Improving Object Pose Estimation with Line Features in MR

Moyang Li
ETH Zurich
moyali@ethz.ch

Yifan Jiang
University of Zürich
yifan.jiang@uzh.ch

Yuqiao Huang
University of Zürich
yuqiao.huang@uzh.ch

Deqing Song
University of Zürich
deqing.song@uzh.ch

Abstract

Object pose estimation plays a critical role in augmented reality (AR) applications, particularly in environments with low-texture surfaces, such as SBB train doors. This work investigates the challenges of using LIMAP for pose estimation with domain shifts, demonstrating its limitations. We introduce a robust, generalizable approach leveraging the dense feature matcher (GIM) to enhance the performance of feature matching under domain-shifted conditions and provide real-time implementation on Microsoft HoloLens, demonstrating its potential for practical AR applications. The results show the effectiveness of GIM for robust feature matching and pose estimation. Supplementary material and code are available on the project page: https://yukhaau.github.io/course-showcase.

1. Introduction

Object pose estimation is essential for enabling seamless augmented reality (AR) and Mixed Reality (MR) experiences, especially in industrial contexts, such as SBB train systems. Accurate and robust pose estimation is critical for applications like maintenance, inspection, and navigation. In real-world settings, many objects lack sufficient texture or distinctive features, making accurate and stable pose estimation challenging for traditional point-based methods. These challenges are intensified in environments with low-texture surfaces or significant variations in appearance due to environmental changes.

To address these limitations, researchers have explored the use of keypoint-free methods and line feature based method. OnePose ++ [9] provides keypoint-free one-shot object pose estimation without CAD models, while line features are particularly effective in environments where the object exhibits strong edges, corners, or linear patterns. For instance, Hierarchical Localization (hloc) [19, 20] and LIMAP [13] have demonstrated the potential of line features for robust pose estimation. Despite their strengths, line-based methods suffer from sensitivity to domain shifts and the inclusion of unrelated lines from the surrounding environment, which often leads to degraded performance. This limitation restricts their applicability in dynamic or complex industrial scenarios. Object detection and segmentation techniques also play a critical role in guiding feature matching and pose estimation. Modern frameworks such as YOLOv8 [10] enable fast and accurate bounding box predictions, which are essential for isolating objects of interest.

This project builds on these advancements and addresses their limitations by introducing two methods and making comprehensive comparisons. The first method utilizes YOLOv8 for bounding box predictions to remove outlier lines and perform LIMAP-based line feature matching, which shows limitations under domain shift. Thus, we further incorporate a dense feature matcher (GIM) [22] to improve the robustness of feature matching under domain shifts, mitigating the challenges posed by feature mismatches and environmental variability.

Our contributions include: (1) a comprehensive performance evaluation of line-based methods for object pose estimation in low-texture environments, (2) the development of a generalizable matching approach that enhances robustness and accuracy in the presence of domain shifts, and (3) a real-time implementation of the proposed pipeline on Microsoft HoloLens 2, demonstrating its practical applicability in augmented reality scenarios. Experimental results validate the effectiveness of our approach, highlighting significant improvements in localization accuracy and robustness across various challenging conditions.

2. Related Works

(Low Texture) Feature Matching. Feature matching is a critical task in visual localization and pose estimation, particularly in low-texture environments. Hierarchical Localization (HLoc) [19, 20] combines image matching and Structure-from-Motion (SfM) to provide a fast, accurate, and scalable solution for both indoor and outdoor environments, enabling it to achieve high accuracy across a variety of benchmarks, making it a foundational tool for improving localization accuracy in complex environments. Then, learning-based line detectors have been developed to tackle the problem of wireframe parsing [1]. Re-

cent deep learning-based detectors, such as L-CNN [25] and its successors, have achieved impressive results in detecting general line segments. Matching of detected line segments often relies on learning-based descriptors [24, 25].Some methods also exploit hybrid point-line [17] and line-junction structures to enhance matching robustness. LIMAP [13] introduces a framework for leveraging line features, focusing on improving matching accuracy by reconstructing 3D line maps and exploiting geometric constraints. However, line-based methods often struggle with domain shifts and unrelated features, limiting their robustness. GIM (Generalizable Image Matcher) [22] addresses these challenges by learning dense feature representations from large-scale datasets, including internet videos. By generalizing well across diverse environments, GIM improves robustness in feature matching, even in low-texture or domain-shifted conditions. This makes it particularly suitable for industrial scenarios where traditional methods often fail. OnePose++ [9] builds on these advancements by incorporating both point and line features, enhancing 6-DoF pose estimation in low-texture and texture-less scenarios. Its ability to robustly integrate multiple types of features makes it a valuable tool for challenging industrial contexts, such as those involving SBB train systems.

Line reconstruction. Line reconstruction involves generating 3D line maps to model the geometric structure of a scene. Bartoli and Sturm [2, 3] pioneered a full Structure-from-Motion (SfM) pipeline for line segments. Global topological constraints were later introduced to build wireframe models in works such as [12]. Learning-based approaches [12] have then emerged as effective solutions for 3D line reconstruction, integrating weak epipolar constraints and graph clustering to improve accuracy. ELSR [23] recently employed planes and points to guide line matching, offering an alternative for precise 3D map construction. Checking weak epipolar constraints over exhaustive matches and graph clustering, and introducing the Line3D++ software remains the top choice [7, 14] for acquiring 3D line maps so far.

Object Detection and Segmentation. Accurate detection and segmentation of objects are vital for guiding feature matching and pose estimation. YOLO (You Only Look Once) [18], particularly its latest iteration YOLOv8, provides a fast and accurate object detection framework that excels in identifying objects and generating bounding boxes in real time. Mask R-CNN [8] extends object detection to instance segmentation, providing pixel-level accuracy essential for extracting object regions in cluttered scenes. Semantic SAM [11] emerged as a powerful tool for semantic segmentation, offering zero-shot generalization to unseen objects. This capability makes it valuable for segmenting objects in low-texture environments, where traditional models struggle.

Synthetic Data Generation Synthetic data has become an indispensable tool for training models in domain-shifted and low-texture environments. BlenderProc [5] facilitates the generation of realistic synthetic datasets from CAD models, providing diverse and controlled training data. By augmenting real-world datasets with synthetic counterparts, BlenderProc enhances the robustness of feature matching pipelines.

3. Methods

The proposed pipeline is described as shown in Figure 1. The pipeline utilizes the generated synthetic dataset and can use both GIM and LIMAP to locate the target object, which in our case is the SBB train door.

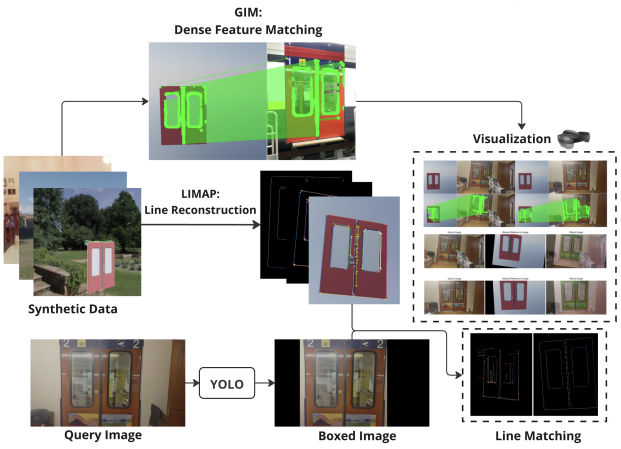


Figure 1. **Pipeline of GIM-based Method and LIMAP.** The top part shows the pipeline of GIM-based feature matching. GIM-based method first conducts dense feature matching between the reference image and query images, then uses matched features to realize image warping and pose estimation. The bottom part presents the pipeline of LIMAP-based method. LIMAP-based method first extracts both line features and point features, then utilizes YOLO bounding box to remove outliers. LIMAP-based method use filtered features to realize localization.

3.1. GIM-based Matching and Localization

Generalizable Image Matcher (GIM) [22] employ internet video to enhance the robustness and generalizability in challenging scenarios.

Feature Matching. We use GIM-DKM model to perform the dense feature matching between the reference image and query images. As shown in Figure 1, the reference image is a front-view image of synthetic dataset, while the query images are from real-world dataset captured with iPad and HoloLens. Thus, there is the large domain shift between the reference image and query images, which causes obstacles to feature matching and further localization.

Localization. We perform robust absolute pose estimation with LO-RANSAC [4] followed by non-linear refinement to realize localization. Given the depth $\mathbf{D}_{\text{ref}}$, camera intrinsics $\mathbf{K}_{\text{ref}}$ of the reference camera, intrinsics $\mathbf{K}_{\text{query}}$ of the query camera, and matched features pairs $\{\mathbf{P}^c_{\text{ref}}, \mathbf{P}^c_{\text{query}}\}$ at the camera coordinate, we compute the 3D coordinates at the world coordinate of keypoints in the reference image by (we suppose that the world coordinate is same with the camera coordinate of the reference image):

$$\mathbf{P}^w_{\text{ref}} = \mathbf{K}^{-1}_{\text{ref}} \mathbf{P}^c_{\text{ref}}, \tag{1}$$

and we have

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathbf{K}_{\text{query}} [\mathbf{R}|\mathbf{t}] \mathbf{P}^w_{\text{ref}}, \tag{2}$$

$$\mathbf{P}^c_{\text{query}} = \begin{bmatrix} x/z \\ y/z \end{bmatrix}. \tag{3}$$

Then we utilize the pose estimation[1] of COLMAP [21] to realize localization.

3.2. LIMAP-based Matching and Localization

LIMAP is a localization toolbox utilizing line features, initially introduced in the paper 3D Line Mapping Revisited [13].

Reconstruction As shown in Figure 1, given a sequence of synthetic images, LIMAP [13] first extracts point and line features with SuperPoint [6] and DeepLSD [15] respectively, uses GlueStick [16] for point and line feature matching, and JLinkage for vanishing point estimator. Then it performs triangulation to reconstruct 3D lines, point clouds, and saves these 3D geometry information for future localization.

Line Feature Matching. During localization, LIMAP performs feature matching between synthetic reference images and real-world query images with the same feature extraction and matching process as demonstrated in the reconstruction part. Then it will associate the 2D lines / points of query images and 3D lines / points stored during the reconstruction.

Query Image Bounding with YOLO. To improve the quality of feature matching, the pipeline removes noisy line features outside the target object in the query image. This is achieved by leveraging a bounding box proposed by YOLO [18]. To enable bounding box generation, the pipeline trains a YOLOv8 [10] network. The YOLO pipeline structure is illustrated in Figure 2.

Localization. With the association of 2D lines / points and 3D lines / points, the pipeline utilizes the pose estimation of COLMAP [21] to realize localization.

[1] pycolmap.estimate_and_refine_absolute_pose()

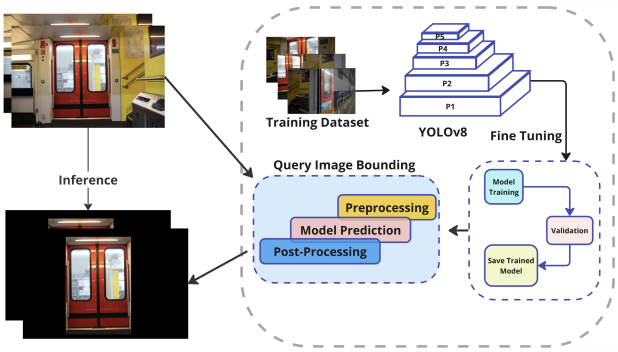


Figure 2. **YOLOv8 Bouding Box Pipeline.** The figure illustrates the image inference process and YOLOv8 model training pipeline. The query image undergoes the preprocessing and is then fed into the trained YOLOv8 model. The training dataset of annotated images is used to fine-tune the model through a process involving model training, validation and model saving. Once fine-tuned, the model processes the query image by performing tasks of preprocessing, model prediction, and post-processing to identify and bound relevant features in the image. The resulting inference highlights SBB door areas within the query image, bounds regions around detected objects and remove the uninterested areas.

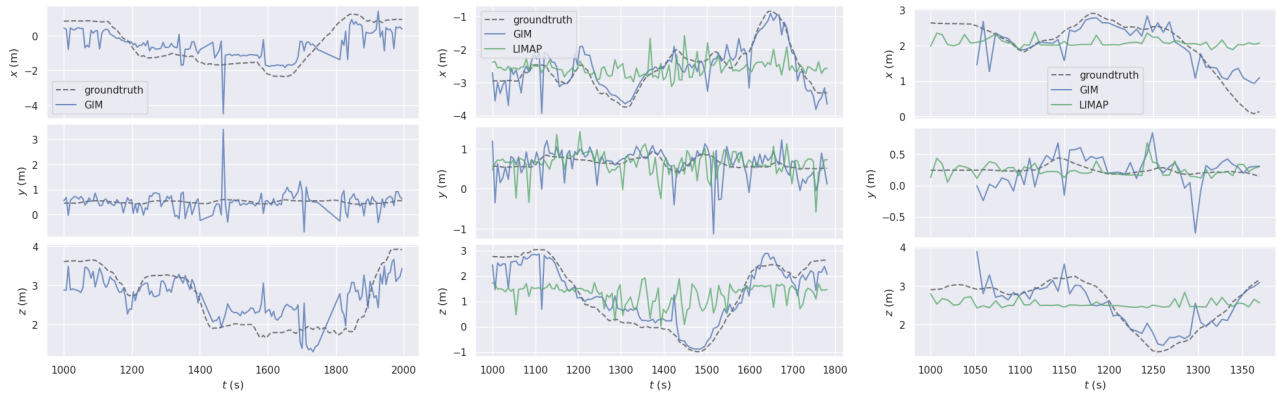
3.3. Hololens and Workstation Deployment

For the deployment of out final pipeline onto Hololens 2, we used a workstation to host the localization and feature-matching algorithms. It processes input data, computes results, and communicates with the HoloLens. While HoloLens 2 acts as the client for visualization. It captures an image and sends it to the workstation, then waits for processed data and renders the results in Unity. A TCP socket-based communication protocol is implemented for reliability. Real-time information is also displayed in HoloLens 2 for monitoring the connection, processing progress and data flow.

4. Experiments

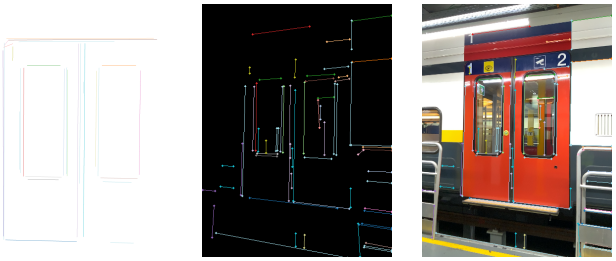
4.1. Datasets and Metrics

Datasets. We use synthetic and real-world datasets to evaluate the performance of our method. We use BlenderProc [5] with the CAD model of the SBB doors to generate synthetic datasets. For each group of synthetic images, the generator first sets the scene background using a random or specified HDR image from polyhaven.com. Next, the CAD model of the target object is positioned within the scene. The scene camera is programmed to randomly move around while maintaining its focus on the point of interest, i.e., the target object, capturing snapshots from various directions. Following this process, the generator produces a set of synthetic images, all sharing the same background but captured from different camera poses. The correspond-



seq1 seq2 seq3

Figure 3. **Qualitative Comparisons of Trajectory Estimation.** Given the reference image from synthetic dataset and a sequence of real-world images, GIM-based method is able to realize stable feature matching and more accurate pose estimation. LIMAP-based method fails to achieve accurate point and line feature matching due to large domain shift, thus leading to the failure on all sequences.



(a) w/o Bounding Box

(b) w/ Bounding Box

Figure 4. **Comparison of LIMAP 2D-3D Line Matching.** Rows represent: (a) without Bounding Box and (b) with Bounding Box. Columns correspond to line reconstruction, 2D lines extraction of query images, and query images, respectively. Lines with same color in the line reconstruction (1st column) and extracted 2D lines (2nd column) denote 2D-3D line correspondence. The comparison reveals that applying bounding boxes significantly reduces noise outside the door areas. However, the large domain shift between the synthetic reference images and real-world query images lead to lots of mismatched lines.

ing camera pose information is also recorded and stored in the dataset for further usage. Besides, we capture several sequences of real-world data with iPad and HoloLens. We use 3 sequences of iPad datasets to evaluate the performance of feature matching and localization. We perform online testing with HoloLens datasets.

Metrics. We evaluate the performance regarding feature matching and localization with commonly used metrics, absolute trajectory error (ATE). The ATE error metric is commonly used for trajectory estimation evaluations in the visual odometry community. Due to the lack of groundtruth matched features of all datasets, we warp the synthetic reference image to the real-world query image and conduct qualitative comparisons for the overlap of SBB doors in both images.

4.2. Experimental Results

Table 1. **Quantitative Comparison of Localization on Synthetic dataset.** LIMAP-based method achieves better performance in terms of translation and rotation due to the introduction of line features.

Method	Translation (m)	Rotation (deg)
Point-based	0.082	0.396
LIMAP-based [13]	**0.032**	**0.145**

Table 2. **Comparison of Localization on Real-world Dataset.** x denotes that the method fails in this sequence. GIM-based method realizes smaller absolute trajectory error (m) on 3 sequences due to its generalizability on datasets with domain shift.

Method	seq1	seq2	seq3
LIMAP-based [13]	x	1.368 $\pm$ 0.499	0.958 $\pm$ 0.417
GIM-based [22]	**1.005** $\pm$ 0.473	**0.659** $\pm$ 0.378	**0.564** $\pm$ 0.346

Matching Warping

(a) Front View

Matching Warping

(b) Lateral View

Matching Warping

(c) Partial Observation

Matching Warping

(d) Internal and Partial Observation

Figure 5. **Comparison of GIM Matching and Warping Results for Four Cases.** Each row corresponds to a different case: (a) to (d). In the warping results of each row, the leftmost image represents the query image, the middle image shows the synthetic SBB train door used as the reference, and the rightmost image displays the result of warping the synthetic SBB train door onto the real-world query image. Across all perspectives, GIM demonstrates its effectiveness in accurately identifying the SBB train doors and performing precise feature matching and warping, showcasing its robustness and reliability in handling challenging scenarios.

Quantitative evaluation results. We evaluate the performance of the GIM-based [22] method and LIMAP-based [13] methods in terms of trajectory estimation on both synthetic datasets and real-world datasets. Table 1 shows the results of LIMAP-based method [13] against point-based method (SuperPoint [6] + HLoc [19]) on synthetic dataset. It demonstrates the superiority of LIMAP-based method on low-texture images such as the SBB door. The introduction of line features improves the robustness of localization. Moreover, we conduct experiments of LIMAP-based method on real-world datasets. Table 2 show that LIMAP-based method fails on the real-world dataset due to large domain shift between the reference image and the query images. There are lots of mismatched line features even with the help of YOLO bounding box. However, the GIM-DKM model can realize more accurate feature matching and stable localization on datasets with large domain shift, as shown in Table 2.

Figure 6. Hololens 2 shows an intuitive button for pipeline activation. Also, during the process, clear output information is provided to either indicate the processing progress or give instructions for better usage of the system.

Qualitative evaluation results. We show the trajectory of 3 real-world sequences in Figure 3. GIM-based method can estimate the overall motion on all 3 sequences thanks to the powerful generalizability of GIM [22] features compared with LIMAP-based method (Superpoint [6] and DeepLSD [15]). To further compare the robustness of LIMAP and GIM, we visualize the results of feature matching in Figure 4 and Figure 5. Figure 4 presents the results for the line matching of LIMAP [13]. Since there is the large domain shift between the synthetic reference image and real-world query images, we can observe many mismatched line features even with the YOLO bounding box, demonstrating that LIMAP fails to handle the domain shift between the matched images. Figure 5 shows that GIM performs accurate feature matching even with the large domain shift. With the points correspondence, we can compute the homography matrix and conduct image warping from the reference image to the query image. Figure 5 shows good overlap between the SBB door of warped reference image and of query images, which further reveals the high accuracy of feature matching for GIM [22]. Besides, we also conduct online experiment with HoloLens and the result is shown in Figure 6.

5. User Study

To assess the user experience with our hololens application, eight participants were invited to conduct a user experience questionnaire. The evaluation process consisted of

three main stages: first, participants were introduced to the project background and workflow pipeline. Next, they were invited to use the hololens in real time to scan an SBB door using our application. Finally, participants completed the questionnaire to give evaluation results in various aspects of the workflow pipeline and provide feedback.

The questionnaire aimed to evaluate participants' overall satisfaction with our pipeline, the perceived acceptance of the results, the reasonableness of the project, and detailed ratings on functionality, result quality, efficiency (response speed), robustness, and usability.

All participants provided positive feedback on the workflow pipeline, indicating that the HoloLens application was generally well-received. The bar chart(Figure 7) illustrates the results of overall satisfaction, perceived reasonableness and the rating scores on pipeline detailed evaluation.

All participants thought our pipeline is generating an acceptable result and expressed high satisfaction with the pipeline.In addition to the overall feedback, participants rated specific aspects of the pipeline, including functionality, result quality, efficiency(response speed), robustness and usability. Most ratings exceeded 4 on a 1–5 scale, except for robustness, where participants found it difficult to provide an objective judgement.

Some participants also provided qualitative suggestions for further improvement. One participant suggested showing "door hologram in 3D", while the other recommended providing additional guidance during the workflow. Based on these insights, the pipeline has been enhanced to be able to localize the SBB door, which is a critical step toward rendering the door hologram in the HoloLens.

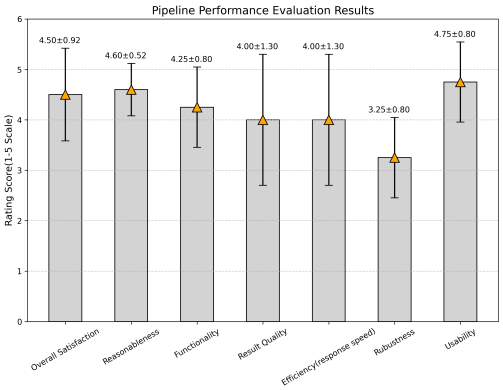


Figure 7. Pipeline Performance Evaluation Results on a 5-point Likert Scale. Users rated most dimensions at or above 4, indicating that the pipeline's performance exceeded their satisfaction levels in Overall Satisfaction, Reasonableness, Functionality, Result Quality, Efficiency and Usability. However, for the Robustness dimension, users expressed uncertainty, stating that they were unable to provide a confident evaluation. This might due to the limited time users had to experience the HoloLens application, making it difficult to form a thorough assessment in the pipeline robustness.

6. Conclusion and Future Works

In this project, we investigate the performance of line features and generalizable image matcher on localization. LIMAP fails to extract robust feature matching due to severe domain shift, but GIM shows good generalizability of feature matching under challenging scenarios.

As we only use one reference image for feature matching, pose estimation is not accurate enough when objects are occluded and vulnerable to mismatched feature points. In the future, we will utilize several reference images to estimate multiple candidate camera poses and use all candidates to vote for final poses to improve the performance of localization.

References

[1] Cuneyt Akinlar and Cihan Topal. Edlines: Real-time line segment detection by edge drawing (ed). In *2011 18th IEEE International Conference on Image Processing*, pages 2837–2840, 2011. 1

[2] Adrien Bartoli and Peter F. Sturm. Structure-from-motion using lines: Representation, triangulation, and bundle adjustment. *Comput. Vis. Image Underst.*, 100:416–441, 2005. 2

[3] Adrien Bartoli, Mathieu Coquerelle, and Peter Sturm. A Framework For Pencil-of-Points Structure-From-Motion. In *European Conference on Computer Vision*, pages 28–40, Prague, Czech Republic, 2004. Springer-Verlag. 2

[4] Ondřej Chum, Jiří Matas, and Josef Kittler. Locally optimized ransac. In *Pattern Recognition: 25th DAGM Symposium, Magdeburg, Germany, September 10-12, 2003. Proceedings 25*, pages 236–243. Springer, 2003. 3

[5] Maximilian Denninger, Dominik Winkelbauer, Martin Sundermeyer, Wout Boerdijk, Markus Knauer, Klaus H. Strobl, Matthias Humt, and Rudolph Triebel. Blenderproc2: A procedural pipeline for photorealistic rendering. *Journal of Open Source Software*, 8(82):4901, 2023. 2, 3

[6] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superpoint: Self-supervised interest point detection and description. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 224–236, 2018. 3, 5

[7] Shuang Gao, Jixiang Wan, Yishan Ping, Xudong Zhang, Shuzhou Dong, Yuchen Yang, Haikuan Ning, Jijunnan Li, and Yandong Guo. Pose refinement with joint optimization of visual points and lines. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2888–2894. IEEE, 2022. 2

[8] Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Girshick. Mask r-cnn. In *ICCV*, 2017. 2

[9] Xingyi He, Jiaming Sun, Yuang Wang, Di Huang, Hujun Bao, and Xiaowei Zhou. Onepose++: Keypoint-free one-shot object pose estimation without CAD models. In *Advances in Neural Information Processing Systems*, 2022. 1, 2

[10] Glenn Jocher, Jing Qiu, and Ayush Chaurasia. Ultralytics YOLO, 2023. 1, 3

[11] Feng Li, Hao Zhang, Peize Sun, Xueyan Zou, Shilong Liu, Jianwei Yang, Chunyuan Li, Lei Zhang, and Jianfeng Gao. Semantic-sam: Segment and recognize anything at any granularity. *arXiv preprint arXiv:2307.04767*, 2023. 2

[12] Kai Li, Jian Yao, Li Li, and Yahui Liu. 3d line segment reconstruction in structured scenes via coplanar line segment clustering. In *ACCV Workshops*, 2016. 2

[13] Shaohui Liu, Yifan Yu, Rémi Pautrat, Marc Pollefeys, and Viktor Larsson. 3d line mapping revisited. In *Computer Vision and Pattern Recognition (CVPR)*, 2023. 1, 2, 3, 4, 5

[14] Yicheng Luo, Jing Ren, Xuefei Zhe, Di Kang, Yajing Xu, Peter Wonka, and Linchao Bao. Learning to construct 3d building wireframes from 3d line clouds. *arXiv preprint arXiv:2208.11948*, 2022. 2

[15] Rémi Pautrat, Daniel Barath, Viktor Larsson, Martin R. Oswald, and Marc Pollefeys. Deeplsd: Line segment detection and refinement with deep image gradients. In *Computer Vision and Pattern Recognition (CVPR)*, 2023. 3, 5

[16] Rémi* Pautrat, Iago* Suárez, Yifan Yu, Marc Pollefeys, and Viktor Larsson. GlueStick: Robust image matching by sticking points and lines together. In *International Conference on Computer Vision (ICCV)*, 2023. 3

[17] A. Pumarola, A. Agudo, L. Porzi, A. Sanfeliu, V. Lepetit, and F. Moreno-Noguer. Geometry-Aware Network for Non-Rigid Shape Prediction from a Single View. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 2

[18] J Redmon. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016. 2, 3

[19] Paul-Edouard Sarlin, Cesar Cadena, Roland Siegwart, and Marcin Dymczyk. From coarse to fine: Robust hierarchical localization at large scale. In *CVPR*, 2019. 1, 5

[20] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. SuperGlue: Learning feature matching with graph neural networks. In *CVPR*, 2020. 1

[21] Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 3

[22] Xuelun Shen, Zhipeng Cai, Wei Yin, Matthias Müller, Zijun Li, Kaixuan Wang, Xiaozhi Chen, and Cheng Wang. Gim: Learning generalizable image matcher from internet videos, 2024. 1, 2, 4, 5

[23] Dong Wei, Yi Wan, Yongjun Zhang, Xinyi Liu, Bin Zhang, and Xiqi Wang. Elsr: Efficient line segment reconstruction with planes and points guidance. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 15786–15794, 2022. 2

[24] Yifan Xu, Weijian Xu, David Cheung, and Zhuowen Tu. Line segment detection using transformers without edges. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4257–4266, 2021. 2

[25] Yichao Zhou, Haozhi Qi, and Yi Ma. End-to-end wireframe parsing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 962–971, 2019. 2