Project list
Hands-On AI Based 3D Vision

Intro:

In this document, you will find the list of proposed projects for the course **"Hands-On AI Based 3D Vision"** for the 2025 Summer Semester.

Each project focuses on different core topics we explored during the lectures, from **neural fields** to **generative models**, **Gaussian splatting**, and **learning-based 3D reconstruction**.

You're encouraged to pick a project that sparks your curiosity and aligns with your interests. Whether you're passionate about rendering novel views with NeRF, experimenting with point cloud transformers, or diving into diffusion-based 3D generation, there's something for you!

Creativity is key!
While each project comes with a set of objectives and a baseline, you are more than welcome to go beyond the minimum requirements. We value originality, so feel free to introduce your ideas, improvements, or extensions, especially if they're inspired by recent papers or applications.

Resources

Each project includes a list of relevant research papers and links to code repositories or datasets to get you started. These are meant to guide you, but they are not exhaustive. You're encouraged to explore further, draw inspiration from the other projects listed, and refer to external resources and tools.

Support

If you have questions or need help at any stage of your project, feel free to contact us at ta_3dvision@listserv.uni-tuebingen.de, using the subject line:

[3D Vision 2025] project assistance

Submission Instructions

To submit your work, please send an email to ta_3dvision@listserv.uni-tuebingen.de, with the subject:
[3D Vision 2025] project submission

Make sure to attach:

- A **report** following the template provided here: overleaf template (remember to fill in all the author information, and uncomment "\cvprfinalcopy" line before submitting)

- A **link to your Git repository** containing the code

Project Guidelines:

- **Team Composition:** Groups of 2 students.

- **Deliverables:** A presentation slide deck and a demonstration of the implemented project.

- **Deadlines:** 22/07/2025, 12:00 PM (noon)

The Report

The report should clearly demonstrate your understanding of the project's theoretical background and your reasoning behind the design and implementation choices. It should include:

- A brief **motivation and plan of attack**

- Key **implementation details**

- A thoughtful **discussion of results**

Don't just show what you did. Explain *why* you did it, and what insights you gained. Reference topics from the course or your own additional research to support your conclusions.

Project 1: *Shape Encoder for Hyunuan3D-2*

Theme: 3d Object Generation

Project Description:
Hunyuan3D-2 is a state-of-the-art model for 3D object generation. While the codebase provides a powerful pipeline for generating high-quality 3D assets, it does not include the shape encoder used during training. According to the paper, the authors rely on 3DShape2Vec as the base for this component.

The goal of this project is to bridge this gap by implementing and training a shape encoder that maps 3D shapes back to their corresponding latent representations. This will allow a full encode-decode pipeline and enable applications like reconstruction, editing, and embedding-based retrieval.

Key Objectives:

- Implement and integrate a 3D shape encoder based on 3DShape2Vec.
- Generate a dataset of (latent code, 3D shape) pairs using the Hunyuan3D-2 generator.
- Train and evaluate the encoder to regress latent codes from shapes.

Suggested Approach:

- Use the Hunyuan3D-2 pipeline to generate a dataset of 3D shapes along with their corresponding latent vectors.
- Format and preprocess the data to be compatible with the 3DShape2Vec architecture.
- Train the encoder on this dataset and evaluate it using reconstruction fidelity (e.g., Chamfer Distance between generated shapes and re-encoded outputs).
- Optionally test applications such as shape interpolation or latent space clustering.

Recommended Reading:

- https://github.com/Tencent-Hunyuan/Hunyuan3D-2
- https://arxiv.org/abs/2501.12202
- [3DShape2Vec paper or repo – insert link here]

Tools & Datasets:

- Hunyuan3D-2 codebase [GitHub link]
- 3DShape2Vec architecture [link to paper/code]
- ShapeNet (optional, for generalization testing)

Expected Deliverables:

- A functional shape encoder that can regress latent codes from generated shapes.
- A set of experiments demonstrating reconstruction ability and evaluating the encoder's accuracy.
- A slide presentation summarizing the approach, results, and insights.

Project 2: Garment Extraction from Images via 3D Reconstruction and Segmentation

Theme:
3D Clothing Reconstruction and Segmentation

Project Description:
This project aims to develop a pipeline for extracting 3D garment models from single RGB images. The approach involves two main stages:

1. **3D Avatar Reconstruction:**
   Utilize the Human-3Diffusion model to reconstruct a realistic 3D avatar from a single image. This model generates high-fidelity 3D representations, capturing detailed geometry and texture, including clothing and accessories.
2. **3D Clothing Segmentation:**
   Apply a 3D clothing segmentation model, such as CloSe-Net, to the reconstructed avatar to segment and extract individual garments. CloSe-Net is trained on the CloSe-D dataset, which includes fine-grained annotations for 18 distinct clothing categories.

By combining these stages, the project seeks to create a system capable of isolating and extracting 3D garment models from images, which can be valuable for applications in virtual try-on, digital fashion, and avatar customization.Deep Fashion3D

Key Objectives:

- Implement the Human-3Diffusion model to reconstruct 3D avatars from single images.
- Integrate the CloSe-Net segmentation model to identify and extract individual garments from the reconstructed avatars.
- Evaluate the accuracy and quality of the extracted garments in terms of geometry and segmentation fidelity.human-3diffusion

Suggested Approach:

- **Data Preparation:**
  Collect a set of images suitable for avatar reconstruction.
- **3D Reconstruction:**
  Use the Human-3Diffusion model to generate 3D avatars from the collected images.
- **Segmentation:**
  Apply the CloSe-Net model to the reconstructed avatars to segment and extract individual garments.CloSe
- **Evaluation:**
  Assess the quality of the extracted garments using metrics such as segmentation accuracy and geometric fidelity. You can also exploit existing datasets by rendering

3D assets and then test the reconstruction capabilities.

Tools & Datasets:

- Human-3Diffusion codebase: [Link to code]
- CloSe-Net and CloSe-D dataset: [Link to code and dataset]
- Find additional datasets for testing and evaluation if needed.

Expected Deliverables:

- A functional pipeline that takes a single image as input and outputs segmented 3D garment models.
- A report detailing the methodology, implementation, and evaluation results.
- A presentation summarizing the project's objectives, approach, and findings.

Recommended Reading:

- Human-3Diffusion: Realistic Avatar Creation via Explicit 3D Consistent Diffusion Models
- CloSe: A 3D Clothing Segmentation Dataset and Model

Project 3: Bridging 2D and 3D — Adapting Visual Transformers for Point Cloud Understanding

Theme:
3D Point Cloud Analysis via Adapting 2D Visual Transformers

Project Description:
This project aims to explore the adaptation of pre-trained 2D visual transformers for 3D point cloud analysis. Given the abundance of large-scale 2D image datasets and the relative scarcity of extensive 3D point cloud datasets, leveraging 2D pre-trained models can be a promising approach. The goal is to investigate methods that effectively transfer knowledge from 2D to 3D domains, enhancing tasks such as classification and segmentation of point clouds.

Key Objectives:

- Understand and implement methodologies that adapt 2D visual transformers for 3D point cloud tasks.
- Evaluate the performance of these adapted models on standard 3D datasets.

Suggested Approach:

- Study existing frameworks like Adapt PointFormer and Pix4Point, which propose mechanisms to adapt 2D transformers for 3D data.
- Implement a selected method, ensuring the adaptation process is well-understood and documented.
- Train and evaluate the adapted model on benchmark 3D datasets such as ModelNet40 or ScanObjectNN.

Expected Deliverables:

- A functional implementation of the adapted transformer model for 3D point cloud analysis.
- A comprehensive evaluation report detailing performance metrics and comparisons with baseline models.
- A presentation summarizing the methodology, results, and insights gained during the project.

Recommended Reading:

- Adapt PointFormer
- Pix4Point
- Pointnet/Pointnet++
- PointCLIP

Project 4: Enhancing NeRF Robustness in Dynamic and Occluded Environments

Theme:
Robust NeRFs in the Wild

Project Description:
Neural Radiance Fields (NeRFs) have demonstrated impressive capabilities in novel view synthesis. However, their performance can degrade in dynamic scenes or environments with significant occlusions. This project focuses on developing techniques to enhance NeRF robustness in such challenging conditions, ensuring high-quality reconstructions despite dynamic elements and occlusions. (e.g. Nerf On the Go)

Key Objectives:

- Investigate methods to improve NeRF performance in dynamic and occluded scenes.
- Implement and evaluate a selected approach on relevant datasets.

Suggested Approach:

- Review recent advancements like NeRF On-the-go and RANRAC, which propose solutions for handling dynamic scenes and occlusions.
- Implement one of these methods, focusing on its application to real-world scenarios.
- Test the implemented method on datasets featuring dynamic scenes, assessing improvements in reconstruction quality.

Expected Deliverables:

- An implemented solution enhancing NeRF robustness in challenging environments.
- An evaluation report comparing the enhanced NeRF with standard NeRF implementations.
- A presentation detailing the approach, findings, and potential areas for further improvement.

Recommended Reading:

- NeRF On-the-go: https://arxiv.org/abs/2405.18715
- RANRAC: https://arxiv.org/abs/2312.09780
- 3d-vision-transformers
- KFD-NeRF:https://arxiv.org/abs/2407.13185

Project 5: Enhancing SLAM3R with Semantic Segmentation for Improved Scene Understanding

Theme:
Monocular SLAM Enhancement

Project Description:
SLAM3R is a real-time monocular SLAM system that reconstructs dense 3D scenes from RGB videos by regressing 3D pointmaps without explicit camera pose estimation. This project aims to augment SLAM3R by integrating semantic segmentation, enabling the system to not only reconstruct geometry but also understand the semantic context of the scene. Incorporating semantic information can enhance scene understanding, facilitate object-level reconstruction, and improve the robustness of SLAM in dynamic environments.SLAM3R

Key Objectives:

- Integrate a semantic segmentation module into the SLAM3R pipeline to label reconstructed pointmaps with semantic classes.
- Evaluate the impact of semantic integration on the accuracy and robustness of SLAM3R in various environments.

Suggested Approach:

- **Semantic Segmentation Integration:**
  Incorporate a pre-trained semantic segmentation model (e.g., DeepLabV3+, Mask2Former) to process input frames and generate semantic labels.
- **Semantic Label Propagation:**
  Develop a method to project 2D semantic labels onto the 3D pointmaps generated by SLAM3R, effectively creating semantically labeled 3D reconstructions.
- **Evaluation:**
  Assess the enhanced SLAM3R system on datasets with ground truth semantic labels, measuring improvements in scene understanding and reconstruction quality.

Expected Deliverables:

- An enhanced SLAM3R system capable of producing semantically labeled 3D reconstructions in real-time.
- A comprehensive report detailing the integration process, challenges faced, and evaluation results.
- A presentation summarizing the project's objectives, methodology, and findings.

Recommended Reading:

- SLAM3R: Real-Time Dense Scene Reconstruction from Monocular RGB Videos

- DeepLabV3+: Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation
- Mask2Former: Masked-attention Mask Transformer for Universal Image Segmentation

- Segment Anything

Project 6: Unsupervised Stereo Reconstruction Using DUSt3R Without Camera Calibration

Theme:
Unsupervised Stereo Reconstruction

Project Description:
Traditional multi-view stereo (MVS) methods require known camera intrinsics and extrinsics to reconstruct 3D scenes, limiting their applicability in scenarios where such information is unavailable. DUSt3R introduces a novel approach to MVS by regressing dense 3D pointmaps directly from image pairs without prior knowledge of camera parameters. This project aims to develop a pipeline using DUSt3R for reconstructing scenes from uncalibrated image collections, enabling 3D reconstruction in more flexible and unconstrained settings.DUSt3R MUSt3R

Key Objectives:

- Implement the DUSt3R pipeline to reconstruct 3D scenes from image pairs without known camera parameters.
- Develop a global alignment strategy to integrate pairwise reconstructions into a coherent 3D model.
- Evaluate the reconstruction quality on various datasets and compare it with traditional MVS methods.

Suggested Approach:

- **DUSt3R Implementation:**
  Set up the DUSt3R framework and process image pairs to generate dense 3D point maps.
- **Global Alignment:**
  Implement the global alignment procedure proposed in DUSt3R to merge pairwise pointmaps into a unified 3D reconstruction.
- **Evaluation:**
  Test the pipeline on datasets with varying characteristics (e.g., indoor, outdoor, textured, textureless) and assess reconstruction accuracy and completeness.

Expected Deliverables:

- A functional pipeline capable of reconstructing 3D scenes from uncalibrated image collections using DUSt3R.
- An evaluation report comparing the performance of the DUSt3R-based pipeline with traditional MVS methods.
- A presentation highlighting the project's goals, methodology, and results.

Recommended Reading:

- DUSt3R: Geometric 3D Vision Made Easy
- MV-DUSt3R+: Single-Stage Scene Reconstruction from Sparse Views In 2 Seconds

Project 7:Enhancing Boundary Precision in 3D Clothing Segmentation

Theme:

3D Clothing Segmentation and Boundary Refinement

Project Description:

This project aims to improve the precision of clothing segmentation boundaries in 3D point cloud data. Traditional image-based methods, such as SAM, offer sharp boundaries but are limited by predefined classes, view inconsistencies, and multi-stage processing. Conversely, 3D point cloud segmentation provides consistent results with single feedforward inference but often suffers from blurry boundaries. By integrating contrastive boundary learning techniques, this project seeks to enhance boundary sharpness in 3D clothing segmentation, addressing the limitations of existing approaches.

Key Objectives:

1. Prepare ground truth (GT) boundary labels for segmentation tasks.
2. Implement contrastive loss mechanisms to enhance boundary delineation.
3. Compare the enhanced model's performance against existing models like CloSeNet and its CLIP-enabled variant.
4. Analyze and evaluate results across multiple datasets.
5. [Optional] Explore post-processing techniques for boundary refinement.

Suggested Approach:

- **Data Preparation:**
  - Utilize datasets such as 2K2K, 3DHumans, CloSeD, Thuman2.0 and GarmentCode.
  - Generate GT boundary labels, possibly referencing methodologies from "Push-the-Boundary" (Section 4).
- **Model Implementation:**
  - Incorporate contrastive boundary learning frameworks, drawing insights from "Contrastive Boundary Learning for Point Cloud Segmentation" .
  - Integrate contrastive loss functions to enhance feature discrimination at boundaries.[Contrastive Boundary Learning for Point Cloud Segmentation]
- **Evaluation:**

    - Benchmark the improved model against CloSeNet and its CLIP-enabled version.
    - Assess performance metrics across datasets.

- **Optional Enhancement:**

    - Implement post-processing techniques to further refine segmentation boundaries.

Expected Deliverables:

- A trained model demonstrating improved boundary precision in 3D clothing segmentation.
- Comparative analysis reports highlighting performance gains over existing models.
- A comprehensive presentation summarizing methodologies, findings, and potential future work.

Recommended Reading:

- CloSe: A 3D Clothing Segmentation Dataset and Model
- Contrastive Boundary Learning for Point Cloud Segmentation
- Push-the-Boundary: Techniques for Preparing GT Boundary Labels

Project 8: Digitalizing Real-Life Objects with State-of-the-Art 3D Reconstruction Methods

Theme:

3D Reconstruction from Real-World Images

Project Description:

This project aims to evaluate the performance of cutting-edge 3D reconstruction models when applied to real-life images captured using standard mobile phones. While recent advancements in 3D reconstruction have demonstrated impressive results using synthetic or controlled datasets, their effectiveness on real-world, casually captured images remains to be thoroughly assessed. By capturing images of everyday objects and processing them through various state-of-the-art models, this project seeks to compare their capabilities in terms of geometry accuracy, texture fidelity, and overall reconstruction quality.

Key Objectives:

1. Capture images of diverse real-world objects using a mobile phone.
2. Apply multiple state-of-the-art 3D reconstruction methods to the captured images. Evaluate and compare the reconstructed 3D models using quantitative metrics and qualitative assessments.

Suggested Approach:

- **Data Collection:**
    - Select a variety of everyday objects differing in shape, texture, and material.
    - Capture multiple images of each object from different angles using a standard mobile phone.

- **3D Reconstruction Methods:**

    - Implement and utilize the following models:
        - **3DTopia**: A two-stage text-to-3D generation model utilizing hybrid diffusion priors for high-quality asset generation.
        - **InstantMesh**: A feed-forward framework for efficient 3D mesh generation from a single image, based on the LRM/Instant3D architecture.
        - **LGM (Large Multi-View Gaussian Model)**: Designed for high-resolution 3D content creation from text or single images using multi-view Gaussian features.

        - **TripoSR**: An open-source model for fast 3D reconstruction from a single image, developed collaboratively by Tripo AI and Stability AI.
        - **ZeroShape**: A regression-based approach achieving state-of-the-art zero-shot generalization in shape reconstruction.
        - **Hunyuan3D-2**: Combines state-of-the-art shape generation and texture synthesis to create high-resolution 3D assets from text and images.

- **Evaluation Metrics:**
    - Quantitative assessments using metrics such as Chamfer Distance and Intersection over Union (IoU).
    - Qualitative evaluations focusing on visual fidelity, texture accuracy, and geometric consistency.

Expected Deliverables:

- A dataset comprising images of real-world objects captured via mobile phone.
- Reconstructed 3D models generated by each of the selected methods.
- A comprehensive comparative analysis report detailing the performance of each method.
- A presentation summarizing the methodology, findings, and conclusions.

Recommended Reading:

- 3DTopia: GitHub Repository
- InstantMesh: GitHub Repository
- LGM: GitHub Repository
- TripoSR: Official Website
- ZeroShape: ArXiv Paper
- Hunyuan3D-2: Official Website

Project 9: Physics-Integrated Gaussians for Dynamic Scene Simulation

Theme:
3D Scene Reconstruction and Physics-Based Simulation

Project Description:
This project aims to bridge the gap between static 3D scene representations and dynamic physics simulations. Traditional 3D reconstruction techniques, such as Gaussian Splatting, excel at capturing static scenes from multi-view images. However, they lack the capability to simulate physical interactions and dynamics. By integrating physics-based simulation into the 3D Gaussian framework, we can enable realistic motion and interaction within reconstructed scenes. This approach allows for the generation of novel motions and behaviors, such as object deformation, collision, and material-specific responses, directly within the reconstructed environment.

Key Objectives:

- Understand and implement the PhysGaussian framework for integrating physics into 3D Gaussian representations.
- Capture and reconstruct new scenes containing dynamic elements, such as humans or materials with interesting physical properties (e.g., jelly, sand).
- Simulate and visualize realistic physical interactions within these scenes, demonstrating the capabilities of the integrated system.

Suggested Approach:

1. **Familiarization:**
    - Study the PhysGaussian method, focusing on how it integrates Newtonian dynamics within 3D Gaussians.
    - Experiment with sample scenes provided by the PhysGaussian project to understand the workflow and capabilities.PhysGaussian
2. **Data Acquisition:**
    - Capture multi-view images of scenes containing dynamic elements, such as a person interacting with objects or materials exhibiting interesting physical behaviors.
    - Ensure accurate camera pose estimation for each captured scene.
3. **Scene Reconstruction:**
    - Use Gaussian Splatting techniques to reconstruct the static 3D representation of the captured scenes.
    - Integrate the reconstructed scenes into the PhysGaussian framework.

4. **Physics Simulation:**
    - Define material properties for different elements within the scene (e.g., elasticity, viscosity).
    - Simulate physical interactions, such as collisions, deformations, and material-specific responses, using the integrated physics engine.

5. **Visualization and Analysis:**
    - Render the simulated scenes to visualize the dynamic behaviors.
    - Analyze the realism and accuracy of the simulations, comparing them to expected physical behaviors.

Expected Deliverables:

- A set of reconstructed 3D scenes with integrated physics simulations demonstrating dynamic behaviors.
- A comprehensive report detailing the methodology, implementation, results, and analysis of the simulations.
- A presentation summarizing the project's objectives, approach, findings, and conclusions.

Recommended Reading:

- PhysGaussian: Physics-Integrated 3D Gaussians for Generative Dynamics
- Gaussian Splatting for Real-Time Radiance Field Rendering
- Material Point Method for Physics-Based Simulation

Project 10: Benchmarking Multi-View Image Generation for 3D Consistency Without Ground Truth

Theme:
Evaluation Metrics for Multi-View Image Generation

Project Description:
This project aims to develop a benchmarking methodology for assessing the 3D consistency of multi-view image generation systems without the need for ground truth 3D data. Traditional evaluation metrics often rely on ground truth, which is scarce or non-existent in many real-world scenarios, or even worse, it does not make sense to evaluate against a GT in a generative context since many outputs can be considered correct. By focusing on the consistency of geometry and textures across different views, this project seeks to provide a reliable assessment framework that reflects the perceptual and structural coherence of generated images.

Key Objectives:

- Investigate existing metrics for evaluating multi-view image consistency without ground truth.
- Develop or adapt a benchmarking pipeline that assesses 3D consistency based on geometric and texture coherence across views.
- Validate the proposed benchmarking method on various multi-view image generation models.MEt3R
- Bonus: Find new creative ways to evaluate Multi View Systems. Benchmarking is a very active field, and a contribution in this direction could be a good step towards a publication.

Suggested Approach:

1. **Literature Review:**
    - Study existing metrics such as MEt3R, which evaluates multi-view consistency using dense 3D reconstructions from image pairs without requiring camera poses.
    - Examine methods like ConsistNet (this is just an example. You will need more papers) that enforce 3D consistency in multi-view image generation through architectural innovations.
2. **Method Development:**
    - Design a benchmarking pipeline that uses feature-based comparisons (e.g., using DINO features) to assess consistency across generated views.
    - Implement warping techniques to align images from different views and measure discrepancies in geometry and texture.
3. **Evaluation:**
    - Apply the benchmarking method to various multi-view image generation models.
    - Analyze the results to determine the effectiveness of the benchmarking approach in capturing inconsistencies.

Expected Deliverables:

- A comprehensive benchmarking pipeline for evaluating multi-view image generation consistency without ground truth.
- A detailed report documenting the methodology, experiments, results, and analyses.
- A presentation summarizing the project's objectives, approach, findings, and conclusions.

Recommended Reading:

- MEt3R: Measuring Multi-View Consistency in Generated Images.
- ConsistNet: Enforcing 3D Consistency for Multi-view Images Diffusion.

Project 11: Comparative Analysis of Surface Reconstruction Methods from Oriented Point Clouds

Theme:
Surface Reconstruction from Point Clouds

Project Description:
Surface reconstruction from oriented point clouds is a fundamental problem in 3D computer vision and graphics, with applications in robotics, augmented reality, and digital manufacturing. Recent advances in Poisson-based reconstruction and deep learning-based methods have significantly improved reconstruction quality, robustness, and scalability. This project aims to ***compare four state-of-the-art surface reconstruction techniques*** — Poisson Surface Reconstruction (PSR), Screened Poisson Reconstruction (SPSR), Deep Geometric Prior (DGP), and Neural Stochastic Screened Poisson Reconstruction (NS-SPSR)—on publicly available datasets such as the ***ABC dataset (used in Point2Surf)***.

The project will evaluate reconstruction ***accuracy, robustness to noise, computational efficiency, and scalability*** across different object complexities. Students will implement or adapt existing implementations of these methods and benchmark them under controlled conditions.

Key Objectives:

- Implement or adapt at least four reconstruction algorithms:
  - Poisson Surface Reconstruction (PSR) (Kazhdan et al., 2006)
  - Screened Poisson Surface Reconstruction (SPSR) (Kazhdan et al., 2013)
  - Deep Geometric Prior for Surface Reconstruction (DGP) (Williams et al., 2019)
  - Stochastic Poisson Surface Reconstruction (Sellan and Jacobson et al, 2022)
  - Neural Stochastic Screened Poisson Reconstruction (NS-SPSR) (Wu et al., 2023)
  - [Optional] Stochastic Poisson Reconstruction with One Solve (using Geometric Gaussian Processes) (Holalkere et al., 2025)
- Benchmark these methods on datasets like ABC, ShapeNet, or ScanNet, evaluating:
  - Geometric accuracy (e.g., Chamfer and Hausdorff distances)
  - Robustness to noise and outliers
  - Computational efficiency and scalability

Suggested Approach:

- **Dataset Preparation:**
  - Use datasets like ABC, ShapeNet, or ScanNet. (to start you can use https://www.cg.tuwien.ac.at/research/publications/2020/erler-2020-p2s/)
  - Generate synthetic noise variations for robustness testing.

- **Method Implementation:**
    - Use Open3D for PSR/SPSR and PyTorch-based code for DGP/NS-SPSR.
    - Ensure consistent input (oriented point clouds) and evaluation metrics.
- **Evaluation:**
    - Compare outputs quantitatively (e.g., error metrics) and qualitatively (visual comparisons).
    - Visualize differences (e.g., hole-filling, noise handling).
- **Analysis:**
    - Draw insights on trade-offs between optimization- and learning-based methods.
    - Discuss the potential for hybrid approaches.

Expected Deliverables:

- Codebase with implementations/adaptations of the selected methods
- Benchmarking results (tables, charts, visualizations)
- Report discussing methodology, results, and conclusions
- Presentation summarizing the work

Recommended Reading:

- [Poisson Surface Reconstruction (Kazhdan et al., 2006)]
- [Screened Poisson Surface Reconstruction (Kazhdan et al., 2013)]
- [Deep Geometric Prior (Williams et al., 2019)]
- [Neural Stochastic Screened Poisson Reconstruction (Wu et al., 2023)]
- [Stochastic Poisson Reconstruction with One Solve (Holalkere et al., 2025)]
- [Point2Surf & ABC Dataset (Erler et al., 2020)]

Project 12: 3D Gaussian Splatting and Compression for Efficient and Accelerated Novel View Synthesis

Theme:
Efficient 3D Scene Representation and Rendering

Project Description:
3D Gaussian Splatting has emerged as a breakthrough technique for real-time radiance field rendering, enabling high-quality novel view synthesis with real-time performance. However, the computational and memory costs of dense Gaussian representations remain a challenge for practical deployment. This project explores ***state-of-the-art compression and optimization techniques*** for 3D Gaussian Splatting, aiming to reduce storage and accelerate rendering while preserving visual fidelity.

Students will compare methods like ***Compressed 3D Gaussian Splatting, Mini-Splatting, and LightGaussian***, evaluating their trade-offs in ***quality, speed, and memory efficiency*** on standard datasets. The project will involve implementing or adapting existing techniques, benchmarking their performance, and analyzing their suitability for real-time applications like AR/VR and robotics.

Key Objectives:

- **Implement/adapt at least 4 of the key Gaussian Splatting compression techniques (including original 3DGS):**
  - *3D Gaussian Splatting for Real-Time Radiance Field Rendering (Kerbl et al., 2023)* – Baseline method for real-time radiance fields.
  - *Mini-Splatting: Representing Scenes with a Constrained Number of Gaussians (Fang et al., 2024)*
  - *Compressed 3D Gaussian Splatting for Accelerated Novel View Synthesis (Niedermayr et al., 2024)*
  - *CompGS: Smaller and Faster Gaussian Splatting with Vector Quantization (Navaneet and Meibodi et al., 2024)*
  - *LightGaussian: Unbounded 3D Gaussian Compression with 15x Reduction and 200+ FPS (Fan and Wang et al., 2024)*
  - [Optional] *3D Gaussian Rendering Can Be Sparser: Efficient Rendering via Learned Fragment Pruning (Ye et al., 2024)*
- **Benchmark on datasets (e.g., Mip-NeRF 360, Tanks & Temples) using:**
  - Rendering quality (PSNR, SSIM, LPIPS)
  - Computational efficiency (FPS, memory usage)
  - Compression ratios (storage savings vs. quality loss)
- **Analyze trade-offs between compression methods and identify optimal use cases**

Suggested Approach:

- **Dataset Preparation:**
    - Use standard novel-view synthesis datasets (e.g., Mip-NeRF 360, Tanks & Temples)
    - Preprocess data for Gaussian Splatting (e.g., COLMAP for point clouds).
- **Method Implementation:**
    - Start with the original 3DGS codebase https://github.com/graphdeco-inria/gaussian-splatting.
    - Integrate ***Mini-Splatting, LightGaussian, or CompGS*** for comparison
    - Ensure consistent evaluation metrics (e.g., rendering speed, quality)
- **Evaluation:**
    - Quantitative: Compare PSNR/SSIM, FPS, and memory footprint
    - Qualitative: Visualize rendering artifacts or improvements
- **Analysis:**
    - Determine which methods best balance speed and quality.
    - Discuss potential hybrid approaches (e.g., pruning + quantization).

Expected Deliverables:

- **Code repository** with implemented/adapted compression techniques (at least 4).
- ***Benchmark results*** (tables, plots, and visual comparisons).
- ***Final report*** documenting methodology, results, and insights.
- **Presentation** summarizing key findings.

Recommended Reading:

- *3D Gaussian Splatting for Real-Time Radiance Field Rendering (link: https://arxiv.org/pdf/2308.04079.pdf)*
- *Mini-Splatting: Representing Scenes with a Constrained Number of Gaussians (link: https://arxiv.org/pdf/2403.14166.pdf)*
- *Compressed 3D Gaussian Splatting for Accelerated Novel View Synthesis (link: https://arxiv.org/pdf/2401.02436.pdf)*
- *CompGS: Smaller and Faster Gaussian Splatting with Vector Quantization (link: https://arxiv.org/pdf/2311.18159.pdf)*
- *LightGaussian: Unbounded 3D Gaussian Compression with 15x Reduction and 200+ FPS (link: https://arxiv.org/pdf/2311.17245.pdf)*
- [Optional] *3D Gaussian Rendering Can Be Sparser: Efficient Rendering via Learned Fragment Pruning (link: https://openreview.net/pdf?id=IVqzbuLfoL)*

Project 13: Real-time Neural Radiance Field Rendering: A Comprehensive Evaluation of Acceleration Techniques

Theme:
Neural Rendering Optimization and Real-time View Synthesis

Project Description:
Neural Radiance Fields (NeRFs) have revolutionized 3D scene reconstruction and novel view synthesis; however, their computational demands limit their application in real-time scenarios. This project conducts a comprehensive evaluation of seven state-of-the-art NeRF acceleration and compression techniques:

- ***KiloNeRF*** (distributes computation across thousands of tiny MLPs)
- ***PlenOctrees*** (enables real-time rendering through octree-based caching)
- ***AdaNeRF*** (implements adaptive sampling to optimize computation)
- ***EfficientNeRF*** (streamlines network architecture for faster inference)
- ***Instant NGP*** (leverages multiresolution hash encoding for speed)
- ***Compact NGP*** (enhances compression via learned hash probing)
- [Optional] ***Lagrangian Hashing*** (introduces advanced compression for neural fields)

Students will implement and compare these methods across multiple dimensions: rendering speed (FPS), memory efficiency (VRAM usage), reconstruction quality (PSNR/SSIM), and training time. The evaluation will use standard datasets (Blender, Mip-NeRF 360) to identify optimal approaches for different applications, from mobile AR to high-fidelity visual effects.

Key Objectives:

- **Implement and adapt seven acceleration techniques:**
    - KiloNeRF's distributed MLP architecture
    - PlenOctrees' caching mechanism
    - AdaNeRF's adaptive sampling
    - EfficientNeRF's streamlined network
    - Instant NGP's hash encoding
    - Compact NGP's learned probing
    - [Optional] Lagrangian Hashing's compression
- **Benchmark performance across:**
    - Rendering speed (FPS at 1080p/4K)
    - Memory efficiency (GPU VRAM consumption)
    - Visual quality (PSNR, SSIM, LPIPS)
    - Training convergence time
- **Develop a decision framework for selecting methods based on:**
    - Hardware constraints (mobile vs. desktop)
    - Quality requirements (real-time vs. offline)
    - Scene complexity (simple objects vs. large environments)

Suggested Approach:

- **Experimental Setup:**

    - Standardized testing environment (GPU specification, OS)
    - Dataset preparation (Blender synthetic, Mip-NeRF 360 real-world)
    - Baseline implementation using Nerfstudio
- **Implementation Strategy:**
    - ***Phase 1***: Basic implementations (Instant NGP, PlenOctrees)
    - ***Phase 2:*** Advanced methods (KiloNeRF, AdaNeRF)
    - ***Phase 3:*** Compression techniques (Compact NGP, Lagrangian Hashing)
- **Evaluation Protocol:**
    - Quantitative: Frame rate, memory usage, quality metrics
    - Qualitative: Visual artifacts analysis
    - Computational: Training time profiling
- **Analysis Framework:**
    - Method clustering by performance characteristics
    - Identification of complementary techniques
    - Development of hybrid approach recommendations

Expected Deliverables:

- **Modular codebase with all seven methods implemented (last one optional)**
- **Comprehensive benchmark results including:**
    - Performance comparison tables
    - Frame rate plots across resolutions
    - Memory-quality tradeoff visualizations
- **Technical report featuring:**
    - Methodology documentation
    - Results analysis
    - Practical implementation guidelines
- **Presentation materials summarizing:**
    - Key findings
    - Visual comparisons
    - Application recommendations

Recommended Reading:

- *KiloNeRF: Speeding up Neural Radiance Fields with Thousands of Tiny MLPs* (link: https://arxiv.org/pdf/2103.13744.pdf)
- *PlenOctrees for Real-time Rendering of Neural Radiance Fields* (link: https://arxiv.org/pdf/2103.14024.pdf)
- *AdaNeRF: Adaptive Sampling for Real-time Rendering of Neural Radiance Fields* (link: https://arxiv.org/pdf/2207.10312.pdf)
- *EfficientNeRF - Efficient Neural Radiance Fields* (link: https://arxiv.org/pdf/2206.00878.pdf)
- *Instant neural graphics primitives with a multiresolution hash encoding* (link: https://arxiv.org/pdf/2201.05989.pdf)
- *Compact Neural Graphics Primitives with Learned Hash Probing* (link: https://arxiv.org/pdf/2312.17241.pdf)
- *Lagrangian Hashing for Compressed Neural Field Representations* (link: https://arxiv.org/pdf/2409.05334.pdf)

Project 14: Training-Free 4D Reconstruction using Easi3R

Theme: Dynamic Scene Understanding and Attention Mechanisms

Project Description:
This project will focus on a deeper analysis of the dynamic object segmentation component of the Easi3R method. While Easi3R demonstrates a novel way to extract motion information from attention maps, there's room to explore the quality of the segmentation and potential enhancements. This project will investigate factors affecting segmentation accuracy and propose modifications to improve its robustness.

Key Objectives:
- Implement the dynamic object segmentation part of the Easi3R method.
- Analyze the sensitivity of the segmentation to different parameters and video characteristics.
- Propose and evaluate a modification to the segmentation method to improve Easi3R's performance.

Suggested Approach:
1. **Implement Dynamic Object Segmentation:**
    - Implement the attention map extraction and aggregation as described in the Easi3R paper.
    - Implement the dynamic attention map computation and segmentation.
    - Potentially: Implement the Easi3R attention mechanism for MASt3r
2. **Sensitivity Analysis**
    - Evaluate the impact of the attention threshold (α) on the segmentation and reconstruction results.
    - Analyze how performance varies with different types of motion (e.g., fast vs. slow, large vs. small objects).
    - Investigate the effect of video quality (e.g., resolution, noise) on the segmentation.
3. **Segmentation Enhancement:**
    - Propose a modification to the segmentation method like incorporating temporal information more effectively, adding spatial regularization to the segmentation mask, or exporing different ways to combine the attention maps.
    - Implement the proposed modification.
4. **Evaluation**
    - Compare your adapted method to the original Easi3R and other baselines

Recommended Reading
- Dust3R
- Easi3R
- MonST3R
- DAS3R
- CUT3R

Project 15: Surface Reconstruction from Gaussian Splatting

Theme: 3D Scene Representation and Mesh Extraction

Project Description:
3D Gaussian Splatting has emerged as a powerful technique for novel view synthesis, offering real-time rendering capabilities. However, extracting a high-quality mesh from the unstructured set of 3D Gaussians remains a challenge. This project will explore and compare different approaches for mesh extraction from 3DGS representations.

Key Objectives:
- Implement/adapt and compare Gaussian Splatting methods for surface reconstruction
- Investigate techniques to improve the quality and accuracy of the extracted meshes.
- Establish a benchmark for evaluating mesh extraction quality and efficiency.
- Compare the performance of different methods based on the benchmark.

Suggested Approach:
1. Method Selection and Implementation:
   - Review the literature for methods for surface reconstruction from Gaussian Splatting
   - Implement/adapt multiple surface extraction methods
2. Benchmark Design:
   - Select a variety of scenes for benchmakring
   - Define quantitative metrics for evaluating the methods like geometric accuracy, mesh quality, rendering quality, computational time, memory usage etc

Expected Deliverables
- Implementation of mesh extraction methods
- Benchmarking pipeline
- Quantitative and qualitative evaluation results
- Report and presentation summarizing your finidigns

Recommended Reading
- SuGaR
- 2DGS
- Gaussian Opacity Fields

Project 16: Comparative Analysis of Point Cloud Transformer Architectures

Theme: 3D Point Cloud Processing and Transformer Networks

Project Description:
This project will involve a comparative study of different transformer-based architectures for point cloud processing. You will implement or utilize existing implementations of selected architectures, establish a benchmark, and analyze their performance on standard point cloud tasks.

Key Objectives:
- Select and implement (or find implementations) of different point cloud transformer architectures.
- Establish a benchmark for evaluating the architectures on tasks like classification and segmentation
- Analyze the strengths and weaknesses of each architecture based on your findings.

Suggested Approach:
1. Architecture Selection
    - Review the literature for different point cloud transformer architectures
    - Choose a handful of architectures that you want to evaluate
2. Implementation / Adaption
    - Implement or adapt existing implementations of your chosen architectures
3. Benchmark Design
    - Select standard point cloud tasks you want to analyze such as classification or segmentation
    - Select appropriate datasets for evaluation, such as ModelNet40, ShapeNetPart, S3DIS etc
    - Define evaluation metrics that help you analyze the strengths and weaknesses of the different architectures
4. Experiments and Analysis
    - Train and evaluate the chosen architectures
    - Analyze the results and discuss potential trade-offs
    - Investigate the impact of different transformer components like attention mechanisms etc

Expected Deliverables:
- Implementations (or adapted implementations) of the chosen point cloud transformer architectures
- Benchmarking pipeline
- Quantitative results and visualizations
- A report and presentation summarizing the comparative analysis

Recommended Reading:
- Point Transformer v1, v2, v3
- Point Cloud Transformer
- OctFormer

Proposing Your Own Research Project (as soon as possible)

Theme:
Independent Research Proposal

Overview:
If you have a research idea that aligns with the themes of our course, such as 3D vision, neural representations, generative models, or scene understanding. You are encouraged to develop and propose your own project. This initiative allows you to explore areas of personal interest while contributing original work to the field.

Steps to Propose Your Project:

1. **Develop Your Idea:**
    - Identify a specific problem or question within the course themes that intrigues you.
    - Conduct a literature review to understand the current state of research and identify gaps or areas for improvement.
    - Define clear objectives and outcomes for your project.

2. **Prepare a Brief Proposal:**
   Draft a concise document (1–2 pages) that includes:
    - Project Title: A clear and descriptive title.
    - Research Question or Problem: What specific issue will your project address?
    - Objectives: What do you aim to achieve?
    - Methodology: Outline the approach and techniques you plan to use.
    - Expected Outcomes: What results do you anticipate?
    - Relevance: How does your project relate to the course themes?

3. **Consult with Teaching Assistants (TAs):**
    - Schedule a meeting with the course TAs to discuss your proposal.
    - Be prepared to explain your idea clearly and answer any questions they may have.

    - Seek feedback and be open to suggestions or modifications.

4. **Revise and Finalize:**

    - Incorporate the feedback received from the TAs into your proposal.
    - Ensure your project is feasible within the course timeframe and resources.
    - Submit the finalized proposal to the TAs for final approval before beginning work.

Important Notes:

- All independent projects must be approved by the TAs before starting.
- Projects should align with the course objectives and themes.
- Originality and creativity are encouraged; however, ensure your project is achievable within the course's scope.

Final Notes

- Students are encouraged to explore additional resources and tools beyond the recommended readings.
- Consultations with course TAs are advised to ensure project alignment with course objectives.
- Final presentations will be held in person. Date on the 3D Vision website.