

WISE: An Adaptive YOLO Ensemble for Accurate E-Waste Object Detection



August 25, 2025

Abstract

Electronic waste (E-waste), containing hazardous chemicals such as lead and mercury, and valuable precious metals, has increased in the last few decades. Reaching 62 million tons in 2022, it continues to accelerate and manual sorting is unable to keep up. Only 22.3% of the 62 million tons was collected and recycled, leaving the rest in landfills. There is an urgency for automated techniques to help. Object detection algorithms, traditionally, focus on improving accuracy end up weighing false negatives and false positives the same. However here, FNs - like batteries being labeled as non-E-waste are far more hazardous.

This paper proposes (1) a new metric calibrated to help pick the best algorithm for E-waste detection, this provides a basis upon which even future algorithms can be evaluated (2) re-visits existing algorithm evaluations for the specific purpose of E-waste detection. We evaluate several standard, tweaked, ensemble methods (3) finally, we propose a smart ensemble (**WISE: Waste-focused Integrated Smart Ensemble**), an ensemble, whose weights are learned using machine learning, to minimize the costs/impact of E-waste disposal. The grand goal of these are to help improve public health and re-usability of precious metals, by enabling more efficient recovery and processing of E-waste.

Keywords: Object detection, You Only Look Once (YOLO), Electronic Waste (E-waste), Artificial Intelligence (AI), Ensemble Methods, Threshold calibration

Introduction

Electronic waste, or E-waste, consists of numerous hazardous materials including plastics, lead, mercury, cadmium, arsenic, etc. (Vats & Singh, 2014). E-waste that end up in landfills is responsible for multiple health hazards, especially in developing countries. Examples of these hazards include fetal loss, prematurity, low birth weight, abnormal thyroid function, neuro-behavioral disturbances, and geno toxicity (Noel-Brune et al., 2013). Additionally,

27 the e-waste that ends up in landfills consists of valuable metals – aluminum (Al), gold (Ag),
28 palladium (Pd), platinum (Pt) (Vats & Singh, 2014) – that can be extracted and reused, saving
29 money. On just a single TV board, 7% of its salvage value comes from silver, 33% from gold,
30 and 7% from palladium (Fornalczyk et al., 2013). In a mobile phone, 11% comes from silver,
31 71% comes from gold, and 11% comes from palladium, totaling to 93% of the phone's salvage
32 value comes from just precious metals (Fornalczyk et al., 2013).

33 In 2022 we generated 62 million metric tons of E-waste and this is expected to rapidly
34 escalate to 74 million metric tons by 2030 (Singh & Parimala S, 2025). Preventing hazardous
35 material from getting into landfills is almost the only way to stop these chemicals eventu-
36 ally finding their way into the eco-system and eventually into animal and human bodies.
37 Facilities and techniques for extracting the valuable metals from e-waste and disposing of
38 the rest appropriately exist. However, the challenge lies in identifying the E-wastes and
39 keeping up with volume. We currently depend on manual sorting, afforded in part by cheap
40 labor from developing and under-developed countries to help prevent hazardous material
41 from getting into landfills.

42 The ethical issues involved in dispatching the E-waste from developed nations to less
43 fortunate localities cannot be overstated. In developed nations, these issues and the
44 serious medical ailments caused by E-waste are often not seen everyday and becomes a
45 hidden problem that even conscientious citizens are not reminded of everyday. These less
46 fortunate localities are often the least equipped environments that simply don't have the
47 medical resources to address the issues created by these imported toxic substances. Many
48 governments have recognized this issue and have attempted to use systems and regulations
49 that cut down on e-Waste. Even with such efforts and often challenged funding, growth
50 in E-waste seems to, so far, out pacing our efforts to prevent it entering our landfills and
51 eco-system.

52 Automated techniques, though not perfect, can be of huge benefit in augmenting manual
53 sorting. Current research has already developed artificial intelligence (AI) models that have
54 been trained and evaluated for generic object-detection. Pictures of labels objects are
55 processed and used to train neural network models. Training involves the incremental
56 adjustment of the neural network's weights to match the known labels. Once trained, the
57 final network with its adjusted weights is used to label, un-labeled objects, that is, detect
58 objects. These neural net models also create bounding boxes for each object and estimate
59 labels for each object. This can even facilitate the extraction of e-waste more easily rather
60 than debating whether an object should be classified under this name.

61 The problem with existing algorithms and their benchmarks is that they focus primarily
62 on accuracy or at most precision and recall, as their measures of success. Accuracy is well
63 warranted for generic object detection, where both false positives (FP) and false negatives
64 (FN) are equally bad. Other variants like precision and recall are well warranted when

only FPs (eg. screening resumes) or FNs (eg. screening for cancer) are to be minimized, 65
respectively. Looking at the problem from the perspective of E-waste, would it be better to 66
claim that a banana peel is e-waste, not allowing it to go to the landfill directly (routing it to 67
manual sorting), or claim that a battery, full of toxic chemicals, is not e-waste, allowing it to 68
go directly to the landfill? The clear objective would be allow FPs like identifying the banana 69
peel as e-waste in lieu of allowing FNs like batteries to reach the landfill. Choosing a standard 70
measure like recall will not help either, since it can be maximized by labeling everything as 71
positive, thereby making FNs zero! But that will make automation useless since it will keep 72
over-burdening manual sorting - the exact reason we seek automation/augmentation. 73

In our research, we analyze several algorithms. The first four algorithms are standard 74
and popular implementations (YOLOv3, YOLOv4, YOLOv5, YOLOv8). The second four are 75
our tweaks to the first four to reduced FNs (YRFNv3, YRFNv4, YRFNv5, YRFNv8). We create 76
two more ensembles that we call them EYOLO and EYRFN, these estimate labels by polling 77
the 8 previous variants and the 4 reduced FN variants. Finally, we create an ensemble whose 78
weights are learned by machine learning, and call it WISE - Waste-focused Integrated Smart 79
Ensemble. Since this is a machine learning algorithm, we do in-sample comparisons with 80
other methods and also make in-sample versus out of sample comparison to ensure that 81
there is no over-fitting. 82

The rest of this paper is organized as follows. We begin below by providing a background 83
on object detection. In the section, Models, that follows, different pre-trained YOLO models 84
we use are described. The section on Methodology gathers all of the paper's contributions 85
including the tweaked models to reduce FNs; the new performance measure (EMC)'s 86
description and reasoning; the ensemble methods and the new method smart ensemble 87
method we have developed, WISE: Waste-focused Integrated Smart Ensemble. Results are 88
summarized in the section before our concluding remarks in the last section. 89

Object Detection 90

Object detection is the process of analyzing an image, localizing objects, and classifying 91
what those objects are. It is utilized in numerous applications such as face, pedestrian (Fig. 1) 92
and object detections (Li & Cao, 2020). Convolutional Neural Networks (CNN), are a specific 93
type of neural networks that play a very significant role in the object detection space. 94
They are essentially classical feed forward neural networks with additional layers called 95
convolutional layers. These convolutional layers help in any kind of computer vision by 96
applying trained averaging kernels. This kernel performs computations on the image's pixel 97
values, or representations of the color and intensity. By performing these computations, 98
the CNN is able to create a feature map which consists of channels, or representations of 99
the different learned features - lines, edges, curves. 100

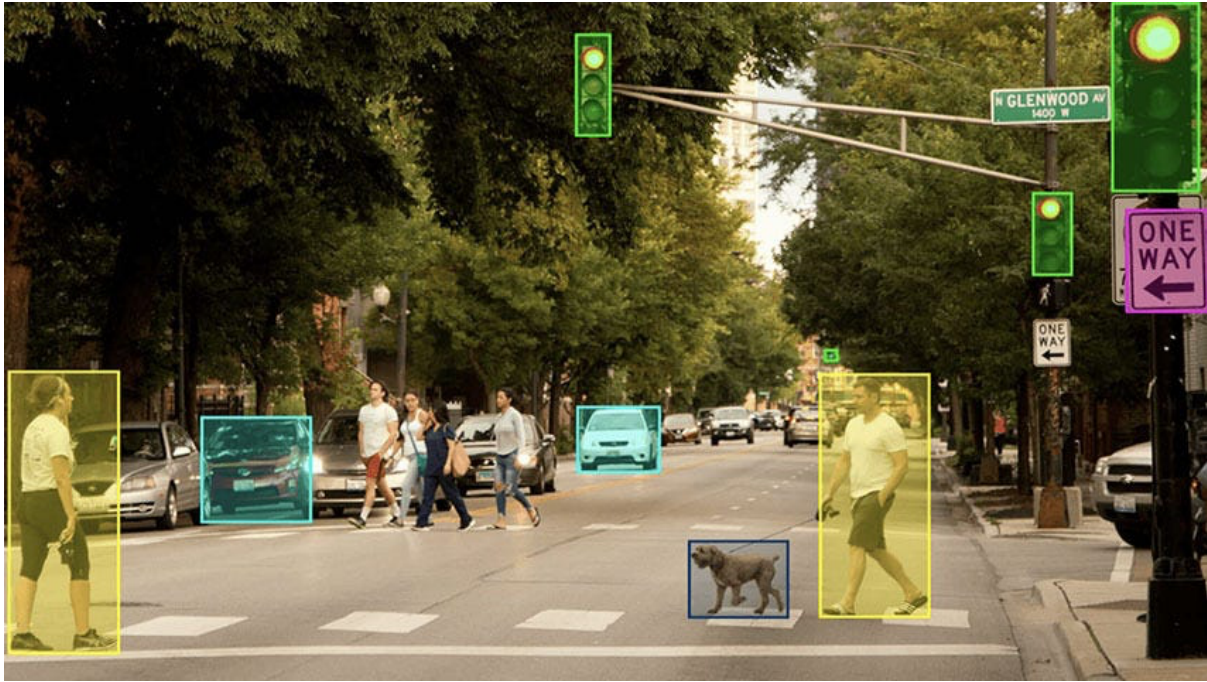


Figure 1: Object Detection

101 When an image is processed to look for certain objects, the image is first transformed
102 into a form that can be passed on as an input to a neural network (NN). This means the
103 image must be processed, resized and normalized for the neural network to take in the
104 image. A trained CNN can then be used to identify objects using intermediate steps that
105 rely on extracted visual features – edges, corners, textures (Parti, 2024). A Region Proposal
106 Network (RPN), a smaller CNN utilizing the extracted visual features, can be used to create
107 bounding boxes around where objects are likely to be located. Objects encapsulated in
108 these bounding boxes are identified and localization is used to calibrate bounding boxes to
109 exclude unnecessary parts of the image. Non-maximum suppression (NMS), a mathematical
110 algorithm, keeps only the most confident detections, excluding false positives and raising
111 accuracy. The bounding boxes are further refined and object identification is ultimately
112 finalized (Parti, 2024).

113 There are two main categories of CNN object detectors: two-stage and one-stage (Wu
114 et al., 2024). In the two-stage detector, it first makes object proposals – guesses on where
115 the object appears in the image. The second stage is classifying the object and refining
116 the bounding box. By contrast, one-stage detectors completely skip this object proposal
117 step. They go straight to predicting the bounding boxes and classifying the object. Overall,
118 two-stage detectors are more accurate but slower in comparison to one-stage detectors
119 (Wu et al., 2024).

120 You Only Look Once (YOLO) is a one-stage detector that uses a CNN architecture (Figure
121 2). It uses the CNN to predict where bounding boxes for objects should be placed. It then
122 assigns the probability of each object in a bounding box being an object of specific class,

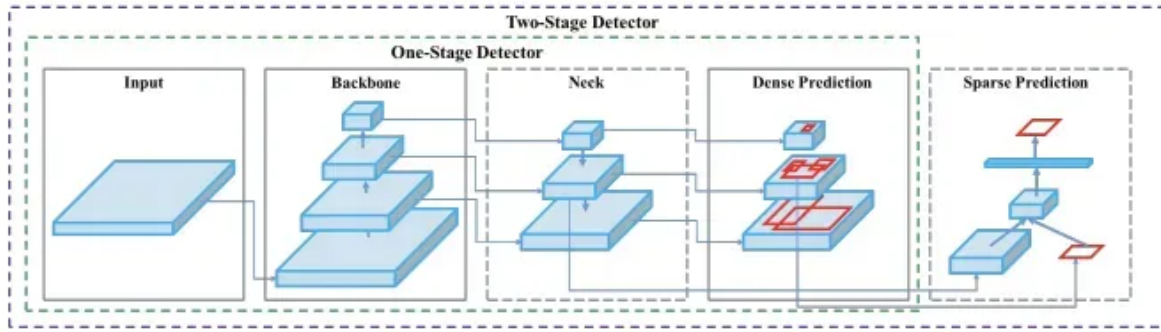


Figure 2: One Stage Vs Two Stage

which are called class-probabilities. For example, the probability the image of an animal 123
 could be classified as a dog with probability 0.1 and a cat with probability 0.9. YOLO is 124
 known for its computational speed and high accuracy. There are numerous versions of 125
 YOLO which all have distinct advantages and disadvantages. YOLO will be the primary 126
 focus of this paper due to the availability of data and popularity of this technique. 127

Models

128

There are 8 versions of the YOLO starting with version 1 proposed by (Redmon et al., 2016). 129
 The same Versions 2 and 3 are improved versions with residual blocks (Redmon & Farhadi, 130
 2018). YOLO version 4 was released by A. Bochkovskiy, who took over after Redmon retired 131
 (Bochkovskiy et al., 2020). It included major performance improvements and new training 132
 tricks. 133

Versions 5 (Jocher & Ultralytics, 2020) and Version 8 (Jocher & Ultralytics, 2023) were de- 134
 veloped by Ultralytics, and do not have an official research paper. Version 5 re-implemented 135
 YOLO in PyTorch while Version 8, improved 5, by unifying the code base and modernized 136
 the architecture. YOLO version 6 was developed by Meituan (Meituan Vision AI Department, 137
 2022), was optimized for industrial deployment and does not have a formal research paper 138
 either. YOLO version 7 (Wang et al., 2022) introduced several new features like extendable 139
 trainable bag-of-freebies and architectural refinements. The most popular amongst the 140
 YOLOs are Version 3 (last official Redmon release), Version 4 (huge leap in accuracy), Version 141
 5 (industry standard, ease use) and Version 8 (latest, cutting edge, unified framework). 142

YOLOv3

143

The backbone structure, the part of the neural network that extracts features, utilized by 144
 YOLOv3s is Darknet-53 (Cheng et al., 2021; Redmon & Farhadi, 2018). Darknet53 is a CNN 145
 with 53 layers which uses residual connections (Figure 3), allowing for input to skip over 146

147 layers to improve efficiency. Convolution blocks, layers of the CNN that extract features,
 148 use a sequence of Convolution, Batch Normalization, and Leaky ReLU, or otherwise called
 149 CBL. Convolution uses kernels to detect features, Batch Normalization allows CNNs to
 150 stabilize and train faster, and Leaky ReLU – a mathematical function – helps the CNN learn
 151 complex patterns .

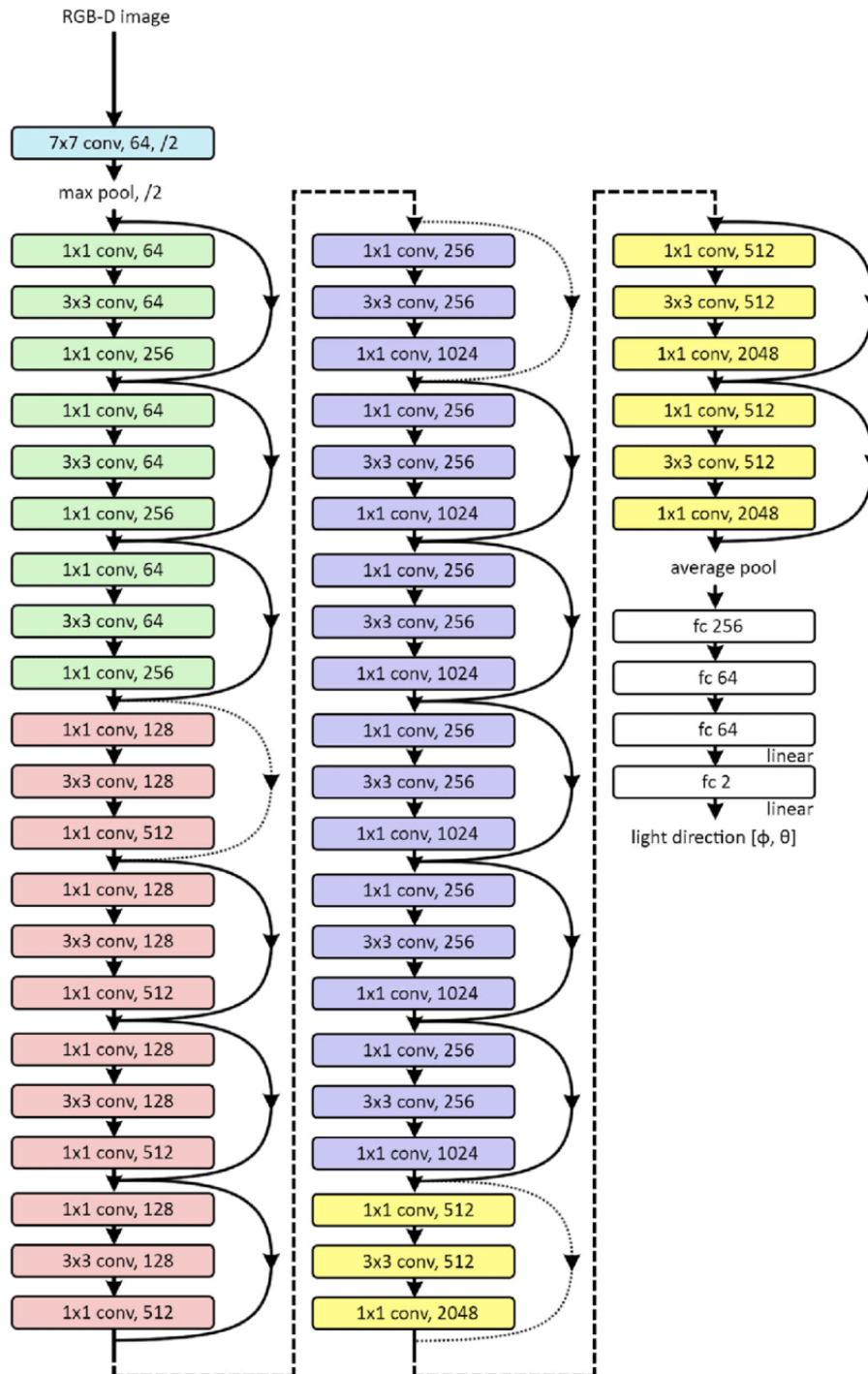


Figure 3: Darknet53 architecture

152 The head structure, or the end of the model, of YOLOv3 predicts where objects are
 153 at three scales: small, medium, and large. In other words, this helps the model detect

small objects, medium objects, and large objects. Anchor boxes, or boxes that have been predefined at certain sizes and shapes, are placed in every grid cell to facilitate bounding box predictions. Finally, the bounding boxes are placed around the target object(s).

YOLOv4 157

YOLOv4's backbone structure is CSPDarknet53 (Bochkovskiy et al., 2020). Darknet53 extracts the features while Cross Stage Partial (CSP) splits channels from the feature map such that half the channels are able to use the residual connections and the other half continue through the several layers. This ensures that there's a balance between performance and efficiency. In the neck structure – the part of the object detection model that combines details that have been collected, allowing for a better understanding of the image – of YOLOv4, Spatial Pyramid Pooling (SPP) – a pooling module, reducing the size of the feature map and keeping on essential parts, in the CNN – takes the feature maps that were created from CSPDarknet53, and underscores important patterns. It also looks at the feature maps in different sizes, helping the SPP pick up minute details as well as the larger context. Path Aggregation Network (PAN), a component of the CNN, focuses on locating where an object occurs in the image.

The head structure for YOLOv4 is the same head structure used for YOLOv3 due to its efficiency. Although the head structures are the same, YOLOv4 has improved backbone and neck structures in comparison to YOLOv3.

YOLOv5s 173

The backbone structure, the part of the neural network that extracts features, utilized by YOLOv5s is inspired by CSPDarknet53 (Feng et al., 2023; Jocher & Ultralytics, 2020). Additional convolution layers are used to extract the image features to a greater extent.

To improve its performance, YOLOv5s records its mistakes which can fall into one of three categories: bounding box regression loss, where the bounding box is placed away from the correct spot from where the object is located, as seen in Figure 4; target confidence loss, where YOLOv5s claims an object is apparent when in reality, it's not; and classification loss, when an object is mislabeled.

YOLOv8s 182

YOLOv8s uses a backbone inspired by CSP. When the channels of the feature map are created, instead of CSP splitting the channels up, YOLOv8s uses Cross Stage Partial with

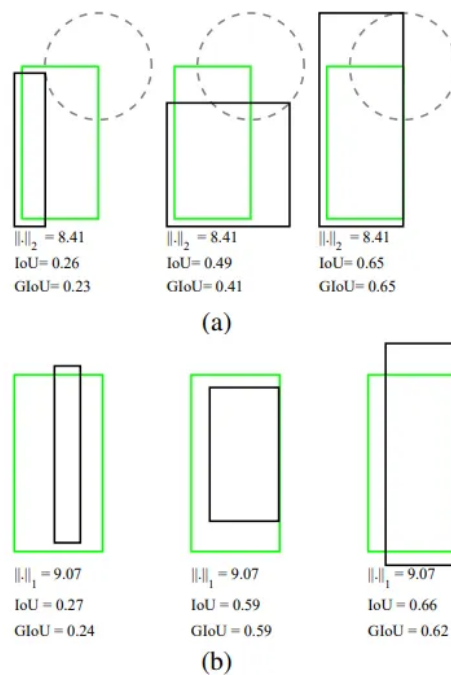


Figure 4: Bounding boxes

185 Fused Layers (C2f). C2f splits the feature map channels and fuses, or combines, the features
 186 from the channels progressively. By gradually fusing channels, the object detector is able
 187 to grab richer details rather than fusing all the channels in one shot like CSP does. A Spatial
 188 Pyramid Pooling-Faster (SPPF) module, a component in the CNN, is used to see objects
 189 in different sizes. Essentially, it uses windows – parts of the image – of different sizes so
 190 then, it's able to grab minute details as well as the bigger picture. For example, a small
 191 window can be used to identify that a dog's fur is more wavy than straight. By contrast, a
 192 big window would look at the dog entirely, and identify that the object is in the shape of a
 193 dog.

194 In the head of YOLOv8s, Upsample layers (U layers, in Figure 5 are used to raise the
 195 resolution of the feature maps. Doing this helps preserve more details, which can result
 196 in greater precision and accuracy. Three separate branches, or pathways, are used for
 197 objectness, predicting if an object is apparent; classification, identifying what object that is;
 198 and regression, predicting where the bounding box should be placed.

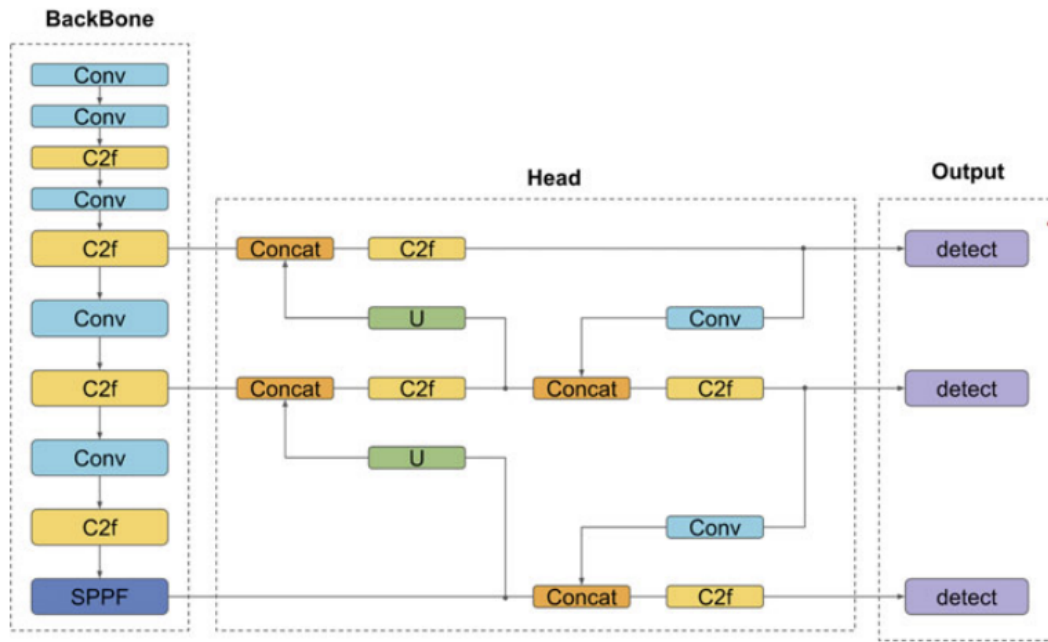


Figure 5: YOLOv8 architecture

Methodology

199

Standard evaluation measures

200

Very often, when analyzing object detection models, the common metrics utilized are 201 accuracy, precision and recall. Sometimes, though more rarely, the F1 scores. Accuracy 202 measures the percentage of predictions that are correct. Accuracy, given in Equation (1), 203 is the most popular measure for classification and is perfect for cases where making any 204 mis-classification is equally bad. All misclassifications hurt the measure the same way. 205

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

In the above, equation, True Positive (TP): Object detector successfully detects e-waste. 206 True Negative (TN): When a non-E-waste is identified by the object detector successfully 207 as non-E-waste. False Positive (FP): When non-E-waste is identified by the object detector 208 incorrectly as E-waste. False Negative (FN): When E-waste is identified by the object 209 detector wrongly as non-E-waste. 210

Precision, given by Equation (2), is also called the positive predictive value and measures 211 how many were correct, among all those predicted positive. In other words, it does not 212 care about false negatives (FNs). False positives (FPs) alone, hurt the measure. This is very 213 useful in applications such as resume screening where even missing out a good resume is 214

215 not as bad as hiring a person who is not the right fit.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

216 Recall, given in Equation (3), is also called Sensitivity or the True positive rate. It asks
217 how many the model correctly detected out of all the objects that were actually positive,
218 that is in our case, e-waste. It tells us the total detected E-waste fraction among all E-
219 wastes. In other words it does no care about FPs and only FNs hurt the measure. This is
220 perfect for applications such as cancer screening. In our case though, we do care about
221 FPs as well, just less so. That is, labeling a tennis ball as E-waste, that is FP, is bad, but not
222 nearly as bad as labeling a battery as a non-E-waste, that is FN. Moreover, recall can be
223 trivially maximized by labeling every item an E-waste. Hence recall can never be used as an
224 objective when training ML methods.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

225 Finally, the F1-score, Equation (4) balances precision and recall, and is relevant in appli-
226 cations where both are equally expensive.

$$F_1 = 2 \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

227 **Reduced FN models**

228 Appending to the list of models described in the *Models* section, we create a duplicate
229 of every model, but lower the confidence levels from 0.5 to 0.25, favoring detection. By
230 lowering the confidence level, the number of FNs is guaranteed to decrease since it increases
231 the overall positivity rate. Confidence level is the minimum confidence required for the
232 object detector to make the final decision that the target object is in the picture. We name
233 these models YRFNv3, YRFNv4, YRFNv5, YRFNv8 respectively. YRFN standing for YOLO
234 with Reduced False Negatives.

235 **EMC: A New Performance Metric for E-waste detection**

236 As we have already discussed, each of the usual measures of success used in classification
237 do not lend themselves well in E-waste detection. We seek a new metric that weights the
238 FNs substantially greater than the FNs - but not completely nullifying the FPs. Accuracy,
239 for example, undermines this situation completely, considering both FNs and FPs the same.

240 We seek a measure that gives us an estimate of the total cost of mis-identification. This

lends it self well to our case since the actual cost of extra manually sorting a FP should 241
 be contrasted against the estimated environmental cost of sending a FN to landfill. These 242
 numbers will allow us to combine the FNs and FPs in our new metric. 243

To determine these costs, we looked at the cost of off-setting the environmental conse- 244
 quences and the costs of inspection of false positives. Looking at a cost benefit analysis 245
 literature (Yang et al., 2021) it was highlighted that the cost to offset the environmental 246
 consequences was on average 4 USD/kg of e-waste, in 2021. This adjusted for inflation 247
 this is 5.95 USD/kg today (2025). Averaging common trash product weights (Empa - Swiss 248
 Federal Laboratories for Materials Science and Technology, 2025), we obtain 11.71 USD/item. 249
 To find the cost of false positives, we take the average worker’s hourly salary, and assuming 250
 they spend five minutes for each non-e-waste product, we evaluate the appropriate cost. 251
 According to (Wikipedia contributors, 2024), the average recycling worker’s hourly salary in 252
 China is \$0.17, while it is 32,000/year according to ZipRecruiter in the US. Though a wide 253
 variation, averaging as assuming objects per minute, give us 0.67 USD/item. These are 254
 rough estimates and hence it is important to note that this is only for comparison purposes. 255
 Our new method’s calibration and performance do not depend on the exact values, only 256
 that we would need non-zero values for both. 257

Below is the new metric that we propose, the E-waste misclassification cost (EMC): 258

$$EMC = \frac{11.71 FN + 0.67 FP}{TP + FN} \quad (5)$$

Ensembles 259

An ensemble method is the combining of multiple models to make a prediction - wisdom of 260
 the crowd. Different models may give different results on whether they’re able to detect an 261
 object or not. To make the final decision, a voting system is used such that if more models 262
 claim that an object is apparent, then the final prediction is that the object is present and 263
 vice-versa. 264

In this study, we created an ensemble of YOLOv3, YOLOv4, YOLOv5, YOLOv8 and 265
 another ensemble of RFNv3, YRFNv4, YRFNv5, YRFNv8. If at least three of the models 266
 detected e-waste as present, then the final decision would be that e-waste is there. We 267
 call the EYOLO and EYRFN, these estimate labels by polling the 4 YOLOs and the 4 YRFNs, 268
 respectively. 269

WISE: Waste-focused Integrated Smart Ensemble 270

Consider an ensemble that does not allow equal weighting. That is, the wisdom of the 271
 crowd is not equally weighted. Better an underlying method, the more its weight. Further, 272

273 we will let machine learning pick the weights that minimize the total EMC as defined by
274 Equation (5). Note that since we are learning these weights to minimize a loss, we will have
275 to train the weights on a training subset of the data and test it on the rest of the test set.
276 We will use a standard randomization of 80/20 to split the data for testing versus training.
277 For comparisons to other methods above, that did not require training, we will report and
278 compare averages over the entire dataset. However, to ensure that the learning is not
279 overfitting, we will compare the EMC of the test set to that of the training to ensure that
280 magnitudes are similar.

281 **Dataset**

282 A mixture of multiple datasets were used to test. We did not pick images that were used for
283 original YOLO training. We utilized Kaggle's e-waste dataset which consisted of mouses,
284 mobiles, microwaves, and keyboards. There were 300 images of each giving a total of 1200
285 images of e-waste. Another 1200 images of non-e-waste were mixed with the e-waste
286 images. These non-e-waste images came from images.cv and roboflow. They all fell under
287 various categories like handbags, chairs, spoons, books, racquets, and umbrellas. Each
288 model was given 2400 images for testing, 1200 consisting of e-waste and 1200 consisting of
289 non-e-waste. To evaluate, the number of true positives, false positives, true negatives, and
290 false negatives were counted.

291 All datasets, preprocessing scripts, and trained models are publicly available at our
292 GitHub repository (GitHub Contributors, 2025).

293 **Results**

294 For each of the methods: classical - YOLOv3, YOLOv4, YOLOv5, YOLOv8; tweaked - YRFNv3,
295 YRFNv4, YRFNv5, YRFNv8; ensemble - EYOLO, EYRFN; Tables 1 to 10 show the confusion
296 matrices. As one can observe the traditional YOLOs, though impressive for object detection,
297 do not fair too well for the purposes of E-waste detection. Very high false negatives (FNs)
298 will essentially cause too many toxic substances to enter landfills and find their ways into
299 living organisms. The YRFNS each have relatively lower FNs. What is also surprising is that
300 the cutting edge YOLO, that is version 8, which is the go-to algorithm today for object
301 detection, is far from the best for E-waste detection.

302 Table 11 shows the confusion matrix for newly developed smart ensemble - WISE (Waste-
303 focused Integrated Smart Ensemble). It has the lowest FNs amongst all methods compared.
304 The learned weights were 0 for six of the 8 method (4 YOLOs + 4 YRFNs). The only two
305 non-zero weights were YRFNv4: 0.6 and YRFNv8: 0.4 with threshold almost just above zero.

Predicted	True	
	E-waste	Not E-waste
E-waste	0	0
Not E-waste	1200	1200

Table 1: YOLOv3 Confusion Matrix

Predicted	True	
	E-waste	Not E-waste
E-waste	0	0
Not E-waste	1200	1200

Table 2: YRFNv3 Confusion Matrix

Predicted	True	
	E-waste	Not E-waste
E-waste	756	4
Not E-waste	444	1196

Table 3: YOLOv4 Confusion Matrix

Predicted	True	
	E-waste	Not E-waste
E-waste	851	11
Not E-waste	349	1189

Table 4: YRFNv4 Confusion Matrix

Predicted	True	
	E-waste	Not E-waste
E-waste	209	5
Not E-waste	991	1195

Table 5: YOLOv5 Confusion Matrix

Predicted	True	
	E-waste	Not E-waste
E-waste	373	14
Not E-waste	827	1186

Table 6: YRFNv5 Confusion Matrix

Predicted	True	
	E-waste	Not E-waste
E-waste	176	6
Not E-waste	1024	1194

Table 7: YOLOv8 Confusion Matrix

Predicted	True	
	E-waste	Not E-waste
E-waste	340	22
Not E-waste	860	1178

Table 8: YRFNv8 Confusion Matrix

Predicted	True	
	E-waste	Not E-waste
E-waste	90	0
Not E-waste	1110	1200

Table 9: EYOLO Confusion Matrix

Predicted	True	
	E-waste	Not E-waste
E-waste	203	0
Not E-waste	997	1200

Table 10: EYRFN Confusion Matrix

Predicted	True	
	E-waste	Not E-waste
E-waste	906	25
Not E-waste	294	1175

Table 11: EYRFN Confusion Matrix

306 A final comparison of all methods in terms of the popular measures of success and our
307 EMC measure are shown in Table 12. As one can see, the smart ensemble out performs
308 every other method significantly. Moreover the magnitude of EMC is in USD per object
309 and can hence easily lend itself to a relatable intuitive value. Table 12, also bolds the best
310 values for each measure of success.

Model	TP	FP	TN	FN	Accuracy	Precision	Recall	EMC
YOLOv3	0	0	1200	1200	0.50	NA	0.00	11.71
YRFNv3	0	0	1200	1200	0.50	NA	0.00	11.71
YOLOv4	756	4	1196	444	0.81	0.99	0.63	4.33
YRFNv4	851	11	1189	349	0.85	0.99	0.71	3.41
YOLOv5	209	5	1195	991	0.59	0.98	0.17	9.67
YRFNv5	373	14	1186	827	0.65	0.96	0.31	8.08
YOLOv8	176	6	1194	1024	0.57	0.97	0.15	10.00
YRFNv8	340	22	1178	860	0.63	0.94	0.28	8.40
EYOLO	90	0	1200	1110	0.54	1.00	0.08	10.83
EYRFN	203	0	1200	997	0.58	1.00	0.17	9.73
WISE - Smart(learned)	906	25	1175	294	0.87	0.97	0.76	2.88

Table 12: Performance comparison of different models.

311 Complete confusion matrices and additional performance breakdowns are provided in
312 our GitHub repository (GitHub Contributors, 2025).

313 Conclusion

314 The lower the EMC, the better the model is at detecting e-waste and reducing overall
315 cost. Though the measure, EMC, has been weight against the estimated cost of FNs against
316 that of FPs, the goal is not the cost reduction but the reduction of E-waste that goes into
317 landfills, for any given budget. This is critical especially since the ever growing volume of
318 E-waste far exceeds the rate at which budgets to tackle E-waste do not grow.

Many surprising observations were made in the results section. Classical cutting- 319
edge methods like YOLOv8, have terrible performance for E-waste detection though their 320
performance for general object detection is commendable. Further thought, would reveal 321
that this is not totally surprising since they were trained and evaluated on a generic object 322
detection context and with accuracy as the primary measure. Hence it might not be 323
fair to compare, or to use such methods for E-waste detection. Next, tweaked methods 324
out perform the classical methods whose default parameters would have been carefully 325
chosen by the authors. This again shows that even parameters calibrated for generic object 326
detection do not work well for E-waste detection. 327

Finally, our smart ensemble does surprisingly and fortunately well in the context of 328
E-waste detection. It not only has a significant cost reduction, it also has the lowest number 329
of FNs in the entire comparison set. While a methodology that reduces FNs to zero while 330
not labeling every object as FP is ideal, that would be for future research. 331

References 332

- Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). Yolov4: Optimal speed and accuracy of 333
object detection. <https://arxiv.org/abs/2004.10934> 334
- Cheng, X., Cheng, Y., Gong, Z., & Chen, Y. (2021). An improved small object detection method 335
based on yolov3. *Multimedia Tools and Applications*, 80(19), 29329–29344. <https://doi.org/10.1007/s11042-021-10605-5> 336
337
- Empa - Swiss Federal Laboratories for Materials Science and Technology. (2025). Ewaste 338
guide [Accessed: 2025-08-24]. 339
- Feng, M., Jiang, T., Wang, J., Yang, L., & Zhang, Y. (2023). A multi-scale traffic object detection 340
algorithm for road scenes based on improved yolov5. *Sensors*, 23(4), 2045. <https://doi.org/10.3390/s23042045> 341
342
- Fornalczyk, A., Willner, J., Francuz, K., & Cebulski, J. (2013). E-waste as a source of valuable 343
metals. *Archives of Materials Science and Engineering*, 63(2). 344
345
346
- Jocher, G., & Ultralytics. (2020). Yolov5. <https://github.com/ultralytics/yolov5> 347
- Jocher, G., & Ultralytics. (2023). Yolov8. <https://github.com/ultralytics/ultralytics> 348
- Li, K., & Cao, L. (2020). A review of object detection techniques. *2020 5th International 349
Conference on Electromechanical Control Technology and Transportation (ICECTT)*, 350
385–390. <https://doi.org/10.1109/ICECTT50890.2020.00091> 351
- Meituan Vision AI Department. (2022). Yolov6: A single-stage object detection framework 352
for industrial applications. <https://github.com/meituan/YOLOv6> 353
- Noel-Brune, M., Goldizen, F. C., Neira, M., van den Berg, M., Lewis, N., King, M., Suk, W. A., 354
Carpenter, D. O., Arnold, R. G., & Sly, P. D. (2013). Health effects of exposure to e-waste. 355
The Lancet Global Health, 1(2), e70. [https://doi.org/10.1016/S2214-109X\(13\)70020-2](https://doi.org/10.1016/S2214-109X(13)70020-2) 356

- 357 Parti, A. (2024). Understanding object detection: A comprehensive guide [Retrieved March
358 18, 2024, from <https://pareto.ai/blog/object-detection>].
- 359 Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-
360 time object detection. *Proceedings of the IEEE Conference on Computer Vision and*
361 *Pattern Recognition (CVPR)*, 779–788. <https://doi.org/10.1109/CVPR.2016.91>
- 362 Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. [https://arxiv.org/](https://arxiv.org/abs/1804.02767)
363 [abs/1804.02767](https://arxiv.org/abs/1804.02767)
- 364 Singh, Y., & Parimala S. (2025, March 26). E-waste management. [https://doi.org/10.2139/](https://doi.org/10.2139/ssrn.5196804)
365 [ssrn.5196804](https://doi.org/10.2139/ssrn.5196804)
- 366 Vats, M. C., & Singh, S. K. (2014). E-waste characteristic and its disposal. *International Journal*
367 *of Ecological Science and Environmental Engineering*, 1(2), 49–61.
- 368 Wang, C.-Y., Bochkovskiy, A., & Liao, H.-Y. M. (2022). Yolov7: Trainable bag-of-freebies sets
369 new state-of-the-art for real-time object detectors. [https://arxiv.org/abs/2207.](https://arxiv.org/abs/2207.02696)
370 [02696](https://arxiv.org/abs/2207.02696)
- 371 Wikipedia contributors. (2024). *Electronic waste in guiyu* [Wikipedia article]. Retrieved
372 August 24, 2025, from https://en.wikipedia.org/wiki/Electronic_waste_in_Guiyu
- 373 Wu, P., Chai, B., Nie, X., Yan, L., Wang, Z., Zhou, Q., Wang, B., Peng, Y., & Li, H. (2024, June 21).
374 Enhanced object detection: A study on vast vocabulary object detection track for
375 v3det challenge 2024. <https://doi.org/10.48550/arXiv.2406.09201>
- 376 Yang, W.-D., Sun, Q., & H-G., N. (2021). Cost-benefit analysis of metal recovery from e-
377 waste: Implications for international policy. *Waste Management*, 123, 42–47. [https:](https://doi.org/10.1016/j.wasman.2021.01.023)
378 [//doi.org/10.1016/j.wasman.2021.01.023](https://doi.org/10.1016/j.wasman.2021.01.023)

The phrasing “less fortunate localities” in the first few pages of the paper seems a bit informal and nonspecific. Perhaps the reader could change this wording to something like “socioeconomically disadvantaged” or “underprivileged,” which seem more apt?

The literature review section in the introduction should be considerably expanded (at least two paragraphs, if not more).

- Has AI been used before in the context of waste classification? Could you give some other literature that offers similar contexts?
- This is mainly because the methods/extensions offered in the paper are not overly mathematically groundbreaking, but the application is very novel and timely. Hence, this paper will seem stronger if more of the relevancy and novelty in the *applications* are emphasized.

There also did not appear to be much critical comparison with prior work on cost-sensitive classification or domain-specific ensembles, perhaps in other similar contexts/applications.

Did you preprocess (e.g., normalize) the data samples before using them? This may have been done to create the dataset, and YOLO implementations usually do this automatically (I think), but this should be specified and stated in the text.

- If these were not done automatically by the implementation, you should have implemented this.

Why do there appear to be 0 true positives for some results in the confusion matrices? This seems perhaps a bit odd.

- If this is a mistake, then please correct it. But if this is *not* a mistake, you should explain why this is the case.

WISE: An Adaptive YOLO Ensemble for Accurate E-Waste Object Detection

September 6, 2025

Abstract

1

Electronic waste (E-waste), containing hazardous chemicals such as lead and mercury, and valuable precious metals, has increased in the last few decades. Reaching 62 million tons in 2022, it continues to accelerate and manual sorting is unable to keep up. Only 22.3% of the 62 million tons was collected and recycled, leaving the rest in landfills. There is an urgency for automated techniques to help. Object detection algorithms, traditionally, focus on improving accuracy end up weighing false negatives and false positives the same. However here, FNs - like batteries being labeled as non-E-waste are far more hazardous.

2

3

4

5

6

7

8

9

This paper proposes (1) a new metric calibrated to help pick the best algorithm for E-waste detection, providing a basis upon which even future algorithms can be evaluated; (2) revisits existing algorithm evaluations for the specific purpose of E-waste detection. We evaluate several standard, tweaked and ensemble methods; (3) finally, we propose a smart ensemble (**WISE: Waste-focused Integrated Smart Ensemble**) whose weights are learned using machine learning, to minimize the costs/impact of E-waste disposal. The grand goal of these are to help improve public health and re-usability of precious metals by enabling more efficient recovery and processing of E-waste.

10

11

12

13

14

15

16

17

Keywords: Object detection, You Only Look Once (YOLO), Electronic Waste (E-waste), Artificial Intelligence (AI), Ensemble Methods, Threshold calibration

18

19

Introduction

20

Electronic waste, or E-waste, consists of numerous hazardous materials including plastics, lead, mercury, cadmium, arsenic, etc. (Vats & Singh, 2014). E-waste that end up in landfills is responsible for multiple health hazards, especially in developing countries. Examples of these hazards include fetal loss, prematurity, low birth weight, abnormal thyroid function, neuro-behavioral disturbances, and genotoxicity (Noel-Brune et al., 2013). Additionally,

21

22

23

24

25

26 the E-waste that ends up in landfills consists of valuable metals – aluminum (Al), gold (Ag),
27 palladium (Pd), platinum (Pt) (Vats & Singh, 2014) – that can be extracted and reused, saving
28 money. On just a single TV board, 7% of its salvage value comes from silver, 33% from gold,
29 and 7% from palladium (Fornalczyk et al., 2013). In a mobile phone, 11% comes from silver,
30 71% comes from gold, and 11% comes from palladium, totaling to 93% of the phone’s salvage
31 value comes from just precious metals (Fornalczyk et al., 2013).

32 In 2022 we generated 62 million metric tons of E-waste and this is expected to rapidly
33 escalate to 74 million metric tons by 2030 (Singh & Parimala S, 2025). Preventing hazardous
34 material from getting into landfills is almost the only way to stop these chemicals eventually
35 finding their way into the ecosystem and eventually into animal and human bodies. Facilities
36 and techniques for extracting the valuable metals from E-waste and disposing of the rest
37 appropriately exist. However, the challenge lies in identifying the E-wastes and keeping
38 up with volume. We currently depend on manual sorting, afforded in part by cheap labor
39 from developing and underdeveloped countries to help prevent hazardous material from
40 reaching landfills.

41 The ethical issues involved in dispatching the E-waste from developed nations to so-
42 cioeconomically disadvantaged nations cannot be overstated. In developed nations, these
43 issues and the serious medical ailments caused by E-waste are often not seen every day
44 and becomes a hidden problem that even conscientious citizens are not reminded of ev-
45 ery day. These underprivileged localities are often the least equipped environments that
46 simply don’t have the medical resources to address the issues created by these imported
47 toxic substances. Many governments have recognized this issue and have attempted to
48 use systems and regulations that cut down on E-waste. Even with such efforts and often
49 challenged funding, growth in E-waste seems to, so far, outpace our efforts to prevent it
50 from entering our landfills and ecosystems.

51 Automated techniques, though not perfect, can be of huge benefit in augmenting manual
52 sorting. Current research has already developed artificial intelligence (AI) models that have
53 been trained and evaluated for generic object-detection. Pictures of labeled objects are
54 processed and used to train neural network models. Training involves the incremental
55 adjustment of the neural network’s weights to match the known labels. Once trained, the
56 final network with its adjusted weights is used to label, unlabeled objects, that is, detect
57 objects. These neural net models also create bounding boxes for each object and estimate
58 labels for each object. This can even facilitate the extraction of E-waste more easily rather
59 than debating whether an object should be classified under this name.

60 The problem with existing algorithms and their benchmarks is that they focus primarily
61 on accuracy or at most precision and recall, as their measures of success. Accuracy is well
62 warranted for generic object detection, where both false positives (FP) and false negatives
63 (FN) are equally bad. Other variants like precision and recall are well warranted when

only FPs (eg. screening resumes) or FNs (eg. screening for cancer) are to be minimized, 64
respectively. Looking at the problem from the perspective of E-waste, would it be better to 65
claim that a banana peel is E-waste, not allowing it to go to the landfill directly (routing it 66
to manual sorting), or claim that a battery, full of toxic chemicals, is not E-waste, allowing 67
it to go directly to the landfill? The clear objective would be to allow FPs like identify- 68
ing the banana peel as E-waste in lieu of allowing FNs like batteries to reach the landfill. 69
Choosing a standard measure like recall will not help either, since it can be maximized 70
by labeling everything as positive, thereby making FNs zero! But that will make automa- 71
tion useless since it will keep over-burdening manual sorting - the exact reason we seek 72
automation/augmentation. 73

Specifically in terms of literature on AI for waste detection, deep learning models have 74
been used for solid waste classification (Adedeji & Wang, 2019; Majchrowska et al., 2022; 75
Oza et al., 2025). The *TrashNet* (Thung & Yang, 2017) data set was created as a part of a 76
student-led project in 2017 and has been one of the popular datasets for trash classification 77
benchmarking. However, this data set is not relevant for E-waste classification since the 78
only six classes included are glass, paper, cardboard, plastic, metal and trash. It contains 79
2527 images and has been used for bench marking in some research papers (Khan et al., 2024; 80
Rahim et al., 2024). Another popular direction of research has been in identifying *waste in* 81
the wild. The *Taco* data set (Proença & Simões, 2020, 2023) provides 1500 images of trash in 82
various locations, from the beaches to city streets. These are to be used to train algorithms 83
that can find trash from pictures taken from, say, an automated garbage collector (Fan et al., 84
2023; Promboonruang et al., 2024; Song et al., 2025). Research on these two trash related 85
applications, even though not directly related to E-waste, demonstrate the unacceptable 86
performances of generic methods for specific application domains. They remind us of the 87
technological and the societal need to develop application-specific methods especially for 88
waste related topics. 89

YOLO family of detectors, though created in 2016 (Redmon et al., 2016), only became 90
popular and became the go-to choice recently. These one-stage detectors have been shown 91
to be very effective specifically for integration into real-time waste sorting, for example 92
with robotic arms (Ibrahim et al., 2023; Paudel et al., 2024). However, in terms of effective 93
sorting methods for identifying E-waste, there exists only one recent paper (Rajeev et al., 94
2025) that simply compares different YOLO methods in terms of accuracy and the reported 95
poor performances underscore the need for improved methods. 96

In our research, we analyze several algorithms. The first four algorithms are standard 97
and popular implementations (YOLOv3, YOLOv4, YOLOv5, YOLOv8). The second four are 98
our tweaks to the first four to reduce FNs (YRFNv3, YRFNv4, YRFNv5, YRFNv8). We create 99
two more ensembles that we call EYOLO and EYRFN, these estimate labels by polling the 100
8 previous variants and the 4 reduced FN variants. Finally, we create an ensemble whose 101
weights are learned by machine learning, and call it WISE - Waste-focused Integrated Smart 102

103 Ensemble. Since this is a machine learning algorithm, we do in-sample comparisons with
104 other methods and also make in-sample versus out-of-sample comparison to ensure that
105 there is no overfitting.

106 Prior waste detection or E-waste detection studies predominantly optimize for accuracy,
107 precision or recall rather than application level costs of mistakes. None, to our knowledge,
108 consider the application level nuances of differentiating costs of FP and FN. There are no
109 proposed methodology improvements or enhancements specifically for detecting E-waste.
110 However, in classical Machine Learning, both (a) minimizing expected misclassifications
111 cost (EMC), when costs of FP and FN differ (Domingos, 1999; Elkan, 2001; Sheng & Ling,
112 2006) and (b) creating better performing ensemble are popular research topics (Bodla
113 et al., 2017; Lin et al., 2017; Solovyev et al., 2021; Wu & Zhu, 2013). Our contribution is to
114 (1) create a new metric calibrated to help pick the best algorithm for E-waste detection,
115 providing a basis upon which even future algorithms can be evaluated; (2) revisits existing
116 algorithm evaluations for the specific purpose of E-waste detection. We evaluate several
117 standard, tweaked and ensemble methods; (3) finally, we propose a smart ensemble (**WISE:**
118 **Waste-focused Integrated Smart Ensemble**) whose weights are learned using machine
119 learning, to minimize the costs/impact of E-waste disposal.

120 The rest of this paper is organized as follows. We begin below by providing a background
121 on object detection. In the *Models* section, different pre-trained YOLO models we used are
122 described. The section on *Methodology* gathers all of the paper's contributions including
123 the tweaked models to reduce FNs; the new performance measure (EMC)'s description and
124 reasoning; the ensemble methods and the new method smart ensemble method we have
125 developed, WISE: Waste-focused Integrated Smart Ensemble. Results are summarized in
126 the section before our concluding remarks in the last section.

127 **Object Detection**

128 Object detection is the process of analyzing an image, localizing objects, and classifying
129 what those objects are. It is utilized in numerous applications such as face, pedestrian (Fig. 1)
130 and object detections (Li & Cao, 2020). Convolutional Neural Networks (CNN), are a specific
131 type of neural networks that play a very significant role in the object detection space.
132 They are essentially classical feed forward neural networks with additional layers called
133 convolutional layers. These convolutional layers help in any kind of computer vision by
134 applying trained averaging kernels. This kernel performs computations on the image's pixel
135 values, or representations of the color and intensity. By performing these computations,
136 the CNN is able to create a feature map which consists of channels, or representations of
137 the different learned features – lines, edges, curves.

138 When an image is processed to look for certain objects, the image is first transformed

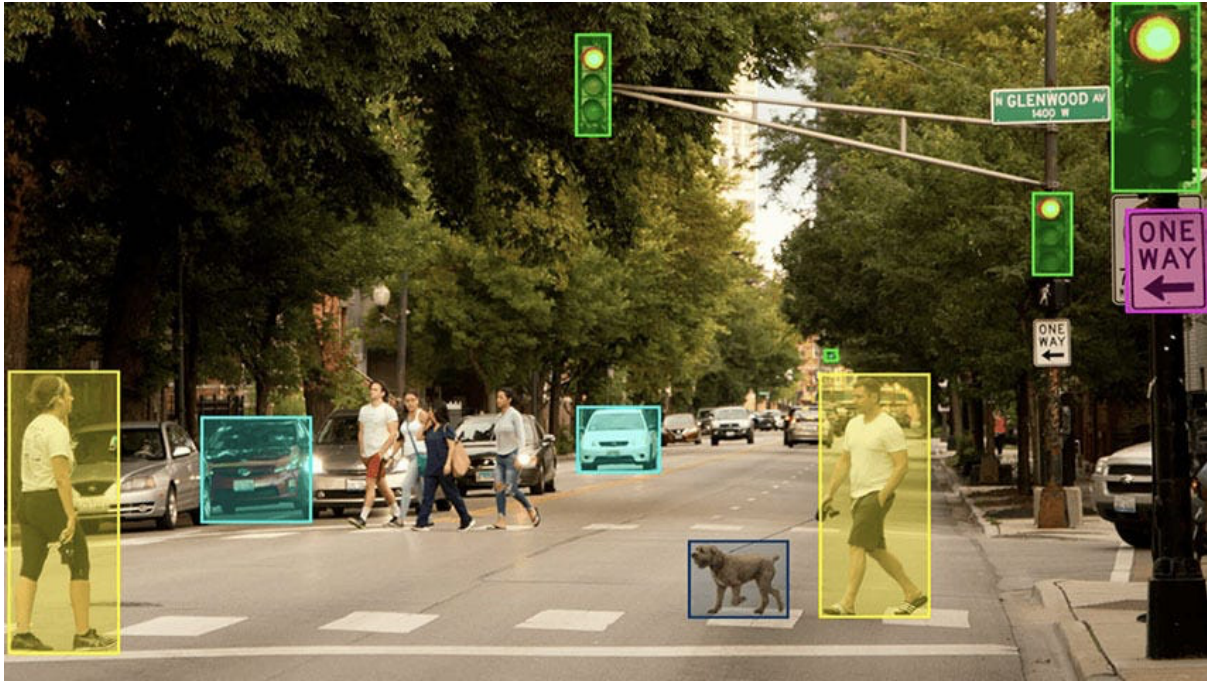


Figure 1: Object Detection¹

into a form that can be passed on as an input to a neural network (NN). This means the 139 image must be processed, re-sized and normalized for the neural network to take in the 140 image. All inputs are also normalized to $[0, 1]$ by dividing by 255. A trained CNN can 141 then be used to identify objects using intermediate steps that rely on extracted visual 142 features – edges, corners, textures (Parti, 2024). A Region Proposal Network (RPN), a 143 smaller CNN that utilizes extracted visual features, can be used to create bounding boxes 144 around potential objects locations. Objects encapsulated in these bounding boxes are then 145 identified and localization is used to calibrate bounding boxes to exclude unnecessary parts 146 of the image. Non-maximum suppression (NMS), a mathematical algorithm, keeps only the 147 most confident detections, excluding false positives and improves accuracy. The bounding 148 boxes are further refined and object identification is finalized (Parti, 2024). 149

There are two main categories of CNN object detectors: two-stage and one-stage (Wu 150 et al., 2024). In the two-stage detector, it first makes object proposals – guessing where 151 the object appears in the image. The second stage classifies the object and refines the 152 bounding box. By contrast, one-stage detectors completely skip this object proposal step. 153 They go straight to predicting the bounding boxes and classifying objects. Overall, two- 154 stage detectors are more accurate but computational much more expensive and slower 155 in comparison to one-stage detectors (Wu et al., 2024). The benefits offered by improved 156 one-stage methods and the costs associated with the two-stage classifiers, have eventually 157 made one-stage classifiers much more popular. 158

You Only Look Once (YOLO) is the most popular one-stage detector that uses a CNN 159

¹Extracted from Potter, 2022.

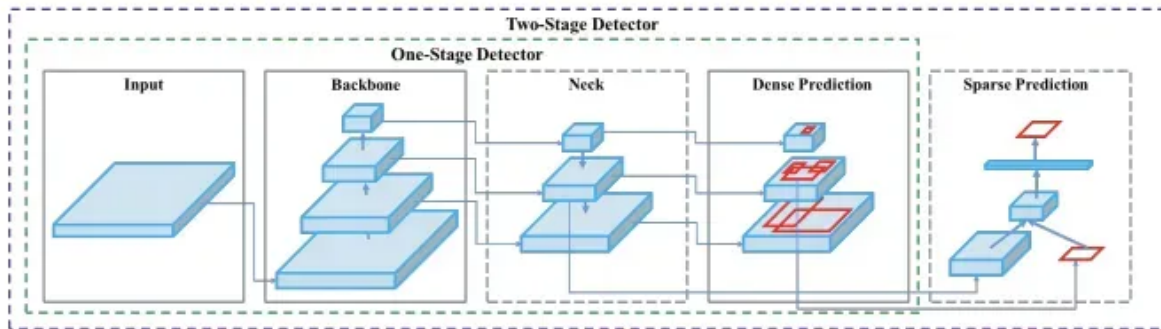


Figure 2: One Stage Vs Two Stage²

160 architecture (Figure 2). It uses the CNN to predict where bounding boxes for objects should
 161 be placed. It then assigns the probability of each object in a bounding box being an object of
 162 specific class, which are called class-probabilities. For example, the probability the image
 163 of an animal could be classified as a dog with probability 0.1 and a cat with probability
 164 0.9. YOLO is known for its computational speed and high accuracy. There are numerous
 165 versions of YOLO which all have distinct advantages and disadvantages. YOLO will be the
 166 primary focus of this paper due to the availability, open-source license and the popularity
 167 of this technique.

168 Models

169 There are eight versions of the YOLO starting with version 1 proposed by (Redmon et al.,
 170 2016). The same Versions 2 and 3 are improved versions with residual blocks (Redmon &
 171 Farhadi, 2018). YOLO version 4 was released by A. Bochkovskiy, who took over after Redmon
 172 retired (Bochkovskiy et al., 2020). It included major performance improvements and new
 173 training tricks.

174 Versions 5 (Jocher & Ultralytics, 2020) and Version 8 (Jocher & Ultralytics, 2023) were de-
 175 veloped by Ultralytics, and do not have an official research paper. Version 5 re-implemented
 176 YOLO in PyTorch while Version 8, improved 5, by unifying the code base and modernized
 177 the architecture. YOLO version 6 was developed by Meituan (Meituan Vision AI Department,
 178 2022), was optimized for industrial deployment and does not have a formal research paper
 179 either. YOLO version 7 (Wang et al., 2022) introduced several new features like extendable
 180 trainable bag-of-freebies and architectural refinements. The most popular among the YO-
 181 LOs are Version 3 (last official Redmon release), Version 4 (huge leap in accuracy), Version 5
 182 (industry standard, ease of use) and Version 8 (latest, cutting edge, unified framework).

²Extracted from Solawetz, 2024

YOLOv3

183

The backbone structure, the part of the neural network that extracts features, utilized by YOLOv3s is Darknet-53 (Cheng et al., 2021; Redmon & Farhadi, 2018). Darknet53 is a CNN with 53 layers which uses residual connections (Figure 3), allowing for input to skip over layers to improve efficiency. Convolution blocks, layers of the CNN that extract features, use a sequence of Convolution, Batch Normalization, and Leaky ReLU, or otherwise called CBL. Convolution uses kernels to detect features, Batch Normalization allows CNNs to stabilize and train faster, and Leaky ReLU – a mathematical function – helps the CNN learn complex patterns. All YOLO methods include normalization of each channel to $[0, 1]$ and do not have per-channel mean/standard deviation standardizations.

The head structure, or the end of the model, of YOLOv3 predicts where objects are at three scales: small, medium, and large. In other words, this helps the model detect small objects, medium objects, and large objects. Anchor boxes, or boxes that have been predefined at certain sizes and shapes, are placed in every grid cell to facilitate bounding box predictions. Finally, the bounding boxes are placed around the target object(s).

YOLOv4

198

YOLOv4's backbone structure is CSPDarknet53 (Bochkovskiy et al., 2020). Darknet53 extracts the features while Cross Stage Partial (CSP) splits channels from the feature map such that half the channels are able to use the residual connections and the other half continue through the several layers. This ensures that there's a balance between performance and efficiency. In the neck structure – the part of the object detection model that combines details that have been collected, allowing for a better understanding of the image – of YOLOv4, Spatial Pyramid Pooling (SPP) – a pooling module, reducing the size of the feature map and keeping on essential parts, in the CNN – takes the feature maps that were created from CSPDarknet53, and underscores important patterns. It also looks at the feature maps in different sizes, helping the SPP pick up minute details as well as the larger context. Path Aggregation Network (PAN), a component of the CNN, focuses on locating where an object occurs in the image.

The head structure for YOLOv4 is the same head structure used for YOLOv3 due to its efficiency. Although the head structures are the same, YOLOv4 has an improved backbone and neck structure in comparison to YOLOv3.

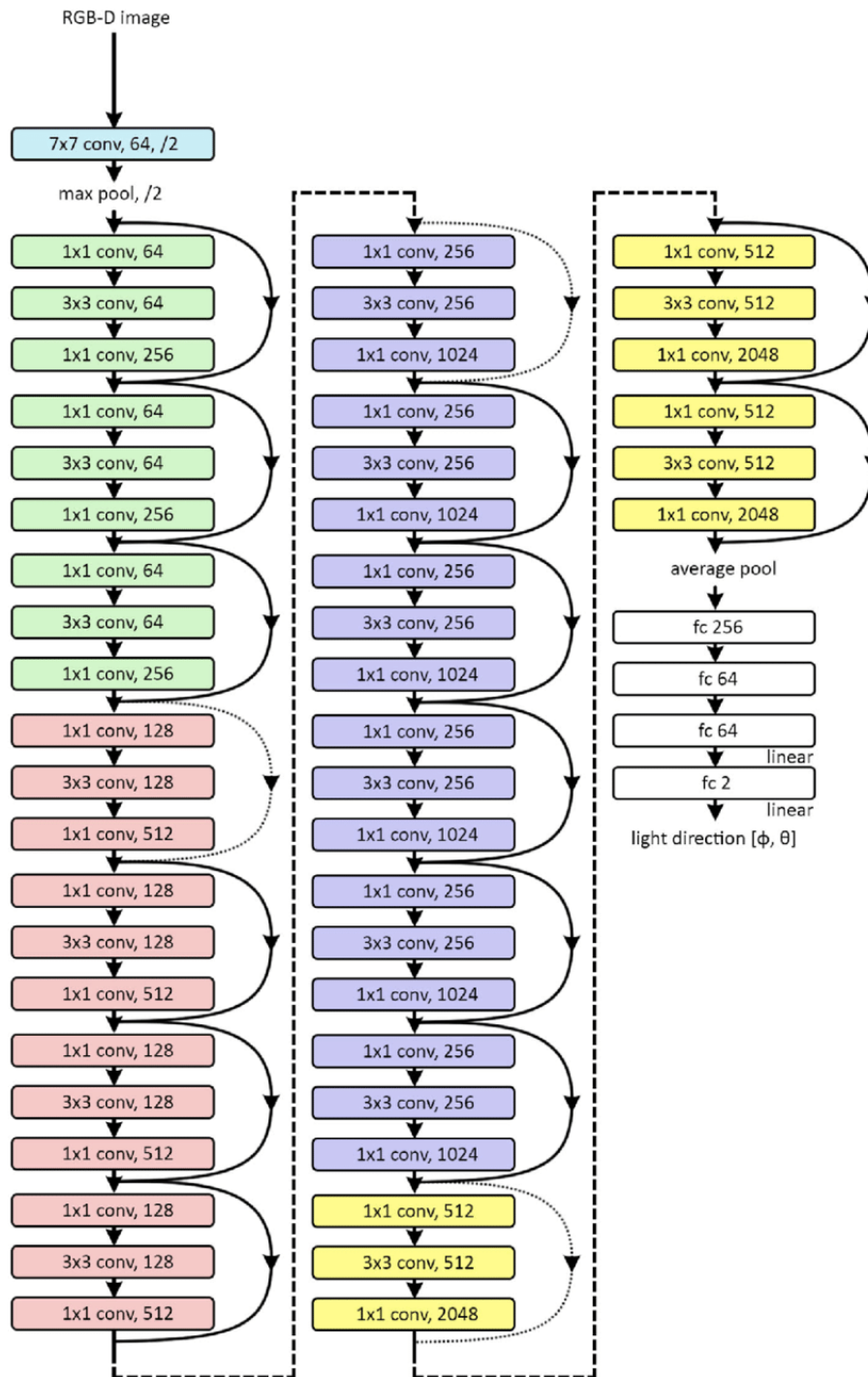


Figure 3: Darknet53 architecture³

214 **YOLOv5s**

215 The backbone structure, the part of the neural network that extracts features, utilized
 216 by YOLOv5s is inspired by CSPDarknet53 (Feng et al., 2023; Jocher & Ultralytics, 2020).
 217 Additional convolution layers are used to extract the image features to a greater extent.

³Extracted from Kán and Kaufmann, 2019

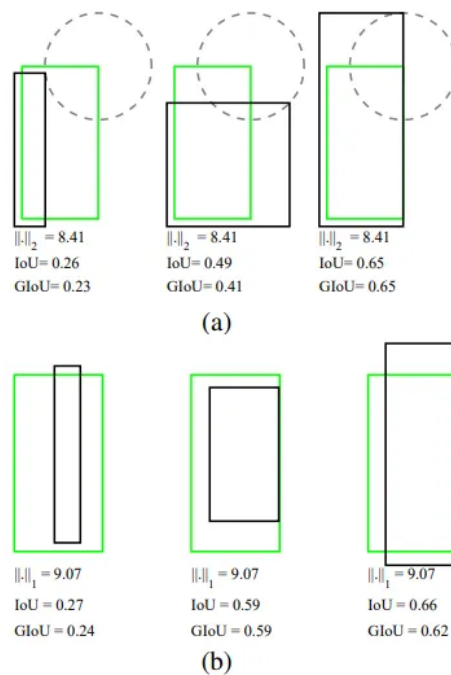


Figure 4: Bounding boxes⁴

To improve its performance, YOLOv5s records its mistakes which can fall into one of three categories: bounding box regression loss, where the bounding box is placed away from the correct spot from where the object is located, as seen in Figure 4; target confidence loss, where YOLOv5s claims an object is apparent when in reality, it's not; and classification loss, when an object is mislabeled.

YOLOv8s

YOLOv8s uses a backbone inspired by CSP. When the channels of the feature map are created, instead of CSP splitting the channels up, YOLOv8s uses Cross Stage Partial with Fused Layers (C2f). C2f splits the feature map channels and fuses, or combines, the features from the channels progressively. By gradually fusing channels, the object detector is able to grab richer details rather than fusing all the channels in one shot like CSP does. A Spatial Pyramid Pooling-Faster (SPPF) module, a component in the CNN, is used to see objects in different sizes. Essentially, it uses windows – parts of the image – of different sizes so then, it's able to grab minute details as well as the bigger picture. For example, a small window can be used to identify that a dog's fur is more wavy than straight. By contrast, a big window would look at the dog entirely, and identify that the object is in the shape of a dog.

In the head of YOLOv8s, Upsample layers (U layers in Figure 5) are used to improve the

⁴Extracted from "Bounding Box Regression Loss", n.d.

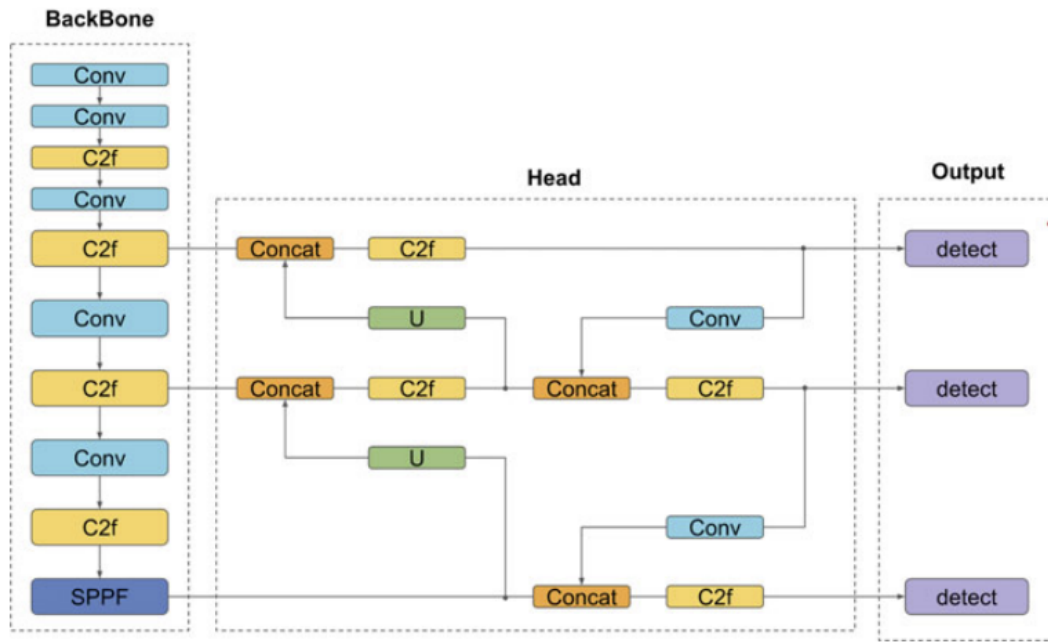


Figure 5: YOLOv8 architecture⁵

236 resolution of the feature maps. Doing this helps preserve more details, which can result
 237 in greater precision and accuracy. Three separate branches, or pathways, are used for
 238 objectness, predicting if an object is apparent; classification, identifying what object that is;
 239 and regression, predicting where the bounding box should be placed.

240 Methodology

241 Standard evaluation measures

242 Very often, when analyzing object detection models, the common metrics utilized are
 243 accuracy, precision and recall. Sometimes, though more rarely, the F1 scores. Accuracy
 244 measures the percentage of predictions that are correct. Accuracy, given in Equation (1),
 245 is the most popular measure for classification and is perfect for cases where making any
 246 mis-classification is equally bad. All misclassifications hurt the measure the same way.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

247 In the above, equation, True Positive (TP): Object detector successfully detects E-waste.
 248 True Negative (TN): When a non-E-waste is identified by the object detector successfully
 249 as non-E-waste. False Positive (FP): When non-E-waste is identified by the object detector

⁵Extracted from Jain et al., 2025

incorrectly as E-waste. False Negative (FN): When E-waste is identified by the object detector wrongly as non-E-waste. 250 251

Precision, given by Equation (2), is also called the positive predictive value and measures how many were correct, among all those predicted positive. In other words, it does not care about false negatives (FNs). False positives (FPs) alone, hurt the measure. This is very useful in applications such as resume screening, where even missing out a good resume is not as bad as hiring a person who is not the right fit. 252 253 254 255 256

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

Recall, given in Equation (3), is also called Sensitivity or the True positive rate. It asks how many the model correctly detected out of all the objects that were actually positive, that is in our case, E-waste. It tells us the total detected E-waste fraction among all E-wastes. In other words, it does not care about FPs and only FNs hurt the measure. This is perfect for applications such as cancer screening. In our case though, we do care about FPs as well, just less so. That is, labeling a tennis ball as E-waste, that is FP, is bad, but not nearly as bad as labeling a battery as a non-E-waste, that is FN. Moreover, recall can be trivially maximized by labeling every item an E-waste. Hence recall can never be used as an objective when training ML methods. 257 258 259 260 261 262 263 264 265

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

Finally, the F1-score, Equation (4) balances precision and recall, and is relevant in applications where both are equally expensive. 266 267

$$F_1 = 2 \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

Reduced FN models 268

Appending to the list of models described in the *Models* section, we create a duplicate of every model, but lower the confidence levels from 0.5 to 0.25, favoring detection. By lowering the confidence level, the number of FNs is guaranteed to decrease since it increases the overall positivity rate. Confidence level is the minimum confidence required for the object detector to make the final decision that the target object is in the picture. We name these models YRFNv3, YRFNv4, YRFNv5, YRFNv8 respectively. YRFN stands for YOLO with Reduced False Negatives. 269 270 271 272 273 274 275

276 **EMC: A New Performance Metric for E-waste detection**

277 As we have already discussed, each of the usual measures of success used in classification
278 do not lend themselves well in E-waste detection. We seek a new metric that weights the
279 FNs substantially greater than the FPs - but not completely nullifying the FPs. Accuracy, for
280 example, undermines this situation completely, considering both FNs and FPs the same.

281 We seek a measure that gives us an estimate of the total cost of mis-identification. This
282 lends itself well to our case since the actual cost of extra manually sorting a FP should be
283 contrasted against the estimated environmental cost of sending a FN to landfill. These
284 numbers will allow us to combine the FNs and FPs in our new metric.

285 To determine these costs, we looked at the cost of offsetting the environmental conse-
286 quences and the costs of inspection of false positives. Looking at a cost benefit analysis
287 literature (Yang et al., 2021) it was highlighted that the cost to offset the environmental
288 consequences was on average 4 USD/kg of E-waste, in 2021. Adjusted for inflation, this is
289 5.95 USD/kg today (2025). Averaging common trash product weights (Empa - Swiss Federal
290 Laboratories for Materials Science and Technology, 2025), we obtain 11.71 USD/item. To
291 find the cost of false positives, we take the average worker's hourly salary, and assuming
292 they spend five minutes on each object, we evaluate the appropriate cost. According to
293 (Wikipedia contributors, 2024), the average recycling worker's hourly salary in China is
294 \$0.17, while it is 32,000/year according to ZipRecruiter in the US. Though a wide variation,
295 averaging five objects per minute, give us 0.67 USD/item. These are rough estimates and
296 hence it is important to note that this is only for comparison purposes. Our new method's
297 calibration and performance do not depend on the exact values, only that we would need
298 non-zero values for both.

299 Below is the new metric that we propose, the E-waste misclassification cost (EMC):

$$\text{EMC} = \frac{11.71 FN + 0.67 FP}{TP + FN} \quad (5)$$

300 **Ensembles**

301 An ensemble method is the combining of multiple models to make a prediction - wisdom of
302 the crowd. Different models may give different results on whether they're able to detect an
303 object or not. To make the final decision, a voting system is used such that if more models
304 claim that an object is apparent, then the final prediction is that the object is present and
305 vice-versa.

306 In this study, we created an ensemble of YOLOv3, YOLOv4, YOLOv5, YOLOv8 and
307 another ensemble of RFNv3, YRFNv4, YRFNv5, YRFNv8. If at least three of the models
308 detected E-waste as present, then the final decision would be that E-waste is there. We

call the EYOLO and EYRFN, these estimate labels by polling the 4 YOLOs and the 4 YRFNs, 309
respectively. 310

WISE: Waste-focused Integrated Smart Ensemble 311

Consider an ensemble that does not allow equal weighting. That is, the wisdom of the 312
crowd is not equally weighted. Better an underlying method, the more its weight. Further, 313
we will let machine learning pick the weights that minimize the total EMC as defined by 314
Equation (5). 315

Note that since we are learning these weights to minimize a loss, we will have to train 316
the weights on a training subset of the data and test it on the rest of the test set. We will 317
use a standard randomization of 80/20 to split the data for testing versus training. For 318
comparisons to other methods above, that did not require training, we will report and 319
compare the averages over the entire dataset. This obviously risks that the performance 320
improvements could be due to over-fitting. To ensure that this is not the case, we will 321
compare the EMC of the test set to that of the training to ensure that magnitudes are 322
similar. 323

Dataset 324

A mixture of multiple datasets was used to test. We did not pick images that were used for 325
original YOLO training. We utilized Kaggle’s E-waste dataset which consisted of mobile 326
phones, microwaves, keyboards and mice. There were 300 images of each giving a total of 327
1200 images of E-waste. Another 1200 images of non-E-waste were mixed with the E-waste 328
images. These non-E-waste images came from images.cv and roboflow. They all fell under 329
various categories like handbags, chairs, spoons, books, racquets, and umbrellas. Each 330
model was given 2400 images for testing, 1200 consisting of E-waste and 1200 consisting 331
of non-E-waste. To evaluate, the number of true positives, false positives, true negatives, 332
and false negatives were counted. 333

All datasets, preprocessing scripts, and trained models are publicly available at our 334
GitHub repository (GitHub Contributors, 2025). 335

Results 336

For each of the methods: classical - YOLOv3, YOLOv4, YOLOv5, YOLOv8; tweaked - YRFNv3, 337
YRFNv4, YRFNv5, YRFNv8; ensemble - EYOLO, EYRFN; Tables 1a to 5b show the confusion 338
matrices. On these tables, each of the columns denote ground truth (hence summing to 339

340 1200), while the rows denote model predictions. As one can observe the traditional YOLOs,
 341 though impressive for object detection, do not fair too well for the purposes of E-waste
 342 detection. More specifically YOLOv3 identifies every object as a non-E-Waste resulting in
 343 1200 TNs and 1200 FNs. This is potentially due to overfitting during the YOLOv3 training.

344 While YOLOv4, YOLOv5 and YOLOv8 perform better, they still weigh FPs and FNs the
 345 same. Very high false negatives (FNs) will essentially cause too many toxic substances to
 346 enter landfills and find their ways into living organisms. The YRFNS each have relatively
 347 lower FNs. What is also surprising is that the cutting edge YOLO, that is version 8, which is
 348 the go-to algorithm today for object detection, is far from the best for E-waste detection.

Predicted	True	
	E-waste	Not E-waste
E-waste	0	0
Not E-waste	1200	1200

1a: YOLOv3 Confusion Matrix

Predicted	True	
	E-waste	Not E-waste
E-waste	0	0
Not E-waste	1200	1200

1b: YRFNv3 Confusion Matrix

Predicted	True	
	E-waste	Not E-waste
E-waste	756	4
Not E-waste	444	1196

2a: YOLOv4 Confusion Matrix

Predicted	True	
	E-waste	Not E-waste
E-waste	851	11
Not E-waste	349	1189

2b: YRFNv4 Confusion Matrix

Predicted	True	
	E-waste	Not E-waste
E-waste	209	5
Not E-waste	991	1195

3a: YOLOv5 Confusion Matrix

Predicted	True	
	E-waste	Not E-waste
E-waste	373	14
Not E-waste	827	1186

3b: YRFNv5 Confusion Matrix

349 Table 6 shows the confusion matrix for newly developed smart ensemble - WISE (Waste-
 350 focused Integrated Smart Ensemble). It has the lowest FNs amongst all methods compared.
 351 The learned weights were 0 for six of the 8 method (4 YOLOs + 4 YRFNs). The only two
 352 non-zero weights were YRFNv4: 0.6 and YRFNv8: 0.4 with threshold almost just above zero.

353 A final comparison of all methods in terms of the popular measures of success and our
 354 EMC measure are shown in Table 7. As one can see, the smart ensemble out performs every
 355 other method significantly. Moreover the magnitude of EMC is in USD per object and can
 356 hence easily lend itself to a relatable intuitive value. Table 7, also bolds the best values for
 357 each measure of success.

358 Complete confusion matrices and additional performance breakdowns are provided in
 359 our GitHub repository (GitHub Contributors, 2025).

Predicted	True	
	E-waste	Not E-waste
E-waste	176	6
Not E-waste	1024	1194

4a: YOLOv8 Confusion Matrix

Predicted	True	
	E-waste	Not E-waste
E-waste	340	22
Not E-waste	860	1178

4b: YRFNv8 Confusion Matrix

Predicted	True	
	E-waste	Not E-waste
E-waste	90	0
Not E-waste	1110	1200

5a: EYOLO Confusion Matrix

Predicted	True	
	E-waste	Not E-waste
E-waste	203	0
Not E-waste	997	1200

5b: EYRFN Confusion Matrix

Conclusion

360

The lower the EMC, the better the model is at detecting E-waste and reducing overall cost. 361
 Though the measure, EMC, has been weighs the estimated cost of FNs against that of FPs, 362
 the goal is not primarily cost reduction. The goal is the reduction of E-waste that goes into 363
 landfills, for any given budget. This is critical especially since the ever growing volume of 364
 E-waste far exceeds the rate at which budgets to tackle E-waste do not grow. 365

Many surprising observations were made in the results section. Classical cutting- 366
 edge methods like YOLOv8, have terrible performance for E-waste detection though their 367
 performance for general object detection is commendable. Further thought, would reveal 368
 that this is not totally surprising since they were trained and evaluated on a generic object 369
 detection context and with accuracy as the primary measure. Hence it might not be 370
 fair to compare, or to use such methods for E-waste detection. Next, tweaked methods 371
 outperform the classical methods whose default parameters would have been carefully 372
 chosen by the authors. This again shows that even parameters calibrated for generic object 373
 detection do not work well for E-waste detection. 374

Finally, our smart ensemble does surprisingly and fortunately well in the context of 375
 E-waste detection. It not only has a significant cost reduction, it also has the lowest number 376
 of FNs in the entire comparison set. While a methodology that reduces FNs to zero while 377
 not labeling every object as FP is ideal, that would be for future research. 378

References

379

Adedeji, O., & Wang, Z. (2019). Intelligent waste classification system using deep learning 380
 convolutional neural network. *Procedia Manufacturing*. [https://doi.org/10.1016/j.](https://doi.org/10.1016/j.promfg.2019.05.086) 381
 promfg.2019.05.086 382

Predicted	True	
	E-waste	Not E-waste
E-waste	906	25
Not E-waste	294	1175

Table 6: WISE Confusion Matrix

Model	TP	FP	TN	FN	Accuracy	Precision	Recall	EMC
YOLOv3	0	0	1200	1200	0.50	NA	0.00	11.71
YRFNv3	0	0	1200	1200	0.50	NA	0.00	11.71
YOLOv4	756	4	1196	444	0.81	0.99	0.63	4.33
YRFNv4	851	11	1189	349	0.85	0.99	0.71	3.41
YOLOv5	209	5	1195	991	0.59	0.98	0.17	9.67
YRFNv5	373	14	1186	827	0.65	0.96	0.31	8.08
YOLOv8	176	6	1194	1024	0.57	0.97	0.15	10.00
YRFNv8	340	22	1178	860	0.63	0.94	0.28	8.40
EYOLO	90	0	1200	1110	0.54	1.00	0.08	10.83
EYRFN	203	0	1200	997	0.58	1.00	0.17	9.73
WISE - Smart(learned)	906	25	1175	294	0.87	0.97	0.76	2.88

Table 7: Performance comparison of different models.

- 383 Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). Yolov4: Optimal speed and accuracy of
384 object detection. <https://arxiv.org/abs/2004.10934>
- 385 Bodla, N., Singh, B., Chellappa, R., & Davis, L. S. (2017). Soft-nms - improving object detection
386 with one line of code. *Proceedings of the IEEE International Conference on Computer*
387 *Vision (ICCV)*, 5561–5569. [https://openaccess.thecvf.com/content_ICCV_2017/](https://openaccess.thecvf.com/content_ICCV_2017/papers/Bodla_Soft-NMS_-_Improving_ICCV_2017_paper.pdf)
388 [papers/Bodla_Soft-NMS_-_Improving_ICCV_2017_paper.pdf](https://openaccess.thecvf.com/content_ICCV_2017/papers/Bodla_Soft-NMS_-_Improving_ICCV_2017_paper.pdf)
- 389 *Bounding box regression loss* [Author not listed; wiki page]. (n.d.). CloudFactory Computer
390 Vision Wiki. Retrieved September 6, 2025, from [https://wiki.cloudfactory.com/](https://wiki.cloudfactory.com/docs/mp-wiki/loss/bounding-box-regression-loss)
391 [docs/mp-wiki/loss/bounding-box-regression-loss](https://wiki.cloudfactory.com/docs/mp-wiki/loss/bounding-box-regression-loss)
- 392 Cheng, X., Cheng, Y., Gong, Z., & Chen, Y. (2021). An improved small object detection method
393 based on yolov3. *Multimedia Tools and Applications*, 80(19), 29329–29344. <https://doi.org/10.1007/s11042-021-10605-5>
- 394
- 395 Domingos, P. (1999). Metacost: A general method for making classifiers cost-sensitive. *Pro-*
396 *ceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery*
397 *and Data Mining (KDD)*, 155–164. <https://doi.org/10.1145/312129.312220>

- Elkan, C. (2001). The foundations of cost-sensitive learning. *Proceedings of the 17th International Joint Conference on Artificial Intelligence (IJCAI)*, 973–978. <https://cseweb.ucsd.edu/~elkan/rescale.pdf> 398 400
- Empa - Swiss Federal Laboratories for Materials Science and Technology. (2025). *Ewaste guide* [Accessed: 2025-08-24]. 401 402
- Fan, J., Cui, L., & Fei, S. (2023). Waste detection system based on data augmentation and yolo_ec. *Sensors*, 23(7), 3646. <https://doi.org/10.3390/s23073646> 403 404
- Feng, M., Jiang, T., Wang, J., Yang, L., & Zhang, Y. (2023). A multi-scale traffic object detection algorithm for road scenes based on improved yolov5. *Sensors*, 23(4), 2045. <https://doi.org/10.3390/s23042045> 405 406 407
- Fornalczyk, A., Willner, J., Francuz, K., & Cebulski, J. (2013). E-waste as a source of valuable metals. *Archives of Materials Science and Engineering*, 63(2). 408 409
- GitHub Contributors. (2025). *Wise* - <https://github.com/vishnumuthuraman/wise.git> [Accessed: 2025-08-24]. 410 411
- Ibrahim, A. N., Youssif, E. S., Hassanien, A. E., Hossam, A. A., Abdo, S. S., & Saleh, A. M. D. (2023). Object detection-based automatic waste segregation using robotic arm. *International Journal of Advanced Computer Science and Applications (IJACSA)*. <https://doi.org/10.14569/IJACSA.2023.0141164> 412 413 414 415
- Jain, H., Jain, J., Shrivastava, V., Pandey, A., & Sharma, S. (2025). Yolov8 and its advancements. *International Journal of Research Publication and Reviews*, 6(5), 660–666. Retrieved September 6, 2025, from <https://ijrpr.com/uploads/V6ISSUE5/IJRPR44604.pdf> 416 417 418
- Jocher, G., & Ultralytics. (2020). *Yolov5*. <https://github.com/ultralytics/yolov5> 419
- Jocher, G., & Ultralytics. (2023). *Yolov8*. <https://github.com/ultralytics/ultralytics> 420
- Kán, P., & Kaufmann, H. (2019). Deeplight: Light source estimation for augmented reality using deep learning. *The Visual Computer*, 35, 873–883. <https://doi.org/10.1007/s00371-019-01666-x> 421 422 423
- Khan, H. A., Naqvi, S. S., Alharbi, A. A. K., Alotaibi, S., & Alkathami, M. (2024). Enhancing trash classification in smart cities using federated deep learning. *Scientific Reports*, 14, 11816. <https://doi.org/10.1038/s41598-024-62003-4> 424 425 426
- Li, K., & Cao, L. (2020). A review of object detection techniques. *2020 5th International Conference on Electromechanical Control Technology and Transportation (ICECTT)*, 385–390. <https://doi.org/10.1109/ICECTT50890.2020.00091> 427 428 429
- Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2999–3007. <https://doi.org/10.1109/ICCV.2017.324> 430 431 432
- Majchrowska, S., Mikołajczyk, A., Ferlin, M., Klawikowska, Z., Plantykw, M. A., Kwasigroch, A., & Majek, K. (2022). Deep learning-based waste detection in natural and urban environments. *Waste Management*, 138, 274–284. <https://doi.org/10.1016/j.wasman.2021.12.001> 433 434 435 436
- Meituan Vision AI Department. (2022). *Yolov6: A single-stage object detection framework for industrial applications*. <https://github.com/meituan/YOLOv6> 437 438

439 Noel-Brune, M., Goldizen, F. C., Neira, M., van den Berg, M., Lewis, N., King, M., Suk, W. A.,
440 Carpenter, D. O., Arnold, R. G., & Sly, P. D. (2013). Health effects of exposure to e-waste.
441 *The Lancet Global Health*, 1(2), e70. [https://doi.org/10.1016/S2214-109X\(13\)70020-2](https://doi.org/10.1016/S2214-109X(13)70020-2)

442 Oza, P., Agrawal, S., Kapadia, M., & Raotole, O. (2025). Solid waste classification using deep
443 neural network: A transfer learning approach. *Earth Science Informatics*, 18, 255.
444 <https://doi.org/10.1007/s12145-025-01743-x>

445 Parti, A. (2024). Understanding object detection: A comprehensive guide [Retrieved March
446 18, 2024, from <https://pareto.ai/blog/object-detection>].

447 Paudel, P., Shrestha, S., Shrestha, S., Gurung, S., & Adhikari, S. (2024). Automated waste
448 sorting with delta arm and yolov8 detection. *Journal of Artificial Intelligence and
449 Capsule Networks*, 6(3), 299–315. <https://doi.org/10.36548/jaicn.2024.3.004>

450 Potter, R. (2022, August 23). How bounding box object detection is taking over and what
451 to do about it [Analytics on Medium]. Retrieved September 6, 2025, from <https://medium.com/analytics/how-bounding-box-object-detection-is-taking-over-and-what-to-do-about-it-9130b6d6c1f1>

452
453

454 Proença, P. F., & Simões, P. (2020). TACO: Trash annotations in context for litter detection.
455 *arXiv*. <https://arxiv.org/abs/2003.06975>

456 Proença, P. F., & Simões, P. (2023). TACO dataset website [Accessed 2025-09-05].

457 Promboonruang, S., Boonrod, T., & Ratchatawetchakul, Y. (2024). Efficient waste detection
458 and classification based on yolov5 models. *Engineering Access (MIJET)*, 10(1), 51–58.
459 <https://doi.org/10.14456/mijet.2024.7>

460 Rahim, L. A. A., Abidin, N. A. Z., Aminuddin, R., Samah, K. A. F. A., Ibrahim, A. Z. M., Yusoh, S. D.,
461 & Nasir, S. D. N. M. (2024). YOLOv5 model-based real-time recyclable waste detection
462 and classification system. In *Innovations in smart cities applications volume 7 (sca*
463 *2023)* (pp. 44–54, Vol. 906). Springer, Cham. [https://doi.org/10.1007/978-3-031-](https://doi.org/10.1007/978-3-031-53824-7_5)
464 [53824-7_5](https://doi.org/10.1007/978-3-031-53824-7_5)

465 Rajeev, P. A., Dharewa, V., Lakshmi, D., Vishnuvarthanan, G., Giri, J., Sathish, T., & Alrashoud,
466 M. (2025). Advancing e-waste classification with customizable YOLO-based deep
467 learning models. *Scientific Reports*, 15, 18151. [https://doi.org/10.1038/s41598-025-](https://doi.org/10.1038/s41598-025-94772-x)
468 [94772-x](https://doi.org/10.1038/s41598-025-94772-x)

469 Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-
470 time object detection. *Proceedings of the IEEE Conference on Computer Vision and
471 Pattern Recognition (CVPR)*, 779–788. <https://doi.org/10.1109/CVPR.2016.91>

472 Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. [https://arxiv.org/](https://arxiv.org/abs/1804.02767)
473 [abs/1804.02767](https://arxiv.org/abs/1804.02767)

474 Sheng, V. S., & Ling, C. X. (2006). Thresholding for making classifiers cost-sensitive. *Pro-*
475 *ceedings of the AAAI Conference on Artificial Intelligence*, 476–481. [https://cdn.aaai.](https://cdn.aaai.org/AAAI/2006/AAAI06-076.pdf)
476 [org/AAAI/2006/AAAI06-076.pdf](https://cdn.aaai.org/AAAI/2006/AAAI06-076.pdf)

477 Singh, Y., & Parimala S. (2025, March 26). E-waste management. [https://doi.org/10.2139/](https://doi.org/10.2139/ssrn.5196804)
478 [ssrn.5196804](https://doi.org/10.2139/ssrn.5196804)

- Solawetz, J. (2024, January 4). *What is yolov4? a detailed breakdown*. Roboflow. Retrieved 479
September 6, 2025, from [https://blog.roboflow.com/a-thorough-breakdown-of- 480](https://blog.roboflow.com/a-thorough-breakdown-of-yolov4/)
[yolov4/](https://blog.roboflow.com/a-thorough-breakdown-of-yolov4/) 481
- Solovyev, R., Wang, W., & Gabruseva, T. (2021). Weighted boxes fusion: Ensembling boxes 482
from different object detection models. *Image and Vision Computing*, 107, 104117. 483
<https://doi.org/10.1016/j.imavis.2021.104117> 484
- Song, L., Yu, H., Yang, Y., Tong, Y., Ren, S., & Ye, C. (2025). Yolo-vg: An efficient real-time 485
recyclable waste detection network. *Journal of Real-Time Image Processing*. [https: 486](https://doi.org/10.1007/s11554-025-01655-5)
[//doi.org/10.1007/s11554-025-01655-5](https://doi.org/10.1007/s11554-025-01655-5) 487
- Thung, G., & Yang, M. (2017). Trashnet [Accessed 2025-09-05]. 488
- Vats, M. C., & Singh, S. K. (2014). E-waste characteristic and its disposal. *International Journal* 489
of Ecological Science and Environmental Engineering, 1(2), 49–61. 490
- Wang, C.-Y., Bochkovskiy, A., & Liao, H.-Y. M. (2022). Yolov7: Trainable bag-of-freebies sets 491
new state-of-the-art for real-time object detectors. [https://arxiv.org/abs/2207. 492](https://arxiv.org/abs/2207.02696)
02696 493
- Wikipedia contributors. (2024). *Electronic waste in Guiyu* [Wikipedia article]. Retrieved 494
August 24, 2025, from [https://en.wikipedia.org/wiki/Electronic_waste_in_Guiyu 495](https://en.wikipedia.org/wiki/Electronic_waste_in_Guiyu)
- Wu, P., Chai, B., Nie, X., Yan, L., Wang, Z., Zhou, Q., Wang, B., Peng, Y., & Li, H. (2024, June 21). 496
Enhanced object detection: A study on vast vocabulary object detection track for 497
v3det challenge 2024. <https://doi.org/10.48550/arXiv.2406.09201> 498
- Wu, T., & Zhu, S.-C. (2013). Learning near-optimal cost-sensitive decision policy for object 499
detection. *Proceedings of the IEEE International Conference on Computer Vision* 500
(ICCV), 753–760. <https://doi.org/10.1109/ICCV.2013.98> 501
- Yang, W.-D., Sun, Q., & H-G., N. (2021). Cost-benefit analysis of metal recovery from e- 502
waste: Implications for international policy. *Waste Management*, 123, 42–47. [https: 503](https://doi.org/10.1016/j.wasman.2021.01.023)
[//doi.org/10.1016/j.wasman.2021.01.023](https://doi.org/10.1016/j.wasman.2021.01.023) 504

Response Letter

1

Dear Editor,

2

We are glad that you believe our paper has a decent chance at successful peer review and we sincerely thank you for the early feedback that has helped me create a stronger version. We are attaching a new revised version to this email. This version broadens the literature review, clarifies image normalization steps and clarifies several parts. Details of additions/modifications follow.

3

4

5

6

7

- **Feedback:** *The phrasing “less fortunate localities” in the first few pages of the paper seems a bit informal and nonspecific. Perhaps the reader could change this wording to something like “socioeconomically disadvantaged” or “underprivileged,” which seem more apt?*

8

9

10

11

Response: We agree that this is too informal and have formalized the language. It now reads:

12

13

“The ethical issues involved in dispatching the E-waste from developed nations to socioeconomically disadvantaged nations cannot be overstated. In developed nations, these issues and the serious medical ailments caused by E-waste are often not seen every day and becomes a hidden problem that even conscientious citizens are not reminded of every day. These underprivileged localities are often the least equipped environments that simply don’t have the medical resources to address the issues created by these imported toxic substances”.

14

15

16

17

18

19

20

- **Feedback:** *The literature review section in the introduction should be considerably expanded (at least two paragraphs, if not more).*

21

22

* *Has AI been used before in the context of waste classification? Could you give some other literature that offers similar contexts?*

23

24

* *This is mainly because the methods/extensions offered in the paper are not overly mathematically groundbreaking, but the application is very novel and timely. Hence, this paper will seem stronger if more of the relevancy and novelty in the applications are emphasized. There also did not appear to be much critical comparison with prior work on cost-sensitive classification or domain-specific ensembles, perhaps in other similar contexts/applications.*

25

26

27

28

29

30

Response: Thanks. We have added three paragraphs to the Introduction, on pages 3 and 4. This broadens the literature review and makes the paper stronger.

31

32

The first paragraph (Pg 3. “Specifically in terms of...”) reviews the work on generic waste detection and describes the common datasets used there and why those datasets

33

34

35 are not relevant for our purposes). The second paragraph (Pg 3. “YOLO family of
36 detectors...” describes the studies on YOLO detectors for waste identification and
37 the one paper for E-waste identification, that we found. This paper was published in
38 Scientific Reports (Nature) and simply underscores the importance of this problem.
39 The final paragraph (Pg 4. “Prior waste detection...”), summarizes prior work and
40 contrasts our contribution to the prior discussed work.

41 • **Feedback:** *Did you preprocess (e.g., normalize) the data samples before using them? This*
42 *may have been done to create the dataset, and YOLO implementations usually do this*
43 *automatically (I think), but this should be specified and stated in the text.*
44 ** If these were not done automatically by the implementation, you should have imple-*
45 *mented this.*

46 **Response:** Thanks for pointing this out. YOLO implementations include a normaliza-
47 tion step and we have now detailed this in a couple of places in the paper.

48 • **Feedback:** *Why do there appear to be 0 true positives for some results in the confusion*
49 *matrices? This seems perhaps a bit odd.*
50 ** If this is a mistake, then please correct it. But if this is not a mistake, you should explain*
51 *why this is the case.*

52 **Response:** We double-checked the code and results and this not a mistake. The issue
53 was that the simpler YOLOv3, just ended up identifying all objects as non-E-waste. It
54 is just that YOLOv3 was not as good as the later versions of YOLO, for our application.
55 We have added this clarifying explanation to the results section. We thank the editor
56 for pointing this out.

Title: WISE: An Adaptive YOLO Ensemble for Accurate E-Waste Object Detection

Final Recommendation: Accept with Minor Revisions

This paper addresses a timely and socially significant challenge by improving the automated detection of electronic waste (e-waste) using object detection models. The authors emphasize that standard evaluation metrics like accuracy, precision, and recall are inadequate for e-waste detection. To address this, the authors propose a novel cost-sensitive metric, the E-waste Misclassification Cost (EMC), which weights FNs more heavily than FPs based on estimated environmental and labor costs. Besides, the paper also introduces WISE, a smart ensemble method that learns weights for multiple YOLO variants to minimize EMC.

Review Feedback and Recommendations:

The paper makes several valuable contributions. First, the introduction of the EMC metric is both innovative and practical, offering a meaningful way to evaluate models in a context where misclassification costs are asymmetric. This aligns well with real-world priorities in waste management and could influence future research beyond e-waste detection. Second, the application of ensemble methods shows clear empirical benefits, with WISE outperforming all baseline models. The paper is well-structured, the methodology is clearly explained, and the results are presented transparently with confusion matrices and performance comparisons. The emphasis on reducing environmental harm through technical means is commendable and socially relevant.

However, to achieve the standard as an accepted Convergence Journal paper, some revisions are required:

1. **Contextualization of YOLO Architectures:**
The descriptions of the YOLO model are detailed but somewhat mechanical. It would strengthen the paper to briefly contextualize these models within the broader literature, this would help readers understand not just what was implemented, but why these versions were selected.
2. **Clarity on WISE's Learning Process:**
The machine learning process used to train WISE's weights is mentioned but not sufficiently detailed. A brief explanation of the optimization algorithm and how overfitting was mitigated would improve reproducibility and methodological transparency.
3. **Discussion of Limitations:**
The paper would be strengthened by a dedicated section discussing limitations. For instance, the dataset, while useful, is relatively small and may not represent all e-waste scenarios. Generalizability to real-world, cluttered, or occluded environments is not addressed. It is suggested that an Introduction to Limitations subsection be added before the conclusion to discuss dataset scope, cost assumptions, and real-world applicability.
4. **Language and Reference:**

Some sections, particularly the introduction and background, contain long sentences that could be broken down for better readability. Minor grammatical errors and repetitive phrasing occur throughout the whole paper. For the References part, the primary issues are a lack of formatting consistency and several instances of incomplete or incorrect information.

5. Formatting

Some minor issues are showing in the document, such as the display of the paragraph shown as follows:

... more hazardous.

This paper proposes (1) a new metric calibrated to help pick the best waste detection, providing a basis upon which even future algorithms can (2) revisits existing algorithm evaluations for the specific purpose of E-waste. We evaluate several standard, tweaked and ensemble methods; (3) finally a smart ensemble (**WISE: Waste-focused Integrated Smart Ensemble**) are learned using machine learning, to minimize the costs/impact of E-waste. The grand goal of these are to help improve public health and re-usability of metals by enabling more efficient recovery and processing of E-waste.

Keywords: Object detection, You Only Look Once (YOLO), Electronic Waste (E-waste), Artificial Intelligence (AI), Ensemble Methods, Threshold calibration

Introduction

Electronic waste, or E-waste, consists of numerous hazardous materials such as lead, mercury, cadmium, arsenic, etc. (Vats & Singh, 2014). E-waste that

Also, the tables displayed in the reference part:

References 379

Adediji, O., & Wang, Z. (2019). Intelligent waste classification system using deep learning convolutional neural network. *Procedia Manufacturing*, *https://doi.org/10.1016/j.promfg.2019.05.086* 380
381
382

Predicted	True	
	E-waste	Not E-waste
E-waste	906	25
Not E-waste	294	1175

Table 6: WISE Confusion Matrix

Model	TP	FP	TN	FN	Accuracy	Precision	Recall	EMC
YOLOv3	0	0	1200	1200	0.50	NA	0.00	11.71
YRFNv3	0	0	1200	1200	0.50	NA	0.00	11.71
YOLOv4	756	4	1196	444	0.81	0.99	0.63	4.33
YRFNv4	851	11	1189	349	0.85	0.99	0.71	3.41

The author could check the required submission format guidelines provided by the Convergence journal, especially on the citation styles and other formatting requirements, spacing, pictures, etc.

WISE: An Adaptive YOLO Ensemble for Accurate E-Waste Object Detection

Reviewer: [Redacted by Managing Editor]

Originality & Significance – Does the paper contribute new insights or perspectives to the field?

- The paper aims to develop a new AI-based model to better identify E-waste to support safer disposal and better reuse. The model developed, WISE, demonstrates, albeit with no statistical evidence, that it is superior to models that currently exist.

Clarity & Structure – Is the argument well-organized and easy to follow? Are ideas clearly presented?

- The Introduction is written very well and is clear in building the rationale for the paper.
- The statement “Research on these two trash related applications, even though not directly related to E-waste, demonstrate the unacceptable performances of generic methods for specific application domains” (lines 85-87) after discussing the two other datasets being used for trash classification, seems unsubstantiated. In fact, these seem like great proof-of-concept trash classification systems being used in a different field where learnings can be applied for developing one specifically for E-waste.
- “There are no proposed methodology improvements or enhancements specifically for detecting E-waste.” (lines 108-109) It is unclear whether this is in relation to the prior E-waste detection studies for accuracy, precision, recall or differentiating costs of FP and FN? This line needs to mention what is exactly missing for the E-waste context.
- The end of the Introduction becomes a little confusing in that it is unclear what the actual purpose of the paper is. Lines 97-105 discuss developing WISE, but then lines 113-119 discuss other aims/contributions of the study. It would be best to consolidate these into one paragraph to build up the rationale for the study. Further lines 277-284 also discuss the purpose/aims of the paper. All of these paragraphs need to be consolidated so that there is a clear aim/s of the paper in the one place.
- Research papers usually have all the background information required to understand the paper in the Introduction. This paper has subheadings and subsequent information for “Object Detection” and “Models” etc after the Introduction. Authors should consider consolidating this within the Introduction to fit a standard manuscript structure. Similarly, this would warrant the removal of paragraph starting from line 120-126 as this is not needed for a manuscript as the structure should be self-explanatory.
- All figure legends could be improved by provide more information eg. Figure 1: Object Detection. An example of use of object detection is shown where the yellow

boxes indicate X, the blue boxes indicate Y, etc. This information should be supported by what is written in the text.

- “The benefits offered by improved one-stage methods” (lines 156-157) is unsubstantiated as the benefits have not been described.
- Mention of Figure 2 in-text should likely come during discussion of two-stage and one-stage CNNs i.e. paragraph starting line 150. Currently it is being mentioned when talking about YOLO but the figure has no reference to YOLO. This is the case for Figure 4 Bounding boxes as well – perhaps should come at their first mention?
- I am unsure whether such detailed descriptions of each YOLO is required. Perhaps this could be displayed in a table with dot points instead to highlight the similarities and differences between the four described?
- I feel tables 1a-5b could be better presented by having all in the one table, with the models on the left and the results in columns. This would be more intuitive for comparisons between the groups. Having read on, I see this has been completed in Table 7 which makes tables 1a-5b redundant, and therefore tables 1a-5b should be removed.

Use of Evidence & Research Methods – Are sources appropriately cited? Is their methodology sound and well-explained?

- The paper uses a lot of evidence to support why the specific models were chosen to be tested in this study.
- There is no use of statistical analyses to determine whether findings are statistically significant which means that it is not possible to say whether WISE is better or not than what already exists at identifying E-waste.

Engagement with Literature – Does the paper demonstrate an understanding of relevant research in the field? Do they acknowledge known results and connect their findings well to them?

- There is good engagement with published literature throughout the paper, until it comes to the Conclusion. The Conclusion requires much more support from literature to place the findings of this study in the wider context of the literature landscape for this topic area. Authors should find literature that supports their findings or describes how the results are surprising or differ from previously published studies.

Grammar & Language

- There are minor grammatical errors throughout e.g. at the start of the Introduction, paragraph starting line 150, paragraph starting line 285. A readthrough of the entire paper, perhaps by a supervisor, is suggested.
- Abbreviations are used without prior definition which can make the paper difficult to follow e.g. FNs in Abstract (line 8), YOLO (line 90).

- There are some concerns around the use of jargon making the paper hard to follow. Authors should consider defining specific terms at their first use to support ease of reading. This is especially crucial as understanding these terms is critical to understanding the premise of the paper. Some examples that require defining include “bounding boxes”, “YOLO”, “performing ensemble”.

Final Recommendation:

- Accept with major revisions (acceptance conditional on satisfactory **major** revisions)

WISE: An Adaptive YOLO Ensemble for Accurate E-Waste Object Detection

September 30, 2025

Abstract 1

Electronic waste (E-waste) - which contains hazardous chemicals (e.g., lead, mercury) and valuable 2
precious metals - has increased tremendously over the past few decades. E-waste reached 62 million metric 3
tons in 2022 and continues to accelerate. Manual sorting is unable to keep up. Only 22.3% was documented 4
as collected and recycled, leaving the rest to enter landfills. There is an urgency for automated techniques 5
to help. Object detection algorithms, traditionally, focus on improving accuracy but end up weighing false 6
negatives and false positives the same. However here, False Negatives (FNs) - like batteries being labeled as 7
non-E-waste are far more hazardous. 8

This paper has three contributions. First, we develop a new metric calibrated to help pick the best 9
algorithm for E-waste detection. This metric provides a basis upon which even future algorithms can be 10
evaluated. Second, we revisit existing algorithm evaluations for the specific purpose of E-waste detection. 11
We evaluate several standard, tweaked and ensemble methods. Finally, we propose a smart ensemble (**WISE:** 12
Waste-focused Integrated Smart Ensemble) whose weights are learned using machine learning, to minimize 13
the costs/impact of E-waste disposal. The grand goal of these is to help improve public health and reusability 14
of precious metals by enabling more efficient recovery and processing of E-waste. 15

16 *Keywords:* Object detection, You Only Look Once (YOLO), Electronic Waste (E-waste), Artificial
17 Intelligence (AI), Ensemble Methods, Threshold calibration

18 **Introduction**

19 Electronic waste, or E-waste, consists of numerous hazardous materials including plastics, lead,
20 mercury, cadmium, arsenic, etc. (Vats & Singh, 2014). E-waste that ends up in landfills is responsible
21 for multiple health hazards, especially in developing countries. Examples of these hazards include
22 fetal loss, prematurity, low birth weight, abnormal thyroid function, neurobehavioral disturbances,
23 and genotoxicity (Noel-Brune et al., 2013). Additionally, the E-waste that ends up in landfills
24 consists of valuable metals – aluminum (Al), gold (Ag), palladium (Pd), platinum (Pt) (Vats &
25 Singh, 2014) – that can be extracted and reused, saving money. On just a single TV board, 7% of
26 its salvage value comes from silver, 33% from gold, and 7% from palladium (Fornalczyk et al.,
27 2013). In a mobile phone, 11% comes from silver, 71% comes from gold, and 11% comes from
28 palladium, totaling 93% of the phone’s salvage value comes from just precious metals (Fornalczyk
29 et al., 2013).

30 In 2022, we generated 62 million metric tons of E-waste and this is expected to rapidly escalate
31 to 74 million metric tons by 2030 (Singh & Parimala S, 2025). Preventing hazardous material from
32 getting into landfills is critical. It is almost the only way to stop these chemicals eventually finding
33 their way into animal and human bodies. Facilities and techniques for extracting the valuable metals
34 from E-waste and disposing of the rest appropriately exist. However, the challenge lies in identifying
35 the E-wastes and keeping up with volume. We currently depend on manual sorting, afforded in part
36 by cheap labor from developing and underdeveloped countries to help prevent hazardous material

from reaching landfills. 37

The ethical issues involved in dispatching the E-waste from developed nations to socioeco- 38
nically disadvantaged nations cannot be overstated. In developed nations, medical ailments 39
caused by E-waste are often not seen every day. As a result, the problem remains hidden and even 40
conscientious citizens rarely notice it. In contrast, underprivileged localities are often the least 41
equipped environments to cope with these imported toxic substances. They also lack the medical 42
resources necessary to tackle the rising health crisis. Many governments have recognized this issue 43
and have attempted to use systems and regulations that cut down on E-waste. Even with such efforts 44
and often challenged funding, growth in E-waste seems to, so far, outpace our efforts to prevent it 45
from entering our landfills and ecosystems. 46

Automated techniques, though not perfect, can be of huge benefit in augmenting manual sorting. 47
Current research has already developed artificial intelligence (AI) models that have been trained 48
and evaluated for generic object-detection. Pictures of labeled objects are processed and used to 49
train neural network models. Training involves the incremental adjustment of the neural network's 50
weights to match the known labels. Once trained, the final network with its adjusted weights is 51
used to label unlabeled objects, that is, detect objects. These neural net models also create bounding 52
boxes for each object and estimate labels for each object. Bounding boxes are tight rectangles 53
around detected objects. This can even facilitate the extraction of E-waste more easily rather than 54
debating whether an object should be classified under this name. 55

The problem with existing algorithms and their benchmarks is that they focus primarily on 56
accuracy or at most precision and recall, as their measures of success. Accuracy is well warranted 57
for generic object detection, where both false positives (FP) and false negatives (FN) are equally bad. 58
Other variants like precision and recall are well warranted when only FPs (eg. screening resumes) 59

60 or FNs (eg. screening for cancer) are to be minimized, respectively. From an E-waste perspective,
61 it is less harmful to label a banana peel as E-waste, not allowing it to go to the landfill directly
62 (routing it to manual sorting). It is much more harmful to label a battery, full of toxic chemicals,
63 as non-E-waste, allowing it to directly enter the landfill and the ecosystem. The clear objective
64 would be to allow FPs like identifying the banana peel as E-waste in lieu of allowing FNs like
65 batteries to reach the landfill. Choosing a standard measure like recall will not help either, since
66 it can be maximized by labeling everything as positive, thereby making FNs zero! But that will
67 make automation useless since it will keep over-burdening manual sorting - the exact reason we
68 seek automation/augmentation.

69 Specifically in terms of literature on AI for waste detection, deep learning models have been
70 used for solid waste classification (Adedeji & Wang, 2019; Majchrowska et al., 2022; Oza et al.,
71 2025). The *TrashNet* (Thung & Yang, 2017) data set was created as a part of a student-led project in
72 2017 and has been one of the popular datasets for trash classification benchmarking. However, this
73 data set is not relevant for E-waste classification since the only six classes included are glass, paper,
74 cardboard, plastic, metal and trash. It contains 2527 images and has been used for benchmarking in
75 some research papers (Khan et al., 2024; Rahim et al., 2024). Another popular direction of research
76 has been in identifying *waste in the wild*. The TACO data set (Proença & Simões, 2020, 2023)
77 provides 1500 images of trash in various locations, from the beaches to city streets. These are to
78 be used to train algorithms that can find trash in pictures taken from, say, an automated garbage
79 collector (Fan et al., 2023; Promboonruang et al., 2024; Song et al., 2025). Findings from these two
80 trash-related applications demonstrate that off-the-shelf general-purpose detectors trained on natural
81 image corpora, can perform poorly. These are however great proof-of-concept trash classification
82 systems being used in related applications. Motivated by this, we frame the E-waste identification

problem as a cost-sensitive identification problem and focus on fine-tuning and developing variants 83
that explicitly minimize our costs. 84

YOLO (You Only Look Once) family of detectors, though created in 2016 (Redmon et al., 85
2016), only became popular and became the go-to choice recently. These one-stage detectors 86
have been shown to be very effective specifically for integration into real-time waste sorting, for 87
example with robotic arms (Ibrahim et al., 2023; Paudel et al., 2024). However, in terms of effective 88
sorting methods for identifying E-waste, there exists only one recent paper (Rajeev et al., 2025) that 89
simply compares different YOLO methods in terms of accuracy and the reported poor performances 90
underscore the need for improved methods. 91

Prior waste detection or E-waste detection studies predominantly optimize for accuracy, precision 92
or recall rather than application level costs of mistakes. None, to our knowledge, consider the 93
application level E-waste misclassification costs that differentiate FP and FN. However, in classical 94
Machine Learning, both (a) minimizing expected misclassifications cost (EMC), when costs of 95
FP and FN differ (Domingos, 1999; Elkan, 2001; Sheng & Ling, 2006) and (b) creating better 96
performing ensemble are popular research topics (Bodla et al., 2017; Lin et al., 2017; Solovyev 97
et al., 2021; Wu & Zhu, 2013). 98

Contribution and outline 99

In this paper we, 100

(1) Create a new metric, E-waste Misclassifications Cost (EMC), calibrated to help pick the 101
best algorithm for E-waste detection, providing a basis upon which even future algorithms can be 102
evaluated. EMC weights the FNs substantially greater than the FPs but does not completely nullify 103

104 the FPs. We compute the coefficients based on accounting cost estimates. We seek a measure that
105 gives us an estimate of the total cost of misidentification, lending itself well to our case since the
106 additional cost of manually sorting a FP should be contrasted against the estimated environmental
107 cost of sending a FN to landfill.

108 (2) Revisit existing algorithm evaluations for the specific purpose of E-waste detection. We
109 evaluate several standard, tweaked and ensemble methods. We analyze several algorithms. The
110 first four algorithms are standard and popular implementations (YOLOv3, YOLOv4, YOLOv5,
111 YOLOv8). The second four are our tweaks to the first four to reduce FNs (YRFNv3, YRFNv4,
112 YRFNv5, YRFNv8). We create two more ensembles that we call EYOLO and EYRFN. These
113 estimate labels by polling the 8 previous variants and the 4 reduced FN variants.

114 (3) Finally, we propose a smart ensemble (**WISE: Waste-focused Integrated Smart Ensemble**)
115 whose weights are learned using machine learning, to minimize the costs/impact of E-waste disposal.
116 Ensemble models combine multiple models to make a final recommendation or decision. Since this
117 is a machine learning algorithm, we do in-sample comparisons with other methods and also make
118 in-sample versus out-of-sample comparison to ensure that there is no overfitting.

119 **Object Detection**

120 Object detection is the process of analyzing an image, localizing objects, and classifying what
121 those objects are. It is utilized in numerous applications such as face, pedestrian (Fig. 1) and
122 object detections (Li & Cao, 2020). Convolutional Neural Networks (CNNs), are a specific type of
123 neural networks that play a very significant role in the object detection space. They are essentially
124 classical feed forward neural networks with additional layers called convolutional layers. These

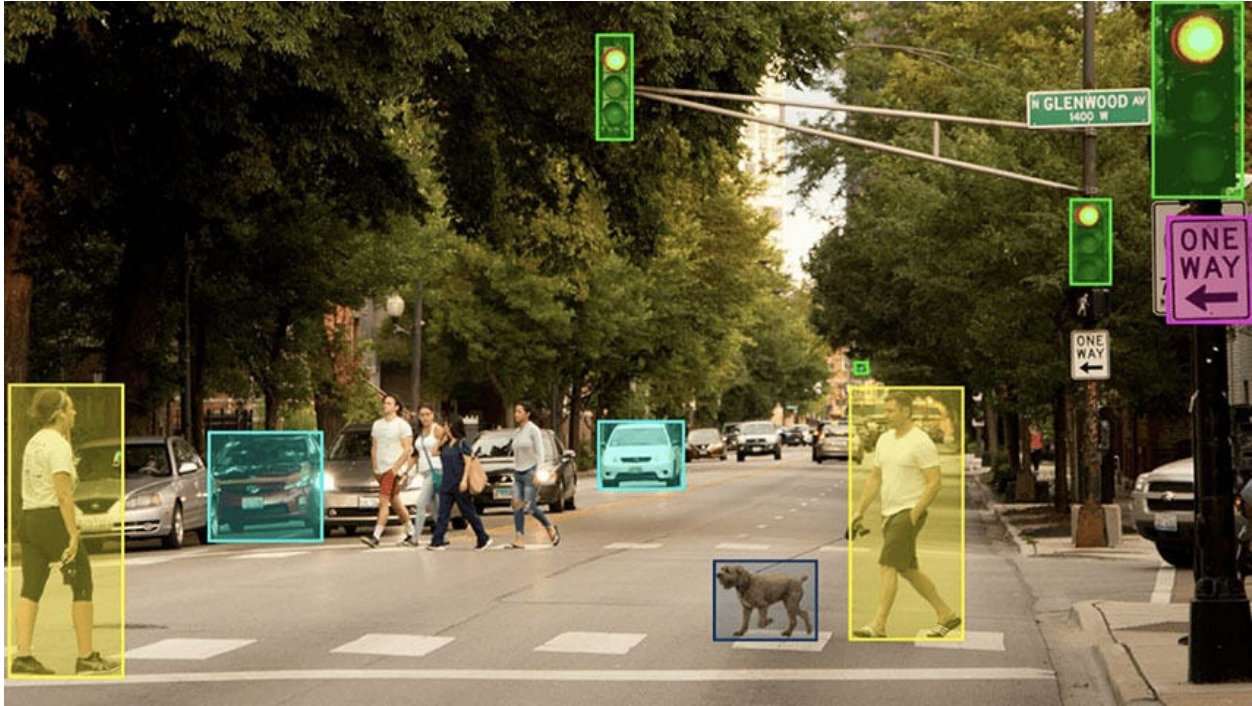


Figure 1: Object Detection example with colored bounding boxes: yellow = person, blue = vehicle, green = traffic light. Boxes illustrate localization outputs prior to non-maximum suppression. ¹

convolutional layers help in any kind of computer vision by applying trained averaging kernels. 125
This kernel performs computations on the image's pixel values, or representations of the color and 126
intensity. By performing these computations, the CNN is able to create a feature map which consists 127
of channels, or representations of the different learned features – lines, edges, curves. 128

When an image is processed to look for certain objects, it is first transformed into a form that can 129
be passed on as an input to a neural network (NN). This means the image must be processed, re-sized 130
and normalized for the neural network to take in the image. All inputs are also normalized to $[0, 1]$ 131
by dividing by 255. A trained CNN can then be used to identify objects using intermediate steps 132
that rely on extracted visual features – edges, corners, textures (Parti, 2024). A Region Proposal 133
Network (RPN), a smaller CNN that utilizes extracted visual features, can be used to create bounding 134
boxes around potential object locations. Objects encapsulated in these bounding boxes are then 135

¹Extracted from Potter, 2022.

136 identified and localization is used to calibrate bounding boxes to exclude unnecessary parts of
137 the image. Non-maximum suppression (NMS), a mathematical algorithm, keeps only the most
138 confident detections, excluding false positives and improves accuracy. The bounding boxes are
139 further refined and object identification is finalized (Parti, 2024).

140 There are two main categories of CNN object detectors: two-stage and one-stage (Wu et al.,
141 2024). In the two-stage detector (Figure 2), the algorithm first makes object proposals — guessing
142 where the object appears in the image. The second stage classifies the object and refines the bounding
143 box. By contrast, one-stage detectors completely skip this object proposal step. They go straight
144 to predicting the bounding boxes and classifying objects. Overall, two-stage detectors are more
145 accurate but computationally much more expensive and slower in comparison to one-stage detectors
146 (Wu et al., 2024). One-stage detectors achieve favorable speed/accuracy trade-offs, often 2–5 times
147 faster, making them practical for real-time applications. Prior work quantifies these trade-offs across
148 architectures and scales (Huang et al., 2017; Liu et al., 2016).

149 You Only Look Once (YOLO) is the most popular one-stage detector that uses a CNN archi-
150 tecture. It uses the CNN to predict where the bounding boxes for objects should be placed. It then
151 assigns the probability of each object in a bounding box being an object of specific class, which are
152 called class-probabilities. For example, the probability the image of an animal could be classified
153 as a dog with probability 0.1 and a cat with probability 0.9. YOLO is known for its computational
154 speed and high accuracy. There are numerous versions of YOLO which all have distinct advantages
155 and disadvantages. YOLO will be the primary focus of this paper due to the availability, open-source
156 license and the popularity of this technique.

²Extracted from Solawetz, 2024

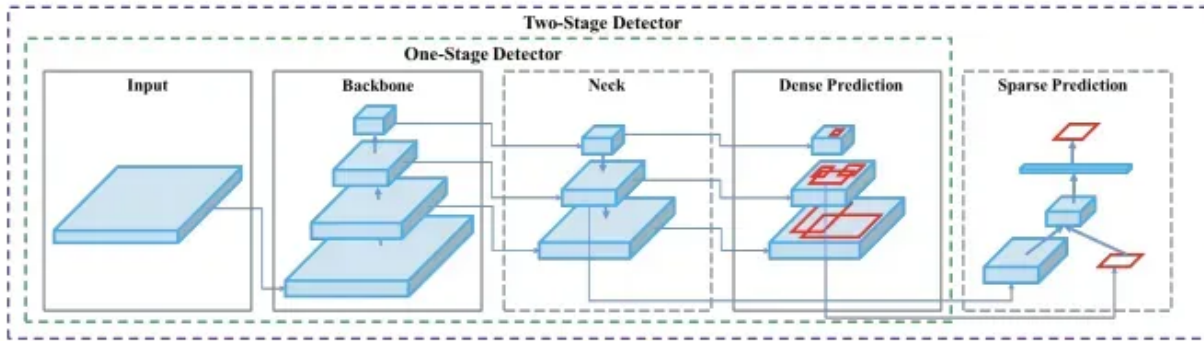


Figure 2: One Stage Vs Two Stage detection: two-stage proposes regions then classifies/refines; one-stage predicts boxes/classes in one pass ²

Models

157

There are eight versions of the YOLO starting with version 1 proposed by (Redmon et al., 2016). 158
 Versions 2 and 3 are improved versions with residual blocks (Redmon & Farhadi, 2018). YOLO 159
 version 4 was released by A. Bochkovskiy, who took over after Redmon retired (Bochkovskiy et al., 160
 2020). It included major performance improvements and new training tricks. 161

Versions 5 (Jocher & Ultralytics, 2020) and Version 8 (Jocher & Ultralytics, 2023) were 162
 developed by Ultralytics, and do not have an official research paper. Version 5 re-implemented 163
 YOLO in PyTorch while Version 8, improved 5, by unifying the code base and modernized the 164
 architecture. YOLO version 6 was developed by Meituan (Meituan Vision AI Department, 2022), 165
 was optimized for industrial deployment and does not have a formal research paper either. YOLO 166
 version 7 (Wang et al., 2022) introduced several new features like extendable trainable bag-of- 167
 freebies and architectural refinements. The most popular among the YOLOs are Version 3 (last 168
 official Redmon release), Version 4 (huge leap in accuracy), Version 5 (industry standard, ease of 169
 use) and Version 8 (latest, cutting edge, unified framework). 170

171 **YOLOv3**

172 The backbone structure, the part of the neural network that extracts features, utilized by YOLOv3s
173 is Darknet-53 (Cheng et al., 2021; Redmon & Farhadi, 2018). Darknet53 is a CNN with 53 layers
174 which uses residual connections (Figure 3), allowing for input to skip over layers to improve effi-
175 ciency. Convolution blocks, layers of the CNN that extract features, use a sequence of Convolution,
176 Batch Normalization, and Leaky ReLU, or otherwise called CBL. Convolution uses kernels to detect
177 features, Batch Normalization allows CNNs to stabilize and train faster, and Leaky ReLU – a mathe-
178 matical function – helps the CNN learn complex patterns. All YOLO methods include normalization
179 of each channel to $[0, 1]$ and do not have per-channel mean/standard deviation standardizations.

180 The head structure, or the end of the model, of YOLOv3 predicts where objects are at three
181 scales: small, medium, and large. In other words, this helps the model detect small objects, medium
182 objects, and large objects. Anchor boxes, or boxes that have been predefined at certain sizes and
183 shapes, are placed in every grid cell to facilitate bounding box predictions. Finally, the bounding
184 boxes are placed around the target object(s).

185 **YOLOv4**

186 YOLOv4's backbone structure is CSPDarknet53 (Bochkovskiy et al., 2020). Darknet53 extracts
187 the features while Cross Stage Partial (CSP) splits channels from the feature map such that half the
188 channels are able to use the residual connections and the other half continue through the several
189 layers. This ensures that there's a balance between performance and efficiency. In the neck structure
190 – the part of the object detection model that combines details that have been collected, allowing
191 for a better understanding of the image – of YOLOv4, Spatial Pyramid Pooling (SPP) – a pooling

³Extracted from Kán and Kaufmann, 2019

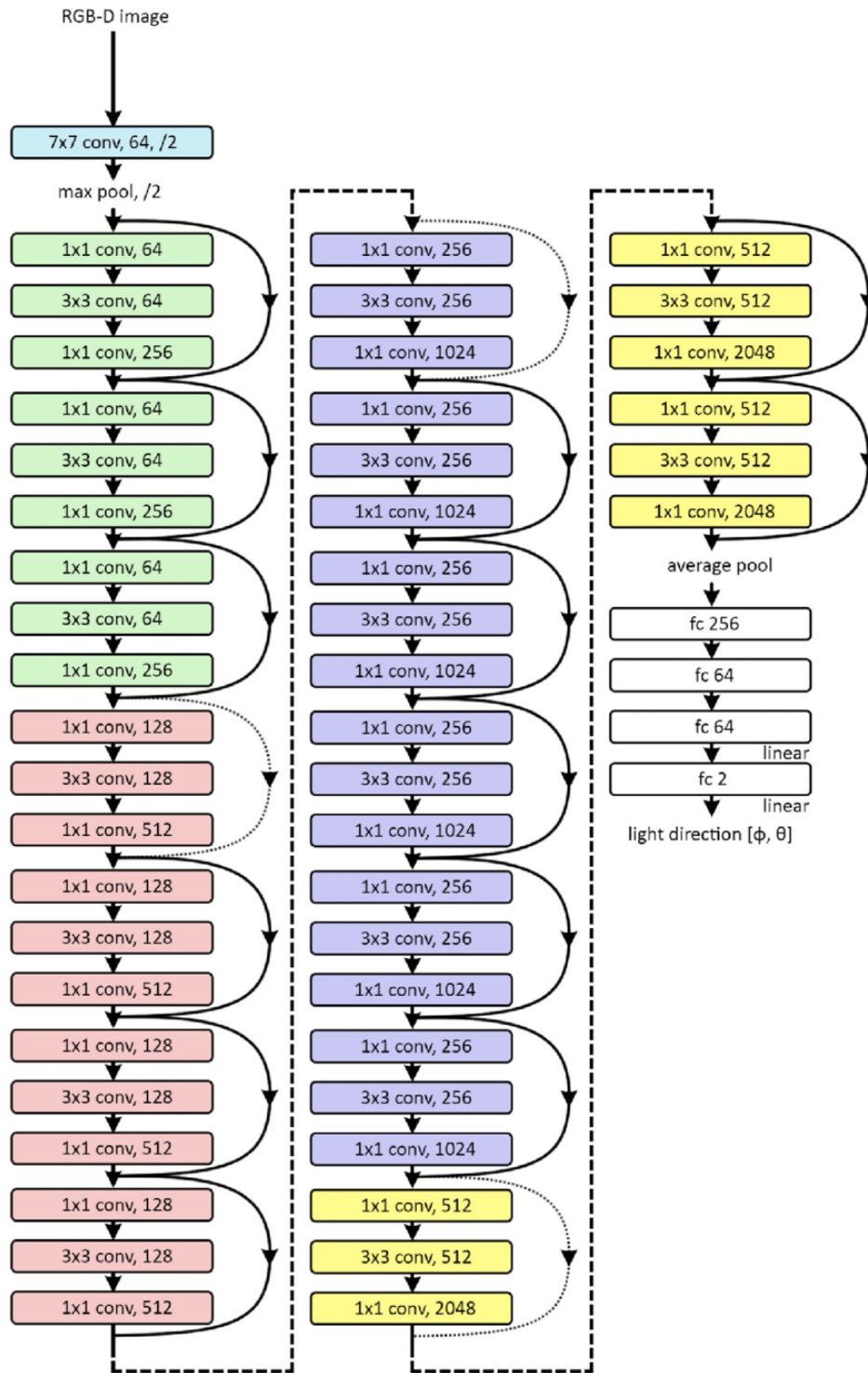


Figure 3: Darknet53 architecture: Deep convolutional network applied to RGB-D input, using residual connections across multiple layers to extract hierarchical features for downstream object detection tasks.³

192 module, reducing the size of the feature map and keeping on essential parts, in the CNN – takes the
193 feature maps that were created from CSPDarknet53, and underscores important patterns. It also
194 looks at the feature maps in different sizes, helping the SPP pick up minute details as well as the
195 larger context. Path Aggregation Network (PAN), a component of the CNN, focuses on locating
196 where an object occurs in the image.

197 The head structure for YOLOv4 is the same head structure used for YOLOv3 due to its efficiency.
198 Although the head structures are the same, YOLOv4 has an improved backbone and neck structure
199 in comparison to YOLOv3.

200 **YOLOv5s**

201 The backbone structure, the part of the neural network that extracts features, utilized by YOLOv5s is
202 inspired by CSPDarknet53 (Feng et al., 2023; Jocher & Ultralytics, 2020). Additional convolution
203 layers are used to extract the image features to a greater extent.

204 To improve its performance, YOLOv5s records its mistakes which can fall into one of three
205 categories: bounding box regression loss, where the bounding box is placed away from the correct
206 spot from where the object is located, as seen in Figure 4; target confidence loss, where YOLOv5s
207 claims an object is apparent when in reality, it's not; and classification loss, when an object is
208 mislabeled.

209 **YOLOv8s**

210 YOLOv8s uses a backbone inspired by CSP. When the channels of the feature map are created,
211 instead of CSP splitting the channels up, YOLOv8s uses Cross Stage Partial with Fused Layers

⁴Extracted from “Bounding Box Regression Loss”, n.d.

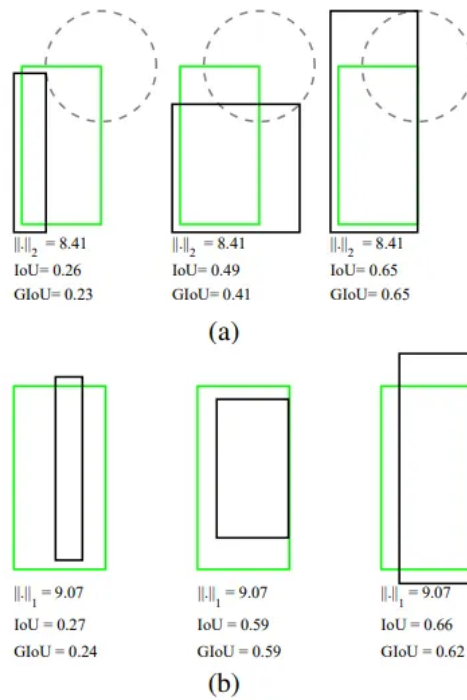


Figure 4: Bounding boxes: Intersection over Union (IoU) and Generalized IoU (GIoU) in bounding box regression, showing how GIoU better accounts for spatial alignment when boxes don't overlap perfectly, improving training stability.⁴

212 (C2f). C2f splits the feature map channels and fuses, or combines, the features from the channels
213 progressively. By gradually fusing channels, the object detector is able to grab richer details rather
214 than fusing all the channels in one shot like CSP does. A Spatial Pyramid Pooling-Faster (SPPF)
215 module, a component in the CNN, is used to see objects in different sizes. Essentially, it uses
216 windows – parts of the image – of different sizes so then, it's able to grab minute details as well as
217 the bigger picture. For example, a small window can be used to identify that a dog's fur is more
218 wavy than straight. By contrast, a big window would look at the dog entirely, and identify that the
219 object is in the shape of a dog.

220 In the head of YOLOv8s, Upsample layers (U layers in Figure 5) are used to improve the
221 resolution of the feature maps. Doing this helps preserve more details, which can result in greater
222 precision and accuracy. Three separate branches, or pathways, are used for objectness, predicting
223 if an object is apparent; classification, identifying what object that is; and regression, predicting
224 where the bounding box should be placed.

⁵Extracted from Jain et al., 2025

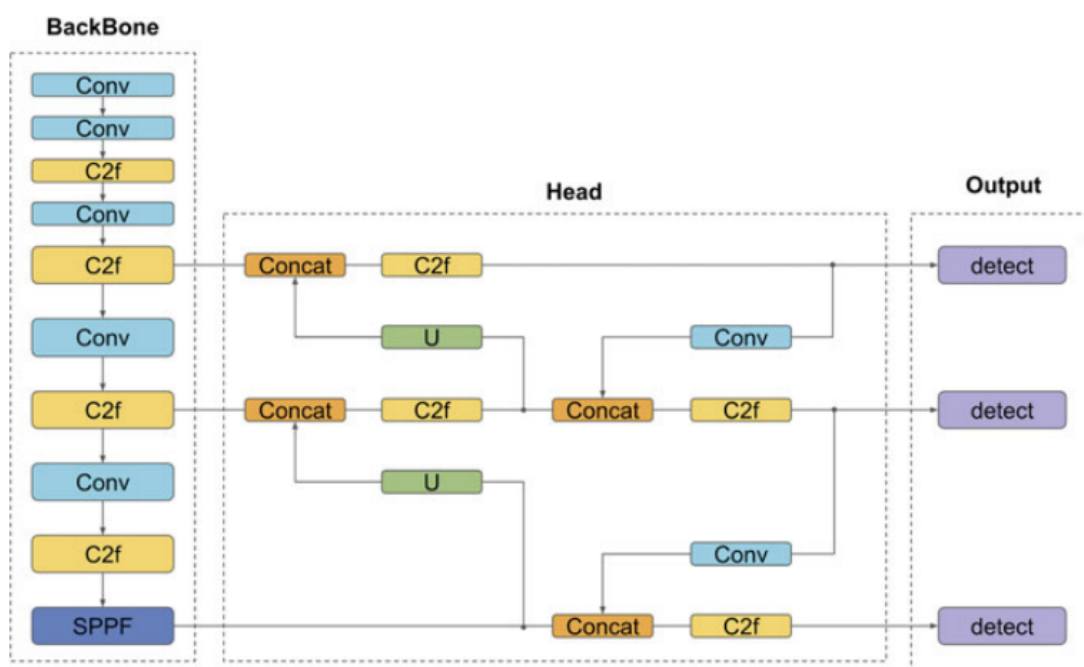


Figure 5: YOLOv8 architecture: the backbone extracts multi-scale features, the head performs feature aggregation, and the output layer generates object detection predictions⁵

Version, Year	Backbone	Neck	Head	Key innovations / training updates	Pros / Cons (esp. for E-waste)
YOLOv3 2018	Darknet-53 (residual)	FPN-like multi-scale fusion	3-scale heads with anchors	Stable baseline; incremental over YOLOv2; residual connections improve depth	+ Robust and widely reproduced baseline. – Weaker recall for small objects, limiting E-waste sensitivity.
YOLOv4 2020	CSPDarknet-53	SPP + PAN	3-scale heads with anchors	Introduced “bag of freebies/costlies,” CIOU/GIOU, mosaic augmentation, stronger training recipe	+ Major accuracy gain, better on small/medium objects. – Higher complexity and compute demand.
YOLOv5s 2020	CSP-inspired (PyTorch)	PANet	3-scale heads with anchors	PyTorch re-implementation; small “s” variant optimized for speed; auto-augment, EMA; easy deployment	+ Flexible, lightweight, edge-ready. – “s” favors speed over recall; in our tests, default thresholds missed subtle E-waste (see YRFNv5).
YOLOv8s 2023	C2f (CSP-fused)	SPPF + PAN	Decoupled obj/cls/reg heads (anchor-free default)	Modernized architecture; anchor-free prediction; cleaner training pipeline; improved feature reuse	+ Strong precision, efficient pipeline. – In our results, default setup under-recalled E-waste; improved when FN-reduction applied (YRFNv8).

Table 1: Concise comparison of YOLO variants evaluated. “s” variants (v5s, v8s) are chosen to match real-time, resource-constrained conveyor-belt settings.

A summary of the comparisons are captured in Table 1. As discussed in the preceding subsections, 225 YOLOv3, v4, v5s, and v8s each represent critical points in the evolution of the YOLO family of 226 detection algorithms. v3 is the canonical baseline. v4 is the first significant performance jump. v5s is 227 the widely adopted PyTorch re-implementation. Finally, v8s is the most recent unified architecture. 228 We pick the “s” variants for v5 and v8 to align better with the real-time, resource-constrained, 229 conveyor-belt based sorting scenario. This spectrum of methods allows our analysis to capture both 230 historical and state-of-the-art behaviors under the E-waste misclassifications cost framework. 231

Methodology 232

Standard evaluation measures 233

Very often, when analyzing object detection models, the common metrics utilized are accuracy, 234 precision and recall. Sometimes, though more rarely, the F1 scores. Accuracy measures the 235 percentage of predictions that are correct. Accuracy, given in Equation (1), is the most popular 236 measure for classification and is perfect for cases where making any misclassification is equally 237 bad. All misclassifications hurt the measure the same way. 238

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

In the above, equation, True Positive (TP): Object detector successfully detects E-waste. True 239 Negative (TN): When a non-E-waste is identified by the object detector successfully as non-E-waste. 240 False Positive (FP): When non-E-waste is identified by the object detector incorrectly as E-waste. 241 False Negative (FN): When E-waste is identified by the object detector wrongly as non-E-waste. 242

243 Precision, given by Equation (2), is also called the positive predictive value and measures how
244 many were correct, among all those predicted positive. In other words, it does not care about false
245 negatives (FNs). False positives (FPs) alone, hurt the measure. This is very useful in applications
246 such as resume screening, where even missing out a good resume is not as bad as hiring a person
247 who is not the right fit.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

248 Recall, given in Equation (3), is also called Sensitivity or the True positive rate. It asks how
249 many the model correctly detected out of all the objects that were actually positive, that is in our
250 case, E-waste. It tells us the total detected E-waste fraction among all E-wastes. In other words,
251 it does not care about FPs and only FNs hurt the measure. This is perfect for applications such as
252 cancer screening. In our case though, we do care about FPs as well, just less so. That is, labeling a
253 tennis ball as E-waste, that is FP, is bad, but not nearly as bad as labeling a battery as a non-E-waste,
254 that is FN. Moreover, recall can be trivially maximized by labeling every item an E-waste. Hence
255 recall can never be used as an objective when training ML methods.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

256 Finally, the F1-score, Equation (4) balances precision and recall, and is relevant in applications
257 where both are equally expensive.

$$F_1 = 2 \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

Reduced FN models

258

Appending to the list of models described in the *Models* section, we create a duplicate of every 259 model, but lower the confidence levels from 0.5 to 0.25, favoring detection. By lowering the 260 confidence level, the number of FNs is guaranteed to decrease since it increases the overall positivity 261 rate. Confidence level is the minimum confidence required for the object detector to make the 262 final decision that the target object is in the picture. We name these models YRFNv3, YRFNv4, 263 YRFNv5, YRFNv8 respectively. YRFN stands for YOLO with Reduced False Negatives. 264

EMC: A New Performance Metric for E-waste detection

265

As we have already discussed, each of the usual measures of success used in classification do 266 not lend themselves well to E-waste detection. Accuracy, for example, undermines this situation 267 completely, considering both FNs and FPs the same. Our proposed new measure seeks to balance 268 FNs and FPs by accounting for the costs of each type of misclassification. 269

To determine these costs, we looked at the cost of offsetting the environmental consequences 270 and the costs of inspection of false positives. Looking at a cost benefit analysis literature (Yang 271 et al., 2021) it was highlighted that the cost to offset the environmental consequences was on average 272 4 USD/kg of E-waste, in 2021. Adjusted for inflation, this is 5.95 USD/kg today (2025). Averaging 273 common trash product weights (Empa - Swiss Federal Laboratories for Materials Science and 274 Technology, 2025), we obtain 11.71 USD/item. To find the cost of false positives, we take the 275 average worker's hourly salary, and assuming they spend five minutes on each object, we evaluate 276 the appropriate cost. According to (Wikipedia contributors, 2024), the average recycling worker's 277 hourly salary in China is \$0.17, while it is 32,000/year according to ZipRecruiter in the US. Though a 278

279 wide variation, averaging five objects per minute, give us 0.67 USD/item. These are rough estimates
280 and hence it is important to note that this is only for comparison purposes. Our new method's
281 calibration and performance do not depend on the exact values, only that we would need non-zero
282 values for both.

283 Below is the new metric that we propose, the E-waste misclassification cost (EMC):

$$\text{EMC} = \frac{11.71 FN + 0.67 FP}{TP + FN} \quad (5)$$

284 **Ensembles**

285 An ensemble method is the combining of multiple models to make a prediction - wisdom of the
286 crowd. Different models may give different results on whether they're able to detect an object or
287 not. To make the final decision, a voting system is used such that if more models claim that an
288 object is apparent, then the final prediction is that the object is present and vice-versa.

289 In this study, we created an ensemble of YOLOv3, YOLOv4, YOLOv5, YOLOv8 and another
290 ensemble of RFNv3, YRFNv4, YRFNv5, YRFNv8. If at least three of the models detected E-waste
291 as present, then the final decision would be that E-waste is there. We call them EYOLO and EYRFN.
292 These ensembles estimate labels by polling the 4 YOLOs and the 4 YRFNs, respectively.

293 **WISE: Waste-focused Integrated Smart Ensemble**

294 Consider an ensemble that does not allow equal weighting. That is, the wisdom of the crowd is
295 not equally weighted. The better an underlying method, the more its weight. Further, we will let
296 machine learning pick the weights that minimize the total EMC as defined by Equation (5).

More specifically, let $x_i \in \{0, 1\}^8$ denote the vector of model-level detections (YOLOv3, 297 YOLOv4, YOLOv5s, YOLOv8s, and YRFNv3, YRFNv4, YRFNv5s, YRFNv8s) for image i , and 298 let $y_i \in \{0, 1\}$ be the ground truth label (1 for E-waste, 0 for non-E-waste). WISE learns a weight 299 vector $w \in \mathbb{R}^8$ and bias b by minimizing the expected misclassification cost with an ℓ_1 penalty: 300

$$\min_{w,b} \frac{1}{N} \sum_{i=1}^N \left(c_{\text{FN}} y_i [1 - \hat{y}_i] + c_{\text{FP}} (1 - y_i) \hat{y}_i \right) + \lambda \|w\|_1, \quad (6)$$

where the prediction is made as 301

$$\hat{y}_i = \mathbf{1}[w^\top x_i + b \geq 0]. \quad (7)$$

We set $c_{\text{FN}} = 11.71$ and $c_{\text{FP}} = 0.67$ (as derived from Equation (5), the EMC metric). Optimiza- 302 tion is performed with the Adam optimizer (batch size 256, learning rate 10^{-3}), with early stopping 303 on a stratified validation split. The threshold 0 is without loss of generality since the bias b can be 304 appropriately adjusted. 305

The learned solution is sparse, with only two non-zero weights: YRFNv4 receives 0.6 weight 306 and YRFNv8 receives 0.4 weight. This sparsity results from the ℓ_1 penalty and is consistent with 307 the high performance of these models in Table 2. 308

Note that since we are learning these weights to minimize a loss, we will have to train the 309 weights on a training subset of the data and test it on the rest of the test set. We will use a standard 310 randomization of 80/20 to split the data for testing versus training. We also tune λ over the grid 311 $\{0, 10^{-4}, 10^{-3}, 10^{-2}\}$. For comparisons to other methods above, that did not require training/testing 312 splits, we will report and compare the averages over the entire dataset. This obviously risks that the 313

314 performance improvements could be biased due to the inclusion of in-sample and out-of-sample
315 data. To ensure that this is not the case, we will compare the EMC of the test set to that of the
316 training to ensure that magnitudes are similar.

317 **Dataset**

318 A mixture of multiple datasets was used to test. We did not pick images that were used for
319 original YOLO training. We utilized Kaggle’s E-waste dataset which consisted of mobile phones,
320 microwaves, keyboards and mice. There were 300 images of each giving a total of 1200 images
321 of E-waste. Another 1200 images of non-E-waste were mixed with the E-waste images. These
322 non-E-waste images came from images.cv and roboflow. They all fell under various categories like
323 handbags, chairs, spoons, books, racquets, and umbrellas. Each model was given 2400 images for
324 testing, 1200 consisting of E-waste and 1200 consisting of non-E-waste. To evaluate, the number
325 of true positives, false positives, true negatives, and false negatives were counted.

326 All datasets, preprocessing scripts, and trained models are publicly available at our GitHub
327 repository (GitHub Contributors, 2025).

328 **Results**

329 For each of the methods: classical - YOLOv3, YOLOv4, YOLOv5, YOLOv8; tweaked - YRFNv3,
330 YRFNv4, YRFNv5, YRFNv8; ensemble - EYOLO, EYRFN; Table 2 shows the confusion matrices
331 for each of these methods. As one can observe the traditional YOLOs, though impressive for object
332 detection, do not fare too well for the purposes of E-waste detection. More specifically YOLOv3
333 identifies every object as a non-E-Waste resulting in 1200 TNs and 1200 FNs. This is potentially

due to overfitting during the YOLOv3 training. 334

While YOLOv4, YOLOv5 and YOLOv8 perform better, they still weigh FPs and FNs the same. 335
Very high false negatives (FNs) will essentially cause too many toxic substances to enter landfills 336
and find their ways into living organisms. The YRFNs each have relatively lower FNs. What is 337
also surprising is that the cutting edge YOLO, that is version 8, which is the go-to algorithm today 338
for object detection, is far from the best for E-waste detection. 339

Table 2 also shows the confusion matrix for newly developed smart ensemble - WISE (Waste- 340
focused Integrated Smart Ensemble). It has the lowest FNs amongst all methods compared. The 341
learned weights were 0 for six of the 8 method (4 YOLOs + 4 YRFNs). The only two non-zero 342
weights were YRFNv4: 0.6 and YRFNv8: 0.4 with threshold almost just above zero. 343

A final comparison of all methods in terms of the popular measures of success and our EMC 344
measure are shown in Table 2. As one can see, the smart ensemble out performs every other method 345
significantly. Moreover the magnitude of EMC is in USD per object and can hence easily lend itself 346
to a relatable intuitive value. Table 2, also bolds the best values for each measure of success. 347

Complete confusion matrices and additional performance breakdowns are provided in our 348
GitHub repository (GitHub Contributors, 2025). 349

Statistical Significance Comparisons 350

While raw performance metrics discussed in the previous section provide useful summaries, they do 351
not by themselves establish whether observed differences are statistically significant. To rule out 352
the possibility that they could have been due to chance variations, in this section, we conducted 353
pairwise comparisons using McNemar's test (McNemar, 1947). The McNemar's test is specifically 354

Model	TP	FP	TN	FN	Accuracy	Precision	Recall	EMC
YOLOv3	0	0	1200	1200	0.50	NA	0.00	11.71
YRFNv3	0	0	1200	1200	0.50	NA	0.00	11.71
YOLOv4	756	4	1196	444	0.81	0.99	0.63	4.33
YRFNv4	851	11	1189	349	0.85	0.99	0.71	3.41
YOLOv5	209	5	1195	991	0.59	0.98	0.17	9.67
YRFNv5	373	14	1186	827	0.65	0.96	0.31	8.08
YOLOv8	176	6	1194	1024	0.57	0.97	0.15	10.00
YRFNv8	340	22	1178	860	0.63	0.94	0.28	8.40
EYOLO	90	0	1200	1110	0.54	1.00	0.08	10.83
EYRFN	203	0	1200	997	0.58	1.00	0.17	9.73
WISE - Smart(learned)	906	25	1175	294	0.87	0.97	0.76	2.88

Table 2: Performance comparison of different models.

355 designed to evaluate differences between two classifiers tested on the same set of data points.

356 Given two models A and B , McNemar’s test constructs a 2×2 contingency table based on
357 whether each classifier prediction is correct or incorrect. The key quantities are n_{10} (instances where
358 A is correct but B is wrong) and n_{01} (instances where A is wrong but B is correct). Under the null
359 hypothesis that both classifiers have equal error rates, these counts should be approximately equal.

360 The test statistic is defined as

$$\chi^2 = \frac{(|n_{01} - n_{10}| - 1)^2}{n_{01} + n_{10}},$$

361 which asymptotically follows a χ^2 distribution with one degree of freedom. A small p -value indicates
362 that the classifiers differ significantly in their performance.

363 Figure 6 presents a heatmap of the pairwise McNemar p -values. It is displayed on a $-\log_{10}(p)$
364 scale for easier readability - large values are significant and colored yellow. As shown, almost

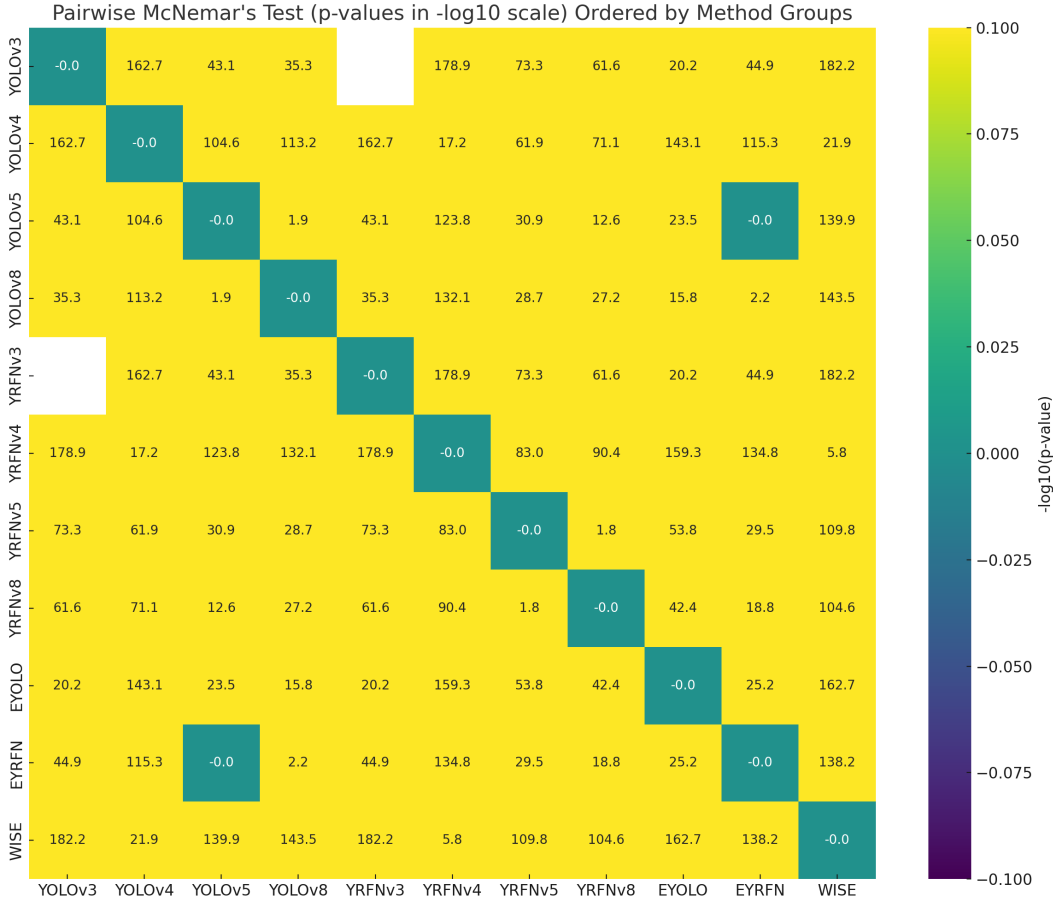


Figure 6: Pairwise McNemar’s test results across all methods. The heatmap displays $-\log_{10}(p)$ values, where larger magnitudes correspond to stronger evidence against the null hypothesis of equal performance.

all off-diagonal entries exhibit extremely small p -values, well below conventional significance 365 thresholds (e.g., $p < 0.05$). This establishes that the observed performance gaps across methods are 366 not attributable to chance. 367

Methods that have identical predictions (**YOLOv3 vs. YRFNv3**) appear visually as blank/white 368 in the heat map since in such cases the McNemar’s statistic is undefined (no discordant pairs). We 369 also notice that EYRFN is not statistically different from YOLOv5. 370

These findings are reassuring that the comparative rankings from Table 2 are not due to chance. 371 Without such an analysis, improvements cannot be trusted to have come from well-grounded 372

373 performance improvements.

374 **Limitations and future work**

375 Our 2,400-image corpus mixes curated E-waste and non-E-waste classes but is modest in size and
376 lacks nuances of in-the-wild images. There might be cluttering, stacking, off-the-frame issues etc.
377 on sorting lines and conveyor belts. In-situation tests and refinements are a subject of future studies.

378 Assumptions on the cost accounting for EMC uses approximate unit costs from literature and
379 wage proxies. We do adjust the costs for inflation. However, we have one EMC measure that is
380 calibrated to costs in the United States, which produces the largest quantity of E-waste in the world.
381 However, we would need to refine this measure of different economies in the future.

382 Finally, we have not studied any of the real-time constraints like latency and throughput. Fur-
383 ther improvements can help create ensembles that also consider the underlying models' latency
384 and throughput to combine them to yield a smart ensemble that matches the expected real-time
385 performance thresholds will be the subject of future studies.

386 **Conclusion**

387 The lower the EMC, the better the model is at detecting E-waste and reducing overall cost. Though
388 the measure, EMC, weighs the estimated cost of FNs against that of FPs, the goal is not primarily
389 cost reduction. The goal is the reduction of E-waste that goes into landfills, for any given budget.
390 This is critical especially since the ever growing volume of E-waste far exceeds the rate at which
391 budgets to tackle E-waste do not grow.

392 Many surprising observations were made in the results section. Classical cutting-edge methods

like YOLOv8 (Jain et al., 2025), have terrible performance for E-waste detection, although their 393 performance for general object detection is commendable. Further thought would reveal that this is 394 not totally surprising, since they were trained and evaluated on a generic object detection context 395 and with accuracy as the primary measure. Hence, it might not be fair to compare, or to use such 396 methods for E-waste detection. Next, tweaked methods outperform the classical methods whose 397 default parameters would have been carefully chosen by the authors. This again shows that even 398 parameters calibrated for generic object detection do not work well for E-waste detection. We also 399 confirmed that almost all model pairs differ in a statistically robust way and more importantly our 400 smart ensemble, WISE, is statistically better than all the other models. 401

Our results align with reports that off-the-shelf detectors can degrade under domain shift 402 (Solovyev et al., 2021). Our results also align with broader ML findings that cost-sensitive training 403 and ensembling can improve decision quality when costs are asymmetric (Domingos, 1999; Elkan, 404 2001; Wolpert, 1992). In E-waste operations, prioritizing lower FNs as opposed to FPs by an 405 EMC-minimizing ensembles like WISE is a principled path forward. Our smart ensemble does 406 surprisingly and fortunately well in the context of E-waste detection. It not only has a significant cost 407 reduction, it also has the lowest number of FNs in the entire comparison set. While a methodology 408 that reduces FNs to zero while not labeling every object as FP is ideal, that would be for future 409 research. 410

411 **References**

- 412 Adedeji, O., & Wang, Z. (2019). Intelligent waste classification system using deep learning convo-
413 lutional neural network. *Procedia Manufacturing*. [https://doi.org/10.1016/j.promfg.2019.05.](https://doi.org/10.1016/j.promfg.2019.05.086)
414 086
- 415 Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). Yolov4: Optimal speed and accuracy of
416 object detection. <https://arxiv.org/abs/2004.10934>
- 417 Bodla, N., Singh, B., Chellappa, R., & Davis, L. S. (2017). Soft-nms – improving object detection
418 with one line of code. *Proceedings of the IEEE International Conference on Computer Vision*
419 *(ICCV)*, 5561–5569. [https://openaccess.thecvf.com/content_ICCV_2017/papers/Bodla_Soft-](https://openaccess.thecvf.com/content_ICCV_2017/papers/Bodla_Soft-NMS_-_Improving_ICCV_2017_paper.pdf)
420 [NMS_-_Improving_ICCV_2017_paper.pdf](https://openaccess.thecvf.com/content_ICCV_2017/papers/Bodla_Soft-NMS_-_Improving_ICCV_2017_paper.pdf)
- 421 *Bounding box regression loss* [Author not listed; wiki page]. (n.d.). CloudFactory Computer Vision
422 Wiki. Retrieved September 6, 2025, from [https://wiki.cloudfactory.com/docs/mp-wiki/loss/](https://wiki.cloudfactory.com/docs/mp-wiki/loss/bounding-box-regression-loss)
423 [bounding-box-regression-loss](https://wiki.cloudfactory.com/docs/mp-wiki/loss/bounding-box-regression-loss)
- 424 Cheng, X., Cheng, Y., Gong, Z., & Chen, Y. (2021). An improved small object detection method
425 based on yolov3. *Multimedia Tools and Applications*, 80(19), 29329–29344. [https://doi.org/](https://doi.org/10.1007/s11042-021-10605-5)
426 [10.1007/s11042-021-10605-5](https://doi.org/10.1007/s11042-021-10605-5)
- 427 Domingos, P. (1999). Metacost: A general method for making classifiers cost-sensitive. *Proceedings*
428 *of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data*
429 *Mining (KDD)*, 155–164. <https://doi.org/10.1145/312129.312220>
- 430 Elkan, C. (2001). The foundations of cost-sensitive learning. *Proceedings of the 17th International*
431 *Joint Conference on Artificial Intelligence (IJCAI)*, 973–978. [https://cseweb.ucsd.edu/](https://cseweb.ucsd.edu/~elkan/rescale.pdf)
432 [~elkan/rescale.pdf](https://cseweb.ucsd.edu/~elkan/rescale.pdf)

Empa - Swiss Federal Laboratories for Materials Science and Technology. (2025). Ewaste guide 433
 [Accessed: 2025-08-24]. 434

Fan, J., Cui, L., & Fei, S. (2023). Waste detection system based on data augmentation and yolo_ec. 435
Sensors, 23(7), 3646. <https://doi.org/10.3390/s23073646> 436

Feng, M., Jiang, T., Wang, J., Yang, L., & Zhang, Y. (2023). A multi-scale traffic object detection 437
 algorithm for road scenes based on improved yolov5. *Sensors*, 23(4), 2045. [https://doi.org/](https://doi.org/10.3390/s23042045) 438
 10.3390/s23042045 439

Fornalczyk, A., Willner, J., Francuz, K., & Cebulski, J. (2013). E-waste as a source of valuable 440
 metals. *Archives of Materials Science and Engineering*, 63(2), 87–92. [http://www.amse.](http://www.amse.acmsse.h2.pl/vol63_2/6325.pdf) 441
[acmsse.h2.pl/vol63_2/6325.pdf](http://www.amse.acmsse.h2.pl/vol63_2/6325.pdf) 442

GitHub Contributors. (2025). Wise – <https://github.com/vishnumuthuraman/wise.git> [Accessed: 443
 2025-08-24]. 444

Huang, J., Rathod, V., Sun, C., Zhu, M., et al. (2017). Speed/accuracy trade-offs for modern 445
 convolutional object detectors. *CVPR*. 446

Ibrahim, A. N., Youssif, E. S., Hassanien, A. E., Hossam, A. A., Abdo, S. S., & Saleh, A. M. D. 447
 (2023). Object detection-based automatic waste segregation using robotic arm. *International* 448
Journal of Advanced Computer Science and Applications (IJACSA). [https://doi.org/10.](https://doi.org/10.14569/IJACSA.2023.0141164) 449
 14569/IJACSA.2023.0141164 450

Jain, H., Jain, J., Shrivastava, V., Pandey, A., & Sharma, S. (2025). Yolov8 and its advancements. 451
International Journal of Research Publication and Reviews, 6(5), 660–666. Retrieved 452
 September 6, 2025, from <https://ijrpr.com/uploads/V6ISSUE5/IJRPR44604.pdf> 453

Jocher, G., & Ultralytics. (2020). Yolov5. <https://github.com/ultralytics/yolov5> 454

Jocher, G., & Ultralytics. (2023). Yolov8. <https://github.com/ultralytics/ultralytics> 455

456 Kán, P., & Kaufmann, H. (2019). Deeplight: Light source estimation for augmented reality using deep
457 learning. *The Visual Computer*, 35, 873–883. <https://doi.org/10.1007/s00371-019-01666-x>

458 Khan, H. A., Naqvi, S. S., Alharbi, A. A. K., Alotaibi, S., & Alkhatami, M. (2024). Enhancing trash
459 classification in smart cities using federated deep learning. *Scientific Reports*, 14, 11816.
460 <https://doi.org/10.1038/s41598-024-62003-4>

461 Li, K., & Cao, L. (2020). A review of object detection techniques. *2020 5th International Conference
462 on Electromechanical Control Technology and Transportation (ICECTT)*, 385–390. <https://doi.org/10.1109/ICECTT50890.2020.00091>

463

464 Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection.
465 *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2999–3007.
466 <https://doi.org/10.1109/ICCV.2017.324>

467 Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2016). Ssd:
468 Single shot multibox detector. *ECCV*.

469 Majchrowska, S., Mikołajczyk, A., Ferlin, M., Klawikowska, Z., Plantykowski, M. A., Kwasigroch, A.,
470 & Majek, K. (2022). Deep learning-based waste detection in natural and urban environments.
471 *Waste Management*, 138, 274–284. <https://doi.org/10.1016/j.wasman.2021.12.001>

472 McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions
473 or percentages. *Psychometrika*, 12(2), 153–157.

474 Meituan Vision AI Department. (2022). Yolov6: A single-stage object detection framework for
475 industrial applications. <https://github.com/meituan/YOLOv6>

476 Noel-Brune, M., Goldizen, F. C., Neira, M., van den Berg, M., Lewis, N., King, M., Suk, W. A.,
477 Carpenter, D. O., Arnold, R. G., & Sly, P. D. (2013). Health effects of exposure to e-waste.
478 *The Lancet Global Health*, 1(2), e70. [https://doi.org/10.1016/S2214-109X\(13\)70020-2](https://doi.org/10.1016/S2214-109X(13)70020-2)

- Oza, P., Agrawal, S., Kapadia, M., & Raotole, O. (2025). Solid waste classification using deep 479
neural network: A transfer learning approach. *Earth Science Informatics*, 18, 255. <https://doi.org/10.1007/s12145-025-01743-x> 480
481
- Parti, A. (2024). Understanding object detection: A comprehensive guide [Retrieved March 18, 482
2024, from <https://pareto.ai/blog/object-detection>]. 483
- Paudel, P., Shrestha, S., Shrestha, S., Gurung, S., & Adhikari, S. (2024). Automated waste sorting 484
with delta arm and yolov8 detection. *Journal of Artificial Intelligence and Capsule Networks*, 485
6(3), 299–315. <https://doi.org/10.36548/jaicn.2024.3.004> 486
- Potter, R. (2022, August 23). *How bounding box object detection is taking over and what to do about* 487
it [Analytics on Medium]. Retrieved September 6, 2025, from [https://medium.com/analytics/](https://medium.com/analytics/how-bounding-box-object-detection-is-taking-over-and-what-to-do-about-it-9130b6d6c1f1) 488
[how-bounding-box-object-detection-is-taking-over-and-what-to-do-about-it-9130b6d6c1f1](https://medium.com/analytics/how-bounding-box-object-detection-is-taking-over-and-what-to-do-about-it-9130b6d6c1f1) 489
- Proença, P. F., & Simões, P. (2020). TACO: Trash annotations in context for litter detection. *arXiv*. 490
<https://arxiv.org/abs/2003.06975> 491
- Proença, P. F., & Simões, P. (2023). TACO dataset website [Accessed 2025-09-05]. 492
- Promboonruang, S., Boonrod, T., & Ratchatawetchakul, Y. (2024). Efficient waste detection and 493
classification based on yolov5 models. *Engineering Access (MIJET)*, 10(1), 51–58. <https://doi.org/10.14456/mijet.2024.7> 494
495
- Rahim, L. A. A., Abidin, N. A. Z., Aminuddin, R., Samah, K. A. F. A., Ibrahim, A. Z. M., Yusoh, 496
S. D., & Nasir, S. D. N. M. (2024). YOLOv5 model-based real-time recyclable waste 497
detection and classification system. In *Innovations in smart cities applications volume 7 (sca* 498
2023) (pp. 44–54, Vol. 906). Springer, Cham. https://doi.org/10.1007/978-3-031-53824-7_5 499

500 Rajeev, P. A., Dharewa, V., Lakshmi, D., Vishnuvarthanan, G., Giri, J., Sathish, T., & Alrashoud, M.
501 (2025). Advancing e-waste classification with customizable YOLO-based deep learning
502 models. *Scientific Reports*, *15*, 18151. <https://doi.org/10.1038/s41598-025-94772-x>

503 Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time
504 object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern
505 Recognition (CVPR)*, 779–788. <https://doi.org/10.1109/CVPR.2016.91>

506 Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. [https://arxiv.org/abs/
507 1804.02767](https://arxiv.org/abs/1804.02767)

508 Sheng, V. S., & Ling, C. X. (2006). Thresholding for making classifiers cost-sensitive. *Proceedings
509 of the AAAI Conference on Artificial Intelligence*, 476–481. [https://cdn.aaai.org/AAAI/
510 2006/AAAI06-076.pdf](https://cdn.aaai.org/AAAI/2006/AAAI06-076.pdf)

511 Singh, Y., & Parimala S. (2025, March 26). E-waste management. [https://doi.org/10.2139/ssrn.
512 5196804](https://doi.org/10.2139/ssrn.5196804)

513 Solawetz, J. (2024, January 4). *What is yolov4? a detailed breakdown*. Roboflow. Retrieved Septem-
514 ber 6, 2025, from <https://blog.roboflow.com/a-thorough-breakdown-of-yolov4/>

515 Solovyev, R., Wang, W., & Gabruseva, T. (2021). Weighted boxes fusion: Ensembling boxes
516 from different object detection models. *Image and Vision Computing*, *107*, 104117. [https:
517 //doi.org/10.1016/j.imavis.2021.104117](https://doi.org/10.1016/j.imavis.2021.104117)

518 Song, L., Yu, H., Yang, Y., Tong, Y., Ren, S., & Ye, C. (2025). Yolo-vg: An efficient real-time
519 recyclable waste detection network. *Journal of Real-Time Image Processing*. [https://doi.org/
520 10.1007/s11554-025-01655-5](https://doi.org/10.1007/s11554-025-01655-5)

521 Thung, G., & Yang, M. (2017). Trashnet [Accessed 2025-09-05].

- Vats, M. C., & Singh, S. K. (2014). E-waste characteristic and its disposal. *International Journal of Ecological Science and Environmental Engineering*, 1(2), 49–61. 522
523
- Wang, C.-Y., Bochkovskiy, A., & Liao, H.-Y. M. (2022). Yolov7: Trainable bag-of-freebies sets 524
new state-of-the-art for real-time object detectors. <https://arxiv.org/abs/2207.02696> 525
- Wikipedia contributors. (2024). *Electronic waste in Guiyu* [Wikipedia article]. Retrieved August 526
24, 2025, from https://en.wikipedia.org/wiki/Electronic_waste_in_Guiyu 527
- Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, 5(2), 241–259. 528
- Wu, P., Chai, B., Nie, X., Yan, L., Wang, Z., Zhou, Q., Wang, B., Peng, Y., & Li, H. (2024). 529
Enhanced object detection: A study on vast vocabulary object detection track for v3det 530
challenge 2024 (version 3). <https://doi.org/10.48550/arXiv.2406.09201> 531
- Wu, T., & Zhu, S.-C. (2013). Learning near-optimal cost-sensitive decision policy for object 532
detection. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 533
753–760. <https://doi.org/10.1109/ICCV.2013.98> 534
- Yang, W.-D., Sun, Q., & H-G., N. (2021). Cost-benefit analysis of metal recovery from e-waste: 535
Implications for international policy. *Waste Management*, 123, 42–47. [https://doi.org/10. 536
1016/j.wasman.2021.01.023](https://doi.org/10.1016/j.wasman.2021.01.023) 537

WISE: An Adaptive YOLO Ensemble for Accurate E-Waste Object Detection

September 30, 2025

Abstract

1

Electronic waste (E-waste) - which contains hazardous chemicals (e.g., lead, mercury) and valuable precious metals - has increased tremendously over the past few decades. E-waste reached 62 million metric tons in 2022 and continues to accelerate. Manual sorting is unable to keep up. Only 22.3% was documented as collected and recycled, leaving the rest to enter landfills. There is an urgency for automated techniques to help. Object detection algorithms, traditionally, focus on improving accuracy but end up weighing false negatives and false positives the same. However here, False Negatives (FNs) - like batteries being labeled as non-E-waste are far more hazardous.

2

3

4

5

6

7

8

This paper has three contributions. First, we develop a new metric calibrated to help pick the best algorithm for E-waste detection. This metric provides a basis upon which even future algorithms can be evaluated. Second, we revisit existing algorithm evaluations for the specific purpose of E-waste detection. We evaluate several standard, tweaked and ensemble methods. Finally, we propose a smart ensemble (**WISE: Waste-focused Integrated Smart Ensemble**) whose weights are learned using machine learning, to minimize the costs/impact of E-waste disposal. The grand goal of these is to help improve public health and reusability of precious metals by enabling more efficient recovery and processing of E-waste.

9

10

11

12

13

14

15

16 *Keywords:* Object detection, You Only Look Once (YOLO), Electronic Waste (E-waste), Artificial
17 Intelligence (AI), Ensemble Methods, Threshold calibration

18 **Introduction**

19 Electronic waste, or E-waste, consists of numerous hazardous materials including plastics, lead,
20 mercury, cadmium, arsenic, etc. (Vats & Singh, 2014). E-waste that ends up in landfills is responsible
21 for multiple health hazards, especially in developing countries. Examples of these hazards include
22 fetal loss, prematurity, low birth weight, abnormal thyroid function, neurobehavioral disturbances,
23 and genotoxicity (Noel-Brune et al., 2013). Additionally, the E-waste that ends up in landfills
24 consists of valuable metals – aluminum (Al), gold (Ag), palladium (Pd), platinum (Pt) (Vats &
25 Singh, 2014) – that can be extracted and reused, saving money. On just a single TV board, 7% of
26 its salvage value comes from silver, 33% from gold, and 7% from palladium (Fornalczyk et al.,
27 2013). In a mobile phone, 11% comes from silver, 71% comes from gold, and 11% comes from
28 palladium, totaling 93% of the phone's salvage value comes from just precious metals (Fornalczyk
29 et al., 2013).

30 In 2022, we generated 62 million metric tons of E-waste and this is expected to rapidly escalate
31 to 74 million metric tons by 2030 (Singh & Parimala S, 2025). Preventing hazardous material from
32 getting into landfills is critical. It is almost the only way to stop these chemicals eventually finding
33 their way into animal and human bodies. Facilities and techniques for extracting the valuable metals
34 from E-waste and disposing of the rest appropriately exist. However, the challenge lies in identifying
35 the E-wastes and keeping up with volume. We currently depend on manual sorting, afforded in part
36 by cheap labor from developing and underdeveloped countries to help prevent hazardous material

from reaching landfills.

37 |

The ethical issues involved in dispatching the E-waste from developed nations to socioeco- 38
nically disadvantaged nations cannot be overstated. In developed nations, medical ailments 39
caused by E-waste are often not seen every day. As a result, the problem remains hidden and even 40
conscientious citizens rarely notice it. In contrast, underprivileged localities are often the least 41
equipped environments to cope with these imported toxic substances. They also lack the medical 42
resources necessary to tackle the rising health crisis. Many governments have recognized this issue 43
and have attempted to use systems and regulations that cut down on E-waste. Even with such efforts 44
and often challenged funding, growth in E-waste seems to, so far, outpace our efforts to prevent it 45
from entering our landfills and ecosystems. 46

Automated techniques, though not perfect, can be of huge benefit in augmenting manual sorting. 47
Current research has already developed artificial intelligence (AI) models that have been trained 48
and evaluated for generic object-detection. Pictures of labeled objects are processed and used to 49
train neural network models. Training involves the incremental adjustment of the neural network's 50
weights to match the known labels. Once trained, the final network with its adjusted weights is 51
used to label unlabeled objects, that is, detect objects. These neural net models also create bounding 52
boxes for each object and estimate labels for each object. Bounding boxes are tight rectangles 53
around detected objects. This can even facilitate the extraction of E-waste more easily rather than 54
debating whether an object should be classified under this name. 55

The problem with existing algorithms and their benchmarks is that they focus primarily on 56
accuracy or at most precision and recall, as their measures of success. Accuracy is well warranted 57
for generic object detection, where both false positives (FP) and false negatives (FN) are equally bad. 58
Other variants like precision and recall are well warranted when only FPs (eg. screening resumes) 59

60 or FNs (eg. screening for cancer) are to be minimized, respectively. From an E-waste perspective,
61 it is less harmful to label a banana peel as E-waste, not allowing it to go to the landfill directly
62 (routing it to manual sorting). It is much more harmful to label a battery, full of toxic chemicals,
63 as non-E-waste, allowing it to directly enter the landfill and the ecosystem. The clear objective
64 would be to allow FPs like identifying the banana peel as E-waste in lieu of allowing FNs like
65 batteries to reach the landfill. Choosing a standard measure like recall will not help either, since
66 it can be maximized by labeling everything as positive, thereby making FNs zero! But that will
67 make automation useless since it will keep over-burdening manual sorting - the exact reason we
68 seek automation/augmentation.

69 Specifically in terms of literature on AI for waste detection, deep learning models have been
70 used for solid waste classification (Adedeji & Wang, 2019; Majchrowska et al., 2022; Oza et al.,
71 2025). The *TrashNet* (Thung & Yang, 2017) data set was created as a part of a student-led project in
72 2017 and has been one of the popular datasets for trash classification benchmarking. However, this
73 data set is not relevant for E-waste classification since the only six classes included are glass, paper,
74 cardboard, plastic, metal and trash. It contains 2527 images and has been used for benchmarking in
75 some research papers (Khan et al., 2024; Rahim et al., 2024). Another popular direction of research
76 has been in identifying *waste in the wild*. The TACO data set (Proença & Simões, 2020, 2023)
77 provides 1500 images of trash in various locations, from the beaches to city streets. These are to
78 be used to train algorithms that can find trash in pictures taken from, say, an automated garbage
79 collector (Fan et al., 2023; Promboonruang et al., 2024; Song et al., 2025). Findings from these two
80 trash-related applications demonstrate that off-the-shelf general-purpose detectors trained on natural
81 image corpora, can perform poorly. These are however great proof-of-concept trash classification
82 systems being used in related applications. Motivated by this, we frame the E-waste identification

problem as a cost-sensitive identification problem and focus on fine-tuning and developing variants that explicitly minimize our costs.

YOLO (You Only Look Once) family of detectors, though created in 2016 (Redmon et al., 2016), only became popular and became the go-to choice recently. These one-stage detectors have been shown to be very effective specifically for integration into real-time waste sorting, for example with robotic arms (Ibrahim et al., 2023; Paudel et al., 2024). However, in terms of effective sorting methods for identifying E-waste, there exists only one recent paper (Rajeev et al., 2025) that simply compares different YOLO methods in terms of accuracy and the reported poor performances underscore the need for improved methods.

Prior waste detection or E-waste detection studies predominantly optimize for accuracy, precision or recall rather than application level costs of mistakes. None, to our knowledge, consider the application level E-waste misclassification costs that differentiate FP and FN. However, in classical Machine Learning, both (a) minimizing expected misclassifications cost (EMC), when costs of FP and FN differ (Domingos, 1999; Elkan, 2001; Sheng & Ling, 2006) and (b) creating better performing ensemble are popular research topics (Bodla et al., 2017; Lin et al., 2017; Solovyev et al., 2021; Wu & Zhu, 2013).

Contribution and outline

In this paper we,

(1) Create a new metric, E-waste Misclassifications Cost (EMC), calibrated to help pick the best algorithm for E-waste detection, providing a basis upon which even future algorithms can be evaluated. EMC weights the FNs substantially greater than the FPs but does not completely nullify

104 the FPs. We compute the coefficients based on accounting cost estimates. We seek a measure that
105 gives us an estimate of the total cost of misidentification, lending itself well to our case since the
106 additional cost of manually sorting a FP should be contrasted against the estimated environmental
107 cost of sending a FN to landfill.

108 (2) Revisit existing algorithm evaluations for the specific purpose of E-waste detection. We
109 evaluate several standard, tweaked and ensemble methods. We analyze several algorithms. The
110 first four algorithms are standard and popular implementations (YOLOv3, YOLOv4, YOLOv5,
111 YOLOv8). The second four are our tweaks to the first four to reduce FNs (YRFNv3, YRFNv4,
112 YRFNv5, YRFNv8). We create two more ensembles that we call EYOLO and EYRFN. These
113 estimate labels by polling the 8 previous variants and the 4 reduced FN variants.

114 (3) Finally, we propose a smart ensemble (**WISE: Waste-focused Integrated Smart Ensemble**)
115 whose weights are learned using machine learning, to minimize the costs/impact of E-waste disposal.
116 Ensemble models combine multiple models to make a final recommendation or decision. Since this
117 is a machine learning algorithm, we do in-sample comparisons with other methods and also make
118 in-sample versus out-of-sample comparison to ensure that there is no overfitting.

119 **Object Detection**

120 Object detection is the process of analyzing an image, localizing objects, and classifying what
121 those objects are. It is utilized in numerous applications such as face, pedestrian (Fig. 1) and
122 object detections (Li & Cao, 2020). Convolutional Neural Networks (CNNs), are a specific type of
123 neural networks that play a very significant role in the object detection space. They are essentially
124 classical feed forward neural networks with additional layers called convolutional layers. These

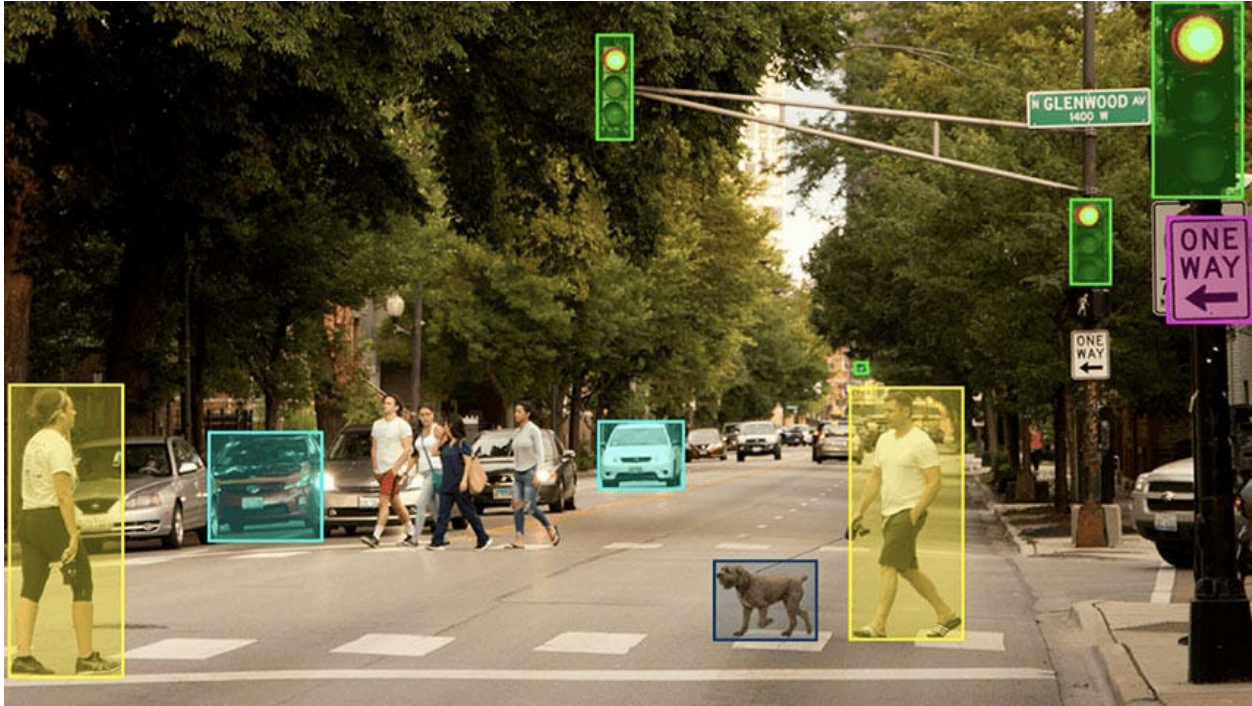


Figure 1: Object Detection example with colored bounding boxes: yellow = person, blue = vehicle, green = traffic light. Boxes illustrate localization outputs prior to non-maximum suppression.¹

convolutional layers help in any kind of computer vision by applying trained averaging kernels. 125
This kernel performs computations on the image's pixel values, or representations of the color and 126
intensity. By performing these computations, the CNN is able to create a feature map which consists 127
of channels, or representations of the different learned features – lines, edges, curves. 128

When an image is processed to look for certain objects, it is first transformed into a form that can 129
be passed on as an input to a neural network (NN). This means the image must be processed, re-sized 130
and normalized for the neural network to take in the image. All inputs are also normalized to $[0, 1]$ 131
by dividing by 255. A trained CNN can then be used to identify objects using intermediate steps 132
that rely on extracted visual features – edges, corners, textures (Parti, 2024). A Region Proposal 133
Network (RPN), a smaller CNN that utilizes extracted visual features, can be used to create bounding 134
boxes around potential object locations. Objects encapsulated in these bounding boxes are then 135

¹Extracted from Potter, 2022.

136 identified and localization is used to calibrate bounding boxes to exclude unnecessary parts of
137 the image. Non-maximum suppression (NMS), a mathematical algorithm, keeps only the most
138 confident detections, excluding false positives and improves accuracy. The bounding boxes are
139 further refined and object identification is finalized (Parti, 2024).

140 There are two main categories of CNN object detectors: two-stage and one-stage (Wu et al.,
141 2024). In the two-stage detector (Figure 2), the algorithm first makes object proposals — guessing
142 where the object appears in the image. The second stage classifies the object and refines the bounding
143 box. By contrast, one-stage detectors completely skip this object proposal step. They go straight
144 to predicting the bounding boxes and classifying objects. Overall, two-stage detectors are more
145 accurate but computationally much more expensive and slower in comparison to one-stage detectors
146 (Wu et al., 2024). One-stage detectors achieve favorable speed/accuracy trade-offs, often 2–5 times
147 faster, making them practical for real-time applications. Prior work quantifies these trade-offs across
148 architectures and scales (Huang et al., 2017; Liu et al., 2016).

149 You Only Look Once (YOLO) is the most popular one-stage detector that uses a CNN archi-
150 tecture. It uses the CNN to predict where the bounding boxes for objects should be placed. It then
151 assigns the probability of each object in a bounding box being an object of specific class, which are
152 called class-probabilities. For example, the probability the image of an animal could be classified
153 as a dog with probability 0.1 and a cat with probability 0.9. YOLO is known for its computational
154 speed and high accuracy. There are numerous versions of YOLO which all have distinct advantages
155 and disadvantages. YOLO will be the primary focus of this paper due to the availability, open-source
156 license and the popularity of this technique.

²Extracted from Solawetz, 2024

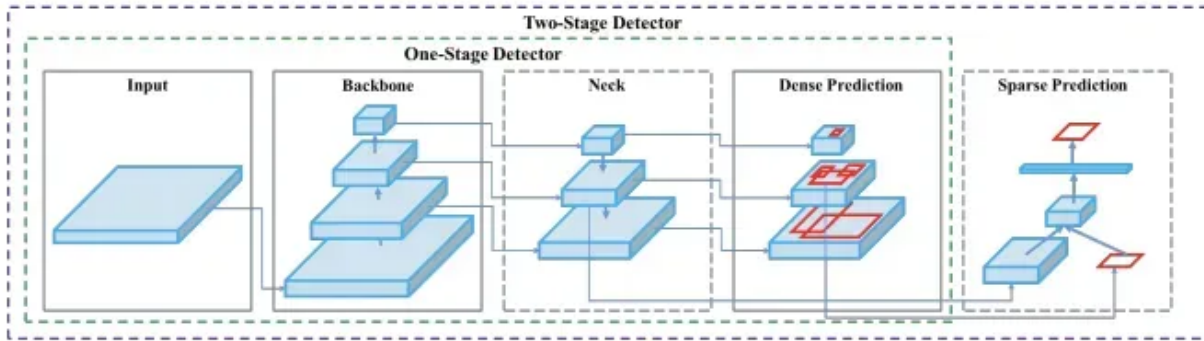


Figure 2: One Stage Vs Two Stage detection: two-stage proposes regions then classifies/refines; one-stage predicts boxes/classes in one pass ²

Models

157

There are eight versions of the YOLO starting with version 1 proposed by (Redmon et al., 2016). 158

Versions 2 and 3 are improved versions with residual blocks (Redmon & Farhadi, 2018). YOLO 159

version 4 was released by A. Bochkovskiy, who took over after Redmon retired (Bochkovskiy et al., 160

2020). It included major performance improvements and new training tricks. 161

Versions 5 (Jocher & Ultralytics, 2020) and Version 8 (Jocher & Ultralytics, 2023) were 162

developed by Ultralytics, and do not have an official research paper. Version 5 re-implemented 163

YOLO in PyTorch while Version 8, improved 5, by unifying the code base and modernized the 164

architecture. YOLO version 6 was developed by Meituan (Meituan Vision AI Department, 2022), 165

was optimized for industrial deployment and does not have a formal research paper either. YOLO 166

version 7 (Wang et al., 2022) introduced several new features like extendable trainable bag-of- 167

freebies and architectural refinements. The most popular among the YOLOs are Version 3 (last 168

official Redmon release), Version 4 (huge leap in accuracy), Version 5 (industry standard, ease of 169

use) and Version 8 (latest, cutting edge, unified framework). 170

171 **YOLOv3**

172 The backbone structure, the part of the neural network that extracts features, utilized by YOLOv3s
173 is Darknet-53 (Cheng et al., 2021; Redmon & Farhadi, 2018). Darknet53 is a CNN with 53 layers
174 which uses residual connections (Figure 3), allowing for input to skip over layers to improve effi-
175 ciency. Convolution blocks, layers of the CNN that extract features, use a sequence of Convolution,
176 Batch Normalization, and Leaky ReLU, or otherwise called CBL. Convolution uses kernels to detect
177 features, Batch Normalization allows CNNs to stabilize and train faster, and Leaky ReLU – a mathe-
178 matical function – helps the CNN learn complex patterns. All YOLO methods include normalization
179 of each channel to $[0, 1]$ and do not have per-channel mean/standard deviation standardizations.

180 The head structure, or the end of the model, of YOLOv3 predicts where objects are at three
181 scales: small, medium, and large. In other words, this helps the model detect small objects, medium
182 objects, and large objects. Anchor boxes, or boxes that have been predefined at certain sizes and
183 shapes, are placed in every grid cell to facilitate bounding box predictions. Finally, the bounding
184 boxes are placed around the target object(s).

185 **YOLOv4**

186 YOLOv4's backbone structure is CSPDarknet53 (Bochkovskiy et al., 2020). Darknet53 extracts
187 the features while Cross Stage Partial (CSP) splits channels from the feature map such that half the
188 channels are able to use the residual connections and the other half continue through the several
189 layers. This ensures that there's a balance between performance and efficiency. In the neck structure
190 – the part of the object detection model that combines details that have been collected, allowing
191 for a better understanding of the image – of YOLOv4, Spatial Pyramid Pooling (SPP) – a pooling

³Extracted from Kán and Kaufmann, 2019

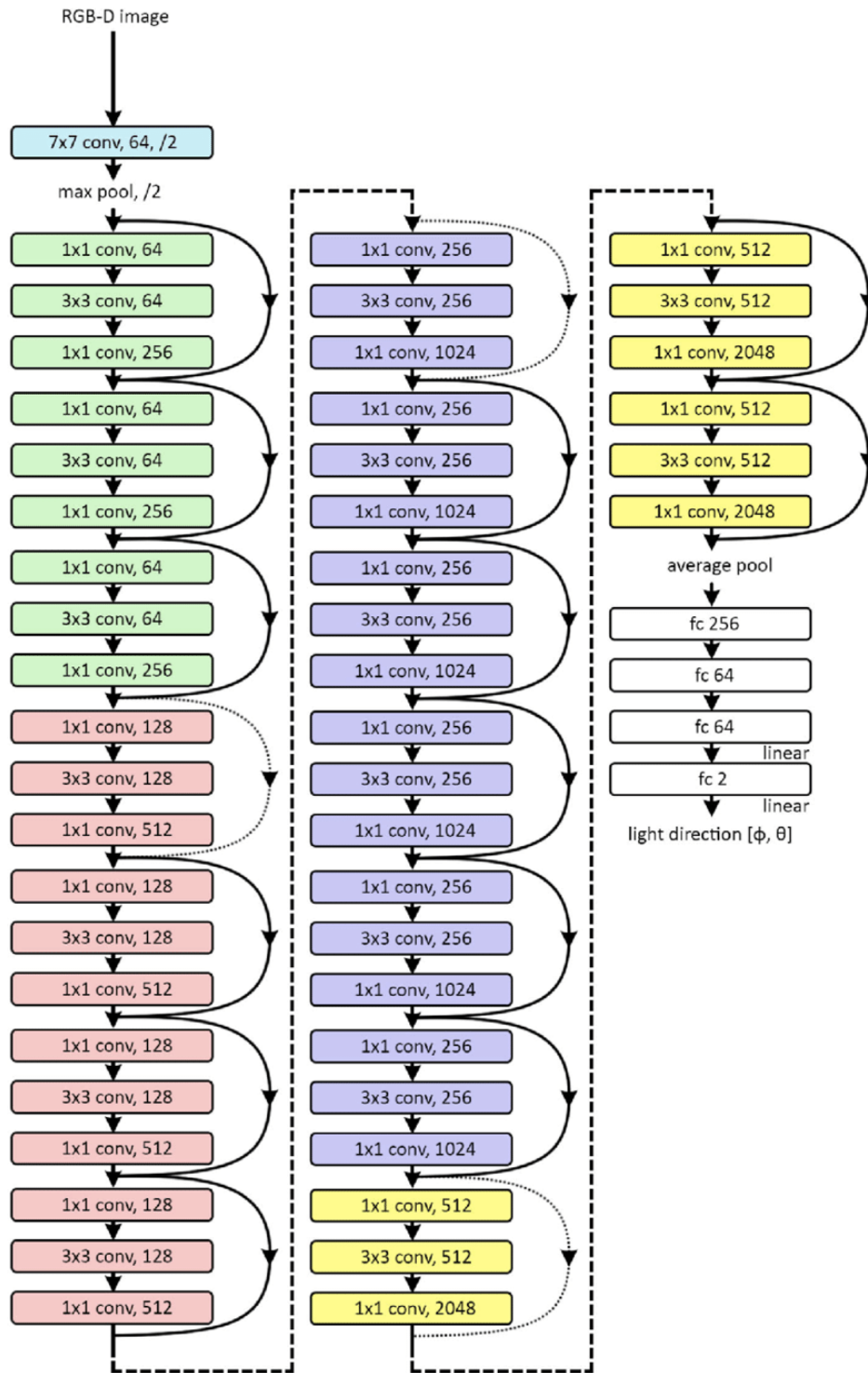


Figure 3: Darknet53 architecture: Deep convolutional network applied to RGB-D input, using residual connections across multiple layers to extract hierarchical features for downstream object detection tasks.³

192 module, reducing the size of the feature map and keeping on essential parts, in the CNN – takes the
193 feature maps that were created from CSPDarknet53, and underscores important patterns. It also
194 looks at the feature maps in different sizes, helping the SPP pick up minute details as well as the
195 larger context. Path Aggregation Network (PAN), a component of the CNN, focuses on locating
196 where an object occurs in the image.

197 The head structure for YOLOv4 is the same head structure used for YOLOv3 due to its efficiency.
198 Although the head structures are the same, YOLOv4 has an improved backbone and neck structure
199 in comparison to YOLOv3.

200 **YOLOv5s**

201 The backbone structure, the part of the neural network that extracts features, utilized by YOLOv5s is
202 inspired by CSPDarknet53 (Feng et al., 2023; Jocher & Ultralytics, 2020). Additional convolution
203 layers are used to extract the image features to a greater extent.

204 To improve its performance, YOLOv5s records its mistakes which can fall into one of three
205 categories: bounding box regression loss, where the bounding box is placed away from the correct
206 spot from where the object is located, as seen in Figure 4; target confidence loss, where YOLOv5s
207 claims an object is apparent when in reality, it's not; and classification loss, when an object is
208 mislabeled.

209 **YOLOv8s**

210 YOLOv8s uses a backbone inspired by CSP. When the channels of the feature map are created,
211 instead of CSP splitting the channels up, YOLOv8s uses Cross Stage Partial with Fused Layers

⁴Extracted from “Bounding Box Regression Loss”, n.d.

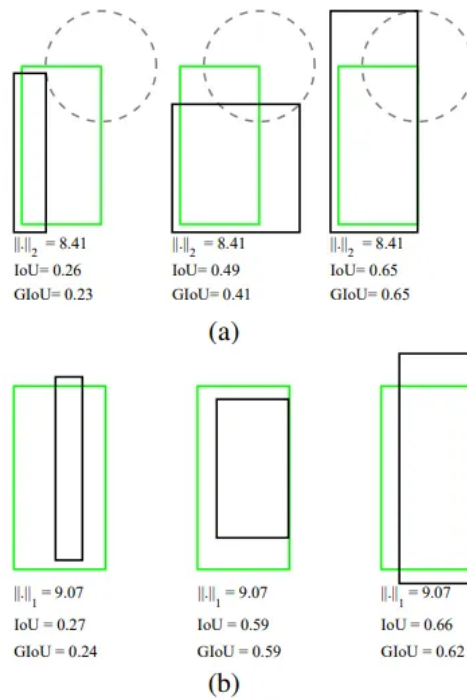


Figure 4: Bounding boxes: Intersection over Union (IoU) and Generalized IoU (GIoU) in bounding box regression, showing how GIoU better accounts for spatial alignment when boxes don't overlap perfectly, improving training stability.⁴

212 (C2f). C2f splits the feature map channels and fuses, or combines, the features from the channels
213 progressively. By gradually fusing channels, the object detector is able to grab richer details rather
214 than fusing all the channels in one shot like CSP does. A Spatial Pyramid Pooling-Faster (SPPF)
215 module, a component in the CNN, is used to see objects in different sizes. Essentially, it uses
216 windows – parts of the image – of different sizes so then, it's able to grab minute details as well as
217 the bigger picture. For example, a small window can be used to identify that a dog's fur is more
218 wavy than straight. By contrast, a big window would look at the dog entirely, and identify that the
219 object is in the shape of a dog.

220 In the head of YOLOv8s, Upsample layers (U layers in Figure 5) are used to improve the
221 resolution of the feature maps. Doing this helps preserve more details, which can result in greater
222 precision and accuracy. Three separate branches, or pathways, are used for objectness, predicting
223 if an object is apparent; classification, identifying what object that is; and regression, predicting
224 where the bounding box should be placed.

⁵Extracted from Jain et al., 2025

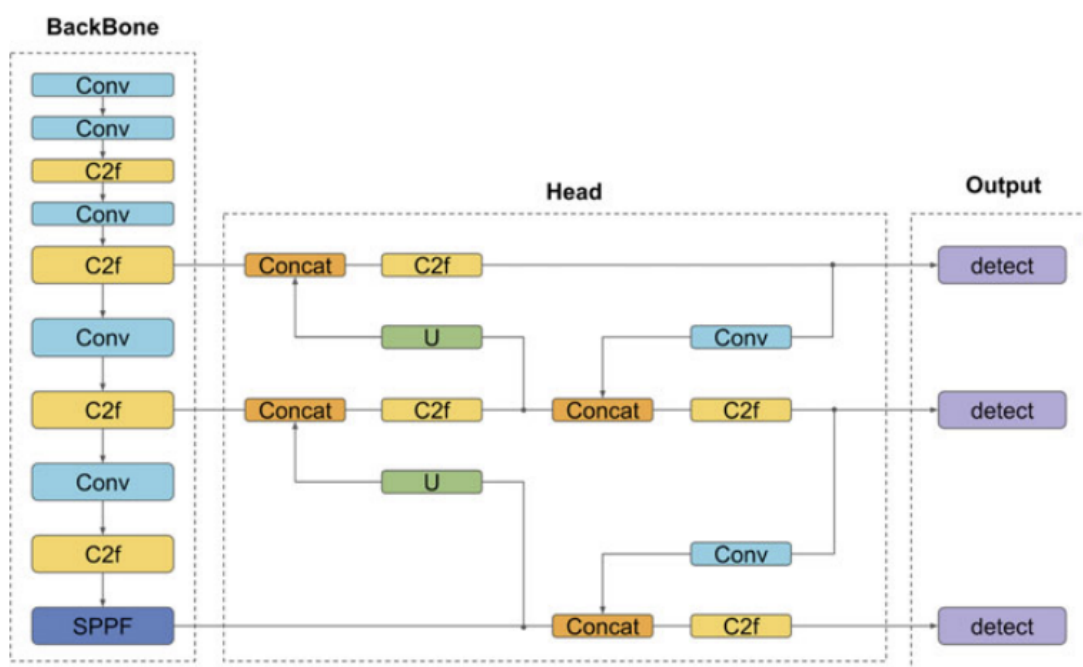


Figure 5: YOLOv8 architecture: the backbone extracts multi-scale features, the head performs feature aggregation, and the output layer generates object detection predictions⁵

Version, Year	Backbone	Neck	Head	Key innovations / training updates	Pros / Cons (esp. for E-waste)
YOLOv3 2018	Darknet-53 (residual)	FPN-like multi-scale fusion	3-scale heads with anchors	Stable baseline; incremental over YOLOv2; residual connections improve depth	+ Robust and widely reproduced baseline. – Weaker recall for small objects, limiting E-waste sensitivity.
YOLOv4 2020	CSPDarknet-53	SPP + PAN	3-scale heads with anchors	Introduced “bag of freebies/costlies,” CIoU/GIoU, mosaic augmentation, stronger training recipe	+ Major accuracy gain, better on small/medium objects. – Higher complexity and compute demand.
YOLOv5s 2020	CSP-inspired (PyTorch)	PANet	3-scale heads with anchors	PyTorch re-implementation; small “s” variant optimized for speed; auto-augment, EMA; easy deployment	+ Flexible, lightweight, edge-ready. – “s” favors speed over recall; in our tests, default thresholds missed subtle E-waste (see YRFNv5).
YOLOv8s 2023	C2f (CSP-fused)	SPPF + PAN	Decoupled obj/cls/reg heads (anchor-free default)	Modernized architecture; anchor-free prediction; cleaner training pipeline; improved feature reuse	+ Strong precision, efficient pipeline. – In our results, default setup under-recalled E-waste; improved when FN-reduction applied (YRFNv8).

Table 1: Concise comparison of YOLO variants evaluated. “s” variants (v5s, v8s) are chosen to match real-time, resource-constrained conveyor-belt settings.

A summary of the comparisons are captured in Table 1. As discussed in the preceding subsections, YOLOv3, v4, v5s, and v8s each represent critical points in the evolution of the YOLO family of detection algorithms. v3 is the canonical baseline. v4 is the first significant performance jump. v5s is the widely adopted PyTorch re-implementation. Finally, v8s is the most recent unified architecture. We pick the “s” variants for v5 and v8 to align better with the real-time, resource-constrained, conveyor-belt based sorting scenario. This spectrum of methods allows our analysis to capture both historical and state-of-the-art behaviors under the E-waste misclassifications cost framework.

Methodology

Standard evaluation measures

Very often, when analyzing object detection models, the common metrics utilized are accuracy, precision and recall. Sometimes, though more rarely, the F1 scores. Accuracy measures the percentage of predictions that are correct. Accuracy, given in Equation (1), is the most popular measure for classification and is perfect for cases where making any misclassification is equally bad. All misclassifications hurt the measure the same way.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

In the above, equation, True Positive (TP): Object detector successfully detects E-waste. True Negative (TN): When a non-E-waste is identified by the object detector successfully as non-E-waste. False Positive (FP): When non-E-waste is identified by the object detector incorrectly as E-waste. False Negative (FN): When E-waste is identified by the object detector wrongly as non-E-waste.

243 Precision, given by Equation (2), is also called the positive predictive value and measures how
244 many were correct, among all those predicted positive. In other words, it does not care about false
245 negatives (FNs). False positives (FPs) alone, hurt the measure. This is very useful in applications
246 such as resume screening, where even missing out a good resume is not as bad as hiring a person
247 who is not the right fit.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

248 Recall, given in Equation (3), is also called Sensitivity or the True positive rate. It asks how
249 many the model correctly detected out of all the objects that were actually positive, that is in our
250 case, E-waste. It tells us the total detected E-waste fraction among all E-wastes. In other words,
251 it does not care about FPs and only FNs hurt the measure. This is perfect for applications such as
252 cancer screening. In our case though, we do care about FPs as well, just less so. That is, labeling a
253 tennis ball as E-waste, that is FP, is bad, but not nearly as bad as labeling a battery as a non-E-waste,
254 that is FN. Moreover, recall can be trivially maximized by labeling every item an E-waste. Hence
255 recall can never be used as an objective when training ML methods.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

256 Finally, the F1-score, Equation (4) balances precision and recall, and is relevant in applications
257 where both are equally expensive.

$$F_1 = 2 \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

Reduced FN models 258

Appending to the list of models described in the *Models* section, we create a duplicate of every 259 model, but lower the confidence levels from 0.5 to 0.25, favoring detection. By lowering the 260 confidence level, the number of FNs is guaranteed to decrease since it increases the overall positivity 261 rate. Confidence level is the minimum confidence required for the object detector to make the 262 final decision that the target object is in the picture. We name these models YRFNv3, YRFNv4, 263 YRFNv5, YRFNv8 respectively. YRFN stands for YOLO with Reduced False Negatives. 264

EMC: A New Performance Metric for E-waste detection 265

As we have already discussed, each of the usual measures of success used in classification do 266 not lend themselves well to E-waste detection. Accuracy, for example, undermines this situation 267 completely, considering both FNs and FPs the same. Our proposed new measure seeks to balance 268 FNs and FPs by accounting for the costs of each type of misclassification. 269

To determine these costs, we looked at the cost of offsetting the environmental consequences 270 and the costs of inspection of false positives. Looking at a cost benefit analysis literature (Yang 271 et al., 2021) it was highlighted that the cost to offset the environmental consequences was on average 272 4 USD/kg of E-waste, in 2021. Adjusted for inflation, this is 5.95 USD/kg today (2025). Averaging 273 common trash product weights (Empa - Swiss Federal Laboratories for Materials Science and 274 Technology, 2025), we obtain 11.71 USD/item. To find the cost of false positives, we take the 275 average worker's hourly salary, and assuming they spend five minutes on each object, we evaluate 276 the appropriate cost. According to (Wikipedia contributors, 2024), the average recycling worker's 277 hourly salary in China is \$0.17, while it is 32,000/year according to ZipRecruiter in the US. Though a 278

279 wide variation, averaging five objects per minute, give us 0.67 USD/item. These are rough estimates
280 and hence it is important to note that this is only for comparison purposes. Our new method's
281 calibration and performance do not depend on the exact values, only that we would need non-zero
282 values for both.

283 Below is the new metric that we propose, the E-waste misclassification cost (EMC):

$$\text{EMC} = \frac{11.71 FN + 0.67 FP}{TP + FN} \quad (5)$$

284 **Ensembles**

285 An ensemble method is the combining of multiple models to make a prediction - wisdom of the
286 crowd. Different models may give different results on whether they're able to detect an object or
287 not. To make the final decision, a voting system is used such that if more models claim that an
288 object is apparent, then the final prediction is that the object is present and vice-versa.

289 In this study, we created an ensemble of YOLOv3, YOLOv4, YOLOv5, YOLOv8 and another
290 ensemble of RFNv3, YRFNv4, YRFNv5, YRFNv8. If at least three of the models detected E-waste
291 as present, then the final decision would be that E-waste is there. We call them EYOLO and EYRFN.
292 These ensembles estimate labels by polling the 4 YOLOs and the 4 YRFNs, respectively.

293 **WISE: Waste-focused Integrated Smart Ensemble**

294 Consider an ensemble that does not allow equal weighting. That is, the wisdom of the crowd is
295 not equally weighted. The better an underlying method, the more its weight. Further, we will let
296 machine learning pick the weights that minimize the total EMC as defined by Equation (5).

More specifically, let $x_i \in \{0, 1\}^8$ denote the vector of model-level detections (YOLOv3, YOLOv4, YOLOv5s, YOLOv8s, and YRFNv3, YRFNv4, YRFNv5s, YRFNv8s) for image i , and let $y_i \in \{0, 1\}$ be the ground truth label (1 for E-waste, 0 for non-E-waste). WISE learns a weight vector $w \in \mathbb{R}^8$ and bias b by minimizing the expected misclassification cost with an ℓ_1 penalty:

$$\min_{w,b} \frac{1}{N} \sum_{i=1}^N \left(c_{\text{FN}} y_i [1 - \hat{y}_i] + c_{\text{FP}} (1 - y_i) \hat{y}_i \right) + \lambda \|w\|_1, \quad (6)$$

where the prediction is made as

$$\hat{y}_i = \mathbf{1}[w^\top x_i + b \geq 0]. \quad (7)$$

We set $c_{\text{FN}} = 11.71$ and $c_{\text{FP}} = 0.67$ (as derived from Equation (5), the EMC metric). Optimization is performed with the Adam optimizer (batch size 256, learning rate 10^{-3}), with early stopping on a stratified validation split. The threshold 0 is without loss of generality since the bias b can be appropriately adjusted.

The learned solution is sparse, with only two non-zero weights: YRFNv4 receives 0.6 weight and YRFNv8 receives 0.4 weight. This sparsity results from the ℓ_1 penalty and is consistent with the high performance of these models in Table 2.

Note that since we are learning these weights to minimize a loss, we will have to train the weights on a training subset of the data and test it on the rest of the test set. We will use a standard randomization of 80/20 to split the data for testing versus training. We also tune λ over the grid $\{0, 10^{-4}, 10^{-3}, 10^{-2}\}$. For comparisons to other methods above, that did not require training/testing splits, we will report and compare the averages over the entire dataset. This obviously risks that the

314 performance improvements could be biased due to the inclusion of in-sample and out-of-sample
315 data. To ensure that this is not the case, we will compare the EMC of the test set to that of the
316 training to ensure that magnitudes are similar.

317 **Dataset**

318 A mixture of multiple datasets was used to test. We did not pick images that were used for
319 original YOLO training. We utilized Kaggle's E-waste dataset which consisted of mobile phones,
320 microwaves, keyboards and mice. There were 300 images of each giving a total of 1200 images
321 of E-waste. Another 1200 images of non-E-waste were mixed with the E-waste images. These
322 non-E-waste images came from images.cv and roboflow. They all fell under various categories like
323 handbags, chairs, spoons, books, racquets, and umbrellas. Each model was given 2400 images for
324 testing, 1200 consisting of E-waste and 1200 consisting of non-E-waste. To evaluate, the number
325 of true positives, false positives, true negatives, and false negatives were counted.

326 All datasets, preprocessing scripts, and trained models are publicly available at our GitHub
327 repository (GitHub Contributors, 2025).

328 **Results**

329 For each of the methods: classical - YOLOv3, YOLOv4, YOLOv5, YOLOv8; tweaked - YRFNv3,
330 YRFNv4, YRFNv5, YRFNv8; ensemble - EYOLO, EYRFN; Table 2 shows the confusion matrices
331 for each of these methods. As one can observe the traditional YOLOs, though impressive for object
332 detection, do not fare too well for the purposes of E-waste detection. More specifically YOLOv3
333 identifies every object as a non-E-Waste resulting in 1200 TNs and 1200 FNs. This is potentially

due to overfitting during the YOLOv3 training. 334

While YOLOv4, YOLOv5 and YOLOv8 perform better, they still weigh FPs and FNs the same. 335
Very high false negatives (FNs) will essentially cause too many toxic substances to enter landfills 336
and find their ways into living organisms. The YRFNs each have relatively lower FNs. What is 337
also surprising is that the cutting edge YOLO, that is version 8, which is the go-to algorithm today 338
for object detection, is far from the best for E-waste detection. 339

Table 2 also shows the confusion matrix for newly developed smart ensemble - WISE (Waste- 340
focused Integrated Smart Ensemble). It has the lowest FNs amongst all methods compared. The 341
learned weights were 0 for six of the 8 method (4 YOLOs + 4 YRFNs). The only two non-zero 342
weights were YRFNv4: 0.6 and YRFNv8: 0.4 with threshold almost just above zero. 343

A final comparison of all methods in terms of the popular measures of success and our EMC 344
measure are shown in Table 2. As one can see, the smart ensemble out performs every other method 345
significantly. Moreover the magnitude of EMC is in USD per object and can hence easily lend itself 346
to a relatable intuitive value. Table 2, also bolds the best values for each measure of success. 347

Complete confusion matrices and additional performance breakdowns are provided in our 348
GitHub repository (GitHub Contributors, 2025). 349

Statistical Significance Comparisons 350

While raw performance metrics discussed in the previous section provide useful summaries, they do 351
not by themselves establish whether observed differences are statistically significant. To rule out 352
the possibility that they could have been due to chance variations, in this section, we conducted 353
pairwise comparisons using McNemar's test (McNemar, 1947). The McNemar's test is specifically 354

Model	TP	FP	TN	FN	Accuracy	Precision	Recall	EMC
YOLOv3	0	0	1200	1200	0.50	NA	0.00	11.71
YRFNv3	0	0	1200	1200	0.50	NA	0.00	11.71
YOLOv4	756	4	1196	444	0.81	0.99	0.63	4.33
YRFNv4	851	11	1189	349	0.85	0.99	0.71	3.41
YOLOv5	209	5	1195	991	0.59	0.98	0.17	9.67
YRFNv5	373	14	1186	827	0.65	0.96	0.31	8.08
YOLOv8	176	6	1194	1024	0.57	0.97	0.15	10.00
YRFNv8	340	22	1178	860	0.63	0.94	0.28	8.40
EYOLO	90	0	1200	1110	0.54	1.00	0.08	10.83
EYRFN	203	0	1200	997	0.58	1.00	0.17	9.73
WISE - Smart(learned)	906	25	1175	294	0.87	0.97	0.76	2.88

Table 2: Performance comparison of different models.

355 designed to evaluate differences between two classifiers tested on the same set of data points.

356 Given two models A and B , McNemar’s test constructs a 2×2 contingency table based on

357 whether each classifier prediction is correct or incorrect. The key quantities are n_{10} (instances where

358 A is correct but B is wrong) and n_{01} (instances where A is wrong but B is correct). Under the null

359 hypothesis that both classifiers have equal error rates, these counts should be approximately equal.

360 The test statistic is defined as

$$\chi^2 = \frac{(|n_{01} - n_{10}| - 1)^2}{n_{01} + n_{10}},$$

361 which asymptotically follows a χ^2 distribution with one degree of freedom. A small p -value indicates

362 that the classifiers differ significantly in their performance.

363 Figure 6 presents a heatmap of the pairwise McNemar p -values. It is displayed on a $-\log_{10}(p)$

364 scale for easier readability - large values are significant and colored yellow. As shown, almost

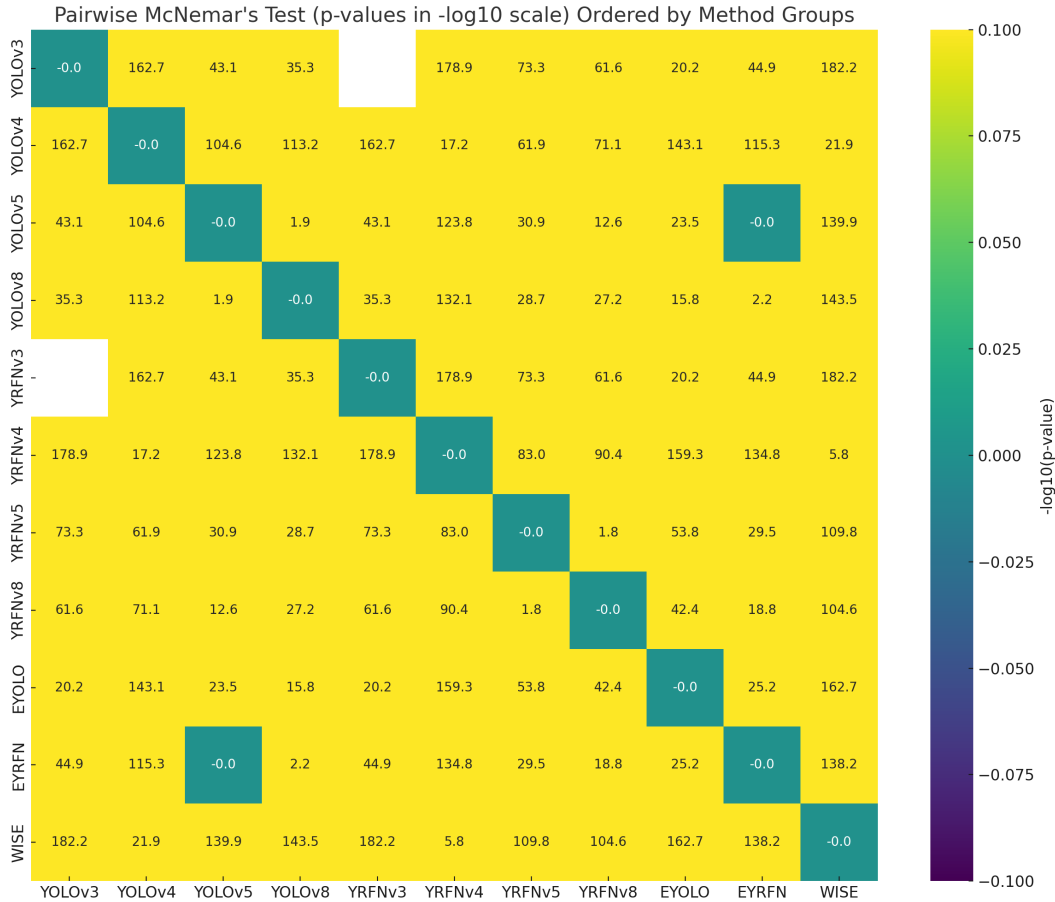


Figure 6: Pairwise McNemar’s test results across all methods. The heatmap displays $-\log_{10}(p)$ values, where larger magnitudes correspond to stronger evidence against the null hypothesis of equal performance.

all off-diagonal entries exhibit extremely small p -values, well below conventional significance thresholds (e.g., $p < 0.05$). This establishes that the observed performance gaps across methods are not attributable to chance.

Methods that have identical predictions (**YOLOv3 vs. YRFNv3**) appear visually as blank/white in the heat map since in such cases the McNemar’s statistic is undefined (no discordant pairs). We also notice that EYRFN is not statistically different from YOLOv5.

These findings are reassuring that the comparative rankings from Table 2 are not due to chance. Without such an analysis, improvements cannot be trusted to have come from well-grounded

373 performance improvements.

374 **Limitations and future work**

375 Our 2,400-image corpus mixes curated E-waste and non-E-waste classes but is modest in size and
376 lacks nuances of in-the-wild images. There might be cluttering, stacking, off-the-frame issues etc.
377 on sorting lines and conveyor belts. In-situation tests and refinements are a subject of future studies.

378 Assumptions on the cost accounting for EMC uses approximate unit costs from literature and
379 wage proxies. We do adjust the costs for inflation. However, we have one EMC measure that is
380 calibrated to costs in the United States, which produces the largest quantity of E-waste in the world.
381 However, we would need to refine this measure of different economies in the future.

382 Finally, we have not studied any of the real-time constraints like latency and throughput. Fur-
383 ther improvements can help create ensembles that also consider the underlying models' latency
384 and throughput to combine them to yield a smart ensemble that matches the expected real-time
385 performance thresholds will be the subject of future studies.

386 **Conclusion**

387 The lower the EMC, the better the model is at detecting E-waste and reducing overall cost. Though
388 the measure, EMC, weighs the estimated cost of FNs against that of FPs, the goal is not primarily
389 cost reduction. The goal is the reduction of E-waste that goes into landfills, for any given budget.
390 This is critical especially since the ever growing volume of E-waste far exceeds the rate at which
391 budgets to tackle E-waste do not grow.

392 Many surprising observations were made in the results section. Classical cutting-edge methods

like YOLOv8 (Jain et al., 2025), have terrible performance for E-waste detection, although their 393
performance for general object detection is commendable. Further thought would reveal that this is 394
not totally surprising, since they were trained and evaluated on a generic object detection context 395
and with accuracy as the primary measure. Hence, it might not be fair to compare, or to use such 396
methods for E-waste detection. Next, tweaked methods outperform the classical methods whose 397
default parameters would have been carefully chosen by the authors. This again shows that even 398
parameters calibrated for generic object detection do not work well for E-waste detection. We also 399
confirmed that almost all model pairs differ in a statistically robust way and more importantly our 400
smart ensemble, WISE, is statistically better than all the other models. 401

Our results align with reports that off-the-shelf detectors can degrade under domain shift 402
(Solovyev et al., 2021). Our results also align with broader ML findings that cost-sensitive training 403
and ensembling can improve decision quality when costs are asymmetric (Domingos, 1999; Elkan, 404
2001; Wolpert, 1992). In E-waste operations, prioritizing lower FNs as opposed to FPs by an 405
EMC-minimizing ensembles like WISE is a principled path forward. Our smart ensemble does 406
surprisingly and fortunately well in the context of E-waste detection. It not only has a significant cost 407
reduction, it also has the lowest number of FNs in the entire comparison set. While a methodology 408
that reduces FNs to zero while not labeling every object as FP is ideal, that would be for future 409
research. 410

411 **References**

- 412 Adedeji, O., & Wang, Z. (2019). Intelligent waste classification system using deep learning convo-
413 lutional neural network. *Procedia Manufacturing*. [https://doi.org/10.1016/j.promfg.2019.05.](https://doi.org/10.1016/j.promfg.2019.05.086)
414 086
- 415 Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). Yolov4: Optimal speed and accuracy of
416 object detection. <https://arxiv.org/abs/2004.10934>
- 417 Bodla, N., Singh, B., Chellappa, R., & Davis, L. S. (2017). Soft-nms – improving object detection
418 with one line of code. *Proceedings of the IEEE International Conference on Computer Vision*
419 *(ICCV)*, 5561–5569. [https://openaccess.thecvf.com/content_ICCV_2017/papers/Bodla_Soft-](https://openaccess.thecvf.com/content_ICCV_2017/papers/Bodla_Soft-NMS_--_Improving_ICCV_2017_paper.pdf)
420 [NMS_--_Improving_ICCV_2017_paper.pdf](https://openaccess.thecvf.com/content_ICCV_2017/papers/Bodla_Soft-NMS_--_Improving_ICCV_2017_paper.pdf)
- 421 *Bounding box regression loss* [Author not listed; wiki page]. (n.d.). CloudFactory Computer Vision
422 Wiki. Retrieved September 6, 2025, from [https://wiki.cloudfactory.com/docs/mp-wiki/loss/](https://wiki.cloudfactory.com/docs/mp-wiki/loss/bounding-box-regression-loss)
423 [bounding-box-regression-loss](https://wiki.cloudfactory.com/docs/mp-wiki/loss/bounding-box-regression-loss)
- 424 Cheng, X., Cheng, Y., Gong, Z., & Chen, Y. (2021). An improved small object detection method
425 based on yolov3. *Multimedia Tools and Applications*, 80(19), 29329–29344. [https://doi.org/](https://doi.org/10.1007/s11042-021-10605-5)
426 [10.1007/s11042-021-10605-5](https://doi.org/10.1007/s11042-021-10605-5)
- 427 Domingos, P. (1999). Metacost: A general method for making classifiers cost-sensitive. *Proceedings*
428 *of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data*
429 *Mining (KDD)*, 155–164. <https://doi.org/10.1145/312129.312220>
- 430 Elkan, C. (2001). The foundations of cost-sensitive learning. *Proceedings of the 17th International*
431 *Joint Conference on Artificial Intelligence (IJCAI)*, 973–978. [https://cseweb.ucsd.edu/](https://cseweb.ucsd.edu/~elkan/rescale.pdf)
432 [~elkan/rescale.pdf](https://cseweb.ucsd.edu/~elkan/rescale.pdf)

Empa - Swiss Federal Laboratories for Materials Science and Technology. (2025). Ewaste guide 433
 [Accessed: 2025-08-24]. 434

Fan, J., Cui, L., & Fei, S. (2023). Waste detection system based on data augmentation and yolo_ec. 435
Sensors, 23(7), 3646. <https://doi.org/10.3390/s23073646> 436

Feng, M., Jiang, T., Wang, J., Yang, L., & Zhang, Y. (2023). A multi-scale traffic object detection 437
 algorithm for road scenes based on improved yolov5. *Sensors*, 23(4), 2045. [https://doi.org/](https://doi.org/10.3390/s23042045) 438
 10.3390/s23042045 439

Fornalczyk, A., Willner, J., Francuz, K., & Cebulski, J. (2013). E-waste as a source of valuable 440
 metals. *Archives of Materials Science and Engineering*, 63(2), 87–92. [http://www.amse.](http://www.amse.acmsse.h2.pl/vol63_2/6325.pdf) 441
[acmsse.h2.pl/vol63_2/6325.pdf](http://www.amse.acmsse.h2.pl/vol63_2/6325.pdf) 442

GitHub Contributors. (2025). Wise – <https://github.com/vishnumuthuraman/wise.git> [Accessed: 443
 2025-08-24]. 444

Huang, J., Rathod, V., Sun, C., Zhu, M., et al. (2017). Speed/accuracy trade-offs for modern 445
 convolutional object detectors. *CVPR*. 446

Ibrahim, A. N., Youssif, E. S., Hassanien, A. E., Hossam, A. A., Abdo, S. S., & Saleh, A. M. D. 447
 (2023). Object detection-based automatic waste segregation using robotic arm. *International* 448
Journal of Advanced Computer Science and Applications (IJACSA). [https://doi.org/10.](https://doi.org/10.14569/IJACSA.2023.0141164) 449
 14569/IJACSA.2023.0141164 450

Jain, H., Jain, J., Shrivastava, V., Pandey, A., & Sharma, S. (2025). Yolov8 and its advancements. 451
International Journal of Research Publication and Reviews, 6(5), 660–666. Retrieved 452
 September 6, 2025, from <https://ijrpr.com/uploads/V6ISSUE5/IJRPR44604.pdf> 453

Jocher, G., & Ultralytics. (2020). Yolov5. <https://github.com/ultralytics/yolov5> 454

Jocher, G., & Ultralytics. (2023). Yolov8. <https://github.com/ultralytics/ultralytics> 455

456 Kán, P., & Kaufmann, H. (2019). Deeplight: Light source estimation for augmented reality using deep
457 learning. *The Visual Computer*, 35, 873–883. <https://doi.org/10.1007/s00371-019-01666-x>

458 Khan, H. A., Naqvi, S. S., Alharbi, A. A. K., Alotaibi, S., & Alkhatami, M. (2024). Enhancing trash
459 classification in smart cities using federated deep learning. *Scientific Reports*, 14, 11816.
460 <https://doi.org/10.1038/s41598-024-62003-4>

461 Li, K., & Cao, L. (2020). A review of object detection techniques. *2020 5th International Conference
462 on Electromechanical Control Technology and Transportation (ICECTT)*, 385–390. <https://doi.org/10.1109/ICECTT50890.2020.00091>

463

464 Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection.
465 *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2999–3007.
466 <https://doi.org/10.1109/ICCV.2017.324>

467 Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2016). Ssd:
468 Single shot multibox detector. *ECCV*.

469 Majchrowska, S., Mikołajczyk, A., Ferlin, M., Klawikowska, Z., Plantykowski, M. A., Kwasigroch, A.,
470 & Majek, K. (2022). Deep learning-based waste detection in natural and urban environments.
471 *Waste Management*, 138, 274–284. <https://doi.org/10.1016/j.wasman.2021.12.001>

472 McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions
473 or percentages. *Psychometrika*, 12(2), 153–157.

474 Meituan Vision AI Department. (2022). Yolov6: A single-stage object detection framework for
475 industrial applications. <https://github.com/meituan/YOLOv6>

476 Noel-Brune, M., Goldizen, F. C., Neira, M., van den Berg, M., Lewis, N., King, M., Suk, W. A.,
477 Carpenter, D. O., Arnold, R. G., & Sly, P. D. (2013). Health effects of exposure to e-waste.
478 *The Lancet Global Health*, 1(2), e70. [https://doi.org/10.1016/S2214-109X\(13\)70020-2](https://doi.org/10.1016/S2214-109X(13)70020-2)

- Oza, P., Agrawal, S., Kapadia, M., & Raotole, O. (2025). Solid waste classification using deep neural network: A transfer learning approach. *Earth Science Informatics*, 18, 255. <https://doi.org/10.1007/s12145-025-01743-x>
- Parti, A. (2024). Understanding object detection: A comprehensive guide [Retrieved March 18, 2024, from <https://pareto.ai/blog/object-detection>].
- Paudel, P., Shrestha, S., Shrestha, S., Gurung, S., & Adhikari, S. (2024). Automated waste sorting with delta arm and yolov8 detection. *Journal of Artificial Intelligence and Capsule Networks*, 6(3), 299–315. <https://doi.org/10.36548/jaicn.2024.3.004>
- Potter, R. (2022, August 23). *How bounding box object detection is taking over and what to do about it* [Analytics on Medium]. Retrieved September 6, 2025, from <https://medium.com/analytics/how-bounding-box-object-detection-is-taking-over-and-what-to-do-about-it-9130b6d6c1f1>
- Proença, P. F., & Simões, P. (2020). TACO: Trash annotations in context for litter detection. *arXiv*. <https://arxiv.org/abs/2003.06975>
- Proença, P. F., & Simões, P. (2023). TACO dataset website [Accessed 2025-09-05].
- Promboonruang, S., Boonrod, T., & Ratchatawetchakul, Y. (2024). Efficient waste detection and classification based on yolov5 models. *Engineering Access (MIJET)*, 10(1), 51–58. <https://doi.org/10.14456/mijet.2024.7>
- Rahim, L. A. A., Abidin, N. A. Z., Aminuddin, R., Samah, K. A. F. A., Ibrahim, A. Z. M., Yusoh, S. D., & Nasir, S. D. N. M. (2024). YOLOv5 model-based real-time recyclable waste detection and classification system. In *Innovations in smart cities applications volume 7 (sca 2023)* (pp. 44–54, Vol. 906). Springer, Cham. https://doi.org/10.1007/978-3-031-53824-7_5

500 Rajeev, P. A., Dharewa, V., Lakshmi, D., Vishnuvarthanan, G., Giri, J., Sathish, T., & Alrashoud, M.
501 (2025). Advancing e-waste classification with customizable YOLO-based deep learning
502 models. *Scientific Reports*, *15*, 18151. <https://doi.org/10.1038/s41598-025-94772-x>

503 Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time
504 object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern
505 Recognition (CVPR)*, 779–788. <https://doi.org/10.1109/CVPR.2016.91>

506 Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. [https://arxiv.org/abs/
507 1804.02767](https://arxiv.org/abs/1804.02767)

508 Sheng, V. S., & Ling, C. X. (2006). Thresholding for making classifiers cost-sensitive. *Proceedings
509 of the AAAI Conference on Artificial Intelligence*, 476–481. [https://cdn.aaai.org/AAAI/
510 2006/AAAI06-076.pdf](https://cdn.aaai.org/AAAI/2006/AAAI06-076.pdf)

511 Singh, Y., & Parimala S. (2025, March 26). E-waste management. [https://doi.org/10.2139/ssrn.
512 5196804](https://doi.org/10.2139/ssrn.5196804)

513 Solawetz, J. (2024, January 4). *What is yolov4? a detailed breakdown*. Roboflow. Retrieved Septem-
514 ber 6, 2025, from <https://blog.roboflow.com/a-thorough-breakdown-of-yolov4/>

515 Solovyev, R., Wang, W., & Gabruseva, T. (2021). Weighted boxes fusion: Ensembling boxes
516 from different object detection models. *Image and Vision Computing*, *107*, 104117. [https:
517 //doi.org/10.1016/j.imavis.2021.104117](https://doi.org/10.1016/j.imavis.2021.104117)

518 Song, L., Yu, H., Yang, Y., Tong, Y., Ren, S., & Ye, C. (2025). Yolo-vg: An efficient real-time
519 recyclable waste detection network. *Journal of Real-Time Image Processing*. [https://doi.org/
520 10.1007/s11554-025-01655-5](https://doi.org/10.1007/s11554-025-01655-5)

521 Thung, G., & Yang, M. (2017). Trashnet [Accessed 2025-09-05].

- Vats, M. C., & Singh, S. K. (2014). E-waste characteristic and its disposal. *International Journal of Ecological Science and Environmental Engineering*, 1(2), 49–61. 522
523
- Wang, C.-Y., Bochkovskiy, A., & Liao, H.-Y. M. (2022). Yolov7: Trainable bag-of-freebies sets 524
new state-of-the-art for real-time object detectors. <https://arxiv.org/abs/2207.02696> 525
- Wikipedia contributors. (2024). *Electronic waste in Guiyu* [Wikipedia article]. Retrieved August 526
24, 2025, from https://en.wikipedia.org/wiki/Electronic_waste_in_Guiyu 527
- Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, 5(2), 241–259. 528
- Wu, P., Chai, B., Nie, X., Yan, L., Wang, Z., Zhou, Q., Wang, B., Peng, Y., & Li, H. (2024). 529
Enhanced object detection: A study on vast vocabulary object detection track for v3det 530
challenge 2024 (version 3). <https://doi.org/10.48550/arXiv.2406.09201> 531
- Wu, T., & Zhu, S.-C. (2013). Learning near-optimal cost-sensitive decision policy for object 532
detection. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 533
753–760. <https://doi.org/10.1109/ICCV.2013.98> 534
- Yang, W.-D., Sun, Q., & H-G., N. (2021). Cost-benefit analysis of metal recovery from e-waste: 535
Implications for international policy. *Waste Management*, 123, 42–47. [https://doi.org/10. 536
1016/j.wasman.2021.01.023](https://doi.org/10.1016/j.wasman.2021.01.023) 537

Response Letter

1

Dear Editorial Team,

2

We are glad to see the conditional acceptance of the paper and would like to sincerely thank you for quick and constructive feedback that has helped us create a stronger version. We are attaching a new revised version. This version addresses all the raised concerns. A point-by-point reply to each of the reviewers' comments are provided below.

3

4

5

6

Response to Reviewer 1 Comments

7

R1-01

8

Contextualization of YOLO Architectures: The descriptions of the YOLO model are detailed but somewhat mechanical. It would strengthen the paper to briefly contextualize these models within the broader literature, this would help readers understand not just what was implemented, but why these versions were selected.

9

10

11

12

We agree that this will tie the existing descriptions of the YOLO methods to why these versions were selected. We have now added this paragraph after the YOLO descriptions that ties them together:

As discussed in the preceding subsections, YOLOv3, v4, v5s, and v8s each represent critical points in the evolution of the YOLO family of detection algorithms. v3 is the canonical baseline. v4 is the first significant performance jump. v5s is the widely adopted PyTorch re-implementation. Finally, v8s is the most recent unified architecture. We pick the “s” variants for v5 and v8 to align better with the real-time, resource constraint, conveyor-belt based sorting scenario. This spectrum of methods allows our analysis to capture both historical and state-of-the-art behaviors under the E-waste misclassifications cost framework.

R1-02

13

Clarity on WISE's Learning Process: The machine learning process used to train WISE's weights is mentioned but not sufficiently detailed. A brief explanation of the optimization algorithm and how overfitting was mitigated would improve reproducibility and methodological transparency.

14

15

16

17

We now, in the “WISE: Waste-focused Integrated Smart Ensemble” section, fully specify the learning objective, regularization and calibration/thresholding, making reproduction straightforward. The complete formulation of the optimization problem is included. The code we uploaded on github and referenced in the previous version did have this implemented. We do agree that writing out the specifics in the paper does help understand the details.

18 **R1-03**

19 *Discussion of Limitations: The paper would be strengthened by a dedicated section dis-*
20 *cussing limitations. For instance, the dataset, while useful, is relatively small and may not*
21 *represent all e-waste scenarios. Generalizability to real-world, cluttered, or occluded environ-*
22 *ments is not addressed. It is suggested that an Introduction to Limitations subsection be added*
23 *before the conclusion to discuss dataset scope, cost assumptions, and real-world applicability.*

24

We have now included this subsection:

Our 2,400-image corpus mixes curated E-waste and non-E-waste classes but is modest in size and lack nuances of images that might be obtained in-the-wild. There might be cluttering, stacking, off-the-frame issues etc. on sorting lines and conveyor belts. In-situation tests and refinements are a subject of future studies.

Assumptions on the cost accounting for EMC uses approximate unit costs from literature and wage proxies. We do adjust the costs for inflation. However, we have one EMC measure that is calibrated to costs in the United States, that produces the largest quantity of E-waste in the world. However, we would need to refine this measure of different economies in the future.

Finally, we have not studied any of the real-time constraints like latency and throughput. Further improvements can help create ensembles that also consider the underlying models’ latency and throughput to combine them to yield a smart ensemble that matches the expected real-time performance thresholds, will be the subject of future studies.

25 **R1-04**

26 *Language and Reference: Some sections, particularly the introduction and background,*
27 *contain long sentences that could be broken down for better readability. Minor grammatical*
28 *errors and repetitive phrasing occur throughout the whole paper. For the References part, the*
29 *primary issues are a lack of formatting consistency and several instances of incomplete or*
30 *incorrect information.*

We have fixed all the typos and corrected the minor grammatical errors. We have tried our level best to correct the formatting and make it closer to the APA style requirements (including 12-pt, Times New Roman, double spaced etc.). We hope all formatting issues are now resolved.

R1-05 Formatting Some minor issues are showing in the document, such as the display of the paragraph shown as follows: 31
32

Also, the tables displayed in the reference part: 33

The author could check the required submission format guidelines provided by the Convergence journal, especially on the citation styles and other formatting requirements, spacing, pictures, etc. 34
35
36

We corrected the abstract margins. We have also double checked the citation styles and other formatting requirement. We are using Latex, but could not find a latex template for this journal and used standard APA Style template. We sincerely hope there are no other formatting issues.

Response to Reviewer 2 Comments 37

R2-01 38

The statement “Research on these two trash related applications, even though not directly related to E-waste, demonstrate the unacceptable performances of generic methods for specific application domains” (lines 85-87) after discussing the two other datasets being used for trash classification, seems unsubstantiated. In fact, these seem like great proof-of-concept trash classification systems being used in a different field where learnings can be applied for developing one specifically for E-waste. 39
40
41
42
43
44

Thanks for flagging this confusion. Our intent was not to compare e-waste to other waste datasets. Rather, we meant that off-the-shelf, general-purpose vision methods, like YOLO, that are trained on very broad images, often under-perform on specific applications - like the trash datasets. We wanted to use a closely related setting like waste detection to demonstrate this.

As you have suggested, we’ve softened our claim and revised the paragraph to state this clearly.

R2-02 45

46 “There are no proposed methodology improvements or enhancements specifically for
47 detecting E-waste.” (lines 108-109) It is unclear whether this is in relation to the prior E-waste
48 detection studies for accuracy, precision, recall or differentiating costs of FP and FN? This
49 line needs to mention what is exactly missing for the E-waste context.

■ We have now clarified this by rewriting the previous sentence to “None, to our knowl-
edge, consider the application level E-waste miss-classification costs that differentiate
FP and FN.”

50 **R2-03**

51 The end of the Introduction becomes a little confusing in that it is unclear what the actual
52 purpose of the paper is. Lines 97-105 discuss developing WISE, but then lines 113-119 discuss
53 other aims/contributions of the study. It would be best to consolidate these into one paragraph
54 to build up the rationale for the study. Further lines 277-284 also discuss the purpose/aims of
55 the paper. All of these paragraphs need to be consolidated so that there is a clear aim/s of the
56 paper in the one place.

■ Thanks for catching this. We now see the confusion and have cleaned it up. We have
consolidated all of these into a new subsection titled “Our Contributions.”

57 **R2-04**

58 Research papers usually have all the background information required to understand the
59 paper in the Introduction. This paper has subheadings and subsequent information for “Object
60 Detection” and “Models” etc after the Introduction. Authors should consider consolidating this
61 within the Introduction to fit a standard manuscript structure. Similarly, this would warrant
62 the removal of paragraph starting from line 120-126 as this is not needed for a manuscript as
63 the structure should be self-explanatory.

■ We have now made “Object Detection” and “Models”, both subsections of the Intro-
duction and removed the lines 120-126.

64 **R2-05**

65 All figure legends could be improved by provide more information eg. Figure 1: Object
66 Detection. An example of use of object detection is shown where the yellow boxes indicate X,
67 the blue boxes indicate Y, etc. This information should be supported by what is written in the
68 text.

For each of the figures 1 to 5, we have now added more information in the legend.

R2-06 69

“The benefits offered by improved one-stage methods” (lines 156-157) is unsubstantiated as the benefits have not been described. 70
71

We have now added that “One-stage detectors achieve favorable speed/accuracy trade-offs, often 2–5 times faster, making them practical for real-time applications. Prior work quantifies these trade-offs across architectures and scales.” With the relevant citation.

R2-07 72

Mention of Figure 2 in-text should likely come during discussion of two-stage and one-stage CNNs i.e. paragraph starting line 150. Currently it is being mentioned when talking about YOLO but the figure has no reference to YOLO. This is the case for Figure 4 Bounding boxes as well – perhaps should come at their first mention? 73
74
75
76

You are right, the Figure 2 reference was misplaced. We have fixed it. Figure 4 reference was rightly placed when referencing bounding box regression losses. However since the previous caption for Figure 4 was less descriptive, it felt misplaced. We have now fixed the caption.

R2-08 77

I am unsure whether such detailed descriptions of each YOLO is required. Perhaps this could be displayed in a table with dot points instead to highlight the similarities and differences between the four described? 78
79
80

We completely understand your perspective. An earlier version of the manuscript,

did not have these and was circulated to two faculty members at the University of Texas and the University of North Carolina. Both suggested that we include some description and history of the YOLO since our ensembles were fundamentally built on the underlying YOLO methods and hence they argued that this was required.

Moreover, Referee 1 has also made a comment to further include some details on why we choose these methods to compare against. The existing descriptions of the methods and their historical context allowed us to illustrate why these we picked. We sincerely hope that you understand.

However, your point on a table describing similarities and differences, we believe, will add additional value to the reader. We have not included that in Table 1.

81 **R2-09** *I feel tables 1a-5b could be better presented by having all in the one table, with the*
82 *models on the left and the results in columns. This would be more intuitive for comparisons*
83 *between the groups. Having read on, I see this has been completed in Table 7 which makes*
84 *tables 1a-5b redundant, and therefore tables 1a-5b should be removed.*

Agreed. We have removed Tables 1a-5b.

85 **R2-11** *There is no use of statistical analyses to determine whether findings are statistically*
86 *significant which means that it is not possible to say whether WISE is better or not than what*
87 *already exists at identifying E-waste.*

We looked into this and realized that this was a very critical thing to add to the paper. We managed to do pair-wise statistical significance tests for each pair of models and have now included an entire new subsection titled “Statistical Significance Comparisons.” This section discussed the McNemar test we did, and summarizes the results as a heatmap (below) that shows that almost all pair-wise comparisons are significant. Especially, it confirms that our method, WISE is statistically better than all the other methods.

88 **R2-12** *There is good engagement with published literature throughout the paper, until it*
89 *comes to the Conclusion. The Conclusion requires much more support from literature to place*
90 *the findings of this study in the wider context of the literature landscape for this topic area.*
91 *Authors should find literature that supports their findings or describes how the results are*
92 *surprising or differ from previously published studies.*

We have added that our results align with reports that off-the-shelf detectors can degrade under domain shift and also that we align with the broader ML findings that cost sensitive algorithms improve decision quality. We have included several appropriate citations now.

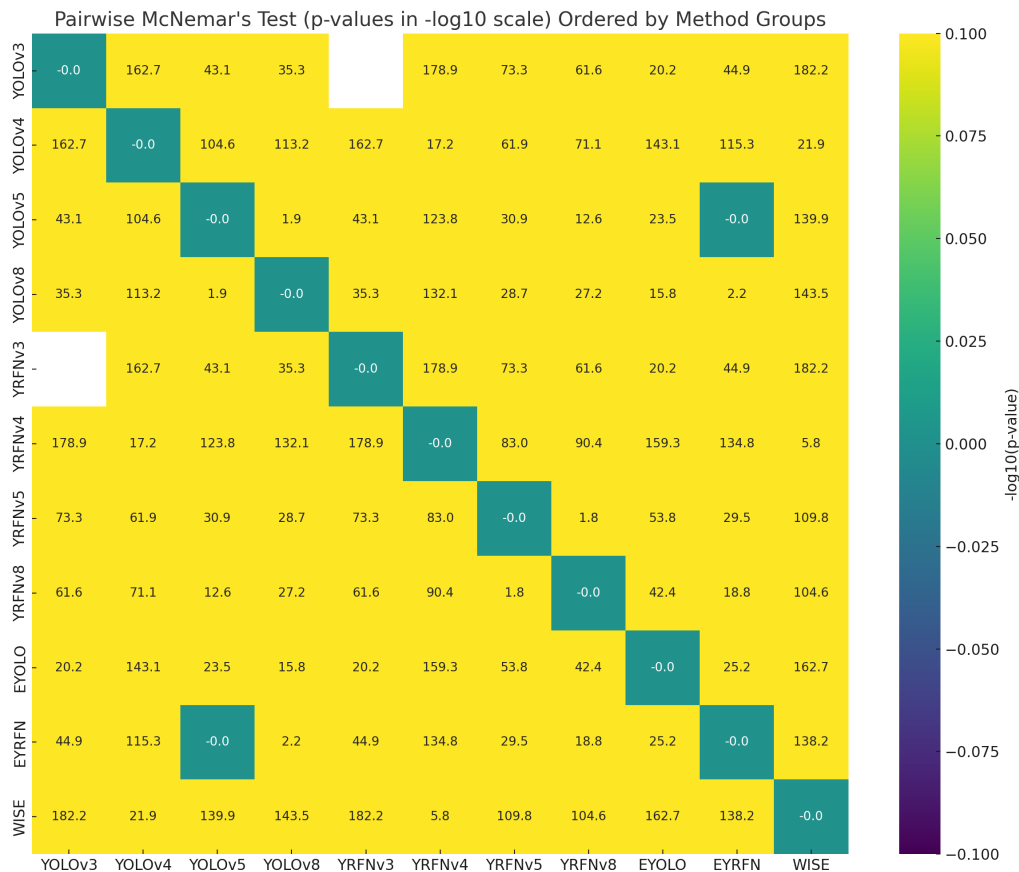


Figure 1: Pairwise McNemar’s test results across all methods. The heatmap displays $-\log_{10}(p)$ values, where larger magnitudes correspond to stronger evidence against the null hypothesis of equal performance.

R2-13 *There are minor grammatical errors throughout e.g. at the start of the Introduction, paragraph starting line 150, paragraph starting line 285. A read through of the entire paper, perhaps by a supervisor, is suggested.* 93
94
95

■ We have completed a complete read through and found the references minor grammatical errors. We believe we have fixed all of these.

R2-14 96

Abbreviations are used without prior definition which can make the paper difficult to follow e.g. FNs in Abstract (line 8), YOLO (line 90). 97
98

■ We have fixed these.

R2-15 99

There are some concerns around the use of jargon making the paper hard to follow. Authors should consider defining specific terms at their first use to support ease of reading. This is 100
101

102 *especially crucial as understanding these terms is critical to understanding the premise of the*
103 *paper. Some examples that require defining include “bounding boxes”, “YOLO”, “performing*
104 *ensemble”.*

We have gone through the paper and have added clarifications when terms like Bounding boxes are used first. For example we have added “Bounding boxes are tight rectangles around detected objects” and that “Ensemble models combine multiple models to make a final recommendation or decision.”

Second review report: WISE: An Adaptive YOLO Ensemble for Accurate E-Waste Object Detection

Decision: Accepted with minor revisions

The authors have done an excellent job addressing the reviewers' comments, significantly improving the clarity, methodological transparency, and overall rigor of the manuscript. The addition of the limitations section, statistical significance testing, and improved contextualization of YOLO variants strengthens the paper substantially. The revisions demonstrate a thoughtful and thorough response to feedback.

Review Details:

While the paper is now in a strong state, the following points could further enhance its impact and clarity:

1. **EMC Metric Justification:**

The EMC metric is a cornerstone of this work, but the cost assumptions. For instance, the environmental cost of a false negative (11.71 USD/item) is derived from an average weight and a single study. A brief sensitivity analysis or a discussion of how EMC might vary with different cost assumptions (e.g., by region or waste type) would strengthen the metric's robustness and generalizability.

2. **Dataset Diversity and Generalizability:**

The dataset, while well-curated, is limited to four types of e-waste (mobile phones, microwaves, keyboards, mice) and a set of common non-e-waste items. The paper would benefit from a clearer acknowledgment of how this composition might affect model performance in real-world settings where e-waste is more diverse (e.g., circuit boards, cables, monitors). A brief discussion on potential domain shift and how WISE might perform on more heterogeneous e-waste streams would be valuable.

3. **Clarity in Writing and Flow:**

While the language has improved, some sections remain dense and could be broken down for better readability. For example, the "Object Detection" and "Models" subsections, though now part of the introduction, still read like a technical manual in places. Consider using more topic sentences and summaries to guide the reader through the technical details.

4. **Figure and Table Accessibility:**

The figures and tables are informative, but some captions could be more self-contained. For instance, Figure 5 (YOLOv8 architecture) is detailed but may be challenging for readers unfamiliar with neural network diagrams. A high-level summary in the caption would make it more accessible.

5. **Broader Impact and Ethical Considerations:**

The paper touches on ethical issues related to e-waste export but could expand on how automated systems like WISE might mitigate or exacerbate these issues. For example, could such systems reduce reliance on manual labor in disadvantaged regions, and if so, what are the socioeconomic implications?

WISE: An Adaptive YOLO Ensemble for Accurate E-Waste Object Detection

Reviewer: [Redacted by Managing Editor]

Second review October 8, 2025

Authors have done a good job in revising the paper in line with the feedback they received. The paper has improved, however some further comments need to be addressed before the paper is suitable for publication:

- The Abstract is currently unstructured and difficult to follow. Authors should consider that a reader would be reading the Abstract first, and therefore it needs to be written in a standalone way where the reader can follow without requiring any further information from the paper.
- The sentence “Manual sorting is unable to keep up” in the Abstract (line 4) is quite sudden to raise a critical component of this paper. More information regarding sorting through E-waste to determine what is safe/not is required, before launching into the fact that current methods are insufficient.
- Similarly, the concept of false negatives and positives (lines 6-7) is raised without any definition or background. Further the abbreviation for false negatives is used at the second mention, rather than the first of false negatives.
- “This paper has three contributions” (line 9) - Contributions to what? Do authors mean to literature in E-waste?
- The Abstract does not state what the aim of the paper was, before it gets into the results/contributions. This is required as it is unclear what the significance of a metric, or the algorithms, or developed ensemble have to the research topic, without the aims being stated.
- “In 2022, we generated 62 million metric tons of E-waste” (line 30) – who does we refer to? Is this globally or a specific country?
- The now included information on the currently available E-waste sorting methods is such an improvement to the paper.
- Kernels in line 125 should be defined so the reader knows what a kernel is.
- Figure 4’s legend does not explain what a and b refer to.
- Figure 5’s legend does not explain the abbreviations used in the figure.
- Table 1 is a great addition to the paper.
- The concept of F1 scores (line 235) needs to be explained.

- “Very high false negatives (FNs) will essentially cause too many toxic substances to enter landfills and find their ways into living organisms”. (lines 336-337). This sentence does not belong in the Results and is more of a Discussion point. Authors need to be cautious about only stating what was found in the Results and any interpretations should be left for the Discussion. Please read through the Results again to ensure no Discussion is included.
- Use and purpose of McNemar’s test should be described in the Methods, with only the actual findings shown in the Results.
- The Conclusion is very long. Some of it eg the discussions comparing this study’s findings to literature, actually belong in the Discussion. The Conclusion should be reserved for a concise wrap-up of the findings and what it means for the research area.
- Several abbreviations are used only once eg RPN, NMS, and therefore do not need to be abbreviated at all.
- It is unclear what the footnotes relate to ie Extracted from author, year. What do these refer to and why are they not included as standard references?
- Authors might want to decide whether they choose to not use contractions throughout the paper such as “don’t” and “it’s” to improve professionalism.
- Grammatically, there is a change in tense throughout, which can make the paper difficult to follow. It will likely be more effective if authors stick to using past tense when describing their own methods and findings to differentiate their work from what is already known in this area.

Final Recommendation:

- Accept with minor revisions