EUROPEAN CONFERENCE ON COMPUTER VISION

CPT-VR: Improving Surface Rendering via Closest Point Transform with View-Reflection Appearance

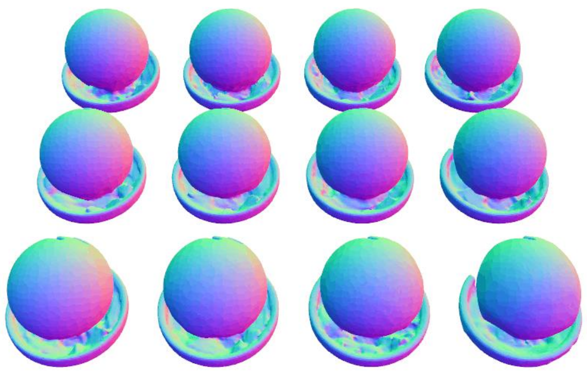
Zhipeng Hu, Yongqiang Zhang, Chen Liu, Lincheng Li, Sida Peng, Xiaowei Zhou, Changjie Fan, Xin Yu

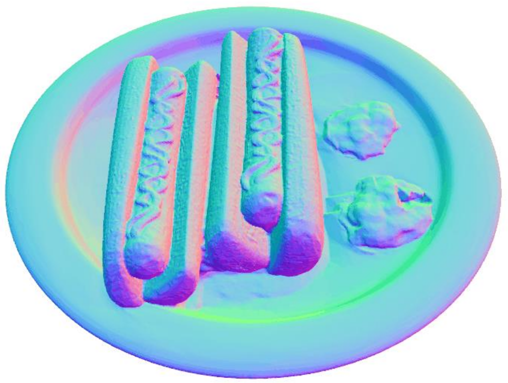
NetEase Fuxi AI Lab, The University of Queensland, Zhejiang University

Board ID 215, Paper ID 9272

1 Overview

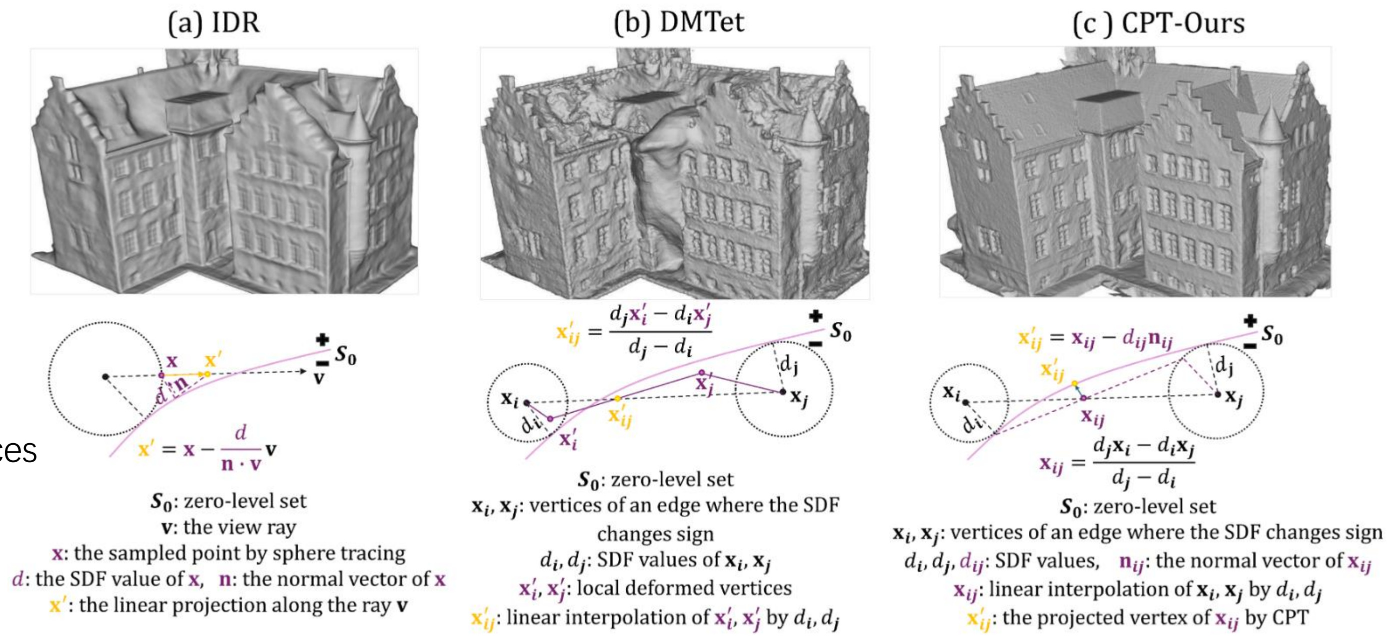
- **Differentiable surface rendering via Closest Point Transform**
  - Projects points exactly on the current surface
  - Address the point deviation issue caused by linear interpolation
  - Not need a differentiable Marching Cubes thanks to **CPT**
- **View-Reflection appearance**
  - **View** rays handle the regions without highlights
  - **Reflection** rays based on the points from CPT are more accurate
  - Combination of View-Reflection rays and CPT capture specular regions well.
- **1-point background**
  - No longer be dependent on any prior BG. Knowledge for surface rendering
  - Simple but efficient for reconstruction of the foreground





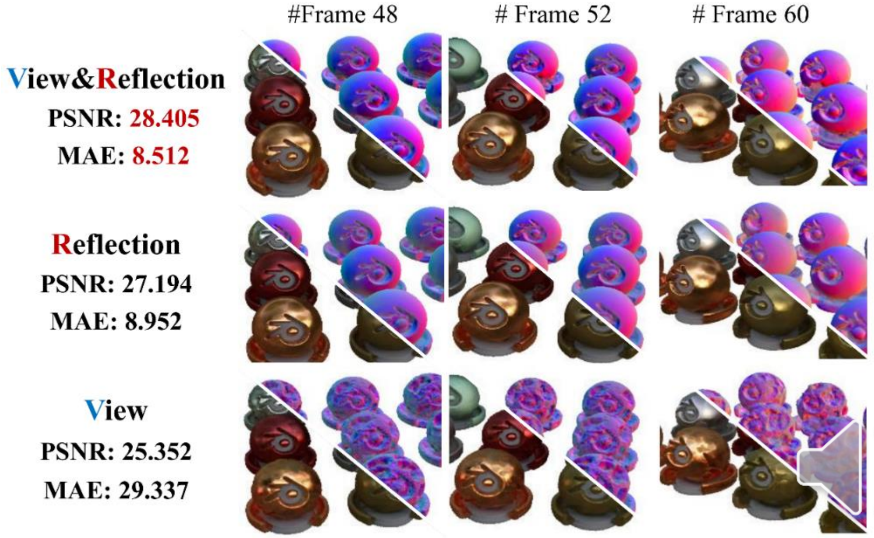
2 Motivation

- Point deviation caused by linear approximation
  - **IDR:** Sphere tracing, linear projection along the ray
  - **DMTet:** Linear interpolation with local deformed vertices
  - Inaccurate gradient backpropagation
  - **CPT:** Projects points exactly on the current surface

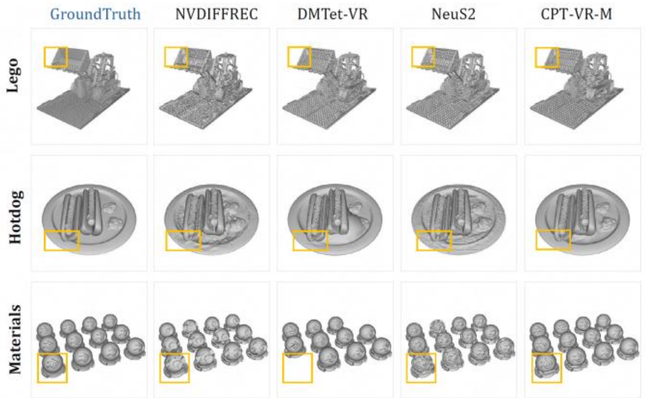


- Only view or reflection rays hardly capture specular regions
  - Only View: Cannot handle specular highlights
  - Only Reflection:
    - Diminish its capability to reconstruct non-specular areas
    - Backpropagation errors due to the deviated points
  - **View-Reflection & CPT**: Handles specular highlights well

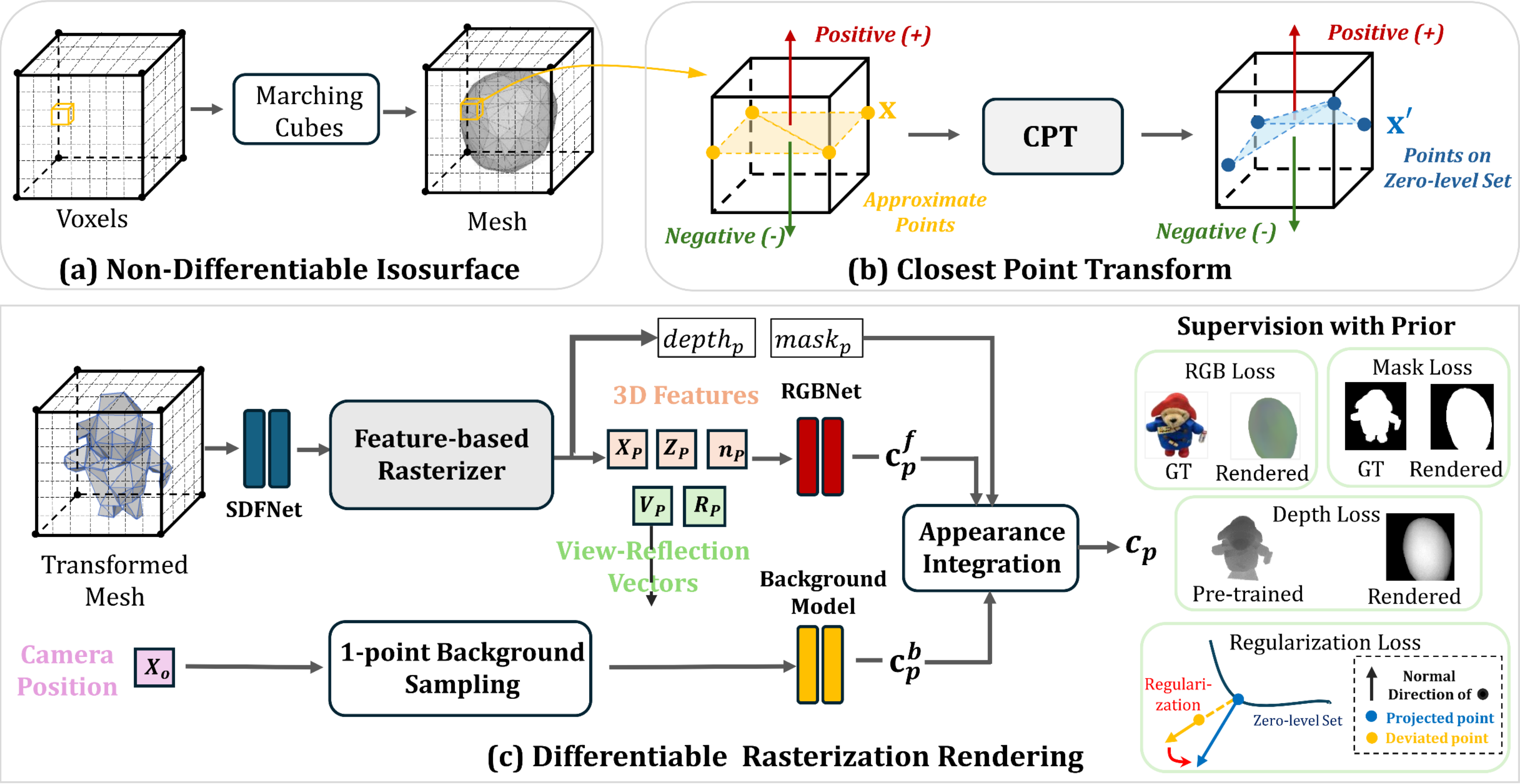
VR gets better normals and RGB



More accurate surface in specular regions or high-frequency regions

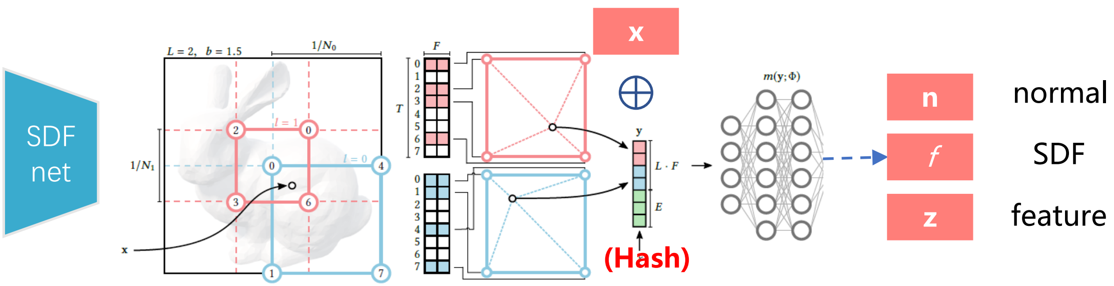


3 Method

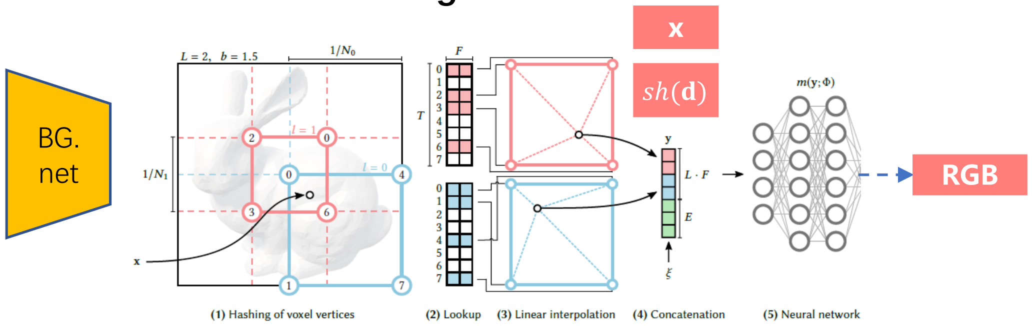


Voxels
Marching Cubes
Mesh
(a) Non-Differentiable Isosurface
Positive (+)
X
Approximate Points
Negative (-)
CPT
Positive (+)
X′
Points on Zero-level Set
Negative (-)
(b) Closest Point Transform
Transformed Mesh
SDFNet
Feature-based Rasterizer
depth_p
mask_p
3D Features
RGBNet
View-Reflection Vectors
Camera Position
1-point Background Sampling
Background Model
Appearance Integration
Supervision with Prior
RGB Loss
GT
Rendered
Mask Loss
GT
Rendered
Depth Loss
Pre-trained
Rendered
Regularization Loss
Regulari-zation
Zero-level Set
Normal Direction of
Projected point
Deviated point
(c) Differentiable Rasterization Rendering

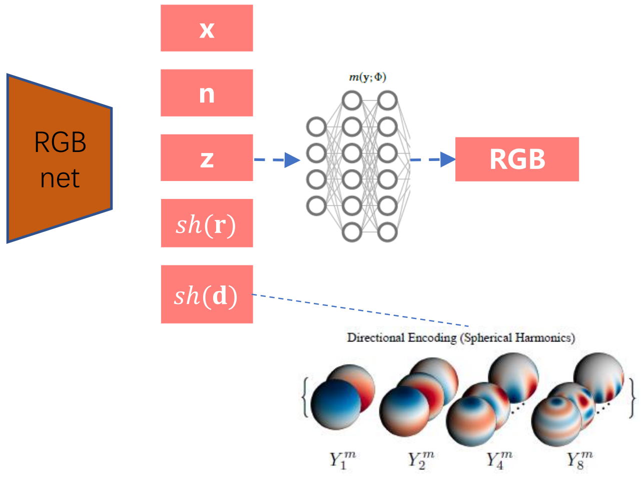
4 Networks

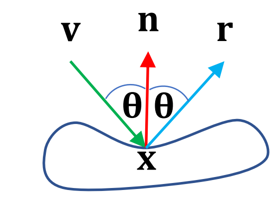


Hashgrid-SDF



Hashgrid-BG

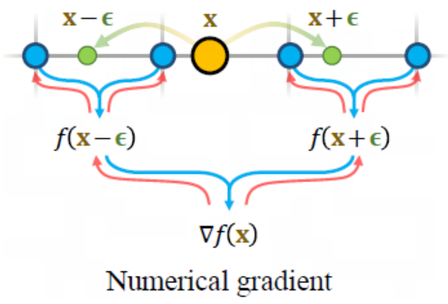




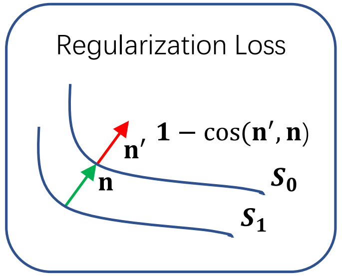
$$\mathbf{r} = 2(-\mathbf{v} \cdot \mathbf{n})\mathbf{n} + \mathbf{v}$$

View-Reflection Spherical Harmonics

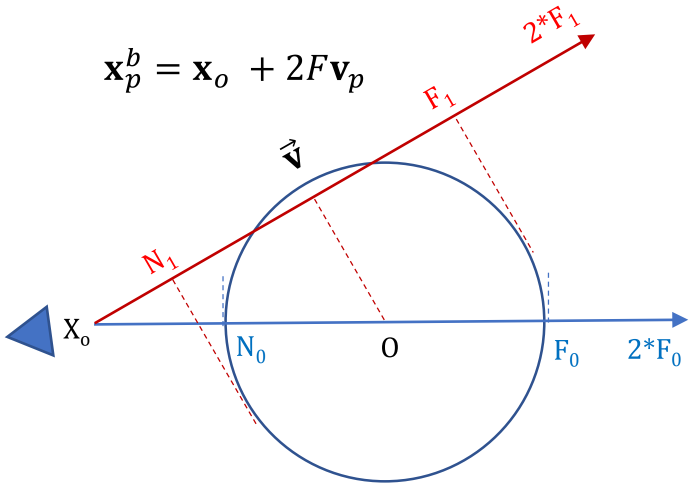
Progressive Numerical Gradient



Normal Consistency



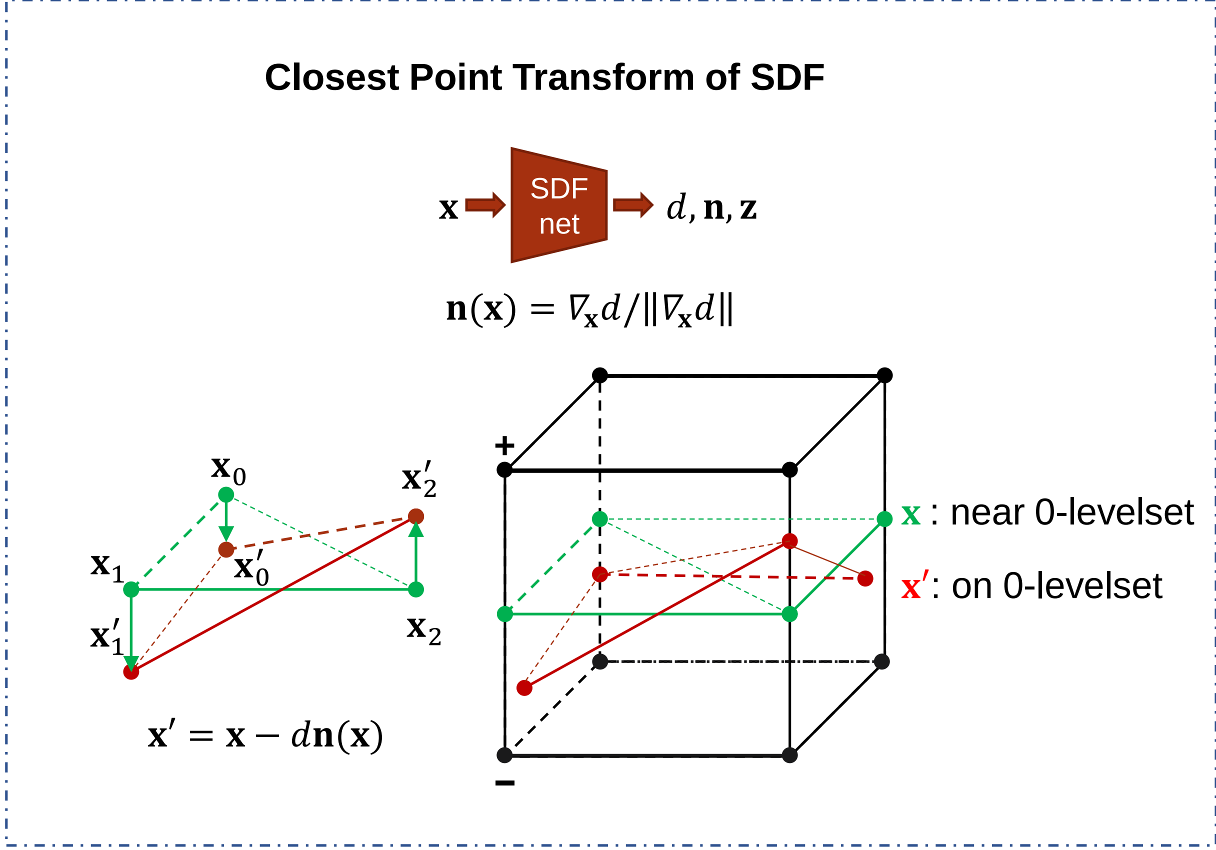
$$\mathbf{x}_p^b = \mathbf{x}_o + 2F\mathbf{v}_p$$

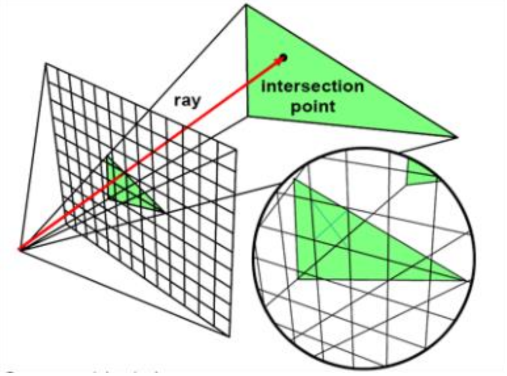


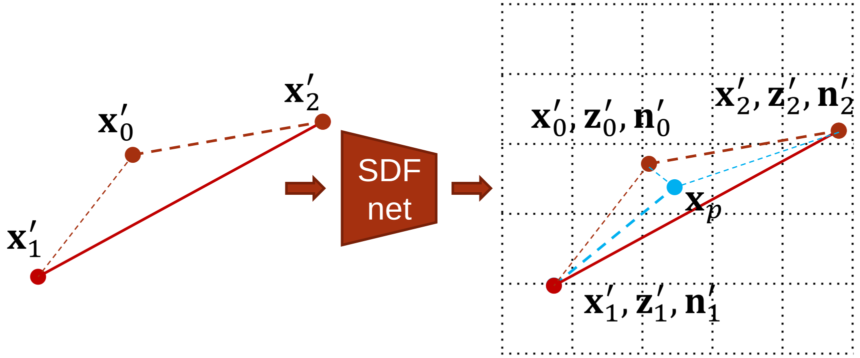
1-Point Background Sampling

5 More Details

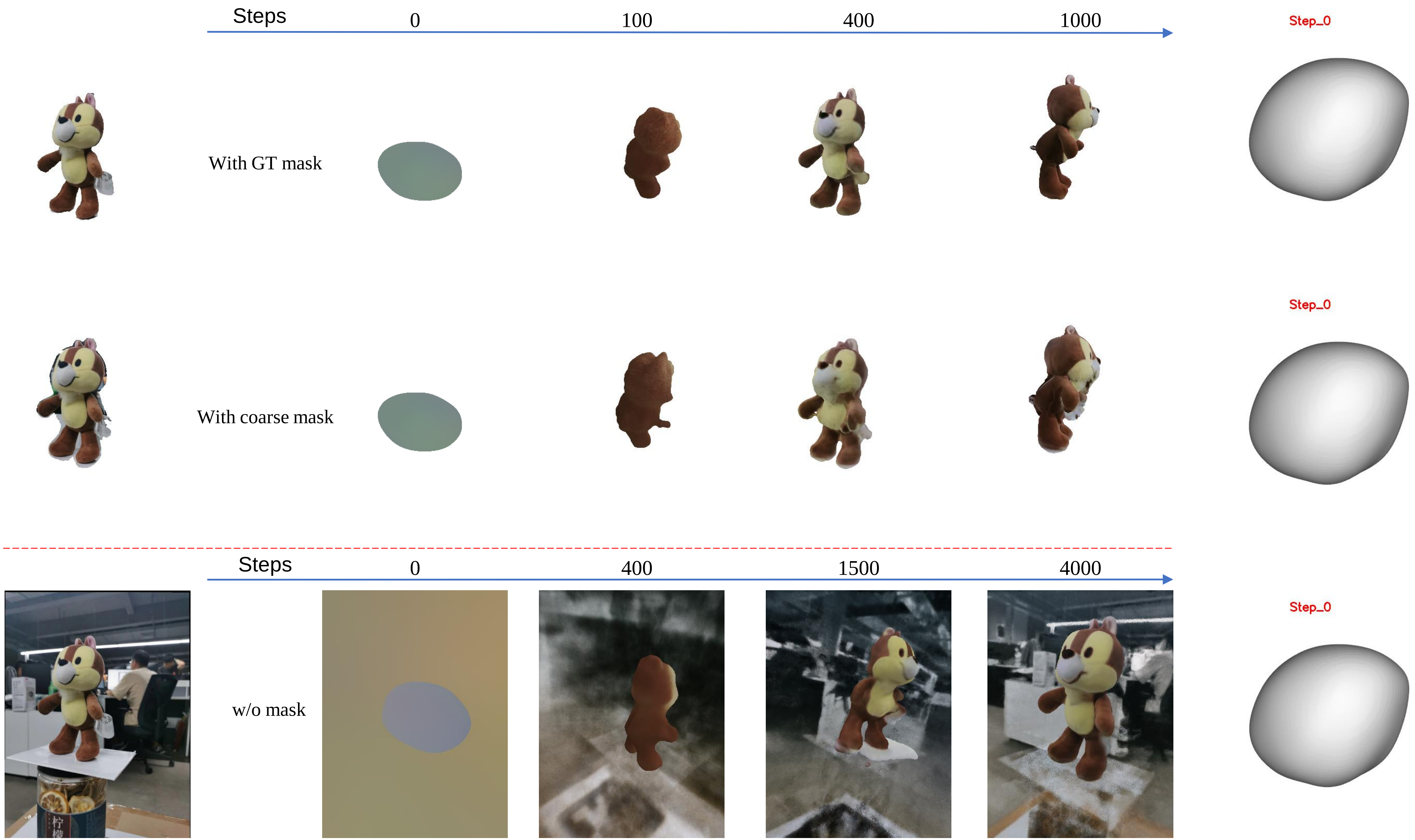
- **Marching Cubes**
  - Sphere-like initialization
  - Extract the mesh at 256^3 or 512^3 per epoch
  - Grid resolution randomly adjusted in [-3,3]
- **Closest Point Transform**
  - Moves the deviated point along the opposite direction of n by its SDF value
  - Run CPT in every iteration
- **Feature-based Rasterisation**
  - Nvdiffrast
  - Barycentric coordinate per view ray
  - Appearance Integration $\mathbf{c}_p = M_p \mathbf{c}_p^f + (1 - M_p)\mathbf{c}_p^b.$







6 Training examples

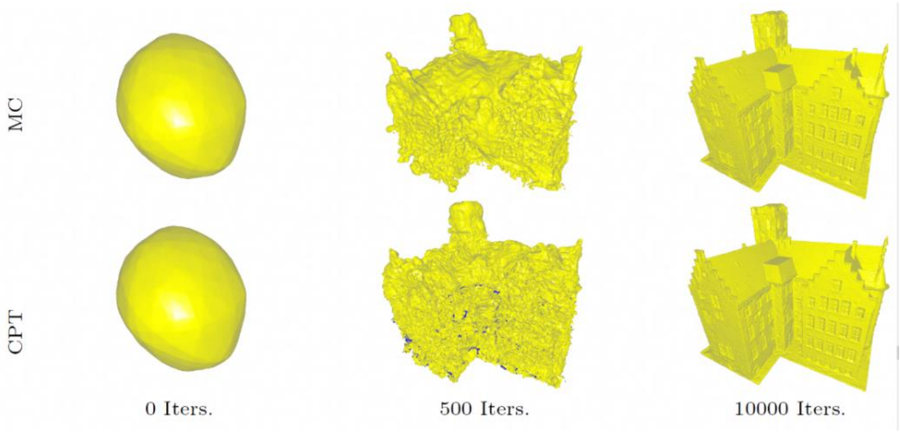


7 Experiments on DTU

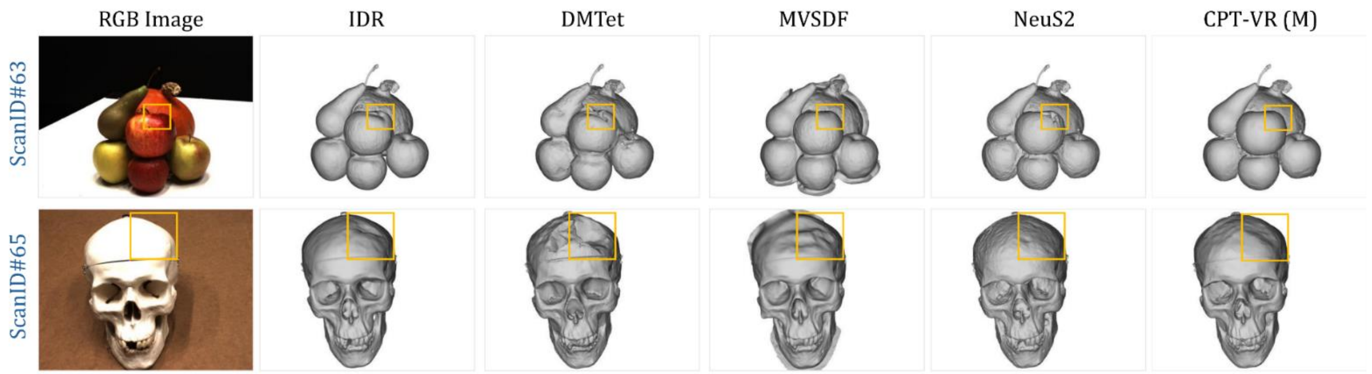
DTU dataset

- Lower Chamfer Distance
- More accurate specular regions
- Stable convergence
  - Some occurrences of overlap and flipping in the early stages
  - Barely visible by the end of the training

Training process



DTU dataset



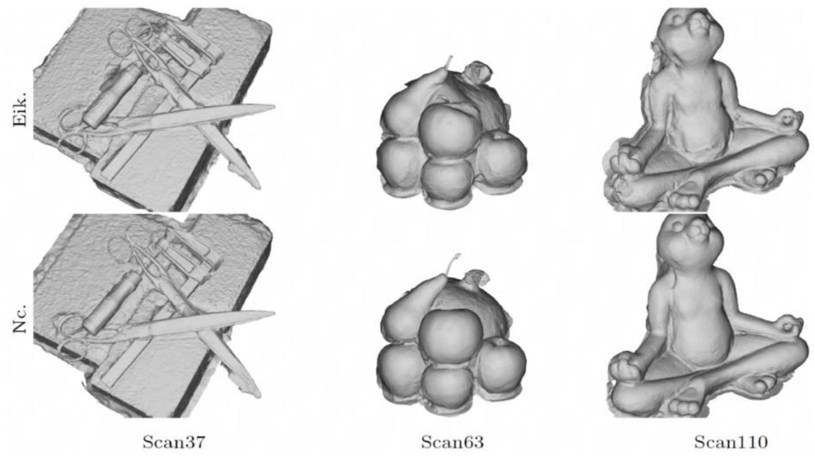
Quantitative results on DTU

ScanID	24	37	40	55	63	65	69	83	97	105	106	110	114	118	122	Mean
DMTet [20]	1.84	1.73	1.05	0.61	1.66	1.37	0.68	1.28	3.82	1.07	0.91	1.15	0.37	0.54	0.55	1.24
IDR [35]	1.63	1.87	0.63	0.48	1.04	0.79	0.77	1.33	1.16	0.76	0.67	0.90	0.42	0.51	0.53	0.90
MVSDF [38]	0.83	1.76	0.88	0.44	1.11	0.90	0.75	1.26	1.02	1.35	0.87	0.84	0.34	0.47	0.46	0.89
FastMESH [40]	0.65	1.48	0.57	0.40	1.48	0.77	0.56	0.86	0.84	0.94	0.72	0.81	0.52	0.49	0.54	0.77
RegSDF [36]	0.59	1.41	0.63	0.42	1.34	0.62	0.59	0.89	0.91	1.02	0.60	0.59	0.29	0.40	0.38	0.71
HF-NeuS [27] (w/ M)	1.11	1.28	0.61	0.47	0.97	0.68	0.62	1.34	0.91	0.73	0.53	1.82	0.38	0.54	0.51	0.83
LOD-NeuS [42] (w/ M)	0.65	0.91	0.37	0.48	1.05	0.87	0.82	1.22	0.95	0.69	0.56	1.30	0.42	0.58	0.57	0.76
NeuS2 [26] (w/ M)	0.56	0.76	0.49	0.37	0.92	0.71	0.76	1.22	1.08	0.63	0.59	0.89	0.40	0.48	0.55	0.70
NeuS2 [26] (w/ BG1)	0.69	0.86	0.83	0.34	1.05	0.68	0.65	1.05	1.12	0.67	0.58	1.41	0.37	0.50	0.52	0.75
Ours (w/ M)	0.55	0.72	0.35	0.38	0.85	0.59	0.54	0.81	0.83	0.67	0.53	0.59	0.32	0.38	0.37	0.56
Ours (w/ BG1)	0.43	0.73	0.36	0.32	0.93	0.61	0.61	0.89	1.02	0.68	0.48	0.73	0.32	0.37	0.37	0.59

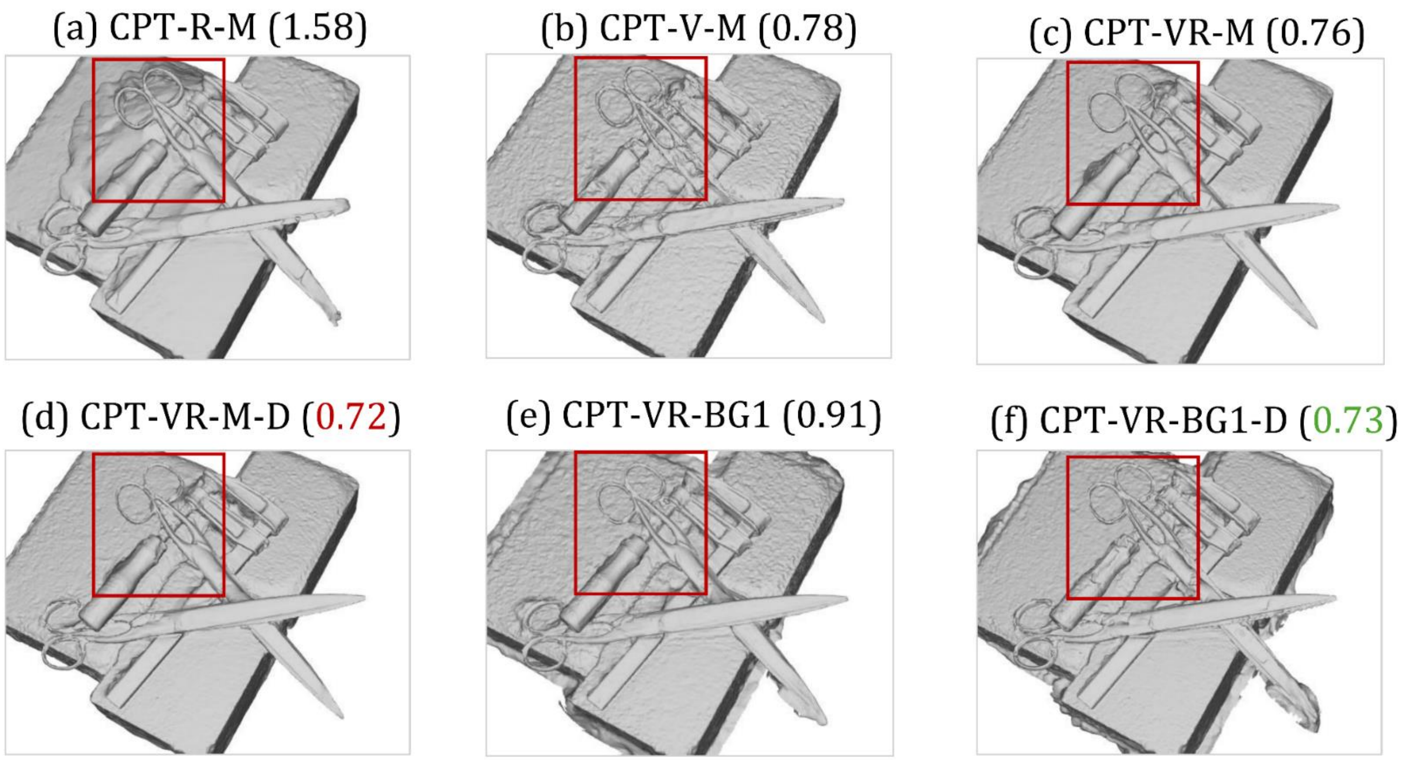
8 Analysis on DTU

- View-Reflection > View> Reflection
- Reflection does not work well for others
- Depth supervision improves the quality
- Normal consistency > Eikonal loss
- MC by 256^3 is efficient
- MC by 512^3 get best shape

Normal consistency V.S. Eikonal loss



Ablation studies on DTU



Impact of view-reflection vectors

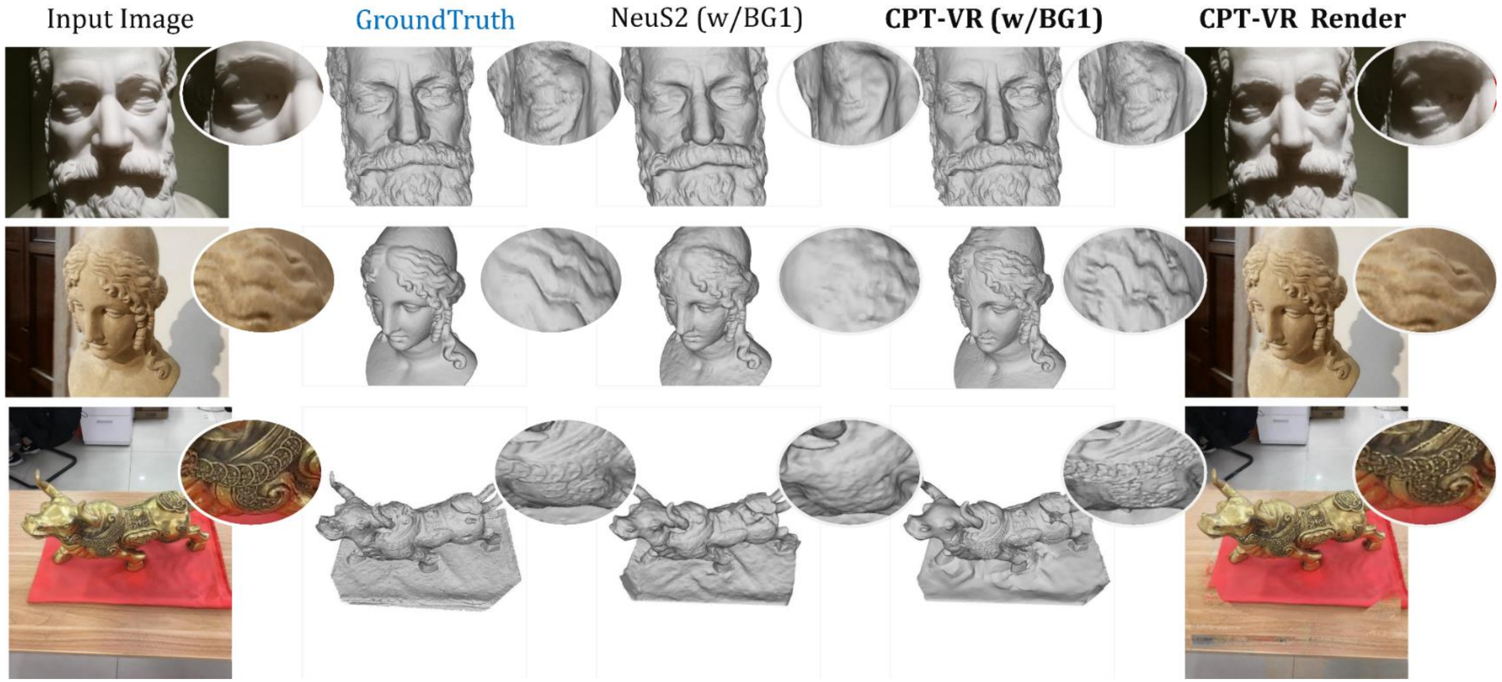
Method	IDR	IDR	DMTet	DMTet	Ours	Ours	Ours
View Vector (V)	✓	✓	✓	✓	✓		✓
Reflection Vector (R)		✓		✓		✓	✓
CD↓	0.90	1.68	1.24	1.63	0.72	0.83	0.65

Efficiency with different settings

Method	Experimetal Settings		Training Time (m) ↓	CD ↓
	Res. of Input Image	Res. of MC		
CPT-VR-M	800 × 600	256	15.3	0.65
CPT-VR-M-D	800 × 600	256	14.4	0.56
	1600 × 1200	512	57.7	0.59
CPT-VR-BG1	800 × 600	256	21.5	0.66
CPT-VR-BG1-D	800 × 600	128	13.7	0.70
	800 × 600	256	23.4	0.63
	1600 × 1200	512	76.1	0.59

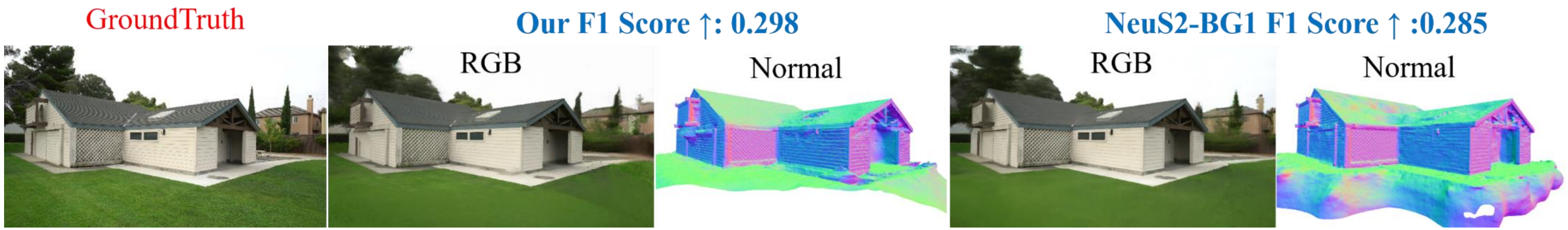
9 More results

Results with complex BG. on BlendedMVS



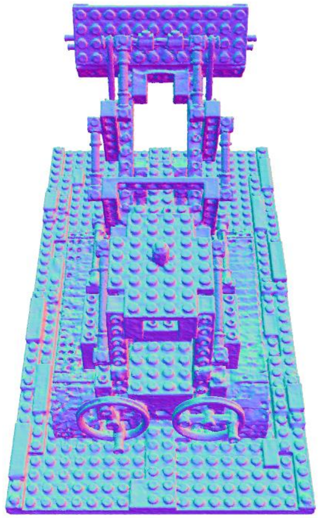
Single-Image-to-3D

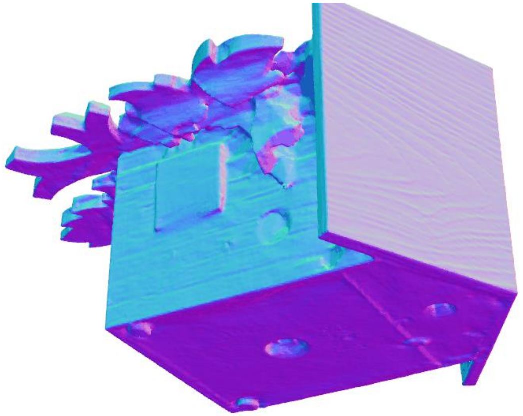
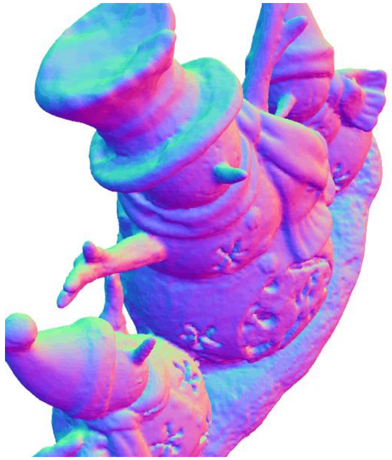
Our method can be adapted for reconstruction in the wild



10 Conclusions

- A novel surface rendering method CPT-VR.
  - Closest Point Transform: Correct the deviated points approximated by linear interpolation
  - View-Reflection & CPT: Roubust to the specular highlights
  - 1-point Background: Independent on any prior BG. knowledge for surface rendering
- Experiments on serveral datasets show our model gets better reconstruction results.
- Extensibility of Single-Image-to-3D Generation



THANK YOU