*applied sciences*

MDPI

*Article*

# DeepGait: A Learning Deep Convolutional Representation for View-Invariant Gait Recognition Using Joint Bayesian

**Chao Li [1,*], Xin Min [1], Shouqian Sun [1], Wenqian Lin [1] and Zhichuan Tang [2]**

[1] College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China; minx@zju.edu.cn (X.M.); ssq@zju.edu.cn (S.S.); linwq@zju.edu.cn (W.L.)

[2] Industrial Design Institute, Zhejiang University of Technology, Hangzhou 310023, China; ttzzcc@zju.edu.cn

* Correspondence: superli@zju.edu.cn

Academic Editor: José Antonio Iglesias Martínez
Received: 3 January 2017; Accepted: 15 February 2017; Published: 23 February 2017

**Abstract:** Human gait, as a soft biometric, helps to recognize people through their walking. To further improve the recognition performance, we propose a novel video sensor-based gait representation, DeepGait, using deep convolutional features and introduce Joint Bayesian to model view variance. DeepGait is generated by using a pre-trained "very deep" network "D-Net" (VGG-D) without any fine-tuning. For non-view setting, DeepGait outperforms hand-crafted representations (e.g., Gait Energy Image, Frequency-Domain Feature and Gait Flow Image, etc.). Furthermore, for cross-view setting, 256-dimensional DeepGait after PCA significantly outperforms the state-of-the-art methods on the OU-ISR large population (OULP) dataset. The OULP dataset, which includes 4007 subjects, makes our result reliable in a statistically reliable way.

**Keywords:** deep convolutional features; gait representation; Joint Bayesian; cross-view gait recognition; gait identification; gait verification

## 1. Introduction

Biometrics refer to the use of intrinsic physical or behavioral traits in order to identify humans. Besides regular features (face, fingerprint, iris, DNA and retina), human gait, which can be obtained from people at larger distances and at low resolution without subjects' cooperation has recently attracted much attention. It also has a vast application prospect in crime investigation and wide-area surveillance. For example, criminals usually wear gloves, dark sun-glasses, and face masks to invalidate finger print, eyes, and face recognition. In such scenarios, gait recognition is the only useful and effective identification method. Previous research [1,2] has shown that human gait, specifically the walking pattern, is difficult to disguise and unique to each person.

In general, video sensor-based gait recognition methods are divided into two families: appearance-based [3–7] and model-based [8–10]. Appearance-based methods focus on the motion of human body and usually operate on silhouettes of gait. They extract the gait descriptors from the silhouettes. The general framework of appearance-based methods usually consists of silhouette extraction, period detection, representation generation, and recognition. Model-based gait recognition focuses more on the extraction of the stride parameters of subject that describe the gait by using the human body structure. The model-based methods usually require high resolution images as well as being computationally expensive, while gait recognition needs to be real-time and effective at low resolution. Our proposed work falls in the category of appearance-based methods. It differs from the majority of contributions in the field in that the Deep Learning (DL) framework is used to extract gait representation compared with well engineered features such as the widely used average silhouette representations: Gait Energy Image (GEI) [3], Gait Flow Image (GFI) [5], Gait Entropy Image (GEnI), Masked GEI based on GEnI (MGEI) [4], and Frequency-Domain Feature (FDF) [6,7]. However, the performance of gait recognition is often influenced by several covariates such as clothing, walking speed, observation views, and carrying bags. For appearance-based methods, view changes are the most problematic covariates. Therefore, we propose a more discriminative appearance-based representation, DeepGait and introduce Joint Bayesian to deal with the view change problems. Numerous experiments were conducted for both non-view variance and cross-view settings on the OU-ISIR large population (OULP) dataset [11] to validate the effectiveness of our proposed method.

### *1.1. Proposal of Deep Convolutional Gait Representation*

Inspired by the deep learning breakthroughs in the image domain [12–14] where rapid progress has been made in the past few years in feature learning, and various pre-trained deep convolutional models [12,13,15] were made available for extracting image and video features, DeepGait was proposed. These features are the activations of the network's last few fully-connected layers which perform well in the other vision tasks [14–17]. A convolutional neural network (CNN) has been successfully demonstrated in many research fields, such as face recognition [18–20] and human action recognition [15] which are relevant to gait recognition. However, to the best of our knowledge, few studies have applied deep learning features in video sensor-based human gait recognition except for [21,22]. In this paper, we proposed a novel gait representation, DeepGait based on VGG-D [12] features using max-pooling on each gait cycle. If the gait video sequence has more than one cycle, we just choose the first one. Our proposed DeepGait differs from [21] in two ways: (1) they first needed to compute the traditional gait representations (GEI, FDF), and regard them as the input data while we just used the original silhouette images; (2) their net needed to be trained on the gait dataset while ours just used the pre-trained VGG-D model without any fine-tuning.

### *1.2. Joint Bayesian for Modeling View Variance*

When dealing with view change problems, several appearance-based approaches are proposed: (1) the view transformation model (VTM) [23,24]; (2) the view-invariant feature-based approaches [21,25]; and (3) multiview gallery-based approaches [26,27]. On the OULP dataset, VTM-based methods are widely used: [24] proposed a generative approach which is a kind of VTM-based methods and makes use of transformation consistency measures (TCM+); [23] further proposed a quality-dependent VTM (wQVTM). Recently, a view-invariant feature-based approach (GEINet) [21] was proposed and achieved the best performance. We introduce Joint Bayesian [28] to model the view variance which differs from the above approaches. For comparison, the unsupervised Nearest Neighbor classifier based on euclidean distance (NN) is also adopted as a baseline method. In order to evaluate the compactness of DeepGait, PCA is used to project the representation into lower dimensions. Furthermore, we choose the right $K = 256$ components to strike a balance between recognition performance and computational complexity when using Joint Bayesian.

### *1.3. Overview*

Our contributions include: (1) introducing deep learning for gait recognition and proposal of a new gait representation which outperforms traditional gait representations when the gallery and probe gait sequences are from the same view (non-view setting); (2) model view variance using Joint Bayesian when the gallery and probe gait sequences are from different views (cross-view setting); (3) improved recognition performances on the OULP dataset for non-view and cross-view settings; (4) making public the trained Joint Bayesian model, test codes and experimental results for further comparison.

Figure 1 shows the overview of our method. The outline of the paper is organized as follows. Section 2 introduces DeepGait, Joint Bayesian for identification and verification tasks, and some evaluation criteria. Section 3 presents the experimental results on the OULP dataset. Section 4 offers our conclusion.

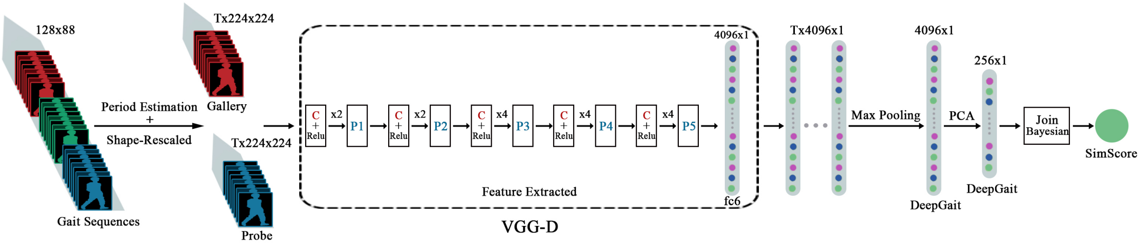


**Figure 1.** An illustration of the proposed gait recognition process. (C: convolution, P: max-pooling, T: gait period).

## 2. Proposed Method

### *2.1. Deep Convolutional Gait Representation*

#### 2.1.1. Gait Period Estimation

Similar to the other appearance-based gait recognition methods, the first step for DeepGait generation is gait period detection. As in [6,11], we calculated the Normalized Auto Correlation (NAC) of each normalized gait sequence along the temporal axis:

$$NAC(N) = \frac{\sum_{x,y}\sum_{n=0}^{N_{total}-N-1} S(x,y,n)S(x,y,n+N)}{\sqrt{\sum_{x,y}\sum_{n=0}^{N_{total}-N-1} S(x,y,n)^2}\sqrt{\sum_{x,y}\sum_{n=0}^{N_{total}-N-1} S(x,y,n+N)^2}} \tag{1}$$

where $NAC(N)$ stands for the autocorrelation for the $N$ frame shift which can quantify periodic gait motion. $N_{total}$ is the number of frames in each gait sequence. $S(x,y,n)$ is the silhouette gray value at position of (x, y) on the $n$-th frame. Empirically, for the natural gait period, the domain of N is set to be [20, 40] and the gait period is estimated as:

$$T_{gait} = arg \max_{N\in[20,40]} NAC(N) \tag{2}$$

where $T_{gait}$ is the gait period. We have made the code and result (large deviations was manually modified) public in Supplementary Materials.

#### 2.1.2. Network Structure

In this paper, a state-of-the-art deep convolutional model (VGG-D) [12] which consists of 19 parameterized layers (16 convolutional layers and 3 fully connected layers) was adopted. Figure 1 shows its' partial structure. VGG-D evaluated very deep convolutional networks using an architecture with very small (3 × 3) convolution filters, which achieved a significant improvement on ImageNet Large-Scale Visual Recognition Challenge 2014 (ILSVRC-2014) [12].

#### 2.1.3. Supervised Pre-Training

By leveraging a large auxiliary labeled dataset to train a deep convolutional model, the high-level learned features from the pre-trained model have sufficient discrimination ability in some image-based classification tasks [16]. To evaluate the efficacy of learned features on gait recognition task, we trained VGG-D net using ImageNet dataset (classification annotations only) [13]. The training procedure generally followed Simonyan et al. [12]. Namely, based on mini-batch stochastic gradient descent, the back-propagation algorithm is used to optimize the softmax-regression objection function [29]. In this paper, we did not fine-tune the model using any gait dataset, because deep convolution features using the pre-trained model had already shown a significant improvement compared to traditional hand-crafted gait representations for non-view setting.

### 2.1.4. Feature Extraction

In order to extract deep learned features for gait representation generalization, the size of input gait silhouette images must be compatible with VGG-D's input size which is known as 224 × 224 pixel size. We first rescaled each image to fixed size. Features were then computed by forward propagating a mean-subtracted and size-fixed (224 × 224) gait image through 16 convolutional/pooling layers and 2 fully connected layers using Caffe, a open source CNN library [30]. According to the other vision tasks [14–17], the first fully connected layer's ( $fc6$) features outcome the other layers' features. Unless otherwise specified, we extracted the 4096-dimensional $fc6$ features as deep convolutional features for gait representation generalization.

### 2.1.5. Representation Generalization and Visualization

Inspired by Gait Energy Image (GEI) which is obtained by simply averaging the silhouette sequence over one gait period and can capture both the spatial and temporal information [3,21] , we make use of max-pooling method over one gait period's $fc6$ features to combine the spatio-temporal information. Another version of $fc6$ features with average-pooling has been tested in our experiments and showed inferior performance, which suggests the DeepGait is valid. In the i-th gait period, if there are T silhouette images, we can generate T $fc6$ features. The j-th deep convolutional gait representation (DeepGait) element of 4096-dimensional representation can then be created from maxing the $fc6$ features by using Equation (3).

$$DeepGait_{i,j} = \max_{k=0}^{T-1} fc6_{i,j,k} \tag{3}$$

Examples of the 256-dimensional DeepGait from the OULP dataset after dimension reduction (in Section 2.2.3) and L2-normalization are shown in Figure 2.

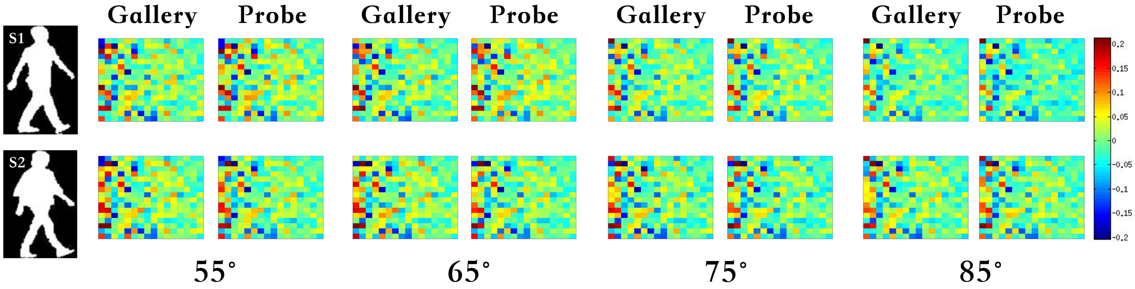


**Figure 2.** Examples of the 256-dimensional DeepGait after dimension reduction under four observation views (55°, 65°, 75°, 85°) . S1 and S2 represent two different subjects, separately. We rearrange the vector as 16 × 16 matrix for the convenience of visualization. Approximately 25% features are non-zero values. Different colors stand for different values.

## *2.2. Gait Recognition*

Usually, gait recognition can be divided into two major tasks: gait verification and gait identification as in face recognition [18–20]. Gait verification is used for verifying whether two input gait sequences (Gallery, Probe) belong to the same subject. In this paper, we calculated the similar score (*SimScore*) using Joint Bayesian to evaluate the similarity of two given sequences. Euclidean distance was also adopted as a baseline method for comparison. In gait identification, a set of subjects are gathered (The gallery), and it aims to decide which of the gallery identities are similar to the probe at test time. Under the closed set identification condition [31], a probe sequence is compared with all the gallery identities, then the identity which has the largest *SimScore* is the final result.

### 2.2.1. Gait Verification Using Joint Bayesian

Joint Bayesian [28] technique was widely and successfully used for face verification [18,19,32]. In this paper, we modeled the extracted DeepGait (after mean-subtracted) by summing two independent Gaussian variables as:

$$x = \mu + \varepsilon \tag{4}$$

where $x$ represents a mean-subtracted DeepGait vector. For a better performance, $L_2$- normalization was applied for DeepGait. $\mu$ is gait identity following a Gaussian distribution $N(0, S_\mu)$. $\varepsilon$ stands for different gait variations (e.g., view, clothing and carrying bags etc.) following a Gaussian distribution $N(0, S_\varepsilon)$. Joint Bayesian models the joint probability of two gait representations using the intra-class variation (I) or inter-class variance (E) hypothesis, $P(x_1, x_2|H_I)$ and $P(x_1, x_2|H_E)$. Given the above prior from Equation (4) and the independent assumption between $\mu$ and $\varepsilon$, the covariance matrix of $P(x_1, x_2|H_I)$ and $P(x_1, x_2|H_E)$ can be derived separately as:

$$\Sigma_I = \begin{bmatrix} S_\mu + S_\varepsilon & S_\mu \\ S_\mu & S_\mu + S_\varepsilon \end{bmatrix} \tag{5}$$

$$\Sigma_E = \begin{bmatrix} S_\mu + S_\varepsilon & 0 \\ 0 & S_\mu + S_\varepsilon \end{bmatrix} \tag{6}$$

$S_\mu$ and $S_\varepsilon$ are two unknown covariance matrices which can be learned from the training set using the Expectation Maximization (EM) algorithm. During the testing phase, the likelihood ratio ($r(x1, x2)$) is regarded as the similar score (*SimScore*):

$$SimScore(x_1, x_2) = r(x_1, x_2) = log\frac{P(x_1, x_2|H_I)}{P(x_1, x_2|H_E)} \tag{7}$$

$r(x_1, x_2)$ is efficiently obtained with the following closed-form process:

$$r(x_1, x_2) = x_1^T A x_1 + x_2^T A x_2 - 2x_1^T G x_2 \tag{8}$$

where $A$ and $G$ are two final result models, which can be obtained by using simple algebra operations between $S_\mu$ and $S_\varepsilon$. Please refer to [28] for more details. We also make public our trained model ($A$ and $G$) and testing codes in Supplementary Materials for further comparison.

Euclidean distance is also adopted as a baseline method for comparison and the similar score (*SimScore*) can be calculated as:

$$SimScore(x_1, x_2) = -||\frac{x_1}{||x_2||} - \frac{x_1}{||x_2||}|| \tag{9}$$

Finally, *SimScore* is compared with a threshold value to verify whether $x_1$ and $x_2$ belong to the same subject.

### 2.2.2. Gait Identification

For gait identification, the probe sample $x_p$ is classified as class $i$, if the final *SimScore* with all the gallery ($x_i$) is the maximum as shown in Equation (10).

$$i = arg \max_{i \in [0, N_{gallery-1}]} SimScore(x_i, x_p) \tag{10}$$

where $N_{gallery}$ is the number of training subjects. In the experiments, we just used the first period of the gait sequence.

### 2.2.3. Dimension Deduction by PCA

The dimension of DeepGait is relatively large (4096) which makes the training process of Joint Bayesian computationally expensive. In order to compute efficiently and evaluate the compactness of DeepGait, we used PCA to project the representation into lower dimensions. PCA can capture the principle components of the origin space. Among all the gallery dataset, we calculated a transformation matrix ($E_{PCA}$) using singular value decomposition for its within-class scatter matrix. The transformation matrix's dimension is $M \times K$, where M is DeepGait's origin dimension, and K is the number of components.

After PCA, for baseline method (euclidean distance), the *SimScore* is calculated as:

$$SimScore(x_1, x_2) = -||\frac{E_{PCA}x_1}{||E_{PCA}x_1||} - \frac{E_{PCA}x_2}{||E_{PCA}x_2||}|| \tag{11}$$

For Joint Bayesian, the *SimScore* is calculated as:

$$SimScore(x_1, x_2) = log\frac{P(E_{PCA}x_1, E_{PCA}x_2|H_I)}{P(E_{PCA}x_1, E_{PCA}x_2|H_E)} \tag{12}$$

## *2.3. Evaluation Criteria*

The recognition performance was evaluated using four metrics: (1) Cumulative Match Characteristics (CMC) curve; (2) rank-1 and rank-5 identification rates; (3) the Receiver Operating Characteristic (ROC) curve of False Acceptance Rates (FAR) and Ralse Rejection Rates (FRR); and (4) Equal Error Rates (EERs). CMC curve, and rank-1/rank-5 identification rates were used for the identification task while ROC curve and EERs were used for the verification task.

## **3. Experiment**

The proposed method was evaluated on the OU-ISIR large population (OULP) dataset which has over 4000 subjects and contains high-quality silhouette images with view variations [11]. The experiments were conducted with two main settings: non-view setting and cross-view setting. For the first setting, all the subjects were used to evaluate the performance of our proposed DeepGait, so that the result could be reliable in a statistical manner. For the second setting, we used a subset of the OULP dataset following the protocol of [21,23,24] for comparison. For further comparison, experimental results, learning models, and test codes are released in Supplementary Materials.

## *3.1. Comparisons of Different Gait Representations for the Non-View Setting*

In this section, we aimed at comparing the performance of our proposed DeepGait with some state-of-the-art gait representations (e.g., GEI, FDF, MGEI, GEnI and GFI) in a statistically reliable manner. The unsupervised whole dataset (NN) classifier was chosen for the sake of all the subjects being used for testing. When we exchanged the gallery and the probe, 2-fold cross validation was adopted. Based on the video sensor's recorded view (55°, 65°, 75°, 85°), we reported the results of comparison in Table 1.

**Table 1.** Comparison of rank-1 (%) and rank-5 (%) identification rates with different gait representations on the whole dataset (NN). GEI: Gait Energy Image; MGEI: Masked GEI based on GEnI; GEnI: Gait Entropy Image; FDF: Frequency-Domain Feature; GFI: Gait Flow Image.

| Rank-1/Rank-5 | Dataset | #Subjects | DeepGait | GEI | MGEI | GEnI | FDF | GFI |
|---|---|---|---|---|---|---|---|---|
| **rank-1** | **View-55** | 3,706 | **90.6** | 85.3 | 79.3 | 75.1 | 83.1 | 61.9 |
| | **View-65** | 3,770 | **91.2** | 85.6 | 83.2 | 77.3 | 84.7 | 66.6 |
| | **View-75** | 3,751 | **91.2** | 86.1 | 84.6 | 79.1 | 86.0 | 69.3 |
| | **View-85** | 3,249 | **92.0** | 85.3 | 83.9 | 80.7 | 85.6 | 69.8 |
| | **Mean** | | **92.3** | 85.6 | 82.8 | 78.1 | 84.9 | 66.9 |
| **rank-5** | **View-55** | 3,706 | **96.0** | 91.8 | 89.3 | 85.5 | 91.0 | 75.5 |
| | **View-65** | 3,770 | **96.0** | 92.3 | 91.5 | 87.7 | 92.3 | 79.5 |
| | **View-75** | 3,751 | **96.1** | 92.2 | 92.0 | 88.8 | 92.5 | 81.3 |
| | **View-85** | 3,249 | **96.5** | 92.6 | 91.9 | 89.3 | 92.3 | 81.9 |
| | **Mean** | | **96.2** | 92.2 | 91.2 | 87.8 | 92.0 | 79.6 |

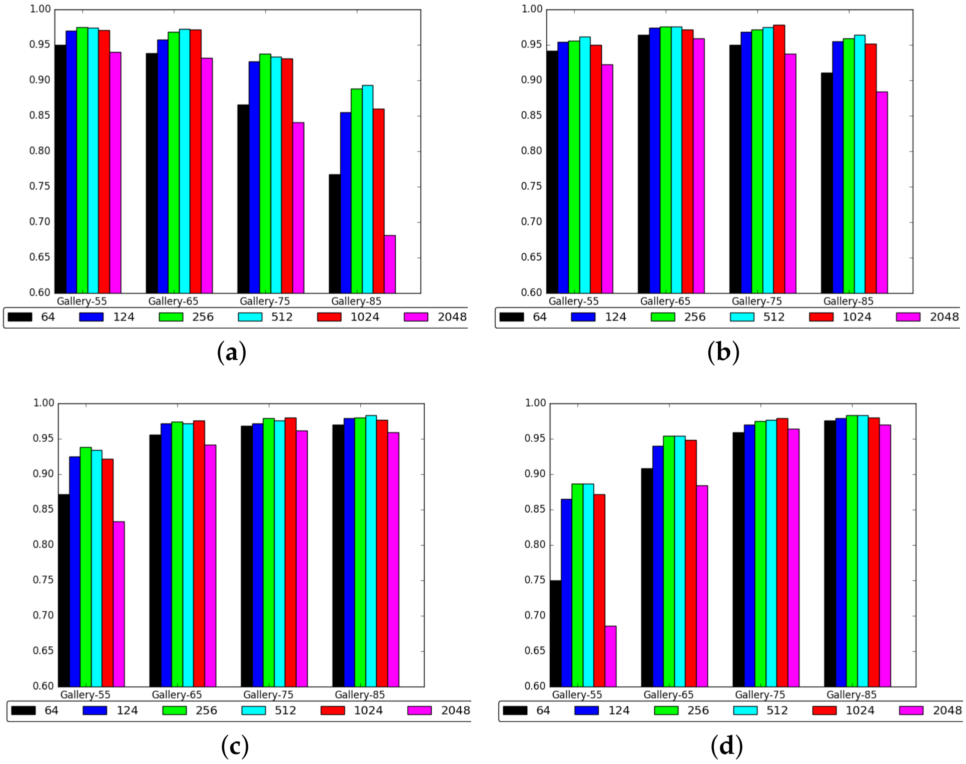
As result, **DeepGait**, using the simple classify method (NN), retained powerful discrimination even over large population condition and outperforms other famous representations. From the four observed views' result, the performance of Deep Gait, GEI and FDF is nearly the same under different observation view. Our proposed DeepGait is independent of view change.

### *3.2. Results for the Cross-View Setting*

In the following two subsections, we chose 1912 subjects containing two gait sequences (Gallery, Probe), and the subset was further divided into two groups of the same number of subjects, one for training while the other one for testing. Following the protocol of [21,23,24] (publicly available at http://www.am.sanken.osaka-u.ac.jp/BiometricDB/dataset/GaitLP/Benchmarks.html), five 2-fold cross validations were performed. During each training phase, 956 × (956-1) = 912,980 intra-class samples and 956 × 1 = 956 inter-class samples were used for training Joint Bayesian. Due to the limited space, the gallery dataset are fixed at three views (55°, 65°, 75°) when we show the CMC and ROC curves.

#### 3.2.1. Number of Components Selection for Joint Bayesian

As we know, the dimension of DeepGait is 4096, and high dimension means that more training data are needed for model learning when Joint Bayesian [28] is used for gait recognition. In fact, number of training samples is often limited in gait recognition, therefore, the dimension of DeepGait needs to be reduced. Due to the powerful discrimination of our proposed DeepGait, we can achieve a competitive performance even in a low dimension after PCA. Experiments of different number of components were performed with Joint Bayesian, so that we could choose the right $K$ components, where $K$ is the number of components, to strike a balance between recognition performance and computational complexity. Figure 3 shows the results of different $K$ components under different combinations of Gallery and Probe views.



(**a**) (**b**)

(**c**) (**d**)

**Figure 3.** Rank-1 identification rates of different number of components after PCA under different Gallery-view and Probe-view combinations. (JB). (**a**) Probe-55; (**b**) Probe-65; (**c**) Probe-75; (**d**) Probe-85.

We can see that $K = 2048$, achieved the worst performance due to under-fitting. The training samples are insufficient when Joint Bayesian was used with high dimension. When dealing with the lowest dimension ($K = 64$), our proposed method still achieved competitive performances among three cross-view combinations (55:65, 65:75, 75:85). Further, we found that $'K = 256'$ achieved almost the same result with $'K = 512'$ under all the cross-view conditions while $'K = 256'$ has half the number of components. For the best balance of performance and computing cost, we finally set $K = 256$ when Joint Bayesian is used in the following experiments.

### 3.2.2. Comparisons with the State-of-the-Art Methods

The proposed method is further compared with other state-of-the-art methods [21,23,24] in cross-view gait recognition. Muramatsu et al. [23,24] proposed the evaluation criteria and five 2-fold cross validations were performed to reduce the effect of random grouping in their experiments. Ref. [24] proposed a generative approach which is a View Transformation Model (VTM) based on transformation consistency measures (TCM+). Ref. [23] further proposed a quality-dependent VTM (wQVTM). Shiraga et al. [21] designed a convolutional neural network for cross-view gait recognition. They reported two kinds of results which mainly differ in input data (GEI, FDF), and the two methods are referred to as GEINet and w/FDF, respectively [21].

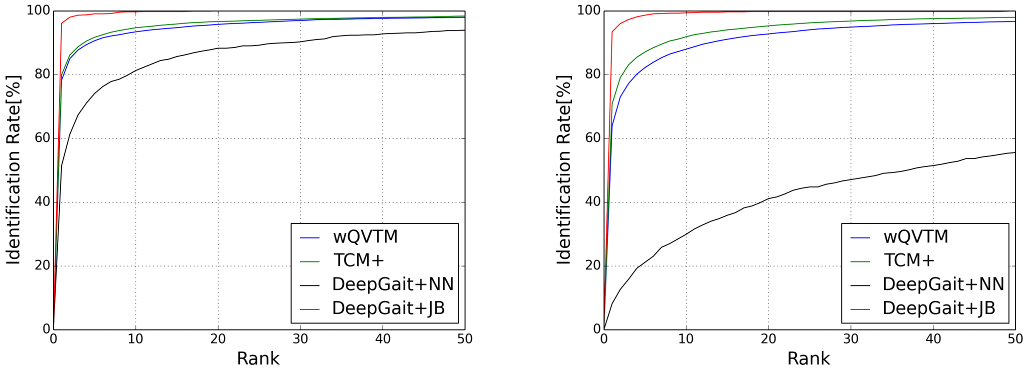
### A. Comparisons for identification task

The performance of our proposed method, 256-dimensional DeepGait with Joint Bayesian (DeepGait + JB) was firstly evaluated in identification task. 4096-dimensional DeepGait with nearest neighbor classifier based on euclidean distance (DeepGait + NN) is also adopted as a baseline method. We summarize the rank-1 and rank-5 identification rates in Table 2. CMC curves are also shown in Figure 4.

As a result, **DeepGait + JB** significantly outperformed the three state-of-the-art methods for all the view combinations. Even with simple classifier NN, DeepGait still achieved competitive performances for four side litter view difference combinations (65:75, 75:85).

**Table 2.** Comparison of rank-1 (%) and rank-5 (%) identification rates with other existent methods in different cross-view settings.

| Gallery View | Method | Rank-1 [%] | | | | Rank-5 [%] | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 55 | 65 | 75 | 85 | 55 | 65 | 75 | 85 |
| 55 | GEINet | (94.7) | 93.2 | 89.1 | 79.9 | | | | |
| | w/FDF | (92.7) | 91.4 | 87.2 | 80.0 | | | | |
| | TCM+ | | 79.9 | 70.8 | 54.5 | | 91.7 | 87.1 | 79.3 |
| | wQVTM | | 78.3 | 64.0 | 48.6 | | 90.6 | 82.2 | 73.9 |
| | DeepGait + NN | (92.7) | 51.5 | 8.2 | 2.9 | (97.2) | 74.1 | 21.1 | 9.3 |
| | DeepGait + JB | (97.4) | **96.1** | **93.4** | **88.7** | (99.2) | **99.1** | **98.6** | **97.1** |
| 65 | GEINet | 93.7 | (95.1) | 93.8 | 90.6 | | | | |
| | w/FDF | 92.3 | (93.9) | 92.2 | 88.6 | | | | |
| | TCM+ | 81.7 | | 79.5 | 70.2 | 92.1 | | 90.2 | 86.8 |
| | wQVTM | 81.5 | | 79.2 | 67.5 | 91.9 | | 90.2 | 84.8 |
| | DeepGait + NN | 48.5 | (94.4) | 73.7 | 34.3 | 70.2 | (97.6) | 88.8 | 56.9 |
| | DeepGait + JB | **97.3** | (97.6) | **97.2** | **95.4** | **99.5** | (99.5) | **99.3** | **99.2** |
| 75 | GEINet | 91.1 | 94.1 | (95.2) | 93.8 | | | | |
| | w/FDF | 88.8 | 92.6 | (93.4) | 91.9 | | | | |
| | TCM+ | 71.9 | 80.0 | | 79.0 | 88.1 | 91.4 | | 90.3 |
| | wQVTM | 70.2 | 80.0 | | 78.2 | 87.1 | 91.4 | | 89.9 |
| | DeepGait + NN | 7.5 | 76.3 | (94.5) | 89.2 | 18.7 | 92.3 | (97.6) | 96.6 |
| | DeepGait + JB | **93.3** | **97.5** | (97.7) | **97.6** | **99.1** | **99.3** | (99.4) | **99.1** |
| 85 | GEINet | 81.4 | 91.2 | 94.6 | (94.7) | | | | |
| | w/FDF | 80.9 | 88.4 | 92.2 | (93.2) | | | | |
| | TCM+ | 53.7 | 73.0 | 79.4 | | 79.6 | 87.9 | 91.2 | |
| | wQVTM | 51.1 | 68.5 | 79.0 | | 75.6 | 85.7 | 91.1 | |
| | DeepGait + NN | 2.8 | 37.2 | 90.5 | (94.8) | 9.9 | 60.9 | 96.5 | (97.8) |
| | DeepGait + JB | **89.3** | **96.4** | **98.3** | (98.3) | **98.3** | **99.3** | **99.1** | (99.1) |



(**a**) (**b**)

**Figure 4.** *Cont.*

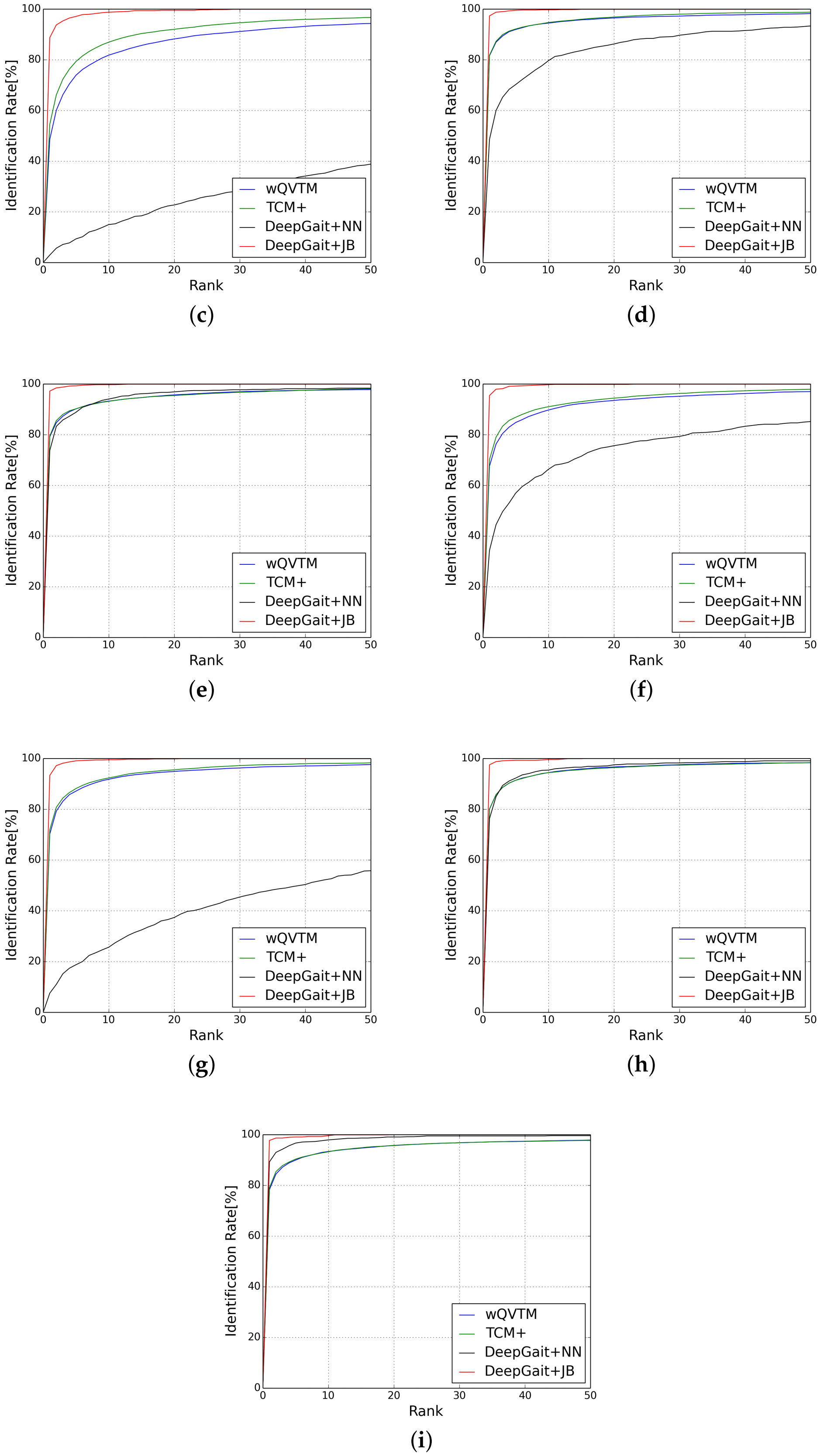


**Figure 4.** Cummulative Match Characteristics (CMC) curves under different cross-view settings. (**a**) G-55:P-65; (**b**) G-55:P-75; (**c**) G-55:P-85; (**d**) G-65:P-55; (**e**) G-65:P-75; (**f**) G-65:P-85; (**g**) G-75:P-55; (**h**) G-75:P-65; (**i**) G-75:P-85.

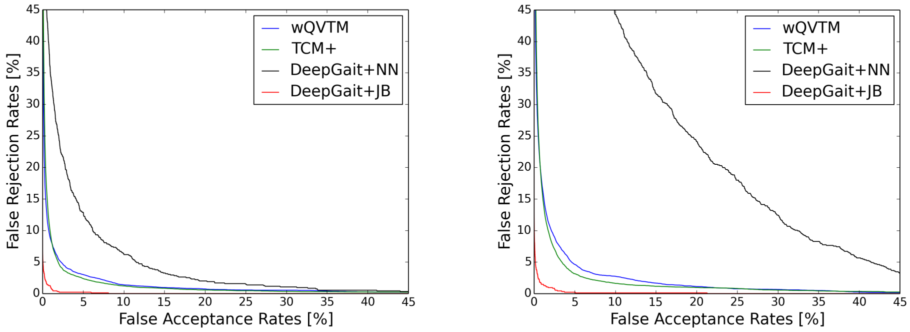
### B. Comparisons for verification task

We used the same protocol as the identification task and summarize the EERs for verification task in Table 3. We also referred DeepGait based on euclidean distance as DeepGait + NN for the sake of consistency.

We find that our proposed method also achieved the best EERs in all cases, especially in cases with large view variance. More specifically, our proposed method improved from 2.5% to 1.9% compared to the best method (GEINet) where the probe view was 85° and gallery view was 55°. Under the exchanged view condition, EERs improved from 2.4% to 1.6%. When comparing DeepGait + NN with **DeepGait + JB**, we can conclude that Joint Bayesian well models the view variance while simple euclidean distance can not well deal with cross-view test in verification task. Figure 5 shows more details of ROC curves.

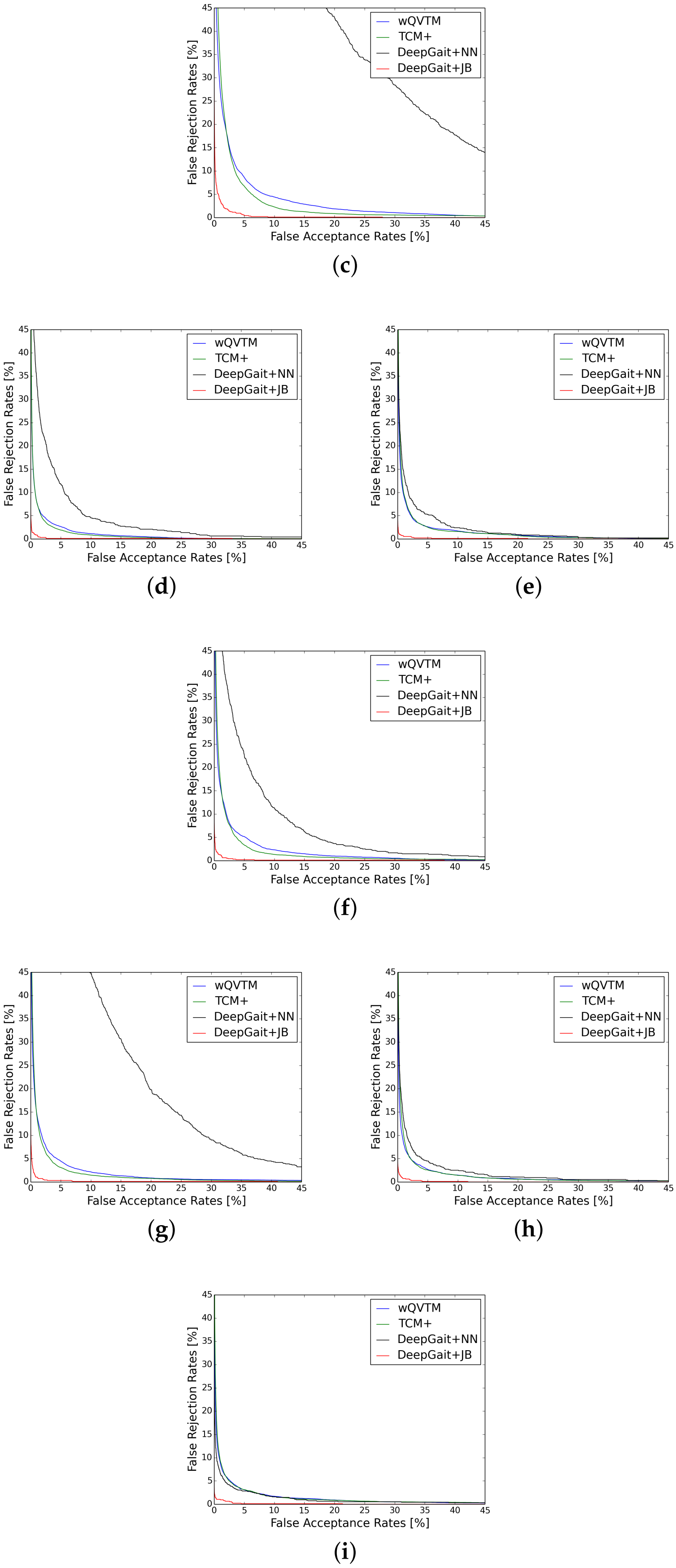
**Table 3.** Comparison of EERs (%) with other existent methods under different cross-view settings.

| Gallery View | Method | 55 | 65 | 75 | 85 |
|---|---|---|---|---|---|
| 55 | GEINet | (1.3) | 1.4 | 1.7 | 2.5 |
| | w/FDF | (1.9) | 2.0 | 2.3 | 2.9 |
| | TCM+ | | 3.2 | 4.0 | 5.7 |
| | wQVTM | | 3.6 | 4.8 | 6.5 |
| | DeepGait + NN | (2.9) | 7.9 | 21.6 | 29.4 |
| | DeepGait + JB | (0.8) | **1.0** | **1.3** | **1.9** |
| 65 | GEINet | 1.2 | (1.0) | 1.3 | 1.6 |
| | w/FDF | 1.7 | (1.4) | 1.7 | 2.2 |
| | TCM+ | 3.0 | | 3.4 | 4.2 |
| | wQVTM | 3.5 | | 3.4 | 5.1 |
| | DeepGait + NN | 7.2 | (3.1) | 5.1 | 10.6 |
| | DeepGait + JB | **0.8** | (0.6) | **0.7** | **1.2** |
| 75 | GEINet | 1.5 | 1.2 | (1.2) | 1.4 |
| | w/FDF | 2.0 | 1.5 | (1.6) | 1.7 |
| | TCM+ | 4.0 | 3.4 | | 3.8 |
| | wQVTM | 4.7 | 3.7 | | 3.8 |
| | DeepGait + NN | 19.9 | 4.6 | (2.7) | 3.4 |
| | DeepGait + JB | **1.1** | **0.8** | (0.8) | **1.0** |
| 85 | GEINet | 2.4 | 1.6 | 1.2 | (1.1) |
| | w/FDF | 2.5 | 1.9 | 1.6 | (1.4) |
| | TCM+ | 5.5 | 4.4 | 3.7 | |
| | wQVTM | 6.5 | 4.9 | 3.7 | |
| | DeepGait + NN | 28.5 | 10.0 | 3.4 | (2.3) |
| | DeepGait + JB | **1.6** | **0.9** | **0.9** | (1.0) |



(**a**) (**b**)

**Figure 5.** *Cont.*



(**c**)

(**d**) (**e**)

(**f**)

(**g**) (**h**)

(**i**)

**Figure 5.** Receiver Operative Characteristics (ROC) curves under different cross-view settings. (**a**) G-55:P-65; (**b**) G-55:P-75; (**c**) G-55:P-85; (**d**) G-65:P-55; (**e**) G-65:P-75; (**f**) G-65:P-85; (**g**) G-75:P-55; (**h**) G-75:P-65; (**i**) G-75:P-85.

## 4. Conclusions

In this paper, we have proposed a new video sensor-based gait representation, DeepGait, for gait recognition and the performance is evaluated on the OU-ISIR large population dataset. For the same view setting, DeepGait has been reported to achieve significantly better performance than previous hand-crafted gait representations (GEI, MGEI, GEnI, FDF, GFI) even with NN classifier based on euclidean distance. The results are reported in a statistically reliable manner, due to a large number in the dataset. Furthermore, Joint Bayesian is used for model the view variance for cross-view setting. We also find DeepGait in 256-d after PCA best balances performance and computing cost with Joint Bayesian. For the cross-view setting, our proposed method significantly outperformed the state-of-the-art methods for both verification and identification tasks. Even with large view variance, our proposed method achieved the best rank-1 identification rate of 88.7%/89.3% and the best EERs of 1.9%/1.6% with (G-55: P-85)/(G-85: P-55), respectively.

For future research, we will evaluate our proposed method against other variances (e.g., clothing, carrying bags and a wider view variation).

**Supplementary Materials:** The experimental results and diffferent gait representations are available online at Zenodo DOI:10.5281/zenodo.321246 (https://doi.org/10.5281/zenodo.321246).

**Acknowledgments:** The authors would like to thank OU-ISIR for providing access to the OU-ISIR Large Population Gait Database. This study is partly supported by the Department of Science and Technology of Zhejiang Province (No. 2015C31051).

**Author Contributions:** C.L. conceived and designed the experiments; S.S. and Z.T. supervised the work; W.L. and X.M. analyzed the data; C.L. and X.M. wrote the paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| DeepGait | Gait representation based on deep convolutional features |
| GEI | Gait energy image |
| MGEI | Masked gait energy image based on gait entropy image |
| GEnI | Gait entropy image |
| GFI | Gait flow image |
| FDF | Frequency-Domain feature |
| CMCs | Cumulative match characteristics |
| ROC | Receiver operating characteristic |
| JB | Joint Bayesian |
| NN | Nearest neighbor classifier based on euclidean distance |
| OULP | the OU-ISIR large population dataset |

## References

1. Murray, M.P.; Drought, A.B.; Kory, R.C. Walking patterns of normal men. *J. Bone Jt. Surg. Am.* **1964**, *46*, 335–360.
2. Cutting, J.E.; Kozlowski, L.T. Recognizing friends by their walk: Gait perception without familiarity cues. *Bull. Psychon. Soc.* **1977**, *9*, 353–356.
3. Man, J.; Bhanu, B. Individual recognition using gait energy image. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 316–322.
4. Bashir, K.; Xiang, T.; Gong, S. Gait recognition without subject cooperation. *Pattern Recognit. Lett.* **2010**, *31*, 2052–2060.
5. Lam, T.H.; Cheung, K.H.; Liu, J.N. Gait flow image: A silhouette-based gait representation for human identification. *Pattern Recognit.* **2011**, *44*, 973–987.

6. Makihara, Y.; Sagawa, R.; Mukaigawa, Y.; Echigo, T.; Yagi, Y. Gait recognition using a view transformation model in the frequency domain. In *European Conference on Computer Vision*; Springer: Berlin, Germany, 2006; pp. 151–163.
7. Bashir, K.; Xiang, T.; Gong, S. Gait recognition using gait entropy image. In Proceedings of the 3rd International Conference on Crime Detection and Prevention (ICDP 2009), IET, London, UK, 2–3 December 2009; pp. 1–6.
8. Luo, J.; Tang, J.; Tjahjadi, T.; Xiao, X. Robust arbitrary view gait recognition based on parametric 3D human body reconstruction and virtual posture synthesis. *Pattern Recognit.* **2016**, *60*, 361–377.
9. Bhanu, B.; Han, J. Model-based human recognition—2D and 3D gait. In *Human Recognition at a Distance in Video*; Springer: Berlin, Germany, 2010; pp. 65–94.
10. Nixon, M.S.; Carter, J.N.; Cunado, D.; Huang, P.S.; Stevenage, S. Automatic gait recognition. In *Biometrics*; Springer: Berlin, Germany, 1996; pp. 231–249.
11. Iwama, H.; Okumura, M.; Makihara, Y.; Yagi, Y. The ou-isir gait database comprising the large population dataset and performance evaluation of gait recognition. *IEEE Trans. Inf. Forensics Secur.* **2012**, *7*, 1511–1521.
12. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556 .
13. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*; NIPS Foundation Inc.: South Lake Tahoe, UV, USA, 2012; pp. 1097–1105.
14. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus Convention Center Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
15. Tran, D.; Bourdev, L.; Fergus, R.; Torresani, L.; Paluri, M. Learning spatiotemporal features with 3d convolutional networks. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 4489–4497.
16. Donahue, J.; Jia, Y.; Vinyals, O.; Hoffman, J.; Zhang, N.; Tzeng, E.; Darrell, T. DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition. *Comput. Vision Pattern Recognit.* **2013**, 647–655, arXiv:1310.1531.
17. Zhou, B.; Lapedriza, A.; Xiao, J.; Torralba, A.; Oliva, A. Learning deep features for scene recognition using places database. In *Advances in Neural Information Processing Systems*; NIPS Foundation Inc.: South Lake Tahoe, UV, USA, 2014; pp. 487–495.
18. Sun, Y.; Wang, X.; Tang, X. Deep learning face representation from predicting 10,000 classes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus Convention Center Columbus, OH, USA, 23–28 June 2014; pp. 1891–1898.
19. Sun, Y.; Chen, Y.; Wang, X.; Tang, X. Deep learning face representation by joint identification-verification. In *Advances in Neural Information Processing Systems*; NIPS Foundation Inc.: South Lake Tahoe, UV, USA, 3–7 December 2014; pp. 1988–1996.
20. Sun, Y.; Wang, X.; Tang, X. Deeply learned face representations are sparse, selective, and robust. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 2892–2900.
21. Shiraga, K.; Makihara, Y.; Muramatsu, D.; Echigo, T.; Yagi, Y. Geinet: View-invariant gait recognition using a convolutional neural network. In Proceedings of the 2016 IEEE International Conference on Biometrics (ICB), Halmstad, Sweden, 13–16 June 2016; pp. 1–8.
22. Wolf, T.; Babaee, M.; Rigoll, G. Multi-view gait recognition using 3D convolutional neural networks. In Proceedings of the IEEE International Conference on Image Processing, Phoenix, AZ, USA, 25–28 September 2016; pp. 4165–4169.
23. Muramatsu, D.; Makihara, Y.; Yagi, Y. View transformation model incorporating quality measures for cross-view gait recognition. *IEEE Trans. Cybern.* **2016**, *46*, 1602–1615.
24. Muramatsu, D.; Makihara, Y.; Yagi, Y. Cross-view gait recognition by fusion of multiple transformation consistency measures. *IET Biom.* **2015**, *4*, 62–73.
25. Kale, A.; Chowdhury, A.K.R.; Chellappa, R. Towards a view invariant gait recognition algorithm. In Proceedings of the IEEE Conference on IEEE Advanced Video and Signal Based Surveillance, Miami, FL, USA, 21–22 July 2003; pp. 143–150.

26. Bodor, R.; Drenner, A.; Fehr, D.; Masoud, O.; Papanikolopoulos, N. View-independent human motion classification using image-based reconstruction. *Image Vision Comput.* **2009**, *27*, 1194–1206.
27. Iwashita, Y.; Baba, R.; Ogawara, K.; Kurazume, R. Person identification from spatio-temporal 3D gait. In Proceedings of the 2010 International Conference on IEEE Emerging Security Technologies (EST), Canterbury, UK, 6-7 September 2010; pp. 30–35.
28. Chen, D.; Cao, X.; Wang, L.; Wen, F.; Sun, J. Bayesian face revisited: A joint formulation. In *European Conference on Computer Vision*; Springer: Berlin, Germany, 2012; pp. 566–579.
29. LeCun, Y.; Boser, B.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.; Jackel, L.D. Backpropagation applied to handwritten zip code recognition. *Neural Comput.* **1989**, *1*, 541–551.
30. Jia, Y.; Shelhamer, E.; Donahue, J.; Karayev, S.; Long, J.; Girshick, R.; Guadarrama, S.; Darrell, T. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM International Conference on Multimedia*; ACM: New York, NY, USA, 2014; pp. 675–678.
31. Learned-Miller, E.; Huang, G.B.; RoyChowdhury, A.; Li, H.; Hua, G. Labeled faces in the wild: A survey. In *Advances in Face Detection and Facial Image Analysis*; Springer: Berlin, Germany, 2016; pp. 189–248.
32. Cao, X.; Wipf, D.; Wen, F.; Duan, G.; Sun, J. A practical transfer learning algorithm for face verification. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 3208–3215.