Contents lists available at ScienceDirect

Waste Management

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

Towards automatic waste containers management in cities via computer vision: containers localization and geo-positioning in city maps

Paula Moral [a], Álvaro García-Martín [a,*], Marcos Escudero-Viñolo [a], José M. Martínez [a], Jesús Bescós [a], Jesús Peñuela [b], Juan Carlos Martínez [b], Gonzalo Alvis [b]

[a] Video Processing and Understanding Lab, Universidad Autónoma de Madrid, 28049 Madrid, Spain
[b] URBASER S.A., Camino de las Hormigueras 171, 28031 Madrid, Spain

ARTICLE INFO

Keywords:
Waste container localization
Deep Learning
Computer Vision
Object detection

ABSTRACT

This paper describes the scientific achievements of a collaboration between a research group and the waste management division of a company. While these results might be the basis for several practical or commercial developments, we here focus on a novel scientific contribution: a methodology to automatically generate geo-located waste container maps. It is based on the use of Computer Vision algorithms to detect waste containers and identify their geographic location and dimensions. Algorithms analyze a video sequence and provide an automatic discrimination between images with and without containers. More precisely, two state-of-the-art object detectors based on deep learning techniques have been selected for testing, according to their performance and to their adaptability to an on-board real-time environment: EfficientDet and YOLOv5. Experimental results indicate that the proposed visual model for waste container detection is able to effectively operate with consistent performance disregarding the container type (organic waste, plastic, glass and paper recycling,…) and the city layout, which has been assessed by evaluating it on eleven different Spanish cities that vary in terms of size, climate, urban layout and containers' appearance.

1. Introduction

Waste management—together with stimulating innovation in recycling and limiting the use of landfilling, is one of the three main objectives of European Union (EU) waste policy for protecting the environment and the human health while promoting its transition to a circular economy (European Comision, 2022). In this region with over 447 million inhabitants representing around 16% of the world's gross domestic product, waste management is a challenging task. For instance, in 2018 the total waste generated in the EU by all households added up to 698 million tonnes, an enormous amount which collection and management involves enormous costs not only economic, but also in terms of human resources, time and environmental impact.

Efficient and effective urban planning and infrastructure monitoring are key stages for the adequate management of urban infrastructures. The former aims to ensure that all users and maintenance services have convenient and safe access to and through infrastructure at the smallest cost in time and resources, whereas the latter refers to the continuous assessment of the infrastructure status for its adequate maintenance.

Generally, the planning of waste collection infrastructure follows solid and well-established premises to place and locate containers in cities. However, situations such as the replacement of old infrastructure by a new one with different capabilities, human interaction processes, or changes in urban regulation or scope may lead to outdated plannings, resulting in under or over availability of containers. Similarly, the cost in time and resources of continuous human-monitoring is considerable and, generally, underestimated in the service budget, leading to an ineffective management process and to a late (or absent) reaction to infrastructure harm or deterioration. In any case, a preliminary waste container location and monitoring protocol is an unavoidable step for the effective management and efficient collection planning of waste.

Nowadays, many cities have opted to digitally transform themselves in response to some of the greatest global challenges of our time: population growth, pollution, scarcity of resources, water management, and energy efficiency. To that aim, they rely on Information and Communication Technologies and Big Data Analytics to aid managing of urban

* Corresponding author.
E-mail addresses: paula.moral@uam.es (P. Moral), alvaro.garcia@uam.es (Á. García-Martín), marcos.escudero@uam.es (M. Escudero-Viñolo), josem.martinez@uam.es (J.M. Martínez), j.bescos@uam.es (J. Bescós), jpenuela@urbaser.com (J. Peñuela), jcmartinez@urbaser.com (J.C. Martínez), galvis@urbaser.com (G. Alvis).

https://doi.org/10.1016/j.wasman.2022.08.007
Received 22 April 2022; Received in revised form 8 August 2022; Accepted 10 August 2022
Available online 16 August 2022

services: from transport services to energy generation and water supply. Their goal: to effectively and sustanibly manage cities while reducing energy consumption and CO2 emissions and increasing the well-being of their inhabitants.

Following this spirit, this work proposes a methodology to automatically generate geo-located waste container maps. Its main novelty relies on the use of computer vision algorithms to identify the geographic localization and dimensions of waste containers. The advantages gained from the use of computer vision algorithms over systems such as receptive antennas or Radio Frequency Identification (RFID) are that they are easier to implement, require a much simpler infrastructure and do not require significant maintenance (Gu et al., 2017). RFID technologies use tags attached to objects that emit radio waves to track and identify them. This involves higher economic costs, including the costs for adding tags to containers, their maintenance, and the RFID readers (Wen et al., 2018).

Ideally, the location of waste containers in urban areas should not vary, especially without the knowledge of the companies in charge of garbage collection. However, in many situations this is not the case. The automatic collection of the location of waste containers responds, then, to several factors. For instance, in the case of public authorities that outsource waste management to private companies and want to verify that the agreed requirements are met, by obtaining the current locations of the containers. Another example comes from the updating of out-of-date maps, through the inclusion of additional waste container deployments or container renovations. Likewise, in the case of waste collection companies that aspire to establish a new deployment in a city whose container locations have not been provided. In addition, the proposed methodology provides the basis for exploring container classifications, which would allow, among other things, to identify of the current condition of the containers. This procedure is also useful, for instance, when conducting audits on compliance with the agreed collections (full/empty container), on the replacement of damaged containers, the differentiation between diverse types of waste, etc. In small population areas with less urban development and events, the proposed method presents more limitations than in most densely populated cities. Large urban centers are subject to different factors that alter the location of the containers, such as continuous growth, new needs and legislation (e.g. on recycling), competition by comparison between neighboring areas regulated by different waste collection services, cases of loss or vandalism, etc. Below are presented some practical cases in which the proposal would solve the need to know the location of waste containers. (Salazar-Adams, 2021 and Bel and Sebo, 2021) discuss the need of audits and evaluation of the waste collection service by an external company. (Salazar-Adams, 2021) focuses on the evaluation of the management on waste sector performance in Mexico, with the motivation of determine the collection companies more efficient; on the other hand, (Bel and Sebo, 2021) analyses the competition by comparing quality in neighboring areas of Barcelona administered by different waste management companies. Another example is (Slavík et al., 2021), an study on the importance of the continuous optimization of waste container distribution related to the needs of households, like changing the density of containers, the distance between the address point and the container, and selecting container locations that respect the habits of households in order to decrease the total collection costs; Lastly, different cities provide users with applications to report if their containers have been stolen or damaged (Seattle government, 2022), or to request new collection points (Miami government, 2022), something that our application could detect without depending on the arbitrary predisposition of individuals.

Towards this goal, it is required the training and setup of an automatic localization method for the identification of the geographical location of waste containers via technological means—usually referred as geo-positioning. For the training of such method, this work has access to enormous amounts of geo-positioned images: from public-accessible images available in tools such as Google Street View to user images uploaded to their social networks profiles. These huge databases can be automatically indexed using Application Programming Interfaces (APIs), cover practically every inhabited place on the planet, allow remote exploration on a global scale, and constitute one of the largest annotated databases of our time. However, for the adequate deployment of an automatic system for the automatic location of containers, training the detection system using these geo-positioned worldwide databases is discouraged for several reasons. Firstly, captured containers are not homogeneously distributed: the most populated areas, those with a higher socioeconomic level and those with easier access, usually have a greater number of samples than the rest. In addition, the temporal coverage of these databases is limited, as the extent of the covered space makes the continuous acquisition of images from the same place impractical, resulting in outdated captures: the containers may change in position, appearance, and state, while their acquired reference images in the database remain unchanged, obsolete, as the continuous updating of the databases is unattainable by the acquisition systems of these world-scale geo-positioned databases.

Alternatively, for the effective training of an automatic localization method, the controlled capture of containers using specific resources that permit covering the environments of interest with greater spatial and temporal density is often preferred. One of the prototypical examples of such approaches is the use of cameras placed in geo-positioned cars. Waste containers captured by these cameras can be approximately geo-positioned by *transitivity* using the geo-position of the vehicle. Ideally, updated versions of the containers can be obtained by simply passing by them and capturing their updated appearances using the camera-equipped and geo-positioned vehicle. However, even in local environments such as medium-size cities, the exploration of all the images captured by such vehicles and the manual annotation of the video-captured containers is an arduous task, with enormous economic and temporal costs.

To dramatically reduce these costs, it is proposed an automatic container localization method that provides containers' geo-positions by detecting the containers in images captured by a geo-positioned vehicle. The proposed method provides an automatic division into locations with and without containers by solely analysing geo-located visual information, i.e., the caputerd images. An example of the method output and of its accuracy compared against human-annotations is provided in Fig. 1. This Figure shows that the division between locations with and without containers enables establishing optimal waste collection routes. Likewise, it allows to implement several other applications, such as the identification of unbalanced container distributions when contrasted with the city census, or the evaluation of containers status or level of deterioration. To the best of our knowledge, This proposal is the first that describes an automatic localization of containers in waste management scenarios. Moreover, through experiments conducted in eleven Spanish cities that vary in terms of size, climate, urban layout and containers' appearance, the learning model that supports the proposed container detector has proven successful operation with consistent performance regardless of the type of container (organic waste, plastic, glass and paper recycling, etc) and of the urban framework.

The contribution presented in this work is a novel methodology to automatically generate geo-located waste container maps. To this aim, ConvNets object detectors from the literature with high performances in terms of precision and inference speed are trained for the container detection task using images captured in a known Global Positioning System (GPS) position. This allows the container characterization by its dimension, localization in the image, GPS coordinates or visual feature. With this information the waste container maps are described, allowing for a large number of possible future applications.

The rest of the paper is organized as follows. Firstly, Section 2 reviews the related state of the art. Then, Section 3 describes the proposed system. Afterwards, the experimental results are analysed in Section 4, and finally, Section 5 concludes the paper.

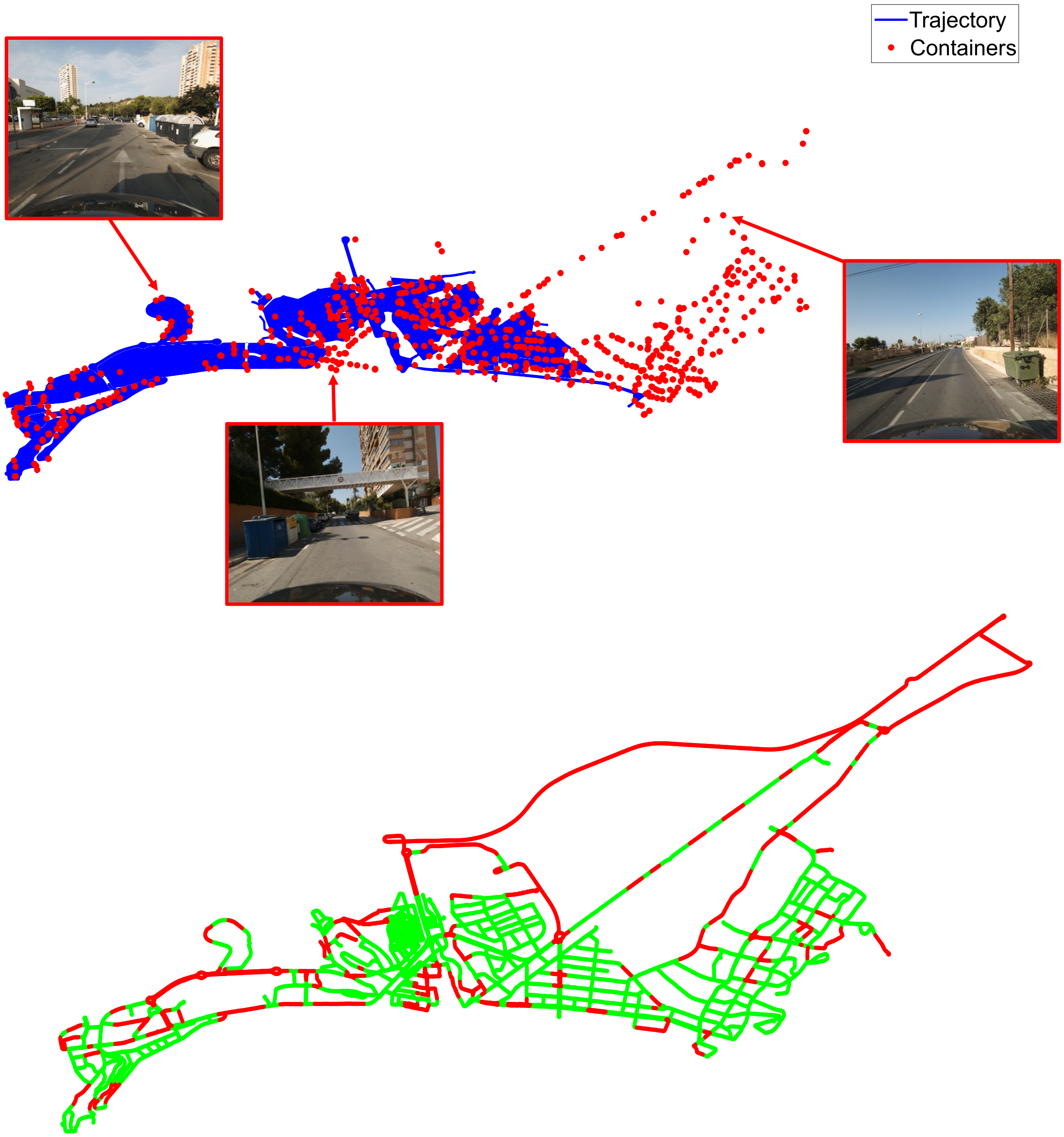

Fig. 1. Automatic container localization in the city of Benidorm, Spain. Top: Vehicle geo-positioned trajectory (in blue), human-annotated camera-captured containers (in red) and three examples of container appearances (surrounding blobs). Bottom: Automatically generated division between geo-positioned locations with (green) and without (red) containers for the proposed method. Note how the automatic localization of containers provided by the method highly correlates with human annotations. Furthermore, the method also proves to have high portability potential, as to avoid over-fitting to the target city, it was not trained using human annotations of this city but from ten other Spanish cities.

2. Literature review

2.1. *Automatic object detection in visual signals*

Automatic object detection in visual signals is the task of assigning a label and a position to objects that appear in an image or in a video sequence. Label assignment is a classification problem (a score is assigned to each object according to the considered object classes) and the position assignment, also called location, is a regression problem (locations in the image are usually represented with rectangles, commonly name Bounding Boxes -BBs-). Through this process, valuable information for the conceptual understanding of images and videos is provided, enabling a wide range of higher abstraction applications, including robot vision (Bai et al., 2020), autonomous driving (Arnold et al., 2019), human–computer interaction (Chakraborty et al., 2017), content-based image retrieval (Dubey, 2021), security/monitoring/video surveillance (Raghunandan et al., 2018), or augmented reality (Rana and Patel, 2019).

Among existing object detection methods, those based on convolutional neural networks (ConvNets)—the preferred tool for Deep Learning in Computer Vision, represent the current state-of-the-art. In fact, ConvNets-based object detection has become a dominant topic in computer vision according to the number of researchers and works in the field. The majority of ConvNets state-of-the-art methods can be categorized into two main genres (Jiao et al., 2019; Zhao et al., 2019; Liu et al., 2020): one-stage object detectors (e.g. Single Shot MultiBox Detector (SSD) (Liu et al., 2016), You Only Look Once (YOLO) v1-v5 (Redmon et al., 2016; Redmon et al., 2017; Redmon and Farhadi, 2018; Bochkovskiy et al., 2020; Jocher et al., 2020), Efficient Detector (EfficientDet) (Tan et al., 2020), Retina Network (RetinaNet) (Lin et al., 2020)) and two-stage object detectors (e.g. Region-Based Convolutional Neural Network (R-CNN) (Girshick et al., 2014), Fast R-CNN (Girshick, 2015), Faster R-CNN (Ren et al., 2017), Cascade R-CNN (Cai and Vasconcelos, 2018)). Whereas two-stage object detectors aim to improve the detection accuracy, one-stage object detection methods have been developed in an attempt to balance the trade-off between performance (in terms of accuracy) and inference speed; hence, they are getting more attraction due to their applicability to real-world problems that demand efficient solutions.

2.2. Aiding waste management using Deep Learning

The huge development of Deep Learning, unprecedented in the field of Computer Vision, is taking place in many fields. In the scope of managing systems in smart cities, recent applications of this technology include automatic control of intelligent street lights (Yaman and Karakose, 2019) or roadside occupation surveillance systems (Ho et al., 2019). Several Computer Vision and Deep Learning techniques have also been explored for waste management, waste planning and waste processing applications. In (Nowakowski and Pamula, 2020), authors propose an image recognition system for the identification and classification of the electrical and electronic equipment waste (specifically, refrigerators, washing machines and monitors), using pictures carefully taken by users of the collection system. According to the category of the waste, e-waste collection companies are then able to deploy a tailored collection plan. Likewise, (Wang et al., 2021) describes an image-based waste classifier focused on reducing the cost of waste classification, monitoring and collection. To facilitate the subsequent waste disposal, the waste is classified into nine categories (*kitchen waste*, *other waste*, *hazardous waste*, *plastic*, *glass*, *paper or cardboard*, *metal*, *fabric* and *other recyclable waste*) before to garbage collection. A similar approach is presented in (Mao et al., 2021) where six different waste categories (*cardboard*, *glass*, *metal*, *paper*, *plastic*, and *trash*) are classified using adapted or customized ConvNets. The adaptation techniques include data augmentation and the use of genetic algorithms to automatically set-up the hyper-parameters of the final fully-connected layer used for object classification. Another interesting approach (Ramírez et al., 2020) describes a method for dumpster classification into seven categories using ConvNets. There, the evaluation dataset compiles cropped versions of labeled images of dumpsters.

Building on top of these methods, there is a small group of approaches that focus on the combined detection and recognition of specific waste for its costumed collection and manipulation. Among them, (Mehendale et al., 2021) proposes an automatic medical waste separator that detects it and categorizes waste into one of the four considered categories (*gloves*, *mask*, *syringe* and *cotton*). Alike, (Wang et al., 2019) proposes a vision-based robot, in this case intended for the recycling of construction waste. The robot is able to detect scattered nails and screws in real time, so that it can automatically collect them and promote the site's construction safety while reducing the squandering of construction material. Meanwhile, (Aitken et al., 2018) presents a vision-based robotic arm for nuclear waste management.

This paper proposes a novel approach for the localization of waste containers. The proposed method leverages detection results from ConvNets object detectors trained on top of deep learned features. This detectors are those from the literature that have been shown to give good results in terms of precision and processing speed, in particular YOLOv5 or EfficientDet. One-stage object detection is used to localize waste containers in images captured in a known GPS position. The final result is the automatic generation of container maps in smart cities scenarios.

3. Materials and methods

3.1. A conceptual approach for the route collection planning of waste containers

The workflow of the envisioned waste management planning system, including stages for the image-based container detection, the container localization and the container characterization, is shown in Fig. 2. During a preliminary non-optimal schedule of the waste collection routes, an on-board camera positioned in the geo-located truck is capturing video frames (temporally correlated consecutive images).

This work proposes a methodology to process these images in an online server using a previously trained image recognition software that localizes the containers. Depending on the on-board capabilities, this system could also work as an on-board application. Once the containers are detected in the images, and thanks to the collected geo-located visual information from on-board cameras, it is possible to obtain the containers localization in terms of GPS coordinates. The data obtained in these two stages, such as the container's dimension, its location in the image, its GPS coordinate and its visual features, lead to the container characterization.

After the geo-positioning and image-based characterization of every container, a plethora of additional applications could benefit from this. For instance the optimization of waste collection routes, actualization of out-of date maps, audits by municipal government of waste management companies, new developments in non-annotated cities, etc.

3.2. Container detection based on Deep Learning

Given that the method proposed in this paper includes object detection techniques to localize waste containers in geo-positioned images, this section summarizes the fundamentals of object detectors based on Deep Learning. ConvNets translate pixels of an image into learnable features at different granularity levels. As aforementioned, detectors can be classified into two main categories: one-stage and two-stage detectors. Two-stage architectures include a feature extraction network followed by a Region Proposal Network (RPN). The outcome of the feature extraction network is a set of Regions of Interest (RoI) which are usually processed by applying RoI pooling to obtain the feature related to the proposals. The objects classification and localization stages are applied on these processed features to yield the final detections. The RPN module is highly computational demanding and rarely operates in real time. However, as their main advantage, two-stage detectors usually present high accuracy in classification and localization. Differently, one-stage detectors predict BBs straightly from input images without a region proposal step. This offers faster processing speed, allowing applications to run in real time. Regarding existing one-stage detectors, EfficientDet (Tan et al., 2020) and YOLOv5 (Jocher et al., 2020) are among the top performing ones in terms of precision and Frames Per Second (*FPS*), allowing the use in on-line systems. Their use in the core of the proposed approach is compared in Section 4.3.

EfficientDet attemps to addres the low accuracy limitation present in one-stage systems compared to two-stage ones by studying various architecture design options. It includes an EfficientNet backbone pretrained using ImageNet (Russakovsky et al., 2015), a weighted bidirectional feature pyramid network (BiFPN) (Tan et al., 2020) and a localization prediction network. The BiFPN is used to avoid the problem of aggregating features at different resolutions (low-level and high-level features).

YOLOv5 is based on the evolution of the YOLO family. This version includes a Cross Stage Partial Network (CSPNet) (Wang et al., 2020) backbone and relies on path-aggregation neck (PANet) (Liu et al., 2018). YOLOv5 is fully implemented in Pytorch from scratch instead of being a modification of the original Darknet from previous YOLO versions. It

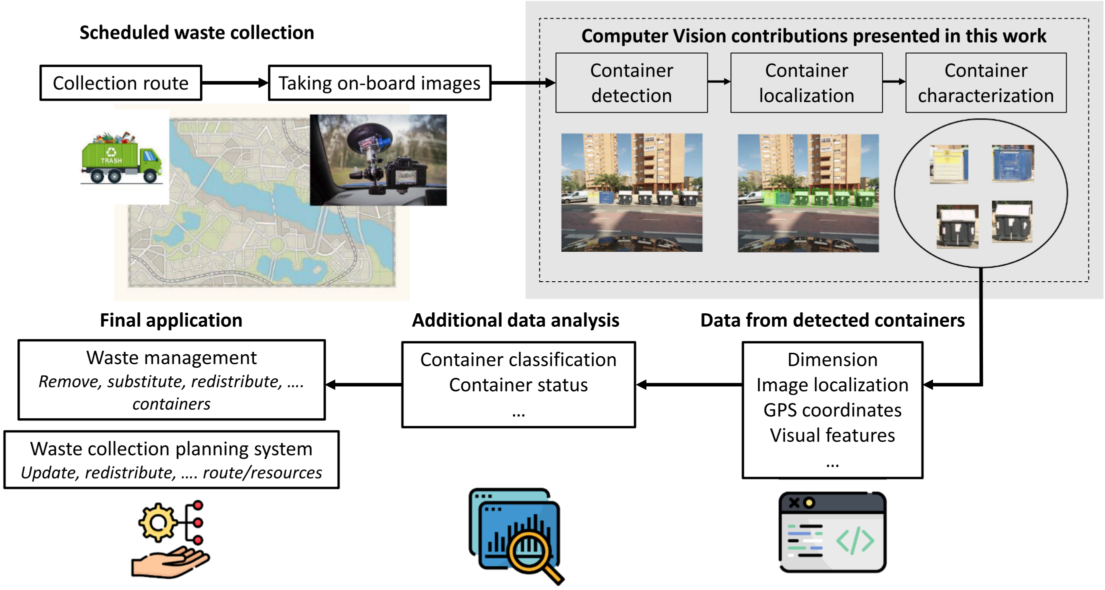
Scheduled waste collection
Collection route
Taking on-board images
Computer Vision contributions presented in this work
Container detection
Container localization
Container characterization
Final application
Additional data analysis
Data from detected containers
Waste management
Remove, substitute, redistribute, …. containers
Waste collection planning system
Update, redistribute, …. route/resources
Container classification
Container status
…
Dimension
Image localization
GPS coordinates
Visual features
…

Fig. 2. Computer vision contributions as part of the potential applications in intelligent waste collection planning systems.

includes mosaic data augmentation and auto learning BBs anchors to improve both the speed and accuracy of previous versions.

3.3. Dataset

The dataset used in this work is composed of images taken from a camera mounted on a vehicle in different cities in Spain. These images contain cars, buildings, traffic signs, vegetation, sidewalks and other urban objects but, for the purpose of this work, only waste containers are considered.

3.3.1. Annotation of video frames including containers

For each sequence, images containing at least one *Container* have been manually annotated. All the containers in the video have been annotated at least but only once, i.e., if the same container appears in several images, only one of them is assigned to the positive set or *Container*-set. Therefore, there are several images with containers that may have been assigned to the *Not-Container* or negative set. To reduce the impact of this annotation noise in the training and validation processes, frames close to a positive one in the evaluation of the methods' performances are ignored. Towards this goal, an uncertainty spatial area of 100-meters around each frame in the *Container* set is defined. Video frames with GPS coordinates in this area are assigned an *Uncertainty* label, finally creating a three-sets partition: *Container*, *Not-Container* and *Uncertainty* (see Fig. 3 for an example of each set). Fig. 3 (c) and (d) are *Uncertainty* images where containers are present, so they cannot be included in the *Not-Container* set, but due to the distance to the camera, they cannot be included in the *Container* set either, as noise would be

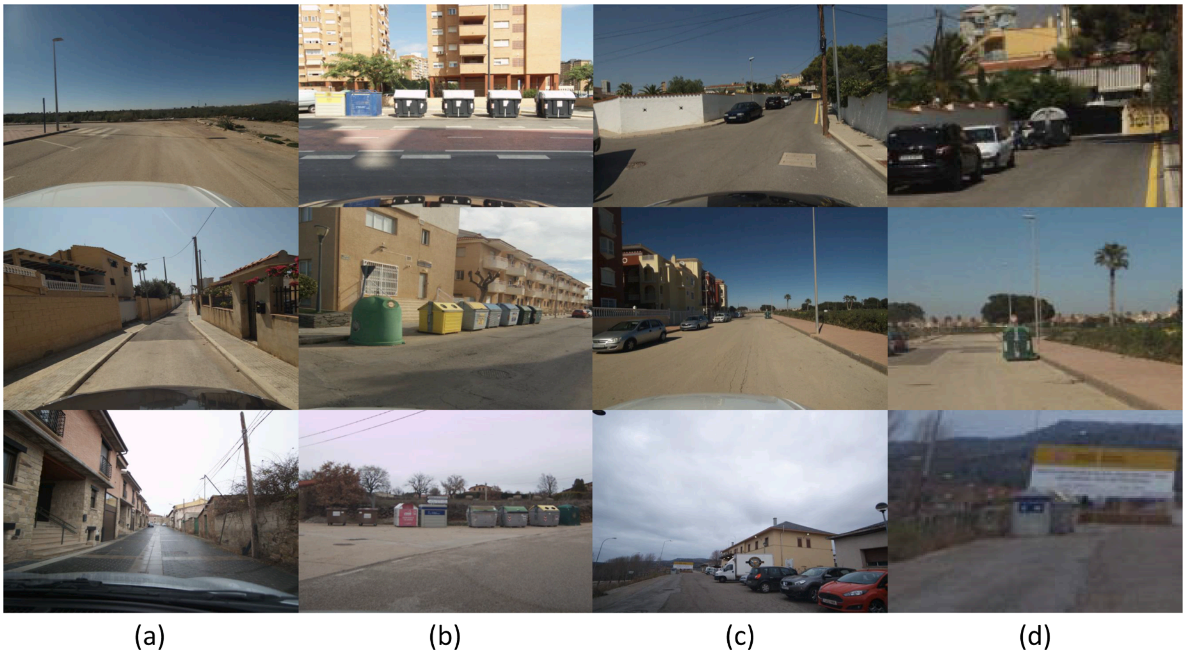

Fig. 3. Example of dataset sets: *Not-Container* (a), *Container* (b), *Uncertainty* (c), same *Uncertainty* example with zoom in the containers area (d). First row are images from Benidorm, second row from Los Alcázares and third row are images from Soria.

introduced during the model training process. In order to see clearly the container, a zoom-in has been applied from (c) to (d). Only frames of the two first sets are used for the training and performance assessment of the evaluated video object detection methods.

3.3.2. Annotation of the containers in the video frames

The spatial location of containers in frames of the *Container* set has been manually annotated using the online tool Alpha Make Sense (Skalski, 2019). For object annotation, the workflow within the tool consists of: images uploading; definition of the list of labels/ classes in the corpus (in this case the only label will be *Container*); then, for each image, the waste containers that appear can be annotated by drawing a BBs enclosing each waste container and the coordinates of the two vertices of each drawn BBs are stored in local files; finally, the files with the annotations are exported following a YOLO (Redmon et al., 2016), Extensible Markup Language (XML) or Comma-Separated Values (CSV) format according to user preferences.

3.3.3. Dataset of waste containers

The dataset consists of video frames or images from 11 cities, summing up a total of 147,160 images divided into 135,192 images in the *Not-Container* set and 11,968 images in the *Container* set. 31,499 containers have been annotated in images of the *Container* set, resulting in an average of 2.6 containers per image in the *Container* set.

The distribution of images and annotations per city are collected in Table 1 whereas captured examples of 4 of the cities are included in Fig. 4.

4. Experimental Results

4.1. Experimental setup

Two tasks are used to evaluate the performance of the proposed system: container detection and containers geo-positioning. For the former complete and representative data is available for the quantitative evaluation, whereas for the latter there is no annotation for several images with containers so that the visual determination of whether a container is present in an image or not may be controversial; hence, this work proposes to generate only qualitative results in the shape of container maps such as the one depicted in Fig. 1.

The EfficientDet detector is trained using 2 as coefficient for the compound scaling method. The backbone is pre-trained on Common Objects in Context (COCO) dataset (Lin et al., 2014), and is then trained for waste container detection during 50 epochs using a learning rate of 0.001 and a batch size of 8 images. An Adamw optimizer (Loshchilov and Hutter, 2017) with a weight decay coefficient of 0.01 is used. For data augmentation, just horizontal flips and normalization operations are applied.

The YOLOv5 detector, the CSPNet backbone (Wang et al., 2020) and the PANet (Liu et al., 2018) for feature extraction are fully trained from scratch using regular Stochastic Gradient Descend with Momentum (SGD) as the optimizer function and 16-samples batches. Then, it is included an initial linear warm-up stage for the first 3 epochs with a starting learning rate of 0.01 and ending in 0.2, the learning rate for training that is decayed every 10 epochs by 0.0005. For data augmentation padding, random crop, horizontal flips, Hue Saturation Value (HSV) color and normalization operations are used.

The models training and evaluation stage have been conducted using PyTorch 1.7.0 DL framework ((Paszke et al., 2017)) running on a Personal Computer using an 8 Cores Central Processing Unit (CPU), 60 GigaBytes (GB) of Random Access Memory (RAM) and a NVIDIA TITAN RTX 24 GB Graphics Processing Unit (GPU).

For estimating the performance of container detection, a Leave-One-Out Cross-Validation procedure is followed. For each target city a model is trained and validated using the data (images and container annotations) from all the other cities (80% for the Train Set and 20% for the Validation Set). The so trained model is tested on the target city data (Test Set). Therefore, the trained model has not been fed the target data during training, so it is possible to evaluate the extrapolation capacity of a model trained on ten other cities, yielding a reliable and unbiased estimate of the performance. Thereby, eleven models are trained. The per-city results are averaged for a global performance considering all the cities. Training in all cities except the one to be used as a test set serves to get the expected result of how the final system trained in all cities will work when used in a new city without annotations.

4.2. Evaluation metrics

The performance of detection algorithms is usually evaluated in terms of their efficiency and their effectiveness (Liu et al., 2020). Efficiency is usually measured through detection speed in FPS, whereas for effectiveness two factors need to be considered: the detection performance or goodness of the detections—in terms of correctness and completeness, and their spatial accuracy—area enclosed by the BBs with respect to that annotated for the object. Spatial accuracy is usually measured in terms of the overlapping ratio between ground truth (*bg*) and detected BBs (*b*) using Intersection over Union (*IoU*) as stated in Eq. 1.

$$IoU(b,bg) = \frac{area(b \cap b_g)}{area(b \cup b_g)} \tag{1}$$

Regarding detection performance, as every detection of the explored detection algorithms is associated a confidence score, performance needs to be assessed under different confidence values which define a detection ranking: those assigned a larger confidence are those with a highest (expected) likelihood to be correct.

Establishing a confidence threshold is equivalent to selecting a subset

Table 1
Dataset and EfficientDet vs YOLOv5 results.

				EfficientDet			YOLOv5		
City	#Not-Container set	#Container set	#Containers	FPS	Validation Set (AP)	Test Set (AP)	FPS	Validation Set (AP)	Test Set (AP)
Benidorm	8,926	1,095	2,764	16.4	0.89	0.91	61.3	0.98	0.94
Brunete	5,051	391	1,394	16.6	0.89	0.89	69.4	0.98	0.94
Burela	3,553	1,251	3,276	16.3	0.91	0.66	63.7	0.99	0.71
Burgos	19,588	2,890	6,900	16.3	0.92	0.73	63.3	0.98	0.78
Cadiz	7,981	891	2,752	16.0	0.90	0.85	76.9	0.98	0.94
Cercedilla	9,448	428	1,347	15.8	0.89	0.90	70.4	0.98	0.94
Los Alcazares	5,132	1,197	2,614	15.7	0.88	0.90	62.1	0.98	0.96
Sant Feliu	15,054	955	3,203	16.4	0.89	0.77	62.1	0.98	0.79
Segovia	4,645	959	2,386	15.8	0.89	0.92	74.6	0.98	0.96
Soria	15,814	1,111	3,054	16.2	0.87	0.90	61.7	0.98	0.96
Toledo	40,000	800	1,809	16.6	0.89	0.77	74.6	0.99	0.91
Total/ Average	135,192	11,968	31,499	16.2	0.89	0.84	67.3	0.98	0.89

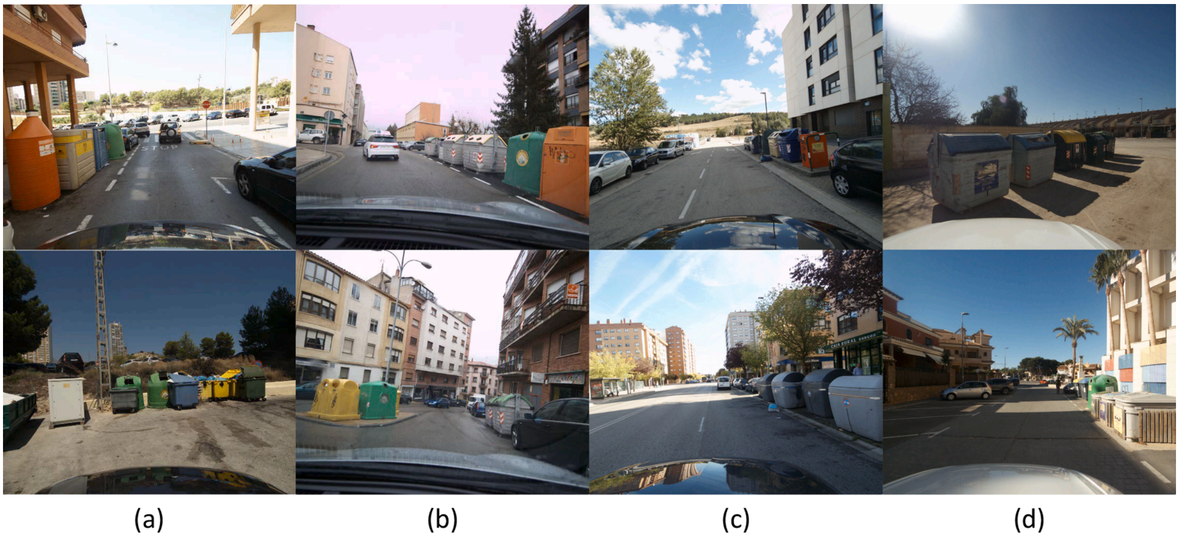

Fig. 4. Visual examples of the cities and of intra- and inter-city container design variation. From left to right: Benidorm (a), Soria (b), Burgos (c) and Los Alcázares (d).

of the ranked detections as correct. For a given threshold, a set of detections is obtained and by comparing this set with the (fixed) set of ground truth BBs, one can define correct object instances (True Positive, *TP*, or True Negative, *TN*), missed object instances (False Negative, *FN*) and false positive instances (False Positive, *FP*). According to the cardinal number of these sets, classical performance measures such as Precision (*P*)—the fraction of all examples which are from the positive class, and Recall (*R*)—the fraction of all positive examples that are successfully classified, can be obtained.

4.3. Results for container detection

As for different confidence score hypothesis the values of *P* and *R* may change, the overall detection performance of a detector is usually measured considering performance curves or graphs and it is usually condensed into a single scalar value using one of the existing global abstractions measures. Among them, the most commonly used metric to assess detectors performance is the Average Precision (AP), which is obtained by finding the area under the precision-recall curve (Russakovsky et al., 2015).

Table 1 compiles efficiency and effectiveness performance of the 11

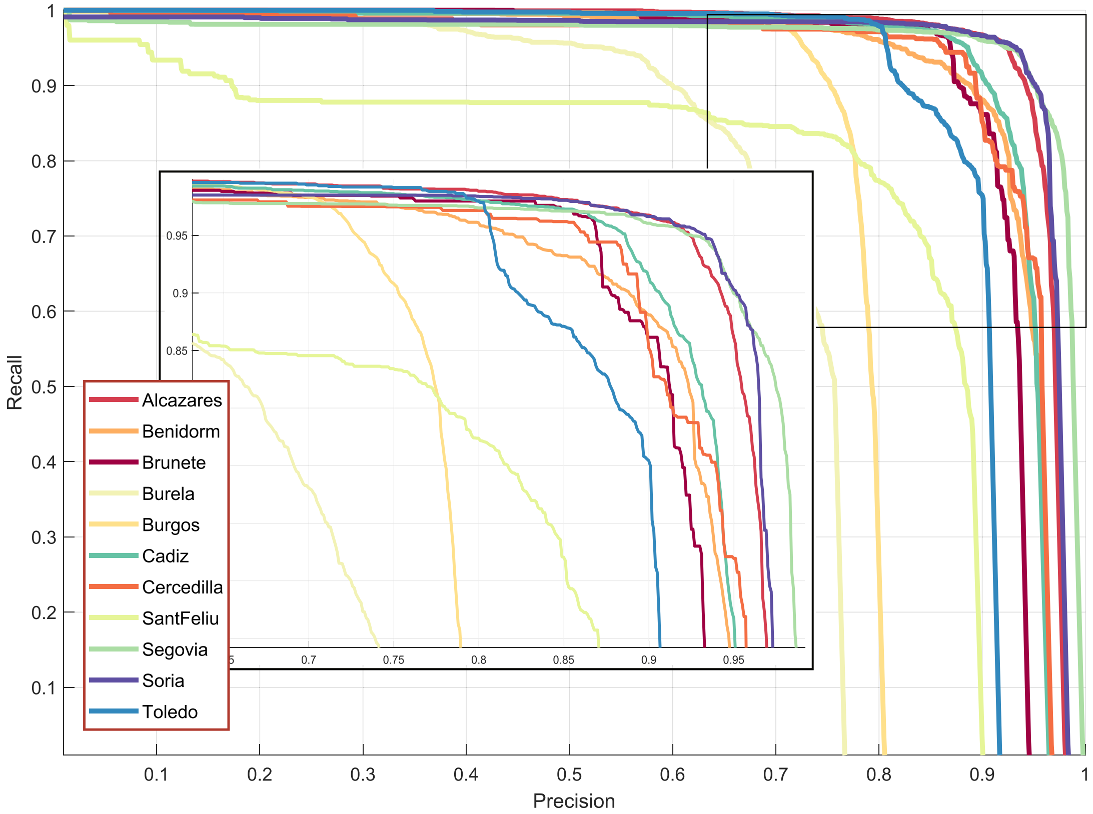

Fig. 5. Precision-recall curves for the 11 trained YOLOv5 models (one for each city). Better viewed in color.

trained EfficientDet and YOLOv5 container detection models. AP is extracted from precision-recall curves (see Fig. 5 for YOLOv5 models) after filtering out all detections with an *IoU* < 0.5 to their closest ground truth one. From Table 1, one can observe that YOLOv5 is not only faster, but also more accurate than EfficientDet. Moreover, AP for the Validation Set of all cities is over 0.87 for EfficientDet and over 0.98 for YOLOv5. Even though the validation examples were not observed during training, the domain and appearances of the containers in the Validation set are the same as those used for training. Regarding the Test set—containing images of the target city, the performance of both detectors is consistent. For EfficientDet, AP ranges between 0.66 and 0.92 with an average AP value of 0.8 and a median AP value of 0.89. As aforementioned, the YOLOv5 detector performs better: the precision-recall curves in Fig. 5 indicate stable and highly accurate container detection for 8 of the cities, with false positive detections being more common for images from Burgos and Burela, and Burela's and Sant Feliu's containers being harder to detect than those in the rest of the cities, resulting in a higher number of false negative detections; hence yielding lower recall values. In any case, the areas under these curves result in top performing AP values ranging between 0.71 and 0.96, with an average AP value of 0.89 and a median AP value of 0.94.

For both detectors, the cities that result in worst transferred performance are Burela and Burgos, which may be caused by the different appearance of the containers of these cities with respect to those in the rest. Besides this domain gap, the rest of the causes for container miss-detection are mainly caused by strong occlusions (see EfficienDet qualitative results for Los Alcázares in Fig. 6), or illumination artifacts in the image. An example of this type of failure can be observed in the qualitative results obtained for Benidorm in Fig. 6, where both detectors fail. Regarding false detections, they mainly emerge from the back of some cars, as in EfficienDet detection in Los Alcázares (see Fig. 6).

4.4. Results for containers geo-positioning

As containers are only annotated once and the video is captured at a regular frame rate, the presence of containers in a considerable part of the *Uncertainty* set of images of the sequence is ensured (see Section 3.3.1). Although results at container level suggest high performance also in the detection of container images, the lack of a complete set of annotations for all the images precludes an exhaustive image-wise quantitative evaluation of the method. Fully annotating a city may be also controversial—besides extremely resource demanding, as for some cases, it may be visually questionable to determine whether a container is present in an image or not. For this reason, the frames classification in *Container* and *Not-Container* sets is evaluated only qualitatively by providing container maps for some of the analyzed cities.

Fig. 1 and Fig. 7 show automatically generated container maps for Benidorm, Toledo and Cercedilla cities. Top maps of the Figures present the vehicle geo-positioned trajectory and the containers' human-provided annotations. The bottom part of the Figures include the automatically generated container maps that present the areas with and without containers.

At first glance, comparing the automatically obtained container areas with the manually annotated positions, it can be observed that there is a container area for every human-annotated container position. The extent of these container areas is an indicator of the visibility of the container which usually starts several frames before the human annotated position and ends several frames after.

From the automatically generated container maps useful information can be obtained for different applications. For instance, in the case of Toledo containers are concentrated in two areas, and that there is a large extension without containers (see Table 1 where the *Not-Container* set has 40,000 images against the 800 of the *Container*-set). Therefore, a future application can use this information to define optimal waste collection routes taking into account this information and avoiding this container-free areas. Furthermore, these maps can be also used to define

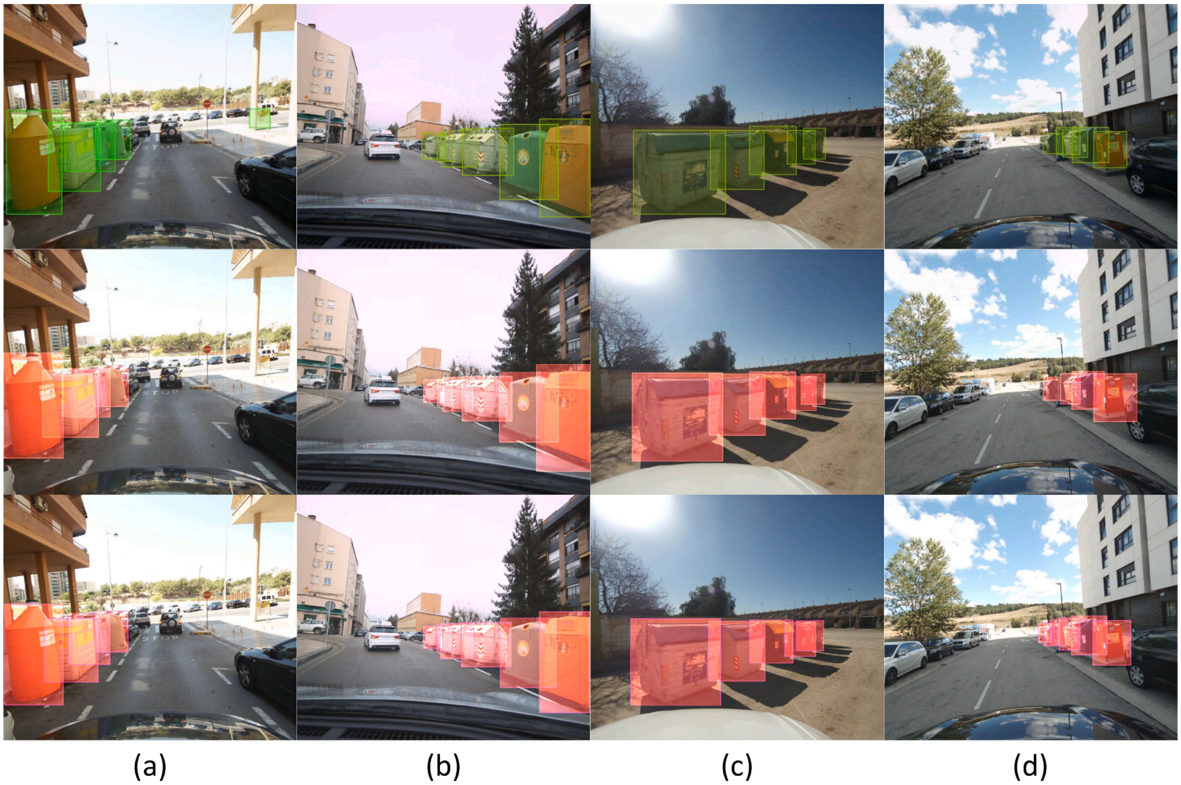

Fig. 6. Qualitative examples of detections of EfficienDet (second row), YOLOv5 (third row) and associated ground-truth (first row). From left to right: Benidorm (a), Soria (b), Burgos (C) and Los Alcázares (d).

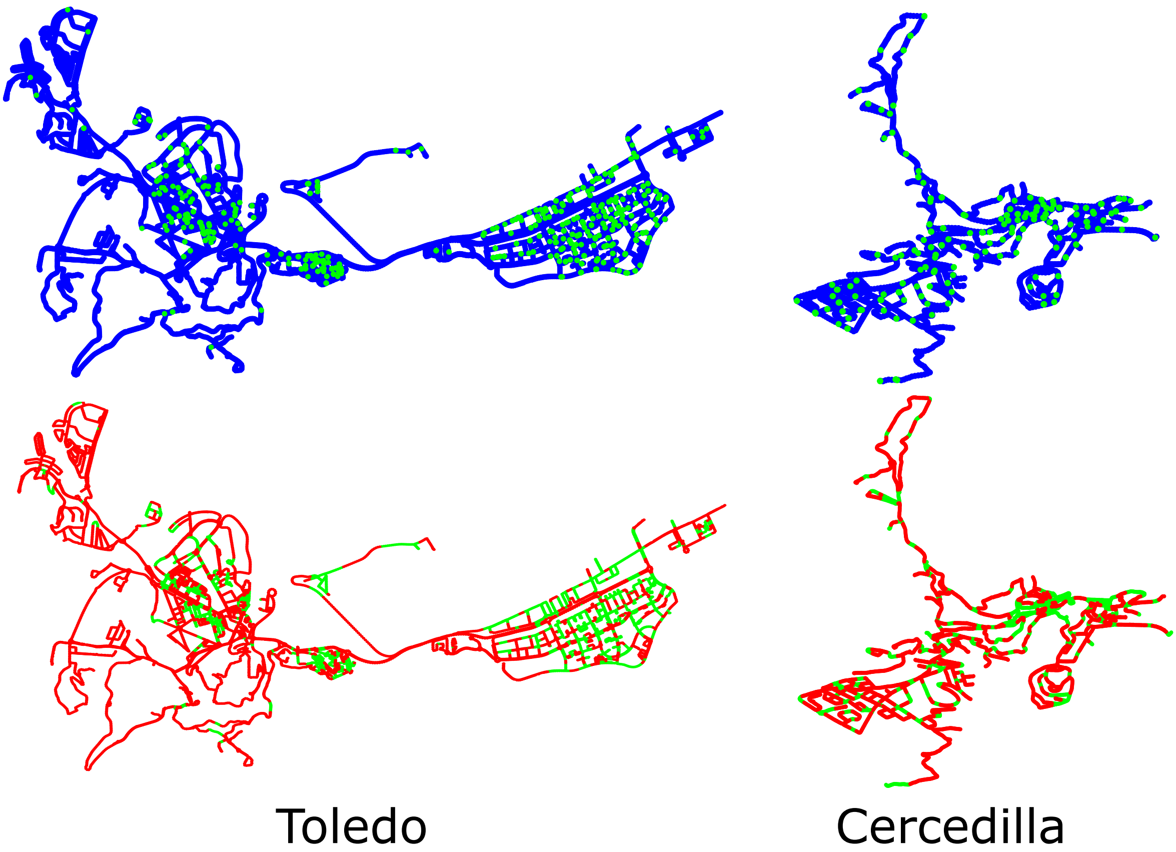

Fig. 7. Automatic container localization in the cities of Toledo (right) and Cercedilla (left). Top: Vehicle geo-positioned trajectory (in blue), human-annotated camera-captured containers (in green). Bottom: Automatically generated division between geo-positioned locations with (green) and without (red) containers for YOLOv5 detection models. Note that for each of these models, human-annotated samples for the target city were not used for training.

optimal collection waste routes, to identify hot spots for vandalism surveillance and to detect uneven distributions of containers.

5. Conclusions

Many cities are opting for using visual analysis technologies based on Deep Learning to manage in an efficient and sustainable way, while reducing energy consumption and emissions of greenhouse gases and increasing the well-being of their inhabitants. In this vein, this paper presents a method for the automatic localization of waste containers in street video sequences captured by a moving vehicle. It allows to generate an automatic division into locations with and without containers from geo-located visual information, leading to geo-located container maps at city scale. These geo-positioned maps enable different applications for the management, collection planning and monitoring of urban waste.

In the core of the proposed method lies a state-of-the-art object detection method to locate containers in images. To select the optimal one for this purpose, the performance of two recent top-performing one-stage object detector methods based on Deep Learning are compared: YOLOv5 and EfficientDet. Reported results indicate that the YOLOv5 detector is the best performing one both in terms of accuracy and efficiency. On average, YOLOv5 obtains an *average precision* value of 0.89 and a detection speed of 67.3 *frames per second*. The representativeness of the experiments is guaranteed by the diversity of the analyzed scenarios, covering sequences of eleven Spanish cities that vary in size, number of inhabitants, urban architecture, climate and container's design and appearance. The obtained results indicate stable and accurate containers' detection and prove that the proposed method is expected to operate similarly in new cities, as they have been obtained using ten cities to train it and a different eleventh one to test it.

Further research might extend the system by differentiating between existing types of containers according to the collected waste nature. This would refine the generated container maps allowing to optimize the waste routes for waste-specific trucks. In addition, different applications could also be developed, such as the evaluation of the state or level of deterioration of the containers, or the identification of unbalanced distributions of containers, contrasted with the city census.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work has been supported by URBASER S.A. and the Universidad Autónoma de Madrid under project REVGA of the *"Segunda Edición del Programa de Fomento de la Transferencia del Conocimiento"* call.

References

Aitken, J.M., Veres, S.M., Shaukat, A., Gao, Y., Cucco, E., Dennis, L.A., Fisher, M., Kuo, J. A., Robinson, T., Mort, P.E., 2018. Autonomous nuclear waste management. IEEE Intell. Syst. 33, 47–55. https://doi.org/10.1109/MIS.2018.111144814.

Arnold, E., Al-Jarrah, O.Y., Dianati, M., Fallah, S., Oxtoby, D., Mouzakitis, A., 2019. A survey on 3d object detection methods for autonomous driving applications. IEEE Trans. Intell. Transp. Syst. 20, 3782–3795. https://doi.org/10.1109/TITS.2019.2892405.

Bai, Q., Li, S., Yang, J., Song, Q., Li, Z., Zhang, X., 2020. Object detection recognition and robot grasping based on machine learning: A survey. IEEE Access 8, 181855–181879. https://doi.org/10.1109/ACCESS.2020.3028740.

Bel, G., Sebo, M., 2021. Watch your neighbor: Strategic competition in waste collection and service quality. Waste Management 127, 63–72. URL: https://www.sciencedirect.com/science/article/pii/S0956053X21002300, doi: 10.1016/j.wasman.2021.04.032.

Bochkovskiy, A., Wang, Chien-Yao andMark Liao, H.Y., 2020. Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934v1, 2020.

Cai, Z., Vasconcelos, N., 2018. Cascade r-cnn: Delving into high quality object detection. In: in: 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6154–6162. https://doi.org/10.1109/CVPR.2018.00644.

Chakraborty, B., Sarma, D., Bhuyan, M., MacDorman, K., 2017. A review of constraints on vision-based gesture recognition for human-computer interaction. IET Comput. Vision 12. https://doi.org/10.1049/iet-cvi.2017.0052.

Dubey, S.R., 2021. A decade survey of content based image retrieval using deep learning. IEEE Trans. Circuits Syst. Video Technol. 1–1 https://doi.org/10.1109/TCSVT.2021.3080920.

European Comision, 2022. Waste and recycling. https://ec.europa.eu/environment/topics/waste-and-recycling_en. Accessed: 22/04/2022.

Girshick, R., 2015. Fast r-cnn. In: in: 2015 IEEE International Conference on Computer Vision (ICCV), pp. 1440–1448. https://doi.org/10.1109/ICCV.2015.169.

Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In: in: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 580–587. https://doi.org/10.1109/CVPR.2014.81.

Gu, F., Ma, B., Guo, J., Summers, P.A., Hall, P., 2017. Internet of things and big data as potential solutions to the problems in waste electrical and electronic equipment management: An exploratory study. Waste Management 68, 434–448. URL: https://www.sciencedirect.com/science/article/pii/S0956053X17305299, doi: 10.1016/j.wasman.2017.07.037.

Ho, G.T.S., Tsang, Y.P., Wu, C.H., Wong, W.H., Choy, K.L., 2019. A computer vision-based roadside occupation surveillance system for intelligent transport in smart cities. Sensors 19. URL: https://www.mdpi.com/1424-8220/19/8/1796, doi: 10.3390/s19081796.

Jiao, L., Zhang, F., Liu, F., Yang, S., Li, L., Feng, Z., Qu, R., 2019. A survey of deep learning-based object detection. IEEE Access 7, 128837–128868. https://doi.org/10.1109/ACCESS.2019.2939201.

Jocher, G., Stoken, A., Borovec, J., NanoCode012, ChristopherSTAN, Changyu, L., Laughing, tkianai, Hogan, A., lorenzomammana, yxNONG, AlexWang1900, Diaconu, L., Marc, wanghaoyang0106, ml5ah, Doug, Ingham, F., Frederik, Guilhen, Hatovix, Poznanski, J., Fang, J., Yu, L., changyu98, Wang, M., Gupta, N., Akhtar, O., PetrDvoracek, Rai, P., 2020. ultralytics/yolov5: v3.1 - Bug Fixes and Performance Improvements. URL: doi: 10.5281/zenodo.4154370, doi:10.5281/zenodo.4154370.

Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2020. Focal loss for dense object detection. IEEE Trans. Pattern Anal. Mach. Intell. 42, 318–327. https://doi.org/10.1109/TPAMI.2018.2858826.

Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C. L., 2014. Microsoft coco: Common objects in context. European conference on computer vision. Springer, pp. 740–755.

Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, Xinwang, Pietikäinen, M., 2020. Deep learning for generic object detection: A survey. Int. J. Comput. Vision 128, 261–318. https://doi.org/10.1007/s11263-019-01247-4.

Liu, S., Qi, L., Qin, H., Shi, J., Jia, J., 2018. Path aggregation network for instance segmentation. In: in: 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8759–8768.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C., 2016. Ssd: Single shot multibox detector. In: Proceedings of the European Conference on Computer Vision (ECCV). Springer, pp. 21–37.

Loshchilov, I., Hutter, F., 2017. Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101.

Mao, W.L., Chen, W.C., Wang, C.T., Lin, Y.H., 2021. Recycling waste classification using optimized convolutional neural network. Resources, Conservation and Recycling 164, 105132. URL: https://www.sciencedirect.com/science/article/pii/S0921344920304493, doi: 10.1016/j.resconrec.2020.105132.

Mehendale, N., Mehendale, N., Sule, V., Tamhankar, C., Kaveri, S., Lakade, N., 2021. Computer vision based medical waste separator. Social Science Research Network. https://doi.org/10.2139/ssrn.3857802.

Miami government, 2022. Replace garbage or recycling cart. https://www.miamigov.com/My-Home-Neighborhood/Garbage-Recycling/Replace-Lost-or-Damaged-RecyclingGarbage-Cart-Bin. Accessed: 05/08/2022.

Nowakowski, P., Pamula, T., 2020. Application of deep learning object classifier to improve e-waste collection planning. Waste Management 109, 1–9. URL: https://www.sciencedirect.com/science/article/pii/S0956053X20302105, doi: 10.1016/j.wasman.2020.04.041.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A., 2017. Automatic differentiation in pytorch.

Raghunandan, A., Mohana, Raghav, P., Aradhya, H.V.R., 2018. Object detection algorithms for video surveillance applications. In: in: 2018 International Conference on Communication and Signal Processing (ICCSP), pp. 0563–0568. https://doi.org/10.1109/ICCSP.2018.8524461.

Ramírez, I., Cuesta-Infante, A., Pantrigo, J.J., Montemayor, A.S., Moreno, J.L., Alonso, V., Anguita, G., Palombarani, L., 2020. Convolutional neural networks for computer vision-based detection and recognition of dumpsters. Neural Comput. Appl. 32, 13203–13211. https://doi.org/10.1007/s00521-018-3390-8.

Rana, K., Patel, B., 2019. Augmented reality engine applications: A survey, in. In: 2019 International Conference on Communication and Signal Processing (ICCSP), pp. 0380–0384. https://doi.org/10.1109/ICCSP.2019.8697999.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection. In: in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779–788. https://doi.org/10.1109/CVPR.2016.91.

Redmon, J., Farhadi, A., 2017. Yolo9000: Better, faster, stronger, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6517–6525. doi:10.1109/CVPR.2017.690.

Redmon, J., Farhadi, A., 2018. Yolov3: An incremental improvement. arXiv preprint arXiv:1412.1628, 2018.

Ren, S., He, K., Girshick, R., Sun, J., 2017. Faster r-cnn: Towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. 39, 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, Michael Berg, A.C., Fei-Fei, L., 2015. Imagenet large scale visual recognition challenge. Int. J. Comput. Vision 115, 211–252. https://doi.org/10.1007/s11263-015-0816-y.

Salazar-Adams, A., 2021. The efficiency of municipal solid waste collection in mexico. Waste Management 133, 71–79. URL: https://www.sciencedirect.com/science/article/pii/S0956053X21003767, doi: 10.1016/j.wasman.2021.07.008.

Seattle government, 2022. Report a missing or damaged container. https://www.seattle.gov/utilities/your-services/collection-and-disposal/your-collection-day/missing-or-damaged-container. Accessed: 05/08/2022.

Skalski, P., 2019. Alpha make sense. https://www.makesense.ai/. Accessed: 22/04/2022.

Slavík, J., Dolejš, M., Rybová, K., 2021. Mixed-method approach incorporating geographic information system (gis) tools for optimizing collection costs and convenience of the biowaste separate collection. Waste Management 134, 177–186. URL: https://www.sciencedirect.com/science/article/pii/S0956053X2100386X, doi: 10.1016/j.wasman.2021.07.018.

Tan, M., Pang, R., Le, Q.V., 2020. Efficientdet: Scalable and efficient object detection. In: 2020 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10778–10787. https://doi.org/10.1109/CVPR42600.2020.01079.

Wang, C., Qin, J., Qu, C., Ran, X., Liu, C., Chen, B., 2021. A smart municipal waste management system based on deep-learning and internet of things. Waste Management 135, 20–29. URL: https://www.sciencedirect.com/science/article/pii/S0956053X21004621, doi: 10.1016/j.wasman.2021.08.028.

Wang, C.Y., Liao, H.Y.M., Wu, Y.H., Chen, P.Y., Hsieh, J.W., Yeh, I.H., 2020. Cspnet: A new backbone that can enhance learning capability of cnn. In: in: 2020 IEEE Conference on Computer Vision and Pattern Recognition workshops (CVPRW), pp. 390–391.

Wang, Z., Li, H., Zhang, X., 2019. Construction waste recycling robot for nails and screws: Computer vision technology and neural network approach. Automation in Construction 97, 220–228. URL: https://www.sciencedirect.com/science/article/pii/S0926580518302218, doi: 10.1016/j.autcon.2018.11.009.

Wen, Z., Hu, S., De Clercq, D., Beck, M.B., Zhang, H., Zhang, H., Fei, F., Liu, J., 2018. Design, implementation, and evaluation of an internet of things (iot) network system for restaurant food waste management. Waste Management 73, 26–38. URL: https://www.sciencedirect.com/science/article/pii/S0956053X17309376, doi: 10.1016/j.wasman.2017.11.054.

Yaman, O., Karakose, M., 2019. New approach for intelligent street lights using computer vision and wireless sensor networks. In: 2019 7th International Istanbul Smart Grids and Cities Congress and Fair (ICSG), pp. 81–85. https://doi.org/10.1109/SGCF.2019.8782330.

Zhao, Z.Q., Zheng, P., Xu, S.T., Wu, X., 2019. Object detection with deep learning: A review. IEEE Trans. Neural Networks Learn. Syst. 30, 3212–3232. https://doi.org/10.1109/TNNLS.2018.2876865.