GAN Prior based Null-Space Learning for Consistent Super-Resolution

\- *AAAI* 2023 ***Oral*** -

Yinhuai Wang, Yujie Hu,
Jiwen Yu, Jian Zhang

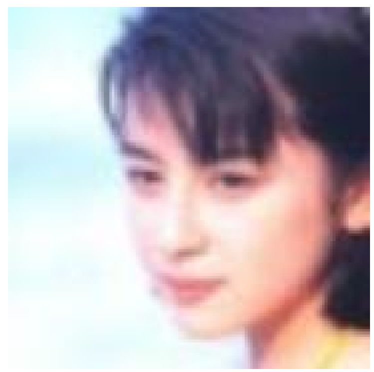
Outline

- Authorship
- Background
- Method
- Experiments
- Conclusion

Background

Noise-free Image Restoration

x



$y = \mathbf{A}x$

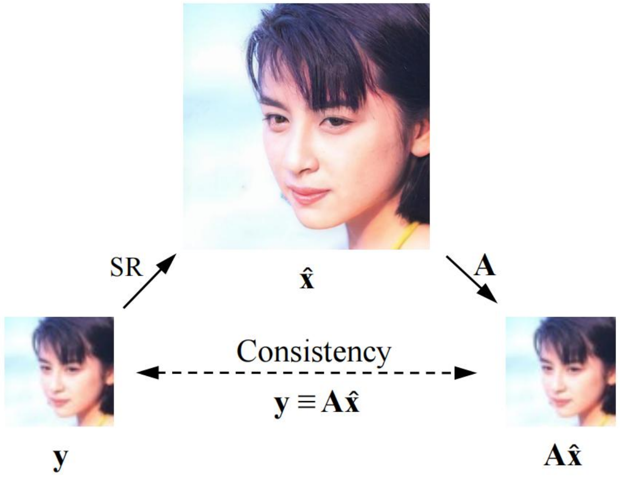
$\hat{x} = \mathbf{A}^{-1}y$

Background

Noise-free Image Restoration

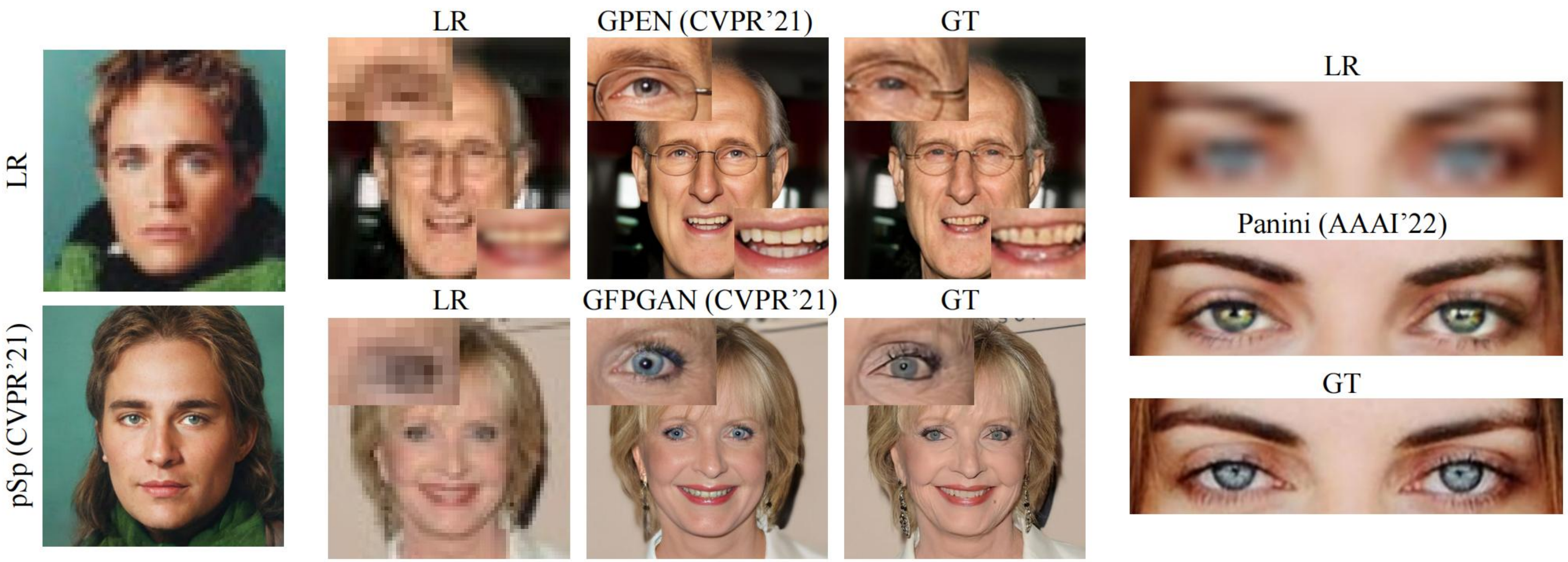
Consistency: $y \equiv \mathbf{A}\hat{x}$

Realness: $\hat{x} \sim p(x)$



Background

GAN-based SR



(a) (b) (c)

Background

Range-Null Space Decomposition (RND)

given a non-zero linear operator $\mathbf{A}$, it usually has at least one pseudo-inverse $\mathbf{A}^{\dagger}$ that satisfies:

$$\mathbf{A}\mathbf{A}^{\dagger}\mathbf{A} \equiv \mathbf{A}$$

It can be obtained by SVD.

Background

Range-Null Space Decomposition (RND)

$\mathbf{A}^{\dagger}\mathbf{A}$ be seen as the operator that projects samples to the range space of $\mathbf{A}$, since $\mathbf{A}\mathbf{A}^{\dagger}\mathbf{A} \equiv \mathbf{A}$. While $\mathbf{I} - \mathbf{A}^{\dagger}\mathbf{A}$ can be seen as the operator that projects samples to the null-space of $\mathbf{A}$, since $\mathbf{A}(\mathbf{I} - \mathbf{A}^{\dagger}\mathbf{A}) \equiv 0$.

Background

Range-Null Space Decomposition (RND)

$$x = \mathbf{A}^{\dagger}\mathbf{A}x + (\mathbf{I} - \mathbf{A}^{\dagger}\mathbf{A})x$$

range space *null space*

Method

In real world, x is unknown.

$$x = \mathbf{A}^{\dagger}\mathbf{A}x + (\mathbf{I} - \mathbf{A}^{\dagger}\mathbf{A})x$$

$$\Downarrow$$

$$x = \mathbf{A}^{\dagger}y + (\mathbf{I} - \mathbf{A}^{\dagger}\mathbf{A})\hat{x}_r$$

Method

In real world, $\mathbf{A}$ is unknown.

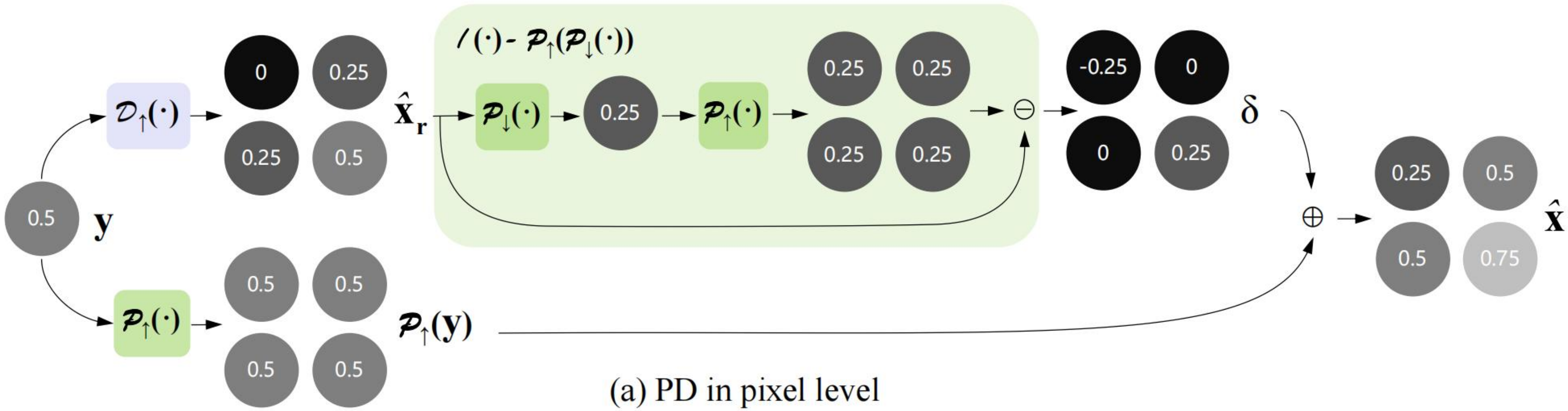
"We observe that many downsampling methods with antialiasing share very similar results."

$$x = \mathbf{A}^{\dagger} y + (\mathbf{I} - \mathbf{A}^{\dagger}\mathbf{A})\hat{x}_r$$

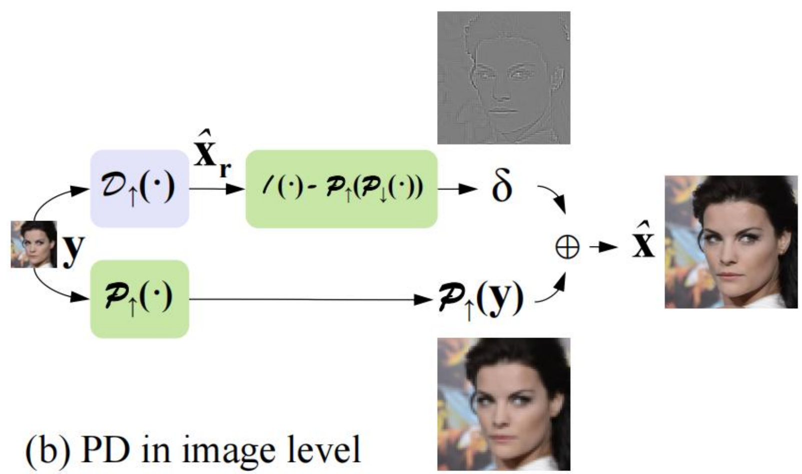
$$\Big\Downarrow \quad \textit{Pooling-based Decomposition}$$

$$x = \mathcal{P}_{\uparrow} y + \hat{x}_r - \mathcal{P}_{\uparrow}\left(\mathcal{P}_{\downarrow}\left(\hat{x}_r\right)\right)$$

Method



(a) PD in pixel level



(b) PD in image level

Experiment

Comparison with CEM [CVPR22]

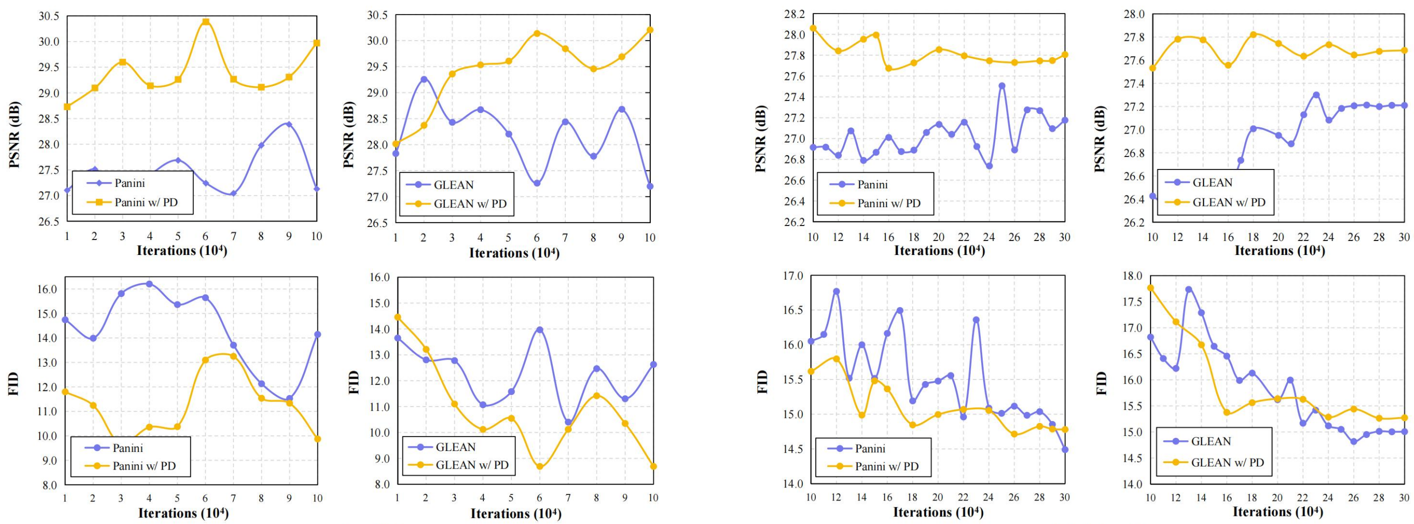
$$\mathbf{A}^{\dagger} = \mathbf{A}^{T}(\mathbf{A}\mathbf{A}^{T})^{-1}$$

Method	PSNR(LR)↑	Time↓
CEM	42.2	31.8ms
PD	**145.7**	**0.68ms**

Table 1: **Validation of the *consistency***. To compare PD with CEM, we calculate the *consistency* strictly following their theories, respectively. The result shows that the implementation of PD is faster and more precise.

Experiment

8x and 16x human face SR



(a) 8× SR (b) 16× SR

Figure 3: **Convergence curves**. Part (a) for the 8× face SR and (b) for the 16× face SR. With PD, both GLEAN and Panini yield significantly higher PSNR and comparable FID.

Experiment

Dataset	Method	PSNR↑	SSIM↑	FID↓
Face	Panini	27.13	0.729	14.15
	Panini w/ PD	**29.97**	**0.801**	**9.87**
	GLEAN	27.20	0.74	12.63
	GLEAN w/ PD	**30.21**	**0.81**	**8.69**
Cat	Panini	22.36	0.596	129.2
	Panini w/ PD	**23.52**	**0.623**	**118.9**
	GLEAN	22.74	0.588	62.92
	GLEAN w/ PD	**22.94**	**0.597**	**58.95**
Church	Panini	19.27	0.483	**67.98**
	Panini w/ PD	**19.80**	**0.491**	69.20
	GLEAN	19.59	0.485	24.49
	GLEAN w/ PD	**19.99**	**0.500**	**24.03**

Table 2: **8× SR on different categories**. The use of PD significantly improves the PSNR, SSIM, and FID in most cases. It is worth noting that PD is parameter-free with negligible computational cost.

Experiment

Method	PSNR↑	SSIM↑	MS-SSIM↑	FID↓
PULSE	21.68	0.676	0.596	42.71
pSp	18.91	0.680	0.526	39.88
GFPGAN	25.17	0.761	0.804	24.34
GPEN	26.07	**0.784**	0.820	31.89
Panini	27.18	0.758	0.843	**14.49**
Panini w/ PD	**27.81**	0.771	**0.851**	14.78
GLEAN	27.21	0.743	0.843	15.01
GLEAN w/ PD	27.69	0.754	0.848	15.27

Table 3: **Comprehensive comparison on 16× face SR**. We compare Panini and GLEAN and their PD-based versions with state-of-the-art face SR methods. The involvement of PD significantly elevates all consistency metrics, i.e., PSNR, SSIM, and MS-SSIM. We attribute the slight rise of FID to the training stochasticity. Actually, the FID is comparable during training, as can be seen in Fig. 3.

Experiment

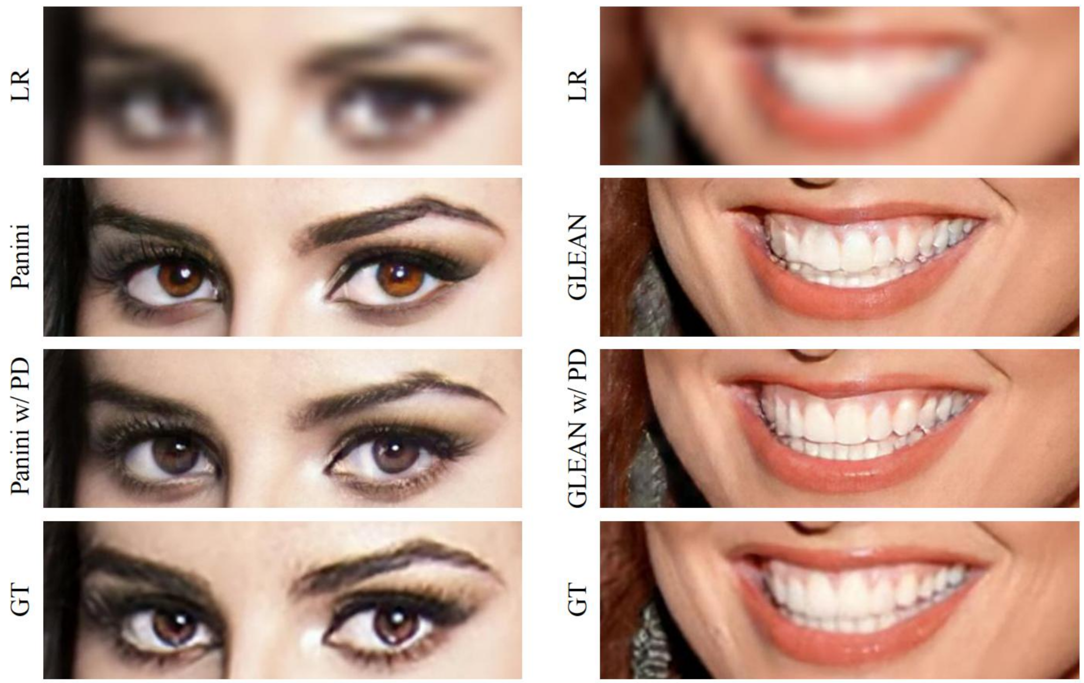


Figure 5: **Qualitative results on 16× face SR**. The use of PD can eliminate color deviation and reduce structural inconsistencies.

Experiment

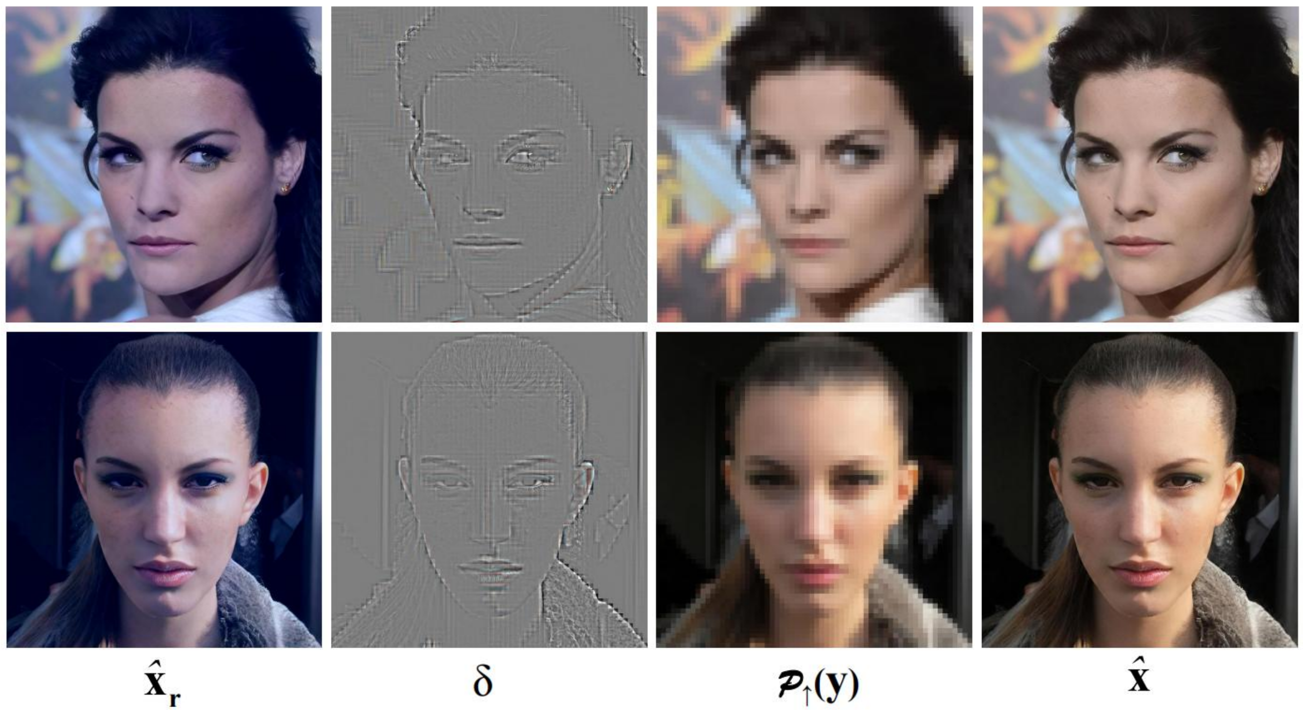
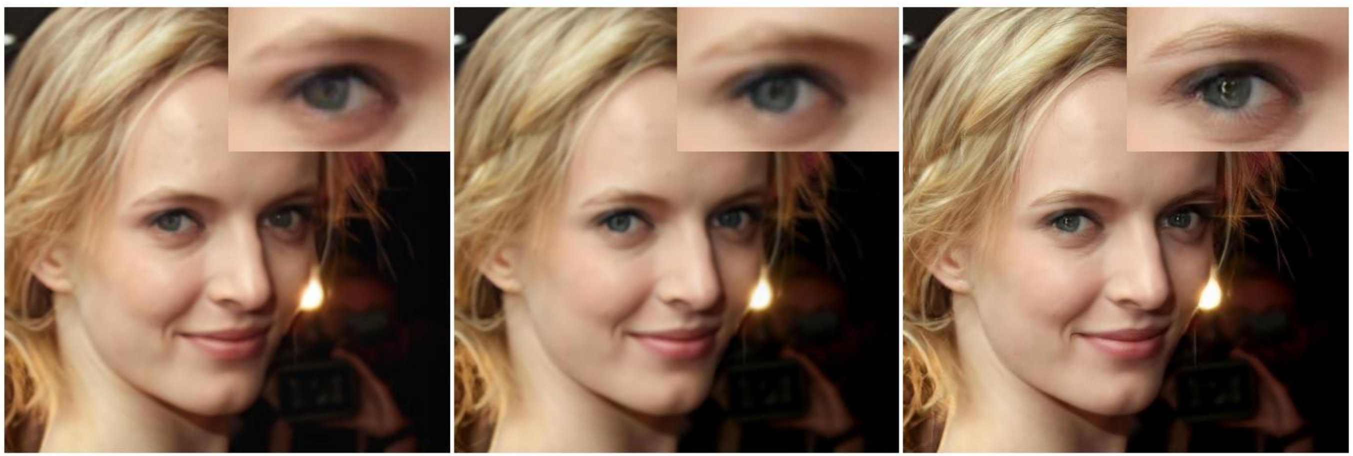


Figure 6: **Visualization of PD.** $\hat{\mathbf{x}}_r$ represent the raw prediction of GAN prior network, δ is the high-frequency part of $\hat{\mathbf{x}}_r$. $\mathcal{P}_{\uparrow}(\mathbf{y})$ denotes the low-frequency contents inherited from LR image. The final result $\hat{\mathbf{x}}$ is yielded by adding $\mathcal{P}_{\uparrow}(\mathbf{y})$ with δ. **(Zoom-in for the best view)**

Experiment

Better generalization results w/o pixel-level loss



Panini GLEAN PDN(Ours)

Figure 7: **Results on unseen downsamplings**. PDN yields clearer results when facing unseen downsamplings. Here the networks are all trained on $8\times$ bicubic(alias) and tested on $8\times$ bicubic(antialias).

Experiment

Better generalization results w/o pixel-level loss

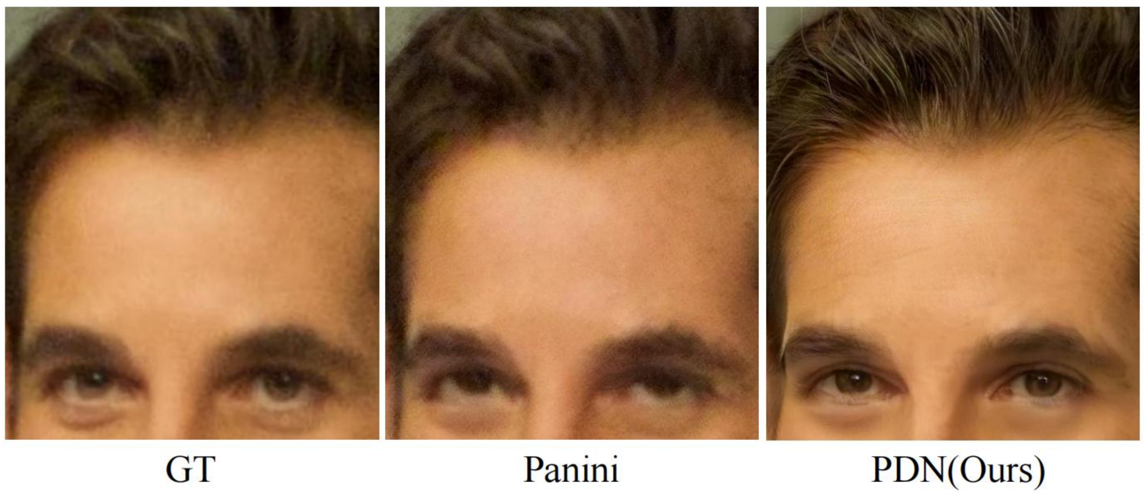
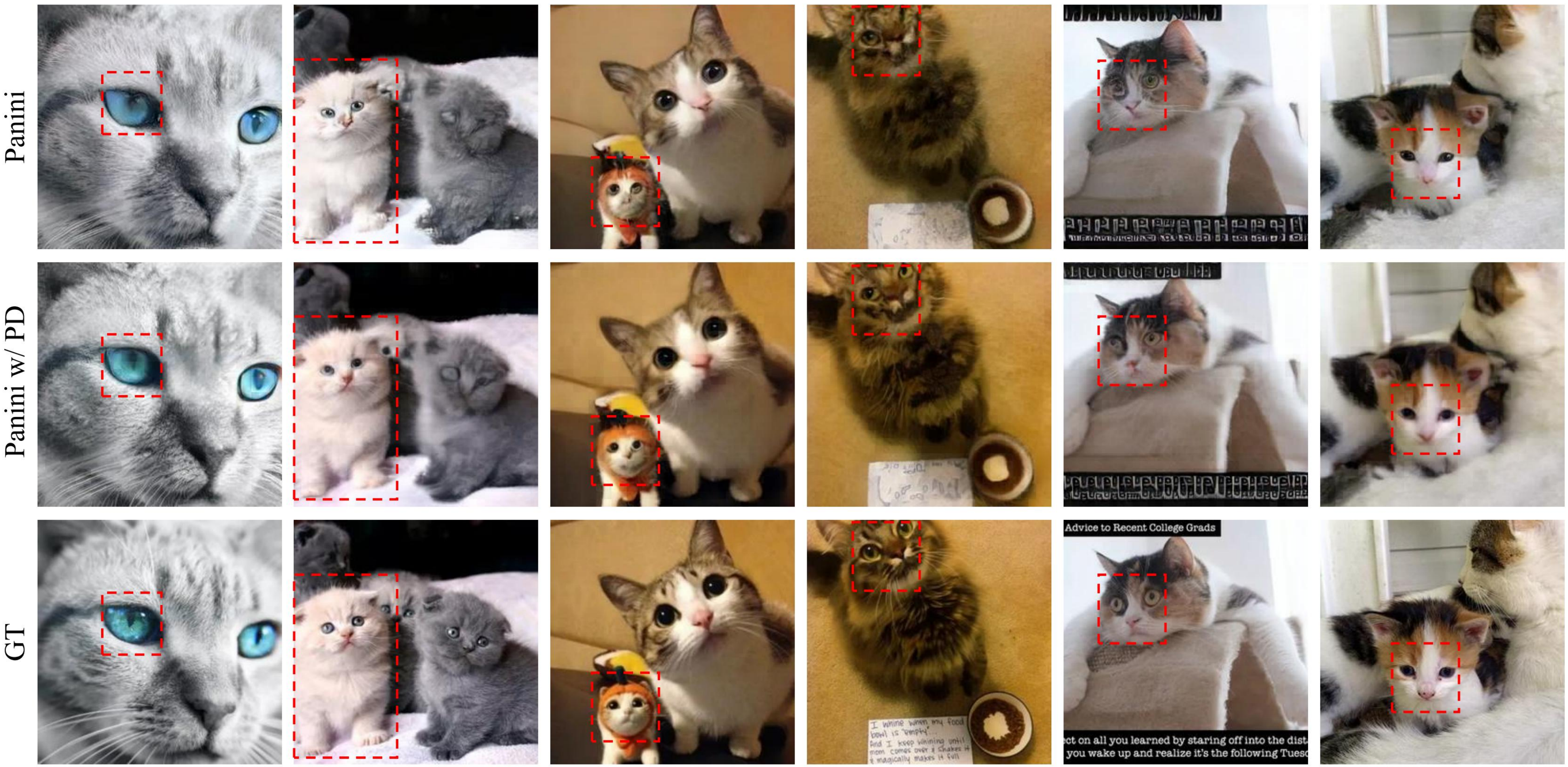
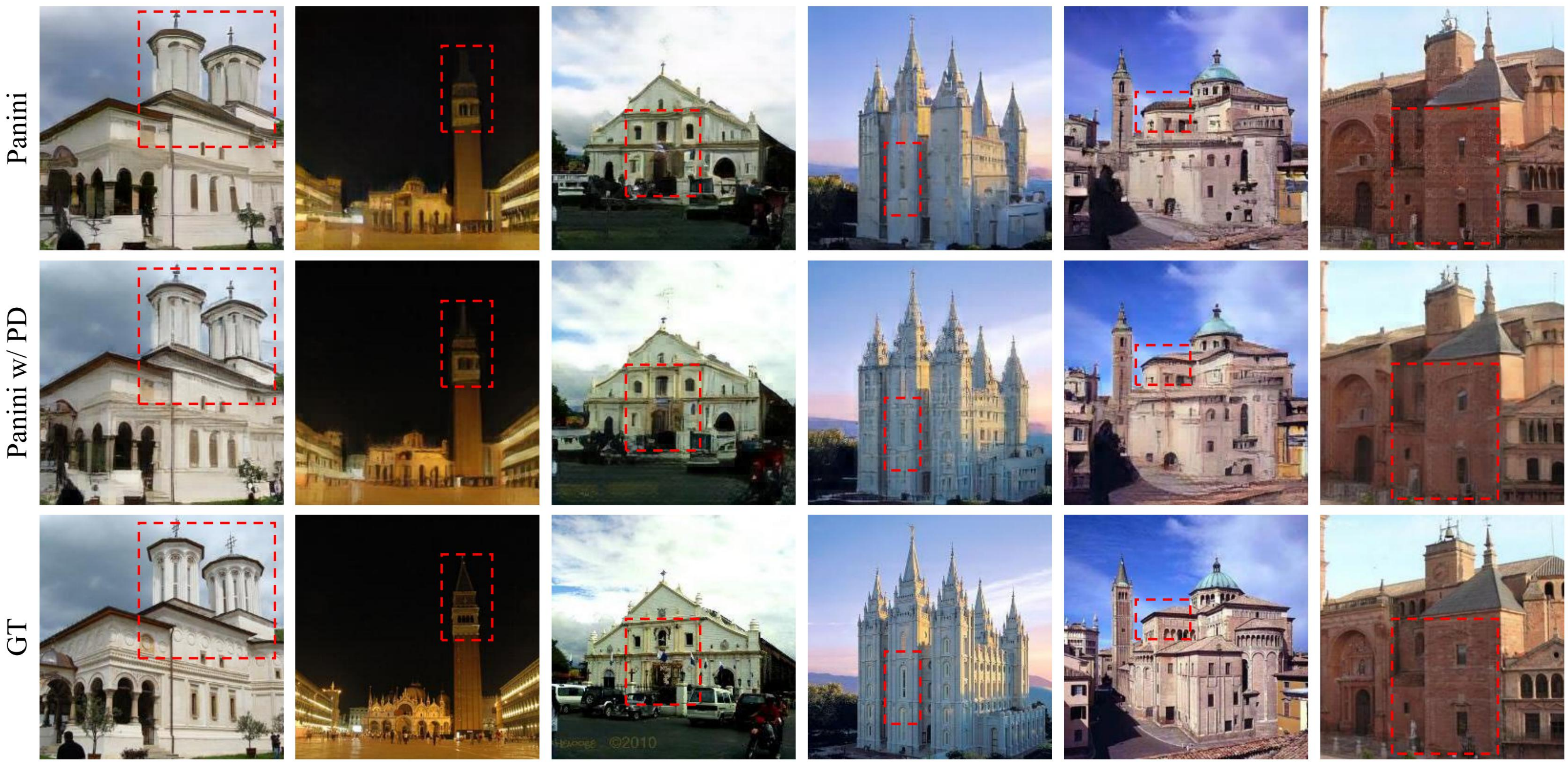


Figure 8: **Results on real-world degradation**. We can see that GLEAN tends to replicate the degradation that GT suffers, while PDN is not affected and tends to generate clear results. Note PDN only uses 16× bicubic(antialias) to synthesize LRs for training, without any simulated degradation.

More results



More results



Conclusion

- A novel method to eliminate inconsistencies for GAN prior based super-resolution networks.

- It can be applied to different backbones, accelerating their training convergence and yielding better consistency.

- It also shows potential in dealing with unseen downsamplings or real-world degradation.

- Hard to deal with hybrid distortions.