Gait Anonymization with Preservation of Appearance and Motion

by

TIEU Thi Ngoc Dung

Dissertation

submitted to the Department of Informatics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

The Graduate University for Advanced Studies, SOKENDAI
September 2021

Advisor

Prof. Isao ECHIZEN
National Institute of Informatics (NII),
The Graduate University for Advanced Studies, SOKENDAI,
The University of Tokyo

Sub-advisors

Prof. Junichi YAMAGISHI
National Institute of Informatics (NII)
The Graduate University for Advanced Studies, SOKENDAI

Advisory Committee

Prof. Shinichi SATOH
National Institute of Informatics (NII)
The University of Tokyo
The Graduate University for Advanced Studies, SOKENDAI
Prof. Yinquiang ZHENG
The University of Tokyo
Prof. Koichi ITO
Tohoku University
Prof. Yu YI
National Institute of Informatics (NII)
The Graduate University for Advanced Studies, SOKENDAI

Acknowledgements

First of all, I would like to deeply express my warm and sincere thanks to my supervisor, Professor Isao Echizen. I fully admire his knowledge, research skills, and the way he encouraged me to overcome obstacles. I also would like to thank him for his financial support during my study at SOKENDAI and NII that helped me so much for my life in Japan. Therefore, I could concentrate on my research.

I would like to express my appreciation to my sub-advisor, Professor Junichi Yamagishi. Discussing with him was truly helpful and inspiring for the new potential of technology.

I also would like to send my thanks to the other committee members: Professor Yinqiang Zheng, Professor Shin'ichi Satoh, Professor Koichi Ito, and Professor Yu Yi for their invaluable comments and suggestions.

I would like to express my gratitude to Ms. Yumiko Seino, Ms. Miki Nihei, Ms. Mai Saito, NII staff, and SOKENDAI staff for their support. They have enthusiastically assisted me with official procedures during my study at SOKENDAI and NII.

I would like to thank postdocs and students of the Echizen lab. I felt comfortable and helpful when discussing with them about research as well as daily life. Talking with them also helped me to balance research and life.

I owe my love thanks to my family for their constant love and support.

Publication List

Journal papers

1. **N.-D. T. Tieu**, H. H. Nguyen, H.-Q. Nguyen-Son, J. Yamagishi, and I. Echizen, "Spatio-Temporal Generative Adversarial Network for Gait Anonymization," Journal of Information Security and Applications, vol. 46, pp. 307–319, June 2019.

2. Noboru BABAGUCHI, Isao ECHIZEN, Junichi YAMAGISHI, Naoko NITTA, Yuta NAKASHIMA, Kazuaki NAKAMURA, Kazuhiro KONO, Fuming FANG, Seiko MYOJIN, Zhenzhong KUANG, Huy H. NGUYEN, **N.-D. T. TIEU**, Preventing Fake Information Generation Against Media Clone Attacks, IEICE Transactions on Information and Systems, 2021

3. Isao ECHIZEN, Noboru BABAGUCHI, Junichi YAMAGISHI, Naoko NITTA, Yuta NAKASHIMA, Kazuaki NAKAMURA, Kazuhiro KONO, Fuming FANG, Seiko MYOJIN, Zhenzhong KUANG, Huy H. NGUYEN, **N.-D. T. TIEU**, "Generation and Detection of Media Clones", IEICE Transactions on Information and Systems, 2021, Volume E104.D, Issue 1, Pages 12-23, 2021

Conference papers

1. **N.-D. T. Tieu**, J. Yamagishi, and I. Echizen, "Color Transfer to Anonymized Gait Images while Maintaining Anonymization," Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2020.

2. **N.-D. T. Tieu**, H. H. Nguyen, F. Fang, J. Yamagishi, and I. Echizen, "An RGB Gait Anonymization Model for Low-Quality Silhouettes," Asia-Pacific Signal and

Information Processing Association Annual Summit and Conference (APSIPA ASC), pp.1686-1693, Nov. 2019.

3. H. H. Nguyen., **N.-D. T. Tieu**, H.-Q. Nguyen-Son, J. Yamagishi, and I. Echizen. "Transformation on Computer-Generated Facial Image to Avoid Detection by Spoofing Detector". International Conference on Multimedia and Expo (ICME), pp. 1-6, June 2018.

4. H. H. Nguyen, **N.-D. T. Tieu**, H.Q. Nguyen-Son, V. Nozick, J. Yamagishi, and I. Echizen, "Modular Convolutional Neural Network for Discriminating between Computer-Generated Images and Photographic Images," International Conference on Availability, Reliability and Security (ARES), ACM, 2018.

5. H.-Q. Nguyen-Son, **N.-D. T. Tieu**, H. H. Nguyen, J. Yamagishi, and I. Echizen, "Identifying Computer-Translated Paragraphs using Coherence Features". Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation (PACLIC 32), 2018

6. H.-Q. Nguyen-Son, **N.-D. T. Tieu**, H. H. Nguyen, J. Yamagishi, and I. Echizen, "Identifying Computer Generated Text Using Statistical Analysis", Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), pp.1504-1511, Dec.2017

7. **N.-D. T. Tieu**, H. H. Nguyen, H.-Q. Nguyen-Son, J. Yamagishi, and I. Echizen, "An Approach for Gait Anonymization Using Deep Learning," IEEE Workshop on Information Forensics and Security (WIFS), pp. 1–6, Dec. 2017. [Best paper award]

Abstract

Video of walking subjects can be easily captured in many ways such as by surveillance cameras in the public areas, by personal devices (e.g. smartphones). Such data consists of sensitive information of subjects in videos like routines, activities, behaviors. Meanwhile, there are many reasons for sharing such data among organizations or companies such as for law enforcement, forensics, research. However, an attacker may crosslink a subject in a video to a subject in another video by using gait recognition systems, and therefore the sensitive information of subjects in videos may be disclosed. This thesis investigates anonymizing the gait (walking pattern) of a subject in a video so that the gait recognition systems incorrectly recognize the anonymized gait while preserving the appearance and motion of the subject for data analysis. The preservation of appearance means ensuring the naturalness of the body shape and preserving the garment color of the original gait images. The preservation of motion consists of guaranteeing the naturalness of movement and retaining the moving direction of the original gait. In this dissertation, several techniques based on deep learning are proposed to solve three problems that step by step adapt to the wild (real application): (1) Gait anonymization on binary gait video, (2) Gait anonymization on RGB gait video, (3) Incomplete silhouette gait anonymization. The evaluation was conducted on an available gait dataset and with several advanced gait recognition methods. The experimental results demonstrate that the proposed models achieve the promised performance in privacy protection as well as can generate quality gait images.

Most current gait recognition approaches exploit silhouette sequences of subjects, which are sequences of binary images containing the shape and style working of each subject, as the gait's feature. This thesis thus, initially, investigates binary gait anonymization. This research is aimed at investigating whether we can find a way to fool gait recognition systems by directly modifying the shape of silhouettes of the original gait. It also explores how to apply deep learning to ensure the naturalness of the body shape and the gait movement. A fast and easy-to-implement method is introduced. This model anonymizes a gait by modifying the

original body shape of that gait with the use of another gait, named “noise gait”, to add to that gait, and therefore, the anonymized gait is partly similar to the original gait and partly similar to the noise gait. To do this, a contour vector containing coordinates of pixels on the contour of a gait at each frame. A convolutional neural network(CNN) based model is designed that takes the contour vector of the original gait and that of the noise gait as two outputs and produces a modified contour vector. Then this modified contour vector is used to generate the anonymized gait.

Although most of the current gait recognition research is based on the silhouette sequence of subjects, most uploaded/shared videos are RGB videos. The development of an approach for RGB gait anonymization is essential and therefore this thesis, secondly, considers how to anonymize RGB gait. This research investigates how to design a deep learning model to modify the shape of RGB gait by using random noise to remove the original identity information of a gait in a video while preserving the appearance and motion of the original subject. To solve this problem, a deep learning-based generative model is proposed that consists of one generator and two discriminators, that is special discriminator and temporal discriminator. First, a random noise in the gait distribution is created by traditional GAN from a random vector in a Gaussian distribution. Then, the random noise is added to a gait that we wish to anonymize. Two discriminators are to improve the quality of generation results. The purpose of the spatial discriminator network is to discriminate a real gait image with a generated gait image. It does this by distinguishing the body shape of a real gait and that of a synthesized one at each frame. The purpose of the temporal discriminator network is to ensure the moving smoothness of the anonymized gait.

This thesis also considers how to generate seamless anonymized gait when incomplete silhouette gaits are used because gait recognition systems are still able to recognize such gaits with high accuracy. In this thesis, a complete silhouette is defined as a seamless silhouette and an incomplete silhouette is defined 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 and that of the background are the same, or the gait is partly occluded by another object. The incomplete silhouettes usually occur in the real application. The two previous models, one for binary and one for RGB gait, are unable to generate the seamless silhouette gait when the incomplete silhouette gaits are used. The gait anonymization research on incomplete silhouettes aims to change the shape of an incomplete silhouette while producing a seamless silhouette. This research is also aimed at generating

the fine texture color regardless of the quality of the original silhouettes. To this end, the proposed model consists of two networks, anonymization network, and colorization network. The anonymization network is based on the DCGAN model and trained by the complete silhouette dataset to generate seamless silhouettes. The colorization network transfers the garment colors of original gait images to anonymized gait images without extracting the garment colors of the original gait images to avoid missing the original colors.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Motivation and Overview

1.1.1 Privacy Concerns

Due to advances in computing technology, media content can be recorded and shared in many ways. Video data can be captured with surveillance cameras on the street by governments and private companies. To give some examples, there are over 4800 government surveillance cameras in Washington, D.C. [15] and more than 4 million closed-circuit television (CCTV) cameras in the United Kingdom. A resident of London is caught on CCTV cameras about 300 times a day on average [16]. In addition, the development of smart devices has enabled videos to be recorded by private individuals using such devices as smartphones and cameras. The data recorded often contains sensitive information such as the behaviors, routines, activities, and affiliations of the people caught on camera.

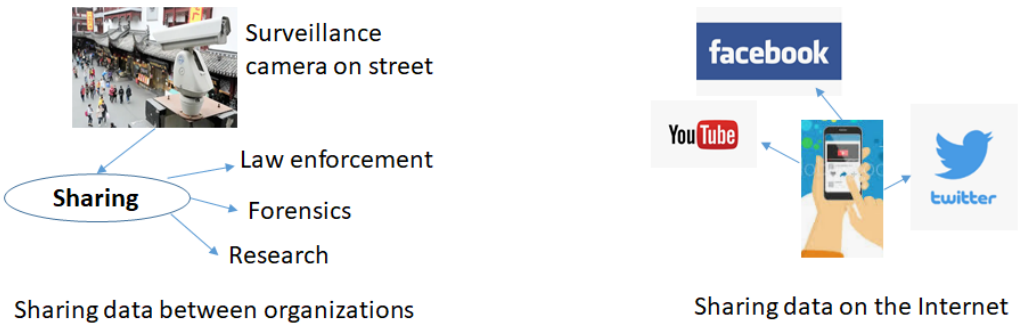


Fig. 1.1 Videos of people walking can be captured and shared in many ways.

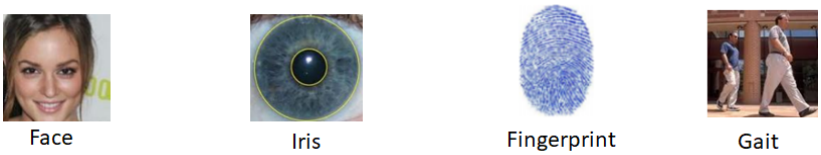


Fig. 1.2 Example biometrics that can be used to recognize a person in a video.

Although there are many reasons for sharing such data among organizations or companies such as for law enforcement, forensics, and research, a n attacker may use gait recognition systems to crosslink a subject in a video to a subject in another video, which could result in disclosure of sensitive information about that person. Therefore, protecting the privacy of people when sharing data is becoming more essential every year. The Europe General Data Protection Regulation (GDPR) in 2018, requires clear and explicit consent from the data subject if their personal data is to be processed beyond the legitimate purpose for which that data was collected. However, if the data is anonymized, such data is not subject to the GDPR, meaning that companies can use anonymized data without consent. Studies of hiding individual identities when sharing data while preserving the information needed for data analysis have been conducted in both academia and industry [11, 16]. These studies have focused on *anonymization* or *de-identification*.

1.1.2 Gait Anonymization

Biometrics is the science of recognizing an individual on the basis of inherent physical and/or biological traits associated with a person [17]. Fingerprint, iris, face, ear, hand geometry, palm print, finger vein geometry, gait, voice, and signature are biometrics widely used in many applications [18–20]. People in videos can be recognized using several biometrics such as face, iris, fingerprint, and gait, as illustrated in Fig. 1.2 .

Among those traits, gait has become important because it can be detected at low resolution [21] and recognized at a distance without physical contact or the person's cooperation [22, 23] while most other biometrics could be identified only at a close distance or with physical contact. While face anonymization has been intensively studied (existing research on face anonymization will be reviewed in Chapter 2), there has not been much gait anonymization research. Gait refers to the manner in which a person walks. It is a complex dynamic activity with movements made using a combination of hundreds of muscles and body joints. These movements are integrated in the sense that they must happen within a specific time. Normal

Fig. 1.3 Sample gaits at different viewing angles.

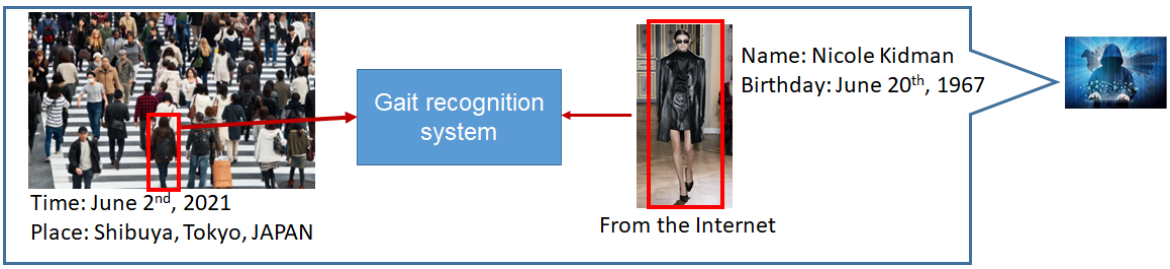


Fig. 1.4 Example of privacy violation resulting from use of gait recognition system.

human walking has been discovered to be cyclical movement; i.e., the motion repeats at a stable frequency [23–25]. It has been defined as the forward thrust of the legs, in which the alternating movements of different parts of the body consume less energy [26]. Fig. 1.3 shows sample gaits at different viewing angles.

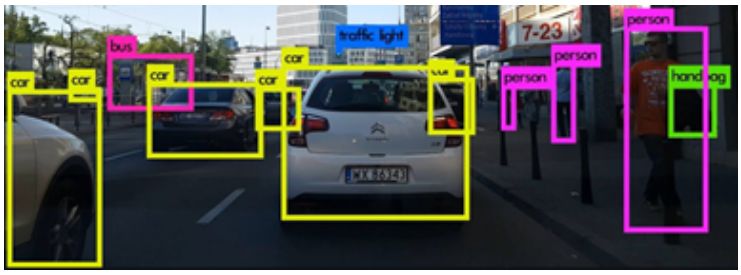
Moreover, with the rapid development of gait recognition research, a state-of-the-art gait recognition method can achieve very high performance. Therefore, a person in a video may be unintentionally recognized with high accuracy by a gait recognizer. Figure 1.4 shows an example of privacy violation resulting from the linking a gait in one video to a gait in another video. In this example, the video data recorded by a surveillance camera was shared to a third party for data analysis. The attacker is assumed to be authorized to access these data and be able to acquire a video of a specific person from the Internet. By using a gait recognition system, the attacker can find the specific person in the shared video, resulting in the disclosure of sensitive information about this person such as behaviors and routines.

In addition to privacy protection, data utility is an important aspect of gait anonymization because it is necessary in many research areas [27], for instance, object detection [28, 29],

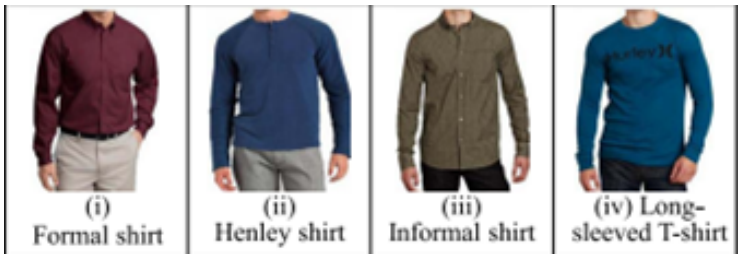
action prediction [30], gender recognition, and clothes recognition [31, 32]. Example research areas requiring data utility preservation in gait anonymization are illustrated in Fig. 1.5.

From the discussion above, we can see that protecting the privacy of a person in a shared video is necessary. Research in this area has been ongoing for several decades and has resulted in such techniques as object blurring [4], object removing [5, 7], and object replacement [6]. However, these techniques focus more on privacy protection than on data utility preservation.

(a) Action prediction [30].



(b) Object detection [28].



(c) Clothing recognition [31].

Fig. 1.5 Preservation of data utility for data analysis.

1.2 Research Objectives

The research described in this thesis was aimed at anonymizing the gait of a person in a video so that it is incorrectly recognized by gait recognition systems while preserving the original appearance and motion of the person. Preservation of appearance means ensuring the naturalness of the body shape and preserving the c colors. Preservation of motion means guaranteeing the naturalness of movement and retaining the direction of movement. Three

problems that step-by-step adapt to the wild have been solved in this thesis. In particular, this research focused on binary gait anonymization, RGB gait anonymization, and incomplete silhouette gait anonymization.

- **Binary gait anonymization:** Most gait recognition approaches use the silhouette sequence of a person, which is a sequence of binary images containing the shape and walking pattern of the person, as the gait's feature. Research on binary gait anonymization has aimed at finding a way to fool gait recognition systems by directly modifying the silhouette shapes of the gait. It has also explored the use of deep learning to ensure the naturalness of the body shape and to retain the viewing angle of the original gait movement. Studies on binary gait anonymization have addressed three research questions in particular:

  1. Can gait recognition systems be fooled by modifying the shapes of the silhouettes in a gait sequence?

  2. How can naturalness of the body shapes in the generated gait be ensured?

  3. How can the moving direction be retained after eliminating the identifying information?

- **RGB gait anonymization:** Silhouette sequences have been used as the gait feature in most gait recognition research, but most videos shared on the Internet or displayed on CCTV are RGB videos. Development of an approach for RGB gait anonymization is thus necessary. Although the binary gait anonymization model achieved promising results for both anonymization and video quality, it is limited to black and white videos and has not solved well the consistency among frames of the generated gait. In addition, the performance of the binary gait anonymization model depends on how the noise gait used to remove the identity of the person in the original video is chosen. Studies on RGB gait anonymization have explored ways to overcome these problems of the binary gait anonymization model. Studies on RGB gait anonymization have addressed two research questions in particular:

  1. How can the shapes of silhouettes in a gait sequence be modified without using another gait?

  2. Does using a generative adversarial network (GAN) improve the quality, especially consistency among frames, of anonymized gait videos?

3. How can the color from original RGB gait images be transferred to binary anonymized gait images?

- **Incomplete silhouette gait anonymization:** A complete silhouette is a seamless silhouette, and an incomplete silhouette is one with one or more parts of the body missing. Incomplete silhouettes usually occur when the color of the foreground is similar to that of the background or when a person is occluded. Gait recognition systems are still able to recognize such gaits with the high accuracy because incomplete silhouettes may occur at some frames of a gait sequence and some parts of human body can be used to recognize the subject. To anonymize incomplete silhouette gaits, a foreground extraction process is used to obtain the gait of the person in the video. This process may result in some of the person's body parts being missing. The models used for binary gait anonymization and RGB gait anonymization are unable to generate a seamless gait when the silhouettes are incomplete. In the research described here, the use of a random noise to remove the identity of the original gait was investigated as a means to improve the anonymization success rate. Studies on incomplete silhouette gait anonymization have addressed three research questions in particular:

  1. How can the shapes of silhouettes in a gait sequence be modified by using a random noise?

  2. How can a seamless silhouette be produced from an incomplete silhouette?.

  3. How can the clothing colors in original gait images be transferred to anonymized gait images when the colors cannot be extracted for the missing parts of the incomplete silhouette?

1.3 Challenges

Although many algorithms have been developed for gait anonymization [4–7], they cannot achieve the two requirements for gait anonymization: privacy protection and data utility preservation. Gait anonymization is challenging for several reasons.

Unlike most other biometrics such as face, fingerprint, iris (for which the biometric feature consists of only spatial information), the biometric feature of gait consists of spatial and temporal information. The identity of a gait must be synchronously removed from all frames in the gait sequence, which may cause inconsistencies among the frames in the generated gait.

Fig. 1.6 Common challenges in gait anonymization.

Although a person walking may be viewed at different angles, the anonymization algorithm should not be specific to each view. Moreover, the person's clothes may be colorful, and retaining tiny textured colors is not easy. Finally, the performance of gait anonymization depends on the results of object segmentation. In many cases, the segmentation process cannot extract a seamless object, especially when the foreground color is similar to the background color or when the object is occluded.

1.4 Original Contributions

The research described in this thesis makes the following original contributions

- **This is the first research on gait anonymization aimed at preservation of both appearance and motion**. Previous studies on gait anonymization focused more on privacy protection such as by obscuring the body of a person [33] and by pixelating or blurring the whole body [4, 5, 34]. Such methods do not address the problem of preserving the original appearance and motion. As far as we know, the work introduced in this thesis is the first study on gait anonymization aimed at preservation of both appearance and motion.

- **Binary gait anonymization:** Most gait recognition approaches take the silhouette sequence as the gait's feature. This study initially investigated binary gait anonymization.

The main idea is to use a "noise gait", that is, another gait, to modify the body shape of the gait. This is done by using a contour vector containing the coordinates of the pixels on the contour of the body shape as gait the feature. A convolutional network (CNN) with two inputs is designed. One input is the contour coordinates of the original gait, and the other is the contour coordinates of the noise gait. The output is a modified contour vector, which is used to produce a corresponding anonymized gait.

- **RGB gait anonymization:** Although most gait recognition research has focused on the silhouette sequence, most videos uploaded and/or shared are RGB videos. Development of an approach for RGB gait anonymization is essential, and the research reported in this thesis addressed the anonymization of RGB gaits. A model composed of one generator and two discriminators is proposed. To alter the shape of the body of a gait, a random noise created by a traditional GAN from a random vector that is sampled in a Gaussian distribution is added to the gait. A spatial discriminator network and a temporal discriminator network are used to improve the quality of the generated results. The purpose of the spatial discriminator network is to discriminate a real and fake gait image at each frame. In other words, this network distinguishes the shape of the real gait from that of a fake gait. The purpose of the temporal discriminator network is to ensure that the motion of a generated gait is smooth by distinguishing the temporal information of a real gait from that of a fake gait.

- **Incomplete silhouette gait anonymization:** Incomplete silhouettes are caused by the foreground extraction process and occur when the color of a body part is similar to the background color or the gait is partially occluded, which commonly occurs in real applications. Incomplete silhouette gaits may be recognized accurately since several parts of the human body can be used to recognize the person in a video. The models for binary gait anonymization and RGB gait anonymization are unable to generate an anonymized seamless silhouette gait when incomplete silhouette gaits are used. The research described in this thesis was aimed at developing a model for gait anonymization that is robust to the silhouette quality of the original gait. To this end, a model consisting of two networks, an anonymization network and a colorization network, was developed. The anonymization network removes the identity of the original gait by the addition of a random noise sampled from Gaussian distribution. It is based on the deep convolutional GAN (DCGAN model) and is trained on a complete silhouette dataset to generate seamless silhouettes. The colorization network transfers

the clothing colors of the original images to the anonymized images without extracting the clothing colors to avoid missing the original colors.

1.5 Relationships among Three Studies

The relationships among three studies (binary gait anonymization, RGB gait anonymization, and incomplete silhouette gait anonymization) are presented in two domains: *adaptation to the wild* and *network architecture*.

1.5.1 Adaptation to the Wild

The relationships in the adaptation to the wild domain are expressed through the achievements of each study, as shown in Fig. 1.7.

The study on binary gait anonymization, for which the input is binary gait images, investigated ways to fool gait recognition systems by modifying the shapes of the silhouettes in a gait sequence. It demonstrated anonymization ability while partly ensuring the naturalness of the anonymized gait images and consistency among the frames of the anonymized gait. The findings provide a base for further research on RGB gait and incomplete silhouette gait data anonymization.

The study on RGB gait anonymization explored anonymization of RGB videos, which are usually used in real application. It demonstrated anonymization ability while maintaining clothing colors, consistency among frames, and moving direction. It partly ensured naturalness of the anonymized gait images (e.g., it was unable to generate seamless gait images when incomplete silhouette gaits were used).

The study on incomplete silhouette gait anonymization addressed the problem of incomplete silhouette gaits, which may occur in real applications and which has not been completely solved by binary gait anonymization and RGB gait anonymization. The results demonstrated that a gait can be anonymized while preserving the original attributes even if incomplete silhouettes are used.

1.5.2 Network Architecture

The relationships in the network architecture domain are expressed through the use of the same base model. Gait generator G of binary gait anonymization is based on an autoencoder

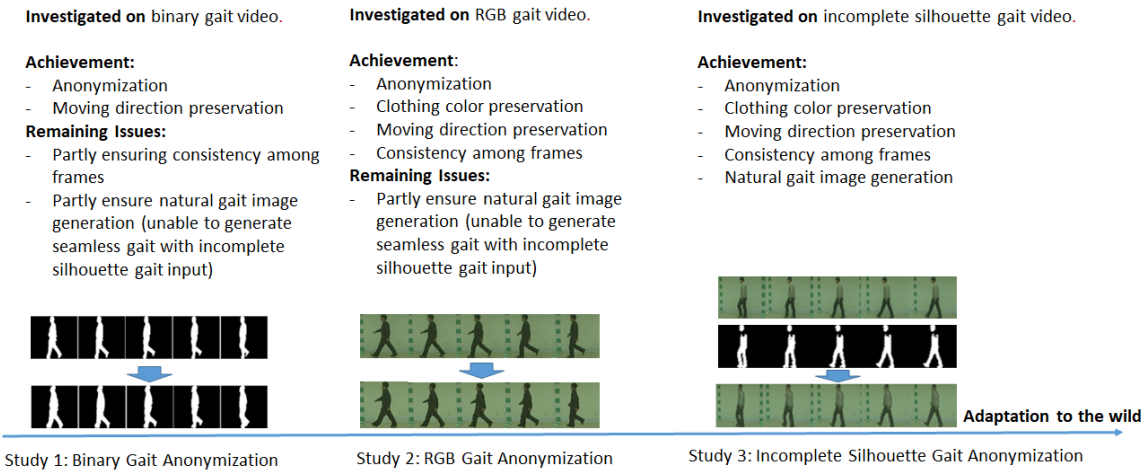


Fig. 1.7 Relationship in adaptation to the wild domain.

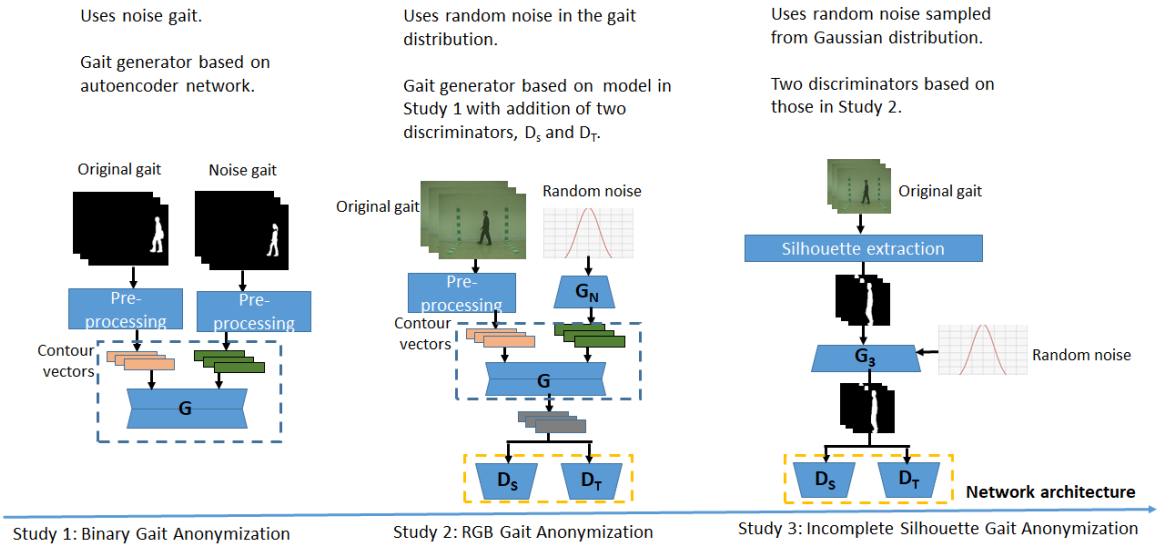


Fig. 1.8 Relationships among three studies in network architecture domain.

network, and the gait generator of RGB gait anonymization is based on generator G. The models of RGB gait anonymization and incomplete silhouette gait anonymization used a spatial discriminator and a temporal discriminator to achieve a natural body shape and natural movement. The relationships among the three studies in the network architecture domain are summarized in Fig. 1.8.

1.6 Organization

The remaining chapters of this thesis are organized as follows.

Chapter 2 provides a detailed literature review of existing image and video anonymization technologies and then discusses the issues related to their application to gait anonymization. This chapter also briefly describes three gait recognition methods that were used to evaluate the anonymization performance of the proposed gait anonymization models. Finally, the gait dataset used in the experiments and the metrics used for evaluation are explained.

Chapter 3 describes the model and implementation of the binary gait anonymization method used in the research described in this thesis. The required pre-processing of raw data before passing the model and the post-processing used to obtain the final videos are explained in detail. The effect of choosing the noise gait on the generated results is also described.

Chapter 4 presents an algorithm using a spatio-temporal generative adversarial network (ST-GAN). The benefit of using a random noise in the gait distribution rather than using the noise gait to remove the identity of the real gait is explained. Detailed descriptions of the proposed generator and two discriminators for improving the quality of the generated results are provided. An analysis is given of the trade-off between anonymization and naturalness, which depends on how much noise is added to the original gait.

Chapter 5 shows that anonymizing an incomplete silhouette gait is necessary because gait recognition systems are able to accurately identify incomplete silhouette gaits. This chapter also explores the effect of incomplete silhouette gaits on the quality of the generated gaits and presents the model proposed for modifying the shape of an incomplete silhouette while being able to produce a seamless silhouette. A colorization model that is able to transfer the colors of missing parts of the original gait to the anonymized gait is described.

Chapter 6 summarizes the key points of the research and provides the research direction for future work.

Chapter 2

Literature Review

This chapter provides a detailed literature review of existing image and video anonymization technologies then gives discussion and demonstration of issues of such the technologies when applied to the gait anonymization. This chapter also briefly presents three gait recognition methods that are used to evaluate the anonymization performance of the proposed gait anonymization models. Finally, the available gait dataset used in experiment and metrics for evaluation are also explained.

2.1 Terminology

In this section, the terminology used in this thesis is explained to ensure better understanding of the contents.

- **Subject** The term 'subject(s)' refers to the person ID(s) used for training and testing.
- **Viewing angle** The terms viewing angle refers to the angle of the subject relative to the camera, as illustrated in Fig. 1.3.
- **Complete Silhouette** Foreground extraction is used to obtain the gait of the person in the video. This process may result in some body parts of the extracted gait being missing. A complete silhouette is defined as a seamless silhouette, as shown in Fig. 2.1.
- **Incomplete Silhouette**

  An incomplete silhouette is defined as a silhouette with one or more parts of the body missing. An incomplete silhouette may occur if the color of a body part is similar

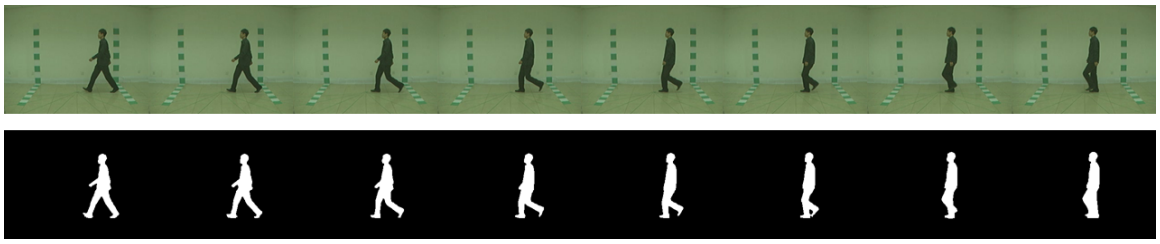


Fig. 2.1 Top row contains RGB gait images; bottom row contains complete silhouette images.

to the background color and if the gait is partially occluded. Samples of incomplete silhouettes are shown in Fig. 2.2.

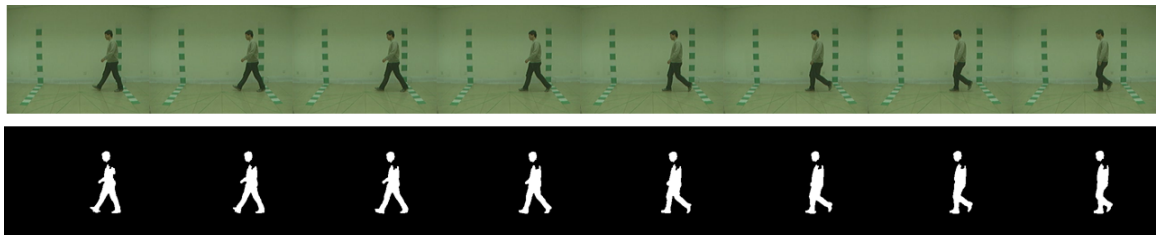


Fig. 2.2 Top row contains RGB gait images; bottom row contains incomplete silhouette images.

2.2 Gait Recognition

The purpose of gait recognition systems is to recognize subjects in videos by using their walking pattern. Given a gallery that contains a set of gait samples, such systems predict the identity of a query gait (probe). In the current literature, gait recognition has two major approaches including model-based and model-free [35–38]. Model-based methods use information about static and dynamic parameters of parts of the body such as upper limbs and lower limbs while model-free methods use either the normalized silhouette or the average silhouette as the gait feature. A sample normalized silhouette extracted from a raw gait sequence is shown in Fig. 2.3. Most gait recognition research has tackled the main challenges of changes in viewing angle (the angle of the person relative to the camera), clothing, walking surface, and walking pattern due to carrying objects or suffering from injury. Generally, the model-free methods seem to be less affected by gait video quality and are thus invariant with those changes. In addition, model-free methods have been proven to have lower computational cost than model-based ones [39]. For these reasons, three model-free

models were chosen to evaluate the gait anonymization success rate for the models presented in this thesis.

Fig. 2.3 Top row contains RGB gait images; bottom row contains normalized silhouette images.

The rest of this subsection briefly summarizes the three model-based gait recognition methods used to evaluate the anonymization success rate. The first one was proposed by Zheng et al. [12]. The main idea is to transform the view of the probe gait to that of the gallery view. Its implementation was reported to be easy and rapid. It uses gait energy images (GEIs) computed from a silhouette sequence as gait features. GEI representation was first defined by Han and Bhanu [40]. It is the average of a silhouette sequence and is computed using

$$g(x,y) = \frac{1}{T}\sum_{t=1}^{T} I_t(x,y), \tag{2.1}$$

where T is the number of frames in the sequence, I_t is the silhouette image at frame t, and x and y are values in $2D$ image coordinates. An example silhouette sequence and its GEI are shown in Fig. 2.4

Fig. 2.4 Gait energy image (GEI): images on left side represent silhouette sequence; image on right side is GEI of sequence.

There are two main processes in their method: signature registration and gait recognition. Partial least squares (PLS), a supervised dimension reduction technique, is applied to each process to extract a feature from the feature GEI of the original gait. With the registration

phase, after the feature is extracted (using PLS), singular value decomposition is used to build a vector transform model (VTM). With the identification phase, this pre-trained transform model is used to convert the viewing angle of the query gait to that of the registered gait set. Finally, the similarity of two gaits is determined by using the L1-norm value between features of these two gaits representing the two subject images under the same viewing angle. The more similar the two gaits, the smaller the distance.

The second gait recognition method was developed by Wu et al. [13]). It uses a deep learning-based model to recognize the most discriminative changes in gait patterns, which suggest changes in human identity. The model takes the GEIs of two gaits as input and outputs the similarity between them. The model is trained by feeding positive and negative sample sets to the network. A positive sample is obtained by randomly picking two gait sequences with different view angles of the same person. A negative one is obtained by picking two gait sequences with different view angles of two different people. This method was reported to be the first CNN-based gait recognition method. Its average gait recognition rate reached 94.1%.

The third gait recognition method is GaitSet [14], introduced by Chao et al. They asserted that the silhouette at each position has a unique appearance, so a silhouette contains information about its position. Therefore, the order of silhouette in a gait sequence is not important. The silhouettes in a gait sequence are thus and they aggregated into a set from which temporal information is to extracted. They also introduced a deep learning-based method for learning discriminative information from a set of gait silhouettes. It takes silhouettes of a gait sequence as input. The gait information of the elements in a set are aggregated by feeding the silhouettes into a set pooling operation that uses a set of convolution and pooling blocks. This set of feature maps is split into strips by using independent fully connected (FC) layers for each pooled feature. Finally, the FC layers are used to map the pooled feature to the discriminative space. This method was reported to be fairly flexible in terms of the number of silhouette frames and the order of silhouettes in the gait sequence. The average accuracy reach 95%.

2.3 Group-based Image/Video Anonymization Methods

Existing methods for anonymizing images and video can be divided into group-based and individual-based ones. The group-based methods hide the identity in a group of individuals

and are commonly used in relational database anonymization. These methods are robust against re-identification attacks by linking or inference. There are two main individual-based methods: k-anonymity and t-closeness. There have been many reports of k-anonymity being used for face image anonymization while there has been only one of t-closeness being used.

2.3.1 k-anonymity

k-anonymity was originally introduced for relational databases to protect against linking attacks [41]. It requires that at least k records have the same non-identification values, and therefore each record in a table must be similar to at least other (k-1) records with non-identification value.

Newton et al. [1] introduced the use of the k-anonymity concept for face image anonymization. The main idea of the k-anonymity algorithm is that, from a closed set of k images, a surrogate image is produced, and then all images in the set are replaced by the surrogate image to obtain an anonymized image set. This means that the performance of face recognition is theoretically limited to 1/k. The algorithm ensures that the images in the original set cannot be linked to the images in the anonymized set, so it is robust against linking attacks.

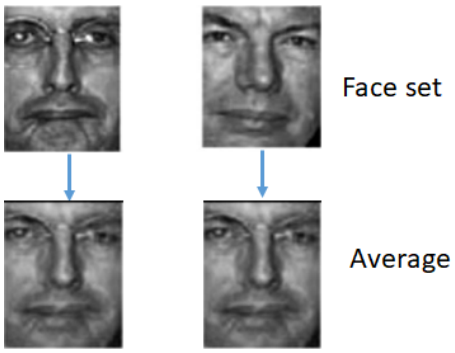


Fig. 2.5 k-anonymity for face image database (images from [1]).

More advanced face anonymization techniques that satisfy the k-anonymity requirement have been proposed [42–44]. In the k-Same-Net approach proposed by Meden et al. [2], a face image set is first clustered into a group. Then a pre-trained model is used to separate the identity attributes from the non-identity attributes, such as race and gender. Finally, from the identity attributes of all images in the original set and the non-identity attributes of the face image to be anonymize, a deep generative model is used to produce an anonymized face image.

Fig. 2.6 Examples of k-anonymity face anonymization results (reprinted from [2]).

2.3.2 t-closeness

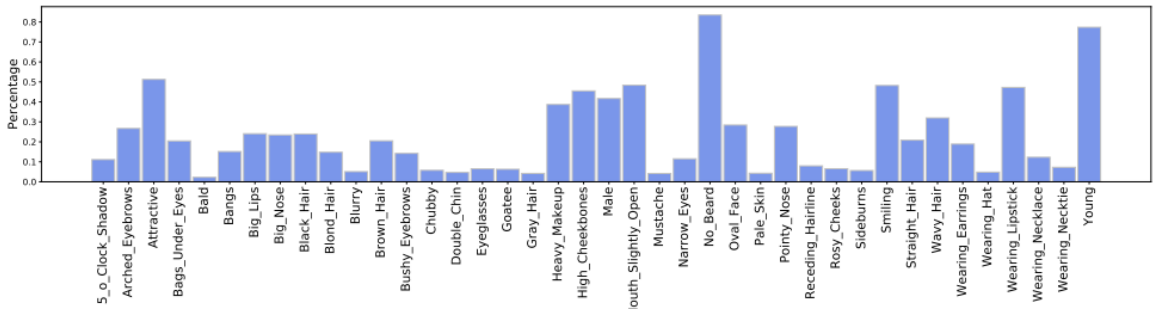


Fig. 2.7 Face attributes and their distribution (images from [3]).

The purpose of using k-anonymity is to prevent the joining of two datasets to achieve re-identification. However, if the sensitive attribute is the same for most records in class, the attacker can still infer the sensitive value of an individual in that class without re-identify the individual [45]. To address this problem, Li et al. [45] proposed using t-closeness, which updates k-anonymity on the basis of the correspondence of the sensitive attribute to the distribution of sensitive values. This requires that the distribution of sensitive values in an equivalence class be close to their distribution in the database.

The use of the t-closeness concept was introduced to face image anonymization by Li et al. [3]. They aimed to an generate anonymized face dataset so that the distribution of sensitive attributes of the anonymized dataset was close to the distribution of those of the original dataset. They defined the sensitive attributes of the face as the nose, eyes, lips, and so on. The distribution of face attributes in the CelebA dataset [46] is shown in Fig.2.7, and example anonymized face images generated with their model are shown in Fig.2.8. This

method prevents linkage from the anonymized identity to the real identity. It is also robust against inference attacks.

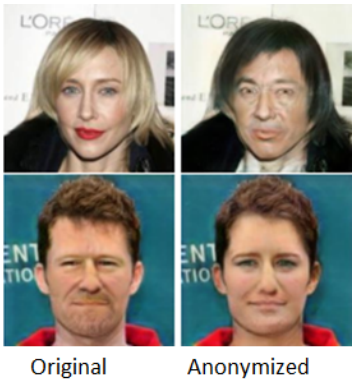


Fig. 2.8 Example anonymized gait images (from [3]).

2.4 Individual-based Image/Video Anonymization Methods

The methods in this group remove the identity of the person in the original image/video so that recognition systems are unable to link the anonymized identity to the real identity.

2.4.1 Naive Methods

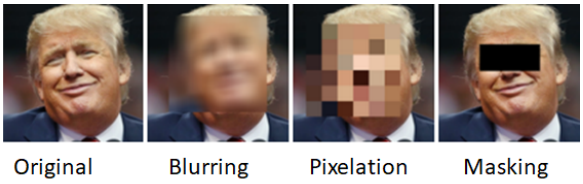


Fig. 2.9 Original image and anonymized images generated by naive methods (reprinted from Li et al. [3]).

Several methods have been proposed for face image anonymization and privacy protection of a person in a video, including blurring, pixelation, masking, and replacement. [6, 47, 48]. In the masking approach, after the face region in the image is located, it or part of the face region is simply masked by assigning the pixels to zero or one. With pixelation, the resolution of a privacy sensitive region is reduced. The image is divided into blocks, the average value of the pixels in every block is computed, and the pixels in each block are replaced by the

average value. This reduces the resolution of the privacy sensitive regions in the image. Pixelation is commonly used in television for anonymization of criminal suspects, witnesses, and bystanders.

Fig. 2.10 Examples of blurring: original images are shown on the left; corresponding blurred images are shown on the right (reprinted from Agrawal et al. [4]).

Blurring is a simple method based on obfuscating the privacy sensitive regions in an image by using image filters. The filters modify the pixel values of privacy sensitive regions by using neighboring pixels. Depending on the application, this method can be used to obscure the faces and bodies of people and even the background in images and videos [6, 49, 50]. The use of blurring has also been proposed for gait anonymization.

Qureshi [33] presented a blurring method in which a separate video bitstream is generated for each foreground object in the raw video. This enables the original video to be reconstructed from the various object-video streams without data loss. During reconstruction, each video bitstream can be rendered in several ways, for instance, obscuring people identities using a silhouette representation or simply not showing an object-video stream at all.

Agrawal and Narayanan [4] introduced a method for gait anonymization that combines two transformations: exponential blurring of the pixels of the target person in the video and line integral convolution. These two transformations smooth and blur the boundary of the individual and thereby remove the gait information that can be used for identification. The output color for each pixel composing the individual is a combination of the neighboring

average colors. Example original images and the images produced by their method are shown in Fig. 2.10.

A method presented by Chen et al. [5] is aimed at hiding the identity of a person in a video by obscuring the person's image. To this end, they proposed a pseudo-geometric model using an edge motion history image. First, the contour of person's body is extracted at each frame, and then an obscured body image is obtained at the current frame by adding the body image at the previous frame. This results in the whole body being obscured while preserving the structure and movement information. As shown in Fig. 2.11, the people in the image end up looking like ghosts.

Othman et al. [51] and Ruchaud et al. [52] developed methods for gender anonymization while maintaining some information for body shape, actions, and motion recognition.

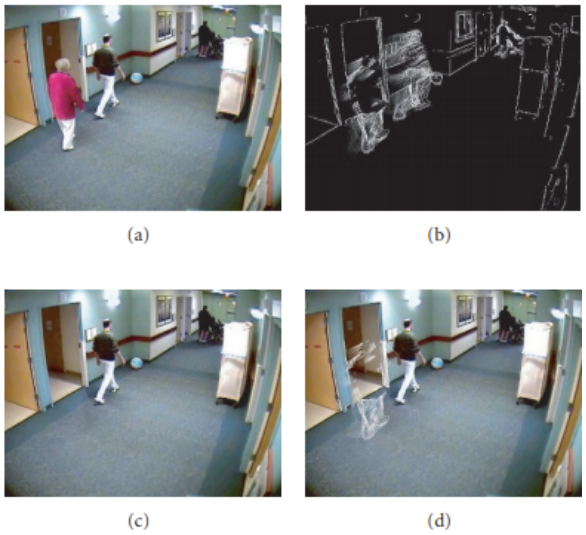


Fig. 2.11 Example of obscuring people: (a) original image, (b) edge motion history image, (c) original image restored but with woman in pink shirt removed, (d) final obscured image (reprinted from Chen et al. [5]).

Another anonymization method was introduced by Fan et al. [6]. Motivated by the need for privacy protection in medical videos, they used digital human models to change the appearances of the people in videos into virtual human objects. The resulting final video streams protect the privacy sensitive information of individuals in the video scene. Example results are shown in Fig. 2.12.

These naive methods focus more on privacy protection than on original attributes preservation.

Fig. 2.12 Example of human privacy protection using digital human model [6]: (a) original image, (b) image generated by replacing people in original image with virtual human objects.

2.4.2 Image Inpainting

Inpainting methods are aimed at reconstructing damaged regions in an image so that the modification is undetectable. These methods use information from the surrounding area to fill in the damaged regions. Inpainting technique have been used for both image [53, 54] and video [55, 56] applications.

Inpainting techniques are used in object removal approaches to protect the privacy of an individual in an image or video. These approaches first remove the person whose privacy is to be protected. Then inpainting is used to reconstruct the region damaged by the removal. The object removal approach was proposed by Granados et al. [7], and Chen et al. [5] to protect an individual in an image or video. Example images illustrating removal of a person from a video are shown in Fig.2.13.

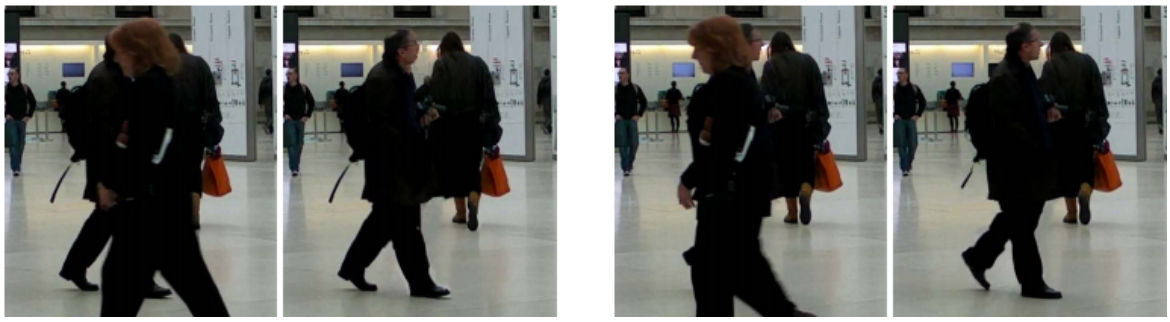


Fig. 2.13 Example images illustrating removal of person from a video: first and third images are original images; second and fourth images are resulting images after removal of foreground person (reprinted from Granados [7]

.

Inpainting techniques are unable to generate natural images when applied to face image anonymization. This is because, in order to protect the privacy information, a large mask (such as a back rectangle) must be used. As a result shown in Fig.2.14, there are not enough clues for use in generating a natural image by inpainting.

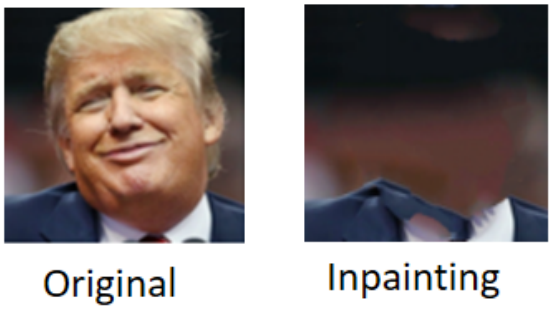


Fig. 2.14 Original face image and anonymized image generated by inpainting (reprinted from Li et al. [3].

2.4.3 Texture Modification

Fig. 2.15 Example of face image anonymization by replacing texture of original face image (reprinted from Samarzija et al. [8]).

With the texture modification approach, images are anonymized by modifying the image texture while preserving the object shape. Samarzija et al. [8] proposed an algorithm for anonymizing a face image by replacing its texture with that of another face image that has the same pose. The shape of the original face is retained while the image texture is changed. This process is illustrated in Fig. 2.16. Zhang et al. [9] presented a method for iris image anonymization. The iris region is blurred so that the image texture is changed while the shape of the iris is preserved. Although such methods are effective for anonymizing the face and iris, they do not change the gait pattern because the gait feature is based on the shape of the body (e.g., the silhouette shapes).

Fig. 2.16 Example anonymized iris image generated by blurring iris region (reprinted from Zhang et al. [9]).

2.4.4 Deep Learning

Several methods combine the idea of inpainting with a generative model to remove the identity of a face image. These methods remove most of the information from a face image, resulting in either a damaged image (which can be used as a landmark face image as shown in Fig. 2.17 or in a masked face image as shown in Fig. 2.18). A generative model is then used to produce a new image from the damaged image.

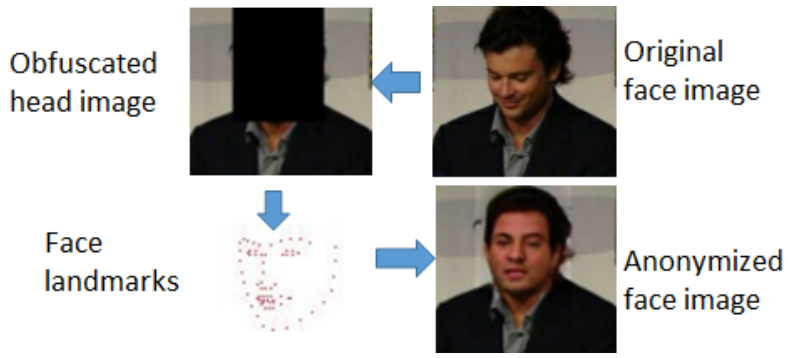


Fig. 2.17 Face image anonymization proposed by Sun et al. [10].

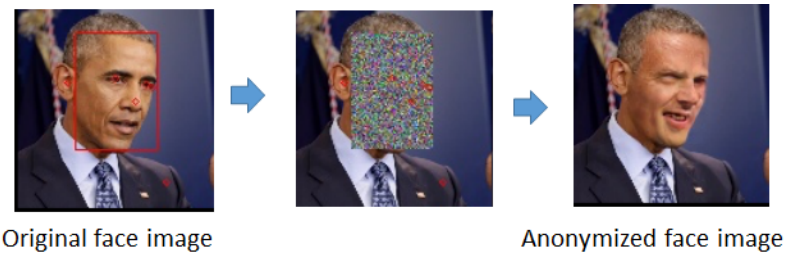


Fig. 2.18 Face image anonymization proposed by Hukkelås et al. [11].

Although these methods have been shown to generate natural face images, they do not ensure consistence among the frames of the generated gait sequence because the models

used are not designed to control temporal information and too much information is removed. Figure 2.19 shows an example of inconsistency among consecutive frames (the person is wearing glasses in some frames but not in other frames).

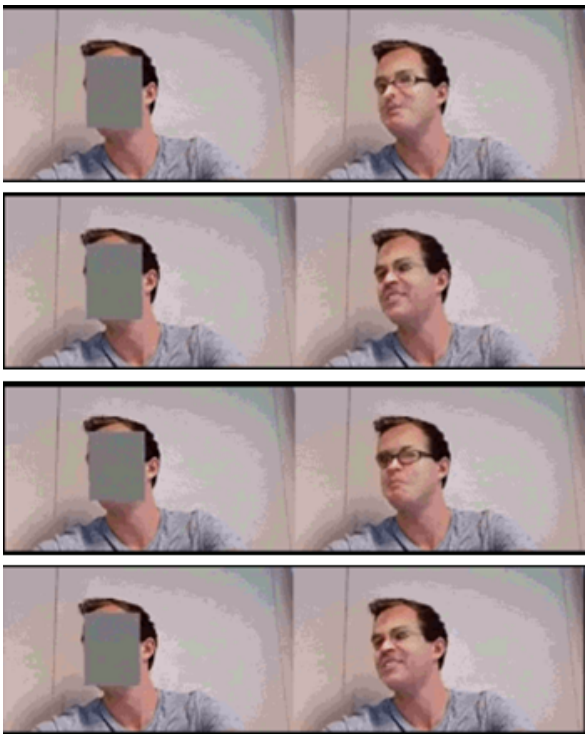


Fig. 2.19 Inconsistency among consecutive frames generated using method proposed by Hukkelås et al. [11].

Several deep learning models have been proposed for anonymizing faces in images and videos by using face verification [57–59]. Face verification is used to maximize the distance between a test face image and the generated face image, thereby forcing the model to remove the facial identity, as shown in Fig. 2.20.

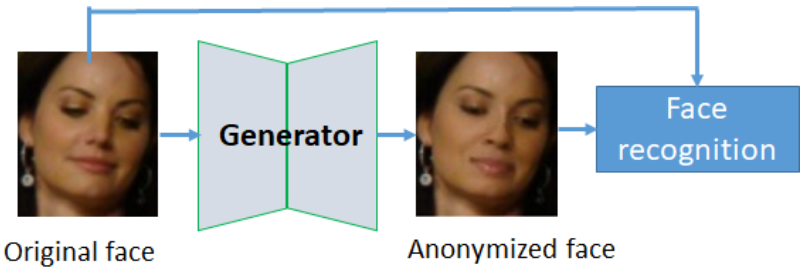


Fig. 2.20 Face verification is used to maximize distance between test face image and generated face image.

However, these models are not applicable to gait anonymization because the input for gait verification are binary images, and these models output gray images that must be converted into binary images using a binarization function. However, a binarization function is a discrete function, so it has no gradient, as illustrated in Fig.2.21.

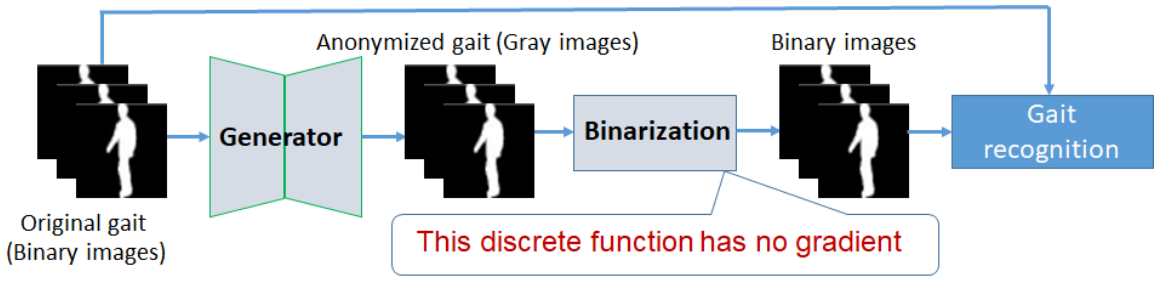


Fig. 2.21 Gait verification cannot be used to maximize distance between test face image and generated face image.

2.5 Issues with Existing Image/Video Anonymization Methods

Although group-based methods are robust against linking and inference attacks, they are not good at preserving the original attributes, as can be seen in Figs. 2.6 and 2.8. The individual-based methods ensure unlinkability from the anonymized identity to the real identity. Moreover, they are better at preserving the original attributes than the group-based methods. Therefore, the research described in this thesis focused on the individual-based approach; the group-based approach is left for future research.

The comprehensive review of the use of individual-based methods for image and video anonymization in section 2.4 demonstrates that such methods are not suitable for gait anonymization they (1) do not generate natural images, (2) are not applicable to gait anonymization, (3) do not change the gait pattern, and (4) do not guarantee consistency among frames. The models proposed in this thesis for gait anonymization address the problems with existing methods. Noise is added to the original gait to modify the shape of the silhouette, which enables gait recognition systems to be fooled. Reconstruction loss is used to ensure that the moving direction of the original gait is preserved. The reconstruction loss and the use of a spatial discriminator guarantee the naturalness of the body shape of the

generated gait. The temporal discriminator ensures consistency among the frames of the anonymized gait.

2.6 Performance Analysis of Gait Anonymization Models

To evaluate the performance of the proposed models, two metrics were used: *naturalness* and *success rate*, which are defined in subsections 2.6.1 and 2.6.2. The robustness of the proposed models against re-identification attacks was also evaluated.

2.6.1 Naturalness

The naturalness metric was to evaluate preservation of the original attributes. The preservation of appearance means ensuring the naturalness of the body shape and of the clothing colors in the original gait images. The preservation of motion consists of guaranteeing the naturalness of movement and retaining the moving direction of the original gait. Subjective evaluation was used to measure naturalness.

The mean opinion score (MOS) was used for the subjective evaluation. MOS test is commonly used to assess the quality of generated media from the human perspective [60],[61]. For example, MOS testing was used in the Wavenet model [62] to evaluate preferences for generated audio waveforms and to assess the quality of images created with a semantic rectifying GAN (SR-GAN) [63]. After watching a pair of anonymized and original video, evaluators evaluate the data utility degree of the anonymized gait with the score ranging from 1 to 5 (1: Bad, 2: Poor, 3: Fair, 4: Good, 5: Excellent).

2.6.2 Success Rate

The success rate is the percentage of gaits successfully anonymized. It was computed similar to the way that Sharif [64] used. It is typically stated as the ratio of the number of anonymized gaits, which are incorrectly recognized to the total number of anonymized gaits, with S defined as the set of gaits correctly recognized by the system, S' as the number of anonymized gaits in S, and M as the number of gaits in S' incorrectly recognized.

$$success\_rate(\%) = \frac{M}{|S|} \times 100\% \tag{2.2}$$

2.6.3 Robustness against Re-identification Attacks

A re-identification attack [65, 66] is an attack attempting to recover anonymized data in order to restore the identity information. In the research reported in this thesis, a commonly used method was used to try to reverse the identity information from anonymized gaits. To demonstrate that the proposed gait anonymization methods are robust against re-identification attacks, both re-identified gait images and the identification accuracy on such gaits are presented.

2.7 Available Dataset

To evaluate the proposed gait anonymization methods, the CASIA-B gait dataset [67], a huge, widely used dataset with multiple viewing angles and different challenging conditions, was used. This dataset contains 110 gait sequences for 124 individuals, with 11 viewing angles each ($0^0, 18^0, \ldots, 180^0$). The frame images of the videos are RGB. An example showing the 11 viewing angles is presented in Fig. 2.22, while an example showing the multiple clothing colors is presented in Fig. 2.23.

Although the models introduced in this thesis do not require that all gait sequences have a fixed length, the models can be implemented more easily and trained more quickly if the sequences all have the same length. Therefore, another dataset was created from the original CASIA-B dataset that contains sequences of the same length, with each sequence having 50 frames. The 50 frames were taken from the end of longer sequences. This new dataset was used for training, validating, and testing the gait recognition systems and the proposed gait anonymization models.

Fig. 2.22 Example of 11 viewing angles in CASIA-B dataset.

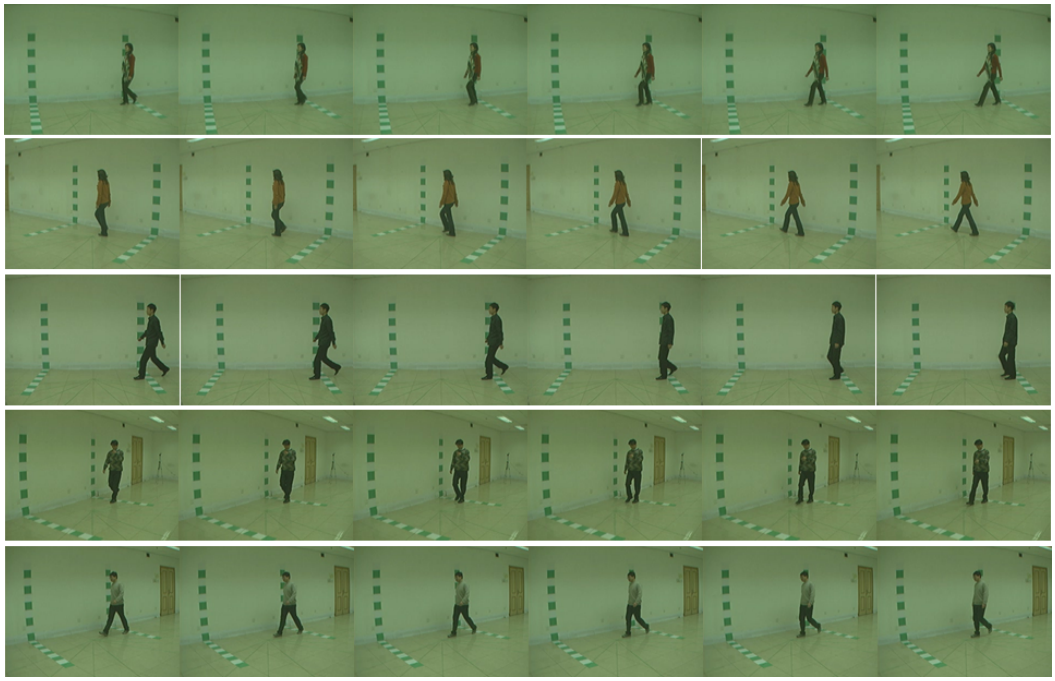


Fig. 2.23 Examples of clothing colors in CASIA-B dataset.

Chapter 3

Binary Gait Anonymization

3.1 Introduction

Most of the current model-based gait recognition approaches explore silhouette sequences of subjects that are a sequence of binary images containing the shape and style working of each subject as the gait's feature. In this thesis, gait anonymization while preserving the gait's naturalness is divided into two tasks anonymization and colorization as illustrated in Fig. 3.1. The first task, anonymization, takes binary silhouettes as the inputs, and output the binary anonymized gait images. This chapter focuses on the second task, anonymization. This chapter aims to investigate that it is possible to fool gait recognition systems by modifying the body shape and how to apply deep learning to ensure the naturalness of the anonymized gait in terms of shape and movement. In the other words, this chapter aims at generating the binary anonymized gait from binary real gait so that the anonymized gait does not consist of the identity of the real gait and the anonymized gait moves smoothly and its shape looks natural. The idea of the method in this chapter is modifying the body shape of a gait that we desire to anonymize by adding to it another gait named "noise gait". Therefore, the coordinates of the contour body (or contour coordinates) of a gait at each frame are used, and then contour vectors are computed from these contour coordinates. A convolutional network is designed, named BiGait-ANET, which has two inputs: the contour vector of the gait to be anonymized and that of the noise gait and the output is the modified contour vector. The anonymized gait is produced with these modified contour vectors by filling these contour vectors.

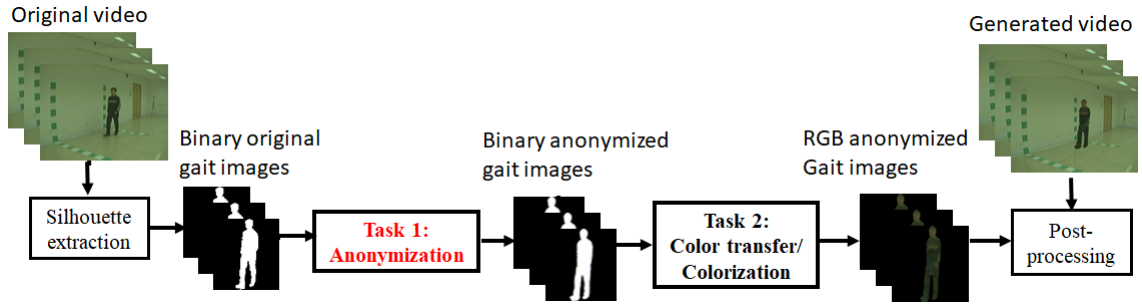


Fig. 3.1 Gait anonymization study in this chapter focuses on anonymization task.

3.2 Methodology

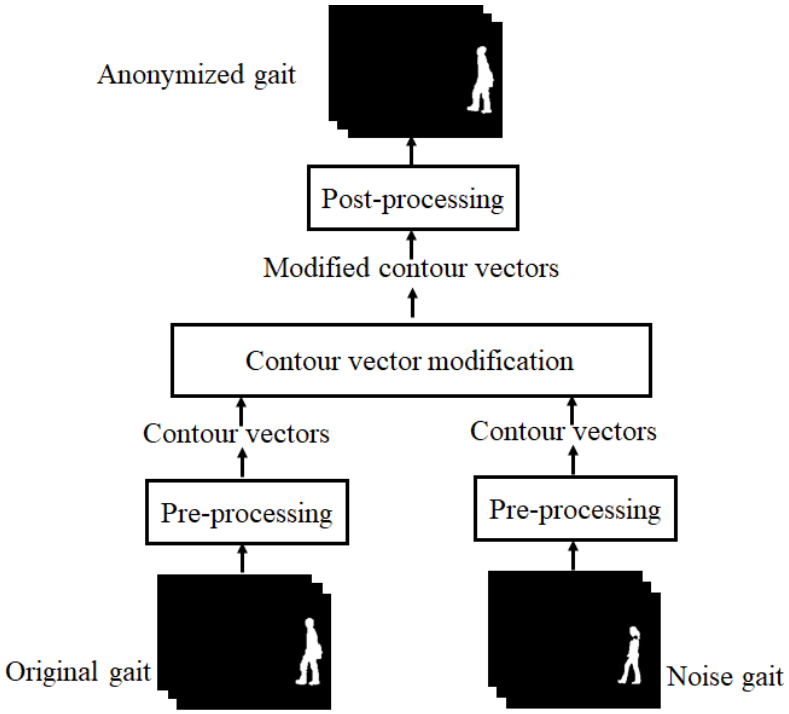


Fig. 3.2 The proposed method includes three steps: Pre-processing, Contour vector modifying (CNN), and Post-processing.

To generate the anonymized gait, the original gait is mixed with the noise gait using a CNN-based model. To modify the body shape of a gait, contour coordinates of the body of a gait at each frame rather than the gait image are used. The proposed model consists of three steps and is presented in Fig.3.2.

Step 1 - Pre-processing: The step firstly extracts the contour of silhouettes of original and noise gait, the original gait, and the noise gait. Then, these contours are converted to vectors, which are called contour vectors.

Step 2 - Contour vector modification: Two contour vectors created from Step 1 become the inputs of the CNN, which takes over the change of the original gait, and the output here is the modified contour vector.

Step 3 - Post-processing: The contour vector generated by Step 2 is passed through the post-processing to produce the anonymized gait, and then the final gait is then pasted into the original video at the similar position to the original gait.

In the remaining part of this section, The detail about the above three steps as well as how to select the noise gait will be explained.

3.2.1 Pre-processing

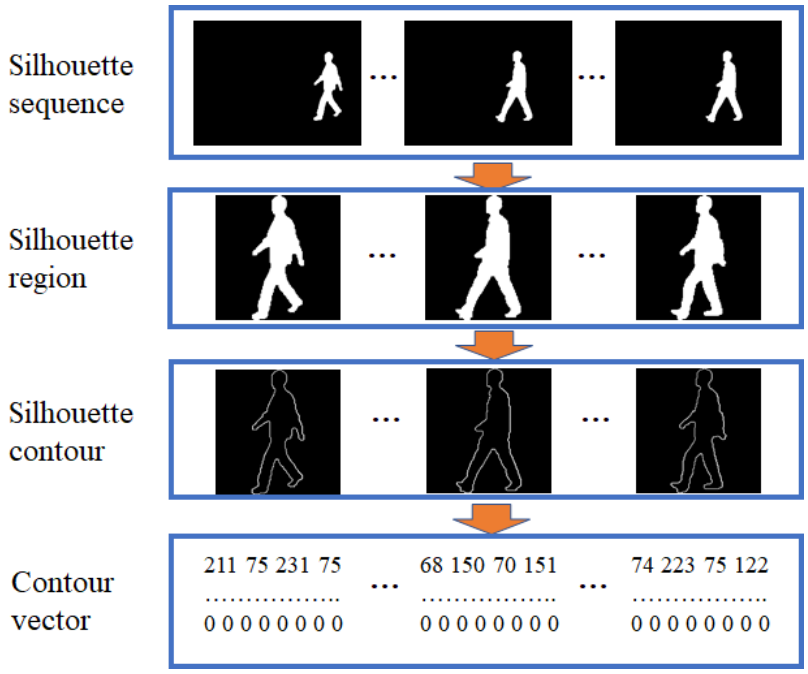


Fig. 3.3 The flow of the pre-processing step.

The crucial idea of the proposed model for binary gait anonymization is altering the body shape of the original gait by adding to it a noise gait. Therefore, silhouette regions are firstly cropped, and then the extracted images are resized to the same size of 240×240.

The frames with the size of 240×240 containing silhouettes are cropped and normalized, then coordinates of pixels on the silhouette contours were computed. Next, such coordinates are transformed to vectors that are used as the inputs of Step 2. The length of these vectors was fixed to 4000. Each vector consists of two coordinates of each pixel on the contour and is corresponding to 2000 pixels on the contour. Note that there are no contours that have the pixel number larger than 2000 in the dataset. For contours that have less than 2000 pixels, the zeros padding is added to the rest of the vector. Therefore, the body shape of a silhouette is now represented by a vector of length 4000 that consists of two parts: one contains coordinates of contour pixels, one contains the zero-padded area. Fig.3.3 illustrates all steps of pre-processing phase.

3.2.2 Contour Vector Modification

This section explains more detail about how to modify the original gait and the network for this step is showed in Fig.3.4. This model takes 2 inputs: the original gait's contour vector and the noise gait's contour vector. The output is modified contour vector. The two inputs go through two shared weights networks to get the abstract representation. The network can be expressed in the following formula:

$$\Phi_1(X_1) = ReLU(W_1 * X_1 + b_1) \tag{3.1}$$

$$\Phi_1(X_2) = ReLU(W_1 * X_2 + b_1) \tag{3.2}$$

The network composes of convolutional functions. The first two convolutional functions include weights matrix W_1 multiplied with two vectors X_1, X_2 and the bias b_1, where X_1, X_2 are two input contour vector of original and noise gait, respectively. Nonlinear operations $ReLU(x) = \max(x,0)$ follow these convolutional functions. The goal of this model is to produce the modified contour vector, so these two functions are then combined into one network as follows.

$$\Phi_2(X_1,X_2) = ReLU(W_2 * (\Phi_1(X_1) + \Phi_1(X_2)) + b_2) \tag{3.3}$$

where W_2, b_2 are weights matrix and bias of the combined network. All parameters of this the contour vector modification network are computed by optimizing the loss function below:

$$\frac{1}{D_{X_1}}\left(||\Phi_2(X_1,X_2)-X_1||^2+\alpha||\Phi_2(X_1,X_2)-X_2||^2\right) \tag{3.4}$$

This loss function forces the output to be partly similar to the original gait and partly similar to the noise gait. The first term is to force the generated gait to be similar to the original gait and hence it maintains the naturalness of the gait images. The second term is to force the output to be partly similar to the noise gait and an adjustable parameter indicates how much the noise gait added to the original one. In the experiments, α was set equal to 3. D_{X_1} is the length of the contour vector and equals 4000 in our case.

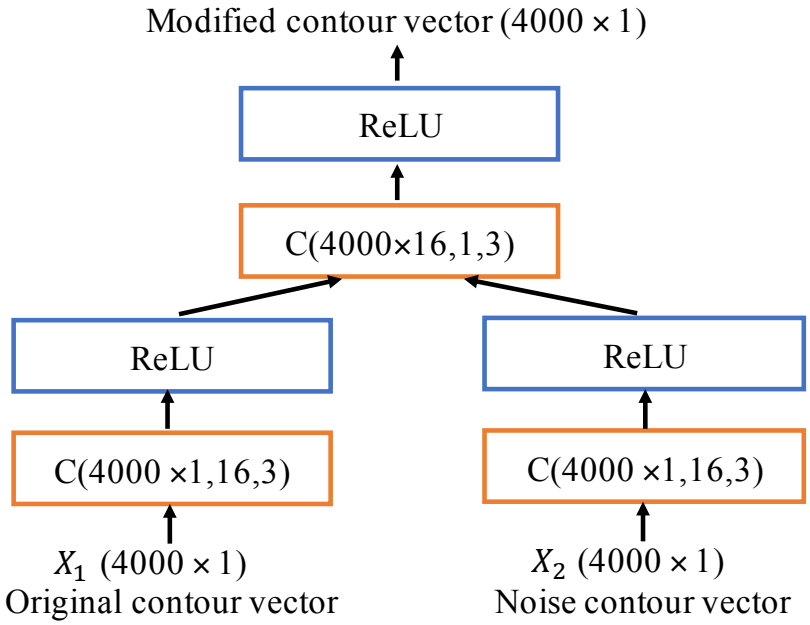


Fig. 3.4 The architecture of the propose model BiGait-NET: C are convolution.

3.2.3 Post-processing

This process is to generate the final video in which the original gait is removed and the anonymized gait is placed in the same position as the original one. This process firstly creates an image of contour from the modified contour vector generated by Step 2, then the region inside the contour in this image is filled to get the image of the modified silhouette. Finally, the silhouette of the original gait in the original video is replaced by the modified one.

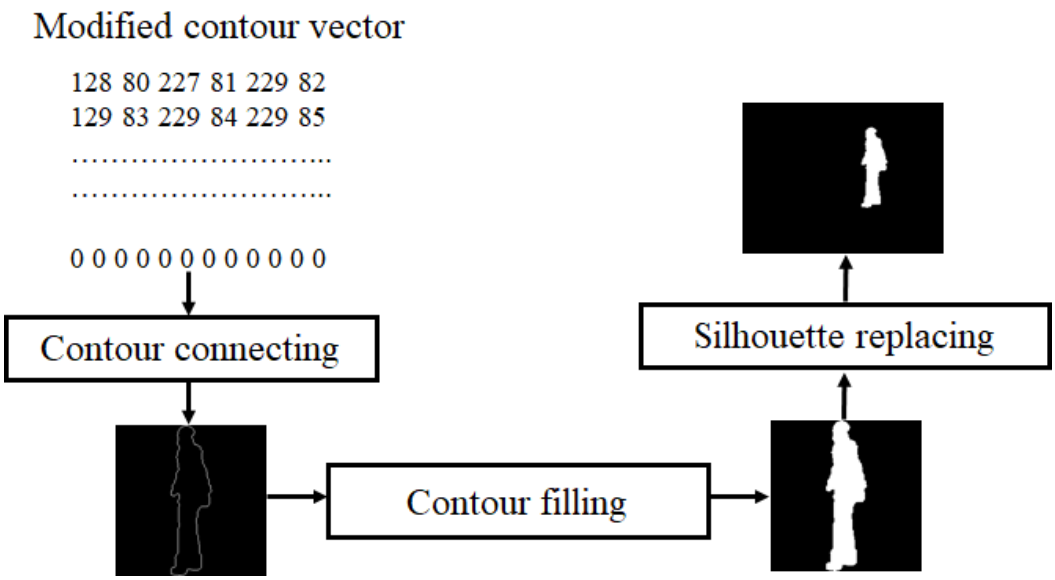


Fig. 3.5 The flow of post-processing.

3.2.4 How to Select the Noise Gait

Selection of the noise gait is one of the important aspects of model BiGait-ANET because it influences the anonymization success and the naturalness of the generated gait. The role of noise gait is to hide the original identity of a gait to be anonymized, however, it may make the naturalness of a generated gait decrease. Therefore, the optimal noise gait is a gait the can hide the identity of the original gait while ensuring much naturalness of the anonymized gait as possible. To choose such optimal noise gaits, two cases are considered: (1) the viewing angles of the original gait and the noise gait differ; (2) the viewing angles of the original gait and the noise gait are the same. The result of first case is shown in Fig.3.6, while that of the second case is in Fig.3.7. In those figures, the noise gaits, original gaits, and the generated gaits are in the first, second, and third row, respectively.

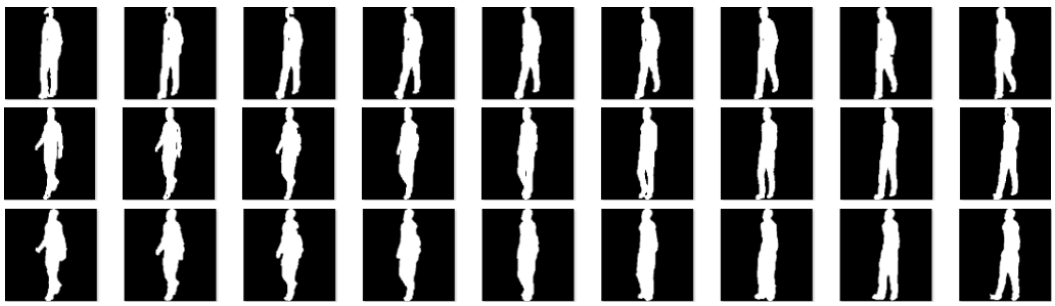


Fig. 3.6 The viewing angle of the original gait and that of the noise gait are the same.

Fig. 3.7 The viewing angle of the original gait and that of the noise gait differ.

With the comparison of the generated gaits in two cases, it is clear that the gait generated with the noise gait in the same viewing angles (Fig.3.7) looks more realistic than that generated with the noise gait in the different viewing angles (Fig.3.6), therefore, in experiments, the noise gait is chosen so that it is not the original gait but has the same view angle as original gaits.

3.3 Experimental Results

The dataset that includes 124 subjects is divided into four non-overlapping parts as shown in Table 3.1. The first part containing 50 subjects is to train the three gait recognition systems. The second part has 24 subjects and is kept to train BiGait-ANET network. The third part containing 10 subjects is to validate the proposed model BiGait-ANET. The remaining part with 40 subjects is preserved for testing. Table 3.1 shows the detail of dataset organization.

Silhouette images of anonymized gaits generated by BiGait-ANET and the evaluation of the proposed model performance with naturalness, success rate and robustness is presented in four following subsections.

Table 3.1 Dataset organization.

Tasks	**Num. of subs**	**Num. of seqs.**	**Num. of frames**
Training the proposed model	24	2640	132,000
Training the gait recognition systems	50	5,500	275,000
Validation	10	1,100	55,000
Testing	40	4,400	220,000

3.3.1 Generation Results

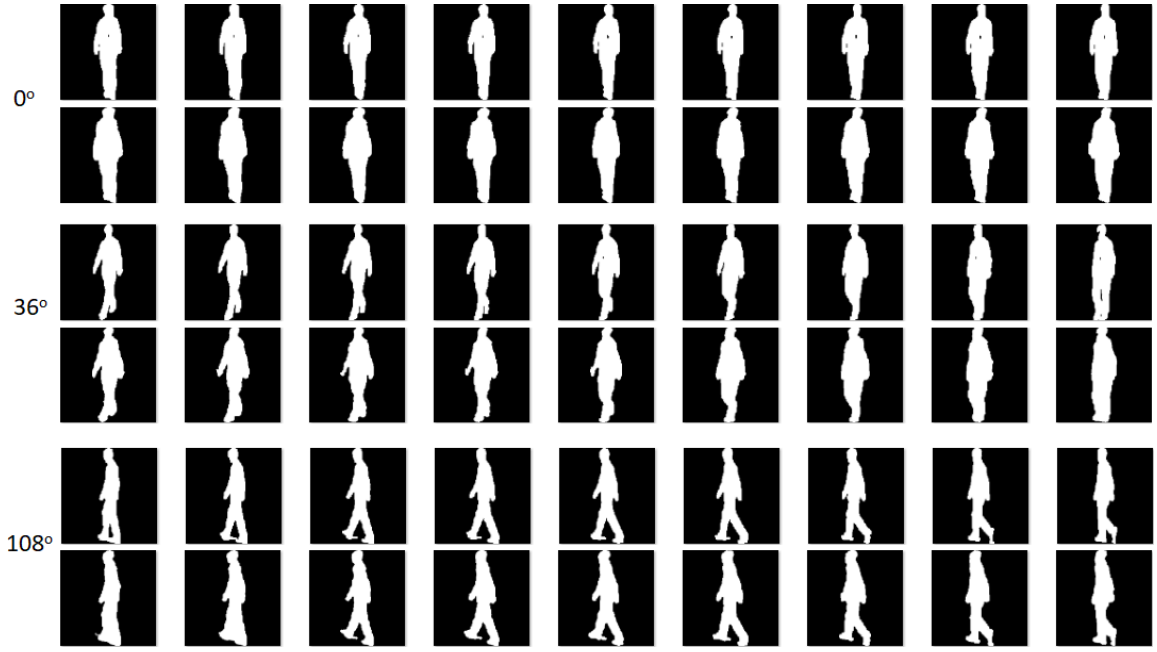


Fig. 3.8 Silhouettes of original gaits and anonymized gaits with various viewing angles: the first rows show the silhouettes of original gaits, the second rows show silhouettes of anonymized gaits.

Fig.3.8 illustrates the generation results of the model BiGait-ANET for one subject under different viewing angles. This figure demonstrates that the proposed model is able to keep the naturalness of the generated gaits in terms of movement and shape. However, the shape of gaits with the side viewing angles look more natural that that at the frontal viewing angles (0^0,180^0).

3.3.2 Naturalness

The naturalness evaluation of the anonymized gaits was done by MOS test and there were 30 evaluators who attended this test. In detail, each evaluator was given 30 random pairs of original and anonymized gait videos and each video is 10 seconds long. After watching each pair, they gave a score for the naturalness of the corresponding anonymized video in a five-point scale score (1: Bad, 2: Poor, 3: Fair, 4: Good, 5: Excellent). The result of MOS scores for each viewing angle are demonstrated in Fig.3.9. From this figure, the findings are withdrawn:

(1) The naturalness of generated gaits evaluated by the human perspective of viewing angle 180^0 got the lowest with 2.94, and that of viewing angle 54^0 was the highest score with 3.73.

(2) The naturalness of generated gaits evaluated by human perspective at the side viewing angles is higher than that at the frontal viewing angles. This result is coherent with the generation results shown in subsection 3.3.1.

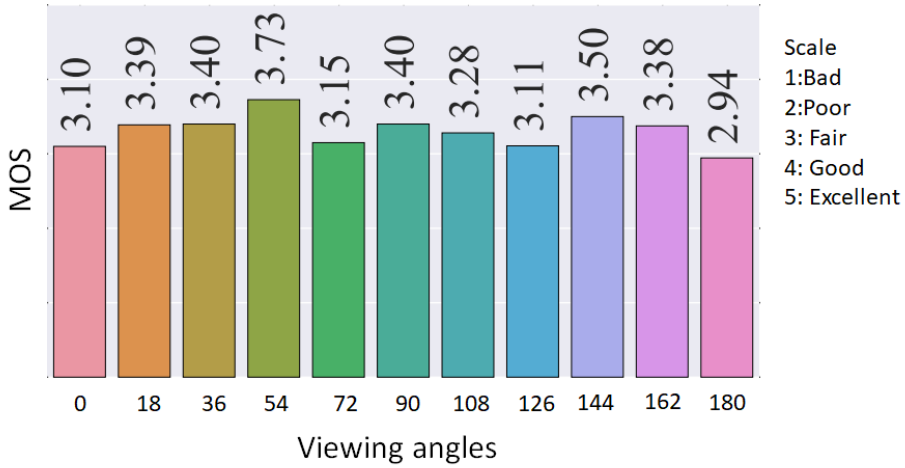


Fig. 3.9 MOS of the generated gaits.

3.3.3 Success Rate

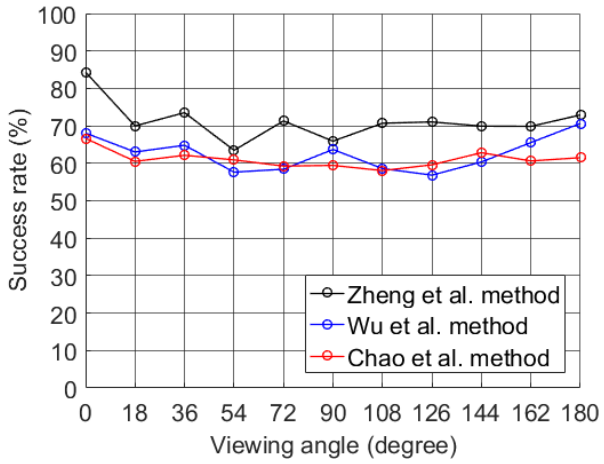
The success rate of the proposed model BiGait-ANET measured for three gait recognition systems is shown in Table 3.2, Table 3.3, and Table 3.4. The comparison of success rate with three gait recognition systems is summarized in Fig.3.10.

Table 3.2 The average success rate (%) of the proposed model BiGait-ANET evaluated with Zheng's [12] method.

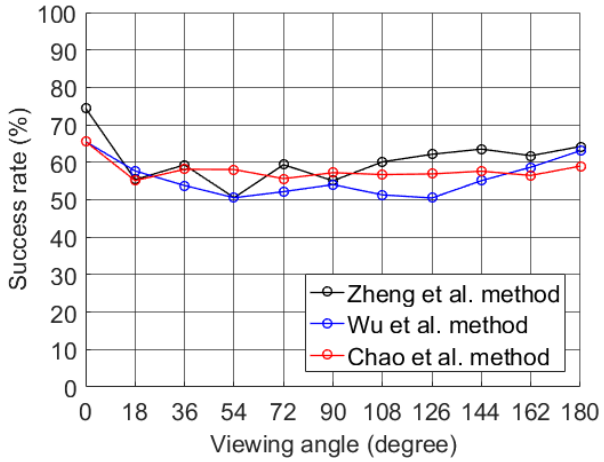
Top	**Viewing angles**										
	$\mathbf{0^0}$	$\mathbf{18^0}$	$\mathbf{36^0}$	$\mathbf{54^0}$	$\mathbf{72^0}$	$\mathbf{90^0}$	$\mathbf{108^0}$	$\mathbf{126^0}$	$\mathbf{144^0}$	$\mathbf{162^0}$	$\mathbf{180^0}$
top – 1	84.35	69.98	73.52	63.47	71.31	65.91	70.70	71.04	69.93	69.87	72.91
top – 3	74.47	55.39	59.31	50.56	59.35	55.07	60.11	62.15	63.53	61.77	64.22

With the experimental results, two findings are withdrawn:

(1) The success rates are higher than 51% with top-1 and higher than 44% with top-3 for all three gait recognition systems. That means the proposed model BiGait-ANET is able to achieve the promise anonymization success.



(a) Comparison of success rates with top-1



(b) Comparison of success rates with top-3

Fig. 3.10 The comparison of success rates with three gait recognition systems.

(2) The success rates with Zheng's method are the lowest for all viewing angles because this gait recognition system is the least robust. The success rates with Chao's method are the highest since this gait recognition model used the sequence of silhouettes as the input that captures much more identity information than the GEI feature, which is used in Zheng's and Wu's method.

(3) Comparing Fig.3.9 and Fig.3.10 show that the success rate and the naturalness seem to be somehow inversely proportional to each other, in general. However, this rule seems to be incorrect for some cases, for instance with view angle 0^0, due to low performances of gait identification systems at those viewing angles.

Table 3.3 The average success rate (%) of the proposed model BiGait-ANET evaluated with Wu's [13] method.

Top	**Viewing angles**										
	$\mathbf{0^0}$	$\mathbf{18^0}$	$\mathbf{36^0}$	$\mathbf{54^0}$	$\mathbf{72^0}$	$\mathbf{90^0}$	$\mathbf{108^0}$	$\mathbf{126^0}$	$\mathbf{144^0}$	$\mathbf{162^0}$	$\mathbf{180^0}$
top – 1	68.06	63.02	64.81	57.63	58.42	63.68	58.58	56.84	60.35	65.54	70.63
top – 3	65.54	57.69	53.80	50.53	52.15	54.03	51.29	50.47	55.10	58.70	63.10

Table 3.4 The average success rate (%) of the proposed model BiGait-ANET evaluated with Chao's [14] method.

Top	**Viewing angles**										
	$\mathbf{0^0}$	$\mathbf{18^0}$	$\mathbf{36^0}$	$\mathbf{54^0}$	$\mathbf{72^0}$	$\mathbf{90^0}$	$\mathbf{108^0}$	$\mathbf{126^0}$	$\mathbf{144^0}$	$\mathbf{162^0}$	$\mathbf{180^0}$
top – 1	66.65	60.51	62.14	60.91	59.26	59.44	58.05	59.60	62.83	60.63	61.50
top – 3	65.55	55.12	58.13	58.07	55.64	57.23	56.72	56.90	57.62	56.49	58.99

3.3.4 Robustness against Re-identification Attack

This subsection investigates whether the proposed gait anonymization model is vulnerable to re-identification attacks. This issue is analyzed with a non-machine-learning-based method and machine-learning-based method. Since the gait generator network consists of the non-reversible ReLU activation function, the proposed method can be considered as a one-way function. In addition to this, as shown in Fig. 3.11, the difference between original and anonymized gait images varies at each frame. These properties make it impossible to find a common formula to reverse the anonymized gait to the original gait.

Fig. 3.11 XOR images of original and anonymized gait images that show the differences between original gait images and anonymized ones.

This subsection is also aimed at investigating whether we can find a machine learning model for re-identification attacks. Since model-free gait recognition systems take silhouettes that are binary images as inputs, if we stack a gait recognition system on the top of the re-identification model to force this model to restore the identity of the original gait from the anonymized gait, the output of the re-identification model must be binarized. However,

the function that converts the output of the re-identification model to the binary images is a discrete function that has no gradient, and therefore the gradient of the whole model cannot be updated.

A traditional denoising autoencoder network [68] was used to try to restore the identity of original gait from an anonymized gait. We used 20 subjects in 40 anonymized subjects to train this network and used the remaining 20 anonymized subjects to evaluate this network. In order to explore the robustness of the proposed gait anonymization model presented in this chapter, we compute the identification accuracy of three gait recognition systems on the re-identified gaits. Table 3.5 shows the this accuracy and a sample of re-identified result is shown in Fig. 3.12. Both Table 3.5 and Fig. 3.12 demonstrates that our model is robust to the re-identification attack.

Table 3.5 The average identification accuracy (%) of the re-identified gaits evaluated with Zheng's method, Wu's method, and Chao's method.

Top	**Viewing angles**										
	$\mathbf{0^0}$	$\mathbf{18^0}$	$\mathbf{36^0}$	$\mathbf{54^0}$	$\mathbf{72^0}$	$\mathbf{90^0}$	$\mathbf{108^0}$	$\mathbf{126^0}$	$\mathbf{144^0}$	$\mathbf{162^0}$	$\mathbf{180^0}$
top – 1	4.05	3.51	3.60	4.40	3.81	3.54	3.60	3.10	3.60	3.87	3.87
top – 3	7.57	8.24	7.47	7.60	7.61	6.92	7.60	7.03	6.93	6.67	7.47

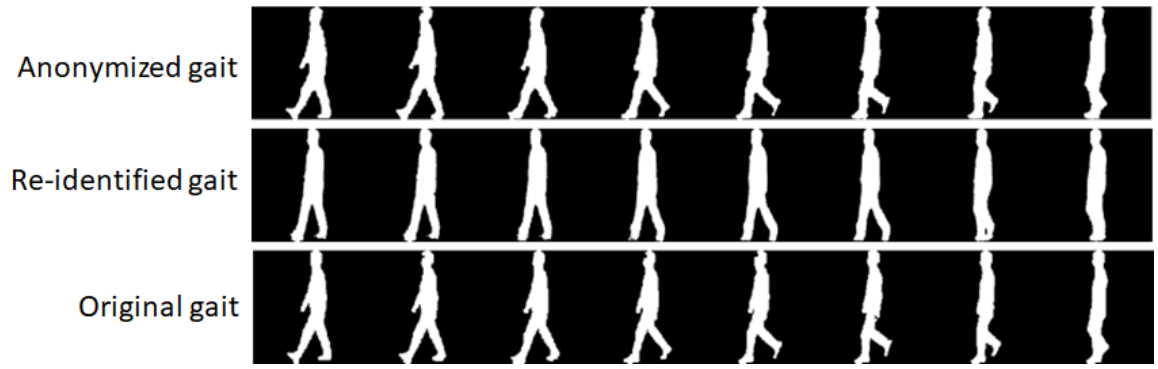


Fig. 3.12 The generation results of re-identification attack of BiGait-ANET with traditional denoising autoencoder.

3.4 Summary

In this chapter, gait anonymization on binary gait videos has been explored by using deep learning. A CNN-based method was designed in this chapter, which is fast and easy-to-implement, to alter the gait in a video so that a gait recognition system is unable to identify

anonymized gaits, while still preserving the appearance and the motion of the original gait. The scope of this chapter is just to solve the problem with the inputs are silhouette sequences, but it is the first research on gait anonymization while preserving the original appearance and motion.

Although the proposed method is is fast and easy-to-implement the experimental results are promising with the success rate, which is evaluated on three gait identification systems, can achieve 84.35% at most with the top-1 identification and the highest naturalness score is 3.73 in the MOS scale. Another finding is that success rate and the naturalness seem to be somehow inversely proportional to each other, in general. However, this trend is incorrect for some cases, for instance with viewing angles 0^0 and 180^0, due to low gait recognition performance at those viewing angles.

Chapter 4

RGB Gait Anonymization

4.1 Introduction

In Chapter 3, the model BiGait-ANET was proposed to anonymized a gait in the binary video. Though the model BiGait-ANET achieved promising results of both anonymization success and generated video quality, it is limited to black and white videos and has not solved well the consistency among frames of the generated gait. In addition to this, the performance of this model depends on the way to choose the noise gait, which is used to erase the identity features of a gait. In this chapter, gait anonymization while preserving the gait's naturalness focus on both tasks anonymization and colorization as shown in Fig. 4.1. Three research questions that have not been solved well by model BiGait-ANET will be considered and tackled in this chapter, that is 1) How to improve anonymization performance?, (2) How to ensure consistency among frames?, (3) How to transfer colors in original gait images to anonymized gait images?.

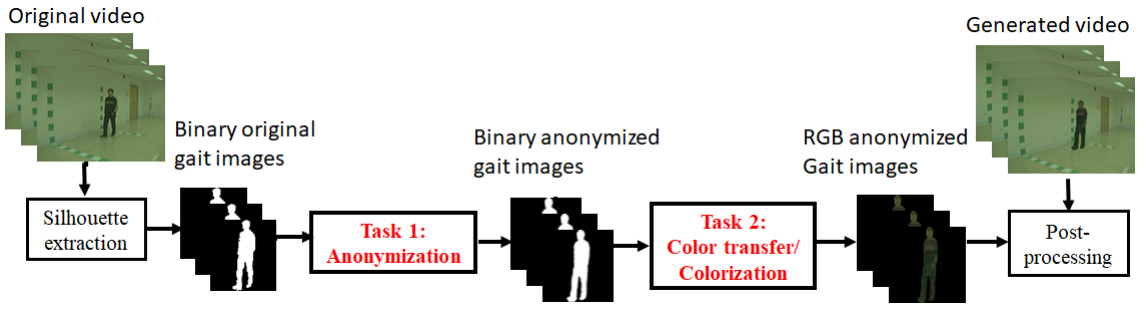


Fig. 4.1 Gait anonymization study in this chapter consists of two tasks anonymization and colorization.

Recent progress in the deep neural network has been receiving extreme success in various topics for instance image synthesis, image classification, recommendation, and security. Amongst various techniques of image synthesis, generative adversarial network (GAN) [69] shows enormous potential and becomes the most popular. There have been many modifications of GAN applied to numerous problems such as image synthesis [70, 71], image editing [72, 73], image super-resolution [60, 74].

Motivated by GAN, we propose a method for gait anonymization on RGB videos so that the gait is incorrectly recognized by gait recognition systems while remaining the naturalness in terms of shape, movement, and colors. This type of anonymized gaits can be applied in many situations: people want to upload their videos on to the internet, but also want to protect their personal information from the gait recognition; when a video captured a suspect is displayed on media channel, his/her gait should be anonymized; the people showed in the human surveillance systems, their gaits must not be recognized by the gait recognition systems.

In chapter 3, a CNN-based model, named BiGait-ANET, was proposed for gait anonymization. This method uses a convolutional neural network to anonymize the original gait, which is desired to anonymize, by using a noise gait. The noise gait was chosen so that it is not the same identity as the original gait, but in the same viewing angle as the original gait. n addition to this, using the noise gait, the anonymization success may decrease in the case that gait recognition systems confuse the original gait with noise gait. Therefore, the selection of the perfect noise gait is not easy. Moreover, this model does not guarantee to generate a natural-looking gait, especially for frontal and back viewing angle because the reconstruction loss is not strong enough to ensure that.

In this chapter, a novel method is introduced to address the two above problems. In this model, the noise gait is replaced by random noise in gait distribution that is generated by the traditional GAN model [69] from a random vector. This traditional GAN is trained before being assigned to the gait generator network, spatio-temporal generative adversarial network (ST-GAN), for natural anonymized gait generation. The model ST-GAN composes of several networks: one generator and two discriminators. The generator takes two inputs, that is the original gait and the random noise to synthesize the anonymized gait, meanwhile, two discriminators are to discriminate the actual gait and the synthesized gait. A colorization algorithm that transfers colors in an original gait image to a synthesized gait image is also

proposed, and therefore the proposed method can be applied to the color dataset instead of the binary one as presented in Chapter 3.

The performance of the proposed model was evaluated on dataset CASIA-B [67] with two metrics success rate and naturalness as presented in the previous Chapter. Three gait recognition systems were used [12–14] as the "black-box" to measure the success rate. To assess the naturalness of the anonymized gait, we use human annotators. The experimental results show that our model can generate the anonymized gaits that obtain a higher naturalness as well as success rate.

Contributions in this chapter are as follows.

- The proposed model can anonymize gait by adding random noise to remove the original identity.
- The proposed model can preserve the appearance and motion better than the BiGait-ANET due to the usage of spatio-temporal discriminators.

4.2 Generative Adversarial Network

A traditional generative adversarial network (GAN) that first is introduced by Goodfellow et al.[69] includes one generator and one discriminator. The generator is trained to generate new examples from a random vector, and the discriminator is trained to discriminate a real sample from the fake sample. The generator and discriminator are trained together until the denerator can fool the discriminator about half the time, at which the generator network is able to generate plausible examples.

Though many extensions for GANs have been proposed to generate very natural images [60, 75–77], not many GANs-based approaches for video generation. Saito et al. [78] proposed a Temporal Generative Adversarial network (TGAN) for video generation. TGAN consists of two generators (a temporal generator and an image generator), and a discriminator. The goal temporal generator is to generate a latent sequence from a random vector. This generator is followed by the image generator, which synthesizes image at each frame of the sequence. Tulyakov et al. proposed a MoCoGAN model to produce a video without a priming image [79]. The key idea is using motion and content. The sampling from the content subspace was accomplished by sampling from a Gaussian distribution. The sampling from the motion subspace was performed using an RNN. There are two discriminators. The image discriminator, which aims to distinguish real from fake single frames, is based on

convolutional neural network (CNN) architecture. The video discriminator, which aims to distinguish the real from the fake videos, is based on spatio-temporal CNN architecture. Vondrick et al. [80] proposed a Generative Adversarial Network for Video (VGAN) presented a model that consists of two generators to generate background and foreground separately. Both generators take a random noise as input. In order to distinguish the real or generated video, they used the spatio-temporal convolutional network, which takes the input as the sequence of images. Most existing methods use the input of random noise vectors to produce a video, but, research in this thesis wants to modify a given gait so that the generated gaits are incorrectly recognized by the recognition systems.

Meanwhile, Yan et al. [81] attempted to synthesize videos of a walking person from sequences of his skeleton and his static image. However, gait anonymization research tries to alter the pattern of one gait from the original video, therefore, Yan et al. method cannot be applied to the gait anonymization.

4.3 Methodology

4.3.1 Definitions and Notations

Definition 1 *(Contour vector): the contour vector of a frame is a* $\mathbb{N}^{4000}$ *vector whose elements are the coordinates of the pixels on the contour of the frame.*

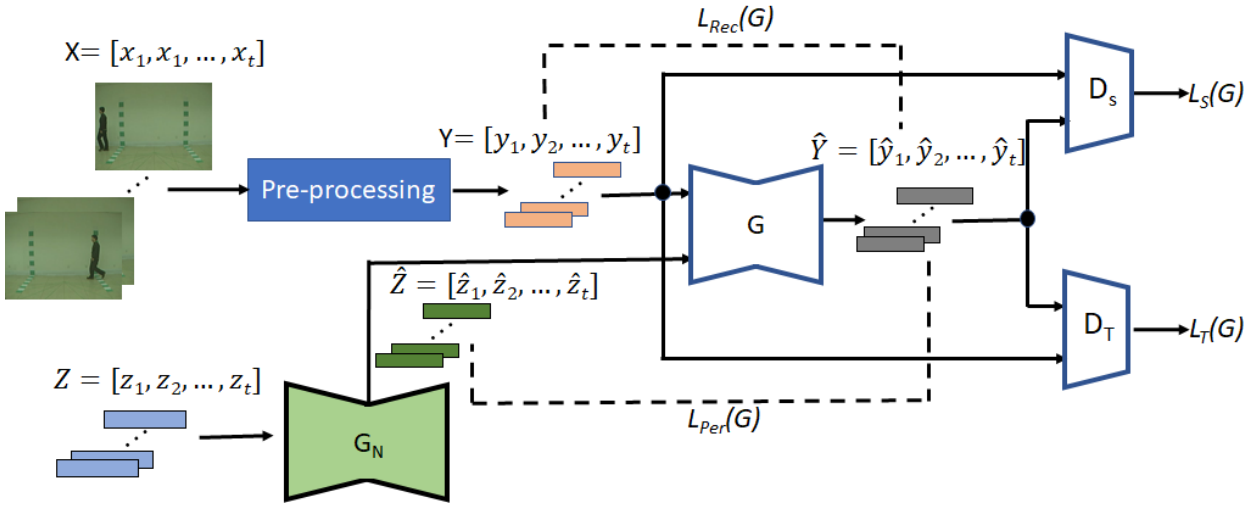
Definition 2 *(Contour sequence): the contour sequence of a gait is the sequence of contour vector of its frames.*

Definition 3 *(Noise contour): is a contour vector which is "added" to a contour vector of a gait for anonymizing that frame. The noise contour is generated by a Noise Generation Network.*

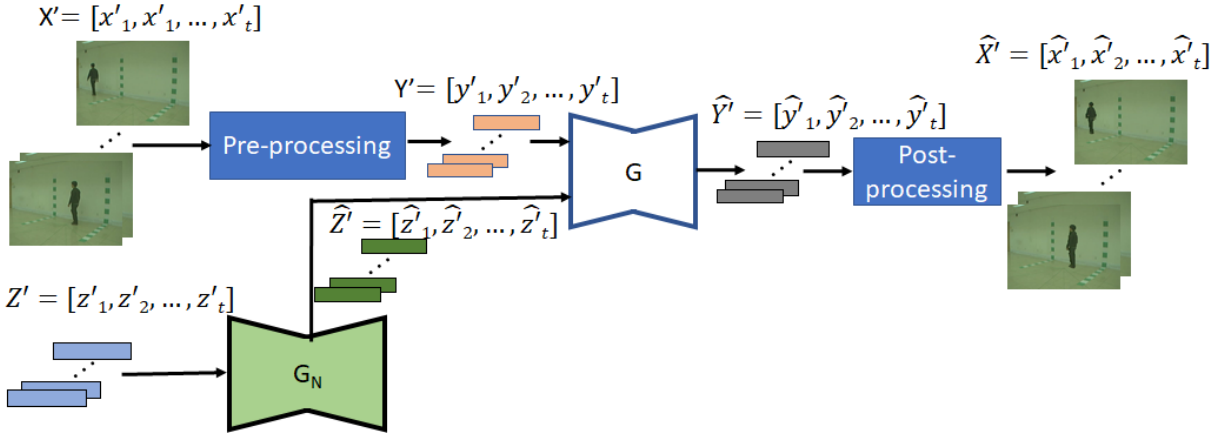
Definition 4 *(Random noise): is the sequence of noise contours, which will be "added" to the original gait to anonymize it.*

Definition 5 *(Random seed): is a normal distributed random vector that is used as the input of the Noise Generation Network for generating a noise contour.*

Table.4.1 presents the notations used in this chapter



(a) Training phase: Only generator G, spatial discriminator D_S, and temporal discriminator D_T are trained in this phase. Noise generator G_N is pre-trained.



(b) Generation phase.

Fig. 4.2 Overview of the training phase and the generation phase of the proposed model.

Notation	Meaning
x_i	the i^{th} frame of an original gait
X	the frame sequence of an original gait, $X = [x_1, x_2, \ldots, x_t]$
y_i	the contour vector of the i^{th} frame of an original gait
Y	the contour sequence of an original gait, $Y = [y_1, y_2, \ldots, y_t]$
z_i	the i^{th} *random seed*
Z	a sequence of noise seed, $Z = [z_1, z_2, \ldots, z_t]$
$\hat{y}_i$	the contour vector of the i^{th} frame of an anonymized gait
$\hat{Y}$	the contour sequence of the an anonymized gait, $\hat{Y} = [\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_t]$
$\hat{z}_i$	the noise contour for the i^{th} for anonymizing i^{th} frame of a gait
$\hat{Z}$	the noise contour sequence for a gait, $\hat{Z} = [\hat{z}_1, \hat{z}_2, \ldots, \hat{z}_t]$

Table 4.1 The notations used throughout the chapter.

4.3.2 Overview of the Proposed Method

The proposed model in this chapter consists of three steps as follows.

Step 1 (Pre-processing): Extracts the contour vectors of silhouettes of a gait. These vectors have the same length that is fixed to 4000, corresponding to 2000 pixels on the contour. If a contour has less than 2000 pixels, adding zero-padding at the end is necessary.

Step 2 (Contour vector modifying): Generates the modified contour vectors from contour vectors of the original gait and random noise.

Step 3 (Post-processing): Generates the anonymized gait from modified contour vectors. Then the original gait in the original video is replaced with the anonymized gait.

In our approach, we inherit Step 1 of our proposed method presented in Chapter3, but, the input of our Step 1 is the color videos instead of the binary ones as presented in chapter 3. For step 3, we also do the same way as presented in chapter 3. However, gaits produced by model ST-GAN are then colorized to output the final RGB anonymized gaits. Step 2 presents our two main contributions. One is to improve the success of anonymization, we use the *random noise* instead of the *noise gait* adopted in model BiGait-ANET. Another is to preserve more information of appearance and motion in anonymized gaits, we propose a novel model, ST-GAN, which contains one generator network and two discriminator networks, spatial discriminator and temporal discriminator.

Figure.4.2 shows the proposed model for the training phase and generation phase. The training phase is explained as follows. Firstly, contour vector Y of the original gait is

computed from the original gait X by the pre-processing to obtain. Secondly, the random noise $\hat{Z}$ is created from the random seeds Z by using the noise generator G_N. Finally, the modified contour vectors $\hat{Y}$ are generated with the gait generator G from the original contour sequence Y and the random noise $\hat{Z}$. Two discriminators D_S and D_T guarantee the quality of generated gait by learning the differences of the shape and the smoothness of the original and the generated gaits, respectively.

The process of the generation phase is as follows, the contour vector Y' of the original video X' and the random noise $\hat{Z}'$ made from the random seed Z' are fed into the generator G to produce the modified contour vector $\hat{Y}'$. This modified contour vector is then entered the post-processing to yield the anonymized gait.

4.3.3 Noise Generation

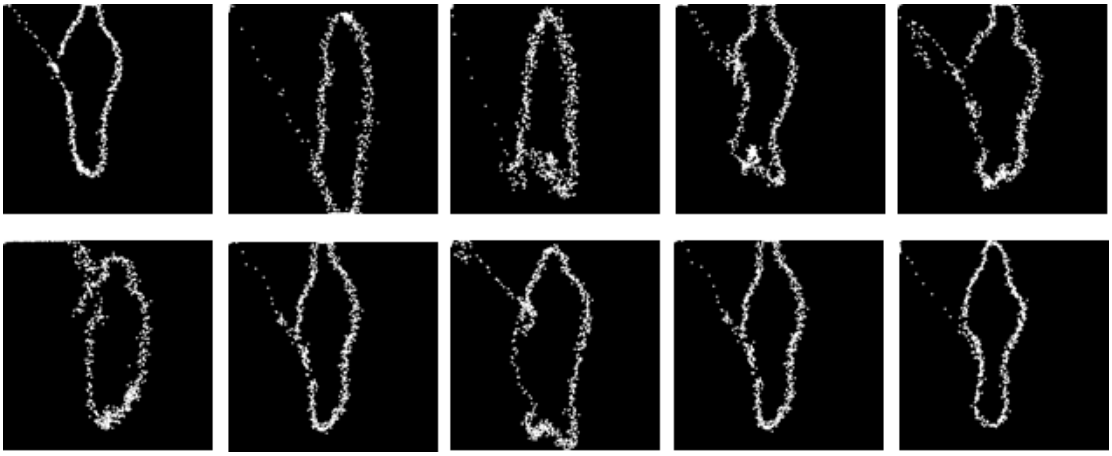


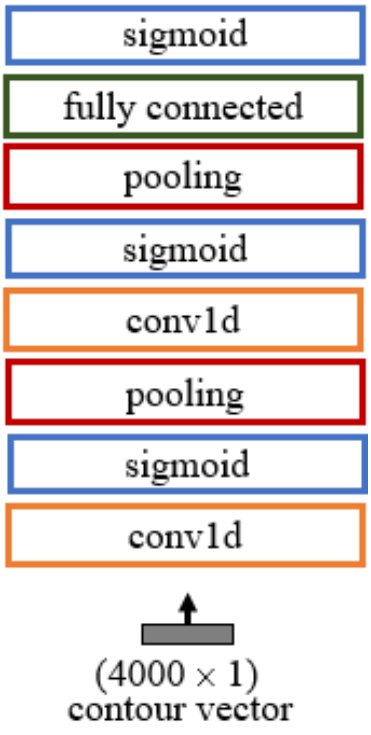
Fig. 4.3 Visulization of noises generated by noise generator G_N .

Using noise gait as in model BiGait-ANET may cause the reduction of anonymization success rate in the case that gait recognition systems recognize the noise gait as the original gait. Therefore, it is difficult to choose the optimal noise gait, which meets properties: not the original gait but in the same view with the original gait and similar length sequence to the original gait. In order to solve these problems, but still maintain the rule that the noise should have the gait shape, the traditional GAN model was adopted to create a noise in the gait distribution (the *random noise*) from normal distributed random seeds. Since our ST-GAN takes inputs as contour vectors of the gait, the random noise must be a vector. That means the original gaits are passed through the pre-processing step before using as the positive examples in the noise generation model. This noise generation network is trained and then

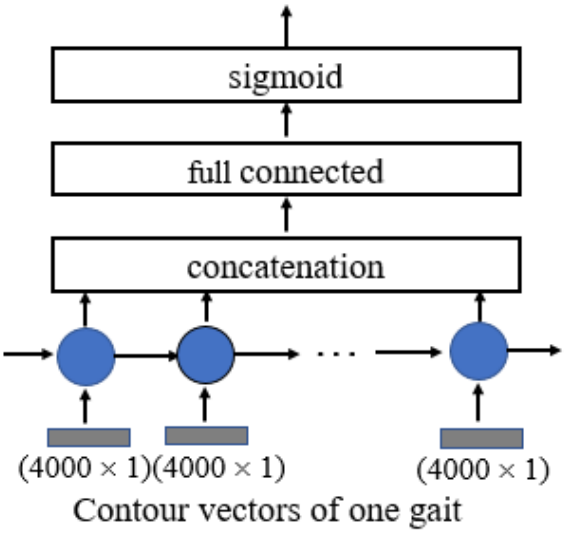
stacked to the main network, ST-GAN. For simplicity, in this research, only the first noise contour z_1 is produced from a random seed. The remaining noise contours $z_2, z_3, \ldots, z_t$ are copied from z_1. Fig.4.3 shows the visualization of some noises generated by G_N.

4.3.4 Anonymization Network

In order to obtain the high quality of the naturalness of the anonymized gaits, two discriminators are sticked to the gait generator: spatial discriminator network D_S and temporal discriminator network D_T. Fig.4.4 illustrates the spatial discriminator and temporal discriminator, respectively.



(a) architecture of the spatial discriminator.



(b) architecture of the temporal discriminator.

Fig. 4.4 The architecture of the discriminators.

The goal of temporal discriminator network is ensuring the motion smoothness of a generated gait by distinguishing the temporal information of a real gait and that of a fake gait. To do so, a LSTM network is used and this network takes a gait sequence as its input. The results at each node are connected into one vector and then passed into a sigmoid function.

Therefore, This flow ensures that the temporal discriminator can assess the naturalness of movement of the whole sequence.

The goal of the spatial discriminator network is to discriminate between a real and fake gait images. In other words, this network distinguishes the shape of a real gait and that of fake one at each frame, therefore, this discriminator is used to increase the naturalness of an anonymized gait in the terms of body shape. Figure.4.4a illustrates this network architecture that consists of convolution layers, fully connected layers, and a sigmoid function.

With the gait anonymization generator, there are multiple options to design the structure of the generator. The generator G of model ST-GAN has the same architecture as the modified contour generation presented in Chapter 3, which is based on an autoencoder network. In network of generator G, the encoder takes contour vectors of the original gait and random noise generated by noise generator G_N as two inputs. These two inputs are then merged by the sum operator before being passed to the decoder.

The whole model ST-GAN is trained as the GAN model, that means G, D_S, and D_T are trained to solve the min-max problem with the objective function as follows:

$$\begin{aligned}
&\min\max L(G,D_S,D_T) = E_{y\sim p_y(y)}[logD_S(Y)] \\
&+E_{y\sim p_y(y),z\sim p_z(z)}[log(1-D_S(G(Y,G_N(Z))))] \\
&+E_{y\sim p_y(y)}[logD_T(Y)] \\
&+E_{y\sim p_y(y),z\sim p_z(z)}[log(1-D_T(G(Y,G_N(Z))))]
\end{aligned} \tag{4.1}$$

Practically, to solving (4.1) the two discriminators D_S, D_T, and the gait generator G are trained alternatively. For more detail, firstly, the generator network is fixed, and two discriminators are trained by a maximum of the two below loss functions:

$$\begin{aligned}
&L(D_S) = E_{y\sim p_y(y)}[logD_S(Y)] \\
&+E_{y\sim p_y(y),z\sim p_z(z)}[log(1-D_S(G(Y,G_N(Z))))]
\end{aligned} \tag{4.2}$$

$$\begin{aligned}
&L(D_T) = E_{y\sim p_y(y)}[logD_T(Y)] \\
&+E_{y\sim p_y(y),z\sim p_z(z)}[log(1-D_T(G(Y,G_N(Z))))]
\end{aligned} \tag{4.3}$$

Then, two discriminator networks are untrained, and the generator is trained to minimize the loss function:

$$L_S(G) = E_{y\sim p_y(y),z\sim p_z(z)}[log(1 - D_S(G(Y, G_N(Z))))] \tag{4.4}$$

$$L_T(G) = E_{y\sim p_y(y),z\sim p_z(z)}[log(1 - D_T(G(Y, G_N(Z))))] \tag{4.5}$$

The reconstruction loss also is minimized by the generator G to preserve the viewing angle and action (here is walking) of the gait that to be anonymized and l_1 loss function is employed for this purpose:

$$L_{Rec}(G) = E_{y\sim p_y(y),z\sim p_z(z)}[\| Y - G(Y, G_N(Z)) \|_1] \tag{4.6}$$

Finally, a perturbation loss is used in order to force the generated gait is partly close to the random noise. This loss ensures that the generated gait can fool gait recognition systems.

$$L_{Per}(G) = E_{y\sim p_y(y),z\sim p_z(z)}[\| Z - G(Y, G_N(Z)) \|_1] \tag{4.7}$$

The gait generator is trained to minimize the total loss function of (4.4), (4.5), (4.6), (4.7):

$$L(G) = L_S(G) + L_T(G) + L_{Rec}(G) + \alpha * L_{Per}(G) \tag{4.8}$$

Here α is an adjustable parameter to balance the trade-off between the naturalness and success rate of the proposed model.

4.3.5 Colorization Algorithm

There have been some researches for synthesizing the colorful objects from the original objects [76], [82], [75]. These methods aim to generate static images, therefore, they cannot be applied to our research, whose input is a video. This is because there are not any parts to control consistency among frames, or the relationship of the scenes between the consecutive frames, in these methods. In this section, an algorithm that transfers the colors of the original gait image to the binary anonymized gait image is introduced. For being simple, we assumed

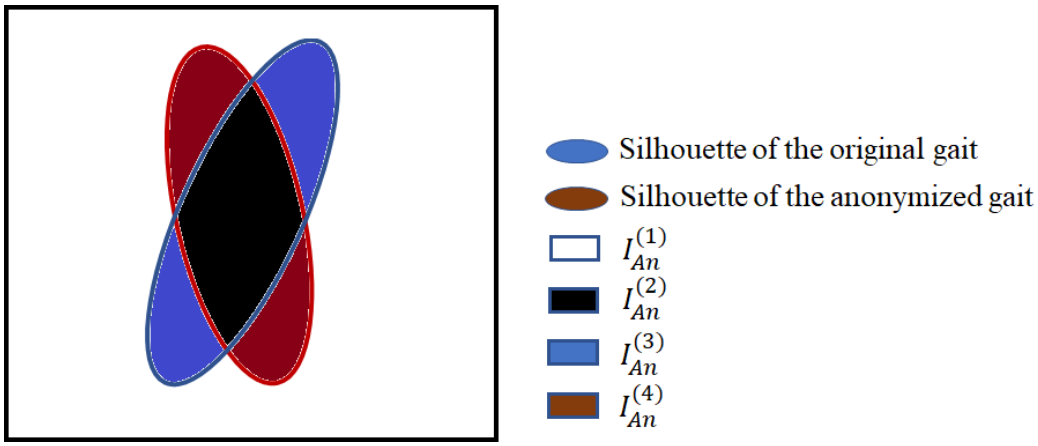


Fig. 4.5 Colorization.

that the background in the raw video of a walking person is static. The notations for colorizing algorithm are as follows. I_{Bg} is denoted the background image, I_{An} is denoted the $t-th$ frame of the final gait, and I_{Or} is the $t-th$ frame of the gait we wish to anonymize. S_{Or} and S_{An} are denoted as the silhouettes of these frames, respectively. Let (i, j) be a pixel coordinate. The goal of the colorizing algorithm now is given I_{Bg}, I_{Or}, S_{Or}, S_{An}, this algorithm yields I_{An}.

To this end, we divide I_{An} in to four regions as shown in Fig. 4.5: The first one, $I_{An}^{(1)}$, belongs to the background region. $I_{An}^{(1)} : \{(i, j) \notin S_{Or} \cup S_{An}\}$; The second one, $I_{An}^{(2)}$, is the overlapping region between original gait silhouette and the anonymized gait silhouette $I_{An}^{(2)} : \{(i, j) \in S_{Or} \cap S_{An}\}$; The third one, $I_{An}^{(3)}$, is the region in the original gait silhouette, but, not in the anonymized gait silhouette $I_{An}^{(3)} : \{(i, j) \in S_{Or} \backslash I_{An}^{(2)}\}$; The fourth one, $I_{An}^{(4)}$, is the region in the anonymized gait silhouette, but not in the original gait silhouette $I_{An}^{(4)} : \{(i, j) \in S_{An} \backslash I_{An}^{(2)}\}$. The colorized anonymized gait is achieved as follows.

$$I_{An}(i, j) = \begin{cases} I_{Bg}(i, j), & \text{if } (i, j) \in I_{An}^{(1)} \\ I_{Or}(i, j), & \text{if } (i, j) \in I_{An}^{(2)} \\ I_{Bg}(i, j), & \text{if } (i, j) \in I_{An}^{(3)} \\ I_{Or}(i', j'), & \text{if } (i, j) \in I_{An}^{(4)} \end{cases} \tag{4.9}$$

where (i', j') is the nearest pixel to the (i, j).

Note that the colorizing algorithm here applies for each frame but the original frame is used as the reference, therefore, the relationship of the scenes between the consecutive frames is maintained.

4.4 Experiment Results

The dataset of 124 subjects is divided into four non-overlapping parts. The first part, whose the number of subjects is 50, is to train three gait identification systems. The second part has 10 subjects and is kept to train G_N network. The third part containing 24 subjects is to train the ST-GAN. The remaining part with 40 subjects is preserved for testing our model. We also show the performance comparison of this model and the model BiGait-ANET (baseline) presented in Chapter 3. To this end, two kinds of evaluation metrics are conducted: (1) naturalness that measures the degree of appearance and motion preservation, (2) success rate. The proposed model was run with several hyperparameters α and it was found that the trade-off between the success rate and naturalness can be controlled well with $\alpha = 0.3$. Therefore, this hyperparameter was used in comparing with the baseline, the model BiGait-ANET, which is detailed in sections 4.1, 4.2, and 4.3. The discussion of the impact of hyperparameter α on the proposed model is also presented in section 4.4.

4.4.1 Generation Results

This section shows the visualization of anonymized gaits.

(1) Some of the generated samples by model ST-GAN and the model BiGait-ANET are showed in Fig. 4.6 and Fig. 4.7. Both figures demonstrate that the anonymized gaits rendered by model BiGait-ANET look less natural because the head of a generated gait is distorted, especially, at the view angle 0^0 and 180^0, and this problem is solved well by using model ST-GAN.

(2) The colorized frames showed in Fig.4.7 and Fig.4.9 demonstrate that the color of consecutive frames is consistent.

(3) Gaits synthesized the gaits by ST-GAN, but, using directly the random vectors generated from the normal distribution are also shown in Fig. 4.8. In this case the G_N was not assigned to the ST-GAN . The result is seen in Fig. 4.8 presents that removing G_N makes the generated gaits less natural, and therefore, G_N plays an important role to generate more natural gaits compared with the model BiGait-ANET.

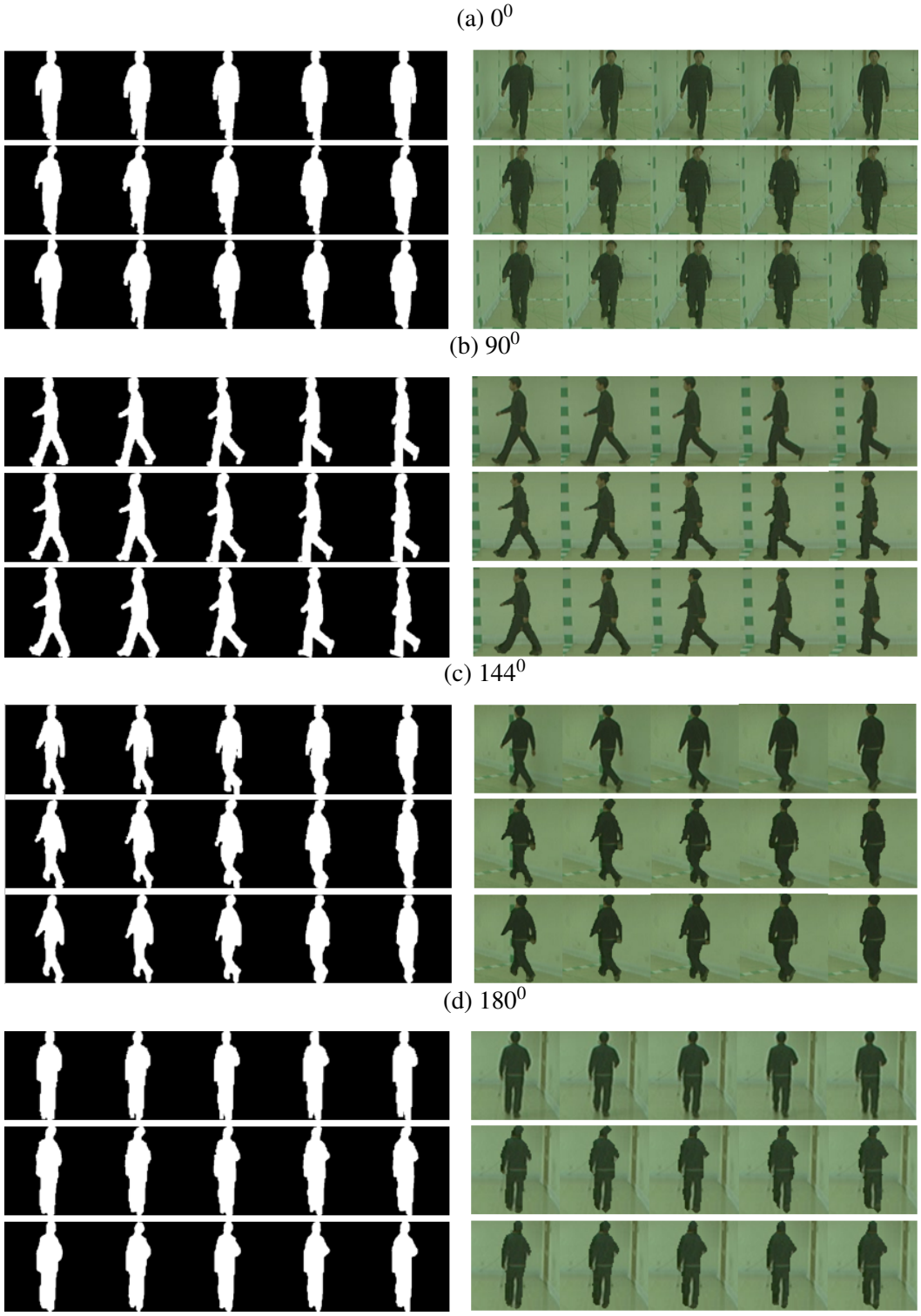


Fig. 4.6 Original and anonymized gait generated by the model ST-GAN and the model BiGait-ANET under various view angles: original gaits are in the first rows, anonymized gaits generated by the model BiGait-ANET are in the second rows, and the anonymized gaits generated by the model ST-GAN are in the third rows.

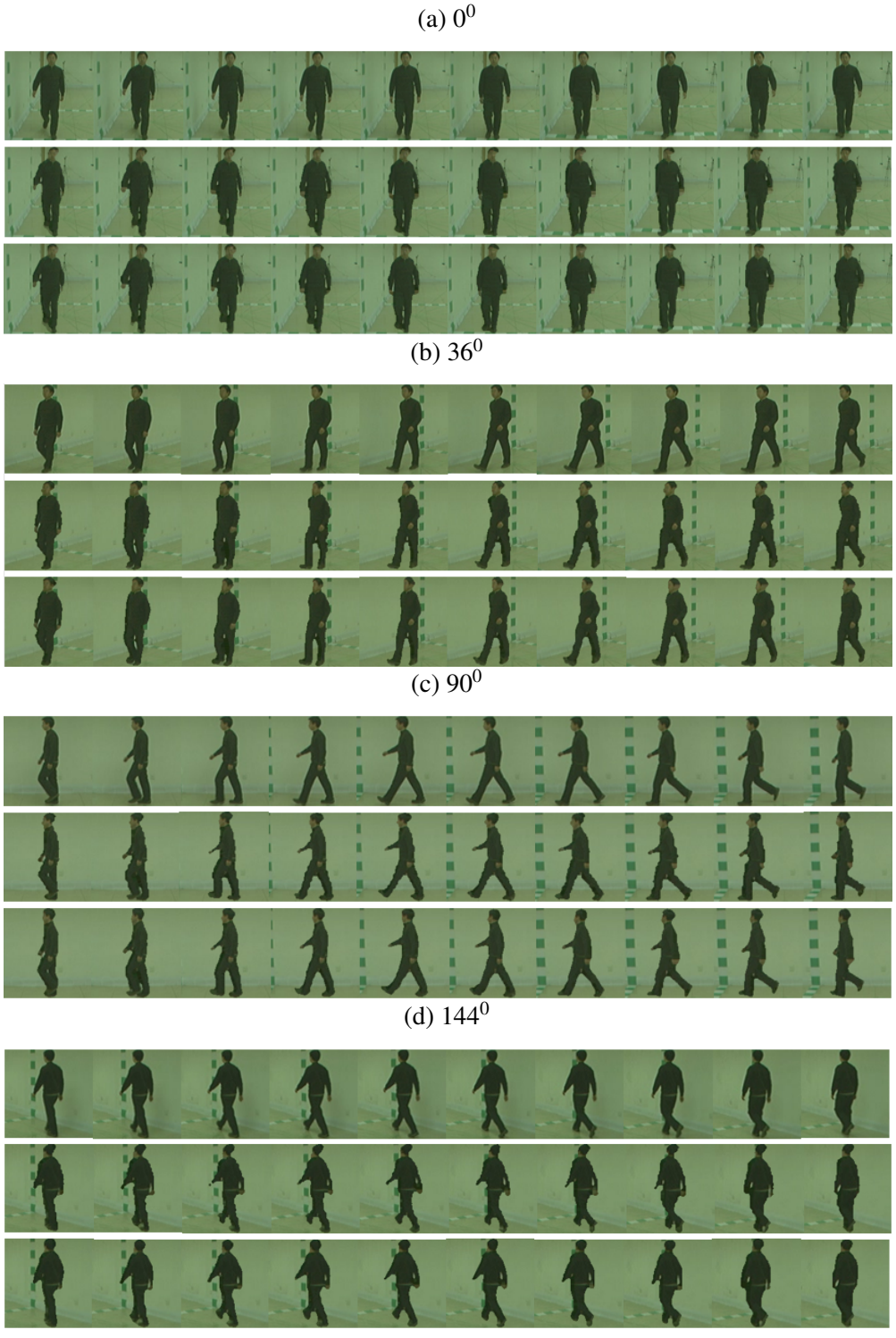
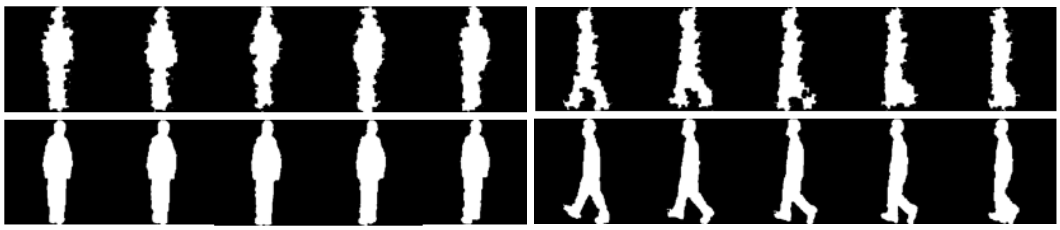


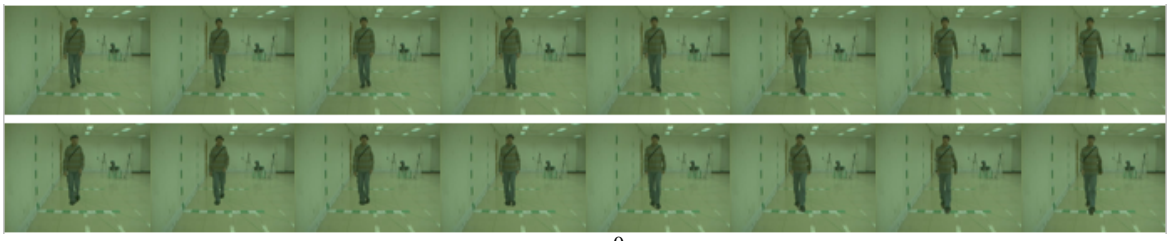
Fig. 4.7 Original and anonymized gait generated by the model ST-GAN and the model BiGait-ANET under various view angles: the first rows show original gaits, second rows show anonymized gaits generated by the model BiGait-ANET and the third rows show the anonymized gaits generated by the model ST-GAN.



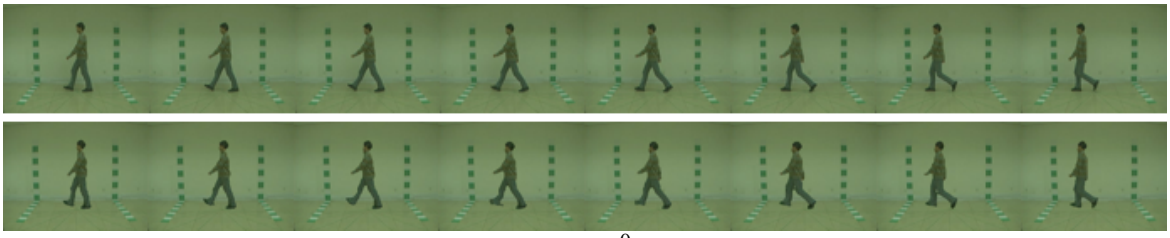
(a) 0^0 (b) 90^0

Fig. 4.8 Silhouette of the anonymized gaits generated from the random seed in the normal distribution (the first rows) and from the random noise generated with noise generation (the second rows).

(a) 0^0



(b) 90^0



(c) 126^0

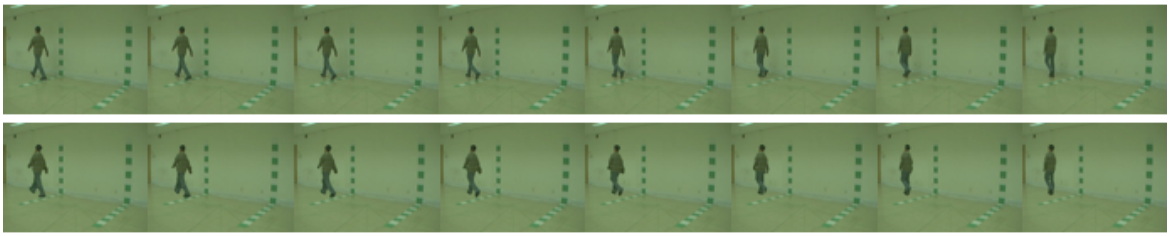


Fig. 4.9 Original and the final anonymized gait video in which the gaits are under various view angles: the first rows show original gaits, second rows show anonymized gaits generated by the model ST-GAN.

4.4.2 Naturalness

We measure the naturalness of the binary anonymized gaits, which are generated by ST-GAN model because the binary gaits influence the quality of the color gait. We also

measure the naturalness of the color anonymized gaits, which are obtained by colorizing the binary anonymized gaits.

For the binary anonymized gaits, we asked 40 volunteers (none of them are authors) with different backgrounds to distinguish between anonymized gait generated by each model and real gait. Haft of them tested for the baseline model and others tested for our proposed model. We gave each volunteer 60 random videos, in which half is from the real set and half is from the anonymized set. They watched each video and answered whether it is real or generated. Fig.4.10 shows the percentage of correct classification by the human of each model. From this figure, we can withdraw three conclusions.

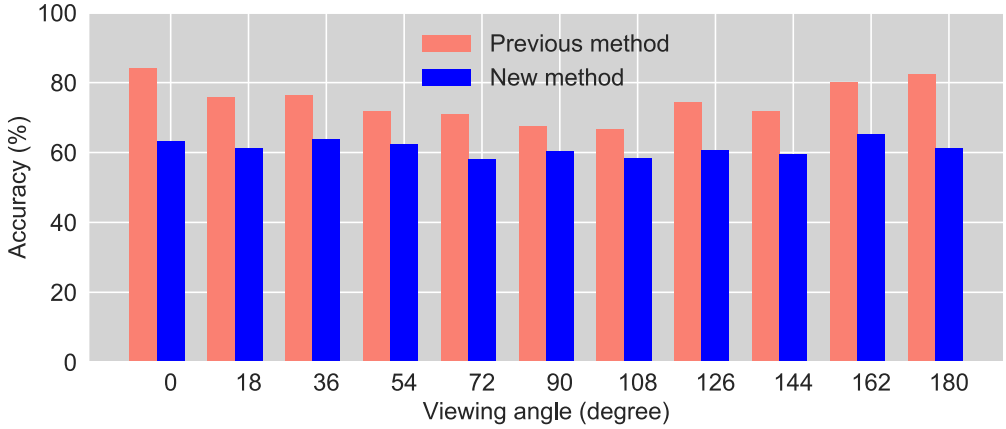


Fig. 4.10 Classification by human.

(1) The anonymized gaits synthesized by model ST-GAN are more natural than those generated by model BiGait-ANET. (2) The naturalness of the anonymized gaits rendered by the proposed model is approximately equal to that of the real gait because the correct classification is from 58% to 65% depend on each view angle. These demonstrate that ST-GAN is able to generate natural anonymized gaits.

(3) The naturalness of the proposed method is quite equal for all views. That means the distortion of the generated gaits in front views of the baseline is solved well by ST-GAN.

The colorizing method uses the information of the nearest pixel. This method, sometimes, is not successful, especially, for coloring the face. Therefore, we did not perform the same test with the binary anonymized gaits. Instead, we conducted the MOS (mean opinion score) test. We ask 20 people among 40 volunteers above. For each volunteer, we picked up randomly 60 pairs of videos of color anonymized gait and corresponding original gait. Haft anonymized gait video is generated by ST-GAN model and the haft is from the BiGait-ANET

model. After watching each pair, they gave a score for the naturalness of the corresponding anonymized video in a five-point scale score. MOS scores are shown in Fig. 4.11. This result demonstrates two conclusions:

(1) The anonymized gaits produced by the model ST-GAN look more realistic than those produced by model BiGait-ANET.

(2) The scores of the degree from 54 to 180 trend to higher than those of the degree from 0 to 36. That is because the colorized face is not natural.

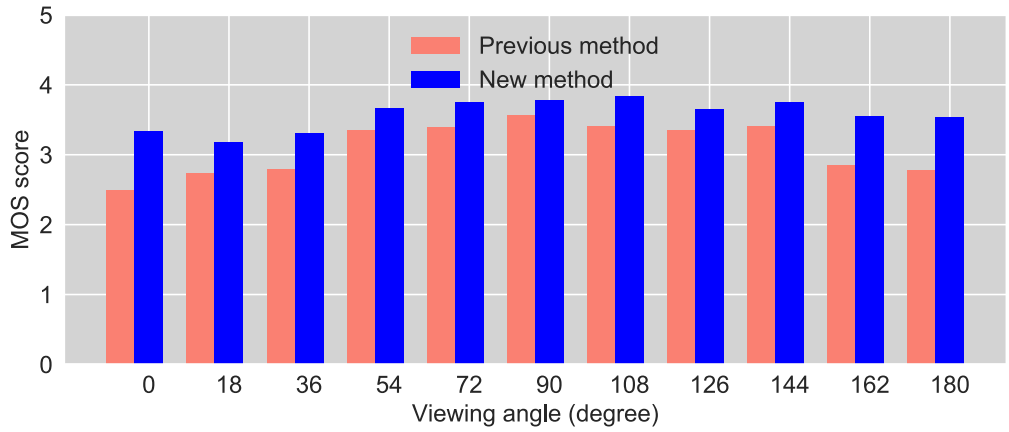


Fig. 4.11 Mean Opinion Score Result.

4.4.3 Success Rate

The success rates of the proposed model ST-GAN measured for three gait recognition systems, by Zheng et al [83], by Wu et al. [13], and by Chao et al. [14] with top-1 and top-3 identification are shown in Table 4.2, Table 4.3, and Table 4.4. The comparison of success rate with three gait recognition systems is summarized in Fig.4.12.

Table 4.2 The average success rate (%) of the proposed model ST-GAN with Zheng's [12] method.

Top	**Viewing angles**										
	$\mathbf{0^0}$	$\mathbf{18^0}$	$\mathbf{36^0}$	$\mathbf{54^0}$	$\mathbf{72^0}$	$\mathbf{90^0}$	$\mathbf{108^0}$	$\mathbf{126^0}$	$\mathbf{144^0}$	$\mathbf{162^0}$	$\mathbf{180^0}$
top – 1	84.77	72.36	76.44	70.33	74.84	71.37	78.07	72.45	71.35	71.54	73.16
top – 3	77.93	65.18	63.34	57.89	64.14	63.49	66.71	65.01	65.14	64.84	65.57

Table 4.3 The average success rate (%) of the proposed model ST-GAN evaluated with Wu's [13] method.

Top	**Viewing angles**										
	$\mathbf{0}^0$	$\mathbf{18}^0$	$\mathbf{36}^0$	$\mathbf{54}^0$	$\mathbf{72}^0$	$\mathbf{90}^0$	$\mathbf{108}^0$	$\mathbf{126}^0$	$\mathbf{144}^0$	$\mathbf{162}^0$	$\mathbf{180}^0$
top – 1	69.86	65.93	66.87	65.63	62.61	68.16	67.21	60.84	64.33	65.55	71.07
top – 3	66.36	61.63	55.72	56.15	57.78	60.45	55.42	51.78	57.73	60.80	64.29

Table 4.4 The average success rate (%) of the proposed model ST-GAN evaluated with Chao's [14] method.

Top	**Viewing angles**										
	$\mathbf{0}^0$	$\mathbf{18}^0$	$\mathbf{36}^0$	$\mathbf{54}^0$	$\mathbf{72}^0$	$\mathbf{90}^0$	$\mathbf{108}^0$	$\mathbf{126}^0$	$\mathbf{144}^0$	$\mathbf{162}^0$	$\mathbf{180}^0$
top – 1	72.59	72.13	71.48	72.11	68.27	64.36	65.65	68.25	70.56	67.03	68.57
top – 3	70.93	65.66	66.40	66.87	66.06	62.15	60.46	62.58	66.69	65.75	66.03

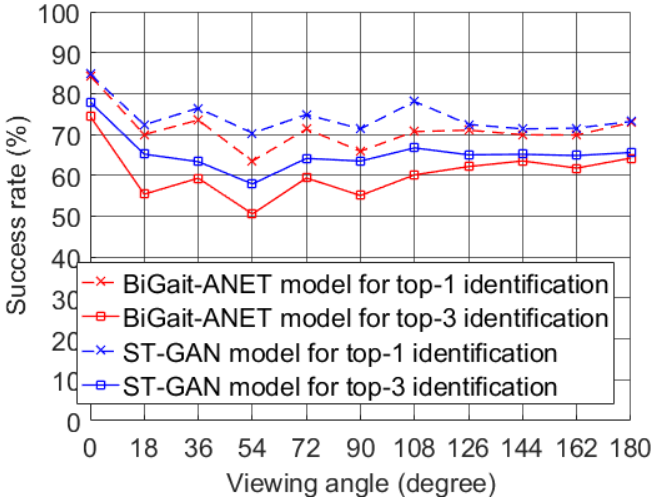
The Fig.4.12a show the success rate of the anonymized gaits generated by each model computed by Zheng's method with top-1 and top-3, respectively, Fig. 4.12b show those computed by Wu method with top-1 and top-3, respectively, and Fig. 4.12c show those computed by Chao method with top-1 and top-3, respectively. These figures illustrate that:

(1) The success rate of the model ST-GAN at $\alpha = 0.3$ is higher than that of model BiGait-ANET at all viewing angles with three gait recognition systems. This illustrates that removing the identity features of one gait by a random noise is better than by a noise gait that is used in the model BiGait-ANET.

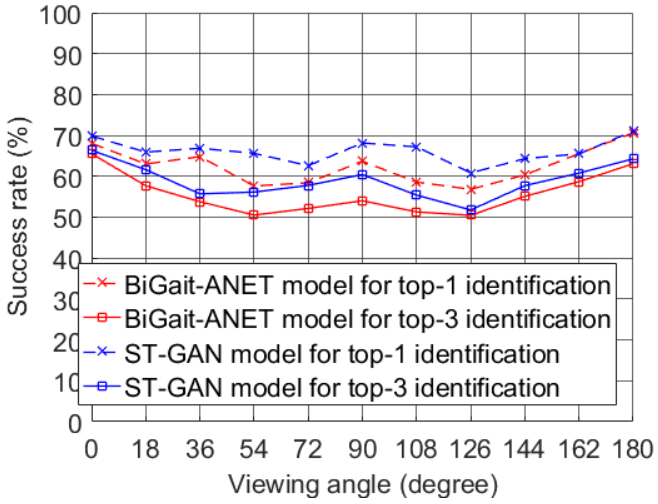
(2) The differences between the values of success rate of the two models decreases with respect to the frontal and back viewing angles because the anonymized gaits synthesized by the model BiGait-Net are distorted at these views.

(3) The success rate evaluated by Zheng's method is better than that evaluated by Wu's method and Chao's method. The reason is that Wu's method is more robust than Zheng's method.

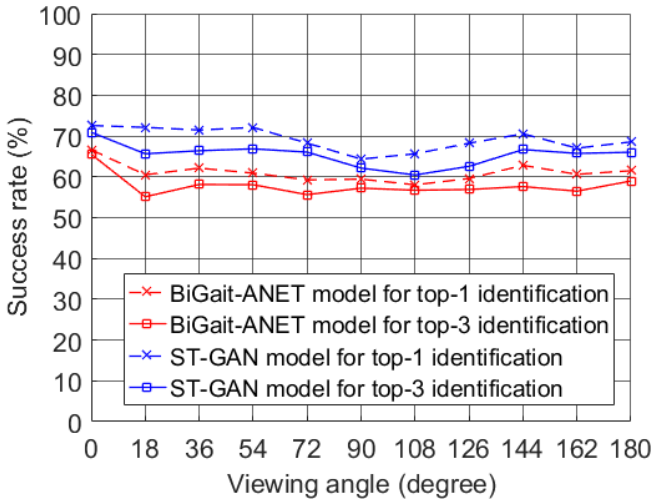
(4) The success rate with Chao's method is not much different between the model BiGait-ANET and the model ST-GAN because Chao's method uses a silhouette sequence as the input that captures more temporal information than GEI, which is used in gait recognition systems proposed by Zheng and Wu.



(a) Zheng's method



(b) Wu's method

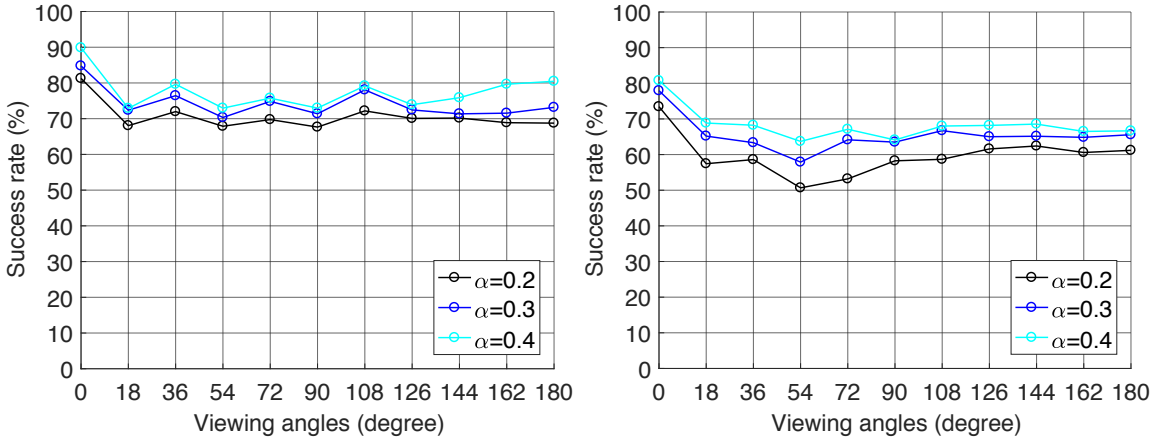


(c) Chao's method

Fig. 4.12 Success rate comparison of the model ST-GAN and the model BiGait-ANET.

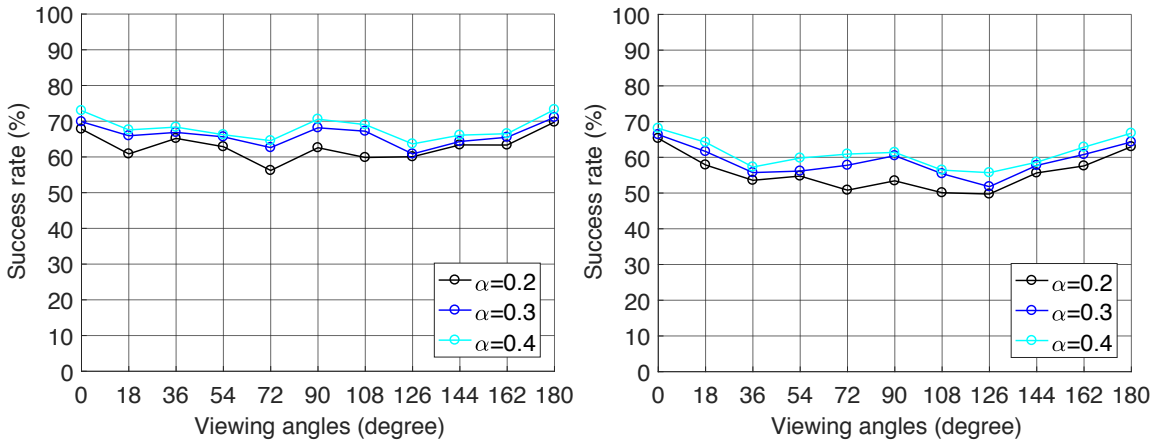
4.4.4 Impact of α

In order to analyze the role of parameter α on our model, we compute generate the anonymized gaits with α ranging from 0.2 to 0.4, and then we also evaluate the success rate corresponding to those α. Fig.4.13 and Fig.4.14 illustrate the impact of α on success rate with respect to Zheng's method and Wu's method, respectively. These this figure demonstrates that the success rate increase when α increase. Fig.4.15 visualizes the anonymized gaits generated with several α and we can realize that the anonymized gaits look less natural when α increases. This means the success rate is inversely proportional to naturalness in our model and parameter α is the factor to control this trade-off.



(a) Success rate at top-1 (b) Success rate at top-3

Fig. 4.13 Impact of α on success rate with Zheng's method.



(a) Success rate at top-1 (b) Success rate at top-3

Fig. 4.14 Impact of α on success rate with Wu's method.

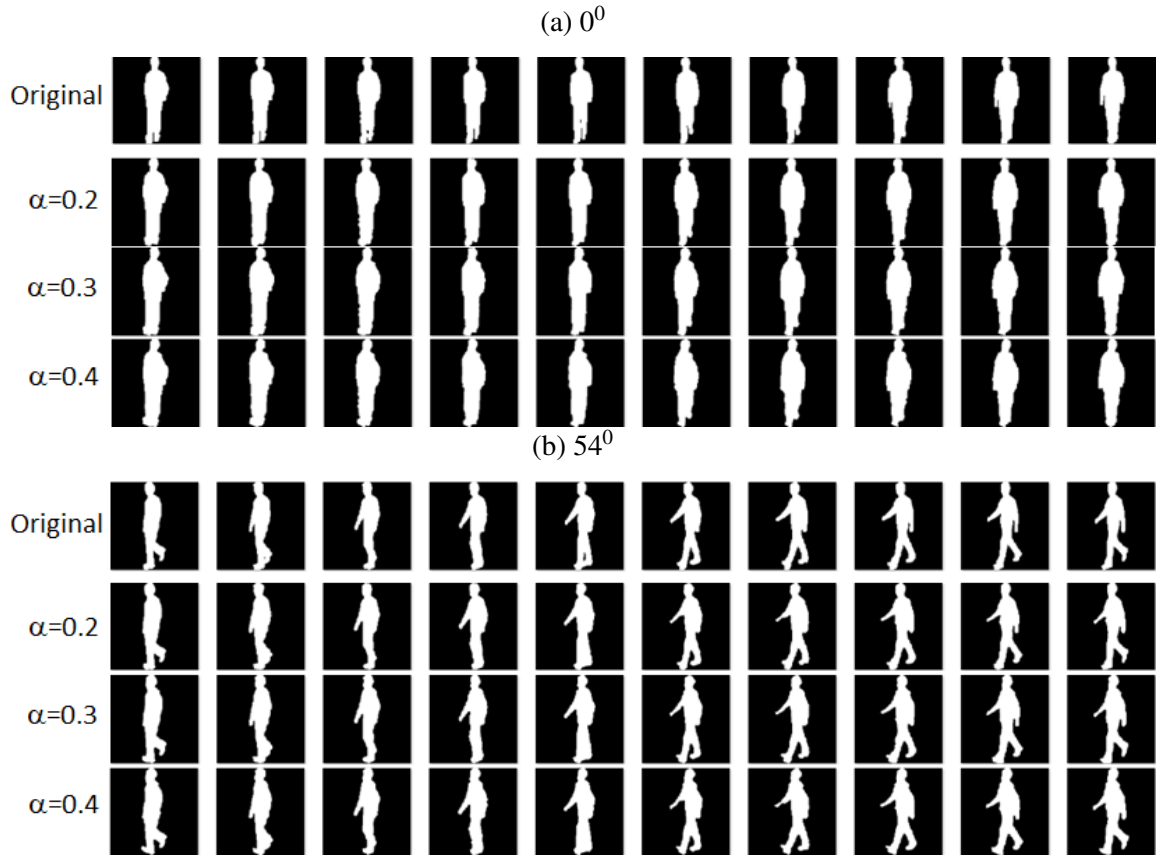


Fig. 4.15 Anonymized gaits generated with various α.

4.4.5 Robustness against Re-identification Attack

Table 4.5 The average identification accuracy (%) of the re-identified gaits evaluated with Zheng's method, Wu's method, and Chao's method.

Top	**Viewing angles**										
	$\mathbf{0^0}$	$\mathbf{18^0}$	$\mathbf{36^0}$	$\mathbf{54^0}$	$\mathbf{72^0}$	$\mathbf{90^0}$	$\mathbf{108^0}$	$\mathbf{126^0}$	$\mathbf{144^0}$	$\mathbf{162^0}$	$\mathbf{180^0}$
top – 1	5.05	3.81	4.70	4.45	3.91	3.78	4.50	4.10	3.90	3.57	3.65
top – 3	7.24	7.35	8.32	7.90	7.58	6.92	7.61	7.15	6.82	6.50	7.55

We also investigate whether we can find a machine learning model for re-identification attack. Since model-free gait recognition systems take silhouettes that are binary images as inputs, if we stack a gait recognition system on the top of the re-identification model to force this model to restore the identity of the original gait from the anonymized gait, the output of the re-identification model must be binarized. However, the function that converts the output of the re-identification model to the binary images is a discrete function that has no gradient, and therefore the gradient of the whole model cannot be updated.

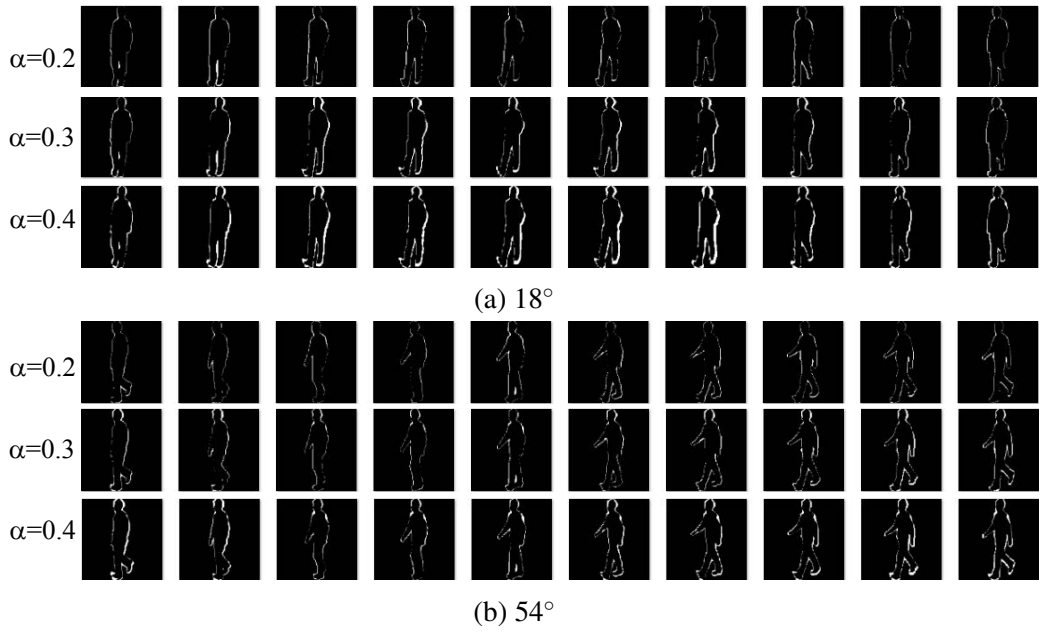


Fig. 4.16 XOR images for three values of α.

We tried using a traditional denoising autoencoder network [68] to restore the original identity given an anonymized gait. We used 20 subjects in 40 anonymized subjects to train this network and used the remaining 20 anonymized subjects to evaluate this network. In order to explore the robustness of the proposed gait anonymization model presented in this chapter, we compute the identification accuracy of three gait recognition systems on the re-identified gaits. Table 4.5 shows the this accuracy and a sample of re-identified result is shown in Fig. 4.17. Both Table 4.5 and Fig. 4.17 demonstrates that our model is robust to the re-identification attack.

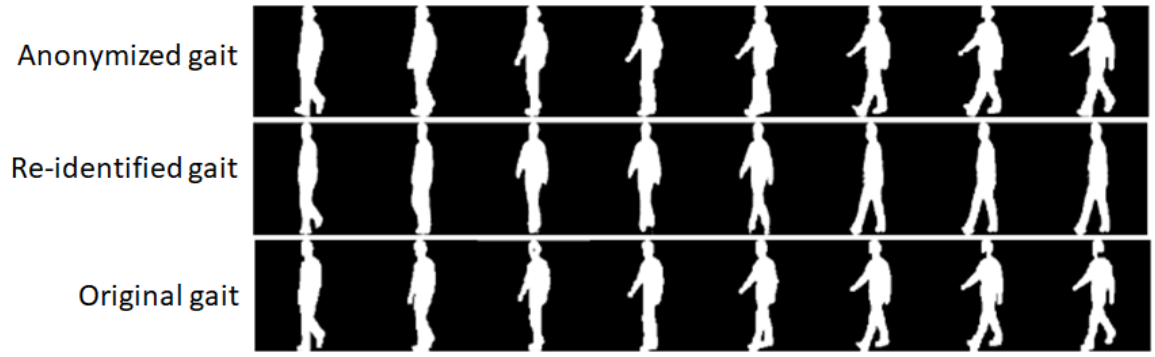


Fig. 4.17 The generation results of re-identification attack of model ST-GAN using traditional denoising autoencoder.

4.5 Summary

In this chapter, model ST-GAN has been introduced to tackle the issues that have not been solved by the model BiGait-ANET. Three issues were considered in this research: (1) How to improve anonymization performance, (2) How to ensure consistency among frames, (3) How to transfer colors in original gait images to anonymized gait images. To solve these issues, the model ST-GAN removes the identity of a gait by combining that gait with a random noise generated in the gait distribution and uses two discriminators to ensure the naturalness of anonymized gait in the terms of body shape and smoothness of movement. A colorization algorithm is also proposed to transfer colors of RGB original gait images to binary anonymized gait images. The experiments in this research demonstrated that the model ST-GAN can solve the above three issues.

There are two limitations to model ST-GAN. First, the model ST-GAN synthesizes anonymized gaits from the silhouette of original gaits, therefore, the quality of anonymized gaits is affected by the silhouette extraction process that may extract the low-quality silhouettes. Second, the colorizing algorithm fills the color of one pixel based on the colors of the nearest pixels, therefore, colors in the missing parts of original gait images that are caused by the silhouette extraction process cannot be transferred to anonymized gait. In the next chapter, these two limitations will be considered and completely solved.

Chapter 5

Incomplete Silhouette Gait Anonymization

5.1 Introduction

One of the important phases in gait recognition systems is subject extraction. In order to obtain a gait of the subject in the video, the foreground extraction process is applied. This process may cause the missing of some body parts of the extracted gait, this is named incomplete silhouette in this thesis. Because the incomplete silhouettes may occur at some frames, the gait recognition systems still use the complete silhouettes to recognize the subjects in a video. In addition to this, some research can use some parts of the human body to recognize subjects in videos. For instance, Gabriel-Sanz et al. [84] used lower parts of the human body as the feature for gait recognition, Similarly et al. adopted the head and feet of the gait for recognition, meanwhile, Rida et al. [85] used top and bottom parts as the gait feature The model ST-GAN is able to generate the natural anonymized gait from original complete silhouette gaits. However, anonymized gait looks unnatural if the input of this model is a gait of incomplete silhouettes. This is because the architecture of the gait generator of ST-GAN is based on an autoencoder network that tries to produce output so that it is close to the network input. The model is able to modify the shape of a gait but unable to synthesis the missing parts that occur in an incomplete silhouette. The goal of this chapter is to consider and solving the following research question that has not been addressed by models in Chapter 3 and Chapter 4: (1) How to anonymize incomplete silhouette gait?; (2)

How to generate seamless gait from incomplete silhouette gait?; (3) How to transfer colors for missing parts?.

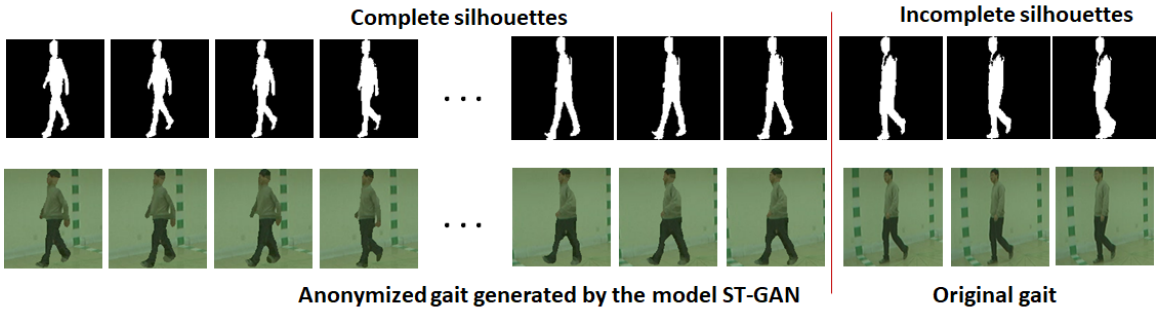


Fig. 5.1 The consistency among frames does not guarantee when only complete silhouettes are anonymized.

This chapter, thus, focus on anonymizing incomplete silhouette gaits while preserving the gait's naturalness with two tasks (anonymization and colorization) are solved as displayed in Fig.5.2

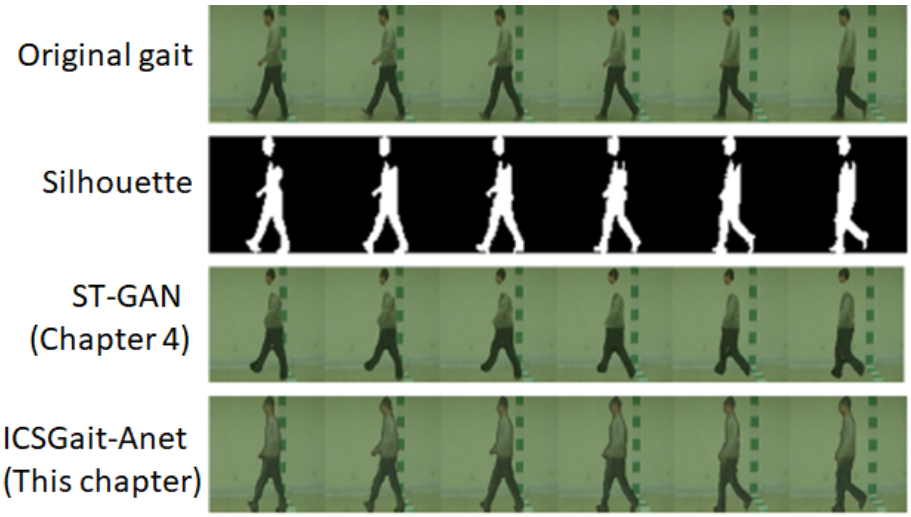


Fig. 5.2 This chapter focuses on anonymizing incomplete silhouette gait and solves both tasks anonymization and colorization.

While there have been various reports of GAN-based models for high-quality image generation, there have been only a few models for video synthesis. Vondrick et al.[80] introduced a model VGAN to generate a video from a random vector. The model consists of two generators, one is to produce background and the other is to generate foreground. Holden et al. [86] proposed a model to create a new motion of the human skeleton. Firstly,

they represented the human motion in the hidden layer by developing a network based on an autoencoder. The decoder of this network was trained to generate natural motion, then they assigned another network to the decoder of that autoencoder network to yield the new sequence of motion. To give more relevant research, Yan et al. [81] combines a Siamese network with a GAN model to generate a motion video of the human body from a sequence of a body skeleton and a RGB body image.

Deep neural networks have been successful in many research areas [87, 88], especially in transferring the garments in a reference image to a target image. Lassner al.[89] 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. [90] developed a method for overlaying the clothing in a product image on a person in a query image while Neuberger et al. [91] 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.

The proposed gait anonymization model aims to generate seamless anonymized gaits from incomplete silhouette gaits, which has not been completely addressed by model BiGait-ANET as well as model ST-GAN. The model introduced in this chapter uses two independent networks. The first one, named anonymization network, is to erase the original identity of a gait that we wish to anonymize, meanwhile the second one, named the colorization network, is to transfer colors in original gait images to anonymized gait images generated by the first network. The anonymization network directly adds random noise to a binary silhouette sequence of an original gait to erase the identity of that gait and produces a silhouette sequence of the anonymized gait. Anonymizing binary silhouettes rather than color ones presses the network to remove the original identity and not care about changing colors in the gait images. In order to ensure that the network can generate seamless silhouette gaits regardless of the quality of the original silhouette, the anonymization network is based on the model DCGAN and trained with the complete silhouette. The colorization network has two inputs, one is the output of the anonymization network and one is the RGB original gait images.

In order to evaluate the performance of our method, we use the dataset CASIA-B [67] with two metrics as presented in Chapter2. Three gait recognition systems were used [12–14] to measure the success rate, while subjective evaluation was used to assess the naturalness of the anonymized gait.

5.2 Methodology

The proposed model ICSGait-ANET includes two networks, anonymization network and colorization network as shown in Fig. 5.3. The first one, the anonymization network (A-NET), is to remove the identity of the original gait, and the second one, the colorization network (C-NET), is to transfer colors in original gait images to the anonymized gait images generated by the first network. The anonymization network adds a random noise vector R to its input, which is the sequence of binary silhouettes X of the gait we wish to protect to hide the identity of this gait. This network then outputs the sequence of silhouette Z of the anonymized gait.

5.2.1 Model Overview

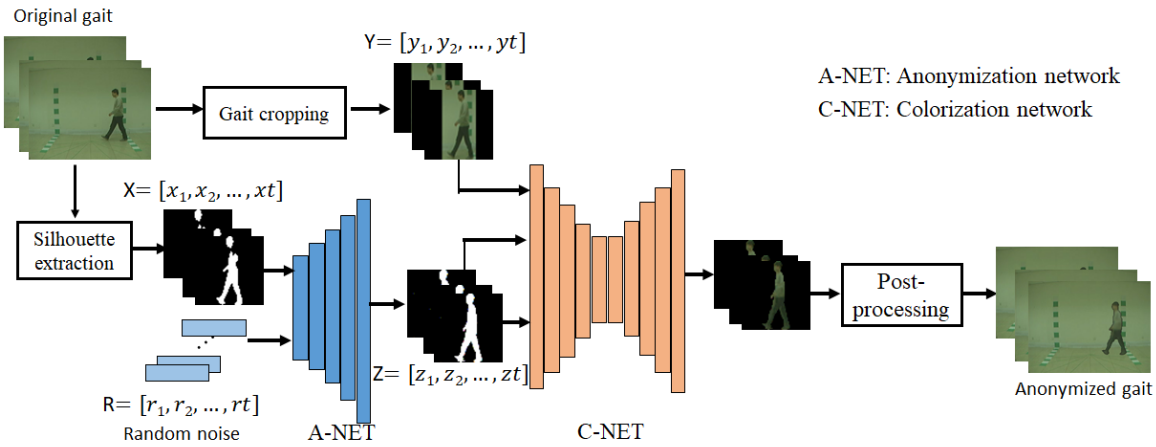


Fig. 5.3 Overview of model ICSGait-ANET.

In order to retain colors in the images of the gait we wish to protect as many as possible, the colorization network was designed with the Siamese structure that has two inputs: the output of ANET, and the RGB original gait image Y. Because this network is aimed at transferring colors of the original gait image to the binary anonymized gait image, only foreground colors are needed. However, if we remove the colors in the background region, the colors of the parts missing in the original gait images are impossible to be extracted. To address this issue, the original gait image was cropped along with the background. The cropping original gait image procedure is summarized as follows. Firstly, the object (gait) position at each frame was detected by using the pre-trained model YOLO [28], then the RGB original gait was cropped along with the background. Next, zero-padding was added to

the cropped image so that the width and the height of the image are the same and we obtain a square image. Finally, the image is resized to get an image with the size of 64x64x3.

In order to remove most colors in the background region, the original gait image is pixel-wise multiplied by the output of A-NET. The C-NET takes the result of this multiplication and the output of A-NET as two inputs. The original gait in the raw video is replaced by the RGB anonymized gait generated with C-NET by using the same method presented in chapter 4.

5.2.2 Anonymization Network

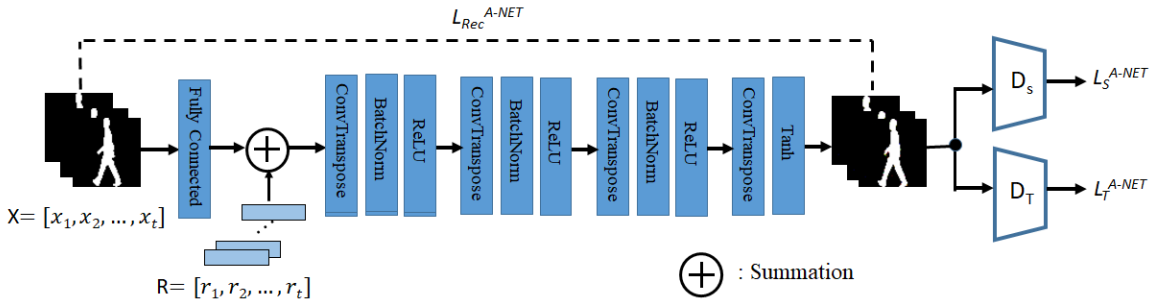
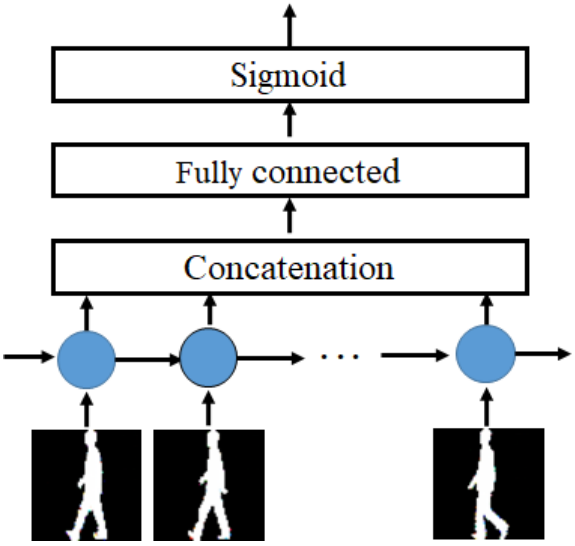


Fig. 5.4 Anonymization network architecture.

The goal of the anonymization network, A-NET, is to remove the original identity of the gait we wish to anonymize while preserving the appearance and motion of the original gait. Figure 5.4 describes the architecture of A-NET. This network is based on DCGAN model that generates images in space of the training dataset from random noise. However, defer from the traditional DCGAN that takes a random noise as the input, A-NET in this thesis has two inputs, the sequence of binary silhouettes X of the gait we wish to protect and a random noise vector whose role is to hide the identity this gait. Inherited from the model ST-GAN presented in chapter 4, A-NET is assigned by two discriminators, a spatial discriminator, and a temporal discriminator to force the gait generator network to produce a natural gait. The spatial discriminator is to maximize the distance from the shape of a generated gait to that of the real gait. The spatial discriminator takes an image of size 64x64x3 as the input and this network consists of convolution layers followed by a sigmoid function. The purpose of the temporal discriminator is to separate the movement of a synthesized gait from that of a real gait. The architectures of these discriminators are similar to those presented in chapter

4. The architectures of the spatial is illustrated in Figs. 5.5a, and temporal discriminators is shown in 5.5b.

(a) Architecture of spatial discriminator.



(b) Architecture of temporal discriminator.

Fig. 5.5 Architecture of two discriminators.

As discussed above, this chapter is aimed at generating natural anonymized gaits regardless of the silhouette quality of the original gait. Therefore, complete silhouette gaits were adopted to train A-NET. Two loss functions are used to train this network:

$$L_S^{A-NET} = E_{x\sim p_x(x),r\sim p_r(r)}[log(1-D_S(f_A(X,R)))] \tag{5.1}$$

$$L_T^{A-NET} = E_{x\sim p_x(x),r\sim p_r(r)}[log(1-D_T(f_A(X,R)))] \tag{5.2}$$

The reconstruction loss is also applied in order to preserve the viewing angle and walking action of the gait that we wish to anonymize.

$$L_{Rec}^{A-NET} = E_{x\sim p_x(x), r\sim p_r(r)}[\| X - f_A(X,R) \|_1] \tag{5.3}$$

where $f_A(.)$ is the output of A-NET.

A-NET was trained by optimizing the objective function below

$$L^{A-NET} = L_S^{A-NET} + L_T^{A-NET} + L_{Rec}^{A-NET} \tag{5.4}$$

5.2.3 Colorization Network

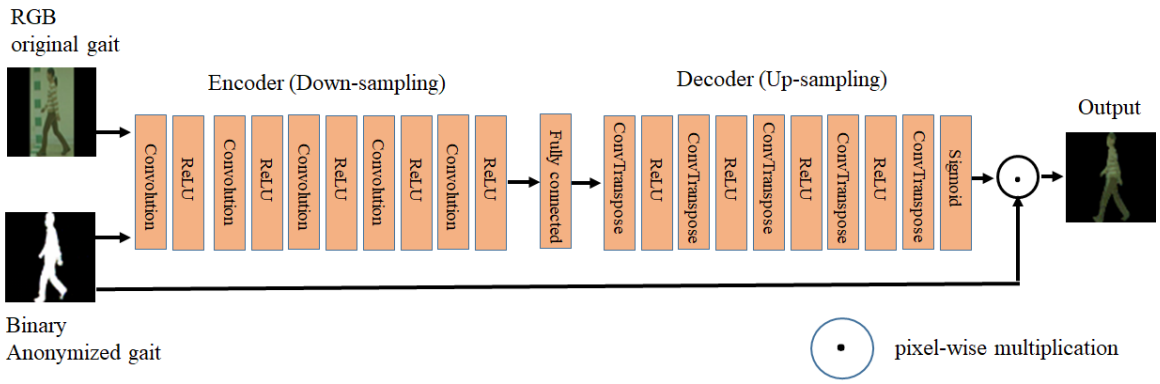


Fig. 5.6 Architecture of colorization network: The encoder compresses the input information into the hidden layer, and the decoder decodes the feature map of the hidden layer. Pixel-wise multiplication reforms the shape of the decoder output to that of the binary anonymized gait.

The colorization network C-NET is aimed at transferring the colors in the original gait images to the gait images synthesized by A-NET. Because only colors in the original subject image is transferred, the background colors are not needed. However, if we remove the colors of the background, it is unable to extract the colors of the parts missing in the original gait images occurring when colors of the subject garment is the same as the background color or the subject is partly occluded by another object. Therefore, the original gait image with the background was cropped.

Figure 5.6 illustrates the C-NET architecture. The 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. Color original gait images are captured along with the background from the original video, and the binary anonymized gait images are generated by the anonymization network. Because we aim at color transfer while preserving the body shape of gaits, 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 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 the colors of the remaining region so that coherent colors and textures are retained between the two regions.

Since this model is aimed at overlaying the colors of original gait images on binary anonymized gait images, 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. 3.4. 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. This network is aimed at transferring 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.

The center region is defined as the overlapping region between the two input gaits and the edge region is 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. 5.7. Because the binary anonymized gait is obtained by modifying the shape of the original gait while keeping the same phase, 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.

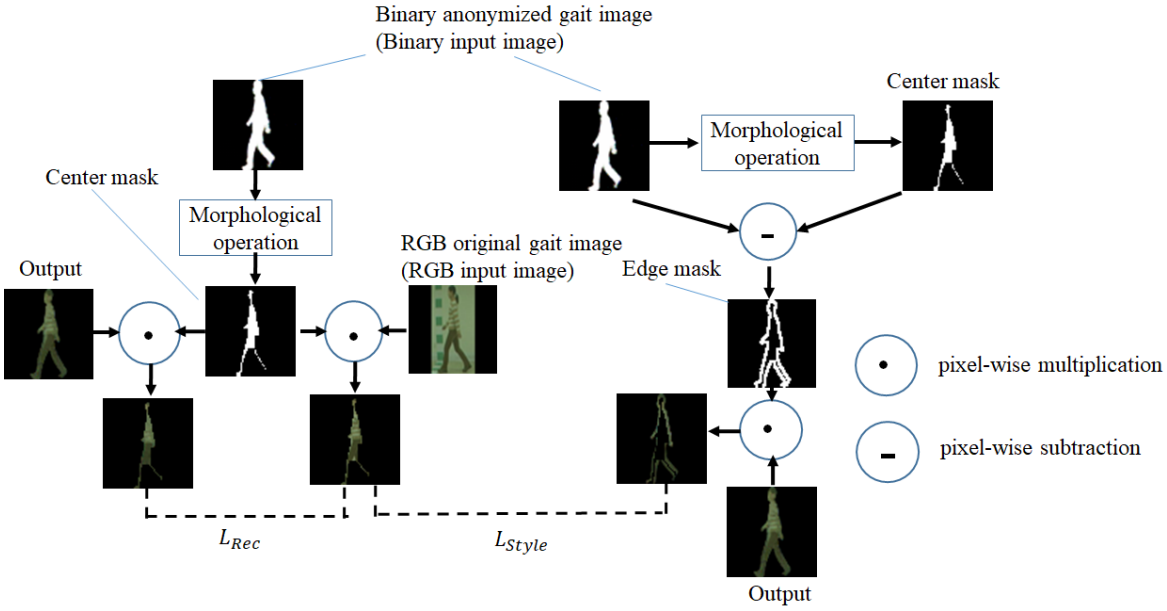


Fig. 5.7 Reconstruction loss L_{Rec} matches the center region of the RGB original gait image to that of the output while style loss L_{Style} matches the center region of the RGB original gait image to the edge region of the output.

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. 5.7, 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.

$$L_{Rec} = ||Mp(x_{bi}) \odot x_{rgb} - Mp(x_{bi}) \odot \Phi(x_{rgb}, x_{bi})||_1, \tag{5.5}$$

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.

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. [92] in generating beautiful stylized and textured images and its use in several studies [93, 94], 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.

$$Gr_{ij}(x) = \langle F_i, F_j \rangle = \sum_k F_{ik} F_{jk} \tag{5.6}$$

where k is the element of each channel.

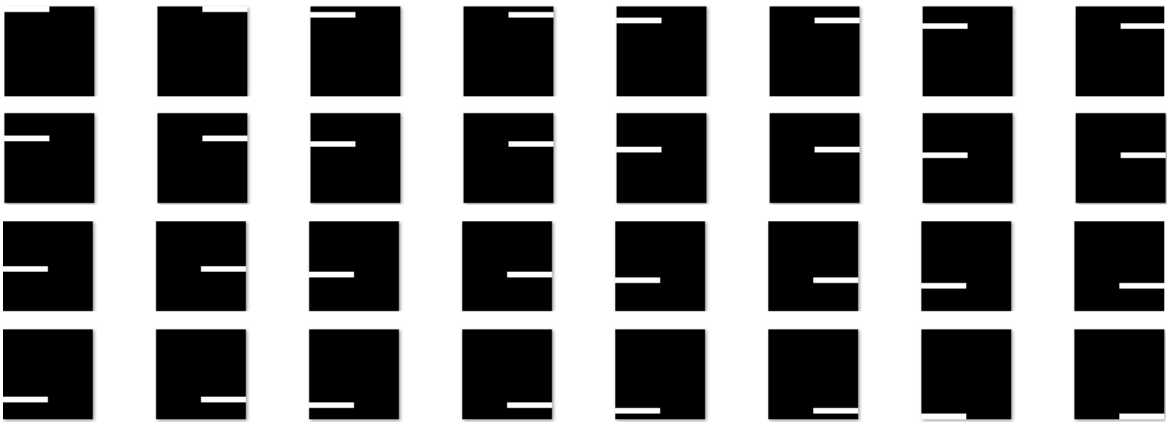


Fig. 5.8 Mask images were used to compute the style loss.

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 of the same size as the input images, as shown in Fig. 5.8, 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 central region and one in the edge region. The color of each patch in the edge region is generated on the basis of 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:

$$L_{Style} = \sum_{l} || \frac{1}{M_l} Gr(Mp(x_{bi}) \odot x_{rgb} \odot m_l) - \frac{1}{N_l} Gr((x_{bi} - Mp(x_{bi})) \odot \Phi(x_{bi}, x_{bi}) \odot m_l) ||_1, \tag{5.7}$$

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

$$M_l = \sum_{p} (Mp(x_{bi}) \odot m_l) \tag{5.8}$$

$$N_l = \sum_{p} ((x_{bi} - Mp(x_{bi})) \odot m_l) \tag{5.9}$$

where p is the element of each patch.

5.3 Experimental Results

Table 5.1 Dataset organization.

Tasks	Num. of subs	Num. of seqs.	Num. of frames
Training the gait recognition systems	50	5,500	275,000
Training A-NET	10	1,100	55,000
Training C-NET	16	17,600	88,000
Validation	8	880	44,000
Testing	40	4,400	220,000

CASIA-B gait dataset [67]. The dataset was divided into five non-overlapping groups as shown in Table 5.1. The first one consists of 50 subjects, equivalent to 5500 sequences,

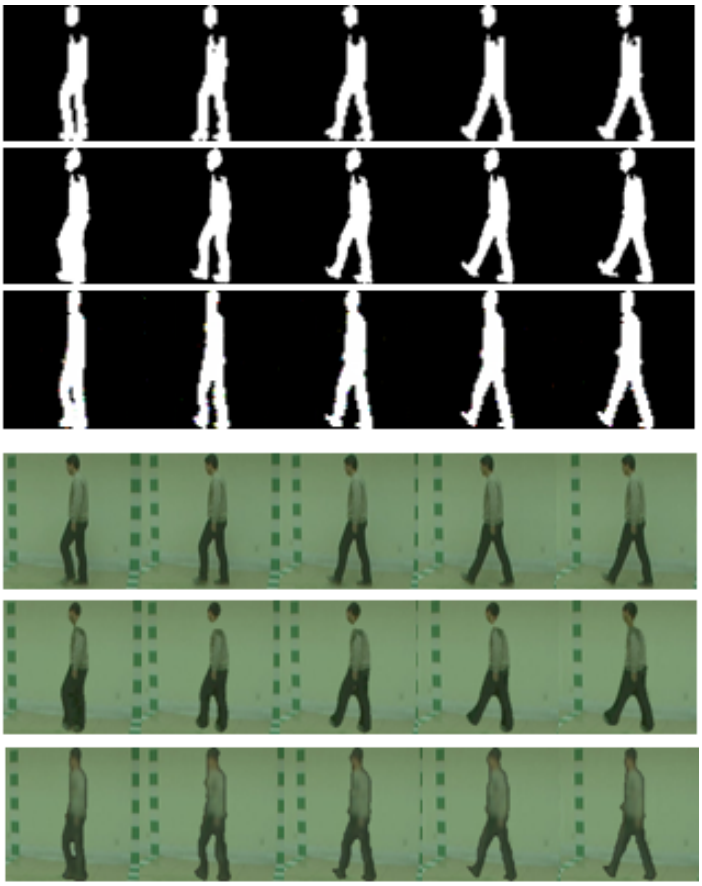


Fig. 5.9 Generation results produced from incomplete silhouette gaits by the model ICSGait-ANET and the model ST-GAN (Chapter 4) with viewing angle 108°: top rows, middle rows, top rows show original gaits, generated images of model ST-GAN, generated images of the ICSGait-ANET, respectively.

which was to train three gait recognition systems. The second one that consists of 10 subjects, equivalent to 1100 sequences was to train A-NET. The third one, which consists of 16 subjects, equivalent to 1760 sequences was to train C-NET. The fourth one, consisting of 8 individuals, equivalent to 880 sequences was for validation. The rest composes of the 40 subjects, equivalent to 4400 sequences, were reserved for testing.

5.3.1 Generation Results

This subsection gives the comparison in generation results of the model introduced in this chapter and the model ST-GAN in Chapter 4. This subsection also shows many generation results of binary/RGB anonymized gait images when complete/incomplete silhouette gait is used as well as the final result frame images, in which the generated gait are placed in

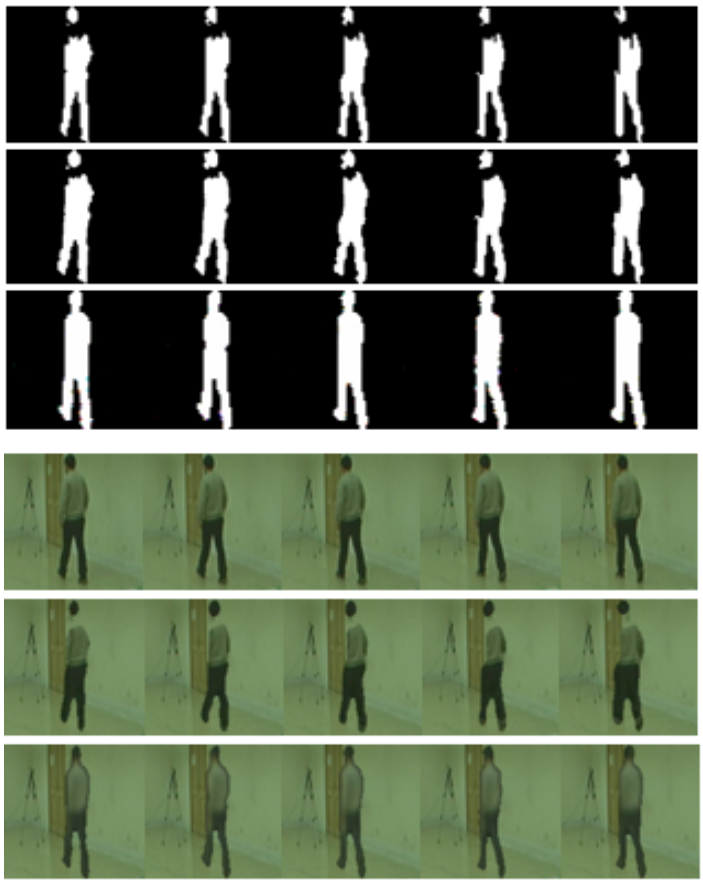


Fig. 5.10 Generation results produced from incomplete silhouette gaits by the model ICSGait-ANET and the model ST-GAN (Chapter 4) with viewing angle 144°: top rows, middle rows, top rows show original gaits, generated images of model ST-GAN, generated images of the ICSGait-ANET, respectively.

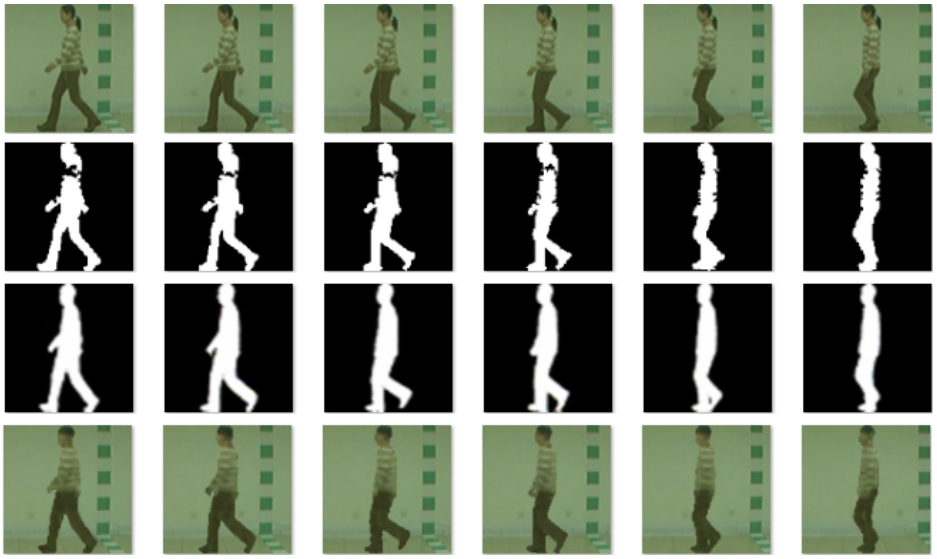


Fig. 5.11 Generation results produced from incomplete silhouette gaits by the model ICSGait-ANET for viewing angle 90°: Two top rows are RGB original gait images and silhouette of original gait, respectively; two bottom rows are silhouette of anonymized gait synthesized with A-NET and RGB anonymized gait images synthesized with C-NET, respectively.

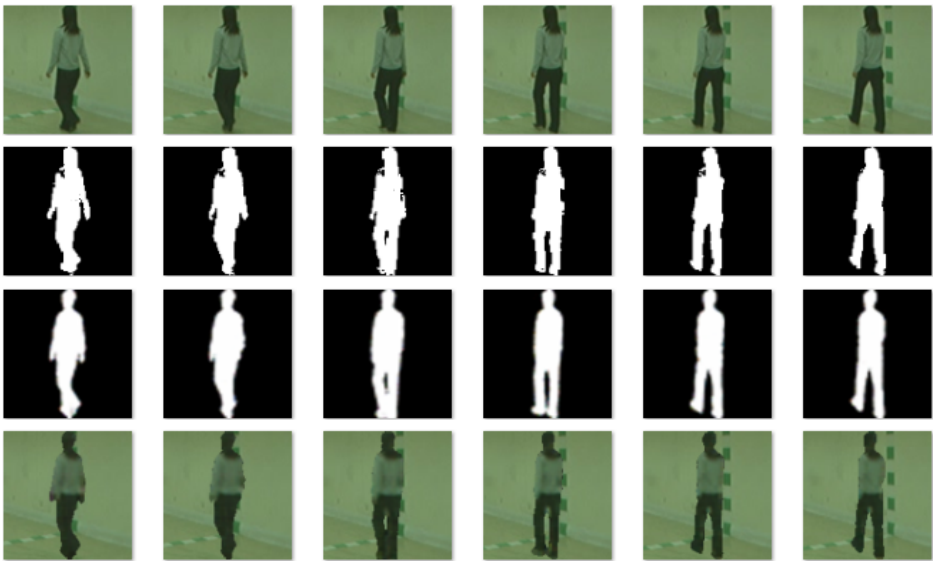


Fig. 5.12 Generation results produced from complete silhouette gaits by the model ICSGait-ANET for viewing angle 144°: Two top rows are RGB original gait images and silhouette of original gait, respectively; two bottom rows are silhouette of anonymized gait synthesized with A-NET and RGB anonymized gait images synthesized with C-NET, respectively.

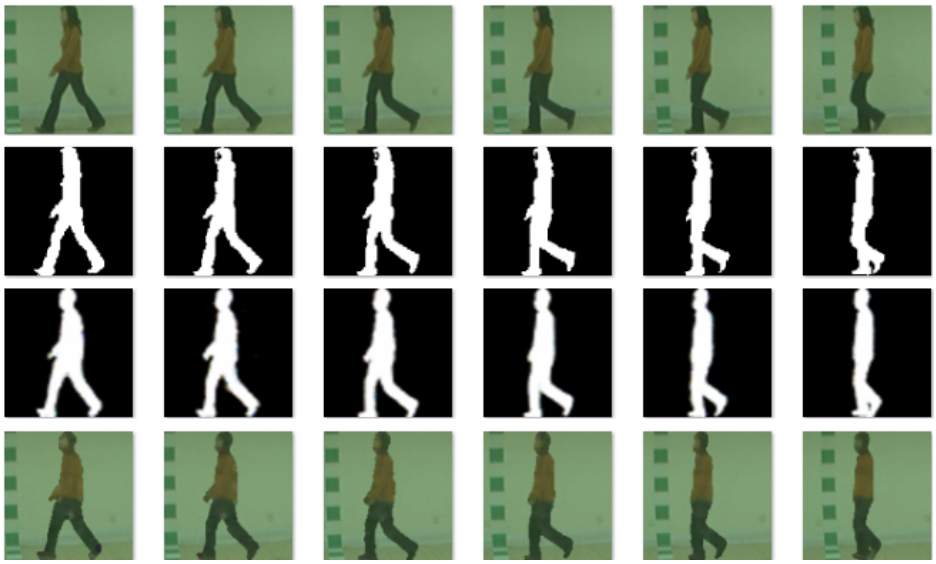


Fig. 5.13 Generation results produced from complete silhouette gaits by the model ICSGait-ANET for viewing angle 90°: Two top rows are RGB original gait images and silhouette of original gait, respectively; two bottom rows are silhouette of anonymized gait synthesized with A-NET and RGB anonymized gait images synthesized with C-NET, respectively.

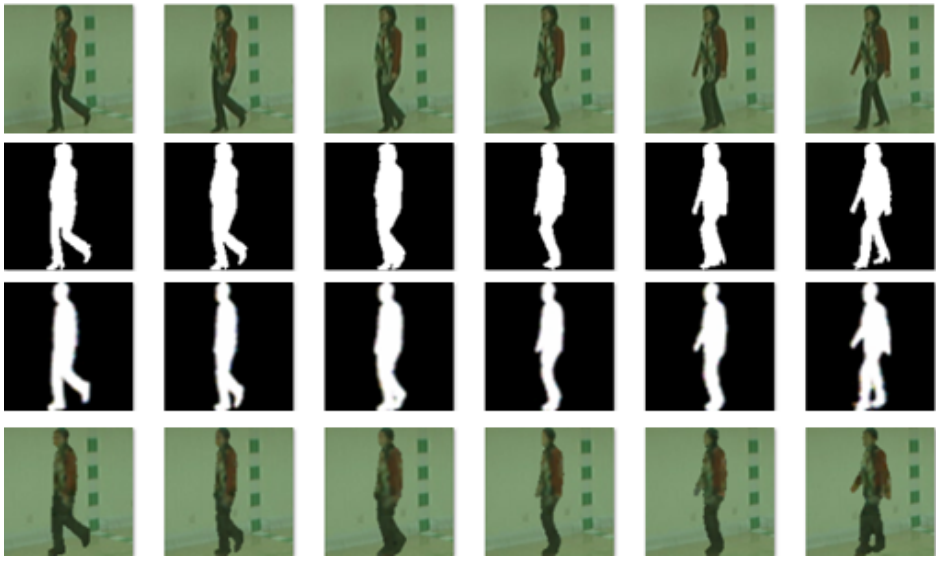
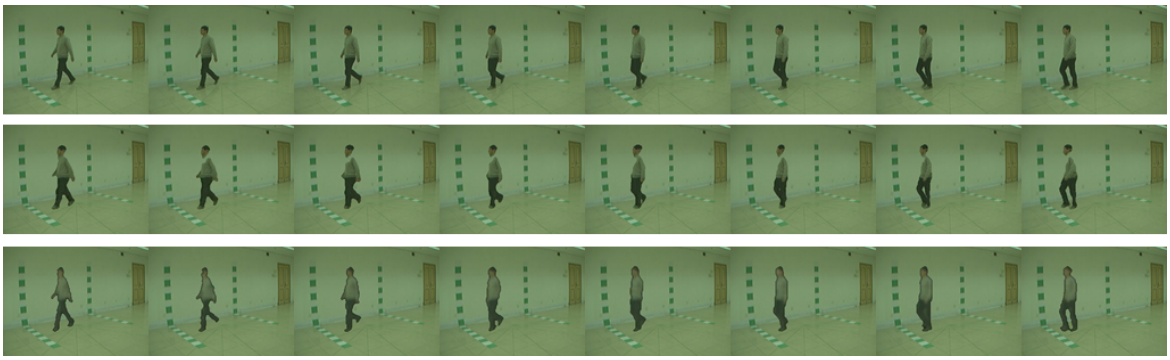
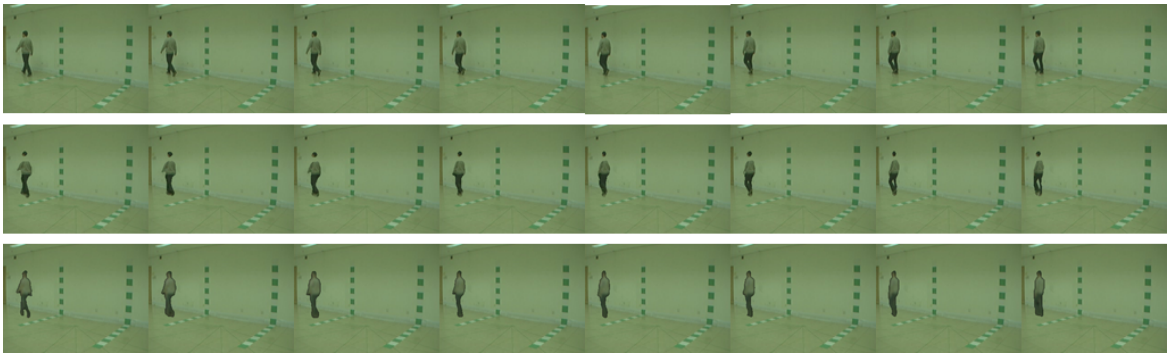


Fig. 5.14 Generation results produced from complete silhouette gaits by the model ICSGait-ANET for viewing angle 72°: Two top rows are RGB original gait images and silhouette of original gait, respectively; two bottom rows are silhouette of anonymized gait synthesized with A-NET and RGB anonymized gait images synthesized with C-NET, respectively.



(a) Viewing angles: 54°



(b) Viewing angles: 126°

Fig. 5.15 Generation results produced from incomplete silhouette gaits by model ICSGait-ANET and model ST-GAN of 54° and 126°: top rows show original gaits, middle rows show anonymized gaits of model ST-GAN, and bottom rows show anonymized gaits of model ICSGait-ANET.

the original scene. The Fig. 5.9 and Fig. 5.10 are comparison of generation results between two models when incomplete silhouettes are used. The Fig.5.11, Fig.5.12, Fig.5.13, Fig.5.14 are samples of results produced by the proposed model when both incomplete and complete silhouette are used. Meanwhile, Fig. 5.15 shows the final frame images.

There are 4 findings pointed here:

1) A-NET is able to produce a seamless anonymized gait image even with incomplete silhouettes of original gaits, which is not completely solved with both models BiGait-ANET and ST-GAN.

2) A-NET can change the temporal information (e.g., hand movement) of a gait.

3) C-NET can transfer colors in RGB original gait images to the binary anonymized gait images for both complete and incomplete silhouettes.

4) The blurry faces in generated gait images is because that the size of input gait images of ICSGait-Net is quite tiny (64x64x3).

5.3.2 Naturalness

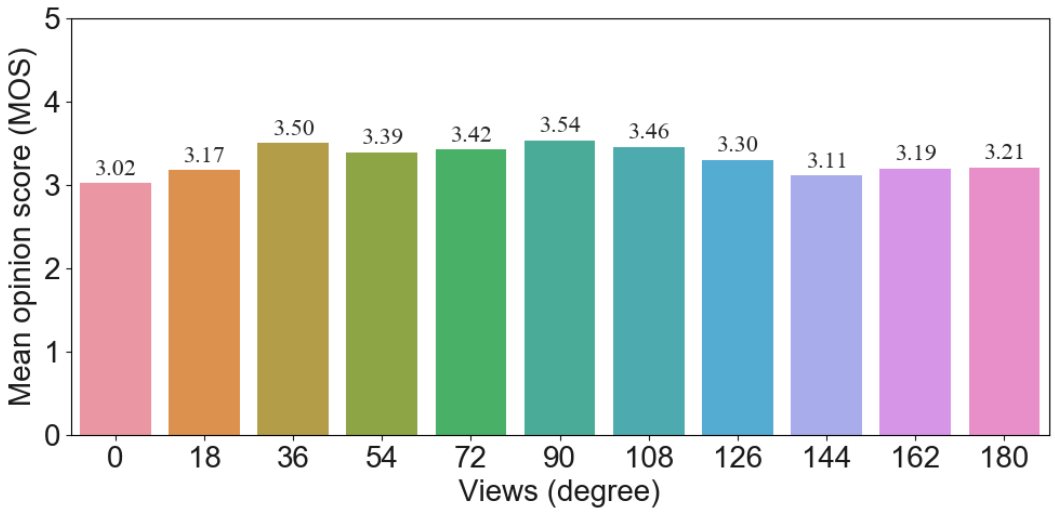


Fig. 5.16 MOS scores of anonymized gaits generated with ICSGait-Net.

The MOS test was used to assess the degree of data utility preservation of anonymized gaits synthesized by model ICSGait-Net. There were 20 evaluators joining this test and each of them evaluates 60 random anonymized gait videos. The evaluators were required to give a score of the degree of data utility preservation of synthesized gait video with the criteria of body shape, gait movement, and garment colors of gait images (compared with the

corresponding original video) with a five-point scale (1: Bad, 2: Poor, 3: Fair, 4: Good, 5: Excellent). The MOS score is shown in Fig. 5.16.

5.3.3 Success Rate

The success rate of the model ICSGat-ANet and model ST-GAN measured for three gait recognition systems is shown in Table 5.2, Table 5.3, and Table 5.4. The comparison of success rate with three gait recognition systems is summarized in Fig.5.17.

Table 5.2 The average success rate (%) of the proposed model ICSGait-ANET with Zheng's method.

top	**Viewing angles**										
	$\mathbf{0}^0$	$\mathbf{18}^0$	$\mathbf{36}^0$	$\mathbf{54}^0$	$\mathbf{72}^0$	$\mathbf{90}^0$	$\mathbf{108}^0$	$\mathbf{126}^0$	$\mathbf{144}^0$	$\mathbf{162}^0$	$\mathbf{180}^0$
top – 1	83.39	86.44	95.01	94.19	92.58	95.05	93.96	95.91	95.27	83.12	80.48
top – 3	76.86	78.58	89.72	91.59	90.89	89.87	90.46	88.73	84.18	78.95	77.92

Table 5.3 The average success rate (%) of the proposed model ICSGait-ANET evaluated with Wu's method.

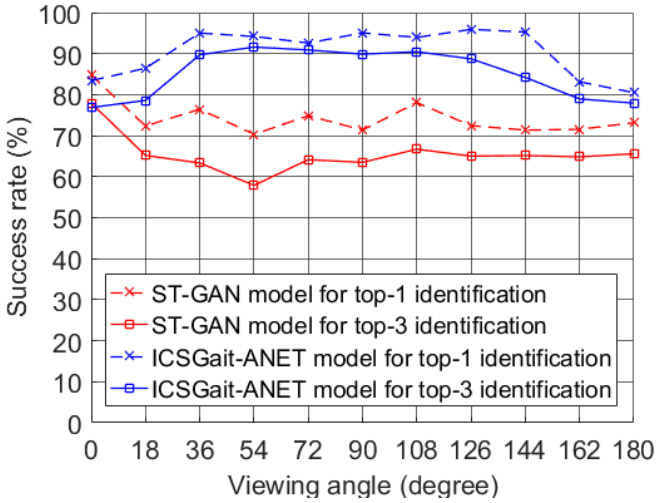
Top	**Viewing angles**										
	$\mathbf{0}^0$	$\mathbf{18}^0$	$\mathbf{36}^0$	$\mathbf{54}^0$	$\mathbf{72}^0$	$\mathbf{90}^0$	$\mathbf{108}^0$	$\mathbf{126}^0$	$\mathbf{144}^0$	$\mathbf{162}^0$	$\mathbf{180}^0$
top – 1	70.52	84.20	87.34	88.77	83.91	89.02	87.86	89.59	90.64	75.97	72.57
top – 3	67.78	68.36	75.85	80.27	78.32	80.97	79.18	80.95	82.80	72.69	71.01

Table 5.4 The average success rate (%) of the proposed model ICSGait-ANET evaluated with Chao's method.

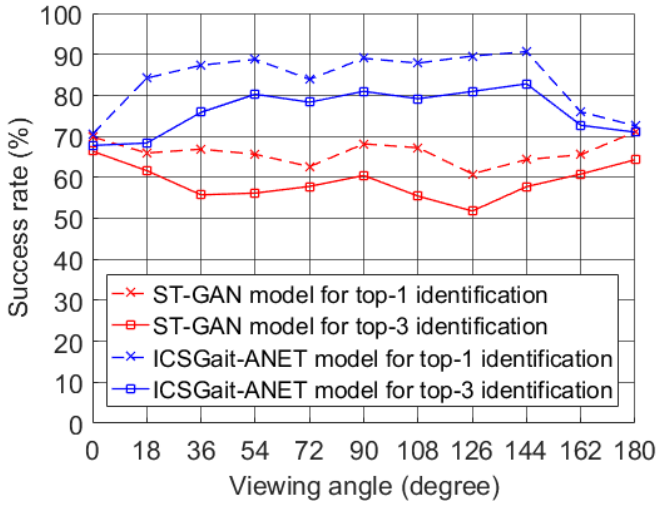
Top	**Viewing angles**										
	$\mathbf{0}^0$	$\mathbf{18}^0$	$\mathbf{36}^0$	$\mathbf{54}^0$	$\mathbf{72}^0$	$\mathbf{90}^0$	$\mathbf{108}^0$	$\mathbf{126}^0$	$\mathbf{144}^0$	$\mathbf{162}^0$	$\mathbf{180}^0$
top – 1	79.98	78.06	78.48	80.30	77.78	70.97	76.19	79.40	81.65	76.01	75.61
top – 3	78.66	76.69	77.57	76.06	73.65	69.84	74.05	75.26	77.02	75.28	74.87

The Fig.5.17a show the success rate of the anonymized gaits generated by each model computed by Zheng's method at top-1 and top-3, respectively, Fig. 5.17b show those computed by Wu method at top-1 and top-3, respectively, and Fig. 5.17c show those computed by Chao method with top-1 and top-3, respectively.

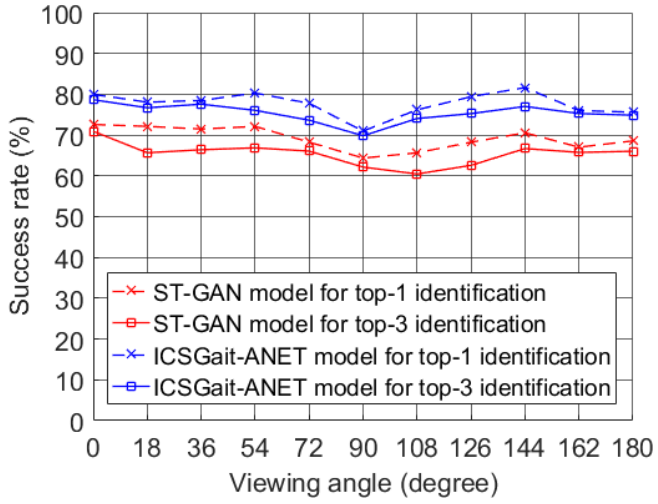
The success rate of model ICSGait-ANET was dramatically better than that of the model ST-GAN for the side viewing angles (from 36° to 144°) because with such viewing angles, temporal information (limbs movement) is one of key feature of gait biometric . Comparison



(a) Zheng's method



(b) Wu's method



(c) Chao's method

Fig. 5.17 Success rate comparison of the model ICSGait-ANET and the model ST-GAN.

of the anonymized gaits generated by the model ICSGait-ANET with those produced by model ST-GAN demonstrates that, in many cases, the model ICSGait-ANET changed the hand position while model ST-GAN did not. This deference between two models is because that the model ST-GAN needs to convert the random noise to the gait distribution before adding it to the original gait, while the model ICSGait-ANET directly adds the random noise to the original gait.

5.3.4 Robustness against Re-identification Attack

We also investigate whether we can find a machine learning model for re-identification attacks. Since model-free gait recognition systems take silhouettes that are binary images as inputs, if we stack a gait recognition system on the top of the re-identification model to force this model to restore the identity of the original gait from the anonymized gait, the output of the re-identification model must be binarized. However, the function that converts the output of the re-identification model to the binary images is a discrete function that has no gradient, and therefore the gradient of the whole model cannot be updated.

We tried using a traditional denoising autoencoder network [68] to restore the identity of the original gait from an anonymized gait. We used 20 subjects in 40 anonymized subjects to train this network and used the remaining 20 anonymized subjects to evaluate this network. In order to explore the robustness of the proposed gait anonymization model presented in this chapter, we compute the identification accuracy of three gait recognition systems on the re-identified gaits. Table 5.5 shows the this accuracy and a sample of re-identified result is shown in Fig. 5.18. Both Table 5.5 and Fig. 5.18 demonstrates that our model is robust to the re-identification attack.

Table 5.5 The average identification accuracy (%) of the re-identified gaits evaluated with Zheng's method, Wu's method, and Chao's method.

Top	**Viewing angles**										
	$\mathbf{0}^0$	$\mathbf{18}^0$	$\mathbf{36}^0$	$\mathbf{54}^0$	$\mathbf{72}^0$	$\mathbf{90}^0$	$\mathbf{108}^0$	$\mathbf{126}^0$	$\mathbf{144}^0$	$\mathbf{162}^0$	$\mathbf{180}^0$
top – 1	3.97	4.85	3.77	3.77	4.62	5.08	5.00	4.40	4.17	4.83	4.33
top – 3	7.94	7.79	7.68	7.25	7.39	7.46	8.00	8.81	8.00	7.17	8.00

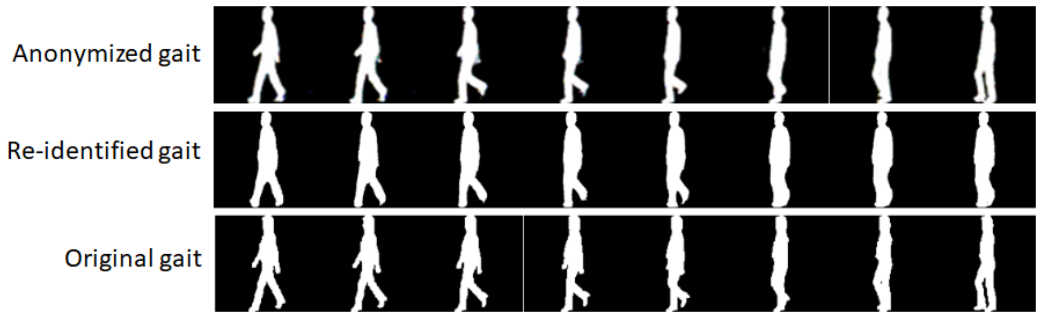


Fig. 5.18 Generation results of re-identification attack of ICSGait-ANET using traditional denoising autoencoder network.

5.4 Summary

The intensive evaluation demonstrates that the model ICSGait-ANET can generate natural anonymized gait regardless of the quality of the original gait that has not been completely solved by the model BiGait-ANET and ST-GAN. The model ICSGait-ANet includes two networks, one is to hide the original identity of a gait that we wish to anonymize with using a random noise, and the other is for colorization of binary anonymized gait images. The success rates of the ICSGait-ANet model are significantly higher than those of model BiGait-ANET and mode ST-GAN.

Chapter 6

Conclusion

6.1 Summary

Due to advances in computing technology, media content can be recorded and shared in many ways. Videos of people walking can be captured by surveillance cameras on the street by governments and private companies. Such data often contain sensitive information about the individuals such as their behaviors, routines, activities, and affiliations. There are many reasons for sharing such data among organizations or companies such as for law enforcement, forensics, and research. However, an attacker could crosslink a person in a video to a person in another video by using gait recognition systems, resulting in the disclosure of sensitive information about that person.

Another essential aspect in addition to protecting sensitive information is preserving the utility of the data as it can be used in many research areas [27], for instance, object detection [28, 29], action prediction [30], gender recognition, and clothing recognition [31, 32]. Therefore, research on gait anonymization that protects the privacy of people walking captured on video against gait recognition systems while maintaining the original attributes is essential. This thesis has focused on such research. Preservation of appearance means ensuring the naturalness of the body shape and preserving the clothing color of the original gait images. Preservation of motion consists of guaranteeing the naturalness of movement and retaining the moving direction of the original gait. Through the various analyses using signal processing, machine learning, and computer vision techniques, the following contributions are made:

- **This is the first research on gait anonymization aimed at preservation of both appearance and motion**: Previous research on privacy protection of people captured on video focused on such methods as object removal, object replacement, and pixelating or blurring the body. Such methods generally focus on privacy concerns and do not address the problem of guaranteeing the naturalness of the generated videos. As far as we know, the work introduced in this thesis is the first study on gait anonymization aimed at preservation of both appearance and motion. The results can be used to prevent sensitive information about people walking captured on video from being revealed by gait recognition while taking into account preservation of the original attributes so that can be used for data analysis.

- **Binary gait anonymization:** Most gait recognition approaches take the silhouette sequences of people walking. These sequences of binary image contain the person's shape and walking style, which are the gait's features. This work described in this thesis thus initially investigated binary gait anonymization. The key idea of the model proposed for doing this is to alter the body shape of a gait so that the anonymized gait is partly similar to the original gait and partly similar to another gait, named a "noise gait." The proposed model has two inputs: the contour vector of the gait we wish to anonymize and that of the noise gait. It produces a modified contour vector, which is then used to produce an anonymized gait. Although this model shows promising performance, the anonymized gait looks less natural from the frontal view. Since adding a "noise gait" to the original gait may make the success rate decreased if gait recognition systems are unable to distinguish the original gait from the noise gait, selecting the perfect noise gait is not easy.

- **RGB gait anonymization:** Although most gait recognition research has focused on the silhouette sequence, most videos uploaded and/or shared are RGB videos. Development of an approach for RGB gait anonymization is essential, and the research reported in this thesis addressed the anonymization of RGB gaits. The ST-GAN model, which includes one generator and two discriminators were proposed. This model removes the identity of the original gait by adding a random noise in the gait space to the original gait. Two discriminators, a spatial discriminator, and a temporal discriminator, are used to improve the quality of the generated results. The goal of the spatial discriminator network is to determine whether the gait image is real or fake at each frame. A convolutional neural network is used for this discriminator. The goal of

the temporal discriminator network is to ensure that the motion of a generated gait is smooth by distinguishing the temporal information of a real gait from that of a fake gait. An LSTM-based architecture is used for this discriminator.

- **Incomplete silhouette gait anonymization:** The work described in this thesis also investigated how to generate a seamless anonymized gait when the silhouette gaits used are incomplete. The focus was on improving the naturalness of the anonymized gaits as well as improving the anonymization success rates. The model proposed for doing this consists of two networks, an anonymization network and a colorization network. The anonymization network is based on the DCGAN model and is trained on a complete silhouette dataset to generate seamless silhouettes while the colorization network transfers the clothing colors in the original gait images to the anonymized gait images without extracting the clothing colors to avoid missing the original colors.

From intensive analysis and evaluation, the following findings were obtained:

- The proposed models are able to generate a high-quality anonymized gaits with a high success rate of anonymization that are robust against re-identification attacks.
- Modifying the shapes of the silhouettes in a gait sequence can fool gait recognition systems.
- Using a random noise to modify the shape of the body achieves a higher success rate than using a noise in the gait distribution that is added to the shape of the body.
- Using a deep learning technique ensures the naturalness of the anonymized gait in terms of body shape, gait movement, and clothing color in the gait image.
- The success rate of anonymization is inversely proportional to gait recognition accuracy.
- In a gait anonymization model, the data utility preservation is inversely proportional to the success rate of anonymization.

6.2 Discussion

The studies described in this thesis on gait anonymization were aimed at protecting the privacy of people walking who are captured on video against gait recognition systems

while maintaining the original attributes. However, a person in a video can be recognized by several other biometrics such as face, fingerprint, iris, and voice with high accuracy [18–20]. Therefore, to completely protect a person in a video from recognition systems, gait anonymization techniques should be integrated with other biometric anonymization techniques.

Beside privacy protection, data utility is an important aspect of gait anonymization because it is necessary for many research areas [27], for instance, object detection [28, 29], action prediction [30], gender recognition, and clothing recognition [31, 32]. In this research, human perspective evaluations were conducted to measure the degree of data utility preservation in terms of appearance and motion. Assessing the data utility preservation by machine should be also considered and conducted in future gait anonymization research. Such measurement can be assessed by human action recognition, human detection, clothing recognition techniques.

Though the gait dataset utilized in this thesis is huge and widely used, the environment represented by the data, such as static backgrounds and viewing angles, has been controlled so that the people in the video can be extracted exactly. Therefore, when applying the proposed models to real applications in which the background is dynamic and a person may move in various viewing angles, image inpainting techniques should be used to fill the gaps in the background that occur when the shape of a person in a video is changed.

6.3 Future Work

The work described in this thesis is the first research on gait anonymization while preserving gait naturalness. Although techniques were proposed for solving several problems in this area, there are still many problems to be addressed. Future work includes improving the proposed techniques. There are several potential new research directions:

- **Re-identification attacks on gait anonymization:** The proposed gait anonymization models have been proven to be robust against re-identification attacks that use a commonly used approach as well as a traditional denoising autoencoder. However, an intensive study on such attacks on gait anonymization models is necessary in order to make gait anonymization models more robust.

- **Group-based gait anonymization:** Investigating gait anonymization based on k-anonymity and t-closeness, which make anonymization models robust against linking attacks and reference attacks.
- **Model-based gait anonymization:** The proposed gait anonymization models are based on model-free gait recognition; therefore, research on gait anonymization that relies on model-based gait recognition should be considered so that gait anonymization is more generally applicable.
- **Body-based gait anonymization:** The proposed gait anonymization models alter the identity of a gait by modifying the body shape. Research based on modifying each body part of a gait should be conducted for gait anonymization and should use body generation techniques.
- **Gait anonymization in the wild:** Although the gait dataset used in this thesis is widely used, is huge, and has multiple viewing angles, there is only one person walking in each video. It is essential that gait anonymization research use more challenging datasets, such as ones with videos with occluded people and videos with people carrying things.

References

[1] E. Newton, L. Sweeney, and B. Malin, "Preserving privacy by de-identifying face images," IEEE Transactions on Knowledge and Data Engineering, vol. 17, no. 2, pp. 232–243, 2005.

[2] B. Meden, Ž. Emeršič, V. Štruc, and P. Peer, "k-same-net: k-anonymity with generative deep neural networks for face deidentification," Entropy, vol. 20, no. 1, p. 60, 2018.

[3] T. Li and L. Lin, "Anonymousnet: Natural face de-identification with measurable privacy," in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 56–65, 2019.

[4] P. Agrawal and P. J. Narayanan, "Person de-identification in videos," IEEE Transactions on Circuits and Systems for Video Technology, vol. 21, pp. 299–310, March 2011.

[5] 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, p. 075427, Jan 2007.

[6] J. Fan, H. Luo, M.-S. Hacid, and E. Bertino, "A Novel Approach for Privacy-Preserving Video Sharing," in 14th ACM international conference on Information and knowledge management, CIKM 2005, (Bremen, Germany), pp. 609–616, ACM, Oct. 2005.

[7] M. Granados, J. Tompkin, K. Kim, O. Grau, J. Kautz, and C. Theobalt, "How not to be seen — object removal from videos of crowded scenes," Comput. Graph. Forum, vol. 31, p. 219–228, May 2012.

[8] B. Samarzija and S. Ribaric, "An approach to the de-identification of faces in different poses," in 2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 1246–1251, IEEE, 2014.

[9] H. Zhang, H. Zhou, W. Jiao, J. Shi, Q. Zang, J. Sun, and J. Zhang, "Biological features de-identification in iris images," in 2018 15th International Symposium on Pervasive Systems, Algorithms and Networks (I-SPAN), pp. 67–71, IEEE, 2018.

[10] Q. Sun, L. Ma, S. J. Oh, L. Van Gool, B. Schiele, and M. Fritz, "Natural and effective obfuscation by head inpainting," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5050–5059, 2018.

[11] H. Hukkelås, R. Mester, and F. Lindseth, "Deepprivacy: A generative adversarial network for face anonymization," in International Symposium on Visual Computing, pp. 565–578, Springer, 2019.

[12] 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), pp. 2073–2076, Sept 2011.

[13] 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 Transactions on Pattern Analysis and Machine Intelligence, vol. 39, pp. 209–226, Feb 2017.

[14] H. Chao, K. Wang, Y. He, J. Zhang, and J. Feng, "Gaitset: Cross-view gait recognition through utilizing gait as a deep set," IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1–1, 2021.

[15] D. J. Solove, Nothing to hide: The false tradeoff between privacy and security. Yale University Press, 2011.

[16] S. Ribaric, A. Ariyaeeinia, and N. Pavesic, "De-identification for privacy protection in multimedia content: A survey," Signal Processing: Image Communication, vol. 47, pp. 131–151, 2016.

[17] A. Ross and A. K. Jain, "Human recognition using biometrics: an overview," in Annales Des Télécommunications, vol. 62, pp. 11–35, Springer, 2007.

[18] Y. Sun, C. Fookes, N. Poh, and M. Tistarelli, "Cohort normalization based sparse representation for undersampled face recognition," in Proceedings of the Workshop on Face Analysis: The Intersection of Computer Vision and Human Perception as part of the 11th Asian Conference on Computer Vision, pp. 1–13, Cardiff University, 2012.

[19] C. McCool, V. Chandran, S. Sridharan, and C. Fookes, "3d face verification using a free-parts approach," Pattern Recognition Letters, vol. 29, no. 9, pp. 1190–1196, 2008.

[20] C. Fookes, D. Chen, R. Lakemond, and S. Sridharan, "Robust facial feature extraction and matching," Journal of Pattern Recognition Research, vol. 7, no. 1, pp. 140–154, 2012.

[21] L. Lee and W. Grimson, "Gait analysis for recognition and classification," in Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition, pp. 155–162, 2002.

[22] J. Wang, M. She, S. Nahavandi, and A. Kouzani, "A review of vision-based gait recognition methods for human identification," in 2010 International Conference on Digital Image Computing: Techniques and Applications, pp. 320–327, Dec 2010.

[23] T. Yamada, S. Gohshi, and I. Echizen, "Privacy visor: Method based on light absorbing and reflecting properties for preventing face image detection," in 2013 IEEE International Conference on Systems, Man, and Cybernetics, pp. 1572–1577, Oct 2013.

[24] D. Zhang, "Automated biometrics: technnologies and systems.," Springer, 2000.

[25] A. K. Jain, R. Bolle, and S. Pankanti, "Biometrics: personal identification in networked society.," Springer, 1999.

[26] M. A. E., "The biomechanics of skipping gaits: a third locomotion paradigm?.," Proceedings Biological sciences, p. 1227–1235, 1998.

[27] M. Paul, S. M. Haque, and S. Chakraborty, "Human detection in surveillance videos and its applications-a review," EURASIP Journal on Advances in Signal Processing, vol. 2013, no. 1, pp. 1–16, 2013.

[28] 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), pp. 779–788, 2016.

[29] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," CoRR, vol. abs/1804.02767, 2018.

[30] Y. Kong, D. Kit, and Y. Fu, "A discriminative model with multiple temporal scales for action prediction," in European conference on computer vision, pp. 596–611, Springer, 2014.

[31] M. Yang and K. Yu, "Real-time clothing recognition in surveillance videos," in 2011 18th IEEE International Conference on Image Processing, pp. 2937–2940, IEEE, 2011.

[32] S. C. Hidayati, C.-W. You, W.-H. Cheng, and K.-L. Hua, "Learning and recognition of clothing genres from full-body images," IEEE transactions on cybernetics, vol. 48, no. 5, pp. 1647–1659, 2017.

[33] F. Z. Qureshi, "Object-video streams for preserving privacy in video surveillance," in 2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance, pp. 442–447, Sept 2009.

[34] M. Ivasic-Kos, A. Iosifidis, A. Tefas, and I. Pitas, "Person de-identification in activity videos," in 2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 1294–1299, May 2014.

[35] J. P. Singh, S. Jain, S. Arora, and U. P. Singh, "Vision-based gait recognition: A survey," IEEE Access, vol. 6, pp. 70497–70527, 2018.

[36] R. G. Birdal, A. Sertbaş, and B. Mïhendisliği, "Human identification based on gait analysis: A survey," in 2018 3rd International Conference on Computer Science and Engineering (UBMK), pp. 489–493, 2018.

[37] C. Wan, L. Wang, and V. V. Phoha, "A survey on gait recognition," vol. 51, Aug. 2018.

[38] L.-F. Liu, W. Jia, and Y.-H. Zhu, "Survey of gait recognition," in Emerging Intelligent Computing Technology and Applications. With Aspects of Artificial Intelligence (D.-S. Huang, K.-H. Jo, H.-H. Lee, H.-J. Kang, and V. Bevilacqua, eds.), (Berlin, Heidelberg), pp. 652–659, Springer Berlin Heidelberg, 2009.

[39] 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), pp. 320–327, 2010.

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

[41] P. Samarati and L. Sweeney, "Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression," in Technique report, SRI International, 1998.

[42] A. Jourabloo, X. Yin, and X. Liu, "Attribute preserved face de-identification," in 2015 International conference on biometrics (ICB), pp. 278–285, IEEE, 2015.

[43] B. Meden, R. C. Mallı, S. Fabijan, H. K. Ekenel, V. Štruc, and P. Peer, "Face deidentification with generative deep neural networks," IET Signal Processing, vol. 11, no. 9, pp. 1046–1054, 2017.

[44] B. Meden, Z. Emersic, V. Struc, and P. Peer, "k-same-net: Neural-network-based face deidentification," in 2017 International Conference and Workshop on Bioinspired Intelligence (IWOBI), pp. 1–7, 2017.

[45] N. Li, T. Li, and S. Venkatasubramanian, "t-closeness: Privacy beyond k-anonymity and l-diversity," in 2007 IEEE 23rd International Conference on Data Engineering, pp. 106–115, 2007.

[46] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in 2015 IEEE International Conference on Computer Vision (ICCV), pp. 3730–3738, 2015.

[47] M. Boyle, C. Edwards, and S. Greenberg, "The effects of filtered video on awareness and privacy," in Proceedings of the 2000 ACM conference on Computer supported cooperative work, pp. 1–10, 2000.

[48] C. Neustaedter, S. Greenberg, and M. Boyle, "Blur filtration fails to preserve privacy for home-based video conferencing," ACM Transactions on Computer-Human Interaction (TOCHI), vol. 13, no. 1, pp. 1–36, 2006.

[49] C. Zhang, Y. Rui, and L.-w. He, "Light weight background blurring for video conferencing applications," in 2006 International Conference on Image Processing, pp. 481–484, IEEE, 2006.

[50] A. Frome, G. Cheung, A. Abdulkader, M. Zennaro, B. Wu, A. Bissacco, H. Adam, H. Neven, and L. Vincent, "Large-scale privacy protection in google street view," in 2009 IEEE 12th international conference on computer vision, pp. 2373–2380, IEEE, 2009.

[51] A. Othman and A. Ross, "Privacy of facial soft biometrics: Suppressing gender but retaining identity," in Computer Vision - ECCV 2014 Workshops (L. Agapito, M. M. Bronstein, and C. Rother, eds.), (Cham), pp. 682–696, Springer International Publishing, 2015.

[52] N. Ruchaud and J. L. Dugelay, "De-genderization by body contours reshaping," in 2017 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA), pp. 1–6, Feb 2017.

[53] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in Proceedings of the 27th annual conference on Computer graphics and interactive techniques, pp. 417–424, 2000.

[54] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in 2009 IEEE 12th international conference on computer vision, pp. 2272–2279, IEEE, 2009.

[55] A. R. Abraham, A. K. Prabhavathy, and J. D. Shree, "A survey on video inpainting," International Journal of Computer Applications, vol. 56, no. 9, 2012.

[56] K. A. Panchal and M. Holia, "A survey: Different techniques of video inpainting," International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE), pp. 2923–2928, 2014.

[57] O. Gafni, L. Wolf, and Y. Taigman, "Live face de-identification in video," in Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 9378–9387, 2019.

[58] Y. Wu, F. Yang, Y. Xu, and H. Ling, "Privacy-protective-gan for privacy preserving face de-identification," Journal of Computer Science and Technology, vol. 34, no. 1, pp. 47–60, 2019.

[59] Z. Ren, Y. J. Lee, and M. S. Ryoo, "Learning to anonymize faces for privacy preserving action detection," in Proceedings of the european conference on computer vision (ECCV), pp. 620–636, 2018.

[60] C. Ledig, L. Theis, F. Huszár, 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), pp. 105–114, July 2017.

[61] 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), pp. 3915–3923, 2018.

[62] A. v. d. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, "Wavenet: A generative model for raw audio," 2016. arxiv:1609.03499.

[63] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. P. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," 2016.

[64] M. Sharif, S. Bhagavatula, L. Bauer, and M. K. Reiter, "Accessorize to a Crime: Real and Stealthy Attacks on State-of-the-Art Face Recognition," in ACM SIGSAC, pp. 1528–1540, 2016.

[65] J. Henriksen-Bulmer and S. Jeary, "Re-identification attacks—a systematic literature review," International Journal of Information Management, vol. 36, no. 6, Part B, pp. 1184–1192, 2016.

[66] D. Su, H. T. Huynh, Z. Chen, Y. Lu, and W. Lu, "Re-identification attack to privacy-preserving data analysis with noisy sample-mean," in Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, p. 1045–1053, Association for Computing Machinery, 2020.

[67] 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, pp. 441–444, 2006.

[68] Y. Bengio, "Learning deep architectures for ai," Found. Trends Mach. Learn., vol. 2, p. 1–127, Jan. 2009.

[69] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2, NIPS'14, (Cambridge, MA, USA), pp. 2672–2680, MIT Press, 2014.

[70] A. Nguyen, A. Dosovitskiy, J. Yosinski, T. Brox, and J. Clune, "Synthesizing the preferred inputs for neurons in neural networks via deep generator networks," in Advances in Neural Information Processing Systems 29 (D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, eds.), pp. 3387–3395, Curran Associates, Inc., 2016.

[71] C. Li and M. Wand, "Precomputed real-time texture synthesis with markovian generative adversarial networks," in ECCV, 2016.

[72] 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 (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds.), pp. 5967–5976, Curran Associates, Inc., 2017.

[73] J. Zhu, P. Krähenbühl, E. Shechtman, and A. A. Efros, "Generative visual manipulation on the natural image manifold," in ECCV (5), vol. 9909 of Lecture Notes in Computer Science, pp. 597–613, Springer, 2016.

[74] C. Yang, X. Lu, Z. Lin, E. Shechtman, O. Wang, and H. Li, "High-resolution image inpainting using multi-scale neural patch synthesis," in CVPR, pp. 4076–4084, IEEE Computer Society, 2017.

[75] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2242–2251, Oct 2017.

[76] J.-Y. Zhu, R. Zhang, D. Pathak, T. Darrell, A. A. Efros, O. Wang, and E. Shechtman, "Toward multimodal image-to-image translation," in Advances in Neural Information Processing Systems 30 (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds.), pp. 465–476, Curran Associates, Inc., 2017.

[77] Z. Wang, Q. She, and T. E. Ward, "Generative adversarial networks in computer vision: A survey and taxonomy," ACM Comput. Surv., vol. 54, Feb. 2021.

[78] M. Saito, E. Matsumoto, and S. Saito, "Temporal generative adversarial nets with singular value clipping," in ICCV, 2017.

[79] S. Tulyakov, M. Liu, X. Yang, and J. Kautz, "Mocogan: Decomposing motion and content for video generation," CoRR, vol. abs/1707.04993, 2017.

[80] C. Vondrick, H. Pirsiavash, and A. Torralba, "Generating videos with scene dynamics," in Proceedings of the 30th International Conference on Neural Information Processing Systems, NIPS'16, (USA), pp. 613–621, Curran Associates Inc., 2016.

[81] 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, MM '17, (New York, NY, USA), pp. 199–207, ACM, 2017.

[82] P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," CoRR, vol. abs/1611.07004, 2016.

[83] 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, pp. 2073–2076, Sept 2011.

[84] S. Gabriel-Sanz, R. Vera-Rodriguez, P. Tome, and J. Fierrez, "Assessment of gait recognition based on the lower part of the human body," in 2013 International Workshop on Biometrics and Forensics (IWBF), pp. 1–4, 2013.

[85] I. Rida, A. Bouridane, S. Al Kork, and F. Bremond, "Gait recognition based on modified phase only correlation," in Image and Signal Processing (A. Elmoataz, O. Lezoray, F. Nouboud, and D. Mammass, eds.), (Cham), pp. 417–424, Springer International Publishing, 2014.

[86] 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, 2016.

[87] 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, pp. 92:1–92:36, Sept. 2018.

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

[89] 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), pp. 853–862, 2017.

[90] 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), pp. 7543–7552, June 2018.

[91] 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), pp. 5184–5193, June 2020.

[92] 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), pp. 2414–2423, June 2016.

[93] 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), pp. 7982–7991, June 2019.

[94] 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, pp. 1349–1357, June 2016.