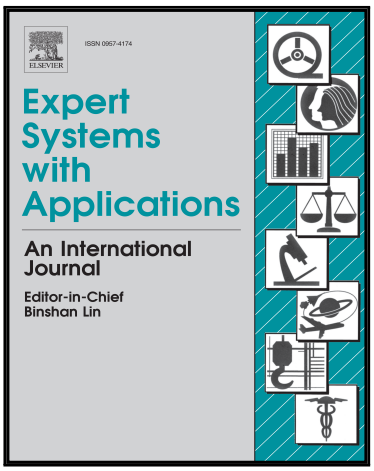
Accepted Manuscript



Deep learning in material recovery: development of method to create training database

Carlos Vrancken , Phil Longhurst , Stuart Wagland

PII: S0957-4174(19)30093-4
DOI: https://doi.org/10.1016/j.eswa.2019.01.077
Reference: ESWA 12478

To appear in: *Expert Systems With Applications*

Received date: 5 September 2018
Revised date: 25 November 2018
Accepted date: 30 January 2019

Please cite this article as: Carlos Vrancken , Phil Longhurst , Stuart Wagland , Deep learning in material recovery: development of method to create training database, *Expert Systems With Applications* (2019), doi: https://doi.org/10.1016/j.eswa.2019.01.077

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Highlights

- Developed method reduces need for material samples
- Capturing images using multiple illuminations has the biggest impact on performance
- Material recognition with deep convolutional network reaches human-level accuracy

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Deep learning in material recovery: development of method to create training database

*Carlos Vrancken, Phil Longhurst, Stuart Wagland**

Carlos Vrancken

School of Water, Energy and Environment, Cranfield University, Cranfield, Bedfordshire, MK43 0AL, UK – E-mail: **c.vrancken@cranfield.ac.uk**

Phil Longhurst

School of Water, Energy and Environment, Cranfield University, Cranfield, Bedfordshire, MK43 0AL, UK – E-mail: **p.j.longhurst@cranfield.ac.uk**

*** Stuart Wagland (Corresponding author)**

School of Water, Energy and Environment, Cranfield University, Cranfield, MK43 0AL, UK; Phone: +44 (0)1234 750111 Ext. 2404; E-mail: s.t.wagland@cranfield.ac.uk;

ACCEPTED MANUSCRIPT

1 ABSTRACT

Increasing the rate of material identification, separation and recovery is a priority in resource management and recovery, and rapid, low cost imaging and interpretation is key. This study uses different combinations of cameras, illuminations and data augmentation techniques to create databases of images to train deep neural networks for the recognition of fibre materials. Using a limited set of 24 material samples sized 1,200 cm^2, it compares the outcome of reducing them to 30 cm^2. The best classification accuracies obtained range from 76.6% to 77.5% indicating it is possible to overcome problems such as limited available materials, time, or storage capabilities, by using a setup with 5 cameras, 5 lights and applying simple software image manipulation techniques. The same method can be used to create deep neural network training databases to recognise a wider range of materials typically found in solid waste streams, in real-time. Furthermore, it offers flexibility as the classification cameras could be deployed at different stages within solid waste processing plants, providing feedback for process control, with the potential of increasing plant efficiency and reducing costs.

2 KEYWORDS

Material recognition; deep neural network; machine learning; waste management; material recovery

3 INTRODUCTION

Priorities to achieve reductions in waste, retaining materials in use within the economy for longer are evident within many government policies and initiatives such as the circular economy (“Four legislative proposals on waste - Think Tank,” 2018; “New EU targets for recycling,”

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

2018; Council of the European Union, 1999; European Parliament., 2009). Targets for collecting separate materials and recycling municipal wastes to reduce landfill reflect this global awareness. Consequently, innovation is needed to advance material identification and separation technologies within current solid waste treatment processes (ISWA, 2016; Sedlak, 2017).

In recent years, many sector studies have used material flow analysis to understand better the volume and composition of wastes (Moriguchi & Hashimoto, 2016; Velis et al., 2013) thus assisting decision making (Turner, Williams, & Kemp, 2016) and improving material recovery and treatment (Al Sabbagh, Velis, Wilson, & Cheeseman, 2012; Allesch & Brunner, 2015, 2017; Arena & Di Gregorio, 2014; Habib, Schibye, Vestbø, Dall, & Wenzel, 2014; Pivnenko, Laner, & Astrup, 2016; Stanisavljevic & Brunner, 2014; Tonini, Dorini, & Astrup, 2014). Knowing the composition of solid waste within process streams in real-time reduces uncertainty (Laner, Rechberger, Feketitsch, & Fellner, 2016; Rechberger, Cencic, & Frühwirth, 2014), eliminating the input-output lag during flow analysis (Pivnenko et al., 2016; Zoboli, Laner, Zessner, & Rechberger, 2016). Integrating material composition sensors at different stages of waste management processes can provide feedback to stakeholders and improve the safety and efficiency of material recovery plants (Vrancken, Longhurst, & Wagland, 2017). This study uses off-the-shelf cameras with visible spectrum sensors, identified previously as the most suitable method amongst existing technologies to measure critical quality indicators for waste treatment. This research advances existing image analysis techniques and applies this to waste streams where it has already been shown to provide good indications of material composition and fuel properties (Peddireddy, Longhurst, & Wagland, 2015; Wagland, Dudley, Naftaly, & Longhurst, 2013; Wagland, Veltre, & Longhurst, 2012).

Applying deep learning techniques (LeCun, Bengio, & Hinton, 2015) to train convolutional neural networks (CNN) (Krizhevsky, Sutskever, & Hinton, 2012; Simonyan & Zisserman, 2014) to recognize objects and materials in images is shown to achieve classification rates close to, or even surpassing, human performance (Russakovsky et al., 2015). Typically, CNN are trained with a database of labelled images, containing many examples of each classification category. Trained CNN are then used to classify materials with data from a separate set of test images.

In general, the accuracy of CNN increases with the number of training images (Bell, Upchurch, Snavely, & Bala, 2015), thus data collection is an important step. Dataset augmentation is regularly used to improve classification accuracy of smaller datasets, applying transformations such as warping, scaling and colour changing. However, exploiting more images in the training dataset has the disadvantage of increasing the time and memory required, resulting in industry concerns (He & Sun, 2015). To address this, techniques such as reusing a previously trained network and parallel computing can help reduce the computational burden whilst improving accuracy (Rawat & Wang, 2017).

Whilst others are investigating and developing network architectures to tackle geometric variance issues (Jaderberg, Simonyan, Zisserman, & kavukcuoglu, 2015; Laptev, Savinov, Buhmann, & Pollefeys, 2016), here we apply a method to generate new training databases using controlled image capture and optimized data augmentation. Classification rates obtained by CNN are negatively affected by large geometric changes in the images, such as rotation (Ciresan et al., 2011; Gong, Wang, Guo, & Lazebnik, 2014; Razavian, Azizpour, Sullivan, & Carlsson, 2014) which would be a common occurrence when receiving material images from streams on a conveyor belt. Comparing the effects of dataset augmentation with different degrees of image rotation and oversampling (Buda, Maki, & Mazurowski, 2017), allows us to improve recognition

ACCEPTED MANUSCRIPT

performance when using images captured in different conditions such as viewing angles and illuminations.

This paper focuses on data capture and augmentation techniques, and it reports how to improve classification rates whilst minimizing the number of training images compared to that of standard datasets (Russakovsky et al., 2015). A series of tests is performed using different methods to acquire and process images then used for training neural networks. Understanding the features that constitute optimal approaches has the potential to lead to robust and efficient learning that recognises materials in recovery facilities, using existing, fast performing deep neural network architectures whilst reducing recognition time and storage requirements.

Whilst the positive effects of training with more images, obtained either through image capture or data augmentation, have been discussed before, this paper sets up a systematic study of image data capture and data augmentation. The experimental results clearly indicate and quantify the effects of increasing the sizes of material samples and using different viewing angles and illuminations, whilst recommending the degrees of image rotation and overlapping to achieve the best possible accuracies. The training datasets are used with AlexNet (Krizhevsky et al., 2012), a well-known deep neural network, to help understand the performance effects of the different experiments. Whilst the proposed framework was designed to identify solid waste material, the experimental framework and results can be transferred to other purposes, using the optimized datasets with different machine learning methodologies and for other applications.

4 MATERIALS AND METHODS

4.1 MATERIAL SAMPLES

The study limited the number of material samples grouping these into 2 main classes: paper and cardboard, with each class containing multiple different items to provide intra-class variance

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

information. Paper samples were represented by white, printed, crumpled, and shredded office paper; brown packaging filler paper; red, white towel, and hygienic tissue paper; glossy magazine paper; folded and crumpled newspaper. Whereas, cardboard samples were represented by plain and printed brown corrugated cardboard; a selection of household food boxes. A total of 24 material samples, including both paper (14 samples, Fig. A2) and cardboard (10 samples, Fig. A3) were used for training, and another 20 material samples were used for testing purposes, consisting of mixed paper (14 samples, Fig. A4) and mixed cardboard (6 samples, Fig. A5).

4.2 IMAGE CAPTURE

The experimental setup used to collect the library of raw images, consisted of 5 Sony DSC-WX500 cameras, labelled C1 to C5, and 5 Gantom GT21 light sources, labelled L1 to L5 (Fig. A1). In spherical coordinates, cameras C1, C2, C3, and C4 were placed at a zenith angle of 30 degrees and azimuth angles of 0, 90, 180 and 270 degrees, respectively, with camera C5 placed at a zenith angle of 0 degrees. All the cameras were mounted pointing towards the centre, with material samples placed, at a focal distance of 1,095 mm. The captured images had a resolution of 4,896 by 3,264 pixels, representing the 43 by 29 cm field of view.

The entire field of view of the cameras, the target area, was covered with photographic green screen material, on top of which the individual samples were placed before being photographed. The green background material was cleaned manually after each material sample was photographed to prevent cross-contamination. During the experiments, each camera was controlled individually to capture images with different viewing angles, using either camera C1, C2, C3, C4 or C5.

The five light sources L1 to L5 were placed next to each of the respective cameras, all set to maximum brightness with a colour temperature of 5,600 K. Each light source illuminated the

target area with a uniform beam diameter of 560 mm. During the experiments, each light was controlled individually to capture images with 6 different illuminations using L1, L2, L3, L4, L5, or a configuration where all the lights were powered, labelled L6.

In the first step of the methodology, shown in Fig. 1, a collection of 30 raw images per sample was captured using all possible combinations of 5 cameras and 6 illuminations. The second step in the method used only the pictures taken by cameras and lights defined by the experimental design. For example, when the experiment considered only 1 camera and 1 light, a single image was selected per material sample, i.e. the one captured using camera C5 and light C5. When the experiment considered all 5 cameras and 6 lights configurations, 30 raw images per material sample were selected.

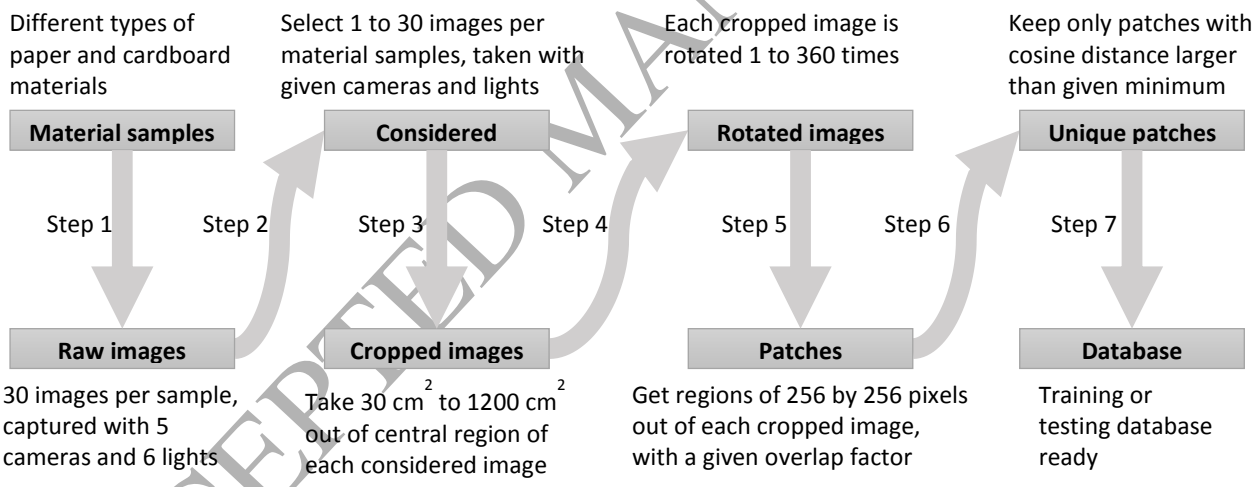


Fig. 1. Steps taken to create the training and testing databases.

4.3 IMAGE PROCESSING

Every image was subjected to a series of transformation, implemented in C++ using OpenCV ("OpenCV library 3.1.0," 2015). The third step cropped images to a smaller size, keeping the central region and discarding the outside. Fourthly, images were rotated, which was implemented

using Lanczos resampling over an 8 by 8 pixels neighbourhood. This allowed the rotation of cropped images with respect to the centre at different angles.

Each cropped and rotated image was then split into patches (256 by 256 pixels), with varying ranges of overlap between patches. A hue, saturation, and value (HSV) colour detection routine was applied to each patch to ensure that only the patches without the green background were kept. All these processes were implemented to run on a single CPU except for the patch extraction step which was multi-threaded and ran on 4 parallel processes. The time taken in each step was measured and recorded for comparison purposes.

4.4 DEEP NEURAL NETWORK

AlexNet (Krizhevsky et al., 2012) was the convolutional neural network architecture used. This had previously been fine-tuned with the MINC database (Bell et al., 2015), with the output layer fc8 modified to predict 2 types of materials. The framework used to implement, train and test the neural networks was Caffe version 1.00 (Jia et al., 2014) running in parallel on four nVidia K80 GPUs. The initial learning rate was set at 10^{-5} and reduced by a factor of 10 for every 2,000 iterations. To reduce overfitting, where the network would perform well on the training data but fail on the test patches, the dropout factor in AlexNet was set to 0.5.

The sixth method step tested each patch by applying the previously trained MINC neural network. The considered output was the fc7 network layer, a 4096-dimensional vector, which was used to calculate the cosine distance to each of the other patches (Bell & Bala, 2015). The cosine distance between two similar images is smaller than the calculated between two unique images. This step ensures the datasets contain only unique patches by requiring a minimum cosine distance value between them. Additionally, each time a patch was fed into the network for training, it was cropped to 227 by 227 pixels from a random region. This reduces overfitting as it

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

avoids duplicate images which would force the network to memorise data instead of learning features.

In the final step, the unique patches were stored into databases for training or testing purposes. A training database was generated for each experiment. Each training step was performed for a minimum of 4,000 iterations or 20 epochs, which meant that each patch in the training database was used at least 20 times. To prevent overfitting, the trained networks were tested every 4 epochs and the training step was considered complete when the accuracy testing stopped improving.

4.5 ACCURACY TESTING

A single testing database was generated from 20 samples of paper and cardboard materials that had not been included in the training database. The paper class, see Fig. A4, included white paper (4 samples containing a mix of blank, printed, crumpled, and shredded material), brown paper, tissue paper (3 samples containing a mix of white and red materials), magazine paper (2 samples of different pages), and newspaper (4 samples of different pages, folded and crumpled). The cardboard class (Fig. A5) included corrugated cardboard, and household food cardboard (5 samples with mixed materials).

The test patches were obtained from random 227 by 227 pixels regions within the raw images, equivalent to 2 by 2 cm, with no overlapping. Only those patches with a cosine distance of 0.2 or more to the other patches in the database were kept. The two material classes were represented in equal number in the testing database, with a total of 18,496 patches. Network accuracy tests were performed on this single testing database, and every testing patch was assigned a probability of being either paper or cardboard by the trained network. The class with the highest probability

ACCEPTED MANUSCRIPT

was interpreted as the predicted material for each patch in the testing database, which was then used to calculate the mean classification accuracy of the trained network.

4.6 DATA ANALYSIS

For each test, the predicted material for every patch in the testing database was compared to the actual material class, and mean accuracy was calculated and expressed as the percentage of correct predictions. Each experiment was performed three times, and the standard deviation of the mean accuracy was calculated. When running the experiments on different crop sizes, a combined value was obtained by averaging the mean accuracies of each crop size studied. Significant differences between results were determined with 1-way ANOVA followed by Dunnett's tests, and 2-way ANOVA followed by Tukey's tests, using 95% confidence intervals.

4.7 EXPERIMENTAL DESIGN

A series of experiments to evaluate the effects of the illumination and viewing angles configurations, the size of the material samples, software image rotation, patch overlap, and the individuality and the number of training patches were designed. These are shown in Table A1 as a summary of experimental setups and the different configurations tested. The experiments consisted of creating a training database of patches using their specific configuration, training a deep neural network with the created databases, and finally applying the trained networks to the testing database to obtain the mean accuracies. Each experiment was performed three times to estimate the uncertainty of the results.

Experiment 1 compares the accuracy obtained with different minimum cosine distances (0.1, 0.2, 0.3, 0.4, and 0.5) between the training patches. Experiment 2 compares the effects of cropping images to the equivalent physical size of 30, 120, 240, 720 and 1,200 cm^2. Then experiment 3 uses software techniques to increase the number of images, rotating the images by

0, 1, 3, 5, 10, 15, 30, and 90 degrees. Experiment 4 overlaps the extracted patches by 0, 12.5, 25, 37.5, 50, 62.5, 75, and 87.5 percent. Experiments 5 use different hardware configurations to use 1 to 6 illuminations, and experiment 6 uses 1 to 5 cameras.

Finally, experiments 7, 8 and 9 use combinations of techniques to increase the number of training images. Experiment 7 combines the software techniques, rotation and overlapping, using the best results obtained in experiments 3 and 4. Experiment 8 combines hardware techniques, using multiple cameras and illuminations, using the results obtained in experiments 5 and 6. Finally, experiment 9 uses a combination of both software and hardware techniques, including rotation, overlapping, multiple cameras and multiple illumination configurations. All the experiments are performed on 30 cm^2 and 1,200 cm^2 cropped image sizes for comparisons.

5 RESULTS AND DISCUSSION

5.1 EXPERIMENT 1: PATCH UNIQUENESS

Using different cosine distances between the training patches shows no significant effect on the resulting accuracies for either 30 cm^2 or 1,200 cm^2 crop sizes (Fig. A6). When averaging the results from both crop sizes, the highest accuracy was obtained with a minimum cosine distance of 0.2, which was therefore used in the following experiments.

The effects of different minimum cosine distance values on the classification rates are more pronounced when working with 30 cm^2 images than with the larger 1,200 cm^2 crops. When using small crops, a cosine distance of 0.2 increases the average classification by 2% or more compared to using a cosine distance of 0.1. When using larger crops, the cosine distance has a smaller effect on classification rates, as the best cosine distance value is 0.1. This is supported by results from other studies (Bell et al., 2015).

ACCEPTED MANUSCRIPT

5.2 EXPERIMENT 2: CROP SIZES

Increasing the number of patches by using larger crop sizes results in higher mean classification accuracies as expected (Table 1). The largest gain was obtained when increasing the crop size from 30 cm^2 to 120 cm^2 for a relative increment of 15% of the classification rate. Further increases in size results in diminishing gains and no statistically significant differences.

Table 1. Baseline results show the effects on mean classification accuracy of using different crop sizes to generate the training patches.

Crop size (cm^2)	30	120	240	720	1,200
Number of training patches	34	178	389	1,052	1,731
Mean accuracy	61.9%	71.0%	72.5%	73.6%	74.5%
Standard deviation	0.1%	1.0%	0.3%	0.3%	0.9%

As expected, the performance of the neural network improved with more data, although the full potential of the configuration may yet to be reached. Taking pictures of more material samples increases the number of training patches and this is likely to produce higher accuracy results than obtained in this experiment. The results of this experiment were used as a baseline for comparison with the classification accuracies obtained in the following experiments.

5.3 EXPERIMENT 3: SOFTWARE ROTATION

Using software rotation to increase the number of patches of 30 cm^2 crops results in a significant difference only when comparing the baseline result. This uses only the original image with 4 (original plus 3 generated) images per sample. In the case of 1,200 cm^2 crop sizes, no significant differences are found from the number of rotations, but a maximum mean accuracy of

ACCEPTED MANUSCRIPT

76.2% was achieved with 12 rotations, using 10,154 patches. Averaging the results from 30 cm^2 and 1,200 cm^2 crop sizes, gives the highest accuracy with 12 rotations.

Gains were larger when working with the smaller, 30 cm^2 crops, with a relative increment of mean accuracy of 6%, whilst the benefit reduced to 4% when working with the larger, 1,200 cm^2 crop sizes. The effect of software rotations shows that this technique is beneficial. However, there are no gains by rotating an image more than 12 times, at 30 degrees each step. Furthermore, whilst the mean accuracy of 68.5% obtained using 68 patches from 4 rotations (Fig. A7) of the 30 cm^2 crop sized images is lower, it is not significantly different to the baseline 71% obtained using 120 cm^2 crop size (Table 1).

5.4 EXPERIMENT 4: PATCH OVERLAP

When working with 30 cm^2 crop sized images, the results show significant differences between the 0% and 12.5% overlap, as well as using the 0% and 25% overlap. The maximum accuracy of 68.6% obtained using a 25% overlap (Fig. A8) from 73 patches, represents a relative gain of 10% compared to not using patch overlapping. This is not significantly lower than the baseline result which uses 240 cm^2 crop sizes (Table 1). However, when using 1,200 cm^2 crop sizes there are no significant differences between any of the different overlapping factors, with the maximum mean accuracy obtained without overlapping. When averaging results from the 30 cm^2 and 1,200 cm^2 crop sizes, the highest accuracy was with 25% overlapping.

Even though more training patches could be generated by increasing the overlapping factor, the classification rates obtained are no different or lower than when using 25% overlap, for both crop sizes in the experiments. This means that the extra patches generated by overlapping the training images contain no extra information that the network can use for training. The same effect is observed in experiment 3, where rotating more than 12 times is not beneficial.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

5.5 EXPERIMENT 5: NUMBER OF LIGHTS

When working with 30 cm^2 crop sized images, increasing the number of different illuminations of the target area from 1 to 4, 5 or 6 achieves significant benefit. Using 5 different illuminations (Fig. A9) gives the maximum mean accuracy of 69.6% with 174 training patches. This represents a relative improvement of 10% from using just 1 illumination. Whilst this result is lower than the baseline 71% accuracy obtained using 120 cm^2 crop size with no augmentation techniques (Table 1) it is not significantly different.

In the case of 1,200 cm^2 crop sized images, increasing the number of different illuminations from 1 to any other number results in a significant improvement, with a maximum accuracy of 76.9% when using 6 illuminations and 11,863 patches. This represents a relative gain of 5% compared to using 1 illumination. Using 2 illuminations results in the same mean accuracy of 76.9% but with larger variability and a standard deviation of 0.9% compared to 0.4% with 6 illuminations.

Averaging the results from using both crops sizes, shows that the highest accuracy can be obtained from using 5 illuminations. Even though more training patches are generated using 6 illuminations, the classification rates obtained are lower and not significantly different when using 5. This suggests that the extra patches generated from images taken with a sixth illumination contain no extra information that the network can use for training.

5.6 EXPERIMENT 6: NUMBER OF CAMERAS

Using the 30 cm^2 crop sized images there are significant differences when taken from multiple cameras as variation occurs with the number of cameras. This is notable except when comparing the results between 3 and 5 cameras. The maximum accuracy of 71.2% (Fig. A10) from 4 cameras and 150 training patches, represents a relative improvement of 14% to that of just 1

ACCEPTED MANUSCRIPT

camera. The resulting mean accuracy is not significantly different to the baseline results using 240 cm^2 crop sizes (Table 1).

Using 1,200 cm^2 crop sizes and increasing the number of cameras from 1 to 5 results in the only significant difference This achieves a maximum accuracy of 75.5% from 9,732 patches, with a relative gain of 3% compared to using just 1 camera. Averaging the results from 30 cm^2 and 1,200 cm^2 crop sizes, shows the highest accuracy is again obtained with 4 cameras. Even though more training patches are generated with 5 cameras, the classification rates are not significantly different for either crop size in the experiments. This means that the extra patches generated by considering images taken with a fifth camera contain no additional information that can be used for network training.

5.7 EXPERIMENT 7: COMBINED SOFTWARE TECHNIQUES

Combining the best result from Section 4.3 of increasing the overlapping factor to 25%, and the best result from Section 4.4, plus applying the software techniques of rotating the 30 cm^2 crop sized images by 30 degrees; this leads to a mean accuracy of 70.4%, when using 199 training patches (Fig. 2). This achieves a significantly better result than the baseline accuracy of 61.9% obtained using only 1 rotation and no overlapping (Table 1). It is also significantly better than the 67.9% accuracy obtained using 12 rotations and no overlapping (Fig. A7). However, it is not significantly different to the baseline results when using 120 cm^2 crop sizes with no augmentation techniques (Table 1), or to that of the mean accuracy of 68.6% obtained using only 1 rotation and 25% overlap (Fig. A8). This indicates that overlapping the patches by 25% is more beneficial than rotating the images by 30 degrees.

When combinations of rotations by 30 degrees and 25% overlapping is applied to the 1,200 cm^2 crop size images, 15,022 training patches were extracted. The mean accuracy obtained was

75.8% (Fig. 3), which is higher but not significantly different to the baseline result of 74.5% obtained when using no rotation or overlapping (Table 1). This represents a relative gain in accuracy of 2%, i.e. lower than the gain of 14% on obtained the 30 cm^2 crops. However, this difference is not enough to compensate for the difference in crop sizes, as the result from 1,200 cm^2 crop sized images is significantly better than when applying the pair of software techniques to 30 cm^2 crops.

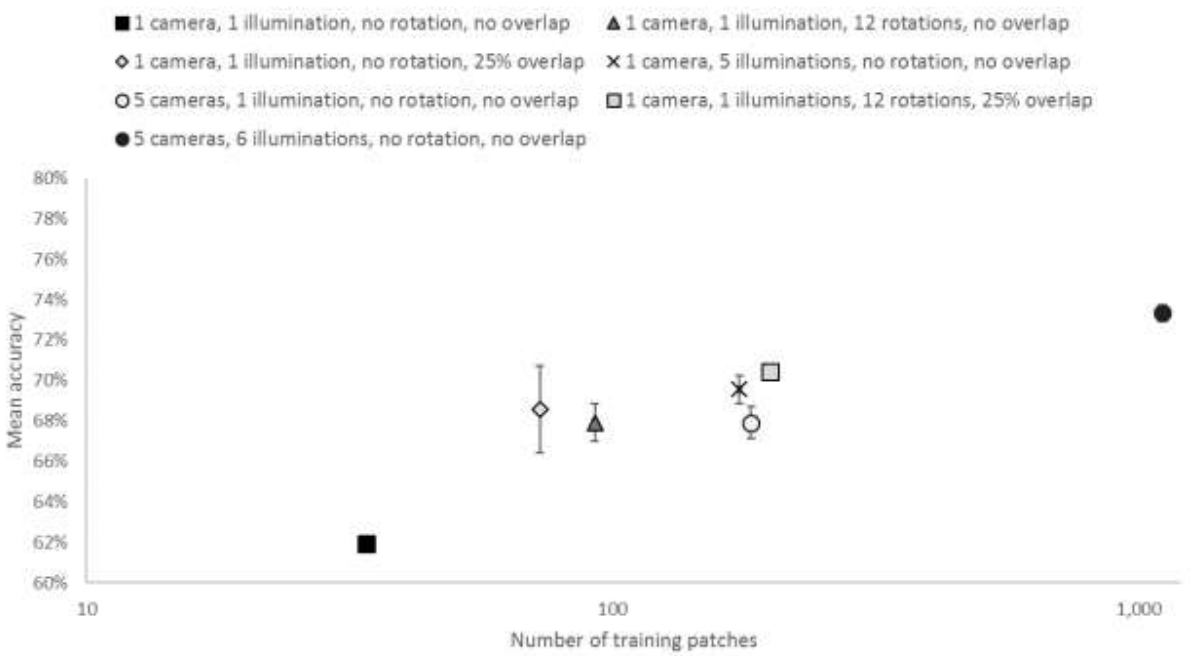


Fig. 2. Effects of different techniques to increase the number of training patches and improve accuracy when using 30 cm^2 crops, on a log-10 scale. Vertical bars indicate standard deviation.

5.8 EXPERIMENT 8: COMBINED HARDWARE TECHNIQUES

To understand the extent to which combining hardware techniques can address the effects of a limited data, this experiment was repeated twice. Firstly, 20 images were captured using a combination of 4 cameras and 5 illuminations with the best results from Sections 4.6 and 4.5, respectively. Secondly, 30 images were captured with all 5 cameras and 6 different illuminations.

ACCEPTED MANUSCRIPT

Using 4 cameras and 5 illuminations resulted in a mean accuracy of 73.2%, using 753 patches from 30 cm^2 crops, and in 75.8% using 40,042 patches from 1,200 cm^2 crop sized images. The experiment with 5 cameras and 6 lights, using 30 cm^2 crop sizes, resulted in a mean accuracy of 73.3% by using 1,106 patches (Fig. 2); and when using 1,200 cm^2 crop sizes, 58,642 training patches were extracted to achieve 76.2% mean accuracy (Fig. 3).

In contrast to experiments 5 and 6, where adding a single camera or a single light was neither significant nor detrimental to the mean accuracy, the combination of an extra camera with an extra light provided useful information for training the neural network. The results from using the combination of 5 cameras and 6 lights were better than using 4 cameras and 5 lights, though the differences were not statistically significant.

The results obtained using 30 cm^2 crop sized images taken with the combination of 5 cameras and 6 lights were significantly better than using 5 cameras and 1 light (Fig. A10), or 1 camera and 6 lights (Fig. A9). However, there was not significant difference to using 4 cameras and 1 light, or 1 camera and 5 lights. These results indicate that both cameras and lights are equally important when working with a limited data set, and that a minimum of 4 cameras and 5 lights is required to get the most benefit from the proposed setup.

The mean accuracy of 73.3% obtained using the combination of 5 cameras and 6 lights on the 30 cm^2 crop sized images was lower but not significantly different than the baseline mean accuracy of 74.5% obtained from 1,200 cm^2 crop size images with no augmentation techniques (Table 1). This means that combining hardware techniques overcomes the issues of having a limited data set for CNN training.

When working with 1,200 cm^2 crop sizes, 5 cameras and 6 lights, the mean accuracy obtained was significantly better than using 5 cameras and 1 illumination (Fig. 3). Whilst it was lower, it

was not significantly different from using 1 camera and 5 illuminations. This suggests that when more material samples are available, additional illumination is more important than adding more cameras.

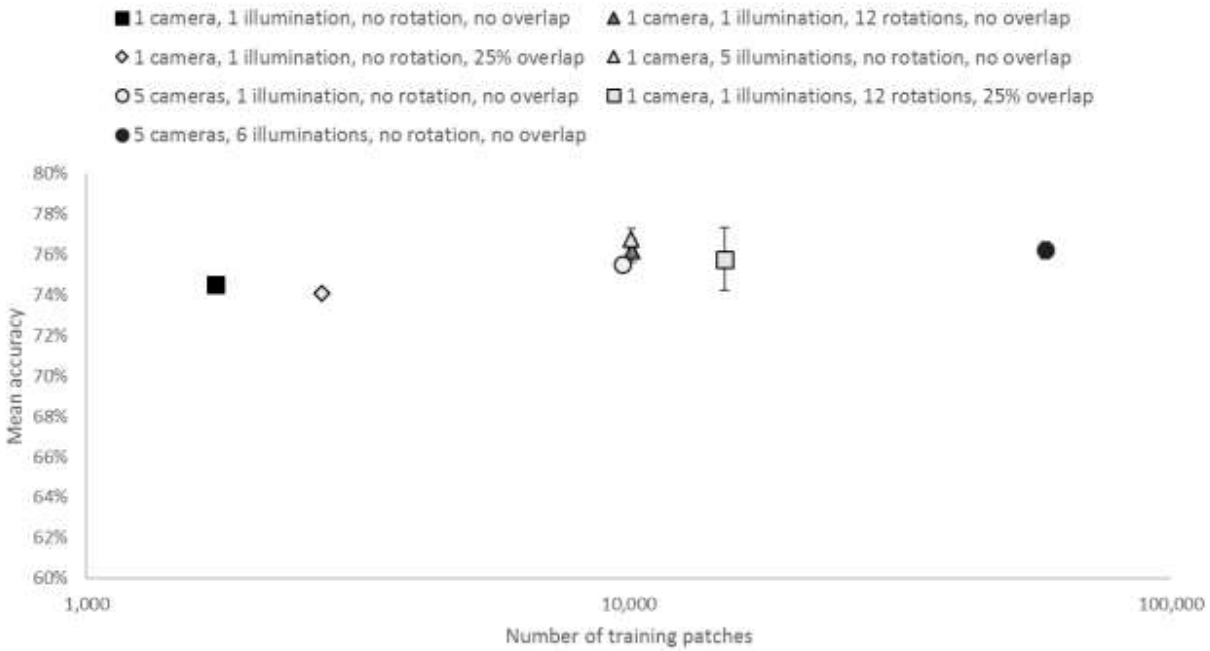


Fig. 3. Effects of different techniques to increase the number of training patches and improve accuracy when using 1,200 cm^2 crops, on a log-10 scale. Vertical bars indicate standard deviation.

5.9 EXPERIMENT 9: COMBINED SOFTWARE AND HARDWARE TECHNIQUES

The combination of using 5 cameras, 6 illuminations, 12 rotations, and a 25% patch overlap on 30 cm^2 crop sized images results in a mean accuracy of 76.6% from 8,309 patches. This result is significantly better than either using a combination of 12 rotations and 25% overlap, or a combination of 5 cameras and 6 illuminations (Fig. 4). This indicates that both pairs of techniques are similarly important in improving accuracy if working with a limited data set.

When the combination was applied to 1,200 cm^2 crop sized images, the accuracy achieved was 77.5% from 561,070 patches, which was significantly different to only using 12 rotations and

ACCEPTED

25% overlap, but not to using only 5 cameras and 6 lights. The implication is that when more data is available, this pair of hardware techniques is more beneficial to the resulting accuracy than the combination of software techniques.

Using this combination of software and hardware techniques on 30 cm^2 crop sized images was better, but not significantly different, than the baseline accuracy of 74.5% obtained when using 1,200 cm^2 crop sized images with no augmentation. Furthermore, the mean accuracy of 76.6% obtained on 30 cm^2 crop sizes was lower but not significantly different than the 77.5% obtained when the same techniques were applied to 1,200 cm^2 crop sizes. This indicates the full potential of the method, as the combination of hardware and software techniques compensates for the lack of raw material and a reduced data set.

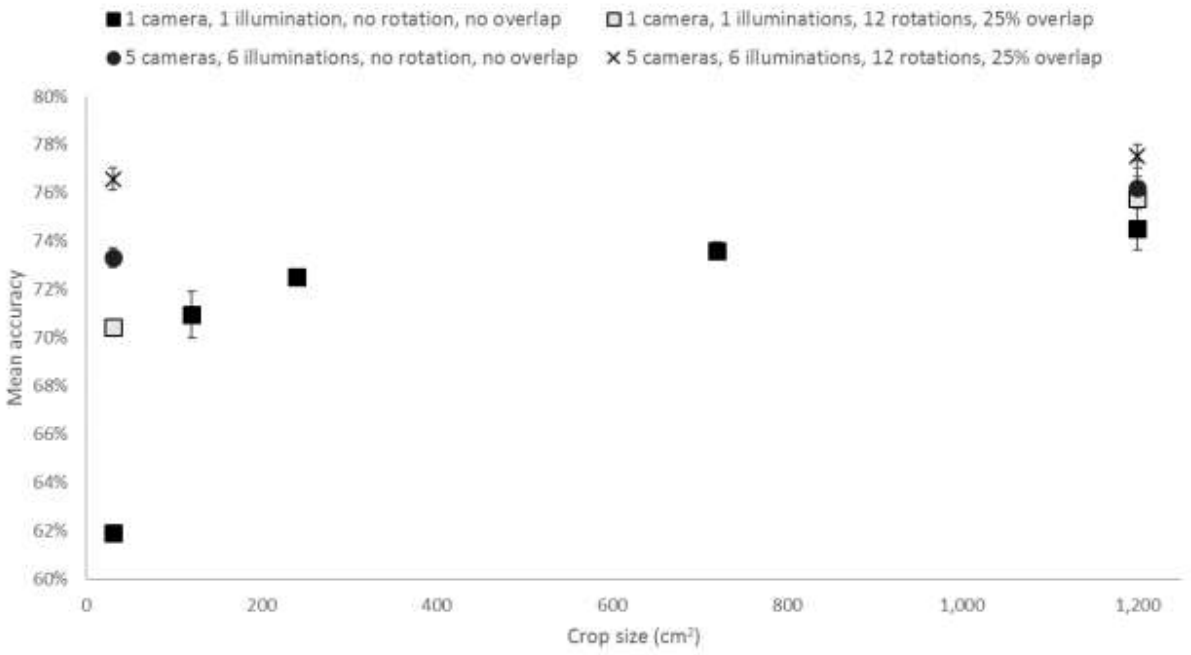


Fig. 4. Combined effects of different techniques when applied to different crop sizes. Vertical bars indicate standard deviation.

ACCEPTED MANUSCRIPT

5.10 TIME AND STORAGE REQUIREMENTS

The configuration used in each experiment affected the classification accuracy, as well as the size of the database file and the time taken to complete (Table 2). The time to capture the 30 pictures of each sample with all the possible combinations of 5 cameras and 6 illuminations took an average of 7.5 minutes. This translates to an average of 15 seconds per capture, including the time to configure the cameras, the lights and the wireless transfer for further computer processing.

The average computing time to rotate a 30 cm^2 crop size image was 0.003 seconds, whilst rotating a 1,200 cm^2 crop size image takes 0.147 seconds. The time required to extract the patches depended on the source material, as an image with greater variability generates more unique patches, which then requires more unique comparisons and file creations. The time to create a training database file depends on the number of unique patches extracted, with an average of 0.025 seconds per patch. Finally, the training step for each experiment took 28 minutes, except for experiment 9 using 1,200 crop size images, which took 4 hours 20 minutes to complete due to the larger number of training patches.

The rotation and patch overlapping techniques had a relatively small impact on the total time, adding a maximum of 4 minutes when working with 30 cm^2 crop sized images captured with all the cameras and illuminations (experiment 9 vs experiment 8). The best result using 30 cm^2 crop size images obtained with the combination of software and hardware techniques (experiment 9), took 3 hours and 5 minutes longer than when using no augmentation techniques (experiment 2), whilst improving the mean accuracy by 24% requires 185 MB additional database storage.

Table 2. Best configuration and mean accuracy obtained with each experiment, showing the average time and file space required, for both 30 cm^2 and 1,200 cm^2 crop sized images.

Experiment	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Number of cameras	1	1	1	1	5	1	5	5
Number of illuminations	1	1	1	5	1	1	6	6
Number of rotations*	1	12	1	1	1	12	1	12
Patch overlap (%)	0	0	25	0	0	25	0	25
30 cm^2 crop sized images								
Mean accuracy	61.9%	67.9%	68.6%	69.6%	67.9%	70.4%	73.3%	76.6%
Average total time (min)	34.3	34.3	34.3	58.5	58.5	34.4	215.1	219.0
Average database size (MB)	7	17	14	33	35	38	211	1,630
1,200 cm^2 crop sized images								
Mean accuracy	74.5%	76.2%	74.1%	76.8%	75.5%	75.8%	76.2%	77.5%
Average total time (min)	40.5	45.5	40.9	85.6	85.5	52.6	544.5	1035.1
Average database size (MB)	331	1,945	520	1,936	1,860	2,877	11,232	108,116

*: The number of rotations includes the original image.

5.11 ACCURACY VS CONFIGURATION AND MATERIAL SIZE

Comparing the individual effects in classification accuracy for each of the 5 main variables studied in the experiments, the single most beneficial change was to increase the size of the images from 30 cm^2 to 1,200 cm^2, resulting in a relative gain of 20.3% (Table 2, experiment 2). When working with 30 cm^2 crops, the largest relative gain obtained was 12% by using multiple illuminations. However, combining the two hardware techniques resulted in a relative accuracy increase of 18%, i.e. not significantly lower than the 20.3% improvement obtained by increasing the crop size to 1,200 cm^2.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Furthermore, combining the hardware technique proved more beneficial than the pair of software techniques, which only managed to improve the results by a relative 13%. On the other hand, the combination of both hardware and software techniques with 30 cm^2 crops resulted in a relative increment of 24% to reach a mean accuracy of 76.6%. This result was the best one obtained with the 30 cm^2 crop sized images, and significantly better than only increasing the crop size to 1,200 cm^2.

The effects of any of the applied techniques were less pronounced when working with the larger, 1,200 cm^2 crop sized images. The largest impact was again obtained using multiple illuminations, improving the results by a relative 1.5%. This higher result was not significantly better than applying either the two hardware techniques (relative gain of 1.1%), or the two software techniques (relative gain of 0.8%). The combination of hardware and software techniques resulted in a relative increment of 2% in accuracy.

The maximum benefit obtained in all the experiments was by increasing the size of the materials from 30 cm^2 to 1,200 cm^2 and applying a combination of the four techniques. This resulted in a relative increase of 25.2% in classification accuracy. The trade-offs were the total time taken of over 17 hours more than the baseline, and the storage requirements which increased from the 7 MB baseline to 108,116 MB.

The mean classification accuracy of 77.5% reached levels of performance typically obtained with databases containing a similar number of training images (Bell et al., 2015; Buda et al., 2017). Furthermore, when working with the 30 cm^2 and the combination of four techniques, the maximum accuracy of 76.6% was lower but it was obtained with a much smaller training database, exceeding the expected results. Using the proposed data augmentation techniques

ACCEPTED MANUSCRIPT

reduces the relative difference in accuracy between using 30 cm^2 and 1,200 cm^2 from 20% to 1%, in both cases reaching human-level accuracy (Wiebel, Valsecchi, & Gegenfurtner, 2013).

6 CONCLUSIONS

This study shows that even if the available time, storage capability or material samples are very limited, the method extracts as much as information for training and increases the neural network performance to almost the same level expected from a much larger dataset. As opposed to other applications where context data, such as the presence of other objects or materials, orientation, and size can be used to identify a material or an object, in this study the deep neural network is successfully trained only with the characteristics of material surfaces as captured with a camera enhanced with multiple illuminations.

The proposed image capture and data augmentation methods could be applied to generate training databases for the recognition of more material classes, such as plastics, textiles, glass, metal, and minerals, or even to objects, for example for the recognition of hazardous items such as batteries and gas canisters. Whereas this study used two categories with wide intra-class variability, the classes could be split so that, for example, white paper, old newspaper, tissue, and magazine would be classified as different types of materials, providing useful information to recycling industries.

ACCEPTED MANUSCRIPT

7 APPENDIX

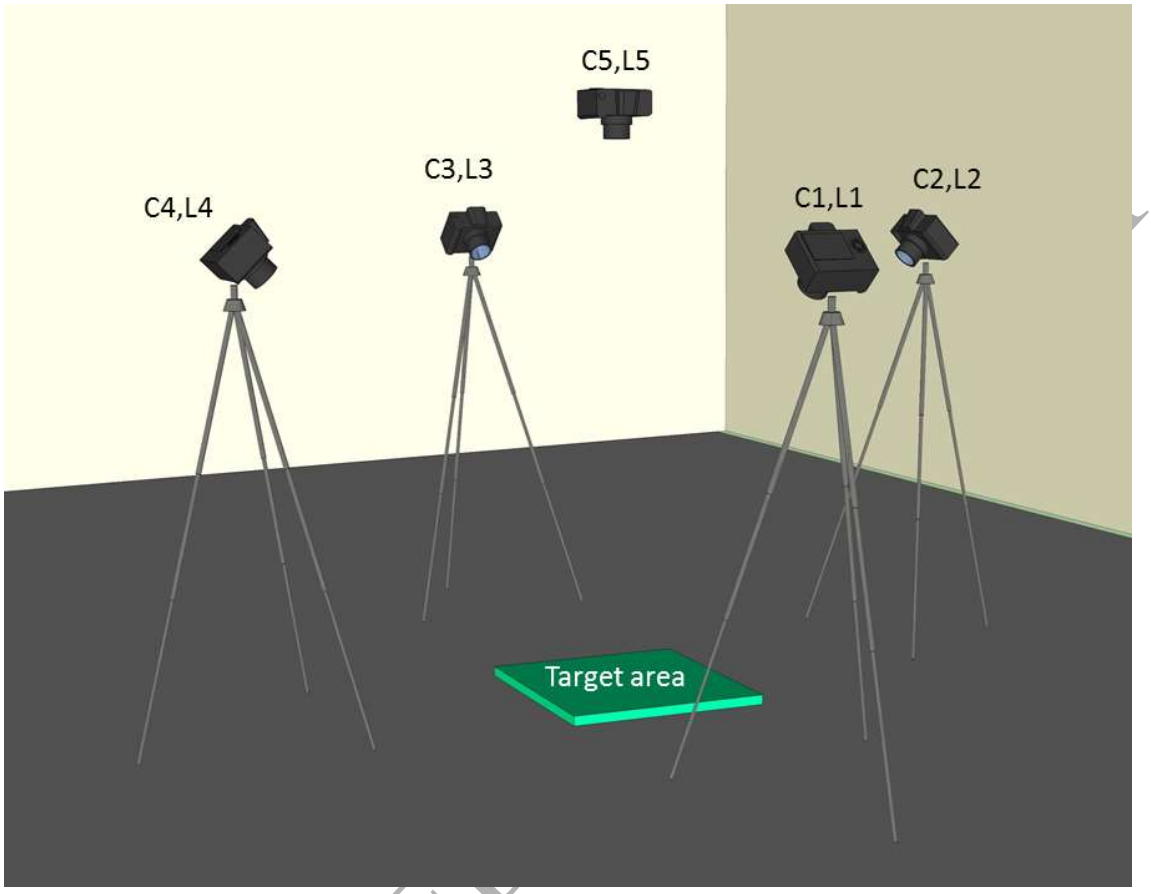


Fig. A1. Schematics depicting the image capture hardware configuration which includes 5 cameras (C1 – C5) and 5 lights (L1 – L5). The lights (not shown) are located next to the cameras, also pointing towards the target area.

ACCEPTED

Fig. A2. Paper samples used for training, illuminated with light L5 and photographed with camera C5.

Fig. A3. Cardboard samples used for training, illuminated with light L5 and photographed with camera C5.

ACCEPTED MANUSCRIPT

Fig. A4. Paper samples used for testing, illuminated with light L5 and photographed with camera C5.

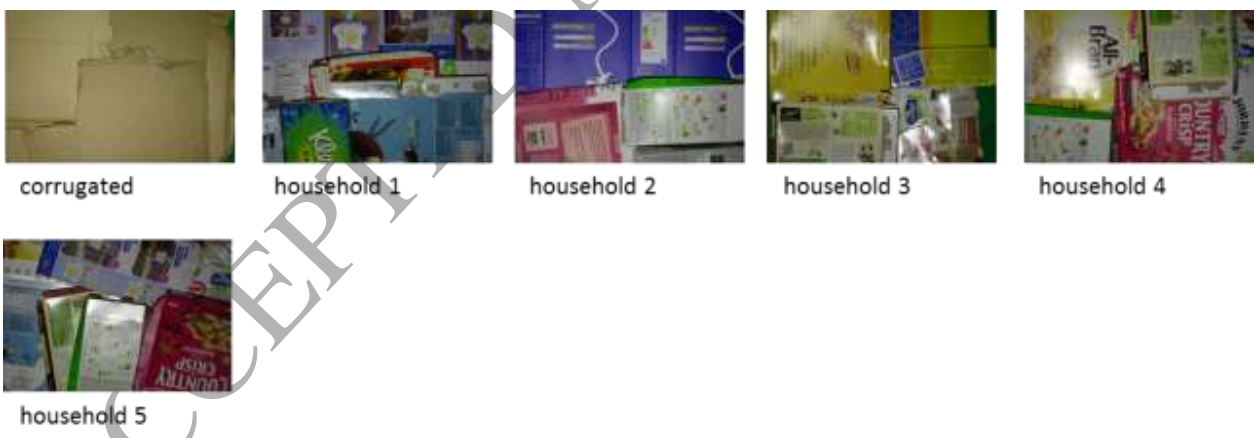


Fig. A5. Cardboard samples used for testing, illuminated with light L5 and photographed with camera C5.

ACCEPTED MANUSCRIPT

Table A1. Summary of experimental setups and the different configurations tested.

Experiment	Cosine distance	Cropped size (cm^2)	Number of rotations (including original)	Overlapping (%)	Lights	Cameras
1	0.1, 0.2, 0.3, 0.4, 0.5	30, 1200	1	0	L5	C5
2	0.2*1	30, 120, 240, 720, 1200	1	0	L5	C5
3	0.2*1	30, 1200	1, 4, 12, 24, 36, 72, 120, 360	0	L5	C5
4	0.2*1	30, 1200	1	0, 12.5, 25, 37.5, 50, 62.5, 75, 87.5	L5	C5
5	0.2*1	30, 1200	1	0	L5, L5+L1, L5+L1+L2, L5+L1+L2+L3, L5+L1+L2+L3+L4, L5+L1+L2+L3+L4+L6	C5
6	0.2*1	30, 1200	1	0	L5	C5, C5+C1, C5+C1+C2, C5+C1+C2+C3, C5+C1+C2+C3+C4
7	0.2*1	30, 1200	12*3	25*4	L5	C5
8	0.2*1	30, 1200	1	0	L1, L2, L3, L4, L5, L6*5	C1, C2, C3, C4, C5*6
9	0.2*1	30, 1200	12*3	25*4	L1, L2, L3, L4, L5, L6*5	C1, C2, C3, C4, C5*6

Note: *n indicates value(s) based on the results from experiment n.

ACCEPTED MANUSCRIPT

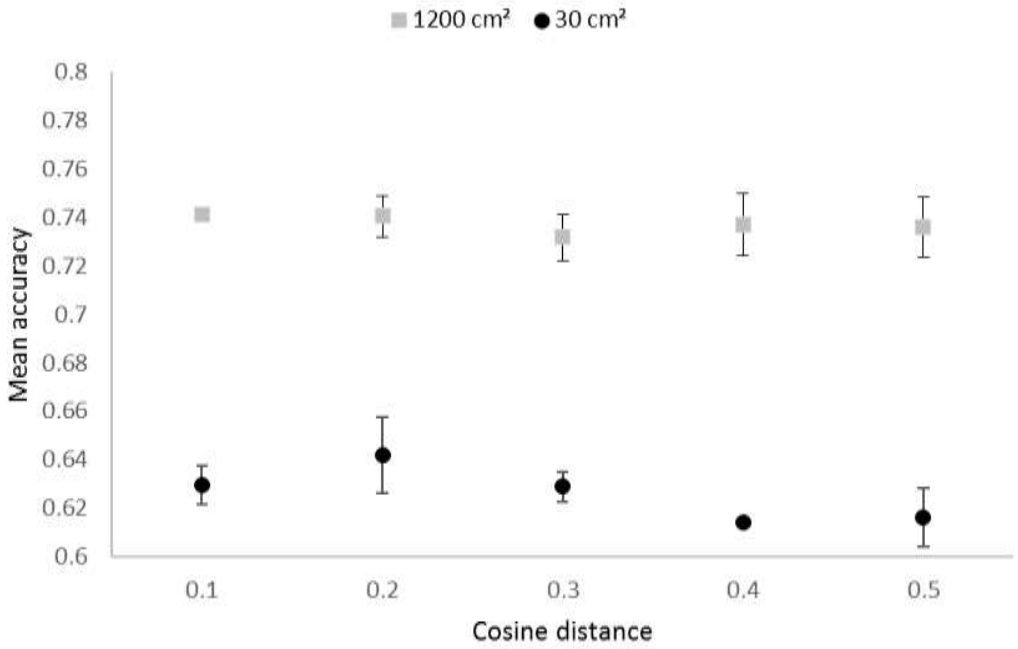


Fig. A6. Effects of using different cosine distances between training images, for different crop sizes. The vertical bars denote standard deviation.

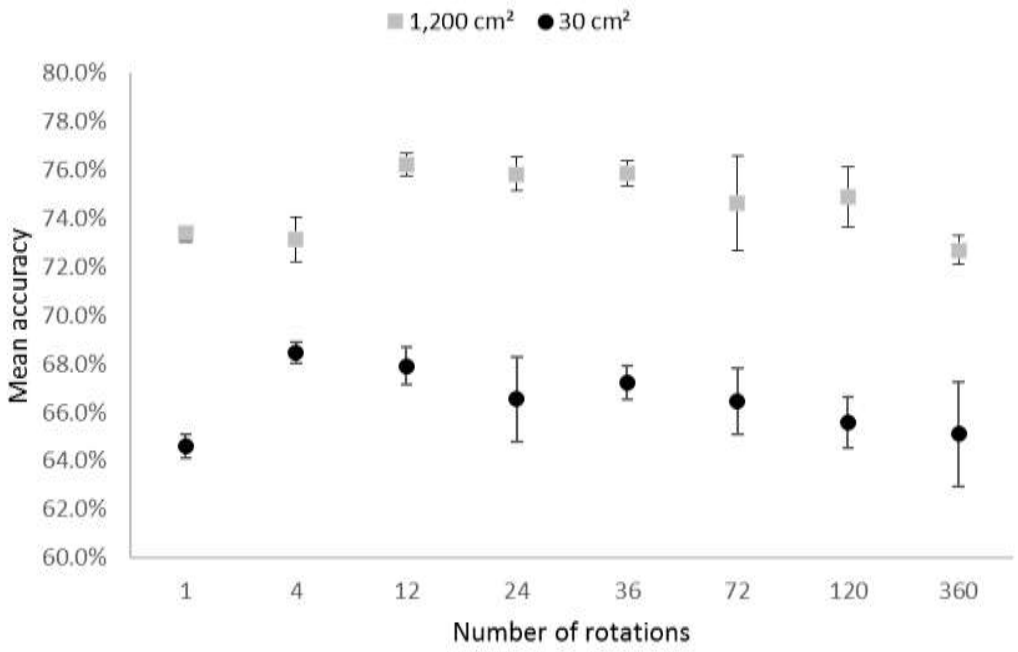


Fig. A7. Effects of using software rotations to increase the number of training images, for different crop sizes. The vertical bars denote standard deviation.

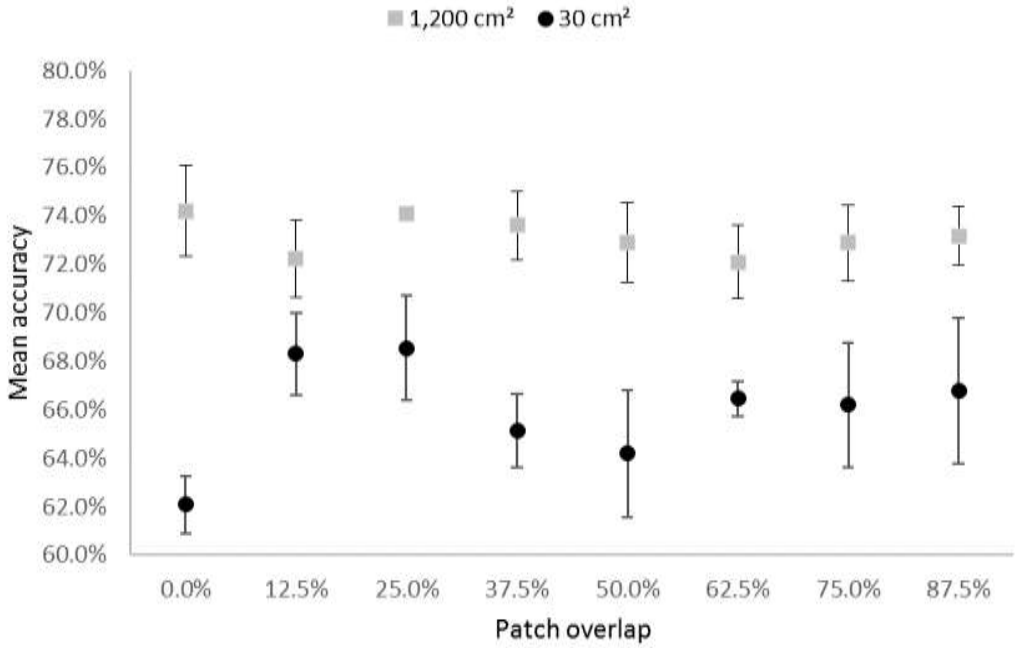


Fig. A8. Effects of using different patch overlapping factors to increase the number of training images, for different crop sizes. The vertical bars denote standard deviation.

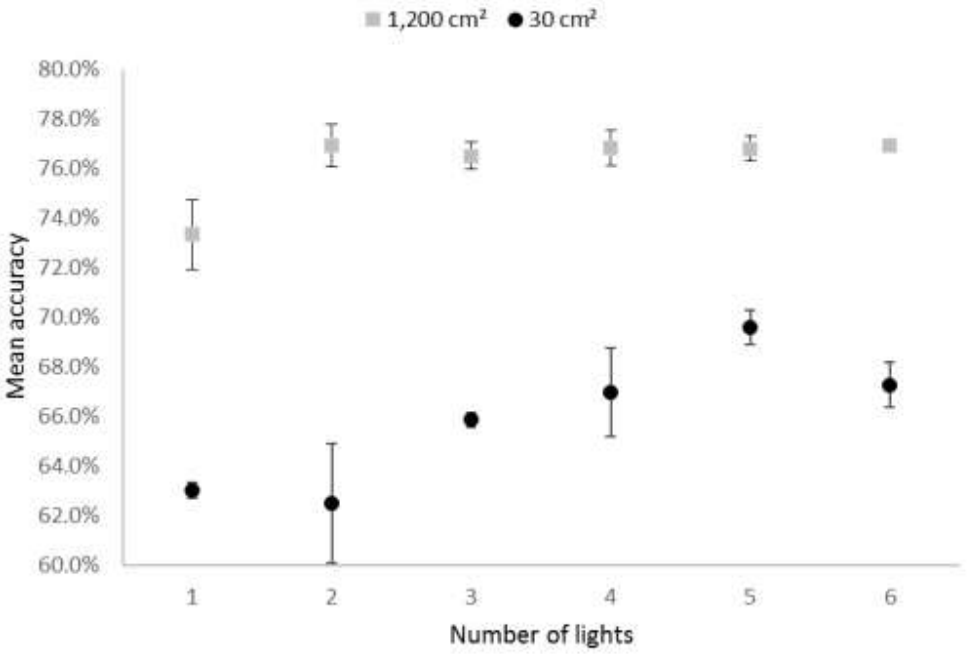


Fig. A9. Effects of using different number of lights to increase the number of training images, for different crop sizes. The vertical bars denote standard deviation.

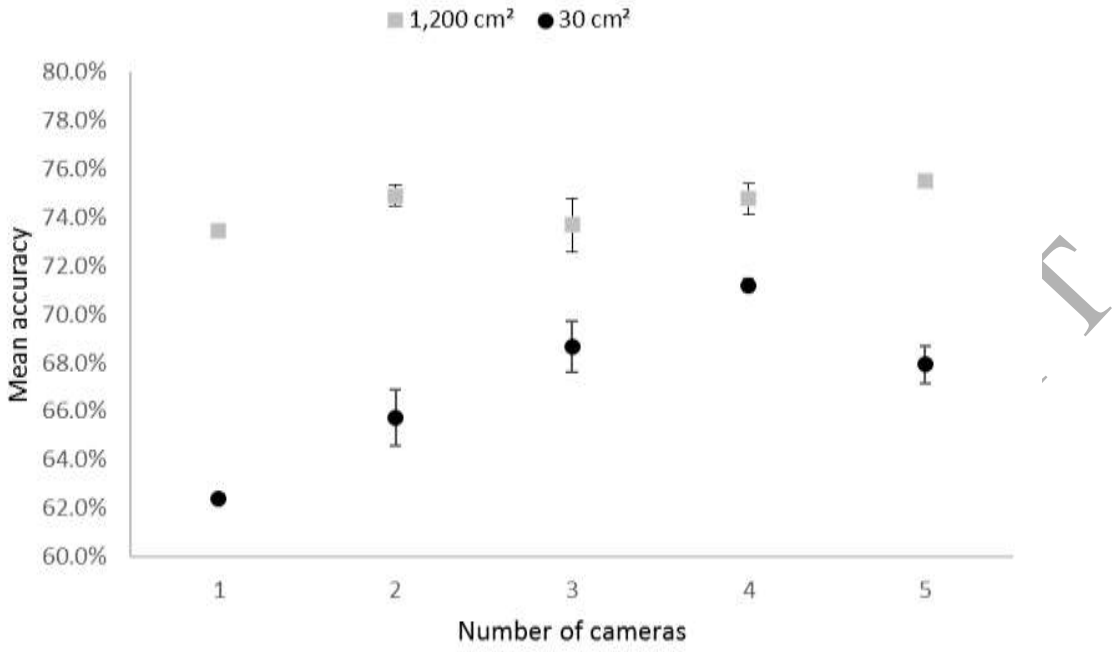


Fig. A10. Effects of using different number of cameras to increase the number of training images, for different crop sizes. The vertical bars denote standard deviation.

8 REFERENCES

Al Sabbagh, M. K., Velis, C. A., Wilson, D. C., & Cheeseman, C. R. (2012). Resource management performance in Bahrain: a systematic analysis of municipal waste management, secondary material flows and organizational aspects. *Waste Management & Research*, *30*(8), 813–824. https://doi.org/10.1177/0734242X12441962

Allesch, A., & Brunner, P. (2015). Material Flow Analysis as a Decision Support Tool for Waste Management: A Literature Review. *Journal of Industrial Ecology*, *19*(5), 753–764. https://doi.org/10.1111/jiec.12354

Allesch, A., & Brunner, P. (2017). Material Flow Analysis as a Tool to improve Waste Management Systems: The Case of Austria. *Environmental Science & Technology*, *51*(1), 540–551. https://doi.org/10.1021/acs.est.6b04204

Arena, U., & Di Gregorio, F. (2014). A waste management planning based on substance flow analysis. *Resources, Conservation and Recycling*, *85*, 54–66. https://doi.org/10.1016/j.resconrec.2013.05.008

Bell, S., & Bala, K. (2015). Learning visual similarity for product design with convolutional neural networks. ACM Transactions on Graphics, 34(4), 98:1-98:10. https://doi.org/10.1145/2766959

Bell, S., Upchurch, P., Snavely, N., & Bala, K. (2015). Material recognition in the wild with the materials in context database. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 3479–3487). https://doi.org/10.1109/CVPR.2015.7298970

Buda, M., Maki, A., & Mazurowski, M. A. (2017). A systematic study of the class imbalance problem in convolutional neural networks. *ArXiv:1710.05381 [Cs, Stat]*. Retrieved from http://arxiv.org/abs/1710.05381

Circular economy package: Four legislative proposals on waste - Think Tank. (2018). Retrieved April 24, 2018, from http://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_BRI(2018)614766

Ciresan, D. C., Meier, U., Masci, J., Gambardella, L. M., Schmidhuber, J., Cireşan, D. C., … Schmidhuber, J. (2011). Flexible, High Performance Convolutional Neural Networks for Image Classification. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence* (pp. 1237–1242). https://doi.org/10.5591/978-1-57735-516-8/IJCAI11-210

Council of the European Union. (1999). Council Directive 1999/31/EC of 26 April 1999 on the landfill of waste. *Official Journal of the European Communities. L 182/1*, 1–19.

European Parliament. (2009). *Directive 2009/28/EC of the European Parliament and of the Council of 23 April 2009*. Retrieved from http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:140:FULL:EN:PDF

Gong, Y., Wang, L., Guo, R., & Lazebnik, S. (2014). Multi-scale Orderless Pooling of Deep Convolutional Activation Features. In *Computer Vision – ECCV 2014* (pp. 392–407). Springer, Cham. https://doi.org/10.1007/978-3-319-10584-0_26

Habib, K., Schibye, P. K., Vestbø, A. P., Dall, O., & Wenzel, H. (2014). Material Flow Analysis of NdFeB Magnets for Denmark: A Comprehensive Waste Flow Sampling and Analysis Approach. *Environmental Science & Technology*, *48*(20), 12229–12237. https://doi.org/10.1021/es501975y

He, K., & Sun, J. (2015). Convolutional neural networks at constrained time cost. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 5353–5360). https://doi.org/10.1109/CVPR.2015.7299173

ISWA. (2016). *Circular Economy: Resources and opportunities*. International Solid Waste Association. Retrieved from http://www.iswa.org/fileadmin/galleries/Task_Forces/Final_Task_Force_Report.pdf

Jaderberg, M., Simonyan, K., Zisserman, A., & kavukcuoglu, koray. (2015). Spatial Transformer Networks. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems 28* (pp. 2017–2025). Curran Associates, Inc. Retrieved from http://papers.nips.cc/paper/5854-spatial-transformer-networks.pdf

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., … Darrell, T. (2014). Caffe: Convolutional Architecture for Fast Feature Embedding. *ArXiv:1408.5093 [Cs]*. Retrieved from http://arxiv.org/abs/1408.5093

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 25* (pp. 1097–1105). Curran Associates, Inc. Retrieved from http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

Laner, D., Rechberger, H., Feketitsch, J., & Fellner, J. (2016). A Novel Approach to Characterize Data Uncertainty in Material Flow Analysis and its Application to Plastics Flows in

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Austria. *Journal of Industrial Ecology*, *20*(5), 1050–1063. https://doi.org/10.1111/jiec.12326

Laptev, D., Savinov, N., Buhmann, J. M., & Pollefeys, M. (2016). TI-POOLING: Transformation-Invariant Pooling for Feature Learning in Convolutional Neural Networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 289–297). https://doi.org/10.1109/CVPR.2016.38

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Moriguchi, Y., & Hashimoto, S. (2016). Material Flow Analysis and Waste Management. In *Taking Stock of Industrial Ecology* (pp. 247–262). Springer, Cham. https://doi.org/10.1007/978-3-319-20571-7_12

OpenCV library 3.1.0. (2015). Retrieved May 2, 2018, from https://docs.opencv.org/3.1.0/

Peddireddy, S., Longhurst, P. J., & Wagland, S. T. (2015). Characterising the composition of waste-derived fuels using a novel image analysis tool. *Waste Management*, *40*, 9–13. https://doi.org/10.1016/j.wasman.2015.03.015

Pivnenko, K., Laner, D., & Astrup, T. F. (2016). Material Cycles and Chemicals: Dynamic Material Flow Analysis of Contaminants in Paper Recycling. *Environmental Science & Technology*, *50*(22), 12302–12311. https://doi.org/10.1021/acs.est.6b01791

Rawat, W., & Wang, Z. (2017). Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Computation*, *29*(9), 2352–2449. https://doi.org/10.1162/neco_a_00990

Razavian, A. S., Azizpour, H., Sullivan, J., & Carlsson, S. (2014). CNN Features off-the-shelf: an Astounding Baseline for Recognition. *ArXiv:1403.6382 [Cs]*. Retrieved from http://arxiv.org/abs/1403.6382

Rechberger, H., Cencic, O., & Frühwirth, R. (2014). Uncertainty in Material Flow Analysis. *Journal of Industrial Ecology*, *18*(2), 159–160. https://doi.org/10.1111/jiec.12087

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., … Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, *115*(3), 211–252. https://doi.org/10.1007/s11263-015-0816-y

Sedlak, D. (2017). From the Ontonagon Boulder to the Circular Economy. *Environmental Science & Technology*, *51*(4), 1941–1942. https://doi.org/10.1021/acs.est.7b00430

Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. *ArXiv:1409.1556 [Cs]*. Retrieved from http://arxiv.org/abs/1409.1556

Stanisavljevic, N., & Brunner, P. H. (2014). Combination of material flow analysis and substance flow analysis: A powerful approach for decision support in waste management. *Waste Management & Research*, *32*(8), 733–744. https://doi.org/10.1177/0734242X14543552

The circular economy package: new EU targets for recycling | News | European Parliament. (2018, April 16). Retrieved April 24, 2018, from http://www.europarl.europa.eu/news/en/headlines/society/20170120STO59356/the-circular-economy-package-new-eu-targets-for-recycling

Tonini, D., Dorini, G., & Astrup, T. F. (2014). Bioenergy, material, and nutrients recovery from household waste: Advanced material, substance, energy, and cost flow analysis of a waste refinery process. *Applied Energy*, *121*, 64–78. https://doi.org/10.1016/j.apenergy.2014.01.058

Turner, D. A., Williams, I. D., & Kemp, S. (2016). Combined material flow analysis and life cycle assessment as a support tool for solid waste management decision making. *Journal of Cleaner Production*, *129*, 234–248. https://doi.org/10.1016/j.jclepro.2016.04.077

Velis, C. A., Wagland, S., Longhurst, P., Robson, B., Sinfield, K., Wise, S., & Pollard, S. (2013). Solid recovered fuel: materials flow analysis and fuel property development during the mechanical processing of biodried waste. *Environmental Science & Technology*, *47*(6), 2957–65. https://doi.org/10.1021/es3021815

Vrancken, C., Longhurst, P. J., & Wagland, S. T. (2017). Critical review of real-time methods for solid waste characterisation: Informing material recovery and fuel production. *Waste Management*, *61*, 40–57. https://doi.org/10.1016/j.wasman.2017.01.019

Wagland, S. T., Dudley, R., Naftaly, M., & Longhurst, P. J. (2013). Determination of renewable energy yield from mixed waste material from the use of novel image analysis methods. *Waste Management*, *33*(11), 2449–2456. https://doi.org/10.1016/j.wasman.2013.06.021

Wagland, S. T., Veltre, F., & Longhurst, P. J. (2012). Development of an image-based analysis method to determine the physical composition of a mixed waste material. *Waste Management*, *32*(2), 245–248. https://doi.org/10.1016/j.wasman.2011.09.019

Wiebel, C. B., Valsecchi, M., & Gegenfurtner, K. R. (2013). The speed and accuracy of material recognition in natural images. *Attention, Perception, & Psychophysics*, *75*(5), 954–966. https://doi.org/10.3758/s13414-013-0436-y

Zoboli, O., Laner, D., Zessner, M., & Rechberger, H. (2016). Added Values of Time Series in Material Flow Analysis: The Austrian Phosphorus Budget from 1990 to 2011. *Journal of Industrial Ecology*, *20*(6), 1334–1348. https://doi.org/10.1111/jiec.12381

ACCEPTED MANUSCRIPT