

Advancing Arbitrary-Scale Image Super-Resolution: Introducing Residual In Residual Dense Networks with a Novel Local Implicit Image Function

Yi-Yu Chen ^{*}, Yi-Leh Wu ^{*}

Abstract

The progress in image super-resolution has seen significant advancements due to the emergence of deep convolutional neural networks. While most researchers concentrate on training models for specific scales, only a few delve into creating models adaptable to various scales. This study builds upon prior research that focuses on achieving arbitrary resolution with a single model. The model under consideration employs an auto-encoder structure. The encoder extracts feature maps from the input image, while the decoder reconstructs these feature maps to the resolution specified by the user. Referred to as RRDN-NLIIF (Residual in Residual Dense Networks with Novel Local Implicit Image Function), our experimental results demonstrate its superior performance over the benchmark model in terms of the PSNR metric.

Keywords: Arbitrary-scale Super Resolution, Convolutional Neural Networks, Implicit Neural Representation, Positional Encoding.

1 Introduction

Super resolution involves enhancing the resolution of a low-resolution image by up-sampling it to a higher resolution. The advent of deep convolutional neural networks has led to the introduction of numerous models, all demonstrating outstanding performance.

Dong et al. proposed SRCNN[11]. It was the first model that used deep learning skills in super resolution territory. This model had a simple structure, with merely three convolutional layers.

Kim et al. found that with extremely high learning rates and residual structure, they can train a deeper network and perform better. They proposed VDSR [12]. Later, they also discovered the importance of recursive-supervision and brought up a new network DRCN.[13].

^{*} National Taiwan University of Science and Technology, Taiwan.

Typically, researchers focus on training individual models for specific scales. Although these models deliver impressive results, their lack of flexibility is notable. To obtain images at various scales, significant time investment is required for training each specific model.

Hu et al. introduced the initial arbitrary scale up-sampling model, MetaSR[1], which exhibited excellent performance on images within the training distribution. However, its effectiveness diminished notably when generating large-scale images beyond the training distribution, as indicated by experiments in [2].

Building upon this foundation and drawing inspiration from recent advancements in 3D reconstruction utilizing implicit neural representation [3-8], Chen et al. put forth a novel arbitrary scale up-sampling model. They introduced the concept of Local Implicit Image Function (LIIF)[2] as a key element in their approach.

In LIIF representation (Equation 1), each continuous image $I^{(l)}$ was represented as a 2D feature map $M^{(l)} \in \mathbb{R}^{H \times W \times D}$. z was the feature factors, and $x \in X$ was a 2D coordinate used to query the predicted RGB signal $s \in S$ in the continuous image domain. Their final model RDN-LIIF showed a splendid performance, and became the state-of-the-art (SOTA) work in 2021.

$$s = f_{\theta}(z, x) \quad (1)$$

Upon scrutinizing the model RDN-LIIF proposed by Chen et al.[2], we identified areas for potential improvement. In this study, we enhanced the feature extraction model by substituting RDN[9] with a more robust alternative capable of capturing finer details. Additionally, we bolstered the up-sampling model to effectively leverage these detailed features. Experimental results demonstrate that our modified model surpasses the original one in terms of the PSNR (dB) index. The contributions made in this work are outlined below.

1. We replace the original feature extraction model RDN with the Residual in Residual Dense Block Network (RRDN)[10], as proposed by Wang et al.
2. Drawing inspiration from techniques employed by Xu et al.[11], we enhance the up-sampling module to create a more potent method, termed NLIIF.

The rest of this paper is organized into the following sections: Section 2 presents a literature review on image super-resolution. Section 3 describes our proposed model. Section 4 showcases experimental results and discusses encountered challenges. Finally, Section 5 provides conclusions and explores potential future work.

2 Related Work

2.1 Single scale super-resolution

Dong et al. introduced SRCNN [11], the first model to apply deep learning techniques to the field of super-resolution. This model had a simple architecture consisting of only three convolutional layers. Kim et al. later demonstrated that using extremely high learning rates and a residual structure allowed for training deeper networks with improved performance, leading to the

development of VDSR [12]. They further recognized the significance of recursive supervision and proposed a new network, DRCN [13].

Some researchers have approached the super-resolution (SR) problem using adversarial methods. Ledig et al. were the first to introduce SRGAN [14], a model distinguished not only by its adversarial framework but also by its choice of loss function. While models trained with L2 loss tend to achieve high PSNR scores, they often lose fine high-frequency details during reconstruction. To address concerns about the limitations of L2 loss, Ledig et al. incorporated perceptual loss and adversarial loss into their model. Their experiments showed that for large upscaling factors (4x), their model produced images with higher perceptual fidelity, even though its PSNR scores were lower. This shift in focus led to greater recognition of perceptual quality in SR research. Following the 2018 PIRM-SR challenge [15], the field diverged into two approaches—one prioritizing improvements in PSNR/SSIM scores and the other using the perceptual index (PI) as a benchmark. In this paper, we adopt PSNR as our evaluation standard.

Building on SRGAN [14], Lim et al. refined the architecture by eliminating redundant modules, resulting in a larger and more efficient network called EDSR [16]. A key innovation of their work was the removal of batch normalization (BN) layers. While BN layers typically accelerate convergence and help prevent overfitting, they were found to degrade the contrast information in training images, ultimately hindering performance in super-resolution tasks. Lim et al.’s experiments demonstrated that their model achieved higher PSNR scores compared to previous approaches.

Around the same time, Zhang et al. introduced RDN [17], a model that integrated the strengths of both residual blocks and dense connection blocks, significantly enhancing super-resolution quality. This architecture later became a standard framework for various arbitrary-scale SR models.

2.2 Arbitrary-scale super-resolution

MDSR [16], introduced by Lim et al., was the first model capable of generating images at multiple resolutions. However, since it was limited to producing only three specific scales, its classification as an arbitrary-scale SR model is debatable.

Hu et al. introduced a fully arbitrary-scale super-resolution model consisting of two main components: a feature extraction module and an upscaling module. They selected RDN [17] as the optimal feature extractor and incorporated MetaSR, an upscaling module leveraging meta-learning techniques. As a result, their model was aptly named Meta-RDN.

Similar to Meta-RDN, Chen et al. introduced an autoencoder-based structure, where a feature extraction module served as the encoder and a feature reconstruction module acted as the decoder. They also utilized RDN [17], as Hu et al. did, to extract image features. However, Chen et al. designed a different decoder, LIIF, which bridges discrete images and continuous representations. This functionality was achieved using a multi-layer perceptron (MLP). Their experiments demonstrated impressive PSNR scores and exceptional performance, enabling upscaling factors of up to $\times 30$.

2.3 Positional encoding in super-resolution

Initially, positional encoding was simply a technique used in transformer architectures [18] to embed positional information into discrete vectors. However, Mildenhall et al. expanded its application in NeRF [19], utilizing positional encoding to map coordinate information into a higher-dimensional space. This allowed the MLP structure in their model to capture high-frequency

details more effectively, resulting in more detailed and less distorted 3D reconstructions. Following their approach, many researchers [20-23] adopted similar methods and achieved promising results.

Recognizing the link between 3D reconstruction and super-resolution tasks, Xu et al. drew inspiration from LIIF [2] and incorporated the positional encoding technique into their SR model. They introduced UltraSR [10], which, according to their experiments, outperformed LIIF [2].

3 Proposed Model

Our proposed model, illustrated in Figure 1, is named RRDN-NLIIF, an abbreviation for Residual in Residual Dense Networks with Novel Local Implicit Image Function. In comparison to the baseline model, our model exhibits two distinct modifications: the alteration of the encoder and the enhancement of the decoder.

3.1 Encoder

We employ RRDN[10], introduced by Wang et al., as our encoder, and the structural overview is depicted in Figure 2. RRDN serves as a feature extraction module within ESGAN[10].

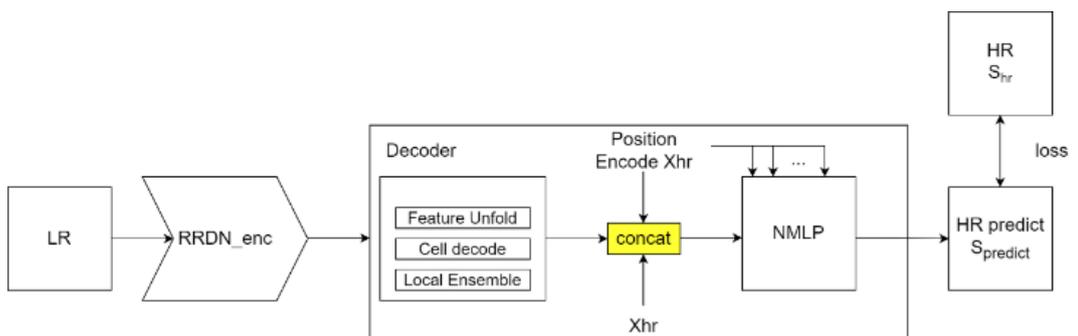


Figure 1: Residual in Residual Dense Networks with Novel Local Implicit Image Function.

ESGAN, an enhanced model derived from SRGAN[12], emerged victorious in the 2018 PIRM-SR[13] challenge. Both models share the concept of utilizing GAN for solving the SR problem and employing perceptual loss to enhance training outcomes. While ESGAN demonstrated remarkable performance in terms of perceptual indices (PI), it exhibited relatively lower PSNR scores compared to other models.

Our primary focus is on improving PSNR, and research akin to ESGAN, which emphasizes PI, may seem unrelated. However, we posit that the feature extraction module is a universal component. This implies that a feature extraction module effective in a PI-based model can also enhance the performance of a PSNR-based model. Section 3 will provide experimental results validating our assumption.

RRDN bears a resemblance to RDN, leveraging the capabilities of residual and dense blocks. Starting with the initial convolutional layer, a feature F_0 undergoes 23 RRDBs (details of RRDBs

in Section 2.2) to yield F_{23} . After a convolutional layer, F_{23} is combined with F_0 through residual learning, resulting in the final feature for reconstruction.

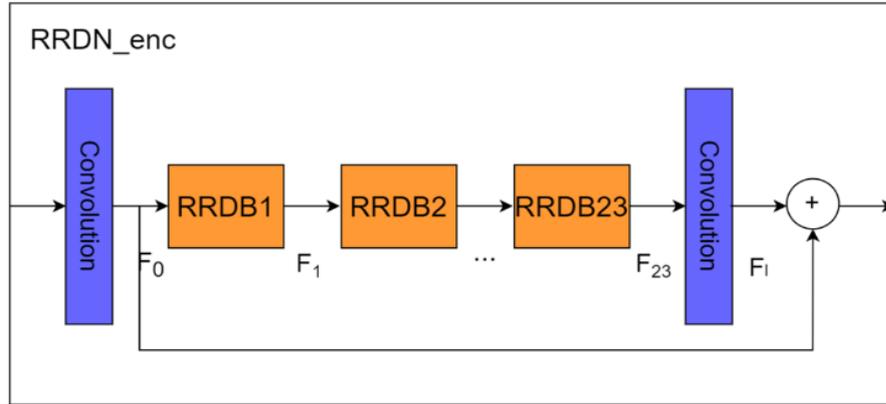


Figure 2: Residual in Residual Dense Networks Encoder

3.2 RRDB and DB

Figure 3 illustrates the processes within the Residual in Residual Dense Block (RRDB). The RRDB comprises three Dense Blocks (DBs). The output of a DB is scaled by a factor of 0.2 and added to its input for residual learning, subsequently becoming the input for the next DB in the sequence.

As depicted in Figure 4, each DB contains the smallest unit, a Convolutional Block (CB). Similar to the Residual Dense Block (RDB) in RDN[9], a CB consists of a convolutional layer and a leaky ReLU activation layer. Additionally, akin to the RDB, the output of a CB serves as the input for the next CB in sequence and potentially for subsequent CBs. The features generated from all CBs undergo dimensional reduction, are scaled by a factor of 0.2, and are then added to the original input for residual learning.

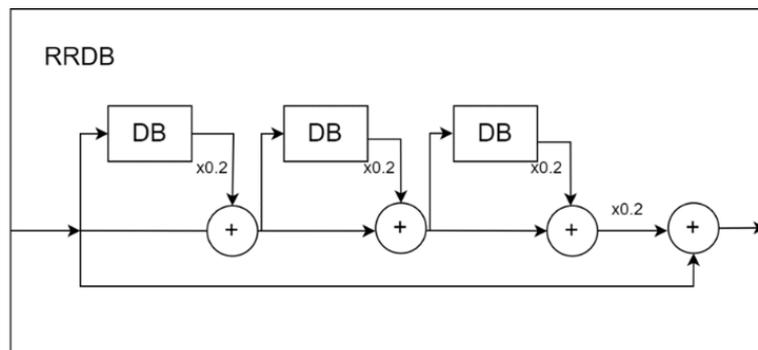


Figure 3: Residual in Residual Dense Block

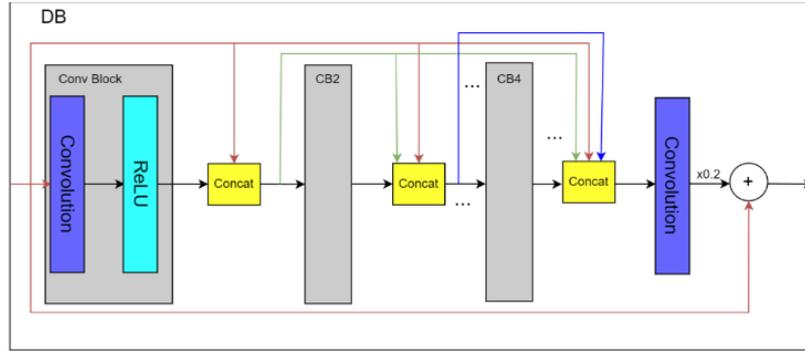


Figure 4: Dense Block

3.3 Decoder

Xu et al. introduced UltraSR[11] as an arbitrary-scale super-resolution model, representing an enhancement of RDN-LIIF. In this improved version, they integrated the entire decoder while incorporating additional operations such as residual learning and positional encoding.

In our approach, we draw inspiration from UltraSR and construct our version of the decoder. The initial modification involves making the positional encoded coordinate the input of a Multi-Layer Perceptron (MLP) and each linear layer within the MLP (refer to Figure 1). The implementation details are outlined in Equation 2.

$$s = f_{\theta}(z, [x, c], \psi(x)) \quad (2)$$

$$\psi(x) = (\sin(\omega_1 x_h), \cos(\omega_1 x_w) \dots \sin(\omega_{48} x_h), \cos(\omega_{48} x_w)) \quad (3)$$

$\psi(x)$ denotes the positional encode coordinates. Frequency parameters $\omega_1, \omega_2 \dots \omega_{48}$ are set to $2e^n$, $n \in 1, 2 \dots 48$. x is divided to two parts according to the height and width of the given resolution, separately x_h and x_w .

Compared to the MLP from RDN-LIIF, our improved MLP (denoted as NMLP, in short new MLP) reaches double the depth of it. We design a BB (in short of basic block), which consists of two linear layers and two activation layers. Apart from this, we set a residual learning structure for features go through two linear layers. Figure 5 shows what our propose NMLP looks like.

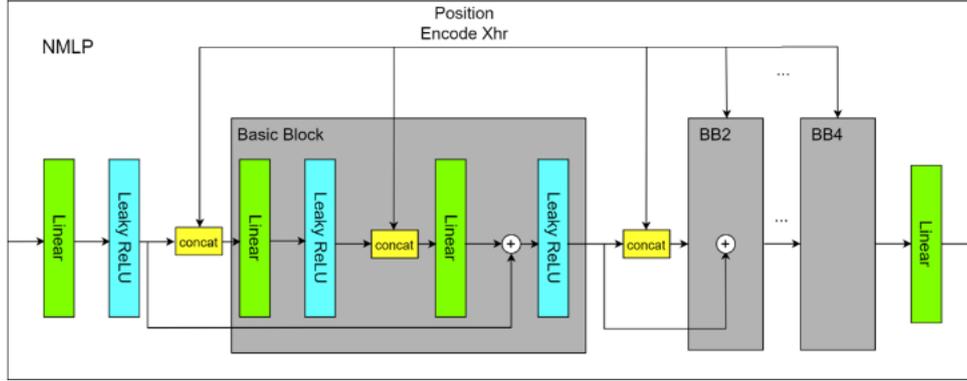


Figure 5: New Multi-Layer Perceptron

4 Experiments

4.1 Set up

We adopt identical training parameters as outlined in [2] to maintain consistency and isolate other variables, showcasing the robustness of our model. The dataset used is the same as [2], namely DIV2K[14], which constituted the dataset for the NTIRE 2017 Challenge[24]. DIV2K comprises 900 images in 2K resolution, along with their $\times 2$, $\times 3$, and $\times 4$ down-sampled counterparts, generated using the default bicubic interpolation setting in the `imresize` function of MATLAB. During training, the first 800 high-resolution (HR) images from DIV2K are utilized, while the last 100 HR images are reserved for testing. Additionally, we employ four standard benchmark datasets, namely SET5[25], SET14[26], B100[27], Urban100[28], CelebAHQ [29], and CelebA [30] to evaluate the performance of our model.

To assess the testing results on DIV2K, we directly compute the Peak Signal-to-Noise Ratio (PSNR) between the output images and the ground-truth images. For testing on benchmark datasets, we calculate the PSNR specifically on the Y channel and apply the border-shaving technique, consistent with [2]. Table 1 provides a summary of the training parameters and environmental settings

Table 1: Training settings

Epoch	1000
Batch size	16
Optimizer	Adam
Loss Function	L1
Initial Learning rate	1×10^{-4}
Learning rate variation	decays by 0.5 every 200 epochs
LR size	48×48

The experiments were conducted on Ubuntu 18.04, utilizing Python 3.7 and CUDA 11.3.

Training of the model was performed using GeForce RTX3090 and RTX4090.

To evaluate the model's performance on scales encountered during training, we employ the corresponding down-sampled datasets provided by DIV2K. For unseen scales, ranging from $\times 6$ to $\times 30$, we crop the high-resolution (HR) image to dimensions divisible by the scale, down-sample it accordingly, and utilize the resulting images as input. The images generated by each model are depicted in Figure 6.

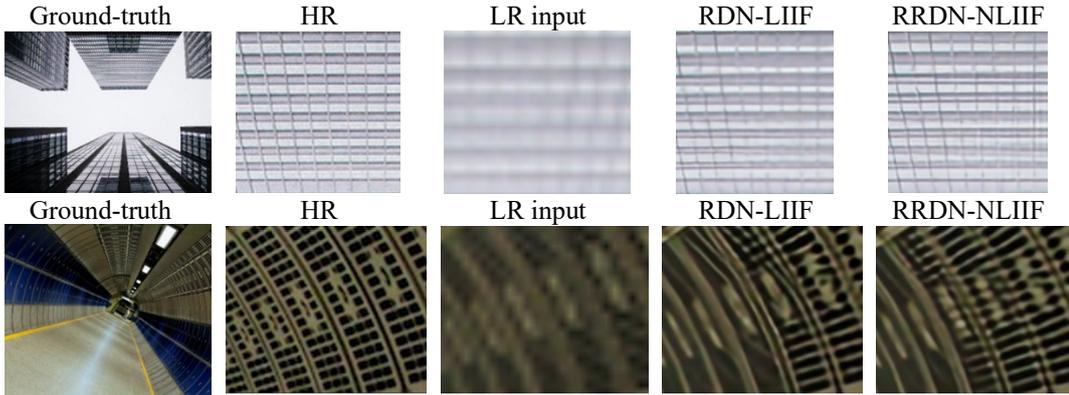


Figure 6: Visual result of image 0845x3.png in DIV2K with 3x amplify and image 78.png in Urban100 with 4x amplify

4.2 Result

Table 4: Comparisons on DIV2K test set.

Method	In-distribution			Out-of-distribution				
	$\times 2$	$\times 3$	$\times 4$	$\times 6$	$\times 12$	$\times 18$	$\times 24$	$\times 30$
Bicubic	31.01	28.22	26.66	24.82	22.27	21.00	20.19	19.59
Meta-RDN	35.00	31.27	29.25	26.88	23.73	22.18	21.17	20.47
RDN-LIIF	34.99	31.26	29.27	26.99	23.89	22.34	21.31	20.59
RRDN-NLIIF	35.04	31.32	29.32	27.04	23.93	22.37	21.33	20.62

Table 5: Comparisons on BENCHMARK datasets

Dataset	Method	In-distribution		Out-of-distribution		
		$\times 2$	$\times 3$	$\times 4$	$\times 6$	$\times 8$
Set5	RDN	38.24	34.71	32.47	-	-
	Meta-RDN	38.22	34.63	32.38	29.04	26.96
	RDN-LIIF	38.17	34.68	32.50	29.15	27.14
	RRDN-NLIIF	38.21	34.73	32.57	29.26	27.28
Set14	RDN	34.01	30.57	28.81	-	-
	Meta-RDN	33.98	30.54	28.78	26.51	24.97
	RDN-LIIF	33.97	30.53	28.80	26.64	25.15
	RRDN-NLIIF	34.02	30.60	28.88	26.66	25.20
B100	RDN	32.34	29.26	27.72	-	-
	Meta-RDN	32.33	29.26	27.71	25.90	24.83

	RDN-LIIF	32.32	29.26	27.74	25.98	24.91
	RRDN-NLIIF (ours)	32.35	29.28	27.76	26.00	24.95
Ur-ban100	RDN	32.89	28.80	26.61	-	-
	Meta-RDN	32.92	28.82	26.55	23.99	22.59
	RDN-LIIF	32.87	28.82	26.68	24.20	22.79
	RRDN-NLIIF (ours)	33.05	28.92	26.77	24.28	22.87

In line with [2], we designate the outcome generated from the last epoch model as our final result. The experimental results presented in Tables 1 and 2 highlight the superior performance of our model compared to other models. Notably, on the DIV2K dataset, the x3 scale exhibits the most substantial improvement, achieving a 0.06dB higher Peak Signal-to-Noise Ratio (PSNR) compared to the original model. Across the four benchmark datasets, excluding B100, our model consistently outperforms the original model by approximately 0.1dB in terms of PSNR. These findings demonstrate that our model not only addresses the challenge of low PSNR scores on in-distribution scales but also maintains strong performance on out-of-distribution scales.

4.3 Ablation Study

For the ablation study, we show that both the substitution of encoder and decoder make contribute to the improvement of PSNR. Therefore, we make two extra models. RRDN-LIIF (Figure 8) is the model that only changes the encoder, and RDN-NLIIF (Figure 9) is the model that only changes the decoder. The comparison of PSNR performance is in Table 3 and Table 4.

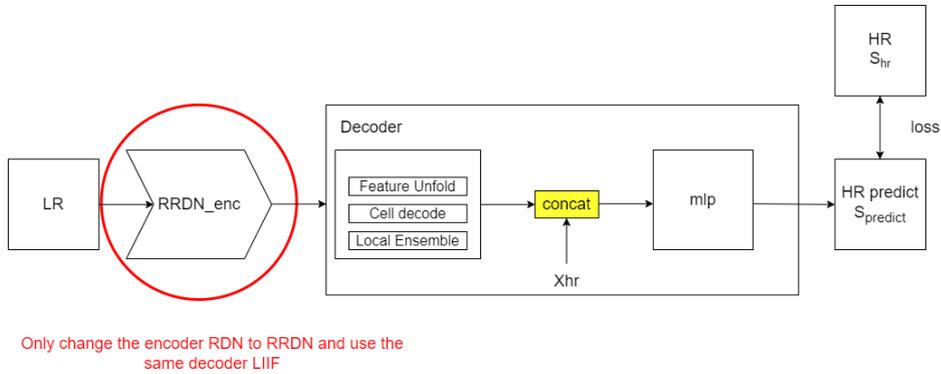


Figure 8: RRDN-LIIF

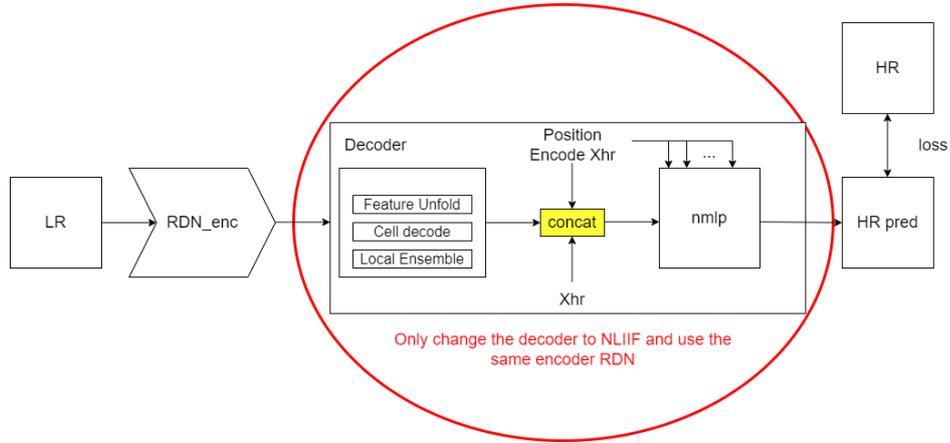


Figure 9: RDN-NLIIF

Table 6: Comparisons on DIV2K test set.

Method	In-distribution			Out-of-distribution				
	×2	×3	×4	×6	×12	×18	×24	×30
RRDN-LIIF	35.03	31.29	29.30	27.01	23.89	22.34	21.30	20.59
RDN-LIIF	35.03	31.30	29.32	27.03	23.93	22.37	21.33	20.62
RRDN-NLIIF	35.04	31.32	29.32	27.04	23.93	22.37	21.33	20.62

Table 7: Comparisons on BENCHMARK datasets

Dataset	Method	In-distribution			Out-of-distribution	
		×2	×3	×4	×6	×8
Set5	RRDN-LIIF	38.22	34.72	32.55	29.28	27.24
	RDN-NLIIF	38.22	34.72	32.54	29.23	27.24
	RRDN-NLIIF	38.21	34.73	32.57	29.26	27.28
Set14	RRDN-LIIF	34.09	30.54	28.88	26.67	25.18
	RDN-LIIF	34.02	30.59	28.88	26.68	25.18
	RRDN-NLIIF	34.02	30.60	28.88	26.66	25.20
B100	RRDN-LIIF	32.34	29.28	27.75	25.99	24.93
	RDN-LIIF	32.34	29.28	27.76	26.00	24.94

	RRDN-NLIIF	32.35	29.28	27.76	26.00	24.95
Urban100	RRDN-LIIF	33.04	28.91	26.77	24.26	22.84
	RDN-NLIIF	32.97	28.82	26.78	24.28	22.89

Based on the test outcomes, it is evident that replacing both the encoder and decoder has a beneficial impact on the model. The proposed final model, RRDN-NLIIF, consistently exhibits superior PSNR performance.

Nevertheless, it was observed that the improvement in PSNR achieved by RRDN-LIIF and RDN-NLIIF did not entirely manifest in our proposed model, RRDN-NLIIF. Further enhancements to the encoder and decoder are imperative to address this discrepancy.

5 Conclusions and Future Work

In this study, we introduce a novel model, RRDN-NLIIF, inspired by a resilient arbitrary-scale super-resolution model developed by Chen et al. [2]. Employing nearly identical parameters and training configurations as the original model, our model achieves the highest Peak Signal-to-Noise Ratio (PSNR) compared to other models across both the DIV2K datasets and the four benchmark datasets at various scales. Our model successfully addresses the limitations of the original model on in-distribution scales, demonstrating robust performance.

However, as observed in Figure 7, there are still areas of concern in the images generated by our model. Additionally, based on our ablation experiments, the PSNR improvements from both the model utilizing the enhanced encoder and the model using only the improved decoder do not fully translate into the performance of our final model. Therefore, further enhancements are necessary.

For future work, we envision two avenues for improvement. First, we consider conducting more fine-tuning on our model to refine its performance. Second, we explore the possibility of identifying a more robust feature extraction structure capable of extracting finer details from images. Notably, the transformer technique has gained popularity in Single Image Super-Resolution (SISR) models, exemplified by HAT [19] proposed by Chen et al. This emerging approach may offer a promising pathway to elevate arbitrary-scale SR models to new heights of performance.

Acknowledgement

This work was partially supported by the Ministry of Science and Technology, Taiwan, under the Grants No. MOST 111-2218-E-011-011-MBK, NSTC 112-2221-E-011-118, and NSTC 113-2221-E-011-116.

References

- [1] X. Hu, H. Mu, X. Zhang, Z. Wang, T. Tan, and J. Sun, "Meta-SR: A magnification-arbitrary network for super-resolution," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 1575-1584. 2019.
- [2] Y. Chen, S. Liu, and X. Wang, "Learning continuous image representation with local implicit image function," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 8628-8638. 2021.
- [3] J. J. Park, P. Florence, J. Straub, R. Newcombe, and S. Lovegrove, "Deepsdf: Learning continuous signed distance functions for shape representation," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 165-174. 2019.
- [4] L. Mescheder, M. Oechsle, M. Niemeyer, S. Nowozin, and A. Geiger, "Occupancy networks: Learning 3d reconstruction in function space," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 4460-4470. 2019.
- [5] Z. Chen and H. Zhang, "Learning implicit fields for generative shape modeling," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5939-5948. 2019.
- [6] S. Saito, Z. Huang, R. Natsume, S. Morishima, A. Kanazawa, and H. Li, "Pifu: Pixel-aligned implicit function for high-resolution clothed human digitization," in Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2304-2314. 2019.
- [7] C. Jiang, A. Sud, A. Makadia, J. Huang, M. Nießner, and T. Funkhouser, "Local implicit grid representations for 3d scenes," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6001-6010. 2020.
- [8] V. Sitzmann, J. Martel, A. Bergman, D. Lindell, and G. Wetzstein, "Implicit neural representations with periodic activation functions," *Advances in Neural Information Processing Systems*, vol. 33, pp. 7462-7473, 2020.
- [9] X. Wang et al., "Esrgan: Enhanced super-resolution generative adversarial networks," in Proceedings of the European conference on computer vision (ECCV) workshops, pp. 0-0. 2018.
- [10] X. Xu, Z. Wang, and H. Shi, "Ultrasr: Spatial encoding is a missing key for implicit image function-based arbitrary-scale super-resolution," *arXiv preprint arXiv:2103.12716*, 2021.
- [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*, vol. 38, no. 2, pp. 295-307, 2015.
- [12] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1646-1654. 2016.
- [13] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," in Proceedings of the IEEE conference on computer vision and pattern

- recognition, pp. 1637-1645. 2016.
- [14] C. Ledig 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, pp. 4681-4690. 2017.
- [15] Y. Blau, R. Mechrez, R. Timofte, T. Michaeli, and L. Zelnik-Manor, "The 2018 PIRM challenge on perceptual image super-resolution," in Proceedings of the European Conference on Computer Vision (ECCV) Workshops, pp. 0-0. 2018.
- [16] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 136-144. 2017.
- [17] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2472-2481. 2018.
- [18] A. Vaswani et al., "Attention is all you need," Advances in neural information processing systems, vol. 30, 2017.
- [19] 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," Communications of the ACM, vol. 65, no. 1, pp. 99-106, 2021.
- [20] L. Liu, J. Gu, K. Zaw Lin, T.-S. Chua, and C. Theobalt, "Neural sparse voxel fields," Advances in Neural Information Processing Systems, vol. 33, pp. 15651-15663, 2020.
- [21] K. Zhang, G. Riegler, N. Snavely, and V. Koltun, "Nerf++: Analyzing and improving neural radiance fields," arXiv preprint arXiv:2010.07492, 2020.
- [22] K. Schwarz, Y. Liao, M. Niemeyer, and A. Geiger, "Graf: Generative radiance fields for 3d-aware image synthesis," Advances in Neural Information Processing Systems, vol. 33, pp. 20154-20166, 2020.
- [23] S. Wizadwongsa, P. Phongthawee, J. Yenphraphai, and S. Suwajanakorn, "Nex: Real-time view synthesis with neural basis expansion," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8534-8543. 2021.
- [24] E. Agustsson and R. Timofte, "NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study," in 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1122-1131, doi: 10.1109/CVPRW.2017.150. 21-26 July 2017 2017.
- [25] M. Bevilacqua, A. Roumy, C. Guillemot, and M.-L. Alberi Morel, "Low-Complexity Single-Image Super-Resolution based on Nonnegative Neighbor Embedding," in British Machine Vision Conference (BMVC), Guildford, Surrey, United Kingdom, <https://hal.inria.fr/hal-00747054/document>
https://hal.inria.fr/hal-00747054/file/bmvc_final_submitted.pdf. [Online]. Available: <https://hal.inria.fr/hal-00747054>. [Online]. Available: <https://hal.inria.fr/hal-00747054>, 2012-

09 2012.

- [26] R. Zeyde, M. Elad, and M. Protter, "On Single Image Scale-Up Using Sparse-Representations. 2010, pp. 711-730.
- [27] 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, vol. 2, pp. 416-423 vol.2, doi: 10.1109/ICCV.2001.937655. 7-14 July 2001 2001.
- [28] 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), pp. 5197-5206, doi: 10.1109/CVPR.2015.7299156. 7-12 June 2015 2015.
- [29] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of gans for improved quality, stability, and variation," arXiv preprint arXiv:1710.10196, 2017.
- [30] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in Proceedings of the IEEE international conference on computer vision, pp. 3730-3738. 2015.
- [31] X. Chen, X. Wang, J. Zhou, and C. Dong, "Activating More Pixels in Image Super-Resolution Transformer," arXiv preprint arXiv:2205.04437, 2022.
- [32] Y. Chen and Y. Wu, "Arbitrary-Scale Image Super Resolution Using Residual in Residual Dense Networks with New Local Implicit Image Function," in Proceedings of the 14th IIAI International Congress on Advanced Applied Informatics, 2023.