Neurocomputing 505 (2022) 315–338

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Survey paper

Intra-class variations with deep learning-based gait analysis: A comprehensive survey of covariates and methods

Anubha Parashar [a,*], Rajveer Singh Shekhawat [b], Weiping Ding [c,*], Imad Rida [d]

[a] School of Computer Science and Engineering, Manipal University Jaipur, Jaipur, Rajasthan, India
[b] School of Computer Science and Engineering, Manipal University Jaipur, Rajasthan, India
[c] School of Information Science and Technology, Nantong University, China
[d] BMBI Laboratory, University of Technology of Compiègne, 60200 Compiègne, France

ARTICLE INFO

Article history:
Received 20 November 2021
Revised 21 June 2022
Accepted 10 July 2022
Available online 16 July 2022
Communicated by Zidong Wang

Keywords:
Gait recognition
Biometrics
Covariates
Deep learning
Computer vision
Video surveillance

ABSTRACT

Gait recognition is an essential biometric technique that recognizes humans at a distance through their unique walking style. In the present era of deep learning, automated gait analysis has become easier with an increase in processing power. However, the recognition accuracy is affected by many covariates such as clothing conditions, carrying objects, varying viewing angles, occlusion, walking speed variations, and thus, it remains a challenging problem. For this complex problem, huge datasets are required to train for given conditions and predict new situations; thus, deep learning is preferred. In this review paper, we categorize various gait covariates which have been recently handled. There are various approaches, but the most effective approach is deep learning; hence in this paper, we include the most used deep learning approaches for each covariate condition found in the literature. Further, we highlight open problem areas handling these covariates and offer some suggestions about their better handling. Based on the review and our understanding of all the gait pipelines employed in deep learning, we have suggested a comprehensive and universal deep learning pipeline that can handle most gait covariates rather than customized deep learning pipelines. The methods of handling gait covariates are summarized according to appearance, pose, and sensors. A comprehensive comparison of reviewed approaches for real-time scenarios in terms of their novelty, benefits, and limitations is then offered, which led us to identify open research problems related to gait covariates. In the end, the paper concludes with the challenges identified and prospects.

1. Introduction

Biometrics is the automated identification of individuals depending on their physiological (face, fingerprint, iris, retina) and behavioral (sign, gait, postural stability) characteristics [227]. It is the most reliable method for the establishment of identity as it offers a high degree of accuracy and robustness. Physical features such as fingerprints, iris, hair color, hand geometry, and behavioral ones such as voice and accent, gait, grid, or striking keys on keyboards, can uniquely distinguish a person from others. Out of these, the human gait draws the attention of most researchers due to its objective and passive nature of observation and widely available surveillance camera networks. Surveillance applications include airport security checks which can be useful for criminal identification, customer authentication in banks, and detection of intrusion in sensitive areas [230]. Gait recognition has significant advantages over other biometrics due to its passive observation method and the uniqueness of the gait. It can even work well with low-quality of gait videos. It is also impossible to copy or masquerade the gait of another person. Criminals typically wear masks, dark sunglasses, and gloves to defy identification, but imitating gait is very hard.

In this review paper, the human gait we focus on can be defined as a synchronized combination of cyclic movements leading to an individual's unique locomotion. One key concept of gait recognition is the gait cycle, defined as the time between initial contact and terminal swing of human locomotion. Fig. 1 consists of a series of images of human locomotion, depicting one gait cycle. As the natural walking pattern of a person may be intermittent, it is enough to accept just one phase known as the entire gait cycle of human locomotion. So one needs to process further by collecting

* Corresponding authors.
E-mail addresses: anubhaparashar1025@gmail.com (A. Parashar), dwp9988@163.com (W. Ding).

https://doi.org/10.1016/j.neucom.2022.07.002
0925-2312/

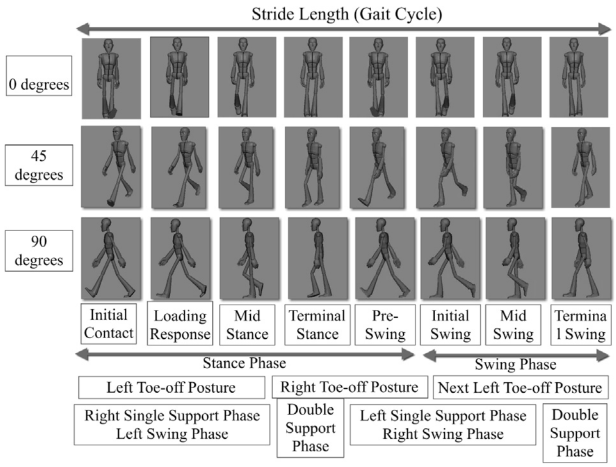

Fig. 1. Gait cycle captured at three different angles.

all such gait cycles under varied conditions to identify the individual uniquely.

The gait recognition methods still have limitations and thus are still evolving, especially under covariate conditions. A covariate is an external mechanism intended to alter the natural gait of an individual and constrain how a person moves. This review addresses the issue of gait recognition under the presence of covariates. These covariates are normal in everyday life and can greatly affect gait recognition efficiency. For example, numerous covariates attributed to the subject itself, such as clothing, carry situations, or variations in view angle, speed variation, walking surface, occlusions, shadow, or edge detection flaws, have been shown to influence gait recognition accuracy drastically. Fig. 2 depicts detailed measures of gait parameters. Major intraclass variations in these parameters lead to covariate conditions.

It is difficult to establish an effective technique for covariates for gait identification invariant to viewing angle adjustments and different circumstances of wear or holding. The Gait generation process can also be affected by internal covariates caused by illness, pregnancy, or old age. Few pieces of research have shown gait recognition methods in the presence of covariates by analyzing the external covariates. Some factors usually correlated with the impact of covariates have been reviewed and provided their effects. Deep learning techniques have recently attracted enormous attention from the computer vision community. The evolution of machine learning technologies has resulted in deep learning models that can handle complex data with minimal preprocessing and greater accuracy for large datasets. Apart from the covariate aspect, we also tried to delineate the deep learning architectures successfully used for handling covariates as deep learning techniques are very effective.

1.1. Motivation

Very few review papers on gait biometrics highlight covariates using the deep learning approaches. Most survey papers before us focused on camera-based gait recognition systems instead of pose and sensor-based approaches. Not many papers covered appearance, pose, and sensor-based approaches together. Earlier reviews have overlooked several essential gait datasets with covariates [225226]. For this reason, we began identifying and organizing papers on this theme for the most recent deep learning approaches used in gait recognition to handle covariates. Some studies on gait recognition like [216 217218219220221], helped us to understand the work done by earlier authors. However, many significant issues were not addressed in these studies. These studies did not give importance to covariates conditions, especially in covariates such as speed variations, view, and spatial–temporal aspects, which significantly affect the outcomes of gait recognition.

1.2. Contributions

This paper provides a comprehensive review of the current literature on vision, pose, and sensor-based gait recognition methods and explores the potential areas for researchers. The major contributions of this work are as below:

1. This review paper proposes ideas to overcome the covariate conditions.
2. Timeline of 150 papers with covariate categories is summarized, and for each reviewed paper, we listed the dataset used and the accuracy achieved.
3. Interpretations of covariates papers published in the past five years with a ratio of most adopted datasets and gait recognition pipeline.
4. It highlights the most popular techniques for gait recognition using deep learning detailing the problem statement and novelty of work.
5. This paper discusses possible adversarial attacks on gait recognition datasets and anonymization methods.
6. The work includes a discussion of various research questions and gaps in their proposed solution.
7. Deliberating views on various challenges faced due to covariates, deep learning architectures, suitable datasets related to real-time analytics, application-specific approach, protection, privacy, and anonymity.

1.3. Organization

After a quick introduction and overview of gait recognition aspects, the remainder of the sections are drawn up as follows; In Section 4, approaches to gait recognition are discussed along with the methodology adopted and deep learning pipeline. In Section 5, a review of covariates using deep learning-based gait biometrics approaches (model-free and model-based). In this section, covariate factors that affect gait recognition are extensively investigated, starting with a problem statement, novel adopted gait recognition techniques using deep learning, its benefits, and limitations. This section also provides a comprehensive treatise on the datasets used for gait recognition. In Section 4, we provide research gaps identified by us while reviewing. For every gap, we proposed solutions for gait covariates along with the deep learning pipeline proposed by us. This section also discusses novel outcomes with the analysis of the most adopted techniques used in the literature. In Section 5, the challenges and prospects of gait recognition and its allied aspects are summarized. The conclusion

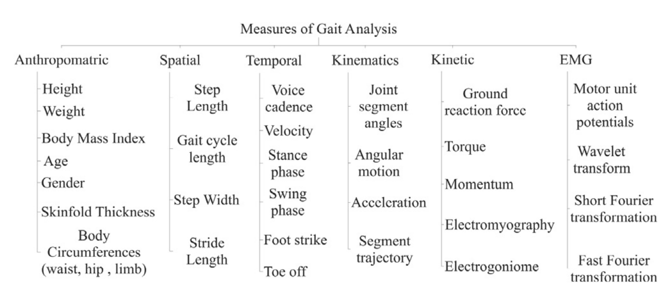

Fig. 2. Detailed measures of gait parameters..

of this paper is in Section 6. Appendix A contains a list of acronyms that are used in this paper.

2. Deep Learning techniques for gait recognition

Several gait detection strategies have been suggested in the last decade. These techniques are classified into two groups: model-free and model-based techniques, as depicted in Fig. 3. Model-free methods concentrate on the silhouette of the entire human body [228]. These approaches are not computationally expensive and are effective with low-resolution images [229]. On the other hand, model-based methods concentrate mostly on obtaining the stride parameters of the subject that define the gait by utilizing the motion of joints of the human body. The approaches that focus on the model-based technique appear to be more challenging. As first joints need to be located precisely, they are costly in computing and require high resolution images.

2.1. Methodology adopted

This review article identifies model-free and model-based approaches to gait recognition using deep learning. This paper classifies gait recognition approaches into three categories: vision, pose, and sensor-based. The current study explores the covariate conditions that influence the impact of gait recognition after studies on the state-of-art. Fig. 4 describes a systematic evaluation procedure of covariate-based gait recognition techniques.

Deep learning models can build multiple layers of feature hierarchies by finding a wide range of features from low-level feature sets. Deep learning models' layers are generic and automated with customizable parameters. Several recent studies using deep learning algorithms in gait recognition have demonstrated promising outcomes. Fig. 5 depicts the prominent deep learning methods for gait recognition used by the literature in this article. One of the popular deep learning architectures is CNN. CNN is good at extracting features from images, but it takes a considerable time to get trained, as a large dataset is required and takes colossal storage space. The deep hierarchical structure of CNN typically involves billions of parameters, which requires more time and computing resources to execute the task. It is essential to reduce the time a deep CNN model takes to process large datasets in gait recognition systems for real-time security and surveillance applications.

Rauf et al. [1] proposed a new architecture to speed up the process and decrease models based on CNN and later test to recognize gait. They proposed a fully connected network, which is smaller in size and has better speed than regular CNN. But their proposed technique has a high computational cost as the algorithm's complexity is more time-consuming. McLaughlin et al. [2], offered a CNN network with recurrent networks in the last layer, which captures spatial and temporal features using a convolutional neural network. Their method efficiently and accurately extracts the frame from features of a gait sequence, but the computational cost of their proposed model is very high. Sokolova et al. [3] proposed a deep learning method that takes optical flow images as input. Their proposed approach is good at extracting spatial information, and they used the complete body silhouette instead of making joints of different body parts separately. However, the method proposed by the authors fails to capture temporal information; therefore, there is a loss in temporal features. In [4], they proposed a multi-pipeline convolution neural network that takes optical flow images as input. Their CNN model captures spatial features efficiently. The disadvantage of using this model is that their model did not capture temporal features. The work in [5], mentions that gait recognition with low-resolution cameras is challenging. They proposed a deep CNN network that extracts features, takes GEI as input, and is trained by a neural network. Their CNN model captured spatial features efficiently, but their method did not capture temporal characteristics. Carley et al. [6], proposed an auto-correlation method that learns features from various viewing angles. Their proposed architecture is optimal for extracting pose features and is very efficient in computation. However, the accuracy of their method is very low.

For recognizing gait, various processing steps need to be incorporated like segmentation of subject, feature learning and feature extraction, similarity check between subjects, and classification. Chunfeng et al. [7] proposed a CNN-based GaitNet, which performs segmentation of subjects, feature learning, feature extraction, similarity check between subjects, and at last, classification. Their method can learn distinct information from raw gait images. Their model is not good at capturing the temporal features and requires a fusion model to improve the accuracy. Habiba et al. [8] proposed a fusion-based pre-trained DNN (VGG19 and AlexNet) and fuzzy entropy controlled skewness network. Their model captures spatial features efficiently. The fusion of two proposed networks increases accuracy, but their fusion model could not extract temporal features.

Zhaopeng et al. [97] proposed a network-based capsule network to recognize gait. Their network captures spatial features (like captured by CNN) and preserves the changes in features like clothing, view variation, and multi-walking situations. The capsule networks are deemed to be better and more efficient than simple CNN because capsule networks can hold all types of information (problems of overfitting) that get generated at the time of pooling.

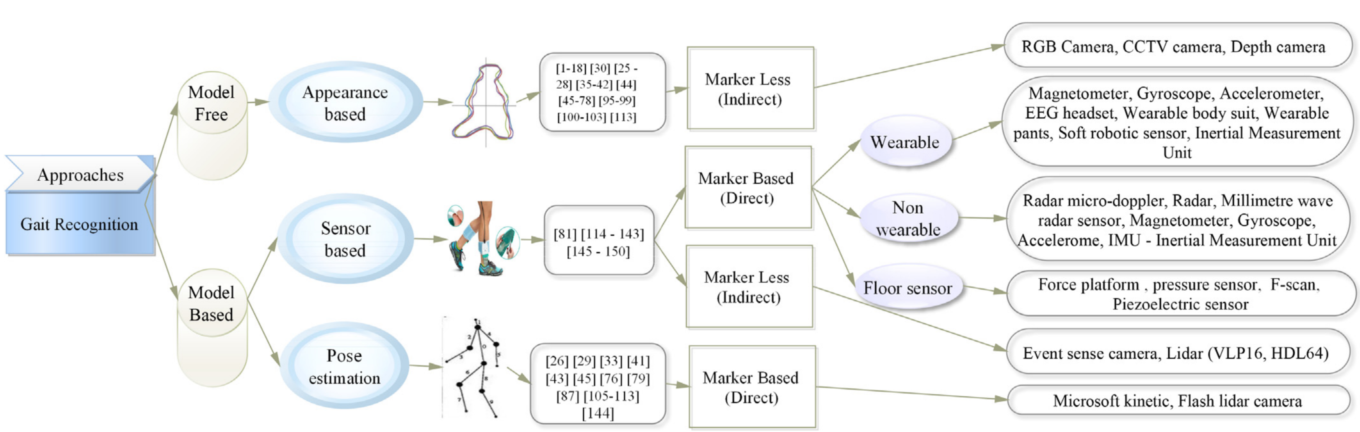

Fig. 3. Approaches of gait recognition..

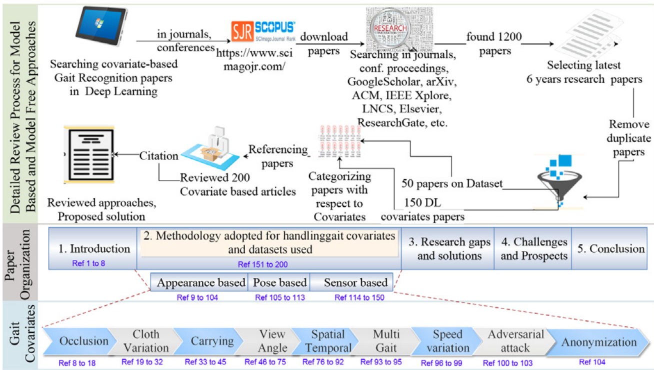

Fig. 4. A detailed review process of cited papers..

The Capsule network is not giving good accuracy when GEI is given as input. Makihara et al. [98] proposed a spatial and temporal deformable CNN. Their proposed method performs efficiently and has good generalization capability on various datasets. Their method is less effective on smaller datasets, real-time data, and with different covariate conditions. However, they did not test their algorithm on different covariate conditions. Sun et al. [24] authors proposed a faster RCNN to extract features from the subject's clothing and locate people in the silhouette images. The proposed architecture is scalable and effective on any covariates changes to increase accuracy. Their proposed method is complex and consumes more time as the regions generated outside the CNN use a selective search algorithm.

3. Covariate classification and suitable approaches

Current gait recognition methods work well in the absence of covariates, but the classification rate decreases in the presence of covariate conditions. Fig. 6 depicts various covariate conditions covered in this paper.

3.1. Model free approaches

Model-free approaches are also known as appearance-based approaches [1–28] [30–32] [34–42] [44–78] [80] [82–99] [100–104]. In this method, the depiction of features attempts to process the entire shape or motion of the silhouettes for an individual. There are two major pros of model-free function depiction. First, model-free approaches can give results even when camera resolution is insufficient, so it is possible to deploy surveillance cameras far from the subject. Second, its cost of computing is lower than the representation of model-based features. Thus, these methods are more common. The big disadvantage of model-free feature representation is that it is dependent on various covariates conditions. Covariate conditions are given in Fig. 7, which significantly impact the gait of a person. At starting, the gait recognition rate will decrease, but after setting the correct parameters of deep learning algorithms, the recognition rate starts increasing. Model-free techniques consider the entire silhouette of a subject to obtain the gait features for recognition purposes and are chosen in most functional implementations.

3.1.1. Occlusion

The methods for determining gait against occlusion are divided into two categories. First is the reconstruction-free approaches [11], which concentrate on extracting the features from a gait silhouette or an average of them, such as the gait energy image (GEI). The features derived from these methods achieve better performance than simply passing the silhouettes for training. In case of a high degree of occlusion, reconstruction-free approaches cannot be used where the calculation of gait cycles is challenging. Second approach is the reconstruction-based methods [9–10] [12–18]. Approaches in this category include the restoration of occluded people, operated on several gait periods in which certain frames are partly occluded. These methods are challenging to implement when all frames are heavily occluded in a series. The main drawback of this approach is that restored silhouette sequences often deteriorate the discriminating capacity of a person after reconstruction. Therefore, gait recognition output is adversely affected. The timeline of papers reviewed for occlusion is shown in Fig. 8, with the dataset used and the accuracy obtained.

A complete gait cycle is often unavailable due to occlusion, which makes gait recognition difficult. M. Babaee et al. [9] proposed a method that converts an incomplete gait cycle by creating an incomplete GEI and later reconstructed a full GEI using deep autoencoders. There is an improvement in the result by reconstructing the frames and regenerating incomplete GEI, but they did not consider clothing and other covariates. The authors in [10], reconstructed complete GEIs from a few frames of the gait cycle. For this, a DNN method was proposed to make a full gait cycle. According to their study, reconstruction of silhouettes does not work if the partial gait cycle is below 20%. In [11], the authors proposed a partial gait cycle using FCNN to recognize subjects upon the availability of an incomplete frame. As they are taking partial gait cycles and, in many instances, even making GEI with a single frame drastically reduces the accuracy. Li-min et al. [12], proposed a VGG-16, LSTM, and GAN to identify occlusion and reconstruct a complete gait cycle by predicting the frames. They used a deep learning method that reconstructs the missing frames and leads to higher identification accuracy. Their algorithm does not collect temporal features. Maryam et al. [13], proposed a non-standard periodic GEI and constructed it with no advanced knowledge of the gait cycle. Therefore gait cycles are no longer needed for making GEI. This system is more efficient and more straightforward as preprocessing is not required.

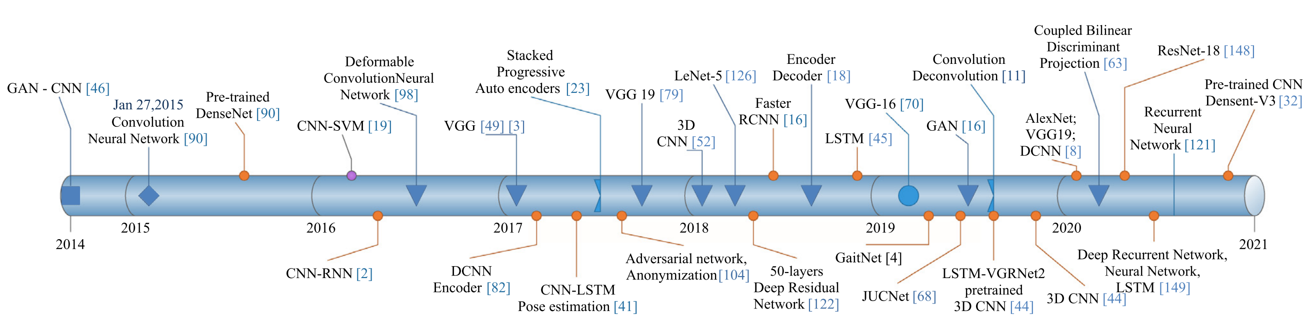

Fig. 5. Deep learning techniques used in gait recognition..

Fig. 6. Overview of covariate conditions.

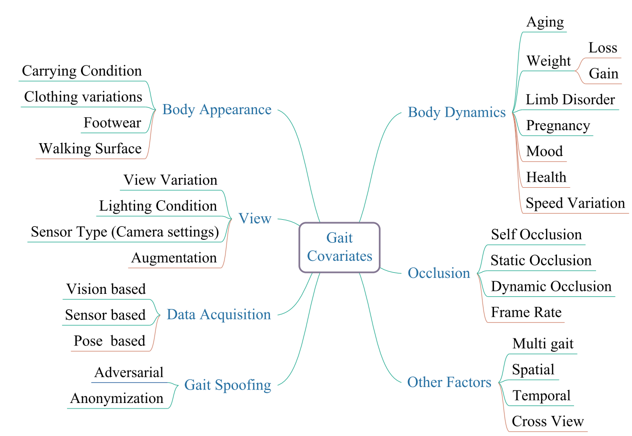

Fig. 7. Influencial covariate conditions affecting gait recognition rate..

The Gait recognition rate degrades due to incomplete GEI. Li et al. [14] used GAN restoration of the GEI method. There is an improvement in the result when the generator is trained with the fusion of mean square error and adversarial loss, but the computation cost of their algorithm is high. Recognition is highly affected by the small area of silhouette due to occlusion. Dhritimaan et al. [15] proposed a GAN-based method that generates occluded silhouettes. Wasserstein GAN is used in constructing sparse features to improve the ability of gait-based classification. The authors assumed they knew the occluded location beforehand but did not explain it. Also, they did not compare the different percentages of the partial gait cycle with respect to the reconstruction error.

GEI features are affected when variations like carrying conditions occur in existing gait recognition methods. In [16], the authors proposed a GAN-based method to make carried objects disappear in order to avoid occlusion caused by these objects. Their method is robust against carrying variations. A large dataset is required for training a 3D gait recognition method in order to train the system. However, their network is simple (regression-based - GEINet) and accurate for smaller data. It is unsuitable for real-time applications requiring a large dataset to train the system.

Occlusion and view angle adversely affect gait recognition accuracy, and GEI loses the temporal information. Uddin et al. [17] proposed a multitask GAN for feature reconstruction. A novel period energy image method is given to conserving temporal features. They combined GANs with multitask learning to achieve competitive results. This process is highly beneficial for mitigating errors due to the variation of view angles in gait. It is not possible to generate a period energy image by attaining stable training because of the high dimensionality of GANs. Body parts of the subject often get occluded due to pillars, other people walking, trees, vehicles, or beams. Wang et al. [18] proposed a deep GAN to reconstruct silhouette sequence from occlusion. Their method can evaluate gait recognition with no advanced knowledge of the gait cycle as it does not take occlusion position. They did not use occlusion with multiple view variations.

3.1.2. Clothing variation and carrying condition

Clothes of a person can vary frequently and create a significant obstacle to the gait identification process depending on vision and acoustics. Clothes do not significantly affect the motion unless they are bulky or particularly tight and restrict the natural walking motion. Similar is the case with footwear which affects the walking pattern if they are high heels. Recognition rates will decline as participants walk in heavy clothes, possibly due to a difference in appearance rather than in gait if appearance-based approaches are used. Heavy clothing may alter the speed of a person. The timeline of papers reviewed for clothing variation is shown in Fig. 9, with the dataset used and the accuracy obtained.

Yeoh et al. [19] used a convolution neural network to reveal discriminating features of gait from GEIs. The authors employed CNN to create a complete deep learning model based on gait identification which is invariant to clothing variation, but their proposed

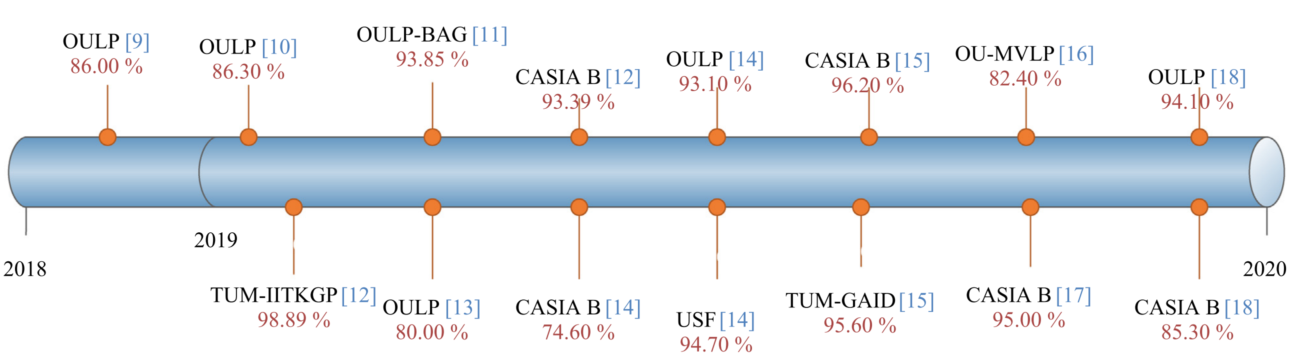

Fig. 8. Timeline of occlusion papers with dataset and accuracy details..

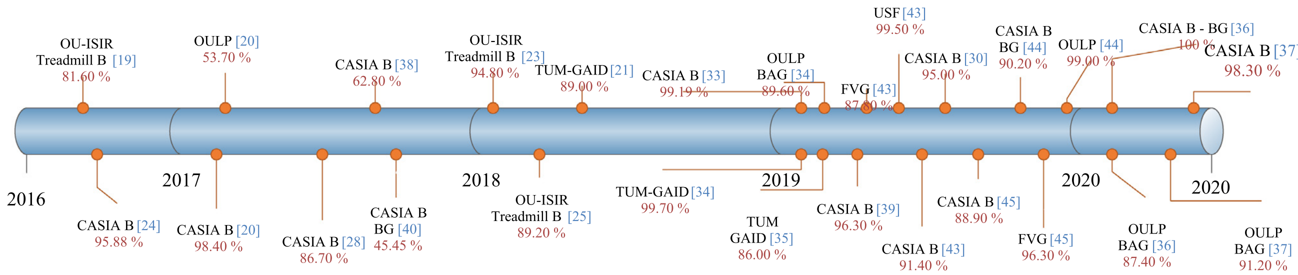

Fig. 9. Timeline of clothing and carrying variations with dataset and accuracy details.

algorithm accuracy is low. Recognition accuracy depends on the static features of intraclass variations, and changes in these features might adversely affect gait recognition accuracy. Therefore, a reliable deep learning-based model which effectively selects GEI-based features is proposed in [20]. The augmentation technique is used to overcome intraclass gait-based changes in features and is efficient on small cloth-based datasets for training. However, the accuracy of their proposed method declines on larger datasets such as OU-ISIR. Castro et al. [21] proposed a deep learning pipeline consisting of CNN and an optical flow as a feature extractor for making the model learn high-level descriptor from a low-level descriptor. Even with low-resolution, depth images outperform the other methods. The accuracy of the network was improved greatly when the optical flow method was used instead of an input based on silhouette. Their proposed method is not good in extracting temporal features because CNN is not efficient in capturing temporal features. CNN relies on datasets to represent output via local hierarchical features, which increases semantic complexity. CNN applies brute force computation, and GPU is the best hardware for acceleration. Critical parameter manipulation of CNNs is important to improve the results. Yeoh et al. [23] assessed the network's four fronts: AlexNet, two-dimensional Convolutional Neural Network, ResNetlm, and CaffeNet. The parameters were selected by keeping the major research interests in mind, so they perfectly tuned the deep learning model and provided optimal results. However, they did not consider different view angles.

The clothing and hairstyle of the subjects play a major role in their gait-based identification, and a gait recognition algorithm should be invariant to such changes. To obtain discriminative and robust features, Hefei et al. [25] proposed a technique by combining two approaches, namely latent-semantic-based analysis and attention-mechanism for clothing-invariant gait recognition. Then they fused the two approaches for higher-level representation, which improved the performance. The authors used a CNN-based method that learns high-level features from raw input data and highlights the important regions of subjects. Their attention mechanisms emphasize local information, and latent semantic variables play an essential role. Although GEI is the most popular representation for gait, it loses spatial and sequential information. Yan et al. [27] used ConvNets for learning rich features from the training set. They utilized the advantage of GEI-based descriptors, computed on the gait cycle; then, features are collected using a CNN and later trained on a multitask learning model. Their approach is computationally complex and time-consuming. Their GEIs are not apt whenever there is a shapeshift due to changes in view, occlusion, or carrying conditions. Alotaibi et al. [28] developed a specialized DCNN architecture. The proposed method is effectively calculating gait recognition and is invariant towards clothing covariate. Few training parameters are required for training the network effectively. Tanh was used as an alternative activation function to ReLU. DCNN failed to attain high accuracy in the proposed system, and the authors used less training data. Yao et al. Hawas et al. [30] deployed CNN with optical flow images to build unique features. Features are measured automatically using CNN layers. The authors, when trained on 90 degrees and tested on 180 degrees, the system gave an accuracy of 6.32%, which is very low.

Human gait is hampered by carrying conditions, and the recognition rate may decrease as people hold objects. The weight itself will modify the stride biomechanics and cause the multiple arm joints to assume different roles than normal gait by altering the gait shape, which will also change the stance. The weight may also add mass to the subject and will affect the gait acceleration. The timeline of papers reviewed for carrying conditions is shown in Fig. 9, with the dataset used and the accuracy obtained.

Carrying conditions imparts a vital part in determining the accuracy of gait recognition. Yang et al. [33] proposed coordinate, angle, position, distance-based skeleton, and GEI for eliminating interferences in carrying. Their method increased gait recognition accuracy and performed better in recognizing subjects with changes in carrying conditions. Their proposed model is time-consuming and uses less data to train, so their model suffers underfitting. In [34], the authors proposed a technique based on recognizing gait using disentangle-representation learning that considers both are carrying covariates and identity features. They efficiently designed an autoencoder from GEIs and considered the covariates to reconstruct the loss part which occurred due to it, but they are not using a good classifier. Castro et al. [35] provided a pragmatic approach using state-of-the-art CNN and an applied dataset for carrying conditions. Their model did not perform well on smaller datasets. Li et al. [36] proposed a technique based on recognizing gait using disentangle-representation learning that considers both covariates and identity features. The authors efficiently designed an autoencoder from GEIs and considered the covariates to reconstruct the loss part, but the classifier is not up to the mark in the model used.

Gait recognition is affected when there is an unpredictability of carrying location, such as the presence of multiple locations, side, back, or front. Li et al. [37] used a generator for gait estimation and generated the templates devoid of a carried object. Then, they passed the generated images into a discriminator for recognition. Unlike GANs, which generate and estimate the template devoid of a carried object, the authors have used an AB-GAN, which selectively removes the unnecessary carrying part from an image leaving other parts unchanged. Suppose the subject is in an overcoat or raincoat covering the entire body. It is not preserving identity as the entire gait template must be regenerated to form a proper GEI making the system perform like an original GAN. Also, they did not consider all the view angles and clothing types. Yu et al. [38] proposed a GaitGAN for clothing, carrying, and viewing invariant gait data. A significant benefit of this approach is that the view angle is not required beforehand. Also, identification information is preserved. Their approach is computationally complex, and the accuracy falls when the difference between viewing angles of training and test data is large.

Shiqi et al. [39] proposed a GAN model to regenerate a canonical side view of a subject with clothes (normal clothing) and devoid of any carriable objects. They opted for a multi-loss technique for optimization, which increased their inter-class variation, thus, decreasing intra-class variation. Also, the viewing angle was not predetermined for producing a gait image. Gait features are adversely affected when the view angle changes, and this is because the local features suffer and change. Removal of carrying parts is done effortlessly, but unwarranted errors occur in the GEI formation when they try to regenerate the body parts that did not carry objects. In [40], the authors proposed an autoencoder (model-based) deep architecture. In the proposed technique, the influence of bags and coats (carrying and clothing) is very less as the view angle of GEI is converted to 90 degrees. In clothing, especially coats, the results are very low. The performance of their method is limited and cannot handle clothing, view, and carrying variance. The accuracy is not good since they are not capturing temporal information through data GEI. Liao et al.

Zhang et al. [43] proposed LSTM autoencoder-based gait network, which unequivocally disentangles the features of appearance and poses from RGB images. Such approaches are useful when covariates like clothing conditions are invariant. One instance of such a scenario is using motion dynamics in vision-related problems, but they did not work on all view angles. Daksh et al. [44] proposed a LSTM-VGRNet2 pre-trained three-dimensional CNN model. Its performance is better than the existing methods. Due to obtaining fewer temporal features, their method is less accurate. Zhang et al. [45] proposed LSTM and autoencoder network for extracting features. Autoencoder networks use RGB images for input, feeding it to a canonical network based on the pose feature extraction technique. They considered appearance-based features for clothing and carrying conditions. However, their results of the carrying condition are not good.

3.1.3. View angle variation

Reorientation of camera position and human subject is a typical problem for vision modalities. The alignment and position of the gallery (training samples) and probe (testing samples) video data from the camera are different for an individual, which lowers the performance of the gait recognition algorithm. The timeline of papers reviewed for view angle variation is shown in Figs. 10 and 11 with the dataset used and the accuracy obtained. Inter-class differences must be as large as possible, whereas the intraclass difference must be as small as possible. Tong et al. [22] proposed a DNN coupled network to increase the accuracy by minimizing the intraclass differences. They used the CNN model and considered cross-view covariate conditions but did not test their proposed algorithm for all viewing angles. The accuracy of their method is 68.17% at 36 degrees, which is low.

Cheheb et al. [42] proposed a GEI-based sparse auto-encoder for extracting the features of various views. They made an efficient deployment of auto-encoders for gait-based recognition. Their proposed approach recorded results with low accuracy and did not discuss how they have enhanced the model for clothing and carrying variation. In [46], authors propose that the view angle of the subject can vary widely and arbitrarily, which adversely influences gait recognition accuracy. They proposed a GAN and CNN-based network. GAN is used for view transformation, and CNN is used for recognition. Their model is robust towards the view angle and subject variation; however, they did not consider all the viewing angles. Also, they did not show the reconstruction using GAN. Shiraga et al. [47] proposed a gait recognition method using CNN to automate the feature extraction. They proposed a network-GEINet for recognizing gait by feeding GEI (input) to CNN. Their technique is effective when the view angle changes between the probe and gallery are small. They endured translation issues because of the black box qualities of CNN. Classical machine learning methods do not use gait recognition in a multi-view environment without capturing spatial and temporal features. In [48], the authors proposed a method using 3D DCNN for multi-view based on gait recognition. This method gives enhanced performance. It is an efficient feature extractor, but this approach cannot work with small datasets. However, their approach is highly complex and high computational cost. Li et al. [49] proposed a method that uses a "pretrained VGG-D" (DCNN) network with no fine-tuning of the parameters. Even though the method runs on a very low dimension, it gives good results. They did not consider cross-variation, cross-view, carrying conditions, and clothing variation.

Gait recognition system performance is affected by view angle. Appearance-based techniques widely depend on the feature extraction parts of GEIs. Primarily these techniques suffer from noise and other covariates. To increase recognition performance, Chao et al. [50] used deep learning for feature extraction. They proposed a pre-trained VGG-D joint bayesian-based model for view variation and improved recognition performances in view and cross-view variations. The proposed method did not discuss other variances (carrying of bags, wider view variations, and clothing). When the training data is in a particular view angle and testing data is in another angle, there is a drop in the performance due to a large viewpoint angle. Jia et al. [51] combined CNN and feature optimizers for achieving view-invariant gait recognition. In moderate variations of view angle, CNN gains discriminative features. The proposed method did not discuss other variances (occlusion, clothing, carrying bags). Thapar et al. [52]proposed a 3-dimensional DCNN for recognizing a human in multi-view conditions. Their model extracts spatial information. Their method is time-consuming and requires data labeling, which is not possible instantly, and incomplete/ unlabeled data shows poor performance. Zhang et al. [53] proposed a normalizing VN-GAN-based network to learn and generate the best features. Their method achieved great performance as their GAN-based model generated

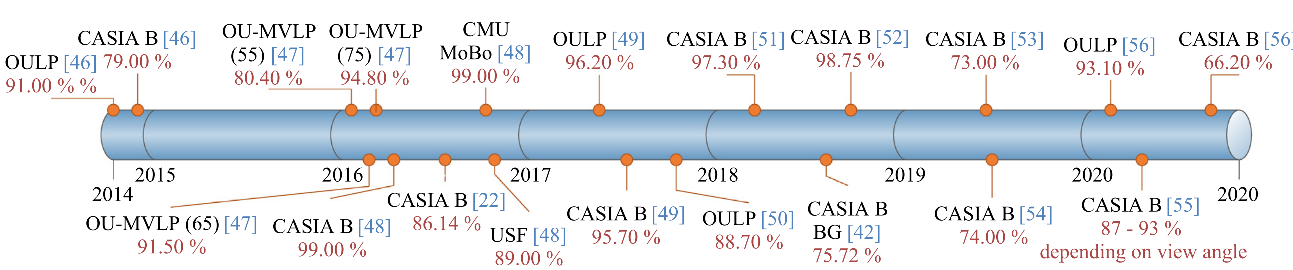

Fig. 10. Timeline of view angle variations with dataset and accuracy details.

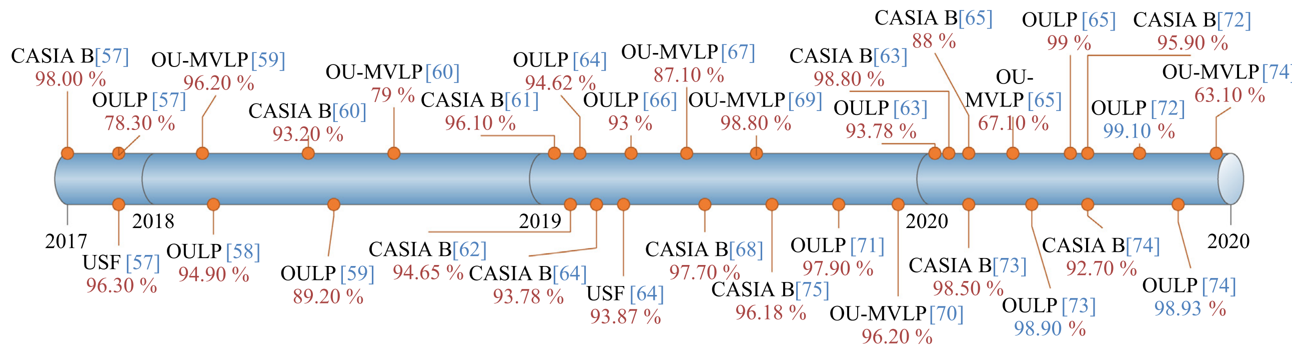

Fig. 11. Timeline of view angle variations with dataset and accuracy details.

real-world results, making it easy to recognize the subjects. Their method does not consider carrying and clothing (coat) conditions.

Zhang et al. [54] proposed a view transformation GAN with the help of a uniform model. Their model works on every view angle. Their model produced GEIs, and their proposed GAN visualized these GEIs. They achieved high performance because their model was view-invariant. They did not consider carrying and clothing conditions. Through the proposed VTGAN, covariates affect identification accuracy. Muqing et al. [55] proposed RNN based network that does not have an adverse effect on the view angle of silhouettes and makes the view-invariant gait recognition system. Their technique can accomplish robust results under various viewpoints, reasonable for real-time gait identification. However, they did not consider other covariates like clothing and carrying conditions. Wang et al. [56] proposed a GAN-based method to learn the global and local features. They did not use a pixel-wise loss function in their proposed method, which effectively found the local and global features. They provided an efficient view transform method using GAN; however, cross-view recognition accuracy is not satisfactory. There is a scope for increasing the performance ability of big datasets.

Similarity learning is important when comparing inter and intraclass differences to recognize gait more robustly. Zifeng et al. [57] proposed a DCNN network to evaluate the inter and intraclass differences. The average recognition rate is high when the cross-view angle is large (>36 degrees). Their result shows less accuracy in a multi-view environment because their algorithm cannot properly implement inter and intraclass differences. In [58], the authors propose a CNN-based siamese network, which takes sequence images of arbitrary length as input through the siamese network. Their proposed method combines the advantage of CNN and the siamese network, but they did not consider all the viewing angles, clothing, or carrying conditions. CNN extracts the features arbitrarily from various silhouette parts in [59]. Still, CNN is not spatially heterogeneous, which results in the loss of discriminatory data as various body parts show variation in movement and shape. CNN extracts features from different human parts arbitrarily without considering spatial heterogeneousness. This causes the loss of discriminatory information as different human parts vary in shape and movement constraints. They proposed a network of attention-based embedded gait pipelines and have addressed the problem of spatial heterogeneity. Clothing and carrying conditions largely change the appearance of a subject, but they did not address the problem of carrying bags and clothing variations.

Large variations in the view angle cause low gait recognition and affect the appearance of a subject. CNN efficiently captures view-invariant angles on a smaller dataset but not on larger ones. BingZhang et al. [60] proposed a discriminant GAN framework for reconstructing the gait silhouettes from all the viewing angles. Their proposed method efficiently learns view-invariant features that can be used in cross-view variations, and their method can scale well to larger datasets. Thus, making it practical for large-scale surveillance applications. However, the accuracy is low on the OU-MVLP dataset. Cross-view, horizontally partitioning is an effective method, but current literature is not learning features based on part level. Luo et al. [61] proposed a sequential multiscale architecture for extracting discriminatory gait features for cross view conditions. Segmentation has a great impact on data. If data quality is low and occlusion exists, then recognition performance decreases. So, their proposed method reduces the effect of segmentation on low resolution and occluded data. They tried to capture temporal data (with attentive temporal pooling), providing diverse weights and silhouettes variations. This approach is robust in real-time application but does not consider clothing or carrying conditions. In [62], the authors proposed color-mapped contour-based images of gait for various cross-view factors. The proposed method protects temporal and spatial differences in regular walking patterns of human gait. However, they did not consider the occlusion covariate.

If the walking direction of gait is different from that of the gait registered, a change in the view angle alters the gait model-free features. Thus, recognition suffers from view variations. Ben et al. [63] implemented the projection of a coupled bi-linear discriminant to align GEI along with the view. The proposed method analyses convergence and sidesteps the under-sampling problem; however, their method faces the issue of overfitting in CNN due to the limit of the labeled dataset. In [64], the authors used a restrictive triplet network to have less impact on the view angle variations. Restrictive triplet network is robust against view variations, which shows that their method can be used in real-time applications. However, they did not consider clothing or carrying covariates. Also, they did not show the performance on any of the view angles, and the loss function used suffers from the problems of hard negative mining.

Existing deep learning techniques frequently depend on the loss/optimization function, which suffers from the problem of hard negative mining. Zhang et al. [65] proposed an efficient loss function known as the angle center loss function for discriminative features of gait. The loss function proposed by them is stoic to a lot of different sizes of temporal windows and local parts. Their loss function learns various sub-centers for each angle, unlike the center loss function, which finds a center for every identity. On a large-scale dataset like OUMVLP, their proposed loss function performs less efficiently because of the loss of information in gait energy images, as it does not capture spatial and temporal features. Zhang et al. [66] proposed a multi-channel convolutional neural network for tackling sequential image sets in parallel. Also, they constructed a novel representation method for gait features with the help of gait silhouettes, ”tri- tuple gait silhouettes.” Their proposed method eradicated the challenge of a sharp reduction in data scale, which happens when GEIs are constructed. Also, the constructed method does not need preprocessing and image merging during image segmentation. Their method is not feasible in the real-time analysis as there are so many silhouettes, more training time, and high computational complexity.

While recognizing humans from gait, either spatial or temporal information is taken. Temporal data (gait-based sequence) is not collected from the gait-based template, due to which the performance of the system degrades. Chao et al. [67] proposed a deep network of GaitSet for Cross-view gait recognition. This network takes advantage of various individual gait frames for learning human gait-based information. They calculated the performance on multiple views and walking conditions, and their technique is stoic to gait images. The biggest advantage of their approach is that the proposed method can integrate gait images from various other videos of different situations but evaluate the performance of only a limited number of silhouettes. Though their approach handled cross-view conditions very well, the approach is not good enough to handle cross-carrying and cross clothing conditions. In [68], the authors proposed a cross-gait-based deep network to identify gait frames. They proposed a quintuplet loss function that increases the performance of the system. They ignored the capability of unique gait representations. Thus, their method is confusing when discriminating various silhouettes with the same carrying, clothing, and illumination conditions. Takemura et al. [69] proposed a pre-trained siamese deep network to recognize a human. Spatial information is collected as they are using convolution-based networks. However, the methods which aim to synthesize gait features (generative methods) are not very accurate. The problem of hard negative mining persists with triplet and contrastive loss functions.

A convolutionThe convolution neural network is sensitive to temporal displacement but can capture minor spatial displacements. As in CNN, matching the differences in silhouettes is evaluated based on the final layer, so they do not provide good results in the presence of very large view variant situations. In [70], the authors proposed a pairwise spatial transformer network for reducing the extra features which are not aligning because of view variation. Their proposed method can improve cross-view gait recognition performance using GAN to preserve generated gait identification information of images. They did not consider all the view angles and their method achieved less accurate results in the CASIA B dataset. Yuqi et al. [71] proposed a joint CNN-based framework that takes less time in training. Efficiency is important in real-time applications. Videos can be affected by noise and illumination variations, increasing complexity, but they do not handle this problem. In [72], the authors aimed to improve recognition rateso and proposed a deep convolution-based LSTM network for gait recognition. Their method was robust in capturing the spatiotemporal feature. However, it did not resolve real application problems related to gait recognition as their method lacked regular sample distribution and their use of small training data.

The walking style of individuals differs with view angle variation, adding to an increase in the complexity of recognizing cross-view gait. The authors in [73], proposed a new method for view-invariant gait representation in cross-view conditions. Their method can capture spatiotemporal features in human gait. Their proposed method is fast as there is no requirement for preprocessing steps such as gait cycle detection and segmentation. It can detect human gait directly from a video database. Computational complexity is high for prediction as it must match with a large number of frames (the number of pixels will be more because no preprocessing is done). Xu et al. [74] proposed a novel cross-view human recognition method based on pairwise-spatial transform networks. In the larger part, their proposed method is reducing not only intrasubject variation but also inter-subject variation. In [75] the authors state that people carrying bags or changing coats in cross-view have lower accuracy in the identification process. They proposed a consecutive horizontal dropout method for human features using deep networks. Their network performance is robust in a cross-view environment, and their method prevents overfitting problems, but they did not consider all the viewing angles.

3.1.4. Spatial and temporal factors affecting gait

It makes perfect sense to get both spatial and temporal information to recognize gait because gait consists of specific sequences of movements in the spatio-temporal domain. The timeline of papers reviewed for spatial and temporal factors is shown in Fig. 12, with the dataset used and the accuracy obtained. Gait energy images capture the spatial information efficiently, but it does not capture temporal information, due to degradation in performance when recognizing humans. The authors in [76], proposed a RNN-based LSTM network for capturing the temporal features. Their method supports the network to train with unlabeled data and is robust when dealing with view transformation. It is not efficient in handling posture and clothing changes.

It is difficult to collect robust spatial features from depth video. The authors in [77], proposed a novel CNN and a RNN to capture spatial and temporal features for gait recognition. Their method captures four-dimensional features. Their method down-sampled the large features collected without losing much information,

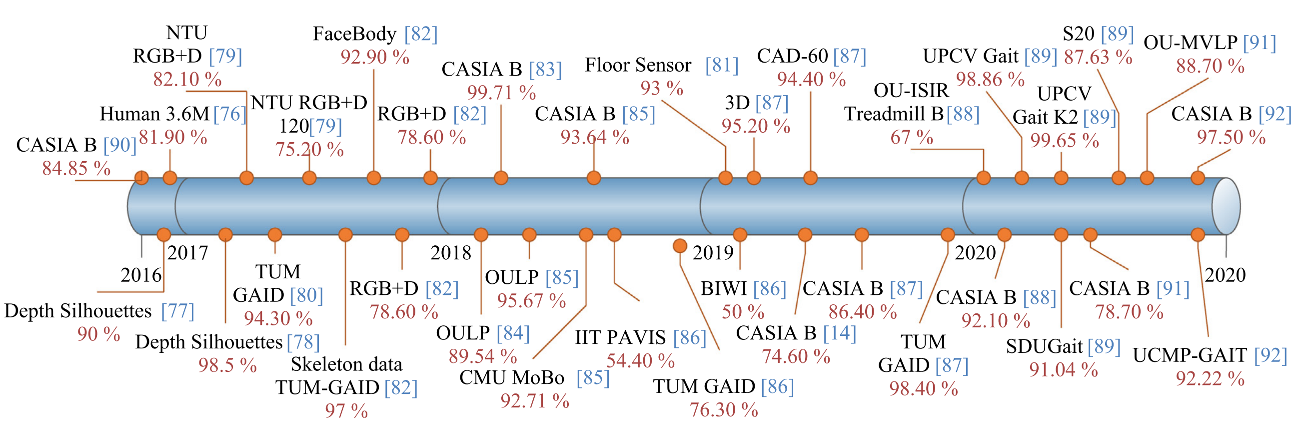

Fig. 12. Timeline of spatial and temporal factors with dataset and accuracy details.

which could be challenging, but they successfully implemented it. Uddin et al. [78] proposed a convolution neural network for collecting features from depth videos. They first extracted local features from depth images and then used optical features to gather spatial and temporal characteristics. Their method performed well by achieving higher accuracy in the absence of covariates. However, they did not consider covariates conditions and did not focus on real-time gait pattern analysis in noisy and complex environments.

View variation makes it difficult to capture spatial and temporal details in skeleton features. The authors in [79], proposed AlexNet using a skeleton sequence into an image and performed an end-to-end learning DCNN. The skeleton sequence image contains spatial–temporal information. Alexnet captures spatial features efficiently. However, its performance decreases in capturing temporal features. In [80], the authors proposed a CNN to capture features from optical flow images to learn high-level descriptors. They have used a three-dimensional convolutional neural network that automatically captured the temporal features and made differentiating easy among different subjects. Their method increased computational complexity as they used a three-dimensional convolutional neural network that takes more time.

Using spatiotemporal data, it is easy to recognize human gait patterns when footsteps are captured from floor sensors. The authors in [81], proposed a method that captures detailed features from the raw data of footstep. Their method is very robust in identifying gait and can precisely differentiate among minute differences in walking patterns of different people. Their proposed method is expensive, and floor sensors cannot be deployed everywhere. Nikolaos et al. [82] proposed a recurrent DNN to learn the high-level sequential data from low-resolution depth silhouettes. They introduced regularization with a hard temporal-attention technique to tackle the small sample size problem. The biggest advantage of a depth sensor is that it is invariant of light and insensitive to color and appearance changes. Their technique captures both spatial and temporal features. However, their method suffers from intraclass variations, and their technique fails in the presence of occlusion.

Convolution neural networks suffer from a loss in temporal information, whereas Long Short-Term Memory suffers from a loss in spatial features. Batchuluun et al. [83] use a shadow-based CNN-LSTM and DCNN to capture spatial and temporal features. Their proposed method is efficient in capturing spatial and temporal information. Their method is not considering the other covariates conditions. Liu et al. [84] proposed a GEI-based SNN, which extracts vigorous spatial and temporal features. Their proposed siamese neural architecture directly learns the effective gait features and uses a parallel neural network for computing the similarities between two human gaits. They have difficulty in handling posture and clothing changes. In real-world scenarios, person appearances are diverse and unconstrained because of changes in posture, makeup, and clothing. In [85], the authors proposed a DNN-based spatial, temporal network, and gait energy image as input. Robust method against view-variations because it obtains both spatial and temporal features. They did not consider covariates conditions. Also, they did not focus on real-time gait pattern analysis in noisy and complex environments.

Depth images are good at collecting spatial and temporal features, but depth cameras are unavailable at all places due to a shortage of available data. Nikolaos et al. [86] proposed a temporal-based attention network with Bernoulli sigmoid function, which collects the features frame-by-frame, and the network is trained using reinforcement learning. Due to the scarcity of data, their proposed technique, which uses pre-trained models, has proven effective in such problems (problems with lack of data). They split the RGB depth images into transfer learning models from the dataset. They did not consider other covariates like view angle and clothing. Also, their method did not have good accuracy.

The multi-view environment creates many hurdles in gait recognition. Francesco et al. [87] proposed a DNN-based spatial, temporal-based approach. They designed LSTM-based temporal DNN, which learns dependency related to time for graph-like structure. Their proposed network is efficient in feature extraction. Long short-term networks are well suited in this scenario because of their capabilities of processing changes across time. The computation cost of their method is high because it generates a model from the silhouette, therefore, high complexity. Their technique does not perform well on small datasets. Xinhui et al. [88] proposed a graph-based spatial–temporal attention-network. Their proposed spatial, temporal graph technique takes less computational time as the system can intuitively make connections with the edges of the graph. The recognition rate is very low in coat-related covariate conditions. In [89], the authors proposed three-dimensional skeleton points using DCNN. Their deep network consists of many layers/ pipelines to differentiate inter and intraclass features along with the cross-class association. For prediction, computational complexity is high as it must match a large number of frames.

Xiaofang et al. [90] proposed a densely connected neural network as the basic algorithm for transfer learning, referred to as DenseNet-based transfer learning. Firstly, this method introduces spatial information of gait through inputting GEI, and then it extracts gait features through DenseNet-based transfer learning. The method in this paper not only reduces the number of parameters but also reduces the storage overhead. It also speeds up the model training speed and improves the recognition rate, but their method did not capture temporal features. The focal Convolution Layer is used to enhance the fine-grained learning of the part-level spatial features. Its core idea is to enable the top convolution kernel to focus on more local details inside each part of the input frame, intuitively exploiting more fine-grained partial information Fan et al. [91] proposed GaitPart, which is collecting temporal features. They have used focal convolutional neural layers to collect the finest features. The features collected by them also include spatial features. They presented a new approach regarding spatial and temporal gait recognition that collects fine-grained features and how every part of the human body requires different modeling to determine the best feature. They did not try to run their algorithm on real-time data.

Due to the presence of less light in coal mines, low-resolution images are obtained, and the recognition rate decreases. To overcome the disadvantage of low-resolution images, Xiaoyang et al. [92] proposed a dense CNN and a stacked-convolutional autoencoder. The proposed network is used for extracting spatiotemporal features in gait recognition. The system is robust in a complicated environment when there is a change in carrying condition, view angle, and variations in hat and clothing attire of people working in mines. Their proposed technique has a high computational cost as the complexity of the algorithm is more. There is a scope for using more images in the dataset and conducting more deep models to enhance accuracy.

3.1.5. Covariates in multi gait and multimodal approaches

In multigait approaches, there exists more than one person, and in multimodal approaches combination of multiple biometrics, approaches can be combined, for example, gait with fingerprint or face. Existing datasets do not contain real-time scenarios like multiple persons walking, occlusion with multiple people, or objects and multimodal data of the same person. Delgado et al. [93] analyzed practical and difficult circumstances where more than one person appears and built up a system that can create datasets from current ones. Their framework helps researchers construct a groundbreaking dataset, unlike all other state-of-the-art approaches, opening new challenges for researchers when they can cope with gait analysis issues utilizing one topic per series. They did not work on occlusion covariate.

Current literature in gait focuses on single subject recognition, but in real-time, there exist multiple persons due to which recognition rate decreases. Xin et al. [94] proposed a model based on attribute-discovery, which considers gait recognition of multiple people, i.e., when multiple people are walking together. Their model requires fewer parameters to be considered, which performs well with fewer data and avoids overfitting issues. Their proposed method is not good at capturing the intraclass variability and cannot represent all the classes. Also, their model is time-consuming and cannot work with occlusion. There are various biometrics feature collection strategies to identify a human, but combining two types of feature collecting approaches is still challenging, and very few pieces of literature have taken advantage of such combinations. Chao et al. [95] proposed a combination of DNN with machine learning feature-based methods with an adaptive late fusion strategy. They try to find points and collect features that can be better classified through gait. It is effectively detecting humans due to a multimodal adaptive late fusion strategy. The authors did not apply any optimization methods for improving the time and efficiency of algorithms.

Existing literature uses the representation of Binary Energy Maps for extracting gait features. This representation has too much noise while extracting the features, leading to lower accuracy. Castro et al. [31] proposed a CNN network that provides an empirical-based comparison for the multimodal fusion of features for recognizing gait. The proposed technique is independent of the filter as they are directly taking images from input data without the preprocessing step to get optimal features for recognizing gait. The computational cost of their model is also less as it does not use a preprocessing step. There is a decrease in the performance because models cannot obtain unique flow vectors resulting in a view-dependent model. Mehmood et al. [32] proposed a CNN-based densenet-201 model for extracting the features and used hybrid selection for reducing the features. Pretrained CNN models are used for automatically extracting features, unlike in classical methods (to calculate the shape of geometric). Their proposed system fails when the training data is small and computational cost and time increase due to the fusion process.

3.1.6. Speed variation

Many factors influence recognition accuracy, like walking speed. A shift of speed or pacing is how one walks, allowing to change all time-based functions. In addition, natural characteristics are often vulnerable to modification. For example, it is widely understood that shifts in phase duration and joint angles will arise as the walking pace changes. Peak values of instantaneous land reaction force often vary at various rates. The work in [96] states that traditional approaches use silhouette features like Gait Energy Image, which averages the gait sequence in one gait cycle. This feature representation efficiently captures the spatial features, but detailed features are lost, including temporal features. They proposed a mutual subspace approach (CNN-based generalized subspace method). Their proposed method efficiently captures the detailed features of the sequences lost in traditional approaches. The computational complexity and cost are high, with less recognition rate.

Gait recognition is challenging when dealing with a person's speed changes. Sometimes the system fails to detect the same person when he/she is with a different prop and has a slight change in speed. For example, a faulty system might consider the subject an imposter if they carry their phone and have a slouchy silhouette. Liang et al. [99] proposed a deformable network that minimizes intrasubject distortion changes in posture. Their proposed method diminishes posture changes, thus making intra-subject changes very small and increasing gait recognition performance. They did not consider inter-subject deformation to get a more accurate network for discriminative posture features.

3.2. Model based approaches

Model-based methods are classified into two groups: pose-based [26,29,33,41,43,45,76,79,87,105–113,144] and sensor-based [81,114–143,145–150] approaches. Model-based feature representation models usually utilize distance or joint angle on the human body for gait identification after designing the entire human body. The aim is to efficiently and accurately detect key positions. Most methods are complicated and costly in computing as these methods need to calculate key points in every frame. Model-based representation does not depend upon the size or vision of characteristics. But it depends on the quality of the video. They can cope with the numerous intraclass variations induced by various covariates that influence the appearance of the items, such as clothes, carrying, and viewing angle. But, calculating the key points in every frame is not a trivial job, and thus computational complexity of these methods is very high.

3.2.1. Covariates in pose-based approaches

As part of creating computer vision for gait recognition, another unique breakthrough enhances the accuracy of gait recognition. Pose-based approaches assist in executing gait recognition through collecting features by joint angles. The timeline of papers reviewed for pose-based is shown in Fig. 13, with the dataset used and the accuracy obtained.

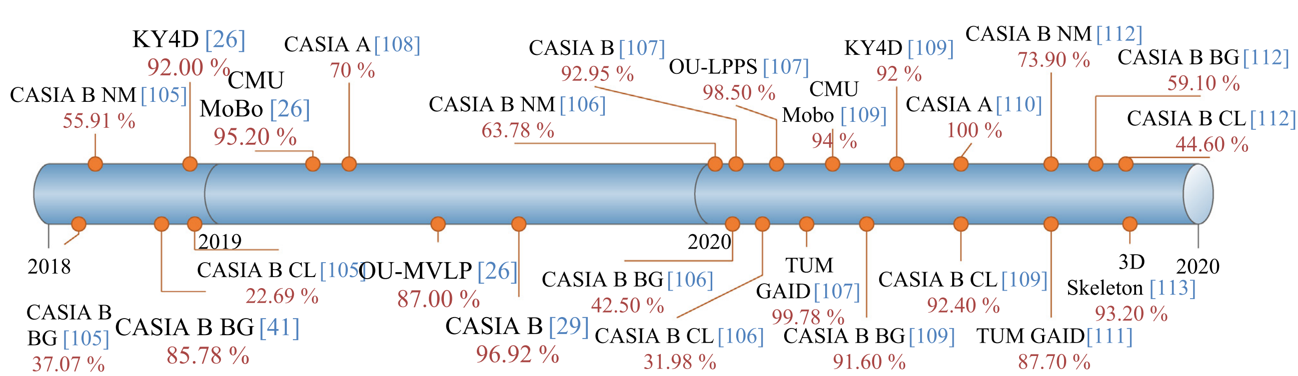

Fig. 13. Timeline of pose-based papers with dataset and accuracy details..

Variations in viewing angles, clothing, and carrying conditions degrade the gait recognition rate. Model-based approaches are quite efficient in handling such variations. Weizhi et al. [105] proposed a three-dimensional convolution neural network and LSTM network capture spatial and temporal features from two-dimensional images. Their model is good at capturing spatial and temporal features. The three-dimensional pose-based method can extract features more precisely in view variation than two-dimensional pose-based methods. But, their model has very low accuracy. Initial techniques to recognize gait were mostly based on appearance. In appearance-based techniques, parameters were taken from silhouettes. It is easy to compute the silhouettes and gives good recognition accuracy but gets highly affected due to covariates. In contrast to appearance-based, model-based approaches are not much affected due to covariates but require a high computational cost and do not give good recognition accuracy with low-resolution images. Liao et al. [106] proposed a model based on pose features that use CNN for gait recognition of 3-dimensional human pose. Since their model uses 3-dimensional pose estimates, it is invariant to view angles and various factors changes. Also, using a three-dimensional pose for spatial and temporal features improves the accuracy of the model. Their proposed model effectively represents gait features and robustness toward covariate condition variations. But their model has very low accuracy and a high computational cost as it converts two-dimensional joint points to three-dimensional ones. There is a big fluctuation in the accuracy when the angle difference between the train and test set becomes 90 degrees.

Early techniques do not consider the entire height of the subject, due to which there is a decrease in recognition accuracy. Sokolova et al. [107] considered the entire height and joint angle points. They proposed a novel pose-based CNN approach that considers the entire height of a silhouette and joint angles. Their method is not much affected due to covariates, especially viewing angles. Their consideration of the entire height and the joint angle points improves the recognition accuracy instead of considering the entire silhouettes. Their model has a lower recognition rate because their technique only uses motion while eliminating all the color information. Tavares et al. [108] proposed the PifPaf technique that extracts features from noisy environments with noises. They are using images from high-resolution cameras and thus get the best pose features. Their model has given poor results and has low accuracy. Also, they did not consider covariate situations. Their proposed model is costly as it requires high-resolution cameras, and high-resolution cameras are not installed at all places.

Luo et al. [109] targeted difficulties in real three-dimensional structured data and proposed a hierarchical temporal memory network using convolution and recurrent neural networks. Their proposed method uses an estimation technique to detect body shape, images of body-parsing, and virtual garments. Due to the use of all these techniques, their model performs well in object and clothing variations. Their method captures both spatial and temporal features, which is more time-consuming. This algorithm performs well only on large-sized datasets. Hossen et al. [110] proposed a RNN-based method. Recurrent neural networks with gated recurrent units architecture are very powerful in capturing the temporal dynamics pose sequence of the human body and performing recognition. The authors also designed a low-dimensional gait feature descriptor based on the 2D coordinates of human pose information proven to be invariant to various covariate factors and effective in representing the dynamics of various gait patterns. Their results are robust as they found effective gait features for the view variant environment. Nevertheless, they did not consider the cross-view conditions.

Current methods based on skeleton have achieved less accuracy as they tackle normal and noise data while recognizing gait. In [111], authors proposed a method based on skeleton using the siamese network along with autoencoder networks. Their model is effective for covariates like variations in clothes, time, and carrying objects. Their method can reconstruct common trajectories and perform accurate gait recognition with images with side view angles. They are not dealing with the cross-view condition. Li et al. [112] proposed a model based on a joint relationship mapping pyramid to capture spatial and temporal features. The computational cost of their approach is very high. Kooksung et al. [113] proposed a new approach for automatically extracting the features with the help of an autoencoder based on RNN. Their proposed method is better at recurrent neural network features than principal component analysis and singular value decomposition. Their recurrent neural network performance is very low because they did not consider sequential data while decreasing dimensionality. In [222], the authors extracted the 3D joint feature and 3D bone feature based on 3D skeleton data and designed a dual graph convolutional network to remove corresponding gait features and fuse them at the feature level. Their method works well on view variation. In [223], authors proposed a GaitGraph that combines skeleton poses with Graph Convolutional Network (GCN) to obtain a modern model-based approach for gait recognition. The main advantages are a cleaner, more elegant extraction of the gait features and the ability to incorporate powerful Spatio-temporal modeling using GCN. In [224], authors proposed a Pairwise Graph Convolutional Network to capture gait features from skeletons. Their method is suitable when identical-view and to cross-dress cases exist.

Luo et al. [26] proposed a novel 3-dimensional CNN based on gait recognition that is robust against clothing and view variations. The authors concentrated on making a robust system for real-time surveillance scenarios invariant from clothing, viewing, and carrying conditions. Their approach requires high computational complexity because of the computation of pose features used by their method and, thus, is costly. [29] proposed a multi-stage CNN for identification and verification. The system is GEI skeleton-based. They have addressed covariate conditions of clothing variation and view variation, and their application can be used for real-world data. In their model, more tuning is required to make it view-invariant. Generally, model-based methods are more robust to clothes and view invariance, but it is not so in this case. Liao et al. [41] proposed a pose-based spatio-temporal model for extracting spatial–temporal features. Changes in carrying and clothing do not have any impact on body joints. In clothing, especially coats, the results are very low. The computational cost of the system is more.

3.2.2. Sensor based approaches

Challenges with optical camera-based gait recognition methods include background motion, illumination problems, and display dependence of the extracted features, which do not exist in wearable sensing-based gait technologies. Sensor-based, signal processing, and data acquisition approaches play a vital role in authenticating the user using gait signature and are less susceptible to covariates. Sensor-based gait recognition techniques are shown in Table 1, which include covariates considered, the dataset used, the accuracy obtained, the type of sensor used, and deep learning architecture.

Wearable-sensors and nonwearable-sensors: There are two categories of data acquisition technologies. Those are wearable and non-wearable sensing devices. Sensor instruments connected to individual joints are classified as wearable, and non-wearable sensors are mounted on the floor or walls to learn objective gait patterns. The measurement of stride-related parameters of biomechanics is the main reason behind achieving gait signatures/features. Hannink et al. [114] tried to gather the best gait features

Table 1
Gait performance parameters under sensor-based approaches

Cite	Covariate consider	Dataset	Accuracy	Sensor used	Gait Pipeline
[144]	Spatial Temporal	eGaIT - embedded	80.78	Wearable sensors	DNN
[115]	Adversarial	GoogleNet	82	Accelerometer, Gyroscope	DCNN
[116]	Person re- identification	SZTAKI-LGA-DB	96.84	Lidar (VLP16, HDL64)	CNN
[117]	Walking Pattern	OULP	96	Accelerometer, Gyroscopes	CNN
[118]	Motion Pattern	10 subjects	97.06	Wearable accelerometer	DCNN
[119]	Pose	ZJU-GaitAcc	93	Accelerometer	DNN
[120]	Wearable Devices	McGill dataset	96.6.	Accelerometer, Gyroscope	CNN
		OU-IS	67.9		
[121]	Multimodal authentication	EID-M EEG	99.57	EEG Headset, Accelerometer	RNN; LSTM
				Gyroscope, Magnetometer	
[122]	Walking Pattern	22 subjects Treadmill	98	Radar micro-doppler	Residual network
[123]	Nonlinear gait Dynamical	ZJU-GaitAcc	92.2	Accelerometer	RBF network
[124]	Recognition	34 participants	90	Accelerometer	CNN
[125]	Recognition	McGill University	71	Accelerometer, Gyroscope	CNN
[126]	Real-Time constraint	Grayscale R-D maps	99	Radar sensor	DCNN - LeNet-5
[127]	Recognition	ZJU-gaitAcc	98.7	Wearable accelerometer	RCNN
[128]	Abnormal Gait Patterns	Original dataset	87.97	IMU, Gyroscope	LSTM-CNN
	Tiptoe, Hemiplegic				
[129]	Recognition	OU-IS	94.8	Accelerometer, Gyroscope	DNN
[130]	Recognition	Own dataset	94.25	Accelerometer, Gyroscope	CNN
[131]	Spatial Temporal	DVS128-Gait EV casia	96	Dynamic Vision Sensor	DNN
[132]	Multimodal	CASIA B	91.3	Camera, Magnetometer	3D CNN, LSTM
[133]	Abnormalities	No dataset	94.1	Radar	Deep learning
[134]	Recognition	OU-IS	89	Gyroscope, Accelerometer	RNN
[135]	Recognition	Calibration dataset	21.22	Wearable pants	Encoder decoder
[135]	Temporal	12 participants	84.88	F-scan	LSTM
[137]	Temporal, Noisy image	3D Flash lidar camera	84.88	RGB camera	Skeleton detector
[138]	Temporal	DFNAPAS	80.2	Accelerometer	CNN
		SisFall	74.58		LSTM
		UniMiB-SHAR	82		
		ASLH	81.72		
[139]	Data scarcity	CNU	82.53	Accelerometer	CNN
		OU-IS	82.53	Gyroscope	
[140]	Clothes	Identification	93.5	Accelerometer	CNN
		Authentication	93.7	Gyroscope	RNN-LSTM
[141]	Context	UIR	94.2	Accelerometer	Encoder
	human action Interference	HHAR	87.9	Gyroscope	Decoder
[142]	Centre of pressure	Treadmill-pressure data	99.9	Force platform pressure sensor	CNN - Resnet
[143]	Variations in activities	SBHAR	97	Wearable sensors	CNN
		UniMiB	93.6		
		REALDISP	94		
[144]	2-D Vision abnormalities	3D walking/ 97.2	97.2	Microsoft Kinect sensor	GAN - LSTM
[145]	Spatial Temporal	Signals of a walking	91	Accelerometer	CNN
[146]	Foot–ankle kinematics	PART VI - 3D motion	96.02	Soft robotic sensors	ANN-LSTM
[147]	Security Authorize	SIIT-CN A, B, D, F, G	98.4	Kinect sensor	CNN
		C	100		
		E	98.9		
[148]	Privacy, Light conditions	mmWave-radar waves	90	mmWave sensor	CNN-ResNet18
[149]	Preserving privacy	12 person data	89	Millimetre wave radar	DRNN LSTM
[150]	Recognition	Raw ankle data-IMU	80	Accelerometer, Gyroscope IMU	CNN

from the data collected by inertial sensors. They proposed a Deep CNN architecture to translate abstract information from a wearable sensor. The implementation of the framework can be generalized and flexible. The collection of representative training data is time-consuming and sometimes infeasible. Adversarial perturbations are vulnerable, and because of them, gait features are not collected properly. Vinay et al. [115] proposed a deep learning architecture to reduce the effect of attack by adversarial perturbations for providing a secure authentication gait-based system. Their proposed method is beneficial and does not degrade authentication performance when attacked by adversarial perturbations. When they conducted tests on the system with minor attacks, the accuracy of the algorithm was decreased by 40%. Thereby, the accuracy of their authenticating system is inadequate. Velodyne HDL-64E LiDAR was prevalent for identifying humans in the last few years. However, the cost and weight of the sensor are very high, due to which it does not support real-time surveillance applications.

Bence et al. [116] proposed a Lidar-based GEI descriptor that effectively cost-effectively measures gait signature than previously used technologies such as Velodyne-VLP-16 LiDAR-based scanner. Their method produced better results than other methods as it works on the motion of an entire body. Low-cost, user-friendly sensor-based gait recognition technology embedded into the wearable device is always in demand. Nguyen et al. [117] proposed a CNN-based authentication system to recognize gait by designing wearable sensors with a gyroscope and accelerometer. Their model portrayed good accuracy but did not consider different contexts, such as overcoming any covariates or taking the entire motion of the body. Handcrafted feature extraction techniques could be erroneous and may be subject and situation-oriented. An inertial sensor that collects motion data of human gait has a typical structure (problem-specific). While training the system with such data may generate entirely different results, it restricts this manual handcrafted feature extracting technique's generalization capability. Dehzangi et al. [118] proposed a deep CNN-based network for extracting discriminative gait features. High accuracy is achieved due to the gathering of complementary discriminative gait signatures, and it effectively expands the transformations via 2D time frequency. They require labeled signals for every subject because signals in the human-robot interface are usually noisy and user-dependent. Gadaleta et al. [150] are gathering data from accelerometer and gyroscope sensors. IMU (ankle worn) is used to capture raw motion data. The collected data of IMU sensor is sent to the pre-processing stage, which pre-processes raw data and is utilized in training signal processing tools. They have used CNN for extracting the feature.

3.2.3. Adversarial attacks and anonymization

The biometric function is how secure it is to spoofing assaults. Snooping on user behavior applies to exploiting "artificial patterns" to obtain illegitimate access to confidential or secure resources. For instance, the possible attacks on the gait recognition system are shown in Fig. 14. In early techniques, imposters attack gait recognition systems by mimicking the walking and clothing patterns. Meijuan et al. [100] use spoofing attacks by replacing the existing dataset with produced image samples. The authors proposed a method based on the generative adversarial network to make fake images of subjects. Their method effectively deceived the GaitSet network. They computed the similarity score of samples in the gallery and the attacking video. If the efficiency of calculating the similarity score is good, then the performance of imposters attack could be done.

Traditional cameras could not capture the changes in intensity. Event-based cameras can capture the intensity changes, but the output images still suffer from unrealistic, noisy, and low-resolution information. Wang et al. [101] proposed a reconstruction method avoid unrealistic, noisy, and low-resolution information by using unsupervised upsampling adversarial learning. The accuracy of a model depends on image resolution. If the images used for feature extraction are of low quality, the accuracy goes down. So, they reconstructed the images to increase the image quality for better accuracy of their model. However, they did not consider the possibility of the effects of event stack forms on the performance of EventSR. Gait recognition systems have higher accuracy than other biometrics, but there is still scope for exploring the robustness of adversarial systems. Ziwen et al. [102] proposed a temporal-based attack using sparse adversarial. Attacks in their model achieved good imperceptibility and a high attack success rate. They have a limited capacity of trained generators, due to which some of the generated adversarial samples are far from the normal distribution of gait silhouette. Xue et al. [103] proposed a GAN for capturing the details in time-serial data, which is lost in capturing the adjoining frame rate. The proposed loss function, margin ratio loss, improves the gait classifier's performance. They did not try their algorithm on various covariates like carrying and clothing conditions. They have worked on very small datasets. They must have evaluated large datasets such as OU-ISIR MVLP to get generalized results.

Information anonymization is one method to keep sensitive details confidential on the internet. That is, the method of anonymizing data sets so individuals can no longer be marked. A randomized, anonymized gait is developed by mixing pre-trained CNNs with noise gait pictures. Tieu et al. [104] modified gait silhouettes such that the gait of a person remains unidentified, but still, the naturalness of gait remains the same. Their proposed method of gait anonymization can prevent unauthorized gait. The method developed achieves anonymization with gait naturalness when high-quality silhouette gaits are used. However, the results look unnatural when low-quality ones are used. They proposed to slightly deform human gait silhouettes by mixing an input silhouette with another called a noise silhouette. However, their method only focuses on the static aspect of human gait, although the dynamic aspect is also important and privacy-sensitive. To

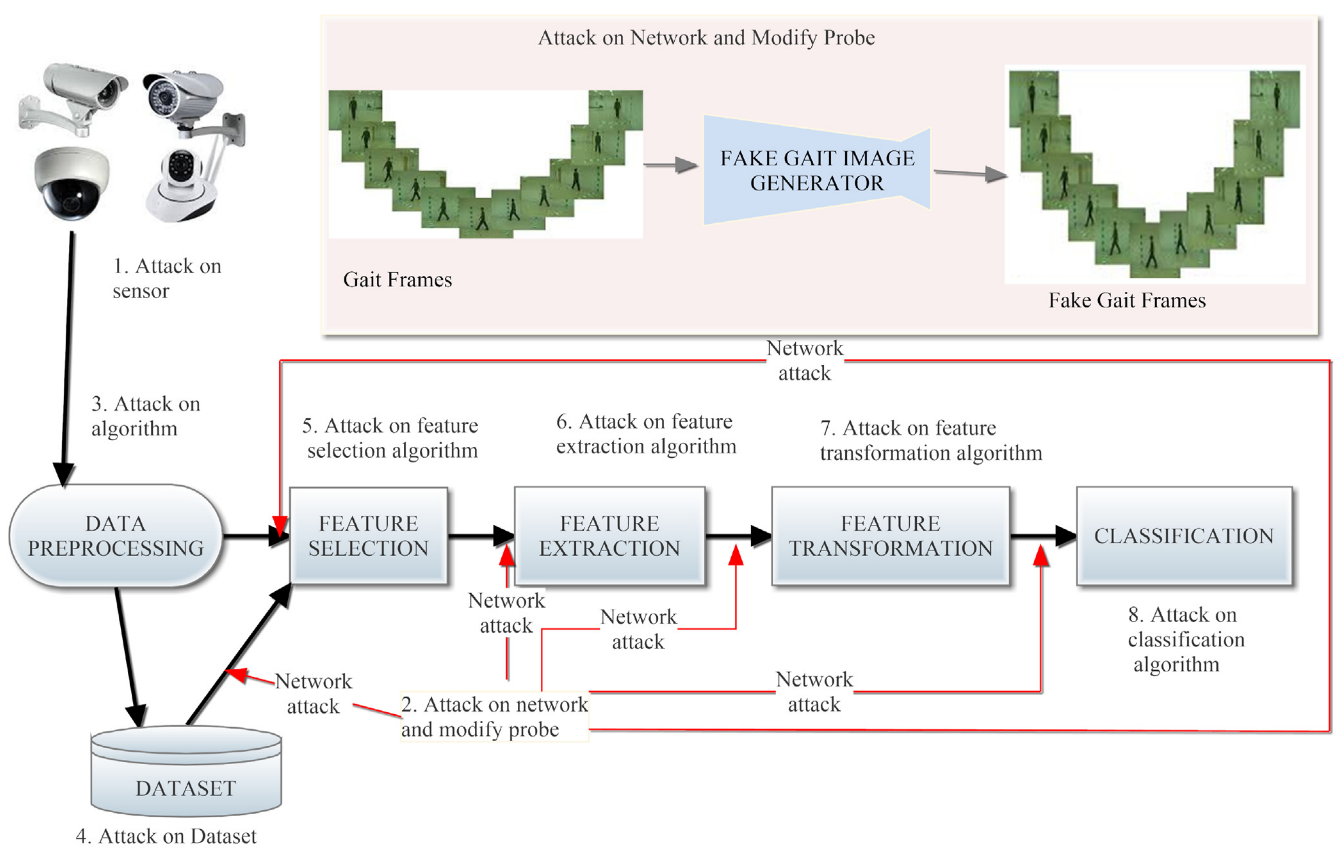

Fig. 14. Covariate attacks on the network to modify probe with the gallery images.

Table 2
Gait datasets for analysis of multiple covariate conditions

Cite	Year	Dataset name	Covariates	Subjects	Sequences	Environment
[151]	2012	OU-ISIR Treadmill B	Cloth, Speed variation	122	1870	Inside
[152]	2012	OULP	Cloth variation	4007	1872	Inside
[153]	2017	OU-ISIR OULP-Age	Spatial Temporal	63486	31923	Inside
[154]	2018	OULP-Bag	Carrying condition	62528	29097	Inside
[155]	2018	OU-MVLP	View	63486	28	Inside
[156]	2020	OU-LPPS Pose Seq.	View; Pose	10307	25	Inside
[157]	2014	OU-IS	Sensor	744	2976	Inside
[158]	2003	CASIA-A	3 View	20	240	Outside
[159]	2006	CASIA-B	View, Cloth variation	124	13640	Outside
[160]	2006	CASIA-C	Speed variation	153	1530	Inside
[161]	2011	TUM-IITKGP	Occlusion, Carry	35	840	Inside
[162]	2014	TUM-GAID -RGB-D	Spatial Temporal	305	30 GB jpg files	Indoor
[163]	2012	USF	Cross view	122	1870	Outside
[164]	2001	CMU MoBo	Carrying, Cross-view, speed	25	600	Treadmill
[165]	2010	KY4D	Pose, Clothes	42	168	Inside
[166]	2014	Human3.6 M	Spatial Temporal	11	3600000	Inside
[167]	2014	NTU RGB + D	Spatial Temporal	60	56880	Inside
[168]	2019	NTU RGB + D 120	Spatial Temporal	120	114480	Inside
[169]	2014	BIWI	Spatial Temporal	50	100	Inside
[170]	2012	IIT PAVIS	Spatial Temporal	Real Time	25 key points	Inside
[171]	2017	3D	Spatial Temporal	1	16	Inside
[172]	2012	CAD-60	Spatial Temporal	60	57600	Inside
[173]	2015	UPCV Gait	Spatial Temporal	1	Points	Inside
[174]	2013	UPCV Gait K2 S20	Spatial Temporal, Occlusion	18	Points	Inside
[175]	2016	SDUGait	Spatial Temporal	52	1040	Inside
[176]	2014	iLIDS-VID	Recognition, View	300	600	Inside
[177]	2011	PRID-2011	Multiple person	475	1134	Inside
[178]	2012	SZU RGB-D	Recognition, Pedestrian	4835	14505	Inside
[179]	2014	AVAMVG	Recognition, View	20	200	Inside
[180]	2014	3D SKELETON	Pose	10	557	Inside
[181]	2017	Ev-RW	Speed	13	26	Inside, Outside
[182]	2012	Caltech	Recognition	15	90	Inside, Outside
[183]	2017	INRIA	Recognition	405	810	Outside
[184]	2008	ETH	Recognition	853528	8580	Inside
[185]	2015	eGaIT - embedded	Wearable sensor	17	101	Inside
[186]	2015	GoogleNet	Sensor-based	60000	–	Inside
[187]	2018	SZTAKI-LGA-DB	Rador-based identification	54	10	Outside
[188]	2015	ZJU-GaitAcc	Pose - sensor	175	11	Inside
[189]	2013	McGill dataset	Wearable sensor	20	40	Inside
[190]	2016	EID-M EEG	Sensor-based	52	11	Inside
[191]	2020	DVS128-Gait + EV	Sensor-based	21	100	Inside
[192]	2019	Calibration dataset	Sensor-based	32	166	Inside
[193]	2016	3D Flash lidar camera	Temporal; Noisy image	16300	48900	Inside
[194]	2017	SisFall	Temporal	19	38	Inside
[195]	2017	UniMiB-SHAR	Temporal	30	11771	Inside
[196]	2016	ASLH	Temporal sensor based	63	378	Inside
[197]	2015	CNU	Sensor-based	495	4	Inside
[198]	2018	SIIT-CN	Sensor-based	Real Time	–	Inside
[199]	2011	Large population	Carrying, Cloth footwear	115	2128	In, Out
[200]	2000	Small population	Carrying, clothing, footwear	12	–	Inside
[138]	2020	DFNAPAS	Temporal sensor based	Real Time	–	Inside
[43]	2019	FVG	Pose	226	2856	Inside
[101]	2020	ESIM-RW	Recognition	4148	8296	Inside
[99]	2016	MARS	Identification	Real Time	–	Inside
[141]	2020	HHAR	Human action	–	–	Inside
[141]	2020	UIR	Human action	–	–	Inside
[146]	2020	PART VI – 3D motion	Foot–ankle data	–	–	Inside
[143]	2020	SBHAR	Activity variation	–	–	Inside
[143]	2020	UniMiB	Activity variation	–	–	Inside
[201]	2019	CASIA E	Carrying, Clothing, View	1014	101400	Inside

use a noise gait, one must be confident that the gait recognition device would not mistake the initial gait with the noise gait. Furthermore, due to the small naturalness loss function, the anonymized gait looks less natural, particularly for viewing angles of 0 and 180 degrees.

3.3. Dataset for various covariates

Collecting huge sample data to train and evaluate gait recognition methods and covering all the covariate conditions is important. More technologies to analyze gait have been introduced recently; this segment explores gait databases developed after 1990 in Table 2. An analysis of the free accessible gait datasets is provided by comparing them with their published year, dataset name, number of subjects in the dataset, sequences, the environment in which the dataset is captured, and the link of publicly available datasets. To date, no dataset includes all the covariate conditions. So, if the researchers try their algorithm on different datasets, the situation is very different from the real-time scenarios, and thus, the performance of the real-time situation differs.

4. Research gaps and proposed solutions

While studying many research articles and implementing some of the proposed solutions, we encountered many unanswered questions, and the same are summarized below as research problems. Appropriate answers to these questions can result in a useful extension of learning and knowledge in this area.

1. Which approaches (Appearance, Pose, or Sensor-based) should be chosen for a gait-based surveillance system?
2. Which covariates are least handled?
3. How to deal with a variety of covariate conditions?
4. How many papers are handling the fusion of covariates?
5. What is the current level of accuracy for recent papers? Is gait recognition feasible in a real-time situation?
6. Which deep learning pipeline is used in the most recent papers? Which deep learning architectures are appropriate for gait surveillance?
7. How many datasets are available, and what is the minimum number of datasets needed to test the algorithm? Is there a dataset including all the real-time scenarios for a gait-based surveillance system? Which is it better to use; a 3D or 2D dataset?
8. Which platform (Fog/Edge) is best for the deep learning model?
9. What are the privacy concerns of a gait surveillance system, and how can they be addressed?
10. Does the gait surveillance system support real-time analytics?

4.1. Research gaps

The above research problems address several research gaps that are explained below.

• Appearance-based methods are less complex, but they have a low level of accuracy. Pose-based methods are precise, but they require a lot of computing power. Wearable sensor-based methods are accurate, but they are not practical everywhere because they require human contact, and non-wearable sensors are not precise.

• Handling covariates is still difficult as the accuracy of the gait recognition algorithm is low.

• Covariates such as temporal features, multi-gait, and speed variation have received little attention.

• Current works in the literature consider one type of covariate at a time as no gait algorithm can handle all the covariate conditions in real-time. Also, the accuracy of these algorithms is low.

• The accuracy achieved so far ranges from 40% to 100%. Full accuracy (100 percent) is usually achieved without covariate conditions when a dataset is manually preprocessed. The presence of covariates, particularly in occlusion and view angle (0 and 180 degrees), impacts the accuracy of the gait recognition algorithm, which is unacceptable in a real-time scenario.

• Most papers use automated CNN layers to obtain the feature set. However, hyper-tuning such CNN parameters remain difficult and vary from dataset to dataset. For gait surveillance systems, pre-trained deep networks are frequently used. Still, the results are unacceptable for a few datasets (real-time datasets). For a few datasets, which are manually preprocessed, the accuracy is high.

• Various gait datasets cover all the covariate conditions for different individuals, but no single dataset contains all the covariate conditions with the same individuals. Three to four datasets are likely to be required to validate the algorithm. However, because there is no single gait dataset that covers all real-time scenarios for a gait-based surveillance system, no condition is still designed to make the algorithm work in the real-time scenario. 3D datasets are more convenient because they capture more details about the subject. Due to the high cost of 3D capturing sensors, only a small number of gait 3D datasets are available.

• Edge computing is more secure than fog computing. However, edge computing is only supported by a small number of deep learning architectures. The optimization of these deep networks is critical.

• Privacy is still a concern in gait recognition, and no literature is available that focuses on security issues.

• Real-time analytics remains difficult due to the lack of optimized deep architectures and datasets covering all covariates.

4.2. Proposed solutions to handle gait covariates

Few approaches that have been used to date utilize model-based algorithms to collect gait biometrics. Today, computing capacity allows the exploration of 3D gait biometrics with the help of 3D methods. 2D database and pose recognition systems are computationally simple and quick but appear to have limitations due to dependency. The 3D gait dataset and techniques can overcome common problems like 3D gait data providing more details than 2D data. Even 3D data can effectively deal with occlusion, noise, and varying view angles. Deep learning models collect either spatial or temporal features in most gait recognition approaches. However, most of the temporal information is lost while using CNN. Both spatial and temporal information needs to be collected to get better performance.

The proposed methodology suggests a fusion-based deep learning pipeline to obtain maximum features and uses anonymization to make a robust and secure Gait Recognition System (GRS). The user identity can be hidden through anonymization. This technique is essential as the stored dataset contains the user identity. An anonymized silhouette looks the same as the original silhouette (that is, the naturalness of silhouettes persists). In case the dataset is hacked, the hacker will not know whether the dataset is encrypted or not. We propose a specialized deep CNN architecture for gait recognition that collects spatial and temporal features. It is depicted in Fig. 15, and it is less susceptible to a variety of covariate conditions in gait recognition.

Following are the features of the proposed deep CNN architecture:

• A fusion of appearance-based and pose-based methods is proposed to optimize the solutions. Features must be calculated with appearance-based methods to use low computational power and then be passed to pose-based methods in the deep learning pipeline to achieve better results.

• More real-time datasets should be developed so that the training can be done rigorously, and a large number of covariates can be handled to achieve better accuracy of the gait recognition algorithm. Also, multi-gait problems can be solved with real-time datasets.

• LSTM has been proposed as a part of a deep learning pipeline to capture temporal features and speed variation.

• Multiple optimized gait pipelines have been proposed to capture all the covariate conditions in real-time.

• Hyperparameter tuning is done to a deep learning pipeline for real-time datasets so that they are not manually preprocessed for training gait recognition algorithms.

• More techniques should be evolved to capture 3D datasets or convert 2D datasets into 3D datasets to make the GRS more reliable and cost-effective.

• Edge computing is more secure than fog computing; therefore, different GPU-enabled microcontrollers can optimize deep networks for GRS like jetson nano.

• Anonymization strategies have been proposed to make the GRS more secure.

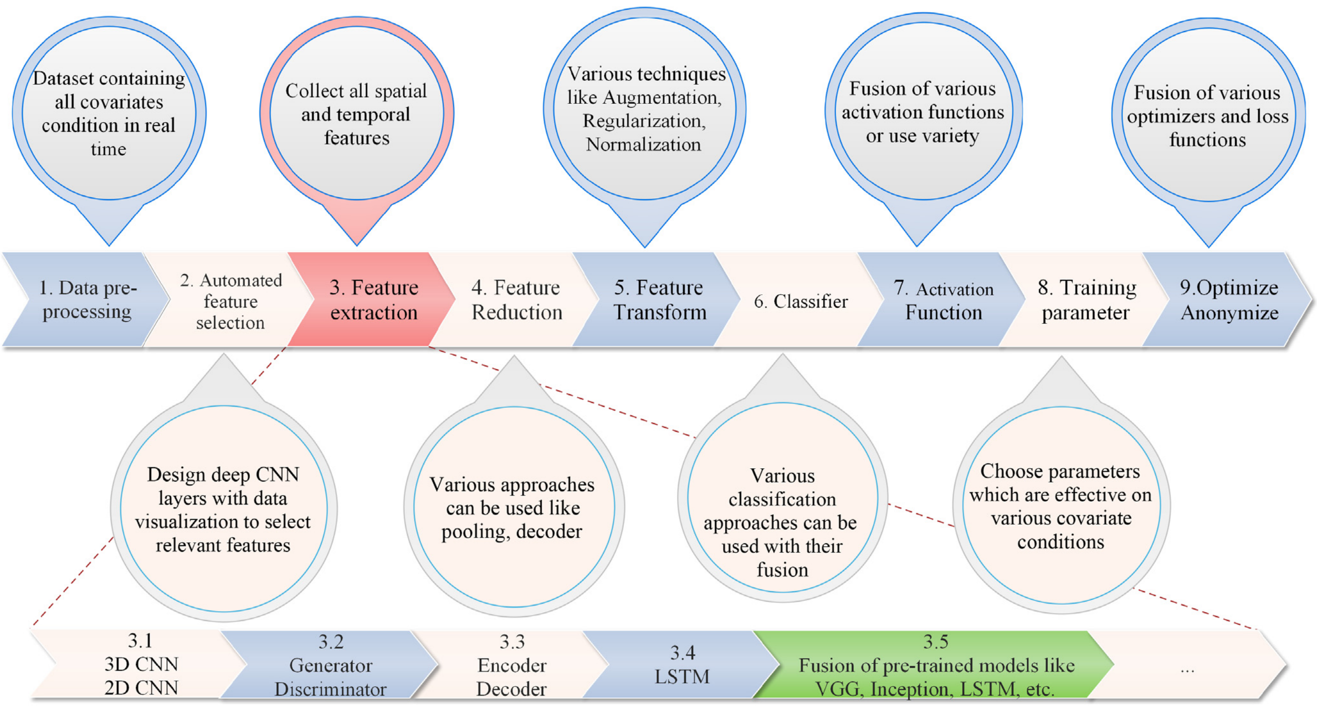

Fig. 15. Proposed deep learning pipeline to deal with gait covariates.

• Real-time analytics has been proposed using powerful hardware architectures and designing optimized deep learning networks that work on large datasets covering all the covariates conditions.

4.3. Salient outcomes of the review

This section will look at the outcomes in terms of the most used covariate techniques and gait pipelines. Next, we provide details about the most popular datasets used by these deep learning pipelines in research papers published in the last five years. In Fig. 16, the ratio of gait research papers (in the form of a pie chart) that has been published in the previous five years is given.

Most of the work on these papers is done by using sensor-based and clothing-based variations. Variability in speed and the inclusion of multi-gait are simply overlooked. On the other hand, view variation, spatial & temporal features, and pose-based work have received some attention.

According to the results, deep learning models significantly outperform machine learning models, which is why deep learning has appeared in several papers in the past five years. When deep learning is employed, a convolutional neural network is the most used approach for automatic feature extraction. The most widely used deep learning approaches for gait recognition are depicted in Fig. 17. CNN is implemented in 35% of papers, LSTM, GANs, and so on are all applied in the remaining 65% of papers.

Various datasets are used to train the gait algorithm against a wide range of covariates. Fig. 18 depicts a ratio of the most widely used gait datasets available for gait recognition. The most adopted gait dataset is CASIA B, followed by OULP, then various wearable sensors, followed by TUM-GAID, OU-MVLP, and others.

4.4. Potential applications of gait biometrics

When deciding on applications of appearance, pose, and sensor-based approaches, one frequently becomes perplexed about the best option. To make this clear, we have provided a clear bifurcation between applications to choose the best approach. Fig. 19 shows a few detailed examples of application-specific approaches. The benefits and limitations of approaches based on appearance, pose, and sensor-based were discussed in Section 4.

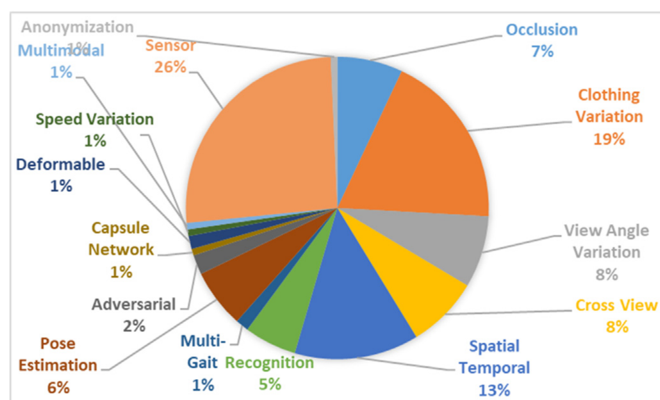

Fig. 16. Ratio of covariate papers based on Appearance, Pose, and Sensors.

Appearance-based methods do not rely on video quality and require little processing power. As a result, these approaches are used where low-resolution cameras and no GPU facilities are available. Pose-based methods are better for handling covariates than appearance-based methods but have high computational complexity. Wearable and non-wearable sensors are the two types used in sensor-based methods. Wearable sensors have higher accuracy than non-wearable sensors, but they require human intervention, so they are not applicable everywhere. Even though non-wearable sensors are inaccurate, they are widely used because they do not require human intervention. The traditional biometric methods of identification have faced many difficulties and limitations. For example, it is not advisable to remove masks for security checks at airports in a pandemic. Due to the limitation of technology, people had to remove face masks to get themselves identified

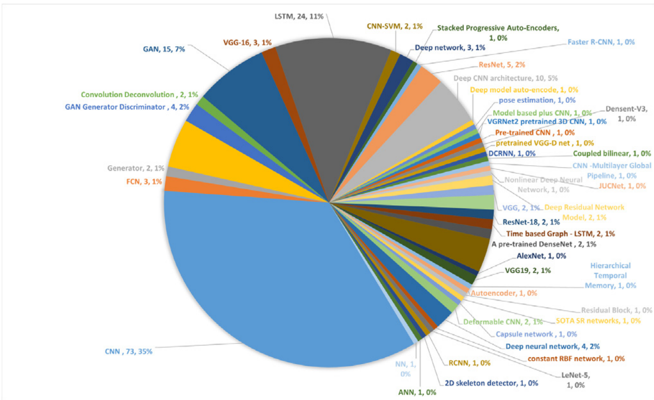

Fig. 17. Ratio of the most adopted deep learning approaches for gait recognition.

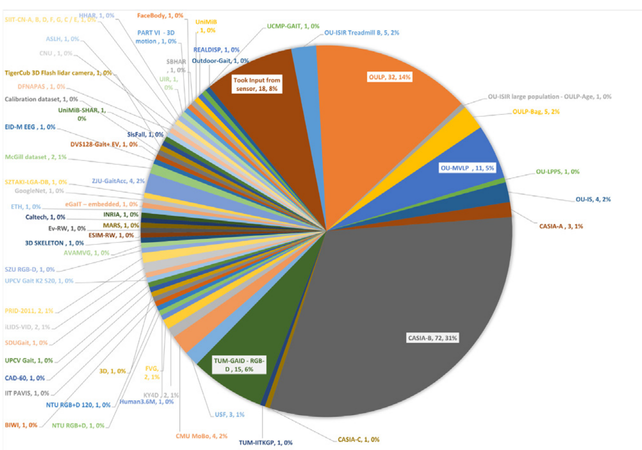

Fig. 18. Ratio of the most adopted gait datasets.

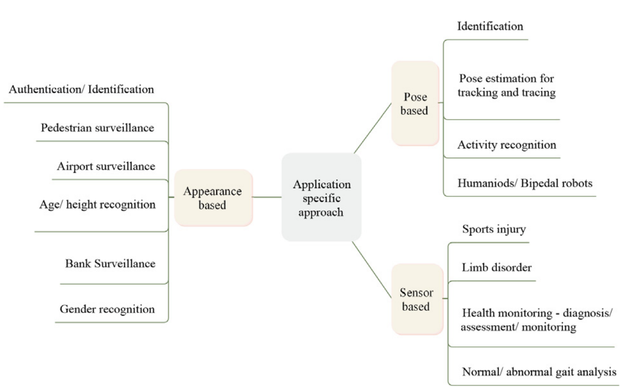

Fig. 19. Application-specific approach based on Appearance, Pose, and Sensor-based.

in the current scenario. Had there been gait surveillance, removing face masks would not be necessary. The unobtrusive method of the gait biometric renders it uniquely suited for surveillance purposes.

5. Challenges and future prospects

The capabilities of GRS have improved over time, and many researchers continue to enhance them. Overcoming significant difficulties involves computational platforms, DNN architectures, data sets, real-time analytics, privacy, security aspects, and aiding to the complexity and covariate conditions. In the following subsections, we brief about these points.

5.1. Covariates

As far as covariates are concerned, there are still shortcomings in the techniques available. The conventional methods are not able to handle covariates effectively. The latter poses various challenges to empower gait-recognition systems to subsume the complexity of covariates and especially significantly impact the gait recognition rate. Among the most prominent adverse effects of covariates on gait recognition is a change in walk style, speed, and spatio-temporal aspects. Our proposed framework (Fig. 17) comprises a fusion of various deep learning models to handle covariates and privacy issues and efficiently utilize various gait datasets to improve the accuracy apart from other useful performance parameters of GRS.

5.2. Deep learning architectures

Deep learning frameworks available perform well in a wide variety of situations to handle various covariate conditions. However, several questions arise about their use. The first is difficulties in making a sound choice of a suitable framework. Next, what do deep learning models specifically learn? Nevertheless, how can adversarial attacks easily deceive the GRS? Lastly, what is a minimal deep neural net configuration that can attain a particular degree of accuracy on a given dataset, and how can the best multiple architectures be fused to lead to an overall improvement?.

5.3. Suitable dataset

Existing biometric datasets for identification (such as CASIA, OU-ISIR, TUM, and USF) still cover a limited portion of all real-time scenarios. There are several situations in which sensors are acquiring data that stops operating, leading to incomplete datasets. Further, people in real-time situations are in millions, e.g., global terrorist tracking. To achieve better biometrics performance, interoperable surveillance systems acquire comprehensive datasets of many people with covariates that should be operational.

5.4. Fog/Edge computing platforms

For a GRS, deep learning models need good computation power and large memory and, thus, are usually deployed on public/private clouds. Most efforts so far have focused on enhancing the performance and accuracy of these models in the cloud, but now there is a need to optimize them to work at Edge or at most in Fog to

Table 3
Acronyms used in this article

Acronym	Full form
[202] CNN	Convolutional Neural Network
[203] DNN	Deep Neural Network
[204] FCNN	Fully Convolutional Neural Network
[204] DCNN	Deep Convolutional Neural Network
[205] SNN	Siamese Neural Network
[206] RNN	Recurrent Neural Network
[207] RCNN	Recurrent Convolutional Neural Network
[208] LSTM	Long Short-Term Memory
[209] GAN	Generative Adversarial Network
[210] AB-GAN	Alpha Blending GAN
[211] VGG	Visual Geometry Group
[212] GEI	Gait Energy Image
[213] SGEI	Skeleton Gait Energy Image
[214] ReLU	Rectified Linear Unit
[215] RGB	Red Green Blue

enable these models to perform well. We found certain algorithms in the literature that reduce the learned size and complexity of the models. Thus, these lightweight architectures can be used on embedded products, providing a CPU/memory-efficient model, and opening new applications.

5.5. Protection, privacy, and anonymity

Protection of the dataset is important in gait identification systems. Adversarial attacks, template attacks, noise attacks, and presentation attacks will harm the datasets and gait recognition systems. Many attempts have been made to provide robust spoofing identification, but not all of them have been reliable. The latest instances of thefts of biometric data have posed privacy issues. Their gait may derive any details regarding the identity/age/gender of the user. Research on authentication schemes is important to ensure the protection of data with visual images.

5.6. Real-time analytics

In real-life scenarios, apart from the accuracy of recognition, responding to intrusions or abnormal situations in a timely manner is critical. For large amounts of video, data uploaded to the cloud would hamper real-time responses even from cloud-based analytics. In such cases, edge/fog analytics have become more relevant and are trending these days. Powerful edge devices with customized accelerators for lightweight deep learning and encryption tasks are thus emerging to handle such demanding situations.

6. Conclusion

Gait biometrics has acquired the center stage in surveillance systems enabled by increased connectivity and data rates complemented by a rise in IT installations in many sensitive areas, including smart cities. It is vital to recognize individuals, in real-time, at many sensitive places like airports, railway stations, bus stations, hospitals, and public areas using prevalent biometrics in such systems. Nonetheless, many barriers have impeded the commercialization of biometric use in every environment. Behavioral gait biometrics has evolved as a potential future security measure. It is unobtrusive, imperceptible, and applicable at a distance, even with low-quality video footage. The method is very robust since it is difficult to imitate the gait of others. This paper sheds light on the deep learning approaches used in handling gait covariates and the datasets employed by various researchers. Compared to earlier surveys, this article has offered a detailed review of covariate-based gait recognition approaches reported from 2015 to 2020 and has explained the terminologies in gait recognition systems. This paper also covered a comprehensive review of all the covariate conditions based on Appearance, Pose, and Sensor-based. Security issues (adversarial attacks) and the methods to overcome them (anonymization methods) are discussed. The problem statement of each targeted paper has been discussed with the novel approach adopted by the authors. Details of more than 50 datasets have been provided, including the available link to access the said datasets. Despite recent breakthroughs in the biometric gait techniques, exploring various covariate approaches further by creating broader and more challenging datasets apart from methods, which help train deep learning models to perform better in real-time scenarios. Covariates conditions need to be handled more efficiently using powerful models and fusion of one or more deep learning pipelines. Such systems can be more effectively deployed, as automatic systems, offering comprehensive, reliable, and secure with enhanced privacy and protection in time to come.

CRediT authorship contribution statement

Anubha Parashar: Conceptualization, Data acquisition, Analysis and interpretation of data, Investigation, Visualization, Writing - original draft. **Rajveer Singh Shekhawat:** Research Supervisor, Reviewing, Editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Acronyms used in this article

The acronyms used in this article are listed in Table 3.

References

[1] M. Rauf, C. Song, Y. Huang, L. Wang, N. Jia, Knowledge transfer between networks and its application on gait recognition, IEEE International Conference on Digital Signal Processing (DSP) 2016 (2016) 492–496.

[2] N. McLaughlin, J. Martinez del Rincon, P. Miller, Recurrent Convolutional Network for Video-Based Person Re-identification, IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2016 (2016) 1325–1334.

[3] Sokolova, A., & Konushin, A. (2017). Gait recognition based on convolutional neural networks. International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, 42

[4] M.J. Marín-Jiménez, F.M. Castro, N. Guil, F. de la Torre, R. Medina-Carnicer, Deep multi-task learning for gait-based biometrics, IEEE International Conference on Image Processing (ICIP) 2017 (2017) 106–110.

[5] P. Nithyakani, A. Shanthini and G. Ponsam, Human Gait Recognition using Deep Convolutional Neural Network, 2019 3rd International Conference on Computing and Communications Technologies (ICCCT), 2019, pp. 208–211.

[6] C. Carley, E. Ristani, C. Tomasi, Person Re-Identification From Gait Using an Autocorrelation Network, IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) 2019 (2019) 2345–2353.

[7] Chunfeng Song, Yongzhen Huang, Yan Huang, Ning Jia, Liang Wang, GaitNet: An End-to-end Network for Gait Based Human Identification, Pattern Recogn. 96 (2019), https://doi.org/10.1016/j.patcog.2019.106988 106988.

[8] H. Arshad, M.A. Khan, M.I. Sharif, M. Yasmin, J.M. Tavares, Y. Zhang, S.C. Satapathy, A multilevel paradigm for deep convolutional neural network features selection with an application to human gait recognition, Expert Syst. (2020).

[9] M. Babaee, L. Li and G. Rigoll, Gait Recognition from Incomplete Gait Cycle, 2018 25th IEEE International Conference on Image Processing (ICIP), 2018, pp. 768–772.

[10] M. Babaee, L. Li, G. Rigoll, Gait Energy Image Reconstruction from Degraded Gait Cycle Using Deep Learning, in: Proceedings of the European Conference on Computer Vision (ECCV) Workshops, 2018.

[11] M. Babaee, Y. Zhu, O. Köpüklü, S. Hörmann, G. Rigoll, Gait energy image restoration using generative adversarial networks, in: 2019 IEEE International Conference on Image Processing (ICIP), IEEE, 2019, pp. 2596–2600.

[12] L.M. Xia, H. Wang, W.T. Guo, Gait recognition based on Wasserstein generating adversarial image inpainting network, J. Central South Univ. 26 (10) (2019) 2759–2770.

[13] M. Babaee, L. Li, G. Rigoll, Person identification from partial gait cycle using fully convolutional neural networks, Neurocomputing 338 (2019) 116–125.

[14] X. Li, Y. Makihara, C. Xu, Y. Yagi, M. Ren, Make the Bag Disappear: Carrying Status-invariant Gait-based Human Age Estimation using Parallel Generative Adversarial Networks, in: 2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS), 2019, pp. 1–9.

[15] Das, D., Agarwal, A., Chattopadhyay, P., & Wang, L. (2019). RGait-NET: An Effective Network for Recovering Missing Information from Occluded Gait Cycles.

[16] Y. He, J. Zhang, H. Shan, L. Wang, Multi-Task GANs for View-Specific Feature Learning in Gait Recognition, IEEE Trans. Inf. Forensics Secur. 14 (1) (Jan. 2019) 102–113.

[17] M.Z. Uddin, D. Muramatsu, N. Takemura, M.A.R. Ahad, Y. Yagi, Spatio-temporal silhouette sequence reconstruction for gait recognition against occlusion, IPSJ Trans. Comput. Vision Appl. 11 (1) (2019) 1–18.

[18] K. Wang, L. Liu, Y. Lee, X. Ding, J. Lin, Nonstandard periodic gait energy image for gait recognition and data augmentation, in: Chinese Conference on Pattern Recognition and Computer Vision (PRCV), Springer, Cham, 2019, pp. 197–208.

[19] T. Yeoh, H.E. Aguirre, K. Tanaka, Clothing-invariant gait recognition using convolutional neural network, in: 2016 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), IEEE, 2016, pp. 1–5.
[20] M. Alotaibi, A. Mahmood, Reduction of Gait Covariate Factors Using Feature Selection and Sparse Dictionary Learning, IEEE International Symposium on Multimedia (ISM) 2016 (2016) 337–340.
[21] F.M. Castro, M.J. Marín-Jiménez, N.G. Mata, N.P. Blanca, Automatic Learning of Gait Signatures for People Identification, IWANN (2017).
[22] S. Tong, Y. Fu, H. Ling, Verification-based pairwise gait identification, in: 2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), IEEE, 2017, pp. 669–673.
[23] T. Yeoh, H.E. Aguirre, K. Tanaka, Clothing-invariant gait recognition using convolutional neural network, International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS) 2016 (2016) 1–5.
[24] Sun, Y., & Liu, Q. (2018). Attribute recognition from clothing using a Faster R-CNN based multitask network. Int. J. Wavelets Multiresolution Inf. Process., 16
[25] H. Ling, J. Wu, P. Li, J. Shen, Attention-Aware Network With Latent Semantic Analysis for Clothing Invariant Gait Recognition, Comput., Mater. Continua (2019).
[26] J. Luo, T. Tjahjadi, View and Clothing Invariant Gait Recognition via 3D Human Semantic Folding, IEEE Access 8 (2020) 100365–100383.
[27] Yan, C., Zhang, B.,& Coenen, F. (2015, October). Multi-attributes gait identification by convolutional neural networks. In 2015 8th International Congress on Image and Signal Processing (CISP) (pp. 642-647). IEEE.
[28] M. Alotaibi, A. Mahmood, Improved gait recognition based on specialized deep convolutional neural network, Comput. Vis. Image Underst. 164 (2017) 103–110.
[29] L. Yao, W. Kusakunniran, Q. Wu, J. Zhang, Z. Tang, Robust CNN-based gait verification and identification using skeleton gait energy image, in: 2018 digital image computing: techniques and applications (DICTA), IEEE, 2018, pp. 1–7.
[30] A.R. Hawas, H.A. El-Khobby, M. Abd-Elnaby, A. El-Samie, E. Fathi, Gait identification by convolutional neural networks and optical flow, Multimedia Tools Appl. 78 (18) (2019) 25873–25888.
[31] F.M. Castro, M.J. Marin-Jimenez, N. Guil, N. Pérez de la Blanca, Multimodal feature fusion for CNN-based gait recognition: an empirical comparison, Neural Comput. Appl. 32 (17) (2020) 14173–14193.
[32] Mehmood, Asif & Khan, Muhammad & Sharif, Muhammad & Khan, Sajid & Shaheen, Muhammad & Saba, Tanzila & Riaz, Naveed & Ashraf, Imran. (2020). Prosperous Human Gait Recognition: An End-to-End System based on Pre-trained CNN Features Selection. Multimedia Tools and Applications.
[33] Yang, Fengjia & Jiang, Xinghao & Sun, Tanfeng & xu, ke. Gait Recognition with Clothing and Carrying Variations Based on GEI and CAPDS Features 2019.
[34] X. Li, Y. Makihara, C. Xu, Y. Yagi, M. Ren, Joint intensity transformer network for gait recognition robust against clothing and carrying status, IEEE Trans. Inf. Forensics Secur. 14 (12) (2019) 3102–3115.
[35] F.M. Castro, N. Guil, M.J. Marín-Jiménez, J. Pérez-Serrano, M. Ujaldón, Energy-based tuning of convolutional neural networks on multi-GPUs, Concurrency and Computation: Practice and Experience 31 (21) (2019) e4786.
[36] X. Li, Y. Makihara, C. Xu, Y. Yagi, M. Ren, Gait recognition via semi-supervised disentangled representation learning to identity and covariate features, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 13309–13319.
[37] Xiang Li, Yasushi Makihara, Chi Xu, Yasushi Yagi, Mingwu Ren, Gait recognition invariant to carried objects using alpha blending generative adversarial networks, Pattern Recogn. 105 (2020) 107376.
[38] S. Yu, GaitGAN: Invariant Gait Feature Extraction Using Generative Adversarial Networks, IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) 2017 (2017) 532–539.
[39] Shiqi Yu et al., GaitGANv2: Invariant Gait Feature Extraction Using Generative Adversarial Networks, Pattern Recogn. (2018).
[40] S. Yu et al., Invariant feature extraction for gait recognition using only one uniform model, Neurocomputing 239 (2017) 81–93.
[41] Liao, R., and others (2017, October). Pose-based temporal-spatial network (PTSN) for gait recognition with carrying and clothing variations. In Chinese conference on biometric recognition (pp. 474–483). Springer, Cham.
[42] I. Cheheb, N. Al-Maadeed, S. Al-Madeed, A. Bouridane, Investigating the use of autoencoders for gait-based person recognition, in: 2018 NASA/ESA Conference on Adaptive Hardware and Systems (AHS), IEEE, 2018, pp. 148–151.
[43] Zhang, Z., and others Gait recognition via disentangled representation learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4710–4719).
[44] Daksh Thapar, Gaurav Jaswal, Aditya Nigam, Chetan Arora, Gait metric learning Siamese network exploiting dual of spatio-temporal 3D-CNN intra and LSTM based inter gait-cycle-segment features, Pattern Recogn. Lett. (2019).
[45] Z. Zhang, L. Tran, F. Liu, X. Liu, On learning disentangled representations for gait recognition, IEEE Trans. Pattern Anal. Mach. Intell. (2020).
[46] W. Swee, C. Dave, and S. Bingquan, "GAIT RECOGNITION FOR PERSON TRACKING ACROSS CAMERA," Comput. Vis. – ECCV 2014 Work., vol. volume 892, 2014.
[47] Shiraga, K., and others (2016, June). Geinet: View-invariant gait recognition using a convolutional neural network. In 2016 international conference on biometrics (ICB) (pp. 1–8). IEEE.
[48] T. Wolf, M. Babaee, G. Rigoll, Multi-view gait recognition using 3D convolutional neural networks, in: 2016 IEEE International Conference on Image Processing (ICIP), IEEE, 2016, pp. 4165–4169.
[49] X. Zhang, S. Sun, C. Li, X. Zhao, and Y. Hu, "DeepGait: A Learning Deep Convolutional Representation for Gait Recognition," in Biometric Recognition. CCBR 2017. Lecture Notes in Computer Science, vol 10568. Springer, Cham, vol. 1, no. c, 2017, pp. 447–456.
[50] Li, C., Min, and others (2017). DeepGait: A learning deep convolutional representation for view-invariant gait recognition using joint Bayesian. Appl. Sci., 7(3), 210.
[51] N. Jia, V. Sanchez, C.T. Li, Learning optimised representations for view-invariant gait recognition, in: 2017 IEEE International Joint Conference on Biometrics (IJCB), IEEE, 2017, pp. 774–780.
[52] D. Thapar, A. Nigam, D. Aggarwal, P. Agarwal, VGR-net: A view invariant gait recognition network, in: 2018 IEEE 4th international conference on identity, security, and behavior analysis (ISBA), IEEE, 2018, pp. 1–8.
[53] P. Zhang, Q. Wu, J. Xu, in: VN-GAN: identity-preserved variation normalizing GAN for gait recognition. In 2019 international joint conference on neural networks (IJCNN), IEEE, 2019, pp. 1–8.
[54] P. Zhang, Q. Wu, J. Xu, in: VT-GAN: View transformation GAN for gait recognition across views. In 2019 International Joint Conference on Neural Networks (IJCNN), IEEE, 2019, pp. 1–8.
[55] M. Deng et al., Human gait recognition based on deterministic learning and knowledge fusion through multiple walking views, J. Franklin Inst. 357 (4) (2020) 2471–2491.
[56] Y. Wang, C. Song, Y. Huang, Z. Wang, L. Wang, Learning view invariant gait features with Two-Stream GAN, Neurocomputing 339 (2019) 245–254.
[57] Z. Wu et al., A comprehensive study on cross-view gait based human identification with deep cnns, IEEE Trans. Pattern Anal. Mach. Intell. 39 (2) (2016) 209–226.
[58] Q. Chen, Feature map pooling for cross-view gait recognition based on silhouette sequence images, IEEE International Joint Conference on Biometrics (IJCB) 2017 (2017) 54–61.
[59] Y. Huang, J. Zhang, H. Zhao, L. Zhang, in: Attention-based network for cross-view gait recognition. In International Conference on Neural Information Processing, Springer, Cham, 2018, pp. 489–498.
[60] Hu, B., Gao, Y., Guan, Y., Long, Y., Lane, N., & Plötz, T. (2018). Robust Cross-View Gait Identification with Evidence: A Discriminant Gait GAN (DiGGAN) Approach on 10000 People. ArXiv, abs/1811.10493.
[61] S. Luo, S. Feng, H. Pan, J. Yin, X. Zhang, in: A sequence-based multi-scale network for cross-view gait recognition. In 2019 6th International Conference on Systems and Informatics (ICSAI), IEEE, 2019, pp. 1179–1183.
[62] Linda, G. & Govindarajan, Themozhi & Bandi, Sudheer. (2019). Color-Mapped Contour Gait Image for Cross-View Gait Recognition Using Deep Convolution Neural Network. International Journal of Wavelets, Multiresolution and Information Processing.
[63] X. Ben, C. Gong, P. Zhang, R. Yan, Q. Wu, W. Meng, Coupled bilinear discriminant projection for cross-view gait recognition, IEEE Trans. Circuits Syst. Video Technol. 30 (3) (2019) 734–747.
[64] S.B. Tong, Y.Z. Fu, H.F. Ling, Cross-view gait recognition based on a restrictive triplet network, Pattern Recogn. Lett. 125 (2019) 212–219.
[65] Y. Zhang, Y. Huang, S. Yu, L. Wang, Cross-view gait recognition by discriminative feature learning, IEEE Trans. Image Process. 29 (2019) 1001–1015.
[66] X. Wang, J. Zhang, W.Q. Yan, Gait recognition using multichannel convolution neural networks, Neural Comput. Appl. 32 (18) (2020) 14275–14285.
[67] Chao, H., He, Y., Zhang, J., & Feng, J. (2019, July). Gaitset: Regarding gait as a set for cross-view gait recognition. In Proceedings of the AAAI conference on artificial intelligence (Vol. 33, No. 01, pp. 8126–8133).
[68] K. Zhang, W. Luo, L. Ma, W. Liu, H. Li, Learning joint gait representation via quintuplet loss minimization, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 4700–4709.
[69] N. Takemura, Y. Makihara, D. Muramatsu, T. Echigo, Y. Yagi, On Input/Output Architectures for Convolutional Neural Network-Based Cross-View Gait Recognition, IEEE Trans. Circuits Syst. Video Technol. 29 (9) (Sept. 2019) 2708–2719, https://doi.org/10.1109/TCSVT.2017.2760835.
[70] R. Zhang, D. Yin, Z. Zhou, Z. Cao, F. Meng, and B. Hu, "Improving Cross-View Gait Recognition With Generative Adversarial Networks Rui," in 2019 Scientific Conference on Network, Power Systems and Computing (NPSC 2019), 2019, vol. 3, no. Npsc, pp. 43–47.
[71] Y. Zhang, Y. Huang, L. Wang, S. Yu, A comprehensive study on gait biometrics using a joint CNN-based method, Pattern Recogn. 93 (2019) 228–236.
[72] Xiuhui Wang, Weiqi Yan, Human Gait Recognition Based on Frame-by-Frame Gait Energy Images and Convolutional Long Short Term Memory, Int. J. Neural Syst. (2019).
[73] Muhammad Khan et al., A non-linear view transformations model for cross-view gait recognition, Neurocomputing 402 (2020).
[74] C. Xu, Y. Makihara, X. Li, Y. Yagi, J. Lu, Cross-view gait recognition using pairwise spatial transformer networks, IEEE Trans. Circuits Syst. Video Technol. 31 (1) (2020) 260–274.
[75] C. Cai, Y. Zhou, Y. Wang, October). CHD: Consecutive Horizontal Dropout for Human Gait Feature Extraction, in: Proceedings of the 2019 8th International Conference on Computing and Pattern Recognition, 2019, pp. 89–94.
[76] Y. Feng, Y. Li, J. Luo, in: Learning effective gait features using LSTM. In 2016 23rd international conference on pattern recognition (ICPR), IEEE, 2016, pp. 325–330.

[77] A. Haque, A. Alahi, L. Fei-Fei, Recurrent attention models for depth-based person identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1229–1238.
[78] Uddin, M.Z., Khaksar, W., & Torresen, J. (2017, November). A robust gait recognition system using spatiotemporal features and deep learning. In 2017 IEEE international conference on multisensor fusion and integration for intelligent systems (MFI) (pp. 156–161). IEEE.
[79] C. Li, P. Wang, S. Wang, Y. Hou, W. Li, in: Skeleton-based action recognition using LSTM and CNN. In 2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), IEEE, 2017, pp. 585–590.
[80] Castro, F.M., and others. Evaluation of CNN architectures for gait recognition based on optical flow maps. In 2017 International Conference of the Biometrics Special Interest Group (pp. 1–5). IEEE.
[81] O. Costilla-Reyes, R. Vera-Rodriguez, P. Scully, K.B. Ozanyan, Analysis of spatio-temporal representations for robust footstep recognition with deep residual neural networks, IEEE Trans. Pattern Anal. Mach. Intell. 41 (2) (2018) 285–296.
[82] Karianakis, N., Liu, Z., Chen, Y., & Soatto, S. (2017). Person depth reid: Robust person re-identification with commodity depth sensors. arXiv preprint arXiv:1705.09882
[83] G. Batchuluun, H.S. Yoon, J.K. Kang, K.R. Park, Gait-Based Human Identification by Combining Shallow Convolutional Neural Network-Stacked Long Short-Term Memory and Deep Convolutional Neural Network, IEEE Access 6 (2018) 63164–63186.
[84] W. Liu, C. Zhang, H. Ma, S. Li, Learning efficient spatial-temporal gait features with deep learning for human identification, Neuroinformatics 16 (3) (2018) 457–471.
[85] S. Tong, Y. Fu, X. Yue, H. Ling, Multi-view gait recognition based on a spatial-temporal deep neural network, IEEE Access 6 (2018) 57583–57596.
[86] Karianakis, and others (2018). Reinforced Temporal Attention and Split-Rate Transfer for Depth-Based Person Re-identification: 15th European Conference, Munich, Germany, September 8–14, 2018.
[87] Francesco Battistone, Alfredo Petrosino, TGLSTM: a Time based Graph Deep Learning Approach to Gait Recognition, Pattern Recogn. Lett. 126 (2018).
[88] Wu, X., and others (2019, November). Spatial-temporal graph attention network for video-based gait recognition. In Asian Conference on Pattern Recognition (pp. 274–286). Springer, Cham.
[89] T. Huynh-The, C.H. Hua, N.A. Tu, D.S. Kim, Learning 3D spatiotemporal gait feature by convolutional network for person identification, Neurocomputing 397 (2020) 192–202.
[90] X. Wu, T. Yang, Z. Xia, Gait Recognition Based on Densenet Transfer Learning, Ijset. Net 9 (1) (2015) 1–14.
[91] C. Fan et al., GaitPart: Temporal Part-Based Model for Gait Recognition, IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2020) 14213–14221.
[92] X. Liu, J. Liu, Gait Recognition Method of Underground Coal Mine Personnel Based on Densely Connected Convolution Network and Stacked Convolutional Autoencoder, Entropy 22 (2020).
[93] R. Delgado-Escaño et al., MuPeG-The Multiple Person Gait Framework, Sensors 20 (2020).
[94] X. Chen, J. Weng, W. Lu, J. Xu, Multi-gait recognition based on attribute discovery, IEEE Trans. Pattern Anal. Mach. Intell. 40 (7) (2017) 1697–1710.
[95] C. Zhu, X.C. Yin, Effective human detection via multi-model classification and adaptive late fusion, Int. J. Wavelets Multiresolut. Inf. Process. 16 (02) (2018) 1840012.
[96] A. Sakai et al., Gait Recognition Based on Constrained Mutual Subspace Method with CNN Features, in: 16th International Conference on Machine Vision Applications, 2019.
[97] Z. Xu, W. Lu, Q. Zhang, Y. Yeung, X. Chen, Gait recognition based on capsule network, J. Vis. Commun. Image Represent 59 (2019) 159–167.
[98] Y. Makihara, D. Adachi, C. Xu, Y. Yagi, Gait recognition by deformable registration, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 561–571.
[99] L. Zheng, Z. Bie, Y. Sun, J. Wang, C. Su, S. Wang, Q. Tian, Mars: A video benchmark for large-scale person re-identification, in: European conference on computer vision, Springer, Cham, 2016, pp. 868–884.
[100] M. Jia, H. Yang, D. Huang, Y. Wang, October). Attacking gait recognition systems via silhouette guided GANs, in: Proceedings of the 27th ACM International Conference on Multimedia, 2019, pp. 638–646.
[101] L. Wang, T.K. Kim, K.J. Yoon, Eventsr: From asynchronous events to image reconstruction, restoration, and super-resolution via end-to-end adversarial learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 8315–8325.
[102] He, Z., Wang, W., Dong, J., & Tan, T. (2020). Temporal sparse adversarial attack on gait recognition. arXiv e-prints, arXiv-2002.
[103] W. Xue, H. Ai, T. Sun, C. Song, Y. Huang, L. Wang, Frame-GAN: Increasing the frame rate of gait videos with generative adversarial networks, Neurocomputing 380 (2020) 95–104.
[104] N.D.T. Tieu, H.H. Nguyen, H.Q. Nguyen-Son, J. Yamagishi, I. Echizen, in: An approach for gait anonymization using deep learning. In 2017 IEEE Workshop on Information Forensics and Security (WIFS), IEEE, 2017, pp. 1–6.
[105] W. An, R. Liao, S. Yu, Y. Huang, P.C. Yuen, in: Improving gait recognition with 3d pose estimation. In Chinese Conference on Biometric Recognition, Springer, Cham, 2018, pp. 137–147.
[106] R. Liao, S. Yu, W. An, Y. Huang, A model-based gait recognition method with body pose and human prior knowledge, Pattern Recogn. 98 (2020) 107069.
[107] A. Sokolova, A. Konushin, Pose-based deep gait recognition, IET Biometrics 8 (2) (2019) 134–143.
[108] H.L. Tavares, J.B.C. Neto, J.P. Papa, D. Colombo, A.N. Marana, Tracking and re-identification of people using soft-biometrics, in: 2019 XV Workshop de Visão Computacional (WVC), IEEE, 2019, pp. 78–83.
[109] J. Luo, T. Tjahjadi, Gait Recognition and Understanding Based on Hierarchical Temporal Memory Using 3D Gait Semantic Folding, Sensors (Basel, Switzerland), 2020, p. 20.
[110] M.M. Hasan, H.A. Mustafa, Multi-level feature fusion for robust pose-based gait recognition using RNN, Int. J. Comput. Sci. Inf. Secur. (IJCSIS) 18 (1) (2020).
[111] W. Sheng, X. Li, Siamese denoising autoencoders for joints trajectories reconstruction and robust gait recognition, Neurocomputing 395 (2020) 86–94.
[112] Li, Na & Zhao, Xinbo & Ma, Chong. (2020). A model-based Gait Recognition Method based on Gait Graph Convolutional Networks and Joints Relationship Pyramid Mapping.
[113] K. Jun, D.W. Lee, K. Lee, S. Lee, M.S. Kim, Feature extraction using an RNN autoencoder for skeleton-based abnormal gait recognition, IEEE Access 8 (2020) 19196–19207.
[114] J. Hannink, T. Kautz, C.F. Pasluosta, K.-G. Gasmann, J. Klucken, B.M. Eskofier, Sensor-Based Gait Parameter Extraction With Deep Convolutional Neural Networks, IEEE J. Biomed. Heal. Informatics 21 (1) (Jan. 2017) 85–93.
[115] V. Prabhu and J. Whaley, "Vulnerability of deep learning-based gait biometric recognition to adversarial perturbations," CVPR 2017 CV-COPS Work., vol. 1, no. 2, 2017.
[116] B. Gálai and C. Benedek, "Gait Recognition with Compact Lidar Sensors," in Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, 2017, vol. 6, pp. 426–432.
[117] K.-T. Nguyen, T.-L. Vo-Tran, D.-T. Dinh, and M.-T. Tran, "Gait Recognition with Multi-region Size Convolutional Neural Network for Authentication with Wearable Sensors," in Lecture Notes in Computer Science, vol. 10646 LNCS, 2017, pp. 197–212.
[118] O. Dehzangi, M. Taherisadr, R. ChangalVala, IMU-Based Gait Recognition Using Convolutional Neural Networks and Multi-Sensor Fusion, Sensors 17 (12) (Nov. 2017) 2735.
[119] G. Giorgi, F. Martinelli, A. Saracino, and M. Sheikhalishahi, "Try Walking in My Shoes, if You Can: Accurate Gait Recognition Through Deep Learning," in Lecture Notes in Computer Science, vol. 10489 LNCS, 2017, pp. 384–395.
[120] Y. Zhao, S. Zhou, Wearable Device-Based Gait Recognition Using Angle Embedded Gait Dynamic Images and a Convolutional Neural Network, Sensors 17 (3) (Feb. 2017) 478.
[121] X. Zhang, L. Yao, C. Huang, T. Gu, Z. Yang, Y. Liu, DeepKey: A Multimodal Biometric Authentication System via Deep Decoding Gaits and Brainwaves, ACM Trans. Intell. Syst. Technol. 11 (4) (Jul. 2020) 1–24.
[122] S. Abdulatif, F. Aziz, K. Armanious, B. Kleiner, B. Yang, and U. Schneider, "A Study of Human Body Characteristics Effect on Micro-Doppler-Based Person Identification using Deep Learning," Arxiv, no. November, pp. 1–6, Apr. 2018.
[123] W. Zeng, J. Chen, C. Yuan, F. Liu, Q. Wang, and Y. Wang, "Accelerometer-based gait recognition via deterministic learning," in 2018 Chinese Control And Decision Conference (CCDC), Jun. 2018, no. September, pp. 6280–6285.
[124] W. Yuan, L. Zhang, "Gait Classification and Identity Authentication Using CNN," in 18th Asia Simulation Conference, AsiaSim 2018, Kyoto, Japan 946 (2018) 119–128.
[125] M. Gadaleta and M. Rossi, "IDNet: Smartphone-based gait recognition with convolutional neural networks," Pattern Recognit., vol. 74, pp. 25–37.
[126] S. Abdulatif, Q. Wei, F. Aziz, B. Kleiner, and U. Schneider, "Micro-doppler based human-robot classification using ensemble and deep learning approaches," in 2018 IEEE Radar Conference (RadarConf18), Apr. 2018, pp. 1043–1048.
[127] G. Giorgi, F. Martinelli, A. Saracino, M. Sheikhalishahi, Walking Through the Deep: Gait Analysis for User Authentication Through Deep Learning, in: IFIP International Conference on ICT Systems Security and Privacy Protection, 2018, pp. 62–76.
[128] J. Gao, P. Gu, Q. Ren, J. Zhang, X. Song, Abnormal Gait Recognition Algorithm Based on LSTM-CNN Fusion Network, IEEE Access 7 (2019) 163180–163190.
[129] R. Delgado-Escano, F.M. Castro, J.R. Cozar, M.J. Marin-Jimenez, N. Guil, An End-to-End Multi-Task and Fusion CNN for Inertial-Based Gait Recognition, IEEE Access 7 (2019) 1897–1908.
[130] D. Jung et al., "Deep Neural Network-Based Gait Classification Using Wearable Inertial Sensor Data," in 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jul. 2019, pp. 3624–3628.
[131] Y. Wang et al., "EV-Gait: Event-Based Robust Gait Recognition Using Dynamic Vision Sensors," in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2019, vol. 2019-June, pp. 6351–6360.
[132] P. Kumar, S. Mukherjee, R. Saini, P. Kaushik, P.P. Roy, D.P. Dogra, Multimodal Gait Recognition With Inertial Sensor Data and Video Using Evolutionary Algorithm, IEEE Trans. Fuzzy Syst. 27 (5) (May 2019) 956–965.
[133] S.Z. Gurbuz, M.G. Amin, Radar-Based Human-Motion Recognition With Deep Learning: Promising Applications for Indoor Monitoring, IEEE Signal Process. Mag. 36 (4) (Jul. 2019) 16–28.
[134] Fernandez-Lopez, Liu-Jimenez, Kiyokawa, and Wu, "Recurrent Neural Network for Inertial Gait User Recognition in Smartphones," Sensors, vol. 19, no. 18, p. 4054, Sep. 2019.

[135] D. Kim, M. Kim, J. Kwon, Y.-L. Park, S. Jo, Semi-Supervised Gait Generation With Two Microfluidic Soft Sensors, IEEE Robot. Autom. Lett. 4 (3) (Jul. 2019) 2501–2507.
[136] A. Turner, S. Hayes, The Classification of Minor Gait Alterations Using Wearable Sensors and Deep Learning, IEEE Trans. Biomed. Eng. 66 (11) (Nov. 2019) 3136–3145.
[137] N. Sadeghzadehyazdi, T. Batabyal, N.K. Dhar, B.O. Familoni, K.M. Iftekharuddin, and S.T. Acton, "GlidarCo: gait recognition by 3D skeleton estimation and biometric feature correction of flash lidar data," pp. 1–11, May 2019.
[138] R. Delgado-Escaño, F.M. Castro, J.R. Cózar, M.J. Marín-Jiménez, N. Guil, and E. Casilari, "A cross-dataset deep learning-based classifier for people fall detection and identification," Comput. Methods Programs Biomed., vol. 184, no. December, p. 105265, Feb. 2020.
[139] L. Tran, D. Choi, Data Augmentation for Inertial Sensor-Based Gait Deep Neural Network, IEEE Access 8 (2020) 12364–12378.
[140] Q. Zou, Y. Wang, Q. Wang, Y. Zhao, Q. Li, Deep Learning-Based Gait Recognition Using Smartphones in the Wild, IEEE Trans. Inf. Forensics Secur. 15 (2020) 1.
[141] M.-C. Lee, Y. Huang, J.J.-C. Ying, C. Chen, and V.S. Tseng, ''DeepIdentifier: A Deep Learning-Based Lightweight Approach for User Identity Recognition," in Lecture Notes in Computer Science, vol. 11888 LNAI, no. March, Springer International Publishing, 2019, pp. 389–405.
[142] P. Terrier, Gait Recognition via Deep Learning of the Center-of-Pressure Trajectory, Appl. Sci. 10 (3) (Jan. 2020) 774.
[143] L. Chen, Y. Zhang, L. Peng, "METIER: A Deep Multi-Task Learning Based Activity and User Recognition Model Using Wearable Sensors," Proc, ACM Interactive, Mobile, Wearable Ubiquitous Technol. 4 (1) (Mar. 2020) 1–18.
[144] J. Luo, T. Tjahjadi, Multi-Set Canonical Correlation Analysis for 3D Abnormal Gait Behaviour Recognition Based on Virtual Sample Generation, IEEE Access 8 (2020) 32485–32501.
[145] S. Kitic, G. Puy, P. Perez, and P. Gilberton, "Scattering features for multimodal gait recognition," in 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Nov. 2017, vol. 2018-Janua, pp. 843–847.
[146] S. Davarzani et al., Closing the Wearable Gap-Part VI: Human Gait Recognition Using Deep Learning Methodologies, Electronics 9 (5) (May 2020) 796.
[147] P. Limcharoen, N. Khamsemanan, C. Nattee, View-Independent Gait Recognition Using Joint Replacement Coordinates (JRCs) and Convolutional Neural Network, IEEE Trans. Inf. Forensics Secur. 15 (2020) 3430–3442.
[148] Z. Meng et al., "Gait Recognition for Co-existing Multiple People Using Millimeter Wave Sensing," in Proceedings of the AAAI Conference on Artificial Intelligence, 2020, no. Vol 34 No 01.
[149] P. Zhao et al., "mID: Tracking and Identifying People with Millimeter Wave Radar," in 2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS), May 2019, pp. 33–40.
[150] M. Gadaleta, L. Merelli, and M. Rossi, "Human authentication from ankle motion data using convolutional neural networks," in 2016 IEEE Statistical Signal Processing Workshop (SSP), Jun. 2016, pp. 1–5.
[151] Y. Makihara et al., The OU-ISIR Gait Database Comprising the Treadmill Dataset, IPSJ Trans. Comput. Vis. Appl. 4 (2012) 53–62, https://doi.org/10.2197/ipsjtcva.4.53.
[152] H. Iwama, M. Okumura, Y. Makihara, Y. Yagi, The OU-ISIR Gait Database Comprising the Large Population Dataset and Performance Evaluation of Gait Recognition, IEEE Trans. Inf. Forensics Secur. 7 (5) (Oct. 2012) 1511–1521.
[153] C. Xu, Y. Makihara, G. Ogi, X. Li, Y. Yagi, J. Lu, The OU-ISIR Gait Database comprising the Large Population Dataset with Age and performance evaluation of age estimation, IPSJ Trans. Comput. Vis. Appl. 9 (1) (Dec. 2017) 24.
[154] M.Z. Uddin et al., The OU-ISIR Large Population Gait Database with real-life carried object and its performance evaluation, IPSJ Trans. Comput. Vis. Appl. 10 (1) (Dec. 2018) 5.
[155] N. Takemura, Y. Makihara, D. Muramatsu, T. Echigo, Y. Yagi, Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition, IPSJ Trans. Comput. Vis. Appl. 10 (1) (Dec. 2018) 4.
[156] W. An et al., Performance Evaluation of Model-based Gait on Multi-view Very Large Population Database with Pose Sequences, IEEE Trans. Biometrics, Behav. Identity Sci. (2020) 1.
[157] T.T. Ngo, Y. Makihara, H. Nagahara, Y. Mukaigawa, Y. Yagi, The largest inertial sensor-based gait database and performance evaluation of gait-based personal authentication, Pattern Recognit. 47 (1) (Jan. 2014) 228–237.
[158] Liang Wang, Tieniu Tan, Huazhong Ning, Hu. Weiming, Silhouette analysis-based gait recognition for human identification, IEEE Trans. Pattern Anal. Mach. Intell. 25 (12) (Dec. 2003) 1505–1518.
[159] Shiqi Yu, Daoliang Tan, and Tieniu Tan, "A Framework for Evaluating the Effect of View Angle, Clothing and Carrying Condition on Gait Recognition," in 18th International Conference on Pattern Recognition (ICPR'06), 2006, vol. 4, pp. 441–444.
[160] Daoliang Tan, Kaiqi Huang, Shiqi Yu, and Tieniu Tan, "Efficient Night Gait Recognition Based on Template Matching," in 18th International Conference on Pattern Recognition (ICPR'06), 2006, vol. 3, pp. 1000–1003.
[161] A. Roy, S. Sural, J. Mukherjee, G. Rigoll, Occlusion detection and gait silhouette reconstruction from degraded scenes, Signal, Image Video Process. 5 (4) (2011) 415–430.
[162] M. Hofmann, J. Geiger, S. Bachmann, B. Schuller, G.R. Institute, The TUM gait from audio, image and depth (GAID) database, J. Vis. Commun. Image Represent 25 (1) (2014) 195–206.
[163] S. Sarkar, P.J. Phillips, Z. Liu, I.R. Vega, P. Grother, K.W. Bowyer, The humanID gait challenge problem: data sets, performance, and analysis, IEEE Trans. Pattern Anal. Mach. Intell. 27 (2) (Feb. 2005) 162–177.
[164] R. Gross and J. Cohn, "The CMU Motion of Body (MoBo) Dataset," Cmu-Ri-Tr-01-18, no. June, pp. 1–11, 2001.
[165] Y. Iwashita, R. Baba, K. Ogawara, and R. Kurazume, "Person Identification from Spatio-temporal 3D Gait," in 2010 International Conference on Emerging Security Technologies, Sep. 2010, pp. 30–35.
[166] C. Ionescu, D. Papava, V. Olaru, C. Sminchisescu, Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments, IEEE Trans. Pattern Anal. Mach. Intell. 36 (7) (Jul. 2014) 1325–1339.
[167] A. Shahroudy, J. Liu, T.-T. Ng, and G. Wang, "NTU RGB+D: A Large Scale Dataset for 3D Human Activity Analysis," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2016, vol. 2016-Decem, pp. 1010–1019.
[168] J. Liu, A. Shahroudy, M.L. Perez, G. Wang, L.-Y. Duan, and A. Kot Chichung, "NTU RGB+D 120: A Large-Scale Benchmark for 3D Human Activity Understanding," IEEE Trans. Pattern Anal. Mach. Intell., pp. 1–1, May 2019, doi: 10.1109/TPAMI.2019.2916873.
[169] M. Munaro, A. Fossati, A. Basso, E. Menegatti, L. Van Gool, One-Shot Person Re-identification with a Consumer Depth Camera, in: S. Gong, M. Cristani, S. Yan, C.C. Loy (Eds.), Person Re-Identification, vol. 56, Springer, London, London, 2014, pp. 161–181.
[170] I.B. Barbosa, M. Cristani, A. Del Bue, L. Bazzani, and V. Murino, "Re-identification with RGB-D Sensors," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 7583 LNCS, no. PART 1, 2012, pp. 433–442.
[171] E. Cippitelli, E. Gambi, and S. Spinsante, ''Human Action Recognition with RGB-D Sensors," in Motion Tracking and Gesture Recognition, InTech, 2017
[172] Jaeyong Sung, C. Ponce, B. Selman, and A. Saxena, "Unstructured human activity detection from RGBD images," in 2012 IEEE International Conference on Robotics and Automation, May 2012, pp. 842–849.
[173] D. Kastaniotis, I. Theodorakopoulos, C. Theoharatos, G. Economou, and S. Fotopoulos, "A framework for gait-based recognition using Kinect," Pattern Recognit. Lett., vol. 68, pp. 327–335, Dec. 201.
[174] D. Kastaniotis, I. Theodorakopoulos, G. Economou, and S. Fotopoulos, "Gait-based gender recognition using pose information for real time applications," in 2013 18th International Conference on Digital Signal Processing (DSP), Jul. 2013, pp. 1–6.
[175] Y. Wang, J. Sun, J. Li, and D. Zhao, "Gait recognition based on 3D skeleton joints captured by kinect," in 2016 IEEE International Conference on Image Processing (ICIP), Sep. 2016, vol. 2016-Augus, pp. 3151–3155.
[176] T. Wang, S. Gong, X. Zhu, and S. Wang, "Person Re-identification by Video Ranking," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 8692 LNCS, no. PART 4, 2014, pp. 688–703.
[177] M. Hirzer, & others, "Person Re-identification by Descriptive and Discriminative Classification," in Lecture Notes in Computer Science, vol. 6688 LNCS, 2011, pp. 91–102.
[178] Y.-R. Li, S. Yu, and S. Wu, "Pedestrian detection in depth images using framelet regularization," in 2012 IEEE International Conference on Computer Science and Automation Engineering (CSAE), May 2012, vol. 2, pp. 300–303.
[179] P.L. Mazzeo, P. Spagnolo, T.B. Moeslund, Activity Monitoring by Multiple Distributed Sensing, vol. 8703, Springer International Publishing, Cham, 2014.
[180] R. Vemulapalli, F. Arrate, and R. Chellappa, "Human Action Recognition by Representing 3D Skeletons as Points in a Lie Group," in 2014 IEEE Conference on Computer Vision and Pattern Recognition, Jun. 2014, pp. 588–595.
[181] E. Mueggler, H. Rebecq, G. Gallego, T. Delbruck, D. Scaramuzza, The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and SLAM, Int. J. Rob. Res. 36 (2) (Feb. 2017) 142–149.
[182] P. Dollar et al., Pedestrian Detection: An Evaluation of the State of the Art, IEEE Trans. Pattern Anal. Mach. Intell., Apr. 34 (4) (2012) 743–761.
[183] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark," in 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Jul. 2017, vol. 2017-July, pp. 3226–3229.
[184] A. Ess, B. Leibe, K. Schindler, and L. Van Gool, "A mobile vision system for robust multi-person tracking," in 2008 IEEE Conference on Computer Vision and Pattern Recognition, Jun. 2008, pp. 1–8.
[185] J. Barth et al., Stride Segmentation during Free Walk Movements Using Multi-Dimensional Subsequence Dynamic Time Warping on Inertial Sensor Data, Sensors 15 (3) (Mar. 2015) 6419–6440.
[186] I.J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and Harnessing Adversarial Examples," 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc., pp. 1–11, Dec. 2014.
[187] C. Benedek, B. Galai, B. Nagy, Z. Janko, Lidar-Based Gait Analysis and Activity Recognition in a 4D Surveillance System, IEEE Trans. Circuits Syst. Video Technol. 28 (1) (Jan. 2018) 101–113.
[188] Yuting Zhang, Gang Pan, Kui Jia, Lu. Minlong, Yueming Wang, Wu. Zhaohui, Accelerometer-Based Gait Recognition by Sparse Representation of Signature Points With Clusters, IEEE Trans. Cybern. 45 (9) (Sep. 2015) 1864–1875.
[189] J. Frank, S. Mannor, J. Pineau, D. Precup, Time Series Analysis Using Geometric Template Matching, IEEE Trans. Pattern Anal. Mach. Intell. 35 (3) (Mar. 2013) 740–754.

[190] S. Keshishzadeh, A. Fallah, and S. Rashidi, "Improved EEG based human authentication system on large dataset," in 2016 24th Iranian Conference on Electrical Engineering (ICEE), May 2016, pp. 1165–1169.
[191] G. Gallego et al., "Event-based Vision: A Survey," IEEE Trans. Pattern Anal. Mach. Intell., pp. 1–1, 2020, doi: 10.1109/TPAMI.2020.3008413.
[192] B. Kwolek et al., Calibrated and synchronized multi-view video and motion capture dataset for evaluation of gait recognition, Multimed. Tools Appl., Nov. 78 (22) (2019) 32437–32465.
[193] R. Horaud, & others, "An overview of depth cameras and range scanners based on time-of-flight technologies," Mach. Vis. Appl., vol. 27, no. 7, pp. 1005–1020, Oct. 2016.
[194] A. Sucerquia, & others "SisFall: A Fall and Movement Dataset," Sensors, vol. 17, no. 12, p. 198, Jan. 2017.
[195] D. Micucci, M. Mobilio, P. Napoletano, UniMiB SHAR: A Dataset for Human Activity Recognition Using Acceleration Data from Smartphones, Appl. Sci. 7 (10) (Oct. 2017) 1101.
[196] A. Özdemir, An Analysis on Sensor Locations of the Human Body for Wearable Fall Detection Devices: Principles and Practice, Sensors 16 (8) (Jul. 2016) 1161.
[197] T. Hoang, D. Choi, T. Nguyen, On the Instability of Sensor Orientation in Gait Verification on Mobile Phone, in: Proceedings of the 12th International Conference on Security and Cryptography, 2015, pp. 148–159.
[198] N. Khamsemanan, C. Nattee, N. Jianwattanapaisarn, Human Identification From Freestyle Walks Using Posture-Based Gait Feature, IEEE Trans. Inf. Forensics Secur. 13 (1) (Jan. 2018) 119–128, https://doi.org/10.1109/TIFS.2017.2738611.
[199] B. Bhanu, V. Govindaraju, Multibiometrics for Human Identification, vol. 9780521115, Cambridge University Press, Cambridge, 2011.
[200] J.B. Hayfron-Acquah, M.S. Nixon, J.N. Carter, Automatic gait recognition by symmetry analysis, Pattern Recognit. Lett. 24 (13) (Sep. 2003) 2175–2183.
[201] Y. Zhang, Y. Huang, L. Wang, S. Yu, A comprehensive study on gait biometrics using a joint cnn-based method, Pattern Recogn. 93 (September 2019) 228–236.
[202] Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, L.D. Jackel, Backpropagation applied to handwritten zip code recognition, Neural Comput. 1 (4) (1989) 541–551.
[203] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.
[204] Lin, M., Chen, Q., & Yan, S. (2014). Network In Network. CoRR, abs/1312.4400.
[205] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, R. Shah, Signature verification using a siamese time delay neural network, in: Advances in neural information processing systems, 1993, p. 6.
[206] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning representations by back-propagating errors, Nature 323 (6088) (1986) 533–536.
[207] M. Liang, X. Hu, Recurrent convolutional neural network for object recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3367–3375.
[208] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.
[209] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, Y. Bengio, Generative adversarial nets, in: Advances in neural information processing systems, 2014, p. 27.
[210] Kwak, H., & Zhang, B.T. (2016). Generating images part by part with composite generative adversarial networks. arXiv preprint arXiv:1607.05387.
[211] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
[212] J. Han, B. Bhanu, Individual recognition using gait energy image, IEEE Trans. Pattern Anal. Mach. Intell. 28 (2) (2005) 316–322.
[213] Luo, J., Zhang, J., Zi, C., Niu, Y., Tian, H., & Xiu, C. (2015). Gait recognition using GEI and AFDEI. International Journal of Optics, 2015.
[214] Xu, B., Wang, N., Chen, T., & Li, M. (2015). Empirical evaluation of rectified activations in convolutional network. arXiv preprint arXiv:1505.00853.
[215] J.C. Maxwell, 1. On the equilibrium of elastic solids, Proc. R. Soc. Edinb. 2 (1851) 294–296.
[216] P. Connor, A. Ross, Biometric recognition by gait: A survey of modalities and features, Computer vision and image understanding 167 (2018) 1–27.
[217] A.S. Alharthi, S.U. Yunas, K.B. Ozanyan, Deep learning for monitoring of human gait: A review, IEEE Sens. J. 19 (21) (2019) 9575–9591.
[218] P. Connor, A. Ross, Biometric recognition by gait: A survey of modalities and features, Computer vision and image understanding 167 (2018) 1–27.
[219] C. Prakash, R. Kumar, N. Mittal, Recent developments in human gait research: parameters, approaches, applications, machine learning techniques, datasets and challenges, Artif. Intell. Rev. 49 (1) (2018) 1–40.
[220] M. Mao, Y. Song, Gait Recognition Based on 3D Skeleton Data and Graph Convolutional Network, IEEE International Joint Conference on Biometrics (IJCB) 2020 (2020) 1–8, https://doi.org/10.1109/IJCB48548.2020.9304916.
[221] J.P. Singh, S. Jain, S. Arora, U.P. Singh, A survey of behavioral biometric gait recognition: Current success and future perspectives, Arch. Comput. Methods Eng. 28 (1) (2021) 107–148.
[222] M. Mao, Y. Song, Gait Recognition Based on 3D Skeleton Data and Graph Convolutional Network, IEEE International Joint Conference on Biometrics (IJCB) 2020 (2020) 1–8.
[223] T. Teepe, A. Khan, J. Gilg, F. Herzog, S. Hörmann, G. Rigoll, GaitGraph: graph convolutional network for skeleton-based gait recognition, in: 2021 IEEE International Conference on Image Processing (ICIP), IEEE, 2021, pp. 2314–2318.
[224] K. Xu, X. Jiang, T. Sun, Gait Identification Based on Human Skeleton with Pairwise Graph Convolutional Network, in: 2021 IEEE International Conference on Multimedia and Expo (ICME), 2021.
[225] I. Rida, N. Almaadeed, S. Almaadeed, Robust gait recognition: a comprehensive survey, IET Biometrics 8 (1) (2019) 14–28.
[226] I. Rida, X. Jiang, G.L. Marcialis, Human Body Part Selection by Group Lasso of Motion for Model-Free Gait Recognition, IEEE Signal Process. Lett. 23 (1) (Jan. 2016) 154–158.
[227] I. Rida, N. Al-Maadeed, S. Al-Maadeed, S. Bakshi, A comprehensive overview of feature representation for biometric recognition, Multimedia Tools Appl. 79 (7) (2020) 4867–4890.
[228] I. Rida, L. Boubchir, N. Al-Maadeed, S. Al-Maadeed, A. Bouridane, Robust model-free gait recognition by statistical dependency feature selection and globality-locality preserving projections, in: 2016 39th International Conference on Telecommunications and Signal Processing (TSP), IEEE, 2016, pp. 652–655.
[229] I. Rida, N. Al Maadeed, G.L. Marcialis, A. Bouridane, R. Herault, G. Gasso, in: Improved model-free gait recognition based on human body part. In Biometric Security and Privacy, Springer, Cham, 2017, pp. 141–161.
[230] Rida, I. (2019). Towards human body-part learning for model-free gait recognition. arXiv preprint arXiv:1904.01620.