Training Generative Image Super-Resolution Models by Wavelet-Domain Losses Enables Better Control of Artifacts


Cansu Korkmaz

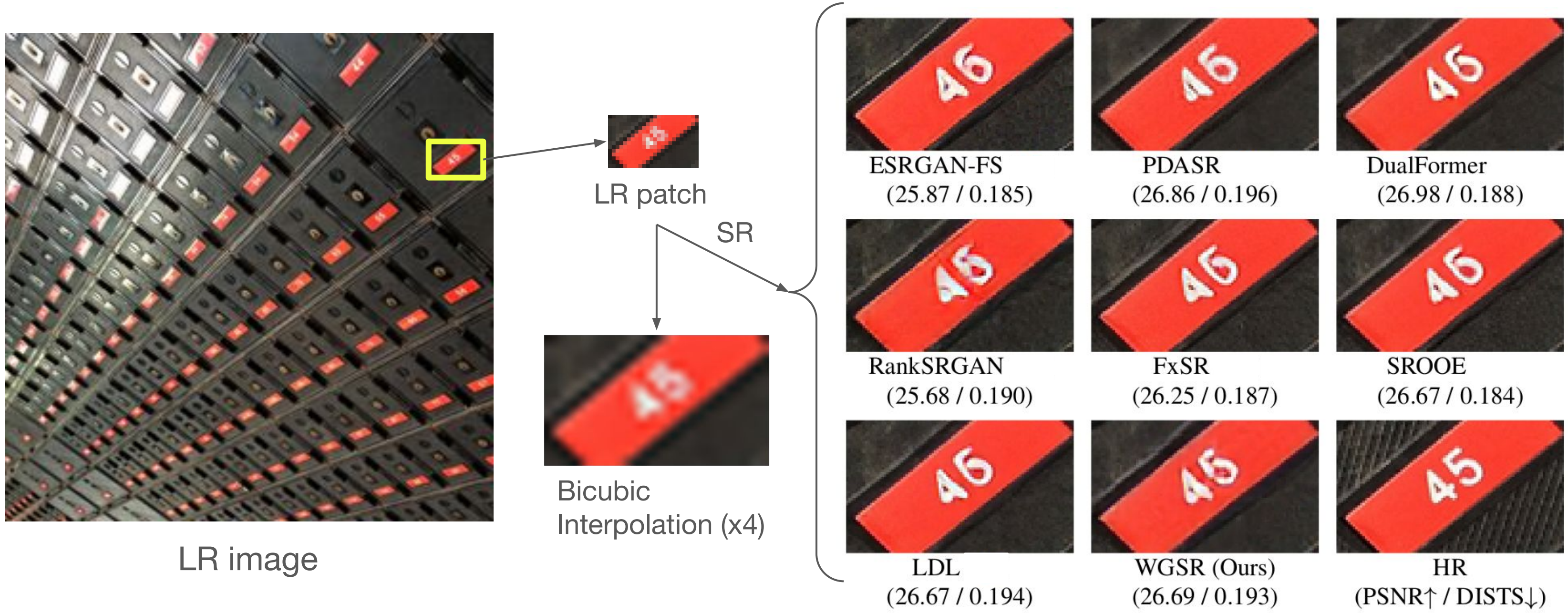

A. Murat Tekalp


Zafer Dogan

College of Engineering and KUIS AI Center, Koc University

Project Page : https://github.com/mandalinadagi/WGSR/

Problem Statement



while trying to recover high-frequency image details, generative methods hallucinate → **causes artifacts**

optimizing for PSNR, DISTS and other **quantitative scores alone does not prevent such hallucination**

Fundamental question: Can a model learn to distinguish genuine image details from artifacts?

Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In IEEE Conf. on Comp. Vision and Patt. Recog. (CVPR), pages 5197–5206, 2015.

KOÇ UNIVERSITY

Motivation & Contributions

We propose a novel **GAN-SR framework** that **uses wavelet-domain losses** to suppress hallucinations and artifacts for a better perception-distortion (PD) trade-off.

without wavelet loss

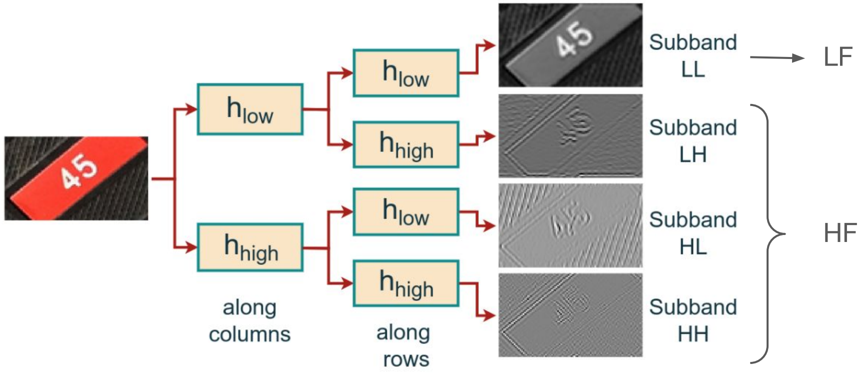
with wavelet loss

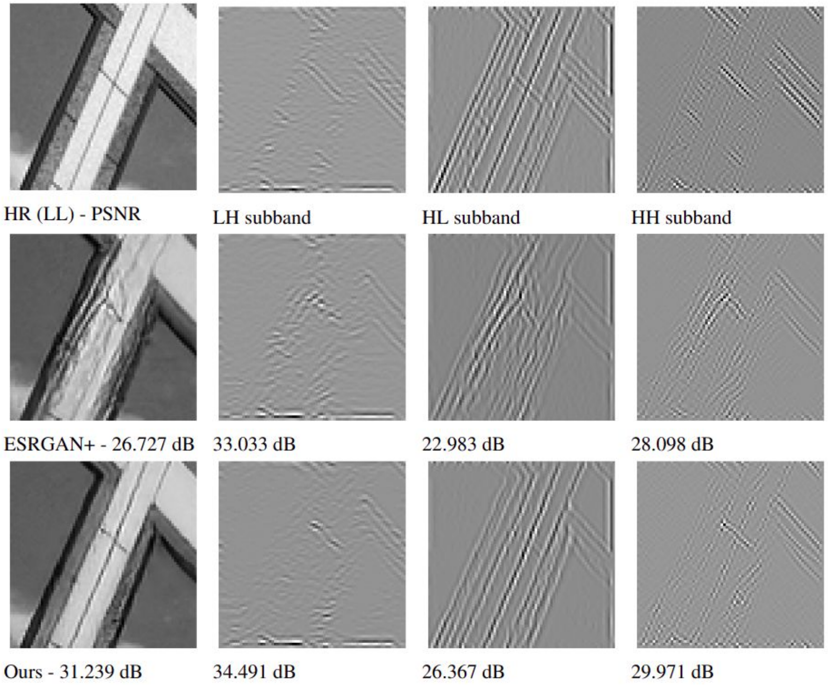
Our main contributions:

1. **Wavelet-domain fidelity loss** sensitive to scale and orientation of local structures
2. **Wavelet-domain discriminator** for adversarial training to control HF artifacts
3. Wavelet-loss guided training with DISTS perceptual loss (instead of LPIPS) significantly improves fidelity

https://github.com/mandalinadagi/WGSR/

Rationale for using Wavelet-Domain Losses





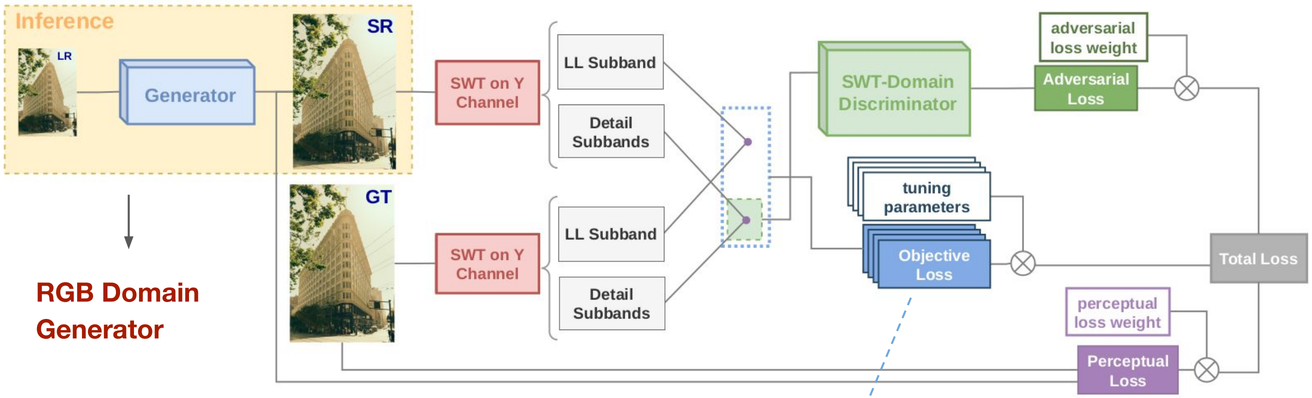
Stationary Wavelet Transform (SWT):
translation-invariant image decomposition → low-frequency (LL) and several high-frequency (LH, HL, HH) subbands

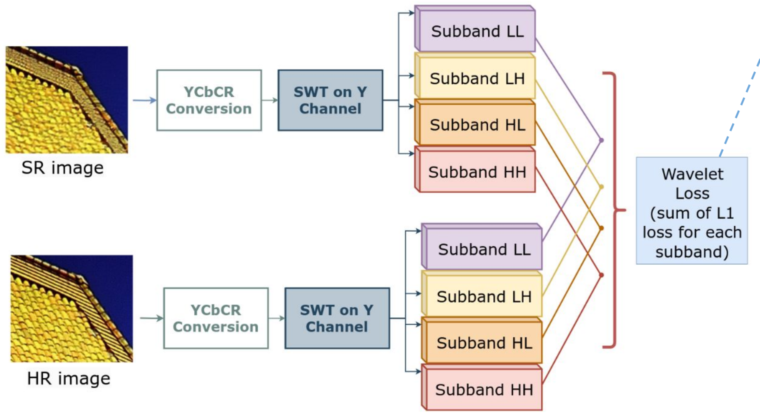
LL subband has significant effect on fidelity of SR results

HF contents aligned with LL subband need to be reconstructed to achieve photorealistic images

Bjorn Jawerth and Wim Sweldens. An overview of wavelet based multiresolution analyses. SIAM Review, 36(3):377–412, 1994
N. C. Rakotonirina and A. Rasoanaivo. Esrgan+: Further improving enhanced super-resolution generative adversarial network. In IEEE Int. Conf. on Acoust., Speech and Signal Processing (ICASSP), pages 3637–3641, 2020

WGSR: Wavelet-Loss Guided SR Framework





SWT Domain Discriminator:

Discriminator tasked to learn how "real" (i.e. aligned with LL) are the generated HF details

HF SWT subbands are sparse → simplifies the discriminator's task

$$L_G = L_{SWT} + \lambda_{adv} \cdot L_{adv,G} + \lambda_{perc} \cdot L_{perc}$$

$$L_{SWT} = \mathbb{E}\Big[\sum_j \lambda_j \big\| SWT(G(x))_j - SWT(y)_j \big\|_1\Big]$$

$$L_{adv,G} = -\mathbb{E}\big[log(1 - D(SWT(y)_*))\big] - \mathbb{E}\big[log D(SWT(G(x))_*)\big]$$

$$L_D = -\mathbb{E}\big[log D(SWT(y)_*)\big] - \mathbb{E}\big[log(1 - D(SWT(G(x))_*))\big]$$

KOÇ UNIVERSITY

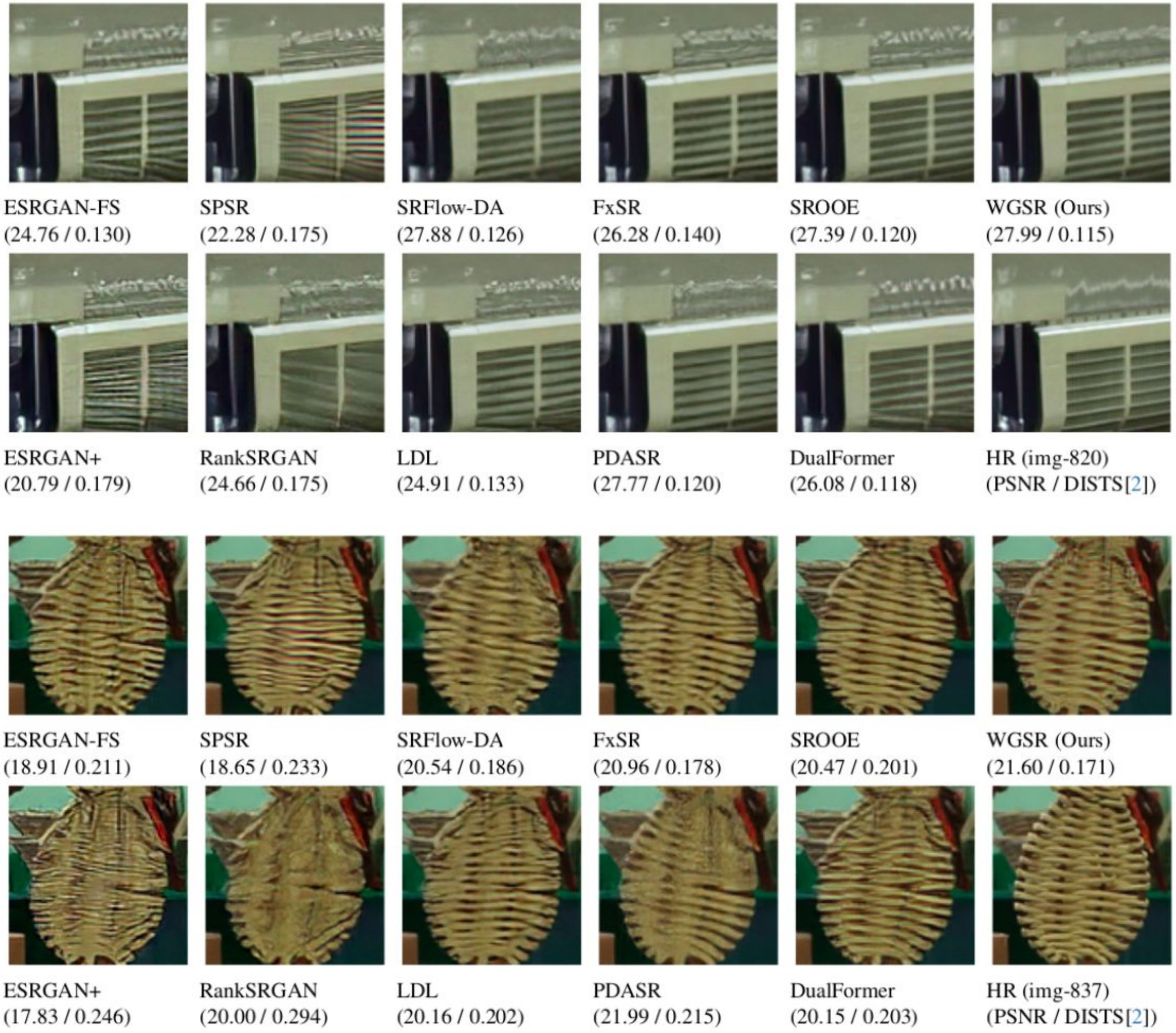
Quantitative Comparison

WGSR improves perceptual quality and reconstruction accuracy, yields best perceptual scores for NIQE, NRQM, and PI

Benchmark	Metric	ESRGAN-FS [12]	ESRGAN+ [50]	SPSR [40]	RankSRGAN [67]	SRFlow-DA [23]	LDL [31]	FxSR [47]	PDASR [71]	DualFormer [38]	SROOE [48]	WGSR (1-lvl)	WGSR (2-lvl)
Dataset		DF2K	DIV2K	DIV2K	DIV2K	DF2K	DIV2K	DIV2K	DIV2K	DIV2K	DF2K	DIV2K	DIV2K
BSD100	PSNR ↑	25.389	24.653	25.546	25.043	26.335	26.142	26.179	**26.879**	26.527	26.364	26.471	26.372
	SSIM ↑	0.658	0.614	0.659	0.639	0.684	0.682	0.685	**0.703**	0.691	0.693	0.696	0.684
	LPIPS ↓	0.166	0.211	0.161	0.183	0.191	0.163	0.157	0.187	0.158	**0.153**	0.187	0.174
	LPIPS-VGG ↓	0.269	0.313	0.263	0.285	0.286	0.244	0.253	0.272	0.242	**0.241**	0.283	0.282
	DISTS ↓	0.119	0.151	0.118	0.129	0.145	0.118	0.118	0.136	0.119	**0.116**	0.137	0.132
	NIQE ↓	3.386	3.675	3.261	**2.903**	3.603	3.383	3.386	3.902	3.957	3.684	3.428	3.243
	NRQM ↑	8.706	8.702	8.703	8.791	8.561	8.623	8.680	8.608	8.617	8.644	8.792	**8.793**
	PI ↓	2.402	2.531	2.335	2.086	2.631	2.473	2.422	2.779	2.796	2.576	**2.053**	2.065
	LR-PSNR ↑	39.910	41.530	40.990	37.510	**49.920**	43.690	49.260	49.830	42.306	49.610	49.046	48.915
Urban100	PSNR ↑	24.556	23.235	24.795	24.121	25.632	25.491	25.668	**26.279**	25.686	25.939	25.779	25.606
	SSIM ↑	0.743	0.707	0.747	0.719	0.763	0.767	0.772	**0.785**	0.773	0.779	0.781	0.777
	LPIPS ↓	0.124	0.143	0.119	0.143	0.129	0.110	**0.109**	0.123	0.115	0.108	0.135	0.135
	LPIPS-VGG ↓	0.222	0.248	0.216	0.249	0.241	**0.197**	0.204	0.223	0.200	0.199	0.243	0.243
	DISTS ↓	0.090	0.104	0.085	0.106	0.115	**0.082**	0.087	0.102	0.085	0.085	0.108	0.101
	NIQE ↓	3.803	3.639	3.686	3.712	4.361	3.777	3.801	4.012	4.148	3.906	3.526	**3.326**
	NRQM ↑	6.652	6.571	6.631	6.756	6.479	6.582	6.608	6.540	6.518	6.552	6.827	**7.406**
	PI ↓	3.590	3.562	3.549	3.278	3.918	3.617	3.603	3.750	3.831	3.695	3.266	**3.112**
	LR-PSNR ↑	40.170	39.200	40.420	36.390	49.790	44.570	48.310	**50.900**	41.367	48.570	48.250	48.125
DIV2K	PSNR ↑	28.073	26.770	28.190	27.196	28.954	28.959	29.022	29.707	29.250	29.312	29.188	**29.857**
	SSIM ↑	0.770	0.743	0.772	0.740	0.789	0.795	0.798	0.810	0.802	0.803	0.804	**0.820**
	LPIPS ↓	0.116	0.133	0.110	0.145	0.123	0.101	0.103	0.123	**0.097**	0.103	0.104	0.111
	LPIPS-VGG ↓	0.226	0.242	0.218	0.250	0.253	**0.199**	0.212	0.237	0.202	0.200	0.210	0.213
	DISTS ↓	0.058	0.067	0.055	0.067	0.075	**0.053**	0.057	0.076	0.056	0.054	0.054	0.056
	NIQE ↓	2.953	2.911	2.952	**2.576**	3.828	2.966	3.064	3.439	3.237	3.464	2.888	2.943
	NRQM ↑	6.724	6.721	6.694	6.828	6.519	6.610	6.671	6.560	6.611	6.543	**6.870**	6.452
	PI ↓	3.137	3.126	3.158	**2.891**	3.650	3.213	3.231	3.462	3.340	3.516	3.107	3.602
	LR-PSNR ↑	42.915	38.407	42.565	37.758	50.151	45.900	50.514	**51.690**	51.088	42.950	49.057	49.076

Visual Comparison

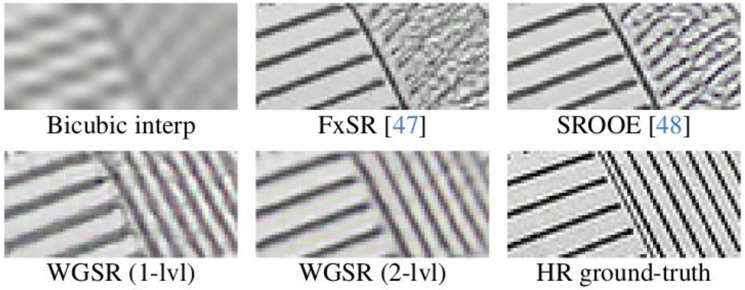
WGSR reconstructs genuine image details with high reconstruction accuracy including the regions with regular patterns and areas containing fine details



KOÇ UNIVERSITY

Parameters of Wavelet Loss

Wavelet Decomposition Level



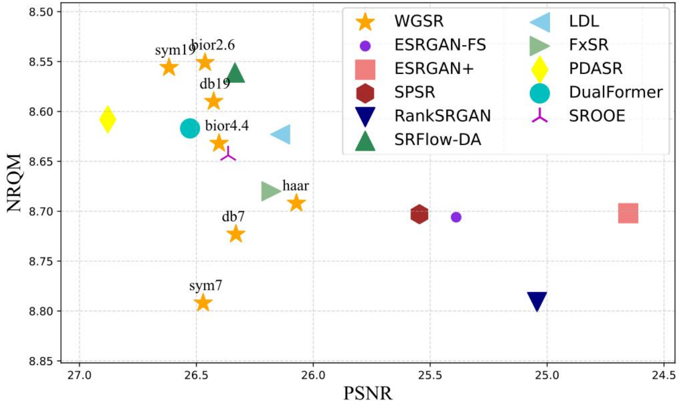
number of SWT levels used depends on scale and orientation of structures in LR images

SOTA GAN SR methods FxSR and SROOE unable to recover correct structure

WGSR 1-level recovers correct orientation of lines but some visible aliasing remains

WGSR 2-level recovers genuine details and structures when level-2 (mid) HF subbands are penalized more in fidelity losses

Choice of Wavelet Family



PD trade-off performance varies according to choice of wavelet family

best objective quality → Symlet “sym19” filter

the best perceptual quality → Daubechies “db7” filter

best trade-off point on the PSNR-NRQM plane → Symlet “sym7” filter

Seung Ho Park, Young Su Moon, and Nam Ik Cho. Flexible style image super-resolution using conditional objective. IEEE Access, 10:9774–9792, 2022.
Seung Ho Park, Young Su Moon, and Nam Ik Cho. Perception-oriented single image super-resolution using optimal objective estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 1725–1735, 2023.

Ablation Study

#	l_1	$L_{adv,G}$	L_{perc}	PSNR ↑	PI ↓
0	RGB	RGB	LPIPS	24.506	2.543
1	RGB	RGB	DISTS	24.812	2.502
2	SWT	RGB	DISTS	25.622	2.501
3	RGB	SWT	DISTS	24.746	**2.443**
4	SWT	SWT	LPIPS	<u>25.859</u>	2.466
5	SWT	SWT	DISTS	**26.331**	<u>2.453</u>

DISTS instead of LPIPS improves 0.3 dB objective performance and 1.6% perceptual performance

SWT fidelity loss improves objective performance by 1 dB

Increasing weight of adversarial loss increases perceptual quality

Selection of weights leads to different PD trade-off points

Proposed combination of losses (#5) achieves +2 dB PSNR gain and better PI score compared to baseline (#0)

D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In IEEE Int. Conf. on Computer Vision. (ICCV), pages 416–423 vol.2, 2001

KOÇ UNIVERSITY

Conclusions

Propose a novel training methodology for GAN-based SR models→ weighted combinations of wavelet losses

Characterization of HF details vs. artifacts better learned by Y-channel wavelet loss functions compared to RGB or Fourier losses

SWT-domain discriminator allows better control of optimization landscape to suppress artifacts

WGSR outperforms existing GAN-SR methods **quantitatively and qualitatively**, **provides better PD trade-off performance**

Extensible: Any off-the-shelf GAN-SR model can be plugged into the proposed framework to benefit from wavelet-loss guidance

THANK YOU

https://github.com/KUIS-AI-Tekalp-Research-Group/image-super-resolution