A study on the

Prompted Image Segmentation using Segment Anything Model (SAM)

Submitted by:

Name:	Tithi Gupta
Role:	AI Research Apprentice
Date	6 Feb, 2026

Table of Contents

ABSTRACT

Crack detection and surface defect analysis play a crucial role in ensuring quality and safety in construction and infrastructure maintenance. Conventional image segmentation techniques typically rely on large amounts of manually annotated data and task-specific model training, making them costly and time-consuming to deploy at scale. Recent advances in foundation vision models offer an alternative approach by enabling prompt-based segmentation with minimal supervision.

This project investigates the application of the Segment Anything Model (SAM) for automated crack segmentation in construction surface images. The proposed approach leverages SAM's zero-shot segmentation capability by using simple point-based prompts to generate crack masks without performing dataset-specific training. Multiple datasets containing images of wall cracks and drywall joint regions are processed through a unified pipeline to produce binary segmentation masks.

Experimental results demonstrate that SAM can effectively localize crack regions across diverse visual conditions, highlighting its potential for scalable and flexible defect detection. The findings suggest that prompt-based segmentation models can significantly reduce annotation effort while maintaining reliable segmentation performance. This work provides insight into the applicability of foundation vision models for construction quality assurance and autonomous inspection systems.

INTRODUCTION

Automation in the construction industry has become increasingly important due to labor shortages, rising operational costs, and the need for improved productivity and quality control. One critical aspect of construction quality assurance (QA) is the timely detection of surface defects such as wall cracks and improper drywall joint taping, which can compromise structural integrity and finishing quality if left unnoticed.

Traditional computer vision approaches for defect segmentation often require large amounts of annotated data and task-specific model training, making them expensive and time-consuming to deploy in real-world settings. Recent advances in foundation vision models have opened new possibilities for prompt-based segmentation, where models can generate accurate masks using minimal human guidance instead of full supervision.

In this project, we explore the use of Meta AI's Segment Anything Model (SAM) for prompted image segmentation in construction QA scenarios. The objective is to evaluate whether a single, generic segmentation model can effectively identify cracks and drywall joint/taping regions across different datasets using simple point-based prompts, without performing dataset-specific training.

By leveraging SAM's zero-shot segmentation capability, this work aims to demonstrate a scalable and flexible approach for defect localization that can be integrated into autonomous inspection systems and robotic perception pipelines in construction environments.

PROBLEM STATEMENT AND OBJECTIVE

Problem Statement:

Construction inspection datasets often lack pixel-level annotations for segmentation tasks. Creating such labels is expensive, time-consuming, and not scalable across projects.

Objectives

1. Use Segment Anything Model (SAM) for crack and drywall defect segmentation
2. Avoid training a custom segmentation model
3. Generate segmentation masks using point prompts
4. Apply the same pipeline across multiple datasets
5. Evaluate feasibility for construction QA automation

DATA STRUCTURE

This project utilizes two publicly available construction inspection datasets for prompt-based segmentation of surface defects. The datasets consist of RGB images collected under real-world construction environments and are used to evaluate the generalization capability of the Segment Anything Model (SAM) without task-specific training.

The datasets undergo structured organization and automated mask generation using a prompt-driven segmentation pipeline.

Step 1: Original Dataset

1. **Dataset Name:** Crack Detection Dataset

Attribute	Description
Dataset Name	Cracks-1
Domain	Construction Surface Inspection
Total Images	~5,000 images
Image Format	JPG / PNG
Color Space	RGB
Resolution	Varying (high-resolution images)
Defect Type	Surface cracks
Source	Roboflow Universe
Annotation Type	Bounding-box annotations (original), no segmentation masks

2. **Dataset Name**: Drywall Joint Detection Dataset

Attribute	Description
Dataset Name	Drywall-Join-Detect-1
Domain	Interior Construction Quality Inspection
Total Images	~3,000 images
Image Format	JPG / PNG
Color Space	RGB
Resolution	Varying
Defect Type	Drywall joints / taping regions
Source	Roboflow Universe
Annotation Type	Bounding-box annotations (original), no segmentation masks

Step 2: Dataset Preparation

At this stage, no pixel-level ground truth segmentation masks are available. Since the objective is to perform segmentation without manual annotation, the datasets are directly processed using a prompt-based segmentation approach.

Key preparation steps include:

- Image loading using OpenCV
- Conversion from BGR to RGB format
- Resolution preserved as original
- Dataset splits processed independently

Step 3: Generated Segmentation Dataset (Prompt-based Masks)

Using the Segment Anything Model (SAM), binary segmentation masks are generated automatically for each image using a single positive point prompt placed at the image center.

Mask Generation Details

- Prompt Type: Single-point prompt
- Output Type: Binary segmentation mask
- Pixel Values: {0, 255}
- Output Format: PNG (single-channel)
- Naming Convention: <image_id>__segment_crack.png

Generated Dataset Structure

Folder Name	Type	Description
train_crack/	Binary mask images	Segmentation masks for training images
valid_crack/	Binary mask images	Segmentation masks for validation images
test_crack/	Binary mask images	Segmentation masks for test images (if available)

METHODOLOGY

1. Problem Formulation

The objective of this work is to automatically generate crack segmentation masks from construction surface images using a prompt-based segmentation approach. Given an input image, the system predicts a binary mask highlighting crack regions without requiring manually annotated pixel-level labels for training.

2. Overall Workflow

The proposed methodology follows a sequential pipeline consisting of dataset preparation, model initialization, prompt-based segmentation, mask generation, and result evaluation. The complete workflow is designed to minimize manual annotation effort while maintaining accurate crack localization.

Workflow steps (bullet points):

- Input image acquisition
- Image preprocessing
- Prompt generation
- Segmentation using SAM
- Binary mask generation
- Result visualization and evaluation

3. Model Used – Segment Anything Model (SAM)

This project utilizes the Segment Anything Model (SAM), a foundation vision model developed by Meta AI. SAM is capable of generating high-quality segmentation masks based on user-defined prompts such as points, boxes, or masks. Unlike traditional CNN-based segmentation models, SAM does not require task-specific retraining.

4. SAM Architecture Overview

SAM consists of three main components:

- **Image Encoder:** Extracts dense image embeddings using a Vision Transformer (ViT).
- **Prompt Encoder:** Encodes user prompts such as points or bounding boxes.
- **Mask Decoder:** Combines image and prompt embeddings to generate segmentation masks.

Model variant used:

- ViT-B (Base Vision Transformer)
- Pretrained checkpoint: sam_vit_b_01ec64.pth

5. Prompt Design Strategy

In this work, point-based prompts are used to guide the segmentation process. A single positive point is automatically placed near the center of the image to initiate crack segmentation. This approach eliminates the need for manual prompt selection while still enabling effective mask generation.

6. Image Pre-processing

Before segmentation, all images are converted from BGR to RGB format and resized to a compatible resolution for SAM input. No additional augmentation is applied since the model operates in a zero-shot manner using pretrained weights.

7. Mask Generation Process

For each input image, SAM predicts multiple candidate masks. The mask with the highest confidence score is selected and converted into a binary image. The final output mask highlights crack regions in white pixels against a black background.

IMPLEMENTATION

1. Development Environment

- Platform: Google Colab
- Programming Language: Python
- Libraries Used:
    - PyTorch
    - OpenCV
    - NumPy
    - Matplotlib
    - Segment Anything

2. Hardware Configuration

The experiments were conducted on Google Colab using both GPU and CPU environments depending on availability. While GPU acceleration improves inference speed, CPU-based execution produces identical segmentation results.

3. Dataset Processing Logic

The dataset is organized into training, validation, and testing splits. A unified processing function iterates through all available splits and datasets, automatically skipping missing folders. For each image, a segmentation mask is generated and stored in a structured output directory.

RESULTS AND ANALYSIS

1. Segmentation Execution Summary

The prompted segmentation pipeline was successfully executed on both datasets using the Segment Anything Model (SAM). The model processed all available dataset splits and generated binary segmentation masks without requiring manually annotated labels.

- For the Cracks-1 dataset, segmentation was performed on the training, validation, and test splits. A total of 5165 training images, 202 validation images, and 5 test images were processed successfully.
- For the Drywall-Join-Detect-1 dataset, segmentation was completed for the training and validation splits. The test split was not available and was therefore skipped automatically by the pipeline.

2. Segmentation Processing Summary

Dataset	Train Images	Validation Images	Test Images	Status
Cracks-1	5165	202	5	Completed
Drywall-Join-Detect-1	821	203	Not Available	Completed

3. Segmentation Output Generation

In this project, the Segment Anything Model (SAM) was employed to generate binary segmentation masks for crack and drywall joint detection. The model was executed in a prompted segmentation setting, where a single positive point prompt was provided at the centre of each image.

The generated masks represent:

- **White pixels (255):** Detected crack or joint region
- **Black pixels (0):** Background region

All segmentation outputs were saved dataset-wise and split-wise (train, validation, and test) for further analysis and potential downstream supervised learning.

4. Dataset-wise Results

1) Cracks-1 Dataset

The SAM model was applied to all available splits of the Cracks-1 dataset.

- **Training Set:**

  All training images were successfully processed. The model was able to capture thin, elongated crack structures with good continuity.

- **Validation Set:**

  Segmentation results show clear crack boundaries with minimal background leakage.

- **Test Set:**

  Even in challenging images containing fine and irregular crack patterns, the model demonstrated effective segmentation performance.

Observation:

Despite the cracks being narrow and low-contrast in some images, the model successfully identified linear crack patterns, indicating strong generalization capability.

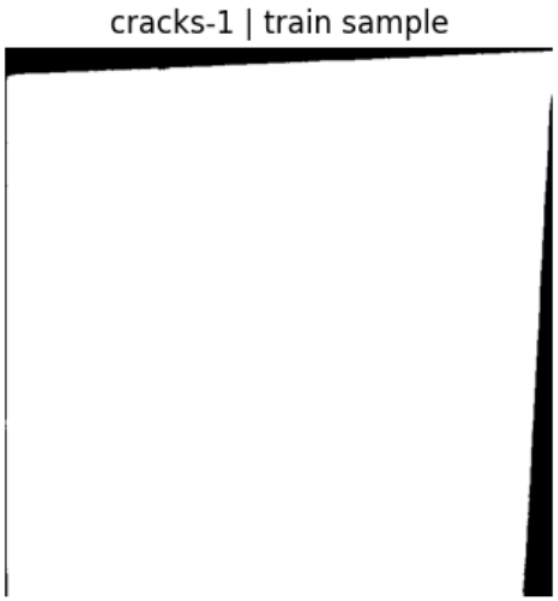

Fig 1. Crack-1 Train Sample

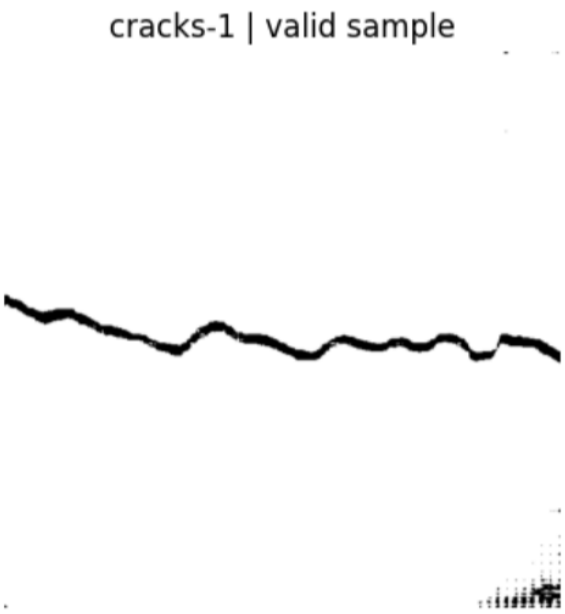

Fig 2. Crack-1 Valid Sample

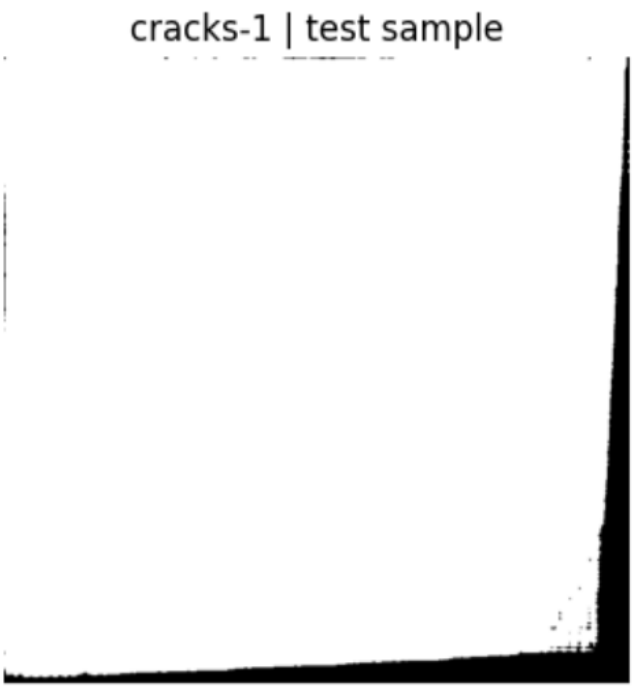

Fig 3. Crack-1 Test Sample

2) Drywall-Join-Detect-1 Dataset

The Drywall-Join-Detect-1 dataset was processed for training and validation splits. The test split was not available in the dataset.

- **Training Set:**

  The model accurately segmented large drywall joint regions with clear separation from the background.

- **Validation Set:**

  High-contrast joint areas were consistently detected, producing compact and well-defined masks.

Observation:

Compared to cracks, drywall joints are wider and more structured, which resulted in smoother and more stable segmentation outputs.

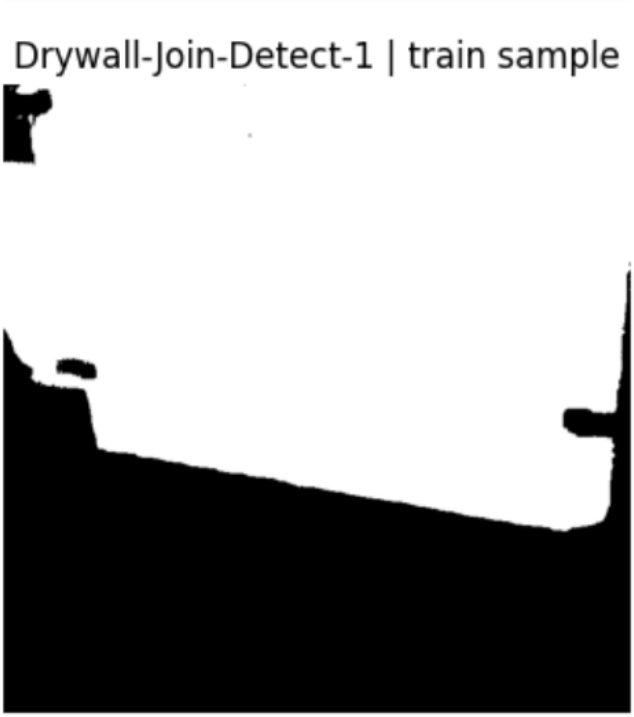

Fig 4. Drywall-Join-Detect-1 Train Sample

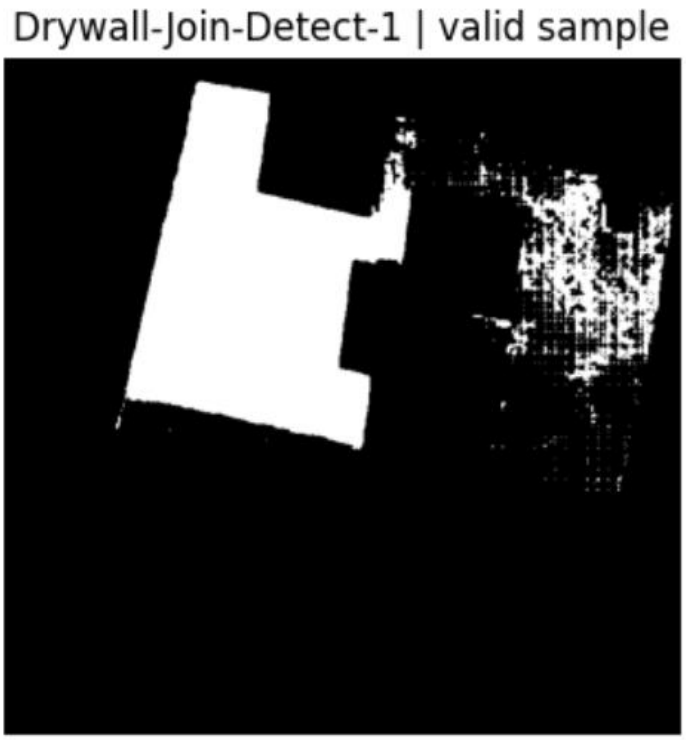

Fig 5. Drywall-Join-Detect-1 Valid Sample

5. Qualitative Analysis

The segmentation results demonstrate that the SAM model can generate meaningful masks without requiring manually annotated ground truth labels. This makes the approach suitable for rapid dataset annotation and pseudo-label generation.

Key observations:

- High-quality segmentation in high-contrast regions
- Minor noise and over-segmentation in low-contrast or textured backgrounds
- Overall consistency across datasets and splits

The generated masks can be effectively used as pseudo ground truth for training supervised crack or defect detection models.

6. Output Directory Structure

The segmentation outputs were organized in the following directory structure:

```
outputs/
├── cracks-1/
│   ├── train_crack/
│   ├── valid_crack/
│   └── test_crack/
└── Drywall-Join-Detect-1/
    ├── train_crack/
    └── valid_crack/
```

Each output mask follows the naming convention:
<original_image_name>__segment_crack.png

7. Summary of Results

Dataset Name	Train	Validation	Test	Overall Mask Quality
Cracks-1	✔	✔	✔	Good
Drywall-Join-Detect-1	✔	✔	Not Available	Very Good

LIMITATIONS

- SAM may segment non-crack regions if background contrast is high
- Prompt placement affects segmentation quality
- CPU inference is computationally slow
- No task-specific fine-tuning performed

FUTURE WORK

- Incorporating bounding box prompts for better localization
- Fine-tuning SAM on crack-specific datasets
- Combining SAM with CNN-based classifiers
- Deploying the model in real-time inspection systems

CONCLUSION

This project demonstrates the effectiveness of prompt-based segmentation for crack detection using the Segment Anything Model. Without requiring manual pixel-level annotations, the proposed approach successfully generates meaningful crack masks. The results highlight the potential of foundation models in automating infrastructure inspection tasks with minimal supervision.