Neural 360

The State of the Art

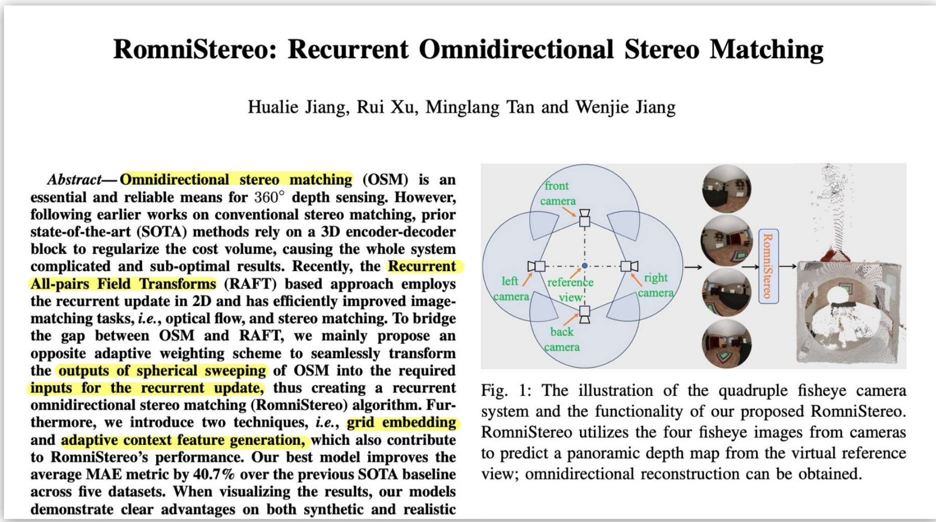
Luiz Velho
IMPA

Topics

- Neural 360 Stereo
  - Multiview 3D Reconstruction
- Neural 360 Splat
  - Differentiable Volumetric Rendering
- 360 CNNs
  - Omnidirectional Image Analysis
- Neural 360 Interiors
  - Environment Scene Construction
- Neural 360 Diffusion
  - Generative Content Creation

RomniStereo

Paper



RomniStereo: Recurrent Omnidirectional Stereo Matching

Hualie Jiang, Rui Xu, Minglang Tan and Wenjie Jiang

***Abstract*— Omnidirectional stereo matching (OSM) is an essential and reliable means for 360° depth sensing. However, following earlier works on conventional stereo matching, prior state-of-the-art (SOTA) methods rely on a 3D encoder-decoder block to regularize the cost volume, causing the whole system complicated and sub-optimal results. Recently, the Recurrent All-pairs Field Transforms (RAFT) based approach employs the recurrent update in 2D and has efficiently improved image-matching tasks, *i.e.*, optical flow, and stereo matching. To bridge the gap between OSM and RAFT, we mainly propose an opposite adaptive weighting scheme to seamlessly transform the outputs of spherical sweeping of OSM into the required inputs for the recurrent update, thus creating a recurrent omnidirectional stereo matching (RomniStereo) algorithm. Furthermore, we introduce two techniques, *i.e.*, grid embedding and adaptive context feature generation, which also contribute to RomniStereo's performance. Our best model improves the average MAE metric by 40.7% over the previous SOTA baseline across five datasets. When visualizing the results, our models demonstrate clear advantages on both synthetic and realistic**

Fig. 1: The illustration of the quadruple fisheye camera system and the functionality of our proposed RomniStereo. RomniStereo utilizes the four fisheye images from cameras to predict a panoramic depth map from the virtual reference view; omnidirectional reconstruction can be obtained.

- IEEE Robotics and Automation Letters (Volume: 9, Issue: 3, March 2024)

Neural Network

- Architecture

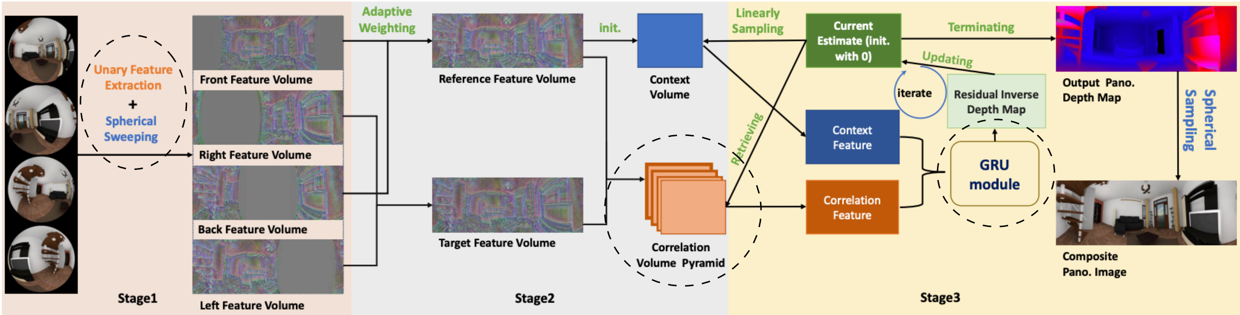
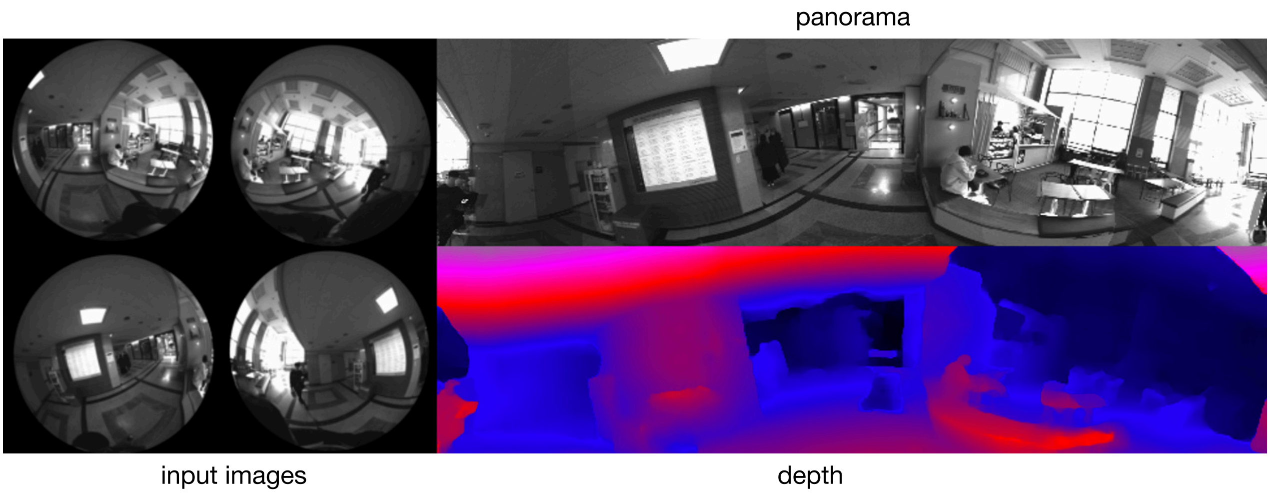


Fig. 2: Our Proposed Recurrent Omnidirectional Stereo Matching Framework.

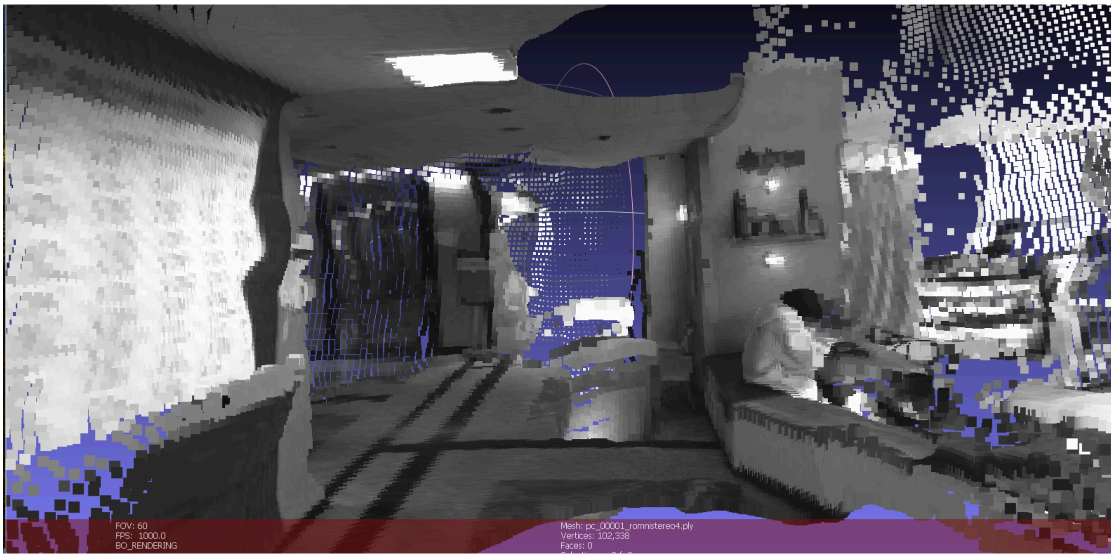
Test

- Interior Scene



Results

- 3D Point Cloud



Splatter 360

Paper

Splatter-360: Generalizable 360 Gaussian Splatting for Wide-baseline Panoramic Images

Zheng Chen[1*], Chenming Wu[2*], Zhelun Shen[2], Chen Zhao[2], Weicai Ye[3], Haocheng Feng[2], Errui Ding[2], Song-Hai Zhang[1†]

[1]Tsinghua University, [2]Baidu VIS, [3]Zhejiang University

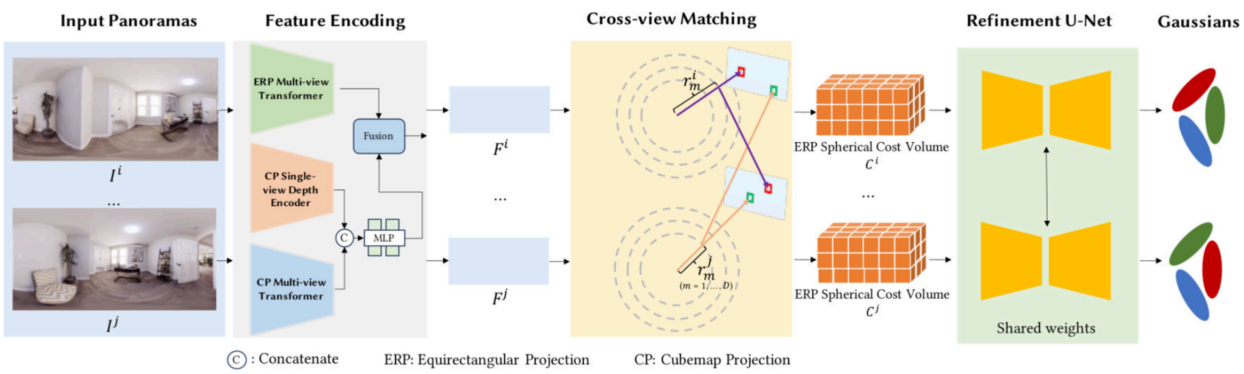
Accepted to CVPR 2025

[*]Equal contribution [†]Corresponding author

Paper arXiv Code

- Project Page

Method

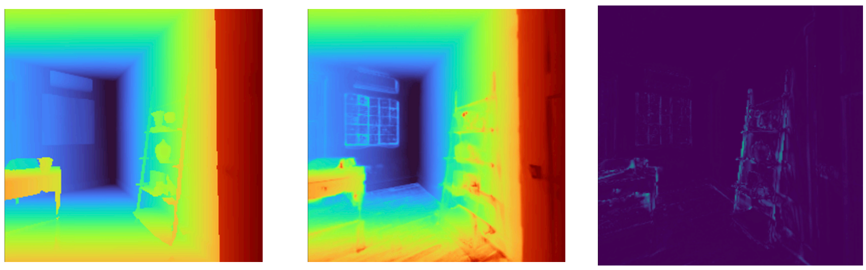


Our Splatter-360 processes 360° panoramic images using a bi-projection encoder that extracts features from both equirectangular projection (ERP) and cube-map projection (CP) through multi-view transformers. These features are used for spherical cost volume construction, and multi-view matching is performed between the reference and source views in spherical space. Next, a refinement U-Net is applied to enhance the spherical cost volume, yielding refined cost volumes and more accurate spherical depth estimations. These refined outputs are then fed into the Gaussian decoder, which produces pixel-aligned Gaussian primitives for synthesizing novel views.

Apartment Scene

- Image Detail

- Depth



gt gen error

Generated Panorama

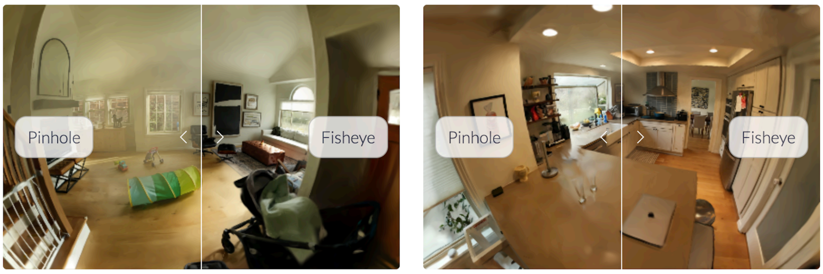
Ground Truth

3D GRUT

Paper

3DGUT: Enabling Distorted Cameras and Secondary Rays in Gaussian Splatting

Qi Wu[1,*] Janick Martinez Esturo[1,*] Ashkan Mirzaei[1,2]
Nicolas Moenne-Loccoz[1] Zan Gojcic[1]
[1] NVIDIA [2] University of Toronto [*] Equal Contribution

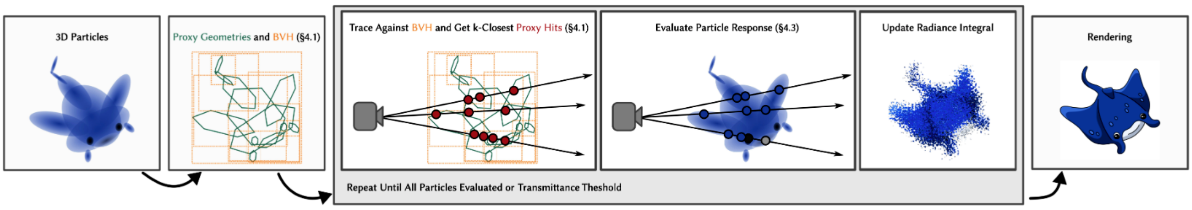


We extend 3D Gaussian Splatting (3DGS) to support nonlinear camera projections and secondary rays for simulating phenomena such as reflections and refractions. By replacing EWA splatting rasterization with the Unscented Transform, our approach retains real-time efficiency while accommodating complex camera effects like rolling shutter.

- CVPR 2025

Method

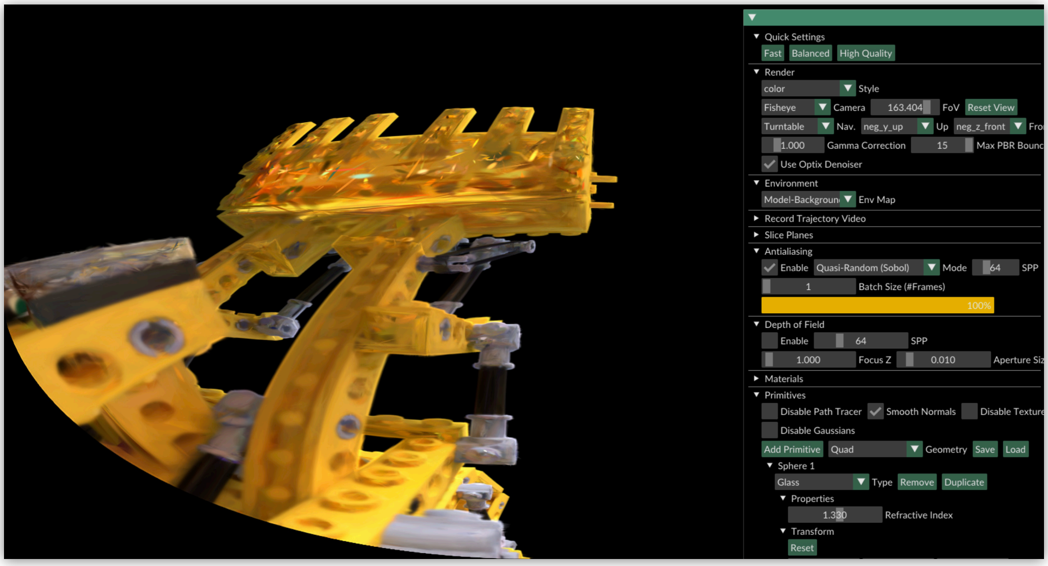
- Pipeline



Given a set of 3D particles, we first build the corresponding bounding primitives and insert them into a BVH. To compute the incoming radiance along each ray, we trace rays against the BVH to get the next k particles. We then compute the intersected particles' response and accumulate the radiance. The process repeats until all particles have been evaluated or the transmittance meets a predefined threshold and the final rendering is returned.

Lego Model

- GRUT Playground



DarSwin Unet

Paper

DarSwin-Unet: Distortion Aware Encoder-Decoder Architecture

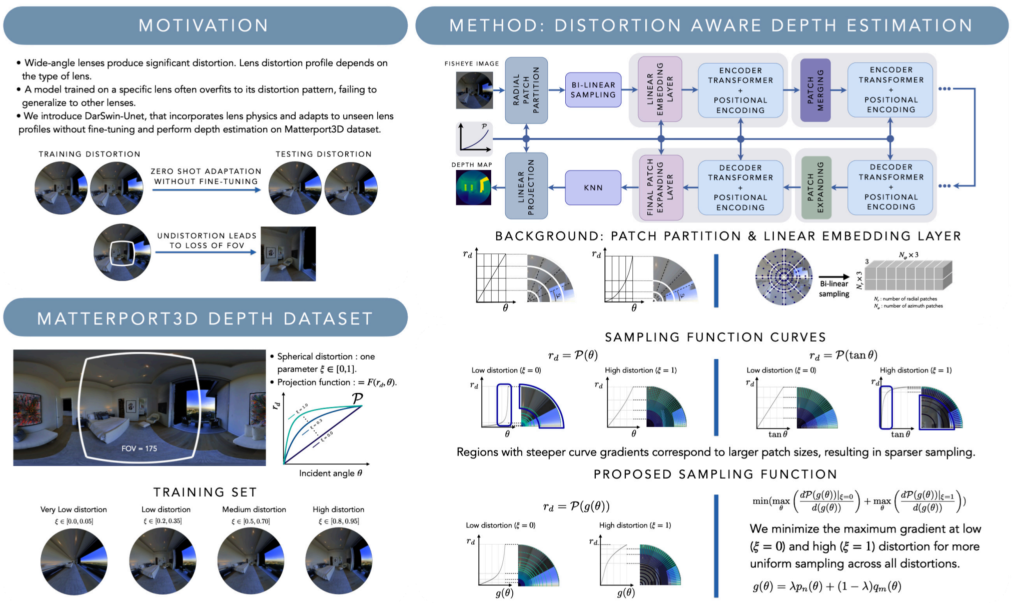
Akshaya Athwale   Ichrak Shili   Émile Bergeron   Ola Ahmad
Jean-François Lalonde

UNIVERSITÉ LAVAL   THALES Building a future we can all trust

Wide-angle fisheye images are becoming increasingly common for perception tasks in applications such as robotics, security, and mobility (e.g. drones, avionics). However, current models often either ignore the distortions in wide-angle images or are not suitable to perform pixel-level tasks. In this paper, we present an encoder-decoder model based on a radial transformer architecture that adapts to distortions in wide-angle lenses by leveraging the physical characteristics defined by the radial distortion profile. In contrast to the original model, which only performs classification tasks, we introduce a U-Net architecture, DarSwin-Unet, designed for pixel level tasks. Furthermore, we propose a novel strategy that minimizes sparsity when sampling the image for creating its input tokens. Our approach enhances the model capability to handle pixel-level tasks in wide-angle fisheye images, making it more effective for real-world applications. Compared to other baselines, DarSwin-Unet achieves the best results across different datasets, with significant gains when trained on bounded levels of distortions (very low, low, medium, and high) and tested on all, including out-of-distribution distortions. We demonstrate its performance on depth estimation and show through extensive experiments that DarSwin-Unet can perform zero-shot adaptation to unseen distortions of different wide-angle lenses.

- WACV 2025

Method



Project

- Code

1. Download pre-trained swin transformer model

- [Get pre-trained model in this link for grp1] (https://hdrdb-public.s3.valeria.science/darswin-unet//home-local2/akath.extra.nobkp/DarSwin-Unet/gr1/epoch_499_1.pth)
- [Get pre-trained model in this link for grp2] (https://hdrdb-public.s3.valeria.science/darswin-unet//home-local2/akath.extra.nobkp/DarSwin-Unet/gr2/epoch_499_2.pth)
- [Get pre-trained model in this link for grp3] (https://hdrdb-public.s3.valeria.science/darswin-unet//home-local2/akath.extra.nobkp/DarSwin-Unet/gr3/epoch_499_3.pth)
- [Get pre-trained model in this link for grp4] (https://hdrdb-public.s3.valeria.science/darswin-unet//home-local2/akath.extra.nobkp/DarSwin-Unet/gr4/epoch_499_4.pth)

2. Environment

- Please prepare an environment with python=3.7, and then use the command "pip install -r requirements.txt" for the dependencies.

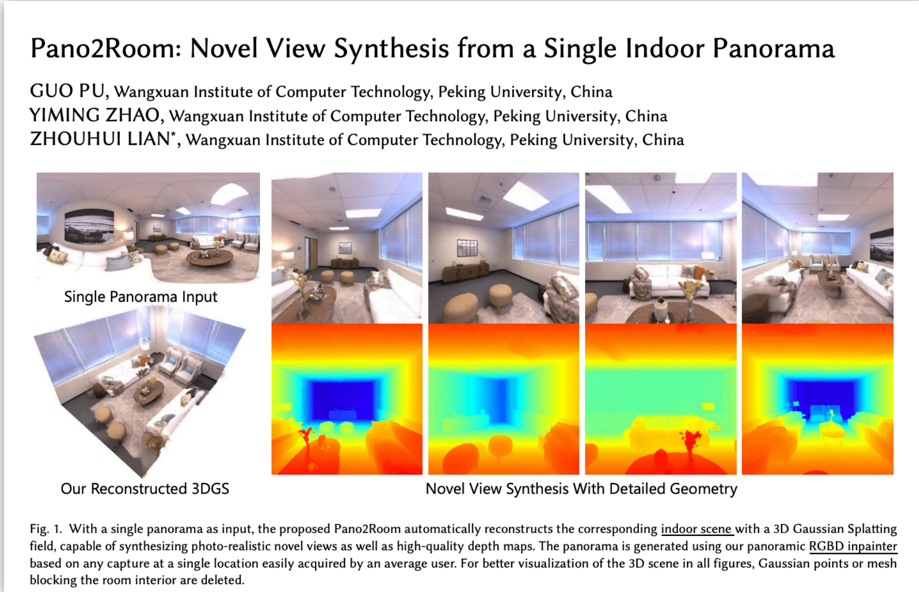
2. Training

- Knn matrix generation : Use matrix.sh to generate a KNN matrix for your desired dataset (with distortion parameter) to determine inverse projection from polar to cartesian
- Run ./train.sh

```
python train.py --dataset Synapse --cfg configs/exp1_64_1.yaml --grp gp1\
    --root_path $SLURM_TMPDIR/data/M3D_low --max_epochs 501 \
    --output_dir  /home/prongs/scratch/darswin_g_theta/grp1_resume \
```

Pano2Room

Paper



Pano2Room: Novel View Synthesis from a Single Indoor Panorama

GUO PU, Wangxuan Institute of Computer Technology, Peking University, China
YIMING ZHAO, Wangxuan Institute of Computer Technology, Peking University, China
ZHOUHUI LIAN*, Wangxuan Institute of Computer Technology, Peking University, China

Fig. 1. With a single panorama as input, the proposed Pano2Room automatically reconstructs the corresponding indoor scene with a 3D Gaussian Splatting field, capable of synthesizing photo-realistic novel views as well as high-quality depth maps. The panorama is generated using our panoramic RGBD inpainter based on any capture at a single location easily acquired by an average user. For better visualization of the 3D scene in all figures, Gaussian points or mesh blocking the room interior are deleted.

- SIGGRAPH Asia 2024

Method

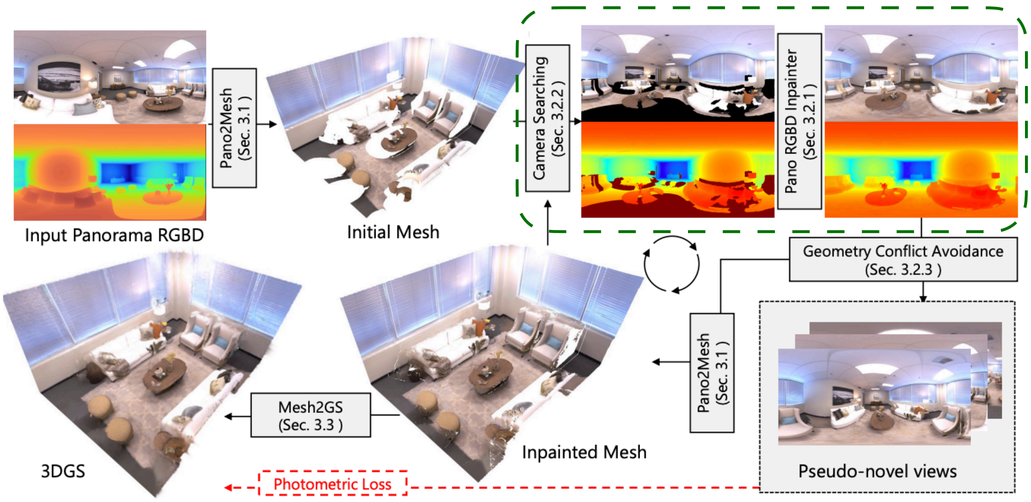


Fig. 2. An overview of Pano2Room. With a panorama as input, we first predict the geometry of the panorama using the panoramic RGBD inpainter. Then we synthesize the initial mesh using a Pano2Mesh module. Next, we iteratively search for cameras with the least view completeness, and under the searched viewpoint, we render the existing mesh to obtain panoramic RGBDs with missing areas. To complete each rendered RGBD, we use the panoramic RGBD inpainter to generate new textures and predict new geometries. The new textures/geometries are iteratively fused into the existing mesh if no geometry conflict is introduced. Finally, the inpainted mesh is converted to a 3DGS and trained with collected pseudo novel views.

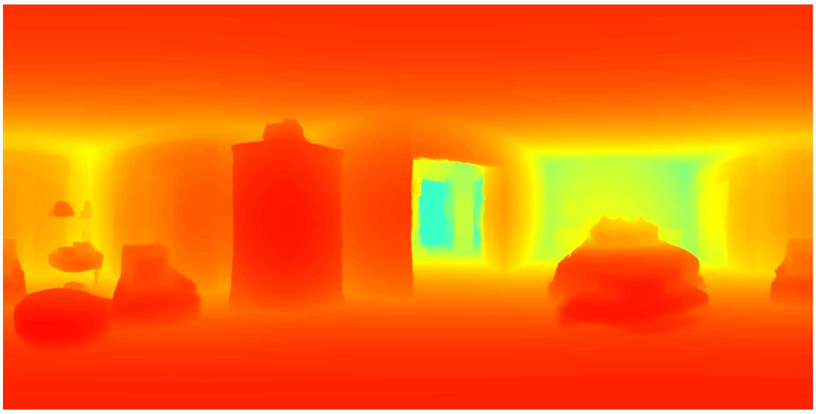
360 Panorama

- Inpainted

- Rendered 3DGS

Depth

- Inpainted



- Rendered 3DGS

Matrix-3D

Paper

Matrix-3D: Omnidirectional Explorable 3D World Generation

Zhongqi Yang[1*], Wenhang Ge[2*], Yuqi Li[1,3*], Jiaqi Chen[1†], Haoyuan Li[1†], Mengyin An[1], Fei Kang[1], Hua Xue[1], Baixin Xu[1], Yuyang Yin[1], Eric Li[1], Yang Liu[1], Yikai Wang[4], Hao-Xiang Guo[1‡], Yahui Zhou[1]

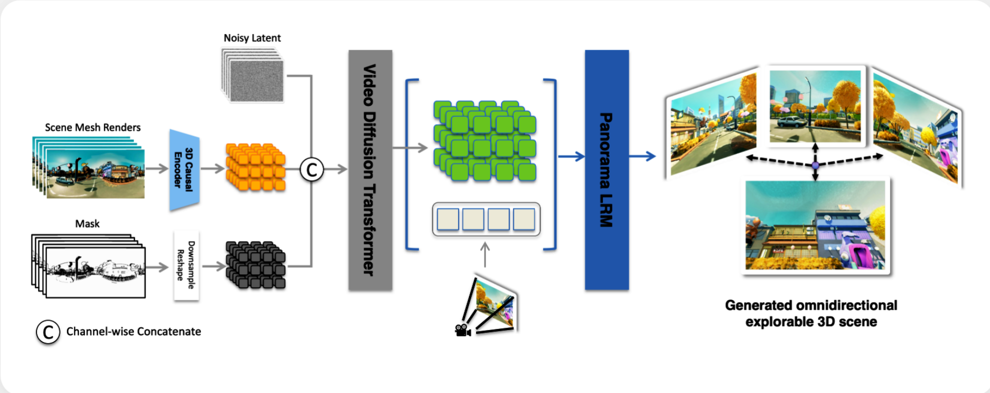
[1]*Skywork AI*
[2]*Hong Kong University of Science and Technology (Guangzhou)*

Matrix-3D utilizes panoramic representation for wide-coverage omnidirectional explorable 3D world generation that combines conditional video generation and panoramic 3D reconstruction.

- arXiv 2025

Method

Overview of Matrix-3D

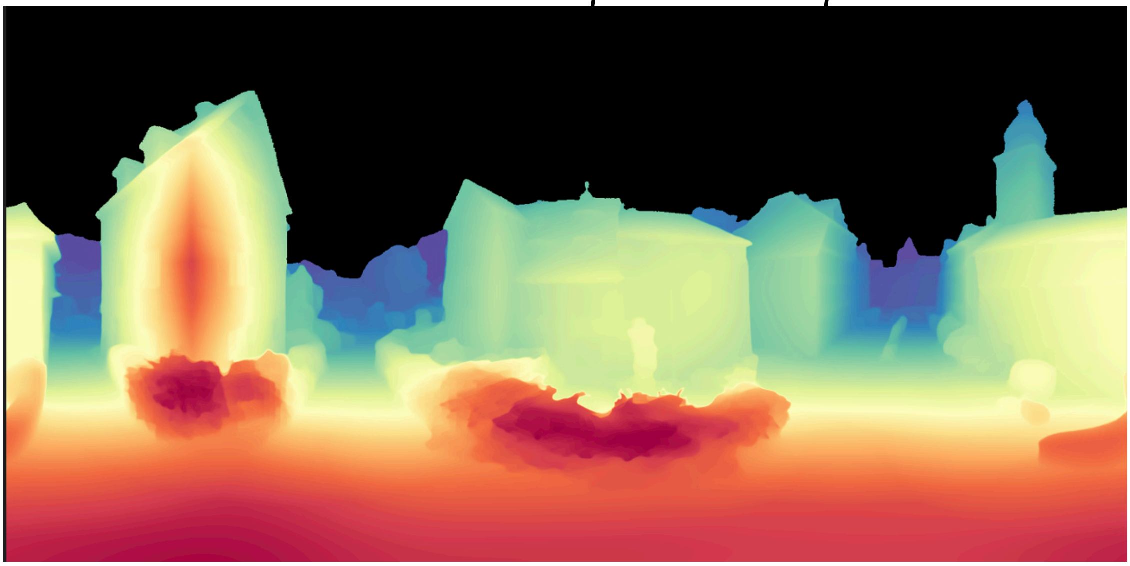


Given trajectory guidance in the form of scene mesh renderings and corresponding masks—obtained by rendering an estimated mesh along a user-defined camera trajectory—we train an image-to-video diffusion model to generate high-quality panoramic videos that precisely follow the specified trajectory. The generated 2D panoramic content is then lifted into an omnidirectional, explorable 3D world using a large-scale panorama reconstruction model.

Generated Panorama

- *Prompt: a medieval village, half-timbered houses, cobblestone streets, lush greenery, clear blue sky, detailed textures, vibrant colors, high resolution.*

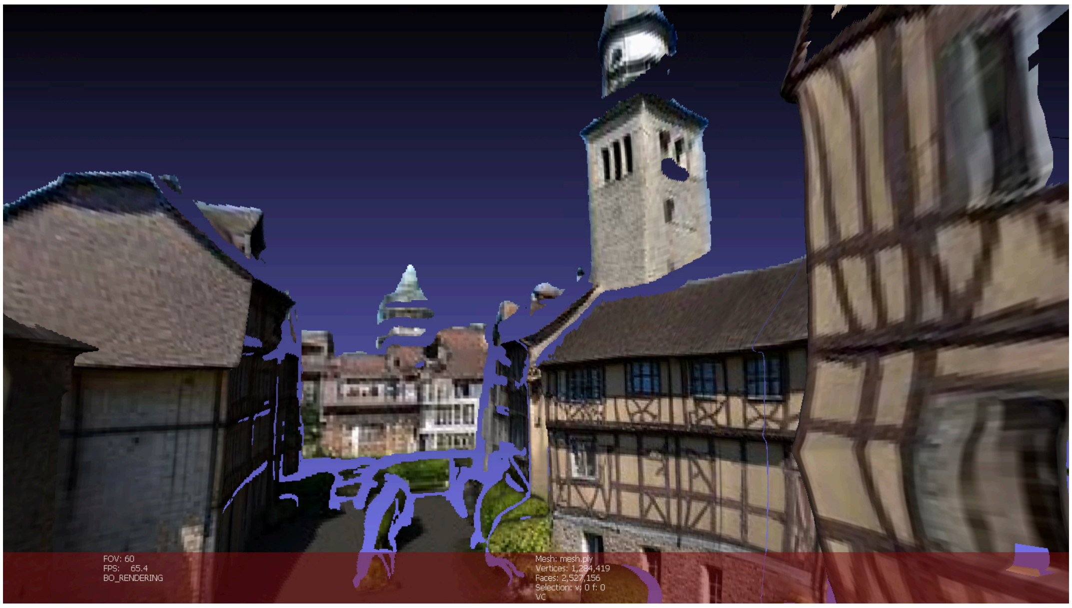
Scene Depth Map



- ***Prompt:*** *a medieval village, half-timbered houses, cobblestone streets, lush greenery, clear blue sky, detailed textures, vibrant colors, high resolution.*

3D Mesh Model

3D View



Video

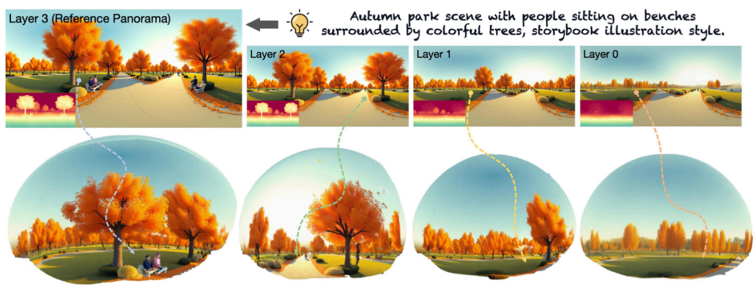
Other Papers

Layer Pano 3D

LayerPano3D: Layered 3D Panorama for Hyper-Immersive Scene Generation(SIGGRAPH 2025)

This repository is the official implementation of LayerPano3D

Project page | **Paper** | **Dataset** | **Huggingface**



Shuai Yang*, Jing Tan*, Mengchen Zhang, Tong Wu, Yixuan Li, Gordon Wetzstein, Ziwei Liu, Dahua Lin

VGGT

VGGT: Visual Geometry Grounded Transformer

Jianyuan Wang[1, 2], Minghao Chen[1, 2], Nikita Karaev[1, 2]
Andrea Vedaldi[1, 2], Christian Rupprecht[1], David Novotny[2]
[1]Visual Geometry Group, University of Oxford, [2]Meta AI

CVPR 2025 Best Paper Award

Paper Code Demo Slides