PET-NeuS: Positional Encoding Tri-Planes for Neural Surfaces

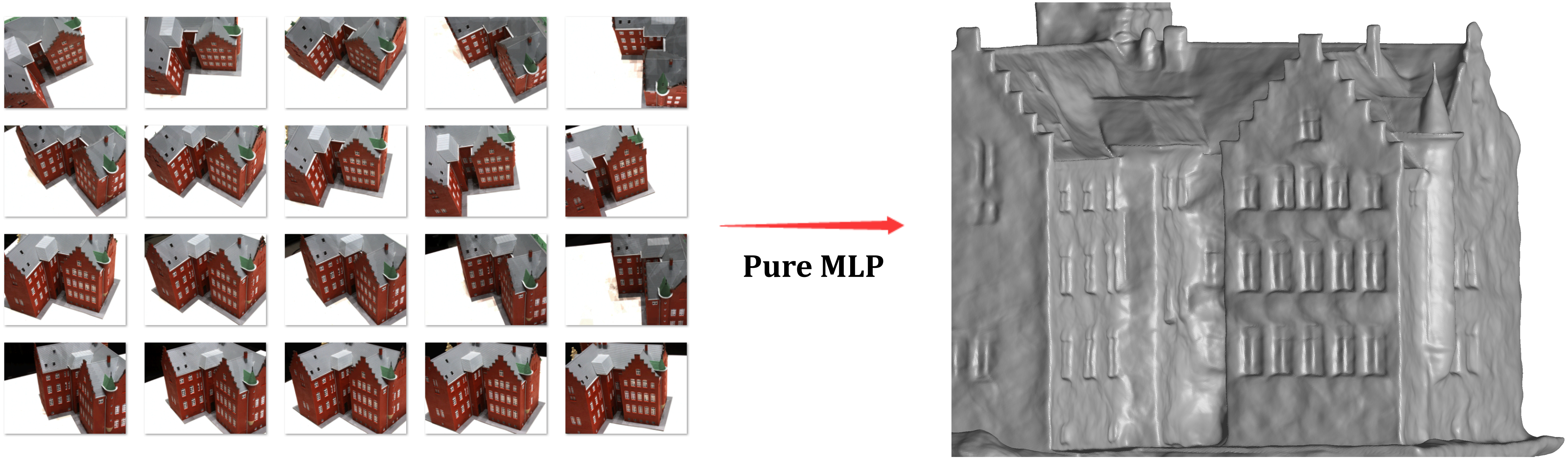
Yiqun Wang, Ivan Skorokhodov, Peter Wonka

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

KAUST
King Abdullah University of Science and Technology

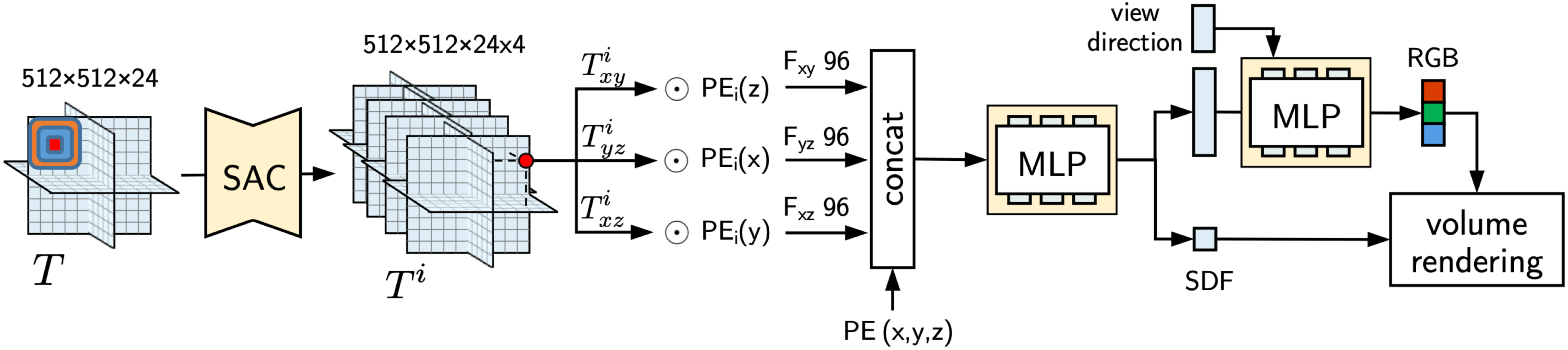
1929
CHONGQING UNIVERSITY

Background



Neural Surface Reconstruction from Multi-View Images

PET-NeuS



Solution:

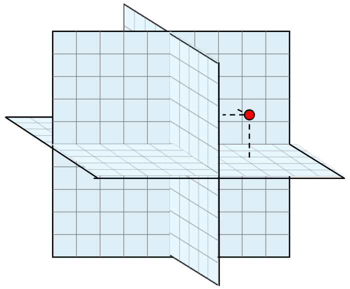
- Integrate tri-plane data structure into neural implicit surface reconstruction framework
- Modulate tri-plane features using positional encoding (PE)
- Utilize multi-scale self-attention convolution to produce tri-plane features

Integrate tri-planes into neural implicit surfaces

Tri-planes:

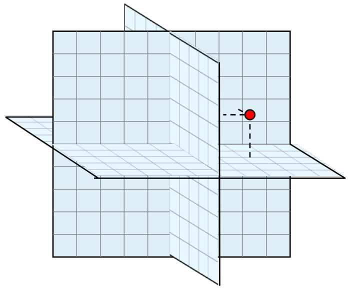
$$T = (T_{xy}, T_{yz}, T_{xz}) \qquad T_* \in \mathbb{R}^{R \times R \times n_f}$$

Tri-plane features at a coordinate:



$$T(x, y, z) = \boldsymbol{w} \in \mathbb{R}^{3n_f} \qquad \boldsymbol{w} = (\boldsymbol{w}_{xy}, \boldsymbol{w}_{yz}, \boldsymbol{w}_{xz})$$

Integrate tri-planes into neural implicit surfaces



Signed Distance Function (SDF):

$$(x, y, z) \in \mathbb{R}^3 \qquad \mapsto \qquad \mathsf{S}(x, y, z) = d_s \in \mathbb{R}$$

Estimate SDF using Tri-planes + MLP:

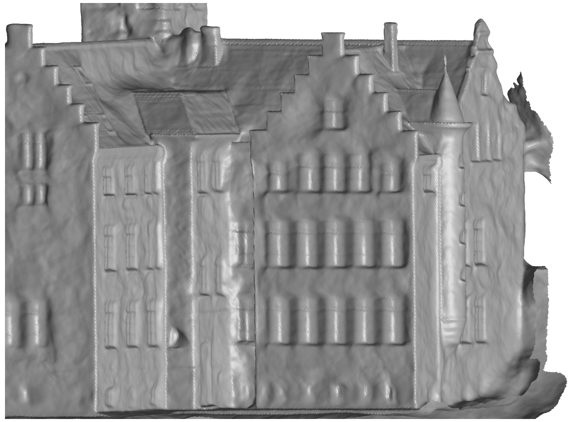
$$(d_s, \boldsymbol{u}) = \mathrm{MLP}_d(\boldsymbol{w}_{xy}, \boldsymbol{w}_{yz}, \boldsymbol{w}_{xz})$$

Density for Volume Rendering:

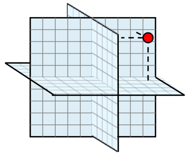
$$\sigma(x, y, z) = s\left(\Psi_s\left(\mathsf{S}(x, y, z)\right) - 1\right) \nabla \mathsf{S}(x, y, z) \cdot \boldsymbol{v}_d$$

Issue

Reference Image



Pure MLP (NeuS)



Tri-planes + MLP

Noise Interference

Modulate tri-planes features using Positional Encoding

3D Fourier series decomposition:

$$f(x, y, z) = \sum_{k=-K}^{K} \sum_{n=-N}^{N} \sum_{m=-M}^{M} a_{mnk} \Theta_m^x \Theta_n^y \Theta_k^z \qquad \Theta_t^x = \begin{cases} \cos(tx) & t > 0 \\ 1 & t = 0 \\ \sin(tx) & t < 0 \end{cases}$$

- Sine and cosine waves from different dimensions x, y, z are entangled with each other through multiplications.

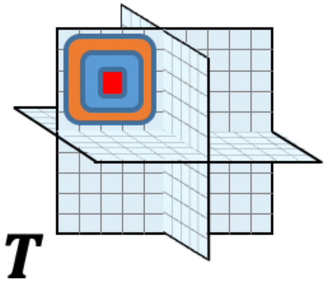
Modulate tri-planes features using Positional Encoding

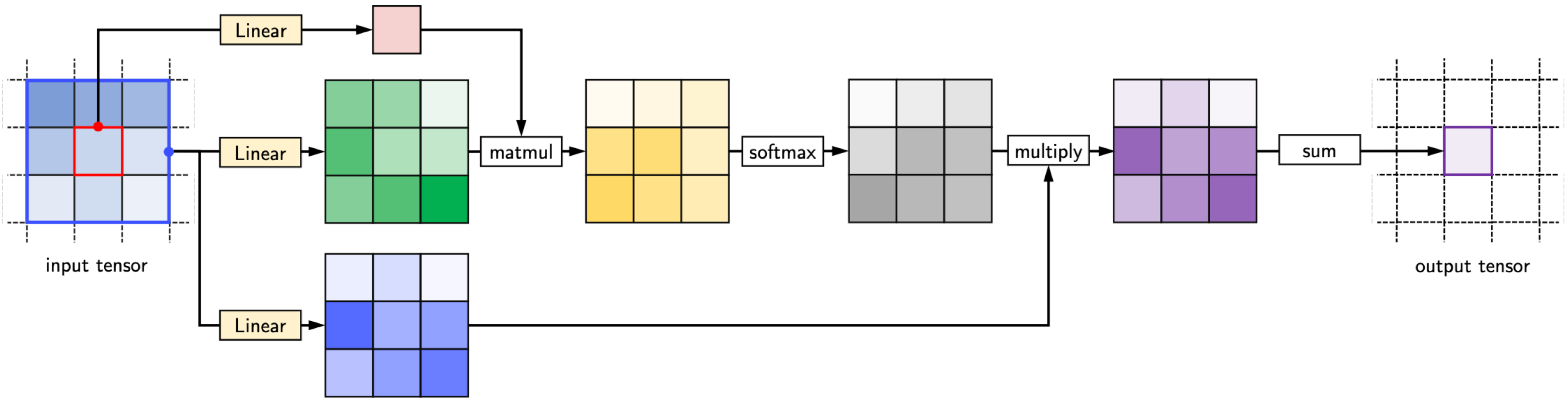
$$f(x,y,z) \approx \mathrm{MLP}\left(\left\{\begin{matrix} \cos(mx) \\ \sin(mx) \\ g_m(y,z)\cos(mx) \\ g'_m(y,z)\sin(mx) \\ \cos(ny) \\ \sin(ny) \\ h_n(x,z)\cos(ny) \\ h'_n(x,z)\sin(ny) \\ \cos(kz) \\ \sin(kz) \\ w_k(x,y)\cos(kz) \\ w'_k(x,y)\sin(kz) \end{matrix}\right\}_{mnk}^{flatten}\right)$$

$$\mathsf{S}(x,y,z) = \mathrm{MLP}\left(\begin{bmatrix} \mathrm{PE}(x,y,z) \\ T_{xy}(x,y) \odot \mathrm{PE}(z) \\ T_{yz}(y,z) \odot \mathrm{PE}(x) \\ T_{xz}(x,z) \odot \mathrm{PE}(y) \end{bmatrix}\right)$$

- We derive $f(x,y,z)$ as a combination of sine and cosine functions and approximate implicit function with an MLP with sine/cosine inputs
- This approximation is directly applicable to our SDF function by utilizing an MLP that takes tri-plane features modulated by positional encoding as inputs

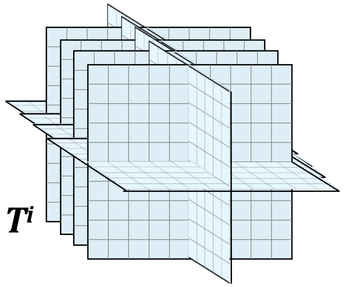
Utilize multi-scale self-attention convolution





- Perform convolution with different window sizes in the spatial domain

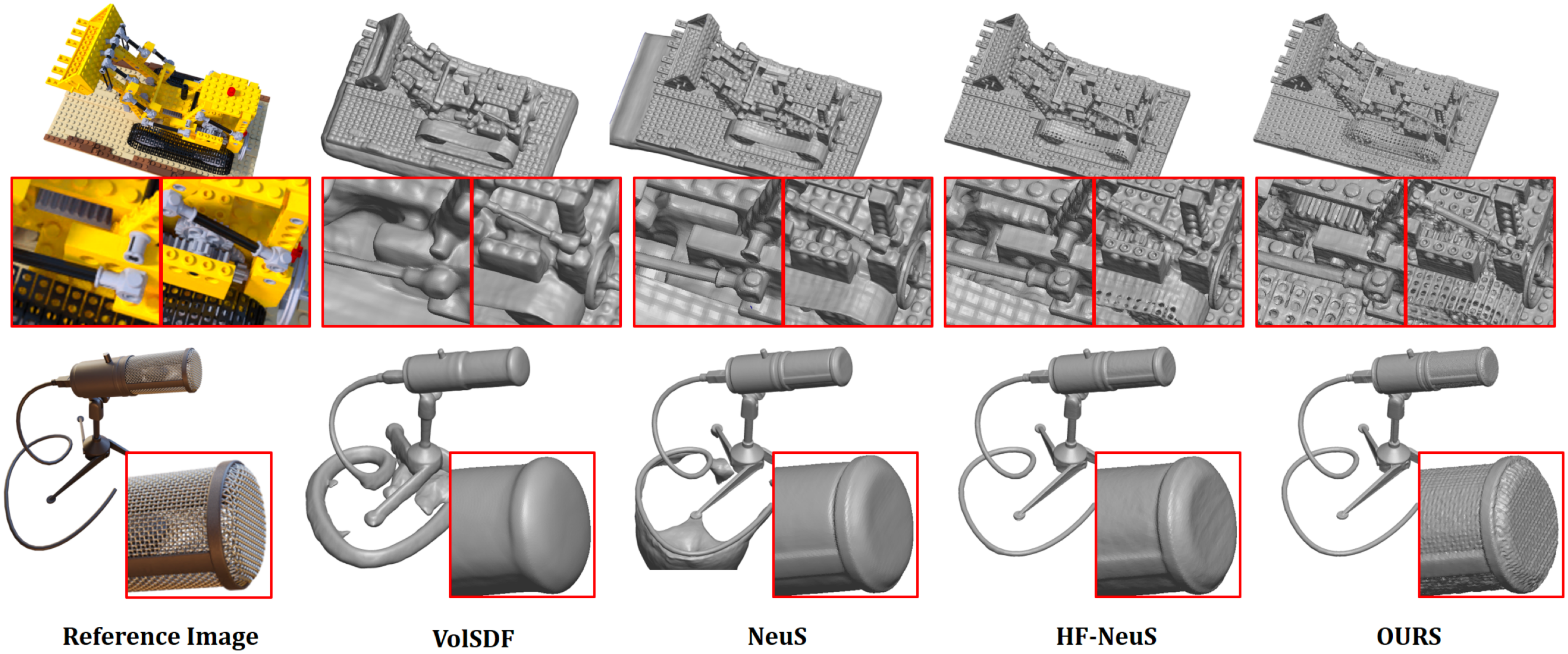
Utilize multi-scale self-attention convolution



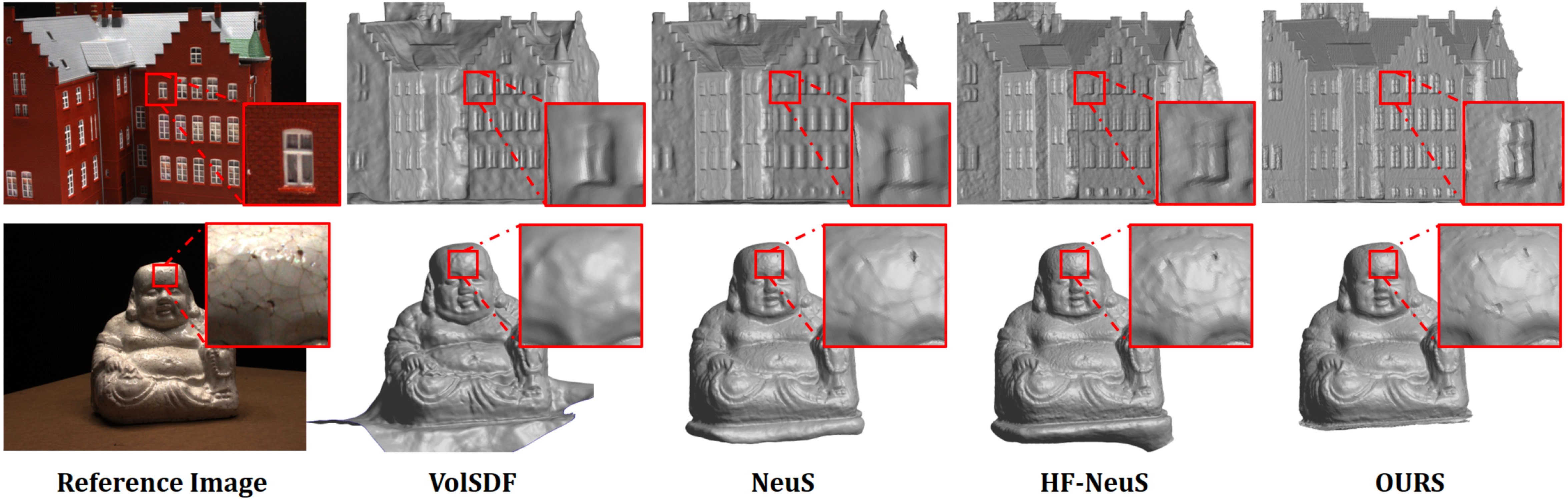
$$T = \text{concat}\left[T^i\right]_{i=0}^{3}, \quad T^i = \left\{T^i_{xy}, T^i_{yz}, T^i_{xz}\right\}$$

- We take the output tri-plane features produced by the i^{th} self-attention convolution (SAC), and concatenate them all together with the original tri-plane features to form the final tri-plane representation for different frequency bands.

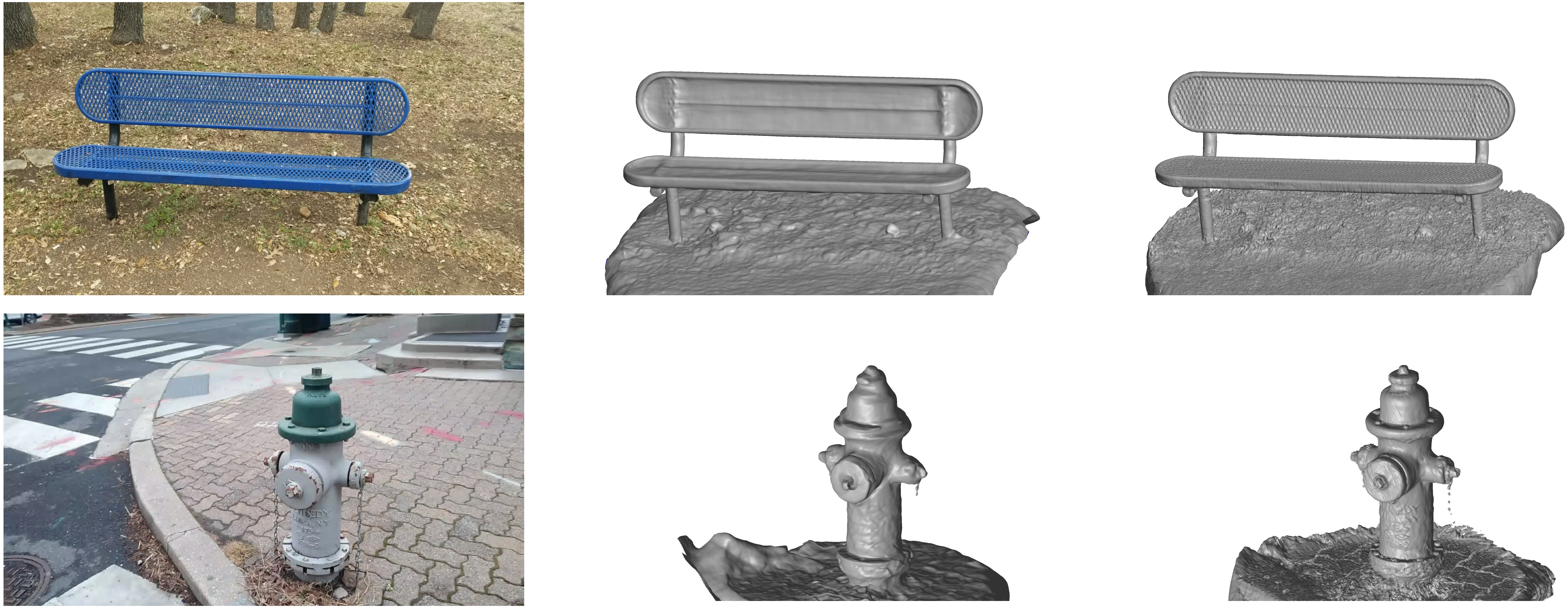
Results



Results

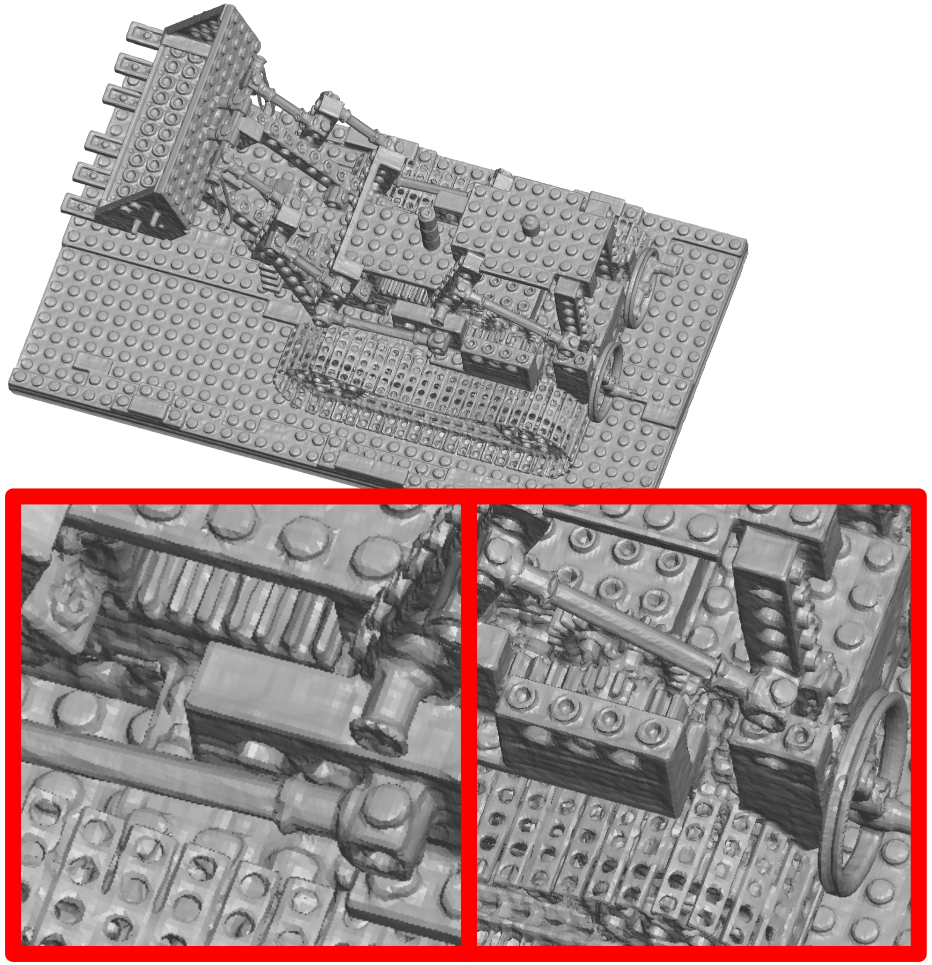


Results



Reference Image NeuS OURS

Thanks for listening



Project:

https://github.com/yiqun-wang/PET-NeuS