Journal of Sustainable Metallurgy
https://doi.org/10.1007/s40831-025-01172-6

RESEARCH ARTICLE

Computer Vision-Based Aluminum Scrap Grade Classification and Detection for Upcycling

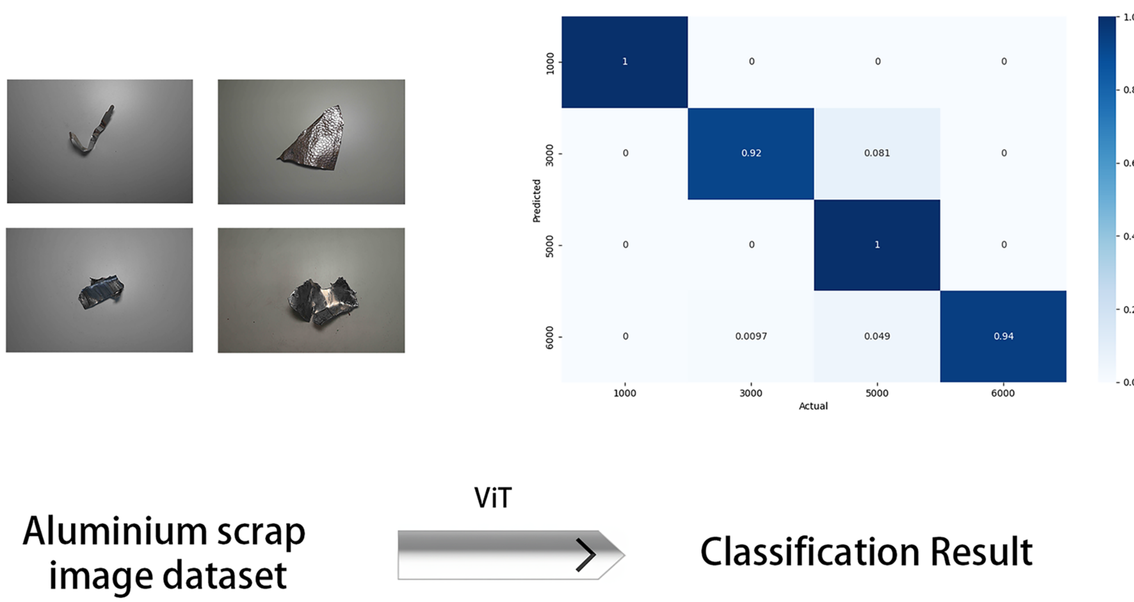
Yijun Quan[1] · Matthew Dunn[2] · Giovanni Montana[1] · Zushu Li[1]

Received: 11 September 2024 / Accepted: 11 June 2025

Abstract

The mechanical properties of aluminum alloys vary significantly across different grades due to differences in their chemical compositions. Therefore, when aluminium scraps are properly sorted during recycling, they can be more easily returned to their specific applications and yield a considerably higher value compared to unsorted scraps. Using computer vision for aluminum scrap sorting could potentially enhance the efficiency of the recycling process. In this work, we test and verify whether computer vision-based methods can be successfully deployed for commercial aluminum scrap grade classification using images only, on both the pre-shredding and post-shredding scraps. Image datasets of both pre-shredding and post-shredding scraps are carefully curated. The experimental results and the feature analysis show that the computer vision-based methods can classify the post-shredding scraps with high accuracy (98.4%) using the traces left by the shredding process, showing the potential of applying computer vision-based classification for accurate commercial aluminum sorting.

Graphical Abstract



Keywords Image classification · Scrap sorting · Feature analysis · Computer vision-based aluminum scrap separation · Circular economy

The contributing editor for this article was Hiromichi Takebe.

Extended author information available on the last page of the article

Introduction

As a fundamental material extensively used across various manufacturing sectors, aluminum is also highly recyclable, with 75% of aluminum produced historically still in circulation today [1, 2]. Recycling aluminum can save over 90% of the energy required to produce virgin aluminum from bauxite ores [3]. Therefore, advancements in aluminum recycling would greatly contribute to achieving environmental sustainability goals.

Different aluminum alloy families differ in chemical compositions, including elements, such as copper, magnesium, silicon, manganese, and zinc. These compositional differences result in distinct mechanical properties, making specific alloys better suited to particular applications. When recycled aluminum scrap is sorted by alloy grade, it can be more effectively reused in its original applications. However, if different alloy grades are mixed without sorting, the resulting scrap blend has an imprecise and inconsistent chemical composition. Producing secondary aluminum from such mixtures typically requires either the addition of virgin aluminum to dilute contaminants or the use of extra refining steps, both of which can be energy intensive. Producing virgin aluminum from bauxite ore has a global average energy consumption of about 20 kWh/kg, accounting for both the bauxite ore refining and aluminum smelting process [4, 5]. The secondary aluminum refining steps can demand substantial energy depending on the method employed. In contrast, if the scrap composition is more consistent with the requirements of the intended product, it can be reused through remelting, which consumes significantly less energy, around 2.8 kWh/kg. Therefore, effective sorting of aluminum alloys helps preserve energy and prevent downcycling, bringing both economic and environmental benefits.

Existing methods developed for aluminum scrap sorting can be mainly characterized into three groups: 1. Eddy Current Separation (ECS)-based methods [6–8], 2. Laser-Induced Breakdown Spectroscopy (LIBS)-based methods [9–11], and 3. X-ray Fluorescence (XRF)-based methods [12, 13]. Each of these methods presents its own set of challenges. ECS-based methods employ a changing magnetic field to sort the scraps. Different aluminum grades, due to their varying conductivity and density, could be thrown out at different distances for separation. However, the effectiveness of the ECS-based methods may be restrained by the shape and size of the scraps, leading to incorrect sorting and reduced purity. Both LIBS- and XRF-based methods are costly to implement at an industrial scale, making them difficult to be economically justified for recyclers to adopt [14]. These methods also typically require surface preparation to expose a clean section of the scrap surface for accurate analysis, which is impractical for processing the often dirty and irregular commercial scrap in high volumes. Consequently, despite their potential for precise composition analysis, the high cost, additional operational requirements, and limited throughput make them economically less viable, often even less attractive than downcycling unsorted scrap for commercial recyclers. Therefore, to promote wider adoption of scrap sorting and reduce the energy and resource demand associated with secondary aluminum production, there is an urgent need for a more cost-effective and scalable sorting method.

Given the latest advancements in computer vision, particularly in image classification, segmentation, and object detection, computer vision-based artificial intelligence techniques find widespread application in various industrial settings [15–17]. Compared to more specialized sensors, many computer vision-based methods rely on relatively simple setups using low-cost digital cameras, making the overall system more accessible and economically viable. As a result, computer vision approaches are increasingly contributing to cost reduction and improved efficiency in various manufacturing processes. Given these strengths, computer vision holds strong potential as a low-cost solution for aluminum scrap grade sorting, providing an accessible technology to enable the upcycling of aluminum scrap. Research has been done to use computer vision or combine computer vision with other magnetic sensors for cast and wrought aluminum scrap separation [18, 19] and classification of aluminum grades with samples prepared in a laboratory environment [20]. These studies have demonstrated the successful deployment of neural network-based methods on aluminum scrap sorting. However, success achieved with clean, controlled samples in a lab environment or in tasks focused on cast-wrought separation does not directly answer the question of whether using computer vision and neural networks is effective to classify commercial aluminum scrap under real industrial conditions.

To answer this question, we use this work to evaluate the feasibility of using computer vision to classify both pre-shredding and post-shredding commercial aluminum scraps. As the terminology implies, pre-shredding scrap refers to material that has not yet undergone the shredding process, whereas post-shredding scrap has passed through a shredder. Exceptions to this classification are discussed in more detail in Sect. 3. For this study and to support future research, we carefully curated image datasets of pre-shredding and post-shredding aluminum scraps of various grades, sourced directly from an industrial setting. Using this dataset, we evaluated the performance of various neural network architectures in classifying aluminum scrap by alloy class. Our experimental results demonstrate that neural networks can achieve high classification accuracy for post-shredding scrap using only standard RGB images captured with off-the-shelf digital cameras. Feature map analyses reveal that the models leverage both surface and edge characteristics of the scrap, such as texture, material folds, and cutting marks introduced by the shredder, as discriminative features. These observations suggest that the models are capable of extracting meaningful cues that correlate with underlying differences in alloy properties, such as mechanical strength and deformation behavior. These findings confirm the potential of computer vision as a low-cost, scalable approach to alloy classification in aluminum scrap sorting. To further enable advancements in this area, we also created a multi-object dataset consisting of images containing mixed-grade aluminum scraps. This dataset is designed to support object detection and instance segmentation tasks, which are essential for developing fully automated, vision-based sorting systems. Together, these contributions lay a strong foundation for future work in deploying computer vision technologies for efficient, grade-based aluminum recycling.

The rest of the paper is organized as follows: Sect. 2 revises the related works on computer vision methods and their application in waste recycling. Section 3 presents the setup of our data acquisition pipeline and the details of the curated datasets. The experiment and the results are presented in Sect. 4 followed by a discussion in Sect. 5. A conclusion is summarized in Sect. 6.

Background

Image classification is a well-known and extensively researched task in computer vision. It involves identifying the class label of an object using visual cues, such as the features of an object class. Traditionally, these features were extracted by methods designed manually, requiring human designers to have expert knowledge of the task to identify the relevant information about each object class, which is not always possible. However, neural networks can extract image features automatically [21] without the need for complex, handcrafted class-specific feature designs. Instead, improvements in algorithm design now focus more on the architecture of neural networks, enabling them to better explore the dataset and uncover the underlying statistical patterns related to each object class. This broadens the range of tasks to which image classification can be applied.

Over the years, several notable deep neural network architectures have been proposed. VGG-Net [22] uses multiple very small convolutional filters (3×3) to extract features from a very deep structure. To address the problem of gradient vanishing in deep neural network training, He et al. [23] proposed the Residual Neural Networks (ResNet) by introducing the skip connections between convolutional layers. Inspired by the recent success of the transformer architecture with its attention mechanism [24] in natural language processing, Dosovitskiy et al. proposed a Vision Transformer (ViT) architecture [25] for computer vision tasks by deploying the attention mechanism on the patches divided from an image. The above-mentioned architectures show promising performance on various image classification tasks, including the tests on public datasets like ImageNet [26], CIFAR-100 [27], and Oxford Flowers-102 [28]. This demonstrates that deep neural network-based classifiers can generalize effectively across different tasks. Advances in this research field have significantly enhanced the applicability and performance of deep neural network-based image classification in various real-life applications.

This improved performance has also drawn the interest of researchers in waste recycling. Compared to other specialized sensor-based waste separation techniques, camera-based computer vision methods offer a simpler setup and often a more cost-effective solution. These methods have the potential for large-scale deployment, with most classification algorithms capable of running in real time. In order to facilitate these applications, dedicated image datasets corresponding to the material in interest are required to train the neural networks. For example, to enable the classification of drinking waste, the Drinking Waste dataset [29] collects more than 4800 images of 4 different types of drinking waste, namely aluminum cans, glass bottles, PET bottles, and HDPE containers. To provide a benchmarking platform for household waste classification, Majchrowska et al. [30] combine 6 publicly available image datasets to form a dataset with 21000 waste instances of 8 different categories following the recycling rules in the city of Gdansk in Poland. These dedicated datasets enable researchers in the corresponding field to extensively develop and test computer vision-based algorithms, potentially improving recycling efficiency.

Aluminum is a crucial material in recycling, with rising demand, but it presents increasing challenges for effective recycling [31], highlighting the need for improved sorting and processing methods [32]. In response to these challenges, recent research has explored the application of computer vision-based classification techniques to improve metal scrap recycling and support more efficient downstream manufacturing processes. A few successful applications of computer vision systems have been demonstrated in the context of steel scrap recycling. Gao et al. [33] treat the steel scrap thickness estimation as a classification problem by constructing a categorical steel scrap image dataset with RGB-D images. The 3440 images from this dataset are categorized into 4 different thickness ranges: ≤ 5 mm, 5–9 mm, 9–12 mm, and ≥ 12 mm. Gao et al. managed to achieve an F1-score of 74% using a feature extractor designed for the point cloud data. Xu et al. [34] also conducted classification on steel scraps but with 2D images and the steel scraps are categorized based on the factors influencing the scrap purchase prices. An overall accuracy of 92.4% is achieved in the steel scrap classification task by applying deep neural networks to images from both laboratory environments and steel mills.

While significant progress has been made in steel scrap classification, computer vision-based classification for aluminum scrap remains relatively underexplored. Diaz-Romero et al. [18] evaluated the performance of various convolutional neural network architectures on the classification of the cast and wrought aluminum scraps. Further to using computer vision-based techniques only, Williams et al. [19] combine magnetic induction spectroscopy with machine vision to perform cast and wrought classification. Given that aluminum alloys are categorized by chemical composition and mechanical properties, grade-level classification is highly desirable for scrap upcycling. However, the current study has only been done in the lab environment with Huang et al. [20] used a ResNet18-based classifier for lab-made aluminum blocks of three grades (1060, 5052, and 6061). While the experimental results on the laboratory-prepared samples are promising, the performance of deep neural networks in classifying aluminum scrap grades from commercially or industrially sourced materials remains largely unexplored

Aluminum Scrap Dataset

In order to investigate the feasibility and conduct computer vision-based commercial aluminum scrap classification for both post-shredding and pre-shredding scraps, corresponding image datasets need to be curated. This was done in collaboration with Total Metal Recovery Ltd., a local aluminum recycler in Worcestershire, UK. The recycler handles both pre-consumer and post-consumer scrap, ensuring the collected samples are representative of typical industrial and commercial waste streams. Scrap samples were obtained through the recycler's routine operations, providing a realistic reflection of materials commonly found in the recycling industry. Each collected item was analyzed using a handheld XRF device to determine its composition and assign it to the appropriate alloy series. Three distinct image datasets were curated: the Pre-Shredding Dataset, the Post-Shredding Dataset, and the Mixed Post-Shredding Dataset. An overview of these datasets is provided in Table 1. All images were captured at a resolution of 4000 × 6000 pixels and saved in JPEG format.

For post-shredding aluminum scraps, we curated samples from the 1000, 3000, 5000, and 6000 series. Notably, the 1000 series scraps in this dataset are not processed by a shredder. This is because the 1000 series scraps commonly found at scrap yards have more uniform shapes due to their typical usage. Most of these scraps have sizes and weights similar to the post-shredding scraps from other alloy series, allowing for effective sorting and recycling without the need for shredding. For this study, we treat them together with the post-shredding scraps from other series. The post-shredding scraps were sourced from a range of origins, including both industrial and domestic waste streams, such as automotive parts and construction materials. These samples were collected directly from the shredder outputs at a recycling facility. All post-shredding samples were relatively compact, with no dimension exceeding 20 cm, consistent with standard shredded scrap sizes in industrial circulation. This small form factor is well suited for handling by automated sorting systems, such as air jets or robotic arms. The ability to accurately classify post-shredding aluminum scrap opens up the potential for integrating automated sorting solutions directly into existing shredding and processing pipelines.

The images of the post-shredding samples were taken in a controlled environment. Each scrap was placed against a uniform background and illuminated by an adjustable LED light with a fixed color temperature of 5000 Kelvin. A Canon EOS R50 camera with a fixed 16-mm F2.8 lens was used to capture the images. The camera was mounted on a fixed tripod, positioned directly above the scraps. Consistent exposure settings including aperture size and ISO speed settings were used for all images, ensuring uniform lighting conditions across the dataset. 1000 images for each aluminum grade series

Table 1 The summary of the pre-shredding and post-shredding aluminum scrap image datasets curated for this study

	Aluminum series	Condition	Total number of images
Post-shredding dataset	1000 3000 5000 6000	Indoor and controlled lighting	4000
Pre-shredding dataset	3000 5000 6000	Outdoor and handheld camera	900

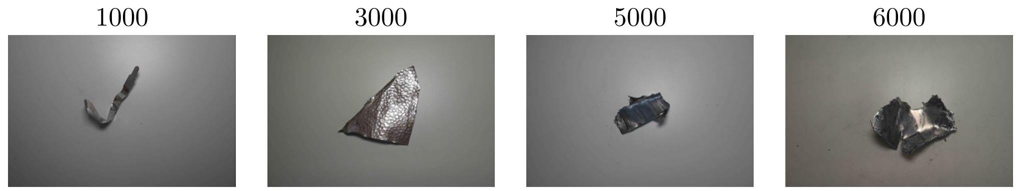


Fig. 1 Example images of the post-shredding scraps included in our dataset. The set of images is taken with a uniform background under controlled lighting conditions

are included in the dataset. Some example images of the scraps curated for this dataset are shown in Fig. 1.

For pre-shredding scraps, we collected samples from the 3000, 5000, and 6000 series, sourced from a variety of domestic and industrial waste streams, including gas cylinders, vehicle parts like engine casings, and aluminum cladding from construction materials. Due to the large and irregular sizes of these items, images were captured directly at the scrap yard before any processing. Scraps were randomly selected from raw piles immediately after unloading from transport vehicles. Each piece was placed on clean ground to ensure isolated imaging, with no overlapping materials. Photographs were taken outdoors using a Canon EOS R50 camera with a 16-mm lens, handheld in auto-exposure mode to maintain consistent brightness across varying lighting conditions. In total, 900 images were collected across the three alloy classes. Sample images are shown in Fig. 2. Since pre-shredding scraps are still in large, unprocessed forms, accurate alloy classification at this stage could enable earlier and more efficient segregation within the recycling process.

Experiment

The meticulously curated datasets offer a platform to benchmark the performance of various deep neural networks on aluminum scrap. To assess the effectiveness of aluminum scrap grade classification, we train and test deep neural network-based image classifiers using the Post-shredding Dataset and the Pre-shredding Dataset.

Neural Network Models

Three neural network architectures are tested on both datasets: VGG-Net ([22]), ResNet ([23]), and ViT ([25]). Specifically, we used VGG16 and ResNet18, the 16-layer and 18-layer versions of VGG-Net and ResNet, respectively. For the ViT classifier, we used the ViT16 model, which partitions the input image into multiple 16×16 patches and applies the attention mechanism to them. Despite being the lightweight versions of their respective architecture families, these models demonstrated satisfactory classification performance on public datasets like CIFAR-100 [27]. The output dimension of the last fully connected layer for the three networks is set to match the number of classes for each corresponding classification task: 3 for the pre-shredding scrap classification and 4 for the post-shredding scrap classification.

Image Processing

All three networks take images of size $224 \times 224 \times 3$ as inputs. Since the original images from the dataset have a resolution of 4000×6000 pixels, we first resize them to 768×512 pixels to preserve most of the image details. Then, a 384×384 pixel section is randomly cropped from the resized images. Random cropping is applied after resizing to ensure that each cropped image includes a portion of the scrap. If cropping were done before resizing, there is a risk that the resulting image would contain only background. The random cropping operation not only reduces the input image size but also serves as a crucial data augmentation step. It allows the neural networks to explore different parts of the scrap to extract features, while the cropped image retains the semantic content of the original input. Finally, the image

Fig. 2 Example images of the pre-shredding scraps included in our dataset

size is further reduced from 384 × 384 to 224 × 224 through resizing, in order to match the input size requirements of the neural network models.

To enhance model generalization and mitigating the risk of overfitting, standard image augmentation techniques were applied to the training data, including random horizontal and vertical flipping, as well as random color jittering, which involves randomly adjusting the brightness, contrast, saturation, and hue levels of the input images. For random horizontal and vertical flipping, we applied a 50% chance of flipping along each axis independently. For color jittering, we use the `ColorJitter()` function from PyTorch's `torchvision.transforms` module, randomly adjusting the brightness, contrast, and saturation of the images within the range of 0.95 to 1.05. For hue adjustment, we apply a random jitter within the range of −0.1 to 0.1. These augmentations help expose the model to a wider variety of visual conditions, making the trained networks more robust and reducing the risk of overfitting. After augmentation, the images are normalized using the standard ImageNet [26] mean and standard deviation values [0.485, 0.456, 0.406] and [0.229, 0.224, 0.225] for the RGB channels, respectively, following common practice. For the training setup, 80% of the images from both datasets were randomly selected to form the training dataset, while the remaining 20% were split into validation and test sets equally.

Experiment Setups

We trained the three networks using Cross-Entropy as the loss function. A batch size of 256 was used for VGG-Net and ResNet for loading the training data while the ViT was trained with a batch size of 32 images. The Root Mean Squared Propagation (RMSProp) algorithm was used as the optimizer, with learning rates set to 10^{-5}, 10^{-4}, and 3×10^{-4} for VGG-Net, ResNet, and ViT, respectively. The learning rates are fine-tuned, and the chosen values are reported based on the experimental results. Each network was trained for 100 epochs on the training datasets for the pre-shredding and post-shredding tasks. The experiment was implemented using Python 3.7 and PyTorch 1.12.0 library. The experiment was conducted on a desktop computer equipped with an Nvidia 3090 GPU with 24GB of VRAM.

Evaluation Indicators

We use Precision, Recall, and F1-score for each class as the metrics for evaluation. The Precision (P), Recall Rate (R) and F1-score are defined as follows:

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \quad , \tag{1}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \quad , \tag{2}$$

$$\mathrm{F1} = 2 \times \frac{P \times R}{P + R} \quad , \tag{3}$$

where TP, FP, and FN are true positives, false positives, and false negatives, respectively.

Results and Feature Analysis

The precision, recall, and F1-scores of the classification results for the post-shredding and pre-shredding aluminum

Table 2 Classification results for the post-shredding aluminum scraps using VGG16, ResNet18, and ViT16 image classifiers

class	Precision			Recall			F1-score		
	VGG	ResNet	ViT	VGG	ResNet	ViT	VGG	ResNet	ViT
1000	0.985	1	1	0.970	0.995	1	0.977	0.997	**1**
3000	1	0.976	1	0.891	0.745	0.919	0.942	0.845	**0.958**
5000	0.966	0.854	0.904	0.966	0.989	1	**0.966**	0.917	0.950
6000	0.924	1	1	1	1	0.942	0.961	**1**	0.970

The highest F1-scores are shown in bold

Table 3 Classification results for the pre-shredding aluminum scraps using VGG16, ResNet18, and ViT16 image classifiers

class	Precision			Recall			F1-score		
	VGG	ResNet	ViT	VGG	ResNet	ViT	VGG	ResNet	ViT
3000	0.581	0.639	0.694	0.862	0.793	0.862	0.694	0.708	**0.769**
5000	0.591	0.593	0.471	0.464	0.571	0.571	0.520	**0.582**	0.516
6000	0.654	0.679	0.667	0.500	0.559	0.412	0.567	**0.613**	0.509

The highest F1-scores are shown in bold

scrap grade classification tasks are reported in Tables 2 and 3.

From the classification results on the post-shredding scraps, we can observe that all three classifiers generally perform well for aluminum scraps from the 1000 series. This is because the scraps in this series have more uniform shapes, which makes it easier for the classifiers to extract visual features corresponding to the class. However, the 3000 series poses the most difficulties for the classifiers among the other three classes. While this class achieves perfect precision with VGG-Net and ViT classifiers, the recall rates are generally lower than the results from other classes, indicating that the classifiers tend to misidentify 3000 series scraps as scraps from other series. This indicates the scraps from the 3000 series may share some common features with the 5000 and 6000 series scraps. Nonetheless, recall and precision rates of near or above 90% are observed for all classes, demonstrating the effectiveness of computer vision-based image classification for aluminum scrap. Among the three tested classifiers, ViT achieves the best overall performance in terms of F1-scores and it has an average accuracy of 98.3% for the classification task. Compared to convolutional neural networks, the ViT model divides images into multiple small patches and applies the attention mechanism to all these patches. This approach enables the model to capture both local features within each patch and overall features through the attention mechanism. This could explain why the ViT model performs slightly better than VGG-Net and ResNet in the post-shredding aluminum classification task. The confusion matrix of the classification results from ViT is shown in Fig. 3.

Compared to the classification results for the post-shredding scraps, the performance of the three classifiers on pre-shredding scraps is less satisfactory. Both precision and recall rates are significantly lower for pre-shredding scraps, except for the recall rate of the 3000 series. This higher precision is due to the fact that most 3000 series samples in the dataset are from construction wastes or radiator fins commonly found at scrapyards. These scraps have distinct features that the classifier can recognize with high confidence, resulting in high recall rates for these scraps. However, the 3000 series also includes a variety of scraps with different features, and similar to the result from post-shredding scraps, these features may be similar to the ones shown in scraps of the 5000 and 6000 series. This leads to misidentification which slightly lowers the recall rate as the classifiers struggle to recognize other scraps of this series.

Despite this, the classification results still surpass the chance level, with the confusion matrix of classification results from ViT shown in Fig. 4. This indicates that deep learning methods managed to extract some distinctive features that can be used to distinguish scraps of different series. However, the diverse uses and the irregular shapes of these scraps make it challenging for the classifiers to generalize these features effectively. In contrast, post-shredding scraps have more uniform shapes within each class, making them easier to recognize. A larger dataset of pre-shredding scraps could help the classifiers learn more features related to each class and improve classification results. However, this would require more support from aluminum recyclers, as each recycler may be limited to their local supplies.

In addition to evaluating the classification results, we also visualize the VGG-Net classifier's feature map to understand how the image classifiers extract features to achieve high performance in post-shredding scrap classification. Example input images and two of their corresponding feature maps are shown in Fig. 5. Bright colors highlight the areas where features are present.

The two Feature Map 1 shown in Fig. 5 reveal that the VGG-Net classifier places significant emphasis on regions of the scrap surface that exhibit higher feature activation responses, as indicated by brighter colors. These areas

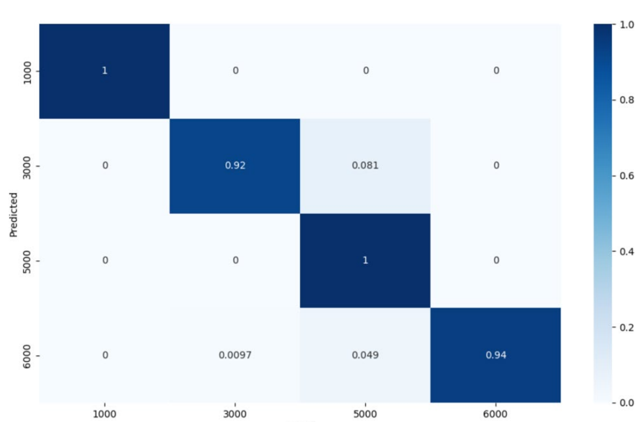


Fig. 3 Confusion matrix for the post-shredding aluminum scrap grade classification

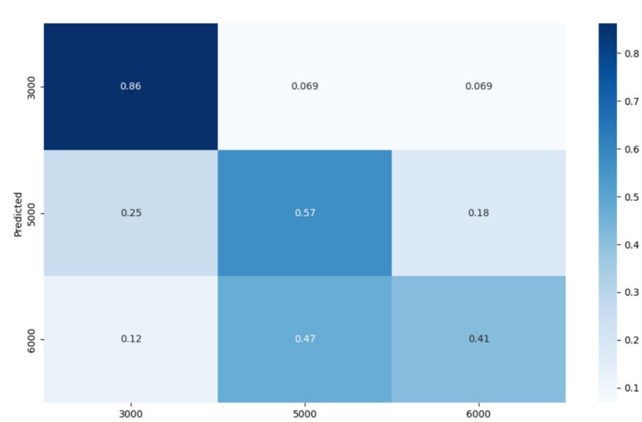


Fig. 4 Confusion matrix for the pre-shredding aluminum scrap grade classification

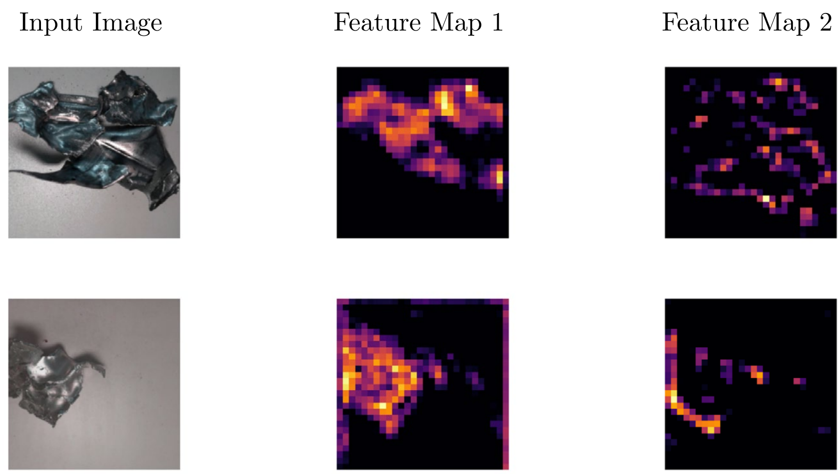


Fig. 5 CNN feature analysis of the post-shredding aluminum scraps of 5000 series (top) and 6000 series (bottom)

correspond to visible dents, deformations, or regions with complex surface geometry. Such patterns suggest that the network relies heavily on surface-related visual cues, most notably, texture variation, depth discontinuities, and material folding, for distinguishing between different types of aluminum scrap. These features likely arise from mechanical handling or the physical characteristics of the scrap itself and provide the model with rich, discriminative information.

In comparison, Feature Map 2 reveals that the model responds strongly to edge-related features, with the highlighted area shown around the edge of the scraps. These are presumed to be associated with shredder-induced marks. Such cutting marks tend to reflect underlying mechanical properties of the material, such as hardness, ductility, and yield strength. Because the scraps are processed using the same shredder, the differences in alloy composition manifest as distinct visual and structural patterns at the edges, providing the model with a secondary set of discriminative cues.

Together, these observations indicate that the classifier is highly sensitive to fine-grained morphologic characteristics. The combined use of surface texture and edge detail enables the model to perform accurate discrimination between aluminum alloys of different grades. This capacity is particularly important for the task of classification, where subtle variations in appearance, driven by different mechanical properties of the materials, must be detected from standard digital images. Ultimately, these results highlight the potential of deep learning-based computer vision systems to perform complex material classification tasks with a high degree of accuracy, even when working with scraps sourced directly from real-world commercial and industrial operations.

Discussion

While this work demonstrates the feasibility of using computer vision for grade-based classification of post-shredding aluminum scrap, significant challenges remain before such systems can be deployed in fully automated, industrial-scale scrap sorting applications. As a data-driven approach, the performance of deep learning models depends heavily on the quality, diversity, and representativeness of the training data. Although the samples we collected are representative of commercial aluminum alloys and reflect real-world variability collected at the time, they still represent a limited subset of the much broader range of conditions and material states encountered in industrial settings. Expanding the dataset to include more diverse samples, including different contamination levels and surface conditions, will be necessary to improve generalization and robustness.

While expanding and diversifying the dataset is the most straightforward and reliable path to improving model accuracy and resilience, we recognize that assembling a truly comprehensive dataset would likely require broad collaboration across the recycling and aluminum industries. Such large-scale data collection efforts can be both costly and logistically challenging.

However, recent advancements in machine learning offer promising alternatives, particularly as these methods are increasingly adapted for industrial applications. Techniques such as unsupervised and self-supervised learning, specifically tailored to industrial use cases [35, 36], as well as reinforcement learning with human feedback [37], provide opportunities to enhance model performance without relying exclusively on large volumes of annotated data. These approaches can help models learn more generalized representations from limited data or adapt to new conditions with minimal additional supervision. By integrating such strategies into future research, it may be possible to substantially

Table 4 The summary of the mixed post-shredding aluminum scrap image dataset

	Aluminum series	Number of pieces per image	Total number of images
Mixed post-shredding dataset	3000	6 or 7	707
	5000		
	6000		

improve the practicality and scalability of computer vision-based scrap classification systems while reducing the burden of data curation.

While we recognize that computer vision-based algorithms can effectively leverage distinct image features to classify aluminum alloys of different grades, this approach remains largely empirical and lacks the compositional insights provided by analytical sensors, such as XRF or LIBS. Therefore, a promising direction for future research is the integration of computer vision with analytical sensing technologies. This hybrid approach would combine the speed and efficiency of computer vision with the precise compositional analysis offered by sensors, ultimately enabling faster and more accurate aluminum scrap sorting.

Moreover, while our classification results are promising, they represent only an initial step toward practical deployment. In real-world industrial environments, aluminum scrap is rarely presented as isolated, individual pieces. Instead, post-shredding scrap typically appears in bulk mixtures, with many pieces of varying grades overlapping or partially occluded. Consequently, it is essential to move beyond simple classification of individual samples toward more complex computer vision tasks such as object detection, instance segmentation, and multi-object tracking. These tasks would enable the identification and classification of individual scrap pieces within mixed batches, a crucial capability for automated sorting systems.

To support future work in this direction, we curated a multi-object dataset featuring images with multiple pieces of post-shredding scraps of mixed grades. This dataset was created to support future work on mixed scrap sorting. We mixed post-shredding scraps from the 3000, 5000, and 6000 series in each image, with the total number of scraps per image ranging from 6 to 7 pieces. These images were taken under controlled lighting conditions similar to those used for the single-piece scrap dataset. The camera was set up on a fixed tripod, and the scraps were placed against a uniform background. Each image is labeled with the number of instances for each aluminum series included. A total of 707 images, featuring 4643 instances of post-shredding scraps, are included in the dataset. A summary of the dataset is shown in Table 4 while some example images are shown in Fig. 6.

Fig. 6 Example image in the mixed scrap dataset: (left) 5 pieces of 6000 series scraps mixed with 1 piece of 3000 series scrap, (right) 1 piece of 5000 series scrap, and 1 piece of 6000 series scrap mixed with 5 pieces of 3000 series scraps

Conclusion

In this work, we investigated the feasibility of using computer vision-based image classification for commercial aluminum scrap grade classification. We constructed two datasets for the image classification task: one for pre-shredding aluminum scraps and another for post-shredding aluminum scraps. By evaluating three different neural network image classifier models, namely the VGG-Net, ResNet, and ViT, on both datasets, we observed that while classifier performance was suboptimal on pre-shredding scraps, the deep neural networks achieved high accuracy on post-shredding scraps, with ViT attaining 98.4% accuracy for the classification task. Feature analysis of the classification results suggests that texture and edge information significantly contribute to classification accuracy. These findings demonstrate the potential of computer vision-based image classification for aluminum scrap grade classification and sorting, which could significantly enhance recycling efficiency and reduce energy consumption in the recycling process. Although the experiments in this study are limited to aluminum scrap grade classification, a multi-object dataset was also created to support future research on sorting mixed aluminum scraps of different grades. Future work will focus on developing multi-object detection capabilities to facilitate the automated sorting of mixed aluminum scrap, as well as exploring unsupervised and self-supervised learning techniques to improve model performance while minimizing the reliance on extensive image annotation.

Funding We wish to acknowledge the support of the Henry Royce Institute for Advanced Materials through the Industrial Collaboration Program, funded by a grant provided by the Engineering and Physical Sciences Research Council (EPSRC) EP/X527257/1. The authors also would like to thank the support from EPSRC (EP/V011804/1, EP/S018107/1).

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Bertram M, Martchek KJ, Rombach G (2009) Material flow analysis in the aluminum industry. J Indus Ecol. https://doi.org/10.1111/j.1530-9290.2009.00158.x
2. Kvande H, Drabløs PA (2014) The aluminum smelting process and innovative alternative technologies. J Occup Environ Med 56:23–32
3. Green JA (2007) Aluminum recycling and processing for energy conservation and sustainability. ASM International, Ohio
4. Institute IA Metallurgical alumina refining energy intensity. https://international-aluminium.org/statistics/metallurgical-alumina-refining-energy-intensity/. Accessed 06 Sep 2025
5. Institute IA Primary aluminium smelting energy intensity. https://international-aluminium.org/statistics/primary-aluminium-smelting-energy-intensity/. Accessed 06 Sep 2025
6. Kercher S, Webb M (1982) Scrap processing by eddy current separation techniques. Res Conserv 8(1):61–74
7. Wang D, Ma X, Zhi X, Zhang S (2013) Research review of scrap metals eddy current separation technology. Sens Trans 158(11):242
8. Ruan J, Qian Y, Xu Z (2014) Environment-friendly technology for recovering nonferrous metals from e-waste: Eddy current separation. Resour Conserv Recycl 87:109–116
9. Sattler H, Yoshida T (1993) New sorting system for recycling of magnesium and its alloys after use. In: First International Conference on Processing Materials for Properties, pp 861–864
10. Campanella B, Grifoni E, Legnaioli S, Lorenzetti G, Pagnotta S, Sorrentino F, Palleschi V (2017) Classification of wrought aluminum alloys by artificial neural networks evaluation of laser induced breakdown spectroscopy spectra from aluminum scrap samples. Spectrochim Acta, Part B 134:52–57
11. Díaz-Romero DJ, Eynde S, Sterkens W, Eckert A, Zaplana I, Goedemé T, Peeters J (2022) Real-time classification of aluminum metal scrap with laser-induced breakdown spectroscopy using deep and other machine learning approaches. Spectrochim Acta, Part B 196:106519
12. Krotkov M, Satayev I, Shepelev D (1993) Technology and equipment for X-ray radiometric sorting of secondary metallic raw materials. In: XVIII International Mineral Processing Congress, Sydney, Australia, pp 1271–1275
13. Kelly S, Apelian D (2016) Automotive aluminum recycling at end of life: a grave-to-gate analysis. Center for Resource Recovery and Recycling (CR3), Metal Processing Institute, Worcester Polytechnic Institute, 2016
14. Raabe D, Ponge D, Uggowitzer PJ, Roscher M, Paolantonio M, Liu C, Antrekowitsch H, Kozeschnik E, Seidmann D, Gault B et al (2022) Making sustainable aluminum by recycling scrap: the science of "dirty'' alloys. Prog Mater Sci 128:100947
15. Zhou L, Zhang L, Konz N (2022) Computer vision techniques in manufacturing. IEEE Trans Syst Man Cybernet: Syst 53(1):105–117
16. Kakani V, Nguyen VH, Kumar BP, Kim H, Pasupuleti VR (2020) A critical review on computer vision and artificial intelligence in food industry. J Agri Food Res 2:100033
17. Xu S, Wang J, Shou W, Ngo T, Sadick A-M, Wang X (2021) Computer vision techniques in construction: a critical review. Arch Comput Methods Eng 28:3383–3397
18. Díaz-Romero D, Sterkens W, Van den Eynde S, Goedemé T, Dewulf W, Peeters J (2021) Deep learning computer vision for the separation of cast- and wrought-aluminum scrap. Resour Conserv Recycl 172:105685
19. Williams KC, O'Toole MD, Peyton AJ (2024) Classification of wrought and cast aluminium using magnetic induction spectroscopy and machine vision. In: 2024 IEEE Sensors Applications Symposium (SAS), pp 1–6. https://doi.org/10.1109/SAS60918.2024.10636579
20. Huang B, Liu J, Zhang Q, Liu K, Li K, Liao X (2022) Identification and classification of aluminum scrap grades based on the resnet18 model. Appl Sci 12(21):11133
21. Lecun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11):2278–2324
22. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556
23. 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
24. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. Adv Neural Inform Proc Syst 30
25. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, et al (2020) An image is worth 16x16 words: transformers for image recognition at scale. arXiv preprint arXiv:2010.11929
26. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L (2009) Imagenet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp 248–255
27. Krizhevsky A (2009) Learning multiple layers of features from tiny images
28. Nilsback M-E, Zisserman A (2008) Automated flower classification over a large number of classes. In: 2008 Sixth Indian Conference on Computer Vision, Graphics and Image Processing, pp 722–729
29. Serezhkin A (2020) Drinking waste classification. Kaggle. https://www. kaggle. com/datasets/arkadiyhacks/drinking-wasteclassification. Accessed 06 June 2021
30. Majchrowska S, Mikołajczyk A, Ferlin M, Klawikowska Z, Plantykow MA, Kwasigroch A, Majek K (2022) Deep learning-based waste detection in natural and urban environments. Waste Manage 138:274–284
31. Billy RG, Müller DB (2023) Aluminium use in passenger cars poses systemic challenges for recycling and GHG emissions. Resources Cons Recycl 190:106827. https://doi.org/10.1016/j.resconrec.2022.106827

32. Van den Eynde S, Bracquené E, Diaz-Romero D, Zaplana I, Engelen B, Duflou JR, Peeters JR (2022) Forecasting global aluminium flows to demonstrate the need for improved sorting and recycling methods. Waste Manage 137:231–240. https://doi.org/10.1016/j.wasman.2021.11.019
33. Gao Z, Lu H, Lei J, Zhao J, Guo H, Shi C, Zhang Y (2023) An RGB-d-based thickness feature descriptor and its application on scrap steel grading. IEEE Trans Instrument Measure 72:1–14. https://doi.org/10.1109/TIM.2023.3328089
34. Xu W, Xiao P, Zhu L, Zhang Y, Chang J, Zhu R, Xu Y (2023) Classification and rating of steel scrap using deep learning. Eng Appl Artif Intell 123:106241
35. Fatima Z, Zardari S, Tanveer MH (2024) Advancing industrial object detection through domain adaptation: a solution for industry 5.0. Actuators. https://doi.org/10.3390/act13120513
36. Intxausti E, Skoaj D, Cernuda C, Zugasti E (2024) A methodology for advanced manufacturing defect detection through self-supervised learning on X-ray images. Appl Sci. https://doi.org/10.3390/app14072785
37. Ouyang L, Wu J, Jiang X, Almeida D, Wainwright C, Mishkin P, Zhang C, Agarwal S, Slama K, Ray A et al (2022) Training language models to follow instructions with human feedback. Adv Neural Inf Process Syst 35:27730–27744

Authors and Affiliations

Yijun Quan[1] · Matthew Dunn[2] · Giovanni Montana[1] · Zushu Li[1]

✉ Yijun Quan
yijun.quan@warwick.ac.uk

[1] WMG, University of Warwick, Coventry CV4 7AL, UK

[2] Total Metal Recovery Ltd., 138 Hanbury Road, Bromsgrove B60 4JC, UK