k-Same-Net: Neural-Network-Based Face Deidentification

Blaž Meden*, Žiga Emeršič*, Vitomir Štruc† and Peter Peer*
*Faculty of Computer and Information Science, University of Ljubljana
Večna pot 113, SI-1000 Ljubljana, Slovenia
Email: {blaz.meden, ziga.emersic, peter.peer}@fri.uni-lj.si
†Faculty of Electrical Engineering, University of Ljubljana
Tržaška cesta 25, SI-1000 Ljubljana, Slovenia
Email: vitomir.struc@fe.uni-lj.si

***Abstract*—An increasing amount of video and image data is being shared between government entities and other relevant stakeholders and requires careful handling of personal information. A popular approach for privacy protection in such data is the use of deidentification techniques, which aim at concealing the identity of individuals in the imagery while still preserving certain aspects of the data after deidentification. In this work, we propose a novel approach towards face deidentification, called k-Same-Net, which combines recent generative neural networks (GNNs) with the well-known k-anonymity mechanism and provides formal guarantees regarding privacy protection on a closed set of identities. Our GNN is able to generate synthetic surrogate face images for deidentification by seamlessly combining features of identities used to train the GNN model. Furthermore, it allows us to guide the image-generation process with a small set of appearance-related parameters that can be used to alter specific aspects (e.g., facial expressions, age, gender) of the synthesized surrogate images. We demonstrate the feasibility of k-Same-Net in comparative experiments with competing techniques on the XM2VTS dataset and discuss the main characteristics of our approach.**

1. Introduction

In recent years an ever increasing amount of image and video data are being recorded, stored or processed worldwide. Key factors that contribute towards this development are personal gadgets, such as mobile phones or tablets, as well as other imaging devices (such as surveillance systems, security cameras and web-cams), which make capturing images and video footage an easy task. While these trends have made our life easier in many aspects, care also need to be taken that the captured data is not misused and the privacy of people visible in the imagery is adequately protected.

Privacy protection is especially important, when image and video data is recorded and shared between government entities and other relevant stakeholder, which may be inclined to exploit the data for purposes different from those for which they were recorded. The main issue here is how to perform secure data sharing, while at the same time preventing possible cases of misuse. As illustrated in Fig. 1, a common approach to address this problem is deidentification, which conceals personal identifiers present in data and thus prevents the recovery and misuse of identity information (e.g. prevents face recognition).

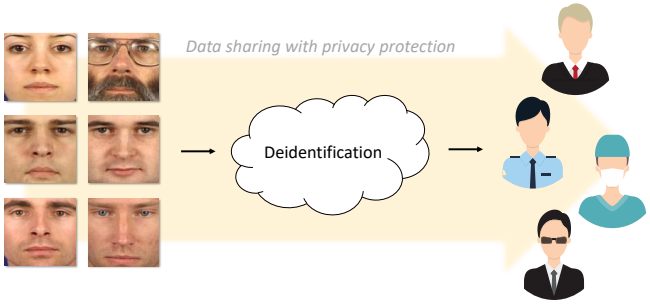

Figure 1: The motivation behind deidentification: to prevent misuse of personal information and ensure privacy protection when data is shared between government entities or other relevant stakeholders, the data needs to be appropriately deidentified before being shared.

As emphasized by Newton et al. in [26] and Gross et al. in [8], early deidentification techniques mostly included naive approaches, such as blacking-out, pixelation or blurring, which are not very effective nor suitable for this task. Blacking-out, for example, puts a black patch over the original face image to conceal identity. While this guarantees anonymity, it also removes all non-identity related information – including characteristics that could be useful for further analysis, but do not rely on identity information. Pixelation and blurring are also considered unsuitable for deidentification, as they are prone to imitation attacks (i.e. parrot attack [26]), where a probe image is simply subjected to the same degradation process as the deidentified images and can then be again recognized reliably.

More recent techniques from the literature try to overcome the limitations outlined above and provide formal guarantees regarding the anonymity of the deidentified data, e.g., [36], [21], [20]. We build on these techniques and present in this paper a novel deidentification approach called, k-Same-Net. The proposed approach exploits a re-

978-1-5386-2064-9/17/$31.00

cent class of generative models, i.e., generative neural networks (GNNs), and combines them with a formal anonymity scheme. The generative model is capable of synthesizing natural, realistic-looking surrogate faces for deidentification and parameterizes some of the visual characteristics of the synthesized data. This property makes it possible to retain certain attributes of the original data, while replacing sensitive personal traits with synthetic content. We demonstrate the feasibility of the proposed approach on a closed of facial images and show comparative results for competing techniques from the literature. We observe that k-Same-Net generates convincing surrogate faces without artifacts and is flexible enough to ensure that selected aspects of the data (such as facial expressions, age or gender) can be retained even after deidentification if needed.

To summarize, the main contributions of this work are:

- We introduce a novel algorithm for face deidentification, called k-Same-Net, that is based on a formal privacy protection scheme and relies on generative neural networks (GNNs).
- We present qualitative results that show how our approach is capable of preserving certain aspects of the data after deidentification.

The remainder of the paper is structured as follows. In Section 2 we review the related work including existing deidentification approaches and the recently introduced generative neural networks. In Section 3 we introduce our new deidentification approach called k-Same-Net combining generative models and a formal anonymity model. In Section 4 we discuss the experimental part and describe the datasets used for experimentation. We also provides some deidentification examples and highlight the merits of our approach in this section. Section 5 concludes the paper and suggests some future research directions.

2. Background and Related Work

In this section we briefly describe relevant work from the literature. We first discuss existing deidentification techniques and then proceed to a short overview of generative deep models.

K-anonymity and face deidentification: Many of the existing deidentification techniques from the literature are based on formal privacy-protection schemes, such as k-anonymity [36], L-diversity [21] or t-closeness [20], which provide theoretical guarantees about the anonymity of the deidentified data. In the field of face deidentification, k-anonymity is likely the most popular among the existing schemes and inspired the family of so-called k-Same deidentification algorithms, e.g., [26], [8], [10].

The main idea of the k-Same family of algorithms is illustrated in Fig. 2. The algorithms take a closed set of N images as input $\mathcal{I} = \{I_1, I_2, \ldots, I_N\}$ and produce a set of N deidentified images $\mathcal{D} = \{D_1, D_2, \ldots, D_N\}$ that cannot be linked to the inputs in an unambiguous manner. Here, anonymity is ensured by identifying clusters of k images in the input-image set $\mathcal{I}$ and replacing all k images

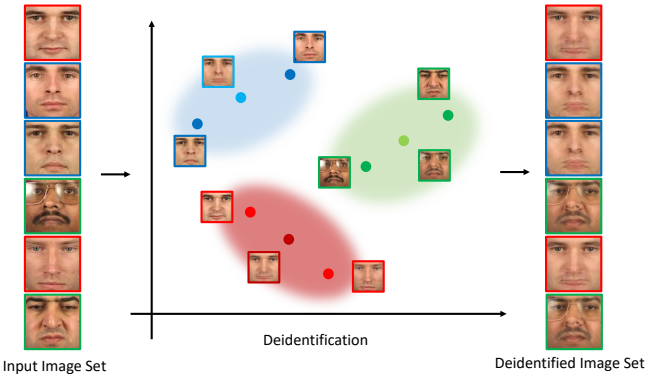

Figure 2: Illustration of the idea behind the k-anonymity mechanisms. The input images set $\mathcal{I}$ on the left is mapped to the deidentified image set $\mathcal{D}$ on the right. Anonymity is ensured by replacing k images from $\mathcal{I}$ with the same surrogate image. To preserve some of the information content of the original images, the surrogate images are computed as cluster centroids of the original images in $\mathcal{I}$. The figure shows an example for $k = 2$.

of each cluster with the corresponding cluster centroid. As a result, the deidentified image set $\mathcal{D}$ contains k copies of each computed centroid and, consequently, each image (or centroid) in $\mathcal{D}$ bears similarities with all k images in the corresponding cluster. The outlined procedure makes it impossible to link any individual from $\mathcal{D}$ to the individuals from $\mathcal{I}$ with a probability higher than $1/k$ and provides formal guarantees with respect to the anonymity of the deidentified data. Note that these guarantees apply only if the images in $\mathcal{I}$ belong to exactly N identities and, hence, each subject in $\mathcal{I}$ is represented with a single image only.

The original k-Same algorithm, proposed by Newton et al. in [26], operates directly in the pixel space and, therefore, preserves visual characteristics of all k images of each cluster in the cluster centroids. The motivation for the algorithm comes from the fact that: *i)* replacement of all images in the k-sized clusters of $\mathcal{I}$ with the same surrogate images ensures anonymity, and *ii)* selecting the cluster centroids (of similar faces) as the surrogates minimizes information loss during deidentification. These properties are illustrated in Fig. 2, where sample deidentification results for the original k-Same approach are presented. As can be seen, the deidentified images still preserve some of the visual information contained in the original images, but also exhibit ghosting effects that appear as a consequence of poor alignment of the images in $\mathcal{I}$.

To address these limitations an extension of the k-Same algorithm was presented by Gross et al. in [9]. The algorithm, named k-Same-Model or k-Same-M, extends the idea of k-anonymity to Active Appearance Models (AAMs) and applies the deidentification procedure in the AAM parameter space. Because AAMs ensure better alignment between images, the surrogate faces feature almost no ghosting effects and appear more realistic. Nevertheless, some potentially useful information (pertaining, for example, to facial expres-

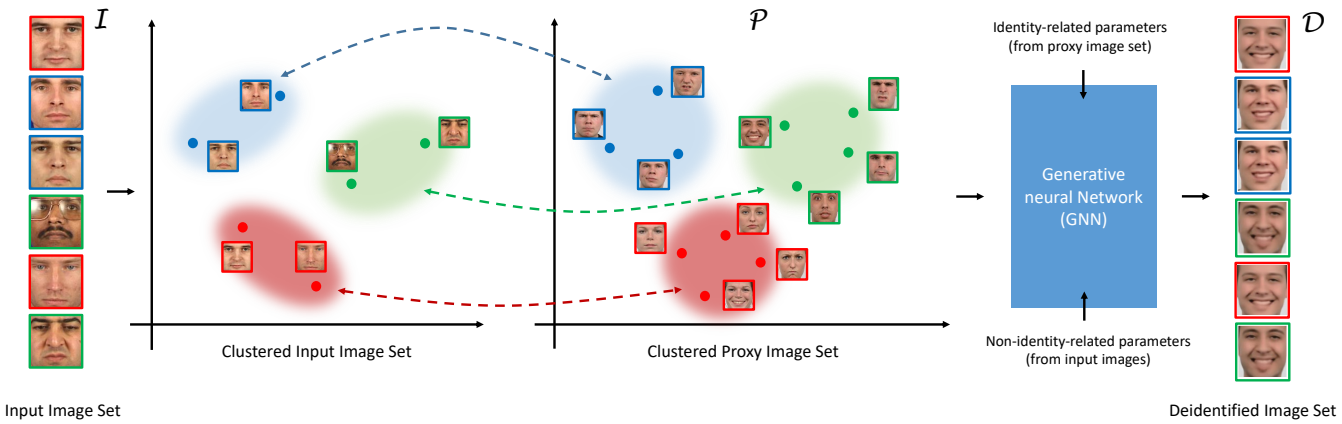

Figure 3: Overview of the k-Same-Net deidentification approach. Similar to other k-Same algorithms, each image in the input set $\mathcal{I}$ on the left is mapped to an image in the deidentified image set $\mathcal{D}$ on the right with k images from $\mathcal{I}$ mapping to the same image in $\mathcal{D}$. The surrogate faces in $\mathcal{D}$ are generated by a GNN that is trained to produce identities from a proxy image set $\mathcal{P}$. Other visual characteristics of the generated surrogate faces (pertaining, for example, to facial expressions, age, gender, etc.) are defined by a set of non-identity related parameters of the GNN and depending on the application can be easily modified during the deidentification with k-Same-Net.

sions) may still get lost during the deidentification process due to the averaging step.

Many extensions of the above approach have been proposed in the literature focusing mainly on improving the naturalness of the deidentified faces and preservation of as much of the non-identity-related information in the original images as possible. These include the k-Same-select [8] and k-Diff-furthest [35] to name a few.

For more information on face deidentification the reader is referred to [34], where a recent survey on this topic is presented.

Generative deep neural networks: Generative deep neural networks (GNNs) represent recent generative models capable of synthesizing artificial naturally looking images of any object and are, therefore, also highly suitable for the task of deidentification.

Goodfellow et al. [7], for example, proposed so-called Generative Adversarial Networks (GANs), which combine two contradictive deep architectures: a basic generative model that synthesizes artificial images and a second discriminator network that tries to classify the synthesized image is either real or artificially generated. The main idea here is to train the discriminator network as a standard classifier to distinguish between two image sources (real or artificial) and the generative network as a generative model capable of fooling the discriminator network. Back-propagation is used with both the discriminator and the generator network to find how the generators parameters should be changed in order to make the generated samples slightly more challenging for the discriminator. Once the training is completed, the generator network outputs images that are indistinguishable from real images for the discriminator and also look visually convincing for human observers.

Dosovitskiy et al. [4] introduced a generative neural network (GNN), capable of drawing 2D images of 3D objects given the object style, viewpoint and color. The network architecture used in this work is identical to the standard Convolutional Neural Network (CNN) architectures commonly used for classification, but is turned "upside-down", which makes it possible to generate synthetic images given high-level information. Thus, instead of learning a classification problem, the authors demonstrate how to generate images from their high-level descriptions. The input parameters consist of three vectors: one-hot encoding of the model identity (which defines the style), azimuth and elevation of the camera position, and the parameters of additional artificial transformations applied to the images. The higher network layers first build a shared, high dimensional hidden object representation from the input parameters, whereas the latter layers generate an image and object segmentation mask from the shared hidden representation.

While several other generative models have been introduced recently in the literature, e.g., [16], [30], [2], we build on the work of Dosovitskiy et al. [4] in this paper and build our k-Same-Net algorithm around this class of GNNs. Note, however, that the same idea could be extended to other model architectures as well.

3. Neural-Network-Based Deidentification

In this section, we introduce the proposed k-Same-Net algorithm. We first present a short overview of the approach, then discuss how deidentification and information preservation is achieved and finally describe the generative part of k-Same-Net including its architecture and training.

3.1. k-Same-Net Overview

A high-level overview of the k-Same-Net approach is presented in Fig. 3. Similar to other algorithms from the

k-Same family, our approach operates on a closed set of N input images $\mathcal{I}$ corresponding to N distinct identities. The algorithm maps the input set $\mathcal{I}$ to a target set of deidentified images $\mathcal{D}$, but unlike existing techniques relies on an additional proxy set of images $\mathcal{P}$ to implement the mapping. With our approach, formal anonymity guarantees are again ensured by replacing clusters of k-images from $\mathcal{I}$ with the same surrogate faces. However, different from competing techniques from the literature (such as [26] or [9]), these surrogate faces are not generated through image or model-parameter averaging, but are synthesized with a generative neural network (GNN) and, therefore, bear no visual similarities with the original images from $\mathcal{I}$. More importantly, potentially useful information of the original images is preserved with k-Same-Net by exposing a set of appearance-related parameters at the input side of the GNN that affect the visual characteristics of the synthesized images (e.g., facial expression, age, gender, etc.) but not the identity. This approach differs significantly from competing solutions in this field, as useful information is "added back" to the deidentified images (as needed), instead of preserving parts of the appearance of the original images explicitly.

3.2. Deidentification with k-Same-Net

Consider a set of N input images $\mathcal{I} = \{I_1, \ldots, I_N\}$ belonging to N identities and a second (proxy) set of M images $\mathcal{P} = \{P_1, \ldots, P_M\}$ corresponding to Q identities, where $M, Q \geq N$. Furthermore, assume that our goal is to map the images in $\mathcal{I}$ to a target set of deidentified images $\mathcal{D} = \{D_1, \ldots, D_N\}$ in such a way that no relation between the subjects in $\mathcal{I}$ and the images in $\mathcal{D}$ can be established without ambiguity.

With the original k-Same algorithm, clusters of k images are generated from $\mathcal{I}$ based on image similarities and used to define surrogate faces D_i (for $i = 1, \ldots, N$) for deidentification (see Section 2 for details). It is straight forward to extend this approach to proxy clusters defined over $\mathcal{P}$ as long as the number of generated clusters for the image sets $\mathcal{I}$ and $\mathcal{P}$ is the same and a one-to-one correspondence between the clusters is established. Replacing all k images of each cluster of $\mathcal{I}$ with the same surrogate images achieves so-called k-anonymity, where linking a deidentified image to one of the identities in $\mathcal{I}$ is limited to a guess with a success probability of $1/k$, regardless of how the surrogate faces are defined[1]. It is, therefore, possible to compute surrogate faces for deidentification from a proxy image set $\mathcal{P}$, that can in general contain an arbitrary number identities, Q, as long as the same number of clusters can be computed as for the image set $\mathcal{I}$, i.e., $Q \geq N$. Under these conditions, the proxy set $\mathcal{P}$ can be used as the training set for the GNN, and synthetic face images produced by the GNN can serve as surrogate faces for deidentification.

It needs to be noted that surrogate faces could also be generated based on a single identity from the proxy set $\mathcal{P}$ for each cluster of $\mathcal{I}$. However, for practical reasons, we generate the synthetic images of the k-Same-Net approach based on multiple (i.e., k) identities and, therefore, deal only with artificial identities and not synthetic images of real people present in our training data.

3.3. k-Same-Net and Data Utility

While the main goal of deidentificaton is facilitating data anonymity, the current trend in this area is to also ensure suitable levels of data utility after deidentification. With k-Same-Net we preserve (or better said retain) some of the information content in the original images by probing the input images for certain characteristics and then feed the results of this probing procedure to the GNN to generate images in accordance with the identified (or desired) characteristics. For example, if our goal is to preserve facial expressions after deidentification, we first recognize the facial expressions in the input data and then generate the surrogate faces in accordance with the identified expressions. This procedure is performed for each image separately, so after deidentification images from the same cluster of $\mathcal{I}$ are deidentified with a surrogate face of the same target identity but may differ in terms of facial expressions. Such an approach allows us to devise selective deidentification schemes tailored towards specific target applications, where only certain visual attributes of the input images are preserved, while others are removed completely.

3.4. GNN Architecture and Algorithm Summary

The main component of our k-Same-Net approach is the generative neural network (GNN) recently proposed in [4] and later extended for face synthesis by M.D. Flynn[2]. The network consist of a hierarchy of fully-connected and deconvolutional layers and once trained is able to generate synthetic surrogate faces D given: *i)* information about the k identities in the relevant proxy cluster of $\mathcal{P}$ encoded in the vector $\mathbf{x}$, and *ii)* information about the non-identity related appearance characteristics of D encoded in the appearance-parameter vector $\mathbf{y}$:

$$D = \text{GNN}(\mathbf{x}, \mathbf{y}). \tag{1}$$

Training of the GNN requires an appropriately annotated training set with labels spanning all appearance characteristics that need to be altered during image generation. While there is no strict limit with respect to the number of input parameters and appearance variations our GNN can handle, it is necessary that suitable labels exist for all images present in the training set.

A summary of the complete k-Same-Net algorithm is given by Algorithm 1.

4. Experiments and Results

In this section we present qualitative experiments to demonstrate the feasibility of our k-Same-Net deidentification approach. We first discuss the datasets used for network

1. The reader is referred to [26] for a formal proof.

2. https://zo7.github.io/blog/2016/09/25/generating-faces.html

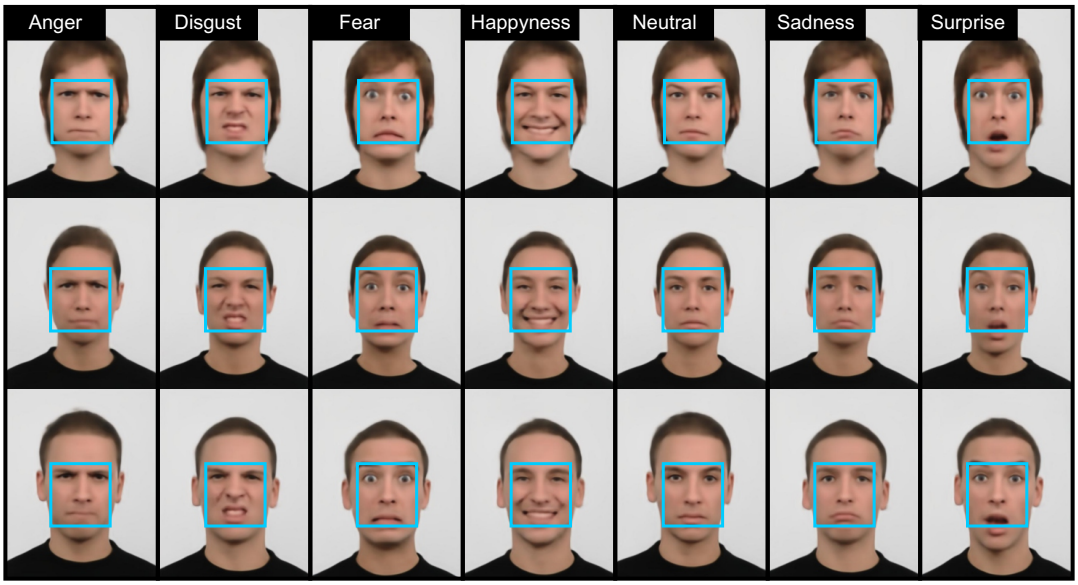

Figure 4: Examples of synthetic images generated by the generative neural network (GNN). The GNN can produce various facial expressions for every identity. Each synthesized face shown is a mixture of k identities from the training (or proxy) set with $k = 2$ for the presented examples. Note that all images appear natural and show no visible artifacts (such as ghosting or other non-naturally looking patterns).

Algorithm 1: k-Same-Net

Input : Input image set $\mathcal{I}$, proxy image set $\mathcal{P}$, parameter k, trained GNN
Output: Deidentified image set $\mathcal{D}$

Compute clusters of k-images from $\mathcal{I}$
Compute clusters of k-identities from $\mathcal{P}$
Define correspondence f between clusters of $\mathcal{I}$ and $\mathcal{P}$
for *each image* $I_i \in \mathcal{I}$ **do**
    Encode k-identities from $\mathcal{P}$ in identity vector $\mathbf{x}$ based on cluster correspondence f
    Define appearance-parameter vector $\mathbf{y}$ based on relevant recognition procedure applied to I_i
    Generate deidentified surrogate face D_i based on GNN and add to $\mathcal{D}$
end

training and experimentation and then present deidentification examples that illustrate some of the key characteristics of k-Same-Net.

4.1. Datasets

To train the GNN needed for k-Same-Net we use the RaFD dataset [18]. RaFD is a high quality image dataset, containing 67 subjects with 8 different facial expression (anger, disgust, fear, happiness, sadness, surprise, contempt and neutral). Each subject is captured under three different gaze directions and from five camera angles with all 8 facial expressions. From these images we select only frontal images displaying 57 adult subjects for our training procedure, resulting in a training set size of 456 facial images.

The RaFD dataset is highly suitable for training the generative network for k-Same-Net since it includes aligned high-quality facial imagery taken in a controlled environment with uniform background and most importantly because it ships with facial-expression annotations that can be used to demonstrate some of the advantages of our deidentification technique.

To evaluate our deidentification approach we use the XM2VTS dataset [24]. We take only a few images/frames from this dataset to illustrate how the k-Same-Net compares to other deidentification techniques and how neural-network-based deidentification can be applied to facial images.

4.2. Network Training

We train our GNN with images from the RaFD dataset and rely on the implementation of M.D. Flynn available from GitHub[3] for the training. The network is trained with back-propagation using stohastic gradient descent and the Adam optimizer. The batch size is set to 16, the learning rate to 0.001 and the number of epochs is limited to 500 [23]. The training requires approximately 24 hours on a desktop CPU with 32GB of ram and a TitanX GPU.

Once the network is trained, it is able to output facial images of artificial identities with various facial expressions as seen in the examples in Fig. 6. The synthesized images shown here are generated by mixing two identities from the training set (or in other words, selecting $k = 2$), so

3. https://github.com/zo7/deconvfaces

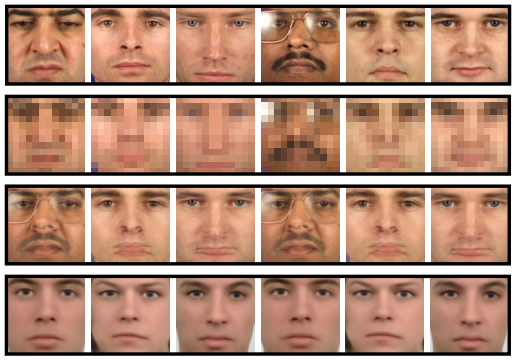

Figure 5: Deidentification results (from top to bottom): the original images, pixelated images, the k-Same algorithm ($k = 2$), and the k-Same-Net approach ($k = 2$).

none of the depicted subjects represents a real person. The blue boxes show the image area that represents the surrogate faces, D_i, needed for deidentification.

4.3. Deidentification Examples

We now present a few illustrative deidentification results based on a selection of images from the XM2VTS dataset. In Fig. 5 the top row shows original XM2VTS images that form our input set $\mathcal{I}$ and the second row shows the same set of images in the same orther but deidentified with a naive approach, where the images in $\mathcal{I}$ are simply down-sampled to hide the identities. Note that while some facial feature are concealed, a simple imitation attack may suffice to successfully link the naively deidentified images to the originals. The third row of images shows sample results for the original k-Same algorithm for $k = 2$. Here, anonymity is guaranteed and it is impossible to link any image from the deidentified image set to the originals with a probability higher than 0.5 (i.e., $1/k = 1/2$). However, the quality and naturalness of the deidentified images is questionable as artifacts appear due to misalignment of the original images. Our approach, shown in the last row of Fig. 5, also comes with anonymity guarantees and produces natural and realistic deidentification results without artifacts.

A key characteristic of the k-Same-Net approach is the possibility of preserving specific non-identity related information (such as facial expressions) of the original images, while concealing facial identity cues with the artificially generated surrogate faces. Fig. 6 illustrates this characteristic on another set of images from the XM2VTS dataset. Here, the top row again shows original images with two new images that were not present in Fig. 5 that are highlighted red. These two images form a cluster and are, therefore, replaced with the same surrogate face images during deidentification, as shown in the second row of Fig. 5, where our k-Same-Net approach was applied, but the same appearance-parameter vector $\mathbf{y}$ (see Eq. (1)) was used for all images. The last row shows an oracle-type of experiment under the assumption that the facial expression of the input images are known and illustrates how facial expressions can be preserved if needed.

Figure 6: Preserving data utility. The top row shows the original images, the second row shows images deidentified with k-Same-Net without preserving any of the input information and the last row shows k-Same-Net results where the facial expression of the originals was retained in the deidentified images. Note that the two images that are marked red belong to the same cluster ($k = 2$) and have been deidentified in the last row using the same artificial target identity, but a different facial expression.

The two images from the input set that are marked with red are still deidentified using the same artificial target identity, but now exhibit different facial expressions. Under the assumption that the facial expressions cannot be exploited for identity inference, the upper bound on the reidentification performance still equals to $1/k$.

Note that in practical implementations the analytical part of the k-Same-Net approach that estimates the properties and visual characteristics of the input images would have to be implemented with existing techniques. Due to recent advancements in the areas of facial landmarking [15], [37], [33], [29], [13], [38], facial expression recognition [14], [3], [25], [1], [28], [6], age estimation [39], [32], [11], [27], [17], gender recognition [12], [22] and a like [19], [31], there are several techniques with open-source implementations available that can be used for this task. To improve the analytical part of our approach, an additional analysis of other biometric traits could be added (e.g. soft biometrics [5]).

5. Conclusion

In this work we combined generative neural networks (GNNs) and a formal privacy protection scheme to perform face deidentification. The approach entitled k-Same-Net utilizes a GNN to generate synthetic face images for deidentification, but enables preservation of selected non-identity related features (such as facial expressions, gender or age). Our future work will focus on assessing issues related to contextual information, which is known to degrade the effectiveness of deidentification. Since generative models represent the current state-of-the-art in the field of content generation, we expect that more generative approaches will be seen in the field of deidentifcation following a similar methodology.

Acknowledgments

This research was supported in parts by the ARRS (Slovenian Research Agency) Research Programme P2-0250 (B) Metrology and Biometric Systems, the ARRS Research Programme P2-0214 (A) Computer Vision and the Croatian HRZZ as part of the DePPSS 6733 project on De-identification for Privacy Protection in Surveillance Systems.

References

[1] S. Afshar and A. Ali Salah. Facial expression recognition in the wild using improved dense trajectories and fisher vector encoding. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2016.

[2] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. *CoRR*, abs/1606.03657, 2016.

[3] A. Dapogny, K. Bailly, and S. Dubuisson. Pairwise conditional random forests for facial expression recognition. In *The IEEE International Conference on Computer Vision (ICCV)*, December 2015.

[4] A. Dosovitskiy, T. J. Sprigenberg, and B. T. Learning to generate chairs with convolutional neural networks. In *CVPR*, pages 1538–1546. IEEE, 2015.

[5] Z. Emersic, V. Struc, and P. Peer. Ear recognition: More than a survey. *Neurocomputing*, in press:1–22, 2017.

[6] R. Gajšek, V. Štruc, S. Dobrišek, and F. Mihelič. Emotion recognition using linear transformations in combination with video. In *Speech and intelligence: proceedings of Interspeech 2009*, pages 1967–1970, Brighton, UK, 2009.

[7] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial networks. *arXiv:1406.2661*, 2014.

[8] R. Gross, E. Airoldi, B. Malin, and L. Sweeney. Integrating utility into face de-identification. In *Proceedings of the 5th International Conference on Privacy Enhancing Technologies*, PET'05, pages 227–242, Berlin, Heidelberg, 2006. Springer-Verlag.

[9] 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)*, pages 161–161, June 2006.

[10] R. Gross, L. Sweeney, F. de la Torre, and S. Baker. Semi-supervised learning of multi-factor models for face de-identification. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2008.

[11] Z. Huo, X. Yang, C. Xing, Y. Zhou, P. Hou, J. Lv, and X. Geng. Deep age distribution learning for apparent age estimation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2016.

[12] S. Jia and N. Cristianini. Learning to classify gender from four million images. *Pattern Recognition Letters*, 58:35 – 41, 2015.

[13] A. Jourabloo and X. Liu. Large-pose face alignment via cnn-based dense 3d model fitting. In *Proc. IEEE Computer Vision and Pattern Recogntion*, Las Vegas, NV, June 2016.

[14] H. Jung, S. Lee, J. Yim, S. Park, and J. Kim. Joint fine-tuning in deep neural networks for facial expression recognition. In *The IEEE International Conference on Computer Vision (ICCV)*, December 2015.

[15] V. Kazemi and J. Sullivan. One millisecond face alignment with an ensemble of regression trees. In *CVPR*, 2014.

[16] D. P. Kingma and M. Welling. Auto-encoding variational bayes. *CoRR*, abs/1312.6114, 2013.

[17] J. Konda and P. Peer. Estimating peoples age from face images with convolutional neural networks. In *25th International Electrotechnical and Computer Science Conference*, September 2016.

[18] O. Langner, R. Dotsch, G. Bijlstra, D. Wigboldus, S. Hawk, and A. van Knippenberg. Presentation and validation of the radboud faces database. *Cognition&Emotion*, 24(8):1377–1388, 2010.

[19] G. Levi and T. Hassner. Age and gender classification using convolutional neural networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2015.

[20] N. Li, T. Li, and S. Venkatasubramanian. t-closeness: Privacy beyond k-anonymity and l-diversity. In *2007 IEEE 23rd International Conference on Data Engineering*, pages 106–115, April 2007.

[21] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam. L-diversity: Privacy beyond k-anonymity. *ACM Trans. Knowl. Discov. Data*, 1(1), Mar. 2007.

[22] J. Mansanet, A. Albiol, and R. Paredes. Local deep neural networks for gender recognition. *Pattern Recognition Letters*, 70:80 – 86, 2016.

[23] B. Meden, R. C. Mallı, S. Fabijan, H. K. Ekenel, V. Struc, and P. Peer. Face deidentification with generative deep neural networks. *IET Signal Processing*, 2017.

[24] K. Messer, J. Kittler, M. Sadeghi, S. Marcel, C. Marcel, S. Bengio, F. Cardinaux, C. Sanderson, J. Czyz, L. Vandendorpe, S. Srisuk, M. Petrou, W. Kurutach, A. Kadyrov, R. Paredes, B. Kepenekci, F. B. Tek, G. B. Akar, F. Deravi, and N. Mavity. Face verification competition on the xm2vts database. In *Proceedings of the 4th International Conference on Audio- and Video-based Biometric Person Authentication*, AVBPA'03, pages 964–974, Berlin, Heidelberg, 2003. Springer-Verlag.

[25] A. Mollahosseini, B. Hasani, M. J. Salvador, H. Abdollahi, D. Chan, and M. H. Mahoor. Facial expression recognition from world wild web. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2016.

[26] E. M. Newton, L. Sweeney, and B. Malin. Preserving privacy by de-identifying face images. *IEEE Transactions on Knowledge and Data Engineering*, 17(2):232–243, Feb 2005.

[27] Z. Niu, M. Zhou, L. Wang, X. Gao, and G. Hua. Ordinal regression with multiple output cnn for age estimation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[28] X. Peng, Z. Xia, L. Li, and X. Feng. Towards facial expression recognition in the wild: A new database and deep recognition system. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2016.

[29] X. Peng, S. Zhang, Y. Yang, and D. N. Metaxas. Piefa: Personalized incremental and ensemble face alignment. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 3880–3888, Dec 2015.

[30] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *CoRR*, abs/1511.06434, 2015.

[31] R. Ranjan, V. M. Patel, and R. Chellappa. Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. *CoRR*, abs/1603.01249, 2016.

[32] R. Ranjan, S. Zhou, J. Cheng Chen, A. Kumar, A. Alavi, V. M. Patel, and R. Chellappa. Unconstrained age estimation with deep convolutional neural networks. In *The IEEE International Conference on Computer Vision (ICCV) Workshops*, December 2015.

[33] S. Ren, X. Cao, Y. Wei, and J. Sun. Face alignment at 3000 fps via regressing local binary features. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1685–1692, June 2014.

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

[35] Z. Sun, L. Meng, and A. Ariyaeeinia. Distinguishable de-identified faces. In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, volume 04, pages 1–6, May 2015.

[36] L. Sweeney. K-anonymity: A model for protecting privacy. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, 10(5):557–570, Oct. 2002.

[37] Z. Zhang, P. Luo, C. C. Loy, and X. Tang. *Facial Landmark Detection by Deep Multi-task Learning*, pages 94–108. Springer International Publishing, Cham, 2014.

[38] X. Zhu, Z. Lei, X. Liu, H. Shi, and S. Z. Li. Face alignment across large poses: A 3d solution. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[39] Y. Zhu, Y. Li, G. Mu, and G. Guo. A study on apparent age estimation. In *The IEEE International Conference on Computer Vision (ICCV) Workshops*, December 2015.