2020 25th International Conference on Pattern Recognition (ICPR)
Milan, Italy, Jan 10-15, 2021

Learning Disentangled Representations for Identity Preserving Surveillance Face Camouflage

Jingzhi Li[1,2], Lutong Han[1,2], Hua Zhang[1,2], Xiaoguang Han[3]
Jingguo Ge[1], Xiaochun Cao[1,2,4]
[1]Institute of Information Engineering, CAS, Beijing, China
[2]School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China
[3]Shenzhen Research Institute of Big Data, Shenzhen 518000, China
[4]Cyberspace Security Research Center, Peng Cheng Laboratory, Shenzhen 518055, China
Email: lijingzhi@iie.ac.cn, zhanghua@iie.ac.cn, hanxiaoguang@cuhk.edu.cn, caoxiaochun@iie.ac.cn

***Abstract*—In this paper, we focus on protecting the facial privacy for people under the surveillance scenarios, by changing some visual appearances of the faces while keeping them recognizable by the current face recognition systems. This is a challenging problem because we need to retain the most important structures of the captured facial images, while modify the salient facial regions to protect personal privacy.To address this problem, we introduce a novel individual face protection model, which can camouflage the face appearance from the perspective of human visual perception and preserve the identity features of faces used for face authentication. To that end, we develop an encoder-decoder network architecture which can separately disentangle the facial feature representation into an appearance code and an identification code. Specifically, we first randomly divide the input face image into two groups, the source and target sets, where the identity and appearance codes can be correspondingly extracted.Then, we recombine the identity and appearance codes to synthesize a new face, which has the same identity as the source subject. Finally, the synthesized faces are employed to replace the original face to protect the individual privacy.Note that our model is end-to-end with a multi-task loss function, which can better preserve the identity and stabilize the training process.Experiments conducted on Cross-Age Celebrity dataset demonstrate the effectiveness of our model and validate our superiority in terms of visual quality and scalability.**

I. Introduction

With the popularity of surveillance camera equipments, billions of photos and videos are generated every day. Although the deployments of surveillance have brought the convenience and enhance the security of people, which may also capture the individual faces and raise the risks of privacy leakage. The privacy concerns are growing as time goes on, and there are no signs of this trend slowing down.Furthermore, a series of laws and regulations have restricted the usability of face images. Thus, facial anonymization technology has attracted widespread public attentions, which has become a hot topic in the field of computer vision.

The traditional approaches [1]–[3] on face anoymization are mainly obfuscation-based, which attempts to protect facial privacy by applying naive transformations, e.g. blurring, pixelization, and masking.Moreover, Newton et al. [4] introduced the concept of K-anonymity into face anonymity, which aims to obfuscate the face based on the average face. Note that these existing models have seriously destroyed the availability of data, which limits their application range, especially for video surveillance system.In recent years, with superior performance achieved by deep learning in image synthesis, facial anonymization researches focus on the methods based on generative models. These algorithms produce random face images to replace the original ones, which ignored the personal identification information. However, Under surveillance scenarios, we require an individual facial protection model to protect the privacy information and the face should be traceable.

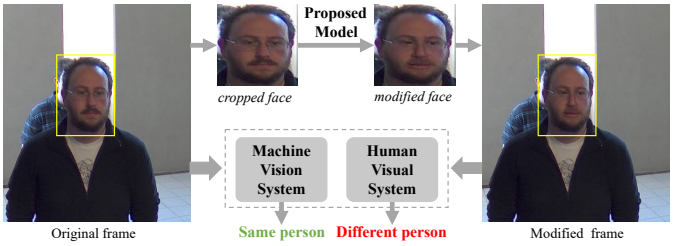

Fig. 1. Identification-preserved face camouflage for surveillance video. The left image is the original frame from surveillance video, and the right image is the modified frame by replacing the target face. Moreover, these two images are identified to be the same person with the facial recognition system, while we can observe their visual differences. The source image is downloaded from the literature [5].

To that end, we need to develop a novel facial anonymization model, which should be preserve the identity of the face. However, the identity preserving face anonymization under the surveillance scenarios is still a challenge task. First, there exist several types of external factors, e.g. pose, emotion, illumination, and even background. While traditional models on this task focus on changing the specific face regions without considering these external conditions. Then, a face may show different scales, which require the face anonymization model to generate the consistency results. Last but not the last, there exists a gap between the human perception and machine recognition system. We need to find the balance between personal information protection and data usability.

To address the issues above, in this paper, we present a new face camouflage model, which aims to maximally decorrelate individuals by human vision while the face identification

978-1-7281-8808-9/20/$31.00 ©2020 IEEE

2020 25th International Conference on Pattern Recognition (ICPR) | 978-1-7281-8808-9/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICPR48806.2021.9412636

system can be recognized. For example, Figure 1 shows that we first extract the face from the surveillance video, and then generate a new face based on our model, which is used to replace the original face to protect his privacy. This is a new type of face privacy protection method, which enables the anonymous faces to retain biometric identification information but have significant visual differences.To achieve this, we develop an encoder-decoder network architecture that can separately disentangle the person feature representation into an appearance code and an identification code. Specifically, we first randomly divide the face image into two groups, the source set and the target set, where the source set is used to extract the identity code and the target set provides the appearance code. We defined identification vector that mostly encodes biometric identification information related semantics, and the face contour, background. While the appearance vector is defined as that captures any other features, e.g., skin color, hair, eyebrow, nose. Then, we recombine the identity and appearance codes to synthesize a new face, which has the same identity with the source subject. Finally, the synthesized faces are used to replace the original face to protect the privacy of individual. Furthermore, our model is trained end-to-end with a multi-task loss function, which can better preserve the identity and stabilize. Experiments conducted on Cross-Age Celebrity dataset demonstrate the effectiveness of our model and validate our superiority in terms of visual quality and scalability.

The contributions of this paper are highlighted below: i) We introduce a new concept and method of face privacy, which distinguishes human perception from machine recognition. The proposed method can change the appearance of face while persevering the identification information. ii) We develop a new face privacy protection model for video surveillance, which can learn the explicit and complementary appearance and identification features to generate high-quality face images. Moreover, our model could be trained end-to-end with only identity labels. iii) We have conducted several experiments to validate the effectiveness of our model, and the experimental results demonstrate that our model could achieve a better performance on face camouflage.

II. Related work

A. Face privacy protection

The development of the face protection technology has experienced 3 phases, ranging from naive transformation type, k-same type, and face replacement type. Additionally, encryption technology also be applied to face protection, which intend to protect regions of interest (ROI) by using the scalability provisions of the used codec.

Earlier works on face protection involved applying naive transformations, known as the ad-hoc approaches [1]–[3]. These approaches are the most common use in our daily life, such as Map Street View, Television news, which obfuscating sensitive information with masking or pixelization or blurring. The simple and direct occlusion method seriously harms the data's availability. And it has been shown to be identified with face-recognition software [6].

Another representative anonymous methods, which called k-Same family, are based on the k-anonymity frame [7]. These algorithms exploit face pixel information or face feature from a set of k closet facial images to obtain the average face. In this way, the query face is anonymized among at lease k candidates, which guarantee the face recognition accuracy is below 1/k. To preserve face attributes like gender and expression, the k-Same-Select algorithm [8] divided the face set into different categories before the application of original k-same algorithm. The k-same-M algorithm [6] applied the k-same algorithm to Active Appearance Models (AAMs), in this way attempts to generate more naturalness of de-identified faces. In [9], [10], the use of k-furthest faces from the gallery to generate average faces is presented, which obviously reduced the dependence of k value. With the improvement of generation model, a generative neural network was used to directly generates faces based on the cluster attributes in k-same-net [11], [12]. These methods focus on face protection and have notable limitations on face images' usability.

More recently, deep neural networks have been used for face anonymous. In particular, the generative adversary networks (GANs) [13] inspire a new vein of face anonymized methods. The dedicated GAN is used in [14] to synthesize full-body images for de-identification, and the face areas is randomly generated. The work of [15] use GAN-based head inpainting technique to generate obscured faces, the latter can ensure privacy-sensitive information is thoroughly removed from the original face. Original faces are replaced in [16], [17] using GAN-based automatic encoder, which preserving certain attribute recognition. [18] present a face attribute transfer model to generate the de-identificated face. In the GAN-based methods of [19]–[21], the face with similar appearance to the person was generated. [22]–[24] extract face attributes of the input face, and then partial attributes are selected to synthesize the anonymous face. These methods aim to minimize identity information, while preserving expression, gender, race, or other facial attributes.

Furthermore, the method generated adversarial face images by operating on the image spatial domain [25]. In work [26], [27], Mirjalili et al. attempted to build Semi-Adversarial Network (SAN) to synthesize face images with obfuscating the gender information, while preserve the matching utility. In comparison to our work, they can only obtain identity-preserving face images in case of gender perturbations, the diversity of anonymity is limited (which would be critical), and the results are not as natural as ours.

B. Face generation

With the rapid development of deep generative models, the generated faces can be more realistic-looking and natural. Commonly used face generation models include Variational Auto-encoders (VAE) [28], Generative Adversarial Networks (GAN) [13], and Many variants of GANs. Recent advances such as DCGAN [29], Progressive GAN [30], StyleGAN [31]

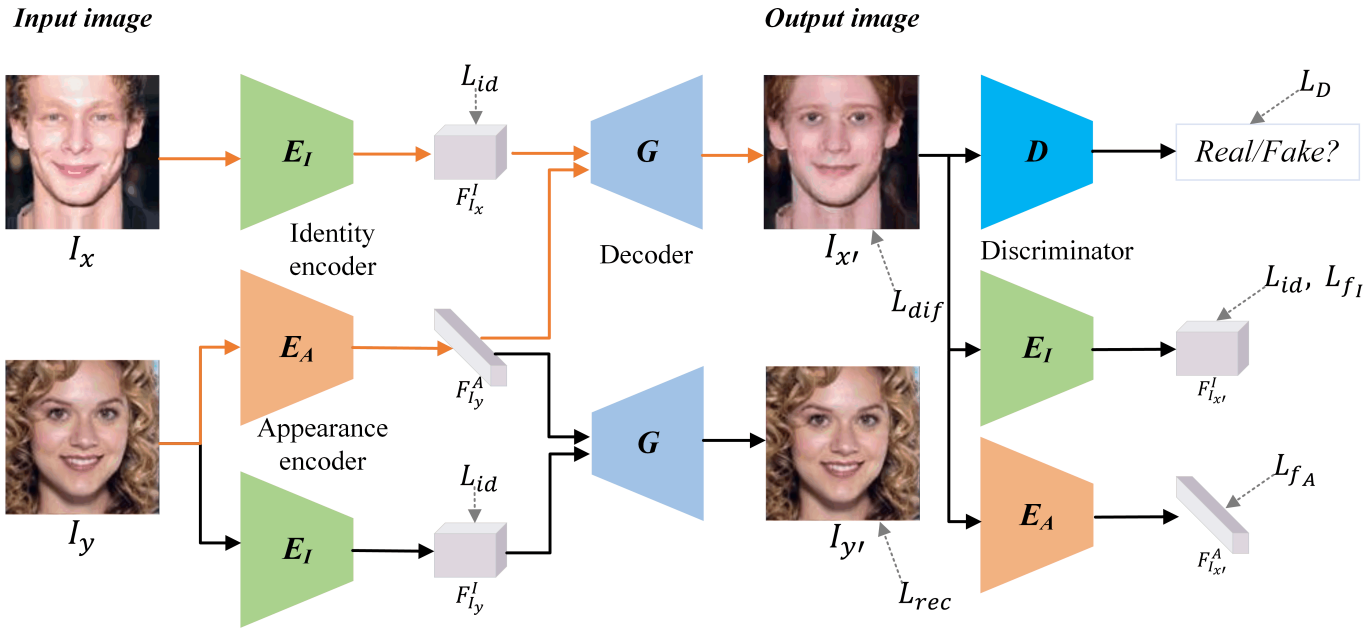

Fig. 2. A schematic overview of individual face protection model. We first disentangle identity and appearance code from the input face images, and then train the individual face protection model by using the self-identity face generation and cross-identity face generation. In the test phase, we only used networks following the orange arrow to generate the privacy protected face image. The loss functions are drawn indicated by dotted arrows.

and CycleGAN [32] have improved the stability of training and the quality of synthesized face images. Meanwhile, many works [33]–[36] have been proposed to generate controllable face images by using image transformation or semantic editing. These methods have separated the facial attributes of the input face, and the generators synthesize the face by adopting the target attributes. Face synthesis methods in [33]–[35] aim to generate face images of the same person with multiple expressions, attributes, diversity and so on. Different from those generate the same face, we apply the identification codes in the latent vector to manipulate the original face can only be recognized by the face recognition system, but not by humans.

III. Method

A. Overview

Given an unprotected face image, our goal is to learn a face modifier that significantly change each face's appearance while retaining their identification. In this work, we disentangled facial attributes into two parts: identification and appearance. The identification attributes that mostly encodes biometric attributes related semantics, and the face contour, background. The appearance attributes that captures any other features, e.g., skin color, hair, eyebrow, nose etc. Our proposed encoder-decoder network architecture is based on generative adversarial network as shown in Figure 2, which is composed of four modules: i) An identity encoder E_I is used to extract the identification features; ii) An appearance encoder E_A is used to extract the appearance features; iii) A generator G is developed to generate the disguised face with the latent codes; iv) A discriminator D is introduced to distinguish between generated images and real ones.

Specifically, given a source face image $\mathbf{I_x}$ and a random face $\mathbf{I_y}$, we first extract the identity representation $\mathbf{F^I_{I_x}} \in \mathbb{R}^{h\times w\times C_s}$ by using E_I from the source image $\mathbf{I_x}$, where h,w and C_s indicate the with, height and channel, respectively. And, the appearance code $\mathbf{F^A_{I_y}} \in \mathbb{R}^{1\times C_a}$ is extracted based on E_A from the appearance image $\mathbf{I_y}$, where C_a is the dimension of appearance code. After that, G takes both latent codes $\mathbf{F^I_{I_x}}$ and $\mathbf{F^A_{I_y}}$ for producing a new face $\mathbf{I_{x'}}$. Additional discriminators and loss functions are deployed for allowing the learning of the above four modules. In training phase, we utilize two kinds of mapping: self-identity face generation and cross-identity face generation. The former can be considered as an automatic encoder, which optimize the generator. The latter can effectively control the direction of generated images and improve the quality of generated images. The details of our proposed network architecture and learning strategies will be discussed in the following subsections.

B. Identity Preserving Face Generation

In this section, we discuss the processing of disentangling the identity code and the appearance code using the feature encoder E_I and E_A, respectively. Extracting identification information from the input face is a supervised training process. Specifically, we chose a group of face images labeled with $(\mathbf{I_i}, l_i)$, where l_i is the identity label of input face $\mathbf{I_i}$. Then, a pre-trained neural network is developed to train a face recognition system. Thus, we use the softmax loss as our identification loss:

$$L_{id} = E(-log(p(l_i|\mathbf{I}_i))), \tag{1}$$

where $p(l_i|\mathbf{I}_i)$ indicates the probability of the face $\mathbf{I_i}$ belonging to the class l_i. Consider the goal of our task, beside the identity information we also need to preserve the spatial face structure including the facial contour and background. Thus, we use the last convolution layer of backbone as the identity feature representation, whose dimension is $h \times w \times C_s$.

Since it is a difficult task to obtain the appearance label for each training face, we can not conduct the appearance encoder E_A as the identify encoder E_I. Moreover, we also need to remove the identity features from the appearance code which should only contain the semantic descriptions on the face. Inspired by recent work on generate adversarial network [32], [37], we use an unsupervised manner to optimize the neural network by introducing image and feature reconstruction loss. In details, we set a self-identity face generation module in the training stage, whose goal is to reconstruct the original face by feeding the identity and appearance code of the same person as shown in Fig. 2. And we extract the features of the last second full connected layer as the appearance code $\mathbf{F}^{\mathbf{A}}_{*}$, whose dimension is $1 \times C_a$. This face reconstruction could be seen as an traditional auto-encoder, which plays an important role in constraining the whole generator. We define the image reconstruction loss function as:

$$L_{rec} = E(||\mathbf{I_y} - G(\mathbf{F}^{\mathbf{I}}_{\mathbf{I_y}}, \mathbf{F}^{\mathbf{A}}_{\mathbf{I_y}})||^2), \tag{2}$$

where G is the face generator.

In the cross-identity face generation, the input identification features and appearance features come from different individuals. Hence, synthetic face images have no reference. To improve the controllable of this model, we introduce the feature reconstruction loss defined as the Euclidean distance between the feature codes in the latent space. These perceptual losses urge the generated images close to the features of input face in the same feature space. Thus, we extract two latent codes from the generated images, and employ the feature reconstruction loss L_{f_I} and L_{f_A} to obtain the consistency feature representation.

$$L_{f_I} = E(||\mathbf{F}^{\mathbf{I}}_{\mathbf{I_x}} - \mathbf{F}^{\mathbf{I}}_{\mathbf{I_{x'}}}||^2), \tag{3}$$

$$L_{f_A} = E(||\mathbf{F}^{\mathbf{A}}_{\mathbf{I_y}} - \mathbf{F}^{\mathbf{A}}_{\mathbf{I_{x'}}}||^2), \tag{4}$$

This feature reconstruction loss could be seen as a regular, which can guarantee the appearance to be transferred on to the synthesis face.

Moreover, to strengthen the visual difference between the generated face and the original face, a reconstruction loss term is used. L_{dif} measures the pixel-wise dissimilarity between the input face $\mathbf{I_x}$ and the output face $\mathbf{I_{x'}}$from the cross-identity face generation, which is used to enforce the reconstructed face away from the original face is far in pixel space:

$$L_{dif} = -E(||\mathbf{I_x} - G(\mathbf{F}^{\mathbf{I}}_{\mathbf{I_x}}, \mathbf{F}^{\mathbf{A}}_{\mathbf{I_y}})||^2), \tag{5}$$

For the face generator G, the extracted identity and appearance are used to generate the new face. Similarly with the traditional GAN, we use the adversarial loss to match the distribution of generated images to the real data distribution:

$$L_D = E(log(D(\mathbf{I_y})) + log(1 - D(G(\mathbf{F}^{\mathbf{I}}_{\mathbf{I_x}}, \mathbf{F}^{\mathbf{A}}_{\mathbf{I_y}})))), \tag{6}$$

C. Full Objectives

For jointly training all networks in the framework, we use the final synthesis loss function L_{tota}, which is a weighted sum of multiple parts:

$$L_{tota} = \lambda_1 L_D + \lambda_2 L_{id} + \lambda_3 L_{rec} + \lambda_4 L_{f_I} + \lambda_5 L_{f_A} + \lambda_6 L_{dif}, \tag{7}$$

where L_D is the discriminator's loss, L_{id} is the identity loss, L_{rec} is the reconstruction loss for the self-identity face generation, L_{f_I} and L_{f_A} are the feature reconstruction losses applied to the cross-identity generated image, L_{dif} is the pixel-wise loss. λ_i are weights of related loss items to control the importance. Since we only need the identity label of input face, the whole framework could be trained end-to-end.

IV. Experiments

A. Dataset

We adopt the Cross-Age Celebrity Dataset (CACD) [38] to evaluate our proposed face camouflage model. CACD dataset is the largest public cross-age database collected from the Internet Movie DataBase (IMDB). Moreover, it contains 163,446 color images of 2,000 celebrities with age annotations ranging from 14 to 62 years old. As described by previous works [39]–[41], they usually divide the images into different number of groups according to the age. Similarly, we use the same setting to separate the images into 5 groups: 11-20, 21-30, 31-40, 41-50, and 50+. To train our model, we randomly select 80% of the images from all the groups as the training set, and then the rest 20% as the testing set.

B. Experimental setup and evaluation metrics

Since our task involves two subtasks, which are the quality of generated face images and the ability of preserving person identification. To evaluate the quality of the generated images, fidelity is usually employed as the measurement for any image generation task. We adopt a common and effective metric Frechet Inception Distance (FID) [42], [43] to evaluate the quality of our generated face camouflage results. Moreover, the lower FID indicates the better of the generation. Furthermore, Structural Similarity (SSIM) [44] is also introduced to measure the structural similarity between generated and real images.

To validate the ability of identity preservation, we employ the face verification to evaluate the identity preservation of generated results. The distances between the test images and the generated faces are computed to measure their similarity. And the accuracy is employed as the evaluation criterion.

C. Implementation details

Our proposed neural network is implemented based on PyTorch. Specifically, the appearance encoder E_a is constructed based on ResNet-50 [45], which is pre-trained on

imagenet [46]. Instead of using all the layers in ResNet-50 [45], we remove the global average pooling layer and the fully connected layer, and then an adaptive max pooling layer is added to output the final appearance representations, whose size is $2048 \times 4 \times 1$. While E_I is composed of four convolutional layers followed by four residual blocks [45] and the size of structure representations is set to $128 \times 64 \times 32$. For the image generator G, it consists of four residual blocks followed by two adaptive instance normalization layers and four convolutional layers. The discriminator D is developed following the popular conditional adversarial networks [47], whose scales of the input image are set to: 28×28, 56×56, and 112×112. In the step of training, we first resize all the input images into 112×112. For the identification persevering model, it has been pre-trained based on the training data. And E_a is used SGD to optimize with learning rate 0.0001 and momentum 0.9. While we employ Adam [48] to optimize the structure encoder E_I, the generator G, and the discriminator D, whose learning rate is set to 0.0001.

Since face images are captured under unconstrained conditions, we first detect the face, and then align them to train a deep convolutional neural network for our face representation. To that end, we extract 68 facial landmarks from each face using the ensemble of regression trees method [49]. Based on the detected face key points, image normalization is performed to align the training faces. Specifically, the face is rotated in the image plane to make it upright based on the detected eye lines. Then, we find the center point on the face by taking the mid-point between the leftmost and the rightmost landmarks, while the center points of eyes and mouth are localized by averaging all the landmarks on the eyes and mouth regions, respectively. We then translate along the x-axis based on center points. To fix the face location along the y-axis, we require the eye line to be placed at 45% of image height from the top of the image, the mouth line is placed at 25% of image height from the bottom of the image, the aligned image is scaled to 128×128, and the center 112×112 region is the final normalized image.

D. Qualitative Analysis

To demonstrate the high quality of generated images and the robustness of our introduced framework, we conduct a series of qualitative experiments on CACD datasets. Moreover, we develop a baseline model, which is removed the identity supervision in the training process. The experimental results are shown in Figure 3, we randomly choose sixteen identities with different views, expressions, and ages as the source images from the training dataset, and randomly select eight different persons as the reference images. After that, our aim is to generate a new face referring the given reference images, which has the same identification with the source image but shows the difference appearances. The results show that our proposed model can preserve most of the identification related appearances and reenact a salient feature different face images. From the experimental results, we can observe that the generated faces of our method are photo-realistic and identification preserving, where the facial expression and global structure are consist with the source images. While our model focuses on altering the salient facial regions, e.g. the lip (the column (l)), noise (the column (j)), and eyes (the column (k)). When there exists an age gap between the source image and reference images, our introduced model could transfer these age-related facial features to the generated images.

Furthermore, to validate the robustness and generalization of our model, we conduct experiments by generalizing our proposed model to unknown identifications. Unknown identification is defined as that the identification of source images is not observed during training phase. Under this experiment circumstance, our face camouflage model is still able to generate the high-quality face images for these unseen identification. Some experimental results are shown in Figure 4, the first column lists the reference image, and the first row shows the source images. We can observe that the face appearance from unknown identification is altered while the source identities are still well preserved. The experimental results demonstrate that our model can be expanded to unseen identities, which makes it more practical in the real-word applications.

E. Quantitative Analysis

As the complementary evaluation for the qualitative results, we also introduce the quantitative analysis to demonstrate the effectiveness of our model. In this subsection, Frechet Inception Distance (FID) [42], [43] and Structural SIMilarity (SSIM) [44] is employed to evaluate the realism and diversity of generated images. Specifically, FID is used to measure the distance between the distribution of generated images and the real images. This score is sensitive to visual artifacts, which can be used to indicates the realism of generated images. Moreover, SSIM is used to compute the intra-class similarity, which can reflect the diversity of generation images.

To conduct the experiments, we randomly select one image for each celebrity as the source image, and then another five images with different identification is chosen as the reference images. Thus, we will achieve 5 generated images for each identification. For FID, we first compute the distances between the source images and the reference images, and then we evaluate how close between the generated images and the original inputs. We also introduce two recent face generation models PA-GAN [40] and IPCGANs [39] as the baselines. Similarly, we use the same setting to compute the SSIM. The experimental results are shown in Table I, our method has achieved a significantly improvement than other methods on both realism and diversity. Note that our model achieve a higher SSIM score, which is due that our model can capture the appearance features of various reference images.

F. Identification Preservation

Since one goal of our model is to preserve the identification information of generated images. To evaluate the performance of our model, we conduct experiments from two aspects: The first aspect is $1:1$ face verification, which is developed based on the predefined face pairs. The other aspect is $1:N$ face

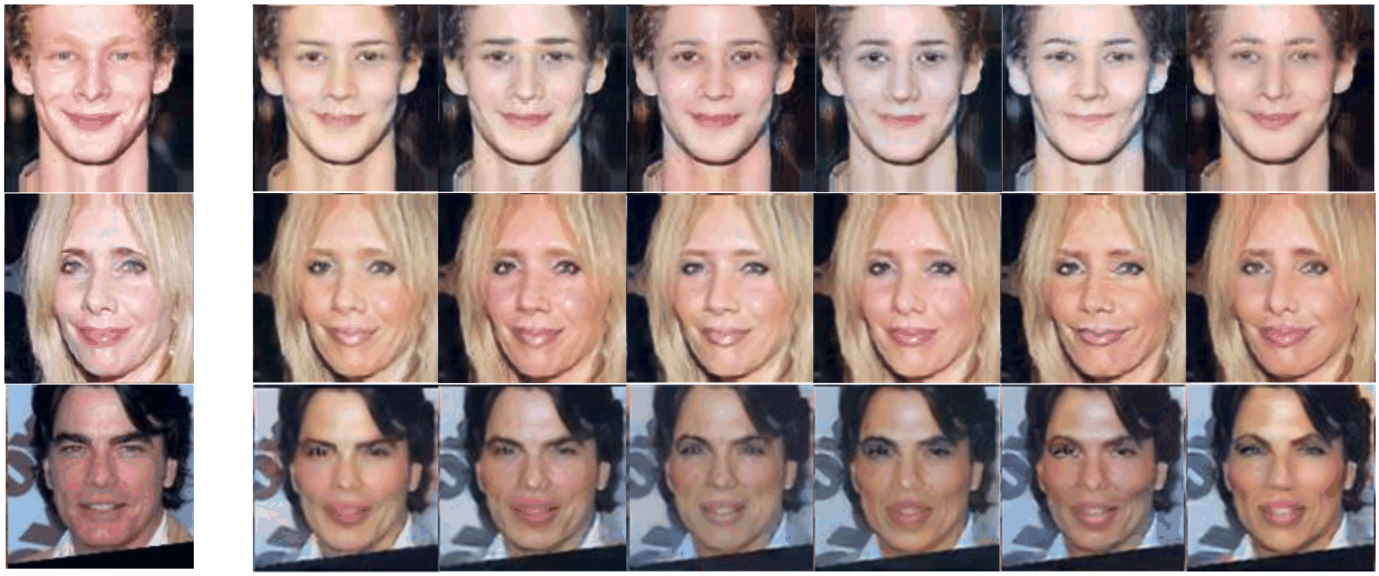

Fig. 3. Examples of our generated images on the CACD datasets. All images are sampled from the test sets. The first column indicates the source images which provides the identity information, while the left rows show the generated new face referencing with different appearance code.

Fig. 4. Examples of generated faces. The first and third row indicate the source images which provides the identity information, while the left rows show the generated new face referencing with the same appearance code.

TABLE I
COMPARISON RESULTS OF FID (LOWER IS BETTER) AND SSIM (HIGHER IS BETTER) TO EVALUATE REALISM AND DIVERSITY OF THE REAL AND GENERATED IMAGES ON CACD.

Methods	Realism (FID)	Diversity (SSIM)
Original real images	11.86	0.37
PA-GAN [40]	138.21	0.11
IPCGANs [39]	87.16	0.12
our model (appearance)	70.68	0.19
our model (identification)	71.23	0.29

identification, which requires to recognize the inputs from N predefined identifications.

TABLE II
COMPARISON RESULTS OF FACE VERIFICATION TO EVALUATE THE IDENTIFICATION PRESERVING OF GENERATED IMAGES ON CACD.

Methods	Face Verification	Face Identification
Original Real Images	97.05%	76.41%
Human	85.75%	-
our model (w/o)	82.63%	42.43%
our model	92.63%	85.64%

For face verification, we conduct the verifications between the test images and the generated faces. To that end, we adopt an open-source face analysis tool InsightFace [50] to achieve the verification scores and the threshold is set as 0.9247 (@FAR=1e-5). Specifically, for each input test image, we generate 10 images given reference images with distinct identification labels. After that, we randomly select the faces to develop the positive face pair (same identification) and negative pair (different identifications). There are $4,000$ positive face pairs and $4,000$ negative face pairs. The average verification rates are shown in Table II. Two baselines are introduced to validate the performance of our model, the original indicates that we generate the face pairs based on the real images without using the generated face. And "our model (w/o) " denotes that we remove the identification persevering module from our model. We can observe that our model achieves satisfactory average face verification rate demonstrating the ability to preserve the identification feature while altering the appearance of faces.

For face identification, we compute the distance between the generated images and the gallery images, and then we transfer the label of nearest instance to the query images. In detail, we randomly choose 5 generated images for each identification, and all the training images and the rest generated images are treated as the gallery. As we can see that our proposed model

shows its ability to preserve the identification. The superior performance of our model is because of adding the generated images into the gallery.

G. User study

To evaluate the quality of generated faces from the perspective of human perception, we employ 25 volunteers to evaluate the generated results of our model. Specifically, 200 random face pairs (100 positive pairs and 100 negative pairs) composed of real and generated images are chosen as input. Then, each volunteer is required to judge whether the face pair is the same person. After that, we compute the average accuracy over all volunteers to measure the quality of generated images, as shown in Table II.

V. Conclusion

In this paper, we explores a new problem which is to protect the privacy of face images in public video surveillance, while preserving the face identity. To that end, we propose an identification-preserved camouflage algorithm. Our approach combines the appearance distance and the identification similarity between the anonymous face and the original one, and realizes the purpose of recognition by machine. Experiments on the face generation and recognition show that the proposed face anonymization method can preserve the identification, and obtain the good quality of generated faces. In addition, our approach allows users to choose any appearances for face protection. In the future, we will focus on generating more realism faces.

VI. Acknowledgements

Supported by the National Key R&D Program of China under Grant 2018YFB0803701, National Natural Science Foundation of China (No.U1936210, U1736219, 61971016), Beijing Natural Science Foundation (No.4202084), The Open Research Fund from Shenzhen Research Institute of Big Data, under Grant No.2019ORF01010, Peng Cheng Laboratory Project of Guangdong Province PCL2018KP004.

References

[1] J. Coutaz, J. Crowley, and F. Berard, "Things that see: Machine perception for human computer interaction," *Communications of the ACM*, vol. 43, no. 3, pp. 54–64, 2000. 1, 2

[2] C. G. Neustaedter and S. Greenberg, *Balancing privacy and awareness in home media spaces*. Citeseer, 2003. 1, 2

[3] 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*, 2000, pp. 1–10. 1, 2

[4] E. M. 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. 1

[5] Y. Wong, S. Chen, S. Mau, C. Sanderson, and B. C. Lovell, "Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition," in *IEEE Biometrics Workshop, Computer Vision and Pattern Recognition (CVPR) Workshops*. IEEE, June 2011, pp. 81–88. 1

[6] R. Gross, L. Sweeney, F. De la Torre, and S. Baker, "Model-based face de-identification," in *2006 Conference on computer vision and pattern recognition workshop (CVPRW'06)*. IEEE, 2006, pp. 161–161. 2

[7] L. Sweeney, "k-anonymity: A model for protecting privacy," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 05, pp. 557–570, 2002. 2

[8] R. Gross, E. Airoldi, B. Malin, and L. Sweeney, "Integrating utility into face de-identification," in *International Workshop on Privacy Enhancing Technologies*. Springer, 2005, pp. 227–242. 2

[9] L. Meng and Z. Sun, "Face de-identification with perfect privacy protection," in *2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. IEEE, 2014, pp. 1234–1239. 2

[10] L. Meng, Z. Sun, A. Ariyaeeinia, and K. L. Bennett, "Retaining expressions on de-identified faces," in *2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. IEEE, 2014, pp. 1252–1257. 2

[11] B. Meden, Z. Emersic, V. Struc, and P. Peer, "κ-same-net: Neural-network-based face deidentification," in *2017 International Conference and Workshop on Bioinspired Intelligence (IWOBI)*. IEEE, 2017, pp. 1–7. 2

[12] 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. 2

[13] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680. 2

[14] K. Brkic, I. Sikiric, T. Hrkac, and Z. Kalafatic, "I know that person: Generative full body and face de-identification of people in images," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2017, pp. 1319–1328. 2

[15] Q. Sun, L. Ma, S. Joon 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*, 2018, pp. 5050–5059. 2

[16] J. Chen, J. Konrad, and P. Ishwar, "Vgan-based image representation learning for privacy-preserving facial expression recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 1570–1579. 2

[17] Z. Ren, Y. Jae Lee, and M. S. Ryoo, "Learning to anonymize faces for privacy preserving action detection," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 620–636. 2

[18] Y. Li and S. Lyu, "De-identification without losing faces," in *Proceedings of the ACM Workshop on Information Hiding and Multimedia Security*, 2019, pp. 83–88. 2

[19] 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. 2

[20] O. Gafni, L. Wolf, and Y. Taigman, "Live face de-identification in video," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 9378–9387. 2

[21] J. Song, Y. Jin, Y. Li, and C. Lang, "Learning structural similarity with evolutionary-gan: A new face de-identification method," in *2019 6th International Conference on Behavioral, Economic and Socio-Cultural Computing (BESC)*, 2019, pp. 1–6. 2

[22] H. Hao, D. Güera, A. R. Reibman, and E. J. Delp, "A utility-preserving gan for face obscuration," *arXiv preprint arXiv:1906.11979*, 2019. 2

[23] T. Li and L. Lin, "Anonymousnet: Natural face de-identification with measurable privacy," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 0–0. 2

[24] S. Shirai and J. Whitehill, "Privacy-preserving annotation of face images through attribute-preserving face synthesis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 0–0. 2

[25] E. Chatzikyriakidis, C. Papaioannidis, and I. Pitas, "Adversarial face de-identification," in *2019 IEEE International Conference on Image Processing (ICIP)*, 2019, pp. 684–688. 2

[26] V. Mirjalili, S. Raschka, A. Namboodiri, and A. Ross, "Semi-adversarial networks: Convolutional autoencoders for imparting privacy to face images," in *2018 International Conference on Biometrics (ICB)*. IEEE, 2018, pp. 82–89. 2

[27] V. Mirjalili, S. Raschka, and A. Ross, "Flowsan: privacy-enhancing semi-adversarial networks to confound arbitrary face-based gender classifiers," *IEEE Access*, vol. 7, pp. 99 735–99 745, 2019. 2

[28] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013. 2

[29] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015. 2

[30] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of gans for improved quality, stability, and variation," *arXiv preprint arXiv:1710.10196*, 2017. 2
[31] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2019, pp. 4401–4410. 2
[32] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232. 3, 4
[33] J. Bao, D. Chen, F. Wen, H. Li, and G. Hua, "Towards open-set identity preserving face synthesis," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 6713–6722. 3
[34] Y. Shen, B. Zhou, P. Luo, and X. Tang, "Facefeat-gan: a two-stage approach for identity-preserving face synthesis," *arXiv preprint arXiv:1812.01288*, 2018. 3
[35] Y. Shen, P. Luo, J. Yan, X. Wang, and X. Tang, "Faceid-gan: Learning a symmetry three-player gan for identity-preserving face synthesis," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 821–830. 3
[36] T. Xiao, J. Hong, and J. Ma, "Elegant: Exchanging latent encodings with gan for transferring multiple face attributes," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 168–184. 3
[37] Z. Zheng, X. Yang, Z. Yu, L. Zheng, Y. Yang, and J. Kautz, "Joint discriminative and generative learning for person re-identification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2019, pp. 2138–2147. 4
[38] B.-C. Chen, C.-S. Chen, and W. H. Hsu, "Cross-age reference coding for age-invariant face recognition and retrieval," in *European conference on computer vision*, 2014, pp. 768–783. 4
[39] Z. Wang, X. Tang, W. Luo, and S. Gao, "Face aging with identity-preserved conditional generative adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7939–7947. 4, 5, 6
[40] H. Yang, D. Huang, Y. Wang, and A. K. Jain, "Learning face age progression: A pyramid architecture of gans," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 4, 5, 6
[41] Z. He, M. Kan, S. Shan, and X. Chen, "S2gan: Share aging factors across ages and share aging trends among individuals," in *The IEEE International Conference on Computer Vision (ICCV)*, 2019. 4
[42] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," in *Advances in Neural Information Processing Systems*, 2017, pp. 6626–6637. 4, 5
[43] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of gans for improved quality, stability, and variation," *arXiv preprint arXiv:1710.10196*, 2017. 4, 5
[44] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004. 4, 5
[45] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778. 4, 5
[46] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105. 4
[47] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134. 5
[48] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014. 5
[49] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1867–1874. 5
[50] J. Deng, J. Guo, X. Niannan, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *CVPR*, 2019. 6