Quality Aware Generative Adversarial Networks

Parimala Kancharla, Sumohana S. Channappayya
Indian Institute of Technology Hyderabad
Email: {ee15m17p100001,sumohana}@iith.ac.in

Objective and Approach

- Goal: To design quality aware regularizers for training Generative Adversarial Networks.
- Approach: **Using the full reference IQA algorithm (SSIM) and No reference IQA algorithm (NIQE).**

Background

- Generative Adversarial Networks (GANs) [1] are generative models designed to learn the probability distribution of data that is aided by adversarial learning.
- A GAN is composed of two models: the generator model $G(z;\theta_G)$ and the discriminator model $D(x;\theta_D)$.
- **Objective function:**

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

- **WGAN [2]**

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

Lipschitz constraint enforced by weight clipping $w\epsilon[-c,c]$.

- **WGAN-GP [3]**
- Duality :

$D_w$is 1-lipschitz function w.r.t $||.||_2 norm \implies ||\nabla D_w(x)||_2 \leq 1 \ \forall x \epsilon R^n$.

- Gradient penalty for discriminator loss to impose the lipschitz constraint.

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

- Where the distribution of $\hat{X}$ is taken to be the uniform distributions on lines connecting points drawn from $p_{g_\theta(z)}$ and p_r .

Structural Similarity Index (SSIM) [4]

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

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,

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

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

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

The valid distance metric can be derived from the components of the SSIM index by $C_3 = C_2/2$ [5].

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

This is a bounded metric : $0 \leq d^Q(x,y) \leq \sqrt{2}$.

Structural Similarity Index

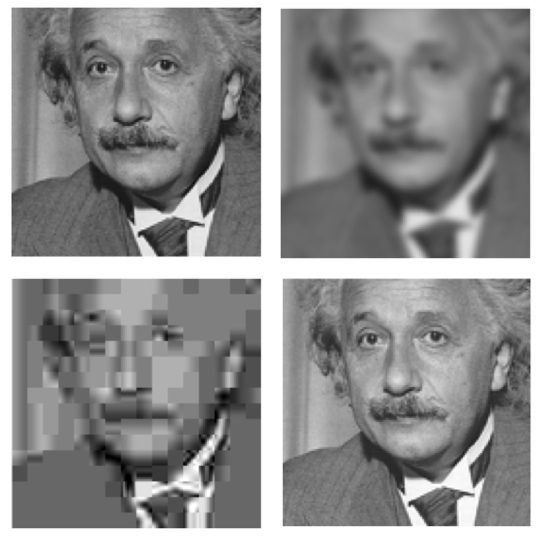

Figure: Same MSE! but different perceptual quality.

MSCN Coefficients

Figure: MSCN Coefficients

Empirical Histogram of MSCN coefficients

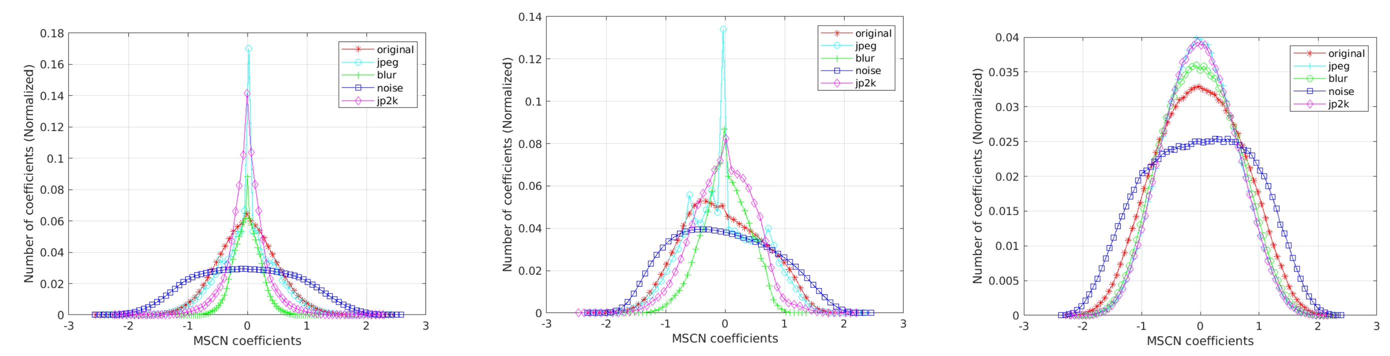

Figure: The empirical histograms of MSCN coefficients

NIQE (No reference Image quality evaluator) [7]

- NIQE builds a statistical model for the class of pristine natural scenes.
- Specifically, it models MSCN pristine image coefficients [6] using a GGD and AGGD.

$$\text{Mean subtracted contrast normalization:} \hat{I}(i,j) = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + 1}$$

- The parameters of these GGD and AGGD models are in turn modeled using a Multivariate Gaussian (MVG) distribution.
- NIQE score: the "distance" of its MVG parameters μ_T, Σ_T from the pristine MVG parameters.

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

QAGAN -SSIM

- The underlying image metric needs to be perceptually correlated.

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

The overall discriminator loss function is

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

where λ_1 and λ_2 are empirically chosen.

QAGAN -NIQE

- Motivation: To preserve the MSCN coefficients of the discriminator gradient map.
- The naturalness of a test discriminator gradient image T is computed to be its "distance" from the pristine image gradient class.

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

- 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.

Qualitative and Quantitative results

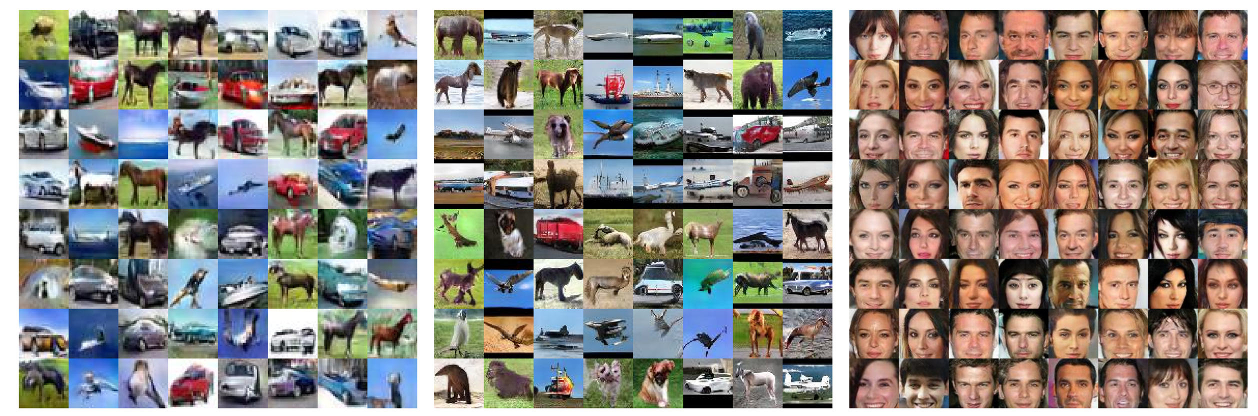

Figure: Images generated using QAGANs with quality aware distance metric regularizer (SSIM).

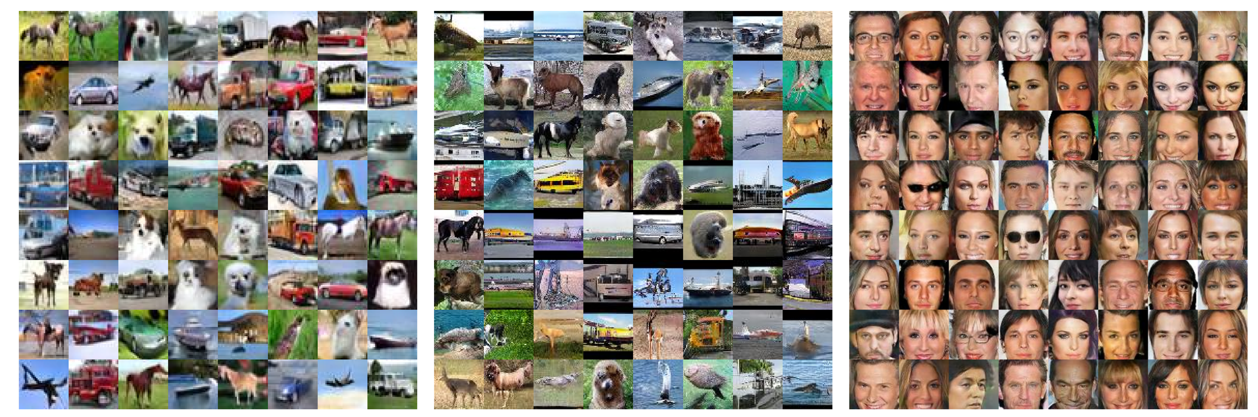

Figure: Images generated using QAGANs with quality aware distance metric regularizer (SSIM).

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**

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)	$\mathbf{7.87 \pm 0.027}$	$\mathbf{12.4697 \pm 0.068}$

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)	$\mathbf{7.87 \pm 0.027}$	$\mathbf{12.4697 \pm 0.068}$

- IS and FID scores on CelebA, STL-10 and CIFAR-10 datasets respectively.

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.

References

- https://www.iith.ac.in/∼lfovia/
- Ian Goodfellow et al. Generative adversarial nets. In: Advances in neural information processing systems.
- Martin Arjovsky, Soumith Chintala, and Lon Bottou. Wasserstein generative adver- sarial networks. In: International Conference on Machine Learning . 2017
- Ishaan Gulrajani et al. Improved training of wasserstein gans. In: Advances in Neural Information Processing Systems . 2017, pp. 57675777.
- Zhou Wang et al. Image quality assessment: from error visibility to structural similarity. In: IEEE transactions on image processing 13.4 (2004), pp. 600612
- Dominique Brunet, Edward R Vrscay, and Zhou Wang. On the mathematical properties of the structural similarity index. In: IEEE Transactions on Image Processing 21.4 (2011), pp. 14881499
- Daniel L Ruderman. The statistics of natural images. In: Network: computation in neural systems 5.4 (1994), pp. 517548
- Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a completely blind image quality analyzer. In: IEEE Signal Processing Letters 20.3 (2013), pp. 209212