Wide Feature Projection with Fast and Memory-Economic Attention for Efficient Image Super-Resolution

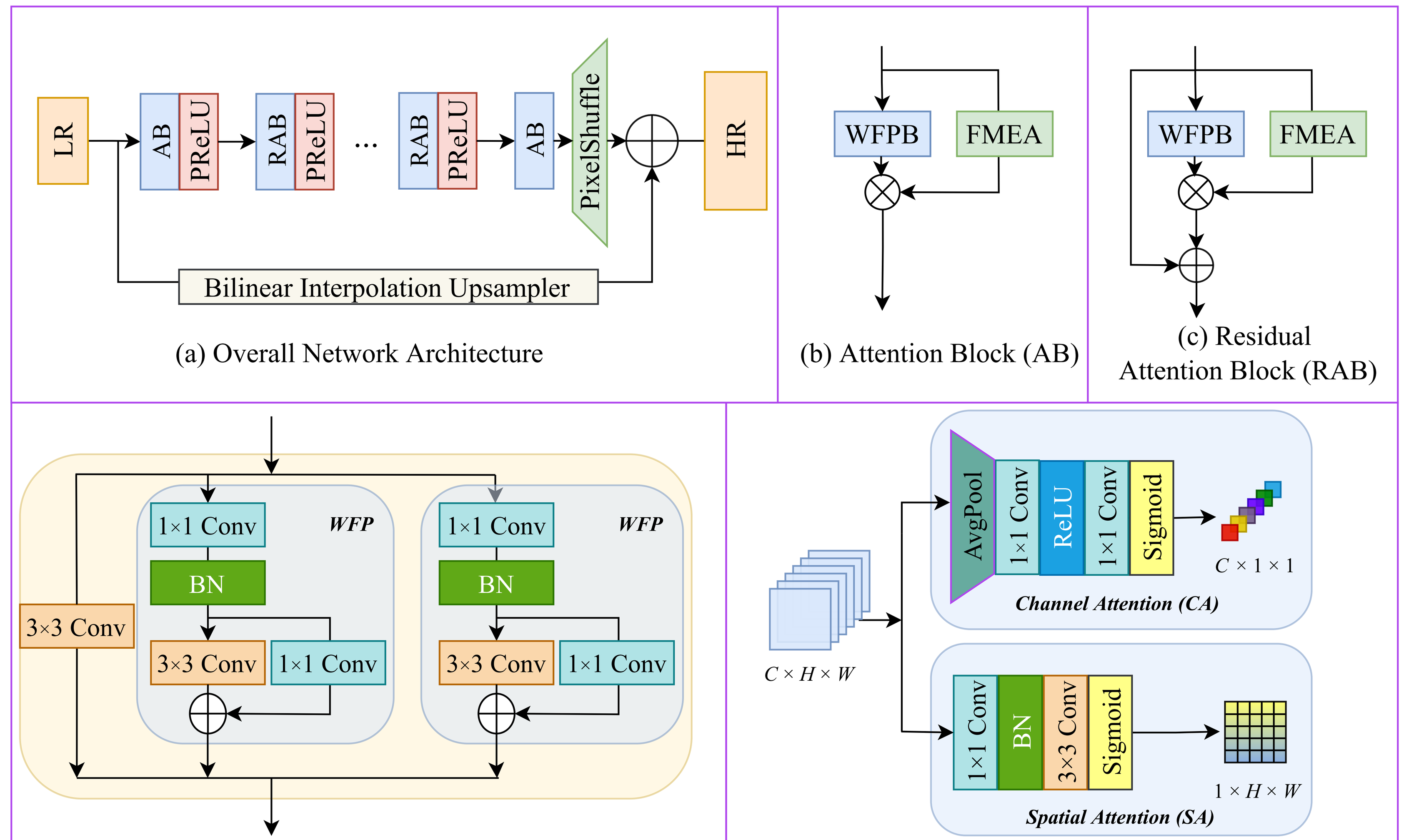
Minghao Fu1 Dongyang Zhang1 Min Lei1,2
Kun He1,2 Changyu Li1 Jie Shao1,2

[1]University of Electronic Science and Technology of China Chengdu, China
[2]Sichuan Artificial Intelligence Research Institute Yibin, China

BMVC 2022

Introduction:

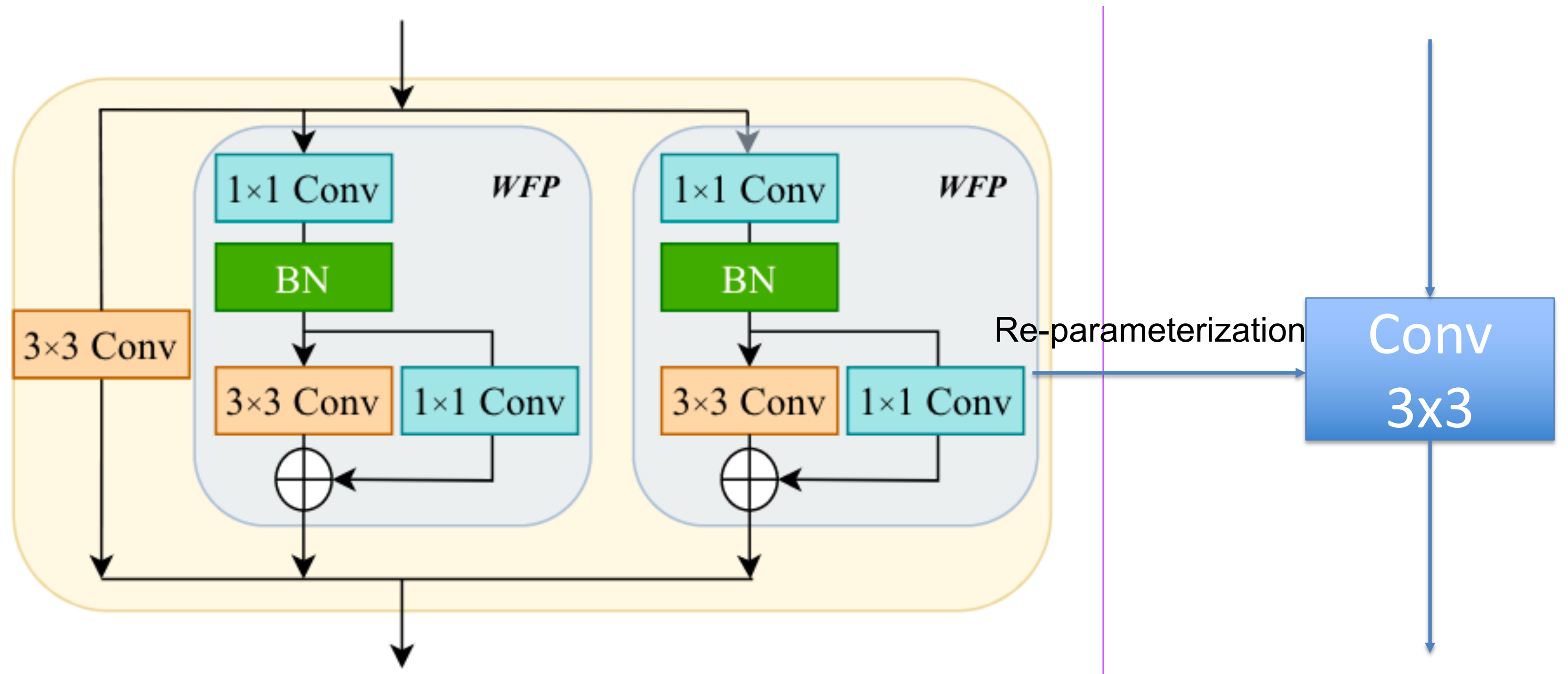
- Issue: Wide feature maps often lead to difficulty in reaching convergence due to information redundancy.
- Issue: Previous attention structures were computationally complex and occupied much memory.
- The main contributions of our work can be summarized as follows:
  - We propose a Wide Feature Projector (WFP) module to train a wider network. Our experiments show that WFP could largely alleviate training difficulties in the larger expanding ratio than the common expand-and-squeeze structure.
  - We explore the factors of memory variation in attention by analyzing the running memory of representative structures. Through our careful analysis, we introduce a Fast and Memory-Economic Attention (FMEA) module that costs little resources with much improved performance.
  - Equipped with FMEA and re-parameterization strategy, we further design a Wide Feature Projection network (WFPN). With lower memory consumption and computation complexity, WFPN demonstrates a comparable performance to state-of-the-art lightweight SR methods.



(a) Overall Network Architecture (b) Attention Block (AB) (c) Residual Attention Block (RAB) (d) Wide Feature Projection Block (WFPB) (e) Fast and Memory-Economic Attention (FMEA)

Wide Feature Projection Block (WFPB)

- Since expanding network width with the re-parameterization is free, can we get more benefit from over-parameterization by further expanding wide features?
- Using a channel-wise compression as a bridge. Re-parameterization strategy is adopted for the fast inference.



(d) Wide Feature Projection Block (WFPB)

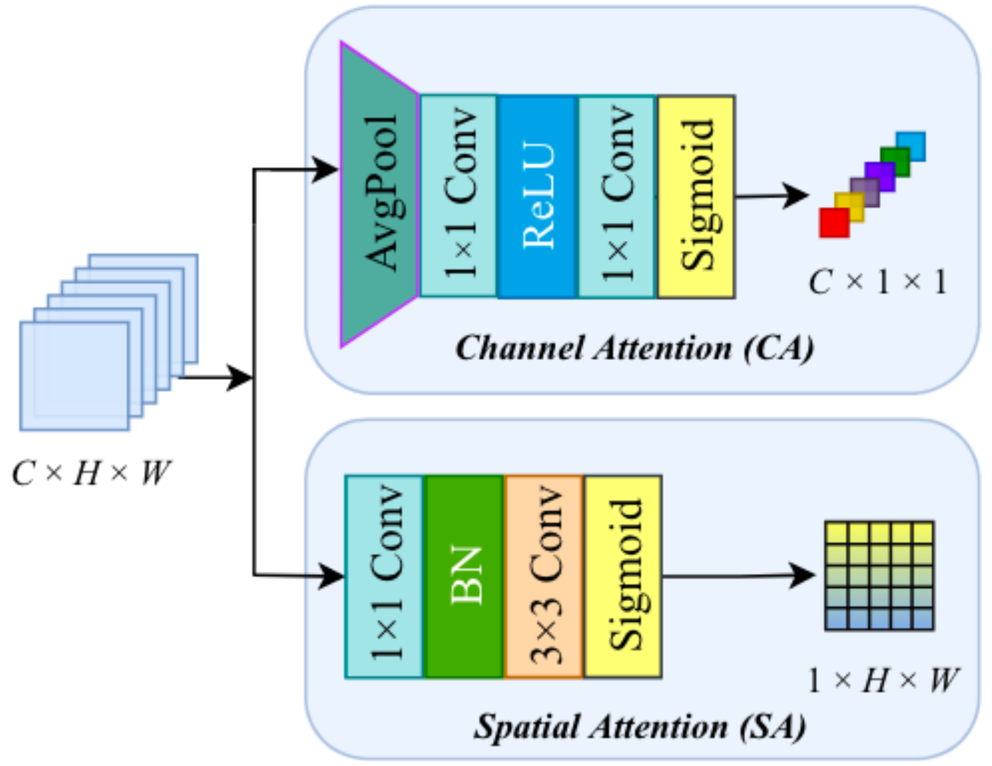
Memory consumption analysis

- To directly display memory consumption rather than analyzing it from experiences, we construct models with typical attention structures and identity mapping.
- Conclusions from table as follows: 1) Multi-branch structure leads to an increment of M_{kept}. 2) Identity has to be kept in memory to conduct other computation. 3) Identity memory sharing. 4) A valid method to reduce peak memory consumption is to avoid generating a large attention map.

Identity		Attention		Identity + Attention	
Base Block	Memory (M)	Base Block	Memory (M)	Base Block	Memory (M)
$Conv(x)$	39.92	$Conv(x) \times CA(x)$	54.27	$Conv(x) \times CA(x) + x$	54.27
$Conv(x) + x$	54.26	$Conv(x) \times SA(x)$	54.49	$Conv(x) \times SA(x) + x$	54.49
$Conv(x) + 2x$	69.00	$Conv(x) \times PA(x)$	69.00	$Conv(x) \times PA(x) + x$	69.00

Fast and memory-economic attention

- Costing little resources with much improved performance.
- Justifying our memory analysis.



(e) Fast and Memory-Economic Attention (FMEA)

Summary/Conclusion

- Wpropose a Wide Feature Projection Block (WFPB) based on Wide Feature Projection (WFP) to train a wider network,
- we design a Fast and MemoryEconomic Attention (FMEA) to minimize resources of attention mechanism.
- WFPN achieves better performance than previous methods with low resource consumption and a fast running speed.

Dataset	Scale	Bicubic	CARN [2] (1592K)	IMDN [14] (715K)	RFDN [21] (550K)	RLFN [17] (543K)	WFPN (ours) (633K)
Set5	×2	33.66 / 0.9299	37.76 / 0.9590	38.00 / 0.9605	38.05 / 0.9606	38.07 / 0.9607	38.08 / 0.9607
	×4	28.42 / 0.8104	32.13 / 0.8937	32.21 / 0.8948	32.24 / 0.8952	32.24 / 0.8952	32.25 / 0.8954
Set14	×2	30.24 / 0.8688	33.52 / 0.9166	33.63 / 0.9177	33.68 / 0.9184	33.72 / 0.9187	33.69 / 0.9179
	×4	26.00 / 0.7027	28.60 / 0.7806	28.58 / 0.7811	28.61 / 0.7819	28.62 / 0.7813	28.65 / 0.7813
B100	×2	29.56 / 0.8431	32.09 / 0.8978	32.19 / 0.8996	32.16 / 0.8994	32.22 / 0.9000	32.24 / 0.9002
	×4	25.96 / 0.6675	27.58 / 0.7349	27.56 / 0.7353	27.57 / 0.7360	27.60 / 0.7364	27.62 / 0.7367
Urban100	×2	26.88 / 0.8403	31.92 / 0.9256	32.17 / 0.9283	32.12 / 0.9278	32.33 / 0.9299	32.29 / 0.9285
	×4	23.14 / 0.6577	26.07 / 0.7837	26.04 / 0.7838	26.11 / 0.7858	26.17 / 0.7877	26.19 / 0.7878
Manga109	×2	31.01 / 0.8923	38.36 / 0.9765	38.88 / 0.9774	38.88 / 0.9773	–	38.90 / 0.9773
	×4	26.66 / 0.7512	30.47 / 0.9084	30.45 / 0.9075	30.58 / 0.9089	–	30.55 / 0.9085