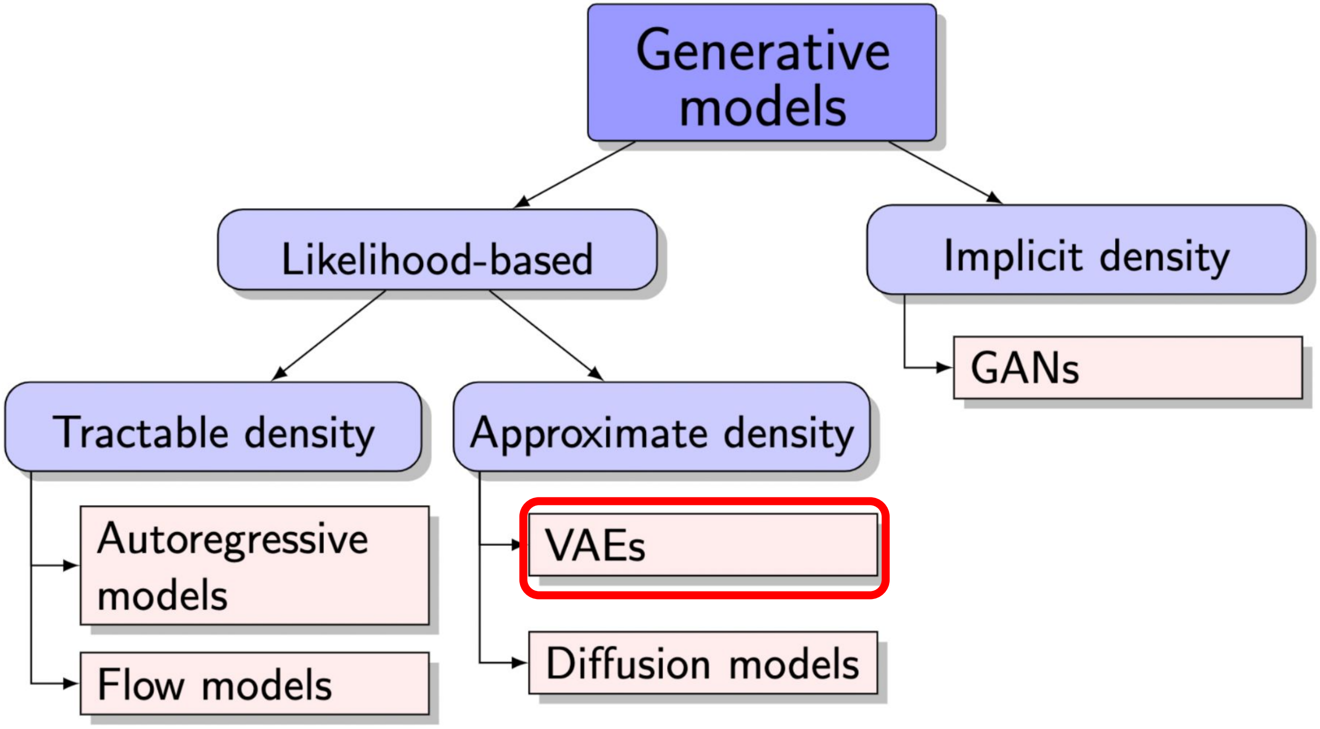

VAE

Today’s program

- Problem Formulation
- VAE
- VQ-VAE
- VQ-GAN
- D-VAE
- DALL-E

Generative Models zoo



Problem statement

We are given i.i.d. samples $\{\mathbf{x}_i\}_{i=1}^n \in X$ (e.g. $X = \mathbb{R}^m$) from unknown distribution $\pi(\mathbf{x})$.

Goal

We would like to learn a distribution $\pi(\mathbf{x})$ for

- evaluating $\pi(\mathbf{x})$ for new samples (how likely to get object $\mathbf{x}$?);
- sampling from $\pi(\mathbf{x})$ (to get new objects $\mathbf{x} \sim \pi(\mathbf{x})$).

Challenge

Data is complex and high-dimensional. E.g. the dataset of images lies in the space $X \subset \mathbb{R}^{\text{width} \times \text{height} \times \text{channels}}$.

Divergences

Fix probabilistic model $p(\mathbf{x}|\boldsymbol{\theta})$ – the set of parameterized distributions.
Instead of searching true $\pi(\mathbf{x})$ over all probability distributions, learn function approximation $p(\mathbf{x}|\boldsymbol{\theta}) \approx \pi(\mathbf{x})$.

What is a divergence?

Let $\mathcal{S}$ be the set of all possible probability distributions. Then $D : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$ is a divergence if

- $D(\pi||p) \geq 0$ for all $\pi, p \in \mathcal{S}$;
- $D(\pi||p) = 0$ if and only if $\pi \equiv p$.

General divergence minimization task

$$\min_{\boldsymbol{\theta}} D(\pi||p),$$

where $\pi(\mathbf{x})$ is a true data distribution, $p(\mathbf{x}|\boldsymbol{\theta})$ is a model distribution.

f-divergence family

$$D_f(\pi||p) = \mathbb{E}_{p(\mathbf{x})} f\left(\frac{\pi(\mathbf{x})}{p(\mathbf{x})}\right) = \int p(\mathbf{x}) f\left(\frac{\pi(\mathbf{x})}{p(\mathbf{x})}\right) d\mathbf{x}.$$

Here $f : \mathbb{R}_+ \to \mathbb{R}$ is a convex, lower semicontinuous function satisfying $f(1) = 0$.

Name	$D_f(P\|Q)$	Generator $f(u)$
Kullback-Leibler	$\int p(x) \log \frac{p(x)}{q(x)}\, \mathrm{d}x$	$u \log u$
Reverse KL	$\int q(x) \log \frac{q(x)}{p(x)}\, \mathrm{d}x$	$-\log u$
Pearson χ^2	$\int \frac{(q(x)-p(x))^2}{p(x)}\, \mathrm{d}x$	$(u-1)^2$
Squared Hellinger	$\int \left(\sqrt{p(x)} - \sqrt{q(x)}\right)^2 \mathrm{d}x$	$(\sqrt{u}-1)^2$
Jensen-Shannon	$\frac{1}{2} \int p(x) \log \frac{2p(x)}{p(x)+q(x)} + q(x) \log \frac{2q(x)}{p(x)+q(x)}\, \mathrm{d}x$	$-(u+1)\log \frac{1+u}{2} + u \log u$
GAN	$\int p(x) \log \frac{2p(x)}{p(x)+q(x)} + q(x) \log \frac{2q(x)}{p(x)+q(x)}\, \mathrm{d}x - \log(4)$	$u \log u - (u+1)\log(u+1)$

MLE

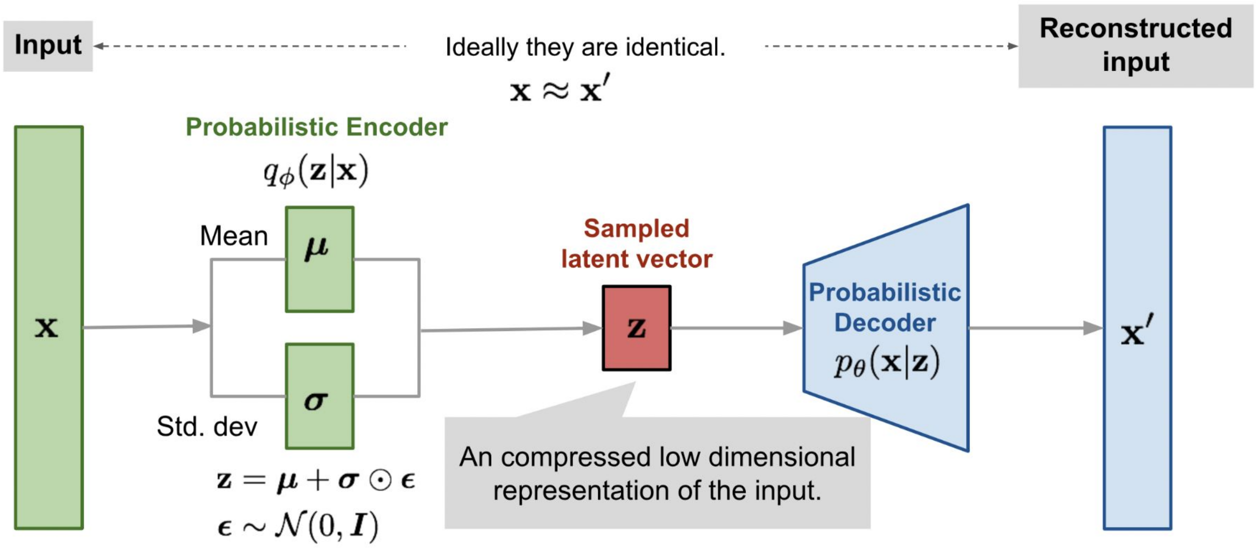
$$
\begin{aligned}
KL(\pi||p) &= \int \pi(\mathbf{x}) \log \frac{\pi(\mathbf{x})}{p(\mathbf{x}|\boldsymbol{\theta})} d\mathbf{x} \\
&= \int \pi(\mathbf{x}) \log \pi(\mathbf{x}) d\mathbf{x} - \int \pi(\mathbf{x}) \log p(\mathbf{x}|\boldsymbol{\theta}) d\mathbf{x} \\
&= -\mathbb{E}_{\pi(\mathbf{x})} \log p(\mathbf{x}|\boldsymbol{\theta}) + \text{const} \\
&\approx -\frac{1}{n} \sum_{i=1}^{n} \log p(\mathbf{x}_i|\boldsymbol{\theta}) + \text{const} \to \min_{\boldsymbol{\theta}} .
\end{aligned}
$$

Maximum likelihood estimation is equivalent to minimization of the Monte-Carlo estimate of forward KL.

Missed part

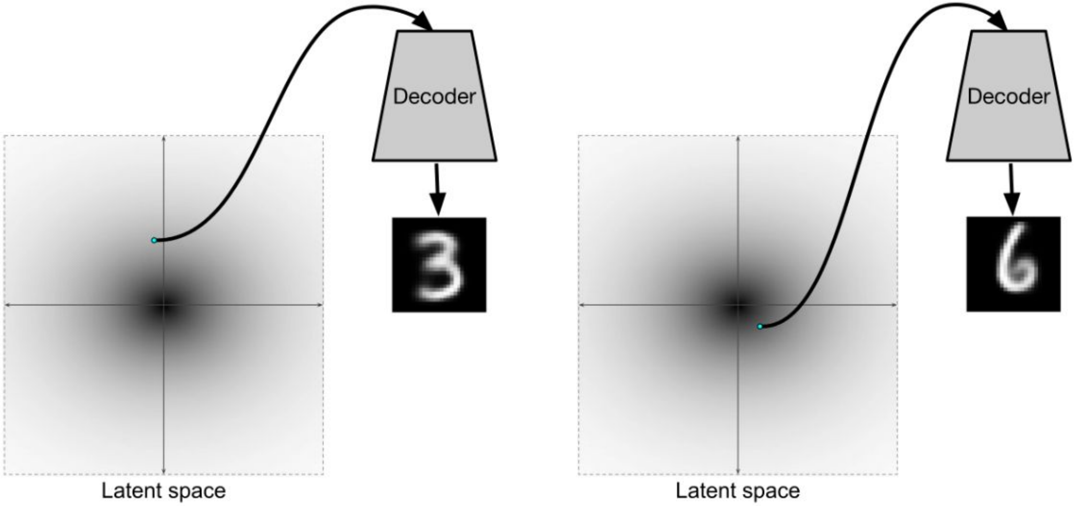
LVM, ELBO, EM-algorithm, Reparameterization trick…
But not today

VAE

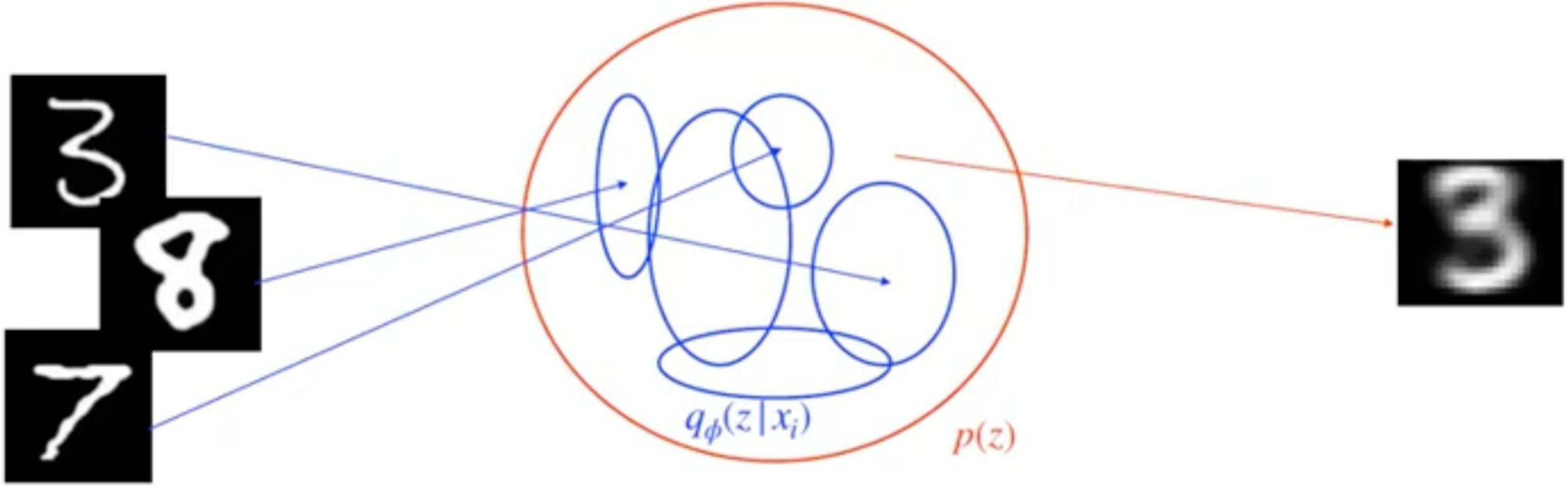


$$
\begin{aligned}
L_{\text{VAE}}(\theta, \phi) &= -\log p_\theta(\mathbf{x}) + D_{\text{KL}}(q_\phi(\mathbf{z}|\mathbf{x}) \| p_\theta(\mathbf{z}|\mathbf{x})) \\
&= -\mathbb{E}_{\mathbf{z} \sim q_\phi(\mathbf{z}|\mathbf{x})} \log p_\theta(\mathbf{x}|\mathbf{z}) + D_{\text{KL}}(q_\phi(\mathbf{z}|\mathbf{x}) \| p_\theta(\mathbf{z})) \\
\theta^*, \phi^* &= \arg\min_{\theta,\phi} L_{\text{VAE}}
\end{aligned}
$$

VAE

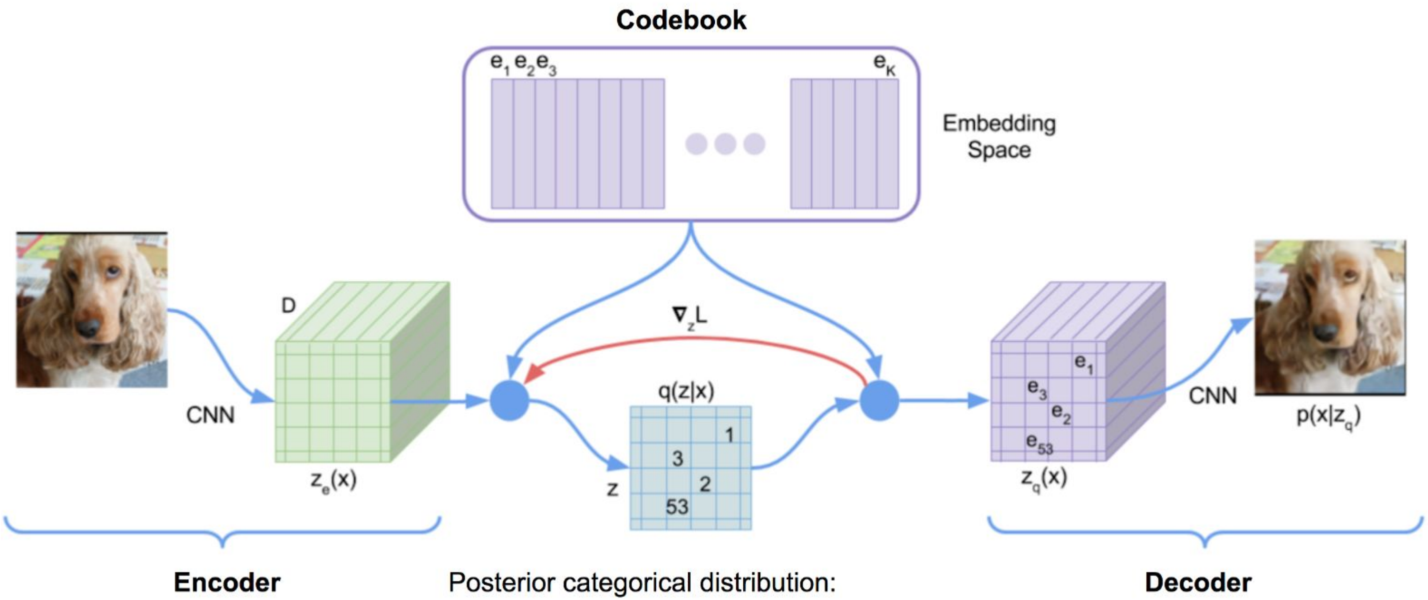


VAE



VAE

VQ-VAE

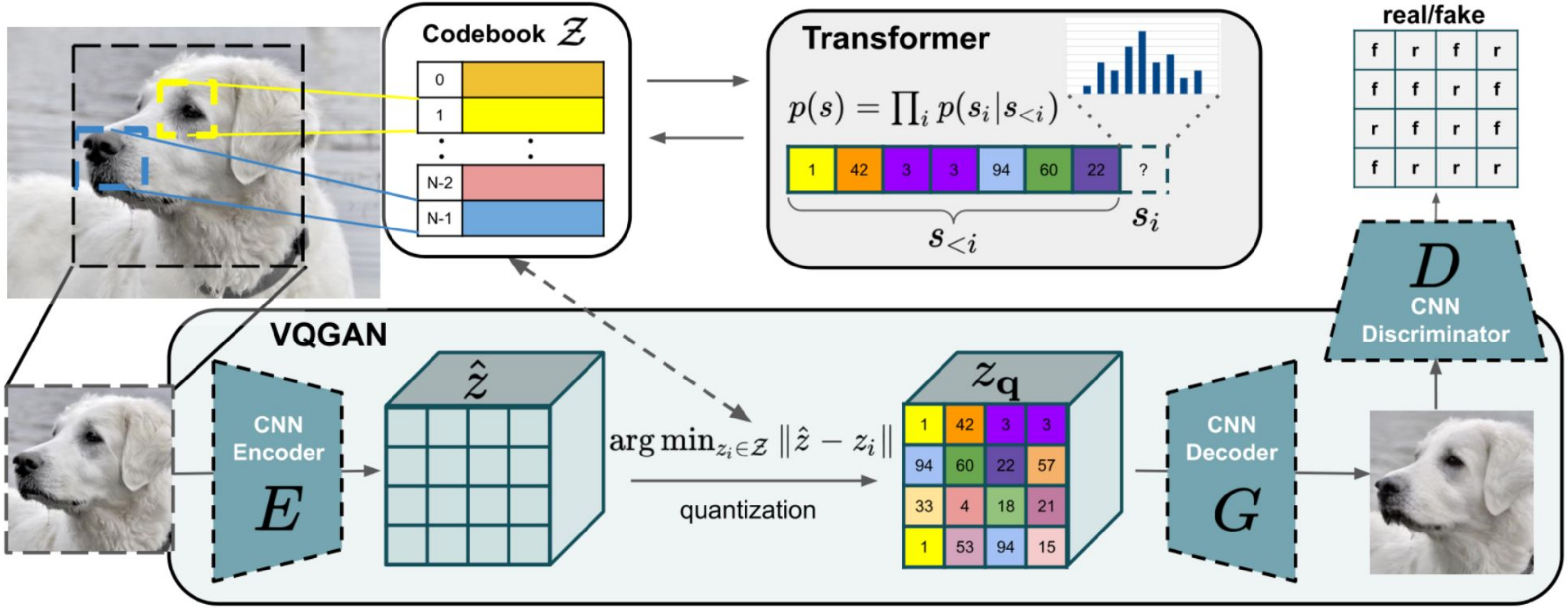


Posterior categorical distribution:

$$q(\mathbf{z} = \mathbf{e}_k|\mathbf{x}) = \begin{cases} 1 & \text{if } k = \arg\min_i \|\mathbf{z}_e(\mathbf{x}) - \mathbf{e}_i\|_2 \\ 0 & \text{otherwise.} \end{cases}$$

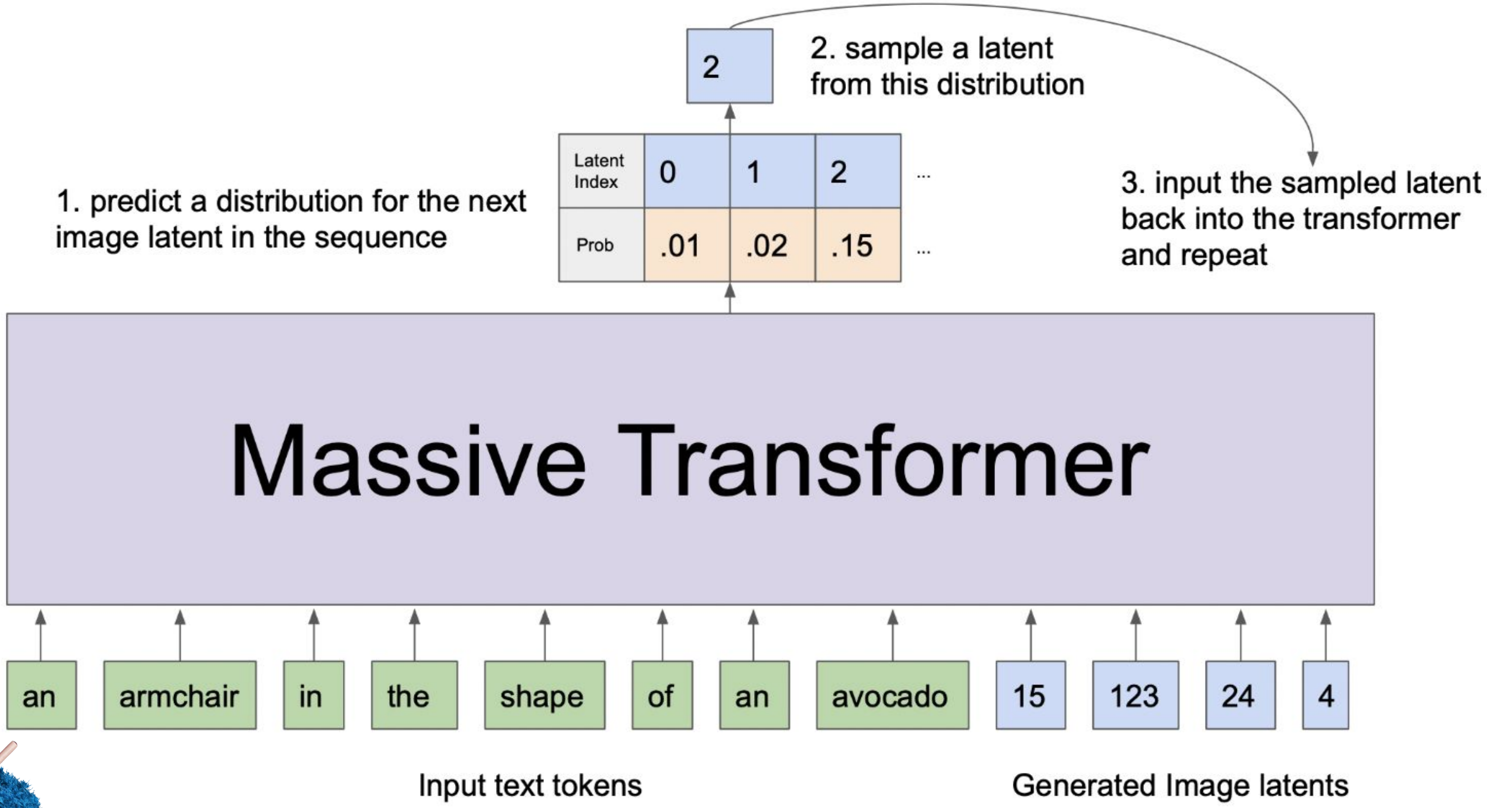
$$L = \underbrace{\|\mathbf{x} - D(\mathbf{e}_k)\|_2^2}_{\text{reconstruction loss}} + \underbrace{\|\text{sg}[E(\mathbf{x})] - \mathbf{e}_k\|_2^2}_{\text{VQ loss}} + \underbrace{\beta\|E(\mathbf{x}) - \text{sg}[\mathbf{e}_k]\|_2^2}_{\text{commitment loss}}$$

VQ-GAN



Codebook $\mathcal{Z}$
0
1
N-2
N-1
Transformer
$p(s) = \prod_i p(s_i|s_{<i})$
1 42 3 3 94 60 22 ?
$s_{<i}$
s_i
real/fake
f r f r
f f r f
r f r f
f r r r
D
CNN Discriminator
VQGAN
CNN Encoder
E
$\hat{z}$
$\arg\min_{z_i \in \mathcal{Z}} \|\hat{z} - z_i\|$
quantization
$z_{\mathbf{q}}$
1 42 3 3
94 60 22 57
33 4 18 21
1 53 94 15
CNN Decoder
G

VQ-GAN


2
2. sample a latent
from this distribution
Latent Index
0
1
2
...
Prob
.01
.02
.15
...
1. predict a distribution for the next image latent in the sequence
3. input the sampled latent back into the transformer and repeat
Massive Transformer
an
armchair
in
the
shape
of
an
avocado
15
123
24
4
Input text tokens
Generated Image latents

VQ-VAE

Samples 1024x1024

Samples diversity

VQ-VAE (Proposed) BigGAN deep

VQ-GAN

D-VAE

If $\mathbf{z}$ is a discrete random variable we cannot differentiate through it.

Gumbel-Max trick

Let $G_k \sim \text{Gumbel}$ for $k = 1, \ldots, K$, i.e. $G = -\log(\log u)$, $u \sim \text{Uniform}[0, 1]$. Then a discrete random variable

$$z = \arg\max_k (\log \pi_k + G_k), \quad \sum_k \pi_k = 1$$

has a categorical distribution $z \sim \text{Categorical}(\boldsymbol{\pi})$ $(P(z = k) = \pi_k)$.

Problem: We still have non-differentiable $\arg\max$ operation.

Gumbel-Softmax relaxation

$$z_k = \frac{\exp((\log \pi_k + G_k)/\tau)}{\sum_{j=1}^{K} \exp((\log \pi_j + G_j)/\tau)}, \quad k = 1, \ldots, K.$$

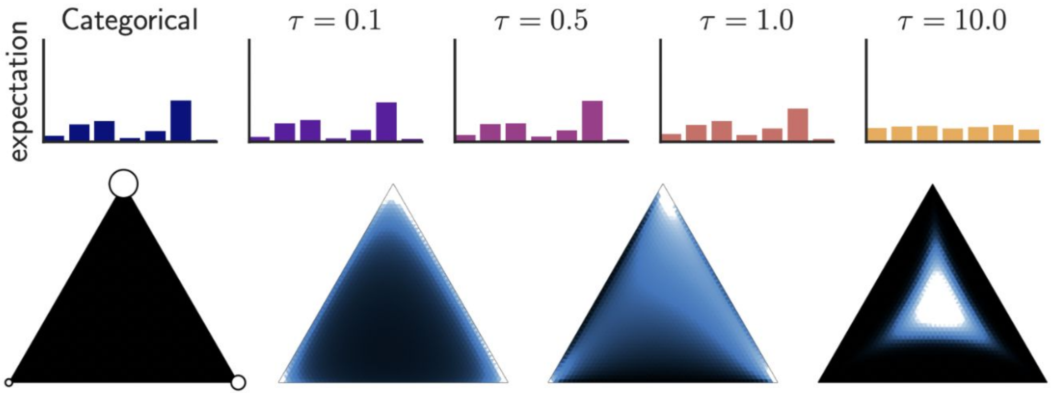
Here τ is a temperature parameter.

Gumbel-Softmax relaxation

Concrete distribution = continuous + discrete

$$z_k = \frac{\exp((\log \pi_k + G_k)/\tau)}{\sum_{j=1}^{K} \exp((\log \pi_j + G_j)/\tau)}, \quad k = 1, \ldots, K.$$

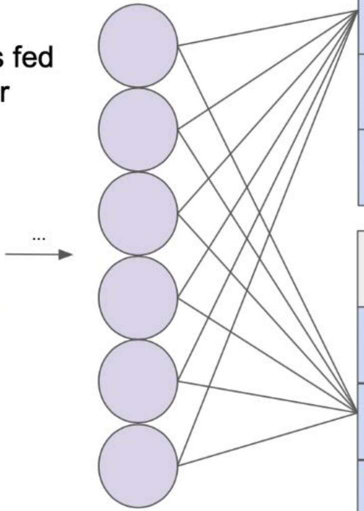
Here τ is a temperature parameter. Now we have differentiable operation.



(a) $\lambda = 0$ (b) $\lambda = 1/2$ (c) $\lambda = 1$ (d) $\lambda = 2$

D-VAE

1. an image is fed to the encoder network



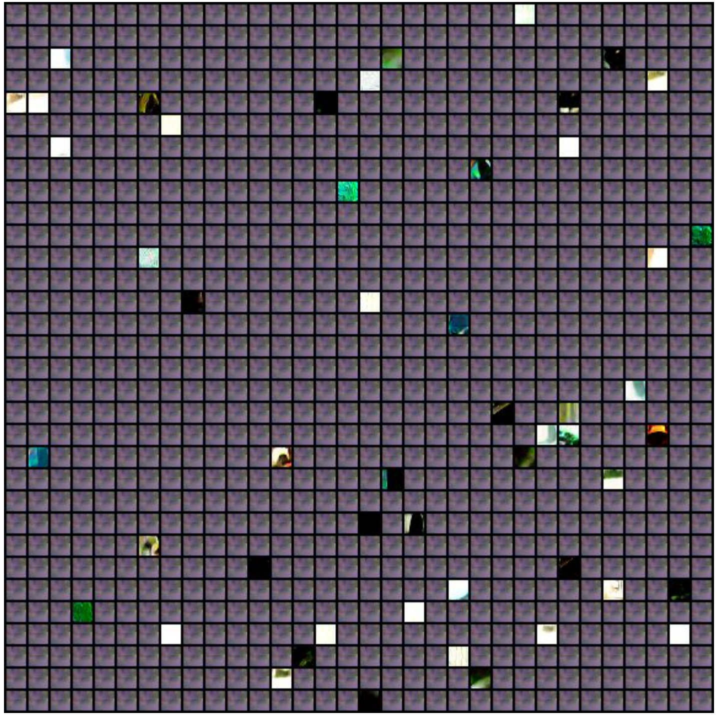
2. the encoder outputs distributions over codebook vectors

Latent Index	Prob
0	0.2
1	0.3
2	0.4
3	0.1

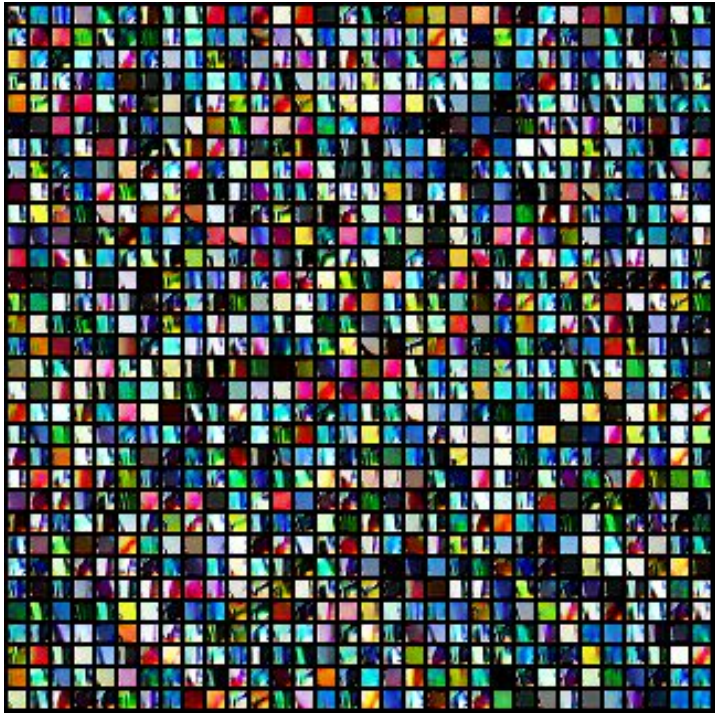
Latent Index	Prob
0	0.4
1	0.2
2	0.15
3	0.25

Latent Index	Codebook Vector
0	[0.01, -2.3, 5.6, 0.04, -0.1, 8.92, 3.24, ...]
1	[5.4, 0.65, 0.2, 4.6, 8.9, -2.43, 0.07, ...]
2	[9.78, 0.67, -3.4, 0.2, -1.0, 7.2, 13.8, ...]
3	[2.45, -8.9, 0.3, 2.04, -0.89, 19.1, 0.3, ...]

Codebook visualization

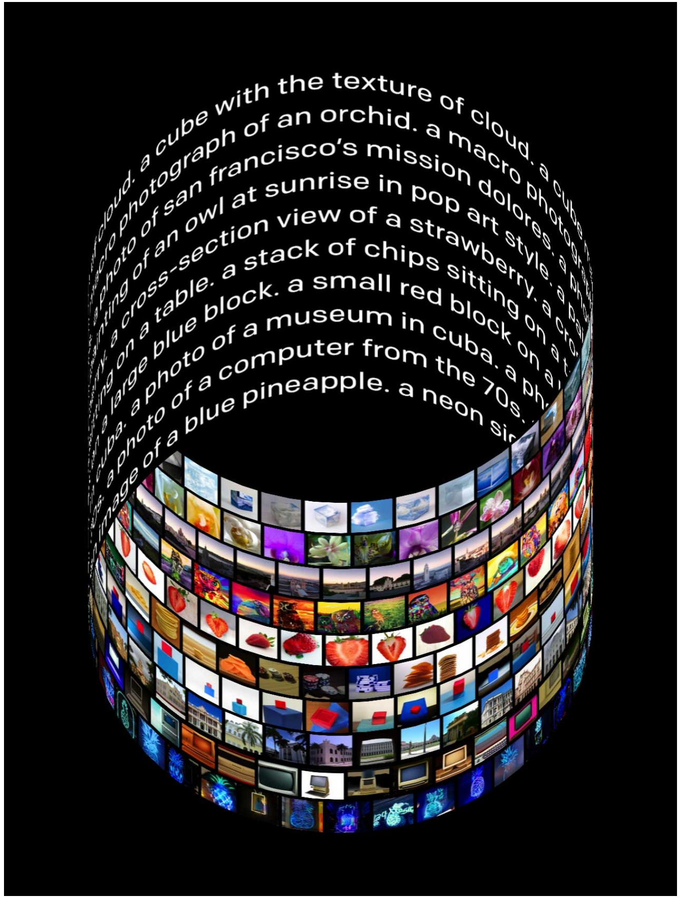


Naive codebook visualization



Gumbel codebook visualization

DALL-E


a cube with the texture of cloud.
photograph of an orchid. a macro
a photo of san francisco's mission dolores.
an owl at sunrise in pop art style.
a cross-section view of a strawberry.
on a table. a stack of chips sitting on a
a large blue block. a small red block on a
a photo of a museum in cuba.
a photo of a computer from the 70s.
a blue pineapple. a neon

Questions?

LinkedIn: Kiram Al-Kharba
tg: @al_kharba