MVSplat:

Efficient 3D Gaussian Splatting from Sparse Multi-View Images

Reviewer: Vitor Pereira Matias

Archeologist: Vitor Pereira Matias

Hacker: Davi

PhD Student: Veronika

Reviewer

Vitor Pereira Matias

Hot topic!

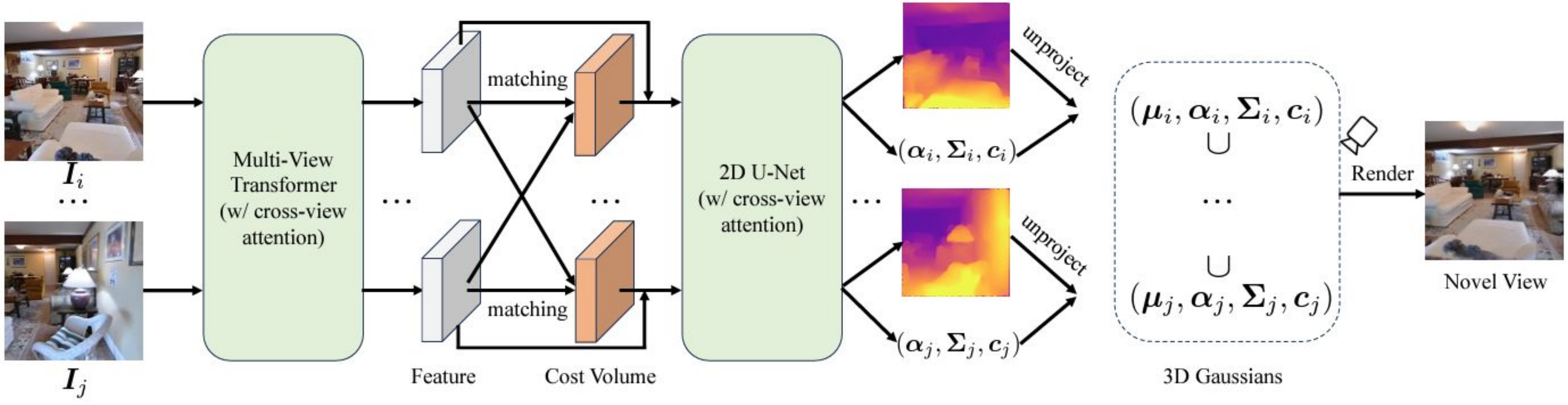
GS + (Large models, transformers, attention, NeRF, Diffusion)

Title	Date	Citations
TrackNeRF	20 Aug 2024	1
V3D	11 Mar 2024	24
Flash3D	6 Jun 2024	6
GS-LRM	30 Apr 2024	29
LGM	7 Feb 2024	121
DNGaussian	24 Mar 2024	27
GRM	21 Mar 2024	51
PixelSplat	19 Dec 2023	76

Summary

Summary:

Explain the key ideas, contributions, and their significance.

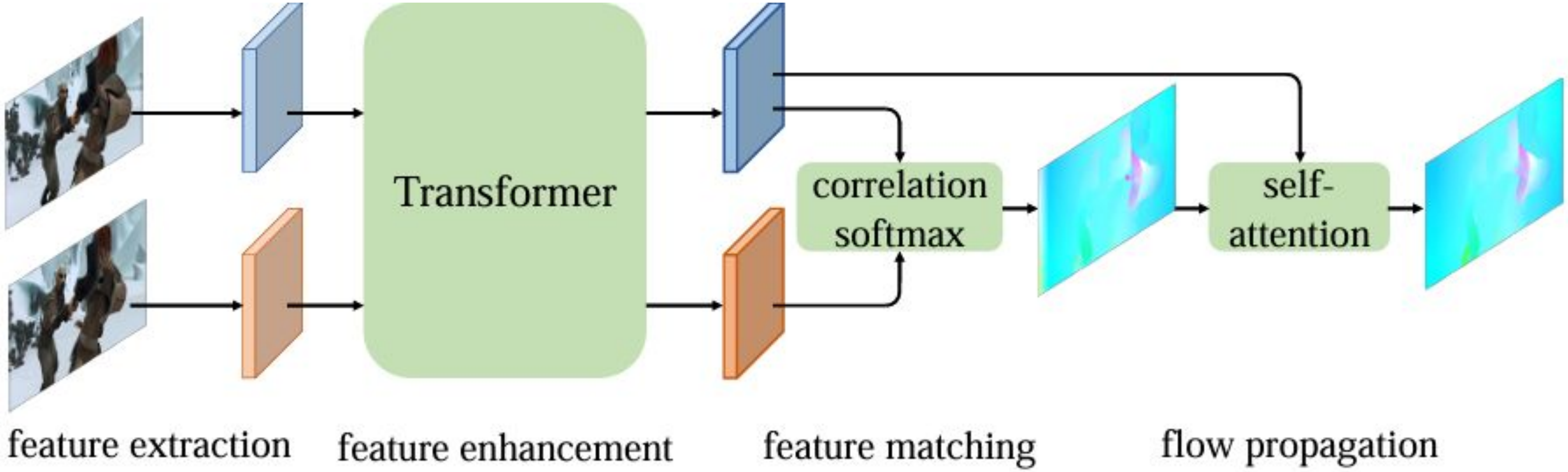


Summary

- Multiview transformer
  - transformer = Swin transformer

Summary:

Explain the key ideas, contributions, and their significance.



sourcer: gmflow

Summary

- **Cost computation:** The cost volume expresses how well a pixel i in image I matches the same pixel in the second image I shifted by vector l.

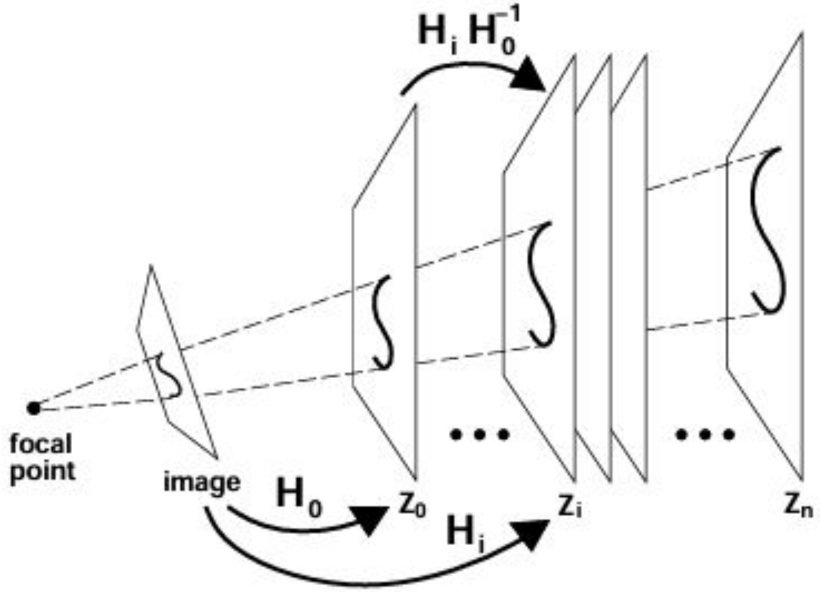


Figure 1: Illustration of the space-sweep method. Features from each image are backprojected onto successive positions $Z = z_i$ of a plane sweeping through space.

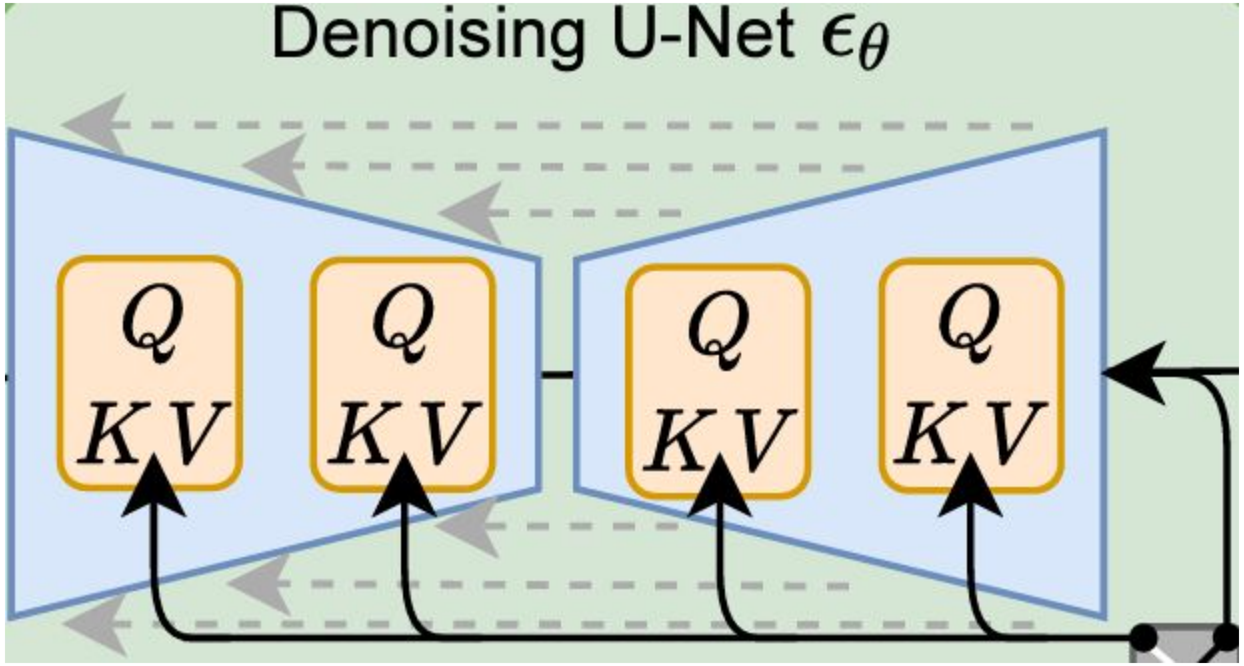
source: A Space-Sweep Approach to Multi-Image Matching

Summary:

Explain the key ideas, contributions, and their significance.

Summary

- Refinement



source: diffusion networks

Summary:

Explain the key ideas, contributions, and their significance.

Summary

- Depth estimation
  - refining is done with u-net as well

Summary:

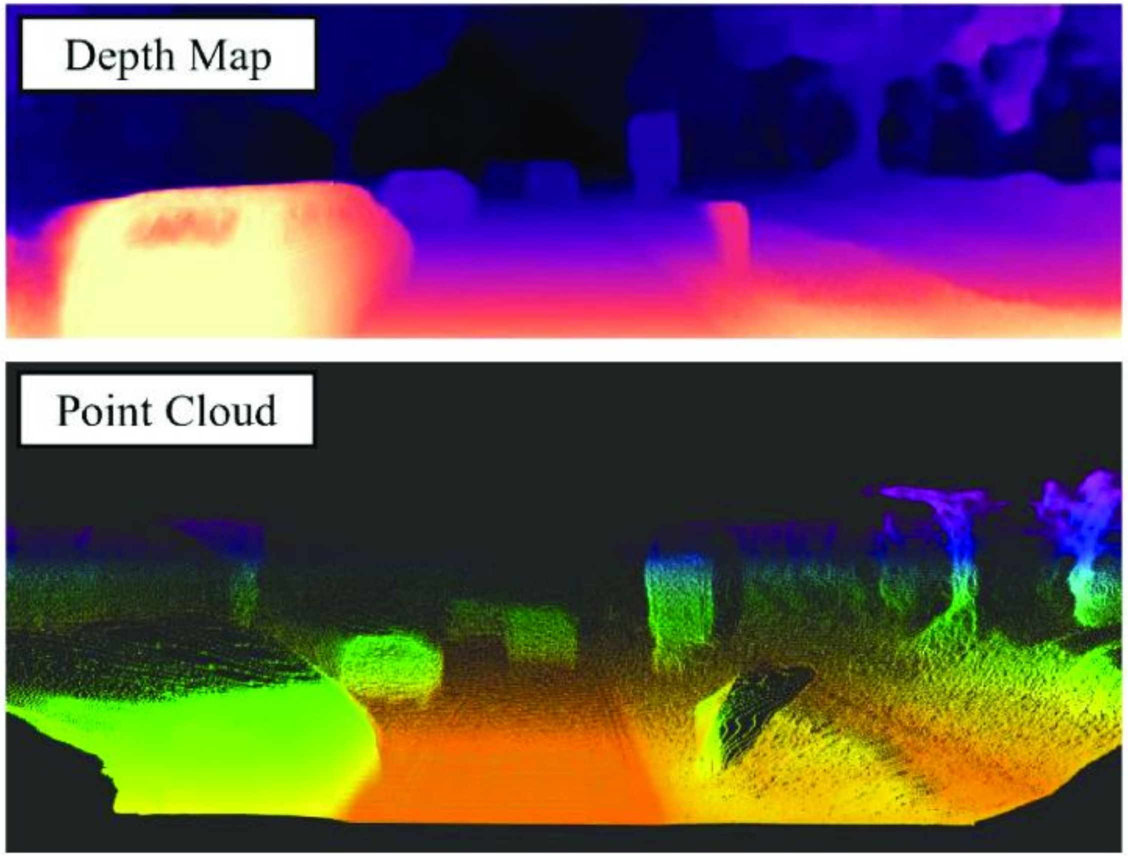
Explain the key ideas, contributions, and their significance.

Depth estimation. We use the `softmax` operation to obtain per-view depth predictions. Specifically, we first normalize the refined cost volume $\hat{\boldsymbol{C}}^i$ in the depth dimension and then perform a weighted average of all depth candidates $\boldsymbol{G} = [d_1, d_2, \cdots, d_D] \in \mathbb{R}^D$:

$$\boldsymbol{V}^i = \mathrm{softmax}(\hat{\boldsymbol{C}}^i)\boldsymbol{G} \in \mathbb{R}^{H \times W}. \quad (6)$$

Summary

- Gaussian generation



Summary:

Explain the key ideas, contributions, and their significance.

src

Strenghts

- 10x fewer parameters
- 2x faster
- better outputs than pixelsplat
- less pos-processing than pixelsplat
- better gneralization for N images

Strengths:

What about the paper provides value? --

Strenghts

Strengths:

What about the paper provides value? --

Table 1: SOTA results

Method	Time (s)	Param (M)	RealEstate10K [54]			ACID [21]		
			PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
pixelNeRF [49]	5.299	28.2	20.43	0.589	0.550	20.97	0.547	0.533
GPNR [35]	13.340	9.6	24.11	0.793	0.255	25.28	0.764	0.332
AttnRend [10]	1.325	125.1	24.78	0.820	0.213	26.88	0.799	0.218
MuRF [44]	0.186	**5.3**	26.10	0.858	0.143	28.09	0.841	0.155
pixelSplat [1]	0.104	125.4	25.89	0.858	0.142	28.14	0.839	0.150
MVSplat	**0.044**	12.0	**26.39**	**0.869**	**0.128**	**28.25**	**0.843**	**0.144**

Strenghts

Strengths:

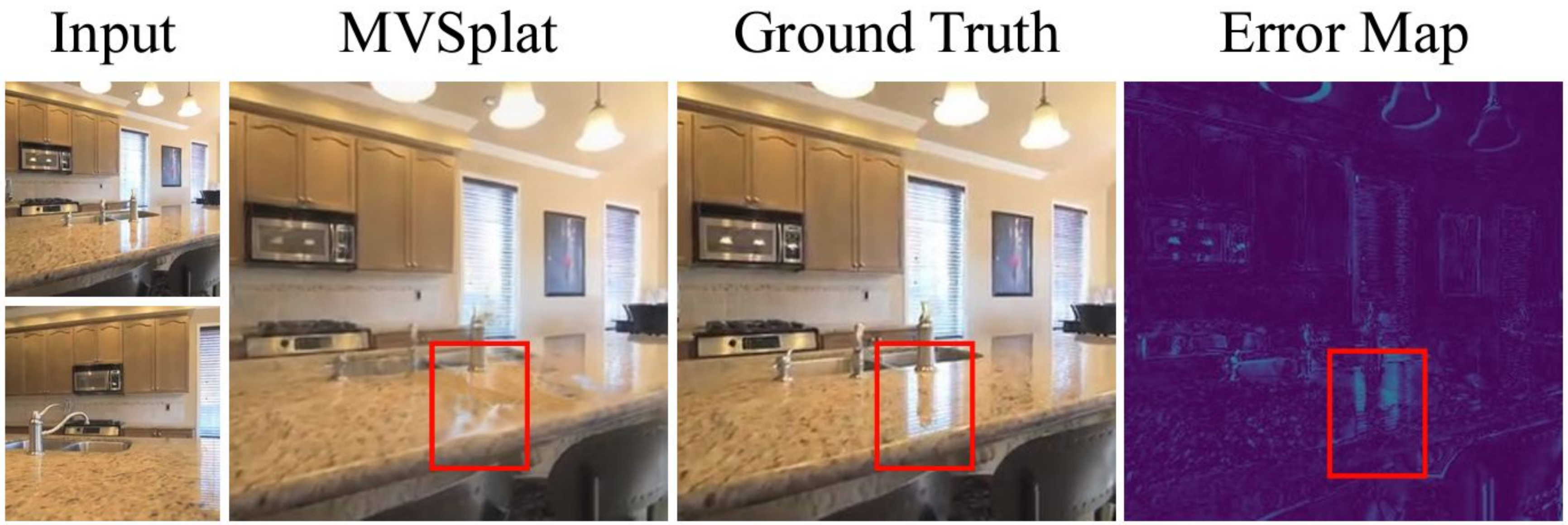
What about the paper provides value? --

Table 2: Better generalization

Training data	Method	ACID [21]			DTU [17]		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
RealEstate10K [54]	pixelSplat [1]	27.64	0.830	0.160	12.89	0.382	0.560
	MVSplat	**28.15**	**0.841**	**0.147**	**13.94**	**0.473**	**0.385**

Weaknesses

- Need camera poses
- Non-lambertian surfaces (mirrors and glasses)



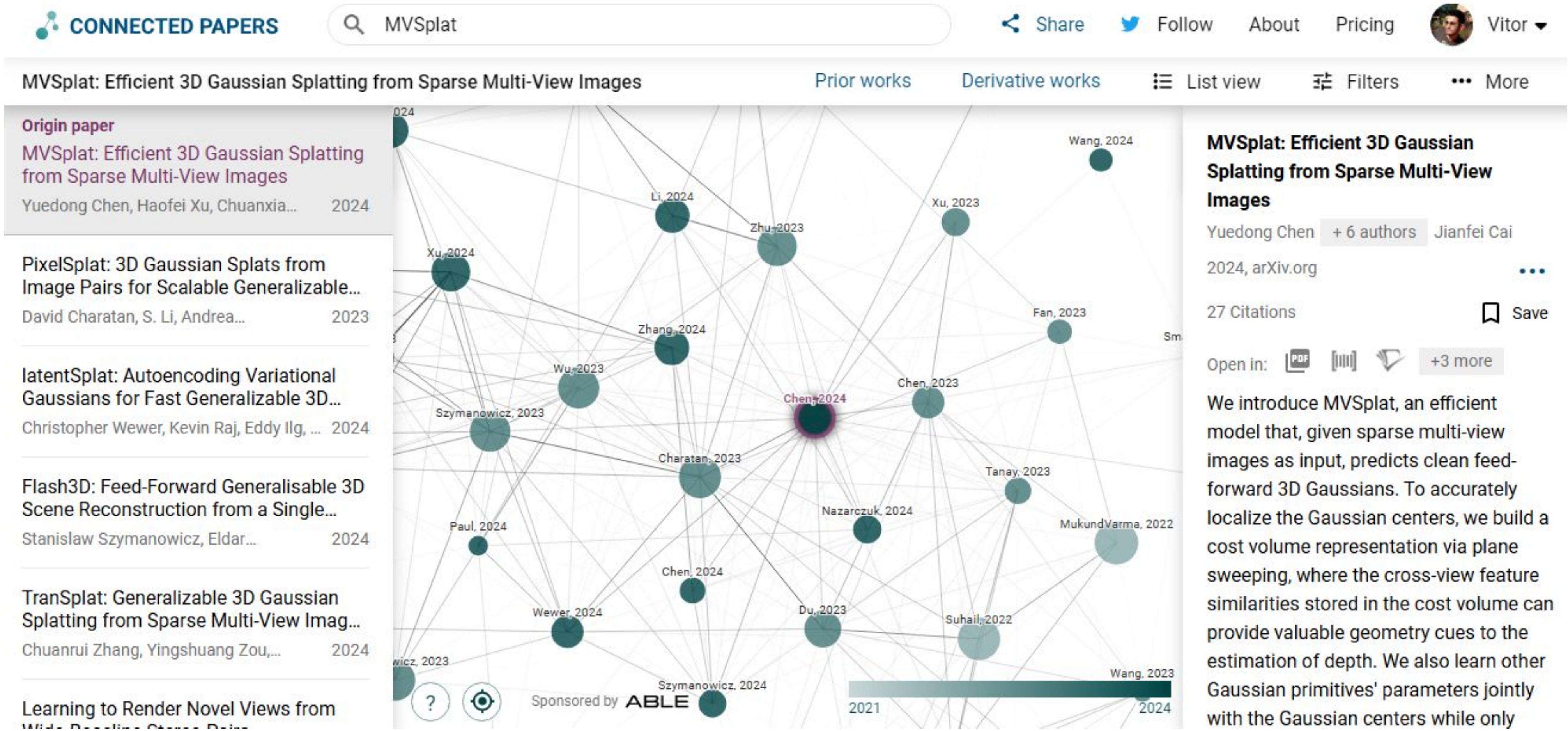
Weaknesses:

What detracts from the contributions?

Archeologist

Vitor Pereira Matias

Citations graph



connectedpapers

Foundations

- NeRF/Gaussian Splatting

- Depth estimation
- Multi-View Stereo (as features)
  - Gmflow (1/2)
  - Swin transformer
- Cost volume
  - Space-Sweep, Unifying Flow, Mvsnet
- (Cost volume, depth) Refinement (2d u-net)/(diffusion)

Foundations: Math

- inverse depth domain
- Warping cnn features
- From depth to 3d point clouds
- From depth softmax to Opacity
- Conv applied to image feature + cost volume + original images returns Covariance and Color
- Loss: LPIPS

Prior/concurrent works

Nerfs

- pixelNerf
- MuRF
- AttnRend
- GPNR

Splats

- pixelSplat
- LaRa
- GS-LRM
- GPS-Gaussian

Derivative works

Nerfs

- TrackNeRF (noisy sparse)

Splats

- HumanSplat
- V3D (diffusion) (git)
- Flash3D
- NoPoSplat

Similar, not cited

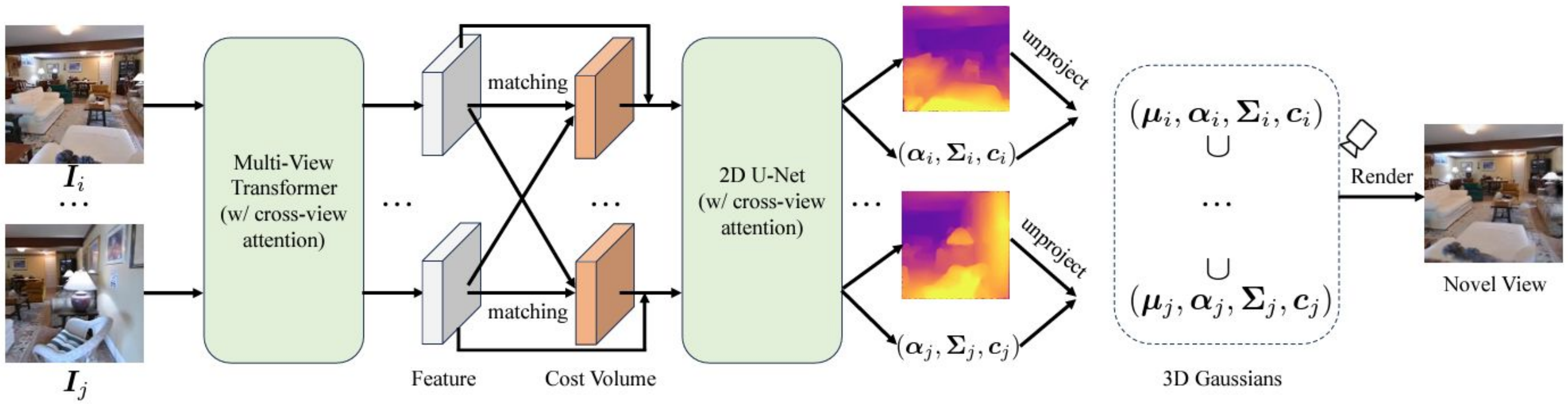
Splats

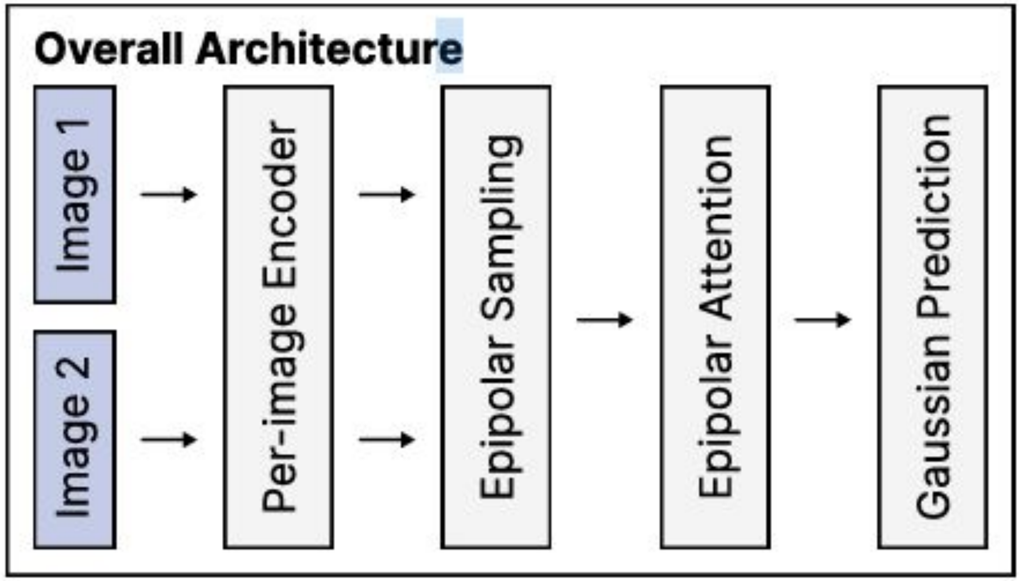
- DNGaussian
- Ges: Generalized exponential splatting
- SparseGS
- FSGS

Other methods

- Dust3r
- Spann3r
- PR-LRM

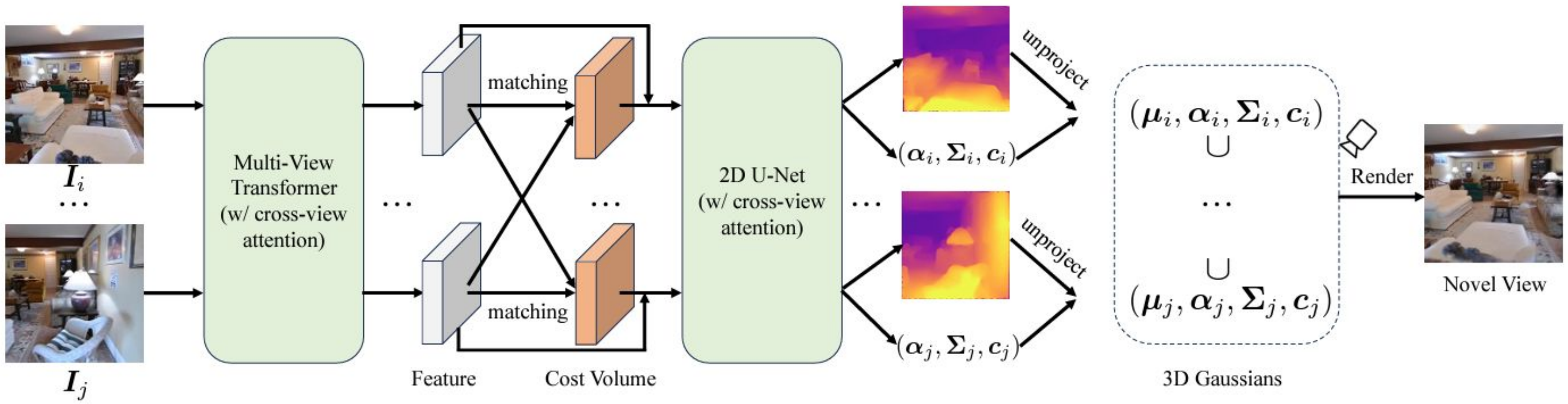
MVSplat vs PixelSplat

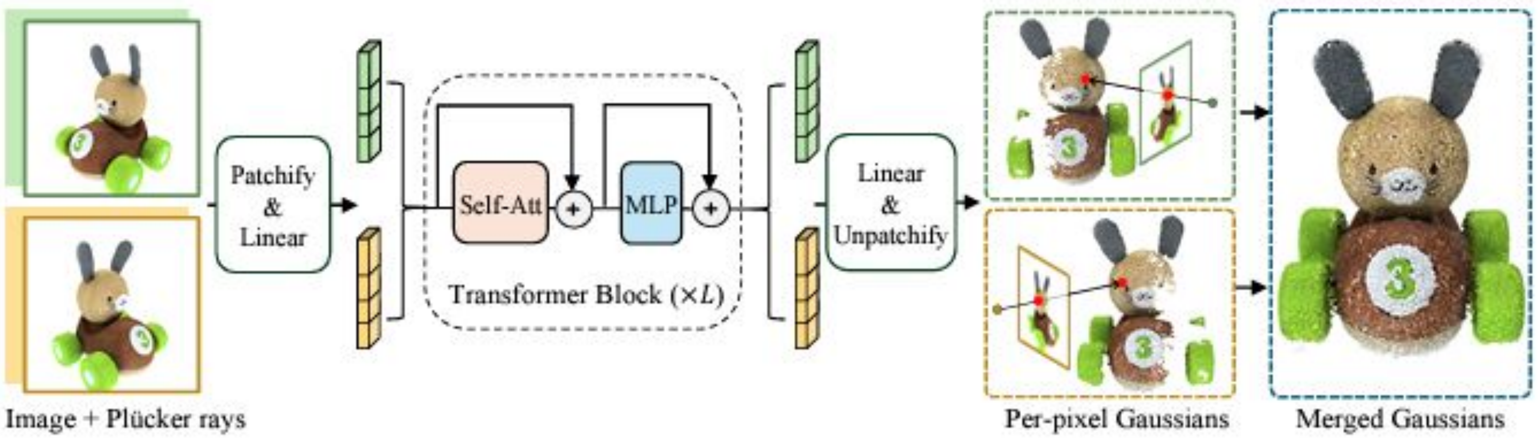




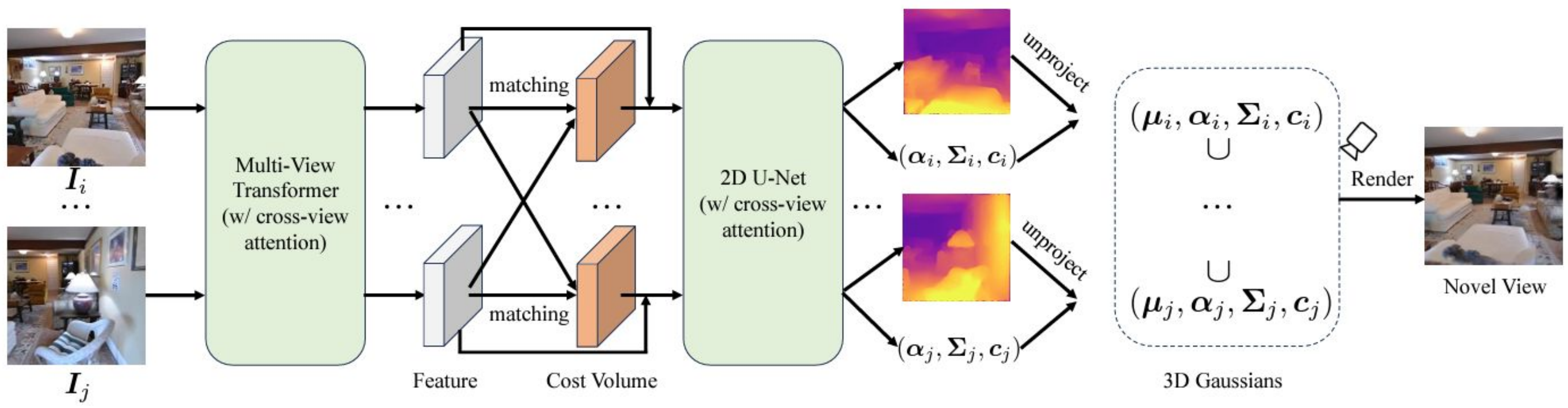
MVSplat vs PixelSplat

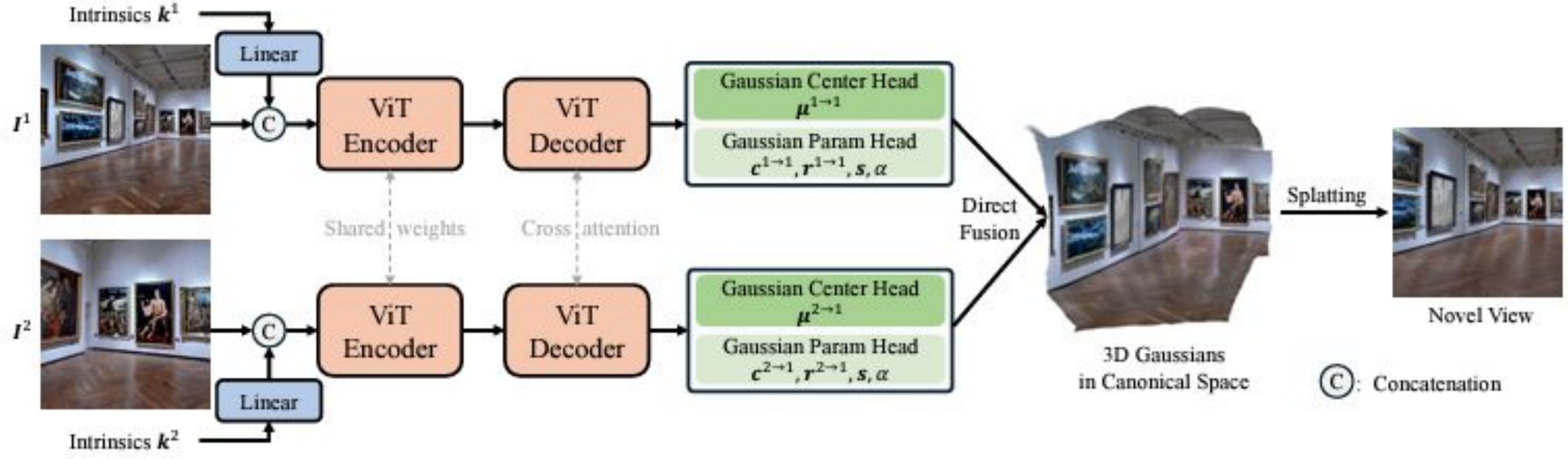
MVSplat vs GS-LRM



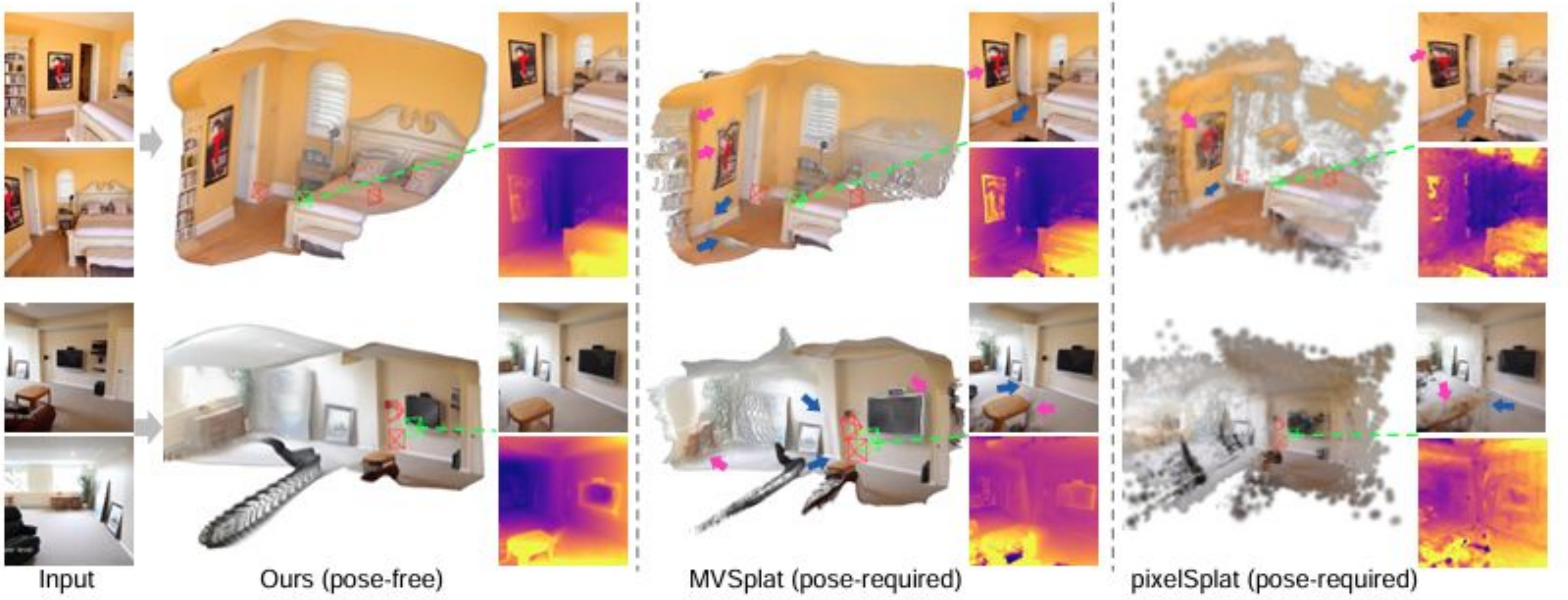


MVSplat vs NoPoSplat





MVSplat vs NoPoSplat



Hacker

Davi

PhD Student

Veronika

Problem: MVSplat might be less effective on non-Lambertian and reflective surfaces

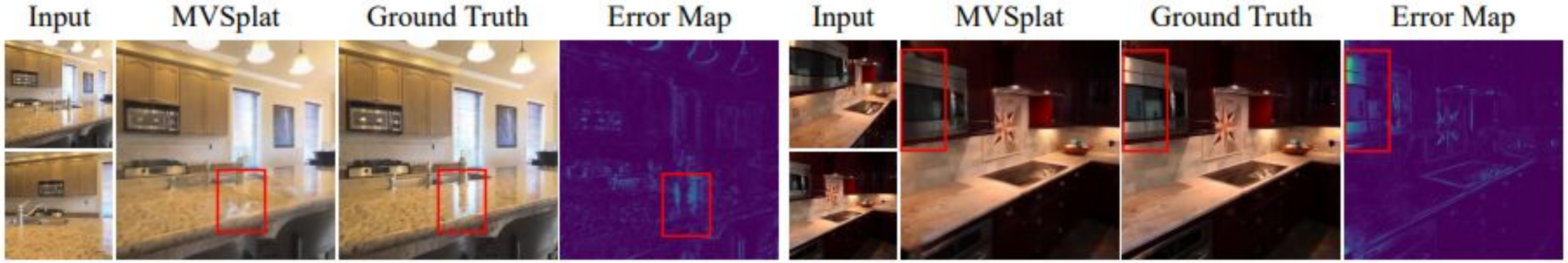


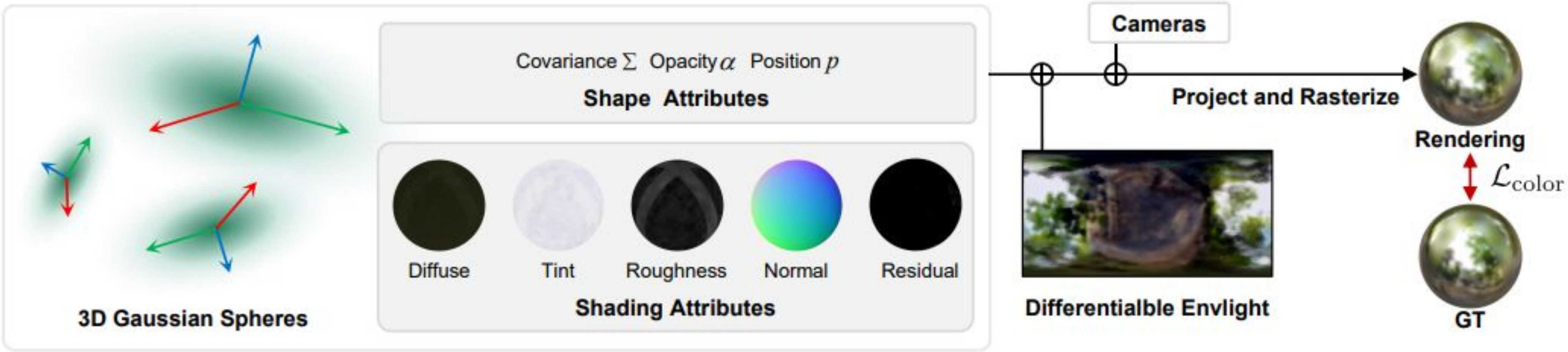
Fig. B: Failure cases. Our MVSplat might be less effective on the non-Lambertian and reflective surfaces.

Project: MVSplat with Shading Functions for Reflective Surfaces

- train the model with more diverse datasets
- combine with article GaussianShader: 3D Gaussian Splatting with Shading Functions for Reflective Surfaces

GaussianShader

- novel method that applies a simplified shading function on 3D Gaussians to enhance the neural rendering in scenes with reflective surfaces while preserving the training and rendering efficiency



Research plan

- incorporate GaussianShader code into MVSplat
- train on various datasets:
  a. NeRF Synthetic
  b. reflective objects datasets: Shiny Blender and Glossy Synthetic
  c. real-world large-scale scenes: Tanks and Temples
  d. datasets previously used
- compare results and, if necessary, adjust models
- check if MVSplat loses effectiveness significantly

References

[1] CVPR Reviewer Guidelines CVPR 2024 Reviewer Tutorial Slides