International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. #, No.# (Mon-20..)

Deep Clean: A Weakly Supervised Waste Localization System Using Deep Convolutional Neural Network

Mohd Anjum[1], M. Sarosh Umar[2], and Sana Shahab[3]

[1]*Department of Computer Engineering, Aligarh Muslim University, Aligarh, India*
[2] *Department of Computer Engineering, Aligarh Muslim University, Aligarh, India*
[3]*Department of Business Administration, Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia*

E-mail address: mohdanjum@zhcet.ac.in, saroshumar@email.com, sshahab@pnu.edu.sa

Received ## Mon. 20##, Revised ## Mon. 20##, Accepted ## Mon. 20##, Published ## Mon. 20##

Abstract: Nowadays, waste dumping on the city streets has become more frequent, especially in developing countries due to the exponential growth of waste generation. This dumping directly affects the city cleanliness, damages resident's health, and pollutes the surrounding environment, air, and water. The conventional waste dump detection and collection method involve humans who visit the streets and spots and manually confirm if any dump is obtained. This method requires a considerable number of employees and manual operations, which demands a significant amount of energy, time, and money. Additionally, the random appearance of waste dump on streets cannot be controlled through the conventional method. The research proposes the automated waste dump detection and localization method using deep learning to overcome these disadvantages. In this method, the weakly supervised learning approach is implemented using a deep convolutional neural network model. The deep convolutional neural network is trained for two categories: waste and no waste, using a manually constructed dataset and tested for the above categories and localizing waste dump in images. The model performance is evaluated through matrices for classification, and a survey is conducted to assess the accuracy mask generated by the model for waste localization. The precision, recall, F-score, accuracy, and MCC matrices are 0.9708, 0.9848, 0.9778, 0.9776 and 0.9553, respectively. The average score from the survey for generated masks is obtained 3.9. The performance matrices result imply that the model performs outstanding for classification with an accuracy of 97.76 percent and is significantly good for localization with an average score of 3.9. Additionally, the study demonstrates two approaches for the practical application of the implemented model. (i) Citizen oriented approach: It integrates mobile application with the model. (ii) Internet of Things oriented approach: It integrates the existing surveillance system and the model.

Keywords: Municipal solid waste, Deep learning, Weakly supervised learning, Deep convolutional neural network, Max pooling

1. Introduction

Municipal solid waste (MSW) has emerged as a fundamental concern at the local, national, and global level as it is directly associated with human health, environmental pollution, and sustainable development of metropolitan areas. Generally, MSW contains the non-hazardous materials generated from daily life activities, but residuals yield from different industrial, residential, and commercial activities are also considered. According to the World Bank report [1], the current MSW generation at the global level is approximately 1.3 billion tons per year. It will increase by around 2.2 billion tons per annum by 2025; India is the most significant contributor to this enormous quantity. The MSW amount grows with population growth, urban migration, socio-economic development, and rising living standards. The increasing volume has become a severe problem in developing countries like India, as the local government lacks fund allocation, resources, and advanced technology to manage the generated MSW. Due to this, it is observed that the waste is dumped illegally on city streets, which severely affects the cleanliness of the city, resident's health, and the surrounding environment [2]. The analysis of last decade waste data shows that the quantity of waste is increasing in multiple folds, so the dumping cases are occurring more frequently [3]. Also, many studies fundamentally demonstrate that proper management is essential to control and monitor waste dumping. Therefore, dump detection is one of the primary concerns of municipal bodies to provide effective and efficient street cleaning services. The preliminary information about the waste location will

E-mail:author's email

significantly improve the collection process in time and energy. Numerous studies have sufficient evidence that if there are waste dumps on streets or roadside; residents do not hesitate to do more dumping. Therefore, it is required to detect the waste as soon as it is thrown to maintain cleanliness. If streets are free from waste clouds, every individual will pay attention before throwing anything, and ultimately a good habit will be developed for waste disposal.

Due to the adverse effect of open dumping of waste in residential areas, it is necessary to detect the dumped waste automatically for immediate action of its removal and further monitoring. Nowadays, to decrease the frequent cases of open dumping, many municipalities run various programs based on social media portals, online surveillance systems, and voluntary dumps reported by citizen and implement their rules in terms of levy or punishment. In residential areas, open dumping hot spots are overseen by municipality employees and reported to the administration whenever any waste dumping is found. In the existing municipal solid waste management (MSWM) system, especially in developing countries, waste surveillance tasks such as dump identification, reporting, and monitoring involve humans, so this system incurs a substantial operating cost, energy, and high time. Now, an advanced technology-based efficient, and cost-effective method is required to solve the problems mentioned above. The study combines the MSWM and current technology to produce excellent results. The proposed approach implements the deep convolution neural network (DCNN) to detect and localize the waste dump in the image automatically.

Machine learning (ML) has potential applications in MSWM, especially detection of waste dumping on the street, waste classification for recycling and disposal, garbage bin level detection and monitoring, and waste generation forecasting involving various influencing factors. It refers to the set of techniques and algorithms from artificial intelligence that aim to develop a system that can learn from training and takes decisions based on experience without human intervention. Deep learning (DL) is an epoch-making area of ML that exploits the artificial neural network with multiple layers and is also called a deep neural network. It provides extraordinary computing techniques, so it is most prevalent in computer science research. The convolutional neural network (CNN) is a significant deep neural network with substantial computational power to analyze the image data. They have emerged as a fundamental solution for an extensive range of problems such as photo tagging, medical diagnosis, language translation, virtual assistant, self-driving car, image caption and colour generation etc. [4].

Additionally, with the evolution of fast computing hardware and CNN, waste detection and classification can be performed efficiently with greater precision. For waste detection, localization, and classification, the system must be significantly resilient and capable of functioning correctly in random environment and variable background. The literature exhibits that supervised and unsupervised learning have been implemented using different types of CNNs [5]. These networks have shown many flexible functionalities for recognition, localization, and categorization [5]. The research implements the weakly supervised learning approach using DCNN to detect the waste dump on the street. The implemented DCNN is also integrated with the mobile app to exhibit the practical application. Additionally, it is also be combined with a city surveillance system to monitor the dump on the streets without human intervention.

The remaining structure of the study is stated as follows. Section 2 presents the review of published research in recent years, titled: Application of DL in MSWM. Data set construction and implemented DCNN are discussed in section 3. Evaluation criteria are formulated in section 4 to examine the outcomes. Section 5 illustrates the integration process of DCNN with other systems to depict the practical applications. Section 6 demonstrates the conclusion of the study and the associated discussion.

2. Literature Review

The literature review demonstrates the application of DL in MSWM studies published in the last few years. It critically analyses and compares the studies that apply the various approaches of DL to detect, classify and localize the waste in the image. DL techniques have the capabilities to recognize and learn features directly from the image. This characteristic has significantly improved image detection and classification. In 2012, a deep CNN AlexNet [6] had been proposed and implemented that has exhibited remarkable classification accuracy in the Large-Scale Visual Recognition Challenge [7]. AlexNet has also been successfully implemented in voice recognition and analysis of medical image [6]. According to the reviewed literature, it is identified that various image recognition approaches have been implemented in MSWM, waste processing and recycling. After the emergence of different deep CNN, numerous studies have been published with significant accuracy to classify the waste using image data. In [8], state of the art CNN architecture AlexNet is trained on the collected data set to detect the waste and classify it into various categories to segregate before disposal.

Similarly, other existing architecture, YOLOv2 and YOLOv3, are utilized to recognize the garbage bin and differentiate the disposable and recyclable materials [9], [10]. A mobile application system is developed to collect the e-waste. The system relies on the CNN to detect the type of electronic waste and faster region CNN to predict shape and size used to determine the number of collecting vehicles and their capacity [11]. Additionally, A faster region-based CNN is exploited to count the garbage bags [12]. In [13], the transfer learning approach combines the three models VGG19, DenseNet169, and NASNetLarge to classify the waste image into six categories. Various CNN architectures with the different number of layers [14], [15] and other CNN such as improved ResNext [4], DenseNet121 [16], Auto Encoder Network [17], Multilayer Hybrid CNN [18] have been proposed and implemented to detect and classify the waste in different categories, namely organic, inorganic, medical, plastic, glass, metal, paper, cardboard and trash etc. Waste classification is generally utilized to segregate different types of waste before the recycling process. Different types of CNN are also significantly utilized in the various task of MSWM other than waste detection and classification. A Long Short Term Memory CNN is trained to forecast the waste generation considering various influencing factors [19][20] and carbon dioxide concentration inside the garbage bin [21]. A deep CNN is proposed to predict the per capita waste generation [22] and demolition waste for three categories reusable, recyclable and landfill [23]. A hardware architecture equipped with a camera, microcontroller and servo motor is developed to segregate the different types of waste materials automatically. The hardware is managed by custom software based on the ResNet-34 algorithm with multi-feature fusion and a new activation function [24][25]. Similar hardware is utilized in addition to LoRa communication protocol and TensorFlow object detection model in place of ResNet-34 to design smart trash bin [26].

Additionally, CNN has been successfully implemented to detect different types of polyethene [27] and defect in potatoes [28]. Moreover, a robot is built to pick the garbage on the grass. The robot operations rely on custom software-based SegNet and ResNet, where SegNet differentiate the ground from other objects, and ResNet detects and locates the objects [29]. According to the review literature, it can be confidently concluded that most studies are concentrated on waste classification. Therefore, the study proposed a DCNN architecture to detect and localize the waste dump on city streets. The model can be utilized in reducing the waste dumping frequency, real-time monitoring and assessing street cleanliness. Additionally, image location data can be used for collection vehicle routing and scheduling.

3. METHODOLOGY

Generally, a fully supervised learning approach is applied to train the DCNN to solve the recognition and localization problems. In this approach, a predictive model is trained from a large labelled image dataset, where the label of each data point (pixel) specifies its ground truth [30]. If this technique is applied in the above problems, then each pixel should be manually tagged with its ground truth, which is a very tedious process and involves high cost and time. Therefore, this article implements the weakly supervised learning approach to overcome the drawbacks mentioned above. In weak supervision, the predictive model is trained according to the image label, where the label represents the class of individual images. Here, a label is given to every image, not to each pixel.

The article proposes an architecture of DCNN and implements it to localize the waste regions present in an image. The DCNN is trained for two categories, namely waste and non-waste, using constructed dataset. In the testing phase, the network predicts the input test image category by global average pooling over the segmented probabilities generated by the model. The network predicts the probabilities of waste for every pixel in an image. The main goal of this DCNN architecture is to enable the weakly supervised algorithm to take advantages of global average pooling. The probabilities masks generated by the model is further processed. Here, the noisy edges in output images are removed if the region enclosed is smaller than a predefined threshold; this threshold is 25 pixels. This model identifies the waste dump on the street if the test image contains the waste areas. The remarkable performance of the implemented network shows its utilization scope in smart city development.

A. Data Set Construction

In image recognition, massive research has been carried out due to the availability of massive, annotated datasets, namely ImageNet, CIFAR10, and MS-COCO. After the rejuvenation of DCNNs, image recognition research is increased multi-fold in the last few years. Contrary to all published datasets for machine learning research, the extensive database of ImageNet contains more than 10 million classified natural images with approximately 1000+ categories. These datasets provide the backbone for the training of DCNNs to substantially increase the accuracy of object recognition and segmentation problems [31]. Among these large categories, no such class is available to train the DCNN for waste detection and localization. Additionally, no other waste dataset is publicly accessible to obtain research objective, and the own dataset is built in previous studies. Therefore, a dataset is handcrafted that comprises two categories: waste and non-waste, to accomplish the research goal.

The constructed dataset contains 5000 images in each category. These images are collected from different internet sources and manual capturing. All three authors have done manual label the collected images according to the criteria: an image is placed in one of the dataset categories if it is labelled the same at least by two people; otherwise, it is discarded. Following the above process, 5000 images are labelled as waste and 5000 images as non-waste to form the final dataset. The dataset has no ambiguity in categorization within the human perception of recognition. Deep classified data distribution characteristics play an essential role in learning more general and accurate models [31]. The selected images have a natural background and various environmental conditions that train the system more robustly and improve the waste detection accuracy. The study divides the dataset into two distinct sets; namely, the training set includes 75% images of the total from each category, and the remaining 25% are put in the testing set.

B. Implemented DCNN

CNN is multi-layered architecture composed of consecutive convolutional and pooling layers in succession with fully connected layer(s) like a conventional multilayer neural network. Convolution layers take the dot product of the sliding filter or kernel and apply a non-linear activation function. The calculated dot product outputs are convolved feature or feature map. CNNs are multi-layered neural networks generally used in recognizing and localizing visual patterns directly from image pixels. The CNN architecture can exploit the two-dimensional structure of the input image. It is obtained by making weighted connections between consecutive convolutional layers and making any pooling that generates the translation invariant feature maps. They are very robust for distortion and minimal pre-processing. The CNNs are mainly used in object recognition and localization in an image or video with remarkable performance and accuracy in real-time.

Additionally, they have wide application in image segmentation and restoration, vision-based obstacle detection and avoidance. In this paper, the max-pooling along with global averaging pooling CNN is implemented. This type of network consists of consecutive convolutional and pooling in alternation. This proposed network comes under a large class of models known as Multi-Stage Hubel-Wiesel Architectures that determine position-specific simple small regions with local receptive fields like convolutional layers and complex regions that carry out operations like pooling layers. Figure 1 illustrates the proposed CNN architecture.

The convolutional layer of the network consists of linear filters in the succession of a non-linear activation function to detect the input. The CNNs are acknowledged as potential networks due to parameter sharing and integration of convolutional and pooling layers. They have combined convolutional and pooling layer. In image processing, convolution operation implies a filter operation that extracts the feature map from the input. After convolution operation, max pooling is performed to shorten the feature map by exploiting neighbouring pixels of the same nature from the feature map—this deep structure of alternatively connected convolution and pooling layer built forward inference pass. In CNN, elementary matrix multiplication operations are performed, and these operations do not have any branch operation. Hence, these operations are implemented through parallel computation, which can be most efficiently executed on GPUs. Additionally, CNN is advantageous in terms of training and the number of parameters. They have significantly fewer parameters and easy to train than the fully connected network of the same hidden layers.

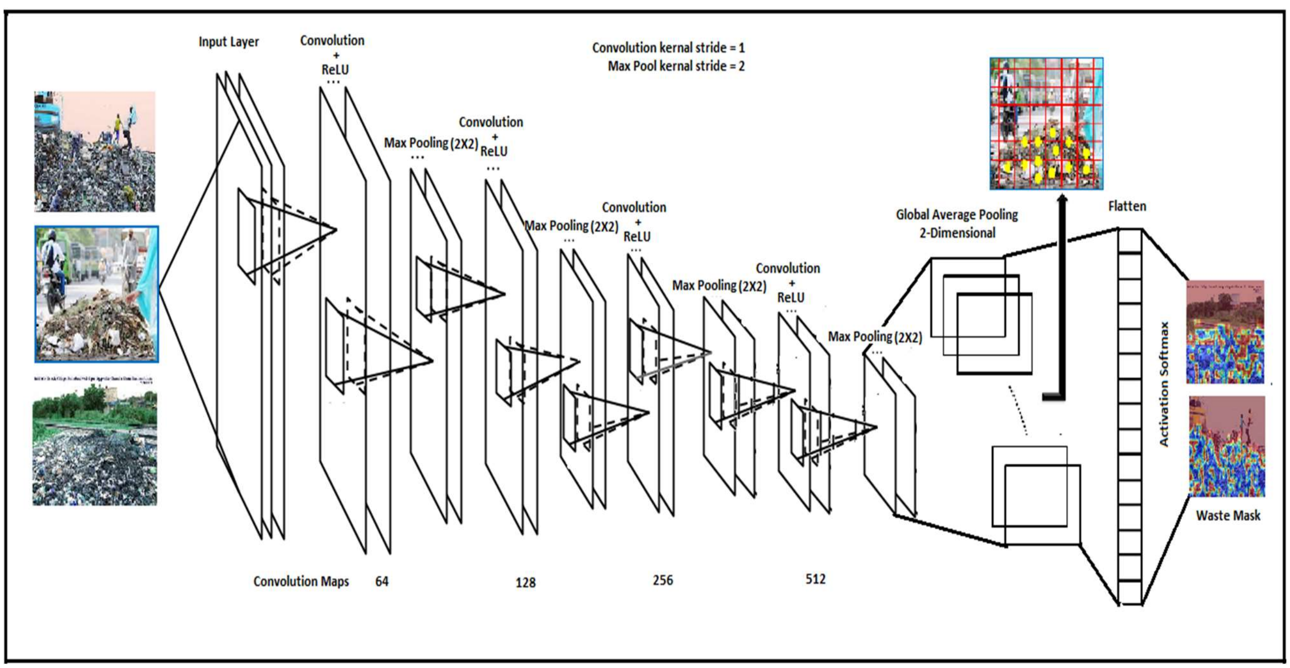

Figure 1. Convolution neural network

Input Layer: This layer of CNN defines the size of the input image and contains directly extracted values of raw pixels. The image size is described by the height, width and number of colour channels (1 and 3 for grayscale and colour images, respectively) of that image. Contrary to neural networks that take the input data as a vector, data is fed as a multi-channel image. This layer also performs the additional function of data pre-processing. There are many ways of data pre-processing such as simple rescaling or normalization, mean subtraction, normalization and principal component analysis (PCA) and whitening. In a simple rescaling process, input data is rescaled independently in all data dimension so that the output data vector has the range [0, 1] or [-1, 1] according to the dataset. Mean subtraction is generally applied when data is stationary; it subtracts the average of training data set from each feature in the input data. Geometrically it centralizes the data cloud nearby origin along each dimension. PCA and whitening is a type of pre-processing in which data is concentrated, as explained above. Then covariance matrix is calculated to find the correlation structure in the dataset.

Convolution layer: It is the primary building block of CNN, consisting of independent filters that convolute independently with the image and generate feature maps. It consists of several parameters such as filters and strides. The size of data and complexity determines the count of convolutional layers.

Filters and strides: The Convolutional layer of the CNN has a two-dimensional layer of neurons mapped over the small segments of the input images. Neurons set up the weights over these segments to learn the features from these areas. The set of applied weight is known as a filter. The filter is shifted in the horizontal and vertical direction for performing the same calculation in each area. The step size of the shift is called a stride.

Feature Maps: When the filter walks on the input image or output of the previous layer, it utilizes the same set of weights for performing the convolution operation to generate the feature maps. Consequently, the count of the feature maps is equal to the filters. All feature maps have a distinct set of weights and neurons of the same map utilizing similar weights for different input regions [5]. All these filters are initialized randomly and become our parameters which will be learned by the network subsequently.

Batch Normalization Layer: The primary function of this layer is to accelerate the network training and decrease the responsiveness of the network towards weights initialization and used between convolutional layer and nonlinearities such as rectified linear units (ReLU). The first step of this layer function is to normalize the activations of all the filters by subtracting the batch average of the inputs, followed by dividing batch standard deviation. This layer normalizes the activations and gradients passing through the network to engender training task straight forward and effortless.

ReLU Layer: This layer consists of a non-linear activation function and generally used after the convolutional and batch normalization layer. The function of this layer is to perform the threshold operation for all inputs; if any input is below zero, then set to zero, otherwise equal to the input. Mathematically, the activation function is given as:

$$f(x) = \{x; x \geq 0 \text{ else } 0$$

It does not alter the dimension of the input data matrix.

Max pooling: This layer reduces the dimension of data output from the convolutional layer (downsampling), decreasing the number of links to the successive layer, generally fully connected layer. This layer reduces the number of parameters learned in previous layers. This layer does not also carry out any learning. Therefore, it shows the speedy converging by choosing the superior invariant features that make better the generalization capability of the network [5].

Fully connected layer: One or more fully connected layers are present after convolutional and pooling layers that connect all the neurons of it to all the previous layer neurons. It discovers the patterns to classify the images by amalgamation of the features learned in the previous layers. It generally learns the non-linear function in this feature space. Additionally, the name-value argument is used to select the value of regularization parameters to improve the learning rate and reduce overfitting.

Global average pooling: In traditional CNNs, the lower layer of the network implements the convolution function. In the classification problem, the two-dimensional feature maps output from the last convolutional layer are converted into a vector and given as input to the fully connected layers, followed by a regression layer. This structure superimposes the CNN architecture with the conventional neural network classifier. The convolutional layer works as a feature extractor, and the output features are categorized traditionally.

In this research paper, global average pooling is implemented instead of the conventional fully connected CNN layer. In this strategy, the average is taken for every feature map. The final output vector is given as input to

E-mail:author's email

the softmax layer instead of directly substituting fully connected layers on top of the feature map. Global average pooling is commonly used in CNN over fully connected layers by establishing correspondences between feature maps and classes. Hence, the feature maps directly give the classes, so it is interpreted as class confidence maps. Additionally, it has no parameter to optimize, so no overfitting is shown [32] and integrates all spatial information, so this layer is more flexible for translating input spatial data.

Output layer (SoftMax and Classification layers): In categorization, the fully connected layer is fed by the output of the adjacent SoftMax and classification layer. It is a generalized logistic function for multiple classes and known as a normalized exponential function. In categorization, each category is represented by one neuron in the output layer. Each neuron uses the SoftMax activation function to give the posterior class probability as output [5].

Loss function: It is preferable to use cross-entropy over the classification and mean square error when performing categorization, prediction and quality assessment of CNN. The cross-entropy loss function computes the distance between the empirical data distribution and the model predictive distribution. The cross-entropy for binary classification is given as:

$$\mathrm{C} = -\mathrm{b} * \ln(\mathrm{p}) - (1-\mathrm{b}) * \ln(1-\mathrm{p})$$

where, p – Predicted probability for an observation O that it belongs to the class C.
b – Binary indicator (0 or 1) if class label C is the true class of an observation.

4. EVALUATION CRITERIA

The implemented DCNN model categorizes the constructed waste dataset into two classes, namely waste and non-waste. Additionally, the model localizes the regions into the images of the waste class, which comprises the waste. The model performance for categorization is measured through evaluation parameters, while a survey is conducted to assess the localization accuracy. The evaluation parameters and survey procedure are demonstrated as follows.

A. Evaluation parameters

All categorization prediction by the DCNN model is compared with ground truth which is the actual category label. The confusion matrix is represented in Table I that quantifies the correctly and incorrectly predicted samples.

TABLE I. CONFUSION MATRIX FOR BINARY CLASSIFICATION OF A WASTE DATA SET

		Prediction	
		Waste	***No Waste***
Ground Truth	***Waste***	True Positive (TP)	False Negative (FN)
	No Waste	False Positive (FP)	True Negative (TN)

TP and TN represent the correctly classified samples, while FP and FP exhibit the incorrectly predicted samples.

The study computes the following matrices to evaluate the performance of the DCNN model. The formulas of performance matrices are displayed in Table II.

TABLE II. CONFUSION MATRIX FOR BINARY CLASSIFICATION OF A WASTE DATA SET

Evaluation parameters	Formula
Precision: It is correctly classified positive samples with respect to all positive predictions.	$\mathrm{Pr.} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}$
Recall: It is correctly classified positive samples with respect to the actual positive class.	$\mathrm{Re.} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}$
F-Score: It is the harmonic mean of the precision and the recall assessing the success of the model.	$\mathrm{F-Score} = \frac{2 * \mathrm{Pr.} * \mathrm{Re.}}{\mathrm{Pr.} + \mathrm{Re.}}$
Accuracy: It is all correctly classified samples with respect to all the samples.	$\mathrm{Acc.} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{FP}+\mathrm{TN}+\mathrm{FN}}$
Matthews Correlation Coefficient (MCC): It computes the correlation by considering the actual class and predicted class as two binary variables. A high correlation between actual and predicted classes indicates a better prediction. The formula of MCC is given as below. $\mathrm{MCC.} = \frac{\mathrm{TP} * \mathrm{TN} - \mathrm{FP} * \mathrm{FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})}}$	

B. Survey for localization evaluation

The individual class of waste images does not exist in all published benchmark image databases, including ImageNet, CVonline, etc. Therefore, model training and testing cannot be done for the benchmark dataset. Consequently, results cannot be shown and compared quantitatively with ground truth. In this study, a practical procedure is followed to comprehend and demonstrate the results qualitatively. A survey is conducted to compare the masked images with corresponding original images, and the overlapping of masks over waste is graded on the Likert scale of 5. A set of random 1000 images is constructed from the test dataset and partitioned into 20 groups, each of 50 images. Each group is surveyed and graded by an individual for the following question. How correctly the mask is overlap with the waste in the original image? After comparison, the individual assigns one of the discrete scores 1, 2, 3, 4, and 5 to each image; these numbers represent no, poor, moderate, high, and complete overlapping, respectively. Finally, the average score is computed to depict the outcomes on a scale of 5. Additionally, the histogram is also constructed to show the scores graphically.

5. RESULTS AND DISCUSSION

The DCNN model training utilizes 7500 images while testing is performed over 2500 images. Table III displays the correctly and incorrectly classified samples for both waste and no waste categories. The classification performance of the model is demonstrated in Table IV using performance matrices. The DCNN model performs remarkably outstanding with a categorization accuracy of 97.76 percent. The value of MCC indicates that the predicted class is highly similar to the actual class. It can be concluded that predictions are good.

TABLE III. CONFUSION MATRIX FOR BINARY CLASSIFICATION OF A WASTE DATA SET

		Prediction	
		Waste	*No Waste*
Ground Truth	*Waste*	1231	19
	No Waste	37	1213

TABLE IV. SCORES OF PERFORMANCE EVALUATION PARAMETERS

Class	Precision	Recall	F-Score	Accuracy	MCC
Waste	0.9708	0.9848	0.9778	0.9776	0.9553

The survey is conducted to assess the performance of the DCNN model for the localization of waste. Figure 2 depicts the graph of obtained scores and the corresponding number of images. The masks generated by the model do not overlap with waste in 2.3 percent images while partially overlap in 4.3 percent. The 20.3 percent images have moderate overlap with waste, while more than 72 percent have high or complete overlapping. The average score is computed 3.9 out of 5, which implies that most of the mask correctly overlap with waste.

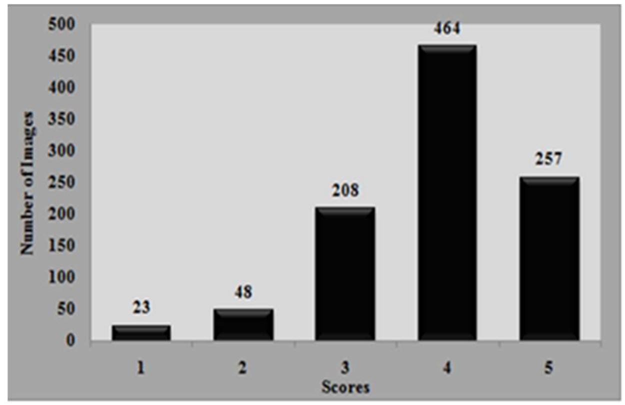

Figure 2. Scores of surveyed images

Figure 3 demonstrates and compares the few sample images arbitrary chosen from the post-test dataset with corresponding actual images. The probabilities of waste are illustrated through colour variation. Additionally, it is visible in the sample images that masks are substantially overlapped with the waste regions.

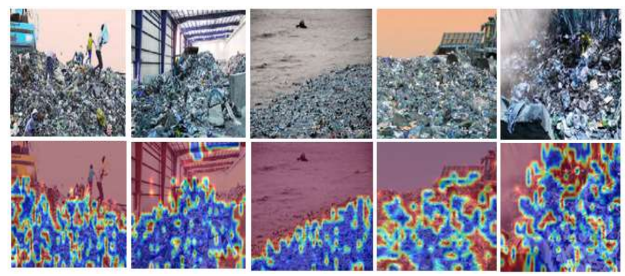

Figure 3. Images with localized waste regions and corresponding original images

6. PRACTICAL APPLICATIONS

Nowadays, smart city has emerged as a modern concept of urban development, where recent technologies are exploited to design, develop and operate prominent services. Therefore, many countries worldwide have started developing smart cities or making a substantial effort to rebuild the existing cities. Smart cities have a wide range of smart services from governance to the environment. MSWM is one of the essential services to maintain the clean environment around the city. The implemented approach can be integrated with other platforms to develop practical applications. The suggested framework can be effectively utilized in real-time monitoring and assessing street cleanliness. The study explains the two methods of integration as follows:

A. Integration with mobile: A citizen-oriented approach

Currently, mobile devices are one of the most prevailing communication and data transfer instruments due to exponential growth in their number and availability worldwide. The computing power of mobile devices has increased multi-folds along with significant additional features. Additionally, ubiquitous internet connectivity with high-speed data transfer services makes these devices more usable to perform numerous tasks. Therefore, mobile phones can be effectively used for data collection with supplementary information, such as the GPS or current location of the device. Furthermore, mobile applications have completely transformed the way of interaction and make them easy to use. Nowadays, almost everyone has a permanent mobile device, so reporting of activities can be performed multiple times without recalling. The implemented DCNN can be integrated with the mobile app to make waste dump collection services more ubiquitous and accessible.

The study introduces the citizen-oriented approach to develop the practical application, which involves the citizens reporting the waste dumps on streets. The approach develops a framework comprising three components: mobile application, waste database, and DCNN module. The citizens use the mobile application to capture the waste dump images and send them to the waste database with additional information such as waste location and address. These images are given as input to the DCNN module to detect and localize the waste. After processing from the DCNN module, location and address information of those images are extracted in which waste dump has been detected and localized. The collecting vehicle collects the waste from these identified locations. The designed framework is shown in figure 4.

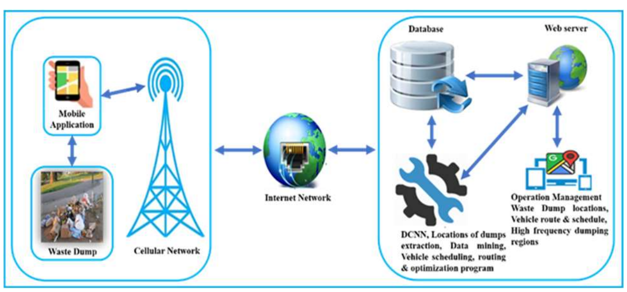

Figure 4. The architecture of DCNN and mobile application integration

B. *Integration with surveillance system: IoT oriented approach*

A smart city is an emerging approach to urban development with improved operational efficiencies. It combines information and communication technology (ICT) and the Internet of Things (IoT) to provide services, manage resources, and connect citizens efficiently and sustainable manner. Smart city has an extensive network of surveillance cameras and other sensors to provide improved and efficient transportation and mobility, providing citizens safety and monitoring the infrastructures. The output of these cameras can be combined with implemented DCNN to detect and monitor the waste dumps on streets.

The study presents an IoT-oriented approach that utilizes the frame from surveillance camera videos at regular intervals as an input image to the DCNN module. The DCNN module performs processing on the input image and identifies the waste dumps along with their localization. Now, the camera address of those images is determined, which comprise the waste dumps. The collecting vehicle utilizes these addresses to pick the waste. The design of the whole system is depicted in figure 5.

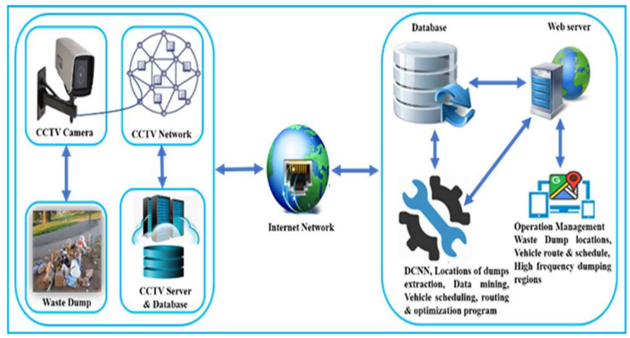

Figure 5. The architecture of DCNN and surveillance camera integration

Furthermore, route optimization algorithms are also applied to output location data to determine the optimized route and schedule for collecting vehicles. Additionally, data mining techniques can also be utilized to determine the waste dumping patterns in different regions of the city. After analyzing these patterns, municipal authorities can launch the awareness program in more frequent waste dumping regions. This location-enabled framework will contribute to sustainable development, especially in developing countries, and it will also help to maintain the clean environment and streets of the city.

7. CONCLUSION

The study implements the weakly supervised learning approach using DCNN to detect and localize the waste dumps on the city streets. The first contribution, a dataset, is manually constructed, which comprises 10,000 images for two categories collected from different internet sources and captured through a camera. Secondly, a DCNN architecture is designed and implemented to exhibit the model success. The results analysis demonstrates that the DCNN model has outstanding performance for waste dump detection and is remarkably good for waste localization. The model achieves an accuracy of 97.76 percent for waste detection and attains a score of 3.9 in the survey for localization. Finally, practical applications demonstrate that the DCNN model has significant importance to create a working real-time waste monitoring system, which is essential for smart city development.

REFERENCES

[1] S. Kaza, L. C. Yao, P. Bhada-Tata, and F. Van Woerden, *What a Waste 2.0: A Global Snapshot of Solid Waste Management to 2050*, vol. 1. Washington: World Bank Publications, The World Bank Group, 1818 H Street NW, Washington, DC 20433, 2018.

[2] P. Ping, G. Xu, E. Kumala, and J. Gao, "Smart Street Litter Detection and Classification Based on Faster R-CNN and Edge Computing," *Int. J. Softw. Eng. Knowl. Eng.*, vol. 30, no. 4, pp. 537–553, 2020, doi: 10.1142/S0218194020400045.

[3] L. A. Guerrero, G. Maas, and W. Hogland, "Solid waste management challenges for cities in developing countries," *Waste Manag.*, vol. 33, no. 1, pp. 220–232, 2013, doi: 10.1016/j.wasman.2012.09.008.

[4] A. H. Vo, L. Hoang Son, M. T. Vo, and T. Le, "A Novel Framework for Trash Classification Using Deep Transfer Learning," *IEEE Access*, vol. 7, pp. 178631–178639, 2019, doi: 10.1109/ACCESS.2019.2959033.

[5] J. Nagi *et al.*, "Max-pooling convolutional neural networks for vision-based hand gesture recognition," in *2011 IEEE International Conference on Signal and Image Processing Applications, ICSIPA 2011*, 2011, pp. 342–347, doi: 10.1109/ICSIPA.2011.6144164.

[6] A. Krizhevsky, Ii. Sulskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Adv. Neural Inf. Process. Syst.*, vol. 60, no. 6, pp. 84–90, 2012, doi: 10.1145/3065386.

[7] O. Russakovsky *et al.*, "ImageNet Large Scale Visual Recognition Challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, 2015, doi: 10.1007/s11263-015-0816-y.

[8] A. Dabholkar, B. Muthiyan, S. Srinivasan, S. Ravi, H. Jeon, and J. Gao, "Smart illegal dumping detection," *Proc. - 3rd IEEE Int. Conf. Big Data Comput. Serv. Appl. Big Data Serv. 2017*, pp. 255–260, 2017, doi: 10.1109/BigDataService.2017.51.

[9] M. Valente, H. Silva, J. M. L. P. Caldeira, V. N. G. J. Soares, and P. D. Gaspar, "Detection of waste containers using computer vision," *Appl. Syst. Innov.*, vol. 2, no. 1, pp. 1–13, 2019, doi: 10.3390/asi2010011.

[10] S. Kumar, D. Yadav, H. Gupta, O. P. Verma, I. A. Ansari, and C. W. Ahn, "A novel yolov3 algorithm-based deep learning approach for waste segregation: Towards smart waste management," *Electron.*, vol. 10, no. 1, pp. 1–20, 2021, doi: 10.3390/electronics10010014.

[11] P. Nowakowski and T. Pamuła, "Application of deep learning object classifier to improve e-waste collection planning," *Waste Manag.*, vol. 109, pp. 1–9, 2020, doi: 10.1016/j.wasman.2020.04.041.

[12] K. Mikami, Y. Chen, J. Nakazawa, Y. Iida, Y. Kishimoto, and Y. Oya, "DeepCounter: Using deep learning to count garbage bags," *Proc. - 2018 IEEE 24th Int. Conf. Embed. Real-Time Comput. Syst. Appl. RTCSA 2018*, pp. 1–10, 2019, doi: 10.1109/RTCSA.2018.00010.

[13] G. L. Huang, J. He, Z. Xu, and G. Huang, "A combination model based on transfer learning for waste classification," *Concurr. Comput.* , vol. 32, no. 19, pp. 1–12, 2020, doi: 10.1002/cpe.5751.

[14] K. Ahmad, K. Khan, and A. Al-Fuqaha, "Intelligent Fusion of Deep Features for Improved Waste Classification," *IEEE Access*, vol. 8, pp. 96495–96504, 2020, doi: 10.1109/ACCESS.2020.2995681.

[15] A. A. A. G. S. Altikat, "Intelligent solid waste classification using deep convolutional neural networks," *Int. J. Environ. Sci. Technol.*, pp. 1–8, 2021, doi: 10.1007/s13762-021-03179-4.

[16] W. L. Mao, W. C. Chen, C. T. Wang, and Y. H. Lin, "Recycling waste classification using optimized convolutional neural network," *Resour. Conserv. Recycl.*, vol. 164, no. August 2020, p. 105132, 2021, doi: 10.1016/j.resconrec.2020.105132.

[17] M. Toğaçar, B. Ergen, and Z. Cömert, "Waste classification using AutoEncoder network with integrated feature selection method in convolutional neural network models," *Meas. J. Int. Meas. Confed.*, vol. 153, 2020, doi: 10.1016/j.measurement.2019.107459.

[18] Y. Chu, C. Huang, X. Xie, B. Tan, S. Kamal, and X. Xiong, "Multilayer hybrid deep-learning method for waste classification and recycling," *Comput. Intell. Neurosci.*, vol. 2018, pp. 1–9, 2018, doi: 10.1155/2018/5060857.

[19] M. Cubillos, "Multi-site household waste generation forecasting using a deep learning approach," *Waste Manag.*, vol. 115, pp. 8–14, 2020, doi: 10.1016/j.wasman.2020.06.046.

[20] D. Niu, F. Wu, S. Dai, S. He, and B. Wu, "Detection of long-term effect in forecasting municipal solid waste using a long short-term memory neural network," *J. Clean. Prod.*, vol. 290, pp. 1–8, 2021, doi: 10.1016/j.jclepro.2020.125187.

[21] A. Hussain *et al.*, "Waste Management and Prediction of Air Pollutants Using IoT and Machine Learning Approach," *Energies*, vol. 13, no. 15, p. 3930, Aug. 2020, doi: 10.3390/en13153930.

[22] F. Fasano, A. S. Addante, B. Valenzano, and G. Scannicchio, "Variables Influencing per Capita Production, Separate Collection, and Costs of Municipal Solid Waste in the Apulia Region (Italy): An Experience of Deep Learning," *Int. J. Environ. Res. Public Health*, vol. 18, no. 2, p. 752, Jan. 2021, doi: 10.3390/ijerph18020752.

[23] L. A. Akanbi, A. O. Oyedele, L. O. Oyedele, and R. O. Salami, "Deep learning model for Demolition Waste Prediction in a circular economy," *J. Clean. Prod.*, vol. 274, no. 2020, 2020, doi: 10.1016/j.jclepro.2020.122843.

[24] Z. Kang, J. Yang, G. Li, and Z. Zhang, "An Automatic Garbage Classification System Based on Deep Learning," *IEEE Access*, vol. 8, pp. 140019–140029, 2020, doi: 10.1109/ACCESS.2020.3010496.

[25] M. W. Rahman, R. Islam, A. Hasan, N. I. Bithi, M. M. Hasan, and M. M. Rahman, "Intelligent waste management system using deep learning with IoT," *J. King Saud Univ. - Comput. Inf. Sci.*, no. xxxx, 2020, doi: 10.1016/j.jksuci.2020.08.016.

[26] T. J. Sheng *et al.*, "An Internet of Things Based Smart Waste Management System Using LoRa and Tensorflow Deep Learning Model," *IEEE Access*, vol. 8, pp. 148793–148811, 2020, doi: 10.1109/ACCESS.2020.3016255.

[27] K. M. Bobulski J, "Waste classification system using image processing and convolutional neural networks," *Int. Work. Artif. Neural Networks, 2019 Jun 12, Springer, Cham.*, pp. 350–361, 2019, [Online]. Available: http://dx.doi.org/10.1007/978-3-030-20518-8_30.

[28] S. Jagtap, C. Bhatt, J. Thik, and S. Rahimifard, "Monitoring potato waste in food manufacturing using image processing and internet of things approach," *Sustain.*, vol. 11, no. 11, pp. 1–12, 2019, doi: 10.3390/su11113173.

[29] J. Bai, S. Lian, Z. Liu, K. Wang, and D. Liu, "Deep learning based robot for automatically picking up garbage on the grass," *IEEE Trans. Consum. Electron.*, vol. 64, no. 3, pp. 382–389, 2019.

[30] Z.-H. Zhou, "A brief introduction to weakly supervised learning," *Natl. Sci. Rev.*, vol. 5, pp. 44–53, 2018, doi: 10.1093/nsr/nwx106.

[31] H. C. Shin *et al.*, "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016, doi: 10.1109/TMI.2016.2528162.

[32] M. Lin, Q. Chen, and S. Yan, "Network In Network," pp. 1–10, 2013, doi: 10.1109/ASRU.2015.7404828.

Mohd Anjum is a research scholar at the Department of Computer Engineering, Aligarh Muslim University, India. He has worked as Assistant Professor from 2012 to 2015 at same. His current research interest focuses on Waste Management, the Internet of Things and Machine Learning.

Mohammad Sarosh Umar is a professor at the Department of Computer Engineering, Aligarh Muslim University, India. He has teaching/research and industrial experience of more than 32 years in India and abroad. He has delivered more than 60 invited talks in Information Technology, E-Governance, E-Commerce, Security and Software Engineering. He has various articles and research papers to his credit, and he has presented papers/delivered invited talks in India and abroad. His research interests are User Authentication using Graphical Methods, Computer Security and Software Engineering and he has published three patents.

Dr Sana Shahab is an Assistant Professor in the College of Business Administration, Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia. Her current research focuses on interdisciplinary applications of Statistics, Computer and Management Science to serve the broad areas of problem-solving and decision-making in the organization.