E-ISSN: 2583-9667
Indexed Journal
Peer Reviewed Journal
https://multiresearchjournal.theviews.in

Received: 02-04-2024
Accepted: 03-06-2024

INTERNATIONAL JOURNAL OF ADVANCE RESEARCH IN MULTIDISCIPLINARY

Volume 2; Issue 3; 2024; Page No. 530-533

Develop and integrate a system that combines both face and gait recognition for enhanced security

[1]Ravindra Suresh Kamble and [2]Dr. Amit Singhal

[1]Research Scholar, Monad University, Hapur, Uttar Pradesh, India
[2]Professor, Monad University, Hapur, Uttar Pradesh, India

Corresponding Author: Ravi Shankar Mishra

Abstract

An important biometric method, gait recognition may identify people from a distance by their distinct gait pattern. Since the advent of deep learning, computing power has steadily increased, making automated gait analysis a breeze. Unfortunately, there are a number of factors that may impact the accuracy of identification, such as the wearer's attire, the things they are holding, the viewing angle, the presence or absence of occlusion, and malicious alterations to the datasets. Deep learning is the method of choice for this complicated issue since it can handle massive datasets. Here, we built a deep learning pipeline to process the gait dataset and a number of gait factors. Instead of using separate deep learning pipelines for each of the gait variables, as was proposed in the work, we have built a universal and thorough pipeline based on what we know about all of the gait pipelines; we have also tested our results on new datasets. By doing so, we may use a single deep learning algorithm to handle almost all of the variables. This includes dealing with gait variables based on appearance and pose-based methods. A number of strategies have been put forth in recent years that combine two types of biometrics-face, which is a physical biometric, and gait, which is a behavioural biometric-in an effort to determine if this combination outperforms methods that use just one of these. In this article, we address the pros and cons of these systems, provide an overview of some of the most well-known methods to this problem, and look at the possibilities for future studies and applications of this technology.

Keywords: Deep learning, biometric, technology, pipelines, breeze

1. Introduction

The way a person moves while walking or running is called their gait pattern. A person's gait is the pattern of their limb movements while they walk. The discovered gait pattern is quite distinctive and personal. In the criminal justice system, this pattern is taken as proof. The gait pattern seems to have a significant role in determining a person's height, health, age, and gender. A person's gait is their typical manner of moving from one place to another. One of the unconscious behaviours that might be used to identify a person is walking. Every individual has their own unique gait that reflects their innate biology. Many things influence how a person walks, including their gender, the shoes they wear, their mental stability, how fast they walk, any health problems they may have, their age, if they are pregnant, and the lighting conditions in their environment. Most incidents involving dacoits, thieves, murderers, robbers, vandals, kidnappers, etc., exhibit the gait pattern. It seems that a common pattern connecting the crime and the criminal is gait pattern in crime sequence. Additionally, it is useful for figuring out how many people are believed to have been engaged in the crime.

There are two main approaches to gait analysis: the model-based technique and the model-free method. The model-based method makes use of the strain generated by the subject's stride parameters, while the model-free method makes use of the subject's physical posture, style of motion, and gait silhouette. These days Because it examines features including stability, originality, universality, and durability, gait pattern analysis is crucial for forensic reasons. Personal identity, health problems, and walking endurance may all be improved using gait pattern analysis. It is possible to determine a person's identity by drawing conclusions about:

Muscle strength is directly proportional to one's age, making it a key component in determining gait patterns. It is possible to tell a person's age by their stride pattern. The elderly often walk by dragging the ground under their toes. In contrast, toddlers and preschoolers exhibit clumsy toe and ball of the foot motions. The average step length for those aged 65 and above will be less. Along with the footsteps, the stick traces suggest an elderly guy.

The length of a person's stride is a good indicator of their gender. In comparison to males, females have shorter step lengths. Females often have steps that are 18–22 inches long, whereas men typically have steps that are 25–28 inches long. Compared to men, girls have a narrower ball of the foot. Barefoot walking also leaves a mark on the toes of certain Indian ladies who wear rings there. As a sign of their heritage, women's ring use is particularly prevalent among Indian women.

A tall footprint with a ladder length of up to 30 inches indicate that the person is tall, but the length of the footprint alone cannot reveal their height. The one that is shorter will have fewer steps.

2. Problem Description

1. In the developing technological era, there are diverse methods to identify the person. Even though there are lots of constraints, the needed requirements are attended and given priority by the research community
2. Started working on a new convolutional neural network for deep learning to identify people in general without their awareness. Crime prevention, attendance tracking, vehicle identification, and similar fields may all benefit from this technology.
3. Convolutional neural networks are used to train the system, which involves recording videos in both natural and artificial lighting conditions, then converting them into silhouettes.
4. Thirdly, the deep learning convolutional neural network's performance is evaluated using CASIA datasets B and C.
5. The dataset is refined and memory space is reduced using the following methods: grouping dataset, systematic random removal, blur image removal based on the Laplacian algorithm, and contour based broken image removal.
6. In the final activation layer, a mix of the softmax and sparsemax functions is used to replace the softmax function, which is used to enhance the deep learning Convolutional Neural Network (CNN).

3. Review of Literature

Sun *et al*. [2014] [11] discovered that real CCTV face photographs are often LR images, which significantly hinders the performance of existing face recognition algorithms and makes face recognition impractical to use. Combining HR and LR facial photos was the main technical focus of this inquiry. This research detailed a Classifier Share Deep Network for Low Resolution Face Recognition (LRFR) with Multi-Hierarchy Loss (CS-MHL-Net) as a potential solution to this issue. To begin, since comparing failure and its variants does not result in network convergence and a decrease in discrepancy, it is recommended that HR and LR use a similar classification method to share the related weights, which can be seen as the class centre, and thus further reduce the domain distance between them. In order to maximise the utilisation of intermediate features and loss constraints, this paper utilised multi-hierarchy loss in intermediate layers. The objective was to decrease the disparity between HR and LR intermediate functions after maximum pooling and to avoid a decrease in reliability caused by the excessive use of intermediate functions. We confirm that LFW and SC-face are efficient, and we show that this method outperforms sample-based methods.

In realistic video surveillance, Gao *et al*. [2020] [1] found that large distances between objects and monitoring cameras typically impact the reliability of the face regions of interest, which in turn impair recognition performance. Present methods often consider holistic representations, ignoring supplementary information from different patch sizes. This work presents a multi-scale MSPRFL approach to address the challenge of low-resolution face recognition. First, using a training dataset and a more robust resolution, the suggested MSPRFL technique employs multi-level knowledge to precisely characterise each patch. By merging the detection outcomes of all patches, this makes advantage of the knowing resolution-robust picture characteristics to narrow the resolution gap. Lastly, it employs an ensemble optimisation approach to learn scale weights, with the goal of combining the findings from many scales in order to comprehend the complimentary discriminatory potential of different patch scales.

4. Objectives of the study

1. Develop and integrate a system that combines both face and gait recognition for enhanced security.
2. Design and train deep learning models specifically tailored for face and gait recognition tasks.

5. Research Methodology

Well standardized bench marking CASIA dataset B and CASIA dataset C are available in internet, as open access is taken and used from the site http://www.cbsr.ia.ac.cn/english/Gait%20Databases.asp for training and testing the convolutional neural network. Two different datasets captured in day time and in night time are used to train the images.

Dataset B consists of 124 person silhouettes with 11 views including various angles, clothing and accessories. It is a large multi view dataset for gait process. It is captured during the day light with normal video camera set at various degrees of angles to capture the walking styles simultaneously.

Dataset C consists of 153 person silhouettes with various walking conditions like normal walking with and without bags, slow walking and fast walking. The videos are collected during night time with infra-red thermal camera. It is a standard, well refined and validated database collected during 2005 and currently used for research purpose for finding gait features and other gait analysis.

To begin person classification, the video must be transformed into a silhouette. It takes almost a thousand silhouettes to transform a 12-second video. While 32% are utilised for testing, 68% are used for silhouette training on this exam. It is not possible to feed the neural network a full dataset all at once. Consequently, the datasets are organised into batches. Epoch stands for the dataset's forecast. For instance, with a batch size of 20 and a data set containing 1000 photos, the epoch should execute 50 iterations. Classification is then applied to CASIA datasets B and C. Colour, size, and form are used to evaluate each image's size, and if necessary, they are converted to greyscale.

6. Results and data interpretation

A deep learning convolutional neural network, developed in Python, is used to conduct the study. Keras is a package that operates on the network architecture is executed by importing the free and default library package, Tensorflow.

The tensorflow keras package generates a deep learning convolutional neural network, and its architecture is shown in table 1 below. The settings, output picture form, and layer type are all detailed. Feature extraction is done using the convolutional and max pooling layers, and person categorisation is done using the layers that follow flatten. In order to simplify things and cut down on training time, the dropout range is set to half, or 0.5. We have set the Epoch settings to 25. We have set the batch size to 50. In order for the deep learning convolutional neural network to acquire weight, it is run 25 times using the provided dataset. Call backs are triggered after 25 repetitions. The epochs are terminated sooner and the best weights are recorded in a separate file on disc based on the loss accuracy. You may use the best weight file to make predictions about people whenever you want after training the dataset. At any point after training, it may be used for human categorisation. It is possible to load the optimal weight file several times in order to forecast the individual

Table 1: Architecture of deep convolutional neural network obtained from Keras package

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 128, 128, 32)	896
activation_1 (Activation)	(None, 128, 128, 32)	0
conv2d_2 (Conv2D)	(None, 126, 126, 32)	9248
activation_2 (Activation)	(None, 126, 126, 32)	0
conv2d_3 (Conv2D)	(None, 124, 124, 32)	9248
activation_3 (Activation)	(None, 124, 124, 32)	0
max_pooling2d_1 (MaxPooling2)	(None, 62, 62, 32)	0
dropout_1 (Dropout)	(None, 62, 62, 32)	0
conv2d_4 (Conv2D)	(None, 60, 60, 32)	9248
activation_4 (Activation)	(None, 60, 60, 32)	0
conv2d_5 (Conv2D)	(None, 58, 58, 64)	18496
activation_5 (Activation)	(None, 58, 58, 64)	0
conv2d_6 (Conv2D)	(None, 56, 56, 32)	18464
activation_6 (Activation)	(None, 56, 56, 32)	0
max_pooling2d_2 (MaxPooling2)	(None, 28, 28, 32)	0
dropout_2 (Dropout)	(None, 28, 28, 32)	0
conv2d_7 (Conv2D)	(None, 26, 26, 32)	9248
activation_7 (Activation)	(None, 26, 26, 32)	0
conv2d_8 (Conv2D)	(None, 24, 24, 64)	18496
activation_8 (Activation)	(None, 24, 24, 64)	0
conv2d_9 (Conv2D)	(None, 22, 22, 32)	18464
activation_9 (Activation)	(None, 22, 22, 32)	0
max_pooling2d_3 (MaxPooling2)	(None, 11, 11, 32)	0
dropout_3 (Dropout)	(None, 11, 11, 32)	0
flatten_1 (Flatten)	(None, 3872)	0
dense_1 (Dense)	(None, 160)	619680
activation_10 (Activation)	(None, 160)	0
dropout_4 (Dropout)	(None, 160)	0
dense_2 (Dense)	(None, 30)	4830
activation_11 (Activation)	(None, 30)	0

Total params: 736,318
Trainable params: 736,318
Non-trainable params: 0

The accuracy rate and time are affected by the number of parameters. The settings primarily in the convolutional and dense layers are tuned in order to decrease the training time and accommodate the employment of silhouettes. According to the results of running Keras Tensorflow, there are a total of 7,36,318 training parameters.

Images from CASIA datasets B and C were used in this study. The enhanced deep learning convolutional neural network receives the datasets and uses them for training and testing purposes. The precision improves in direct proportion to the number of epochs. Accuracy and validation accuracy are recorded after callbacks in the network evaluation. We classify the people using best weights and record the top ten results. The average is computed to determine the approximate degree of accuracy in the person's categorisation or prediction.

Table 2: Dataset classification with training time and accuracy

Dataset	No. of Persons trained	No. of Silhouettes	Epoch	Accuracy	Training Time (Hours)
Casia B	124	11,18,373	25	99.12	51.895
Casia C	153	1,00,346	25	99.23	6.058

The table 2 elaborates the number of persons walking styles are captured in the CASIA dataset B and CASIA dataset C, number of silhouettes used for training and testing, number of epochs used for training the dataset, accuracy and the training time in hours. Initially the time for training the machine was too high. As training increases, the classification accuracy also increases.

The person count is high for CASIA dataset C but the silhouettes count is less while compared to CASIA dataset B. In CASIA dataset B, 124 persons images are trained and the silhouettes count ranges 11,18373. In CASIA dataset C, 153 persons images are trained and the silhouettes ranges 1,00,346.

The accuracy level for CASIA dataset B starts from 68.21% and in 25 epochs and call backs, it gradually increased to 99.12% for testing. Overall time taken for training the CASIA dataset B is 51.895 hours.

The accuracy level for CASIA dataset C starts from 76.5% and in 25 epochs and callbacks, it gradually increased to 99.23% for testing. Overall time taken for training is 6.058 hours.

Table 3: Classification accuracy of Casia dataset B

Prediction/Dataset	CASIA Dataset B	Timing (sec)
1	96.83	65
2	96.92	70
3	97.93	80
4	99	62
5	97.96	90
6	96.92	75
7	97.85	69
8	97.91	68
9	97.96	98
10	99	63
Average	97.828	74

Table 3 states the accuracy level and the prediction timing in hours for the last 10 predictions of CASIA dataset B from the saved best weights. The accuracy level varies from 96.73% to 99%. The speed of the last ten predictions was more or less stable. The timing surrounds from 1 minute to 2 minutes i.e.) the timing ranges from 62 seconds to 98 seconds.

7. Conclusion

In order to build a system that can function better in a broader variety of settings than a classifier that uses just one of these biometrics the face and gait a behavioural biometric-this study aimed to augment and integrate the two. A system that could recognized both faces and gait was so suggested. In order to identify the unknown individual, a decision-level fusion method was used, in which the gait classifier was given the top matches from the face classifier. In order to recognize faces, we used the eigenfaces method and a Bayesian inference-based classifier from our earlier work. To recognize gaits, we used a model-based strategy, rather than a holistic one, and we used data from an optoelectronic motion capture system to identify which gait features were most relevant for recognition.

An overview of gait analysis, deep learning, and convolutional neural networks was provided at the outset of this study. As an obvious first step in doing quality research, a number of literature reviews were reviewed. Cameras with a survey lens and cell phones are used to record the footage. Silhouettes are created from videos. Public databases such as CASIA datasets B and C were used in the proposal and implementation of a deep learning convolutional neural network. The sample size is decreased by the use of several sampling procedures, such as the grouping method and the systematic random elimination approach. The activation function is tweaked to make it more accurate and to boost performance. In order to compare the two networks' performance, a hybrid of the softmax and sparsemax functions is used in the final activation layer. This is then conducted using different sampling techniques.

8. References

1. Jain A, Hong L, Pankanti S. Biometric identification. Communications of the ACM. 2020;43(2):90-98.
2. Zhao W, Chellappa R, Phillips PJ, Rosenfeld A. Face recognition: A literature survey. ACM Computing Surveys (CSUR). 2023;35(4):399-458.
3. Jain AK, Li SZ. Handbook of face recognition. Vol. 1. New York: Springer; c2021.
4. Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001). 2001;1:1-2. IEEE.
5. Bagherian E, Rahmat RWO. Facial feature extraction for face recognition: A review. In: 2008 International Symposium on Information Technology; 2008;2:1-9. IEEE.
6. Tian YL, Kanade T, Cohn JF. Evaluation of Gabor-wavelet-based facial action unit recognition in image sequences of increasing complexity. In: Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition; c2022. p. 229-234. IEEE.
7. Huang GB, Mattar M, Berg T, Learned-Miller E. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Computer Vision and Image Understanding; c2018.
8. Wolf L, Hassner T, Maoz I. Face recognition in unconstrained videos with matched background similarity. In: CVPR 2011; c2011. p. 529-534. IEEE.
9. Setty S, Husain M, Beham P, Gudavalli J, Kandasamy M, Vaddi R, *et al*. Indian movie face database: A benchmark for face recognition under wide variations. In: 2023 Fourth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG); c2023. p. 1-5. IEEE.
10. Ng HW, Winkler S. A data-driven approach to cleaning large face datasets. In: 2014 IEEE International Conference on Image Processing (ICIP); c2014. p. 343-347. IEEE.
11. Sun Y, Wang X, Tang X. Deep learning face representation from predicting 10,000 classes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; c2014. p. 1891-1898.
12. Klare BF, Klein B, Taborsky E, Blanton A, Cheney J, Allen K, *et al*. Pushing the frontiers of unconstrained face detection and recognition: IARPA Janus Benchmark A. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; c2015. p. 1931-1939.
13. Parkhi OM, Vedaldi A, Zisserman A. Deep face recognition. In: Proc. BMVC; c2015.
14. Kemelmacher-Shlizerman I, Seitz SM, Miller D, Brossard E. The Megaface benchmark: 1 million faces for recognition at scale. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; c2016. p. 4873-4882.