Journal of Engineering Science and Technology Review 14 (5) (2021) 151 - 155

JOURNAL OF
Engineering Science and Technology Review

www.jestr.org

Research Article

Gait-Skeleton: Skeleton Based Convolutional Neural Network Gait Recognition

Sk Md Alfayeed* and Baljit Singh Saini

Department of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab, India

Received 23 March 2021; Accepted 28 October 2021

Abstract

Gait recognition is an image/video-based biometric approach used to determine the identity of individuals according to their style of walk, the direction of walk, or the manner of the walk. The efficiency of gait recognition is often impaired by covarying conditions like changes in clothing, angle of view, and carriage condition. Most of the methods for gait recognition depict a person using silhouette images in every frame, which is under the Appearance-based approach. However, pictures of silhouettes will lack fine grains and most papers fail to acknowledge how these figures are obtained in complicated scenes. Moreover, silhouette pictures not only include gait features, but also other visual clues. However, this method cannot be treated as the best recognition of the gait. Another method is called the Model-Based approach where we utilize the RGB images to generate skeletons over the images to recognize the gait patterns most efficiently. Here, we propose Gait-Skeleton with the combination of Convolutional Neural Network (CNN) in order to achieve a contemporary model-based technique for gait recognition. The primary benefits are a smoother, smarter extraction of the gait features and the ability to integrate strong spatial modelling with CNN. The famous CASIA-B data set experiments demonstrate that our technique archives the finest in model-based gait recognition performance.

Keywords: Gait, Convolutional Neural Network (CNN), Skeleton.

1. Introduction

Biometric is described as a study to identify a person's specific characteristics with their fundamental physical characteristics such as iris, face, hand geometry, fingerprint, recognition of voice, and retina or behavioral characteristics such as gait, typing patterns, or gestures. Human gait is known for understanding human movement and monitoring people's walking activity. Human Identity based on gait has been a hot subject in biometrics in recent years. Current biometric methods such as iris, hand mechanics, and finger identification, needed the subject's physical interactions and cooperation. They are not capable of distinguishing an individual from a remote distance. They can't operate in challenging conditions, including low light and poor visibility [1]. For example, criminals typically wear masks and gloves to get rid of typical recognition such as iris, fingerprint, facial. Gait recognition is the only valuable and efficient means of detection in such situations. In addition, understanding the gait is heavily dependent on both the patterns of human gestures and the human body shape [2]. The human gait incorporates special features that eliminate the drawbacks of these current approaches. Gait recognition is considered to be a promising biometric solution for the next decade because of the extensive utilization in forensic identification and criminal detection [1].

Two methods can be used for gait recognition: model-based and appearance-based or model-free. The model-based method relies on the structure of the human body to pull out the gait with dynamic gait parameters (stride and speed) [3]. The main drawback of the model-based approach is using high-resolution images, which is computationally costly. The model-free method specifically focused on the gait silhouettes [4] or gait sequences [5] and focus on human body movements. Gait descriptors or patterns can be extracted from the gait silhouette. Compared to the model-based methods, the model-free method is applied directly to a gait sequence. But the benefit of model-based identification is that it deals with the RGB photographs that are suited to public inspection or the outside.

Two methods can be used for gait recognition: model-based and appearance-based or model-free. The model-based method relies on the structure of the human body to pull out the gait with dynamic gait parameters (stride and speed) [3]. The main drawback of the model-based approach is using high-resolution images, which is computationally costly. The model-free method specifically focused on the gait silhouettes [4] or gait sequences [5] and focus on human body movements. Gait descriptors or patterns can be extracted from the gait silhouette. Compared to the model-based methods, the model-free method is applied directly to a gait sequence. But the benefit of model-based identification is that it deals with the RGB photographs that are suited to public inspection or the outside.

The implementation of an appropriate method that is invariable in many configurations, including adjustment of viewing angles, and the usage or lack of the subject under the circumstances, is a difficult challenge in gait recognition. Thus, under these covariates or variations, we regard the issue of gait recognition as the key field. These covariates also exist in actual circumstances and can greatly impact gait recognition efficiency.

*E-mail address: s.m.alfayeed@gmail.com

ISSN: 1791-2377 © 2021 School of Science, IHU. All rights reserved.
doi:10.25103/jestr.145.17

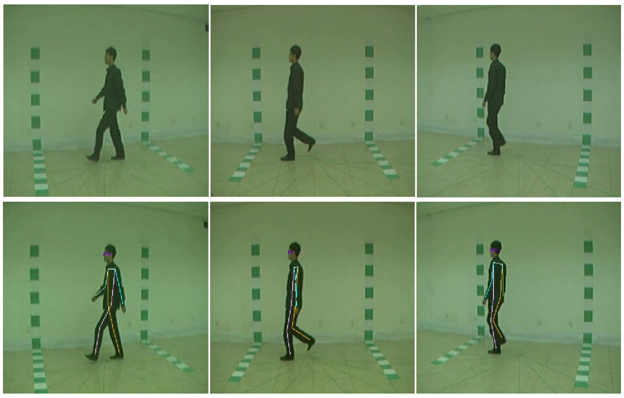

Fig. 1. Upper sample images of Model-free approach and lower sample images of Model-based approach.

The principle of minimum potential energy is based on displacement as the basic unknown quantity. The potential energy function is established to solve the settlement, while the ground reaction is calculated from the displacement. However, it is rarely reported to solve the foundation beam based on the principle of minimum complementary energy. The reason is that the premise of using the stress variation method is to set the known stress not only to satisfy the stress boundary condition but also to satisfy the equilibrium equation, which is difficult to achieve for general problems. But it is satisfactory for the half-space elastic foundation beam, and this method can directly solve the ground reaction, avoiding the error of the reaction obtained by the second calculation of the displacement.

Although an ideal foundation calculation model has been established, while with the development of large-scale construction projects, the interaction of ground and foundation beams, slabs, and other foundations require more in-depth theoretical studies. How to solve the internal force and displacement of the foundation beam more accurately and conveniently has become the key problem.

1.1 Problem statement

Most researchers have solved the issue of gait identification using a hard-craft feature extraction process and a deep-learning technique [6]. In the last couple of years, deep learning has attracted recognition because of the drawbacks of the hard-craft extraction process in various data sets [1]. Deep learning has the potential to perform well in larger and complex datasets; it has become popular for action and signals identification, image classification, and natural language processing in computer vision fields [8]. Deep learning gait recognition provides great results with promising future usage, especially the convolutional neural network. Therefore, this paper proposes Gait-Skeleton, a model-based multilayer deep CNN approach combined with gait skeleton pose estimation. The pose estimate substitutes the silhouette extraction from earlier methods. The skeleton-based representation provides actual identification while still using less sensitive personal information.

1.2 Objectives

The major objectives of this paper are as follows:

- To develop a robust model-based approach utilizing CNN's powerful spatial and temporal modeling.
- To recognize human gait without individual intervention, a robust CNN and Gait Skeleton-based approach, Gait-Skeleton is developed in this work.

1.3 Organization

The remainder of the paper is structured as follows as Section II gives some ideas and a summary of previous related researches. Section III provides an explanation of the proposed methodology. Section IV gives elaboration and analysis of the results with figures and tables. Finally, Section V delivers the conclusion along with the future scope

2. Discussion

In 2016, Shiraga et al. [15] proposed a convolutional neural network (CNN) model by feeding the most frequent image-based gait energy image (GEI) to recognize human gait called GEINet. They performed the analysis to illustrate the efficacy in both co-operative and unco-operative settings in terms of cross-view gait recognition using the OU-ISIR (wide population) dataset. Wolf et al. [13] presented a convolutional neural network by using 3D convolutions for Gait Recognition with several views capturing Spatio-temporal characteristics. The gray-scale and optical flow input formats are used that improve color invariance. The findings are comparable to better efficiency, particularly for wide sight differences, in contrast to previous approaches. Zhang et al. [19] developed a gait recognition system based on a siamese neural network for the efficient and discriminatory extraction of human gait features. Unlike traditional convolutional neural networks, the siamese neural network can use distance metrics learning to derive the similarity metric from the pairs of the gait of the same person or from a different person.

In 2017, Seyfioglu et al. [14] demonstrated the potential for radar technology to distinguish a huge number of classes of human unaided and aided motion. This research shows the ability to distinguish radar against various types of supported and unassisted travel. Deep microDoppler features are used to achieve an 89 percent accurate classification with a 3-layer auto-encoder layout and 17 percent improvement over the traditional SVM with 127 predetermined features.

In 2018, Gadaleta and Rossi [16] introduced a smartphone motion signals-based user authentication framework called IDNet. The aims of this application are the identification of the target user by use of the gyroscope and accelerometer signals from the smartphone inside the front pocket. IDNet is the first approach that uses a profound deep learning approach for gait recognition.

In 2019, Xu et al. [20] proposed a model based on the capsule network considering similar mid-level features and matching local features. They used the OU-ISIR B dataset (treadmill) and CASIA-B dataset for their experiments. They stated that their method exceeded the previous state-of-the-art outcomes. Min et al. [1] proposed various activation function-based CNN gait recognition models using CASIA-B all view dataset. Their method achieved 98.8% accuracy with 100 epochs iteration time for each model. Bonetto et al. [22] demonstrated a gait analysis system for anomaly detection using the RNN Seq2Seq model and CNN classifier model. They used a smartphone mounted on the chest for capturing live human gate videos and images. Their architecture was able to detect anomalies in 100% of the cases.

3. Methodology

In this section, our approach to learning discriminatory information from a sequence of human poses is described. Fig 2 shows the overall pipeline.

3.1 Notations A human skeleton is referred to as G = (W, F) in which W = {w1,. . wN } is the set of O nodes that are joint, and F is an adjacence matrice set of edges which represent bones, B ∈ S O×O, alternatively, with Bj,i = 1 in connection with the edges of wj and wi. Since H is unspanned, B is symmetrical.

Thus A structurally and *Y*-functionally describes the input gauge, $Yu \in S\ O \times D$ being at u time the node features. The D-dimension feature *Y* comprises the 2D coordinate and trust. A weight matrix that can be learned at layer *m* of a system is referred to as $\Theta(m) \in S\ Dm \times Sm+1$.

where E_f is the elastic modulus of ground, μ_f is the Poisson’s ratio of ground. In Eq. (2), *p*(*x*) is the unknown function that satisfies the equilibrium condition:

3.2 Convolutional neural network (CNN) was inspired by the human brain cell neurons. CNN has been adapted impeccably in many fields of science, including image matching, facial recognition, human identification. A CNN can consist of convolutional layers, pooling layers, normalization layers such as fully connected layers, hidden layers, as well as input and output layers. The extraction of a certain kind of function is the task of each of these convolutional layers. The network creates a higher degree of capacity as the depth of the network grows as each layer relies on the previous layers.

The layers of the CNN update rule can be applied on inputs from skeletons, as determined in features *Y* and graph *B*, to features at time *u* as: $Y\ (m+1)\ u = \sigma\ \ (E\tilde{}\ -\ 1/2\ B\tilde{}\ E\tilde{}\ -\ 1/2\ Y\ (m)\ u\ \Theta\ (m)$ where $B\tilde{} = B + J$ is the Skeleton Graph with additional self-loops to retain identity traits; $\sigma(\cdot)$ is the activation function and $E\tilde{}$ is the diagonal degree. The GCNs are diagonal degree matrixes with *A*: The word $D - E\tilde{} - 1/2\ B\tilde{}\ E\tilde{} - 1/2\ Y\ (m)\ u$ can intuitively be interpreted from messages delivered by direct neighbors as approximate spatial average aggregation.

3.3 Activation function We used the activation function for accelerated training after each convolutional layer and fully connected layer. The sigmoid activation function historically has provided excellent results for the neural network, but it has drawbacks called slow to converge and vanishing gradient problems. ReLU has now become popular in deep learning because it is capable of delivering successful training networks and solving the issue of varnishing gradients.

The Sigmoid function is primarily used because the probability is only between the 0 and 1 range, and hence it is the correct choice for the models where we have to estimate the probability as an output [9].

$$Sig(i) = 1/1+e\ -i \qquad (1)$$

The Rectified Linear Unit (ReLU) is one of the most commonly used activation functions in deep learning models. The function receives any negative input, it returns 0, but for any positive value i, it returns that value back [10]. So, it can be written as:

$$ReLU(i) = max(0,i) \qquad (2)$$

Exponential Linear Unit (ELU) is an activation function that solves some of the issues with ReLUs and leaves some of the positive elements in place. For this activation function, a recommended α value can be selected between 0.1 and 0.3 [18].

$$\begin{aligned} ELU(i) = \ & i \qquad if\ i > 0 \\ & \alpha(e\ i -1)\ if\ i < 0 \end{aligned} \qquad (3)$$

Leaky Rectified Linear Unit (LeakyReLU) is an activation function that also has an alpha α value like ReLU, the alpha value is preferred between 0.1 to 0.3. In this function, there is no “dead ReLU” (or “dying ReLU”) problem. When the ReLU has values under 0, this completely blocks learning in the ReLU because of gradients of 0 in the negative part [18].

$$\begin{aligned} & LReLU(i) = i \quad if\ i > 0 \\ & \alpha i \ \ if\ i \leq 0 \end{aligned} \qquad (4)$$

Parametric Rectified Linear Unit (PReLU) was introduced to overcome the shortcomings of ReLU (dying ReLU problem) and LeakyReLU (inconsistent predictions for negative input values). LeakyReLU allows a small, nonzero gradient when the unit is inactive. PReLU takes this concept further by transforming the leakage coefficient into a parameter learned along with the other neural network parameters [17].

$$\begin{aligned} & PReLU(i) = i, \quad if\ i > 0 \\ & \alpha i,\ otherwise \end{aligned} \qquad (5)$$

3.4 Pose Extraction We determine the human position in each frame for extracting features from raw input images. The purpose of the pose estimation is to detect *O* key points from an image $J \in SX \times I \times 3$ (i.e. hip, knee, shoulder, etc.) This is solved with the state-of-the-art method [7], by evaluating *O* heat maps $\{I1, I2, . . . , IO\ \}$ of size $X' \times I'$ *a* heatmap *Io*, where the *o-th* key point is located. The optimum location of the *Io* heat maps results in a position of the *vn* keypoint defining the edges *W*.

HRNet [11] as a 2D human pose estimator is used in our technique. We use the supplied COCO dataset network [12] pre-trained. There are 17 key points in the COCO Dataset. No set of bones or edges E is presented, but we have a widely used setup. Some samples after pose estimation are shown in Fig 2.

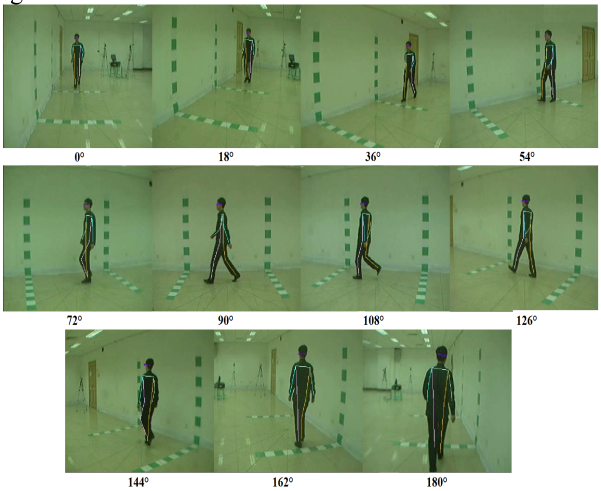

Fig. 2. Pose estimated skeleton-based sequences (0° to 180°).

4. Implementation details

The primary architectural structure of the Network follows the design suggested under the CNN with our case adaptations. The network consists of blocks of CNN. The unit consists of a Convolution, accompanied by a traditional 2D Convolution in the temporal domain. The network consists of several CNN layers in a sequence, followed by two convolutional layers of size 32x3x3 and 64x3x3, two activation functions layers, two max pool layers of size (2x2), and two dense layers of size 256 with (sigmoid function) and 124 (with softmax function). Each model is trained with 25 epochs and a batch size of 128. The accuracy graph is shown in Fig 4.

We use several single augmentation strategies to increase the skeletal graph. Firstly, we change the sequence order, which can be understood as the backward person. Secondly, by the graph center of gravity, we mirror the skeletal graph along a vertical axis. This increase causes the person to walk the other way. In order to make our network more resilient to estimate inaccuracies, we add small Gaussian noise to each joint and the same joint.

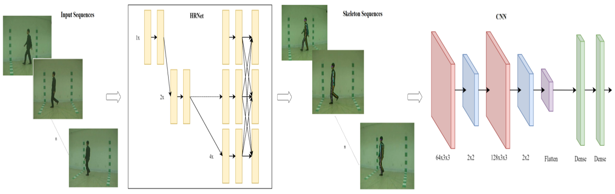

Fig. 3. Implementation flow overview.

Table 1. CNN Network architecture.

Layers	Value
Conv2D	32x3x3
MaxPool	2x2
Dropout	0.5
Conv2D	64x3x3
MaxPool	2x2
Dropout	0.5
Flatten	-
Dense	256
Dense	124

5. Experiments and results

The RGB images/videos do not come with the majority of available gait datasets because they are adaptable to gait processes based on GEIs/silhouettes. We therefore cannot assess the largest public gait dataset OU-MVLP [14], which is a contrast with other techniques in the assessment of the frequently used dataset CASIA-B.

4.1 Dataset The CASIA-B dataset [23] is frequently used in most recent researches. Inside the dataset, there are a total of 15004 videos of 124 subjects of 11 different angle views (0° to 180°) having six normal walking sequences (nm 1, nm 2, nm 3, nm 4, nm 5, nm 6) and two walking sequences with carrying bag and wearing a coat (bg 1, bg 2, cl 1, cl 2).

Therefore, the training set includes the first 74 subjects out of 124 subjects, while the other 50 subjects are the test set. The four first sequences of the normal condition (nm 1 to nm 4) are kept on the folder of all three test sets, while the remaining six sequences of the test are grouped together (nm 5 to nm 6, bg 1 to bg 2, and cl 1 to cl 2).

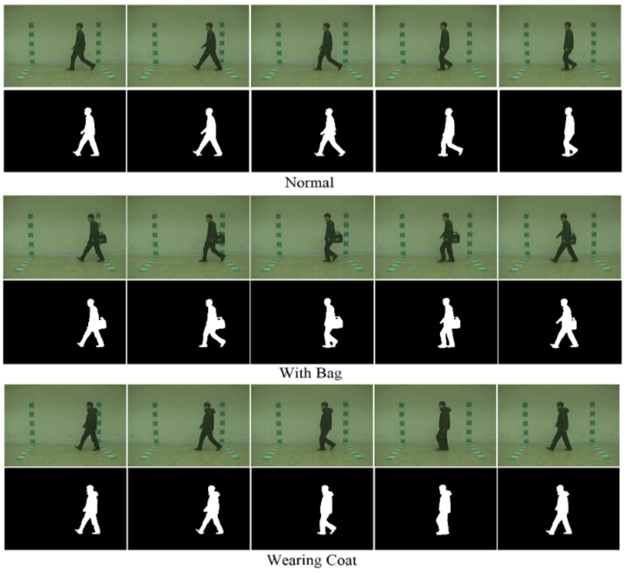

Fig. 4. CASIA-B Dataset samples in 90° view (nm-2, bg-2, cl-2).

All the images are sent HRNet skeleton generator for skeleton attached sequences. Then those sequences are fed into CNN to get the results.

4.2 Equipment For the experiment the Python codes with TensorFlow and Keras library are used using Jupyter Notebook. The physical machine with 8GB RAM, 250GB SSD, NVIDIA GT 740 Graphic Card is used.

5.1 Comparison with other methods Table 2 displays a comparison of Gait-Skeleton with PoseGait [24], representing the only post-recognition technique using hand-craft posing features. In all cross-view and walking conditions, our approach reveals radical improvements. This demonstrates the supremacy of our architecture as a feature extractor with all methods using a comparable performance extractor.

The most successful models currently have features based on appearance. In **Table 3**, we compare the appearance, model, and strategic approach. The first three techniques [15,16,20] all use images of the silhouettes as their feature extraction. Notably, we can still archive competition results against such appearance-based procedures with our skeleton-based feature extraction.

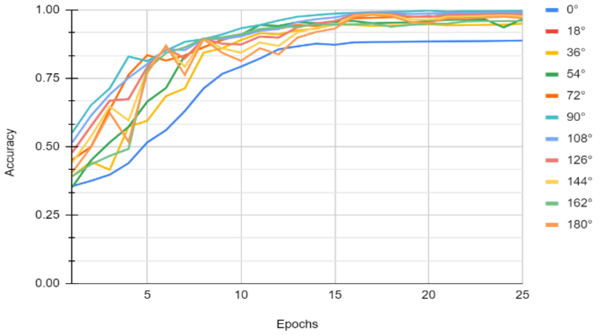

Fig. 5. Comparison between PoseGait and our approach Gait-Skeleton.

Table 2. Comparison between PoseGait and our approach Gait-Skeleton.

Views	Accuracy (%)	
	PoseGait [24]	**Gait-Skeleton**
0°	49.71	74.77

0°	61.63	87.95
18°	67	81.99
36°	66.74	84.37
54°	60.85	87.81
72°	59.3	91.31
90°	62.53	89.63
108°	61.48	87.95
126°	67.36	86.27
144°	62.03	84.8
162°	47.63	84.59
180°	49.71	74.77
Mean	60.56	**85.58**

Table 3. Result comparison table.

Type	Methods	Accuracy (%)
Appearance-base	**GEINet [15]**	**97.98**
	IDENet [16]	94
	Capsule Network [20]	74.44
Model-based	PoseGait [24]	43.3
	Gait-Skeleton	**85.58**

6. Conclusions

In this article, we introduce a new approach Gait-Skeleton for gait recognition using a skeleton-based sequence. To extract the 2D skeleton pose by using a human skeleton pose estimator, and to extract the gait information taking into consideration the intrinsic graphical structure of the skeleton. Moreover, State-of-the-art findings show modular gait recognition and competitive results against appearance-based techniques in gait recognition in studies performed on the well-known CASIA-B database [23]. Though the accuracy of our method is not greater than appearance-based approaches, it is still better than other model-based approaches.

References

1. Min PP, Sayeed S, Ong TS. Gait recognition using deep convolutional features. In2019 7th International Conference on Information and Communication Technology (ICoICT) 2019 Jul 24 (pp. 1-5). IEEE.
2. Alotaibi M, Mahmood A. Improved gait recognition based on specialized deep convolutional neural network. Computer Vision and Image Understanding. 2017 Nov 1;164:103-10.
3. Bashir K, Xiang T, Gong S. Gait recognition using gait entropy image.
4. Han J, Bhanu B. Individual recognition using gait energy image. IEEE transactions on pattern analysis and machine intelligence. 2005 Dec 19;28(2):316-22.
5. Alharthi AS, Yunas SU, Ozanyan KB. Deep learning for monitoring of human gait: A review. IEEE Sensors Journal. 2019 Jul 15;19(21):9575-91.
6. P Nikam and V Raut. Improved MANET security using Elliptic curve cryptography and EAACK. *In*: Proceedings of 2015 International Conference on Computational Intelligence and Communication Networks, Jabalur, India. 2015, p. 1125-9.
7. Cheng B, Xiao B, Wang J, Shi H, Huang TS, Zhang L. Higherhrnet: Scale-aware representation learning for bottom-up human pose estimation. InProceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2020 (pp. 5386-5395).
8. Das D, Chakrabarty A. Human gait recognition using deep neural networks. InProceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies 2016 Mar 4 (pp. 1-6).
9. Sharma, S., Activation functions in neural networks. Towards Data Science, 2017, 6.
10. DanD, Rectified Linear Units (ReLU) in Deep Learning. 2018. Retrieved from https://www.kaggle.com/dansbecker/rectified-linear-units-relu-in-deep-learning
11. Sun K, Xiao B, Liu D, Wang J. Deep high-resolution representation learning for human pose estimation. InProceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2019 (pp. 5693-5703).
12. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P. and Zitnick, C.L., 2014, September. Microsoft coco: Common objects in context. In European conference on computer vision (pp. 740-755). Springer, Cham.
13. Wolf T, Babaee M, Rigoll G. Multi-view gait recognition using 3D convolutional neural networks. In2016 IEEE International Conference on Image Processing (ICIP) 2016 Sep 25 (pp. 4165-4169). IEEE.
14. Takemura N, Makihara Y, Muramatsu D, Echigo T, Yagi Y. Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition. IPSJ Transactions on Computer Vision and Applications. 2018 Dec;10(1):1-4.
15. Shiraga K, Makihara Y, Muramatsu D, Echigo T, Yagi Y. Geinet: View-invariant gait recognition using a convolutional neural network. In2016 international conference on biometrics (ICB) 2016 Jun 13 (pp. 1-8). IEEE.
16. Gadaleta M, Rossi M. Idnet: Smartphone-based gait recognition with convolutional neural networks. Pattern Recognition. 2018 Feb 1;74:25-37.
17. Wikipedia. Rectifier (neural networks). Retrived from https://en.wikipedia.org/wiki/Rectifier_(neural_networks)#Parametric_ ReLU
18. Casper Hansen, Activation Functions Explained - GELU, SELU, ELU, ReLU and more. Deep Learning. 2019.
19. Zhang C, Liu W, Ma H, Fu H. Siamese neural network based gait recognition for human identification. In2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2016 Mar 20 (pp. 2832-2836). IEEE.
20. Xu Z, Lu W, Zhang Q, Yeung Y, Chen X. Gait recognition based on capsule network. Journal of Visual Communication and Image Representation. 2019 Feb 1;59:159-67.
21. RH Jhaveri. MR-AODV: A solution to mitigate blackhole and grayhole attacks in AODV based MANETs. *In*: Proceedings of the 3rd International Conference on Advanced Computing and Communication Technologies, Rohtak, India. 2013, p. 254-60.
22. Bonetto R, Soldan M, Lanaro A, Milani S, Rossi M. Seq2Seq RNN based Gait Anomaly Detection from Smartphone Acquired Multimodal Motion Data. arXiv preprint arXiv:1911.08608. 2019 Nov 19.
23. Yu S, Tan D, Tan T. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In18th International Conference on Pattern Recognition (ICPR'06) 2006 Aug 20 (Vol. 4, pp. 441-444). IEEE.
24. Yu S, Tan D, Tan T. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In18th International Conference on Pattern Recognition (ICPR'06) 2006 Aug 20 (Vol. 4, pp. 441-444). IEEE.