POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Executive Summary of the Thesis

Synthetic data augmentation for illegal landfill detection in remote sensing images

Laurea Magistrale in Computer Science and Engineering - Ingegneria Informatica

Author: Federico Gibellini

Advisor: Prof. Piero Fraternali

Academic year: 2022-2023

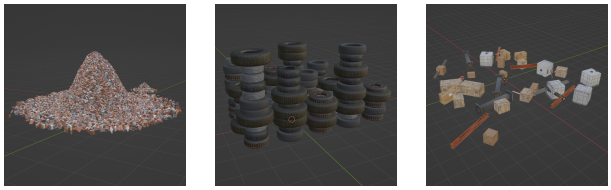
1. Introduction

In the context of environmental risk assessment, illegal landfills represent a serious health threat because of the numerous pollutants they may release, both through leachate and through combustion fumes. Therefore, they must be detected as soon as possible and action must be promptly taken to appropriately dispose of their waste [1]. For this reason, the European Union launched, in December 2022, a 3-year project, named PERIVALLON, aimed at leveraging advanced technologies, such as Artificial Intelligence and Computer Vision, to develop tools and solutions for counteracting this phenomenon and its related crimes [2]. This thesis is part of such project.

Recent studies have demonstrated that it is possible to address the illegal landfill detection problem as a binary classification task on Remote Sensing images, i.e., images collected via satellites or airborne vehicles [1, 3, 4]. However, these approaches, which adopt data-hungry methods such as Convolutional Neural Networks, usually require a significant amount of data, which is easy to obtain for the negative case, whereas it usually incurs prohibitive costs for positive samples. Therefore, it is necessary either to train models with little data, for example adopting weakly or self-supervised approaches [3], or to extend the training set by means of data augmentation strategies [1].

This thesis follows the latter approach and focuses on the practice of synthetic data augmentation. In this context, we designed a pipeline to introduce synthetic waste instances in Remote Sensing images by means of Blender, a popular 3D modelling tool [5]. This pipeline combines geographic and semantic information to determine realistic locations for instance placement and is able to arrange objects in 3 structures, *Heaps*, *Stacks* and *Scattered objects*, for 3 waste categories from the AerialWaste dataset [3], *Tires*, *Rubble/excavated earth and rocks* (henceforth, *Rubble*) and *Bulky items* (Figure 1).

The designed process allows to create synthetic images in averagely $\approx$ 10 seconds per image, with instances appearing in totally realistic locations in 85% of the cases. Furthermore, fine-tuning experiments highlight that extending the training set with synthetic images usually results in increases in F1 and recall, in spite of decreases in accuracy and precision. This result can be accounted as an improvement, when considering that, with the aim of aiding human inspectors in detecting hazardous waste sites, recall might be privileged over other metrics.



(a) *Heap* of *rubble* (b) *Stacks* of *tires* (c) *Scattered bulky items*

Figure 1: Object structures and waste categories

2. Method

The designed pipeline consists of a sequence of phases (Figure 2) to render a Blender virtual scene where instances of waste structures are created and placed on top of a background image. Images are extracted from the AerialWaste dataset [3], a collection of almost 11,000 aerial and satellite images from 3 different sources and suitable for different tasks, such as binary and multi-label classification for waste detection.
The following sections describe in detail the pipeline phases, always assuming the reference coordinate system to be right-handed and following the z-up convention.

2.1. Scene setup

The first phase starts with the introduction of a *Plane* mesh, a simple rectangular mesh with only 4 vertices, coinciding with the rectangle ones. This mesh, which is intended to host the background image as texture, is created as a square with a 210-meter side, since all images from AerialWaste [3] cover a geographic area of these dimensions in spite of their variable resolutions. After applying the texture, an orthographic camera is introduced on the z axis, pointing down towards the plane mesh and located sufficiently distant from the mesh not to intersect its projection plane with any of the eventual instances. Then, a white ambient light is added to the scene and a directional light is introduced in a fixed position, to emulate the effect of sun illumination.
After having defined the location for implanting an instance (Section 2.2) and created the instance itself (Section 2.3), the augmented image is obtained via rendering this scene from the point of view of the orthographic camera.

2.2. Instance placement

This phase aims at identifying the dimensions and rotation of a rectangular area where an instance can be created with the approach described in Section 2.3. This rectangle must be in a feasible region according to a combination of geographic and semantic information about the chosen background image. This information is provided by several inputs, each of which allowing the creation of a binary mask identifying legal and illegal regions. The overall feasible regions are obtained as the intersection of the legal regions of all the input masks, as shown in Figure 3. The Figure also shows that the overall legal regions are then eroded by the minimum required instance size and the center for the output rectangle is searched withing such eroded regions in order to guarantee the rectangle to be completely contained within a legal region.
The following subsections describe the reasons and the methods to create each of the 5 input masks used in the designed pipeline.

2.2.1. Land cover

Land cover information is used to exclude all those areas that never or hardly ever contain garbage in AerialWaste [3] positive samples. This information is extracted from the DUSAF7.0 dataset, publicly available on the online geographic portal of Regione Lombardia [6], the reference governmental institution for the geographic area described in the background images. This dataset is a digital map, with polygons classifying land regions based on their us-

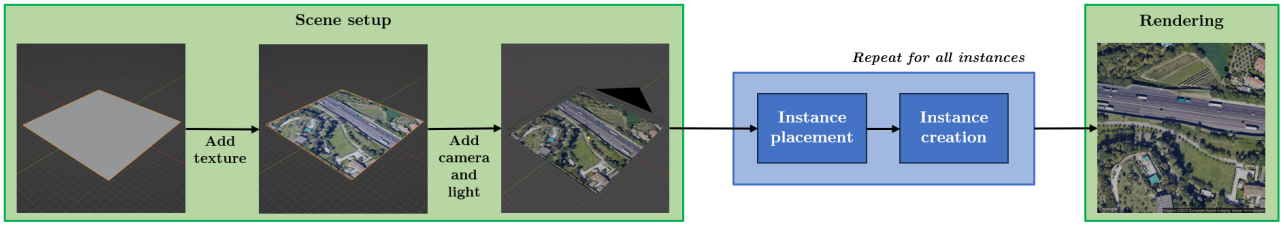


Figure 2: The designed pipeline and its phases, with focus on *Scene setup* and *Rendering*.

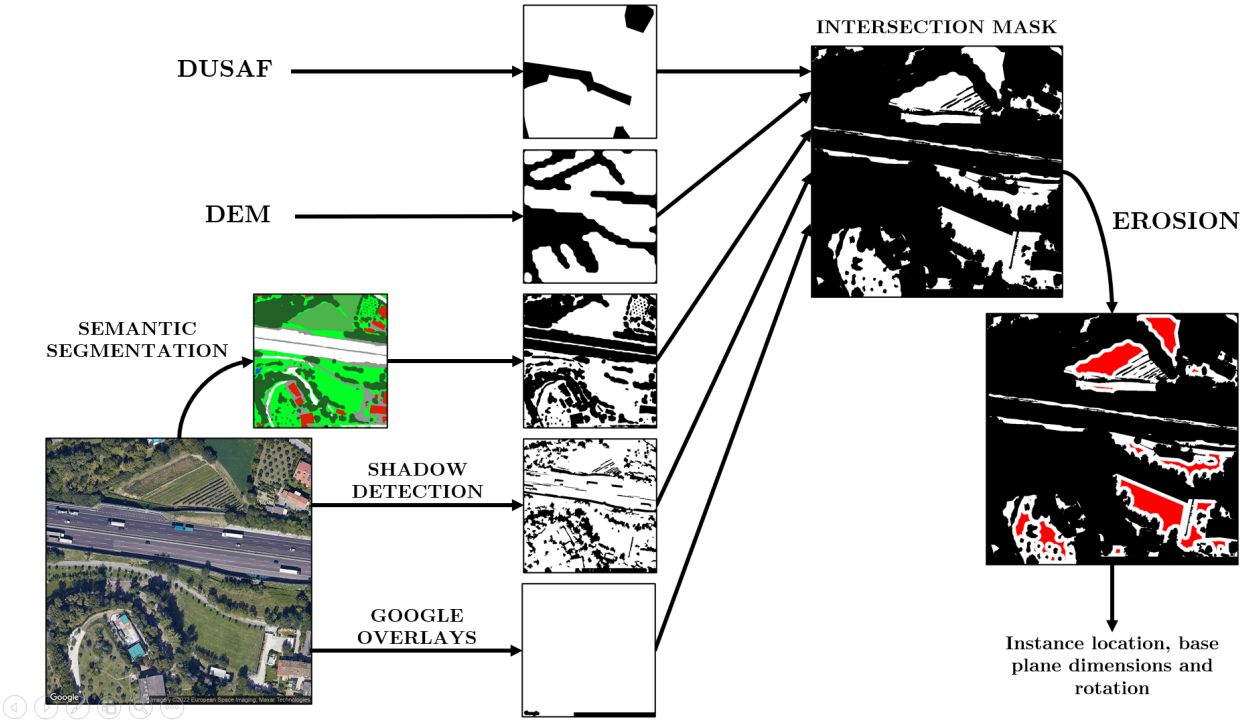


Figure 3: The *Instance placement* phase: input mask combination and effects of erosion on legal regions. The center of the output rectangle is found in the red areas.

age. For each category in this classification, to build this input mask, we extracted the number of positives containing such class and visually inspected all the related samples for categories appearing in at most 100 images. If no or significantly few instances were found, the category was deemed removable and its polygons were considered illegal regions. Table 1 shows a list of all removable land cover classes along with the number of positive images containing them.

2.2.2. Digital Terrain Model (DTM)

Slope information can be used to map as illegal areas with a steep terrain or presenting a sensitive altitude discontinuity, following the notion that waste would be unlikely to remain in such areas being it prone to falling from its position. In order to create a mask for such mapping, the necessary information was extracted from the DTM5x5, the most recent and accurate Digital Terrain Model (DTM) provided by Regione Lombardia through its online geographic portal [6]. This dataset, which consists in a grid of altitude data, allows to compute a slope value for each point of the grid. Such value was chosen to be the maximum among those computed with respect to each of the reference point 8 neighbors. After obtaining this information for all points of the grid portion included in the image, illegal locations were defined to be those with a slope higher than 20°.

2.2.3. Google overlays

The AerialWaste dataset includes images downloaded from Google Earth and, therefore, watermarked with two overlays: a Google logo in the bottom left corner and a copyright box in the bottom right one. Being these overlays clearly illegal image portions for placement, they must be included in a specific mask. Since the logo is always in the same exact position, it can be covered with a manually-built polygon. Instead, for copyright boxes, their top left corner can be identified via template matching on the "Imagery" word, which is always at the beginning of the writing in the box. Then, a rectangle can be drawn given the bottom right corner, which always coincides with the bottom right corner of the whole image.

2.2.4. Semantic Segmentation

A network for semantic segmentation was trained to extract semantic information from the

ID	Name	Occ.
222	Fruit trees and berry plantations	96
511	Water courses	94
12123	Technological units	58
223	Olive groves	50
2242	Other wood cultures	49
31311	High-density coppice mixed forest	33
3223	Bank vegetation	33
5122	Artificial water bodies	27
331	Beaches, dunes, sands	27
31112	High-density tall-tree broad-leaved forest	25
3222	Riverbed vegetation	17
12121	Hospital units	11
411	Inland marshes	11
5123	Water bodies from extractive aquifer activities	11
3121	Mid-high-density coniferous forest	9
5121	Natural water bodies	7
314	Recent reforestations	4
31122	Low-density tall-tree broad-leaved forest	3
332	Bare rocks	3
3111	High-density broad-leaved forest	3
3114	Chestnut groves	2
31312	High-density tall-tree mixed forest	2
3112	Low-density broad-leaved forest	1
333	Sparsely vegetated areas	1
2313	Water meadows	1
3122	Low-density coniferous forest	0
31321	Low-density coppice mixed forest	0
3211	Natural grasslands without arboreal or shrub-like species	0
3212	Natural grasslands with arboreal or shrub-like species	0
123	Port areas	0

Table 1: Removable land cover categories. The last column displays the number of positive images (occurrences) containing a specific category

background images and discard frequent locations where a garbage instance would be totally unrealistic. The network was trained on OpenEarthMap [7], a dataset for land cover mapping containing 5000 sub-meter-resolution images from all over the world and manually annotated with 8 classes. Among these, the classes of *Road*, *Tree*, *Water* and *Buildings* were deemed to identify illegal regions.

2.2.5. Shadow detection

The mask for shadow detection was inserted to avoid positioning instances across light and shadow areas, which would result in a rendering artifact motivated by the impossibility of illuminating only a portion of an instance. The adopted shadow detection algorithm is an adjusted version of the one presented in [8] and centered on the computation of the *SSSI* index, a value describing, for each pixel, its likelihood of belonging to a shadow on the basis of its spatial and spectral features. In our version, such index is computed as:

$$SSSI = \frac{PC1 + B}{R + G + 1} \tag{1}$$

where *PC1* is the first principal component from *PCA*, whereas R, G and B are the 3 image channels. Then, the index is averaged across the image on a cluster basis and a relative threshold is applied to the range of averaged values. After visually inspecting the results, the threshold was set to 60% of the index value range, thus defining as shadow and, therefore, illegal pixels with a value higher than such threshold.

2.3. Instance creation

In this phase, waste object models are arranged in either of 3 structures, *Heaps*, *Stacks* and *Scattered objects*, thanks to *Geometry nodes*, a Blender feature allowing to procedurally modify an object geometry. Specific configurations of this feature also allow to cover a mesh with other 3D models. In our pipeline, the chosen mesh is a rectangle whose dimension, rotation and scene position are determined with the method described in Section 2.2.

Even though, eventually, the process for generating any structure converges into applying a *Geometry node* configuration as a mesh modifier, such process differs based on the type of structure to create, as shown in Figure 4. In particular, for the cases of *Stacks* and *Scattered objects*, this generation procedure is extremely simple and consists in adding to the scene a *Plane* mesh and applying to it the modifier. Instead, for the *Heap* case, Blender does not natively provide a base mesh with hill-like protrusions, which are necessary to obtain anything similar to a heap.

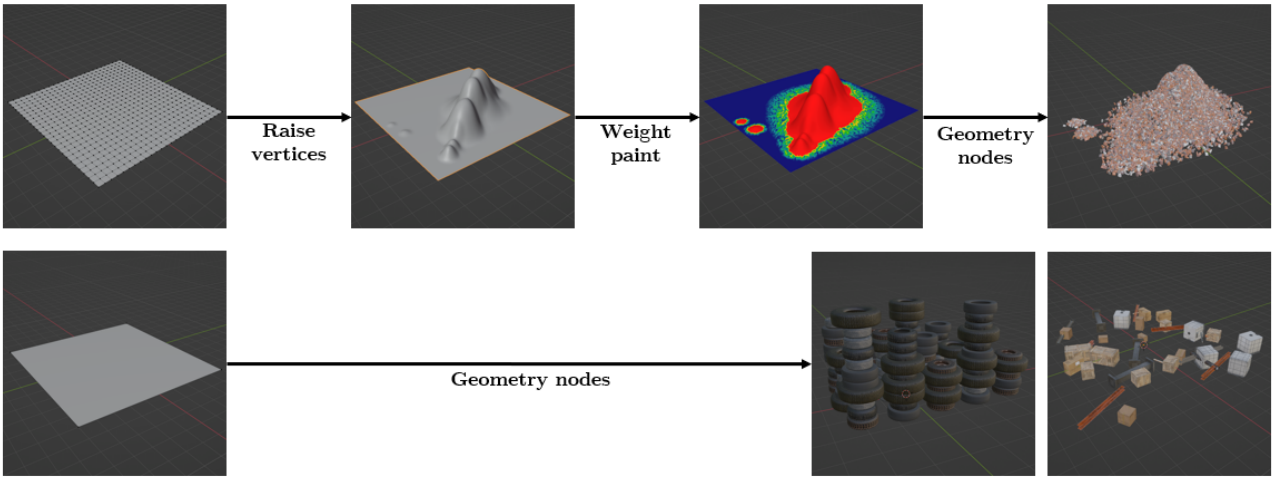


Figure 4: Representation of the instance creation phase. Top line: the procedure to create *Heaps*. Bottom line: the procedure for *Stacks* and *Scattered objects*.

Therefore, a pre-modelling procedure has to be conducted. The base mesh imported in the scene is the *Grid*, i.e., a rectangle with vertices also on the inside. Then, a given number of such vertices is randomly selected and, for them, the z coordinate is changed to a positive value, thus moving the selected vertices higher and forming protrusions. Subsequently, the mesh is applied a *Weight paint*, which allows to cover with models only the just-created protrusions when integrated in the *Geometry node* configuration. After these pre-processing steps, the mesh is ready to receive the modifier.

3. Experiments

Experiments were performed to evaluate the designed pipeline and its output synthetic images. In some of these experiments, instance generation was additionally constrained with limitations in location or dimensions. In particular, instances could be forced to locate in a central portion of the image or to have a greater size. Furthermore, in order to ensure realism in the created instances, the implanted combinations of structure and waste types were restricted in all experiments: *Tires* can be present in *Stacks* or *Heaps*, whereas *Rubble* can only appear in *Heaps* and *Bulky items* can only be *Scattered*.
The first set of experiments aimed at evaluating computation efficiency of the proposed approach by calculating an average augmentation time per image. For these experiments, 4 sets of 100 images each were created, with 3 standard-sized instances per image. Three of these sets contained instances of a single waste type, whereas the last one contained waste from all categories. The computed averages highlight that *Rubble* instances are accountable for most of the computational effort, costing 3 to 4 times more than other instances. This behavior can be mainly attributed to the *heap* generation procedure, since it affects also *Tire* instances. Results also show that, overall, an image can be created in averagely ≈ 10 seconds.
Then, the location placement process was qualitatively evaluated, with an experiment involving a set of 100 synthetic images containing a total of 315 standard-sized instances. These instances were visually inspected and classified in 3 categories based on their location. This classification revealed that only less 4.5% of all images were positioned in *illegal* locations, i.e., where input masks should prevent placement, thus highlighting mask flaws. In addition, 10.5% was in a *borderline* position, i.e., in legal regions according to the inputs without being totally physically feasible, as in the case of instances crossing shadow and light areas. Finally, 85% of the instances were placed in totally *legal* areas.
Following studies shifted attention to the behavior of augmented images with relation to the network proposed in [4], for which a model trained on AerialWaste [3] is publicly available.
Synthetic images were further evaluated by inference with such model. For this experiments, 4 sets of 100 images each were generated, differing in instance size and location and exploring all

the available combinations given the pairs *centered/not-centered* and *standard-sized/greater-sized*. Results show that the network seems to be biased towards greater and centered instances, even though performing poorly in all cases.
Finally, the model was involved in a set of fine-tuning experiments, where the AerialWaste [3] training set was extended a variable number of synthetic images, though leaving the validation and test set unaltered and composed of only real images. Furthermore, the training set was extended with multiple sets of synthetic images to analyze, once more, the effect of the 4 combinations of instance constraints. These experiments highlight that fine-tuned models tend to have, with respect to the baseline model, higher F1 and recall metrics, while showing lower accuracy and precision. The best results were obtained by extending the training set wit 500 synthetic images with greater and centered instances.

4. Conclusions

This thesis addressed the problem of illegal landfill detection as a binary classification task, aiming to solve its major issue of positive sample scarcity with a method for synthetic data augmentation. This method, which consists in a totally-automated pipeline for implanting various waste structures in Remote Sensing images, allows to artificially create positive samples with multiple landfill instances in averagely ≈ 10 seconds per image. This pipeline also combines geographic and semantic information from each image to define the most realistic location for instance placement, eventually positioning less than 4.5% of all instances in unfeasible areas.
The effectiveness of the generated images was evaluated via some fine-tuning experiments on the network proposed by Torres et al. [3], for which a model trained on the AerialWaste dataset is publicly available. These experiments highlighted that extending the training set with synthetic images usually results in a minor increase in F1 score and a significant one in recall. This result can be considered an improvement, in spite of a minor decrease in accuracy, when considering the eventual practical aim of the trained models, i.e., to aid human inspectors in detecting seriously dangerous landfills.
Given the potential of the proposed method, future work can proceed with its extension to other waste categories or to other tasks, such as multi-label classification and segmentation, even though still requiring improvements in photo-realism and input accuracy.

References

[1] C. Padubidri, A. Kamilaris, and S. Karatsiolis, "Accurate detection of illegal dumping sites using high resolution aerial photography and deep learning," in *2022 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*, pp. 451–456, IEEE, 2022.

[2] "Introduction to PERIVALLON project." `https://perivallon-he.eu/introduction-to-perivallon-project-2/`. Accessed: 2023-10-30.

[3] R. N. Torres and P. Fraternali, "AerialWaste dataset for landfill discovery in aerial and satellite images," *Scientific Data*, vol. 10, no. 1, p. 63, 2023.

[4] R. N. Torres and P. Fraternali, "Learning to identify illegal landfills through scene classification in aerial images," *Remote Sensing*, vol. 13, no. 22, p. 4520, 2021.

[5] "blender.org - Home of the Blender project - Free and Open 3D Creation Software." `https://www.blender.org`. Accessed: 2023-11-05.

[6] "Home - Geoportale della Lombardia." `https://www.geoportale.regione.lombardia.it`. Accessed: 2023-11-09.

[7] J. Xia, N. Yokoya, B. Adriano, and C. Broni-Bediako, "OpenEarthMap: A Benchmark Dataset for Global High-Resolution Land Cover Mapping," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, January 2023.

[8] M. Kakooei and Y. Baleghi, "Shadow detection in very high resolution rgb images using a special thresholding on a new spectral–spatial index," *Journal of Applied Remote Sensing*, vol. 14, no. 1, pp. 016503–016503, 2020.