# Searching for Waste with Computer Vision

* Hello my name is Julian Hernandez, and I am a computer science student at California State University, Sacramento. Today I am going to show you how we can use computer vision to help find uncollected litter.

# The Trash Problem

* And there’s plenty of it to be found, in 2012 the world bank estimated that municipal solid waste production is doubling roughly every decade.
* 270 species harmed or killed
* Electronic waste has been found to leak toxic heavy metals into the environment
* $200 billion each year globally
* 5% of GHG emissions are from waste management
  + While it’s not a lot compared to transportation or electricity, it will need to be decreased to reach our climate targets

# **Literature Review –** Previous research has focused on two areas…

* Two areas:
  + Sorting waste and making robots that can go clean up trash
* Sorting only is good for waste that makes it into the trash bin, and does nothing to help clean up litter
* And while robotics have made shown a lot of promise, they will need a lot more improvement before they’re ready to go out and clean on their own. Additionally, they will likely need to be proven to function at scale in other industries before waste management can make use of clean up robots.

# **Goal of this research –** With this in mind, we are seeking to answer two questions

* Where is waste located?
* How much there is?
* For the first we use image metadata such as coordinates, and camera orientation to estimate each item’s location
* For the second question, we use something called recall to estimate how much total waste is out there is based on our detections

# Trash Annotation in Context Dataset – For our input images we used…

* TACO is a standard format
* Free license
* 1500 official 1700 unofficial
* Each picture has a mask and label with 60 categories
* Context in hand vs on the ground
* We can annotate new images, which will be important later when we want to improve accuracy

# Taco Annotations

* There’s decent amount of annotations, but compared to

# Mapillary **-** For our first output dataset we are using

* Dash cam images converted
* Personally info scrubbed
* Can scan a state or city
* Traditional methods could take weeks

# Clean up images**–** For our second dataset we are using images from along the American River parkway.

* 150 images taken
* These were collected during a cleanup and taken using a standard smartphone camera
* Similar to what a volunteer organization could take

# Mask Regional Convolutional Neural Network – For this we are using an algorithm called…

* Mask R-CNN is a neural network architecture that was made open source by Facebook in 2018
  + This algorithm can learn to detect any type of object if it has enough sample input images to learn from
* In the past it has been used by humanitarian organizations to quickly map out buildings
* And in the medical field to automate nuclei detection in cells
* What we want to do
  + Feed it a set of images to teach it how to detect litter
  + Then run the program on two different datasets, one to answer each of our questions

# Feature Extraction – The way it learns

* Original images are given to the program with outlined objects such as this picture with some trash in a river
* First it learns features such as size, shape, and color that it can use from the outlined area to distinguish waste from the background
* Next it finds what features it can use from the background to accurately reject false predictions
  + As we will see later, it can struggle with background objects that it hasn’t seen before and that resemble trash

# **Metrics –** To measure how well our model performs we get a summed total for four categories

* **True positives** are accurate detections
* **True negatives** which are background space
* **False positives** are objects that are mislabeled as waste
* **False negatives** is litter that are missed

We use these summed totals to calculate:

* **Precision** tells us how many of our detections were correct
* **Recall** tells us how much of the total waste we were able to find

# **Results -** While Mask RCNN failed to accurately detect waste from vehicle mounted images, it worked very well with our sac state cleanup dataset

* 3 Likely causes for poor performance:
  1. TACO only has smartphone pics so learned features don’t translate to a new perspective
  2. Because they’re from a car there’s less pixels to work with and detections will be noisy
  3. And most importantly we found many false positives from new unfamiliar backgrounds

# **Misclassification –** Specifically we found that cars were routinely mislabeled are soda cans with pop tabs for their wheels

* We can see what features both share such as metallic texture, bright coloring, and rounded edges
* To improve this more images with cars in the background will need to be uploaded to the TACO dataset, so it learn what features are needed to distinguish between a can and a car

# **Future Work –**

# **Conclusion –** In conclusion,

* While the technology isn’t far enough along to search for new areas to clean up, but it can be used to measure the impact of cleanup activities and provide insight into what’s being cleaned up
* In the future, as the TACO dataset increases in size and new images are uploaded from a vehicles perspective we expect the accuracy to increase and new possibilities to be come to light
* Because global waste production isn’t predicted to plateau until the end of the century, new technologies are needed that can help us to cleanup waste at the rate that we are littering it.

# Questions

* Thank you very much are there any questions?

**Searching for Waste with Computer Vision**

Hello, my name is Julian Hernandez, and I am a computer science student at California State University, Sacramento. Today I am going to show you how we can use computer vision to help find uncollected litter. And there’s plenty of it to be found, in 2012 the world bank estimated that global waste production is doubling roughly every decade. This has a wide range of negative effects. Plastic alone has been found to harm or kill over 270 different species of animals and electronic waste, when not recycled properly, can leak toxic heavy metals into the ground and water system. Waste management costs over $200 billion each year and accounts for about 5% of our total greenhouse gas emissions. Which while this doesn’t seem like a lot compared to transportation or electricity generation, it will need to be decreased to reach our climate targets.

**Previous Research**

Previous research has focused on two areas: using computer vision to help automatically sort waste and building robots that can clean it up. While better sorting methods are needed this doesn’t help with waste that doesn’t make it into the trash bin. Additionally, robotics needs to make a lot more progress before they’re ready to go out and start cleaning on their own.

**Research Questions**

With this in mind, we are seeking to answer two questions: Where is waste located? And how much there is? To answer the first question, we are using image metadata such as coordinates, and camera orientation to estimate each item’s location. Then we save this data to a geojson file and calculate the summed totals for each category of object and plot it on a map. This data could help researchers quickly identify trends of where waste is being left and how that changes over time.

**Mask RCNN**

The algorithm we chose to use is called Mask Regional Convolutional Neural Network or (Mask RCNN) which was made open source by Facebook in 2018. This can detect any type of object if it has enough sample input images to learn from. In the past it has been used to quickly map out buildings in disaster locations and to automatically detect nuclei in cells. What we want to do is teach it how to see litter then run it on two different collections of images to answer each of our questions.

**Feature Extraction**

The way it learns is using feature extraction. Images are given to the program with outlined objects such as this picture of waste in a creek bed. First it learns what features, such as size, shape, and color, that it can use to distinguish between different categories. Next it finds what background features it can use to accurately reject false predictions. As we will see later, this is important as it has no understanding of the real world except what is present in the training images.

**TACO**

We decided to use Trash Annotations in Context for our training dataset which provides images of waste in a standard format used for computer vision research. All the images are provided under free license and new ones can be added and annotated, allowing anyone to expand the dataset as needed overtime. It consists of 1500 images and almost 5000 objects that have been individually outlined. There are also an additional 1700 images that have been annotated and can be used but not aren’t officially validated by the creators yet. Every picture contains a mask outlining each piece of waste with a label that places it into one of 60 categories. Additionally, context is added so that it that can understand the difference between a plastic bottle in someone’s and a waste bottle on the street.

**Mapillary**

For our first output dataset we chose to use images provided by Mapillary. Mapillary converts community submitted dash cam footage into a collection of images covering the road network and scrubs any personally identifiable information such as faces and license plates. The hope here is that if detection is accurate enough, Mask RCNN could quickly scan whole cities or states for trash. This is something that is nearly infeasible using traditional searching methods, as it would take many people possibly months to cover the same area.

**Cleanup dataset**

The second output dataset consists of about 150 hand collected images of waste that were cleaned up at Sacramento State.

The photos were taken using a standard smartphone camera before waste was discarded, in total we removed a full standard trash bag’s worth of waste, collecting images all along the way.

This dataset is very similar to what could be collected by a volunteer organization during their cleanup time to estimate the total waste that was picked up.

**Metrics**

To measure how well our model performs we get a summed total for four categories: true positives which are accurate detections, true negatives are backspace space, false positives are objects that have been mislabeled as waste, and false negatives which are litter that has been missed. Then we use these summed totals to calculate precision and recall. Precision tell us how likely a given detection is to be correct, and recall tells us what percentage of the total waste was found. Previous research usually can achieve a score of around 80% for each of these measurements.

**Results**

Discrepancy between vehicle images and phone images

While Mask RCNN failed to accurately detect waste from Mapillary images, we found that it worked very well with handheld smartphone images. There’s three likely causes for this discrepancy between our two datasets: 1 TACO only contains images taken from smartphones so it may not be able to translate learned features to new perspectives. 2 because the Mapillary images were taken from a car they are going to be inherently further away, and a lower resolution mean that our detections are going to be a lot more noisy. And lastly, we found many false positives from new unfamiliar objects in the background.

**Mislabeling**

Specifically, we found that cars were routinely mislabeled as soda cans with pop tabs for tires, making up over half of our false positive detections.

We can infer what features it might be getting confused on, they both have metallic texturing, bright colors, and rounded edges.

And because the training dataset contains almost no images with cars in the background, it would have no understanding of what a car is and how to distinguish it from a can.

This one issue was the largest cause of false positives, and fixing it would vastly increase the precision of our model, making it a much more viable option for scanning images from Mapillary.

**Annotations and future work**

**Next step background foreground, understand features**

Our next step is to add new images to TACO with cars in the background and litter in the foreground which it allow it understand what features are needed to distinguish between a car and a can.

**Using the pubic annotator**

We will be using the publicly available annotator at tacodataset.org that allows you to outline new images of litter.

**Anyone can use it**

This is something that anyone here who is interested in helping improve this technology could use without any specialized knowledge or experience.

**Will help accuracy and surrounding environment**

This will help improve trash detection accuracy, and open new possibilities for making our surrounding environment cleaner.

**Conclusion**

In conclusion, while the technology isn’t far enough along to search for new areas to clean up, it can be used to measure the impact of cleanup activities and provide insight to into geographic trends.

In the future, as the TACO dataset increases in size, we expect accuracy to improve especially in unfamiliar scenes.

Lastly, because global waste production isn’t predicted to plateau until the end of the century, new technologies such as the one I’ve shown you today, are needed to help us to cleanup waste faster than we are littering it.

Thank you. Here is my poster, take a moment to review and I can take question.

Overview

**Searching for Waste with Computer Vision**

**Previous research**

**Research question**

**Mask RCNN**

**TACO**

**Mapillary**

**Cleanup dataset**

**Metrics**

**Results**

**Mislabeling**

**Annotations and Future Work**

**Conclusion**