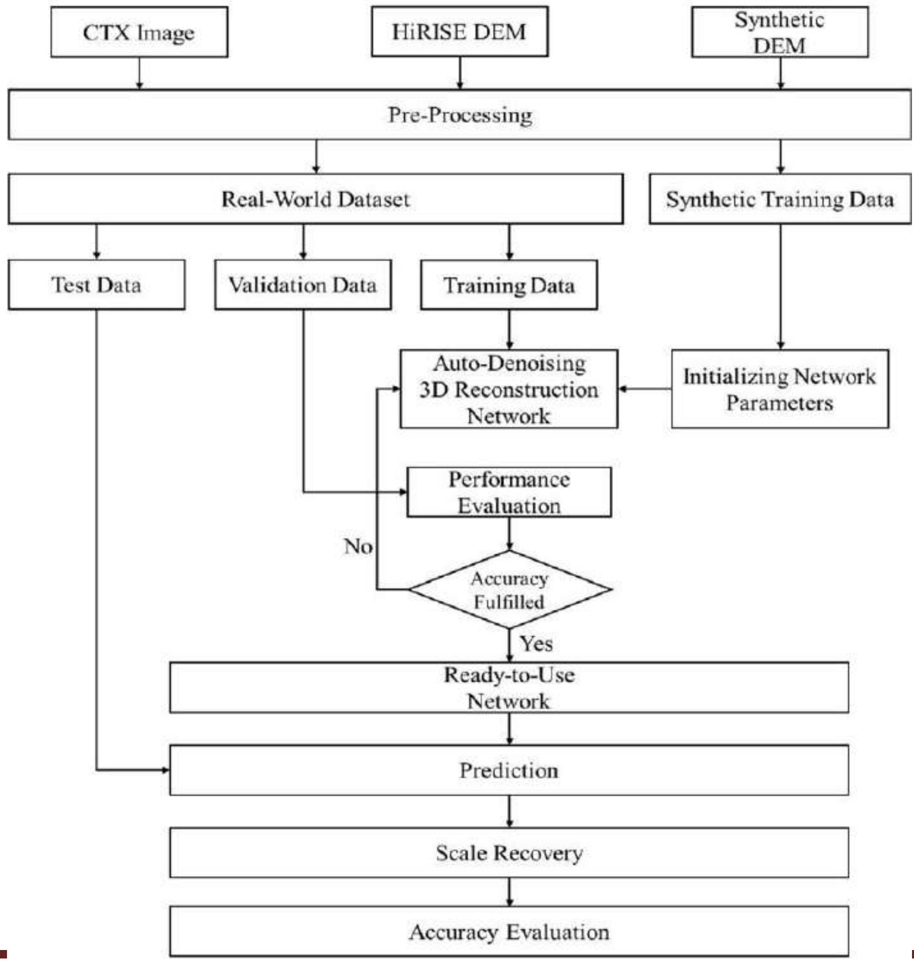
CTX Image
HiRISE DEM
Synthetic DEM
Pre-Processing
Real-World Dataset
Synthetic Training Data
Test Data
Validation Data
Training Data
Auto-Denoising 3D Reconstruction Network
Initializing Network Parameters
Performance Evaluation
No
Accuracy Fulfilled
Yes
Ready-to-Use Network
Prediction
Scale Recovery
Accuracy Evaluation

Fig.6. Flowchart of the CNN training process

VI. RESULTS

Even though we didn’t have a huge dataset and access to high-end PC, but from the above results, we can infer that CNN performs better than a simple DNN, as CNN is able to model spatial features we can assume that a camera model leaves particular kind of spatial noise in the image which can be modeled mathematically.

Table2. Summary of the model generated

LAYERS	SHAPE
input1(InputLayer)	(None, 100, 100, 3)
batchnormalization1	(None, 100, 100, 3)
conv2d1(Conv2D)	(None, 100,100, 16)
batchnormalization1	(None, 100, 100, 3)
conv2d1(Conv2D)	(None, 98, 98, 16)
maxpooling2d1	(None, 46, 46, 16)
conv2d1(Conv2D)	(None, 46, 46, 32)

batchnormalization1	(None, 46, 46, 32)
conv2d1(Conv2D)	(None, 42, 42, 32)
maxpooling2d1MaxPooling 2	(None, 16)
Dense2(dense)	(None,512)
Dense3(dense)	(None,512)
softmax1(dense)	(None, 7)

A dense neural network based and a convolution neural network based methods yielded accuracy values of 82% and 91% respectively.

Table.4. Confusion Matrix for DNN

	0	1	2	3	4	5	6
0	0.94	0.02	0.01	0.02	0	0	0.01
1	0.02	0.92	0.02	0.01	0.03	0	0
2	0.01	0.02	0.82	0.05	0.02	0.03	0.05
3	0.02	0.01	0.05	0.85	0.04	0.03	0
4	0	0.03	0.02	0.04	0.91	0	0
5	0	0	0.03	0.03	0	0.92	0.02
6	0.1	0	0.05	0	0	0.02	0.92

VII. CONCLUSION AND FUTURE SCOPE

This paper has proposed a robust methodology to trace back an image to the camera model that was used to capture it. A dense neural network based and a convolution neural network based methodologies have been employed. Both methods yielded useful results with accuracy values of 82% and 91% respectively. Variations of CNN and a much deeper network with skip connection can be used to achieve higher accuracy. We can also use pre trained networks like the Inception module, GoogleNet to harness the benefits of transfer learning. An interesting route would be to apply Capsule nets to model spatial relationships more concretely.

REFERENCES

[1]. M. C. Stamm, M. Wu, and K. J. R. Liu, "Information forensics: An overview of the first decade," IEEE Access, vol. 1, pp. 167–200, May 2013.
[2]. A. Piva. (2013). An overview on image forensics. ISRN Signal Process. [Online]. Available: https://
[3]. www.hindawi.com/journals/isrn/2013/496701/
[4]. M. Kirchner and T. Gloe, "Forensic camera model identification," in Handbook of Digital Forensics of Multimedia Data and Devices. Hoboken, NJ: Wiley, 2015, pp. 329–374.
[5].] A. Swaminathan, M. Wu, and K. J. R. Liu, "Digital image forensics via intrinsic fingerprints," IEEE Trans. Inf. Forensics Security, vol. 3, no. 1, pp. 101–117, Mar. 2008.
[6]. H. Cao and A. C. Kot, "Accurate detection of demosaicing regularity for digital image forensics," IEEE Trans. Inf. Forensics Security, vol. 4, no. 4, pp. 899–910, Dec. 2009.
[7]. C. Chen and M. C. Stamm, "Camera model identification framework using an ensemble of demosaicing features," in Proc. IEEE Int. Workshop Information Forensics and Security, 2015, Rome, Italy, pp. 1–6

[8]. T. Filler, J. Fridrich, and M. Goljan, “Using sensor pattern noise for camera model identification,” in Proc. IEEE Int. Conf. Image Processing, 2008, San Diego, CA, pp. 1296–1299.
[9]. T. H. Thai, R. Cogranne, and F. Retraint, “Camera model identification based on the heteroscedastic noise model,” IEEE Trans. Image Process., vol. 23, no. 1, pp. 250–263, Jan. 2014.
[10]. E. Kee, M. K. Johnson, and H. Farid, “Digital image authentication from JPEG headers,” IEEE Trans. Inf. Forensics Security, vol. 6, no. 3, pp. 1066–1075, Sept. 2011.
[11]. F. Marra, G. Poggi, C. Sansone, and L. Verdoliva, “Evaluation of residual-based local features for camera model identification,” in Proc. Int. Conf. Image Analysis and Processing, 2015, pp. 11–18.
[12]. M. Kharrazi, H. T. Sencar, and N. Memon, “Blind source camera identification,” in Proc. Int. Conf. Image Processing (ICIP), 2004, pp. 709–712.
[13]. C. C. Loy, T. Xiang and S. Gong, ”Multi-camera activity correlation analysis”, Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 19881995, Jun. 2009.
[14]. X. Wang, ”Intelligent multi-camera video surveillance: A review” in Pattern Recognit. Lett., vol. 34, pp. 3-19, 2013, [online] Available: http://www.sciencedirect.com/science/article/pii/S016786551200219X.
[15]. N. Gheissari, T. B. Sebastian and R. Hartley, ”Person reidentification using spatiotemporal appearance”, Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), vol. 2, pp. 1528-1535, Jun. 2006.
[16]. M. Farenzena, L. Bazzani, A. Perina, V. Murino and M. Cristani, ”Person re-identification by symmetry-driven accumulation of local features”, Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 2360-2367, Jun. 2010.

[17]. D. S. Cheng, M. Cristani, M. Stoppa, L. Bazzani and V. Murino, ”Custom pictorial structures for re-identification”, Proc. Brit. Mach. Vis. Conf., pp. 1-11, Sep. 2011.
[18]. S. Liao, Y. Hu, X. Zhu and S. Z. Li, ”Person re-identification by local maximal occurrence representation and metric learning”, Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 2197-2206, Jun. 2015.
[19]. S. Tan, F. Zheng and L. Shao, ”Dense invariant feature based support vector ranking for person re-identification”, Proc. IEEE Global Conf. Signal Inf. Process., pp. 687-691, Dec. 2015.
[20]. T. Matsukawa and E. Suzuki, ”Person re-identification using CNN features learned from combination of attributes”, Proc. Int. Conf. Pattern Recognit., pp. 2428-2433, Dec. 2016.
[21]. L. Wu, C. Shen and A. van den Hengel, ”PersonNet: Person re-identification with deep convolutional neural networks”, arXiv:1601.07255, 2016, [online] Available: https://arxiv.org/abs/1601.07255.
[22]. R. Varior, M. Haloi and G. Wang, ”Gated Siamese convolutional neural network architecture for human re-identification”, Proc. Eur. Conf. Comput. Vis., pp. 791-808, Sep. 2016.
[23]. Zhang et al., ”AlignedReID: Surpassing human-level performance in person re-identification”, arXiv:1711.08184, 2017, [online] Available: https://arxiv.org/abs/1711.08184.
[24]. Sun, L. Zheng, W. Deng and S. Wang, ”SVDNet for pedestrian retrieval”, Proc. IEEE Int. Conf. Comput. Vis., pp. 3800-3808, Oct. 2017.
[25]. Lin, L. Ren, J. Lu, J. Feng and J. Zhou, ”Consistent-aware deep learning for person re-identification in a camera network”, Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 3396-3405, Jul. 2017.

[26]. Zhang, Q. Zhang, X. Wei, Y. Zhang and Y. Xia, ”Person reidentification with triplet focal loss”, IEEE Access, vol. 6, pp. 7809278099, 2018.
[27]. Hu, S. Liao, Z. Lei, D. Yi and S. Z. Li, ”Exploring structural information and fusing multiple features for person re-identification”,
[28]. Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops, pp. 794799, Jun. 2013.
[29]. Bazzani, M. Cristani and V. Murino, ”Symmetry-driven accumulation of local features for human characterization and re-identification”, Comput. Vis. Image Understand., vol. 117, no. 2, pp. 130-144, 2013.
[30]. Yang, J. Yang, J. Yan, S. Liao, D. Yi and S. Z. Li, ”Salient color names for person re-identification”, Proc. Eur. Conf. Comput. Vis., pp. 536-551, Sep. 2014.