

**DeepFakes (DeepFakes, 2020)** is a classical deep learning autoencoder-based image manipulation method widely used for face swapping on both still images and video sequences. As described above, in the learning phase, two autoencoders with a shared encoder are trained to reconstruct the images of source and target face. To create a fake image, the encoded source image is passed as input to the target image decoder. This software was used to create manipulated videos of celebrities, some of them pornographic (Brandon, 2018). Its presentation on Reddit in 2017 brought wide interests to the dangers of AI supported video manipulation and coined the term deepfake.

**FaceApp (Faceapp, 2020)** is a smartphone app that essentially lets human beings rework pictures of their faces to make them appear younger or older, or change their gender or hairstyle. It uses neural networks to edit the images by letting users bend reality in a few ways, such as by adding a smile, shaving off a few years or applying a beautifying 'hotness' filter. While it does not perform the face swap described above, it uses very similar technology. It became available in February 2017 and thus paved the way for the true deepfake software later in the same year. While originally geared towards photo manipulation, it is now also able to process videos.

**FakeApp (Fakeapp, 2020)** is an open source desktop application that allows the user to create AI manipulated videos. It uses social media, image search engines, and videos to insert someone else's face onto preexisting videos frame by frame. It has been already used for the production of fake news, fake surveillance videos, and malicious hoaxes. So far it is the most widely used software for creating deepfakes. It requires a Windows PC with a powerful graphics card to run properly, a requirement which most gaming PCs fulfill. FakeApp is based on Google's TensorFlow (Abadi et al., 2016) deep learning technology.

**FaceSwap (Kowalski, 2020)** is a traditional computer graphics-based approach for face replacement in videos. In this method, sparse facial landmarks are detected to extract the face region in an image. These landmarks are then used to fit a 3D template model which is projected back onto the target image by minimizing the distance between the projected shape and localized landmarks. Finally, the rendered model is blended with the image and color correction is applied. Unlike the other programs described here, FaceSwap does not rely on machine learning. While FaceSwap is now able to process videos, the results are typically less realistic than those of the AI-based systems.

**Face2Face (Thies et al., 2016)** is a facial reenactment system that transfers the expressions of a person in a source video to another person in a target video, while maintaining the identity of the target person. In this method, faces are compressed into a low-dimensional expression space, where expressions can be easily transferred from the source to the target. While is conceptually similar to FakeApp, it tends to produce manipulated videos of a higher quality. However, by now even better software is available.

### **The second generation of Deepfake software**

As described above, the first generation of deepfake software required a large number of training images to function properly. Consequently, these programs are impractical for creating manipulated videos of an average person. For that reason, most deepfake videos that

were created for entertainment purposes featured famous actors of which many images are publicly available.

However, a new generation of deepfake software no longer has this restriction. This is due to the use of generative adversarial networks (GANs) (The Verge, 2020). GANs are a type of neural network similar to autoencoders. However, in a GAN the detector and the generator networks work against each other. The task of the generator is to create variants of images that are similar but not identical to original inputs by adding random noise. Using these generated images, the detector network, which is also called discriminator in this context, is trained as shown in Figure 6. In this manner, the discriminator network becomes very good at recognizing variants of the same image, such as a face seen from many different angles, even if there are few original images to train from.

Note that unlike the first generation, which contains easy to use software, this program and most of the second generation of deepfake generators are research codes which require considerable technical skill to be used successfully. However, it would certainly be possible to create user-friendly software from them. The most prominent programs making use of this concept for the second generation of deepfake software, are:

**Few-Shot Face Translation GAN (Shaoanlu, 2019)** is an efficient implementation of combined unsupervised image-to-image translation (FUNIT) with spatially-adaptive denormalization (SPADE). It transforms an image from a source domain to look like an image from a target domain using only a single source image and a small set of target images by simultaneously learning to translate between images. This enables a generalization of unseen source and target domains with the following normalization on an input image segmentation layer used to preserve semantic information. As the name indicates, this program specializes in implementing the concept of training from few samples using GANs, as described above.

**NeuralTextures (Thies et al., 2019)** is a GAN based facial reenactment technique. In this method, a generative model is trained to learn the neural texture of a target person using original video data. The objective of the GAN is a combination of adversarial and photometric reconstruction loss. This software places special emphasis on analyzing and copying skin texture. As a result, skin created through the face swap looks much more realistic, thus making the manipulation harder to detect.

**FaceSwap-GAN (Shaoanlu, 2020)** is an advanced approach based on denoising autoencoder combined with adversarial losses and attention mechanisms for face swapping. It employs perceptual loss to improve direction of eyeballs to be more realistic and consistent with input face and attention mask that helps with handling occlusion, eliminating artifacts, and producing natural skin tone. FaceSwap-GAN is a highly sophisticated code that combines the advantages of the previous two systems. It also automatically corrects unrealistic looking eye positions.

**Identity-aware CycleGAN (Fang et al., 2020)** is the most recent and advanced face photosketch synthesis approach taking into consideration tiny image nuances in order to produce identity-capable faces. The model used applies a perceptual loss to supervise the image

generation network. It improves photo-sketch synthesis by paying more attention to the synthesis of key facial regions, such as eyes and nose, which are important for identity recognition. In practical terms, it combines the capabilities of FaceSwap-GAN and Face2Face, and it produces the most realistic looking DeepFake videos so far

**StyleGAN (Karras et al., 2019)** is an alternative generator architecture for generative adversarial networks, inspired by the existing image style transfer methods, enables automatic learning and unsupervised separation of high-level attributes, including pose and human face identity and stochastic variation in the generated images, e.g., freckles, hair, with the intuitive scale-specific control of the faked image synthesis process. The StyleGAN generator not only shows high distribution quality metrics, but also demonstrates better interpolation properties, and disentangles the latent factors of face variations.

**Pixel2Style2Pixel Framework (Richardson et al., 2020)** is a generic image-to-image style translation framework, which is based on a novel encoder network that directly generates a series of style vectors which are used in a pre-trained StyleGAN generator. This enables direct embed real images into discovered style space with no requirements for additional manual optimization and providing improved performance in the reconstruction of a fake image. The framework demonstrated its ability to align a face image to a frontal pose without any labeled data, generate multi-modal results for ambiguous tasks such as conditional face generation from segmentation maps, and construct high-resolution images from corresponding low-resolution and bad-quality images.

**InterFaceGAN (Shen et al., 2020)** is a novel framework for semantic face editing by interpreting the latent semantics learned by GANs utilizing different semantics encoded in the latent space of GANs for photo-realistic face-image synthesis leading to more precise control of facial attributes. The framework allows for manipulating gender, age, expression, the presence of eyeglasses and the face pose, and is able to fix the artifacts accidentally generated by GAN models. The proposed method achieves superior quality in real image manipulation and face synthesis. It spontaneously brings a disentangled and controllable facial attribute representation.

**StyleRig (Tewari et al., 2020)** is an advanced extension of the StyleGAN generator that, in addition to generated photo-realistic portrait images of faces with eyes, teeth, hair and context (neck, shoulders, background), enables a rig-like control over semantic face parameters that are interpretable in 3D, such as face pose, expressions, and scene illumination. The StyleRig extension employs a combination of three-dimensional morphable face models that offer control over the semantic parameters with the powerful StyleGAN-based faces generation that opens the possibility for real-3D-world faked-face creation, e.g., relief masks, artificially-animated human-like robots and even real human faces (Crystal et al., 2020).