

Generating Large-Scale Segmentation Data from Sparse Annotations using 3D Reconstruction and Gaussian Splatting

Lorenz Krefft¹, Ludwig Hoegner², Cornelius Preidel³

¹ Hochschule München, Departement of Geoinformation, lorenz.krefft@hm.edu
² Hochschule München, Departement of Geoinformation, ludwig.hoegner@hm.edu
³ Hochschule München, Department of Civil Engineering, cornelius.preidel@hm.edu



Methodology

Pixel-wise annotations remain a major bottleneck for training modern semantic segmentation models. This paper introduces a geometry-aware pipeline for amplifying sparse manual annotations into large-scale pixel-wise training datasets using 3D reconstruction and Gaussian splatting. The proposed approach transfers limited 2D annotations into a shared 3D representation, enabling the generation of geometrically consistent novel views with automatically derived labels. The pipeline significantly reduces manual annotation effort while enabling the efficient creation of diverse, consistently labeled

datasets. In an example application, we expand a dataset by a factor of 222 using only 24 manually annotated images. We evaluate the proposed method on crack segmentation, a challenging task characterized by thin and irregular structures. Experimental results on an independent test dataset show that models trained on synthetically generated data achieve performance comparable to models trained on fully manually annotated datasets. Furthermore, combining synthetic and real data improves segmentation performance.

1

Data Collection

Photogrammetric scan of concrete specimens with cracks.

- ~ 15- 30 images per scan
- Camera parameters (intrinsic & extrinsic)
- Sparse point cloud
- Dense point cloud

2

Manual Annotation (Sparse)

Cracks are annotated in a minimal number of views.

The goal is to annotate each crack at least once.

3

Label Transfer to 3D

The dense point cloud is classified based on the Sparse annotations.

The 3D point cloud is projected onto the image plane of the sparse annotations. The point cloud is classified using the information from these annotations.

4

Synthetic Image Generation (Gaussian Splatting)

New camera poses are rendered using Gaussian splatting.

Pose variation:

- Translation: + 5 cm (x,y,z)
- Rotation: + 5° (yaw, pitch,roll)

5

Label Transfer to Synthetic Images

Classified 3D points are reprojected to 2D and converted to masks.

Rendered

Generated Label

Projection of 3D points

Create 2 px circles

Morphological Closing

6

Dataset Generation

Sliding window (448 x 448 px, 10% overlap) is applied. Only tiles wit ≥ 10% crack pixels are kept.

Tile 448 x 448 px

Label

Resulting Datasets

Dataset	Train Images	Val Images
Original Images Dataset (from photogrammetric Scan)	978	245
Synthetically Generated Dataset	13,072	3,269
Crack Segmentation Dataset	9,603	1,695
Crack Segmentation Dataset + Synthetically Generated Dataset	22,675	4,964

7

Model Training

Model	Training Data
Model A	Trained on Original Images Dataset
Model B	Trained on Synthetically Generated Dataset
Model C	Trained on Crack Segmentation Dataset
Model D	Trained on Crack Segmentation Dataset + Synthetically Generated Dataset

Identical network architectures (DeepLabV3+ with Mix Vision Transformer) and training parameters.

8

Evaluation

Crack Dataset

4,029 images (train + val + test)

All models are evaluated on the independent Crack Dataset

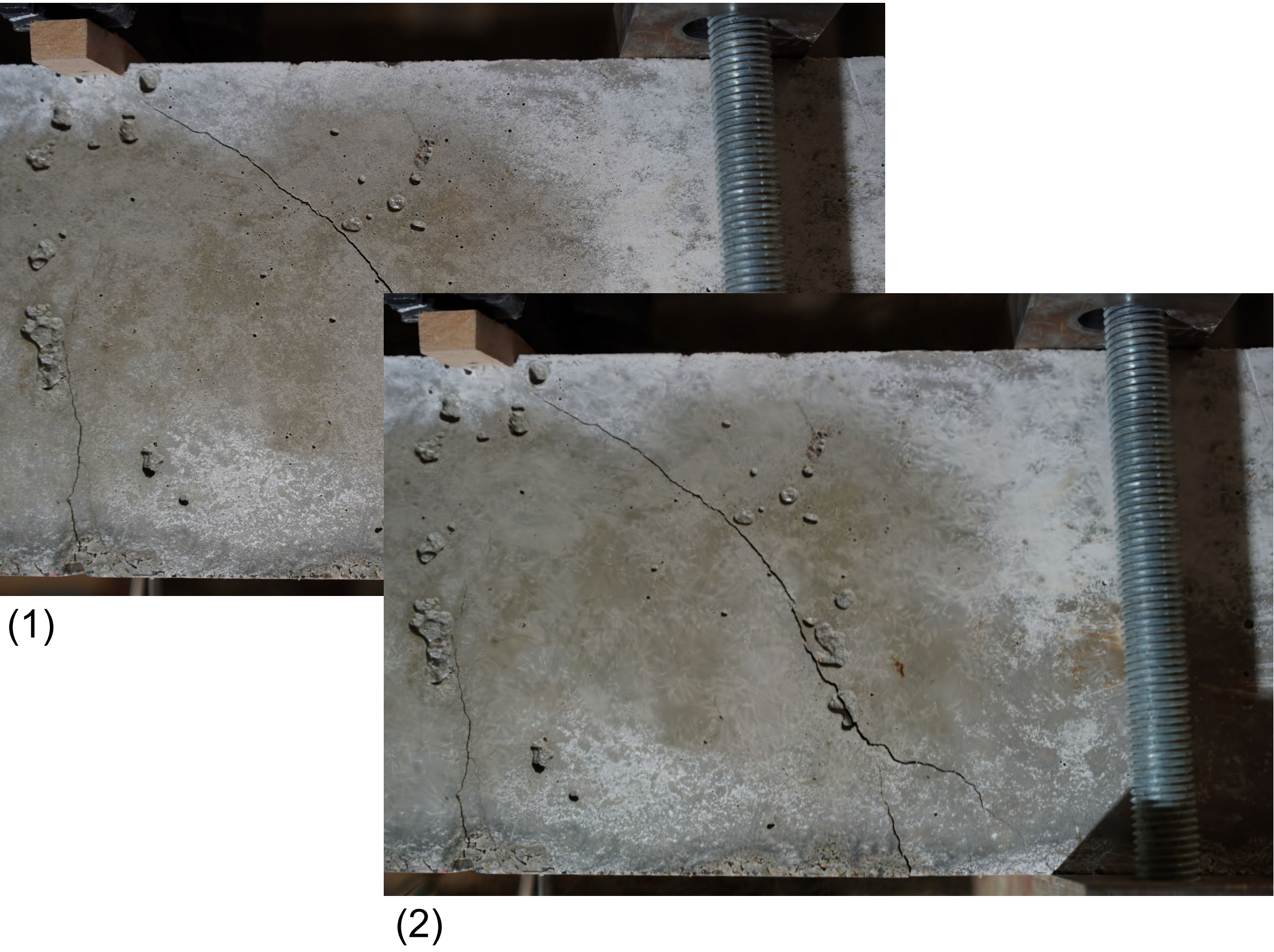


Figure 1 shows an original photogrammetric image of a test specimen with cracks. Figure 2 shows the same representation of the scene using Gaussian splatting.

Testing

Four models' with a DeepLabV3+ decoder and a Mixed Vision Transformer encoder were trained using the same setup and differs only in terms of the training data. Based

on the different training data, the models produce different performance metrics. Evaluation was always performed on a consistent, independent validation dataset.

	Intersection over Union Arithmetic mean	Intersection over Union Standard deviation
Original Images Dataset	0.3801	± 0.1807
Synthetically generated Dataset	0.5317	± 0.1466
Crack Segmentation Dataset	0.5345	± 0.1752
Crack Segmentation Dataset & Synthetically generated Dataset	0.5498	± 0.1551