Quality Aware Generative Adversarial Networks

NeurIPS 2019

Parimala Kancharla, Sumohana S. Channappayya

Department of Electrical Engineering, IIT Hyderabad

Outline

- Background
  - Generative Adversarial Network
  - Wasserstein Generative Adversarial Network (WGAN).
  - WGAN-GP
  - Banach WGAN
- Image Quality Assessment
  - Structural Similarity Index (SSIM).
  - Natural Image Quality Evaluator (NIQE)
- Proposed QAGANs
  - QAGAN based on SSIM
  - QAGAN based on NIQE
- References

Background

Generative Adversarial Network [1]

- Suppose the data points (x) are coming from some distribution $p_{data}(x)$ and z is the random noise vector drawn from standard Normal distribution $p_z(z)$.
- $G(z;\theta_g)$: Generator with parameters θ_g.
- $D(x;\theta_d)$: Discriminator with parameters θ_d.

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets ", NIPS-2014

Background

Generative Adversarial Network [1]

- Suppose the data points (x) are coming from some distribution $p_{data}(x)$ and z is the random noise vector drawn from standard Normal distribution $p_z(z)$.
- $G(z; \theta_g)$: Generator with parameters θ_g.
- $D(x; \theta_d)$: Discriminator with parameters θ_d.

Objective function

- D and G play the following two player min-max game with value function $V(G; D)$:

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)}[log(D(x))] + E_{z \sim p_z(z)}[log(1 - D(G(z)))]$$

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets ", NIPS-2014

Background

Wasserstein Generative Adversarial Networks [Arjovsky et al., 2017]

The primal form: $W(p_r, p_\theta) = \inf_{\gamma \in \Gamma(p_r, p_\theta)} E_{(x,y)\sim\gamma}[||x - y||_2]$

$$W(p_r, p_\theta) = \max_{w \in W} E_{x\sim p_r}[D_w(x)] - E_{z\sim p_z}[D_w(g_\theta(z)],\ w \in [-c, c].$$

Wasserstein Generative Adversarial Networks [Arjovsky et al., 2017]

$$\text{The primal form: } W(p_r, p_\theta) = \inf_{\gamma \in \Gamma(p_r, p_\theta)} E_{(x,y)\sim\gamma}[||x - y||_2]$$

$$W(p_r, p_\theta) = \max_{w \in W} E_{x\sim p_r}[D_w(x)] - E_{z\sim p_z}[D_w(g_\theta(z)],\ w \in [-c, c].$$

WGAN-GP [Gulrajani et al., 2017]

$$|D_w(x) - D_w(y)| \leq ||x - y||_2 \implies ||\nabla D_w(x)||_2 \leq 1 \quad \forall x \in R^n.$$

$$L = E_{x\sim p_r}[D_w(x)] - E_{z\sim p_z}[D_w(g_\theta(z)] + \lambda E_{\hat{x}}(||\nabla D_w(\hat{x})||_2 - 1)^2$$

Background

Wasserstein Generative Adversarial Networks [Arjovsky et al., 2017]

The primal form: $W(p_r, p_\theta) = \inf_{\gamma \in \Gamma(p_r, p_\theta)} E_{(x,y)\sim\gamma}[||x - y||_2]$

$$W(p_r, p_\theta) = \max_{w \in W} E_{x\sim p_r}[D_w(x)] - E_{z\sim p_z}[D_w(g_\theta(z)],\; w \in [-c, c].$$

WGAN-GP [Gulrajani et al., 2017]

$$|D_w(x) - D_w(y)| \leq ||x - y||_2 \implies ||\nabla D_w(x)||_2 \leq 1 \quad \forall x \in \; R^n.$$

$$L = E_{x\sim p_r}[D_w(x)] - E_{z\sim p_z}[D_w(g_\theta(z)] + \lambda E_{\hat{x}}(||\nabla D_w(\hat{x})||_2 - 1)^2$$

Banach WGAN [Adler and Lunz, 2018]

Generalized the theory of WGAN-GP to Banach spaces.

$$L = \lambda E_{x\sim p_r}[D_w(x)] - E_{z\sim p_z}[D_w(g_\theta(z)] + \lambda E_{\hat{x}}\left(\frac{1}{\gamma}||\nabla D_w(\hat{x})||_{B^*} - 1)\right)^2$$

Challenges

Challenges in Generative Adversarial Networks:

- **Low visual quality.**
- Mode collapse.
- Generalization.
- Disentanglement.
- Stability in training.

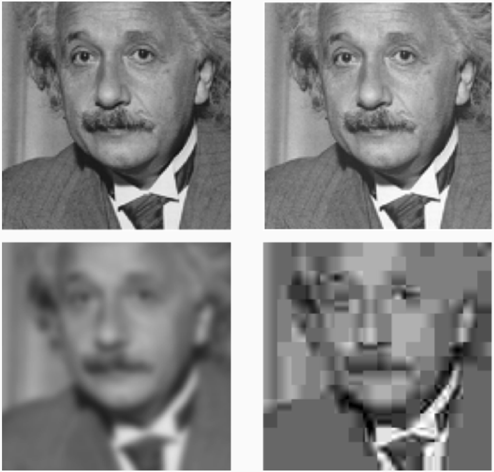
Major contributions

- We make explicit use of objective image quality assessment (IQA) metrics and their variants for regularizing WGAN with gradient penalty (WGAN-GP), and propose Quality Aware GANs (QAGANs).
- Specifically we have used a full reference image quality assessment algorithm (SSIM) and natural image quality evaluator (NIQE).

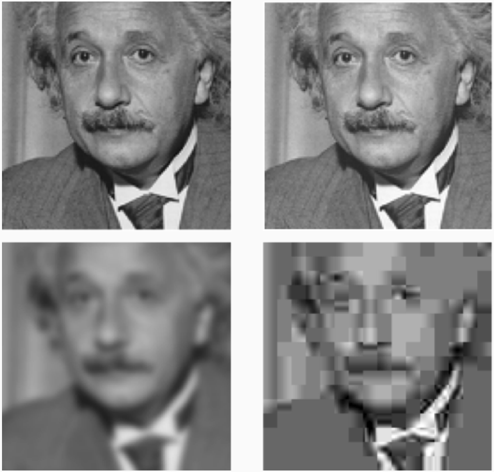
$$(WGAN - GP) + \binom{SSIM'}{NIQE} \Longrightarrow \binom{QAGAN - SSIM}{QAGAN - NIQE}$$

- We demonstrated state-of-the-art performance on CIFAR-10, STL10 and CelebA datasets for non-progressive GANs.

What is Quality Assessment?



Why Quality Assessment?



All distorted images have same mean squared error (MSE)! L^p norms fail! [2]

[2] Z. Wang, and A. C. Bovik. "Mean squared error: love it or leave it? A new look at signal fidelity measures."IEEE Signal Processing Magazine 26.1 (2009): 98-117.

The Structural Similarity Index (SSIM) [3]

$$\text{SSIM}(P_{(i,j)}, T_{(i,j)}) = L(P_{(i,j)}, T_{(i,j)}).C(P_{(i,j)}, T_{(i,j)}).S(P_{(i,j)}, T_{(i,j)}),$$

where P, T refer to the pristine and test image respectively. $L(P_{(i,j)}, T_{(i,j)})$, $C(P_{(i,j)}, T_{(i,j)})$, $S(P_{(i,j)}, T_{(i,j)})$ are the local luminance, contrast and structure scores at pixel (i, j) respectively. Further,

$$L(P_{(i,j)}, T_{(i,j)}) = \frac{2\mu_P(i,j)\mu_T(i,j) + C_1}{\mu_P^2(i,j) + \mu_T^2(,j) + C_1}$$

$$C(P_{(i,j)}, T_{(i,j)}) = \frac{2\sigma_P(i,j)\sigma_T(i,j) + C_2}{{\sigma_P}^2(i,j) + {\sigma_T}^2(i,j) + C_2}$$

$$S(P_{(i,j)}, T_{(i,j)}) = \frac{\sigma_{PT}(i,j) + C_3}{\sigma_P(i,j)\sigma_T(i,j) + C_3}$$

[3] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing , 13(4):600âĂŞ612

Metric form of SSIM [4]

The constants C_1, C_2, C_3 are used to avoid division-by-zero issues. For simplicity, $C_3 = C_2/2$ in the standard implementation which leads to

$$\text{SSIM}(P_{(i,j)}, T_{(i,j)}) = L(P_{(i,j)}, T_{(i,j)}).CS(P_{(i,j)}, T_{(i,j)}),$$

where

$$CS(P_{(i,j)}, T_{(i,j)}) = \frac{2\sigma_{PT}(i,j) + C_2}{\sigma_P{}^2(i,j) + \sigma_T{}^2(i,j) + C_2}.$$

The valid distance metric can be derived from the components of the SSIM index.

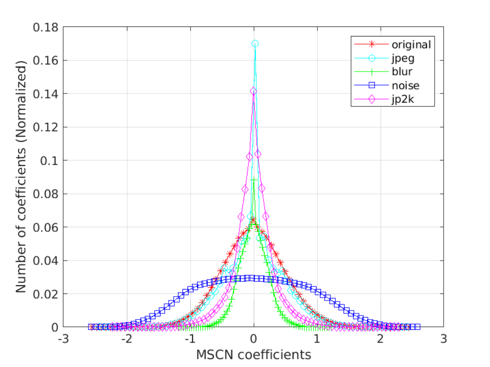
$$d^Q(P_{(i,j)}, T_{(i,j)}) = \sqrt{2 - L(P_{(i,j)}, T_{(i,j)}) - CS(P_{(i,j)}, T_{(i,j)})}$$

[4] Brunet, Dominique, Edward R. Vrscay, and Zhou Wang. "On the mathematical properties of the structural similarity index."IEEE Transactions on Image Processing 21.4 (2011): 1488-1499

The statistics of natural images [5]

- Natural images are distinctive because they contain unique statistical signature.
  Mean subtracted contrast normalization coefficients (MSCN)

$$\hat{I}(i,j) = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j)}$$



Natural scenes

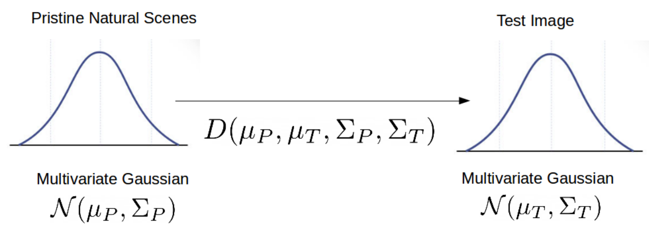
Natural Image Quality Evaluator (NIQE) [6]

- NIQE builds a statistical model for MSCN coefficients of the class of pristine natural scenes and uses the model′s parameters for quality estimation.

[6] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. âĂIJMaking a âĂIJcompletely blindâĂİ image quality analyzerâĂİ. In: IEEE Signal Processing Letters 20.3 (2013), pp. 209âĂŞ212

Natural Image Quality Evaluator (NIQE) [6]

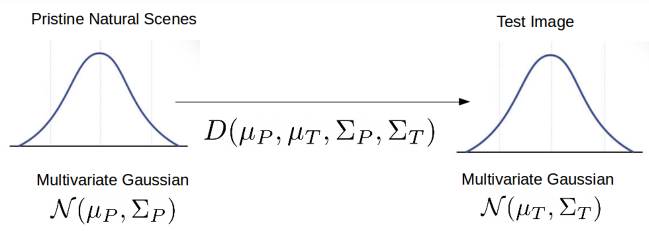
- NIQE builds a statistical model for MSCN coefficients of the class of pristine natural scenes and uses the model's parameters for quality estimation.



[6] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. âĂIJMaking a âĂIJcompletely blindâĂİ image quality analyzerâĂİ. In: IEEE Signal Processing Letters 20.3 (2013), pp. 209âĂŞ212

Natural Image Quality Evaluator (NIQE) [6]

- NIQE builds a statistical model for MSCN coefficients of the class of pristine natural scenes and uses the model′s parameters for quality estimation.



Natural Image Quality Evaluator (NIQE)

$$D(\mu_P, \mu_T, \Sigma_P, \Sigma_T) = \sqrt{(\mu_P - \mu_T)^T \left(\frac{\Sigma_P + \Sigma_T}{2}\right)^{-1} (\mu_P - \mu_T)}.$$

[6] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. âĂIJMaking a âĂIJcompletely blindâĂİ image quality analyzerâĂİ. In: IEEE Signal Processing Letters 20.3 (2013), pp. 209âĂŞ212

QAGAN-SSIM

- We propose to use the perceptually correlated SSIM derived metric as the underlying image metric in the wasserstein distance.
- This implies that the discriminator should be lipschitz with respect to SSIM metric ($d^Q(x, y)$).
- As this is a bounded metric between $0 \leq d^Q(x, y) \leq \sqrt{2}$.

$$|D(X) - D(Y)| \leq d^Q(X, Y)$$

- SSIM gradient penalty term makes the discriminator quality aware.

QAGAN-SSIM

$$\text{SSIM GP} = E_{X\sim P_r, Y\sim P_G}\left[\left(\frac{|D(X)-D(Y)|}{d^Q(X,Y)}\right)-1\right]^2.$$

$$L_d = \min_{D\epsilon\mathcal{D}}\left(E_{z\sim P_z}D(G(z)) - E_{X\sim P_r}D(X)\right) + \lambda_1 E_{\hat{x}\sim P_{\hat{x}}}(||\nabla_{\hat{x}}D(\hat{x})||_2 - 1)^2 +$$
$$\lambda_2 E_{X\sim P_r, G(z)\sim P_G}\left[\left(\frac{|D(X)-D(G(z))|}{d^Q(X,G(z))}\right)-1\right]^2,$$

where λ_1 and λ_2 are empirically chosen. where $d^Q(x,y)$ is the metric SSIM, that is

$$d^Q(X,Y) = \sqrt{2 - L(X,Y) - CS(X,Y)}$$

Motivation to impose Quality aware gradient penalty

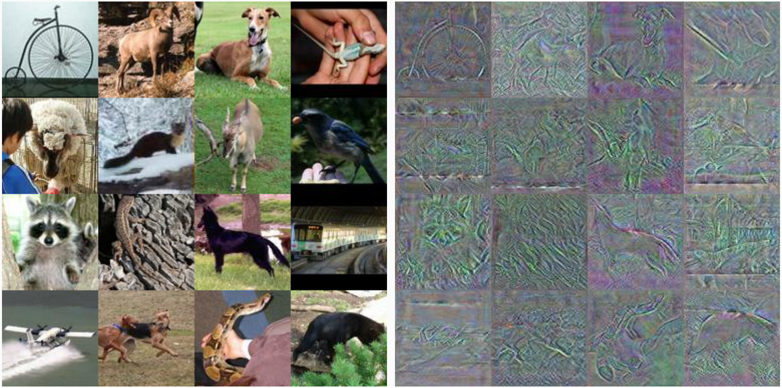
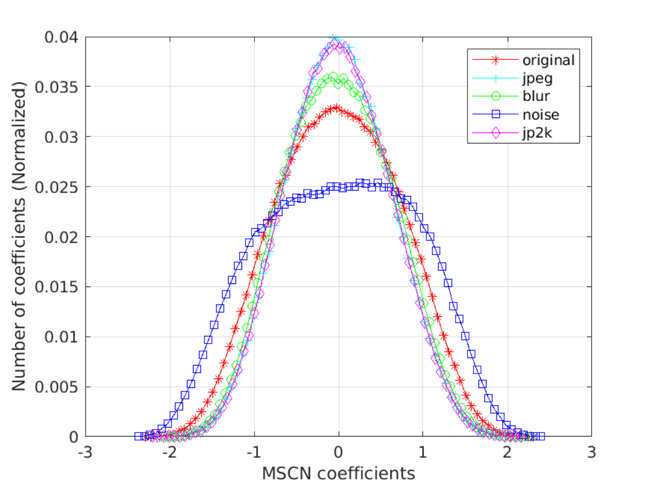


Figure: Real images and their corresponding discriminator gradient maps

Motivation to impose Quality aware gradient penalty



MSCN coefficients of the discriminator gradient maps.

Estimating the pristine model:

- The core idea of enforcing statistical signature onto the discriminator gradient map remains the same as the gradient maps are computed from the smooth function.
- The reference model parameters μ_P,Σ_P are estimated from the real discriminator gradient maps.

Quality aware gradient penalty

The naturalness of a test discriminator gradient map (T)

$$||(T|\mu_P, \Sigma_P)||_{\text{NIQE}} := \sqrt{(\mu_P - \mu_T)^T \left(\frac{\Sigma_P + \Sigma_T}{2}\right)^{-1} (\mu_P - \mu_T)},$$

where μ_T, Σ_T are the model parameters of the test image's multivariate gaussian model.

Discriminator Loss Function

$$L_d = \min_{D \epsilon \mathcal{D}} \left(E_{z \sim P_z} D(G(z)) - E_{X \sim P_r} D(X) \right) + \lambda_1 E_{\hat{x} \sim P_{\hat{x}}} (||\nabla_{\hat{x}} D(\hat{x})||_2 - 1)^2 + \lambda_2 E_{\hat{x} \sim P_{\hat{x}}} (||(\nabla_{\hat{x}} D(\hat{x})|\mu_P, \Sigma_P)||_{\text{NIQE}}),$$

where λ_1 and λ_2 are hyper parameters chosen empirically. As before, $\hat{x}$ is sampled from a line joining the real and fake image distributions.

Results - CelebA dataset

Table1: FID scores on the **CelebA** dataset (64×64).

Model	FID
Real Faces (CelebA)	1.09
WGAN-GP	12.89
Banach WGAN	10.5
MMD GAN-rep-b	6.79
QAGAN (SSIM)	**6.421**
QAGAN (NIQE)	**6.504**

Qualitative Results

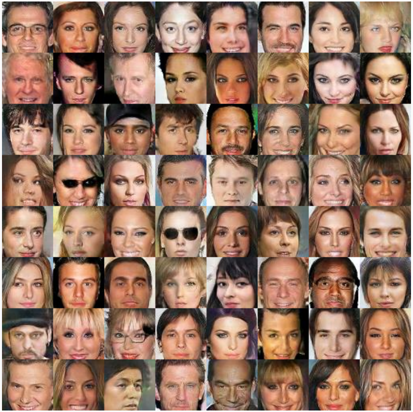


Figure: Randomly sampled images generated using QAGAN(SSIM) for CelebA dataset (64×64).

Qualitative Results

Figure: Randomly sampled images generated using QAGAN(NIQE) for CelebA dataset (64×64).

Results - STL10 dataset

Table2: IS and FID on the **STL-10** dataset (48×48).

Model	IS	FID
Real Data (48×48)	26.08 ± 0.26	7.9
WGAN-GP	9.05 ± 0.12	55.1 ± 0.0
SNGAN	9.10 ± 0.04	40.10 ± 0.50
MMD GAN-rep	9.36 ± 0.0	36.67 ± 0.0
QAGAN (SSIM)	$\mathbf{9.29 \pm 0.05}$	$\mathbf{19.77 \pm 0.0091}$
QAGAN (NIQE)	$\mathbf{9.1720 \pm 0.08}$	$\mathbf{19.45 \pm 0.0013}$

Qualitative Results

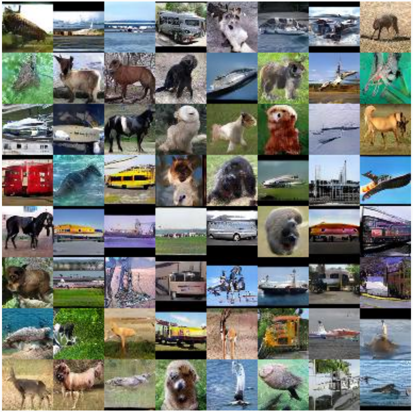


Figure: Randomly sampled images generated using QAGAN(SSIM) for STL10 dataset (64×64).

Qualitative Results

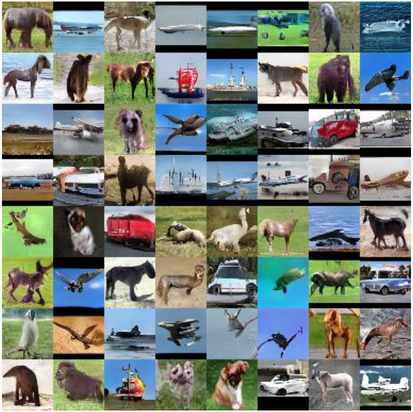


Figure: Randomly sampled images generated using QAGAN(NIQE) for STL10 dataset (64×64).

Table1: Inception Score (IS) and Fréchet Inception Distance (FID) computed from 50,000 samples of the **CIFAR-10** dataset (32×32). Scores that are unavailable are marked with a '-'.

Model	**IS**	**FID**
Real data	11.24 ± 0.12	7.80
DCGAN	6.16 ± 0.07	-
WGAN-GP	7.86 ± 0.10	40.2 ± 0.0
CTGAN	8.12 ± 0.12	-
SNGAN	8.12 ± 0.12	21.5 ± 0.21
$W^{-\frac{3}{2},2}$ - Banach WGAN	8.26 ± 0.07	-
L^{10} - Banach WGAN	8.31 ± 0.07	-
MMD GAN-rep-b	8.29 ± 0.0	$16.21 \pm 0,0$
QAGAN (SSIM)	$\mathbf{8.37 \pm 0.04}$	$\mathbf{13.91 \pm 0.105}$
QAGAN (NIQE)	$\mathbf{7.87 \pm 0.027}$	$\mathbf{12.4697 \pm 0.068}$

Qualitative Results

Figure: Randomly sampled images generated using QAGAN(SSIM) for CIFAR10 dataset (32×32).

Qualitative Results

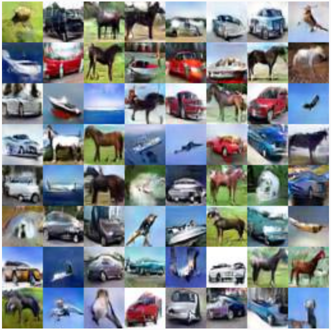


Figure: Randomly sampled images generated using QAGAN(NIQE) for CIFAR10 dataset (32×32).

Conclusions

- Based on insights from both FR and NR IQA metrics, we have proposed two novel regularization approaches for the WGAN-GP framework.
- The key takeaway from our work is that the unique local structural and statistical signature of pristine natural images must be preserved in the generated images.
- We believe that this work opens up new and exciting directions in image and video generative modeling, given the plethora of excellent QA metrics.

References

- code : https://github.com/lfovia/QAGANS
- P. Kancharla, S. S. Channappayya, Quality Aware Generative Adversarial Networks,accepted to the Thirty-third Conference on Neural Information Processing Systems (NeurIPS) 2019, Vancouver, Canada.

References I

Adler, J. and Lunz, S. (2018).
Banach wasserstein gan.
In *Advances in Neural Information Processing Systems*, pages 6754–6763.

Arjovsky, M., Chintala, S., and Bottou, L. (2017).
Wasserstein generative adversarial networks.
In *International Conference on Machine Learning*, pages 214–223.

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. C. (2017).
Improved training of wasserstein gans.
In *Advances in Neural Information Processing Systems*, pages 5767–5777.