Towards a floating plastic waste early warning system

Gábor Paller[1] and Gábor Élő[1]

[1] *Széchenyi István University, Information Society Research & Education Group, Egyetem tér 1. Győr, Hungary*
paller.gabor@sze.hu,elo@sze.hu

Keywords: image processing, waste detection, machine learning

Abstract: Plastic waste in living waters is a worldwide problem. One particular variant of this problem is floating plastic waste, e.g. plastic bottles or bags. Rivers often carry large amount of floating plastic waste, due to unauthorized or not properly maintained waste dumps installed in the rivers' flood plain. It is of utmost importance that environmental protection agencies be aware of such large-scale plastic pollutions so that they can initiate appropriate countermeasures. This paper presents two iterations of an early warning system designed to alert environmental protection agencies of plastic waste pollution. These systems are based on processing camera images but while the first iteration uses motion detection for identifying relevant images, the second iteration adopted a machine learning algorithm deployed in edge computing architecture. Better selectivity of the machine learning-based solution significantly eases the burden on the operators of the early warning system.

1 INTRODUCTION

By quantity, plastic is the most important riverine pollutant. Depending on the measurement source plastic pollution is estimated to be between 50% and 70% of the entire solid pollution material (Aytan et al., 2020) (Castro-Jiménez et al., 2019). The size of plastic segments vary, from nano plastic (<0.1 µm) to macroplastic (>5 cm) but eventually larger segments break down to micro and nano plastics and enter the food chain (van Emmerik and Schwarz, 2020). Therefore it is advantageous to eliminate the plastic pollution while it still consists of larger segments.

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 pollutions, usually when the river is flooding. Typical polluting item is a plastic bottle. Figure 1 shows a plastic bottle pollution case on the Szamos river, Hungary which is probably a result of an improperly handled waste dump. Enviromental protection agencies react to large-scale pollution cases by mobilizing heavy equipment, e.g. barges and excavators. 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.

- Detection distance is relatively long. The rivers where we want to perform the detection are quite wide, 30-50 meters is 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.

Considering these difficulties we chose observation in the visible light domain. This approach does have its drawbacks too. The current implementation works only in the daytime and observation in the night is definitely a requirement. This requirement can be implemented with a strong infrared searchlight but it is not in the content of our current research.

2 EARLY WARNING SYSTEM ITERATIONS

2.1 First iteration: 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, 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 regards to motion detection parameters so we built our own motion detection camera and its server backend. 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.
- Camera server is a web application implemented in Spring/Java 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 440000 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. Figures 2 and 3 show such a larger floating debris and the difference mask that triggered the image capture. Green areas are masked out from the motion detection.

After a lengthy configuration tuning process 4000 pixel threshold was chosen. This is the number of pixels changed in the image that triggers a capture. Considering the image size of 1024x640 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.

2.2 Second iteration: deep learning

The first iteration 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 colors and shapes therefore 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 recognize floating plastic waste is not a new idea, (van Lieshout et al., 2020) also took this approach. The camera setup and the classfication requirements are different, however. 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").

Our second iteration is still expected to cover as large 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. For this reason we did not expect any classification of the floating waste.

We experimented with the YOLOv3 (Redmon and Farhadi, 2018) and Faster R-CNN (Ren et al., 2015) deep neural networks. We could not achieve reliable object detection with YOLOv3, imprecise localisation was experienced. This is in line with YOLOv3

Figure 1: Macroplastic pollution consisting mostly of plastic bottles, Szamos river, Olcsva, Hungary, 2020 June

Figure 2: Floating island of debris

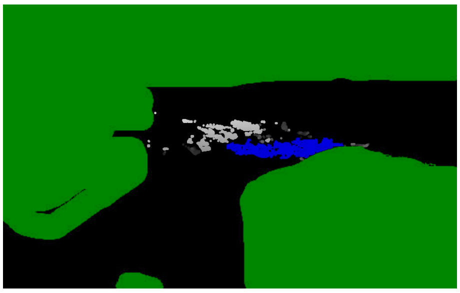

Figure 3: Floating island of debris, mask image

authors' own paper which notes that YOLOv3 struggles to get the boxes perfectly aligned with the object.

In 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 ResNet-101 backbone commonly used with Faster R-CNN, so we fell back to ResNet-50-FPN model. The implementation came from Torchvision, the model was pre-trained on COCO train2017 dataset, the last 3 layers of the backbone was allowed to train and we trained for 11 epochs. The threshold confidence level was set to a relatively low value, 0.25. We expect that this low level will 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] to determine the bounding area of the relevant object and came from the following sources.

- Live images of plastic waste floating on the river. Some of the images were collected by camera crews while others came from our own camera from the first iteration. We did not have enough high-quality images of this kind in our disposition therefore we needed to add additional images from diverse sources.
- Plastic waste collected from the river and photoed in front of neutral backgrounds. Figure 4 shows such a training image with bounding area annotation shown.
- Plastic waste images in natural settings (e.g. seashores) 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 pixel. There is no scaling augmentation as the model's Feature Pyramid Network takes care of scaling the training data during inference.

The result is demonstrated in Figure 5. The algorithm cannot recognize every waste item but it recognizes enough many so that the warning can be triggered.

Further analysis was done on 5 video footages (Table 1) 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 distance (5-50 meters), plastic and non-plastic at the same time. By viewing

[1] https://www.robots.ox.ac.uk/~vgg/software/via/

Figure 4: Typical training image with bounding area annotation shown

the selected section of the video footage, we counted the recognizable 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 recognized The human viewer considers the item a plastic waste but the algorithm does not locate it correctly.

Miscategorized 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 footages 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. Therefore the counts can be considered only approximations.

In addition, we calculated the following metrics.

$$NRT = \frac{TNR}{TNR+TR} 100\% \quad (1)$$

where TNR and TR are the sum of the recognized and not recognized item counts in all the frames of the analysed segment.

$$MR = \frac{TM}{TR} 100\%. \quad (2)$$

where TM is the sum of all the miscategorized items in all the frames of the analysed segment.

The results can be seen in Table 2. The results demonstrate that the algorithm does not recognize a large amount of relevant objects. In case of footage #1, more than half of the items were not recognized. The items that were recognized, however, are numerous enough to trigger an alert therefore the early warning system performs its intended function. The large number of miscategorized items (up to 15%) is also a concern. More detailed analysis reveals, however, that this is mostly due to the chaotic waste mass where even human viewers have trouble distinguishing and categorizing items. Footage #5 has no such problem because no such mass of waste is present in that footage. We therefore consider the DNN-based image recognition algorithm a great leap toward more reliable waste recognition and we expect that its performance will improve as more training images accumulate during field operation.

The early warning system's architecture was updated to accomodate the functions needed to operate the deep neural network (Figure 6). Compared to the first iteration 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 could pick individual plastic waste objects. A Foscam surveillance camera with 18x optical zoom was employed.

- Camera unit is now responsible of running the trained DNN in inference mode. This required significant hardware update 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. This required however to swap the Raspberry Pi 3 with an industrial PC equipped with NVIDIA GPU. The surveillance camera and the camera server are connected with ONVIF protocol which lets the camera server rotate the camera so that the camera's view can scan the entire observation area.

- Camera server got additional functions related to selecting and annotating images in which a relevant object was not recognized, initiating training on the training server and updating the DNN weight files on the camera units.

- 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 relatively strong GPU (6GB GPU memory with the current model).

Figure 5: Detected waste in the Olcsva video footage

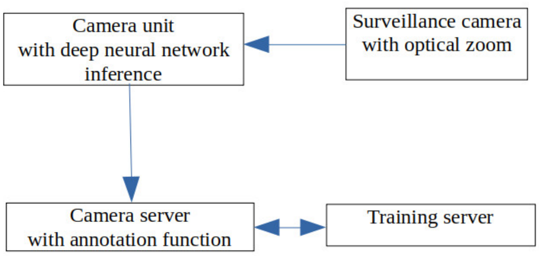

Figure 6: Updated architecture of the second iteration

3 CONCLUSIONS

We presented two iterations of a floating plastic waste early warning system. The second iteration featuring a DNN increased the selectivity of the system significantly generating much less false alarms. We were not entirely satisified with the performance of the algorithm as significant number of relevant items were not recognized and miscategorized but we still consider the second iteration a major leap forward. Due to bandwidth limitations the system can only be deployed in edge computing architectural style which requires relatively powerful GPU in the camera unit in case of the second iteration.

ACKNOWLEDGEMENTS

We thank the support of the PET Kupa initiative [2] which provided us with field experience, countless images and items of floating plastic waste. We also thank the Kalóz Kikötő in Bodrogkisfalud, Hungary that provided space for us to mount our experimental equipment.

REFERENCES

Aytan, U., Pogojeva, M., and Simeonova, A. (2020). *Marine Litter in the Black Sea.*

[2] https://petkupa.hu/

ID	Description	Video recorder	Distance	Frames analysed
#1	Individual plastic waste (mostly bottles) floating down the river bend	Professional camera crew with optical zoom	30-50 meters	300
#2	Mass of mixed floating waste (wood, plastic) captured at a dam	Drone	5-15 meters	50
#3	Mass of mixed floating waste (wood, plastic) captured at a dam	Drone	5-15 meters	30
#4	Mass of mixed floating waste (wood, plastic) near to a river bank	Person filming from a boat	5-15 meters	24
#5	River segment with floating non-plastic waste	Drone	10-20 meters	60

Table 1: Video footages analysed for detection efficiency

ID	TR	TNR	TM	NRT	MR
#1	608	817	12	57.3%	1.97%
#2	1057	603	153	36.3%	14.47%
#3	309	129	31	29.4%	10.03%
#4	277	107	11	27.86%	3.97%
#5	0	0	0	N/A	N/A

Table 2: Results of the video footage analysis

Castro-Jiménez, J., González Fernández, D., Fornier, M., Schmidt, N., and Sempéré, R. (2019). Macro-litter in surface waters from the Rhone River: Plastic pollution and loading to the NW Mediterranean Sea. *Marine Pollution Bulletin*, 146.

Gundupalli, S. P., Hait, S., and Thakur, A. (2017). A review on automated sorting of source-separated municipal solid waste for recycling. *Waste Management*, 60:56–74. Special Thematic Issue: Urban Mining and Circular Economy.

Lechner, A., Keckeis, H., Lumesberger-Loisl, F., Zens, B., Krusch, R., Tritthart, M., Glas, M., and 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, https://www.youtube.com/watch?v=TLyK_aIu3fc.

Redmon, J. and Farhadi, A. (2018). YOLOv3: An incremental improvement.

Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. In Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc.

van Emmerik, T. and Schwarz, A. (2020). Plastic debris in rivers. *WIREs Water*, 7(1):e1398.

van Lieshout, C., van Oeveren, K., van Emmerik, T., and Postma, E. (2020). Automated river plastic monitoring using deep learning and cameras. *Earth and Space Science*, 7(8):e2019EA000960.