

Team Copy-Waste - Project Vlog 5

Team Member (Re)introductions

- Nolan - Machine Learning Lead
- Will - Back-End Services Lead
- Rishabh - Front-End Services Lead

Vision:

Our project has two objectives; reduce risks to the public, waste management employees and facilities and reduce the cost to municipalities and waste management companies caused by rare and severe contaminants in recycling waste collection.

Mission

We strive to achieve our vision by automating the detection of severe and rare contaminants in the recycling stream

Business Need

Our business need, to reiterate, is to reduce risk and cost to municipalities and waste management employees. As recycling rates increase, contamination and risk becomes a growing concern.

Fires are a major concern at waste facilities: Here in Regina there were two serious incidents in 2021. This is suspected to be caused by batteries, which is currently not detectable as there is no image dataset available.

Current: Status: Green

Project Issues

No issues at this time

Project Changes

- Mitacs Project Approved
- IP contracts signed with Industry Partner
- Bin Detector completed and handed off to Industry Partner

Individual Contributions (Feb 1 - Mar 8):

Team:

- Completed and updated project documentation
- Auditing 90,000 images of contaminants in recycling collection (30,000 each)
- Pair programming sessions and code reviews

- Meetings with Industry Partner to discuss edge cases, platform integration and perform code reviews.

Will:

- Auditing grocery bag annotations
- Adding the ability to properly export transformed annotation files
- Layering original and artificial objects
- Refactoring
 - Moving functions
 - Changing variable/function names
 - Extracting logic for more reusable code
- Auditing and annotating yard waste images
- Change exporting format to match that of incoming data

Rishabh:

- Handling pasting multiple objects with unique transformations
- Layering existing and artificial objects within the augmented image
- Re-trigger transformations when an edge case flag is raised
- Exporting augmented annotations in COCO format
- Conducting User Testing on the dashboard
- Implementing graphs into the dashboard
- Auditing and annotating yard waste images
- Begin logging performance metrics as the data augmentation is running

Nolan:

- Process for converting Polygon formats
- Detecting and adjusting overlapped annotation masks.
- Implemented argument flags for edge cases, allow augmentation pipeline to reprocess any images that resulted in an edge case.
- Generated additional metrics for the bin tip detector.
- Transition management of bin tip detector to industry partner.
- Created GIS review of our project dashboard.
- Experiment with data augmentation pipeline parameters
- Create function to handle environment variables
- Export data directly to S3 Buckets

Bin Tip Detector

The bin detector has been continuously evaluated for real-world performance over the past 4 weeks. It has performed within expected parameters and management of the project has been

transitioned to our industry partner. All future development will be handled by our industry partner.

Simple Copy-Paste

In the past 4 weeks, we audited 90,000 images of the current recycling contaminant dataset, alongside our industry partner to ensure the best performance through our data augmentation pipeline. All of the auditing resulted in a validated dataset we can trust for our project

As you can see all we needed to do was click on each image which contains inaccurate annotations and then press space bar to view the next set of images. Many of these annotations were found to be too small, containing the wrong object, too dark, or too blurry. All of these inaccuracies contribute to hindering the performance of detection models created through our copy paste pipeline.

The first augmented dataset we will be creating is for the contaminant yard waste. The yard waste images available to us contained a variety of objects such as branches, leaves, soil, and grass clippings. We've separated yard waste into these categories to optimize our model's performance. We identified that the highest occurrence amongst these categories to be leaves. Therefore, the first augmented dataset will consist of leaves.

Over the course of several weeks of testing and development we added features that allow original objects and artificial objects to be layered randomly in the resulting image. This helps simulate real-world scenarios where objects are dumped on top of each other in random orientations. This required a number of functions to be developed that handle layering and clipping of segmentation masks, tests for edge cases and automatic reprocessing of rejected augmentations.

When visualizing the annotations, you can see that the artificially placed leaves on the right clip sections of the segmentation mask on the black garbage bag. This is significant as overlapping segmentation masks may not describe objects accurately which are layered.

We also added a number of functions that handle data formatting and conversions. As well as exporting the data back to the same format in which it was received. Our method results in a directory of augmented images and an annotations file.

Waste Management Dashboard

We conducted user testing on the dashboard with our industry partner and gained an understanding of what is valuable to the user. We conducted this with our industry partner. We would have preferred to test this with Waste Management and Municipality workers who are our north star customers. However, this requires coordination with a third-party and we were unable to reach these groups of users. But we are confident with the testing performed as our industry partner has had several conversations with gathered feedback from our north star customer.

With user testing our goal is to evaluate the utility, usability, and desirability of the dashboard from the user's perspective.

And to do this, our strategy is to:

1. First provide some background and establish the dashboard's purpose
2. Then complete a list of tasks which are designed to simulate the experiences a user will have as they perform their regular checks with the dashboard.
3. Then, finally they get experience the dashboard at their own accord to provide additional feedback

Some of our key findings from this were:

- Explicitly stating why information is coloured with green, yellow, and red
- Adding Legends to the Map displayed on the collection day pages
- Users are interested in seeing contamination rates across the neighbourhoods on the map. This could be showcased using a range of colour, where rates are high the colour should be darker.

Finally, we concluded that the useful, and desirable, however, there needs to be further changes to improve its usability

Knowledge Management Overview

We have continued to perform weekly agile sprints, while managing our kanban board and recording meeting minutes to ensure we are progressing effectively.

Additionally, Mitacs approved our project and allowing us to utilize the allocated research funds for costs such as annotations and AWS credits.

Next up

Team

- Refactoring and Code Reviews
- Large scale data ingestion from Industry Partner's dataset
- Finalize project documentations, poster, and commercial

Rishabh

- Continue adding logging metrics functionality
- Assist with scaling up the augmented training dataset
- Update importing format from s3 buckets as required by the training pipeline
- Implement / update experiences on the dashboard as requested by users

Nolan

- Training Mask R CNN models using augmented datasets
- Generate performance metrics for resulting Models

- Continue integration with industry partner and set up automated retraining of models
- Code reviews and documentation.

Will

- Scaling up the augmented dataset
- Ingesting the audited contaminant dataset for augmentation.
- Exporting augmented image datasets to s3

Team Reflection

- **Does the team feel "on track"? (reiterate the above colour status)**
 - We believe we are still on track
- **What progress does the team particularly feel good (great) about?**
 - We feel really great about completing the Bin Tip detector, the project has been transferred to our industry partner.
 - Our progress on our copy-paste implementation has progressed rapidly. We have produced small scale augmented datasets and have started producing large scale datasets. We are ready to start training models using our augmented datasets.
- **What barriers (if any) does the team feel is a current impediment to success?**
 - Potential for unexpected problems with model training.
- **What help (if any) does the team require to move positively forward?**
 - Guidance and resources from our industry partner regarding Sagemaker and Mask R CNN training.
- **What questions or concerns does the team have (if any)?**
 - We are concerned about potential problems with model training and performance. We know there are a number of unknown factors related to experimental/ cutting edge software development and we have created procedures to mitigate certain risks. Our project is reaching a critical point where all of our efforts over the past 6 months are coming together and justified.