Self-Calibrated Efficient Transformer for Lightweight Super-Resolution

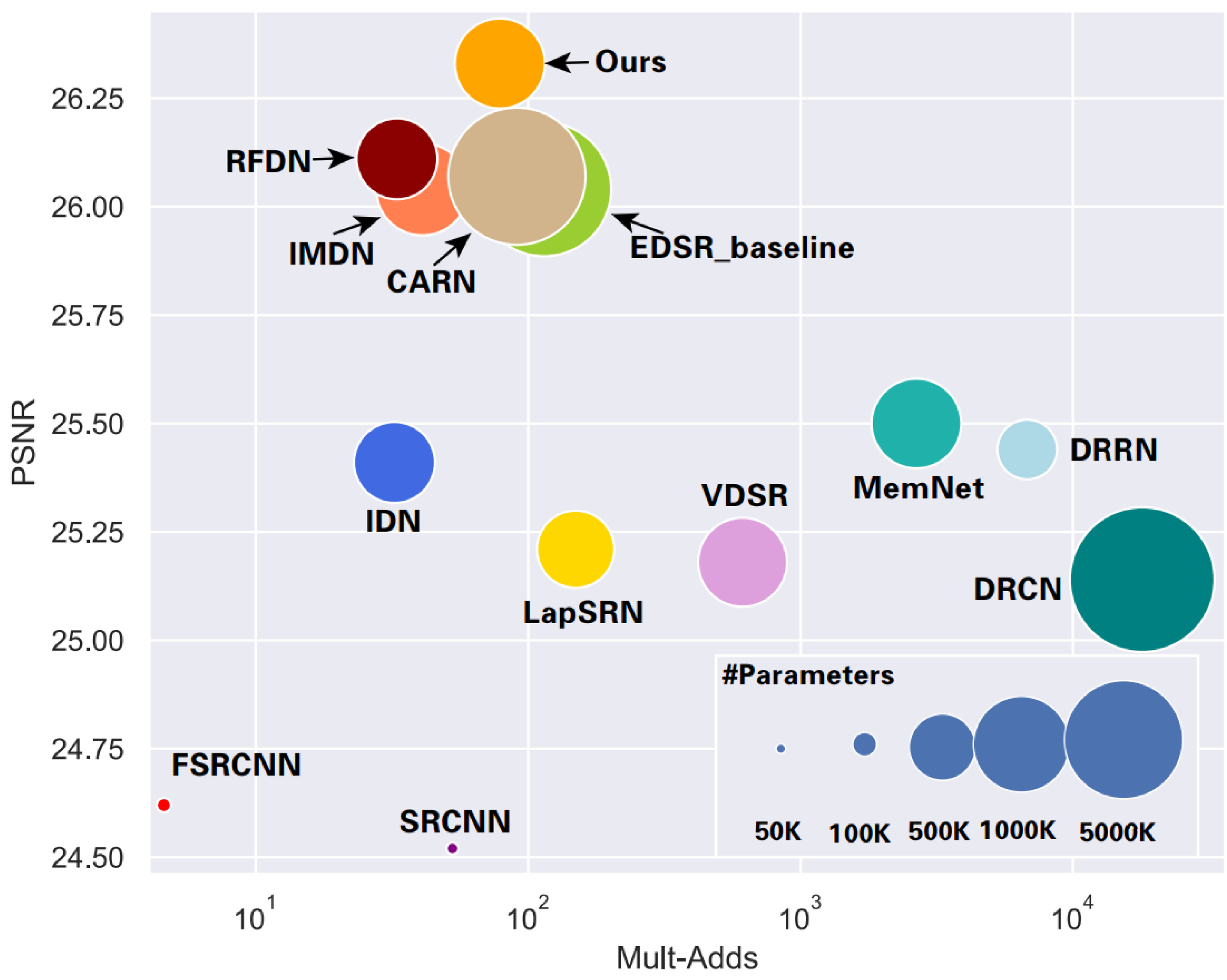
Wenbin Zou[1,*], Tian Ye[2,*], Weixin Zheng[3,*], Yunchen Zhang[4], Liang Chen[1,†], Yi Wu[1]
Fujian Normal University[1], Jimei University[2], Fuzhou University[3], China Design Group[4]

CVPR JUNE 19-24 2022 NEW ORLEANS • LOUISIANA

One-Sentence Summary:

We propose the Self-Calibrated Efficient Transformer Network (SCET), which effectively combines the efficient pixel attention mechanism with the transformer to achieves excellent results with few parameters.

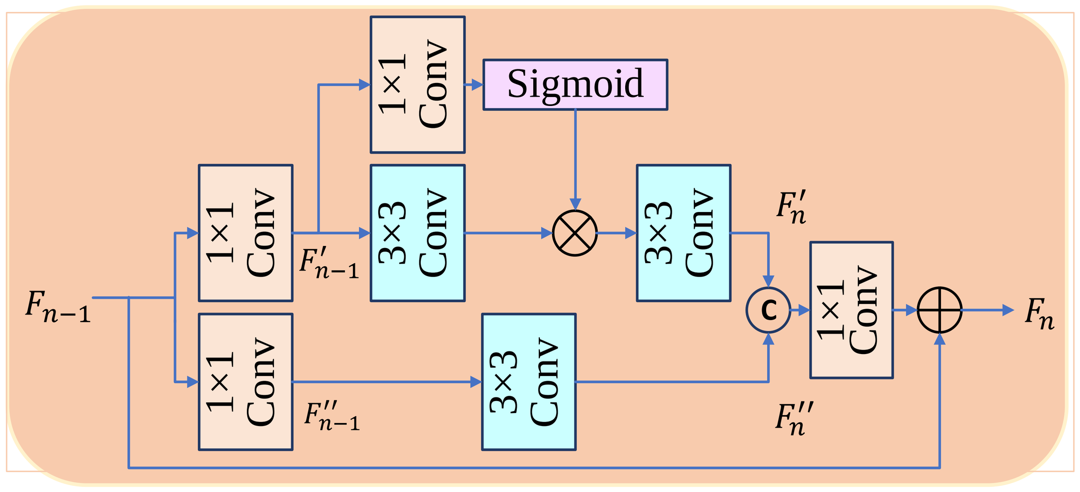
Motivations:



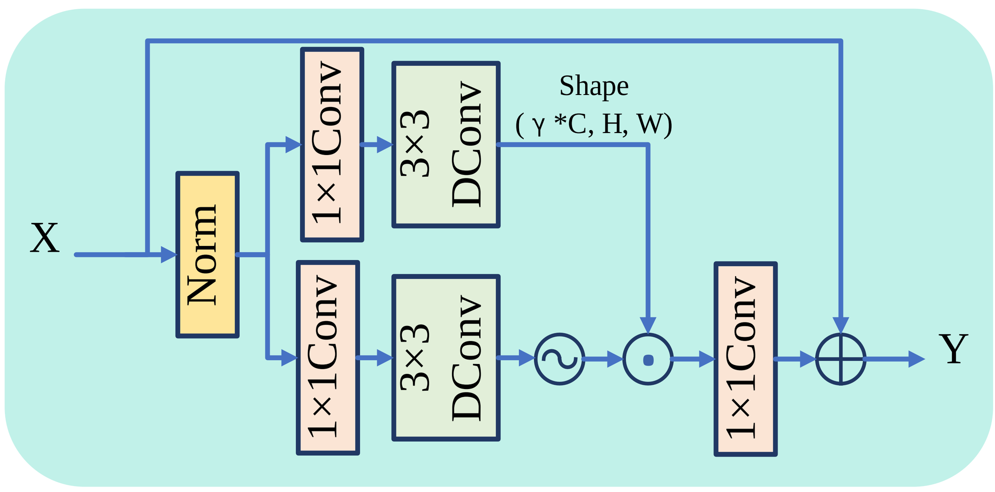
- Most of CNN-based SR networks have a large number of parameters, resulting in the limitation of the application of SR technology in edge devices.
- Most of lightweight SR methods focus on local contextual information and do not consider global similar textures, leading to problems such as artifacts in the recovered image.
- The limited receptive field of convolution operation is difficult to capture globally similar features, resulting in a poor trade-off between performance and complexity.

How to design a **lightweight** transformer to **effectively** perform single image super-resolution?

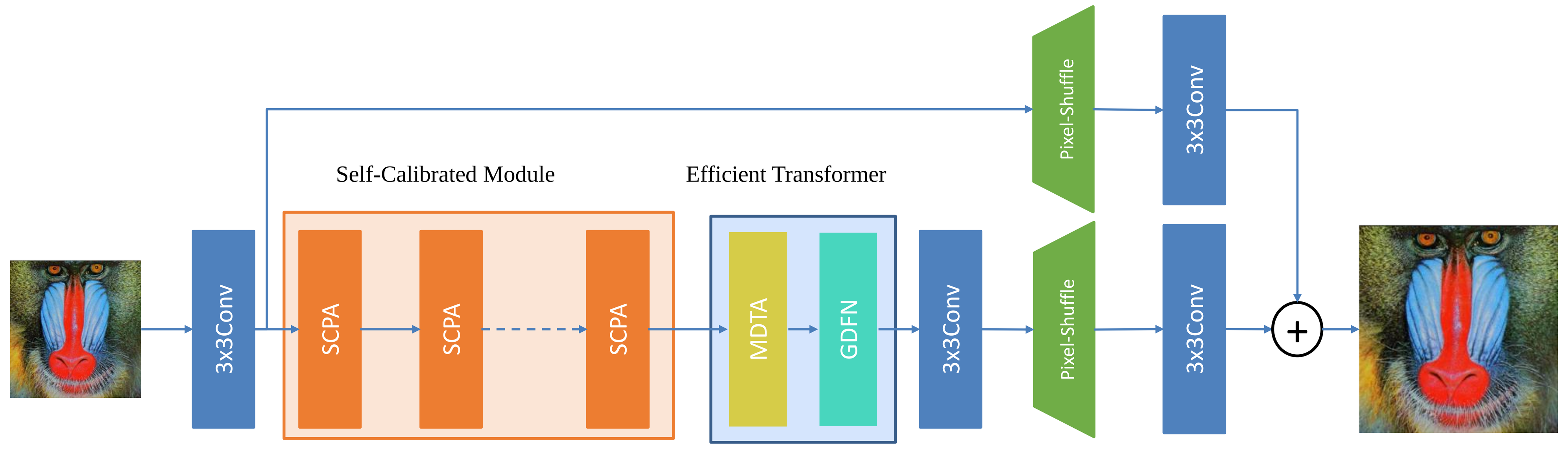
Self-Calibrated Efficient Transformer Network (SCET)

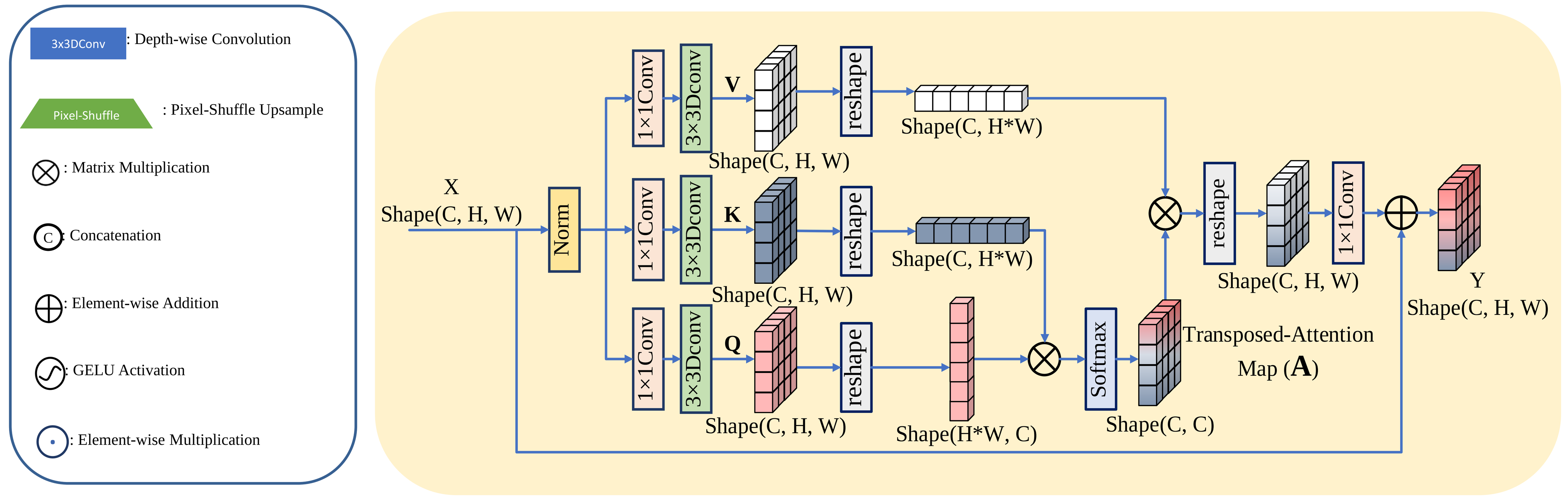


(a). Self-Calibrated convolutions with Pixel Attention (SCPA)



(c). Gated-Dconv feed-forward network (GDFN)





(b). Multi-Dconv Head Transposed Attention (MDTA)

- We introduce the efficient transformer design to the lightweight SISR task, effectively exploiting to the property that the transformer module can capture long-range dependencies, avoiding the problem of wrong textures generated by current lightweight SR methods.
- We design the SC module as the high-performance extractor. Compared with the information distillation mechanism in the IMDB block, the SC module employs a more efficient feature propagation strategy, achieving better performance with fewer parameters and less computational effort.

Quantitative Results:

Method	Scale	Params	Set5 PSNR/SSIM	Set14 PSNR/SSIM	B100 PSNR/SSIM	Urban100 PSNR/SSIM	Manga109 PSNR/SSIM
Bicubic	×2	-	33.66/0.9299	30.24/0.8688	29.56/0.8431	26.88/0.8403	30.80/0.9339
SRCNN [11]		8K	36.66/0.9542	32.45/0.9067	31.36/0.8879	29.50/0.8946	35.60/0.9663
VDSR [20]		666K	37.53/0.9587	33.03/0.9124	31.90/0.8960	30.76/0.9140	37.22/0.9750
DRRN [34]		298K	37.74/0.9591	33.23/0.9136	32.05/0.8973	31.23/0.9188	37.88/0.9749
DRCN [21]		1,774K	37.63/0.9588	33.04/0.9118	31.85/0.8942	30.75/0.9133	37.55/0.9732
IDN [19]		553K	37.83/0.9600	33.30/0.9148	32.08/0.8985	31.27/0.9196	38.01/0.9749
CARN [2]		1,592K	37.76/0.9590	33.52/0.9166	32.09/0.8978	31.92/0.9256	38.36/0.9765
IMDN [18]		694K	38.00/0.9605	33.63/0.9177	32.19/0.8996	32.17/0.9283	38.88/0.9774
PAN [51]		261K	38.00/0.9605	33.59/0.9181	32.18/0.8997	32.01/0.9273	38.70/0.9773
RFDN [28]		534K	38.05/0.9606	33.68/0.9184	32.16/0.8994	32.12/0.9278	38.88/0.9773
A^2F-M [42]		999K	38.04/0.9607	33.67/0.9184	32.18/0.8996	32.27/0.9294	38.87/0.9774
SCET (Ours)		683K	38.06/0.9615	33.78/0.9198	32.24/0.9006	32.38/0.9299	39.86/0.9821

Qualitative Results:

- Visual comparison for × 4 SR.