Resources, Conservation & Recycling 172 (2021) 105685

Contents lists available at ScienceDirect

Resources, Conservation & Recycling

journal homepage: www.elsevier.com/locate/resconrec

Full length article

Deep learning computer vision for the separation of Cast- and Wrought-Aluminum scrap

Dillam Díaz-Romero [a,b,*], Wouter Sterkens [a,b], Simon Van den Eynde [a], Toon Goedemé [b], Wim Dewulf [a], Jef Peeters [a]

[a] Department of Mechanical Engineering-KU Leuven, 3001 Leuven, Belgium
[b] PSI-EAVISE-KU Leuven, 2860 Sint-Katelijne-Waver, Belgium

ARTICLE INFO

Keywords:
Artificial intelligence
Automatic sorting
Scrap recycling
Cast and wrought Aluminum
Deep learning computer vision
Object detection and recognition

ABSTRACT

In consequence of the electrification and the increased adoption of lightweight structures in the automotive industry, global demand for wrought Aluminum (Al) is expected to rise while demand for cast Al will stagnate. Since cast alloys can only be converted to wrought alloys by energy-intensive processes, the most promising strategy to avoid the emergence of excess Al cast alloys scrap is to sort cast from wrought alloys. To date, the separation of complex mixes of non-ferrous metals often implies the use of either or both sink-float techniques and/or X-ray fluorescence (XRF) based sorting. Therefore, the presented research develops an efficient method to classify cast and wrought (C&W) alloys in a real-time system with a conveyor belt using transfer learning methods, such as fine-tuning and feature extraction. Five CNNs are evaluated to classify C&W alloys using colour and depth images and transfer learning methods. In addition, the early fusion and late fusion of colour and depth images of C&W Al are investigated. For early fusion, data is added as an extra input channel to the first convolution layer of the CNN, and for later fusion, the images are fed in two separate subnetworks with the same architecture, where the parameters of the fully-connected layers are concatenated in both subnetworks. Our approach shows that late fusion CNN DenseNet allows obtaining the best performances and can achieve up to 98% accuracy.

1. Introduction

Due to the growing international demand for Aluminum (Al), Al is currently more produced than all other non-ferrous metals, and it is the second most used metal after steel (Cullen and Allwood, 2013). Al alloy compositions have been used for a long time, primarily due to their properties in various industries, such as transportation, aerospace, construction, packaging, electronics, and the automotive industry. The outstanding strength-to-weight ratio of Al has led to significant weight reductions in the automotive and aerospace sector, reducing large amounts of fuel consumption in the use phase (Association European Aluminium, 2013). However, the primary Al production itself is still responsible for a significant impact on the environment. In 2013 alone, its production was responsible for 3.5% of the global electricity consumption and 1% of the global CO_2 emissions (Cullen and Allwood, 2013).

With the growing demand for Al, there is an increasing need to develop and improve Al alloys' recycling processes that preserve metal alloys at their highest economic value. Recycled or "secondary" Al leads to 85% less solid waste, 95% less CO_2 emissions, 99% less CO emissions, and compared to the primary production, it almost eliminates the emissions of polycyclic aromatic hydrocarbons. Hence, compared to primary Al, secondary Al has a significantly lower impact on various relevant environmental issues, such as resource depletion, greenhouse gas emissions, toxicity, and ocean acidification (Schlesinger, 2013; The Economist, 2007). However, recycling schemes for Al are under pressure as the balance between supply and demand of secondary Al is at a tipping point. The demand for wrought alloys continues to rise for various applications, while cast alloys' demand is expected to stagnate in the near future (Modaresi, 2015).

The European Al Association (EAA) (European Aluminum, 2019) EAA 2050 vision could substantially contribute to the Al industry to decarbonize the European economy. However, the biggest challenge in achieving the use of secondary Al for various products is, according to

* Corresponding author.
E-mail addresses: dillam.diaz@kuleuven.be, diazrome@ualberta.ca (D. Díaz-Romero).

https://doi.org/10.1016/j.resconrec.2021.105685
Received 16 October 2020; Received in revised form 20 April 2021; Accepted 15 May 2021
Available online 28 May 2021

the EAA, not the availability of Al scrap. Instead, the EEA stresses that state-of-the-art Al sorting methods need to be substantially improved and be able to sort the waste stream into its respective alloy with a purity that allows the use of these alloys for applications that currently require virgin Al (European Aluminum, 2019). Therefore, the development and industrial implementation of automated shredding and separation technologies are necessary to reduce Al waste and improve the limited purity of recycled Al.

1.1. Separation of Cast and Wrought Aluminum

35-40% of the total Al consumption is governed by the transportation industry, which currently consists of a complex mix of materials (Soo et al., 2018). Their efficient separation often implies shredding, leaving a very high fraction of mixed material to be sorted from electrical equipment (WEEE) and end-of-life vehicles (ELVs). When not dismantled for depollution or reuse, these end-of-life products undergo a size reduction to facilitate further sorting steps. In this sorting process, the first step is commonly removing iron by magnetic separation (Mesina et al., 2005). Subsequently, sink-float separation methods are frequently applied to remove non-metallic materials, such as plastics, wood, and rubber from the metals. Thereafter, heavy media separation processes using ferrosilicon suspensions at densities of 3200–4200 kg/m3 are often used to separate the heavy metals, such as stainless steel, copper, brass zinc, from the light non-ferrous metals. To purify the light non-ferrous metals and separate all non-metals, eddy current separation methods are commonly applied.

The further sorting of the light and heavy non-ferrous metals is currently either done by manual sorting or by automated sorting techniques. For example, colour-based sorting can be used to sort copper and brass from other heavy non-ferrous metals (Koyanaka and Kobayashi, 2010). Alternative methods are X-ray Fluorescence (XRF) and X-ray Transmission (XRT) systems, which can be used to separate metals either based on the fluorescence spectra, which are specific for every atomic number, or on the material density. However, Al alloys do not differ significantly in colour, and both the density and fluorescence spectra of the alloy elements Si and Mg and Al themselves differ insufficiently to be able to implement XRF or XRT for sorting. Consequently, the separation of Al scrap based on their alloy is not possible with the sorting methods currently implemented on an industrial scale.

Nonetheless, Cast and Wrought (C&W) Al alloys have complex and very distinct inclusions of a variety of alloying and tramp elements. Whereas the alloy elements contribute to specific and desired properties, the tramp elements are contaminants that do not contribute to the desirable quality. Alloying and tramp elements generally constitute up to 20% of cast alloys' total weight and 10 of wrought alloys' total weight (Paraskevas et al., 2015). As a result, Al scrap cannot be used directly for the production of wrought alloys. The production of recycled Al can be achieved by either or both, sorting and/or diluting it with primary Al and/or the addition of the desired alloying elements (Capuzzi and Timelli, 2018; Paraskevas et al., 2015). As an alternative to dilution, an electrolytic process, commonly referred to as the Hoopes process, can also be used to remove magnesium, manganese, silicon, iron, zinc, copper, and chromium and purify the aluminum. However, the Hoopes process is not commonly implemented on an industrial scale or not an option to be implemented, as it is highly energy-intensive and, hence, a costly process (Haraldsson and Johansson, 2018).

Therefore, different approaches have been investigated to sort Al alloys based on their content of alloying and tramp elements. In 1983, the Bureau of Mines developed a technology that uses the thermo-mechanical behaviour of C&W alloys for scrap separation (Ambrose et al., 1983). With this technique, aluminum scrap is first heated to a temperature between 550°C and 580°C in a small pot furnace and then dropped into a laboratory jaw crusher. In this process, the wrought material retains its shape, while 100% of the cast aluminum was pulverized, allowing it to separate both by sieving (Ambrose et al., 1983). However, since those processes require even more energy than the primary production process of aluminum (17-18 kWh/kg compared to 14 kWh/kg), it is not considered economically viable (Gaustad et al., 2012).

In 2005, the Delft University of Technology developed and tested an automated scrap with a combined electro-dual energy X-ray transmission sensor. The system demonstrated to be able to classify between 15 and 45 pieces of the C&W Al with an accuracy of 68% and 50%, respectively (Mesina et al., 2005). In the past decade, Laser-Induced Breakdown Spectroscopy (LIBS) systems have also been investigated to increase the robustness and cost-efficiency of automated sorting for Al alloys (Capuzzi and Timelli, 2018; Gurell et al., 2012; Werheit et al., 2011). For example, Werheit *et al.* proposed in 2011 a fast single piece identification by three-dimensional LIBS scanning. In this research, a total of 60 and 168 C&W Al scrap pieces were used for the classification measurements. After discarding 20% of the data as outliers, low and high silicon alloyed Al pieces were identified with the correctness of over 96%. Results also demonstrated that the LIBS system required for this classification a high analytical precision, which limits its use in an on-line system (Werheit et al., 2011).

As an alternative to the LIBS and XRF systems, Koyanaka and Kobayashi proposed in 2010 an automatic classification of light metal scrap using differences in bulk density and three-dimensional (3D) shape (Koyanaka and Kobayashi, 2010). The research used computer vision techniques combined with a 3D linear camera and a mechanical balance to sort 104 pieces of cast Al, 192 pieces of wrought Al, and 246 pieces of Mg. The system classified C&W Al with an accuracy of approximately 90%. This system demonstrates the ability to use shape parameters, such as width, height, volume, and projected area of irregular-shaped metal parts to classify C&W Al, but has the limitation that just one piece can be analyzed at once.

With the knowledge that shape parameters can be used, it is valuable to investigate how computer vision techniques could be used to obtain a higher-level understanding based on images using computers as a source. Nowadays, computer vision is shifting from hand-crafted methods to deep neural network methods, using convolutional neural networks (CNNs) as a technique to teach a system to learn by example (LeCun et al., 1998). Therefore, CNNs have been used as an effective solution for waste sorting and surface imperfection detection due to their capacity of detection and accuracy (Bobulski and Kubanek, 2019; Konovalenko et al., 2020; Li et al., 2016). However, the number of images required to train these non-line models on a specified task is primarily unavailable and/or complex and, hence, costly to annotate with high accuracy.

Therefore, the presented research integrates existing approaches by combining RGB and 3D images using feature concatenation and deep feature extraction to classify objects. In addition, because the training of CNNs typically requires a large and correctly annotated dataset, the use of transfer learning algorithms, such as fine-tuning and feature extraction, is investigated to boost classifying the performance of C&W Al when only a relatively small dataset is available. Finally, different Deep Learning models are compared to identify which models are most suited for Al alloy classification based on 3D and RGB linear cameras.

Our study makes two novel contributions to the research area of aluminum recognition:

1) We have developed a data-analysis pipeline and Deep Learning algorithms to accurately detect C&W Al by using RGB, I, D images as data sources; and
2) We have evaluated the potential of Deep Learning models for detecting C&W Al in an online application.

The remainder of this paper is organized as follows. In Section 2, we summarize background information and related work using Deep Learning to classify recycling. In Section 3, we outline our data-collection procedure, our image processing methods, and Al classifications using the fusion of 3D and RGB images. In Section 4, we explain our

evaluation methodology and present our results. Section 5 concludes the paper and discusses future work.

2. Related work

2.1. Deep learning in recycling

The development of CNNs has made it possible to improve and rely on computer vision systems for broad applications in image classification, segmentation, and detection regardless of the high variations in visual information, which made it impossible to manually tune statistical methods (Chu et al., 2018; He et al., 2016; Shao et al., 2017). The use of CNNs has mainly been investigated in the context of object recognition and classification in RGB images. Bobulski and Kubanek proposed to use an RGB digital system for classifying plastic waste in four different classes (polyethylene terephthalate, high-density polyethylene, polypropylene, and polystyrene) by using a six layers CNN with a classifying accuracy of over 92% (Bobulski and Kubanek, 2019). Additionally, Dang et al. (Dang et al., 2019) demonstrated and compared the use of AlexNet, GoogleNet, VGGNet, and ResNet CNNs to separate metallic debris such as brass, copper, and Al by using a high-resolution RGB camera. Their results showed the possibility of classifying metallic materials with an accuracy of up to 96%, while the accuracy was highly dependent on the number of classes in which the materials had to be classified. It is essential to notice that this research focused on classifying materials that are easy to distinguish based on the colour or shape, while in practice, many materials that are essential to be sorted for recycling, such as C&W Al alloys, are less distinguishable.

In contrast to single-sensor systems, the use of multi-sensors, such as multi-camera systems (RGB, Spectral, 3D, thermal), scanner sensors (XFR, LIBS), or combinations between cameras and sensors could collect more detailed information about the recycled material. Combined systems allow researchers to improve object detection and recognize a higher number of objects simultaneously (Ophoff et al., 2019; Schwarz et al., 2015). In using multi-sensor systems, Chu et al. proposed a multi-layer hybrid Deep Learning method for classifying paper, plastic, metal, glass, and other recycling waste (Chu et al., 2018). The study deployed a high-resolution RGB camera to capture 100 debris images and an induction sensor to facilitate the sorting of solid debris. An algorithm based on AlexNet CNN was used to extract image features, and a multi-layer perceptron method was applied to classify between other items or recyclable waste, achieving an overall classification accuracy of over 90%. Therefore, in this document, CNNs have been proposed as a novel way to classify C&W Al by using a multi-camera RGB and a 3D camera system.

2.2. Related work to RGB-D fused images

Compared to object recognition using images, RGB camera-based recognition provides limited information about objects' appearance in different scenes (Shao et al., 2017). In the recycling field, the use of RGB camera images taken on the conveyor belt could complicate the partition of the foreground and the object's background when the object has similar colours and textures as the conveyor. Additionally, detections on RGB images in real-time applications are often affected by the illuminance changes. As shown in the RGB images in Fig. 2, the two Al scraps have similar colours compared to the RGB+D image in the case of the C&W classification, making their identification harder without proper lighting. RGB+D is known as the fusion of the RGB image and 3D image. Fused images provide improved scene awareness due to depth information per each pixel, leading to a more accurate and robust performance on vision applications (Shao et al., 2017).

In recent decades, RGB+D images applied to CNNs have been used in diverse applications such as object recognition, label recognition, robot localization, navigation, pedestrian detection, and recycling. In 2015, Zhang developed a relative orientation and position detection based on an RGB-D image (Zhang et al., 2015). Furthermore, RGB-D object recognition using a pre-trained CNN was proposed by Schwarz et al. for instantaneous recognition and pose estimation of objects on planar surfaces (Schwarz et al., 2015). Wang et al. proposed a computer vision system based on 100 RGB-D images on a dusty conveyor belt as a segmentation method for segmenting construction debris in harsh environments (Wang et al., 2017).

Moreover, new studies have been conducted to evaluate the most efficient way to fuse RGB and 3D images. Shao et al. investigated Deep Learning's employment for extracting and fusing features from RGB-D data (Shao et al., 2017). The research compared in-depth features from RGB-D feature concatenation and in-depth features from RGB-D fusion. In RGB-D feature concatenation, the features are extracted from the last layer of the two CNNs that are trained based on RGB data and depth image (late fusion). Then, the characteristics extracted from the last layer of the two CNNs in parallel are concatenated, as shown in Fig. 4. The image is merged in the deep feature case before it is fed to a CNN (early fusion). Furthermore, Ophoff et al. investigated the determination of the optimal fusion point in the network, from which they concluded that fusing towards the late layers provides the best results (Ophoff et al., 2019).

3. Material & methods

3.1. Materials

548 mixed Al scrap samples with different shapes (e.g. compact, bar, sheet, pipe, and irregular), an average mass of 37g, and sizes between 10 and 280 mm were collected from a Belgian recycling facility. The samples were gathered randomly from a highly mixed shredded non-ferrous mixture of almost exclusively aluminum alloys, also known as twitch. For this study, 428 wrought and 120 cast Al scrap samples were collected from the twitch fraction. For the ground truth of the used classifiers, a 548 shredded scrap metal dataset was developed by combining images captured on the conveyor belt using a Niton™ XL2 XRF analyzer, suitable for scrap metal identification for cross-analyzing all the samples.

3.2. Methods

Fig. 1 represents the approach chosen for this study. First, the images are captured by using our computer vision system, then, using traditional computer vision, the Region Of Interests (ROIs) are extracted. Once the images are detected and cropped, the images are fed into CNN to classify C&W Al.

3.2.1. RGB & D cameras

548 × 3 (RGB, Intensity, and Depth) images per metal scrap piece are captured on a conveyor belt by using the constant light system METAPHASE UL-CLL709-W-24Z, a Dalsa 2K line Gige vision™ colour camera with a 52 Frame Rate (fps) and two LMI GOCATOR 2340 3D laser line profile sensors with a scan rate of 5 kHz synchronized by an LMI GOCATOR MASTER 810. Before capturing the photos, the colour camera was calibrated to ensure the camera output is uniform and set at a correct white balance.

3.2.2. ROI extraction

The programming language Python V. 3.7 and the OpenCV library are used to pre-process the colour and depth images for identifying the ROI, as well as for image merging and analysis of results. Furthermore, the Pytorch library is used to modify the network architectures. Since the images contain multiple scrap metals, an ROI extraction step is used to locate all the scrap metals in the images, as shown in Fig. 1. For the ROI extraction, algorithms are used to distinguish the objects from the background (Dang et al., 2019; Suzuki, 1985).

The first step to detect the ROI is to apply a morphological

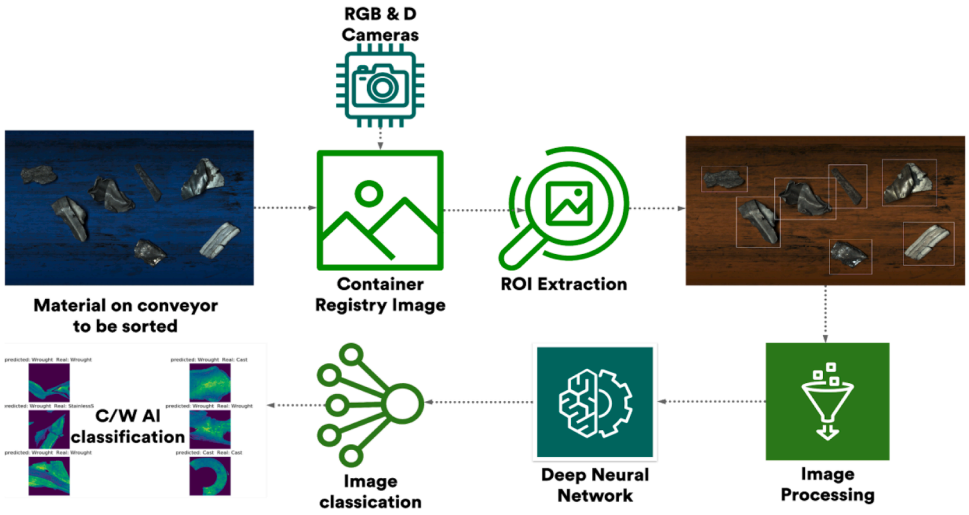

Fig. 1. Flowchart of the proposed approach to classify Cast and Wrought Aluminum.

transformation, such as image smoothing with a Gaussian filter, erode, dilate, and Canny edge detection (Lorsakul and Suthakorn, 2007). Thereafter, adaptive thresholding for transforming is applied to convert the grayscale into a binary image.

The threshold algorithm can be used to create a binary image. When the pixel value is more significant than a threshold value, it is assigned a value (most likely white). In all other cases, it is assigned another value (most likely black) (Mordvintsev and Abid, 2014). However, the global value as a threshold might not be good enough in all the conditions where the image has different light conditions in a different section of the image, such as the Al scraps in our dataset with irregular reflections and shapes. In those cases, it is recommended to use an adaptive threshold (Mordvintsev and Abid, 2014). The adaptive threshold is an algorithm that calculates the threshold for a small region of the image. The resulting image has multiple thresholds for different regions of the same image, which helps avoid thresholding out saturated pixels in the resulting binary image.

The *cv2.ADAPTIVE_THRESH_GAUSSIAN_C* adaptive thresholding algorithm was used to transform RGB images to grayscale; the algorithm uses a weighted sum of neighbourhood values where weights are a Gaussian window. In the performed experiments, the following parameters were used: Thresholding type (*THRESH_BINARY*) and BlockSize (5-pixel). Afterwards, the obtained binary image is passed to a *findContour* function, which returns the object contours used for cropping the objects from the gathered images.

3.2.3. Image processing

Once all the D, I, RGB images have been cropped, the following three additional subsets of images are created: Depth Image Heat map (DHEAT), RGB+D, and RGB+I mixes, as shown in Fig. 2. These types of images are generated, as they have also been used with success in previous research to improve network classification (Ophoff et al., 2019; Schwarz et al., 2015; Shao et al., 2017). A DHEAT subset is generated by converting the D images to RGB using a Viridis perceptually uniform sequential colour-map. RGB+D and RGB+I mixes are generated by merging the images using early fusion and late fusion (Shao et al., 2017), as shown in Fig. 4.

3.3. Network architecture

Deep neural networks have been very successful in image classification and object recognition (Chu et al., 2018; Dang et al., 2019; He et al., 2016; Shao et al., 2017; Werheit et al., 2011). CNN is a class of deep neural networks inspired by the visual cortex system (Hubel and Wiesel, 1968) and is mainly used to analyze visual imagery. CNNs are composed of convolutional and pooling layers. Convolutional layers help to extract a map of different features or parameters, highlighting the presence of a given characteristic in the image, such as geometrical, distance, colour, and shape information or to detect the edge or execute the sharpening of the object. Pooling layers are applied to select the largest values on the feature maps and use them as inputs to subsequent layers. Typically, at the end of CNN's, two or three layers are fully-connected in order to generate the output of class probabilities.

To evaluate and compare the efficiency of C&W Al classification for the seven datasets and different state-of-the-art image classification CNNs, the following five CNNs are used: **VGG16** (Simonyan and Zisserman, 2014), **AlexNet** (Krizhevsky, 2014), **ResNet-18** (He et al., 2016), **DenseNet** (Huang et al., 2017), and **SqueezeNet** (Iandola et al., 2016). **AlexNet, ResNet-18**, and **VGG16** were used in previous research to separate metal debris (Chu et al., 2018; Dang et al., 2019). DenseNet,

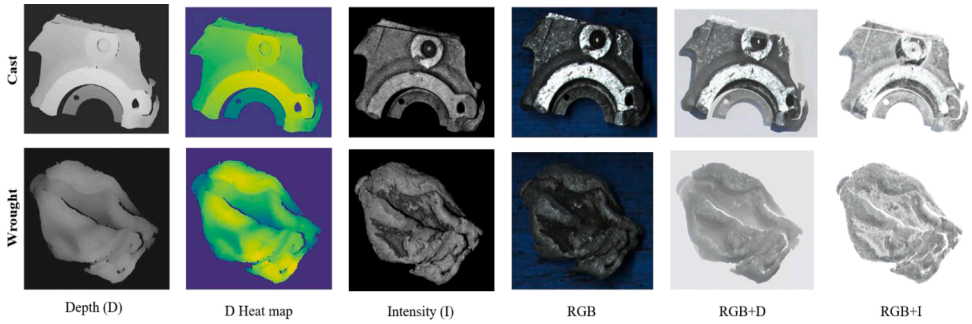

Fig. 2. Example of Cast and Wrought Aluminum images used in this study.

VGG16, and ResNet-18 have been used in prior research to classify solid and recyclable waste (Aral et al., 2018; Chu et al., 2018), and **SqueezeNet** was used to detect bottle statues for recycling (Mwangi and Mokoena, 2019). As shown in Table 1, the used CNN architectures strongly differ in their model size (in MB). The classification error rate for classifying the ImageNet dataset in the large-scale visual recognition challenge (Russakovsky et al., 2015) and several convolutions and fully-connected layers. The Top-5 error represented the percentage when the classifier did not include its correct best five guesses. As shown by the size, SqueezeNet is the smallest CNN architecture, but it has been demonstrated to achieve AlexNet-level accuracy. The advantage of the network's smaller size is that it allows real-time detections on embedded devices, such as the NVIDIA Jetson TX2 (Sanchez et al., 2018). SqueezeNet architecture has eight fire modules, each of which is made up of a 1×1 filter known as a squeeze convolution layer, which is fed into an expanded layer that has a mix of 1×1 and 3×3 convolutions filters (Iandola et al., 2016). Those fire modes attempt to decrease the number of parameters in the CNN while trying to preserve its accuracy. Differently, DenseNet architecture uses 3 transitions and 2 DenseBlocks. A DenseBlock can be made up of multiple groups of 1×1 and 3×3 convolution filters, where the feature map remains constant, but the number of filters changes in between them. In the DenseNet CNN, the transition layers are 1×1 convolution and 2×2 polling layers between the DenseBlocks that are responsible for the subsampling when normalizing the input layer by re-centering and rescaling (Batch normalization). DenseNet architecture has the advantage to reuse features, which has been reported to perform well when training data is insufficient (Huang et al., 2017).

3.3.1. *Pre-trained network and preprocessing methodologies*

Due to the absence of qualitative and large datasets and the high cost to develop such datasets, transfer learning's applicability is investigated. Since two approaches have been successfully implemented in previous research (He et al., 2016; Schwarz et al., 2015), the two transfer learning methods have also been investigated: fine-tuning and feature extraction, which both were pre-trained on the 1000-class ImageNet dataset.

In Fig. 3, a hypothetical example of a CNN containing two convolutional layers, one pooling layer, and two fully-connected layers is shown. For fine-tuning (shown in blue in Fig. 3), the training starts with a pre-trained model, and all the model parameters of the network are updated for a new task, essentially re-training the entire model. Feature extraction (represented in orange in Fig. 3) starts with a pre-trained model but only updates the parameters of the final layer, from which the prediction is derived (Li and Hoiem, 2017).

It is necessary to pre-process the input data to match the pre-trained weights and the neural network architecture to use transfer learning. For this all pre-trained models, except for the 3-channel RGB images of the shape ($3 \times 224 \times 224$), are normalized using the mean = [0.485, 0.456, 0.406] and std = [0.229, 0.224, 0.225]. During the re-training for both *fine-tuning* and *feature extraction*, data image argumentations are applied using 75% Random Resized crop, Random Vertical and Horizontal rotation, as this has been demonstrated to improve the image classification (Perez and Wang, 2017; Wong et al., 2016) and to reduce the influence of incorrect ROI extraction.

3.3.2. *Fusion network architectures*

In this investigation, early fusion and late fusion for classifying C&W Al are investigated (see Fig. 4). For early fusion, data is added as an extra input channel to the first convolution layer of the CNN, while the additional channels' parameters are initialized using a copy of the R channel from the original pre-trained RGB parameters. In the case of late fusion, a network was developed in which the images are fed in two separate subnetworks with the same architecture, and where the parameters of the layers fully-connected are concatenated in both subnetworks to have a unified output. In addition, the proposed architectures use the same Cross-Entropy Loss function that combines *LogSoftmax* and *Negative Log-likelihood Loss* (NLLloss) in one single class, which is helpful when the training data has an unbalanced training set (Paszke et al., 2019).

Table 1
Summary Analysis of CNN Architectures on Model Size, Classification Error Rate, and Model Depth.

CNN Model	Develop by	Size (M)	Top-5 error (%)	Model Description	# Layers
ALEXNET (2012)	Krizhevsky (Krizhevsky, 2014)	242	15.3	5 conv + 3 fc layers	8
VGG16 (2014)	Siminyan et al., (Simonyan and Zisserman, 2014)	537	9.62	13 conv + 3 fc layers	16
RESNET-18 (2015)	Xie et al. (Xie et al., 2017)	44.8	10.92	17 conv + 1 fc layers	18
SQUEEZENET (2016)	Iandola et al.(Iandola et al., 2016)	4.8	9.38	2 conv + 8 fire + 1fc	18
DENSENET BC (2017)	Huang et al. (Huang et al., 2017)	104.1	5.92	5 conv + 3 trans +2 db +1 fc	100

"Conv," "DB," "tran," "fc" indicate convolutional, DenseBlock, transition and fully-connected layers, respectively.

4. Results and discussion

The datasets of 548 images are randomly split in 70% of training, 10% of validation, and 20% of test set images. All networks were trained for 200 epochs, with tests running every 75 batches on the validation for the training of the CNNs a single GPU: NVIDIA Tesla V100 32 GB, and for the testing a CPU: Intel® Xeon® E5-2698 v4 2.2 GHz with 512 GB DDR4 RDIMM memory are used.

4.1. *Classification of C&W Al*

Fig. 5 shows the comparison of the f1-scores, accuracy, recall, and precision using the two transfer learning methods to classify C&W Al. In this figure, the columns represent the metrics, while the rows represent the analyzed image. Each subplot contains the obtained results for the five CNN architectures. The formulas used for calculating these metrics are shown in Eqs. (1-3). This formula was applied to every class and then averaged to provide a means of evaluating the quality of sorting. The accuracy rate is calculated by the number of correctly classified images divided by the total number of images.

Nevertheless, this metric cannot be used as primary due to an imbalance of C&W Al classes. For unbalanced classes, it is helpful in determining the f1-score, which is the harmonic mean of precision and recall, as it gives a better measure to assess the number of incorrectly classified cases. Since with a higher recall, the marginal benefit of Al scrap remains high and leads to a lower cost of action for recycling scraps. It was decided to evaluate recall for the envisaged online recycling application.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{1}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2}$$

$$F1 - score = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

True-positive (TP) means that the data has the label "Cast" and the prediction has the same outcome, while True-negative (TN) occurs when

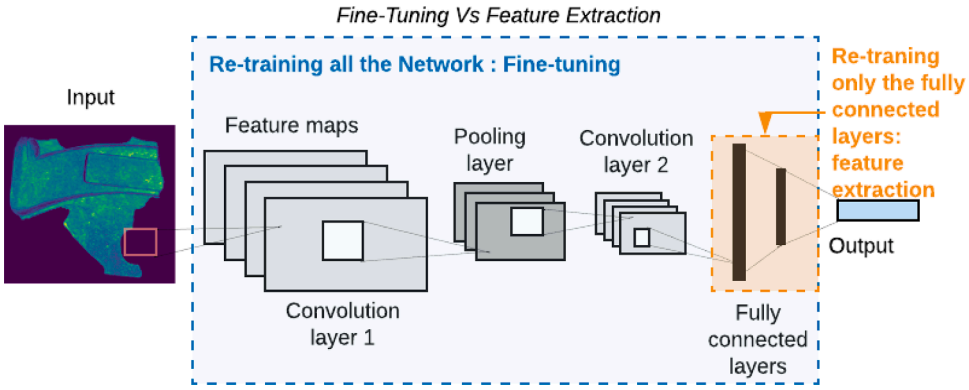

Fig. 3. An example of a CNN containing two convolutional layers, one pooling layer, and two fully-connected layers, explaining the difference between transfer learning by fine-tuning and feature extraction (He et al., 2016; Schwarz et al., 2015).

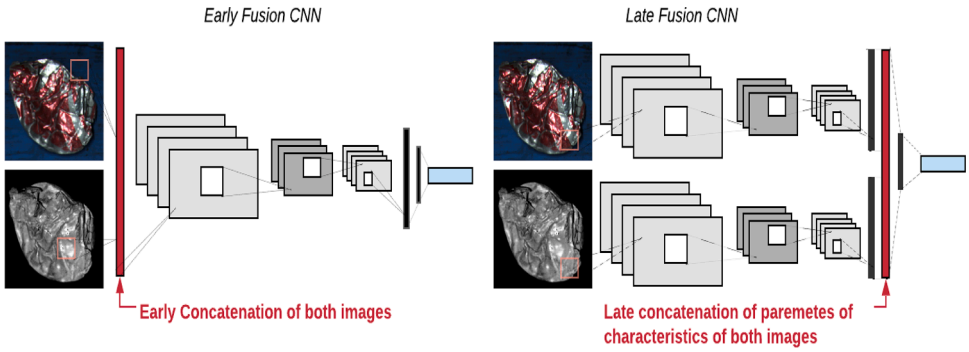

Fig. 4. Implemented fusion architectures on the left early fusion and on the right late fusion compared with each other.

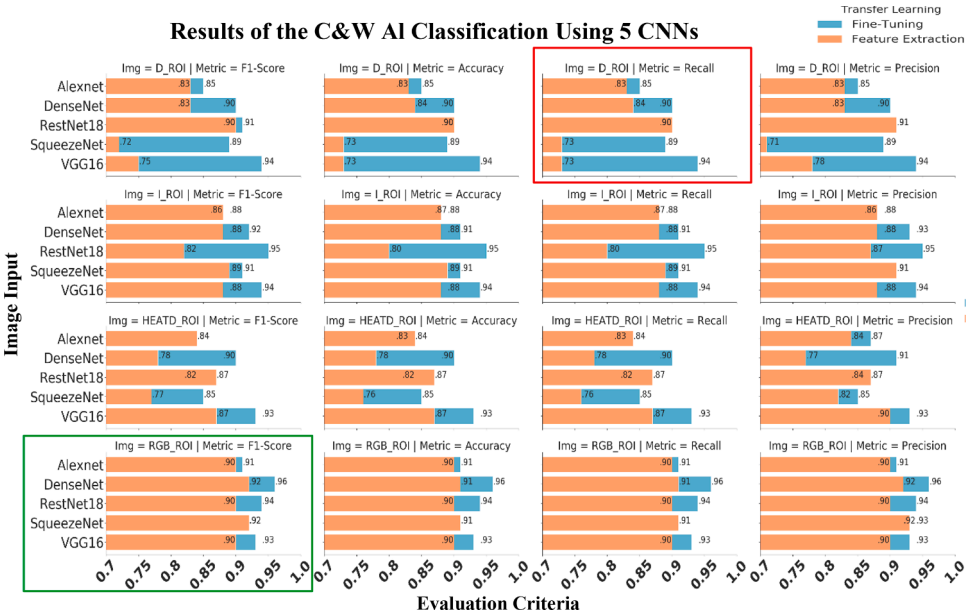

Fig. 5. Results of C&W Al classification. The figure has 16 subplots in a 4 × 4 matrix, the columns represent the metrics (F1-Score, Accuracy, Recall, and Precision), and rows present the analyzed image (D_ROI, I_ROI, HEATD_ROI, and RGB_ROI). The first two rows show the results of early fusion, and the remaining rows show the results of the late fusion. Equal to Fig. 6, the transfer learning results are represented by Fine-tuning and Feature extraction in orange and blue, respectively.

the data's label is "Wrought" and the prediction confirms that it is "Wrought." False-positive (FP) means that the data's label is "wrought," but that it was wrongly classified as "Cast." False-negative (FN) is the case when the data's label is "Cast," but the algorithm predicts "Wrought'.

In total, 40 experiments are first carried out to determine which image type is more efficient for classifying C&W Al. Fig. 5 shows the comparison of the five CNN models, transfer learning, and this for the four different metrics used. Using f1-score and recall as metrics shows that RGB images have the best performance to classify C&W Al, with

results for the different networks all above 90%, with further improvements by adopting the fine-tuning transfer learning methodology, as shown in the green bounding box in Fig. 5 The best results are achieved by feeding RGB images through a fine-tuned DenseNet, with an F1-score of 96%. The D and Heat map images showed the lowest performance with an overall efficiency of 72% to 94%, as depicted in the red bounding box in Fig. 5. Overall, these experiments demonstrate that using CNNs with RGB, I, and D is a novel and promising alternative for classifying C&W Al with high accuracy.

4.2. Comparative Results of Early and Late Fusion to Classify Al

Since the best classification performances were obtained with RGB images, an additional 40 experiments are performed to evaluate early and late fusion of RGB images with either intensity or depth images (and not all possible combinations). As shown in Fig. 6 and marketed with a green bounding box, the best classification results are obtained by combining RGB and depth images, using the fine-tuning transfer learning method and a late fusion approach. Compared to only using RGB images, an increase of 2% to 5% is obtained for all metrics. Minor to no improvements is obtained by fusing the RGB and I images. The assumption is that the computer vision system captures the intensity image using a fixed lighting system which may provide redundant or not additional information. In future experiments, datasets of images with variable lighting conditions of samples of different colours will be developed. Moreover, the increase in recall and precision is more prominent when using any fusion type. It can be noted that the early fusion using AlexNet, RestNet18, and SqueezeNet networks with D images perform optimally, achieving better results than an RGB-only network. The best classifier for early and late fusion RGB+D images is DenseNet, but for late fusion, it is RestNet18. Overall, fine-tuning performance is better in all used models. However, the performance of DenseNet and RestNet18 using RGB+D images does not vary regardless of the used transfer learning methods. The presented experiments demonstrate that RGB+D and RGB+I improve classification networks by increasing accuracy, recall, and precision. The improvements suggest that late fusion reduces the strong bias compared to the early fusion, where initializing the network's weight with the pre-trained weight of a pure RGB network may not be representative of the D image. Our research suggests that late fusion using depth images introduced lower-level features (such as edges and basic shapes), which helped in classifying C&W Al scrap pieces. The deeper we advance through the network, the more abstract the features become and the more subsamples related to a spatially specific Al class.

4.3. C&W Al classification applied to the recycling industry

Table 2 shows the comparison of the time needed for classifying 82 pieces of C&W Al, showing the: Load Time, which is the time that the system takes to crop and load the images to the CNN, the Prediction Time, which defines the time that the system requires for predicting C&W AL classes, and the Total Time, which is defined as the sum of Load and Prediction Time. As can be seen in the **bold** values in Table 2, DenseNet is the fastest classifier for early fusion and SqueezeNet for later fusion. In contrast, VGG16 and AlexNet are the slowest classifiers for early fusion and late fusion, respectively, as shown in Table 2 by the *italic* values. However, it is essential to notice that these timing results are from non-optimized implementations of the different neural networks. As the obtained performances are similar, it is considered that the runtime can be significantly improved by optimizing the implementations.

To evaluate the potential of Deep Learning computer vision for classifying C&W Al in an online sorting application, it is important to consider what would be the maximum throughput that could be achieved with such a system. Firstly, all scrap pieces are weighted to define an average mass of 37g. Using this value, it is assumed that the DenseNet classifier has to be fed by 82 pieces per second. This classifier was chosen since it performed the fastest, achieving the highest accuracy. Finally, the Machine capacities per hour were calculated as 10.635 Ton/h and 5.881 Ton/h for early and late fusion. The developed system capacity is depending on the accuracy levels. Calculations demonstrate that using late fusion leads to a better classification, while the system capacity decreases by almost half. Nonetheless, with the achieved capacity of more than five ton/h, the classification results demonstrate the feasibility of using late fusion techniques to sort aluminum. Further research on new recycling methods will be needed to maximize Al scrap purity, improve recovery yields, and reduce processing costs.

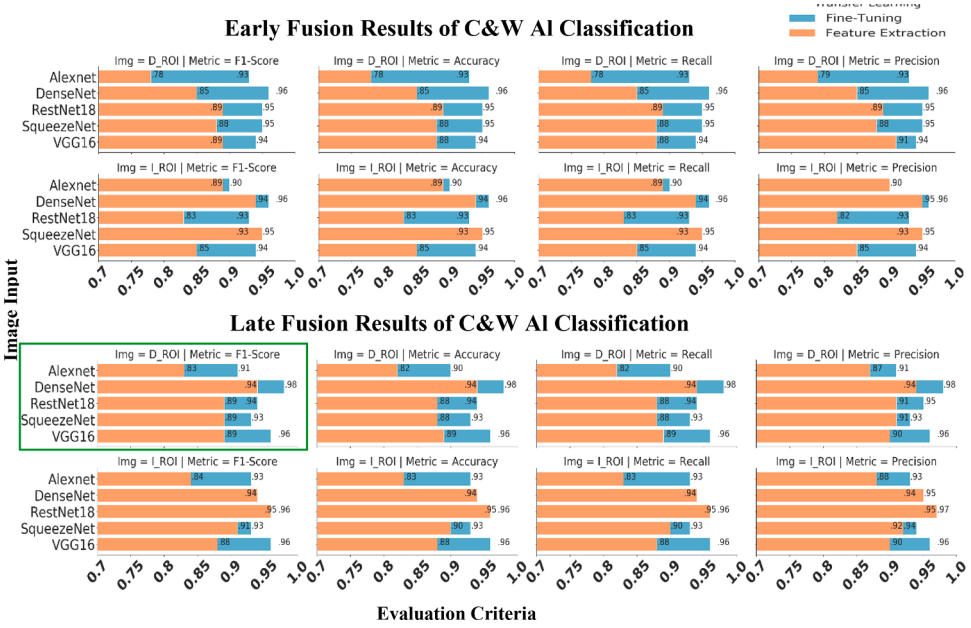

Fig. 6. RGB+D, RGB+I fusion results of C&W Al classification.

Table 2
Summary Analysis of CNN Architectures Runtime to Classify C&W Al.

		Load Time Avg. (s)		Predict Time Avg. (s)		Total Time Avg. (s)		Time per Flake Avg. (s)	
CNN model	#. Scrap pieces	Early F.	Late F.	Early F.	Late F.	Early F.	Late F.	Early F.	Late F.
ALEXNET	82	0.0031	0.0060	1.09	2.07	1.093	*2.081*	0.0133	0.0253
DENSENET	82	0.0033	0.0063	1.02	1.85	**1.027**	1.857	0.0125	0.0226
RESNET-18	82	0.0031	0.0060	1.05	1.90	1.058	1.907	0.0129	0.0232
SQUEEZENET	82	0.0030	0.0057	1.19	1.79	1.194	**1.793**	0.0145	0.0218
VGG16	82	0.0033	0.0062	1.33	1.84	*1.339*	1.846	0.0163	0.0226

"F," "s," "Avg." indicate fusion, seconds, and average.

5. Conclusions

The study investigates the benefits and limitations of image fusion and identifies techniques and practices in image processing and Deep Learning that are most suited using data from 3D and RGB linear cameras. To address the problem of small datasets in the field of recycling, the use of transfer learning is investigated. In addition, the potential of combining RGB, Depth, and Intensity images from a colour and triangulation 3D camera through data fusion is investigated. The results obtained with the CNN DenseNet and a fine-tuning, pre-trained and late fusion approach show that the developed method is up-and-coming and efficient. Best results were obtained with the CNN DenseNet and late fusion of RGB+Depth images, demonstrating accuracy of 98% for classifying C&W Al. Hence, the presented method is considered a viable method for alloy-based Al sorting, encompassing significant potential to increase the technical feasibility and economic viability of enhanced Al sorting for wrought-to-wrought recycling.

The presented research is also considered an important first step towards improved non-ferrous metal sorting and recycling. However, if Al scrap could be classified with more detail than Cast vs. Wrought into strategically pre-selected categories containing only a tiny variety of alloying elements, more recycled Al can be used directly in the production of wrought alloys. In the transition to a more circular economy, the use of Deep Learning for the sorting of aluminum alloys can be a suitable strategy to eliminate the threat of scrap surplus and significantly increase the value of recycled aluminum. Therefore, in future research, a larger dataset of different aluminum alloys will be gathered and labelled to investigate the applicability of the developed method for the classification of C&W Al, as well as the recognition of the specific Aluminum alloy and other non-ferrous materials, such as magnesium. In addition, different strategies will be investigated how the developed computer vision system can be integrated into an advanced sorting system, possibly using combinations of distinct sorting techniques, such as LIBS and XRT/F, to increase either or both the throughput and accuracy for the sorting of complex metal scrap mixture.

CRediT authorship contribution statement

Dillam Díaz-Romero: Conceptualization, Investigation, Methodology, Software, Validation, Data curation, Visualization, Writing - original draft. **Wouter Sterkens:** Conceptualization, Investigation. **Simon Van den Eynde:** Data curation, Visualization. **Toon Goedemé:** Methodology, Resources, Supervision, Writing - review & editing. **Wim Dewulf:** Supervision, Writing - review & editing. **Jef Peeters:** Supervision, Writing - review & editing, Funding acquisition.

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.

Acknowledgments

This work has received funding from the European Institute of Innovation and Technology (EIT), a body of the European Union, under the Horizon 2020, the EU Framework Programme for Research and Innovation in the AUtomatic SOrting of mixed scrap Metals (AUSOM) project (https://www.ausomproject.eu/). Thanks to F. Arslan for constructive discussions and support.

References

Ambrose, F., Brown, R.D., Montagna, D., Makar, H.V., 1983. Hot-crush technique for separation of cast- and wrought-aluminum alloy scrap. Conserv. Recycl. 6, 63–69. https://doi.org/10.1016/0361-3658(83)90018-8.
Aral, R.A., Keskin, Ş.R., Kaya, M., Hacıömeroğlu, M., 2018. Classification of trashnet dataset based on deep learning models. 2018 IEEE International Conference on Big Data (Big Data). IEEE, pp. 2058–2062.
Association European Aluminium, 2013. Aluminium in cars–unlocking the light-weighting potential. Eur. Alum. Assoc. Bruss.
Bobulski, J., Kubanek, M., 2019. Waste classification system using image processing and convolutional neural networks. International Work-Conference on Artificial Neural Networks. Springer, pp. 350–361.
Capuzzi, S., Timelli, G., 2018. Preparation and melting of scrap in aluminum recycling: a review. Metals 8, 249.
Chu, Y., Huang, C., Xie, X., Tan, B., Kamal, S., Xiong, X., 2018. Multilayer hybrid deep-learning method for waste classification and recycling. Comput. Intell. Neurosci. 2018.
Cullen, J.M., Allwood, J.M., 2013. Mapping the global flow of aluminum: From liquid aluminum to end-use goods. Environ. Sci. Technol. 47, 3057–3064.
Dang, T.L., Cao, T., Hoshino, Y., 2019. Classification of Metal Objects Using Deep Neural Networks in Waste Processing Line. ICIC Int 15, 1901–1912.
European Aluminum, 2019. VISION 2050 - European Aluminium [WWW Document]. URL https://www.european-aluminium.eu/vision-2050/(accessed 7.21.20).
Gaustad, G., Olivetti, E., Kirchain, R., 2012. Improving aluminum recycling: a survey of sorting and impurity removal technologies. Resour. Conserv. Recycl. 58, 79–87.
Gurell, J., Bengtson, A., Falkenström, M., Hansson, B.A.M., 2012. Laser induced breakdown spectroscopy for fast elemental analysis and sorting of metallic scrap pieces using certified reference materials. Spectrochim. Acta Part B At. Spectrosc. 74, 46–50.
Haraldsson, J., Johansson, M.T., 2018. Review of measures for improved energy efficiency in production-related processes in the aluminium industry–From electrolysis to recycling. Renew. Sustain. Energy Rev. 93, 525–548.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.
Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708.
Iandola, F.N., Han, S., Moskewicz, M.W., Ashraf, K., Dally, W.J., Keutzer, K., 2016. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. ArXiv Prepr. ArXiv160207360.
Konovalenko, I., Maruschak, P., Brezinová, J., Viňáš, J., Brezina, J., 2020. Steel surface defect classification using deep residual neural network. Metals 10, 846.
Koyanaka, S., Kobayashi, K., 2010. Automatic sorting of lightweight metal scrap by sensing apparent density and three-dimensional shape. Resour. Conserv. Recycl. 54, 571–578.
Krizhevsky, A., 2014. One weird trick for parallelizing convolutional neural networks. ArXiv Prepr. ArXiv14045997.
LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86, 2278–2324.
Li, W., Breier, M., Merhof, D., 2016. Recycle deep features for better object detection. ArXiv Prepr. ArXiv160705066.
Li, Z., Hoiem, D., 2017. Learning without forgetting. IEEE Trans. Pattern Anal. Mach. Intell. 40, 2935–2947.
Lorsakul, A., Suthakorn, J., 2007. Traffic sign recognition using neural network on OpenCV: toward intelligent vehicle/driver assistance system. In: 4th International Conference on Ubiquitous Robots and Ambient Intelligence. Citeseer, pp. 22–24.
Mesina, M.B., De Jong, T.P.R., Dalmijn, W.L., 2005. Scrap stainless steel detection using a pulsed electromagnetic field. Int. J. Miner. Process. 76, 21–31.
Modaresi, R., 2015. Dynamics of aluminum use in the global passenger car system: challenges and solutions of recycling and material substitution.

Mordvintsev, A., Abid, K., 2014. Opencv-python tutorials documentation. Obtenido Httpsmedia Readthedocs Orgpdfopencv-Python-Tutroalslatestopencv-Python-Tutroals Pdf.
Mwangi, H.W., Mokoena, M., 2019. Using deep learning to detect polyethylene terephthalate (PET) bottle status for recycling. Glob. J. Comput. Sci. Technol.
Ophoff, T., Van Beeck, K., Goedemé, T., 2019. Exploring RGB+ Depth fusion for real-time object detection. Sensors 19, 866.
Paraskevas, D., Kellens, K., Dewulf, W., Duflou, J.R., 2015. Environmental modelling of aluminium recycling: a Life Cycle Assessment tool for sustainable metal management. J. Clean. Prod. 105, 357–370.
Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., 2019. Pytorch: An imperative style, high-performance deep learning library. ArXiv PreprArXiv191201703.
Perez, L., Wang, J., 2017. The effectiveness of data augmentation in image classification using deep learning. ArXiv PreprArXiv171204621.
Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., 2015. Imagenet large scale visual recognition challenge. Int. J. Comput. Vis. 115, 211–252.
Sanchez, J., Soltani, N., Chamarthi, R., Sawant, A., Tabkhi, H., 2018. A novel 1d-convolution accelerator for low-power real-time cnn processing on the edge. 2018 IEEE High Performance Extreme Computing Conference (HPEC). IEEE, pp. 1–8.
Schlesinger, M.E., 2013. Aluminum recycling. CRC press.
Schwarz, M., Schulz, H., Behnke, S., 2015. RGB-D object recognition and pose estimation based on pre-trained convolutional neural network features. 2015 IEEE International Conference on Robotics and Automation (ICRA). IEEE, pp. 1329–1335.
Shao, L., Cai, Z., Liu, L., Lu, K., 2017. Performance evaluation of deep feature learning for RGB-D image/video classification. Inf. Sci. 385, 266–283.
Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. ArXiv Prepr. ArXiv14091556.
Soo, V.K., Peeters, J., Paraskevas, D., Compston, P., Doolan, M., Duflou, J.R., 2018. Sustainable aluminium recycling of end-of-life products: A joining techniques perspective. J. Clean. Prod. 178, 119–132.
Suzuki, S., 1985. Topological structural analysis of digitized binary images by border following. Comput. Vis. Graph. Image Process. 30, 32–46.
The Economist, 2007. The price of virtue [WWW Document]. URL https://www.economist.com/leaders/2007/06/07/the-price-of-virtue (accessed 7.20.20).
Wang, C., Liu, S., Zhang, J., Feng, Y., Chen, S., 2017. RGB-D Based Object Segmentation in Severe Color Degraded Environment. CCF Chinese Conference on Computer Vision. Springer, pp. 465–476.
Werheit, P., Fricke-Begemann, C., Gesing, M., Noll, R., 2011. Fast single piece identification with a 3D scanning LIBS for aluminium cast and wrought alloys recycling. J. Anal. At. Spectrom. 26, 2166–2174.
Wong, S.C., Gatt, A., Stamatescu, V., McDonnell, M.D., 2016. Understanding data augmentation for classification: when to warp? 2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA). IEEE, pp. 1–6.
Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K., 2017. Aggregated residual transformations for deep neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1492–1500.
Zhang, J., Yang, X., Song, G.-M., Chen, T.-Y., Zhang, Y., 2015. Relative orientation and position detections based on an RGB-D sensor and dynamic cooperation strategies for jumping sensor nodes recycling. Sensors 15, 23618–23639.