Particle Size Estimation in Mixed Commercial Waste Images Using Deep Learning

Phongsathorn Kittiworapanya*, Kitsuchart Pasupa[†] and Peter Auer[‡]

*[†]Faculty of Information Technology, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand [‡]Department of Mathematics and Information Technology, Montanuniversität Leoben, Leoben 8700, Austria Email: *60070055@kmitl.ac.th, [†]kitsuchart@it.kmitl.ac.th, [‡]auer@unileoben.ac.at

Abstract—We assessed several state-of-the-art deep learning algorithms and computer vision techniques for estimating the particle size of mixed commercial waste from images. In waste management, the first step is often coarse shredding, using the particle size to set up the shredder machine. The difficulty is separating the waste particles in an image, which can not be performed well. This work focused on estimating size by using the texture from the input image, captured at a fixed height from the camera lens to the ground. We found that EfficientNet achieved the best performance of 0.72 on F_1 -Score and 75.89% on accuracy.

Index Terms—Deep Learning, Computer Vision, Waste Management, Mixed Commercial Waste, Size Estimation

I. INTRODUCTION

With the growing amount of mixed commercial waste, waste management plans become necessary [1]. The first step in waste management is coarse shredding for treating solid municipal and commercial waste [2]–[4], followed by one or more screening stages. Shredding reduces the particle size. To optimize the shredding machine performance, the size and material of waste particle needs to be known [5]. Generally, human operators manually estimate size and classify material using screens, magnetic separators or sensor-based sorting machines.

With the improvement of computer and robotic technology, smart waste factory concepts automated waste management, reaching higher sorting and recycling rates. Recycling and Recovery of Waste 4.0 (ReWaste 4.0) [6] is a concept in which a smart waste factory consists of several waste treatment plants, which perform different tasks and are interconnected via data stream and logistic systems (*e.g.* sorting plants, production plants for solid recovered fuels, *etc.*). This connection enables dynamic process control and various actuator systems actively intervene in the processes. The new industry 4.0 approaches (*e.g.* robotic, big data, digital networking, *etc.*) will be used in the waste management to achieve smart waste factory goals.

We describe a method to estimate the size of waste particles from images of mixed commercial waste, using machine learning and computer vision techniques.

The paper was organized as follows. Section II describes related work on waste estimation using computer vision and

† Corresponding Author

machine learning, Section III explains methods used, including data acquisition, data preprocessing, algorithms for size estimation and experiments, Section IV discusses the results, and Section V concludes.

II. RELATED WORK

Many advanced computer vision and machine learning techniques have been used to estimate the size of waste particles. To characterize waste particles, Di Maria *et al.* [7] estimated the particle size distribution based on image analysis, *e.g.* image descriptors, to find image signatures (feature vectors), auto-correlation, circular filters and granulometry (a set of morphological filters at different scales). Kandlbauer *et al.* [8] analysed individual waste particle images with a set of image transformations to detect particles and a regression model to estimate sizes. Zhang *et al.* [9] and Dunnu *et al.* [10] used dimensionless coefficients, *e.g.* roundness or ratios between the projected area of the object, and the area of a bounding box, to describe the shape and size of particles.

Some of these works estimated object size of other than the waste particle by using an image. Ponce *et al.* [11] analysed images to estimate the size of olive fruit, *e.g.* image segmentation by combining the mask of the global threshold using Otsu's method with the separated inverted value (V) and saturation (S) channels of the image in HSV color space, and estimate the size using a linear estimation model. Oo *et al.* [12] estimated the size of strawberries using image analysis with kite geometry, *e.g.* parallograms in the shape of a traditional kite [13], which is similar in shape to a strawberry, and calculated the size from the segmentation mask.

None of these works estimated particle sizes from input images of mixed commercial waste. It is complex and difficult to detect each particle in waste, *e.g.* the label on each can and bottle and overlapping particles, can degrade waste particle detection, and estimated size from detected particles will not be accurate. Since convolutional neural networks (CNNs) archived significant success for texture classification [14], our hypothesis was that we could estimate waste particle size, using image texture analysis with deep learning techniques, *i.e.* ResNet-18, ResNet-50 [15], VGG-16 [16], MobileNet-V2 [17] and EfficientNet [18].



Fig. 1. Data acquisition: (a) Waste heap imaged by a camera mounted at a fixed height, H, from ground to camera optical center, (b) Collection of mixed waste images, and (c) Stitched image from (b) showing the partitions.

III. METHODOLOGY

Here we describe the acquisition and description of the dataset, data preprocessing and augmentation, and algorithms for particle size estimation on mixed commercial waste images.

A. Data acquisition

Since there is no publicly available dataset for size estimation from images of mixed commercial waste, we collected and created a new dataset. Our Mixed Commercial Waste Size dataset contains mixed waste images with different particle size ranges, including 20–40, 40–60, 60–80, 80–100, 100– 200, and 200–400 mm. We formed six heaps of objects in the desired range of sizes. The camera was a Nikon D7500, taking 2, 784×1 , 856 pixel RGB images. It was mounted at a fix height from the ground to the camera optical center. Each image contained particles with similar sizes. Fig. 1 shows the data acquisition process, and Fig. 2 shows samples of mixed commercial waste image.

After we collected the images, we found that some parts of one image overlapped with other images. This problem could bias the experiment, because some sections of overlapped images might appear multiple times in test data. Therefore, we created a new dataset by stitching all images, in each size range, using Image Composite Editor [19]. Then, we split the stitched image into many 720×720 pixel images. Table I shows the distribution of our dataset.



Fig. 2. Sample of mixed commercial waste, grouped in six size ranges: (a) 20-40 mm, (b) 40-60 mm, (c) 60-80 mm, (d) 80-100 mm, (e) 100-200 mm, and (f) 200-400 mm.

TABLE I Distribution of dataset

Size range [mm]	Number of images	Fraction [%]
20-40	17	18
40-60	14	15
60-80	16	17
80-100	12	12
100-200	25	26
200-400	12	12
Total	96	100

B. Data pre-processing and augmentation

After we split the images, the RGB images were resized, without changing the aspect ratio, and zero padded to 224×224 pixel squares to match ResNet-18, ResNet-50, VGG-16 and EfficientNet requirements or 256×256 pixels for MobileNet-V2. To match the software requirements, images were normalized, so that pixels were floating point values in (0, 1).

After preprocessing, we randomly split the dataset into k non-overlapping partitions, with k = 5, using nested cross-validation [20], [21]. The partitions, that were selected for training data, were augmented to increase the number of images. To augment the dataset, a one or more of these operations—FancyPCA with $\sigma = 0.6$ [22], horizontal and vertical flipping, image transpose, cutout [23] with square patch size randomly set from 120 to 200 pixels and color channel shuffling—were applied to the images. Although the data set was augmented, the proportion of each class remained the same.

FancyPCA augmented image by altering the intensities of RGB channels, while maintaining the features and detail of

the image. The steps were:

- Images were reformatted to match the software input requirements, *i.e.* $w \times h$ matrices of RGB triples were converted to wh vectors of RGB triples,
- Principle Component analysis was used to find the eigenvalues, P, and eigenvectors, λ from the covariance matrix,
- · Eigenvectors were sorted in order of the eigenvalues,
- A random value, α , was drawn from a zero mean gaussian with standard deviation, σ . We added multiples of the principal components with magnitudes proportional to the corresponding eigenvalues times, α . Therefore, to each RGB image pixel, $I_{xy} = [I_{xy}^R, I_{xy}^G, I_{xy}^B]$, we added the following quantity:

$$[P_1, P_2, P_3][\alpha_1\lambda_1, \alpha_2\lambda_2, \alpha_3\lambda_3]^T$$
(1)

where P_i and λ_i are i^{th} eigenvectors and eigenvalues of the 3 × 3 co-variance matrix of RGB pixel values, respectively.

C. Algorithms for particle size estimation

We started with the selected state-of-the-art CNN models, including ResNet-18, ResNet-50, VGG-16, MobileNet-V2 and EfficientNet. Particle size was classified as belonging the size ranges listed in Table I.

D. Experiments

We evaluated the performance of size estimation, using the selected CNN models trained and tested on our dataset. Since the data was limited, we used nested cross-validation. It used an outer k iteration cross-validation loop to divide data into training and test folds, and an inner loop iterated over each of the k models. We divided the dataset into five partitions, retaining the proportion of each size range and evaluated the tuned algorithms to ensure that we used every sample for evaluating the algorithm.

The five selected CNN models were fine-tuned from the pretrained ImageNet [24] classification. Models were trained for 300 epochs, using Stochastic Gradient Descent (SGD), to find optimal parameters, with Nesterov momentum [25] set to 0.9 and learning rate to 0.0001. We selected the optimal models for each inner-loop of nested cross-validation to evaluate the model performance with data in the test fold on the outer-loop.

IV. RESULTS AND DISCUSSION

Table II reports the average performance across each of the outer test partitions by using evaluation matrices—Precision, Recall, F_1 -Score, and Accuracy—for the selected algorithms. We found that all algorithms were reasonably successful and generally assigned the correct size range to an image. EfficientNet performed best on every evaluation metric, with 0.72 on F_1 -Score and 75.89% on accuracy, which benefitted from applying compound scaling to all three dimensions of the CNN: width, depth and resolution. It was followed by ResNet-50 with 0.71 on F_1 -Score and 74.95% on accuracy. Fig. 3 shows the confusion matrix comparing EfficientNet

and VGG-16, calculated by summing the predicted and actual values on each testing iteration from the outer-loop of nested cross-validation. EfficientNet performed better than VGG-16 for the smallest particles (20–40 mm), but neither algorithm performed well on larger particles, between 80–100 mm to 200–400 mm.

TABLE II EVALUATION SCORES FOR FIVE ALGORITHMS

Model	Precision	Recall	F ₁ -Score	Accuracy [%]
ResNet-18 ResNet-50 VGG-16 EfficientNet MobileNet-V2		$ \begin{vmatrix} 0.65 \pm 0.11 \\ 0.72 \pm 0.15 \\ 0.70 \pm 0.07 \\ \textbf{0.73} \pm \textbf{0.10} \\ 0.71 \pm 0.13 \end{vmatrix} $	$ \begin{vmatrix} 0.62 \pm 0.11 \\ 0.70 \pm 0.16 \\ 0.67 \pm 0.09 \\ 0.72 \pm 0.08 \\ 0.69 \pm 0.14 \end{vmatrix} $	$\begin{array}{c} 66.68 \pm 13.13 \\ 74.95 \pm 16.47 \\ 73.89 \pm 09.28 \\ \textbf{75.89} \pm \textbf{11.08} \\ 71.84 \pm 15.22 \end{array}$

				VGG-16							
20-40	15 15.3%	1 1.0%	1 1.0%	0 0.0%	0 0.0%	0 0.0%	17 88.2% 11. <mark>8%</mark>				
40-60	2 2.0%	11 11.2%	1 1.0%	0 0.0%	0 0.0%	0 0.0%	14 78.6% <mark>21.4%</mark>				
d 60-80	0 0.0%	2 2.0%	14 14.3%	0 0.0%	1 1.0%	0 0.0%	17 82.4% 17. <mark>6%</mark>				
redicte 80-100	0 0.0%	0 0.0%	2 2.0%	6 6.1%	4 4.1%	1 1.0%	13 46.2% 53.8%				
P 100-200	0 0.0%	0 0.0%	1 1.0%	1 1.0%	20 20.4%	3 3.1%	25 80.0% <mark>20.0%</mark>				
200-400	0 0.0%	0 0.0%	0 0.0%	0 0.0%	2 2.0%	10 10.2%	12 83.3% <mark>16.7%</mark>				
	17 88.2% 11.8%	14 78.6% 21.4%	19 73.7% <mark>26.3%</mark>	7 85.7% 14.3%	27 74.1% <mark>25.9%</mark>	14 71.4% 28.6%	98 77.6% 22.4%				
	20-40	20-40 40-60 60-80 80-100 100-200 200-400 Actual									
EfficientNet											
			Eff	ficient	let						
20 - 40	17 17,3%	0 0,0%	Eff 0.0%	ficientN 0 0.0%	let 0.0%	0 0.0%	17 100% 0.0%				
40-60 20-40	17 17.3% 2 2.0%	0 0.0% 9 9.2%	Eff 0.0% 3.1%	ficientN 0.0% 0.0%	let 0.0% 0.0%	0 0.0% 0.0%	17 100% 0.0% 14 64.3% 35.7%				
d 60-80 40-60 20-40	17 17.3% 2 2.0% 0 0.0%	0.0% 9.2% 0.0%	Eff 0.0% 3.1% 13.3%	FicientN 0.0% 0.0% 0.0% 2.0%	0 0 0.0% 0 0.0% 2	0.0% 0.0% 0.0%	17 100% 0.0% 14 64.3% 35.7% 17 76.5% 23.5%				
redicted 80-100 60-80 40-60 20-40	17 17.3% 2 2.0% 0 0.0%	0 0.0% 9 9.2% 0 0.0%	Eff 0,0% 3,1% 13,1% 13,3%	FicientN 0.0% 0.0% 2.0% 8 8.2%	0 0.0% 0 0.0% 2 2.0% 4 4.1%	0 0.0% 0.0% 0.0%	17 100% 0.0% 14 64.3% 35.7% 17 76.5% 23.5% 13 61.5% 38.5%				
Predicted 100-200 80-100 60-80 40-60 20-40	17,3% 2,0% 0,0% 0,0% 0,0%	0.0% 99.2% 0.0% 0.0%	0 .0% 3 .1% 1.3 .1% 1.0% .0%	FicientN 0.0% 0.0% 2.0% 8.2% 1.0%	0 0.0% 0 0.0% 2.0% 4.1% 22.4%	0 0.0% 0.0% 0.0% 0.0% 2.0%	17 100% 0.0% 14 64.3% 35.7% 17 75 23.5% 13 61.5% 38.5% 25 8.50% 12.0%				
Predicted 200-400 100-200 80-100 60-80 40-60 20-40	17 17.3% 2.0% 0.0% 0.0% 0.0%	0.0% 99.2% 0.0% 0.0% 0.0%	0 0.0% 3 3.1% 13.3% 13.3% 0.0% 0.0%	FicientN 0 0.0% 2 2.0% 8 8.2% 1 1.0% 2 2.0%	0 0.0% 0 0.0% 2 2.0% 4 4.1% 222.4% 22.4%	0.0% 0.0% 0.0% 0.0% 2.0% 88.2%	17 100% 0.0% 14 64.3% 23.5% 17 76.5% 23.5% 61.5% 38.5% 25 88.0% 12.0% 25 88.0% 12.0% 25 88.0% 12.0%				
Predicted 200-400 100-200 80-100 60-80 40-60 20-40	177,3% 2,0% 0,0% 0,0% 0,0% 0,0% 0,0% 199 819,5%	0.0% 99.2% 0.0% 0.0% 0.0% 0.0%	Eff 0.0% 3.1% 13.3% 1.0% 0.0% 0.0% 7.65% 23.5%	FicientN 0,0% 0,0% 2,0% 8,2% 1,0% 2,0% 1,0% 2,0% 1,3% 31.5%	let 0,0% 0,0% 2,0% 4,1% 22,4% 2,0% 2,0% 3,3% 2,3%	0.0% 0.0% 0.0% 0.0% 2.0% 8.2% 10 8.0% 20.0%	17 100% 0.0% 14 64.3% 35.7% 76.5% 38.5% 13 61.5% 38.5% 12.0% 12 65 88.0% 12.0% 12 65 88.0% 12.0% 33.3% 98 78.6% 21.4%				

Fig. 3. Confusion Matrix comparing VGG-19 (top) and EfficientNet (bottom).

V. CONCLUSION AND FUTURE WORK

With the growing amount of mixed commercial waste, the concept of Smart Waste Factory was introduced to allow higher sorting and recycling rates. An automatic system (*e.g.*

robotic, big data, networking, *etc.*) is needded to improve waste management. This work estimated the waste particle size, using mixed commercial waste, captured from cameras mounted at a fixed height above the waste. We tested several state-of-the-art deep learning models, *i.e.* ResNet-15, ResNet-18, VGG-16, MobileNet-V2, and EfficientNet, to classify waste into a set of size ranges. EfficientNet performed best on all metrics.

ACKNOWLEDGMENT

This work was supported by ASEAN-European Academic University Network [grant number ICM-2018-11794] and the Austrian Research Promotion Agency (FFG) [project ReWaste 4.0, grant number 860884]. The images used in this paper were captured by Ahmad Almasri when working at the Chair of Automation, Montanuniversität Leoben.

REFERENCES

- S. Kaza, L. C. Yao, P. Bhada-Tata, and F. Van Woerden, *What a Waste* 2.0. Washington, DC, USA: Washington, DC: World Bank, Sep. 2018.
- [2] R. Sarc and K. E. Lorber, "Production, quality and quality assurance of Refuse Derived Fuels (RDFs)," *Waste Management*, vol. 33, no. 9, pp. 1825–1834, Sep. 2013.
- [3] R. Sarc, K. E. Lorber, R. Pomberger, M. Rogetzer, and E. M. Sipple, "Design, quality, and quality assurance of solid recovered fuels for the substitution of fossil feedstock in the cement industry," *Waste Management & Research*, vol. 32, no. 7, pp. 565–585, Jun. 2014.
- [4] R. Sarc, I. Seidler, L. Kandlbauer, K. Lorber, and R. Pomberger, "Design, quality and quality assurance of solid recovered fuels for the substitution of fossil feedstock in the cement industry – Update 2019," *Waste Management & Research*, vol. 37, no. 9, pp. 885–897, Jul. 2019.
- [5] K. Khodier, S. A. Viczek, A. Curtis, A. Aldrian, P. O'Leary, M. Lehner, and R. Sarc, "Sampling and analysis of coarsely shredded mixed commercial waste. Part I: procedure, particle size and sorting analysis," *International Journal of Environmental Science and Technology*, vol. 17, no. 2, pp. 959–972, Feb. 2020.
- [6] R. Sarc, A. Curtis, L. Kandlbauer, K. Khodier, K. E. Lorber, and R. Pomberger, "Digitalisation and intelligent robotics in value chain of circular economy oriented waste management – A review," *Waste Management*, vol. 95, pp. 476–492, Jul. 2019.
- [7] F. Di Maria, F. Bianconi, C. Micale, S. Baglioni, and M. Marionni, "Quality assessment for recycling aggregates from construction and demolition waste: An image-based approach for particle size estimation," *Waste Management*, vol. 48, pp. 344–352, Feb. 2016.
- [8] L. Kandlbauer, K. Khodier, D. Ninevski, and R. Sarc, "Sensor-based particle size determination of shredded mixed commercial waste based on two-dimensional images," *Waste Management*, vol. 120, pp. 784–794, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0956053X2030619X
- [9] Z. Zhang, J. guo Yang, X. Su, and L. hua Ding, "Analysis of large particle sizes using a machine vision system," *Physicochemical Problems* of *Mineral Processing*, 2013.

- [10] G. Dunnu, T. Hilber, and U. Schnell, "Advanced Size Measurements and Aerodynamic Classification of Solid Recovered Fuel Particles," *Energy Fuels*, vol. 20, no. 4, pp. 1685–1690, Jul. 2006.
- [11] J. M. Ponce, A. Aquino, B. Millán, and J. M. Andújar, "Olive-Fruit Mass and Size Estimation Using Image Analysis and Feature Modeling," *Sensors*, vol. 18, no. 9, p. 2930, Sep. 2018.
- [12] L. M. Oo and N. Z. Aung, "A simple and efficient method for automatic strawberry shape and size estimation and classification," *Biosystems Engineering*, vol. 170, pp. 96–107, Jun. 2018.
- [13] M. D. Villiers, "The role and function of a hierarchical classification of quadrilaterals," *For the Learning of Mathematics*, vol. 14, no. 1, pp. 11–18, 1994. [Online]. Available: http://www.jstor.org/stable/40248098
- [14] S. Fujieda, K. Takayama, and T. Hachisuka, "Wavelet Convolutional Neural Networks for Texture Classification," *arXiv*, Jul. 2017. [Online]. Available: https://arxiv.org/abs/1707.07394v1
- [15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, Jun. 2016, pp. 770–778. [Online]. Available: https://doi.ieeecomputersociety.org/10.1109/CVPR.2016.90
- [16] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv*, Sep. 2014. [Online]. Available: https://arxiv.org/abs/1409.1556v6
- [17] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings* of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Jun. 2018, pp. 4510–4520.
- [18] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proceedings of the 36th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 97. PMLR, 09–15 Jun 2019, pp. 6105–6114. [Online]. Available: http://proceedings.mlr.press/v97/tan19a.html
- [19] "Image Composite Editor Microsoft Research," Nov. 2016, [Online; accessed 16. Mar. 2021]. [Online]. Available: https://www.microsoft.com/en-us/research/product/computationalphotography-applications/image-composite-editor
- [20] G. C. Cawley and N. L. Talbot, "On over-fitting in model selection and subsequent selection bias in performance evaluation," *The Journal of Machine Learning Research*, vol. 11, p. 2079–2107, Aug. 2010.
- [21] S. Raschka, "Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning," arXiv, Nov 2018. [Online]. Available: https://arxiv.org/abs/1811.12808v3
- [22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, p. 84–90, May 2017. [Online]. Available: https://doi.org/10.1145/3065386
- [23] T. DeVries and G. W. Taylor, "Improved Regularization of Convolutional Neural Networks with Cutout," *arXiv*, Aug. 2017. [Online]. Available: https://arxiv.org/abs/1708.04552v2
- [24] J. Deng, W. Dong, R. Socher, L. Li, Kai Li, and Li Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2009, pp. 248–255.
- [25] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, "On the importance of initialization and momentum in deep learning," in *Proceedings* of the 30th International Conference on International Conference on Machine Learning - Volume 28, ser. ICML'13. JMLR.org, 2013, p. III-1139-III-1147.