

Self-Attention Generative Adversarial Networks

Seho Kim

Jan 18, 2020

- Image synthesis
 - Remarkable progress: Emergence of GANs
 - GANs based on deep convolutional networks
 - SOTA ImageNet GAN Miyato and Koyama, 2018
 - Self-attention
 - Self-Attention Generative Adversarial Networks(SAGANs)
 - Enforcing good conditioning of GAN generators using the spectral normalization technique
 - Inception Score (IS) from 36.8 to **52.52**
and reducing Frechet Inception Distance (FID) from 27.62 to **18.65**

- Generative Adversarial Networks
 - Great success in various image generation tasks
; image-to-image translation, image super-resolution and text-to-image synthesis
 - Known to be unstable and sensitive to the choices of hyperparameters
 - Attempt to stabilize the GAN training dynamics and improve the sample diversity
 - Limiting spectral norm of the weight matrices in the discriminator
- Attention Models
 - Capture global dependencies
 - Self-attention as a non-local operation

Background

- Spectral Normalization
 - Controls the Lipschitz constant by literally constraining the spectral norm of each layer

Spectral Norm of Matrix A (L2 Matrix Norm of A)

$$\sigma(A) := \max_{\mathbf{h}: \mathbf{h} \neq 0} \frac{\|\mathbf{A}\mathbf{h}\|_2}{\|\mathbf{h}\|_2} = \max_{\|\mathbf{h}\|_2 \leq 1} \|\mathbf{A}\mathbf{h}\|_2 \quad (1)$$

- Equivalent to the largest singular value of A

$$\|g\|_{Lip} = \sup_{\mathbf{h}} \sigma(\nabla g(\mathbf{h})) = \sup_{\mathbf{h}} \sigma(W) = \sigma(W), \text{ where, } g(\mathbf{h}) = W\mathbf{h} \quad (2)$$

$$\begin{aligned} \|f\|_{Lip} : \|f\|_{Lip} &\leq \|\mathbf{h}_L \mapsto W^{L+1}\mathbf{h}_L\|_{Lip} \cdot \|a_L\|_{Lip} \cdot \|\mathbf{h}_{L-1} \mapsto W^L\mathbf{h}_{L-1}\|_{Lip} \\ &\cdots \|a_1\|_{Lip} \cdot \|\mathbf{h}_0 \mapsto W^1\mathbf{h}_0\|_{Lip} = \prod_{l=1}^{L+1} \|\mathbf{h}_{l-1} \mapsto W^l\mathbf{h}_{l-1}\|_{Lip} = \prod_{l=1}^{L+1} \sigma(W^l) \end{aligned} \quad (2.1)$$

- Spectral Normalization
 - Normalize the spectral norm of the weight matrix W so that it satisfied the Lipschitz constraint $\sigma(W) = 1$

$$\bar{W}_{SN}(W) := W/\sigma(W) \quad (3)$$

$$\sigma(\bar{W}_{SN}(W)) = 1 \quad (3.1)$$

- The Hinge Version of the Adversarial Loss

$$\begin{aligned} V_D(\hat{G}, D) = & \mathbb{E}_{\mathbf{x} \sim q_{data}(\mathbf{x})} [\min(0, -1 + D(\mathbf{x}))] \\ & + \mathbb{E}_{\mathbf{z} \sim p(\mathbf{z})} [\min(0, -1 - D(\hat{G}(\mathbf{z})))] \end{aligned}$$

$$V_G(G, \hat{D}) = -\mathbb{E}_{\mathbf{z} \sim p(\mathbf{z})} [\hat{D}(G(\mathbf{z}))] \quad (4)$$

Self-Attention Module

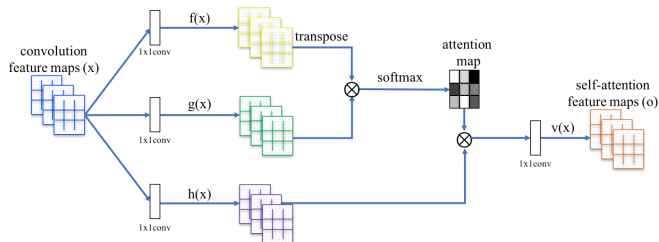
$$\beta_{j,i} = \frac{\exp(s_{ij})}{\sum_{i=1}^N \exp(s_{ij})}, \text{ where } s_{ij} = \mathbf{f}(\mathbf{x}_i)^T \mathbf{g}(\mathbf{x}_j) \quad (5)$$

$$\mathbf{o} = (\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_j, \dots, \mathbf{o}_N) \in \mathbb{R}_{C \times N},$$

$$\text{where, } \mathbf{o}_j = \mathbf{v} \left(\sum_{i=1}^N \beta_{j,i} \mathbf{h}(\mathbf{x}_i) \right), \quad \mathbf{h}(\mathbf{x}_i) = \mathbf{W}_h \mathbf{x}_i, \quad \mathbf{v}(\mathbf{x}_i) = \mathbf{W}_v \mathbf{x}_i \quad (6)$$

$$\mathbf{y}_i = \gamma \mathbf{o}_i + \mathbf{x}_i \quad (7)$$

Self-Attention Generative Adversarial Networks



Hinge version of the adversarial loss

$$L_D = -\mathbb{E}_{(x,y) \sim p_{data}} [\min(0, -1 + D(x, y))] \\ - \mathbb{E}_{z \sim p_z, y \sim p_{data}} [\min(0, -1 - D(G(z), y))]$$

$$L_G = -\mathbb{E}_{z \sim p_z, y \sim p_{data}} D(G(z), y), \quad (8)$$

Techniques to Stabilize the Training of GANs

- Spectral normalization in the generator as well as in the discriminator
- The two-timescale update rule (TTUR)
 - Spectral normalization (SN) for both generator and discriminator
 - Restricting the spectral norm of each layer
 - Does not require extra hyper-parameter tuning
 - Computational cost is relatively small
 - Imbalanced learning rate for generator and discriminator updates
 - Regularization of the discriminator often slows down the GAN's learning process
 - Using separate learning rates (TTUR)
 - Produce better results given the same wall-clock time

- Evaluation metrics
 - Inception Score (IS) and Frechet Inception Distance (FID)
- Network structures and implementation details
 - Evaluating the proposed stabilization techniques
 - Self-attention mechanism
 - Comparison with the state-of-the-art

- Self-Attention Generative Adversarial Networks (SAGANs)
 - The self-attention module
 - Spectral normalization and TTUR