ETH zürich

Department of Computer Science

Improving the Ground Truth: MegaDepth in 2023

Alexander Veicht, Felix Yang, Andri Horat, Deep Desai
Supervisor: Philipp Lindenberger

1 Motivation

MegaDepth [1], a dataset of unstructured images featuring popular tourist landmarks, was introduced in 2018. By leveraging structure from motion (SfM) and multi-view stereo (MVS) techniques along with data cleaning methods, MegaDepth generates camera poses and depth maps for each image.

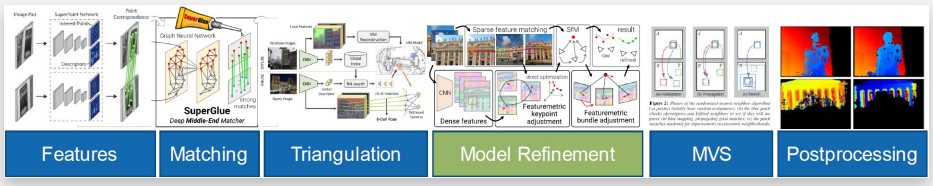
Problems:

- MegaDepth suffers from limitations like degenerate camera poses, incomplete depth maps, and inaccuracies caused by unregistered images or noise in the pipeline.
- The original MegaDepth dataset only makes use of a very small fraction of all available images because COLMAP fails to register the remaining ones.

Our solution:

We propose a refined MegaDepth ground-truth pipeline using recent deep learning-based methods such as Hierarchical Localization [2], SuperPoint [3], SuperGlue [4] and Pixel-Perfect Structure-from-Motion [5] to address the issues.

2 Refined MegaDepth Ground-Truth Pipeline



Original MegaDepth Pipeline:

- **Features**: SIFT
- **Matching:** NN-ratio
- **Triangulation**: COLMAP
- **Model refinement**: N/A
- **MVS**: COLMAP
- **Postprocessing**:
  - **Image processing**
  - **Semantic segmentation**: PSPNet
  - **Ordinal maps**

Refined MegaDepth Pipeline:

- **Features**: SuperPoint
- **Matching**: SuperGlue
- **Triangulation**: COLMAP
- **Model refinement**: PixSfM
- **MVS**: COLMAP
- **Postprocessing**:
  - **Image processing**
  - **Semantic segmentation:** SegFormer
  - **Ordinal maps**

3 Overlap Metrics

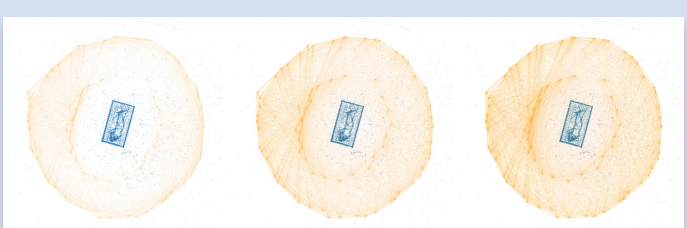
Problem:

The overlap between two images gives a prior for the difficulty of the image pair. The existing overlap score that is obtained from the sparse correspondences is very noisy and degrades heavily if some features are not matched across all images.

Sparse overlap score: 26.6%
Cosine-weighted dense overlap score: 42.2%
Dense overlap score: 78.8%

Our solution:

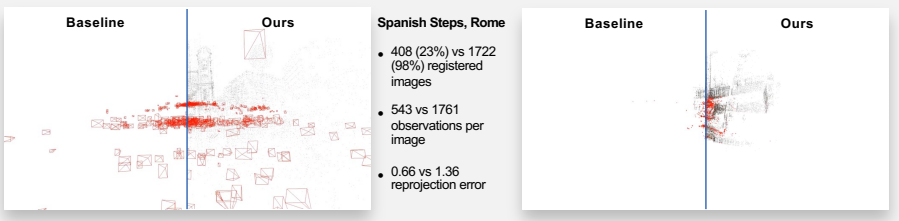
Our newly proposed dense overlap score overcomes these flaws by leveraging information from the dense reconstruction. We also use a cosine disparity weighted dense overlap to penalize drastic view changes.0



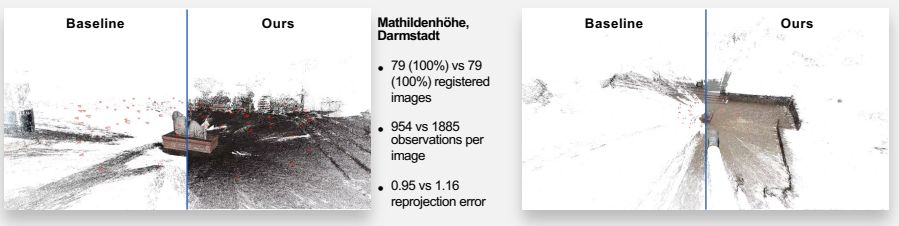
From left to right: sparse overlap, cosine-weighted dense overlap and dense overlap

4 Results

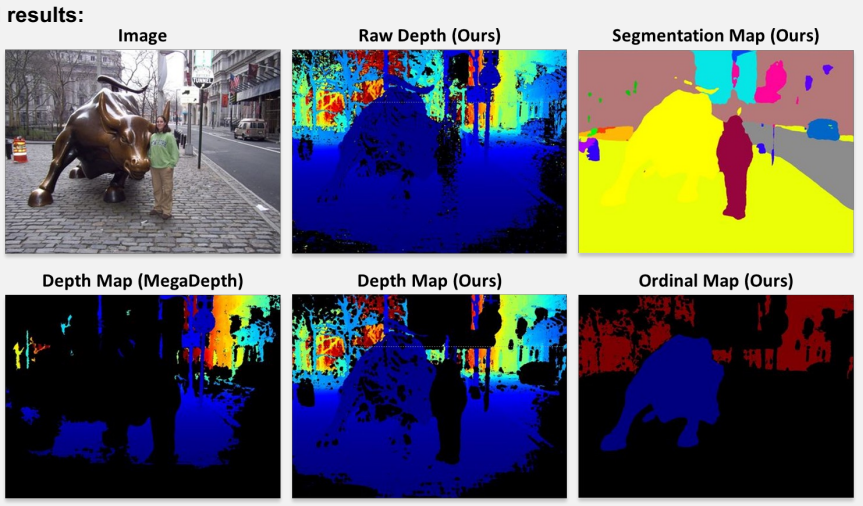
Sparse models:



Dense models:

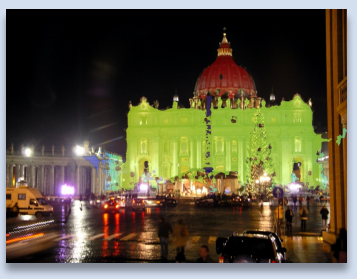


Final results:



5 Insights

- Symmetry is a significant limitation, demonstrated by the presence of "ghost towers" in sparse reconstruction, where symmetrical objects are falsely represented.
- Rotated images are not registered, which can result in failed reconstruction for scenes with a mixture of upside-down or atypically rotated images.
- SuperPoint captures more keypoints in the background, while LoFTR captures more keypoints in the foreground.
- Concatenating features from SuperPoint and LoFTR may allow for the utilization of both foreground and background information in reconstruction.
- Our refined dataset features much more nighttime images.



References

1. Li, Zhengqi, and Noah Snavely. "Megadepth: Learning single-view depth prediction from internet photos." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.
2. Sarlin, Paul-Edouard, et al. "From coarse to fine: Robust hierarchical localization at large scale." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019.
3. DeTone, Daniel, Tomasz Malisiewicz, and Andrew Rabinovich. "Superpoint: Self-supervised interest point detection and description." *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2018.
4. Sarlin, Paul-Edouard, et al. "Superglue: Learning feature matching with graph neural networks." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020.
5. Lindenberger, Philipp, et al. "Pixel-perfect structure-from-motion with featuremetric refinement." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021.