More to Perceptual Loss in Super Resolution
— Scope for Future —

Akella Ravi Tej

I. Abstract

This is a research proposal for **Designing Optimal Objective Functions for a Generative Model** as an extension of my undergraduate thesis project on **Super Resolution as a Supervised Learning Problem**. Designing an optimal objective function will yield better results in comparison to sub-optimal loss functions which leave checkerboard artifacts or texture patterns on the generated outputs. Pixel-wise loss functions are sensitive to geometric transformations and more tolerant to blurring, hence correlate poorly with perceptual quality. While perceptual loss functions show better correlation to human perception, they are often sub-optimal loss functions for a given task. This proposal is about developing a novel approach for designing a better class of perceptual functions that can be trained in an iterative manner alongside the generative model.

II. Background

Most of the work in super resolution (SR) in the last four years [1][2][4][5][7] concentrates on training deep neural networks for achieving state-of-the-art performance in PSNR metric. Almost all of these methods use pixel-wise loss functions leading to blurred results, demonstrating the inability of point estimates to capture the multi-modality of conditional distribution [3]. This drawback of point estimates called regression-to-the-mean problem, can be attributed to the unstable nature of high frequency information to geometric deformations under standard euclidean metric. It is ironic as PSNR has been demonstrated to correlate poorly with human perception but increases with minimization of pixel-wise loss. This is why networks trained on pixel-wise losses offer state-of-the-art performance on PSNR metric, but fail to produce outputs that are indistinguishable from natural high resolution (HR) images.

This advanced my research in a new direction with the objective of generating SR images that are indistinguishable from natural HR images. This correlates well with the objective function of generative adversarial networks (GAN) [10], where a discriminator network is trained jointly with the generator network. The difficulty of training a GAN from scratch is alleviated by combining the GAN loss function with mean squared error (MSE). Although training a GAN is a relatively more unstable task [11][12], it generates more realistic SR images. To avoid the pitfalls of MSE, perceptual loss functions were proposed that correlate better with human perception. The only class of perceptual loss functions that exist in the literature are deep convolutional architectures (show stability to small geometric deformations and are rich feature extractors) trained for a closely related auxiliary task (the feature representations learned from the auxiliary domain are assumed to generalize to the new domain) on a diverse dataset (provides better generalization to unseen data). SR models trained using a combination of perceptual loss and adversarial training [3][6][8][9] yield state-of-the-art results in Mean Opinion Scores (MOS). These results however are not state-of-the-art in PSNR, corroborating that PSNR is not a good metric to measure the performance of an SR model.

Using the latent representations of a network trained for an auxiliary task as the perceptual loss also transfers some high frequency information from of the auxiliary domain to the new domain. This results in unwanted artifacts and texture patterns that make the generated outputs easily distinguishable from their respective targets. For instance, minimizing the distance in feature space of VGG19 [13] implants a few characteristic traits (that correlate well with features used to discriminate between classes in ImageNet [14]) in the generator network which can be seen in the form of subtle textures in the generated outputs.

III. Research Methodology and Goals

I would like to research on novel methods to train the Perceptual loss function coupled with the generator-discriminator network. As the perceptual loss function is trained along with the generator-discriminator network, it is not expected to induce unwonted artifacts or texture patterns from other domains in the generated images. I believe this approach can help the generator network tap into the true distribution of natural HR images. These methods can further be generalized to any generative network leading to more stable adversarial training.

A perceptual function trained to distinguish SR images from target HR images will learn to extract features that differentiate between the two. The latent representation of this network would hence form a better perceptual loss function as it does not add unnecessary artifacts that would in turn help it better differentiate between the two classes. As the objective of the perceptual loss network is same as the discriminator network of a GAN, a well trained discriminator network becomes an ideal candidate for the perceptual loss function. However, the generator and discriminator networks get better at their respective tasks simultaneously, making it impossible to use the optimal perceptual loss function in combination with adversarial training.

A. *Iterative Approach*

I am proposing an iterative approach to train the perceptual function in various stages. As suggested in [9], let us first train the model using a combination of adversarial training and a sub-optimal perceptual loss. Upon reaching convergence, the discriminator network is chosen as the new perceptual loss function (still sub-optimal as this network is trained in conjunction with a sub-optimal generator network). The model is trained again from scratch using a combination of adversarial training and the new perceptual loss. This process is carried out until both the discriminator network and the perceptual network become equally good at their common objective.

IV. Conclusion

The proposed approach takes the best performing SR model and adapts its perceptual loss network in the process of training the generator thereby improving its performance. The proposed approach is speculated to produce more realistic HR images and stand as the new state-of-the-art on MOS metric.

REFERENCES

[1] C. Dong, C. C. Loy, K. He, and X. Tang. Learning a deep convolutional network for image super-resolution. In ECCV, pages 184–199, 2014. 1, 2, 3, 5, 6, 7, 8

[2] C. Dong, C. C. Loy, K. He, and X. Tang. Image super-resolution using deep convolutional networks. IEEE TPAMI, 2015. 6, 7

[3] J. Bruna, P. Sprechmann, and Y. LeCun. Super-resolution with deep convolutional sufficient statistics. In ICLR, 2016.
W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In CVPR, 2016.

[4] Kim, J., Lee, J.K., Lee, K.M.: Accurate image super-resolution using very deep convolutional networks. CVPR (2016)

[5] Kim, J., Lee, J.K., Lee, K.M.: Deeply-recursive convolutional network for image super-resolution. CVPR (2016)

[6] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In ECCV, 2016.

[7] Y. Tai, J. Yang, and X. Liu. Image super-resolution via deep recursive residual network. In CVPR, 2017.

[8] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in IEEE Conference on Computer Vision and Pattern Recognition, 2017. 4

[9] M. S. M. Sajjadi, B. Scholkopf, and M. Hirsch. Enhancenet: Single image super-resolution through automated texture synthesis. In ICCV, pages 4491–4500, 2017. 3

[10] 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, pp. 2672–2680.

[11] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training gans," in Advances in Neural Information Processing Systems, 2016, pp. 2226–2234.

[12] M. Arjovsky and L. Bottou, "Towards principled methods for training generative adversarial networks," NIPS 2016 Workshop on Adversarial Training, 2016.

[13] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In International Conference on Learning Representations (ICLR), 2015. 2, 3, 4, 5

[14] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248–255.