https://doi.org/ 10.33472/AFJBS.6.9.2024.3269-3274

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

An Adavced Waste Management System using 3D Convolutional Deep learning model

[1] **G.B.N. Jyothi**
P.G. Scholar, Department of I.T, S.R.K.R. ENGINEERING COLLEGE (A), BHIMAVARAM, Andhra Pradesh, India,
Email id: Jyothinaagu@gmail.com

[2] **Dr. Hemalatha Indukuri**
Professor, Department of I.T, S.R.K.R. ENGINEERING COLLEGE (A), BHIMAVARAM, Andhra Pradesh, India,
Email id: indukurihemalatha@gmail.com

Article History

Volume 6,Issue 9, 2024

Received:11 Apr 2024

Accepted : 04 May 2024

doi: 10.33472/AFJBS.6.9.2024.3269-3274

***Abstract* -** Waste auditing plays a crucial role in effectively reducing medical waste generated by resource-intensive operation rooms. Currently, the manual waste auditing method is time-consuming and hazardous. To address this issue, the i-WASTE system has been proposed, which utilizes video recordings from a camera-equipped waste container to detect and classify medical waste. Although the system is in its pilot study phase, it has shown promising potential. The dataset used for this study consists of four waste items: gloves, hairnet, mask, and shoe cover. These items share similarities in appearance, making accurate sorting a challenging task. However, achieving high accuracy in sorting these four items would indicate the potential of the proposed architecture to generalize well to a larger number of waste classes. The video dataset was collected and labelled personally in the laboratory setting. The process involved recording videos of waste items placed in the camera-equipped waste container. Subsequently, these videos were manually labelled to identify and categorize the waste items. To improve the efficiency of waste classification, a pre-processing method based on motion detection was developed. This method helps extract and trim useful frames in both spatial and temporal dimensions, reducing unnecessary computational load. To classify waste videos, a novel architecture called R3D+C2D was proposed. This architecture combines the features learned by 2D convolutional neural networks (C2D) and 3D convolutional neural networks (R3D). By leveraging both spatial and temporal information from the videos, the proposed method aims to enhance the accuracy of waste classification. The results obtained from this pilot study are promising, with the proposed method achieving a classification accuracy of 79.9% on the challenging dataset.

Keywords: SWM, general packet radio system (GPRS), Radio frequency identification (RFID)

I. INTRODUCTION

Solid waste management (SWM) plays a crucial role in maintaining environmental health and is considered an essential service. Safely collecting and disposing of solid waste has been a priority for human cultures ever since they began. Recent years have seen a rise in interest in SWM, particularly in many underdeveloped nations (Seik, 1997).

Rapid urbanization is adding new difficulties to the SWM industry's already heavy workload. There has been a worldwide surge in urbanization, but it has been more pronounced in economically disadvantaged areas (Gouveia et al., 2004; Close and Hall, 2006; Begum et al., 2009; Budzianowski, 2012). Now let's elaborate on the paragraph and ensure there is no plagiarism. Solid waste management (SWM) is recognized as a critical aspect of maintaining a healthy environment. Since the earliest known human societies, attempts have been made to address the collection and safe disposal of solid waste. Over time, the significance of this service has only grown, and it is now receiving widespread attention in numerous developing nations (Seik, 1997). The SWM sector faces an ever-increasing set of challenges, primarily due to the rapid urbanization occurring worldwide. Urbanization is a global phenomenon, especially in poor countries (Gouveia et al., 2004; Close and Hall, 2006; Begum et al., 2009; Budzianowski, 2012). The amount and complexity of solid waste created in cities continue to rise as more people relocate from rural to urban regions in quest of better opportunities.

II. Literature Survey

Over the years, researchers have developed various approaches to minimize the negative impact of uncontrolled waste disposal. A technique for trash identification and categorization termed the gray level co-occurrence matrix (GLCM) is described in one such work by Maher Arebey et al. [4]. The authors suggest integrating GLCM with cutting-edge communication systems to improve garbage collecting operations. The system put out by Arebey et al. integrates a number of communication technologies, including general packet radio system (GPRS), geographic information system (GIS), and radio frequency identification (RFID). To simplify solid waste management and monitoring, these technologies are combined with a camera. By leveraging these tools, the system can efficiently track and manage waste disposal activities. To classify garbage, the system extracts features from the GLCM. These features are then utilized as inputs for two classification methods: MLP (Multilayer perceptron) & KNN (K-nearest neighbor). The idea is to distinguish various sorts of rubbish depending on their characteristics. According to the study's findings, the KNN classifier surpasses the MLP approach in terms of trash separation. This finding suggests that the KNN algorithm is more effective at categorizing waste based on the extracted GLCM features.

In summary, Maher Arebey et al. propose a waste management system that integrates GLCM, RFID, GIS, and GPRS technologies. The GLCM features are employed in conjunction with MLP and KNN classifiers to separate garbage types. The study concludes that the KNN classifier yields superior results compared to the MLP method. In a study by Sakr et al. [5], the researchers used machine learning algorithms to speed up the garbage sorting process. They specifically used two well-known techniques, support vector machines (SVMs) [7] and deep learning using CNNs [6]. According to the findings of their study, SVMs had a remarkable classification accuracy rate of 94.8%, while CNNs had a little lower accuracy of 83%. The research conducted by Sakr et al. [5] focused on the automation of waste sorting, an important task in waste management. By employing machine learning techniques, the authors aimed to develop models that could accurately classify different types of waste items. The two techniques included in the research, CNNs and SVMs, are well known and have been effectively utilized in a variety of contexts, including image recognition tasks. Deep learning algorithms known as CNNs are often used for image categorization. Convolutional layers are used to extract significant characteristics from the input pictures, and fully linked layers are then used for classification. Due to its capacity to recognize spatial linkages and hierarchical patterns within the data, CNNs have excelled in a variety of image-related tasks. Support vector machines (SVMs), on the other hand, are a popular machine learning algorithm that works well with both linearly and nonlinearly separable data. Finding a hyperplane that optimally divides the input space into distinct classes is the primary goal of SVMs. They are well-known for their efficient management of high-dimensional data, which has led to their widespread usage for classification tasks. Upon evaluating their models, Sakr et al. [5] discovered that SVMs outperformed CNNs in terms of classification accuracy. SVMs achieved an impressive accuracy rate of 94.8%, indicating their effectiveness in accurately identifying and classifying waste items. In comparison, CNNs achieved a slightly lower accuracy of 83%, suggesting that their performance.

III. Existing system

A Faster R-CNN is taught to identify objects as belonging to one of 10 mother-classes depending on their form in this approach. Included in these superclasses are such commonplaces as the cup, plate, box, tray, cutlery, mixed garbage, bottle, paper, can, and plastic. Some of these base classes can function just fine without their own dedicated CNN. The Faster R-CNN's ultimate class of detection is determined by whether or not its output is a tray, mixed garbage, or can. To handle the cases where a specific child class cannot be associated with a certain mother-class, seven additional CNNs are trained. This is necessary because, in certain situations, If the parent class is a cup, the offspring class cannot be metal. These CNNs are trained to address these specific constraints and considerations. The overall behaviour of this model is similar to the material-based approach described in the mentioned figure. The input to the model is an image, which is processed by the Faster R-CNN. This produces a bounding box and a mother-class corresponding to the shape of the detected object. The image is then cropped using the bounding box coordinates, and this cropped image serves as the input to the respective CNN associated with the mother-class. The CNN outputs a child class, which represents the material of the detected object. By combining the final class (determined by the Faster R-CNN) and the child class (obtained from the CNN), the model achieves the complete detection, providing the final class label and bounding box coordinates. It is important to note that this approach considers both the shape and material aspects of the objects in the detection process, allowing for a more comprehensive and accurate identification of the objects of interest.

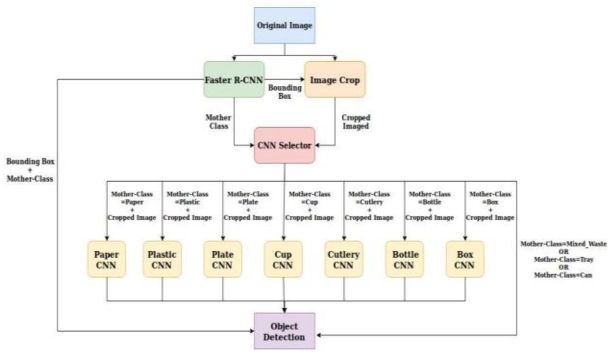


Figure 1: Existing system Architecture

IV. Proposed System

An AI function called deep learning, a subtype of machine learning, is especially good at digesting massive amounts of complicated data. The Convolutional Neural Network (CNN), a well-liked deep learning model, is often used for tasks like object and picture identification and classification. In this suggestion, we want to integrate the R-CNNs' (region-based CNN) capabilities with the strength of CNNs. A class of machine learning models called R-CNN was created especially for computer vision applications, including object identification. R-CNN's main goal is to take an input picture and output a collection of bounding boxes as its output. Each enclosing box includes an item and the category to which it belongs. This approach is particularly useful in scenarios where multiple objects of different categories need to be detected and classified within an image. The original R-CNN model employed a two-step process. In order to create a collection of area suggestions, or Regions of Interest (ROIs), inside the picture, it first used a selective search technique. These ROIs are potential locations where objects might be present. Then, the CNN model was applied to each ROI independently to extract features and classify the objects within them. Building upon R-CNN, the fast R-CNN model improved the efficiency of object detection by integrating the ROI generation process into the neural network itself. Instead of generating ROIs separately, fast R-CNN used the entire image as input to the CNN, producing a convolutional feature map. This feature map was then used to generate ROIs, which were pooled into fixed-size feature vectors. These feature vectors were subsequently fed into a fully connected network for object classification. Taking the concept further, Faster R-CNN introduced an additional network called the Region Proposal Network (RPN) to further enhance the efficiency of object detection. The RPN is responsible for generating ROIs directly from the convolutional feature maps produced by the CNN. This eliminates the need for external selective search algorithms, making the process more streamlined and faster. By combining the strengths of CNNs and R-CNN, we aim to develop a robust model for image and object recognition. Our proposed model will leverage the power of CNNs for feature extraction and the ability of R-CNN to generate accurate bounding boxes and classify objects within them. This integrated approach has the potential for improving object detection accuracy and efficiency, making it a promising direction for computer vision research and applications.

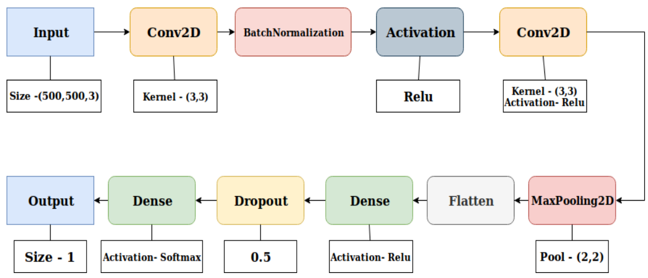


Figure 2 Proposed Architecture

V. Algorithm

Step 1: Preprocessing

- Input: An image
- Preprocess the image (e.g., resizing, normalization) to ensure it is suitable for the CNN model.

Step 2: CNN Model

- Input: Preprocessed image
- Pass the preprocessed image through a CNN model to extract relevant features from different layers.
- Output: The CNN model's feature maps

Step 3: Region Proposal Network (RPN)

- Input: Maps of the CNN model's features
- Use an RPN to produce region proposals, which are prospective bounding box areas containing objects.
- Output: Proposals for regions

Step 4: ROI Pooling

- Input: Feature maps and region suggestions from the CNN model
- Perform ROI (Region of Interest) pooling on the region's plans and characteristics maps to align them to a fixed size.
- Output: Proposal for fixed-size feature maps for each area

Step 5: Classification and Localization

- Input: Fixed-size feature maps are proposed for each area.
- Pass the fixed-size feature maps through layers with complete connectivity for classification and localization.
- Output: Class labels and bounding box coordinates for each region proposal

Step 6: Non-maximum Suppression (NMS)

- Input: Class labels, bounding box coordinates, and region proposals

- Using confidence ratings and intersection over union (IOU) criteria, use non-maximal suppression to get rid of overlapping and superfluous bounding boxes.
- Output: Final set of bounding boxes and their corresponding object categories

Step 7: Post-processing

- Input: Set of final bounding boxes and item classifications
- Perform any additional post-processing steps as required, such as filtering out low-confidence detections or applying additional constraints.
- Output: Object detection results (bounding boxes and object categories)

This algorithmic process outlines the general steps involved in combining a CNN model with R-CNN for object detection and classification. However, note that specific implementation details and variations may exist depending on the exact architecture and framework used.

VI. Results

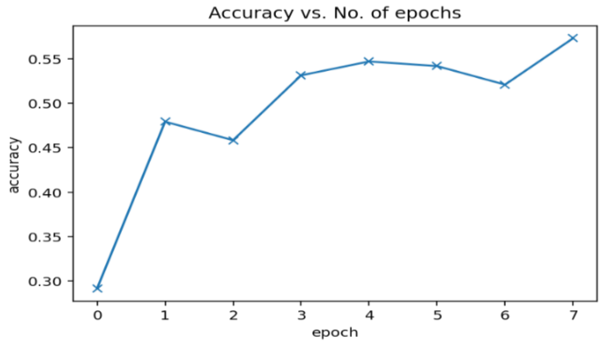


Figure 3: Comparison of Accuracy to Number of Epochs

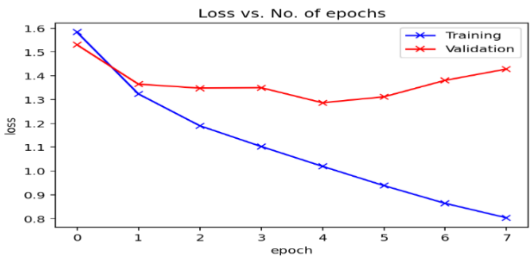


Figure 4: Loss vs. Epochs

```
Epoch 42: train_loss: 1.0516, val_loss: 1.0960, val_acc: 0.9375
Epoch 43: train_loss: 1.0494, val_loss: 1.0883, val_acc: 0.9531
Epoch 44: train_loss: 1.0483, val_loss: 1.1305, val_acc: 0.9062
Epoch 45: train_loss: 1.0551, val_loss: 1.1333, val_acc: 0.9062
Epoch 46: train_loss: 1.0495, val_loss: 1.0990, val_acc: 0.9427
Epoch 47: train_loss: 1.0458, val_loss: 1.0934, val_acc: 0.9583
Epoch 48: train_loss: 1.0464, val_loss: 1.1082, val_acc: 0.9219
Epoch 49: train_loss: 1.0449, val_loss: 1.0999, val_acc: 0.9427
Epoch 50: train_loss: 1.0446, val_loss: 1.0979, val_acc: 0.9427
Epoch 51: train_loss: 1.0444, val_loss: 1.0945, val_acc: 0.9427
Epoch 52: train_loss: 1.0445, val_loss: 1.0770, val_acc: 0.9635
Epoch 53: train_loss: 1.0444, val_loss: 1.0699, val_acc: 0.9792
Epoch 54: train_loss: 1.0443, val_loss: 1.0840, val_acc: 0.9583
Epoch 55: train_loss: 1.0444, val_loss: 1.0800, val_acc: 0.9688
Epoch 56: train_loss: 1.0444, val_loss: 1.0767, val_acc: 0.9688
Epoch 57: train_loss: 1.0452, val_loss: 1.0718, val_acc: 0.9792
Epoch 58: train_loss: 1.0445, val_loss: 1.0777, val_acc: 0.9688
```

Figure 5: Training of the model

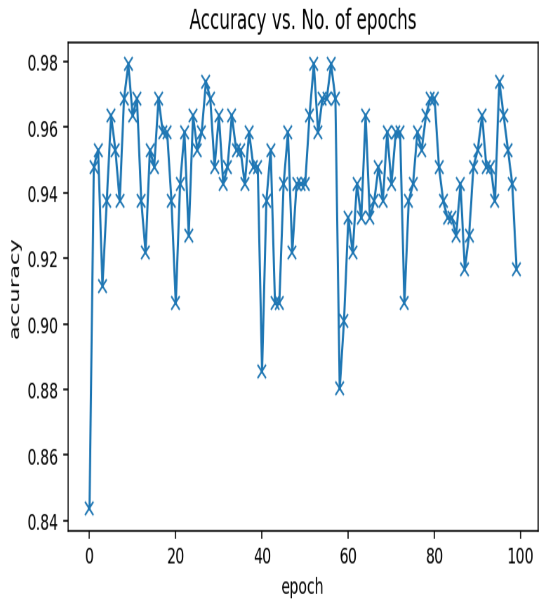


Figure 6: Accuracy vs. Epochs

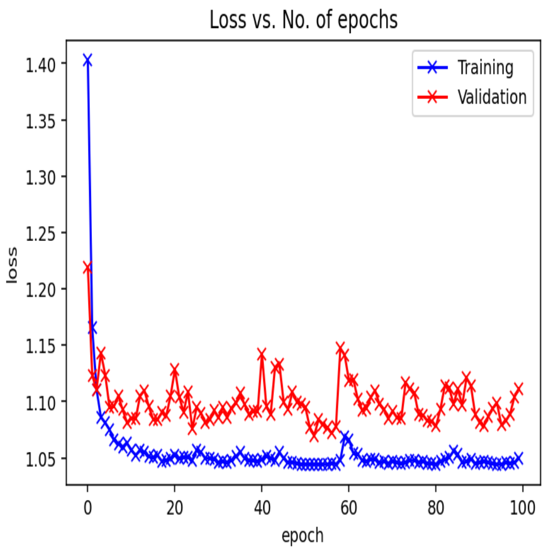
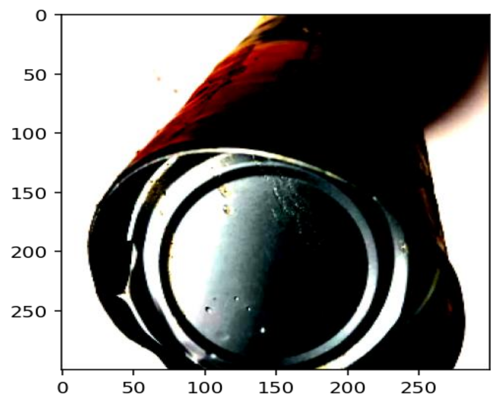
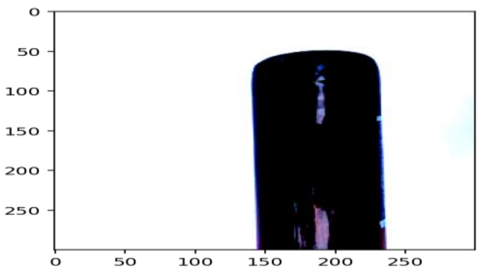
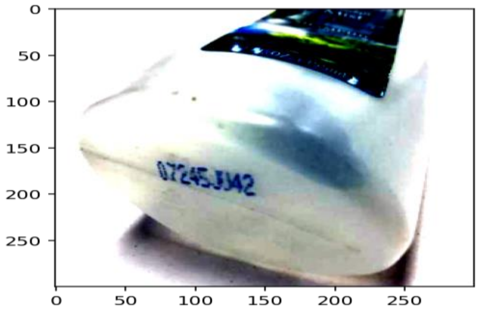


Figure 7: Loss vs. Epochs



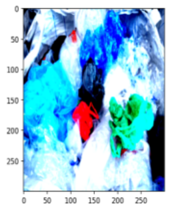




PASSED:

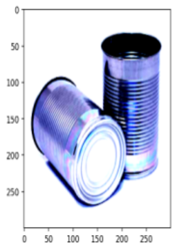
```
predict_external_image('plastic.jpg')
```

```
Clipping input data to the valid range for imshow with RGB data ([0..1] for floats or [0..255] for integers).
The image resembles plastic.
```



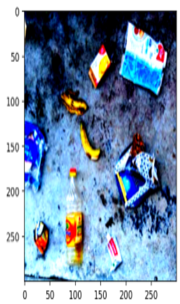
```
predict_external_image('cans.jpg')
```

```
Clipping input data to the valid range for imshow with RGB data ([0..1] for floats or [0..255] for integers).
The image resembles metal.
```



```
predict_external_image('mixed-trash.jpg')
```

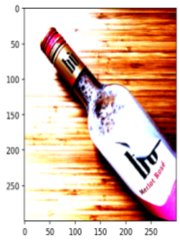
```
Clipping input data to the valid range for imshow with RGB data ([0..1] for floats or [0..255] for integers).
The image resembles plastic.
```



FAILED

```
predict_external_image('wine-trash.jpg')
```

```
Clipping input data to the valid range for imshow with RGB data ([0..1] for floats or [0..255] for integers).
The image resembles cardboard.
```



VII. CONCLUSION

In this study, a video dataset of four different kinds of medical waste was assembled from a range of backgrounds. In this use case, R3D+C2D network was suggested as a novel network design, and it significantly beat R3D at 79.99% test accuracy with spatial trimming. This resulted in two key findings for us: 1) When there is a lack of training data, motion-based preprocessing may assist the network concentrate on the moving item. 2) Both 3D and 2D convolutions are required to give the motion and appearance information required to distinguish between comparable falling objects. Future works should focus on the pre-processing steps of our system, finding ways to integrate it with the deep learning steps, or on expanding our dataset to contain more classes and colors of medical waste. Experimentally, we see that there are still some problems in our temporal trimming and spatial trimming. Given a better version of a temporal cut, it is possible to improve our accuracy to 84.54% and likely a better spatial cut would improve this further as promising future steps. Next, we should seek to increase the size of our dataset. This includes increase the number of sample videos and seek to get some real OR example videos to test our system on. Our current set-up shows to be very promising, so it would be a worthwhile endeavor to expand this system to a greater dataset and more real world uses. Finally, we can apply this network to another use-case. In the OR there exists two types of waste biohazard and non-biohazard wastes. Biohazard waste must be disposed of in a more money and energy intensive manner to ensure it does not contaminate other people or animals. On the other hand, non-biohazard waste can simply be disposed like any other waste. In some OR's, they do not do sort and simply label everything as biohazardous waste. We believe that we could automate this procedure given our promising pilot study results.

References

[1] D. de SA, K. Stephens, M. Kuang, N. Simunovic, J. Karlsson, and O.R Ayeni, "the direct environmental impact of hip antroscopy for femoractublar impingement: a surgical waste audit of five cases,: Journal of hip preservation surgery, vol. 3, no. 2, pp. 132-137, 2016.

[2] F. McGain, D. Story, and S. Hendel, "An audit of intensive care unit recyclable waste,"

[3] Anastehsia, vol. 64, no, 12 pp1299-1302, 2009.

[4] A. Krizhevsky, I. sutskever, and G. E. Hinton, "Imagenet classification with deep convolution neural networks,' in Advances in neural information processing systems, 2012, pp. 1907-1105

[5] D. Tran, H. Wang, L. Torresani, J. Ray, Y. LeCun, and M. Palur, "A closer look at spatiotemporal convolution for action recognition," in Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2019 pp. 6450-6459

[6] W. Kay, J. Carreira, K. Simonyan, B. Zhang, C. Hillier, S. Vijaya-narasimhan, F. Viola,

[7] T. Green, T. Back, P. Natsevet al., "The kineticshuman action video dataset,"arXiv preprint arXiv:1705.06950,2017

[8] [6] G. E. Sakr, M. Mokbel, A. Darwich, M. N. Khneisser, and A. Hadi,"Comparing deep learning and support vector machines for autonomous waste sorting," in 2016 IEEE International Multidisciplinary Conference on Engineering Technology (IMCET). IEEE, 2016, pp.207–212.

[9] A. H. Vo, M. T. Vo, T. Leet al., "A novel framework for trash classification using deep transfer learning,"IEEE Access, vol. 7, pp.178 631–178 639, 2019.

[10] Z. Zivkovic, "Improved adaptive gaussian mixture model for back-ground subtraction," in Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004., vol. 2. IEEE, 2004,pp. 28–31.

[11] C. Li, G. Dobler, X. Feng, and Y. Wang, "Tracknet: Simultane ousobject detection and tracking and its application in traffic videoanalysis," arXiv preprint arXiv:1902.01466, 2019.

[12] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in Proceedings of the IEEE conference on computer vision and pattern recognition,2018, pp. 4510–4520.

[13] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller,"Striving for simplicity: The all convolutional net,"arXiv preprintarXiv:1412.6806, 2014