Paper ID 0942

BEIJING JIAOTONG UNIVERSITY

University of Science and Technology Beijing

ACM multimedia

Ottawa, Canada OCT 29 – NOV 3, 2023

Kernel Dimension Matters: to Activate Available Kernels for Real-time Video Super-Resolution

Shuo Jin[1,2], Meiqin Liu[1,2*], Chao Yao[3], Chunyu Lin[1,2] and Yao Zhao[1,2]

[1] Institute of Information Science, Beijing Jiaotong University, Beijing, China

[2] Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing, China

[3] School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing, China

Motivation

- It is crucial to split channels with dynamic and static information for efficient processing, which represents the redundant complements and temporal offsets respectively.

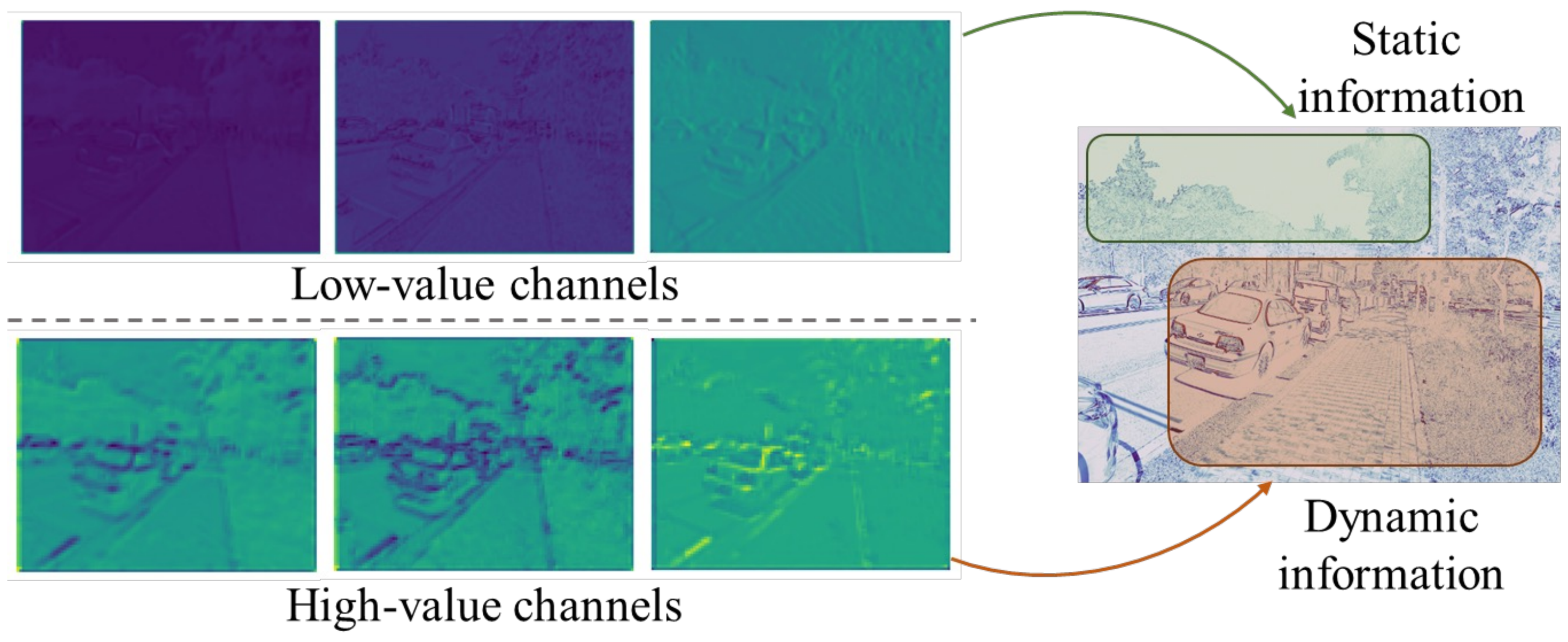

Fig.1 Different value channels.

Contribution

- We propose a Kernel Split Network (KSNet) for RTVSR, including unidirectional & bidirectional recurrent paradigms, which reconstructs the 720 × 1080 video with 31 ms/frame and achieve superior results over other SOTA methods.
- We design a kernel-split strategy to discriminate the channels of high-value and low-value, and apply the re-parameterized convolutions on the high-value channels. This strategy enables the exploration of high-value dynamic information and the representation of convolution layers, while reducing the dimensional complexity.
- We adopt multiple flows on the deformable alignment module to light the training burden of the deformable convolution and enhance the motion representation with affordable calculation costs.

Methodology

- Given a sequence of LR frames $\{x_1, x_2, \dots, x_n, \dots, x_N\}$ where $x_n \in \mathbb{R}^{C\times W\times H}$, the HR frames $\{y_1, y_2, \dots, y_n, \dots, y_N\}$ are constructed in a frame-to-frame pipeline. The input frames are propagated in the first recurrent branch (green area) and then refined in the other recurrent branch (blue area).

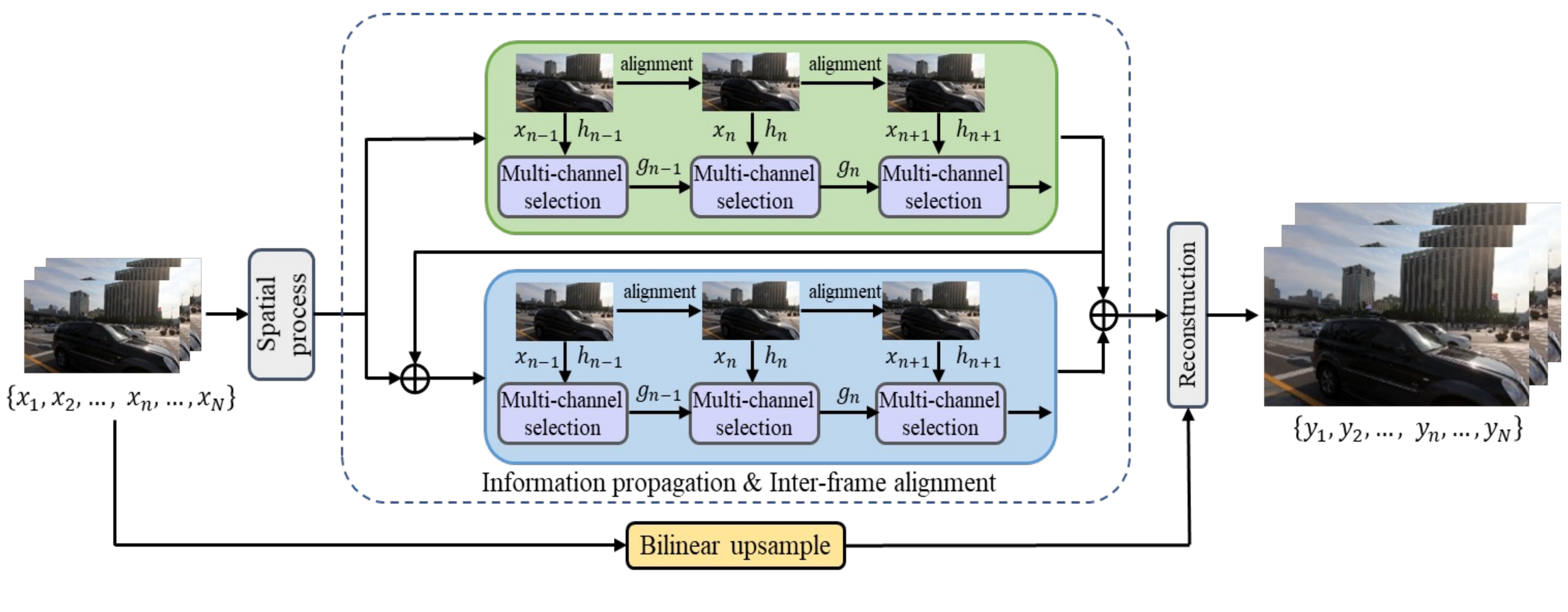

Fig.2 Overview of Kernel Split Network (KSNet).

Multi-channel Selection Unit (MSU)

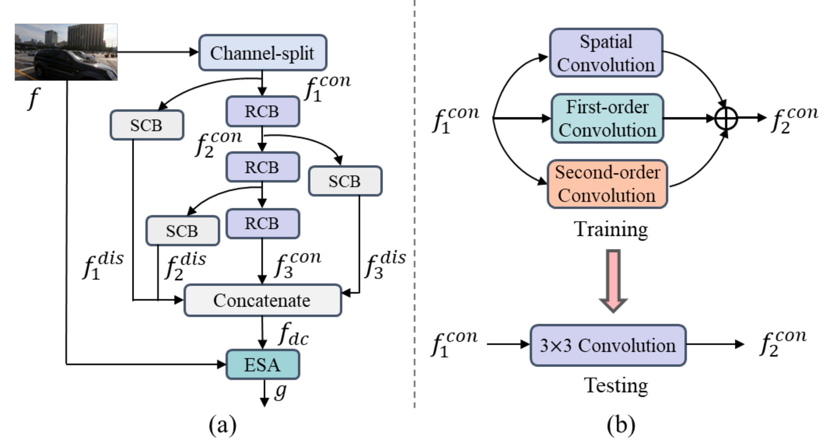

Fig.3 Multi-channel Selection Unit.

MSU splits the high-value and low-value channels to merge the kernel dimensions hierarchically.

$$\sigma_2(\sigma_1(k)) = W_2 * (W_1 * k + B_1) + B_2$$

$$W_{\mathrm{pr}}^{:,:,i+\lfloor\lfloor\frac{H-h}{2}\rfloor,j+\lfloor\frac{W-w}{2}\rfloor} = W_{\mathrm{ps}}^{:,:,i,j}$$
$$B_{\mathrm{pr}} = B_{\mathrm{ps}}$$

$$W_{\mathrm{ps}} = \mathrm{perm}(W_1) * W_2 \qquad W_{re} = W_p + W_{\mathrm{one}} + W_{\mathrm{sec}}$$
$$B_{\mathrm{ps}} = W_2 * \mathrm{rep}(B_1) + B_2 \qquad B_{re} = B_p + B_{\mathrm{one}} + B_{\mathrm{sec}}$$

Multiple Flow Deformable Alignment (MFDA)

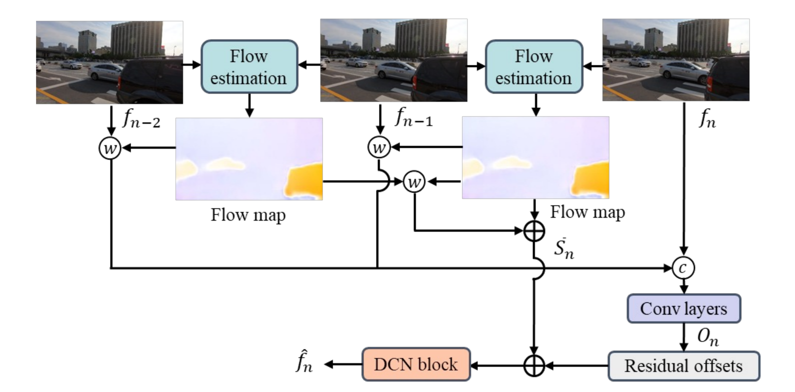

Fig.4 Multiple Flow Deformable Alignment.

MFDA aligns the motion area, where the optical flow is viewed as residual offsets for deformable alignment.

$$s_{n\to n-1} = \mathrm{ME}(f_n, f_{n-1})$$
$$s_{n-1\to n-2} = \mathrm{ME}(f_{n-1}, f_{n-2})$$

$$\bar{s}_n = s_{n\to n-1} + \mathrm{MC}(s_{n-1\to n-2}, s_{n\to n-1})$$
$$\bar{f}_n = \mathrm{MC}(f_{n-2}, s_{n-1\to n-2}) + \mathrm{MC}(f_{n-1}, s_{n\to n-1})$$
$$O_n = C(\mathrm{cat}(\bar{f}_n, f_n))$$
$$\hat{f}_n = D(O_n + \bar{s}_n)$$

Experiments

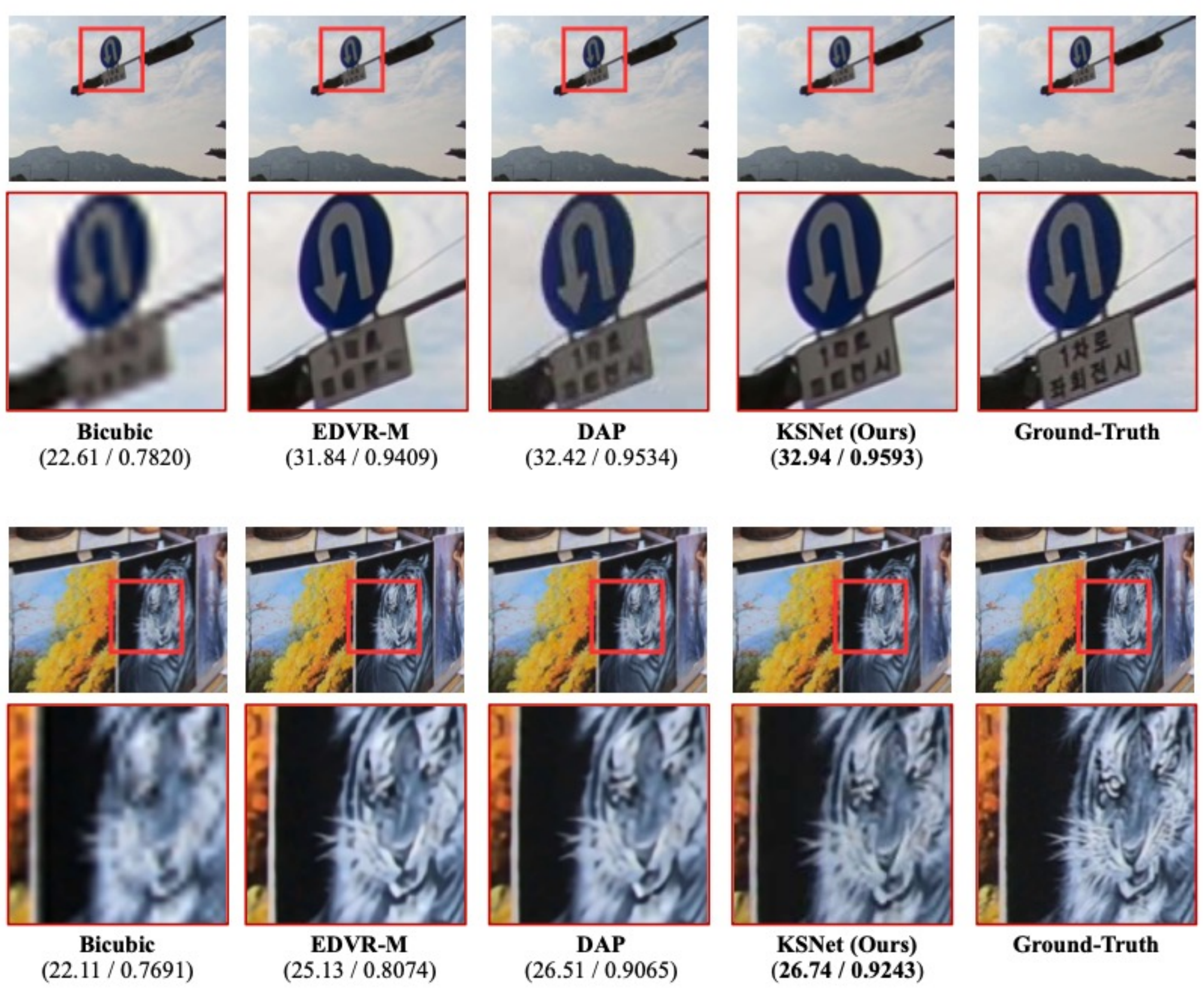

Fig.5 Visual comparisons with SOTA methods on REDS dataset.

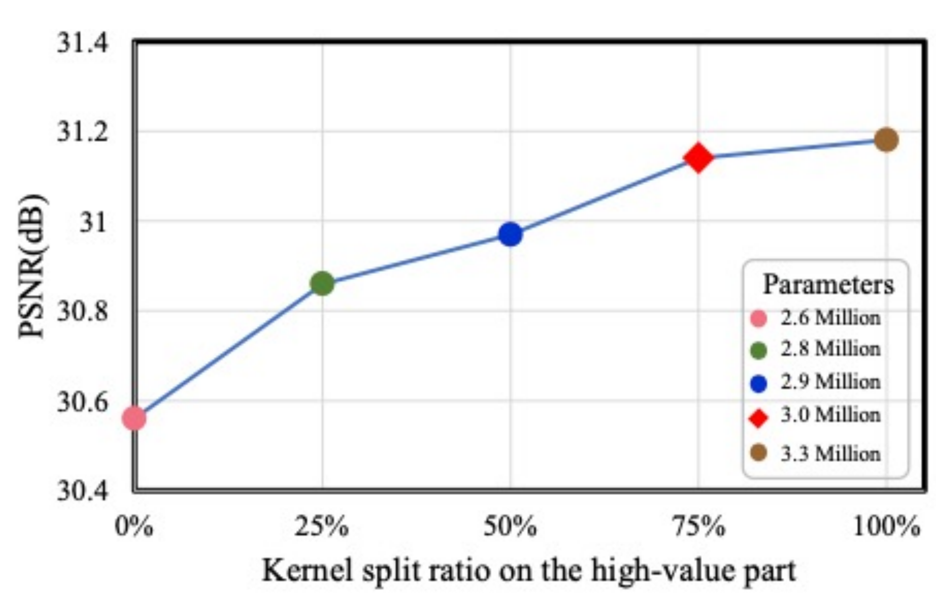

Fig.6 Comparisons on the reconstruction performance and kernel split ratio.

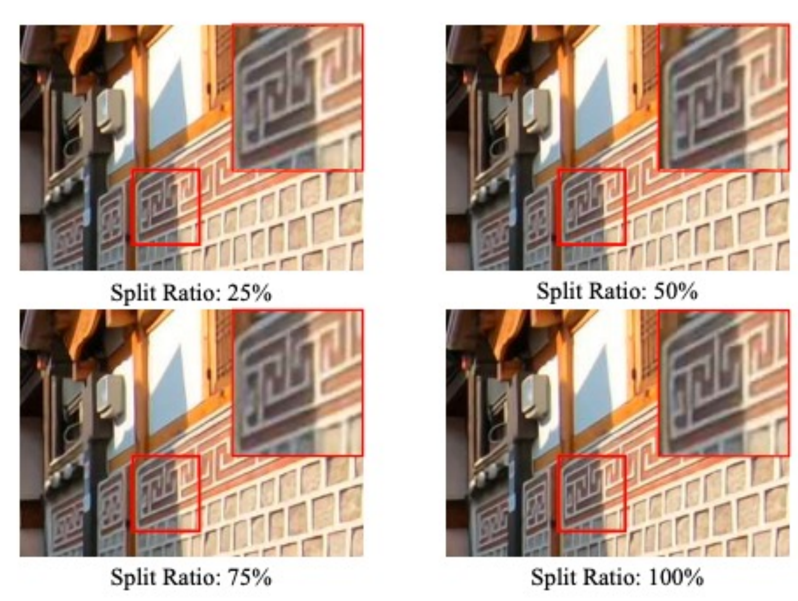

Fig.7 Comparisons on the visual quality of different split ratios.

Tab.1 Quantitative comparison with different methods.

Methods	Params(M)	FLOPs(G)	fps(1/s)	Run(ms)	Test datasets		
					Vid4	REDS4	Vimeo90K-T
Bicubic	-	-	-	-	23.78/0.6347	26.14/0.7292	31.30/0.8687
VESPCN [2]	-	-	28.6	35	25.35/0.7557	-	-
RLSP [6]	4.2	503.7	21.7	46	27.05/0.8139	-	36.49/0.9403
RRN [9]	3.4	387.5	22.2	45	27.09/0.8185	-	-
DAP-128 [5]	-	330.0	26.3	38	-	30.59/0.8703	37.29/0.9476
RSDN [7]	6.2	713.2	10.6	94	**27.22/0.8249**	-	37.23/0.9471
EDVR-M [24]	3.3	925.7	8.6	116	27.10/0.8186	30.53/0.8699	37.33/0.9484
KSNet-uni (Ours)	**3.0**	**296.9**	**32.3**	**31**	27.14/0.8208	**30.69/0.8724**	**37.34/0.9490**
ASSL-bi [34]	2.7	210.2	32.3	31	27.03/0.8163	30.72/0.8783	36.71/0.9410
RSCL-bi [26]	**2.7**	**210.2**	32.3	31	27.16/0.8213	30.99/0.8831	36.83/0.9421
BasicVSR [3]	6.3	330.0	15.9	63	**27.24/0.8251**	**31.42/0.8909**	37.53/0.9450
KSNet-bi (Ours)	3.0	296.9	**32.3**	**31**	27.22/0.8245	31.14/0.8862	**37.54/0.9503**

Tab.2 Ablation study on re-parameterization strategy.

Model	Params-training(M)	Params-testing(M)	FLOPs-training(T)	FLOPs-testing(T)	Run(ms)	fps(1/s)	PSNR(dB)
KSNet-uni-plain	2.9	2.9	296.9	296.9	31	32.3	30.50
KSNet-bi-plain	2.9	2.9	296.9	296.9	31	32.3	30.94
KSNet-uni (Ours)	3.4	2.9	320.9	296.9	31	32.3	30.69
KSNet-bi (Ours)	3.4	2.9	320.9	296.9	31	32.3	31.14