ImplicitDeepfake
Plausible Face-Swapping through Implicit Deepfake Generation

G. Stanishevskii, J. Steczkiewicz, T. Szczepanik, S. Tadeja, J. Tabor, P. Spurek
Jagiellonian University, Faculty of Mathematics and Computer Science, Kraków, Poland
University of Cambridge, Department of Engineering

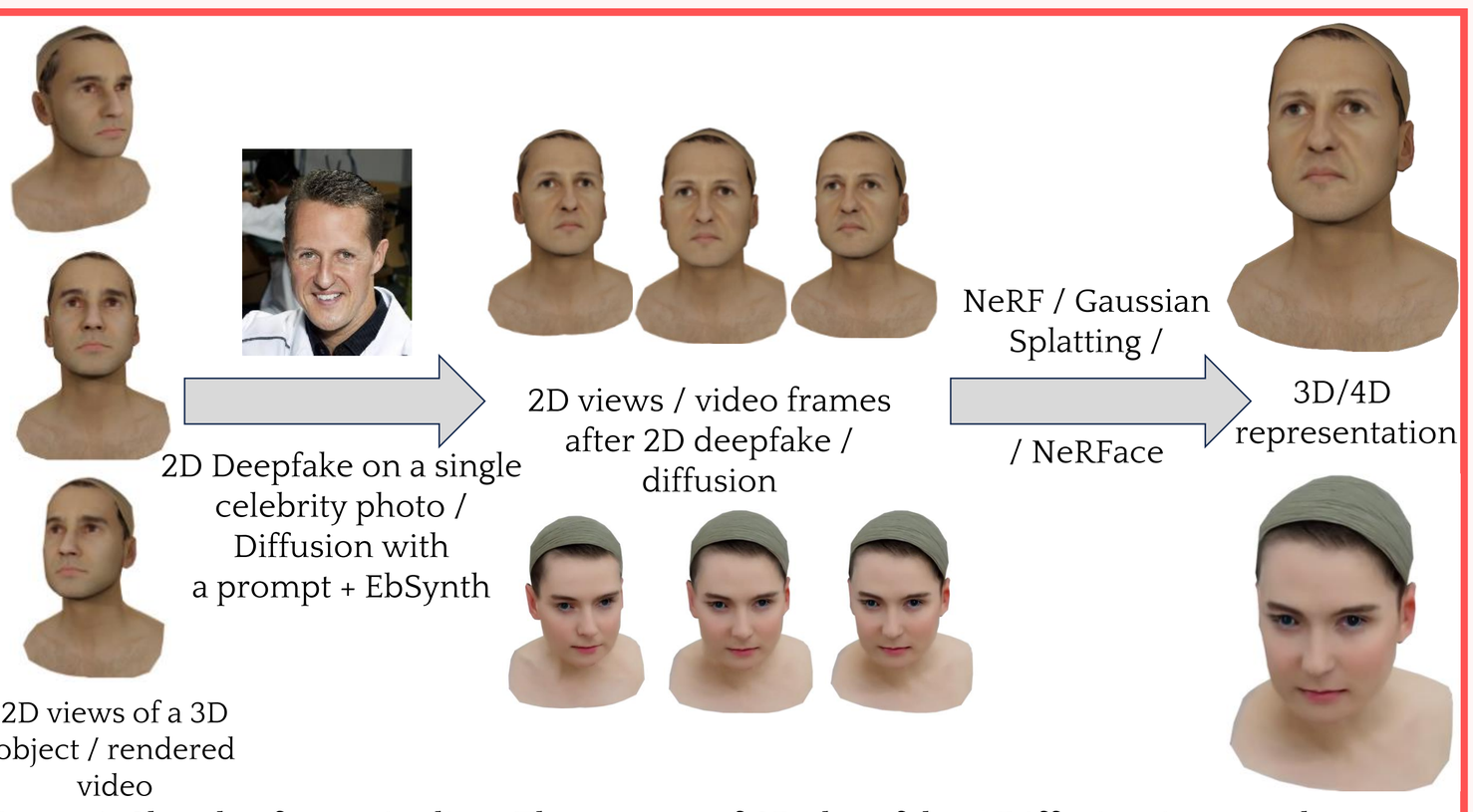

Figure 1. Sketch of our pipeline. The output of 2D deepfake / Diffusion is passed as input to a 3D/4D rendering technique.

Overview

A range of notable advances have been made in areas of deepfake generation, detection and diffusion models in recent years. To extend this growing body of research, we present ImplicitDeepfake - a first model that produces a 3D deepfake, using state-of-the-art machine learning-based rendering methods such as Neural Radiance Fields (NeRFs)[1] and Gaussian Splatting (GS)[2]. Our model can be also extended with NeRFace[4] to create dynamic avatars and change their facial expressions. Ultimately, we can use diffusion models[5] and Example-based Image Synthesis (EbSynth)[6] to modify face avatars with simple text props.

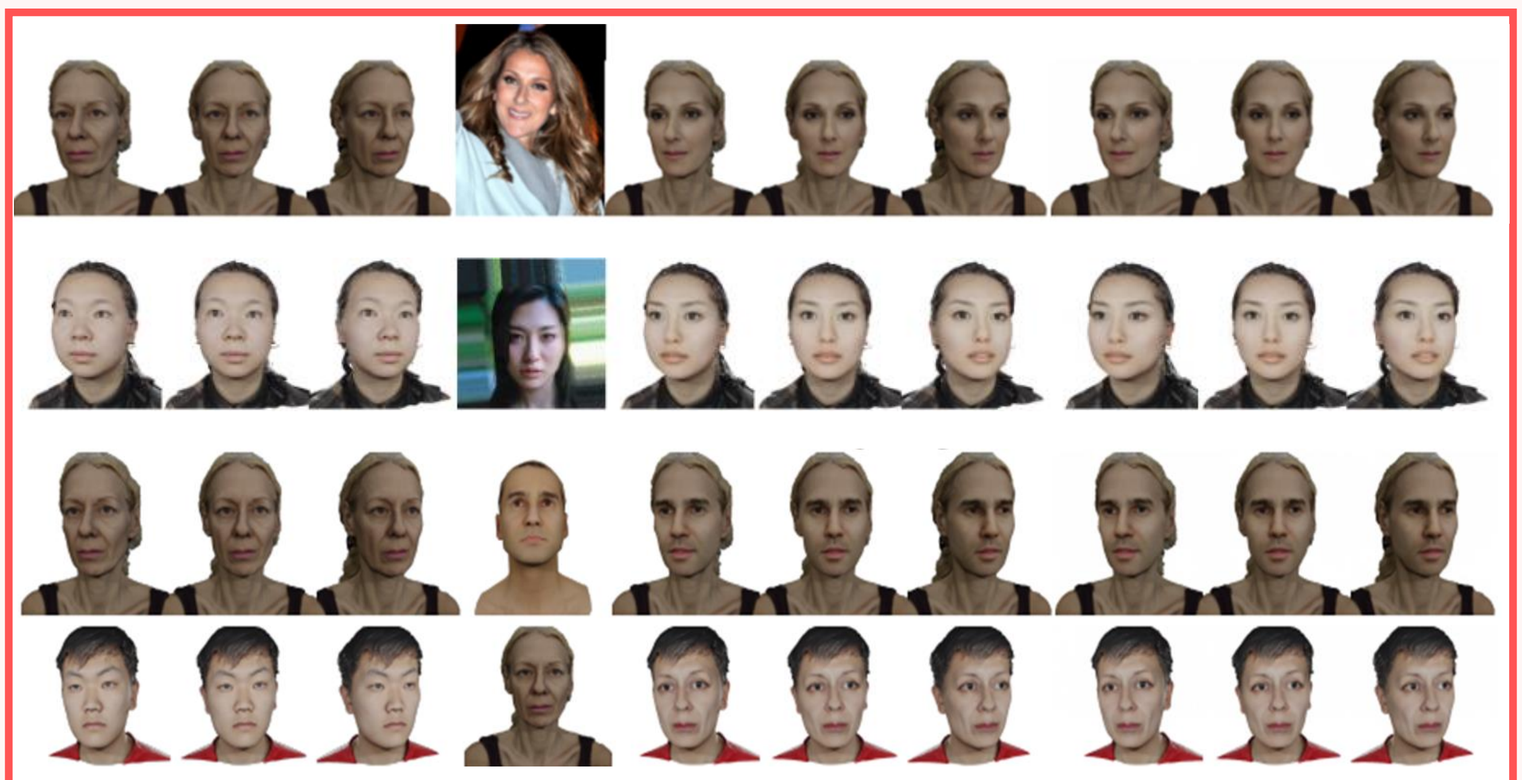

Figure 2 Results of ImplicitDeepfake using NeRF and GS. The first three images show an anonymous 3D face, the fourth is a celebrity photo. The next three images display base poses processed by a classical 2D deepfake, and finally rendered in 3D. The last two rows illustrate our method's ability to handle dissimilar faces, highlighting gender differences.

Our method

Hybrid of the classical deepfake model and neural rendering that can use either NeRF or GS models. In general, we can rely on any other approach to produce novel views from 2D images.

Seamless pipeline

The combination of 2D deepfake / Diffusion and NeRF, GS and NeRFace works perfectly together, resulting in plausible outcomes of noticeably high quality. We should stress that the user can choose any 2D deepfake to achieve similar results. For diffusion, we used Stable Diffusion.

Figure 3. Results of ImplicitDeepfake training using NeRFace: original video frames (first three images), a celebrity photo (fourth image), NeRFace generated results (three next images), and the same images with altered facial expressions.

Reproduction

A set of photos or video with camera positions of a 3D face is passed through deepfake model[3] or diffusion model[5]. In this way, we obtain input for the 3D/4D rendering method using either a single photo of the target person or a given prompt in the case of diffusion.

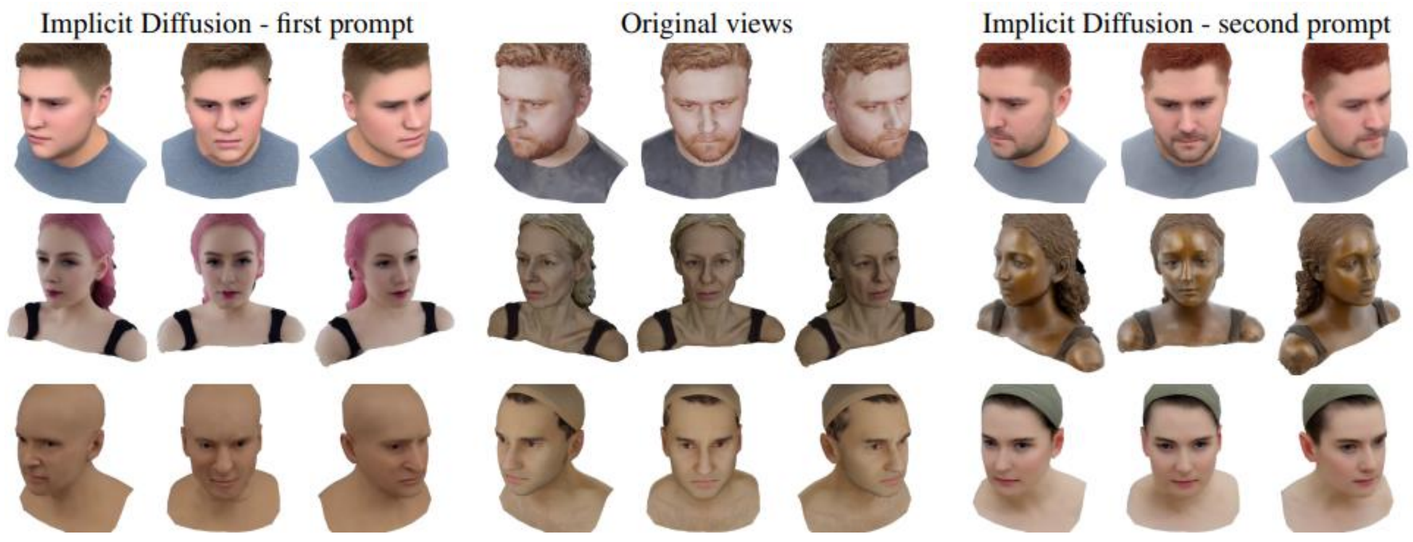

Figure 4. Results of Implicit Diffusion for three different faces. Each row shows the original avatar and two final 3D models generated using two different prompts.

References

[1] Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., and Ng, R. Nerf: Representing scenes as neural radiance fields for view synthesis. In ECCV. 2020.
[2] Kerbl, B., Kopanas, G., Leimkühler, T., and Drettakis, G. 3d gaussian splatting for real-time radiance field rendering. ACM Transactions on Graphics, 42(4), 2023
[3] Groshev, A., Maltseva, A., Chesakov, D., Kuznetsov, A., and Dimitrov, D. Ghost—a new face swap approach for image and video domains. IEEE Access, 10:83452–83462, 2022. doi: 10.1109/ACCESS.2022.3196668.
[4] Gafni, G., Thies J., Zollhöfer, M., Nießner, M. NeRFace: Dynamic Neural Radiance Fields for Monocular 4D Facial Avatar Reconstruction. In CVPR 2021
[5] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Bjorn Ommer. High-resolution image syn- ¨ thesis with latent diffusion models. CoRR, abs/2112.10752, 2021.
[6] Ondˇrej Jamriska, ˇ Sˇ arka Sochorov ´ a, Ond ´ ˇrej Texler, Michal Luka´c, Jakub Fi ˇ ser, Jingwan Lu, Eli Shechtman, and Daniel ˇ Sykora. Stylizing video by example. ´ ACM Transactions on Graphics, 38(4), 2019.