BourGAN: Generative Networks with Metric Embeddings

Chang Xiao Peilin Zhong Changxi Zheng

Columbia University

Generative Adversarial Networks (GANs)

- Training data is a set of data items $X = \{x_i\}_{i=1}^n$ where each x_i is drawn from an unknown data distribution $\mathcal{X}$.
- A generative network G learns to map from a known latent space distribution $\mathcal{Z}$ (typically a standard Gaussian distribution) to $\mathcal{X}$.
- Training is accomplished by introducing an additional network D whose goal is to distinguish from the generated samples to the real data items drawn from $\mathcal{X}$.
- The generator G is trained by solving a minmax optimization with the following objective.

$$L_{\text{gan}}(G, D) = \mathop{\mathbb{E}}_{x\sim\mathcal{X}}[\log D(x)] + \mathop{\mathbb{E}}_{z\sim\mathcal{Z}}[\log(1 - D(G(z)))] .$$

This objective is minimized over G and maximized over D.

- The objective is optimized when the distribution of $G(z)$ is $\mathcal{X}$.

Mode Collapse

- *Mode collapse* is one of the most prominent issues in optimizing GANs.
- It is a phenomenon that a GAN can only capture a few modes of $\mathcal{X}$.
- The generated samples lack the diversity as shown in the real dataset.

Modes in Metric Space

We consider the *geometric* interpretation of modes:

- The modes of a data distribution should be viewed under a specific distance metric of data items.
- Different metrics may lead to different partitions of modes.
- We address the problem of mode collapse in a general metric space.

Logarithmic Pairwise Distance Distribution (LPDD). We propose to use the pairwise distance distribution of data items to reflect the mode structure in a dataset.

- Consider a metric space $(\mathbb{M}, \mathrm{d})$, and a distribution $\mathcal{X}$ over $\mathbb{M}$.
- Two independent samples x, y are drawn from $\mathcal{X}$, and $\eta = \log(\mathrm{d}(x, y))$.
- We call the distribution of η conditioned on $x \neq y$ the logarithmic pairwise distance distribution (LPDD) of $\mathcal{X}$.

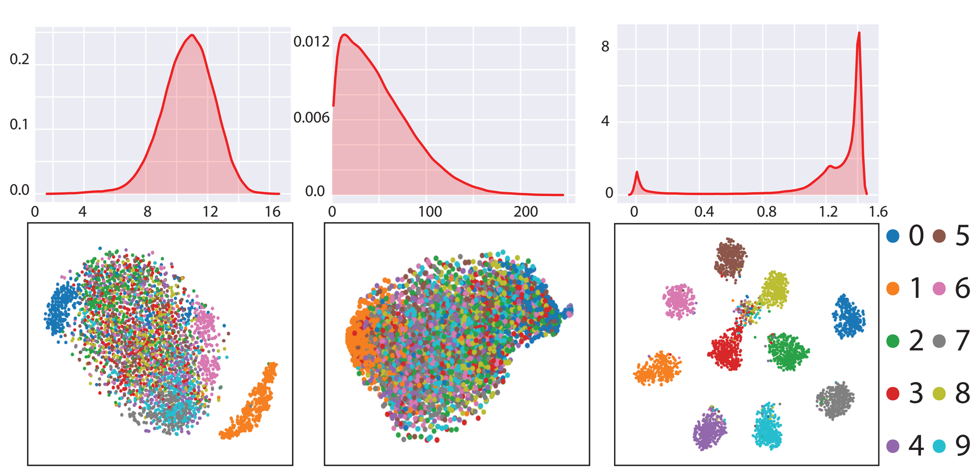

Figure 1: **(Top)** Pairwise distance distribution on MNIST dataset under different distance metrics. Left: ℓ_2 distance, Middle: Earth Mover's distance (EMD), Right: classifier distance. **(Bottom)** t-SNE visualization of data items.

Then, ideally, we want to solve:

$$\min_G \max_D L_{\text{gan}}(G, D)$$
$$\text{s.t. LPDD of } G(z) \text{ is as the same as LPDD of } \mathcal{X}.$$

Latent Space Distribution

We question the commonly used multivariate Gaussian that generates random vectors for the generator network. In the presence of separated modes, drawing random vectors from a single Gaussian may lead to arbitrarily large gradients of the generator, and a better choice is by using a *mixture of Gaussians*.

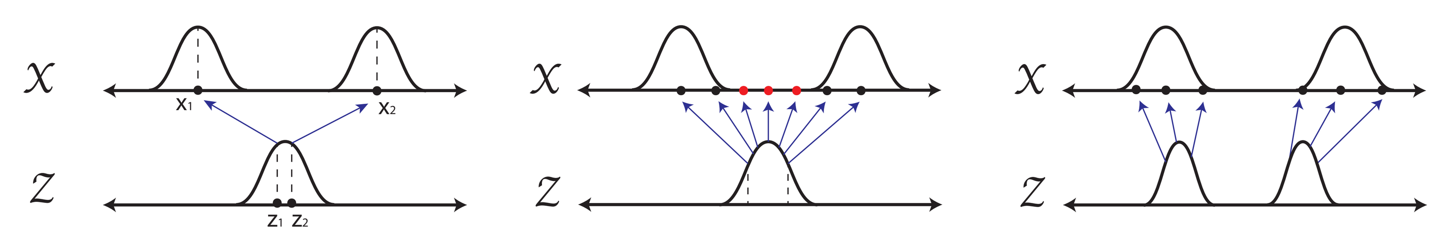

Figure 2: **Multi-mode challenge.**

BourGAN

Theorem.(Bourgain's theorem) *Consider a finite metric space* (Y, d) *with* $m = |Y|$. *There exists a mapping* $g : Y \to \mathbb{R}^k$ *for some* $k = O(\log^2 m)$ *such that* $\forall y, y' \in Y, \mathrm{d}(y, y') \le \|g(y) - g(y')\|_2 \le \alpha \cdot \mathrm{d}(y, y')$, *where* α *is a constant satisfying* $\alpha \le O(\log m)$.

Our training algorithm contains two stages:

- Using Bourgain's theorem to construct a latent space distribution $\mathcal{Z}$ (mixture of Gaussians).
- Training generative network G with additional distance loss.

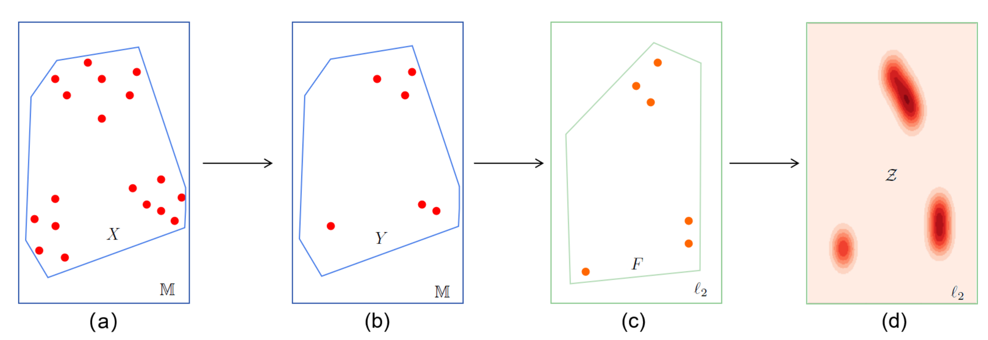

Figure 3: **Latent space distribution construction.** (a) $\mathcal{X}$ is an unknown data distribution over metric space $(\mathbb{M}, \mathrm{d})$. Each training data item in X is drawn independently from $\mathcal{X}$. (b) We sub-sample a small number of data items Y uniformly at random from X. (c) By Bourgain's theorem, we can embed set Y (from metric space $\mathbb{M}$) to a set F in ℓ_2 space such that the pairwise distance is well preserved. (d) We regard each point in F as a center of a Gaussian distribution, and construct latent space distribution $\mathcal{Z}$ as the mixture of these Gaussian distributions.

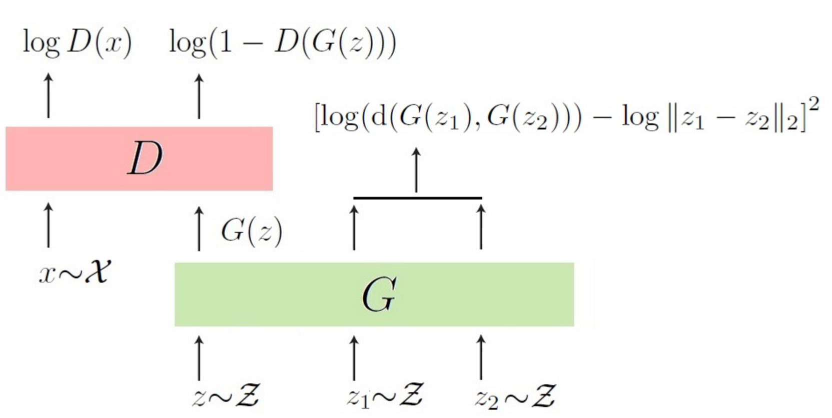

Figure 4: **Training generative networks with distance loss.**

We define distance loss to be

$$L_{\text{dist}}(G) = \mathbb{E}_{z_i, z_j\sim\mathcal{Z}}\left[\left(\log(d(G(z_i), G(z_j))) - \log(\|z_i - z_j\|_2)\right)^2\right].$$

Our new objective is $L(G, D) = L_{\text{gan}}(G, D) + \beta \cdot L_{\text{dist}}(G)$. We still try to minimize it over G and maximize it over D.

Experiments

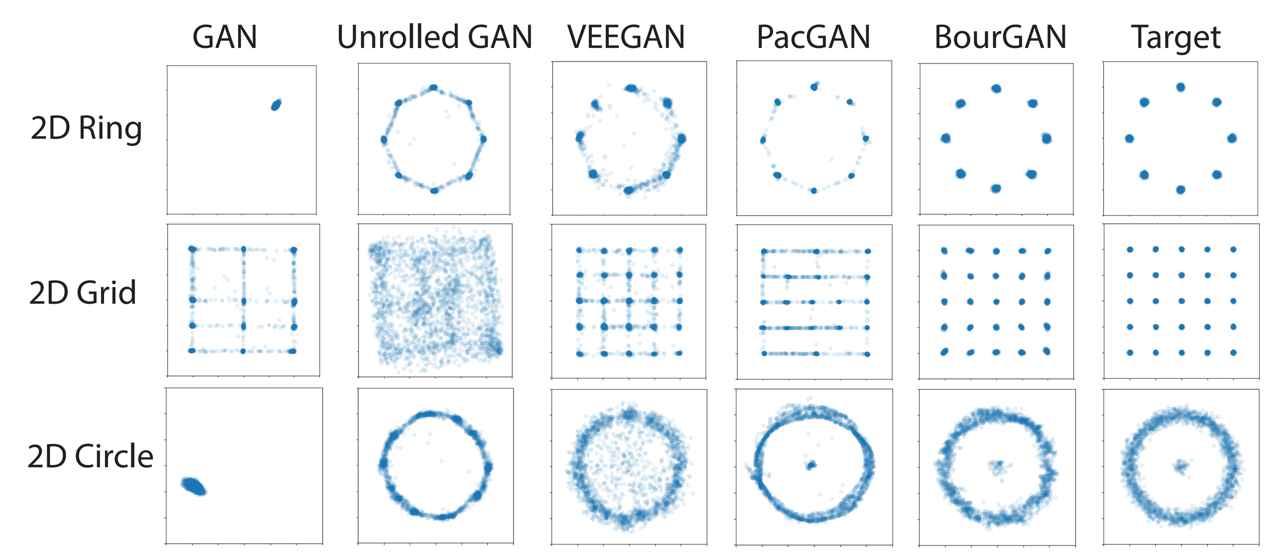

Figure 5: **Synthetic data tests**. In all three tests, our method clearly captures all the modes presented in the targets, while producing *no* unwanted samples located between the regions of modes.

	2D Ring		2D Grid		2D Circle	
	#modes (max 8)	low quality	#modes (max 25)	low quality	center captured	low quality
GAN	1.0	0.06%	1.617	17.70%	No	0.14%
Unrolled	7.6	12.03%	14.9	95.11%	No	0.50%
VEEGAN	8.0	13.23%	24.4	22.84%	Yes	10.72%
PacGAN	7.8	1.79%	24.3	20.54%	Yes	1.38%
BourGAN	8.0	0.12%	25.0	4.09%	Yes	0.35%

Table 1: Statistics of Experiments on Synthetic Datasets

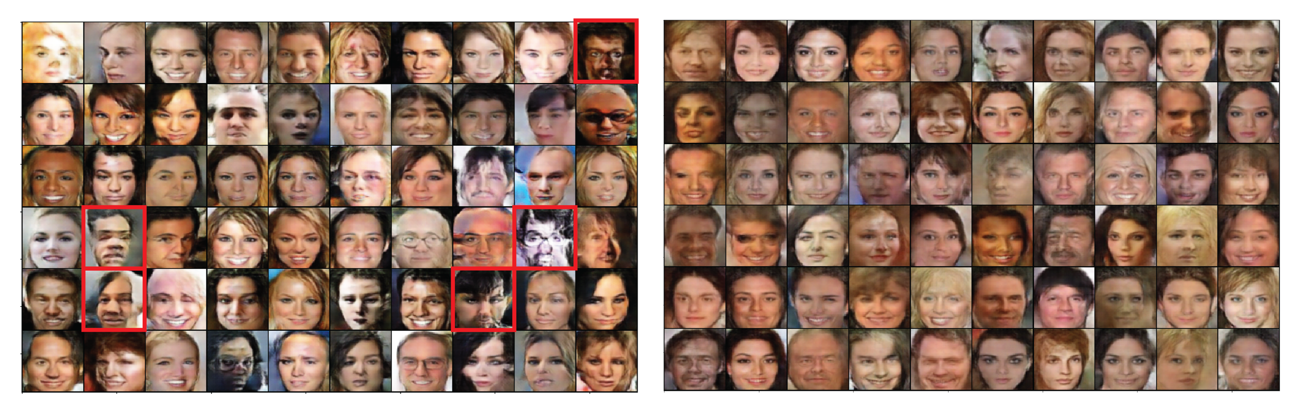

Figure 6: **Qualitative results** on CelebA dataset using DCGAN (Left) and BourGAN (Right) under ℓ_2 metric. It appears that DCGAN generates some samples that are visually more implausible (e.g., red boxes) in comparison to BourGAN. Results are fairly sampled at random, not cherry-picked.

Acknowledgements

We thank Daniel Hsu, Carl Vondrick and Henrique Maia for the helpful feedback. Chang Xiao and Changxi Zheng are supported in part by the National Science Foundation (CAREER-1453101, 1717178 and 1816041) and generous donations from SoftBank and Adobe. Peilin Zhong is supported in part by National Science Foundation (CCF-1703925, CCF-1421161, CCF-1714818, CCF-1617955 and CCF-1740833), Simons Foundation (#491119 to Alexandr Andoni) and Google Research Award.