EPiC Series in Built Environment

Volume 3, 2022, Pages 380–388

ASC2022. 58th Annual Associated Schools of Construction International Conference

# Single Image based Volume Estimation for Dump Trucks in Earthmoving using Machine Learning Approach

**Mercan Deniz, Master, Faria Alam, Master, Chenxi Yuan, Ph.D., Hoo Sang Ko, Ph.D. and H. Felix Lee, Ph.D.**
Southern Illinois University Edwardsville
Edwardsville, Illinois

Earthmoving is one of the key activities in most heavy civil construction projects. The dump truck is one primary construction vehicle for earthmoving. Two popular approaches are currently used to estimate earthmoving volume by trucks, i.e., manually counting the number of loaded trucks and weighing loaded trucks on a scale station. Considering both methods are either error-prone, time-consuming, or costly, this study aims to estimate different earth volumes in dump trucks from a single image using the machine learning approach. By establishing a pre-trained deep learning neural network from 3663 images with sixteen different volumes of the earth using a scaled dump truck model, the proposed approach is tested to estimate the truckload in a quantitative manner in real-time. Another 1221 images are used for verification in six case combinations out of the sixteen different volumes. The preliminary results show that the classification accuracy by using the pre-trained network is 100% if the volume gap between adjacent classes is more than 5%, while 76.67% if the volume gap is 1%. The preliminary test results show a great potential that the proposed methods could be applied to the field and provide a fast and accurate estimate of truckload with minimal cost.

**Key Words:** Machine learning, Image-based volume estimation, Earthmoving, Productivity

## Introduction

Earthmoving is a base operation in field construction. Dump trucks are a major means used for earthmoving operations (Jabri & Zayed, 2017). Determining the amount of load moved by dump trucks is required to track the earthmoving operations' productivity and is used for financial settlements within the earthmoving contractors. There are currently two types of earthmoving quantity statistics methods, i.e., manually counting the overall number of truckloads moved (i.e., Overall Truck Number Counting (OTC)) or weighing trucks at load weigh stations (Single Truck Load Weighing Stations (STW)). OTC can be used for scenarios like airport constructions which have a large quantity of overall transportation that is made with low unit load price and uniform trucks (Moselhi & Alshibani, 2009). It is based on only checking by humans whether the truck is full or empty and the number of trucks. Since full load is determined by the human eye, there can be minor errors in this method which can lead to a major loss for contractors because the method does not provide a realistic quantity. Different quantities (usually from 100% to 60% loaded trucks) are counted as fully loaded trucks. This method comes with inevitable human error and high labor cost. Overall Truck Number Counting is limited under conditions like unfamiliar altitude and extreme weather, which are threatened labor safety (Guo et al., 2016; Yi & Chan, 2017). On the other hand, STW can provide more accurate statistics about the amount of truckload (Lin et al., 2017). However, it is applicable for cases with a small amount of total transportation in general (Lee & Chow, 2011). It can also be useful for cases that have multi-party contracting, a small amount of earthmoving, and/or expensive highway transportation charges (Fekpe et al., 1993). This method, however, has disadvantages. In addition to the land and facility necessary for STW, a single truck scale costs between $35,000 to $100,000 along with a maintenance cost in the following period (Carlton Scale, 2021). It needs to be replaced after fulfilling its lifetime which is not so long to compensate for the cost. Scales are placed in stations on highways which require trucks to get in line first to be weighted and then stop on the scale. Therefore, these stations can cause a bottleneck which ends up with traffic jams on busy highways, especially during peak times (Lee & Chow, 2011). These bottlenecks in stations also cause delays in the truck flow in each project and affect the cost efficiency of all projects which use these stations for weighing their trucks (Samandar et al., 2018). Therefore, the current implementations have their limitations, such as high labor and economic costs, limited application environment, continuous maintenance requirement, and transportation interruption (Liu et al., 2019).

A comparison study by Liu et al. (2019) proposes the framework of the novel, automated earthmoving quantity statistics that mainly applies vision-based deep learning for full/empty-loaded truck classification as the core work and counts full-load trucks. It utilizes the field-equipped surveillance video system and deep learning convolutional neural network (CNN) related image recognition models to achieve unmanned and non-contact truckload condition judgment. The study compares 12 deep learning models constructed by four classical CNNs and two transfer learning methods to classify empty and fully loaded trucks. Isolated projects in an open construction site with uniform and uncovered trucks are required for the vision-based earthmoving quantity statistics method proposed in their study. Although this study successfully validated the feasibility in use of deep learning approach for vision-based full/empty-loaded truck classification, it is still a binary classification problem, which solves the problem of Yes and No question. This study steps further to evaluate the possibility in use of the deep learning approach for quantitatively estimating the earth volume in uniform and uncovered dump trucks. By addressing this research question, the associated cost or error in earthmoving volume can be reduced and the total number of trucks used in earthmoving projects can be optimized, especially for large construction areas. Besides, if the use of the deep learning approach in earthmoving operations can be validated to a quantitative level, the value of this research can be extended to various relevant application scenarios, which includes but not limited to roads, airports, ports, buildings, utilities, etc.

## Methodology and Experiments

This study aims to estimate different earth volumes in dump trucks from a single image using the machine learning approach. By establishing a pre-trained deep learning neural network from images with different volumes of the earth in the dump trucks, the proposed approach is designed to estimate the truckload in a quantitative manner. Deep learning is a subfield of machine learning. It is based on artificial or convolutional neural networks, inspired by the human brain. It allows computers and/or machines to solve problems by learning large data sets. These data sets can contain images, text, or sound (LeCun et al., 2015). The proposed methodology consists of four main parts as shown in the Figure below: data collection, code preparation, training process (including pre-processing, training, and testing), and comparison (see Figure 1). The details will be explained in the following of this section.

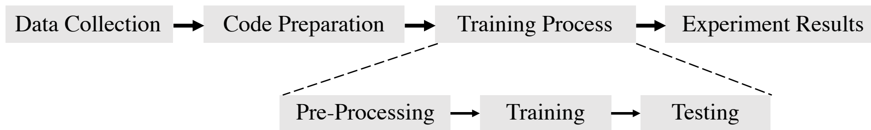


Figure 1. The overall workflow

## *Data Collection*

The data collection was obtained from a toy truck which has the identical features as a real scaled dump truck. To make the experiment more realistic, photos were taken at different places and different times of the day, so the different lights and backgrounds would be used within the data set. Two types of fill materials were used: play sand from Home Depot and the soil from the backyard. The light was set to the natural sunlight, and the background and fill material colors were set to the same tone. For twelve different load ranges, photo images were collected, which includes the empty load case (441 photos), 200ml load case (437 photos), 400ml case (478 photos), 600ml load case (395 photos), 610ml load case (184 photos), 620ml load case (207 photos), 630ml load case (200 photos), 640ml load case (231 photos), 650ml load case (225 photos), 700ml load case (200 photos), 750ml load case (276 photos), 800 ml load case (373 photos), 850 ml load case (276 photos), 900 ml load case (303 photos), 950 ml load case (265 photos) and 1000 ml load case (393 photos). With the camera held in the landscape orientation for each image taken, the angle and the height were not fixed but all borders of the dump truck were required to be visible in the photo images. The original image is 3456x2304. Figure 2 shows sample images with different earth volumes.

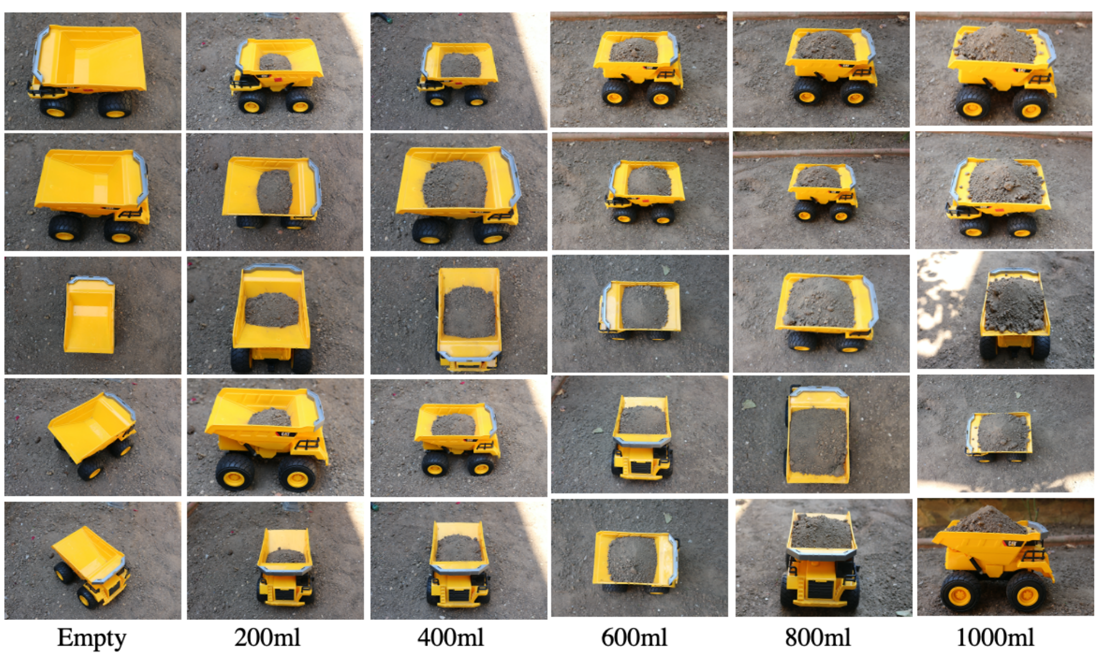


Figure 1. Sample images with different earth volumes

## *Code Creation in Matlab*

In the code creation process in Matlab (MathWorks, 2021), a pre-trained network architecture with layer replacement is used to conduct a convolutional deep learning process, which is divided into the following six main parts: load and explore image data, define network architecture, specify training options, train network, and classify validation images and compute the accuracy (see Figure 3).

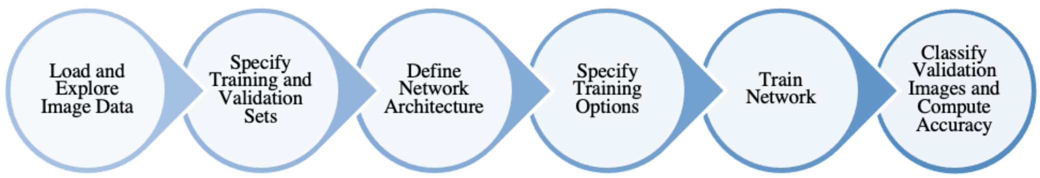


**Figure 2:** The general structure of network models

### *Loading and exploring the image data*

The sample data are loaded as an image data store by specifying the data set path and using the "*imageDatastore" Matlab* command. The image datastore enables the network to store large image data and efficiently read batches of images from folders during the training of the convolutional neural network.

### *Specifying training and validation sets*

The data are divided into two different sets, i.e., a training set and a validation set. While 75% of the total number of images are used as a training set, the remaining 25% are used as a validation set.

### *Defining the network architecture*

Liu et al. (2019) found that the pre-trained network named VGG16 is working fast in terms of training and testing/validation time while it also has a high testing/validation accuracy to determine if a truck is either empty loaded or fully loaded. Since the results of the paper are promising VGG16 is selected to be used in this experiment. VGG16 is a convolutional neural network model proposed by Simonyan and Zisserman (2014), which has 16 layers. For this experiment, only the architecture of the network model is used, which has 41 layers (see Figure 4). The layers are started with the image input layer and followed by five convolutional groups, three fully connected layer groups, ending with the classification layer as the output (MathWorks, 2021). The details of each layer group are as follows.

Image Input Layer is where the image size is specified. The image size in this study is specified as 224 x 224. The first two convolutional group includes two convolutional 2D layers, and each is followed by a relu layer. The following three convolutional groups include three convolutional 2D layers and each is followed by a relu layer. Convolutional groups are divided from each other by max-pooling layers. Convolutional 2D Layer is where the filter size and number of filters are specified. In this study, the filter size is specified as 3 x 3 with the number of filters as 64, 128, 256, 512, and 512 respectively (see Figure 4). For max-pooling 2D layer, the size of the rectangular region is [2, 2], while the name-value pair argument is specified as "Stride". The first two fully-connected layer groups include one fully connected layer each followed by a relu layer and a drop out 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 approach, both dropout layers have a probability of “0.5”. For each experiment case, the number of output classes of the network is changed based on the experiment requirement.

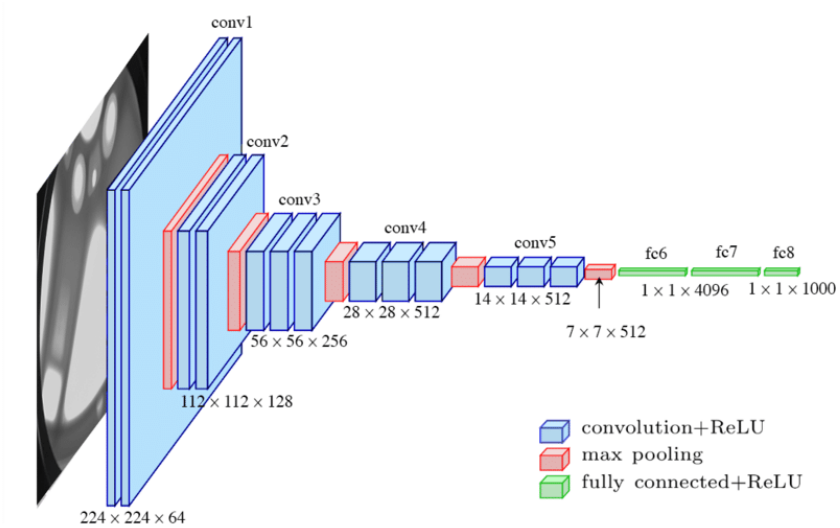


**Figure 3:** Pretrained network architecture

The training options that have been used in this approach are listed as followed. The training options “Solver Name”, “Plots”, “Max Epochs”, “Mini Batch Size”, “Shuffle”, “Validation Data”, “Validation Frequency”, and “Initial Learning Rate” are specified as “sgdm” (stochastic gradient descent with momentum), “training-progress”, “100”, “every-epoch”, “imds” (image data store), “5”, and “0.0001” respectively. The rest of the training options are set as default values.

### *Specifying the training options*

All options have been specified except the preferred default settings. Training options include “Solver Name”, “Plots and Display”, “Mini-Batch Options”, “Validation”, “Solver Options”, “Gradient Clipping”, “Sequence Options”, “Hardware Options”, and “Check Points”. The specified options are discussed in the following subsections.

### *Training the network*

Specified training data set, network architecture layers and training options are used to train the network.

### *Classifying the validation images and computing the accuracy*

The validation data set is classified by using the trained network. Accuracy is the percentage of labels that the network correctly predicts.

An example of the Matlab code can be found from the shared link below:
https://www.dropbox.com/sh/jfoit87ahjju068/AAA8TiBLI1GQVePT6lMQNHrCa?dl=0

## *Training Process*

### *Pre-processing*

In this process, there are three main steps which are pre-processing, training, and validation. While the pre-processing step is done by the XnView App, training and validation steps are performed in MatLab. The collected image data are resized according to the requirements of the pre-trained network model. The original size of all images is 3456 x 2304. For the VGG16 convolutional neural network model, the input required is the size of 224 x 224 so all the images are resized to 224 x 224 for this study.

### *Training and Testing*

The training command is given as “net= trainNetwork (imds, layers, options)” in the network. The command means that the network is using 75% of the “imds” data store with the specified layers to train the network with specified training options. During the training process, a plot is progressing along with the training.

For testing, the network is using the remaining 25% of the “imds” data store with the “YPred = classify (net, imdsTest)” command. The accuracy of the network is also calculated by the validation results. The accuracy calculation is made by the comparison between the network’s predicted label and the original label of the image. The validation is also made with a stand-alone function (pre-trained network) that is created with the training network. In this process, labeled images are randomly given to the function. The system predicts the labels, and the results are compared for the accuracy calculation. The original volume and predicted volume comparison can be seen in the experiment results section.

## *Experimental Results*

The proposed Deep Learning method is applied to six different case combinations. The accuracy of the stand-alone (pre-trained) function is 100% with the usage of the network model which was created for nine cases (50 ml apart) and 76.67% for 10 ml apart cases. For the 50 ml apart cases test, 18 random and new (non-added in the training process) images are given to the system to predict the classes. For the 10 ml apart cases test, 30 random and new (non-added in the training process) images are given to the system to predict the classes. The pre-trained network model can detect every random image with a 5% difference of the total volume correctly but not the 1% difference of the total volume. The detailed information and the validation accuracy results of each case are shown in Table 1 below. The training progress plots can be found in Figure 5.

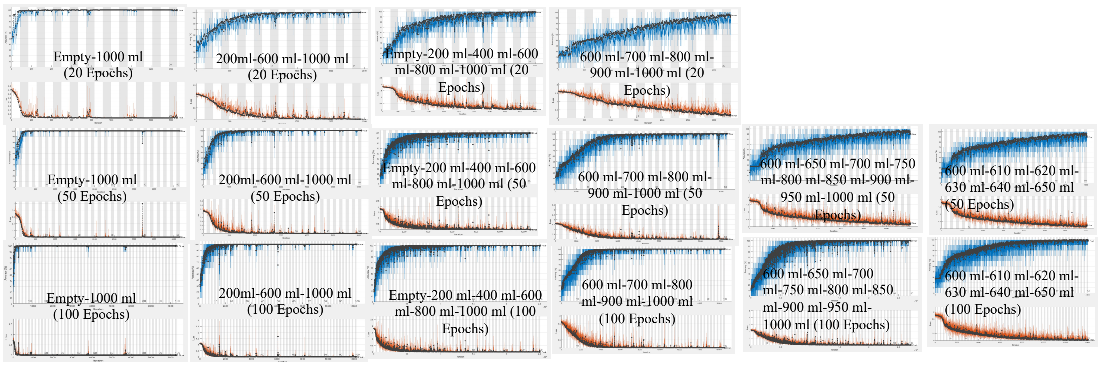


Figure 4. The training progress plots

In Table 1, it can also be observed that the number of epochs is in direct proportion to the accuracy. The results show that the number of images and number of epochs is in direct proportion to the number of iterations. As much as the number of iterations increased the time that the system needs to complete the training is also increasing, while the accuracy is improved. In consideration of 100 epochs runs, the training progress plots in Figure 5 show that as much as the volume difference decreases the network needs more time to improve the validation accuracy. While the 20 epoch runs gave an acceptable accuracy with a meaningful time consumption for a large difference of volume between classes, the 50 epoch runs gave a higher accuracy with an acceptable time consumption for each case. Thus, the suggested maximum epoch for the training is 50 epochs due to the possible overfitting issue in 100 epochs.

Table 1.

Experiment results

| **Case combinations** | **Max Epoch** | **Iteration** | **Time** | **Accuracy** |
|---|---|---|---|---|
| Empty-1000ml | 20 | 1660 | 30 min 16 sec | 99.40% |
| | 50 | 4150 | 68 min 28 sec | 100% |
| | 100 | 8300 | 133 min 40 sec | 100% |
| 200ml-600ml-1000ml | 20 | 2440 | 56 mins 3 sec | 100% |
| | 50 | 6100 | 141 min 18 sec | 100% |
| | 100 | 12200 | 283 min 40 sec | 100% |
| 600ml-700ml-800ml-900ml-1000ml | 20 | 3320 | 96 min 20 sec | 94.30% |
| | 50 | 8300 | 246 min 11 sec | 100% |
| | 100 | 16600 | 484 min 34 sec | 100% |
| Empty-200ml-400ml-600ml-800ml-1000ml | 10 | 2510 | 101 min 46 sec | 93.17% |
| | 20 | 5020 | 205 min 53 sec | 98.53% |
| | 50 | 12550 | 523 min 32 sec | 100% |
| | 100 | 25100 | 1053 min 39 sec | 100% |

| | | | | |
|---|---|---|---|---|
| 600ml-650ml-700ml-750ml-800ml-850ml-900ml-950ml-1000ml | 50 | 13500 | 574 min 34 sec | 99.70% |
| | 100 | 27100 | 1202 min 35 sec | 100% |
| 600ml-610ml-620ml-630ml-640ml-650ml | 50 | 7050 | 174 min 28 sec | 89.72% |
| | 100 | 14100 | 362 min 21 sec | 98.01% |

## Discussions and Future Works

In this study, the authors investigated the feasibility in the use of the deep learning approach for quantitatively estimating the earth volume in uniform and uncovered dump trucks. The toy truck was used to validate the concept. Total 4884 images were collected, with 3663 images used for training and the remaining 1221 images for validation. The preliminary results show that the pre-trained network architecture model is sensitive to the 5% difference in total volume but not the 1% difference in the total volume. 100% accuracy is achieved for all cases when volume difference in different classification is 5%, i.e., 50ml out of max load 1000mL or larger. Even for the last case where the volume difference is 1%, i.e, 10ml apart from 600 to 650, the pre-trained network still gives 98.01% accuracy with 100 epochs and 89.72% accuracy with 50 epochs. The results of 10 ml apart cases' validation test show that the network cannot classify the middle class correctly.

Meanwhile, to avoid the overfitting issues using a larger number of epochs, authors are implementing a few experiments using Transfer Learning with a smaller number of epochs, i.e., 30 versus 50. It is found that for cases that have no more than 3 classes, the accuracy is quite close to the results in Table 1, plus the training efficiency is improved obviously. As the classes increase to 4~5, the accuracy drops by 5~10%. Besides, the accuracy becomes very sensitive to the images if more irrelevant background features are included. Therefore, authors need to design a comprehensive testing plan, to investigate the possible major factors that will affect the training performance and volume estimation accuracy, such as image size, image quality, the suggested number of images as the training dataset, the optimal maximum epochs in the training model, the validation frequency, minimal batch size, etc. Since the ultimate goal for the research team is to automate this volume estimate process for real-time applications in the earthwork field application, the future study will investigate the challenges and performance if the proposed approach is applied to the field setting of the earthmoving operations, which might include but not limited to the light conditions under different weather or from different shooting angles, different size/colors of trucks in the images, varied shade effects due to the earthwork.

## References

Carlton Scale. (2021). How Much Does A Truck Scale Cost. Retrieved from https://www.carltonscale.com/much-truck-scale-cost/

Fekpe, E. S., Clayton, A. M., & Alfa, A. S. (1993). Aspects of performance of truck weigh stations. *Canadian Journal of Civil Engineering, 20*(3), 380-385.

Guo, C., Xu, J., Wang, M., Yan, T., Yang, L., & Sun, Z. (2016). Study on oxygen supply standard for physical health of construction personnel of high-altitude tunnels. *International journal of environmental research and public health, 13*(1), 64.

Jabri, A., & Zayed, T. (2017). Agent-based modeling and simulation of earthmoving operations. *Automation in Construction, 81*, 210-223.

Lee, J., & Chow, G. (2011). Operation analysis of the electronic screening system at a commercial vehicle weigh station. *Journal of Intelligent Transportation Systems, 15*(2), 91-103.

Lin, H., Xiang, H., Wang, L., & Yang, J. (2017). Weighing method for truck scale based on neural network with weight-smoothing constraint. *Measurement, 106*, 128-136.

Liu, Q., Feng, C., Song, Z., Louis, J., & Zhou, J. (2019). Deep learning model comparison for vision-based classification of full/empty-load trucks in earthmoving operations. *Applied Sciences, 9*(22), 4871.

MathWorks. (2021). Deep Learning with Images. Retrieved from https://www.mathworks.com/help/deeplearning/deep-learning-with-images.html?s_tid=CRUX_lftnav

Moselhi, O., & Alshibani, A. (2009). Optimization of earthmoving operations in heavy civil engineering projects. *Journal of Construction Engineering and Management, 135*(10), 948-954.

Samandar, M. S., Williams, B. M., & Ahmed, I. (2018). Weigh Station Impact on Truck Travel Time Reliability: Results and Findings from a Field Study and a Simulation Experiment. *Transportation research record, 2672*(9), 120-129.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Yi, W., & Chan, A. P. (2017). Effects of heat stress on construction labor productivity in Hong Kong: a case study of rebar workers. *International journal of environmental research and public health, 14*(9), 1055.