ISSN 2683-345X
journal homepage: http://ijiemjournal.uns.ac.rs/

International Journal of Industrial Engineering and Management

Volume 14 / No 1 / March 2023 / 41 - 50

Original research article

Deep Learning Approach for Volume Estimation in Earthmoving Operation

F. Alam[a], H. S. Ko[a,*], H. F. Lee[a], C. Yuan[b]

[a] Department of Industrial Engineering, Southern Illinois University Edwardsville, Edwardsville, IL, USA;
[b] Department of Construction Management, Southern Illinois University Edwardsville, Edwardsville, IL, USA

ABSTRACT

Earthmoving is a significant activity in most heavy structural designing projects. Earthmoving volume is typically assessed by counting the number of stacked trucks and weighing them on a scale; however, these strategies are error-prone and costly. To address the challenge, this study investigated a deep learning approach for estimating earth volume from photo images of loaded trucks. First, a basic classification model with one convolutional layer has been developed to estimate earth volume by classifying the images into different levels. Next, we applied transfer learning to a pre-trained deep convolutional neural network in order to improve classification performance. For evaluation of the approach, the models have been trained and tested by using images of miniature trucks loaded with different amounts of earth, ranging between 0 and 1000 ml up to six classes at 200 ml intervals. The experimental results showed that the pre-trained network with transfer learning achieved more than 90% accuracy in most cases. The results indicate that the proposed approach has the potential in estimating earth volume in trucks in real-time with minimal intervention by taking images.

ARTICLE INFO

Article history:

Received August 5, 2022
Revised February 25, 2023
Accepted March 1, 2023
Published March 8, 2023

Keywords:
Convolutional Neural Network;
Deep Learning;
Earthmoving Operation;
Transfer Learning;
Volume Estimation

*Corresponding author:
Hoo Sang Ko
hko@siue.edu

1. Introduction

Major civil engineering projects include earthmoving activities, which have long been the focus of research into predicting their output before starting site work. Calculating the quantity of earth transported by dump trucks is necessary for tracking the productivity of earthmoving operations and for financial settlements among earthmoving companies. Counting the number of loaded trucks and weighing loaded trucks on a scale station are two popular approaches for estimating earthmoving volume by vehicles or trucks; however, these procedures are prone to errors, take a long time, and are expensive. Furthermore, owing to intra-class variability in how construction tasks are generally performed and the length of each work step, it is often essential to record many cycles of operations in order to generate a comprehensive analysis of operational efficiency. Traditional time studies are not only time-consuming, but they also need a substantial amount of time spent on manually processing data. Because of the physical limits or biases of the observer, the repetitious data processing process might also deteriorate the quality of the procedure. It is not feasible to interpret the relation between activity duty cycles and productivity, or fuel usage and emissions, without a complete activity analysis. Several techniques have been devised to produce such

http://doi.org/10.24867/IJIEM-2023-1-323

estimations, with varying degrees of effectiveness [1]. Monthly billing data, monthly peak load measurements, the transformer's peak load analysis [2], and existing diversified load curves [3] may all be used to estimate loads. Several strategies have been used to model radial networks, including neural network and fuzzy load estimates [4], sensitivity-based zonal methods [5], [6], and a real-time distribution circuit load modeling strategy based on stochastic load models [7].

With the development of computer vision technology, detection of vehicle types through image processing and pattern recognition has been widely used [8]. The vehicle classification system based on machine vision can be embedded in current traffic cameras. It has many advantages, such as convenient installation, easy maintainability, and small areas of occupation. Besides, the data obtained from the system can be used for research and process for other purposes. With the rapid advancement of the graphics processing unit (GPU), the computing power for image processing has been greatly enhanced, which also in turn brought the fast advancement of deep learning. Compared with traditional feature extraction algorithms, deep learning has a better adaptability and universal applicability. In recent years, the technology of deep learning has been successfully applied to the segmentation, detection, and recognition of objects in images and videos, such as face and pedestrian recognition [9].

This study aims to estimate the earth volume with high accuracy in real time by using photo images of trucks taken from different distances and angles. To achieve the goal, we will develop deep learning models that classify the images into different volume levels. First, we will begin with a basic convolutional neural network (CNN) with one convolutional layer. Next, we will train a complex, pre-trained deep CNN with our images in order to improve the classification performance. Particularly, we will use a popular pre-trained model, VGG16 [10], to obtain a network that can learn complex patterns from the images; transfer learning (TL) will be applied to train the complex classification model with our truck images promptly. For evaluation of the approach, the models will be trained and tested by using images of miniature trucks loaded with different amounts of earth, ranging between 0 and 1000 ml up to six classes at 200 ml intervals. We expect that the deep learning models based on CNN will be effective because of their ability to extract high-level features, such as edges and areas, from the input images that are valuable for accurate prediction. Compared to traditional techniques for earth volume estimation, which require manual work and weighing station, the proposed approach will be cost-effective by automating the process and providing real time feedback.

The rest of this paper is organized as follows. Section 2 presents the research works related to this research. In Section 3, models of volume classification based on CNN with one convolutional layer and a pre-trained deep CNN with transfer learning are proposed. The proposed methods are evaluated and analysed by the experiments in Section 4. Finally, Section 5 draws the conclusion of this research.

2. Literature Review

Various traditional and artificial intelligence (AI)-based models have been presented in literature. A vision-based activity identification framework was designed that focused on excavators and dump trucks working together with the four key modules: equipment tracking, individual equipment action recognition, interaction analysis, and post-processing [11]. The experimental findings confirmed not only the viability of the suggested strategy but also the statistical significance of the interaction analysis. One of the main tools used in earthmoving activities are dump trucks [12]. A detection technique for mining truck loading volume based on deep learning and image recognition had been proposed [13]. The objective was to integrate artificial intelligence technology and image identification technology into the detection of mining truck loading volume. The images were pre-classified using the VGG16 deep neural network model, and the classification results were shown along with the possibility of each category. A machine vision technique to determine the volume of rock mixture was proposed [14]. The highly challenging nickel mineral system was used to show a generic machine vision technique for on-line load determination of rock mixtures. An optimization model that utilized genetic algorithms, linear programming, and geographic information systems for earthmoving activities was presented [15]. In this paper it had been stated that, in order to determine whether a truck was full or empty and how many trucks there were, overall truck number counting (OTC) relied only on human inspection. Since full load was judged by the human eye, the approach might contain inaccuracies that caused contractors to incur a significant loss since the method didn't give a realistic quantity.

Recently, deep learning models, particularly CNN-based models have been studied in various image classification applications. An autonomous CNN architecture was suggested based on evolutionary algorithms for image classification [16]. A new method was suggested for improving a weight-smoothing constraint neural network (WSCNN) and a weighing method for a truck scale based on WSCNN [17]. According to the test results, the weighing mistakes of the truck scale with WSCNN were considerably smaller than those of the neural network-based model of the nonlinear system by implementing the proposed data induction approach. An artificial neural network (ANN) model was developed that used basic predictors to forecast the condition level of earthmoving vehicles [18]. The model's performance was compared to the predictive accuracy of discriminant analysis (DA). The model's predicted accuracy was greater than 94 percent, according to the validation procedure. A computer vision-based algorithm was demonstrated for detecting single actions of earthmoving construction equipment [19]. For a given video collected from various angles, scales, and illuminations, the model distinguished specific motions of the construction equipment. A vehicle classification approach using a pre-trained deep model such as VGG16 was implemented to pre-train the deep model with efficiency-oriented system settings and to increase the chance of using the model in minimum-capacity datasets by optimizing over-fitting limits [20]. A vehicle make and model recognition (VMMR) framework was proposed based on deep feature extraction from VGG16 model, followed by the feature reduction and classification [21]. Deep features were extracted from the image of the vehicle through the layers of VGG16. The suggested algorithm has an automated feature that works for image classification without domain knowledge. A framework was proposed that includes field monitoring systems and sight deep learning for full/empty-load truck categorization [22]. The proposed framework's fundamental model was evaluated for practicality and recognized for model selection recommendations in potential field earthmoving quantity statistics implementation. Deep learning algorithms' major drawbacks include their dependency on a large number of training images and the need for appropriate deep network architecture optimization. These problems were resolved by transfer learning and the VGG architecture was made more suitable for multi-classification problems [23]. A novel technique was provided for layer-wise tuning, and image classification was used to identify the data slices that were the most useful. To reliably distinguish construction equipment, a deep CNN trained by transfer learning was proposed, which involved transferring the knowledge of models learned in other domains with a huge quantity of training data to the construction sector [24]. The fundamentals of CNNs, including a discussion of the many layers that were employed by using traffic sign identification as an example, were explained by [25]. They described the constraints of the general problem and introduced methods and implementation software created to attain the greatest performance on a "labeled" dataset. Transfer learning was implemented to fine-tune the parameters of the pre-trained network (VGG19) for image classification tasks [26]. All these studies show the potential of deep learning models trained by transfer learning in various image classification tasks, which is in line with our research.

The previous studies presented the frameworks which were implemented to classify the full/empty load of earthmoving trucks under a certain scenario. Moreover, in the previous works the images were not pre-processed enough to extract truck information in advance to recognize the load weight of the trucks. In our work, the models were developed to classify multiple classes of different amounts of earth, and they were able to achieve good performance by applying the pretrained deep CNN with transfer learning.

3. Methodology

3.1 Deep Learning - Convolutional Neural Network

Artificial neural networks are algorithms inspired by the structure and function of the brain. These neural networks attempt to simulate the actions of the human brain-albeit far from matching its ability-allowing it to "learn" from large amounts of data. While a neural network with a single layer can still make approximate predictions, additional hidden layers can help to optimize and refine for accuracy. It drives many artificial intelligence (AI) applications and services that improve automation, performing analytical and physical tasks without human intervention. The methods are representation-learning methods with multiple levels of representation, obtained by composing simple but non-linear modules, each of which transforms the representation at one level (starting with the raw input) into a representation at a higher, slightly more abstract level.

Deep learning allows computational models that are composed of multiple processing layers in depth to learn representations of data with multiple levels of abstraction. Deep learning is a subset of machine learning, which is essentially a neural network with three or more layers. Deep learning eliminates some of the data pre-processing that is typically involved with machine learning. This algorithm can ingest and process unstructured data, like text and images, and it automates feature extraction, removing some of the dependency on human experts.

The proposed deep learning model is developed in six main steps as shown in Figure 1: collecting data, choosing the model, building the model, training the model, evaluating the model, and making prediction.

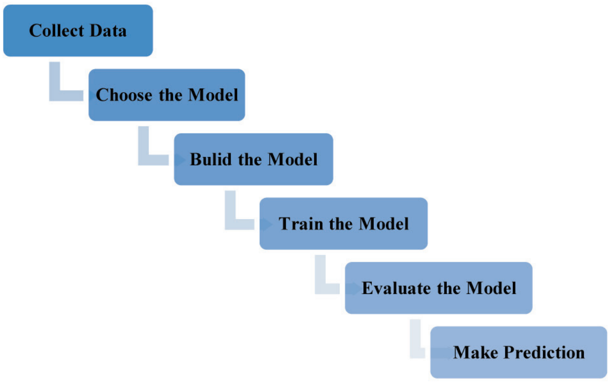

Figure 1. Deep Learning Methodology

3.1.1 Collecting Data

We included numerous images of the same load that were captured from diverse views and angles which were of different shapes and distributions in the body. Some images had earth in cone shapes while others did in flat, inclined or irregular shapes. For developing the models, images were obtained from a miniature truck which has the identical features as a real scaled dump truck. The miniature truck was loaded with different amounts of earth, ranging between 0 and 1000 ml, up to six classes at 200 ml intervals shown in Table 1. To make the experiment more realistic, photos were taken at different places and angles at different times of the day, so the different lights and backgrounds would be used within the dataset. Two types of fill materials were used: sand (store-bought) and soil (from the backyard). However, the light was set to the natural sunlight, and the background and fill material colours were always set to the same tone. Images were collected at six different load levels, approximately 400 images for each. The images of each class have been illustrated in Figure 2. With the camera held in the landscape orientation for each image taken, the angle and the height were not fixed as long as all borders of the dump truck were within the camera image.

Table 1. Number of Images for Each Case

Volume	Number of Images
200 ml	437
400 ml	478
600 ml	395
800 ml	373
1000 ml	393
Empty	441

3.1.2 Choosing the Model

3.1.2.1 A CNN with One Convolutional Layer

First, for the experiment, a CNN model with one convolutional layer was chosen, which is the smallest unit of a deep neural network. In this model, seven layers are used to create a simple CNN as depicted in Figure 3 below. The details of each layer are as follows. Image input layer is where the image size is specified. For this model, the image size is specified as 224x224x3. Convolutional 2D layer is where the filter size and number of filters are specified. In this model, the filter size is specified as 3x3 with the number of filters as 64. ReLU layer, also known as the rectified linear unit, is one of the most common activation functions. Max pooling 2D layer is a down-sampling operation that reduces the three-dimensional size of the feature map and removes three-dimensional information noises. Down-sampling enables the network to increase the number of filters in deeper convolutional layers without increasing the

Figure 2. Images of dump trucks for 200ml, 400ml, 600ml, 800ml, 1000ml, and empty loads

required number of computations per layer. In this model, the size of the rectangular region is [2,2] with a stride of 2. Fully connected layer combines all the features learned by the previous layers across the image to identify the larger patterns. The output size, which is equal to the number of classes, is specified in this layer. The number of classes varies between 2 and 6 in this study. Softmax layer is chosen as an activation layer that normalizes the output of the fully connected layer. Classification layer is the final layer that uses the probabilities collected from the softmax activation function for each input to allocate the input to one of the similarly exclusive classes and compute the loss.

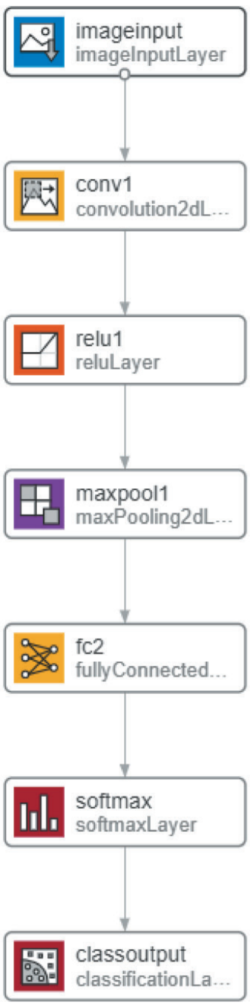

Figure 3. A CNN with One Convolutional Layer

3.1.2.2 A Pre-trained Network with Transfer Learning

Liu et al. [21] has reported that the pre-trained network named VGG16 worked fast in terms of training and testing/validation time while it also had a high testing/validation accuracy in determining whether a truck is empty or fully loaded. Since the result was promising, VGG16 is selected as the pre-trained network in this study. VGG16 is a deep CNN model that has been developed for the ImageNet competition. It was pre-trained with the large-scale ImageNet dataset. In order to apply it to estimation of earthmoving, transfer learning (TL) is required to extract features from the truck load images and fine-tune the pre-trained network for classification of the truck load images. TL is a machine learning method to train an existing model, which has been trained for another classification problem, to learn for a new classification problem through the transfer of knowledge that has already been learned from the old problem.

VGG16 is composed of 16 layers with learnable weights but 41 layers in total. Figure 4 shows the overall architecture of the VGG16.

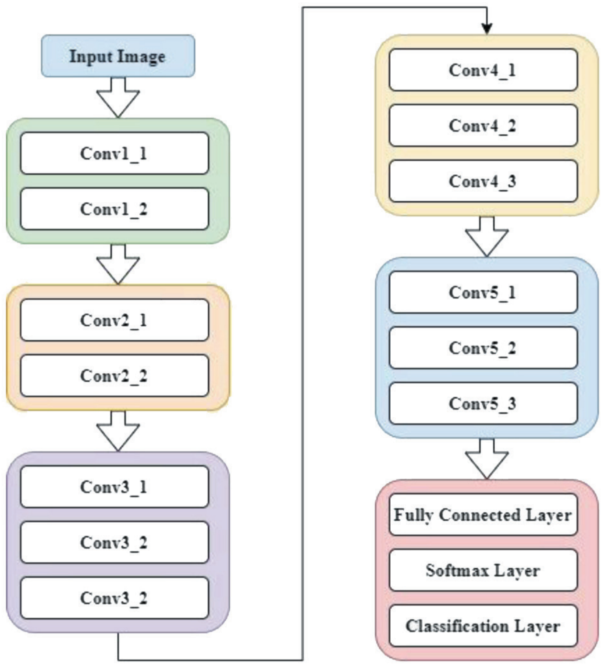

Figure 4. VGG-16 Model Architecture

3.1.3 Building the Model

The deep learning model with transfer learning is built based on a pre-trained network architecture (VGG16) with layer replacement. The models are developed in the following steps as shown in Figure 5: loading and exploring image data, specifying or partitioning the dataset into training and validation sets, defining network architecture, specifying training options, training the network, classifying validation images and computing the accuracy.

The architecture begins with the image input layer, followed by five convolutional groups and three fully connected layer groups, and ends with the classification layer as the output. The details of each layer group are as follows:

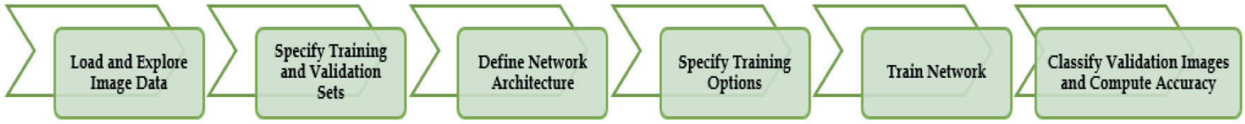

Figure 5. Model Building Procedure

The image input layer is where the image size is specified. For this network, the image size is specified as 224x224. The first two convolutional groups include two convolutional 2D layers each followed by a ReLU layer, while the following three convolutional groups include three convolutional 2D layers each followed by a ReLU layer. Convolutional groups are separated from each other by max pooling layers. The convolutional 2D layer is the layer where the filter size and the number of filters are specified. In this network, the filter size is set to 3x3 while the number of filters is specified as 64, 128, 256, 512, and 512, respectively, for each of the five convolutional groups. For the max pooling 2D layer, the size of the rectangular region is [2,2] with a stride of 2.

The first two fully connected layer groups include one fully connected layer, each followed by a ReLU layer and a dropout layer, while the last fully connected group includes one fully connected layer followed by a softmax layer. The dropout layer randomly sets input elements to zero with a given probability. In this network, both dropout layers have a dropout probability of 0.5. In the output layer, the number of output classes of the network varies from 2 to 6.

3.1.4 Training the Model

The model is trained to predict the true class among the 6 classes of the earth volume in the dump truck. To train the network for this classification task, the stochastic gradient descent with momentum method was selected along with a set of training parameters, including learning rate, validation frequency, and mini batch size, obtained by a grid search. From the original dataset, 75% of the data are used for training and the remaining 25% of the data for validation.

3.1.5 Evaluating the Model

Overfitting occurs when a statistical model or machine learning algorithm captures the noise of the data. Intuitively, it happens when the model or the algorithm fits the data too well. Overfitting results in a model with a high accuracy for the training dataset but poor results on new datasets. Such a model is not of any use in the real world as it is not able to predict outcomes for new cases, particularly in cases where it is applied. When it comes to evaluating a model, accuracy and other classification performance measures on a separate dataset must be considered for validating and assessing the effectiveness of the model. In this study, not only accuracy but also precision, recall, and F1 score will be used as performance measures.

3.1.6 Making Prediction

Once the training is completed, the trained network can be used to predict or estimate the classified volume for input images from a camera. The prediction performance on a new dataset indicates a better approximation of how the model will perform in the real world. The models were tested by using a new set of images of the six classes.

4. Experimental Results

Images for training have been collected from the six classes as shown in Table 1. In pre-processing the data, the collected images were scaled and resized to fit the requirements of the pre-trained network model. All images were cropped to a square, 1700x1700 pixels and then resized 224x224 as required by VGG16. For training, learning rate, validation frequency, and mini batch size were set to 0.0001, 10, and 30, respectively.

4.1 Experimental Results from CNN with One Convolutional Layer

The simple CNN model for volume classification was tested in five scenarios, from 2 classes (i.e., 1000 ml vs. empty) to 6 classes where the volume changed from 0 to 1000 ml at 200 ml intervals. Table 2 shows the results obtained for the five scenarios, including the classification accuracy and training time. The table also contains training parameters obtained by a grid search, including validation frequency, maximum number of epochs for training, and mini batch size.

The experimental results show the accuracy of the model was 95.78% for 2 classes, 30.83% for 3 classes, 23.49% for 4 classes, 18.27% for 5 classes, and 15.57% for 6 classes. The results show that the simple CNN is not capable of learning patterns in complex scenarios with more than 2 classes, which calls for more complex networks.

4.2 Experimental Results from Pre-trained Network with Transfer Learning

The pre-trained VGG16 with transfer learning for volume classification was tested in the same five scenarios, from 2 classes to 6 classes. Table 3 shows the results along with the training parameters obtained by

Table 2. Experimental Results from CNN with One Convolutional Layer

Scenario	Classes	Number of Images	Validation Frequency	Max Epochs	Minibatch Size	Train Validation	Accuracy	Training Time
2 Classes	1000ml Empty	314 352	5	10	10	75%-25%	95.78%	3 min 55 sec
3 Classes	200ml 600ml 10000ml	348 316 314	5	10	10	75%-25%	30.83%	10 min 52 sec
4 Classes	200ml 600ml 1000ml Empty	348 316 314 352	5	10	10	75%-25%	23.49%	13 min 8 sec
5 Classes	200ml 400ml 600ml 1000ml Empty	348 382 316 314 352	5	10	10	75%-25%	18.27%	19 min 28 sec
6 Classes	200ml 400ml 600ml 800ml 1000ml Empty	348 382 316 298 314 352	5	10	10	75%-25%	15.57%	28 min 33 sec

Table 3. Experimental Results from Pre-Trained Network Model

Scenario	Classes	Number of Images	Validation Frequency	Max Epochs	Minibatch Size	Train Validation	Accuracy	Training Time
2 Classes	1000ml Empty	314 352	10	30	30	75%-25%	100%	13 min 48 sec
3 Classes	200ml 600ml 10000ml	348 316 314	10	30	30	75%-25%	99.18%	13 min 1 sec
4 Classes	200ml 600ml 1000ml Empty	348 316 314 352	10	30	30	75%-25%	97.89%	56 min 37 sec
5 Classes	200ml 400ml 600ml 1000ml Empty	348 382 316 314 352	10	30	30	75%-25%	91.80%	50 min 41 sec
6 Classes	200ml 400ml 600ml 800ml 1000ml Empty	348 382 316 298 314 352	10	30	30	75%-25%	88.38%	85 min 2 sec

a grid search. The experimental results show the accuracy of the model was 100% for 2 classes, 99.18% for 3 classes, 97.89% for 4 classes, 91.80% for 5 classes, and 88% for 6 classes. The results show that the pre-trained network is capable of distinguishing between the six classes. Although training time for the pre-trained network is much longer than that for the simple CNN, the times are reasonable and shortened significantly by transfer learning.

After training is completed, the trained network was tested with additional datasets. Table 4 shows the result of this deployment test, which is very promising with an error rate less than 5% for 6 classes scenarios. In addition, root mean squared error (RMSE) and normalized RMSE (NRMSE) of each scenario are presented in Table 4 to show the estimated numerical error.

In addition, the classification performance of the pre-trained network with transfer learning was evaluated by precision, recall, and F1 score, which are defined as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (1)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (2)$$

$$\text{F1 score} = \frac{2*\text{Recall}*\text{Precision}}{\text{Recall+Precision}} \quad (3)$$

where TP = true positive, TN = true negative, FP = false positive, and FN = false negative. All the classes against each other for each case have been compared, resulting in the pairwise comparisons, using the three metrics. The results for each case are presented in Table 5.

To sum up, the experimental results showed that the pre-trained CNN with TL was able to recognize all 2 classes truckload in the dataset with 100% of ac-

Table 4. Deployment Test Results by Pre-Trained Model

Scenario	Classes	Number of Images	Number of Misclassified Images	Error Rate	RMSE	NRMSE
2 Classes	1000ml	79	0	0%	0	0
	Empty	89	0			
3 Classes	200ml	89	1	2.02%	56.91	0.071
	600ml	79	4			
	1000ml	79	0			
4 Classes	200ml	89	9	4.76%	66.37	0.066
	600ml	79	6			
	1000ml	79	1			
	Empty	89	0			
5 Classes	200ml	89	5	3.44%	40.82	0.041
	400ml	96	8			
	600ml	79	2			
	1000ml	79	0			
	Empty	89	0			
6 Classes	200ml	89	2	4.96%	49.45	0.049
	400ml	96	8			
	600ml	79	11			
	800ml	75	2			
	1000ml	79	2			
	Empty	89	0			

Table 5. Additional Deployment Test Results by Pre-Trained Model

Scenario	Recall	Precision	F1-Score
2 Classes	100%	100%	100%
3 Classes	98.01%	98.13%	98.04%
4 Classes	95.45%	95.47%	95.35%
5 Classes	96.59%	96.70%	96.60%
6 Classes	94.92%	95.05%	94.90%

curacy and above 90% of accuracy for the cases of up to five classes. For the case of 6 classes, the CNN performed reasonably with an accuracy close to 90%. One thing to note is that the number of iterations in training is directly proportional to the number of images and epochs. As the number of iterations increases, so does the time required for the system to complete the training. In addition, the training time usually took longer as the number of classes increased as shown in Table 3 e.g., 13 min 48 sec for two classes and 56 min 37 sec for four classes. In our experiments, however, the training time for the most complex scenario (six classes) took 85 min 2 sec (in a machine with Intel(R) Core(TM) i9-9900K CPU @3.60GHz), 32 GB RAM, and an NVIDIA GeForce RTX 2080 GPU with 8 GB) and the testing time was only 2 sec, which shows that the proposed model can be trained in a reasonable time and deployed for earth volume estimation in real-time.

5. Conclusion

On construction sites, the present techniques for estimating the earth volume in dump trucks may be expensive or incorrect, while taking up a significant amount of time and effort. This research aimed to overcome these problems with two different methods: a simple CNN with one convolutional layer and a pre-trained deep CNN with transfer learning. The CNNs classified the amount of the load in a dump truck via photo images by extracting informative features from the images.

As mentioned earlier, numerous images of the same load with different shapes and distributions were used for the experiments where some images showed the earth in cone shapes while others in flat, inclined, or irregular shapes. The network trained by many such images and ground-truth labels should be able to obtain features that help to estimate the (3D) load from the 2D photo images. The results clearly showed that the pre-trained CNN with transfer learning performs significantly better than the simple CNN. The CNN with one convolutional layer worked well only for the two classes scenario. However, it performed poorly for the rest of the scenarios with more than two classes. The pre-trained deep CNN with transfer learning showed promising results for every scenario that could properly estimate earth volume in trucks without the need for a scale, addressing the drawbacks of present approaches. This could minimize the expense or inaccuracy associated with earthmoving volume, as well as optimize the overall number of trucks utilized in earthmoving operations, particularly for large construction sites. With this model, the validation accuracy was more than 90% for up to 5 classes scenario and 88% for 6 classes scenario. Additionally, the deployment test results showed an error rate less than 5% in every scenario. The experiments also showed that the number of images and epochs is directly proportional to the number of iterations. Additionally, as the number of iterations increased, so did the amount of time required for the system to finish the training. Another finding was that if the volume difference between the classes is reduced, classification would become more challenging and hence take longer. Although the experiments had been conducted with insufficient images to train such a network, our network was validated by the validation results and it was possible due to using the pretrained network (VGG). The performance could be further improved by training with more images with various shapes and views.

We will train the proposed deep CNN by using more labelled-image datasets from the actual construction sites to refine the network model to obtain better test results. The model will be deployed on the construction sites and utilize images captured by a camera to estimate earth volume in real time. To achieve the goal, the proposed model would need to be trained and tested with actual images that contain not only the earth but also debris of different types of materials such as rocks, concrete, gravel, etc., which would make earth volume estimation more accurate yet challenging. Also in the future, we can implement Regression in Edge Impulse to train our model. Regression models may be trained with the images and show a predicted weight since they can learn to take in any sort of input and return a numeric output.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] B. Mohammadi, Y. Guan, R. Moazenzadeh and M.J.S. Safari, “Implementation of hybrid particle swarm optimization-differential evolution algorithms coupled with multi-layer perceptron for suspended sediment load estimation,” Catena, vol. 198, 105024, 2021, doi: 10.1016/j.catena.2020.105024.

[2] V. Borozan and N. Rajakovic, "Minimum loss distribution network configuration: analyses and management," in 14th

International Conference and Exhibition on Electricity Distribution. Part 1. Contributions (IEE Conf. Publ. No. 438), Birmingham, UK, 1997, pp. 18/1-18/5 vol.6, doi: 10.1049/cp:19970631.

[3] R. P. Broadwater, A. H. Khan, H. E. Shaalan and R. E. Lee, "Time varying load analysis to reduce distribution losses through reconfiguration," in IEEE Transactions on Power Delivery, vol. 8, no. 1, pp. 294-300, Jan. 1993, doi: 10.1109/61.180349.

[4] M. Wang, "Manufacturing capacity evaluation of smart job-shop based on neural network," International Journal of Simulation Modelling, vol. 20, no. 4, pp. 778 - 789, 2021, doi: 10.2507/IJSIMM20-4-CO19.

[5] J. Wan and K. N. Miu, "Load estimation in radial electric power distribution networks using limited measurements," in IEEE International Symposium on Circuits and Systems: Emerging Technologies for the 21st Century, Geneva, Switzerland, 2000, pp. 517–528.

[6] J. Wan and M. K. Nan, "A Zonal Load Estimation Method for Unbalanced, Radial Distribution Networks," IEEE Power Engineering Review, vol. 22, no. 6, 58, 2022, doi: 10.1109/MPER.2002.4312293.

[7] A. K. Ghosh, D. L. Lubkeman and R. H. Jones, "Load modeling for distribution circuit state estimation," IEEE Transactions on Power Delivery, vol. 12, no. 2, pp. 999–1005, 1997, doi: 10.1109/61.584427.

[8] B. Coifman, D. Beymer, P. McLauchlan and J. Malik, "A real-time computer vision system for vehicle tracking and traffic surveillance," Transportation Research Part C: Emerging Technologies, vol. 6, no. 4, pp. 271–288, 1998, doi: 10.1016/S0968-090X(98)00019-9.

[9] X. Wang, W. Zhang, X. Wu, L. Xiao, Y. Qian and Z. Fang, "Real-time vehicle type classification with deep convolutional neural networks," Journal of Real-Time Image Processing, vol. 16, no. 1, pp. 5–14, 2019, doi: 10.1007/s11554-017-0712-5.

[10] Z. Omiotek and A. Kotyra, "Flame image processing and classification using a pre-trained vgg16 model in combustion diagnosis," Sensors (Switzerland), vol. 21, no. 2, pp. 1–15, 2021, doi: doi.org/10.3390/s21020500.

[11] J. Kim, S. Chi and J. Seo, "Interaction analysis for vision-based activity identification of earthmoving excavators and dump trucks," Automation in Construction, vol. 87, pp. 297–308, 2018, doi: 10.1016/j.autcon.2017.12.016.

[12] A. Jabri and T. Zayed, "Agent-based modeling and simulation of earthmoving operations," Automation in Construction, vol. 81, pp. 210–223, 2017, doi: 10.1016/j.autcon.2017.06.017.

[13] L. Bi, Y. L. Li and Z. H. Guo, "Study on the estimation of ore loading quantity of truck based on deep convolutional neural network," Gold Science and Technology, vol. 27, no. 1, pp. 112-120, 2019, doi: 10.11872/j.issn.1005-2518.2019.01.112.

[14] J. Tessier, C. Duchesne and G. Bartolacci, "A machine vision approach to on-line estimation of run-of-mine ore composition on conveyor belts," Minerals Engineering, vol. 20, no. 12, pp. 1129-1144, 2007, doi: 10.1016/j.mineng.2007.04.009.

[15] O. Moselhi and A. Alshibani, "Optimization of earthmoving operations in heavy civil engineering projects," Journal of Construction Engineering and Management, vol. 135, no. 10, pp. 948-954, 2009, doi: 10.1061/(asce)0733-9364(2009)135:10(948).

[16] Y. Sun, B. Xue, M. Zhang, G. G. Yen and J. Lv, "Automatically Designing CNN Architectures Using the Genetic Algorithm for Image Classification. IEEE Transactions on Cybernetics, vol. 50, no. 9, pp. 3840–3854, 2020, doi: 10.1109/TCYB.2020.2983860.

[17] H. Lin, H. Xiang, L.Wang and J. Yang, "Weighing method for truck scale based on neural network with weight-smoothing constraint," Measurement, vol. 106, pp. 128–136, 2017, doi: 10.1016/j.measurement.2017.02.038.

[18] M. Marinelli, S. Lambropoulos and K. Petroutsatou, "Earthmoving trucks condition level prediction using neural networks," Journal of Quality in Maintenance Engineering, vol. 20, no. 2, pp. 182–192, 2014, doi: 10.1108/JQME-09-2012-0031.

[19] M. Golparvar-Fard, A. Heydarian and J.C. Niebles, "Vision-based action recognition of earthmoving equipment using spatio-temporal features and support vector machine classifiers," Advanced Engineering Informatics, vol. 27, no. 4, pp. 652–663, 2013, doi: 10.1016/j.aei.2013.09.001.

[20] V. Sowmya and R. Radha, "Efficiency-Optimized Approach-Vehicle Classification Features Transfer Learning and Data Augmentation Utilizing Deep Convolutional Neural Networks," International Journal of Applied Engineering Research, vol. 15, no. 4, pp. 372–376, 2020.

[21] S. Naseer, S. M. A. Shah, S. Aziz, M. U. Khan and K. Iqtidar, "Vehicle Make and Model Recognition using Deep Transfer Learning and Support Vector Machines," in 2020 IEEE 23rd International Multitopic Conference (INMIC), Bahawalpur, Pakistan, 2020, pp. 1-6, doi: 10.1109/INMIC50486.2020.9318063.

[22] Q. Liu, C. Feng, Z. Song, J. Louis and J. Zhou, “Deep Learning Model Comparison for Vision-Based Classification of Full/Empty-Load Trucks in Earthmoving Operations,” Applied Sciences, vol. 9, no. 22, p. 4871, Nov. 2019, doi: 10.3390/app9224871.

[23] N. M. Khan, N. Abraham and M. Hon, "Transfer Learning With Intelligent Training Data Selection for Prediction of Alzheimer’s Disease," IEEE Access, vol. 7, pp. 72726-72735, 2019, doi: 10.1109/ACCESS.2019.2920448.

[24] H. Kim, H. Kim, Y. W. Hong and H. Byun, "Detecting Construction Equipment Using a Region-Based Fully Convolutional Network and Transfer Learning," Journal of Computing in Civil Engineering, vol. 32, no. 2, 04017082, 2018, doi: 10.1061/(ASCE)CP.1943-5487.0000731.

[25] L. Pigou, S. Dieleman, P.J. Kindermans and B. Schrauwen, "Sign Language Recognition Using Convolutional Neural Networks," In: Agapito, L., Bronstein, M., Rother, C. (eds) Computer Vision - ECCV 2014 Workshops. ECCV 2014. Lecture Notes in Computer Science, 2015, vol 8925, doi: 10.1007/978-3-319-16178-5_40.

[26] M. Shaha and M. Pawar, "Transfer Learning for Image Classification," in 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2018, pp. 656-660, doi: 10.1109/ICECA.2018.8474802.