

AnySR: Realizing Image Super-Resolution as Any-Scale, Any-Resource

Wengyi Zhan¹, Mingbao Lin², Chia-Wen Lin³, *Fellow, IEEE*, and Rongrong Ji¹, *Senior Member, IEEE*

Abstract—In an effort to improve the efficiency and scalability of single-image super-resolution (SISR) applications, we introduce AnySR, to rebuild existing arbitrary-scale SR methods into any-scale, any-resource implementation. As a contrast to off-the-shelf methods that solve SR tasks across various scales with the same computing costs, our AnySR innovates in: 1) building arbitrary-scale tasks as any-resource implementation, reducing resource requirements for smaller scales without additional parameters; 2) enhancing any-scale performance in a feature-interweaving fashion, inserting scale pairs into features at regular intervals and ensuring correct feature/scale processing. The efficacy of our AnySR is fully demonstrated by rebuilding most existing arbitrary-scale SISR methods and validating on five popular SISR test datasets. The results show that our AnySR implements SISR tasks in a computing-more-efficient fashion, and performs on par with existing arbitrary-scale SISR methods. For the first time, we realize SISR tasks as not only any-scale in literature, but also as any-resource. Our code is available at <https://github.com/CrispyFeSo4/AnySR>.

Index Terms—Super-resolution, any-scale, any-resource.

I. INTRODUCTION

SINGLE image super-resolution (SISR) is the process of reconstructing a low-resolution (LR) image into a high-resolution (HR) one rich in detail. The ill-posed nature has made SISR one of the most challenging tasks in low-level computer vision. The past decades have witnessed the advent of many classic studies that continuously address such a challenge. Typical works such as EDSR [1], RCAN [2], and RDN [3], serving as network backbones, effectively extract both texture and semantic information from LR images and lay the foundation for HR reconstruction. The pixel-shuffle

upsampling method [4] takes a step further to enhance the image reconstruction quality. Standing upon the shoulders of these pioneering works, the majority of SISR methods [1], [2], [3] adhere to a straightforward and efficient paradigm in which features are extracted by feature extractors, followed by upscaling modules for reconstruction.

Albeit the encouraging achievement, most of the above methods are confined to addressing the image reconstruction of a fixed upsampling scale, and are therefore stuck in repeatedly rebuilding an SR network, when the fully trained model cannot accommodate the resolutions in demand. Actually, there is a significant demand for multi-scale SR tasks. Training individual models for respective resolutions seems to be a great waste of computing resources and fails to adapt to real-time online deployment. Reflecting on this, many recent researchers have shifted their focus to integrating any-scale tasks into only one integrated SR model. For example, Meta-SR [9] learns to predict coordinates and scale-oriented convolutional weights to generate HR images. Taking into consideration the neighborhoods in the LR image and a scale-related cell size, the subsequent LIIF [10] utilizes an MLP to predict the pixel values of the queried coordinates in an HR one. To enhance the representational capabilities, LTE [11] further introduces a local texture estimator that transforms coordinate information into the Fourier domain. The very recent SRNO [12] and OPE-SR [13] enable HR tasks of continuous scales by respectively incorporating Neural Operators [14] and introducing an orthogonal position encoding upscale module.

With the fast development of the computer vision community, more stringent requirements on efficient and scalable vision applications have been put forward [15], [16], [17], [18]. Delving into a reflection on most of the existing arbitrary-scale SR methods, we realize that they are not fully excavated to realize the above advocacy. Though succeeding in tackling SR tasks across various scales in one single network, they also cause additional resource overhead, given our empirical observation that lower-resolution scales can be dealt with more easily. To verify this, we have explicitly conducted an extensive literature review and made a summary in Fig. 1, outlining the performance trend *w.r.t.* the SR scale and network complexity. We can find that a network with relatively lower complexity is sufficient to ensure satisfactory performance for SR tasks with smaller reconstruction scales. For example, when performing a $\times 2$ SISR task, ECBSR [5] obtains 33.86 dB PSNR with only 1.38 GFLOPs consumption while EDSR [1] achieves a slightly improved performance of 34.21 dB PSNR by eating up totally 93.89 GFLOPs. However, as the scale increases a lot, a more complex network is

Received 27 May 2024; revised 9 October 2024; accepted 9 November 2024. Date of publication 19 November 2024; date of current version 25 November 2024. This work was supported in part by the National Science and Technology Major Project under Grant 2022ZD0118202; in part by the National Science Fund for Distinguished Young Scholars under Grant 62025603; in part by the National Natural Science Foundation of China under Grant U21B2037, Grant U22B2051, Grant U23A20383, Grant 62176222, Grant 62176223, Grant 62176226, Grant 62072386, Grant 62072387, Grant 62072389, Grant 62002305, and Grant 62272401; and in part by the Natural Science Foundation of Fujian Province of China under Grant 2021J06003 and Grant 2022J06001. The associate editor coordinating the review of this article and approving it for publication was Dr. Xinfeng Zhang. (*Corresponding author: Rongrong Ji.*)

Wengyi Zhan and Rongrong Ji are with the Key Laboratory of Multimedia Trusted Perception and Efficient Computing, Ministry of Education of China, Xiamen University, Xiamen 361005, China (e-mail: zhanwy@stu.xmu.edu.cn; rrji@xmu.edu.cn).

Mingbao Lin is with Skywork AI, Singapore 118222 (e-mail: linmb001@outlook.com).

Chia-Wen Lin is with the Department of Electrical Engineering and the Institute of Communications Engineering, National Tsing Hua University, Hsinchu 30013, Taiwan (e-mail: cwlin@ee.nthu.edu.tw).

Digital Object Identifier 10.1109/TIP.2024.3497807

1941-0042 © 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See <https://www.ieee.org/publications/rights/index.html> for more information.

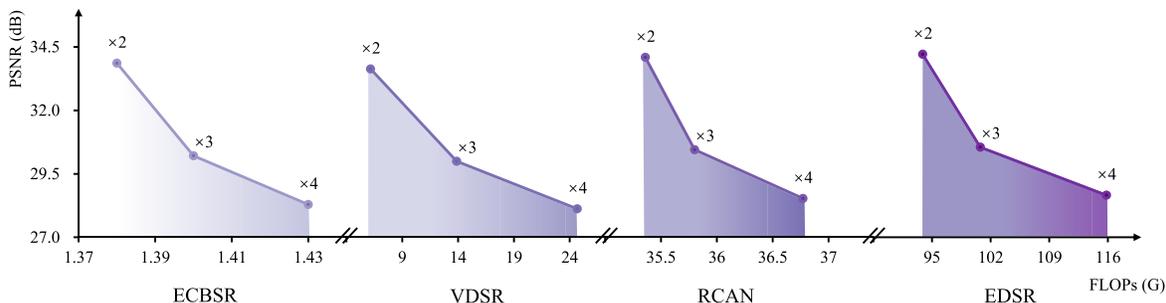


Fig. 1. PSNR performance v.s. computing cost measured as GFLOPs for existing methods of ECBSR [5], VDSR [6], RCAN [2] and EDSR [1]. For fairness, all reported data is borrowed from [7] and the test set is DIV2K [8].

required for the sake of a better performance. For instance, only 28.29 dB PSNR is gained by the method of ECBSR under 1.43 GFLOPs while the result increases to 28.67 dB PSNR by EDSR when more GFLOPs of 115.83 are taken to construct a $\times 4$ SISR task.

As analyzed in the existing research [19], the platforms to perform SISR tasks are often featured with limited and varying computing resources over time, due to the potential resource occupation of other applications. On account of the efficiency and scalability of SR applications, more efforts are actually needed to take the edge off computing expenses. As a consequence, we take a step back in this paper and launch the first attempt to realize image super-resolution not only at any-scale in the literature but also with any-resource, which to our best knowledge by far has never been noticed and resolved.

To this end, we propose AnySR in this paper, a general method with “SR” in the method’s name indicating not only Super-Resolution task but also any-Scale and any-Resource implementation. To be more concrete, we reorganize scale information into multiple sub-groups, each of which consists of similar scale sizes. In this context, we design multiple networks with varying levels of complexity. For cases where computational resources are limited, networks with lower complexity are assigned to handle smaller-scale SISR tasks to ensure inference efficiency, while more complex networks are used for larger-scale SISR tasks to achieve better performance. Further, to circumvent the introduction of cumbersome parameters, we train and infer all networks in a parameter-sharing manner [20], [21] in which the weights of smaller networks become parts of larger ones. Such a fashion also conveniences the case of abundant computational resources, where opting for the most complex network produces superior reconstruction results for all scales. Importantly, unlike traditional methods [20], [21] that forward a batch of samples multiple times during training, we forward each batch a time and thus do not enlarge the training costs.

Conventional practices have often employed the same feature extraction approach across images of various scales [9], [10]. However, tasks at different scales demand specific adjustments to their features tailored to their respective scales, which can also, to some extent, compensate for the performance loss brought about by reduced computational resources and mutual weight influence from parameter-sharing. For better scale information injection, we enhance any-scale feature in a feature-interweaving fashion, by repeating and inserting

scale pairs into features at regular intervals. These operations decouple mutual weight influence and bring “any-scale” implementation closer to the performance of the original arbitrary-scale network. Our AnySR method can be incorporated into most existing arbitrary-scale SR methods, and extensive experiments on typical test datasets have demonstrated its effectiveness.

Overall, the major contributions of this paper are three-fold:

- We present AnySR, a versatile approach designed to adapt existing SISR methods to function seamlessly across “any-scale” and with “any-resources”.
- Any-scale enhancement through feature-interweaving fashion ensures sufficient scale information and correct feature/scale processing.
- Extensive experiments show that when integrated with existing arbitrary-scale SR methods, the proposed AnySR approach consistently achieves comparable performance in a more efficient manner.

II. RELATED WORK

A. Arbitrary-Scale Super-Resolution

Over the past few years, many classic convolutional neural networks (CNNs) based methods, such as SRCNN [22], SRResNet [23], EDSR [1], and RDN [3], have been proposed and shown commendable promise in SR tasks. To further improve SR performance, more methods incorporate residual blocks [6], [24], [25], dense blocks [3], [26], [27], and other techniques [28], [29], [30], [31], [32], [33]. Additionally, certain SR approaches leverage attention mechanisms, such as self-attention [34], [35], channel attention [36], [37], [38], and non-local attention [39], [40], [41]. Yet, most of these methods are tailored to particular scales, constraining their versatility and adaptability.

Therefore, recent researchers have turned their attention to arbitrary-scale SR tasks, aiming to tackle SR problems with arbitrary scales within a unified model. For example, Meta-SR [9] innovatively introduces the first arbitrary-scale meta-upscale module, predicting convolutional filters based on coordinates and scales to generate HR images. Subsequently, employing implicit neural representation, LIIF [10] predicts the RGB value at an arbitrary query coordinate by incorporating image coordinates and features from the backbone around that point. Further enhancements have been achieved by techniques like LTE [11], which introduces a

local texture estimator characterizing image textures in Fourier space. Additionally, recent advancements in the SR field have introduced novel structures. For instance, SRNO [12] incorporates Neural Operators, and OPE-SR [13] introduces orthogonal position encoding (OPE), an extension of position encoding. These innovations have demonstrated substantial performance improvements.

B. Efficient Image Super-Resolution

In recent years, numerous approaches [42], [43], [44], [45], [46] have emerged to develop efficient super-resolution networks. Building upon the pioneering work of SRCNN [22], FSRCNN [42] substantially accelerates the SISR networks. This acceleration is achieved by taking the original LR image as input without bicubic interpolation, using smaller sizes of convolution kernels, and incorporating a deconvolutional layer at the final stage of the network for upsampling. LapSRN [47] incrementally reconstructs the sub-band residuals of HR images through the utilization of the Laplacian pyramid. CARN [43] achieves improved efficiency by ingeniously designing cascading residual networks with group convolution. IMDN [48] presents information multi-distillation blocks featuring a contrast-aware channel attention layer. In parallel, RFDN [44] optimizes the architecture using a feature distillation mechanism through the proposed residual feature distillation block. In the wake of RFDN, RLFN [49] adds more channels to compensate for discarded feature distillation branches and introduces the residual local feature block, resulting in enhanced inference speed and superior performance with fewer parameters. On the other hand, FMEN [45] introduces the re-parameterization technique, expanding the optimization space during training through re-parameterizable building blocks [50], without incurring additional inference time. HAT [38] proposes same-task pre-training, which directly performs pre-training on a larger-scale dataset such as ImageNet [51] based on the same task. Although it requires careful control of the pre-training iteration count and the use of an appropriately small learning rate, this approach is straightforward and leads to significant performance improvements.

III. METHODOLOGY

A. Preliminaries

The general pipeline of our AnySR method acts in accordance with most existing methods [1], [38], [52] as in the upper half of Fig. 2, typically highlighted with 1) A 3×3 convolutional layer to convert a given LR image $I_{LR} \in \mathbb{R}^{H \times W \times 3}$ into a shallow feature $f_s \in \mathbb{R}^{H \times W \times C_{in}}$ where C_{in} represents the channel number. 2) N consecutive blocks for deep feature extraction, standing out as the ‘‘AnySR’’ block in this paper, along with a common convolutional layer and a global residual connection to the shadow feature f_s . 3) An upsampler $U(\cdot)$ to reconstruct an HR version of the LR image, denoted as $I'_{HR} \in \mathbb{R}^{H' \times W' \times 3}$ where $H' > H$ and $W' > W$. For ease of the following representation, we use $F(\cdot)$ to denote feature extraction operations in 1) and 2).

Given an LR image I_{LR} , the process of constructing its HR version I'_{HR} under the arbitrary-scale setting [9], [10], [52] can

Algorithm 1 AnySR Training

Input: LR image I_{LR} , HR image I_{HR} , sorted and split scale set $S = S_1 \cup S_2 \cup \dots \cup S_T$, a pre-trained arbitrary-scale SR model F with weights Θ_F , training step K .

- 1: **for** $k = 1 \rightarrow K$ **do**
- 2: $t \sim [1, 2, \dots, T]$;
- 3: $(s^h, s^w) \sim S_t$;
- 4: $t \leftarrow T$ with probability p ;
- 5: Forward I_{LR} and (s^h, s^w) to F with weights $\Theta_F[1 : |\Theta_{F_t}|]$;
- 6: Obtain reconstructed HR image I'_{HR} via Eq. (5);
- 7: $F \leftarrow F - \beta \nabla \|I_{HR} - I'_{HR}\|_1$. % $\Theta_F[1 : |\Theta_{F_t}|]$ are updated.
- 8: **end for**
- 9: **return** F

be formulated as

$$I'_{HR} = U(F(I_{LR}, S; \Theta_F); \Theta_U), \quad (1)$$

where Θ_F denotes the learnable parameters of feature extractor $F(\cdot)$, taking as inputs LR image I_{LR} and upsampling scale set $S = \{(s_i^h, s_i^w)\}_{i=1}^n$, Θ_U denotes the learnable parameters of upsampler $U(\cdot)$ taking the outputs of $F(\cdot)$ as inputs.

As analyzed in Sec. I, such an arbitrary-scale pipeline constrains the efficiency and scalability of applying SR models. A less complex network can sometimes adequately handle a smaller reconstruction scale. Consequently, allocating a uniform inference overhead for all scales manifests insufficient scalability to accommodate running platforms of time-varying computational resources, and presents poor efficiency when dealing with smaller reconstruction scales. Contemplating this, in Sec. III-B, we rebuild current methods as an implementation of ‘‘any-resource’’, to enable the selection of a smaller network in the SR model based on the abundance of computing resources. Then, we continue proposing ‘‘feature-interweaving’’ in Sec. III-C to enhance the performance under our ‘‘any-resource’’ implementation.

B. Any-Resource Implementation

The upper part of Fig. 2 manifests our any-resource implementation. Considering the close relationship between scaling factor (s_i^h, s_i^w) and reconstruction difficulty, we rearrange scale set S in an ascending order and partition it into T groups:

$$S_1 \cup S_2 \cup \dots \cup S_T = S \quad \text{and} \quad S_1 \cap S_2 \cap \dots \cap S_T = \emptyset. \quad (2)$$

Correspondingly, T distinct feature extraction networks $\{F_t\}_{t=1}^T$ with various complexities are then constructed. Each network F_t is affiliated with parameters of a particular size as Θ_{F_t} , indicating the complexity. We have $|\Theta_{F_{t+1}}| > |\Theta_{F_t}|$.

A natural approach is to utilize network F_t to process the SR tasks within group S_t such that 1) Smaller-scale tasks can be efficiently fulfilled through a structurally simple network; 2) Larger-scale ones benefit from complex networks. Thus, the process of reconstructing group S_t tasks becomes

$$I'_{HR} = U(F_t(I_{LR}, S_t; \Theta_{F_t}); \Theta_U). \quad (3)$$

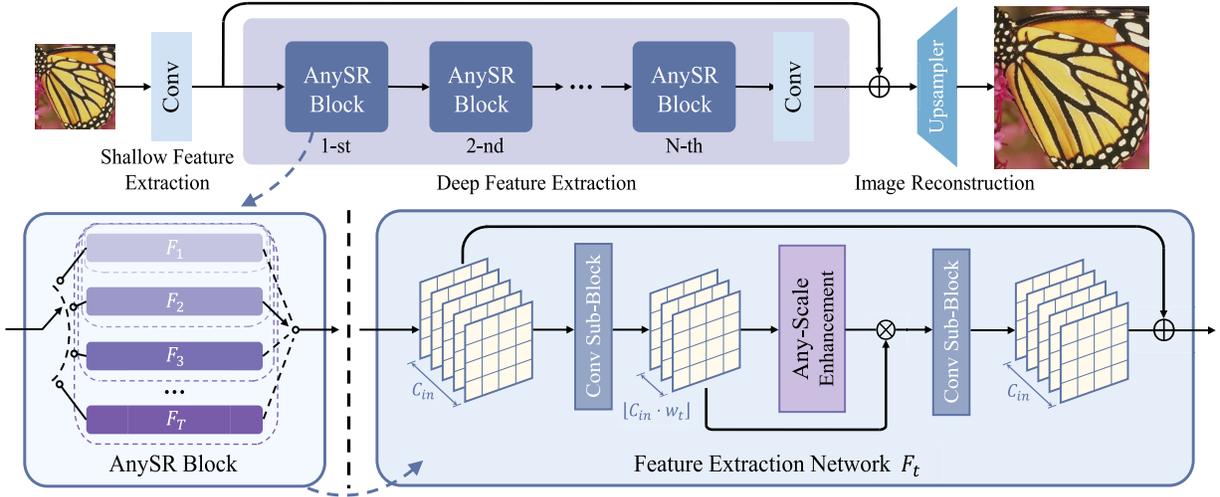


Fig. 2. Overview of the proposed AnySR: Includes Shallow Feature Extraction, Deep Feature Extraction with AnySR Blocks, and Image Reconstruction. AnySR Block automatically selects sub-net F_t based on task complexity for feature extraction and enhancement.

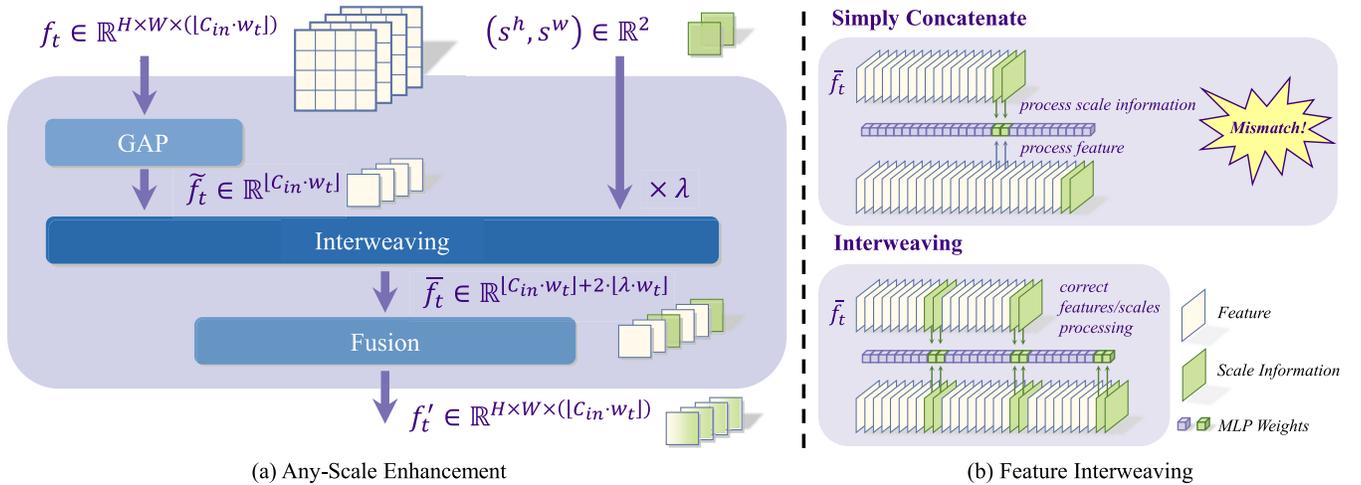


Fig. 3. Framework of our any-scale enhancement including (a) any-scale enhancement pipeline and (b) feature-interweaving illustration.

1) *All-in-One Training*: Although the above process greatly conserves computational resources, it defects in that 1) T networks have to be repeatedly trained like traditional SR methods, and that 2) additional parameters are introduced compared with arbitrary-scale SR methods. To solve this issue, we follow [20] and [21] to train and infer all networks in a parameter-sharing manner by the following set of constraints:

$$\begin{aligned} \Theta_{F_1} &= \Theta_F[1 : |\Theta_{F_1}|], \\ \Theta_{F_2} &= \Theta_F[1 : |\Theta_{F_2}|], \\ &\vdots \\ \Theta_{F_T} &= \Theta_F[1 : |\Theta_{F_T}|] = \Theta_F, \end{aligned} \quad (4)$$

where $\Theta_F[1 : |\Theta_{F_i}|]$ represents the first $|\Theta_{F_i}|$ filters of Θ_F , making F_i a subnet of F as $\Theta_{F_1} \subset \Theta_{F_2} \subset \dots \subset \Theta_{F_T} \subseteq \Theta_F$.

Consequently, the process of our AnySR to reconstruct multiple scales can be unified as follows:

$$I'_{HR} = U(F(I_{LR}, S_i; \Theta_F[1 : |\Theta_{F_i}|]); \Theta_U). \quad (5)$$

Algorithm 1 shows how to train AnySR. At each training iteration, we randomly forward a subnet of the original model F and update only the corresponding weights $\Theta_F[1 : |\Theta_{F_i}|]$. It realizes multiple networks of different resources in one training, and introduces no additional parameters. In Line 4 of Algorithm 1, the subnet is probably reset to F for the consideration of retaining the original ability of F , which performs multiple scale tasks through the entire weights Θ_F . Note that, we forward each batch once, a distinction from existing weight-sharing methods [20], [21] that perform multiple forwards, for the consideration of 1) better training efficiency, and 2) already fully-pretrained weights Θ_F .

C. Any-Scale Enhancement

Though any-resource implementation boosts the scalability and efficiency of SR tasks, its parameter-sharing puts the performance at risk, due to fewer parameters for smaller scales and mutual weight influence among different scales. It is of great necessity to enhance the reconstruction results of smaller

networks and bring them closer to the performance of the original arbitrary-scale network.

Previous research [52] has proven beneficial from linking extracted features with scales, ultimately optimizing the performance of the upsampling module. We realize scale information can be well excavated and ameliorated under our any-resource implementation for the following reasons: 1) The weights $\Theta_F[1 : |\Theta_{F_t}|]$ are particularly trained to deal with the scale set S_t , leading to scale-aware features; 2) Partial weights are shared among different subnets, injecting information of other scales. Therefore, we achieve any-scale enhancement by emphasizing the significance of customized handling for features across different scales. Specifically, as shown in Fig. 2, inside each ‘‘AnySR’’ block lies two sub-blocks with the first/second reducing/increasing the channels. We choose to inject better scale information, in a feature-interweaving fashion as illustrated in Fig. 3(a), into the outputs of the first sub-block, denoted as $\tilde{f}_t \in \mathbb{R}^{H \times W \times (\lfloor C_{in} \cdot w_t \rfloor)}$, where $w_t = \frac{|\Theta_{F_t}|}{|\Theta_F|}$, and $\lfloor \cdot \rfloor$ denotes the floor function.

1) *Feature-Interweaving*: Features $\tilde{f}_t \in \mathbb{R}^{H \times W \times (\lfloor C_{in} \cdot w_t \rfloor)}$ from F_t would first go through a global average pooling (GAP) to get a variant $\tilde{f}_t \in \mathbb{R}^{\lfloor C_{in} \cdot w_t \rfloor}$ where the scale information $(s^h, s^w) \in S_t$ is formally injected.

To this end, one naive fashion, as illustrated in the upper half of Fig. 3(b), is to follow existing methods [10], [53] which simply concatenate (s^h, s^w) at the rear of \tilde{f}_t to form $\tilde{f}_t = [\tilde{f}_t, s^h, s^w] \in \mathbb{R}^{\lfloor C_{in} \cdot w_t \rfloor + 2}$. Then, a two-layer MLP, with weights $W_1 \in \mathbb{R}^{2 \cdot C_{in} \times (C_{in} + 2)}$, $W_2 \in \mathbb{R}^{C_{in} \times 2 \cdot C_{in}}$ and a ReLU layer inserted between, is created, followed by a Sigmoid function to weight the original features f_t as

$$f'_t = f_t \odot \text{Sigmoid}(W_2[1 : \lfloor C_{in} \cdot w_t \rfloor, :]) \cdot \text{ReLU}(W_1[:, 1 : \lfloor C_{in} \cdot w_t \rfloor + 2] \cdot \tilde{f}_t), \quad (6)$$

where $W_1[:, 1 : \lfloor C_{in} \cdot w_t \rfloor + 2] \in \mathbb{R}^{2 \cdot C_{in} \times (\lfloor C_{in} \cdot w_t \rfloor + 2)}$ and $W_2[1 : \lfloor C_{in} \cdot w_t \rfloor, :] \in \mathbb{R}^{\lfloor C_{in} \cdot w_t \rfloor \times 2 \cdot C_{in}}$ are the shared MLP weights for the network F_t . f'_t then serves as the input of the second sub-block in the ‘‘AnySR’’ block.

Though this naive approach facilitates the interaction between features and scale information, two notable limitations, as we analyze, arise: 1) Insufficient scale information. Compared to a total of $\lfloor C_{in} \cdot w_t \rfloor$ channels in image features, where $\lfloor C_{in} \cdot w_t \rfloor \gg 2$ signifies a limited influence of scale information on the weighted features, the scale information takes up only 2 dimensions. 2) Inappropriate scale processing. For the MLP weights $W_1 \in \mathbb{R}^{2 \cdot C_{in} \times (C_{in} + 2)}$, the sub-weights $W_1[:, \lfloor C_{in} \cdot w_t \rfloor + 1 : \lfloor C_{in} \cdot w_t \rfloor + 2] \in \mathbb{R}^{2 \cdot C_{in} \times 2}$ are used to process the scale information $(s^h, s^w) \in S_t$. For $(s^h, s^w) \in S_{t+1}$, the weights to process scales are $W_1[:, \lfloor C_{in} \cdot w_{t+1} \rfloor + 1 : \lfloor C_{in} \cdot w_{t+1} \rfloor + 2] \in \mathbb{R}^{2 \cdot C_{in} \times 2}$ and $W_1[:, \lfloor C_{in} \cdot w_t \rfloor + 1 : \lfloor C_{in} \cdot w_t \rfloor + 2] \in \mathbb{R}^{2 \cdot C_{in} \times 2}$ are now responsible for processing image features. If off-the-shelf methods [10], [53] are still applied here, as a result, they may weaken the impact of scale information and thus fail to well excavate customized features at different scales. Experimental demonstration has been provided in Table VII.

To better align features with scale information, for network F_t , our feature-interweaving inserts scale pair (s^h, s^w) for λ

times into positions $(\lfloor C_{in} \cdot i / \lambda \rfloor + 2i - 1, \lfloor C_{in} \cdot i / \lambda \rfloor + 2i)$ where $i = 1, 2, \dots, \lfloor \lambda \cdot w_t \rfloor$, of the pooled features $\tilde{f}_t \in \mathbb{R}^{\lfloor C_{in} \cdot w_t \rfloor}$, which leads to the following concatenated features $\tilde{f}_t \in \mathbb{R}^{\lfloor C_{in} \cdot w_t \rfloor + 2 \cdot \lfloor \lambda \cdot w_t \rfloor}$:

$$\begin{aligned} \tilde{f}_t = & \left[\tilde{f}_t[1 : \lfloor C_{in} \cdot 1 / \lambda \rfloor], s^h, s^w, \right. \\ & \tilde{f}_t[\lfloor C_{in} \cdot 1 / \lambda \rfloor + 1 : \lfloor C_{in} \cdot 2 / \lambda \rfloor], s^h, s^w, \\ & \dots, \\ & \tilde{f}_t[\lfloor C_{in} \cdot (i - 1) / \lambda \rfloor + 1 : \lfloor C_{in} \cdot i / \lambda \rfloor], s^h, s^w, \dots, \\ & \left. \tilde{f}_t[\lfloor C_{in} \cdot \lfloor \lambda \cdot w_t \rfloor / \lambda \rfloor + 1 :], s^h, s^w \right]. \quad (7) \end{aligned}$$

Notably, the scale pair (s^h, s^w) is not always appended at the end of \tilde{f}_t . For Convenience, we define the fixed interval between these insertions as $C_{fix} = C_{in} / \lambda$, and refer to the last portion of the feature \tilde{f}_t as C_{last} . When $C_{last} < C_{fix}$, the scale pair is not appended, ensuring correct feature and scale processing. Conversely, when $C_{last} = C_{fix}$, the scale pair is appended to the end of the feature.

Then, weights W_1 in the two-layer MLP become $W_1 \in \mathbb{R}^{2 \cdot C_{in} \times (C_{in} + 2 \cdot \lambda)}$, and the weighted features in Eq. (6) is:

$$f'_t = f_t \odot \text{Sigmoid}(W_2[1 : \lfloor C_{in} \cdot w_t \rfloor, :]) \cdot \text{ReLU}(W_1[:, 1 : \lfloor C_{in} \cdot w_t \rfloor + 2 \cdot \lfloor \lambda \cdot w_t \rfloor] \cdot \tilde{f}_t). \quad (8)$$

As a consequence, the weights $W_1[:, \lfloor C_{in} \cdot i / \lambda \rfloor + 2i - 1 : \lfloor C_{in} \cdot i / \lambda \rfloor + 2i] \in \mathbb{R}^{2 \cdot C_{in} \times 2}$, in which $i = 1, 2, \dots, \lfloor \lambda \cdot w_t \rfloor$ consistently deal with the scale information whatever the specific value of t . In conclusion, our feature-interweaving overcomes the issues of traditional methods [10], [53] by taking into account two key operations: 1) repeating the scale pair (s^h, s^w) λ times to ensure sufficient scale information for network F_t ; 2) inserting scale pairs into features in regular intervals to ensure the correct feature/scale processing. The efficacy of our feature-interweaving will be shown in Sec. IV.

IV. EXPERIMENTAL RESULTS

A. Experiment Settings

1) *Training Details*: We primarily rely on existing well-established arbitrary-scale SR models, including Meta-SR [9], LIIF [10], ArbSR [52], and SRNO [12], for AnySR rebuilding. Without loss of generality, experiments on top of two famous feature extraction backbones, including EDSR [1] and RDN [3], are conducted to validate the universality of AnySR.

During specific implementation, we configure the network number $T = 4$ and $\{w_t\}_{t=1}^T = \{0.5, 0.7, 0.9, 1.0\}$, that is, the smallest scale group S_1 enables a reduction in inference cost by up to 50% compared with the original. The reset probability p in Line 4 of Algorithm 1 is configured at 0.6, and the hyperparameter λ in feature-interweaving is set as 4 for EDSR and 8 for RDN. Accordingly, the scale groups are designated as $S_1 = \{1.1, 1.2, \dots, 1.7\}$, $S_2 = \{1.8, 1.9, \dots, 2.5\}$, $S_3 = \{2.6, 2.7, \dots, 3.2\}$ and $S_4 = \{3.3, 3.4, \dots, 4.0\}$.

We train AnySR for 500 epochs from the pre-trained models, with an initial learning rate 10^{-5} decayed by 0.5 every 200 epochs. As pre-trained models for some existing methods (Meta-SR: EDSR; ArbSR: EDSR, RDN) are not available, we re-implement and re-train them following the experiment

TABLE I

PSNR (dB) AND LPIPS COMPARISON OF EXISTING ARBITRARY-SCALE SR METHODS AND THEIR ANYSR VARIANTS, HIGHLIGHTED BY † (THROUGH DIFFERENT SUBNETS) AND ‡ (ANYSR-RETRAINED VERSION VIA THE LARGEST NETWORK). * DENOTES OUR RE-IMPLEMENTATION. FOLLOWING [13], ANYSR PERFORMANCE IS HIGHLIGHTED IN BLUE IF WITHIN 0.15dB (PSNR) / 0.01 (LPIPS) OF THE VANILLA METHOD, AND RED IF IMPROVED

Backbone Networks	Methods	Set5 [54]			Set14 [55]			B100 [56]			Urban100 [57]			Manga109 [58]		
		×2	×3	×4	×2	×3	×4	×2	×3	×4	×2	×3	×4	×2	×3	×4
EDSR	Metric	PSNR↑														
	MetaSR*	37.89	34.35	32.08	33.57	30.29	28.52	32.14	29.07	27.54	32.07	28.10	25.97	38.36	33.39	30.37
	MetaSR†	37.85	34.40	32.17	33.44	30.28	28.55	32.07	29.06	27.56	31.84	28.09	26.05	38.30	33.47	30.46
	MetaSR‡	37.89	34.40	32.17	33.58	30.32	28.55	32.14	29.08	27.56	32.16	28.19	26.05	38.41	33.50	30.46
	LIIF	37.99	34.40	32.24	33.66	30.34	28.62	32.17	29.10	27.60	32.15	28.22	26.15	36.19	31.47	28.81
	LIIF†	37.88	34.34	32.21	33.53	30.32	28.61	32.10	29.08	27.59	31.90	28.21	26.19	35.99	31.45	28.80
	LIIF‡	37.98	34.37	32.21	33.62	30.35	28.61	32.15	29.09	27.59	32.11	28.25	26.19	36.26	31.53	28.80
	ArbSR*	37.97	34.39	32.07	33.70	30.31	28.56	32.19	29.11	27.55	32.18	28.13	25.96	38.52	33.63	30.46
	ArbSR†	37.87	34.34	32.05	33.46	30.28	28.53	32.10	29.07	27.53	31.73	27.99	25.92	38.33	33.53	30.35
	ArbSR‡	37.93	34.35	32.05	33.65	30.30	28.53	32.17	29.08	27.53	32.13	28.07	25.92	38.44	33.60	30.35
	SRNO	38.15	34.53	32.39	33.83	30.50	28.79	32.27	29.29	27.67	32.63	28.58	26.50	39.01	33.91	30.88
	SRNO†	38.04	34.50	32.38	33.74	30.47	28.78	32.19	29.15	27.64	32.27	28.45	26.45	38.78	33.77	30.83
	SRNO‡	38.12	34.51	32.38	33.81	30.49	28.78	32.25	29.17	27.64	32.52	28.50	26.45	38.96	33.83	30.83
	Metric	LPIPS↓														
	MetaSR*	0.022	0.054	0.078	0.040	0.089	0.126	0.055	0.102	0.139	0.044	0.100	0.147	0.012	0.036	0.062
	MetaSR†	0.023	0.053	0.076	0.041	0.087	0.125	0.055	0.101	0.137	0.044	0.098	0.144	0.012	0.034	0.059
	MetaSR‡	0.022	0.053	0.076	0.040	0.087	0.125	0.055	0.101	0.137	0.042	0.096	0.144	0.012	0.034	0.059
	LIIF	0.040	0.093	0.135	0.071	0.164	0.231	0.113	0.231	0.310	0.049	0.124	0.187	0.074	0.104	0.136
	LIIF†	0.041	0.098	0.143	0.074	0.169	0.238	0.113	0.237	0.320	0.053	0.132	0.199	0.075	0.106	0.141
	LIIF‡	0.041	0.096	0.139	0.072	0.167	0.235	0.113	0.235	0.316	0.049	0.125	0.189	0.075	0.105	0.137
ArbSR*	0.022	0.052	0.075	0.038	0.085	0.123	0.054	0.097	0.136	0.038	0.088	0.138	0.011	0.033	0.059	
ArbSR†	0.023	0.052	0.074	0.041	0.085	0.123	0.055	0.097	0.135	0.042	0.090	0.139	0.011	0.033	0.059	
ArbSR‡	0.022	0.052	0.074	0.038	0.085	0.123	0.054	0.097	0.135	0.037	0.088	0.139	0.011	0.033	0.059	
SRNO	0.040	0.092	0.137	0.070	0.162	0.226	0.111	0.229	0.303	0.045	0.118	0.175	0.015	0.045	0.077	
SRNO†	0.041	0.092	0.138	0.072	0.164	0.227	0.112	0.230	0.304	0.048	0.120	0.176	0.016	0.046	0.078	
SRNO‡	0.040	0.093	0.138	0.071	0.163	0.227	0.112	0.230	0.304	0.047	0.119	0.176	0.016	0.046	0.078	
RDN	Metric	PSNR↑														
	MetaSR	38.22	34.63	32.38	33.98	30.54	28.78	32.33	29.26	27.71	32.92	28.82	26.55	39.18	34.14	31.03
	MetaSR†	38.13	34.57	32.51	33.94	30.50	28.83	32.25	29.21	27.74	32.69	28.63	26.69	39.06	34.07	31.27
	MetaSR‡	38.19	34.72	32.51	33.94	30.59	28.83	32.30	29.27	27.74	32.94	28.92	26.69	39.23	34.32	31.27
	LIIF	38.17	34.68	32.50	33.97	30.53	28.80	32.32	29.26	27.74	32.87	28.82	26.68	39.26	34.21	31.20
	LIIF†	38.15	34.58	32.49	33.84	30.49	28.82	32.27	29.21	27.74	32.59	28.62	26.68	39.02	34.02	31.22
	LIIF‡	38.19	34.66	32.49	33.99	30.54	28.82	32.31	29.27	27.74	32.79	28.81	26.68	39.19	34.20	31.22
	ArbSR*	38.10	34.57	32.26	33.83	30.46	28.68	32.26	29.19	27.64	32.47	28.45	26.23	38.81	34.02	30.87
	ArbSR†	38.01	34.43	32.31	33.61	30.37	28.70	32.17	29.11	27.66	32.13	28.14	26.30	38.61	33.70	30.94
	ArbSR‡	38.05	34.59	32.31	33.80	30.47	28.70	32.27	29.20	27.66	32.57	28.52	26.30	38.82	34.01	30.94
	SRNO	38.32	34.84	32.69	34.27	30.71	28.97	32.43	29.37	27.83	33.33	29.14	26.98	39.47	34.62	31.58
	SRNO†	38.18	34.73	32.72	34.03	30.60	28.98	32.33	29.27	27.81	32.94	28.87	26.95	39.25	34.33	31.53
	SRNO‡	38.27	34.83	32.72	34.23	30.69	28.98	32.39	29.34	27.81	33.23	29.10	26.95	39.43	34.58	31.53
	Metric	LPIPS↓														
	MetaSR	0.021	0.051	0.075	0.037	0.082	0.120	0.053	0.098	0.132	0.037	0.088	0.130	0.037	0.032	0.053
	MetaSR†	0.022	0.052	0.074	0.038	0.085	0.121	0.055	0.100	0.132	0.038	0.091	0.129	0.038	0.033	0.053
	MetaSR‡	0.021	0.050	0.074	0.037	0.083	0.121	0.053	0.098	0.132	0.037	0.088	0.129	0.037	0.032	0.053
	LIIF	0.039	0.092	0.133	0.067	0.157	0.224	0.111	0.227	0.303	0.043	0.110	0.170	0.074	0.101	0.130
	LIIF†	0.040	0.093	0.136	0.071	0.165	0.233	0.113	0.233	0.312	0.046	0.122	0.187	0.074	0.103	0.135
	LIIF‡	0.040	0.094	0.137	0.071	0.165	0.232	0.113	0.235	0.313	0.045	0.119	0.182	0.074	0.103	0.133
ArbSR*	0.021	0.052	0.075	0.037	0.084	0.122	0.054	0.098	0.135	0.036	0.087	0.134	0.011	0.033	0.057	
ArbSR†	0.021	0.052	0.074	0.039	0.086	0.122	0.054	0.098	0.134	0.038	0.089	0.128	0.011	0.033	0.055	
ArbSR‡	0.021	0.052	0.074	0.037	0.085	0.122	0.054	0.097	0.134	0.035	0.083	0.128	0.011	0.032	0.055	
SRNO	0.039	0.090	0.133	0.066	0.157	0.222	0.105	0.222	0.297	0.039	0.106	0.162	0.015	0.043	0.073	
SRNO†	0.040	0.093	0.133	0.068	0.161	0.224	0.106	0.223	0.298	0.041	0.111	0.163	0.015	0.044	0.074	
SRNO‡	0.039	0.091	0.133	0.067	0.160	0.224	0.106	0.224	0.298	0.040	0.107	0.163	0.015	0.043	0.074	

TABLE II

PSNR (dB) AND LPIPS COMPARISON FOR FRACTIONAL-SCALE SR TASKS OF SRNO [12] AND THEIR ANYSR VARIANTS, HIGHLIGHTED BY \dagger (THROUGH DIFFERENT SUBNETS) AND \ddagger (ANYSR-RETRAINED VERSION VIA THE LARGEST NETWORK). ANYSR PERFORMANCE IS HIGHLIGHTED IN BLUE IF WITHIN 0.15dB (PSNR) / 0.01 (LPIPS) OF THE VANILLA METHOD, AND RED IF IMPROVED

Backbone Network	Methods	Set5			Set14			B100			Urban100			Manga109		
		$\times 1.8$	$\times 2.6$	$\times 3.6$	$\times 1.9$	$\times 2.9$	$\times 3.8$	$\times 1.9$	$\times 2.7$	$\times 3.5$	$\times 1.8$	$\times 2.6$	$\times 3.7$	$\times 1.3$	$\times 2.6$	$\times 3.3$
EDSR	Metric	PSNR \uparrow														
	SRNO	39.12	35.72	33.14	34.42	30.74	29.06	32.79	29.87	28.33	33.94	29.83	27.02	45.29	35.71	32.93
	SRNO \dagger	39.04	35.68	33.12	34.34	30.70	29.04	32.73	29.84	28.31	33.60	29.72	26.99	45.01	35.57	32.87
	SRNO \ddagger	39.11	35.70	33.12	34.39	30.71	29.04	32.78	29.85	28.31	33.85	29.77	26.99	45.28	35.63	32.87
	Metric	LPIPS \downarrow														
	SRNO	0.028	0.070	0.119	0.048	0.127	0.203	0.078	0.184	0.274	0.031	0.089	0.154	0.010	0.032	0.065
	SRNO \dagger	0.028	0.071	0.118	0.050	0.129	0.204	0.079	0.185	0.275	0.033	0.091	0.155	0.011	0.032	0.065
SRNO \ddagger	0.028	0.071	0.118	0.049	0.129	0.204	0.079	0.185	0.275	0.032	0.090	0.155	0.011	0.032	0.065	
RDN	Methods	Set5			Set14			B100			Urban100			Manga109		
		$\times 1.8$	$\times 2.6$	$\times 3.1$	$\times 1.8$	$\times 2.7$	$\times 3.5$	$\times 1.8$	$\times 2.6$	$\times 3.7$	$\times 1.8$	$\times 3.1$	$\times 3.8$	$\times 1.6$	$\times 1.8$	$\times 3.4$
	Metric	PSNR \uparrow														
	SRNO	39.31	35.92	34.65	35.35	31.50	29.73	33.54	30.28	28.21	34.67	28.88	27.35	42.39	40.80	33.30
	SRNO \dagger	39.21	35.86	34.51	35.24	31.40	29.73	33.46	30.20	28.20	34.31	28.63	27.32	42.20	40.61	33.27
	SRNO \ddagger	39.29	35.92	34.67	35.37	31.53	29.73	33.52	30.26	28.20	34.59	28.85	27.32	42.39	40.78	33.27
	Metric	LPIPS \downarrow														
SRNO	0.028	0.068	0.115	0.047	0.123	0.198	0.075	0.178	0.268	0.026	0.080	0.141	0.010	0.030	0.062	
SRNO \dagger	0.028	0.070	0.115	0.048	0.127	0.200	0.076	0.180	0.270	0.028	0.084	0.142	0.011	0.032	0.062	
SRNO \ddagger	0.028	0.069	0.115	0.047	0.124	0.200	0.076	0.180	0.270	0.027	0.081	0.142	0.010	0.031	0.062	

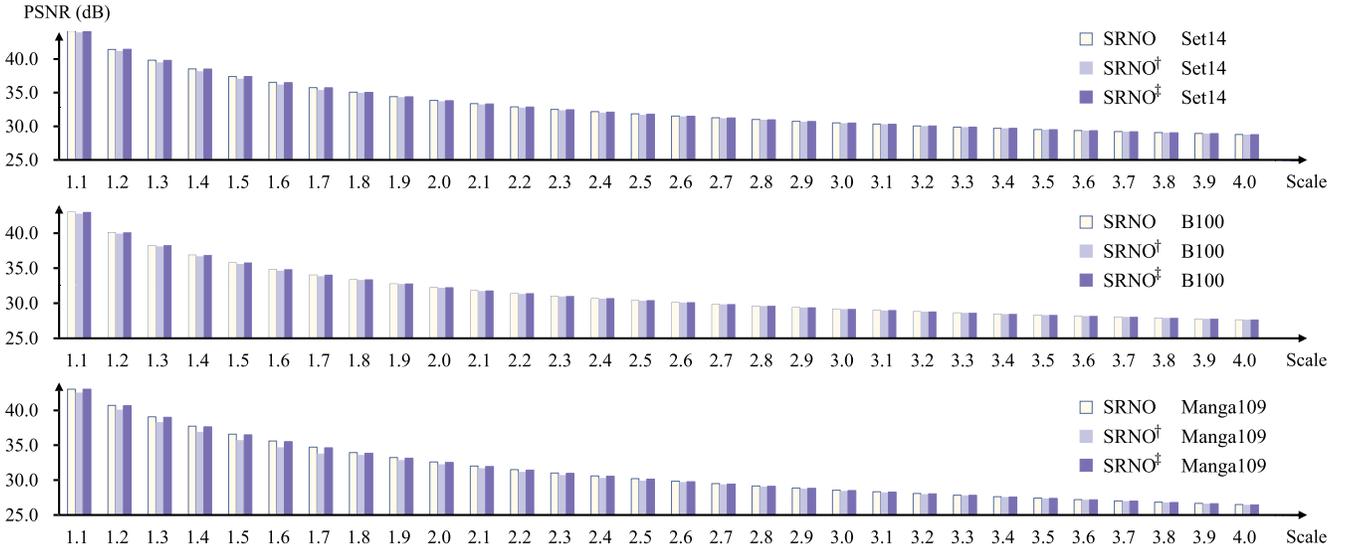


Fig. 4. PSNR(dB) comparisons across all scales on different datasets of arbitrary-scale SR model SRNO [12], its AnySR variants (through different subnets) highlighted by \dagger , and AnySR-retrained version (through the largest network) denoted by \ddagger .

settings outlined in the papers [9], [52]. To ensure a fair performance evaluation, we keep identical settings and configurations used for the original models in our training process. The patch size is 50×50 for Meta-SR and ArbSR, 48×48 for LIIF, and 128×128 for SRNO, with a batch size of 8 per GPU for EDSR and 4 per GPU for RDN. We employ an ℓ_1 loss [1] and the Adam optimizer [59].

2) *Dataset*: For training, current practices [38] use DF2K (including DIV2K [8] and Flickr2K [60]) as the training set, or pre-training on ImageNet1K [51] to enhance performance. For fairness, we choose to follow the traditional training

regimen directly on the DIV2K dataset to ensure a direct comparison with previous studies and to maintain consistency with the experimental protocols established in the literature.

For evaluation, we assess the PSNR [61] and LPIPS [62] on the most common evaluation datasets, including Set5 [54], Set14 [55], B100 [63], Urban100 [57], and Manga109 [58].

B. Main Results

1) *Quantitative Results*: Extensive experiments are conducted using in-distribution and out-of-distribution scales, and

TABLE III

PSNR (dB) AND LPIPS COMPARISON FOR OUT-OF-DISTRIBUTION SCALES OF ARBITRARY-SCALE SR METHODS AND THEIR ANYSR VARIANTS, HIGHLIGHTED BY † (THROUGH DIFFERENT SUBNETS) AND ‡ (ANYSR-RETRAINED VERSION). ANYSR PERFORMANCE IS MARKED IN BLUE IF WITHIN 0.15dB (PSNR) / 0.01 (LPIPS) OF THE BASELINE, AND RED IF IMPROVED

Backbone	Method	Set5					Set14					Urban100				
		×6	×12	×18	×24	×30	×6	×12	×18	×24	×30	×6	×12	×18	×24	×30
EDSR	Metric	PSNR↑														
	MetaSR	28.48	24.17	22.01	20.97	20.06	26.30	22.99	21.44	20.43	19.60	23.57	20.74	19.47	18.71	18.17
	MetaSR†‡	28.50	24.20	21.96	20.94	20.01	26.31	22.99	21.45	20.42	19.57	23.61	20.75	19.47	18.71	18.15
	LIIF	28.96	24.56	22.48	21.37	20.58	26.45	23.16	21.61	20.66	19.81	23.79	20.89	19.62	18.84	18.29
	LIIF†‡	28.94	24.59	22.51	21.39	20.56	26.45	23.19	21.61	20.65	19.83	23.79	20.89	19.61	18.84	18.29
	ArbSR	28.39	23.76	21.69	20.64	19.75	26.19	22.57	20.99	19.93	19.22	23.47	20.37	19.08	18.33	17.82
	ArbSR†‡	28.37	23.73	21.66	20.62	19.73	26.15	22.56	20.96	19.91	19.19	23.46	20.36	19.06	18.32	17.80
	SRNO	29.06	24.62	22.52	21.37	20.54	26.55	23.19	21.65	20.68	19.77	24.08	21.10	19.75	18.96	18.40
	SRNO†‡	29.05	24.57	22.51	21.31	20.52	26.55	23.15	21.61	20.65	19.73	24.01	21.01	19.67	18.88	18.33
	Metric	LPIPS↓														
	MetaSR	0.123	0.211	0.263	0.281	0.297	0.194	0.266	0.302	0.319	0.333	0.224	0.340	0.383	0.400	0.410
	MetaSR†‡	0.123	0.207	0.262	0.282	0.297	0.193	0.264	0.299	0.315	0.330	0.222	0.338	0.381	0.396	0.407
	LIIF	0.204	0.379	0.510	0.587	0.626	0.330	0.487	0.567	0.625	0.672	0.287	0.489	0.603	0.668	0.709
	LIIF†‡	0.210	0.389	0.513	0.598	0.635	0.336	0.493	0.573	0.630	0.676	0.292	0.495	0.609	0.672	0.712
	ArbSR	0.133	0.265	0.319	0.333	0.343	0.180	0.279	0.318	0.330	0.339	0.229	0.350	0.389	0.401	0.404
	ArbSR†‡	0.136	0.269	0.321	0.337	0.348	0.183	0.281	0.319	0.331	0.341	0.233	0.352	0.392	0.403	0.406
SRNO	0.210	0.394	0.525	0.619	0.636	0.332	0.496	0.579	0.637	0.687	0.281	0.491	0.610	0.677	0.719	
SRNO†‡	0.209	0.392	0.529	0.620	0.641	0.329	0.491	0.578	0.636	0.687	0.279	0.485	0.605	0.674	0.719	
RDN	Metric	PSNR↑														
	MetaSR	28.95	24.36	22.04	20.99	20.01	26.57	23.10	21.50	20.51	19.55	24.06	20.96	19.62	18.83	18.25
	MetaSR†‡	28.96	24.28	22.02	20.98	20.04	26.55	23.03	21.44	20.47	19.51	24.03	20.91	19.52	18.73	18.15
	LIIF	29.15	24.86	22.66	21.50	20.57	26.64	23.24	21.73	20.78	19.85	24.20	21.15	19.80	19.00	18.44
	LIIF†‡	29.14	24.89	22.64	21.49	20.62	26.65	23.28	21.76	20.76	19.90	24.18	21.15	19.81	19.02	18.44
	ArbSR	28.45	23.57	21.53	20.45	19.55	26.18	22.44	20.83	19.79	19.09	23.55	20.27	18.97	18.24	17.73
	ArbSR†‡	28.51	23.57	21.52	20.43	19.56	26.22	22.46	20.84	19.78	19.09	23.60	20.29	18.97	18.23	17.73
	SRNO	29.40	24.94	22.73	21.50	20.65	26.76	23.33	21.82	20.78	19.96	24.42	21.35	19.95	19.15	18.55
	SRNO†‡	29.45	24.87	22.76	21.45	20.62	26.74	23.26	21.75	20.79	19.88	24.38	21.27	19.86	19.06	18.46
	Metric	LPIPS↓														
	MetaSR	0.114	0.195	0.248	0.272	0.295	0.184	0.257	0.293	0.308	0.326	0.199	0.320	0.370	0.389	0.403
	MetaSR†‡	0.106	0.192	0.244	0.267	0.292	0.182	0.257	0.292	0.306	0.326	0.196	0.317	0.368	0.388	0.402
	LIIF	0.197	0.361	0.491	0.566	0.601	0.319	0.472	0.551	0.604	0.657	0.263	0.457	0.574	0.644	0.690
	LIIF†‡	0.210	0.372	0.501	0.579	0.610	0.330	0.485	0.560	0.615	0.666	0.278	0.475	0.590	0.657	0.701
	ArbSR	0.132	0.267	0.316	0.333	0.340	0.179	0.278	0.316	0.327	0.335	0.229	0.351	0.388	0.398	0.401
	ArbSR†‡	0.128	0.266	0.317	0.334	0.341	0.176	0.278	0.316	0.327	0.336	0.224	0.349	0.387	0.398	0.400
SRNO	0.206	0.387	0.514	0.591	0.628	0.323	0.485	0.565	0.623	0.674	0.267	0.474	0.595	0.665	0.711	
SRNO†‡	0.203	0.378	0.503	0.582	0.624	0.322	0.479	0.559	0.612	0.663	0.263	0.461	0.582	0.654	0.703	

we analyze the quantitative performance below to show the efficacy of our proposed AnySR.

a) *In-distribution scales*: Table I presents the PSNR and LPIPS performance on integer scales, showing the comparison between AnySR variants with different subnets dealing with individual scales and AnySR-retrained largest network for all scales. Additionally, Table II shows the quantitative results on fractional scales, with three fractional scales randomly selected for each dataset on the SRNO model [12]. More PSNR results of AnySR across all 30 scales ($\times 1.1$ - $\times 4.0$) are visualized in Fig. 4, to provide a comprehensive overview of the results.

From Table I and Table II, there is a limited performance drop at smaller scales with an average decrease of 0.15dB

(PSNR) / 0.05 (LPIPS). With only this minimal performance loss, we can transform existing methods into an any-scale, any-resource implementation by dynamically selecting the network based on available computational resources, and we believe this trade-off is worthwhile. When selecting the entire network, AnySR shows only a negligible performance drop across all scales, with an average PSNR difference of merely 0.05dB, and the average LPIPS difference is even less than 0.01. This indicates that AnySR-retrained models can effectively preserve the original model's performance during the rebuilding process, which can also be visually observed in Fig. 4. Such good performance is closely related to the reset probability p in Line 4 of Algorithm 1, investigated in Sec. IV-E.

HR (Dataset, Img ID, Scale)

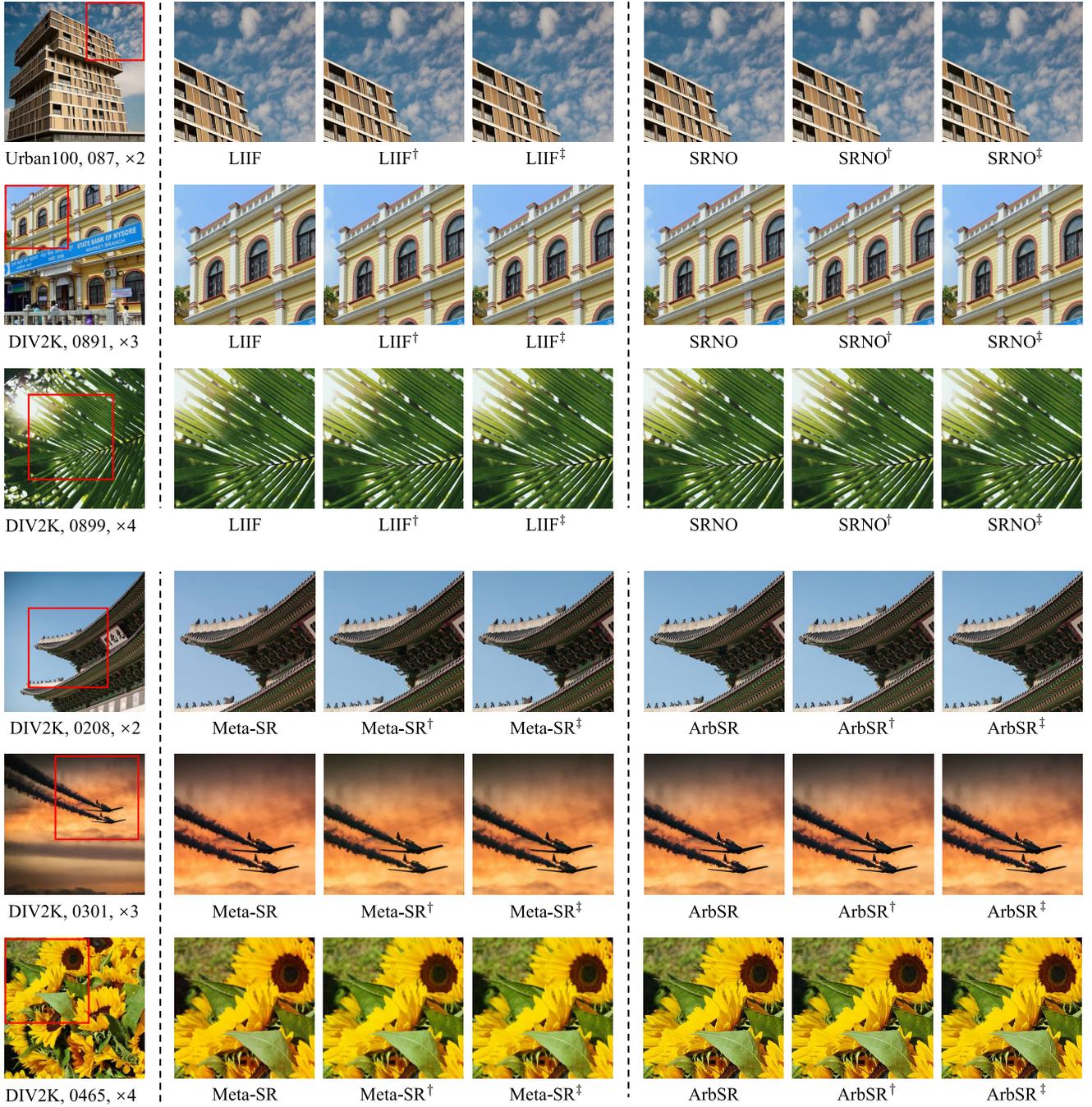


Fig. 5. Visualization of existing arbitrary-scale SR models LIIF [10], SRNO [12], Meta-SR [9], ArbSR [52], their AnySR variants (through different subnets) highlighted by \dagger , and AnySR-retrained version (through the largest network) denoted as \ddagger .

b) Out-of-distribution scales: Table III presents the performance of out-of-distribution scales. For scales near the training range ($\times 6$, $\times 12$), AnySR shows overall metric improvements. However, as the scales deviate from the training range, there is a noticeable performance drop. Nevertheless, we observe across all scales, the AnySR-retrained version achieves better LPIPS scores, indicating that the images generated by AnySR maintain a high level of visual quality. Overall, these results demonstrate that our AnySR preserves, and in some cases surpasses, the reconstruction capability of the original network, allowing for stable, high-quality outputs

across devices with varying computational resources through the dynamic selection of different reconstruction networks.

2) Qualitative Results: A large number of qualitative experiments have been carried out, demonstrating that AnySR can generate competitive, high-fidelity visual images while reducing computational costs.

a) In-distribution scales: Fig. 5 makes a comparison between off-the-shelf arbitrary-scale SR models and their corresponding AnySR variants on DIV2K [8] and Urban100 [57] under the two specific settings including:



Fig. 6. Visualization comparison of SRNO [12], its AnySR variant (highlighted by \dagger), and AnySR-retrained version (denoted by \ddagger) in representative scenarios. Line 1: dense and fine grains/textures; Line 2: animal fur textures; Line 3: dense and complex lines.

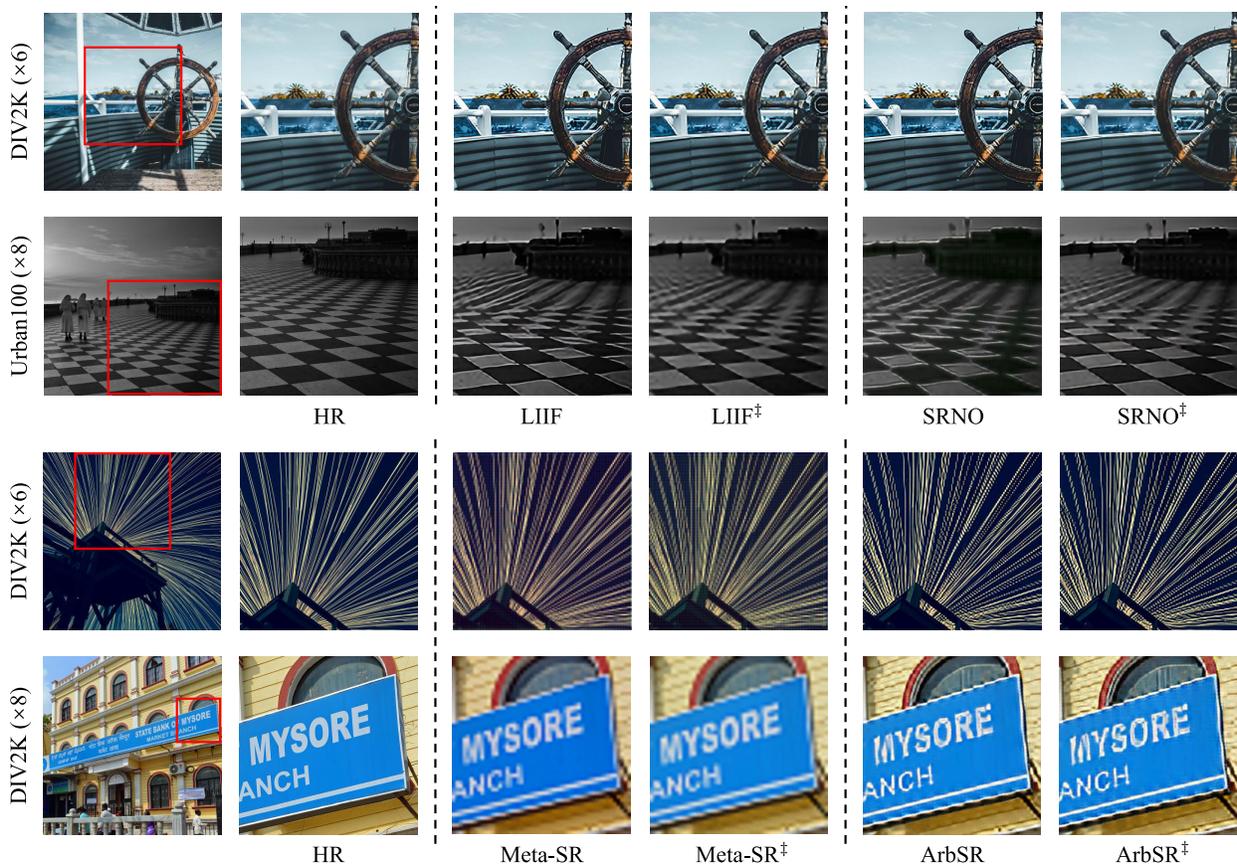


Fig. 7. Visualization results for out-of-distribution scales of existing arbitrary-scale SR models and their AnySR-retrained version (denoted by \ddagger).

different subnets for different scales and the entire network for all scales.

Qualitative results of the models are satisfactory, as both AnySR \dagger and AnySR \ddagger avoid generating artifacts or distortions. As shown in Line 2 and Line 3 of Fig. 5, both clear building edges and rich textural details on the leaves are retained.

We conduct evaluations of the model's reconstruction ability across different scales using representative scenarios. In Line 1 of Fig. 6, we task our AnySR with reconstructing fine-grained, densely arranged objects to assess its handling of detail,

yielding commendable results. We can clearly observe the frost on blueberries, the dew on leaves, and the texture on coral surfaces. In Line 2, AnySR successfully reconstructs complex animal patterns (zebra stripes) and realistic, delicate animal fur (feathers of the swan and the eagle). In Line 3, our model accurately reconstructs the intricate and intertwined lines of buildings, resulting in clear and organized outputs.

All reconstructed results demonstrate the robust reconstruction capability of AnySR. Even at the $\times 2$ scale, where computational costs are reduced by nearly 30%, the model still

TABLE IV

COMPARISON OF THE NUMBERS OF PARAMETERS, PSNR PERFORMANCES, AND FLOPS COUNT FOR EXISTING ARBITRARY-SCALE SR METHODS AND THEIR ANYSR VARIANTS (THROUGH DIFFERENT SUBNETS) HIGHLIGHTED BY †, ANYSR-RETRAINED VERSION (THROUGH THE LARGEST NETWORK) DENOTED BY ‡. PARAMETERS AND FLOPS ARE MEASURED WITH CONSISTENT PIXEL SIZE FOR UPSAMPLING SR IMAGES ACROSS ALL SCALES

Model	Params (M)	×2		×3		×4	
		PSNR (dB)	FLOPs (G)	PSNR (dB)	FLOPs (G)	PSNR (dB)	FLOPs (G)
MetaSR	1.22 (100.00%)	33.57 (+0.00)	140.20 (100.00%)	30.29 (+0.00)	62.03 (100.00%)	28.52 (+0.00)	35.01 (100.00%)
MetaSR†	1.50 (122.95%)	33.44 (−0.13)	97.73 (69.71%)	30.28 (−0.01)	55.45 (89.40%)	28.55 (+0.03)	35.01 (100.00%)
MetaSR‡	1.50 (122.95%)	33.58 (+0.01)	140.20 (100.00%)	30.32 (+0.03)	62.03 (100.00%)	28.55 (+0.03)	35.01 (100.00%)
LIIF	1.22 (100.00%)	33.66 (+0.00)	140.20 (100.00%)	30.34 (+0.00)	53.03 (100.00%)	28.62 (+0.00)	35.01 (100.00%)
LIIF†	1.50 (122.95%)	33.53 (−0.13)	97.73 (69.71%)	30.32 (−0.02)	45.23 (85.30%)	28.61 (−0.01)	35.01 (100.00%)
LIIF‡	1.50 (122.95%)	33.62 (−0.04)	140.20 (100.00%)	30.35 (+0.01)	53.02 (100.00%)	28.61 (−0.01)	35.01 (100.00%)
ArbSR	1.22 (100.00%)	33.70 (+0.00)	140.92 (100.00%)	30.31 (+0.00)	62.34 (100.00%)	28.56 (+0.00)	35.14 (100.00%)
ArbSR†	1.50 (122.95%)	33.46 (−0.24)	101.86 (72.28%)	30.28 (−0.03)	57.26 (91.85%)	28.53 (−0.03)	35.14 (100.00%)
ArbSR‡	1.50 (122.95%)	33.65 (−0.05)	140.92 (100.00%)	30.30 (−0.01)	62.34 (100.00%)	28.53 (−0.03)	35.14 (100.00%)
SRNO	1.22 (100.00%)	33.83 (+0.00)	141.28 (100.00%)	30.50 (+0.00)	62.50 (100.00%)	28.79 (+0.00)	35.28 (100.00%)
SRNO†	1.50 (122.95%)	33.74 (−0.09)	97.93 (69.25%)	30.47 (−0.03)	55.54 (88.86%)	28.78 (−0.01)	35.28 (100.00%)
SRNO‡	1.50 (122.95%)	33.81 (−0.02)	141.28 (100.00%)	30.49 (−0.01)	62.50 (100.00%)	28.78 (−0.01)	35.28 (100.00%)

TABLE V

COMPARISON OF THE NUMBER OF PARAMETERS AND FLOPS OF A SINGLE RESNET BLOCK OF EDSR [1] IN SRNO [12], ITS ANYSR VARIANTS AND ITS ANYSR-RETRAINED VERSION

Setting	Metric	Conv	Relu	GAP	FC	ALL
×2, SRNO	Params (K)	73.86	0.00	-	-	73.86
	FLOPs (M)	8860.00	3.84	-	-	8863.84
×2, SRNO†	Params (K)	73.86	0.00	0.00	17.60	91.46
	FLOPs (M)	6090.00	2.64	2.64	0.024	6095.30
×2, SRNO‡	Params (K)	73.86	0.00	0.00	17.60	91.46
	FLOPs (M)	8860.00	3.84	3.84	0.035	8867.72
×3, SRNO	Params (K)	73.86	0.00	-	-	73.86
	FLOPs (M)	3920.00	1.70	-	-	3921.70
×3, SRNO†	Params (K)	73.86	0.00	0.00	17.60	91.46
	FLOPs (M)	3500.00	1.51	1.51	0.031	3503.05
×3, SRNO‡	Params (K)	73.86	0.00	0.00	17.60	91.46
	FLOPs (M)	3920.00	1.70	1.70	0.035	3923.44
×4, SRNO	Params (K)	73.86	0.00	-	-	73.86
	FLOPs (M)	2220.00	0.96	-	-	2220.96
×4,SRNO†	Params (K)	73.86	0.00	0.00	17.60	91.46
	FLOPs (M)	2220.00	0.96	0.96	0.035	2221.96
×4,SRNO‡	Params (K)	73.86	0.00	0.00	17.60	91.46
	FLOPs (M)	2220.00	0.96	0.96	0.035	2221.96

achieves visual results on par with the original model. This reflects AnySR’s ability to deliver high-quality reconstructed images, even under varying computational resource conditions.

b) *Out-of-distribution scales*: Images from the test set are downsampled and subsequently input into both the original model and its AnySR-retrained version for reconstruction, with the results compared against the corresponding high-resolution (HR) images, as illustrated in Fig. 7. Since the entire network is chosen under the out-of-distribution scale setting, the AnySR† results are omitted. For LIIF and SRNO, which perform well on out-of-distribution scales, our AnySR achieves notable improvements while fully retaining their reconstruction capabilities. For example, in Line 2 of Fig. 7,

AnySR reconstructs more regular paving stones with fewer distortions and deformations. For Meta-SR and ArbSR, which originally shows mediocre performance on out-of-distribution scales, AnySR training leads to performance improvements. This is particularly evident in Meta-SR, as shown in Lines 3 and 4 of Fig. 7. In the out-of-distribution scale setting, Meta-SR’s reconstruction results exhibit noticeable color discrepancies, but after retraining with AnySR, the model corrects the color inconsistencies and reduces image blurring.

C. Complexity Analysis

We perform a thorough comparison on SR tasks at different scales, encompassing the number of parameters, PSNR performance, and FLOPs, for existing arbitrary-scale SR methods and their AnySR variants of different subnets to solve different scales and the entire network to solve all scales. All results are evaluated on top of the EDSR [1] backbone network on the Set14 dataset [55] as shown in Table IV. Since the encoder contains 16 ResNet blocks, we calculate the parameters and FLOPs for a single image, for a single ResNet block under different settings, as shown in Table V. Taking SRNO as an example, we conduct a thorough and detailed analysis of the parameter count and computational cost.

In terms of parameters, our AnySR variant introduces only 0.28M additional weights (22.95% of the original 1.22M parameters). As shown in Table V, these additional weights come from the fully connected layers, which are primarily dedicated to any-scale enhancement.

As for computational cost, as shown in Table V, despite the introduction of fully connected layers and GAP operations, the FLOPs are almost negligible compared to those required by the convolutional layers of the original model. Instead, the selection of smaller subnets leads to significant savings in FLOPs. For example, when performing the ×2 task with selected subnetworks, each ResNet block saves around 2800M

TABLE VI

PSNR (dB) AND LPIPS COMPARISON OF SRNO USING SMALLER SUBNETS (DENOTED AS F_i) TO PERFORM OUT-OF-DISTRIBUTION SCALE SR TASKS. UNDERLINES INDICATES THE BEST PERFORMANCE

Backbone	Method	Set5					Set14					Urban100				
		$\times 6$	$\times 12$	$\times 18$	$\times 24$	$\times 30$	$\times 6$	$\times 12$	$\times 18$	$\times 24$	$\times 30$	$\times 6$	$\times 12$	$\times 18$	$\times 24$	$\times 30$
EDSR	Metric	PSNR \uparrow														
	SRNO	<u>29.06</u>	<u>24.62</u>	<u>22.52</u>	<u>21.37</u>	<u>20.54</u>	<u>26.55</u>	<u>23.19</u>	<u>21.65</u>	<u>20.68</u>	<u>19.77</u>	<u>24.08</u>	<u>21.10</u>	<u>19.75</u>	<u>18.96</u>	<u>18.40</u>
	F_1	28.64	24.32	22.31	21.09	20.18	26.26	22.98	21.50	20.51	19.63	23.53	20.74	19.47	18.73	18.20
	F_2	28.92	24.55	22.44	21.26	20.41	26.46	23.10	21.58	20.62	19.72	23.83	20.92	19.61	18.82	18.28
	F_3	29.00	24.57	22.45	21.29	20.51	26.52	23.14	21.61	20.66	19.74	23.96	20.99	19.66	18.87	18.33
	F_4 (SRNO \ddagger)	29.05	24.57	22.51	21.31	20.52	<u>26.55</u>	23.15	21.61	20.65	19.73	24.01	21.01	19.67	18.88	18.33
	Metric	LPIPS \downarrow														
	SRNO	0.210	0.394	<u>0.525</u>	0.619	<u>0.636</u>	0.332	0.496	0.579	0.637	0.687	0.281	0.491	0.610	0.677	<u>0.719</u>
	F_1	0.220	0.403	0.548	0.625	0.641	0.342	0.508	0.595	0.652	0.698	0.307	0.523	0.641	0.701	0.737
	F_2	0.213	0.398	0.526	0.624	0.650	0.333	0.498	0.581	0.640	0.689	0.289	0.497	0.617	0.683	0.725
F_3	<u>0.209</u>	0.394	0.527	<u>0.618</u>	0.641	0.331	0.494	0.579	0.637	<u>0.686</u>	0.282	0.489	0.608	0.676	0.720	
F_4 (SRNO \ddagger)	<u>0.209</u>	<u>0.392</u>	0.529	0.620	0.641	<u>0.329</u>	<u>0.491</u>	<u>0.578</u>	<u>0.636</u>	0.687	<u>0.279</u>	<u>0.485</u>	<u>0.605</u>	<u>0.674</u>	<u>0.719</u>	
RDN	Metric	PSNR \uparrow														
	SRNO	29.40	<u>24.94</u>	22.73	<u>21.50</u>	<u>20.65</u>	<u>26.76</u>	<u>23.33</u>	<u>21.82</u>	20.78	<u>19.96</u>	<u>24.42</u>	<u>21.35</u>	<u>19.95</u>	<u>19.15</u>	<u>18.55</u>
	F_1	29.01	24.72	22.46	21.32	20.49	26.53	23.20	21.70	20.64	19.81	24.02	21.06	19.73	18.94	18.37
	F_2	29.20	24.81	22.56	21.28	20.56	26.65	23.24	21.72	20.68	19.82	24.22	21.18	19.80	19.00	18.41
	F_3	29.23	24.76	22.64	21.30	20.61	26.67	23.23	21.76	20.75	19.83	24.25	21.19	19.82	19.00	18.42
	F_4 (SRNO \ddagger)	<u>29.45</u>	24.87	<u>22.76</u>	21.45	20.62	26.74	23.26	21.75	<u>20.79</u>	19.88	24.38	21.27	19.86	19.06	18.46
	Metric	LPIPS \downarrow														
	SRNO	0.206	0.387	0.514	0.591	0.628	0.323	0.485	0.565	0.623	0.674	0.267	0.474	0.595	0.665	0.711
	F_1	0.218	0.402	0.532	0.599	0.636	0.337	0.502	0.585	0.643	0.686	0.294	0.511	0.626	0.688	0.727
	F_2	0.210	<u>0.375</u>	<u>0.501</u>	<u>0.571</u>	<u>0.614</u>	0.328	0.484	0.562	0.619	0.666	0.274	0.476	0.590	0.658	0.704
F_3	0.210	0.384	0.513	0.578	0.621	0.328	0.487	0.562	0.617	0.668	0.273	0.477	0.592	0.658	0.704	
F_4 (SRNO \ddagger)	<u>0.203</u>	0.378	0.503	0.582	0.624	<u>0.322</u>	<u>0.479</u>	<u>0.559</u>	<u>0.612</u>	<u>0.663</u>	<u>0.263</u>	<u>0.461</u>	<u>0.582</u>	<u>0.654</u>	<u>0.703</u>	

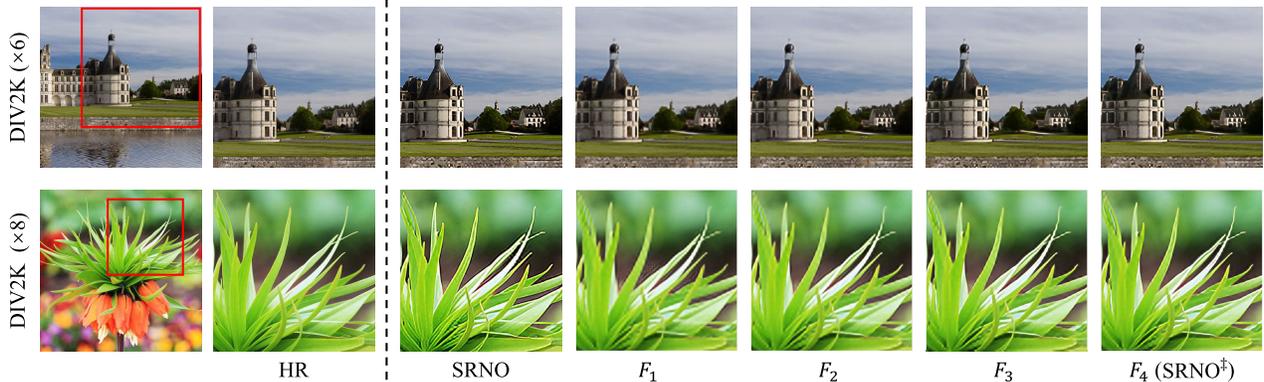


Fig. 8. Visualization results of SRNO using smaller subnets (denoted as F_i) to perform out-of-distribution scale SR tasks on the EDSR backbone.

FLOPs in the convolutional layers, while FLOPs introduced by the GAP operation and fully connected layers are less than 3M. The total number of FLOPs (97.93G) constitutes only 69.25% of the original 141.28G, with a minimal performance loss of 0.09 dB, making it suitable for devices with limited computational power to output high-quality results.

In case of ample computing resources, AnySR maintains a good performance-to-consumption balance. When performing the $\times 4$ task with the entire network, Table V shows that each ResNet block introduces negligible 0.995M FLOPs. For the entire backbone, the FLOPs of the AnySR-retrained version

(35.28G) are approximately equal to those of the original model, implying no computing introduction.

D. Subnet Exploration

We evaluate smaller subnets in handling high magnification factors. Table VI and Fig. 8 present the quantitative and qualitative results of SRNO and its AnySR variants with subnets of varying sizes, denoted as F_i , on out-of-distribution scales. Smaller subnets perform lower-scored PSNR, as reflected in the visual results with less detail recovery such as distant lighthouses and the fine edges of plant leaves. However, it improves as subnet size increases. The LPIPS scores show

TABLE VII

ABLATION RESULTS OF PSNR(DB) AND LPIPS ON ALTERNATIONS OF ANYSR INCLUDING THE “w/o ASE” VARIANT REMOVING ANY-SCALE ENHANCEMENT (ASE) AND THE “w/o FI” VARIANT REPLACING FEATURE-INTERWEAVING (FI) WITH A SIMPLE CONCATENATION. UNDERLINES INDICATES THE BEST PERFORMANCE

Settings	Set5			Set14			B100			Urban100			Manga109		
	×2	×3	×4	×2	×3	×4	×2	×3	×4	×2	×3	×4	×2	×3	×4
Metric	PSNR↑														
AnySR [†]	<u>38.04</u>	34.50	<u>32.38</u>	<u>33.74</u>	<u>30.47</u>	<u>28.78</u>	<u>32.19</u>	<u>29.15</u>	<u>27.64</u>	<u>32.27</u>	<u>28.45</u>	<u>26.45</u>	<u>38.78</u>	<u>33.77</u>	<u>30.83</u>
AnySR [†] (w/o ASE)	38.01	<u>34.51</u>	32.34	33.72	30.43	28.73	32.14	29.11	27.61	<u>32.27</u>	28.44	26.43	38.70	33.74	30.79
AnySR [†] (w/o FI)	38.00	34.47	32.33	33.71	30.42	28.73	32.14	29.12	27.61	32.25	28.43	26.42	38.75	33.72	30.81
AnySR [‡]	<u>38.12</u>	34.51	<u>32.38</u>	<u>33.81</u>	<u>30.49</u>	<u>28.78</u>	<u>32.25</u>	<u>29.17</u>	<u>27.64</u>	<u>32.52</u>	<u>28.50</u>	<u>26.45</u>	<u>38.96</u>	<u>33.83</u>	<u>30.83</u>
AnySR [‡] (w/o ASE)	38.10	<u>34.52</u>	32.34	33.76	30.45	28.73	32.21	29.12	27.61	32.51	28.47	26.43	38.90	33.81	30.79
AnySR [‡] (w/o FI)	38.10	34.48	32.33	33.76	30.45	28.73	32.22	29.14	27.61	32.50	28.46	26.42	38.91	33.80	30.81
Metric	LPIPS↓														
AnySR [†]	<u>0.041</u>	<u>0.092</u>	<u>0.138</u>	<u>0.072</u>	<u>0.164</u>	<u>0.227</u>	<u>0.112</u>	<u>0.230</u>	<u>0.304</u>	<u>0.048</u>	<u>0.120</u>	<u>0.176</u>	<u>0.016</u>	<u>0.046</u>	<u>0.078</u>
AnySR [†] (w/o ASE)	0.045	0.096	0.142	<u>0.072</u>	0.167	0.230	0.116	0.235	0.305	0.053	0.121	0.179	0.019	0.053	0.080
AnySR [†] (w/o FI)	0.044	0.095	0.139	0.075	0.167	0.229	0.115	0.235	0.305	0.049	0.122	0.179	0.018	0.049	0.079
AnySR [‡]	<u>0.040</u>	<u>0.093</u>	<u>0.138</u>	<u>0.071</u>	<u>0.163</u>	<u>0.227</u>	<u>0.112</u>	<u>0.230</u>	<u>0.304</u>	<u>0.047</u>	<u>0.119</u>	<u>0.176</u>	<u>0.016</u>	<u>0.046</u>	<u>0.078</u>
AnySR [‡] (w/o ASE)	0.043	0.096	0.142	0.073	0.165	0.230	0.115	0.234	0.305	0.052	0.121	0.179	0.018	0.052	0.080
AnySR [‡] (w/o FI)	0.043	0.095	0.139	0.072	0.165	0.229	0.114	0.233	0.305	0.049	0.121	0.179	0.018	0.048	0.079

TABLE VIII

ABLATION RESULTS OF PSNR (DB) ON THE RESET PROBABILITY p . UNDERLINES INDICATES THE BEST PERFORMANCE

Settings	Set5			Set14			B100			Urban100			Manga109		
	×2	×3	×4	×2	×3	×4	×2	×3	×4	×2	×3	×4	×2	×3	×4
	AnySR Variants														
AnySR [†] (p=0)	38.01	34.45	32.37	33.73	30.45	28.77	32.18	29.13	27.64	32.25	28.44	26.44	38.77	33.77	30.82
AnySR [†] (p=0.2)	38.03	34.45	32.37	33.73	30.45	28.76	32.18	29.13	27.64	32.25	28.44	<u>26.45</u>	38.76	33.77	30.82
AnySR [†] (p=0.4)	38.03	34.47	32.37	33.72	30.45	28.77	<u>32.19</u>	29.13	<u>27.65</u>	32.26	28.43	26.44	38.77	33.77	<u>30.83</u>
AnySR [†] (p=0.6)	<u>38.04</u>	<u>34.50</u>	<u>32.38</u>	<u>33.74</u>	<u>30.47</u>	<u>28.78</u>	<u>32.19</u>	<u>29.15</u>	<u>27.65</u>	<u>32.27</u>	<u>28.45</u>	<u>26.45</u>	<u>38.78</u>	33.77	<u>30.83</u>
AnySR [†] (p=0.8)	38.03	34.49	<u>32.38</u>	33.73	30.46	28.77	32.18	29.14	<u>27.65</u>	32.26	28.44	<u>26.45</u>	<u>38.78</u>	<u>33.78</u>	30.82
	AnySR-Retrained Version														
AnySR [‡] (p=0)	38.06	34.47	32.37	33.76	30.46	28.77	32.21	29.14	27.64	32.48	28.46	26.44	38.91	33.82	30.82
AnySR [‡] (p=0.2)	38.07	34.47	32.37	33.78	30.46	28.76	32.22	29.14	27.64	32.50	28.47	<u>26.45</u>	38.93	33.83	30.82
AnySR [‡] (p=0.4)	38.08	34.48	32.37	33.78	30.46	28.77	32.22	29.14	<u>27.65</u>	32.50	28.47	26.44	38.94	33.83	<u>30.83</u>
AnySR [‡] (p=0.6)	<u>38.12</u>	<u>34.51</u>	<u>32.38</u>	<u>33.81</u>	<u>30.49</u>	<u>28.78</u>	<u>32.25</u>	<u>29.17</u>	<u>27.65</u>	<u>32.52</u>	<u>28.50</u>	<u>26.45</u>	<u>38.96</u>	33.83	<u>30.83</u>
AnySR [‡] (p=0.8)	38.11	<u>34.51</u>	<u>32.38</u>	33.80	<u>30.49</u>	28.77	32.24	29.16	<u>27.65</u>	32.51	28.49	<u>26.45</u>	<u>38.96</u>	<u>33.84</u>	30.82

slight variation, indicating that AnySR’s subnets generally maintain good perceptual quality in the generated images.

E. Ablation Studies

We conduct ablation studies to validate the effectiveness of individual components of AnySR. All ablation experiments are performed on SRNO [12], using EDSR [1] as the backbone. We compare our AnySR with two variants: 1) “w/o ASE”: AnySR without any-scale enhancement; 2) “w/o FI”: AnySR replacing feature-interweaving with a simple concatenation. Also, we vary the value of reset probability p in Line 4 of Algorithm 1, to show its importance.

1) *Any-Scale Enhancement*: Any-scale enhancement constitutes one of the fundamental branches encompassed within our overarching research framework. By injecting, excavating, and ameliorating sufficient scale information, we realize customized handling for features at different scales. To validate

the effectiveness of ASE, we train the network by removing this component, and the results are presented in Table VII. It is evident that “w/o ASE” leads to a certain performance drop.

2) *Feature-Interweaving*: Feature-interweaving considers the non-uniformity of shared weights and overcomes the mutual weight influence across different scales by repeating and inserting scale pairs at regular intervals. We study this mechanism by substitute the feature-interweaving fashion with a simple concatenation in earlier methods [10], [53]. With feature-interweaving, a better performance is achieved in Table VII.

3) *Reset Probability p* : As mentioned in Sec. III-B, we reset subnet F_i to the entire network F with a certain probability p to retain the original ability. We conduct ablation experiments with different values of $p = \{0, 0.2, 0.4, 0.6, 0.8\}$. Note that $p = 0$ indicates the removal of the reset mechanism. The results are presented in Table VIII. A low p , in particular for

$p = 0$, leads to minimal impact on enhancing the performance of the entire network during inference. Hence, opting for $p = 0.6$ or 0.8 is deemed more appropriate. Nevertheless, it is observed that with $p = 0.8$, the performance improvement for the entire network inference is slightly higher than $p = 0.6$, and a higher probability incurs higher training costs. Considering the trade-off between the performance of the entire network inference and network training costs, we ultimately select $p = 0.6$ as the experimental setting for the study presented in this paper.

V. CONCLUSION AND FUTURE WORK

In this paper, we presented AnySR, a simple yet versatile approach to transform existing arbitrary-scale super-resolution methods into implementations that adapt to any scale and resource availability. By rebuilding arbitrary-scale tasks into an any-resource implementation, we enable the completion of smaller-scale SISR tasks with reduced computational resources and no additional parameters. To maintain performance, we enhance any-scale features through a feature-interweaving fashion, ensuring sufficient scale information and correct feature/scale processing. Extensive experiments on benchmark datasets demonstrate the efficiency and scalability of our AnySR method in arbitrary-scale SISR applications.

An alternative approach is resorting to a more complex NAS (Neural Architecture Search), which may achieve better performance and will be our major future exploration.

REFERENCES

- [1] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1132–1140.
- [2] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis.*, 2018.
- [3] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2472–2481.
- [4] W. Shi et al., "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1874–1883.
- [5] X. Zhang, H. Zeng, and L. Zhang, "Edge-oriented convolution block for real-time super resolution on mobile devices," in *Proc. 29th ACM Int. Conf. Multimedia*, Oct. 2021, pp. 4034–4043.
- [6] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1646–1654.
- [7] S. Wang, J. Liu, K. Chen, X. Li, M. Lu, and Y. Guo, "Adaptive patch exiting for scalable single image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2022, pp. 292–307.
- [8] E. Agustsson and R. Timofte, "NTIRE 2017 challenge on single image super-resolution: Dataset and study," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1122–1131.
- [9] X. Hu, H. Mu, X. Zhang, Z. Wang, T. Tan, and J. Sun, "MetaSR: A magnification-arbitrary network for super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1575–1584.
- [10] Y. Chen, S. Liu, and X. Wang, "Learning continuous image representation with local implicit image function," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 8624–8634.
- [11] J. Lee and K. H. Jin, "Local texture estimator for implicit representation function," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 1919–1928.
- [12] M. Wei and X. Zhang, "Super-resolution neural operator," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 18247–18256.
- [13] G. Song, Q. Sun, L. Zhang, R. Su, J. Shi, and Y. He, "OPE-SR: Orthogonal position encoding for designing a parameter-free upsampling module in arbitrary-scale image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 10009–10020.
- [14] N. Kovachki et al., "Neural operator: Learning maps between function spaces," 2021, *arXiv:2108.08481*.
- [15] T. Liang, J. Glossner, L. Wang, S. Shi, and X. Zhang, "Pruning and quantization for deep neural network acceleration: A survey," *Neuro-computing*, vol. 461, pp. 370–403, Oct. 2021.
- [16] Y. Han, G. Huang, S. Song, L. Yang, H. Wang, and Y. Wang, "Dynamic neural networks: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 11, pp. 7436–7456, Nov. 2022.
- [17] A. Gholami, S. Kim, Z. Dong, Z. Yao, M. W. Mahoney, and K. Keutzer, "A survey of quantization methods for efficient neural network inference," in *Proc. Low-Power Comput. Vis. (LPCV)*, 2022, pp. 291–326.
- [18] G. Menghani, "Efficient deep learning: A survey on making deep learning models smaller, faster, and better," *ACM Comput. Surveys*, vol. 55, no. 12, pp. 1–37, Dec. 2023.
- [19] B. Chen et al., "Arm: Any-time super-resolution method," in *Proc. Eur. Conf. Comput. Vis.*, 2022, pp. 254–270.
- [20] Y. Zhang et al., "Real-time image demoiring on mobile devices," in *Proc. Int. Conf. Learn. Represent.*, 2023, pp. 1–13.
- [21] T. Yang et al., "MutualNet: Adaptive ConvNet via mutual learning from different model configurations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 811–827, Jan. 2023.
- [22] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.
- [23] C. Ledig et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 105–114.
- [24] L. Cavigelli, P. Hager, and L. Benini, "CAS-CNN: A deep convolutional neural network for image compression artifact suppression," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 752–759.
- [25] K. Zhang, Y. Li, W. Zuo, L. Zhang, L. Van Gool, and R. Timofte, "Plug-and-play image restoration with deep denoiser prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 6360–6376, Oct. 2022.
- [26] X. Wang et al., "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proc. Eur. Conf. Comput. Vis. Workshops*, 2018, pp. 1–16.
- [27] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 7, pp. 2480–2495, Jul. 2021.
- [28] Y. Chen and T. Pock, "Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1256–1272, Jun. 2017.
- [29] W. Cheng, M. Zhao, Z. Ye, and S. Gu, "MFAGAN: A compression framework for memory-efficient on-device super-resolution GAN," 2021, *arXiv:2107.12679*.
- [30] X. Deng, Y. Zhang, M. Xu, S. Gu, and Y. Duan, "Deep coupled feedback network for joint exposure fusion and image super-resolution," *IEEE Trans. Image Process.*, vol. 30, pp. 3098–3112, 2021.
- [31] X. Jiang et al., "FABNet: Frequency-aware binarized network for single image super-resolution," *IEEE Trans. Image Process.*, vol. 32, pp. 6234–6247, 2023.
- [32] C. Ma, P. Yu, J. Lu, and J. Zhou, "Recovering realistic details for magnification-arbitrary image super-resolution," *IEEE Trans. Image Process.*, vol. 31, pp. 3669–3683, 2022.
- [33] Q. Cai et al., "HIPA: Hierarchical patch transformer for single image super resolution," *IEEE Trans. Image Process.*, vol. 32, pp. 3226–3237, 2023.
- [34] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "SwinIR: Image restoration using Swin transformer," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 1833–1844.
- [35] Y. Xiao, Q. Yuan, K. Jiang, J. He, C.-W. Lin, and L. Zhang, "TTST: A top-k token selective transformer for remote sensing image super-resolution," *IEEE Trans. Image Process.*, vol. 33, pp. 738–752, 2024.
- [36] T. Dai, J. Cai, Y. Zhang, S.-T. Xia, and L. Zhang, "Second-order attention network for single image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 11057–11066.

- [37] B. Niu et al., "Single image super-resolution via a holistic attention network," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 191–207.
- [38] X. Chen, X. Wang, J. Zhou, Y. Qiao, and C. Dong, "Activating more pixels in image super-resolution transformer," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 22367–22377.
- [39] D. Liu, B. Wen, Y. Fan, C. C. Loy, and T. S. Huang, "Non-local recurrent network for image restoration," in *Proc. Adv. Neural Inform. Process. Syst.*, 2018, pp. 1673–1682.
- [40] Y. Mei, Y. Fan, and Y. Zhou, "Image super-resolution with non-local sparse attention," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 3516–3525.
- [41] J.-N. Su, M. Gan, G.-Y. Chen, W. Guo, and C. L. P. Chen, "High-similarity-pass attention for single image super-resolution," *IEEE Trans. Image Process.*, vol. 33, pp. 610–624, 2024.
- [42] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 391–407.
- [43] N. Ahn, B. Kang, and K.-A. Sohn, "Fast, accurate, and lightweight super-resolution with cascading residual network," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 252–268.
- [44] J. Liu, J. Tang, and G. Wu, "Residual feature distillation network for lightweight image super-resolution," in *Proc. Eur. Conf. Comput. Vis. Workshops*, 2020, pp. 41–55.
- [45] Z. Du, D. Liu, J. Liu, J. Tang, G. Wu, and L. Fu, "Fast and memory-efficient network towards efficient image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2022, pp. 852–861.
- [46] J. Zhang, Z. Wang, H. Wang, J. Zhou, and J. Lu, "Anycost network quantization for image super-resolution," *IEEE Trans. Image Process.*, vol. 33, pp. 2279–2292, 2024.
- [47] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep Laplacian pyramid networks for fast and accurate super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5835–5843.
- [48] Z. Hui, X. Gao, Y. Yang, and X. Wang, "Lightweight image super-resolution with information multi-distillation network," in *Proc. 27th ACM Int. Conf. Multimedia*, Oct. 2019, pp. 2024–2032.
- [49] F. Kong et al., "Residual local feature network for efficient super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2022, pp. 765–775.
- [50] X. Ding, X. Zhang, N. Ma, J. Han, G. Ding, and J. Sun, "RepVGG: Making VGG-style ConvNets great again," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 13728–13737.
- [51] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
- [52] L. Wang, Y. Wang, Z. Lin, J. Yang, W. An, and Y. Guo, "Learning a single network for scale-arbitrary super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 4781–4790.
- [53] J. Cao et al., "CiaoSR: Continuous implicit attention-in-attention network for arbitrary-scale image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 1796–1807.
- [54] M. Bevilacqua, A. Roumy, C. Guillemot, and M.-L.-A. Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in *Proc. Brit. Mach. Vis. Conf.*, 2012, pp. 135.1–135.10.
- [55] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Proc. Int. Conf. Curves Surf.*, 2012, pp. 711–730.
- [56] 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 *Proc. 8th IEEE Int. Conf. Comput. Vision. (ICCV)*, Jul. 2001, pp. 416–423.
- [57] J.-B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 5197–5206.
- [58] K. Aizawa et al., "Building a manga dataset 'manga109' with annotations for multimedia applications," *IEEE Multimedia*, vol. 27, no. 2, pp. 8–18, Jun. 2020.
- [59] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
- [60] R. Timofte et al., "NTIRE 2017 challenge on single image super-resolution: Methods and results," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1110–1121.
- [61] A. K. Jain, *Fundamentals of Digital Image Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, 1989.
- [62] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 586–595.
- [63] R. Timofte, V. De, and L. V. Gool, "Anchored neighborhood regression for fast example-based super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1920–1927.



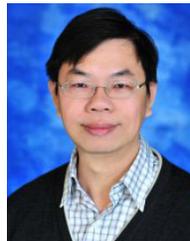
Wengyi Zhan received the bachelor's degree in intelligence science and technology from Xiamen University, Xiamen, China, in 2023, where she is currently pursuing the master's student with the School of Informatics, specializing in intelligence science and technology.

Her current research interests include diffusion models for consistent video/image generation.



Mingbao Lin received the B.S. degree from Fuzhou University, Fuzhou, China, in 2016, and the joint M.S. and Ph.D. degree in intelligence science and technology from Xiamen University, Xiamen, China, in 2022.

He is currently a Research Scientist with Skywork AI, Singapore, and an Adjunct Industry Supervisor with Xiamen University. His current research interests include developing low-latency multimodal interaction systems, such as text, audio, image, and video.



Chia-Wen Lin (Fellow, IEEE) received the Ph.D. degree in electrical engineering from National Tsing Hua University (NTHU), Hsinchu, Taiwan, in 2000. He is currently a Distinguished Professor with the Department of Electrical Engineering and the Institute of Communications Engineering, NTHU. His research interests include image/video processing and computer vision. He served as a Fellow Evaluation Committee Member (2021–2023), a BoG Member-at-Large (2022–2024), and a Distinguished Lecturer (2018–2019) of the IEEE Circuits and Systems Society. He received two best paper awards from VCIP in 2010 and 2015. He was the Chair of the IEEE ICME Steering Committee (2020–2021). He served as the TPC Co-Chair for IEEE ICIP 2019, IEEE ICME 2010, and PCS 2022; and the General Co-Chair for IEEE VCIP 2018 and PCS 2024. He served as an Associate Editor for IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, IEEE TRANSACTIONS ON MULTIMEDIA, IEEE MULTIMEDIA, and *Neural Networks*.



Rongrong Ji (Senior Member, IEEE) is currently a Nanqiang Distinguished Professor with Xiamen University and the Director of the Office of Science and Technology, Xiamen University, and the Analytics and Computing Laboratory. He has published more than 50 articles in ACM/IEEE TRANSACTIONS, including IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE and IJCV; and more than 100 full papers on top-tier conferences, such as CVPR and NeurIPS. His publications have got over 20K citations in

Google Scholar. His research interests include computer vision, multimedia analysis, and machine learning. He is an Advisory Member for Artificial Intelligence Construction in the Electronic Information Education Committee of the National Ministry of Education. He was a recipient of the Best Paper Award of ACM Multimedia in 2011. He was awarded the National Science Foundation for Excellent Young Scholars in 2014, the National Ten Thousand Plan for Young Top Talents in 2017, and the National Science Foundation for Distinguished Young Scholars in 2020. He served as the Area Chair for top-tier conferences, such as CVPR and ACM Multimedia.