Supplementary of

A Fast, Simple, and Flexible Scale Informative Feature Transform Module for Arbitrary Scale Image Super-Resolution

Aupendu Kar*
Dolby Laboratories, Inc
Bengaluru, India
mailtoaupendu@gmail.com

Prabir Kumar Biswas
Indian Institute of Technology Kharagpur
Kharagpur, India
pkb@ece.iitkgp.ac.in

1. More Details of Super-resolution Under Homographic Transformation

Homographic degradation is relevant in several real-world applications where images undergo geometric distortions, such as aerial or satellite imaging, surveillance footage, and handheld camera captures with lens distortion or perspective shifts. These distortions are often modeled using homographic transformations, and restoring such images is crucial for downstream tasks like object detection or scene understanding.

We use our proposed upscaling module to perform the homographic transformation on low-resolution image I_{LR} to recover it from geometrical distortion. M is the transformation matrix that transform I_{LR} to get back geometric distortion-free super-resolved image I_{SR}. We assume I_{LR} has gone through the distortions like rotation R, scaling S, sheering H, and projection P. M^{-1} is the degradation matrix which degrades high-resolution image I_{HR}. M^{-1} is composed as follows.

$$
\begin{aligned}
M^{-1} &= HRSP, \\
H &= \begin{pmatrix} 1 & h_x & 0 \\ h_y & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \\
R &= \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}, \\
S &= \begin{pmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{pmatrix}, \\
P &= \begin{pmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ p_x & p_y & 1 \end{pmatrix},
\end{aligned}
\tag{1}
$$

*Work conducted while Aupendu Kar was affiliated with the Indian Institute of Technology Kharagpur.

Baseline	Positional Info.	M-matrix Info.	Transform ($\mathcal{W}_a x$)	Transform ($\mathcal{W}_a x + \mathcal{W}_b$)	mPSNR in dB
Baseline-1	✗	✗	✗	✗	30.15
Baseline-2	✗	✗	✗	✓	31.09
Baseline-3	✗	✓	✗	✓	31.13
Baseline-4	✓	✓	✗	✓	31.18
Baseline-5	✓	✗	✓	✗	31.17
Baseline-6	✓	✗	✗	✓	**31.48**

Table 1. Ablation study on proposed upscaling module's components during super-resolution under homographic transformation. We use the DIV2K Validation dataset for this experiment. 100 validation images are center cropped into 384×384 and 100 randomly selected M^{-1} are used to generate the low-resolution images. Masked PSNR (mPSNR) are used for performance evaluation [4].

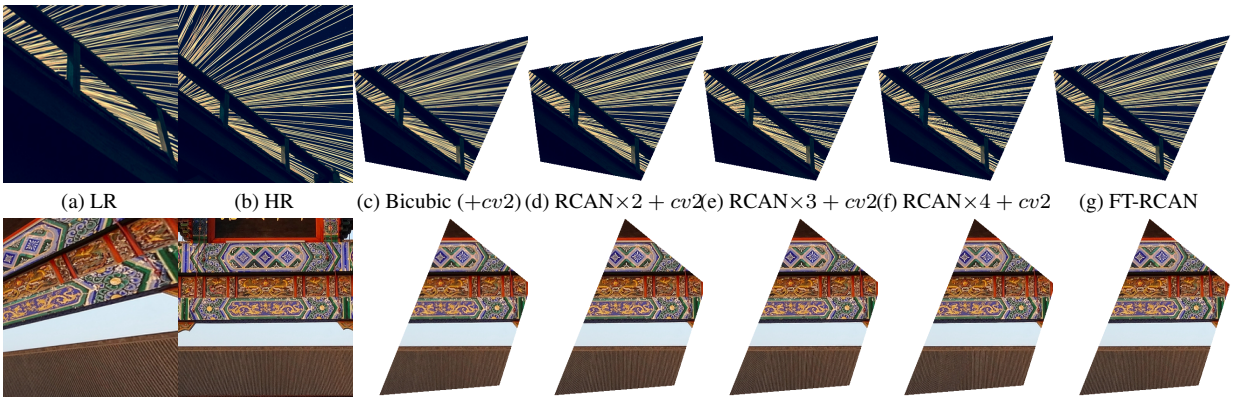
h_x and h_y are sheering along x and y direction. θ is the rotation angle. s_x and s_x scaling factor along x and y direction. t_x, t_y, p_x, p_y are the projection parameters.

2. Qualitative Performance on Homographic Transformation

Fig. 1 shows the qualitative performance comparison of our proposed upscaling module as compared to other baseline techniques. We witness that our proposed approach performs better than different baselines in preserving the textures in the image.

3. Ablation Study of Homographic Transformation Model

Table 1 shows six different ablation studies on each component of the upscaling module while we use it for super-resolution under homographic transformation. Baseline-1 is the nearest neighbor interpolation-based homographic transformation using the M matrix. Baseline-2 adopts the proposed feature transform ($\mathcal{W}_a x + \mathcal{W}_b$), and it shows around 1 dB PSNR improvement, which shows the effec-



(a) LR (b) HR (c) Bicubic (+$cv2$) (d) RCAN×2 + $cv2$ (e) RCAN×3 + $cv2$ (f) RCAN×4 + $cv2$ (g) FT-RCAN

(h) LR (i) HR (j) Bicubic (k) RDN×2 + $cv2$ (l) RDN×3 + $cv2$ (m) RDN×4 + $cv2$ (n) FT-RDN

Figure 1. Qualitative performance analysis of super-resolution under homographic transformation.

tiveness of the feature transform. Baseline-3 utilizes the information of M during transformation parameter estimation, $\mathcal{W}_a$, and $\mathcal{W}_b$. However, it shows a negligible improvement. Baseline-4 incorporates the positional-shift information along with M matrix information. It also provides a marginal improvement. However, if we only use the information of positional-shift information, as shown in Baseline-6, it significantly improves the performance. Therefore, we only use positional-shift information in our approach. Baseline-5 is the same as Baseline-6 except for the transformation equation. In Baseline-5, we only multiply the features with $\mathcal{W}_a$ and do not perform the shift using $\mathcal{W}_b$. The table shows that Baseline-5 performs poorly as compared to Baseline-6. This is because the shift using $\mathcal{W}_b$ helps to boost the high-frequency features, which are generally degraded due to the homographic transformation operation.

4. More Results on Integer Scale Super-Resolution

Table 2 shows the performance of our model for integer scale super-resolution for four different test sets. The PSNR is calculated on the Y channel of the YCbCr color space.

Dataset	Method	In-distribution			Out-of-distribution	
		×2	×3	×4	×6	×8
Set5	RDN [6]	38.24	34.71	32.47	-	-
	Meta-RDN [2]	38.22	34.63	32.38	29.04	26.96
	LIIF-RDN [1]	38.17	34.68	32.50	29.15	27.14
	LTE-RDN [3]	38.23	34.72	32.61	29.32	27.26
	OPE-RDN [5]	37.60	34.59	32.47	29.17	27.22
	FT-RDN (Ours)	38.17	34.73	32.53	29.21	27.18
Set14	RDN [6]	34.01	30.57	28.81	-	-
	Meta-RDN [2]	33.98	30.54	28.78	26.51	24.97
	LIIF-RDN [1]	33.97	30.53	28.80	26.64	25.15
	LTE-RDN [3]	34.09	30.58	28.88	26.71	25.16
	OPE-RDN [5]	33.39	30.49	28.80	26.65	25.17
	FT-RDN (Ours)	34.01	30.60	28.86	26.71	25.22
B100	RDN [6]	32.34	29.26	27.72	-	-
	Meta-RDN [2]	32.33	29.26	27.71	25.90	24.83
	LIIF-RDN [1]	32.32	29.26	27.74	25.98	24.91
	LTE-RDN [3]	32.36	29.30	27.77	26.01	24.95
	OPE-RDN [5]	32.05	29.19	27.72	25.96	24.91
	FT-RDN (Ours)	32.34	29.29	27.78	26.02	24.95
Urban100	RDN [6]	32.89	28.80	26.61	-	-
	Meta-RDN [2]	32.92	28.82	26.55	23.99	22.59
	LIIF-RDN [1]	32.87	28.82	26.68	24.20	22.79
	LTE-RDN [3]	33.04	28.97	26.81	24.28	22.88
	OPE-RDN [5]	31.78	28.63	26.53	24.06	22.70
	FT-RDN (Ours)	32.96	28.92	26.77	24.27	22.88

Table 2. Quantitative comparison on benchmark datasets. RDN trains different models for different scales. MetaSR, LIIF, LTE, OPE, and our FT use one model for all scales and are trained with continuous random scales uniformly sampled in ×1–×4.

References

[1] Yinbo Chen, Sifei Liu, and Xiaolong Wang. Learning continuous image representation with local implicit image function. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8628–8638, 2021. 2

[2] Xuecai Hu, Haoyuan Mu, Xiangyu Zhang, Zilei Wang, Tieniu Tan, and Jian Sun. Meta-sr: A magnification-arbitrary network for super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1575–1584, 2019. 2

[3] Jaewon Lee and Kyong Hwan Jin. Local texture estimator for implicit representation function. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1929–1938, 2022. 2

[4] Sanghyun Son and Kyoung Mu Lee. Srwarp: Generalized image super-resolution under arbitrary transformation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7782–7791, 2021. 1

[5] Gaochao Song, Qian Sun, Luo Zhang, Ran Su, Jianfeng Shi, and Ying He. Ope-sr: Orthogonal position encoding for designing a parameter-free upsampling module in arbitrary-scale image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10009–10020, 2023. 2

[6] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2472–2481, 2018. 2