

Guidelines for Pre-processing Concrete Surface Images for Structural Health Monitoring

(Code Repository, Dataset, and Theoretical Overview)

Mahdi Asgarinejad ^a, Milad Farmahini Farahani ^a, Maryam Bitaraf ^{a 1}, Amirmohammad Ramezaniapour ^a

^a School of Civil Engineering, College of Engineering, University of Tehran, Tehran, Iran

1. Project Overview

Structural Health Monitoring (SHM) of concrete structures has become an essential area of research in recent years. Automated image-based damage detection—particularly using computer vision and deep learning—offers significant potential for early and reliable identification of cracks, bug-holes, and other surface defects.

A crucial stage in such computer vision pipelines is image pre-processing, which directly affects the accuracy, generalization, and efficiency of subsequent detection models. This repository provides a comprehensive framework for image pre-processing of concrete surface images, including a curated dataset, detailed codes, and illustrative sample results.

For theoretical details of the employed methods and algorithms, please refer to the accompanying document “*Theoretical Overview*”.

This repository consists of three main components:

- **Dataset** (raw references, labeled JSON, optimized samples)
- **Codes** (Google Colab notebook + modular Python functions)
- **Image Sample Results** (Some outputs of preprocessing techniques)

2. Dataset

The dataset has been structured to facilitate supervised machine learning (ML) and deep learning (DL) tasks in SHM. It supports flexible labeling formats, enabling users to select between:

1. **Cracks only**
2. **Bug-holes only**
3. **Combined cracks and bug-holes**

The dataset includes the following parts:

- (a) Referenced datasets: Four publicly available concrete surface datasets, sizes ranging from 227×227 to 4032×3024 pixels, comprising more than 50,000 images.
- (b) Optimized image samples: 1,599 images (256×256) processed with Color Multi-Level Thresholding.
- (c) Annotation file: A JSON file in LabelMe format containing polygons, labels, and references for the 1,599 images.

¹ Corresponding author.

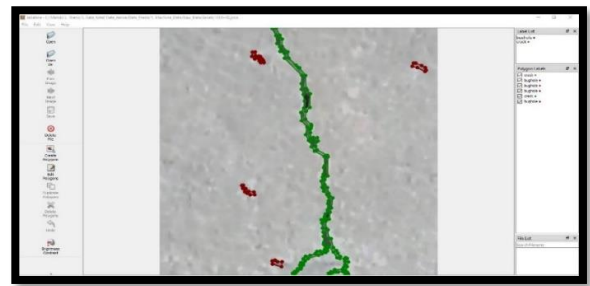
E-mail address: Maryam.bitaraf@ut.ac.ir



(a) Original Image



(b) Optimized Image Sample



(c) Labeling Sample

3. Codes

A step-by-step Google Colab notebook and modular Python scripts are provided for pre-processing. The pipeline includes 10 stages:

1. Environment setup and required libraries installation
2. Import libraries and mount Google Drive
3. Routine pattern recognition methods
4. Image preprocessing and augmentation functions
5. Image denoising methods
6. Dimensionality reduction methods
7. Intra-image fusion methods
8. Color content optimization methods
9. Data reconstruction methods
10. Assessment metrics and evaluation

How to use:

- **Update dependencies via `requirements.txt`**
- **Run Step 1 and Step 2 (environment setup)**
- **Select and execute the desired preprocessing functions**
- **Refer to *Theoretical Overview* for explanations of each function**
- **Step 9 provides evaluation procedures**
- **Some sample of case studies can help in method selection**

4. Image Sample Results

A selection of processed images is provided to illustrate different preprocessing outputs. Techniques demonstrated include:

- Bilateral and classical filtering
- Color Multi-Level Thresholding
- Edge detection (global and local)
- Binary images (global and local thresholding)
- Histogram-enhanced images

- Shadow removal
- Mask generation
- Text-label format export (TXT)

These examples can help researchers identify suitable preprocessing strategies for their own studies.

5. Installation & Usage

Clone this repository and install the required dependencies:

```
git clone https://github.com/YourUsername/Concrete-SHM-Preprocessing.git
cd Concrete-SHM-Preprocessing
pip install -r requirements.txt
```

For Colab users: simply open the provided notebook and follow the step-by-step instructions.

6. Citation

If you use this dataset or code in your research, please cite as follows:

```
@dataset {Asgarinejad2025,
author    = {Mahdi Asgarinejad and Milad F. Farahani and Maryam Bitaraf and A. Ramezaniapour},
title     = {Guidelines for Pre-processing Concrete Surface Images for SHM},
year      = {2025},
publisher = {University of Tehran},
URL       = {https://github.com/...}}
```

A DOI will be added upon Zenodo release.

7. License

This project is released under the License with institutional approval by the University of Tehran. See “LICENSE” for details.

8. Acknowledgments

Thanks to Prof. Elias Ebrahimzadeh, Mr. Mohammad Javad Kamrani, and Miss A'zam Zangoie for their contributions to this dataset and code.

9. References

Please see the *REFERENCES* file for full citations. The base article for this work is:

“Asgarinejad M., Bitaraf M., and Ramezaniapour A.M, Color image segmentation based multilevel thresholding for detection of cracks and bug-holes in concrete surface images using optimization methods, *Advances in Structural Engineering*, Volume ??, Pages ???, DOI: ??.”

10. Contact

For questions or collaborations, please contact:

Mahdi Asgarinejad

Email: M.Asgarinejad76@ut.ac.ir

GitHub:???