Multimedia Tools and Applications
https://doi.org/10.1007/s11042-020-09579-x

Human detection and tracking with deep convolutional neural networks under the constrained of noise and occluded scenes

Ejaz Ul Haq[1] · Huang Jianjun[1] · Kang Li[1] · Hafeez Ul Haq[2]

Received: 19 April 2020 / Revised: 31 July 2020 / Accepted: 11 August 2020
Published online: 17 August 2020

Abstract
Human detection and tracking is a key aspect in surveillance system due to its importance in timely identification of person, recognition of human activity and scene analysis. Convolutional neural networks have been widely used approach in detection and tracking related tasks. In this paper, a robust framework is presented for the human detection and tracking in noisy and occluded environments with the aid of data augmentation techniques. In addition, a softmax layer and integrated loss function is used to improve the detection and classification performance of the proposed model. The primary focus is to perform the human detection task in unconstrained environments. The implemented system outperforms the state-of-the-arts methods which can be validated from the experimental results.

Keywords Human Detection · Deep learning · Data augmentation techniques

1 Introduction

Human detection and tracking is well known problem that has been studied frequently in the computer vision literature. Automatic people detection is a vital task that has many important practical applications in video surveillance system, transportation control, security system and managing crowded scenes [47]. Due to its wide applicability, it can be divided into two aspects: human verification and human identification. The anterior focus on determining whether a human exist in a set of different objects in an image and video or not, while the human identification aim to recognizes and locate the person amongst a set of various objects

✉ Huang Jianjun
huangjin@szu.edu.cn

[1] ATR Key Laboratory, School of Electronics and Information Engineering, Shenzhen University, Shenzhen, China

[2] Fujian Normal University, Fuzhou, China

[51]. Despite of enough research done in the field of human identification still it is considered a challenging task in many real time applications. Human poses are a non-rigid appearance which differentiates divergent groups of people from each other in terms of size, shape and body structure. In addition, noise, occlusions and surrounding environment greatly affect the human detection and tracking performance in a real time scenario [26].

In the recent decades, deep learning based methods are extensively used for the human identification in computer vision. Convolutional neural network has emerged as a most prevalent method for figuring out real time problems. Different vigorous and discriminative configurations of convolutional neural networks have applied in solving image processing, passenger flow calculation, crowd counting and object detection tasks [28]. In fact, neural network is a sub category of artificial intelligence having ability to learn efficient features from large dataset of human. Recently, convolutional neural networks have been proved an optimal solution for human identification in dense scenes. The most contending and significant development in human identification has been concealed in the recent era which established on convolutional neural network [9]. The convolution and pooling operations in convolutional neural networks can adequately locate and identify human using the extracted features in occluded scenes and unconstrained surroundings [34].

Large number of training data is required for CNNs to obtain most favorable and desirable output. Data augmentation (DA) is a combination of different low-complexity techniques that pervasively used in the literature to minimize the over fitting problems in training a deep neural network [46]. Data augmentation offers a feasible solution to train convolutional neural network using a dense dataset. However, images collected from videos, search engines and movies in dense datasets are often superimposed by noise, low intensity and blur effects. Under the effect of abnormal constrained, sometimes it is difficult to locate the features required for human identification which make it challenging to find human regions [11]. Human identification related task based on CNNs has acquired exceptional detection and tracking rates, however a vast area of studied done by different researchers in the literature have limitation in considering the complex scenes. DA methods have been examined by several researchers to increase the training dataset and reproduce more patterns to fulfill the obligation of deep neural network [37, 39].

In [20], the authors proposed a visual attentive model based on bottom line mechanism which aims to explore objects more effectively and proficiently than the surrounding regions on the sights or images. The main focus of the proposed model is to understand the human response of processing the visual stimuli during the detection process. In the whole process, the proposed model is independent of the image data and prior facts of the image sight. Observing the human-approach, the meaningful patches and directional content stimulate the human's visual neurons and visual attentive sense helps the human visual system to recognize and identify salient objects. On the experimental observation and results, the proposed method is considered effective and reliable compared to state-of-the-art techniques. In [21], a novel framework is proposed for underwater saliency detection which integrates together local contrast, pattern distinctness and Quaternionic Distance Based Weber Descriptor (QDWD).In the proposed method, local contrast is used to highlight the contrast difference between the salient objects and background to resolve the problems of low-contrast in underwater image or scene.PCA coordinates are used to estimate the pattern distinctness in order to minimize the impact of complex underwater lighting characteristics. Pattern distinctness can effectively reduce the influence of complex lighting in underwater image. QDWD is computed independently of images scenes in the first step to overcome the allusion associated with the underwater vision. The main purpose of computing QDWD is to obtain the directional prompts for any underwater image. In the final step, local contrast, pattern distinctness and QDWD are integrated for underwater saliency detection. Based on the experimental results, it is concluded that proposed method perform better in underwater saliency detection compare to state-of-the-art techniques. In [22], the authors introduces a database comprises of underwater images which assist in salient object detection. The database including total of 8600 underwater images of 430 individual objects, and is called Marine Underwater Environment Database (MUED). All 430 individual objects have different illuminations; pose variation, complicated backgrounds, numerous salient objects and water turbidity etc. This dataset is publicly available providing assistance to researchers in technical and academic fields whose research concern is saliency object detection using underwater images. MUED database also include manually labeled images for better understanding the application and utilization of underwater image processing and underwater computer vision. MUED database provide wide flexibility in accuracy, scale, background and ground truth information, therefore it can be broadly employed in different state-of-the-arts algorithms used for saliency object detection.

Below is the contribution of our work:

- In this paper, we proposed a method based on deep learning which is able to solve human detection and tracking issues in more complex environment mainly caused by different sizes and poses, missing parts, variation in intensity level, corrupted parts and cropped information.
- Encouraged by DA methods and to generate more complex situations in the training phase, in this paper, an aggressive data augmentation is used to reproduce a large number of complex human images, to learn deep human representation from large scale data.
- The deep CNNs model is trained with different transformations of human images in different poses, variation in intensity level, blurred and corrupted parts and noisy regions.
- The human images and the occluded images generated as result of DA methods are merged together in the training phase, to efficiently learn the deep human features.
- A softmax layer is introduced in the deep CNN model to improve the detection and classification accuracy.
- Our proposed method is compared with several methods used in the literature for the human detection and tracking and the experimental results revealed that our proposed technique outperform the state-of-the-art methods.

The contribution of the rest of the paper is follows. Section 2 presents the previous work done. Our proposed method deep CNN and modified Data augmentation method is discussed in Section 3. Section 4 presents the results and experiments. Conclusion is drawn in Section 5.

2 Related work

In this section we have overviewed the literature work published, related to human detection and tracking. Numerous works has been done in the literature which briefly explains the human detection and tracking task. Human detection and tracking has remained a popular field in computer vision which mainly consists of two parts: human detection and human identification. Many techniques have been proposed in the literature for efficient feature extraction and classification of the extracted feature to detect and track human.

Deep learning methods have efficiently combine the feature extraction and classification task in same end-to-end way. Normally, extracting human features from low intensity images requires image enhancement techniques which improves the illumination level and minimize the superimposed noise on the features. However, using enhancement methods also eliminate useful features of low frequency which can be recovered by applying histograms processing approaches [14] and many image intensity enhancement methods [5, 6, 18, 40]. Histogram of oriented gradient (HOG) has been extensively used for the people detection [36], face detection [23] and recognition and age calculation [16]. The HOG method calculates the histogram features and gradient strength of each pixel in the image.

Chahyati et al. [4] proposed a deep learning based model for the identification and tracking of multiple human in a scene. The tracking of human was done using a combination of two methods, Euclidean distance and neural network model. The proposed method in [4] achieves higher accuracy and satisfactory results. Yang et al. [50] presents a deep learning approach for the detection of human faces using multi view and numerous variation of face. Tian el al. [44] comes up with a deep learning model for the accurate detection of human in dense scenes, occluded environment and weakly labeled features. Zhang et al. [53] proposed a cascaded network based on deep learning which manipulate the implicit relation of face detection corresponding to its position in a blurred and noisy environment. Brunetti et al. [2] proposed a combination of deep learning and machine learning approach with higher accuracy and low computational time for real time human identification and tracking. However, this method suffers from low speed. Lan et al. [25] presented a method based on sparse network which focus to remove the unusual feature and the proposed model is extended into a kernel based network to minimize the non-linear similarity amongst different features. However, this method is unable to perform well on real time detection and tracking.

A lot of research has been done to enhance the human detection performance in unimpeded environment and many of them achieve good detection and tracking accuracy. Guo et al. [15] proposed a deep learning model for human recognition considering images in both visible lights and distorted scene. The proposed model achieves very effective experimental results using real time data with different intensity level. An.F et al. [1] presents a model based on the combination of CNN and long–short-term memory (LSTM) for accurate identification of different facial expressions. Lv et al. [33] presents a data augmentation based approach using artificial network to generate a large number of maligned images of face for the proficient training of deep CNN model. The experimental outputs validate the robustness and accuracy of the proposed method. Huang et al. [17] put forward an easy and adaptive approach based on deep learning for the extraction of features from most informative frames while a pixel method was used to process the easy frames. Supreeth et al. [43] presents a novel hybrid approach based on the background subtraction and clustering method for the identification of multiple objects in videos. The result shows that the proposed method obtained satisfactory classification accuracy under the effect of shadows and occlusions in the video background. Girshik et al. [13] used R-CNN method working on the forward pass phenomena for each pixel to detect and label objects in 2-dimentional images. Du et al. [8] presented a fused deep neural network design for accurate and vigorous pedestrian detection.

Yahia et al. [41] proposed a deep learning structure for human detection and tracking. The proposed deep learning structure improves the existing limitations in the state-of-the-arts and enhances the execution time of the images. The human detection systems can be used further in many real time applications. Mateus [35] presented a novel and effective approach based on deep learning for human detection tasks. In the first step, an aggregate channel features (ACF) are cascaded with deep CNN in order to achieve faster human and pedestrian detection. Then a mixture of asymmetric Gaussian function is proposed for defining human aware constraints. B.Kim et al. [3] proposed a human detection model based on deep convolutional neural network which take an image as input to classify human and pedestrians. The proposed framework based on VGG-16 architecture for human and pedestrians detection using an INRIA dataset. The experimental results shows that human are detected and the proposed method achieves better state-of-the-art accuracy. Sun et al. [48] comes up with a human detection technique based on time-division ultra-wideband multiple-input multiple-output (MIMO) which overcome the poor positioning limitation of current methods. In the first step, delay and summation imaging method is used on all the echoes of the channel. Then sidelobe interference is removed through using maximum coherence factor. Band-pass incoherence factor (BICF) and variance factor (VF) are calculated for the channel data to differentiate between static objects and human. The proposed method performs well in tracking and detection of human. The main aim of work done in [12] is to design a new classifier which is able to differentiate between human and the background in occluded environment. HOG (Histograms of oriented gradients) is jointly used with SVM (support vector machine) and IL (logic inference) for the purpose of designing new classifier. First, the image is divided into twelve parts and for each part feature is extracted and classified using SVM classifier. This scheme helps in locating the different parts of human body in image. Second, the twelve partitioned are joined together and IL classifier is used to differentiate between the background region and human. Several publicly available datasets were used to conduct the experimental work. H. Jeon et al. [19] solve the issue of human detection in harsh situation through proposing a deep learning based method integrated with triangular patterns. Triangular pattern is used to extract more detailed and stable features from the local region. Extracted features are fed into deep neural network as input for the detection of human in dense and occluded environment. A brightness aware model is proposed in K.N.Renu et al. [38] work for human detection based on deep learning. The proposed model is able to detect human in day and night scenario and in different illumination conditions.

Inspired from deep learning, we proposed a deep CNN model with softmax layer to improved human detection and tracking performance. A modified Data augmentation technique is used to train the CNN model and to learn an efficient human identification under the effect of noise and occluded scenes.

3 Proposed method based on deep CNN and data augmentation techniques

Data and information captured through camera or collected from search engines often consist of massive noise and low intensity level. These parameters considerably degrade the performance of human detection and tracking techniques. Recently deep learning significantly enhances the performance of many tracking algorithms. In this paper a deep CNN model is proposed considering different complex parameters including low luminous level, shadow effect, superimposed noise, blurred background and missed information for the appearance of human. Deep CNNs have the ability to extract useful information from massive training datasets. So a rich training datasets consist of different scenarios training set is much needed to train deep CNNs model for effective detection and tracking. Data augmentation posses valuable techniques for generating a samples training dataset with different deviation to train efficiently deep CNN model for the enhanced performance of human tracking and detection. Data augmentation techniques produce preprocessing data for training deep CNNs model. DA techniques simply transform the data to different states such as background change, variation in intensity level, flipping of certain amount of data and crop of information to some extent. In this work, DA methods are applied to human datasets to generate samples of training data with different illumination level, since after deep CNN is train with the generated samples dataset to improve the human recognition in different situations.

Recently deep learning acquired very good accuracy in image processing and objects detection. Many researchers and literature work used numerous deep architectures to train deep CNNs models. In this paper we used deep learning model based on You Only Look Once (YOLO), called YOLO9000 or YOLO v2 [10] which allow us to train jointly on detection and classification. The key idea is to use the large amount of data to explore the capacity of current detection and tracking systems. The key focus to use this architecture is to eliminate the over fitting problem by introducing batch normalization which lead to convergence of the model. Furthermore, the YOLO9000 model work on the input image size of 416*416, the filters size adjusts the model to work on higher resolution images. Table 1 gives the detail configuration of the model. The model work on 19 convolutional layers and 5 max pooling layers. 1*1 filter sizes are used to compress the features between convolutional layers of 3*3 sizes [29]. YOLO9000 model works on the idea of anchor boxes which is used to detect the bounding boxes. YOLO9000 model scans the entire image at once and detect the human by creating a bounding box. This model is able to detect both small objects as well as large objects in an image. K-means clustering algorithm is used to select the best priors over the bounding box in the training dataset. K-mean clustering algorithm is run several times over the training dataset and the average value is selected. Softmax layer is used to predict that images in dataset that are mutually exclusive. However, Data augmentation techniques aid in presenting samples dataset in different variations, so the softmax layer doesn't assume the mutual exclusive phenomena in the dataset, thus improve the overall classification accuracy. K-means algorithm gives more error on larger bounding boxes compare to small boxes using Euclidean distance.

The technical novelty of the proposed work in this paper includes a hybrid method based on YOLO9000 and data augmentation techniques for the human detection task. YOLO9000 model is used to detect and track human in more complex situations including low luminous level, shadow effect, superimposed noise, blurred background and missed information for the appearance of human, which are generated as a result of data augmentation to evaluate the performance of proposed network. Literature studies revealed that YOLO9000 is solely used for object and human detection and tracking task. However, using YOLO9000 with data augmentation improves its detection and tracking accuracy performance in occluded and blurred environment. We also introduced softmax layer in the YOLO9000 CNN model which further improves the detection and classification performance. An integrated softmax loss and localization loss is used as loss function in this work which improves the discriminative power of deeply extracted features. Figure 1 demonstrates the working flow of our proposed framework in this paper.

The following details are important to discuss in the proposed model:

3.1 Data augmentation

We have used different data augmentation methods in the proposed work to reduce over fitting and maximize the size of data. Various data augmentation techniques such as translation,

Table 1 Detail Configuration of the Proposed Model

Layer	Filters	Output	Stride
Convolutional Layer	32	224*224	3*3
Maxpooling Layer		112*112	2*2/2
Convolutional Layer	64	112*112	3*3
Maxpooling Layer		56*56	2*2/2
Convolutional Layer	128	56*56	3*3
Convolutional Layer	64	56*56	1*1
Convolutional Layer	128	56*56	3*3
Maxpooling Layer		28*28	2*2/2
Convolutional Layer	256	28*28	3*3
Convolutional Layer	128	28*28	1*1
Convolutional Layer	256	28*28	3*3
Maxpooling Layer		14*14	2*2/2
Convolutional Layer	512	14*14	3*3
Convolutional Layer	256	14*14	1*1
Convolutional Layer	512	14*14	3*3
Convolutional Layer	256	14*14	1*1
Convolutional Layer	512	14*14	3*3
Maxpooling Layer		11*11	2*2/2
Convolutional Layer	1024	11*11	3*3
Convolutional Layer	512	11*11	1*1
Convolutional Layer	1024	11*11	3*3
Convolutional Layer	512	11*11	1*1
Convolutional Layer	1024	11*11	3*3
Convolutional Layer	1000	11*11	1*1
Softmax Layer			

rotation and flipping are frequently used to increase the number of images in training dataset for the training of YOLO9000. Input images are rotated on different angles such as 90,180 and 270 degree in rotation based data augmentation methods. In flipping based data augmentation methods, the input images are mirrored from horizontal and vertical direction. Literature studies consider image translations as data augmentation method but it gives wrong result due to its limitation in detection and classification tasks.

3.2 YOLO9000

YOLO9000 is an object detection method that can detect over 9000 objects classes in real time. It is more accurate and faster detector compared to past works. This model use batch normalization to improve the convergence and eliminate different form of regularization techniques which in turns help in convergence. Significant improvement can be seen in accuracy on adding batch normalization in convolutional layers. Batch normalization phenomena also overcome the dropout problem without over fitting. This version of YOLO has the ability to adjust its filter size so that it works better and properly on high resolution input. All fully connected layers are removed in YOLO9000 and use anchor boxes to predict the bounding boxes. One Pooling layer is removes so that to get high resolution output from convolutional layer. Anchor boxes improve the prediction ability of the model more than thousand per image. Different values of scales and sizes used for anchor boxes, were predefined without taking into context any prior data or information. With anchor boxes, there can be seen slight improvement in recall for YOLO9000 compare to other previous model which aim that YOLO9000 has more room to improve.

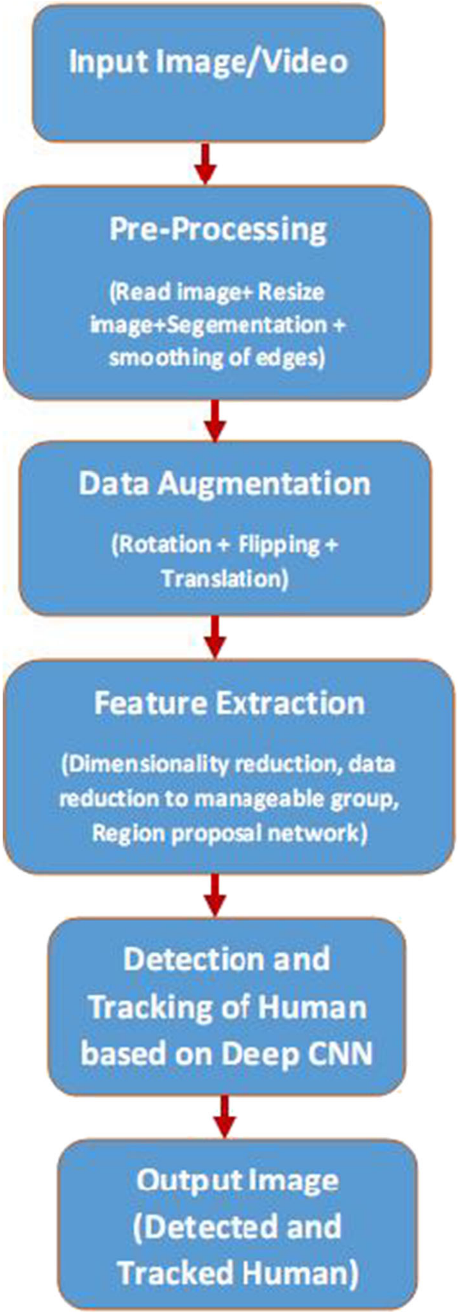

Fig. 1 Framework of our proposed method

K-means clustering algorithm is effectively used giving much better intersection over union score for object detection. The proposed model of YOLO9000 offer accurate detection and also it is considered fast than state-of-the-art networks. Many detection algorithms based on VGG-16 as feature extractor model but it offers complexity. This version of YOLO based on

GoogleNet architecture and it requires fewer operations to process an image than other previous architecture. It can be concluded that YOLO9000 can keep good balance between the accuracy and the complexity offered by the network.

Prediction and classification is jointly trained by the proposed YOLO9000 mechanism. Labeled images are used for the detection related tasks such as predicting the coordinates of bounding boxes. These labeled images are also used to predict the specific category of each image. However during the training process, the detection and classification images are mixed based on WordTree graph which make YOLO9000 a stronger model to detect and classify each class of object.

3.3 K-mean clustering

It is considered one of the finest unsupervised machine learning algorithms that is extensively used to solve clustering problems. The algorithm follows a simple way to classify certain data by assuming a fixed number of k clusters. K-centers are defined, each k-center belong to a specific cluster. In order to obtain much better result, the best way is to replace these centers far away from each other. In the next step, each point of a specific dataset is placed to its nearest center. The early stage grouping is done when no point is remained without association to a center. K new centriods are re-calculated after getting a cluster from previous step process. Same data points and the nearest centers are bind together after getting k new centriods. A loop can be observed after repeating this process and k centers constantly changing its location and at one stage there can be no further changes noticed from k centers. This algorithm aims at minimizing the square error function which can be given by:

$$Z(V) = \sum_{i=1}^{c} \sum_{j=1}^{c_i} \left(\left|\left| a_i - b_j \right|\right| \right)^2 \tag{1}$$

Where $V = \{v_1, v_2, v_3, \ldots, v_c\}$ represent center of clusters, $\|a_i - b_j\|$ represent the Euclidean distance between a_i and b_j. c_i shows the total number of data points and c shows the number of cluster center.

K-mean clustering algorithm is fast, robust and relatively effective and efficient method.

3.4 Loss function

In this work total loss function is the sum of localization loss and softmax loss. During the training of data samples using YOLO9000 deep model, an integrated softmax loss and localization loss is used as loss function for the deep model, which simply improves the classification accuracy. The loss function improves the discriminative power of the deeply extracted features. Cross validation is used to keep balances between the two losses and predict the total number of matched bounding boxes.

Softmax loss and cross entropy loss are two interchangeable term commonly used. Cross entropy loss is referred as softmax loss when softmax is used as output of the convolutional neural network model. Therefore Softmax loss can be given by Eq. (2)

$$L_{softmax} = -\sum_{i=1}^{c} y_i . log\widehat{y}_i \tag{2}$$

Where $\widehat{y}_i$ represent the $i-th$ scalar value of the model output, y_i represent the target value.

The localization loss can be calculated by estimating the smooth L1 loss between the ground true box and the predicted bounding box value. The smooth L1 loss is simply used for obtaining bounding box on object or human detection system. The localization loss can be formulated from following expression (3)

$$l_{loc} = \sum_{i \in x,y,w,h} smooth_{L1}\left(t_i^u - v_i\right) \tag{3}$$

$$smooth_{L1} = \begin{cases} 0.5x^2, & if\ |x| < 1 \\ |x| - 0.5, & otherwise \end{cases}$$

v represent the ground true bounding box and t_u shows the value for predicted bounding box of human class. The width and height of bounding box can be represented by w and h respectively while x and y shows the left corner location of bounding box.

The total loss function can be given by following equation:

$$LF = L_{softmax} + L_{loc} \tag{4}$$

LF denotes the loss function, $L_{softmax}$ represent the softmax loss and L_{loc} shows the localization loss.

4 Results and experiments

INRIA human dataset It is publicly reachable and standard dataset used for the human detection and tracking. The INRIA human dataset introduced by Navneet Dalal and Bill Trigs and it consist of training and testing data. In our experiments we used 2416 images for training the model and 1126 images for testing purposes. The people in this dataset are usually in standing position having different orientation, pose and occluded background and also much of these images are taken in crowd scenes.

Pedestrian parsing on surveillance scenes (PPSS) dataset This dataset consist of total 3673 images generated from 171 videos of different scenes. 2064 images of people in this dataset are occluded with different orientation, poses and having variation while 1609 images are not occluded. This dataset is larger than currently available datasets about 20 times. Images from first 100 video surveillance scenes are used for training the proposed model while images from last 71 video scenes are used for testing purpose.

Four important parameters are studied to evaluate the performance of the proposed convolutional neural network, that are true negative (tn), true positive (tp), false positive (fp) and false negative (fn). Further, these four outcomes are used to compute the evaluation metrics. The efficiency and effectiveness of the proposed deep CNN model is measured based on these evaluation metrics. The effectiveness and performance of the proposed method is measures through following indices.

Specificity It defines ability of model to precisely classify and detect the actual brain tumor and can be expressed by following mathematical relation

$$Specificity = \frac{tn}{tn + fp} \tag{5}$$

Sensitivity It defines the ability of the model to perfectly classify the relevant brain tumors and below mathematical relation defines it

$$Sensitivity = \frac{tp}{tp + fn} \tag{6}$$

Precision It calculates the true positive measures and can be computed from following expression

$$Precision = \frac{tp}{tp + fp} \tag{7}$$

Accuracy It determines the ability of system to differentiate the different types of brain tumor correctly. Following relation is used to calculate the accuracy

$$Accuracy = \frac{tp + fn}{tp + tn + fp + fn} \tag{8}$$

F1 measure It is measure of a test's accuracy which considers both precision and sensitivity. The following mathematical expression defines F1 measure:

$$F1\ measure = 2 * \frac{sensitivity * precision}{(sensitivity + precision)} \tag{9}$$

In the current study, the performance of the proposed deep CNN model is compared with different machine learning and deep learning models used for human detection. Head-mounted projective display (HMPD), twin background model (TBGM) + histogram oriented gradient (HOG) and VGG-16 are the three model used to validate the performance of the proposed deep CNN model on INRIA human and Pedestrian Parsing on Surveillance Scenes (PPSS) dataset. Figures 2 and 3 shows the confusion matrix of different methods on INRIA human dataset and PPSS dataset, respectively. Table 2 shows the evaluation metrics in terms of sensitivity, specificity, precision and accuracy for different methods on INRIA human dataset and PPSS dataset. It can be observed from Table 2 that our proposed model achieves higher detection accuracy on both INRIA human dataset and PPSS dataset, which prove the effectiveness of our model.

The effect of data augmentation techniques on INRIA human dataset and Pedestrian Parsing on Surveillance Scenes (PPSS) dataset has also been evaluated. The performance and accuracy of the proposed method is compared with different state of the art methods on INRIA human and Pedestrian Parsing on Surveillance Scenes (PPSS) datasets [32]. The proposed method is also validated under the effect of softmax loss and localization loss. We use publicly available datasets to evaluate the effectiveness of the proposed method used in this work.

We focus to detect and track human using different datasets with noisy and occluded environment. For this purpose we have generated multiple scenarios images with different intensity level, flipping, illumination level, noise and occluded background. We used collection of data augmentation techniques to change original data and reproduce new dataset with unconstrained surroundings. Figure 4 shows collection of different data augmentation methods which are used in our work. Our proposed deep learning model is trained with images having

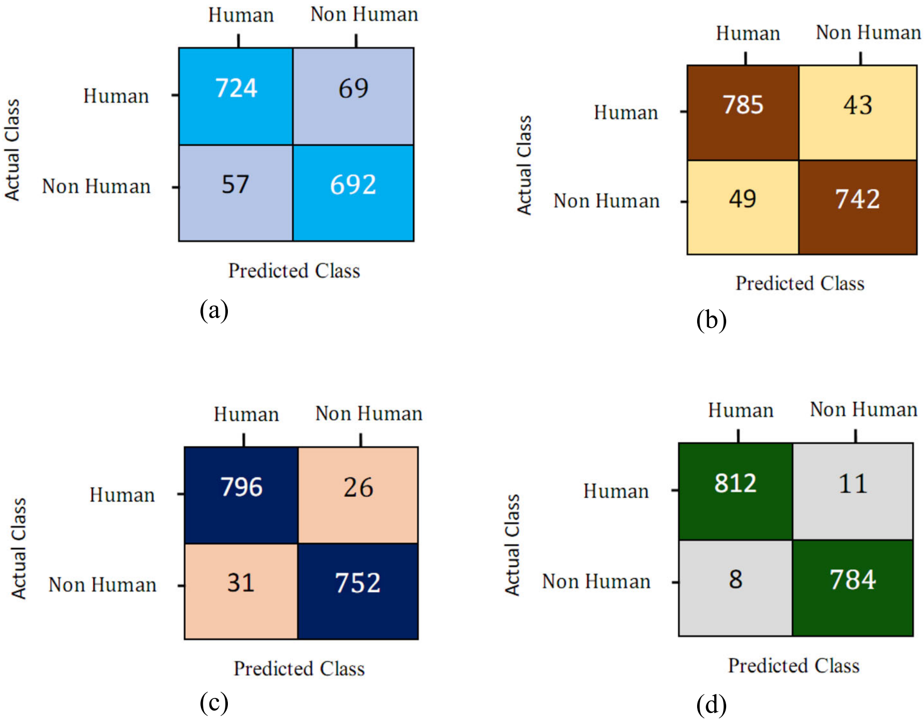

Fig. 2 Confusion matrix of different methods on INRIA human dataset (**a**) HMPD method (**b**) TBGM+HOG (**c**) VGG-16 (**d**) Our proposed Method

different intensity level as given in Fig. 4a. Furthermore, combination of DA techniques such as flipping, noise and histogram are proposed to train our model. The resultant images are given in Fig. 4b. Fig. 4c shows images with cropped information to efficiently learn the features in training phase. Added to this, occluded human images are also used to train our model as shown in Fig. 4d. The algorithm aims at detection and tracking of human in blurred and occluded environment. INRIA human and Pedestrian Parsing on Surveillance Scenes (PPSS) datasets are used in this work. After applying data augmentation methods on INRIA human and Pedestrian Parsing on Surveillance Scenes (PPSS) datasets, we conducted several experiments to detect and track human in these images having complex situation. Figure 5 shows the result of the proposed method to detect and track human in occluded and blurred environment.

Table 3 shows the performance of the proposed deep CNN model on both INRIA human and Pedestrian Parsing on Surveillance Scenes (PPSS) datasets in terms of specificity, sensitivity, precision and accuracy.

Table 4 consolidates the overall performance of proposed deep CNN model on both INRIA human and Pedestrian Parsing on Surveillance Scenes (PPSS) datasets.

4.1 Training of the proposed method

We used GPU GeForce GTX 1080 Ti to train our model for human detection and tracking. A publicly available deep learning model is implemented and trains it on Tensor-Flow. Different

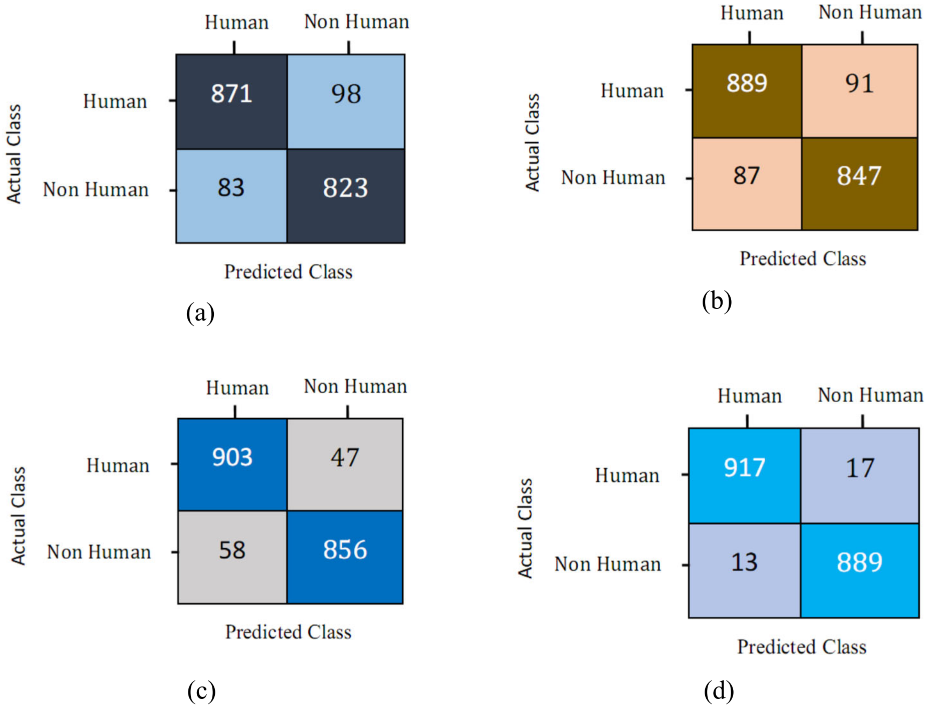

Fig. 3 Confusion matrix of different methods on PPSS dataset (**a**) HMPD method (**b**) TBGM+HOG (**c**) VGG-16 (**d**) Our proposed Method

data augmentation techniques and the combination of softmax loss and normalization loss contribute to achieve enhanced accuracy and detection performance. YOLO9000 deep learning model is used for the detection and tracking of human using the train and test dataset as it outperform the state-of-the-arts techniques. Deep learning model is trained over the dataset generated by different data augmentation methods in unconstrained environment for 135 epochs. Several data augmentation methods and YOLO9000 is integrated together in this work to enhance the overall generalization performance of the deep network. Over fitting issue in the loss function is limited by utilizing weight decay parameter. 0.05 is selected initially as a nominal value for learning rate and later reduces by the factor of 0.1. The coefficient of

Table 2 Evaluation metrics in terms of sensitivity, specificity, precision and accuracy for different methods on INRIA human dataset and PPSS dataset

Metrics Dataset	Method	TP	TN	FP	FN	Sensitivity	Specificity	Precision	F1 score	Accuracy
INRIA human dataset	HMPD	724	692	57	69	0.912	0.923	0.927	0.919	91.8%
	TBGM+HOG	785	742	47	43	0.948	0.940	0.943	0.952	94.4%
	VGG-16	796	752	31	26	0.968	0.960	0.962	0.966	96.4%
	Our Proposed	812	784	8	11	0.986	0.989	0.990	0.969	98.8%
PPSS dataset	HMPD	871	823	83	98	0.893	0.908	0.898	0.896	90.7%
	TBGM+HOG	889	847	87	91	0.907	0.906	0.910	0.908	90.8%
	VGG-16	903	856	58	47	0.950	0.936	0.939	0.943	94.3%
	Our Proposed	917	889	13	17	0.981	0.985	0.986	0.983	98.3%

(a) Examples with different intensity level

(b) Examples with Flipping, Blur, Noise and Histogram

(c) Examples with cropped regions

(d) Examples with occluded region

Fig. 4 **a**, **b**, **c**, **d** Images reproduced using different Data Augmentation method

Fig. 5 **a**, **b**, **c** Human detection and tracking based on proposed method

momentum is also selected as 0.8. Cosine distance is used to calculate the similarity index after converting the features to different dataset using data augmentation methods.

Table 3 Performance of the proposed deep CNN model on both INRIA human and PPSS datasets

Metrics Method	Dataset	Sensitivity	Specificity	Precision	Accuracy
Proposed Deep CNN Architecture	INRIA human Dataset	0.986	0.989	0.990	98.8%
	PPSS Dataset	0.981	0.985	0.986	98.3%

Table 4 Overall performance of proposed deep CNN model on INRIA human and PPSS dataset

Metrics Dataset	Epoch	Training accuracy	Validation accuracy	Training loss	Validation loss
INRIA human Dataset	5	0.83	0.837	0.94	0.77
	10	0.856	0.862	0.57	0.45
	15	0.912	0.917	0.42	0.36
	20	0.946	0.951	0.35	0.23
	25	0.957	0.963	0.15	0.07
	30	0.968	0.977	0.055	0.035
	35	0.973	0.980	0.021	0.017
	40	0.985	0.989	0.0125	0.013
PPSS Dataset	5	0.859	0.868	0.87	0.74
	10	0.868	0.873	0.63	0.58
	15	0.926	0.931	0.38	0.35
	20	0.944	0.950	0.21	0.16
	25	0.961	0.968	0.03	0.01
	30	0.975	0.981	0.024	0.017
	35	0.980	0.987	0.0113	0.0112
	40	0.986	0.990	0.0110	0.0109

The softmax loss and normalization loss is integrated in this work. Several experiments implemented on INRIA human dataset and Pedestrian Parsing on Surveillance Scenes (PPSS) datasets. Firstly both datasets were trained to check the accuracy using only software loss. Later both softmax loss and localization loss were used to evaluate the effectiveness on training both datasets. It can be observed from the results in Table 5 that the combination of both losses achieves higher accuracy compare to using both losses separately.

4.2 Training of the proposed method with data augmentation techniques

The combination of different data augmentation techniques are used to train our proposed network to evaluate the effectiveness and reliability of the deep model. We used six different combinations to train our model and assess the accuracy on INRIA human and Pedestrian Parsing on Surveillance Scenes (PPSS) datasets. Following data augmentation methods were used in this work:

1. Different intensity level
2. Flipping, Blurring, Noise and Histogram
3. Cropped Region
4. Occluded area
5. Combination of intensity level, Flipping, Blurring, Noise and Histogram, Cropped Region, Occluded area

Table 5 Effect of Loss function on the Accuracy

Loss	Accuracy %	
	INRIA Dataset	PPSS Dataset
Softmax Loss	98.17	98.43
Localization Loss	98.54	98.61
Softmax Loss+ Localization Loss	99.29	99.32

6. No augmentation method

Table 6 presents the detail classification accuracy of using different data augmentation methods on INRIA human and Pedestrian Parsing on Surveillance Scenes (PPSS) datasets. We achieved only 91.3% classification rate on INRIA human dataset with no data augmentation methods. For case 1, 3 and 4 the classification rate increases gradually. Combining different data augmentation techniques, we can achieve improved classification rate. This can be observed from case 2 and 5. For case 5, we have achieved highest classification rate through combining multiple DA techniques. Thus it can be deduced that using more DA techniques the classification performance can be considerably enhanced. Combination of data augmentation techniques can be used to enhance the human detection and tracking performance in unconstrained environment.

For the Pedestrian Parsing on Surveillance Scenes (PPSS) dataset, it can also be observed that the classification accuracy decreases to 90.4% with no data augmentation techniques, whereas the classification accuracy progressively increases on fusing more data augmentation methods. For case 5 and 2, highest classification accuracy is achieved which can be observed from Table 3. Accordingly, our proposed deep model achieve better classification rate on both INRIA human and Pedestrian Parsing on Surveillance Scenes (PPSS) datasets. Figure 6 represent the graphical representation of the accuracies achieved through using data augmentation techniques on INRIA human and Pedestrian Parsing on Surveillance Scenes (PPSS) datasets.

5 Comparison of the proposed method with state-of-the-art techniques

In this section we analyze the comparison of state-of-the-art techniques used for human detection and tracking with our proposed method. Our main focus is to detect and track human in more difficult environment. We used combination of different data augmentation techniques in order to generate more complex human images in the training data which further enables the deep network to learn human detection features in unconstrained situations. Several studies have been done in the literature for human detection and tracking having different classification and accuracy rate. Authors in [27] adopted scale aware Fast R-CNN method which can detect human from different and multiple ranges. The proposed framework integrates the large-scale sub networks and small scale sub networks into a single architecture. Finally,

Table 6 Classification accuracy using different data augmentation methods on INRIA human and Pedestrian Parsing on Surveillance Scenes (PPSS) datasets

Data Augmentation Techniques	Accuracy %	
	INRIA Dataset	PPSS Dataset
1.Different Intensity Level	91.5	92.1
2.Flipping, Blurring, Noise and Histogram	97.9	98.2
3.Cropped Region	94.7	94.8
4.Occluded area	95.2	95.7
5.Combination of intensity level, Flipping, Blurring, Noise and Histogram, Cropped Region, Occluded area	99.3	98.8
6.No augmentation method	91.3	90.4

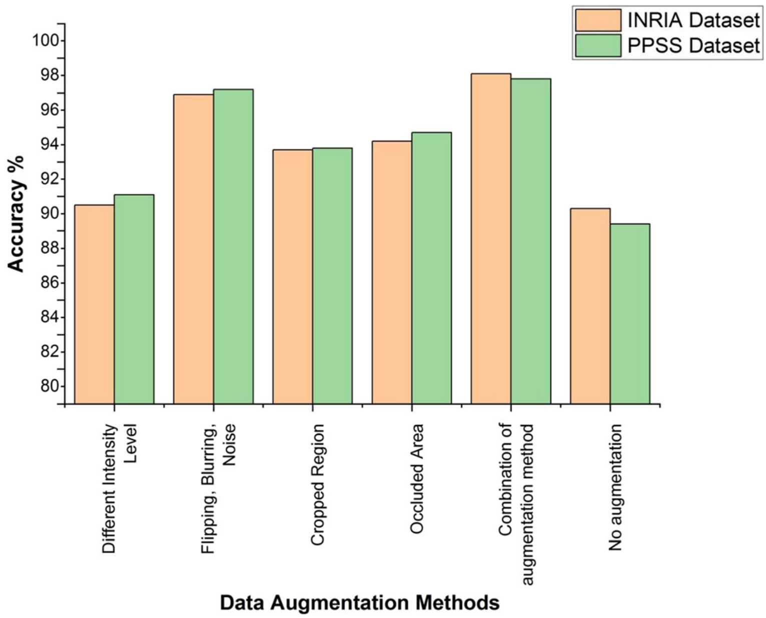

Fig. 6 Graphical representation of Accuracies obtained through Data Augmentation Techniques

detection results are generated combining the outputs of all sub networks. In [31] authors presented a single shot multibox detector (SSD) method for pedestrian detection. The SSD convolutional neural network extracts the shallow features and then fuses this extracted information with the convolutional layer of deep semantic information. Finally the human are detected in the static images. Pre-selection boxes are used in the proposed method having different ratios which improve the detection ability of the overall model. Authors in [52] present a novel baseline based on region proposal network and boosted forest method for human detection tasks. In the proposed pipeline, the main focus of region proposal network was to compute the convolutional features map and bounding boxes. A cascaded boosted forest was adopted to classify the features extracted by region proposal network. Apparently, scheme used in [52] lead to effective output and results. In [7] authors proposed a hybrid method based on V-Disparity and convolutional neural network for human detection and tracking application. V-Disparity method is used to extract region of interest and fed into convolutional neural network as an input. In the proposed work, convolutional neural network are combined with recursive neural networks to effectively learn extracted features and classify them. The authors in [30] proposed a method based on Faster R-CNN combined with skip pooling to handle the human detection problems. The region proposal network architecture of Faster R-CNN is extended to multi-layer framework and then integrates with skip pooling. Skip pooling network extract multiples region of interest from lower layers and then fed these to higher layers without taking into account the intermediate layers. The work in [24] proposed a compression method which based on teacher-student framework and conventional random forest (RF) for doing human detection task. In the compression network, the student shallow RF is trained by mimicking the teacher RF performance through using the soft version of teacher RF output. Authors in [42] used L1 norm based human detection network which achieves a total of 17% increase in accuracy. C.Xu et al. [49] proposed improved mask R-

CNN method for fast human detection in real time which obtained an accuracy of 88%. W.wang et al. [45] proposed a significant method based on two streams unified deep network for human detection as it consume less computational time and enhanced human detection accuracy. Although different human detection algorithms have limitation of detecting humans in heavily occluded scenes. The possible solution to this mentioned problem is to include multiple human images in dense and occluded environment in order to train the deep learning model. We focus in our work to include different difficult human images using data augmentation methods such as flipping, cropping, histogram and noisy information to generate these images. Besides, including complex images in the dataset we achieve significant improved detection rate of 99.3% on INRIA dataset and 98.8% on PPSS dataset using a combination of multiple DA techniques. This reveal that deep model has capability of learning complex features and obtain considerable better detection rate under the constrained of imperfect human data. Table 7 shows the comparative analysis of the proposed method with state of the art techniques.

6 Conclusion

In this paper we have evolved a deep learning model for the robust human detection and tracking under the constrained of occluded environment. We have used combination of different data augmentation techniques for generating complex circumstances for the human appearance. Softmax loss and localization loss is used as a fused loss function which simply improved the reliability and efficiency of the deep model. The experimental results on INRIA human dataset and Pedestrian Parsing on Surveillance Scenes (PPSS) datasets shows that the proposed model improves the human detection and tracking in the noisy and occluded environment. Finally we obtain 99.3% on INRIA human dataset and 98.8% on Pedestrian Parsing on Surveillance Scenes (PPSS) datasets using deep learning framework which prove the efficiency of our proposed technique. In future, this work can be further proceed following the initial study is to do experiments with different hyper parameters and analyzing the performance with using different network architectures. In next phase we will utilized the results of human detection achieved in this work for human action recognition to identify each specific activity of human. We will also extend the proposed method for object detection in occluded and noisy environment. Different models such as VGG 16, ResNet etc. can also be

Table 7 Comparison with state-of-the-art methods

Method	Accuracy %
J.Li et al. [27]	90.68%
S.Liu et al. [31]	90.2%
Zhang et al. [52]	93.1%
D.Tran et al. [7]	88.9%
J.Liu et al. [30]	91%
S.Kim et al. [24]	27.60%
A.Salvaraj et al. [42]	92%
C.Xu et al. [49]	88%
W.Wang et al. [45]	91.4%
Our Method	99.05%

trained on the pedestrian dataset to evaluate the performance of all models. Furthermore, we will also train different pedestrian datasets (for example Caltech) on various deep models and test the performance. Moreover, for future perspective we plan to enhance the detection and tracking performance of our proposed method in a scene where both human and background are similar or pedestrian is occluded with other objects. At the moment, proposed method achieved quite rational computational speed without minimizing the detection and tracking accuracy but still it is needed to apply on embedded system. In addition, we also plan to integrate the proposed method with Visual Saliency in order to detect and track motion patterns of pedestrian and human in more effective way, which have many real time applications for video surveillance.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. An F, Liu Z (2019) Facial expression recognition algorithm based on parameter adaptive initialization of CNN and LSTM. Vis. Comput. 35:1–16. https://doi.org/10.1007/s0037 1-019-01635 -4
2. Brunetti A, Buongiorno D, Francesco G, Bevilacqua V (2018) Neurocomputing computer vision and deep learning techniques for pedestrian detection and tracking : A survey. Neurocomputing 300:17–33
3. Bubryur Kim, N Yuvaraj, KR Sri Preethaa,R Santhosh, A Sabari (2020). Enhanced pedestrian detection using optimized deep convolution neural network for smart building surveillance, Soft Computing.https://doi.org/10.1007/s00500-020-04999-1
4. Chahyati D, Fanany MI, Arymurthy AM (2017) Tracking people by Detection Using CNN features. Procedia Comput Sci 124:167–172
5. Coltuc, D; Bolon, P (1999). Strict ordering on discrete images and applications. In Proceedings of the IEEE International Conference on Image Processing, Kobe, Japan, 24–28 October 1999; pp. 150–153
6. Coltuc D, Bolon P, Chassery J-M (2006) Exact histogram specification. IEEE Trans Image Process 15: 1143–1152
7. Tran Thi Dinh, Nguyen Dinh Vinh, Jeon Jae Wook (2018). Robust Pedestrian Detection via a Recursive Convolution Neural Network, SNPD 2018, June 27–29 2018, Busan
8. X Du, M El-Khamy, J Lee, and L Davis (2017). “Fused dnn: A deep neural network fusion approach to fast and robust pedestrian detection,” in WACV
9. Dundar A, Jin J, Martini B, Culurciello E (2017) Embedded streaming deep neural networks accelerator with applications. IEEE TransNeural Netw & Learning Syst 28(7):1572–1583
10. Farhadi, A and Redmon, J (2016). YOLO9000: better, Faster, Stronger
11. Felzenszwalb PF, Girshick RB, McAllester D, Ramanan D (2010) Object detection with discriminatively trained part-based models. IEEE Trans. Pattern Anal Mach Intell 32:1627–1645
12. Flores Calero MJ, Aldás M, Lázaro J, Gardel A, Onofa N, Quinga B (2019) Pedestrian Detection under partial occlusion by using logic inference, HOG and SVM. IEEE Latin America Transactions 17(09):1552–1559. https://doi.org/10.1109/TLA.2019.8931190
13. R Girshick, J Donahue, T Darrell, and J Malik (2014). “Rich feature hierarchies for accurate object detection and semantic segmentation,” in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 580–587
14. Gonzalez, RC; Woods, RE (2010). Digital image processing, 3rd ed.; Prentice Hall: New Jersey
15. Guo K, Wu S, Xu YF (2017) Face recognition using both visible light image and near-infrared image and a deep network. CAAI Trans Intell Technol 2(1):39–47
16. Hajizadeh, MA; Ebrahimnezhad, H (2011). Classification of age groups from facial image using histograms of oriented gradients. In Proceedings of the 7th Iranian Conference on Machine Vision and Image Processing, Iran University of Science and Technology (IUST), Tehran, Iran, 16–17 November 2011; pp. 1–5
17. Huang, C, Lucey, S, Ramanan, D (2017). Learning policies for adaptive tracking with deep feature cascades. In: Computer Vision Foundation

18. Jen, TC; Hsieh, B; Wang, SJ (2005). Image contrast enhancement based on intensity-pair distribution.In Proceedings of the IEEE International Conference on Image Processing, Genova, Italy, 11–14 September 2005; pp. 1–4
19. H Jeon, VD Nguyen and JW Jeon (2019). "Pedestrian Detection Based on Deep Learning," IECON - 45th annual conference of the IEEE industrial electronics society, Lisbon, Portugal, 2019, pp. 144–151, doi: https://doi.org/10.1109/IECON.2019.8927417
20. Jian M, Lam K, Dong J, Shen L (2015) Visual-patch-attention-aware saliency Detection. IEEE Transactions on Cybernetics 45(8):1575–1586. https://doi.org/10.1109/TCYB.2014.2356200
21. Jian M, Qi Q, Dong J, Yin Y, Lam K-M (2018) Integrating QDWD with pattern distinctness and local contrast for underwater saliency detection. Journal of Visual Communication and Image Representation, Volume 53:31–41. https://doi.org/10.1016/j.jvcir.2018.03.008
22. Jian M, Qi Q, Yu H, Dong J, Cui C, Nie X, Zhang H, Yin Y, Lam K-M (2019) The extended marine underwater environment database and baseline evaluations. Applied Soft Computing, Volume 80:425–437. https://doi.org/10.1016/j.asoc.2019.04.025
23. Karaaba, M; Surinta, O; Schomaker, L; Wiering, MA (2015). Robust face recognition by computing distances from multiple histograms of oriented gradients. In Proceedings of the IEEE Symposium Series on Computational Intelligence, Cape Town International Convention Center, Cape Town, South Africa, 7–10 December 2015; pp. 203–209
24. Kim S, Kwak S, Ko BC (2019) Fast Pedestrian Detection in surveillance video based on soft target training of shallow random Forest. IEEE Access 7:12415–12426. https://doi.org/10.1109/ACCESS.2019.2892425
25. Lan X, Ma AJ, Yuen PC, Chellappa R (2015) Joint sparse representation and robust feature-level fusion for multi-cue visual tracking IEEE trans. Image Process 24(12):5826–5841
26. Lei, Z, Chu, R, He, R, Liao, S, Li, SZ (2007). Face recognition by discriminant analysis with Gabor tensor representation. In: International Conference on Biometrics, pp. 87–95. Springer, Berlin
27. Li J, Liang X, Shen S, Xu T, Yan S (2018) Scale-aware fast R-CNN for pedestrian detection. IEEE Trans Multimedia 20(4):985–996
28. R Lienhart and J Maydt (2002). "An extended set of haar-like features for rapid object detection," in ICIP
29. M Lin, Q Chen, and S Yan (2013). Network in network. arXiv preprint arXiv:1312.4400
30. J Liu, X Gao, N Bao, J Tang and G Wu (2017), "Deep convolutional neural networks for pedestrian detection with skip pooling," 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, pp. 2056–2063, doi: https://doi.org/10.1109/IJCNN.2017.7966103
31. Shu-an Liu, Shi Lv, Hailin Zhang, Jun Gong (2019). Pedestrian Detection Algorithm Based on the Improved SSD, The 31th Chinese Control and Decision Conference (2019 CCDC), IEEE
32. P Luo, X Wang, and X Tang (2013). Pedestrian parsing via deep Decompositional neural network, in proceedings of IEEE international conference on computer vision (ICCV)
33. Lv JJ, Cheng C, Tian GD, Zhou XD, Zhou X (2016) Landmark perturbation-based data augmentation for unconstrained face recognition. Signal Process Image Commun 47:465–475
34. Madbouly AMM, Mostafa M-SM, Wafy M (2015) Performance assessment of feature detector-descriptor combination. Int J ComputSciIssues 12(5):87–94
35. Mateus A, Ribeiro D, Miraldo P, Nascimento JC (2019) Efficient and robust Pedestrian Detection using deep learning for human-aware navigation. Robot Auton Syst 113:23–37
36. Pawłowski, P, Piniarski, K (2015). D ˛abrowski, A. Pedestrian detection in low resolution night vision images. In Proceedings of the IEEE Signal Processing: Algorithms, Architectures, Arrangements, and Applications, Pozna ´n, Poland, 23–25 September 2015; pp. 185–190
37. J Redmon, S Divvala, R Girshick, and A Farhadi (2016). "You only look once: unified, real-time object detection," in CVPR
38. KN Renu Chebrolu and PN Kumar (2019). "Deep learning based Pedestrian Detection at all light conditions," 2019 International Conference on Communication and Signal Processing (ICCSP), Chennai, pp. 0838–0842, doi: https://doi.org/10.1109/ICCSP.2019.8698101
39. D Ribeiro, A Mateus, JC Nascimento, and P Miraldo (2016). "A real-time pedestrian detector using deep learning for human-aware navigation,"arXiv:1607.04441
40. Rivera AR, Ryu B, Chae O (2012) Content-aware dark image enhancement through channel division. IEEE Trans Image Process 21:3967–3980
41. Yahia Fahem Said, Mohammad Barr (2019). Pedestrian Detection for Advanced Driver Assistance Systems using Deep Learning Algorithms, IJCSNS International Journal of Computer Science and Network Security, VOL.19 No.9, September 2019
42. Anandamurugan Selvaraj, Jeeva Selvaraj ,Sivabalakrishnan Maruthaiappan, Gokulnath Chandra Babu, Priyan Malarvizhi Kumar (2020). L_1 norm based pedestrian detection using video analytics technique, An international journal of computational intelligence, 22 February 2020. https://doi.org/10.1111/coin.12292

43. Supreeth HSG, Patil CM (2018) Efficient multiple moving object detection and tracking using combined background subtraction and clustering. Signal Image Video Process 15:1097
44. Y Tian, P Luo, X Wang, and X Tang (2015). "Deep learning strong parts for pedestrian detection," in ICCV
45. Wentong Wang , Lichun Wang , Xufei Ge, Jinghua Li and Baocai Yin (2020), Pedestrian Detection Based on Two-Stream UDN, Appl. Sci. , 10, 1866; doi:https://doi.org/10.3390/app10051866
46. Wang, X, Wang, K, Lian, S (2019). A survey on face data augmentation. arXiv :1904.11685
47. Wojek C, Dollar P, Schiele B, Perona P (2012) Pedestrian detection:An evaluation of the state of the art. IEEE Trans. Pattern Anal. Mach.Intell. 34(4):743
48. Xinxin S, Liangnian J, Qinghua L (2019) Detection of stationary humans using time-division UWB MIMO through-wall radar. The Journal of Engineering 2019(20):6799–6802. https://doi.org/10.1049/joe.2019.0542
49. Chenchen Xu, Guili Wang, Songsong Yan,Jianghua Yu, Baojun Zhang, Shu Dai, Yu Li, and Lin Xu, Fast Vehicle and Pedestrian Detection Using Improved Mask R-CNN (2020). Hindawi, Mathematical Problems in Engineering,Volume 2020, Article ID 5761414, 15 pages. https://doi.org/10.1155/2020/5761414
50. S Yang, P Luo, CC Loy, and X Tang (2015). "From facial parts responses to face detection: A deep learning approach," in ICCV
51. Z Yang and R Nevatia (2016). "A multi-scale cascade fully convolutional network face detector," in ICPR
52. Zhang, L Lin, X Liang, K He (2016). Is faster r-cnn doing well for pedestrian detection? In: European Conf. Computer Vision (ECCV), pp. 443–457.
53. Zhang K, Zhang Z, Li Z, Qiao Y (2016) Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Process Lett 23(10):1499–1503

Ejaz Ul Haq received B.S degree in Electrical Engineering from University of Engineering and Technology, Peshawar, Pakistan in 2014 and finished M.S. degree in Electrical Engineering from Xiamen University of Technology, China. He is currently PhD student. His research areas include Image Processing, Data Mining, Computer Vision, AI, power system, smart grid and Control system.

Huang Jianjun received both his Bachelor and Ph.D degree from Xidian University. Now he is a professor of Shenzhen University. His research interests include compressed sensing, intelligent recognition technology, and embedded systems.

Kang Li is s currently Lecturer in Shenzhen University. Her Research area includes medical imaging, compressed sensing and MRI.

Hafeez Ul Haq is currently Master degree student in Fujian Normal University. His research area include Compressed sensing, Molecular Biology and MRI.