IMPROVING THE VISUAL QUALITY OF GENERATIVE ADVERSARIAL NETWORK (GAN) - GENERATED IMAGES USING THE MULTI-SCALE STRUCTURAL SIMILARITY INDEX.

Parimala Kancharla, Sumohana S. Channappayya
Lab For Video & Image Analysis (LFOVIA)
Indian Institute of Technology Hyderabad
Emails: ee15m17p100001@iith.ac.in, sumohana@iith.ac.in

1. Objective and Approach

- Goal: Simple yet effective method to improve the visual quality of Generative Adversarial Network (GAN) [1] generated images.
- Approach: Image quality assessment metric is introduced into the loss function of GAN to guarantee the local structural and statistical integrity.

2. Boundary Equilibrium Generative Adversarial Network (GAN)

- Generative Adversarial Networks (GANs) 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:**

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

- BEGAN [2] is an extension of GAN, where the discriminator block is replaced with an autoencoder.
- The loss function $L(x)$ for autoencoder is defined as follows .

$$L(x) = |x - D(x)|^n; n = 1, 2.$$

- The objective function of BEGAN then becomes

$$L_D = L(x) - k_t L(G(z)) \text{ for } \theta_D,$$
$$L_G = L(G(z)) \text{ for } \theta_G,$$
$$k_{t+1} = k_t + \lambda_k(\gamma L(x) - L(G(z))) \text{ for training step } t.$$

- The parameters θ_G and θ_D are updated by minimizing the loss functions L_D and L_G respectively.
- k_t is the variable to control how much emphasis should be put on $L(G(z))$ during gradient descent.
- λ_k is the proportional gain for k_t.

3. Proposed MS-SSIM index Constrained BEGAN

- The autoencoder in BEGAN architecture allowed us to use the full reference image quality assessment metric.
- Multi scale Structural Similarity index (MS-SSIM) [3] is an image quality assessment technique, which measures the structural loss between two images.
- In the proposed method, the loss function of the BEGAN's discriminator is modified to be a weighted average of MAD and 1-(MS-SSIM).
- **Proposed loss function:**

$$L(x) = \lambda_1 L_1(x) + \lambda_2 L_2(x),$$
$$L_1(x) = |x - D(x)|^n,$$
$$L_2(x) = 1 - (\text{MS-SSIM}(x, D(x))).$$

- Where λ_1 and λ_2 are normalized weights given to each of the metrics. Specifically, $0 \leq \lambda_1, \lambda_2 \leq 1$ and $\lambda_2 = 1 - \lambda_1$.
- The proposed method is evaluated for various combinations of λ_1 and λ_2.

4. Quantitative Results

- The Frechet Inception Distance (FID) is used to quantify the quality of the generated samples.

$$\text{FID}(x,g) = ||\mu_x - \mu_g||_2^2 + \text{Tr}(\Sigma_x + \Sigma_g - 2(\Sigma_x\Sigma_g)^{\frac{1}{2}}),$$

- (μ_x,Σ_x) and (μ_g,Σ_g) are the mean vector and the covariance matrix of the sample embeddings from the real and generated distributions respectively.
- NIQE [4] is a popular no reference image quality assessment technique based on natural scene statistics.
- FID and NIQE scores are negatively correlated with the visual quality.

Model Parameters		FID	NIQE
λ_1(MAD)	λ_2(1-(MS-SSIM))		
1	0	77.41	8.53
0.9	0.1	72.91	8.93
0.5	0.5	**64.96**	**7.61**
0.1	0.9	71.35	8.54
0	1	70.72	8.83

Model Parameters		FID	NIQE
λ_1 (MAD)	λ_2 (1-(MS-SSIM))		
1	0	235.89	9.72
0.9	0.1	245.57	8.42
0.5	0.5	**205.03**	**7.33**
0.1	0.9	268.52	8.14
0	1	235.00	8.52

Table: Proposed BEGAN results on CelebA dataset (left) and Stanford cars dataset (right).

5. Qualitative Results

Figure: BEGAN-MAD ($\lambda_1 = 1$, $\lambda_2 = 0$) based approach.

Figure: BEGAN-MAD+MS-SSIM ($\lambda_1 = 0.5$, $\lambda_2 = 0.5$) based approach.

Figure: BEGAN-MAD ($\lambda_1 = 1$, $\lambda_2 = 0$) based approach.

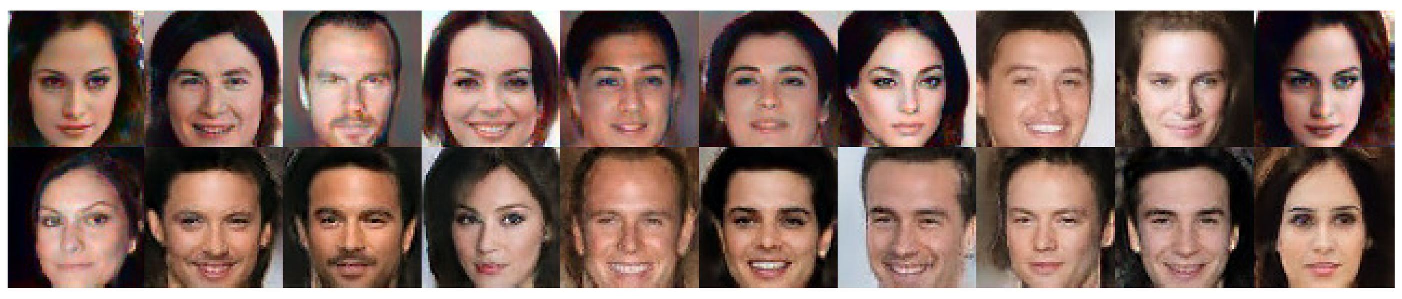

Figure: BEGAN-MAD+MS-SSIM ($\lambda_1 = 0.5$, $\lambda_2 = 0.5$) based approach.

- Randomly selected BEGAN generated images trained on the CelebA dataset and the Stanford Cars dataset.

6. Performance Evaluation

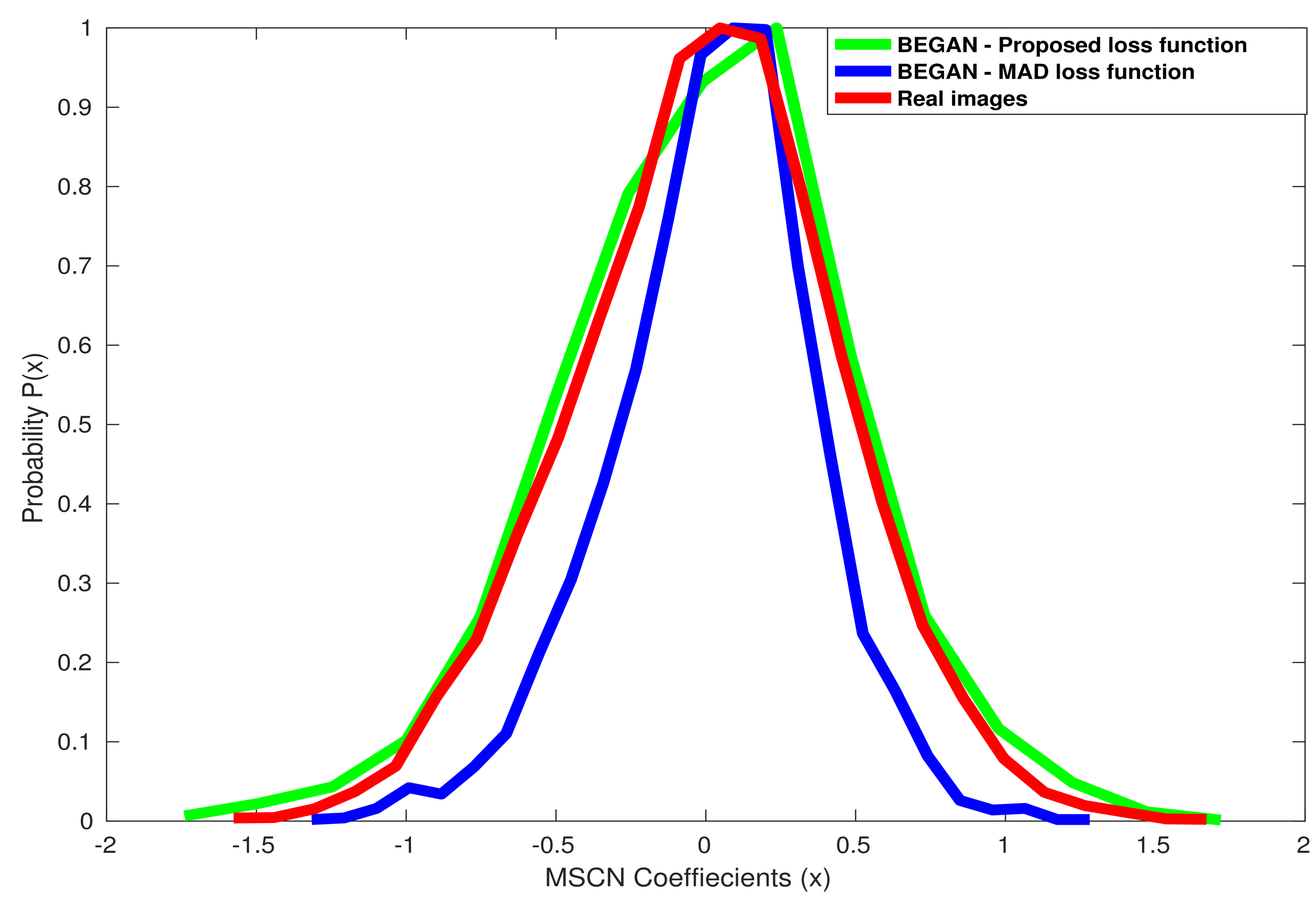

Figure: Normalized histograms of mean subtracted contrast normalized coefficients.

7. Conclusions and Future work

- We have explicitly integrated an image quality assessment model into the image generation model.
- Demonstrated that results are promising qualitatively and quantitatively.
- Future work: Build on these preliminary results by leveraging the rich literature on natural scene statistical models.

8. References

- I.Goodfellow, J.Pouget-Abadie, M.Mirza, B.Xu, D.Warde-Farley, S.Ozair, A.Courville, and Y.Bengio, "Generative adversarial nets," in Advances in Neural Information Processing Systems-2014
- D. Berthelot, T. Schumm, and L. Metz, BEGAN: boundary equilibrium generative adversarial networks, CoRR, vol. abs/1703.10717, 2017.
- Z. Wang, E. P. Simoncelli, and A. C. Bovik, Multi- scale structural similarity for image quality assessment, in Thirty-Seventh Asilomar Conference on, vol. 2. Ieee, 2003, pp. 13981402.
- A. Mittal, R. Soundararajan, and A. C. Bovik, Making a completely blind image quality analyzer, IEEE Signal Processing Letters, vol. 20, no. 3, pp. 209212, 2013