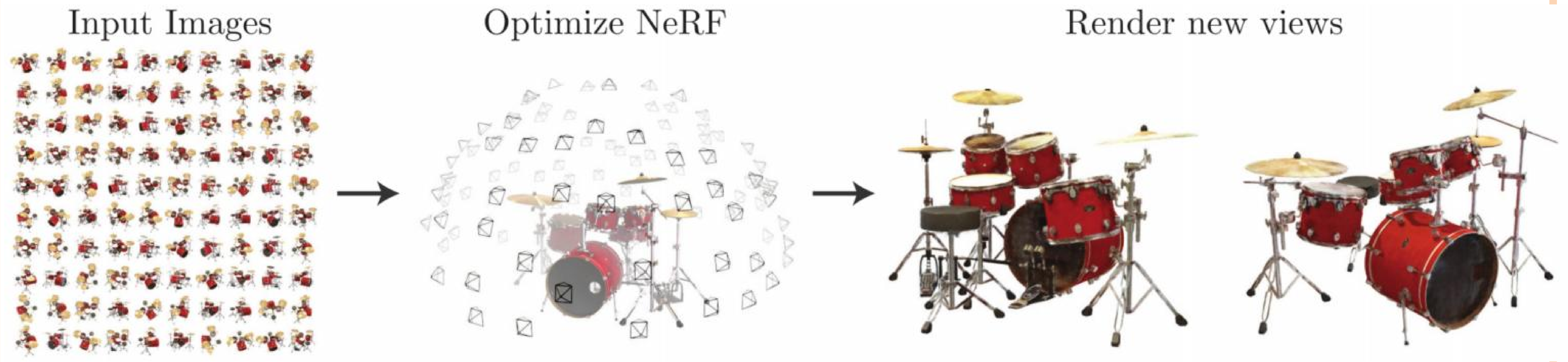
# REGULARIZING NEURAL RADIANCE FIELDS FROM SPARSE RGD INPUTS


Qian Li, Franck Multon, Adnane Boukhayma

*Inria, Univ. Rennes, CNRS, IRISA, M2S, France*

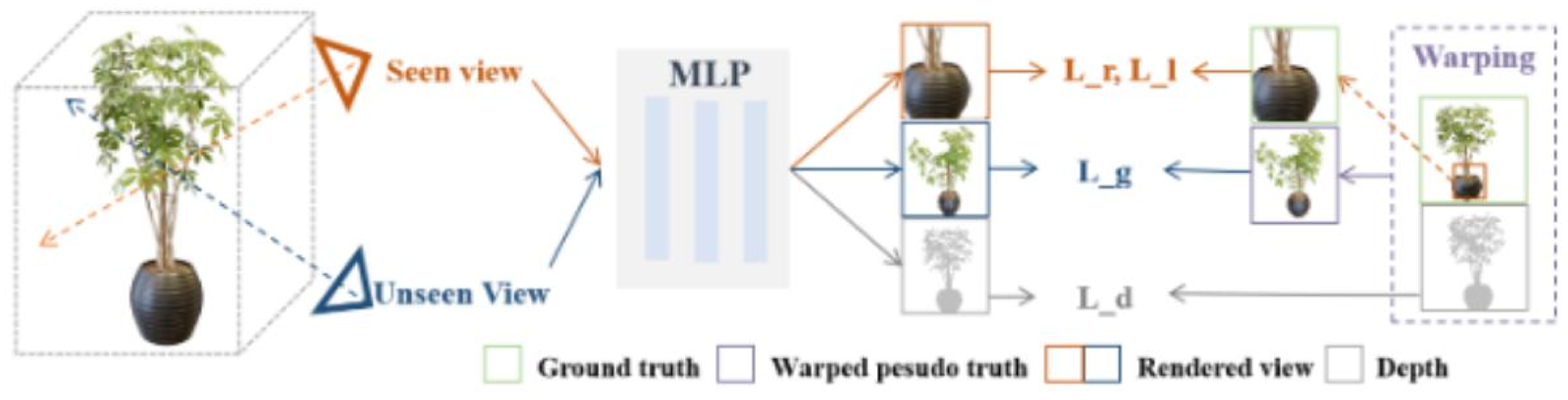

## PROBLEM AND MOTIVATAION

NeRF achieves photo-realistic renderings when given dense inputs, while its' performance drops dramatically with the decrease of training views' number.



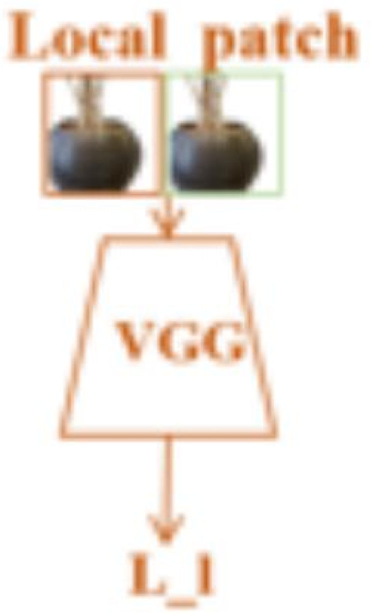

Our work aims to improve neural radiance fields (NeRF) from sparse inputs.

## METHOD




Overview of the proposed framework. Our regularizations include photo-metric loss L_r, global geometry loss L_g, local patch appearance loss L_l and depth loss L_d.

## ALGORITHM

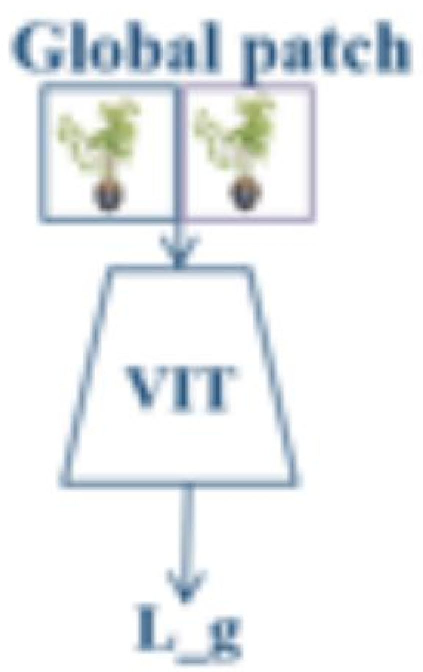



Our contribution includes:

1. we propose a global sampling strategy with a geometry regularization utilizing warped images as augmented pseudo-views to encourage geometry consistency across multi-views.

$$L_g = \|\phi(I_w) - \phi(I_r)\|_2^2$$

Iw is the warped patch of pseudo-views, Ir is the rendered patch, and $\Phi$ denotes an image encoder.




2. we int roduce a local patch sampling scheme with a patch-based regularization for appearance consistency.

$$L_l = \|\varphi(I_t) - \varphi(I_r)\|_2^2$$

I_t is the warped patch of pseudo-views, I_r is the rendered patch, and $\varphi$ denotes an image encoder.

3. our method exploits depth information for explicit geometry regularization.

$$L_d = \|M \odot (D_r - D_t)\| \quad \hat{D}_r = \frac{1}{\sum_{z=1}^{N} T_z \alpha_z} \sum_{z=1}^{N} T_z \alpha_z t_z$$

D r is the rendered depth from NeRF,
Dt is the ground truth depth, and M denotes masks.

## RESULTS

The proposed approach outperforms existing baselines on real benchmarks DTU datasets from sparse inputs and achieves the state of art results.

| Method | Setting | PSNR↑ | | | SSIM↑ | | | LPIPS↓ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 3-view | 6-view | 9-view | 3-view | 6-view | 9-view | 3-view | 6-view | 9-view |
| SRF [22] | Trained on DTU | 15.32 | 17.54 | 18.35 | 0.671 | 0.730 | 0.752 | 0.304 | 0.250 | 0.232 |
| PixelNeRF[10] | | 16.82 | 19.11 | 20.40 | 0.695 | 0.745 | 0.768 | 0.270 | 0.232 | 0.220 |
| MVSNeRF[11] | | 18.63 | 20.70 | 22.40 | 0.769 | 0.823 | 0.853 | 0.197 | 0.156 | 0.135 |
| FWD [17] | | 21.98 | - | - | 0.791 | - | - | 0.208 | - | - |
| mip-NeRF [9] | Optimized per Scene on DTU | 8.68 | 16.54 | 23.58 | 0.571 | 0.741 | 0.879 | 0.353 | 0.198 | 0.092 |
| TensoRF [6] | | 13.77 | 15.84 | 17.27 | 0.545 | 0.614 | 0.662 | 0.382 | 0.296 | 0.267 |
| DictNeRF [7] | | 11.85 | 20.63 | 23.83 | 0.633 | 0.778 | 0.823 | 0.314 | 0.201 | 0.173 |
| RegNeRF [8] | | 18.89 | 22.20 | 24.93 | 0.745 | 0.841 | 0.884 | 0.190 | 0.117 | 0.089 |
| DSNeRF [12] | | 16.9 | 20.60 | 22.30 | 0.57 | 0.75 | 0.81 | 0.45 | 0.29 | 0.24 |
| **Ours** | | 22.02 | 24.16 | 25.74 | 0.802 | 0.829 | 0.858 | 0.135 | 0.151 | 0.076 |

**Table 1**: Comparison of the average PSNR, SSIM, and LPIPS of reconstructed images on the DTU [16] dataset, using 3/6/9 views for training. The higher the better for both PSNR and SSIM. The lower the better for LPIPS [26]. The purple color represents the performance of ranking, the darker the better.

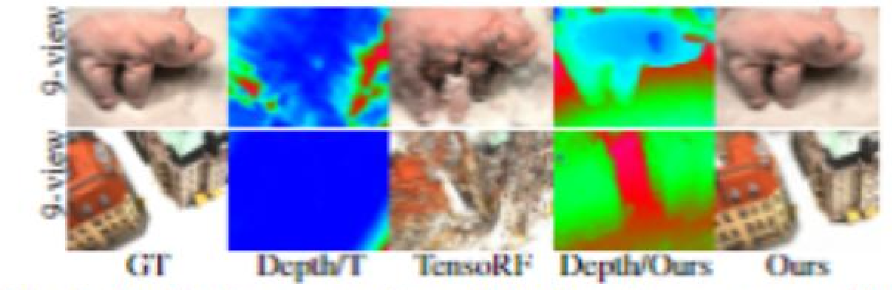



**Fig. 3**: Qualitative comparison of depth and appearance with TensoRF [6] from 9 input views on the DTU dataset [16].

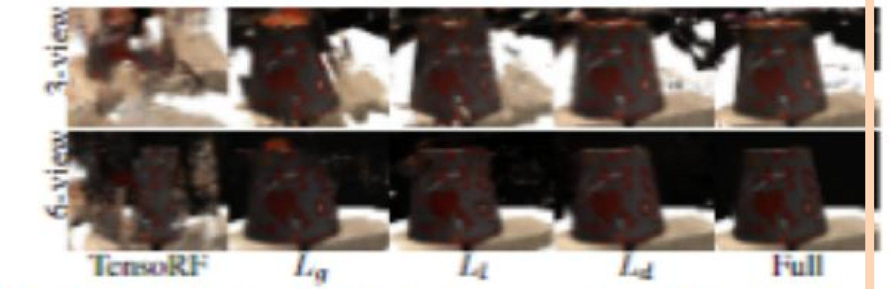



**Fig. 4**: Qualitative ablations of proposed regularization from 3/6 inputs on the DTU dataset [16].

## CONCLUSION

- ✓ This paper introduces a novel approach to improve neural radiance fields(NeRF) from sparse RGBD inputs.
- ✓ Our work achieves state of art performance across different metrics.
- ✓ However, one of the limitations is that it requires depth. Future work could explore how to use monocular depth estimated by networks. More researches explore implicit surface recon- struction from sparse RGBD inputs.