DM^{3}Net: Dual-Camera Super-Resolution via Domain Modulation and Multi-scale Matching Supplementary Materials

1. Experiments Settings Supplement

Datasets. We conduct our experiments on 3 publicly available real-world dual-camera super-resolution datasets: **DuSR-Real**, **RealMCVSR-Real**, and **CameraFusion-Real**. These datasets differ in terms of degradation levels, alignment quality, and resolution. DuSR-Real provides well-aligned image triplets captured simultaneously with iPhone 13 dual-lens cameras, covering diverse scenes at a resolution of 1792×896. RealMCVSR-Real features more challenging degradations such as motion blur, while CameraFusion-Real offers the highest resolution (3584×2560) but includes minor misalignment. We adopt these datasets to evaluate the generalization and robustness of our model under various real-world scenarios.

Implementation Details. During training stage, we apply random flipping and 90° rotations for data augmentation. The batch size is fixed to 4, and the LR patch size is set to 128×128. The model is trained for 400 epochs using the Adam optimizer [1] with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The initial learning rate is 1×10^{-4} and is decayed to 5×10^{-5} after 250,000 iterations. All experiments are implemented using PyTorch [2] and conducted on a single NVIDIA A100 40G PCIE GPU.

2. Lightweight Version of DM^{3}Net.

To enable on-device deployment, we also design a compact variant of DM^{3}Net. Concretely, we reduce the number of MLP layers and the embedding dimensionality in the Global Prior Extractor, decrease the convolutional depth of both the LR and Ref Encoders, and prune several cascaded residual blocks across the network. With these modifications, the total parameter count drops from 18.64 M to just 7.53 M.

3. Quantitative Comparison Supplement

In previous studies [3–5], it has been commonly observed that training with only the L_1 or Charbonnier loss yields superior metrics but worse visual quality. Accordingly, we also re-trained the comparing Dual-camera SR models using only the L_1 or Charbonnier loss, denoted as -ℓ versions. It is important to note that, since our method include an additional L_{domain} to supervise the domain-aware embeddings, we use both the Charbonnier loss and domain-aware loss to train DM^{3}Net-ℓ model. Tables 1, 2 and 3 lists the quantitative results of the -ℓ version models on DuSR-Real, RealMCV-Real, and CameraFusion-Real datasets. It can be observed that our DM^{3}Net-ℓ consistently achieves the best performance across most evaluation metrics.

4. More Ablation Study

4.1. Ablation on Key Pruning

We also investigate the impact of the **sampling interval** and **threshold** in Key Pruning on both the performance and efficiency of the model. Figure 1 shows the curves of PSNR and inference time with varying thresholds under different sampling intervals on the DuSR-Real dataset. It can be observed that a higher threshold generally leads to better PSNR but also results in increased inference time. This trend is more evident when the sampling interval is smaller. To balance the model efficiency and performance, we set the sampling interval to 16 and the threshold to 0.7.

4.2. Visual Comparison

Fig 2 presents the visual comparison of the ablation study on multi-scale matching. It is observed that adopting our multi-scaling matching outperforms using single-scale matching on 1/4, 1/2, or 1 scale. Fig 3 presents the visual comparison of the ablation study on domain modulation. Our method that uses both $\mathbf{z}$ and $\mathbf{z}_{\text{gt}}$ obtains the correct color while the other two models generate color distortion.

5. Complexity Analysis

We present the latency and number of parameters in Table 4. Our method is faster than SwinIR, TTSR, and MASA-SR in terms of latency. Latency indicates the time required to generate one HR result (1792×896) using one NVIDIA A100 GPU.

Table 1. Quantitative comparisons of -ℓ version models the on DuSR-Real Dataset. The best metrics are in bold.

2*Method	Full-Image			Center-Image	Corner-Image
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR/SSIM	PSNR/SSIM
DCSR-ℓ	26.25	0.8576	0.209	28.63 / 0.8934	25.71 / 0.8454
SelfDZSR-ℓ	25.71	0.8337	0.205	26.30 / 0.8368	25.58 / 0.8326
KeDuSR-ℓ	27.66	0.8890	0.177	**29.58** / 0.9303	27.24 / 0.8750
DM^{3}Net-ℓ	**27.79**	**0.8902**	**0.173**	29.50 / **0.9347**	**27.41 / 0.8751**
DM^{3}Net-s-ℓ	27.70	0.8873	0.182	29.58 / 0.9318	27.29 / 0.8721

Table 2. Quantitative comparisons of -ℓ version models on RealMCVSR-Real. The best metrics are in bold.

2*Method	Full-Image			Center-Image	Corner-Image
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR/SSIM	PSNR/SSIM
DCSR-ℓ	26.00	0.8018	0.312	27.38 / 0.8315	25.67 / 0.7917
SelfDZSR-ℓ	25.28	0.7800	0.279	25.33 / 0.7746	25.33 / 0.7818
KeDuSR-ℓ	27.05	0.8406	0.238	29.25 / 0.9191	26.56 / **0.8139**
DM^{3}Net-ℓ	**27.11**	**0.8415**	**0.236**	**29.27 / 0.9254**	**26.62** / 0.8130
DM^{3}Net-s-ℓ	27.02	0.8403	0.241	29.20 / 0.9241	26.54 / 0.8118

Table 3. Quantitative comparisons of -ℓ version models on CameraFusion-Real Dataset. The best metrics are in bold.

2*Method	Full-Image			Center-Image	Corner-Image
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR/SSIM	PSNR/SSIM
DCSR-ℓ	25.38	0.7977	0.346	26.34 / 0.8106	25.17 / 0.7934
SelfDZSR-ℓ	25.88	0.7852	0.284	26.91 / 0.7960	25.66 / 0.7816
KeDuSR-ℓ	27.53	0.8292	0.322	30.48 / 0.8656	26.93 / 0.8169
DM^{3}Net-ℓ	**27.93**	**0.8427**	**0.282**	**32.10 / 0.9180**	**27.16 / 0.8174**
DM^{3}Net-s-ℓ	27.73	0.8403	0.288	31.72 / 0.9116	26.99 / 0.8162

Table 4. Comparison of the model parameters and latency.

	SwinIR	Real-ESRGAN	TTSR	MASA-SR	DCSR	SelfDZSR	KeDuSR	DM^{3}Net	DM^{3}Net-s
Params (M)	11.75	16.70	6.25	4.02	3.19	0.52	5.63	18.64	7.53
Latency (s)	4.609	0.113	7.067	6.013	1.172	0.793	0.836	1.404	1.387

Table 5. Generalization Evaluation with the models trained on DuSR-Real dataset.

Method	**RealMCVSR-Real** PSNR↑ / SSIM↑	**CameraFusion-Real** PSNR↑ / SSIM↑
SwinIR	24.00 / 0.7738	25.01 / 0.7755
Real-ESRGAN	23.76 / 0.7680	23.94 / 0.7114
MASA-SR	25.18 / 0.7757	25.38 / 0.7724
TTSR	24.86 / 0.7796	24.50 / 0.7653
DCSR	24.96 / 0.7807	24.82 / 0.7530
SelfDZSR	24.73 / 0.7741	24.79 / 0.7253
KeDuSR	26.21 / 0.8189	26.78 / 0.7909
DM^{3}Net	**26.66 / 0.8284**	**27.26 / 0.8172**

6. Generalization Evaluation

We evaluate the generalization ability of different approaches on the RealMCVSR-Real and CameraFusion-Real datasets using models trained on DUSR-Real dataset. As shown in Table 5, our method outperforms the competing approaches, demonstrating the superior generalization capability of DM^{3}Net. This advantage can be attributed to our matching mechanism and the extraction of domain-aware embeddings, both of which are independent of the training dataset.

7. More Visual Comparisons

We present more visual comparative results in Figure 4 and Figure 5. Our method demonstrates superior performance in terms of structure details and color fidelity.

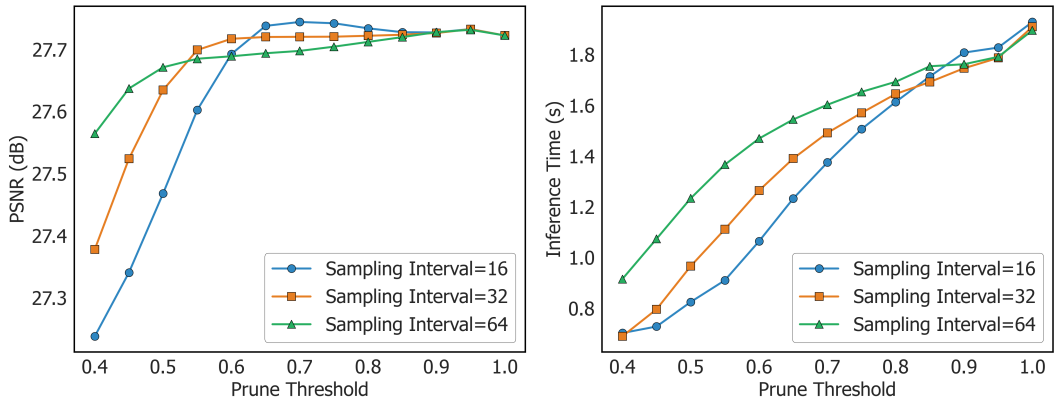

Figure 1. PSNR and inference time in relation to varying thresholds under different sampling intervals. Tests are conducted on DuSR-Real dataset on NVIDA A100 GPU.

Figure 2. Visual comparisons of ablation study on multi-scale matching. 1/4, 1/2, and 1 denote matching at respective scales.

8. Application and Limitations

The application of our DM^{3}Net is for computational photography in multi-camera systems such as smartphones, drones, and action cameras. A current limitation of our method lies in the relatively large model parameter size, and its inference speed is not yet real-time. In future work, we will plan to reduce the model parameters and increase the speed.

References

[1] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 1

[2] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, pages 8024–8035, 2019. 1

[3] Tengfei Wang, Jiaxin Xie, Wenxiu Sun, Qiong Yan, and Qifeng Chen. Dual-camera super-resolution with aligned attention modules. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2001–2010, 2021. 1

[4] Huanjing Yue, Zifan Cui, Kun Li, and Jingyu Yang. Kedusr: real-world dual-lens super-resolution via kernel-free matching. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 6881–6889, 2024.

[5] Zhilu Zhang, Ruohao Wang, Hongzhi Zhang, Yunjin Chen, and Wangmeng Zuo. Self-supervised learning for real-world super-resolution from dual zoomed observations. In *Proceedings of the European Conference on Computer Vision*, 2022. 1

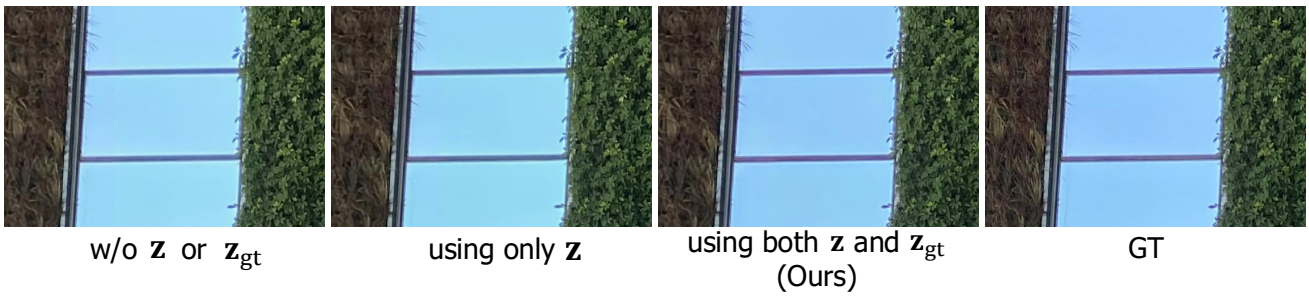

Figure 3. Visual comparisons of ablation study on domain modulation.

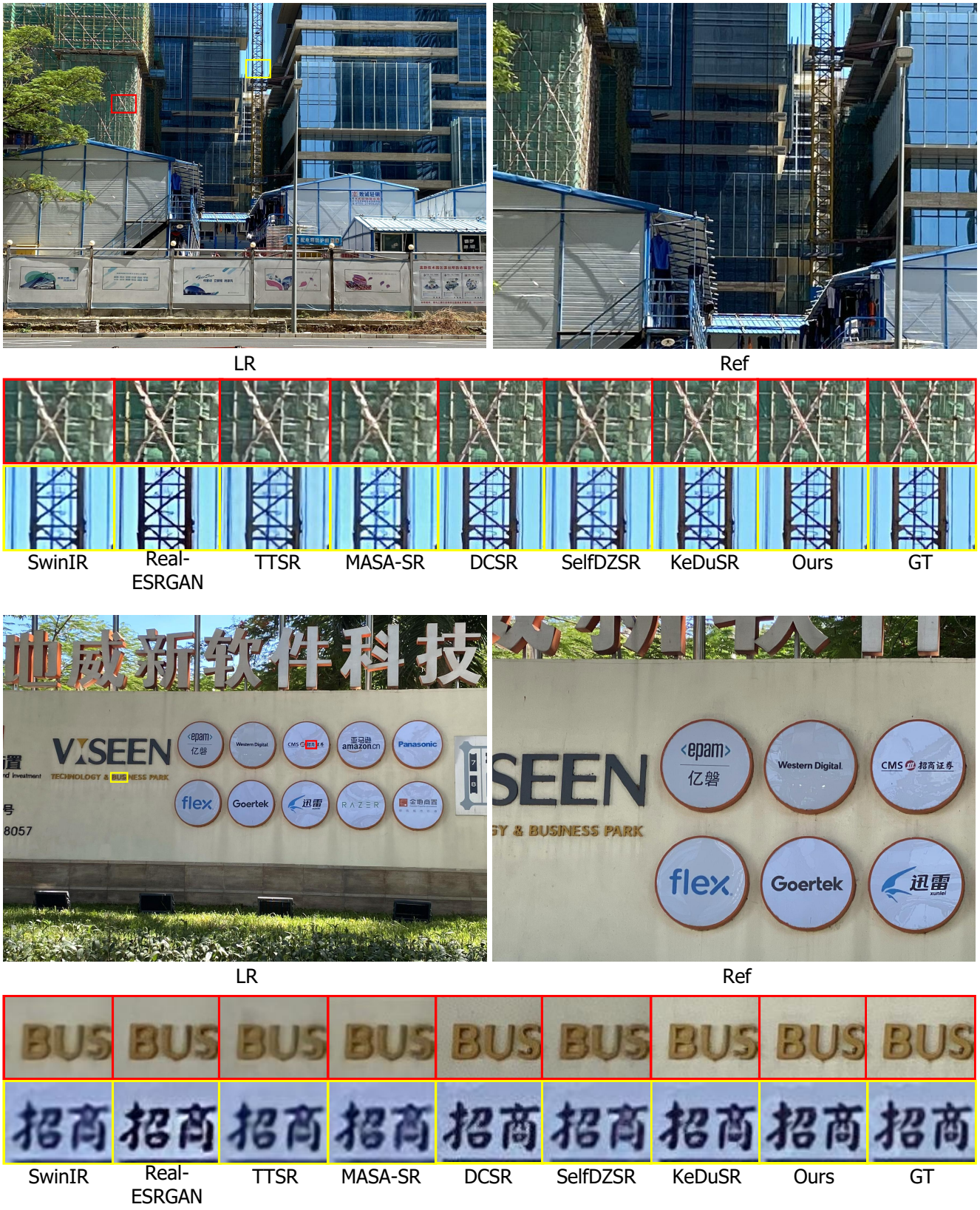

Figure 4. Visual comparisons on CameraFusion-Real.

Figure 5. Visual comparisons on RealMCVSR-Real.