Enhancing Gender Privacy
with Photo-realistic Fusion of Disentangled Spatial Segments

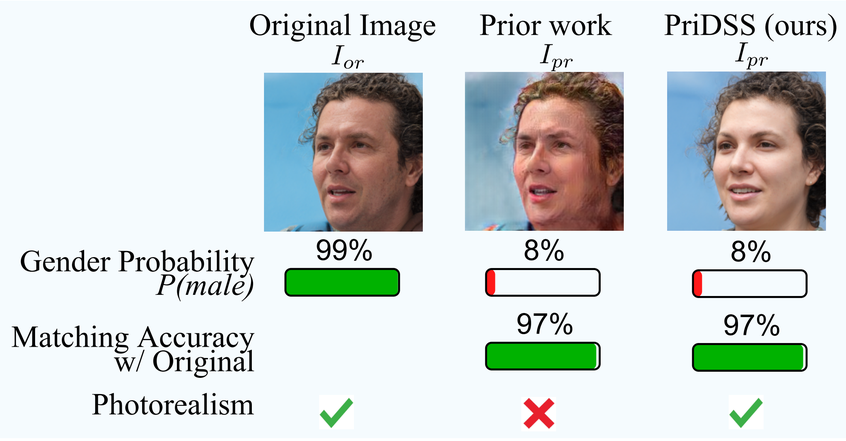
Peter Rot, Janez Križaj, Peter Peer, Vitomir Štruc

University of Ljubljana

SCAN ME

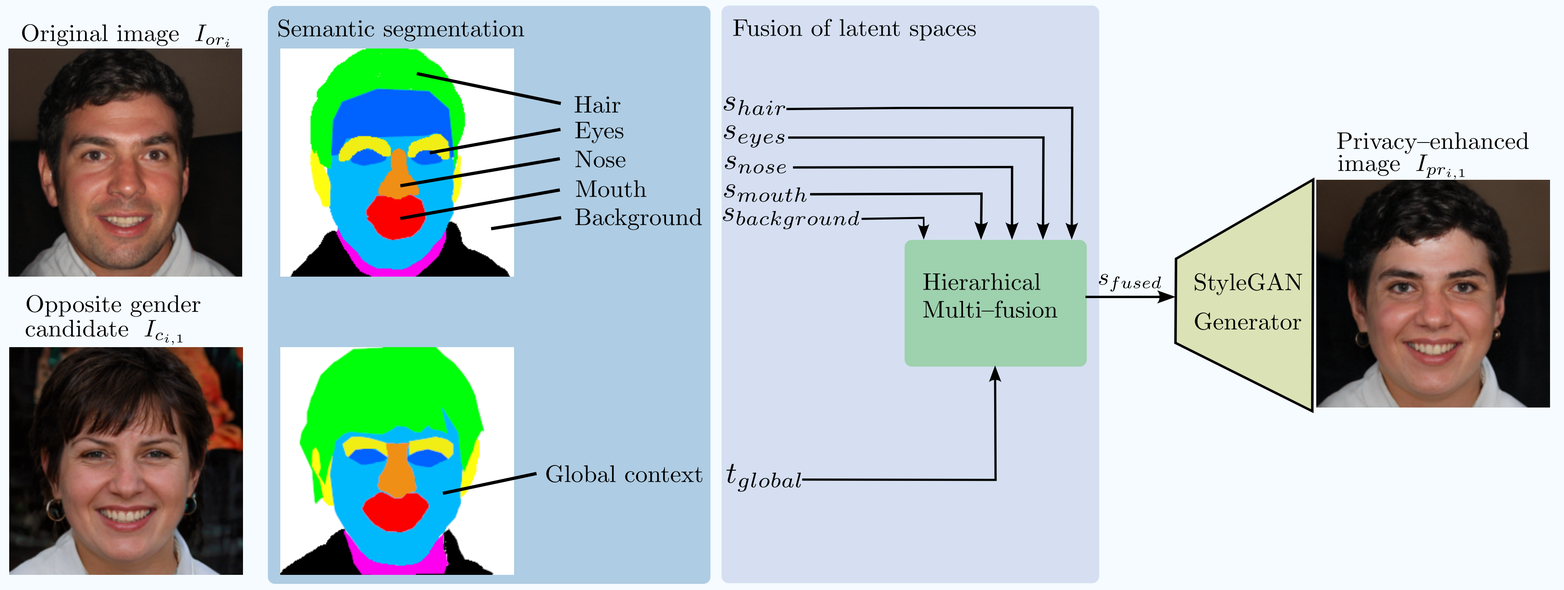
Introduction

- Besides identity, faces carry information about soft biometrics (e.g. gender, ethnicity, age, etc.), which are considered sensitive personal information.
- To disable the extraction of gender but still maintain usability of images for face verification, we propose PriDSS.
- The main advantage of PriDSS over existing methods is the assured photorealism of privacy-enhanced images.



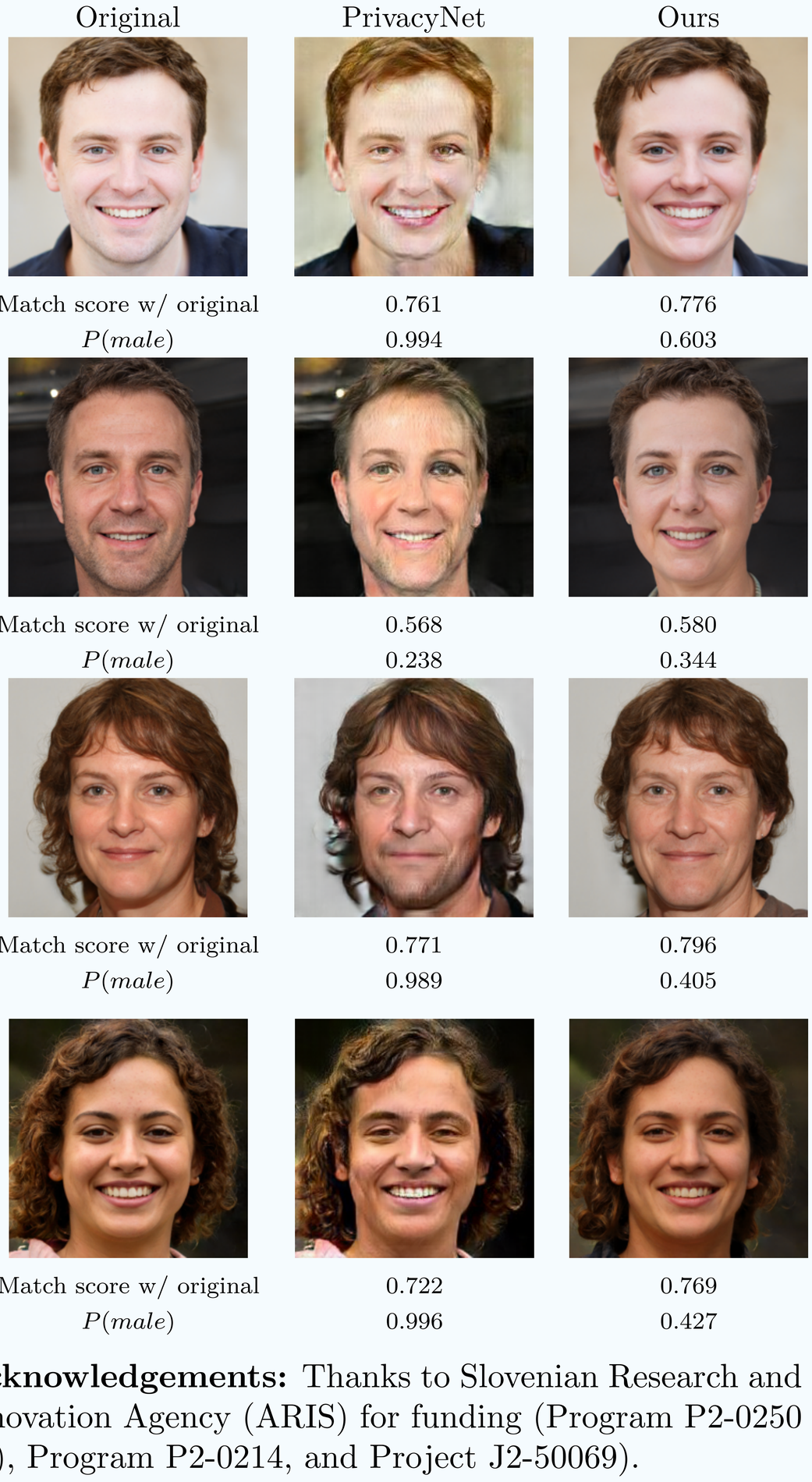
Methodology

- Discriminative identity information primarily resides in eyes, nose, and mouth.
- Other regions (chin, cheeks, etc.) constitute *global context*, containing soft biometrics.
- PriDSS combines (i) identity-related information with (ii) global context from the opposite gender.
- Information fusion occurs in the disentangled StyleGAN space, from which an image is generated using StyleGAN generator.
  By considering multiple images of the opposite gender, PriDSS generates many candidate images from which the best candidate is selected.



- The best candidate is selected using *privacy-gain identity-loss* coefficient (PIC), defined as $PIC = |2P(male) - 1| - SIM_{score}$, where SIM_{score} denotes cosine similarity with original image.

Qualitative results



Quantitative results

In comparison to the state-of-the-art PrivacyNet, our approach achieves competitve results in terms of identity preservation and privacy enhancement of gender, while assuring 2× better photorealism in terms of FID score. The results are presented in Table 1.

Performance indicator	Original	PrivacyNet	PriDSS (ours)
Gender (AUC)	0.981	0.5400	0.5900
Verification (EER) w/ Original	n/a	0.0070	0.0680
Verification (FNMR@FMR10^{-1}) w/ Original	n/a	0.0005	0.0010
Photo-realism (FID) w/ Original	n/a	57.499	25.386

Table 1: Performance evaluation and SoTa comparison.

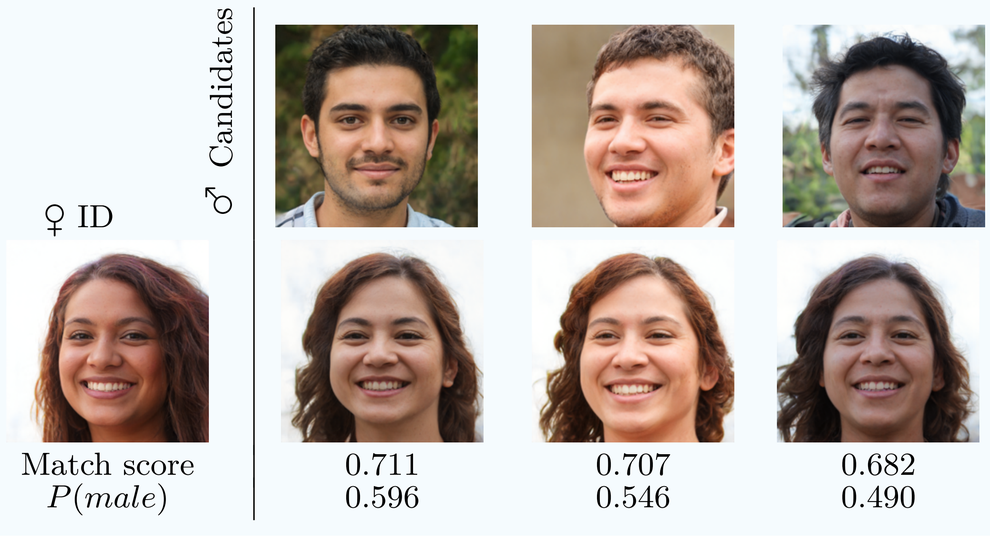
Privacy-enhancing techniques are often susceptible to image restoration attempts, which undermine their effectiveness. We evaluated the robustness to such attempts using PrivacyProber (PP) framework using three PP implementations (D - denoising, I - inpainting, and A - autoencoder). As shown in Table 2, neither of the compared methods is vulnerable to restoration attempts to a high degree.

Input image		PrivacyNet	PriDSS (ours)
Privacy enhanced		0.540	0.590
Recovered with	PP–D	0.545	0.603
	PP-I	0.540	0.602
	PP–A	0.550	0.554

Table 2: Robustness to image restoration attempts.

Ablation study

We explore the impact of candidate-image selection.



Acknowledgements: Thanks to Slovenian Research and Innovation Agency (ARIS) for funding (Program P2-0250 (B), Program P2-0214, and Project J2-50069).

ICASSP 2024 KOREA
14-19. April, Seoul