We thank the reviewers for their careful review and insightful comments. We address the comments in the following.

Limited nature of the contribution:

Through this work, our primary goal is to demonstrate that statistics of natural scenes, and the ability of IQA algorithms to quantify the naturalness of a scene have an important role to play in the generative modeling of natural images. The WGAN-GP framework provides us a good setting to convey our idea effectively. Nevertheless, we believe that the core idea of imposing a "naturalness" constraint in generating natural scenes would be effective wherever the discriminator function is smooth. This includes the 1-Lipschitz functions in WGAN-GP and PGGAN, hinge-loss objective function based Self-attention GAN and Spectrally normalized GAN, CT-GAN and WGAN-LP etc. Also, since we impose no constraints on the generator, we expect it to work well in the conditional GAN setting too.

To further justify our claim and implement reviewer suggestions, we have applied the proposed regularizers to the PGGAN architecture (both original and growing) at resolutions of 128×128 and 256×256 on the CelebA dataset, and show the results in Fig. 1. Interestingly and importantly, we observed that the proposed regularizers resulted in faster convergence and improved visual quality of the generated images. We hope that these results also address concerns about the effectiveness of QAGANs at higher resolutions. While memory and time constraints limited our testing to a resolution of 256×256 and 6K iterations, we are optimistic that our method would work at higher resolutions as well.

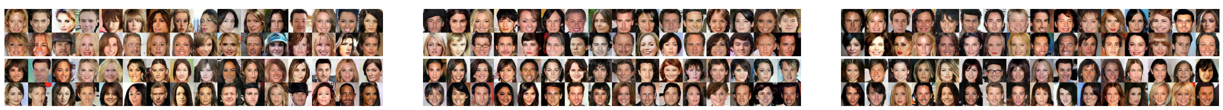

Figure 1: Top: 128×128. Bottom: 256×256. Left: PGGAN, $FID_{128\times128}$ = 64.50, $FID_{256\times256}$ = 62.86. Center: PGGAN with SSIM, $FID_{128\times128}$ = 47.46, $FID_{256\times256}$ = 38.324. Right: PGGAN with NIQE, $FID_{128\times128}$ = 49.80, $FID_{256\times256}$ = 44.84. These are results after 6K iterations on the CelebA dataset.

Comparison with the work by Kancharla and Channappayya, ICIP 2018 [Kancharla2018]:

While their work is similar in spirit, we present several fundamental differences in the following. First, our work clearly discusses the issues with the direct application of IQA algorithms as cost functions and proposes novel perceptual quality regularizers that are fine-tuned to the GAN framework - either that nicely fit the GAN math framework (SSIM-based) or that capture/model the local statistics of discriminator gradients (NIQE-based). Kancharla2018 on the other hand does a straightforward application of the MS-SSIM index and uses NIQE only for performance evaluation and not as a cost function. Next, our work presents a systematic stability analysis in the WGAN-GP setting and guarantees stability (please see supplementary material) while Kancharla2018 only presents empirical analysis. Also, they mention instabilities when the MS-SSIM term is given higher weightage. Further, we have conducted detailed experimental analysis and validation in our work. Finally, a qualitative comparison is shown in Fig. 2 from which it is clear that our method not only generates images with better structural information but also has greater diversity. This can be attributed to two main factors: our quality based regularizers and the improvements due to WGAN-GP relative to BEGAN.

Figure 2: Left image: Montage from Kancharla2018 (permission obtained from IEEE), with FID = 205. Right image: Montage from QAGAN with NIQE with FID = 86 (50K iterations). Note improved structural information and diversity.

Clarifications:

- We do not intend to portray that the SSIM index has not been used as a cost function in the literature. Rather, what we want to convey is that while IQA algorithms are indeed very effective, their usage as cost functions has not been as widespread as one would like due to their typically unwieldy mathematical formulation.
- Since the SSIM index can be negative, it no longer satisfies the requirement of a metric in the mathematical sense (i.e., $x, y \in \mathcal{X}$ for some set $\mathcal{X}$, $d(x, y) \geq 0$). We do not imply that the boundedness renders it an invalid metric.
- WGAN-GP uses the average of the error norm between the real and fake samples (without correspondence) as one of the elements of the cost function (Proposition 1, primal form in [Gul+17]). We reason that d^Q would be a better choice than error norm (in the average sense) for measuring the *perceptual distance* between the real and fake image sets. We also observed that average $d^Q(X, Y)$ values reduce with iterations.
- We have presented a convergence/stability analysis (for any $\lambda > 0$) of the proposed regularizers in the supplementary material provided with the initial submission. We point the reviewer to Sections 1 and 3 in that document.
- The λs were not tuned individually for FID and IS results. There were no stability issues with variation in λs. Nevertheless, our choice of λs is based on performance.
- We will incorporate presentation improvements in the final version if the submission is accepted.

Again, we thank the reviewers for their insightful comments that has led to important discussions and clarifications.