Varying Loss

0.5
1.0
1.0
1.5
2.0
2.5
0.004
0.002
5

In GANs, a regularizer:

I-GAN loss

Generator loss

Discriminator loss

Discriminator

Generator

Real images

Generator

Fake sample

Latent codes

Real sample

Noise

InfoGAN

InfoGAN [1] learns disentangled and interpretable representations by maximizing the mutual information between a subset of the latent variables and the GAN generated sample. This is done through the addition of an extra term to the objective function.

**Background: GANs and Mutual Information**

GANs are trained by a two-player minimax game between Discriminator $D$ and Generator $G$ with value function $V_{GAN}(D, G)$:

$$\min_D \max_G V_{GAN}(D, G) = E_{x\sim p_{data}}[\log D(x)] + E_{z\sim p_z}[\log(1 - D(G(z)))]$$

The mutual information (MI) $I(X; Y)$ between random variables $X$ (all images) and $Y$ (real/fake label) is:

$$I(X; Y) = H(X) - H(X|Y) = D_{KL}(p_{X|Y} || p_X \otimes p_Y)$$

As the MI is intractable, we use a variational lower bound on $I(c; G(z, c))$:

$$I(c; G(z, c)) \geq E_{z\sim p_z}[E_{x\sim p_{data}}[\log q(y|x)]] + H(c) = L_I(G, Q)$$

In GANs, $G$ minimizes the same variational lower bound on $I(X; Y)$ [5]:

$$I(X, Y) \geq E_{x\sim p_{data}}[E_{z\sim p_z}[\log q(y|x)]] + H(Y) = E_{z\sim p_z}[\log q(y|z = 1|x) + E_{x\sim p_{data}}[\log(1 - q(y = 1|x)]) + H(Y)$$

**Method: Information Regularizing GANs**

Maximize the MI between Generator output $G(z, c)$ and latent codes $c$ as a regularizer:

$$\min_G \max_D V_{G}^{\text{GAN}}(D, G) = V_{GAN}(D, G) - \lambda L_I(G, Q)$$

**InfoGAN vs Vanilla GAN**

(a) Varying $c_1$ on InfoGAN - digit

(b) Varying $c_2$ on a Vanilla GAN

(a) Varying $c_3$ on InfoGAN - thickness

(b) Varying $c_2$ on a Vanilla GAN

(a) Varying $c_2$ on InfoGAN - tilt

(b) Varying $c_3$ on a Vanilla GAN

**Stability Analysis and Mutual Information**

Stability on Info-WGAN

Stability on Info-WGAN

Stability on Info-WGAN

**Wasserstein GANs (WGANs) [2]** optimise Wasserstein distance. Benefits:

1. This metric is less prone to model collapse and vanishing gradients.
2. Introduction of weight clipping to enforce Lipschitz constraint.

$$\min_G \max_D V_{WGAN}^{\text{D, G}}(D, G) = E_{x\sim p_{data}}[D(x)] - E_{z\sim p_z}[D(G(z))']$$

$$\min_G \max_D V_{WGAN}^{\text{D, G}}(D, G) = V_{WGAN}(D, G) - \lambda L_I(G, Q)$$

**Findings:** Info-WGAN performs satisfactorily with discrete latent codes although it finds difficulties interpreting the continuous ones.

**MINE • GANs**

MINE [3] is a lower bound on the MI, obtained from the Donsker-Varadhan representation of the KL divergence by restricting function $T$ to be parametrized by a neural net.

$$D_{KL}(P_{X,Y} || P_X \otimes P_Y) \geq \sup_{T \in \Theta} E_{P_{X,Y}}[T] - \log E_{P_X \otimes P_Y}[e^T]$$

**Findings:** MineGAN is only able to learn the continuous codes and is harder to train: tricks were needed to stabilize the MI.

**Future work: Disentanglement of VAEs**

Variational Autoencoders (VAEs) [4]:

- Have a more continuous and smooth latent space.
- Provide a more structured and interpretable latent space.
- Can perform interpolation in the latent space.

$$L_{MINE} = E_{x\sim p(x)}[\log Q(x|z) - \log Q(z|x)] \rightarrow L_{VAE} = L_{MINE} + \lambda L_I$$

- Can we predict the appropriate number of latent codes in an unsupervised manner for different datasets?