Investigation of possible improvements for pseudo-labelling for object detection datasets

Shon s232192

April 11th 2024

1 Introduction

The development of robust object detection systems is pivotal in advancing computational models applicable across a spectrum of industries, including security, automotive navigation, and healthcare diagnostics. Such systems rely heavily on the quality and comprehensiveness of the training datasets used. Conventionally, the preparation of these datasets involves extensive manual annotation, a process that is not only time-consuming but also susceptible to inaccuracies inherent in human judgment. Pseudo-labelling, a semi-supervised learning approach that utilizes outputs from pre-trained deep learning models to generate labels, presents a compelling alternative. This technique promises to accelerate the dataset annotation process while potentially maintaining, or even improving, the accuracy of the data provided to object detection systems. This report explores enhancements to pseudo-labelling methods aimed at boosting their effectiveness in object detection. The research focuses on three primary objectives: 1. Dataset Creation: Streamlining the development of event-based video datasets specifically designed for object detection, providing a robust basis for model training and evaluation. 2. Performance Evaluation: Evaluating the accuracy and reliability of pseudo-labelling by comparing it to manual annotations, with an emphasis on practical applications. 3. Innovative Improvements: Proposing and testing novel methods to refine the pseudo-labelling process, enhancing its precision and efficiency for broader dataset usage. Through targeted manual annotation and the utilization of advanced detection models like YOLOv5 for initial label generation, this study aims to strengthen pseudo-labelling techniques and their application in the field of computer vision.

2 Materials and Methods

This section provides a detailed description of the methodologies and tools utilized in our study to enhance pseudo-labelling techniques for object detection datasets. Our research approach combines advanced automated labeling technologies with traditional manual annotation methods to optimize both accuracy and efficiency.

2.1 Overview of Tools and Techniques

We employed several key technologies, each chosen for their strengths in handling specific aspects of data annotation and tracking in object detection.

- Pseudo-labelling: This semi-supervised learning technique leverages a model trained on a limited dataset to generate labels for a larger, unlabeled dataset. The use of pseudo-labelling facilitates the expansion of training data significantly without the proportional increase in manual labeling effort. The effectiveness of this approach hinges on the initial model's accuracy, as errors can perpetuate and amplify throughout training iterations. Ongoing refinement of these labels through subsequent iterations helps in enhancing the overall model's performance and its ability to generalize to new, unseen data.

- YOLOv5: As a cutting-edge object detection system, YOLOv5 offers several advantages for automated labeling, including high-speed processing capabilities and scalability. It features an advanced neural network architecture that incorporates elements like Cross-Stage Partial networks (CSPNet), Spatial Pyramid Pooling (SPP), and Path Aggregation Network (PAN) to optimize both speed and accuracy. These elements allow YOLOv5 to deliver high-quality object detection even in challenging conditions, making it invaluable for initial label generation and iterative model improvements.

- LabelMe: This web-based annotation tool is pivotal for manual labeling, providing a flexible and intuitive interface for delineating object boundaries with high precision. LabelMe supports detailed polygon annotations, allowing users to outline objects with complex shapes effectively. This precision is crucial for creating accurate ground truth datasets necessary for validating and training machine learning models.

- OpenCV Trackers: We utilized OpenCV trackers for their robustness in maintaining object identity across video frames. This capability is essential in continuous monitoring applications where objects must be tracked through various movements and occlusions. OpenCV offers several tracker options, such as CSRT and KCF, which are optimized for different tracking environments and offer a balance between speed and tracking accuracy.

2.2 Methodological Application

The methodological approach was designed to capitalize on the strengths of each tool across different data settings, ensuring both robustness and accuracy in annotations:

- Controlled Environments: In settings such as laboratories and clear daylight conditions, YOLOv5 was primarily utilized for its speed and accuracy to automatically generate initial labels across a broad dataset. This was particularly effective due to the stable and predictable environmental conditions, which are ideal for automated detection. The initial labels generated by YOLOv5 were subsequently reviewed and refined using LabelMe, where manual annotators corrected any discrepancies and added fine details to ensure the highest degree of accuracy.

- Field Data: The variable and often unpredictable conditions encountered in field settings necessitated a flexible approach to labeling. Here, YOLOv5 was again used to provide a first pass of labels, identifying objects that were clearly visible. Due to environmental factors such as changing lighting and potential obstructions, each automated label was critically reviewed using LabelMe. This allowed annotators to make necessary adjustments based on their expert judgment and real-world nuances that automated systems might overlook.

- Challenging Conditions: Nighttime and other low-light environments posed the greatest challenge due to reduced visibility and the higher likelihood of detection errors. In these scenarios, the reliability on automated tools like YOLOv5 was reduced, and greater emphasis was placed on manual labeling with LabelMe. OpenCV trackers were employed strategically to maintain continuity of object identification where possible. The trackers helped in mapping the movement and transformation of objects across frames, which was crucial for consistent tracking and labeling in video sequences.

- Integration of Techniques: Across all environments, the integration of automated and manual labeling processes ensured that our datasets were not only large and diverse but also accurately annotated. This dual approach leveraged the speed and breadth of coverage provided by YOLOv5 with the precision and adaptability of manual labeling by human annotators using LabelMe.

3 Results

This section presents the findings from the extensive annotation of thousands of images using both LabelMe for manual annotations and YOLOv5 for automated pseudo-labeling. The results highlight critical insights into the comparative accuracy and efficiency of these two approaches, particularly in identifying structural defects such as cracks and spallings. Detailed discussions will follow on the precision of manual labeling versus the automated methods, emphasizing discrepancies observed in the annotations and their implications for the use of pseudo-labeling in object detection datasets.
Three photographs (Figure.1) were selected to illustrate a notable phenomenon: the inability of YOLOv5 to consistently detect the entirety of crack formations. Each photograph presents a crack that manifests as a darkened area at the initiation point, which gradually transitions to a lighter shade as the crack progresses. This gradation in visual contrast appears to impede YOLOv5's detection algorithm.

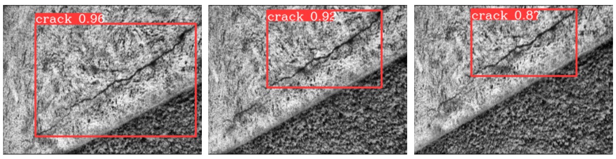

Figure 1: Generated by YOLOv5

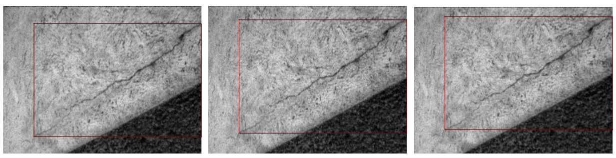

Figure 2: Generated by LabelMe

Manual annotations with LabelMe revealed a high degree of precision in outlining these cracks in their full extent, adhering closely to their nuanced visual progression. Conversely, YOLOv5's annotations were observed to be incomplete, often either terminating prematurely or failing to initiate at the correct point of the crack's origin. This suggests that the gradient change in color and contrast within the crack's structure presents a challenge for the automated detection system, which relies on defined patterns of contrast and color consistency for object identification.

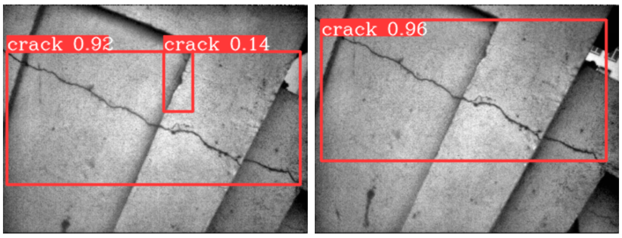

Figure 3: Generated by YOLOv5

Further analysis of YOLOv5's output elucidates an inconsistency in crack detection across contiguous frames. In an illustrative pair of consecutive images processed by YOLOv5, an initial frame displays two distinct cracks identified by the model. However, a subsequent frame from the same sequence reveals that only one of these cracks is detected. This discrepancy not only underscores the variability in YOLOv5's recognition capabilities but also accentuates the necessity for manual review and correction. Such instances exemplify the challenges automated systems face in maintaining detection consistency across a temporal sequence, which is critical for applications reliant on precise structural assessments. These findings highlight the indispensable role of human oversight in validating and refining the results of automated object detection algorithms, ensuring the reliability and integrity of the data.

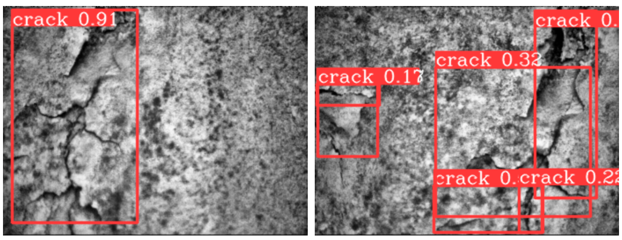

Figure 4: Generated by YOLOv5

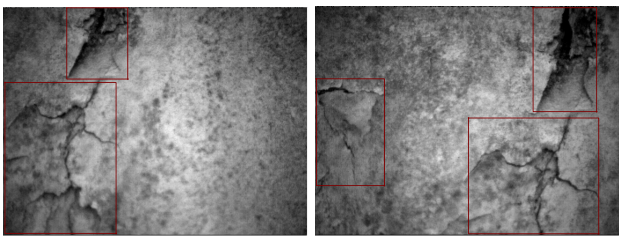

Figure 5: Generated by LabelMe

A salient disparity is observed in the labeling approach adopted by manual annotation via LabelMe as opposed to that by YOLOv5. Manual annotation demonstrates a propensity for recognizing continuity within a crack, often treating it as a single entity even when the crack exhibits variable characteristics along its length. In contrast, YOLOv5's algorithmic propensity is to segment the same crack into multiple distinct entities. This divergence can be attributed to the fundamental differences between the holistic human perception of structural features and the algorithmic processing inherent to YOLOv5. While human annotators integrate contextual cues and the visual continuity of a crack, YOLOv5 operates on a set of learned patterns that may not fully encapsulate such nuances, particularly when a crack's appearance varies subtly along its trajectory.
This contrast underscores a critical aspect of machine learning models: the tendency to analyze based on localized patterns, which can lead to fragmentation in the detection of extended linear features like cracks. YOLOv5's base model, trained on a diverse but finite set of data, may not have encountered sufficient variance within individual crack formations to learn the concept of continuity robustly. This limitation highlights the importance of curating training datasets that encompass a comprehensive representation of the target anomalies and suggests that augmentation of the training data with a wider range of continuous feature presentations could mitigate this issue. It also reinforces the necessity for iterative model tuning and the incorporation of domain-specific knowledge into the training process, potentially through techniques such as transfer learning or fine-tuning with a specialized dataset.

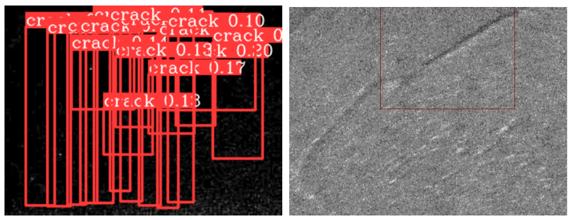

Figure 6: Left: nighttime, Right: after processed

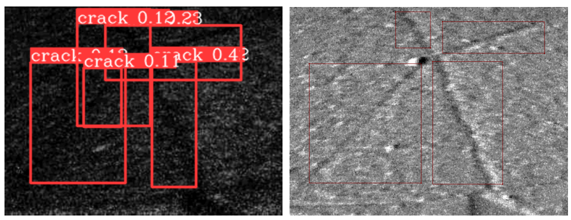

Figure 7: Left: nighttime, Right: after processed

Our comparative evaluation revealed notable variations in YOLOv5's performance under different nighttime conditions. In extremely low-light environments, the model struggled significantly, as highlighted by a series of near-black images where YOLOv5's labels diverged

sharply from manual annotations. This discrepancy underscores the model's reliance on visible features for accurate object detection. In contrast, when night scenes were moderately illuminated, YOLOv5 demonstrated improved accuracy, suggesting its capability to adapt to better-lit conditions. This variability highlights the model's sensitivity to lighting and indicates that with sufficient artificial illumination, its performance can approach the reliability of manual annotations.
The primary challenge for YOLOv5 in dark conditions stems from its dependence on discernible visual cues, which are often absent in nighttime imagery. The typical training datasets, which may not adequately represent such challenging scenarios, result in a model ill-prepared for the complexities of nocturnal environments. This issue is exacerbated by reduced visibility and the nuanced dynamics of nighttime lighting.
However, the model's success under moderately lit conditions reinforces the potential for YOLOv5 to operate effectively across a broader range of lighting scenarios, provided there is enough light to define critical features. This adaptability calls for a comprehensive training approach that encompasses a spectrum of lighting conditions, enhancing the model's robustness. Exploring image enhancement techniques during preprocessing and considering specialized night-vision capabilities are promising strategies to boost YOLOv5's low-light efficacy.

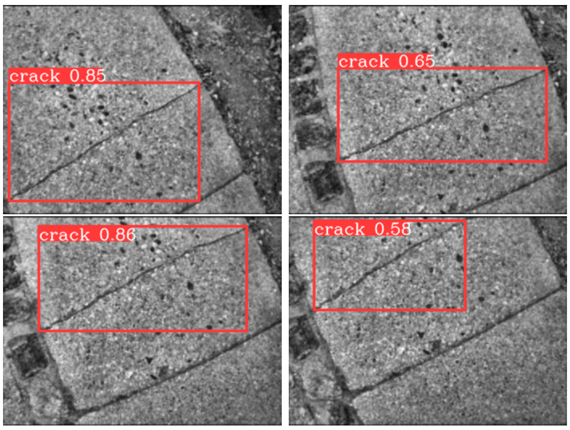

Figure 8: A series of labeled crack frames

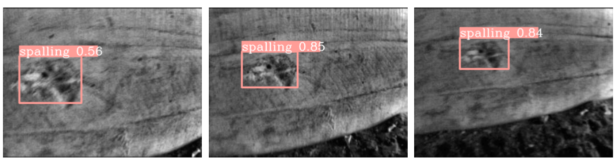

Figure 9: A series of labeled spalling frames

Despite its challenges, YOLOv5's contributions to object detection must be acknowledged, especially in ideal conditions where the model demonstrates exceptional accuracy and speed. In environments with clear visibility and high-contrast objects, YOLOv5 consistently outperforms manual labeling efforts, showcasing its capacity for rapid and precise annotations. This capability is particularly beneficial when dealing with large datasets, where manual annotation would be prohibitively time-consuming and less consistent.
The robust performance of YOLOv5 in favorable conditions suggests that with further refinement, the model's reliability could extend to a broader range of scenarios, including those with complex lighting and object details. The integration of automated systems like YOLOv5 with the meticulous oversight of human annotators forms a complementary approach, enhancing the overall quality and reliability of the data. As technology progresses, leveraging the strengths of both methods will be essential in developing comprehensive and accurate object detection datasets.

4 Discussions

Throughout this study, we faced several challenges with pseudo-labelling, particularly evident in YOLOv5's inconsistent ability to detect subtle and gradational features in object boundaries across diverse environmental conditions. The automated system often failed to capture the entirety of defects such as cracks, leading to incomplete or incorrect labeling. To mitigate these issues, we implemented a strategic mix of manual and automated labeling techniques tailored to the specific needs of different data environments.
In field settings, we bolstered the robustness of our dataset by manually annotating 10 percents of the collected data, carefully selecting samples that showcased a wide range of scenarios. These manually labeled instances were integrated with available public datasets, creating a hybrid training set that enhanced the learning models' exposure to varied real-world conditions. For laboratory data, which allows for more controlled conditions, we ensured comprehensive learning by including all images used in creating the dataset within the training set. This approach facilitated "perfect learning," where the models could thoroughly understand the specific characteristics of defects under stable conditions.
Furthermore, the integration of OpenCV trackers required thoughtful adjustment to maintain continuity across frames. By selectively labeling only 10 to 15 key frames that depicted complete defects, we optimized the trackers' ability to follow the defect across sequences. This selective manual labeling, combined with the use of a median filter, significantly reduced the instances where the defect would disappear or become obscured in consecutive frames, enhancing the overall accuracy of the tracking process.
These methodological adjustments significantly improved the reliability of our pseudo-labelling process,

underscoring the importance of a hybrid approach that leverages both manual precision and automated efficiency. As we continue to refine these techniques, future research will explore the integration of advanced machine learning strategies to further enhance the accuracy and adaptability of pseudo-labelling in facing the complexities of real-world applications.

5 Conclusions

Throughout this report, we have critically examined the efficacy of pseudo-labeling for object detection, specifically targeting the identification of structural defects such as cracks and spallings. Employing both YOLOv5's advanced detection capabilities and the precise manual annotation facilitated by LabelMe, we have uncovered the intrinsic strengths and limitations of automated versus human-led processes in diverse environmental conditions.

Our analysis confirms that while YOLOv5 performs exceptionally under conditions of clear visibility, its efficacy diminishes in low-light scenarios where manual annotations provide superior precision. This highlights the crucial role of manual validation, particularly in complex imaging contexts. The observed tendency of YOLOv5 to fragment detections, contrasted with the continuity preserved by manual labeling, underscores the need for training datasets that are nuanced and context-aware, reflective of the real-world challenges object detection systems must navigate.

Moving forward, the development of object detection models should focus on integrating specific domain nuances, leveraging machine learning advancements to adapt to various environmental conditions. Additionally, the annotation workflow should ideally integrate automated and manual methods to harness the strengths of both, thereby enhancing the quality and efficiency of the resulting datasets.

In conclusion, automated labeling tools like YOLOv5 represent significant progress in the field of object detection but are not a standalone solution. The ongoing need for human expertise in ensuring data quality is undeniable. By continuing to refine these tools and methodologies, and by fostering an environment of continuous improvement and innovation, we can further the capabilities of semi-supervised machine learning in object detection. It is through such dedicated efforts that we will push the boundaries of what can be achieved in this dynamic field of artificial intelligence.