CS598PS Final Report
FRC-GAN

Ufuk Soylu, Ulas Kamaci, Berk Iskender

***Abstract*—Recently, generative models are used to obtain a framework to provide photo-realistic images for super-resolution by learning the distribution of the dataset. Despite having visually attractive images, they do not provide scientifically reliable results, and there isn't any method developed for super-resolution which takes the resolution into consideration. In this context, we propose a new loss function for generative adversarial training that produces images correlated at different spatial frequencies with their ground truth high frequency counterparts. We compare the training results obtained by using our loss function with the state-of-the art results and see that our loss function provides much better results in higher frequencies while keeping the same quality for lower frequencies. Additionally, our loss function eliminates the high frequency artifacts generated by current GAN models.**

I. Introduction

Super-resolution (SR) problem is the task of estimating a high-resolution (HR) image from its low-resolution (LR) counterpart and it gathered significant consideration from the computer vision research community [1], [2]. SR has the ill-posed nature due to under-determined structure. Absent texture details in LR image lead to missing texture details in reconstructed SR image and there are infinitely many reconstructions for a given LR image which makes SR an ill-posed problem.

Classical approaches to solve SR is based on filtering approaches, e.g. linear, bi-cubic or Lanczos filtering [3]. These approaches are very fast but they yield solutions with overly smooth textures. There are more powerful methods than classical approaches such as in Glasner et al. [4], they exploit redundancies across scales within the image in self-similarity paradigm where self dictionaries are learned. Gu et al. [5] proposed a convolutional sparser coding approach that improves consistency by considering whole image not only overlapping patches.

There are many other approaches like neighborhood embedding approaches [6], combining an edge-directed SR based on a gradient profile prior [7]. However, in this work, we study convolutional neural network based SR algorithms which have superior performance. In Wang et al. [8], they used their feed-forward network architecture based on the learned iterative shrinkage and thresholding. Dong et al. [9] trained three layer fully connected network to learn a mapping between bi-cubic interpolation to high resolution image. Later, it was shown that learning the up-scaling filters directly increases performance both in terms of accuracy and speed [10], [11]. However, these methods commonly minimize MSE between the reconstructed HR and the ground truth. Even though minimizing MSE maximizes the peak signal-to-noise-ratio (PSNR), it lacks the ability to capture perceptually relevant differences [12]. As depicted in Figure 1, the MSE-based solutions lead to overly smooth reconstructions because of the pixel-wise average of possible solutions. Therefore Ledig et al. [12], which is of also greatest interest for this work, suggests to use generate adversarial networks to push the reconstructions closer to the natural image manifold to produce perceptually more convincing reconstructions.

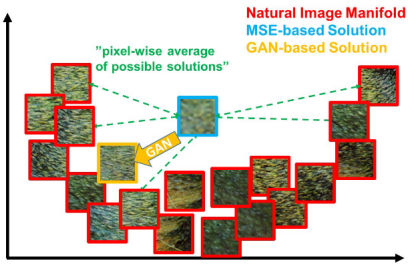


Fig. 1: Illustration of patches from the natural image manifold and super-resolved patches obtained with MSE and GAN from [12].

Ledig et al. [12] provides a framework for using GANs in the context of super-resolution. They developed a new perceptual loss where it consists of two parts: (i) content loss and (ii) adversarial loss. It can be seen as a regularized loss form where content loss as the data fidelity is calculated based on MSE and adversarial loss as the regularizer is calculated based on the discriminator output.

In this work, we develop a new loss function by further developing Ledig et al. [12]'s perceptual loss. In the super-resolution literature including deep learning based algorithms, there isn't any loss metric that is related to any kind of resolution definition. This is a fundamentally missing component since the SR problem naturally tries to increase the lower resolution to the higher resolution.

To introduce a loss metric based on a resolution definition, Fourier ring correlation (FRC) [13], which measures the degree of correlation of two images at different spatial frequencies, is chosen to be implemented since recent work suggests that they can be utilized in image restoration tasks such as denoising and deconvolution to design powerful algorithms [14]. Although this metric has been heavily used in certain microscopy applications, it didn't gather any attention for deep learning research or other imaging modalities.

II. Dataset

In this work, the dataset provided by Visual Object Classes Challenge 2012 (VOC2012) [15] is used for generative network training and testing. The dataset can be used for classification/detection, segmentation, action classification and person layout taster. There are a significant number of different types of visual objects in realistic scenes in the data-set. In our setting, the images from the data-set are considered as high resolution ground truth data. Low resolution counter parts are obtained by applying an anti-aliasing filter followed by an down-sampling operation.

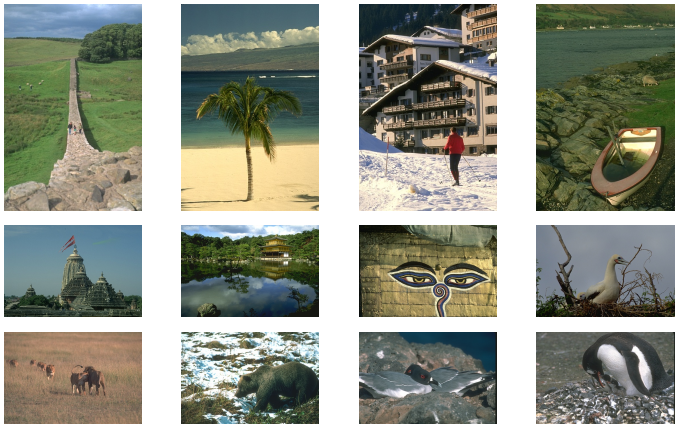


Fig. 2: Several examples from the dataset [15].

III. Background

A. Generative Adversarial Networks

Adversarial network architecture is introduced by Goodfellow et al. [16]. They define a discriminator network D_{θ_D} and a generator network G_{θ_G} which these networks are used to optimize in an alternating manner to solve the adversarial min-max problem:

$$\min_{\Theta_G}\max_{\Theta_D}\mathbb{E}_{I^{HR}\sim p_{train}(I^{HR})}\left[logD_{\theta_D}(I^{HR})\right] + \mathbb{E}_{I^{LR}\sim p_G(I^{LR})}\left[log(1-D_{\theta_D}(G(_{\theta_G})I^{HR}))\right]$$

The fundamental insight behind this formulation is that it allows one to train a generative model G with the goal of tricking a differentiable discriminator D that is also trained simultaneously to distinguish super-resolved images from real HR. By following this approach, a generator can be learnt to create solutions that are highly similar to real HR i.e. close to the natural image manifold. This leads to perceptually superior solutions and this is in contrast to SR reconstructions produced by minimizing over MSE error.

B. Fourier Ring Correlation

FRC measures the normalized cross-correlation in the Fourier domain as a function of spatial frequency:

$$\mathrm{FRC}(r_i) = \left(\sum_{r\in r_i} F_1(r)\cdot F_2(r)^*\right) / \left(\sqrt{\sum_{r\in r_i} F_1^2(r)\cdot F_2^2(r)}\right) \quad (1)$$

where F_1 and F_2 are the 2D Fourier transforms of the two images in polar coordinates, r is the radius in the frequency domain where origin corresponds to DC, and r_i is the set of spatial frequencies included in the i^{th} ring.

Obtaining the FRC curve for a pair of images using (1) is illustrated in Figure 3. Two rings depicted in the Figures 3c-3d correspond to the vertical lines given in Figure 3e. Since it is normalized, the maximum value of FRC is 1, and we are taking the real value of the complex correlation. FRC curve can be used to determine the resolution of an image by using a threshold. Common thresholds include 0.5, 0.143 and 2-sigma criterion. See [17] for a detailed discussion of different threshold criteria.

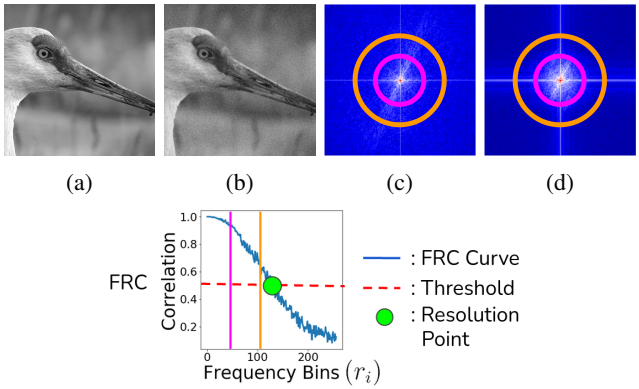


Fig. 3: (a) Image 1 (b) Image 2 (c) Fourier transform of image 1 (d) Fourier transform of image 2 (e) FRC curve.

FRC, together with its 3D extension Fourier shell correlation (FSC), have been used for determining the resolution in various microscopy applications such as fluorescence light microscopy [18], cryo-electron microscopy [13], [19], and X-ray imaging microscopy [20]. It has also recently been utilized as a tool in image restoration problems like blind deconvolution and denoising [14].

IV. Proposed Approach

A. Network Architecture

The network architecture is chosen as in Figure 4. It is the identical network structure as [12]. By this way, a fair comparison of our new loss function to their loss function will be obtained.

Specifically, there is a deep generator network with B residual blocks with identical layout. There are two convolutional layers with small 3×3 kernels and 64 feature maps followed by batch-normalization layers and parametric ReLU as the activation layers. The resolution of the input image is increased with two trained sub-pixel convolution layers. In order to distinguish real HR from super-resolved images, there is a discriminator which is shown in Figure 4. It contains eight convolutional layers with an increasing number of 3×3 filter kernels, increasing by a factor of 2 from 64 to 512 kernels. Additionally, image resolution is reduced by using strided convolutions.

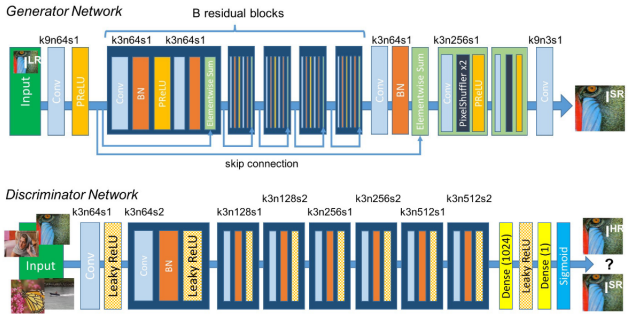


Fig. 4: Architecture of Generator and Discriminator from [12] where k is the kernel size, n is the number of feature maps and s is the stride.

B. Generator Losses

1) FRC loss: To implement FRC loss, we modified the loss function existing in SR-GAN [12] and obtained l_{mod} as follows:

$$l_{mod} = l_{content} + \lambda l_{\text{Adv}} + \gamma l_{\text{FRC}}, \tag{2}$$

where the l_{FRC} is given as

$$\begin{aligned} l_{\text{FRC}} &= |\mathbf{1} - \text{Re}\{\text{FRC}(\mathbf{r})\}| \\ &= \sum_i (1 - \text{Re}\{\text{FRC}(r_i)\})^2 \end{aligned} \tag{3}$$

where $\mathbf{r} \in \mathbb{R}^d$, $\mathbf{1} \in \mathbb{R}^d$.

Since autograd of Pytorch [21] does not allow automatic differentiation over complex numbers, we had to implement the FRC loss by explicitly computing the real and imaginary components of the complex values in the FRC computations. To this end, we implemented complex multiplication, addition, and division modules which make use of 2D real valued tensors where each dimension stores the real and the imaginary components, respectively. By doing so, we were able to utilize the automatic differentiation to update the generator parameters.

2) L2-loss in image domain: In (2), $l_{content}$ includes L2-loss in image domain and is defined as

$$l_{MSE} = \frac{1}{r^2 W H} \sum_{x=1}^{rW} \sum_{y=1}^{rH} (I_{x,y}^{HR} - G_{\theta_G}(I_{x,y}^{LR}))^2$$

where r is the down-sampling factor, W is the width of LR, H is the height of LR.

3) Content loss using a pre-trained network: The second component of (2) is instead of relying pixel-wise losses, we use a pre-trained 19 layer VGG and define VGG loss as the euclidean distance between final feature representations:

$$l_{VGG} = \frac{1}{W_{i,j} H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}(I_{x,y}^{HR}) - \phi_{i,j}(G_{\theta_G}(I_{x,y}^{LR})))^2$$

where $\phi_{i,j}$ indicate the feature map, $W_{i,j}$ and $H_{i,j}$ describe the dimensions of the respective feature maps.

4) TV loss: Another component that is used in training the generator is total variation regularization which is defined as

$$l_{TV} = \sum_{i,j} \sqrt{I_{i+1,j}^{SR} - I_{i,j}^{SR}|^2 + |I_{i,j+1}^{SR} - I_{i,j}^{SR}|^2}$$

5) Adversarial loss: In addition to previous losses, there is the adversarial loss which encourages our generator network to favor solutions on the natural image manifold.

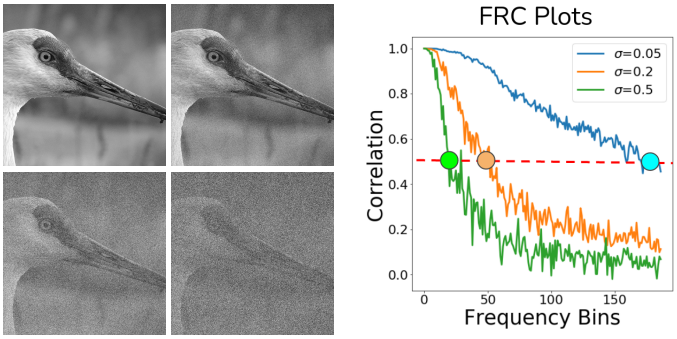
$$l_{Adv} = \sum_{n=1}^{N} -log D_{\theta_D}(G_{\theta_G}(I^{LR}))$$

V. Experiments and Results

A. FRC experiments

Here we present two experiments that provide intuition on how FRC works.

1) Noise addition: In this experiment we add white Gaussian noise with standard deviations (STD) of 0.05, 0.2, and 0.5 to a reference image and plot at the resulting FRC curves. The images and the FRC curves are given in Figure 5.



(a) (b)

Fig. 5: (a) Reference and the noisy images (b) FRC curves.

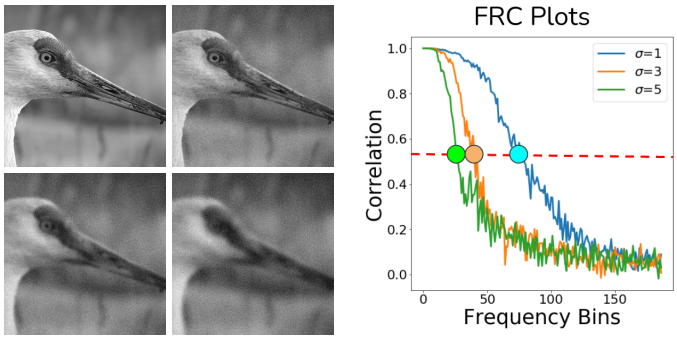
The first observation in Figure 5b is that, as the spatial frequency increases, FRC decreases. The explanation is that white noise has the same energy at all frequencies, however, natural images tend to have less energy at higher frequencies. So, the higher frequency content gets buried under noise. The second observation is that the FRC gets lower as the noise level gets higher. The energy argument also applies here. Higher noise energy leads to lower correlation.

2) Blurring and noise addition: In this experiment we first apply Gaussian blur with kernel STDs of 1, 3, and 5 pixels to a reference image, then add white Gaussian noise with the STD of 0.05, and plot the resulting FRC curves. The images and the FRC curves are given in Figure 6.

As higher levels of blur suppresses high frequencies in the signal more, we see lower FRC values for higher levels of blur.

B. Verification of the effect of higher frequency content on the PSNR and SSIM metrics

It is well-known that the most of the energy is located around the lower frequencies in the Fourier domain for natural images. A typical example is illustrated in Fig. 3.



(a) (b)

Fig. 6: (a) Reference and the blurry images (b) FRC curves.

Also, cumulative fraction of total energy contained in the circle of radius r in the Fourier domain for all test dataset is shown in Fig. 7. Moreover, our specific implementation of

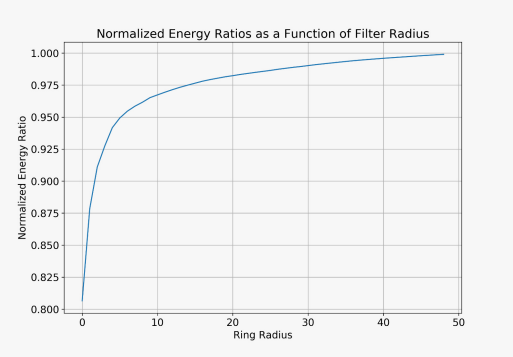


Fig. 7: Cumulative fraction of total energy contained in radius r for all test dataset.

the FRC loss puts larger importance on the higher frequencies since we compute the MSE loss with respect to 1 for every r, which is the maximum possible correlation value, and the FRC curve decays as r increases. A possible concern that we had was related to FRC loss causing suboptimal performance over the lower frequencies, since it does not take how much of energy is located at any radius r into account and treats correlations at each radius equivalently. Firstly, we tried to show that it is vitally important to recover higher frequencies. For this purpose, we conducted the following experiment: We reconstructed the images by only using the frequency components included in the circle or radius r and discarded the rest of the frequencies. Then, we obtained the average SSIM and PSNR values for each r over the test dataset that we used for assessing the performance of our proposed method. By doing so, we were able to verify that a significant improvement can be obtained by recovering high frequencies better. PSNR and SSIM vs. r plots are provided in Fig. 8 and 9. Even after including the 97% of the total energy, it is only possible to obtain an SSIM around 0.6 and a PSNR of 23 dB. This signals the importance of correct recovery of high frequency components.

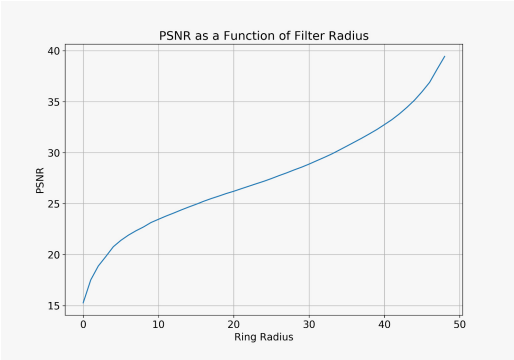


Fig. 8: Average PSNR (in dB) vs. r for the test dataset.

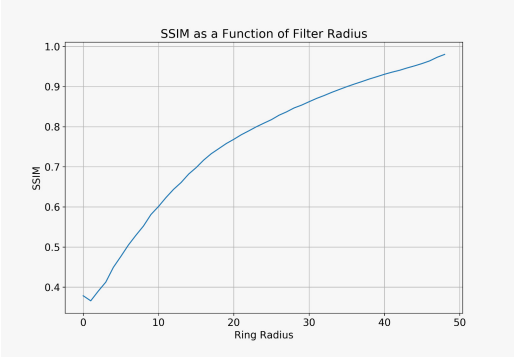


Fig. 9: Average SSIM vs. r for the test dataset.

C. Comparison across different methods

1) $l2$ ***FRC loss:*** After training SRGAN and our modification SRGAN-FRC on the patches obtained from BSDS300 dataset, we compared the mean FRC curves obtained from the test dataset. For this task, 4x upscaling was needed to perform super-resolution. In our comparison, we also added the results obtained using traditional interpolation methods of bilinear and bicubic. The comparison can be seen in Fig. 10.

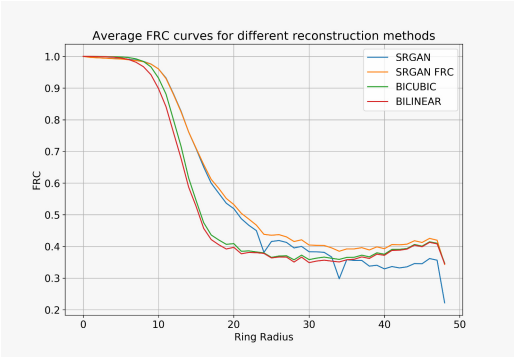


Fig. 10: Average FRC curves for different super-resolution methods.

Without the FRC loss, l_{FRC}, even bilinear and bicubic interpolations provide better correlations compared to SRGAN for higher frequency rings. This can be caused by the hallucinated details in high frequencies and artificial background textures that occur for almost all test images for plain SRGAN implementation.

Once l_{FRC} is incorporated into the training procedure, we were able to obtain superior performance over all values of r compared to all other methods. While improving over the higher frequencies, the method does not suffer any drawbacks on lower frequencies. Results for several random patches obtained from the test dataset can be observed in Fig. 11. An important note is that due to the instabilities after the FRC loss implementation in training phase, we were able to use 1000 epochs for SRGAN training but only 650 epochs for training the SRGAN-FRC with the same learning rate and optimizer.

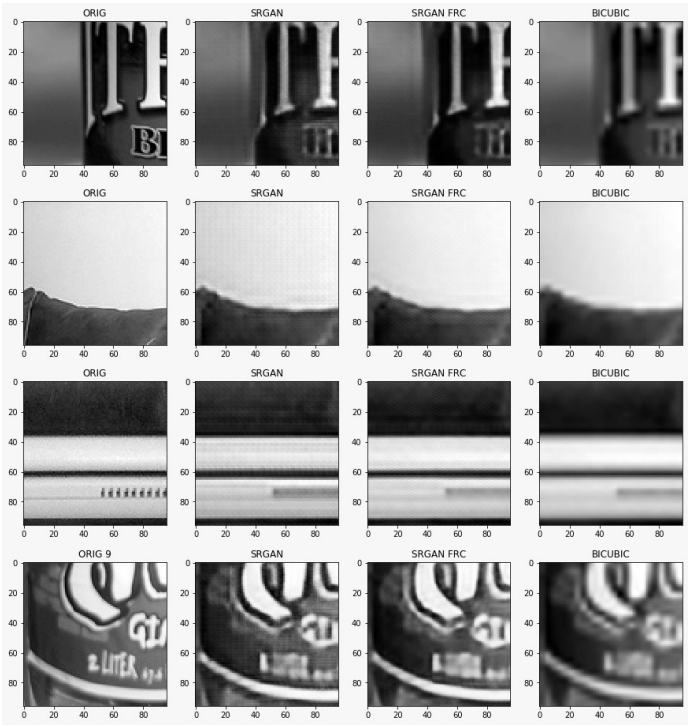


Fig. 11: Comparison of several patches obtained from the test dataset for each method: (i) Original high resolution, (ii) SRGAN, (iii) SRGAN-FRC, and (iv) bicubic interpolation.

As expected, bicubic interpolation provides the most smooth result. SRGAN seems to solve this problem by providing sharper features, however, it adds an artificial high frequency background texture to almost all patches. Also, as shown in Fig. 10, these sharper features are not correlated with the original image content in the Fourier domain. While suppressing these texture artifacts to a great extent, SRGAN-FRC results still provide sharper images.

*2) **Concentrated** $l2$ **FRC loss**:* For mission critical applications, it can be desired to recover certain frequency bands, and hence certain resolution levels, especially better. The loss that we propose is flexible in that sense. Just as applying a filtered error term in spatial domain, it is possible to constrain our loss to be computed on certain frequency domain rings. Assuming the set $D = \{r_1, \ldots, r_n\}$ as the radii of the rings, the loss takes the following form

$$l_{\text{FRC}} = \sum_{r_i \in D} (1 - \text{Re}\{\text{FRC}(r_i)\})^2 \tag{4}$$

where $\mathbf{r} \in \mathbb{R}^d$, $\mathbf{1} \in \mathbb{R}^d$.

Considering the image size as $m \times m$, we conducted experiments by selecting the set D as the rings with radii between $r = m/6$ and $r = m/3$. The comparison of the average FRC curves for concentrated and regular implementations of $l2$ FRC losses is provided in Fig. 12.

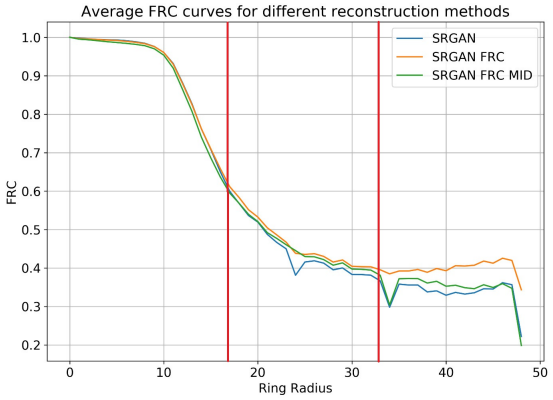


Fig. 12: Average FRC curves for concentrated and regular loss functions. Concentrated range of r is between the vertical red lines.

As expected, the concentrated loss provided suboptimal performances for the high frequencies due to not including those radii in our loss. In the concentrated region, both concentrated and regular $l2$-norm losses provide comparable results. However, our expectation was to obtain better results which was not realized through this experiment. Nevertheless, the reason can still be related to the instabilities of the training process for concentrated case and we are planning to conduct more experiments regarding this method.

*3) $l1$ **FRC loss**:* Instead of computing the squared error, we also experimented on using the absolute error as

$$\begin{aligned} l_{\text{FRC}} &= ||\mathbf{1} - \text{Re}\{\text{FRC}(\mathbf{r})\}||_1 \\ &= \sum_i |1 - \text{Re}\{\text{FRC}(r_i)\}| \end{aligned} \tag{5}$$

where $\mathbf{r} \in \mathbb{R}^d$, $\mathbf{1} \in \mathbb{R}^d$. This version of loss function turned out to be unstable during the training process compared to the squared error and we were able to perform training using 250 epochs. A comparison of average FRC curves can be seen in Fig.13.

We observed slightly suboptimal performance in terms of FRC values when we used $l1$-norm error. However, this suboptimality can be attributed to the unstable nature of the training process and less number of epochs used because of it. Main drawback of $l1$ loss was a shift in the color of the outputs which can be seen in Fig. 14.

*4) **Other performance metrics comparison**:* Finally, we computed the average PSNR (in dB) and SSIM over test set results to compare the different loss functions and methods on common performance metrics. To focus on best performing

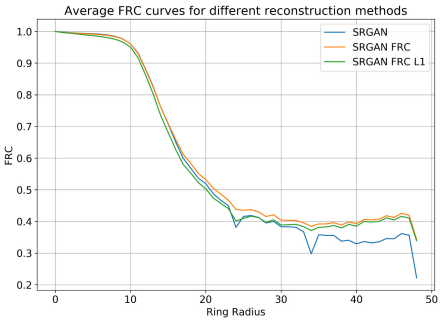


Fig. 13: Average FRC curves for different loss functions.

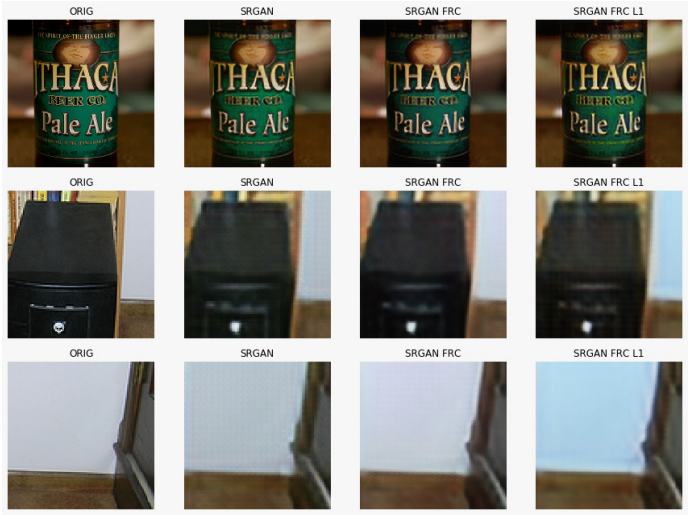


Fig. 14: Comparison of a full image (Top) and two patches obtained from the test dataset for each method: (i) Original high resolution, (ii) SRGAN, (iii) SRGAN-FRC, and (iv) SRGAN-FRC L1.

methods, concentrated $l2$ norm FRC loss and bilinear interpolation results are not displayed. As shown in Table I, the best results for both PSNR and SSIM is obtained using $l2$ FRC and $l1$ FRC losses, respectively.

TABLE I: Average accuracy results for test images for the algorithms 4x super-resolution.

	SRGAN	SRGAN-FRC	SRGAN-FRC L1	Bicubic
PSNR (dB)	24.162	24.079	**24.245**	24.087
SSIM	0.699	**0.704**	0.690	0.689

VI. Discussion and Conclusions

Recently, generative models provided a framework to provide photo-realistic images for super-resolution by learning the distribution of the dataset. However, having visually plausible images does not necessarily translate into scientifically reliable results, and usually, methods developed for super-resolution in this context do not take the resolution directly into consideration. Considering these aspects, we proposed a new loss function for GAN training procedure that produces images correlated at different spatial frequencies with their ground truth high frequency counterparts.

References

[1] Q. Yang, R. Yang, J. Davis, and D. Nistér, "Spatial-depth super resolution for range images," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2007, pp. 1–8. 1

[2] D. Chowdhuri, K. Sendhil Kumar, M. R. Babu, and C. P. Reddy, "Very low resolution face recognition in parallel environment," *IJCSIT) International Journal of Computer Science and Information Technologies*, vol. 3, no. 3, pp. 4408–4410, 2012. 1

[3] C. E. Duchon, "Lanczos filtering in one and two dimensions," *Journal of applied meteorology*, vol. 18, no. 8, pp. 1016–1022, 1979. 1

[4] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *2009 IEEE 12th international conference on computer vision*. IEEE, 2009, pp. 349–356. 1

[5] S. Gu, W. Zuo, Q. Xie, D. Meng, X. Feng, and L. Zhang, "Convolutional sparse coding for image super-resolution," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1823–1831. 1

[6] R. Timofte, V. De Smet, and L. Van Gool, "Anchored neighborhood regression for fast example-based super-resolution," in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 1920–1927. 1

[7] J. Sun, Z. Xu, and H.-Y. Shum, "Image super-resolution using gradient profile prior," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2008, pp. 1–8. 1

[8] Z. Wang, D. Liu, J. Yang, W. Han, and T. Huang, "Deep networks for image super-resolution with sparse prior," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 370–378. 1

[9] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2015. 1

[10] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *European conference on computer vision*. Springer, 2016, pp. 391–407. 1

[11] 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 *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1874–1883. 1

[12] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. 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*, 2017, pp. 4681–4690. 1, 2, 3

[13] W. Saxton and W. Baumeister, "The correlation averaging of a regularly arranged bacterial cell envelope protein," *Journal of microscopy*, vol. 127, no. 2, pp. 127–138, 1982. 1, 2

[14] S. Koho, G. Tortarolo, M. Castello, T. Deguchi, A. Diaspro, and G. Vicidomini, "Fourier ring correlation simplifies image restoration in fluorescence microscopy," *Nature communications*, vol. 10, no. 1, pp. 1–9, 2019. 1, 2

[15] "Visual object classes challenge 2012," http://host.robots.ox.ac.uk/pascal/VOC/voc2012/, accessed: 2020-12-10. 2

[16] 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. 2

[17] M. Van Heel and M. Schatz, "Fourier shell correlation threshold criteria," *Journal of structural biology*, vol. 151, no. 3, pp. 250–262, 2005. 2

[18] N. Banterle, K. H. Bui, E. A. Lemke, and M. Beck, "Fourier ring correlation as a resolution criterion for super-resolution microscopy," *Journal of structural biology*, vol. 183, no. 3, pp. 363–367, 2013. 2

[19] G. Harauz and M. van Heel, "Exact filters for general geometry three dimensional reconstruction," *Optik (Stuttgart)*, vol. 73, no. 4, pp. 146–156, 1986. 2

[20] J. Vila-Comamala, Y. Pan, J. Lombardo, W. M. Harris, W. Chiu, C. David, and Y. Wang, "Zone-doubled fresnel zone plates for high-resolution hard x-ray full-field transmission microscopy," *Journal of Synchrotron Radiation*, vol. 19, no. 5, pp. 705–709, 2012. 2

[21] "Automatic differentiation package - torch.autograd¶." [Online]. Available: https://pytorch.org/docs/stable/autograd.html 3