GANs Powered by Autoencoding — A Theoretic Reasoning

Zhifei Zhang, Yang Song, and Hairong Qi

THE UNIVERSITY OF TENNESSEE KNOXVILLE

ICML @ Sydney
Thirty-fourth International Conference on Machine Learning

Goal:

Many recent works integrate GANs with autoencoding (AE) networks to generate photo-realistic images conditioned on input images. It has been observed empirically that the addition of the AE network, besides generating images under certain condition, helps alleviate the mode missing and instability problems during training. **The main objective of this work is not to propose new solutions, but rather to provide theoretical reasoning on why the AE-based GAN structure is able to remedy the mode missing and instability issues.** We further show that by adding an adaptive decay variable to the adversarial error, the instability issue caused by competition between the generator and discriminator is largely alleviated.

Mode Missing:

The mode missing problem is mainly caused by the insufficient punishment on the condition of

$$p_g(x) < p_x(x).$$

- $p_g(x)$ --- the probability that a sample x is generated by the generator.
- $p_x(x)$ --- the probability that a sample x appears in the real data.

It indicates that a real sample is with lower probability to be generated --- mode missing. To solve this problem, an extra penalty could be added to emphasize the cost on mode missing.

Instability:

- **Gradient vanishing** --- the gradient back propagated to the generator is 0 if the discriminator is perfect.
- **Gradient explosion** --- the gradient back propagated to the generator may oscillate divergently when the generator is trained to approach the optimum.

Reasoning:

To avoid the model missing, we can update the generator to

$$\textbf{maximize } \mathbf{E}_{x\sim p_x}[\log p_g(x)],$$

which fits p_g to p_x without mode missing. However, we only have observations drawn from p_x and p_g instead of analytical expression, so direct comparison between p_x and p_g is intractable.

Intuitively, we can compare two arbitrary unknown distributions by the Monte Carlo method.

Hungarian: Assume two large but finite sample sets $\mathbb{X}_x$ and $\mathbb{X}_g$, of the same size, n, randomly drawn from p_x and p_g, respectively. Suppose the Hungarian assignment function is $\mathcal{H} : \mathbb{X}_x \to \mathbb{X}_g$ based on a distance metric $\mathcal{L}(x, \mathcal{H}(x))$, $x \in \mathbb{X}_x$ and $\mathcal{H}(x) \in \mathbb{X}_g$, such that $\mathbb{E}_{x\in\mathbb{X}_x}[\mathcal{L}(x, \mathcal{H}(x))]$ is minimized. Then the distance between p_x and p_g can be measured by

$$\mathbb{E}_{x\in\mathbb{X}_x}[\mathcal{L}(x, \mathcal{H}(x))]. \quad (1)$$

Ideally, if $p_x = p_g$, Eq. 1 achieves its minimum. In practice, the training dataset could be considered as $\mathbb{X}_x$, and $\mathbb{X}_g$ consists of the generated samples.

Problems of Hungarian matching:

- High computational complexity, $O(n^3)$.
- Access all the training and generated samples at the same time, preventing the mini-batch learning.

An alternative implement by AE:

Learn a network that maps x to $\mathcal{H}(x)$, i.e.,

$$\mathcal{H}(x) = G(E(x))$$

Where G denotes the generator, and E is an encoder. The Hungarian method is implemented by an AE.

The objective of GANs incorporating AE:

$$\mathbb{E}_{x\sim p_x}[\log(D(x)(1 - D(\mathcal{H}(x)))) + \lambda\mathcal{L}(x, \mathcal{H}(x))]$$

Experimental Results:

Comparse GAN+AE to GANs without AE

Notation	Method
GAN	Vanilla GAN (Goodfellow et al., 2014)
LSGAN	Least square GAN (Mao et al., 2016)
GMMN	Minimize MMD (Li et al., 2015)
GAN+MMD	Incorporate MMD as loss function
GAN+MB	Mini-batch (Salimans et al., 2016)
GAN+UR	Unrolled GAN (Metz et al., 2016)
GAN+AE	Incorporate AE to GAN

For fair comparison, we implement each method with the same architecture of Conv and Deconv networks.

MNIST after 1 and 10 epochs:

GAN LSGAN GMMN GAN+MMD GAN+MB GAN+UR GAN+AE

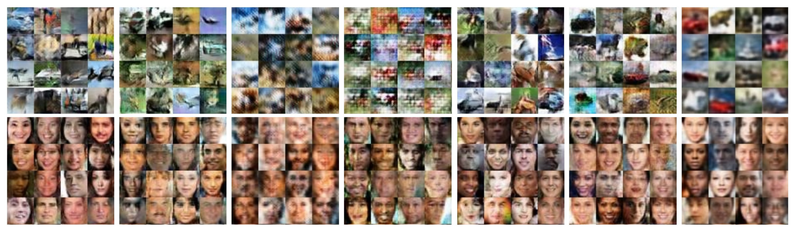
Compositional MNIST with 1000 classes:

GAN LSGAN GMMN GAN+MMD GAN+MB GAN+UR GAN+AE

Epoch	GAN	LSGAN	GMMN	GAN+MMD	GAN+MB	GAN+UR	GAN+AE
20	154	142	409	301	516	136	**52**
50	82	69	408	76	79	41	**39**

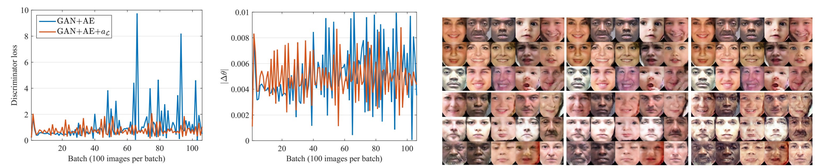
Natural images and face images:

GAN LSGAN GMMN GAN+MB GAN+UR GAN+AE AE



Instability Evaluation by adding the decay variable:

$$\mathbb{E}_{x\sim p_x}[\log(D(x)(1 - \alpha_{\mathcal{L}}D(\mathcal{H}(x)))) + \lambda\mathcal{L}(x, \mathcal{H}(x))]$$



Left: comparison on discriminator loss. Middle: absolute gradient from the discriminator. Right: generated images after 10, 50, and 100 batches from GAN+AE (top) and AE+GAN+αL (bottom).