An RGB Gait Anonymization Model for Low-Quality Silhouettes

Ngoc-Dung T. Tieu[1], Huy H. Nguyen[1], Fuming Fang[2], Junichi Yamagishi[1,2], Isao Echizen[1,2,3]
[1]SOKENDAI, Kanagawa, Japan; [2]National Institute of Informatics, Tokyo, Japan
[3]The University of Tokyo, Japan

***Abstract*—Gait anonymization while maintaining naturalness is used for protecting a person's identity against gait recognition systems when a video of the person walking is uploaded to social media. There has been some research on gait anonymization, but only for high-quality silhouette gaits. We present an RGB gait anonymization model for low-quality silhouette gaits that can generate natural, seamless anonymized gaits for which the original silhouettes cannot be extracted correctly. Our model includes two main networks. The first one, a deep convolutional generative adversarial network, is used to anonymize the original gait by adding to it a random noise vector. By training on high-quality silhouette data, this network can generate a high-quality anonymized silhouette sequence from a low-quality silhouette one. Restricting its input to binary silhouette sequences instead of color gaits forces it to focus on anonymizing the gait rather than changing body color. The second main network, which follows the first one, colorizes the anonymized silhouette sequence generated by the first network by using the color of the original gait. Evaluation in terms of success rate and naturalness demonstrated that our model can anonymize gaits while maintaining naturalness.**

I. Introduction

Gait anonymization while maintaining naturalness is becoming increasingly important due to the growth of social media and substantial improvements in gait recognition systems [1], [2]. Previous research on gait anonymization used gait datasets for which the silhouettes could be extracted correctly. That is, high-quality silhouette gaits were used. However, there are many cases in which silhouettes cannot be extracted correctly (e.g., the silhouette is missing one or more body parts) for some video frames or for the whole sequence. We have developed a model for anonymizing such low-quality silhouette gaits.

There have been several studies on anonymizing gait information. A naive approach is to visually obscure the region containing the person by blocking out, blurring, or pixelization [3] to frustrate viewers, but this makes the output video unnatural. Chen et al. [4] anonymized the whole body by using edge motion history images, with the person obscured in the final image. To the best of our knowledge, there have been only two studies on using gait contours for anonymization [1], [2]. One focused on binary gaits, and the other focused on color gaits. These studies achieved anonymization with gait naturalness when high-quality silhouette gaits were used. When low-quality ones were used, the anonymized gaits looked unnatural. Fig. 1 shows a sample of an anonymized gait obtained by low-quality silhouette gaits generated using the spatio-temporal generative adversarial network (ST-GAN) of Tieu et al. [2] and using our proposed model.

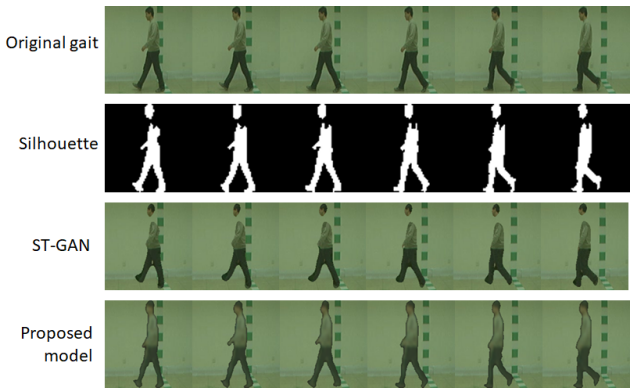


Fig. 1: Anonymized gait generated from low-quality silhouette gait.

Our proposed gait anonymization model solves the problem of using low-quality silhouette gaits, which was not completely solved by using contour-based approaches. It uses two independently trained networks. The first one anonymizes the original gait, and the second one then colorizes the anonymized gait by using the color in the original gait image. The first network uses the binary silhouette sequence of the original gait and a random noise vector to remove the identity of the original gait and outputs an anonymized silhouette sequence. Using a binary silhouette sequence rather than a color silhouette sequence makes the network focus on anonymizing the gait rather than changing its color. By using the deep convolutional generative adversarial network and training on a high-quality silhouette dataset, this network can generate a high-quality anonymized silhouette sequence from a low-quality silhouette one. The second network takes the output of the first network and uses the original RGB gait image to restore the color of the original gait image as much as possible.

We evaluated the proposed model on the CASIA-B gait dataset [5] using two metrics: success rate and naturalness. The success rate was calculated as in previous studies [1], [2], and gait naturalness was estimated by using the mean opinion score (MOS). We used the gait recognition system presented by Zheng et al. [6] as a black-box system for measuring the success rate. The success rate with our model was significantly higher than that with Tieu et al.'s model for the side views, and the MOS score for anonymized gaits generated from low-

quality silhouette gaits ranged from 2.70 to 3.37.

Our contributions are as follows:

(1) We present a network that can generate natural anonymized gaits from low-quality silhouette gaits.

(2) We present a network that can transfer the color of the original gait image to the silhouette sequence of the anonymized gait.

(3) The gait anonymization success rate for side views with our model was significantly higher than that with the state-of-the-art method.

After discussing related work in Section 2, we describe the proposed model in Section 3 and present the experimental results in Section 4. We summarize the key points and mention future work in Section 5.

II. Related Work

A. Gait Recognition

Gait recognition systems are aimed at recognizing people on the basis of their gait, which represents the manner and pattern of a person walking, as observed in a video. They estimate the identity of a probe sample, given a gallery samples registered to the system [7], [8], [9]. The two main approaches to gait recognition in the current state-of-the-art are model-free and model-based methods. The model-based methods use information about body parts (arms, legs, limbs, thighs, etc.) while the model-free ones use a single template computed from a sequence of silhouettes. Among the model-free methods, which are more relevant to our work, Zheng et al. [6] proposed a robust, easy-to-implement, rapid method [1]. We thus used their method as a black-box system for evaluating the success rate of our proposed model. Their method uses gait energy images (GEIs), which are obtained by averaging all the silhouettes [7], as the gait feature for their system . To address the problem of multi-view gait recognition, they transform the gait feature in the probe into that in the gallery view. Fig. 2 shows a sample silhouette sequence and its GEI.

Fig. 2: Gait energy image (GEI): images on left are silhouette sequence; image on right is sequence GEI.

B. Gait Anonymization

The human gait has become a privacy concern due to rapid advances in gait recognition research [10]. The identities of people in a video can now be revealed by gait recognition systems. However, there have been few studies on anonymizing gait information. Agrawal et al. conducted research on body pixelation and on blurring by applying a blurring filter to the region containing the person [3]. Another anonymization method uses an edge motion history image [4] to blur the person's entire body. Such methods are not suitable for uploading and sharing video on social media due to the unnatural appearance of the final image.

As mentioned in the Introduction, the only two studies we know of [1], [2] that were aimed at anonymizing gaits while maintaining naturalness focused on gait contour used either binary gaits or color gaits. The methods developed achieve anonymization with gait naturalness when high-quality silhouette gaits are used. However, the results look unnatural when low-quality ones are used.

C. Generative Adversarial Example

Deep neural networks have achieved success in a wide range of areas [11], [12], especially in media object generation. Among the various types of generative models, the generative adversarial network (GAN) proposed by Goodfellow et al. [13] has attracted much attention. Its development inspired research in various areas such as style transfer and texture synthesis [14], [15], image editing [16], [17], and motion generation [18], [19].

While various GAN-based models have been reported for high-quality image synthesis [20], [21], only a few GAN-based model have been introduced for video generation. Vondrick et al. reported a model for video generation (VGAN) [22] that can produce a video from random noise vectors sampled from a Gaussian distribution. The vectors are input to two generators, one for background generation and the other for foreground generation. Holden et al. [18] trained an autoencoder network to represent human motions at the hidden units of the network and then stacked another network on the top of the autoencoder to create a new motion sequence. They focused on the human skeleton while our objective is to generate whole body sequences. In other related research [23], a Siamese structure network and a GAN model were used to produce a video of the motions of the entire human body as derived from a human skeleton sequence and a single whole body image.

III. Proposed Model

A. Model Overview

Before overviewing our proposed model, we briefly review the research on color gait anonymization reported by Tieu et al. [2], which is closely related to the research reported here. The method developed in that study uses an encoder-decoder structure generator to produce binary anonymized gaits from binary original gaits by adding random noise. To achieve gait naturalness, they stacked a spatial discriminator and a temporal discriminator on top of the gait generator. To colorize the anonymized gaits, they used a colorization method in which the color of the nearest pixel is used. Although their method achieves anonymization with naturalness when using high-quality silhouette gaits, it is unable to achieve naturalness when using low-quality ones. This is because the encoder-decoder structure generator attempts to generate output similar to the input. The method thus modifies the contour of the original gait but does not generate the body parts missing in the original image. To overcome this problem, we have developed

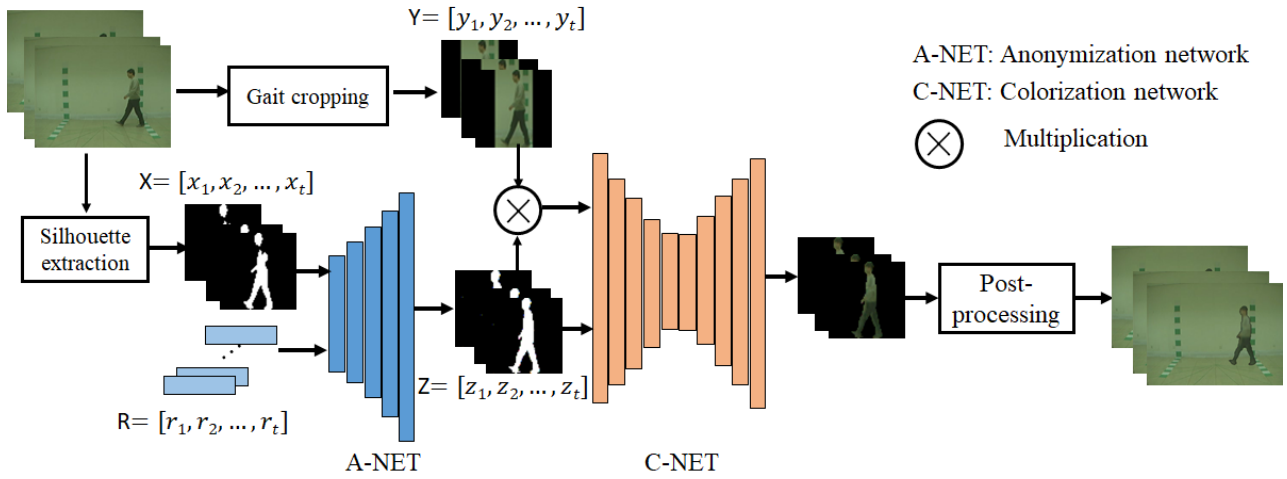


Fig. 3: Overview of proposed model.

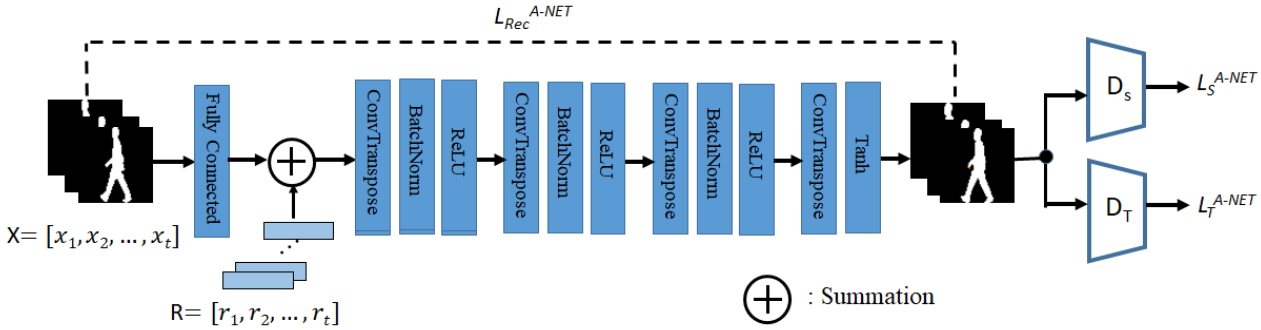


Fig. 4: Anonymization network architecture.

a model that uses one network for binary gait anonymization and another one for binary anonymized gait colorization. The model uses the same discriminators as in the previous study to handle the quality of the gait generator.

As shown in Fig. 3, our model includes two independently trained networks. The first network, the anonymization network (A-NET), anonymizes the original gait while the second network, the colorization network (C-NET), following the first one colorizes the gait generated by the first network with the color of the original gait image. The anonymization network takes binary silhouette sequence X of the original gait (obtained by background subtraction) and uses random noise vector R to remove the identity of the original gait. It then outputs anonymized silhouette sequence Z.

To maintain the color of the original gait image as much as possible, we use a Siamese structure for the second network, which takes the output of the first network and the original RGB gait image Y as inputs. Because we need only the color of the gait image, the background in the original image should be removed. However, it is impossible to obtain the gait color from the raw video if the gait's silhouette cannot be extracted correctly. Therefore, we use the original gait image including the background. We use a pre-trained YOLO model (version 3) [24] to detect the object (gait) position in each frame. After extracting the position of the gait, we crop the gait along with the background. We add zeros padding around the cropped image so that we obtain a square image with the size equal to the height of the cropped image. We then resize this image to 64x64x3. To remove the background information from the original image, we multiply the image by the output of the anonymization network. The result of this operation and the output of the anonymization network are input to C-NET. The RGB anonymized gait synthesized by C-NET is then placed in the original position in the background using the same method presented by Tieu et al. [2].

B. Anonymization Network

The aim for the anonymization network is to produce an anonymized gait from an original one while maintaining the naturalness and movement direction of the original gait. The architecture of A-NET is shown in Fig. 4. Motivated by the potential of DCGAN [25], which uses random noise to generate an image in the same distribution as the training dataset, we use the original DCGAN model for A-NET. However, we modified it to use two inputs: the binary silhouette sequence of the original gait and a random noise vector, which is added

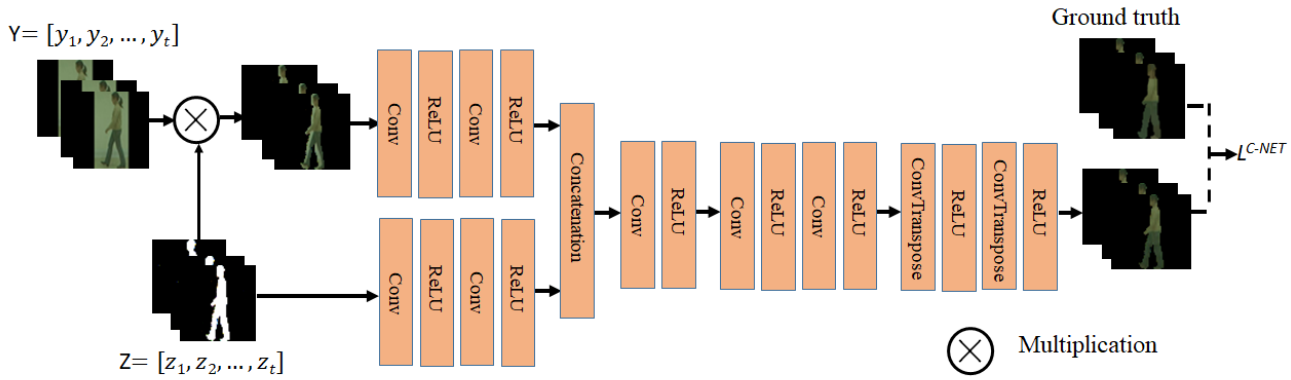
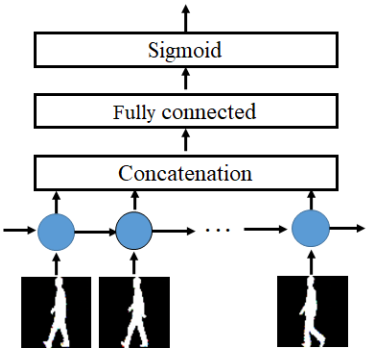


Fig. 5: Colorization network architecture.

to the original gait to remove the gait's identity. Inherited from ST-GAN model [2], we also stack a spatial discriminator and a temporal discriminator on the anonymization network. The aim with the spatial discriminator network is to distinguish the shape of a real gait and the shape of a synthesized gait. The input to this network is an image of size 64x64x3. Its architecture includes a two-dimension convolution network with a sigmoid function on top. The aim with the temporal discriminator network is to evaluate the smoothness of the anonymized gait. The architectures of this discriminator is similar to that proposed by Tieu et al. [2], but the input is a sequence of images of size 64x64x3. The architectures of the spatial and temporal discriminators are shown in Figs. 6a and 6b, respectively.

(a) Spatial discriminator network.



(b) Temporal discriminator network.

Fig. 6: Discriminator networks.

As mentioned above, our aim is to generate natural anonymized gaits even when the silhouettes of the original gait have low quality. Therefore, we used high-quality silhouette gaits for training A-NET. To train this network, we use two loss functions:

$$L_S^{A-NET} = E_{x\sim p_x(x), r\sim p_r(r)}[log(1 - D_S(f_A(X,R)))] \quad (1)$$

$$L_T^{A-NET} = E_{x\sim p_x(x), r\sim p_r(r)}[log(1 - D_T(f_A(X,R)))] \quad (2)$$

To maintain the viewing angle and action information ("walking") of the original gait, we also use reconstruction loss for training:

$$L_{Rec}^{A-NET} = E_{x\sim p_x(x), r\sim p_r(r)}[\| X - f_A(X,R) \|_1] \quad (3)$$

The objective function of A-NET is given by

$$L^{A-NET} = L_S^{A-NET} + L_T^{A-NET} + L_{Rec}^{A-NET} \quad (4)$$

where $f_A(.)$ is the output of the anonymization network.

C. Colorization Network

The aim with C-NET is to colorize the image of the gait synthesized by A-NET using the color of the original image. Because this network needs to learn the original color, it takes the anonymized gait generated by A-NET and the original gait image as inputs. Clearly, an RGB original gait image must be used, but extracting the gait color may result in the loss of color for any regions where the color of the gait is similar to the background color (it may not be possible to extract an exact silhouette for that region). Therefore, the gait color is taken with the background, as mentioned in Subsection *Model Overview*. Fig. 5 illustrates the architecture of this network.

We initially tried to remove the anonymized gait generated from the A-NET from the input of C-NET, but the generated output looked similar to the multiplied image. This is because C-NET had became an autoencoder network, which tried to generate output that looked like the input. Therefore, we use the Siamese architecture, which takes the A-NET output and

the multiplied image as inputs, for C-NET so that it does not generate an image similar to the multiplied image.

To train this network, we need to create a ground truth dataset. Because we want to transfer the original gait's color to the anonymized gait, we use the colorization method presented by Tieu et al. [2] to colorize the A-NET output. Several samples of the ground truth are shown in Fig. 7

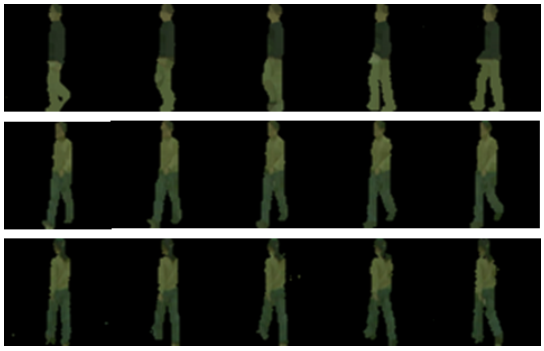


Fig. 7: Ground truth samples.

C-NET is trained by minimizing the loss function:

$$L^{C-NET} = E_{y\sim p_y(y),z\sim p_z(z)}[\| GT - f_C(Y,Z) \|_1] \quad (5)$$

where f_C is the C-NET output and GT is the ground truth.

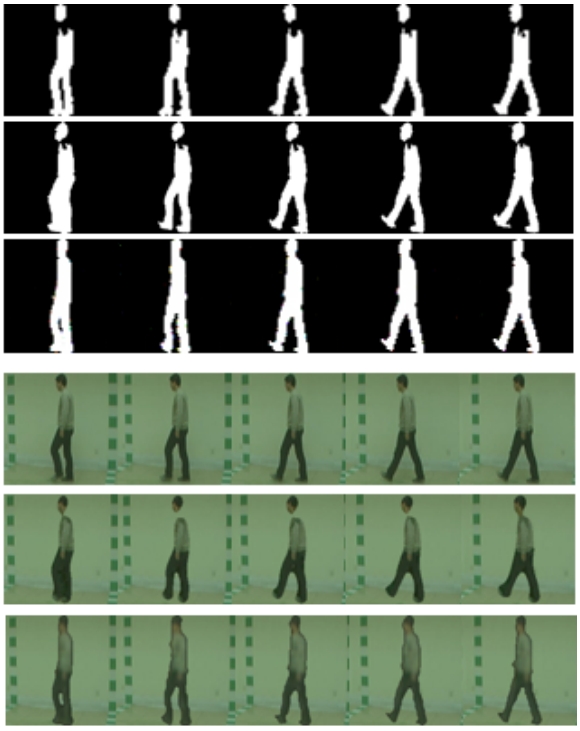
IV. Experimental Results

Our experiments were conducted on the CASIA-B gait dataset [5]. This dataset contains 110 gait sequences for each of 124 individuals recorded at 11 viewing angles ($0°$, $18°$, ..., $180°$). We divided the gait sequences of the 124 individuals into five non-overlapping groups. The first group, containing 5500 sequences (50 individuals), was used for training the gait recognition system. The second group, containing 1100 sequences (10 individuals) was used for training A-NET. The third group, containing 16 individuals (1760 sequences) was used for training C-NET. The fourth group, containing 8 individuals (880 sequences) was used for validation. The remaining 40 individuals (440 sequences) was used for testing.

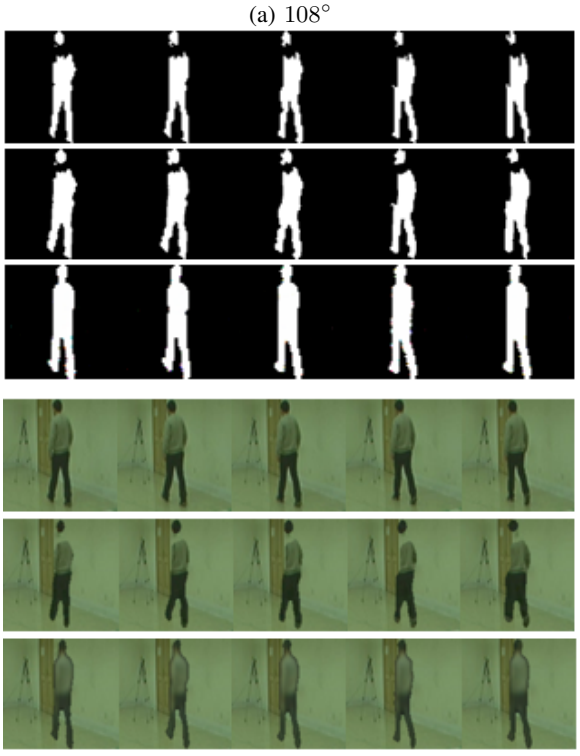
To evaluate the performance of the system, we use two metrics: success rate and naturalness. To evaluate the effectiveness of our proposed model for low-quality silhouette gaits, we used the ST-GAN model [2] as the baseline. For each model, we show the generation results and the success rate metrics. To evaluate the naturalness of gaits generated by our model for low-quality silhouette gaits, we used the MOS.

A. Generation Results

Several of the binary and RGB anonymized gaits generated with our model and the baseline method are shown in Figs. 8, 9, and 10. Our model used A-NET to generate binary anonymized gaits and C-NET to colorize the binary anonymized gaits. As shown by the results for both high-quality and low-quality silhouette gaits.

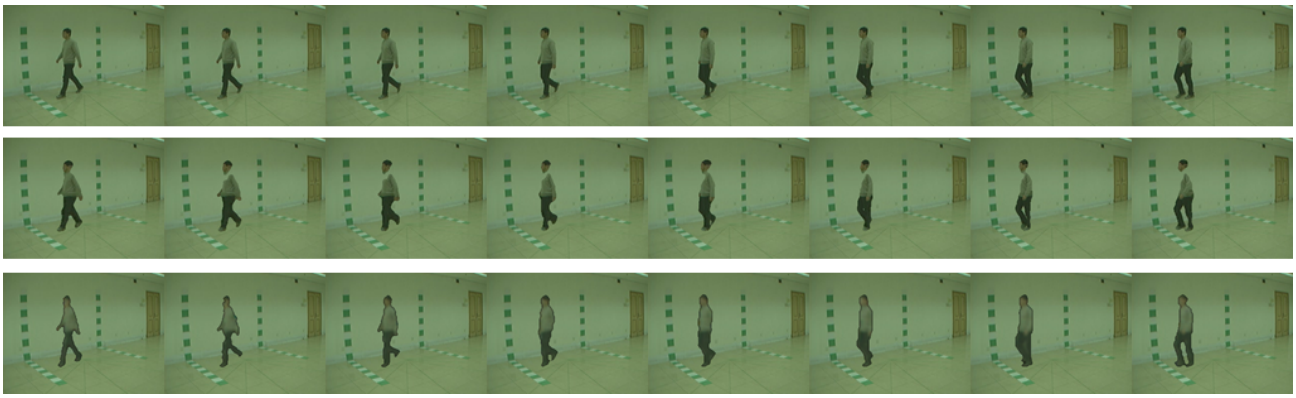


(a) $108°$

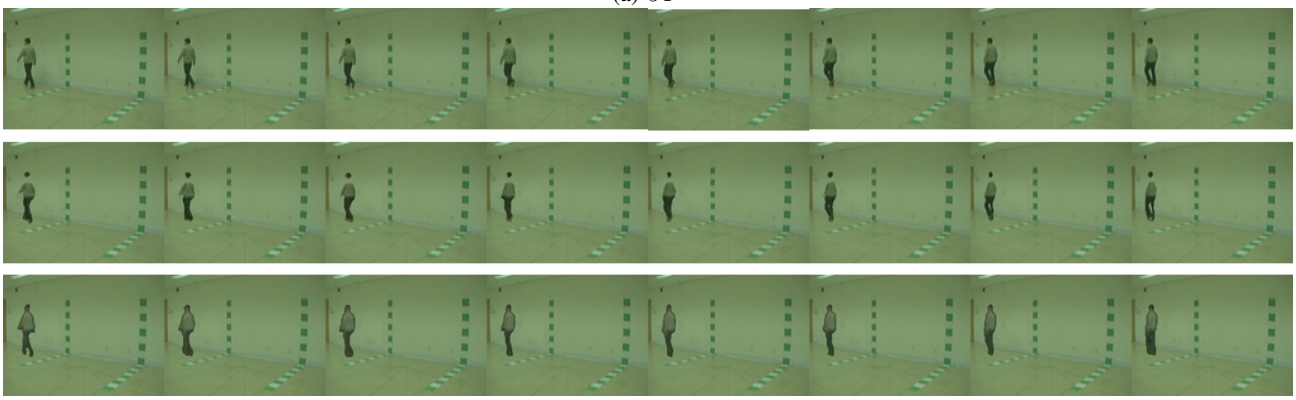


(b) $144°$

Fig. 8: Original and anonymized gaits generated from low-quality silhouette gaits with proposed and baseline (ST-GAN) methods for $108°$ and $144°$ view angles: top rows are original gaits, middle rows are results of baseline method, and bottom rows are results of proposed method.



(a) 54°



(b) 126°

Fig. 9: Original and anonymized gaits generated from low-quality silhouette gaits with proposed and baseline (ST-GAN) methods for 54° and 126° view angles: top rows are original gaits, middle rows are results of baseline method, and bottom rows are results of proposed method.

1) A-NET can generate a natural binary anonymized gait image even when the quality of the original silhouette is low, which is not always the case with the baseline method.

2) A-NET can anonymize not only the shape of the original gait, but, in some cases, can also anonymize the temporal information of the original gait (e.g., hand movement).

3) C-NET can colorize the binary anonymized gait image with the color of the original gait image for both high- and low-quality silhouettes.

4) The faces in the RGB anonymized gait images are blurry because the faces in the ground truth gait images were not clear and the size of each frame of the input gait image used in our model is small (64x64x3).

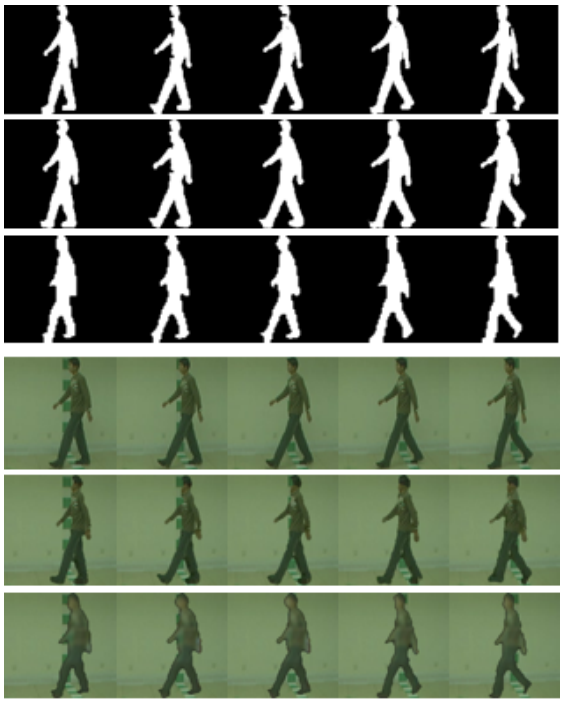
B. Naturalness

To measure the naturalness of anonymized gaits generated with our model, we used the MOS test. This test has been used for evaluating the quality of media generation results from the user's perspective [20], [26] and also for measuring the naturalness of gaits anonymized using state-of-the-art methods [1], [2]. We asked 20 volunteer evaluators to each evaluate 60 random anonymized gait videos. We asked them to rate gait naturalness on the basis of the gait's shape, movement, and color on a five-point scale (1: Bad, 2: Poor, 3: Fair, 4: Good, 5: Excellent). As shown in Fig. 11, the MOS for the naturalness of anonymized gaits based on low-quality silhouette gaits ranged from 2.70 to 3.37 depending on the view angle.

C. Success Rate

The success rate is the rate at which anonymized gaits are not correctly identified. The success rate metric is the same as that used for the baseline: the ratio of the number of anonymized gaits that were not correctly identified to the total number of anonymized gaits. We used the gait recognition system presented by Zheng with top-1 and top-3 identification. We separated our test set into two subsets, low-quality silhouettes (subset 1) and high-quality silhouettes (subset 2). The success rate for subset 1 ranged from 99.94% to 100%. This high rate is due to A-NET not only removing the identity of the original gait but also generating any body parts missing in the original silhouette.

For subset 2, the success rate (Fig. 12) was much higher than that of the baseline for the side views (from 36° to 144°) because temporal information (limbs moving) plays an important role for side views. Comparison of the gaits anonymized with the proposed model with those anonymized with the baseline model revealed that, in many cases, our model modified the hand movement while the baseline did



(a) 90°

Fig. 10: Original and anonymized gaits generated from high-quality silhouette gaits with proposed and baseline (ST-GAN) methods: top rows are original gaits, middle rows are results of baseline method, and bottom rows are results of proposed method.

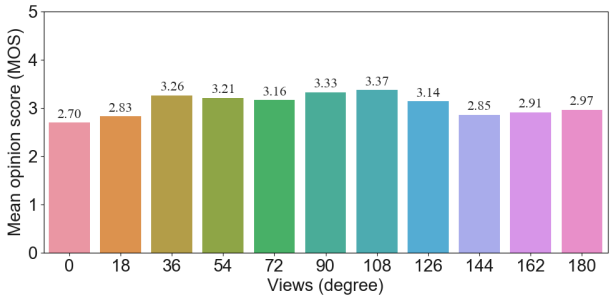


Fig. 11: Mean opinion scores for naturalness of anonymized gaits based on low-quality silhouette gaits.

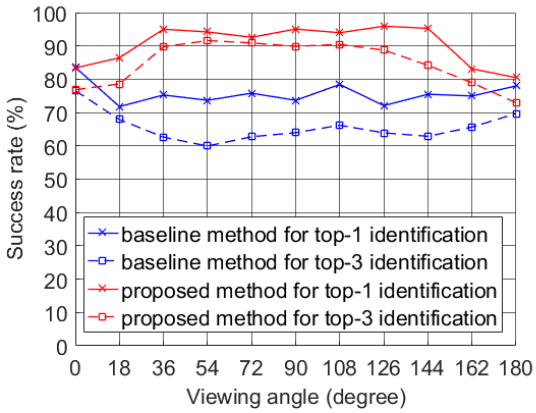
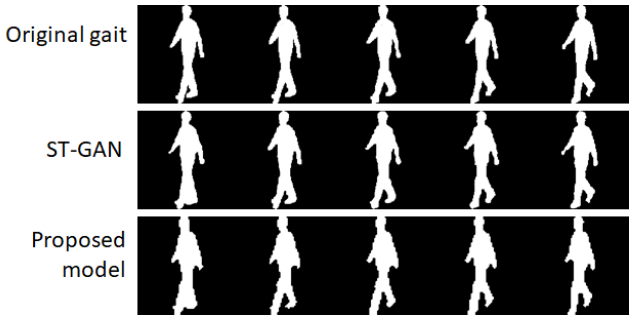
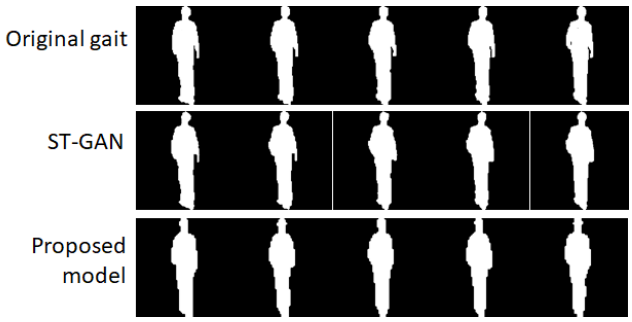


Fig. 12: Success rate comparison between proposed and baseline models.



(a) 54°



(b) 0°

Fig. 13: Silhouettes of original and anonymized gaits generated from low-quality silhouette gaits.

not. In other words, our model modified not only the shape of the gait but also the temporal information while the baseline method modified only the shape of the gait. Sample images are shown in Figs. 8 and 13. The success rate was only slightly higher than that of the baseline for the frontal views (0°, 180°) because the temporal information at these angles is less important, so modification of the shape by the baseline method was more effective.

V. Conclusion

Our proposed RGB gait anonymization model works well even for low-quality silhouette gaits, which is not the case for the state-of-the-art method. Our model uses two networks, one to anonymize the silhouette sequence of the original gait and the other to transfer the color of the original gait image to the anonymized silhouette sequence. Its success rate for low-quality silhouette gaits ranged from 99.94% to 100%, and its success rate for high-quality silhouette gaits was significantly higher than that of the state-of-the-art method for the side views. The mean opinion scores for anonymized gaits based on low-quality silhouette gaits ranged from 2.70 to 3.37. The ability of our model to anonymize gaits while maintaining naturalness makes it well suited for preprocessing videos to

anonymize gaits before they are uploaded to social media.

VI. Acknowledgments

This research was supported by JSPS KAKENHI Grants JP16H06302 and JP18H04120 and by JST CREST Grant JPMJCR18A6, Japan.

References

[1] N.-D. T. Tieu, H. H. Nguyen, H.-Q. N.-S. Nguyen-Son, J. Yamagishi, and I. Echizen, "An approach for gait anonymization using deep learning," in *2017 IEEE Workshop on Information Forensics and Security (WIFS)*, Dec 2017, pp. 1–6.

[2] N.-D. T. Tieu, H. H. Nguyen, H.-Q. Nguyen-Son, and I. E. Junichi Yamagishi, "Spatio-temporal generative adversarial network for gait anonymization," *Journal of Information Security and Applications*, vol. 46, pp. 307–319, June 2019.

[3] A. Senior, *Protecting Privacy in Video Surveillance*. Springer Publishing Company, Incorporated, 2009.

[4] D. Chen, Y. Chang, R. Yan, and J. Yang, "Tools for protecting the privacy of specific individuals in video," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, no. 1, pp. 1–9, Jan 2007.

[5] S. Yu, D. Tan, and T. Tan, "A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition," in *18th International Conference on Pattern Recognition (ICPR'06)*, vol. 4, 2006, pp. 441–444.

[6] S. Zheng, J. Zhang, K. Huang, R. He, and T. Tan, "Robust view transformation model for gait recognition," in *2011 18th IEEE International Conference on Image Processing (ICIP)*, Sept 2011, pp. 2073–2076.

[7] J. Han and B. Bhanu, "Individual recognition using gait energy image," *IEEE Trans. on Pattern Anal. and Mach. Intell.*, pp. 316–322, 2006.

[8] J. Wang, M. She, S. Nahavandi, and A. Kouzani, "A Review of Vision-Based Gait Recognition Methods for Human Identification," in *Digital Image Computing: Techniques and Applications (DICTA)*, 2010, pp. 320–327.

[9] Z. Wu, Y. Huang, L. Wang, X. Wang, and T. Tan, "A Comprehensive Study on Cross-View Gait Based Human Identification with Deep CNNs," *IEEE Trans. on Pattern Anal. and Mach. Intell.*, pp. 209–226, 2017.

[10] C. Wan, L. Wang, and V. V. Phoha, "A survey on gait recognition," *ACM Comput. Surv.*, vol. 51, no. 5, pp. 89:1–89:35, Aug. 2018.

[11] S. Pouyanfar, S. Sadiq, Y. Yan, H. Tian, Y. Tao, M. P. Reyes, M.-L. Shyu, S.-C. Chen, and S. S. Iyengar, "A survey on deep learning: Algorithms, techniques, and applications," *ACM Comput. Surv.*, vol. 51, no. 5, pp. 92:1–92:36, Sep. 2018.

[12] T. Nguyen and A. Takasu, "Npe: neural personalized embedding for collaborative filtering," in *Proceedings of the 27th International Joint Conference on Artificial Intelligence*. AAAI Press, 2018, pp. 1583–1589.

[13] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," *NIPS*, pp. 1–9, 2014.

[14] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 2414–2423.

[15] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European Conference on Computer Vision*, 2016, pp. 694–711.

[16] G. Lample, N. Zeghidour, N. Usunier, A. Bordes, L. DENOYER, and M. A. Ranzato, "Fader networks:manipulating images by sliding attributes," in *Advances in Neural Information Processing Systems 30*, 2017, pp. 5967–5976.

[17] J. Zhu, P. Krähenbühl, E. Shechtman, and A. A. Efros, "Generative visual manipulation on the natural image manifold," in *The European Conference on Computer Vision (ECCV)*, 2016, pp. 597–613.

[18] D. Holden, J. Saito, and T. Komura, "A deep learning framework for character motion synthesis and editing," *ACM Trans. Graph.*, pp. 138:1–138:11.

[19] J. Martinez, M. J. Black, and J. Romero, "On human motion prediction using recurrent neural networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017, pp. 4674–4683.

[20] C. Ledig, L. Theis, F. Huszr, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017, pp. 105–114.

[21] C. Yang, X. Lu, Z. Lin, E. Shechtman, O. Wang, and H. Li, "High-resolution image inpainting using multi-scale neural patch synthesis," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4076–4084.

[22] C. Vondrick, H. Pirsiavash, and A. Torralba, "Generating videos with scene dynamics," in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, 2016, pp. 613–621.

[23] Y. Yan, J. Xu, B. Ni, W. Zhang, and X. Yang, "Skeleton-aided articulated motion generation," in *Proceedings of the 2017 ACM on Multimedia Conference*, 2017, pp. 199–207.

[24] J. Redmon, S. K. Divvala, R. B. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788.

[25] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015. [Online]. Available: http://arxiv.org/abs/1511.06434

[26] A. van den Oord, Y. Li, I. Babuschkin, K. Simonyan, O. Vinyals, K. Kavukcuoglu, G. van den Driessche, E. Lockhart, L. C. Cobo, F. Stimberg, N. Casagrande, D. Grewe, S. Noury, S. Dieleman, E. Elsen, N. Kalchbrenner, H. Zen, A. Graves, H. King, T. Walters, D. Belov, and D. Hassabis, "Parallel wavenet: Fast high-fidelity speech synthesis," in *International Conference on Machine Learning (ICML)*, 2018, pp. 3915–3923.