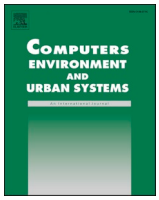
Computers, Environment and Urban Systems 93 (2022) 101752

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/ceus



Using a deep learning model to quantify trash accumulation for cleaner urban stormwater

Gary Conley [a,*], Stephanie Castle Zinn [b], Taylor Hanson [b], Krista McDonald [a], Nicole Beck [a], Howard Wen [b]

[a] 2NDNATURE, 500 Seabright Avenue, Santa Cruz, California 95062, USA
[b] Fuscoe Engineering, 16795 Von Karman, Suite 100, Irvine, California 92606, USA

ARTICLE INFO

Keywords:
Urban trash
Litter
Stormwater
Machine learning
Mask R-CNN

ABSTRACT

With growing understanding of trash impacts on aquatic habitats throughout the world, cities increasingly face regulatory requirements to reduce trash inputs to local waterways and the ocean, but they often rely upon insufficient monitoring data to prioritize and measure trash reduction effectiveness. We present an approach designed to make urban trash monitoring more cost-efficient and align the data collected with critical information needs of cities. We quantified urban trash accumulation along roadsides using vehicle mounted cameras and a deep convolutional neural network model to identify trash in the imagery captured. We compared the trash detection performance of three different models, with the best performing model (Mask R-CNN) achieving 91% recall, 83% precision, and 77% accuracy using data collected along 84 road segments in two California Cities. Trash detection model outputs were interpreted via a statistical model to relate the proportion of image pixels identified as trash to measured trash volumes. The resulting model estimates explained 67% of the variance in measured trash volumes collected on roadsides, which is more than double the variance explained by walking visual assessments. With vastly more efficient data collection compared to the visual assessments, deep learning-based monitoring approaches can provide a stronger basis for understanding urban trash sources, changes over time, and cost-effective compliance with stormwater regulatory requirements.

1. Introduction

Trash transported from city streets (urban litter) via stormwater systems contributes to the degradation of aquatic habitats (Hoellein, Rojas, Pink, Gasior, & Kelly, 2014; Sigler, 2014) and the persistent and expansive accumulations of trash gyres such as the Great Pacific Garbage Patch (Dautel, 2009). In response to this growing environmental threat, several communities throughout the United States, including the State of California (State Water Resources Control Board (SWRCB), 2015), City of New York (New York State Department of Environmental Conservation (NYSDEC), 2015), City of Los Angeles (State Water Resources Control Board (SWRCB), 2015), San Francisco Bay Area (San Francisco Regional Water Quality Control Board (SFRWQCB), 2015), and City and County of Honolulu (Hawaii Department of Health, 2012), have implemented water quality regulations aimed at reducing trash delivery to aquatic systems. To meet compliance with regulatory requirements, stormwater programs typically must demonstrate that trash is either being captured in devices serving the stormwater drainage system, or it is not accumulated on streets and available for transport to municipal stormwater outfalls to receiving waters (State Water Resources Control Board (SWRCB), 2015). While several field protocols have been developed to estimate trash accumulation on city streets (e.g., US Environmental Protection Agency (EPA), 2021; Bay Area Stormwater Management Agencies Association (BASMAA), 2014), they are time consuming to conduct at scale, and these shortcomings are reflected in the severely limited coverage and sampling frequency requirements (State Water Resources Control Board (SWRCB), 2017a, 2017b). Recent work in California indicates that prescribed minimum levels of monitoring effort based on an initial analysis by Bay Area Stormwater Management Agencies Association (BASMAA) (2016) may be inadequate for characterizing trash conditions with high levels of certainty or detecting changes over time (Conley, Beck, Riihimaki, & Hoke, 2019). The spatial and temporal variances of trash accumulation have been previously identified as

* Corresponding author.
E-mail address: gary@2ndnaturewater.com (G. Conley).

https://doi.org/10.1016/j.compenvurbsys.2021.101752
Received 26 June 2021; Received in revised form 22 December 2021; Accepted 23 December 2021
Available online 15 January 2022

potential confounding factors for detecting patterns and changes over time in trash monitoring designs (Wheeler & Knight, 2017). These challenges highlight the benefits of efficient monitoring approaches that allow for more frequent observations, with greater spatial density and coverage where needed.

In addition to helping cities more cost effectively meet regulatory requirements, an efficient and reliable means of measuring urban trash accumulation can provide better information for quantifying trash impacts on communities (Muñoz-Cadena, Lina-Manjarrez, Estrada, & Ramon-Gallegos, 2012), determining trash mitigation effectiveness (Marais & Armitage, 2004; Marais, Armitage, & Wise, 2004), and guide real-time adaptive responses (e.g., Hossain et al., 2019). As cities continue to get smarter, they improve their capacity for more efficient data capture and incorporation of rapidly updated data streams to augment decision making, improve resiliency, and promote citizen well-being (Toli & Murtagh, 2020). Central to this development is the digitization of information flows across spatial and temporal scales to advise municipal staff where problems emerge to allow for quick and effective responses (Batty et al., 2012). Environmentally targeted technologies and monitoring can help smart cities to become sustainable cities (Ahvenniemi, Huovila, Pinto-Seppä, & Airaksinen, 2017), given that a key reason for making a city "smart" is to mitigate the problems generated by the urban population growth and rapid urbanization (Chourabi et al., 2012). This evolution of cities has traditionally centered on technologies such as Internet of Things (IoT), and information communication technology (Silva et al., 2018), but has recently begun to incorporate the use of big data and artificial intelligence (AI) to inform functions like operations and maintenance of urban roads (e.g., Yu et al., 2021). While the combination of drive-by sensing with big spatiotemporal data analytics and AI appears to have great potential for characterizing urban environments (Anjomshoaa, Santi, Duarte, & Ratti, 2020; Gunturi & Shekhar, 2017), image-capture based applications have been underutilized in this emerging research frontier (Yang, Clarke, Shekhar, & Tao, 2020).

Artificial intelligence (AI) and machine learning technologies in environmental monitoring can be particularly useful where there are limited resources for characterizing systems, minimizing risks, and prioritizing actions (Ghannam & Techtmann, 2021; Hino, Benami, & Brooks, 2018; Pyayt, Mokhov, Lang, Krzhizhanovskaya, & Meijer, 2011). Environmental applications often involve image capture and object identification via feature annotation (Zurowietz, Langenkämper, Hosking, Ruhl, & Nattkemper, 2018), with the accuracy of computer vision technology sometimes eclipsing the performance of human experts (He, Zhang, Ren, & Sun, 2015). Image-based classification models can support a wide range of environmental decision support, and growth of AI applications for this purpose has been fueled by rapid improvements to low-cost environmental sensor networks (Okafor, Alghorani, & Delaney, 2020) and cloud-based machine learning platforms and libraries (Roy et al., 2019). These advances make it easier than ever for novice users across a range of fields to apply AI technologies for cost-effective measurements of systems with the potential to provide more data and better insights that can lead to improved environmental outcomes (e.g., Schermer & Hogeweg, 2018).

The problem of trash identification and classification has been recently addressed with deep learning, using Convolutional Neural Networks (CNN) with applications for classifying and sorting trash at various life cycle points (Salimi, Dewantara, & Wibowo, 2018: Tiyajamorn, Lorprasertkul, Assabumrungrat, Poomarin, & Chancharoen, 2019; Adedeji & Wang, 2019). A CNN is a class of deep neural network that resembles processing patterns of the brain with several layers of filters that assess attributes such as shapes, colors, and edge detection to summarize recognition of features in an input image. Most applications reported in the literature are targeted towards trash sorting problems, rather than identification of trash as litter in urban environments. Notable exceptions are the models reported by Mittal, Yagnik, Garg, and Krishnan (2016), who created a CNN-based smartphone app called SpotGarbage, and the work of De Carolis, Ladogana, and Macchiarulo (2020), who used annotated Google Streetview imagery to train a CNN model to detect trash in video streams. Applications of image-based trash classification over wide geographical areas have been limited to primarily aerial imagery captured by drones (Deidun, Gauci, Lagorio, & Galgani, 2018; Hengstmann & Fischer, 2020; Kraft, Piechocki, Ptak, & Walas, 2021). Such applications are well suited to shorelines, river channels, or coastal herbaceous wetlands (e.g., Moore, Hale, Weisberg, Flores, & Kauhanen, 2020; Tharani, Amin, Maaz, & Taj, 2020), where a bird's eye view often provides a good perspective with few obstructions for identifying trash. This perspective can be more constraining in urban environments, where structures, cars, and trees often obstruct the view, particularly near curbsides, where trash tends to collect. In addition, many cities have restrictions on allowed areas and flight altitudes of drones. For these reasons, vehicle mounted cameras may provide a more practical method for capturing imagery to quantify trash accumulation on city streets.

A more efficient method for trash data collection has the potential to dramatically reduce ongoing municipal monitoring costs for regulatory compliance, while also providing data that are more amenable to tracking progress over time and detecting spatial patterns. In this study we present an assessment of a newly developed approach for urban trash monitoring that relies on vehicle mounted cameras for image capture, a deep learning trash detection model for identifying trash, and a regression model for estimating trash volumes from the trash detection model outputs. The objective of this study was to determine whether automated image-based trash monitoring could provide comparable information to human visual surveys for characterizing trash conditions on city streets. To that end, we compared the performance of three deep CNN models to determine the most appropriate one for our application, but focus primarily on comparison of model outputs with field data measurements to quantify trash volumes, rather than on performance differences between models.

2. Methods and data

2.1. Study sites and overview

Urban trash data were collected in the City of Salinas (Monterey County, California; population = 156,000) and Anaheim (Orange County, California; population = 349,000) during the summer and fall of 2020 (Fig. 1). These two cities were selected based on geographic proximity, familiarity with their trash mitigation programs and regulatory requirements, and representation of two distinct climatic regimes in California. Both cities show intensive urban development over most of their geographic extent and approximately half of the area for both cities are categorized as either 'disadvantaged' or 'severely disadvantaged' based on reported household incomes. (US Census Data (ACS: 2012–2016 and ACS: 2014-2018), 2018). Traffic conditions in Salinas made it easier to spread surveyed segments widely throughout the city, while the Anaheim segments where concentrated along a few corridors of the city, where each side of the same road were assessed separately and treated as unique observations.

The study workflow is presented in Fig. 2, consisting of three primary steps: collecting trash field observations as a basis for model training and comparison, training and evaluation trash detection models, and specification of a statistical model for relating the machine learning model outputs to trash volumes. In the sections that follow, we present the candidate trash detection model structures, performance testing metrics, the regression modeling approach, and workflows for data collected in Salinas and Anaheim.

2.2. Trash data collection

For each road segment, three different trash data types were collected by two trash surveyors: visual assessments, trash volume

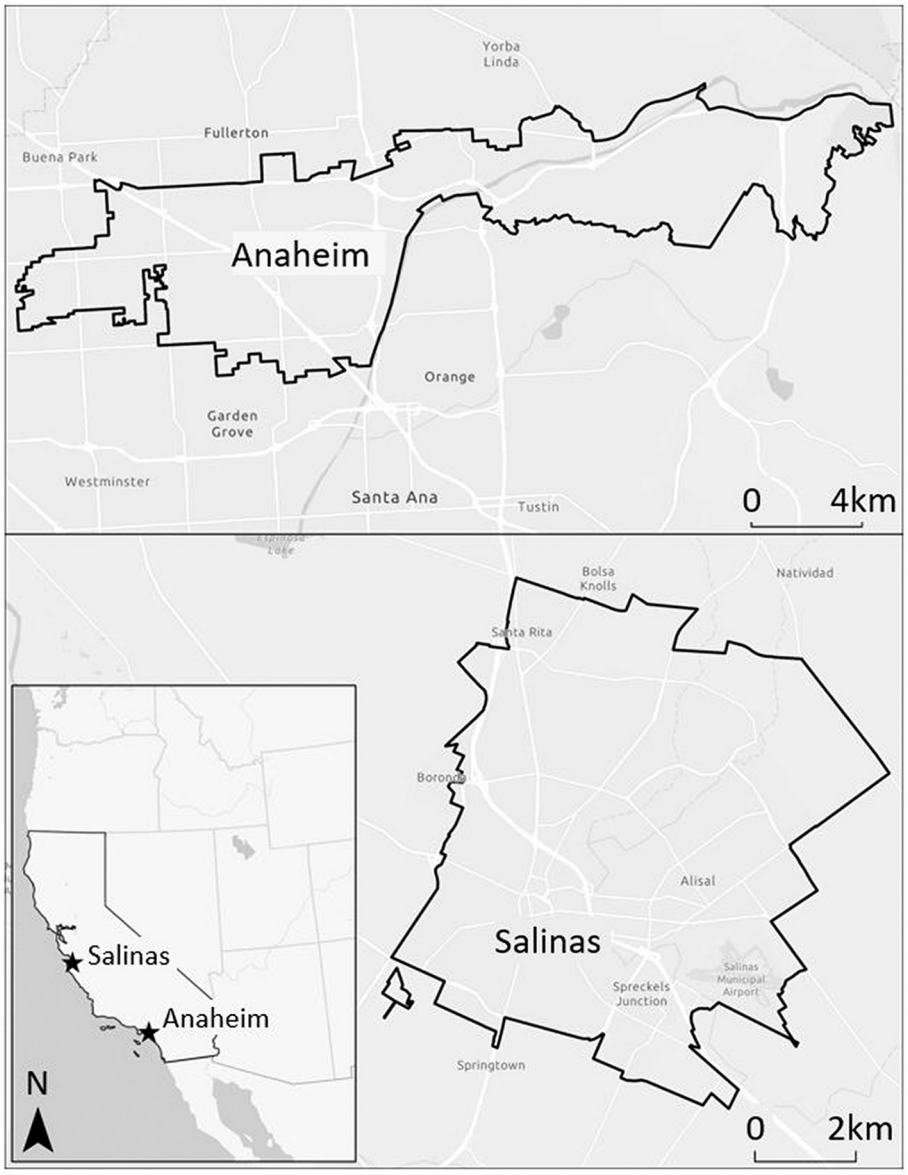


Fig. 1. Study locations in the cities of Salinas and Anaheim, California, USA as defined by their stormwater regulatory boundaries.

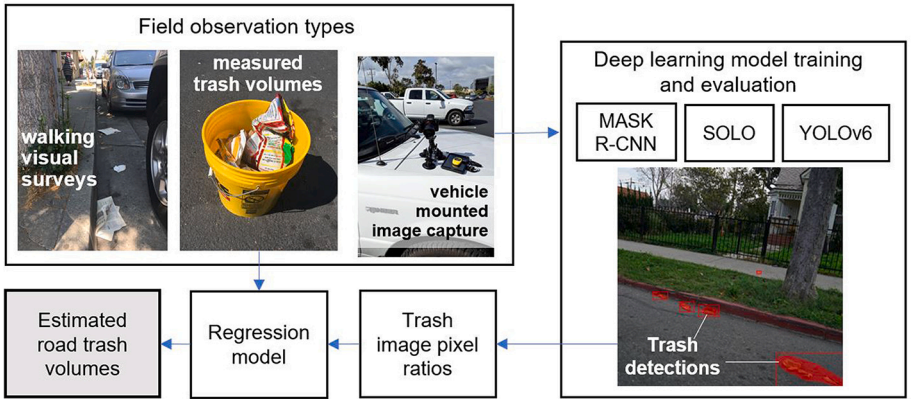


Fig. 2. Study workflow block diagram with field observation types, deep learning models. and translation of model outputs to trash volumes. The models evaluated: Mask R-CNN, SOLO, and YOLOv6, along with each component shown in the diagram are described in the methods sections that follow.

measurements, and roadside imagery. First, the surveyors drove the road segment to capture a set of images with the vehicle mounted camera. Next, they parked the vehicle and walked the segment to visually interpret the level of trash accumulation. Finally, the trash surveyors collected the trash along the road segment using a graduated bucket to measure the total trash volume. Data collection areas included the road shoulder, sidewalk, and any areas immediately adjacent to the sidewalk that was within the camera's field of view. To increase the likelihood of trash accumulation in the streets, surveys were conducted one-two days before street sweeping occurred and no sooner than three days following a storm event. A total of 14 km of road length was surveyed, with 40 segments in Salinas and 44 segments in Anaheim. Individual segment lengths varied from approximately 180 m to 380 m long as broken by city block intersections. Road segments were chosen based on access to the roadside, lack of obstruction by cars, prior knowledge of trash accumulation, and the ability to safely drive slowly enough to capture usable imagery.

2.2.1. Visual trash assessments

Visual assessments of trash along roadways are a qualitative method to measure the accumulated trash available for transport into the storm drain system within a specific area. These assessments are the primary way that California municipalities demonstrate compliance with regulatory requirements per local National Pollutant Discharge Elimination System (NPDES) Permits for areas that are not served by devices to fully capture trash delivered from roadways (State Water Resources Control Board (SWRCB), 2015). For this study, we employed a previously developed visual assessment field protocol, termed the On-Land Visual Trash Assessment (OVTA) (Bay Area Stormwater Management Agencies Association (BASMAA), 2014), which has been accepted by the California State Water Resources Control Board (SWRCB) to comply with water quality permit requirements to reduce trash inputs to storm drain systems (State Water Resources Control Board (SWRCB), 2018) and recently integrated to the US Environmental Protection Agency's Escaped Trash Assessment Protocol (US Environmental Protection Agency (EPA), 2021). The OVTA protocol is similar to the qualitative elements of protocols previously developed by the California Surface Water Ambient monitoring program for evaluating riparian corridors (Moore, Cover, & Senter, 2007), both of which have shown empirical association with downstream trash accumulation. Per the OVTA protocol, road segments are assigned one of four trash condition categories (low, moderate, high, very high) based on observed trash accumulation on the road shoulder, gutter, and sidewalk. Category assignment is dictated by comparison of the observed trash with a set of reference images and narrative descriptions of trash abundance associated with each of the trash condition categories. Data quality was insured with audits of the assessments by field personnel who had been trained in the OVTA field protocol and performed several hundred visual trash assessments prior to initiation of the current study.

2.2.2. Trash volume measurements

As a direct measure of trash accumulation, we collected all the observable trash on each road segment and measured the trash volumes with graduated buckets. Since this study is concerned primarily with mobile trash that can be transported by wind or water into local storm drains, large items such as furniture or appliances were not included as trash in the volume measurements. This exclusion criterion also aligns with the BASMAA OVTA visual assessment protocol. Trash was loaded into the graduated buckets, moderately compressed, and volume recorded according to the fill level and number of buckets needed to collect all trash along the segment. Four of the road segments in Salinas had areas of trash accumulation that were visible from the roadside but were not able accessible for collection. In these instances, we estimated the trash volumes visually.

2.2.3. Trash imagery

Imagery was captured using a 25-megapixel digital camera mounted to the hood of a vehicle at a height of 1.3 m, and driving at constant speed of approximately 30 km/h. The camera was pointed at the roadside perpendicular to the vehicle. Geospatial data for the segments assessed was recorded via a phone app (ArcGIS Collector) and a GPS device synced to the camera. The focal length of the camera lens was 28 mm, and the shutter speed was set at 1/500th of a second to minimize motion blur but still allow an adequate level of sensor sensitivity (ISO) in shadowy areas. The capture rate was set at one image per second, which provided minimal overlap or gaps between the images captured with the vehicle moving at 30 km/h. The distance maintained from the curb was approximately three meters, and care was taken to maintain the same distance from the curbside during image capture so that the elements within the field of view was similar for all road segments.

2.3. Trash detection models

We compared three CNN-based computer vision models: Mask R-CNN, SOLO, and YOLOv6, which are each described in the following sections. From this comparison, we selected the model with best trash detection performance for use in the subsequent steps to quantify trash volumes in a regression model with the field measured trash volume data. Below, we provide a brief description of each model tested along with conceptual schematics depicting the simplified architecture of each (Fig. 3).

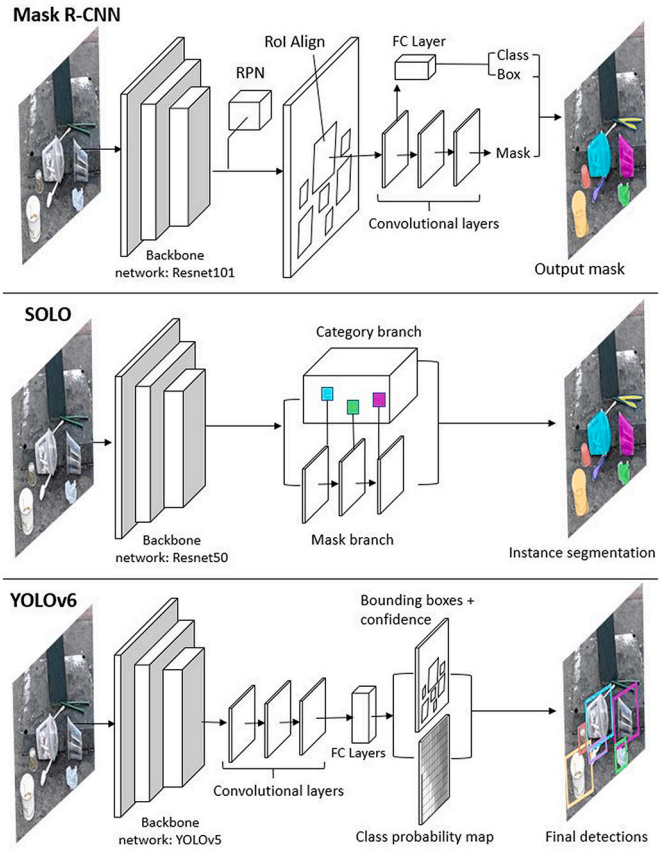


Fig. 3. Trash detection model schematics for Mask R-CNN (after He et al., 2017), SOLO (after Wang, Kong, et al., 2020, Wang, Zhang, et al., 2020), and YOLOv6 (after Redmon et al., 2016). Architecture components differ across models including backbone networks, convolutional layers, region proposal network (RPN), use of region of interest align (RoI), and construction of fully connected (FC) layers.

2.3.1. Mask R-CNN

Mask R-CNN (He, Gkioxari, Dollár, & Girshick, 2017) is an extension of the Fast/Faster R-CNN approaches (Girshick, 2015; Ren, He, Girshick, & Sun, 2015), and uses a set of filters (feature detectors) through multiple convolutions to output feature maps (convolutional layers) from an image (see Fig. 3). The Mask R-CNN architecture uses instance segmentation, which combines object detection with semantic segmentation to classify each pixel into a fixed set of categories, with differentiation of unique object instances (He et al., 2017). Object detection uses a fixed set of object categories and draws a bounding box each time an object appears in the image. Semantic segmentation assigns a class label to each image pixel. With the instance segmentation used in Mask R-CNN, instead of bounding boxes, the model identifies which pixels within the bounding boxes belong to each object. In contrast to semantic segmentation, a separate mask is drawn for each instance detected in the image. It is distinguished from previous efforts by the parallel prediction of masks and class labels, wherein segmentation masks are predicted for each Region of Interest (RoI) at the same time as classification bounding box regressions are performed (Fig. 3). RoI Align (He et al., 2017) is used for pooling the feature maps and building the fully connected layers from the convolutional layers. The Mask R-CNN approach allows pixel by pixel instance segmentation that fully preserves the exact spatial location of detected objects.

2.3.2. SOLO

Segmenting objects by locations (SOLO) takes a unique approach to instance segmentation, with reliance on the concept of 'instance categories' introduced by Wang et al. (2019) and a feature pyramid network (FPN) to distinguish instances with different object sizes (Lin et al., 2017). This approach assigns instance categories to pixels based on the location and size of the instance within the image, in effect changing the mask segmentation problem into a classification problem (Fig. 3). In this way, SOLO avoids to the multiple steps of identifying regions of interest and generating feature maps within each of those bounding boxes. Instance segmentation is performed via two subtasks: category prediction and instance mask generation across a uniform grid overlaid on the image. Wang et al. (2019) point out that this is a simpler, more direct approach compared to those which rely on first separating objects by drawing bounding boxes around them. A key assumption embedded within this approach is that most instances in the same image have either different locations or are of different sizes. Recent work has shown comparable accuracy of SOLO with Mask R-CNN using the COCO dataset (Wang et al., 2019) and recent enhancements have shown it outperforming several algorithms which rely upon regional proposal networks (Wang, Kong, Shen, Jiang, & Li, 2020; Wang, Zhang, Kong, Li, & Shen, 2020).

2.3.3. YOLOv6

You only look once (YOLO) is a rapidly evolving approach which can detect objects in real time, only needing the algorithm to propagate through an image one time (Redmon, Divvala, Girshick, & Farhadi, 2016). As such, a primary advantage of YOLO above other methods is speed for use in video. An image is divided up into grid cells and bounding boxes are predicted for each object in an image with each grid cell predicting objects within itself and calculating confidence scores. Each grid cell predicts class probabilities to estimate the class of an object (see Fig. 3). The algorithm has been iteratively improved since its inception with new versions emerging at a swift pace (Redmon & Farhadi, 2016; Farhadi & Redmon, 2018; Bochkovskiy, Wang, & Liao, 2020; Shao et al., 2021). The most recent addition, YOLOv5-v6, has been employed for this study. Object detection performance has been shown to be comparable to other CNN-based algorithms (Redmon et al., 2016), with both speed and accuracy improvements with progressive versions of the algorithm (Bochkovskiy et al., 2020).

2.3.4. Model training, calibration, and validation testing

The training procedure and imagery was the same for all three models. In all cases, we labeled trash in 1000 images collected in cities throughout Orange County, California to train the network for automatic trash identification, with the imagery split into training and validation datasets, 80% being used for training and 20% for validation testing. Training was performed over 60 epochs, with each epoch passing over all the training images, which included 400 steps in each epoch. Precision was quantified as the number of correct trash detections divided by total trash detections, which describes the ability of the model to correctly identify trash versus other objects within the image. Recall was calculated as the number of correct trash detections divided by the number of actual pieces of trash in the image and describes the model's trash detection sensitivity. Accuracy was calculated as the number of correct predictions divided by the total number of predictions. These metrics were calculated with manual identification of trash in each image for comparison against the model predictions.

Since the primary purpose of the model is to quantify the amounts of trash available for transport into the stormwater system, the models were trained for trash detection (presence/absence) rather than specific object type identification (e.g., cup, straw, etc.) as would be done in a classification problem. Thus, objects identified as trash included very small items such as cigarette butts, medium sized ones such as soda bottles and cans, and large items such as cardboard boxes. Using the broad object category of 'trash' reduced the number of training iterations and the amount of image annotation that would have otherwise been required for more detailed object identification. A total of 22 training runs were completed to arrive at the final model hyperparameter values which achieved acceptable precision and recall performance on the training data.

While there are as many as 50 hyperparameters in the models, only four were changed during model training. We modified learning rate (how much the model changes the weights at each update), the rpn (regional proposal network), anchor size (size of the boxes the region proposal network chooses), and the training schedule (which layers are trained). Training schedule options included: all layers, Resnet stage 3+, Resnet stage 4+, Resnet stage 3+, and heads (RPN, classifier, and mask heads of the network). For the model backbones, Mask R-CNN used resnet101, SOLO used resnet50, and YOLOv6 used YOLOv5.

2.4. Quantifying trash volumes

Outputs from the trained models were processed to quantify number of pixels identified as the class 'trash'. For each image, the pixels constituting objects detected as trash were summed to determine the ratio of trash pixels to non-trash pixels in the image. Since objects closer to the camera occupy more pixel space, we adjusted the image trash pixel ratio (R) based on distance of objects from the camera using the y coordinate at the center of each mask (Eq. (1)),

$$R = \frac{y}{h}\, 4\, p \tag{1}$$

where y is vertical location of the mask in the image, h is the height of the image, and p is the total number of pixels detected as trash in the image. A constant multiplier of four is used based on initial trials which indicated that given the 28-mm lens field of view, and our distance from the roadside, the same object at the bottom of the image took up approximately four-times more pixels compared to one at the top of the image. With this formulation, trash object pixel ratios are normalized relative to the pixel area they would occupy if they were all at the front (bottom) of the image. Images that yielded trash pixel ratios greater than 0.5% were reviewed to determine if non-trash objects had been incorrectly identified as trash, which occurred in approximately 12% of images. These images were manually annotated before additional training runs of the model were performed.

For each road segment observed, we averaged the trash pixel ratio

outputs from all images captured within each segment. Since road segments varied in length, measured trash volumes were normalized to volumes collected per 200 m of roadway. The measured trash volumes were used as the independent variable in a linear regression model to estimate trash volumes from the trash pixel ratio outputs for each road segment. Model coefficients were estimated via ordinary least squares and regression residuals were tested for normality (Anderson-Darling test). We also compared results from the visual surveys that used the OVTA trash accumulation categories to both the measured volumes and the trash detection model outputs via a one-way analysis of variance (ANOVA).

3. Results

3.1. Trash data collection

A total of 84 road segments were assessed in Salinas and Anaheim, with measured trash volumes, visual assessments (OVTA), and photographic imagery collected for each city (approximately 3800 images total). The road segment locations and measured trash volumes are shown for each city in Fig. 4. Data collection was focused mostly along two corridors in Anaheim and was widely distributed throughout the City of Salinas. Average collection times for each data type per 200 m road segment were as follows: trash volume measurements (avg. = 42 min, SD = 16 min); visual assessments (avg. = 20 min, SD = 7 min); image capture (avg. = 21 s, SD = 11 s). These estimates only include the time taken to perform each assessment, so that they include parking and walking road segments for visual assessments. Notwithstanding time differences associated with setup and data management, they show that image capture is 57-fold less time consuming when compared to walking visual assessments. Trash volumes ranged widely across road segments in Salinas, spanning three orders of magnitude, resulting in a strong positive skew to the data distributions with greater observation frequency at the low end of the data range (see Fig. 4). Overall, the Anaheim roads had much less trash than Salinas, with approximately 20% of segments in Anaheim having almost no observable trash present. Though the trash volumes were low, the Anaheim data often showed substantially different levels of trash accumulation on opposite sides of the road. Salinas showed a much broader range of trash accumulation than Anaheim, with many clean and middling road segments, and several where hundreds of liters of trash had accumulated. The heaviest trash accumulation observed in Salinas was along a few road segments adjacent to industrial areas located in the southeast quadrant of the city (Fig. 4).

3.2. Trash detection model performance comparison

Results of the trash detection model performance are provided in Table 2 for training and validation image sets for each model. The Mask R-CNN model outperformed both SOLO and YOLOv6 in terms of recall, precision, and accuracy by substantial margins.

Calibration performance showed SOLO and YOLOv6 were within 3% of one another for all metrics on the training data, but the models diverged in validation testing. YOLOv6 achieved better precision and accuracy results than SOLO, but SOLO had the better validation recall performance. Compared to the closest competitor (SOLO), Mask R-CNN achieved accuracy values that were 21% better on the training data and 25% better on the validation data. Since Mask R-CNN also showed superior performance for recall and precision (Table 2), this model was selected for our application to quantify trash volumes and compare with the field-based approaches. For simplicity, we therefore focus further performance examination on the Mask R-CNN model exclusively.

For the selected Mask R-CNN trash detection model, training performance improved steadily throughout the training iterations to yield a result of 92% precision and 95% recall. Training progress relative to the overall loss function values are shown in Fig. 5 for both calibration and validation image datasets across all training epochs. The overall loss function for this model was an additive linear combination of loss functions that describe performance of individual components: the region proposal network (RPN) separation of background from objects, the RPN localization of objects, Mask R-CNN localization of objects, Mask R-CNN recognition each class of objects, and Mask R-CNN segmentation of

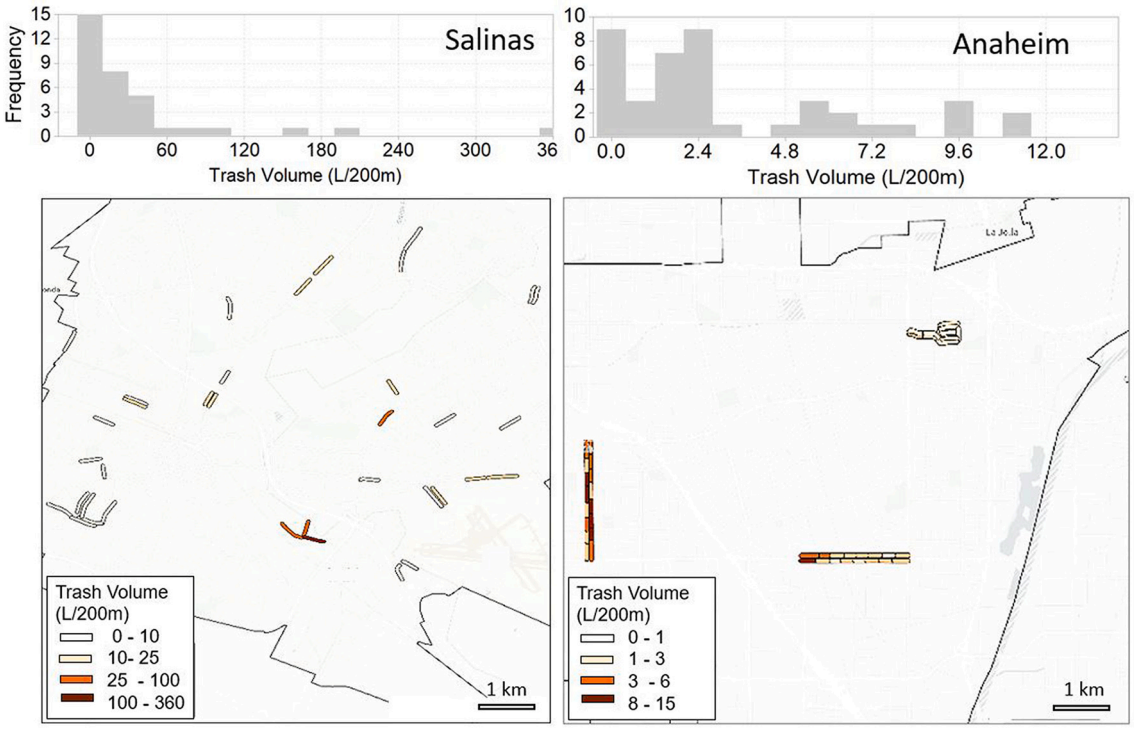


Fig. 4. Trash volume measurements for Salinas and Anaheim. Note difference in data scales for each city which reflects much less trash on the roads in Anaheim.

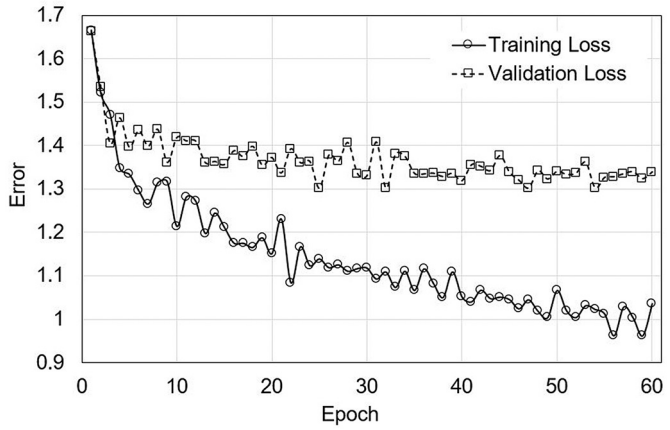


Fig. 5. Mask R-CNN training and validation loss for training epochs, each of which each included 400 steps.

objects. Leveling of the overall loss function slope in the second half of the validation data set indicates stabilization of the model's predictive ability across the 24,000 iterations through the entire dataset (60 epochs, 400 steps each). Validation testing showed a moderate performance decreased to 83% precision and 91% recall. The majority of false positives included vegetation or paint marks on a curb mistaken for trash by the model. Examples of model outputs are illustrated in Fig. 6, with successful trash detections along curbsides, a manhole cover incorrectly identified as trash (false positive detection), and several items embedded in vegetation that were not identified as trash (false negative detection). The manhole cover illustrates one of the few examples of a large object false detection, which can result in overestimation of trash volume accumulation. Although 17% of detections were false positives, most of these objects occupied less than 0.02% of the image pixels, so they had a minimal impact on the overall image pixel ratios and subsequently estimated trash volumes.

3.3. Comparison of the Mask R-CNN trash detection model outputs to visual assessments

Without exception, the segment averaged proportion of image pixels identified as trash was very small – always below 1% of the total image segment pixels, with individual images always below 5%. This provides a challenge for the trash detection model in that there is only a small amount of information in each image to quantify trash objects relative to the background image variation. The model output ratio of trash pixels to overall image pixels (trash pixel ratio) aligned approximately with the visual assessment categories, as shown in Fig. 7. There was, however, substantial overlap in the trash pixel ratios across the OVTA categories, particularly between the moderate and high categories. In terms of comparison with measured trash volumes, a one-way ANOVA showed that the trash survey (OVTA) scores explained 31% of the variance in measured trash volumes (p-value <0.01). This result partly reflects the way that the OVTA trash categories have been defined with wide ranges, especially at the high end of the scale where the 'very high' category has a range of 227–681 l of trash (see Bay Area Stormwater Management Agencies Association (BASMAA), 2014).

3.4. Estimating trash volumes from the trash detection model outputs

Comparison of the Mask R-CNN trash detection model outputs with the measured trash volumes is a direct measure of the model's ability to quantify trash on the road that is available for transport into the storm drain system, which corresponds with the volume-based metrics used for regulatory compliance (State Water Resources Control Board (SWRCB), 2015). We used a linear regression analysis to quantify the degree to which trash pixel ratios output from the Mask R-CNN model could explain the measured trash volumes, with each data point representing an individual road segment (Fig. 8). The model prediction interval spans several orders of magnitude, owing to the non-linear nature of the trash volume data in relationship to the two-dimensional trash pixel data. The model performance statistics and coefficients are listed in Table 1,

Fig. 6. Mask R-CNN trash detection model examples from the Anaheim imagery. Images 1. and 2. are accurate trash detections, 3. shows a false positive detection during initial training; and 4. shows false negative detections.

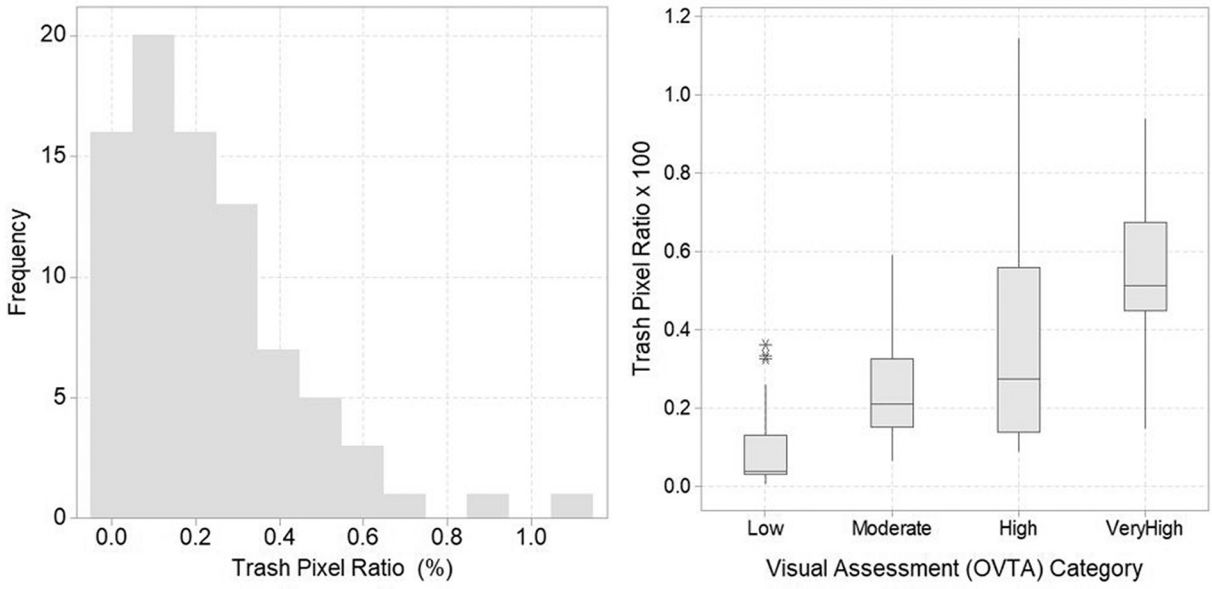


Fig. 7. Trash pixel ratio distribution and relationship between trash model pixel ratio and OVTA visual assessment categories with box plots of trash pixel ratios. Boxes represent the interquartile range of the trash pixel ratios, whiskers are the largest values that are not outliers, and stars are outliers.

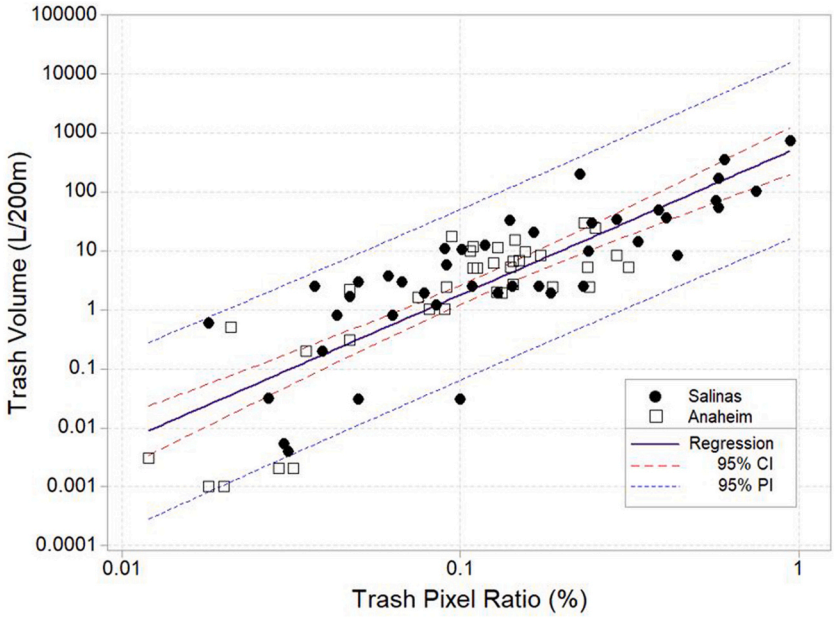


Fig. 8. Regression model to predict trash volumes in Salinas and Anaheim from trash pixel ratios ($n = 84$) with regression slope 95% confidence interval (red dashed lines) and prediction interval (blue dashed curves). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

showing a regression slope significant in exceedance of the 99% confidence level ($p < 0.001$), and approximately 67% of variance explained in the observed trash volume data across the road segments from both cities, with a standard error (S) value of 0.72 l. Model errors showed a normal distribution with constant variance across the range of the data. The regression analysis outputs show that the trash detection model outputs have moderate explanatory power for quantifying trash volumes. Performance of this statistical model would benefit from additional data collection in other cities with a better representation across the range of trash conditions since the current data spread shows most observations concentrated near the middle of the range.

4. Discussion

4.1. Trash model performance

Several factors combine to make trash identification in urban environments a challenging problem for object detection, largely due to variation of background settings, interaction between trash and other objects, and the low number of trash pixels relative to background pixels. Trash detection performance in this study (see Table 2) was comparable and in some cases superior to that achieved in similar applications in the urban environments. Using YOLOv3 De Carolis et al. (2020) showed an average recall performance of 69%, and a precision of

Table 1
Trash detection model parameter ranges, and final values selected during the model training.

		Final Values		
Model Parameter	Range	Mask R-CNN	SOLO	YOLOv6
base learning rate	0.01–0.0001	0.001	0.01	0.01
rpn anchor scales	4–256	16, 32, 64, 128, 256	4,8,16,32,64	N/A
training schedule		heads	heads	heads
backbone		Resnet101	Resnet50	YOLOv5

Table 2
Trash detection model performance comparison for Mask R-CNN, SOLO, and YOLOv6 on training and validation imagery sets.

		Recall	Precision	Accuracy
Training	Mask R-CNN	95%	92%	87%
	SOLO	78%	82%	66%
	YOLOv6	75%	80%	65%
Validation	Mask R-CNN	91%	83%	77%
	SOLO	81%	60%	52%
	YOLOv6	70%	79%	59%

68%. While our results using YOLOv6 were somewhat improved from that effort (70% recall, 79% precision), the Mask R-CNN results were substantially better at 91% recall, 83% precision. Mittal et al. (2016) reported 83% recall, 90% precision, with an accuracy of 88%, which is better than the accuracy reached by the Mask R-CNN on our dataset (77%). Both previous applications were concerned with real-time trash identification, where speed of operation is of greater importance than the current study. Overall, our results add to the evidence provided by these and other researchers that object detection models for in-situ trash identification can achieve good performance that is only slightly inferior to that achieved in trash sorting applications, which do not typically need to contend with such a wide array of object contexts. Melinte, Travediu, and Dumitriu (2020), for example, achieved 95.6% object detection accuracy with 97.6 precision using a Fast CNN model by classifying trash into material types. Like the present study, their model used feature extraction and box predictor modules for classification, but the training data were single items with uniform backgrounds. Similarly, while the model reported by Hossain et al. (2019) is intended for use on city streets, the accuracy of 96% was achieved using images that had uniform backgrounds much like the training images. We would expect better model performance in these cases compared to trash identification urban environments, where background and lighting conditions are variable; and trash is often clustered together or bound up with vegetation or other objects.

Another central challenge is that what we label 'trash' is a diverse array of objects ranging widely in size, shape, color, and texture. Initial trials regularly misidentified several non-trash objects as trash including electrical access covers on sidewalks, construction gravel bags, manholes and spray paint on sidewalks. Model performance is often limited by the number of different types of objects that can be manually annotated in for the training dataset, and additional training imagery to overcome similar model errors would almost certainly improve performance. Many applications reported in the literature rely upon pre-annotated imagery databases, such as TrashNet (Thung, 2020), which contains many thousands of single object images with uniform backgrounds. Further development of databases such as the Garbage in Images (GINI) dataset (Mittal et al., 2016), which contains imagery of in-situ trash in a wide variety of settings, should help to provide a rich training dataset for future application development and make the model training much more efficient.

Table 3
Regression model performance and coefficients.

Model	Coefficient	Std. Error	p-value	F	R2	Adj R2	S
Constant	2.759	0.72	<0.001	159.4	0.68	0.67	0.72
Pixel Ratio	2.503		<0.001				

4.2. Trash volume estimates

The trash CNN model outputs explained the majority of variance in trash volumes (Adj R2 = 0.67), but an important contributor to the unexplained variance is likely the limited information contained in 2-dimensional images for estimating trash volumes (See Table 3.). While we accounted for trash distance from the camera, the calculation was based on limited number of trials with only a few object types and could be refined with additional experimentation. The other important element for determining the proportional number of trash pixels in an image is the orientation of the object, but accounting for orientation in images would pose a greater challenge than characterizing object locations. An alternative approach that avoids these problems would be to identify specific trash objects as separate classes and estimate trash volume based on object type. However, this would likely require substantially more image annotations to achieve comparable performance and given the variance of object sizes of the same class, we currently have no intuition as to whether such an approach would improve the accuracy of trash volume estimates.

Conversion of trash pixel ratios to trash volumes brings the model outputs into alignment with stormwater management information needs: quantifying patterns of trash loading to stormwater systems, prioritizing problem areas, and measuring progress to satisfy regulatory requirements. The results of this study provide encouragement that with additional data collection the relationship between trash pixel ratios from the trash CNN model and measured trash accumulation volumes may be generalizable to other cities. With 67% of the observable trash variance explained by the trash detection model outputs, this more than double the predictive capacity of the BASMAA OVTA visual assessment method, which explained 31% of the observable trash volumes. In the development of the OVTA method, there was no consistent relationship found between visual assessment scores and trash accumulated on sidewalks Bay Area Stormwater Management Agencies Association (BASMAA) (2016), which may partly account for the relatively poor performance for predicting trash volumes.

4.3. Deep learning-based trash monitoring challenges and opportunities

The opportunity to bring deep learning tools to reduce the cost of monitoring and regulatory compliance for municipal stormwater applications is clear and additional research can help to surmount remaining logistical and technical barriers. The recent proliferation of deep learning tools, datasets, and trash identification algorithms can help cities bridge the gap between sustainability frameworks and smart technologies leading to 'smart sustainable cities' of the sort described by

Ahvenniemi et al. (2017). Cost trade-offs associated with different experimental monitoring designs should be explored determine approaches that can best inform management questions (e.g., Wheeler & Knight, 2017). The greater frequency and spatial coverage of observations afforded by imagery interpreted the trash detection model is likely to result in greater statistical power for detecting changes over time and spatial patterns (Conley et al., 2019), but this should be tested explicitly. Future research should also include testing additional trash detection models across a wider array of urban environments, including the impact of parked cars, which obstruct trash from view of the camera. Automation of the image capture and processing workflows will need to be developed to realize the full cost savings of machine learning based trash detection systems, which reflects the growing need for cities to use more data more effectively (Yang et al., 2020).

5. Conclusions

We have presented the results from a monitoring approach designed to make urban trash monitoring more cost-efficient and align the data collected with critical information needs of cities and water quality regulatory agencies. This approach relies on vehicle-based image capture, automated detection of trash via a deep learning trash detection model, and a statistical model to relate within-image trash pixel ratios to measured trash volumes. In a comparison of three deep learning-based object detection algorithms, Mask R-CNN emerged as the best performing alternative. When incorporated to a log-linear regression model, the outputs from the Mask R-CNN trash detection model (image trash pixel ratios) explained 67% of the variance in the trash collected on road segments. This is more than double the variance explained by visual survey methods currently accepted by the California regulatory agencies to support regulatory compliance tracking, with data collection that is nearly 60-fold more efficient.

An improved basis of information for decision making can help cities move towards more sustainable solutions to urban trash impacts on waterways. While the initial setup, model training, and data handling requires substantial time, such an investment may indeed prove worthwhile in the long term. A more cost-effective monitoring approach can facilitate more efficient regulatory compliance and also ensure that the monitoring data support rigorous hypothesis testing to directly address stormwater trash management questions. As water quality regulations related to trash continue to become more common, data collection systems driven by AI can provide better insights at lower costs. With the appropriate analytical tools in place, these data can help cities respond to problems more quicky, identify where to focus efforts, and understand trash control measure effectiveness.

Software availability

The trash detection model and training dataset created as part of this study is available via an online repository and can be accessed by contacting the authors.

Conflicts of interest

The authors have declared no conflicts of interest.

Credit author contributions statement

Gary Conley: Conceptualization, Methodology, Formal analysis, Writing-original draft. Stephanie Castle Zinn: Methodology, Conceptualization, Writing-review & editing. Taylor Hanson: Methodology, Formal analysis, Software. Krista McDonald: Data curation, vidualization. Nicole Beck: Writing - review & editing, Supervision. Howard Wen: Supervision.

Acknowledgements

The authors thank the City of Salinas and the City of Anaheim for their support and vision in working towards trash-free urban waterways. We thank Catherine Riihimaki for her thoughtful review of the manuscript.

References

Adedeji, O., & Wang, Z. (2019). Intelligent waste classification system using deep learning convolutional neural network. *Procedia Manufacturing, 35*, 607–612.

Ahvenniemi, H., Huovila, A., Pinto-Seppä, I., & Airaksinen, M. (2017). What are the differences between sustainable and smart cities? *Cities, 60*, 234–245.

Anjomshoaa, A., Santi, P., Duarte, F., & Ratti, C. (2020). Quantifying the spatio-temporal potential of drive-by sensing in smart cities. *Journal of Urban Technology*, 1–18.

Batty, M., Axhausen, K. W., Giannotti, F., Pozdnoukhov, A., Bazzani, A., Wachowicz, M., Ouzounis, G., & Portugali, Y. (2012). Smart cities of the future. *The European Physical Journal Special Topics, 214*(1), 481–518.

Bay Area Stormwater Management Agencies Association (BASMAA). (2014). San Francisco Bay area stormwater trash generation rates. Final technical report (Prepared by EOA, Inc. June 2014). https://www.waterboards.ca.gov/sanfranciscobay/water_issues/programs/stormwater/MRP/BASMAA_Trash_ Generation_Rates_Final_Report.pdf.

Bay Area Stormwater Management Agencies Association (BASMAA). (2016). *Tracking California's trash project: Evaluation of the on-land visual assessment protocol as a method to establish baseline levels of trash and detect improvements in stormwater quality. State water resources control board grant agreement no. 12-420-550*.

Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). *Yolov4: optimal speed and accuracy of object detection.* arXiv preprint arXiv:2004.10934.

Chourabi, H., Nam, T., Walker, S., Gil-Garcia, J. R., Mellouli, S., Nahon, K., Pardo, T. A., & Scholl, H. J. (2012, January). Understanding smart cities: An integrative framework. In *2012 45th Hawaii international conference on system sciences* (pp. 2289–2297). IEEE.

Conley, G., Beck, N., Riihimaki, C. A., & Hoke, C. (2019). Improving urban trash reduction tracking with spatially distributed Bayesian uncertainty estimates. *Computers, Environment and Urban Systems, 77*, Article 101344.

Dautel, S. L. (2009). *Transoceanic trash: International and United States strategies for the great Pacific Garbage Patch. Golden Gate U. envtl. lj. 3* p. 181).

De Carolis, B., Ladogana, F., & Macchiarulo, N. (2020, May). YOLO TrashNet: Garbage detection in video streams. In *2020 IEEE conference on evolving and adaptive intelligent systems (EAIS)* (pp. 1–7). IEEE.

Deidun, A., Gauci, A., Lagorio, S., & Galgani, F. (2018). Optimising beached litter monitoring protocols through aerial imagery. *Marine Pollution Bulletin, 131*, 212–217.

Farhadi, A., & Redmon, J. (2018, April). Yolov3: An incremental improvement. In *Computer vision and pattern recognition*. Berlin/Heidelberg, Germany: Springer. pp. 1804–02.

Ghannam, R. B., & Techtmann, S. M. (2021). Machine learning applications in microbial ecology, human microbiome studies, and environmental monitoring. *Computational and Structural Biotechnology Journal, 19*, 1092–1107.

Girshick, R. (2015). Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 1440–1448).

Gunturi, V. M., & Shekhar, S. (2017). Big spatio-temporal network data analytics for smart cities: Research needs. In *Seeing cities through big data* (pp. 127–140). Cham: Springer.

Hawaii Department of Health. (2012). National pollutant discharge elimination system (NPDES) permit (66 pp) http://www.honolulu.gov/rep/site/dfmswq/dfmswq_docs/NPDES_permit_2015.pdf.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961–2969).

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision* (pp. 1026–1034).

Hengstmann, E., & Fischer, E. K. (2020). Anthropogenic litter in freshwater environments–Study on lake beaches evaluating marine guidelines and aerial imaging. *Environmental Research, 189*, Article 109945.

Hino, M., Benami, E., & Brooks, N. (2018). Machine learning for environmental monitoring. *Nature Sustainability, 1*(10), 583–588.

Hoellein, T., Rojas, M., Pink, A., Gasior, J., & Kelly, J. (2014). Anthropogenic litter in urban freshwater ecosystems: distribution and microbial interactions. *PLoS One, 9*, 1–13. June 2014 https://doi.org/10.1371/journal.pone.0098485.

Hossain, S., Debnath, B., Anika, A., Junaed-Al-Hossain, M., Biswas, S., & Shahnaz, C. (2019, October). Autonomous trash collector based on object detection using deep neural network. In *TENCON 2019–2019 IEEE region 10 conference (TENCON)* (pp. 1406–1410). IEEE.

Kraft, M., Piechocki, M., Ptak, B., & Walas, K. (2021). Autonomous, onboard vision-based trash and litter detection in low altitude aerial images collected by an unmanned aerial vehicle. *Remote Sensing, 13*(5), 965.

Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2117–2125).

Marais, M., & Armitage, N. (2004). The measurement and reduction of urban litter entering stormwater drainage systems: Paper 2 – Strategies for reducing the litter in

the stormwater drainage systems. Water SA. 30, 483-492. October 2004. https://doi.org/10.4314/wsa.v30i4.5100.

Marais, M., Armitage, N., & Wise, C. (2004). *The measurement and reduction of urban litter entering stormwater drainage systems: Paper 1 - Quantifying the problem using the City of Cape Town as a case study. Water SA. 30, 469-482. October 2004*. https://doi.org/10.4314/wsa.v30i4.5099

Melinte, D. O., Travediu, A. M., & Dumitriu, D. N. (2020). Deep convolutional neural networks object detector for real-time waste identification. *Applied Sciences, 10*(20), 7301.

Mittal, G., Yagnik, K. B., Garg, M., & Krishnan, N. C. (2016). September. Spotgarbage: smartphone app to detect garbage using deep learning. In *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing* (pp. 940–945).

Moore, S., Cover, M., & Senter, A. (2007). *A rapid trash assessment method applied to waters of the San Francisco Bay Region: Trash measurement in streams. Rapid trash assessment protocol, Version 8. San Francisco Bay Regional Water Board, Surface Water Ambient Monitoring Program*, 55 pp.

Moore, S., Hale, T., Weisberg, S. B., Flores, L., & Kauhanen, P. (2020). *California trash monitoring methods field testing report. SFEI Publication #1026*. Richmond, CA: San Francisco Estuary Institute, 89 pp.

Muñoz-Cadena, C., Lina-Manjarrez, P., Estrada, I., & Ramon-Gallegos, E. (2012). An approach to litter generation and littering practices in a Mexico City neighborhood. *Sustainability, 4*, 1733–1754. December 2012 https://doi.org/10.3390/su4081733.

New York State Department of Environmental Conservation (NYSDEC). (2015). State pollutant discharge elimination permit (53 pp) http://www.honolulu.gov/rep/site/dfmswq/dfmswq_docs/NPDES_permit_2015.pdf Accessed November 2018.

Okafor, N. U., Alghorani, Y., & Delaney, D. T. (2020). Improving data quality of low-cost IoT sensors in environmental monitoring networks using data fusion and machine learning approach. *ICT Express, 6*(3), 220–228.

Pyayt, A. L., Mokhov, I. I., Lang, B., Krzhizhanovskaya, V. V., & Meijer, R. J. (2011). Machine learning methods for environmental monitoring and flood protection. *World Academy of Science, Engineering and Technology, 78*, 118–123.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779–788).

Redmon, J., & Farhadi, A. (2016). *YOLO9000: Better, faster, stronger*. arXiv 2016, arXiv:1612.08242.

Ren, S., He, K., Girshick, R., & Sun, J. (2015). *Faster R-CNN: Towards real-time object detection with region proposal networks*.

Roy, A., Qureshi, S., Pande, K., Nair, D., Gairola, K., Jain, P., Singh, S., Sharma, K., Jagadale, A., Lin, Y. Y., & Sharma, S. (2019). Performance comparison of machine learning platforms. *INFORMS Journal on Computing, 31*(2), 207–225.

Salimi, I., Dewantara, B. S. B., & Wibowo, I. K. (2018, October). Visual-based trash detection and classification system for smart trash bin robot. In *2018 international electronics symposium on knowledge creation and intelligent computing (IES-KCIC)* (pp. 378–383). IEEE.

San Francisco Regional Water Quality Control Board (SFRWQCB). (2015). San Francisco Bay region municipal regional stormwater national pollutant discharge eliminations System (NPDES) permit (364 pp) https://www.waterboards.ca.gov/sanfranciscobay/water_issues/programs/stormwater/Municipal/mrpwrittencomments/Revised_Tentative_Order_and_Attachments.pdf Accessed date: November 2018.

Schermer, M., & Hogeweg, L. (2018). Supporting citizen scientists with automatic species identification using deep learning image recognition models. In *Biodiversity information science and standards*.

Shao, H., Pu, J., & Mu, J. (2021). Pig-Posture Recognition Based on Computer Vision: Dataset and Exploration. *Animals, 11*(5), 1295.

Sigler, M. (2014). The effects of plastic pollution on aquatic wildlife: current situations and future solutions. *Water, Air, and Soil Pollution, 225*, 2184. https://doi.org/10.1007/s11270-014-2184-6

Silva, B. N., Khan, M., Jung, C., Seo, J., Muhammad, D., Han, J., Yoon, Y., & Han, K. (2018). Urban planning and smart city decision management empowered by real-time data processing using big data analytics. *Sensors, 18*(9), 2994.

State Water Resources Control Board (SWRCB). (2015). Water quality control plan: Ocean waters of California (Ocean Plan). Effective January 2016. http://www.waterboards.ca.gov/water_issues/programs/ocean/docs/cop2015.pdf.

State Water Resources Control Board (SWRCB). (2017a). Recommended trash assessment minimum level of effort for establishing baseline trash generation levels. June 2017. https://www.waterboards.ca.gov/water_issues/programs/stormwater/docs/trash_implementation/trash_assmnt.pdf.

State Water Resources Control Board (SWRCB). (2017b). *Recommended trash assessment minimum level of effort for establishing baseline trash generation levels (June 2017)*.

State Water Resources Control Board (SWRCB). (2018). California state water resources control board trash implementation program Accessed May 2017. https://www.waterboards.ca.gov/water_issues/programs/stormwater/trash_implementation.html

Tharani, M., Amin, A. W., Maaz, M., & Taj, M. (2020). *Attention neural network for trash detection on water channels*. arXiv preprint arXiv:2007.04639.

Thung, G. (2020). Trashnet. GitHub repository. Available online https://github.com/garythung/trashnet (accessed on 5 May 2021).

Tiyajamorn, P., Lorprasertkul, P., Assabumrungrat, R., Poomarin, W., & Chancharoen, R. (2019). November. Automatic trash classification using convolutional neural network machine learning. In *2019 IEEE international conference on cybernetics and intelligent systems (CIS) and IEEE conference on robotics, automation and mechatronics (RAM)* (pp. 71–76). IEEE.

Toli, A. M., & Murtagh, N. (2020 Jun 2). The concept of sustainability in smart city definitions. *Frontiers in Built Environment, 6*, 77.

US Census Data (ACS: 2012–2016 and ACS: 2014-2018). (2018). https://data.cnra.ca.gov/dataset/dacs-census Accessed June 2021.

US Environmental Protection Agency (EPA). (2021). *Escaped trash assessment protocol reference manual. Final report, April 2021*, 67 pp.

Wang, X., Liu, S., Shen, X., Shen, C. and Jia, J., 2019. Associatively segmenting instances and semantics in point clouds. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4096-4105).

Wang, X., Kong, T., Shen, C., Jiang, Y., & Li, L. (2020, August). Solo: Segmenting objects by locations. In *European conference on computer vision* (pp. 649–665). Cham: Springer.

Wang, X., Zhang, R., Kong, T., Li, L., & Shen, C. (2020). *SOLOv2: Dynamic and fast instance segmentation*. arXiv preprint arXiv:2003.10152.

Wheeler, S. G., & Knight, E. K. (2017). Monitoring considerations for the trash amendments. Prepared by the California ocean science trust for the state water resources control board (SWRCB) and the California ocean protection council (OPC). July 2017. https://www.waterboards.ca.gov/water_issues/programs/stormwater/docs/trash_implementation/monitconsidfortrashamend_july2017.pdf.

Yang, C., Clarke, K., Shekhar, S., & Tao, C. V. (2020). *Big spatiotemporal data analytics: A research and innovation frontier*.

Yu, G., Wang, Y., Hu, M., Shi, L., Mao, Z., & Sugumaran, V. (2021). RIOMS: An intelligent system for operation and maintenance of urban roads using spatio-temporal data in smart cities. *Future Generation Computer Systems, 115*, 583–609.

Zurowietz, M., Langenkämper, D., Hosking, B., Ruhl, H. A., & Nattkemper, T. W. (2018). MAIA—A machine learning assisted image annotation method for environmental monitoring and exploration. *PLoS One, 13*(11), Article e0207498. https://doi.org/10.1371/journal.pone.0207498