Lightweight Wavelet-based Transformer for Image Super-resolution

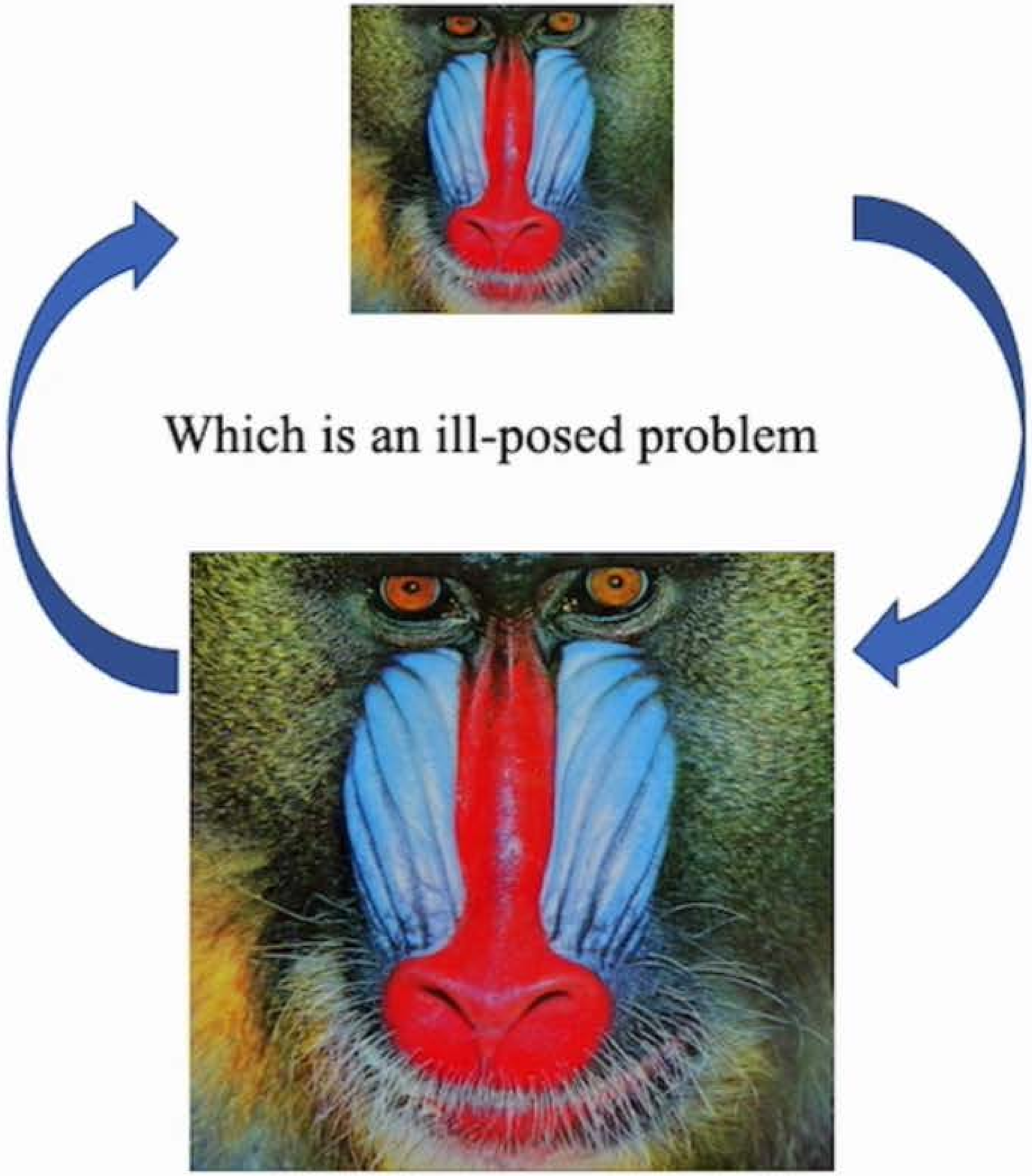
Jinye Ran and Zili Zhang*

College of Computer and Information Science, Southwest University, Chongqing 400715, China

Introduction

- Problem Definition
  - Super-resolution (SR) aims to recover a high-resolution (HR) image from a low-resolution (LR) image counterpart.
  - Pursuing the SR quality of the model while ignoring the lightweight problem.

- Main Challenge
  - Generally, model scale and model quality are a trade-off issue.
  - How to get a better SR quality on an acceptable model scale?



Introduction

- Contribution
  - A Lightweight Transformer Backbone (LTB) is designed to implicitly mine self-similarity information in images to ensure SR quality.
  - The mapping from LR to HR is fitted on the wavelet domain, while the stability of the inverse wavelet transform is guaranteed by Wavelet Coefficient Enhancement Backbone (WCEB).
  - Achieve competitive results on multiple publicly available benchmarks.

Approach

- Overall Architecture

$$F_0 = f_{1*1}(f_{3*3}(I_{LR}))$$

$$I^W = concat(SWT(F_0))$$

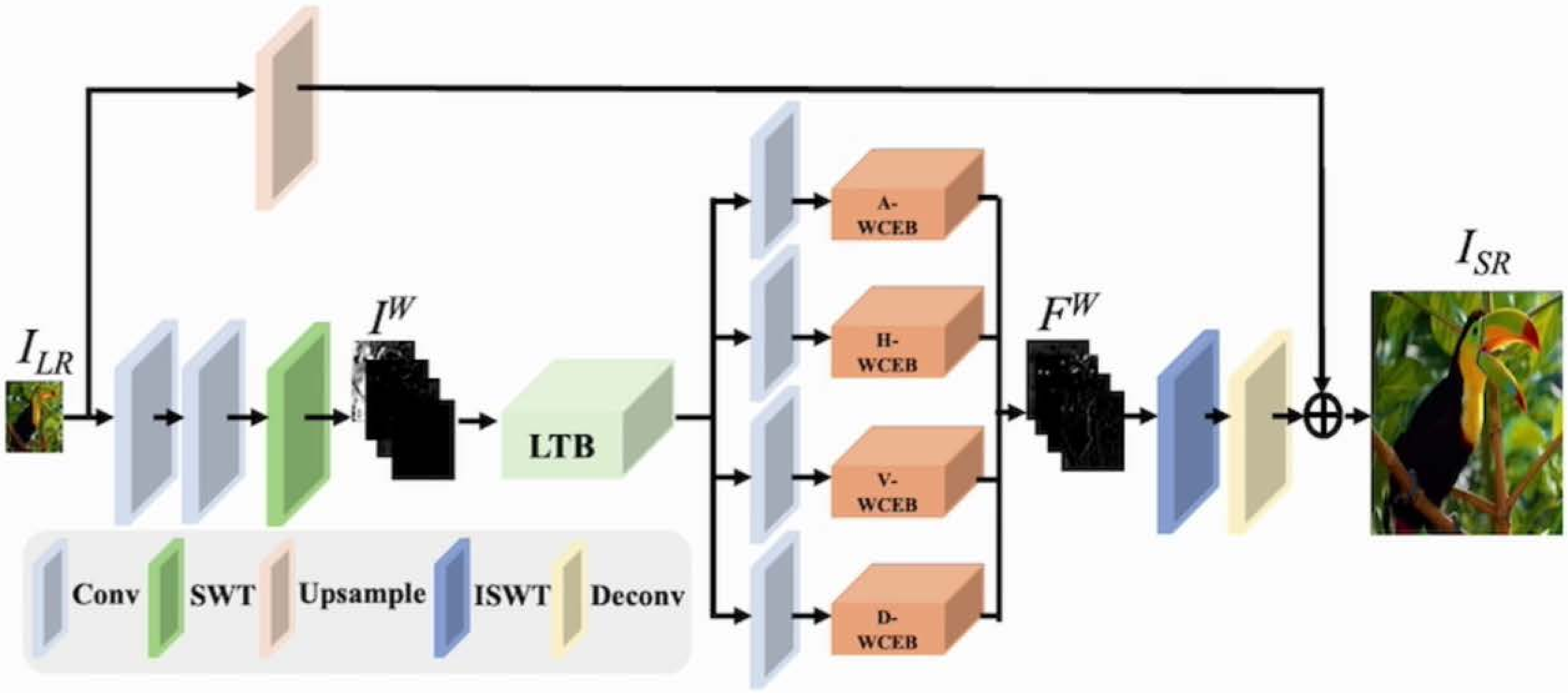
$$F_L = \phi^5(\psi(\phi^8(f_{group}(I^W))))$$

$$F_A, F_H, F_V, F_D = split(F_L)$$

$$F^W = concat(\sigma_{A,H,V,D}(F_A, F_H, F_V, F_D))$$

$$F_D = ISWT(F^W)$$

$$I_{SR} = f_{Deconv}(f_{3*3}(F_d)) + f_{up}(I_{LR})$$



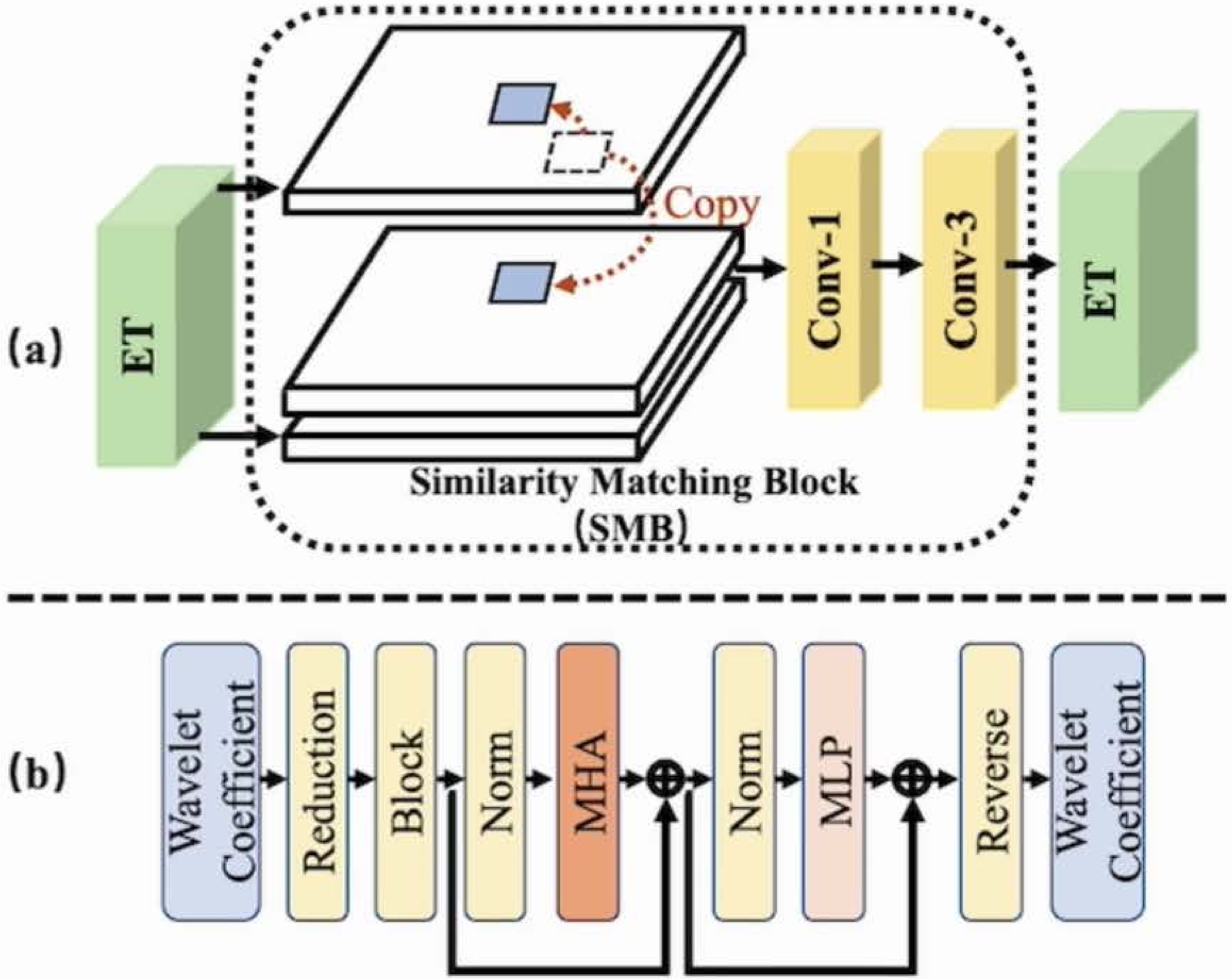
Approach

- Lightweight Transformer Backbone

$$S_{m1} = f_{partitioning}(f_{reduction}(S_i))$$
$$S_{m2} = MHA(Norm(S_{m1})) + S_{m1}$$
$$S_o = f_{reverse}(MLP(Norm(S_{m2})) + S_{m2})$$

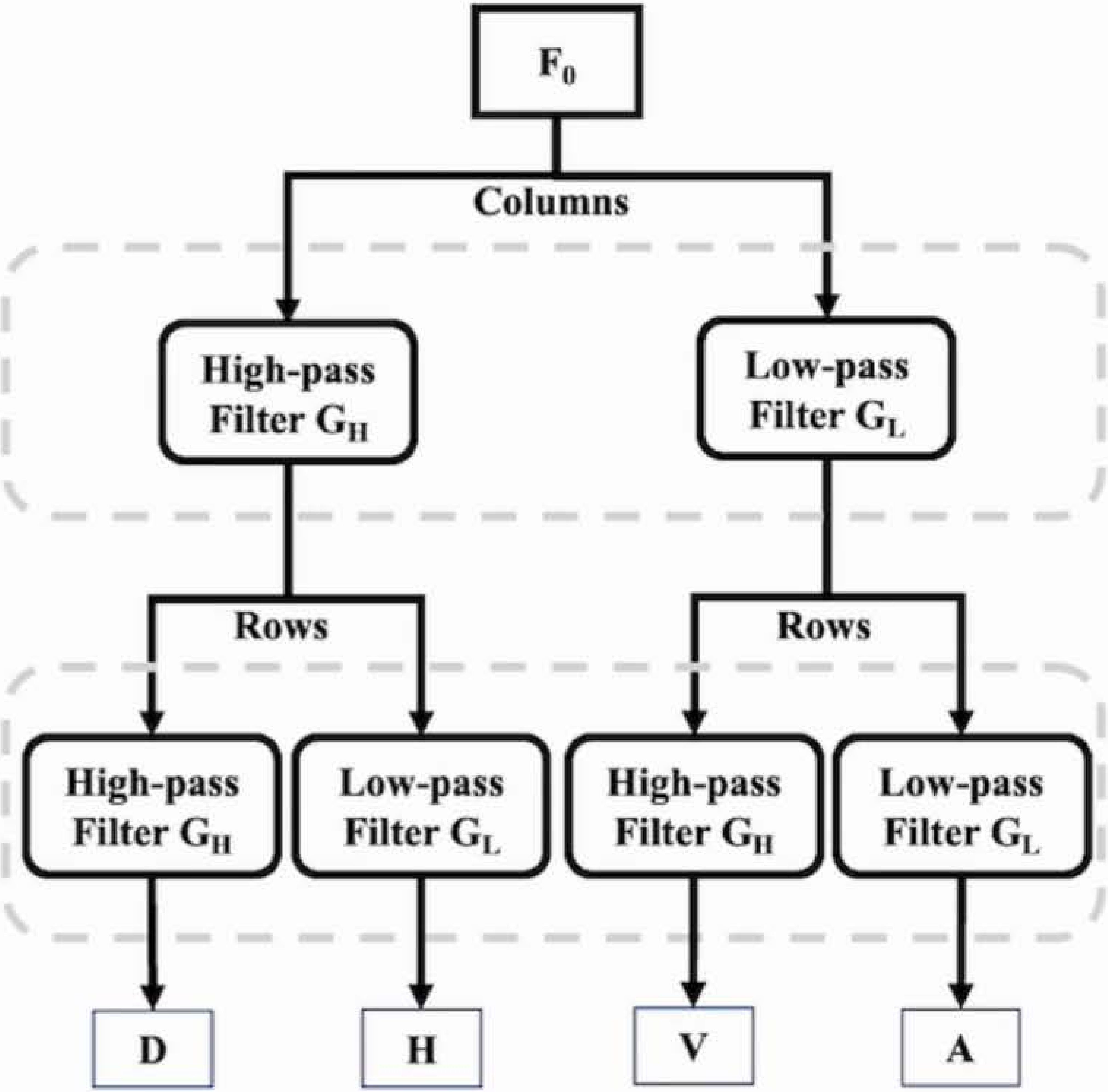
- Similarity Matching Block

$$p_{i1c,j1c} = \arg\max_{p_{i2,j2}} \langle \frac{p_{i1,j1}}{\|p_{i1,j1}\|}, \frac{p_{i2,j2}}{\|p_{i2,j2}\|} \rangle$$
$$s.t. \quad |i1 - i2| + |j1 - j2| \neq 0$$



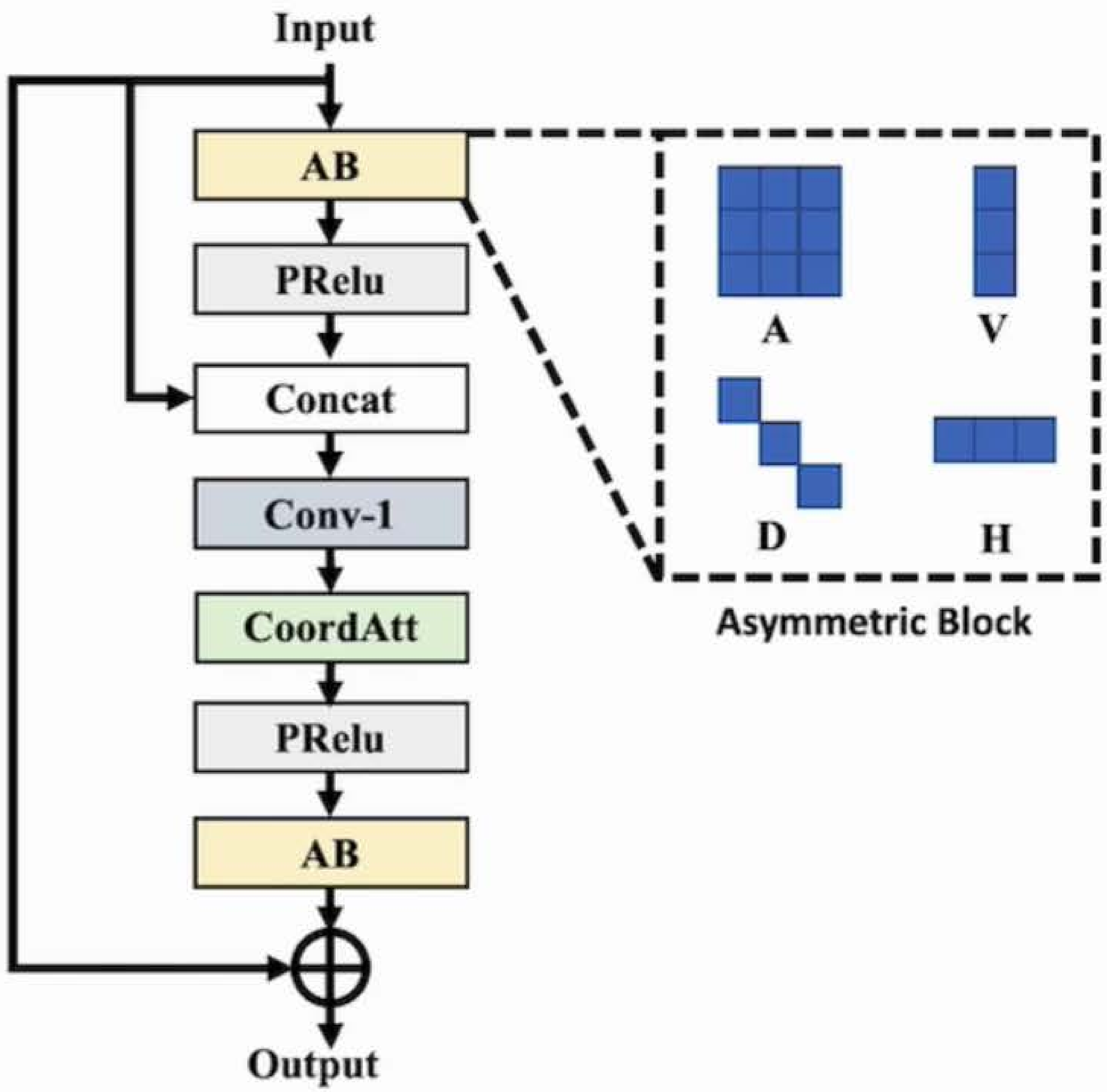
Approach

- Wavelet Transform
  - Stationary wavelet transform
  - Inverse stationary wavelet transform
  - Advantages of wavelet transform



Approach

- Wavelet coefficient Enhancement Backbone
  - Different Asymmetric Block for different wavelet coefficients
  - Different wavelet coefficients' channel redundancy
  - Different wavelet coefficients' structure information



Evaluation

- Benchmark
  - Set5, Set14, BSD100, Urban100, Manga109
- Qualitative evaluation
  - PSNR/SSIM
  - ×2 ×3 ×4
- Quantitative evaluation
  - Subjective visual

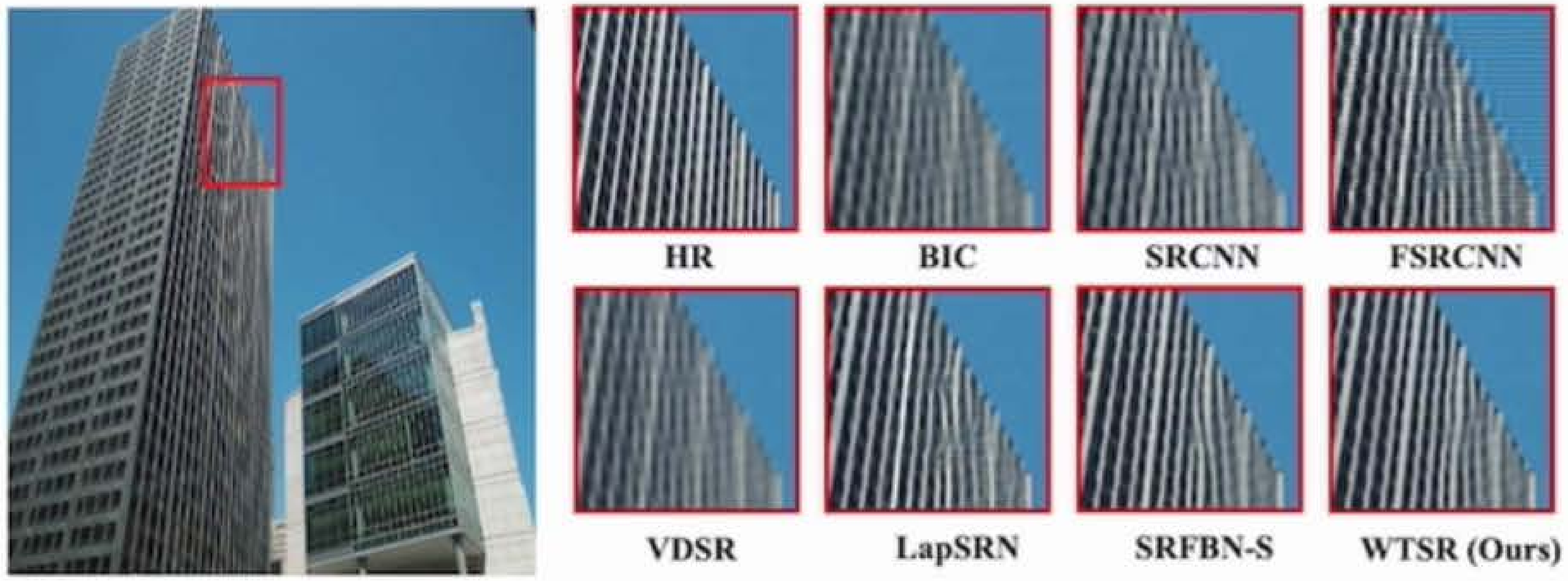


Table 1. Quantitative results of WTSR compared with other lightweight super-resolution network, the best model performance is **highlighted** and the second performance is underlined.

Methods	Scales	Params	Set5 PSNR/SSIM	Set14 PSNR/SSIM	BSD100 PSNR/SSIM	Urban100 PSNR/SSIM	Manga109 PSNR/SSIM
Bicubic		-	33.66/0.930	30.24/0.869	29.56/0.843	26.88/0.840	30.30/0.934
SRCNN[32]		8K	36.66/0.954	32.45/0.907	31.36/0.888	29.50/0.895	35.60/0.966
FSRCNN[6]		13K	37.00/0.956	32.63/0.909	31.53/0.892	29.88/0.902	36.67/0.971
VDSR[15]		666K	37.53/0.959	33.03/0.912	31.90/0.896	30.76/0.914	37.22/0.975
DWSR[8]		374K	37.43/0.957	33.07/0.911	31.80/0.894	31.46/0.916	-/-
LapSRN[19]	x2	813K	37.52/0.959	32.99/0.912	31.80/0.895	30.41/0.910	37.27/0.974
MemNet[31]		678K	37.78/0.960	33.28/0.914	32.08/0.898	31.31/0.920	37.72/0.974
CARN-M[2]		412K	37.53/0.958	33.26/0.914	31.92/0.896	31.23/0.919	-/-
SRFBN-S[20]		483K	37.78/0.960	33.35/0.916	32.00/0.897	31.41/0.921	38.06/0.976
WDRN-S[33]		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-/-
WTSR		528K	**37.95**/0.961	**33.51**/0.915	**32.09**/0.893	**31.91/0.928**	**38.42/0.976**
Bicubic		-	30.39/0.868	27.55/0.774	27.21/0.739	24.46/0.735	26.95/0.856
SRCNN[32]		8K	32.75/0.909	29.30/0.822	28.41/0.786	26.24/0.799	30.48/0.912
FSRCNN[6]		13K	33.18/0.914	29.37/0.824	28.53/0.791	26.43/0.808	31.10/0.921
VDSR[15]		666K	33.66/0.921	29.77/0.831	28.82/0.798	27.14/0.828	32.01/0.934
DWSR[8]		374K	33.82/0.922	29.83/0.831	-/-	-/-	-/-
LapSRN[19]	x3	813K	33.81/0.922	29.79/0.833	28.82/0.798	27.07/0.828	32.21/0.935
MemNet[31]		678K	34.09/0.925	30.00/0.835	28.96/0.800	27.56/0.838	32.51/0.937
CARN-M[2]		412K	33.99/0.924	30.08/0.837	28.91/0.800	27.55/0.839	-/-
SRFBN-S[20]		483K	34.20/0.926	30.10/**0.837**	28.96/0.801	27.66/0.842	33.02/0.94
WDRN-S[33]		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-/-
WTSR		558K	**34.27/0.925**	30.12/0.836	**28.98/0.802**	27.69/0.841	**33.11/0.941**
Bicubic		-	28.42/0.810	26.00/0.703	25.96/0.668	23.14/0.658	24.89/0.787
SRCNN[32]		8K	30.48/0.863	27.50/0.751	26.90/0.710	24.52/0.722	27.58/0.856
FSRCNN[6]		13K	30.72/0.866	27.61/0.755	26.98/0.715	24.62/0.728	27.90/0.861
VDSR[15]		666K	31.35/0.884	28.01/0.767	27.29/0.725	25.18/0.752	28.83/0.887
DWSR[8]		374K	31.39/0.883	28.04/0.767	27.25/0.724	25.26/0.755	-/-
LapSRN[19]	x4	813K	31.54/0.885	28.09/0.770	27.32/0.728	25.21/0.756	29.09/0.890
MemNet[31]		678K	31.74/0.889	28.26/0.772	27.40/0.728	25.50/0.763	29.42/0.894
CARN-M[2]		412K	31.92/0.890	28.42/0.776	27.44/0.730	25.62/0.769	-/-
SRFBN-S[20]		483K	31.98/0.892	28.45/0.778	27.44/0.731	25.71/0.772	29.91/0.901
WDRN-S[33]		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-/-
WTSR		593K	**32.16/0.895**	**28.57/0.781**	**27.56/0.735**	**26.03/0.784**	**30.44/0.908**

Ablation

- Ablation on different Wavelet Transform
  - None
  - DWT
  - SWT
- Ablation on different LTB
  - Similarity Matching Block
  - Efficient Transformer Encoder's partitioning size
  - Efficient Transformer Encoder's order
- Ablation on WCEB
  - Different number of WECM
  - Different type of WECM

Table 2. Comparisons on PSNR/SSIM of WTSR with different wavelet transform. Best results are **highlighted**.

Wavelet Transform Type	Params	PSNR/SSIM Set5	Set14	BSD100	Urban100	Manga109
None	593K	32.01/0.893	27.47/0.781	27.51/0.734	25.84/0.777	20.22/0.905
DWT	593K	27.56/0.790	25.51/0.682	25.54/0.647	22.69/0.635	24.19/0.767
SWT	593K	**32.16/0.895**	**28.57/0.781**	**27.56/0.735**	**26.03/0.784**	**30.44/0.908**

Table 3. Comparisons on PSNR/SSIM of WTSR with different network of LTB. Best results are **highlighted**. The number after T indicates the partitioning size of ET encoder, S represents SMB, and the arrow denotes the direction of data flow.

The Network of LTB	Params	PSNR/SSIM Set5	Set14	BSD100	Urban100	Manga109
T5 → T5	567K	31.92/0.892	27.43/0.779	27.46/0.734	25.78/0.777	30.06/0.904
T8 → T8	569K	31.96/0.892	28.46/0.779	27.48/0.733	25.79/0.776	30.09/0.904
T5 → T8	568K	31.97/0.893	28.44/0.779	27.48/0.733	25.79/0.776	30.04/0.903
T8 → T5	568K	32.00/0.893	28.47/0.779	27.50/0.733	25.87/0.778	30.11/0.904
T8 → S → T5	593K	**32.16/0.895**	**28.57/0.781**	**27.56/0.735**	**26.03/0.784**	**30.44/0.908**

Table 4. Study the effect of each WCEB on PSNR/SSIM, Best results are **highlighted**.

WCEM Type	Params	PSNR / SSIM Set5	Set14	BSD100	Urban100	Manga109
None	416K	31.91 / 0.892	28.44 / 0.778	27.47 / 0.732	25.75 / 0.7735	29.92 / 0.901
A	482K	32.01 / 0.893	28.42 / 0.779	27.48 / 0.734	25.80 / 0.7782	29.99 / 0.924
A + H	519K	32.06 / 0.894	28.54 / 0.780	27.52 / 0.733	25.91 / 0.7791	30.27 / 0.905
A + H + V	556K	32.10 / 0.894	28.54 / 0.781	27.54 / 0.735	25.97 / 0.7823	30.31 / 0.907
A + H + V + D	593K	**32.16/0.895**	**28.57/0.781**	**27.56/0.735**	**26.03/0.784**	**30.44/0.908**

Potential Limitations

- Processing images at arbitrary resolution
    - Before the wavelet-based transformer, there is a fill operation that fills the resolution of the image to an integer multiple of the partitioning size. Although this does not increase the number of parameters in the wavelet-based transformer, it slightly affects the runtime of the model.
- Hyperparametric sensitivity
    - According to the existing experimental data, WTSR is very sensitive to the hyperparameters of the model, which is not conducive to the rapid iteration of the project.

Conclusions

- A lightweight network called WTSR is proposed to extend the application scenarios of super-resolution algorithm.
- In the WTSR, many useful components include LTB, SMB and WCEB, have been proposed to balance the size and accuracy of the network.
- In the future, we will extend the proposed WTSR to specific mobile devices.