Color Transfer to Anonymized Gait Images While Maintaining Anonymization

Ngoc-Dung T. Tieu*, Junichi Yamagishi*†, and Isao Echizen*†‡
* The Graduate University for Advanced Studies, SOKENDAI, Kanagawa, Japan
† National Institute of Informatics, Tokyo, Japan
‡ The Universtity of Tokyo, Tokyo, Japan
E-mail: {dungtieu,jyamagis,iechizen}@nii.ac.jp

***Abstract*—Gait anonymization helps prevent the identification of people by gait recognition systems using videos uploaded to social media. Our current gait anonymization approach is to first modify the silhouette of the gait sequence and then transfer the colors of the skin, hair, clothing, etc. in the original gait images to the modified gait images to produce a final RGB anonymized gait image sequence. Since users typically care about the quality of the generated videos as they want to share them with family and friends in addition to caring about privacy, the generated videos should contain color images that are sharp and finely textured. Existing anonymization models are unable to produce such images. In this paper, we focus on color transfer while maintaining anonymization. This is challenging because the original gait images may consist of multiple colors, the modified gait differs from the original one, there is no real ground truth anonymized gait, and it may be difficult to exactly separate the foreground colors from the background colors in the original gait images. To overcome this problem, we propose transferring the colors without using ground truth and without extracting the colors in the original gait images. In this model, the overlapping region between the two gaits is first located, and the colors in that region in the original images are transferred to the modified images. The colors in the remaining region are interpolated from the color in the overlapping region, so the colors in the overlapping and non-overlapping regions are coherent. Quantitative and qualitative experiments demonstrated that the proposed model is more effective than our previous models with no reduction in anonymization.**

***Index Terms*—Gait; biometric trait; security; gait anonymization; deep learning**

I. Introduction

A person's image in a video sequence can reveal various kinds of privacy-sensitive information [1], and rapid advances in gait recognition have made gait important biometric information. This has increased the risk that videos uploaded and shared on social media will be maliciously exploited to generate fake videos or obtain personal information. Anonymizing gait while maintaining naturalness is one approach to preventing the identification of people by gait recognition systems using videos uploaded to social media. It comprises two main tasks, as shown in Fig. 1: anonymization and color transfer/colorization. Though the anonymization prevents gait recognition systems from identifying the target person, visible artifacts remaining after color transfer/colorization are an important concern, especially if the anonymized video is to be shared.

Our group previously proposed two models for anonymizing RGB gaits. We define a complete silhouette as a seamless silhouette and an incomplete silhouette as a silhouette with one or more parts of the body missing. Incomplete silhouettes are caused by the foreground extraction process and occur when the color of the body part is similar to the background color or the gait is partly occluded by another object. This is because a ton of foreground extraction algorithms rely on the difference between the background and foreground colors [2], [3], [4], [5], [6]. Our first model [7] works well on complete silhouettes but is unable to generate a seamless gait from incomplete silhouettes. Our second model [8] overcomes this drawback. However, the colorization algorithms in both models cannot generate images that are sharp and finely textured, especially when the silhouettes of the original gait are incomplete or a color is unseen in the training data. This is because the colors in the original gait image must be extracted, which may be difficult for incomplete silhouettes, or because artificial ground truths are used to train the model.

We have now focused on improving the second task (color transfer/colorization), as indicated by the red dashed rectangular box in Fig. 1. In this paper, we present a model that can generate sharp and finely textured RGB anonymized gait images from the RGB original gait images and binary anonymized gait images without using ground truths and without extracting the colors of the original gait from the background. Our model includes one encoder followed by one decoder. The model takes the original gait image and the binary anonymized gait image at each frame of the two gait sequences as inputs. The RGB original gait images are captured along with the background from the raw video, and the binary anonymized gait images are generated by the anonymization network (Task 1 in Fig. 1) introduced in our previous report [8]. Because we aim at color transfer while preserving the shape of the anonymized gait, the output of the decoder is pixel-wise multiplied by the binary anonymized gait image to force the network to reform the shape of output to that of the binary anonymized gait. To reduce the visible artifacts in the final results, we use a loss function that matches the output with the original gait instead of an artificial ground truth. Since the shapes of the output and original gait are not the same, the loss function is aimed at preserving the colors in the overlapping region between the two gaits and interpolating

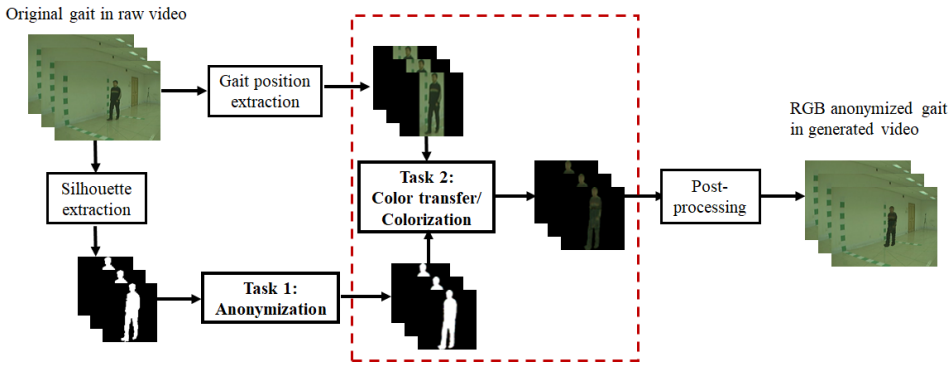


Fig. 1: Flow chart of our previous gait anonymization models. This study focused on Task 2: color transfer/colorization.

the colors of the remaining region so that coherent colors and textures are retained between the two regions.

We evaluated the proposed model on the CASIA-B gait dataset [9] both qualitatively and quantitatively. For quantitative measurement, we used three metrics: structural similarity (SSIM), peak signal-to-noise ratio (PSNR), and success rate. SSIM and PSNR are widely used to measure the quality of generated images while the success rate is used to measure anonymization success. The quantitative and qualitative experiments both demonstrated that our model is more effective than our previous model at transferring the colors from the original gait images to the binary anonymized gait images while preserving anonymization success.

After discussing related work in Section II, we describe the proposed model in Section III and present the results of our experimental evaluation in Section IV. We summarize the key points in Section V.

II. Related Work

A. Gait Recognition

The human gait, i.e., the pattern of walking, has become an important biometric trait that can be used to identify people at a distance [10], [11], [12]. Previously proposed gait recognition methods can be divided into two main approaches: model-free and model-based methods. The model-based methods extract the gait feature on the basis of dynamic or static parameters of body parts while the model-free approaches use silhouette information. The model-free approach is more robust to low-resolution videos and has lower computational cost. We therefore used the model-free approach to evaluate the anonymization performance of our existing models. Specifically, we used a widely used gait identification method introduced by Zheng et al. [13]. They used a gait energy image (GEI) computed from the average silhouette of one gait cycle [10]. Fig. 2 shows an example of silhouettes of one gait cycle and its GEI. They applied a view transformation to the GEI to transform the viewing angle of the probe gait to that of the gallery gait. After applying the view transformation to these gaits, they computed the similarity between them and used it as the distance between them.

Fig. 2: Images on the left represent silhouettes of one gait cycle while that on the right is the gait feature.

B. Gait Anonymization

Our group has conducted several studies on anonymizing gait while retaining naturalness in the generated videos [14], [7], [8]. Two of them were reported on RGB gait datasets. In the first study, we introduced an anonymization network and a color transfer algorithm. The anonymization network anonymizes the binary gait on the basis of the gait's contour, and the color transfer algorithm transfers the colors in the original gait images to the binary anonymized gait images by using the nearest neighbor color. Although the naturalness of the anonymized gaits generated from complete silhouettes were preserved, anonymized gaits generated from incomplete silhouettes looked unnatural.

To overcome this drawback, we introduced another gait anonymization model [8] based on the deep convolutional generative adversarial network (DCGAN) [15] architecture and trained on complete silhouettes to generate the unbroken anonymized gait regardless of the silhouette quality of the original gait. The colorization network used to transfer the colors of the original gait images to the binary anonymized gait images was trained by minimizing the loss function between the model output and the ground truth. Ground truths were created using our first model. However, the colorization network was unable to generate sharp and finely textured colors due to artifacts in the ground truth.

C. Human Garment Generation

Deep neural networks have achieved success in a wide range of areas [16], [17], especially in transferring the garments in a reference image to a target image. Lassner al.[18] introduced a model for generating images of people wearing arbitrary clothing given a body pose image, therefore, their model does not guarantee coherence among the frames of a gait sequence. Motivated by the growing popularity of online shopping, Han

et al. [19] developed a method for overlaying the clothing in a product image on a person in a query image while Neuberger et al. [20] created a model for synthesizing a new image composed of various selected items of clothing to help a customer compare outfits and choose an attractive one.

There have also been more relevant studies that focused on transferring the garments from a person in a given image to the same person in a generated image. Zhao et al.'s model [21] and Ma et al.'s model [22] are aimed at generating an image of a person from a desired viewpoint or in a desired pose from a static image containing that person while keeping the clothing the same. To transfer the clothing from the original image to the generated pose, they combined the U-net convolutional network for biomedical image segmentation [23] with a discriminator that distinguishes between the original and ground truth image pair and the original image and generated image pair. However, in the anonymization problem, there is no real ground truth. Raj et al. [24] presented an algorithm for transferring the clothing of a person in a reference image to another person in the target image while preserving the shape and pose of the target person. After a segmentation process is applied to the reference and target images, the color of the clothing is disentangled from the body pose for each body part. The clothing segmentation of the reference person and body part segmentation of the target person are then combined to generate the desired image. Unfortunately, this algorithm cannot be applied to our problem since we cannot segment the binary anonymized images into body parts.

III. Model

In this section, we present our new model for transferring colors from an RGB original gait image to a binary anonymized gait image as part of the gait anonymization approach. Given an RGB image x_{rgb} containing the original gait and an image x_{bi} of the binary anonymized gait, our model generates an image containing a gait that has the same shape as the gait in x_{bi} and is overlaid by the colors (including those of the skin, hair, clothing, etc.) on the gait in x_{rgb}. Due to the color similarity between the original gait image and the generated gait image, after applying our model to all frames in the gait sequence, we expect to obtain frame-by-frame coherent colors in the final video.

In our previous gait anonymization models, the performance of the colorization task depends on the the color of the original gait extraction process [7] or the quality of the artificial ground truth [8], which results in artifacts in the final video. The newly proposed model overcome this limitation by capturing the RGB original gait without separating it from the background, comparing the color of the output with that of the captured original gait image, and using it to optimize model performance. We present our model architecture in subsection III-A and explain the loss function used to generate the color and texture details of the original gait image in subsection III-B. After obtaining RGB anonymized gait images, we apply the post-processing introduced in previous reports [8], [7] to produce the final video containing the RGB anonymized gait with the original background.

A. Network Architecture

Since our model is aimed at overlaying the colors of the original gait image on the binary anonymized gait image, the model should ideally take the original gait image without the background and the binary anonymized gait as inputs. However, in some cases, the foreground cannot be exactly extracted from the background, e.g., when the silhouette is incomplete. This problem is overcome by taking the original gait image with the background and the binary anonymized gait image at each frame of the two gait sequences as inputs. The network architecture of the proposed model is shown in Fig. 3

A pre-trained YOLO model [25] is used to detect the position of the original gait in the raw video, and then the gait is cropped along with the background. Next, zero padding is added around the cropped image to create a square image with sides equal to the height of the cropped image, and all images are then resized to the same size. A binary anonymized gait sequence is produced using our previous RGB gait anonymization model [8], which is based on the DCGAN [15] architecture and can generate a seamless anonymized gait regardless of the quality of the original silhouette. First, our model concatenates the two inputs and then compresses the information therein by using a convolutional encoder followed by a fully connected hidden layer. Next, a convolutional decoder decodes the encoded feature map of the hidden layer. Our aim is to transfer color while preserving the gait pattern of the anonymized gait. In other words, the shape of the final gait should match that of the binary anonymized gait. To this end, the output of the decoder is pixel-wise multiplied by the binary anonymized gait to obtain the final output.

Different from the colorization network in our previous model [8], the convolutional encoder in our new model is placed before the fully connected layer in order to compress more of the input information into the hidden layer and thereby make the layer more informative for the decoder. Moreover, the pixel-wise multiplication has been moved from the beginning of the network to the end. This was done because the ground truth is not used in the loss function, so applying the pixel-wise multiplication to the decoder output forces the network to reform the shape of the output to that of the binary anonymized gait.

B. Loss Function

We define the center region as the overlapping region between the two input gaits and define the edge region as the region belonging to the binary anonymized gait but not belonging to the center region. The center region is located by applying a morphological operation to the binary input image x_{bi}, and the edge region is located by subtracting the center region from the binary input image, as illustrated in Fig. 4. Because the binary anonymized gait is obtained by modifying the shape of the original gait while keeping the same phase,

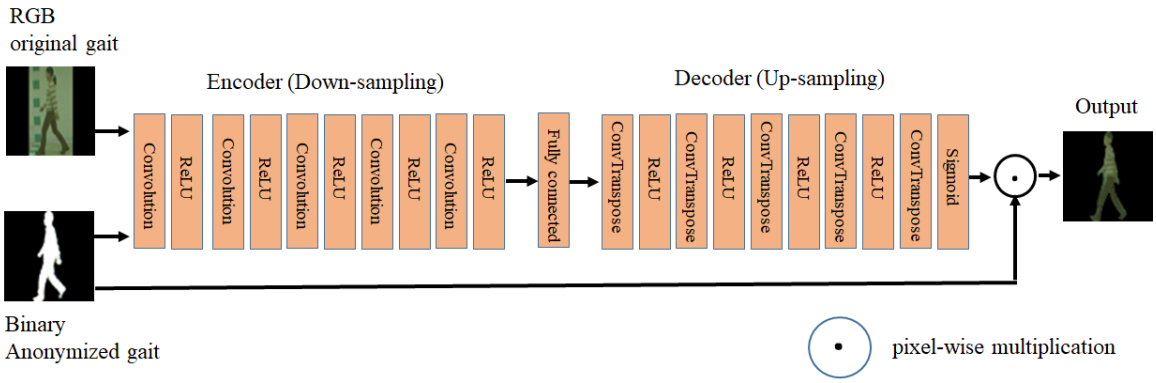


Fig. 3: Network architecture: The encoder compresses the input information into the hidden layer, and the decoder decodes the feature map of the hidden layer. Pixel-wise multiplication reforms the shape of the decoder output to that of the binary anonymized gait.

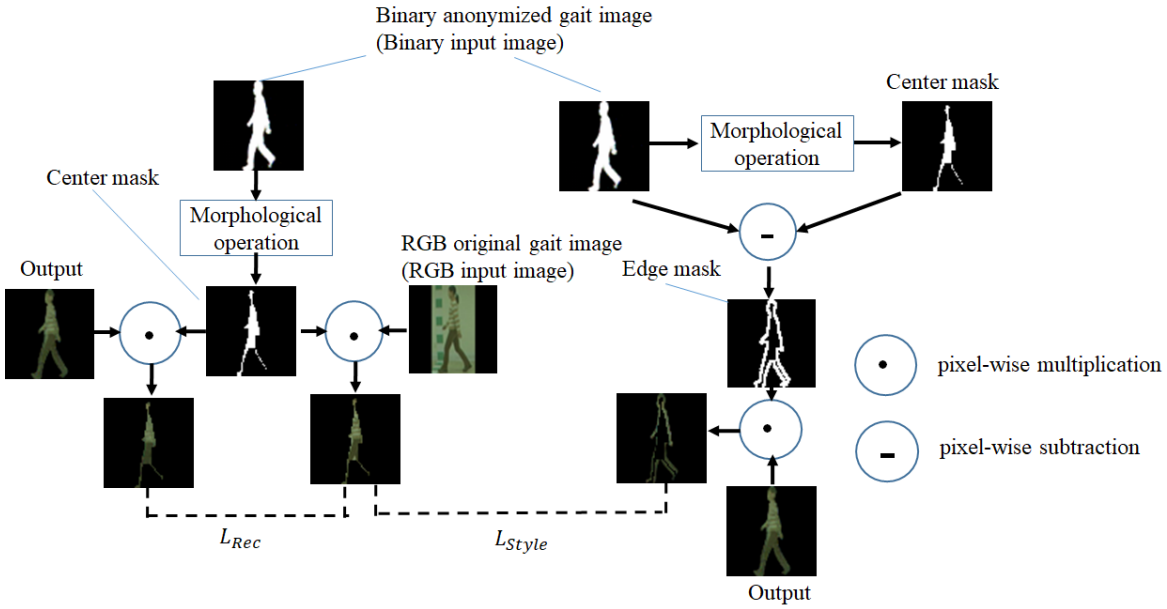


Fig. 4: Reconstruction loss L_{Rec} matches the center region of the RGB original gait image to that of the output while style loss L_{Style} matches the center region of the RGB original gait image to the edge region of the output.

our network tries to reconstruct the color of the center region and then interpolate the color of the edge region from that of the center region. Our model is trained by minimizing an objective function that includes two terms, reconstruction loss and style loss, in order to construct the color of each region. The loss function is described in detail in the rest of this subsection.

1) Reconstruction Loss: Reconstruction loss is used to transfer the color in the center region of the RGB original gait image to that of the binary anonymized gait image. As shown in Fig. 4, a center mask is first created by applying a morphological operation to the binary input image. The center region of the RGB input image and that of the model output are then computed by pixel-wise multiplication between the center mask and each image, respectively. The l_1 loss used to match the two regions is formulated as follows.

$$L_{Rec} = ||Mp(x_{bi}) \odot x_{rgb} - Mp(x_{bi}) \odot \Phi(x_{rgb}, x_{bi})||_1, \quad (1)$$

where Φ is the color transfer network, $Mp(\cdot)$ is the morphological operation, x_{rgb} is the RGB original gait image, x_{bi} is the binary anonymized gait image, and $\odot$ is pixel-wise multiplication.

2) Style Loss: Our task now is to generate the color in the edge region so that it is coherent with the color in the center region. In other words, we need to design a loss function so that the network can capture the color style of the center region

and transfer that color to the edge region. Inspired by the success of Gram matrix loss introduced by Gatys et al. [26] in generating beautiful stylized and textured images and its use in several studies [27], [28], we used this loss to generate the color in the edge region. We denote F_i as the vectorized (flattened) feature map of the i-th channel of input image x. The Gram matrix of x is defined as the inner product between such feature maps.

$$Gr_{ij}(x) = \langle F_i, F_j \rangle = \sum_k F_{ik} F_{jk} \tag{2}$$

where k is the element of each channel.

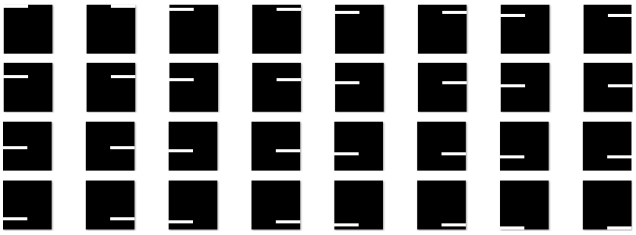


Fig. 5: Masks with the same size as the two input images were used to compute the style loss.

Because the center and edge regions may include all body parts and the color may differ among body parts, we divide these regions into patches. We create 32 masks with the same size as the input images, as shown in Fig. 5, to extract patches. The center and edge regions are pixel-wise multiplied using each mask one by one in order to find the pair of nearest patches, one in the center region and one in the edge region. The color of each patch in the edge region is generated on the basis the color of the nearest patch in the center region. This is enabled by training the model to match the Gram matrix values of these two patches. Since the number of pixels in these two patches differs, the Gram matrix is normalized by dividing it by the number of pixels in each patch. The style loss function is computed by summing the Gram matrix matching of all pairs:

$$L_{Style} = \sum_l || \frac{1}{M_l} Gr(Mp(x_{bi}) \odot x_{rgb} \odot m_l) - \frac{1}{N_l} Gr((x_{bi} - Mp(x_{bi})) \odot \Phi(x_{bi}, x_{bi}) \odot m_l) ||_1, \tag{3}$$

where M_l and N_l are the numbers of pixels in each patch (center and edge, respectively), and l is the index of the mask, l-th m_l.

The number of pixels in each patch is computed using

$$M_l = \sum_p (Mp(x_{bi}) \odot m_l) \tag{4}$$

$$N_l = \sum_p ((x_{bi} - Mp(x_{bi})) \odot m_l) \tag{5}$$

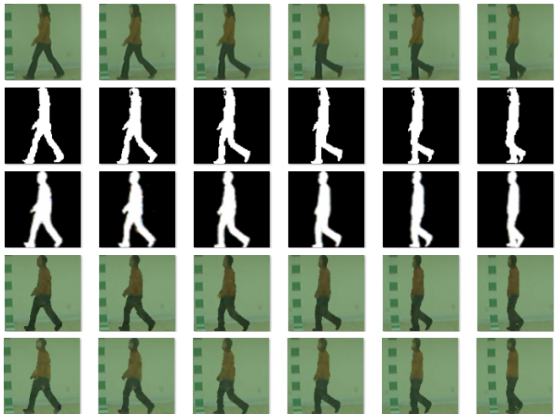
where p is the element of each patch.

IV. Evaluation

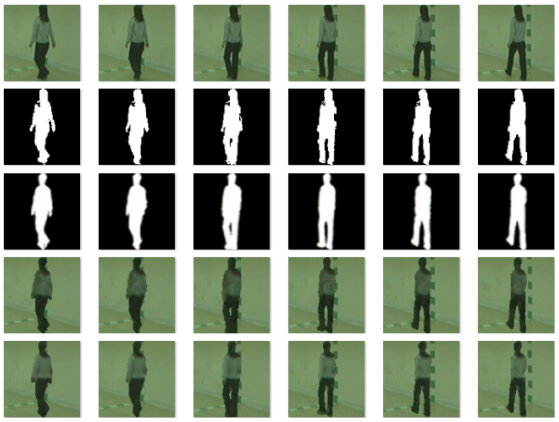
We evaluated our proposed model experimentally using the CASIA-B gait dataset [9]. This dataset contains 110 gait sequences for each of 124 individuals recorded at 11 viewing angles ($0°$, $18°$, $\ldots$, $180°$). We evaluated the model both qualitatively and quantitatively in comparison with the baseline introduced in our previous study [8]. We used three metrics for the quantitative evaluation: SSIM, PSNR, and success rate of anonymization. We divided the dataset into three non-overlapping groups. The first group (30 individuals; 3300 sequences) was used to train the baseline. The second group, (50 individuals; 5500 sequences) was used to train the gait identification system proposed by Zheng et al. [13], which is briefly summarized in subsection II-A. The last group (44 individuals; 4400 sequences) was used as a test set to evaluate the proposed model compared with the baseline. To investigate the effect of the proposed model on different kinds of data, we defined four subsets in the test set: an *incomplete silhouette set* containing incomplete silhouette gaits, a *complete silhouette set* containing complete silhouette gaits, a *plain color set* containing images in which clothing with one color overlays the gait, and a *textural color set* containing images in which clothing with multiple colors overlays the gait.

A. Qualitative Evaluation

Figs. 6 to 8 show the images generated by the proposed and baseline models for a variety of viewing angles. A visual comparison between the results for the two models on the plain color set is shown in Fig. 6. It demonstrates that while both models can transfer a plain color, the color generated by the proposed model looks more similar to the original color. A visual comparison of the results for the textural color set is shown in Fig. 7. It shows that the baseline model produced blurry and coarse images while the proposed model captured the finely textured color in the RGB original gait images and transferred it to the generated images. The baseline model was trained using a loss function that matched the network output with the ground truth, which contained artifacts. In contrast, the proposed model was trained using a loss function that matched the output with the actual gait. In other words, the proposed model compared the generated color and the real color, while the baseline model attempted to minimize the distance between the synthesized color and the artificial color. A visual comparison of the results for the complete silhouette set are shown in Figs. 6, 7a, and 7c. The images generated by the proposed model are better than those generated by the baseline, especially for the textural color set (Figs. 7a and 7c). A visual comparison for the incomplete silhouette set is shown in Fig. 8. For the baseline model, we trained the network on complete silhouettes and used this pre-trained model to colorize the binary anonymized gait images because it is not easy to create the ground truth when the silhouette is incomplete. This led to artifacts in the baseline-generated images when the color to be transferred was not included in the training data. In contrast, the proposed model used the reconstruction and style loss functions to match the output image colors with the



(a) 90°



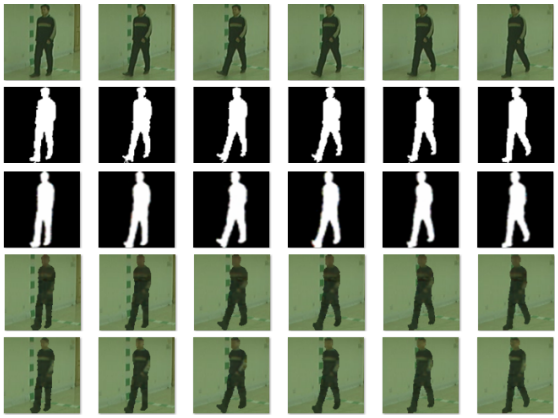
(b) 144°

Fig. 6: Images generated by proposed and baseline models on plain color set. Top rows are RGB original gait images, second rows are binary original gait images, third rows are binary anonymized gait images, and fourth and fifth rows are color transfer results by baseline and proposed models, respectively.

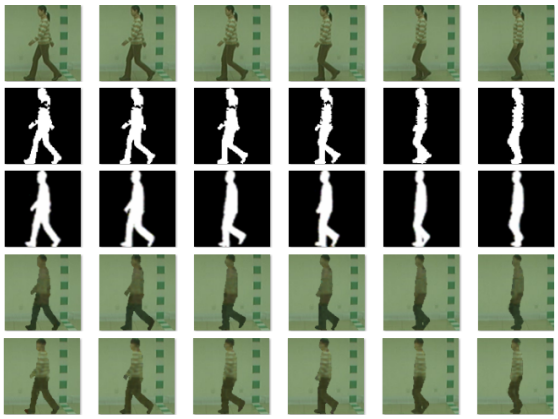
original image colors. Therefore, the proposed model can be trained on both incomplete and complete silhouette gaits. This leads the proposed model can generate images with either a plain color or a finely textured color regardless of the quality of the original gait silhouettes.

B. Quantitative Evaluation

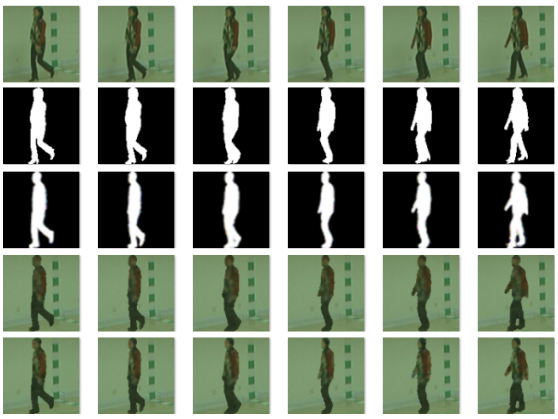
Two of the metrics we used, PSNR and SSIM, are commonly used to measure the similarity between two images. We used them to compare the quality of the images generated by the proposed model with that of the ones generated by the baseline model. PSNR is computed on the basis of the pixel-level mean square error, which measures the difference between corresponding pixels. SSIM [29] is used to assess the structural similarity between images by independent comparisons of three image characteristics: luminance, contrast, and structure. Both metrics are commonly used quality measurements and widely used in image reconstruction [24], [30]. We



(a) 36°

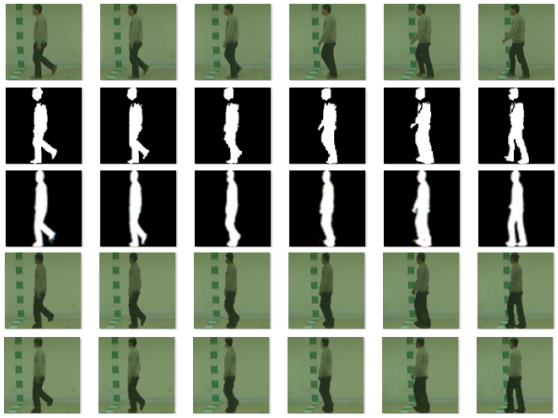


(b) 72°



(c) 72°

Fig. 7: Images generated by proposed and baseline models on textural color set. Top rows are RGB original gait images, second rows are binary original gait images, third rows are binary anonymized gait images, and fourth and fifth rows are color transfer results for baseline and proposed models, respectively.



(a) 90°

(b) 126°

Fig. 8: Color transfer results for proposed and baseline models on incomplete silhouette set. Top rows are RGB original gait images, second rows are binary original gait images, third rows are binary anonymized gait images, and fourth and fifth rows are color transfer results for baseline and proposed models, respectively.

computed these metrics for each pair of images, the generated image and the original image. Table I shows the results for the plain and textural color sets. It shows that the proposed model was more effective than the baseline one in color transfer for both sets. This is consistent with our qualitative evaluation.

As shown in the qualitative evaluation, the proposed model can generate the natural color even if the silhouettes of the original gait are incomplete, whereas the baseline model generates results that look like artifacts. This is consistent with the comparison in Table II, which presents the results of the quantitative evaluation for the complete and incomplete silhouette sets. These results demonstrate that the proposed model is robust against the quality of the silhouettes of the original gait.

The third metric was the success rate of anonymization, which is the ratio of the number of anonymized gaits that were not correctly identified by the gait recognition system and the total number of anonymized gaits. The success rates for the two model were completely the same for every viewing angle and ranged from 79.04% to 93.61% depending on the viewing angle. That is, the success rate remained the same after the color of the anonymized gaits was changed.

TABLE I: PSNR and SSIM of baseline and proposed models for plain and textural color sets.

Method	PSNR		SSIM	
	Plain color	Textural color	Plain color	Textural color
Baseline	24.2417	23.6273	0.9062	0.8970
Proposed	24.7302	24.2822	0.9142	0.9065

TABLE II: PSNR and SSIM of the baseline and proposed models for complete and incomplete silhouette sets.

Method	PSNR		SSIM	
	Incomplete silhouette	Complete silhouette	Incomplete silhouette	Complete silhouette
Baseline	23.9494	24.1035	0.9047	0.9042
Proposed	24.4664	24.8346	0.9126	0.9094

V. Conclusion

We have introduced a model for color transfer in gait anonymization that overcomes the limitations of two previously reported models: an inability to generate natural output when the original gait silhouettes are incomplete and an inability to produce a sharp color, especially a finely textured color. The proposed model was trained using the loss function that includes two terms: reconstruction loss and style loss. The first term is aimed at constructing the color of the center region while the second term is aimed at generating the color of the edge region so that the colors between two regions are coherent. We conducted extensive qualitative and quantitative evaluations on four sets of data: plain color, textural color, incomplete silhouettes, and complete silhouettes. Both evaluations demonstrated that the proposed model is more effective and more robust against silhouette quality for color transfer while preserving the success rate of anonymization.

VI. Acknowledgments

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

References

[1] 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, 2018.

[2] O. Barnich and M. Van Droogenbroeck, "Vibe: A universal background subtraction algorithm for video sequences," *IEEE Transactions on Image Processing*, vol. 20, no. 6, pp. 1709–1724, 2011.

[3] M. Braham and M. Van Droogenbroeck, "Deep background subtraction with scene-specific convolutional neural networks," in *2016 International Conference on Systems, Signals and Image Processing (IWSSIP)*, 2016, pp. 1–4.

[4] X. Liang, S. Liao, X. Wang, W. Liu, Y. Chen, and S. Z. Li, "Deep background subtraction with guided learning," in *2018 IEEE International Conference on Multimedia and Expo (ICME)*, 2018, pp. 1–6.

[5] Y. Yan, H. Zhao, F.-J. Kao, V. Vargas, S. Zhao, and J. Ren, "Deep background subtraction of thermal and visible imagery for pedestrian detection in videos," in *International Conference on Brain Inspired Cognitive Systems (BICS2018)*, 2018, pp. 75–84.

[6] Zhiyu Wang, Hui Xu, Lifeng Sun, and Shiqiang Yang, "Background subtraction in dynamic scenes with adaptive spatial fusing," in *2009 IEEE International Workshop on Multimedia Signal Processing*, 2009, pp. 1–6.

[7] 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.

[8] N. T. Tieu, H. H. Nguyen, F. Fang, J. Yamagishi, and I. Echizen, "An rgb gait anonymization model for low-quality silhouettes," in *2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2019, pp. 1686–1693.

[9] 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.

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

[11] 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.

[12] 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.

[13] 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.

[14] N.-D. T. Tieu, H. H. Nguyen, H.-Q. 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.

[15] 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

[16] 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.

[17] 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.

[18] C. Lassner, G. Pons-Moll, and P. V. Gehler, "A generative model of people in clothing," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 853–862.

[19] X. Han, Z. Wu, Z. Wu, R. Yu, and L. S. Davis, "Viton: An image-based virtual try-on network," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018, pp. 7543–7552.

[20] A. Neuberger, E. Borenstein, B. Hilleli, E. Oks, and S. Alpert, "Image based virtual try-on network from unpaired data," in *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020, pp. 5184–5193.

[21] B. Zhao, X. Wu, Z.-Q. Cheng, H. Liu, and J. Feng, "Multi-view image generation from a single-view," *Proceedings of the 26th ACM international conference on Multimedia*, pp. 383–391, 2018.

[22] L. Ma, X. Jia, Q. Sun, B. Schiele, T. Tuytelaars, and L. Van Gool, "Pose guided person image generation," in *Advances in Neural Information Processing Systems 30*. Curran Associates, Inc., 2017, pp. 406–416.

[23] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Springer International Publishing, 2015, pp. 234–241.

[24] A. Raj, P. Sangkloy, H. Chang, J. Hays, D. Ceylan, and J. Lu, "Swapnet: Image based garment transfer," *European Conference on Computer Vision, ECCV*, pp. 1–17, 2018.

[25] 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.

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

[27] Z. Zhang, Z. Wang, Z. Lin, and H. Qi, "Image super-resolution by neural texture transfer," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019, pp. 7982–7991.

[28] U. Dmitry, L. Vadim, V. Andrea, and L. Victor, "Texture networks: feed-forward synthesis of textures and stylized images," in *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016*, June 2016, pp. 1349–1357.

[29] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[30] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Generative image inpainting with contextual attention," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 5505–5514.