# Photo-Realistic Continuous Image Super-Resolution with Implicit Neural Networks and Generative Adversarial Networks

Muhammad Sarmad*[1], Leonardo Ruspini[2], and Frank Lindseth[1]

[1]Norwegian University of Science and Technology, Trondheim, Norway
[2]Petricore, Norway

## Abstract

The implicit neural networks (INNs) can represent images in the continuous domain. They consume raw (X, Y) coordinates and output a color value. Therefore they can represent and generate images at arbitrarily high resolutions in contrast to convolutional neural networks (CNNs) that output a constant-sized array of pixels. In this work, we show how to super-resolve a single image using an INN to produce sharp and photo-realistic images. We employ a random patch-based coordinate sampling method to obtain patches with context and structure; we use these patches to train the INN in an adversarial setting. We demonstrate that the trained network retains the desirable properties of INNs while the output is sharper compared to previous work. We also show qualitative and quantitative comparisons with INN and CNN baselines on benchmark datasets of DIV2K, Set5, Set14, Urban100, and B100. Our code will be made public at `https://github.com/iSarmad/CiSRGan`.

## 1 Introduction

Image enhancement and super-resolution find applications in various consumer products such as smartphone photography, TV and video, etc. The advent of deep learning and neural networks has enabled advancements in single-image super-resolution (SISR). Convolutional neural networks (CNNs) are the most popular method for SISR [11]. However, the output of CNNs is an array of pixels with a fixed size. Therefore, we need to train a new network for different scaling factors. This strategy can be very inconvenient and time-consuming.

Recently a class of neural networks called implicit neural networks (INNs) has gained attention [33, 25, 28]. These networks can represent an image by storing the color value of each pixel corresponding to a given pixel coordinate [26, 31]. This image representation leads to a continuous model where one can zoom in to a single image arbitrarily by changing the discretization level of the input coordinates.

Chen et al. [8] proposed an INN based method called local implicit image function (LIIF) for SISR. They used a single INN to perform SISR for any scale and achieved arbitrary zooming capability i.e. given a neural network that was trained for scales in the range of 1x to 4x (*we refer to this range as in-scale*), their model can perform super-resolution on 6x and 8x etc (*out-of-scale*). This ability to extrapolate makes LIIF very beneficial for super-resolution. Furthermore, LIIF is on par with CNNs in terms of distortion metrics such as the PSNR [22]. Despite these advantages, LIIF suffers from blurry outputs for *out-of-scale* super-resolution due to the use of pixel-wise loss function. In this work, we propose continuous image super-resolution generative adversarial network (CiSR-GAN) that trains INNs in an adversarial setting for super-resolution, thus improving the perceptual quality and photo-realism of output for out-of-scale SISR. To the best of our knowledge, training implicit network for the task of out-of-scale single image super-resolution in an adversarial setting has not been proposed before.

We compare our method with previous state of the art in INN and CNN based super-resolution methods.

*Corresponding Author: muhammad.sarmad@ntnu.no



https://doi.org/10.7557/18.6285

## 2 Related Works

**Convolutional Neural Network based SISR** Before convolutional neural networks (CNNs) [18, 19, 13], handcrafted algorithms were used to perform single image super-resolution (SISR); e.g., Yang et al. [39] used sparse coding to solve this task. Recently, SISR using CNN has become main stream [20, 27, 23, 37]. SISR can be divided into algorithms that either focus on lowering distortion or improving perceptual quality [6]. Our work focuses on improving the perceptual quality.

**Implicit Neural Networks for SISR** Implicit neural networks (INNs) have recently become popular as a way to represent continuous images and shapes [26, 38, 4, 9, 3, 10]. Occupancy Networks [25] and Deep SDF [28] used INNs for 3D shape representation. Then Sitzman et al. [31], and Tancik et al.[34] showed that the INNs could also be used to represent images with high fidelity. Later works learned GANs using INNs [7, 32, 30, 2]. Local implicit image function (LIIF) [8] recently showed that continuous representation could also be used to perform SISR. The resulting SISR model is agnostic to resolution, and a single model can be used to super-resolve images to any required resolution. LIIF [8] uses the $L_1$ loss to train the network, which renders the output blurry. However, we train our model in the adversarial setting to perform photo-realistic SISR and achieve a better result.

## 3 Method

Consider a low-resolution 2D Image $I_{\downarrow s}$ that consists of arrays of pixels. The high resolution 2D image corresponding to $I_{\downarrow s}$ is given as $I \rightarrow I(x, y) \in \mathbb{R}^{X \times Y}$. Where $I_{\downarrow s}(x, y) \in \mathbb{R}^{\frac{X \times Y}{s}}$, and $s$ is the scaling factor. Each pixel in $I$ has coordinates x and y. Let's assume that a continuous image can be represented by a function $f_\theta$. Then the discrete image $I$ can be represented as:

$$I = f_\theta(c, z), \tag{1}$$

$z$ is the latent vector of the features of low-resolution image $I_{\downarrow s}$. Note that $c = x_{hr} - v$, $x_{hr}$ are the pixel coordinates of image $I$ and $v$ are the coordinates of the feature vector $z$ in the image domain. In this work, $f_\theta$ is the implicit neural (INN).

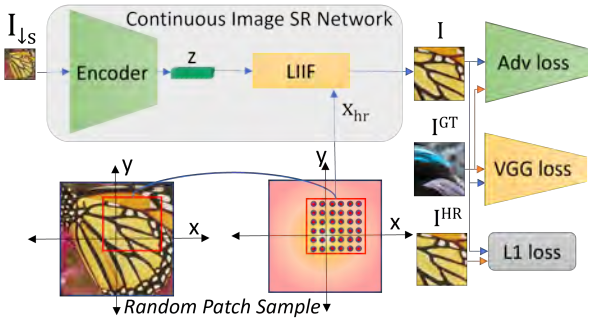


Figure 1: **Training Method:** The low-resolution image $I_{\downarrow s}$ is passed through CNN encoder to get feature vector $z$. A random patch is selected from the coordinate space of desired high resolution image to obtain high resolution coordinates $x_{hr}$. $z$ and $x_{hr}$ are passed through Local implicit function image (LIIF) generator to obtain the super-resolved output image $I$. This $I$ is compared with $I^{GT}$ using adversarial loss ('Adv loss'), perceptual loss ('VGG loss') and with $I^{HR}$ using pixel loss $L_1$.

More specifically, for $f_\theta$ we employ the local implicit image function (LIIF) with default configurations. For details, we refer to the paper [8].

**Training LIIF in an Adversarial Setting** An overview of our approach is shown in Figure. 1. The input image is passed through a convolutional encoder to obtain a latent vector $z$. This latent vector $z$ and the image $I$ coordinates $x_{hr}$ are used to obtain the color values of the pixels at input coordinates $x_{hr}$ using LIIF block [8]. Note that the INN consists of a few multilayer perceptron (MLP) layers that are present inside the LIIF block. We need an output image patch to train the INN using adversarial and perceptual loss. The previous method [8] uses a random set of coordinates from the image. This sampling method works well when the objective is to minimize the pixel-wise loss, e.g., $L_1$. However, looking at only pixels means the contextual information is lost. Therefore, we propose a random patch-based sampling procedure instead of a random point-based sampling method to retain contextual information. We first train LIIF [8] with random patches instead of random points with only a pixel-wise loss. We notice that this random patch-based sampling method performs similar to a random coordinate-based sampling method

in terms of performance.

We use the $L_1$ loss following previous work [8], which trains with only the $L_1$ objective leading to smooth images which blur the textural information for *out-of-scale* super-resolution.

The use of a patch-based sampling procedure permits the use of adversarial loss that is based on generative adversarial network (GAN) [12]. The GAN consists of a generator and a discriminator that compete against each other. The goal of the generator is to generate realistic images, whereas the goal of the discriminator is to get good at classifying generated images as fake. In this joint training, both get better, resulting in realistic image generation. However, instead of using a standard GAN formulation, we use a relativistic GAN formulation instead [16]. This formulation is different from the standard discriminator, which estimates the probability that an input image is real. Instead, the discriminator predicts the probability that a real image is relatively more realistic than a fake one. We define a discriminator network $D_{\theta_D}$, which is optimized in an alternating manner along with generator network $G_{\theta_G}$ to solve the adversarial min-max problem. The relativistic GAN solves the following min-max problem:

$$\min_{\theta_G}\max_{\theta_D}\ \mathbb{E}_X[\log D_{\theta_D}(I^{GT}, G_{\theta_G}(I_{\downarrow s}))]+ \mathbb{E}_X[\log(1 - D_{\theta_D}(G_{\theta_G}(I_{\downarrow s}), I^{GT}))] \tag{2}$$

Note that, $X = (I^{GT}, I_{\downarrow s}) \sim (p_{\text{train}}(I^{GT}), p_G(I_{\downarrow s}))$ and $D_{\theta_D}(I^{GT}, G_{\theta_G}(I_{\downarrow s})) = \sigma(\mathcal{C}(I^{GT}) - \mathbb{E}_{G_\theta(I_{\downarrow s})}[\mathcal{C}(G_{\theta_G}(I_{\downarrow s}))])$ . Where $\mathbb{E}_{G_\theta(I_{\downarrow s})}[.]$ is mean over the generated data in the mini-batch. $\sigma$ is the sigmoid activation function and $\mathcal{C}$ is the output of discriminator before the activation function. For details, we refer to [16].

We also use the perceptual loss that is the distance between the features of a pre-trained VGG network between the predicted image $I$ and the ground-truth image $I^{GT}$ [15]. The complete training objective for the generator is as follows:

$$\mathcal{L}_t = \lambda_1\mathcal{L}_1 + \lambda_2\mathcal{L}_G + \lambda_3\mathcal{L}_{VGG} \tag{3}$$

Where $\mathcal{L}_1, \mathcal{L}_G$ and $\mathcal{L}_{VGG}$ are the content, adversarial and perceptual losses respectively. The $\lambda_1$, $\lambda_2$ and $\lambda_3$ are weighting hyperparameters terms for each of the objectives respectively. We set them following guidelines from previous work [37].

# 4 Experiments

We employed Pytorch for the implementation of all our models [29]. We trained all the networks on an NVIDIA RTX Titan GPU. The code is built on the open-source implementations [8, 35].

**Dataset and Metrics** Like [8], we use the DIV2K dataset with standard split for training and validation [1] for fair comparison. Testing is performed on multiple test datasets including Set5, Set14, Urban100 and B100 [5, 40, 14, 24]. The results for the related works were generated for comparison using pre-trained models provided by Chen et al. [8], and SPSR [23]. We use peak signal-to-noise ration (PSNR) as a metric for comparison. PSNR (measured in dB) is a measure of quality between super-resolved image and ground truth. Even though it is a good measure of distortion, however, it is a poor indicator of perceptual quality [6]. Therefore we additionally report perceptual similarity metric (LPIPS) [41] for comparison with previous works. LPIPS measures the distance in VGG [15] feature space between the super-resolved and the ground-truth image. The lower the distance, the more perceptually similar the super-resolved image is to the ground truth.

**Training Details** Similar to LIIF [8], we use RDN [42] as the encoder, where a feature map $z$ is generated with the same size as the input image. The INN $f_\theta$ is a 5-layer MLP with ReLU activation and hidden dimensions of 256. Encoder and INN act as the generator in our model. The discriminator is based on the architecture used by ESRGAN [37]. We use input patches of 64 x 64 during training. The generator's output is the same as the input patch size, i.e., 64 x 64; therefore, the discriminator is adjusted to cater to an image patch of this size. We use transfer learning and initialize the weights of our generator from a pre-trained RDN-LIIF [8]. We train all models for 75 epochs with batch size 16 on the DIV2K training set. We utilize the Adam [17] optimizer for both generator and discriminator with a learning rate of $1^{-4}$. The weights for $\lambda_1$, $\lambda_2$ and $\lambda_3$ are set to $1^{-2}$, $5^{-3}$ and 1 [37]. For a fair comparison with LIIF, we also train the models from the 1x-4x scale range.

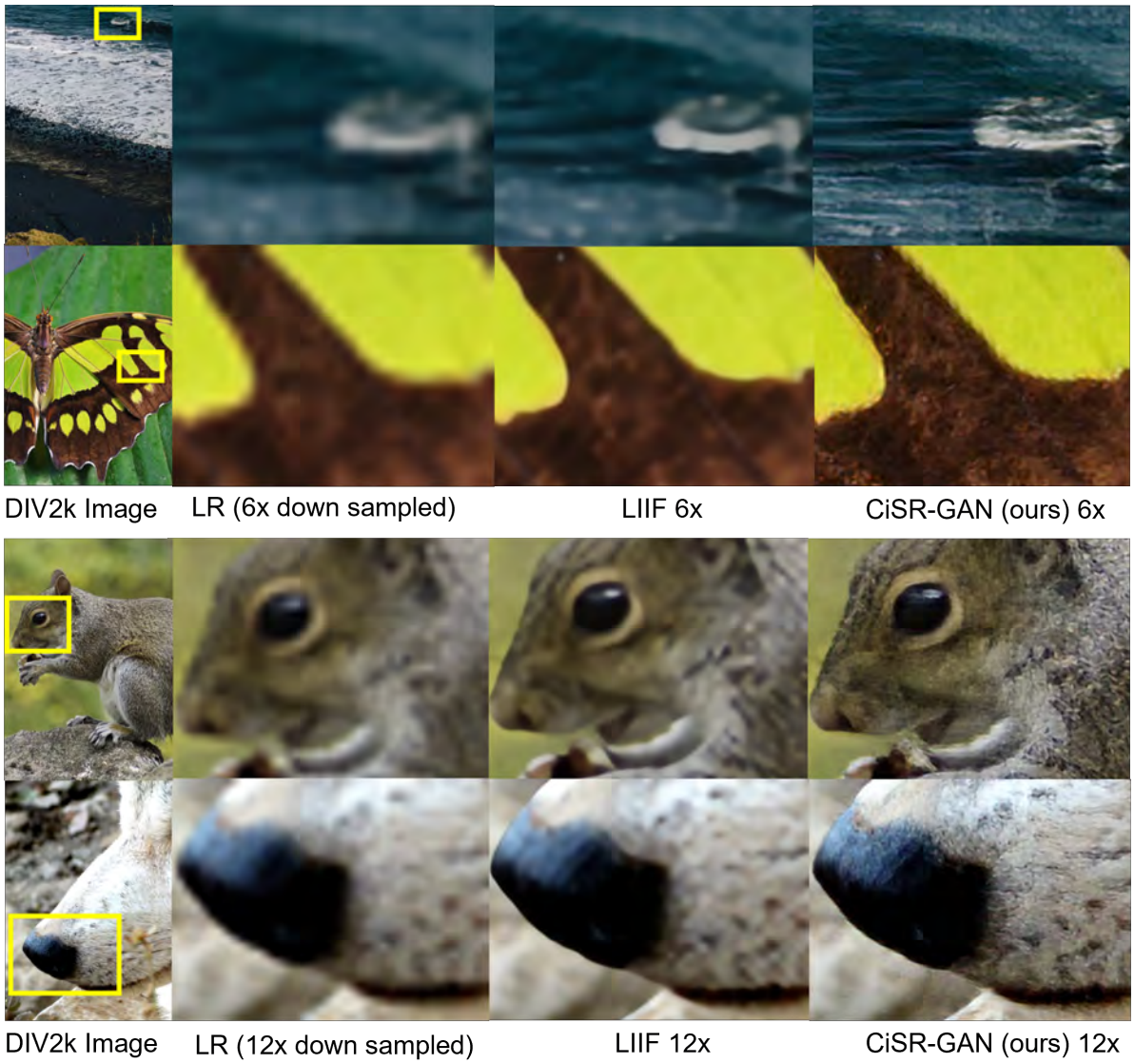


Figure 2: **Out-of-Scale Qualitative Comparison on DIV2K:**. This figure shows the reference image from DIV2k, the low-resolution input image (LR), super-resolved image using LIIF [8] and finally our model's output (CiSR-GAN). LR images are 6x and 12x down-sampled from ground-truth HR images and super-resolved to 6x and 12x in the top 2 and bottom 2 rows respectively demonstrating out-of-scale performance. All models were trained for 1x-4x only therefore we refer to 6x and 12x as *out-of-scale*. From the images we can see that LIIF has a smoothing effect where it blurs out the high-level detail in the images. Comparatively, our models clearly produces sharper results retaining textural details like waves of water, texture in butterfly wings and fine hair of animals.

**Qualitative Analysis**

**Out-of-Scale:** The qualitative results on DIV2K validation set [1] and Set14 [40] test set are shown in Figure. 2 and Figure. 3 respectively. The proposed CiSR-GAN produces realistic images containing textures due to the adversarial and perceptual nature of the objective as compared to the LIIF [8]. LIIF's output is always blurry for *out-of-scale* super-resolution smoothing out

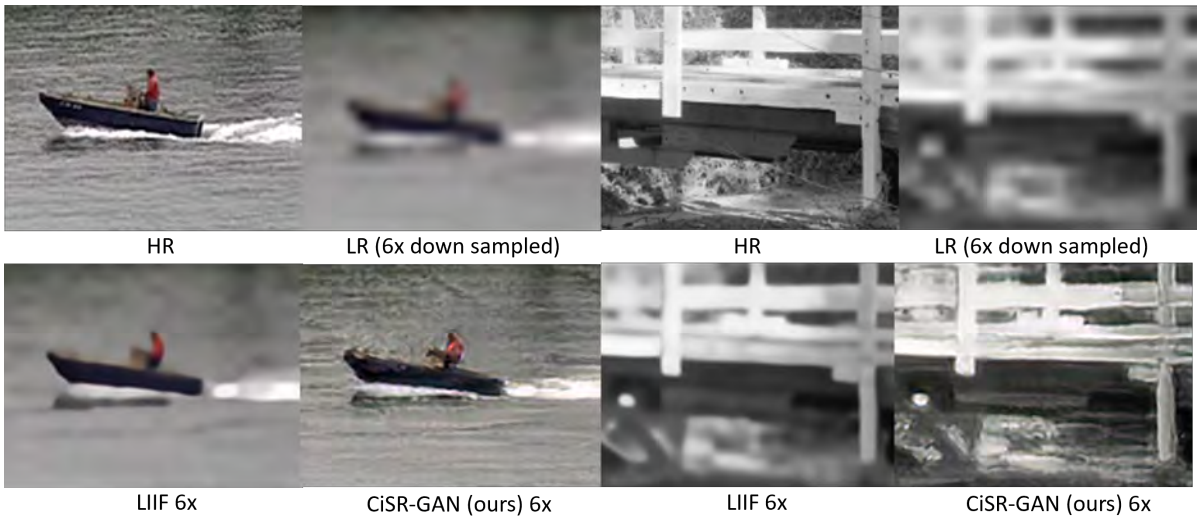


Figure 3: **Out-of-Scale Qualitative Comparison on Set 14:**. This figure shows the high resolution ground truth image (HR), the low-resolution image (LR), super-resolved image using LIIF model [8] and our model's output (CiSR-GAN's). All input images are 6x down-sampled from ground-truth images and super-resolved to 6x. All models were trained for 1x-4x only. We observe the same smoothing effect for LIIF outputs where the high level details such as water waves and texture in the fence has been blurred, while our model retains the high-level details and the image produced is much more realistic than LIIF.

the textural information. At the same time, we also maintain all the desired properties of an implicit network, e.g., a single model can perform super-resolution at higher scales even if the model is not trained for it. All the results presented in the qualitative comparison are for 6x or 12x upsampling to compare with LIIF, whereas we train our models on 1x-4x down-sampled images.

**In-Scale:** Please note that CNN decoder based models [37, 27, 21] are not a direct competitor of our method since they can not perform *out-of-scale* super-resolution. However, we test their performance for *in-scale* super-resolution i.e. for 4x scaling factor for the sake of comprehensiveness. We compare with the best performing recent CNN-based method Structure-Preserving Super Resolution (SPSR) [23], that recently showed great results in retrieving sharp lines and geometry. All images are 4x down-sampled from the ground truth HR images and super-resolved to 4x. The performance is shown in Figure. 4. SPSR model adds edge artifacts like lines or texture to the super-resolved image whereas CiSR-GAN produces more realistic results.

**Quantitative Results**

**CiSR-GAN vs LIIF** We compare our model (CiSR-GAN) with previous work on the DIV2k dataset, as shown in Table. 1. The perceptual similarity metric (LPIPS) is a distance metric; therefore, the lower the value, the better. Whereas the higher the peak signal-to-noise ratio (PSNR), the better. Blau et al. [6] have previously shown that there is a trade-off between distortion and perception, and this can also be observed for our model. CiSR-GAN formulation has lower PSNR values than local implicit image function LIIF [8] as it is trained on the adversarial and perceptual loss. However, it consistently performs better than LIIF in terms of LPIPS metric. Lower LPIPS means that we can expect aesthetically pleasing results from CiSR-GAN. CiSR-GAN can also be evaluated for *out-of-scale* models easily since it is based on an INN. It maintains the edge over LIIF in terms of perceptual metrics for all scales evaluated.

**In-Scale:** We further compare the performance with state-of-the-art methods, including SRGAN, ESRGAN, and SPSR [23, 37, 20]. We notice that CiSR-GAN outperforms all in LPIPS while main-

| Method | Metric | In-Scale | | | Out-of-Scale | | | |
|---|---|---|---|---|---|---|---|---|
| | | ×2 | ×3 | ×4 | ×6 | ×12 | ×24 | ×30 |
| RDN-LIIF [8] | PSNR | **34.99** | **31.26** | **29.27** | **26.99** | **23.89** | **21.31** | **20.59** |
| | LPIPS | 0.0558 | 0.1344 | 0.1947 | 0.2760 | 0.4163 | 0.5506 | 0.5845 |
| CiSR-GAN (ours) | PSNR | 32.01 | 27.95 | 26.30 | 24.27 | 21.67 | 19.52 | 18.92 |
| | LPIPS | **0.0254** | **0.0641** | **0.1016** | **0.1642** | **0.3409** | **0.4839** | **0.5319** |

Table 1: **Distortion vs Perception.** Scaling factor for training is in range ×1–×4. Best values are bold.

| Dataset | Metric | SFTGAN [36] | SRGAN [20] | ESRGAN [37] | SPSR [23] | CiSR-GAN (ours) |
|---|---|---|---|---|---|---|
| **Set5** | LPIPS | 0.0890 | 0.0882 | 0.0748 | 0.0644 | **0.0604** |
| | PSNR | 29.932 | 29.168 | **30.454** | 30.400 | 30.05 |
| **Set14** | LPIPS | 0.4393 | 0.1663 | 0.1329 | 0.1318 | **0.1160** |
| | PSNR | 26.100 | 26.171 | 26.276 | **26.640** | 26.62 |
| **B100** | LPIPS | 0.5249 | 0.1980 | 0.1614 | 0.1611 | **0.1436** |
| | PSNR | 25.961 | 25.459 | 25.317 | 25.505 | **25.72** |
| **Urban100** | LPIPS | 0.4726 | 0.1551 | 0.1229 | 0.1184 | **0.1179** |
| | PSNR | 23.145 | 24.397 | 24.360 | **24.799** | 24.36 |

Table 2: **In-Scale Quantitative comparison with CNNs on benchmark datasets** This table shows CiSR-GAN with other perceptual quality focused methods. Best results are in **bold**. All models have been trained and tested on 4x down-sampled images.

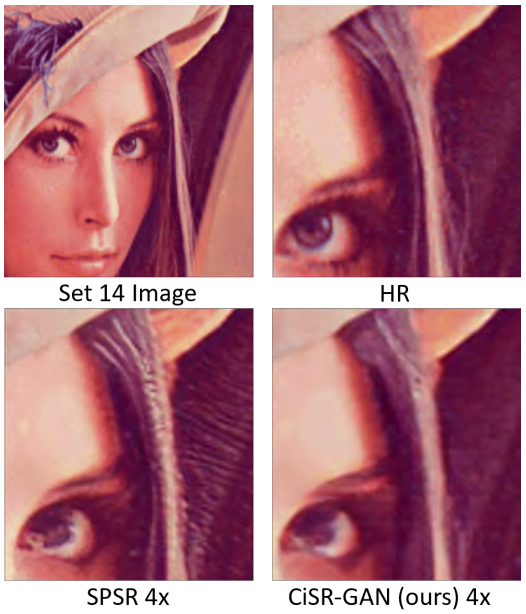


Figure 4: **In-Scale Qualitative Comparison with CNN:**. This figure shows the reference image, the high resolution image (HR), the 4x super-resolved image using Structure-Preserving Super Resolution (SPSR) [23] and our model's output (CiSR-GAN). In the SPSR output, we see lines in the background and artifacts in the eye and the hair whereas CiSR-GAN produces more realistic result.

taining comparable PSNR, as shown in Table. 2. Generally there is large gap between the SPSR and CiSR-GAN based on LPIPS metric, however, the difference is small in the test set Urban100 [14]. This behavior is expected as the gradient guidance based structure priors used in their model encourage the retrieval of lines and geometry that are commonly found in that dataset.

# 5 Conclusion

In this work, we improved the perceptual quality of the implicit neural network based single image super-resolution. The main hindrance in utilizing adversarial losses for continuous image representation models was the random co-ordinate-based sampling procedure adopted by previous works. We proposed to use a patch-based sampling method. Then we trained the implicit neural network with additional objectives based on adversarial and perceptual losses. We demonstrated that the resulting network produces sharp and photo-realistic images while maintaining the desirable properties of the implicit neural networks i.e out-of-scale super-resolution. As future work, our method can also be trained with gradient guidance based structure prior to improve PSNR.

**Acknowledgements.** This work was partially supported by the Norwegian Research Council (grant number 296093) and the members of the SmartRocks joint industry project (ENI AS, Repsol AS, and Chevron Corporation).

# References

[1] E. Agustsson and R. Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017.

[2] I. Anokhin, K. Demochkin, T. Khakhulin, G. Sterkin, V. Lempitsky, and D. Korzhenkov. Image generators with conditionally-independent pixel synthesis. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14273–14282, 2021. DOI: 10.1109/CVPR46437.2021.01405.

[3] M. Atzmon and Y. Lipman. Sal: Sign agnostic learning of shapes from raw data. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2562–2571, 2020. DOI: 10.1109/CVPR42600.2020.00264.

[4] A. Basher, M. Sarmad, and J. Boutellier. Lightsal: Lightweight sign agnostic learning for implicit surface representation. *CoRR*, abs/2103.14273, 2021.

[5] M. Bevilacqua, A. Roumy, C. Guillemot, and M. line Alberi Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In *Proceedings of the British Machine Vision Conference*, pages 135.1–135.10. BMVA Press, 2012. DOI: http://dx.doi.org/10.5244/C.26.135.

[6] Y. Blau and T. Michaeli. The perception-distortion tradeoff. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6228–6237, 2018. DOI: 10.1109/CVPR.2018.00652.

[7] E. R. Chan, M. Monteiro, P. Kellnhofer, J. Wu, and G. Wetzstein. pi-gan: Periodic implicit generative adversarial networks for 3d-aware image synthesis. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5795–5805, 2021. DOI: 10.1109/CVPR46437.2021.00574.

[8] Y. Chen, S. Liu, and X. Wang. Learning continuous image representation with local implicit image function. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8624–8634, 2021. DOI: 10.1109/CVPR46437.2021.00852.

[9] Z. Chen and H. Zhang. Learning implicit fields for generative shape modeling. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5932–5941, 2019. DOI: 10.1109/CVPR.2019.00609.

[10] J. Chibane, M. A. mir, and G. Pons-Moll. Neural unsigned distance fields for implicit function learning. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 21638–21652. Curran Associates, Inc., 2020.

[11] 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*, 38(2):295–307, 2016. DOI: 10.1109/TPAMI.2015.2439281.

[12] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc., 2014.

[13] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.

[14] J.-B. Huang, A. Singh, and N. Ahuja. Single image super-resolution from transformed self-exemplars. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5197–5206, 2015. DOI: 10.1109/CVPR.2015.7299156.

[15] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016.

[16] A. Jolicoeur-Martineau. The relativistic discriminator: a key element missing from standard GAN. In *International Conference on Learning Representations*, 2019.

[17] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'12, pages 1097–1105, USA, 2012. Curran Associates Inc.

[19] Y. LeCun, B. E. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. E. Hubbard, and L. D. Jackel. Handwritten digit recognition with a back-propagation network. In D. S. Touretzky, editor, *Advances in Neural Information Processing Systems 2*, pages 396–404. Morgan-Kaufmann, 1990.

[20] C. Ledig, L. Theis, F. Huszár, 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 *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 105–114, 2017. DOI: 10.1109/CVPR.2017.19.

[21] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee. Enhanced deep residual networks for single image super-resolution. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1132–1140, 2017. DOI: 10.1109/CVPRW.2017.151.

[22] Y. Lu. The level weighted structural similarity loss: A step away from the mse, 2019.

[23] C. Ma, Y. Rao, Y. Cheng, C. Chen, J. Lu, and J. Zhou. Structure-preserving super resolution with gradient guidance. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7766–7775, 2020. DOI: 10.1109/CVPR42600.2020.00779.

[24] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 2, pages 416–423 vol.2, 2001. DOI: 10.1109/ICCV.2001.937655.

[25] L. Mescheder, M. Oechsle, M. Niemeyer, S. Nowozin, and A. Geiger. Occupancy networks: Learning 3d reconstruction in function space. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4455–4465, 2019. DOI: 10.1109/CVPR.2019.00459.

[26] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *European conference on computer vision*, pages 405–421. Springer, 2020.

[27] K. Nazeri, H. Thasarathan, and M. Ebrahimi. Edge-informed single image super-resolution. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pages 3275–3284, 2019. DOI: 10.1109/ICCVW.2019.00409.

[28] J. J. Park, P. Florence, J. Straub, R. Newcombe, and S. Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 165–174, 2019. DOI: 10.1109/CVPR.2019.00025.

[29] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer. Automatic differentiation in pytorch. In *NIPS-W*, 2017.

[30] K. Schwarz, Y. Liao, M. Niemeyer, and A. Geiger. Graf: Generative radiance fields for 3d-aware image synthesis. In H. Larochelle,

M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 20154–20166. Curran Associates, Inc., 2020.

[31] V. Sitzmann, J. Martel, A. Bergman, D. Lindell, and G. Wetzstein. Implicit neural representations with periodic activation functions. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 7462–7473. Curran Associates, Inc., 2020.

[32] I. Skorokhodov, S. Ignatyev, and M. Elhoseiny. Adversarial generation of continuous images. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10748–10759, 2021. DOI: 10.1109/CVPR46437.2021.01061.

[33] K. O. Stanley. Compositional pattern producing networks: A novel abstraction of development. *Genetic programming and evolvable machines*, 8(2):131–162, 2007.

[34] M. Tancik, P. P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singhal, R. Ramamoorthi, J. T. Barron, and R. Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *arXiv preprint arXiv:2006.10739*, 2020.

[35] X. Wang, K. Yu, K. C. Chan, C. Dong, and C. C. Loy. Basicsr, 2020.

[36] X. Wang, K. Yu, C. Dong, and C. Change Loy. Recovering realistic texture in image super-resolution by deep spatial feature transform. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 606–615, 2018. DOI: 10.1109/CVPR.2018.00070.

[37] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In L. Leal-Taixé and S. Roth, editors, *Computer Vision – ECCV 2018 Workshops*, pages 63–79, Cham, 2019. Springer International Publishing.

[38] X. Xu, Z. Wang, and H. Shi. Ultrasr: Spatial encoding is a missing key for implicit image function-based arbitrary-scale super-resolution. *CoRR*, abs/2103.12716, 2021.

[39] J. Yang, J. Wright, T. S. Huang, and Y. Ma. Image super-resolution via sparse representation. *IEEE Transactions on Image Processing*, 19(11):2861–2873, 2010. DOI: 10.1109/TIP.2010.2050625.

[40] R. Zeyde, M. Elad, and M. Protter. On single image scale-up using sparse-representations. In J.-D. Boissonnat, P. Chenin, A. Cohen, C. Gout, T. Lyche, M.-L. Mazure, and L. Schumaker, editors, *Curves and Surfaces*, pages 711–730, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[41] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 586–595, 2018. DOI: 10.1109/CVPR.2018.00068.

[42] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu. Residual dense network for image super-resolution. *CoRR*, abs/1802.08797, 2018.