

A first look at Generative Adversarial Networks

Raphaël Deknop

July 29, 2020

Introduction

In the last decade, machine learning and artificial intelligence have become increasingly ubiquitous. From autonomous vehicles to face recognition or even [beating world-class players in online games](#), it is no wonder that this technology has an enormous potential and can be dangerously powerful. This document is aimed at the general public and intends to provide an intuitive -yet precise- understanding on what Generative Adversarial Networks are, what they are capable of, along with some further reflexions.

1. Discriminative and generative models

In machine learning, two main approaches can be followed. On the one hand, the **discriminative** method may be employed to tackle classifying problems such as assigning the correct label to an image, computing a highly probable output from a (previously unseen) complex input and much more. On the other hand, this model cannot *generate* similar data. An intuitive illustration of discriminative modelling in human behaviour is the capacity of distinguishing chinese characters while remaining unable to correctly draw one of them.

This is where the **generative** one differs. As its name suggests, a model using such an approach can generate data resembling what it has been fed. Here, as long as the input data is carefully selected, no label is required. Let's say that you want your model to generate a Shakespearean poem, it would be counterproductive to train it with some -unlabelled- Edgar Allan Poe or Oscar Wilde writings.

2. Generative Adversarial Networks

Although those two approaches were just presented separately, nothing is forbidding a network to follow both of them. This is where Generative Adversarial Networks come in. In a GAN, a generative network produces candidates while a discriminative one evaluates them. To put it simply, the goal of the generator is to fool the discriminator into thinking that the data provided is real.

This type of network, presented in 2014 by Ian Goodfellow and his collaborators [[Goo+14](#)], has been an enormous step in machine learning. To illustrate that statement, look at this series of generated human faces:



Figure 1: Progress of GANs on human face generation. Source: [Ian Goodfellow's Twitter](#)

The evolution here is astonishing. Let's remember that before 2014, GANs simply didn't exist. There are a bunch of websites showing GAN-generated images of [people](#), [artworks](#), [cats](#) or even [horses](#) produced by StyleGAN2 [Kar+19] for the curious out there. It is effortless -and an interesting exercise- to come up with potential applications and realise that there are plenty of them.

3. Going beyond GANs

There exists plenty of neural network types, generative adversarial networks are just one of them. The reason I presented them here is because I find them to be lying on the right spot between intuitive understanding and current state of the art in AI research. I haven't been technical here, as it is not the intention of this article.

Interestingly enough, GANs have their limitation. For example, it is impossible to tweak parameters on a generated human face picture in order to change, say, hair colour, face expression or nose type. However, that is made possible by Adversarial Latent Autoencoders (ALAEs) [PAD20] which make use of [latent spaces](#).

[Figure 2](#) contains screenshots from a video made by *Two Minute Papers* on ALAEs. Its creator, Karoly Zsolnai-Fehér, demonstrates the use of cursors on the right in order to tweak desired parameters. In the case of [figures 2a](#) and [2b](#) the cursor *mouth-open* is set from low to high. Notice how smile lines appear not only near the mouth but also around the eyes.

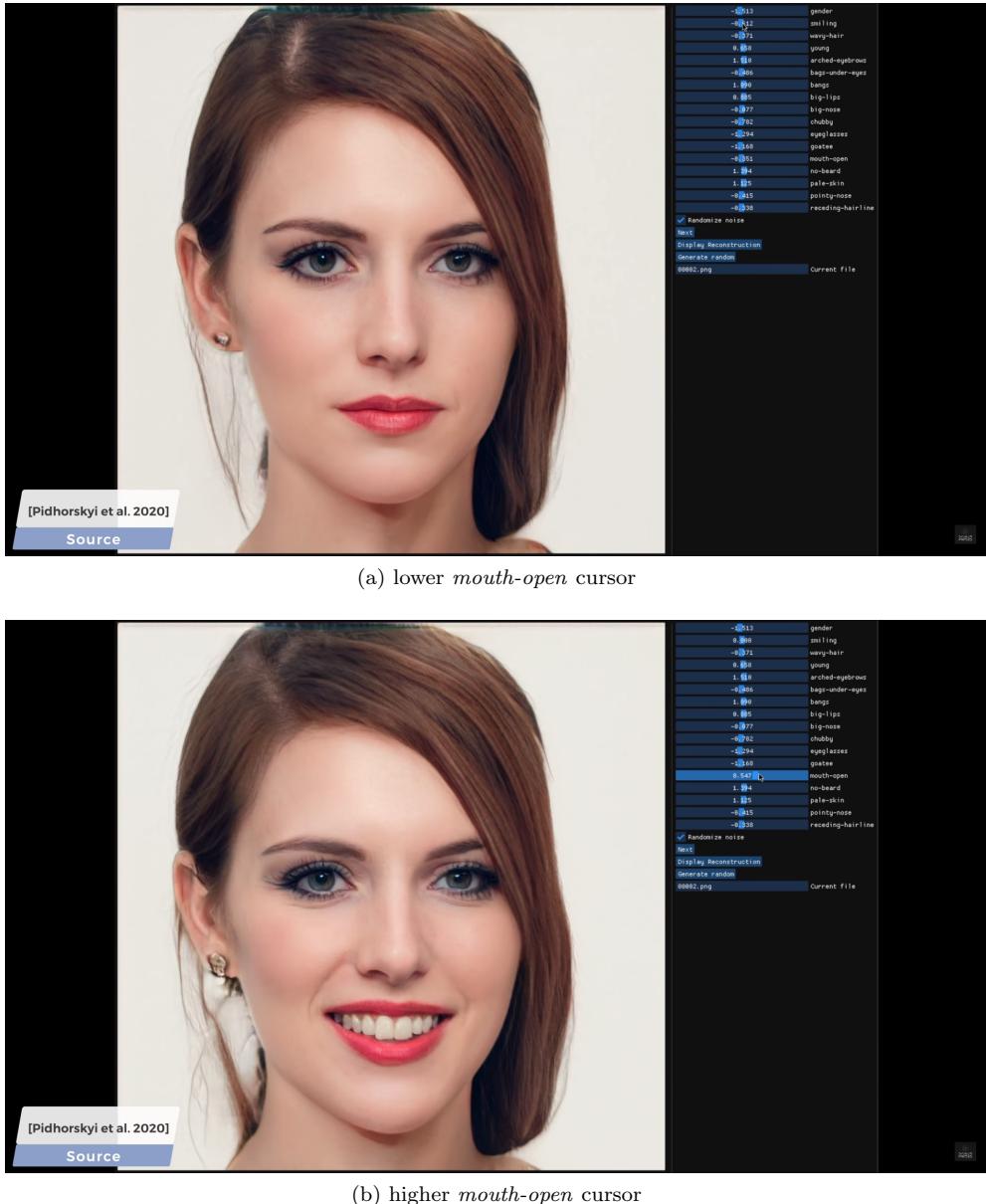


Figure 2: Tweakable generated human faces produced by an ALAE. Screenshots from [TMP's video](#) on the subject.

4. Discussion

Considering the available information and considering the actual progress made in AI and ML research, what could be achievable in the next decade ? In order to answer such a question, several ways of reasoning can be employed. In this case, I will adopt the following:

1. Acknowledge the current advancements in the topic you are studying
2. Think about how (or *if*) they could be improved and polished individually
3. Try to come up with new ways in which they could be combined to produce innovative results

This last point can be eased by merging different topics together. Let's try with machine learning and cinema.

Could creating a movie become a one-person job that would't require actors or any equipment other than a computer ? We are not talking about animation here but rather in the order of giving a script as an input and obtaining a complete realistic movie as an output. We have seen that it is an unchallenging task to generate pictures of human faces and, as illustrated on [figure 2](#), to make it smile. Could we create a latent space of facial expressions ? It seems like we can [[ZS17](#)]. Some more research and training and we are good to go. What about body movement ? Well, considering current papers, it looks promising [[Sia+20](#); [Par+19a](#)]. This could go on with interior design [[Mao+16](#)], landscapes [[Par+19b](#)] (Nvidia set up a site where you can [try it out yourself](#)), text to image synthesis [[Ree+16](#)], natural language processing [[Bro+20](#)] and way more.

While each of all those cited works have their own flaws, I'd recommend looking at [figure 1](#) again. That should give an indication on the room for improvement that exists for those applications. Combining the evoked technologies and letting them mature enough would create the ideal conditions in which our actorless movie could come to reality. The impacts of such an achievement could be huge. For example, if this way of creating films becomes the standard, it's a whole industry that would collapse. This is of course still close to science-fiction but weren't driverless cars science-fiction 10 or 20 years ago ?

5. Conclusion

AI research is making progress at a substantial pace and is shifting our society, the examples we have seen are only the tip of an iceberg whose size we ignore. The possibilities seem endless and as long as sufficient data and computing power are provided, the only limiting factor appears to be our imagination (and values). As with every powerful tool, we need to be vigilant with it, but the future looks fascinating.

References

- [Bro+20] Tom B. Brown et al. *Language Models are Few-Shot Learners*. 2020. arXiv: [2005.14165 \[cs.CL\]](#).
- [Goo+14] Ian J. Goodfellow et al. *Generative Adversarial Networks*. 2014. arXiv: [1406.2661 \[stat.ML\]](#).
- [Kar+19] Tero Karras et al. *Analyzing and Improving the Image Quality of StyleGAN*. 2019. arXiv: [1912.04958 \[cs.CV\]](#).
- [Mao+16] Xudong Mao et al. *Least Squares Generative Adversarial Networks*. 2016. arXiv: [1611.04076 \[cs.CV\]](#).
- [PAD20] Stanislav Pidhorskyi, Donald Adjeroh, and Gianfranco Doretto. *Adversarial Latent Autoencoders*. [preprint]. 2020. arXiv: [2004.04467 \[cs.LG\]](#).
- [Par+19a] Soohwan Park et al. *Learning Predict-and-Simulate Policies From Unorganized Human Motion Data*. 2019. URL: <http://mrl.snu.ac.kr/publications/ProjectICC/ICC.pdf>.
- [Par+19b] Taesung Park et al. *Semantic Image Synthesis with Spatially-Adaptive Normalization*. 2019. arXiv: [1903.07291 \[cs.CV\]](#).
- [Ree+16] Scott Reed et al. *Generative Adversarial Text to Image Synthesis*. 2016. arXiv: [1605.05396 \[cs.NE\]](#).
- [Sia+20] Aliaksandr Siarohin et al. *First Order Motion Model for Image Animation*. 2020. arXiv: [2003.00196 \[cs.CV\]](#).
- [ZS17] Yuqian Zhou and Bertram Emil Shi. *Photorealistic Facial Expression Synthesis by the Conditional Difference Adversarial Autoencoder*. 2017. arXiv: [1708.09126 \[cs.CV\]](#).