香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

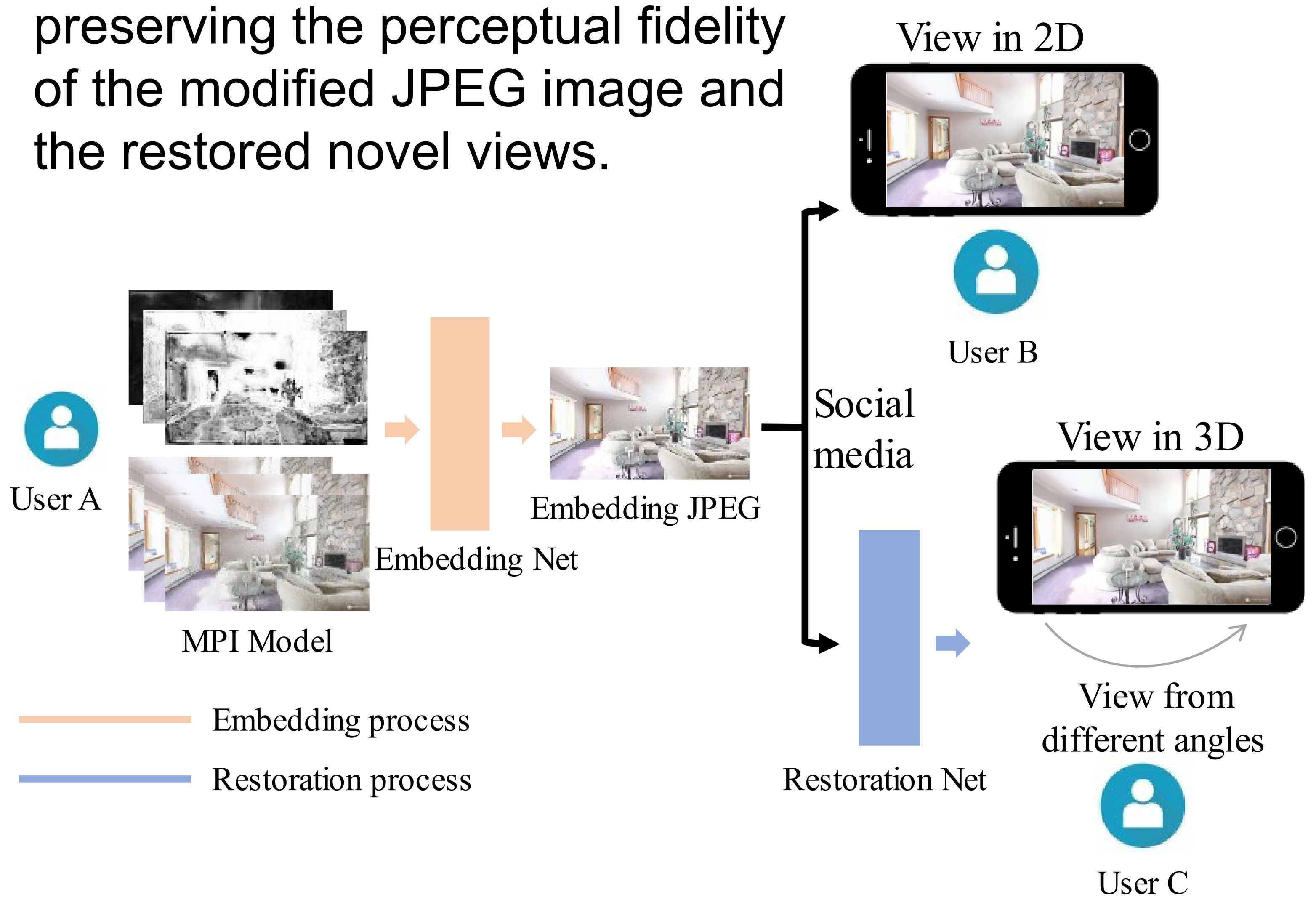
Embedding Novel Views in a Single JPEG Image

Yue Wu* Guotao Meng* Qifeng Chen
The Hong Kong University of Science and Technology

2021 ICCV OCTOBER 11-17 VIRTUAL

Motivation

We propose a novel approach for embedding novel views in a single JPEG image while preserving the perceptual fidelity of the modified JPEG image and the restored novel views.

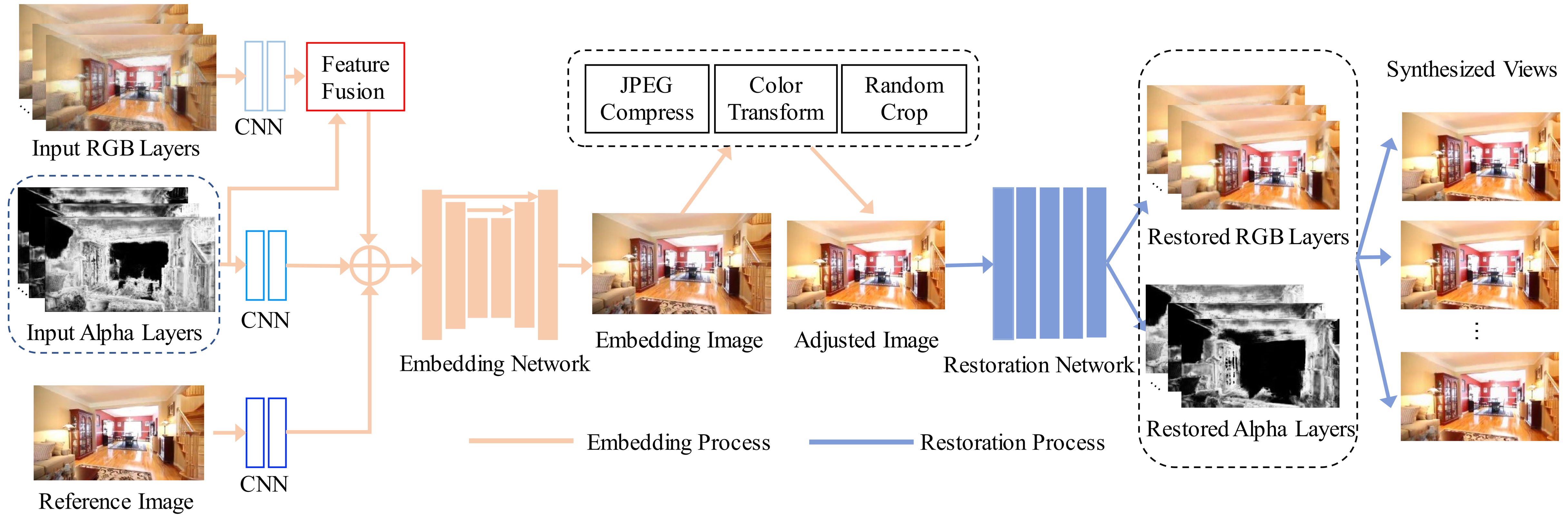


Our approach can embed an MPI into a JPEG image, and then share it on social networks. With our restoration network (as a plugin), users can view this image in 3D from different viewpoints. Meanwhile, users without our restoration network can still view this image as an ordinary 2D image.

Summary

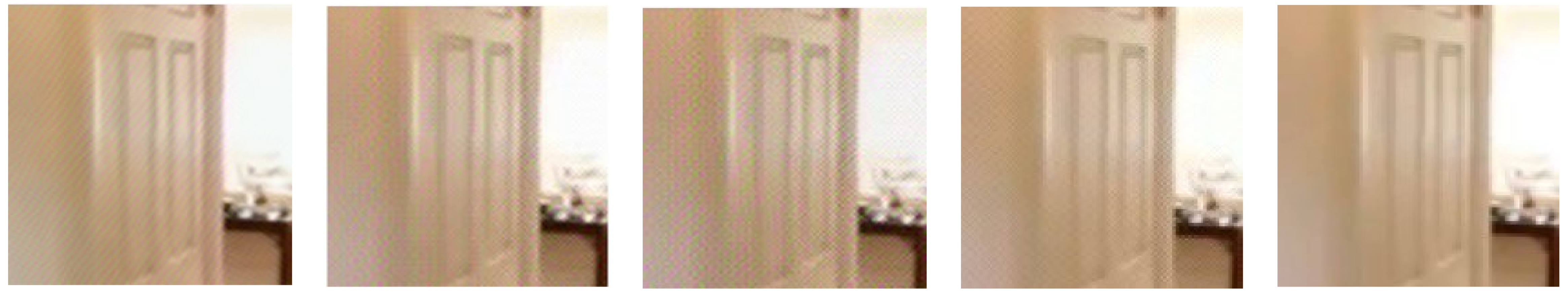
- We propose the first dedicated model that can embed MPIs in JPEG images.
- We introduce adversarial training and frequency domain loss to suppress high-frequency artifacts in the embedding image.
- Our system can embed MPIs in JPEG images in nearly-imperceptible form and restore high-fidelity novel views synthesis. Moreover, our system is robust to a range of image manipulations such as JPEG compression, color adjusting, and cropping.

Method



The framework of the proposed method. First, the features of the RGB layers, the alpha layers, and the reference image are fused to feed into the embedding network to generate the embedding image. Subsequently, the embedding image is compressed and altered. Later the restoration network decodes the embedding image into MPIs. Finally, the novel views can be rendered from different viewpoints with the restored MPIs.

The effectiveness of our proposed loss functions



(a) Video Snapshot[34] (b) W/O frequency (c) W/O adv loss (d) Full model (e) Ground truth

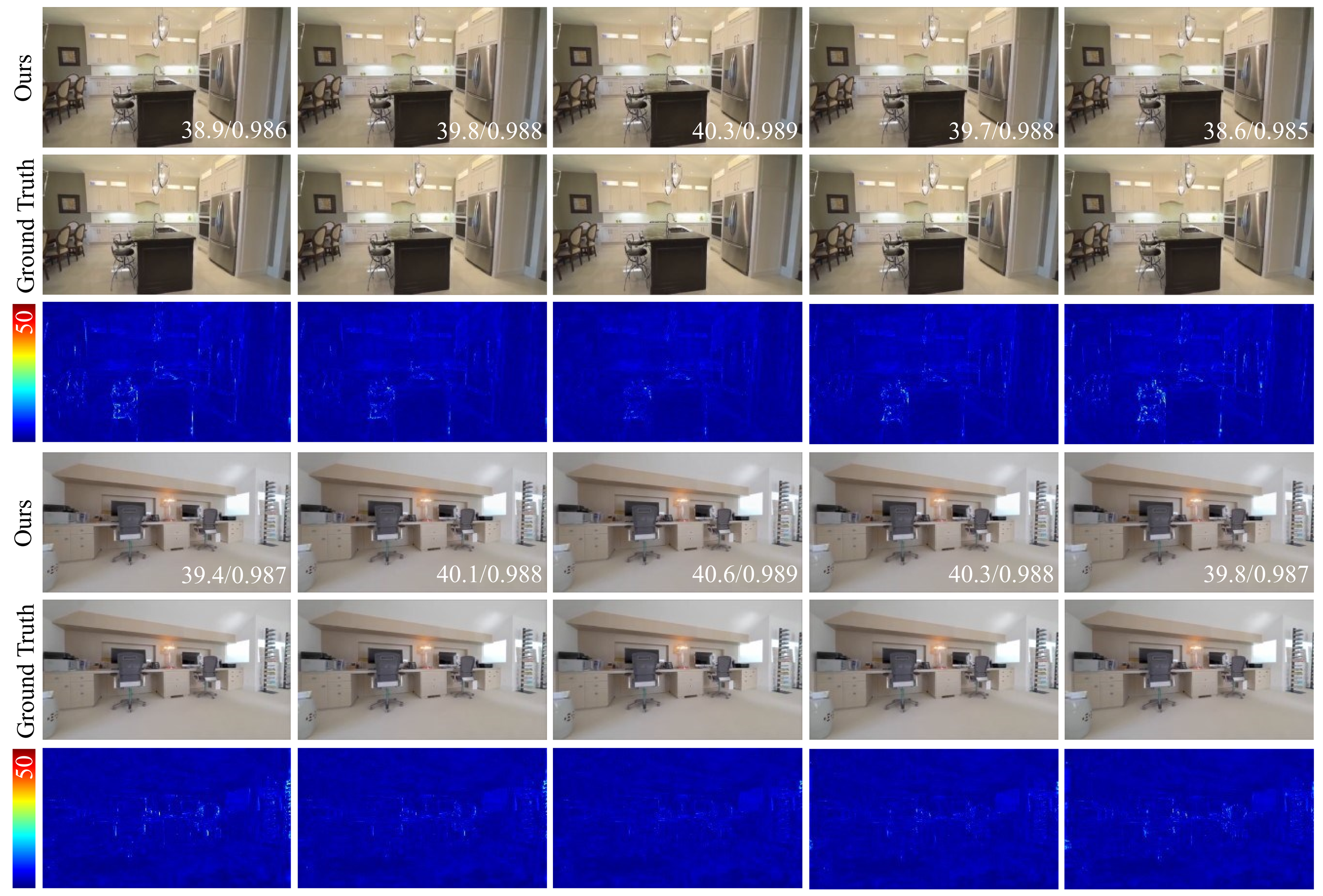
Without the freq loss and the adv loss, the embedding image has evident artifacts.

Quantitative Result

	Stereo-Mag [32]						PB-MPI [21]					
	Embedding			Render			Embedding			Render		
	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓
UNet	0.8830	25.644	0.2654	0.8490	22.676	0.2717	0.8619	26.005	0.2783	0.8016	21.869	0.3057
ResUNet	0.8229	25.780	0.3763	0.8931	24.326	0.2169	0.7871	25.565	0.3785	0.8441	23.071	0.2614
Video Snapshot [34]	0.7939	30.395	**0.1007**	0.8703	27.066	0.1732	0.7705	30.305	0.1045	0.8417	25.885	0.2553
Ours w/o GAN	0.8661	32.884	0.1832	0.9664	35.174	0.0637	0.9537	36.529	0.1215	0.9393	30.435	0.1107
Ours w/o Render	**0.9688**	**33.998**	**0.0926**	0.9492	32.301	0.1137	**0.9695**	**37.650**	**0.0667**	0.8912	27.497	0.1845
Ours w/o Frequency	0.7970	30.911	0.2844	**0.9773**	**38.105**	**0.0509**	0.7983	27.796	0.2349	**0.9513**	**30.831**	**0.0948**
Ours	**0.8941**	**34.616**	0.1736	**0.9750**	**36.683**	**0.0535**	**0.9593**	**36.736**	**0.0951**	**0.9533**	**32.840**	**0.0951**

The comparison of the quality of the embedding images and the rendered novel views.

Qualitative Result



The results of the synthesized images and the difference maps.

Quantitative Result

	SSIM/PSNR	Speed (s)	Model (M)
LDI	0.8426/25.735	42.3	438
Two views (PNG)	0.9016/27.639	0.448	264
Ours (PNG)	0.8953/27.198	0.017	6.3

The comparison between our method and LDI in terms of render quality.

	Model (M)	Recover MPI (s)	Render (s)
SynSin [28]	273	-	0.077
S-MPI [24]	167	0.692	0.706
Stereo [32]	185	0.427	0.441
PB [21]	524	2.483	2.497
Ours	**6.3**	**0.003**	**0.017**

The first column is the model size. The second column is the inference time for generating single MPI. The third column is the time for rendering a single view.

	Render		
	SSIM↑	PSNR↑	LPIPS↓
SynSin [28]	0.7851	24.078	0.2016
S-MPI [24]	0.8084	23.837	0.1905
Ours	**0.8810**	**26.732**	**0.1507**

We compare our method with state-of-the-art single image view synthesis methods. The results demonstrate that embedding the novel views into a single image produces much better results than single image view synthesis.