

## Convolutional Bi-LSTM Based Human Gait Recognition Using Video Sequences

Javaria Amin<sup>1</sup>, Muhammad Almas Anjum<sup>2</sup>, Muhammad Sharif<sup>3</sup>, Seifedine Kadry<sup>4</sup>, Yunyoung Nam<sup>5,\*</sup> and ShuiHua Wang<sup>6</sup>

<sup>1</sup>University of Wah, Wah Cantt, 47040, Pakistan

<sup>2</sup>National University of Technology (NUTECH), Islamabad, 44000, Pakistan

<sup>3</sup>COMSATS University Islamabad, Wah Campus, Wah Cantt, Pakistan

<sup>4</sup>Faculty of Applied Computing and Technology, Noroff University College, Kristiansand, Norway

<sup>5</sup>Department of Computer Science and Engineering, Soonchunhyang University, Asan, 31538, Korea

<sup>6</sup>Department of Mathematics, University of Leicester, Leicester, UK

\*Corresponding Author: Yunyoung Nam. Email: ynam@sch.ac.kr

Received: 12 January 2021; Accepted: 14 February 2021

**Abstract:** Recognition of human gait is a difficult assignment, particularly for unobtrusive surveillance in a video and human identification from a large distance. Therefore, a method is proposed for the classification and recognition of different types of human gait. The proposed approach is consisting of two phases. In phase I, the new model is proposed named convolutional bidirectional long short-term memory (Conv-BiLSTM) to classify the video frames of human gait. In this model, features are derived through convolutional neural network (CNN) named ResNet-18 and supplied as an input to the LSTM model that provided more distinguishable temporal information. In phase II, the YOLOv2-squeezeNet model is designed, where deep features are extricated using the fireconcat-02 layer and fed/passed to the tinyYOLOv2 model for recognized/localized the human gaits with predicted scores. The proposed method achieved up to 90% correct prediction scores on CASIA-A, CASIA-B, and the CASIA-C benchmark datasets. The proposed method achieved better/improved prediction scores as compared to the recent existing works.

**Keywords:** Bi-LSTM; YOLOv2; open neural network; resNet-18; gait; squeezeNet

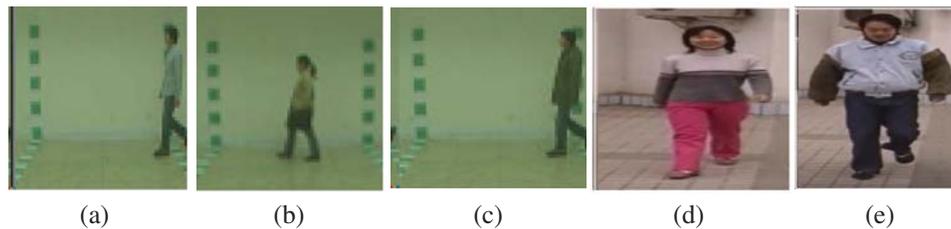
### 1 Introduction

Gait biometric represents a person walking styles and more powerful as compared to other biometrics [1] i.e., iris, palmprint, face, and fingerprint [2], etc. Therefore, it can be utilized for person identification from a long-distance [3]. Human gait with different styles is illustrated in Fig. 1. Gait recognition methodologies have attained more attention in the last two decades in real-time applications such as forensic identification, video surveillance, and crime investigation [4]. In literature, some research works proposed improved feature vectors to discriminate the gait patterns



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

based on the motion [5–7]. The recognition of human body parts in motion is achieving more attention from researchers [8]. However, it is a more challenging and difficult task to accurately track each part of the human body [9]. The appearance-based gait recognition methodologies commonly utilized human silhouettes as input.



**Figure 1:** Human gait with different actions at  $90^\circ$  (a) normal (b) human with a bag (c) human wearing a coat (d) woman (e) man

These approaches might obtain maximum recognition scores when there is less variation in consecutive frames. When the variation increased in the consecutive frames, the performance of these algorithms decreased in real-time applications [10].

The gait features are drastically changed in case of different variations i.e., illumination, view, clothing, and carrying [11]. Model-based features are utilized to track the human body parts and movement [12–14]. The main contribution of the presented approach is based on feature vectors that are extracted from LSTM and ResNet-18 model. The extracted feature vectors contain more prominent discriminative information to classify the different types of human gaits based on fully connected and softmax layers. Furthermore, in phase II classified images are recognized using a proposed modified YOLOv2-ONNX model, which consists of 20 layers that are configured by applying the open neural network (ONNX) model and SqueezeNet architecture as the base-network of the tinyYOLOv2 model. The best recognition results are achieved by extracting deep features using the fireconcat-02 layer to the squeezeNet architecture and further fed as an input to the YOLOv2 model. The proposed method accurately recognizes the different kinds of human gaits.

## 2 Related Work

Several machine learning approaches are used in the literature for human gait recognition (HGR) [15]. For HGR, features play a vital role to extract the discriminant information. Modified Local Optimal Oriented Pattern (MLOOP) features are extracted for HGR, and selected best features from MLOOP features vector [16]. The histogram oriented gradient (HOG) with Harlick features are combined for HGR and tested on the CASIA (A–B) datasets [17]. The Gabor wavelet features are extracted from the input images in different orientations [18] for HGR. The method performance is computed on CASIA (A and B) datasets [19]. The multi-scale LBP and Gabor features are extracted and selected the best features by spectra discriminant analysis-based regression method [20–22]. Principle component analysis (PCA) along with gait energy image (GEI) feature vectors are utilized for human identification [23]. However, it is difficult to recognize the variations in frames such as clothing, angle, and view [24]. To improve the recognition results, the fusion of structural gait profile and the energy shifted image is performed [25]. The deep features are extracted [26] using pre-trained AlexNet and VGG-19 and fused using skewness & entropy. The informative features are selected by the FECS method for HGR. The method is

evaluated on CASIA A, B, and C datasets [27]. The gait flow image & Gaussian image features are extracted to create a features vector and fed to the extended neural network classifier for HGR [28,29]. The stacked progressive work autoencoders (SPAЕ) model is employed for gait recognition at different angles and views, in which some temporal information is missing [30]. GaitSet is applied for the extraction of invariant features for action recognition. The component-based frequency features are extracted for the identification of human actions [31]. The temporal features among the frame might obtain improved results as compared with the GEI [32]. However, classifying the cross-clothing and cross-carrying conditions is still a difficult activity due to changes in human shape and appearance [33]. Feng et al. [34] extracted the heat map from the joint of the human body in an RGB input image instead of utilizing a binary silhouette. The extracted heat maps are supplied further to the LSTM model for temporal feature extraction. Recently, the skeleton and joints of the body are also utilized for the recognition of person identification [35]. It is observed that gait recognition with higher accuracy is still a challenging task [36].

### 3 Proposed Methodology

The proposed model contains two phases; robust feature extraction and classification is a challenging task for human gait recognition. Therefore, in phase I, the Conv-BiLSTM model is developed, in which deep features are extracted from the localized images using Resnet-18 and supplied to the LSTM network to classify the different types of human gaits. In phase II, input images are passed to the proposed YOLOv2-Squeeze model, which extracts deep from the fireconcat-02 layer of the squeeze-Net model and is supplied as an input to the tinyYOLOv2 model for localization/recognition of the different types of human gaits. The proposed model steps are displayed in Fig. 2.

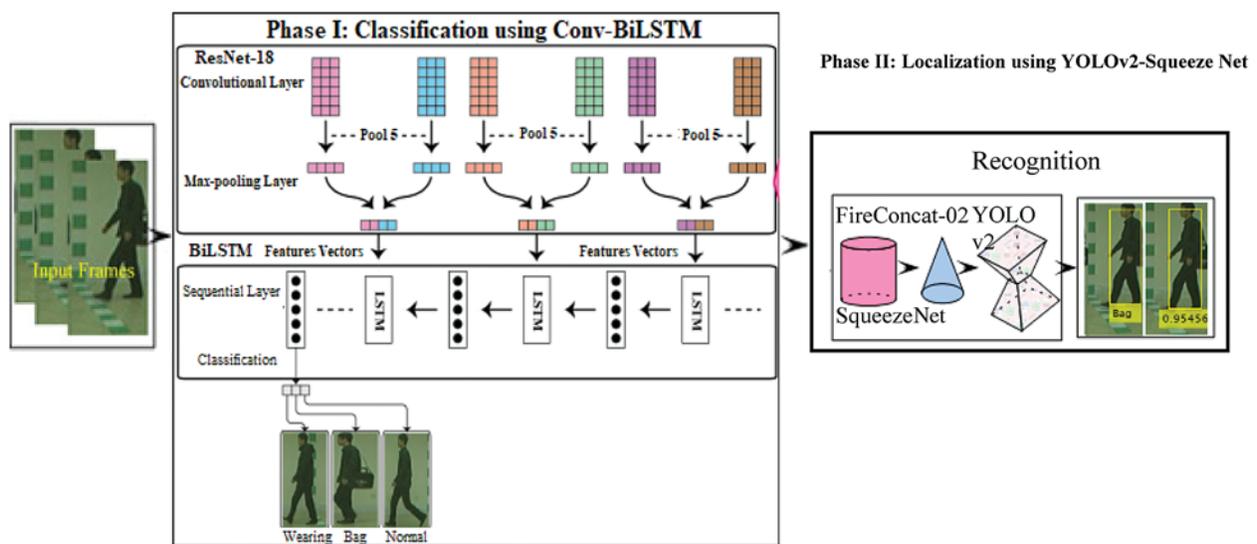


Figure 2: General proposed approach steps

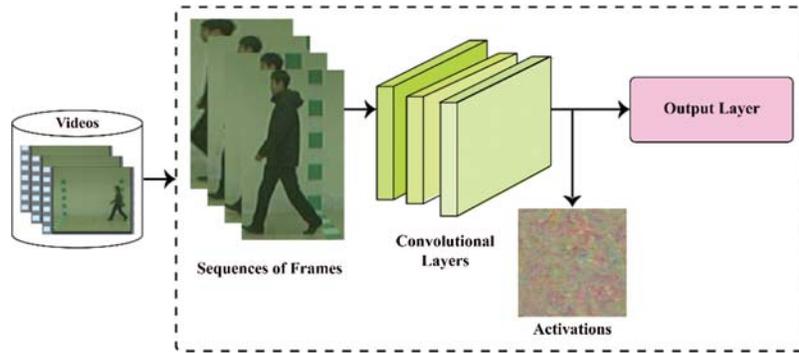
#### 3.1 Proposed Conv-BiLSTM Model for Classification of the Localized Images

The video frames are classified using the proposed Conv-BiLSTM model, in which deep features are extracted from the input frames by the CNN model such as Resnet18. Next, the

sequence structures are restored and output is reshaped into sequence vectors using the unfolding sequence layer. After that, resultant vector sequences are created using BiLSTM and output layers. Finally, assembled both networks into a single network.

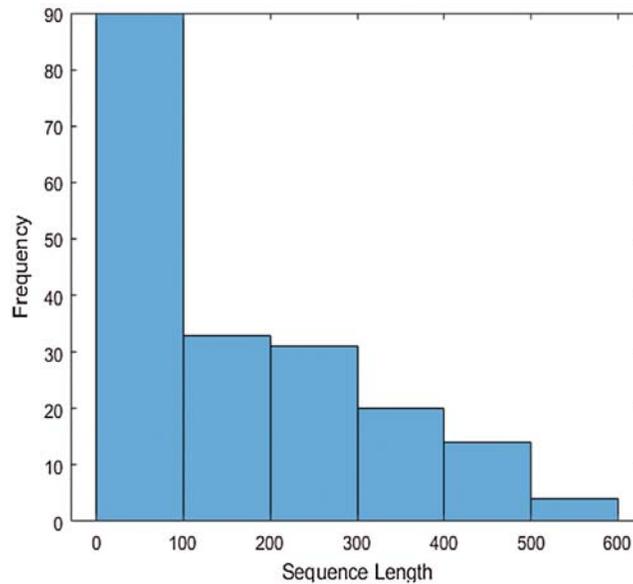
### 3.2 Convolutional Neural Network

The convolutional layers extract the feature vectors from the localized images. These feature vectors are used as the input of the activations function on the last pooling layer of the Resnet18 model as shown in Fig. 3. In the training phase, the model creates padding due to a large sequence of frames which has a negative impact on the accuracy of the gait classification.



**Figure 3:** Feature vectors extraction using Resnet-18 model

To overcome this problem, the classification results are improved by removing the sequences with more than 600-time steps with class labels. The bar length of the histogram represents the selected sequences in Fig. 4.



**Figure 4:** Visualization of the training data sequences

### 3.3 Bidirectional Long Short-Term Memory (BiLSTM) Model

The modified BiLSTM model is used for the classification of human gaits, in which LSTM layers are used for more efficient temporal feature learning. The selection of hyperparameters for model training is done after the extensive experiment as given in [Tab. 1](#).

**Table 1:** Experiment for parameter selection for model training

Hidden units	Minibatch size	Accuracy
1000	4	0.89
1500	8	0.94
<b>2000</b>	<b>16</b>	<b>0.98</b>
2500	32	0.94

[Tab. 1](#), shows the experiment of the parameter's selection, where 2000 hidden Units, 16 batch size is used for the further experiment because increase/decrease the HU obtained accuracy is decreased. The Hyperparameters of the BiLSTM model are stated in [Tab. 2](#).

**Table 2:** Selected hyperparameters of BiLSTM

Hidden units	2000
Dropout	0.5
MiniBatchSize	16
Gradient threshold	2
InitialLearnRate	1e-4
Maximum epochs	30

The model specification is as: Sequence input (1024 dimensions), LSTM layers (2000 hidden units (HU)), 50% dropout, fully connected layers, softmax, and a classification layer. The activation functions of the proposed BiLSTM model are mentioned in [Tab. 3](#).

**Table 3:** BiLSTM layers with corresponding activations

Layers	Activation functions	Learnable
Input sequence	1024	–
BiLSTM	4000	Weights = $16000 \times 1024$ Recurrent bias = $16000 \times 2000$ Bias = $16000 \times 1$
Drop	4000	–
FC	3	Weights = $3 \times 4000$ Bias = $3 \times 1$
Softmax	–	–
Classification	–	–

The LSTM [37] cell has four gates, i.e., input, forget, output gate, and cell candidate. In the LSTM block, three weights are learnable, i.e., input  $f$ , recurrent weights  $R_W$ , and bias  $b$ . The matrices of the learnable weights are expressed mathematically as:

$$f_W = \begin{bmatrix} f_W^{input} \\ f_W^{forget} \\ f_W^{cell\ candidate} \\ f_W^{Output} \end{bmatrix}, \quad Recurrent_W = \begin{bmatrix} Recurrent_W^{input} \\ Recurrent_W^{forget} \\ Recurrent_W^{cell\ candidate} \\ Recurrent_W^{Output} \end{bmatrix}, \quad b = \begin{bmatrix} b_{input} \\ b_{forget} \\ b_{cell\ candidate} \\ b_{Output} \end{bmatrix} \quad (1)$$

The cell state  $c_t$  at the time step ( $t$ ) is written as:

$$cell\ state_t = F_t \odot cell\ state_{t-1} + I_t \odot cell\ candidate_t \quad (2)$$

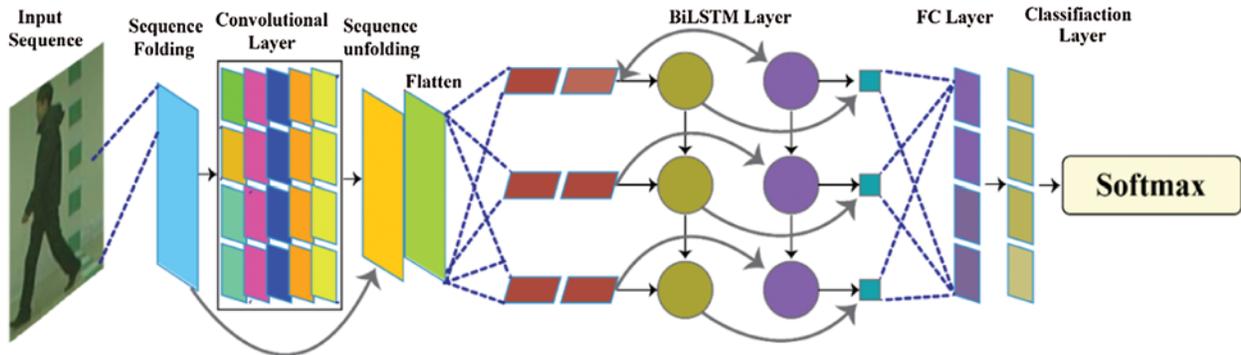
where  $\odot$  denotes Hadamard product. The hidden state  $h_t$  is represented as:

$$hidden\ state_t = Output_t \odot tanh(cell\ state_t) \quad (3)$$

In the LSTM model, based on time steps, feature vectors are computed through LSTM layers and supplied to the next block. The  $n$ th block output is used for the class label prediction, in which HU follows the fully connected, softmax, and the output layers.

### 3.4 Concatenation of CNN and LSTM Models

In the proposed model, LSTM layers are concatenated with CNN layers, in which frames are transformed into a sequence of vectors to classify the human gaits. Fig. 5, shows the steps of the assembled network.

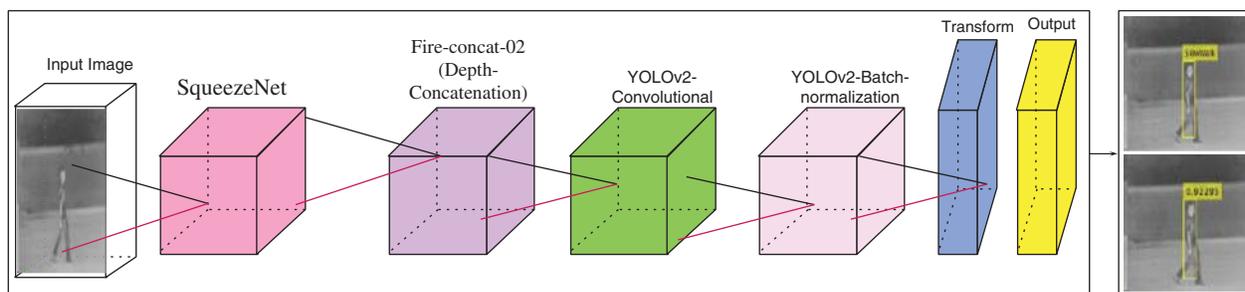


**Figure 5:** Proposed Conv-BiLSTM model

In Fig. 5, input sequences are passed to the convolutional layers, where features are extracted by convolutional operators. The convolutional layers follow the sequence folding layer. The sequence unfolding layer is followed by the flatten layer in which the structure of the sequences is restored and output is reshaped into a vector. The gait classification is performed using the output of BiLSTM followed by fully connected and softmax layers.

### 3.5 Localization of Human Gait Using YOLOv2-SqueezeNet Model

YOLOv2 is fast and effective as compared with recurrent neural network (RCNN) and SSD detectors. Therefore, in this research, YOLOv2-SqueezeNet model is suggested for different types of human gait localization such as female, male, fast walk, slow walk, walk with the bag, normal, and wearing as shown in Fig. 6.



**Figure 6:** YOLOv2-SqueezeNet model for localization

Fig. 6, shows proposed YOLOv2-SqueezeNet model, where features are extracted from fireconcat-02 layer of the SqueezeNet model and passed as an input to the pre-train YOLOv2 detector. The proposed model more accurately localized the required regions with class labels.

The model consists of the 20 layers in which 01 input image, 04 convolutional, 04 ReLU, 01 depth concatenation, 01 max-pooling, of the squeezeNet model, and 02 YOLOv2 convolutional, 02 YOLOv2 batch-normalization, 02 YOLOv2ReLU, 01 YOLOv2 transforms, and 01 YOLOv2-output of the YOLOv2 model. The activation functions of the YOLOv2-SqueezeNet model are shown in Tab. 4. Tab. 5, presents the training hyperparameters.

**Table 4:** Layer wise activations of YOLOv2-SqueezeNet model

Layers	Activations
Data (input image)	$130 \times 130 \times 3$
Conv	$64 \times 64 \times 64$
ReLU	$64 \times 64 \times 64$
Max pool	$31 \times 31 \times 64$
Fire2-squeeze (convolution)	$31 \times 31 \times 16$
Fire2-squeeze (ReLU)	$31 \times 31 \times 16$
02-Fire2-expend (convolution)	$31 \times 31 \times 64$
02-Fire2-expend (ReLU)	$31 \times 31 \times 64$
Fire2-concatenation	$31 \times 31 \times 128$
02-YOLOv2 (convolution)	$31 \times 31 \times 128$
02-YOLOv2 (ReLU)	$31 \times 31 \times 128$
02-YOLOv2 (batch-normalization)	$31 \times 31 \times 128$
YOLOv2-classification	$31 \times 31 \times 2$
YOLOv2-transform	$31 \times 31 \times 2$
YOLOv2-output	—

**Table 5:** Hyperparameters of YOLOv2-SqueezeNet model

Size of the mini-batch	14
Epochs	1000
Frequency validation	50
Rate of the learning	0.001

Tab. 5 presents the hyperparameters that are selected to configure the proposed model for human gait classification, in which mini-batch size is selected 14, 1000 epochs are used for model training because greater than equal to the 1000 epochs model results are consistent.

#### 4 Experimental Setup

Gait recognition is a great challenge due to complex recognition patterns that have been utilized in different fields such as machine learning, robotics, studying, biomedical, visual surveillance, and forensic. Therefore, intelligent recognition and the digital security group designed CASIA (A, B & C) datasets in the national pattern recognition laboratory [38–44].

The presented study is implemented on Matlab 2020RA Toolbox using a Core-i7 desktop Computer with a 740 K Nvidia Graphic Card. 0.5 hold out validation is used for model training. The description of the number of training and testing images are mentioned in Tab. 6.

**Table 6:** Description of training and testing number images in the corresponding datasets

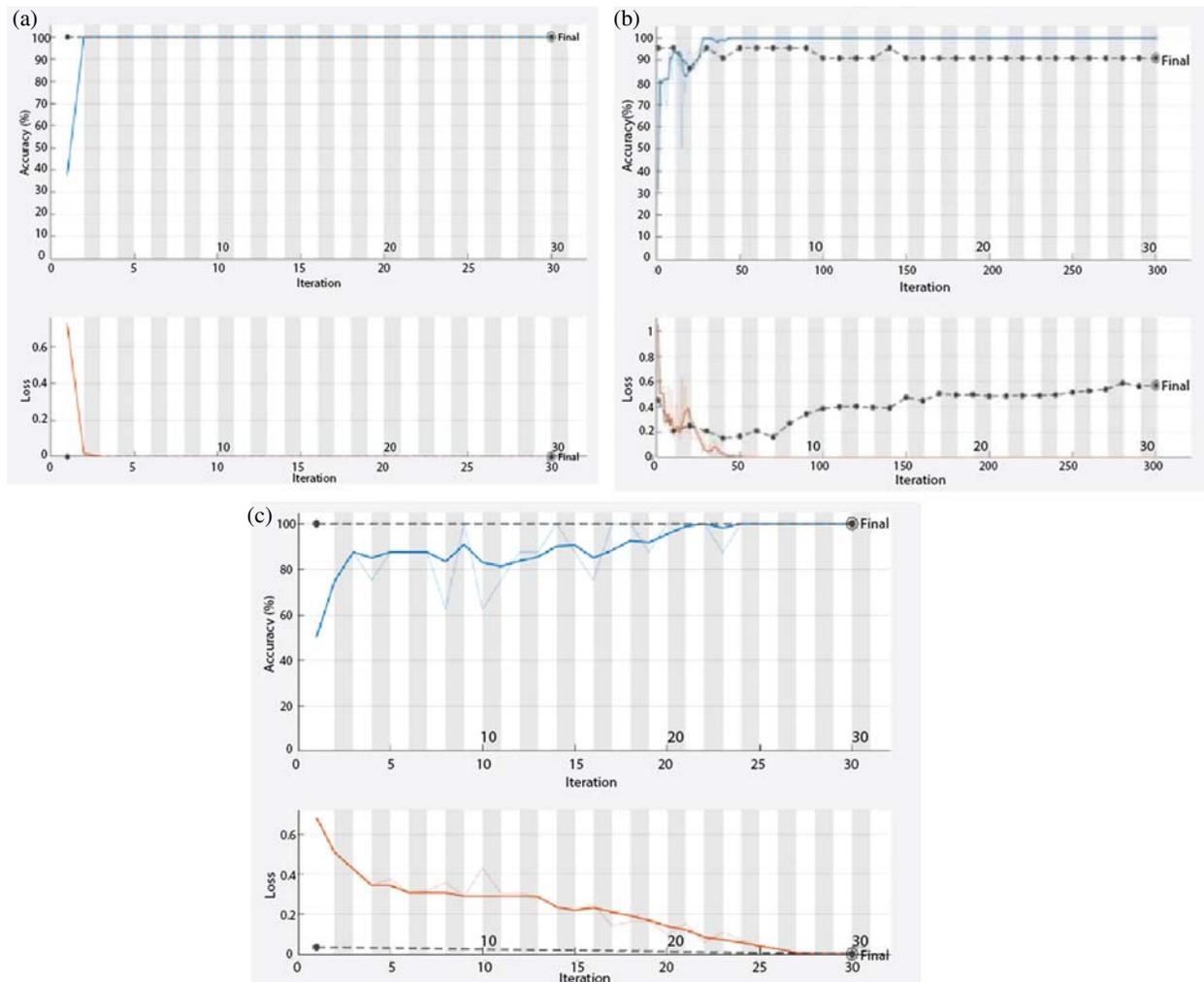
Datasets	Number of images in training and testing
CASIA-A	Total number of: 19139 Training images: 9569 Testing images: 9569
CASIA-B	Total number of: 51036 Training images: 22518 Testing images: 22518
CASIA-C	Total number of: 91800 Training images: 45900 Testing images: 45900

#### 4.1 Results and Discussion

In the developed framework, implement two experiments for the analysis of the proposed approach performance. The first experiment is performed to compute the performance of the YOLOv2-ONNX model and the second experiment is performed for classification results.

#### 4.2 Experiment #01

In this experiment, extracted feature vectors using the Conv-BiLSTM model are passed to the softmax layer for the classification of different types of human gaits such as female/male, bag, wearing, normal, and fast walk, slow walk, normal walk classes of the CASIA-A, CASIA-B and CASIA-C datasets respectively. Fig. 7, represents the proposed approach performance.



**Figure 7:** Training/testing results with respective loss rate (a) CASIA-A (b) CASIA-B (c) CASIA-C (blue line shows training, red shows loss rate, and dotted black line represent validation accuracy)

In Fig. 7, the proposed model achieved 1.00 validation accuracy (VA) on CASIA-A and CASIA-C datasets, while it achieved 0.96 VA on the CASIA-B dataset.

The classification outcomes are stated in the Tabs. 7–9.

Tab. 7, shows experimental results on CASIA-A dataset proposed method achieves 1.00 CPR on two classes of female/male.

**Table 7:** Proposed method results for human gaits recognition on different datasets using CASIA-A dataset

Dataset	Classes	SE	SP	PPV	FPR	FNR	NPV	CPR
CASIA-A	Female	1.00	1.00	1.00	0.00	0.00	1.00	1.00
	Male	1.00	1.00	1.00	0.00	0.00	1.00	1.00

**Tab. 8**, CASIA-B dataset is considered for performance evaluation, where three classes such as Bag, wearing, and normal are involved. The method achieved 0.92 CPR in bag class, 1.00 CPR on wearing, and 0.88 CPR in the normal class.

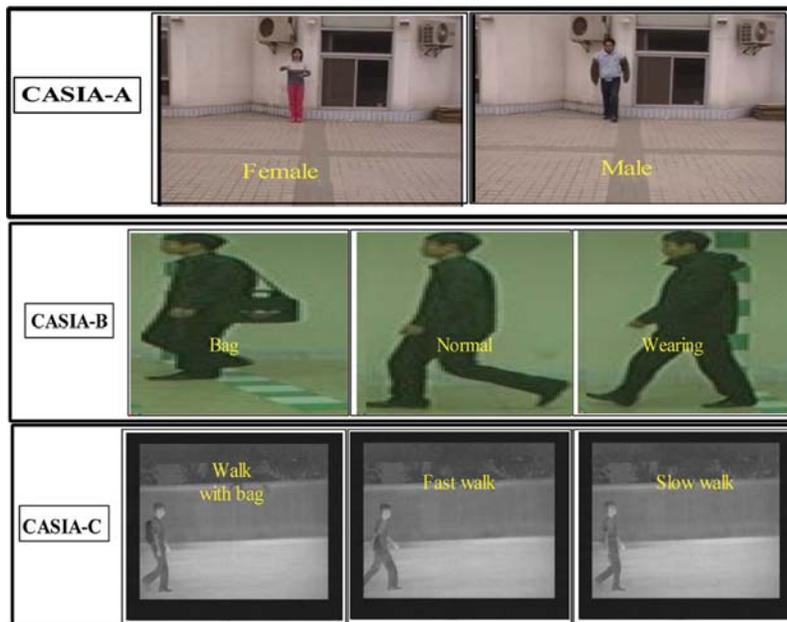
**Table 8:** Proposed method results for human gaits recognition on different datasets using the CASIA-B dataset

Dataset	Classes	SE	SP	PPV	FPR	FNR	NPV	CPR
CASIA-B	Bag	0.94	0.86	0.94	0.13	0.05	0.86	0.92
	Wearing	1.00	1.00	1.00	1.00	0.00	0.00	1.00
	Normal	0.91	0.81	0.92	0.18	0.08	0.78	0.88

The evaluation results in **Tab. 9** shows that, the proposed method achieved 1.00 CPR on the classes of CASIA-C dataset. The outcomes in **Tabs. 7–9**, depicts that the proposed model obtained a 1.00 correct recognition rate (CPR). The recognition outcomes on the CASIA-B dataset are 0.92 CPR on humans with the bag, 1.00 CPR on wearing class, and 0.88 CPR on a normal class. The predicted labels of human gait recognition are shown in **Fig. 8**. The proposed approach comparison is mentioned in **Tab. 10**.

**Table 9:** Proposed method results for human gaits recognition on different datasets using the CASIA-C dataset

Dataset	Classes	SE	SP	PPV	FPR	FNR	NPV	CPR
CASIA-C	Slow walk	1.00	1.00	1.00	0.00	0.00	1.00	1.00
	Fast walk	1.00	1.00	1.00	0.00	0.00	1.00	1.00
	Walk with bag	1.00	1.00	1.00	0.00	0.00	1.00	1.00



**Figure 8:** Predicted labels on benchmark datasets

**Table 10:** Proposed approach results compared with recent approaches

Refs.	YEAR	CASIA-A	CASIA-B
[45]	2020	0.95 CPR	0.92 CPR
[46]	2019	–	0.95 CPR
Proposed method	2020	1.00 CPR	0.96 CPR in which 0.88 (human with bag) and 0.92 (normal)

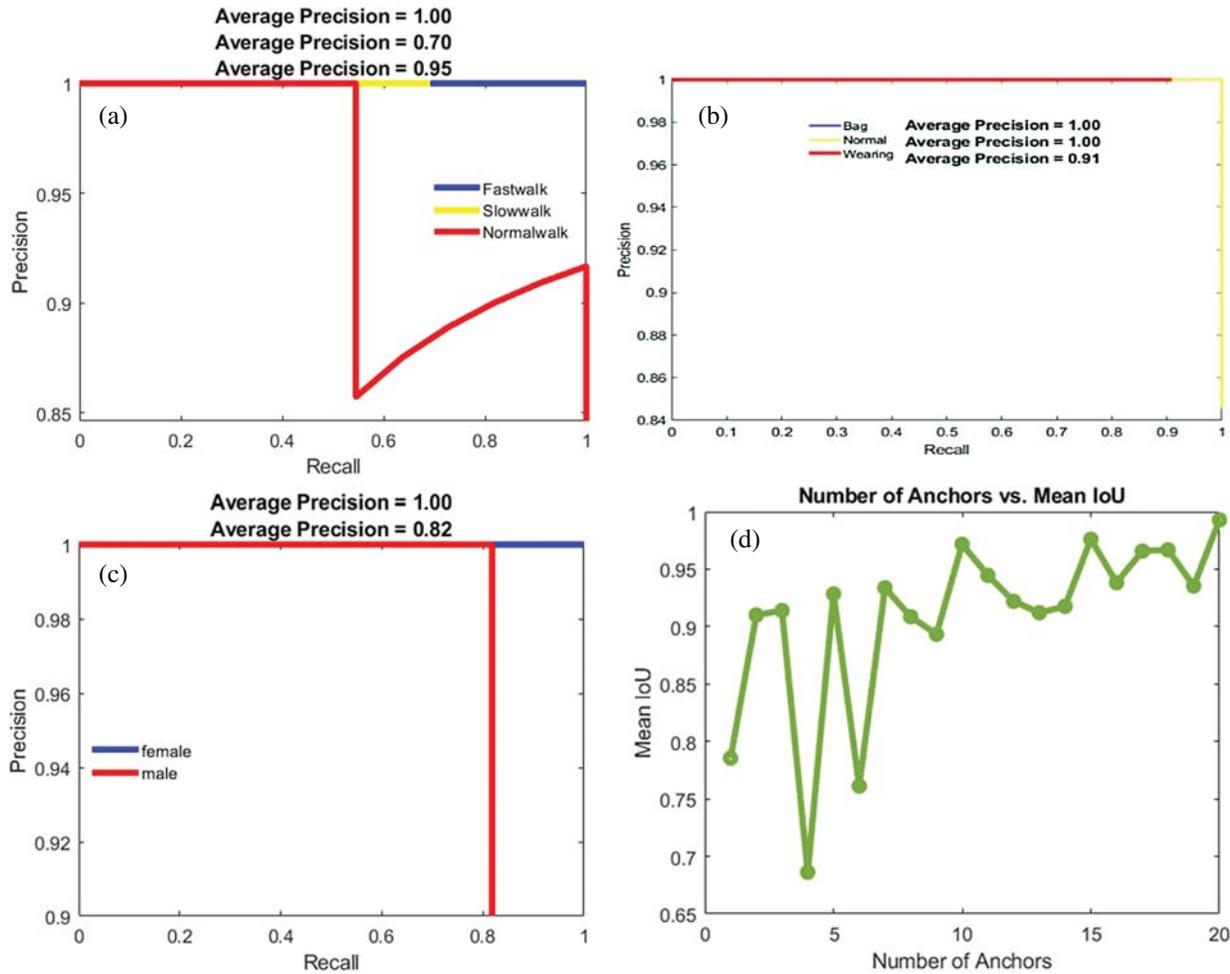
Six recent states of the art approaches are considered for performance evaluation based on some benchmark datasets. In the comparison scenario, the experimental setup is also discussed for existing work with proposed work. Wang et al. [45] used an ensemble learning method for human gait classification on CASIA-A & CASIA-B datasets and achieved results are 0.95 and 0.92 CPR respectively. Wang et al. [46] utilized the LSTM model to learn the sequential patterns of the input images and achieved 0.95 CPR on the CASIA-B dataset. The results in Tab. 11, are compared with the latest methodologies which show the proposed approach performance is superior. The proposed model results are better because of strongest feature vectors are obtained using the Conv-BiLSTM model for the classification of different types of human gaits with maximum CPR and also provided good results on a limited range of the input videos.

**Table 11:** Localization results

Datasets	Classes	mAP
CASIA-A	Female	1.00
	Male	0.82
CASIA-B	Bag	1.00
	Wearing	0.91
	Normal	1.00
CASIA-C	Fast walk	1.00
	Slow walk	0.70
	Normal walk	0.95

### 4.3 Experiment #02

The proposed YOLOv2-ONNX model is validated on CASIA-A, CASIA-B, and CASIA-C in terms of mean average precision (mAP) as mentioned in Tab. 11. The localization outcome according to the respective class labels is graphically depicted in Fig. 9. Tab. 11, shows the proposed approach obtained mAP of 1.00, 0.91, and 1.00 on different classes such as Bag, wearing, and normal of the CASIA-B dataset respectively.

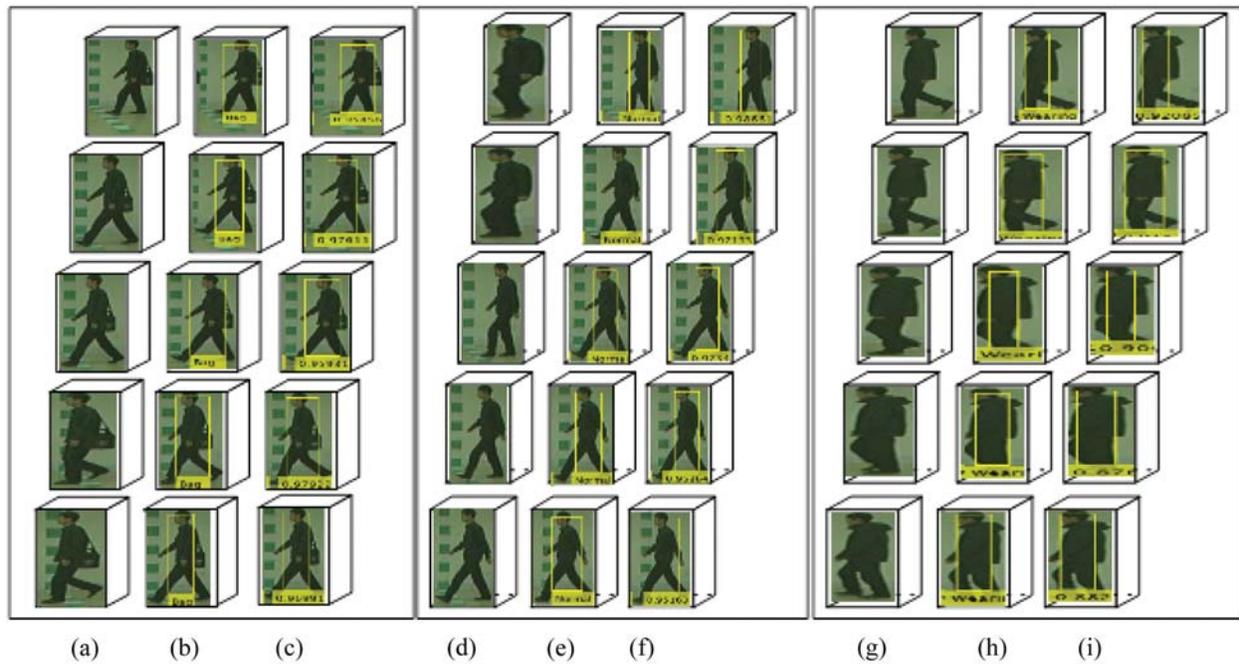


**Figure 9:** Localization results in term of mAP and IoU (a) CASIA-C (b) CASIA-B (c) CASIA-A (d) IoU

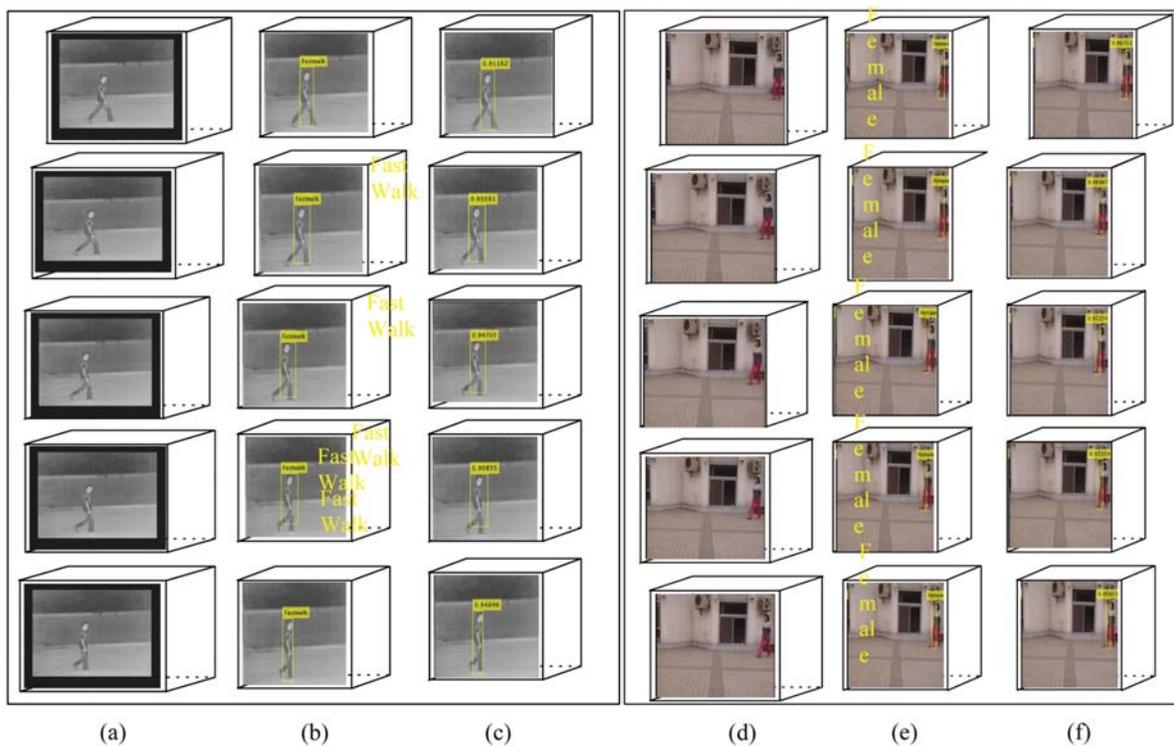
On different classes of the CASIA-C dataset i.e., fast walk, slow walk, and normal walk achieved mAP is 1.00, 0.70, and 0.95 respectively, where on the CASIA-A dataset attained mAP is 1.00 and 0.822 on female and male classes respectively. The proposed method more precisely localizes the different types of human gaits as illustrated in Figs. 10–12.

Fig. 10 shows, maximum achieved predicted scores of 0.979 on bag class, 0.986 on normal class, and 0.928 on wearing class.

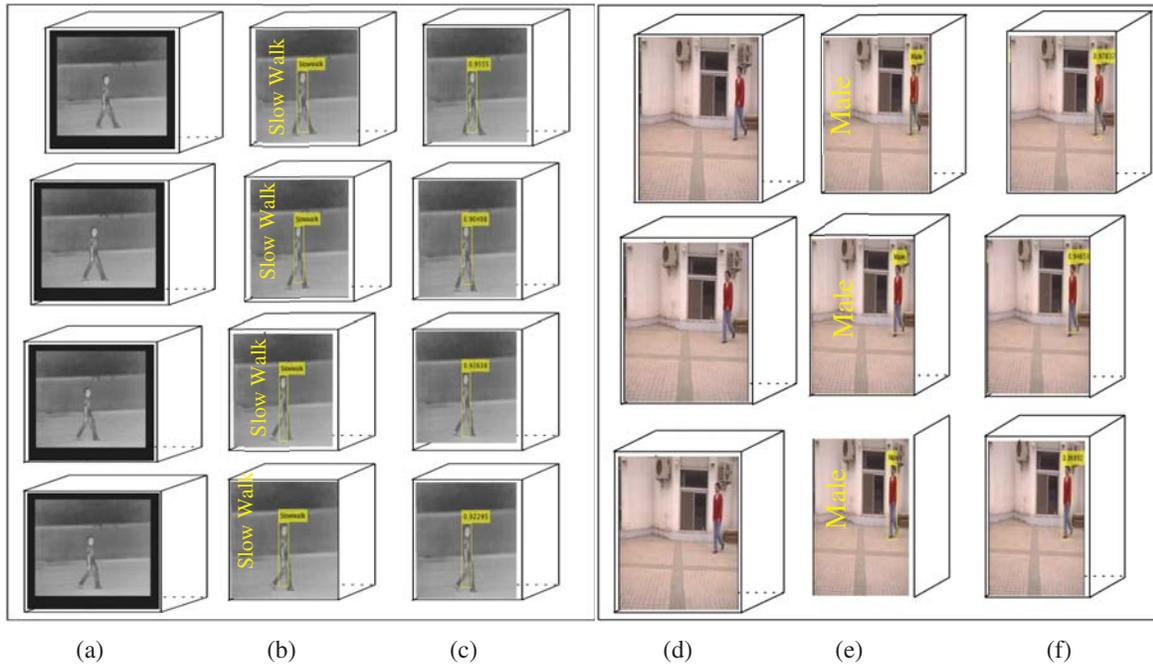
Figs. 10–12 reveals that the suggested approach, the obtained higher predicted scores are 0.948 on the fast walk, 0.972 on female class, 0.955 on slow walk class, and 0.978 on male class.



**Figure 10:** Gait localization (a, d and g) original gait images (b, e and h) gait labels (c, f and i) prediction scores



**Figure 11:** Gait localization (a, d) original gait images (b, e) gait labels (c, f) prediction scores



**Figure 12:** Gait localization (a, d) original gait images (b, e) gait labels (c, f) prediction scores

## 5 Conclusions

Due to differences in the multiple viewpoints of human gaits, the HGR is a difficult activity. Therefore, in this study tinyYOLOv2-SqueezeNet model is developed that more accurately localized the different types of human gaits. The proposed method achieved mAP of 1.00, 0.91, and 1.00 on Bag, wearing, and normal classes of CASIA-B dataset respectively. Whereas 1.00, 0.70, and 0.95 mAP on the fast walk, slow walk, and normal walk of CASIA-C dataset respectively. Similarly, 1.00 and 0.82 mAP on female and male classes of the CASIA-A dataset respectively. Furthermore, this research investigates a features extraction model based on Conv-BiLSTM that more accurately classifies human gaits. The experimentation is performed on CASIA-A, B, and C datasets. The model achieves 1.00 CPR to classify human with coat wearing, 0.92 CPR on a human with bag class and 0.87 CPR in a normal class. The overall CPR including three classes (wearing, bag, and normal) achieved 0.91. The 1.00 CPR achieved on CASIA-A as well as CASIA-C datasets on all classes such as female, male, human with a slow walk, human with a fast walk, human with the bag. The computed results proved that a combination of CNN and BiLSTM provides the highest recognition rate as compared with individual CNN or the LSTM models. The proposed method performance is dependent on a selected number of features; however, some useful features may be ignored. Moreover, video sequences in a low-quality resolution that affect recognition accuracy.

**Funding Statement:** This research was supported by the Korea Institute for Advancement of Technology (KIAT) Grant funded by the Korea Government (MOTIE) (P0012724, The Competency, Development Program for Industry Specialist) and the Soonchunhyang University Research Fund.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

- [1] M. Sannidhan and G. AnanthPrabhu, "A comprehensive review on various state-of-the-art techniques for composite sketch matching," *Imperial Journal of Interdisciplinary Research*, vol. 2, no. 2, pp. 1131–1138, 2016.
- [2] N. V. Boulgouris, D. Hatzinakos and K. N. Plataniotis, "Gait recognition: A challenging signal processing technology for biometric identification," *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 78–90, 2005.
- [3] K. Koide and J. Miura, "Identification of a specific person using color, height, and gait features for a person following robot," *Robotics and Autonomous Systems*, vol. 84, no. 2, pp. 76–87, 2016.
- [4] J. P. Singh, S. Jain, S. Arora and U. P. Singh, "Vision-based gait recognition: A survey," *IEEE Access*, vol. 6, pp. 70497–70527, 2018.
- [5] S. Bharadwaj, T. I. Dhamecha, M. Vatsa and R. Singh, "Computationally efficient face spoofing detection with motion magnification," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition Workshops*, Portland, Oregon, pp. 105–110, 2013.
- [6] S. Yu, T. Tan, K. Huang, K. Jia and X. Wu, "A study on gait-based gender classification," *IEEE Transactions on Image Processing*, vol. 18, no. 8, pp. 1905–1910, 2009.
- [7] A. Veeraraghavan, A. K. Roy-Chowdhury and R. Chellappa, "Matching shape sequences in video with applications in human movement analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 12, pp. 1896–1909, 2005.
- [8] V. Parameswaran and R. Chellappa, "View invariance for human action recognition," *International Journal of Computer Vision*, vol. 66, no. 1, pp. 83–101, 2006.
- [9] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio *et al.*, "Real-time human pose recognition in parts from single depth images," in *CVPR*, Providence, RI, USA, pp. 1297–1304, 2011.
- [10] A. F. Abate, M. Nappi, D. Riccio and G. Sabatino, "2D and 3D face recognition: A survey," *Pattern Recognition Letters*, vol. 28, no. 14, pp. 1885–1906, 2007.
- [11] D. S. Matovski, M. S. Nixon, S. Mahmoodi and J. N. Carter, "The effect of time on gait recognition performance," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 543–552, 2011.
- [12] I. C. Chang and S. Y. Lin, "3D human motion tracking based on a progressive particle filter," *Pattern Recognition*, vol. 43, no. 10, pp. 3621–3635, 2010.
- [13] D. M. Gavrila and L. S. Davis, "3-D model-based tracking of humans in action: a multi-view approach," in *Proc. CVPR IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, San Francisco, CA, USA, pp. 73–80, 1996.
- [14] V. Lepetit and P. Fua, "Monocular model-based 3d tracking of rigid objects: A survey," *Foundations and Trends in Computer Graphics and Vision*, vol. 1, no. 1, pp. 1–89, 2005.
- [15] C. Wan, L. Wang and V. V. Phoha, "A survey on gait recognition," *ACM Computing Surveys (CSUR)*, vol. 51, no. 5, pp. 1–35, 2018.
- [16] R. Anusha and C. D. Jaidhar, "Clothing invariant human gait recognition using modified local optimal oriented pattern binary descriptor," *Multimedia Tools and Applications*, vol. 79, no. 3, pp. 2873–2896, 2020.
- [17] R. Anusha and C. D. Jaidhar, "Human gait recognition based on histogram of oriented gradients and Haralick texture descriptor," *Multimedia Tools and Applications*, vol. 79, no. 11–12, pp. 8213–8234, 2020.
- [18] J. H. Shah, M. Sharif, M. Yasmin and S. L. Fernandes, "Facial expressions classification and false label reduction using LDA and threefold SVM," *Pattern Recognition Letters*, vol. 139, pp. 166–173, 2017.

- [19] H. Guo, B. Li, Y. Zhang, Y. Zhang and W. Li, "Gait recognition based on the feature extraction of Gabor filter and linear discriminant analysis and improved local coupled extreme learning machine," *Hindawi*, vol. 20, pp. 1–9, 2020.
- [20] A. O. Lishani, L. Boubchir, E. Khalifa and A. Bouridane, "Human gait recognition using GEI-based local multi-scale feature descriptors," *Multimedia Tools and Applications*, vol. 78, no. 5, pp. 5715–5730, 2019.
- [21] S. K. Gupta, G. M. Sultaniya and P. Chattopadhyay, "An efficient descriptor for gait recognition using spatio-temporal cues," *Emerging Technology in Modelling and Graphics*, vol. 5, pp. 85–97, 2020.
- [22] R. Vinothkanna and P. Sasikumar, "A novel multimodal biometrics system with fingerprint and gait recognition traits using contourlet derivative weighted rank fusion," in *Int. Conf. on Computational Vision and Bio Inspired Computing*, Cham, Springer, pp. 950–963, 2019.
- [23] D. Pavithra and S. Math, "A review on human gait detection," *Global Journal of Computer Science and Technology*, vol. 2, pp. 1–9, 2019.
- [24] Z. Zhang, L. Tran, X. Yin, Y. Atoum, X. Liu *et al.*, "Gait recognition via disentangled representation learning," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Long Beach, CA, pp. 4710–4719, 2019.
- [25] X. Wang and J. Zhang, "Gait feature extraction and gait classification using two-branch CNN," *Multimedia Tools and Applications*, vol. 79, no. 3, pp. 1–14, 2019.
- [26] R. Ranjan, R. Arya, S. L. Fernandes, E. Sravya and V. Jain, "A fuzzy neural network approach for automatic K-complex detection in sleep EEG signal," *Pattern Recognition Letters*, vol. 115, no. 6, pp. 74–83, 2018.
- [27] H. Arshad, M. A. Khan, M. I. Sharif, M. Yasmin, J. M. R. Tavares *et al.*, "A multilevel paradigm for deep convolutional neural network features selection with an application to human gait recognition," *Expert Systems*, vol. 2, pp. 1–25, 2020.
- [28] P. Arora, S. Srivastava and S. Singhal, "Analysis of gait flow image and gait Gaussian image using extension neural network for gait recognition," in *Deep Learning and Neural Networks: Concepts, Methodologies, Tools, and Applications*, Management Association, I. (Ed.) Hershey PA, USA: IGI Global, pp. 429–449, 2020. Chapter 25, [Online]. <http://www.igi.Global.com>.
- [29] M. M. Hasan, H. A. Mustafa and I. Security, "Multi-level feature fusion for robust pose-based gait recognition using RNN," *International Journal of Computer Science and Information Security*, vol. 18, no. 1, pp. 1–12, 2020.
- [30] R. Liao, S. Yu, W. An and Y. Huang, "A model-based gait recognition method with body pose and human prior knowledge," *Pattern Recognition*, vol. 98, no. 2, pp. 1–11, 2020.
- [31] O. Dehzangi, M. Taherisadr, R. ChangalVala and P. Asnani, "Motion-based gait identification using spectro-temporal transform and convolutional neural networks," in *Advances in Body Area Networks*, Switzerland: Springer, Cham, pp. 407–421, 2019. [Online]. [https://doi.org/10.1007/978-3-030-02819-0\\_31](https://doi.org/10.1007/978-3-030-02819-0_31).
- [32] Y. Zhang, Y. Huang, S. Yu and L. Wang, "Cross-view gait recognition by discriminative feature learning," *IEEE Transactions on Image Processing*, vol. 29, pp. 1001–1015, 2019.
- [33] K. Zhang, W. Luo, L. Ma, W. Liu and H. Li, "Learning joint gait representation via quintuplet loss minimization," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Long Beach, CA, pp. 4700–4709, 2019.
- [34] J. Feng, Y. Wang and S. F. Chang, "3D shape retrieval using a single depth image from low-cost sensors," in *2016 IEEE Winter Conf. on Applications of Computer Vision*, Lake Placid, NY, USA, pp. 1–9, 2016.
- [35] M. M. Hasan and H. A. Mustafa, "Multi-level feature fusion for robust pose-based gait recognition using RNN," *International Journal of Computer Science and Information Security*, vol. 18, no. 1, pp. 1–12, 2020.
- [36] J. Wang, M. She, S. Nahavandi and A. Kouzani, "A review of vision-based gait recognition methods for human identification," in *2010 Int. Conf. on Digital Image Computing: Techniques and Applications*, Sydney, NSW, Australia, pp. 320–327, 2020.

- [37] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [38] S. Zheng, J. Zhang, K. Huang, R. He and T. Tan, “Robust view transformation model for gait recognition,” in *2011 18th IEEE Int. Conf. on Image Processing*. Brussels, Belgium, pp. 2073–2076, 2011.
- [39] S. Yu, D. Tan and T. Tan, “A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition,” in *18th Int. Conf. on Pattern Recognition*, Hong Kong, China, pp. 441–444, 2006.
- [40] R. He, T. Tan and L. Wang, “Robust recovery of corrupted low-rankmatrix by implicit regularizers,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 4, pp. 770–783, 2013.
- [41] S. Zheng, K. Huang, T. Tan and D. Tao, “A cascade fusion scheme for gait and cumulative foot pressure image recognition,” *Pattern Recognition*, vol. 45, no. 10, pp. 3603–3610, 2012.
- [42] S. Zheng, K. Huang and T. Tan, “Evaluation framework on translation-invariant representation for cumulative foot pressure image,” in *2011 18th IEEE Int. Conf. on Image Processing*, Brussels, Belgium, pp. 201–204, 2011.
- [43] L. Wang, T. Tan, H. Ning and W. Hu, “Silhouette analysis-based gait recognition for human identification,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1505–1518, 2003.
- [44] D. Tan, K. Huang, S. Yu and T. Tan, “Efficient night gait recognition based on template matching,” in *18th Int. Conf. on Pattern Recognition*, Hong Kong, China, pp. 1000–1003, 2006.
- [45] X. Wang and W. Q. Yan, “Cross-view gait recognition through ensemble learning,” *Neural Computing and Applications*, vol. 32, no. 11, pp. 7275–7287, 2020.
- [46] X. Wang and W. Q. Yan, “Human gait recognition based on frame-by-frame gait energy images and convolutional long short-term memory,” *International Journal of Neural Systems*, vol. 30, no. 1, pp. 1950027–1950028, 2020.