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
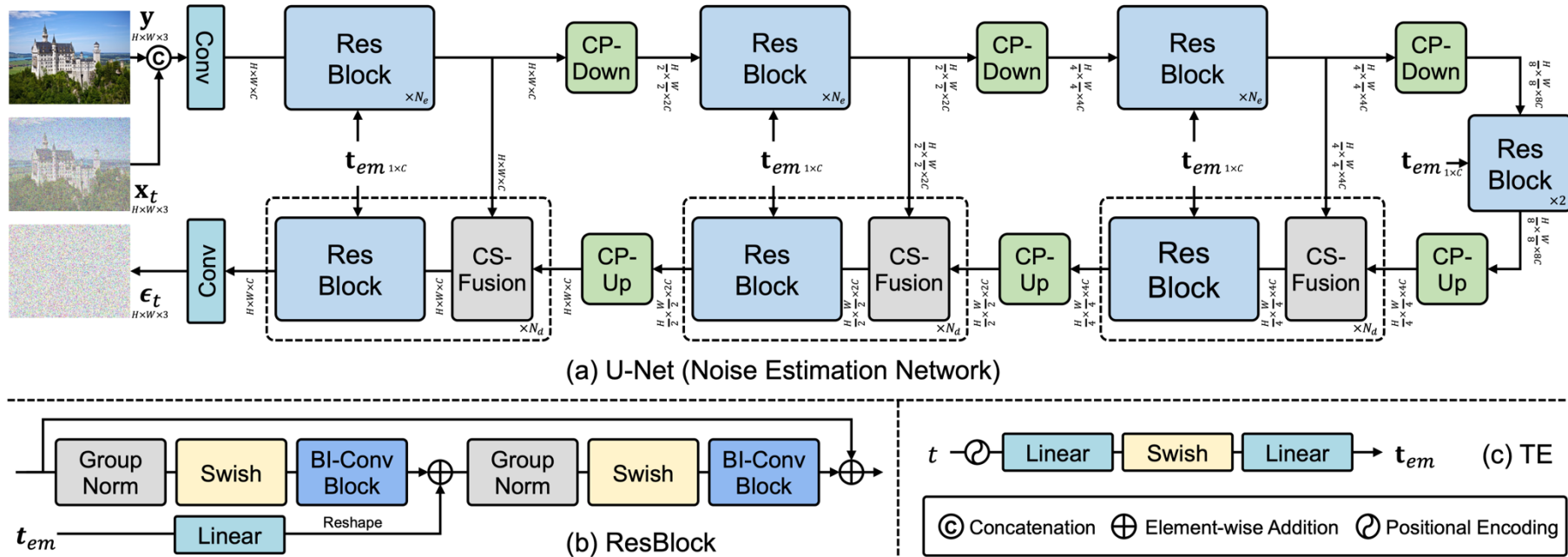
Introduction

Motivation

- **Diffusion models (DMs)** excel in SR tasks but face high costs.

- **Binarization** (1-bit quantization) reduces memory and computation.

- However, the architecture and multi-step iterative design of DM limit its application.

- We propose **BI-DiffSR**, a novel binarized DM for image SR.

| HR | Bicubic | SR3(FP) [1] | BBCU(ICLR'22) [2] | BI-DiffSR(ours) |

[1] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. TPAMI, 2022.
[2] Bin Xia, Yulun Zhang, Yitong Wang, Yapeng Tian, Wenming Yang, Radu Timofte, and Luc Van Gool. Basic binary convolution unit for binarized image restoration network. In ICLR, 2022.

(a) U-Net (Noise Estimation Network)

(b) ResBlock

(c) TE

© Concatenation   ⊕ Element-wise Addition   ⊘ Positional Encoding

Overall

- The overall structure of the noise estimation network in BI-DiffSR.
- **Analysis:** UNet struggles with binarization due to dimension mismatch and fusion.

Method

(a) CP-Down

(b) CP-Up

(c) CS-Fusion

(d) Distribution Compare

(e) Channel Shuffle
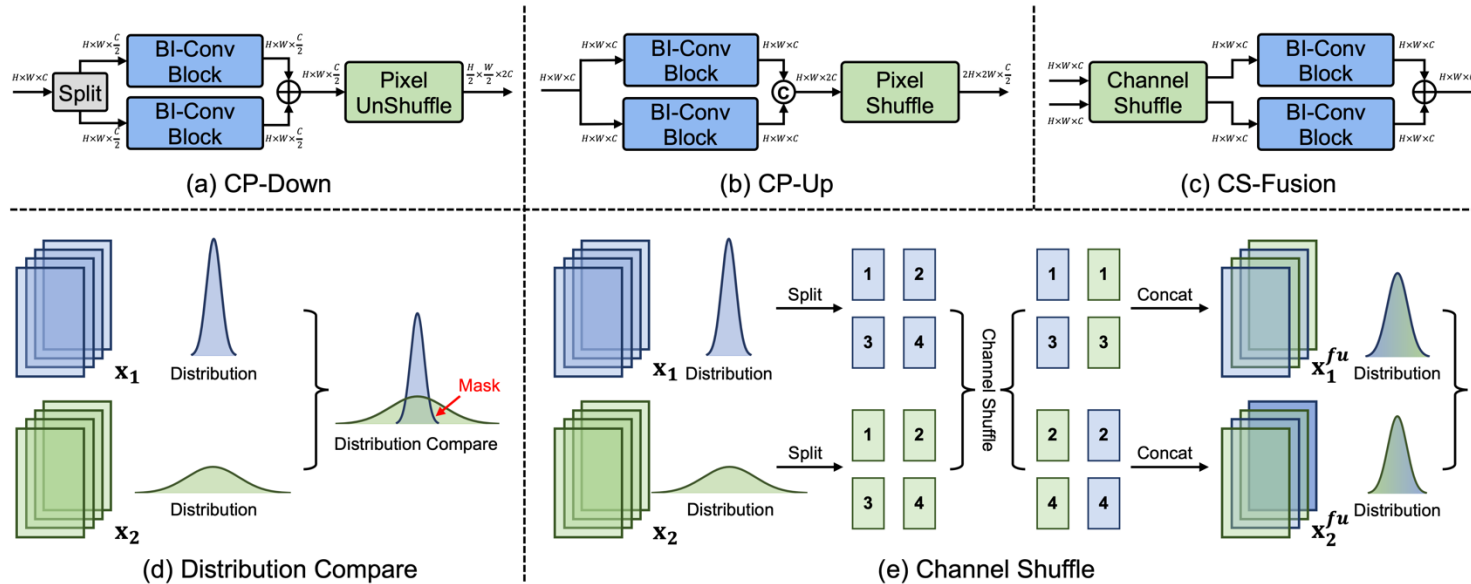
Challenge

- **Dimension Mismatch:** Resolution changes disrupt full-precision propagation.
- **Fusion Difficulty:** Activation range differences hinder feature fusion.

Structure

- **CP-Down/Up:** Enables consistent reshaping, preserving full precision.
- **CS-Fusion:** Uses channel shuffle for balanced distribution, enhancing fusion.

Method

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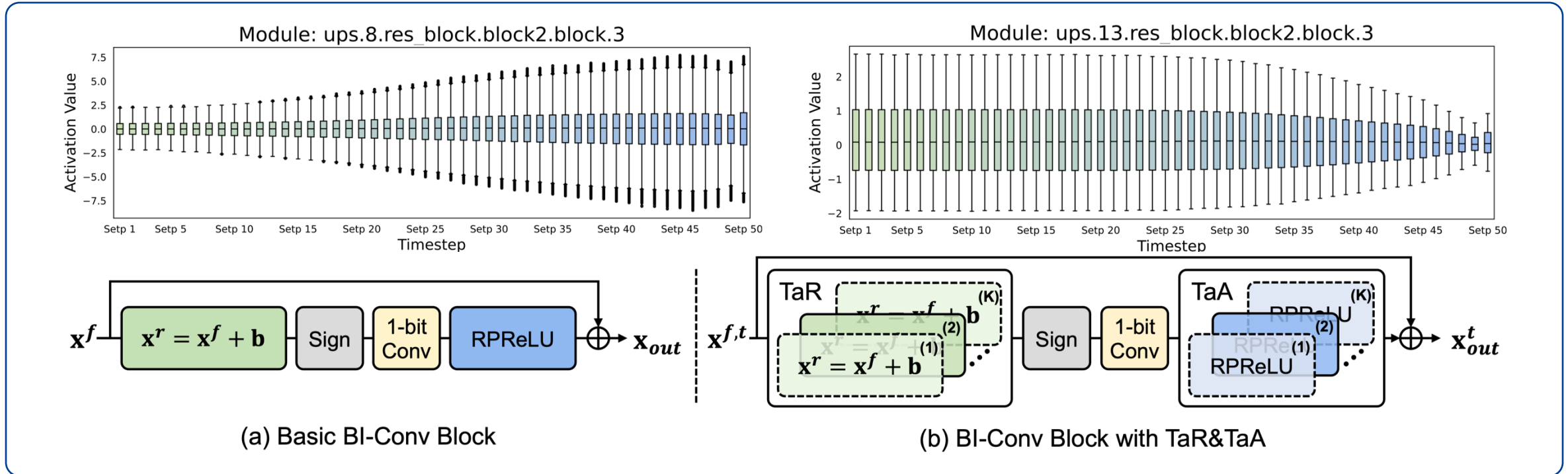
Activation

- **Analysis:** Activation distributions differ in multi-step iterations.
- **TaR/TaA:** Adjusts activations across timesteps, enhancing the binarized module.

Experiments

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.
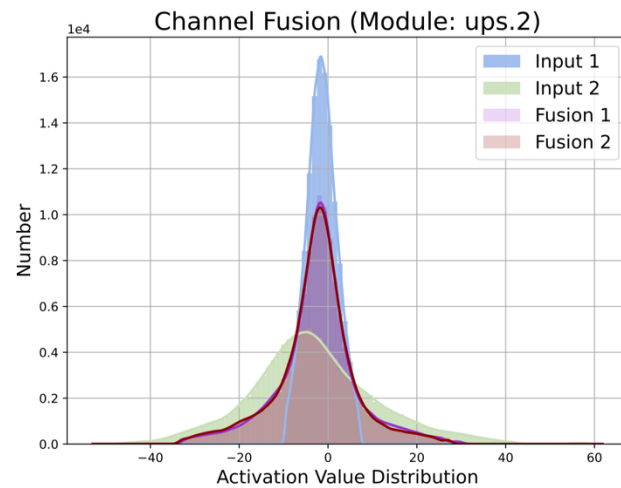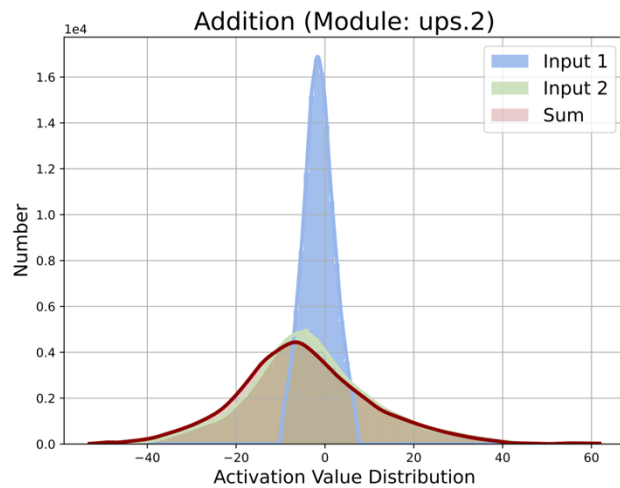
Ablation

- **Break Down:** Each module improves performance.

- **CS-Fusion:** Effectively fuses different features.

- **Visualization:** Distribution.

Experiments

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.

Ablation

- **Timestep-aware (TaA/TaR):** Effective improvements are achieved only when both TaR and TaA are used, with 5 biases and RPReLU providing notable gains.
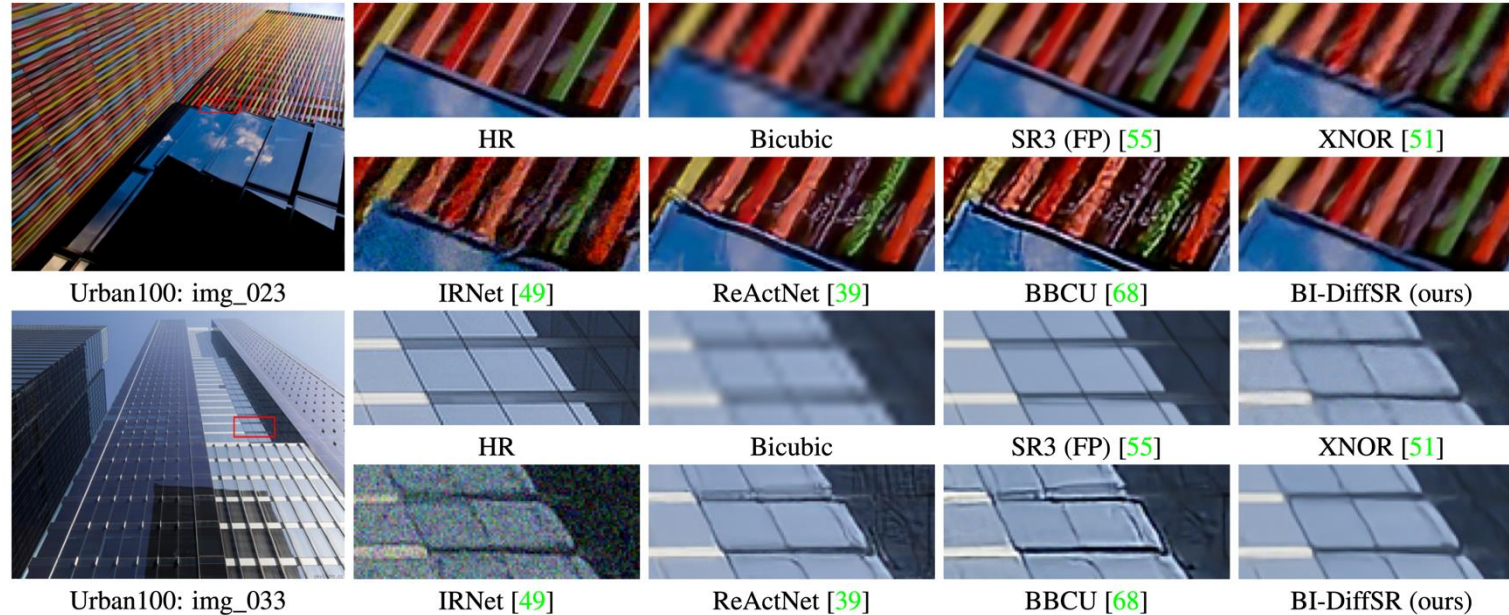
- **Visualization:** Five learnable biases in TaR.
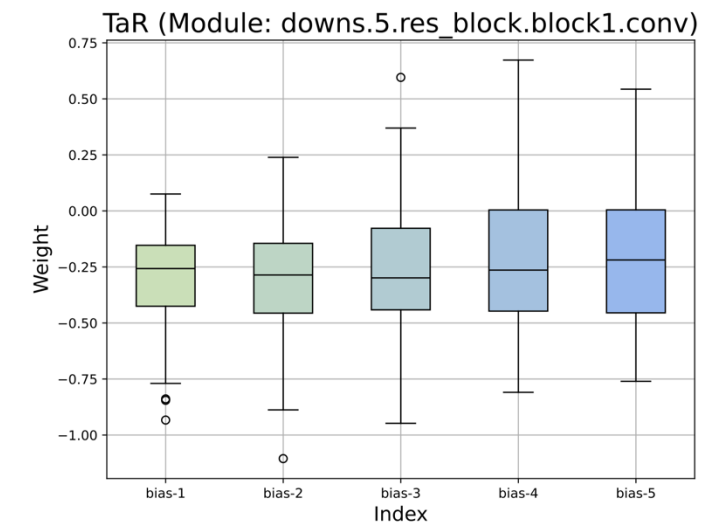
Method	Scale	Params (M)	Ops (G)	Set5			B100			Urban100			Manga109		
				PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Bicubic	×2	N/A	N/A	33.67	0.9303	0.1274	29.55	0.8431	0.2508	26.87	0.8403	0.2064	30.82	0.9349	0.1025
SR3 [55]	×2	55.41	176.41	36.69	0.9513	0.0310	30.41	0.8683	0.0700	30.29	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

Quantitative

- **Best Performance:** Achieves the best results among recent binarization methods.
- **Perceptual Metrics:** Comparable to the full-precision SR3 model in LPIPS.

Urban100: img_023

Urban100: img_033

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

Method	Params (M)	OPs (G)	Simulated Time (s)
SR3 [55]	55.41	176.41	55.37
BI-DiffSR	4.58	36.67	13.00

Visual

- Our method restores clearer images with more texture details.

- The gap between BI-DiffSR and full-precision models is small.

Cost

- Runs much faster than full-precision methods.
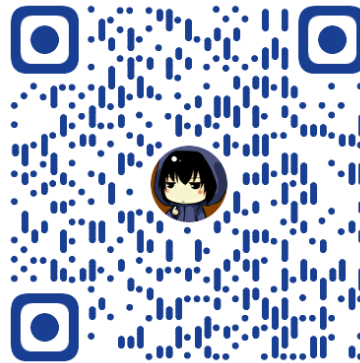
Conclusion

Contribution

We propose **BI-DiffSR**, a novel binarized DM for image SR.

- **Architecture:** Design modules for binarization: **CP-Down**, **CP-Up**, **CS-Fusion**.

- **Activation:** Introduce **TaR** and **TaA** to enhance binarized modules.

- **Performance:** Outperforms SOTA binarization methods.

Poster

- Time: Fri 13 Dec
  11 a.m. - 2 p.m. PST

Project

Thanks!