Bridging Diffusion Models and 3D Representations:
A 3D Consistent Super-Resolution Framework

Yi-Ting Chen[1]   Ting-Hsuan Liao[1]   Pengsheng Guo[2]   Alexander Schwing[3]   Jia-Bin Huang[1]
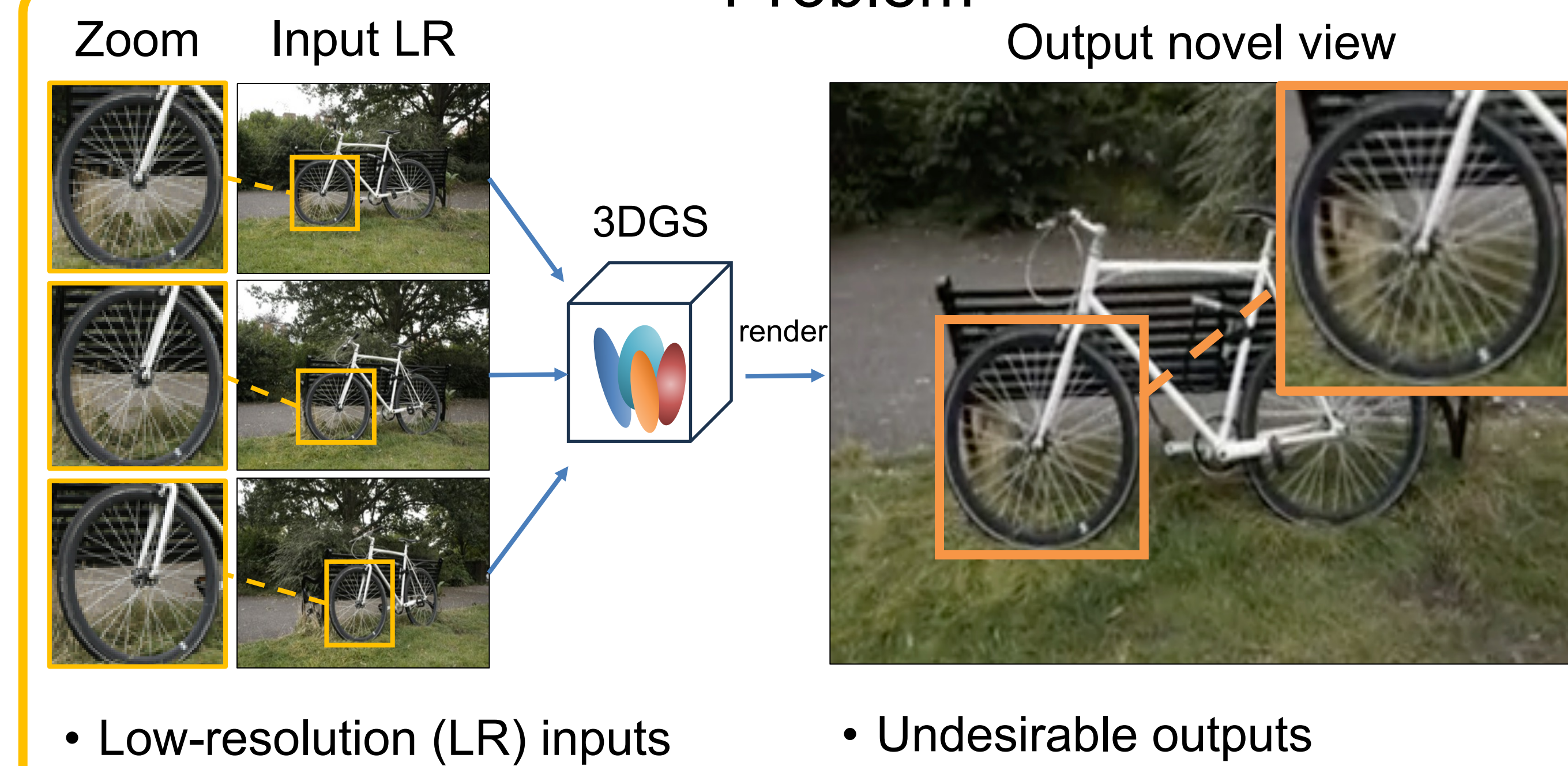
University of Maryland, College Park[1]   Carnegie Mellon University[2]   University of Illinois at Urbana-Champaign[3]
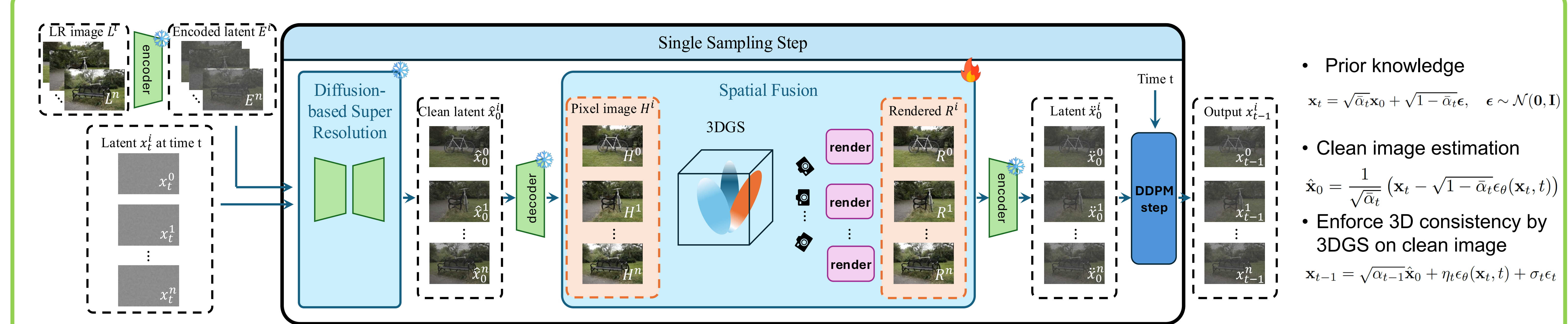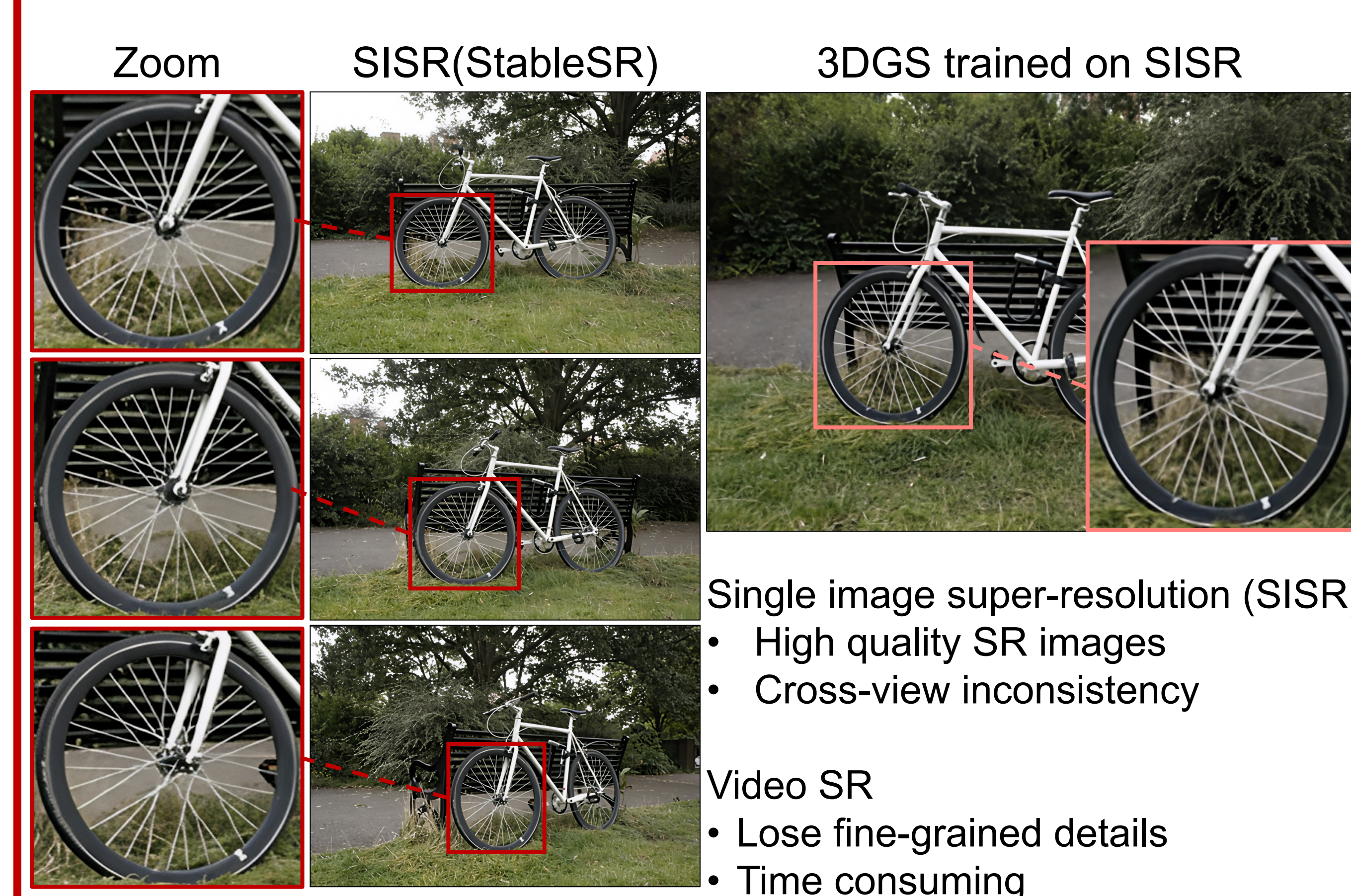
ICCV HONOLULU HAWAII OCT 19-23, 2025

Problem

- Low-resolution (LR) inputs
- Undesirable outputs

3DSR

Single Sampling Step

- Prior knowledge

$$x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, \quad \epsilon \sim \mathcal{N}(0,\mathbf{I})$$

- Clean image estimation

$$\hat{x}_0 = \frac{1}{\sqrt{\bar{\alpha}_t}}\left(x_t - \sqrt{1-\bar{\alpha}_t}\epsilon_\theta(x_t,t)\right)$$

- Enforce 3D consistency by 3DGS on clean image

$$x_{t-1} = \sqrt{\alpha_{t-1}}\hat{x}_0 + \eta_t\epsilon_\theta(x_t,t) + \sigma_t\epsilon_t$$

Challenge

Single image super-resolution (SISR)
- High quality SR images
- Cross-view inconsistency

Video SR
- Lose fine-grained details
- Time consuming
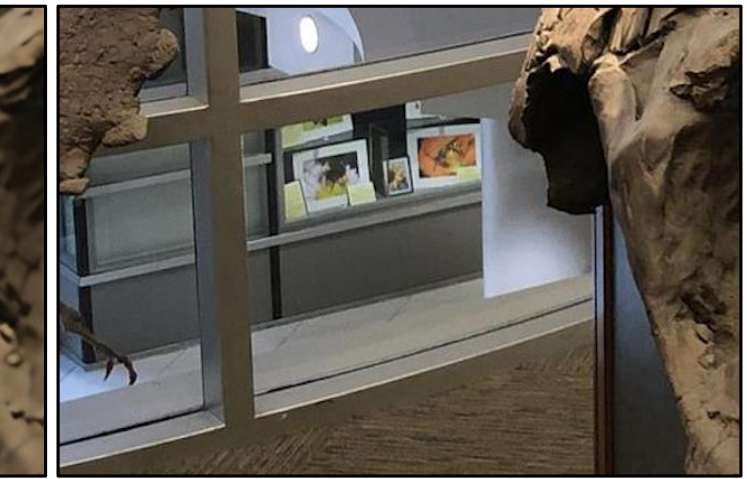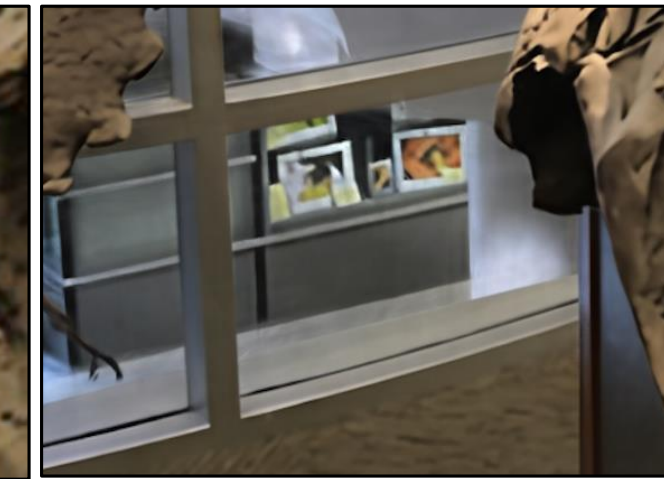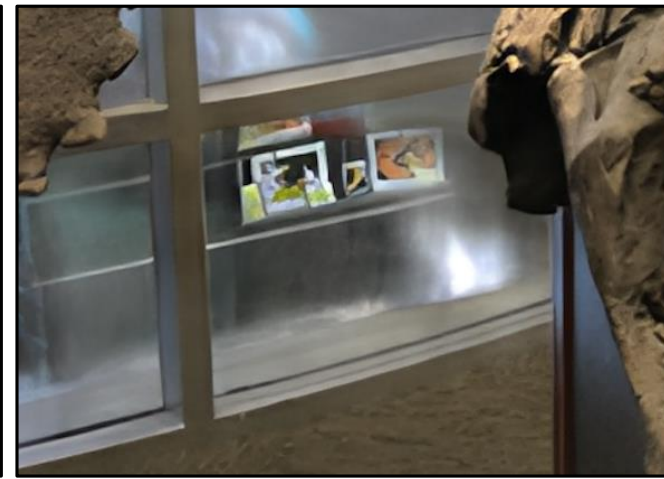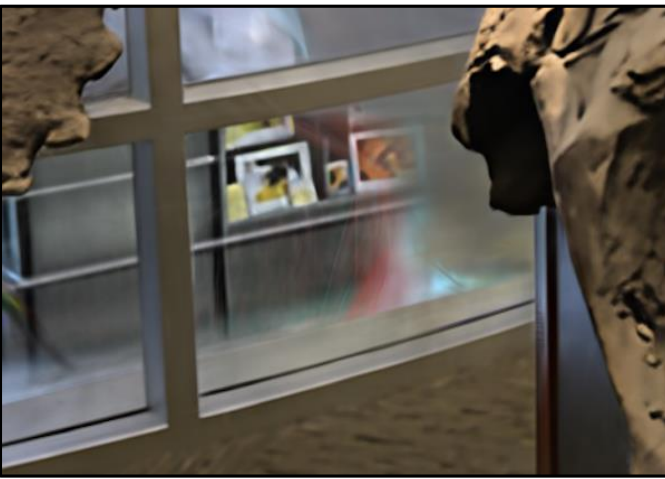
Evaluation

Input LR | StableSR | SuperGaussian | DiSR-NeRF | 3DSR (Ours) | Groundtruth

StableSR | SuperGaussian | 3DSR (Ours) | Groundtruth

LLFF × 8 downsampled and × 4 upsampled

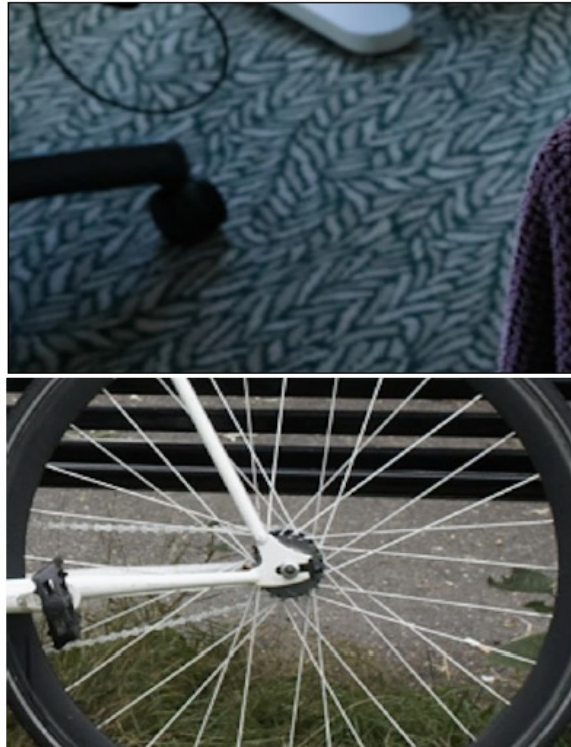| Method	PSNR↑	SSIM↑	LPIPS↓	NIQE↓	MEt3R↓	FID↓
SuperGaussian	23.054	0.725	0.296	4.553	0.541	51.199
DiSR-NeRF	22.504	0.697	0.310	6.293	0.518	54.138
StableSR	22.748	0.717	0.219	4.793	0.531	41.129
3DSR (Ours)	24.212	0.754	0.181	4.632	0.516	20.731

MipNeRF360 × 8 downsampled and × 4 upsampled

Method	PSNR↑	SSIM↑	LPIPS↓	NIQE↓	MEt3R↓	FID↓
SuperGaussian	25.252	0.725	0.303	4.694	0.675	32.652
StableSR	24.313	0.700	0.326	5.177	0.644	44.174
3DSR (Ours)	26.097	0.746	0.222	5.065	0.625	22.438