Semi-Supervised Learning with Active Module for Semantic Segmentation Annotation for Crack Detection

Spring 2023

CEE 8803: Advanced AI for Smart Cities
Final Report

Instructor: Prof. Yichang (James) Tsai

Group 3
Haolin Wang
Jayda Ritchie
Xiaoyue Zhang

May 6th, 2023

Table of Contents

Executive Summary

This report presents a comprehensive study of a semi-supervised learning approach with an active module for semantic segmentation annotation for crack detection in 3D pavement images. The primary goal is to reduce the amount of human effort required for data annotation while maintaining robust and reliable results comparable to those of manually annotated data. The study comprises a review of existing methods, the development of a novel methodology, and an evaluation of its performance.

A thorough literature review identifies a gap in existing methods for crack detection and segmentation, particularly in terms of utilizing unannotated data to strengthen the training model and reduce annotation effort. To address this gap, a semi-supervised learning algorithm with an active module is proposed, containing two models: a student model for feature extraction and a teacher model for model updates using exponential moving average (EMA) weights.

The methodology section discusses data features, data diversity, annotation methods, potential noises/errors, and data splitting. The dataset consists of 3D pavement images collected by the Georgia Tech Sensing Vehicle (GTSV), offering a diverse range of crack types, patterns, and complexities. Data splitting is carefully designed to explore the optimal ratio of labeled/unlabeled data and the impact of active learning on model performance. Also due to the purpose of this project is to validate the proposed frame for annotation assistance, the data is only splitted as the training and testing datasets without the need to finetune the hyperparameters.

A semi-supervised learning module with an active module is designed to validate the proposed active module's ability to guide annotation order and determine the optimal ratio between labeled and unlabeled data. The module is comprised of a student model and a teacher model, with both models employing a U-Net architecture using a ResNet encoder and a DeepLabV3+ decoder. The active module is incorporated to provide annotation guidance throughout the learning process.

The loss function used for all models, except the fully supervised baseline model, is a combination of supervised loss and unsupervised loss with equal weights. Both supervised and unsupervised losses are composed of cross-entropy loss and DICE loss, addressing the unbiased ratio between class[0] (pavement background) and class[1] (crack). DICE loss tackles data imbalance and measures similarities between two images, while cross-entropy measures the difference between two probability distributions for a variable.

Evaluation results indicate that the best performance is achieved using supervised learning. And the active module does not show significant improvement in the annotation process under insufficient training circumstances, and the active criterion may not be appropriately chosen. It is also found that a higher portion of annotated data leads to better performance, regardless of other factors.

In conclusion, the semi-supervised learning approach with an active module for semantic segmentation annotation for crack detection in 3D pavement images shows promise in reducing human effort for data annotation while maintaining reliable results. However, further experimentation and model tuning are necessary to improve performance, particularly in the areas identified through error analysis.

Recommendations for future work include increasing the training times, exploring other possible criteria for the active module, and finding more annotated datasets to provide a more extensive training process.

1. Introduction

The detection of surface cracks is essential for ensuring the safety and serviceability of civil engineering infrastructure. Cracks often serve as early indicators of degradation, and their detection allows for the implementation of preventative measures. Regular inspections and maintenance of aging facilities are crucial in order to reduce the risk of structural deterioration. Traditional crack detection, however, primarily relies on manual inspection, which is heavily dependent on professionals and can be both inefficient and costly. With the advancement of computer vision and image processing techniques, automatic crack detection methods have gradually replaced conventional manual inspection, providing more efficient and objective results.

To utilize machine learning approaches for automatic crack detection, large and diverse datasets are required to train accurate deep learning models. Data annotation is necessary to identify relevant features that provide meaning to the computer. Within the framework of automatic crack detection, semantic segmentation is one of the most common and promising data annotation techniques. Deep learning models using semantic segmentation have demonstrated excellent performance in crack detection tasks. However, obtaining the best performance from these models typically requires fully supervised segmentation methods with large amounts of annotated data, which can be both time-consuming and labor-intensive.

To address this challenge, in this project, we introduce a semi-supervised semantic segmentation network for crack detection that uses a small number of labeled samples and a large number of unlabeled samples. Our aim is to identify the most informative samples for model training, thereby reducing overhead annotation costs while maintaining sufficient accuracy levels. Furthermore, an active learning module is employed to guide the annotation order and determine the optimal ratio of unlabeled to labeled images.

The proposed method consists of a student model and a teacher model that operate on the same network structure. Pseudo-labels are assigned to the pixels of unlabeled images to enable semi-supervised semantic segmentation.

2. Literature Review

2.1. Review

Li et al. (2020) developed a semi-supervised learning-based detection model for pavement cracks, utilizing unlabeled pavement images for model training. The model comprises a fully convolutional segmentation network and a discriminator network, which can use unlabeled road images during training. Zhang et al. (2020) proposed a network consisting of two Generative Adversarial Networks (GANs) designed to produce more accurate results and prevent overfitting. König et al. (2022) developed a classification CNN to create pseudo-labels of cracks, which can be integrated into a crack image segmentation model's training dataset to enhance performance. Kang et al. (2020) combined three computer vision algorithms and proposed a method for automated crack location, segmentation, and quantification. This hybrid approach compensates for the limitations of traditional segmentation algorithms and achieves high accuracy. Li et al. (2020) and Shim et al. (2020) proposed a semi-supervised crack segmentation method based on adversarial learning. The model consists of a segmentation network and a discriminator network, with the segmentation network outputting a prediction map given the input crack image, and the discriminator network differentiating the ground truth label map from the prediction map. For unannotated data, a self-taught strategy is employed using the trained discriminator.

2.2. Needs Improvement

While it is evident from the research that semantic segmentation has been widely used in crack detection and can serve as the foundation for future studies measuring crack morphological features, there are still areas for improvement. For instance, existing semi-supervised learning research (Li et al., 2020) does not convincingly identify which images or features are most informative for model training. Identifying these specific features would allow further reduction in the number of annotated images required for training. Obtaining pixel-level data annotation for semantic segmentation is time-consuming, and reducing overhead costs would make such models more scalable for practical use.

The application of AI for automating the data annotation process is highly beneficial. AI algorithms can process data much faster than humans, significantly decreasing annotation time. Moreover, once properly trained, algorithms can identify patterns with greater accuracy than humans. Additionally, as the amount of available data continues to increase, the manual labeling process becomes increasingly costly and time-consuming. Therefore, the case for using AI in this context is compelling.

3. Proposed Methodology

3.1. Data

3.1.1. Data description

In recent years, the advancement of sensor technologies has led to 3D laser technology becoming the most widely used method for acquiring high-resolution surface data. 3D lasers are less sensitive to lighting conditions (making it easier to remove noise from surface stains) and allow cracks to be more distinguishable (Tsai and Li, 2012). The dataset for our proposed machine learning model was composed of 3D pavement images collected by Hsieh & Tsai (2020). This data was collected by the Georgia Tech Sensing Vehicle (GTSV), where two 3D line laser sensors were mounted onto the vehicle to collect 1000 (longitudinal) x 2080 (transverse) points. The point clouds collected were then transformed into 3D range images. The size of the images is 520 × 1,250 pixels with high resolution (Fig.2). Manually annotated crack images were denoted as ground truth data that was then used for validation for the proposed model. Each image within the dataset was finely annotated and was paired with an unlabeled version of the image. The corresponding binary segmentation image was composed of pixel RGB value of 0 (black) for the background and 255 (white) to identify crack.

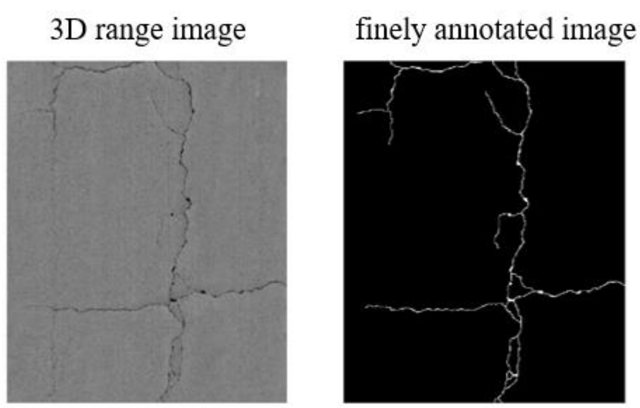

Fig.1 3D crack image (Hsieh & Tsai, 2020)

3.1.2. Data Diversity

Diversity of the dataset is one of the most important factors that can impact the performance of machine learning models. Providing high-quality images with various obstructions, such as shadows, surface roughness, lack of illumination, and the presence of other objects, such as patches and white lane markings in the training set, is a crucial step to producing an accurate and realistic model. The width of the crack should also be considered since it is difficult to distinguish the difference between cracks and background noises as it decreases, as noted in Fig.

3. Besides that, crack type and the complexity of the crack pattern are also critical to the overall accuracy and processing time of the machine learning model.

Fig.2 Cracked Forest Dataset (Shi et al., 2016)

3.1.3. Annotation method and potential noises/errors

Manual annotation for segmented data is liable to noises/errors. Possible potential errors that can arise from human annotation include the following. Firstly, the annotator may over-identify or under-identify the amount of crack pixels present, leading to an overall decrease in accuracy. Furthermore, the dataset may be subject to inconsistent labeling in the event that different annotators label the same object differently. In order to mitigate these issues, implementing quality control measures, such as having multiple annotators review the same data and adopting a consensus-based approach, can help maintain a high level of accuracy and consistency.

3.1.4. Data splitting

Considering that the purpose of this study is to find the most informative samples and determine the overall best ratio of labeled/unlabeled data, the images in the dataset are split into multiple labeled/unlabeled sets for semi-supervised learning. These sets are trained using two frameworks presented in this experiment: active semi-supervised learning and semi-supervised learning.

- Active Semi-Supervised Learning
    - 0/10 - Unsupervised learning with K-means clustering
    - 3/10 - Active Semi-Supervised learning
    - 5/10 - Active Semi-Supervised learning
    - 7/10 - Active Semi-Supervised learning
    - 10/10 - Used as a baseline with supervised learning to test our model
- Semi-Supervised Learning
    - 3/10 - Random Semi-Supervised Learning
    - 5/10 - Random Semi-Supervised learning
    - 7/10 - Random Semi-Supervised learning
    - 10/10 - Used as a baseline with supervised learning to test our model

3.2. AI Methods / Models

3.2.1. Review of Models

The following models and methods will be used for implementing the proposed AI task. For the semi-supervised learning aspect of the AI task, a GAN (Generative Adversarial Network) will be used in the data augmentation process to increase the diversity of the data and assist in preventing overfitting. GAN-based learning has been used in several papers for semi-supervised learning to encourage confusion between labeled and unlabeled examples. With the GAN combined generator and discriminator, the overall model performance can be improved. For semantic segmentation, cross-entropy loss/negative log-likelihood will be used as the loss function for determining the difference between the predicted probability distribution and the true distribution (Kniazieva, 2022).

The models for semi-supervised semantic segmentation that will be explored in this study are fully convolutional networks (FCN) and a standard convolutional neural network (CNN). A fully convolutional network is a type of neural network that provides a segmented image of the original image. FCN takes an image as input and produces a corresponding output image where each pixel in the output image will represent the class of the object that the corresponding pixel in the input image belongs to. FCN architecture is widely used because it allows end-to-end training for a convolutional neural network, preserving spatial information by replacing fully connected layers in traditional CNNs with convolutional layers. A standard convolutional neural network is another model that will be explored for implementing the AI task and will be compared against the FCN. A CNN is able to learn features in images through convolutional layers; the input image is fed into a series of convolutional layers followed by pooling layers to learn high-level features. CNNs are widely used for image segmentation tasks because they can identify and classify different objects or regions based on their texture, color, shape, and other visual features.

3.2.2. AI task definition and AI problem formulation

In this project, we focus on semi-supervised learning, which uses unannotated data to strengthen the training model and reduce annotation effort. We introduce a semi-supervised learning algorithm for crack image segmentation, comprising two models: a student model and a teacher model. The student model extracts image features at multiple scales and leverages contextual information, while the teacher model has the same structure as the student model but uses exponential moving average weights for updates. The student and teacher models work in tandem to improve model performance iteratively.

The main objectives of our project are as follows:

- Identify images requiring more annotation and find the most informative samples for model training.

- Increase the quantity and diversity of finely annotated segmentation data for algorithm use.
- Reduce human effort required in the data annotation process through automation.
  - Decrease overall time and cost of classifying data for machine learning.
  - Minimize subjectivity and human error.
- Ensure that the annotated data remains as robust and reliable as manually annotated data while achieving the objectives above.

3.2.3. Model selection, training, and validation

To validate the proposed active module's capability to guide the annotation order and to determine the best ratio between labeled and unlabeled data to best utilize all the labeled and unlabeled data, the project will design the semi-supervised learning module experiment in the following parts:

- Random Annotation

  Given an unlabeled crack range image dataset, randomly choose:

  - 30% to annotate, then train the data on semi-supervised model;
  - Another 20% to annotate to get a 1:1 dataset, then train the data on semi-supervised model;
  - Again, choose 30% to annotate and train on the model;
  - Finally, annotate all the remaining unlabeled data and train it on a fully supervised learning model.
- Active Annotation

  Given an unlabeled crack range image dataset with the proposed active module, actively choose:

  - 30% to annotate and train on the semi-supervised model;
  - 50% to annotate and train on the semi-supervised model;
  - 70% to annotate and train on the semi-supervised model;
  - Finally, actively annotate all the remaining 30% unlabeled data and train it on a fully supervised learning model.
- Result Comparison and Evaluation

  Compare all six semi-supervised learning results and one supervised learning result to validate the robustness of the proposed framework and to identify possible improvements.

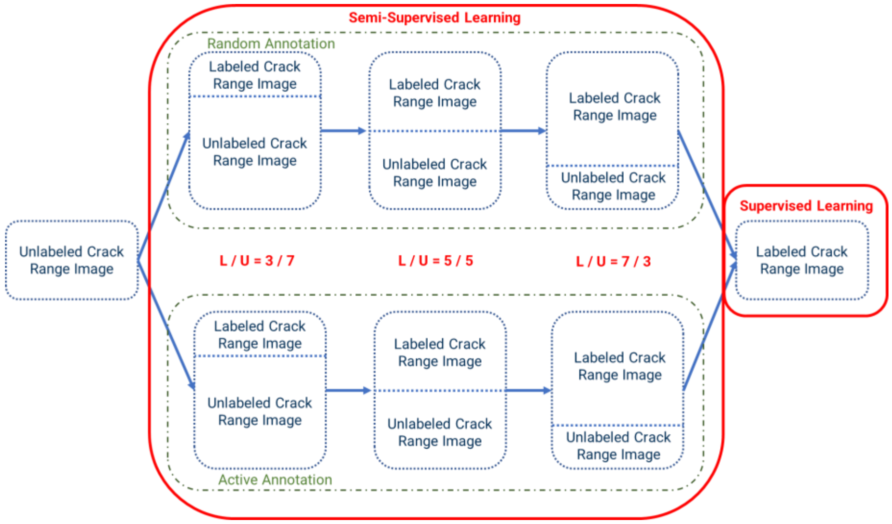

Fig.3 Experiment Design

To reduce the total number of manual annotations required to train a model while utilizing all collected data, semi-supervised learning is employed. This approach uses both labeled and unlabeled data for learning and performance improvement. The experiment incorporates an active learning module to provide annotation guidance, with the goal of determining if implementing active learning impacts the optimal ratio of labeled/unlabeled data.

The semi-supervised module consists of a student model and a teacher model. The student model is trained on both labeled and unlabeled data, while the teacher model is trained only on unlabeled data, providing pseudo-labels to help train the student model. The teacher model is updated using Exponential Moving Average (EMA) by the trained student model. Both the student and teacher models use a U-Net model (Convolutional Neural Network) with a ResNet encoder and a DeepLabV3+ decoder.

The active module has two components: Active Module 1 provides guidance for annotating fully unlabeled data. It utilizes the teacher model from the semi-supervised module to offer unsupervised prediction features on images, then employs K-Means to cluster the unlabeled data and generate the initial annotation queue. Active Module 2 iteratively provides annotation guidance during the semi-supervised process, using the trained teacher model to produce the entropy order of the remaining unlabeled data. This generates the annotation queue according to experiment requirements.

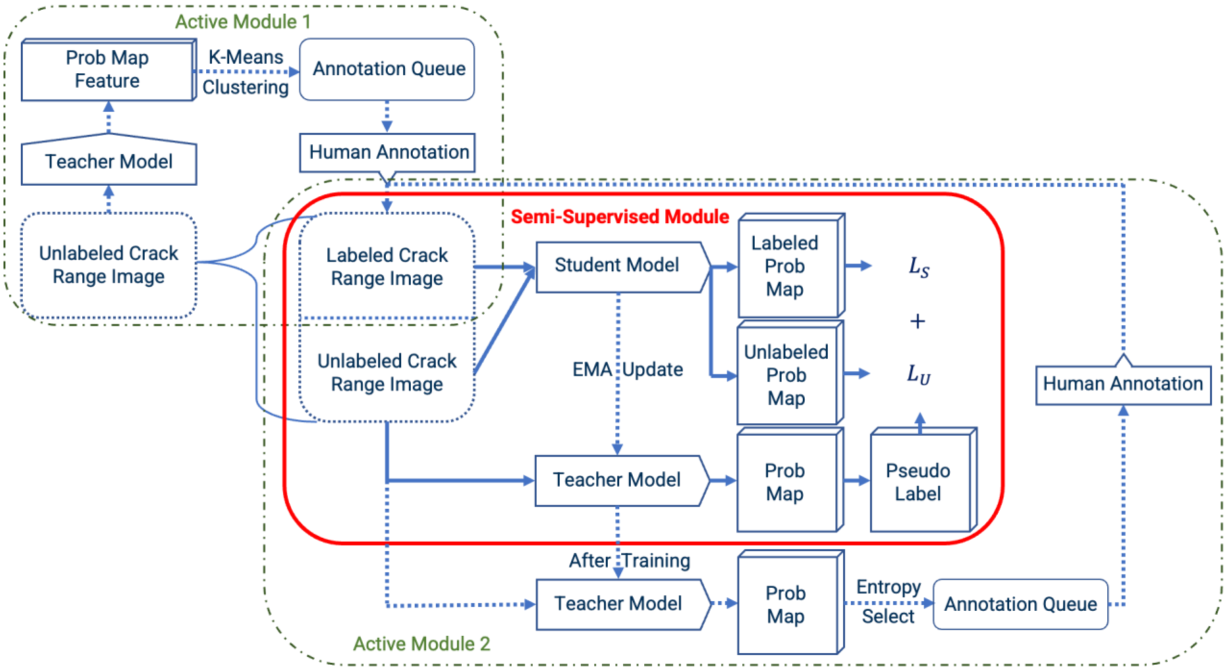

Fig.4 Model Frame Design

3.2.4. The loss function(s) to train the model

For the proposed tasks all models use the loss function that combines loss from supervised and unsupervised learning with the same weight (noted in Eqn 1) except the fully supervised which is considered as the baseline model. Supervised Loss and Unsupervised Loss is a combination of Cross-Entropy Loss and DICE Loss, considering the unbiased ratio between class of pavement background and class of crack.

Cross-entropy loss (or logistic loss) is chosen because it is often one of the most common loss functions for semantic segmentation. It measures the mutual entropy between two probability distributions p and q. It can be used in classification to measure the difference between the ground truth class distribution and the distribution of the predicted classes. Cross entropy loss uses standard supervised pixel-wise cross-entropy loss term evaluated only for the labeled samples shown in Eqn2 and Eqn3.

The Dice coefficient is widely adopted to calculate the similarity between two images in computer vision domain. The indices of dice coefficient is to mesure the overlap between the two sets. It considers the loss information both locally and globally, which is critical for high accuracy. The advantage of using dice loss is that it can very well handle the class imbalance in terms of pixel count for foreground and background.The mathematical equation is defined in Eqn 4.

Mathematical Formula of Loss Function

$$L = \lambda_u L_u + \lambda_s L_s \qquad (Eqn\ 1)$$

L_s : supervised learning loss
L_u : unsupervised learning loss
λ_s: supervised learning loss
λ_u: weight of unsupervised learning loss

The mathematical formula describing the cross entropy loss function in classification for supervised and unsupervised loss is noted below:

$$L_u = -\sum_{c=1}^{M} y_u * log\ p\ (y\ =\ c|x) \qquad (Eqn\ 2)$$

$$L_s = -\sum_{c=1}^{M} y_s * log\ p\ (y\ =\ c|x) \qquad (Eqn\ 3)$$

M: number of classes c
y_u: binary indicator if the class label is c (for unsupervised learning)
y_s: binary indicator if the class label is c (for supervised learning)
$p(y\ =\ c|x)$: probability of the label being c given the input feature vector x.

The mathematical formula describing the dice loss function is noted below:

$$L_D = 1 - \frac{2\sum_{i=1}^{n} p_i y_i}{\sum_{i=1}^{n} {p_i}^2 + \sum_{i=1}^{n} {y_i}^2} \qquad (Eqn\ 4)$$

y_i : real pixel value
p_i : predicted pixel value

4. Evaluation and Discussion

After conducting four refinements in feature extraction and a modification of the model, a total of five testing results are generated, and their performances are evaluated.

4.1. Performance measurement metrics

To understand how the model performs under different annotation methods and data ratios, several metrics are used since the project is a semantic segmentation task. Testing is conducted to calculate the result's mean Intersection over Union (mIoU) and the IoU of class[1] (crack). In addition to these metrics, the predicted results and ground truth are also compared using the following three metrics: True Positive (TP), False Positive (FP), and False Negative (FN). Based on these metrics, precision, recall, and F1-Score are calculated. True Negatives (TN) are not considered in this project since the number of pavement background pixels is much greater than the crack pixels.

- True Positive (TP): The actual pixel is a crack, and the predicted pixel is a crack.
- False Positive (FP): The actual pixel is non-crack, but the predicted pixel is a crack.
- False Negative (FN): The actual pixel is a crack, but the predicted pixel is non-crack.

$$Precision = \frac{TP}{TP + FP} \qquad (Eqn\ 5)$$

$$Recall = \frac{TP}{TP + FN} \qquad (Eqn\ 6)$$

$$F1\ Score = \frac{2Precision * Recall}{Precision + Recall} \qquad (Eqn\ 7)$$

By calculating and comparing the mIoU, IoU, precision, recall, and F1-Score for each test, the impact of different annotation methods and data ratios on the performance of the semi-supervised learning algorithm for crack image segmentation can be assessed. This evaluation will provide valuable insights into the effectiveness of the approach and help identify areas for further improvement or refinement.

4.2. Evaluation results

The table below presents the evaluation results of each experiment, revealing that the active module did not significantly influence the outcomes, while a higher amount of annotated data led to better performance. Possible reasons for cases of false negatives (FN) or false positives (FP) might include the model not being sufficiently trained, with only 5-7 epochs per experiment, or not being trained on enough images due to poorly chosen splits.

Different training methods were used in the experiments, including supervised, semi-supervised, and active semi-supervised with varying degrees of data annotation and training iterations. The results show that the supervised method had the highest mIoU (64.60) and Class[1] IoU (30.20) scores, while the active semi-supervised and semi-supervised methods had lower scores.

Overall, the evaluation results indicate that increasing the amount of annotated data can improve the model's performance, but the active module does not have a significant effect on the training results. However, further experimentation and tuning of the model is necessary to achieve better results.

	Semi-Sup (0.3)	Active-Semi (0.3)	Semi-Sup-Equal Train (0.5)	Semi-Sup-Less Train (0.5)	Active-Semi (0.5)	Semi-Sup (0.7)	Supervised (1.0)
mIoU	63.11	61.47	62.63	62.07	62.32	62.93	64.60
Class[1] IoU	27.40	23.91	26.38	25.23	25.93	27.07	30.20
Precision	0.43	0.49	0.44	0.44	0.44	0.43	0.51
Recall	0.50	0.34	0.46	0.42	0.51	0.51	0.49
F1-Score	0.46	0.41	0.45	0.43	0.47	0.46	0.50
TP	0.30	0.25	0.29	0.28	0.31	0.30	0.33
FP	0.40	0.26	0.37	0.34	0.39	0.40	0.32
FN	0.30	0.48	0.34	0.38	0.30	0.29	0.35

Table 1: Evaluation of Each Experiment.

4.3. Error analysis

In this semantic segmentation task, error analysis is based on the comparison between the original pavement range images and their corresponding predicted segmentation images, masked with True Positive, False Positive, and False Negative. The analysis should focus on the most representative examples in the results.

From the first example (Fig.5), it is clear that the model has difficulty detecting trivial cracks (micro-cracks). This may be due to insufficient training steps or a lack of diverse data on trivial cracks.

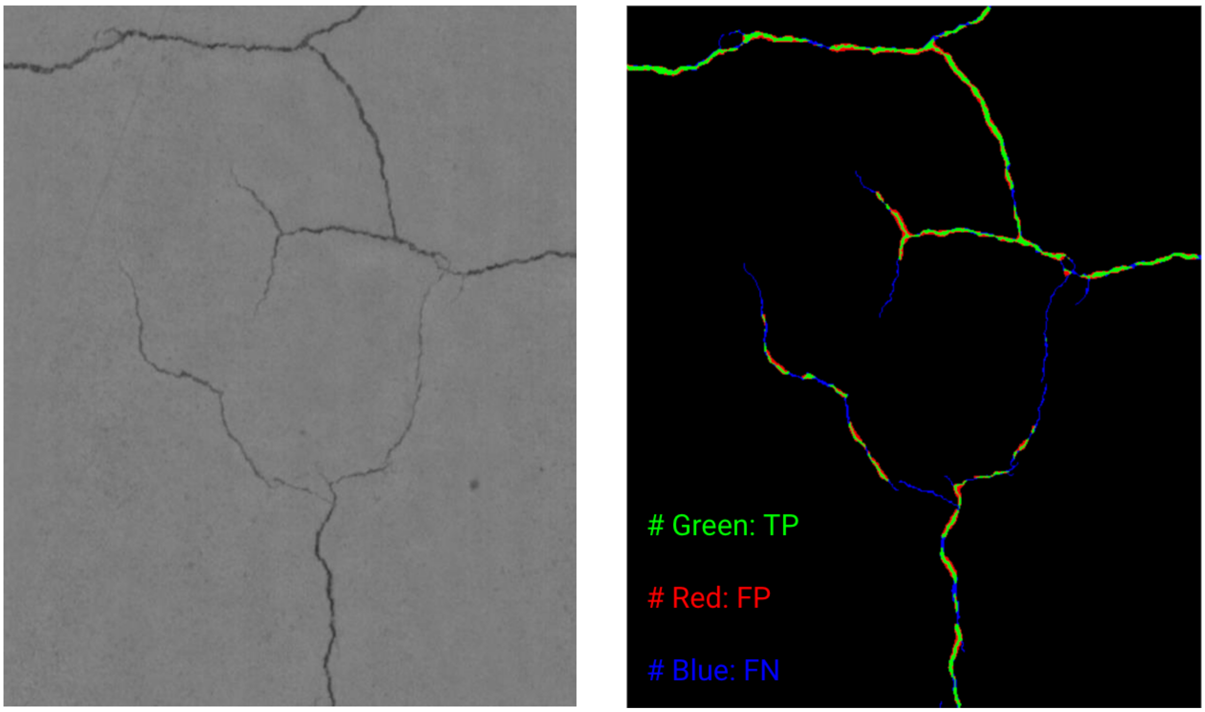

Fig.5 Error Analysis 1

The second example (Fig.6) indicates that the model is not adequately trained to recognize blocking and transverse cracks. It also reveals that the model has limited capability to identify trivial cracks.

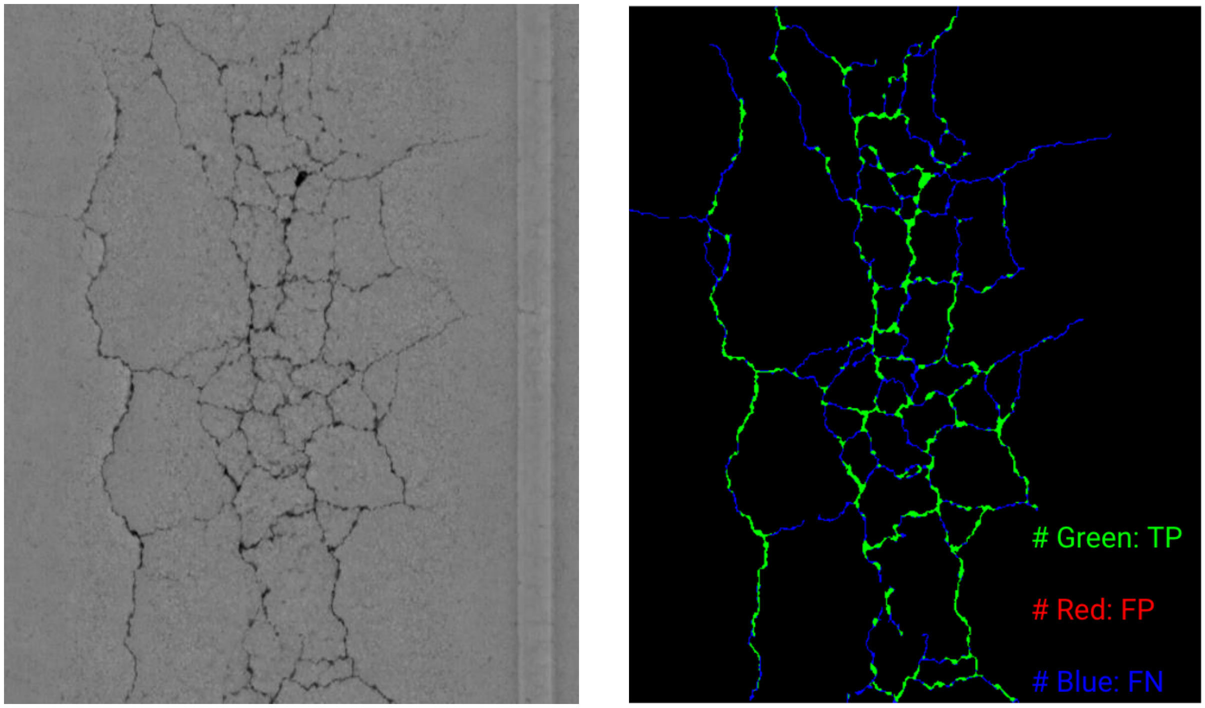

Fig.6 Error Analysis 2

In the third example (Fig.7), the model appears to be more focused on the darker parts of images and not sufficiently trained to detect other types of distress. This issue may arise from inappropriate weights between the two classes in the loss function, causing the model to lean too heavily towards darker pixels.

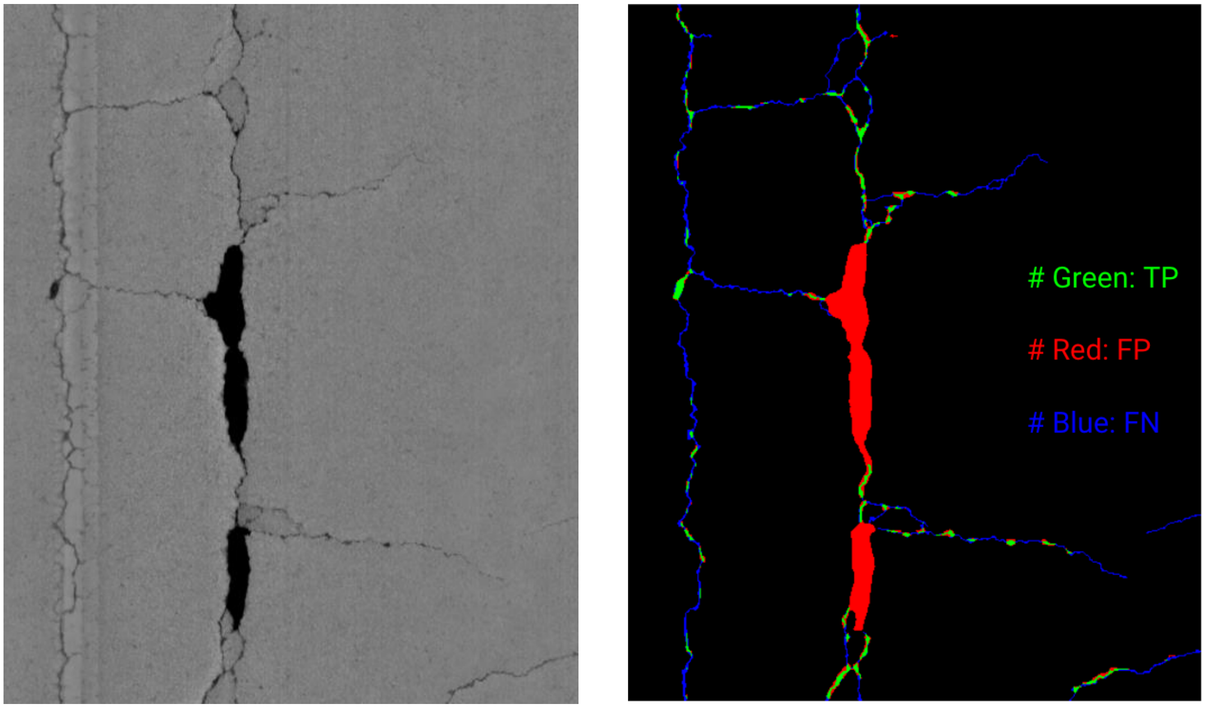

Fig.7 Error Analysis 3

5. Conclusions and Recommendations

Our project has successfully implemented semi-supervised learning combined with an active module on crack detection tasks. Semi-supervised learning reduces the time required for image annotation and active learning module provides a guideline to choose the most informative data to obtain optimal ratio between labeled images and unlabeled image for training. From the experimental results above, the following conclusions can be drawn: 1). The split (trained : untrained) that yielded the best result used 1: 0 split (1132 images) with the active semi approach. 2). Under the insufficient training circumstances, the active module has no significant effect on improving the annotation process, or the active criterion is not appropriately chose. 3). A higher portion of annotated data can lead to a better performances, regardless other factors.

Due to the extreme lack of labels, the semi-supervised learning frameworks commonly need to pay a price in time for higher accuracy. This study can be further extended on the aspects of optimizing the training process to reduce time usage while maintaining high accuracy. Another possible research direction could be to identify how different criteria of active modules can impact the training process.

References

[1] Ali, R., Chuah, J. H., Talip, M. S. A., Mokhtar, N., & Shoaib, M. A. (2021). Automatic pixel-level crack segmentation in images using fully convolutional neural network based on residual blocks and pixel local weights. *Engineering Applications of Artificial Intelligence*, *104*, 104391. https://doi.org/10.1016/j.engappai.2021.104391

[2] Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(12), 2481–2495. https://doi.org/10.1109/TPAMI.2016.2644615

[3] Chen, Q., Huang, Y., Sun, H., & Huang, W. (2021). Pavement crack detection using hessian structure propagation. *Advanced Engineering Informatics*, *49*, 101303. https://doi.org/10.1016/j.aei.2021.101303

[4] Hsieh, Y.-A., & Tsai, Y. J. (2020). Machine Learning for Crack Detection: Review and Model Performance Comparison. *Journal of Computing in Civil Engineering*, *34*(5), 04020038. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000918

[5] König, J., Jenkins, M., Mannion, M., Barrie, P., & Morison, G. (2022). *What's Cracking? A Review and Analysis of Deep Learning Methods for Structural Crack Segmentation, Detection and Quantification* (arXiv:2202.03714). arXiv. http://arxiv.org/abs/2202.03714

[6] Li, H., Wang, W., Wang, M., Li, L., & Vimlund, V. (2022). A review of deep learning methods for pixel-level crack detection. *Journal of Traffic and Transportation Engineering (English Edition)*, *9*(6), 945–968. https://doi.org/10.1016/j.jtte.2022.11.003

[7] Monarch, R., & Manning, C. D. (n.d.). *Human-in-the-Loop Machine Learning: Active Learning and Annotation for Human-Centered AI*.

[8] Sander, J., de Vos, B. D., & Išgum, I. (2020). Automatic segmentation with detection of local segmentation failures in cardiac MRI. *Scientific Reports*, *10*(1), 21769. https://doi.org/10.1038/s41598-020-77733-4

[9] Shi, Y., Cui, L., Qi, Z., Meng, F., & Chen, Z. (2016). Automatic Road Crack Detection Using Random Structured Forests. *IEEE Transactions on Intelligent Transportation Systems*, *17*(12), 3434–3445. https://doi.org/10.1109/TITS.2016.2552248

[10] Tsai, Y.-C. (James), & Chatterjee, A. (2017). Comprehensive, Quantitative Crack Detection Algorithm Performance Evaluation System. *Journal of Computing in Civil Engineering*, *31*(5), 04017047. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000696

[11] Zhang, L., Yang, F., Daniel Zhang, Y., & Zhu, Y. J. (2016). Road crack detection using deep convolutional neural network. 2016 IEEE International Conference on Image Processing (ICIP), 3708–3712. https://doi.org/10.1109/ICIP.2016.7533052

[12] Kniazieva, Y. (2022, November 3). Automated Data Annotation. https://labelyourdata.com/articles/automated-data-annotation-process

[13] Chen S., Yang Y., Hua Y. (2023). Semi-Supervised Active Learning for Object Detection. Electronics, 12(2):375. https://doi.org/10.3390/electronics12020375

[14] Hongxia Li, Weixing Wang, Mengfei Wang, Limin Li, Vivian Vimlund,A review of deep learning methods for pixel-level crack detection, Journal of Traffic and Transportation Engineering (English Edition),Volume 9, Issue 6, 2022, Pages 945-968, ISSN 2095-7564, https://doi.org/10.1016/j.jtte.2022.11.003

[15] Kang, D., Benipal, S.S., Gopal, D.L. and Cha, Y.J., 2020. Hybrid pixel-level concrete crack segmentation and quantification across complex backgrounds using deep learning.

Automation in Construction, 118, p.103291.
[16] König, J., Jenkins, M., Mannion, M., Barrie, P. and Morison, G., 2022. What's Cracking? A Review and Analysis of Deep Learning Methods for Structural Crack Segmentation, Detection and Quantification. arXiv preprint arXiv:2202.03714.
[17] Li, G., Wan, J., He, S., Liu, Q. and Ma, B., 2020. Semi-supervised semantic segmentation using adversarial learning for pavement crack detection. IEEE Access, 8, pp.51446-51459.
[18] Shim, S., Kim, J., Cho, G.C. and Lee, S.W., 2020. Multiscale and adversarial learning-based semi-supervised semantic segmentation approach for crack detection in concrete structures. IEEE Access, 8, pp.170939-170950.
[19] Zhang, K., Zhang, Y. and Cheng, H.D., 2020. Self-supervised structure learning for crack detection based on cycle-consistent generative adversarial networks. Journal of Computing in Civil Engineering, 34(3), p.04020004.