Final Report - Machine Learning for Sewer Deficiency Detection

Amey Dhamgunde, Rory Gao, Zidan Li, Victor Milne
MIE429 Capstone Design for Crozier Consulting Engineers

I. Introduction

A foundational component of residential and commercial land development is the proper maintenance of civil infrastructure systems such as storm and sanitary sewers. Developers must follow specific engineering design and construction standards and projects are subject to extensive review and inspection prior to approval and execution.

Crozier Consulting Engineers has 20 years of experience supporting developers in implementing civil infrastructure systems. One component of their work involves manually reviewing robot-captured CCTV video footage, provided by an inspection contractor, of storm and sanitary sewers. Crozier's employees then report on any cracks, debris, and other deficiencies identified in the footage and offer suggestions on remediating these deficiencies.

Developments can feature many kilometers of mostly deficiency-free infrastructure, but Crozier's employees must review all the given footage for completeness. The current video review process is time-consuming and requires meticulous attention; Crozier aims to reduce the required manual labour during this process. An efficient project workflow is essential in supporting Crozier's future growth and hastening project turnaround times for developers.

This report presents a *machine-learning approach* to reducing manual labour in Crozier's deficiency review process. Our solution must specifically reduce the total amount of time Crozier's employees spend on reviewing the video footage.

II. Description of Data

Two main sources of data are relevant to our work: The Sewer-ML dataset [1] and Crozier's CCTV videos. Sewer-ML is an open-source dataset containing 1.3 million high-quality labelled images of both healthy and defective sewer pipes. Crozier's data is unlabelled; we instead are given the corresponding inspection reports noting the physical locations of any defects found in the videos. Both datasets can be deconstructed into individual frames of labelled sewer images, some of which featuring defects that an ideal automated system should detect. For Crozier's data, we have an additional task of generating labelled images using these reports.

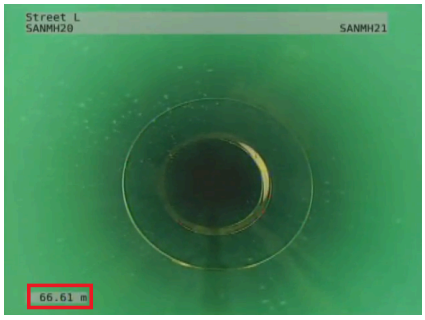

Figure 1: Crozier data - Typical sewer pipe with no defects. Note the distance stamp which tracks the position of the camera along the pipe.

Figure 2: Crozier data - Snapshot of Sewer pipe with no defects. Note the variation in resolution and pipe color.

Crozier's CCTV footage was collected between 2021 and 2023 by different inspection contractors across three residential land development projects in Southern Ontario. Each video involves a human-controlled robot traversing sections of the sewer network with a front-facing

camera. The robot mostly traverses forwards, but may backtrack or fully return to the origin. When encountering service laterals, visible defects, and other regions of interest, the camera stops and rotates to view these features. Given the data collection process, the system must be able to recognize sewer defects from different camera angles, and be resilient to differing recording software and hardware.

Due to the significant effort required to extract training data from Crozier's unlabeled videos, we primarily trained our model using Sewer-ML using both the full dataset and also a subset of the images featuring specific classes of deficiencies unique to Crozier's dataset (see appendix). However, to support future training performed strictly on Crozier's data, we outline a data labelling methodology below.

III. Methods

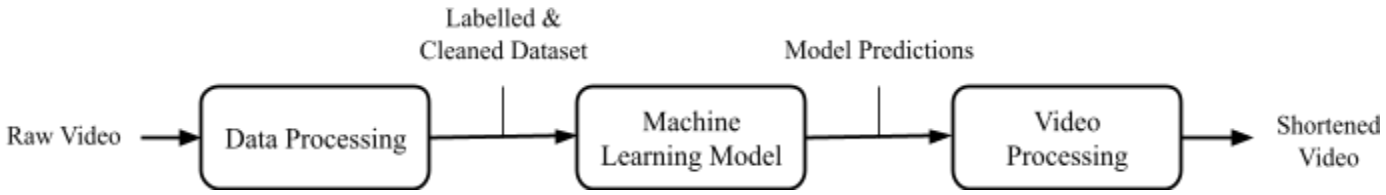

Figure 3: Outline of project methodology

Data Labelling

The challenge of preprocessing and generating training data from Crozier's unlabeled CCTV sewer videos demands a specialized methodological approach. Central to this task is the challenge of associating video frames with documented sewer defects, where existing documentation indicates physical locations within a sewer stretch. CCTV videos include distance markings typically positioned in the bottom left or right of the screen, which denote precise pipe locations. However, reading these meter markings presents a complex Optical Character Recognition (OCR) challenge, complicated by significant variations across different CCTV footage providers—each with unique fonts, text styles, and varying noise levels.

Contemporary OCR methodologies leverage machine learning techniques alongside robust image preprocessing and filtering strategies. In our approach, we used EasyOCR, a readily available Python package that offers a comprehensive suite of machine learning and processing layers designed to detect and output text within images [2]. This off-the-shelf solution provides a reliable framework for extracting critical location information from diverse and challenging video footage.

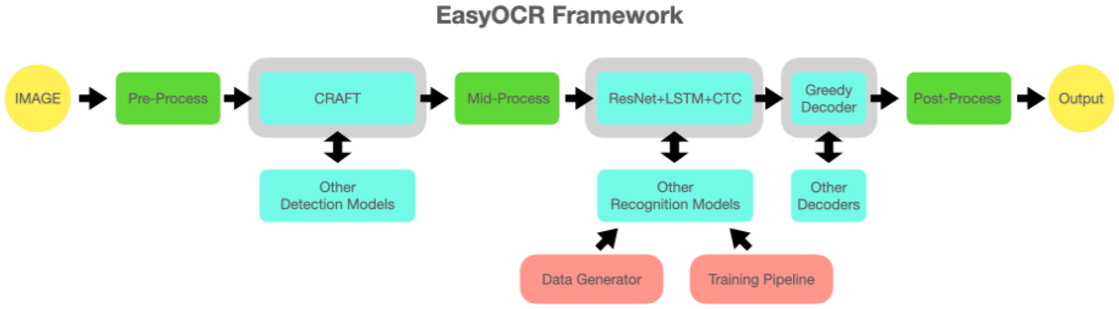

Figure 4: Outline of EasyOCR's components and structure

However, the naive approach of reading distance markings frame by frame often produced inaccurate results. To improve OCR accuracy, we implemented several strategies. Basic preprocessing included resizing images to make small text easier to detect and applying a binary threshold to convert pixels to 0s and 1s, which improved contrast and removed color artifacts. We developed a contour-based detection system to split the text box into individual digit bounding boxes, reading characters one at a time to reduce the chances of missing or hallucinating a nonexistent character. A rules-based system was created to constrain character readings, such as only allowing certain digits to increment by 1 between frames. This helped filter out unreliable or random readings. Finally, we applied a post-processing smoothing function to the entire data timeline to remove local inconsistencies. These techniques significantly improved the reliability of extracting location information from CCTV sewer inspection videos.

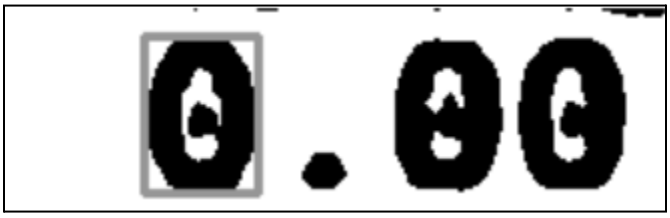

Figure 5: Example of a contour selected from a binary threshold text image

Sewer Image Classification

a. Approach and Model Selection

We approach the sewer image classification task using deep learning techniques. We concentrated on Convolutional Neural Networks (CNNs) for this task. CNNs are a state-of-the-art technology for image classification tasks as the architecture allows the capturing of many visual patterns of interest. This is crucial for our application as there exist many different classes of sewer pipe defects that our model must recognize. Furthermore, CNNs are relatively fast and stable to train and deploy, do not require manual feature engineering, and are computationally efficient for image classification compared to other machine learning approaches.

In our model development process, we systematically evaluated two prominent CNN architectures: ResNet50 and VGG16. These architectures differ fundamentally in their layering methods, with ResNet50 employing residual connections, while VGG16 relies on hierarchical stacks of convolutional and pooling layers. Through performance testing, ResNet50 emerged as the superior model and was selected as our primary approach.

For ResNet50, we explored two weight initialization strategies: training the model from scratch using random weight initialization, and leveraging a pre-trained model sourced from the PyTorch Vision repository. This approach allows us to compare the performance implications of starting with random versus pre-learned weights. Image preprocessing played a crucial role in preparing our input data for the ResNet50 model. We implemented two key preprocessing steps: first, standardizing pixel values to normalize the input data, and second, resizing images to the specific input format of 224x224x3 pixels required by the ResNet architecture. Finally, our hyperparameter tuning process focused on learning rate variation and image augmentation techniques - which were pixel color inversion and additive Gaussian noise.

b. Model Performance Evaluation

We assessed the performance of our models using validation accuracy as the selection criteria. Our exploration showed that starting from a pretrained model offered better performance.

Architecture	Pretrained?	Learning rate	Validation accuracy
ResNet50	Yes	0.001	0.8202
ResNet50	Yes	0.005	0.7539
ResNet50	No	0.001	0.7376

Table 1: A comparison of the top-performing models

Finally, our best chosen model was then fine tuned on a small subset (~100) of crozier data. This was done to impart some generalization to Crozier's datasets.

All code to reproduce our results is in our Github: (biosharp18/MIE429_capstone: Crozier CCTV sewer detection)

c. Model Inference on Crozier Datasets

The next step is to apply our trained CNN model to Crozier's provided CCTV videos. In the deployment scenario, the model processes videos using a frame-by-frame inference approach. Each video undergoes the same preprocessing pipeline previously described, including pixel standardization and resolution scaling to match the model's input requirements.

For each frame, the model predicts the deficiency likelihood. It then produces a detailed plot that visualizes the predicted likelihood of defects across the video's duration. This granular analysis allows for a comprehensive assessment of potential sewer infrastructure issues.

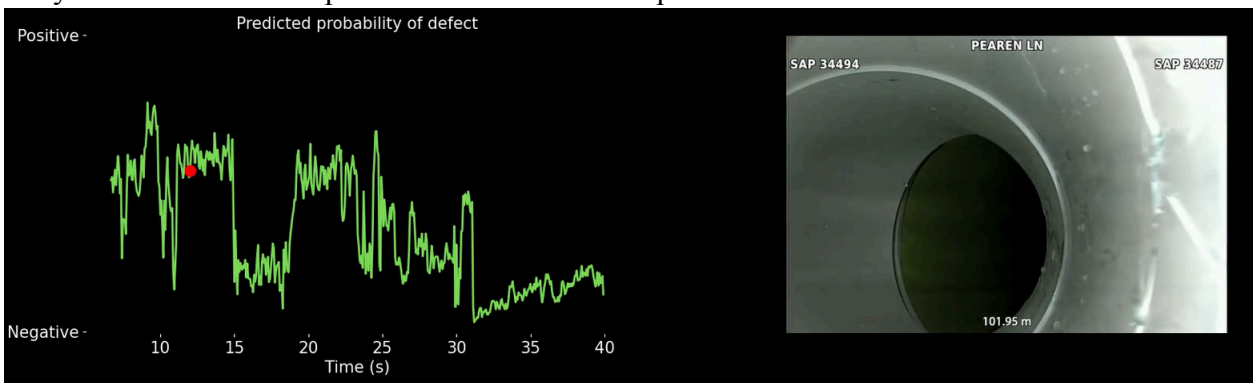

Figure 6: Example of a model prediction plot

d. Video Processing

The video processing stage leverages two Python libraries—OpenCV-Python and MoviePy—to manipulate video playback dynamically. Using the defect likelihood predictions and their corresponding timestamps, we implemented an adaptive video processing technique that adjusts the video playback speed to be inversely proportional to the predicted deficiency likelihood at that frame. The resulting video features a fast playback speed when the model

predicts a low deficiency probability, and slow playback speed when the model predicts a high deficiency probability. We apply gaussian smoothing to the model predictions to allow for smoother, less abrupt speed transitions.

The range of playback speed adjustment factors is specified by a lower and upper bound and should be modified to align with the viewing preferences of Crozier's employees. We used an exponential function to define the speed curve between the upper and lower bounds. Example speed curves are depicted below:

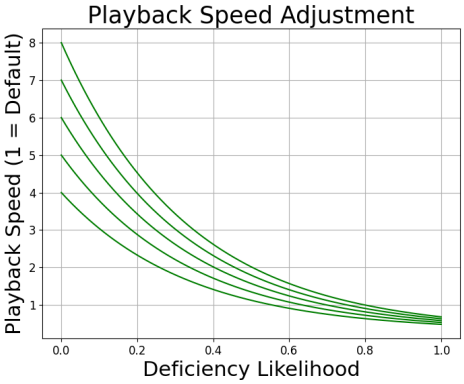

Figure 7: Speed adjustment curves with lower bound $a = 0.3$ *and upper bound* $b = \{4, 5, 6, 7, 8\}$. *The exponential curve is specified by the following function, setting* x *as the predicted deficiency likelihood:*

$$speed(x) = a + (b - a)(e^{-cx}) \qquad (1)$$

where c is an arbitrary constant that can be modified.

Another option is to adjust the playback speed depending on the movement of the video-capturing robot. We note that many deficiencies occur at connections between the main sewer line and the laterals. In addition, the camera panning motion can occur abruptly, but always when the robot is positionally at rest. We can therefore observe the OCR output and slow the playback speed when the robot is stationary. This strategy can be combined with the exponential curve above to achieve a more optimal speed adjustment function for Crozier.

IV. Results

The performance metrics highlight a typical machine learning challenge: model overfitting. The variance between training and validation metrics suggests the model has learned the training data characteristics more than the generalizable patterns. Despite these variations, the validation metrics remain competitive, indicating the model's utility for sewer defect detection. Slight overfitting is not uncommon in specialized computer vision tasks with limited datasets.

Split	Accuracy	Precision	Recall
Train	0.8379	0.8115	0.8356
Validation	0.7764	0.7756	0.6454

Table 2: Best model's performance metrics across training and validation splits.

a. Deployment Evaluation

Quantitatively assessing model performance in the deployment scenario presented significant challenges due to limited ground truth labels. To address this, we adopted a qualitative benchmarking approach focusing on the model's ability to recognize defects in Crozier's sewer inspection videos.

We deployed our best-performing model, selected based on validation accuracy, into our visualization tool. Our analysis reveals good insights into the model's predictive capabilities:

- Strengths: The model demonstrates promising potential in defect detection, successfully identifying crack defects in certain video segments.

- Limitations: Inconsistent performance was observed, with the model failing to consistently label identical defects across different video timestamps.

b. Visualization of Prediction Scenarios

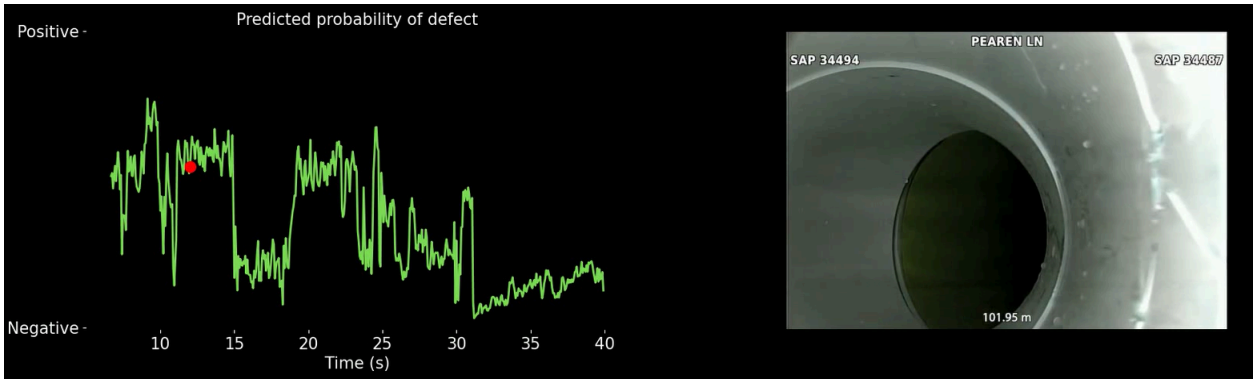

Figure 8: True Positive - High predicted defect probability correctly identifying a crack defect

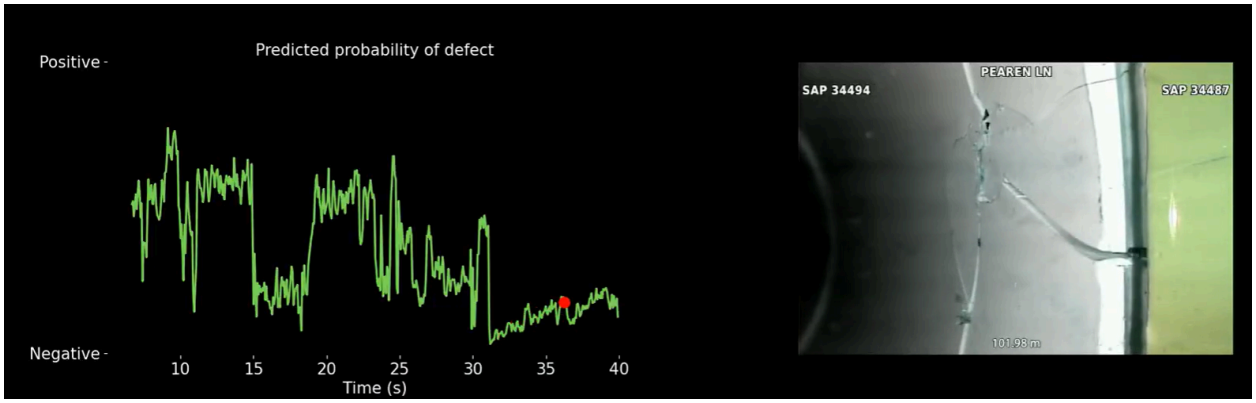

Figure 9: False Negative - Low predicted defect probability when a crack defect is present

Key Takeaways

While the model provides a useful signal for defect detection, we emphasize to Crozier the importance of recognizing its current limitations. The inconsistent performance underscores the need for continued model refinement and careful interpretation of results.

c. Video processing results

We apply video processing to 5 of Crozier's provided videos. We use 50 step frame classification, a lower and upper speed adjustment bound of 0.3x and 8x, and the exponential function specified in equation 1. The length reduction for these examples is between 60% and 80% depending on the frequency and likelihood of predicted deficiencies in the video. We expect these results to be consistent across other videos.

We also estimate the expected time required to load the model, classify the frames in the video, and then process each video, shown in the total runtime column. This value will vary between users, vary based on resource availability, and vary based on the step interval used.

Video	Raw length	Processed length	**Length Reduction**	Model load	Analysis (CPU)	Video processing	**Total runtime**
1	2m 1s	48s	**60%**	25s	20s	47s	**1m 32s**
2	18m 10s	7m 11s	**60%**		1m 32s	5m 27s	**7m 24s**
3	12m 44s	2m 35s	**80%**		2m 0s	4m 43s	**7m 8s**
4	8m 10s	1m 32s	**81%**		1m 30s	3m 16s	**5m 11s**
5	7m 34s	1m 22s	**82%**		1m 15s	2m 37s	**4m 17s**

Table 3: Video processing results. Model load only required once if processing multiple videos

d. OCR Results

Switching focus to the task of labeling the Crozier dataset, the results from the OCR distance reading system were very positive, as most of the outputs conformed well to the reality of the video. The following are six graphs, one from each of the six categories of sewer video provided by Crozier (Sanitary and Storm for each of Colgan, Mattamy Salem, and Prestonvale). Four different font and video styles are present within this sample, showing the generalization ability of the software.

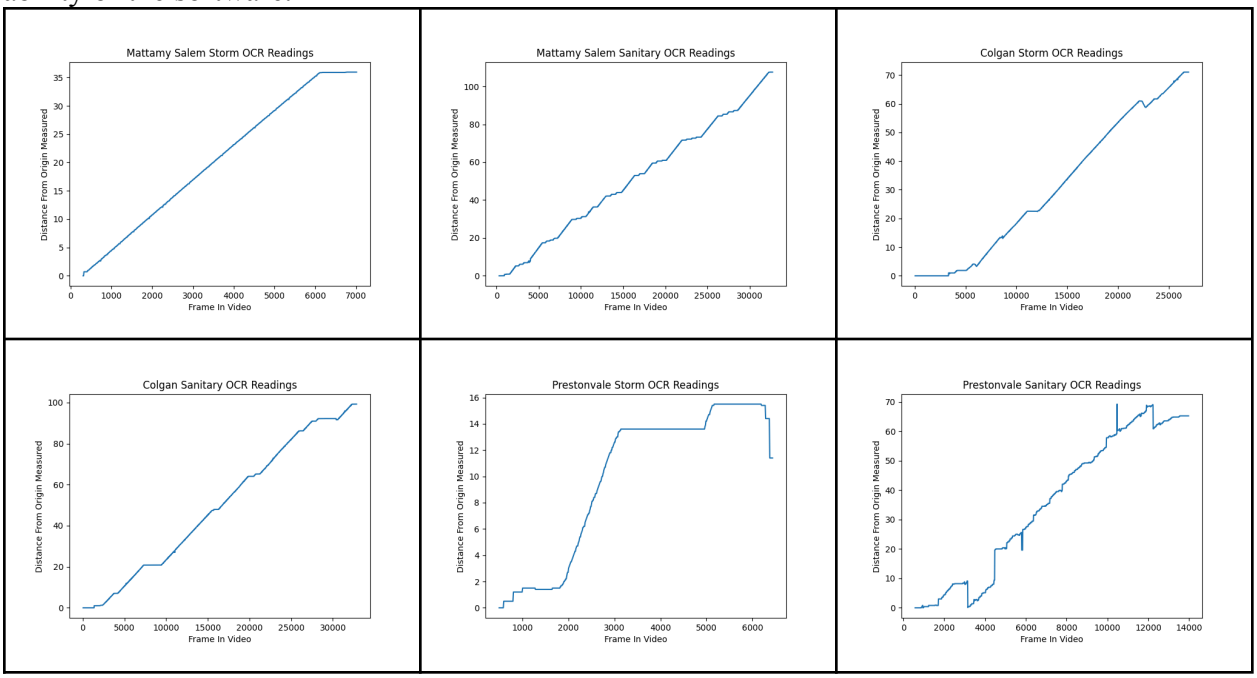

Table 4: Sample distance by frame count graphs from different sources

Visual inspection is a reasonable method of appraising these results, as discontinuities are easy to detect and almost always indicate a mistake on the OCR's part. For another possibly useful metric, one could look at the gaps between frames in metres, showing the outlier results.

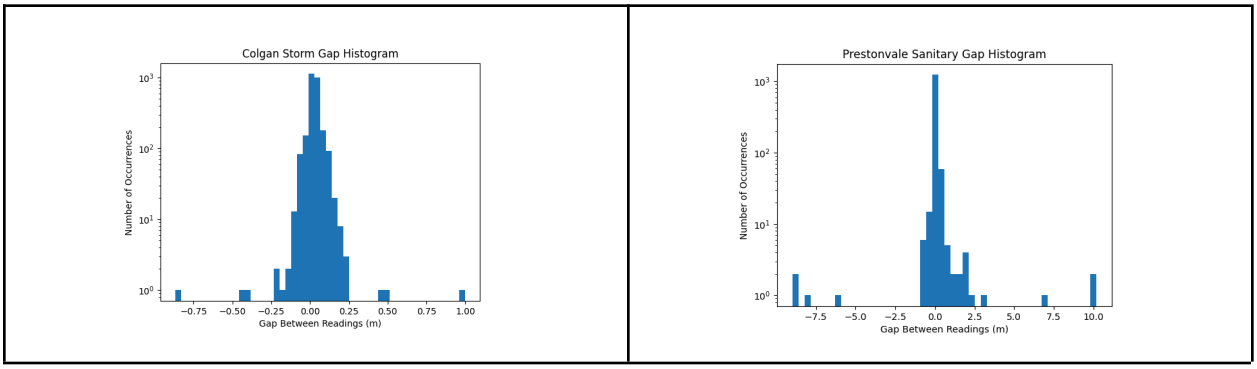

Table 5: Logarithmic scale gap size histograms, lengths in metres between frames

The tables above show the logarithmic frequency of gap sizes between every frame in two videos, one where the OCR system performs well, and one where it doesn't. The prevalence of gap sizes of 5m or greater shows the problems with this particular Prestonvale reading.

V. Discussion

a. Current Utility of our Work for Crozier

The utility of our solution to Crozier's needs depends mainly on the predictive model performance. In general, the model's ability to correctly identify deficiencies allows the playback speed to change dynamically throughout the video. However, we note that even with limited prediction accuracy, the resulting video retains the entirety of the original content, and employees retain the ability to pause the video or adjust the playback speed as desired. Due to the class imbalance within the data, correctly classifying sections of pipe as deficiency-free will offer the greatest reduction in video duration and help achieve Crozier's main objective of reducing manual labour. We believe that our model and resulting output sufficiently achieves the desired goal of the project.

b. Future Model Training and the OCR Framework

The provided model will inevitably require additional training as Crozier expands and supports new, larger projects. The OCR framework aims to support the training process. Though initially planned to be used to gather training data for the current model, the system can still provide value to the client for their future training needs. They have access to many unlabeled CCTV videos and corresponding reports labeling defects, and software that automatically associates the two to create labeled training data can provide a lot of value.

Of course, the system as currently implemented is not perfect, and it struggles on videos from certain providers. The Prestonvale Sanitary footage in particular features very noisy text displays and uses a font that does not seem to align well with EasyOCR's capabilities. While certain types of errors are robustly smoothed over by the postprocessing step, some pitfalls are not currently adjusted for. For example, a system rule only allows tens and hundreds digits to increment when the previous digit is a nine (19 → 20, 099 → 100), but in cases where the nine digit is misread as an eight or another number, the digit is not allowed to increment. As another example, in videos where the distance measure disappears from the screen, the system does not

have a check for this as currently implemented, instead guessing the “most likely” answer from noise.

Given this process is intended for use in training models for aiding client-facing activities, reliability is a key concern. There are certainly improvements that can be made to fix some of the remaining issues with the OCR reader, but human oversight is most important. Since the software outputs graphs of the distance readings for each frame, a human can quickly glance over the output and evaluate accuracy. Reliable readings could then easily be passed to an automated data collection process, while problematic readings could be put to the side for manual review. This fits well with the general theme of the project as a whole, in providing time-saving opportunities while maintaining human evaluation for reliability.

c. Broader Utility of our Work

The methodology used in the inspection and subsequent remediation of sewers can be extended to other engineering domains. In particular, in the natural gas and liquids pipelines industry, a computer vision approach to detecting interior deficiencies could be considered as a lower-impact alternative to other mechanical inspection techniques. Apart from pipelines, any construction project requiring comprehensive inspection could make use of computer vision technology to accelerate the review process necessary to complete the project.

d. Ethical Considerations

Our initial plan in the project proposal was to return a subset of video footage exclusively containing the parts of the footage labelled as defective. However, we recognize that it is important to keep Crozier accountable for the final review and identification of defects. Allowing the model to cut out potentially deficient footage raises significant ethical concerns regarding which party should be held accountable for the mistake. We therefore pivoted our approach to adjusting the playback speed of the video, such false negatives in the model can still be seen through human review.

VI. Implementation

We have developed the components required to automate parts of Crozier’s sewer deficiency detection process. The current prototype can accept input video, generate model predictions, and produce a processed output video. Our pipeline can achieve a 60-80% reduction in processing time through the dynamic video speed adjustment approach.

However, the current implementation exists as a Python script, which presents significant barriers to client adoption. To enhance usability, Crozier will need to develop a front-end interface and create other visualization tools (if needed) for model predictions. We also acknowledge that the current model predictions exhibit substantial noise due to insufficient training and fine-tuning, which can limit the tool's effectiveness within larger land development projects. The OCR methodology aims to address this challenge by allowing Crozier to generate additional, class-specific training data to be used for further training the model.

For the OCR system, the current implementation is not yet an automated end-to-end training data curation process. The current script outputs distance values for every frame in a video, and work remains in exporting these results, reading the defect locations from the inspection reports, and using both sets of data to extract positive and negative labeled training data from the videos.

VII. Conclusion & Future directions

In this project, we have provided Crozier with a machine learning model that has been trained on a subset of Sewer-ML and Crozier-provided data. We're also providing a OCR data curation method for Crozier to continue fine-tuning the model using their own data.

Generally, we addressed the main goal of the project which was to implement a system with the potential to reduce the required manual labour on Crozier's end while leveraging machine learning. The video speedup system we implemented is our form of addressing it. We note that the deficiency prediction model provides great potential for other methods of defect visualization. As for the requirements we set in our proposal, we successfully eliminated the potential to miss potential defects by choosing to process videos by scaling playback speed instead of frame deletion. We also provide flexibility and scalability by providing many useful individual components in our solution which may be assembled by Crozier as they see fit.

The next steps for this project include tying the components we've delivered together into a fully-fledged application accessible to Crozier's employees of all backgrounds. Additionally, our predictive model could benefit from further training and fine-tuning on Crozier data. We hope our machine learning approach will greatly accelerate the video analysis component of the sewer deficiency detection process and offer a scalable solution to meet Crozier's future growth and project labour demands.

VIII. References

[1] J. B. Haurum and T. B. Moeslund, Sewer-ML: a multi-label sewer defect classification dataset and benchmark, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), (2021), 13456-13467. https://arxiv.org/pdf/2103.10895v1.

[2] EasyOCR, Jaided AI, https://github.com/JaidedAI/EasyOCR

IX. Appendix

a. Attribution Table

Name	Contributions
Amey Dhamgunde	Report contributions: - Methods (analysis + justification of model/architecture, etc), Discussion, Conclusion, Project contributions: - Client email communication, model testing environment + boilerplate setup, report writing + presentation creation - Latte Art
Rory Gao	Report contributions: - Implementation, Results, Methods (Model performance eval and Deployment evaluation) Project contributions: - Model training and iteration - Deployment (demo videos) - Ground the coffee beans
Daniel Li	Report contributions: - Introduction - Video processing methodology, results, and discussion Project contributions: - Introduction, project motivation, client background research - Client email communication - Performed manual data labelling on Crozier's dataset - Implemented the video processing methodology
Victor Milne	Report contributions: - Data (Crozier dataset, part of the Sewer-ML text) - All sections in Methods, Results, Discussion, and Implementation related to the OCR work Project contributions: - Initial exploratory research on solutions - Contributed to report and presentation creation - Solely designed and built the OCR distance reading system

b. Classes of Deficiencies in Sewer-ML [1]

Table 1: **Sewer inspection classes.** Overview and short description of each annotation class [17] and the class-importance weights (CIW) [16].

Code	Description	CIW
VA	Water Level (in percentages)	0.0310
RB	Cracks, breaks, and collapses	1.0000
OB	Surface damage	0.5518
PF	Production error	0.2896
DE	Deformation	0.1622
FS	Displaced joint	0.6419
IS	Intruding sealing material	0.1847
RO	Roots	0.3559
IN	Infiltration	0.3131
AF	Settled deposits	0.0811
BE	Attached deposits	0.2275
FO	Obstacle	0.2477
GR	Branch pipe	0.0901
PH	Chiseled connection	0.4167
PB	Drilled connection	0.4167
OS	Lateral reinstatement cuts	0.9009
OP	Connection with transition profile	0.3829
OK	Connection with construction changes	0.4396

We trained the CNN model first using the one entire set of Sewer-ML data. We then retrained using only a subset of deficiency classes corresponding to those featured in Crozier's dataset:

- RB: Cracks, breaks, collapses
- IS: Intruding sealing material
- RO: Roots
- AF: Settled deposits
- BE: Attached deposits
- FO: Obstacle