

# Automated Extraction of Pipe Geometry Using SAM for Mixed Reality Inspection Tasks

Sajjad Einizinab<sup>1,2</sup>, Kourosh Khoshelham<sup>1,2</sup>, Stephan Winter<sup>2</sup>, Philip Christopher<sup>1,2</sup>

<sup>1</sup> Building 4.0 CRC, Caulfield East, 3145, Victoria, Australia

<sup>2</sup> Dept. of Infrastructure Engineering, University of Melbourne, Parkville, 3010, Victoria, Australia  
(einizinabs, k.khoshelham, winter, pbc)@unimelb.edu.au

**Keywords:** Automation, Mixed Reality, Segment Anything Model (SAM), Pipe geometry, Inspection.

## Abstract

Accurate detection and measurement of building elements are essential for efficient automated inspection and quality assessment in construction. This study evaluates the effectiveness of the Segment Anything Model (SAM) for pipe segmentation using a Mixed Reality-based dataset and introduces an automated method for pipe 3D centreline reconstruction and diameter estimation. The impact of the input point prompt distribution and number on segmentation accuracy is analyzed, identifying optimal configurations for improved performance. Using depth data and pose information from the MR device, the proposed approach reconstructs the 3D centreline and estimates pipe diameters with high reliability. The method is evaluated in a real experimental pipe network. The results indicate that the use of five-point prompts in a uniform distribution achieves approximately 90% precision and recall for pipe segmentation, with median position and diameter errors of 33 mm and 10 mm, respectively. The findings highlight the ability of the MR system to achieve accurate pipe positioning and diameter estimation, particularly in pipe networks with moderate complexity and fewer thin pipes, where segmentation and measurement challenges are minimized.

## 1. Introduction

In recent years, the rapid advancement of automation technologies has revolutionized the construction sector, significantly improving the efficiency and quality of construction work (Zhang et al., 2019). One area that can particularly benefit from this transformation is the inspection of building Mechanical, Electrical and Plumbing (MEP) systems including pipes. Pipe inspection involves a precise assessment of the geometric characteristics of the pipes such as their position and size after installation to ensure proper functioning and integration with the other building elements as well as compliance with regulations (Einizinab et al., 2023).

Non-contact sensing methods, including photogrammetry and laser scanning, have advanced MEP inspections by capturing precise 3D geometric data; however, these technologies face limitations due to their high computational requirements and the non-simultaneous nature of data acquisition and inspection processes, limiting their applicability for real-time inspection tasks (Becker et al., 2023). Mixed Reality (MR) offers a more efficient approach by seamlessly integrating virtual and physical elements to enable precise real-time inspections (Einizinab et al., 2023). MR systems equipped with an RGB camera, depth sensor, and inertial measurement unit allow inspectors not only to visualize Building Information Models (BIM) overlaid on real-world structures, enabling effective comparisons between as-built and as-designed models, but also to perform geometric measurements directly within the mixed reality environment (Radanovic et al., 2023).

Despite the advanced capabilities of MR systems, the inspection process relies on manual measurements through the MR device and visual confirmation by the inspectors, which limits efficiency and increases the risk of human error (Einizinab et al., 2023). Leveraging state-of-the-art image segmentation techniques, such as the Segment Anything Model (SAM) (Kirillov et al., 2023) and multiple sensors integrated into

MR devices presents an opportunity to automate the geometric measurement process within MR, reducing the need for manual input while enhancing both accuracy and efficiency. SAM, a promptable vision-language model, can segment objects in images without additional training, making it highly suitable for geometric pipe measurement using MR (Wang et al., 2024). Among the available input prompts for SAM, i.e., points, boxes, and their combinations, point-based prompts stand out in MR-based automated measurements due to their ease of use.

To effectively leverage SAM for automated MR measurements, two critical questions need to be addressed. First, what is the optimal number and distribution of input point prompts required to achieve accurate segmentation of target pipes? Determining this will be crucial to improve the reliability and efficiency of MR-based inspection workflows. Second, after isolating the pipes using the optimal point prompt strategy, how accurately can the MR system position the pipes and measure their size, such as diameter? Addressing these questions will provide insights into improving the integration of SAM with MR systems for precise and automated geometric measurements.

To address these questions, we applied the SAM segmentation method to RGB images captured by an MR device and experimented with varying numbers and distributions of input point prompts to identify the optimal configuration for accurate pipe segmentation. Based on the results, we propose an automated method for reconstructing pipe positions and measuring pipe diameters. In this method, the 2D segmented pipes generated by SAM are used to create a centreline in the image space. Using depth data from the MR device's built-in camera and the device's position and orientation (pose) obtained from the Simultaneous Localization and Mapping (SLAM) technique, a 3D representation of the pipe's centreline is reconstructed and its diameter is estimated. The proposed method is evaluated on a dense pipe system in a real-world scenario by comparing the reconstructed results with ground-truth values obtained from the corresponding BIM model.

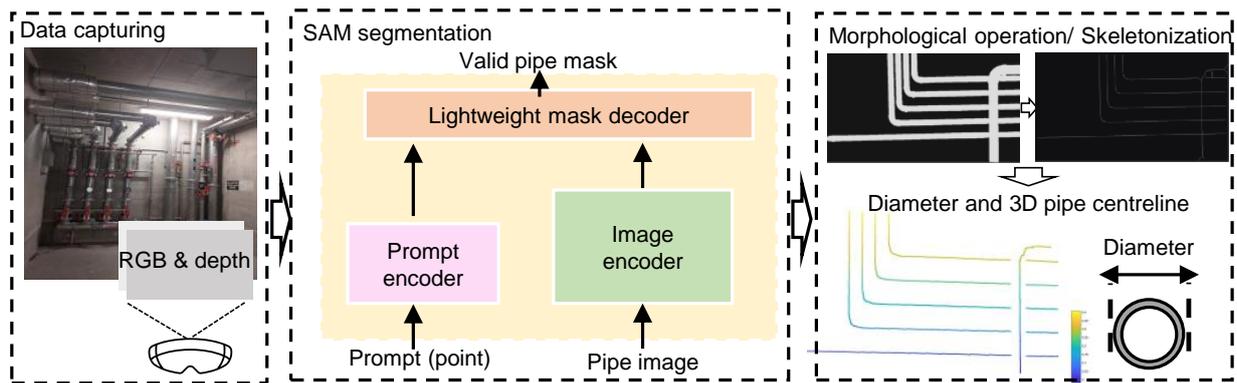


Figure 1. Overview of the proposed method.

## 2. Related works

The segmentation of pipes in images is a critical aspect of automated, non-contact inspection workflows, as it facilitates the accurate identification and delineation of pipe components. Image segmentation is a core task in computer vision, with wide-ranging applications across various domains (Wang et al., 2024). It enables precise object detection and localization within images, which is essential for analyzing and interpreting visual data. Traditional semantic segmentation techniques rely on feature extraction methods such as Histograms of Oriented Gradient (HOG) and Scale-Invariant Feature Transform (SIFT) descriptors to identify significant features (Hussein et al., 2023). This approach involves dividing the image into small patches, which are subsequently classified using local classifiers such as random decision forests or support vector machines (Thoma, 2016). However, the use of small spatial windows often leads to noisy predictions and substantial computational overhead, as each patch must be individually processed.

Recent developments in deep learning have led to the emergence of advanced semantic segmentation techniques, such as U-Net (Ronneberger et al., 2015), Faster R-CNN (Ren et al., 2016), Mask R-CNN (He et al., 2017), DeepLab (Chen et al., 2017), and Pyramid Scene Parsing Network (PSPNet) (Zhao et al., 2017). U-Net employs an encoder-decoder structure to deliver highly accurate segmentation results. Mask R-CNN, an extension of the Faster R-CNN model for object detection, incorporates instance segmentation capabilities. DeepLab utilizes atrous convolutions and fully connected conditional random fields to improve semantic segmentation performance in natural images. Similarly, PSPNet addresses scene parsing tasks by employing pyramid pooling to capture global contextual information and enhance segmentation accuracy. Despite their effectiveness, these models typically require training dataset and are not designed for general-purpose segmentation across a diverse range of applications (Wang et al., 2024).

SAM (Kirillov et al., 2023), one of the first foundational models in computer vision, addresses this limitation by offering a promptable framework for general image segmentation. SAM can segment a variety of objects across different types of images without the need for retraining, making it particularly suitable for pipe detection tasks. Its versatility, combined with its computational efficiency and adaptability, underscores its potential for use in MR applications, especially in building pipe inspections where accurate and flexible segmentation is essential.

Although SAM has introduced significant advancements in im-

age segmentation, its effectiveness in MR-based real-world applications remains to be thoroughly evaluated. Acknowledging the prior studies (Zhang et al., 2023, Wang et al., 2024, Ying et al., 2025) that have examined SAM's capabilities in segmentation tasks, this study addresses two key challenges: the number of input point prompts and their spatial distribution within the scene for MR-based pipe inspection tasks. Furthermore, this study evaluates the performance of SAM-based pipe segmentation for subsequent 3D centreline reconstruction and diameter estimation of pipes using an MR device.

## 3. Method

The proposed method includes three main steps: data capture by the MR device, segmentation using SAM, and geometric reconstruction of the centreline and diameter estimation of the pipes (Figure 1).

### 3.1 Pipe segmentation using SAM

SAM is utilized to identify and segment pipes from the RGB images. SAM is composed of three key components: an image encoder, a prompt encoder (based on points in this study), and a mask decoder (Kirillov et al., 2023). The image encoder is built on a pre-trained Vision Transformer (ViT), which is specifically adapted to handle high-resolution images. After the image and prompt data are encoded, the mask decoder processes the data using a modified transformer decoder. This decoder enables bidirectional cross-attention between the prompts and the image features, enhancing the model's ability to focus on relevant areas. Finally, an upsampling followed by a multi-layer perceptron calculates the mask's foreground probability, producing a segmentation of the pipe structures.

### 3.2 Pipe geometry measurements

The SLAM algorithm built in MR devices integrates sensor data to track the device's position and orientation in real-time. It extracts environmental features and registers consecutive depth images to build a spatial map and estimate the pose of the depth camera in a global coordinate system, with the origin being the point where the first depth image is captured. In this global system, the poses of both the RGB camera and the depth sensor are known, which enables the transformation of each RGB pixel to align with its corresponding depth pixel. In other words, for every pixel in the RGB image, an exact corresponding depth value or an interpolated depth value from the neighboring pixels will be determined.



Figure 2. Schematic representation of the distribution scenarios with four input point prompts (yellow dots): Uniform distribution (left), Clustered distribution (middle), and Boundary distribution (right).

After segmenting pipes from the RGB images using SAM, the resulting binary image is pre-processed with morphological operations to ensure continuous pipe structures. A skeletonization algorithm (Zhang and Suen, 1984) is then applied to extract the centreline of the pipes, reducing them to a single-pixel-wide line.

A distance transform is used to calculate the distance from each centreline pixel to the nearest pipe boundary, which represents the pipe's radius in pixels. These pixel-based radius values are converted to real-world measurements using depth information, pose values, and the camera's intrinsic parameters. The radius is then doubled to determine the pipe diameter at each centreline pixel.

The centreline image contains the pixel-wise diameters, and the depth image provides the distance to the pipe's surface. Using the camera's intrinsic parameters and depth values, pixel coordinates are transformed into 3D coordinates within the MR camera's local frame. The pre-calculated radius of the pipe is added to the depth value to calculate the 3D coordinates of the pipe's centreline, rather than the surface centreline of the pipe. The 3D coordinates are then converted from the camera's local frame to the global coordinate system using the MR camera poses generated by the SLAM algorithm.

### 3.3 Evaluation

The performance of SAM with varying numbers and distributions of input point prompts for pipe segmentation is evaluated using standard image segmentation metrics, including Intersection over Union (IoU), recall, and precision. Ground-truth annotations of pipes are created in multiple sample images to support this evaluation.

For evaluation of SAM segmentation, the primary goal is to identify the optimal number and distribution of input point prompts. We define three distinct point distributions and, for each distribution scenario, we assess SAM performance by varying the number of input point prompts from 1 to 5 and calculate the resulting segmentation metrics. Furthermore, in each distribution scenario, for different numbers of input point prompts, we ensure that the points are selected in similar locations across all cases to maintain a consistent distribution pattern. The three distribution scenarios are as follows:

1. Uniform distribution: Points are evenly distributed across the entire image, centered on the pipes.
2. Clustered distribution: Points are concentrated within a specific region of the image.
3. Boundary distribution: Points are uniformly distributed along the edges or boundaries of the pipe.



Figure 3. Case study environment (left) and its corresponding BIM model (right).

Figure 2 illustrates the schematic representation of the distribution scenarios with four input point prompts. In the uniform and boundary distributions, the points are evenly distributed across the image, but their positions on the pipes differ: one set of points is placed at the center of the pipes, while the other set is located at the edges. In the clustered distribution, points are selected either at the center or the edge of the pipes, but all the points are located within a specific region of the image. In all scenarios, we assign the point prompts to the entire image, leaving some pipes without prompts. This choice influences the segmentation results and allows us to assess the ability of SAM to identify unselected pipes within the image.

The accuracy of pipe diameter and 3D centreline is then assessed by comparing the MR-derived measurements with the corresponding pre-aligned BIM pipe data. The analysis is limited to scenarios where the pipes are accurately segmented by SAM, minimizing the impact of segmentation inaccuracies.

For each point in the MR-derived centrelines, the nearest neighbor in the BIM pipe centrelines is identified, and an error metric is calculated as the distance between corresponding points. This error is evaluated in two aspects: spatial error, which measures the deviation in 3D positions, and diameter error, which quantifies the difference between the estimated and actual pipe diameters, which is extracted from the BIM. The average 3D positioning and diameter errors of the points for each pipe element are used to determine the accuracies for that particular pipe.

## 4. Experimental results and analysis

### 4.1 Experiment design

We utilized an experimental pipe network with a dense layout located on the ground floor of the Melbourne Connect building at the Department of Infrastructure Engineering, University of Melbourne. The network consists of numerous pipe segments ranging in diameters from 15 to 150 mm and pipe spacings

Distribution	Uniform					Clustered					Boundary				
Prompt No.	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Precision	0.98	0.72	0.84	0.84	<b>0.89</b>	0.98	0.97	0.96	0.96	0.96	0.98	0.76	0.86	0.74	0.76
Recall	0.13	0.86	0.86	0.92	<b>0.91</b>	0.13	0.19	0.13	0.17	0.16	0.18	0.69	0.88	0.94	0.92
IoU	0.13	0.64	0.74	0.79	<b>0.82</b>	0.13	0.19	0.13	0.17	0.16	0.18	0.56	0.77	0.70	0.71

Table 1. Evaluation metrics for SAM-based pipe segmentation across three distribution scenarios with varying numbers of input point prompts.

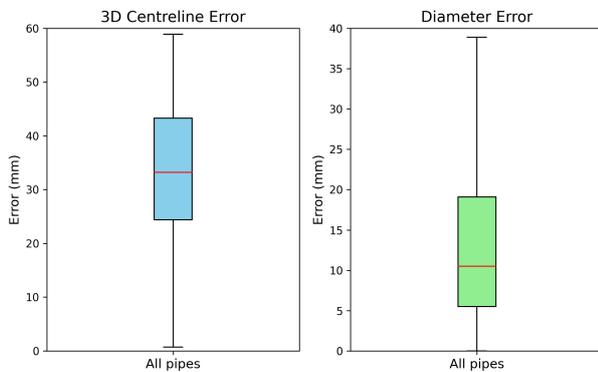


Figure 4. Distribution of errors in 3D centreline reconstruction and diameter estimation of pipes.

varying between 10 mm and higher. The pipe network layout includes numerous intersecting segments, with some areas having pipes positioned so closely that distinguishing them in distant scans might pose a challenge for segmentation. The evaluation focuses on SAM’s segmentation capability under varying distributions and number of input point prompts, as well as 3D centreline reconstruction and diameter estimation of the well-segmented pipes.

Data from the pipe system were collected using Microsoft HoloLens 2. For segmentation evaluation, 15 RGB images and their corresponding depth images, captured at distances ranging from 1 to 2.5 meters from the pipes, were used. The images were annotated using the online tool V7 Darwin (V7Labs, n.d.), with each annotated image converted into a binary mask where pipe pixels were labeled as ‘1’. The annotated images serve as the ground-truth for segmentation evaluation.

Regarding the SAM architecture and inputs, we employed the ViT-H image encoder and the model was implemented on a laptop equipped with a NVIDIA(R) GeForce(R) RTX(TM) 4070 GPU, and 64GB (2×32GB) RAM.

Additionally, ground-truth data for 3D reconstruction and diameter measurements were derived from a BIM model of the environment. A terrestrial laser scanner was used to capture 3D point cloud representations of the space, which were then processed in Autodesk Revit 2023 to generate the BIM model. For evaluation, the BIM model was aligned with the real-world environment, ensuring that each real pipe object had a corresponding virtual object from the aligned BIM. Figure 3 illustrates the experimental environment and its corresponding BIM model.

## 4.2 Results

Table 1 presents the results of SAM-based pipe segmentation in three distribution scenarios: uniform, clustered, and boundary—using varying numbers of input point prompts from 1 to 5. The results demonstrate that the model consistently achieves high precision in majority of the scenarios, indicating that the

most of detected pipe pixels are accurately classified, with minimal false positives, even though some pipes lack point prompts. This suggests strong performance in pipe identification. Figure 5 illustrates a sample RGB image of the pipe layout, along with its ground-truth annotated mask image and the pipes segmented using SAM.

However, recall and IoU values show a clear dependency on the number of input point prompts across all distribution scenarios. In the uniform and boundary distributions, increasing the number of input point prompts leads to a steady improvement in recall and IoU values, reaching high values when more than one point prompt is used. The high recall and IoU in this scenario indicate that the model is able to detect a large proportion of the true pipe pixels, with a good overlap between the predicted and ground-truth pipe regions.

In contrast, the clustered distribution scenario yields relatively low recall and IoU values, regardless of the number of input point prompts. This suggests that concentrating the input point prompts in a specific region of the image limits the segmentation accuracy, as the model struggles to generalize across the entire scene.

The uniform distribution, where input points are evenly distributed across the scene, proves to be the most effective strategy for improving segmentation accuracy. When points are selected along the centreline of the pipes, the accuracy increases significantly compared to selecting points along the edges or boundaries.

Among the assessed scenarios, uniform and boundary distributions with more than three input point prompts exhibit similarly high performance. However, the uniform distribution shows a slight advantage, making it the most reliable choice to achieve optimal segmentation accuracy, particularly as the number of input point prompts increases. Furthermore, the capability of SAM is evident in its ability to accurately isolate pipes even without assigned point prompts among numerous pipes in the image, as shown in Figure 5.

Figure 4 demonstrates the accuracy of the 3D centreline reconstruction and diameter estimation for the detected pipes through SAM. For 3D centreline positioning, the overall range of errors indicates a noticeable spread, with values extending from nearly 0 to 60 mm. The median value, located near the middle of this range, suggests that the majority of errors are concentrated around this central value. The interquartile range is relatively smaller, indicating that most positioning errors fall within a narrower range (25–43 mm). However, the broader range between the minimum and maximum highlights a few instances of significant discrepancies, reflecting outliers in the results.

In terms of diameter estimation, the overall error range is also wide, with the minimum error being close to 0 and the maximum error up to 40 mm. The median value, placed towards the lower end of the distribution, suggests that most of the diameter

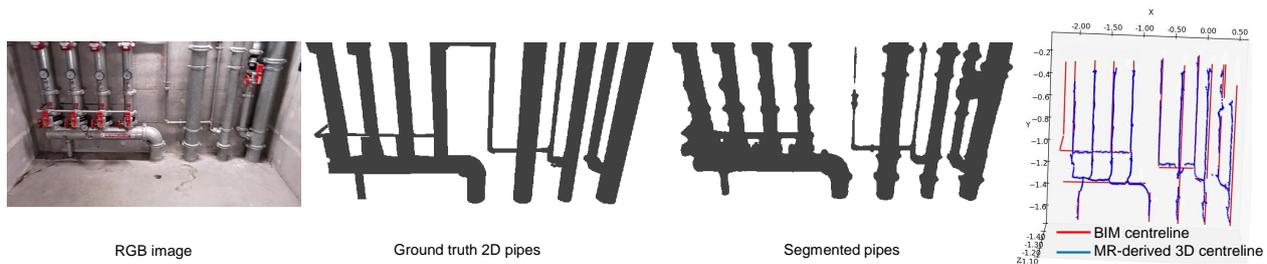


Figure 5. A sample RGB image of the pipe layout, its ground-truth annotated mask image, SAM-segmented pipes, and the aligned 3D pipe centreline from the BIM model overlaid on the reconstructed 3D pipe centreline extracted from MR image.

estimation errors are relatively small. The interquartile range for diameter errors is more compact, indicating that the majority of the diameter estimates are concentrated within a certain range (5–18 mm). Similar to the 3D reconstruction, there are some larger errors in diameter estimation, though these outliers are less frequent and do not represent the majority of cases. In both 3D centreline reconstruction and diameter estimation, greater uncertainties arise in the more complex sections of the pipe layout, particularly where pipes are closely spaced and thinner.

Together, these findings show that while the accuracy for both 3D centreline reconstruction and diameter estimation is generally high, there are occasional outliers in both cases that introduce larger errors. Nonetheless, most of the errors are within a moderate range, suggesting reasonable overall performance in terms of pipe inspection tasks using MR devices. This assertion is supported by the fact that the pipe layout in this study is highly complex and features numerous thin pipes and tightly spaced sections.

It is noteworthy that the accuracy of BIM alignment within the MR visualization directly impacts reconstruction and diameter estimation, as the BIM model serves as the ground-truth for measurement accuracies.

The results demonstrate the reliability of MR-based reconstruction and diameter estimation for pipe elements, highlighting the potential for automated MR-based measurements in building inspection tasks. By implementing the proposed method, automatic detection of pipe positions and diameters within the MR environment and comparison with pre-aligned BIM objects, automatic pipe inspection becomes feasible. However, the actual size and spacing of the pipe network play a crucial role. Given the current reconstruction and diameter estimation accuracy, distinguishing closely spaced and thin pipes remains challenging, making precise measurements in dense pipe layouts more complex. Conversely, for moderate and simpler pipe networks, the method proves to be highly reliable. Figure 5 displays a schematic view of the aligned 3D pipe centreline extracted from the BIM model, overlaid on the reconstructed 3D pipe centreline extracted from the MR data.

## 5. Conclusion and Future Works

In this paper, we evaluated the segmentation capabilities of SAM in identifying pipe elements within images captured by an MR device, focusing on the number and distribution of input point prompts. In addition, we introduced an approach for automated measurement and positioning of pipe elements using MR. By selecting the optimal number and distribution of

input point prompts, the method successfully isolated pipes, reconstructed their 3D centrelines, and estimated their diameters using pose values and depth data from the MR device.

A comprehensive evaluation with a complex pipe network layout in various scenarios identified the best input point prompt strategy for the SAM model. The method demonstrated reliable diameter estimation and 3D centreline reconstruction, demonstrating the potential of MR-based automation for pipe inspection tasks. The results indicate that the proposed automatic method is well suited for inspecting pipe networks, particularly those that are not highly intricate or contain thin pipe elements.

Future research can explore the accuracy of results concerning the distance between objects and the MR device, as understanding this relationship is crucial for refining measurements. Implementing automated measurements in real-world pipe inspection scenarios is another key objective. Additionally, fully integrating SAM-based pipe detection within an MR device presents a promising direction for further development.

## 6. Acknowledgments

This research is supported by Building 4.0 CRC. The support of the Commonwealth of Australia through the Cooperative Research Center Program is acknowledged.

## References

- Becker, S., Einizinab, S., Radanovic, S., Khoshelham, K., Mirzaei, K., Fang, Y., 2023. Reality capture methods for remote inspection of building work. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48, 275–281.
- Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A. L., 2017. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4), 834–848.
- Einizinab, S., Khoshelham, K., Winter, S., Christopher, P., Fang, Y., Windholz, E., Radanovic, M., Hu, S., 2023. Enabling technologies for remote and virtual inspection of building work. *Automation in Construction*, 156, 105096.
- He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask R-CNN. *Proceedings of the IEEE International Conference on Computer Vision*, 2961–2969.

- Hussein, G. S., Elseuofi, S., Dukhan, W. H., Ali, A. H., 2023. A Novel Method for Banknote Recognition Using a Combined Histogram of Oriented Gradients and Scale-Invariant Feature Transform.
- Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y. et al., 2023. Segment anything. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 4015–4026.
- Radanovic, M., Khoshelham, K., Fraser, C., 2023. Aligning the real and the virtual world: Mixed Reality localisation using learning-based 3D-3D model registration. *Advanced Engineering Informatics*, 56, 101960.
- Ren, S., He, K., Girshick, R., Sun, J., 2016. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence*, 39(6), 1137–1149.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, Springer, 234–241.
- Thoma, M., 2016. A survey of semantic segmentation. *arXiv preprint arXiv:1602.06541*.
- V7Labs, n.d. V7-AI data platform for ML teams. <https://www.v7labs.com/>. Accessed on January 27, 2025.
- Wang, Y., Zhao, Y., Petzold, L., 2024. An empirical study on the robustness of the segment anything model (SAM). *Pattern Recognition*, 110685.
- Ying, W., Khoshelham, K., Kemp, J., 2025. Assessment of stone decay in heritage sites using machine learning. *In Proceedings of ISPRS Geospatial Week 2025, Dubai, UAE*.
- Zhang, C., Zhang, C., Kang, T., Kim, D., Bae, S.-H., Kweon, I. S., 2023. Attack-SAM: Towards attacking segment anything model with adversarial examples. *arXiv preprint arXiv:2305.00866*.
- Zhang, D., Zhang, J., Xiong, H., Cui, Z., Lu, D., 2019. Taking advantage of collective intelligence and BIM-based virtual reality in fire safety inspection for commercial and public buildings. *Applied Sciences*, 9(23), 5068.
- Zhang, T. Y., Suen, C. Y., 1984. A fast parallel algorithm for thinning digital patterns. *Communications of the ACM*, 27(3), 236–239.
- Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid scene parsing network. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2881–2890.