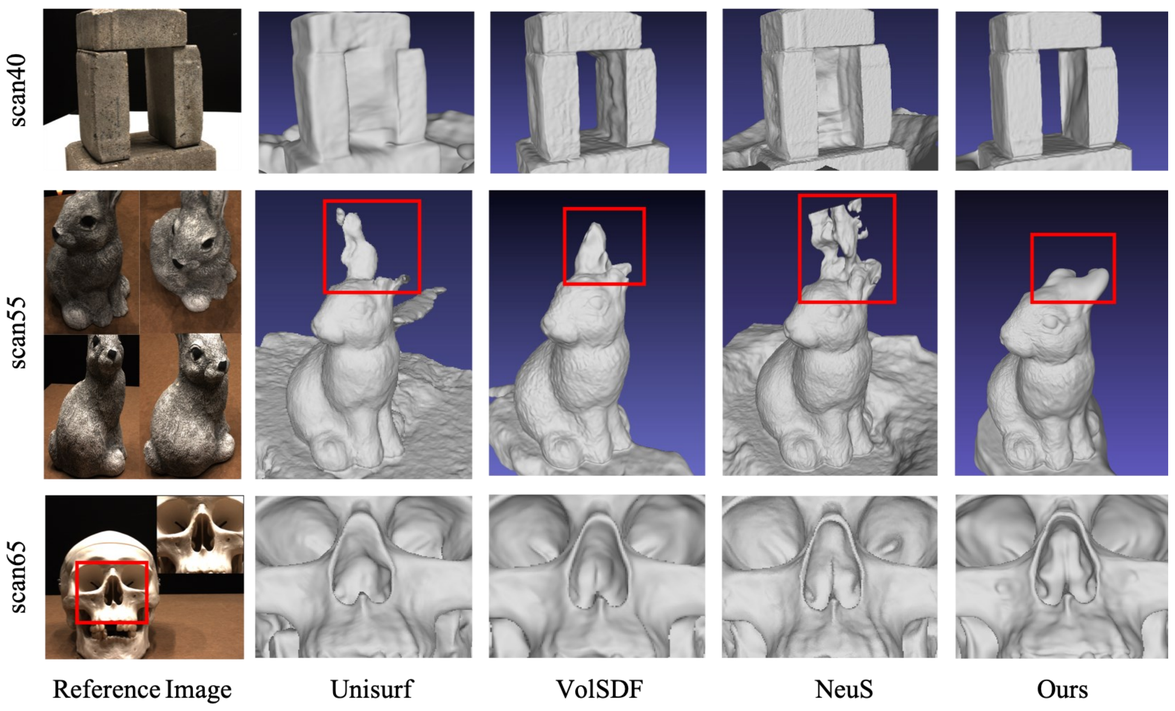
Efficient Implicit SDF and Color Reconstruction via Shared Feature Field

Shuangkang Fang[1*], Dacheng Qi[1*], Weixin Xu[2], Yufeng Wang[1],
Zehao Zhang[1], Xiaorong Zhang[1] , Huayu Zhang[1], Zeqi Shao[1], Wenrui Ding[1]
[1]Beihang University & [2]Megvii Inc. * Indicates equal contribution

Motivation

Scenes with **complex topologies** or **textureless regions** lack **pixel-wise masks** for supervision, making full scene recovery challenging. Bridging this gap is essential for robust scene understanding.



* Comparison with standard methods using approximately half of the training images from the DTU MVS dataset.

Our Contributions:

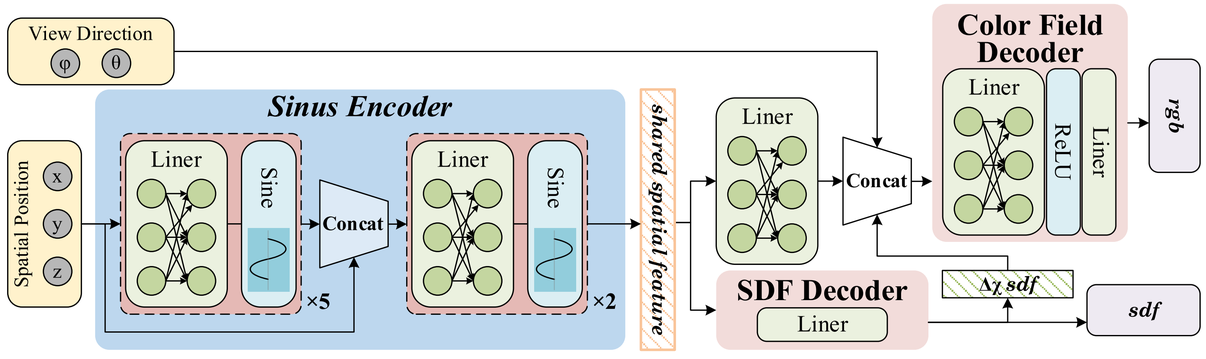
- We propose an end-to-end network tailored for **textureless scenes** and **sparse input** reconstruction.
- We design a **Sinus-ReLU activation function** to reduce artifacts and improve synthesis quality.

Shared Feature Field

Our SFF combines the multiple MLPs paradigm with an **Encoder-Decoder paradigm** and a **mixed Sinus-ReLU activation function paradigm**.

Encoder-Decoder paradigm

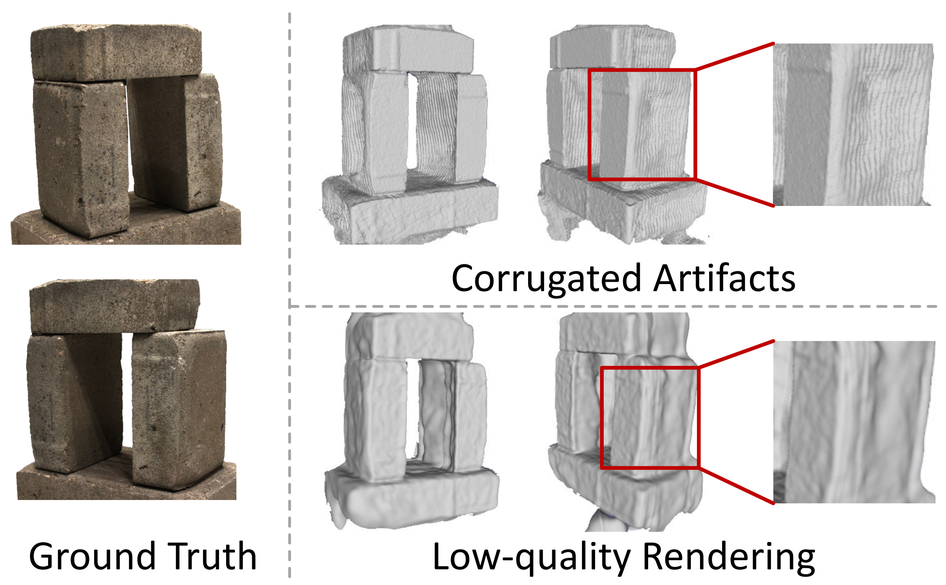
Traditional methods use separate MLPs for encoding and decoding, causing positional ambiguity and poor geometric consistency. We propose a **weight-shared encoder-decoder**, generating **shared spatial features** in a single pass to improve regularization and accuracy.



Compared to the common paradigm, where the color field is conditioned on $[x, v, \nabla_x sdf, F_{sdf}]$, we allow the SDF and color field to share the **same feature** $\boldsymbol{F}$ from the encoder, while obtaining the spatial position through an additional forward pass.

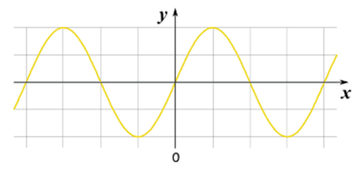
Mixed Sinus-ReLU Activation Function Paradigm

We observe that high-frequency positional encoding of spatial positions introduces **corrugated artifacts** on the mesh, while low-frequency positional encoding leads to **lower-quality novel view rendering**.



To improve view quality and reduce artifacts, we replace positional encoding with a **mixed Sinus-ReLU activation**, using sinus for the encoder and ReLU for the decoders.

In the Sinus activation function, ω_0 is a **trainable parameter** initialized to ***3***. Network weights are initialized using a **uniform distribution**:



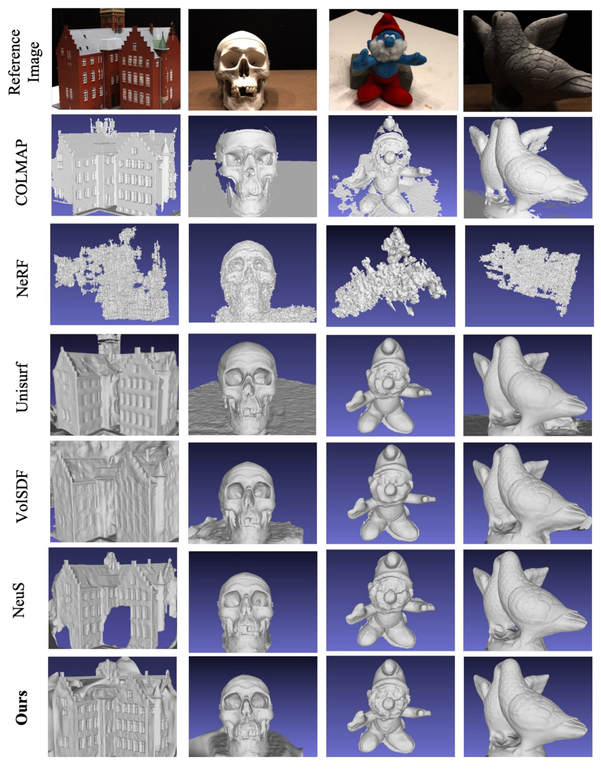
$$Sinus(\mathbf{x}) = \sin(\omega_0 \cdot \mathbf{x})$$

$$W \sim U\left(-\sqrt{\frac{6}{ic}}/\omega_0\,, \sqrt{\frac{6}{ic}}/\omega_0\right),$$

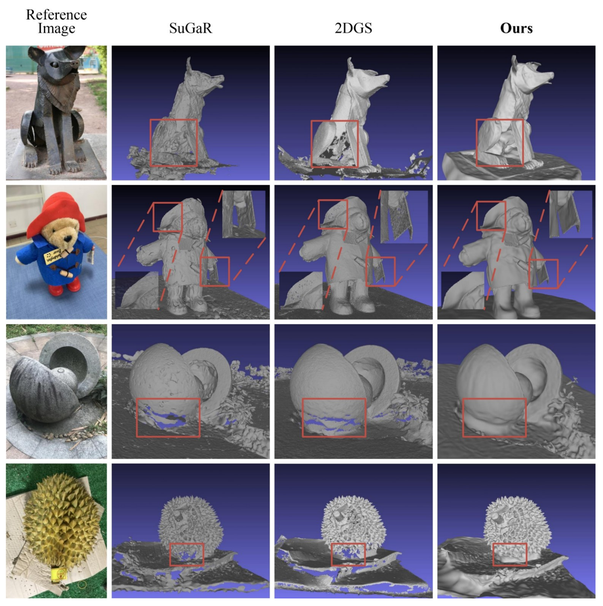
except for the first layer, which is initialized as :

$$W \sim U\left(-\sqrt{\frac{1}{ic}}, \sqrt{\frac{1}{ic}}\right)$$

Experiments



DTU Dataset



BlendedMVS Dataset