Research on Multi-Feature Gait Recognition Based on Deep Learning

Yu Si

Southwest University, ChongQing, 400715, China;

Abstract. In recent years, surveillance cameras have been widely used in public safety and criminal investigation fields. In order to improve recognition accuracy, many monitoring devices are equipped with intelligent systems that can capture and analyze visual data and extract information related to suspicious behavior and personnel identity. Although widely used, common biometric recognition technologies such as facial recognition, fingerprint recognition, and iris recognition also face certain security risks, such as criminals using photos to evade recognition or 3D printed masks to deceive cameras. Unlike this, gait recognition technology verifies identity by analyzing an individual's gait characteristics, which has a wider recognition range and stronger robustness. Gait recognition technology has shown advantages, especially in complex environments and long-distance scenes, and is unaffected by masks and occlusions. Gait data such as video images and 3D motion capture data can effectively extract unique gait features of the human body. Due to the large number of dimensions in gait data, accurate feature extraction is crucial for improving recognition accuracy. Currently, gait recognition technology is mostly based on machine learning and deep learning methods, continuously improving recognition models' performance and application potential through training and optimization on large-scale datasets. The gait recognition algorithm proposed in this article is based on the LSTM-DBN network and evaluated on the CASIA-B dataset. As an important benchmark dataset in cross-view gait recognition, the CASIA-B dataset covers gait sequences from different perspectives.

Keywords: Deep learning; Gait recognition; Gait recognition system; Physical representation.

1. Introduction

In recent years, surveillance cameras have been widely used in fields such as public safety and criminal investigation [1], especially with the increasing number of settings in public places. To enhance security, many monitoring devices are equipped with intelligent systems that can capture and analyze visual information and extract data related to suspicious behavior and identity. The current commonly used biometric recognition technology includes facial, fingerprint, and iris recognition. However, these technologies still face certain security risks [2]. For example, criminals may use certain methods to evade recognition mechanisms or utilize the fault-tolerant nature of facial recognition technology to complete identity verification with the help of others' photos. In addition, the remote images obtained by monitoring devices may also significantly impact the accuracy of recognition. As early as 1994, researchers such as Niyogi and Adelson had shown that human gait, which refers to how people walk, can be an effective recognition feature [3].

Gait recognition technology is a cutting-edge technology in global biometric identification, which uses an individual's walking gait to verify identity. Compared with traditional technologies such as facial recognition, it has stronger security and is particularly suitable for complex environments and long-distance scenes. For example, 3D printed [4] masks can accurately reproduce facial features, which can easily lead to misjudgment by monitoring devices. Therefore, the uniqueness of gait recognition technology lies in distinguishing different individuals by analyzing their body structure and gait. Compared with other biometric technologies, gait recognition has a larger monitoring range and can achieve accurate recognition within tens of meters under ordinary high-definition cameras. Research has shown that everyone's gait is unique[5].

Especially during COVID-19, the popularity of masks has greatly affected facial recognition, especially when low light and facial cover recognition accuracy has declined. When facial features cannot be effectively extracted, gait recognition technology has shown significant advantages. Compared with the traditional image recognition method, gait recognition provides more accurate and reliable authentication functions[5]. In addition, gait recognition technology has significant promise for a wide range of applications, including safety surveillance, forecasting of movement patterns, and human-common interaction, with extensive potential for future development [6].

Gait information covers various forms of data, such as video images, electromyography, and three-dimensional kinematics [7]. Among them, 3D motion capture technology is a high-precision optical motion capture system that can collect and record real-time 3D gait data of the human body and quantitatively analyze gait indicators such as time, distance, and kinematic parameters.

Gait data typically contains many dimensions, and not all dimensions are significantly helpful for identity recognition. Therefore, effective feature extraction can effectively reduce data dimensionality, extract information that is most helpful in distinguishing individuals, and optimize the training time and effectiveness of recognition models [8]. Therefore, before gait recognition, it is particularly crucial to choose appropriate denoising methods and screen effective features. High-quality data processing methods can capture the uniqueness of individual gait, such as walking rhythm and body posture, thereby improving recognition accuracy and enhancing system robustness. Currently, the main algorithms for gait recognition are based on machine learning and deep learning techniques, which require collecting a large amount of feature data and its fine processing, training, and optimization to construct effective training models [9].

However, due to the influence of individual step size, walking speed, and lower limb muscle fatigue on gait characteristics, there are often significant distribution differences between training and testing samples. Therefore, traditional algorithms often fall into local optima, resulting in unsatisfactory classification performance [10]. The current research focuses on improving the recognition accuracy and stability of algorithms for more effective gait analysis.

On this basis, deep learning techniques are widely considered to be very suitable for solving the challenges in gait recognition. Deep learning can automatically extract hierarchical features from raw data, avoiding the complexity and limitations of traditional manual feature extraction processes. Through multi-layer neural networks, deep learning models can learn more abstract gait features, such as individual unique step rhythms, motion trajectories, etc., and effectively handle noise and variability in the data. In addition, deep learning techniques have demonstrated powerful capabilities in processing large-scale datasets, optimizing feature representations during training, and significantly improving recognition accuracy and stability.

In the field of gait recognition, deep learning has demonstrated its outstanding adaptive characteristics, which can maintain excellent classification performance even in the face of different individual gait features, environmental changes, and data noise. Therefore, deep learning can not only effectively reduce the impact of distribution differences during training but also help enhance models' robustness and generalization ability, making gait recognition systems more efficient and reliable[11].

2. Principles of Deep Learning

Over the past few years, the swift advancement in deep learning has spurred the creation of numerous innovative algorithms for gait-based identity verification. However, the advantages and disadvantages of these algorithms need to be validated on a unified and standardized dataset. Therefore, the role of datasets in artificial intelligence and machine learning is crucial. The development and validation of gait recognition models heavily rely on specially designed datasets containing rich covariate factors such as viewpoint changes, clothing types, carried items, different ground types, and indoor and outdoor environmental conditions, to ensure that the algorithm can be fully trained in diverse scenarios. These different factors enhance the model's robustness and ensure its ability to deal with different challenges and complex situations in practical applications.

At present, representative databases in the field of gait recognition include OU-ISIR, OU-ISIR LP Bag, and CASIA Dataset-B. The OU-ISIR dataset contains rich gait information involving 4007 participants of different age groups and gait features, providing important benchmark data support for gait recognition methods. The OU-ISIR LP Bag database mainly focuses on the impact of carrying items on gait recognition, containing gait sequences of 62528 subjects under different carrying conditions, suitable for studying the interference of carrying items on gait recognition. CASIA dataset - b is a commonly used multi-view dataset in the field of gait recognition, containing gait information from 124 participants and covering 11 videos from different perspectives. This dataset considers multiple covariate factors such as perspective changes, item carrying, and clothing, which can effectively evaluate the performance of algorithms in different environments. These datasets provide strong support for the training, validation, and improvement of gait recognition algorithms due to their rich diversity, wide scale, and diverse covariate factors, promoting the progress and innovation of gait recognition technology in practical applications.

LSTM is a special recurrent neural network (RNN). Compared with traditional RNN, LSTM effectively overcomes the limitations of traditional RNN in dealing with long-term dependence by introducing a series of innovative mechanisms, such as input gate, output gate, forgetting gate and memory unit. When facing long time interval events, it is difficult for ordinary RNNs to retain long-term information due to the problem of gradient disappearance or explosion. LSTM solves this problem through these mechanisms, so that it can better handle and predict long time interval events or information in the time series. The LSTM network is designed to simulate the memory mechanism of the human brain, so it can not only remember important information, but also forget irrelevant content, thus improving the performance of the model in time series data. The core component of LSTM is the storage unit (also called "cell state"), which is the memory part of the network and is responsible for saving important and long-term information in the whole network. The storage unit can transfer effective information from one time step to the next, so as to ensure that the network can capture information with a long time span. In parallel, LSTM networks regulate the flow of information through different gating mechanisms. The input gate controls whether the current input data needs to be added to the storage unit, and determines which information is new, important and worth remembering. The output gate determines when the information in the storage unit is output, helping the model generate the final prediction results based on the current storage information. The forgetting gate plays a crucial role in LSTM, which determines which information in the storage unit needs to be forgotten. The Forgotten Gate avoids long-term storage of irrelevant or redundant data by evaluating and screening information that is no longer needed, and ensures that the network only retains the most valuable memory. This mechanism greatly optimizes the long-term storage.

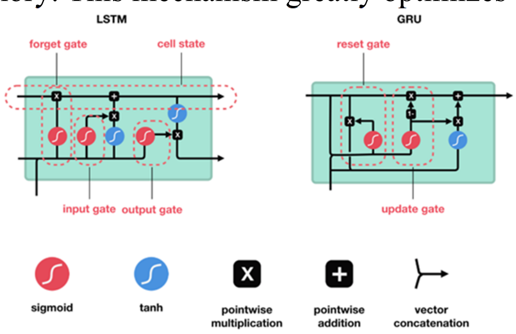

Figure 1. LSTM Unit Structure Diagram

This study's construction of LSTM-DBN aims to extract time sequence features from functional data sets and use the DBN (depth belief network) model for regression analysis and prediction. However, with the increase of the depth and number of parameters of the model, the LSTM-DBN architecture is easy to overfit. To address this issue, we have introduced a Dropout layer. During the training phase, the Dropout layer randomly selects some neurons for "Dropout," which means that these neurons will be excluded from the current forward and backward propagation calculations, thereby reducing the model's dependence on specific neurons. This can improve the model's generalization performance, prevent overfitting, and effectively alleviate the problem. This method plays an important role in enhancing the robustness of the model.

Deep Belief Network (DBN) is a generative deep learning model composed of multiple stacked Restricted Boltzmann Machines (RBM) used to learn advanced feature representations of data. The basic concept of DBN is to use multi-layer unsupervised learning to gradually obtain abstract features of data and then use supervised learning methods to achieve the final classification or regression goals.

3. Comparison of Variable Selection and Recognition Algorithms based on LSTM

Preprocess the original gait image; in gait recognition methods, it is usually avoided to use the original image directly. Use the initial image as input because raw images that have not been cleaned and converted may be affected by various factors, such as light Lines, shadows, background noise, etc. These interfering factors mask the true gait characteristics and may also introduce misleading information. Sexual information has an impact on gait recognition. In order to eliminate or reduce unnecessary variables, CASIA-B is selected in this chapter. Perform a series of image preprocessing operations on the dataset. Weighted average gait energy synthesis from the standard gait contour sequence measurement chart obtained. Further, gait contour sequences with gait energy maps are combined to innovatively propose a hybrid gait model board to ensure that deep spatiotemporal feature information in gait can be extracted more clearly and accurately.

Gait recognition relies on the posture of pedestrians while walking and typically converts the original gait image into a binary form through threshold adaptation. However, the original gait images often contain various noises and interferences, which may cause confusion between the gait of the target pedestrian and the surrounding environment. Therefore, accurately segmenting the target pedestrian from the background becomes a key step in gait image preprocessing. Through effective preprocessing, key information about gait can be highlighted, such as the contour of the target pedestrian and the range of motion of the upper and lower limbs, while reducing noise and interference in the image.

This preprocessing process is crucial for improving the quality of gait features, ensuring that the algorithm focuses on the main features of pedestrians and ignores irrelevant background information. The optimized image data will be more helpful in constructing identity recognition models with high recognition accuracy, thereby improving the performance and robustness of gait recognition systems in practical applications.

The walking process is essentially a periodic reciprocating motion, and the method of extracting pedestrian gait features using gait contour images often loses some spatial information. However, Gait Energy Image (GEI), as an effective representation, can accumulate energy within a walking cycle and fully reflect the relative position changes of contour sequences over time. Specifically, the grayscale value of the gait energy map changes over time, and the increase or decrease in grayscale value represents the size of the pixel energy at that point. This means the degree to which the human body contour covers a certain position throughout the entire walking cycle, thus better revealing the body shape and posture characteristics of the human body.

A gait energy map can effectively capture the motion characteristics during walking. By accumulating multiple time steps, it not only preserves the dynamic characteristics of gait but also effectively enhances the unique identification information related to individual pedestrians. Therefore, GEI provides a more stable and useful feature representation that helps improve the accuracy and robustness of gait recognition systems.

In the experiment, a network consisting of two layers of LSTM was first constructed. A Dropout layer was added after the second layer, followed by another layer of LSTM, and a DBN model was introduced after the third layer, thus forming a complete LSTM-DBN architecture. For the convenience of comparing optimization algorithms in the future, the following hyperparameters were set in this experiment: Dropout dropout rate of 0.21, LSTM layers of 2, and the number of neurons in each layer of 16, 32, and 64, respectively; DBN is set to 2 layers, including 3 RBMs (Restricted Boltzmann Machines), with 64, 32, and 2 neurons in each layer, 20 iterations, a learning rate of 0.05, and 100 reverse fine-tuning iterations.

4. Experimental results and analysis

The CASIA-B dataset is specifically designed for cross-view gait recognition and has been extensively utilized in numerous studies on gait recognition. This paper assesses the performance of the proposed algorithm model using the CASIA-B dataset. The dataset includes data on 124 individuals, each with 10 distinct walking conditions and 11 unique angles for each condition, resulting in 110 gait sequences per individual. To measure recognition accuracy, this article uses the Rank-1 metric.

The CASIA-B dataset does not have an official partitioning method. However, the mainstream partitioning method is shown in Table 4.1. In this experiment, we chose the Large Sample (LT) training strategy to ensure the integrity of the training. Specifically, the experiment used all gait sequences of the first 74 pedestrians out of 124 as the training set. In comparison, the last 50 pedestrians selected the first four normal walking state sequences (NM01-NM04) as the query set, and the remaining six sequences (NM05-NM06, BG01-BG02, CL01-CL02) as the gallery set. This division ensures the model can be trained and validated from diverse perspectives and states.

Table 1. Experimental parameters

Name	parameters
CPU	M4 Pro
GPU	M4 Pro
System	MAC OS
Epochs	1000*100

When analyzing the experimental results of the CASIA-B dataset, it was observed that the model proposed in this paper exhibited high recognition accuracy when dealing with biased frontal perspectives of 0° to 54 ° and 180°. However, when backside (180 °) viewing angles, the accuracy is relatively low. The reason for this phenomenon may be related to the presentation of information about the human body contour from different perspectives.

In a frontal perspective (18 ° to 54 °), the outline of the human body is relatively clear, and the algorithm can extract pedestrian feature information well, thus achieving high recognition accuracy. However, in the front, side, and back perspectives, due to the inability of the camera to capture complete pedestrian dynamic information, the human body contour is relatively blurred, especially in the back perspective, which lacks key information from the front and side perspectives, making it difficult for the algorithm to effectively identify pedestrian features, resulting in a decrease in accuracy.

In addition, the experimental results also showed that when the dataset contains the sample from the same angle, the model shows higher recognition accuracy. This is because the sample provides more consistent functional information from the same perspective, making the model easier to match and compare. When the dataset does not contain samples from the same perspective, the overall recognition accuracy of the model slightly decreases, which may be due to a lack of sufficient multi-perspective information. The model needs more training and optimization to improve its recognition ability when processing different perspectives.

Therefore, for gait recognition from different perspectives, it is necessary to improve the model further to enhance its recognition performance for side and back perspectives or introduce more multi-perspective information to improve overall recognition accuracy. The recognition rate of each perspective exceeds 90%, while the average recognition rate reaches 96.9%.

Table 2. Experimental results

Gallery NM#1-4 Query	0	18	36	54	180
NM#5-6	98.2	97.5	99.5	94.5	97.1

BG#1-2	98.4	95.7	91.2	94.7	92.9
CL#1-2	98.8	94.4	93.5	91.5	90.9

5. Conclusion

In short, the gait recognition algorithm proposed in this paper combines LSTM (short - and long-term memory network) and DBN (deep confidence network) networks, and is tested and verified on the Casia-B dataset. Casia-B data set is an important data set widely used in gait recognition research. It contains gait sequences taken from multiple different angles, and has become one of the standard data sets to measure the performance of gait recognition algorithms. Through the experimental analysis of this dataset, the experimental results in this paper show that the LSTM-DBN model can achieve high recognition accuracy in a positive perspective, especially in a slightly biased perspective. This shows that the model has strong ability and high accuracy in capturing and classifying gait features from a positive perspective. However, the recognition accuracy of the model is relatively low in the side and back view. The reason for this phenomenon may be related to the expression of human contour information in gait images. In the front view, the shape and gait characteristics of the human body are more obvious, which is easy to be recognized and extracted by the model; However, in the side and back view, the contour information of the human body may be occluded or lost, which makes the model unable to effectively extract key gait features. Therefore, how to improve the performance of the model under these perspectives, especially enhance the ability to extract human contour information, is still an important direction in future gait recognition research.

Reference

[1] Yeoh T W, Daolio F, Aguirre H E, et al. On the effectiveness of feature selection methods for gait classification under different covariate factors[J]. Applied Soft Computing, 2017, 61: 42-57.

[2] Yan C, Gong B, Wei Y , et al. Deep multi-view enhancement hashing for image retrieval[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 43(4): 1445-1451.

[3] Zhang E, Zhao Y , Xiong W. Active energy image plus 2DLPP for gait recognition[J]. Signal Processing, 2010, 90(7): 2295-2302.

[4] Wang C, Zhang J, Wang L, et al. Human identification using temporal information preserving gait template[J]. IEEE transactions on pattern analysis and machine intelligence, 2011, 34(11): 2164-2176.

[5] Yao H, Zhang S, Hong R, et al. Deep representation learning with part loss for person re-identification[J]. IEEE Transactions on Image Processing, 2019, 28(6): 2860-2871.

[6] Takemura N, Makihara Y , Muramatsu D, et al. On input/output architectures for convolutional neural network-based cross-view gait recognition[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2017, 29(9): 2708-2719.

[7] Anwary A R, Yu H, Callaway A, et al. Validity and consistency of concurrent extraction of gait features using inertial measurement units and motion capture system[J]. IEEE Sensors Journal, 2020,21(2): 1625-1634.

[8] Tong S, Fu Y , Ling H. Cross-view gait recognition based on a restrictive triplet network[J]. Pattern Recognition Letters, 2019, 125: 212-219.

[9] Khan A, Khan M A, Javed M Y , et al. Human Gait Recognition Using Deep Learning and Improved Ant Colony Optimization[J]. Computers, Materials & Continua, 2022, 70(2): 452-459

[10] Wang X, Yan W Q. Human gait recognition based on frame-by-frame gait energy images and convolutional long short-term memory[J]. International journal of neural systems, 2020, 30(01):1950027.

[11] Choi S, Kim J, Kim W, et al. Skeleton-based gait recognition via robust frame-level matching[J].IEEE Transactions on Information Forensics and Security, 2019, 14(10): 2577-2592.