Face and Human Detection in Low Light for Surveillance Purposes

Mohit Sarin
Electronics and Communication Engg.
NIT Raipur
Chhattisgarh, India
mohitsarin98@gmail.com

Shreya Chandrakar
Electronics and Communication Engg.
NIT Raipur
Chhattisgarh, India
schandrakar.266@gmail.com

Dr. Ramnarayan Patel
Electrical Engg.
NIT Raipur
Chhattisgarh, India
rnpatel.ee@nitrr.ac.in

***Abstract*—Surveillance based on computer vision is the need of the current era. Most of the existing surveillance cameras which are being used by the Security Services and Military Forces works on the sensor-based analysis of conditions, it activates a signal indicating the presence of some movable creature and cannot accurately differentiate between a hare, a deer or a human. The current surveillance methods show high accuracy in presence of daylight but fail miserably in low light conditions because these sensors equipped with normal cameras are not able to capture images with the same accuracy in such conditions. Hence to overcome this problem, we propose a network that is a combination of Human and Face Detection systems to distinguish between different entities and enhance the accuracy of the results. This model works well in extreme conditions such as low-light, blurred images, fog, haze or when most of the body or face is covered.**

***Index Terms*—human detection, convolutional neural network, surveillance methods, face detection**

I. Introduction

What differentiates a human from a machine is its capability to differentiate between things, objects, faces, etc. But, in the past few years machines have developed the capability to do the same. There are human and face detection systems that are not only able to detect but also differentiate between different faces and postures [1]. Earlier biometrics was the only method for a machine to clearly distinguish between two different entities [2]. This included a fingerprint sensing system for all the major control access systems, but with the evidence of technology we have come a long way and iris-based access control has taken up the charge [3]. It surely doesn ́t end here with just biometrics as the only means of distinction. Now, we have Artificially Intelligent Face Recognition which can be embedded in systems as small as our Smartphones and Home Security kits.

Human Detection in the recent past has gained much recognition and in a great way has complemented Face Detection technology over a spectrum of uses. Many countries are using these technologies to detect and report accidents through an automated system that is set up at all the crowded crossings and circles [4]. It is also being successfully implemented in fields like Autonomous Driving Cars, Rescue Services and Surveillance Systems [5].

Out of all of its practical usage, what fascinated us the most was its applicability in the field of Military Surveillance. The current technology that is being used is motion sensing [6]. It is being used by most of the countries to keep a check on the borderline activities. But the precision remains a big question mark due to the extreme conditions under which it works. So, incorporating technologies like low light human detection and blurred image sensing for face detection surely will add on to the accuracy with which we can detect any activity in such sensitive regions of work.

II. Related Work

A. Face Detection

Initial face detection work [7] was focused on robust handcrafted representations and included training of powerful classifiers for the application of machine learning. Around 2016, it utilized structural dependencies present in faces and modeled them using elastic deformation structures, which seems to work well on the datasets present at that time, but its accuracy decreases drastically when tested on UFDD Dataset [8].

Many models have been proposed in the past years but the Viola-Jones face detector [9] is considered a breakthrough in the area of face detection. It uses Haar-like features, based on the AdaBoost Cascade scheme. Nowadays, researchers have achieved high accuracy by integrating multiple hand-crafted features. Headhunter [10], ACF-multiscale [11] and LDCF+ [12] are some of the works which have achieved the finest execution among the traditional strategies. These solutions are capable of real-time detection on CPU, but hand-crafted features lack the robustness to complicated face changes like illumination, expression, pose and occlusion. Therefore, these strategies may not be versatile to low-quality images.

The success of CNN-based methods in various computer vision tasks such as object tracking, autonomous driving has inspired several face detection approaches. Cascade CNN was proposed to address the issue of high variances of face detection and high computational cost. In this algorithm, negative samples are removed at early stages and then the results are refined. Zhang et al. proposed a method named ICC-CNN [13], which is similar to Cascade CNN but it rejects samples in different layers within a single CNN. The advantage of this method is its high computation speed.

The first end-to-end CNN-based object detection method was Faster-RCNN [14]. It is based on a Region Proposal Network (RPN) and a Region Classification Network (RCN). Anchor Boxes were first proposed in Faster-RCNN, and this method remains to be the base-line approach for most of the anchor box based face detectors. The architecture of the RPN network is similar to that of VGG-16 [15].

By the use of Dynamic Bayesian Network (DBN) [16], Grgic et al. set up a few cameras to click facial images and perform face recognition on it. Using a Principal Component Analysis (PCA) method [17], he mounted five cameras over an entrance and collected face information at some different locations to perform face recognition.

Algorithms based on SSD [18] has the advantage of the multi-scale feature maps, as it had prepared scale-variant detectors on diverse layers. One of the disadvantages of SSD is that it is not reasonable for identifying compact small objects, due to its default anchor design. S3FD [19], FaceBoxes [20], Scaleface [21] and HR-ER [22] have recently been proposed to resolve the anchor mismatching issue and increase the recall level of small faces by either enhancing the matching methodology and anchor densities or assigning layers with particular scale ranges.

Zhang et al. [23] suggested Single Shot Scale-Invariant Face Detector (S3FD) to solve anchor-based techniques for the detection of small objects through their new methods of anchor layout. To overcome the issue of a high false-positive rate of small faces, authors have proposed a technique called max-out background technique. S3FD is yet another algorithm for object detection based on the Single Shot Detector (SSD) [18] model for object detection, with VGG-16 [15] as the base network. To improve the detection accuracy, S3FD uses hard negative mining.

B. *Human Detection*

Layne et al. proposed a Symmetry-Driven Accumulation of local features (SDALF) approach with Metric Learning Attributes (MLA) [24].

In the paper, Analysis-Based Gait Recognition for Human Identification, by Wang, L.; Tan, T.; Ning, H.; Hu, W. Silhouette [25], a study was conducted on the detection of human identity gait based on PCA and silhouette analysis. In Statistical Feature Fusion for Gait-Based Human Recognition by Han J., Bhanu, [26], gait-based recognition was further studied and the problems arising due to insufficient gait data were resolved. By applying Multiple Discriminant Analysis (MDA) recognition performance was improved.

For vision-related problems related to face or human identification or detection, HOG methodology is widely applied. Some of its applications can be seen in cases of Pedestrian Detection, Age Estimation, Face Recognition and Gender Classification. Upon the accumulation of the strength and direction of the gradient information for all the pixels within the sub-block, this method constructs histogram features of an image sub-block. Generally, we apply the image enhancement process while doing detection in low light conditions such as night time. Several techniques have already been studied for image enhancement [27]. Traditional methodologies use Histogram Specification (HS) and Histogram Equalization (HE). One of the problems with these methods was that of an increase in the noise, which also results in a decrease of the important low-frequency components.To find a solution to these problems, numerous investigations have been done and researches are being carried out on a variety of intensity mapping-based image improvement techniques and histogram processing. Numerous investigations have been done on distinctive strategies that perform De-Noising followed by image enhancement [28]. Although, these methods have only shown positive results to increase image visibility no concrete results have been put forth that overcomes the persisting issue of human detection at night.

Huang's research has been successful in explaining human detection at night by the use of visible light images [29]. The researchers used continuous visible light video frames to perform human detection. As this method applies to continuous images of a person’s side, the only drawback that comes into light is that of an increased difficulty when the person is moving close to or far away from the camera. Since they need to process several continuous images the procedural complexity and long processing time are some other significant drawbacks. So, to overcome these problems we propose a deep CNN-based human detection model that processes upon a single image for both face and body data.

III. Proposed Method

Our proposed model is a 2-parallel set up to detect the human body as well as a human face simultaneously. Our model works well in extreme conditions such as low-light, blurred images, fog, haze or when most of the body or face is covered. So, it is well suited for surveillance in different conditions.

We have used a CNN based human detection network, with various face detection techniques. The proposed model has been experimented with the use of SSD [18], S3FD [19] and HR-ER [22] method for face detection network.

A. *The CNN Based Network*

The human detection network used in this paper is based on Convolutional Neural Network(CNN). The size of the input image generally depends on camera settings and hence may be different for different cameras. Hence, to ensure a fixed size of the image goes as input to the proposed network, normalization is performed through Bilinear Interpolation to fix the image size of 183 x 103 pixels (height x width). Most of the famous networks such as AlexNet [30] and a few others from the previous work [31] used square input images.

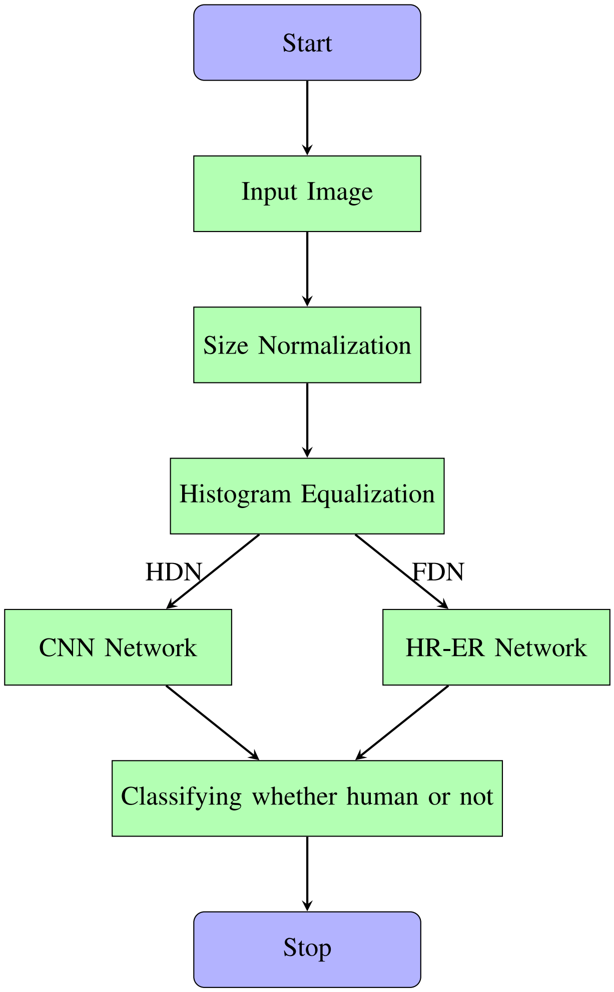

Figure 1. Overview of the model.

HDN - Human Detection Network

FDN - Face Detection Network

The focus of this research is human area, for which the width is generally smaller than height. That's why the normalization of the square-shape size is not done because it stretches the picture to a great extent with regards to its height and width, that misshapes the human region and makes it quite difficult to detect features. Also if we use square size normalization without horizontal stretching too much background area beside the targeted object (here human) will be included, which in turn results in incorrect detection of features. To prevent this, we have used uniform human or background images as the input of CNN layers with 183 x 103 pixels (height x width). For incidence when the object is either too close or far far away from the camera, size normalization proves to be the key to make up for the size change. To enhance the brightness of dark images, brightness normalization using the zero-center method was adopted.

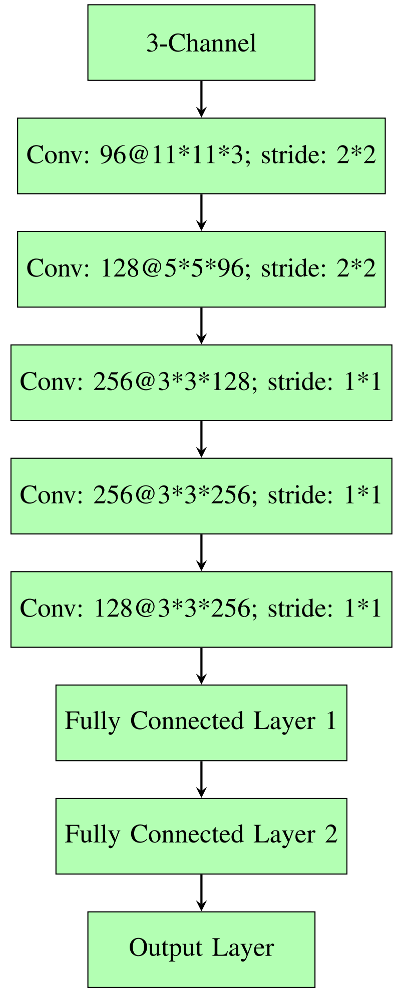

Figure 2. Proposed CNN Model.

The number of filters that are used in every convolutional layer used in our research is smaller as compared to that of the AlexNet structure. Moreover, the number of nodes in fully connected layers are also smaller than those in AlexNet. We have also compared the performances of the algorithm when Histogram Equalization was performed and when it was not. The accuracy seems to be better when Histogram Equalization was done.

The output of this CNN network gives a bounding box over the human. Hence, the image will be classified as a human and background area. Figure 2 shows the structure of CNN used in our experiment. All the details of convolutional layers and fully connected layers are mentioned in Table I.

The 1st convolutional filter size is relatively large as compared to the proposed size in previous studies [32,33] because input images are mostly present in low-quality, blurry and low-light conditions, so it has a greater amount of noise. Therefore, the filter size was increased to prevent invalid features from being produced as a result of noise. The Rectified ReLu layer is used to avoid the Vanishing Gradient Problem that might occur when a sigmoid or hyperbolic function is used as activation in back-propagation for training.

After the Relu layer, we have used the cross channel

TABLE I
PROPOSED CNN ARCHITECTURE USED IN OUR MODEL.

Layer Name	Number of Filters	Size of Feature Map	Size of Kernel	Number of Stride	Number of Padding
Image input layer		183 x 103 x 3			
1st convolutional layer	96	87 x 47 x 96	11 x 11 x 3	2 x 2	0 x 0
ReLU Layer		87 x 47 x 96			
Cross Channel Normalization Layer		87 x 47 x 96			
Max Pooling Layer	1	43 x 23 x 96	3 x 3	2 x 2	0 x 0
2nd Convolutional Layer	128	43 x 23 x 128	5 x 5 x 96	1 x 1	2 x 2
ReLU Layer		43 x 23 x 128			
Cross Channel Normalization Layer		43 x 23 x 128			
Max Pooling Layer	1	21 x 11 x 128	3 x 3	2 x 2	0 x 0
3rd Convolutional Layer	256	21 x 11 x 256	3 x 3 x 128	1 x 1	1 x 1
ReLU Layer		21 x 11 x 256			
4th Convolutional Layer	256		3 x 3 x 256	1 x 1	1 x 1
ReLU Layer		21 x 11 x 256			
5th Convolutional Layer	128	21 x 11 x 128	3 x 3 x 256	1 x 1	1 x 1
ReLU Layer		21 x 11 x128			
Max Pooling Layer	1	10 x 5 x 128	3 x 3	2 x 2	0 x 0
1st fully connected layer		4096			
ReLU Layer		4096			
2nd fully connected layer		4096			
ReLU Layer		1024			
Dropout Layer		1024			
3rd fully connected layer		1024			
Softmax Layer		2			
Classification Layer		2			

normalization layer to carry out channel-wise normalization. The input to the max-pooling layer was the feature map obtained after cross channel normalization. The output of the max-pooling layer is a 96 feature map with a size of 43 x 23.

The two initial layers were implied to extricate low-level picture highlights that included edges of picture or blob texture features. Apart from these three additional layers were brought into consideration for high-level feature extraction.

After processing of the above mentioned five convolutional layers, 128 feature maps with a size of 21 x 11 pixels were finally obtained. The output of these five convolutional layers was given to three fully connected layers, which included 4096, 1024 and 2 neurons respectively. In the 3rd fully connected layer, we have used the softmax function.

As mentioned in previous experimental results [30], the over-fitting problem is the most persistent issue that the CNN-based recognition systems suffer, which may result in low accuracy output over the testing data, despite its high accuracy with training data. To overcome this issue, we used dropout methods [30] and data augmentation techniques to avoid the problem of over-fitting. Before the 3rd fully connected layer, we have used a dropout layer.

To disengage a few neurons connection between two connected layers, the dropout method [30] was used with a dropout probability of 50%.

In the proposed model, both the networks (One for human detection and the other for face detection) were trained on a different dataset. The dataset used for training human detection networks is CVC-14 [34], DNHD-DB1 [35] and Kaist datasets [36]. And we have used the Dark Face open-source dataset [37], UFDD dataset [8] for the training of face detection network.

As per figure 1, the input image will be common for both the network and the networks will run parallelly to give its output. The output from human detection network will be analyzed first and if the human is not detected, as in some cases only face is visible (extreme condition when the human and background color will be almost similar), then face detection network will give us the final result indicating the presence or absence of the human.

For the Face Detection algorithm, we tested our model with SSD [18], S3FD [19] and HR-ER [22]. We have used the architecture as proposed by their creators. Only we have changed the size of the input, which of (160 x 160) in S3FD, [(40 x 40)-(140 x 140)] in HR-ER, but we have used (183 x 103).

The performance of different algorithms on different datasets is compared and the overall performances of the model with different networks can be seen in Table II.

After analyzing the performance of our model, different algorithms for face detection were used and upon analysis, HR-ER was adopted, as it shows better results in all the datasets we have tried with. We have used a similar architecture as proposed by Hu. It is based on RSNet-101. The input to this algorithm is the same as the input of the model as described in Fig 2.

Particularly, when a person hides beside a bush or a tree so that to avoid coming completely under camera surveillance, we can detect his tiny face and confirms that a human is present in that location. Moreover, HR-ER [22] will also work well in those cases where we want to keep the camera in a distant position so that it is not identified by a person. So, we will be having a very tiny image, and hence this algorithm outperforms the other algorithms.

Hence, it was concluded that CNN based human detection works best with HR-ER [22] face detection method.

B. Measurement Criteria

To report the results of our experiment, the human or face area is considered to be positive data whereas the background is considered negative data. To simplify the criteria, we further divide the positive and negative data into True or False. When the background is correctly detected then it gives a True Negative (TN) value high. Whereas, when a face or human is detected it gives a high True Positive (TP) value. On the contrary, if the background is inaccurately reported, we get a False Negative (FN) high and when an inaccurate human or face is recognized as a high False Positive (FP) is obtained.

Here we have used true positive rate to define errors but, rather False Negative Rate (FNR) and False Positive Rate (FPR) can also be used for the same. The FNR can be calculated as 100 −TPR (%) and FPR as 100−TNR (%).

Table II list the TPR as confusion matrices for different datasets [8,34,35,36,37].

IV. Results and Discussions

Fig. 3a

Fig. 3b

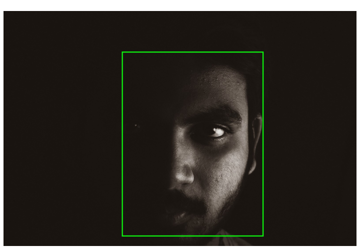

Fig. 3c

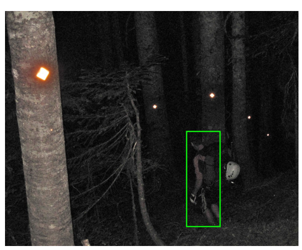

Fig. 3d

Fig. 3e

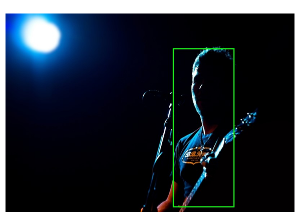

Fig. 3f

Figure 3. Output Images

The three models were tested on four different databases, mostly on the low-light images for both human and face detection network.

After analyzing the results it can be concluded that the CNN+HR-ER model performs the best, with an average TPR of 96.27%. Our Model is tested on different datasets [8,34,35,36,37]. The results and comparisons are mentioned in Table II. On human + face dataset, our model performs appreciably in haze conditions and even for blur and low-quality images.

The results obtained by our proposed model in the extreme conditions such as in a fog, dark and camouflaged background is shown in Figure 3.

Fig. 3a and Fig. 3b represents the performance of our proposed network in the extreme weather conditions and when there is not much difference between the background and human. Detection in these conditions will be very useful for military and surveillance purposes.

While testing on face related datasets [8,36], the human detection algorithm doesn't perform well as expected, hence only the output of face detection algorithm was considered and the presence of a face, indicates the presence of human too. Although this is one of the rare cases, indicating that the surveillance camera is so much nearer to the human that it can only capture its face as shown in Fig. 3c. When the captured image contains complete human, for example in Fig. 3d or 3e, our algorithm works well.

Fig. 3e indicates that our network works well in haze and fog conditions.

V. Fallacy of the Model Proposed

Fig. 4a

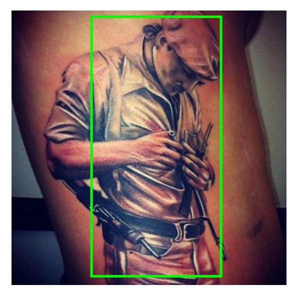

Fig. 4b

Figure 4. Fallacy of the Model

In a very dark environment and no light condition, sometimes the background and human cannot be distinguished appropriately as shown in image Fig. 4a. i.e. the image brightness is so low that it is impossible to detect human or face information even by human eyes. Some of the images where our algorithm doesn't perform well are shown in Fig. 4a and Fig. 4b.

Also in some of the cases, when a surveillance camera is too close to human or the cases where only some part of the face is visible, some extreme cases such as Fig. 4b, which represents a tattoo on a human hand which is held very close to a surveillance camera, give misleading results.

As we have proposed this model for surveillance purposes, hence images such as that of Fig. 4b will be the rarest case.

TABLE II
AVERAGE TESTING ACCURACY FOR HUMAN/FACE DETECTION

METHOD	DATABASE	TPR
Proposed CNN +HR-ER	CVC + UFDD	96.83
	KAIST + UFDD	95.93
	DNHD-DB1 + Dark	96.02
	DNHD-DB1 + UFDD	96.30
Proposed CNN +SSD	CVC + UFDD	92.80
	KAIST + UFDD	91.13
	DNHD-DB1 + Dark	91.22
	DNHD-DB1 + UFDD	91.49
Proposed CNN +S3FD	CVC + UFDD	94.21
	KAIST + UFDD	93.41
	DNHD-DB1 + Dark	93.12
	DNHD-DB1 + UFDD	93.77

VI. CONCLUSIONS

The above-proposed model performs human and face detection one after the other. All the experiments done until now was done separately for human and face, but in practical situations, we need a model that performs both the task one after the other.

So, the model integrates these two tasks and for military purposes, it will be easy to detect humans in low light, rain, haze, and blur. In cases where the human body is highly camouflaged with the background, the face recognition overcomes the issue of non-recognition of human despite it's presence in the frame.

Upon testing the performance of the various combination of networks while changing the blur, noise or contrast level. We have compared various models on publicly available datasets, for humans we have used DNHD-DB1, CVC-14 and KAIST Dataset and for faces we have used Caltech, UMIST, exclusively Dark Dataset and UFDD Dataset. On examining various combinations, we state that the combination of the proposed CNN network for the human + HR-ER method for faces gives the best result.

We have compared different combinations in term of TPR (True Positive Rate) and the results are listed in Table II.

REFERENCES

[1] H. Chung, C. Hou and S. Liang, "Face detection and posture recognition in a real time tracking system," *2017 IEEE International Systems Engineering Symposium (ISSE)*, Vienna, 2017, pp. 1-6.

[2] Sharma, Annu and Chaturvedi, Praveena and Arya, Shwetank. (2015). Human Recognition Methods based on Biometric Technologies. *International Journal of Computer Applications*. 120. 1-7.

[3] A. Kumar and A. R. Asati, "Iris based biometric identification system," *2014 International Conference on Audio, Language and Image Processing*, Shanghai, 2014, pp. 260-265.

[4] Bhumkar, S. P., V. V. Deotare, and R. V. Babar. "Accident avoidance and detection on highways." *International Journal of Engineering Trends and Technology* 3.2 (2012): 247-252.

[5] Bojarski, Mariusz, et al. "End to end learning for self-driving cars." *arXiv preprint arXiv:1604.07316* (2016).

[6] Gill, Tyler, et al. "A system for change detection and human recognition in voxel space using the Microsoft Kinect sensor." *2011 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*. IEEE, 2011.

[7] S. C. Brubaker, J. Wu, J. Sun, M. D. Mullin, and J. M. Rehg. On the design of cascades of boosted ensembles for face detection. *International Journal of Computer Vision*, 77(1- 3):6586, 2008.

[8] Nada, Hajime, et al. "Pushing the limits of unconstrained face detection: a challenge dataset and baseline results." *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. IEEE, 2018.

[9] Viola, Paul, and Michael J. Jones. "Robust real-time face detection." *International journal of computer vision* 57.2 (2004): 137-154.

[10] M. Mathias, R. Benenson, M. Pedersoli, and L. Van Gool, Face detection without bells and whistles, in *European Conference on Computer Vision*, pp. 720735, Springer, 2014.

[11] B. Yang, J. Yan, Z. Lei, and S. Z. Li, Aggregate channel features for multi-view face detection, in IJCB, pp. 18, IEEE, 2014.

[12] E. Ohn-Bar and M. M. Trivedi, To boost or not to boost? on the limits of boosted trees for object detection, in ICPR, pp. 33503355, IEEE, 2016.

[13] K. Zhang, Z. Zhang, H. Wang, Z. Li, Y. Qiao, and W. Liu, Detecting faces using inside cascaded contextual cnn, in ICCV, pp. 31713179, 2017.

[14] Girshick, Ross. "Fast r-cnn." Proceedings of the *IEEE international conference on computer vision*. 2015.

[15] Han, Song, Huizi Mao, and William J. Dally. "Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding." *arXiv preprint arXiv:1510.00149* (2015).

[16] Murphy, Kevin Patrick, and Stuart Russell. "Dynamic bayesian networks: representation, inference and learning." (2002).

[17] Wold, Svante, Kim Esbensen, and Paul Geladi. "Principal component analysis." Chemometrics and intelligent laboratory systems 2.1-3 (1987): 37-52.

[18] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, Ssd: Single shot multibox detector, in ECCV, pp. 2137, Springer, 2016

[19] S. Zhang, X. Zhu, Z. Lei, H. Shi, X. Wang, and S. Z. Li, S 3 fd: Single shot scale-invariant face detector, *arXiv preprint arXiv:1708.05237*, 2017.

[20] S. Zhang, X. Zhu, Z. Lei, H. Shi, X. Wang, and S. Z. Li, Faceboxes: a cpu real-time face detector with high accuracy, *arXiv preprint arXiv:1708.05234*, 2017.

[21] S. Yang, Y. Xiong, C. C. Loy, and X. Tang, Face detection through scale-friendly deep convolutional networks, *arXiv preprint arXiv:1706.02863*, 2017.

[22] P. Hu and D. Ramanan, Finding tiny faces, in CVPR, pp. 15221530, IEEE, 2017.

[23] S. Zhang, X. Zhu, Z. Lei, H. Shi, X. Wang, and S. Z. Li. S3fd: Single shot scale-invariant face detector.

[24] Layne, R.; Hospedales, T.M.; Gong, S. Towards Person Identification and Re-Identification with Attributes. pp. 402412.

[25] Wang, L.; Tan, T.; Ning, H.; Hu, W. Silhouette Analysis-Based Gait Recognition for Human Identification.IEEE Trans. Pattern Anal. Mach. Intel. 2003, 25, 15051518

[26] Han, J.; Bhanu, B. Statistical Feature Fusion for Gait-Based Human Recognition.pp. II-842II-847.

[27] Gonzalez, R.C.; Woods, R.E. Digital Image Processing, 3rd ed.; Prentice Hall: New Jersey, NJ, USA, 2010.

[28] Li, L.; Wang, R.; Wang, W.; Gao, W. A low-light image enhancement method for both denoising and contrast enlarging. pp. 37303734.

[29] Huang, K.; Wang, L.; Tan, T.; Maybank, S. A real-time object detecting and tracking system for outdoor night surveillance. Pattern Recognit. 2008, 41, 432444.

[30] Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks.In Advances in Neural Information Processing Systems 25; Curran Associates, Inc.: New York, NY, USA, 2012; pp. 10971105.

[31] Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. Proc. IEEE 1998, 86, 22782324.

[32] Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. pp. 114.

[33] Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S. Going deeper with convolutions. pp. 19.

[34] Gonzlez, A.; Fang, Z.; Socarras, Y.; Serrat, J.; Vzquez, D.; Xu, J.; Lpez, A.M. Pedestrian detection at day/night time with visible and FIR cameras.

[35] http://dm.dgu.edu/link.html

[36] Hwang, S.; Park, J.; Kim, N.; Choi, Y.; Kweon, I.S. Multispectral pedestrian detection: Benchmark dataset and baseline. pp. 19.

[37] https://flyywh.github.io/CVPRW2019LowLight/