Using computer vision to recognize composition of construction waste mixtures: A semantic segmentation approach

Article *in* Resources Conservation and Recycling · March 2022

DOI: 10.1016/j.resconrec.2021.106022

CITATIONS
49

READS
923

3 authors:

Weisheng Lu
The University of Hong Kong
346 PUBLICATIONS **12,478** CITATIONS
SEE PROFILE

Junjie Chen
The University of Hong Kong
50 PUBLICATIONS **703** CITATIONS
SEE PROFILE

Fan Xue
The University of Hong Kong
159 PUBLICATIONS **4,554** CITATIONS
SEE PROFILE

Using computer vision to recognize composition of construction waste mixtures: A semantic segmentation approach

Weisheng Lu, Junjie Chen *, and Fan Xue
Department of Real Estate and Construction, The University of Hong Kong, Pokfulam, Hong Kong, China

Abstract

Timely and accurate recognition of construction waste (CW) composition can provide yardstick information for its subsequent management (e.g., segregation, determining proper disposal destination). Increasingly, smart technologies such as computer vision (CV), robotics, and artificial intelligence (AI) are deployed to automate waste composition recognition. Existing studies focus on individual waste objects in well-controlled environments, but do not consider the complexity of the real-life scenarios. This research takes the challenges of the mixture and clutter nature of CW as a departure point and attempts to automate CW composition recognition by using CV technologies. Firstly, meticulous data collection, cleansing, and annotation efforts are made to create a high-quality CW dataset comprising 5,366 images. Then, a state-of-the-art CV semantic segmentation technique, DeepLabv3+, is introduced to develop a CW segmentation model. Finally, several training hyperparameters are tested via orthogonal experiments to calibrate the model performance. The proposed approach achieved a mean Intersection over Union (mIoU) of 0.56 in segmenting nine types of materials with a time performance of 0.51s per image. The approach was found to be robust to variation of illumination and vehicle types. The study contributes to the important problem of material composition recognition, formalizing a deep learning-based semantic segmentation approach for CW composition recognition in complex environments. It paves the way for better CW management, particularly in engaging robotics, in the future. The trained models are hosted on GitHub, based on which researchers can further finetune for their specific applications.

Keywords: Construction and demolition waste; Waste composition; Construction waste management; Artificial intelligence; Computer vision; Semantic segmentation.

* Corresponding author.
E-mail address: chenjj10@hku.hk (J. Chen).

1. Introduction

The extensive construction activities in the past few decades have significantly improved our quality of life by materializing buildings and infrastructure. However, construction has also resulted in the skyrocketing amount of construction waste (CW), or referred to as construction and demolition (C&D) waste. In Europe, for example, the construction sector produces 820 million tonnes of wastes annually, accounting for 46% of the total waste streams (Gálvez-Martos et al., 2018; Ku et al., 2020). In 2015, the United States generated 548 million tons of C&D debris, which is more than twice the amount of generated municipal solid waste (USEPA, 2018). In 2016, the United Kingdom generated 66.2 million tonnes of non-hazardous C&D waste, of which 91.0% was recovered (Defra, 2020). In 2019, the generation of CW in Hong Kong (HK) has doubled since 2008, hitting nearly 18 million tonnes per annum (HKEPD, 2020). The mountainous CW calls for better waste processing and management.

Information on the composition of CW is a prerequisite for its proper processing and management (HKEPD, 2019; NSWEPA, 2020). On the one hand, it is of significant value for the operation of construction waste management (CWM) schemes. For example, in Hong Kong, CW is categorized as inert (e.g., concrete, bricks, and sand) or non-inert (e.g., bamboo, wood, and plastics), and the composition of a waste truckload determines which facility will accept it and the levy chargeable (Chen et al., 2021; Lu and Yuan, 2021). On the other hand, the composition information can also be used to enable automated waste segregation (Gundupalli et al., 2017b). With the ability to recognize specific material types, positions, and dimensions, it is viable to replace human workers with intelligent robots to sort CW materials automatically. Properly employed, robots can yield a higher throughput, reduce occupational hazards, and enable production of better-quality recycled materials (Toto, 2019).

Many waste composition sensing technologies have been developed, among which computer vision (CV) stands out for its cost effectiveness, ease of maintenance, and applicability to a wide range of materials. Different waste materials have their unique physicochemical properties regarding absorbance, colors, etc. Such differences in photometric characteristics have determined the different appearance of various waste categories, making them distinguishable through visual recognition. For many years, CV has been explored for the recognition of municipal solid waste (MSW) composition. For example, support vector machines (SVM) (Paulraj et al., 2016), AlexNet (Mittal et al., 2016), and region-based convolutional neural network (R-CNN) (Nowakowski and Pamuła, 2020) have been used to recognize or detect MSW materials at source. At the waste sorting stage, extensive research efforts have been made to incorporate image classification/detection techniques so that robotic systems can automatically segregate waste materials (Mao et al., 2021; Vrancken et al., 2019; Yang and Thung, 2016; Zhang et al., 2021).

Despite the considerable progress achieved, existing studies have been limited to the recognition of MSW in a relatively structured environment. Transfer of such technologies to CWM scenarios in natural settings is difficult for several reasons. Firstly, CW usually comprises a mixture of intermingled bulky materials. The image classification/object detection techniques used in existing studies may not be able to provide composition information with sufficient granularity. Secondly, while MSW is usually processed in orderly indoor facilities, CW segregation is generally performed in complex outdoor environments with variant illumination and a cluttered background. Such complexities and the mixture nature of CW pose great challenges to the composition recognition. Existing studies tend to oversimplify the application environments, aiming only at classifying or detecting individual waste items appearing against a simple, unified background (Huang et al., 2020; Meng and Chu, 2020; Yang and Thung, 2016). Much remains unclear how CV performs in recognizing composition of CW mixtures in complex, cluttered real-life environments.

Semantic segmentation, as a specific CV technique, is promising to the recognition of CW composition. Compared with image classification or object detection, semantic segmentation delivers finer granularity by performing classification in a pixelwise manner (Mansouri, 2019). It can not only detect and identify objects of interest from a relatively unstructured and complex background, but also provide detailed information about the geometry and boundaries of the objects. However, little research attention, if any, has been paid to the application of semantic segmentation in recognizing CW composition. Having been used primarily for objects with explicit structures, it is unclear whether the semantic segmentation technique can recognize CW composition due to the mixed state of CW materials and, if it can, what methodologies should be used to prepare the corresponding dataset and train and calibrate the model.

To fill in the knowledge gap, our study aims to provide a semantic segmentation approach for the recognition of CW mixture in complex and cluttered real-life environments. Through meticulous model training and calibration, and empirical analysis, our study demonstrates for the first time the viability of a deep learning-enabled CV model for segmenting highly unstructured CW in complex environments. The proposed method lays the foundation for future applications such as robotic waste segregation.

2. Literature review

2.1 Computer vision in waste management

Studies have been ongoing for many decades to apply CV in waste management (Gundupalli et al., 2017b). By recognizing waste materials via visual sensors such as cameras, robots can be deployed to automatically segregate desired items from waste streams transported by conveyor belts. Such ideas can be traced back to early 2000s or even before, when Faibish et al. (1997) presented a robotic system with stereo vision to recycle waste paper and Mattone et

al. (2000) proposed a solution for waste packaging classification based on optical sensors. In its early days, visual recognition of waste material relied heavily on hand-engineered features fed to machine learning models such as multilayer perceptron (MLP) (Koyanaka and Kobayashi, 2011), SVM (Wang et al., 2019b), and nearest neighbor (Gundupalli et al., 2018) to reduce problem complexity. However, performance of these models was in general not robust enough to adapt to real-life waste material variations.

The situation has improved since 2012 with the success of end-to-end deep learning techniques (Krizhevsky et al., 2012) made possible by drastically improved computing power and big data. Yang and Thung (2016) applied AlexNet, a deep convolutional neural network (CNN), for the classification of MSW such as paper, glass, and cardboard. Based on their TrashNet dataset (Thung and Yang, 2019), a series of studies have been carried out to significantly improve classification performance from the initial accuracy of below 70% (Yang and Thung, 2016) to over 90%. For example, Mao et al. (2021) improved DenseNet121's classification accuracy on TrashNet by the application of a genetic algorithm for hyperparameter optimization. Zhang et al. (2021) proposed a residual network with a self-monitoring module for recyclable waste classification on the same dataset. Researchers have also tried to locate and detect multiple types of waste material in images. Awe et al. (2017) trained a Fast R-CNN model on TrashNet, achieving a mean average precision of 0.683. Mittal et al. (2016) developed an Android application based on AlexNet that can automatically detect and localize garbage in photographs.

In the construction industry, the potential of CV is receiving increasing attention in CWM. Xiao et al. (2020) compared the performance of extreme learning machine and CNN in classifying typical CW including wood, rubber, bricks and concrete. Lau et al. (2020) developed a (near) real-time image recognition approach based on CNN for the determination of recycled aggregate composition. Ku et al. (2020) proposed a deep learning-based grasping detection method able to decide the optimal grasping pose for robots sorting detected CW materials. Wang et al. (2020; 2019a) presented deep learning models based on Faster R-CNN and Mask R-CNN to detect and segment nails and screws on construction sites. Lukka et al. (2014) and Kujala et al. (2015) presented robotic systems for the segregation and sorting of CW using CV to tackle the issues of material classification and object grasping.

2.2 Strengths and weaknesses

Despite their impressive progress, existing studies may have oversimplified the working conditions of their models. For example, TrashNet, the widely used dataset in the research community (Huang et al., 2020; Meng and Chu, 2020), contains only photos of individual items of MSW against a simple background. However, real-life contexts are more complicated, as the waste materials are always randomly mixed, the illumination is constantly changing, and the background can be cluttered. This is especially true for CW, which is a

mixture of different materials, and is usually processed in the complex outdoor environments (Lu and Yuan, 2012).

In addition, little attention has been paid to the problem of waste segmentation. Compared with classification or detection, semantic segmentation can provide spatial geometry of waste materials at a higher level of granularity, enabling better solutions for composition measurement and robotic sorting. However, while some studies have touched on the waste segmentation problem, they either relied on hand-engineered features (and thus have low robustness and generalizability) (Gundupalli et al., 2017a, 2018), or required input of additional data modality such as depth information and X-ray imagery (Lukka et al., 2014; Zhu et al., 2018), or only focused on small, separate waste items (Sun et al., 2019; Wang et al., 2020).

Semantic segmentation provides opportunities to address the challenges of CW composition recognition. In 2015, Long et al. (2015) introduced CNN to the problem of semantic segmentation by implementing pixelwise classification. Since then, numerous model structures have emerged including U-Net (Ronneberger et al., 2015), FC-DenseNet (Jégou et al., 2017), DeepLab (Chen et al., 2017a; 2017b; 2018), and Mask R-CNN (He et al., 2017). Among them, the DeepLab series is reputable and widely accepted for its solid performance and relatively simple rationale (Li, 2020). The model series proposed groundbreaking techniques such as atrous convolution, and atrous spatial pyramid pooling (ASPP), and have incorporated emerging algorithms such as multi-scale input and encoder-decoder structure to improve accuracy. In 2018, the DeepLabv3+ achieved the state of the art on the PASCAL VOC 2012 dataset.

DeepLab has been applied in diverse domains, including structural condition assessment (Wu et al., 2019; Xu et al., 2020), medical image analysis (Xiao et al., 2018), and autonomous vehicles (Capellier et al., 2018). However, most of these applications have focused on the extraction of individual objects with relatively explicit and clear structures, or “things” (Lin et al., 2014), rather than the segmentation of materials, or “stuff”. It remains unclear whether existing algorithms such as DeepLab can be customized and re-calibrated to segment a highly unstructured, cluttered mixture of different materials, where the structure or distribution is not always clear or cannot even be explicitly represented, from complex environments. Our research fills this gap by formalizing an approach based on DeepLab for effective CW semantic segmentation.

3. Methodology

3.1. Preparing a big dataset of construction waste images

For deep learning, data preparation is far from a trivial task. If the data is not big enough,

useful patterns and generalizable features may be overwhelmed by noisy features. If the data quality is low (e.g., image noise, imbalanced class, and bad annotation), the trained segmentation models will produce biased and unsatisfactory results.

3.1.1. Data collection

To obtain a large collection of mixed CW images in complex outdoor environments, the research team has engaged with the Hong Kong Environmental Protection Department (HKEPD). Since 2006, the HKEPD has operated a construction waste disposal charging scheme. Sensing systems are deployed in all government waste disposal facilities to measure composition of incoming waste loads (Chen et al., 2021; Lu and Yuan, 2021). The sensed data includes not only regular physical indexes such as weight and depth of the waste loads, but also top-down photos captured by cameras.

A dataset of 5,366 photos of waste loads was obtained in the month of October 2019 from one waste disposal facility. The photos were taken by cameras (DS-2CD2025FWD-IHONG KONG 4mm from Hikvision Digital Technology) at the entrance toll gates of the facility, above which a steel ceiling is in place to protect the cameras and truckloads from undesired weather conditions such as raining. The cameras were installed at the upper-rear direction of the trucks. The dataset includes photos taken both during the daytime and at night, and during night operation artificial light sources are used to illuminate the wastes in the truckloads. Fig. 1 (a) shows some example photos of CW loads, clearly depicting the natural cluttered state in which different types of waste materials are randomly mixed in truck buckets. To ensure the robustness of subsequent model training, our photo collection has to reflect the complexity and variation of real-life applications, which mainly come from two sources: (a) illumination and (b) vehicle types. The former is a notable influencing factor for computer vision applications, and as for the latter, the variant appearance and size of different vehicle types directly impact the segmentation performance of construction wastes contained inside.

The hue, saturation, brightness (HSB) color model was used to quantify illumination variation as it can reflect lighting intensity via a separate channel, i.e., the B (brightness) channel, without the influence of hue (H) or saturation (S). This study used the average value of the B channel of an image to characterize its brightness level. Fig. 1 (b) shows the distribution of images over different levels of brightness, revealing a wide range. When the brightness level is below 98, the truckloads are usually in shadow, significantly reducing the illumination. Images with such low brightness account for nearly 25% of the dataset, indicating that it effectively represents the complexity of real-life lighting conditions. Vehicle type is another complex factor as many different types of vehicles are used to carry CW in Hong Kong. Fig. 1 (c) shows the distribution of the dataset by truck tonnage. A larger tonnage indicates a larger dump bucket, and the corresponding truck tends to have a grip mounted onboard (e.g., #2, #4, #7, and #8 in Fig. 1). Fig. 1 (c) shows that the dominant

proportion of trucks are 22~24 tonnes, but also demonstrates a wide coverage of the dataset over other truck tonnages (e.g. 4~6, 8~10, and 36~38 tonnes).

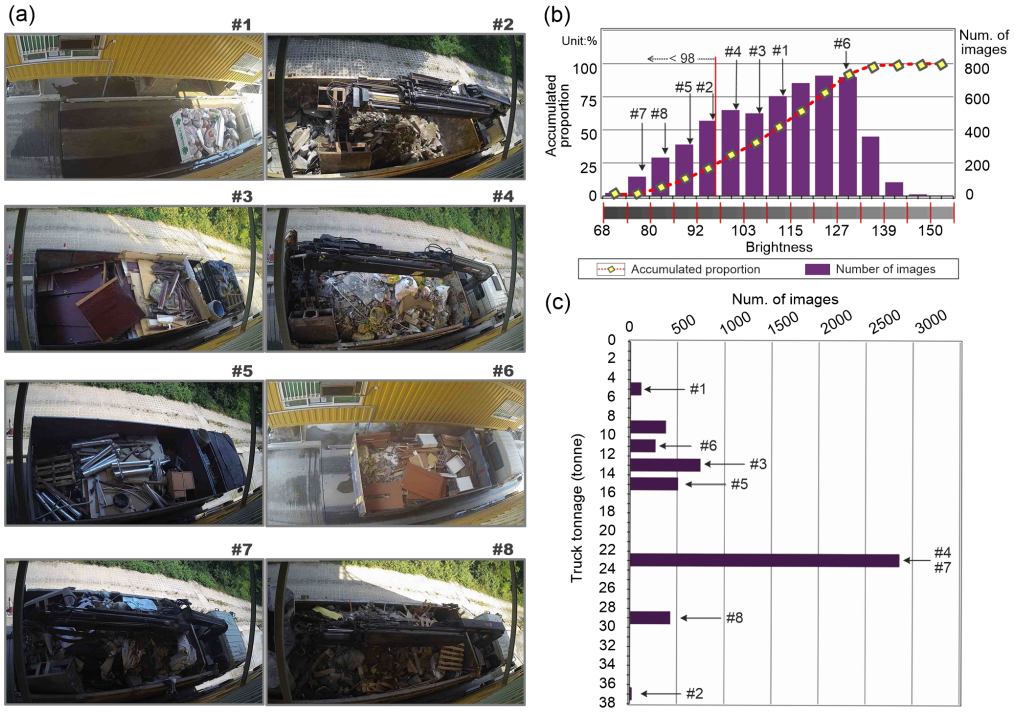

Fig. 1. (a) Example images from the collected dataset, and distribution of the dataset over (b) different levels of brightness and (c) truck tonnage. The corresponding brightness and truck tonnage of images in (a) have been annotated in (b) and (c).

3.1.2. Data preprocessing and cleansing

Data preprocessing and cleansing were implemented to ensure quality of the dataset. First, the images were uniformly rescaled by a factor of k ($\leqslant 1$). While the original high resolution (1920*1080) could potentially ensure segmentation precision by maintaining sharp edges and detail, it would also consume more computation resources making the processing of each image tedious and lengthy. To achieve a compromise between accuracy and computation time, different scaling factors were tested (see Section 4). Second, it was observed that waste materials in the dump buckets became nearly indistinguishable at a brightness below 85. Therefore, samples with extremely low average brightness were removed, resulting in a dataset of 5,022 samples.

3.1.3. Data annotation

For the task of semantic segmentation, data annotation is required to assign image pixels to their corresponding categories. Given the dataset has more than 5,000 images, the research team outsourced this task to professional annotators via the Taobao platform. The annotators were provided with clear guidelines on what objects to label and what the objects look like.

In the CWM system of HK (HKEPD, 2011), construction waste materials can be classified into two major categories, i.e., inert and noninert materials, under which classification can be further extended to greater granularity. For example, inert materials include debris, concrete, earth, etc., and noninert materials include timber, packaging, vegetation, etc. However, after a pre-screening of the dataset, it was found that the presence of certain waste types (e.g., vegetation, bamboo, and bitumen) is rare. Annotating such infrequent waste types in ultra-high granularity requires tremendous resources and efforts while does not provide too much benefits in practice. Therefore, after considering the common local CW types and the annotation operability, nine materials/objects of interest were determined, as shown in Table 1. The classification system not only ensures class labels to be annotated encompass the major waste types under the large categories of inert and noninert materials, but also allow sufficient flexibility to incorporate infrequent waste materials into the type of "other non-inert" or "mixed" wastes. For the annotation tool, the workers were asked to use Python Labelme (Wada, 2019). This required them to draw polygons around the profiles of the materials/objects of interest and select their corresponding class labels.

Table 1. Specification on CW materials and related objects to be annotated.

No.	Types of material/object	Description of the material/object
1	Rock/Stone/Rubble/Debris	✧ Stones, concrete slabs, rocks, and rubble and debris ✧ Tend to be large and bulky
2	Gravel/Concrete/Bricks	✧ Concrete aggregates, gravel, and bricks ✧ Tend to be medium-sized
3	Earth/Slurry/Mud	✧ Sand, clay, and granules ✧ Tend to be small particles
4	Packaging/Fabric/Plastic	✧ All types of bags (plastic, fiber, etc.) ✧ Nylon paulin
5	Wood/Cardboard	✧ Furniture made of wood ✧ Wood board ✧ Boxes made of paper or cardboard
6	Other non-inert	✧ Other non-inert waste (e.g., chairs, appliances, and ladder)
7	Mixed	✧ A mixture of inert and non-inert waste that is difficult to distinguish
8	Grip	✧ The grab bucket of the truck
9	Truck	✧ The part of the truck bucket that is not covered by waste ✧ The front of the truck

To ensure annotation quality, a research team member performed spot checks daily. Annotation results from that day were randomly selected to check for errors or unqualified

labeling. If found, those samples and other images with similar errors were returned for correction and then re-checked. After a month of hard work by 10 annotators, the annotation results were obtained. Fig. 2 (a) and (b) show the quantity of pixels and the number of images over different material/object categories, respectively. One might have noticed that the dataset is highly imbalanced over different classes. The data imbalance may cause the trained model overfit to the majority classes while perform poorly on the minorities. The negative effects can be alleviated by widely-used techniques such as "class weight".

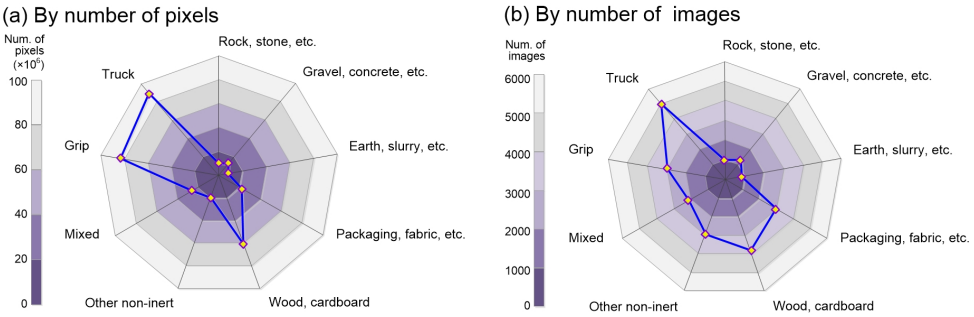

Fig. 2. Distribution of dataset over different materials/objects: (a) by number of pixels; (b) by number of images.

3.2. Developing a construction waste semantic segmentation model

The next step was to develop a CW semantic segmentation model based on a suitable deep learning architecture. There are many off-the-shelf deep semantic segmentation architectures (e.g., U-Net and Mask R-CNN), among which DeepLabv3+ has demonstrated its superiority by achieving the state of the art on the PASCAL VOC 2012 dataset (Chen et al., 2018). Hence, this study adopted DeepLabv3+ as the main architecture of our CW segmentation model.

As shown in Fig. 3, the model comprises an encoder and a decoder. The encoder includes two modules: the backbone network and the atrous spatial pyramid pooling (ASPP) module. The backbone performs initial feature extraction with prevalent CNN structures such as Xception (Chollet, 2017) and ResNet (He et al., 2016). Based on the feature maps provided by the backbone, the ASPP module extracts features at different scales by the so-called atrous convolution with different rates r_i (i = 1, 2, 3), e.g., r_1=6, r_2=12, r_3=18, which are then concatenated with the result of 1×1 convolution (or "conv" for short) operation and the global average pooling feature to achieve scale invariance (Chen et al., 2018). In the decoder, the ASPP output and low-level features retrieved from the backbone are concatenated. The output is gradually increased to the resolution of the original input image by the operations of convolution and upsampling, which finally gives the predicted segmentation result.

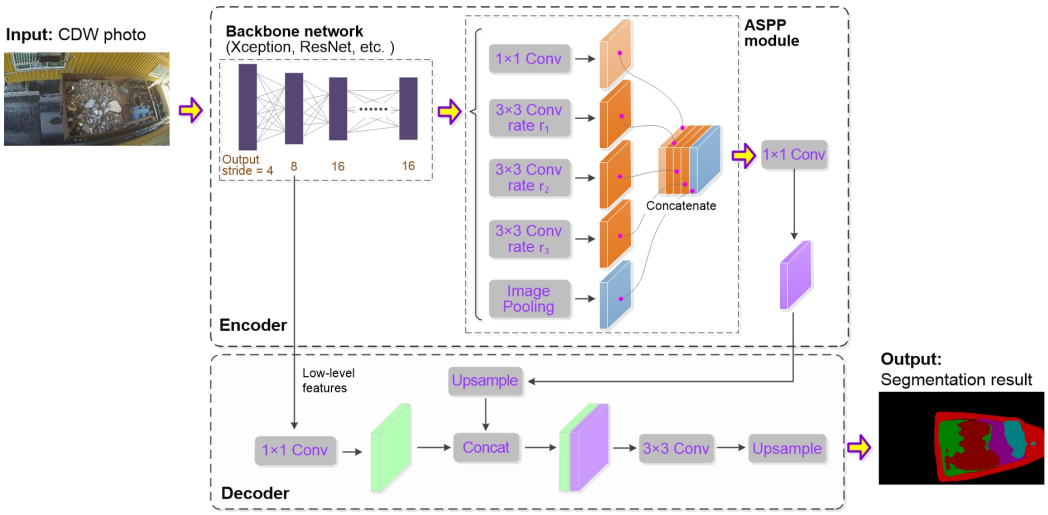

Fig. 3. Structure of the CW semantic segmentation model.

3.3. Training and calibrating model on the big dataset

With the collected dataset, the CW semantic segmentation model can be trained. However, there are several hyperparameters during model training (as will be listed in Section 3.3.1) that can affect the final segmentation performance. Such parameters must be meticulously calibrated and fine-tuned until satisfactory results are obtained. To this end, measurable metrics are required to quantify the model effectiveness, as will be introduced in Section 3.3.2.

3.3.1. Hyperparameters to calibrate

In training DeepLab, several hyperparameters need to be specified. By referring to the training protocol used in the original papers (Chen et al., 2014, 2017a; 2017b; 2018), the research team formed a list of hyperparameters to be calibrated.

- *Backbone network*. The selection of backbone is critical because it directly affects computation complexity and feature extraction. A powerful backbone can extract useful image features in less computation time. This study will evaluate the effectiveness of two reputable CNNs for our model's backbone: Xception and ResNet.
- *Pretrained dataset*. The publicly available models in (Zhu et al., 2020c) were typically pretrained on ImageNet and (or) MS-COCO, which are mainly for the segmentation of objects (or "things"). It would be interesting to see if a model pretrained on a material dataset, such as MINC (materials in context) (Bell et al., 2015), could lead to better CW segmentation performance.
- *Multi-scale (MS) and left-right flip (LR) input*. In the original implementation (Chen et al., 2017b; Li, 2020), higher accuracy was obtained when MS and LR were applied to process the input images at the inference stage. Our study will also investigate whether

the same operation can lead to CW segmentation performance improvement.

- *Output stride (OS)*. OS refers to the ratio of input image resolution to the output resolution of the backbone (Li, 2020). A lower OS can provide feature maps with finer details but will take more time to process. This study will compare results of two different OS, i.e., OS=16 (the corresponding APSS rates r_i are 6, 12, 18), and OS=8 (the corresponding APSS rates r_i are 12, 24, 36).
- *Image resolution*. While higher resolution images reveal sharper material details, which presumably can lead to higher segmentation precision, they also require more time to process. To trade off accuracy and time performance, the influence of different scaling ratios k (i.e., k = 1, 0.521, 0.267) will be investigated.

3.3.2. Evaluation metrics

The model performance is evaluated from two aspects, i.e., segmentation accuracy and time consumption. For time consumption, as quick response is important at the deployment stage, the average inference time per image was used as the evaluation metric. For segmentation accuracy, this study adopts the following four metrics to ensure the comprehensiveness of the evaluation:

$$\mathrm{mIoU} = \frac{1}{N}\sum_{i=1}^{N}\frac{\#TP_i}{\#TP_i + \#FP_i + \#FN_i} \tag{1}$$

$$\mathrm{mF1} = \frac{1}{N}\sum_{i=1}^{N}\frac{2\#TP_i}{2\#TP_i + \#FP_i + \#FN_i} \tag{2}$$

$$\mathrm{mP} = \frac{1}{N}\sum_{i=1}^{N}\frac{\#TP_i}{\#TP_i + \#FP_i} \tag{3}$$

$$\mathrm{mR} = \frac{1}{N}\sum_{i=1}^{N}\frac{\#TP_i}{\#TP_i + \#FN_i} \tag{4}$$

where mIoU, mF1, mP, and mR are, respectively, the mean of intersection over union (IoU), F1-score, precision, and recall across all classes; N is the total number of classes to be labeled; $\#TP_i$, $\#FP_i$, and $\#FN_i$ are, respectively, the total number of true positive (TP), false positive (FP), and false negative (FN) pixels for class i over all images in the dataset. The higher the metrics, the better the segmentation accuracy. The upper limit is 1, which signifies a complete alignment of the predicted composition with the ground truth (i.e., the annotated composition). Among the four metrics, mIoU is the most commonly used index for semantic segmentation evaluation.

4. Experimental studies

The CW segmentation models were trained and calibrated on Amazon Web Services (AWS) Sagemaker p2.xlarge instances with a NVIDIA K80 GPU with 12 GB memory and

p3.2xlarge instances with a NVIDIA Tesla V100 GPU with 16 GB memory. This study accepted the default training protocols used in (Aquariusjay and Zhu, 2019; Chen et al., 2017b), but the hyperparameters listed in Section 3.3.1 were calibrated by us. Among important settings, a "poly" decay policy was used to designate learning rate, and the activation and loss functions were specified as "softmax" and "cross entropy", respectively; to counteract the negative effects induced by the imbalanced dataset, this study assigned weights to different classes inversely proportional to their respective frequencies when calculating the loss; due to limited GPU memory, the pretrained batch-norm weights (i.e., set fine_tune_batch_norm=false), and a small crop size and batch size (e.g., train_crop_size=513×513 and train_batch_size=1) were used. The models were trained by 50,000 steps, during which data augmentation was applied by randomly scaling and flipping (in horizontal direction) the input images. The dataset was randomly split into a training set, a validation set, and a test set according to the ratio of 7.0:1.5:1.5, resulting in 3515, 754, and 753 images in the respective subsets. The models were trained and calibrated on the training and validation sets, and finally evaluated on the test set.

4.1. Orthogonal experimental design

Multiple factors (i.e., hyperparameters) and several levels in each factor, as listed in Section 3.3.1, had to be calibrated. It would be onerous to exhaust all possible hyperparameter combinations (2×2×2×2×3=48). Orthogonal experimental design is an effective way to achieve balance between representativeness and number of experiments. The orthogonal array for four two-level factors and one three-level factor (SAS Institute Inc., 2020) was used to design our hyperparameter calibration experiments, as listed in Table 2. With the design, the required number of experiments was reduced significantly to 12.

Table 2. Orthogonal experimental design for model hyperparameter calibration.

Run	Template [a]	Experimental Factors (Hyperparameters)				
		Backbone	Pretrained on MINC [b]	MS and FL [c]	OS [d]	Resolution [e]
#1	00000	Xception	N	N	16	k=0.267
#2	00111	Xception	N	Y	8	k=0.521
#3	00112	Xception	N	Y	8	k=1
#4	01002	Xception	Y	N	16	k=1
#5	01010	Xception	Y	N	8	k=0.267
#6	01101	Xception	Y	Y	16	k=0.521
#7	10001	ResNet	N	N	16	k=0.521
#8	10012	ResNet	N	N	8	k=1
#9	10100	ResNet	N	Y	16	k=0.267
#10	11011	ResNet	Y	N	8	k=0.521
#11	11102	ResNet	Y	Y	16	k=1
#12	11110	ResNet	Y	Y	8	k=0.267

[a] Combination template of different factors, where the number at *i* digit signifies the corresponding level of the factor at *i* column.

[b] If the model has been pretrained on MINC dataset: Y stands for "Yes", and N stands for "No".

[c] If MS (multi-scale) and LR (left-right flip) is applied at the inference stage: Y stands for "Yes", and N stands for "No".

[d] Output stride (OS) and the corresponding APSS rates: when OS=16, the corresponding APSS rates are 6, 12, 18; when OS=8, the corresponding APSS rates are 12, 24, 36.

[e] Scaling value k to adjust the resolution: the corresponding resolutions for k=1, 0.521, and 0.267 are 1920*1080, 1001*563, and 513*289, respectively.

4.2. Influences of different hyperparameters on model performance

The results of the 12 orthogonal experiments in Table 2 revealed how the different factors (or hyperparameters) would influence the model performance. For example, if interested in the effects of different backbone networks on mIoU, one can compare the average mIoU of experiments that have used the Xception backbone (i.e., run #1, #2, #3, #4, #5, and #6 in Table 2) and the ResNet backbone (i.e., run #7, #8, #9, #10, #11, and #12 in Table 2). The same can be applied to other factors and evaluation metrics. Understanding of the influence of different factors is useful for determining optimal hyperparameter combinations. Table 3 summarizes the resulted performance metrics under different hyperparameter options. It was observed that the segmentation accuracy metrics (i.e., mIoU, mF1, mP, and mR) basically followed the same patterns over the change of hyperparameter options. The strong agreements indicate the true effects of the hyperparameters have been well depicted, ruling out the potential influence of other factors such as accuracy fluctuation. For the wide acceptance of mIoU in the research community, it will be used as our primary metric to measure segmentation accuracy thereafter.

Table 3. Summary of segmentation performance metrics under different hyperparameters.

Hyperparameters	Options	mIoU	mF1	mP	mR	Time (s)
Backbone	Xception	0.547	0.679	0.698	0.675	14.5
	ResNet	0.525	0.659	0.684	0.656	6.2
Pretrained	nMINC	0.530	0.662	0.688	0.659	13.6
	yMINC	0.542	0.676	0.695	0.671	7.1
InputProcess *	nMS	0.536	0.671	0.687	0.668	0.8
	yMS	0.536	0.667	0.695	0.663	22.7
Stride	OS16	0.528	0.660	0.686	0.658	2.5
	OS8	0.544	0.678	0.696	0.673	15.9
Resolution	k0.267	0.518	0.652	0.681	0.648	2.3
	k0.521	0.551	0.684	0.698	0.684	8.9
	k1	0.539	0.670	0.694	0.665	19.8

[*] Hyperparameter indicating whether input processing is applied.

Fig. 4 visualizes the effects of different hyperparameters on model performance. In the figure, each row focuses on the effect of an individual factor, and the three columns demonstrate the corresponding training curve, mIoU, and required inference time for each image. In the first column, each curve depicts the training process of an experiment in Table 2. Training curves of different levels for a factor (e.g., Xception and ResNet for the "backbone" factor) are marked with different essential colors for intuitive comparison. Note that subsampling and

exponential smoothing techniques were applied to make the general trend more perceivable. The original loss, smoothed loss values, and the corresponding errors have been attached in the Supplementary Data for readers' further reference. In the second column, the mIoU is evaluated on the validation set, and is calculated as the average of experiments that have applied the same level of interest. In the third column, the time performance is evaluated on the validation set, where the red dots indicate inference time for different experiments.

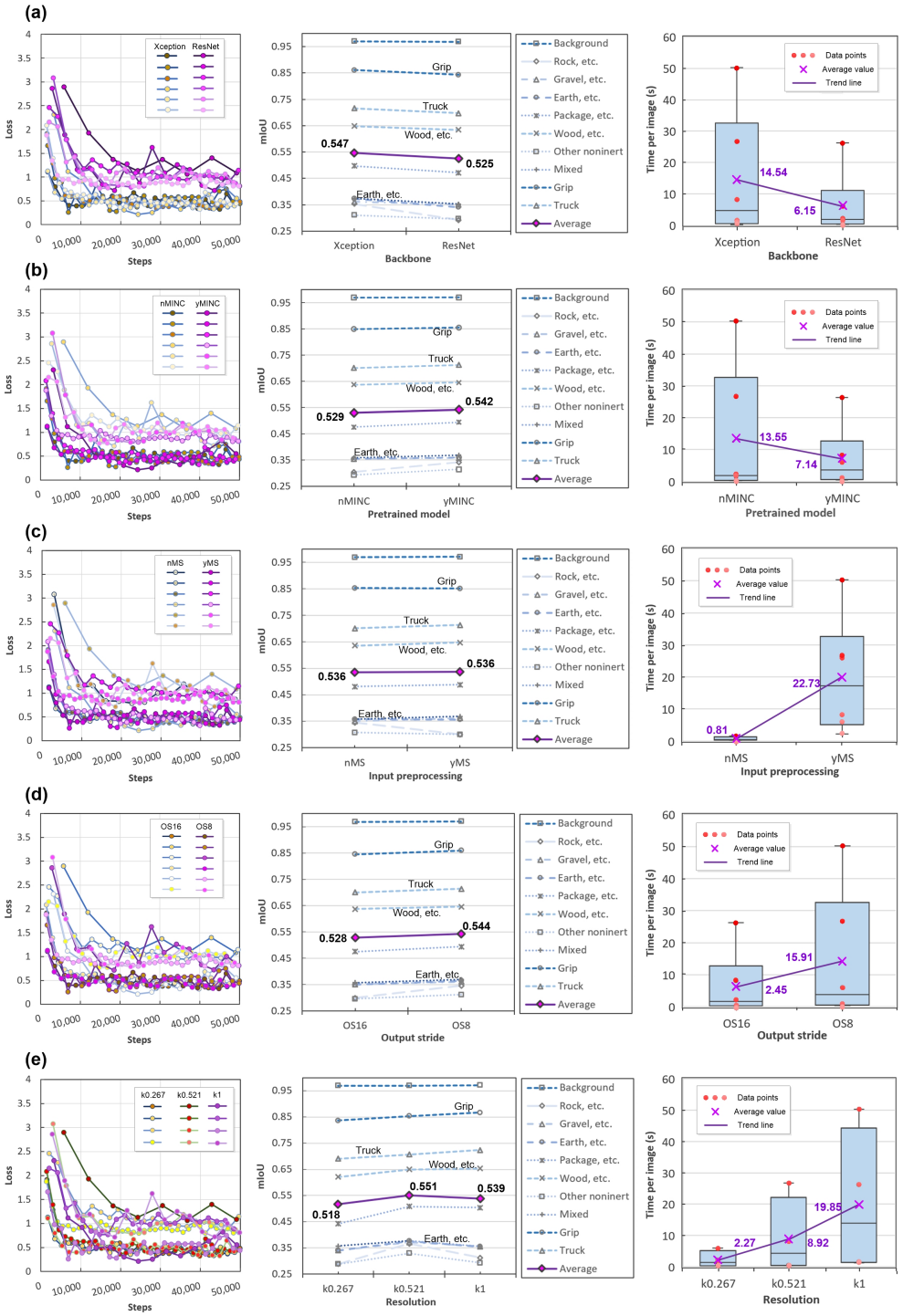

Fig. 4. Effects of (a) backbone networks, (b) pretrained dataset, (c) input preprocessing, (d) output stride, and (e) resolution on performance from three aspects, i.e., training process (first column), mIoU (second column), and time performance (third column).

4.2.1. Effects of backbone network

The two backbone networks in the tests were Xception65 (Zhu et al., 2020b) and ResNet-101 (Zhu et al., 2020a). Quite significant differences in model performance were observed from Fig. 4(a) when different backbones were used. As shown by the training curves in the first column, the training loss of experiments with an Xception backbone is significantly lower than those with ResNet. The mIoU provided by models with Xception backbone is higher for almost all nine types of material/object than by their counterparts. The overall mIoU for Xception is 0.547; 0.022 higher than ResNet. However, the Xception backbone tends to be more time-consuming, as indicated by the graph in the third column.

4.2.2. Effects of pretrained dataset

Fig. 4(b) investigates whether pretraining models on the MINC dataset can improve performance. Regarding the training process, although the difference does not seem as significant as that observed in Fig. 4(a), models pretrained on MINC tend to have lower losses in a general sense. The mIoU in the second column reaffirms the observation, with the overall mIoU of models pretrained on MINC being 0.013 higher that of their counterparts. MINC is a dataset oriented to materials recognition in context, and includes 23 categories of material such as wood, glass, brick and tile. Thus, pretraining the dataset is beneficial for the model in recognizing similar materials in CW, leading to higher segmentation precision. As indicated by the graph in the third column, pretraining on MINC also seems to reduce inference time. The model checkpoints pretrained on MINC and the MINC dataset in PASCAL VOC format have been publicized on GitHub (civilServant-666, 2021). Based on the archive, readers can train their own material recognition models.

4.2.3. Effects of input preprocessing at inference stage

Fig. 4(c) aims to understand whether preprocessing inputs with MS and LF (Li, 2020) at the inference stage can improve segmentation performance. The training curves with MS applied or not are mixed together without observable differentiation. This is because the MS operations are only applied at the inference stage and thus have no influence on the training process. However, unlike (Chen et al., 2017b), meaningful improvement on mIoU was not observed after applying MS and LF, as indicated by the graph in the second column. This could be related to a compromise on number of scales applied in some experiments due to limited GPU resources. Thus, the full potential of MS and LF operation in terms of meaningful improvement may not have been reached. The application of MS did drastically increase the required inference time to 22.73 s to process an image, 28 times that of not applying MS. This is consistent with (Chen et al., 2017b).

4.2.4. Effects of output stride

Fig. 4(d) investigates the influence of OS. Presumably, a lower OS can result in larger feature maps, which then can preserve more details of the objects' boundary and edge information, leading to higher segmentation accuracy. The assumption is validated by our results in the first and second columns of Fig. 4 (d), where training loss of models with OS8 distributes lower, in a general sense, than their counterparts, and the mIoU is 0.16 higher when an output stride of 8 is used. However, larger feature maps require more time to process. This can lead to longer inference time for per image, as indicated by the graph in the third column.

4.2.5. Effects of image resolution

The influence of image resolution is analyzed in Fig. 4(e). It is observed that segmentation performance would be impaired if the images were scaled by a factor k of 0.267 (i.e., into a resolution of 513*289), leading to higher training loss (the first column) and lower average mIoU (the second column). However, scaling the images by a k of 0.521 does not seem to cause a drop in performance, as the training curves of $k = 0.521$ and $k = 1$ are hardly distinguishable, and the average mIoU of $k = 0.521$ is even higher than that of $k = 1$. Regarding time performance, it is no surprise to see the inference time increases with the growth of k, as larger photos require more time to process.

4.3. Performance evaluation of the optimal model

The orthogonal experimental analysis suggests that an Xception backbone pretrained on MINC with an output stride of 8 and input images scaled by 0.521 tends to yield higher segmentation accuracy. Therefore, the hyperparameter combination "backbone = Xception, pretrained_on_MINC = True, OS = 8, and $k = 0.521$" was used to train a presumably optimal model. As the MS and LR operation does not yield observable accuracy improvement but can significantly slow down processing, it was not applied at the inference stage.

Fig. 5 shows the mIoU and confusion matrix of the optimal model on the validation and test sets. The overall mIoUs on validation set and test set are close at 0.562 and 0.555, respectively, while the respective mF1 values are 0.696 and 0.688. The results indicate the model performs well in generalizing to new samples. Regarding mIoU of specific materials/objects, consistent with the orthogonal experiments, the model performs best in segmenting objects such as grip and truck, followed by non-inert materials such as wood, cardboard and packaging, and finally inert materials such as rock, gravel and earth. Fig. 6 shows some segmentation results from the test set. The second and the third columns are the ground-truth and predictive segmentations, respectively, for the raw images in the first column. In the fourth column, the predictive segmentations are overlaid onto the raw images for results visualization.

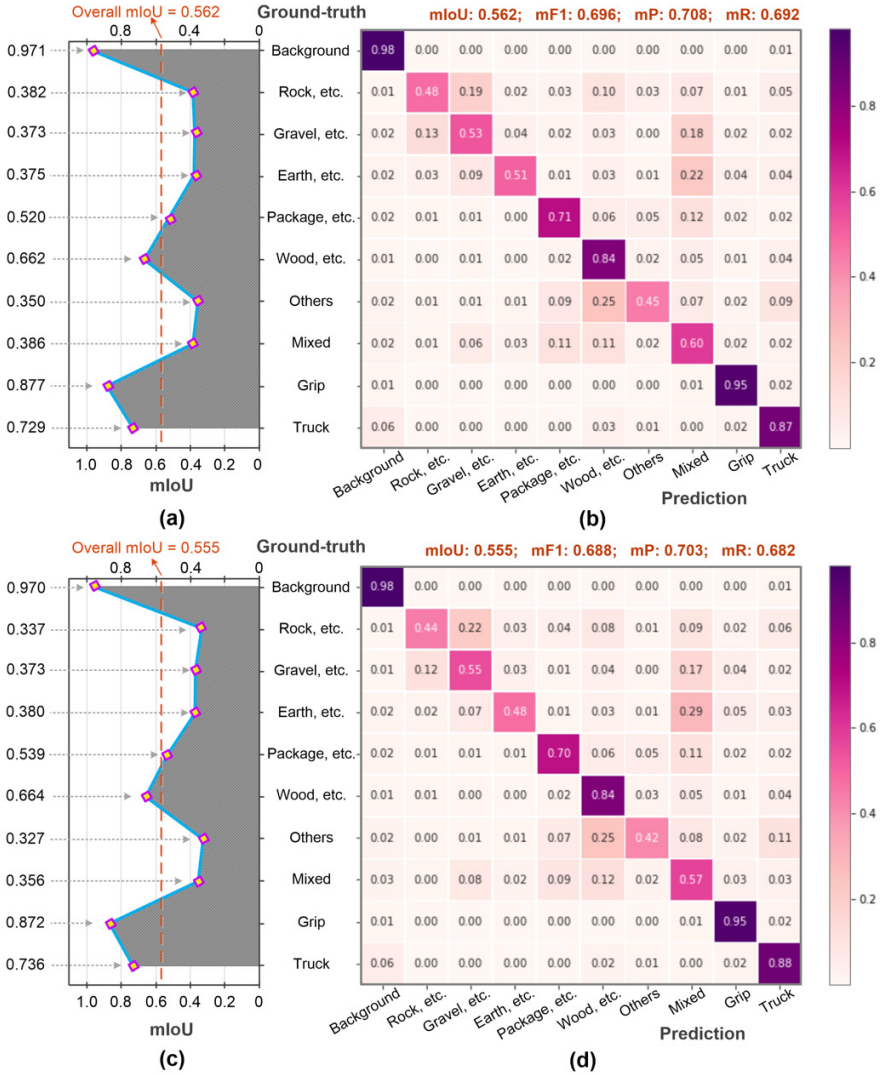

Fig. 5. Segmentation performance of the model, which includes (a) mIoU and (b) confusion matrix on the validation set; and (c) mIoU and (d) confusion matrix on the test set. Note that the values in confusion matrices have been rounded up by two decimals.

Although boundaries of the segmentation could be fine-tuned further, the model successfully identified most of the materials appearing in the dump buckets, even recognizing waste materials at a finer level of detail than the ground-truth annotations. For example, in #2 of Fig. 6, the model correctly recognized the "Earth/Slurry/Mud" (dark yellow) in the middle and rear of the dump bucket, while the "ground-truth" annotation only assigned a rough class label "Mixed" (gray) to the area. In #8, there were actually some relatively large-sized "Rock/Stone/Rubble/Debris" in the area annotated as "Gravel/Concrete/Bricks" (green), which were successfully identified by our model (see the dark red area in the predictive segmentation). Regarding time performance, the model averaged 0.51 s to process an image.

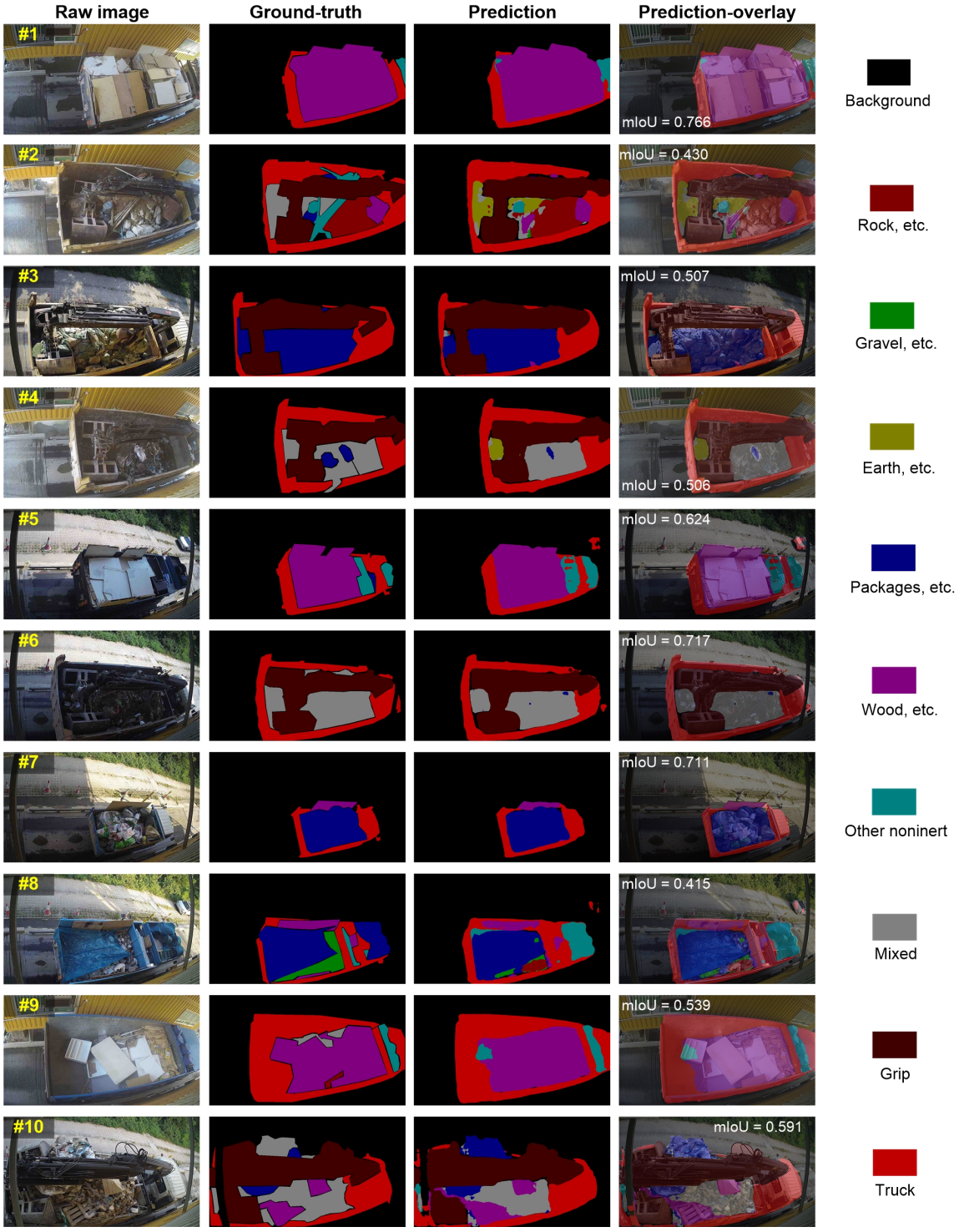

Fig. 6. Segmentation results of examples from the test set.

5. Discussion

5.1. Significance of the research findings

The experimental studies demonstrate an overall mIoU of 0.56 in segmenting CW. When evaluating segmentation performance, the complexity of a specific dataset should be taken into account as some datasets might be more challenging than others. For example, while the

DeepLabv3+ achieved a mIoU of 0.878 on Pascal VOC 2012, its performance on ADE20K is 0.457 (Zhu et al., 2020c). On Pascal Context dataset, the state of the art is 0.562 (Sanjaya et al., 2020). Considering variant scale and clutter nature of the waste materials in our dataset, the task is quite challenging (some materials are even difficult for humans to distinguish), and thus an mIoU of 0.56 is satisfactory. To validate the robustness of the model to complexity and variation of external environments, Figs. 7(a) and (b) analyze the mIoU distribution of all the 1,507 samples from both the validation and test sets over different levels of brightness and types of vehicles. Despite variation of brightness (Fig. 7(a)), the mIoUs of the samples are basically around the same level of 0.5. In the most extreme cases where the illumination is low, the average mIoU still reaches 0.477. With change of vehicle tonnage, the mIoUs slightly fluctuate around 0.5 (Fig. 7(b)). However, the lowest mIoU (0.476) observed at the group (24, 32] is still at the same level as those observed from other groups. This analysis demonstrates the efficacy of the proposed model in recognizing CW composition in its cluttered state and external variation.

Our proposed approach makes significant theoretical contributions to the problem of material composition recognition, particularly for materials with a mixture and cluttered nature in complex environments. Specifically, this paper proposes a three-step methodology leveraging deep learning that involves preparation of the material image dataset, development of the construction waste segmentation model based on DeepLabv3+ structure, and training and calibration of the segmentation model. It has been experimentally demonstrated the effectiveness of deep learning techniques such as DeepLab in segmenting mixtures of bulky CW with a desired accuracy (mIoU = 0.56) and time performance (0.51 s per image). Additionally, the study shows how material recognition performance can be improved by calibrating the semantic segmentation model via orthogonal experimental design. In the case of DeepLabv3+, it is observed that a model pretrained on other material datasets (e.g., MINC) with an Xception backbone, an OS of 8, and high-resolution input images tends to yield higher segmentation accuracy.

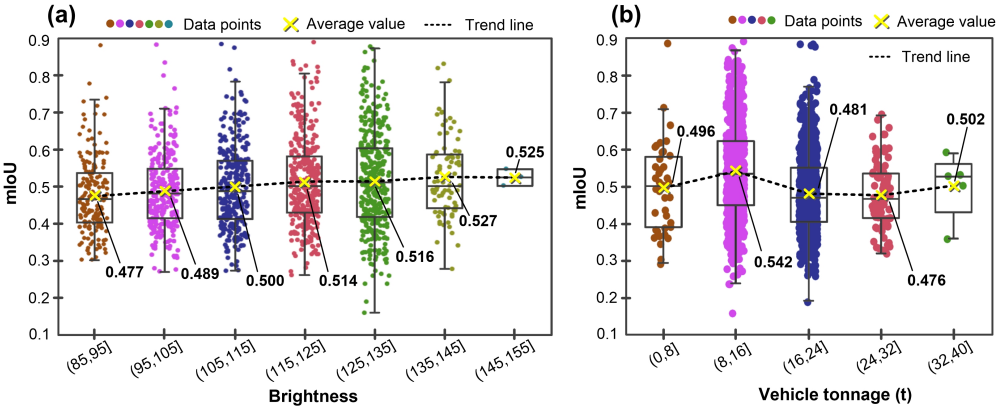

Fig. 7. Distribution of mIoU over different (a) levels of brightness and (b) vehicle tonnages.

This study paves the way for better CW management. On the one hand, it provides a powerful technical tool in the use of CV to gauge material composition in waste dumps, a yardstick metric for CW management in many places. With automatic recognition of dump materials from the surface, composition of the entire dump can be more reliably estimated. On the other hand, the study provides information concerning the categories, position and geometry of the materials in a mixture of CW. This information can be used to drive robots so that segregation in complex environments can be automated. Replacement of human sorting workers with machines will result in fewer occupational accidents, higher efficiency, and better segregation quality.

5.2. Limitations and future work

There is room for further improvement in segmentation performance. First, the complexity of CW composition makes it extremely challenging to resolve ambiguities in certain material types. For example, the "Mixed" type, according to its definition, refers to a mixture of inert and non-inert waste that is difficult to classify. However, a large-scale "mixed" area can comprise many identifiable specific types of materials if observed from a smaller scale. This was the case in #2 of Fig. 6, where the model detected "Earth/Slurry/Mud" material in a region labeled "Mixed" by the annotation workers. Another example is the classes "Rock" and "Gravel", which can be too indistinguishable to be universally recognized. For instance, in #8 of Fig. 6, while the annotator labeled the inert materials in the dump bucket "Gravel", the model recognized some larger materials as "Rock".

Second, the model's ability to depict the materials' edges and boundaries should be improved. Fig. 8 highlights the ground-truth boundaries of different materials/objects, revealing the deviation and inconsistency of predicted boundaries. Future research should consider improving the mIoU by incorporating low-level information on the details of edges and boundaries of the materials/objects. One way to do this might be the integration of some preprocessing techniques such as superpixel segmentation, as it can automatically extract segmentation edges in the images by analyzing the similarity of color, texture, and other properties of image pixels. Another way for potential improvement is the use of emerging boundary refinement schemes such as SegFix (Yuan et al., 2020), which are becoming prevalent to finetune the results of semantic segmentation.

Third, the segmentation precision distributes unevenly across different categories, as shown by Fig. 5. This phenomenon could, on one the hand, be related to the fact that objects, such as grip and truck, and non-inert materials, such as wood and packaging, present more regular shapes and salient features that can be easily identified by the model, leading to higher segmentation accuracy. On the other hand, it might be due to the imbalanced distribution of the dataset over different materials (Fig. 2(a)). Although different class weights have been assigned to counteract the downside brought by the imbalanced dataset, it might also be

interesting for future research to examine if the issue can be further mitigated by IoU-based loss functions such as dice coefficient (Milletari et al., 2016), which has demonstrated its strength in handling imbalanced datasets (Li et al., 2019).

Fourth, compared with other state-of-the-art segmentation techniques (Chen et al., 2019) that can process around 30 image frames per second, the efficiency of the construction waste segmentation model is relatively low. Although the lag-behind time performance can be partially attributed to the differences of the used hardware and input resolution, future research is suggested to refine the model structure for potential efficiency improvement.

Fig. 8. Examples that show inconsistency between the predicted segmentation boundaries and the ground-truth.

6. Conclusions

Information on construction waste composition is of paramount importance for better construction waste management. The potential of computer vision for recognizing construction waste composition has long been acknowledged. However, existing studies have mainly focused on classification or detection of municipal solid waste in simplified environments. This paper presents an approach to tackling the problem of recognizing construction waste composition with high-level granularity in complex environments by leveraging semantic segmentation techniques. Firstly, a big dataset of 5,366 construction waste images capturing a wide range of materials (e.g., rock, gravel, wood and packaging) in varied, real-life outdoor environments was collected, preprocessed and annotated meticulously. Based on the state-of-the-art DeepLabv3+ structure, a construction waste segmentation model was then developed. Finally, the model was trained and calibrated on the construction waste image dataset via orthogonal experimental design.

It was found that smaller output stride, larger image resolution, and pretraining on a material dataset such as MINC can potentially lead to higher segmentation accuracy. Compared with ResNet, an Xception backbone is preferable. Our optimal model achieved an overall mIoU of 0.56 in segmenting nine types of materials/objects, which demonstrates the efficacy of the semantic segmentation approach to recognizing mixtures of bulky construction waste materials in complex environments. The findings pave the way for automated construction

waste segregation by deploying robots in the future. The trained models have been made publicly available in the hope that they can help researchers to train their own material recognition models.

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.

Acknowledgement

This research is jointly supported by the Strategic Public Policy Research (SPPR) Funding Scheme (Project No.: S2018.A8.010.18S) and the Environment and Conservation Fund (ECF) (Project No.: ECF 2019-111) of the Government of the Hong Kong Special Administrative Region.

References

Aquariusjay, Zhu, Y., 2019. Running Deeplab on Pascal Voc 2012 Semantic Segmentation Dataset. https://github.com/tensorflow/models/blob/master/research/deeplab/g3doc/pascal.md (Accessed April 7 2021).

Awe, O., Mengistu, R., Sreedhar, V., 2017. Smart Trash Net: Waste Localization and Classification, arXiv preprint.

Bell, S., Upchurch, P., Snavely, N., Bala, K., 2015. Material Recognition in the Wild with the Materials in Context Database, Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3479-3487.

Capellier, E., Davoine, F., Fremont, V., Ibanez-Guzman, J., Li, Y., 2018. Evidential Grid Mapping, from Asynchronous Lidar Scans and Rgb Images, for Autonomous Driving, 2018 21st International Conference on Intelligent Transportation Systems (ITSC), pp. 2595-2602.

Chen, J., Lu, W., Xue, F., 2021. "Looking beneath the Surface": A Visual-Physical Feature Hybrid Approach for Unattended Gauging of Construction Waste Composition. Journal of Environmental Management 286, 112233. https://doi.org/10.1016/j.jenvman.2021.112233.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2014. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected Crfs. arXiv preprint arXiv:1412.7062.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2017a. Deeplab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected Crfs. IEEE transactions on pattern analysis and machine intelligence 40, 834-848.

Chen, L.-C., Papandreou, G., Schroff, F., Adam, H., 2017b. Rethinking Atrous Convolution for Semantic Image Segmentation. arXiv preprint arXiv:1706.05587.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. Springer International Publishing, Cham, pp. 833-851.

Chen, X., Zhou, Z., Ying, Y., Qi, D., 2019. Real-Time Human Segmentation Using Pose Skeleton Map, 2019 Chinese Control Conference (CCC), pp. 8472-8477.

Chollet, F., 2017. Xception: Deep Learning with Depthwise Separable Convolutions,

Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1251-1258.
civilServant-666, 2021. Deeplab-V3-Models-for-Waste-Material-Recognition. https://github.com/civilServant-666/DeepLab-v3-models-for-waste-material-recognition (Accessed April 25 2021).
Defra, 2020. Uk Statistics on Waste. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/918270/UK_Statistics_on_Waste_statistical_notice_March_2020_accessible_FINAL_updated_size_12.pdf (Accessed April 26 2021).
Faibish, S., Bacakoglu, H., Goldenberg, A.A., 1997. An Eye-Hand System for Automated Paper Recycling, Proceedings of International Conference on Robotics and Automation, pp. 9-14 vol.11.
Gálvez-Martos, J.-L., Styles, D., Schoenberger, H., Zeschmar-Lahl, B., 2018. Construction and Demolition Waste Best Management Practice in Europe. Resources, Conservation and Recycling 136, 166-178. https://doi.org/10.1016/j.resconrec.2018.04.016.
Gundupalli, S.P., Hait, S., Thakur, A., 2017a. Multi-Material Classification of Dry Recyclables from Municipal Solid Waste Based on Thermal Imaging, Waste Management, pp. 13-21.
Gundupalli, S.P., Hait, S., Thakur, A., 2017b. A Review on Automated Sorting of Source-Separated Municipal Solid Waste for Recycling. Waste Management 60, 56-74. https://doi.org/10.1016/j.wasman.2016.09.015.
Gundupalli, S.P., Hait, S., Thakur, A., 2018. Classification of Metallic and Non-Metallic Fractions of E-Waste Using Thermal Imaging-Based Technique. Process Safety and Environmental Protection 118, 32-39. https://doi.org/10.1016/j.psep.2018.06.022.
He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask R-Cnn, Proceedings of the IEEE international conference on computer vision, pp. 2961-2969.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep Residual Learning for Image Recognition, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778.
HKEPD, 2011. Construction Waste Statistics. https://www.epd.gov.hk/epd/misc/cdm/trip.htm (Accessed April, 7 2021).
HKEPD, 2019. Management of Abandoned Construction and Demolition Materials. https://www.aud.gov.hk/pdf_e/e67ch04sum.pdf (Accessed 20 November 2020).
HKEPD, 2020. Hong Kong Waste Treatment and Disposal Statistics. https://www.epd.gov.hk/epd/english/environmentinhk/waste/data/stat_treat.html (Accessed April 7 2021).
Huang, G.-L., He, J., Xu, Z., Huang, G., 2020. A Combination Model Based on Transfer Learning for Waste Classification. Concurrency and Computation: Practice and Experience 32, e5751. https://doi.org/10.1002/cpe.5751.
Jégou, S., Drozdzal, M., Vazquez, D., Romero, A., Bengio, Y., 2017. The One Hundred Layers Tiramisu: Fully Convolutional Densenets for Semantic Segmentation, Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 11-19.
Koyanaka, S., Kobayashi, K., 2011. Incorporation of Neural Network Analysis into a Technique for Automatically Sorting Lightweight Metal Scrap Generated by Elv Shredder Facilities. Resources, Conservation and Recycling 55, 515-523. https://doi.org/10.1016/j.resconrec.2011.01.001.
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet Classification with Deep Convolutional Neural Networks, ADV NEUR IN, pp. 1097-1105.
Ku, Y., Yang, J., Fang, H., Xiao, W., Zhuang, J., 2020. Deep Learning of Grasping Detection

for a Robot Used in Sorting Construction and Demolition Waste. Journal of Material Cycles and Waste Management. 10.1007/s10163-020-01098-z.
Kujala, J.V., Lukka, T.J., Holopainen, H., 2015. Picking a Conveyor Clean by an Autonomously Learning Robot, arXiv preprint arXiv:1511.07608.
Lau Hiu Hoong, J.D., Lux, J., Mahieux, P.-Y., Turcry, P., Aït-Mokhtar, A., 2020. Determination of the Composition of Recycled Aggregates Using a Deep Learning-Based Image Analysis. AUTOMAT CONSTR 116, 103204. https://doi.org/10.1016/j.autcon.2020.103204.
Li, E.Y., 2020. Witnessing the Progression in Semantic Segmentation: Deeplab Series from V1 to V3+. https://towardsdatascience.com/witnessing-the-progression-in-semantic-segmentation-deeplab-series-from-v1-to-v3-4f1dd0899e6e (Accessed April, 7 2021).
Li, X., Sun, X., Meng, Y., Liang, J., Wu, F., Li, J., 2019. Dice Loss for Data-Imbalanced Nlp Tasks. arXiv preprint arXiv:1911.02855.
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.
Long, J., Shelhamer, E., Darrell, T., 2015. Fully Convolutional Networks for Semantic Segmentation, Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3431-3440.
Lu, W., Yuan, H., 2012. Off-Site Sorting of Construction Waste: What Can We Learn from Hong Kong? Resources, Conservation and Recycling 69, 100-108. https://doi.org/10.1016/j.resconrec.2012.09.007.
Lu, W., Yuan, L., 2021. Investigating the Bulk Density of Construction Waste: A Big Data-Driven Approach. Resources, Conservation & Recycling 169, 105480. https://doi.org/10.1016/j.resconrec.2021.105480.
Lukka, T.J., Tossavainen, T., Kujala, J.V., Raiko, T., 2014. Zenrobotics Recycler–Robotic Sorting Using Machine Learning, Proceedings of the International Conference on Sensor-Based Sorting (SBS), pp. 1-8.
Mansouri, I., 2019. Computer Vision Part 6: Semantic Segmentation, Classification on the Pixel Level. https://medium.com/analytics-vidhya/computer-vision-part-6-semantic-segmentation-classification-on-the-pixel-level-ee9f5d59c1c8 (Accessed April, 7 2021).
Mao, W., Chen, W., Wang, C., Lin, Y., 2021. Recycling Waste Classification Using Optimized Convolutional Neural Network. Resources, Conservation and Recycling 164, 105132. https://doi.org/10.1016/j.resconrec.2020.105132.
Mattone, R., Campagiorni, G., Galati, F., 2000. Sorting of Items on a Moving Conveyor Belt. Part 1: A Technique for Detecting and Classifying Objects, Robotics and Computer-Integrated Manufacturing, pp. 73-80.
Meng, S., Chu, W., 2020. A Study of Garbage Classification with Convolutional Neural Networks, 2020 Indo – Taiwan 2nd International Conference on Computing, Analytics and Networks (Indo-Taiwan ICAN), pp. 152-157.
Milletari, F., Navab, N., Ahmadi, S.-A., 2016. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation, 2016 fourth international conference on 3D vision (3DV). IEEE, pp. 565-571.
Mittal, G., Yagnik, K.B., Garg, M., Krishnan, N.C., 2016. Spotgarbage: Smartphone App to Detect Garbage Using Deep Learning, Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 940-945.
Nowakowski, P., Pamuła, T., 2020. Application of Deep Learning Object Classifier to Improve E-Waste Collection Planning. Waste Management 109, 1-9. https://doi.org/10.1016/j.wasman.2020.04.041.
NSWEPA, 2020. Protection of the Environment Operations (Waste) Regulation 2014.

https://legislation.nsw.gov.au/view/html/inforce/current/sl-2014-0666#statusinformation (Accessed August 19 2021).
Paulraj, S.G., Hait, S., Thakur, A., Asme, 2016. Automated Municipal Solid Waste Sorting for Recycling Using a Mobile Manipulator, Proceedings of the Asme International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, 2016.
Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation, International Conference on Medical image computing and computer-assisted intervention. Springer, pp. 234-241.
Sanjaya, Y.C., Gunawan, A.A.S., Irwansyah, E., 2020. Semantic Segmentation for Aerial Images: A Literature Review. Engineering, MAthematics and Computer Science (EMACS) Journal 2, 133-139.
SAS Institute Inc., 2020. Frequently Used Orthogonal Arrays. http://support.sas.com/techsup/technote/ts723_Designs.txt (Accessed April 7 2021).
Sun, L., Zhao, C., Yan, Z., Liu, P., Duckett, T., Stolkin, R., 2019. A Novel Weakly-Supervised Approach for Rgb-D-Based Nuclear Waste Object Detection, Ieee Sensors Journal, pp. 3487-3500.
Thung, G., Yang, M., 2019. Trashnet Dataset. https://github.com/garythung/trashnet (Accessed 9 February 2021).
Toto, D., 2019. Machinex Sells Nine Samurai Units. https://www.wastetodaymagazine.com/article/machinex-samurai-installations/ (Accessed Oct. 2 2021).
USEPA, 2018. Advancing Sustainable Materials Management: 2015 Fact Sheet. https://bit.ly/2yRMTvN (Accessed April 25 2021).
Vrancken, C., Longhurst, P., Wagland, S., 2019. Deep Learning in Material Recovery: Development of Method to Create Training Database, Expert Systems with Applications, pp. 268-280.
Wada, K., 2019. Labelme: Image Polygonal Annotation with Python. https://github.com/zhong110020/labelme (Accessed April, 7 2021).
Wang, Z., Li, H., Yang, X., 2020. Vision-Based Robotic System for on-Site Construction and Demolition Waste Sorting and Recycling. Journal of Building Engineering 32, 101769. https://doi.org/10.1016/j.jobe.2020.101769.
Wang, Z., Li, H., Zhang, X., 2019a. Construction Waste Recycling Robot for Nails and Screws: Computer Vision Technology and Neural Network Approach. AUTOMAT CONSTR 97, 220-228. https://doi.org/10.1016/j.autcon.2018.11.009.
Wang, Z., Peng, B., Huang, Y., Sun, G., 2019b. Classification for Plastic Bottles Recycling Based on Image Recognition, Waste Management, pp. 170-181.
Wu, H., Yao, L., Xu, Z., Li, Y., Ao, X., Chen, Q., Li, Z., Meng, B., 2019. Road Pothole Extraction and Safety Evaluation by Integration of Point Cloud and Images Derived from Mobile Mapping Sensors. ADV ENG INFORM 42, 100936. https://doi.org/10.1016/j.aei.2019.100936.
Xiao, W., Chang, L., Liu, W., 2018. Semantic Segmentation of Colorectal Polyps with Deeplab and Lstm Networks, 2018 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), pp. 1-2.
Xiao, W., Yang, J., Fang, H., Zhuang, J., Ku, Y., 2020. Classifying Construction and Demolition Waste by Combining Spatial and Spectral Features, Proceedings of the Institution of Civil Engineers - Waste and Resource Management. ICE Publishing, pp. 79-90.
Xu, J., Gui, C., Han, Q., 2020. Recognition of Rust Grade and Rust Ratio of Steel Structures Based on Ensembled Convolutional Neural Network. Computer-Aided Civil and

Infrastructure Engineering 35, 1160-1174. https://doi.org/10.1111/mice.12563.
Yang, M., Thung, G., 2016. Classification of Trash for Recyclability Status, CS229 Project Report.
Yuan, Y., Xie, J., Chen, X., Wang, J., 2020. Segfix: Model-Agnostic Boundary Refinement for Segmentation, European Conference on Computer Vision. Springer, pp. 489-506.
Zhang, Q., Zhang, X., Mu, X., Wang, Z., Tian, R., Wang, X., Liu, X., 2021. Recyclable Waste Image Recognition Based on Deep Learning. Resources, Conservation and Recycling 171, 105636. https://doi.org/10.1016/j.resconrec.2021.105636.
Zhu, W., Chen, L., Wang, B., Wang, Z., 2018. Online Detection in the Separation Process of Tobacco Leaf Stems as Biomass Byproducts Based on Low Energy X-Ray Imaging, Waste and Biomass Valorization, pp. 1451-1458.
Zhu, Y., huihui-personal, aquariusjay, 2020a. Pretrained Model: Resnet_V1_101_Beta_Imagenet. http://download.tensorflow.org/models/resnet_v1_101_2018_05_04.tar.gz (Accessed April 7 2021).
Zhu, Y., huihui-personal, aquariusjay, 2020b. Pretrained Model: Xception65_Coco_Voc_Trainaug. http://download.tensorflow.org/models/deeplabv3_pascal_train_aug_2018_01_04.tar.gz (Accessed April 7 2021).
Zhu, Y., huihui-personal, aquariusjay, 2020c. Tensorflow Deeplab Model Zoo. https://github.com/tensorflow/models/blob/master/research/deeplab/g3doc/model_zoo.md (Accessed April 7 2021).