CS 172 Final Project

CS 172 TA Team

Novel View Synthesis for Dynamic Scene

Contact: wangshuo2022@shanghaitech.edu.cn

zhaoyq12022@shanghaitech.edu.cn

Problem setting

- Given a monocular video with camera poses of Dynamic Scenes
- Achieve Novel View Synthesis.
- You should take the optimization speed into consideration.

Dataset and Evaluation Metrics

- Dynamic Scene Dataset
- https://filebox.ece.vt.edu/~chengao/free-view-video/data.zip

- Evaluation metrics:
- Rendering quality: PSNR, LIPIS, SSIM.
- Training speed: The total optimization step and training time(You should also report the GPU device).

Requirements

- You can refer to any speed up strategy of NeRF to implement fast training and fast rendering.
- Code and a technical report are required
- You should at least implement one methods for this setting.
- Show your results with data and also visualization

- Reference: Dynamic View Synthesis from Dynamic Monocular Video, Gao et. al.

Bonus (full score if better results)

- Extend Gaussian Splatting to this setting
- graphdeco-inria/gaussian-splatting: Original reference implementation of "3D Gaussian Splatting for Real-Time Radiance Field Rendering" (github.com)

Animatable Human Avatar

Contact:liushy2023@shanghaitech.edu.cn

Problem setting

- Given multi-view or monocular RGB videos.
- Reconstruct animatable human avatars (animate by any pose).

e.g.

Input

Output

Dataset and Evaluation Metrics

- Dataset
- https://epan.shanghaitech.edu.cn/v/link/view/5a0eb98ad77f4f3e97612df8a6f416d2 (multi-view)
- https://graphics.tu-bs.de/people-snapshot (monocular)

- Evaluation metrics:
- Rendering quality: PSNR, LIPIS, SSIM.

Requirements

- Code of one method about this setting.
- Your technical report.
- Your quantitative results and your visualization.
- Tips: you should utilize human structure priors (SMPL) to make your results better.

- Reference:

- Neural Body: Implicit Neural Representations with Structured Latent Codes for Novel View Synthesis of Dynamic Humans **Sida Peng** et. al
- InstantAvatar: Learning Avatars from Monocular Video in 60 Seconds

Bonus (full score if better results)

- Extend Gaussian Splatting and monocular video to this setting.
- graphdeco-inria/gaussian-splatting: Original reference implementation of "3D Gaussian Splatting for Real-Time Radiance Field Rendering" (github.com)

Indoor scene reconstruction

Contact: xujw2023@shanghaitech.edu.cn

Problem setting

- Given a set of images of an indoor room scene
- Reconstruct the geometry of the scene

- Methods tips:
- Please use neural SDF representation.
- Please try to use the monocular geometry prior(like normal and depth) to enhance the reconstruction quality

Dataset and Evaluation

- Scannet
- wget https://s3.eu-central-1.amazonaws.com/avg-projects/monosdf/data/scannet.tar

- Evaluation metrics:
- Reconstruction quality: Chamfer distance, F-Score

- Please also report the total optimization steps and your training time(and also the GPU device you use)

Requirements

- Code of one method about this setting.
- Your technical report.
- Your quantitative results and your visualization.

- MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction, Yu et. al.

Bonus

- A common approach is to utilize pre-trained model to get the prior, but there may be noise.

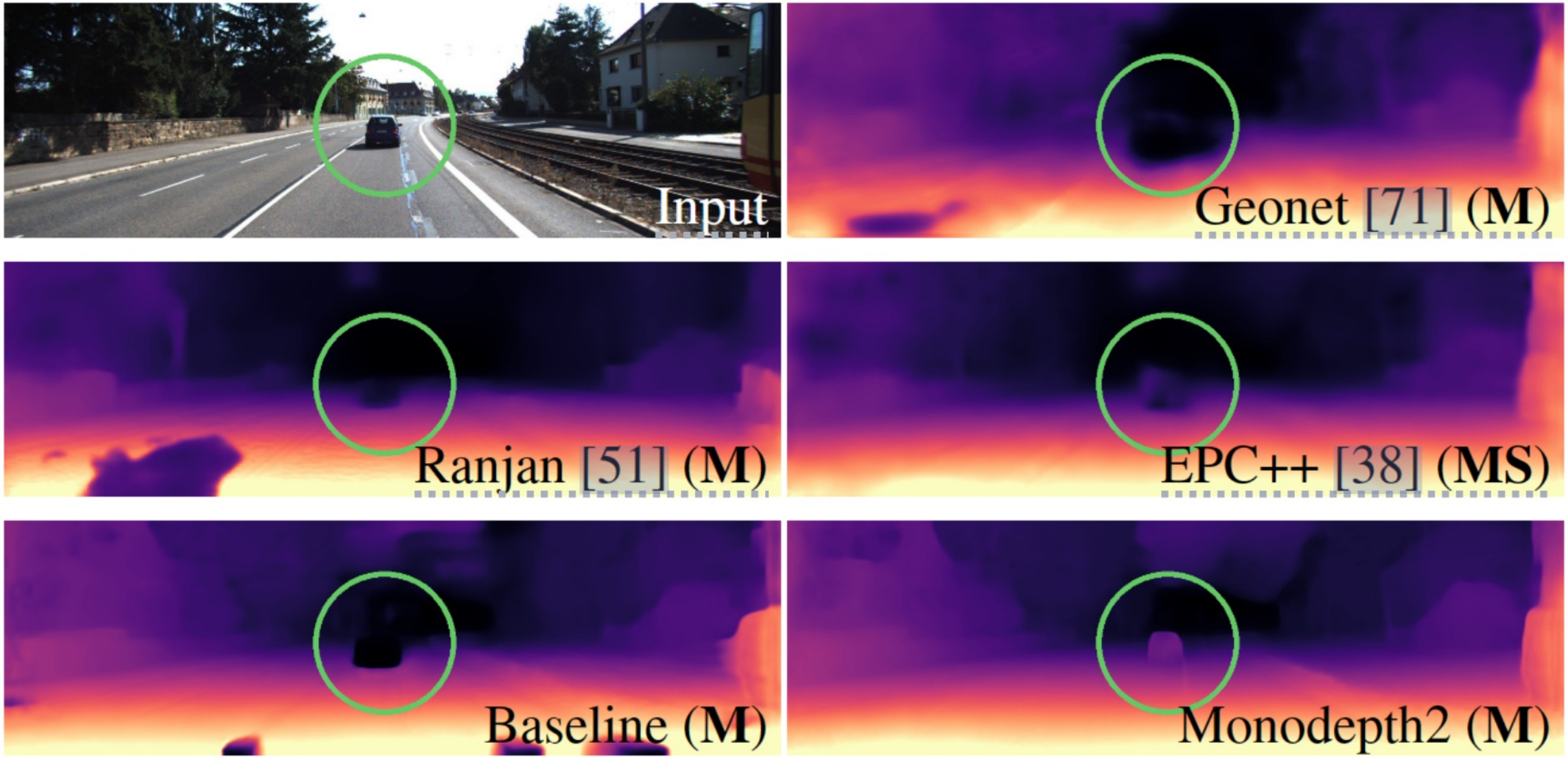
- For example, this predicted monocular depth map is wrong in some region.
- Leverage your imagination, and try to give a solution to this problem.

Disentangling Object Motion for Self-supervised Depth Estimation

Contact: wangry3@shanghaitech.edu.cn

wangchy8@shanghaitech.edu.cn

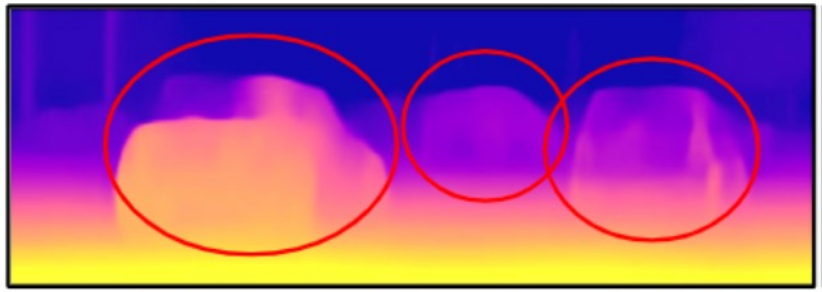
Problem Setting



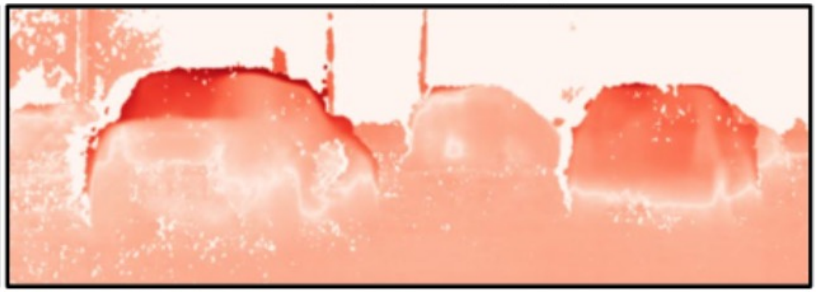
Problem Setting

Input Data from Cityscapes

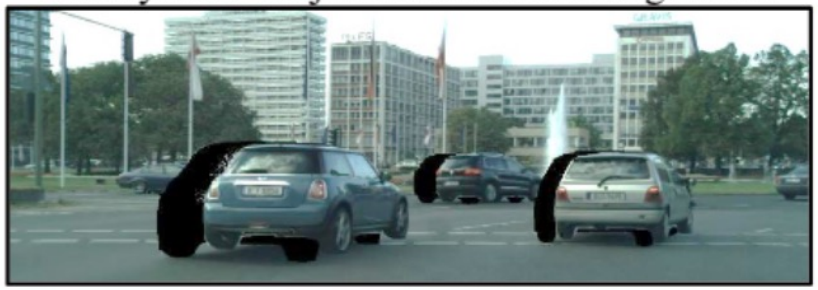
Depth Prediction of Manydepth



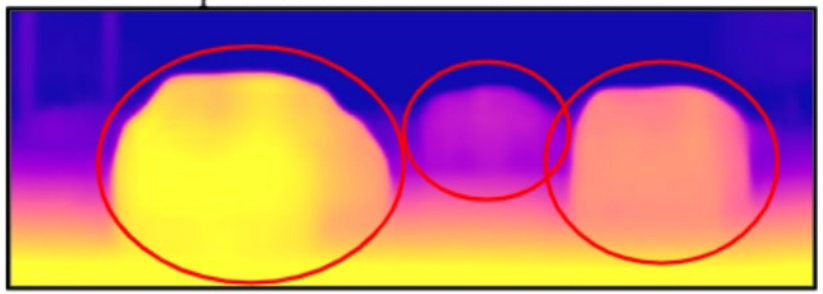
Error Map of Manydepth



Our Dynamic Object Motion Disentanglement



Depth Prediction of Our Method



Error Map of Our Method

Problem Setting

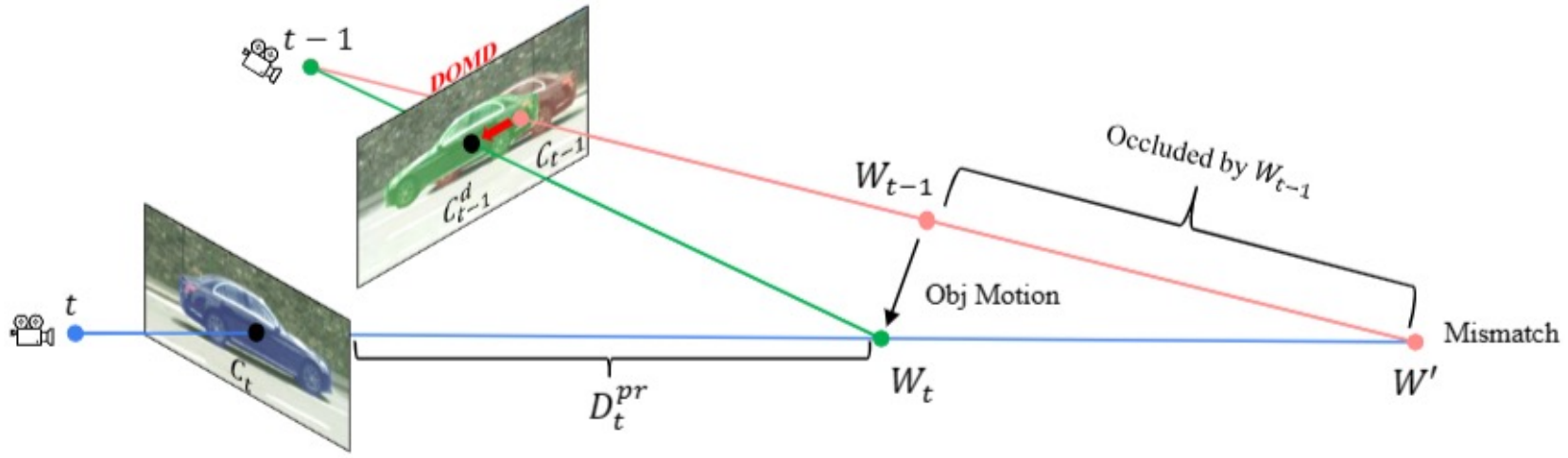
- Problem:
  - Self-supervised depth estimation heavily suffers from the mismatch of the photometric consistency caused by moving objects.

- What you need to do:
  - Find a method to solve this problem.

Previous Works – Segmentation Supervised

- Steps:
  - Mask the movable objects by a pretrained semantic segmentation network.
  - Predict the scene flow of them to correct the projection.
  - Finally, get the well-trained depth, pose, and scene flow networks.

- Reference:
  1. Disentangling Object Motion and Occlusion for Unsupervised Multi-frame Monocular Depth.
  2. Self-Supervised Monocular Depth Estimation: Solving the Dynamic Object Problem by Semantic Guidance.



Previous Works – Unsupervised

- Problem:
  - The ambiguity of camera and objects motion.

- Steps:
  - Predict the scene flow.
  - Regularize the scene flow to movable objects.
  - Finally, get the well-trained depth, pose, and scene flow networks.

- Reference:
  1. SC-DepthV3: Robust Self-supervised Monocular Depth Estimation for Dynamic Scenes.
  2. Unsupervised Monocular Depth Learning in Dynamic Scenes.
  3. EmerNeRF: Emergent Spatial-Temporal Scene Decomposition via Self-Supervision.

Dataset and Evaluation

- Dataset:
  - Cityscapes (√)
  - KITTI (×)

- Evaluation:
  - Report the following metrics: Abs_Rel, Sq_Rel, RMSE, RMSE_log, A1, A2, A3.
  - Show the ablation studies of your proposed method.