



Enoch Kan

Aug 17, 2018 · 5 min read · ✨ Member-only



Search



What The Heck Are VAE-GANs?

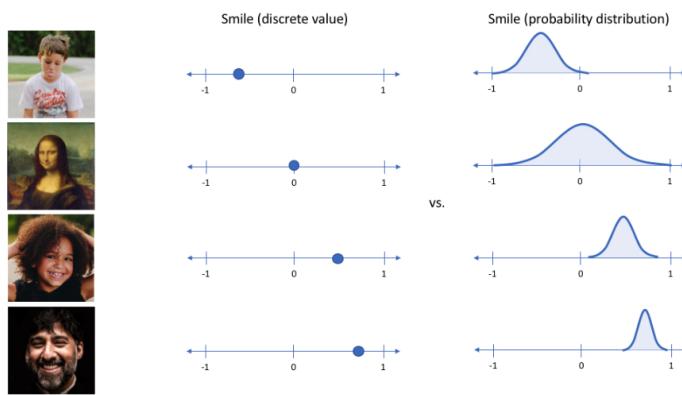
Yep, you read the title correctly. While a few friends of mine are vegans, none of them knew anything about VAE-GANs. VAE-GAN stands for Variational Autoencoder- Generative Adversarial Network (that is one heck of a name.) Before we get started, I must confess that I am no expert in this subject matter (I don't have PhD in electrical engineering, just sayin'). But after reading several research papers and watching [Ian Goodfellow's 30-minute long intro to GANs](#), here is a short (yet concise) summary of my major takeaways:



Image reconstructed by VAE and VAE-GAN compared to their original input images

Variational Autoencoders (VAEs)

The simplest way of explaining variational autoencoders is through a diagram. Alternatively, you can read [Irhum Shafkat's excellent article on Intuitively Understanding Variational Autoencoders](#). At this point I assume you have a general idea of what unsupervised learning and generative models are. The textbook definition of a VAE is that it “provides *probabilistic* descriptions of observations in latent spaces.” In plain English, this means VAEs store latent attributes as probability distributions.



"Variational autoencoders" - Jeremy Jordan

Each input image has features that can normally be described as single, discrete values. Variational autoencoders describe these values as probability distributions. Decoders can then sample randomly from the probability distributions for input vectors. Let me guess, you're probably wondering what a decoder is, right? Let's take a step back and look at the general architecture of VAE.



Enoch Kan

550 Followers

ML Lead @ [measurable.energy](#). Currently based in London, United Kingdom GB

Follow

More from Medium

Ning-Yu Kao

Don't Get Me Wrong, Any... [12ft.io – Is the paywall gone? Yes / No](#)

Isaiah McCall in Yard Couch

Now That YouTube is Dying, Here Are Better Alternatives

KARRI TIRUMALA VINAY

Data Science is Dying?

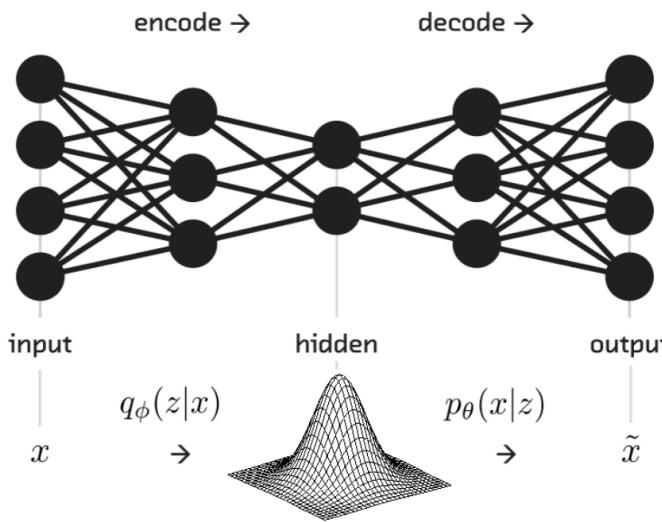
Amy Parker

Top 13 Artificial Intelligence Applications in 2023

Help Status Writers Blog Careers Privacy Terms About Knowable

12ft.io – Is the paywall gone? Yes / No

12ft.io – Is the paywall gone? Yes / No



12ft.io – Is the paywall gone? Yes / No

Typical setup of a variational autoencoder is nothing other than a cleverly designed deep neural network, which consists of a pair of networks: the encoder and the decoder. The **encoder** can be better described as a variational inference network, which is responsible for the mapping of input x to posteriors distributions $q_\theta(z|x)$. Likelihood $p(x|z)$ is then parametrized by the **decoder**, a generative network which takes latent variables z and parameters as inputs and projects them to data distributions $p_\phi(x|z)$.

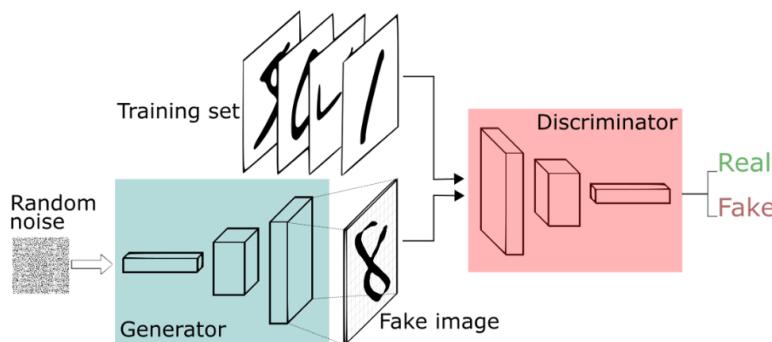
A major drawback of VAEs is the blurry outputs that they generate. As suggested by Dosovitskiy & Brox, VAE models tend to produce unrealistic, blurry samples. This has to do with how data distributions are recovered and loss functions are calculated in VAEs in which we will discuss further below. A 2017 paper by Zhao et. al. has suggested modifications to VAEs to not use variational Bayes method to improve output quality.

12ft.io – Is the paywall gone? Yes / No

Generative Adversarial Networks (GANs)

The dictionary definition of *adversarial* is *involving or characterized by conflict or opposition*. This is in my opinion a very accurate description of what GANs are. Just like VAEs, GANs belong to a class of generative algorithms that are used in unsupervised machine learning. Typical GANs consist of two neural networks, a **generative** neural network and a **discriminative** neural network. A generative neural network is responsible for taking noise as input and generating samples. The discriminative neural network is then asked to evaluate and distinguish the generated samples from training data. Much like VAEs, generative networks map latent variables and parameters to data distributions.

12ft.io – Is the paywall gone? Yes / No



12ft.io – Is the paywall gone? Yes / No

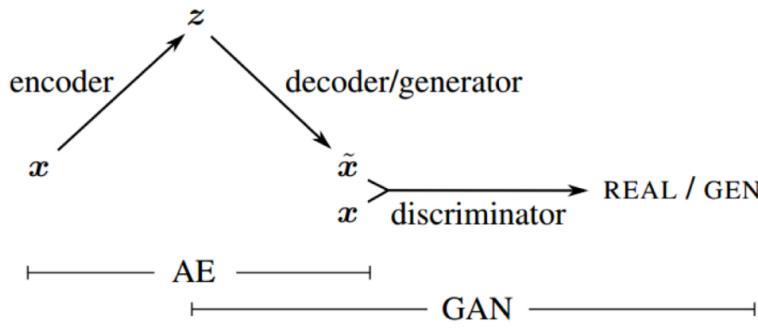
The major goal of generators is to generate data that increasingly “fools” the discriminative neural network, i.e. increasing its error rate. This can be done by repeatedly generating samples that appear to be from the training data distribution. A simple way to visualize this is the “competition” between a cop and a cyber criminal. The cyber criminal (generator) attempts to create online identities that resemble ordinary citizens, while the cop (discriminator) tries to distinguish fake profiles from the real ones.

12ft.io – Is the paywall gone? Yes / No

Variational Autoencoder Generative Adversarial Networks (VAE-GANs)

Okay. Now that we have introduced VAEs and GANs, it's time to discuss what VAE-GANs really are. The term VAE-GAN is first introduced in the paper *“Autoencoding beyond pixels using a learned similarity metric”* by A. Larsen et. al. The authors suggested the combination of variational autoencoders and generative adversarial networks outperforms traditional VAEs.

12ft.io – Is the paywall gone? Yes / No



VAE-GAN architecture, the discriminator from GAN takes input from VAE's decoder

12ft.io – Is the paywall gone? Yes / No

Remember GANs are subdivided into generators and discriminator networks? The authors suggested a GAN discriminator can be used in place of a VAE's decoder to learn the loss function. The motivation behind this modification is as mentioned above, VAEs tend to produce blurry outputs during the reconstruction phase. This “blurriness” is somehow related to the way VAE's loss function is calculated. I am not going into the nitty gritty of how this new loss function is calculated, but all you need to know is this set of equation

$$-\mathcal{L}_{vae} = MSE(D^l(x_{decoder}), D^l(x_{real})) + prior$$

12ft.io – Is the paywall gone? Yes / No

$$-\mathcal{L}_{GAN} = E_{x \sim p_{data}} \log(D(x)) + E_{x \sim p_{model}} \log(1 - D(x))$$

[Learned Losses in VAE-GAN](#)

Now that's a lot of L.S. But jokes aside, the above equations assume the *un* layer of the discriminator have outputs that differ in a Gaussian manner. As a result, calculating the mean squared error (MSE) between the *l*th layer outputs gives us the VAE's loss function. The final output of GAN, $D(x)$, can then be used to calculate its own loss function.



12ft.io – Is the paywall gone? Yes / No

Generative models are now added to the list of AI research by top tech companies such as Facebook. Yann LeCun, a prominent computer scientist and AI visionary once said “*This (Generative Adversarial Networks), and the variations that are now being proposed is the most interesting idea in the last 10 years in ML, in my opinion.*”

Besides VAE-GANs, many other variations of GANs have been researched and implemented. DCGANs, or Deep Convolutional Generative Adversarial Networks, are introduced not long after Ian Goodfellow’s introduction to the original GANs. I am excited to see for generative model to find its role in future AI applications, and potentially improving the qualities of our lives.

12ft.io – Is the paywall gone? Yes / No

Thanks for reading my article.

🕒 -- ⏱ 2

↑ ⌂

More from Towards Data Science

Follow

Your home for data science. A Medium publication sharing concepts, ideas and codes.

12ft.io – Is the paywall gone? Yes / No

Read more from Towards Data Science

Recommended from Medium

Abed Islam

Sentiment Analysis—Sorry Not Sorry

Temperature of Egypt

AMERICAN EGYPTIAN MUSEUM

Sunil kumar Jangir

The outline of Gradient Descent

MLearning.ai

Chier Hu in Mr. Translator

What are the applications of machine translation? Who can take interest for this service?

Mr. Translator



Gavin Hull in MLearning.ai

Building a Neural Network Zoo From Scratch: The Perceptron

MLearning.ai



Philipp Eis... in Towards Data Sci...

How to build a real Besserwissen with BERT text classification for Slack...

Towards Data Sci...



Tristan Rouillard

Levels of annotation automation

Towards Data Sci...



Dylan Alme... in Towards Data Sci...

Combining Bayesian Neural Networks and Ensemble techniques

Towards Data Sci...



Santiago Velez Garcia

ML Regularization Techniques

Towards Data Sci...



[Open in app](#)[Get started](#)

Published in Towards Data Science

You have **2** free member-only stories left this month. [Sign up for Medium and get an extra one](#)



Enoch Kan

[Follow](#)

Aug 16, 2018 · 5 min read · ✨ · ⏰ Listen

[Save](#)

What The Heck Are VAE-GANs?

Yep, you read the title correctly. While a few friends of mine are vegans, none of them knew anything about VAE-GANs. VAE-GAN stands for Variational Autoencoder-Generative Adversarial Network (that is one heck of a name.) Before we get started, I must confess that I am no expert in this subject matter (I don't have PhD in electrical engineering, just sayin'). But after reading several research papers and watching [Ian Goodfellow's 30-minute long intro to GANs](#), here is a short (yet concise) summary of my major takeaways:



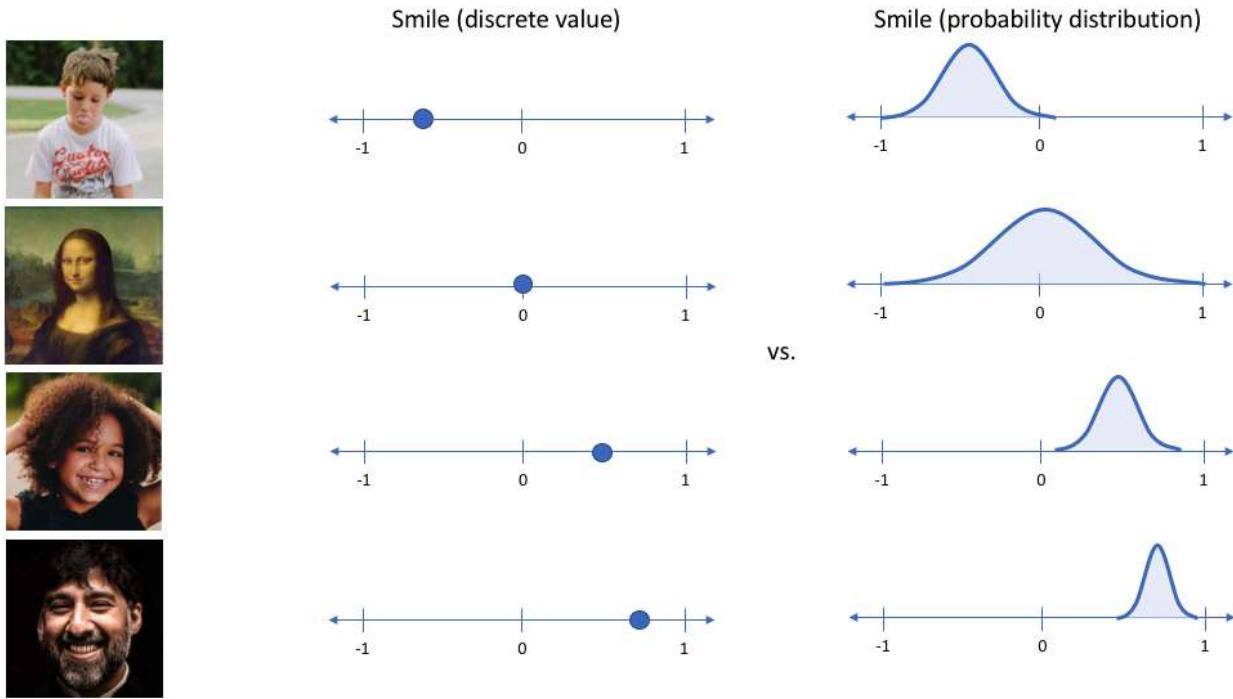
Image reconstructed by VAE and VAE-GAN compared to their original input images

Variational Autoencoders (VAEs)



[Open in app](#)[Get started](#)

of a VAE is that it “provides *probabilistic* descriptions of observations in latent spaces.” In plain English, this means VAEs store latent attributes as probability distributions.



[“Variational autoencoders”](#)- Jeremy Jordan

Each input image has features that can normally be described as single, discrete values. Variational autoencoders describe these values as probability distributions. Decoders can then sample randomly from the probability distributions for input vectors. Let me guess, you’re probably wondering what a decoder is, right? Let’s take a step back and look at the general architecture of VAE.

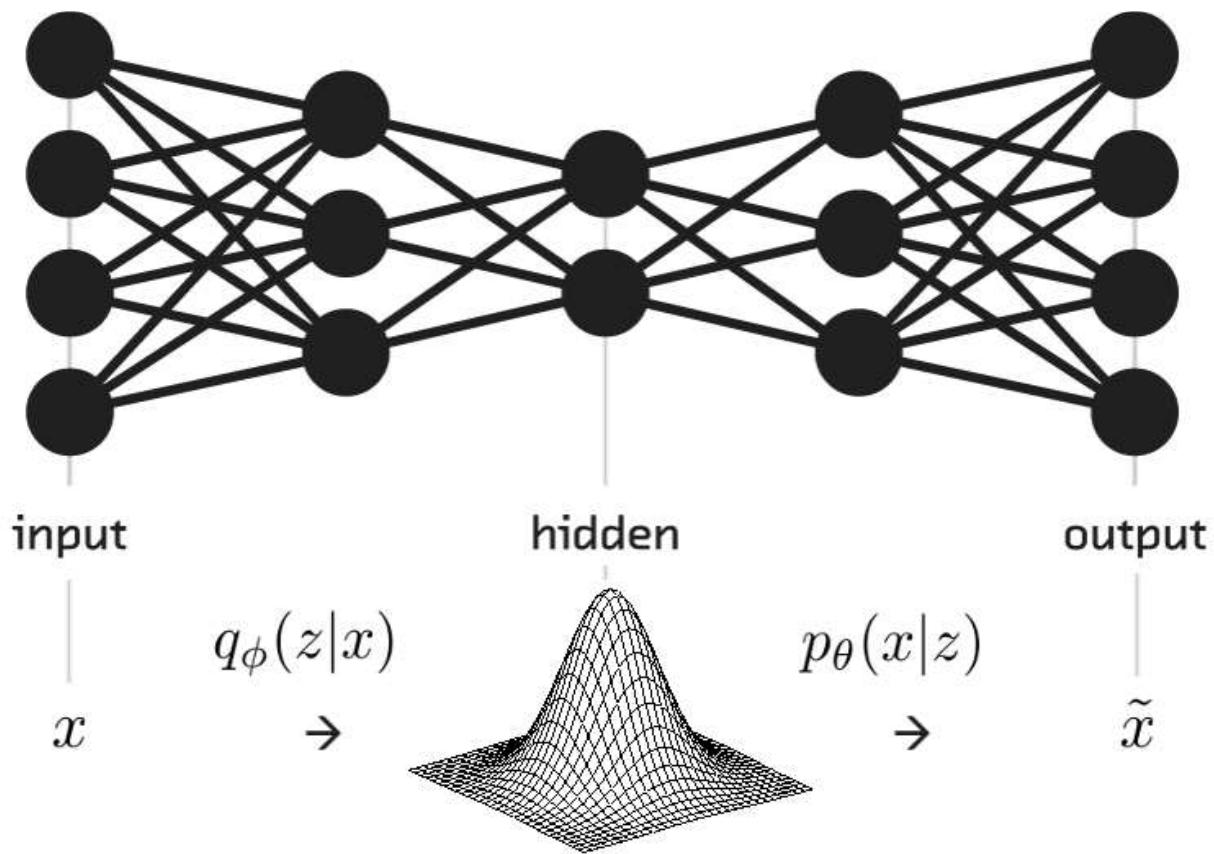
652 | 2





Open in app

Get started



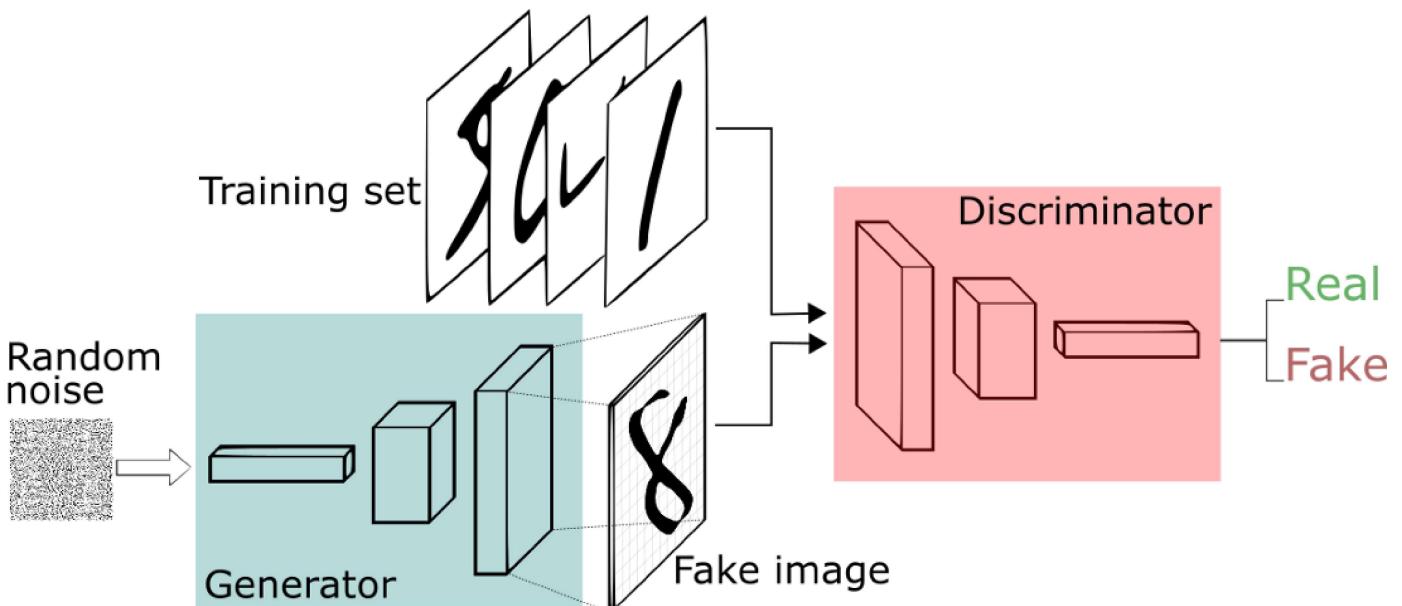
Typical setup of a variational autoencoder is nothing other than a cleverly designed deep neural network, which consists of a pair of networks: the encoder and the decoder. The **encoder** can be better described as a variational inference network, which is responsible for the mapping of input x to posteriors distributions $q_\theta(z|x)$. Likelihood $p(x|z)$ is then parametrized by the **decoder**, a generative network which takes latent variables z and parameters as inputs and projects them to data distributions $p_\phi(x|z)$.

A major drawback of VAEs is the blurry outputs that they generate. As suggested by Dosovitskiy & Brox, VAE models tend to produce unrealistic, blurry samples. This has to do with how data distributions are recovered and loss functions are calculated in VAEs in which we will discuss further below. A 2017 [paper](#) by Zhao et. al. has suggested modifications to VAEs to not use variational Bayes method to improve output quality.



[Open in app](#)[Get started](#)

VAEs, GANs belong to a class of generative algorithms that are used in unsupervised machine learning. Typical GANs consist of two neural networks, a **generative** neural network and a **discriminative** neural network. A generative neural network is responsible for taking noise as input and generating samples. The discriminative neural network is then asked to evaluate and distinguish the generated samples from training data. Much like VAEs, generative networks map latent variables and parameters to data distributions.



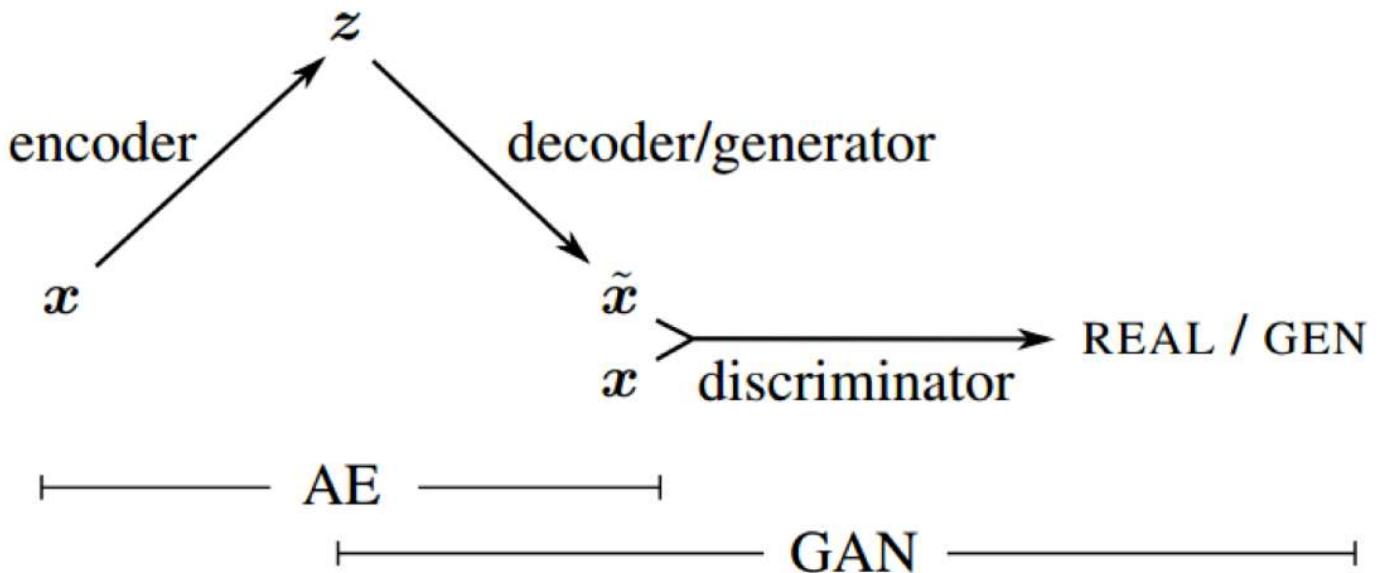
The major goal of generators is to generate data that increasingly “fools” the discriminative neural network, i.e. increasing its error rate. This can be done by repeatedly generating samples that appear to be from the training data distribution. A simple way to visualize this is the “competition” between a cop and a cyber criminal. The cyber criminal (generator) attempts to create online identities that resemble ordinary citizens, while the cop (discriminator) tries to distinguish fake profiles from the real ones.



[Open in app](#)[Get started](#)

Variational Autoencoder Generative Adversarial Networks (VAE-GANs)

Okay. Now that we have introduced VAEs and GANs, it's time to discuss what VAE-GANs really are. The term VAE-GAN is first introduced in the paper "[Autoencoding beyond pixels using a learned similarity metric](#)" by A. Larsen et. al. The authors suggested the combination of variational autoencoders and generative adversarial networks outperforms traditional VAEs.



VAE-GAN architecture, the discriminator from GAN takes input from VAE's decoder

Remember GANs are subdivided into generators and discriminator networks? The authors suggested a GAN discriminator can be used in place of a VAE's decoder to learn the loss function. The motivation behind this modification is as mentioned above.



[Open in app](#)[Get started](#)

to know is this set of equation

$$-\mathcal{L}_{vae} = MSE(D^l(x_{decoder}), D^l(x_{real})) + prior$$

$$-\mathcal{L}_{GAN} = E_{x \sim p_{data}} \log(D(x)) + E_{x \sim p_{model}} \log(1 - D(x))$$

Learned Losses in VAE-GAN

Now that's a lot of L 's. But jokes aside, the above equations assume the l th layer of the discriminator have outputs that differ in a Gaussian manner. As a result, calculating the mean squared error (MSE) between the l th layer outputs gives us the VAE's loss function. The final output of GAN, $D(x)$, can then be used to calculate its own loss function.



Generative models are now added to the list of AI research by top tech companies such as Facebook. Yann Lecun, a prominent computer scientist and AI visionary once said *“This (Generative Adversarial Networks), and the variations that are now being proposed is the most interesting idea in the last 10 years in ML, in my opinion.”*

Besides VAE-GANs, many other variations of GANs have been researched and implemented. DCGANs, or Deep Convolutional Generative Adversarial Networks, are introduced not long after Ian Goodfellow's introduction to the original GANs. I am excited to see for generative model to find its role in future AI applications, and potentially improving the qualities of our lives.



[Open in app](#)[Get started](#)

Sign up for The Variable

By Towards Data Science

Every Thursday, the Variable delivers the very best of Towards Data Science: from hands-on tutorials and cutting-edge research to original features you don't want to miss. [Take a look.](#)

By signing up, you will create a Medium account if you don't already have one. Review our [Privacy Policy](#) for more information about our privacy practices.

+ Get this newsletter

[About](#) [Help](#) [Terms](#) [Privacy](#)

Get the Medium app

