

Ladybird: Quasi-Monte Carlo Sampling for Deep Implicit Field Based 3D Reconstruction with Symmetry

Yifan Xu^{*1}, Tianqi Fan^{*2,3}, Yi Yuan^{†1}, and Gurprit Singh³

¹ Netease Fuxi AI Lab, Hangzhou, China

{xuyifan, yuanyi}@corp.netease.com

² Saarland Informatics Campus, Saarbruecken, Germany

³ MPI for Informatics, Saarbruecken, Germany

{tfan,gsingh}@mpi-inf.mpg.de

1 Validation accuracy

In Table 1, we report the SDF validation accuracy. The experimental setup is the same as that in Section 3.3, and our validation set consists of 6000 images. We see that Grid+FPS results in faster convergence and higher SDF validation accuracy.

Table 1. Validation accuracy of different sampling method.

| Epoch | 1 | 2 | 3 | 5 | 10 | 30 |
|-------------|-------|-------|-------|-------|-------|-------|
| Grid+Random | 0.743 | 0.777 | 0.788 | 0.803 | 0.817 | 0.825 |
| Grid+FPS | 0.803 | 0.859 | 0.872 | 0.888 | 0.905 | 0.917 |

2 Spectrum, more on discrepancy

FPS induces blue-noise behavior by construction. Gaussian Jitter+FPS gives a power spectrum with blue-noise characteristics (Figure 1). However, Jitter+FPS gives higher discrepancy compared to Grid+FPS and worse 3D reconstruction results. Generating good 3D blue noise samples at 256^3 resolution is computationally very expensive. Hence we excluded blue-noise samplers in this work.

The discrepancy depends on the initial sample size, final sample size, and their ratio. In Table 2, we report the Star Discrepancy (x0.01) of different samplers with varying initial sample size. In the original FPS paper [1], the author gave a deterministic bounds on the distance between sample points (Theorem 4.2), which is used to prove that FPS is a uniform sampler. This analysis shields some lights on why FPS results in low-discrepancy, as it could lead to a deterministic bounds on discrepancy.

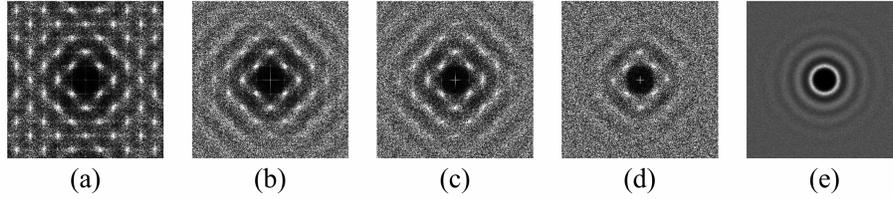


Fig. 1. Power spectra of (a) Grid+FPS, (b) Jitter+FPS ($\sigma = 0.005$), (c) Jitter+FPS ($\sigma = 0.01$), (d) Jitter+FPS ($\sigma = 0.02$), (e) Blue noise.

Table 2. Mean ($\times 0.01$) and standard deviation ($\times 0.01$) of star discrepancy of different samplers. A+B means we first sample $n = 128^2, 256^2, 512^3$ points using sampling method A and then select a subset of size 2048 with method B.

| Initial sample size | Metric | Grid+Random | Grid+FPS | Jitter+FPS | Sobol+FPS |
|---------------------|--------|-------------|----------|------------|-----------|
| 128×128 | Mean | 3.06 | 2.84 | 5.41 | 1.75 |
| | Std | 0.34 | 0.18 | 0.16 | 0.08 |
| 256×256 | Mean | 2.98 | 2.48 | 6.07 | 2.51 |
| | Std | 0.26 | 0.16 | 0.77 | 0.47 |
| 512×512 | Mean | 3.07 | 2.66 | 6.48 | 2.62 |
| | Std | 0.5 | 0.1 | 0.31 | 0.23 |

3 Marching Cube at higher resolution

Using Ladybird configured as in Section 3.5, we run Marching Cube at different resolutions (64^3 and 512^3). Due to the high memory and computation requirement at increased resolution, we only report CD for 100 objects that are randomly sampled from the ShapeNet test dataset. The results are summarized in Table 3.

Table 3. Effect of Marching Cube resolution on the reconstruction results on 100 objects randomly sampled from ShapeNet test set. Metrics are class mean of CD ($\times 0.001$) computed on 2048 points.

| Resolution | Grid+Random | Grid+FPS | Sobol+FPS |
|------------|-------------|----------|-----------|
| 64^3 | 10.79 | 9.04 | 10.20 |
| 512^3 | 10.60 | 8.81 | 9.76 |

4 Limitations

The reconstruction quality of Ladybird is restricted by the input image resolution (currently 137×137). However, issues such as memory, speed and compatibility

with pre-trained image networks need to be considered when increasing the input image resolution. We would like to address the problem of 3D reconstruction from a high resolution image in future work.

Since we need to spatially align the image to the mesh and utilize the corresponding local features, accurate camera pose is crucial to our method (Figure 2). A better camera pose estimation network will lead to significant improvement of our system.

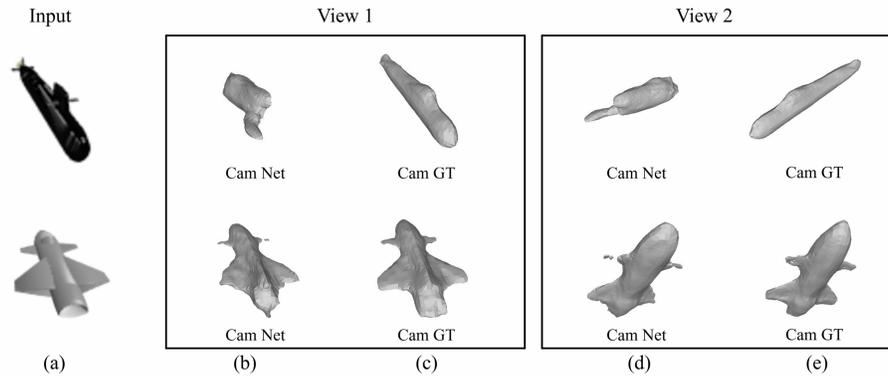


Fig. 2. Inaccurate estimation of camera pose leads to failures in reconstruction. (a) indicates the input images. (b) and (d) are the reconstruction results using estimated camera poses in two different views. (c) and (e) are the reconstruction results using ground truth camera poses in two different views.

References

1. Schlömer, T., Heck, D., Deussen, O.: Farthest-point optimized point sets with maximized minimum distance. In: Proceedings of the ACM SIGGRAPH Symposium on High Performance Graphics. pp. 135–142 (2011)