

Article

Intelligent and Real-Time Detection and Classification Algorithm for Recycled Materials Using Convolutional Neural Networks

Dimitris Ziouzos ^{1,*} , Nikolaos Baras ^{1,*} , Vasileios Balafas ¹ , Minas Dasygenis ¹  and Adam Stimoniariis ² 

¹ Department of Electrical and Computer Engineering, University of Western Macedonia, 50100 Kozani, Greece; v.balafas@uowm.gr (V.B.); mdasygenis@uowm.gr (M.D.)

² Department of Chemical Engineering, University of Western Macedonia, 50100 Kozani, Greece; astimoniariis@uowm.gr

* Correspondence: dziouzos@uowm.gr (D.Z.); nbaras@uowm.gr (N.B.)

Abstract: In recent years, the production of municipal solid waste has constantly been increasing. Recycling is becoming more and more important, as it is the only way that we can have a clean and sustainable environment. Recycling, however, is a process that is not fully automated; large volumes of waste materials need to be processed manually. New and novel techniques have to be implemented in order to manage the increased volume of waste materials at recycling factories. In this paper, we propose a novel methodology that can identify common waste materials as they are being processed on a moving belt in waste collection facilities. An efficient waste material detection and classification system is proposed, which can be used in real integrated solid waste management systems. This system is based on a convolutional neural network and is trained using a custom dataset of images, taken on site from actual moving belts in waste collection facilities. The experimental results indicate that the proposed system can outperform existing algorithms found in the literature in real-world conditions, with 92.43% accuracy.

Keywords: recycling; convolutional neural networks; object detection



Citation: Ziouzos, D.; Baras, N.; Balafas, V.; Dasygenis, M.; Stimoniariis, A. Intelligent and Real-Time Detection and Classification Algorithm for Recycled Materials Using Convolutional Neural Networks. *Recycling* **2022**, *7*, 9. <https://doi.org/10.3390/recycling7010009>

Academic Editors: Francesco Paolo La Mantia and Beatrice Castellani

Received: 17 January 2022

Accepted: 16 February 2022

Published: 18 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Due to the constant increase in the population and lack of possibility of developing available natural resources, the problem of raw materials wastage has started to become a significant issue affecting mankind. According to the United States Environmental Protection Agency's recycling data, 75% of all municipal solid waste (MSW) that may be recycled is wasted or lost [1]. Recycling is a problem that affects all humans on a global basis, and it is closely tied to environmental contamination, economic growth, and even societal norms.

The management of MSW is inextricably linked to environmental sustainability. Excess waste materials are typically burned, stored, or discarded in landfills or even in the sea. As a result, the air, soil, and water have become contaminated, and flora and wildlife have become endangered and extinct. This is especially noticeable in developing countries, where recycling rates are significantly lower. According to research, the quantity of poisons and chemicals in the air in some developing countries is many times higher than what is considered tolerable. It is just a matter of time until people start developing respiratory and neurological disorders as a result of living in an atmosphere with such high levels of pollutants in the air.

Considering that a significant percentage of municipal solid waste can be recycled, it is critical to employ recycling processes in order to reuse materials and thereby reduce environmental issues. Many European Union countries are making progress regarding recycling [2], primarily through introducing legislation to encourage individuals to recycle

and make non-recycling more difficult. Standard waste disposal bins, which formerly collected every material, have been replaced by single-material containers throughout most places. Different materials, other than those required, are not permitted to be disposed of in these containers by law. As a result, while these containers make recycling easier for governments and waste disposal authorities, they make recycling more difficult for citizens.

Citizens' participation in the recycling process is critical in obtaining a better recycling rate. Although richer nations have succeeded in including citizens in the recycling process, they are still falling short of the goals they have planned. The recycling disparity between industrialized and underdeveloped countries is enormous. Many studies have found a link between the habit of recycling and a country's Gross Domestic Product (GDP) [3–5]. Normally, several years are needed for a nation to prosper economically. As a result, a sorting management system is required. The majority of municipal garbage is not segregated at the point of generation and ends up in dump sites. This is why we require a more effective system for separating recyclable materials, and here lies our contribution. We can detect, remove, and sort these objects on a moving belt using machine vision and artificial intelligence (AI). CNN (convolutional neural networks) characteristics are used by the majority of today's top-performing object detection systems. Having a better automated system helps us to send fewer recyclable items to landfills. In this research, we propose a computer vision system having high accuracy detection using CNN, which is one of the most recognized deep learning algorithms in image classification, segmentation and detection. Recyclable material classification is a challenging problem and requires demanding techniques to obtain a reliable dataset. Except dataset collection, it is a necessity to provide a global approach for industrial applications. In this study, we present a more accurate and optimal waste classification technique.

Waste management and the efficient sorting of the materials have been considered as important for ecologically sustainable development worldwide. It is essential for society to reduce waste accumulation by recycling and reusing disposed products. Efficient selective sorting is often implemented to improve recycling and reduce the environmental impact. This problem should be specially treated in developing countries, where waste management is a severe problem for their urbanization and economic development.

Our motivation involves finding an automatic method for sorting waste towards reducing waste and pollution. This will not only have positive environmental effects but also beneficial economic effects. In addition, our system has a great community appeal by adding the value of knowledge and the social stimulus in the separation and disposal of garbage. So, we investigate the different types of neural networks (NN) to classify the garbage waste images into four classes: glass, paper, metal, and plastic. The main goal of this research is to eliminate the risk of incorrect waste material disposal in collection bins. We propose an automatic system based on CNN (convolutional neural networks) that is capable of separating the waste materials based on their category.

The contributions of our research can be summarized as follows:

1. A high-accuracy classification method to separate the recycle materials from the waste is introduced.
2. A low-power and low-cost solution is provided because the proposed algorithm can be implemented in a wide variety of devices, including small form factor embedded devices.
3. A system that, if adopted, can increase the recycling rate and thus positively impact the society and the planet is presented.

The rest of the paper is structured as follows: Section 2 outlines similar research works found in the literature. In Section 3, we describe our proposed methodology. Section 4 presents the dataset, the algorithm and the training process. In Section 5, we provide details regarding the experiments (both synthetic and in real-world conditions) that we conducted and their results. Finally, Section 6 outlines the conclusions of this research.

2. Related Work

Municipal solid waste management processing is a complex research area that may be tackled from a variety of perspectives, including socioeconomic, scientific, ethical, and many others. We have chosen scientific papers that are most similar to our study in terms of waste object detection for our literature review.

In the research work of Ruiz et al., several CNN architectures were used for the automatic classification of waste: VGG-16 [6], VGG-19 [6], ResNet [7], Inception and Inception-ResNet. The experiments were performed on the TrashNet dataset, where the best classification results were achieved using a ResNet architecture with 88.66% average accuracy.

Ahmad et al. [8] proposed an intelligent fusion of deep features for improved waste classification. Utilizing deep architectures, namely AlexNet, GoogleNet, VggNet and ResNet pre-trained on ImageNet, they classified the waste into six classes. The best accuracy was 95.58% using the double fusion (PSO) method. However, their accuracy is worse than the accuracy of our system.

Another similar approach was presented by Khairulbadri et al. in [9]. In their work, the authors developed a material identification system for a vending machine platform. The proposed system is capable of identifying objects in real time, as the user inserts the items into the machine. To achieve this, the authors used a CNN and the YOLO algorithm. Even though, according to the experimental results, the accuracy of the proposed system is 95.9%, considering the fact that the system is configured to only identify bottles and aluminum cans, it does not look promising.

Thung and Yang [10] developed two models: a vector machine (SVM) with SIFT features and a neural network model. SVM has a 63 percent accuracy rate, while CNN has a 23 percent accuracy rate. The goal is to photograph a single element of recycling or garbage and sort it into six categories: glass, paper, metal, plastic, cardboard, and trash.

Similar works were produced by Mittal et al. [11], who detected whether an image contains garbage or not. This project employs the pretrained AlexNet model and achieves a mean accuracy of 87.69%. However, this project aims at segmenting garbage in an image without providing functions of waste classification. In our model, the accuracy is 96.57%. In this work, the waste is divided into two classes: garbage and non-garbage.

In [12], Chu et al. propose a hybrid deep learning method for waste classification to assist in the recycling process. In their system, they used six categories of waste materials: paper, plastic, metal, glass, fruit and vegetables and others. According to the experimental results of their methodology, the claimed accuracy is promising. However, it is not possible to directly compare their results with ours, since the testing of the author's methodology was performed using clear images with a single color background. In contrast, during our tests, we used a plethora of real waste material images, taken directly from the waste material belt in low light conditions. Additionally, the dataset that the authors used was not a customized dataset like ours.

Compared to the projects found in the literature, the algorithm that we propose manages to achieve a superior accuracy when performing object detection. Our system is capable of detecting materials from four different categories, including plastic, paper, aluminum and others. The novelty of the proposed methodology is that it can achieve higher accuracy percentages, due to the customization of the dataset, based on the consumer habits of each individual city.

3. Artificial Intelligence and Object Detection

Artificial intelligence (AI) is the capacity of a program or a computer-controlled machine to accomplish tasks that would normally be performed by intelligent beings [13]. The phrase is widely used to refer to a project aimed at creating systems with human-like cognitive abilities, such as the ability to reason, discern meaning, generalize, and learn from past experiences.

Ever since the invention of the digital computer in the mid-20th century, it has been proved that machines can be programmed to perform extremely complicated jobs with ease, such as finding proofs for scientific equations or playing chess [14]. Despite significant increases in computers' processing power and storage capabilities, no software is yet to equal human flexibility across broader areas or in activities requiring a great deal of common knowledge. However, several systems have surpassed the performance levels of human specialists and professionals in completing specific tasks, and artificial intelligence in this limited sense can be found in applications as diverse as medical diagnosis, computer search engines, and voice or handwriting recognition. One of these applications is the application domain of object detection.

Object detection is a computer vision feature in which numerous items are detected in digital photos or videos [15]. Humans, automobiles, furniture, houses, and animals are among the items that can be detected. Object detection algorithms attempt to answer two fundamental questions:

1. What is the object depicted in a specific picture?
2. What is the location of the object within the picture?

Object identification is accomplished using a variety of methods, including rapid R-CNN, Retina-Net, and Single-Shot MultiBox Detector (SSD). Despite the fact that these approaches have overcome the constraints of data limitation and modeling in object detection, they are not capable of detecting objects in a single algorithm run. The YOLO algorithm has gained popularity over the last few years mainly because it offers a better performance and also accuracy over the other regular object detection techniques.

4. Our Adaptive Model Methodology

The recycling of municipal solid waste (MSW) is a major goal for modern society. Although recycling has reached the desired level in most developed countries, in developing countries recycling is low. How to separate recyclable materials requires new methods to achieve the best possible results. Today's methods with manual sorters expose employees to unhealthy work situations and to a repetitive job, and are quite expensive. To design and implement a system that will help further to automate the process, which has many variables, the "adaptive mode" was developed and designed, as we called it, and is presented in Figure 1. Our proposed model gives the best possible solution to such problems as it emphasizes the purpose of use depending on the specific characteristics of each problem, and in the process of determining the solution the human factor plays an important role.

Our model has 10 stages. For a better understanding of the model, in the flowchart we have divided the process into four parts. The first part concerns the first three stages that assist in the collection of the necessary information for the system, which is why they have a different shape and color. The next four steps are for the development of the prototype. The third part concerns the evaluation of the prototype. If we were not satisfied, then we returned to the third level and continued with the rest, as shown by the dotted dots. The fourth and last part concerns the implementation and maintenance of the system.

According to the adaptive model, the system requirements are analyzed. In this first stage, we record the requirements for the system to function properly. Essentially, we record the requirements that exist in terms of hardware, software, space where the system will be used, etc. In the next step, the process is planned in a general context and according to the possible solutions that currently exist. In this second stage, the procedures that could potentially be utilized are recorded and planned. The third stage is developing the identification of use. At this stage, the user of the system actively participates in solving the problem, expressing views, thoughts, experiences, knowledge and expectations. In this way, special emphasis is placed on the human factor and more variables are taken into account. In the next step, the development of the proposed solutions takes place. At this stage, all possible solutions to the problem are listed. Then, the optimal solution is selected. Once we have chosen the optimal solution, we proceed to the design of this solution. Special emphasis should be placed on this stage since if the foundations are not laid properly we

will not be able to properly develop the system, which is the next step. Once the prototype is properly developed, one proceeds to the evaluation and usability of the system. In this step, one checks if the planned goals has been achieved, as well as the efficiency and usability of the system. It is the pilot application that is the last step before implementation. If the system fails in the evaluation, then we return to the identification of the objectives as shown by the red arrow, we check the identification of the use again knowing now how the system reacts and then we choose another solution. Once one the solution is chosen, one proceeds with the next steps and conducts another evaluation. If the evaluation of the system is positive, then one can proceed to its implementation. This is the penultimate stage for our system. However, knowing that such a system is not static but, on the contrary, is very flexible, the final stage As its maintenance and upgrades.

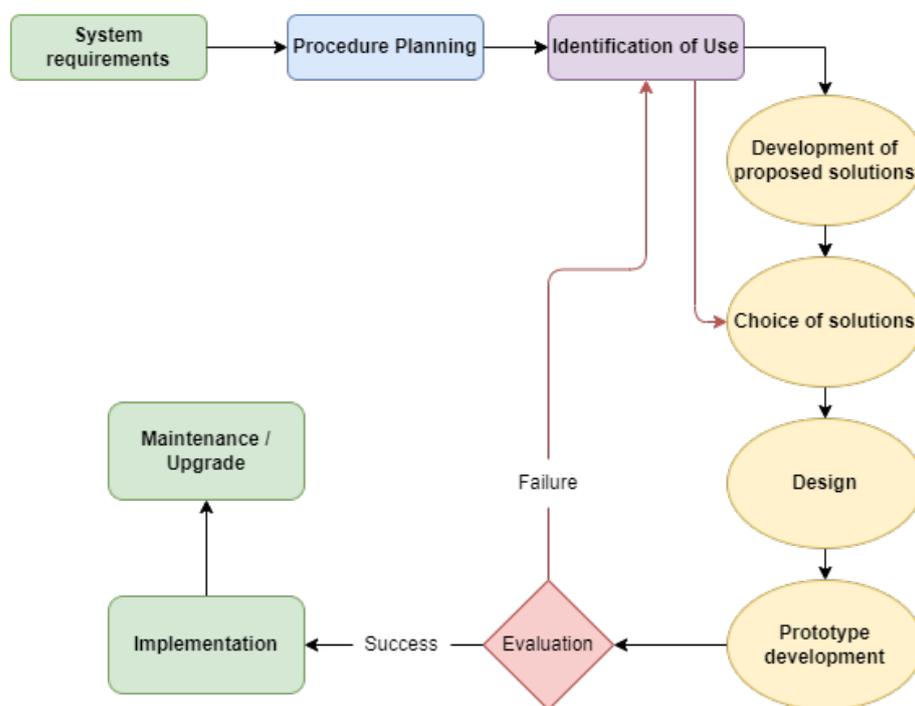


Figure 1. Our Adaptive Model.

The Case Study

To make our model more understandable, we use the “adaptive model” in the presented case study. In the present work, the aim is to implement an automated system for the detection of recyclable materials so that we have the smallest possible deviation in the appropriate choice of materials. For this purpose, our research team conducted field research in the company EPADYM (<http://en.epadym.gr/> (accessed on 16 January 2022)), which is located in the prefecture of Western Macedonia and deals with waste management and recycling of materials in order to collect data in real situations.

To implement the adaptive model, we analyzed the system requirements. At this stage, the research team was guided to the workshop and explained in detail the whole process of sorting recyclable materials that include both mechanical and electronic methods. Particular emphasis was placed on lighting, as it is very low, as well as the speed of the conveyor belt. Another important element is the location of the camera where the lighting and the vibrations from the conveyor belt can distort the results of the research. These elements play an important role in the subsequent selection of the hardware that will be selected in order to obtain the best possible result. One last point that should be given special importance is dust. At the point where the manual sorters work, there is no particular problem concerning the other points of the route since the garbage goes through various procedures. The factory has 100 Mbps fast internet as well as an internal network that

reaches 1000 Mbps. So, using the internet is an easy task. With this as a given, we could modify various systems for real-time object detection.

Knowing these data, we move on to the next stage of the process design. At this stage, the location of the camera installation was selected so that we have the best possible image capture. The location near the handholds was chosen as it is the best possible place to set up the camera. There is not much dust, the belt moves more slowly than other points and the lighting is better than in the other possible points suggested. YOLO Real-Time Object Detection was selected from the literature review and the experience of the research team.

In the third stage, the identification of use is determined. This stage is particularly important as the decisions of the previous stages are reported to the end user and researchers are expected to receive opinions, thoughts, experiences, knowledge and expectations. With the active participation of the end users, through the personal interview-discussion, the final goal of the system becomes clear. At this stage, the factory manager, the engineer, the manual sorters and the maintainers expressed their views on the implementation of the system and the problems they face. It is estimated that a percentage of recyclable materials near 15% escapes the handholds due to fatigue, both physical and mental. In the present work, this is mentioned as a user's view and for the exact reasons corresponding studies should be conducted. In the fourth stage, all the proposed solutions that have emerged after the participation of the end users are recorded. The solutions we had are:

1. Placement of the camera at the point of hand sorting and use of YOLO with specific weights and levels;
2. Placement of the camera before the point of hand sorting and use of YOLO with defaults setting;
3. Placement of two cameras at the point of hand sorting and use of YOLO with specific weights and levels.

Based on the previous stages, we chose the first solution, which in our opinion is the best possible solution.

The design of the selected solution is divided into two parts. In the first, it concerns the equipment and in the second, the software. An IR camera was chosen so that in low-light conditions the desired result can be achieved, which is connected to a laptop computer and is located on a support base next to the conveyor belt. YOLO V3 was selected and configured to obtain the maximum accuracy result.

As soon as the YOLO configuration was completed, the equipment was set up in the factory and the training of the system began. After the end of the training which lasted 3 days, the system was ready for evaluation.

During the evaluation, the results we obtained were not very good, and for this reason, we returned to the third stage which concerns the determination of the use. There we realized that more modifications should be made to YOLO and we should add more light. So, small changes were made to the code and a light projector was added to the entire system. Then we proceeded with the next steps faster than before, since the steps had been taken and we reached the stage of evaluation for the second time where the results were encouraging and therefore we proceeded to the final implementation of the system. The system then needed to be maintained and upgraded.

5. Materials and Methods

The TACO dataset was used as the basis for this study. Trash Annotations in Context (TACO) [16] is a publicly accessible photographic library of garbage outdoors. It includes photographs of garbage photographed in a variety of settings, including sea shores and roadways in England. To train and assess object detection algorithms, these photos were manually labeled and segmented according to a hierarchical taxonomy. Figure 2 shows an illustrative representation of the TACO dataset.



Figure 2. Visualization of the TACO dataset [16].

The entire amount of pictures in the TACO is 1500, which is a minimal quantity for training an object detector capable of detecting the three major classes of paper, plastic, and aluminum. As a result, we gathered 2500 more photos, labeled them, and included them to our dataset. We filmed footage from an actual trash treatment facility in the city of Kozani, Greece, in order to collect these photographs. Figure 3 depicts a frame taken from the live video feed at a moving belt. It is evident that the quality of the image and the waste material itself is lower compared to images found in popular datasets in the literature. It is, however, necessary to use these images as well if we want our system to be capable of identifying waste materials in these conditions.



Figure 3. Annotating waste materials in real-world conditions using the video feed from the moving belt at the waste collection facility in the city of Kozani, Greece.

5.1. Network Architecture

Our main detection model is based on the YOLOv4 [17] network. The architecture of our network is shown in Table 1, and the backbone network we utilized is the CSPdarknet-53 [17] network, combining CSPNet [18] and Darknet-53 [19].

Table 1. The mAP for each class of the trained model.

Class	mAP
Paper	95.42%
Plastic	95.46%
Aluminium	94.27%
Other	84.57%

Models such as DenseNet [20], VGG [6], ResNet [7] etc., are used as feature extractors. They are fine tuned on the detection dataset after being pretrained on image recognition datasets such as ImageNet [21]. These networks, which generate various levels of features with higher semantics as the network becomes deeper (more layers), turn out to be useful for the object detection network's later stages.

Between the backbone and the head, there are extra layers. They are used to remove various feature maps from various backbone points. The neck element may be an FPN [22], PANet [23], or Bi-FPN [24], for example. FPN is used by YOLOv3 to remove features of various sizes from the backbone.

5.2. CSPNet

CSPNet [18] was created to address previous work in the logic method, which necessitates a significant amount of computation in terms of network structure. The problem of high-inference computations is thought to be induced by the duplication of gradient information in network optimization, while CSPNet can minimize computation while maintaining accuracy by incorporating gradient changes into feature maps. It will not only boost the CNN's learning ability, but also reduce computational bottlenecks and memory costs while preserving precision. It is a well-known method to use ResNet as a reference and to add a residue module to the Darknet53 network to solve the deep network gradient problem. Each residue module contains two convolutional layers and one shortcut attachment, as well as duplicate residual modules. The design excludes the pooling and complete link layers; network undersampling is achieved by setting the convolution stride to 2. The image's scale will be cut in half after going through this convolution layer plate. Each convolutional layer includes convolution, BN, and Leaky Relu, as well as a zero padding after each residual module. It has been discovered that CSPDarkNet53 has advantages in target identification, making it a safer alternative for the test model's backbone. The picture classification parameters of CSPDarknet53 are 27.6 M, which is far larger than other algorithms, and the receptive field size is also much larger than other neural networks. As a consequence, it is an excellent base for our proposed flower detection system. The SSP block is also connected to the CSPDarknet53 backbone network, which can generate a fixed-size output independent of input size and can use different dimensions of the same image as input to generate pooling features of the same length. SPP has little effect on the configuration of the network when it is put behind the last convolutional layer; it merely removes the initial pooling layer. It can be used for object identification as well as image recognition. As the SSP block is used in CSPDarknet53, the approval area is greatly increased, the most relevant contextual characteristics are extracted, and network operations are barely slowed. In YOLOv3, PANET [23] was used as a parametric polymerization process for different bone levels for different detector levels instead of FPN. It is a bottom-up direction optimization aimed at promoting information transfer, which can help shorten the information path, improve the feature pyramid, and effectively locate the signal current at low levels to improve the entire feature level. PANET created adaptive feature pools to link the feature grid to all of the feature layers, allowing valuable data from each layer to flow directly into the proposed subnetwork beneath. The architecture's head, which is also an anchor-based detection model, uses YOLOv3 [19]. In the YOLOv3 model, the residual net-work structure is used as a guide to form a deeper network level and multiresolution detection, which enhances the mAP and small object detection effect.

5.3. Loss Function

DIoU [25] loss is the loss function we used, and it is possible to directly minimize the normalized distance between the anchor frame and the target frame to achieve quicker convergence speed. It is often more reliable and smoother when overlapped or also included with the goal box when doing regression. DIoU failure is based on IoU (Intersection over Union) which takes the middle distance of the bounding box into account. Therefore, it is GIoU [26] that can improve the loss of IoU in the case that the gradient does not change without overlapping boundary boxes, which adds a penalty term on the basis of the loss function of IoU.

5.4. Activation Function

The activation function is a function running on the neuron of the neural network, which is responsible for mapping the input of the neuron to the output. Its function is to increase the nonlinear change of the neural network model. In Figure 4, the Mish activation function [27] is shown, in which the original figure can be found in Mish [27].

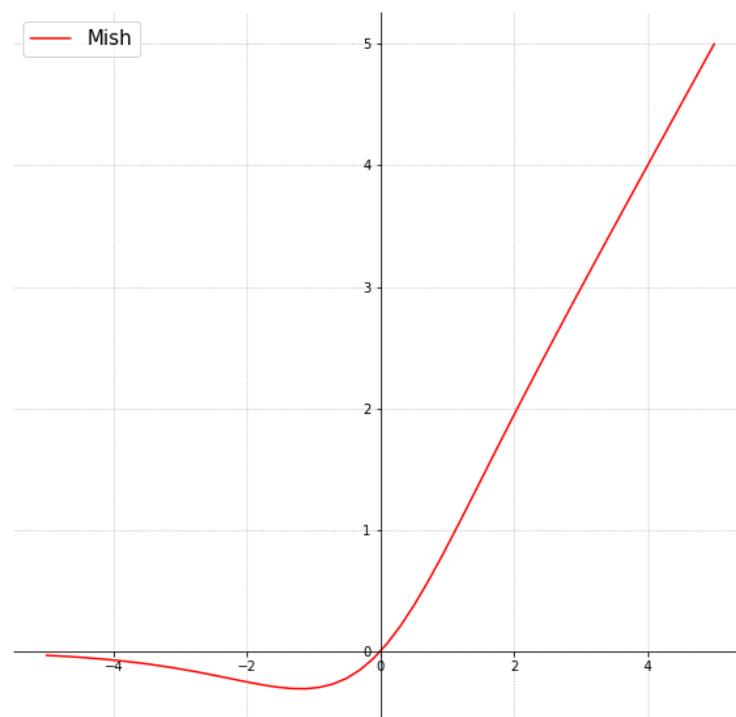


Figure 4. Mish activation function.

The Mish activation function was used in our architecture to replace Leaky Relu, which was a small persistent leak that housed the enhanced Relu function in YOLOv3.

5.5. Data Augmentation

Because the ability to gather datasets is limited, YOLOv4 produces new training samples from existing datasets. As a result, data augmentation operations on existing datasets are required to improve the generalization capabilities of the training model. As a result, varied data augmentation will optimize the use of data gathering in current image processing, which is the key to the object detection framework's effectiveness. YOLOv4 alters the image's amplitude, contrast, color, saturation, and noise through photometric distortion. Geometric distortion also includes random zooming, chopping, spinning, and reverse flipping. In addition to geometric and photometric distortion, YOLOv4 uses image occlusion technologies such as Random Erase, Cutout, Hide and Seek, Grid Mask, and image mixing technologies CutMix and Mosaic, which can be visualized in Figure 5.

Mixed-up data augmentation is the most often utilized method. Mixed-up combines two photos with different ratios to generate frames with varying degrees of certainty, enabling for additional samples to be utilized throughout processing. In comparison, Self Adversarial Training (SAT) is used to mitigate adversarial attacks to a degree. CNN measures LOSS and then uses backpropagation to create the illusion that there is no target in the picture. SAT will boost the model's robustness and generalization potential by performing standard target detection on the changed image.

Figure 5 depicts the image augmentation effect.



Figure 5. Example of Random Erasing for object detection [28].

5.6. Training Process

For the training process, we followed this procedure to prepare our dataset and train our object detection algorithm. Firstly, we separated the images into train, test, and validation folders in proportions of 80%, 10%, and 10%, respectively. In this case, there is third dataset split from the original dataset which is kept hidden from training and evaluation process. Such models have a greater likelihood of generalizing on an unseen dataset than only splitting into training and validation datasets. Then, we converted the image annotations, from the online annotation service that we used to annotate the images, to suitable formats to be readable by the algorithm. Then, we resized the images to fit the size of the neural network and after that, we created the appropriate training configuration file. We adjusted hyper-parameters for the training then we initiated the training process. When the training finished, we evaluated the results using the metrics from the COCO benchmark challenge. Finally, we extracted the results and we saved the trained model.

6. Experimental Setup

In order to evaluate the proposed methodology and examine the performance of the fine tuned YOLO algorithm and our custom dataset, we decided to conduct a series of real-world experiments in the waste collection facility (WCF) located in the city of Kozani, Greece. In the WCF, the waste materials are being processed and separated using a moving belt. To get the best results and the highest possible accuracy, it was necessary to use a high framerate camera capable of capturing video of at least 60 Frames Per Second (FPS). In our setup, we used a camera capable of taking 60 FPS live feed from the waste materials and feed it directly on to the YOLO algorithm. In Figure 6 we present the equipment that we used in order to conduct the experiments; a high performance laptop computer and a high framerate camera. The conditions were far from ideal, mainly because of the lighting of the environment and the condition of the waste materials on the moving belt. Many

materials were destroyed and deformed. We used this non-ideal conditions for our research, because we wanted to study the machine learning techniques taken from photographs “in-vivo” (on field) and not “in-vitro” (in controlled environment within the lab).



Figure 6. Setting up the equipment for the experiments at the Waste Collection Facility in the city of Kozani, Greece.

7. Results and Discussion

The total accuracy of our proposed detection and classification algorithm of recycled materials at IoU threshold at 50%, and Area-Under-Curve for each unique Recall the mean average precision (mAP@0.50) was 92.43%. From the videos that we tested in real-time detection we were able to detect and classify almost all materials that appeared in the recycling rail. In Table 1, we present the mAP for each different class.

The recall of our trained model was 92% and the average IoU was 63.58%.

The experimental results of our methodology were very promising, given the fact that we tested the algorithm using live video feed from real world waste materials as they were being processed on a moving belt. Because of the location-based custom dataset that we used, it is difficult to directly compare our system to others found in the literature, where they use waste material images in ideal conditions. Another reason that demonstrates the difficulty to compare our tool to others is the lack of real-time object detection capabilities of the other tools, as they do not support real-time object detection in such high frame rates.

Training the neural network and using the proposed system requires a significant amount of computational power. When compared to traditional methods of waste detection and classification, the proposed system has many advantages that outweigh its disadvantages. The main disadvantage of our model is the hardware cost and energy consumption. There are ways, however, that we can improve the proposed model and further increase its efficiency.

A viable alternative for improving the suggested system's execution speed and efficiency is to execute the object detection algorithm on an FPGA board, such as Zynq-7000 ARM/FPGA SoC. Over the last few decades, FPGA boards have grown in prominence, particularly in the disciplines of machine vision and robotics [29,30]. FPGA technology has a number of advantages over software operating on a CPU or GPU, including faster execution [31] and lower power consumption [32]. The FPGA hardware, as well as a high-performance conventional server, might be housed in the cloud in the suggested scheme. The server could transfer the garbage detection procedure to the FPGA board since it would collect images for identification from integrated devices in garbage collection plants. Many boards support PCIe connectivity, with a maximum theoretical bandwidth of 15.75 GB/s in version 3.0. The main disadvantage of FPGA boards, however, is the increased engineering cost and complexity. The development process on FPGA boards includes the design of custom hardware circuits. Traditionally, these hardware circuits are described via Hardware Description Languages (HDL), such as VHDL and Verilog, whereas software is programmed via one of a plethora of programming languages, such as Java, C and Python. An upcoming trend on FPGA is High Level Synthesis (HLS) [33,34]. HLS allows the programming of FPGAs using regular programming languages such as OpenCL or C++, allowing for a much higher level of abstraction. Programming FPGAs is still an order of magnitude more complicated than programming instruction-based systems, even when utilizing high-level languages. FPGA boards are only useful for speeding up the detection method at larger scales, in which the system's efficiency overcomes the extra complexity and manufacturing cost.

8. Conclusions

In recent decades, people's increased consumption combined with the Earth's ever growing population has led to large amounts of municipal solid waste. The official numbers suggest that this amount is expected to exceed 2 billion metric tons by 2025. It is clearly a problem that we have to solve in order to have a sustainable and clean environment. According to the experts, the process of recycling is the only viable solution.

The process of recycling, however, requires that the waste materials are separated. Separating the different kinds of materials you typically see in a waste collection facility is a very difficult, time-consuming and often times impossible process. Because of this, the separation procedure of waste materials is typically being performed by the hands of workers. This further increases the already high cost of recycling.

As an attempt to solve this issue, in this paper we present an innovative solution for waste detection in waste collection facilities. By utilizing pre-trained small embedded devices, we can achieve low on-site implementation cost and complexity. In order to recognize recyclable goods on a moving garbage belt, we use computer vision methods and a convolutional neural network (CNN). According to CNN, waste products are classified into four types: paper, plastic, metal, carton, and other. This project's major goal is to address the issue of non-segregated garbage, which affects both developing as well as developed countries. Overall the accuracy of the CNN in our experiments is 92.43 percent, which is quite promising, given the conditions that we conducted our experiments. Our system was capable of detecting waste materials during real-world experiments on a moving belt accurately in real time.

Author Contributions: M.D. was the supervisor of this manuscript and D.Z., N.B., V.B. and A.S. contributed equally to the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The main part of the data is the TACO dataset, which is available online for download at <http://tacodataset.org/> (accessed on 16 January 2022). In our work, the images we collected and manually labeled are available for download in our Github repository.

Acknowledgments: We would like to thank the waste collection factory located in the city of Kozani for allowing us to capture images of the waste materials and build our data set.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CNN	Convolutional neural network
YOLO	You Only Look Once
ML	Machine Learning
SAT	Self Adversarial Training

References

1. United States Environmental Protection Agency. *Recycling Statistics*; United States Environmental Protection Agency: Washington, DC, USA, 2019.
2. Levlin, J.E.; Read, B.; Grossmann, H.; Hooimeijer, A.; Ervasti, I.; Lozo, B.; Julien Saint Amand, F.; Cochaux, A.; Faul, A.; Ringman, J.; et al. *The Future of Paper Recycling in Europe: Opportunities and Limitations*; Paper Industry Technical Association: Bury, UK, 2010.
3. Kinnaman, T.C. The economics of municipal solid waste management. *Waste Manag.* **2009**, *2615*. [[CrossRef](#)] [[PubMed](#)]
4. Scheinberg, A.; Simpson, M. A tale of five cities: Using recycling frameworks to analyse inclusive recycling performance. *Waste Manag. Res.* **2015**, *33*, 975–985. [[CrossRef](#)] [[PubMed](#)]
5. Zaman, A.U. A comprehensive study of the environmental and economic benefits of resource recovery from global waste management systems. *J. Clean. Prod.* **2016**, *124*, 41–50. [[CrossRef](#)]
6. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
7. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
8. Ahmad, K.; Khan, K.; Al-Fuqaha, A. Intelligent Fusion of Deep Features for Improved Waste Classification. *IEEE Access* **2020**, *8*, 96495–96504. [[CrossRef](#)]
9. Khairulbadri, K.; Tomari, M. Development of Vision Based on Recycling Material Identification for Reverse Vending Machine Platform. *Evol. Electr. Electron. Eng.* **2020**, *1*, 368–377.
10. Thung, G.; Yang, M. Classification of Trash for Recyclability Status. 2016. Available online: <http://cs229.stanford.edu/proj2016/report/ThungYang-ClassificationOfTrashForRecyclabilityStatus-report.pdf> (accessed on 16 January 2022).
11. Mittal, G.; Yagnik, K.B.; Garg, M.; Krishnan, N.C. SpotGarbage: Smartphone App to Detect Garbage Using Deep Learning. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Heidelberg, Germany, 12–16 September 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 940–945. [[CrossRef](#)]
12. Chu, Y.; Huang, C.; Xie, X.; Tan, B.; Kamal, S.; Xiong, X. Multilayer Hybrid Deep-Learning Method for Waste Classification and Recycling. *Comput. Intell. Neurosci.* **2018**, *2018*, 5060857. [[CrossRef](#)] [[PubMed](#)]
13. Zhang, C.; Lu, Y. Study on artificial intelligence: The state of the art and future prospects. *J. Ind. Inf. Integr.* **2021**, *23*, 100224. [[CrossRef](#)]
14. Maharaj, S.; Polson, N.; Turk, A. Chess AI: Competing Paradigms for Machine Intelligence. *arXiv* **2021**, arXiv:2109.11602.
15. Kumar, M.; Bansal, M.; Kumar, M. 2D Object Recognition Techniques: State-of-the-Art Work. *Arch. Comput. Methods Eng.* **2020**, *28*. [[CrossRef](#)]
16. Proença, P.F.; Simões, P. TACO: Trash Annotations in Context for Litter Detection. *arXiv* **2020**, arXiv:2003.06975.
17. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
18. Wang, C.Y.; Liao, H.Y.M.; Wu, Y.H.; Chen, P.Y.; Hsieh, J.W.; Yeh, I.H. CSPNet: A new backbone that can enhance learning capability of CNN. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 390–391.
19. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
20. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
21. Deng, J.; Dong, W.; Socher, R.; Li, L.; Li, K.; Li, F.-F. ImageNet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255. [[CrossRef](#)]
22. Lin, T.Y.; Dollar, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.

23. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8759–8768.
24. Tan, M.; Pang, R.; Le, Q.V. Efficientdet: Scalable and efficient object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 10781–10790.
25. Zheng, Z.; Wang, P.; Liu, W.; Li, J.; Ye, R.; Ren, D. Distance-IoU loss: Faster and better learning for bounding box regression. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 12993–13000.
26. Rezatofighi, H.; Tsoi, N.; Gwak, J.; Sadeghian, A.; Reid, I.; Savarese, S. Generalized intersection over union: A metric and a loss for bounding box regression. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–17 June 2019; pp. 658–666.
27. Misra, D. Mish: A Self Regularized Non-Monotonic Activation Function. *arXiv* **2019**, arXiv:1908.08681.
28. Zhong, Z.; Zheng, L.; Kang, G.; Li, S.; Yang, Y. Random erasing data augmentation. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 13001–13008.
29. Nguyen, D.T.; Nguyen, T.N.; Kim, H.; Lee, H.J. A High-Throughput and Power-Efficient FPGA Implementation of YOLO CNN for Object Detection. *IEEE Trans. Very Large Scale Integr. VLSI Syst.* **2019**, *27*, 1861–1873. [[CrossRef](#)]
30. Long, X.; Hu, S.; Hu, Y.; Gu, Q.; Ishii, I. An FPGA-Based Ultra-High-Speed Object Detection Algorithm with Multi-Frame Information Fusion. *Sensors* **2019**, *19*, 3707. [[CrossRef](#)] [[PubMed](#)]
31. Rzaev, E.; Khanaev, A.; Amerikanov, A. Neural Network for Real-Time Object Detection on FPGA. In Proceedings of the 2021 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM), Sochi, Russia, 17–21 May 2021; pp. 719–723. [[CrossRef](#)]
32. Li, J.; Un, K.F.; Yu, W.H.; Mak, P.I.; Martins, R.P. An FPGA-Based Energy-Efficient Reconfigurable Convolutional Neural Network Accelerator for Object Recognition Applications. *IEEE Trans. Circuits Syst. II Express Briefs* **2021**, *68*, 3143–3147. [[CrossRef](#)]
33. Heinsius, L. Real-Time YOLOv4 FPGA Design with Catapult High-Level Synthesis. Master's Thesis, University of Twente, Enschede, The Netherlands, 2021.
34. Ferrandi, F.; Castellana, V.G.; Curzel, S.; Fezzardi, P.; Fiorito, M.; Lattuada, M.; Minutoli, M.; Pilato, C.; Tumeo, A. Invited: Bambu: An Open-Source Research Framework for the High-Level Synthesis of Complex Applications. In Proceedings of the 2021 58th ACM/IEEE Design Automation Conference (DAC), San Francisco, CA, USA, 5–9 December 2021; pp. 1327–1330. [[CrossRef](#)]