ViP-NeRF: Visibility Prior for Sparse Input Neural Radiance Fields

Nagabhushan Somraj   Rajiv Soundararajan

Visual Information Processing (VIP) Lab, Indian Institute of Science, Bengaluru

https://nagabhushansn95.github.io/publications/2023/ViP-NeRF.html

① NeRF with Sparse Inputs

- NeRF [1] typically requires hundreds of images per scene.
- Produces severe distortions when trained with few images.
- Cause: Under-constrained volume rendering equations.

NeRF with dense input views

NeRF with sparse input views

② Prior Work on Few-Shot NeRF

DS-NeRF

- Obtains sparse depth from Structure from Motion (SfM).
- Supervises depth estimated by NeRF using SfM depth.
- Accurate but supervision at sparse keypoints only.

DDP-NeRF

- Completes sparse depth using pre-trained neural network.
- Dense depth supervision provided at every pixel.
- Suffers from generalization issues when generating prior on novel scenes.

- There is a need for dense and reliable prior.
  - We introduce visibility prior.

③ Visibility Prior

Occluded in second view

Visible in both views

Primary View

Secondary View

- Visibility prior indicates if a pixel in primary view is visible in secondary view.
- We constrain the visibility learned by NeRF with this visibility prior.

Why Visibility Prior?

- Related to relative depth.
- Easier to estimate.
- Dense and Reliable.
- Estimated without neural networks.

④ Visibility Prior Estimation

Primary View Image

Secondary View Image

warp to primary view at different depths

Primary View Image

Plane Sweep Volume

Error Maps

white - lower error
black - higher error

threshold minimum error

Visibility Map corresponding to primary view

white - visible
black - occluded

⑤ ViP-NeRF Model Architecture

Secondary Ray

Primary Ray

Primary View

Secondary View

NeRF MLPs

$\mathcal{F}_2$

$\mathcal{F}_2$

shared weights

Volume Rendering

$\hat{\mathbf{c}} \rightarrow \mathcal{L}_{mse}$

$\hat{T}_i$

$T_i \rightarrow \mathcal{L}_v$

$t' \rightarrow \mathcal{L}_{vip}$

Visibility Prior Loss

⑥ Training NeRF with Visibility Prior

$T_i = 1; T'_i = 0$
$T_i = 1; T'_i = 1$

Primary View

Secondary View

$T_i \leftarrow$ Visibility of 3D point in primary view
$T'_i \leftarrow$ Visibility of 3D point in secondary view
$\tau' \leftarrow$ Visibility prior $\in \{0,1\}$

$t' = \sum_i w_i T'_i \leftarrow$ average visibility of the pixel in secondary view

Visibility prior loss:

$\mathcal{L}_{vip} = \|\tau' - t'\|_1 \odot 1_{\{\tau'=1\}}$

Prior reliable when $\tau' = 1$

Sparse depth loss [2] in addition to $\mathcal{L}_{vip}$

How to compute T'_i?

- Naive approach: Shoot ray from secondary view.
  - N^2 MLP queries per pixel instead of N.
- Our approach: NeRF to learn view-dependent visibility $\hat{T}_i, \hat{T}'_i$.

$t' = \sum_i w_i \hat{T}'_i$

$\mathcal{L}_v = \|\hat{T}_i - T_i\|_2^2$

Primary View

Secondary View

⑦ Evaluation and Comparisons

Databases:

- Real Estate-10K [4]
  - 5 scenes, indoor & outdoor
- NeRF-LLFF [5]
  - 8 scenes, challenging scenes

Evaluation Measure: LPIPS

Real Estate - 10K [4]   NeRF - LLFF [5]

Comparison with State-of-the-art Sparse Input NeRFs

Input 1   InfoNeRF   DS-NeRF   ViP-NeRF
Input 2   RegNeRF   DDP-NeRF   Ground Truth

Ablations

ViP-NeRF   ViP-NeRF   w/o sparse depth   w/o vis prior   Ground Truth

Take home message: Add visibility prior to any NeRF model to train it with sparse views.

⑧ References and Acknowledgements

[1] Mildenhall et al., "Representing Scenes as Neural Radiance Fields for View Synthesis", ECCV 2020.
[2] Deng et al., "Depth-Supervised NeRF: Fewer Views and Faster Training for Free", CVPR 2022.
[3] Roessle et al., "Dense Depth Priors for Neural Radiance Fields From Sparse Input Views", CVPR 2022.
[4] Zhou et al., "Stereo Magnification: Learning View Synthesis using Multiplane Images", SIGGRAPH 2018.
[5] Mildenhall et al., "Local Light Field Fusion", SIGGRAPH 2019.