UNDERGRADUATE RESEARCH OPPORTUNITIES PROGRAMME (UROP)

USING COMPUTER VISION TO DETECT HOUSEKEEPING ISSUES ON CONSTRUCTION SITES

by

JOEL LO LIANG ZE

(B.Eng., ISE)

DEPARTMENT OF THE BUILT ENVIRONMENT, COLLEGE OF DESIGN AND ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

Supervisor:

Associate Professor Yang Miang Goh

Semester 2 Academic Year 2022/2023

Abstract

Using computer vision to detect housekeeping issues on construction sites.

The construction industry often involves the storage of materials on-site, which, if not properly maintained, can lead to safety risks and reduced productivity. Traditional methods of site maintenance involve the use of construction workers and supervisors to keep the site tidy, but due to a tight labour market, there is a need to develop computer vision models to reduce reliance on manual labour. This study aims to develop a minimum viable computer vision model that can detect untidiness in construction site images, with a focus on identifying construction materials and waste materials that contribute to untidiness.

To achieve this, we explore the use of data pre-processing techniques for object detection, including image resizing, colour normalisation, and data augmentation. We also compare two state-of-the-art object detection models, DETR and Deformable DETR, to determine their effectiveness in identifying poor housekeeping practices in construction sites. Our experiments show that DETR outperforms Deformable DETR, achieving higher mean average precision scores.

The results of this study demonstrate the potential of computer vision models in improving construction sites maintenance practices and in reducing safety risks on construction sites. The proposed model can be integrated into existing CCTV and body camera systems, enabling real-time monitoring of site cleanliness, and providing actionable insights for construction managers. Future research can focus on expanding the model's capabilities to detect other types of hazards and improving its precision in identifying waste materials contributing to untidiness with more epochs and more images in the training of the DETR model.

Preface

In my report, I will discuss upon the potential of Computer Vision and how it may save lives and money in construction sites. May this report make you a proponent of safety and advanced machine learning technology.

Acknowledgements

I would like to express my deepest gratitude to the following individuals and organisations for their invaluable contributions to this research project on using computer vision to detect housekeeping issues on construction sites.

First and foremost, I would like to extend my thanks to my research supervisor, Associate Professor Yang Miang Goh, for his guidance, encouragement, and expertise throughout the research process. His insightful feedback and constructive criticism were instrumental in shaping the direction and methodology of this study.

I am also grateful to the construction companies that provided access to their sites and facilitated the data collection process. Without their cooperation and support, this research would not have been possible. Special thanks to Housing Development Board for their collaboration.

I would also like to thank the research team members who assisted with various aspects of the research, including data pre-processing, model development, and analysis. Their hard work and dedication were essential in achieving the project's objectives. Special thanks to Michael and Zherui for being my mentors in this research project.

Finally, I would like to acknowledge the support provided by College of Design and Engineering under the National University of Singapore. Their assistance enabled me to carry out this research and advance my understanding of the potential of computer vision in improving construction site safety.

Once again, I express my sincere gratitude to everyone involved in this research project.

Table of Contents

1. INTRODUCTION

1.1 Overview

Housekeeping issues were prevalent across construction sites, creating a risk of slips, trips, and falls. Currently, site cleanliness and maintenance are carried out by construction workers led by supervisors. In Singapore's tight labour situation, artificial intelligence solutions can help supervisors better manage and upkeeping big construction site. The study aims to use object detection to identify poor housekeeping practices.

1.2 Scope

The research seeks to understand how computer vision can help in the detection of poor housekeeping practices without incurring additional manpower. The scope is limited to construction sites in Singapore and their relevant agencies such as the Housing Development Board of Singapore.

1.3 Aims

The research aims to saves lives in construction and reduce the costs to construction companies in Singapore. The eventual goal is to alleviate the site supervisors' work in maintaining a high level of compliance with the safety standards of Singapore. Ultimately, this research aims to develop a minimum viable computer vision model that can detect untidiness in construction site images, with a focus on identifying construction materials and waste materials that contribute to untidiness, poor housekeeping practices and poor safety measures.

1.4 Research Questions

The research focuses on selecting the appropriate computer vision model for use, exploring methods for training, and fine-tuning these models, and identifying the key parameters for optimal detection of poor housekeeping practices.

1.5 Conceptual Framework

The key concepts and variables identified are the usage of transformer-based computer vision models and fine-tuning of said models with custom datasets for identification of poor housekeeping issues with object detection, through the comparison of the inference performance and the suitability of DETR (DEtection TRansformer) (Carion, N., 2020) and Deformable DETR (Zhu, X. 2020) models. By adopting a systematic approach, from data processing to labelling to object detection and evaluation, the research aims to produce valid and reliable results that contribute to the understanding of fine-tuning pre-trained transformer-based models for object detection tasks for identification of poor housekeeping issues.

1.6 Research Paper Synopsis

DETR was used to prove that computer vision can be used to detect poor housekeeping practices albeit with a varying degree of error. More research must be done into other implementations and models to reduce the margin of error for use on construction sites.

2. LITERATURE REVIEW

2.1 Problems

Figure 1

Building construction works involve various trades and necessitate the storage of materials onsite during the construction process. However, improper material storage can lead to safety risks, including slips, trips, falls, and obstructions to access. As shown in Figure 1, these hazards can result in time loss, reduced productivity, and the need for additional labour to rectify the issues. Furthermore, maintaining site cleanliness and safety is particularly challenging given the current tight labour situation, which demands a reduction in worker reliance for site upkeep. As of the timing of this research, there are no official measures in place to automate this system of checks. (LEE QIJIE BRYAN, 2022)

2.2 Theoretical Perspective

The theoretical perspective is that one potential solution to this problem is the development of a computer vision model capable of detecting untidiness in construction site images. While previous efforts by two interns have resulted in the creation of some initial models, further improvements are necessary to achieve a minimum viable model that can effectively classify scenes as tidy or untidy and identify specific waste materials causing untidiness, such as debris, water ponding, misplacements of steel bars and rebars. There is a focus on the implementation of computer vision models in real-time and with a high degree of accuracy.

2. LITERATURE REVIEW

2.3 Supervised Learning

Figure 2

There have been instances of supervised learning being used on constructions sites. Most notably, the implementation of Custom object detection for PPE (Personal Protection Equipment) Detection of Construction Site Workers (Snehilsanyal, N.D.) using YoloV8 by Snehil Sanyal. The implementation uses anchor-based models to detect PPE. It does not conclusively differentiate between good and bad usage of PPE and is only a basic implementation of object detection to detect the absence or presence of PPE. Moreover, it utilises stock images that contains watermarks, as seen in Figure 2, which reduces the accuracy of the model. However, it is useful in laying the foundation for this research in using object detection in construction sites and showcasing the efficacy of object detection models in identifying construction site related equipment.

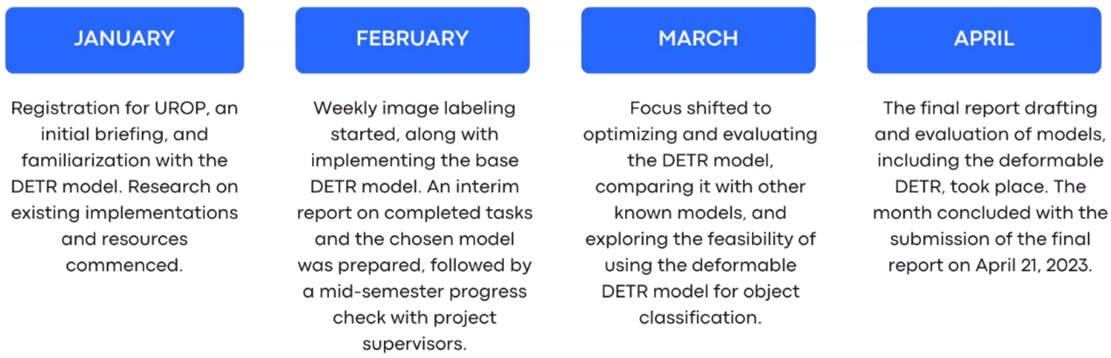
3. Research Approach

3.1 Data Sources

The main data sources are those procured from the construction sites under HDB. Labelling was done with the open-source labelling software, LabelMe. The main labels are debris, water ponding, steel pipes and rebars.

3.2 Process Flow

Figure 3



The main timeline for the research is outlined in Figure 3, by systematically building upon a pretrained DETR model. The research was able to build iteratively and systematically, a computer vision model for detection of poor housekeeping practices in construction sites. Inference using DETR and Deformable DETR models were used to identify which was the better base model to start with.

3. Research Approach

3.2 Process Flow

Figure 4

Figure 5

Running the inference test as follows in the figures above, it can be noted that the average accuracy of the DETR model, shown in Figure 4, is higher than that of the Deformable DETR model, shown in Figure 5, and hence was used as the model of choice for starting the training and testing for object detection of poor housekeeping practices in construction sites.

The steps to train the model are broken down into the following steps:

Data collection: Collect a dataset of images of construction sites, partnered with HDB, with poor housekeeping practices, containing examples of debris, water ponding, steel pipes, and rebars.

Data pre-processing: Use LabelMe to annotate the bounding boxes. Once labelled, resize the images to a fixed size, and split the dataset into training, validation, and testing sets.

3. Research Approach

3.2 Process Flow

Model training: Train a DETR model on the training set using the following steps:

a. Encode the input image using a convolutional neural network (CNN) to generate a feature map.

b. Generate a set of object queries, which are learned vectors that will be used to predict the objects in the image.

c. Apply the transformer network to the feature map and the object queries to predict a set of bounding boxes and labels for each object query.

d. Compute the loss function, which is a combination of the bounding box loss and the label classification loss, and backpropagate the gradients to update the model parameters.

e. Repeat steps b-d for multiple epochs until the model converges. In the research, the model was trained for 100 epochs.

Model evaluation: Evaluate the performance of the trained model on the validation and testing sets using standard metrics such as mean average precision (mAP).

Model deployment: Deploy the trained model to detect poor housekeeping practices in construction sites using debris, water ponding, steel pipes, and rebars as labels in real-world scenarios. This can be implemented with body cameras and CCTVs to aid the site supervisor in detection of these poor housekeeping practices as it happens. Currently, this step is beyond the scope of the current study.

To abstract the low-level details of the implementation of the steps, the pytorch, tensorflow and huggingface libraries were used to build the DETR model.

3.3 Testing and Data Validation

The main testing of the model was validated with completely new images that were not used for training and testing. Data validation of the accuracy of labels are verified with HDB on a biweekly basis with feedback from industrial experts on the good and bad practices identified.

4. OBJECT DETECTION

4.1 Results of experiments

The key results shows that the DETR Model can detect some instances of poor housekeeping practices as well as identify the construction site objects that are responsible for the poor housekeeping practices through object detection. However, given the limited size of training datasets and epochs for training, it is not able to reliably detect poor housekeeping practices and tends to miss some objects, resulting in a moderate level of false positives.

Figure 6

In the example shown in Figure 6, the model was able to identify a segment of the debris and was able to correctly classify it as a poor housekeeping practice, on the left-hand side. However, it defers from the truth, on the right hand-side, and was unable to detect a larger area of the debris.

Figure 7

On further experimentations and testing, it was able to identify poor housekeeping practices that was left out in the initial labelling such as the misplaced rebars in the left-hand side of Figure 7. Unfortunately, it was unable to detect precisely the 2 misplaced steel pipes as shown in the truth labelling in the right-hand side of Figure 7.

4. OBJECT DETECTION

4.2 Performance of Model

The main performance of the DETR Model is evaluated as follows:

Evaluation Metric	IoU Threshold	Object Area	Maximum Detections	Score
Average Precision (AP)	0.50-0.95	All	100	0.005
Average Precision (AP)	0.50	All	100	0.018
Average Precision (AP)	0.75	All	100	0.000
Average Precision (AP)	0.50-0.95	Small	100	-1.000
Average Precision (AP)	0.50-0.95	Medium	100	0.000
Average Precision (AP)	0.50-0.95	Large	100	0.006
Average Recall (AR)	0.50-0.95	All	1	0.019
Average Recall (AR)	0.50-0.95	All	10	0.019
Average Recall (AR)	0.50-0.95	All	100	0.019
Average Recall (AR)	0.50-0.95	Small	100	-1.000
Average Recall (AR)	0.50-0.95	Medium	100	0.000
Average Recall (AR)	0.50-0.95	Large	100	0.021

Table 1

Table 1 shows the evaluation results of the DETR model for object detection of poor housekeeping practices in construction sites. The results are presented in terms of Average Precision (AP) and Average Recall (AR) scores for different IoU thresholds, object areas, and maximum detections. The evaluation results suggest that the model has low precision and recall for most categories of object sizes and IoU thresholds, and it failed to detect any small objects.

4.3 Discussion

Overall, the evaluation results suggest that the DETR model performed poorly in detecting objects with low precision and recall for all object sizes and IoU thresholds but is able to detect poor housekeeping images. This could indicate that the model needs more training data or tuning of hyperparameters to improve its performance before it can go into production. However, as a minimum viable product, it can classify good and poor housekeeping practices.

5. CONCLUSION AND RECOMMENDATIONS

5.1 Assumptions

The key assumption made here is that a lack of object detection of poor instances of housekeeping practices is indicative of good housekeeping practices, which might not always be the case. Moreover, it is assumed that the key contributors to poor housekeeping practices are mainly the 4 labels and not any other parameters, which might limit the usefulness in production where construction companies have additional sets of guidelines on what constitutes a good or bad housekeeping practice.

5.2 Limitations

The key limitation is the lack of data to train on and the lack of epochs.

The training/testing/validation data split is as follows:

- Number of training examples: 128
- Number of validation examples: 37
- Number of test examples: 18

Typically, in large computer vision models, a minimum train sample of 10,000 images would be used to train on. Unfortunately, due to the time constraint and difficulty in manual labelling of the custom construction site datasets. The model is limited in the amount of data it was trained on. Moreover, the number of epochs and cycles used were insufficient as seen from the extremely low recall, which suggests that a longer epoch and training cycle could have been used.

5.3 Future Works

In conclusion, the DETR model has the potential to classify poor and good house keeping practices and single out and label the objects responsible for such classification. However, the precision of the model can be improved with further finetuning and more data points for training. Future works includes porting the model over to production for use in body cameras installed on construction workers and CCTVs that are stationed around construction sites for better surveillance of poor housekeeping practices.

6. REFERENCES

Carion, N. (2020, May 26). End-to-End Object Detection with Transformers. arXiv.org. https://arxiv.org/abs/2005.12872

LEE QIJIE BRYAN (2022). EVALUATION OF HOUSEKEEPING ISSUES ON CONSTRUCTION SITES IN SINGAPORE. ScholarBank@NUS Repository.

Snehilsanyal. (n.d.). GitHub - snehilsanyal/Construction-Site-Safety-PPE-Detection: Custom object detection for PPE Detection of Construction Site Workers. This repo contains notebook for PPE Detection using YoloV8. GitHub. https://github.com/snehilsanyal/Construction-Site-Safety-PPE-Detection

Zhu, X. (2020, October 8). Deformable DETR: Deformable Transformers for End-to-End Object Detection. arXiv.org. https://arxiv.org/abs/2010.04159

7. APPENDICES

7.1 Appendix A - Timeline

Week	Task/Agenda	Item/Deliverable
Week -2 to Week 2	-Registration for UROP	-UROP Sem 2, AY 22/23
Week 3 26/01/2023 – 29/01/2023	-Initial Briefing (260123 1400-1500) -Read up on DETR Model for object classification	-UROP Timeline Summary of readings
Week 4 30/01/2023 – 05/02/2023	-Start weekly labelling of images -Implement base model of DETR for image classification, using pre-existing online data	-Labelme – 500 (vet) -DETR Template model .ipynb file
Week 5 06/02/2023 – 12/02/2023	-Instance segmentation -Familiarise with using onsite GPUs at SDE while attempting to train model	-Labelme – 500 (vet) -Instance segmentation
Week 6 13/02/2023 – 19/02/2023	-Work on DETR Classification Model	-Labelme – 500 (vet) -DETR Base Model

		-Check with prof and Ras on labelling
Recess Week 20/02/2023 – 26/02/2023	-Interim Report (completed tasks and writeup on chosen model) -Work on DETR Classification Model	-Labelme – 500 (vet) -Check with HDB on labelling -Mid-Semester progress check with Prof Goh, Michael, and Zherui
Week 7 27/02/2023 – 05/03/2023	-Work on DETR Classification Model -Data Validation of labelled images	-Labelme – 500 (vet) -DETR Base Model
Week 8 06/03/2023 – 12/03/2023	-Optimisation of DETR Classification Model	-DETR Base Model (Optimised)
Week 9 13/03/2023 – 19/03/2023	-Evaluation of DETR Classification Model	-DETR Final Model (Optimised) -Documentation and Comparison with other known models
Week 10 20/03/2023 – 26/03/2023	-Explore feasibility of using deformable DETR Model for object classification	-Deformable DETR Model
Week 11 27/03/2023 – 02/04/2023	-Start working on the Final Report -Evaluation of Models	-Final Report Draft 1 -Deformable DETR Comparison to DETR
Week 12 03/04/2023 – 09/04/2023	-Report Draft 2 -Wrap up research	-Final Report Draft 2 -Clean up code base and comments
Week 13 10/04/2023 – 16/04/2023	-Finish Citations -Finish Final Report	-Final Draft report -Compilation of Final Deliverables
Week 14 17/04/2023 – 21/04/2023	-Submission of Final Report	-Final Report