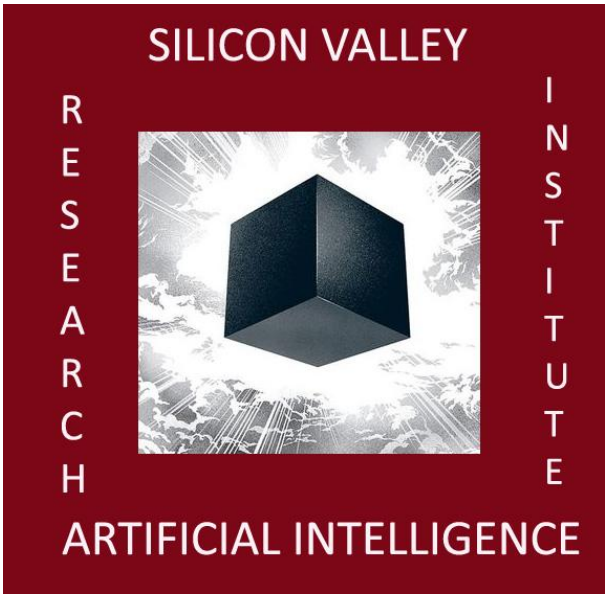

Generative Adversarial Networks: a brief introduction

" Machines may have become good (or, at least, better) at classifying objects in categories (i.e. in recognizing what an object is), but they still lag far behind in drawing an example of a category. There is a difference between recognizing a dog and drawing a dog. If you understood what a dog is, it should be easy for you to sketch what a dog looks like. You have a generative mind: you can classify an object in its category and you can draw a typical object of that category, an object that presumably is not like any specific object that you have seen (unless you are Giotto). In order to implement this "generative" behavior a new approach to machine learning was required.

A recurrent network was used to generate sequences by Hinton's student Ilya Sutskever ("Generating Text with Recurrent Neural Networks", 2011), but recurrent neural networks were clearly limited in their ability to look ahead. In 2014 Hinton's student Alex Graves at the University of Toronto used a LSTM network, more efficient at storing and retrieving information than plain recurrent networs, to generate handwriting. You can enter a text at his webpage "http://www.cs.toronto.edu/~graves/handwriting.cgi" (as of 2017) the system will write it out in human-like handwriting ("Generating Sequences With Recurrent Neural Networks", 2014). This was an important first step.

"Turing Learning" was developed by Roderich Gross at the University of Sheffield: it pits two algorithms against each other, one trying to classify the other while the other is trying to fool the former ("A Coevolutionary Approach to Learn Animal Behavior Through Controlled Interaction", 2013). In a similar fashion in 2014 Ian Goodfellow, one of Bengio's students at the University of Montreal, invented "generative adversarial networks" (GANs), consisting of two neural networks that compete against each other, one trying to fool the other ("Adversarial Examples and Adversarial Training", 2014). Another member of the same lab, Mehdi Mirza,

improved the idea with "conditional adversarial nets" ("Conditional Generative Adversarial Nets", 2014). In 2015 Alec Radford at Indico Data Solutions in Boston proved that an expanded version of GAN's can generate perfectly valid images, except that they are not real ("Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks", 2015).

GANs contain two independent neural networks that behave as adversaries: one (the "discriminator") tries to correctly classify the real images while the other one (the "generator") produces fake images to fool the former (the "discriminator"); the generator needs to improve its ability to "fake" images while the discriminator needs to improve its ability to discriminate fake ones and real ones. The images produced by the generator are only partially random because they have to resemble the real ones. The generator is trying to fool the discriminator while the discriminator is trying to not get fooled by the generator. As they evolve the respective skills, both tend towards the point where the counterfeits and the originals are indistinguishable. The process trains the discriminator to classify more and more accurately. It also, incidentally, trains the generator to produce highly realistic pictures of imaginary objects, which may represent an art in itself.

The wonders of GANs quickly lured legions of researchers. In 2016 a joint team of the University of Michigan (Honglak Lee) and the Max Planck Institute in Germany (Bernt Schiele and Zeynep Akata) employed GANs to generate images from text descriptions ("Generative Adversarial Text to Image Synthesis", 2016). Antonio Torralba's student Carl Vondrick at MIT employed a GAN to predict the plausible evolution of a scene, i.e. to generate a video. This implies understanding what is going on in the scene and inferring what is reasonable to see happen next ("Generating Videos with Scene Dynamics", 2016). Alexei Efros' students at UC Berkeley (including Jun-Yan Zhu and Phillip Isola) created a neural network that can turn the picture of a horse into the picture of a zebra using GANs ("Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks", 2017). Then they used "conditional adversarial networks" to develop the "Pix2pix model" capable of generating images from sketches or abstract diagrams ("Image-to-Image Translation with Conditional Adversarial Networks", 2017). When they released the related Pix2pix software, it started a wave of experiments (many of them by professional artists) in creating images: sketch your desired handbag and the system displays what appears to be a real handbag and even colors it. The conditional adversarial network (cGAN) learns how to map an input image onto an output image, or, if you prefer, how to redraw an image with different attributes. Later in 2017 Ming-Yu Liu's team at Nvidia used a slightly different architecture for their image-to-image translation system UNIT (or UNsupervised Image-to-image Translation), i.e. variational autoencoders coupled with generative adversarial networks ("Unsupervised Image-to-Image Translation Networks", 2017). A few months later Jaakko Lehtinen's team at Aalto University in Finland published a paper showing how GANs can create photorealistic pictures of fake celebrities.

Within a few years, Alec Radford’s original DCGAN (Deep Convolutional Generative Adversarial Network) begat legions of variants: Yann LeCun's Energy-based GAN or EGBAN (2016) at New York University; Léon Bottou's Wasserstein GAN or WGAN at Facebook (2017), Aaron Courville's WGAN-GP at Montreal Institute for Learning Algorithms (2017) that generated photorealistic images of bedrooms; Trevor Darrell's bidirectional GANs or BiGANs at

UC Berkeley (2017); Luke Metz's BEGAN (Boundary Equilibrium GAN) at Google that generated photorealistic faces (2017); Alexei Efros' above-mentioned Cycle-Consistent Adversarial Networks or CycleGANs (2017) at UC Berkeley; Fernando De la Torre's HDCGAN (2018) at Carnegie Mellon University for high resolution pictures; Aaron Courville's "adversarially learned inference model" (2017) that learns both a generation network and an inference network; etc.

GANs are probably more interesting as a model of human intelligence than the inventors realized. Competition is one of the key factors in evolution, and, in particular, in the evolution of the brain. Competition often ends up being collaboration: when two adversaries compete, they indirectly help each other improve. They induce a positive feedback loop on their skills. The fundamental case of competition is perhaps the relationship between the two sexes. Charles Darwin in "The Descent of Man and Selection in Relation to Sex" (1871) and Ronald Fisher in "The Genetical Theory of Natural Selection" (1930) already pointed out that sexual selection could greatly accelerate evolution: the female chooses the male and therefore males are pressured to try and be chosen, and as more and more males qualify the female has to become choosier, pressuring the males to further improve, and so on in an endless positive feedback loop. Geoffrey Miller in "The Mating Mind" (2000) went beyond the tail of the peacock and the song of the thrushes. He speculated that language itself, and therefore mind, is created via a feedback loop of this kind: Miller views the human mind not as a problem solver, but as a "sexual ornament". The human brain's creative intelligence must exist for a purpose, and that purpose is not obvious: survival in the environment does not quite require the sophistication of Einstein's science or Michelangelo's paintings or Beethoven's symphonies. On the other hand, these are precisely the kind of things that the human brain does a lot better than other animal brains. The human brain is much more powerful than it needs to be. Miller explains the emergence of art, science and philosophy by thinking not in terms of survival benefits but in terms of reproductive benefits. Sexual selection shapes not only the animal world but also our own mind and our civilizations.

Alas, GANs are very difficult to train. As an alternative to GANs, Thomas Brox's student Alexey Dosovitskiy showed that one can train a convnet to generate images ("Learning to Generate Chairs, Tables and Cars with Convolutional Networks", 2016); and no adversarial training was used by Qifeng Chen at Stanford University to synthesize photorealistic images ("Photographic Image Synthesis with Cascaded Refinement Networks", 2017)