Gait Virtual Sample Generation Method Based on CNN and DLTL: Postprint

Authors: Zhi Shuangshuang, Zhao Qinghui, Jin Dahai, Tang Jin

Date: 2018-11-29T00:00:00+00:00

Abstract

To address the pressing small-sample problem in gait recognition within counter-terrorism and security fields, we propose a gait virtual sample generation method based on deep convolutional neural network CNN (convolutional and neural network) and DLTL (dual learning and transfer learning). First, low-level responses from a VGG19-based deep convolutional neural network model are employed to extract gait style feature maps, followed by style training on these maps using adversarial network-based dual learning DL (dual learning) to obtain a style feature model; second, high-level responses from the VGG19 model are utilized to extract gait content feature maps, which then learn the style features from the style feature model; finally, transfer learning TL (transfer learning) is applied to acquire gait virtual offset samples. Experimental results indicate that although the overall style of gait virtual samples generated through DLTL style learning undergoes transformation, the human gait features remain preserved, effectively expanding the small-sample capacity; when the number of virtual samples increases to a certain threshold, the gait recognition rate demonstrates improvement. Comparative experiments with existing gait virtual sample generation methods reveal that the proposed algorithm surpasses existing approaches, capable of generating numerous virtual samples while stably enhancing the recognition rate of gait recognition.

Full Text

Preamble

Gait Virtual Sample Generation Method Based on CNN and DLTL

Zhi Shuangshuang[1], Zhao Qinghui[2], Jin Dahai[1], Tang Jin[2]†
(1. Engineering Training Center, Xi' an Polytechnic University, Xi' an Shannxi 710048, China;

2. School of Information Science & Engineering, Central South University, Changsha 410083, China)

Abstract: To address the small sample problem in gait recognition for counterterrorism and security applications, this paper proposes a novel gait virtual sample generation method based on deep CNN (Convolutional Neural Network) and DLTL (Dual Learning and Transfer Learning). First, low-level responses of a VGG19-based deep convolutional neural network model extract gait style feature maps, which are then trained using dual learning (DL) based on adversarial networks to obtain a style feature model. Second, high-level responses of the VGG19 model extract gait content feature maps, which then learn the style features from the style feature model. Finally, transfer learning (TL) is employed to obtain gait virtual migration samples. Experimental results demonstrate that while the overall style of gait virtual samples generated through DLTL style learning changes, the human gait characteristics remain unchanged, effectively expanding small sample capacity. When the number of virtual samples increases to a certain quantity, gait recognition rate improves. Comparative experiments with existing gait virtual sample generation methods show that the proposed algorithm outperforms existing methods, capable of generating virtual samples in large quantities while steadily improving gait recognition accuracy.

Keywords: gait recognition; CNN; DLTL; virtual sample; gait recognition rate

0 Introduction

Gait recognition is a method for extracting biometric features from image sequences of pedestrians walking. Compared with other biometric recognition methods such as DNA, fingerprint, iris, and 2D/3D face recognition, gait recognition offers several advantages: it does not require subject cooperation, can operate at long distances and with low image quality, and is difficult to disguise or conceal [1]. Consequently, identity recognition based on gait features has become a research hotspot in recent years [2,3], with applications in security and counterterrorism receiving significant attention from scholars worldwide [4].

Gait recognition methods can be categorized based on whether they require human pose parameters into model-free (appearance-based statistical methods) and model-based approaches. Model-free gait recognition methods directly extract and match features from gait silhouettes without constructing human model structure data. These can be further subdivided based on silhouette data into 2D, 2.5D, or 3D gait silhouette recognition methods. Model-based approaches, which utilize human models for gait recognition, can employ either 2D models (stick figure models, articulated models, etc.) or 3D human models (3D ellipsoid models, 3D joint skeleton models). While model-free methods can directly use gait silhouette data, model-based methods have slightly higher requirements for image pixels and clarity. Moreover, constructing human models, particularly 3D human models, involves higher computational complexity than model-free approaches. Currently, model-free research represents an important direction in gait recognition.

Model-free gait recognition methods extract statistical information from gait silhouettes within a cycle and directly match contour data reflecting body shape and motion characteristics. The most fundamental matching approach involves synchronizing gait silhouettes according to different temporal sequences and postures, then directly comparing the similarity of gait silhouettes with the same posture across different individuals using metrics such as Euclidean distance or cosine distance. Another common method calculates the mean of gait silhouette sequences to form the Gait Energy Image (GEI) [5], which is then used for matching. Numerous related energy map features have been derived from GEI, including Motion Silhouette Image (MSI) [6], Gait Flow Image (GFI) [7], Color Gait Image (CGI) [8], Frame Difference Energy Image (FDEI) [9], Gait Energy Volume (GEV) [10], Depth Gradient Histogram Energy Image (DGHEI) [11], and Curve-based Gait Color Energy Image (CGCEI) [12]. GEI features are particularly notable for their low computational complexity and effective suppression of image distribution noise through silhouette averaging.

Despite significant progress, model-free gait recognition still faces challenges including contour loss and human shadows caused by multiple covariates, viewpoint differences in gait recognition, and the small sample problem. Due to long gait cycles and difficulties in data collection, acquiring large quantities of gait samples is challenging. Current public gait databases contain only a limited number of samples, with few gait samples per individual, representing the concentrated manifestation of the gait recognition data volume problem.

For small sample problems in image recognition, scholars have proposed methods to generate virtual training samples from given small-scale real samples based on prior knowledge [13]. Existing research includes approaches using BPNN and mega-trend diffusion techniques [14], and methods utilizing original sample distribution functions [15]. Reference [16] proposed using Decentralized Neural Networks (DNN) to generate virtual samples, demonstrating that DNN possesses stronger predictive performance than BPNN. References [17-19] presented virtual sample generation algorithms based on Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Monte Carlo combined with PSO. To address the small sample problem in gait recognition, current gait virtual sample generation methods based on energy maps primarily include synthetic gait templates [20] and adversarial network-based [21] virtual sample generation methods. However, synthetic gait template methods compromise the integrity of energy map contours, preventing large-scale virtual sample generation. Meanwhile, adversarial network-based methods suffer from uncertainty caused by random noise, resulting in poor robustness and instability of generated gait virtual samples.

To better solve the small sample problem in gait recognition, this paper proposes a virtual sample generation method based on deep CNN and DLTL style learning. The method first extracts gait style feature maps using low-level responses of a deep convolutional neural network model, then trains these style feature maps using dual learning based on adversarial networks to obtain a style feature model. Subsequently, gait content feature maps are extracted using high-level responses of the VGG19 model, which then learn the style features from the style feature model. Finally, transfer learning is applied to obtain gait virtual migration samples. This approach can effectively expand gait sample capacity.

1 Gait Virtual Sample Generation Method Based on CNN and DLTL Style Learning

The proposed method originates from practical applications where gait human silhouette segmentation cannot achieve 100% accuracy, and environmental covariate changes affect segmentation position and effectiveness. Based on this observation, we propose a gait virtual sample generation method using CNN and DLTL style learning to alter the style background and climate of gait recognition samples, thereby achieving large-scale generation of gait virtual samples.

The main idea is as follows: First, extract gait style feature maps using low-level responses of a VGG19-based deep convolutional neural network model, then perform style training on these feature maps using dual learning (DL) based on adversarial networks to obtain a style feature model. Second, extract gait content feature maps using high-level responses of the VGG19 model, then enable these content feature maps to learn the style features from the style feature model. The structure of this method is shown in Figure 1 [Figure 1: see original paper].

1.1 VGG19 Model Network Architecture

Deep CNNs map images to a lower-dimensional space through multiple convolutional transformations and downsampling operations, extracting sparse features from images. Due to weight sharing, the number of neurons and parameters is reduced, making training easier. In the current big data environment, sufficiently deep CNN models have become important tools for solving numerous problems in computer vision [22].

The VGGnet proposed by Karen et al. employs 3$×$3 convolution kernels with pooling layers inserted after several convolutional layers. Increasing network depth benefits image classification accuracy, but excessive layers can cause network degradation [23]. VGG19 comprises 16 convolutional layers (Conv1_1-Conv5_4), 5 pooling layers (pool1-pool5), and 3 fully connected layers (Fc6-Fc8), using Rectified Linear Units (ReLU) as activation functions. The VGG19 structure is shown in Figure 2 [Figure 2: see original paper].

1.2 Extraction of Image Content and Style Features

Image features can be divided into content and style, which can be represented using responses from convolutional layers in deep CNN models. Low-level responses in deep CNN models describe image style, while high-level responses describe image content [24].

1.2.1 Image Content Feature Map Extraction To extract image content feature maps, we first generate an image of the same size as the target image using random noise and input it into the CNN model. The response at the l-th convolutional layer of this network model can be denoted as , with dimensions , where represents image height, represents image width, and represents the number of pixel points. Similarly, the target image can be input into the network model to obtain the feature response at this layer.

Since we expect the image content of and to be consistent, the objective function minimizes the L2 norm error:

This error function can be used to compute derivatives for each element of the responses at this layer:

Using the chain rule, we can compute the error derivative with respect to the input image elements to obtain , which is the classic backpropagation algorithm. We use to update to obtain a new l-th layer response that is closer to the target image response , thus extracting the content feature map of the target image.

1.2.2 Image Style Feature Map Extraction To extract image style feature maps, we first construct a feature matrix of size . is obtained by computing the l-th layer response with positional information eliminated, serving as a style descriptor. The element at position (x, y) represents the correlation between the x-th and y-th channel responses. For the style of the target image at each layer, the objective function can be obtained by minimizing the following error function:

We can derive the derivative of this error function with respect to the responses at each layer:

Similarly, we can derive through backpropagation to update , making it closer to the style of , thus extracting the image style feature map.

The classification network model selected in this paper is the VGG19 deep convolutional neural network model. Style feature maps are extracted from the conv1_1, conv2_1, conv3_1, conv4_1, and conv5_1 layers of this model, while gait content feature maps are extracted from the conv4_2 layer.

1.3 Style Training Based on Dual Learning (DL)

The fundamental idea of dual learning (DL) is to construct a closed-loop feedback system with two dual tasks that can obtain feedback information from unlabeled data to improve the training models for dual tasks. Assuming two image sets with different styles, X and Y, we can train two mappings G: X→Y and F: Y→X, where G and F are mutually convertible. Dual learning simultaneously trains both G and F mappings and adds a cycle consistency loss function to encourage and .

As shown in Figure 3 [Figure 3: see original paper], Figure 3(a) illustrates the mapping functions G: X→Y and F: Y→X, along with their corresponding adversarial network discriminators DX and DY. DY encourages the G mapping to transform X into a virtual style indistinguishable from real style samples, and vice versa. To further regularize the mappings, loss functions are introduced. Figure 3(b) shows the forward loss function: x→G(x)→F(G(x))   x, while Figure 3(c) shows the backward loss function: y→F(y)→G(F(y))   y.

For the forward and backward loss functions, we adopt the loss function description from the least squares adversarial network model for mapping function G: X→Y and discriminator DY, defined as follows:

The complete forward loss function for dual learning is:

The complete backward loss function for dual learning is:

The complete incentive regulation loss function for dual learning is:

where Px is the true distribution of x, and Py is the true distribution of y. In equation (7), the forward loss function is minimized when DY encourages the G mapping to transform X into something indistinguishable from Y. Similarly, in equation (8), the backward loss function is minimized when Dx encourages the F mapping to transform Y into something indistinguishable from X.

The style training approach uses a generator to convert style features from set X to set Y according to the G mapping relationship, employing discriminator DY to provide feedback incentives for generated samples, aiming to produce samples similar to real style Y. Similarly, it uses a generator to convert style features from set Y to set X according to the F mapping relationship, employing discriminator DX to provide feedback incentives, aiming to produce samples similar to real style X. By combining both through the loss function, we construct a dual learning model capable of mutual style conversion to generate diverse style transformation feature models. In this paper, dual learning is used for summer and winter style training to obtain winter and summer style feature models.

1.4 Style Learning Based on Transfer Learning (TL)

The essence of transfer learning (TL) is to leverage existing knowledge to learn new knowledge, with the core being the identification of similarities between existing and new knowledge [25]. The style learning method based on TL in this paper involves enabling gait content feature maps to learn from style feature models. The style learning algorithm randomly generates a white noise image as input, keeps the weights of the VGG19 classification network model unchanged, computes the loss values of content feature maps and style pre-trained models at each convolutional layer, and then uses standard backpropagation to update the learning process. It adjusts the input image to minimize both content and style loss values, allowing the gait content feature map to learn different environmental styles from the style pre-trained model through parameter adjustment, generating the final gait virtual sample. The style learning loss function is shown in equation (10):

where represents the style learning loss function, represents the style feature loss function, and represents the gait content feature loss function. denotes the style feature map, denotes the gait content feature map, and denotes the final virtual sample generated. α and β are adjustment parameters representing the weighting factors of the style model and content features in the generated virtual sample, respectively, with $0 \leq \alpha \leq 1$, $0 \leq \beta \leq 1$, and $\alpha + \beta = 1$. When α varies from 10^{-6} to 10^{-1}, larger α values make the final generated image closer to the style image, while smaller α values make it closer to the content image. Therefore, α should not be too large. When $\alpha = 10^{-4}$ and $\beta = 1$ - 10^{-4}, the final generated image is closer to the actual real image, better achieving the desired objective.

The image content feature loss function is:

where l represents the layer number, and F and P represent the feature maps of and in the convolutional neural network, respectively.

The style feature loss function at each layer in the VGG19 model is:

where represents the original image, represents the final generated image, and Al and Gl represent the outputs of the original image and final virtual sample at layer l of the neural network, respectively.

The total style feature loss is:

where wl represents the weight of each layer' s loss in the overall loss.

2 Experiments

The experiments investigate the gait virtual sample generation method based on CNN and DLTL style learning on gait images from the GaitDatabase at 0°, 45°, and 90° orientations. The large-scale original dataset for style feature training uses 1,273 summer images and 854 winter images from Yosemite, with the classification network model being the VGG19-based deep convolutional neural network. The deep learning framework is Torch, the programming language environment is Lua5.1, and the GPU is GTX1060.

The experimental objective is to train summer and winter style features from the Yosemite large-scale original dataset to obtain summer and winter feature models. The style training loss functions based on dual learning during the summer and winter feature training process are shown in Figure 4 [Figure 4: see original paper], where green represents the summer style training loss function, purple represents the summer real sample style function, white represents the winter style training loss function, and blue represents the winter real sample style function.

The loss functions demonstrate that the image style training process gradually stabilizes, with the model training essentially converging at epoch = 40. Initially, the trained style differs significantly from the real sample style, but as the loss function provides feedback regulation, the generated style samples from one class become increasingly similar to the real style samples of the other class, eventually reaching a point where the other class' s discriminator cannot distinguish them.

Using summer and winter feature models based on TL style learning, we perform environmental background style migration. As shown in Figures 5 [Figure 5: see original paper] and 6 [Figure 6: see original paper], the original image → winter → summer DLTL style learning transforms the gait human background environment tones to cool colors, exhibiting winter characteristics with white snow spots on the floor and human body. Subsequent migration to summer background style changes the tones to warm colors with sunlight spots on the ground. Conversely, original image → summer → winter DLTL style learning transforms the background to warm summer tones with sunlight, then to winter style with cool tones and white snow spots.

Figure 7 [Figure 7: see original paper] shows that although the overall style of gait virtual samples generated through DLTL style learning changes, the gait human content remains unchanged. The human gait features are preserved without loss of pedestrian silhouette information; only the environment for gait recognition experiments changes, with variations in color, tone, and saturation. This achieves the goal of increasing within-class samples, expanding small sample capacity, and providing diversity through different background environments, thereby reducing overfitting probability for subsequent work.

The training samples consist of 3 energy maps generated from 30 individuals at 0° orientation, totaling 1,000 original RGB images. Using the CNN and DLTL style learning-based virtual sample generation method, each individual' s virtual sample quantity is increased to 5, 10, 15, 20, 25, 30, 35, 40, 45, and 50. Feature extraction employs the commonly used GEI-based gait feature extraction method, with Euclidean distance as the classifier. PCA dimensionality reduction is applied with a contribution rate above 95%. Test samples consist of 150 real samples collected from these 30 individuals at 0° orientation. Pattern matching is performed using the Euclidean distance classifier, with multiple measurements of Euclidean distance differences averaged to obtain the gait recognition rate shown in Figure 9 [Figure 9: see original paper].

Figure 9 demonstrates that as the number of DLTL style learning virtual samples increases, the recognition rate improves. After the virtual sample quantity reaches a certain number, the recognition rate shows improvement, with the trend suggesting possible saturation at higher quantities. Experimental verification confirms that DLTL style learning not only expands within-class sample numbers but also mitigates the small sample problem in gait recognition to some extent.

Table 1 shows the gait recognition performance of three virtual sample generation methods using different classifiers, with test samples being 150 real samples from 30 individuals at 0° orientation. The results indicate that DLTL style learning outperforms the other two methods.

Table 1. Recognition rate of different classifiers

Virtual Sample Generation	Synthetic Gait Template	DLTL Style Learning
	86.67%	88.67%
	90.67%	86.00%
	89.33%	90.00%
	84.67%	88.00%
		91.33%

Figure 8 [Figure 8: see original paper] compares two virtual energy map sample generation algorithms. Both can improve gait recognition rates within a certain sample quantity range. The synthetic gait template method shows an initial increase followed by a decrease because cropping the foot area excessively compromises energy map contour integrity, preventing large-scale virtual sample generation. The adversarial network-based method exhibits oscillating recognition rates as virtual sample quantity increases, indicating poor robustness and instability that makes it difficult to apply in practical gait recognition systems. This is due to uncertainty caused by random noise, which introduces new limitations to gait virtual sample construction despite enabling large-scale generation.

Table 2. Comparison of three algorithms for gait virtual sample generation

Virtual Sample Generation	Stability	Distortion	Large-scale Generation	Real Sample Enhancement
Synthetic Gait Template	Poor	Yes	No	Initial increase then decrease
Adversarial Network	Poor	No	Yes	Oscillating
DLTL Style Learning	Good	No	Yes	Steady improvement

Table 2 presents an experimental comparison of three gait virtual sample generation algorithms. The CNN and DLTL style learning-based method is optimal among the three, as it not only steadily improves real sample recognition rates but also addresses the small sample problem in gait recognition to a certain extent.

3 Conclusion

This paper proposes a gait virtual sample generation method based on deep convolutional neural networks and DLTL style learning, which effectively improves the small sample problem in gait recognition and outperforms existing gait virtual sample generation methods. Experimental results demonstrate that the method can generate numerous virtual samples with good robustness. The proposed CNN and DLTL style learning-based gait virtual sample method can improve gait recognition rates. Future research will focus on extracting more complex style feature models, enhancing the efficiency of gait content feature map migration learning from style feature models, and efficiently obtaining gait virtual samples with more diverse styles.

References

[1] Iwama H, Muramatsu D, Makihara Y, et al. Gait-based person-verification system for forensics [C]// Proc of the 5th IEEE International Conference on Biometrics: Theory, Applications and Systems. Washington DC: IEEE Computer Society, 2012: 113-120.

[2] Tang Jin, Luo Jian, Tjahjadi T. Robust arbitrary-view gait recognition based on 3d partial similarity matching [J]. IEEE Trans on Image Processing, 2017, 26 (1): 7-22.

[3] Luo Jian, Tang Jin, Tjahjadi T, et al. Robust arbitrary view gait recognition based on parametric 3D human body reconstruction and virtual posture synthesis [J]. Pattern Recognition, 2016 (60): 361-377.

[4] Bouchrika I, Goffredo M, Carter J, et al. On using gait in forensic biometrics [J]. Journal of Forensic Sciences, 2011, 56 (4): 882-889.

[5] Han Ju, Bhanu B. Individual recognition using gait energy image [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2006, 28 (2): 316-322.

[6] Lam T H. W, Lee R S T. A new representation for human gait recognition: Motion silhouettes image (MSI) [M]// Advances in Biometrics. Berlin: Springer, 2006: 612-618.

[7] Lam T H W, Cheung K H, Liu J N K. Gait flow image: A silhouette-based gait representation for human identification [J]. Pattern Recognition, 2011, 44 (4): 973-987.

[8] Wang Chen, Zhang Junping, Wang Liang, et al. Human identification using temporal information preserving gait template [J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 2012, 34 (11): 2164-2176.

[9] Chen Changhong, Liang Jimin, Zhao Heng, et al. Frame difference energy image for gait recognition with incomplete silhouettes [J]. Pattern Recognition Letters, 2009, 30 (11): 977-984.

[10] Sivapalan S, Chen D, Denman S, et al. Gait energy volumes and frontal gait recognition using depth images [C]// Proc of International Joint Conference on Biometrics. Piscataway, NJ: IEEE Press, 2011: 1-6.

[11] Hofmann M, Bachmann S, Rigoll G. 2. 5D gait biometrics using the depth gradient histogram energy image [C]// Proc of the 5th IEEE International Conference on Biometrics: Theory, Applications and Systems. Washington DC: IEEE Computer Society, 2012: 399-403.

[12] Tang Jin, Luo Jian, Tjahjadi T, et al. 2. 5D multi-view gait recognition based on point cloud registration [J]. Sensors, 2014, 14 (4): 6124-6143.

[13] Du Yong, Wang Yu. Generating virtual training samples for sparse representation of face images and face recognition [J]. Journal of Modern Optics, 2016, 63 (6): 536-544.

[14] Li D C, Wu C S, Tsai T I, et al. Using mega-trend-diffusion and artificial samples in small data set learning for early flexible manufacturing system scheduling knowledge [J]. Computers and Operations Research, 2007, 34 (4): 966-982.

[15] Li D C, Lin Yaosan. Using virtual sample generation to build up management knowledge in the early manufacturing stages [J]. European Journal of Operational Research, 2006, 175 (1): 413-434.

[16] Huang Chongfu, Moraga C. A diffusion-neural-network for learning from small samples [J]. International Journal of Approximate Reasoning, 2004, 35 (2): 137-161.

[17] Li D C, Wen I H. A genetic algorithm-based virtual sample generation technique to improve small data set learning [J]. Neurocomputing, 2014 (143): 222-230.

[18] Chen Zhongsheng, Zhu Bao, He Yanlin, et al. A PSO based virtual sample generation method for small sample sets: Applications to regression datasets [J]. Engineering Applications of Artificial Intelligence, 2017 (59): 236-243.

[19] Gong Hongfei, Chen Zhongsheng, Zhu Qunxiong, et al. A Monte Carlo and PSO based virtual sample generation method for enhancing the energy prediction and energy optimization on small data problem: an empirical study of petrochemical industries [J]. Applied Energy, 2017, 197: 405-415.

[21] Wang Kunfeng, Gou Chao, Duan Yanjie, et al. Generative adversarial networks: the state of the art and beyond [J]. Acta Automatica Sinica, 2017, 43 (3): 321-332.

[22] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [EB/OL]. (2015-04-10) [2015-07-19]. http:///arxiv.org/pdf/1409.1556v6.

[23] Zhou Junyu, Zhao Yanming. Application of convolution neural network in image classification and object detection [J]. Computer Engineering and Applications, 2017, 53 (13): 34-41.

[25] Zhang Bo, Shi Zhongzhi, Zhao Xiaofei, et al. A transfer learning based on canonical correlation analysis across different domains [J]. Chinese Journal of Computers, 2015, 38 (7): 1326-1336.

[26] Wang Chenchen, Wang Yalin, Ge Zhongqin, et al. Convolutional neural network-based Chinese ink-painting artistic style extraction [J]. Journal of Graphics, 2017, 38 (5): 754-759.

Note: Figure translations are in progress. See original paper for figures.