Swift Parameter-free Attention Network for Efficient Super-Resolution

Cheng Wan[1*]   Hongyuan Yu[2*]   Zhiqi Li[1*]   Yihang Chen[1]   Yajun Zou[2]

Yuqing Liu[2]   Xuanwu Yin[2]   Kunlong Zuo[2]

[1]Georgia Institute of Technology, [2]Xiaomi Inc

{cwan38, zli3167, ychen3726}@gatech.edu

{yuhongyuan, zouyajun, liuyuqing9, yinxuanwu, zuokunlong}@xiaomi.com

Abstract

Single Image Super-Resolution (SISR) is a crucial task in low-level computer vision, aiming to reconstruct high-resolution images from low-resolution counterparts. Conventional attention mechanisms have significantly improved SISR performance but often result in complex network structures and large number of parameters, leading to slow inference speed and large model size. To address this issue, we propose the Swift Parameter-free Attention Network (SPAN), a highly efficient SISR model that balances parameter count, inference speed, and image quality. SPAN employs a novel parameter-free attention mechanism, which leverages symmetric activation functions and residual connections to enhance high-contribution information and suppress redundant information. Our theoretical analysis demonstrates the effectiveness of this design in achieving the attention mechanism's purpose. We evaluate SPAN on multiple benchmarks, showing that it outperforms existing efficient super-resolution models in terms of both image quality and inference speed, achieving a significant quality-speed trade-off. This makes SPAN highly suitable for real-world applications, particularly in resource-constrained scenarios. Notably, we won the first place both in the overall performance track and runtime track of the NTIRE 2024 efficient super-resolution challenge. Our code and models are made publicly available at https://github.com/hongyuanyu/span.

1. Introduction

Single Image Super-Resolution (SISR) is a well-established task in low-level computer vision, which aims to reconstruct a high-resolution image from a single low-resolution image. This task has broad applicability in enhancing image quality across various domains [16, 37, 43, 44, 48, 49, 57]. The advent of deep learning has led to significant advancements in this field [2, 10, 12, 19, 24, 32, 34, 36, 50, 59]. Recent progress in super-resolution tasks has been largely
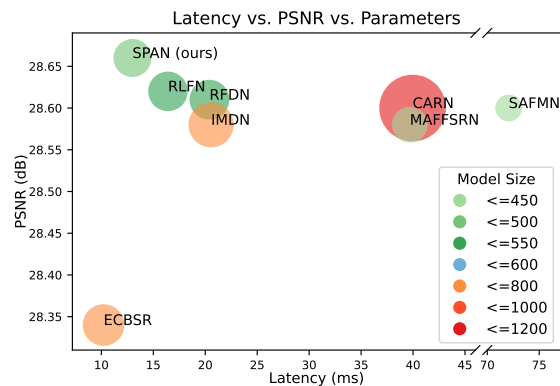
---

*Equal contribution, Corresponding Author

Figure 1. Latency, PSNR, and complexity of model comparison on Set14 dataset in x4 scale factor task.

driven by the attention mechanism. Numerous state-of-the-art super-resolution networks incorporate attention mechanisms or even employ larger vision transformers (ViTs) as the model architecture [6, 8, 20, 27, 32, 35, 42, 53, 60]. These networks emphasize key features and long-distance dependencies between patches through attention maps, capturing a wider range of contextual information to ensure continuity of details and accuracy of edge textures. However, the computational requirements of the attention mechanism, which involve complex network structures and a substantial number of additional parameters, lead to challenges such as large model size and slow inference speed. These challenges limit the applicability of these models, hindering their use in efficient, high-speed computing scenarios, such as SISR tasks on resource-constrained mobile devices.

Numerous existing efficient super-resolution (ESR) techniques have achieved certain successes in enhancing model efficiency. Some models primarily focus on reducing model FLOPs and parameters, accomplishing this through methods like group convolution and depth-wise separable convolution [2, 19, 34]. However, simply reducing FLOPs or parameters sometimes does not lead to a significant improvement in the model's inference speed, and it can also de-

crease model accuracy. Other models [23] reduce model parameter size through feature information sharing and downsizing non-attention branches. But these models still contain many parameters within their complex computational structures, resulting in long running time. To ensure fast inference speed, it is crucial to maintain a simple network topology. However, conventional attention mechanisms often result in more complex network structures. To address this problem, we propose a parameter-free attention mechanism and theoretically demonstrate that our Swift Parameter-free Attention Network (SPAN) can achieve the attention mechanism's role of enhancing high-contribution information and suppressing redundant information through symmetric activation functions and residual connections.

In SPAN, a parameter-free attention mechanism is constructed by passing extracted features through a symmetric activation function around the origin to calculate attention maps directly. This attention mechanism focuses on information-rich regions without the need of additional parameter learning, allowing for rapid and effective feature extraction from shallow to deep layers. The design of symmetric activation functions and residual connections in the modules help to solve issues related to information loss of the parameter-free attention modules. The simplicity of the network structure ensures operational speed, addressing the challenges posed by conventional attention mechanisms.

In summary, our main contributions are as follows:

- We design a novel parameter-free attention mechanism that employs symmetric activation functions and residual connections to enhance high-contribution information and suppress redundant information, thereby simplifying the network structure and improving inference speed without sacrificing accuracy.
- We propose the Swift Parameter-free Attention Network (SPAN), which leverages the parameter-free attention mechanism to achieve rapid and effective feature extraction from shallow to deep layers while maintaining low model complexity and parameter count.
- Through theoretical analysis and experimental validation, we demonstrate the effectiveness and superiority of SPAN in single-image super-resolution tasks, proving its practicality and potential application value in resource-constrained scenarios.

2. Related Work

2.1. Efficient Super Resolution on Image

Most existing ESR models focus on reducing model parameters or FLOPs to improve efficiency. SRCNN [11] introduces an end-to-end mapping for single image super-resolution using a deep convolutional neural network (CNN), optimizing all layers jointly. DRCN [25] presents a deeply-recursive convolutional network with up to 16

recursive layers to enhance super-resolution performance, overcoming training challenges through innovative techniques and achieving substantial improvements over previous methods. LatticeNet [38] introduces the Lattice Block for combining Residual Blocks using a lattice filter bank, showcasing improved performance through this novel combination approach. CARN [3] implements a cascading mechanism on a residual network to create an accurate and lightweight model. IMDN [20] is a lightweight and accurate single image SR model. It extracts hierarchical features and selectively aggregates them using contrast-aware channel attention. RFDN [35] improves over IMDN using more lightweight and flexible feature distillation connections and shallow residual blocks, achieving better SR performance with lower model complexity. However, simply minimizing parameters and FLOPs does not necessarily lead to better model efficiency, especially during inference. There is a need to develop SR models that prioritize faster inference speed rather than just reducing parameters or FLOPs.

To address this, RLFN [27] enhances model compactness and accelerates inference without sacrificing SR restoration quality based on RFDN. They analyze properties of intermediate features and find shallow features are critical for PSNR-oriented models. Based on this, they propose an improved feature extractor to effectively capture edges and details. Further, a multi-stage warm-up training strategy is introduced to speed up model convergence and improve SR restoration accuracy. Omni [53] introduces Aggregation Networks for efficient lightweight image super-resolution. It utilizes Omni Self-Attention to fuse spatial and channel self-attentions. In addition, a multi-scale feature extraction method is introduced to achieve high-quality restoration with low computational cost.

2.2. Attention Mechanism

For ESR task, the application of lightweight attention mechanisms play a significant role for enhancing model performance without substantially increasing complexity. The pivotal role of attention in modern vision models lies in its dynamic re-weighting of features, which directs computational resources to the most salient parts of the input, thus boosting efficiency and efficacy across various tasks [5, 15, 22, 29, 52, 56].

Attention-based super-resolution networks typically require a substantial receptive field to capture both local and global information, thereby enhancing super-resolution performance. However, the utilization of parameterized attention maps can slow down inference speed. In contrast, efficient super-resolution (SR) networks should maintain performance while ensuring rapid inference speed.

We observe that attention maps can be generated without necessitating additional training and parameters [7, 17], yet still contribute positively to the model's performance.
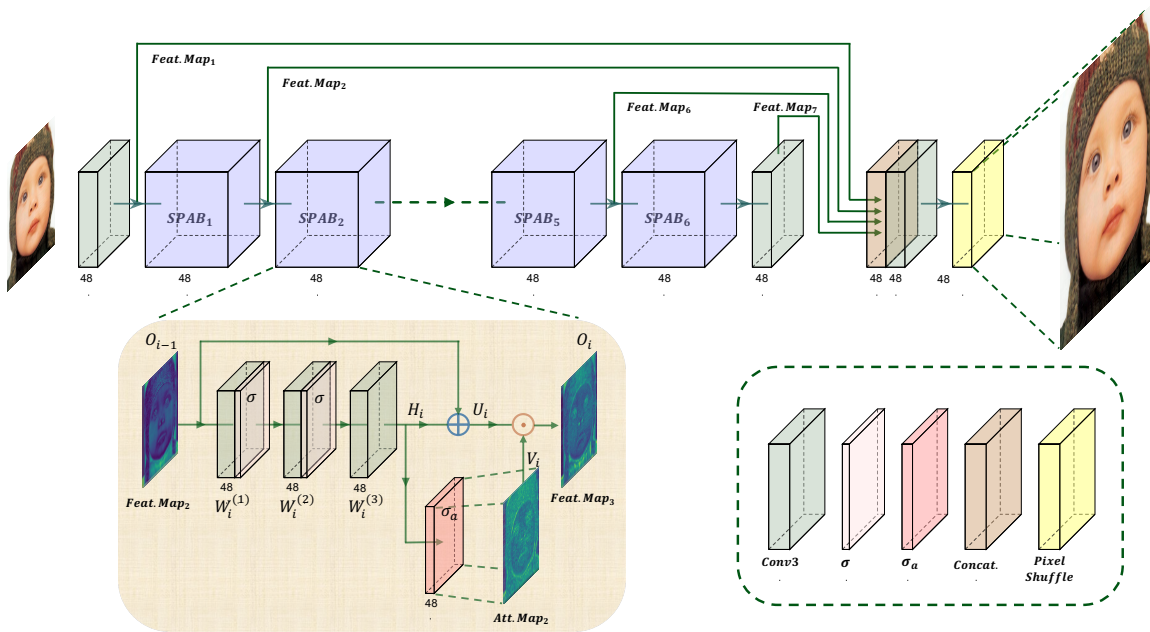
Figure 2. The proposed SPAN architecture. The yellow area indicates the internal structure of each SPAB module. $Att.Map_2$ denotes the generated attention map. Input is a low resolution image, and output is a high resolution image.

The crux of this lightweight attention approach is to maximize the representational power of the super-resolution network within a constrained model budget. By incorporating these principles, we can develop a swift and effective attention mechanism for super-resolution models. Our proposed parameter-free attention mechanism utilizes existing convolutional layers to enhance high-contribution information and suppress redundant information, thus eliminating the need for additional parameter-intensive processes. This not only streamlines the model but also augments its ability to weakly localize objects, a feature that is essential for refining super-resolution techniques [7].

3. Method

In this section, we will first present our proposed method: the Swift Parameter-free Attention Block (SPAB) based on the attention mechanism and the proposed SPAN built using SPABs for super resolution. Then, we will conduct theoretical analysis of SPAB, demonstrating the effectiveness of parameter-free attention mechanisms for super resolution tasks, especially with symmetric activation functions and residual connections.

3.1. Network Architecture.

As shown in Figure 2, SPAN consists of 6 consecutive SPABs and each SPAB block extracts progressively higher-level features sequentially through three convolutional layers with C'-channeled $H' \times W'$-sized kernels (In our model, we choose $H' = W' = 3$.). The extracted features H_i are then added with a residual connection from the input

of SPAB, forming the pre-attention feature map U_i for that block. The features extracted by the convolutional layers are passed through an activation function $\sigma_a(\cdot)$ that is symmetric about the origin to obtain the attention map V_i. The feature map and attention map are element-wise multiplied to produce the final output $O_i = U_i \odot V_i$ of the SPAB block, where $\odot$ denotes element-wise multiplication. We use $W_i^{(j)} \in R^{C' \times H' \times W'}$ to represent the kernel of the j-th convolutional layer of the i-th SPAB block and σ to represent the activation function following the convolutional layer. Then the SPAB block can be expressed as:

$$
\begin{aligned}
O_i &= F_{W_i}^{(i)}(O_{i-1}) = U_i \odot V_i, \\
U_i &= O_{i-1} \oplus H_i, \quad V_i = \sigma_a(H_i), \\
H_i &= F_{c,W_i}^{(i)}(O_{i-1}), \\
&= W_i^{(3)} \otimes \sigma(W_i^{(2)} \otimes \sigma(W_i^{(1)} \otimes O_{i-1})),
\end{aligned}
\tag{1}
$$

where $\oplus$ and $\otimes$ represent the element-wise sum between extracted features and residual connections, and the convolution operation, respectively. $F_{W_i}^{(i)}$ and $F_{c,W_i}^{(i)}$ are the function representing the i-th SPAB and the function representing the 3 convolution layers of i-th SPAB with parameters $W_i = (W_i^{(1)}, W_i^{(2)}, W_i^{(3)})$, respectively. $O_0 = \sigma(W_0 \otimes I_{LR})$ is a C'-channeled $H \times W$ feature map from the C-channeled $H \times W$-sized low-resolution input image I_{LR} undergone a convolutional layer with 3×3 sized kernel W_0. This convolutional layer ensures that each SPAB has the same number of channels as input. The whole SPAN neural network can

be described as

$$I_{\text{HR}} = F(I_{\text{LR}}) = \text{PixelShuffle}[W_{f2} \otimes O],$$
$$O = \text{Concat}(O_0, O_1, O_5, W_{f1} \otimes O_6), \tag{2}$$

where O is a $4C'$-channeled $H \times W$-sized feature map with multiple hierarchical features obtaining by concatenating O_0 with the outputs of the first, fifth, and the convolved output of the sixth SPAB blocks by C'-channeled 3×3-sized kernel W_{f1}. O is processed through a 3×3 convolutional layer to create an r^2C channel feature map of size $H \times W$. Pixel shuffle is a classic upsampling method [46], which can increase the spatial resolution without adding computational complexity by rearranging the elements in the feature maps. Then, this feature map goes through a pixel shuffle module to generate a high-resolution image of C channels and dimensions $rH \times rW$, where r represents the super-resolution factor.

3.2. Parameter-Free Attention Mechanism

In SPAB, we directly obtain the attention map from the higher-level feature information extracted by the convolutional layers through an origin-symmetric activation function σ_a. Because in the branch dedicated to computing the attention map that diverges from the computation of the feature map, there are no modules with trainable parameters except for the activation function, our attention mechanism is parameter-free. In previous super-resolution works, while attention maps can improve model accuracy by allowing the model to selectively focus on the most relevant parts of the feature, calculating attention maps introduces additional parameters, which slows down the model's computation [32, 51, 53] with extra calculations. To enhance the computational efficiency of our model, we draw inspiration from parameter-free attention mechanisms as proposed in [7, 13, 47, 55], and in our attention mechanism, we directly obtain the attention map V_i through a parameter-free activation function σ_a.

In the context of the super-resolution task, the reason why computing the attention map in this way is effective lies in the fact that, in super-resolution tasks, when utilizing attention mechanisms, the neural network should focus on where local information such as complex textures, edges, color transitions, and more is particularly rich, where super-resolution is more challenging with issues such as blurring artifacts tending to occur [39]. Interestingly, these edge and texture information required for attention can be directly detected through the convolutional kernels learned during training [21] and at the same time, they are also the information that networks need to extract in order to accomplish the super-resolution task. Therefore, we can potentially determine the region for attention directly based on the magnitudes of the convolutional layer's output values and obtain the attention map $V_i = \sigma_a(H_i)$ parameter-free directly from

the output of the convolutional layer. This is also reflected in the visualizations of the feature map after using our attention in Figure 4a (compared with Figure 4c), where, after training, the attention map V_i directly computed from the output of the convolutional layer H_i tend to make feature map O_i relatively higher in areas with complex textures and boundaries.

Our parameter-free attention mechanism can be theoretically demonstrated through the following process. It should be noted that because we analyze the role of the residual connection in Section 3.3, we remove the residual connection in SPAB for simplicity in this section and add the residual connection for analyzing in Section 3.3. Without attention, the gradient used to update the i-th SPAB during the training of the model can be expressed as

$$\frac{\partial L}{\partial W_i} = \Pi \frac{\partial F_{W_i}^{(i)}(O_{i-1})}{\partial W_i}$$
$$= \Pi \frac{\partial F_{c,W_i}^{(i)}(O_{i-1})}{\partial W_i}, \tag{3}$$

where L denotes the loss during training and Π represents the product of gradients before $\frac{\partial F_{W_i}^{(i)}(O_{i-1})}{\partial W_i}$ in the gradient chain in back-propagation algorithm. While after adding self attention mechanism, the gradient is

$$\frac{\partial L}{\partial W_i} = \Pi \frac{\partial F_{W_i}^{(i)}(O_{i-1})}{\partial W_i}$$
$$= \Pi \frac{\partial}{\partial W_i}(F_{c,W_i}^{(i)}(O_{i-1}) \odot \sigma_a(F_{c,W_i}^{(i)}(O_{i-1}))) \quad (4)$$
$$= \Pi \frac{\partial F_{c,W_i}^{(i)}(O_{i-1})}{\partial W_i} \odot (H_i \odot \sigma_a'(H_i) + \sigma_a(H_i)).$$

According to Equation 4, it can be found that for information-richer region, corresponding values in feature H_i and $\sigma_a(H_i)$ have larger absolute value and with $\sigma_a'(H_i) > 0$, corresponding absolute value in $H_i \odot \sigma_a'(H_i) + \sigma_a(H_i)$ will be larger to make the information-richer region have more influence on the gradient, so that through the training process, the model will pay more attention to information-rich regions.

3.3. Design Consideration

The idea of computing attention maps directly without parameters from feature extracted by convolutional layers, led to two design considerations for our neural network: the choice of activation function for computing the attention map and the use of residual connections.

Symmetric Activation Function As mentioned in Section 3.1, we choose the activation function symmetric about the origin to compute the attention map. There are two main reasons, firstly, because in feature maps extracted through
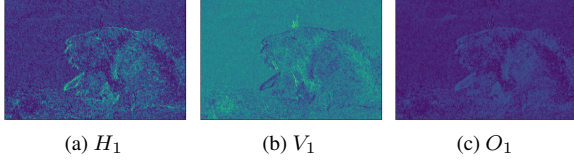
(a) H_1 (b) V_1 (c) O_1

Figure 3. **Attention Weakened by Non-odd activation function:** From left to right, the sequence are the first SPAB's Pre-attention Feature Map H_1, the Attention Map V_1, and the Output Feature Map O_1, when activation function σ_a is sigmoid. In the figures, brighter regions denote larger absolute values.

structure-related convolutional layers, like gradient kernels, sign of values always represents directions and the absolute magnitude represents the feature quantity. To generate the attention map $\sigma_a(H_i)$ directly based on feature quantities, it must roughly holds

$$|\sigma_a(x)| = |\sigma_a(|x|)| = |\sigma_a(-x)|. \quad (5)$$

Second, per Equation 4, our attention method amplifies gradients in information-rich regions and dampens those in information-poor regions. To ensure these effects, $H_i \odot \sigma_a'(H_i)$ and $\sigma_a(H_i)$ must not cancel each other out, thus it is necessary to ensure:

$$x\sigma_a'(x) \cdot \sigma_a(x) > 0$$
$$\overset{\sigma_a'(x)>0}{\Rightarrow} x\sigma_a(x) > 0. \quad (6)$$

The second inequality in Equation 6 is due to the fact that common activation functions are increasing functions. Based on 5 and 6, it can be deduced that σ_a needs to be an odd function symmetric about the origin. In addition, using non-odd activation functions like Sigmoid, which completely filter out negative values, would result in the loss of information for features with large magnitudes but negative values ($|x| \gg 0, x < 0$), as shown in Figure 3. In Figure 3, comparing H_1 and V_1, it can be observed that in H_1, some areas with large absolute feature values become dimmer in $V_1 = \sigma_a(H_i)$, representing a decrease in magnitude, due to the effect of the sigmoid function filtering out negative values. The resulting feature map O_i, obtained using this type of attention map V_1, fails to emphasize the features, resulting in a rather blurred representation compared to initial features.

Residual Connection On the other hand, the attention mechanism enhances higher-level features extracted at each layer, which can lead to a significant loss of information in regions where higher-level features are less prominent. Compared to the results without using attention (Figure 4f), using attention can cause issue of the excessive loss of information with the later SPAB block (Figure 4e), ultimately leading to a decrease in the accuracy of super-resolution results (Table 3). We employ residual connections to address this problem. Through residual connections, we use lower-level features from the input of the SPAB layer to compensate for the excessive loss of information during the feature

LR Input

(a) 1st SPAB (b) 1st SPAB_nores (c) 1st SPAB_empty

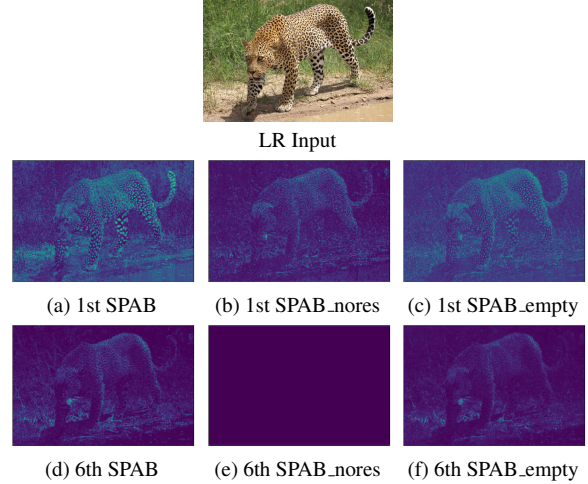(d) 6th SPAB (e) 6th SPAB_nores (f) 6th SPAB_empty

Figure 4. **Residual Connections Improve Attention:** a comparison is made between the output feature maps of the first and sixth SPABs for complete SPAB (a) (d), SPAB with no residual connections (b) (e), and SPAB without attention and residual connections (c) (f). It is observed that the SPAB with attention but lacking residual connections tends to lose a substantial amount of information, which issue is alleviated by incorporating residual connections.

map generation process, and in the final SPAB (Equation 1), $V_i \odot H_i$ is replaced by $V_i \odot (H_i \oplus O_{i-1})$. This will result in the gradient $\frac{\partial L}{\partial W_i}$ for updating W_i during the training process in Equation 4 become

$$\frac{\partial L}{\partial W_i} = \Pi \frac{\partial F_{W_i}^{(i)}(O_{i-1})}{\partial W_i}$$
$$= \Pi \frac{\partial}{\partial W_i}((F_{c,W_i}^{(i)}(O_{i-1}) + O_{i-1}) \odot \sigma_a(F_{c,W_i}^{(i)}(O_{i-1}))) \quad (7)$$
$$= \Pi \frac{\partial F_{c,W_i}^{(i)}(O_{i-1})}{\partial W_i} \odot ((H_i + O_{i-1}) \odot \sigma_a'(H_i) + \sigma_a(H_i)).$$

In this way, because of the existing of O_{i-1} in $\frac{\partial L}{\partial W_i}$, for the trained model, the regions focused on by the attention mechanism are determined not only by the information of the current level but also by the lower-level information output by the previous SPAB. This alleviates the significant loss of information caused by overly focusing on higher-level features.

By comparing the feature maps of the first and sixth layers of the block with no attention (Figure 4c and 4f), the block with attention (Figure 4b and 4e), and the block with improved attention through residual connections (Figure 4a and 4d), we observe that the SPAB block with residual connections can further highlight features compared to the block with no attention. At the same time, it does not suffer from the loss of excessive information seen in the block with attention but without residual connections, ensuring that lower-level information is preserved. Detailed numerical results are shown in Table 3

Scale	Model	Params (K)	Runtime (ms)	Set5 PSNR↑ / SSIM↑	Set14 PSNR↑ / SSIM↑	BSD100 PSNR↑ / SSIM↑	Urban100 PSNR↑ / SSIM↑	Manga109 PSNR↑ / SSIM↑
×2	SRCNN[11]	24	6.92	36.66 / 0.9542	32.42 / 0.9063	31.36 / 0.8879	29.50 / 0.8946	35.74/0.9661
	FSRCNN[12]	12	9.02	36.98 / 0.9556	32.62 / 0.9087	31.50 / 0.8904	29.85 / 0.9009	36.67/0.9694
	VDSR[24]	666	35.37	37.53 / 0.9587	33.05 / 0.9127	31.90 / 0.8960	30.77 / 0.9141	37.22/0.9729
	DRCN[23]	1774	716.45	37.63 / 0.9588	33.04 / 0.9118	31.85 / 0.8942	30.75 / 0.9133	28.93 / 0.8854
	LapSRN[28]	251	53.98	37.52 / 0.9591	32.99 / 0.9124	31.80 / 0.8952	30.41 / 0.9103	37.27/0.9740
	CARN[2]	1592	159.10	37.76 / 0.9590	33.52 / 0.9166	32.09 / 0.8978	31.92 / 0.9256	38.36/0.9765
	IMDN[19]	694	77.34	38.00 / 0.9605	33.63 / 0.9177	32.19 / 0.8996	32.17 / 0.9283	38.88/0.9774
	RFDN[34]	534	74.51	38.05 / 0.9606	33.68 / 0.9184	32.16 / 0.8994	32.12 / 0.9278	38.88/0.9773
	MAFFSRN[42]	402	152.91	37.97 / 0.9603	33.49 / 0.9170	32.14 / 0.8994	31.96 / 0.9268	/
	ECBSR[59]	596	39.96	37.90 / 0.9615	33.34 / 0.9178	32.10 / 0.9018	31.71 / 0.9250	/
	RLFN-S[27]	454	56.09	38.05 / 0.9607	33.68 / 0.9172	32.19 / 0.8997	32.17 / 0.9286	/
	RLFN[27]	527	60.39	38.07 / 0.9607	33.72 / 0.9187	32.22 / 0.9000	32.33 / 0.9299	/
	ShuffleMixer[50]	394	218.36	38.01 / 0.9606	33.63 / 0.9180	32.17 / 0.8995	31.89 / 0.9257	38.83/0.9774
	SAFMN[51]	228	118.07	38.00 / 0.9605	33.54 / 0.9177	32.16 / 0.8995	31.84 / 0.9256	38.71/0.9771
	SPAN-S (ours)	411	**45.08**	38.06 / 0.9608	33.73 / 0.9187	32.21 / 0.9001	32.20 / 0.9288	38.85 / 0.9776
	SPAN (ours)	481	**50.39**	38.08 / 0.9608	33.71 / 0.9183	32.22 / 0.9002	32.24 / 0.9294	38.94 / 0.9777
×4	SRCNN[11]	57	1.90	30.48 / 0.8628	27.49 / 0.7503	26.90 / 0.7101	24.52 / 0.7221	27.58 / 0.8555
	FSRCNN[12]	13	2.22	30.72 / 0.8660	27.61 / 0.7550	26.98 / 0.7150	24.62 / 0.7280	27.90 / 0.8610
	VDSR[24]	666	8.95	31.35 / 0.8838	28.01 / 0.7674	27.29 / 0.7251	25.18 / 0.7524	28.83 / 0.8870
	DRCN[23]	1774	176.59	31.53 / 0.8854	28.02 / 0.7670	27.23 / 0.7233	25.14 / 0.7510	28.93 / 0.8854
	LapSRN[28]	502	66.81	31.54 / 0.8852	28.09 / 0.7700	27.32 / 0.7275	25.21 / 0.7562	29.09 / 0.8900
	CARN[2]	1592	39.96	32.13 / 0.8937	28.60 / 0.7806	27.58 / 0.7349	26.07 / 0.7837	30.47 / 0.9084
	IMDN[19]	715	20.56	32.21 / 0.8948	28.58 / 0.7811	27.56 / 0.7353	26.04 / 0.7838	30.45 / 0.9075
	RFDN[34]	550	20.40	32.24 / 0.8952	28.61 / 0.7819	27.57 / 0.7360	26.11 / 0.7858	30.58 / 0.9089
	MAFFSRN[42]	441	39.69	32.18 / 0.8948	28.58 / 0.7812	27.57 / 0.7361	26.04 / 0.7848	/
	ECBSR[59]	603	10.21	31.92 / 0.8946	28.34 / 0.7817	27.48 / 0.7393	25.81 /0.7773	/
	FDIWN-M[14]	454	-	32.17 / 0.8941	28.55 / 0.7806	27.58 / 0.7364	26.02 / 0.7844	/
	RLFN-S[27]	470	15.16	32.23 / 0.8961	28.61 / 0.7818	27.58 / 0.7359	26.15 / 0.7866	/
	RLFN[27]	543	16.41	32.24 / 0.8952	28.62 / 0.7813	27.60 / 0.7364	26.17 / 0.7877	/
	ShuffleMixer[50]	411	144.22	32.21 / 0.8953	28.66 / 0.7827	27.61 / 0.7366	26.08 / 0.7835	30.65 / 0.9093
	SAFMN[51]	240	72.06	32.18 / 0.8948	28.60 / 0.7813	27.58 / 0.7359	25.97 / 0.7809	30.43/0.9063
	SPAN-S (ours)	426	**12.22**	32.20 / 0.8950	28.64 / 0.7828	27.60 / 0.7368	26.13 / 0.7865	30.60 / 0.9095
	SPAN (ours)	498	**13.67**	32.20 / 0.8953	28.66 / 0.7834	27.62 / 0.7374	26.18 / 0.7879	30.66 / 0.9103

Table 1. Quantitative results of the state-of-the-art ESR models on five benchmark datasets. The approach to evaluating inference time remains consistent with RLFN [27]. The best and second-best results are marked in red and blue colors, and the **bold** numbers represent that the inference speed of our model is the fastest when the performance and number of parameters are similar.

4. Experiments

4.1. Experimental Setup

Datasets and Metrics In accordance with established techniques [30, 32], our models are trained on the DF2K dataset, which is a combination of DIV2K [1] and Flickr2K [33] datasets, comprising a total of 3450 (800 + 2650) high-quality images. We adopt standard protocols to generate LR images by bicubic downscaling of reference HR images. We evaluate our models on 4 different benchmark dataset: Set5 [4], Set14 [58], B100 [40], Urban100 [18] and Manga109 [41], PSNR and SSIM are used as performance on the Y channel of YCbCr space for SR task.

Implementation Details Six SPAB modules with 48-channeled feature maps are depolyed for training procedure. We have employed the re-parameterization method (REP) [9] to improve the efficiency during the inference stage. During each training batch, 64 HR RGB patches are cropped, measuring 256×256, and subjected to random flipping and rotation. The learning rate is initialized at 5×10^{-4} and undergoes a halving process every 2×10^5 iterations. The network undergoes training in for a total of 10^6 iterations on four NVIDIA A100 GPUs, with each run taking approximately 12 hours, and the L1 loss function being minimized through the utilization of the Adam optimizer[26]. We perform the aforementioned training settings twice after loading the trained weights to obtain the optimal results.

4.2. Quantitative Results

In this study, we conduct 2x and 4x upscaling on the SPAN and SPAN-S models across various benchmark tests, and compare their detailed test results with the current state-of-the-art efficient super-resolution models [2, 11, 12, 14, 19, 23, 24, 27, 28, 34, 42, 50, 51, 54, 59]. For detailed results, please refer to Table 1. Across multiple benchmarks, SPAN and SPAN-S exhibit superior performance in terms of PSNR and SSIM compared to other models, especially notable in inference time. With 50K fewer parameters than RLFN and RLFN-S, SPAN and SPAN-S demon-

$\sigma_a(x)$	Learnable	Set14 PSNR/SSIM	BSD100 PSNR/SSIM	Urban100 PSNR/SSIM
Sigmoid(x)	-	28.62/0.7826	27.59/0.7366	26.08/0.7854
Sigmoid$(x) - 0.5$	-	28.63/0.7825	27.60/0.7368	26.10/0.7856
Sigmoid$(ax) - 0.5$	✓	28.62/0.7825	27.60/0.7368	26.12/0.7861
$b \times ($Sigmoid$(ax) - 0.5)$	✓	28.62/0.7826	27.61/0.7367	26.11/0.7860

Table 2. Performance comparison of using different activation functions $\sigma_a(x)$ evaluated on three benchmark datasets.

Model	Runtime (ms)	Set5 PSNR / SSIM	Set14 PSNR / SSIM	BSD100 PSNR / SSIM	Urban100 PSNR / SSIM
SPAN_nores	11.65	31.25/0.8808	28.05/0.7687	27.21/0.7236	25.09/0.7496
SPAN_noatt	10.85	32.07/0.8943	28.56/0.7813	27.57/0.7357	25.97/0.7817
SPAN_empty	10.22	32.06/0.8939	28.56/0.7811	27.56/0.7355	26.02/0.7830
SPAN (ours)	12.22	32.18/0.8950	28.63/0.7825	27.60/0.7368	26.10/0.7856

Table 3. Performance comparison of SPAN_nores, SPAN_noatt, SPAN_empty, and our baseline SPAN. Runtime is the average time of 10 runs on DF2K validation set.

Model	Set5 PSNR / SSIM	Set14 PSNR / SSIM	BSD100 PSNR / SSIM	Urban100 PSNR / SSIM
SPAN_no_rep	32.11/0.8943	28.59/0.7818	27.58/0.7358	26.02/0.7833
SPAN_rep	32.18/0.8950	28.63/0.7825	27.60/0.7368	26.10/0.7856

Table 4. Comparision of our baseline SPAN with re-parameterization and SPAN_no_rep without using the re-parameterization technology for training process.

strate significant advantages in inference speed and in the performance metrics of PSNR and SSIM. As illustrated in Figure 1, through visualizing the relationship between image quality, inference time, and model size, we observe that SPAN achieves significantly higher PSNR than other models at comparable inference speeds; and with similar model parameter counts, it not only performs better but also operates faster. Hence, RLFN and RLFN-S have achieved the best current balance in terms of quality, parameter count, and inference speed.

4.3. Activation Function

In the Section 3.2, we discuss the use of origin-symmetric activation functions $\sigma_a(x)$ in the context of a parameter-free attention mechanism, its role and principles. We experiment with and proposed several different origin-symmetric activation functions, comparing their impact on the ESR performance with the attention mechanism. As shown in Table 2, we conduct experiments to compare the effects of different activation functions. To ensure model speed, we ultimately choose $\sigma_a(x) = $ Sigmoid$(x) - 0.5$ as the activation function, which is simple but effective for the attention mechanism in our SPAN model. The proposed learnable activation functions, serving as an extension to the SPAN model and a discussion point on activation functions, demonstrate potentially higher performance. However, it is observed that the speed tends to decrease with the use of learnable activation functions.

4.4. Ablation Study

In the ablation experiment, we uniformly conduct x4 scale factor experiments on the models with 48 channels.

Residual Connection As discussed in Section 3, we have improved the performance of our model under the non-parametric attention mechanism by incorporating residual connections within each SPAB module to mitigate excessive information loss. To demonstrate the effectiveness of these intra-module residual connections, we conduct experiments by removing them from every SPAB module in our model and comparing the results with our baseline model. The model without residual connections is denoted as SPAN_nores. All experimental settings for this model are kept consistent with our SPAN model.

We thoroughly train both models under identical settings and evaluate their performance on four benchmark datasets: Set5, Set14, BSD100, and Urban100. Table 3 presents the results, highlighting the impact of residual connections on model performance. Notably, these connections enhance image quality while maintaining high inference speed.

Attention Mechanism To assess the efficacy of our attention mechanism's foundation, which utilizes a parameter-free attention process for information feedback, we conduct an experiment by removing the activation function and the element-wise multipication operation from the attention component within the SPAB module. This effectively eliminates the parameter-free attention mechanism, enabling the module to output a direct residual connection with the original input. Maintaining the same experimental setup, we designate this modified model as SPAN_noatt and compare it across various dimensions with the original SPAN model. Table 3 demonstrates the enhanced performance of SPAN over SPAN_noatt, validating the effectiveness of our parameter-free attention mechanism in augmenting network capabilities and accentuating high-frequency features in feature maps.

Combination Module We also perform an extensive validation of the proposed module design. The SPAB module, which integrates residual connections with a parameter-free attention mechanism, achieves efficient direct enhancement. To evaluate this, we remove these two key components in our experiments while maintaining other experimental settings constant. Results from Table 3 demonstrate that our proposed attention module, incorporating residual connections, significantly improves image quality in super-resolution tasks without compromising processing speed. These findings robustly substantiate the effectiveness and practicality of our module design in enhancing super-resolution processing outcomes.

Re-parameterization As Table 4 showed, we have implemented the re-parameterization technique (rep) [9] to enhance the efficiency of the inference phase.

Training Setting As can be seen from Table 5, training the model one more time can improve its performance, but more training rounds will not bring significant gains.
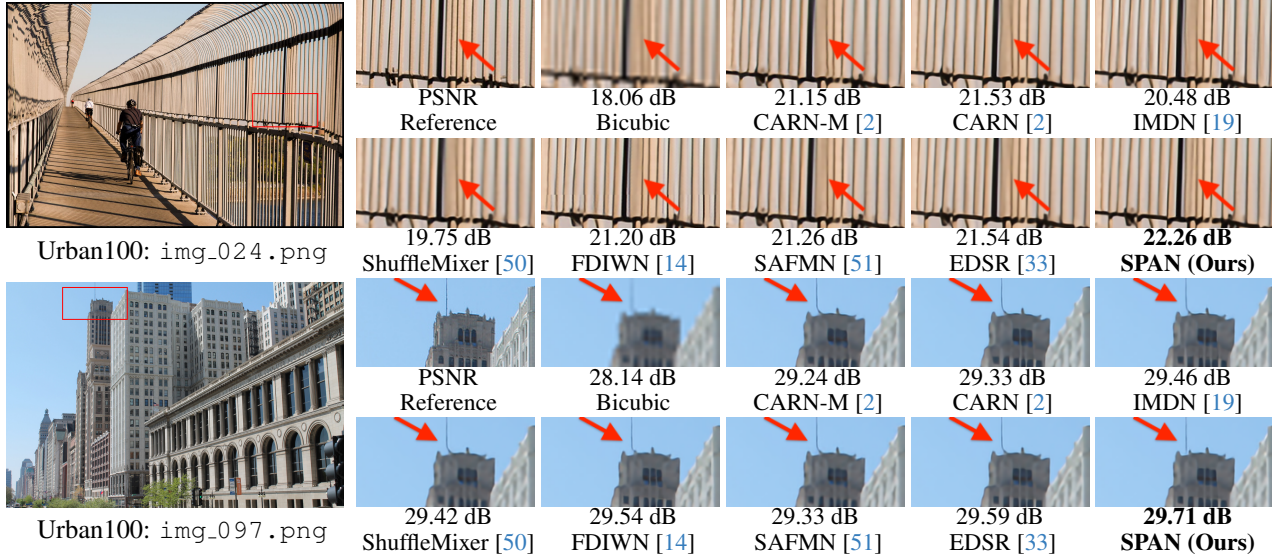
Figure 5. Visual comparison for ×4 SR methods. The patches for comparison are marked with red boxes. (Best viewed by zooming.)

Model	Set5 PSNR / SSIM	Set14 PSNR / SSIM	BSD100 PSNR / SSIM	Urban100 PSNR / SSIM
SPAN_stage1	32.18/0.8950	28.63/0.7825	27.60/0.7368	26.10/0.7856
SPAN_stage2	32.20/0.8953	28.64/0.7828	27.61/0.7370	26.13/0.7864
SPAN_stage3	32.20/0.8952	28.64/0.7828	27.61/0.7371	26.13/0.7865

Table 5. Performance comparison of SPAN in different training stages, SPAN_stage3 represents our latest model.

4.5. SPAN for NTIRE 2024 challenge

Our team won the 1st place in the main track (Overall Performance Track) and the 1st place in the sub-track1 (Inference Runtime Track) of NTIRE 2024 efficient super-resolution challenge [45]. The model structure and training strategy are slightly different from the above. The proposed model has 6 SPABs, in which the number of feature channels is set to 28. DIV2K and LSDIR [31] datasets are used for training in this challenge. During each training batch, 64 HR RGB patches are cropped, measuring 256×256, and subjected to random flipping and rotation. In the training phase, NGswin [8] is used as the teacher model for boosting the restoration performance. The learning rate is initialized at 5×10^{-4} and undergoes a halving process every 2×10^5 iterations. The network undergoes training for a total of 10^6 iterations, with the L1 loss function being minimized through the utilization of the Adam optimizer [26]. We repeat the aforementioned training settings four times after loading the trained weights. Subsequently, fine-tuning is executed using the L1 and L2 loss functions, with an initial learning rate of 1×10^{-5} for 5×10^5 iterations, and HR patch size of 512. We conduct finetuning on four models utilizing both L1 and L2 losses, and employ batch sizes of 64 and 128. Finally, we integrate these four models to obtain the ultimate model. In comparison to the RLFN

Team name	PSNR [val]	PSNR [test]	Ave Time [ms]	Parameters [M]	FLOPs [G]	Overall Ranking
XiaomiMM(ours)	26.94	27.01	**5.592**	**0.151**	9.83	**1**
Cao Group	26.90	27.00	8.372	0.215	13.05	2
BSR	26.90	27.00	9.384	0.218	11.95	3
VPEG_O	26.90	27.01	9.630	0.212	13.86	4
CMVG	26.90	27.01	10.022	0.202	12.17	5
RLFN(Baseline)	26.96	27.07	11.77	0.317	19.67	/

Table 6. Mean Results of 5 runs (With an extra zero-run to warm-up the GPU first) on single NVIDIA GeForce RTX 3090 GPU of NTIRE2024 ESR Challenge. The top five methods are included.

method that secure the first place in the NTIRE2022 Efficient Super-Resolution Challenge, our method significantly outperforms other methods across all metrics while attaining the fastest running time.

5. Conclusion

In this paper, we have presented the Swift Parameter-free Attention Network, an efficient SISR model that addresses the challenges posed by conventional attention mechanisms in terms of complex network structures, slow inference speed, and large model size. SPAN uses a parameter-free attention mechanism to enhance important information and reduce redundancy. Its simple structure, symmetric activation functions, and residual connections ensure high image quality and fast inference speed. Our extensive experiments on multiple benchmarks have shown that SPAN image quality existing efficient super-resolution models in terms of both performance and inference speed, achieving a significant quality-speed trade-off. This makes SPAN highly suitable for real-world applications, particularly in resource-constrained scenarios such as mobile devices. Future research may apply the parameter-free attention mechanism to other computer vision tasks and further optimize the network for greater efficiency.

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2017. 6

[2] Namhyuk Ahn, Byungkon Kang, and Kyung-Ah Sohn. Fast, accurate, and lightweight super-resolution with cascading residual network. In *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part X*, pages 256–272. Springer, 2018. 1, 6, 8

[3] Namhyuk Ahn, Byungkon Kang, and Kyung-Ah Sohn. Fast, accurate, and lightweight super-resolution with cascading residual network. In *Proceedings of the European conference on computer vision (ECCV)*, pages 252–268, 2018. 2

[4] Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie Line Alberi-Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. 2012. 6

[5] Chenjie Cao, Qiaole Dong, and Yanwei Fu. Learning prior feature and attention enhanced image inpainting. In *European Conference on Computer Vision*, pages 306–322. Springer, 2022. 2

[6] Jiezhang Cao, Qin Wang, Yongqin Xian, Yawei Li, Bingbing Ni, Zhiming Pi, Kai Zhang, Yulun Zhang, Radu Timofte, and Luc Van Gool. Ciaosr: Continuous implicit attention-in-attention network for arbitrary-scale image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1796–1807, 2023. 1

[7] Junsuk Choe and Hyunjung Shim. Attention-based dropout layer for weakly supervised object localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2219–2228, 2019. 2, 3, 4

[8] Haram Choi, Jeongmin Lee, and Jihoon Yang. N-gram in swin transformers for efficient lightweight image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2071–2081, 2023. 1, 8

[9] Xiaohan Ding, Xiangyu Zhang, Ningning Ma, Jungong Han, Guiguang Ding, and Jian Sun. Repvgg: Making vgg-style convnets great again. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13733–13742, 2021. 6, 7

[10] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *Computer Vision - ECCV 2014 - 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV*, pages 184–199. Springer, 2014. 1

[11] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13*, pages 184–199. Springer, 2014. 2, 6

[12] Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network.

In *Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II*, pages 391–407. Springer, 2016. 1, 6

[13] Jie Du, Kai Guan, Yanhong Zhou, Yuanman Li, and Tianfu Wang. Parameter-free similarity-aware attention module for medical image classification and segmentation. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2022. 4

[14] Guangwei Gao, Wenjie Li, Juncheng Li, Fei Wu, Huimin Lu, and Yi Yu. Feature distillation interaction weighting network for lightweight image super-resolution. *CoRR*, abs/2112.08655, 2021. 6, 8

[15] Shenyuan Gao, Chunluan Zhou, Chao Ma, Xinggang Wang, and Junsong Yuan. Aiatrack: Attention in attention for transformer visual tracking. In *European Conference on Computer Vision*, pages 146–164. Springer, 2022. 2

[16] C Heltin Genitha and K Vani. Super resolution mapping of satellite images using hopfield neural networks. In *Recent Advances in Space Technology Services and Climate Change 2010 (RSTS & CC-2010)*, pages 114–118. IEEE, 2010. 1

[17] Daniel Haase and Manuel Amthor. Rethinking depthwise separable convolutions: How intra-kernel correlations lead to improved mobilenets. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14600–14609, 2020. 2

[18] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5197–5206, 2015. 6

[19] Zheng Hui, Xinbo Gao, Yunchu Yang, and Xiumei Wang. Lightweight image super-resolution with information multi-distillation network. In *Proceedings of the 27th ACM International Conference on Multimedia, MM 2019, Nice, France, October 21-25, 2019*, pages 2024–2032. ACM, 2019. 1, 6, 8

[20] Zheng Hui, Xinbo Gao, Yunchu Yang, and Xiumei Wang. Lightweight image super-resolution with information multi-distillation network. In *Proceedings of the 27th acm international conference on multimedia*, pages 2024–2032, 2019. 1, 2

[21] Carlo Innamorati, Tobias Ritschel, Tim Weyrich, and Niloy J Mitra. Learning on the edge: Investigating boundary filters in cnns. *International Journal of Computer Vision*, 128:773–782, 2020. 4

[22] Younho Jang, Wheemyung Shin, Jinbeom Kim, Simon Woo, and Sung-Ho Bae. Glamd: Global and local attention mask distillation for object detectors. In *European Conference on Computer Vision*, pages 460–476. Springer, 2022. 2

[23] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Deeply-recursive convolutional network for image super-resolution. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 1637–1645. IEEE Computer Society, 2016. 2, 6

[24] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA,*

June 27-30, 2016, pages 1646–1654. IEEE Computer Society, 2016. 1, 6

[25] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Deeply-recursive convolutional network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1637–1645, 2016. 2

[26] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6, 8

[27] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 766–776, 2022. 1, 2, 6

[28] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 5835–5843. IEEE Computer Society, 2017. 6

[29] Hunsang Lee, Hyesong Choi, Kwanghoon Sohn, and Dongbo Min. Knn local attention for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2139–2149, 2022. 2

[30] 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. In *NeurIPS*, pages 20343–20355, 2020. 6

[31] Yawei Li, Kai Zhang, Jingyun Liang, Jiezhang Cao, Ce Liu, Rui Gong, Yulun Zhang, Hao Tang, Yun Liu, Denis Demandolx, Rakesh Ranjan, Radu Timofte, and Luc Van Gool. Lsdir: A large scale dataset for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 8

[32] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *IEEE/CVF International Conference on Computer Vision Workshops, ICCVW 2021, Montreal, BC, Canada, October 11-17, 2021*, pages 1833–1844. IEEE, 2021. 1, 4, 6

[33] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 1132–1140. IEEE Computer Society, 2017. 6, 8

[34] Jie Liu, Jie Tang, and Gangshan Wu. Residual feature distillation network for lightweight image super-resolution. In *Computer Vision - ECCV 2020 Workshops - Glasgow, UK, August 23-28, 2020, Proceedings, Part III*, pages 41–55. Springer, 2020. 1, 6

[35] Jie Liu, Jie Tang, and Gangshan Wu. Residual feature distillation network for lightweight image super-resolution. In *Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*, pages 41–55. Springer, 2020. 1, 2

[36] Jie Liu, Wenjie Zhang, Yuting Tang, Jie Tang, and Gangshan Wu. Residual feature aggregation network for image super-

resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2359–2368, 2020. 1

[37] Wei Liu, Dahua Lin, and Xiaoou Tang. Hallucinating faces: Tensorpatch super-resolution and coupled residue compensation. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, pages 478–484. IEEE, 2005. 1

[38] Xiaotong Luo, Yuan Xie, Yulun Zhang, Yanyun Qu, Cuihua Li, and Yun Fu. Latticenet: Towards lightweight image super-resolution with lattice block. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXII 16*, pages 272–289. Springer, 2020. 2

[39] Cheng Ma, Yongming Rao, Yean Cheng, Ce Chen, Jiwen Lu, and Jie Zhou. Structure-preserving super resolution with gradient guidance. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7769–7778, 2020. 4

[40] 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 *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, pages 416–423. IEEE, 2001. 6

[41] 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*, 76:21811–21838, 2017. 6

[42] Abdul Muqeet, Jiwon Hwang, Subin Yang, Jung Heum Kang, Yongwoo Kim, and Sung-Ho Bae. Multi-attention based ultra lightweight image super-resolution. In *Computer Vision - ECCV 2020 Workshops - Glasgow, UK, August 23-28, 2020, Proceedings, Part III*, pages 103–118. Springer, 2020. 1, 6

[43] Rakesh C Patel and MV Joshi. Super-resolution of hyperspectral images using compressive sensing based approach. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 1:83–88, 2012. 1

[44] Ronald R Peeters, Pierre Kornprobst, Mila Nikolova, Stefan Sunaert, Thierry Vieville, Grégoire Malandain, Rachid Deriche, Olivier Faugeras, Michael Ng, and Paul Van Hecke. The use of super-resolution techniques to reduce slice thickness in functional mri. *International Journal of Imaging Systems and Technology*, 14(3):131–138, 2004. 1

[45] Bin Ren, Yawei Li, Nancy Mehta, Radu Timofte, et al. The ninth NTIRE 2024 efficient super-resolution challenge report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024. 8

[46] 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 *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1874–1883, 2016. 4

[47] Yuxuan Shi, Lingxiao Yang, Wangpeng An, Xiantong Zhen, and Liuqing Wang. Parameter-free channel attention for

image classification and super-resolution. *arXiv preprint arXiv:2303.11055*, 2023. 4

[48] Amanjot Singh and Jagroop Singh Sidhu. Super resolution applications in modern digital image processing. *International Journal of Computer Applications*, 150(2):0975–8887, 2016. 1

[49] B Sreenivas, B Narasimha Chary, and INDIA KARIMNA-GAR. Processing of satellite image using digital image processing. In *A world forum on Geospatial*, pages 18–21, 2011. 1

[50] Long Sun, Jinshan Pan, and Jinhui Tang. Shufflemixer: An efficient convnet for image super-resolution. *Advances in Neural Information Processing Systems*, 35:17314–17326, 2022. 1, 6, 8

[51] Long Sun, Jiangxin Dong, Jinhui Tang, and Jinshan Pan. Spatially-adaptive feature modulation for efficient image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023. 4, 6, 8

[52] Thanh-Dat Truong, Quoc-Huy Bui, Chi Nhan Duong, Han-Seok Seo, Son Lam Phung, Xin Li, and Khoa Luu. Direcformer: A directed attention in transformer approach to robust action recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20030–20040, 2022. 2

[53] Hang Wang, Xuanhong Chen, Bingbing Ni, Yutian Liu, and Jinfan Liu. Omni aggregation networks for lightweight image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22378–22387, 2023. 1, 2, 4

[54] Xintao Wang, Chao Dong, and Ying Shan. Repsr: Training efficient vgg-style super-resolution networks with structural re-parameterization and batch normalization. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 2556–2564, 2022. 6

[55] Lingxiao Yang, Ru-Yuan Zhang, Lida Li, and Xiaohua Xie. Simam: A simple, parameter-free attention module for convolutional neural networks. In *International conference on machine learning*, pages 11863–11874. PMLR, 2021. 4

[56] Jiyang Yu, Jingen Liu, Liefeng Bo, and Tao Mei. Memory-augmented non-local attention for video super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17834–17843, 2022. 2

[57] Zhi Yuan, Jiong Wu, Sei-ichiro Kamata, Alireza Ahrary, and Peimin Yan. Fingerprint image enhancement by super resolution with early stopping. In *2009 IEEE International Conference on Intelligent Computing and Intelligent Systems*, pages 527–531. IEEE, 2009. 1

[58] 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. Springer, 2010. 6

[59] Xindong Zhang, Hui Zeng, and Lei Zhang. Edge-oriented convolution block for real-time super resolution on mobile devices. In *MM '21: ACM Multimedia Conference, Virtual Event, China, October 20 - 24, 2021*, pages 4034–4043. ACM, 2021. 1, 6

[60] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part VII*, pages 294–310. Springer, 2018. 1