# SPARF: Large-Scale Learning of 3D Sparse Radiance Fields from Few Input Images

Abdullah Hamdi[1,2], Bernard Ghanem[2], Matthias Nießner[1]

Technical University of Munich (TUM)   King Abdullah University of Science and Technology (KAUST)

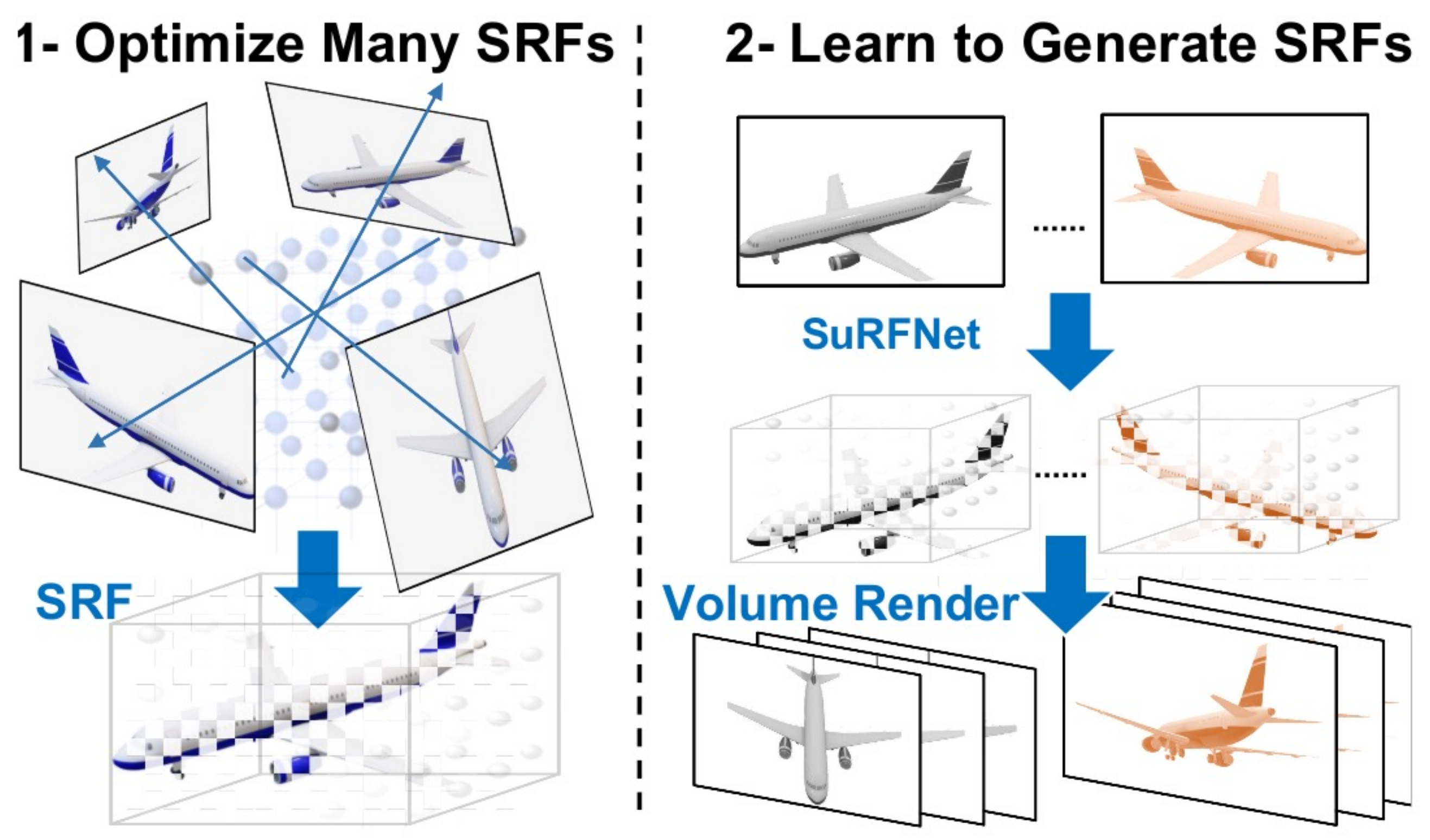
ICCV23 PARIS

website:

## Can we learn 3D priors on Radiance Fields (NeRFs) ?

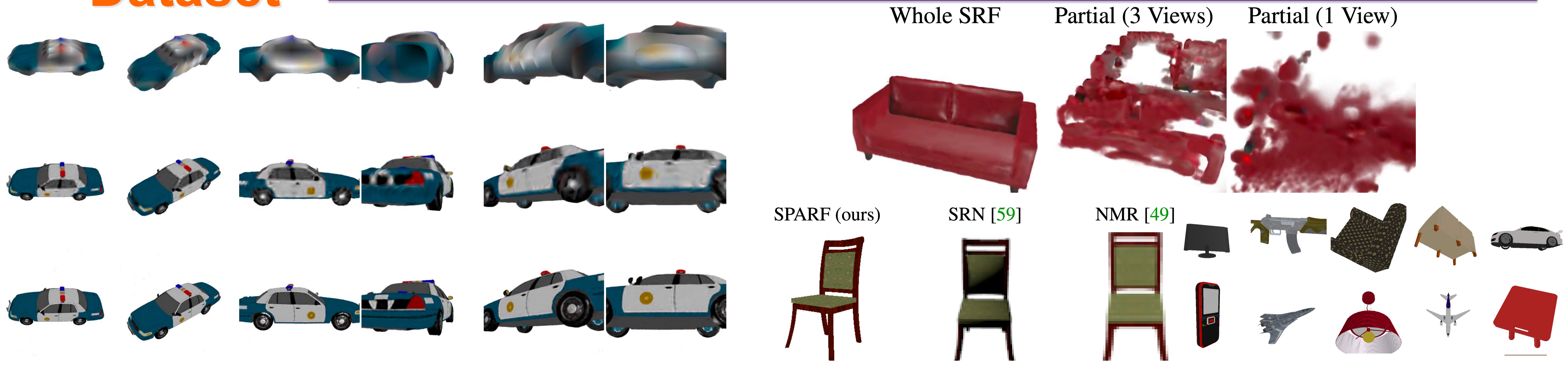
1- Propose SPARF, a large-scale dataset of 3D shapes Plenoxels with multiple voxel resolutions (32, 128, 512)

2- propose SuRFNet, a pipeline to generate SRFs conditioned on input images, achieving SOTA on ShapeNet novel views synthesis from one or few input images.



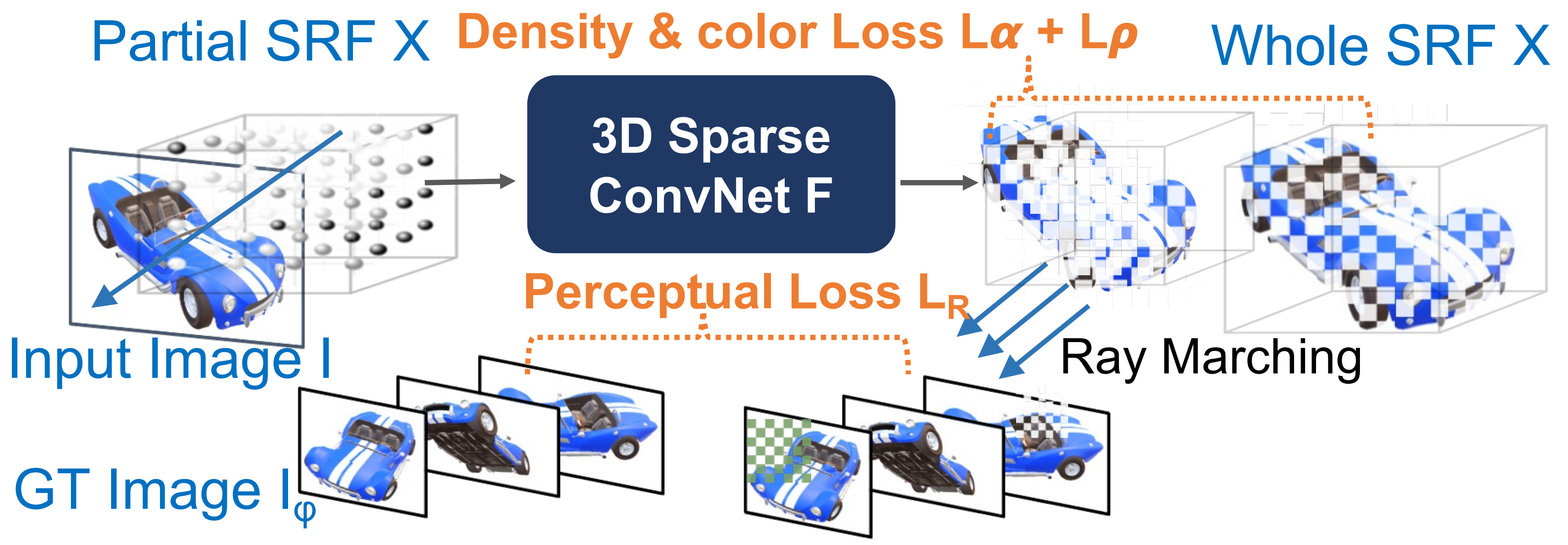
## SPARF Dataset

**1M NeRF, 17M images, 40K shapes**



## Novel Views Synthesis

### Pipeline



$$\text{Loss}_{\mathbf{F}} = L_\alpha + \lambda_\rho L_\rho + \lambda_R L_R$$

$$L_\rho\left(\mathcal{X}, \hat{\mathcal{X}}\right) = \|\mathbf{M}_\alpha \mathbf{F}(\mathcal{X})_\rho - \mathbf{M}_\alpha \hat{\mathcal{X}}_\rho\|_1$$

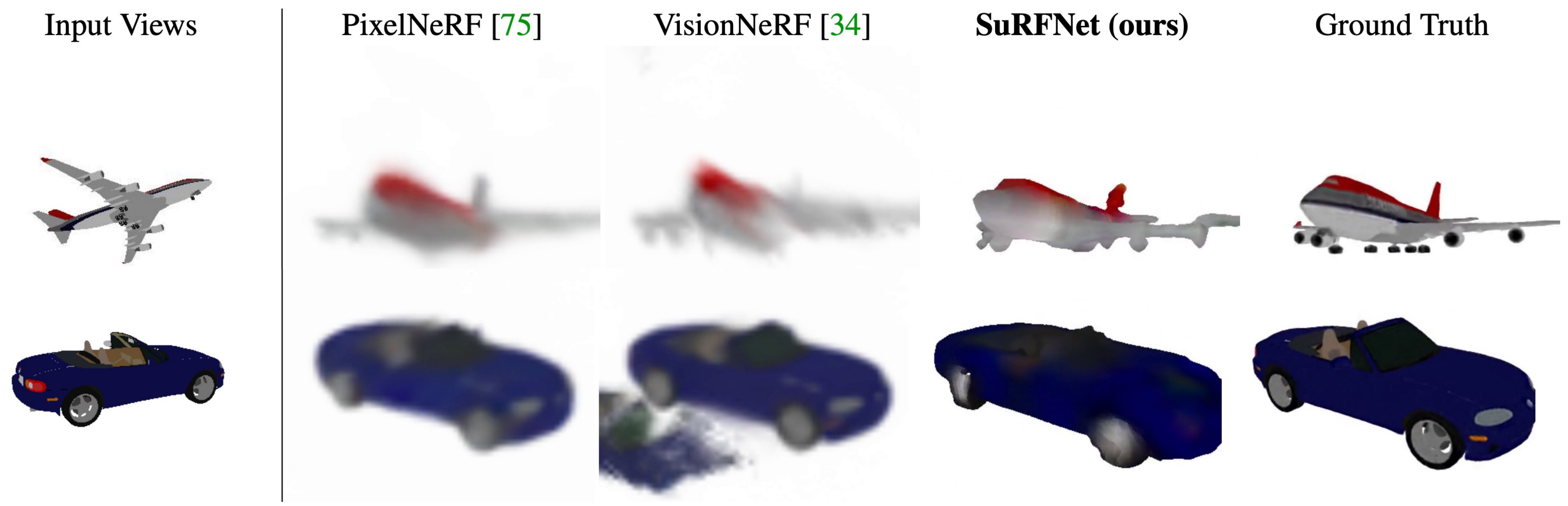
$$\text{s. t.} \quad \mathbf{M}_\alpha = \mathbb{1}(\hat{\mathcal{X}}_\alpha > \alpha_{\text{dense}})$$

$$L_R\left(\mathcal{X}\right) = \|\mathcal{R}_\phi\left(\mathbf{F}(\mathcal{X})\right) - \mathbf{I}_\phi\|_1,$$

$$L_\alpha\left(\mathcal{X}, \hat{\mathcal{X}}\right) = -(\hat{\mathbf{y}}\log(\mathbf{y}) + (\mathbf{1}-\hat{\mathbf{y}})\log(\mathbf{1}-\mathbf{y}))$$

$$\text{s. t.} \quad \hat{\mathbf{y}} = \mathbb{1}\left(\mathcal{S}(\hat{\mathcal{X}}_\alpha) > \alpha_{\text{dense}}\right), \ \mathbf{y} = \mathcal{S}\left(\mathbf{F}(\mathcal{X})\right)_\alpha$$

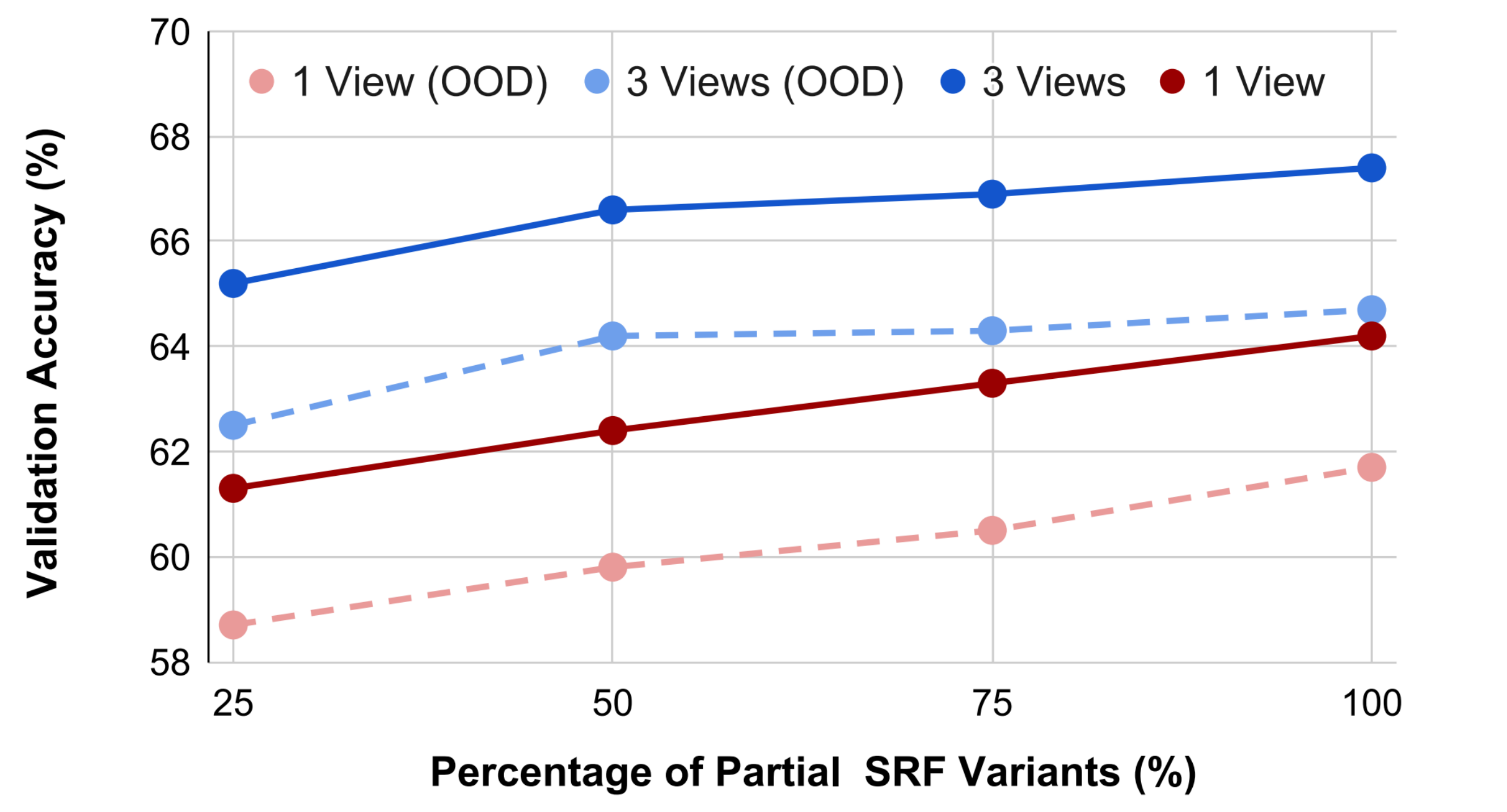
### Results

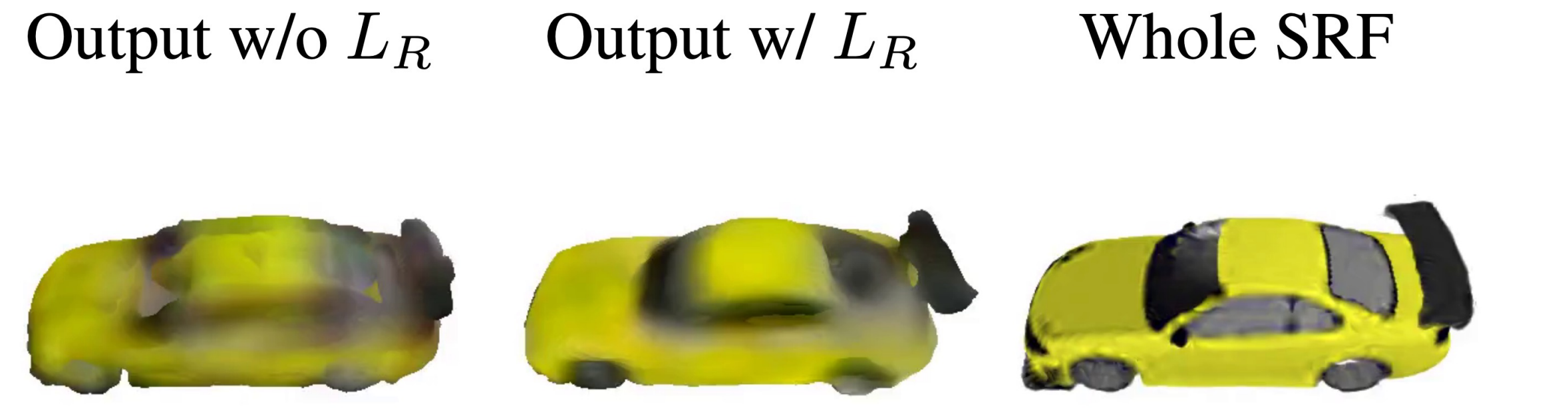


## Analysis

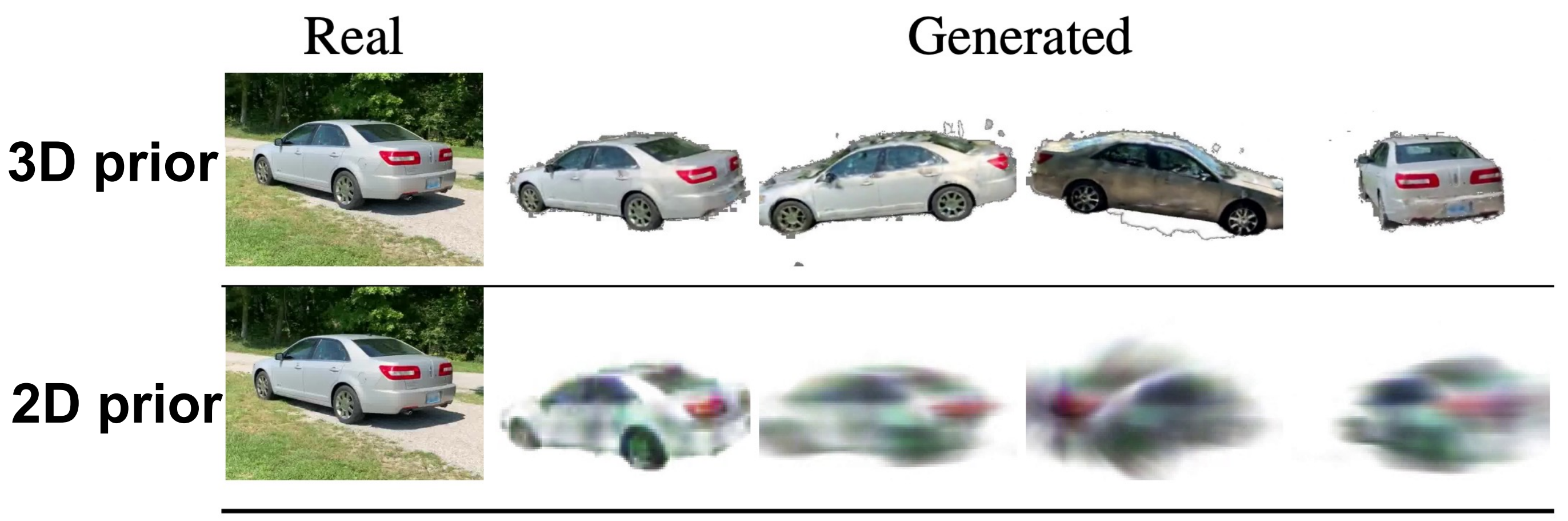
| Baselines | chair | watercraft | rifle | display | lamp | speaker | cabinet | bench | car | airplane | sofa | table | phone | mean |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **SPARF Classes** | | | | | | | | | | | | | |
| Plenoxels [14] (1V) | 9.2 | 11.1 | 11.7 | 8.0 | 13.6 | 8.2 | 10.4 | 10.5 | 7.1 | 12.8 | 9.3 | 9.9 | 8.3 | 10.0 |
| Plenoxels [14] (3V) | 10.7 | 13.3 | 14.9 | 9.7 | 15.8 | 10.4 | 12.4 | 11.6 | 7.1 | 14.6 | 11.6 | 10.8 | 9.7 | 11.7 |
| PixelNerf [63] (1V) | 13.3 | 16.3 | 16.7 | 11.9 | 17.6 | 11.3 | 14.5 | 14.6 | 13.2 | 19.2 | 13.3 | 13.2 | 13.2 | 14.5 |
| PixelNerf [63] (3V) | 13.5 | 16.6 | 16.9 | 12.2 | 17.9 | 11.9 | 14.9 | 14.8 | 13.4 | **19.4** | 13.4 | 13.3 | 13.3 | 14.7 |
| VisionNeRF [28] (1V) | 13.0 | 15.6 | 15.8 | 11.7 | 16.7 | 11.2 | 14.0 | 14.3 | 12.7 | 17.8 | 13.3 | 13.0 | 12.6 | 14.0 |
| **SuRFNet (ours) (1V)** | 11.6 | 16.2 | 17.0 | 12.0 | 16.2 | 12.6 | 17.0 | 13.5 | 16.6 | 17.5 | 14.1 | 10.1 | 15.3 | 14.6 |
| **SuRFNet (ours) (3V)** | **15.3** | **18.3** | **18.8** | **15.0** | **19.0** | **16.6** | **20.0** | **15.6** | **16.6** | 18.5 | **18.1** | **14.9** | **17.8** | **17.3** |

**SPARF Benchmark on Out-Of-Distribution View Synthesis:** One view (1V) and three views (3V) inputs are reported.





| Network | Network FLOPs (G) | Network Inference (ms) | Parameters Number (M) | Rendering Speed (FPS) |
|---|---|---|---|---|
| PixelNeRF [63] | 7.3 | 5.33 | 21.8 | 1.2 |
| VisionNerf [28] | 33.7 | 12.5 | 68.6 | 1.2 |
| SuRFNet (small) | ~15 | 14.4 | 13.4 | 15 |
| SuRFNet (large) | ~100 | 90.0 | 87.3 | 15 |



**Acknowledgements**: This work was supported by the King Abdullah University of Science and Technology (KAUST) and the German Research Foundation (DFG)