EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

# Geometrically Accurate Feed-Forward Gaussian Splatting for Unbounded Scenes

Master Thesis Defense

**César Díaz Blanco**

| | |
|---|---|
| Supervisor: | Zehao Yu<br>*Autonomous Vision Group, University of Tübingen* |
| Second supervisor: | Haofei Xu<br>*Autonomous Vision Group, University of Tübingen* |
| First examiner: | Prof. Dr.-Ing. Andreas Geiger<br>*Autonomous Vision Group, University of Tübingen* |
| Second examiner: | Prof. Dr. Gerard Pons-Moll<br>*Real Virtual Humans, University of Tübingen* |


*Tübingen, 27.04.2026*

# 1

# Introduction

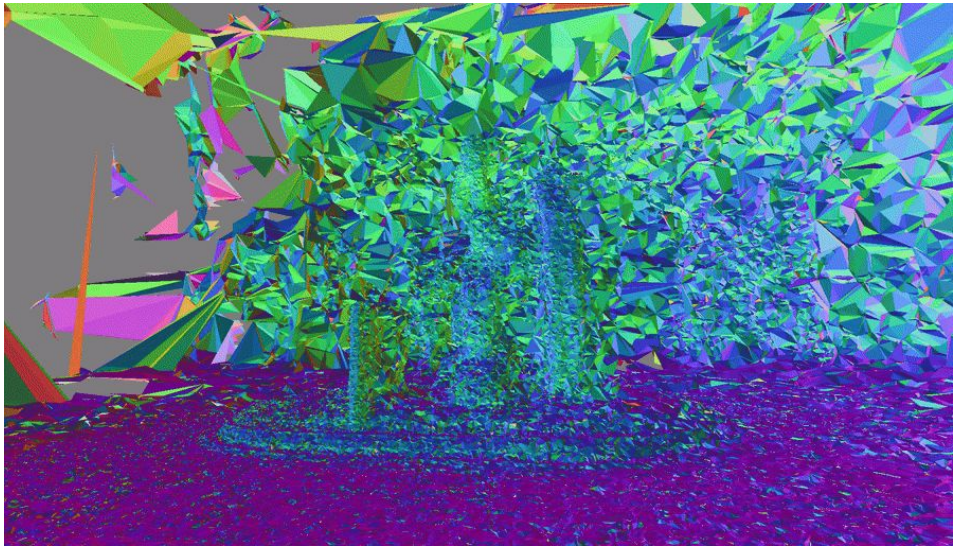
# Feed-Forward Gaussian Splatting

Feed-Forward Gaussian Splatting models take in multi-view images and output a Gaussian Splatting representation

~0.4s

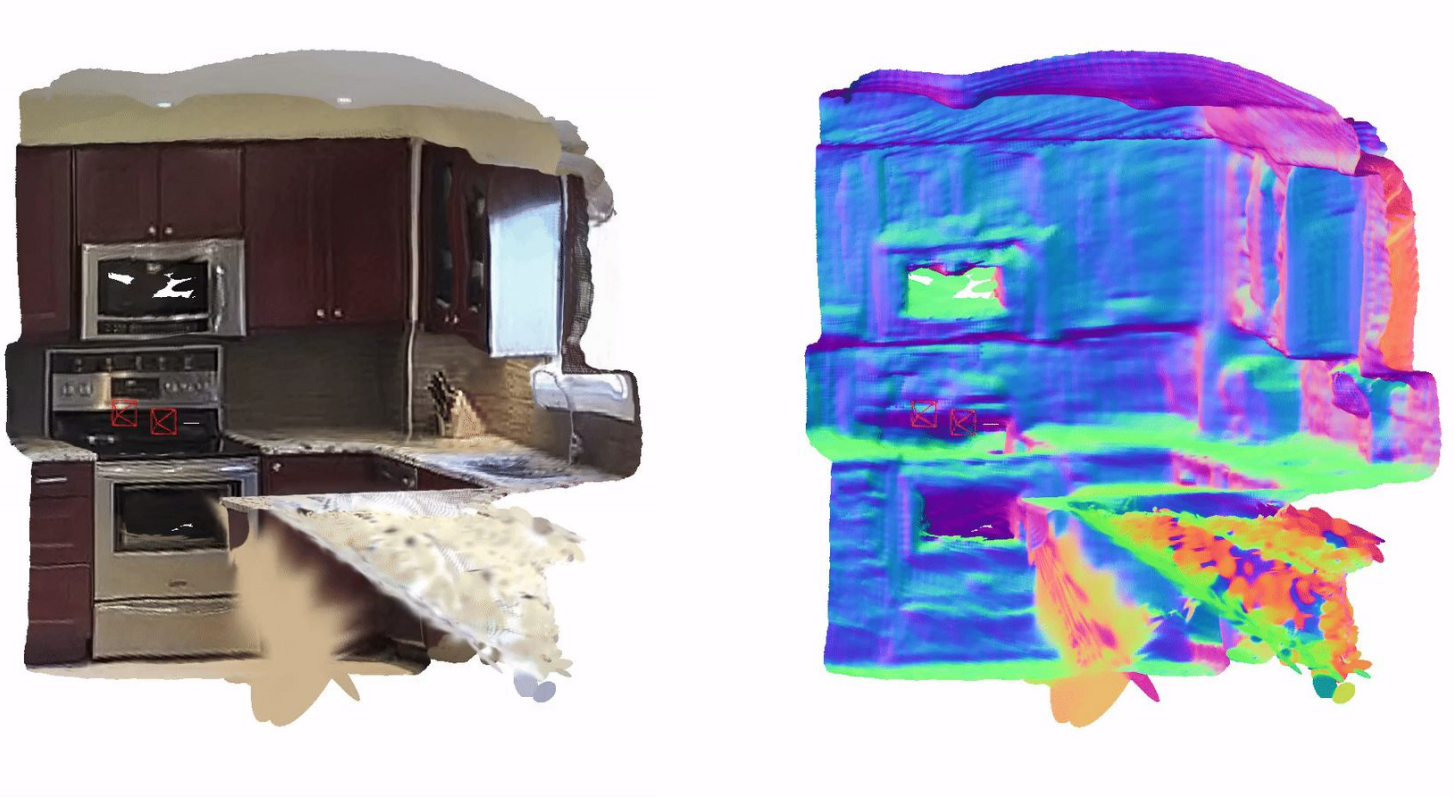
# Motivation

Feed-Forward Gaussian Splatting models achieve great visual quality but fail to achieve geometric quality, especially in the unbounded scene case

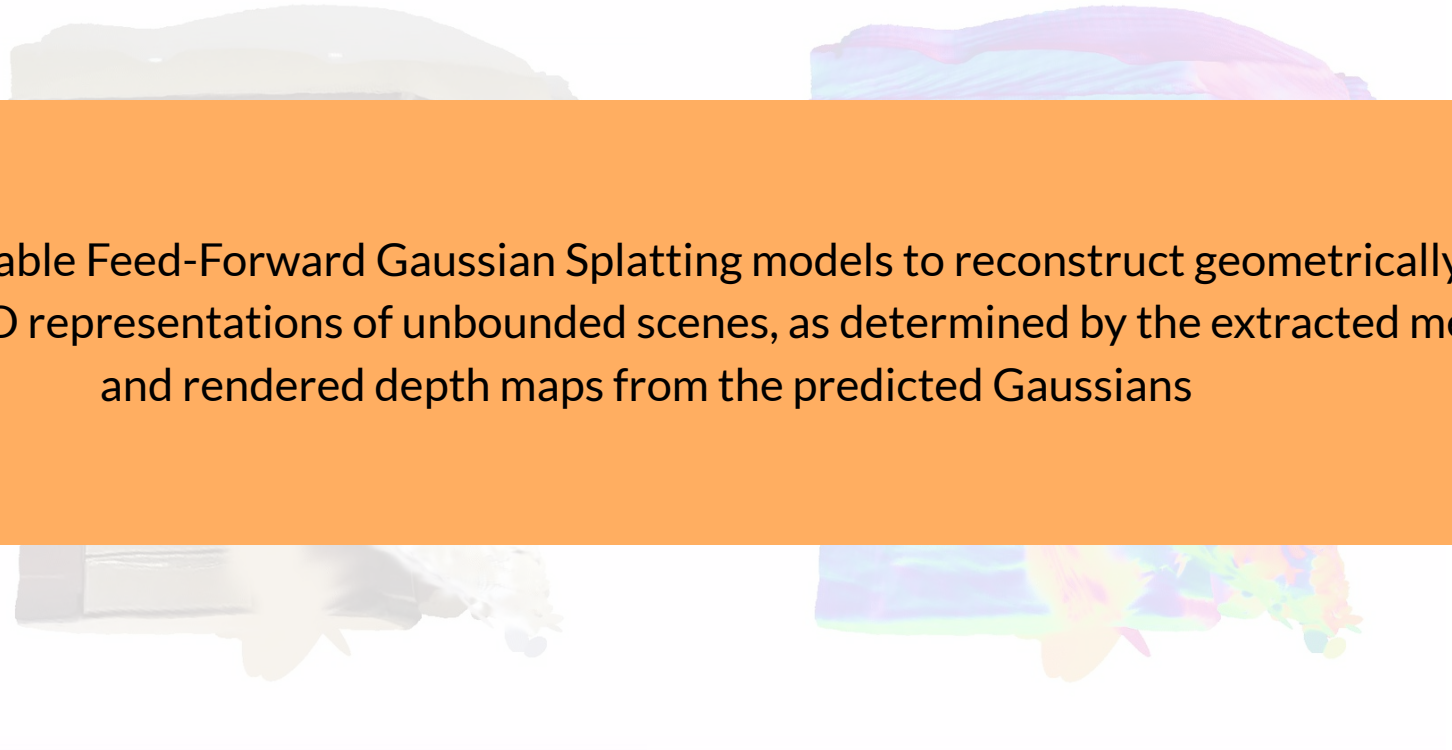


Xu et al.: ReSplat: Learning Recurrent Gaussian Splatting; arXiv 2026

# Motivation



# Motivation



Goal: Enable Feed-Forward Gaussian Splatting models to reconstruct geometrically accurate 3D representations of unbounded scenes, as determined by the extracted mesh and rendered depth maps from the predicted Gaussians

# Index


1. **Introduction**
   - **Data:** Unbounded scenes
   - **Model**
2. **Related Work**
   - Gaussian Splatting Renderers
3. **Method**
   - Our additions to enable geometrically accurate 3D representations
4. **Evaluation**
   - Metrics
   - Extracted mesh and rendered depth maps
   - Findings

# Datasets: Training

## DL3DV

- 10,510 multi-view scenes
- Around half of the scenes are unbounded
- Does not provide complete GT depth

## Depth Anything 3 massive dataset

- 160k environmental scenes
- 5% are outdoor
- 20% come from mixed datasets: indoor or outdoor
- GT depth comes from SfM methods thus incomplete

Ling et al.: DL3DV-10K: A Large-Scale Scene Dataset for Deep Learning-based 3D Vision; CVPR 2024
Lin et al.: Depth Anything 3: Recovering the Visual Space from Any Views; arXiv 2025

# Datasets: Training

DL3DV

- 10,510 multi-view scenes

Main Hypothesis: Feed-Forward Gaussian Splatting models can achieve accurate 3D representations with photometric and self-supervised geometric losses

- 20% come from mixed datasets: indoor or outdoor
- GT depth comes from SfM methods thus incomplete

Ling et al.: DL3DV-10K: A Large-Scale Scene Dataset for Deep Learning-based 3D Vision; CVPR 2024
Lin et al.: Depth Anything 3: Recovering the Visual Space from Any Views; arXiv 2025

# Framework: Training with NVS loss

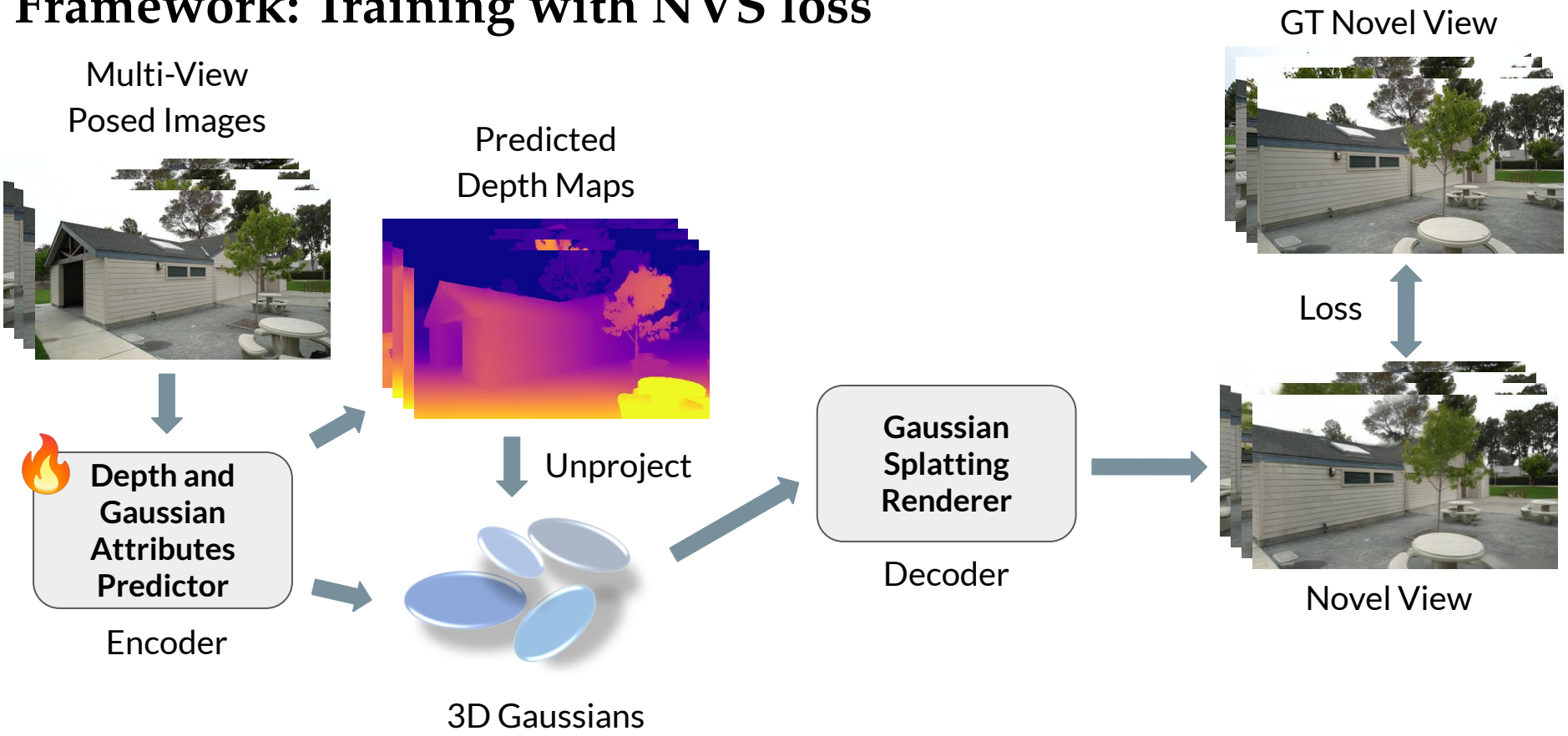



Illustration adapted from: Xu et al.: DepthSplat: Connecting Gaussian Splatting and Depth; CVPR 2025

# Framework: Mesh Extraction

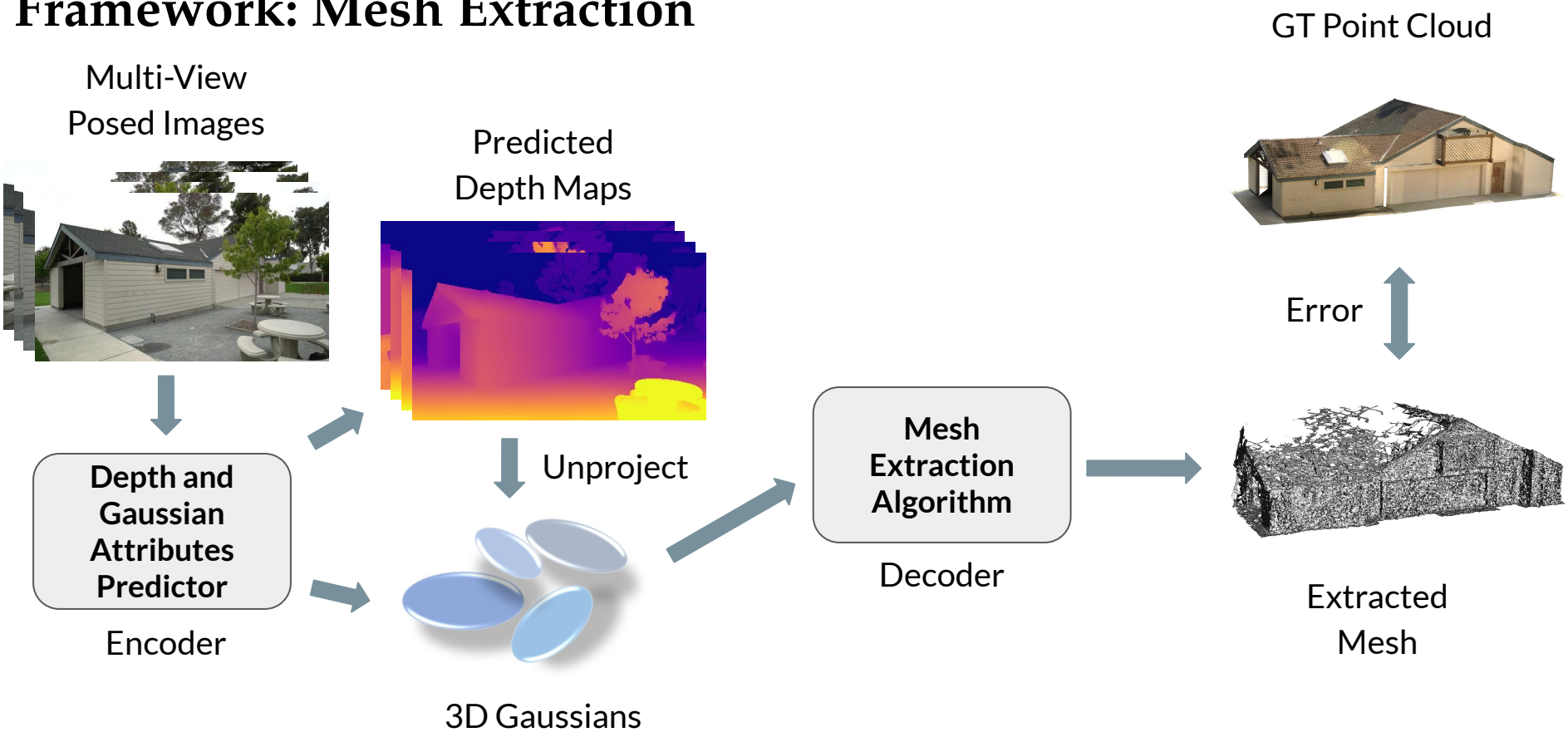



Illustration adapted from: Xu et al.: DepthSplat: Connecting Gaussian Splatting and Depth; CVPR 2025

# Framework: Depth Estimation

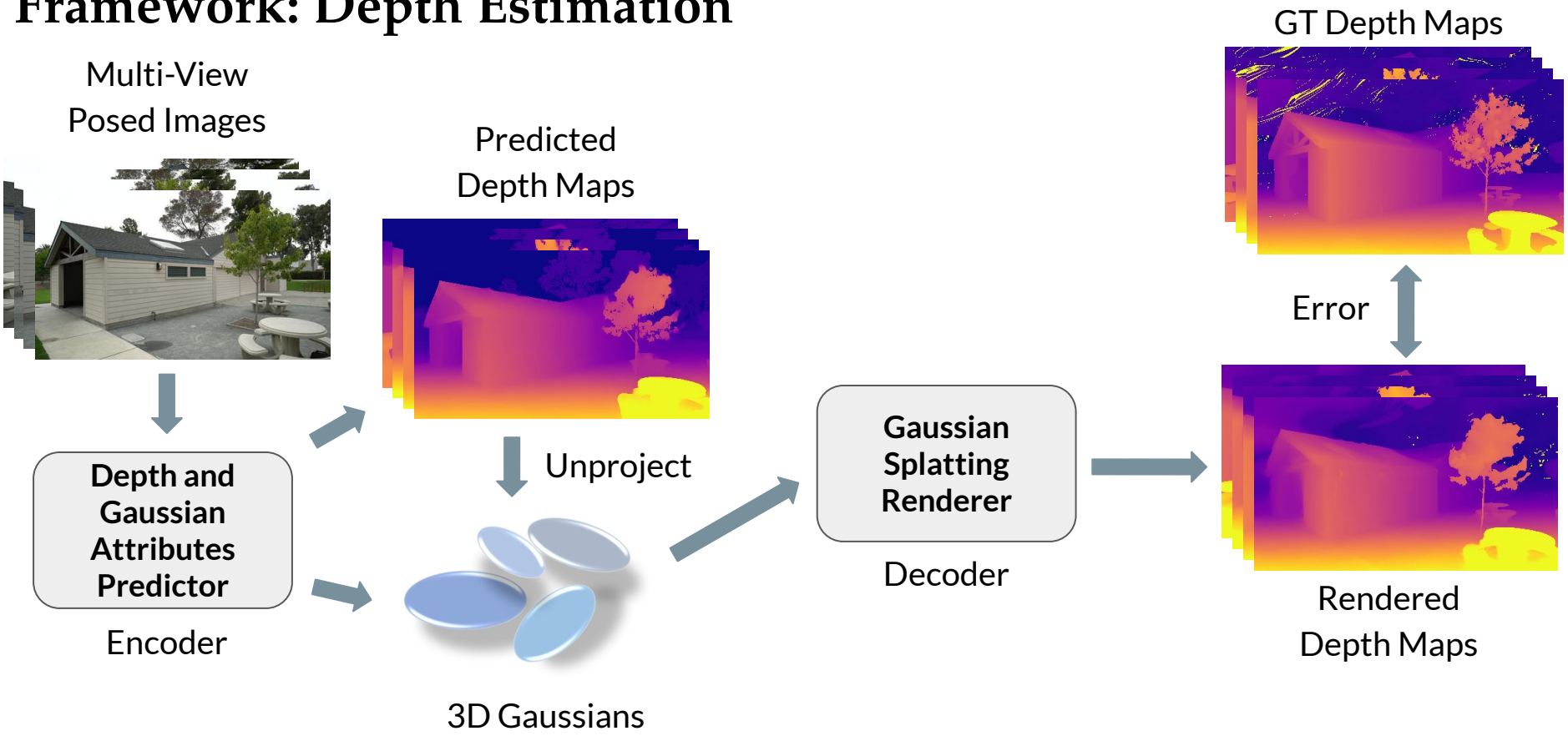



Illustration adapted from: Xu et al.: DepthSplat: Connecting Gaussian Splatting and Depth; CVPR 2025

# Framework: Training with NVS loss

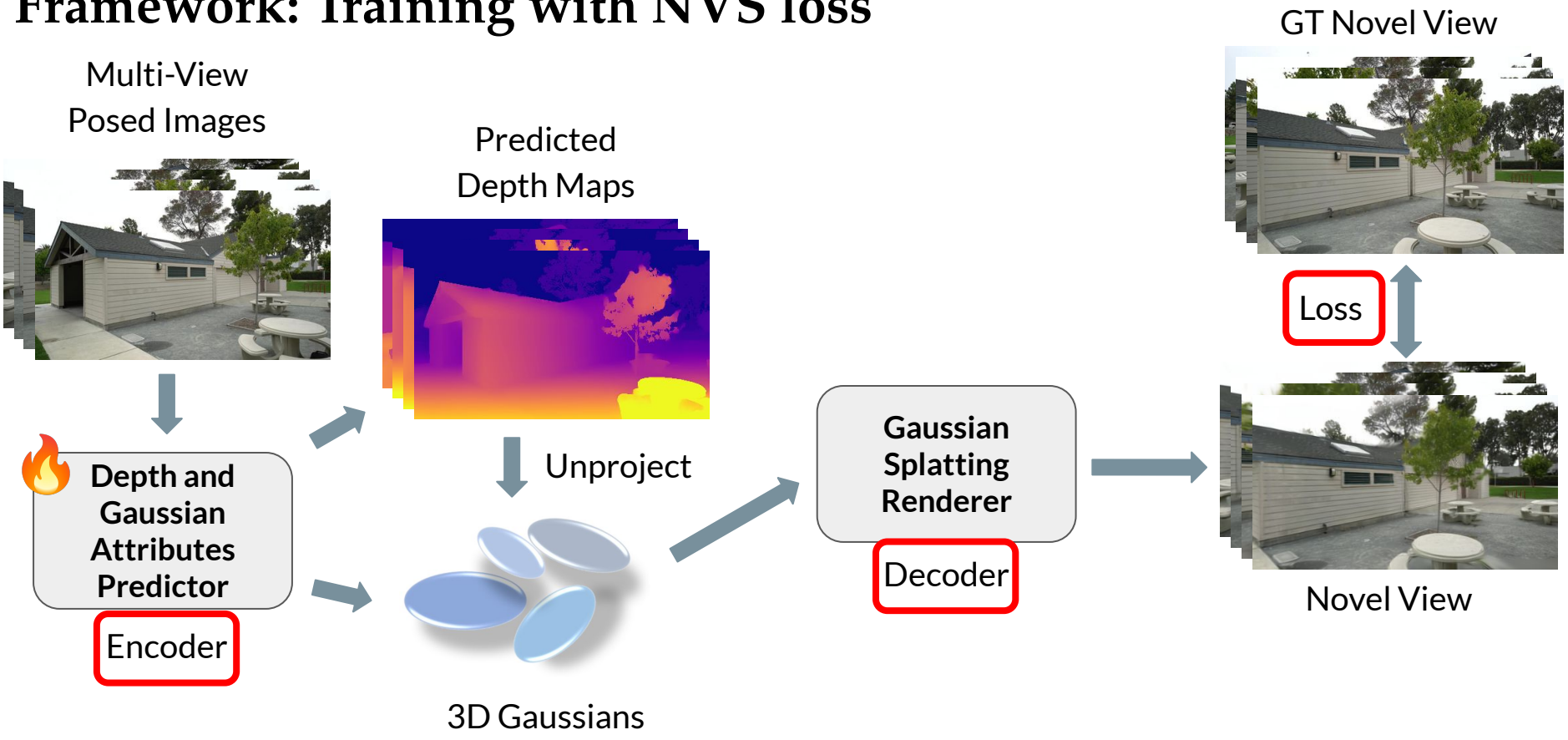
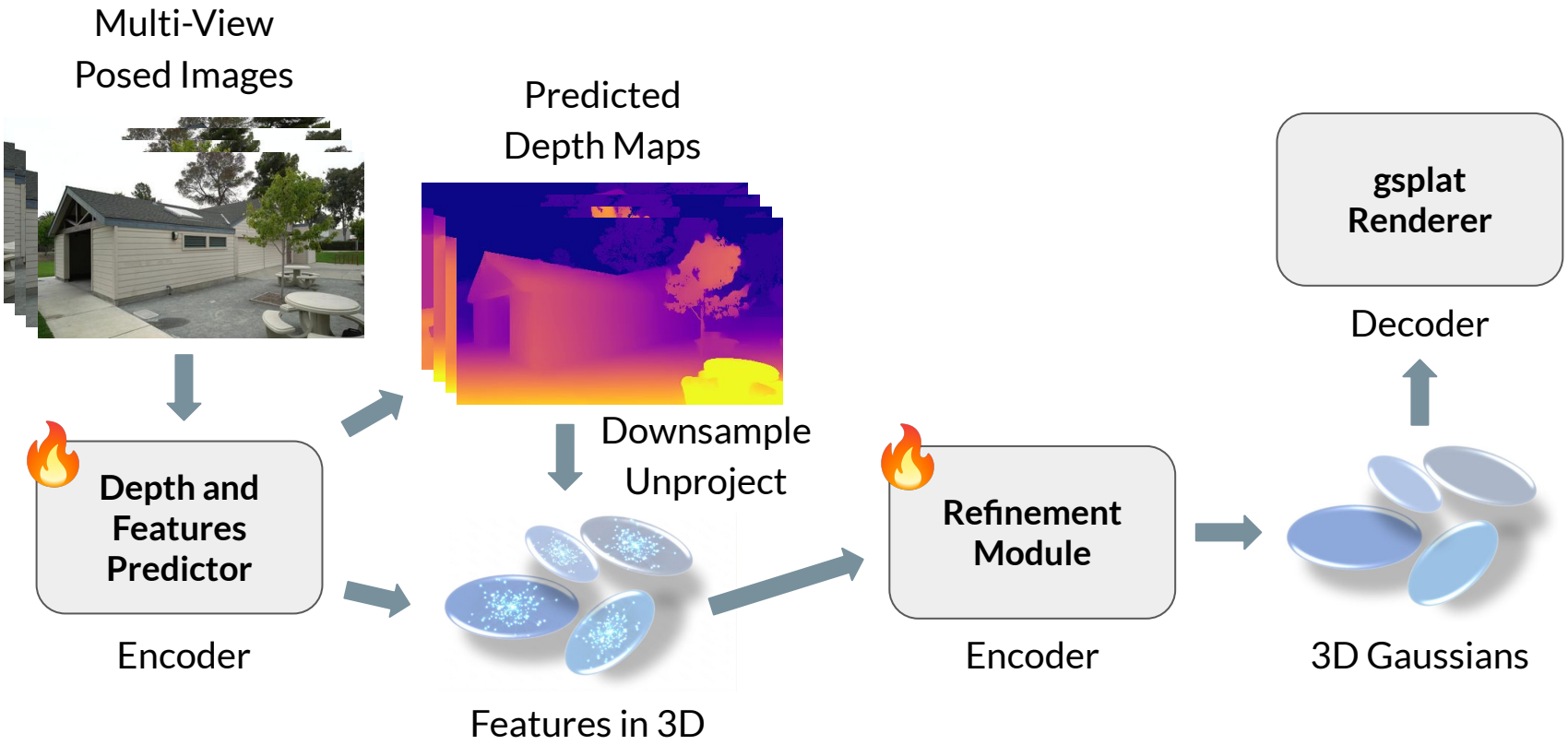



Illustration adapted from: Xu et al.: DepthSplat: Connecting Gaussian Splatting and Depth; CVPR 2025

# 2

# Related Work

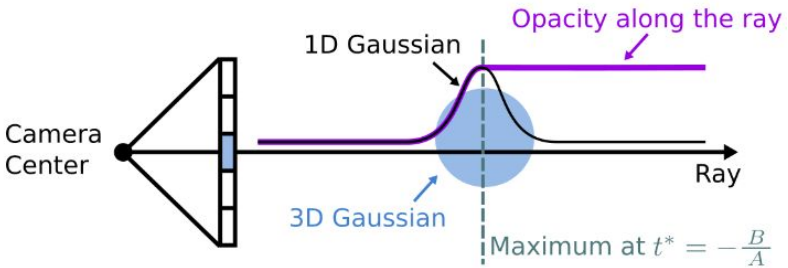
# ReSplat: Learning Recurrent Gaussian Splatting



# Gaussian Opacity Fields

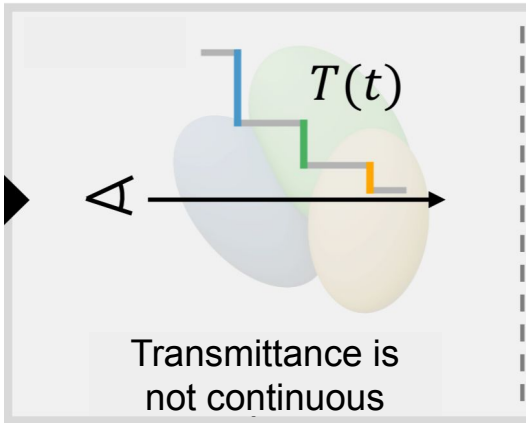
## Contributions

- Color, depth, and normal renderings
- Defines opacity at every 3D position enabling unbounded mesh extraction




## Shortcomings

- Depth rendering is not precise as it does not narrow down on 0.5-transmittance




Yu et al.: Gaussian Opacity Fields: Efficient Adaptive Surface Reconstruction in Unbounded Scenes; SIGGRAPH Asia 2024
Illustration from: Zhang et al.: Geometry-Grounded Gaussian Splatting ; arXiv 2026

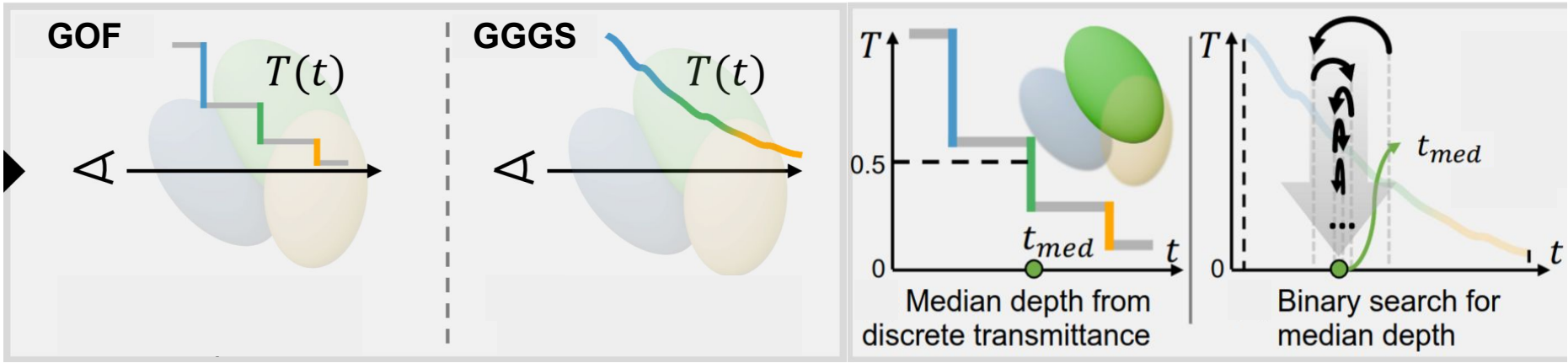
# Geometry-Grounded Gaussian Splatting

**Contributions**

- Produces precise and consistent depth renderings

**Shortcomings**

- Depth rendering process differs from color and normal rendering



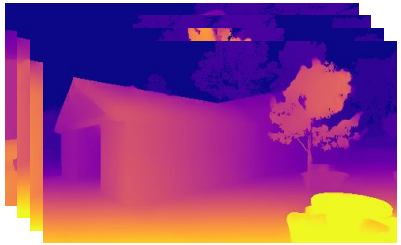
# 3

# Method

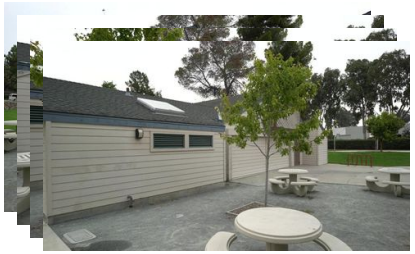
# Framework: Additions

Multi-View
Posed Images

Predicted
Depth Maps



GT Novel View



**2. Self-Supervised Geometry**

Loss

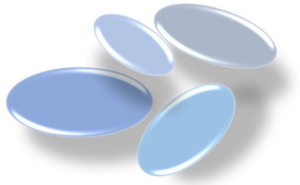
Novel View

**Depth and Gaussian Attributes Predictor**

Encoder

**1. 3D Smoothing Filter**
**3. Finetuning**

Unproject



3D Gaussians

**Gaussian Splatting Renderer**

Decoder

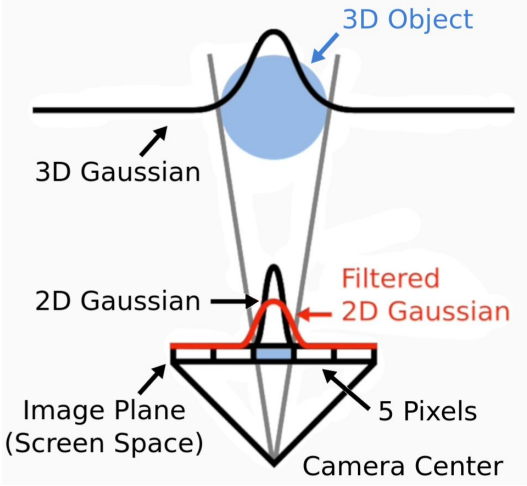
**GOF and GGGS**

Illustration adapted from: Xu et al.: DepthSplat: Connecting Gaussian Splatting and Depth; CVPR 2025

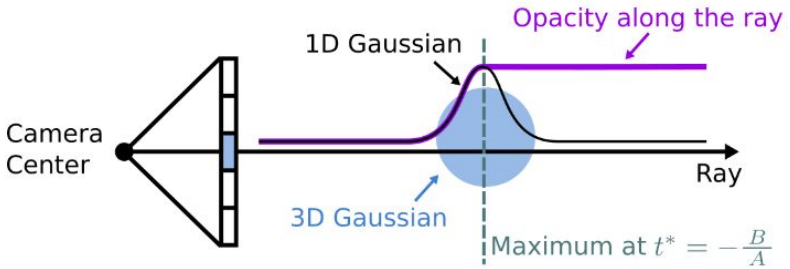
# 1. 3D Smoothing Filter

gsplat GOF

# 1. 3D Smoothing Filter




gsplat




GOF

Illustrations from: Yu et al.: Mip-Splatting: Alias-free 3D Gaussian Splatting; CVPR 2024
Yu et al.: Gaussian Opacity Fields: Efficient Adaptive Surface Reconstruction in Unbounded Scenes; SIGGRAPH Asia 2024

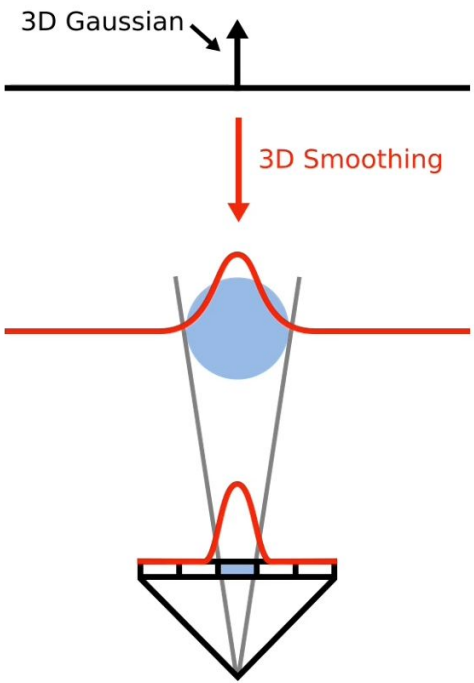
# 1. 3D Smoothing Filter

**MipSplatting on Feed-Forward Gaussian Splatting**

- **Goal:** Eliminate high-frequency artifacts thus stabilizing training when using GOF
- **Issue:** 3D Gaussian frequency exceeds the camera's sampling rate
- **Solution:** Apply a 3D smoothing filter to cap the Gaussian's frequency below the maximal sampling limit

constrain frequency to $\nu_k = \frac{f_n}{d_{k,n}}$

achieved by $\Sigma_k = \Sigma_{k,\text{pred}} + \frac{s}{\nu_k} \cdot \mathbf{I}$




Yu et al.: Mip-Splatting: Alias-free 3D Gaussian Splatting; CVPR 2024

# 2. Self-Supervised Geometry Losses

**Normal consistency loss from 2DGS**

- Ensures gaussians are locally aligned with surfaces
- Encourages alignment between gaussian normals and the gradient of the depth map

$$\mathcal{L}_n = \sum_i \omega_i (1 - \mathbf{n}_i^{\mathrm{T}} \mathbf{N}) \qquad \mathbf{N}(x, y) = \frac{\nabla_x \mathbf{p}_s \times \nabla_y \mathbf{p}_s}{|\nabla_x \mathbf{p}_s \times \nabla_y \mathbf{p}_s|}$$

**Other losses are unstable and redundant**

- Image Warp
- Edge-Aware Depth Smoothness

| GGGS Opt. | w/o $\mathcal{L}_{iw}$ | w/o $\mathcal{L}_n$ | Full |
|---|---|---|---|
| F-Score ↑ | 0.60 | 0.57 | 0.60 |

Huang et al.: 2D Gaussian Splatting for Geometrically Accurate Radiance Fields; SIGGRAPH 2024
Results from: Zhang et al.: Geometry-Grounded Gaussian Splatting ; arXiv 2026

# 3. ReSplat Encoder Fine-Tuning

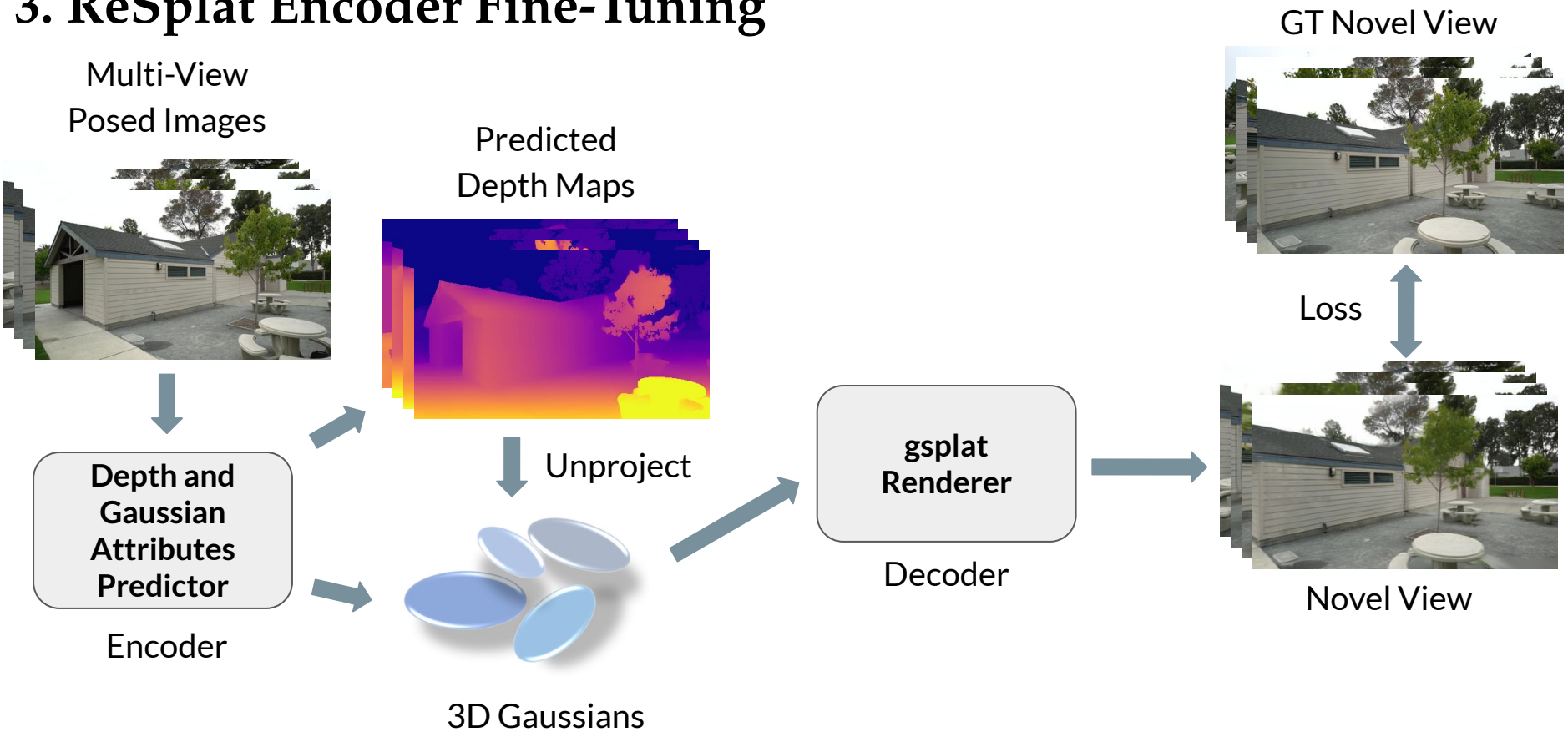
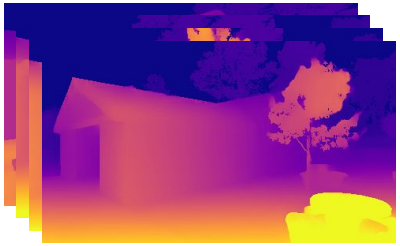



Illustration adapted from: Xu et al.: DepthSplat: Connecting Gaussian Splatting and Depth; CVPR 2025

# 3. ReSplat Encoder Fine-Tuning

Multi-View
Posed Images

Predicted
Depth Maps



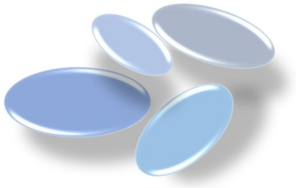
GT Novel View

**Normal Consistency** Loss

**Depth and Gaussian Attributes Predictor**

Encoder

Unproject



3D Gaussians

**GGGS Renderer**

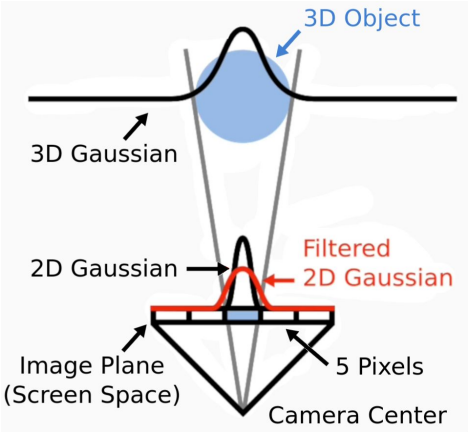
Decoder

Novel View
+ Depth & Normal Maps

Illustration adapted from: Xu et al.: DepthSplat: Connecting Gaussian Splatting and Depth; CVPR 2025

# 3. ReSplat Encoder Fine-Tuning

# 3. ReSplat Encoder Fine-Tuning




**Forward pass: Color rendering in GGGS and gsplat**

- Project 3D Gaussians to 2D view-space Gaussians
- Filter 2D covariances to avoid stability issues

**Backward pass: 2D Gaussian Filtering**

**GGGS**

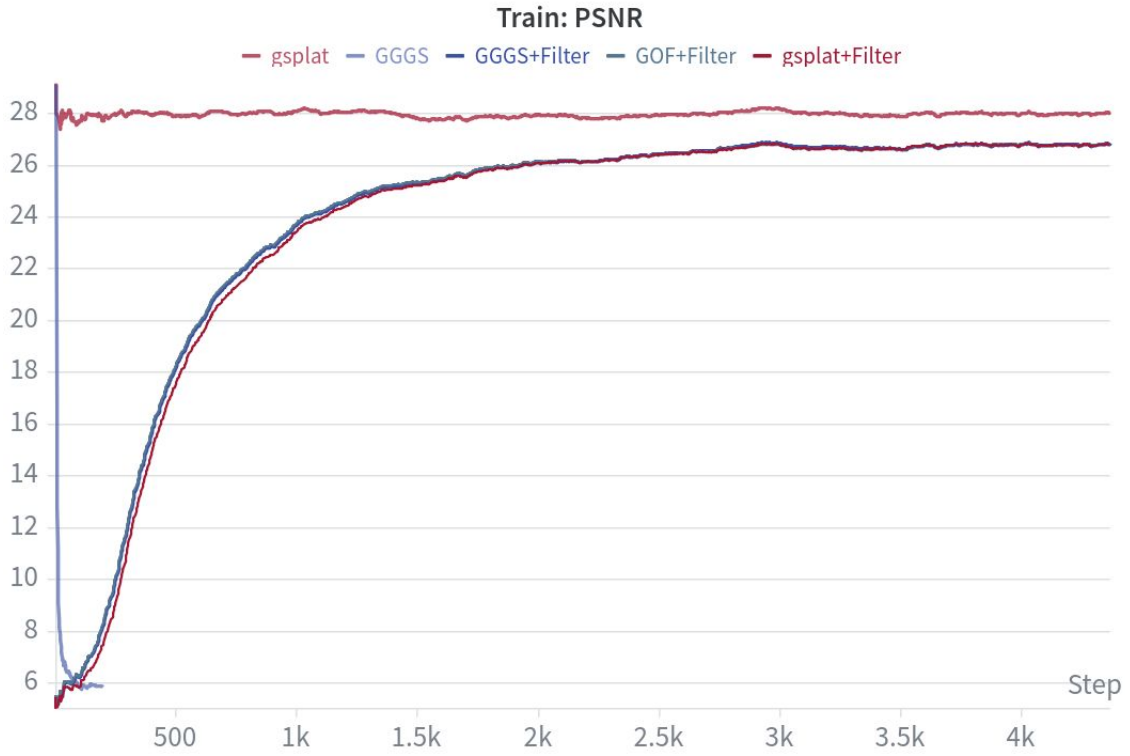
- Computes derivatives through intermediate variables
- Stops gradient flow when dividing by near-zero values

**gsplat**

- Simplifies chain rule math beforehand
- Avoids divisions which could lead to undefined or explosive gradients

Illustration from: Yu et al.: Mip-Splatting: Alias-free 3D Gaussian Splatting; CVPR 2024

# 3. ReSplat Encoder Fine-Tuning



# 4

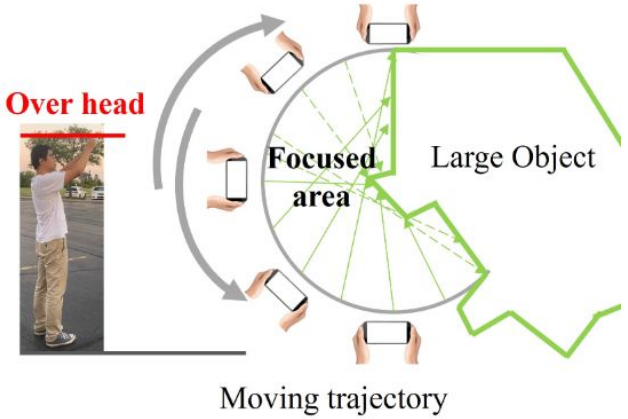
# Evaluation

# Datasets: Testing

- **Tanks and Temples**: Barn, Caterpillar, Courthouse, Ignatius, Meetingroom, Truck
- **MipNeRF360**: bicycle, flowers, garden, stump, treehill
- **DL3DV**'s test set: 10 outdoor scenes

**Context and target images**

We select images which resemble the standardized camera trajectories of DL3DV




Knapitsch et al.: Tanks and Temples: Benchmarking Large-Scale Scene Reconstruction; ACM Transactions on Graphics 2017
Barron et al.: Mip-NeRF360: Unbounded Anti-Aliased Neural Radiance Fields; CVPR 2022
Ling et al.: DL3DV-10K: A Large-Scale Scene Dataset for Deep Learning-based 3D Vision; CVPR 2024

# Methods

- **Encoder Depth Estimation:** Trained on RE10K and DL3DV with photometric loss and fine-tuned on indoor and synthetic datasets with GT depth supervision
- **Encoder**
  - Trained with gsplat as decoder and photometric losses
  - Trained from scratch with GOF and normal consistency loss
  - Fine-tuned to GOF decoder and normal consistency loss
  - Fine-tuned to GGGS decoder and normal consistency loss
- **GGGS per-scene optimization for 4k steps and regularization from 1k steps**
  - With MV losses: Including image-warp like loss
  - Without MV losses: L1, SSIM, and normal consistency loss
- **SurfSplat**
- **Depth Anything 3 Giant**

# Metrics

### Mesh Reconstruction

Tanks and Temples provides GT point clouds and evaluation code. We report:

- F-Score: Percentage of points in reconstruction close enough to the ground truth and viceversa

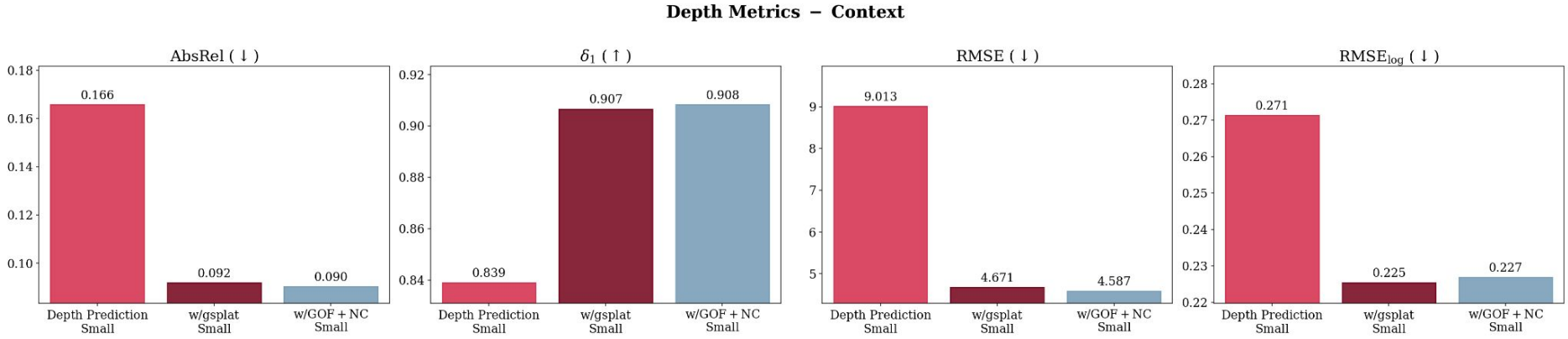
### Depth Estimation

'GT' depth maps are rendered from the per-scene GGGS-optimized gaussian scene. We report:

- Mean Absolute Relative L1 error
- RMSE and log RMSE
- Threshold Accuracy, $\delta_1$

$$\max\left(\frac{d_{\text{gt}}}{d_{\text{rend}}}, \frac{d_{\text{rend}}}{d_{\text{gt}}}\right) < 1.25$$

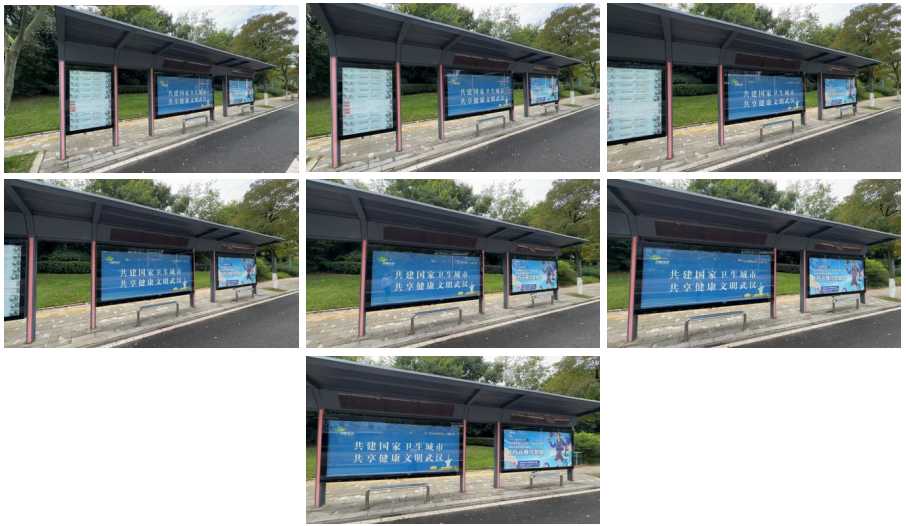
# Finding 1:
# Encoder Trained with Photometric Loss Excels at Depth Accuracy




- Training with photometric and normal consistency losses using GOF decoder does not significantly improve depth accuracy

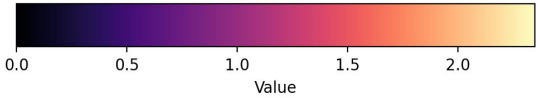
# Finding 1:
# Encoder Trained with Photometric Loss Excels at Depth Accuracy

**Input Images**



# Finding 1:
# Encoder Trained with Photometric Loss Excels at Depth Accuracy

**Absolute Depth Difference, non-metric**

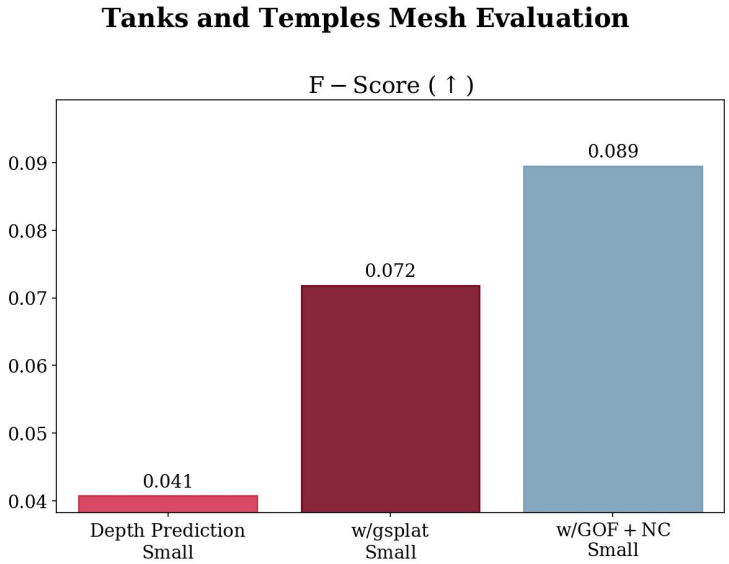



Scene AbsRel:

Depth Predictor
0.1829

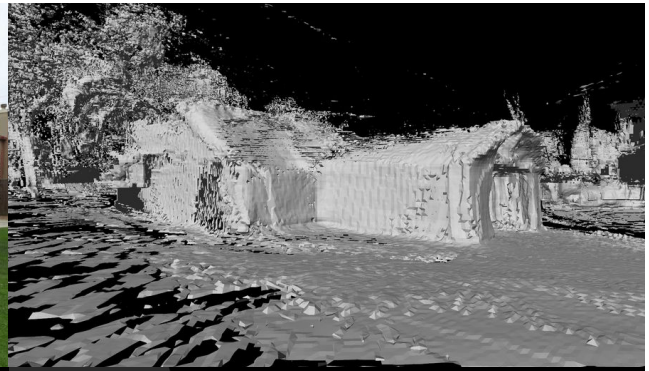
Encoder w/gsplat
0.0612

# Finding 2:
# Photometric and Normal Consistency Loss Enable Accurate Mesh Reconstruction



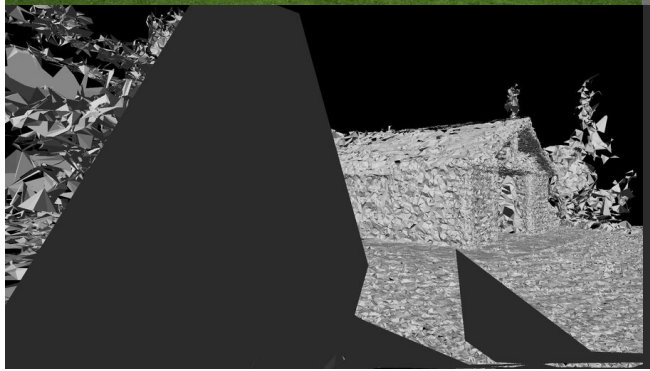
# Finding 2:
# Photometric and Normal Consistency Loss Enable Accurate Mesh Reconstruction

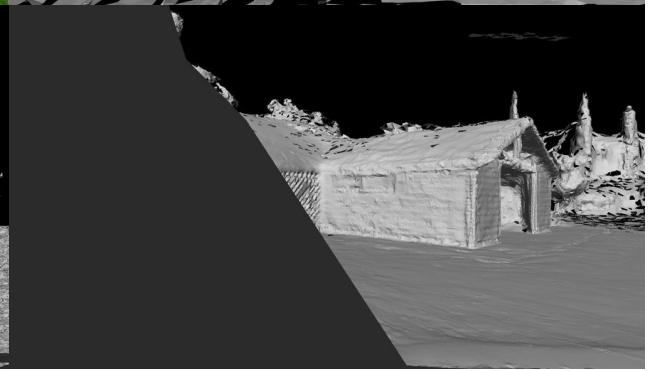
Color



Depth Predictor F-Score: 0.0335

w/gsplat Small F-Score: 0.0682



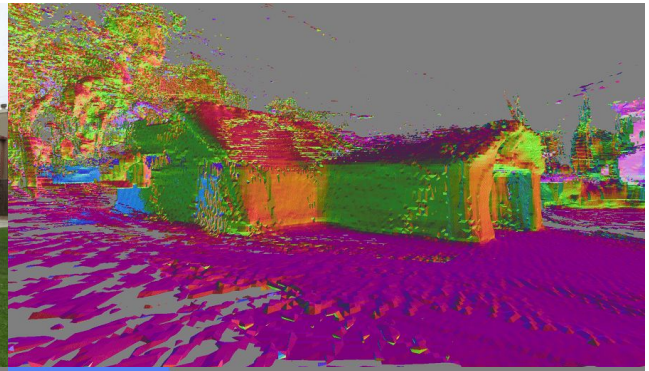


w/GOF Small F-Score: 0.0738

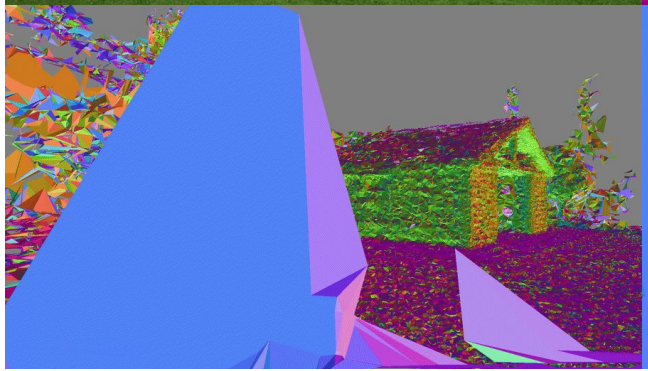
# Finding 2:
# Photometric and Normal Consistency Loss Enable Accurate Mesh Reconstruction

Color

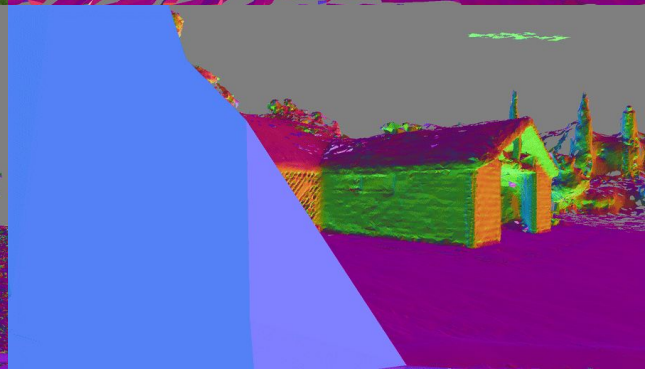
Depth Predictor F-Score: 0.033



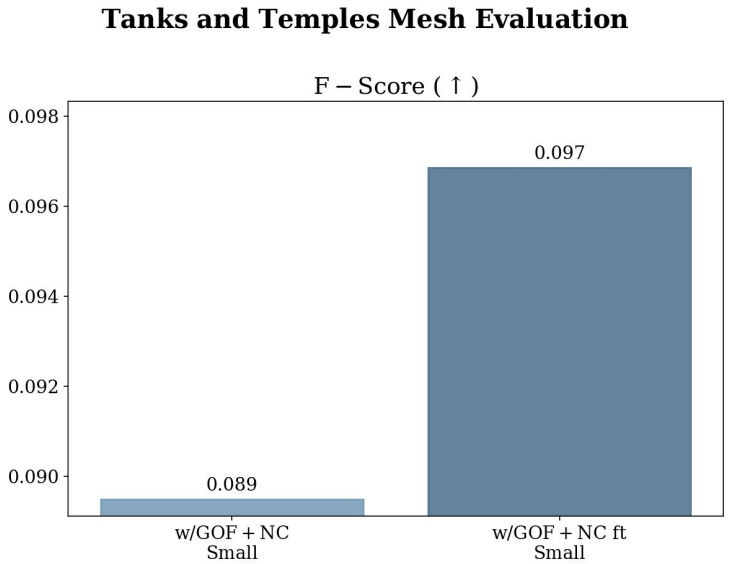
w/gsplat Small F-Score: 0.068



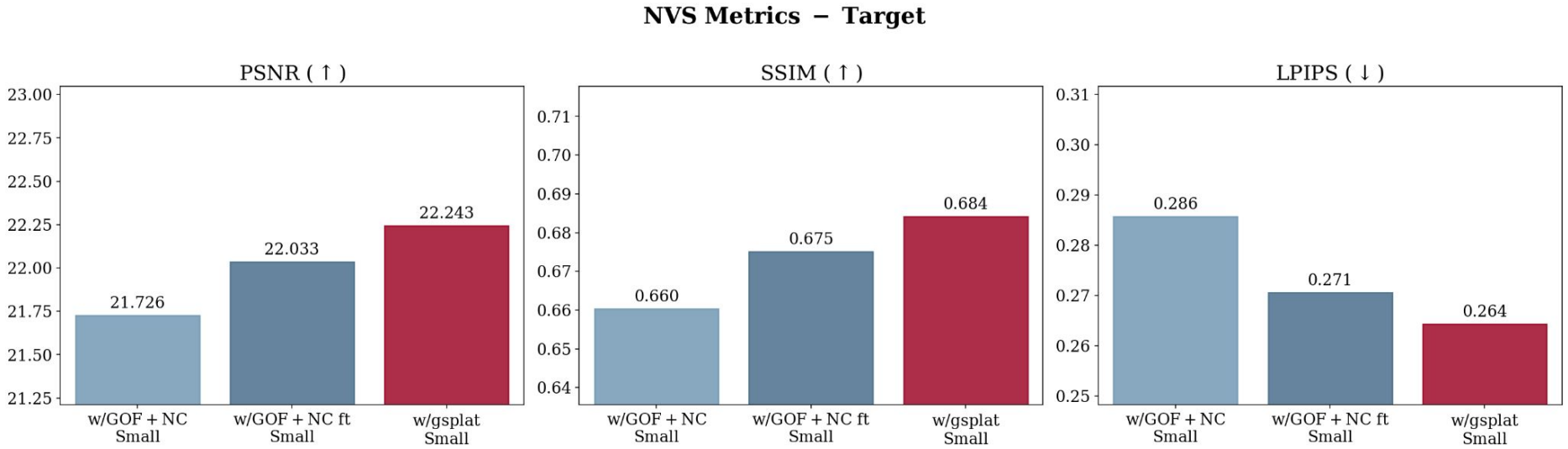
w/GOF Small F-Score: 0.073



# Finding 3:
# Fine-Tuning on Normal Consistency Loss Outperforms Training from Scratch



# Finding 3:
# Fine-Tuning on Normal Consistency Loss Outperforms Training from Scratch



# Finding 3:
# Fine-Tuning on Normal Consistency Loss Outperforms Training from Scratch

GT

w/GOF+NC Small

Target views
PSNR: 20.00

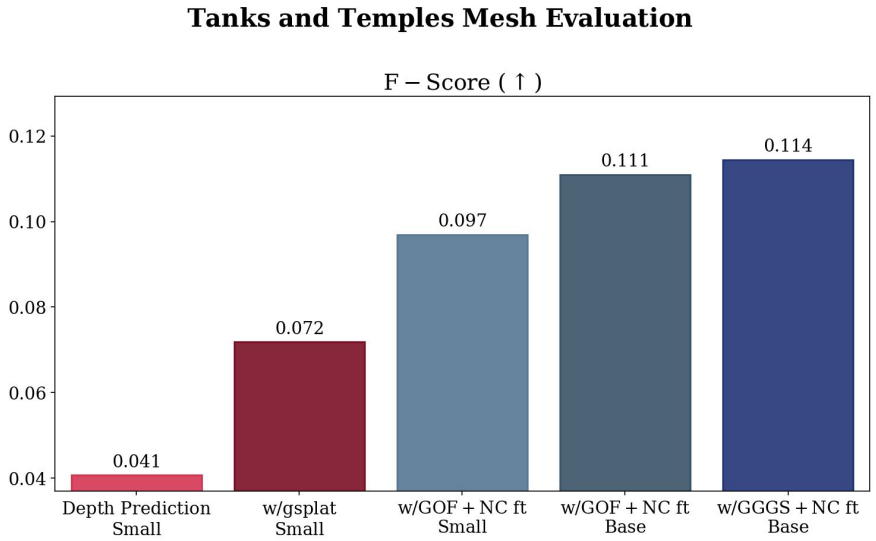
w/GOF+NC ft Small

Target views
PSNR: 20.30

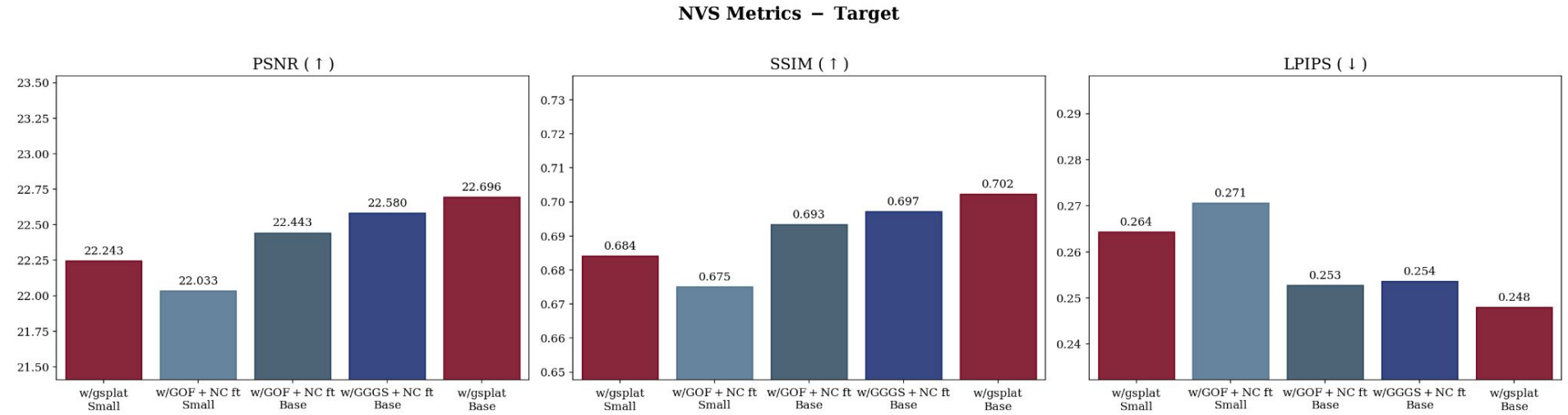
w/gsplat Small

Target views
PSNR: 20.49

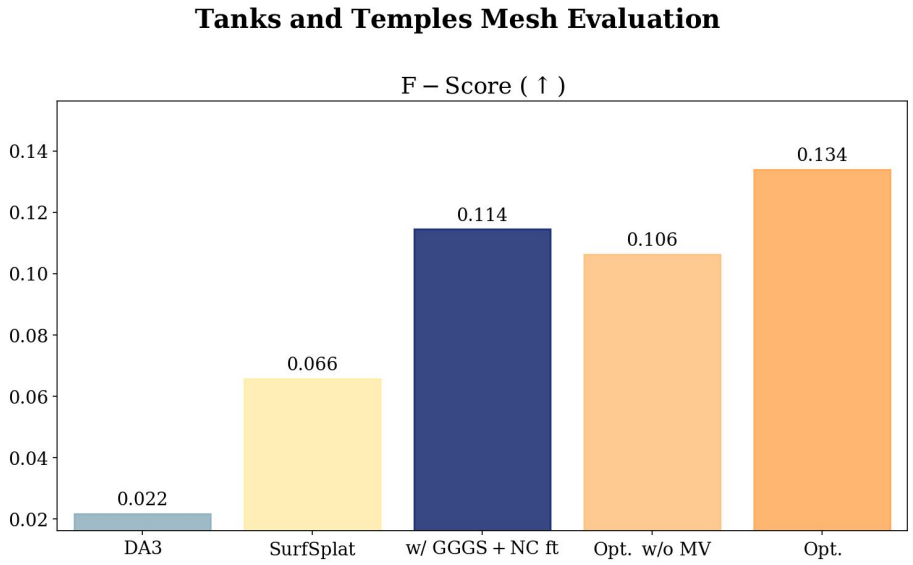
# Finding 4:
# Fine-Tuned Base Encoder with GGGS and Normal Consistency Loss Performs the Best



# Finding 4:
# Fine-Tuned Base Encoder with GGGS and Normal Consistency Loss Performs the Best



# Finding 5:
# Fine-Tuned Base Encoder Outperforms DA3, SurfSplat, and Performs on-par with per-scene Optimization

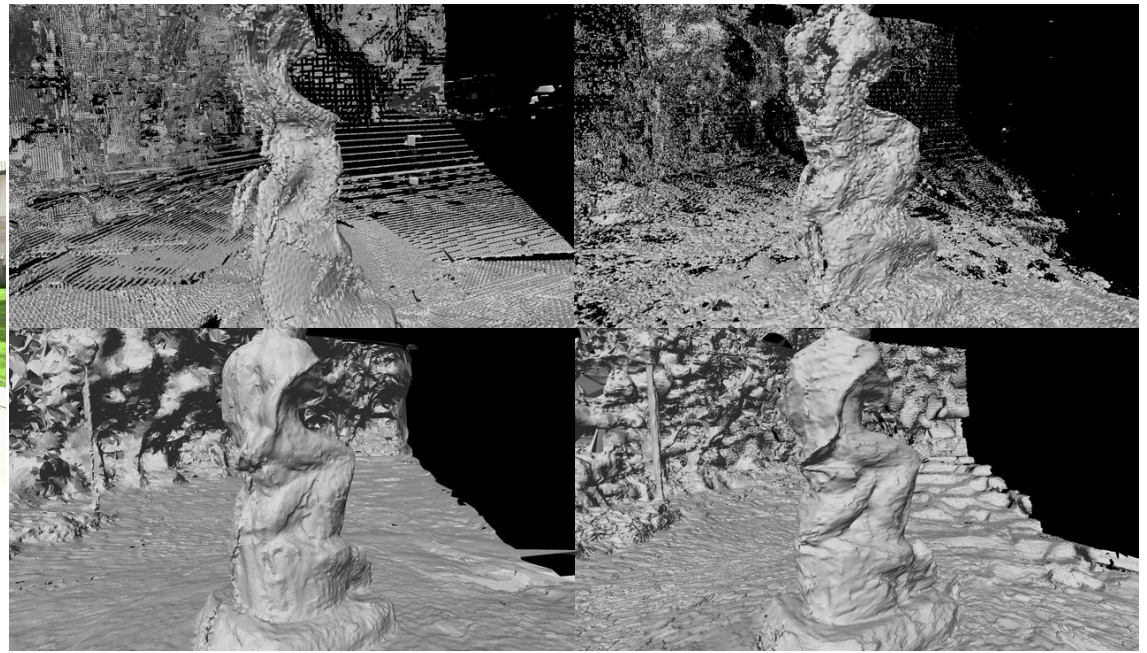


# Finding 5:
# Fine-Tuned ReSplat Encoder Outperforms DA3, SurfSplat, and Performs on-par with per-scene Optimization

Color

DA3. F-Score: 0.03

SurfSplat. F-Score: 0.12



w/GGGS+NC ft. F-Score: 0.17

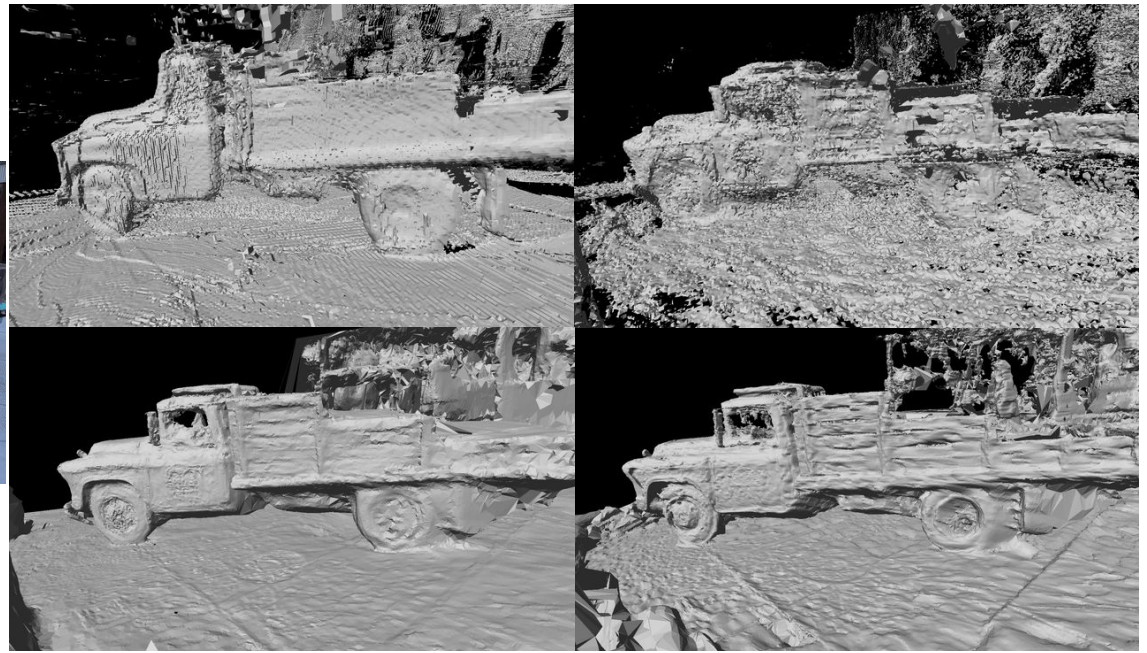
GGGS Opt. F-Score: 0.26

# Finding 5:
# Fine-Tuned ReSplat Encoder Outperforms DA3, SurfSplat, and Performs on-par with per-scene Optimization

Color

DA3. F-Score: 0.05

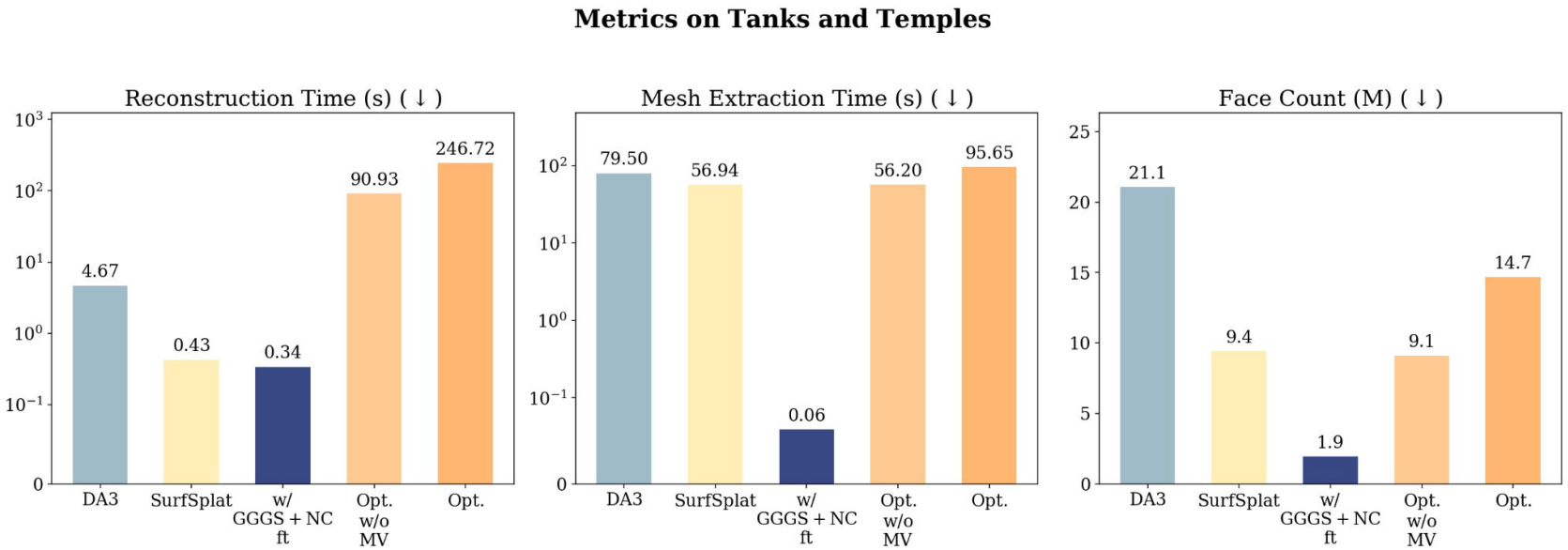
SurfSplat. F-Score: 0.14



w/GGGS+NC ft. F-Score: 0.18

GGGS Opt. F-Score: 0.19

# Finding 5:
# Fine-Tuned ReSplat Encoder Outperforms DA3, SurfSplat, and Performs on-par with per-scene Optimization



# 5

# Conclusion

# Conclusion

1. Feed-Forward Gaussian Splatting models **can** achieve accurate **depth estimation** with photometric losses
2. Normal consistency loss does not improve depth estimation significantly but does improve **mesh reconstruction accuracy**
3. Fine-Tuning Feed-Forward Gaussian Splatting models is not straightforward and depends on the choice of **primitives** and **renderers**
4. Feed-Forward Gaussian Splatting models **achieve state-of-the-art performance in depth estimation and mesh reconstruction tasks**

**Future Work:**

- Compact 3D representations circumvent depth predictor module
- Primitives other than gaussians: triangles or meshes
- New depth priors: Depth Anything V3

# Fin

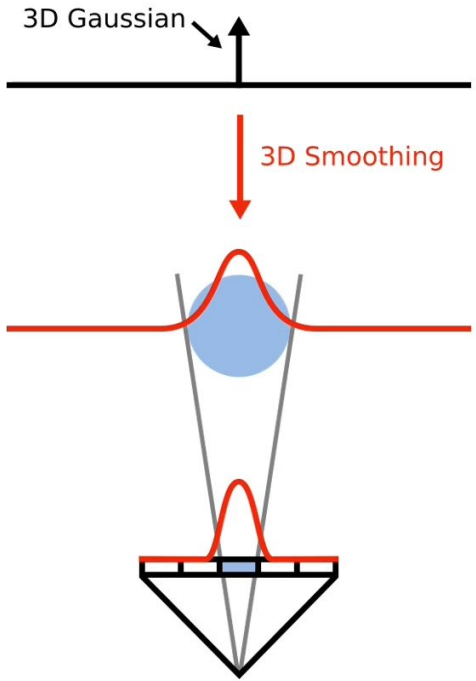
# Questions ?

# Appendix

# 1. 3D Smoothing Filter

**MipSplatting on 3D Gaussian Splatting**

- **Goal**: Eliminate high-frequency artifacts
- **Issue:** 3D Gaussian frequency exceeds the camera's sampling rate
- **Solution:** Apply a 3D smoothing filter to cap the Gaussian's frequency below the maximal sampling limit

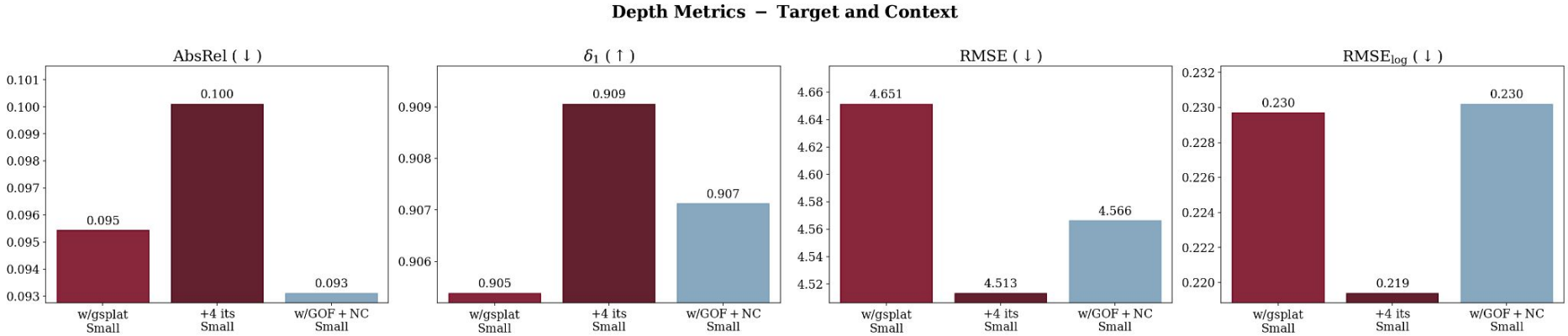



constrain frequency to

$$\hat{\nu}_k = \max\left(\left\{\mathbb{1}_n(\mathbf{p}_k)\cdot\frac{f_n}{d_n}\right\}_{n=1}^{N}\right)$$

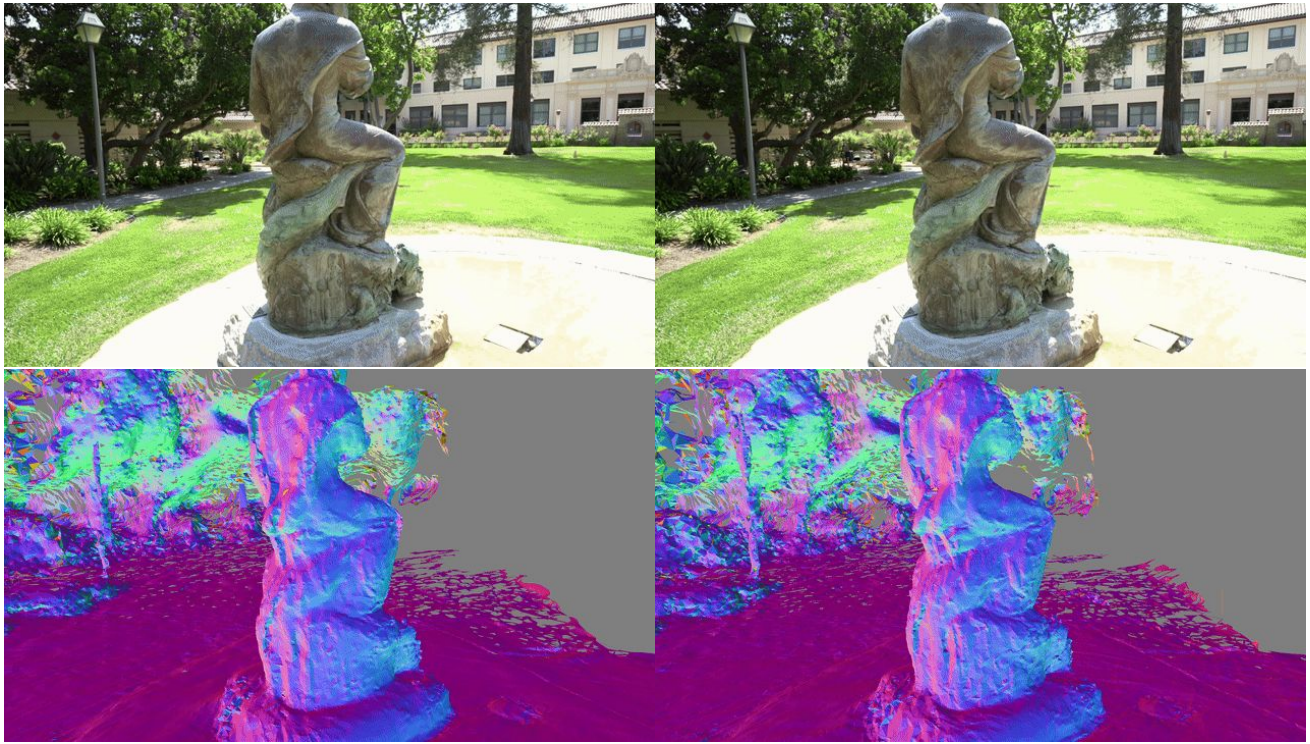
Yu et al.: Mip-Splatting: Alias-free 3D Gaussian Splatting; CVPR 2024

# Finding 1:
# Encoder Trained with Photometric Loss Excels at Depth Accuracy

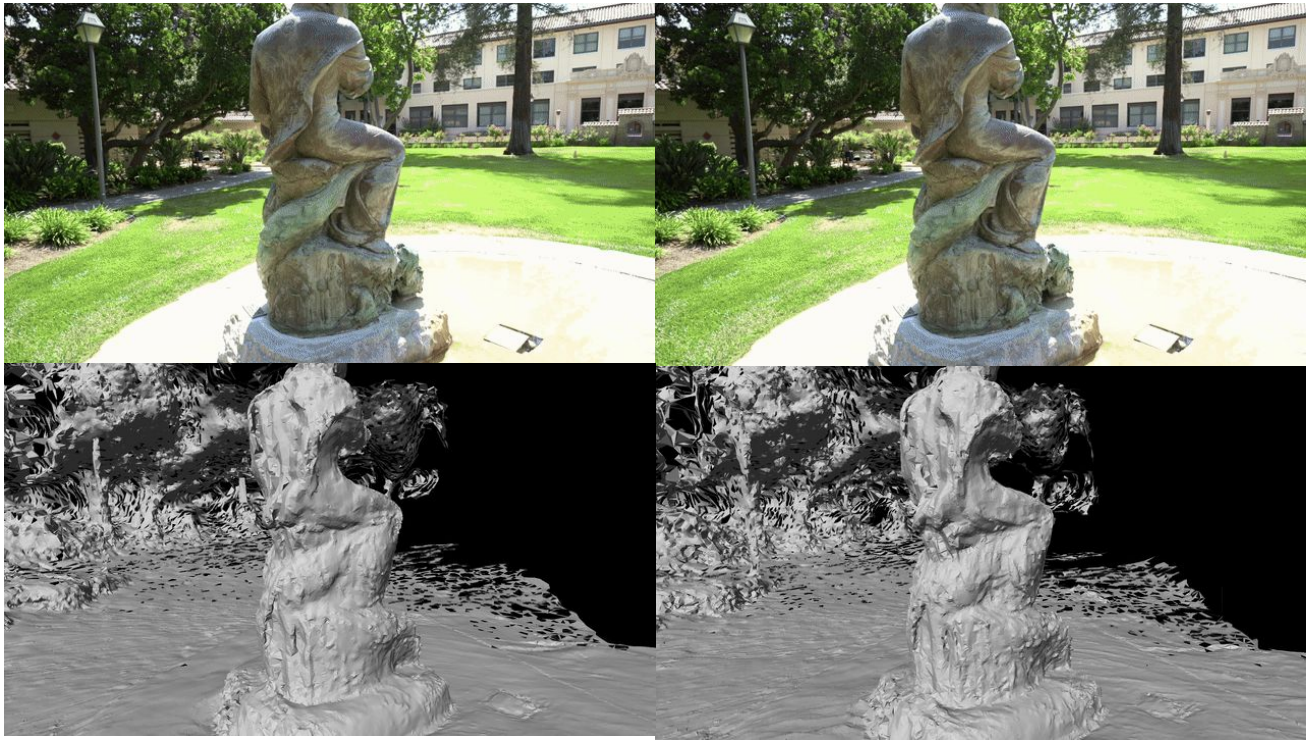



- Refinement module also does not significantly improve depth accuracy

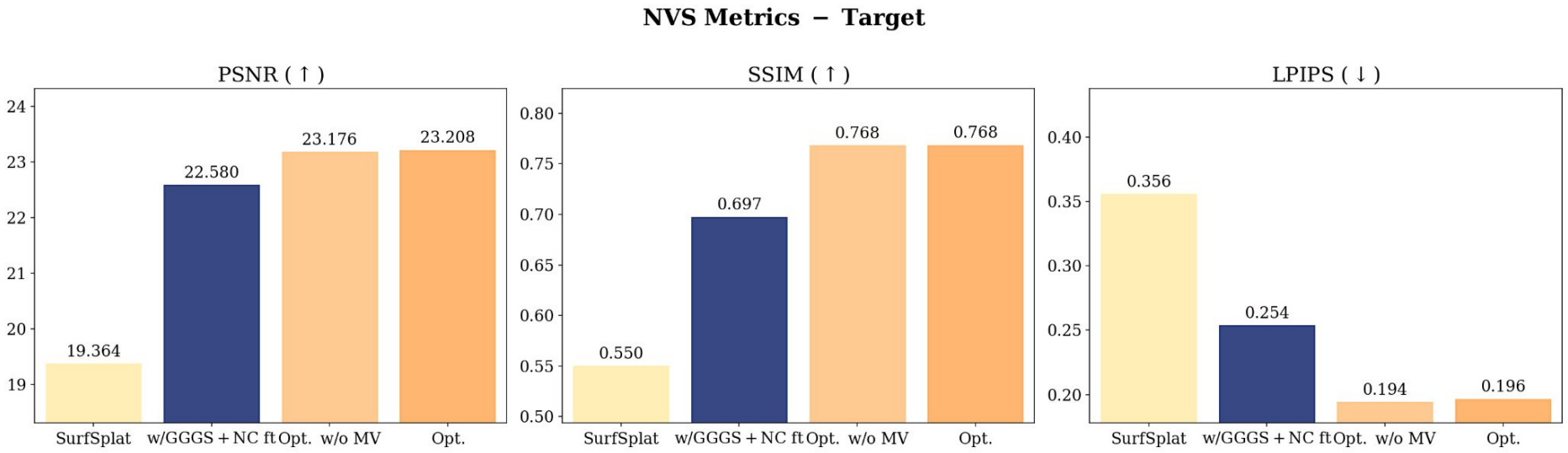
# Finding 3:
# Fine-Tuning on Normal Consistency Loss Outperforms Training from Scratch



# Finding 3:
# Fine-Tuning on Normal Consistency Loss Outperforms Training from Scratch



# Finding 5:
# Fine-Tuned ReSplat Encoder Outperforms DA3, SurfSplat, and Performs on-par with per-scene Optimization



# Abstract

*Existing feed-forward Gaussian Splatting models focus on visual quality, while the few approaches targeting geometric quality are limited to indoor scenes and image pairs. This thesis focuses on enabling feed-forward models to achieve geometrically accurate 3D representations of unbounded scenes from 8 or 16 sparse views at 256 by 448 resolution, evaluated via extracted meshes and rendered depth maps. Leveraging the recent ReSplat model, we make three major contributions towards this goal:*

*First, we stabilize training by applying a 3D filter to the predicted Gaussians, preventing them from exceeding the Nyquist frequency.*

*Second, we explore the behavior of common self-supervised losses for geometric accuracy in feed-forward models. We find that edge-aware depth smoothness and image warp losses lead to unstable training. Instead, we use an edge-aware normal consistency loss.*

*Third, we successfully fine-tune a ReSplat encoder, trained with gsplat as its decoder, using the Gaussian Opacity Fields and the Geometry-Grounded Gaussian Splatting (GGGS) decoders alongside the edge-aware normal consistency loss. These fine-tuned encoders achieve similar geometric accuracy and higher visual quality compared to training from scratch.*

*We evaluate our GGGS fine-tuned encoder and observe strong performance across three key metrics. For mesh reconstruction on the Tanks and Temples dataset, our method improves precision by 73% over the photometric-only encoder, performs on par with per-scene optimization, and surpasses Depth Anything 3 by a factor of three. For depth estimation across MipNeRF360, Tanks and Temples, and DL3DV, our method reduces absolute relative distance by 7% against the photometric-only baseline, 31% against Depth Anything 3, and performs on par with per-scene optimization. These improvements come at a minimal cost to visual fidelity, maintaining a PSNR within 0.12dB of the photometric-only encoder.*