IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 1, 2020, accepted October 5, 2020, date of publication October 12, 2020, date of current version October 30, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3030086

Human Pose Estimation-Based Real-Time Gait Analysis Using Convolutional Neural Network

ALI ROHAN[1], (Associate Member, IEEE), MOHAMMED RABAH[2], TAREK HOSNY[3], AND SUNG-HO KIM[4]

[1]Department of Mechanical, Robotics and Energy Engineering, Dongguk University-Seoul, Seoul 04620, South Korea
[2]Computational Engineering and Analysis Research Group, Turku University of Applied Sciences, 20520 Turku, Finland
[3]Communication Engineering Department, Al-Safwa High Institute of Engineering, High Ministry of Education, Cairo 11837, Egypt
[4]Department of Control and Robotics Engineering, Kunsan National University, Gunsan 54150, South Korea

Corresponding author: Ali Rohan (ali_rohan2003@hotmail.com)

ABSTRACT Gait analysis is widely used in clinical practice to help in understanding the gait abnormalities and its association with a certain underlying medical condition for better diagnosis and prognosis. Several technologies embedded in the specialized devices such as computer-interfaced video cameras to measure patient motion, electrodes placed on the surface of the skin to appreciate muscle activity, force platforms embedded in a walkway to monitor the forces and torques produced between the ambulatory patient and the ground, Inertial Measurement Unit (IMU) sensors, and wearable devices are being used for this purpose. All of these technologies require an expert to translate the data recorded by the said embedded specialized devices, which is typically done by a medical expert but with the recent improvements in the field of Artificial Intelligence (AI), especially in deep learning, it is possible now to create a mechanism where the translation of the data can be performed by a deep learning tool such as Convolutional Neural Network (CNN). Therefore, this work presents an approach where human pose estimation is combined with a CNN for classification between normal and abnormal gait of a human with an ability to provide information about the detected abnormalities form an extracted skeletal image in real-time.

INDEX TERMS Convolutional neural network, deep learning, gait analysis, pose estimation.

I. INTRODUCTION

Almost every person on this planet excluding the disabled persons walks approximately 1 to 3 Km on a daily basis and it is considered a very important factor in defining our physical and mental health. Most people do not consider this importance of walking but it would not be inevitable to say that an inability to walk or to be mobile can dramatically change a person's life. It can cause significant health problems for the short and long term. Some people can move with abnormal gait for many years without any prominent symptoms, whereas for others, abnormal gait can cause injuries, pain that can lead towards a bigger health problem such as Musculoskeletal problems [1], Cardiovascular health issues [2], and Mental health issues [3]. In medical diagnostics, gait analysis is widely used to study underlying pathological gait conditions [4], for diagnosis in chiropractic and osteopathic professions; as hindrances in gait may be indicative of a misaligned pelvis or sacrum [5], [6], comparative biomechanics; to understand about the mechanics of locomotion [7] and biometrics; to recognize people by the way they walk [8]. Aside from clinical applications, gait analysis is also used in professional sports training to optimize and improve athletic performance [9].

There has been extensive research in the recent past to develop a methodology using different techniques for gait analysis. Previously, authors [10] introduced an average silhouette representation for a sequence. The Gait Energy Image (GEI) was proposed in [11]. The average silhouette representation is widely used than other techniques due to its simplicity and it is found to be effective in saving the computation time. In [12], authors presented another gait representation scheme called as Gait Entropy Image, and in [13] Gait using Pal and Pal Entropy (GPPE). A recent work conducted by authors [14], evaluated that the performance of different gait representation templates falls behind the performance level of GEI using large gait datasets. Principal component analysis (PCA) was used in gait recognition [15] to extract the

The associate editor coordinating the review of this manuscript and approving it for publication was Jianqing Zhu.

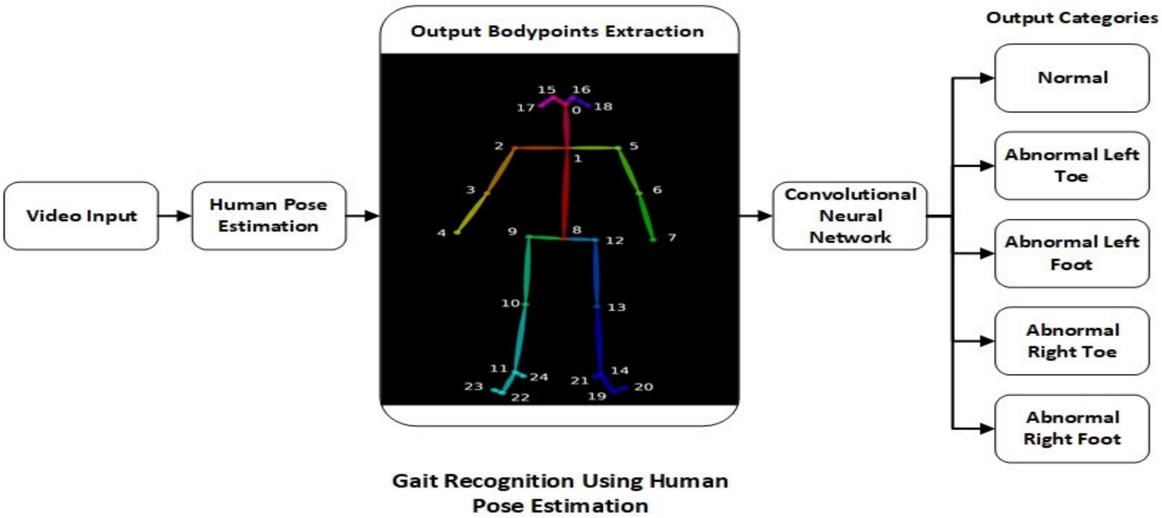


FIGURE 1. **Basic architecture of the proposed approach for gait analysis.**

unique features from gait representation. Linear discriminant analysis (LDA) was used in [16] to reduce the dimensionality of the gait representation data into lower dimensional space-separated classes but the problem with these techniques is that PCA and LDA do not take advantage of 2-dimensional data. In PCA and LDA, an image or 2-D data has to be converted to a 1-dimension vector and this sometimes results in poor recognition performance. Some subspace learning approaches have also been proposed [17]–[20], which started to consider learning the features from an object by considering the representation of higher-order tensors. Subspace learning approaches are widely used approaches for gait recognition. In [21], the authors proposed matrix-based sparse bilinear discriminant analysis (SBDA) as a sparse learning method effective for gait recognition. Also, Locality Preserving Projections (LPP) [22] and Local Fisher Discriminant Analysis (LEFDA) [23] were employed to generate the gait features.

Recently, Convolutional Neural Networks (CNNs) have achieved great results in different fields of pattern recognition, detection, and classification, especially in computer vision. CNN is a class of deep neural networks and is mostly applied to analyze visual imagery. In [24], an efficient deep neural network architecture for visual recognition is proposed and it is named as GoogLeNet. It is a big network that comprises of around 27 layers, and it uses max and average pooling, dropout method, and a softmax classifier. In [25], the authors proposed a method to recognize periodic human actions using CNN. In [26], a large CNN trained using 1.2 million high-resolution images from the ImageNet dataset was introduced. A CNN architecture to speed up the training time for largescale video classification was proposed in [27]. One million YouTube videos belonging to 487 classes of sports were trained using this architecture.

All these above-mentioned studies have shown some advancement in the process of gait analysis but there is the unavailability of a proper gait analysis mechanism for the subjects where problematic constraints such as the view angle, walking speed, clothing, surface, carrying status, shoe, and elapsed time are considered and minimalized in an efficacious way. Therefore, for further progress, this research work proposes an approach where human pose estimation is combined with a deep neural network such as CNN to classify the normal and abnormal gait of a person. The reason to use human pose estimation for gait analysis is that in pose estimation, deep learning-based CNN is used to detect the body points of a person without the worry of any problematic constraint. This gives us skeletal images where body points are joined to form a skeleton of a person. Only these obtained skeletal images are further used for classification and gait analysis. Hence, giving the freedom from using any wearable sensor or device on the subject's body with minimal effect of any problematic constraint.

This paper is divided into the following sections: Section 1 provides an introduction, Section 2 describes the basic architecture of the system; Section 3 contains the results and discussion, and Section 4 concludes the study.

II. BASIC ARCHITECTURE OF THE SYSTEM

The basic architecture of the proposed approach is shown in the Fig.1. The first step involves the recording of a live video of a person's gait movement. Each frame in the video is processed through the human pose estimation algorithm to obtain a skeleton image comprising of 25 body points. The obtained image is given as an input to a CNN trained to classify multiple classes. In this work, we trained the CNN for five types of classes. These classes are *Normal, Abnormal Left Toe, Abnormal Left Foot, Abnormal Right Toe, and Abnormal Right Foot*. Once the trained CNN receives the input skeletal image of a person, the output is given as the image with a label of predicted class.

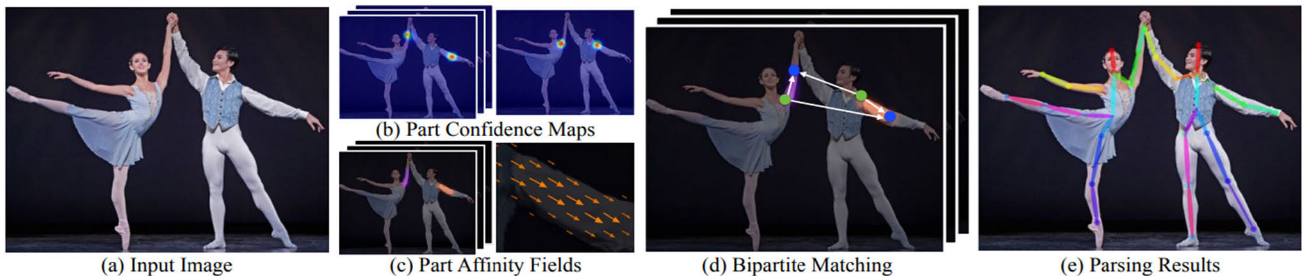


FIGURE 2. **Overall pipeline. (a) The method takes the entire image as the input for a CNN to jointly predict (b) confidence maps for body part detection and (c) PAFs for part association. (d) The parsing step performs a set of bipartite matchings to associate body part candidates. (e) Finally assemble them into full body poses for all people in the image.**

A. HUMAN POSE ESTIMATION

Human pose estimation is the process of predicting the structure of the body pose from a single, typically monocular image and is considered as one of the key problems in the field of computer vision. The importance of the pose estimation is its applicability in serval applications such as higher-level reasoning in the context of human-computer interaction and activity recognition; it is also one of the basic building blocks for marker-less motion capture (MoCap) technology. MoCap technology is useful for applications ranging from character animation to clinical analysis of the gait.

The human pose estimation faces several challenges among which the most significant are: (1) variability of human visual appearance in images, (2) variability in lighting conditions, (3) variability in human physique, (4) partial occlusions due to self-articulation and layering of objects in the scene, (5) complexity of human skeletal structure, (6) high dimensionality of the pose, and (7) the loss of 3d information that results from observing the pose from 2d planar image projections. In the past, there was no approach that could produce satisfactory results in general while dealing with all of the aforementioned challenges but recently researchers have shown very promising results using deep learning tools.

In [28], [29], the authors proposed a method for multi-person 2D pose estimation using Part Affinity Fields (PAF). PAF is a set of 2D vector fields that encode the location and orientation of limbs over the image domain. Given a set of detected body parts (shown as the red and blue points in Fig. 3a), how do we assemble them to form the full-body poses of an unknown number of people? We need a confidence measure of the association for each pair of body part detections, i.e., that they belong to the same person. One possible way to measure the association is to detect an additional midpoint between each pair of parts on a limb and check for its incidence between candidate part detections, as shown in Fig. 3b. However, when people crowd together—as they are prone to do—these midpoints are likely to support false associations (shown as green lines in Fig. 3b). Such false associations arise due to two limitations in the representation: (1) it encodes only the position, and not the orientation, of each limb; (2) it reduces the region of support of a limb to a single point. To address these limitations, the authors presented a novel feature representation called part affinity fields that preserves both location and orientation information across the region of support of the limb (as shown in Fig. 3c). The part affinity is a 2D vector field for each limb, for each pixel in the area belonging to a particular limb, a 2D vector encodes the direction that points from one part of the limb to the other. Each type of limb has a corresponding affinity field joining its two associated body parts. The overall pipeline of the algorithm is shown in Fig.2 and Fig. 4 presents the flow chart. The demonstrated results have significantly improved the issues related to the human pose estimation and it has made possible to utilize human pose estimation in application like gait analysis and this work is a step towards that. Fig. 2 shows the pipeline of the human pose estimation proposed by the authors and used in this work.

FIGURE 3. **Part association strategies. (a) The body part detection candidates (red and blue dots) for two body part types and allconnection candidates (grey lines). (b) The connection results using the midpoint (yellow dots) representation: correct connection (black lines) and incorrect connections (green lines) that also satisfy the incidence constraint. (c) The results using PAFs (yellow arrows). By encoding position and orientation over the support of the limb, PAFs eliminate false associations.**

B. CNN-BASED CLASSIFIER

A CNN-based classifier is used in this work to classify the normal and abnormal gait of a person. The architecture of the CNN used in this work is based on Visual Geometry Group-19 (VGG-19) architecture. Providing more depth in the network, and size of the convolution filters, VGG networks have a clear advantage over other network architecture such as AlexNet. The detail of the architecture is given in Table 1. The CNN is composed of 16 convolutional layers

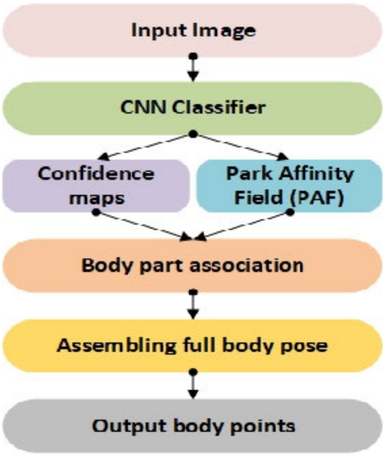


FIGURE 4. Flow chart for Human Pose Estimation.

and 3 fully connected layers. The size of the convolutional filter or respective fields is 3 × 3. Two 3 × 3 convolutional layers cover an effective receptive field of 5 × 5 and three 3 × 3 such layers have a 7 × 7 effective receptive field to incorporate forward flow with three non-linear rectifications instead of a single. The additional non-linearity introduction makes the decision function more discriminative and reduces the number of parameters significantly.

1) CNN TRAINING

The training of the CNN was performed offline by collecting and labeling a dataset comprising of multiple videos of a person walking under different scenarios. The data collection is one of the vital parts in achieving a good classification accuracy. We collected the data by segregating the classes into five categories; Normal, Abnormal Left Toe, Abnormal Left Foot, Abnormal Right Toe, and Abnormal Right Foot. For each category, the walking style of the person was made to change to mimic some injury or problem in feet. For each category we collected 50 videos, each video was recorded using a camera at 60 Frame Per Second (FPS) with a resolution of 1280 × 720 for a duration of 50 seconds. These videos were processed through the human pose estimation algorithm to obtain skeletal images which were further used for training and testing the CNN. 80% of these 50 videos for each category was used for training the CNN while 20% were used for testing the CNN. The network was trained for 45,000 iterations on a PC with specifications mentioned in Table 2. The accuracy achieved for classification is 97.3%. To create a problematic gait, we mimicked the walking style by using some object tied with the toe of the person under experiment, this gave us the mimicked data for the categories of Abnormal Left Toe and Abnormal Right Toe. On the other hand, to mimic some problems in feet we made the person walk in slippers with a foot cast and it gave us the data for the categories of Abnormal Left foot and Abnormal Right Foot. For category Normal, the person walked without any foot cast or object attached to the toe. Fig. 5-9 shows some example images extracted from the recorded videos with their category label. It can be seen the Fig. 5-9 that all of this data was collected inside our laboratory and the person was walking on a treadmill.

TABLE 1. The architecture of the CNN.

Layer (type)	Output Shape
input_1 (Input Layer)	(None, 224, 224, 3)
block1_conv1 (Conv2D)	(None, 224, 224, 64)
block1_conv2 (Conv2D)	(None, 224, 224, 64)
block1_pool (MaxPooling2D)	(None, 112, 112, 64)
block2_conv1 (Conv2D)	(None, 112, 112, 128)
block2_conv2 (Conv2D)	(None, 112, 112, 128)
block2_pool (MaxPooling2D)	(None, 56, 56, 128)
block3_conv1 (Conv2D)	(None, 56, 56, 256)
block3_conv2 (Conv2D)	(None, 56, 56, 256)
block3_conv3 (Conv2D)	(None, 56, 56, 256)
block3_conv4 (Conv2D)	(None, 56, 56, 256)
block3_pool (MaxPooling2D)	(None, 28, 28, 256)
block4_conv1 (Conv2D)	(None, 28, 28, 512)
block4_conv2 (Conv2D)	(None, 28, 28, 512)
block4_conv3 (Conv2D)	(None, 28, 28, 512)
block4_conv4 (Conv2D)	(None, 28, 28, 512)
block4_pool (MaxPooling2D)	(None, 14, 14, 512)
block5_conv1 (Conv2D)	(None, 14, 14, 512)
block5_conv2 (Conv2D)	(None, 14, 14, 512)
block5_conv3 (Conv2D)	(None, 14, 14, 512)
block5_conv4 (Conv2D)	(None, 14, 14, 512)
block5_pool (MaxPooling2D)	(None, 7, 7, 512)
average_pooling2d_1	(None, 1, 1, 512)
flatten (Flatten)	(None, 512)
dense_1 (Dense)	(None, 512)
dropout_1 (Dropout)	(None, 512)
dense_2 (Dense)	(None, 3)

TABLE 2. PC Specifications.

Name	Detail
CPU	AMD Ryzen 7 2700 X Eight-core processor @ 3.70 GHz
Memory	16 GB
GPU	NIVIDIA GeForce GTX 1080 Ti

III. RESULTS AND DISCUSSION

The implementation of the proposed approach presented in this work was done in real-time by distributing the

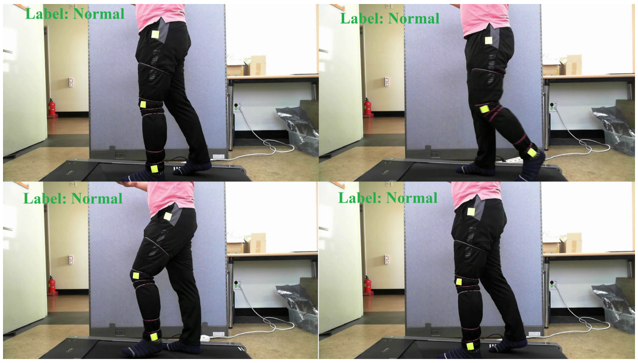


FIGURE 5. Example of the recorded data for category Normal.

FIGURE 6. Example of the recorded data for category Abnormal Left Foot.

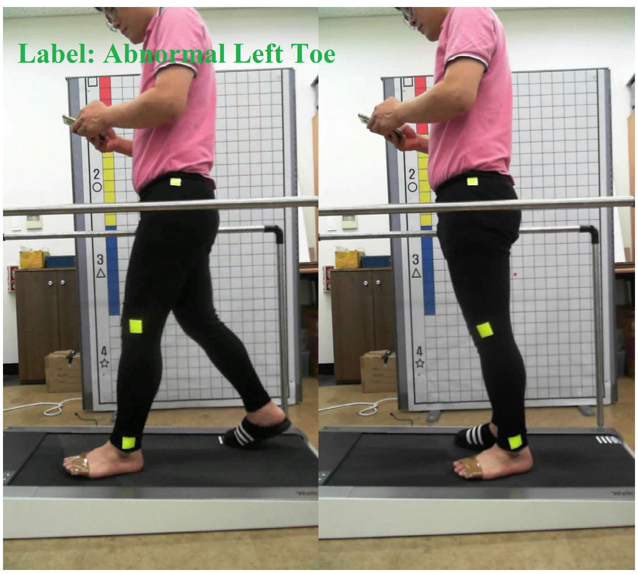


FIGURE 7. Example of the recorded data for category Abnormal Left Toe.

computational load on CPU and GPU, working in parallel to reduce the computational time of the CNN. The recorded time to process a single frame and classifying the category

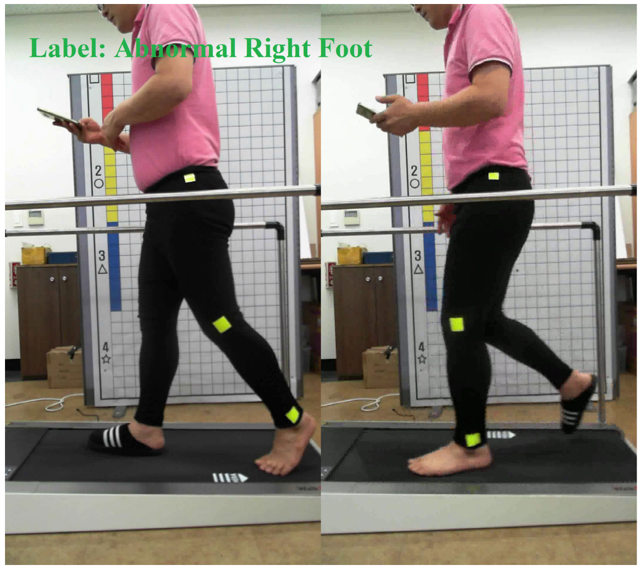


FIGURE 8. Example of the recorded data for category Abnormal Right Foot.

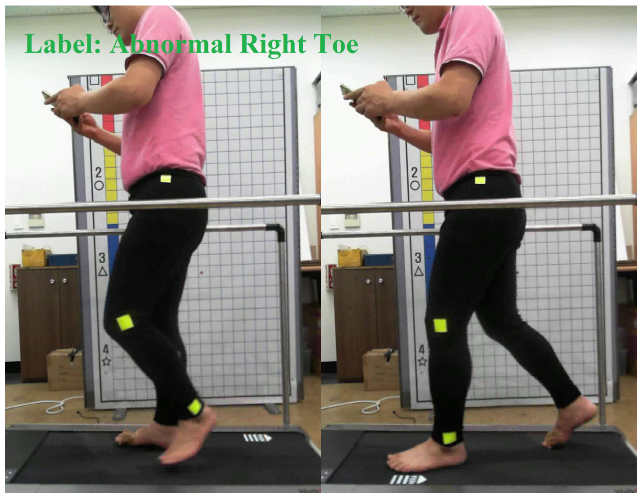


FIGURE 9. Example of the recorded data for category Abnormal Right Toe.

was 47ms. The FPS achieved during this process was almost 20 FPS. We compared the computational time and FPS of our network with a CNN of ResNet50 architecture. For ResNet50, the time it took to process a single frame was almost 12ms higher than proposed CNN with FPS of 15FPS.

The results obtained are shown in Fig. 10-13. The results shown in these figures are respective to the Fig. 6-9. Fig. 10-13 shows the 25 body points skeletal images extracted using human pose estimation. These extracted body points are conjoined to form a respective skeletal image which is further processed through the CNN-based classifier for classification of the categories. It should be noticed that in this work 25 body points are used to classify a gait abnormality problem related to just human feet. The reason to extract and use 25 body points rather than just footpoints is due to the fact that the trained CNN achieved more accurate results when more body points information was provided with a very low confusion rate.

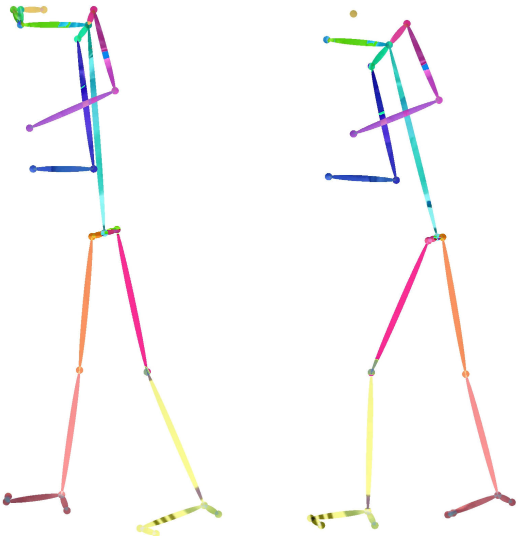


FIGURE 10. Extracted skeletal image using pose estimation for Fig. 6.

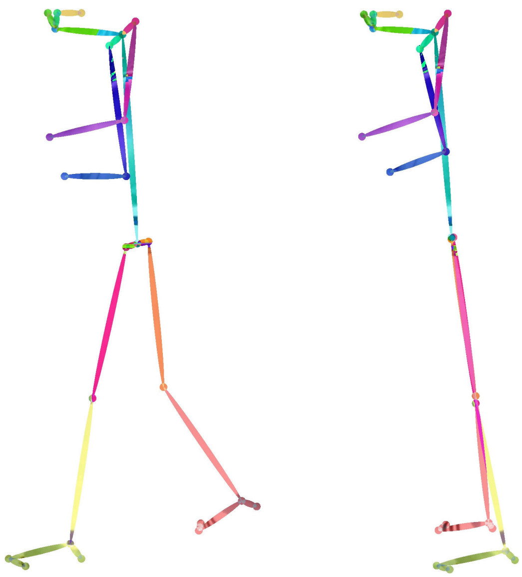


FIGURE 11. Extracted skeletal image using pose estimation for Fig. 7.

To evaluate the performance of the classifier we applied the following three metrics:

(1) Sensitivity: This metric is defined as the proportion of true positives images that are correctly classified by a classifier. This metric is calculated by taking into account the True Positives (*Tpos*) and False Negatives (*Fneg*) of the detected class as given by (1).

$$Sensitivity = \frac{Tpos}{Tpos + Fneg} \tag{1}$$

(2) Precision: This metric is a widely used metric and is defined as the proportion of True Positives among all the detected classes of the system and is given by (2).

$$Precision = \frac{Tpos}{Tpos + Fpos} \tag{2}$$

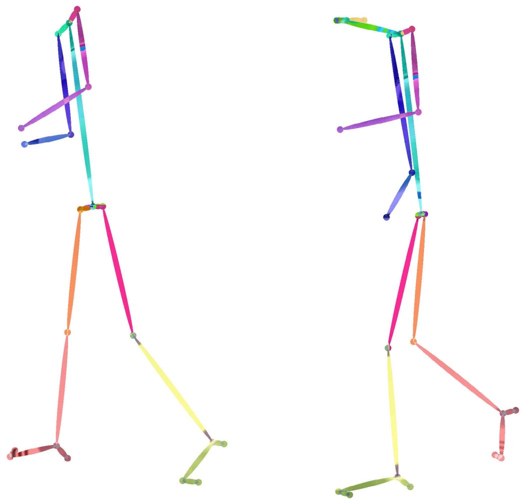


FIGURE 12. Extracted skeletal image using pose estimation for Fig. 8.

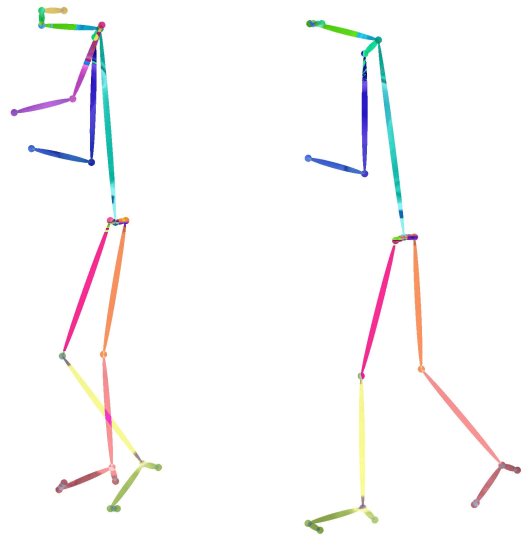


FIGURE 13. Extracted skeletal image using pose estimation for Fig. 9.

(3) Frames-Per-Second (FPS): FPS is the rate at which a classifier is capable of processing incoming camera frames.

In order to compute the overall performance of the classifier, we define a composite score metric [30]. The reason to create such matric is to measure overall performance including the FPS which is a very important parameter to analyze computer vision and image processing applications. Whereas, the only accuracy was used to determine the performance of the CNN-based classifier. This metric consists of a linear combination of Sensitivity and Precision together with the achieved FPS. We parametrize the score with respect to a vector of weights $w \in [0, 1]^3$ as given by (3). We prioritized FPS with a weight of 0.4 over the other two accuracy-related metrics because FPS is a more prominent factor in performance evaluation of the overall system, whereas other parameters were equally weighted with 0.2.

$$Score\,(w) = w_1 \times FPS + w_2 \times Sensitivity + w_3 \times Precision \tag{3}$$

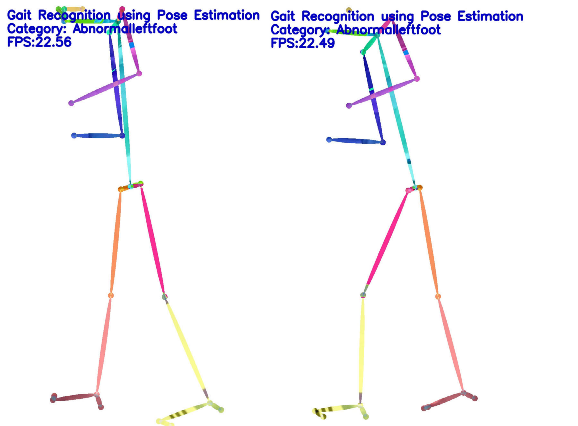


FIGURE 14. **Output result of the CNN-based classifier with respect to Fig.6, and Fig. 10 with detected category as Abnormal Left Foot.**

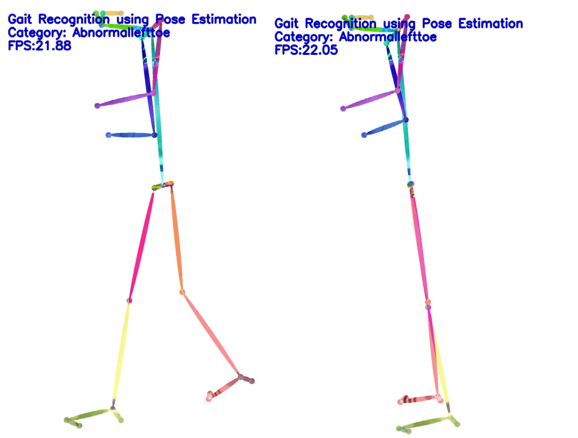


FIGURE 15. **Output result of the CNN-based classifier with respect to Fig.7, and Fig. 11 with detected category as Abnormal Left Toe.**

With *Tpos* = 143, 000, *Fneg* = 3000 and *Fpos* = 4000 out of 150,000 images database obtained from 50 videos. The calculated score with a sensitivity of 0.9795 and a precision of 0.9728 for the classifier and average FPS of 20 was 8.39. We compared our classifier with a ResNet50 architecture with the same database and the score calculated for the ResNet50 was 6.24. More importantly, the accuracy achieved for the classifier was 97.3% higher than ResNet50's 96.5%.

Fig. 14-17 shows the results achieved at the output of the classifier. These results are related to Fig. 6-9 used as an input for human pose estimation and Fig. 10-13 given as an input to the CNN-based classifier. The classifier returns the output with the category name. in this work, the classification threshold was kept at 90%. The CNN only returned the images of the respective class, if the said classification threshold was achieved. Furthermore, the proposed approach was tested by changing the walking speed with different view angles of the person and the results found were equally promising. It is necessary to provide data with different view angles and walking speed while training the classifier in order to create a proper gait analysis mechanism. The advantage of the proposed approach is that it eliminates the impact of other

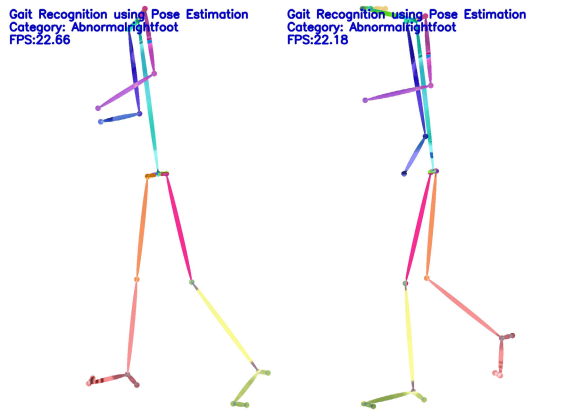


FIGURE 16. **Output result of the CNN-based classifier with respect to Fig.8, and Fig. 12 with detected category as Abnormal Right Foot.**

FIGURE 17. **Output result of the CNN-based classifier with respect to Fig.9, and Fig. 13 with detected category as Abnormal Right Toe.**

problematic constraints such as clothing, surface, carrying status and shoes as it solely considers the human pose based on body point skeletal images.

As mentioned before, the main idea is to propose an approach using CNN, it can be noticed from the results presented that by just putting a minute difference in a person's toe and feet. CNN was able to distinguish between normal and abnormal gait. That proves that if CNN is trained properly with proper constraints, not it can help performing other tasks such as posture analysis and also can easily distinguish between gaits of different people. In the future, we are working on more precise whole-body points posture analysis and this work is one of the successful steps towards that.

IV. CONCLUSION

In this work, an approach to develop an efficient gait analysis mechanism using deep learning tools such as Convolutional Neural Network (CNN)-based classifier is presented. The proposed approach uses the human pose estimation method to classify the abnormalities found in a specific person's gait. By introducing a proper data collection and training scheme for CNN, the proposed approach addresses the problems related to gait analysis methods used previously. The experimental results are promising in solving some of the typical problem constraints involved in the gait analysis system. The accuracy achieved for classifying normal and abnormal gait is 97.3% that proves the applicability of the proposed mechanism. Furthermore, in the future, the system can be used to distinguish different people's gait by adding more data in the training process of the classifier.

REFERENCES

[1] W. Pirker and R. Katzenschlager, "Gait disorders in adults and the elderly: A clinical guide," *Wiener klinische Wochenschrift*, vol. 129, nos. 3–4, pp. 81–95, Feb. 2017, doi: 10.1007/s00508-016-1096-4.

[2] J. R. Jørgensen, D. T. Bech-Pedersen, P. Zeeman, J. Sørensen, L. L. Andersen, and M. Schönberger, "Effect of intensive outpatient physical training on gait performance and cardiovascular health in people with hemiparesis after stroke," *Phys. Therapy*, vol. 90, no. 4, pp. 527–537, 2010.

[3] R. D. Sanders and P. M. Gillig, "Gait and its assessment in psychiatry," *Psychiatry (Edgmont)*, vol. 7, no. 7, pp. 38–43, 2010.

[4] W. Khan and A. Badii, "Pathological gait classification and segmentation by processing the hip joints motion data to support mobile gait rehabilitation," *Res. C Med. Eng. Sci.*, vol. 7, no. 3, 2019, doi: 10.31031/RMES.2019.07.000662.

[5] K. K. Haussler, "Equine manual therapies in sport horse practice," *Veterinary Clinics North America: Equine Pract.*, vol. 34, no. 2, pp. 375–389, Aug. 2018.

[6] P. Gottipati, S. Fatone, T. Koski, P. A. Sugrue, and A. Ganju, "Crouch gait in persons with positive sagittal spine alignment resolves with surgery," *Gait Posture*, vol. 39, no. 1, pp. 372–377, 2014, doi: 10.1016/j.gaitpost.2013.08.012.

[7] M. J. Kennedy, M. Lamontagne, and P. E. Beaulé, "Femoroacetabular impingement alters hip and pelvic biomechanics during gait: walking biomechanics of FAI," *Gait Posture*, vol. 30, no. 1, pp. 41–44, Jul. 2009, doi: 10.1016/j.gaitpost.2009.02.008.

[8] M. J. Marin-Jimenez, F. M. Castro, N. Guil, F. de la Torre, and R. Medina-Carnicer, "Deep multi-task learning for gait-based biometrics," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Beijing, China, Sep. 2017, pp. 106–110, doi: 10.1109/ICIP.2017.8296252.

[9] T. M. Parker, L. R. Osternig, P. van Donkelaar, and L.-S. Chou, "Balance control during gait in athletes and non-athletes following concussion," *Med. Eng. Phys.*, vol. 30, no. 8, pp. 959–967, 2008, doi: 10.1016/j.medengphy.2007.12.006.

[10] Z. Liu and S. Sarkar, "Simplest representation yet for gait recognition: Averaged silhouette," in *Proc. 17th Int. Conf. Pattern Recognit.*, vol. 4, Aug. 2004, pp. 211–214.

[11] J. Han and B. Bhanu, "Individual recognition using gait energy image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 2, pp. 316–322, Feb. 2006.

[12] K. Bashir, T. Xiang, and S. Gong, "Gait recognition using gait entropy image," in *Proc. 3rd Int. Conf. Imag. Crime Detection Prevention (ICDP)*, 2009, pp. 1–6.

[13] M. Jeevan, N. Jain, M. Hanmandlu, and G. Chetty, "Gait recognition based on gait pal and pal entropy image," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 4195–4199.

[14] H. Iwama, M. Okumura, Y. Makihara, and Y. Yagi, "The ou-isir gait database comprising the large population dataset and performance evaluation of gait recognition," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 5, pp. 1511–1521, Oct. 2012.

[15] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cogn. Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991.

[16] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[17] D. Xu, S. Yan, L. Zhang, S. Lin, H.-J. Zhang, and T. S. Huang, "Reconstruction and recognition of tensor-based objects with concurrent subspaces analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 1, pp. 36–47, Jan. 2008.

[18] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, and H.-J. Zhang, "Discriminant analysis with tensor representation," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2005, pp. 526–532.

[19] D. Xu, S. Yan, D. Tao, L. Zhang, X. Li, and H.-J. Zhang, "Human gait recognition with matrix representation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 7, pp. 896–903, Jul. 2006.

[20] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and Gabor features for gait recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, 29, no. 10, pp. 1700–1715, Oct. 2007.

[21] Z. Lai, Y. Xu, Z. Jin, and D. Zhang, "Human gait recognition via sparse discriminant projection learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 10, pp. 1651–1662, Oct. 2014.

[22] X. He and P. Niyogi, "Locality preserving projections," in *Proc. NIPS*, vol. 16, 2003, pp. 153–160.

[23] M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis," *J. Mach. Learn. Res.*, vol. 8, pp. 1027–1061, May 2007.

[24] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[25] E. P. Ijjina and C. K. Mohan, "One-shot periodic activity recognition using convolutional neural networks," in *Proc. 13th Int. Conf. Mach. Learn. Appl.*, Dec. 2014, pp. 388–391.

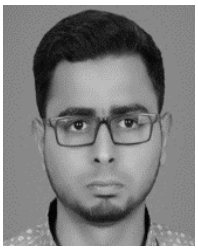
[26] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[27] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1725–1732.

[28] Zhe Cao, Tomas Simon, Shih-En Wei, Yaser Sheikh, "Realtime multi-person 2D pose estimation using part affinity fields," CoRR 2016, *arXiv:1611.08050*. [Online]. Available: https://arxiv.org/abs/1611.08050

[29] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: Realtime multi-person 2D pose estimation using part affinity fields," 2018, *arXiv:1812.08008*. [Online]. Available: https://arxiv.org/abs/1812.08008

[30] A. Rohan, M. Rabah, and S.-H. Kim, "Convolutional neural network-based real-time object detection and tracking for parrot AR drone 2," *IEEE Access*, vol. 7, pp. 69575–69584, 2019, doi: 10.1109/ACCESS.2019.2919332.



ALI ROHAN (Associate Member, IEEE) received the B.S. degree in electrical engineering from The University of Faisalabad, Pakistan, in 2012, and the M.S. and Ph.D. degrees in electrical, electronics, and control engineering from Kunsan National University, South Korea, in 2018 and 2020, respectively. From 2012 to 2013, he worked as a Development Engineer at the Niagara Group of Industries, Pakistan. From 2013 to 2015, he worked as a Project Engineer for Circle Club, Pakistan. From 2015 to 2016, he worked as a Project Manager for Steam Masters, Pakistan, and also as a Lecturer at the Department of Electrical and Telecommunication Engineering, Government College University Faisalabad, Pakistan. From 2016 to 2020, he worked as a Research Associate with the Factory Automation and Intelligent Control Lab., Kunsan National University. He is currently working as an Assistant Professor with the Department of Mechanical, Robotics, and Energy Engineering, Dongguk University, South Korea. His research interests includes machine learning, AI, UAV's, power electronics, fuzzy logic, EV systems, image processing, computer vision, and prognostics and health management (PHM).

MOHAMMED RABAH received the B.A. degree in electronics and communication engineering from the AL-SAFWA High Institute of Engineering, Egypt, in 2015, and the M.S. degree in electronics and information engineering and the Ph.D. degree from Kunsan National University, South Korea, in December 2017 and August 2020, respectively. He is currently working as a Research Engineer with the Turku University of Applied Sciences, Turku, Finland. His research interests includes automation, control and intelligent systems. Furthermore, he is also interested in UAV's applications, fuzzy systems, and deep learning.

TAREK HOSNY received the B.S., M.S., and Ph.D. degrees in communication engineering from the Shoubra Faculty of Engineering, Cairo, Egypt, in 2006, 2014, and 2019, respectively. He is currently a Lecturer with the Al-Safwa High Institute of Engineering, Communication Engineering Department, Cairo, Egypt. His research interests include LI-FI, visible light communication systems, machine learning, and optics.

SUNG-HO KIM received the B.S. degree in electrical engineering from Korea University, in 1984, the M.S. and Ph.D. degrees in electrical engineering from Korea University, in 1986 and 1991, respectively, and the post-doc from Japan Hiroshima University, in 1996. He is currently a Professor with Kunsan National University. His research interests includes fuzzy logic, sensor networks, neural networks, intelligent control systems, renewable energy systems, fault diagnosis systems.