"Looking beneath the surface": A visual-physical feature hybrid approach for unattended gauging of construction waste composition

Junjie Chen, Weisheng Lu, and Fan Xue

Department of Real Estate and Construction, The University of Hong Kong, Pokfulam, Hong Kong SAR, China

Emails: { chenjj10; wilsonlu; xuef } @hku.hk

Abstract

There are various scenarios challenging human experts to judge the interior of something based on limited surface information. Likewise, at waste disposal facilities around the world, human inspectors are often challenged to gauge the composition of waste bulks to determine admissibility and chargeable levy. Manual approaches are laborious, hazardous, and prone to carelessness and fatigue, making unattended gauging of construction waste composition using simple surface information highly desired. This research attempts to contribute to automated waste composition gauging by harnessing a valuable dataset from Hong Kong. Firstly, visual features, called visual inert probability (*VIP*), characterizing inert and non-inert materials are extracted from 1,127 photos of waste bulks using a fine-tuned convolutional neural network (CNN). Then, these visual features together with easy-to-obtain physical features (e.g., weight and depth) are fed to a tailor-made support vector machine (SVM) model to determine waste composition as measured by the proportions of inert and non-inert materials. The visual-physical feature hybrid model achieved a waste composition gauging accuracy of 94% in the experiments. This high performance implies that the model, with proper adaption and integration, could replace human inspectors to smooth the operation of the waste disposal facilities.

Keywords: Construction and demolition waste; Construction waste management; Waste composition; Computer vision; Deep convolutional neural network; Support Vector machine.

1. Introduction

In industrial activities or daily life, human experts often need to "look beneath the surface" of a subject matter using limited information from exterior visual and physical features. An Internet-era example is tele-diagnosis, use of which has increased amid COVID-19 pandemic social distancing (Larner, 2020) and which requires doctors to consult patients and make

diagnoses remotely by video call and with very limited information of medical history of the patients. In infrastructure inspection, engineers need to evaluate structural condition using external appearances (e.g., cracks or other damages) and limited information collected by sensors (Koch et al., 2015; Li et al., 2012). Energy analysts in the petroleum industry use satellite imagery of oil storage facilities to better understand change in reserve volumes (Watts, 2019). Remote sensing has also been used in minefield detection to infer areas of buried unexploded ordnance (Bennett, 1999; Maathuis, 2003).

Similarly in solid waste management, people need to judge the internal composition of waste materials (Sauve and Van Acker, 2020) for subsequent processing (e.g., recycling or landfilling) based on a limited set of features, such as visual appearance, moisture, weight, and volume. With increasing environmental awareness in Mainland China, the authority has begun to advocate municipal solid waste (MSW) segregation at source. In Tier-1 cities such as Shanghai and Beijing, this has been made compulsory (Arantes et al., 2020; Zuo and Yan, 2019). Household MSW generally comprises organic waste, non-recyclable inorganic waste and recyclable waste (Zhang et al., 2010). Within these broad categories are hundreds of different waste materials and it is difficult for both citizens and enforcers, who have little knowledge and experience in garbage classification, to distinguish among them based on appearance. It is no surprise that residents "were almost driven crazy by garbage classification" (Yu, 2019).

Even more challenging is gauging the composition of a mixture of bulky construction waste (CW), also referred to as construction and demolition (C&D) waste. This is exactly the dilemma now confronting many countries and regions, including Hong Kong. In 2006, the Hong Kong Environmental Protection Department (EPD) launched the Construction Waste Disposal Charging Scheme (CWDCS), imposing scaled waste disposal fees on contractors or waste haulers according to the proportion of inert content in a waste dump (e.g., HK$200 per ton for waste dumps with less than 50% inert content, and HK$71 per ton for entirely inert waste content). The CWDCS relies on the reliable and efficient gauging of inert waste proportions. In current practice, human inspectors have to judge whether the composition of an incoming waste dump meets the criteria based on limited data, including weight and depth of the waste and overhead photos. In case of dispute, entire truckloads of waste have to be manually separated and examined. This human-reliant practice is laborious, inefficient, and can be affected by fatigue, sloppiness, or even bribery.

Recent technological advancements offer means to overcome the limitations of human-reliant decision-making. With explosive growth in available data and computing power, knowledge to assist human decision-makers can be extracted using computer vision (CV), data mining, machine learning (ML), and other techniques. Such domain knowledge can be used to train intelligent machines to support or even replace human experts for biomedical image analysis (Tschandl et al., 2020), ambient intelligence-based healthcare (Haque et al., 2020),

infrastructure condition evaluation (Bhola et al., 2018; Chen and Liu, 2021; Wu et al., 2019) and oil storage estimation from satellite imagery (Mubasir, 2020). Emerging technologies can be applied to automate the gauging of CW composition as well. With a big enough dataset, it is viable to find the hidden correlation between surface features of waste dumps and human judgements on their composition. If a machine can learn the correlation, it can then be used to automatically gauge CW composition, not only outperforming human inspectors in terms of efficiency, but also freeing us from such issues as fatigue, sloppiness, and corruption.

Since AlexNet won the ImageNet competition in 2012 (Krizhevsky et al., 2012), convolutional neural network (CNN) has become a state-of-the-art approach to CV-related tasks. Over the course of nearly a decade, it evolved into numerical variants with different architectures, such as DenseNet for image classification, Faster R-CNN for object detection, and DeepLab for semantic segmentation. CNNs and other CV techniques are increasingly being used in waste management for better efficiency and productivity. Nowakowski and Pamuła (2020) applied Faster R-CNN to process smartphone-captured photos to categorize and detect e-wastes, which is beneficial to waste collection planning. Arebey et al. (2012) used image features for automatic waste bin level detection. Manufacturers (BHS, 2017) are incorporating deep CNNs into the sorting lines in waste recovery facilities to replace humans for quality control. However, existing studies mainly focus on recognizing or detecting individual waste items in a structured or semi-structured environment. Few of them, if any, has used CV to understand the internal composition of a dump of mixed waste. In addition, stand-alone use of CV might be biased as it only reflects part of the characteristics of an object via visual appearance. To better estimate the interior of an object, more features should be incorporated. There exist studies (Chu et al., 2018; Koyanaka and Kobayashi, 2011) in the waste sorting sector that integrated visual features with physical properties such as conductivity to detect materials on conveyor belts. However, for waste dumps, it remains unclear what types of physical features should be selected, and how they should be integrated with visual recognition to gauge waste composition.

The aim of this study is to provide a machine learning-based approach to automatically gauge the interior composition of construction waste through hybrid use of its visual appearance and physical features. The contribution is twofold. From a theoretical perspective, this study innovatively integrates visual features extracted by a CNN and physical properties of a waste dump to estimate its interior composition. In our experiments, the accuracy of a visual-physical integrated model significantly outperforms that of a model based solely on visual appearance, thus demonstrating the superiority of the hybrid approach. From a practical point of view, the approach is expected to improve the efficiency and reliability of waste composition gauging and can be deployed to enable a shift towards unattended operations in waste disposal facilities, similarly to recently unmanned kitchens, groceries, and warehouses. The automation frees humans from repetitive and hazardous tasks, enabling them to pursue more value-added jobs.

3

2. Research needs

CW comprises both inert and non-inert contents. Inert construction waste, including concrete, bricks, and rubble, can be reused in future projects. Non-inert materials such as bamboo, plastics, and paper have limited residual value and thus should be landfilled (HKEPD, 2019). To appropriately manage such waste according to its composition, the Hong Kong EPD launched the CWDCS in 2006. As shown in Fig. 1, the CWDCS imposes different levels of levy on contractors/waste haulers based on the amount of inert waste materials they dispose of. A truckload full of inert waste is charged HK$71 per ton at the public fills, while a levy of HK$200 per ton is imposed for disposal in landfill if it comprises less than 50% inert waste by weight. Loads comprising no less than 50% inert waste are directed to off-site sorting facilities (OSF) for a levy of HK$175 per ton.

Since implementation, the CWDCS has undoubtedly played a significant role in CW minimization in Hong Kong (Lu et al., 2015). However, it has been a challenge to efficiently and reliably gauge the CW composition, and thus decide whether the inert content meets the required proportion of 50% by weight. To implement the CWDCS, the EPD set up sensors such as weighbridges, rangefinders and cameras at the OSFs to measure surface features of the incoming waste loads, e.g., tare weight, depth, and photos of the top surface. The image at the upper right corner of Fig. 1 shows how the system works. Human inspectors sit in an office at the OSF entrance and, when a truck arrives, they need to evaluate the inert content proportion in the waste dump based on the given data (e.g., tare weight, depth and photos), and decide if it is acceptable to the facility (i.e., it comprises no less than 50% inert waste). In the event of dispute, they are even required to manually separate the entire truckload of waste to determine its composition. Such human-engaged practice is problematic not only because it is inefficient and laborious, but also owing to the issue of fatigue and sloppiness that can impair the inspectors' judgements, as well as the potential risk of corruption that will undermine the credibility of the system.
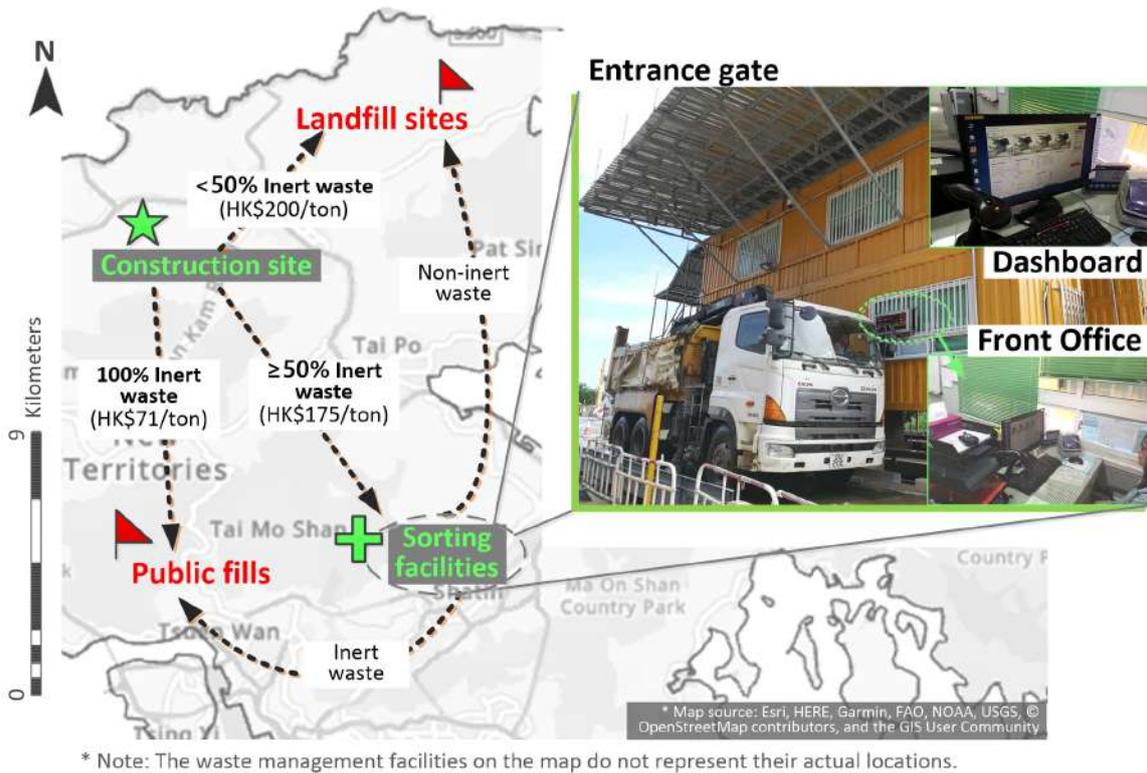
Fig. 1. A schematic diagram showing construction waste management practice in Hong Kong.

The conundrum in Hong Kong is not unique; instead, similar dilemmas surrounding solid waste management exist in Mainland China and around the world. Parallels can be found in a broad range of contexts (e.g., telediagnosis, minefield detection and structure condition assessment) where human experts are relied on to make judgements with very limited relevant surface information. Given humans' physical and mental limits, such judgements are prone to low efficiency while factors such as fatigue and corruption can lower productivity. For better efficiency and reliability in scenarios such as waste composition gauging, an automated approach is needed to replace human experts.

3. Technical viability

Recent technological developments (e.g., data mining, CV and ML) provide opportunities to overcome the limitations of human-reliant CW composition gauging. As society enters an era of big data (Donoho, 2000), big data analytics may offer solutions to problems that are traditionally difficult to solve in a small data context. The magnitude of available data is unprecedented, and by mining it latent knowledge, hidden patterns and unknown correlations can be extracted. Where a dataset is sufficiently large to include human decisions on a matter of interest and the factors that influence their decisions, one can reconstruct, from the dataset, the mapping relationship between the influencing factors and the human decisions. Such insights into how domain experts make decisions have been used to develop expert systems, which can operate independently for medical diagnosis (Malmir et al., 2017), automobile repair (Fang and Fang, 2013), structural condition assessment (Fabianowski et al., 2020) and other

domain-specific purposes.

If big data is the fuel driving the intelligence of machines, then the ML algorithm is the engine. The advent of deep learning in recent years provides powerful tools to process and analyze massive amounts of high-dimensional visual data. In the domain of CV, the processing of images for classification, detection and segmentation was traditionally dependent on domain expertise for feature extraction to reduce the complexity of the data (Mahapatra, 2018). CNN, a representative deep learning algorithm, revolutionized the domain by substituting the traditional problem-solving paradigm with an end-to-end approach. With such an approach, no human intervention is required (O'Mahony et al., 2019); instead, the networks can automatically learn high-level features from the input raw images in an incremental manner (Goodfellow et al., 2016). Owing to its robust performance and being free of feature handcrafting, CNN and its variants have been widely applied to solve, with super-human accuracy, problems that were assumed unsolvable (O'Mahony et al., 2019), e.g., biomedical image analysis (Tschandl et al., 2020), facial expression recognition (He et al., 2021), structure defects recognition (Dorafshan and Azari, 2020), and ambient intelligence-enabled healthcare (Haque et al., 2020).

External visual appearances only reveal some of the characteristics of a subject. To comprehensively characterize the subject, more information such as physical features should be taken into account. This is especially true for solid waste separation as materials sharing similar visual features, such as glass and a transparent plastic sheet, can maintain different physiochemical properties. Previous research has endeavored to use both visual and physical features to improve waste sorting performance. Chu et al. (2018) used waste images associated with numerical information measured by sensors to sort recyclable from other waste items. Koyanaka and Kobayashi (2011) considered both weight and 3D shape of waste fragments for the segregation of metal scraps from end-of-life vehicles. A prevalent ML algorithm for visual-physical feature fusion is support vector machine (SVM), which aims to find the optimal hyperplane that can best separate data samples of different classes. The superiority of SVM in integrating visual features extracted from CNN with other features has been demonstrated by a previous study (Xue et al., 2016).

The operation of the CWDCS over a span of more than 10 years in Hong Kong has resulted in a valuable big dataset comprising not only visual photos and physical properties (e.g., weight, height and depth) of tens of thousands of waste dumps, but also judgements made by inspectors regarding waste composition. With such a big dataset at hand, it is viable to apply emerging technologies such as CNN and SVM to train a hybrid model that can automatically gauge the proportion of inert content in CW.

4. The visual-physical feature hybrid approach

4.1. Visual feature extraction

Inert (e.g., concrete and bricks) and non-inert materials (e.g., wood, plastic and bamboo) in CW are visually distinct. Hence, top-down photos taken by surveillance cameras at OSFs can provide visual clues on whether the percentage of inert waste exceeds a certain level (e.g., 50%).

CNN is applied to exploit the visual differentiation between inert and non-inert contents to automate waste composition gauging. There are many different off-the-shelf CNN architectures, among which DenseNet has demonstrated robust performance in waste recognition by previous studies (Lu and Chen, 2020; Mao et al., 2021). Our study uses DenseNet169 (Huang et al., 2018), a variation of DenseNet, to extract critical high-level visual features for inert content gauging. As shown in Fig. 2, transfer learning technique (Zhang et al., 2019) is applied to train a classification model for which the input is top-down truckload photos captured by surveillance cameras and the output is labels indicating whether the transactions of the concerned truckloads were completed. The model comprises a DenseNet169 structure at the bottom, which has been pre-trained on ImageNet (Stanford Vision Lab et al., 2016), and three self-designed layers on the top, which include a global average pooling layer and two fully-connected layers (with 4,096 and 1 neurons respectively). The neuron of the last fully connected layer uses sigmoid as activation function, and outputs a scale value ranging from 0 to 1. Based on the scale value, a class label is assigned to the corresponding input image: if the value is no less than 0.5, the input image belongs to the "Completed" class (implying a composition with over 50% inert contents); otherwise, the image belongs to the "Rejected" class (implying inert content is less than 50%).
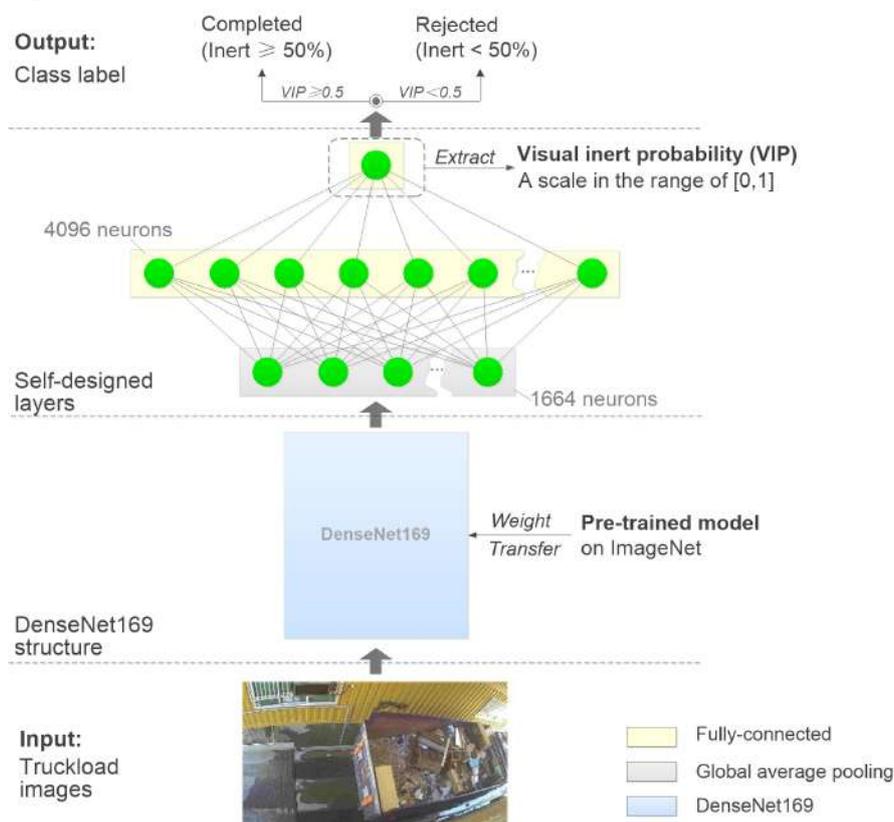
Fig. 2. Model structure based on DenseNet169 for visual feature extraction.

If the model is well trained, the output of the last fully connected layer forms a high-level indicator of proportion of inert content from a visual perspective. It also makes sense from the mathematical point of view, as the value gives the probability of the sample being positive (i.e., completed or inert $\geq 50\%$ in our case). Therefore, the model output scale value is extracted as a high-level visual feature of the concerned truckload, which we refer to as visual inert probability (*VIP*).

4.2. Physical features

Different materials have different physical properties. One of the most common and measurable physical properties is density. Most inert materials are of higher densities than their non-inert counterparts (Lu and Yuan, 2020). Thus, a waste load with more inert content is generally heavier than a load of similar volume with more non-inert content. In practice, it is both difficult and inefficient to directly measure the density of a truckload of construction waste. However, it is possible to reflect the variation of waste composition by developing a set of indirect features based on sensing data (e.g., weight and depth) collected at the OSFs. Here, four such feature indexes are introduced as follows.

(1) *iWD*

The *iWD* index is defined as a ratio of the gross weight of an incoming truck W_{in} to the *waste depth*, as shown by Eq. (1).

$$iWD = \frac{W_{in}}{waste\ depth} \tag{1}$$

$$waste\ depth = H_1 - H_2 \tag{2}$$

where H_1 and H_2 are the respective heights from the surface of the waste and the dump bed to the floor of weighbridge.

(2) *nWD*

The *nWD* index is defined as a ratio of the net weight of construction waste loaded by the incoming truck to the *waste depth*, as shown by Eq. (3).

$$nWD = \frac{W_{in} - W_{tare}}{waste\ depth} \tag{3}$$

where W_{tare} is the tare weight of the truck. Despite the absence of the bottom area of the loading bucket in the denominator, the index can be expected to reflect to a certain extent the overall density of the loaded waste materials, as a higher density usually (but not necessarily) indicates a higher net weight and a shallower waste depth.

(3) *GVW ratio*

The *GVW ratio* is defined in Eq. (4), which has been used by the Hong Kong EPD for

many years to guide the operation of OSFs. This demonstrates the significant role of the index in helping to infer the composition of inert contents, and thus is also considered as a potential physical feature for our model.

$$GVW\ Ratio = \frac{W_{in} - W_{tare}}{PGVW} \tag{4}$$

where *PGVW* is the permitted gross vehicle weight, which can be deemed as a fixed value that can be obtained from the manufacturing specifications of the corresponding truck model.

(4) *iGVW ratio*

The definition of *iGVW ratio* is similar to that of *GVW ratio*, but instead of net weight it uses the gross weight of an incoming truck as the numerator.

$$iGVW\ ratio = \frac{W_{in}}{PGVW} \tag{5}$$

The use of *PGVW* in Eq. (4) and (5) normalizes the respective indexes to allow direct comparison between different types of trucks with different loading capacity.

Among the above four physical features, *iWD* and *nWD* are dimensional variables with the unit of "ton/m", while *GVW ratio* and *iGVW ratio* are dimensionless variables. They were proposed based on an assumption that a truckload of waste with considerable portion of inert substance, given its density, should not be too deep and too light. However, the assumption does not guarantee their actual correlations with the corresponding inert waste proportion. Quantitative statistical analysis is thus required to select suitable ones from the four candidates for the hybrid model, as will be discussed in section 5.3.

4.3. Visual-physical hybrid model

In contrast to standalone visual (*VIP*) or physical features (*iWD*, *nWD*, *GVW ratio* and *iGVW ratio*), hybrid use of these features can allow a more comprehensive consideration of construction waste characteristics, leading to higher gauging accuracy. To train our inert content gauging model, support vector machine (SVM), a ML technique famous for its solid performance in classification problems and only requiring a small amount of training data, is used to fuse both the visual and physical features. Fig. 3 shows structure of the hybrid model based on SVM. The model takes a vector containing both visual and physical features proposed above as input, and then gives a prediction on whether the truckload corresponding to the input features contains more than 50% inert content by weight.
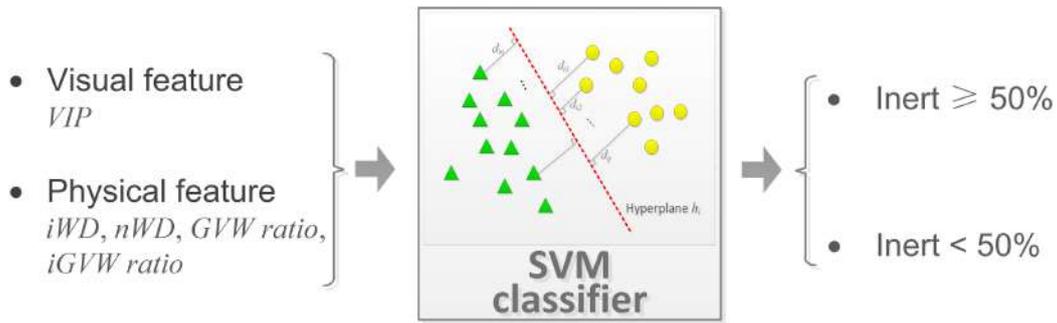
Fig. 3. Visual-physical feature hybrid model based on support vector machine.

5. Experimental studies

5.1. Data collection and preprocessing

Over the past few years, our research team has collected a large construction waste management dataset in Hong Kong (Lu, 2019; Lu and Yuan, 2020). In the dataset, records of truckloads received by the OSFs are of significance for this study. Fig. 4 shows a spreadsheet with information on all incoming trucks for one OSF in October 2019. The recorded data includes not only physical properties of the received waste, such as W_{in}, net weight, $PGVW$ and *weight depth*, but also top-down photos of all truckloads captured by surveillance cameras. Another data field is the record state of each truckload, which indicates whether the concerned load of waste was accepted or rejected by the facility. The record state "Completed" or "Rejected" provides an indicator of the proportion of the inert content—"Completed" implies a proportion of over 50%, while "Rejected" means the opposite.

The data includes 5,347 transaction records of 666 trucks over a duration of one month. Among these, only 296 transactions were rejected for not meeting the criterion of "more than 50% of inert contents by weight". Training ML models on such an imbalanced dataset can lead to overfitting, where the resulting model favors the majority class while performs poorly on the minority. To address this issue, the majority class (i.e., records with the "Completed" label) and the minority class (i.e., records with the "Rejected" label) were respectively downsampled and upsampled. The results were then further processed to remove records with missing data fields (e.g., *waste depth*), which resulted in a dataset of 1127 records. Finally, the dataset was divided into a training set, a validation set and a test set, according to a ratio of 7:1.5:1.5. The detailed composition is shown in Table 1.
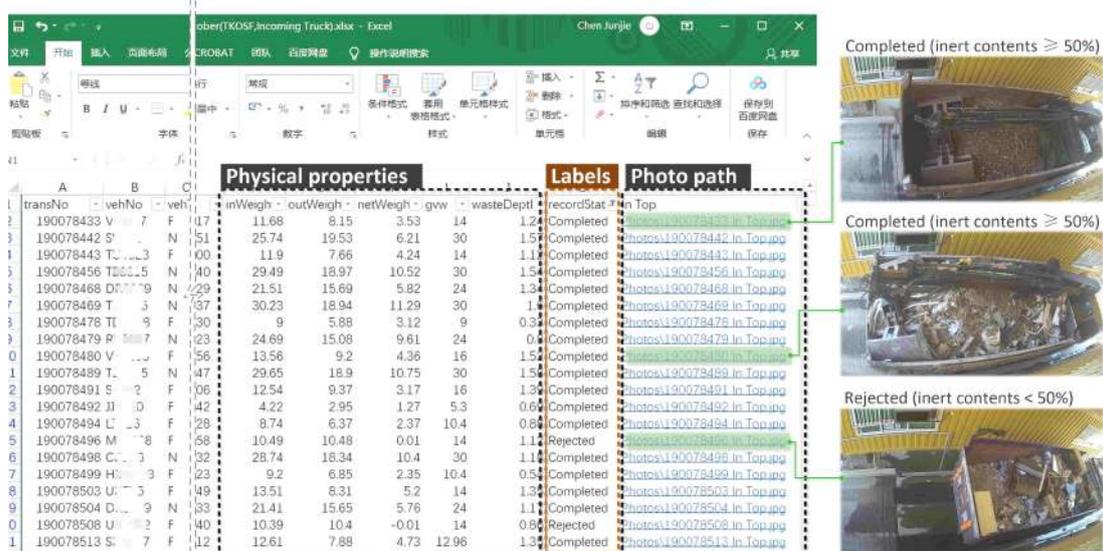
Fig. 4. Data records of received truckloads at an OSF.

Table 1. Composition of the dataset used for experimental studies.

Label	Dataset separation			
	Training	Validation	Test	Total
Completed (inert ≥ 50%)	415	91	91	597
Rejected (inert < 50%)	376	75	79	530
Total	791	166	170	1127

*Note: The numbers are quantities of records.

5.2. Results of visual recognition

A DenseNet169 model was fine-tuned on the training set for "completed (inert ≥ 50%)/rejected (inert < 50%)" classification. The Keras library on the TensorFlow 2 framework was used to train the model. The training parameters were set as follows: Adam algorithm with a learning rate of 0.001 was used for optimization; the used loss function was binary cross-entropy; batch size, step per epoch, and number of epochs were set as 4, 200 and 20, respectively.

Figs. 5 (a) and (b), respectively, depict the change of accuracy and loss during the 20 epochs of training. We observed that, despite the fluctuation, the evolvement of performance on the validation set was in general consistent with that on the training set in the first 10 epochs, both experiencing an increase on accuracy and a drop of loss. The observation indicates the model was being properly trained and was learning generalizable patterns from the training data. However, after 10 epochs, while the model performance kept improving on the training set, its accuracy on the validation set started decreasing drastically; meanwhile, the loss on the validation set remained fluctuating at the same level after 10 epochs. This phenomenon signifies the occurrence of overfitting, where the model performs extremely well on the training set but fails to generalize to the new samples in the validation/test set.
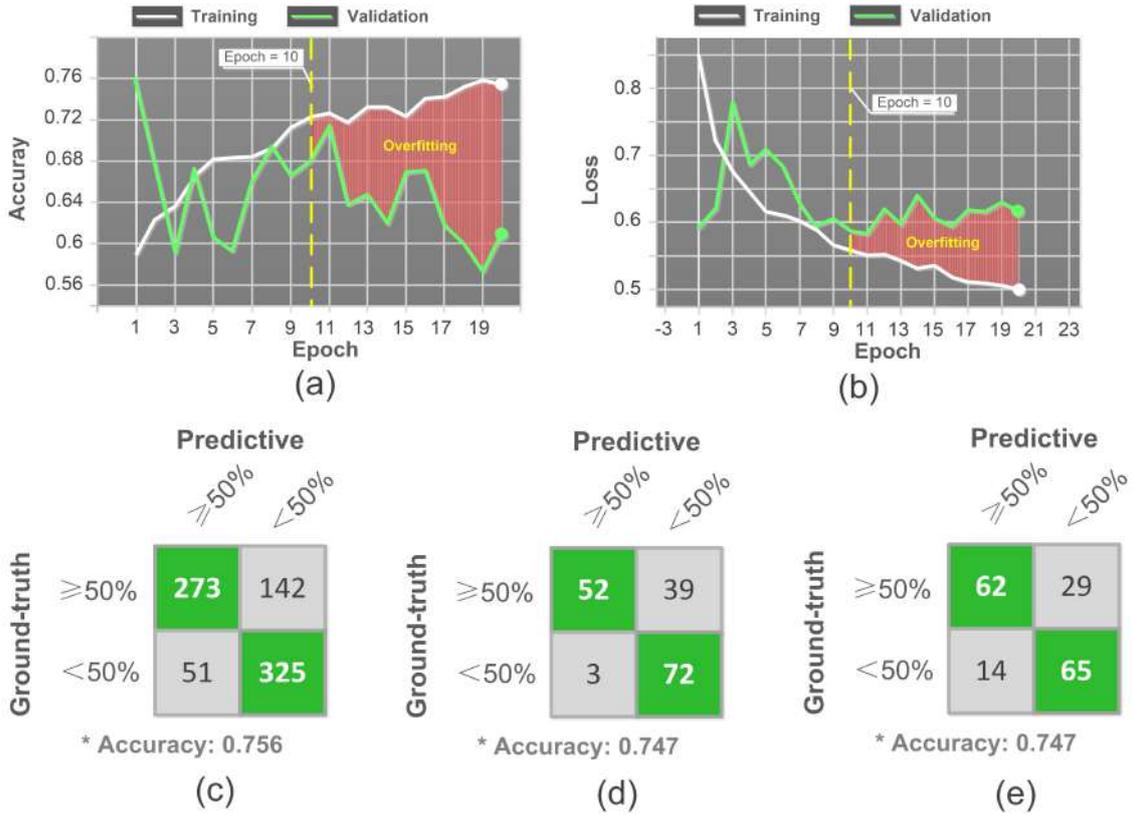
Fig. 5. Change of (a) accuracy and (b) loss during training process; confusion matrix of selected model on (c) training set, (d) validation set and (e) test set.

To prevent overfitting, the "early stopping" strategy was adopted (Brownlee, 2018a). The model after 10 epochs of training was selected for performance evaluation. Figs. 5 (c), (d), and (e) show the confusion matrices of the model on training set, validation set and test set, respectively. It can be observed that the accuracies on the three subsets are all around 0.75, demonstrating the model's ability to generalize to new test samples. Based on the model, the visual features, *VIP*, of all samples in the dataset can be extracted for further analysis.

5.3. Feature correlation analysis

The correlation between the proposed visual/physical features and waste composition was analyzed with Pearson's correlation coefficients. The analysis was performed in IBM SPSS based on the entire dataset with 1127 samples. Through correlation analysis, features with less significant coefficients can be excluded from the hybrid model.

Figs. 6 (a) ~ (e) show the distribution of the five features (1 visual, 4 physical) by box plots. Their Pearson's correlation coefficients with the "Completed/Rejected" labels are listed in Fig. 6 (f). The most significant feature is *VIP* with a coefficient of 0.531, followed by the physical features *nWD* and *iGVW ratio* with coefficients of 0.335 and 0.325, respectively. The correlation of the remaining physical features *iWD* and *GVW ratio* is comparatively weak, and thus we exclude them from the hybrid model. The correlation analysis implies that a truckload

with a higher *VIP*, a higher *net weight*/*waste depth*, and a higher *inweight*/*PGVW* is more likely to contain no less than 50% of inert materials, and hence is more likely to be accepted by an OSF.
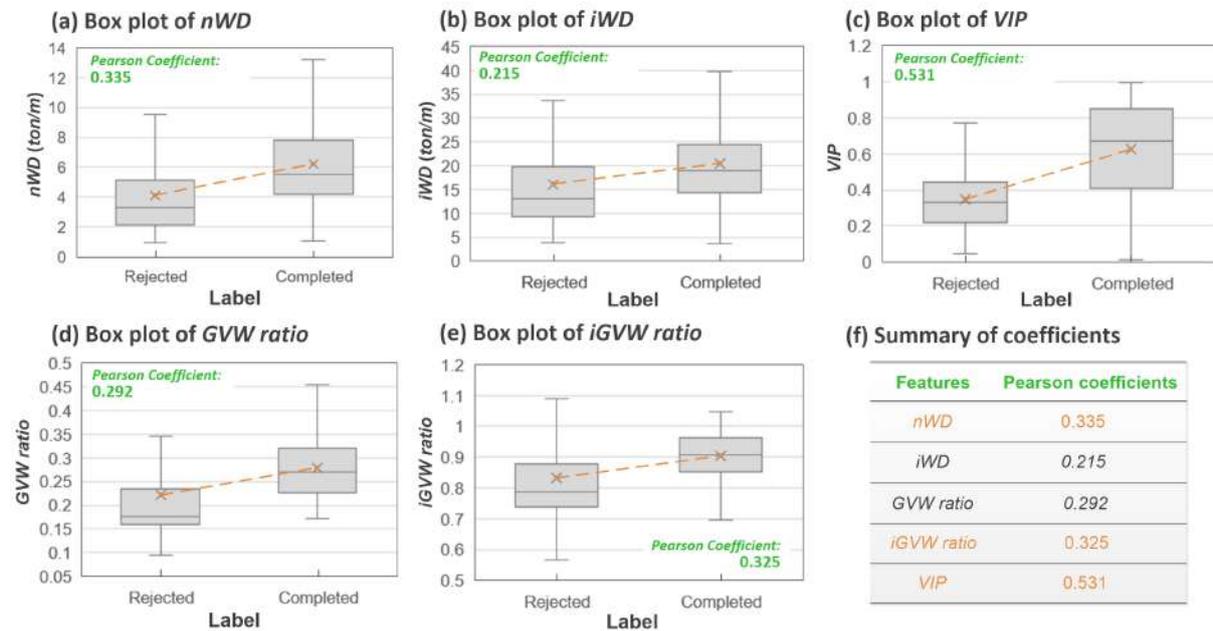
Fig. 6. Box plots of (a) *nWD*, (b) *iWD*, (c) *VIP*, (d) *GVW ratio*, and (e) *iGVW ratio*; (f) a table listing the Pearson's coefficients of all the five features.

5.4. Performance of the hybrid model

A hybrid SVM model integrating the visual feature *VIP* and physical properties *nWD* and *iGVW ratio* was trained to gauge whether a waste dump comprises more than 50% inert material. Due to the differences in dimension and range of distribution, feature values of all samples in the dataset were first normalized to the range of [0, 1] with the Min-Max normalization approach (Codecademy, 2020). The Python package scikit-learn was used to train and test the SVM. Model training and hyperparameter tuning was undertaken with 10-fold cross validation (Brownlee, 2018b). Detailed information of the hyperparameter tuning process can be found in the Supplementary Material.

When hyperparameters were specified as "kernel = rbf, C = 9000, and gamma = 6000", best performance in terms of prediction accuracy (i.e., 0.928) was observed on the training and validation set. The model trained with these hyperparameters was used to predict whether truckloads in the test set comprised more than 50% inert waste by weight. The prediction performance is shown in Fig. 7. The model's predictions had a very high chance (i.e., 94%) of being identical to evaluations made by human inspectors, which are deemed as the ground-truth. Such high accuracy demonstrates the model's promise for deployment in OSFs, replacing humans for efficient and reliable inert proportion gauging.

(a) Confusion matrix

(b) Performance metrics

	Precision	Recall	F1-score	Accuracy
Inert ≥ 50%	0.92	0.98	0.95	0.94
Inert < 50%	0.97	0.90	0.93	

Fig. 7. Performance of the hybrid model in predicting inert waste proportion on test set.

6. Discussion

6.1. A peek into the black box using visual recognition

The visual recognition model based on DenseNet169 achieved an accuracy of 74.7% on the test set, which is quite acceptable. This is especially true given that the input photos have not been preprocessed to remove background areas surrounding the trucks, and that the visual appearance of a waste dump does not necessarily reflect its interior composition. The *VIP* extracted from the visual recognition model correlated to the inert content proportion with a coefficient of 0.531, playing a critical role in the solid performance of the hybrid model. Therefore, it is important to understand what kind of patterns the visual model learnt that are crucial for distinguishing between waste loads with different inert proportions. A neural network has been traditionally deemed as a black box, inside which the learnt patterns are difficult for humans to interpret. Thanks to recent developments in explainable machine learning, it is now possible for us to glance into the black box through emerging techniques such as class activation mapping (CAM) (Zhou et al., 2016) and gradient-weighted CAM (Grad-CAM) (Selvaraju et al., 2017). We used Grad-CAM in this study because, unlike CAM, it does not require modification of the original network and model re-training (Chetoui, 2019). Grad-CAM calculates the importance weight of a feature map with respect to the output class via gradients, and then computes the weighted sum of all the feature maps to produce a heat map indicating salient regions that affect the model's decision.

14
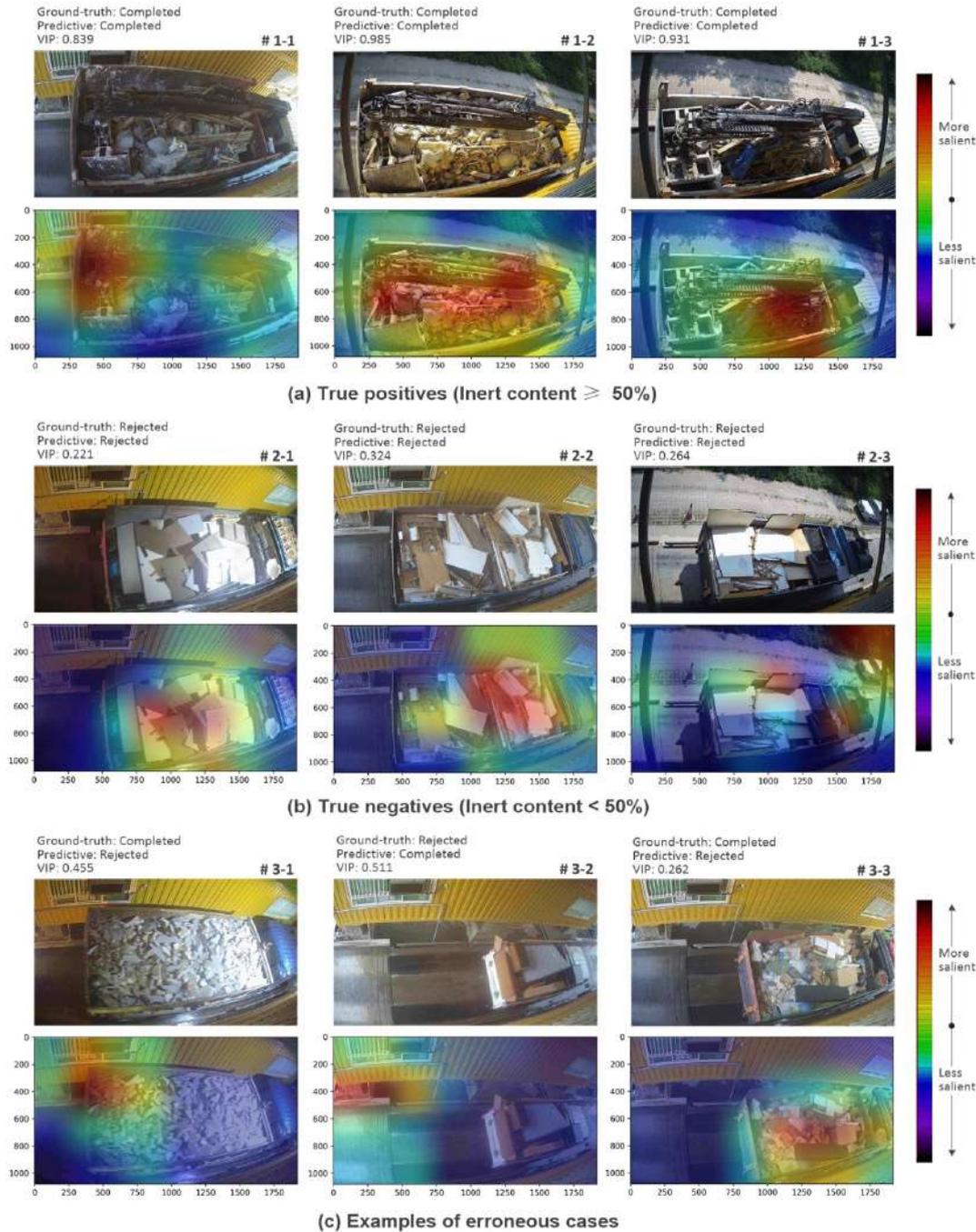
Fig. 8. Grad-CAM heatmaps indicating salient regions/features that affect the model decision.

Figs. 8 (a) and (b) show the heatmaps of the input images when the model correctly identified the proportion of inert contents. As can be seen from the figure, despite the existence of irrelevant background (e.g., ground, vegetation and exterior of toll gate office), the model has learnt to focus on pixel areas where the truck buckets are located, forming a premise for subsequent inert content gauging. For samples with more than 50% inert content (Fig. 8 (a)), image regions with salient features appear to be larger than those observed in samples with less than 50% inert content (Fig. 8 (b)). In addition, it seems the model tends to pay more attention to the grab buckets mounted on the truck, and to correlate these buckets with greater likelihood

of over 50% inert content. The phenomenon can be explained by empirical observations. Specifically, construction waste with greater inert content tends to weigh more, and hence is more likely to be loaded and transported by larger trucks with higher tonnage and grab buckets. The model has also acquired the visual pattern of non-inert materials, and successfully correlated pixel regions with bright colors and relatively smooth textures to the label of "inert < 50%" (Fig. 8 (b)).

Fig. 8 (c) shows Grad-CAM heatmaps of some erroneous cases. The full load of concrete fragments in #3-1 is not a normal case like those displayed in #1-1 ~ #1-3 where the different inert (and non-inert) materials are randomly mixed together. The model may have been confused by the relative infrequency of #3-1, leading to an incorrect judgement. In #3-2, the model fails to detect where the truck bucket is, and thus the incorrect prediction is not surprising. For #3-3, from the human visual perspective, the sample looks as though it contains a large portion of non-inert materials, consistent with the predictive label. However, there might be a large amount of inert content hidden in a lower layer, which cannot be picked up by visual sensors. This highlights the importance of integrating both visual and physical features to achieve a more comprehensive evaluation.

6.2. Adaptability of the proposed approach to broader contexts

Although developed in the Hong Kong context, the proposed approach can be adapted to waste management dilemmas in other countries/regions. The differences between inert and non-inert waste with respect to physical properties and visual appearance are universal, so the visual feature *VIP* and physical features *iWD*, *nWD*, *GVW ratio* and *iGVW ratio* can serve as universal indicators to gauge inert and non-inert contents. With regards to interior waste composition corresponding to each of waste dumps, this can be obtained via approaches applicable to the local context. In our study, such information was inferred from the OSF binary labels indicating whether the waste dumps were "Completed" (and hence comprised more than 50% inert waste) or "Rejected" (comprising less than 50% inert waste). Thus, the model is actually a binary classifier gauging whether the relevant mixed waste is made up of more than 50% inert content. This is not necessarily the case in other countries/regions, as those places might manage construction waste using trisection or quarter division. One can even meticulously segregate and probe into the waste mixture to get a continuous and accurate waste composition value of each truckload. In this way, it is viable to reconstruct a regression model (rather than a classifier) that can make explicit the waste dump composition. In addition, the incoming trucks do not have to be inspected one by one through the toll gates; rather, it is worth exploring to deploy unmanned aerial vehicles (UAV) and object detection models to enable inspection of multiple trucks at the same time. In such a way, the efficiency can be further improved.

6.3. Significance of the research findings

The experimental studies demonstrate a 94% accuracy of our hybrid approach in gauging

composition of CW. To evaluate the performance, results obtained from the field of waste image classification can be used for comparison. State-of-the-art performance of waste classification on TrashNet (Thung and Yang, 2019), a public open trash dataset, is around 96% (Huang et al., 2020; Yang and Li, 2020). The TrashNet dataset mainly consists of photos of individual waste items appeared on relatively simple background. Compared with the classification on TrashNet, it is more difficult to recognize composition of a bulk of mixed waste. Thus, the 94% accuracy of our approach is rather satisfactory.

The visual-physical feature hybrid model proposed in this study contributes to the general problem of "looking beneath the surface". To get to the essence of something, humans often have to make inferences from external appearances. Knowledge of such human inference can be extracted to train intelligent machines. This is demonstrated by our study from a perspective of waste management, which has successfully reconstructed the mapping between the surface features of waste dumps and their interior composition from more than 1,000 pieces of data. While the idiom "seeing is believing" reflects the importance of visual information, our study implies depending merely on vision can bring bias. One possible way to overcome such bias is to consider information from other sources to complement visual features. When visual and physical features were integrated, satisfactory performance was attained. From a practical point of view, the proposed approach facilitates the trend towards unmanned/unattended gauging of CW composition, reducing investment in manpower, improving efficiency, and also ensuring reliability and credibility of waste management systems. Our approach vividly demonstrates how unattended operations, driven mainly by algorithms and big data, can address the problems associated with human-in-the-loop operations.

7. Conclusions

There are many scenarios in which human experts need to make inferences about something from its outward appearance. Likewise, current practice in gauging construction waste composition requires human inspectors to judge whether the inert content in a waste dump exceeds required proportions based on limited information. This practice is inefficient, laborious, hazardous, and can be undermined by fatigue, sloppiness and corruption. To address these limitations, this paper presents an automated approach for inert waste content gauging through the hybrid use of both visual and physical features of construction waste. A fine-tuned CNN model was trained to extract a high-level feature index called visual inert probability (*VIP*) from more than 1,000 photos of waste dumps. The *VIP* demonstrated a correlation coefficient of 0.531 with inert content proportion. It thus can provide important visual clues on the composition of construction waste. Four different physical features were defined based on the collected sensing data (e.g., weight and depth), of which the *nWD* and *iGVW ratio* presented moderate correlations with inert content proportion. Our visual-physical feature hybrid approach achieved 94% accuracy in gauging construction waste composition at an off-site sorting facility in Hong Kong. The accuracy is almost equally as well as human inspectors.

Since it can work automatically without human intervention, this approach is not prone to fatigue or bribery from waste haulers/contractors. The solid performance of the proposed approach and its advantages thus demonstrate its promise in replacing human inspectors at waste disposal facilities for unattended operation.

Future studies are recommended to further facilitate waste composition gauging. First, semantic segmentation techniques such as DeepLabv3 can be applied to recognize the specific types of material on the surface of the waste dumps, which will provide more visual information on waste composition, and thus can potentially lead to higher accuracy. Second, the concept of bulk density can be utilized to better characterize the physical properties of waste dumps. As bulk density is directly correlated to waste composition from a theoretical perspective, incorporating it into the hybrid model can be expected to further improve the model performance.

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.

References

Arantes, V., Zou, C., Che, Y., 2020. Coping with waste: a government-NGO collaborative governance approach in Shanghai. J. Environ. Manage. 259, 109653. https://doi.org/10.1016/j.jenvman.2019.109653.

Arebey, M., Hannan, M.A., Begum, R.A., Basri, H., 2012. Solid waste bin level detection using gray level co-occurrence matrix feature extraction approach. J. Environ. Manage. 104, 9-18. https://doi.org/10.1016/j.jenvman.2012.03.035.

Bennett, J., 1999. Remote sensing of surface UXO with active laser and passive infrared airborne line scanner. https://www.serdp-estcp.org/Program-Areas/Munitions-Response/Land/Sensors/MR-199523 (Accessed 17 February 2021).

Bhola, R., Krishna, N.H., Ramesh, K.N., Senthilnath, J., Anand, G., 2018. Detection of the power lines in UAV remote sensed images using spectral-spatial methods. J. Environ. Manage. 206, 1233-1242. https://doi.org/10.1016/j.jenvman.2017.09.036.

BHS, 2017. Max-AI™ autonomous QC - Youtube. https://www.youtube.com/watch?v=Q7tE_vNYzzU (Accessed 3 November 2020).

Brownlee, J., 2018a. A gentle introduction to early stopping to avoid overtraining neural networks. https://machinelearningmastery.com/early-stopping-to-avoid-overtraining-neural-network-models/ (Accessed 9 February 2021).

Brownlee, J., 2018b. A gentle introduction to k-fold cross-validation. https://machinelearningmastery.com/k-fold-cross-validation/ (Accessed 10 November 2020).

Chen, J., Liu, D., 2021. Bottom-up image detection of water channel slope damages based on superpixel segmentation and support vector machine. Adv. Eng. Inform. 47, 101205. https://doi.org/10.1016/j.aei.2020.101205.

Chetoui, M., 2019. Gradient-weighted class activation mapping - Grad-CAM. https://medium.com/@mohamedchetoui/grad-cam-gradient-weighted-class-activation-mapping-ffd72742243a (Accessed 25 November 2020).

Chu, Y., Huang, C., Xie, X., Tan, B., Kamal, S., Xiong, X., 2018. Multilayer hybrid deep-learning method for waste classification and recycling. Comput. Intel. Neurosc. 2018, 5060857. https://doi.org/10.1155/2018/5060857.

Codecademy, 2020. Min-Max normalization. https://www.codecademy.com/articles/normalization#:~:text=Min-max%20normalization%20is%20one,decimal%20between%200%20and%201 (Accessed 25 November 2020).

Donoho, D.L., 2000. High-dimensional data analysis: the curses and blessings of dimensionality, AMS math challenges lecture. Citeseer, p. 32.

Dorafshan, S., Azari, H., 2020. Deep learning models for bridge deck evaluation using impact echo. Constr. Build. Mater. 263, 120109. https://doi.org/10.1016/j.conbuildmat.2020.120109.

Fabianowski, D., Jakiel, P., Stemplewski, S., 2020. Development of artificial neural network for condition assessment of bridges based on hybrid decision making method – feasibility study. Expert Syst. Appl. 168, 114271. https://doi.org/10.1016/j.eswa.2020.114271.

Fang, X., Fang, K., 2013. Knowledge-based auto repair diagnosis system and application, 2nd International Conference on Advances in Computer Science and Engineering (CSE 2013). Atlantis Press.

Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y., 2016. Deep learning. MIT press Cambridge.

Haque, A., Milstein, A., Fei-Fei, L., 2020. Illuminating the dark spaces of healthcare with ambient intelligence. Nature 585, 193-202. https://doi.org/10.1038/s41586-020-2669-y.

He, L., Chan, J.C., Wang, Z., 2021. Automatic depression recognition using CNN with attention mechanism from videos. Neurocomputing 422, 165-175. https://doi.org/10.1016/j.neucom.2020.10.015.

HKEPD, 2019. Management of abandoned construction and demolition materials. https://www.aud.gov.hk/pdf_e/e67ch04sum.pdf (Accessed 20 November 2020).

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.

Huang, G., Liu, Z., Maaten, L.v.d., Weinberger, K.Q., 2018. Densely Connected Convolutional Networks, arXiv preprint.

Koch, C., Georgieva, K., Kasireddy, V., Akinci, B., Fieguth, P., 2015. A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure. Adv. Eng. Inform. 29, 196-210. https://doi.org/10.1016/j.aei.2015.01.008.

Koyanaka, S., Kobayashi, K., 2011. Incorporation of neural network analysis into a technique for automatically sorting lightweight metal scrap generated by ELV shredder facilities. Resour. Conserv. Recycl. 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.

Larner, A.J., 2020. Cognitive testing in the COVID-19 era: can existing screeners be adapted for telephone use? Neurodegenerative Disease Management. https://doi.org/10.2217/nmt-2020-0040.

Li, S., Zhu, S., Xu, Y.-L., Chen, Z.-W., Li, H., 2012. Long-term condition assessment of suspenders under traffic loads based on structural monitoring system: application to the Tsing Ma Bridge. Structural Control and Health Monitoring 19, 82-101. https://doi.org/10.1002/stc.427.

Lu, W., 2019. Big data analytics to identify illegal construction waste dumping: a Hong Kong study. Resour. Conserv. Recycl. 141, 264-272. https://doi.org/10.1016/j.resconrec.2018.10.039.

Lu, W., Chen, J., 2020. Computer vision for solid waste segregation: a critical review of academic research. Renewable and Sustainable Energy Reviews (Manuscript submitted for publication).

Lu, W., Chen, X., Peng, Y., Shen, L., 2015. Benchmarking construction waste management performance using big data. Resour. Conserv. Recycl. 105, 49-58. https://doi.org/10.1016/j.resconrec.2015.10.013.

Lu, W., Yuan, L., 2020. Investigating the bulk density of construction waste: a big data-driven

approach. Resour. Conserv. Recycl. (Manuscript accepted for publication).

Maathuis, B.H.P., 2003. Remote sensing based detection of minefields. Geocarto. Int. 18, 51-60. https://doi.org/10.1080/10106040308542263.

Mahapatra, S., 2018. Why deep learning over traditional machine learning? https://towardsdatascience.com/why-deep-learning-is-needed-over-traditional-machine-learning-1b6a99177063 (Accessed 11 December 2020).

Malmir, B., Amini, M., Chang, S.I., 2017. A medical decision support system for disease diagnosis under uncertainty. Expert Syst. Appl. 88, 95-108. https://doi.org/10.1016/j.eswa.2017.06.031.

Mao, W., Chen, W., Wang, C., Lin, Y., 2021. Recycling waste classification using optimized convolutional neural network. Resour. Conserv. Recycl. 164, 105132. https://doi.org/10.1016/j.resconrec.2020.105132.

Mubasir, M., 2020. Oil storage tank's volume occupancy on satellite imagery using YOLOv3. https://towardsdatascience.com/oil-storage-tanks-volume-occupancy-on-satellite-imagery-using-yolov3-3cf251362d9d (Accessed 11 November 2020).

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.

O'Mahony, N., Campbell, S., Carvalho, A., Harapanahalli, S., Hernandez, G.V., Krpalkova, L., Riordan, D., Walsh, J., 2019. Deep learning vs. traditional computer vision, Science and Information Conference. Springer, pp. 128-144.

Sauve, G., Van Acker, K., 2020. The environmental impacts of municipal solid waste landfills in Europe: a life cycle assessment of proper reference cases to support decision making. J. Environ. Manage. 261, 110216. https://doi.org/10.1016/j.jenvman.2020.110216.

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad-CAM: visual explanations from deep networks via gradient-based localization, 2017 IEEE International Conference on Computer Vision (ICCV), pp. 618-626.

Stanford Vision Lab, Stanford University, Princeton University, 2016. Imagenet. http://image-net.org/ (Accessed 28 October 2020).

Thung, G., Yang, M., 2019. Trashnet dataset. https://github.com/garythung/trashnet (Accessed 9 February 2021).

Tschandl, P., Rinner, C., Apalla, Z., Argenziano, G., Codella, N., Halpern, A., Janda, M., Lallas, A., Longo, C., Malvehy, J., Paoli, J., Puig, S., Rosendahl, C., Soyer, H.P., Zalaudek, I., Kittler, H., 2020. Human–computer collaboration for skin cancer recognition. Nat. Med. 26, 1229-1234. https://doi.org/10.1038/s41591-020-0942-0.

Watts, W., 2019. How oil traders are using satellites to keep an eye on an increasingly unpredictable market. https://www.marketwatch.com/story/how-oil-traders-are-using-satellites-to-keep-an-eye-on-an-increasingly-unpredictable-market-2019-10-04 (Accessed 17 February 2021).

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.

Xue, D.-X., Zhang, R., Feng, H., Wang, Y.-L., 2016. CNN-SVM for microvascular morphological type recognition with data augmentation. J. Med. Biol. Eng. 36, 755-764. https://doi.org/10.1007/s40846-016-0182-4.

Yang, Z., Li, D., 2020. Wasnet: a neural network-based garbage collection management system. IEEE ACCESS 8, 103984-103993. https://doi.org/10.1109/access.2020.2999678.

Yu, K., 2019. SBS Language | Shanghai residents are being driven crazy by new trash-sorting rules. https://www.sbs.com.au/language/english/audio/shanghai-residents-are-being-driven-crazy-by-new-trash-sorting-rules_1 (Accessed 11 November 2020).

Zhang, D.Q., Tan, S.K., Gersberg, R.M., 2010. Municipal solid waste management in China: status, problems and challenges. J. Environ. Manage. 91, 1623-1633. https://doi.org/10.1016/j.jenvman.2010.03.012.

Zhang, Y., Li, M., Han, S., Ren, Q., Shi, J., 2019. Intelligent identification for rock-mineral microscopic images using ensemble machine learning algorithms. Sensors 19, 3914. https://doi.org/10.3390/s19183914.

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A., 2016. Learning deep features for discriminative localization, Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2921-2929.

Zuo, M., Yan, A., 2019. Shanghai begins new waste sorting era, as China eyes cleaner image. https://www.scmp.com/news/china/society/article/3016801/shanghai-begins-new-waste-sorting-era-china-eyes-cleaner-image (Accessed 2 November 2020).

Supplementary material:

S1. Hyperparameter Tuning for the Hybrid SVM model

When training a support vector machine (SVM) model with Python scikit-learn package, three important hyperparameters need to be specified, i.e., kernel, C and gamma. The "kernel" designates the type of hyperplane used to separate the data. For example, "linear" kernel is effective for linear-separable data while the other kernels (e.g., "poly", and "rbf") are normally used to deal with nonlinear data. C is the regularization parameter that defines the tolerance level of misclassification, while the gamma defines the range of influence a training sample should reach.

We used 10-fold cross validation to tune the above three hyperparameters. The implementation of k-fold cross validation has two benefits. First, it allows full utilization of the dataset for training since a separate validation set is not needed. This is critically important when the amount of available data is limited. Second, it mitigates the adverse influence of different data splits on the performance metrics, thus allowing for a more objective performance evaluation. The training and validation sets in Table 1 of the paper were combined to form a new dataset (including 957 samples in total) for 10-fold cross validation. Table S1 lists the results of 10-fold cross validation under different hyperparameters. When hyperparameters were specified as "kernel = rbf, C = 9000, and gamma = 6000", best performance in terms of prediction accuracy (i.e., 0.928) was observed on the dataset.

Table S1. Cross validation results of the hybrid model trained with different hyperparameters.

No.	Hyperparameters				Performance metrics		
	Kernel	degree[*]	C	gamma	Training time/fold (s)	accuracy	F1-score
1		2	300	10	1.67	0.757	0.756
2	Poly	2	5000	40	98.3	0.771	0.771
3		2	100	0.5	0.01	0.76	0.759
4		3	100	0.4	0.01	0.766	0.764
5		/	100	0.8	0.01	0.85	0.85
6		/	300	10	0.06	0.866	0.866
7	rbf	/	600	120	0.06	0.889	0.889
8		/	600	300	0.03	0.91	0.91
9		/	9000	6000	0.04	0.928	0.927

[*] The hyperparameter is only applicable to "Poly" kernel, which represents the degree of the polynomial kernel