# IDENTITY GENERATION WITH DEEP GENERATIVE MODELS

NUNZIO ALEXANDRO LETIZIA
PiktID, Klagenfurt, Austria
nunzio.letizia@piktid.com

**Abstract** Deep generative models have drawn the attention of the AI community in the last decade. The scalability of neural architectures helps solving multiple relevant problems, e.g., text-to-image generation, otherwise not addressable. In the context of image data privacy, the increasing amount of produced, shared, and stored images imposes new measures to protect personal identity information. At the same time, such protection mechanisms need to preserve the image quality. In this talk, we discuss how PiktID is using some recent deep learning-based techniques for protecting human identities in pictures. We show several examples, and we present interesting use-cases.

**Keywords:**
deep learning,
anonymization,
generative models,
image privacy,
data protection

University of Maribor Press

DOI https://doi.org/10.18690/um.feri.4.2023.3
ISBN 978-961-286-720-1

# Generiranje identitet z globokimi generativnimi modeli

Nunzio Alexandro Letizia

PiktID, Celovec, Avstria
nunzio.letizia@piktid.com

**Sinopsis** Globoki generativni modeli so v zadnjem desetletju pritegnili pozornost skupnosti AI. Razširljivost nevronskih arhitektur pomaga pri reševanju več pomembnih problemov, npr. ustvarjanje besedila v sliko, ki sicer ni naslovljivo. V kontekstu zasebnosti slikovnih podatkov vse večja količina proizvedenih, deljenih in shranjenih slik nalaga nove ukrepe za zaščito informacij o osebni identiteti. Hkrati morajo takšni zaščitni mehanizmi ohraniti kakovost slike. V tem govoru razpravljamo o tem, kako PiktID uporablja nekatere nedavne tehnike, ki temeljijo na globokem učenju, za zaščito človeških identitet na slikah. Prikazujemo več primerov in predstavljamo zanimive primere uporabe.

**Ključne besede:**
globoko učenje,
anonimizacija,
generativni modeli,
zasebnost slike,
zaščita podatkov

https://doi.org/10.18690/um.feri.4.2023.3 DOI
978-961-286-720-1 ISBN

Univerzitetna založba
Univerze v Mariboru

## 1 Introduction

Anonymization is the process of removing personal identifiable information from data, in our case, images. In particular, we will mostly refer to sensitive data as the biometric information in the form of facial and body features.

PiktID aims at substituting biometric information available in images with new synthetic one not belonging to any human. The objective is to ensure security and enhance private protection, thus inhibiting re-identification of the original subjects from anonymized data.

However, to reach the final full anonymization goal, several technological and technical challenges need to be firstly defined, addressed and lastly solved.

## 2 Methodology

Two distinct classes of anonymization solutions have been developed in literature. Classical image processing-based and deep-learning based. The former solutions aim at detecting the sensitive subject and cover the information typically with filters such as Gaussian or bounding boxes applied on specific regions (that can also be defined via segmentation).

However, these techniques have the major flaw of being immediately visible to the human eye, thus significantly decreasing the quality, and of not preserving the statistical information that in some applications may be relevant (e.g., extracting statistical information of a shop's visitors being GDPR compliant) (Yang 2022).

The latter approach, instead, has recently received attention and some initial works based on generative models have been already proposed but only applied to faces (Zhongzheng, 2018). In particular, it is worth mentioning DeepPrivacy (Håkon, 2019) and CFA-NET (Tianxiang, 2021) since they both use Generative Adversarial Networks (GANs) (Goodfellow, 2014) and the second one also uses StyleGAN (Karras, 2020).

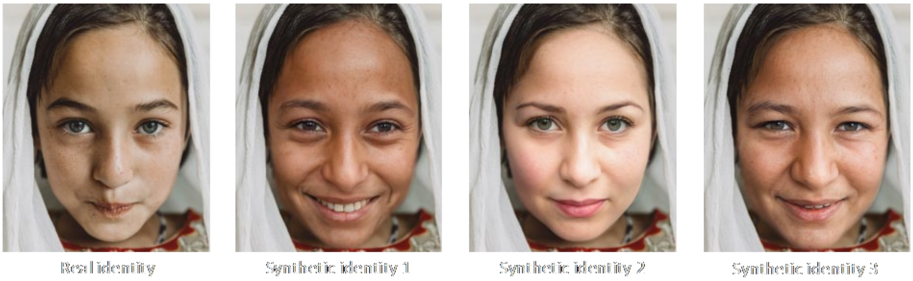

**Figure 1: Examples from the current PiktID face anonymization tool. Quality is preserved.**
Source: own.

Scaling GANs to hair and body anonymization is partially an unsolved task and it seems that diffusion models (Jonathan, 2020) is the most suitable generative model in this scenario. In fact, diffusion models have shown a more versatile generation process compared to GANs, enabling the generation of more complex and diverse images, thanks to the guidance of a language model. However, they still lack precision in details (high-frequency components) and the research community is working towards refinement of models such as Stable Diffusion.

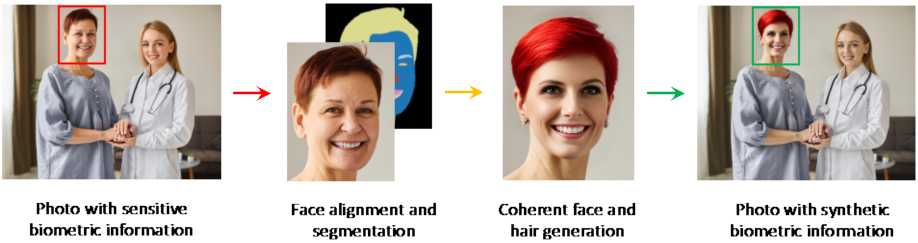
Photo with sensitive biometric information
Face alignment and segmentation
Coherent face and hair generation
Photo with synthetic biometric information

**Figure 2: PiktID workflow for face and hair anonymization. The biometric information is extracted from the vulnerable subject and it is replaced with synthetic one.**
Source: own.

When substituting faces, hair or the full body with synthetic biometric information, it often happens that a mismatch occurs and a background prediction block is needed. For instance, suppose to change the hairstyle in a way that the new hair is much shorter than the original one. The initially covered regions, that are now substituted, require a background prediction block that needs to understand the scene and coherently fill in the region based on the surroundings. Due to the variety of possible scenes, diffusion models can again be exploited to essentially inpaint the missing content with new semantically meaningful one.

It is worth noticing that developing an accurate background prediction technology also enables an extra feature in the product: the possibility in some cases to even remove the entire person (face, hair and body) from the picture in the sensitive region. According to the application, such a solution may be appropriate. Case 1: people in an event, full anonymization is desired since the objective there is to show the affluence protecting personal information. Case 2: undesired person in a group of people, complete cancellation of the person in the picture may be preferred. In the picture below, it is shown an example of cancellation with background prediction. Notice that the background prediction technology can be exploited also to produce partial parts surrounding the body.

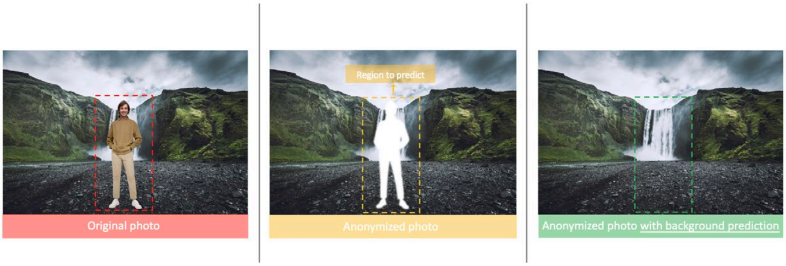

**Figure 3: Example of the proposed background prediction with diffusion models applied to cancellation.**

Source: own.

## References

Yang, Kaiyu and Yau, Jacqueline and Fei-Fei, Li and Deng, Jia and Russakovsky, Olga (2022). “A Study of Face Obfuscation in ImageNet”. *International Conference on Machine Learning (ICML)*

Zhongzheng Ren and Yong Jae Lee and Micheal S. Ryoo (2018). “Learning to Anonymize Faces for Privacy Preserving Action Detection”. *European Conference on Computer Vision (ECCV)*

Håkon Hukkelås, Rudolf Mester, Frank Lindseth (2019). “DeepPrivacy: A Generative Adversarial Network for Face Anonymization”. *Internation Symposium on Visual Computing (ISVC)*

Tianxiang Ma, Dongze Li, Wei Wang and Jing Dong (2021). “Face Anonymization by Manipulating Decoupled Identity Representation”. *CoRR*

Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. (2014) “Generative adversarial nets”. *Advances in neural information processing systems. p. 2672–80*

T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen and T. Aila (2020). "Analyzing and Improving the Image Quality of StyleGAN," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*

Jonathan Ho and Ajay Jain and Pieter Abbeel (2020). “Denoising Diffusion Probabilistic Models”, *Advances in Neural Information Processing Systems*