

VOL. 107, 2023



DOI: 10.3303/CET23107013

Guest Editors: Petar S. Varbanov, Bohong Wang, Petro Kapustenko Copyright © 2023, AIDIC Servizi S.r.l. ISBN 979-12-81206-07-6; ISSN 2283-9216

Enhanced Floating Plastic Waste Detecting on Offsets of River Tisza, Hungary

Gábor Élő*, Gabor Paller

Széchenyi István University, Egyetem tér 1., 9026 Győr, Hungary elo@sze.hu

The topic was to spot and potentially measure the amount of floating waste on the river Tisza and its offsets. It was a two-year long study in which a complete system was designed and built and performed measurements 50-70 % of waste dumped into rivers are plastics. Ranging from micro-plastics (< 0.1 µm) to macro-plastics (>5 cm). The nature of the plastic pollution depends greatly on the source of the pollution. In the river Tisza and its offsets, the pollution is mainly coming from landfills located near the upstream. In the first phase, an experimental motion-detection camera system was developed to try out multiple configurations during the research. The open-source motion software has been implemented, running on Raspberry Pi 3 as data collectors. The system uploaded data into a data server running in the cloud (Azure). The camera system was operating for more than a year and collected over 440,000 pictures. At the end of this phase, the conclusion was that individual plastic objects are not recognisable, only bigger groups of them. On top of this, we have seen that the optical noise is very high, rendering many of the pictures unfit for analysis, but the results still served as a very good starting point for the collection of Al training data. During the second phase, software was experimented. YOLOv3 and Faster R-CNN have been applied, eventually settling for Faster R-CNN with a ResNet-50-FPN base network.

1. Introduction

For the everyday person, the most visible environmental problem is the presence and processing of waste. Urban and countryside residents encounter waste collection, affecting their behaviour and opinions. On the lowest level, packaging and discarding goods is an everyday task, sometimes even from the point of purchase up until the recycling bin. One level above that, people desire to keep their immediate environment waste-free and utilise waste in the best possible way. The media brings news to the global environment and waste processing. A special case of this is when someone is living or working near surface waters, so the global problem of water pollution is affecting them directly as well (Aytan et al., 2020). One of the most visible forms of pollution is the waste dumped into our surface waters. This waste can assemble into islands in the seas or in lakes and can travel down on rivers and from river wash (Klemeš et al., 2021). It's an inherent issue that people living upstream waters pollute the environment of those who live downstream (Castro-Jimenez et al., 2019). Society must accommodate for this basic law of physics. Even if one or two PET bottles would not seem like a big deal upstream, they can compound into mountains of waste on a slower downstream section and can even cause disasters. It is a logical approach to get the waste out as soon as possible, as removing a few PET bottles is easier than lifting out huge piles of garbage using heavy machines (Figure 1). This was the founding statement of the two-year R&D project that aimed to rid the river Tisza and its feeders of plastic waste. The goal was to build a floating device that collects the waste upon its first appearance when there are only a few items to collect. As a result, a watercraft has been made that can optically detect discarded PET bottles and remove them from the river. In this project, our task was the optical detection and early alerting of the appearance of floating PET bottles.



Figure 1: Typical waste in the river

2. Problem analysis: floating macroplastic pollution

Plastic pollution has many sources. (Lechner et al., 2014) highlights micro- and mesoplastic debris resulting from industrial plastic production. In Eastern Hungary, significant macroplastic pollution is caused by improperly handled or outright illegal upstream waste dumps (Ljasuk, 2021). These waste dumps cause large-scale macroplastic pollution, usually when the river is flooding. A typical polluting item is a plastic bottle. Figure 2 shows a debris-moving simulation supporting problem analysis.



Figure 2: Debris moving simulation

This requires preparation time; therefore, a timely warning is very important. Plastic items can be detected in a number of ways, but the observation limitations quickly eliminate most of them. The observation environment has the following properties. The detection distance is relatively long. The rivers where performing the detection are quite wide. Distances of 30-50 m are not uncommon. Although it would be definitely easier to mount the

camera downward-looking (van Lieshout et al., 2020), the water surfaces to monitor do not permit that configuration.

Plastic items to be detected are covered with other materials from the environment. There is almost always a water film on them, and other foreign materials (e.g., dirt or algae) are quite common. These limitations make remote materials testing methods largely unusable. Laser Induced Breakdown Spectroscopy or spectral imaging all require illumination by a special light source (laser or infrared/ultraviolet light source) (Gundupalli et al., 2017), which is very complicated given the significant distance between the observation location and the target object. The water film and other materials covering the targets also make remote materials testing unfeasible.

3. First phase: using motion detection camera

Motion detection security cameras almost always have a feature that detects moving objects in the input image stream. The mechanism is simple. The camera compares the actual image to the previous image (or a limited set of previous images) and calculates the differing pixels between the actual and previous image(s). If the amount of differing pixels is too high, a relevant movement event is triggered and based on the configuration of the camera, the image and the difference mask are saved. Off-the-shelf security cameras have limited configuration options with regard to motion detection parameters, so we built our own motion detection camera and its server backend (Figure 3).



Figure 3: Motion detection camera

It was clear from the beginning that the system must operate in edge computing architecture, i.e., the camera node has to have built-in intelligence to select image candidates where something relevant is happening, as the data connection between the camera unit and its server backend will not be able to transfer all the images taken. Components of this system are the following. Camera unit based on a Raspberry Pi 3 Model B+ single board computer and its Raspberry Pi Camera Module 2. The camera unit runs the motion open-source software that implements the motion detection algorithm and has many configuration parameters to tune this algorithm. The camera unit continuously runs the motion detection, and in case of a movement trigger it saves the actual image and the difference mask to the local SD card. The camera unit also maintains an SSH tunnel to its camera server. The camera unit runs the motion open-source software that implements the motion detection algorithm and has many configuration parameters to tune this algorithm. Camera server is a web application implemented in Spring/Java and deployed into the Azure cloud. The camera server regularly visits the camera units and retrieves the images and the difference masks. The camera server has a web interface that allows authenticated users to browse images. Administrator users can also configure the motion detection parameters. The camera unit was deployed in the harbour of Bodrogkisfalud, Hungary and operated for 13 months. During this time, the camera unit recorded more than 440,000 images. Most of these images were not floating plastic waste but unrelated changes in the input image, e.g., boat traffic of the harbour or even the sun's glitter on the river. When the camera recorded relevant images, those images were related to larger islands of floating debris, sometimes containing plastic waste (Figure 4).



Figure 4: Islands of floating debris

Figure 5 shows such larger floating debris and the different masks that triggered the image capture. Green areas are masked out from the motion detection.



Figure 5: Mask image of floating island

After a lengthy configuration tuning process, 4,000 pixel threshold was chosen. This is the number of pixels changed in the image that triggers a capture. Considering the image size of 1,024x640 pixels, it is clear that individual plastic waste items cannot be detected only if they form a larger block of debris. Even with this quite high threshold, the first iteration generates a large amount of irrelevant images because its selectivity is low.

4. Second phase: Applying Deep Neural Networks (DNN)

The first phase failed in terms of selectivity as it picked up a large number of images where nothing relevant happened. In addition, its sensitivity did not satisfy the requirements either because the pixel threshold was too high to capture individual plastic waste items and lowering the threshold would have generated even more false alarms. As plastic waste is often contaminated by, e.g., dirt and comes in different colours and shapes, we needed an image recognition algorithm able to operate in such a noisy environment. Deep Neural Networks (DNN) were expected to satisfy these requirements. Applying DNN to recognise floating plastic waste is not a new idea (van Lieshout et al., 2020) also took this approach. The camera setup and the classification requirements are different. Their system uses a downward-looking camera, which decreases the distance to the target objects. This setup also results in better resolution, which lets them perform more detailed classification ("plastic"/"not plastic"). The second iteration is still expected to cover as large a water surface as possible, which means that target objects measuring 20-30 cm can be as far as 20-30 meters from the camera. Even if the distance can be partially offset by optical zoom, targets will still look small in the input image. We experimented with the YOLOv3 3 (Redmon and Farhadi, 2018) and Faster R-CNN (Ren et al., 2017) deep neural networks. Implemented reliable object detection with YOLOv3. Imprecise localisation was experienced. In the case of Faster R-CNN, the challenge was that our training machine had only 6GB of GPU memory, which is not sufficient to train with the ResNet-101 backbone commonly used with Faster R-CNN, so that was falling back to the ResNet-50-FPN model. The implementation came from Torchvision. The model was pre-trained on the COCO train2017 dataset; the last 3 layers of the backbone were allowed to train, and we trained for 11 epochs. The threshold confidence level was set to a relatively low value, 0.25. The expectation was that this low level would generate some false positives, but the low quality of the target objects (due to the quite long distance) requires relatively lax recognition. The initial training dataset had 195 images, was annotated with the VGG Image Annotator tool 1 (Dutta et al., 2023) to determine the bounding area of the relevant object and came from the following sources. Some of the images were collected by camera crews, while others came from cameras from the first iteration. Plastic waste collected from the river and captured in front of neutral backgrounds.



Figure 6: Training image

Figure 6 shows such a training image with bounding area annotation shown. Plastic waste images in natural settings (e.g., seashores) were collected from the internet. The system's architecture is constructed in such a way that its user is expected to continuously collect and annotate images. Hence, we expect the training image set to grow. The training image set was augmented by rotating the training images by 90, 180 and 270 degrees. All the images were scaled so that their longer side was 640 pixels. There is no scaling augmentation as the model's Feature Pyramid Network takes care of scaling the training data during inference.



Figure 7: Waste recognition

The result is demonstrated in Figure 7. The algorithm cannot recognise every waste item, but it recognises enough so that the warning can be triggered. Further analysis was done on five video footage taken in different circumstances about real large-scale plastic waste pollution. Each footage is filmed in a river landscape environment and depicts floating debris from larger distances (5-50 m), plastic and non-plastic, at the same time. Typical training image with bounding area annotation shown the selected section of the video footage, we counted the recognisable debris and compared it to the output of the algorithm.

The following categories were considered:

- Recognized: the human viewer considers the item a plastic waste, and the algorithm located it correctly.
- Not recognised: The human viewer considers the item a plastic waste, but the algorithm does not locate it correctly.
- Miscategorised: The human viewer does not consider the item a plastic waste, but the algorithm identifies it as such.

Note that due to the long distance and the quality of the footage, it is not always easy to decide, even for a human viewer, if a certain piece of debris is, e.g., a plastic bottle or a wooden trunk. Also, in the mass of floating debris, it is not always possible to distinguish individual items. More detailed analysis reveals, however, that this is mostly due to the chaotic waste mass where even human viewers have trouble distinguishing and categorising items. DNN-based image recognition algorithm was a great leap toward more reliable waste recognition, and it is expected that its performance will improve as more training images accumulate during field operation. The early warning system's architecture was updated to accommodate the functions needed to operate the deep neural network.

Compared to the first phase, the changes are the following.

- Images are now taken by a professional surveillance camera featuring optical zoom. This is necessary to provide sufficiently detailed images so that the DNN-based algorithm can pick individual plastic waste objects. A Foscam surveillance camera with 18x optical zoom was employed.
- The camera unit is now responsible for running the trained DNN in inference mode. This required significant
 hardware updates as the embedded computer has to be equipped now with a reasonably powerful GPU.
 As the camera unit is an edge node and runs only inference and not training, a GPU with 2GB GPU memory
 is enough. The surveillance camera and the camera server are connected with the ONVIF protocol, which
 lets the camera server rotate the camera so that the camera's view can scan the entire observation area.
- The camera server got additional functions related to selecting and annotating images in which a relevant object was not recognised, initiating training on the training server and updating the DNN weight files on the camera units.
- The training server is a new component that is responsible for running the DNN in training mode once new annotated images are available. It is separated from the camera server as training requires a relatively strong GPU (6GB GPU memory with the current model).

The second iteration featuring a DNN increased the selectivity of the system significantly, generating much fewer false alarms.

5. Conclusions

After concluding of research, the actions performed in phase two have significantly reduced the amount of false detections. On the large water surface area, it is hard to recognise the target objects, so perfect detection is not possible, i.e., always have false negatives (and some false positives). This makes the solution fit for early detection and alerting but not so fit for precise counting. Due to bandwidth limitations, the system can only be deployed in an edge computing architectural style, which requires a relatively powerful GPU in the camera unit in case of the second iteration. Results can be adapted as-is for practical use of early alerting. It is capable of operating ashore, aboard a watercraft or aboard an aircraft (e.g., drone) and on every waste collection area where detection is challenging otherwise (e.g. agricultural lands, forests, roadsides).

References

- Aytan Ü., Pogojeva M., Simeonova A. (Eds.), 2020, Marine Litter in the Black Sea, Turkish Marine Research Foundation (TUDAV) Publication, Istanbul, Turkey, 361.
- Castro-Jimenez J., Gonzalez Fernandez D., Fornier M., Schmidt N., Sempere R., 2019, Macro-litter in surface waters from the Rhone River: Plastic pollution and loading to the NW Mediterranean Sea, Marine Pollution Bulletin, 146, 60-66.
- Dutta A., Gupta A., Zisserman A., 2023, VGG Image Annotator (VIA). https://www.robots.ox.ac.uk/~vgg/software/via/, accessed 18.11.2023.
- Gundupalli S.P., Hait S., Thakur A., 2017, A review on automated sorting of source-separated municipal solid waste for recycling. Waste Management, 60, 56–74.
- Klemeš J.J., Fan Y.V., Jiang P., 2021, Plastics: friends or foes? The circularity and plastic waste footprint, Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 43, 1549-1565.
- Lechner A., Keckeis H., Lumesberger-Loisl F., Zens B., Krusch R., Tritthart M., Glas M., Schludermann E., 2014, The Danube so colourful: A potpourri of plastic litter outnumbers fish larvae in Europe's second largest river. Environmental Pollution, 188, 177–181.
- Ljasuk D., 2021, In the name of the Tisza source to the Black Sea. <www.youtube.com/watch?v=TLyK alu3fc>, accessed 11.10.2023.
- Redmon J., Farhadi A., 2018, YOLOv3: An incremental improvement.Tech report. <arxiv.org/abs/1804.02767>, accessed 11.10.2023.
- Ren S., He K., Girshick R., Sun J., 2017, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39, 1137-1149.
- van Lieshout C., van Oeveren K., van Emmerik T., Postma E., 2020, Automated river plastic monitoring using deep learning and cameras. Earth and Space Science, 7, e2019EA000960.

78