Improving Cross-View Gait Recognition With Generative Adversarial Networks

Rui Zhang
Key Laboratory of Electromagnetic Space Information of CAS
Information Science and Technology, University of Science and Technology of China
Hefei, Anhui 230027, China
rzhang13@mail.ustc.edu.cn

Dong Yin*
Key Laboratory of Electromagnetic Space Information of CAS
Information Science and Technology, University of Science and Technology of China
Hefei, Anhui 230027, China
yindong@ustc.edu.cn
* Corresponding Author

Zhipeng Zhou
Key Laboratory of Electromagnetic Space Information of CAS
Information Science and Technology, University of Science and Technology of China
Hefei, Anhui 230027, China

Zhi Cao
Key Laboratory of Electromagnetic Space Information of CAS
Information Science and Technology, University of Science and Technology of China
Hefei, Anhui 230027, China

Fanjun Meng
Key Laboratory of Electromagnetic Space Information of CAS
Information Science and Technology, University of Science and Technology of China
Hefei, Anhui 230027, China

Bin Hu
Key Laboratory of Electromagnetic Space Information of CAS
Information Science and Technology, University of Science and Technology of China
Hefei, Anhui 230027, China

***Abstract*—The performance of gait recognition can be obviously affected by view angle variation. In this paper, we present a new method which uses a view transformation generative adversarial networks (GAN) to improve performance of dealing with cross-view gait recognition problem. Our proposed method firstly trains a convolutional neural network (CNN) using gait energy image (GEI) for recognition. Then, a GAN model is taken as a generator to transform gait images with variety view angle to unique side view images. In order to preserve the identification information of generated images, the generated images are input into the fixed pre-trained CNN and recognition loss is used to update generator. Finally, we combine the distance matrix of original and generated image and get final recognition results. We conduct experiments to demonstrate the improvement of adding GAN branch on three popular gait dataset. Experimental results show that our method can achieve state-of-the-art performance.**

Keywords—gait recognition, generative adversarial networks,cross view

I. INTRODUCTION

Gait as a kind of behavioral biometric feature is suitable for human identification at a distance. Compared with other kinds of biometric features such as iris, voice, fingerprints and face, gait has attractive advantages. Gait feature can be more easily captured in long-distance and uncontrolled scenarios. It is also hard to fake and doesn't require high resolution videos. Therefore, gait recognition which focuses on the problem of identifying people by the unique way they walk, is expected to be applied to surveillance and criminal investigation using closed-circuit televisions installed in public like streets, stations and shopping malls. In fact, gait recognition has been already used in practical cases in criminal investigations [1, 2].

Approaches to gait recognition are mainly categorized into two groups. The first one is model-based methods [3, 4, 5] which employ modelling human body and local movement patterns of different body parts. The disadvantage of model-based methods is that they require building an accurate human model and relatively high computational cost. The second category of gait approach is appearance-based method [6, 7] which extract appearance features from human silhouettes. With efficiency of feature extraction, the appearance-based methods are dominant in gait recognition approaches. However, appearance-based methods are largely affected by many potential various factors (e.g., viewpoint, clothing, carriages, and walking speed). Among these sources of variation, view angle, which can reduce the recognition accuracy greatly, is one of the significant issues, because people always change their walking direction depending on the destination in the public.

Recently, there are significant numbers of approaches to cross-view gait recognition based on convolutional neural networks (CNNs). Shiraga et al. [8] designed GEINet, which is a CNN with a single gait feature, that is, the gait energy image (GEI) [9], also known as the averaged silhouette [10] for cross-view gait recognition. Their experimental results demonstrated that GEINet outperformed existing generative and discriminative methods without CNN. Wu et al. [11] provided an extensive empirical evaluation in terms of various scenarios, namely, cross-view and cross-walking-condition, with different preprocessing approaches and network architectures. Zhang et al. [12] developed a Siamese neural network based gait recognition framework to automatically extract robust and discriminative gait features for human identification. Yu et al. [13] proposed a method named as GaitGAN which is based on generative adversarial networks (GAN) [14]. In the proposed method, a GAN model is taken as a regressor to generate invariant gait images that is side view images with normal clothing and without carrying bags.

In this paper, we aim to address the cross-view problem using a more capable gait representation network and a well-designed GAN. We propose a feature learning network based on VGG16 [15] and Hard Triplet Loss [16], which can greatly improve recognition accuracy. Considering variations on view angle, clothing and carriages, we then propose a GAN, which acts as a regressor to transform gait image captured with any source of variation into a unique view image. Combining the feature vector of origin and generative images, our method get a significant performance on cross-view gait recognition.

The rest of the paper is structured as follows. In Section 2, we propose our algorithm by first showing network structure and then explaining objective and training strategy. Section 3 provides the experimental results of the proposed method for gait recognition. Finally, we draw our conclusion in Section 4.

II. Proposed Method

In this section, we first introduce gait energy image, which is input of our network in detail. Subsequently, the CNN architecture and the training method are detailed. Finally, the recognition algorithm, which combines the original and generated image distance matrix is proposed to further improve gait recognition performance.

A. Input data

The first step of appearance-based approaches to gait recognition is usually silhouette extraction. In order to distinguish from person re-identification task, the motivation of silhouette extraction is to avoid being affected by clothes' colors and textures. The gait energy image [9] is a widespread feature representation for gait. It is efficiently produced by averaging all the silhouette of a specific person in the same state (e.g., viewpoint, carriages). Therefore, we take GEI as the input and target image of our method. Additionally, to meet network structure requirements, we set GEI size as 96 × 128 pixels.

B. Network structure

Our network structure can be divided into two part: feature extraction and side view image generator. Different to some methods which train a GAN and use generated and original images to optimize a more robust feature extraction, our method apply a fixed feature extraction to optimize GAN. According to our research, gait conversion is not as accurate as color or shape transformation. Using generated image of GEI will lead feature extraction perform worse.

We firstly design a gait recognition network to get a good feature extraction of gait. We choose VGG16 as backbone, because it has better capacity of feature representation than a small network (e.g., AlexNet [17] and other structures used in [8, 10, 11]). Additionally, comparing to deeper network structures such as Densenet [18], VGG16 avoid over fitting due to the reason that GEI only contains gait information and lacks semantic, color or texture features. As shown in Fig. 1, in our method, we set the input channel of the first convolutional layer to 1. According to our research, we change the dimensionality of the last connected layer from 4096 to 64 to reduce computational cost while accuracies decrease a little. In order to reduce the intra-class distance and increase the inter-class distance, we utilize Hard Triplet Loss [16] to update network parameters.

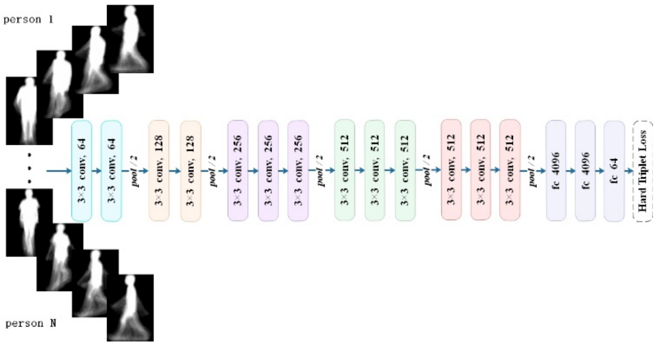

Fig. 1. Gait Recognition Network

According to previous appearance-based approaches, when gallery and query images are from the same angle, the recognition accuracy is usually high. Inspired by the pixel-level domain transfer in Pix2pix [19], we propose cross-view gait GAN to transform the gait data from any view to pictures from the unique angle and preserve identification information. GAN is divided into two networks: generative network and discriminator. As shown in Fig. 2, similar to U-net [20], our generative network mainly contains six convolutional layers (Conv) and six deconvolutional layers (Deconv). Kernel size of all layers is set to 4 and stride of the first four Conv and the latter four Deconv is set to 2, which achieve down-pooling and up-sampling function respectively.

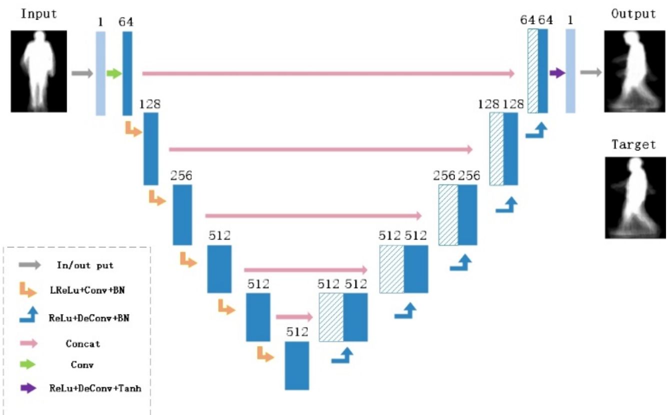

Fig. 2. Cross-View Gait Generative Network

As shown in Fig.3, the discriminator is a real/fake discriminator which is trained to classify whether an image is real or not. If the input image is from a real gait dataset, the discriminator will output 1. Otherwise, it will output 0. Regarding discriminator structure, we use four blocks, each of which consists of a LeakyReLu, a Convolutional layer and a Batch Normalization layer.

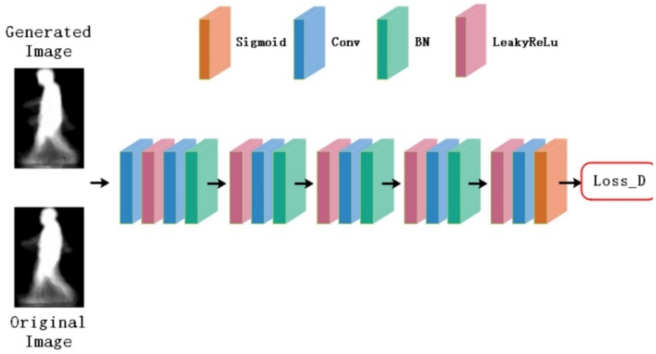

Fig. 3. Discriminator Structure

C. Objective and Training

1） Gait Recognition Objective

Given a probe image and the gallery images, our method uses the hard triplet loss to enlarge the Euclidean distance of images with different labels and reduce distance of images in same identities. At each training batch, we randomly sample a mini-batch of N=KP images, where P and K are the number of identities and images in each category respectively. For each picture, we pick the hardest positive and the hardest negative samples within the batch to form a triplet. Then hard triplet loss is computed by

$$\mathrm{L}_{TH}(\theta;X)=\sum_{i=1}^{P}\sum_{\alpha=1}^{K}\left[\begin{array}{l} m+\max\limits_{p=1\ldots K} D(f_\theta(x_\alpha^i),f_\theta(x_p^i)) \\ -\min\limits_{\substack{j=1\ldots P\\ n=1\ldots K\\ j\neq i}} D(f_\theta(x_\alpha^i),f_\theta(x_n^j)) \end{array}\right]_+ \quad (1)$$

Where f_θ is the output vector of VGG16, and a data point is x_j^i corresponds to the j-th image of the i-th person in the batch. The squared Euclidean distance. m is set to 0.3.

2) GAN Objective

The generator G is trained to produce outputs that cannot be distinguished from real image by an adversarially trained discriminator, D, which is trained to detect the generated image. The objective of GAN can be expressed as

$$\mathrm{L}_{GAN}(G,D)=E_y\left[\log D(y)\right]+E_x\left[\log(1-D(G(x)))\right] \quad (2)$$

Following [19], we mix the GAN objective with L1 distance loss to make output of generator not only fool the discriminator but also near the ground truth.

$$\mathrm{L}_{L1}(G)=E_{x,y}\left[\left\|y-G(x)\right\|_1\right] \quad (3)$$

In order to preserve identification information, we propose an identification loss. At a mini-batch, N=KP images X are firstly input into generator and then generated images $G(X)$ pass through fixed pre-trained VGG16. Generator parameters are optimized by hard triplet loss to reduce the intra-class distance and increase the inter-class distance of generated images.

$$\mathrm{L}_{id}(G)=L_{TH}(\theta;G(X)) \quad (4)$$

Our final GAN objective is

$$G^*=\arg\min_G\max_D L_{GAN}(G,D)+\alpha L_{L1}(G)+\beta L_{id}(G) \quad (5)$$

To sum up, our training process is divided into 3 steps: 1) Produce GEI by averaging person silhouettes from original dataset. 2) Pre-train a VGG16 for gait recognition using gait recognition objective. 3) Train generator and discriminator using GAN objective with VGG16 fixed. For gait recognition VGG16, Stochastic Gradient Descent (SGD) is used to optimize our model. We use an initial learning rate of 0.001 and it divided by 10 every 100 epochs. The weight decay is 0.0005 and the momentum for gradient update is 0.9. For GAN training, we use Adam and set beta1, beta2 and weight decay to 0.5, 0.999 and 0 respectively. Learning rate is 0.0002 for both generator and discriminator. α is set to 0.1. Considering that generator cannot produce good quality fake images and adding hard triplet loss will destroy the balance between generator and discriminator in the early stage of training, we set β to 0 at first 50 epochs, 0.01 at 50~100 epochs, 0.1 at 100~150 epochs and 1 latter.

D. Recognition Algorithm

The purpose of our method is to improve cross-view gait recognition performance by transforming GEIs at all the viewpoints to the GEIs of walking at 90° (side view). In practice, we develop a 6-step gait recognition process as shown in Algorithm 1.

Algorithm 1 Gait recognition process. After 6 steps, the rank-1 gait recognition result is calculated.

Step 1: Gallery images $\boldsymbol{X}_g$ and probe images $\boldsymbol{X}_p$ pass through VGG16 and get feature vector $\boldsymbol{x}_g$ and $\boldsymbol{x}_p$.
Step 2: Gallery images $\boldsymbol{X}_g$ and probe images $\boldsymbol{X}_p$ pass through generator and produce generated images $\widetilde{\boldsymbol{X}}_g$ and $\widetilde{\boldsymbol{X}}_p$.
Step 3: Gallery images $\widetilde{\boldsymbol{X}}_g$ and probe images $\widetilde{\boldsymbol{X}}_p$ pass through VGG16 and get feature vector $\widetilde{\boldsymbol{x}}_g$ and $\widetilde{\boldsymbol{x}}_p$.
Step 4: Calculate the distance matrix $\boldsymbol{M}$ of $\boldsymbol{x}_g$ and $\boldsymbol{x}_p$, the distance matrix $\widetilde{\boldsymbol{M}}$of $\widetilde{\boldsymbol{x}}_g$ and $\widetilde{\boldsymbol{x}}_p$.
Step 5: Calculate the final distance matrix $\boldsymbol{M}_{final}=\boldsymbol{M}+\boldsymbol{\lambda}\widetilde{\boldsymbol{M}}$.
Step 6: According to final distance matrix, get recognition rank list.

After calculating the distance matrix of original and generated images respectively, we add two matrices directly and set $\boldsymbol{\lambda}$ to 0.2. Since the recognition accuracy of the generated images cannot reach the original images, we just fine-tune the original distance matrix using the distance matrix of generated images. Therefore, we set $\boldsymbol{\lambda}$ far less than 1. To some extent, our method can obviously improving the recognition results by applying generative adversarial networks. As shown in Fig. 4, numbers above images represent human ID. The good performance of GAN can improve correct matching from rank-3 of original image to rank-1 of the final result.

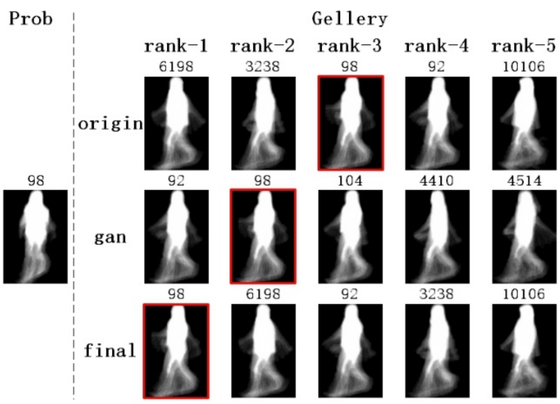

Fig. 4. Recognition Rank List

III. Experimental Results

A. Datasets

To evaluate the proposed method, three datasets are involved. The first one is OU-ISIR Multi-View Large Population with 10307 subjects. The second one is OU-ISIR Large Population Dataset with 4007 subjects and the last dataset is CASIA-B with 124 subjects.

OU-ISIR Multi-View Large Population (OU-MVLP) [21] is the world's largest wide view variation gait database, which includes 10307 subjects (5114 males and 5193 females) from 14 view angles: 0°, 15°, 30°, 45°, 60°, 75°, 90°, 180°, 195°, 210°, 215°, 240°, 255°, 270°.

OU-ISIR Large Population (OULP) Dataset [22] gait dataset is a very large dataset which contains 4007 subjects ranging from 1 to 94 years old. The dataset contains 4 view (55°, 65°, 75°, 85°) and it include two sequences under the normal walking condition.

CASIA-B gait dataset [23] is one of the popular public gait datasets. It consists of 124 subjects (31 females and 93 males) captured from 11 views. The view range is from 0° to 180° with 18° interval between two nearest views. What's more, there are 6 sequences for normal walking (nm), 2 sequences for walking with a bag (bg) and 2 sequences for walking in a coat (cl).

B. Experimental results on OU-MVLP dataset

We follow the evaluation method [21]. 10307 subjects are divided into two disjoint groups of approximately equal size, that is, 5153 training and 5154 testing subjects. We compare the proposed method with 3in+2diff model [21] and evaluate the recognition accuracy for all pairs of the four typical view angles: 0°, 30°, 60°, and 90°. Rank-1 identification rates are shown in Table 1. The results demonstrate that our methods can get an improvement on the dataset with such a large scale and wide view variation. Furthermore, combining GAN, our final method increases all rank-1 accuracies.

TABLE I. Rank-1 accuracy on OU-MVLP.

			Probe				
			0	30	60	90	mean
Gallery	3in+2diff	0	83.9	44.8	16.2	15.6	40.1
		30	51.6	92.1	57.1	41.3	60.5
		60	20.4	55.9	90.3	59.6	41.7
		90	18.9	42.2	60.8	91.9	53.5
	VGG16 (Ours)	0	87.2	60.0	28.1	27.0	50.6
		30	62.3	95.5	71.3	56.8	71.5
		60	35.6	72.0	94.2	71.2	68.3
		90	38.6	62.9	74.5	96.2	68.1
	Final (Ours)	0	**88.0**	**61.7**	**30.8**	**30.2**	**52.7**
		30	**65.8**	**95.6**	**74.3**	**61.1**	**74.2**
		60	**38.7**	**74.7**	**94.5**	**73.9**	**70.5**
		90	**41.5**	**67.3**	**76.5**	**96.2**	**70.4**

C. Experimental results on OULP dataset

In the experiments on OULP dataset, we use a subset of the dataset comprising two walks taken from 1912 subjects to meet the protocols of benchmarks [24]. We train the network using all the view angles for each training subject. During testing, we fix gallery views at 85 degrees and compare all probe view angles. We compare our method with two deep learning approaches: GEINet [8] and PBD [24]. The result is shown in Table II. It can be seen from result that, due to the high accuracy of VGG16, the GAN cannot improve performance on this basis without making recognition worse.

TABLE II. Comparison of rank-1 accuracy on OULP dataset.

Method	Rank-1		
	55	65	75
GEINet [8]	81.4	91.2	94.6
PBD [24]	92.1	96.5	97.8
VGG16 (Ours)	**95.6**	**99.3**	**100**
Final (Ours)	**95.6**	**99.3**	**100**

TABLE III. Comparison of rank-1 accuracy on CASIA-B dataset.

			Probe					
			0	36	72	108	144	180
Gallery NM#1-4	NM #5-6	GaitGAN [13]	47.1	66.1	66.5	65.0	66.1	46.0
		Ours	**57.4**	**75.5**	**73.4**	**70.8**	**76.4**	**58.8**
	BG# 1-2	GaitGAN [13]	33.1	45.7	40.5	39.2	44.6	28.5
		Ours	**45.8**	**57.5**	**53.0**	**50.2**	**61.4**	**46.6**
	CL# 1-2	GaitGAN [13]	11.3	26.7	26.4	24.9	29.2	13.3
		Ours	**31.9**	**40.3**	**38.6**	**37.8**	**40.2**	**30.4**

D. Experimental results on CASIA-B dataset

In our experiments using CASIA-B dataset, the three types of gait data including nm, bg and cl are all involved. We put all sequences of the first 62 subjects into the training set and remaining 62 subjects into the test set. In the test set, the first 4 normal walking (NM#1-4) sequences are used as gallery set. Furthermore, we use NM#5-6, BG#1-2 and CL#1-2 to evaluate as probes respectively. Results are shown in Table 3. We fix probe view angle and average accuracies of different gallery view. Comparing to GaitGAN [13], our method get better performance on each view angle. The results show that besides view variation, our method can deal with different walking conditions. Moreover, improved accuracy also shows that our network can generalize well on such small dataset with only 5000 training images.

IV. Conclusion

In this paper, we propose a new method to improve cross-view gait recognition performance using GAN. We firstly design a CNN model for recognition with VGG16 backbone and Hard Triplet Loss. Then, a GAN is trained to produce images in unique side view from images taken at variety angle. Furthermore, to preserve identification information of generated gait images, we input them into the fixed pre-trained CNN model. Experimental results show that the combination of original and generated images' distance matrices can achieve promising performance for gait recognition.

Acknowledgment

This work was supported by Anhui Key Research and Development Plan (1804a09020049).

References

[1] Bouchrika, I. , Goffredo, M. , Carter, J. , & Nixon, M. . (2011). On using gait in forensic biometrics. Journal of Forensic Sciences,56(4),

8.

[2] Muramatsu, D. , Makihara, Y. , Iwama, H. , Tanoue, T. , & Yagi, Y. . (2013). Gait verification system for criminal investigation. Ipsj Transactions on Computer Vision & Applications, 5, 747-748.

[3] Ariyanto, G., & Nixon, M. S. (2011). Model-based 3D gait biometrics.

[4] Ariyanto, G., & Nixon, M. S. (2012, March). Marionette mass-spring model for 3D gait biometrics. In Biometrics (ICB), 2012 5th IAPR International Conference on (pp. 354-359). IEEE.

[5] Cunado, D., Nixon, M. S., & Carter, J. N. (2003). Automatic extraction and description of human gait models for recognition purposes. Computer Vision and Image Understanding, 90(1), 1-41.

[6] Ben, X., Zhang, P., Meng, W., Yan, R., Yang, M., Liu, W., & Zhang, H. (2016). On the distance metric learning between cross-domain gaits. Neurocomputing, 208, 153-164.

[7] Yu, S., Wang, Q., Shen, L., & Huang, Y. (2016, December). View invariant gait recognition using only one uniform model. In Pattern Recognition (ICPR), 2016 23rd International Conference on (pp. 889-894). IEEE.

[8] Shiraga, K., Makihara, Y., Muramatsu, D., Echigo, T., & Yagi, Y. (2016, June). Geinet: View-invariant gait recognition using a convolutional neural network. In Biometrics (ICB), 2016 International Conference on (pp. 1-8). IEEE.

[9] Han, J., & Bhanu, B. (2006). Individual recognition using gait energy image. IEEE Transactions on Pattern Analysis & Machine Intelligence, (2), 316-322.

[10] Liu, Z., & Sarkar, S. (2004, August). Simplest representation yet for gait recognition: Averaged silhouette. In Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on (Vol. 4, pp. 211-214). IEEE.

[11] Wu, Z., Huang, Y., Wang, L., Wang, X., & Tan, T. (2017). A comprehensive study on cross-view gait based human identification with deep cnns. IEEE Transactions on Pattern Analysis & Machine Intelligence, (2), 209-226.

[12] Zhang, C., Liu, W., Ma, H., & Fu, H. (2016, March). Siamese neural network based gait recognition for human identification. In Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on (pp. 2832-2836). IEEE.

[13] Yu, S., Chen, H., Reyes, E. B. G., & Poh, N. (2017, July). GaitGAN: Invariant Gait Feature Extraction Using Generative Adversarial Networks. In CVPR Workshops (pp. 532-539).

[14] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. In Advances in neural information processing systems (pp. 2672-2680).

[15] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[16] Hermans, A., Beyer, L., & Leibe, B. (2017). In defense of the triplet loss for person re-identification. arXiv preprint arXiv:1703.07737.

[17] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

[18] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017, July). Densely connected convolutional networks. In CVPR (Vol. 1, No. 2, p. 3).

[19] Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. arXiv preprint.

[20] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

[21] Takemura, N., Makihara, Y., Muramatsu, D., Echigo, T., & Yagi, Y. (2017). On input/output architectures for convolutional neural network-based cross-view gait recognition. IEEE Transactions on Circuits and Systems for Video Technology.

[22] Iwama, H., Okumura, M., Makihara, Y., & Yagi, Y. (2012). The ou-isir gait database comprising the large population dataset and performance evaluation of gait recognition. IEEE Transactions on Information Forensics and Security, 7(5), 1511-1521.

[23] Yu, S., Tan, D., & Tan, T. (2006, August). A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In Pattern Recognition, 2006. ICPR 2006. 18th International Conference on (Vol. 4, pp. 441-444). IEEE.

[24] Sokolova, A. , & Konushin, A. . (2017). Pose-based deep gait recognition.