StyleGAN-Structure and Applications

Lucas Fabian Naumann

TU Dresden

lucas_fabian.naumann@mailbox.tu-dresden.de

SUMMER SCHOOL 2023

ai.uni-jena.de

Introduction

StyleGAN [1] is a generative model introduced in 2018 that gained a lot of attention at that time due to its unmatched image quality, especially of human faces. Since then, generative models have witnessed rapid advancements, new diffusion and autoregressive models used in e.g., DALL-E or Midjourney pushed the boundary in image quality even further. Although StyleGAN is not the best model for photorealistic image generation anymore, its unique structure allows fine-grained control over the generated images making it still valuable for a range of state-of-the-art applications. In the remainder of this poster, first, an overview of the StyleGAN architecture is given, then its usage is demonstrated in four selected applications.

(a) (b)

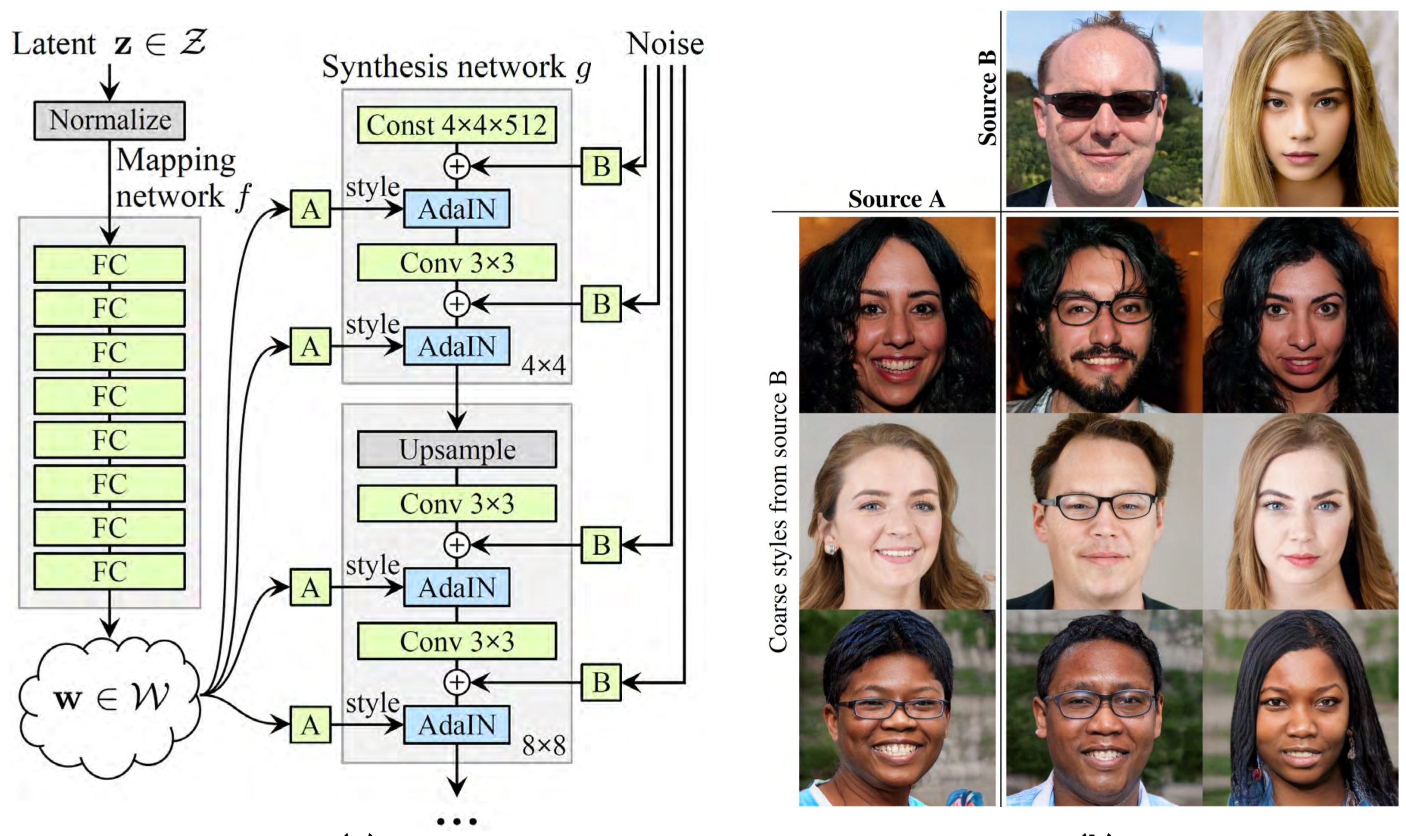
Fig. 1: GAN generated image by a network from 2014 **(a)** [1] and by StyleGAN **(b)** [2].

StyleGAN

StyleGAN incorporates the concept of style transfer into progressive Generative Adversarial Networks (GANs) by using two interconnected networks.

The hierarchical synthesis network constructs an image from 512 constant feature maps of size 4×4 using convolutional layers, upsampling, the addition of noise and adaptive instance normalization (AdaIN) layers. The noise adds stochastic variance to the image, while the AdaIN operations control image features at different granularity levels using a so-called style vector to replace the mean and variance of feature maps. The style vector is generated from a latent input vector through a non-linear mapping network with learned parameters. The corresponding vector space is called style space.

To generate random images within the domain of the training data, a random latent vector is used as input to the mapping network. Given the style vector of an image, its features can be changed by modifying components of this vector giving rise to applications like style mixing. The style vector of any given image can be computed using an encoder allowing modifications of real images.



(a) (b)

Fig. 2: Structure of StyleGAN **(a)**, example for style mixing of high-level styles **(b)**. [2]

StylEx

Generates counterfactual explanations for image classifiers by finding influential dimensions within the style space and presenting images, in which those dimensions are varied, to the user.

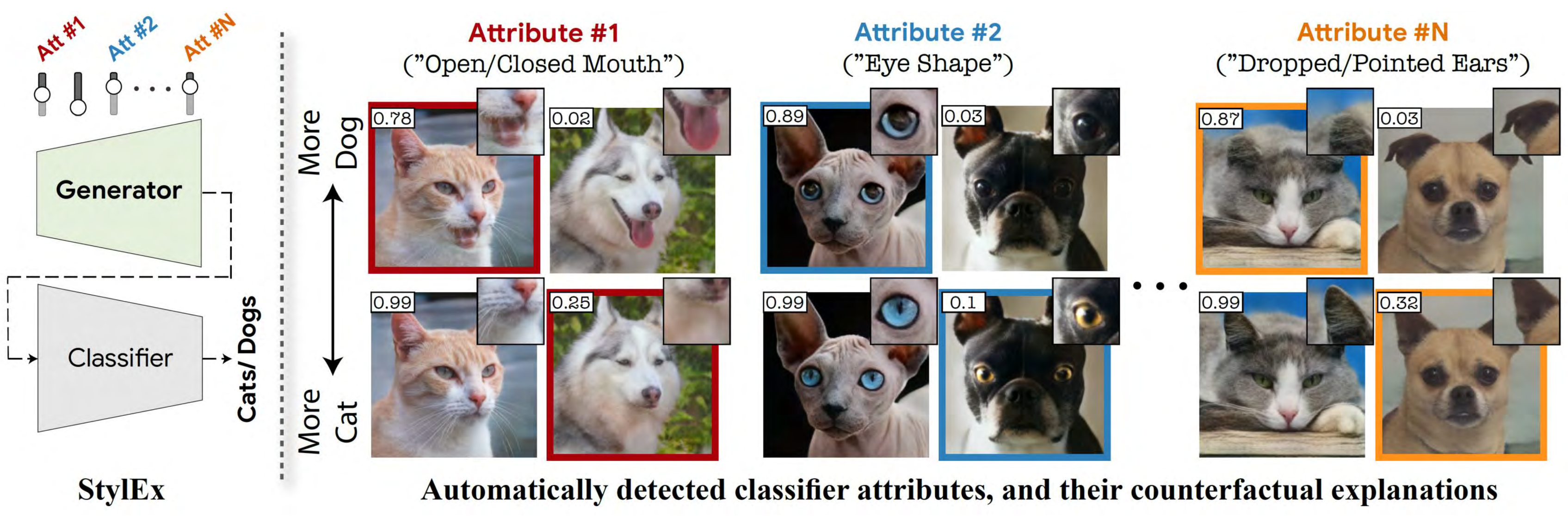


Fig. 3: Schematic structure of StylEx. [3]

Editing in Style

Locally and semantically editing a target image given a reference image by finding relevant style space dimensions using k-means clustering and interpolating the corresponding style vector components of the target to the ones of the reference.

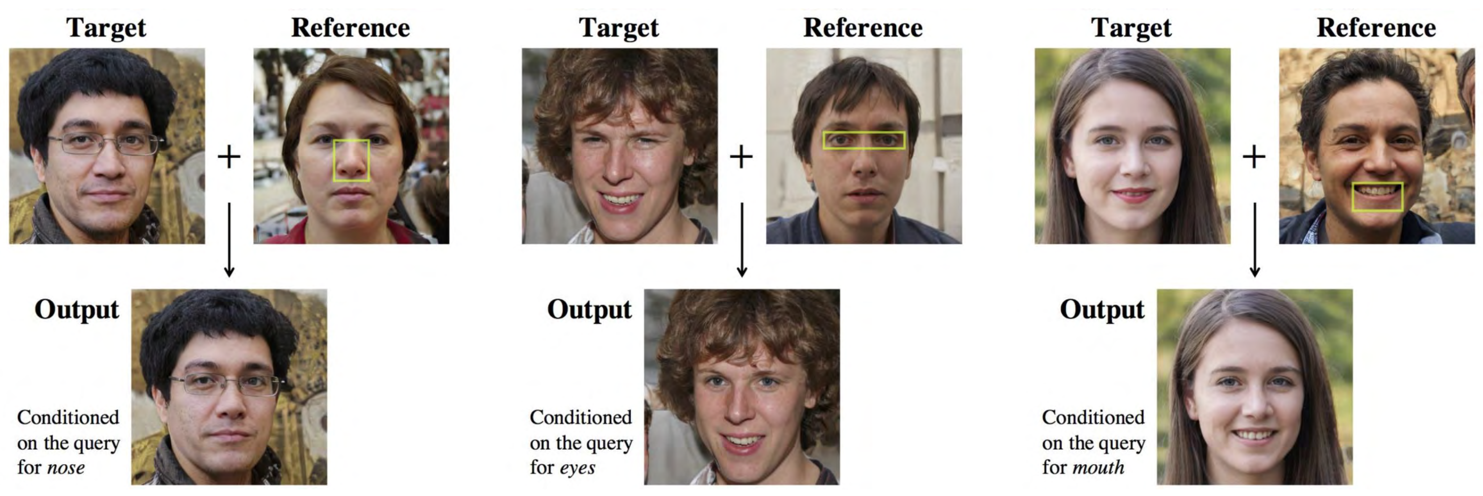


Fig. 4: Local semantic editing by style using a target and a reference image. [4]

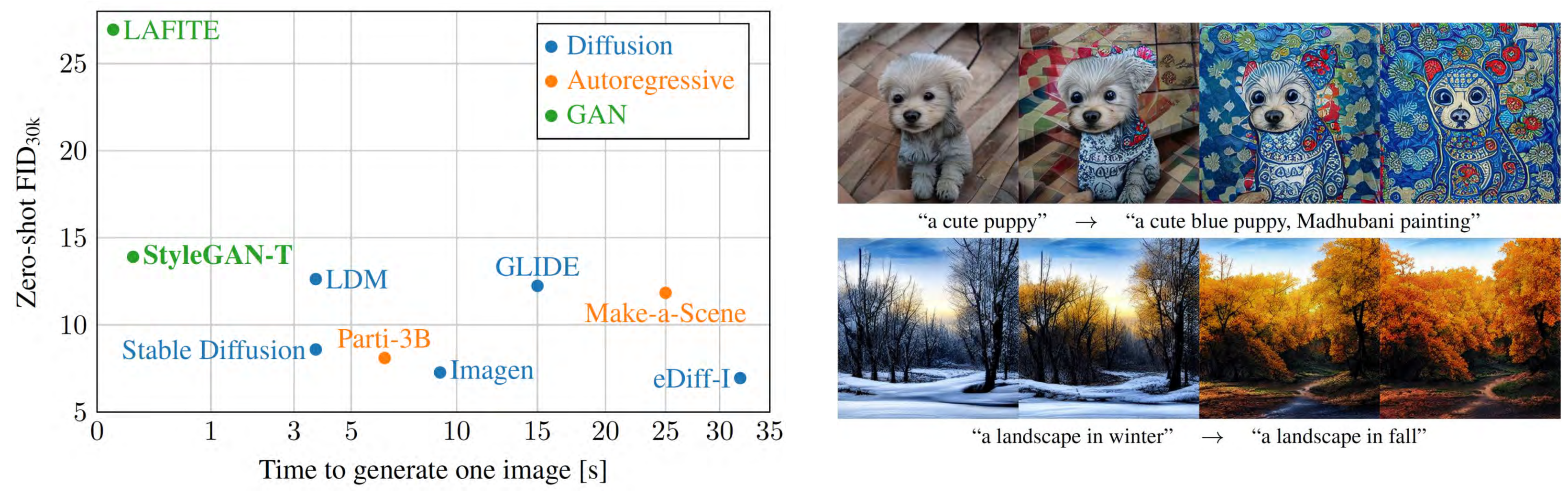
Pose with Style

Reconstructing a person from a single image into different poses by modifying the StyleGAN architecture such that the target pose is the synthetic network input, and the latent vector encodes the source image and a coordinate mapping.

Fig. 5: Comparison of human reposing methods. [5]

StyleGAN-T

Adopts the StyleGAN architecture for large-scale text-to-image synthesis. StyleGAN-T works faster than diffusion or autoregressive models and enables control over the generated images but with a trade-off in terms of reduced image quality.



(a) (b)

Fig. 6: Quality and speed comparison for text-to-image synthesis models **(a)**, modification of generated images with StyleGAN-T **(b)**. [6]

References

[1] Ian Goodfellow et al. "Generative Adversarial Nets". In: *Advances in Neural Information Processing Systems*. Vol. 27. Curran Associates, Inc., 2014.

[2] Tero Karras, Samuli Laine, and Timo Aila. "A Style-Based Generator Architecture for Generative Adversarial Networks". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence 43.12 (2021)*, pp. 4217–4228.

[3] Oran Lang et al. "Explaining in Style: Training a GAN to explain a classifier in StyleSpace". In: *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. 2021, pp. 673–682.

[4] Edo Collins et al. "Editing in Style: Uncovering the Local Semantics of GANs". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020.

[5] Badour AlBahar et al. "Pose with Style: Detail-Preserving Pose-Guided Image Synthesis with Conditional StyleGAN". In: *ACM Transactions on Graphics (2021)*.

[6] Axel Sauer et al. "StyleGAN-T: Unlocking the Power of GANs for Fast Large-Scale Text-to-Image Synthesis". In: *Proceedings of the 40th International Conference on Machine Learning*. Ed. by Andreas Krause et al. Vol. 202. Proceedings of Machine Learning Research. PMLR, 2023, pp. 30105–30118.