Boosting Single Image Super-Resolution via Partial Channel Shifting

Xiaoming Zhang, Tianrui Li, Xiaole Zhao*
Southwest Jiaotong University, China

Abstract

Although deep learning has significantly facilitated the progress of single image super-resolution (SISR) in recent years, it still hits bottlenecks to further improve SR performance with the continuous growth of model scale. Therefore, one of the hotspots in the field is to construct efficient SISR models by elevating the effectiveness of feature representation. In this work, we present a straightforward and generic approach for feature enhancement that can effectively promote the performance of SR models, dubbed partial channel shifting (PCS). Specifically, it is inspired by the temporal shifting in video understanding and displaces part of the channels along the spatial dimensions, thus allowing the effective receptive field to be amplified and the feature diversity to be augmented at almost zero cost. Also, it can be assembled into off-the-shelf models as a plug-and-play component for performance boosting without extra network parameters and computational overhead. However, regulating the features with PCS encounters some issues, like shifting directions and amplitudes, proportions, patterns of shifted channels, etc. We impose some technical constraints on the issues to simplify the general channel shifting. Extensive and throughout experiments illustrate that the PCS indeed enlarges the effective receptive field, augments the feature diversity for efficiently enhancing SR recovery, and can endow obvious performance gains to existing models.

1. Introduction

Single image super-resolution (SISR) is a long-standing and challenging hotspot in the computer vision community, to reconstruct a high-resolution (HR) image from one low-resolution (LR) image. Since an LR image can be degraded from an infinite number of HR images, it is substantially a mathematically ill-posed inverse problem. Convolutional neural networks (CNNs) [25, 24] have demonstrated significant breakthroughs in recent years, by leveraging large-scale datasets, a highly nonlinear mapping from LR and HR image pairs is learned. As a pioneering contribution, Dong *et al.* [7] draw lessons from sparse representation [51, 52] and propose a super-resolution convolutional neural network (SRCNN) to learn an end-to-end nonlinear mapping from LR to HR images. This model is further extended from the perspectives of sparse coding [48], increasing network depth [20] and sharing parameters [21, 42], *etc*. Although these models are capable of reconstructing high-fidelity SR results, their achievements rely immensely on the upscaling of the model scale, including network parameters, depth, and subtle topologies. For instance, EDSR [30] maintains more than 43M network parameters and about 70 layers of network depth. While RCAN [58] facilitates the effectiveness of feature representation via attention mechanism [14] and reduces the parameters to ∼16M, it reaches over 400 layers of depth and significantly lengthens the time of inference. However, simply increasing the model scale also hits the bottleneck of performance improvement, and it is difficult to elevate the utilization efficiency of network assets. Moreover, model training also suffers from the intractable holdback of information weakening as the model scale continues to grow. Therefore, a promising compro-

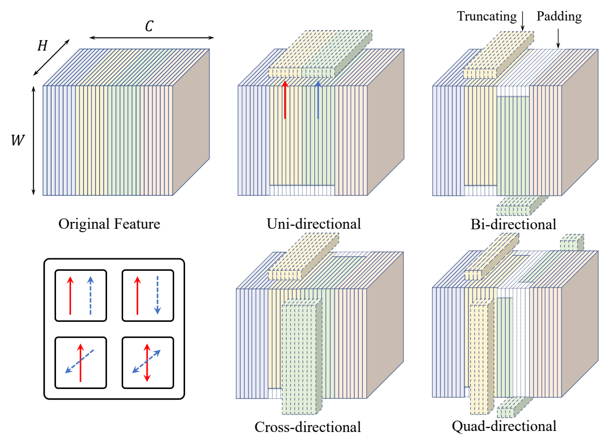

Figure 1: Representative shifting patterns of applying the **PCS** to intermediate features, which show different spatial misalignments that enhance the Effective Receptive Field (ERF) [34] and feature diversity of the models. The pixels removed from the feature grids are discarded outright, and the emptied grids are simply padded with zeros to maintain the grid structure. These misalignments are conducive to expanding the local association among intermediate features while maintaining spatial integrity.

*Corresponding author, email: zxlation@foxmail.com
Code is available at https://github.com/OwXiaoM/PCS

mise can probably be to promote the effectiveness of intermediate features, just like RCAN [58], which is intuitively helpful for further improving the upper bound of SR performance and developing lightweight SR models suitable for real-world scenarios.

At present, the measures to enhance the effectiveness of features mainly include feature fusion [16, 2, 60], wide activation [40, 53], gating mechanism [14, 58, 61] and nonlocality [46, 59, 6]. The most commonly-used trick may be the attention mechanism [14] that adjusts the feature responses towards the most informative and important components of the inputs. The SR community has fully embraced this approach and extended it in various innovative ways [15, 59, 39, 38, 3]. The ideas of these methods mainly focus on adapting the channel-wise attention to other dimensions, such as spatial-wise or layer-wise attention, or exploring the interaction of the attention mechanism [14] in different dimensions. To accelerate model inference, it can also be improved from the perspective of sparse representation [51, 52]. Transformer [44] is a more advanced architecture based on the self-attention mechanism, and has recently been adopted for the tasks of low-level computer vision [29] and made remarkable success. *Although these approaches have made great progress in improving the effectiveness of features, they always introduce extra cost (e.g., network depth or parameters) compared with the benchmark and compromise with the loss of efficiency*.

Both enlarging the scale of SR models and enhancing the effectiveness of intermediate features aim to achieve more accurate nonlinear inference for SR tasks, nevertheless, they are usually accompanied by additional overheads. Different from these two manners, we expect to improve the nonlinear inference of SR models with zero cost by enlarging the receptive field of local mapping and increasing the diversity of hierarchical features, which is simply implemented by a generic operation termed partial channel shifting (**PCS**). Specifically, it displaces a part of the feature channels along two spatial dimensions (H and W), so that the originally aligned features present certain dislocation in spatial dimensions, as shown in Fig. 1. The elements removed from the feature grids are directly discarded, while the emptied grids are simply padded with zeros. However, naive shifting suffers from the problem of infinite enumeration, such as shifting directions and magnitudes, the proportions of the shifted channels, and the combinational patterns of unshifted and shifted channels, *etc*. Therefore, we further investigate some technical constraints of these issues to simplify the application of naive shifting, which results in four PCS patterns as illustrated in Fig. 1. Our benchmark experiments demonstrate the synergistic effect of the PCS on existing SR models. We also use analytical tools [34, 9] to illustrate the role of PCS in enhancing feature effectiveness. In summary, the main contributions of this work include:

(**1**) We propose a simple and effective method of feature enhancement for SISR models, termed as **PCS**, which avails amplifying the Effective Receptive Fields (ERFs) [34] and augmenting the feature diversity without extra parameters and computational overhead.

(**2**) We discuss several technical constraints to mitigate the problem of infinite enumeration, which streamlines the naive shifting and finally distills the state space down to four representative patterns for channel shifting.

(**3**) The benefits of the PCS for model inference and feature enhancement is experimentally analyzed and verified with the assistance of Effective Receptive Fields (ERFs) and Local Attribution Maps (LAM) [9].

(**4**) The benchmark experiments that apply the PCS to off-the-shelf SISR models, including typical lightweight and large-scale ones, provide strong evidence of its effectiveness in promoting SR performance.

2. Related Work

2.1. CNN-based SISR Methods

High-Fidelity Models Since Dong *et al.* put forward the first CNN-based SR model [7], various deep SR models have emerged in the literature. The improvements based on the pioneering work mainly focused on increasing network depth [20, 43] and sharing network parameters [21, 42]. Although the latter can prevent the expansion of model scale, it can not effectively reduce the computational burden due to the recursive computation. An obvious boost in SR fidelity comes from EDSR [30], a CNN model based on modularized residual learning [10, 11], which increases the network depth to nearly 70 layers and parameters up to 43M. This hits the scale bottleneck of general image SR and researchers turn to more effective network topologies, such as RDN [60], RCAN [58], HAN [39], IGNN [63], SAN [6], and NLSN [38], *etc*. Although these large-scale models show performance improvements with parameter decrease, they also pay expenses for, *e.g.*, network depth, global computation, and attention mechanism, which lead to the decline in the efficiency of model inference.

High-Efficiency Models Apart from the difficulty in further expanding the model scale to improve the SR performance of large-scale models, it is also intractable to deploy them in practical scenarios. Lightweight SR methods with high-efficiency feature representation are studied and proposed to ease this problem [2, 26, 18, 27, 35, 23, 50]. These approaches are mainly designed to elevate the compromise between SR performance and model overhead, which is closely related to the effectiveness of feature representation. However, they typically suffer from a limited receptive field due to the locality of convolution operations. This

Note we refer to channel shifting as shifting feature maps on different channels along their spatial dimensions, which is different from [13].

(a) DRRN [42] with PCS.

(b) EDSR [30] with PCS.

Figure 2: Schematic diagram of the PCS applied to the existing SR models. (a) DRRN [42] is a representative **lightweight** model with recursive structure, while (b) EDSR [30] is a typical **largescale** model with modularized residual learning. For EDSR [30], the PCS is embedded into each residual block for periodic feature enhancement. Please note that our PCS does not introduce extra parameters and computation other than negligible overhead for data movement.

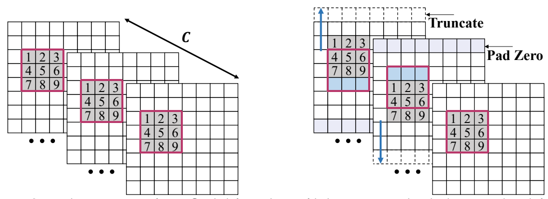

Figure 3: The receptive field is plausibly extended through shifting a part of channels. **Left**: The spatially aligned features provide a regular 3×3 receptive field for 3×3 conv kernels; **Right**: The conv kernels "see" a larger range via staggered feature maps.

issue can be relieved via applying various techniques like non-locality [59, 32, 38, 6] and feature fusion [16, 2, 60], whereas they usually prioritize global long-range computation at the expense of efficiency, and easily perturb the correlation between pixels in local receptive field [32].

2.2. Feature Shifting

Designing compact models with reduced parameters and lower computational complexity is a long-standing issue in the computer vision and deep learning communities. Different from the common-used convolutions (including depthwise separable [5] and pointwise group [57] convolutions), feature shifting [49] aims to sidestep extra parameters and computation entirely, therefore surpassing the sophisticated designs like gating mechanism [14, 58, 61] in terms of improving the tradeoff between efficiency and performance. It has been instantiated as Temporal Shift Module (TSM) [31], ShiftNet [49], and AddressNet [13] for various high-level vision tasks. However, it is rarely used for low-level visual tasks that require accurate and intensive prediction, such as image SR. In this work, we simplify the generic feature shifting by imposing constraints to pace it in SR tasks.

3. Partial Channel Shifting

3.1. Intuition

The receptive field contributes to the effectiveness of the local mapping of neural networks by providing available information within a local range. Generally, a larger receptive field indicates a more accurate local mapping, which can be confirmed by the usual settings of training SR models with image patches of different sizes [29, 8]. Shifting a part of feature channels favors the subsequent mapping to extract information from a larger spatial range, as demonstrated in Fig. 3. We will experimentally show that the effective receptive field can be directly affected by channel shifting, which defers common sense.

On the other hand, the augmented receptive field counts for the final prediction by combining multiple regular receptive fields induced by channel shifting. This indicates that the shifting operation can serve as an approach of feature augmentation, which is critical for low-level vision tasks like image SR, as it provides diversified input for subsequent inference. In Sec. 4.4, we also experimentally verify that more spatial information is explored via displacing a part of feature channels when reconstructing the same target area of an HR image.

3.2. Holdbacks to Shift Low-Level Features

Spatial Locality Images usually show spatial locality in structure: *image pixels are closely related to their neighbors, while they are weakly connected to those far away*. If features are shifted with large magnitudes, it may harm the spatial locality of the image and lead to severe performance degradation of pixel-wise recovery, which is undesirable in low-level vision tasks (*e.g.*, image SR). — H1

Boundary Information As previously mentioned, the shift operation directly truncates the elements removed from the feature grids, while the emptied grids are simply padded with zeros, which is mainly in consideration of efficiency and convenience of implementation. However, simple truncation and zero-padding can lead to the loss of important boundary information in partial channels. Intuitively, it is unfavorable for pixel-level recovery tasks that demand high levels of accuracy and dense prediction. — H2

Data Movement Channel shifting requires moving the data of partial channels along feature grids. Although moving feature data is hardware friendly, it still hinders the efficient inference of deep models, especially for image SR which requires upscaling input LR images and a large amount of data movement. Except for the simple and efficient truncation and zero-filling, the inference efficiency for SR models with channel shifting needs further consideration. — H3

Infinite Enumeration Problem Despite the shift being only performed on a subset of channels, it still exhibits a huge state space. Consider a feature $\mathbf{x}$ with the size of $H \times W \times C$, where H and W denote spatial dimensions and C is the number of channels. If the shifting operation on spatial dimensions is directional, it gives $[(2H+1)\times(2W+1)]^C$ possible shifts. It approaches an infinite enumeration to exhaustively search over this state space for the optimal shift. Besides, it is hard to determine the optimal permutation of shifted and unshifted channels. Also, it cannot be directly learned from model training as permutation operations are always discrete and intricate to optimize [19, 49]. — H4

The above problems demonstrate that it is intractable to directly apply channel shifting, which is extensively proven to be effective in high-level visual tasks, to image SR.

3.3. Constrained Channel Shifting

3.3.1 Formulation

Given a intermediate feature $\mathbf{x} \in \mathbb{R}^{H\times W\times C}$, the operation for channel shifting is defined as:

$$\tilde{\mathbf{x}} = \mathcal{S}_c(\mathbf{x}, \boldsymbol{\pi}), \tag{1}$$

where $\mathcal{S}_c(\cdot)$ denotes the operation of channel shifting, and $\tilde{\mathbf{x}}$ represents the result feature with the same shape as $\mathbf{x}$ due to truncation and zero-padding. $\boldsymbol{\pi}$ indicates a shifting pattern for the entire feature $\mathbf{x}$, which is a tuple of length C:

$$\boldsymbol{\pi} = \{\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots, \boldsymbol{\pi}_C\}. \tag{2}$$

If we rewrite $\mathbf{x}$ as $[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_C]$, where $\mathbf{x}_k (1 \le k \le C)$ denotes the k-th channel of $\mathbf{x}$, then $\boldsymbol{\pi}_k$ implies the shifting pattern for channel $\mathbf{x}_k$. Considering two spatial dimensions and their shifting magnitudes, $\boldsymbol{\pi}_k$ can be a bivariate tuple:

$$\boldsymbol{\pi}_k = \{h_k, w_k\}, \quad 1 \le k \le C, \tag{3}$$

where h_k and w_k denote the movement magnitudes in H and W directions, respectively, and satisfy $-H \le h_k \le H$ and $-W \le w_k \le W$, where the sign indicates the direction of movement. In this way, the subsequent inference will take $\tilde{\mathbf{x}}$ as the input to obtain plausibly enhanced receptive field and augmented feature diversity.

3.3.2 Constraints

We have to limit the state space via restraining $\boldsymbol{\pi}$, according to H4. Nevertheless, more details need to be explicitly highlighted with the consideration of H1 $\sim$ H4 discussed in Sec. 3.2. To this end, we need to make decisions regarding the following three aspects:

Shifting Magnitude: It can be seen from Eqn.(3) that there are $(2H+1)\times(2W+1)$ possible shifts for a feature map of a single channel without any restrictions (including $h_k = w_k = 0$). We set a threshold, $\mathbf{T_m}$, for shifting magnitude to reduce the possible shifts on a single channel, where $0 < \mathbf{T_m} < H$ and $0 < \mathbf{T_m} < W$. Thus $\boldsymbol{\pi}_k$ can be rewritten as:

$$\boldsymbol{\pi}_k = \{h_k, w_k\}, \quad |h_k| < \mathbf{T_m}, \;\; |w_k| < \mathbf{T_m}, \tag{4}$$

where $1 \le k \le C$. Eqn.(4) produces $(2\mathbf{T_m}+1)^2$ feasible $\boldsymbol{\pi}_k$ and lessens the state space for a single channel by a large margin if $\mathbf{T_m} \ll H$ and $\mathbf{T_m} \ll W$. More importantly, it can effectively counteract the potentially detrimental effects of channel shifting on spatial locality (H1).

Shifted Channels: For efficiency reasons, shifting all channels of an intermediate feature is not recommended. Therefore, we also limit the number of shifted channels through a threshold $\mathbf{T_c}$, where $0 < \mathbf{T_c} < C$. To better describe shifting partial channels when C varies, we define the channel shifting ratio as $\gamma = N/C$, where $N \le \mathbf{T_c}$ denotes the number of shifted channels for C-channel features. $\mathbf{T_c}$ and γ also assist in mitigating the loss of boundary information by retaining the boundaries (H2) in unshifted channels. Since $\gamma < 1$, the method is termed as *partial* channel shifting (PCS). The set of the indexes indicating the shifted channels is denoted as $\mathcal{I} = \{1, 2, \ldots, N\}$.

Permutation: The next important issue is to determine the permutation for shifted and unshifted channels. Unlike [49] which introduces extra parameters to learn the optimal permutation in an end-to-end manner, we do not expect our PCS to have additional overhead, and so we decide it by simple heuristics. Although the permutation seems to have impacts on model performance as different channels learn different features, our experiments show a counterintuitively random effect of permutation on model performance. Taking spatial locality (H1) and data movement (H3) into account, we do a very extreme simplification of the permutation: all shifted channels are grouped and only the permutations of these groups are considered, as shown in Fig. 4. For instance, the simplest permutation can be that all shifted channels are assembled together and centered:

$$\boldsymbol{\pi} = \{\overbrace{0, \ldots, 0}^{(C-N)/2}, \overbrace{\hat{\boldsymbol{\pi}}_1, \ldots, \hat{\boldsymbol{\pi}}_k, \ldots, \hat{\boldsymbol{\pi}}_N}^{N}, \overbrace{0, \ldots, 0}^{(C-N)/2}\}, \tag{5}$$

where 0 denotes the unshifted channels or the shifting magnitudes (h_k and w_k in Eqn.(3)) are equal to 0, and $\hat{\boldsymbol{\pi}}_k$ is the channel with $h_k \neq 0$ and/or $w_k \neq 0$. Despite the simple settlement for the permutation, it still exhibits observable performance improvements for SR tasks.

3.4. Application to Existing SR Models

To confirm the effectiveness of our PCS in enhancing the low-level features, we directly apply it to two off-the-shelf SR models, *i.e.*, DRRN [42] and EDSR [30]. Although their performance is not currently optimal, they are representative of the SR community: the former is a **lightweight** SR model with recursive structure and the latter is a **largescale** model with modularized residual learning [10]. The PCS is embedded in the front of the typical local mapping of both models for feature enhancement, as illustrated in Fig. 2. We also supplement the experiments with the lightweight version of EDSR [30], which has only ~3.48% parameters of

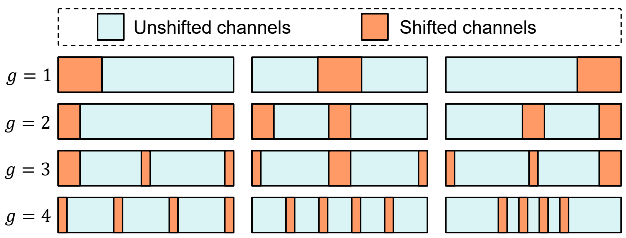

Figure 4: **Permutation**: shifted channels are simply clustered into g groups. However, we only consider the simplified permutations of these groups, since we find that more complex permutations do not necessarily result in noticeable performance gains, but significantly degrade the efficiency of the model.

the largescale version. It is worth noting that, except for the negligible overhead caused by data movement, our PCS does not exert extra parameters or computational burden on these models, nor does it change their network topologies.

To verify the generality of PCS, in the supplementary material, we apply PCS on the recent **state-of-the-art** lightweight and largescale SISR models, including RFDN [33], BSRN [28], VAPSR [62], OISR [12], SAN [6] and NLSN [38], with also impressive gains of performance.

4. Experiments

4.1. Datasets and Metrics

Following [58, 30, 60], we use 800 training images from DIV2K [1] as the training dataset and 5 standard benchmark datasets: Set5 [4], Set14 [54], B100 [36], Urban100 [17] and Manga109 [37] for testing. They comply with the most extensive experimental paradigm and are consistent with the benchmark models. For quantitative assessment, the reconstructed results are measured with PSNR and SSIM [47] metrics on $\mathbf{Y}$ channel of the YCbCr color space.

4.2. Implementation and Training Details

During training, we randomly crop 48×48 patches from the training examples and form a mini-batch of 16 images. The training images are further augmented via horizontal flipping and random rotation of 90, 180, and 270 degrees. We optimize the model by ADAM optimizer [22] with $\beta_1 = 0.9$, $\beta_2 = 0.99$ and $\epsilon = 10^{-8}$. We adopt the common-used L_1 loss to guide model training. The learning rate is initially set to 0.0001 for all layers and halved at every 200 epochs, *i.e.*, it decays in a piece-wise constant manner. We apply a single non-learnable convolution to implement the partial channel shifting, compared with the computational cost of the model itself, its extra memory cost is negligible.

In the subsequent experiments, we retrain DRRN-based models for 600 epochs and EDSR-based models for 1000 epochs. All our models are reimplemented using PyTorch 1.1 and retrained on NVIDIA TITAN-Xp GPUs.

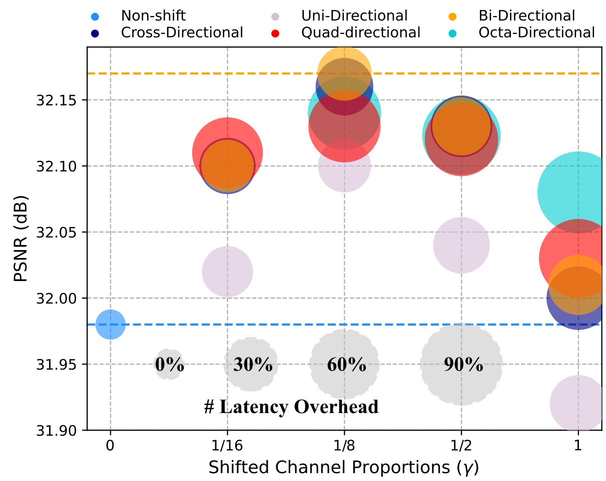

Figure 5: The compromise between typical shifting patterns with EDSR(S, PCS) on Urban100 [17] ($\mathbf{T_m} = 2$, SR$\times$2). We use the latency overhead of model training for comparison and the training time of EDSR(S) as the benchmark. The total training time is reported as the latency overhead when using the vanilla shift. Please note that the octa-directional shift is unavailable for $\gamma = 1/16$, since there are only $64 \times (1/16) = 4$ shifted channels in this case.

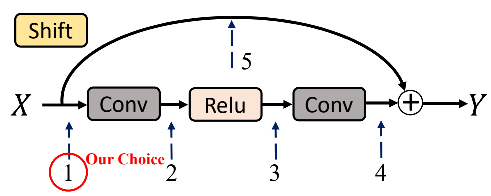

Figure 6: Optional probes of PCS embedded in the building block of EDSR [30]. Intuitively, position **1** should be the best choice as a building block corresponds to a local inference, which is also consistent with our experimental results.

4.3. Ablation Study

Considering the holdbacks in Sec. 3.2, it is desirable to explore feasible strategies to apply the lean **PCS** to image SR tasks, which is capable of boosting SR performance at zero cost except for minimal data movement. It differs from previous work [49, 13, 31, 45, 55] that simply take **PCS** as a trick of feature enhancement for their models.

4.3.1 Shift Patterns

As it is impossible to consider all the enumerations of H4, and the receptive field is impacted by the shifting directions, we analyze according to the shifting channel covering one, two, four, and eight possible directions (W, H, and diagonals) in Table 1, so that the receptive field is expanded from more directions in turn. The experimental results show that shifting directions in specific spatial directions cause only a small difference, and we conjuncture it depends mainly on the distribution and texture of the target datasets, which also demonstrates that moving more channels to cover more possible spatial directions doesn't significantly affect the per-

Table 1: The performance comparison of different channel shifting covering various possible directions of augmenting the **ERFs**. The model is tested on B100 [36] for SR$\times$2, with $\mathbf{T_m} = 1$ and $\gamma = 1/8$. $+/-$ denote two opposite directions, and wh represents the diagonal directions between W and H (PSNR).

w^+	w^-	h^+	h^-	w^+h^+	w^-h^-	w^+h^-	w^-h^+	B100 [36] DRRN	B100 [36] EDSR(S)
								31.99	32.16
✓								32.00	32.18
	✓							32.02	32.18
		✓						32.00	32.18
			✓					32.00	32.19
				✓				32.01	32.18
					✓			32.00	32.17
						✓		32.01	32.18
							✓	32.00	32.18
✓	✓							32.02	32.19
		✓	✓					32.01	32.18
				✓	✓			31.99	32.18
						✓	✓	32.00	32.20
✓		✓						32.02	32.18
✓			✓					32.00	32.19
	✓	✓						32.01	32.19
	✓		✓					32.01	32.19
				✓		✓		32.01	32.18
				✓			✓	32.00	32.19
					✓	✓		31.99	32.18
					✓		✓	31.99	32.18
✓	✓	✓	✓					32.02	32.17
				✓	✓	✓	✓	32.02	32.18
✓	✓	✓	✓	✓	✓	✓	✓	32.00	32.19

Table 2: PSNR comparison of channel permutation. Corresponding to Fig. 4, we test the performance of DRRN(PCS) / EDSR(S, PCS) with quad-directional shifts ($|h_k| = |w_k| = 2, k \in \mathcal{I}$, $\gamma = 1/8$) on B100 [36] (SR$\times$2).

$g = 1$	32.00 / 32.18	32.01/ 32.18	32.00 / 32.19
$g = 2$	31.99 / 32.20	31.99 / 32.19	31.99 / 32.19
$g = 3$	31.98 / 32.18	31.99 / 32.19	31.99 / 32.19
$g = 4$	31.99 / 32.18	32.00 / 32.19	31.99 / 32.19

Table 3: The probes of **PCS** (Fig. 6) show significant performance differences tested on Urban100 [17] (SR$\times$2).

Position	-	1	2	3	4	5
PSNR	31.99	**32.17**	32.12	32.14	32.05	32.09

formance of the SR model. Therefore, we speculate that more complex permutations on shifting directions are unlikely to achieve significant performance gains.

Table 2 compares the simplified permutations of shifted and unshifted channels, where we can not observe distinct differences between these permutations. We thereby choose the simplest permutation that all shifted channels are assembled into $g = 1$ group and centered, for simplicity. As a result, we utilize the approach depicted in Fig. 1 to manage the shifting directions on the channel.

It seems that shifted channels covering more directions contribute to increasing the performance of the model because they expand the ERFs to a greater extent. However, we can observe the opposite of this intuition from Fig. 5, *i.e.*, the bi-directional shift covering only two directions shows better SR performance than the octa-directional shift covering eight directions, while the latter goes with more latency overhead. The counterintuitive observation may imply that more staggered displacement of channels is also harmful to the spatial locality of the image.

4.3.2 Shifting Magnitude

The shifting magnitude of **PCS** should be limited within a certain range according to H1 $\sim$ H3. We investigate several magnitudes of a bi-directional shift with EDSR(S) [30], as shown in Fig. 8(a). We can find that: (1) the PCS endows obvious improvement on the performance of the SR model regardless of the values of γ and $\mathbf{T_m}$; (2) A magnitude of 2 can robustly improve the performance in our experimental setup; (3) The further enlargement in the magnitude exerts a random impact on the performance; (4) There is a high probability that the model will be degraded if the magnitude is very large (*e.g.*, $|h_k|$ or $|w_k| = 8$).

These observations demonstrate that: (1) When the magnitude is small, the positive effect of increasing ERFs is dominant, thereby improving the SR performance; (2) Large magnitudes may destroy the spatial locality and negate the positive impact of ERF enlargement; (3) The problem of spatial locality and boundary information loss becomes severe when the shifting magnitude is very large, resulting in a loss of important contextual information and leading to the degradation of model performance.

In summary, we set $|h_k|$ or $|w_k| = 2$ as suggested in Fig. 8(a) for the follow-up experiments.

4.3.3 Shifted Channels

We use bi-directional shift ($|h_k| = 2, |w_k| = 0, k \in \mathcal{I}$) to assess how shifted channel proportions affect the performance of SR models. As shown in Fig. 8(b): the shifted channel proportion doesn't obviously impact the models, especially for the very deep CNNs (*e.g.*, EDSR(L)). Meanwhile, considering H_3, we set γ as $1/16$ for DRRN(PCS), $1/8$ for EDSR(S, PCS) and $1/16$ for EDSR(L, PCS).

4.3.4 Position of PCS in Residual Block

Possible positions of channel shifting embedded in the residual block of EDSR in Fig. 6, which is a typical local mapping of deep SR models. Our experiments in Table 3 show that the best performance gains can be achieved by embedding channel shifting at the very front of the local mapping, which means the early feature shifting in the Residual Connection is beneficial for the enhancement of the SR model. Note that we do not consider the position

Table 4: The comparison of quantitative results on benchmark datasets. † denotes the reimplementation of the model on DIV2K [1], and Δ refers to the **improvement** by applying PCS.

Method	Scale	Params.	Set5 [4]		Set14 [54]		B100 [36]		Urban100 [17]		Manga109 [37]	
			PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
DRRN [42]†	2	0.30M	37.76	0.9597	33.32	0.9151	31.99	0.8973	31.32	0.9199	37.86	0.9755
DRRN(PCS)	2	0.30M	37.81	0.9597	33.36	0.9155	32.04	0.8977	31.51	0.9216	38.08	0.9759
Δ			**+0.05**	**+0.0000**	**+0.04**	**+0.0004**	**+0.05**	**+0.0004**	**+0.19**	**+0.0017**	**+0.22**	**+0.0004**
EDSR(S) [30]	2	1.37M	37.99	0.9604	33.59	0.9175	32.16	0.8994	31.98	0.9272	38.55	0.9769
EDSR(S, PCS)	2	1.37M	38.01	0.9605	33.60	0.9175	32.19	0.8997	32.17	0.9287	38.61	0.9770
Δ			**+0.02**	**+0.0001**	**+0.01**	**+0.0000**	**+0.03**	**+0.0003**	**+0.19**	**+0.0009**	**+0.06**	**+0.0001**
EDSR(L) [30]	2	40.73M	38.11	0.9602	33.92	0.9195	32.32	0.9013	32.93	0.9351	39.10	0.9773
EDSR(L, PCS)	2	40.73M	38.24	0.9613	34.04	0.9209	32.36	0.9019	32.99	0.9361	39.24	0.9783
Δ			**+0.13**	**+0.0011**	**+0.12**	**+0.0004**	**+0.04**	**+0.0006**	**+0.06**	**+0.0010**	**+0.12**	**+0.0010**
DRRN [42]†	3	0.30M	34.06	0.9246	30.06	0.8367	28.92	0.8003	27.61	0.8392	32.73	0.9392
DRRN(PCS)	3	0.30M	34.16	0.9252	30.10	0.8374	28.95	0.8008	27.70	0.8411	32.95	0.9405
Δ			**+0.10**	**+0.0006**	**+0.04**	**+0.0007**	**+0.03**	**+0.0005**	**+0.03**	**+0.0019**	**+0.22**	**+0.0013**
EDSR(S) [30]	3	1.55M	34.37	0.9270	30.28	0.8418	29.09	0.8052	28.15	0.8527	33.45	0.9439
EDSR(S, PCS)	3	1.55M	34.41	0.9272	30.33	0.8424	29.11	0.8055	28.23	0.8541	33.59	0.9448
Δ			**+0.04**	**+0.0002**	**+0.05**	**+0.0006**	**+0.02**	**+0.0003**	**+0.08**	**+0.0014**	**+0.14**	**+0.0009**
EDSR(L) [30]	3	43.68M	34.65	0.9280	30.52	0.8462	29.25	0.8093	28.80	0.8653	34.17	0.9476
EDSR(L, PCS)	3	43.68M	34.76	0.9299	30.56	0.8469	29.28	0.8101	28.88	0.8672	34.21	0.9489
Δ			**+0.11**	**+0.0019**	**+0.04**	**+0.0007**	**+0.03**	**+0.0008**	**+0.08**	**+0.0019**	**+0.04**	**+0.0013**
DRRN [42]†	4	0.30M	31.81	0.8904	28.34	0.7756	27.39	0.7295	25.58	0.7680	29.65	0.8978
DRRN(PCS)	4	0.30M	31.96	0.8921	28.38	0.7769	27.42	0.7307	25.65	0.7707	29.86	0.9007
Δ			**+0.15**	**+0.0017**	**+0.04**	**+0.0013**	**+0.03**	**+0.0012**	**+0.07**	**+0.0027**	**+0.21**	**+0.0029**
EDSR(S) [30]	4	1.52M	32.09	0.8938	28.58	0.7813	27.57	0.7357	26.04	0.7849	30.35	0.9067
EDSR(S, PCS)	4	1.52M	32.20	0.8949	28.61	0.7826	27.58	0.7361	26.13	0.7881	30.51	0.9087
Δ			**+0.11**	**+0.0011**	**+0.03**	**+0.0013**	**+0.01**	**+0.0004**	**+0.09**	**+0.0032**	**+0.16**	**+0.0010**
EDSR(L) [30]	4	43.10M	32.46	0.8968	28.80	0.7876	27.71	0.7420	26.64	0.8033	31.02	0.9148
EDSR(L, PCS)	4	43.10M	32.51	0.8989	28.83	0.7881	27.72	0.7423	26.69	0.8053	31.10	0.9167
Δ			**+0.05**	**+0.0021**	**+0.03**	**+0.0005**	**+0.01**	**+0.0003**	**+0.05**	**+0.0020**	**+0.08**	**+0.0019**

before the identity connection, as it is not conducive to protecting the complete structure of the feature. As identity mapping can effectively avoid the disturbance of H2, we thus conjuncture that the subsequent nonlinear mapping of the channel shifting should be sufficient and compatible to complement the identity mapping effectively.

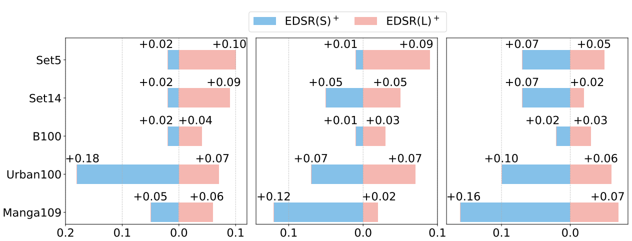

Figure 7: PSNR improvement by the enhancements of PCS on EDSR$^+$, from left to right are SR ($\times 2$, $\times 3$, $\times 4$) respectively.

4.4. Analyses of ERFs and LAMs

The Effective Receptive Field [34] (ERFs) measure the actual receptive field for CNNs. As Fig. 9 demonstrates that when applying bi-directional shift, PCS-enhanced methods bring obvious enlargement to the ERFs.

We also apply Local Attribution Maps (LAM) [9] to examine the involved range of utilized information for the reconstruction of a target area. Fig. 10 illustrates that the PCS-enhanced models activate more spatial information along the shifting directions, while ensuring that local textures are consistent with the direction of channel shifting.

4.5. Benchmark results

The DRRN [42]† is with 1 recursive block, in which 9 residual units are stacked with 128 filters. A bi-directional shift operation with $|h_k| = 2, |w_k| = 0, k \in \mathcal{I}$, and $\gamma = 1/16$ is inserted in the residual unit to form DRRN(PCS), as shown in Fig. 2(a). For EDSR [30], both the small EDSR(S) and large EDSR(L) comply with the original settings of network topology, upon which we built EDSR(S, PCS) and EDSR(L, PCS) by simply inserting a bi-directional shift into each residual block as shown in Fig. 2(b) and Fig. 6, with $|h_k| = 2, |w_k| = 0, k \in \mathcal{I}$, and $\gamma = 1/8$ for EDSR(S, PCS) and $\gamma = 1/16$ for EDSR(L, PCS).

The quantitative results on $\times 2$, $\times 3$ and $\times 4$ upscaling tasks are shown in Table 4. Following [30], we apply the geometric self-ensemble strategy to improve the performance, which is denoted as $^+$, and Fig. 7 shows the PSNR improvement by the enhancements of PCS, and the maximum is **0.18 dB**. Notably, smaller-scale models exhibit more noticeable improvements. The visual comparisons of SR results are in Fig. 11 and the supplementary material. Qualitative and quantitative results show that PCS is adept at handling highly self-similar patterns. Meanwhile, Fig. 8(c) shows the PCS-enhanced methods perform better with bigger training patches, and in Fig. 8(d), the PCS methods converge faster than non-PCS methods, with also obvious improvements.

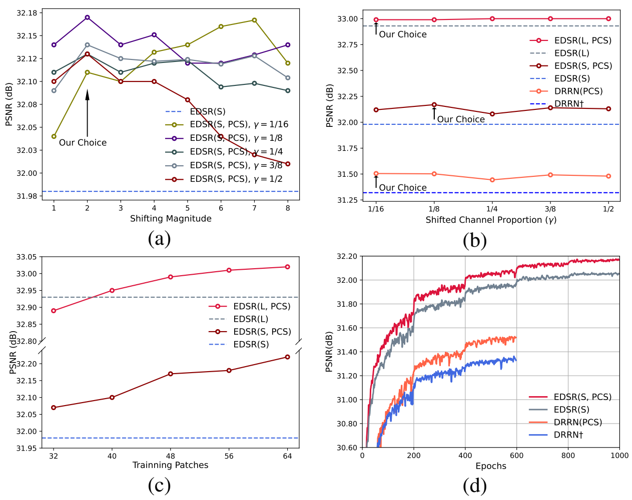

Figure 8: Ablation study on different settings of PCS, and the results are all tested on Urban100 [17] SR($\times 2$): (a) Shifting magnitudes of PCS; (b) Impact of γ on model performance; (c) Effect of different sizes of training patches; (d) Convergence comparison of non-PCS models and PCS-enhanced models.

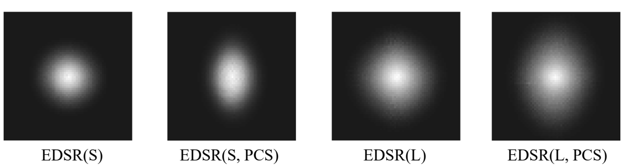

Figure 9: The comparison of ERFs [34] between non-PCS and PCS-enhanced models, and PCS effectively broadens the ERFs.

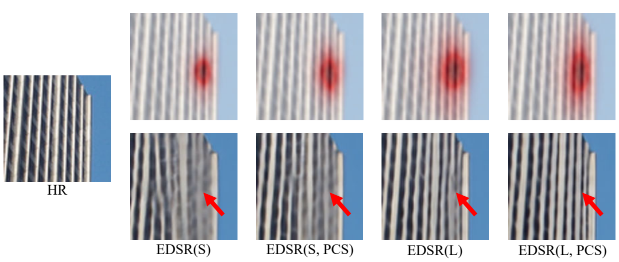

Figure 10: LAM [9] analysis for EDSR [30] with the enhancement of bi-directional shift. The input is img_096 from Urban100 [17] (SR$\times$4). With more activated pixels, the outputs of EDSR(PCS) show better spatial coherence and texture consistency.

5. Discussion

Theoretically, it is impractical (H4) and unnecessary (H1) to exhaust all possible situations in the state space. However, from the perspective of increasing the receptive field of local inference, we consider several typical constraints for applying PCS to SR models, which cover the main factors (*e.g.*, the proportion of shifted channels, shifting magnitude, the permutation of shifted and unshifted channels, coverage of the receptive field by the directions of channel shifting) that are related to the performance and efficiency of models. As the size of the convolution kernel directly affects the receptive field of local mapping, to prevent undue expansion of shift pattern possibilities, we only considered the general settings of 3×3. It is worth noting that the gains offered by PCS diminish as the model scale increases, as larger-scale models already possess adequate vast receptive fields. The supplementary material additionally shows the synergies between PCS and attention mechanisms are complex. Besides, as the identity shortcut is crucial to PCS, applying on models without or with few residual connections is not appropriate (*e.g.*, SRCNN [7], ESPCN [41], ECBSR [56]). For our further research, we will explore the synergies between PCS and Transformer-based SR architectures, and the design of adaptive PCS strategy.

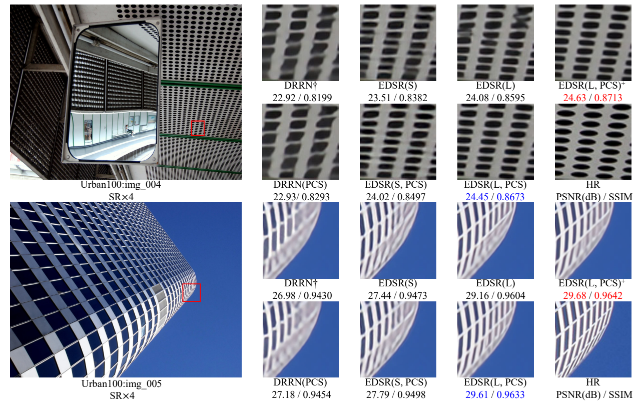

Figure 11: Qualitative comparisons between SR methods.

6. Conclusion

In this paper, we present Partial Channel Shifting (PCS) to enhance SISR networks, which enlarges the effective receptive field of the convolutional network without any extra parameters or Flops. Without any bells and whistles, PCS takes the initiative to feed more spatial information to the convolutional operation, by locally shifting a very small portion of feature channels, which is versatile, generic, plug-and-play, and could be applied in various architectures. It reveals at the same time that compared with directly increasing the long-range dependencies, the enhancement of local feature correlations is also effective in the SISR task. Further applications of PCS into deep neural networks improve the state-of-the-art on multiple benchmarks. Extensive evaluations show that PCS is generic and a highly effective operation over the present parameters-extensive methods for accurate image super-resolution.

Acknowledgment

This work is supported in part by the National Natural Science Foundation of China (Nos. 62102330, 62176221, 62276218), and in part by the Natural Science Foundation of Sichuan Province (Nos. 2022NSFSC0947, 2022NSFSC0945).

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *CVPRW*, pages 126–135, 2017. 5, 7

[2] Namhyuk Ahn, Byungkon Kang, and Kyung-Ah Sohn. Fast, accurate, and lightweight super-resolution with cascading residual network. In *ECCV*, pages 252–268, 2018. 2, 3

[3] Parichehr Behjati, Pau Rodriguez, Carles Fernández, Isabelle Hupont, Armin Mehri, and Jordi Gonzàlez. Single image super-resolution based on directional variance attention network. *Pattern Recognition*, 133:108997, 2022. 2

[4] Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie Line Alberi-Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In *BMVC*, pages 1–10, 2012. 5, 7

[5] François Chollet. Xception: Deep learning with depthwise separable convolutions. In *CVPR*, pages 1251–1258, 2017. 3

[6] Tao Dai, Jianrui Cai, Yongbing Zhang, Shu-Tao Xia, and Lei Zhang. Second-order attention network for single image super-resolution. In *CVPR*, pages 11065–11074, 2019. 2, 3, 5

[7] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE TPAMI*, 38(2):295–307, 2015. 1, 2, 8

[8] Hao Feng, Leijun Wang, Yongming Li, and Anyu Du. LKASR: Large kernel attention for lightweight image super-resolution. *Knowledge-Based Systems*, 252:109376, 2022. 3

[9] Jinjin Gu and Chao Dong. Interpreting super-resolution networks with local attribution maps. In *CVPR*, pages 9199–9208, 2021. 2, 7, 8

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016. 2, 4

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *ECCV*, pages 630–645, 2016. 2

[12] Xiangyu He, Zitao Mo, Peisong Wang, Yang Liu, Mingyuan Yang, and Jian Cheng. Ode-inspired network design for single image super-resolution. In *CVPR*, pages 1732–1741, 2019. 5

[13] Yihui He, Xianggen Liu, Huasong Zhong, and Yuchun Ma. Addressnet: Shift-based primitives for efficient convolutional neural networks. In *WACV*, pages 1213–1222, 2019. 2, 3, 5

[14] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *CVPR*, pages 7132–7141, 2018. 1, 2, 3

[15] Yanting Hu, Jie Li, Yuanfei Huang, and Xinbo Gao. Channel-wise and spatial feature modulation network for single image super-resolution. *IEEE TCSVT*, 30(11):3911–3927, 2019. 2

[16] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *CVPR*, pages 4700–4708, 2017. 2, 3

[17] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *CVPR*, pages 5197–5206, 2015. 5, 6, 7, 8

[18] Zheng Hui, Xinbo Gao, Yunchu Yang, and Xiumei Wang. Lightweight image super-resolution with information multi-distillation network. In *ACMMM*, pages 2024–2032, 2019. 2

[19] Eric Jang, Shixiang Gu, and Ben Poole. Categorical reparameterization with gumbel-softmax. In *ICLR*, 2017. 4

[20] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *CVPR*, pages 1646–1654, 2016. 1, 2

[21] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Deeply-recursive convolutional network for image super-resolution. In *CVPR*, pages 1637–1645, 2016. 1, 2

[22] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 5

[23] Fangyuan Kong, Mingxi Li, Songwei Liu, Ding Liu, Jingwen He, Yang Bai, Fangmin Chen, and Lean Fu. Residual local feature network for efficient super-resolution. In *CVPRW*, pages 766–776, 2022. 2

[24] Yann LeCun, Bernhard Boser, John Denker, Donnie Henderson, Richard Howard, Wayne Hubbard, and Lawrence Jackel. Handwritten digit recognition with a back-propagation network. *NeurIPS*, 2, 1989. 1

[25] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551, 1989. 1

[26] Juncheng Li, Faming Fang, Kangfu Mei, and Guixu Zhang. Multi-scale residual network for image super-resolution. In *ECCV*, pages 517–532, 2018. 2

[27] Wenbo Li, Kun Zhou, Lu Qi, Nianjuan Jiang, Jiangbo Lu, and Jiaya Jia. Lapar: Linearly-assembled pixel-adaptive regression network for single image super-resolution and beyond. *NeurIPS*, 33:20343–20355, 2020. 2

[28] Zheyuan Li, Yingqi Liu, Xiangyu Chen, Haoming Cai, Jinjin Gu, Yu Qiao, and Chao Dong. Blueprint separable residual network for efficient image super-resolution. In *CVPRW*, pages 833–843, 2022. 5

[29] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *ICCVW*, pages 1833–1844, 2021. 2, 3

[30] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *CVPRW*, pages 136–144, 2017. 1, 2, 3, 4, 5, 6, 7, 8

[31] Ji Lin, Chuang Gan, and Song Han. TSM: Temporal shift module for efficient video understanding. In *ICCV*, pages 7083–7093, 2019. 3, 5

[32] Ding Liu, Bihan Wen, Yuchen Fan, Chen Change Loy, and Thomas S Huang. Non-local recurrent network for image restoration. *NeurIPS*, 31, 2018. 3

[33] Jie Liu, Jie Tang, and Gangshan Wu. Residual feature distillation network for lightweight image super-resolution. In *ECCVW*, pages 41–55, 2020. 5

[34] Wenjie Luo, Yujia Li, Raquel Urtasun, and Richard Zemel. Understanding the effective receptive field in deep convolutional neural networks. *NeurIPS*, 29, 2016. 1, 2, 7, 8

[35] Xiaotong Luo, Yuan Xie, Yulun Zhang, Yanyun Qu, Cuihua Li, and Yun Fu. Latticenet: Towards lightweight image super-resolution with lattice block. In *ECCV*, pages 272–289, 2020. 2

[36] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *ICCV*, pages 416–423, 2001. 5, 6, 7

[37] Yusuke Matsui, Kota Ito, Yuji Aramaki, Azuma Fujimoto, Toru Ogawa, Toshihiko Yamasaki, and Kiyoharu Aizawa. Sketch-based manga retrieval using manga109 dataset. *Multimedia Tools and Applications*, 76(20):21811–21838, 2017. 5, 7

[38] Yiqun Mei, Yuchen Fan, and Yuqian Zhou. Image super-resolution with non-local sparse attention. In *CVPR*, pages 3517–3526, 2021. 2, 3, 5

[39] Ben Niu, Weilei Wen, Wenqi Ren, Xiangde Zhang, Lianping Yang, Shuzhen Wang, Kaihao Zhang, Xiaochun Cao, and Haifeng Shen. Single image super-resolution via a holistic attention network. In *ECCV*, pages 191–207, 2020. 2

[40] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *CVPR*, pages 4510–4520, 2018. 2

[41] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *CVPR*, pages 1874–1883, 2016. 8

[42] Ying Tai, Jian Yang, and Xiaoming Liu. Image super-resolution via deep recursive residual network. In *CVPR*, pages 3147–3155, 2017. 1, 2, 3, 4, 7

[43] Ying Tai, Jian Yang, Xiaoming Liu, and Chunyan Xu. MemNet: A persistent memory network for image restoration. In *ICCV*, pages 4549–4557, 2017. 2

[44] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *NeurIPS*, 30, 2017. 2

[45] Guangting Wang, Yucheng Zhao, Chuanxin Tang, Chong Luo, and Wenjun Zeng. When shift operation meets vision transformer: An extremely simple alternative to attention mechanism. In *AAAI*, volume 36, pages 2423–2430, 2022. 5

[46] Xiaolong Wang, Ross Girshick, Abhinav Gupta, and Kaiming He. Non-local neural networks. In *CVPR*, pages 7794–7803, 2018. 2

[47] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE TIP*, 13(4):600–612, 2004. 5

[48] Zhaowen Wang, Ding Liu, Jianchao Yang, Wei Han, and Thomas Huang. Deep networks for image super-resolution with sparse prior. In *ICCV*, pages 370–378, 2015. 1

[49] Bichen Wu, Alvin Wan, Xiangyu Yue, Peter Jin, Sicheng Zhao, Noah Golmant, Amir Gholaminejad, Joseph Gonzalez, and Kurt Keutzer. Shift: A zero flop, zero parameter alternative to spatial convolutions. In *CVPR*, pages 9127–9135, 2018. 3, 4, 5

[50] Chengxing Xie, Xiaoming Zhang, Linze Li, Haiteng Meng, Tianlin Zhang, Tianrui Li, and Xiaole Zhao. Large kernel distillation network for efficient single image super-resolution. In *CVPRW*, pages 1283–1292, 2023. 2

[51] Jianchao Yang, John Wright, Thomas Huang, and Yi Ma. Image super-resolution as sparse representation of raw image patches. In *CVPR*, pages 1–8, 2008. 1, 2

[52] Jianchao Yang, John Wright, Thomas S. Huang, and Yi Ma. Image super-resolution via sparse representation. *IEEE TIP*, 19(11):2861–2873, 2010. 1, 2

[53] Jiahui Yu, Yuchen Fan, Jianchao Yang, Ning Xu, Zhaowen Wang, Xinchao Wang, and Thomas Huang. Wide activation for efficient and accurate image super-resolution. *arXiv preprint arXiv:1808.08718*, 2018. 2

[54] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *International conference on curves and surfaces*, pages 711–730, 2010. 5, 7

[55] Xindong Zhang, Hui Zeng, Shi Guo, and Lei Zhang. Efficient long-range attention network for image super-resolution. In *ECCV*, pages 649–667, 2022. 5

[56] Xindong Zhang, Hui Zeng, and Lei Zhang. Edge-oriented convolution block for real-time super resolution on mobile devices. In *ACMMM*, pages 4034–4043, 2021. 8

[57] Xiangyu Zhang, Xinyu Zhou, Mengxiao Lin, and Jian Sun. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In *CVPR*, pages 6848–6856, 2018. 3

[58] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *ECCV*, pages 286–301, 2018. 1, 2, 3, 5

[59] Yulun Zhang, Kunpeng Li, Kai Li, Bineng Zhong, and Yun Fu. Residual non-local attention networks for image restoration. In *ICLR*, 2019. 2, 3

[60] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *CVPR*, pages 2472–2481, 2018. 2, 3, 5

[61] Xiaole Zhao, Xiafei Hu, Ying Liao, Tian He, Tao Zhang, Xueming Zou, and Jinsha Tian. Accurate mr image super-resolution via lightweight lateral inhibition network. *CVIU*, 201:103075, 2020. 2, 3

[62] Lin Zhou, Haoming Cai, Jinjin Gu, Zheyuan Li, Yingqi Liu, Xiangyu Chen, Yu Qiao, and Chao Dong. Efficient image super-resolution using vast-receptive-field attention. In *ECCVW*, pages 256–272, 2023. 5

[63] Shangchen Zhou, Jiawei Zhang, Wangmeng Zuo, and Chen Change Loy. Cross-scale internal graph neural network for image super-resolution. *NeurIPS*, 33:3499–3509, 2020. 2