FRAMEWORK for Roadside Litter Identification and Face Recognition using Convolutional Neural Networks

Abiga Sansuri.M
Department of ECE
PSG College of Technology
Coimbatore, TN, India
18l101@psgtech.ac.in

Dr.Anusuya K.V
Department of ECE
PSG College of Technology
Coimbatore, TN, India
kva.ece@psgtech.ac.in

Nivethitha.M
Department of ECE
PSG College of Technology
Coimbatore, TN, India
18l134@psgtech.ac.in

Gutta Abhinav Krishna Sai
Department of ECE
PSG College of Technology
Coimbatore, TN, India
18l115@psgtech.ac.in

Abstract— Conventional street cleaning methods involve lots of human resources. Using face detection and image processing techniques, the identification of people who litter on the roadside is made possible. By fixing cameras in crowded places such as bus stands and markets, the proposed system utilizes a deep neural network algorithm called Convolutional Neural Network (CNN), especially Mask Region Convolutional Neural Network (Mask R-CNN). The proposed schema makes use of this deep learning algorithm to analyze the photographs of the streets and detects the litter (if any) in them. Further, in the presence of litter, the schema identifies the person throwing the litter. In-built library files of Python are involved in this process for generating and comparing the face encodings. Reducing operational costs and assisting with better monitoring of the streets can improve the system efficiency tremendously.

Keywords— Cleanliness, Face Recognition, litter detection, MASK - RCNN, Neural Networks

I. INTRODUCTION

Littering is a common occurrence all around the world. Some may even refer to it as chronic. People cannot help but litter everywhere, be it on roads, railway tracks, floors, water bodies, etc. Swachh Bharat Abhiyan – A Government of India mission to make people more responsible towards their surroundings and motivating them to take charge of the cleanliness around them. However, the closest thing to punishment enforced by the government is imposing a fine of rupees five hundred to whoever litters of one's free will according to Section 278 of The Indian Penal Code. Currently, the process of litter detection on streets is not automated and it demands human intervention at almost every level which makes it even more time-consuming. Most of the tasks involved in street cleaning i.e., litter detection and classification can be automated using deep learning techniques. In this paper, an automated workflow is proposed to achieve these tasks. Our proposed framework aims at improving the environment of a city by using object detection and face recognition techniques. By doing so strict laws and measures can be enforced effectively, to make the citizens more responsible towards their environment.

II. LITERATURE SURVEY

In paper [1], the mobile processing component of the proposed framework is considered. The main goal of this component is to process the captured pictures and determine if they are worthy for deeper processing. Further object detection and classification are performed on the captured images. Performing object detection using CNN is one of the alternatives. Optimized variants of CNN such as Region Convolutional Neural Network (R-CNN) [2], [3] have been produced due to the advancements in the field of deep learning. However, the training model which uses R-CNN is expensive in terms of space and time and performs slow object detection [4]. A more enhanced implementation is Fast R-CNN [4], which has significant performance improvements. It delivers at near real-time rates for object recognition using deep neural networks with multiple layers of convolutional and max-pooling layers. Selective search is used to generate predictions and it involves a great deal of time. To overcome this, Faster R-CNN replaces the selective search method with a region proposal network which makes the algorithm much faster [5]. However, pixel-to-pixel alignment between network inputs and outputs cannot be performed by Faster RCNN. To execute coarse spatial quantization for feature extraction, it utilizes the Region of Interest (RoI) Pool. An easy, quantization-free layer known as RoI Align, conserves the exact spatial locations. It is used in Mask R-CNN [6]. Mask R-CNN is theoretically simple. A class label and a bounding-box offset are the results of Faster RCNN. In addition to this, a third branch that outputs the object mask (a binary mask that indicates the pixels at where the object is in the bounding box) is added. The extraction of the spatial layout of an object is carried out using Mask RCNN.

III. Proposed Methodology

With the aid of the above-mentioned literature, the proposed schema is designed and depicted in Fig. 1. The proposed schema automates the task of litter detection by using deep learning algorithms applied on images of streets captured through a camera. A pi camera captures the images at every 20s and feeds them into the processing component that has the primary goal to process the captured pictures and determine their worth for in-depth processing. This result is fed into the Object Detection Block (ODB) that uses the Convolutional Neural Network (CNN), especially the Mask RCNN. The ODB is responsible to differentiate the litter and non-litter objects successfully.

2000 region proposals for the input image are achieved with the help of an individual network with a selective search algorithm [10]. RoI Align layer which is not quantized is utilized to reshape the predicted region proposals. Each region proposal image is deformed into a fixed size of 224x224. These region proposal images are then passed to the trained CNN to obtain a 4096-dimensional feature vector against all the 2000 region proposals, resulting in a 2000x4096 dimensional matrix.

Each region proposal is classified using SVM (Support Vector Machines) [7] for each class. The SVM weights(4096-dimensional) are built into a matrix and it is multiplied with the feature matrix for all the N classes. This results in a matrix that assigns a score to each class to which a region proposal belongs. The class with the maximum score isprovided with a request. Hence, all the 2000 region proposals or bounding boxes in the image are provided with a class label. Out of those many bounding boxes, many of them would be redundant. Hence, the overlapping bounding boxes need to be removed and are accomplished using a Non- maximum suppression algorithm [9].

Using the scores, the system distinguishes between litter and non-litter. For litter, the control is transferred to the face detection block. The inbuilt functions of Python generate and compare the face encodings [12]. The algorithm considers the specific essential measurements of the face such as the color, the size, the eyes slant and the gap between eyebrows, etc. The combined features define the face encoding and are used to identify the specific image (face). Further, the images are compressed and stored in the cloud. The final output displays an image with the bounding box over the trash and the person who threw litter along with labels.

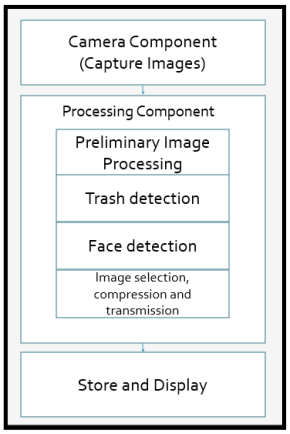

Fig. 1. The proposed system

IV. Implementation

The proposed algorithm is implemented using Python script language. The processing component is made up of several program files and is shown in Fig. 2. The function of each program is narrated below.

- Main Program that consists of trash detection and face detection instructions.
- Mask RCNN folder that consists of python files of the base configuration class and the main mask RCNN model implementation. It also contains common utility functions and classes, display, and visualize functions.
- The training set provided in the trash folder includes all the annotations of JSON files. VGG image annotator [11] was utilized for image annotation.
- A separate folder is made use to record all the images captured by the raspberry pi camera module.
- To train the network, weights are exploited.
- A set of known images are stored in a distinct folder.

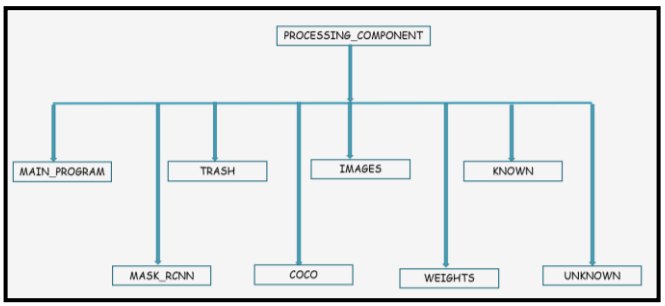

Fig. 2. Processing Component

A. Main Program

As shown in Fig. 3, all the necessary library files such as Pycocotools, Kaggle, Mrcnn, TensorFlow (1.14.0), Keras, NumPy, Scipy, (2.2.0), Face_recognition, Pillow, TensorFlow, Skimage, h5py, Ipython, Cv2, Matplotlib, and Python are initially installed.

B. Trash detection and Face recognition

All the required weights along with the configuration functions of MASK-RCNN are imported. The number of GPU to work on the processing of the image is set, in addition to the number of the image processed per GPU. The configuration details of the MASK_ RCNN are displayed. The device is set to load the neural network on. The model is inspected in training or inferences mode. The validation dataset is loaded and a model is created in inference mode. On creating, all the trained weights are loaded. Further, the images from the directory (wherein the images are stored) are obtained once it is captured by Raspberry Pi every 20 seconds. Object detection on the images is performed and the results are displayed. If the scores of the result are greater than the threshold, then the images processed by the trash detection block are stored in a folder called unknown.

With the help of functions in face recognition, the encodings from the face present in the image are generated. Similarly, encodings for the set of known images are generated. Comparison between the encodings is carried out until a match is found. Eventually, the output is shown with labels and scores. The flowchart illustrating the flow of the face detection process is shown in Fig. 4. The hardware diagram is demonstrated in Fig. 5. The connections to the Raspberry Pi model [8] are depicted in Fig. 6.

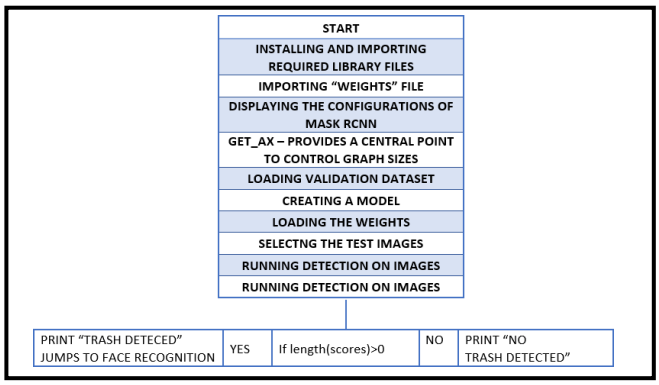

Fig. 3. The Flow of the Main Program

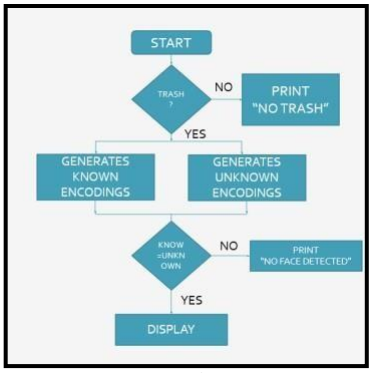

Fig. 4. Flowchart of Face Detection

Fig. 5. The Raspberry pi 4 with camera module

Fig. 6. Connections in raspberry pi 4

V. Results Analysis

The results generated by the system are shown in Fig. 7, 8, 9, 10, 11, and 12. The dimensions of the image to be processed and the dimensions of the image after pre-processing are displayed in the results. In addition to the scores, the message (TRASH DETECTED or NO TRASH DETECTED) is shown as an output. There are four possible illustrations: With trash in the presence of the person, with no trash but in the presence of a person, with trash along with no person, with no trash along with no person. Here, Fig 7,9, 11 represent the outputs with trash in the presence of a face. A bounding box around the face of the person who threw the trash besides their name is also displayed. In addition to this, the processed image is visible on the screen comprising of the bounding box, labels, and binary mask over the trash along with the labels. In case of no trash with no face, the image captured is displayed as shown in Fig. 8 and 10. Additionally, Fig. 12 exhibits a result with no trash but in the presence of a person.

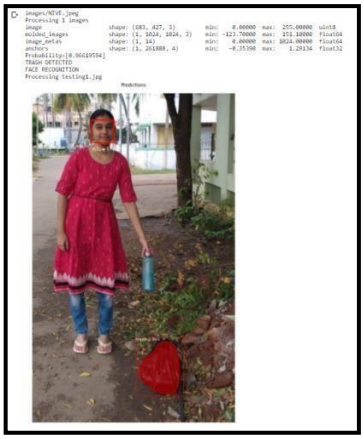

Fig. 7. Output with trash

Fig. 8. Output with no trash

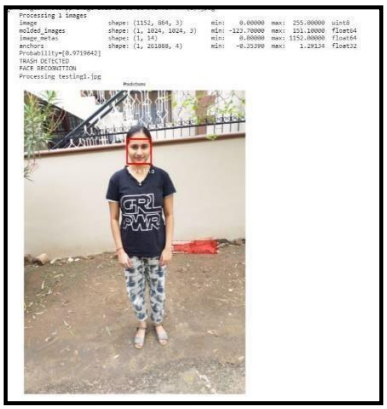

Fig. 9. Output with trash and face detected

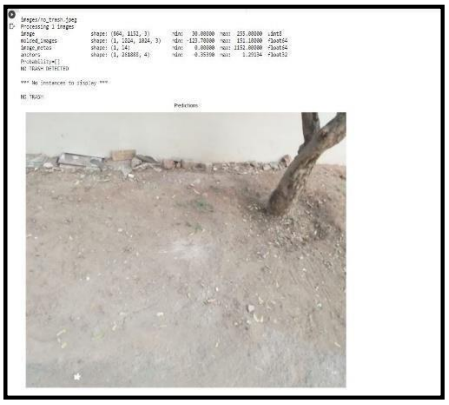

Fig. 10. Output with no trash

Fig. 11. Output with trash and face detected

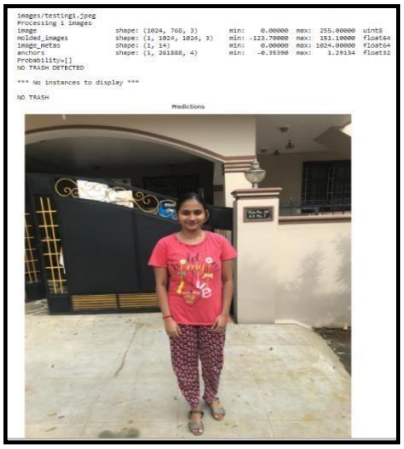

Fig. 12. Output with no trash

VI. Conclusion

Cleanliness plays a vital role at all stages of life. If strict rules are enforced, it will be helpful to create a clean city without taking a toll on our health. With the help of cameras affixed in public places like bus stands, markets, etc. along with this system will assist us to identify violators in optimum time with minimum use of manual resources. This system can be further extended by sending an alert message to the individual who litters and penalizes them monetarily. Also, an e-challan or memo could be issued.

References

[1] Chandni Ramchandani, Rakshith Koravadi Hatwar, ParteekMakkar, Yanki Shah, Pooja Yelure, MagdaliniEirinaki, “A deep learning framework for smart street cleaning” in *2017 IEEE Third International Conference on Big Data Computing Service and Applications*. [Online]. Available: https://ieeexplore.ieee.org/document/7944927/

[2] A. Alzu’bi, A. Amira, and N. Ramzan, “Compact root bilinear CNN for content-based image retrieval,” Aug 2016, pp. 41–45.

[3] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” in Advances in neural information processing systems, 2015, pp. 91–99.

[4] R. Girshick, “Fast R-CNN,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1440–1448.

[5] Lokanath M, Sai Kumar K, and Sanath Keerthi E, “ Accurate object classification and detection by faster RCNN,” in *IOP Conference Series Materials Science and Engineering, November 2017*. [Online]. Available: https://www.researchgate.net/publication/321478922_Accurate_object_classification_and_detection_by_faster-RCNN

[6] Kaiming He, Georgia Gkioxari, Piotr Dollár, Ross Girshick, “Mask RCNN” arXiv:1703.06870,2017.[Online].Available:https://arxiv.org/abs/1703.06870

[7] “Support Vector Machines”. [Online]. Available: https://scikit-learn.org/stable/modules/svm.html

[8] “Raspberry Pi 4 Model B Datasheet, “June 2019. [Online]. Available: https://www.raspberrypi.org/documentation/hardware/raspberrypi/bcm2711/rpi_DATA_2711_1p0_preliminary.pdf

[9] Jan Hosang, Rodrigo Benenson, Bernt Schiele “Learning Non-Maximum Suppression Algorithm,” arXiv:1705.02950 May 2017. [Online]. Available: https://arxiv.org/abs/1705.02950

[10] “Selective search for Object Detection,” September 18, 2017. [Online]. Available:https://www.learnopencv.com/selective-search-for-object-detection-cpp-python/

[11] “Getting Started with VGG Image Annotator for Object Detection Tutorial,” September 25, 2020. [Online]. Available: https://blog.roboflow.com/vgg-image-annotator/

[12]” Recognizing Face using the “face_recognition” library,” March 9, 2020. [Online]. Availabe:https://medium.com/analytics-vidhya/recognising-face-using-the-face-recognition-library-afdb6d86bcf0