Material Measurement Units: Foundations Through a Survey

FEDERICO ZOCCO, Centre for Intelligent Autonomous Manufacturing Systems, Queen's University Belfast, Northern Ireland, UK
SEÁN MCLOONE, Centre for Intelligent Autonomous Manufacturing Systems, Queen's University Belfast, Northern Ireland, UK

Long-term availability of minerals and industrial materials is a necessary condition for sustainable development as they are the constituents of any manufacturing product. In particular, technologies with increasing demand such as GPUs and photovoltaic panels are made of critical raw materials. To enhance the efficiency of material management, in this paper we make three main contributions: first, we identify in the literature an emerging computer-vision-enabled material monitoring technology which we call Material Measurement Unit (MMU); second, we provide a survey of works relevant to the development of MMUs; third, we describe a material stock monitoring sensor network deploying multiple MMUs.

CCS Concepts: • **Applied computing** → **Physical sciences and engineering**; *Environmental sciences*; *Mathematics and statistics*.

Additional Key Words and Phrases: computer vision, machine learning, ecological automation, industrial ecology, circular economy

ACM Reference Format:
Federico Zocco and Seán McLoone. 0000. Material Measurement Units: Foundations Through a Survey. *J. ACM* 0, 0, Article 0 (0000), 19 pages. https://doi.org/00.0000/000000.00000000

1 PROBLEM INTRODUCTION

Modern society provides high quality life standards in developed countries while seeking to improve the conditions of developing areas [4]. At the same time, multiple and complex historical factors have contributed to human population growth [11]. Spreading welfare at large scale relies on the availability of raw materials needed to produce the goods and services supporting society (e.g. drugs, wind turbines, books, the electrical energy consumed to show this document in electronic format) [51]. The uncertainties about the long-term supply of critical raw materials have recently led to the *circular economy* concept, which aims to create manufacturing systems with minimal mineral extraction and minimal waste production [27]. To shift from linear to circular manufacturing, real-time data on materials stocks and flows need to be measured as accurately as possible to guide sustainable decisions [82, 83]. In particular, the goal of this paper is to devise a system for generating data/measurements to answer two sustainability questions: "What is the *amount* of a critical material currently accumulated in a region, e.g. a city or country? What *types* of raw material are accumulated?".

Our contributions:

Authors' addresses: Federico Zocco, Centre for Intelligent Autonomous Manufacturing Systems, Queen's University Belfast, Belfast, Northern Ireland, UK, fzocco01@qub.ac.uk; Seán McLoone, Centre for Intelligent Autonomous Manufacturing Systems, Queen's University Belfast, Belfast, Northern Ireland, UK, s.mcloone@qub.ac.uk.

0004-5411/0000/0-ART0 $15.00
https://doi.org/00.0000/000000.00000000

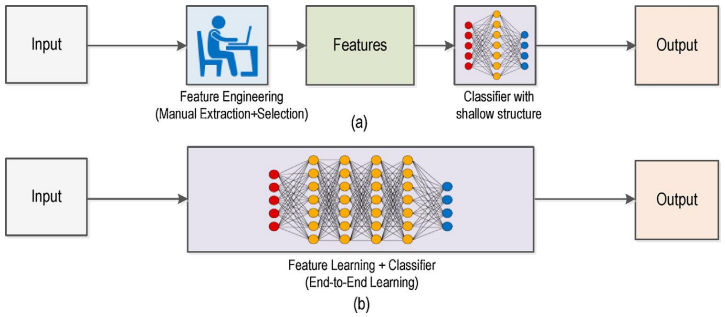

Fig. 1. {Taken from [111]} Main difference between (a) hand-crafted feature methods and (b) representation learning methods.

(1) we identify an emerging monitoring technology to measure material stocks;
(2) we provide a survey of work relevant to this development with an emphasis on its foundations;
(3) we provide a guide towards the design of a material monitoring sensor network.

The paper is organized as follows. Section 2 sets out the background; Section 3 defines the material measurement system and its technological foundations; Section 4 covers system implementation aspects. Finally, Section 5 provides some conclusions.

2 BACKGROUND: COMPUTER VISION

Computer vision is a research area concerned with making useful decisions about real physical objects and scenes based on sensed images [98]. It is a subfield of artificial intelligence and consists of designing a *signal-to-symbol converter* [79]: cameras provide signals (i.e. measurements) about the physical world and the computer vision model converts them into symbolic representations, e.g. the word/symbol "cat" if the image depicts a cat.

Research in computer vision spans more than forty years [106]. State-of-the-art techniques can be divided into two broad categories: hand-designed feature methods (i.e. classic machine learning) and representation learning methods (i.e. deep learning). For the foundations of the field, the reader should refer to well-established books such as [106] for an updated and general treatment, [41] for a focus on the first category of techniques and [46, 119] for a focus on the second category. This section provides the general concepts of the two broad categories.

The main difference between classic hand-crafted feature methods and representation learning methods is depicted in Fig. 1: the former requires an expert to design the algorithm/model capturing the characteristic features of the image of interest, the latter instead assume that the computer learns them during a training phase; as a consequence, the former requires expert domain knowledge and less computing resources than the latter. Given that a deep learning of representations works directly between the input images and the output symbols, the second category is also referred as an end-to-end learning approach.

Hand-crafted feature methods. One of the simplest methods is based on bag-of-words (BoW) [106], also called bag-of-features, which identifies a set of key words (i.e. features) for each image analogously to text documents that are described by the word content. Then, a test image is assigned to the class with the closest word/feature composition. One of the first BoW recognition systems was proposed in [24] and it is shown in Fig. 2. As visible, the feature extraction step of Fig. 1 is expanded there. The image patches are detected using Harris affine *detectors* [77], which in turn are used to compute the scale invariant feature transform (SIFT) *descriptors* [71]. The histogram of

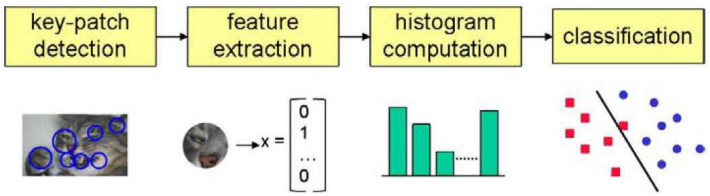

Fig. 2. {Taken from [25]} Typical processing sequence in a bag-of-words vision system.

visual words is used as input vector to a machine learning classifier. Other types of detectors and descriptors are compared in [120].

Bag-of-words models are the simplest because they do not consider the geometric relationships between different parts and features [106]. While this makes them particularly efficient, higher inference accuracy is provided by *part-based* models which focus on the geometric relationships between the constituent parts of the object [38]. More details on part-based modeling can be found in [39].

An approach even more accurate than part-based modeling takes into account the *context* in which the object with its constituent parts occur [86]. Combinations of part-based and context models in the same vision system have also been proposed [22, 104].

As visible in Fig. 2, the pipeline final block is the classifier. One of the simplest classification algorithms is the *k-nearest neighbors* which consists of finding the k training samples closest to the new sample and evaluating its class knowing the class of the neighbors [106]. A library with nearest neighbors algorithms for large training datasets is presented in [81].

The k-nearest neighbors is a non-parametric approach since it does not define a model of learned parameters from the training set. A simple parametric classification algorithm is *multiclass logistic regression* (despite the name, this method is not for regression), which learns a linear model and applies the *softmax* function to the model output to give the probability of having the class C^i given the input feature vector $\boldsymbol{x}$, i.e. $p(C^i|\boldsymbol{x})$ [15].

In some cases there are multiple possible surfaces that correctly divide the training samples into their classes. In these cases kernel *support vector machines* (SVMs) define the decision boundary as the one that maximizes the distance between the training set classes [15]. A survey of kernel-based methods for computer vision can be found in [63].

Another approach consists of using *decision trees* having a graph structure. The key idea of this approach is to divide the complex classification task in simpler tests that are hierarchically organized [106]. For example, assume we have an image of an outdoor garden, to classify it as "outdoor garden" the problem can be split in two subsequent steps: the first answering "Is there the sky at the top?" and, if true, the second answers "Is the bottom part green?" [23].

Representation learning methods. The most used computer vision methods belonging to this second category are based on *convolutional neural networks* (CNNs) [66, 119]. It consists of a network architecture designed to perform computations emulating multiple connected layers of neurons in a fashion similar to the neural network of a human brain. Training a CNN-based computer vision system requires four components: the training/test datasets, the network architecture, the training algorithm and the cost function to be optimized by the training algorithm. Moreover, by definition, a CNN has at least one neural layer that performs the *convolution* operation [46]. An example of a CNN architecture is depicted in Fig. 3: training images are given as input and, once the end-to-end learning is completed, the resulting model gives the conditional probabilities $p(C^i|\boldsymbol{x})$, where C^i is the i-th class (e.g. car, bicycle) and $\boldsymbol{x}$ is the input image. Therefore, the features are

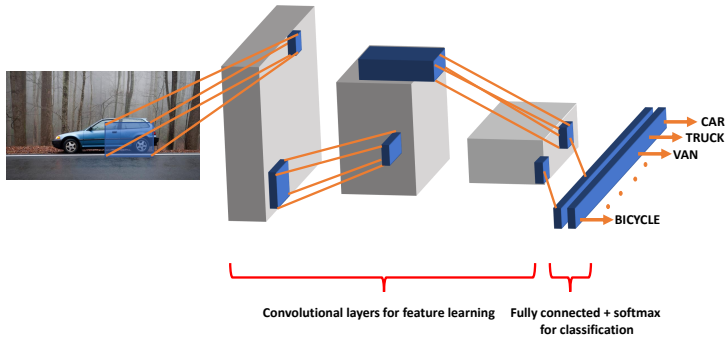

Fig. 3. Example of the architecture of a simple convolutional neural network: from the input image the features are learned with hierarchical levels of abstraction using multiple convolutional layers. Finally, the softmax layer converts the features into classes.

embedded into the model parameters defined through the optimization of the multi-modal cost function. A key reason that has made CNNs particularly successful for computer vision over other neural network architectures is that it accepts a two-dimensional input and, through convolutions, performs two-dimensional operations. Hence the pixels of the input image are processed preserving their original relative position.

Different CNN architectures have been proposed over the years [121]. For example, in [45] firstly several local regions of the input image are identified, secondly a large CNN learns the features of each local region, finally it classifies the content of each region using a linear SVM per class. Subsequently this CNN architecture has been sped-up in [44, 94].

The basic CNN can process images of arbitrary size with the output size of the convolutional layers (also called *feature maps*) being influenced by the image input size. As a consequence, a trained CNN may have the architecture suitable to process images of a fixed size (say 256×256), but not be adaptable to process images of a different size (say 128×128). Therefore, [55] proposes placing a layer after the last convolutional layer in order to make it possible for the network to process different image sizes.

As discussed about Fig. 3, typically the output of the model is the marginal probability $p(C^i|\boldsymbol{x})$. The *fully convolutional* network in [70] gives instead such a probability pixel-to-pixel, i.e. the output is a two-dimensional matrix giving the class of each pixel. Subsequently, the region-based classification approach cited above (i.e. [45]) has been combined in [28] with a fully convolutional network.

A more complete object classification system is proposed in [93], where for a given input image depicting a scene, the model infers both the class of every detected object and a bounding block around them to locate their spatial position. Specifically, the model divides the input image in a grid and, stating the problem as a regression task, for each grid cell it predicts multiple bounding boxes, the confidence for those boxes and the object class probabilities. The authors initially define a smaller CNN network and then, once pre-trained, they convert the model to perform detection adding four convolutional layers, two fully connected layers and increasing the input resolution of the network from 224×224 to 448×448. Subsequently, this algorithm has been sped-up and made more accurate in [69].

For a more comprehensive treatment of CNNs the reader should refer to [46, 119] for the basics, [121] for a review of several variants and [110] for a brief review of the most popular deep learning algorithms for computer vision (not just CNNs).

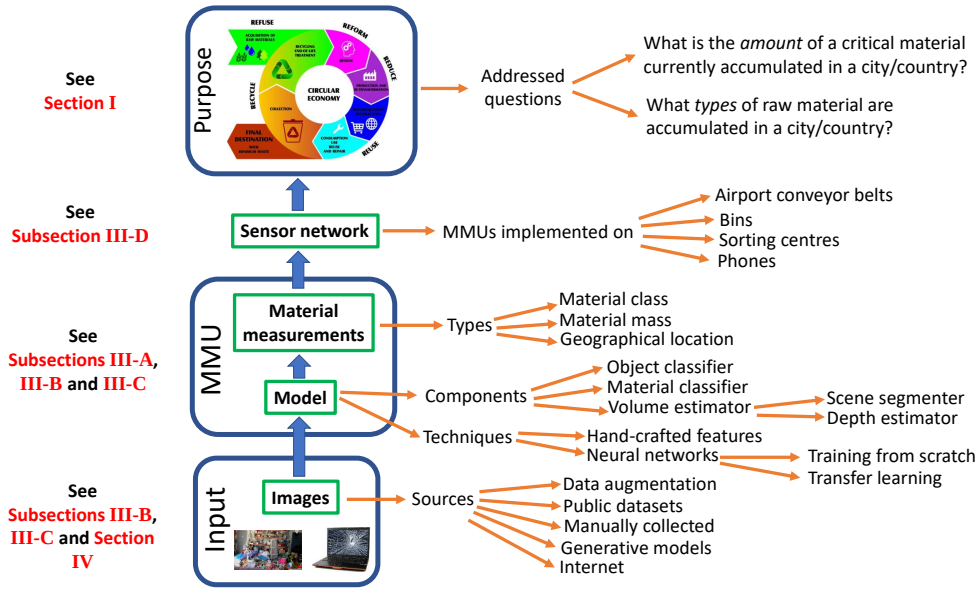

Fig. 4. Overview of the MMU system and its purpose: an MMU receives images as input and seeks to measure material stocks in a desired location. Multiple MMUs could be connected as a sensor network. The corresponding sections of the survey are indicated along the left side[a].

[a]Circular economy image sourced from:
https://www.portoprotocol.com/circular-economy-as-a-way-of-increasing-efficiency-in-organizations/

3 COMPUTER-VISION-ENABLED MEASUREMENTS

3.1 Material Measurement Unit

The monitoring system is called a Material Measurement Unit (MMU) and is defined as follows.

Definition 1. *An MMU is a complex sensor that, through a mathematical model, receives images of objects as input (e.g. RGB images, X-ray images, depth images) and provides as output information about the material composition of the object. The fundamental output measurements are (1) the class of material and (2) the mass of material. Briefly, it is a converter from object images to material measurements such as the class and the mass.*

Figure 4 provides an overview of the MMU context and purpose: the input to the system are "Images" as in the green rectangle at the bottom and they are collected or generated through the "Sources" indicated with the orange arrows on the right side; the images are processed by the MMU internal "Model" (i.e. the second green rectangle); the outputs of the model are "Material measurements" having the "Types" specified by the orange arrows; the fourth green rectangle mentions a "Sensor network", which is realized if multiple MMUs are implemented on different platforms and interconnected; the final "Purpose" of such a sensor network is monitoring the material stocks and flows for a more sustainable natural resources management.

An MMU can be seen as made up of three components, each one dedicated to a specific task:

(1) Component 1 for material recognition
(2) Component 2 for object recognition
(3) Component 3 for volume estimation.

Given that the density of a material is typically a known parameter [1], the volume estimation of a product evaluated by Component 3 permits estimation of the mass through $mass = density \times volume$. When processing an object with hidden parts such as a mobile phone, Component 1 would recognize the plastic of the case and the glass of the screen, but not the internal electronics, whereas Component 2 could recognize the specific model of a phone and read the corresponding full list of materials from a database. If instead waste plastic packaging is being processed, the complex shape of the damaged packaging makes Component 1 preferred to Component 2 because it focuses on the texture of plastic ignoring the unpredictable shape of the damaged object.

The remaining subsections of this section are ordered considering the scale-up of the system: Subsection 3.2 focuses on Component 1 covering works from the computer vision literature on material recognition; Subsection 3.3 adds Components 2 and 3; finally Subsection 3.4 proposes a distributed monitoring system exploiting multiple MMUs. Details on the topics covered by each subsection are captured on the left side of Fig. 4.

3.2 Computer Vision Focusing on Materials or Waste

One of the three components of an MMU is the material recognizer, hence here we discuss previous work on computer vision for material or waste recognition.

In [96] CNNs are trained to recognize the traits between materials using weakly-supervised learning. Particularly relevant is the result that their system is able to segment the scene with masks purely based on the material appearance, which is a local attribute, hence it does not rely on the particular shape of the objects. As noted previously, this can be useful when it comes to automatically sorting trash because products thrown away have a non-standard shape caused by the damage they experience, e.g. an empty plastic bottle could be compressed to save space, a glass bottle could be broken.

Hand-crafted-feature-based material recognition is proposed in [99], then both a generative and a discriminative model are trained from the extracted features. Moreover, to prevent the overfitting of the generative one, a greedy algorithm is designed to add one feature at a time as long as the recognition rate increases. An authors' conclusive suggestion is that the system accuracy could benefit from including the modeling of non-local features, e.g. object shape, correlated with the local surface appearances.

The authors of [112] focus on "waste in the wild" proposing a two-stage scene segmentation to yield a binary waste detection system, i.e. whether there is waste in the scene or not. At the first stage, the full scene is segmented; at the second stage, a zoom-in is performed around the detected waste and the zoomed image is processed for a fine segmentation of the waste shape. The two-stage approach shows an improvement when compared to single-stage segmentations using the same neural models. For example, an MMU could use the accurate segmentation of the second stage as input to a material recognition or volume estimation system to return the type of material or estimate its mass. The authors of [112] emphasize that the task they have considered is binary, i.e. waste or not waste, because of the leak of available images for several classes of waste type that makes currently unfeasible training a segmentation system to recognize the waste type. A similar problem was experienced in [10], where to address the issue a *data augmentation through trash simulation* is proposed: given a set of images of objects taken from the trash, i.e. pieces of trash, a new image is generated randomly combining the initial pieces of trash (say 2-6) as in Fig. 5, thus simulating images taken from the trash rather than images of single pieces.

Orientation histograms such as scale-invariant feature transform (SIFT) [72] and histograms of oriented gradients (HOG) [29] are the most commonly used low-level features for object recognition. Exploiting a kernel view the authors of [19] generalize the definition of such low-level features and give insights on how to define novel variants. Successively, these kernel descriptors have been

Fig. 5. {Taken from [10]} A technique for data augmentation is to create a new image as a random combination of the available images depicting single objects. When single waste items are combined, the resulting new image simulates the trash. Therefore, this data augmentation technique could increase the accuracy of an MMU in processing the trash.

used in [56] for both material recognition and object recognition to investigate whether these two recognition tasks are related or not; in particular they found that using the outputs of an object recognizer improves the material recognizer accuracy, whereas the object recognizer (which focused mainly on the shape) does not help material recognition.

In [68] high-level material categories are learned based on low-level and mid-level features specifically designed for material recognition. They introduce a set of mid-level features to capture the shape, the reflectance, the micro-texture aspect and the color from the image. Using all these features might cause overfitting of the training set and it is not known a-priori which features are the most relevant for material recognition. Therefore, an augmented version of the latent Dirichlet allocation algorithm [17] is developed by the authors to perform a greedy feature selection. Finally the selected features are combined together to build a material recognition system.

In [109] material classification is performed comparing two modeling approaches: the Varma-Zisserman's classifier [68, 108], that uses a bank of filters to process the image patches, and the so called "Joint" classifier, that directly uses the source image patches instead of the filter responses generated by filtering them. The empirical comparison suggested that the "Joint" classifier is more accurate. A comparison in terms of computational time/complexity is not considered in [109], but in the context of MMUs it is an important performance metric as will be highlighted in Subsection 3.4 because small platforms such as phones or microcontrollers might impose computational limits.

3.3 Computer Vision Focusing on Food

Food computing is the research area seeking to make machines able to process food images and extract information such as the type of food in the image, whether there is food or not in the image, how much food is in the image, what is its recipe and how many calories does it contain [78]. Such information helps the machine user (e.g. the user of a mobile phone) to monitor his/her diet and modify the diet for the benefit of his/her health [78]. We observe that *food is an object, i.e. a manufactured product, made of organic materials*, hence research questions and challenges addressed in food computing are closely related to material computing. As we show in this section, a reader can get useful insights from food computing for developing MMUs by simply looking at food as an object, at the recipe as the object material composition, and at the food portion volume estimation as the object component volume estimation.

From food to material recognition. For example, the food recognition approach in [114] uses SIFT descriptors [72] to compute the most likely food types appearing in the frames of a video recording a volunteer eating in a restaurant. Similarly, considering that unused or faulty objects accumulated in private houses might be a valuable source of materials, an MMU could process a video of these objects provided by the household to estimate the type and mass of each detected material.

In [8] a SIFT-based bag-of-features model [85] followed by an SVM classifier is designed after an extensive investigation considering a dataset of 5000 food images organized in 11 classes. The final classification accuracy, of the order of 78%, could be similarly achieved optimizing the model to process images taken from trash, which are complex to classify because the objects frequently have a different shape once thrown away (e.g. a bottle deformed to save space, a package that has been damaged to extract the content).

The authors of [117] use descriptors based on the relative geometric position of the ingredients exploiting the fact that a type of food has ingredients arranged in predictable spatial configurations, e.g. a sandwich has ingredients distributed vertically over multiple layers, a plate of salad has ingredients distributed horizontally all over the plate. A similar modeling approach could effectively exploit the predictable relative position of the components of a manufacturing product, e.g. an electrical machine is composed of a rotor inside a stator, a phone is externally composed of a screen on top of a case, books are multiple layers of sheets.

In [57] the focus was primarily on computationally simple detection models because it was intended to be implemented on mobile phones for a real-time food recognition application. The authors describe how the application works and its performance considering two implementations: the first one uses a bag-of-features model with SURF features [13], the second uses a Fischer vector model [88] with HOG [29]; both use the extracted image features as input for a linear SVM classifier. Similarly, a mobile phone application for material measurement could be used to process images or videos of unused and faulty products accumulated in the user's house; then, for example, products made of critical raw materials [2, 3, 48] could be collected in agreement with the householder for their recycling/re-manufacturing.

Based on the observation that food items often have ingredients distributed in slices (i.e. layers), the authors of [75] propose using "slice" convolutional kernels in a CNN [65] to improve the model classification accuracy. The authors also point out that such an accurate model to date requires memory and computational costs too high to be implemented on devices with limited resources (e.g. on mobile phones as in [57]). The idea of adapting the neural layers to the type of target items can be seen as a method to embed a-priori knowledge in the layers architecture; an alternative approach could be adding a-priori knowledge through fine-grained classes able to catch minor differences between target items (e.g. ravioli vs. dumplings, mobile phone of brand A vs. brand B) by formulating a multi-task loss function as proposed in [113].

Transfer learning [87] has shown to improve the food classification accuracy of neural models compared to models trained from scratch [105, 115]. Similarly, the features learned by a network trained on datasets with images of different types of objects could be fine-tuned (either the whole network or just part of it) on smaller datasets of images of products whose material is of particular interest, e.g. critical raw materials [2, 3, 48]. To exploit transfer learning, over the years, large neural models have been developed, trained on large datasets, made publicly available and ready to use (see Section 4 for details).

Data augmentation [101] could be applied to increase the number of images available for training. An example application for food recognition is proposed in [73], where new training images are generated rotating, translating and rescaling the original ones. Similarly, as in [123], data

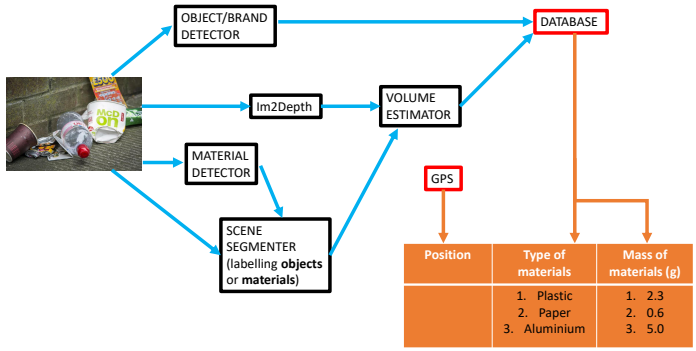

Fig. 6. An example of the system proposed in [76] adapted to measuring material compositions instead of food calories.

augmentation could be used with images of trash pieces to accurately classify their material composition (e.g. paper, plastic).

From food portion to material volume estimation. Given the similarity between processing food and non-food items, below we discuss relevant papers on food calories or food volume estimation.

In [76] the food volume is estimated requiring a single RGB image using the CNN architecture of [35] to work as a virtual depth camera; then, the depth map is converted into a voxel representation; semantic segmentation [118] is performed to identify the pixels corresponding to food and finally, in combination with the voxel representation, the volume of food is estimated. This approach, as pointed out by the developers, simply requires a single image collected "from the wild", hence it is particularly flexible; however, the whole system is quite complex as it combines multiple tasks, each one with its own complexity and with the accuracy/robustness of the whole system depending on the accuracy/robustness of all its components: open-world recognition (i.e. detecting food items from a generic scene), depth measurements from a single RGB image, 3d voxel representation from a 2d image, and scene segmentation. As pointed out by the authors, their system requires further development. Our interest in their system is its application to measure materials. An example of how the system could be adapted to our case is illustrated in Fig. 6.

In [36] a generative adversarial network [47] is used to map the input image depicting a food scene into the corresponding, pixel-by-pixel, energy content. The result is that their model reads the RGB food image and returns as output the energy (i.e. calories) at each pixel, e.g. a pixel without food has zero calories, a pixel belonging to broccoli corresponds to an energy weight lower than a pixel belonging to meat; hence, the energy weight can be seen as an energy density in J/m^3. In common with [76], this approach requires a single input image; a difference is in the fact that here the real size of the food is reconstructed using a marker of known size included in the food image rather than using a neural model as a virtual depth camera. Inspired by this work, an MMU could be implemented training a generative model to map the input RGB image of an object to the corresponding pixel-by-pixel density of material in Kg/m^3; then, the summation of all the weight-densities overlapping each mask of the semantically segmented image gives an estimate of the mass of each labeled material.

Table 1. Examples of books about the raw material composition of complex materials or products.

Material or product	Reference
carbon and graphite materials	[42]
rubber	[34]
nonwoven fabrics	[5]
energy systems	[18]
electrical and electronic materials	[50]
printed electronics	[26]
paper	[12]
plastic	[21]
automobile bodies	[31]
miscellaneous	[6, 7]

An approach based on stereo vision is proposed in [32], hence it requires two images of the food item taken from two different views; after a test on the best compromise between accuracy and efficiency, SURF is chosen for feature extraction; the system requires a reference object placed next to the food to reconstruct the real sizes, and the food should be placed inside an elliptical, flat plate, i.e. bowls are not permitted. Compared to [76], this work has the advantage to be implementable on a computer with limited performance (e.g. a mobile phone). The simplest version implemented in [76] requires the user to select the best labels and exploits the knowledge of the menu of the restaurant the dish belongs to, whereas their flexible and highly automated more complex version is, according to the authors, at a preliminary stage. Note that [32] focuses on food portion volume estimation without evaluating the calories/energy content. Its application could be adapted to estimate the volume of the components of a manufacturing product and then, knowing the density of detected materials, converted into masses. Such an MMU could run in a mobile phone.

From food recipes to material composition of products. An output that MMUs should provide is the list of materials making a target product. In general, we see two approaches to making the system capable of providing such information: (1) embedding it inside the mathematical model during the supervised training phase so that the model learns it (e.g. segmenting the scene labeling the materials as in [14]); (2) providing the MMU with access to a list of materials stored in a database (e.g. as in [76], where the recipe of a detected food item is retrieved from the restaurant menu).

Considering that the recipe of a food item is a list reporting the types and masses of materials/ingredients that compose the desired food type, the methods used in food computing to collect recipes can give insights on how to collect information about the material composition of other manufacturing products. Specifically, we list five methods:

- cookbook-like, i.e. looking at books describing the manufacturing process of the target product such as the ones in Table 1;
- websites, e.g. the manufacturer website providing the technical specifications of its products, websites collecting material compositions similarly to the ones created for recipes (e.g. Allrecipes, RecipeSource);
- research papers reporting the material composition of specific products such as the ones in Table 2;
- documentaries such as "How It's Made"[1];
- performing a chemical composition analysis of the product of interest.

[1]https://en.wikipedia.org/wiki/How_It%27s_Made

Table 2. Examples of published research articles reporting the results from an analysis of the material composition of manufacturing products.

Reference	Analyzed products
[62]	LCD screens
[92]	fluorescent lamps
[95]	computer monitors
[58]	mobile phones
[107]	headset, HDD, SSD
[40]	vehicle batteries

Table 3. Analogy between water and manufacturing networks; the SI units are between squared brackets.

	Displacement	Flow variable	Effort variable
Water network	water mass $[Kg]$	mass flow rate $[Kg/s]$	pressure $[N/m^2]$
Manufacturing network	material mass $[Kg]$	mass flow rate $[Kg/s]$	force $[N]$

3.4 Towards a Sensor Network for Material Stock Monitoring

Definition 2. *A manufacturing network is a set of locations/buildings connected by the exchange of material, e.g. raw material reservoirs, manufacturers, shops, houses, waste sorting centers, recycling centers, landfills.*

An *analogy* between water networks and manufacturing networks can be seen considering both an intuitive and a physical explanation. The intuitive explanation is that both networks are made of nodes (i.e. compartments) that exchange materials (e.g. water, aluminum, plastic) over time and space; the physical explanation is based on the force-voltage physical analogy, also known as Maxwell's analogy [20], as detailed in Table 3. Note that the two networks have a different *effort variable* because within a water network flows a fluid, whereas within a manufacturing network typically flow solid products. Another difference is that a hydraulic network confines the fluid in pipes, whereas a manufacturing network moves the solid materials using transport systems (e.g. trucks, airplanes, ships).

The physical analogy between water and manufacturing networks suggests that MMUs could be used as a sensor network to monitor the flow of materials as the system proposed in [103] for water networks. Multiple platforms equipped with MMUs could be a sensor network for material stock monitoring. An example of the system is shown in Fig. 7 considering the design of an MMU sensor network for a city using Belfast as an example. Below we list three types of platform upon which an MMU could be implemented to realize an MMU sensor network.

MMU in bins. The use of multiple sensors in "smart bins" for tasks such as detecting the waste level or reporting their geographical position have been considered over the years [52, 102]. In [9, 53, 54] a computer vision system is investigated to process internal images of bins; in particular, in these works: (1) the camera is installed on the truck (i.e. a smart truck) and used by a worker when the truck approaches the bin; (2) the image processing returns information only about the waste level. What if the camera is installed directly on each bin? What if we integrate the variables currently monitored in bins [102] with the knowledge of the class and mass of material recognized by an MMU?

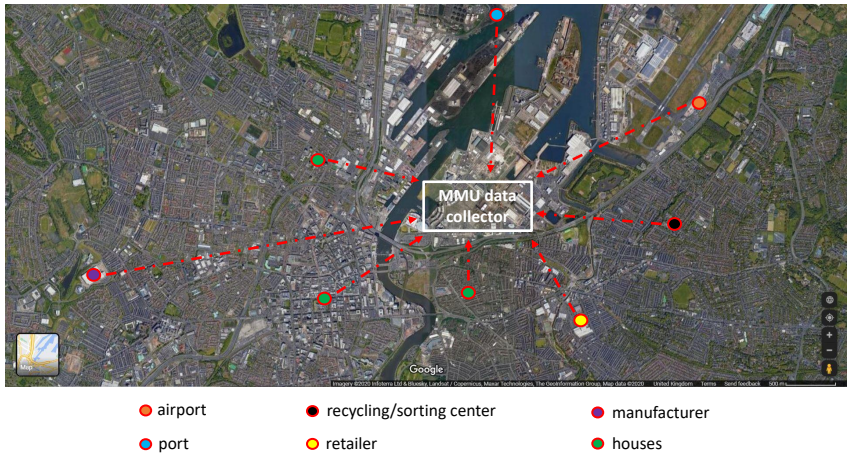

Fig. 7. Example of an MMU sensor network that could be implemented in Belfast.

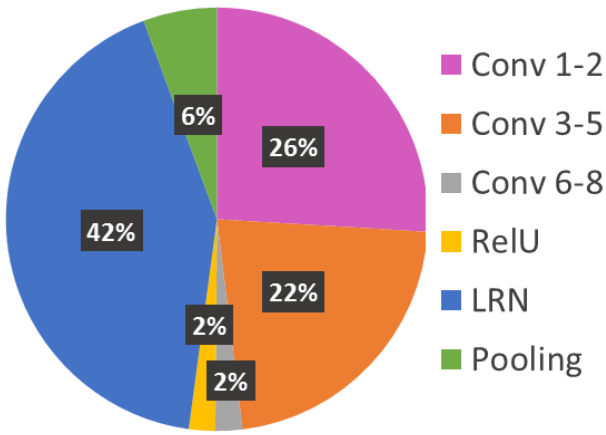

Fig. 8. {Taken from [80]} Computational time required by the different components of a large neural model to be simplified and implemented on a smartphone.

MMU in autonomous sorting systems. Previous works integrating computer vision in autonomous sorting systems have been proposed in [74] for demolition waste, in [64] for electronic waste, in [16] for pomegranate arils, in [89] for plastic granulate, in [97] for municipal solid waste and in [61] in a patent application.

MMU in mobile phones. As far as we know, the only work proposing a mobile phone application for waste detection is [80]: the pre-trained model AlexNet [60] is fine-tuned on a garbage-focused dataset using a GPU and then simplified to be implemented in smartphones; after an analysis of the image processing time required by the different components of the system as shown in Fig. 8, the model with the best compromise between classification accuracy, image processing time and memory size is chosen.

As discussed in Subsection 3.3, computer vision systems proposed for food classification and calories estimation are related to material classification and mass estimation, respectively. In [76]

the system prototype runs on mobile phones and requiring the user to provide a single image, whereas [32, 43] need two images. An interactive application is proposed in [57], whereas the authors of [90] delegate the most complex tasks such as food recognition to a cloud server instead of to the phone CPU.

Implementing MMUs on mobile phones is particularly challenging because the computational expensive task of estimating the mass from images is necessary, whereas it might be avoided in automated sorting facilities or bins through the use of weight scales; moreover, mobile phones have limited computing performance compared to the hardware that can be used in sorting facilities, e.g. GPUs. On the other hand, a mobile phone has the advantage of being portable and cheap. Potentially any owner of a mobile phone could collect material measurements through an MMU mobile application or send images/videos to a central server running an MMU.

4 HANDS-ON DEEP LEARNING

In general, the model of an MMU can be defined in two ways: using a pre-trained model or training a model from scratch. Usually the model reaches a good accuracy if it is at least fine-tuned with images depicting the domain of application. Hence, in Table 4 we list the source paper and the download website of several publicly available datasets containing images of manufacturing products or waste items that might be useful to train/fine-tune MMU models. If the choice is to use pre-trained models rather than training from scratch, the links to pre-trained models available in some machine learning libraries are: MATLAB[2], PyTorch[3], Keras[4] and Caffe[5].

5 DISCUSSIONS AND CONCLUSIONS

Motivated by the increasing concern for long-term materials supply both at local and global scale, we identified in the literature an emerging computer-vision material monitoring technology that we referred to as Material Measurement Unit and we reviewed works relevant to its development. Five main future research paths are summarized below: the first path essentially consists of implementing the most advanced recognition systems on platforms such as bins, mobile phones or sorting centers; the second, third and fourth path are concerned with improving three recognition systems already implemented on specific platforms; the last path consists of adapting systems from one platform to another (e.g. from mobile phones to smart bins).

(1) The systems developed in [96] and [99] for material recognition are based on CNN and hand-crafted features, respectively, and are not implemented on specific platforms. Therefore, their implementation on mobile phones, smart bins and sorting centers could be investigated. Successively, the material recognizer could be combined with an object classifier to improve the system accuracy as done in [56]. Two preliminary questions arise with respect to the CNN-based approach: “Do I train the network from scratch?” and “Are the publicly available datasets sufficiently rich for the target application?”. If the answer to the second question is negative, a valuable contribution to the field would be the development and publication of a new dataset for this purpose.
(2) The mobile phone application of [76] for food calories estimation is, according to the authors, at a preliminary stage. Their system is highly automated, but complex as it involves RGB map estimation, 3d voxel representation, open-world recognition and scene segmentation. Transferring the target application from food calories to material stock monitoring will result in a promising MMU.

[2]https://uk.mathworks.com/help/deeplearning/ug/pretrained-convolutional-neural-networks.html
[3]https://pytorch.org/docs/stable/torchvision/models.html
[4]https://keras.io/api/applications/
[5]https://caffe.berkeleyvision.org/model_zoo.html

Table 4. Publicly available datasets with images of materials and manufacturing products in use or as waste.

Dataset name	Reference	Original task	Waste focused?
Caltech 101[1]	[37]	Object recognition	No
Caltech 256[2]	[49]	Object recognition	No
COIL100[3]	[84]	Object recognition	No
COCO[4]	[67]	Object recognition	No
MINC[5]	[14]	Material recognition	No
ADE20K[6]	[122]	Object recognition	No
Open Images[7]	[59]	Object recognition	No
trashnet[8]	[116]	Material recognition	Yes
TACO[9]	[91]	Material/object recognition	Yes
MJU-Waste[10]	[112]	Object recognition	Yes
Flickr Material Database[11]	[100]	Material recognition	N/A
GINI[12]	[80]	Object recognition	Yes
CUReT[13]	[30]	Material recognition	N/A
ImageNet[14]	[33]	Material/object recognition	N/A

[1] http://www.vision.caltech.edu/Image_Datasets/Caltech101/#Description
[2] http://www.vision.caltech.edu/Image_Datasets/Caltech256/
[3] https://www1.cs.columbia.edu/CAVE/software/softlib/coil-100.php
[4] https://cocodataset.org/#home
[5] http://opensurfaces.cs.cornell.edu/publications/minc/
[6] https://groups.csail.mit.edu/vision/datasets/ADE20K/
[7] https://storage.googleapis.com/openimages/web/index.html
[8] https://github.com/garythung/trashnet
[9] http://tacodataset.org/
[10] https://github.com/realwecan/mju-waste
[11] http://people.csail.mit.edu/celiu/CVPR2010/FMD/index.html
[12] https://github.com/spotgarbage/spotgarbage-GINI/blob/master/README.md
[13] https://www1.cs.columbia.edu/CAVE/software/curet/index.php
[14] http://www.image-net.org/

(3) While the system mentioned in the previous point is based on CNNs, the approach in [32] is based on hand-crafted features, therefore less computationally demanding. In general, [76] could be seen as a more challenging research path to be ready for deployment later than the approach of [32]; however, the latter appears less promising in terms of both accuracy and flexibility.
(4) The mobile phone application of [80] could be improved, for example, using a more advanced neural architecture with a similar computational complexity. Moreover, a high performance central server could communicate with the phone performing the most demanding tasks, i.e. exploiting cloud computing instead of edge computing.
(5) The systems mentioned in the previous three points consider mobile phones. Their implementation for smart bins (i.e. on microcontrollers) could be investigated.

The design of a sensor network will follow when single MMUs are accurate and reliable.

Importance of Benchmarking. Regardless of the preferred research direction, benchmarking the resulting models in a standardized way permits their performance to be effectively assessed; benchmarking is a standard practice among developers of classification systems, which makes it

possible to rank models based on chosen metrics[6,7]. To advance the development of MMUs, the important benchmarking metrics are: (1) the model accuracy in *waste* item classification; (2) the model accuracy in *material* classification; (3) the model accuracy in *mass* estimation; (4) the model computational complexity (e.g. seconds needed to process an image); (5) the model memory storage requirements (e.g. its size in MB).

ACKNOWLEDGMENTS

The first author gratefully acknowledges Irish Manufacturing Research (IMR) for the financial support provided for this work.

REFERENCES

[1] [n.d.]. American Institute of Physics Handbook, Section 2b. https://web.mit.edu/8.13/8.13c/references-fall/aip/aip-handbook.html#sec1
[2] [n.d.]. Congressional Research Service, Critical Minerals and U.S. Public Policy, 2019. https://www.everycrsreport.com/files/20190628_R45810_b3112ce909b130b5d525d2265a62ce8236464664.pdf
[3] [n.d.]. European Commission, Critical Raw Materials Resilience, 2020. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0474
[4] [n.d.]. The United Nations Sustainable Development Goals. https://www.un.org/sustainabledevelopment/sustainable-development-goals/
[5] Wilhelm Albrecht, Hilmar Fuchs, and Walter Kittelmann. 2006. *Nonwoven fabrics: Raw materials, manufacture, applications, characteristics, testing processes.* John Wiley & Sons.
[6] Julian M Allwood and Jonathan M Cullen. 2015. *Sustainable materials without the hot air: Making buildings, vehicles and products efficiently and with less new material.* UIT Cambridge Limited.
[7] Julian M Allwood, Jonathan M Cullen, Mark A Carruth, Daniel R Cooper, Martin McBrien, Rachel L Milford, Muiris C Moynihan, and Alexandra CH Patel. 2012. *Sustainable materials: With both eyes open.*
[8] Marios M Anthimopoulos, Lauro Gianola, Luca Scarnato, Peter Diem, and Stavroula G Mougiakakou. 2014. A food recognition system for diabetic patients based on an optimized bag-of-features model. *IEEE journal of biomedical and health informatics* 18, 4 (2014), 1261–1271.
[9] Maher Arebey, MA Hannan, RA Begum, and Hassan Basri. 2012. Solid waste bin level detection using gray level co-occurrence matrix feature extraction approach. *Journal of environmental management* 104 (2012), 9–18.
[10] Oluwasanya Awe, Robel Mengistu, and Vikram Sreedhar. 2017. Smart trash net: Waste localization and classification. *arXiv preprint* (2017).
[11] Massimo Livi Bacci. 2017. *A concise history of world population.* John Wiley & Sons.
[12] Pratima Bajpai. 2018. *Biermann's handbook of pulp and paper: Raw material and pulp making. Volume 1, 3rd edition.* Elsevier.
[13] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. 2008. Speeded-up robust features (SURF). *Computer vision and image understanding* 110, 3 (2008), 346–359.
[14] Sean Bell, Paul Upchurch, Noah Snavely, and Kavita Bala. 2015. Material recognition in the wild with the materials in context database. In *Proceedings of the IEEE conference on computer vision and pattern recognition.* 3479–3487.
[15] Christopher M Bishop. 2006. *Pattern recognition and machine learning.* springer.
[16] José Blasco, Sergio Cubero, J Gómez-Sanchís, P Mira, and Enrique Moltó. 2009. Development of a machine for the automatic sorting of pomegranate (Punica granatum) arils based on computer vision. *Journal of food engineering* 90, 1 (2009), 27–34.
[17] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent Dirichlet allocation. *Journal of machine learning research* 3, Jan (2003), 993–1022.
[18] Alena Bleicher and Alexandra Pehlken. 2020. The material basis of energy transitions. Elsevier, 1–256.
[19] Liefeng Bo, Xiaofeng Ren, and Dieter Fox. 2010. Kernel descriptors for visual recognition. In *Advances in neural information processing systems.* 244–252.
[20] Wolfgang Borutzky. 2009. *Bond graph methodology: Development and analysis of multidisciplinary dynamic system models.* Springer Science & Business Media.
[21] John Andrew Brydson. 1999. *Plastics materials.* Elsevier.

[6]https://paperswithcode.com/task/image-classification
[7]https://benchmarks.ai/

[22] David J Crandall and Daniel P Huttenlocher. 2007. Composite models of objects and scenes for category recognition. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 1–8.

[23] Antonio Criminisi and Jamie Shotton. 2013. *Decision forests for computer vision and medical image analysis*. Springer.

[24] Gabriella Csurka, Christopher Dance, Lixin Fan, Jutta Willamowski, and Cédric Bray. 2004. Visual categorization with bags of keypoints. In *Workshop on statistical learning in computer vision, ECCV*, Vol. 1. Prague, 1–2.

[25] Gabriela Csurka, Christopher R Dance, Florent Perronnin, and Jutta Willamowski. 2006. Generic visual categorization using weak geometry. In *Toward Category-Level Object Recognition*. Springer, 207–224.

[26] Zheng Cui. 2016. *Printed electronics: Materials, technologies and applications*. John Wiley & Sons.

[27] Jonathan Cullen. 2017. Circular economy: theoretical benchmark or perpetual motion machine? (2017).

[28] Jifeng Dai, Yi Li, Kaiming He, and Jian Sun. 2016. R-FCN: Object detection via region-based fully convolutional networks. In *Advances in neural information processing systems*. 379–387.

[29] Navneet Dalal and Bill Triggs. 2005. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, Vol. 1. IEEE, 886–893.

[30] Kristin J Dana, Bram Van Ginneken, Shree K Nayar, and Jan J Koenderink. 1999. Reflectance and texture of real-world surfaces. *ACM Transactions On Graphics (TOG)* 18, 1 (1999), 1–34.

[31] Geoffrey Davies. 2012. *Materials for automobile bodies*. Butterworth-Heinemann.

[32] Joachim Dehais, Marios Anthimopoulos, Sergey Shevchik, and Stavroula Mougiakakou. 2016. Two-view 3D reconstruction for food volume estimation. *IEEE transactions on multimedia* 19, 5 (2016), 1090–1099.

[33] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*. IEEE, 248–255.

[34] John S Dick and Charles P Rader. 2014. *Raw materials supply chain for rubber products: Overview of the global use of raw materials, polymers, compounding ingredients, and chemical intermediates*. Carl Hanser Verlag GmbH Co KG.

[35] David Eigen and Rob Fergus. 2015. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In *Proceedings of the IEEE international conference on computer vision*. 2650–2658.

[36] Shaobo Fang, Zeman Shao, Runyu Mao, Chichen Fu, Edward J Delp, Fengqing Zhu, Deborah A Kerr, and Carol J Boushey. 2018. Single-view food portion estimation: Learning image-to-energy mappings using generative adversarial networks. In *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 251–255.

[37] Li Fei-Fei, Rob Fergus, and Pietro Perona. 2006. One-shot learning of object categories. *IEEE transactions on pattern analysis and machine intelligence* 28, 4 (2006), 594–611.

[38] Pedro F Felzenszwalb and Daniel P Huttenlocher. 2005. Pictorial structures for object recognition. *International journal of computer vision* 61, 1 (2005), 55–79.

[39] R. Fergus. 2009. Classical methods for object recognition. In *ICCV 2009 Short Course on Recognizing and Learning Object Categories, Kyoto, Japan*. http://people.csail.mit.edu/torralba/shortCourseRLOC/

[40] Tomer Fishman, Rupert J Myers, Orlando Rios, and TE Graedel. 2018. Implications of emerging vehicle technologies on rare earth supply and demand in the United States. *Resources* 7, 1 (2018), 9.

[41] David A Forsyth and Jean Ponce. 2012. *Computer vision: A modern approach; 2nd Edition*. https://eclass.teicrete.gr/modules/document/file.php/TM152/Books/Computer%20Vision%20-%20A%20Modern%20Approach%20-%20D.%20Forsyth,%20J.%20Ponce.pdf

[42] Wilhelm Frohs and Hubert Jaeger. February 2021. *Industrial carbon and graphite materials: Raw materials, production and applications*. John Wiley & Sons.

[43] Junyi Gao, Weihao Tan, Liantao Ma, Yasha Wang, and Wen Tang. 2019. MUSEFood: Multi-sensor-based Food Volume Estimation on Smartphones. In *2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*. IEEE, 899–906.

[44] Ross Girshick. 2015. Fast R-CNN. In *Proceedings of the IEEE international conference on computer vision*. 1440–1448.

[45] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 580–587.

[46] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. [n.d.]. *Deep Learning*. MIT Press. http://www.deeplearningbook.org

[47] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Advances in neural information processing systems*. 2672–2680.

[48] Thomas E Graedel, EM Harper, Nedal T Nassar, Philip Nuss, and Barbara K Reck. 2015. Criticality of metals and metalloids. *Proceedings of the National Academy of Sciences* 112, 14 (2015), 4257–4262.

[49] G Griffin, A Holub, and P Perona. 2007. The caltech-256: Caltech technical report. *vol* 7694 (2007), 3.

[50] KM Gupta and Nishu Gupta. 2015. *Advanced electrical and electronics materials: Processes and applications*. John Wiley & Sons.

[51] Timothy Gutowski, Daniel Cooper, and Sahil Sahni. 2017. Why we use more materials. *Philosophical transactions of the royal society a: mathematical, physical and engineering sciences* 375, 2095 (2017), 20160368.
[52] MA Hannan, Md Abdulla Al Mamun, Aini Hussain, Hassan Basri, and Rawshan Ara Begum. 2015. A review on technologies and their usage in solid waste monitoring and management systems: Issues and challenges. *Waste Management* 43 (2015), 509–523.
[53] MA Hannan, Maher Arebey, Rawshan Ara Begum, and Hassan Basri. 2011. Radio Frequency Identification (RFID) and communication technologies for solid waste bin and truck monitoring system. *Waste management* 31, 12 (2011), 2406–2413.
[54] MA Hannan, Maher Arebey, Rawshan Ara Begum, A Mustafa, and Hassan Basri. 2013. An automated solid waste bin level detection system using Gabor wavelet filters and multi-layer perception. *Resources, Conservation and Recycling* 72 (2013), 33–42.
[55] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE transactions on pattern analysis and machine intelligence* 37, 9 (2015), 1904–1916.
[56] Diane Hu, Liefeng Bo, and Xiaofeng Ren. 2011. Toward Robust Material Recognition for Everyday Objects.. In *BMVC*, Vol. 2. Citeseer, 6.
[57] Yoshiyuki Kawano and Keiji Yanai. 2015. Foodcam: A real-time food recognition system on a smartphone. *Multimedia Tools and Applications* 74, 14 (2015), 5263–5287.
[58] Yumi Kim, Hyunhee Seo, and Yul Roh. 2018. Metal recovery from the mobile phone waste by chemical and biological treatments. *Minerals* 8, 1 (2018), 8.
[59] Ivan Krasin, Tom Duerig, Neil Alldrin, Vittorio Ferrari, Sami Abu-El-Haija, Alina Kuznetsova, Hassan Rom, Jasper Uijlings, Stefan Popov, Shahab Kamali, Matteo Malloci, Jordi Pont-Tuset, Andreas Veit, Serge Belongie, Victor Gomes, Abhinav Gupta, Chen Sun, Gal Chechik, David Cai, Zheyun Feng, Dhyanesh Narayanan, and Kevin Murphy. 2017. OpenImages: A public dataset for large-scale multi-label and multi-class image classification. *Dataset available from https://storage.googleapis.com/openimages/web/index.html* (2017).
[60] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*. 1097–1105.
[61] Nalin Kumar, Manuel Gerardo Garcia, Kanishka Tyagi, et al. 2018. Material sorting using a vision system. US Patent App. 15/963,755.
[62] Ville Lahtela, Sami Virolainen, Andreas Uwaoma, Mari Kallioinen, Timo Kärki, and Tuomo Sainio. 2019. Novel mechanical pre-treatment methods for effective indium recovery from end-of-life liquid-crystal display panels. *Journal of Cleaner Production* 230 (2019), 580–591.
[63] Christoph H Lampert. 2009. *Kernel methods in computer vision*. Now Publishers Inc.
[64] Rapolti Laszlo, Rodica Holonec, Romul Copîndean, and Florin Dragan. 2019. Sorting System for e-Waste Recycling using Contour Vision Sensors. In *2019 8th International Conference on Modern Power Systems (MPS)*. IEEE, 1–4.
[65] Yann LeCun, Patrick Haffner, Léon Bottou, and Yoshua Bengio. 1999. Object recognition with gradient-based learning. In *Shape, contour and grouping in computer vision*. Springer, 319–345.
[66] Yann LeCun, Koray Kavukcuoglu, and Clément Farabet. 2010. Convolutional networks and applications in vision. In *Proceedings of 2010 IEEE international symposium on circuits and systems*. IEEE, 253–256.
[67] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. Microsoft coco: Common objects in context. In *European conference on computer vision*. Springer, 740–755.
[68] Ce Liu, Lavanya Sharan, Edward H Adelson, and Ruth Rosenholtz. 2010. Exploring features in a bayesian framework for material recognition. In *2010 IEEE computer society conference on computer vision and pattern recognition*. IEEE, 239–246.
[69] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. 2016. Ssd: Single shot multibox detector. In *European conference on computer vision*. Springer, 21–37.
[70] Jonathan Long, Evan Shelhamer, and Trevor Darrell. 2015. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3431–3440.
[71] David G Lowe. 1999. Object recognition from local scale-invariant features. In *Proceedings of the seventh IEEE international conference on computer vision*, Vol. 2. IEEE, 1150–1157.
[72] David G Lowe. 2004. Distinctive image features from scale-invariant keypoints. *International journal of computer vision* 60, 2 (2004), 91–110.
[73] Yuzhen Lu. 2016. Food image recognition by using convolutional neural networks (cnns). *arXiv preprint arXiv:1612.00983* (2016).
[74] Tuomas J Lukka, Timo Tossavainen, Janne V Kujala, and Tapani Raiko. 2014. ZenRobotics Recycler–Robotic sorting using machine learning. In *Proceedings of the International Conference on Sensor-Based Sorting (SBS)*. 1–8.

[75] Niki Martinel, Gian Luca Foresti, and Christian Micheloni. 2018. Wide-slice residual networks for food recognition. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 567–576.

[76] Austin Meyers, Nick Johnston, Vivek Rathod, Anoop Korattikara, Alex Gorban, Nathan Silberman, Sergio Guadarrama, George Papandreou, Jonathan Huang, and Kevin P Murphy. 2015. Im2Calories: towards an automated mobile vision food diary. In *Proceedings of the IEEE International Conference on Computer Vision*. 1233–1241.

[77] Krystian Mikolajczyk and Cordelia Schmid. 2002. An affine invariant interest point detector. In *European conference on computer vision*. Springer, 128–142.

[78] Weiqing Min, Shuqiang Jiang, Linhu Liu, Yong Rui, and Ramesh Jain. 2019. A survey on food computing. *ACM Computing Surveys (CSUR)* 52, 5 (2019), 1–36.

[79] Marvin Minsky. 1961. Steps toward artificial intelligence. *Proceedings of the IRE* 49, 1 (1961), 8–30.

[80] Gaurav Mittal, Kaushal B Yagnik, Mohit Garg, and Narayanan C Krishnan. 2016. Spotgarbage: smartphone app to detect garbage using deep learning. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 940–945.

[81] Marius Muja and David G Lowe. 2014. Scalable nearest neighbor algorithms for high dimensional data. *IEEE transactions on pattern analysis and machine intelligence* 36, 11 (2014), 2227–2240.

[82] Rupert J Myers, Tomer Fishman, Barbara K Reck, and TE Graedel. 2019. Unified materials information system (UMIS): An integrated material stocks and flows data structure. *Journal of Industrial Ecology* 23, 1 (2019), 222–240.

[83] Rupert J Myers, Barbara K Reck, and TE Graedel. 2019. YSTAFDB, a unified database of material stocks and flows for sustainability science. *Scientific data* 6, 1 (2019), 1–13.

[84] Sameer A Nene, Shree K Nayar, and Hiroshi Murase. 1996. Columbia Object Image Library (COIL-100). (1996).

[85] Stephen O'Hara and Bruce A Draper. 2011. Introduction to the bag of features paradigm for image classification and retrieval. *arXiv preprint arXiv:1101.3354* (2011).

[86] Aude Oliva and Antonio Torralba. 2007. The role of context in object recognition. *Trends in cognitive sciences* 11, 12 (2007), 520–527.

[87] Sinno Jialin Pan and Qiang Yang. 2009. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering* 22, 10 (2009), 1345–1359.

[88] Florent Perronnin, Jorge Sánchez, and Thomas Mensink. 2010. Improving the fisher kernel for large-scale image classification. In *European conference on computer vision*. Springer, 143–156.

[89] Tadej Peršak, Branka Viltužnik, Jernej Hernavs, and Simon Klančnik. 2020. Vision-Based Sorting Systems for Transparent Plastic Granulate. *Applied Sciences* 10, 12 (2020), 4269.

[90] Parisa Pouladzadeh and Shervin Shirmohammadi. 2017. Mobile multi-food recognition using deep learning. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 13, 3s (2017), 1–21.

[91] Pedro F Proença and Pedro Simões. 2020. TACO: Trash Annotations in Context for Litter Detection. *arXiv preprint arXiv:2003.06975* (2020).

[92] Mahmoud A Rabah. 2008. Recyclables recovery of europium and yttrium metals and some salts from spent fluorescent lamps. *Waste management* 28, 2 (2008), 318–325.

[93] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 779–788.

[94] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. 2015. Faster R-CNN: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*. 91–99.

[95] Luciene V Resende and Carlos A Morais. 2010. Study of the recovery of rare earth elements from computer monitor scraps–Leaching experiments. *Minerals Engineering* 23, 3 (2010), 277–280.

[96] Gabriel Schwartz and Ko Nishino. 2019. Recognizing material properties from images. *IEEE transactions on pattern analysis and machine intelligence* (2019).

[97] AV Seredkin, MP Tokarev, IA Plohih, OA Gobyzov, and DM Markovich. 2019. Development of a method of detection and classification of waste objects on a conveyor for a robotic sorting system. In *Journal of Physics: Conference Series*, Vol. 1359. IOP Publishing, 012127.

[98] LG Shapiro and GC Stockman. 2000. Computer Vision, March 2000.

[99] Lavanya Sharan, Ce Liu, Ruth Rosenholtz, and Edward H Adelson. 2013. Recognizing materials using perceptually inspired features. *International journal of computer vision* 103, 3 (2013), 348–371.

[100] Lavanya Sharan, Ruth Rosenholtz, and Edward Adelson. 2009. Material perception: What can you see in a brief glance? *Journal of Vision* 9, 8 (2009), 784–784.

[101] Connor Shorten and Taghi M Khoshgoftaar. 2019. A survey on image data augmentation for deep learning. *Journal of Big Data* 6, 1 (2019), 60.

[102] Gulshan Soni and Selvaradjou Kandasamy. 2017. Smart garbage bin systems–A comprehensive survey. In *International Conference on Intelligent Information Technologies*. Springer, 194–206.

[103] Ivan Stoianov, Lama Nachman, Andrew Whittle, Sam Madden, and Ralph Kling. 2008. Sensor networks for monitoring water supply and sewer systems: Lessons from Boston. In *Water Distribution Systems Analysis Symposium 2006*. 1–17.
[104] Erik B Sudderth, Antonio Torralba, William T Freeman, and Alan S Willsky. 2008. Describing visual scenes using transformed objects and parts. *International Journal of Computer Vision* 77, 1-3 (2008), 291–330.
[105] Jianing Sun, Katarzyna Radecka, and Zeljko Zilic. 2019. Exploring better food detection via transfer learning. In *2019 16th International Conference on Machine Vision Applications (MVA)*. IEEE, 1–6.
[106] Richard Szeliski. Draft of September 2020. *Computer vision: Algorithms and applications; 2nd Edition*. Springer. https://szeliski.org/Book/
[107] Esther Thiébaud, Lorenz M Hilty, Mathias Schluep, Heinz W Böni, and Martin Faulstich. 2018. Where do our resources go? indium, neodymium, and gold flows connected to the use of electronic equipment in Switzerland. *Sustainability* 10, 8 (2018), 2658.
[108] Manik Varma and Andrew Zisserman. 2005. A statistical approach to texture classification from single images. *International journal of computer vision* 62, 1-2 (2005), 61–81.
[109] Manik Varma and Andrew Zisserman. 2008. A statistical approach to material classification using image patch exemplars. *IEEE transactions on pattern analysis and machine intelligence* 31, 11 (2008), 2032–2047.
[110] Athanasios Voulodimos, Nikolaos Doulamis, Anastasios Doulamis, and Eftychios Protopapadakis. 2018. Deep learning for computer vision: A brief review. *Computational intelligence and neuroscience* 2018 (2018).
[111] Jinjiang Wang, Yulin Ma, Laibin Zhang, Robert X Gao, and Dazhong Wu. 2018. Deep learning for smart manufacturing: Methods and applications. *Journal of Manufacturing Systems* 48 (2018), 144–156.
[112] Tao Wang, Yuanzheng Cai, Lingyu Liang, and Dongyi Ye. 2020. A Multi-Level Approach to Waste Object Segmentation. *Sensors* 20, 14 (2020), 3816.
[113] Hui Wu, Michele Merler, Rosario Uceda-Sosa, and John R Smith. 2016. Learning to make better mistakes: Semantics-aware visual food recognition. In *Proceedings of the 24th ACM international conference on Multimedia*. 172–176.
[114] Wen Wu and Jie Yang. 2009. Fast food recognition from videos of eating for calorie estimation. In *2009 IEEE International Conference on Multimedia and Expo*. IEEE, 1210–1213.
[115] Guangyi Xiao, Qi Wu, Hao Chen, Da Cao, Jingzhi Guo, and Zhiguo Gong. 2019. A Deep Transfer Learning Solution for Food Material Recognition Using Electronic Scales. *IEEE Transactions on Industrial Informatics* 16, 4 (2019), 2290–2300.
[116] Mindy Yang and Gary Thung. 2016. Classification of trash for recyclability status. *CS229 Project Report* 2016 (2016).
[117] Shulin Yang, Mei Chen, Dean Pomerleau, and Rahul Sukthankar. 2010. Food recognition using statistics of pairwise local features. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 2249–2256.
[118] Hongshan Yu, Zhengeng Yang, Lei Tan, Yaonan Wang, Wei Sun, Mingui Sun, and Yandong Tang. 2018. Methods and datasets on semantic segmentation: A review. *Neurocomputing* 304 (2018), 82–103.
[119] Aston Zhang, Zachary C Lipton, Mu Li, and Alexander J Smola. October 2020. Dive into deep learning; Release 0.15.0. (October 2020). https://d2l.ai/
[120] Jianguo Zhang, Marcin Marszałek, Svetlana Lazebnik, and Cordelia Schmid. 2007. Local features and kernels for classification of texture and object categories: A comprehensive study. *International journal of computer vision* 73, 2 (2007), 213–238.
[121] Zhong-Qiu Zhao, Peng Zheng, Shou-tao Xu, and Xindong Wu. 2019. Object detection with deep learning: A review. *IEEE transactions on neural networks and learning systems* 30, 11 (2019), 3212–3232.
[122] Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso, and Antonio Torralba. 2017. Scene Parsing through ADE20K Dataset. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
[123] Federico Zocco and Seán McLoone. [n.d.]. An adaptive memory multi-batch L-BFGS algorithm for neural network training. accepted at the 21st IFAC World Congress, Berlin, Germany, July 12–17, 2020 (arXiv preprint: https://arxiv.org/abs/2012.07434).