Binarized Diffusion Model for Image Super-Resolution

Zheng Chen[1], Haotong Qin[2*], Yong Guo[3], Xiongfei Su[4], Xin Yuan[4], Linghe Kong[1], Yulun Zhang[1*]

[1]Shanghai Jiao Tong University, [2]ETH Zürich, [3]Max Planck Institute for Informatics, [4]Westlake University

Introduction

Diffusion models (DMs) excel in SR tasks but face high costs. **Binarization** (1-bit quantization) reduces memory and computation. However, the architecture and iterative design of DM limit its application. To tackle this, we propose **BI-DiffSR**, a novel binarized DM for SR.

HR | Bicubic | SR3 (FP) | BBCU (ICLR'22) | BI-DiffSR (ours)

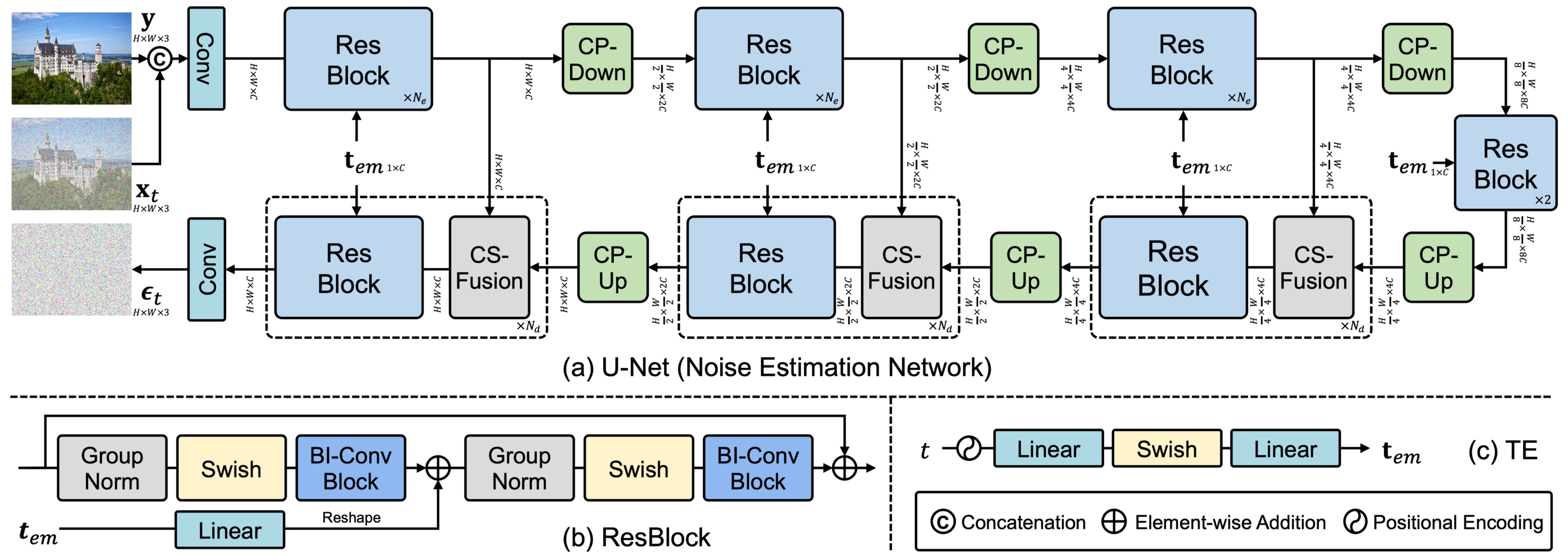
Contribution

- **Architecture:** Design modules for binarization, including consistent-pixel-downsample (**CP-Down**) and upsample (**CP-Up**), and channel-shuffle-fusion (**CS-Fusion**).
- **Activation:** Introduce timestep-aware redistribution (**TaR**) and activation (**TaA**) to adapt activation distributions by timestep, enhancing binarized modules.
- **Performance:** Outperforms SOTA binarization methods and achieves perceptual quality comparable to full-precision models.

Project

Method

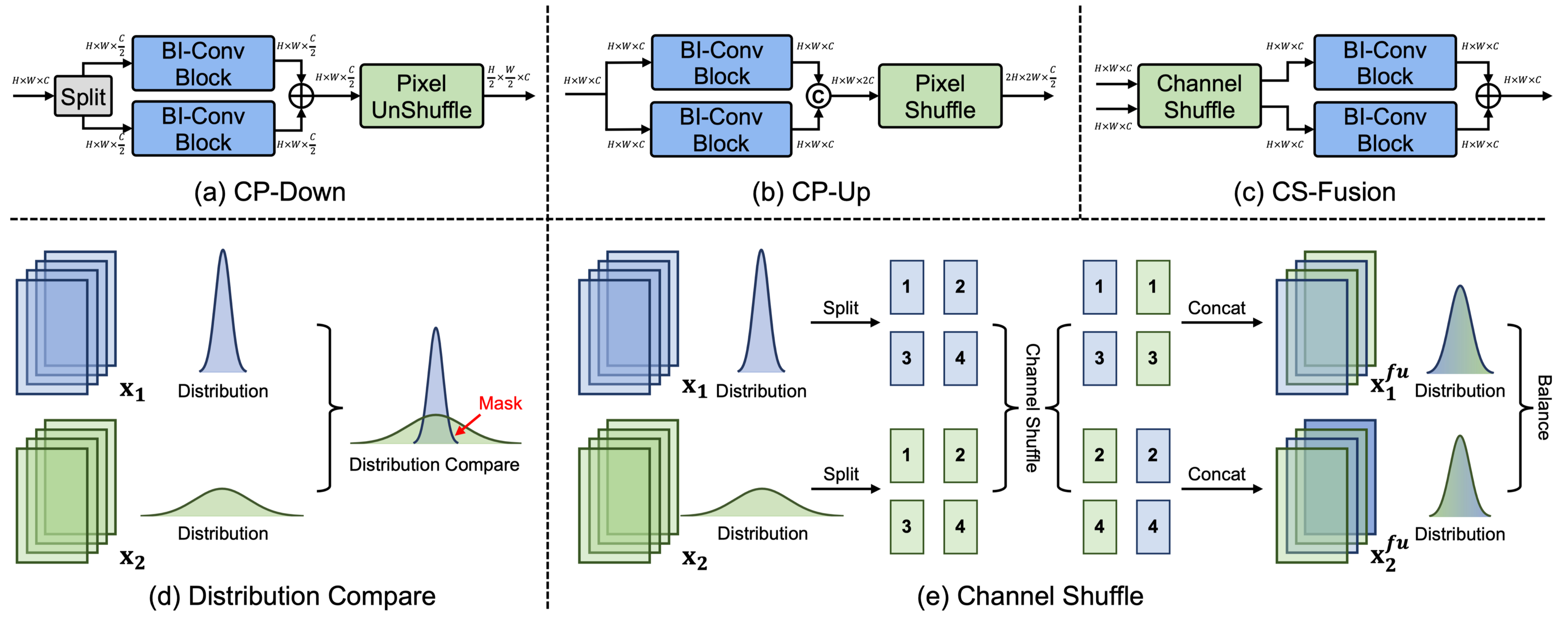
❖ Overall



(a) U-Net (Noise Estimation Network) (b) ResBlock (c) TE

❖ Model Structure

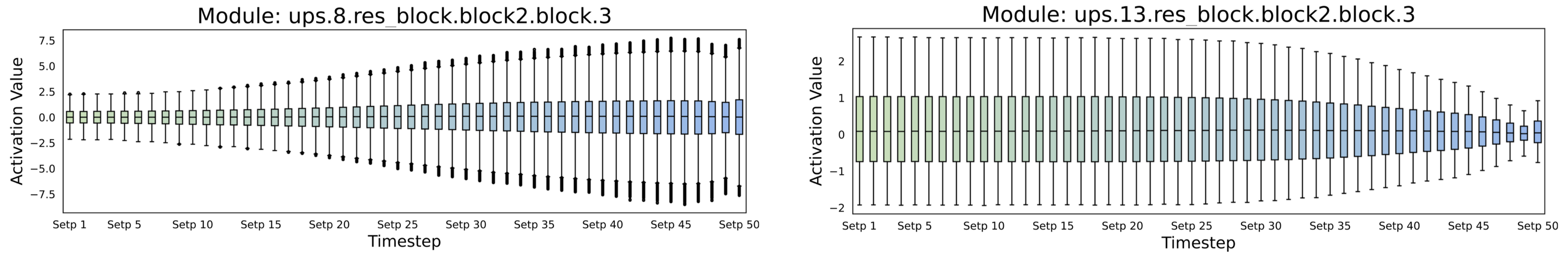
- **Challenge I: Dimension Mismatch.** Frequent changes in feature resolution cause dimension mismatches, blocking full-precision propagation.
- **Challenge II: Fusion Difficulty.** Significant activation range differences between the encoder and decoder hinder effective feature fusion in skip connections.



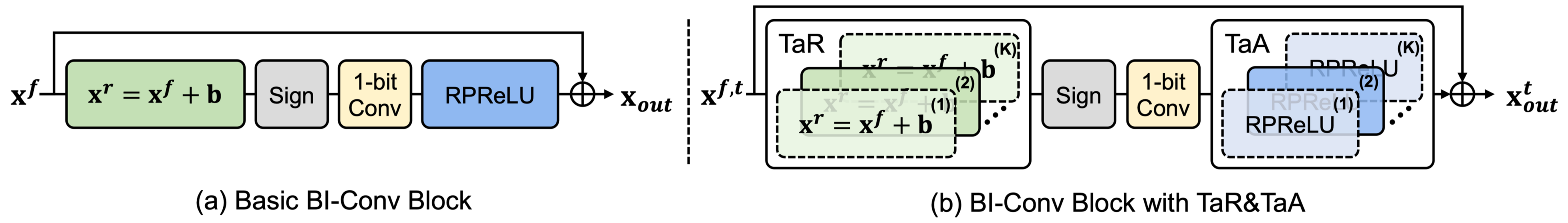
(a) CP-Down (b) CP-Up (c) CS-Fusion (d) Distribution Compare (e) Channel Shuffle

- **CP-Down/Up:** Ensures consistent feature reshaping, allowing identity shortcuts to maintain full-precision information transfer throughout the network effectively.
- **CS-Fusion:** Balances feature distribution by channel shuffle operation, ensuring better distribution matching and promoting more effective feature fusion.

❖ Activation Distribution



- Multi-step iterations in diffusion models change activation distributions, with adjacent timesteps appearing similar and distant ones differing significantly.



(a) Basic BI-Conv Block (b) BI-Conv Block with TaR&TaA

- **TaR/TaA:** Adjusts activations across timesteps, enhancing the binarized module.

Experiments

❖ Ablation Study

Method	Baseline	+Identity	+CP-Down&Up	+CS-Fusion	+TaR&TaA
Params (M)	4.29	4.29	4.29	4.30	4.58
OPs (G)	36.67	36.67	36.67	36.67	36.67
PSNR (dB)	27.66	29.29	31.08	31.99	32.66
LPIPS	0.0780	0.0658	0.0327	0.0261	0.0200

(a) Break-down ablation.

Method	Params (M)	OPs (G)	PSNR (dB)	LPIPS
Add	4.10	33.40	18.89	0.1695
Concat	4.29	36.67	31.08	0.0327
Split	4.30	36.67	29.67	0.0384
CS-Fusion	4.30	36.67	31.99	0.0261

(b) Ablation on feature fusion.

Method	TaR	TaA	Params (M)	Ops (G)	PSNR (dB)	LPIPS
w/o			4.30	36.67	31.99	0.0261
In	✓		4.37	36.67	29.27	0.0337
Out		✓	4.51	36.67	29.13	0.0308
All	✓	✓	4.58	36.67	32.66	0.0200

(c) Ablation on time aware module (TaR and TaA).

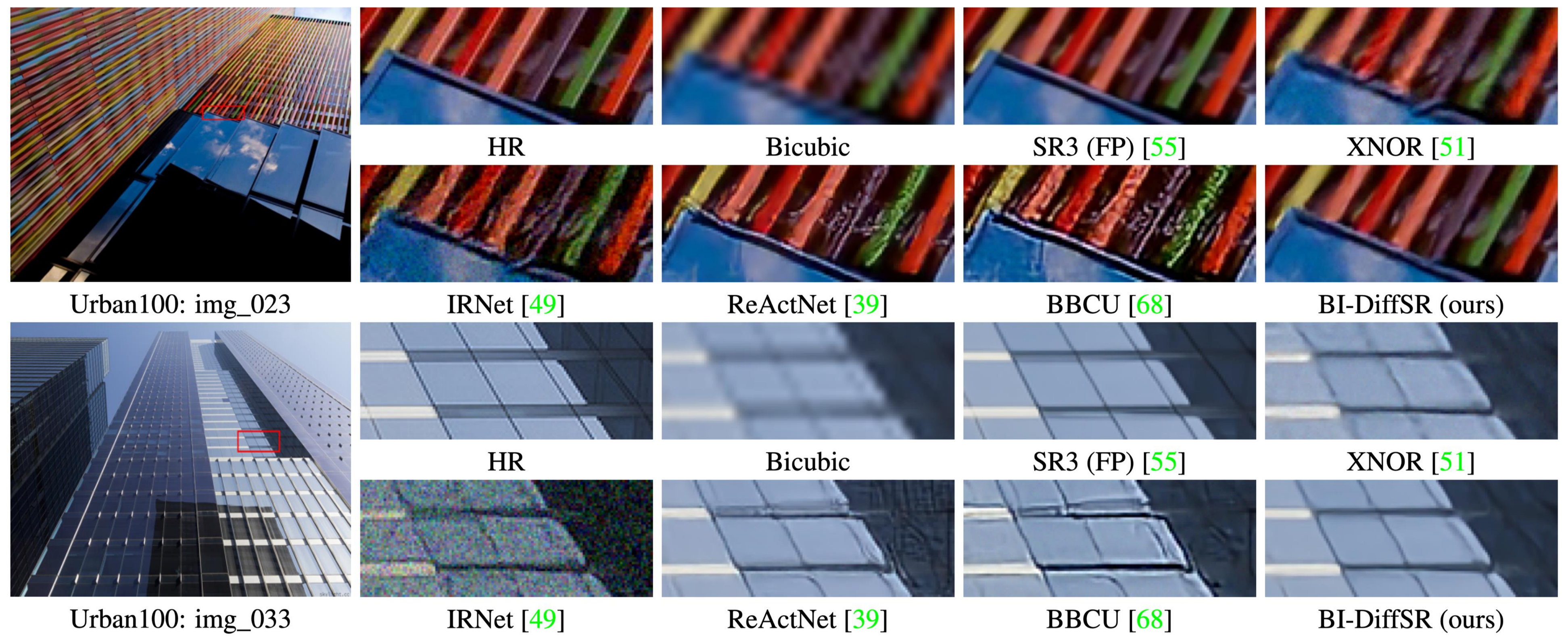
#Pair	1	2	5
Params (M)	4.30	4.37	4.58
OPs (G)	36.67	36.67	36.67
PSNR (dB)	31.99	32.42	32.66
LPIPS	0.0261	0.0229	0.0200

(d) Numbers (#) of bias and RPReLU pair.

❖ Quantitative Results

Method	Scale	Params (M)	Ops (G)	Set5 PSNR	Set5 SSIM	Set5 LPIPS	B100 PSNR	B100 SSIM	B100 LPIPS	Urban100 PSNR	Urban100 SSIM	Urban100 LPIPS	Manga109 PSNR	Manga109 SSIM	Manga109 LPIPS
Bicubic	×2	N/A	N/A	33.67	0.9303	0.1274	29.55	0.8431	0.2508	26.87	0.8403	0.2064	30.82	0.9349	0.1025
SR3 [55]	×2	55.41	176.41	36.69	0.9513	0.0310	30.41	0.8683	0.0700	30.29	0.9060	0.0430	35.11	0.9682	0.0161
BNN [19]	×2	4.78	37.93	13.97	0.5210	0.4529	13.73	0.4553	0.5784	12.75	0.4236	0.5575	9.29	0.3035	0.7489
DoReFa [74]	×2	4.78	37.93	16.43	0.6553	0.2662	16.11	0.5912	0.3972	15.09	0.5495	0.4055	12.35	0.4609	0.5047
XNOR [51]	×2	4.78	37.93	32.34	0.8661	0.0782	27.94	0.7548	0.1665	27.47	0.8225	0.1153	31.99	0.9428	0.0326
IRNet [49]	×2	4.78	37.93	32.55	0.9340	0.0446	27.76	0.8199	0.1115	26.34	0.8452	0.0913	23.89	0.7621	0.1820
ReActNet [39]	×2	4.85	37.93	34.30	0.9271	0.0351	28.36	0.8158	0.0943	27.43	0.8563	0.0731	32.16	0.9441	0.0379
BBCU [68]	×2	4.82	37.75	34.31	0.9281	0.0393	28.39	0.8202	0.0905	28.05	0.8669	0.0620	32.88	0.9508	0.0272
BI-DiffSR (ours)	×2	4.58	36.67	35.68	0.9414	0.0277	29.73	0.8478	0.0682	28.97	0.8815	0.0522	33.99	0.9601	0.0172

❖ Visual Results



Urban100: img_023 | HR | Bicubic | SR3 (FP) [55] | XNOR [51] | IRNet [49] | ReActNet [39] | BBCU [68] | BI-DiffSR (ours)

Urban100: img_033 | HR | Bicubic | SR3 (FP) [55] | XNOR [51] | IRNet [49] | ReActNet [39] | BBCU [68] | BI-DiffSR (ours)