

Supplementary for: “Training a Better Loss Function for Image Restoration”

Anonymous ICCV submission

Paper ID 10510

This document includes additional details that could not be included in the main paper due to the lack of space. This comprises: *a)* manifold visualization for SR-GAN discriminator as compared to our multi-scale discriminators; *b)* qualitative results for the JPEG artefact removal application; *c)* ablation study on the choice of seed images and the number of discriminators of the Multi-Scale Discriminative Feature (MDF) loss; *d)* hyper-parameter tuning for the VGG and LPIPS feature-wise loss functions; and *e)* performance of loss functions as quality predictors. Finally, we provide an HTML report which comprises all the results.

1. Image manifold comparison

In this section, we repeat the experiment conducted in Sec. 4 (of the main paper), instead this time for a fully trained SR-GAN [2] discriminator. This further bolsters our claim that the task-specific discriminators of our MDF loss function learn to detect the generator distortions instead of the entire natural image manifold. This thereby allows our MDF loss function, trained on a single image, to be used to effective feature extractors between the generated and the reference image.

We chose the same sample of 100 natural images from the ILSVRC validation dataset [4]. From these images we generated *a)* JPEG compressed images using a compression quality between 7 and 10, *b)* blurry image samples by downsampling and upsampling the images by a factor of 4 using bi-linear filter and *c)* scrambled images by randomly permuting the pixels on each level of the Laplacian pyramid. Such permutations distort the second-order statistic, but preserve the composition of the spatial spectrum. A trained SR-GAN discriminator is used to extract the latent feature space of each set of images. The feature space for each image is chosen after the Global Average Pooling (GAP) layer of the network. We used t-SNE to reduce the dimensionality of the feature vector to 3 for visualization. Fig. 1 shows the plot of the features from each set of images. The visualization shows that the discriminator of SR-GAN learns the natural image manifold (unlike our multi-scale discriminator) and can discriminate between natural and randomly permuted images. However, it cannot dis-

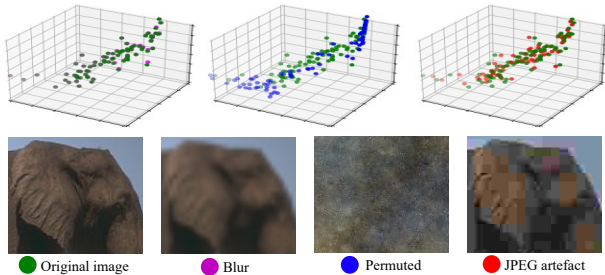


Figure 1: Manifold assumption validation: The figure shows the 3D t-SNE plots of the latent feature vectors extracted from diverse sets of images using an SR-GAN discriminator trained on DIV2K dataset [1]. The SR-GAN discriminator cannot differentiate between the original and jpeg images (right plot), thereby cannot be used as an effective feature extractor to detect and remove distortions.

criminate between the JPEG compressed and original images, making it an inferior feature extractor to detect and remove distortions.

2. JPEG artefact removal results

In this section, we provide qualitative results showing comparison between three sample reconstructed images from the BSD Test Set using our (MDF) loss with various other loss functions for the task of JPEG artefact removal application. The test images are compressed with a quality factor of 10 and a more challenging factor of 7. Fig. 2 shows the results for the compression quality factor 7. The performance of the various loss functions seems to be comparable for the quality factor of 10, however, our model substantially provides artefact removal, especially in the uniform areas of the image for a much challenging codec quality of 7. The same was also observed in the subjective experiment conducted (see Sec. 5.1 of the main paper). Additional qualitative results can be seen in the HTML report attached to the Supplementary Material.

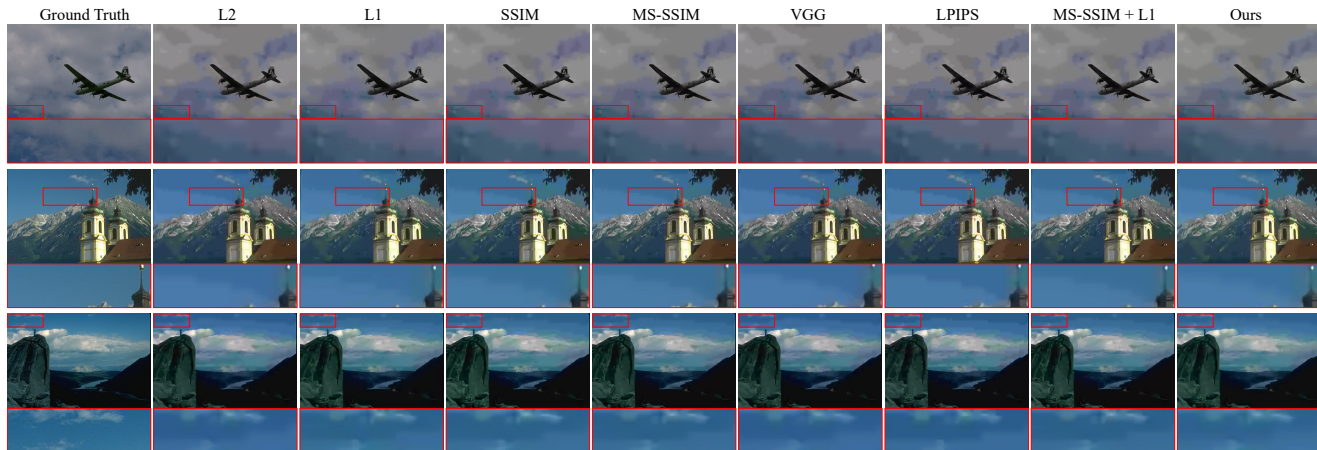


Figure 2: Results for JPEG artefact removal (compression quality = 7) using DnCNN model [5] trained using different losses. Our loss improves artefact reduction, especially in the uniform areas of an image. Qualitative results in terms of PSNR, SSIM and NIQE are reported in Table 2 of the main paper. Best viewed when zoomed.

Table 1: Ablation study on training the SISR model (EDSR) using different scales of our loss. The scale number represents the number of scales included in the MDF loss.

Scales	1	2	3	5	7	8
PSNR \uparrow	22.55	23.89	24.43	24.89	25.27	25.37
SSIM \uparrow	0.51	0.59	0.62	0.66	0.68	0.70
NIQE \downarrow	6.109	5.358	4.953	4.124	3.979	3.635

3. Ablation study

3.1. Scales of Discriminators

Since our MDF loss function comprises a series of discriminators trained on a single image at various scales, we need to select the optimal number of scales (the hyperparameter K in Equation 2 of the main paper) to achieve the best performance. We perform an ablation study on training the EDSR model [3] using only the coarsest scale discriminator and subsequently adding finer scales. We observe a significant increase in quality of the images generated with the increase in the number of discriminators. As shown in Table 1, our loss performs the best when all 8 scales are employed.

3.2. Seed image

Next, we study the effect of using different natural and synthetic images for training our MDF loss function. Fig. 3 shows five seed images including two natural and three synthetic ones that were used to train the discriminators. *Pyramid Permutation* image has been created by a random permutation of pixel order on each level of the Laplacian pyramid. Such permutation distorts image second-order statistics, but preserves the composition of the spatial spectrum. *Pink Noise* image contains $1/f^2$ noise that is typical for natu-

ral images. *Contrast Rings* image contains concentric rings whose contrast is reduced towards the centre to cover the range of edges of all orientations and contrast magnitudes. The results of SISR (EDSR), shown in the bottom part of Fig. 3, indicate that the visual quality of the super-resolved images is the best for natural images and is degraded as the statistics of the training images is distorted. However, from the results for all the applications, the visual quality of the restored images is more dependent on the nature of the distortions added (z^k) than the choice of the seed image.

4. Hyperparameter-tuning for VGG and LPIPS

In Fig. 4 we show the qualitative results for the trade-off between the MSE and LPIPS/VGG network components in the joint loss function. For fair comparison, we conducted a hyper-parameter search over the scalar λ controlling the weight of the feature-wise loss function. We searched over the values in $\{\lambda : \lambda = 10^k, k = -3, \dots, 3\}$. The greater λ parameter is, the more LPIPS/VGG components contribution is. In our experiments across all image restoration applications, we found the best results are produced when $\lambda = 1$ for VGG and $\lambda = 0.1$ for LPIPS loss. Additional qualitative results are provided in the HTML report.

5. Image quality metrics and loss functions

To further investigate the performance of loss functions as quality predictors, we generated a set of images that were distorted by blur, noise, added sinusoidal grating, contrast and brightness changes. The distortions were generated so that they degraded the image in equal steps of PSNR. Fig. 5 presents an example of images with introduced distortions at three PSNR levels. The experiment shows a failure case of PSNR, predicting the same quality even though the dis-

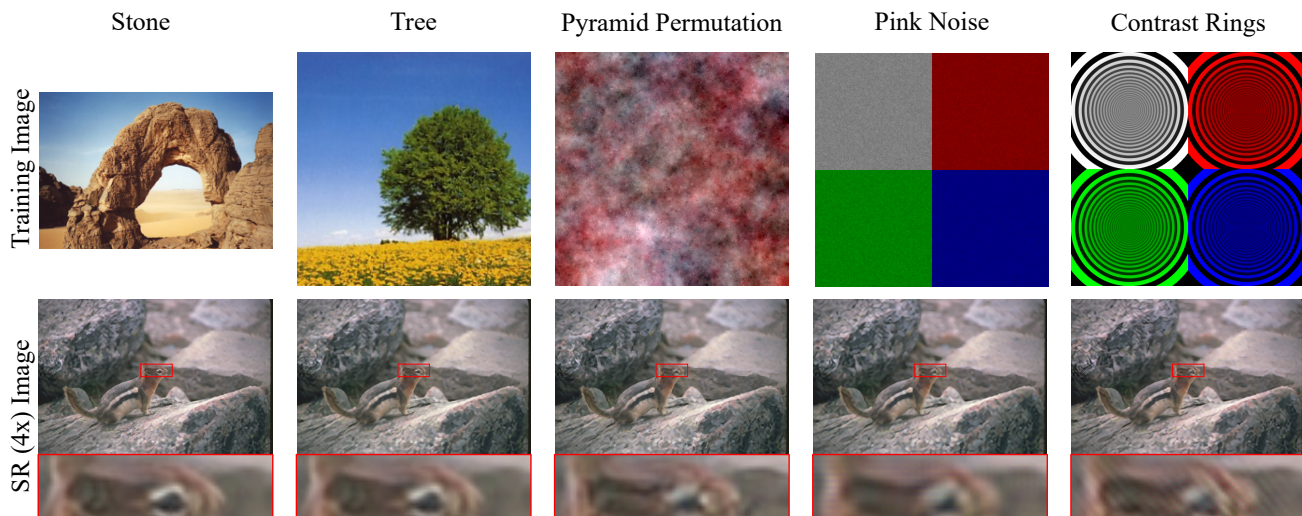


Figure 3: Ablation study on changing the image used for training our MDF loss function. It can be seen that natural images provide visually better results as compared to synthetic images. Best viewed when zoomed.

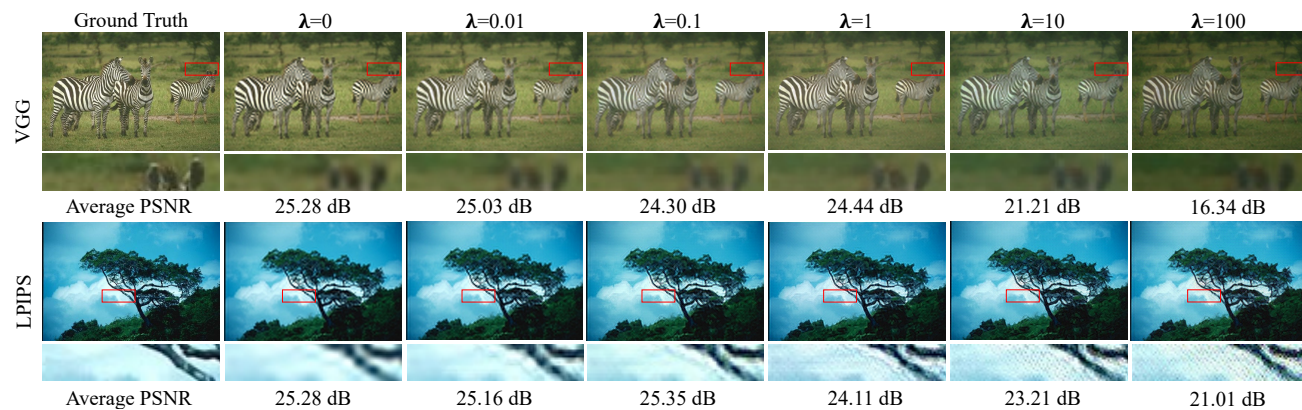


Figure 4: Comparison of the single-image super resolution (SISR) results (EDSR) when trained using a weighted sum of VGG/LPIPS and MSE feature-wise losses: $\text{MSE} + \lambda \text{VGG/LPIPS}$. The average PSNR is reported for the entire test set.

tortions due to contrast and brightness are much less objectionable than the others to a human observer.

In Fig. 6, we show the loss values computed for the increasing amount of distortions of different types for different loss functions. Despite the same PSNR value, the distortions due to noise, blur and added sinusoidal wave are much more noticeable than those due to contrast and brightness change (refer to Fig. 5). The loss functions derived from quality metrics (SSIM, MS-SSIM) and also feature-wise losses (VGG, LPIPS) penalize more the distortions that result in higher degradation of quality. In contrast, MDF losses penalize the most the distortions that are relevant for a given task: blur in case of SISR (MDF SR), blur and noise in case of denoising, and contrast followed by the mixture of all distortions in case of JPEG artifact removal. This is another example demonstrating that an effective loss

(MDF) function does not need to predict image quality.

6. HTML report

In the Supplementary Material, we provide a comprehensive HTML report, showing the results for each loss function across different image reconstruction applications for various datasets. We further provide results for the ablation study and the hyper-parameter selection. The HTML report, including all the inference images, are attached with the supplementary material. Please visit the URL [HTML_Report_Paper_ID_10510.html](https://arxiv.org/html/2010.10510v1) inside the folder named “Report_10510”.

Due to size limitations for the supplementary material, we include the first 30 images from each test set. Images are stored as JPEGs with a quality of 90 to ensure that coding

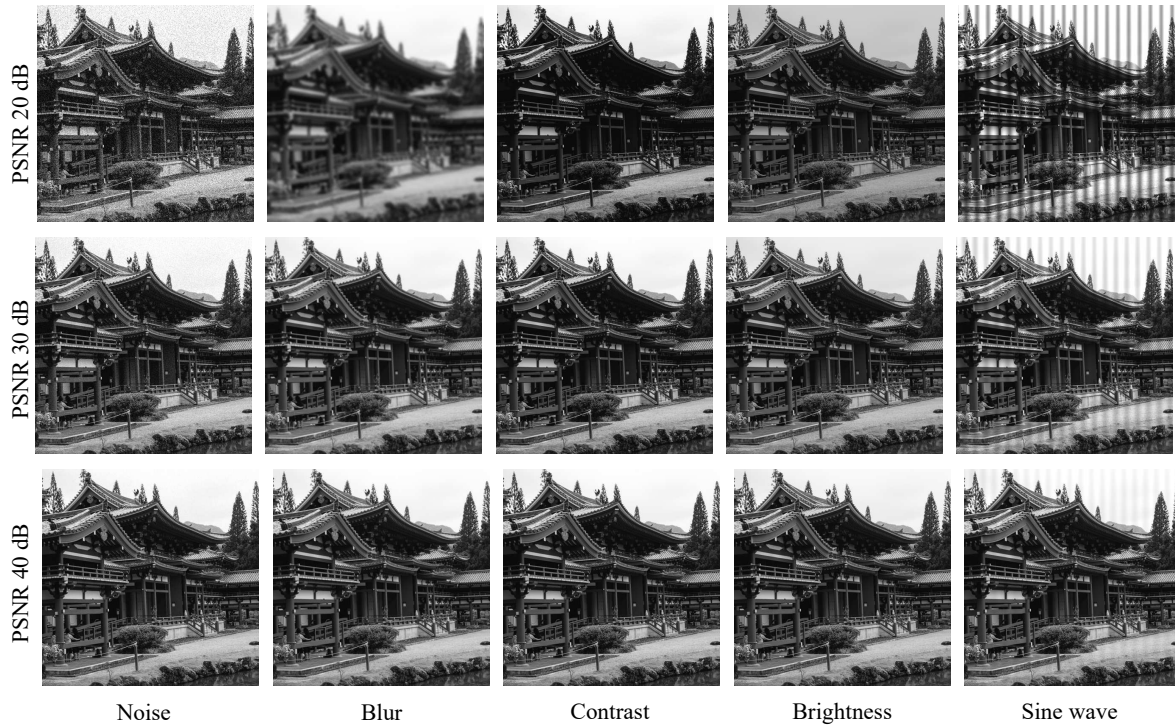


Figure 5: Examples of images used to test the sensitivity of loss functions to different types of distortions. We introduced artifacts so that the each distortion results in the same PSNR level (across each row). Here we provide examples of images at 20 dB, 30 dB and 40 dB. Note that the perceived quality differs between the columns despite the same PSNR level.

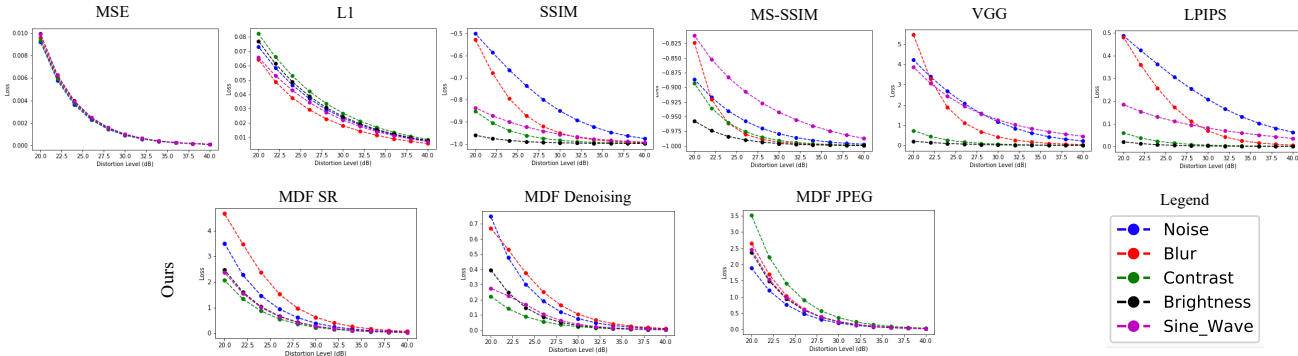


Figure 6: Loss values for the increasing amount of distortions of different types. The distortion levels have been generated to result in equal PSNR values, shown on the x-axis. Despite the same PSNR value, the distortions due to noise, blur and added sinusoidal wave are much more noticeable than those due to contrast and brightness change (refer to Fig. 5). The MDF loss accurately predicts the perceived magnitude of task specific distortions for which it is trained.

distortions do not distort the results. Upon acceptance, the code and the complete set of inference outputs will be made public for the research community.

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceed-*

ings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 126–135, 2017. 1

[2] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690,

2017. 1

[3] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017. 2

[4] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015. 1

[5] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017. 2

486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539