[4]Deep Generative Modeling

Which of Those are Fake?

- All of the faces above are computer generated face

Supervised vs unsupervised Learning

- Supervised learning
  - Data(x, y), x is data and y is label
  - Goal is to learn function to map
  - Examples are classification, regression, object detection, semantic segmentation
- Unsupervised learning
  - x is data, there's no labels
  - Goal is to learn underlying structure of data
  - Examples are clustering, feature or dimensionally reduction, etc...

Generative Modeling

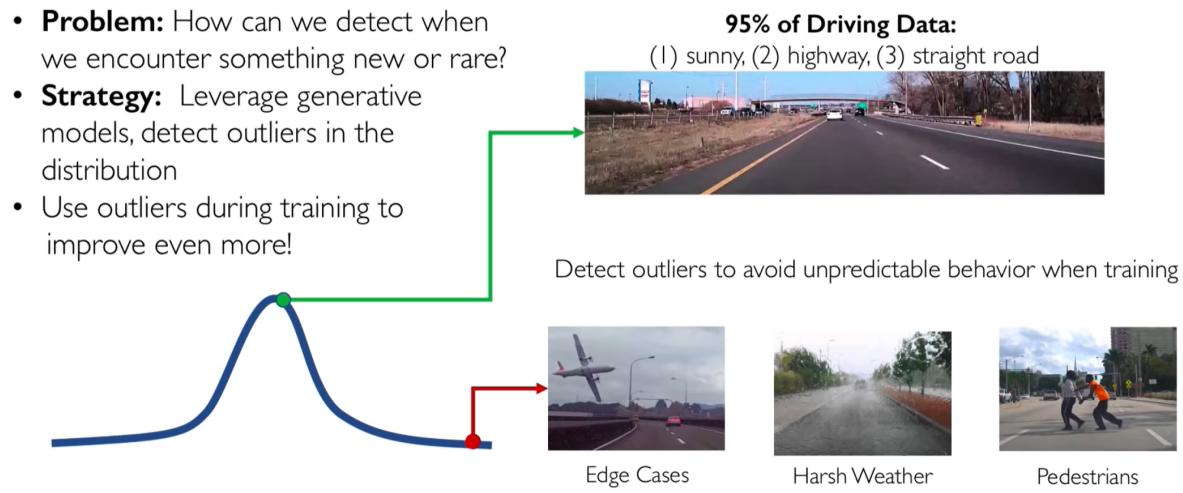
- Goal is to Take input training samples from some distribution and learn a model that represents that distribution

- The core question behind generative modeling is how can we learn $P_{model}(x)$ similar to $P_{data}(x)$

Why we need Generative Modeling

- Capable of undercovering underlying features in dataset

- Outlier detection

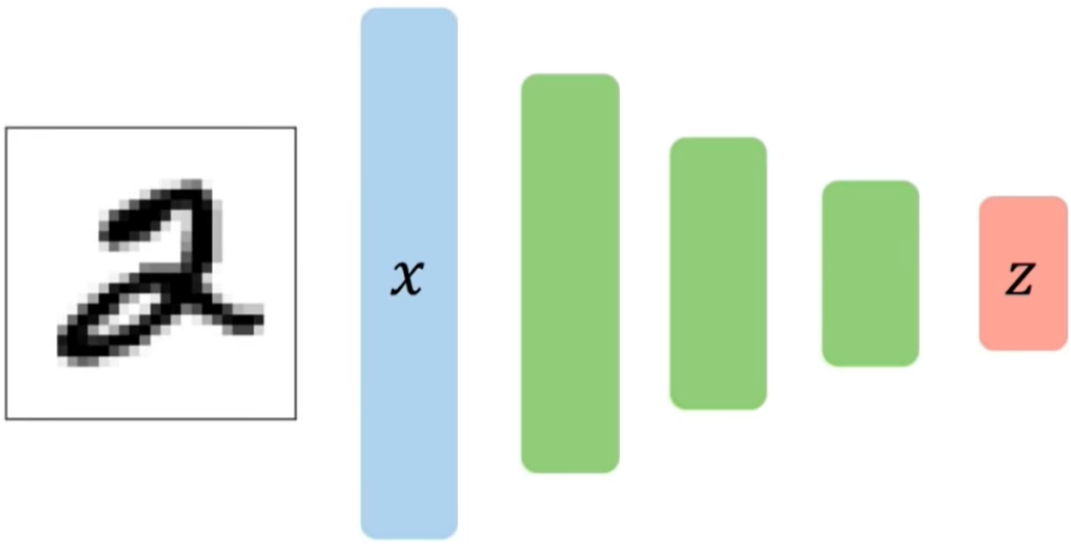


Latent Variables

- The goal is to laern underlying latent varibales even though only given observe data

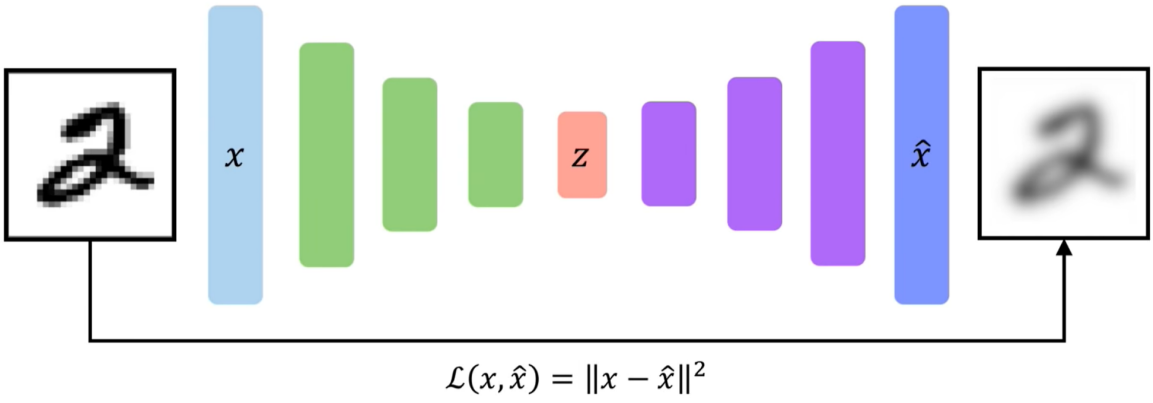
Autoencoder

- Unsubervised approach for learning a low-dimentional feature representation from unlabeled data

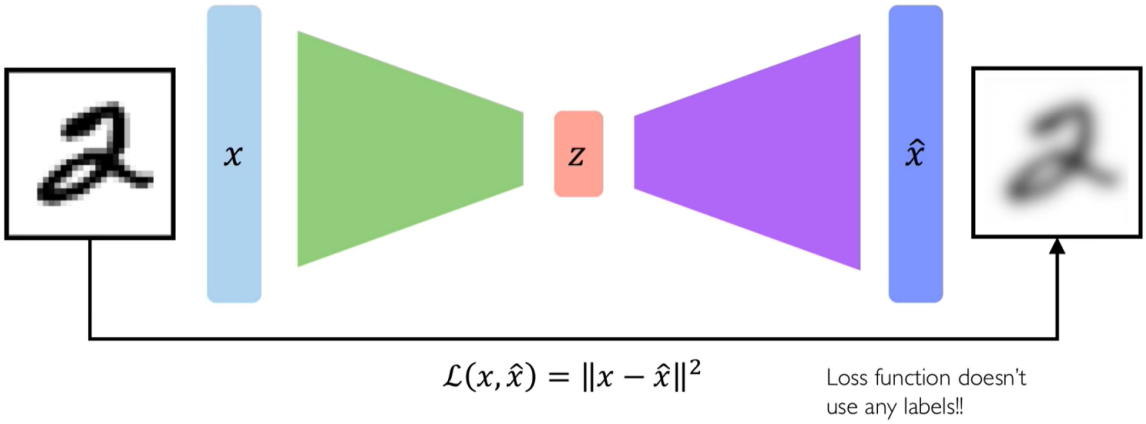


- "Encoder" learns mapping from the data x, to a low-dimentional latent space, z

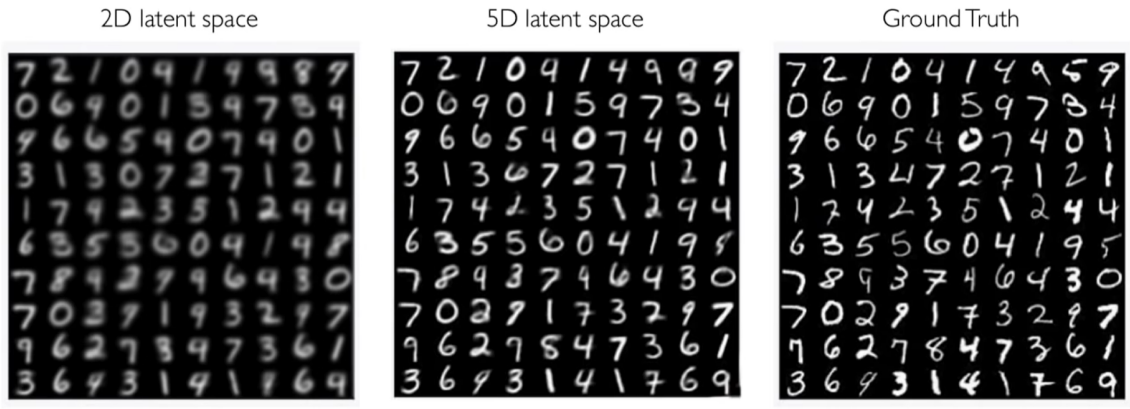
- This is not possible to learn
- Train the model to use these features to resconstruct the original data
- "Decoder" learns mapping back from latent, z, to reconstructed observation, $\hat{x}$



- To train this network, we want to make the reconstructed image as similar from the original image as possible
- Simplfied:



- Autoincoders is a form is compression, smaller latent space will cause larger training bottleneck

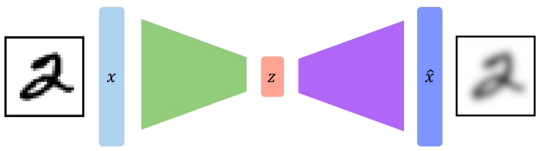


This Image is from a library called mnist

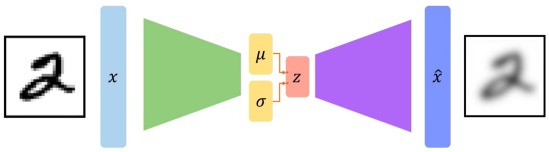
Autoencoder for representation learning

- Bottleneck layers force network to learn a compressed latent represenatation
- Reconstruction loss forces the latent representation to capture or encode as much information about data as possible
- Autoencoding = Automatically encoding data

Variational Autoencoders

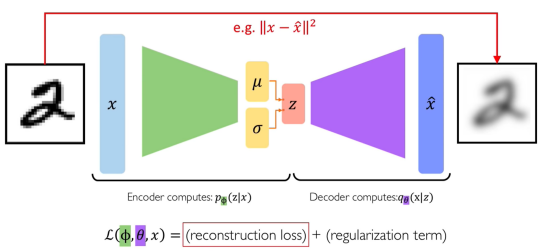


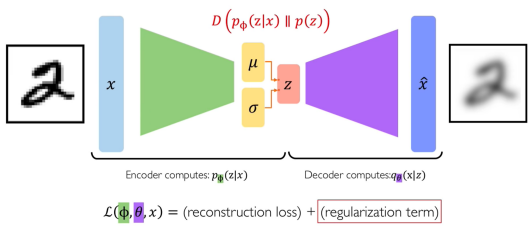
Traditional Autoencoders



Variational Autoencoders

- There is mean and sigma that parameterize a probability distribution for each of the latent variables



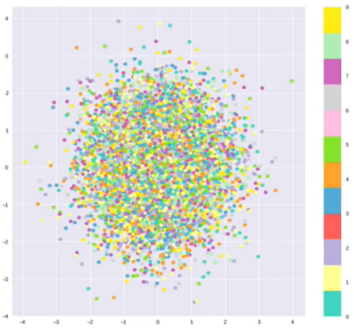


- This helps it not to overfit

Priors on latent Distribution

$$D\left(p_\phi(z|x) \parallel p(z)\right)$$
$$= -\frac{1}{2}\sum_{j=0}^{k-1}\left(\sigma_j + \mu_j^2 - 1 - \log \sigma_j\right)$$

KL-divergence between the two distributions



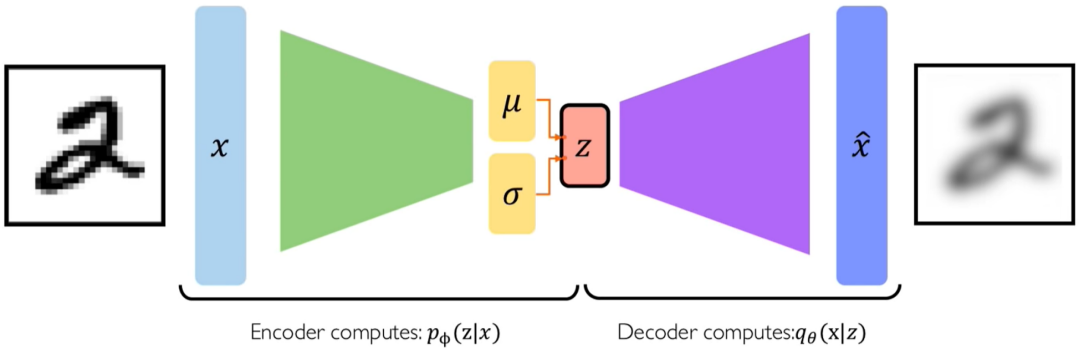
Common choice of prior:

$$p(z) = \mathcal{N}(\mu = 0, \sigma^2 = 1)$$

- Encourages encodings to distribute encodings evenly around the center of the latent space
- Penalize the network when it tries to "cheat" by clustering points in specific regions (ie. memorizing the data)

VAEs computation graph

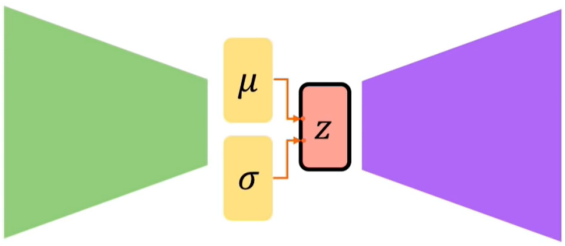
- It cannot backpropagate gradient through sampling layers



$$\mathcal{L}(\phi, \theta, x) = \text{(reconstruction loss)} + \text{(regularization term)}$$

Reparametrizing the sampling layer

- To reparametrize the sampling layer such that the network can be trained end to end



Key Idea:

~~$z \sim \mathcal{N}(\mu, \sigma^2)$~~

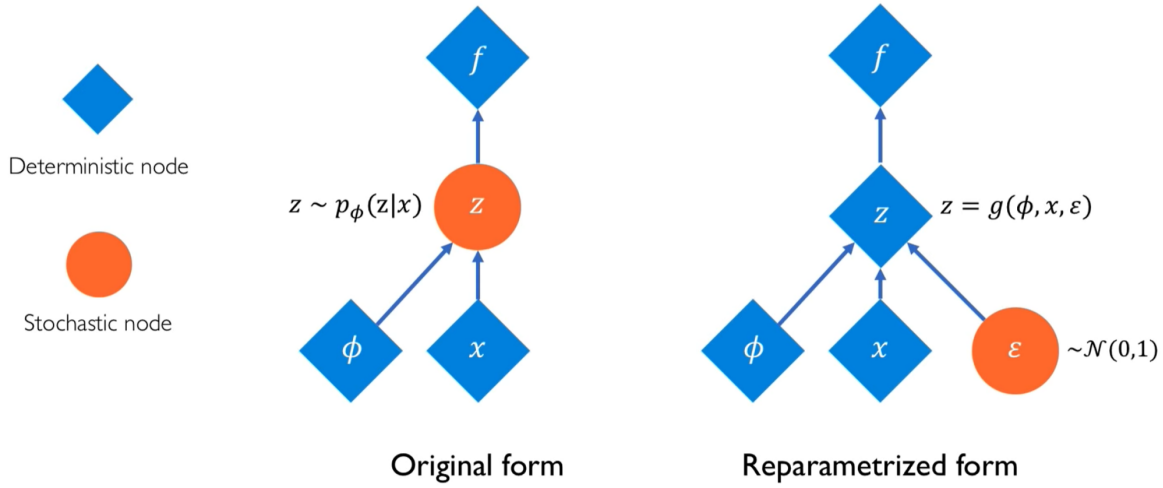
Consider the sampled latent vector z as a sum of

- a fixed μ vector,
- and fixed σ vector, scaled by random constants drawn from the prior distribution

$$\Rightarrow z = \mu + \sigma \odot \varepsilon$$

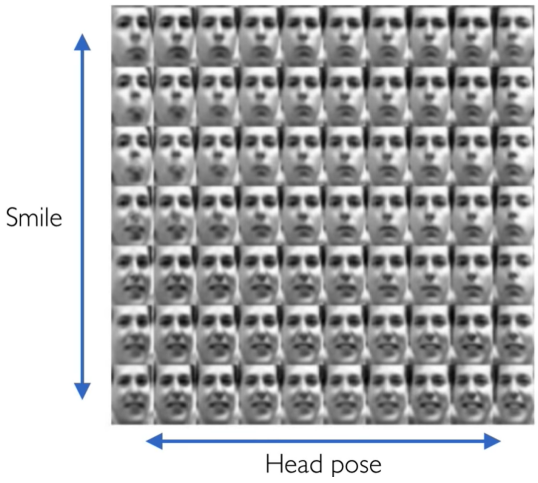
where $\varepsilon \sim \mathcal{N}(0,1)$

- The visualization of reparameterization



Latent Perturbation

- Slowly increase or decrease a single Latent variable keep all other variables fixed
- Different dimensions of z encodes different interpretable latent features
- Ideally we want latent variables that are uncorrelated with each other



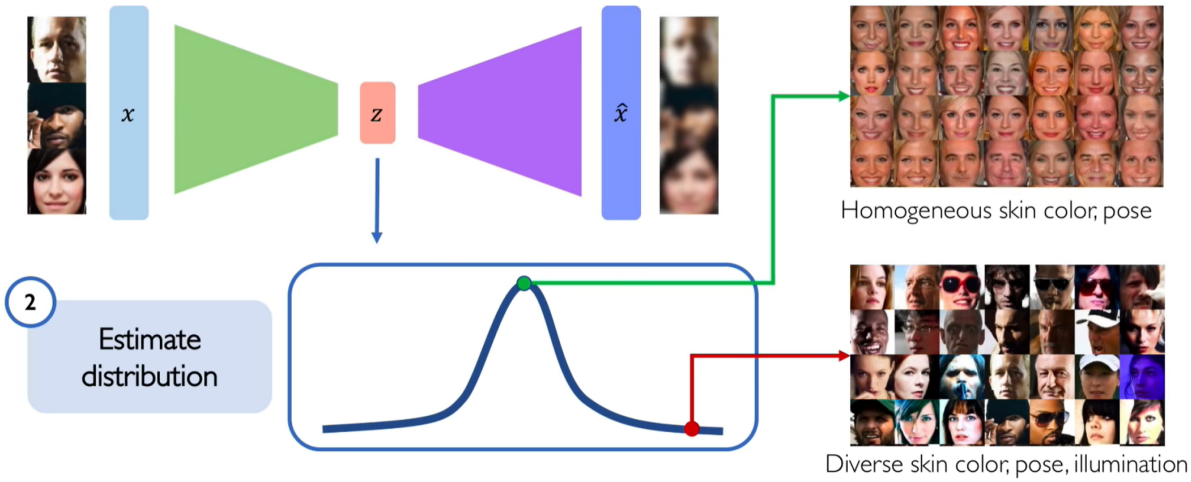
- Enforce diagonal prior on the latent variables to encourage independencies
- This idea is called **disentanglement.** To encourage the network to learn variables that's as independent as possible

Generative model debiasing

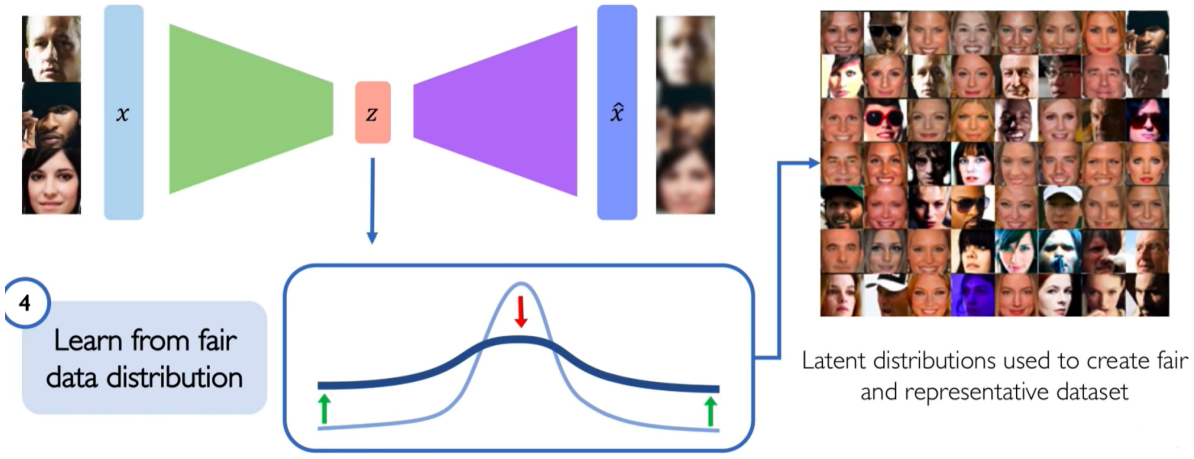
Homogeneous skin color, pose

Diverse skin color, pose, illumination

- we can use generative model to learn the generative model in the dataset
- And uncover which part of the features are underrepresented or overrepresented
- A VAE network is used to learn the underlying network in the training dataset in a unbiased and unsupervised manner
- Certain instances may be overrepresented in the dataset like skin color or pose
    - The possibility of chose certain overrepresented data will be unfairly high
- Certain instances like glasses, shadows, hats may be underrepresented which decrease the possibility of being selected during sampling

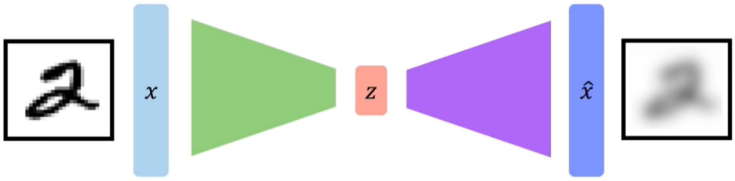


- The algorithm works by using the inferred distribution to adaptively resample the data during training
- Which is used to generate a more balanced and fair training dataset which ultimately result in a unbiased classifier



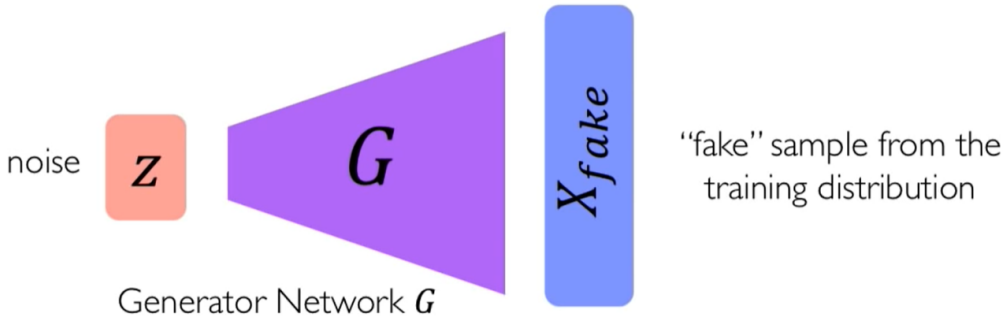
VAEs Summary

1. Compress representation of world to something we can use to learn
2. Reconstruction allows for unsupervised learning (no labels!)
3. Reparameterization trick to train end-to-end
4. Interpret hidden latent variables using perturbation
5. Generating new examples

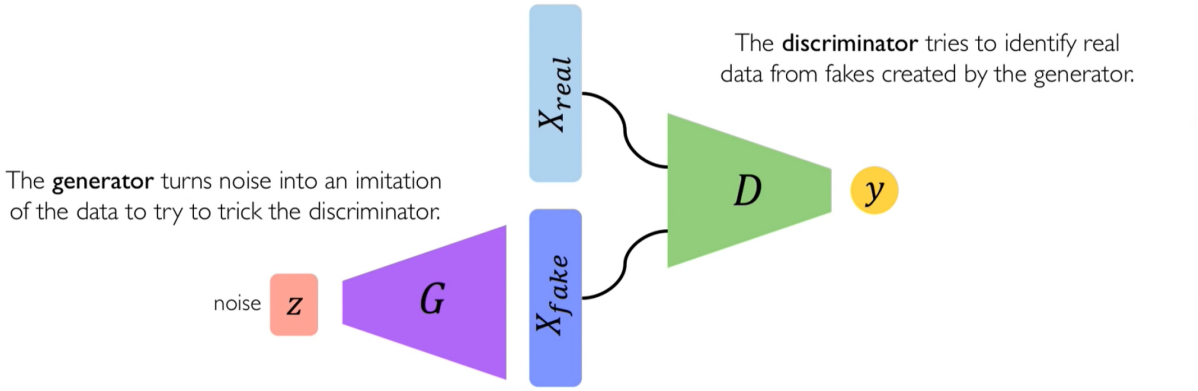


Generative Adversarial Networks (GANs)

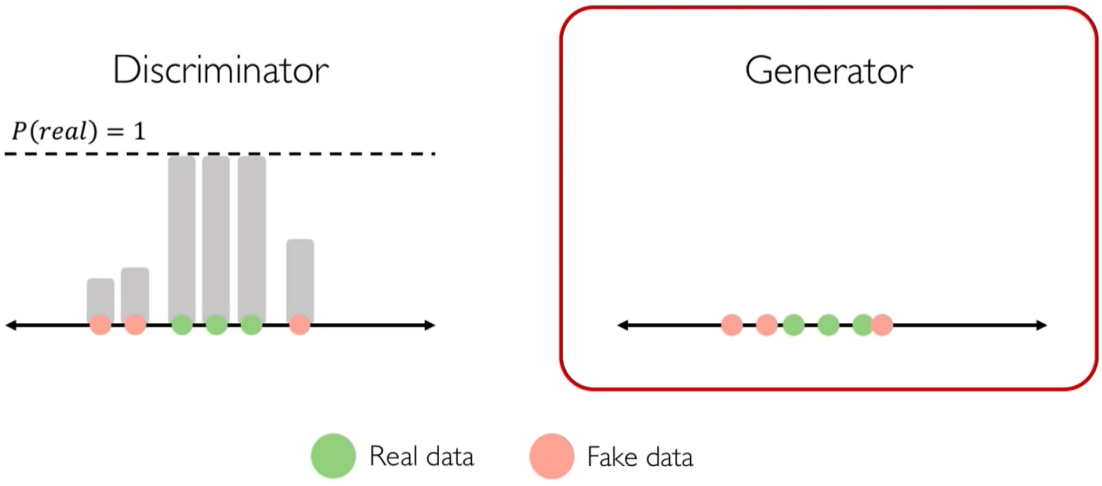
- The idea is to not explicitly model density and instead just sample to generate new instances
- The problem is you cannot sample from complex distribution directly
- Which the solution is to sample from something simple then learn a transformation to the training distribution



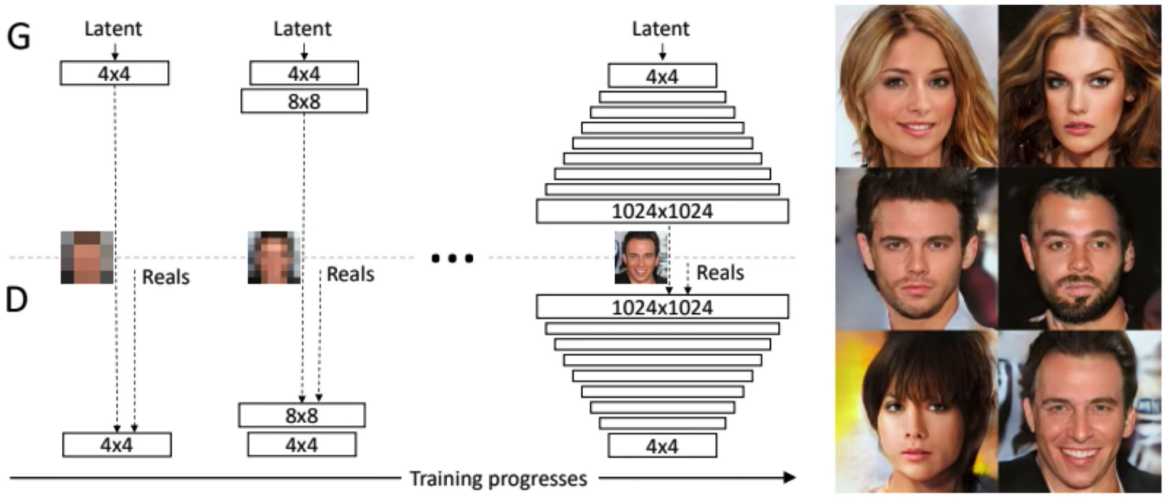
- Generative adversarial networks are ways to make a generative model by having two neural networks complete with each other



- generator creates a fake data and discriminator identities weather it's real or fake
- The better the discriminator the better the generator gets at generating as close to real image as possible



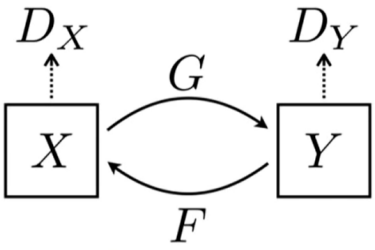
- The discriminator will try and train itself to identify weather the data is real or fake
- And as the generator gets closer and closer to the real data the better the result would be
- Eventually it's going to be hard for the discriminator to identity what's real and what's fake



- The discriminator and generator starts with very low spatial resolution
- As training progresses the layers are incrementally added to make the image look more realistic
- The images below are some fake celebrity faces generated with this approach

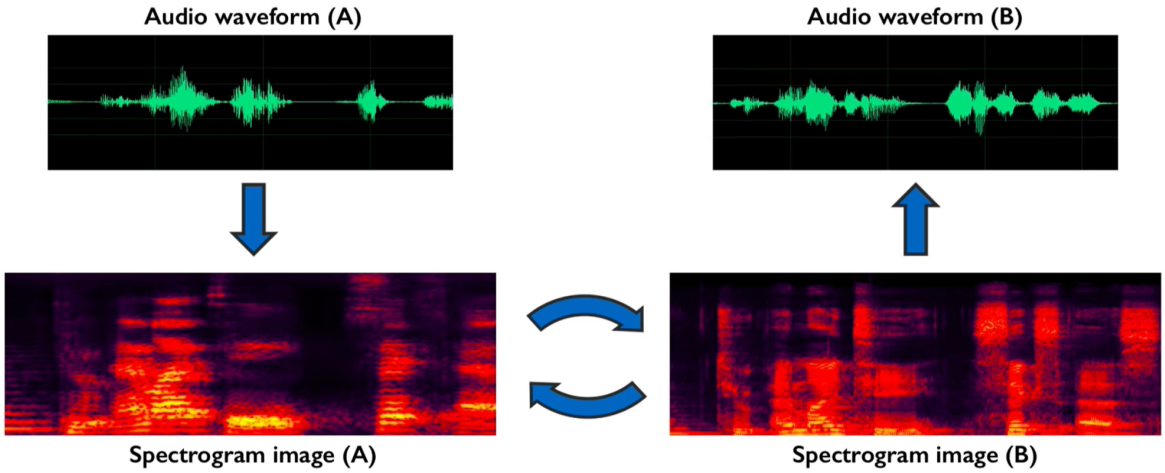
CycleGAN: Domain Transformation

- CycleGAN learns transformation across domains with unpair data



CycleGAN: Transforming Speech

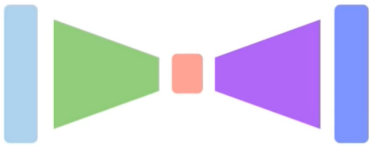
- Transform one representation to another



Summary

Autoencoders and Variational Autoencoders (VAEs)

Learn **lower-dimensional** latent space and **sample** to generate input reconstructions



Generative Adversarial Networks (GANs)

Competing **generator** and **discriminator** networks