Reference-based Image SR with Deformable Attention Transformer

Jiezhang Cao, Jingyun Liang, Kai Zhang, Yawei Li, Yulun Zhang, Wenguan Wang, Luc Van Gool

Computer Vision Lab, ETH Zürich, Switzerland, KU Leuven, Belgium

ECCV TEL AVIV 2022

Background

Reference-based image super-resolution (RefSR) aims to exploit auxiliary reference (Ref) images to super-resolve low-resolution (LR) images.

Two critical challenges:

1. It is difficult to match the correspondence between LR and Ref images when they are significantly different;
2. How to transfer the relevant texture from Ref images to compensate the details for LR images is very challenging.

Methods

Reference-based Deformable Attention

Extract multi-scale features of LR and Ref images:

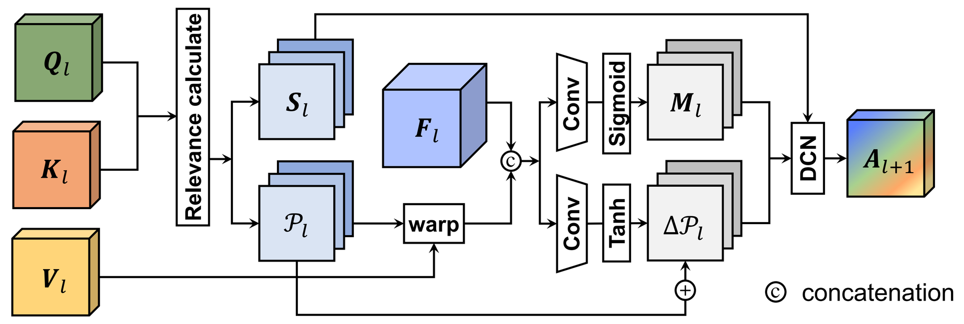
$$Q_l = E_l^q(X_{LR\uparrow}), K_l = E_l^k(X_{Ref}), V_l = E_l^v(X_{Ref}),$$

Correspondence matching. The top K relevant positions in K' can be calculated by

$$P_i = \left[\sigma\left(Q_l'^{\top} K_l'\right)\right]_i = \mathrm{TopK}_j\left(\widetilde{q}_i \cdot \widetilde{k}_j\right)$$

Similarity-aware texture transfer. We calculate a feature at the position p using modified DCN, *i.e.*,

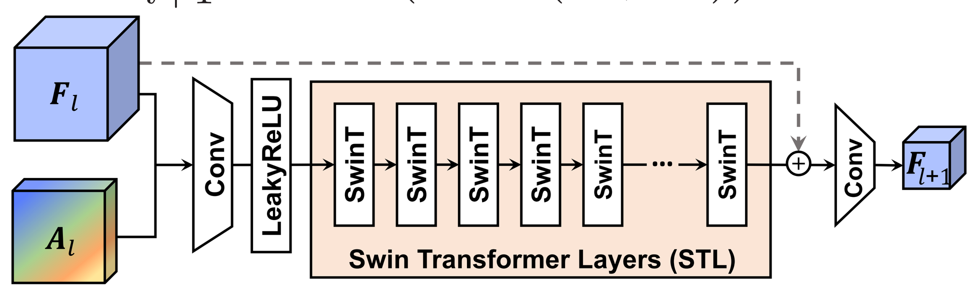
$$A_l(p_i) = \sum_k s_i^k \sum_j w_j V_l(p_i + \Delta p_i^k + p_j + \Delta p_j) m_j$$



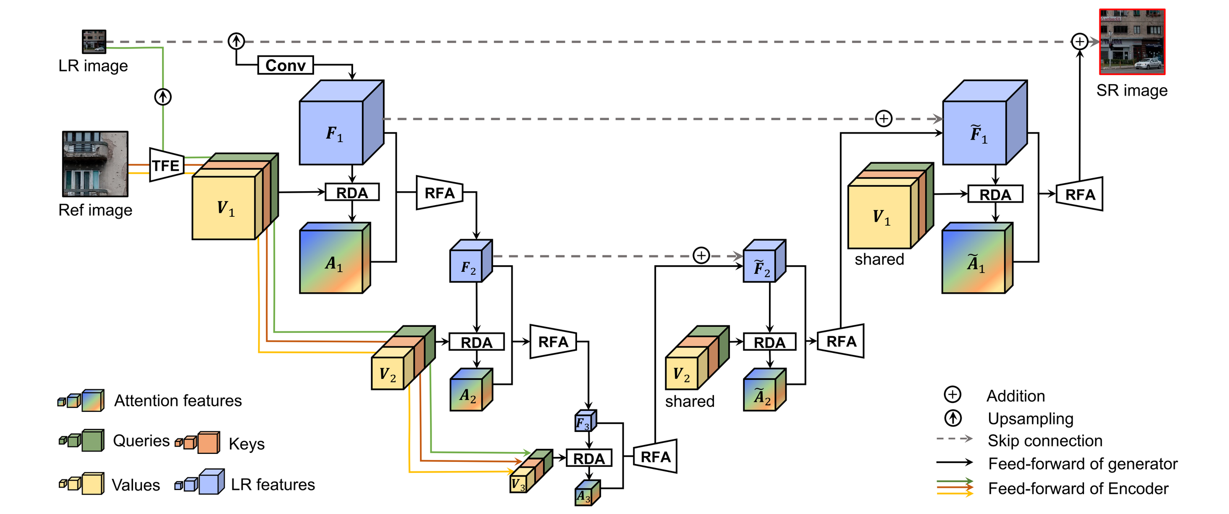
Residual Feature Aggregation

We use Swin Transformer to extract deeper features of the LR and transferred features,

$$F'_{l+1} = \mathrm{STL}(\mathrm{Conv}(F_l, A_l)) + F_l$$



Architecture



At each scale, our model consists of texture feature encoders, a reference-based deformable attention and a residual feature aggregation.

Qualitative Results

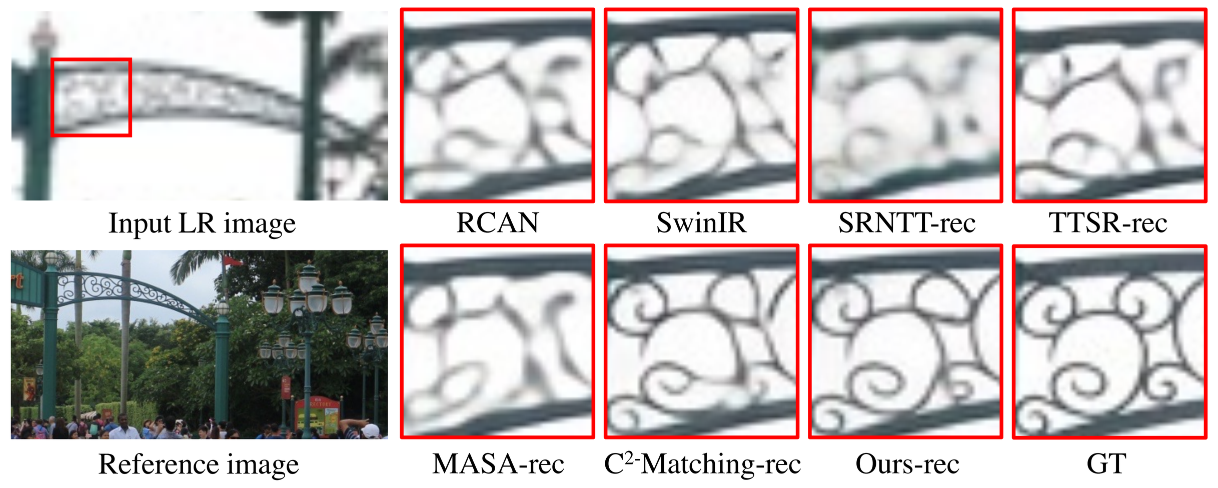


Figure 1: Qualitative comparisons of SISR and RefSR models trained with the reconstruction loss.

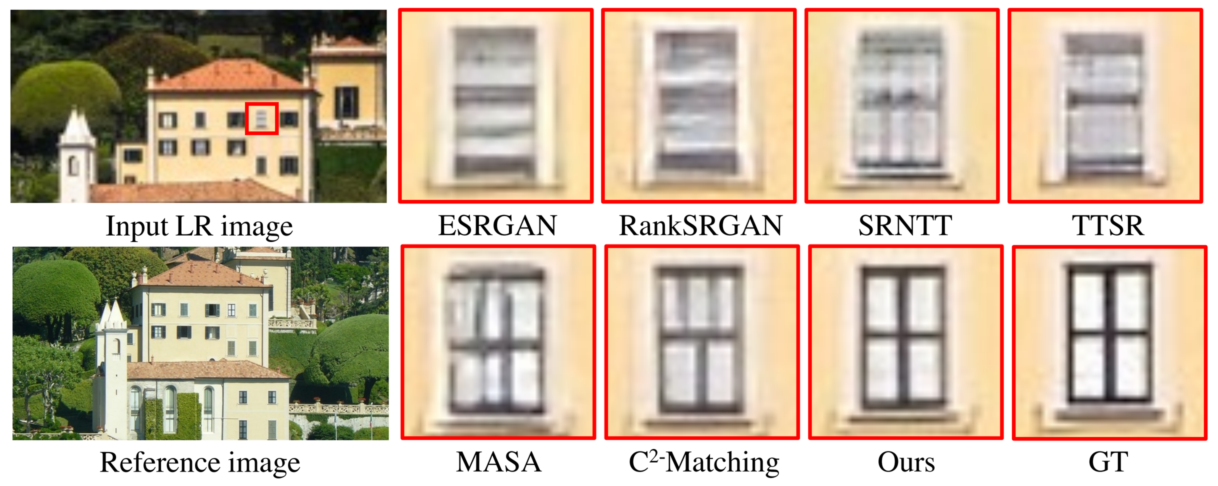


Figure 2: Qualitative comparisons of SISR and RefSR models trained with all loss.

- Our model achieves the best performance on visual quality when trained with the reconstruction loss and all loss.
- Our method is able to transfer more accurate textures from the Ref images to generate SR images with higher quality.

Quantitative Results

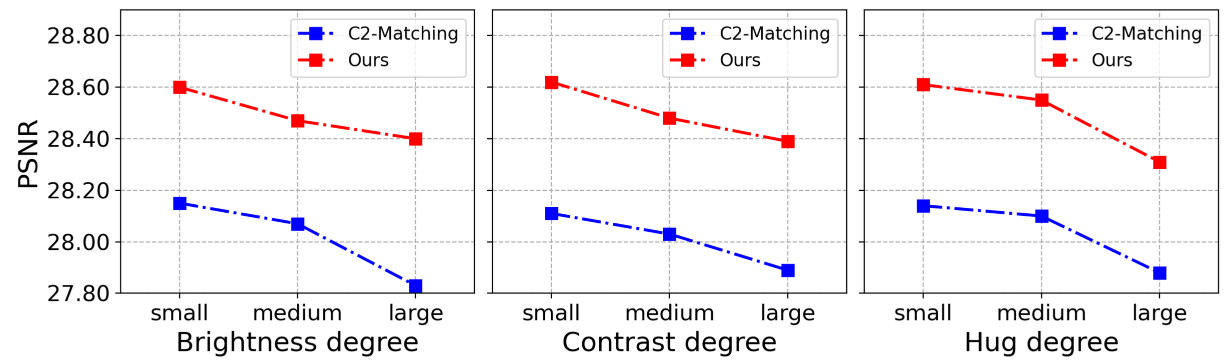
Table 1: Quantitative comparisons (PSNR and SSIM) of SR models trained with only reconstruction loss.

SR paradigms	Methods	CUFED5		Urban100		Manga109		Sun80		WR-SR	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
SISR	SRCNN	25.33	0.745	24.41	0.738	27.12	0.850	28.26	0.781	27.27	0.767
	EDSR	25.93	0.777	25.51	0.783	28.93	0.891	28.52	0.792	28.07	0.793
	ENet	24.24	0.695	23.63	0.711	25.25	0.802	26.24	0.702	25.47	0.699
	RCAN	26.06	0.769	25.42	0.768	29.38	0.895	29.86	0.810	28.25	0.799
	SwinIR	26.62	0.790	26.26	0.797	30.05	0.910	30.11	0.817	28.06	0.797
RefSR	CrossNet	25.48	0.764	25.11	0.764	23.36	0.741	28.52	0.793	-	-
	SRNTT-rec	26.24	0.784	25.50	0.783	28.95	0.885	28.54	0.793	27.59	0.780
	TTSR-rec	27.09	0.804	25.87	0.784	30.09	0.907	30.02	0.814	27.97	0.792
	SSEN-rec	26.78	0.791	-	-	-	-	-	-	-	-
	E2ENT^2-rec	24.24	0.724	-	-	-	-	28.50	0.789	-	-
	MASA-rec	27.54	0.814	26.09	0.786	30.24	0.909	30.15	0.815	28.19	0.796
	C^2-Matching-rec	28.24	0.841	26.03	0.785	30.47	0.911	30.18	0.817	28.32	0.801
	DATSR-rec (Ours)	**28.72**	**0.856**	**26.52**	**0.798**	**30.49**	**0.912**	**30.20**	**0.818**	**28.34**	**0.805**

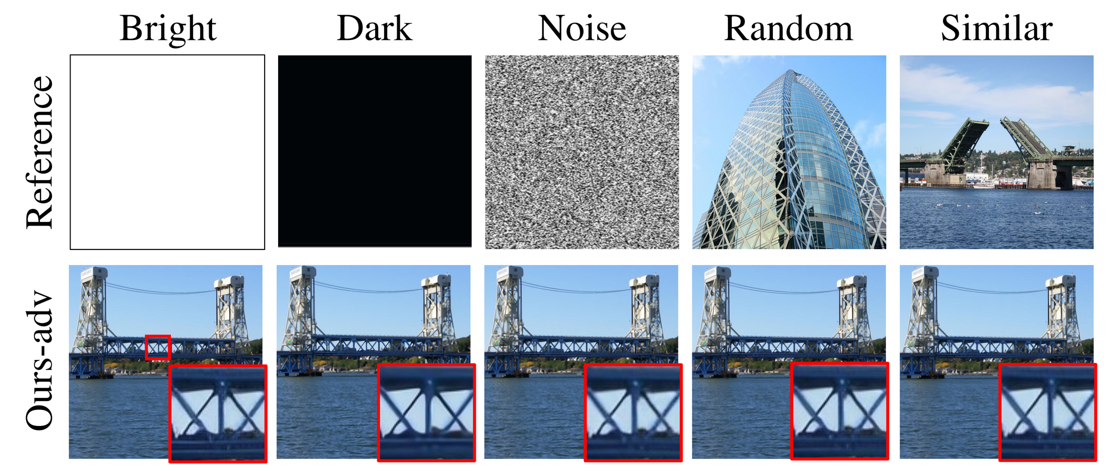
Table 2: Quantitative comparisons (PSNR and SSIM) of SR models trained with all losses.

SR paradigms	Methods	CUFED5		Urban100		Manga109		Sun80		WR-SR	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
SISR	SRGAN	24.40	0.702	24.07	0.729	25.12	0.802	26.76	0.725	26.21	0.728
	ESRGAN	21.90	0.633	20.91	0.620	23.53	0.797	24.18	0.651	26.07	0.726
	RankSRGAN	22.31	0.635	21.47	0.624	25.04	0.803	25.60	0.667	26.15	0.719
RefSR	SRNTT	25.61	0.764	25.09	0.774	27.54	0.862	27.59	0.756	26.53	0.745
	TTSR	25.53	0.765	24.62	0.747	28.70	0.886	28.59	0.774	26.83	0.762
	SSEN	25.35	0.742	-	-	-	-	-	-	-	-
	E2ENT^2	24.01	0.705	-	-	-	-	28.13	0.765	-	-
	MASA	24.92	0.729	23.78	0.712	27.26	0.847	27.12	0.708	25.74	0.717
	C^2-Matching	27.16	0.805	25.52	0.764	29.73	0.893	29.75	0.799	27.80	0.780
	DATSR (Ours)	**27.95**	**0.835**	**25.92**	**0.775**	**29.75**	**0.893**	**29.77**	**0.800**	**27.87**	**0.787**

Conclusion



- Randomly change the brightness, contrast and hue into small, medium and large
- Our model is more robust than C2-Matching under different image transformations.



- Our method has robust performance and high visual quality even if the Ref images have no useful texture information.

References

[1] Yuming Jiang and et al. Robust reference-based super-resolution via c2-matching. In *CVPR*, 2021.

[2] Jingyun Liang, Jiezhang Cao, and et al. Swinir: Image restoration using swin transformer. In *ICCVW*, 2021.

Conclusions

We propose a novel reference-based image super-resolution with deformable attention Transformer, called DATSR. DATSR achieves SOTA performance as it is robust to different brightness, contrast, and color, and still shows good robustness even in some extreme cases. Moreover, DATSR trained with a single Ref image performs better than existing Multi-RefSR methods trained with multiple Ref images.

Contact Information

Web https://www.jiezhangcao.com

Code https://github.com/caojiezhang/DATSR

Email jiezhang.cao@vision.ee.ethz.ch

Phone +86 18814099131