Modern Computer Vision Methods

Introduction Meeting

for WS 2024/25 [IN2107]

Dr. Benjamin Busam, Changxuan Li, Christian Kapeller, Diego Biagini, Felix Tristram, Mert Karaoglu, Mert Kiray, Sen Wang

MCVM Team

Felix
Tristram

Sen
Wang

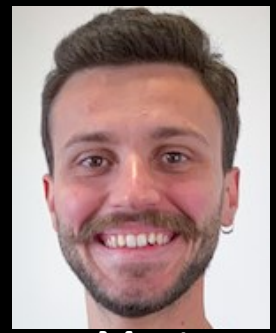
Diego
Biagini

Benjamin
Busam

Christian
Kapeller

Changxuan
Li

Mert
Kiray



Mert
Karaoglu

MCVM

Course Structure

Course Dates

15.10.2024	**Introduction Session**
	Individual Tutor Meetings
26.11.2024	**Presentation Training**
	Individual Tutor Meetings
10.12.2024	**Invited Research Talk**
	Individual Tutor Meetings
07.01.2025	Presentation Slot I
14.01.2025	Presentation Slot II
21.01.2025	Presentation Slot III

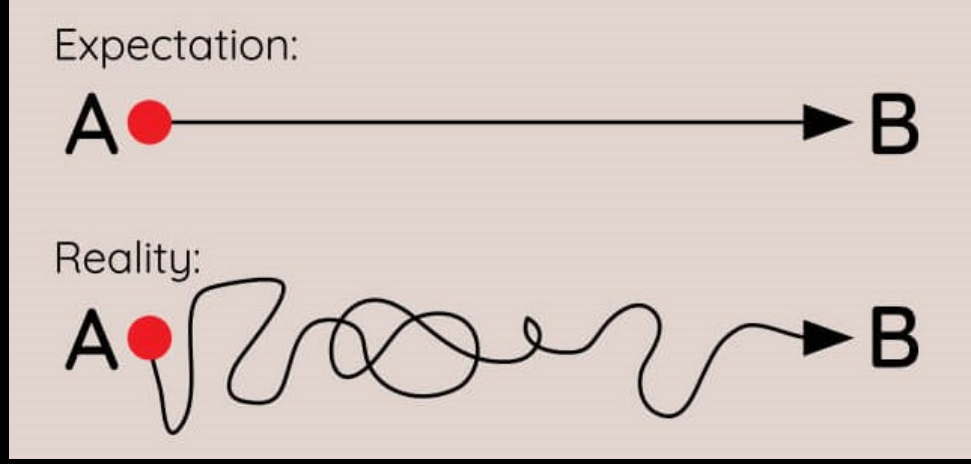
Paper Overview

Authors	Title	Source	Year	Supervisor
Attal, Verbin, Mildenhall, Hedman, Barron, O'Toole, Srinivasan	Flash Cache: Reducing Bias in Radiance Cache Based Inverse Rendering	ECCV	2024	Felix
Xing, Xia, Zhang, Chen, Wang, Wong, Shan	Dynamicrafter: Animating open-domain images with video diffusion priors	ECCV	2024	Diego
Karaev, Rocco, Graham, Neverova, Vedaldi, Rupprecht	Cotracker: It is better to track together	ECCV	2024	Mert Kar.
Fu, Liu, Kulkarni, Kautz, Efros, Wang	COLMAP-Free 3D Gaussian Splatting	CVPR	2024	Mert Kiray
Charatan, Li, Tagliasacchi, Sitzmann	PixelSplat: 3d gaussian splats from image pairs for scalable generalizable 3d reconstruction	CVPR	2024	Changxuan
Xiang, Li, Cheng, Lai, Zhang, Liao, Zeng, Liu	GaussianRoom: Improving 3D Gaussian Splatting with SDF Guidance and Monocular Cues for Indoor Scene Reconstruction	arXiv	2024	Sen
Peng, Tang, Zhou, Wang, Liu, Li, Chellappa	BAGS: Blur Agnostic Gaussian Splatting through Multi-Scale Kernel Modeling	ECCV	2024	Christian

In Person / Virtual – Hybrid

- Generally onsite
- In exceptional cases: virtual via zoom
- Tuesdays at 4pm in MI 03.13.010

What we expect from you



- Interest in Computer Vision
- Independent and pro-active participation
- Actively asking for help [supervisor meetings]
- Coding knowledge
- Illustrating methods with examples / demos

Goals

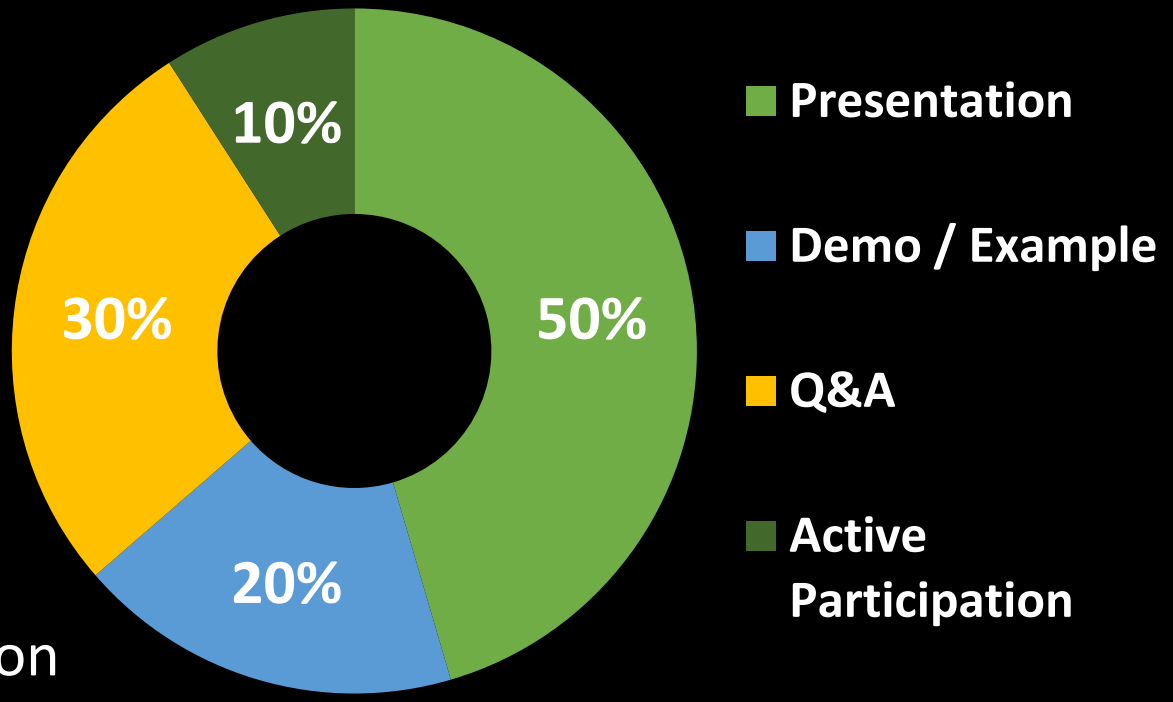
- Scientifically Learning about…
  - State-of-the-art Computer Vision
  - Current research challenges and applications
  - Communicate / discuss on most recent advantages with expert scientists
  - Hands-on experience with available code bases

- Skill training of…
  - Reading / understanding of a scientific work
  - Get overview of scientific field through literature research
  - Research talk in front of an audience, related Q&A

Presentation

- Presentation: 20 ± 2 minutes talk + 10-15 minutes Q&A
- Content should cover
  - Introduction / Relevance of Problem
  - Context / Related Work
  - Main Contribution(s)
  - Experimental Results
    - Hands-on experience with code
  - Discussion
  - Future Work
- Presentation should be self-contained, send slides 2 weeks before
- Attend all talks + active participation in other discussions

Evaluation Criteria



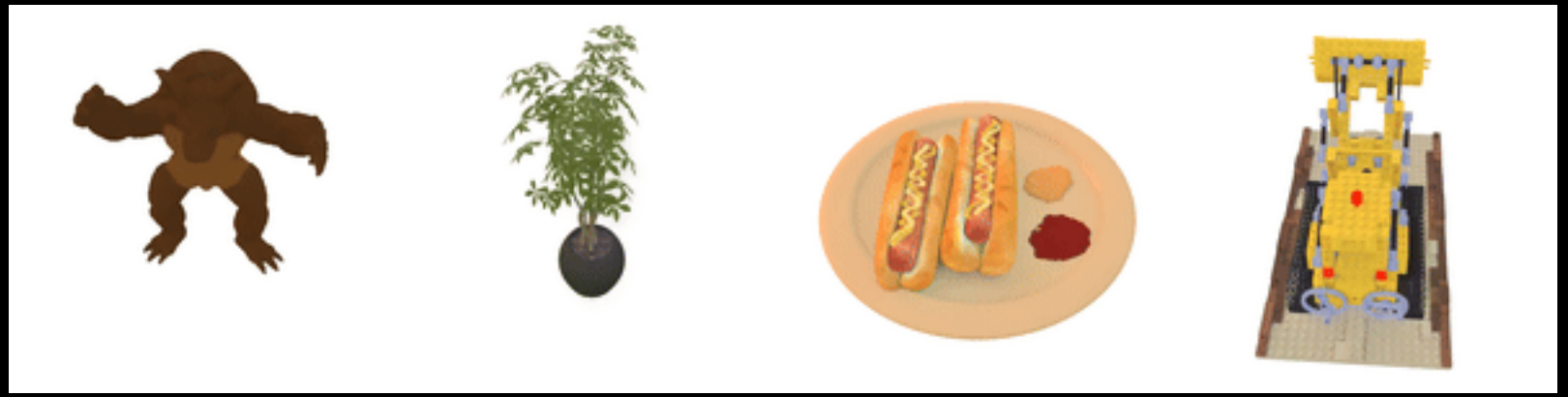
- Quality of Presentation
  - Scientific Content of the Talk + Preparation
  - Quality of the Slides
  - Putting the Topic in Context (Related Work)
- Examples / Hands-on Code
- Scientific Discussion (Q&A)
- Independent Interaction / Active Participation in the Course

Seed Paper Intros

MCVM WiSe 24/25

Flash Cache: Reducing Bias in Radiance Cache Based Inverse Rendering [Felix]

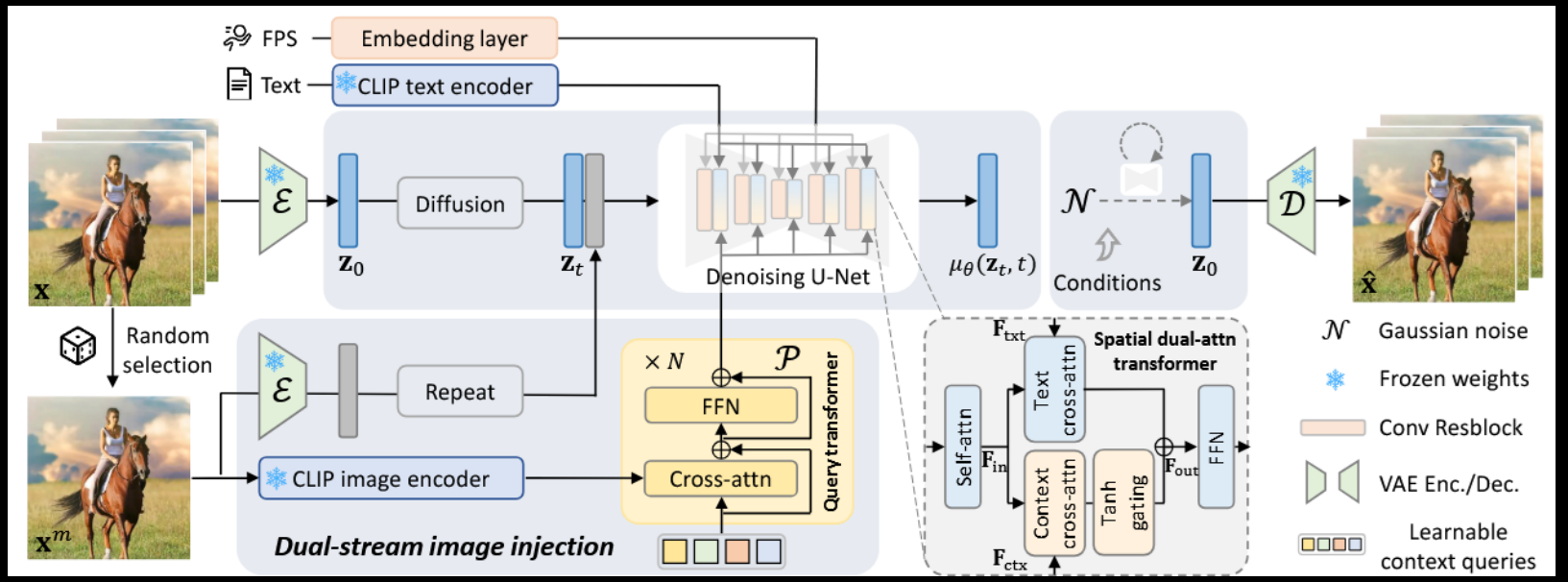
- Reconstruct a scenes geometry, materials and lighting from observed images
- hard to model incoming light in current methods
- Flash Cache uses fast caching and smart secondary rays interactions to speed up this learning



Attal, B., Verbin, D., Mildenhall, B., Hedman, P., Barron, J. T., O'Toole, M., & Srinivasan, P. P. (2024). Flash Cache: Reducing Bias in Radiance Cache Based Inverse Rendering. arXiv preprint arXiv:2409.05867.

DynamiCrafter [Diego]

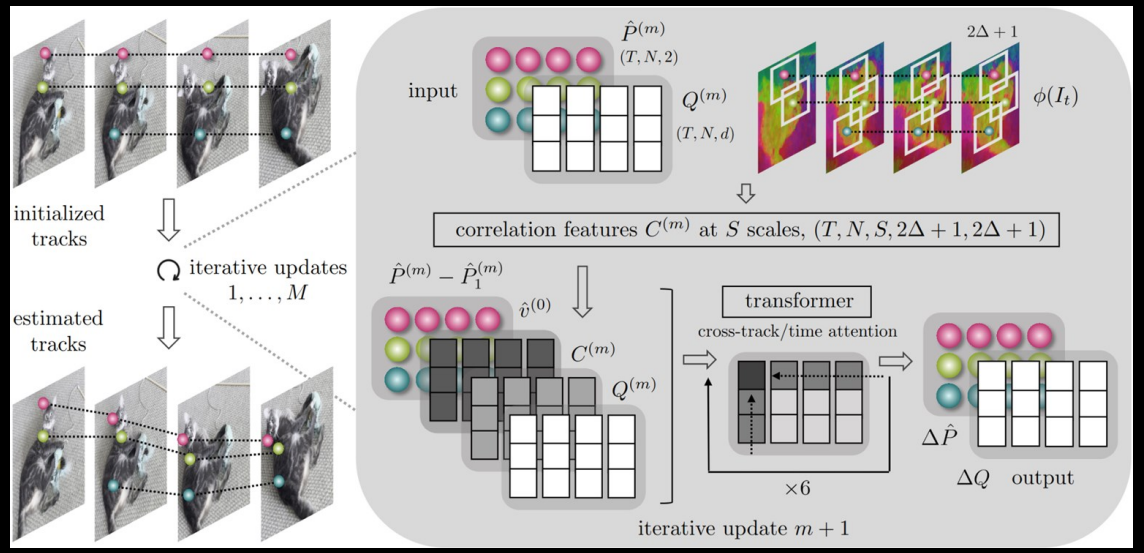
- Diffusion-based video model for Image2Video
- Injects image conditioning through dual-stream mechanism, first projecting image in text-aligned embedding space
- Multi stage training strategy(based on T2I as well)



Xing, J., Xia, M., Zhang, Y., Chen, H., Wang, X., Wong, T. T., & Shan, Y. (2023). Dynamicrafter: Animating open-domain images with video diffusion priors. ECCV 2024

CoTracker [Mert]

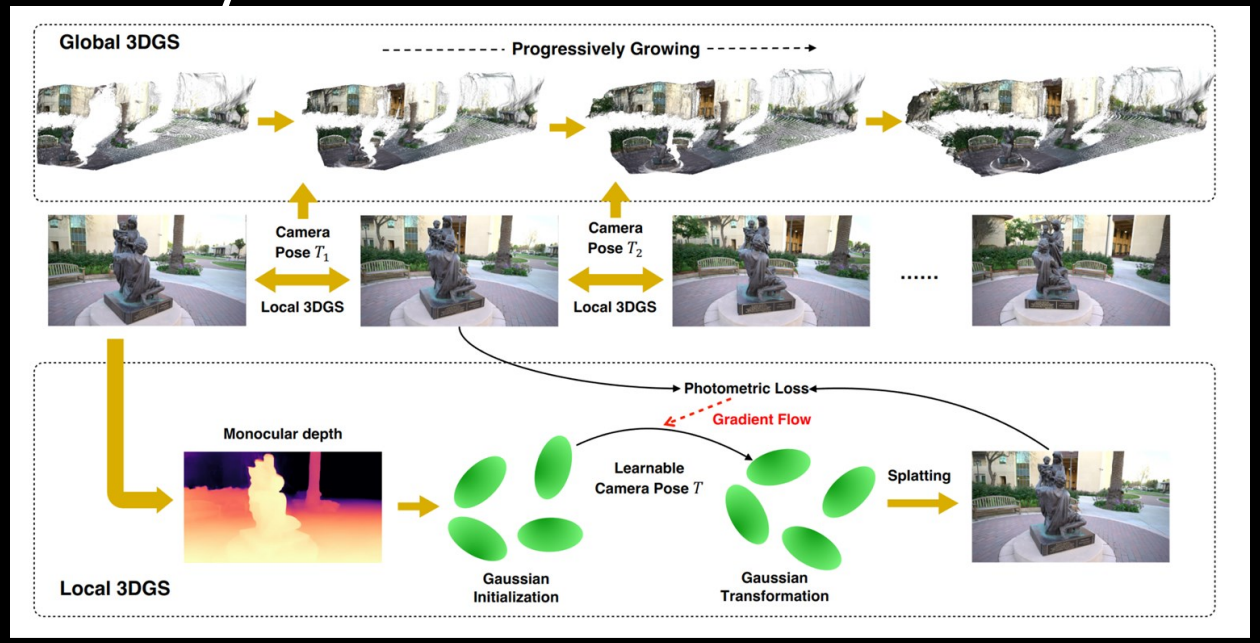
- Long-term point tracking method from RGB videos
- Processes tracklets in temporal windows
- Utilizes cross-track and -time attention for spatial and temporal causality



Karaev, Nikita, Ignacio Rocco, Benjamin Graham, Natalia Neverova, Andrea Vedaldi, and Christian Rupprecht. "Cotracker: It is better to track together." ECCV (2024).

COLMAP-Free 3D Gaussian Splatting [Mert]

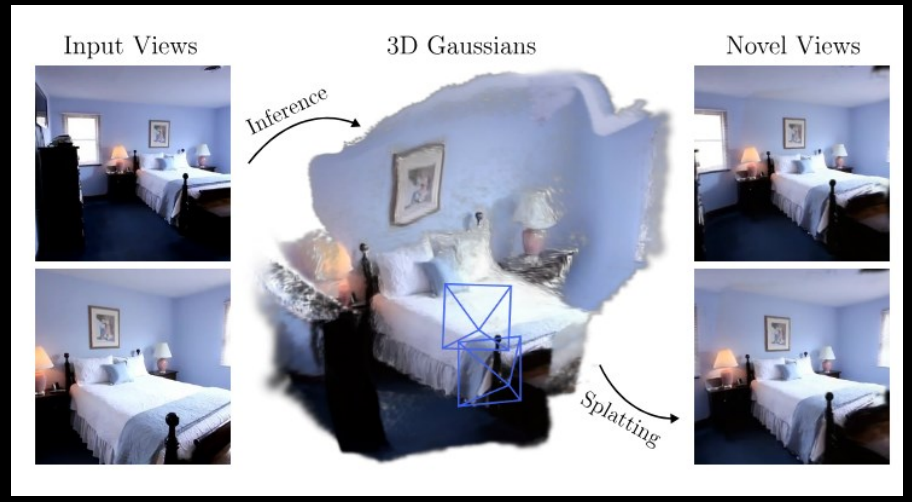
- Progressively grow the 3D Gaussians without the need to pre-compute the camera poses
- Pose estimation by 3D Gaussian transformation in frames t-1 and t

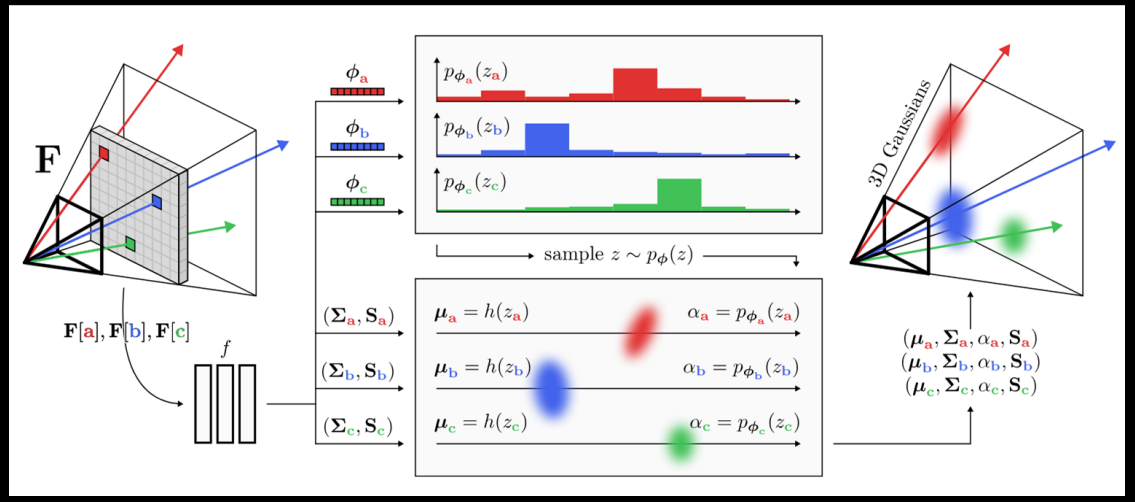


Fu, Yang, Sifei Liu, Amey Kulkarni, Jan Kautz, Alexei A. Efros, and Xiaolong Wang. 'COLMAP-Free 3D Gaussian Splatting'. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 20796–805, 2024.

PixelSplat [Changxuan]

- Feedforward scene reconstruction w/ pair of images by 3DGS
- Scale factor inferred by multiview epipolar transformer
- Dense Gaussian distribution and differentiable sampling to avoid local minima

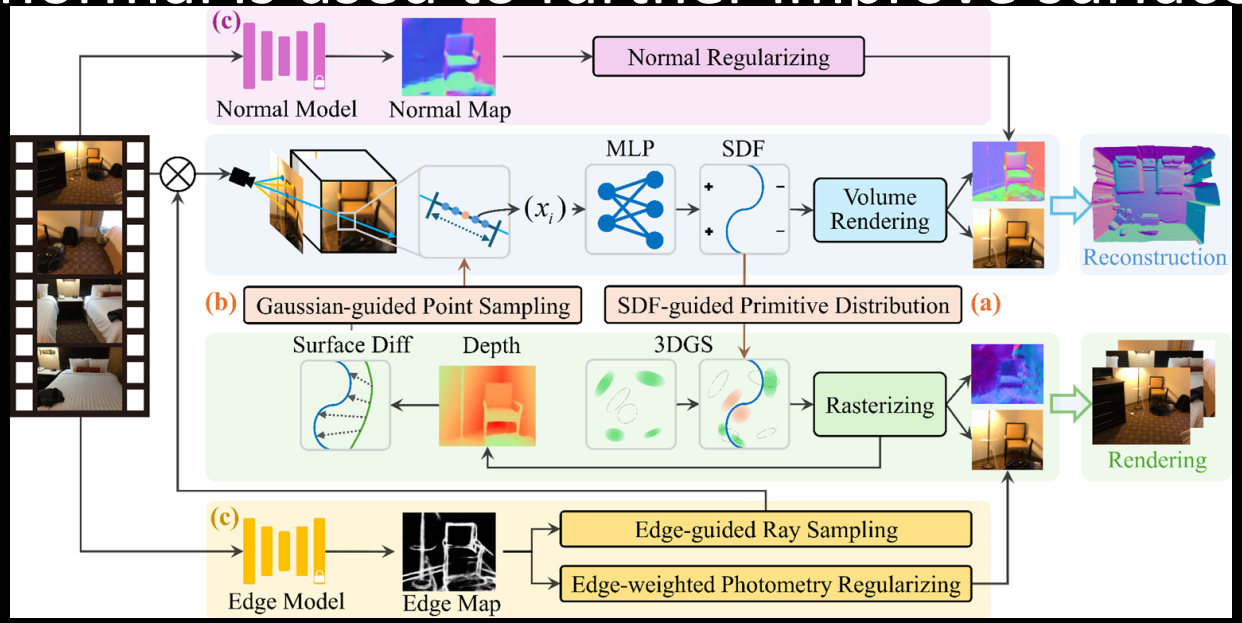




Charatan, David, Sizhe Lester Li, Andrea Tagliasacchi, and Vincent Sitzmann. "pixelsplat: 3d gaussian splats from image pairs for scalable generalizable 3d reconstruction." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 19457-19467. 2024.

GaussianRoom [Sen]

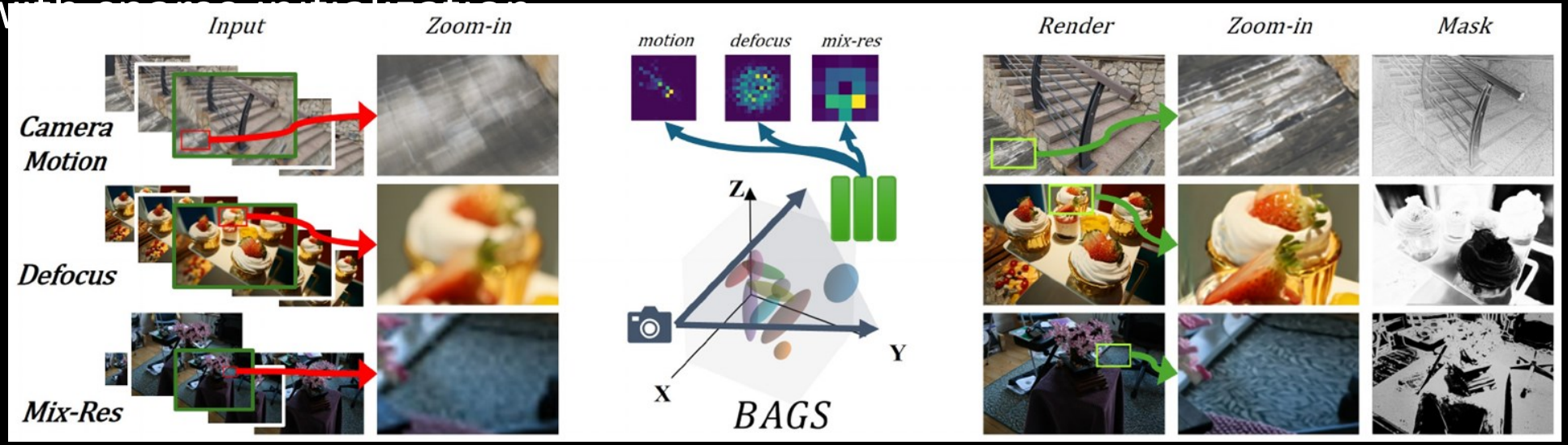
- Integrate neural SDF within 3DGS and form a positive cycle.
- Geometry from the SDF will constrain the Gaussian primitives.
- The rasterized depth from Gaussian provide constraint for SDF.
- Additional normal is used to further improve surface reconstruction.



Haodong Xiang, Xinghui Li, Kai Cheng, Xiansong Lai, Wanting Zhang, Zhichao Liao, Long Zeng, Xueping Liu. GaussianRoom: Improving 3D Gaussian Splatting with SDF Guidance and Monocular Cues for Indoor Scene Reconstruction. arXiv 2024

BAGS - Blur Agnostic Gaussian Splatting through Multi-Scale Kernel Modeling [Christian]

- 3D reconstruction from blurry images (motion, defocus, mix-res)
- Uses a proposal network that estimates per-pixel convolutional blur kernels
- Coarse-to-fine optimization scheme allows fast processing and deals w



Peng, C., Tang, Y., Zhou, Y., Wang, N., Liu, X., Li, D., & Chellappa,BAGS: Blur Agnostic Gaussian Splatting through Multi-Scale Kernel Modeling. ECCV 2024. http://arxiv.org/abs/2403.04926

Next Steps

Paper Selection

https://forms.gle/aV9bur1VyXXwxqWc9

Deadline: November 20, 2024

- We optimize for global happiness

Next Meeting(s): Individual Meetings with Tutors

- Presentation Training: Tuesday, November 26 at 4pm in MI 03.13.010

Questions

E-Mail us on

mcvm@mailnavab.informatik.tu-muenchen.de

Your MCVM Team:

Benjamin Busam, Changxuan Li, Christian Kapeller, Diego Biagini,

Felix Tristram, Mert Karaoglu, Mert Kiray, Sen Wang

Web:

https://www.cs.cit.tum.de/camp/teaching/seminars/modern-computer-vision-methods-ws-2024-25/