A deep transfer learning based approach to detect COVID-19 waste

Ayushman Buragohain | Bhabesh Mali | Santanu Saha | Pranav Kumar Singh

Department of CSE, Central Institute of Technology Kokrajhar, Kokrajhar, Assam, India

Correspondence
Pranav Kumar Singh, Department of CSE, Central Institute of Technology Kokrajhar, Kokrajhar, Assam 783370, India.
Email: snghpranav@gmail.com

COVID-19 or Novel Coronavirus disease is not only creating a pandemic but also created another kind of problem, initiating a group of wastes, which is also called as COVID-19 waste. COVID-19 waste includes the mask, hand gloves, sanitizer bottles, Personal Protective Equipment (PPE) kits, syringes used to vaccinate people, etc. These wastes are now polluting every continent and ocean. Improper disposal of such wastes may increase the rate of spread of contamination. In this regard, we decided to build up a detection model, which will be able to detect some of the COVID-19 wastes. We considered masks, hand gloves, and syringes as the initial wastes to get detected. We collected the dataset manually, annotated the images with these three classes, then trained different CNN models to compare the accuracies of the models for our dataset. We got the best model to be EfficientDet D0, which gives a mean average precision of 0.82. Further, we have also developed a UI to deploy the model, where general users can upload the images and can detect the wastes, controlling the threshold.

KEYWORDS
applied AI learning, CNN, COVID-19, deep transfer learning

1 INTRODUCTION

Coronavirus disease 19 (COVID-19) is a highly transmissible and pathogenic viral infection caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). It has caused a global pandemic and has led to a huge number of loss of lives worldwide.

Following the pandemic, the World Health Organization (WHO) and various institutions of the world have asked the governments of all the countries of the world to request their citizens to maintain proper COVID protocols, which include wearing a mask, sanitizing their hands, etc., to stop the spread of this virus. As a result, many people acquired this new lifestyle of wearing masks and gloves, which eventually lead to the increase of wastes of the single-use mask, gloves, etc., everywhere. It is seen that the people are throwing single-use masks as well as gloves everywhere and in every corner after they are used. This can potentially increase the risk of the spread of the virus at a faster rate as the thrown mask and gloves may be contaminated with the virus. Therefore, the wastes have to be identified and disposed off as early as possible in order to slow the spread.

Motivated from all these, we decided to develop a model which could detect some of these COVID-19 wastes. We initially decided to detect masks, gloves and also syringes that are used for vaccination, using the help of Convolution Neural Network (CNN) with the transfer learning approach. To the best of our knowledge, this is the first study where transfer learning has been used to detect COVID-19 wastes.
We have collected our datasets and annotated them manually into these three classes. Then we trained and tested our dataset with different CNN architectures. We found that with our dataset, we got the highest Mean Average Precision (mAP) for EfficientDet D0\(^1\) model. The model provided us with a mAP@IOU (Intersection over Union) = 0.5 of 0.82.

The rest of the work is organized as follows: Section 2 discusses the related works. Section 3 presents the proposed methodologies that give the details related to the collection of datasets and the models we have used. Section 4 presents the experimental setup details and also discusses about the evaluation metric values and about the User Interface (UI) and finally, the paper concludes in Section 5.

2 | RELATED WORKS

In study\(^2\) the authors proposed a model by modifying the Densenet121 architecture, which was slower in predicting. The authors named the modified version as RecycleNet where they have altered the connection patterns of the skip connections inside dense blocks. They have used the TrashNet dataset. For the transfer learning approach, they have used the pre-trained weights of ImageNet. Though they got the best accuracy on DenseNet121, but the model was slower, and so they modified it to RecycleNet. The accuracy they got with RecycleNet is 81%. In\(^3\) the authors developed a garbage classification model using Transfer learning. Their work used the weights of the pre-trained dataset “ImageNet” along with the Inception-V3 model for training. They have divided the waste into four categories, that is recyclable waste, hazardous waste, household waste and residual waste. They obtained a training accuracy of 99.3% and test accuracy of 93.2%

In study\(^4\) the authors proposed a model which can detect waste objects in the water. The authors named the model as AquaVision, which is based on RetinaNet architecture. They have used the datasets such as TACO (Trash Annotations in Context), TrashNet for transfer learning comparisons. In order to test the model, they manually created a dataset, which they named as AquaTrash dataset. The best mAP they obtained on the AquaTrash dataset is 0.8148. In\(^5\) the authors compared various models on a dataset collected by them manually to detect garbage. The authors have used five different models, that is, Efficient Det-D1, Single-Shot Object Detector (SSD) ResNet \(-\) 50, Faster Region Based Convolutional Neural Networks (RCNN), Center ResNet101 V1, and YOLO (You Only Look Once) v5M for comparison to select the best model. They have used COCO (Common Objects in Context) dataset for the transfer learning approach. The authors obtained the best model as YOLO v5M with mAP of 61.3%.

We have also found several other works related to waste object detection using various models of CNN. But we observed that this is the first time someone has used the transfer learning approach to detect COVID-19 waste.

3 | PROPOSED METHODOLOGY

3.1 | Dataset collection

The dataset was collected manually from various pages on the internet as well as taken manually through a mobile phone device. The device name is RealMe 3 Pro. The dataset contains images of three COVID-19 waste products. They are masks, gloves, and syringes used to vaccinate people. The total number of images in the dataset is 656 images, of which 253 images were of mask, 211 glove images, and 192 syringe images. While collecting the data, we took care of adding any duplicate images.

3.2 | Convolutional neural network models

Convolutional Neural Network is a huge section containing different model architectures for various classification and detection process. Modern CNN models can generally contain two main components: the backbone and the head. The backbone refers to the network which takes input as the image and extracts the feature map upon which the rest of the network is based.

The head generally refers to the part of the network that generates the required predictions. In our task, the head will generate bounding box coordinates as well as class predictions for the class present in the bounding box prediction generated. For our dataset, we have compared three different architectures of the CNN models, which are discussed below.
3.2.1 | RetinaNet

RetinaNet\(^6\) is a one-stage object detection model that is developed by Facebook AI Research (FAIR) that utilizes a focal loss function to address class imbalance during training. We used a pre-trained RetinaNet\(^6\) with ResNet50\(^7\) backbone in our tasks and changed the model head to detect three classes for our required task. RetinaNet also consisted of a Feature Pyramid Network (FPN).\(^8\) FPN is used on top of the backbone, which in our case is a ResNet50 network. FPN\(^8\) is used for constructing a rich multi-scale feature pyramid from one single resolution image. We also trained with the default anchor boxes, and focal loss\(^6\) parameters that were used to train on the COCO dataset for RetinaNet.\(^6\)

3.2.2 | Faster RCNN

Faster RCNN\(^9\) is an object detection architecture presented by Girshick et al. in 2015. The most widely used state-of-the-art version of the R-CNN family — Faster R-CNN\(^9\) was first published in 2015. The RCNN networks are mainly made up of the following components: (a) A region proposal algorithm to generate probable bounding box locations over the images. (b) A backbone network to generate features from an image. (c) A classification network to predict the class of object. (d) A regression network to make the bounding box locations more precise.

The Faster RCNN\(^9\) model used a Region Proposal Network (RPN) to generate a probable bounding box. We used a Faster RCNN pre-trained on COCO dataset. For our case, the backbone was a ResNet50\(^7\) we modified to classification network to generate predictions for three classes.

3.2.3 | EfficientDet

EfficientDet\(^1\) is a CNN architecture used for detection of objects where the components are a BiFPN\(^1\) feature network and an EfficientNet backbone. The BiFPN (Bi-Directional Feature Pyramid Network) network takes in features from an EfficientNet backbone and repeatedly applies top-down and bottom-up directional feature fusion. These fused features are then fed to a class and box network to produce class and bounding box predictions, respectively.

In our case, we have used an EfficientDet D0 model, pre-trained on the COCO dataset. This model uses an EfficientNet B0 model as the backbone to generate features. We modified the classification network to generate three class predictions, and the remaining network was kept the same.

3.3 | System architecture

The proposed annotated dataset is trained with various CNN models, for comparison, in order to get the best model. Firstly, the annotated dataset is split into training and testing data. The training data is used to train the models for computing the final weights. The testing data along with the final weights is then used to compute Mean Average Precision (mAP) for all the models. Finally, based on the mAP metric, the model that obtained the best mAP value is selected for detecting the waste objects. Figure 1 shows the system architecture of the proposed model.
EXPERIMENTAL EVALUATION

4.1 Dataset preprocessing

After collecting the datasets, we annotated the waste objects in the images manually. We basically annotated three classes of objects in the images. They were mainly masks, gloves, and syringes. We annotated the objects in the images with the help of a free software library known as LabelImg.

4.2 Transfer learning

Transfer learning consists of taking features learned on one problem and leveraging them on a new, similar problem. Transfer Learning significantly reduces the training time and also cuts down resource usage. Transfer learning is usually done for tasks where your dataset has too little data to train a full-scale model from scratch.

The general idea is to use the knowledge a model has learned from a task with a lot of available labeled training data in a new task that does not have much data. Instead of starting the learning process from scratch, we start with patterns learned from solving a related task. Including the pre-trained model weights also leads to faster training and lower generalization error.

CNN architectures usually try to detect edges in the earlier layers, shapes in the middle layer and some task-specific features in the later layers. In transfer learning, the early and middle layers are used and we only retrain the latter layers. It helps leverage the labeled data of the task it was initially trained on. This is to say that we keep the backbone of the model unchanged and make only changes in the head of the model.

For our task we used models which were pretrained on the Common Objects in Context (COCO) dataset. The weights of these models were then fine tuned on our dataset.

4.3 Training

After annotating the waste in the images of the complete dataset, we split the dataset into training and testing sets via stratified random split wherein training images pertained to 80% of the data and the remaining 20% was our test set. The images for both training and testing were then resized with a pad keeping the original height, width ratio constant such that the largest size will be 900 pixels and then randomly cropped to $512 \times 512$ pixels for the training dataset and resized to $512 \times 512$ pixels for the validation dataset. In random crop we take a randomly selected patch from the original image and use it for training the models instead of the original image. Random crop has the effect of creating new images from existing images, which helps our model from overfitting. Figure 2 shows how random cropping takes place in an image. Then we trained all the models with the training dataset keeping the batch size for all the models as 8.

Data augmentation was also applied during the training of the model. Images were randomly horizontally flipped, brightness and contrast is reduced or increased by a factor of 0.3, and randomly rotated by a factor of 45. Data augmentation was used to make our model more robust and less susceptible to overfit on the training dataset.

![Figure 2](image-url)

**Figure 2** Illustration of random cropping in an image
validation, data augmentation was not active. While training out the models, the learning rate was kept as 0.0001 for all the models.

### 4.4 Evaluation metrics

Our models were evaluated on the basis of mAP@IOU = 0.5. We have used three different architectures to detect wastes from the images. The models include RetinaNet, EfficientDet D0, and Faster RCNN. The mAP@IOU = 0.5 of all the models was calculated. The mAP score or Mean Average Precision is an evaluation metric that is calculated by taking the mean AP (Average Precision) over all classes or/and overall IoU thresholds, depending on different detection challenges. AP or Average Precision is defined as finding the area under the precision-recall curve. The results of all the models are shown in Table 1. The baseline RetinaNet6 model is trained on a small subset of the whole training dataset. The results are compared, and we found the best model to be as EfficientDet D0 with mAP to be as 0.82 for 40 epochs. The mAP for each epoch of the Efficient Det D0 model is shown in Figure 3A. The loss of both training and validation data while training the EfficientDet D0 model is shown in Figure 3B.

### 4.5 Detection of waste images

After training the models, we calculated the mAP’s of all the models. Based upon mAP value, we got the best model as EfficientDet D0. We have also designed a UI (Figure 4C) to deploy our model to make the process of detection easier. The UI will take the images, and the user will have to adjust the threshold parameter manually to detect the waste objects in the images correctly. The main objective of the user interface is to showcase to the user about the model's detection capability of waste objects in images and also to give information about the threshold change, which will help to detect the objects more fruitfully. Further the user, upon desired results, can deploy the model in a smart garbage collector robot to detect the wastes and collect it. The detection of COVID-19 waste objects by the EfficientDet D0 model is illustrated in Figure 4A and B.

### 5 CONCLUSION

In this work, we proposed a model that helps to detect some of the COVID-19 wastes that are inappropriately disposed off. We have used our own dataset and compared it with three different architectures of CNN using transfer learning approach. We got the best model for our dataset to be as EfficientDet D0 model, which gave us the mAP of 0.82.

| Models            | Backbone       | Epochs | Average Precision at IOU 0.5 |
|-------------------|----------------|--------|-------------------------------|
| RetinaNet(Baseline)6 | Resnet50- FPN7  | 15     | 0.682                         |
| RetinaNet6        | Resnet50- FPN7  | 20     | 0.719                         |
| EfficientDet1     | EfficientNet B0-BiFPN1 | 40     | 0.82                          |
| Faster RCNN9      | Resnet50- RPN9  | 20     | 0.739                         |

**TABLE 1** Summarized information and mAP scores of proposed deep transfer learning techniques

**FIGURE 3** A, mAP upon epoch of EfficientDet D0 model B, Training and validation loss of EfficientDet D0 model
The model can be deployed in a smart garbage collector robot with arms, where initially the robot will click the pictures, and the model will then process the images and detect the waste objects. The detected waste objects with images can also be further used for improving the overall accuracy of the model. Further, the arms can collect the waste materials for safe dumping. It will be helpful to reduce human intervention and will help in preventing infection and also pollution. This can be one of the used cases under Information and Communication Technology (ICT) for Societal challenges. One example of such a similar work is discussed in.10 In the future, we will try to detect other COVID-19 wastes by collecting the appropriate datasets.

PEER REVIEW
The peer review history for this article is available at https://publons.com/publon/10.1002/itl2.327.

DATA AVAILABILITY STATEMENT
Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

ORCID
Ayushman Buragohain  https://orcid.org/0000-0002-1657-3431
Pranav Kumar Singh  https://orcid.org/0000-0001-8987-0229

REFERENCES
1. Tan M, Pang R, Le QV. Efficientdet: Scalable and efficient object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 2020:10781–10790.
2. Bircanoglu C, Atay M, Beşer F, Genç Ö, Kızrak MA. RecycleNet: intelligent waste sorting using deep neural networks. IEEE; 2018:1–7.
3. Cao L, Xiang W. Application of convolutional neural network based on transfer learning for garbage classification. IEEE; 2020:1032–1036.
4. Panwar H, Gupta P, Siddiqui MK, et al. AquaVision: automating the detection of waste in water bodies using deep transfer learning. *Case Stud Chem Environ Eng*. 2020;100026.
5. Patel D, Patel F, Patel S, Patel N, Shah D, Patel V. Garbage detection using advanced object detection techniques. IEEE; 2021:526–531.
6. Lin TY, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. IEEE; 2017:2980–2988.
7. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016:770–778.
8. Lin TY, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. 2017:2117–2125.
9. Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell*. 2016;39(6):1137-1149.
10. Bai J, Lian S, Liu Z, Wang K, Liu D. Deep learning based robot for automatically picking up garbage on the grass. *IEEE Trans Consumer Electron*. 2018;64(3):382-389.

How to cite this article: Buragohain A, Mali B, Saha S, Singh PK. A deep transfer learning based approach to detect COVID-19 waste. *Internet Technology Letters*. 2022;5(3):e327. doi: 10.1002/itl2.327