Boundary Equilibrium GAN

Berthelot, Schumm and Metz
at Google
2017.03

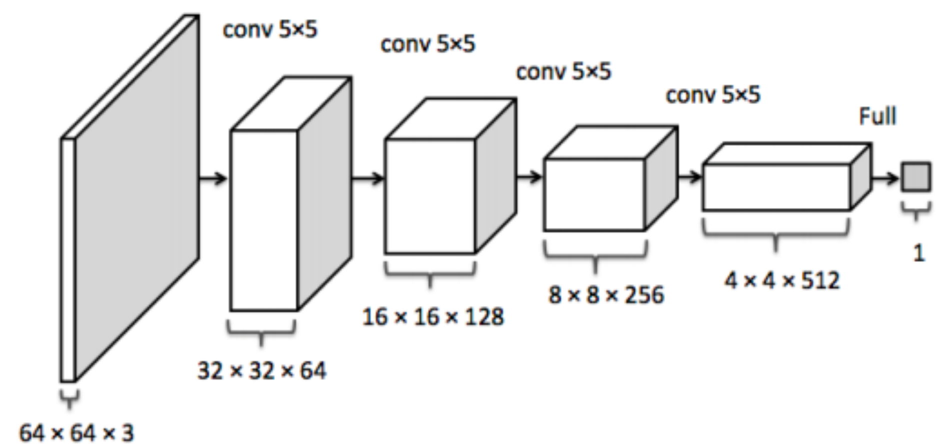
07/12/2018 Ziyi

Boundary Equilibrium GAN

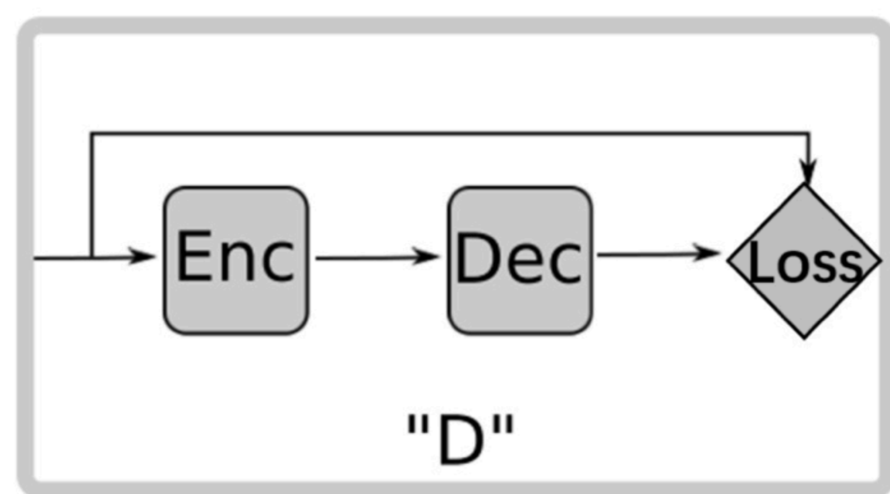
- HighLights
  - AutoEncoder architecture as Discriminator
  - Distribution metric among AE-Loss (rather than data)
  - Boundary Equilibrium and Convergence Measure
- Experiments
- Some Problems

AutoEncoder as Discriminator

- Common GANs use a discriminative network (normally last layer as softmax)



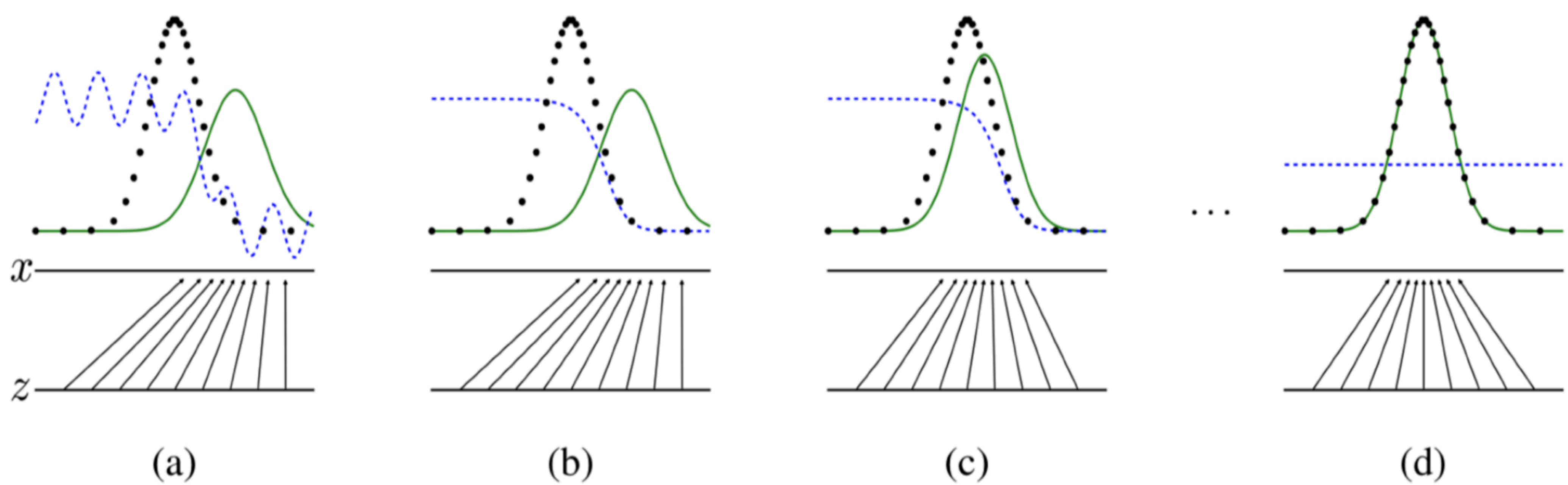
- BEGAN takes an AutoEncoder as discriminator



Change of Distribution Assumption

- Original GAN, model **data-points** into distributions

Blue dashed line —— discriminative distribution (D)
Black dashed line —— data distribution Px
Green solid line —— generative distribution Pg (G)



Lower horizontal line —— domain of z (noise)
Upper horizontal line —— domain of x

Change of Distribution Assumption

- Original GAN, model **data-points** into distributions

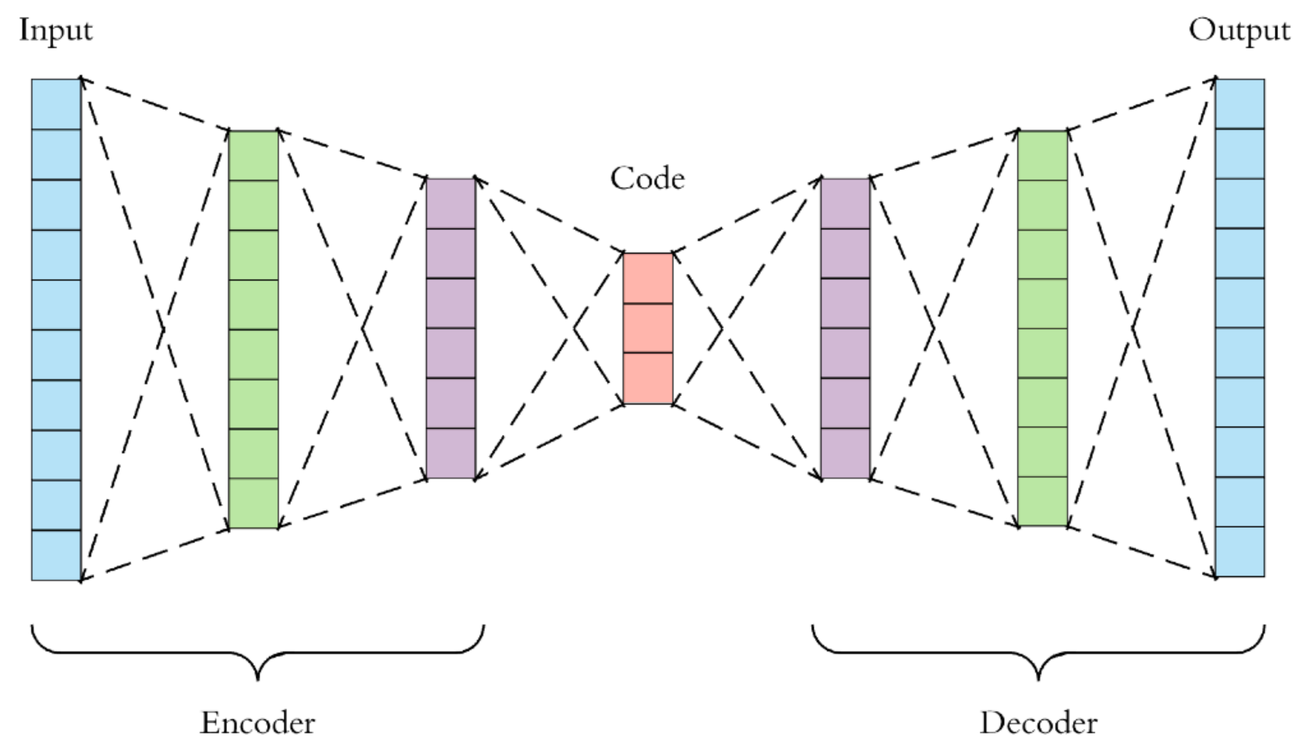
$$\mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))].$$

- Thus G generates samples close to real image distribution

$$\min_G \max_D V(D, G)$$

Change of Distribution Assumption

- BEGAN, model **AutoEncoder Loss** into distributions



$$\mathcal{L}(v) = |v - D(v)|^{\eta} \text{ where } \begin{cases} D : \mathbb{R}^{N_x} \mapsto \mathbb{R}^{N_x} & \text{is the autoencoder function.} \\ \eta \in \{1, 2\} & \text{is the target norm.} \\ v \in \mathbb{R}^{N_x} & \text{is a sample of dimension } N_x. \end{cases}$$

Change of Distribution Assumption

- BEGAN, model **AutoEncoder Loss** into distributions

$$\mathcal{L}(v) = |v - D(v)|^\eta \text{ where } \begin{cases} D : \mathbb{R}^{N_x} \mapsto \mathbb{R}^{N_x} & \text{is the autoencoder function.} \\ \eta \in \{1, 2\} & \text{is the target norm.} \\ v \in \mathbb{R}^{N_x} & \text{is a sample of dimension } N_x. \end{cases}$$

- We can formulate W-distance as

$$W_1(\mu_1, \mu_2) = \inf_{\gamma \in \Gamma(\mu_1, \mu_2)} \mathbb{E}_{(x_1, x_2) \sim \gamma}[|x_1 - x_2|]$$

μ_1, μ_2 represents the distribution of $\mathcal{L}(x)$and $\mathcal{L}(G(z))$

- With Jensen's Inequality, the lower bound was derived

$$\inf \mathbb{E}[|x_1 - x_2|] \geqslant \inf |\mathbb{E}[x_1 - x_2]| = |m_1 - m_2|$$

m_1, m_2 represents the mean of $\mu_1$and μ_2

Change of Distribution Assumption

给定两个一维的正态分布 $\mu_1 = \mathcal{N}(m_1, c_1), \quad \mu_2 = \mathcal{N}(m_2, c_2)$，不难计算出它们之间的 Wasserstein distance：

$$W(\mu_1, \mu_2)^2 = (m_1 - m_2)^2 + (c_1 + c_2 - 2\sqrt{c_1 c_2})$$

> **假设：**
> $\dfrac{c_1 + c_2 - 2\sqrt{c_1 c_2}}{(m_1 - m_2)^2}$ **是常数，或者是关于W的递增函数。**

在该假设下，我们可以只通过最小化 $(m_1 - m_2)^2$ 来最小化 $W(\mu_1, \mu_2)^2$。也就是说，$W(\mu_1, \mu_2) \propto |m_1 - m_2|$。

BEGAN Objectives

- As two options to maximize the lower bound $|m_1 - m_2|$

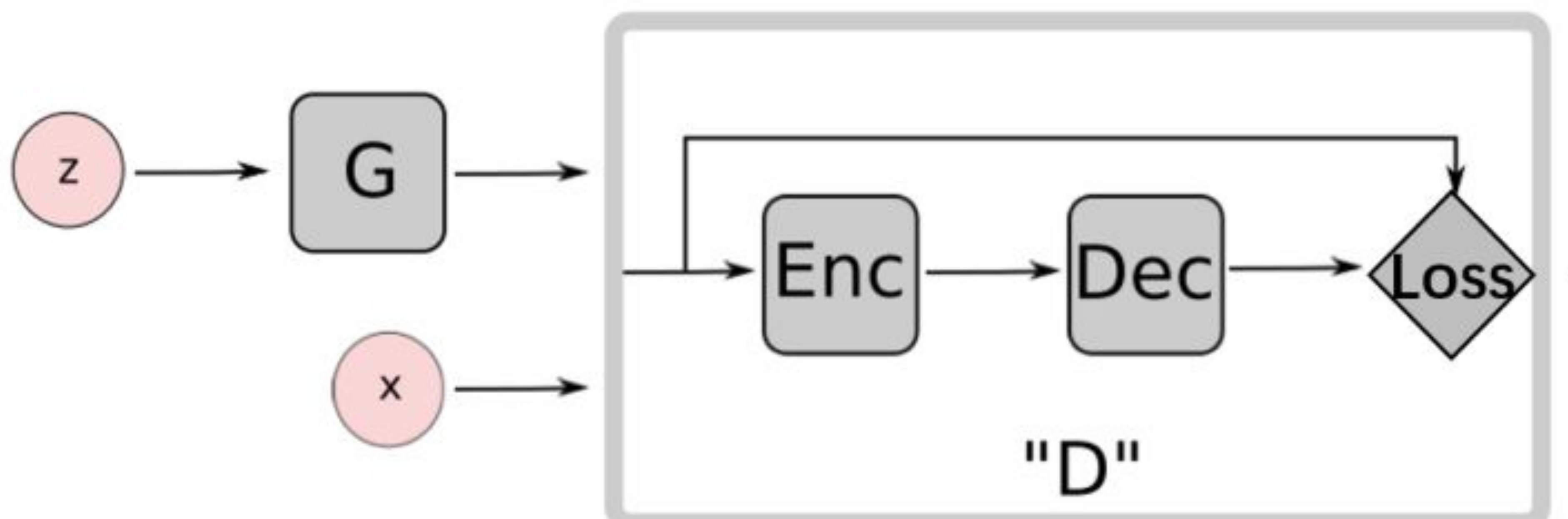
$$(a) \begin{cases} W_1(\mu_1, \mu_2) \geqslant m_1 - m_2 \\ m_1 \to \infty \\ m_2 \to 0 \end{cases} \quad \text{or } (b) \begin{cases} W_1(\mu_1, \mu_2) \geqslant m_2 - m_1 \\ m_1 \to 0 \\ m_2 \to \infty \end{cases}$$

- As minimizing m1 naturally leads to auto-encoding real samples, option (b) is selected

$$\begin{cases} \mathcal{L}_D = \mathcal{L}(x; \theta_D) - \mathcal{L}(G(z_D; \theta_G); \theta_D) & \text{for } \theta_D \\ \mathcal{L}_G = -\mathcal{L}_D & \text{for } \theta_G \end{cases}$$

BEGAN Objectives

- The simple architecture could be simplified as



Balancing Equilibrium

- We can consider equilibrium between G and D when

$$\mathbb{E}\left[\mathcal{L}(x)\right] = \mathbb{E}\left[\mathcal{L}(G(z))\right]$$

- AutoEncoder has two competing goals: 1.autoencode real images 2.tell real from generated.
  Hyper-parameter introduced to balance two goals

$$\gamma = \frac{\mathbb{E}\left[\mathcal{L}(G(z))\right]}{\mathbb{E}\left[\mathcal{L}(x)\right]} \quad \gamma \in [0, 1]$$

Balancing Equilibrium

- About the value of Gamma

$$\gamma = \frac{\mathbb{E}\left[\mathcal{L}(G(z))\right]}{\mathbb{E}\left[\mathcal{L}(x)\right]} \quad \gamma \in [0, 1]$$

$$(b) \begin{cases} W_1(\mu_1, \mu_2) \geqslant m_2 - m_1 \\ m_1 \to 0 \\ m_2 \to \infty \end{cases}$$

- At the beginning of training, $\mathcal{L}(x) >> \mathcal{L}(G(z))$ as real images are far complicated than generated noises, thus Gamma takes value smaller than 1 (actually it doesn't matter)

BEGAN Objectives

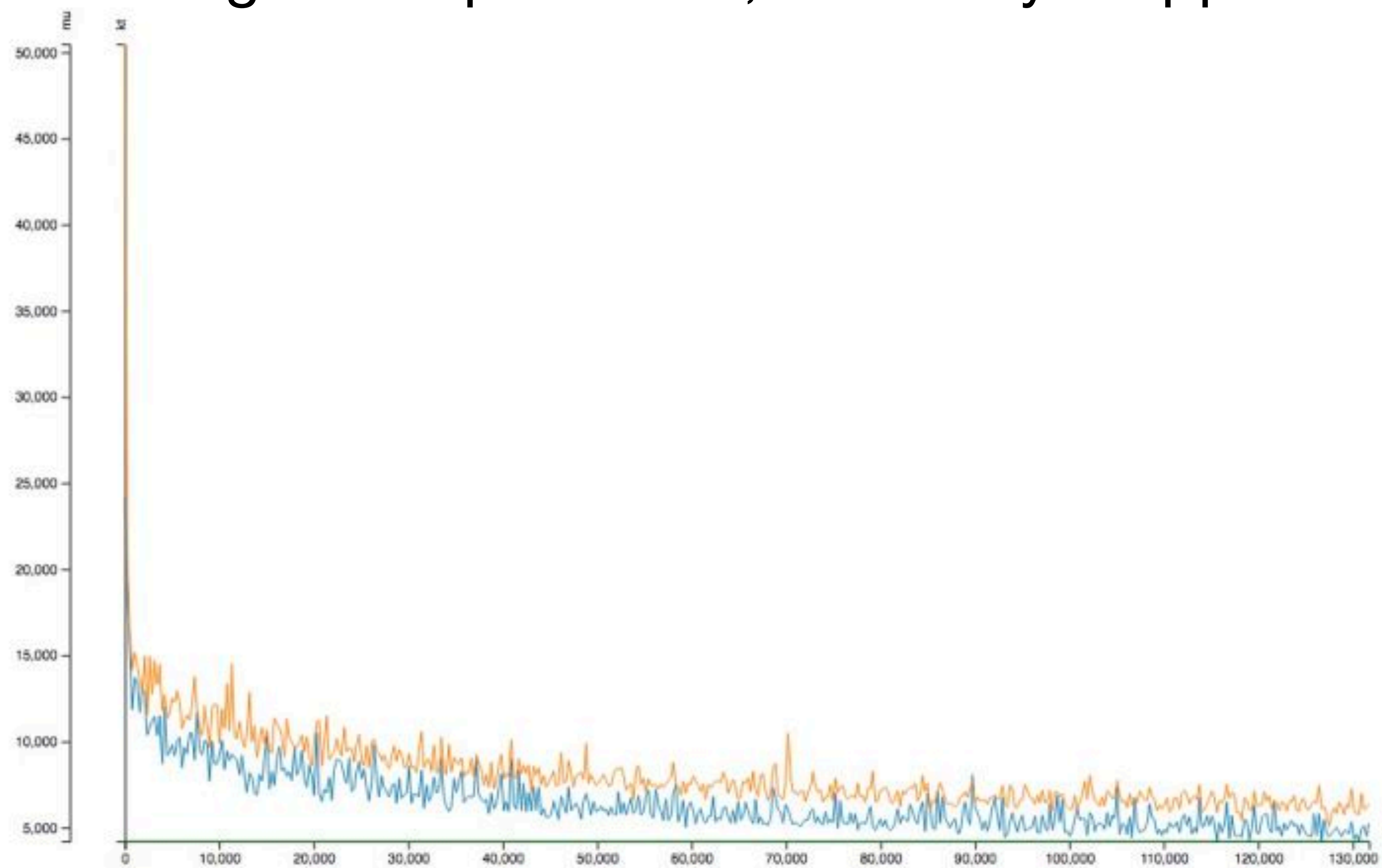
$$\begin{cases} \mathcal{L}_D = \mathcal{L}(x) - k_t.\mathcal{L}(G(z_D)) & \text{for } \theta_D \\ \mathcal{L}_G = \mathcal{L}(G(z_G)) & \text{for } \theta_G \\ k_{t+1} = k_t + \lambda_k(\gamma\mathcal{L}(x) - \mathcal{L}(G(z_G))) & \text{for each training step } t \end{cases}$$

$$k_t \in [0, 1]$$

- Borrow idea from Proportional Control Theory to maintain the Equilibrium, k is introduced to control how much emphasis to put on $\mathcal{L}(G(z_D))$

BEGAN Objectives

- But during real experiments, k is always clipped as 0



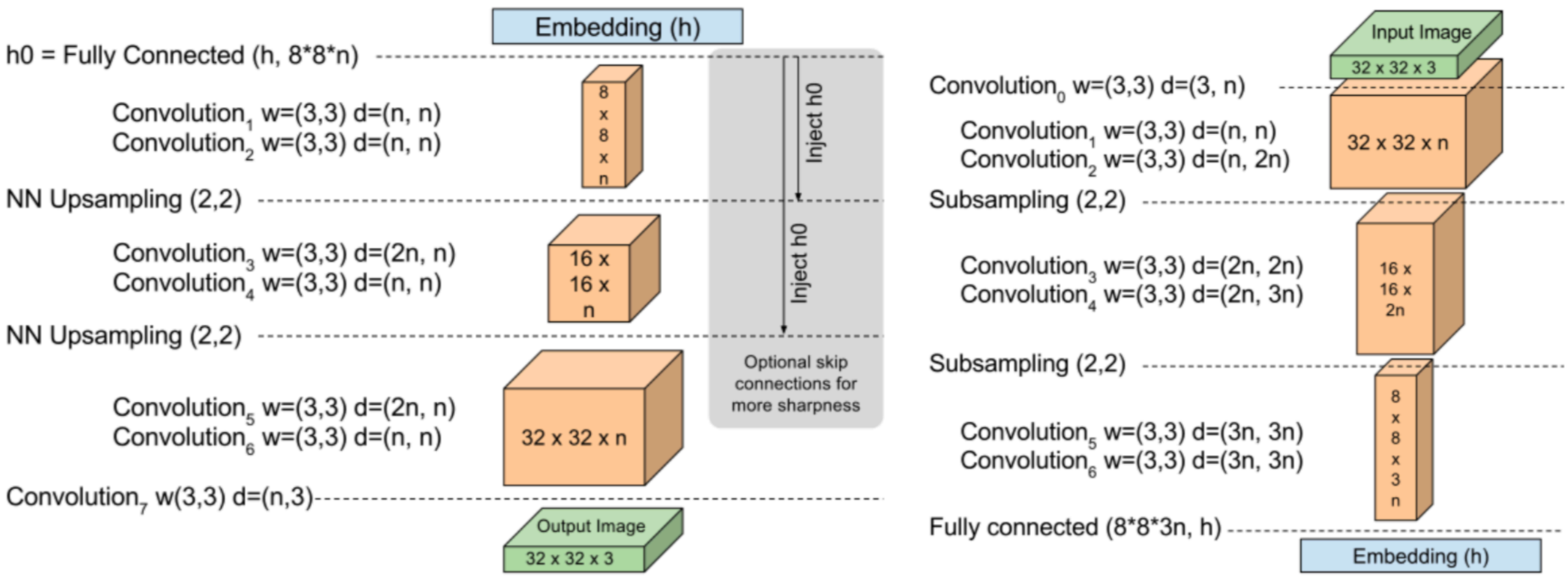
Convergence Measure

- A problem appears in GANs is to determine whether model has reached convergence, “visual inspection”

- Here a convergence measurement is given as

$$\mathcal{M}_{global} = \mathcal{L}(x) + |\gamma\mathcal{L}(x) - \mathcal{L}(G(z_G))|$$

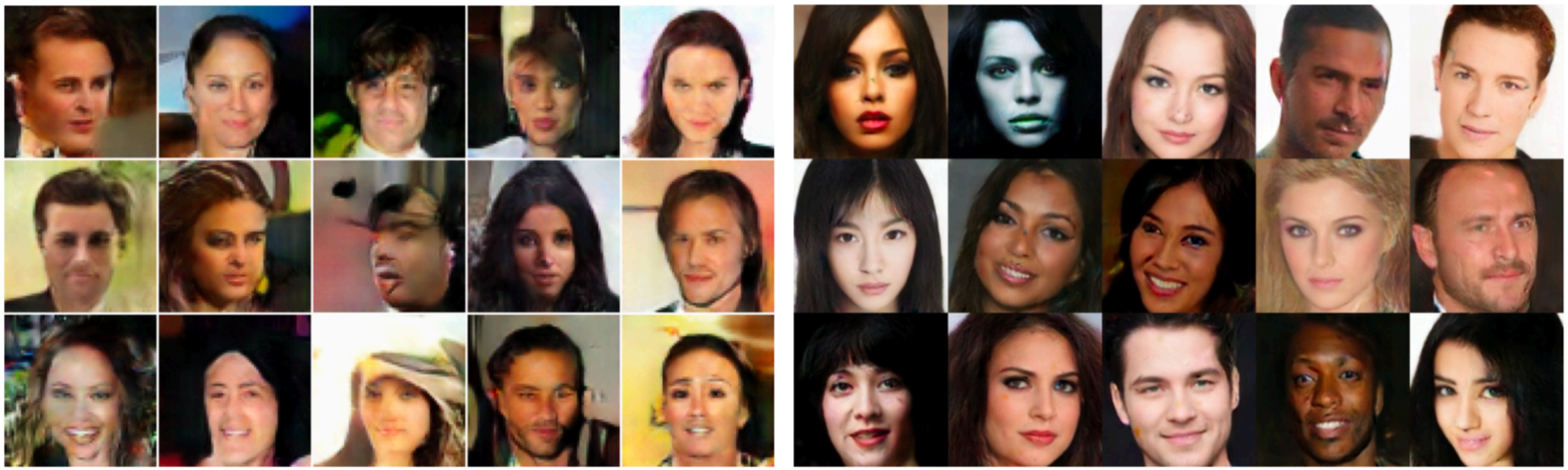
Architecture



(a) Generator/Decoder (b) Encoder

- No Batch-Norm, No dropout, etc

Experiments



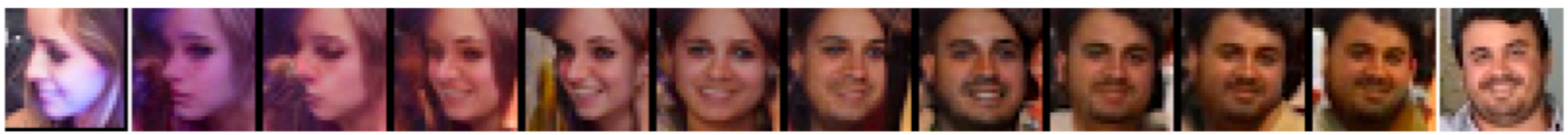
(a) EBGAN (64x64) (b) Our results (128x128)

Figure 2: Random samples comparison

Figure 3: Random 64x64 samples at varying $\gamma \in \{0.3, 0.5, 0.7\}$

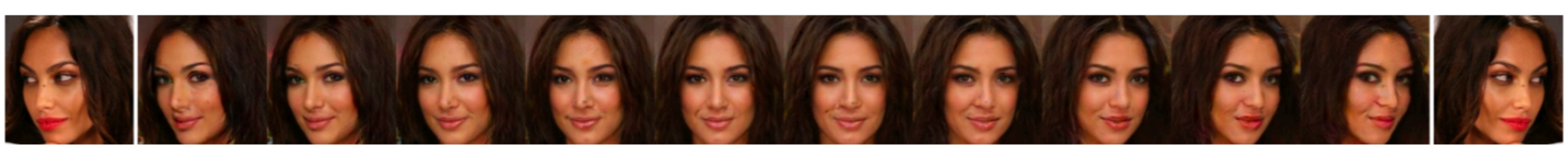
Experiments

(a) ALI [5] (64x64)



(b) Conditional PixelCNN [13] (32x32)

(c) Our results (128x128 with 128 filters)



(d) Mirror interpolations (our results 128x128 with 128 filters)

Figure 4: Interpolations of real images in latent space

Experiments

Method (unsupervised)	Score
Real data	11.24
DFM [19]	7.72
BEGAN (ours)	**5.62**
ALI [5]	5.34
Improved GANs [16]	4.36
MIX + WGAN [2]	4.04

Table 1: Inception scores (higher is better)

In the End

- Pros:
  - AutoEncoder as Discriminator
  - Distribution Assumption upon loss instead of data
  - Equilibrium and Convergence measurement
- Cons:
  - No clear relation with W-distance
  - Mode collapse in face images
  - About the IID assumption on every pixel

In the End

- A metric evaluating the diversity and quality of data