Review

Current Landscape of Generative Adversarial Networks for Facial Deidentification in Dermatology: Systematic Review and Evaluation

Christine Park[1], BA; Hyeon Ki Jeong[1], PhD; Ricardo Henao[2], PhD; Meenal Kheterpal[1], MD

[1]Department of Dermatology, Duke University Medical Center, Durham, NC, United States
[2]Department of Biostatistics and Bioinformatics, Duke University, Durham, NC, United States

Corresponding Author:
Christine Park, BA
Department of Dermatology
Duke University Medical Center
40 Duke Medicine Cir.
Durham, NC, 27710
United States
Phone: 1 7757728063
Email: cp268@duke.edu

Abstract

Background: Deidentifying facial images is critical for protecting patient anonymity in the era of increasing tools for automatic image analysis in dermatology.

Objective: The aim of this paper was to review the current literature in the field of automatic facial deidentification algorithms.

Methods: We conducted a systematic search using a combination of headings and keywords to encompass the concepts of facial deidentification and privacy preservation. The MEDLINE (via PubMed), Embase (via Elsevier), and Web of Science (via Clarivate) databases were queried from inception to May 1, 2021. Studies of incorrect design and outcomes were excluded during the screening and review process.

Results: A total of 18 studies reporting on various methodologies of facial deidentification algorithms were included in the final review. The study methods were rated individually regarding their utility for use cases in dermatology pertaining to skin color and pigmentation preservation, texture preservation, data utility, and human detection. Most studies that were notable in the literature addressed feature preservation while sacrificing skin color and texture.

Conclusions: Facial deidentification algorithms are sparse and inadequate for preserving both facial features and skin pigmentation and texture quality in facial photographs. A novel approach is needed to ensure greater patient anonymity, while increasing data access for automated image analysis in dermatology for improved patient care.

Introduction

Facial Deidentification in Dermatology

Over the last several years, there has been an explosion of artificial intelligence (AI) and deep learning for dermatological image analysis. These tools have demonstrated efficacy in the diagnosis and quantification of skin conditions at par with or surpassing human performance [1,2]. Additionally, there have been use cases in dermatology where the human eye is unable to precisely quantify the burden of disease, while AI can be used to support the clinical decision-making process with better consistency [3,4].

Facial image data are needed for developing models that evaluate attributes such as redness (ie, acne and rosacea models), texture (ie, wrinkles and aging models), pigmentation (ie, melasma, seborrheic keratoses, aging, and postinflammatory hyperpigmentation models), and skin lesions. To advance AI in dermatology, image data are needed at scale. For patient data to be used for research, consent may be obtained; however, for data at scale where this is not possible, adequate deidentification must be applied to images. Traditionally, journals have required facial feature concealment that typically covers the eyes, but these guidelines are largely insufficient to meet the ethical and legal guidelines from the Health Insurance Portability and

Accountability Act for patient privacy and identity protection [5,6]. Facial features, tattoos, jewelry, birthmarks, and other identity-informative background features are additional features that are considered identifying; facial feature deidentification is considered the most challenging task, given a lack of expert consensus and a lack of testing infrastructure and quantitative metrics for adequacy of automatic and manual facial image deidentification algorithms.

Identity protection challenges extend to other industries involved with facial images as well as video privacy. Hence, there have been increasing efforts to propose facial deidentification algorithms in the literature with corresponding use cases. Ideally, the methods should both hide the original identity of participants and preserve data reusability. We hypothesize that automated facial deidentification algorithms currently proposed in the literature may be useful for dermatological research use. To this end, we conducted a systematic review to search for studies reporting facial deidentification and summarized their proposed methodology and application to image analysis in dermatology.

Comparison of Different Facial Deidentification Algorithms

Conventional methods of ad hoc facial deidentification use blur [7], pixelation [8], masking, random swapping, perturbation, and face region replacement [7,9-18] to obfuscate parts or entire images to protect visual privacy. This set of obfuscating techniques prevent the rendering of the original image, but they do not necessarily guarantee preservation of privacy (ie, masks and blur can be removed) and often compromise data utility (ie, preservation of dermatological characteristics with diagnostic value) [19,20]. To test if these techniques protect privacy, studies have explored whether these methods can fool computer and human detection. Many studies have successfully avoided detection by use of computer algorithms but have found that human eyes can easily notice the alteration [21-24]. Furthermore, simply applying distorting filters to images risks identity revelation after reconstruction [13].

The k-anonymity–based algorithms were proposed as one of the original feasible approaches in solving this issue of data utility after deidentification [25]. Briefly, the k-anonymity–based methods and their variations deidentify an image by replacing the face with the average of k images from a given collection of images, and they achieve privacy protection with a rate lower than $1/k$. The most commonly used k-algorithm is from the k-Same family [8,13,17]. However, one of the key issues with the variations of the k-Same algorithm is the introduction of ghosting artifacts caused by the misalignment of images. The ghosting artifacts compromise privacy protection by making the images appear unnatural. The ghosting effect can be overcome by employing a large k in the algorithms, but this requires a large image collection, otherwise it results in a lack of distinction among the deidentified faces; this is because the number of discriminative faces in the deidentified face set is limited by the total number of images divided by k. This is problematic for applications in skin image analysis in dermatology because adequate privacy protection is achieved with averaging a greater number of images, which, in turn, will dilute redness, pigmentation, and other image attributes that are critical to dermatologic data utility. In other words, there is an intrinsic trade-off when choosing k between identifiability and preservation of dermatological features.

The k-Same-M algorithm was developed to eliminate the ghosting effects in order to enhance privacy protection with minimal loss of data utility [26]. This algorithm uses an active appearance model (AAM), which is an algorithm that can reconstruct an image representation based on its shape and texture [26]. In this way, an AAM coupled with the k-based algorithms can help reduce the ghosting effect in the deidentified images by ensuring a better alignment of the synthesized identity onto the original images. However, the reconstructed images from an AAM are still averaged images from the respective data set and, hence, some important aspect of data utility, such as facial expression, could be compromised.

Another technique for achieving facial deidentification is through the use of machine learning methods involving deep neural networks [27-31]. Convolutional neural networks (CNNs) are effective in extracting features from raw faces and, hence, facilitate image transformation into target outcomes. Limitations associated with methods involving CNNs and convolutional autoencoders are that they are time costly because they require a large sample size to be trained and optimized. Specifically for CNNs, these are supervised algorithms that also need labels for ground-truth classifications. Furthermore, the output images are still not natural enough to effectively preserve privacy.

Generative neural networks (GNNs) constitute a novel method to generate realistic face surrogates that can be used for deidentification. This quality can be exploited to retain skin attribute quality from a source image of interest. These also allow for retaining certain aspects of the original data, such as age, gender, and facial expressions, while replacing sensitive personal attributes with artificial objects, such as facial features. GNNs are originally based on generative adversarial networks (GAN), which combine a generative model that produces a synthetic image and a discriminator (ie, critic) network that classifies the synthetic image as either real or artificial. This method works by training the discriminator network as a standard classifier to distinguish between the two image sources as real or artificial and training the generative network as an image-generating model that can fool the discriminator network, with the goal of generating the most realistic-appearing synthetic images [32]. The model is improved in an adversarial manner via back-propagation with both generative and discriminator networks to identify the generator's parameters that should be optimized to make the generated images increasingly challenging for the discriminator. After completion of training, the output images from the generator network should be indistinguishable from the real images for the discriminator as well as look visually convincing for humans [13,25,33-35].

The use of GANs in facial deidentification is intriguing due to their potential for disentanglement of facial features and skin attributes. Theoretically, facial images can be deidentified by a GAN that recognizes facial features, such as eyes, nose, and lips, and then replaces them with features from another facial image, while continuing to preserve the realistic-appearing facial image as well as features of interest, such as redness,

pigmentation, texture, and skin lesions. Hence, based on their high data utility, GANs hold the promise of privacy protection by completely changing image identification by human and automated detection. This study focused on reviewing the GAN-based models published to date for facial deidentification for dermatologic use cases. We also evaluated the performance of top-performing GANs in deidentifying dermatological images while preserving the important facial and skin quality features in these images.

Methods

Search Strategy

We conducted a systematic search using a combination of headings and keywords to encompass the concepts of facial deidentification and privacy preservation. The MEDLINE (via PubMed), Embase (via Elsevier), and Web of Science (via Clarivate) databases were queried from inception to May 1, 2021. We also performed referential backtracking on the most recent studies to ensure inclusion of all relevant articles. Studies of incorrect design and outcomes were excluded during the screening and review process. The search strategies are outlined in Multimedia Appendix 1.

Definitions and Inclusion and Exclusion Criteria

Facial features were defined as identifying features associated with an individual, including the eyebrows, eyes, nose, mouth, and ears. For deidentification in dermatologic use cases, these features are important to remove and replace. The skin was then defined as the remaining facial area bounded by the hairline.

Preservation of skin quality by algorithms was evaluated as to how well the algorithms preserved the quality of the skin tone and texture from the input images. We included studies that focused on variations of the GAN algorithm for the purpose of facial deidentification in images, video, or both. Studies were excluded if they focused on any other facial deidentification algorithms due to low preservation of pixel-level skin quality based on the methodology.

Ethics Approval

This study was approved by the Institutional Review Board (Retrospective cutaneous dermato-oncological conditions treated by dermatology service) for protocol No. Pro00100765. Patient consent was not required due to the nature of this study.

Results

Overview

A total of 18 studies using GAN methodology were included in the final review (Figure 1). Table 1 [36-53] summarizes the different types of GAN algorithms and the goals of all the studies as well as an evaluation of their ability to preserve skin quality (ie, color and texture), capacity for data utility, and demonstration of adequate facial deidentification with human eyes based on the results illustrated in the studies. We then applied two of the best GAN-based algorithms that were publicly available to the SD-260 (260 classes of skin diseases) data set [54], a public data set of images of dermatological conditions, to assess whether the output images appropriately preserved skin quality.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram.
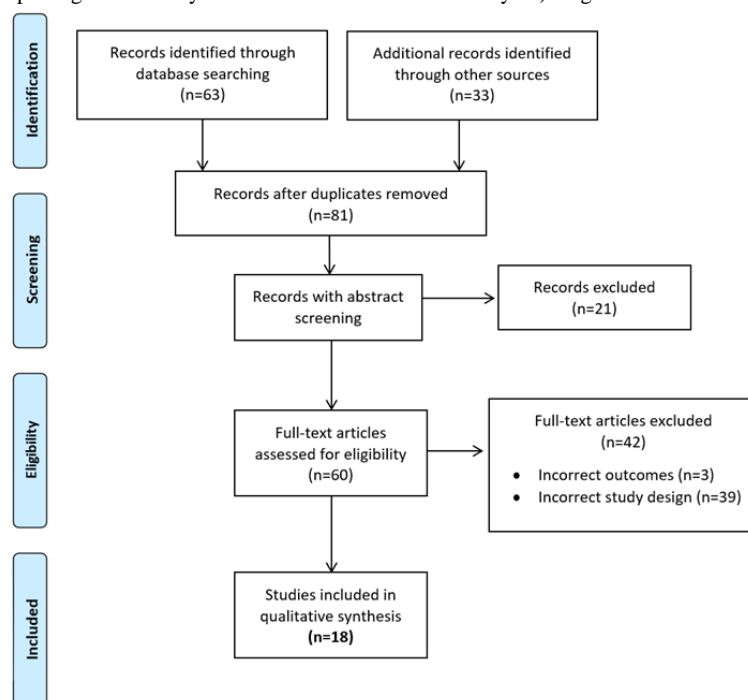
Table 1. Overview of included GAN-based studies.

Author, year	Method of facial dei-dentification	Novelty in proposed method of facial dei-dentification	Skin attribute preservation		Data utility	Facial deidentifi-cation (human)
			Color	Texture		
Pan et al, 2019 [36]	*k*-Same-Siamese-GAN[a]	Maintenance of high resolution of images to preserve their utility	Partial	No	Low	Yes
Song et al, 2019 [37]	Evolutionary GAN	Structural similarity index and the distance between the original face and the deidentified face	Partial	Partial	Low	No
Agarwal et al, 2021 [38]	StyleGAN and GAN	Preservation of emotion and nonbiometric facial attributes of a target face	N/A[b]	No	Low	Yes
Nitzan et al, 2020 [39]	Disentanglement coupled with GAN	Disentanglement of identity from other facial attributes with minimal training	Yes	No	High	No
Lin et al, 2021 [40]	Facial privacy GAN for social robots	Strengthened feature-extraction ability to improve the discriminatory accuracy	Partial	No	Low	Partial
Maximov et al, 2020 [41]	Conditional identity anonymization GAN	Development of a model for image and video anonymization with removal of identifying characteristics of faces and bodies	Yes	No	High	Yes
Brkic et al, 2017 [42]	Conditional GAN	Production of realistic deidentified human images that avoid human- and machine-based recognition	N/A	N/A	Low	N/A
Meden et al, 2017 [43]	Generative neural network	Synthesis of artificial surrogate faces with preservation of nonidentity-related aspects of the data for data use	No	No	Low	Yes
Mirjalili et al, 2017 [44]	Convolutional autoencoder using semiadversarial network	Autoencoder-based transformation of an input face image	N/A	No	Low	No
Radford et al, 2016 [45]	DCGAN[c]	Unsupervised GAN	No	No	Low	No
Wu et al, 2019 [46]	Privacy-protective GAN	Privacy protection, utility preservation, and structure similarity	N/A	Partial	Low	Yes
Hukkelås et al, 2019 [47]	Conditional GAN	Novel generator architecture for face anonymization via synthesis of realistic faces	No	No	Low	Yes
Ren et al, 2018 [48]	Multitask extension of GAN	Deidentification in video with preservation of action	No	No	High	Yes
Sun et al, 2018 [49]	DCGAN	Novel head inpainting obfuscation technique	Partial	No	Low	Yes
Sun et al, 2018 [50]	GAN	New hybrid approach for identity obfuscation in photos via head replacement	Partial	No	Low	Yes
Bao et al, 2018 [51]	GAN	Disentanglement of identity and attributes from faces for recombination into different identities and attributes for identity-preserving face synthesis in open domains	No	No	High	No
Li et al, 2019 [52]	Adaptive embedding integration network	High-fidelity face swapping	Yes	No	High	Yes
Nirkin et al, 2019 [53]	Face-swapping GAN	Face re-enactment with adjustment for pose and expression variations	No	No	High	Yes

[a]GAN: generative adversarial network.

[b]N/A: not applicable; this information was not reported in this study.

[c]DCGAN: deep convolutional generative adversarial network.

XSL•FO

RenderX

Disentanglement-Coupled GAN

One of the algorithms we chose was the disentanglement-coupled GAN presented by Nitzan et al [39]. The goal of this model is to generate an image by combining the identity of a given identity image with the attributes extracted from an attribute image. The author generates 70,000 images using StyleGAN [55], which are then used as the training data set. Identity is preserved by penalizing the identity difference between the identity image and attribute image. Attribute preservation is achieved by penalizing the difference in pixel-level and facial landmarks between identity image and attribute image. The network architecture is illustrated in Figure 2.

The performance of this method was compared against previously published methods, such as latent optimization for representation disentanglement [56], FaceShifter [52], and face-swapping GAN [53], for qualitative assessment; the performance was also compared against the adversarial latent autoencoder (ALAE) method [57] and the pixel2style2pixel (pSp) method [58] for quantitative assessment. Qualitatively, the authors demonstrated that their method showed better preservation for facial expression (ie, attribute image), head shape, and hair (ie, identity image) compared to the other models noted above. Quantitatively, the reconstruction performance was assessed by measuring pixel-wise reconstruction and preservation of semantic features, followed by comparison of the outcome to that of ALAE and pSp methods. This evaluation indicated that the pSp method showed better performance, but the author emphasized that their method was mainly for disentanglement and was not necessarily designed to reconstruct pixel-level information for reconstruction. This indicates that the model was able to replace and preserve realistic facial features, head shape, hair, and expressions due to superior performance of the disentanglement component while compromising pixel-level detail.

When applying the disentanglement-coupled GAN to the SD-260 data set, there were two sources for the input data: one for *identity* and another for *attribute*. For this model, we experimented with whether the attributes, such as redness and pigmentation, of the faces from the dermatological images could be encoded in a new identity. Figure 3A shows the qualitative results derived from the model: in the data set where the images of interest, with redness and pigmentation, are the *attribute* images, there is no transfer of skin features of interest, only transfer of facial positions and expressions. Figure 3B shows that when the images of interest are the *identity* images, features are transferred without pixel-level accuracy to preserve high data utility for dermatology use. Overall, we can see that while the model generates realistic faces, it is unable to preserve pixel-level details of the faces.

Figure 2. Disentanglement scheme. Solid lines indicate data flow and dashed lines indicate data loss. The identity and attribute codes are first extracted from two input images using encoders E_{id} and E_{attr}, respectively. Through the mapping network M, the concatenated codes are mapped to W, the latent space of the pretrained generator G, which, in turn, generates the resulting image. An adversarial loss L_{adv} ensures proper mapping to the W space. Identity preservation is encouraged using L_{id}, which penalizes differences in identity between I_{id} and I_{out}. Attribute preservation is encouraged using L_{rec} and L_{lnd}, which penalize pixel-level and facial landmark differences, respectively, between I_{attr} and I_{out} (reproduced from Nitzan et al [39], with permission from Yotam Nitzan). D_W: discriminator; Elnd: landmark encoder; z: latent code.
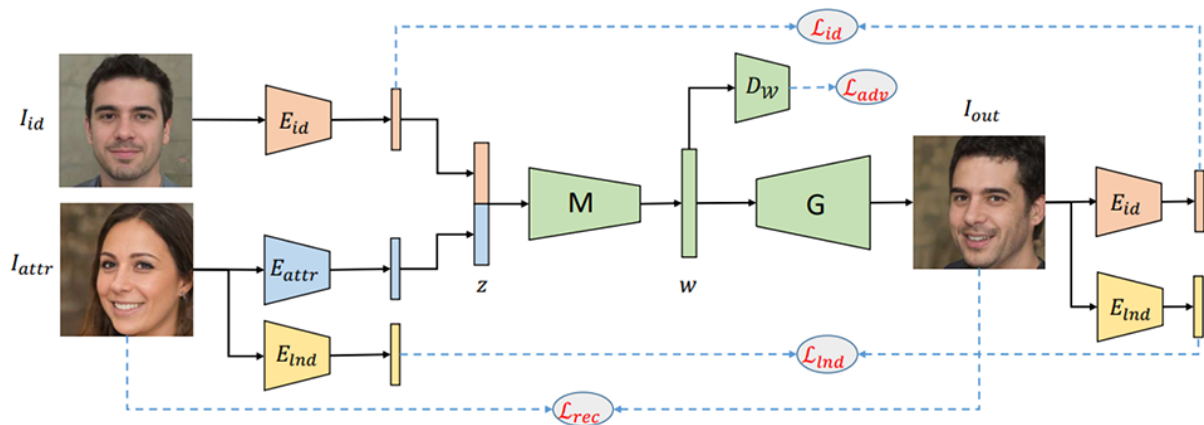
Figure 3. Output using the disentanglement-coupled GAN on dermatological images derived from the SD-260 data set. (A) Identity images assuming the facial pose and alteration of facial features from the attribute images. The attribute images fail to transfer the features of interest (ie, redness and pigmentation). (B) When switching the identity images to the images with features of interest, the model fails to preserve the dermatological features. GAN: generative adversarial network; SD-260: 260 classes of skin diseases.
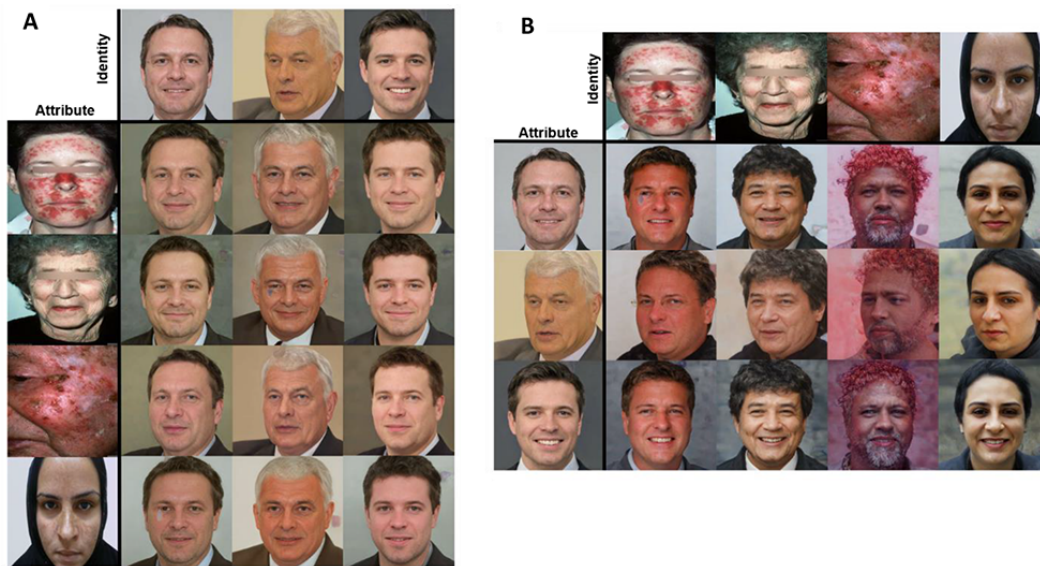
Conditional Identity Anonymization GAN

The goal of this paper was to develop a model that can deidentify images and videos while preserving features for other computer vision tasks, such as detection, tracking, or recognition [41]. The overview of the methodology is as follows. The method first extracted the landmarks of a given image that contained a sparse representation of the face with limited information on the identity. This allowed the generator to adjust to the face shape, which enabled better preservation of the input pose. The authors used only the face silhouette, the mouth, and the bridge of the nose instead of using all 68 landmarks in order to allow the network to freely choose the facial features. The method also extracted masked background images to allow the model to learn to generate faces and not the background. Once the landmark and the background were extracted, the method used a conditional GAN (CGAN) [59] to generate realistic images by encoding the landmark and masked image and combining them with the identity images to feed into the decoder. The generated output image was then fed into the identity discriminator network to prevent the network from generating faces similar to the training data set and to ensure facial anonymization. The model architecture is shown in Figure 4.

The model was trained and evaluated on three public data sets: CelebA (CelebFaces Attributes), MOTS (Multi-Object Tracking and Segmentation), and Labeled Faces in the Wild. The performance of the model was assessed by using face detection and reidentification metrics with other existing methods, such as blurring and pixelization. When compared with a state-of-the-art facial deidentification method by Gafni et al [60], conditional identity anonymization GAN (CIAGAN) showed better deidentification rates by computer detection on two different data sets. The authors concluded that their method can both deidentify the source images better and generate much more diverse images compared to Gafni et al's method.

When we applied the CIAGAN to the SD-260 data set, we first processed the landmarks of the dermatological images. Then, we allowed the model to deidentify each individual's face from the processed landmark and background images. The model was pretrained using 1200 identities from the CelebA data set. Figure 5 shows the result from this model. The qualitative results show a reduction in pixel-level resolution as well as poor preservation of the dermatological attributes of interest in the mid to lower part of the face, while preserving the skin features of interest (ie, redness and pigmentation) in the forehead area. While this is a good method for facial swapping, CGAN at this level fails to preserve significant areas of interest with high-utility pixel-level detail.

Figure 4. CIAGAN model scheme. The model takes the image and its landmarks, the masked face, and the desired identity as input. The generator is an encoder-decoder model where the encoder embeds the image information into a low-dimensional space. The identity given as a one-hot label is encoded via a transposed convolutional neural network and is fed into the bottleneck of the generator. Then, the decoder decodes the combined information of source images and the identities into a generated image. The generator plays an adversarial game with a discriminator in a standard GAN setting. Finally, the identity discriminator network is introduced, whose goal is to provide a guiding signal to the generator about the desired identity of the generated face (reproduced from Maximov et al [41], with permission from Laura Leal-Taixe). CIAGAN: conditional identity anonymization generative adversarial network; GAN: generative adversarial network.
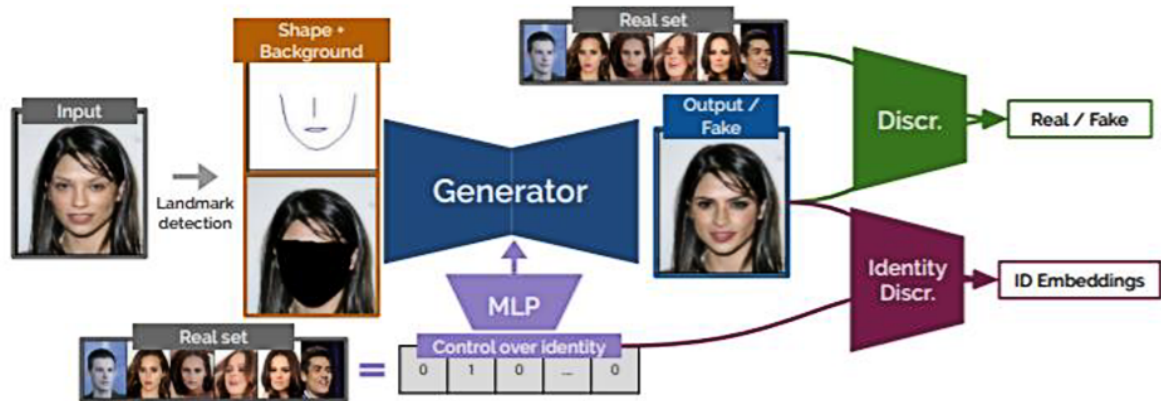
Figure 5. Output using CIAGAN on dermatological images derived from the SD-260 data set. Images on the left serve as source images, and a facial swap is done on the mid and lower part of the face for the images on the right. Generated images are of poor quality and only partially preserve facial attributes. CIAGAN: conditional identity anonymization generative adversarial network; SD-260: 260 classes of skin diseases.

Discussion

Principal Findings

Apart from the conventional facial deidentification methods, many of the advanced algorithms aim to preserve key facial features and expressions while maintaining privacy protection for the input images. Specifically, for GANs, there exist three major general limitations with these algorithms. Firstly, the outputs from these models that use face synthesis exhibit significant similarities between the synthetic and original images [61], which can be detected via human evaluation. Many of the currently existing algorithms are effective at modifying the images to avoid identification by face recognition software [17] but are not good enough to pass deidentification by humans. Thus, additional effort needs to be focused on addressing human detection, such as facial feature swap. Secondly, it is difficult to integrate the synthesized faces smoothly into the original image and make the images look unnatural, which compromises

privacy protection [17,62]. Finally, synthetic faces can decrease data usability due to changes in skin attributes, such tone and texture, and due to changes in patient identity, such as age, gender, and race [13,49,63-65]. Particularly for medical applications, even with the recently developed, well-intentioned algorithms, such as disentanglement and CIAGAN, the existing facial deidentification models fail to precisely and accurately preserve the color and texture of the facial skin for applications in their attempt to protect the identity of individuals with dermatological conditions, such as rosacea, melasma, among others, included in the data sets. Hence, the challenge involved with sharing large data sets that include facial images of patients with dermatological conditions, while adequately protecting their identity, remains unresolved.

The current standards for deidentifying patient images involve blurring, pixelating, and masking out important identifying facial features, such as the eyes and eyebrows [6]. Kuang et al [66] showed that pixelation and blurring demonstrate high

deidentification performance on computer detection compared to other advanced methods, such as privacy-protective GAN [67], natural and effective obfuscation [49], and AnonymousNet [63], which is one of the reasons that they remain as popular methods of facial deidentification. However, these conventional methods are at risk of identity restoration via decoding and reconstruction.

We propose that an ideal facial deidentification algorithm for dermatological application needs to (1) preserve facial architectural (ie, shape and gender) and skin features (ie, color and texture) to maintain data utility, while achieving adequate deidentification, and (2) avoid detection by computer and human analysis. To optimally protect the privacy of individuals in the images, the algorithm must be able to modify the image in a way that will be perceived as unaltered. In other words, the replacement identity will need to fuse well with the original content of the image. However, while altering the original content of the image, the skin attributes have to be preserved well enough so that the data utility of the data set involving the dermatological condition is not lost.

Herein, we demonstrate the utility of GAN-based facial deidentification methods to serve as use cases for AI development in dermatology, such as models quantifying redness (acne, rosacea, dermatitis, etc), pigmentation (melasma, postinflammatory hyperpigmentation, lentigines, etc), and texture (aging-related changes, volumetric assessment for neurotoxins or fillers, etc). While GAN development efforts for facial deidentification are not currently focused on skin-based use cases, focusing future efforts to achieve these goals can lead to an optimal facial deidentification model for dermatology.

Conclusions

Although facial deidentification is a rapidly evolving field with several advanced algorithms for achieving facial deidentification by computer-level recognition, their application to dermatology use cases is currently suboptimal. However, GAN-based models have the potential to preserve skin attributes while replacing facial features that risk detection, holding promise to solve the dilemma of data sharing while preserving patient privacy and identity. Future work should focus on developing a model that can achieve both skin attribute preservation as well as detection avoidance by both computers and humans.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strategy.
[PDF File (Adobe PDF File), 40 KB-Multimedia Appendix 1]

References

1. Du-Harpur X, Watt F, Luscombe N, Lynch M. What is AI? Applications of artificial intelligence to dermatology. Br J Dermatol 2020 Sep;183(3):423-430 [FREE Full text] [doi: 10.1111/bjd.18880] [Medline: 31960407]
2. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017 Feb 02;542(7639):115-118 [FREE Full text] [doi: 10.1038/nature21056] [Medline: 28117445]
3. De A, Sarda A, Gupta S, Das S. Use of artificial intelligence in dermatology. Indian J Dermatol 2020;65(5):352 [doi: 10.4103/ijd.ijd_418_20]
4. Jain A, Way D, Gupta V, Gao Y, de Oliveira Marinho G, Hartford J, et al. Development and assessment of an artificial intelligence-based tool for skin condition diagnosis by primary care physicians and nurse practitioners in teledermatology practices. JAMA Netw Open 2021 Apr 01;4(4):e217249 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.7249] [Medline: 33909055]
5. International Committee of Medical Journal Editors. Protection of patients' rights to privacy. BMJ 1995 Nov 11;311(7015):1272 [FREE Full text] [doi: 10.1136/bmj.311.7015.1272] [Medline: 11644736]
6. Roberts EA, Troiano C, Spiegel JH. Standardization of guidelines for patient photograph deidentification. Ann Plast Surg 2016 Jun;76(6):611-614 [doi: 10.1097/SAP.0000000000000817] [Medline: 27015333]
7. Neustaedter C, Greenberg S, Boyle M. Blur filtration fails to preserve privacy for home-based video conferencing. ACM Trans Comput Hum Interact 2006 Mar;13(1):1-36 [doi: 10.1145/1143518.1143519]
8. Boult TE. PICO: Privacy through invertible cryptographic obscuration. In: Proceedings of the Computer Vision for Interactive and Intelligent Environment Conference. 2005 Presented at: The Computer Vision for Interactive and Intelligent Environment Conference; November 17-18, 2005; Lexington, KY p. 27-38 [doi: 10.1109/cviie.2005.16]
9. Bitouk D, Kumar N, Dhillon S, Belhumeur P, Nayar S. Face swapping: Automatically replacing faces in photographs. In: Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference. New York, NY: Association for Computing Machinery; 2008 Presented at: The Special Interest Group on Computer Graphics and Interactive Techniques Conference; August 11-15, 2008; Los Angeles, CA p. 1-8 [doi: 10.1145/1399504.1360638]
10. Boyle M, Edwards C, Greenberg S. The effects of filtered video on awareness and privacy. In: Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work. 2000 Presented at: 2000 ACM Conference on Computer Supported Cooperative Work; December 2-6, 2000; Philadelphia, PA p. 1-10 [doi: 10.1145/358916.358935]

XSL•FO
RenderX

11. Crowley JL, Coutaz J, Bérard F. Perceptual user interfaces: Things that see. Commun ACM 2000 Mar;43(3):54 [doi: 10.1145/330534.330540]

12. Greenberg S, Kuzuoka H. Using digital but physical surrogates to mediate awareness, communication and privacy in media spaces. Pers Technol 1999 Dec;3(4):182-198 [doi: 10.1007/bf01540552]

13. Gross R, Airoldi E, Malin B, Sweeney L. Integrating utility into face de-identification. In: Proceedings of the 5th International Workshop on Privacy Enhancing Technologies.: International Workshop on Privacy Enhancing Technologies; 2005 Presented at: The 5th International Workshop on Privacy Enhancing Technologies; May 30-June 1, 2005; Cavtat, Croatia p. 227-242 [doi: 10.1007/11767831_15]

14. Gross R, Sweeney L, Torre F, Baker S. Semi-supervised learning of multi-factor models for face de-identification. In: Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition. 2008 Presented at: The 26th IEEE Conference on Computer Vision and Pattern Recognition; June 23-28, 2008; Anchorage, AK p. 1-8 [doi: 10.1109/cvpr.2008.4587369]

15. Hudson SE, Smith I. Techniques for addressing fundamental privacy and disruption tradeoffs in awareness support systems. In: Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work. New York, NY: Association for Computing Machinery; 1996 Presented at: The 1996 ACM Conference on Computer Supported Cooperative Work; November 16-20, 1996; Boston, MA p. 248-257 [doi: 10.1145/240080.240295]

16. Neustaedter C, Greenberg S. Balancing privacy and awareness in home media spaces. In: Proceedings of the 5th International Conference on Ubiquitous Computing. Workshop on Ubicomp Communities: Privacy as Boundary Negotiation. 2003 Presented at: The 5th International Conference on Ubiquitous Computing. Workshop on Ubicomp Communities: Privacy as Boundary Negotiation; October 12, 2003; Seattle, WA p. 1-5 URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.217.1778&rep=rep1&type=pdf

17. Newton E, Sweeney L, Malin B. Preserving privacy by de-identifying face images. IEEE Trans Knowl Data Eng 2005 Feb;17(2):232-243 [doi: 10.1109/tkde.2005.32]

18. Zhao Q, Stasko J. Evaluating image filtering based techniques in media space applications. In: Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work. New York, NY: Association for Computing Machinery; 1998 Presented at: The 1998 ACM Conference on Computer Supported Cooperative Work; November 14-18, 1998; Seattle, WA p. 11-18 [doi: 10.1145/289444.289450]

19. Nousi P, Papadopoulos S, Tefas A, Pitas I. Deep autoencoders for attribute preserving face de-identification. Signal Process Image Commun 2020 Feb;81:115699 [doi: 10.1016/j.image.2019.115699]

20. Oh S, Benenson R, Fritz M, Schiele B. Faceless person recognition: Privacy implications in social media. In: Proceedings of the 14th European Conference on Computer Vision. 2016 Presented at: The 14th European Conference on Computer Vision; October 8-16, 2016; Amsterdam, the Netherlands p. 19-35 [doi: 10.1007/978-3-319-46487-9_2]

21. Liu Y, Zhang W, Yu N. Protecting privacy in shared photos via adversarial examples based stealth. Secur Commun Netw 2017;2017:1-15 [doi: 10.1155/2017/1897438]

22. Oh S, Fritz M, Schiele B. Adversarial image perturbation for privacy protection -- A game theory perspective. In: Proceedings of the 2017 IEEE Computer Vision and Pattern Recognition. 2017 Presented at: The 2017 IEEE Computer Vision and Pattern Recognition; October 22-29, 2017; Venice, Italy p. 1491-1500 [doi: 10.1109/iccv.2017.165]

23. Sim T, Zhang L. Controllable face privacy. In: Proceedings of the 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition. 2015 Presented at: 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition; May 4-8, 2015; Ljubljana, Slovenia p. 1-8 [doi: 10.1109/fg.2015.7285018]

24. Moosavi-Dezfooli SM, Fawzi A, Fawzi O, Frossard P. Universal adversarial perturbations. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. 2017 Presented at: The 2017 IEEE Conference on Computer Vision and Pattern Recognition; July 21-26, 2017; Honolulu, HI p. 86-94 [doi: 10.1109/cvpr.2017.17]

25. Sweeney L. k-anonymity: A model for protecting privacy. Int J Uncertain Fuzziness Knowl Based Syst 2012 May 02;10(05):557-570 [doi: 10.1142/S0218488502001648]

26. Gross R, Sweeney L, Torre F, Baker S. Model-based face de-identification. In: Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop. 2006 Presented at: The 2006 Conference on Computer Vision and Pattern Recognition Workshop; June 17-22, 2006; New York, NY p. 161 [doi: 10.1109/cvprw.2006.125]

27. Taigman Y, Yang M, Ranzato M, Wolf L. DeepFace: Closing the gap to human-level performance in face verification. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. 2014 Presented at: The 2014 IEEE Conference on Computer Vision and Pattern Recognition; June 23-28, 2014; Columbus, OH p. 1701-1708 [doi: 10.1109/cvpr.2014.220]

28. Sun Y, Wang X, Tang X. Deep learning face representation from predicting 10,000 classes. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. 2014 Presented at: The 2014 IEEE Conference on Computer Vision and Pattern Recognition; June 23-28, 2014; Columbus, OH p. 1891-1898 [doi: 10.1109/cvpr.2014.244]

29. Schroff F, Kalenichenko D, Philbin J. FaceNet: A unified embedding for face recognition and clustering. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. 2015 Presented at: The 2015 IEEE Conference on Computer Vision and Pattern Recognition; June 7-12, 2015; Boston, MA p. 815-823 [doi: 10.1109/cvpr.2015.7298682]

30. Amos B, Ludwiczuk B, Satyanarayanan M. OpenFace: A general-purpose face recognition library with mobile applications. CMU School of Computer Science. Pittsburgh, PA: CMU School of Computer Science; 2016 Jun. URL: https://www.cs.cmu.edu/~satya/docdir/CMU-CS-16-118.pdf [accessed 2022-05-12]

31. Wu W, Kan M, Liu X, Yang Y, Shan S, Chen X. Recursive Spatial Transformer (ReST) for alignment-free face recognition. In: Proceedings of the 2017 IEEE International Conference on Computer Vision. 2017 Presented at: 2017 IEEE International Conference on Computer Vision; October 22-29, 2017; Venice, Italy p. 3792-3800 [doi: 10.1109/iccv.2017.407]

32. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. Commun ACM 2020 Oct 22;63(11):139-144 [doi: 10.1145/3422622]

33. Yang X, Li Y, Lyu S. Exposing deep fakes using inconsistent head poses. In: Proceedings of the 2019 IEEE International Conference on Acoustics, Speech and Signal Processing. 2019 Presented at: The 2019 IEEE International Conference on Acoustics, Speech and Signal Processing; May 12-17, 2019; Brighton, UK p. 8261-8265 [doi: 10.1109/icassp.2019.8683164]

34. 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. 2015 Presented at: 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition; May 4-8, 2015; Ljubljana, Slovenia p. 1-6 [doi: 10.1109/fg.2015.7285019]

35. Cao Y, Jia L, Chen Y, Lin N, Yang C, Zhang B, et al. Recent advances of generative adversarial networks in computer vision. IEEE Access 2019;7:14985-15006 [doi: 10.1109/access.2018.2886814]

36. Pan Y, Haung M, Ding K, Wu J, Jang J. k-Same-Siamese-GAN: k-Same algorithm with generative adversarial network for facial image de-identification with hyperparameter tuning and mixed precision training. In: Proceedings of the 16th IEEE International Conference on Advanced Video and Signal Based Surveillance. 2019 Presented at: The 16th IEEE International Conference on Advanced Video and Signal Based Surveillance; September 18-21, 2019; Taipei, Taiwan p. 1-8 [doi: 10.1109/avss.2019.8909866]

37. Song J, Jin Y, Li Y, Lang C. Learning structural similarity with evolutionary-GAN: A new face de-identification method. In: Proceedings of the 6th International Conference on Behavioral, Economic and Socio-Cultural Computing. 2019 Presented at: The 6th International Conference on Behavioral, Economic and Socio-Cultural Computing; October 28-30, 2019; Beijing, China p. 1-6 [doi: 10.1109/besc48373.2019.8962993]

38. Agarwal A, Chattopadhyay P, Wang L. Privacy preservation through facial de-identification with simultaneous emotion preservation. Signal Image Video Process 2020 Nov 27;15(5):951-958 [doi: 10.1007/s11760-020-01819-9]

39. Nitzan Y, Bermano A, Li Y, Cohen-Or D. Face identity disentanglement via latent space mapping. ACM Trans Graph 2020 Dec 31;39(6):1-14 [doi: 10.1145/3414685.3417826]

40. Lin J, Li Y, Yang G. FPGAN: Face de-identification method with generative adversarial networks for social robots. Neural Netw 2021 Jan;133:132-147 [doi: 10.1016/j.neunet.2020.09.001] [Medline: 33217682]

41. Maximov M, Elezi I, Leal-Taix'e L. CIAGAN: Conditional Identity Anonymization Generative Adversarial Networks. In: Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020 Presented at: The 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition; June 13-19, 2020; Seattle, WA p. 5446-5455 [doi: 10.1109/cvpr42600.2020.00549]

42. Brkic K, Sikiric I, Hrkac T, Kalafatic Z. I know that person: Generative full body and face de-identification of people in images. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2017 Presented at: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops; July 21-26, 2017; Honolulu, HI p. 1319-1328 [doi: 10.1109/CVPRW.2017.173]

43. Meden B, Mallı RC, Fabijan S, Ekenel HK, Štruc V, Peer P. Face deidentification with generative deep neural networks. IET Signal Process 2017 Dec;11(9):1046-1054 [doi: 10.1049/iet-spr.2017.0049]

44. Mirjalili V, Raschka S, Namboodiri A, Ross A. Semi-adversarial networks: Convolutional autoencoders for imparting privacy to face images. In: Proceedings of the 2018 International Conference on Biometrics. 2018 Presented at: The 2018 International Conference on Biometrics; February 20-23, 2018; Gold Coast, Australia p. 82-89 [doi: 10.1109/icb2018.2018.00023]

45. Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. In: Proceedings of the International Conference on Learning Representations. 2016 Presented at: The International Conference on Learning Representations; May 2-4, 2016; San Juan, Puerto Rico p. 1-16 URL: https://arxiv.org/abs/1511.06434

46. Wu Y, Yang F, Xu Y, Ling H. Privacy-protective-GAN for privacy preserving face de-identification. J Comput Sci Technol 2019 Jan 18;34(1):47-60 [doi: 10.1007/s11390-019-1898-8]

47. Hukkelås H, Mester R, Lindseth F. DeepPrivacy: A generative adversarial network for face anonymization. In: Proceedings of the International Symposium on Visual Computing. 2019 Presented at: International Symposium on Visual Computing; October 7-9, 2019; Lake Tahoe, NV p. 565-578 [doi: 10.1007/978-3-030-33720-9_44]

48. Ren Z, Lee Y, Ryoo M. Learning to anonymize faces for privacy preserving action detection. In: Proceedings of the 15th European Conference on Computer Vision. 2018 Presented at: The 15th European Conference on Computer Vision; September 8-14, 2018; Munich, Germany p. 639-655 [doi: 10.1007/978-3-030-01246-5_38]

49. Sun Q, Ma L, Oh S, Van GL, Schiele B, Fritz M. Natural and effective obfuscation by head inpainting. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018 Presented at: The 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; June 18-23, 2018; Salt Lake City, UT p. 5050-5059 [doi: 10.1109/cvpr.2018.00530]

50. Sun Q, Tewari A, Xu W, Fritz M, Theobalt C, Schiele B. A hybrid model for identity obfuscation by face replacement. In: Proceedings of the 15th European Conference on Computer Vision. 2018 Presented at: The 15th European Conference on Computer Vision; September 8-14, 2018; Munich, Germany p. 570-586 [doi: 10.1007/978-3-030-01246-5_34]

51. Bao J, Chen D, Wen F, Li H, Hua G. Towards open-set identity preserving face synthesis. In: Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018 Presented at: The 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; June 18-23, 2018; Salt Lake City, UT p. 6713-6722 [doi: 10.1109/cvpr.2018.00702]

52. Li L, Bao J, Yang H, Chen D, Wen F. FaceShifter: Towards high fidelity and occlusion aware face swapping. In: Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020 Presented at: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition; June 14-19, 2020; Virtual p. 1-11 URL: https://arxiv.org/pdf/1912.13457.pdf

53. Nirkin Y, Keller Y, Hassner T. FSGAN: Subject agnostic face swapping and reenactment. In: Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. 2019 Presented at: The 2019 IEEE/CVF International Conference on Computer Vision; 27 October 27-November 2, 2019; Seoul, South Korea p. 7183-7192 [doi: 10.1109/iccv.2019.00728]

54. Yang J, Wu X, Liang J, Sun X, Cheng M, Rosin PL, et al. Self-paced balance learning for clinical skin disease recognition. IEEE Trans Neural Netw Learn Syst 2020 Aug;31(8):2832-2846 [doi: 10.1109/tnnls.2019.2917524]

55. Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019 Presented at: The 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition; June 15-20, 2019; Long Beach, CA p. 4396-4405 [doi: 10.1109/cvpr.2019.00453]

56. Gabbay A, Hoshen Y. Demystifying inter-class disentanglement. In: Proceedings of the 8th International Conference on Learning Representations. 2020 Presented at: The 8th International Conference on Learning Representations; April 26-May 1, 2020; Virtual p. 1-22 URL: https://arxiv.org/pdf/1906.11796.pdf

57. Pidhorskyi S, Adjeroh D, Doretto G. Adversarial latent autoencoders. In: Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020 Presented at: The 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition; June 13-19, 2020; Seattle, WA p. 14092-14101 [doi: 10.1109/cvpr42600.2020.01411]

58. Richardson E, Alaluf Y, Patashnik O, Nitzan Y, Azar Y, Shapiro S. Encoding in style: A StyleGAN encoder for image-to-image translation. In: Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021 Presented at: The 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition; June 20-25, 2021; Nashville, TN p. 2287-2296 [doi: 10.1109/cvpr46437.2021.00232]

59. Mirza M, Osindero S. Conditional generative adversarial nets. ArXiv Preprint posted online on November 6, 2014 [FREE Full text]

60. Gafni O, Wolf L, Taigman Y. Live face de-identification in video. In: Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. 2019 Presented at: The 2019 IEEE/CVF International Conference on Computer Vision; October 27-November 2, 2019; Seoul, South Korea p. 9377-9386 [doi: 10.1109/iccv.2019.00947]

61. Agarwal S, Farid H, Gu Y, He M, Nagano K, Li H. Protecting world leaders against deep fakes. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019 Presented at: The 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition; June 15-20, 2019; Long Beach, CA p. 38-45 URL: https://tinyurl.com/4mk6vfac

62. Letournel G, Bugeau A, Ta V, Domenger J. Face de-identification with expressions preservation. In: Proceedings of the 2015 IEEE International Conference on Image Processing. 2015 Presented at: The 2015 IEEE International Conference on Image Processing; September 27-30, 2015; Quebec City, QC p. 4366-4370 [doi: 10.1109/icip.2015.7351631]

63. Li T, Lin L. Natural face de-identification with measurable privacy. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2019 Presented at: The 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops; June 16-17, 2019; Long Beach, CA p. 56-65 [doi: 10.1109/cvprw.2019.00013]

64. Du L, Yi M, Blasch E, Ling H. Balancing privacy protection and utility preservation in face de-identification. In: Proceedings of the 2014 IEEE International Joint Conference on Biometrics. 2014 Presented at: The 2014 IEEE International Joint Conference on Biometrics; September-October 2, 2014; Clearwater, FL p. 1-8 [doi: 10.1109/btas.2014.6996249]

65. Orekondy T, Fritz M, Schiele B. Automatic redaction of private information in images. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018 Presented at: The 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; June 18-23, 2018; Salt Lake City, UT p. 8466-8475 [doi: 10.1109/cvpr.2018.00883]

66. Kuang Z, Guo Z, Fang J, Yu J, Babaguchi N, Fan J. Unnoticeable synthetic face replacement for image privacy protection. Neurocomputing 2021 Oct;457:322-333 [doi: 10.1016/j.neucom.2021.06.061]

67. Liu Y, Peng J, Yu J, Wu Y. PPGAN: Privacy-preserving Generative Adversarial Network. In: Proceedings of the IEEE 25th International Conference on Parallel and Distributed Systems. 2019 Presented at: The IEEE 25th International

Conference on Parallel and Distributed Systems; December, 4-6, 2019; Tianjin, China p. 985-989 [doi: 10.1109/icpads47876.2019.00150]

Abbreviations

AAM: active appearance model
AI: artificial intelligence
ALAE: adversarial latent autoencoder
CelebA: CelebFaces Attributes
CGAN: conditional generative adversarial network
CIAGAN: conditional identity anonymization generative adversarial network
CNN: convolutional neural network
GAN: generative adversarial network
GNN: generative neural network
LORD: latent optimization for representation disentanglement
MOTS: Multi-Object Tracking and Segmentation
pSp: pixel2style2pixel
SD-260: 260 classes of skin diseases

XSL•FO
RenderX