Signal, Image and Video Processing
https://doi.org/10.1007/s11760-020-01819-9

ORIGINAL PAPER

Privacy preservation through facial de-identification with simultaneous emotion preservation

Ayush Agarwal[1] · Pratik Chattopadhyay[2] · Lipo Wang[3]

Received: 20 July 2020 / Revised: 11 October 2020 / Accepted: 10 November 2020

Abstract
Due to the availability of low-cost internet and other data transmission media, a high volume of multimedia data get shared very quickly. Often, the identity of individuals gets revealed through images or videos without their consent, which affects their privacy. Since face is the only biometric feature that reveals the most identifiable characteristics of a person in an image or a video frame, the need for the development of an effective face de-identification algorithm for privacy preservation cannot be over-emphasized. Existing solutions to face de-identification are either non-formal or are unable to obfuscate identifiable features completely. In this paper, we propose an automated face de-identification algorithm that takes as input a facial image and generates a new face that preserves the emotion and non-biometric facial attributes of a target face. We consider a proxy set of a large collection of artificial faces generated by StyleGAN and select the most appropriate face from the proxy set that has a facial expression and pose similar to that of the target face. The faces in the proxy set are artificially generated, and hence the face selected from this set is completely anonymous. To retain the non-biometric attributes of the target face, we employ a generative adversarial network (GAN) with a suitable loss function that fuses the non-biometric attributes of the target face with the face selected from the proxy set to obtain the final de-identified face. Experimental results emphasize the superiority of our approach over state-of-the-art face de-identification methods.

Keywords Privacy protection · Face de-identification · Emotion clusters · GAN · Mini-Xception

1 Introduction

Over the past few decades, with the advancement of sophisticated camera technologies as well as fast transmission media, a large amount of image/video data gets recorded and shared every day on social media. The development of sophisticated camera technology has made it easy to capture and analyze a high volume of image/video data with ease. Malicious viewers often extract sensitive information from these photos and videos and misuse these, which hurts the privacy of the subjects captured in the photos/videos. However, the privacy preservation methods currently used by most third-party digital data management systems are not effective enough, resulting in leaking of the identity information. Thus, the need for developing a proper approach to preserve the privacy of individuals captured in photos/videos cannot be over-emphasized.

Among the different biometric attributes, the face of a person is known to preserve significant identity information, and thus obfuscating facial identity features before sharing photos/videos over the internet (or other data transmission media) is of utmost necessity. De-identification is a technique to conceal the identity of a subject so that he/she cannot be identified by standard biometric identification mechanisms. Traditional ways of achieving privacy protection such as warping, blurring, pixelization, etc., are based on simple image processing techniques in which several other important non-biometric features, such as emotion and face style also get obscured, along with identity. Also, these approaches do not guarantee that the identity traits have been completely obfuscated. The *k*-same family of de-identification approaches developed later could successfully

✉ Pratik Chattopadhyay
pratik.cse@iitbhu.ac.in

Ayush Agarwal
ayushagarwal@mnnit.ac.in

Lipo Wang
elpwang@ntu.edu.sg

[1] Department of ECE, MNNIT, Allahabad 211004, India

[2] Department of CSE, IIT (BHU), Varanasi 221005, India

[3] School of EEE, NTU, Singapore 639798, Singapore

Published online: 27 November 2020

overcome the drawbacks of the traditional techniques. However, these methods generate de-identified faces by linearly aggregating features from multiple faces without incorporating any learning mechanisms. This causes the generated faces to bear ghostly appearances. Also, multiple input faces can get mapped to the same de-identified face. The advancement of deep neural network architectures has led to the development of a few automated learning-based face de-identification approaches such as [1,2].

In the present work also, we focus on face de-identification from images. Traditional approaches to face de-identification fail to obfuscate identity information at a high resolution, and the de-identified faces generated by these techniques bear significant identity similarity with respect to the original face or with some other faces in the real-world domain. We propose a learning-based algorithm to perform face de-identification by preserving emotion and non-biometric facial style features. The faces generated by our approach are completely anonymous, i.e., they do not have identity similarity with any real face. The main contributions of the proposed work are as follows:

- Ours is the first-ever approach to face de-identification that generates completely anonymous but real-looking de-identified faces.
- Despite generating anonymous faces, our approach is capable of retaining the emotional characteristics and non-biometric facial attributes of the input face at a high resolution, which existing techniques fail to achieve.
- Extensive experimental evaluation presented in the paper shows that our approach outperforms the state-of-the-art machine learning-based face de-identification techniques.

2 Related work

Existing work on face de-identification can be broadly categorized as non-formal methods, formal methods, and deep neural network-based approaches, which are discussed next.

2.1 Non-formal methods

These are the oldest privacy preservation techniques and are very simplistic, focusing mostly on the application of standard image processing techniques like blurring [3], warping [4,5], pixelation, etc. But none of these techniques obfuscate identity information suitably, and hence cannot guarantee complete anonymity. Also, the original face can be easily reconstructed from the output face by applying reverse transformation mechanisms, such as de-blurring and de-pixelation. The extent of blurring, or pixelation, or warping on an image has to be manually controlled and execution of the above transformation operations by a higher degree can obscure facial emotions, style, and other non-biometric features as well, which is not desirable. These factors limit the applicability of non-formal approaches for de-identification in practical situations where privacy preservation is a major concern.

2.2 Formal methods

Unlike the approaches discussed in Sect. 2.1, the formal methods can guarantee anonymity to a certain extent. In [6], Mosaddegh et al. described an approach for photo-realistic face de-identification that was based on borrowing different face features like chin, eyes, nose, etc., from a set of donors' faces and replacing the attributes of original faces with these borrowed ones. The work in [7] presented a reversible de-identification technique that can be used in conjugation with any obfuscation technique. Another popular category of face de-identification techniques was based on the k-same family of approaches, e.g., [8–10]. k-same computes a set of de-identified images in which each element indiscriminately relates to at least k elements of a set of person-specific images. However, the faces generated by this method appear to be ghostly due to the averaging of multiple faces. As an improvement, Active Appearance Model (AAM)-based face de-identification was introduced in [8], which has been seen to successfully eliminate the ghostly appearance on the de-identified face. The q-far de-identification approach proposed in [11] is also an AAM-based approach that incorporates an additional pose estimation step to align the faces to be aggregated. In recent years, several other versions of the k-same algorithm have been proposed such as [12–14]. A short-coming of each of the above techniques is that these are unable to generate unique anonymous de-identified faces for different subjects.

2.3 Deep neural approaches

To date, only a few deep learning-based face de-identification techniques exist in the literature. Among these, [15] proposed by Meden et al. is based on deep generative neural network termed as k-Same Net. The algorithm has three main components: (a) guaranteeing k-anonymity, (b) exploiting a proxy face dataset, and (c) employing a generative network for de-identification. As in the k-Same family of approaches, k-anonymity implies replacing k images in the test set by the same surrogate face from a proxy set. The centroid of the mapped cluster along with the input test face is passed through the generative neural network (GNN) to obtain the de-identified face. The drawback of this approach is that it often fails to retain non-biometric facial characteristics such as emotion, face style, etc. A similar approach was also proposed by the same authors in [16] in which instead of mapping to a cluster constructed from the proxy set, an input face was replaced by a different face from the proxy set, following which a GNN-based de-identification similar to that described in [15] has been performed. However, the approaches described in [16] are also not so effective for use in practical scenarios since the identity of the subject can be visually obtained without much effort even from his/her generated de-identified face. A similar conclusion follows for the work in [17], in which case also the identity of the subject can be visually determined easily. To improve over k-Same Net and other formal de-identification means described above, the privacy protective generative adversarial networks (PPGAN) was proposed in [1]. PPGAN leveraged conditional GAN (cGAN) along with two external modules: (a) a verificator with contrastive loss, and (b) a regulator to preserve a high degree of structural similarity between the input and the de-identified faces. The de-identified image generated by PPGAN is capable of retaining the emotion present in the input face, but its identity obfuscation quality is not appreciably good. In [18], Zhenliang He et al. proposed a generative network called AttributeGAN (abbreviated as AttGAN) that manipulates single or multiple attributes from a face image to generate a new face with desired attributes while preserving other facial details. Although this approach minimizes attribute loss to a certain extent, it requires a gallery dataset with facial images as well as the corresponding attribute annotations, which is not practically available always. An automated approach for deriving the attribute embeddings using neural networks seems to be more suitable for face de-identification with non-biometric attribute preservation.

In this work, we specifically focus on improving the existing solutions by considering a proxy set of anonymous faces generated by StyleGAN [19]. Given an input face, it is mapped to an appropriate face in the proxy set having similar emotion and pose, and next the non-biometric facial attributes of the input face are merged with the proxy face using a GAN. These non-biometric facial style features are obtained in an automated way by passing the face image through a pre-trained ResNet model. Although StyleGAN can potentially generate highly realistic anonymous faces, it is not suitable for performing face de-identification by preserving emotion characteristics. Secondly, due to the presence of a large number of layers in the StyleGAN network, its response time is significantly high. Owing to the above reasons, we have not used StyleGAN directly as our generating network. Instead, we employ a separate deep network, namely, the mini-Xception network, to extract the desired emotion features from a given face. Next, we use the anonymous proxy set of StyleGAN-generated face repository to select a face with similar pose and emotion as that of the input face. Choosing a proxy face with a similar pose reduces the ghostly appearance on the de-identified face, which is one of the main drawbacks of the k-Same family of de-identification approaches. The individual steps of our algorithm are explained in detail in Sect. 3.

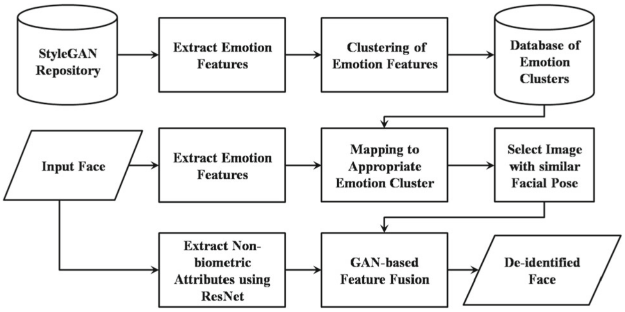
StyleGAN Repository
Extract Emotion Features
Clustering of Emotion Features
Database of Emotion Clusters
Input Face
Extract Emotion Features
Mapping to Appropriate Emotion Cluster
Select Image with similar Facial Pose
Extract Non-biometric Attributes using ResNet
GAN-based Feature Fusion
De-identified Face

Fig. 1 Block diagram of our proposed approach

3 Proposed approach

Face de-identification can be considered as a transformation of an input face image F to another face image $\hat{F}$ using some mapping function, say $\mathbb{T}$. Mathematically, $\mathbb{T}(F) = \hat{F}$, and the function $\mathbb{T}$ must be non-invertible, i.e., $\mathbb{T}^{-1}(\hat{F}) \neq F$. The present problem can be viewed as an application of image-to-image translation, in which finding a suitable transformation function ($\mathbb{T}$) is the major challenge. In recent years, generative neural networks (GNNs) are being extensively used as function approximators for generative modeling in several tasks, such as image-to-image translation [20], style transfer [21,22], image colorization [23], etc. Hence, we also use a conditional generative adversarial network (cGAN) [20], a variant of GNN, in this work. We consider an anonymous face dataset G that consists of faces with all possible emotions as a proxy dataset. A fixed number of emotion clusters is first extracted from this proxy dataset by applying K-means clustering. The optimal number of clusters in the clustering process is determined by plotting an elbow curve (to be discussed in Sect. 4). On completion of the clustering phase, each cluster corresponds to a particular type of facial emotion. Given a test face, we first map it to the appropriate emotion cluster, and next select a proxy face from this cluster possessing a facial pose similar to that of the input face. The repository of artificial faces generated by StyleGAN [19] has been used as a proxy set to determine the above emotion clusters. A block diagram of the overall face de-identification approach is shown in Fig. 1, and the individual steps are explained in the following sub-sections.

3.1 Extract emotion features

The pre-trained model of the mini-Xception convolutional neural network [24] has been used to extract the emotion features from any given face. To extract the emotion embedding vectors from each input face, we consider the 128-dimensional features generated at the fully-connected layer just before the final global average pooling and softmax classification layers of the mini-Xception network. Let us consider that the proxy set G consists of a total of N faces F_1^G, F_2^G, F_3^G, …, F_N^G, and $\mathcal{E}_i$ denotes the emotion feature for the i^{th} face in the repository. Further, let us assume that $\mathcal{E}$ denotes the set of emotion features corresponding to all the N faces, as obtained from the mini-Xception network. Thus, $\mathcal{E} = \{\mathcal{E}_1\ \mathcal{E}_2\ \ldots\mathcal{E}_N\}$. The next step is the clustering of the emotion embedding vectors in the set $\mathcal{E}$ into several emotion clusters, each representing a particular type of emotion. Instead of using the above emotion vectors directly for clustering, we project these into the LPP (Locality Preserving Projection) sub-space and perform the clustering in this transformed space. Projecting the original emotion feature vectors to the LPP sub-space helps in obtaining emotion vectors with reduced dimension, free from noise and redundant attributes [25,26]. We observe that reducing the mini-Xception generated feature dimension to 64 from the 128 can retain more than 90% of the variance present in the feature set $\mathcal{E}$.

3.2 Determining emotion clusters and mapping of facial image to appropriate cluster

The LPP-transformed emotion vectors are next clustered using K-means clustering in such a way that faces with similar emotion vectors are placed in the same group. The appropriate value of K to be used in the clustering is determined by plotting an elbow curve, to be discussed in detail in Sect. 4. Presently, let us assume that we have already determined the optimal value of K from the elbow curve to be used to perform the K-means clustering. The cluster centers corresponding to these clusters (say, $E_{C_1}^G$, $E_{C_2}^G$, …, $E_{C_K}^G$) depict the K unique emotions present in G. The LPP-reduced emotion vector E^T corresponding to any given test image (say, F^T) is next compared with each of the K emotion cluster centers, i.e., $E_{C_1}^G$, $E_{C_2}^G$, …, $E_{C_K}^G$, using a cosine similarity metric, and the best matching emotion cluster is considered for the next step of facial pose matching. If the winning cluster is denoted by C^*, and cos(*) represents the cosine operator, then

$$C^* = \text{argmax}_k cos(E^T, E_k^G), k \in \{C_1, C_2, \ldots, C_K\} \quad (1)$$

3.3 Facial pose matching

The faces images corresponding to the winning cluster C^* will have similar emotional characteristics, but the poses of the individual faces in this cluster are likely to be different from each other. But for generating realistic de-identified faces, the proxy face must have a pose similar to that of the input face. To accomplish this, we employ an additional neural model as discussed in [27]. This is a pre-trained two-layer neural network that predicts the roll, pitch, and yaw angles for any input face in real-time. Let us assume that C^* consists of M faces denoted by $F_1^{C^*}$, $F_2^{C^*}$, …, $F_M^{C^*}$, (where $M << N$). We construct a gallery of facial pose feature vectors for each of these M subjects by concatenating the three angles, i.e., roll, pitch, and yaw for each face, as predicted by the above-mentioned pre-trained network. Let $P_1^{C^*}$, $P_2^{C^*}$, …, $P_M^{C^*}$ denote the facial pose features corresponding to $F_1^{C^*}$, $F_2^{C^*}$, …, $F_M^{C^*}$, respectively, and P^T denotes the facial pose features corresponding to the test image F^T. The face with the best matching pose in cluster C^* is obtained by computing the cosine similarity between P^T and each of $P_m^{C^*}$ (m={1,2,…,M}), and selecting the face (say, F^S) that provides the highest similarity value. The selected face F^S thus has facial pose and emotional characteristics similar to that of the input test face F^T, and F^S will be used as the proxy face for the next step of GAN-based feature fusion. Dividing the proxy set into emotion clusters and mapping of the test face to the appropriate emotion cluster enables the facial pose matching stage to be carried out with only a small subset of the total number of faces in the proxy set, thereby saving significant processing time.

3.4 Non-biometric attribute extraction and GAN-based feature fusion

In this step, our objective is to transfer the non-biometric facial attributes from the input face F^T to F^S in such a way that the style transferred image remains un-identifiable, but it can preserve the emotion and pose of F^T. Let us denote this style-transferred image by F^D, which is also the final de-identified image. To obtain the desired facial attributes of the input face F^T, we employ a ResNet model pre-trained on VGG Face dataset [28] to generate facial attribute embedding or latent vector (denoted by z). This is next fused with F^S by employing a GAN architecture to generate the style transferred de-identified face F^D. A schematic diagram of the network used in our study is shown in Fig. 2.

We use a modified version of the U-net generator with the addition of attribute vector at the bottleneck, and PatchGAN discriminator with instance normalization. The model is trained by considering the following factors: (a) facial attribute loss between F^D and F^T should be as small as possible, (b) facial geometry of F^D and F^T should be similar to each other. To preserve the above conditions we use three loss functions during the training phase, namely, (a) L_{adv} which is the standard adversarial loss, (b) L_{mix} termed as

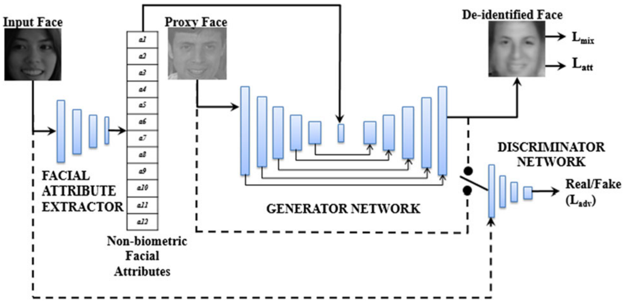

Fig. 2 Schematic diagram of the GAN network

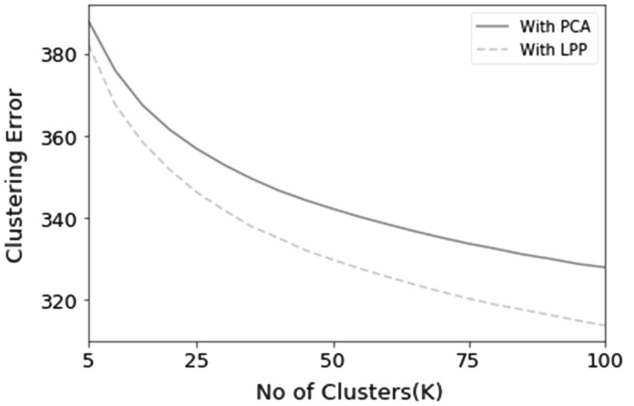

Fig. 3 Elbow curves corresponding to LPP-reduced and PCA-reduced emotion features

the mixing loss (or, reconstruction loss) and computed as the $L1$-norm of the difference between the input face image F^T and the generated de-identified face F^D, and (c) L_{att} termed as the attribute loss and computed as the $L1$-norm of the difference between the attributes of F^T and F^D extracted from the pre-trained ResNet model (refer to Fig. 2). All the above loss terms, i.e., L_{adv}, L_{mix}, and L_{att} are aggregated through weighted summation, and the overall loss is back-propagated through the GAN to adjust its weights. The StyleGAN generated face repository has been used to train this GAN model, and finally, its performance is evaluated on unseen face data. For each input image F^T in the training set, we find the facial attribute feature z and the corresponding proxy face F^S. The generator of the GAN performs a mapping of the form $G : (F^S, z) \mapsto F^D$, where G and D are the functions learned by the generator and the discriminator network, respectively. The three loss functions can be mathematically stated as follows:

$$L_{\text{adv}}(G, D) = \mathbf{E}_{F^T, F^S}[log(D(F^T, F^S))] + \mathbf{E}_{F^T, F^S, z}[log(1 - D(F^T, G(F^S, z)))], \tag{2}$$

$$L_{\text{mix}}(G) = \mathbf{E}_{F^T, F^S, z}[|| F^T - G(F^S, z) ||_{l_1}], \text{ and} \tag{3}$$

$$L_{\text{att}}(G) = \mathbf{E}_{z, \hat{z}}[|| z - \hat{z} ||_{l_1}]. \tag{4}$$

The overall loss L_{face} is thus computed as:

$$L_{\text{face}} = \lambda_1 L_{\text{adv}} + \lambda_2 L_{\text{mix}} + \lambda_3 L_{\text{att}}, \tag{5}$$

where λ_1, λ_2 and λ_3 are experimentally determined.

4 Experiments and results

Our algorithm has been implemented on a system having 64 GB RAM, one i9-18 core processor, and three GPUs: one Titan Xp with 12 GB RAM, 12 GB frame-buffer memory and 256 MB BAR1 memory, and two GeForce GTX 1080 Ti with 11 GB RAM, 11 GB frame-buffer memory and 256 MB of BAR1 memory. As discussed in Sect. 3, we use the StyleGAN generated face data repository as the proxy set of anonymous identities. This repository consists of 30,000 anonymous face images with mixed emotions. As discussed in Sect. 3.2. before computing the emotion clusters the emotion vectors provided by the mini-Xception network are projected to the LPP subspace. Since principal component analysis (PCA) is also a popularly used subspace projection technique, we study the benefits gained by employing LPP-reduced emotion features instead of PCA-reduced features in the subsequent clustering phase. Elbow curves presented in Fig. 3 show improvement in the clustering error as the number of clusters is increased from 5 to 95 in steps of 5, for both LPP and PCA. The clustering error is considered to be the mean of the intra-cluster distances of each point from its nearest cluster center.

It can be seen from the figure that the clustering error obtained from the LPP-reduced emotion features is always less than that obtained by the PCA-reduced features for any value of the number of clusters. This implies that the LPP-reduced emotion features result in obtaining more compact emotion clusters that are separated from each other. It can be also observed from the figure that for both the elbow curves, the rate of decrease of the clustering error is quite low as the number of clusters is increased beyond 50. Thus, from visual observation of Fig. 3, we choose the value of K in K-Means clustering to be 50. As explained before, these, 50 clusters represent 50 different facial emotions. The de-identified faces generated by the GAN after completion of the training phase are presented in the last row of Fig. 4. The first row of the same figure shows a sample of face images from the StyleGAN repository, while the second row corresponds to the proxy faces selected from the repository. Experimentally, we observe that setting the value of λ_1 to 1, λ_2 to 100, and λ_3 to 0.5 helps in carrying out effective style transfer and generation of realistic de-identified faces.

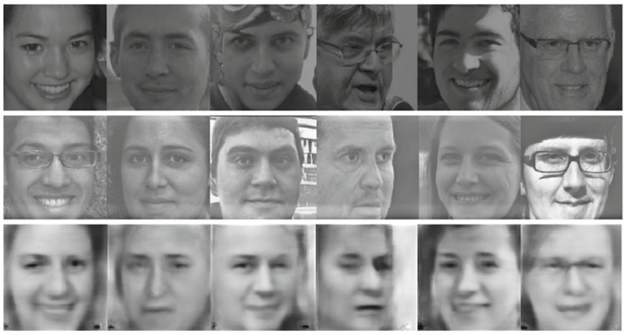

Fig. 4 1st row: input images from StyleGAN repository, 2nd row: selected proxy faces, 3rd row de-identified faces

It can be seen that the de-identified faces retain the non-biometric facial characteristics of the input faces and identity characteristics of the matched faces.

To verify the importance of the facial pose matching stage on the final de-identification result (refer to Sect. 3.3), we show results by carrying out face de-identification under the following two settings: (i) incorporating the facial pose matching phase, (ii) without incorporating the facial pose matching phase. For the second situation, we show results by selecting a proxy face from the mapped emotion cluster having the highest appearance disparity with respect to the F^T. This experiment is performed by training the GAN model on the StyleGAN data using the 50 emotion clusters obtained before, and testing it on the Radboud Faces Database (RaFD) [29], which is unknown to the trained model. This dataset consists of 8040 facial images from 67 subjects with varying emotions and facial poses. The subjects present in this data belong to either of the following five categories: Caucasian male, Caucasian female, Caucasian kid male, Caucasian kid female, and Moroccan Dutch male. The emotions present in RaFD corresponding to each subject are as follows: anger, disgust, fear, happiness, sadness, and neutral. Each emotion is accompanied by three gaze directions, and snapshots of each face are captured from five different viewpoints. Resulting images are shown in Fig. 5, in which the first column corresponds to a set of faces from the RaFD data, while the second and third columns respectively show the de-identified faces generated by the GAN for scenarios (i) and (ii).

From the figure, it can be seen that the de-identified faces look more realistic if the facial pose matching phase is employed before carrying out GAN-based feature fusion. Otherwise, the resulting faces bear a ghostly appearance.

We next make a comparative performance evaluation of our approach with other popular state-of-the-art machine learning-based face de-identification algorithms, namely, the K-SameNet [15] and PPGAN [1]. Three datasets have been used in this study, namely the RaFD [29], the XM2VTS [30], and the CelebA [31]. The XM2VTS data consists of frontal views of the faces of the 295 subjects recorded in four sessions, resulting in a total of 1180 face images. These images are captured at 720×576 pixel resolution and are stored in PPM format. The CelebA [31] is a large-scale face attributes data with 202,599 face images from 10,177 identities, and 40 attribute annotations corresponding to each face. The images in this data cover large pose variations and background clutter. In our next set of experiments, we consider all the images from the RaFD and X2MTS data, and randomly selected 2000 face images from the CelebA data for testing. Qualitative performance comparison of the different de-identification algorithms can be obtained from Fig. 6, which shows a set of faces arranged into rows and columns. Among these, the first four columns respectively correspond to inputs and outputs obtained from the RaFD data, while the next four columns correspond to the inputs and outputs obtained from the XM2VTS data, and the final four columns correspond to the inputs and outputs obtained from the CelebA data. For each of the datasets, the first column shows an input face, and the next three columns respectively present the de-identified faces generated by [1,15], and our proposed approach corresponding to the input face. From visual inspection, it can be seen that our approach performs the best among the three in terms of identity obfuscation for all the datasets used in the study. It also performs quite well in retaining the desired emotions on the generated faces. On the other hand, both PPGAN and K-SameNet fail to remove identity features properly, and hence, the emotion and pose on the generated faces look significantly similar to that of the input face. The effectiveness of our approach over other de-identification algorithms has been observed for the other images present in the three datasets as well.

We also perform a quantitative evaluation of the above de-identification approaches on the three different datasets,

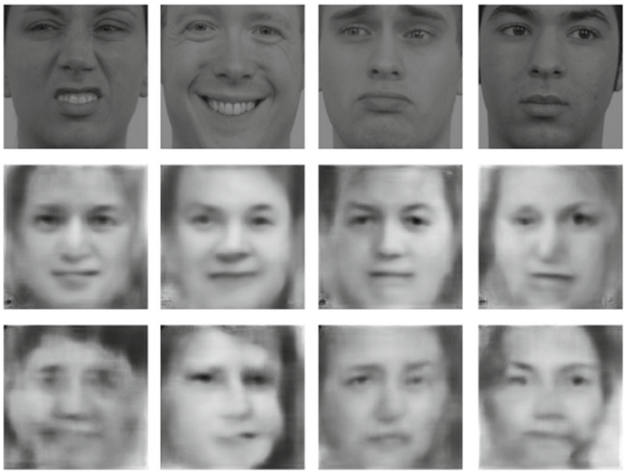

Fig. 5 Qualitative evaluation of our approach: First row corresponds to input images from RaFD data, second row corresponds to output with the facial pose matching step, third row corresponds to output without facial pose matching

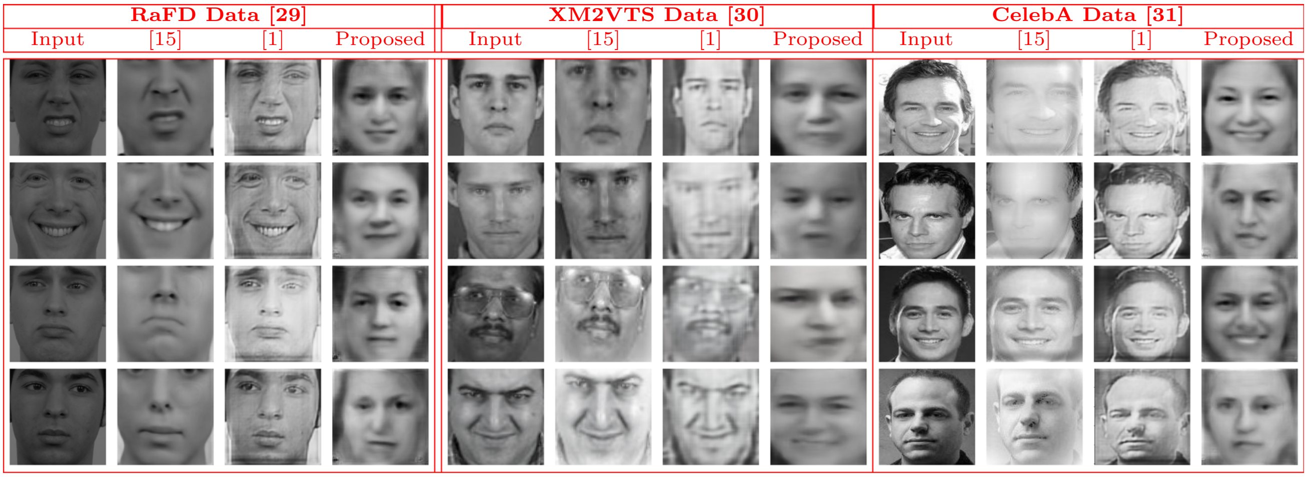

Fig. 6 Qualitative comparison of the performance of different de-identification algorithms on three different data sets

namely [29–31], and present the results in terms of two metrics, namely the *DMOS* and the *FMOS*. The parameter *DMOS* stands for *De-identification Mean Opinion Score* indicating how far a de-identification algorithm is successful in obfuscating the identity of an input face, without altering the emotional characteristics of the face. On the other hand, the parameter *FMOS* indicates *Facial Mean Opinion Score* which is a score depicting how realistic is the appearance of the face generated by a de-identification algorithm. To obtain these quantitative metrics, we conducted a survey in which 60 participants (from outside our research team) were given a set of five images from each of the RaFD, XM2VTS, and CelebA datasets along with the corresponding de-identified faces generated by [1,15], and the proposed approach. The participants were given a day time to provide two ratings for each de-identified face image in the range between 0 to 5. Out of these two ratings, one represents how effectively identity obfuscation has been performed, and the other represents how realistic the output image looks. In the above rating scheme, 0 means the worst possible score, and 5 means the best possible score. The average of the two types of ratings given by all the participants form the DMOS and FMOS, respectively. The corresponding scores obtained for the different algorithms for each dataset are presented in Table 1.

From the table, we observe that the proposed approach always outperforms both PPGAN and K-SameNet in terms of DMOS. The average DMOS given by our approach on the RaFD data is 3.97 which is remarkably better than the corresponding DMOS for [1,15], which are 2.15 and 0.91, respectively. Also, the DMOS of [1,15] on the XM2VTS and CelebA data are significantly lower than those obtained for our approach. While the FMOS of our approach is also higher than [1,15] for the XM2VTS and CelebA data to some extent, in the case of the RaFD data it is slightly less compared to the two other approaches. This is since both [1,15] retain significant appearance similarity with respect to the input face. But as seen from the figure, none of these state-of-the-art techniques can obfuscate identity information properly. Hence, in terms of emotion preservation and identity obfuscation, our method performs the best among the three used in the study.

Table 1 Quantitative evaluation of the different approaches on the two data sets in terms of DMOS and FMOS

Dataset	Evaluation metric	Method		
		[15]	[1]	Proposed
RaFD	DMOS	2.15	0.91	3.97
[29]	FMOS	4.05	4.51	3.31
XM2VTS	DMOS	1.69	2.30	4.10
[30]	FMOS	3.49	3.34	3.62
CelebA	DMOS	2.19	3.11	3.75
[31]	FMOS	2.69	3.21	3.69

5 Conclusions and future work

In this paper, we present an effective face de-identification algorithm that preserves facial pose and emotion. Generation of anonymous de-identified faces is guaranteed due to selecting a proxy face from a large collection of artificial faces having emotion and facial pose similar to that of the input face. The latest learning strategies have been employed to fuse the non-biometric features of the input face with that of the proxy face to generate the de-identified face. Extensive qualitative and quantitative analyses in Sect. 4 show that the proposed approach performs de-identification effectively and also significantly improves over the recent popular learning-based de-identifying approaches. In the future, focus can be given on preserving gender, hair color, etc., as well as improving the efficiency of the algorithm to carry out video de-identification.

Acknowledgements The authors thank NVIDIA for supporting their research with a Titan Xp GPU, and everyone who participated in the survey for computing the opinion scores.

References

1. Wu, Y., Yang, F., Ling, H.: Privacy-protective-GAN for face de-identification (2018). arXiv:1806.08906
2. Ribaric, S., Ariyaeeinia, A., Pavesic, N.: De-identification for privacy protection in multimedia content: a survey. Signal Process. Image Commun. **47**, 131–151 (2016)
3. Frome, A., Cheung, G., Abdulkader, A., Zennaro, M., Wu, B., Bissacco, A., Adam, H., Neven, H., Vincent, L.: Large-scale privacy protection in Google street view. In: Proceedings of the International Conference on Computer Vision, pp. 2373–2380 (2009)
4. Sohn, H., De Neve, W., Ro, Y.M.: Privacy protection in video surveillance systems: analysis of subband-adaptive scrambling in JPEG XR. IEEE Trans Circuits Syst Video Technol **21**(2), 170–177 (2011)
5. Schulz, L.E., Bonawitz, E.B.: Serious fun: preschoolers engage in more exploratory play when evidence is confounded. Dev Psychol **43**(4), 1045 (2007)
6. Mosaddegh, S., Simon, L., Jurie, F.: Photorealistic face de-identification by aggregating donors' face components. In: Proceeding of the Asian Conference on Computer Vision, pp. 159–174. Springer (2014)
7. Tejaswini, G., Venkataramana, T.: A novel reversible de-identification approach for lossless image compression based on reversible watermarking mechanism based on obfuscation process. Int. J. Eng. Comput. Sci. **4**(10), 14817–14823 (2015)
8. Gross, R., Sweeney, L., De la Torre, F., Baker, S.: Model-based face de-identification. In: Proceedings of the Conference on Computer Vision and Pattern Recognition Workshop, pp. 161–161. IEEE (2006)
9. Gross, R., Airoldi, E., Malin, B., Sweeney, L.: Integrating utility into face de-identification. In: Proceedings of the International Workshop on Privacy Enhancing Technologies, pp. 227–242. Springer (2005)
10. Gross, R., Sweeney, L., De La Torre, F., Baker, S.: Semi-supervised learning of multi-factor models for face de-identification. In: Proceedings of the Computer Vision and Pattern Recognition, pp. 1–8. IEEE (2008)
11. Samarzija, B., Ribaric, S.: An approach to the de-identification of faces in different poses. In: Proceedings of the 37th International Convention on Information and Communication Technology, Electronics and Microelectronics, pp. 1246–1251. IEEE (2014)
12. Meng, L., Sun, Z., Ariyaeeinia, A., Bennett, K.L.: Retaining expressions on de-identified faces. In: Proceedings of the 37th International Convention on Information and Communication Technology, Electronics and Microelectronics, pp. 1252–1257. IEEE (2014)
13. Sun, Z., Meng, L., Ariyaeeinia, A.: Distinguishable de-identified faces. In: Proceedings of the 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, vol. 4, pp. 1–6. IEEE (2015)
14. Du, L., Yi, M., Blasch, E., Ling, H.: GARP-face: balancing privacy protection and utility preservation in face de-identification. In: Proceedings of the IEEE International Joint Conference on Biometrics, pp. 1–8. IEEE (2014)
15. Meden, B., Emeršič, Ž., Štruc, V., Peer, P.: k-Same-Net: k-Anonymity with generative deep neural networks for face deidentification. Entropy **20**(1), 60 (2018)
16. Meden, B., Mallı, R.C., Fabijan, S., Ekenel, H.K., Štruc, V., Peer, P.: Face deidentification with generative deep neural networks. IET Signal Process. **11**(9), 1046–1054 (2017)
17. Li, Y., Lyu, S.: De-identification without losing faces. In: Proceedings of the ACM Workshop on Information Hiding and Multimedia Security, pp. 83–88 (2019)
18. He, Z., Zuo, W., Kan, M., Shan, S., Chen, X.: Arbitrary facial attribute editing: only change what you want (2017). CoRR, arXiv:1711.10678
19. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks (2018). CoRR, arXiv:1812.04948
20. Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the Computer Vision and Pattern Recognition (2017)
21. Zhang, L., Ji, Y., Lin, X. and Liu, C.: Style transfer for anime sketches with enhanced residual U-Net and auxiliary classifier GAN. In: Proceedings of the 4th Asian Conference on Pattern Recognition, pp. 506–511. IEEE (2017)
22. Kancharagunta, K.B., Dubey, S.R.: CSGAN: cyclic-synthesized generative adversarial networks for image-to-image transformation (2019). arXiv:1901.03554
23. Cao, Y., Zhou, Z., Zhang, W. and Yu, Y.: Unsupervised diverse colorization via generative adversarial networks. In: Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pp. 151–166. Springer (2017)
24. Arriaga, O., Valdenegro-Toro, M., Plöger, P.: Real-time convolutional neural networks for emotion and gender classification (2017). arXiv:1710.07557
25. Zheng, X., Cai, D., He, X., Ma, W.Y., Lin, X.: Locality preserving clustering for image database. In: Proceedings of the 12th Annual ACM International Conference on Multimedia, pp. 885–891. ACM (2004)
26. Biswas, S.K., Milanfar, P.: Laplacian object: one-shot object detection by locality preserving projection. In: IEEE International Conference on Image Processing, pp. 4062–4066. IEEE (2014)
27. Gualberto, A.: Regressor-face pose (2018). https://github.com/arnaldog12/Deep-Learning/tree/master/problems
28. Parkhi, O.M., Vedaldi, A., Zisserman, A., et al.: Deep face recognition. In: Proceedings of the British Machine Vision Conference, pp. 1–12 (2015)
29. Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D.H., Hawk, S.T., Van Knippenberg, A.D.: Presentation and validation of the radboud faces database. Cogn. Emot. **24**(8), 1377–1388 (2010)
30. Messer, K., Matas, J., Kittler, J., Luettin, J. and Maitre, G.: XM2VTSDB: the extended M2VTS database. In: Proceedings of the 2nd International Conference on Audio and Video-Based Biometric Person Authentication, vol. 964, pp. 965–966 (1999)
31. Liu, Z., Luo, P., Wang, X., Tang, X.: Deep learning face attributes in the wild. In: Proceedings of International Conference on Computer Vision (ICCV), pp. 3730–3738 (2015)