

Article

Deep Learning Model Comparison for Vision-Based Classification of Full/Empty-Load Trucks in Earthmoving Operations

Quan Liu ^{1,2} , Chen Feng ^{1,2}, Zida Song ^{1,2,*} , Joseph Louis ^{3,*} and Jian Zhou ⁴

¹ School of Water Resources and Hydropower Engineering, Wuhan University, Wuhan 430072, China; hapland@whu.edu.cn (Q.L.); whufchen2014@whu.edu.cn (C.F.)

² State Key Laboratory of Water Resources & Hydropower Engineering Science, Wuhan University, Wuhan 430072, China

³ School of Civil and Construction Engineering, Oregon State University, Corvallis, OR 97331, USA

⁴ Changjiang Survey, Planning, Design and Research Co., Ltd., Wuhan 430072, China; zhoujian@cjwsjy.com.cn

* Correspondence: zidasong@whu.edu.cn (Z.S.); joseph.louis@oregonstate.edu (J.L.); Tel.: +86-158-2732-6656 (Z.S.); +1-541-737-2114 (J.L.)

Received: 9 October 2019; Accepted: 11 November 2019; Published: 14 November 2019



Featured Application: Vision-based truck load counting in earthmoving operations, civil engineering management, and intelligent engineering.

Abstract: Earthmoving is an integral civil engineering operation of significance, and tracking its productivity requires the statistics of loads moved by dump trucks. Since current truck loads' statistics methods are laborious, costly, and limited in application, this paper presents the framework of a novel, automated, non-contact field earthmoving quantity statistics (FEQS) for projects with large earthmoving demands that use uniform and uncovered trucks. The proposed FEQS framework utilizes field surveillance systems and adopts vision-based deep learning for full/empty-load truck classification as the core work. Since convolutional neural network (CNN) and its transfer learning (TL) forms are popular vision-based deep learning models and numerous in type, a comparison study is conducted to test the framework's core work feasibility and evaluate the performance of different deep learning models in implementation. The comparison study involved 12 CNN or CNN-TL models in full/empty-load truck classification, and the results revealed that while several provided satisfactory performance, the VGG16-FineTune provided the optimal performance. This proved the core work feasibility of the proposed FEQS framework. Further discussion provides model choice suggestions that CNN-TL models are more feasible than CNN prototypes, and models that adopt different TL methods have advantages in either working accuracy or speed for different tasks.

Keywords: vision-based deep learning; earthmoving operation; dump truck; model comparing; convolutional neural network; transfer learning

1. Introduction

Earthmoving is a ubiquitous operation in construction projects and comprises a significant portion of the project cost and time, especially in the case of heavy civil and linear projects [1–3]. Within earthmoving operations, dump trucks are the major equipment typically used for the conveyance of material [4]. Thus, keeping track of dump truck operations, especially by counting or weighing working trucks, provides important information in the form of field earthmoving quantity statistics (FEQS). FEQS is defined as the statistical basis for the quantification of material moved in terms of quarry shipment, project site earthwork loading, or engineering waste earthwork disposal. FEQS data thus

contributes to field material management and is the primary information for financial settlement with earthmoving contractors. Thus, conducting FEQS is a very important and necessary aspect of managing earthmoving operations that can help avoid practical civil engineering management problems, like earthwork smuggling, financial settlement errors, and erroneous quantity estimation.

Current methods of updating FEQS rely on either manually counting the overall number of truck loads moved or on weighing trucks at load weigh stations. Both methods suffer from numerous disadvantages due to the manual nature of the former, and the high cost and disruption to operation caused in the latter, as described in the related work section. Moreover, current FEQS can be error prone or lack traceable data records, thus reducing validity. Incorrect tracking of truck counts and truck loads leads to numerous issues on construction projects between stakeholders that cause significant troubles for successful project completion. Aside from the problems of reporting inaccurate data, these issues can cause litigious situations between various stakeholders and derail the project management. On the other hand, the advent of advanced computational and artificial intelligence methods provides the possibility of automatically and objectively collecting FEQS data from site surveillance cameras, and informationized approaches make FEQS data cyber-recordable and easy to trace.

This paper thus presents the framework of a novel automated, non-contact, and vision-based FEQS, with the goal of reducing manual effort, costs, and errors to collect truck load information. The core work in the FEQS framework is the full/empty-load classification of earthmoving trucks as counting full-load trucks is an effective statistics strategy under a certain scenario. Vision-based deep learning is applied in the proposed FEQS as the core work solution, and because the potential usable deep learning models are numerous, testing and selection among them is needed. Hence, a comparison study to test the deep learning models' feasibility in full/empty-load truck classification and identify in the FEQS application is developed. Through the deep learning model comparison, the core work feasibility of the proposed framework can be assessed and a practical model choice for FEQS implementation can be suggested.

Thus, the main contributions of the paper can be summarized as: (1) The framework of an automated, non-contact FEQS applying vision-based deep learning is presented, which has advantages toward existed FEQS methods in terms of manual effort, costs, and errors; (2) the core work of the framework, i.e., the classification of full/empty-load trucks in earthmoving operations, is assessed in terms of feasibility through a comparison study that involves multiple deep learning models; and (3) the comparison study results are further discussed to give model choice suggestions for future implementation of the proposed FEQS.

The rest of the paper is as follows: Related work of the current state of art and practice in the domain of earthmoving FEQS and vision-based deep learning is first reviewed to identify the gaps in knowledge and thus to explain the proposed FEQS framework and the authors' research. Then, the methodology is described, followed by the comparison study results and discussion. Finally, the conclusions of this study are provided in terms of its contributions to knowledge and practice, along with limitations and future work of the study.

2. Related Work

Current FEQS methods in the civil engineering industry can be divided into two categories according to the statistical logic employed, i.e., truck counting or truck weighing. Both methods are reviewed to identify their limitations in this section, which stem from the manual effort or cost involved therein. The proposed solution to overcome these limitations is vision-based deep learning, which is also reviewed in this section along with its application in similar domains to highlight the gaps in research that will be targeted by this paper.

2.1. State-of-the-Art FEQS Methods

2.1.1. Counting FEQS

Counting FEQS counts the number of loaded trucks to keep track of the amount of material moved. It is applicable to scenarios that have a large quantity of overall transportation, low unit load price, and uniform trucks. Thus, the counting FEQS is suitable for civil engineering projects with a huge earthmoving demand and without precise single truck load weighing requirements, like hydropower, airport, large-scale landscape transformation [5] etc. In these larger projects, the trucks are generally uniform in capacity and truck loading is managed in a standardized manner [2]. Therefore, full-load trucks are assumed to reach a standard loading quantity, and for FEQS, checking whether a truck is full or empty and counting them is required. Moreover, compared to the large overall quantity, the quantity error of a single truck loading is minor, and this error can be shared and fixed by the overall statistics.

Currently, counting FEQS is mostly dependent on manual recognition of trucks' full/empty state and accounting. Automated solutions have been proposed to aid with keeping count of trucks, such as the use of computer aids to reduce the burden of manual work and human error rate. Tools, like the global positioning system (GPS) and radio frequency identification (RFID), have been applied for vehicle tracking or trip counting [6–9]. However, these tools can only be used to assist manual or achieve vehicle trip counting, and are unable to solve judgmental problems, like checking if a truck is full or empty. Hence, human labor is still required with these tools, and the related human error, high labor cost, and application limitations cannot be fully overcome. In detail, manual accounting applications are limited under harsh conditions, like high altitude, steep mountains, extreme cold, or hot areas, as labor safety may be threatened [10,11]; and there is also potential health damages as workers must suffer noise and vibration caused by trucks moving for a long time [12].

2.1.2. Weighing FEQS

As opposed to the imprecise method of counting trucks, weighing FEQS is a way of collecting statistics regarding the precise weight of material moved by the truck. This information is obtained from contact weighing tools, like the truck scale [13]. Weighing FEQS can obtain relatively accurate truck loading quantities and is applicable to scenarios that have a small amount of total transportation, multi-party contracting, or high unit load price, like earthmoving of small construction projects or highway freight charge [14,15]. In these scenarios, the truck model, load types, and values can be complex. Also, managing different subcontractors requires separate financial settlements, resulting in the need for more precise statistics toward single truck transportation.

The major class of limitations of weighing FEQS are problems imposed by the need for truck scale—a single truck scale costs between \$35,000 to \$100,000, has limited durability, and needs to be replaced after a given operating period [16], resulting in high deployment and maintenance costs. These scales also require trucks to first stop and then conduct the weighing, and can thus be a bottleneck that causes traffic jams during peak times [14], thus disturbing the truck flow and lowering overall transport efficiency [17].

2.2. Needs for Vision-Based FEQS Method

In current FEQS methods, problems have been exposed, such as high labor and economic costs, limited application environment, continuous maintenance requirement, and transportation interruption. Thus, it is of practical significance to develop a better FEQS system that is free of human labor, low cost, and well-adapted to the operating environment. Currently, there is no suitable solution for non-contact truck weighing; although devices that are more advanced than scales have been developed, contact with vehicles and laborious installation are still required [18,19]. Thus, high cost and limited application is unavoidable in weighing FEQS. However, for counting FEQS, as long as problems that need human judgment can be solved by unmanned and low-cost approaches, the goal

of developing a better means of FEQS is possible. Hence, in this paper, the research focus is placed on counting FEQS.

Counting FEQS is applicable for civil engineering projects, like hydropower, airport, large-scale landscape transformation, etc., as they have huge earthmoving demands and do not require precise truck load weighing. For these projects, their sites are generally located in open fields and can be equipped with surveillance camera systems for construction management and safety. Also, for these projects, earthmoving trucks normally are uniform and have just two states of full or empty when in working order, because project investment is adequate, contracting relationships are simple, and the truck loading processes are well-organized to guarantee the earthmoving quality of each time (a partially loaded truck is considered as a fault when working). Moreover, unlike transportation on city roads that strictly forbids dust spreading, in the open field, buckets of trucks do not need to be covered as no residents will be disturbed by dust due to earthmoving. Under this earthmoving operation scenario, through the surveillance camera systems, the truck loading conditions of full or empty can be directly viewed without occlusions, as uncovered trucks have an obvious characteristic difference between full/empty-load conditions, as Figure 1 shows. Thus, by judging the truck loading condition, i.e., the binary classification problem of full/empty-load trucks, the core work of the FEQS framework can be achieved. Also, as machine vision [20] can replace human eyes for truck loading condition judgment with little cost from the collection of video information and is free of human labor, a fully-automated, non-contact full/empty-load classification of earthmoving trucks can be implemented. Hence, a novel vision-based FEQS of truck counting can be proposed, and proper vision-based truck image classification approaches are needed.



Figure 1. Earthmoving trucks in surveillance videos: (a) Full-load truck samples; (b) empty-load truck samples.

2.3. Vision-Based Deep Learning in Related Areas

Machine vision is one branch of artificial intelligence [21] that pertains to the use of machines, including computers and related instruments, to replace human eyes to make observations and judgements about real-world scenes [20]. Currently, the industry adopts deep learning models, like convolutional neural networks (CNN), deep Boltzmann machines, deep belief networks etc.,

to achieve machine vision [22]. Among them, CNN [23] and its improved form, the transfer learning (TL) form [24], are the most popular approaches [25–28].

The deep learning models of CNN or CNN-TL have a wide range of applications, including medical science [29], agriculture [30], geology [31], manufacturing [32], transportation [33], civil engineering [34], and construction safety [35,36]. Studies are also abundant in specific aspects of vehicle management and earthmoving operations. Deep learning CNN and image-collecting tools, like surveillance cameras, have been combined and used for vehicle classification or real-time traffic monitoring [37–40]. CNN model improvements, like adopting the layer skipping strategy for better vehicle classification [41], or using CNN-TL to achieve both detection and classification of vehicles that include dump trucks, cars, and buses [42–44], have been performed. Apart from vehicle classification or detection, CNN is also able to recognize the working or idle state of earthwork machines, like excavators or trucks [45], and CNN-TL can benefit earthmoving operations or related construction management [27,46]. Other non-CNN machine vision methods also have applications in related areas, like vehicle collision prediction or construction machine detection, [47–50]. It can be seen that current vision-based deep learning researches mainly focus on the vehicle classification or state identification of earthwork machinery, and CNN and CNN-TL are widely applied.

Hence, CNN-related deep learning has the potential to replace human judgement for truck classification. Since CNN has numerous types and there are more than one TL methods, testing and selection among different models in truck image classification is necessary, which is merely studied at present.

3. Proposed FEQS Framework and Research Conception

In view of the suitability between the counting FEQS scenario of huge earthmoving demands, uniform uncovered trucks, equipped camera system, and vision-based deep learning, the authors posit that deep learning models of CNN or CNN-TL can achieve full/empty-load truck classification and contribute to unmanned and non-contact FEQS in earthmoving operations. Thus, the FEQS framework of this vision-based conception is proposed and is shown in Figure 2.

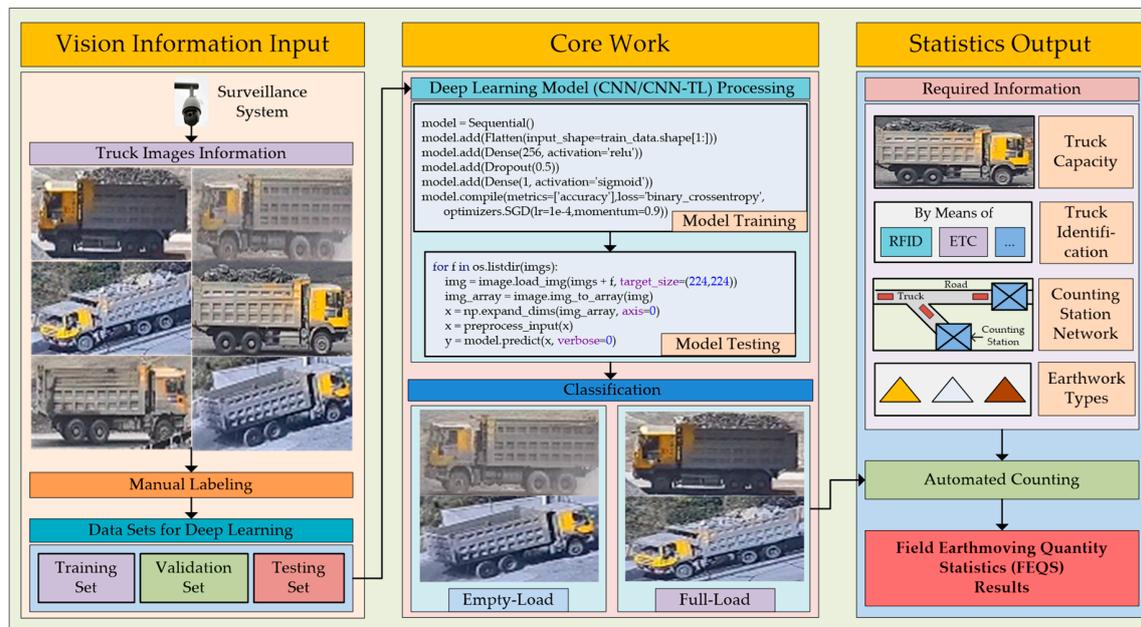


Figure 2. Proposed Field Earthmoving Quantity Statistics (FEQS) framework.

The framework first acquires vision information from the surveillance system and establishes data sets for deep learning by manual labeling. It then applies deep learning CNN-related models for

full/empty-load truck classification judgment. Finally, it combines the necessary information about trucks, truck identification, and the earthmoving project with the truck classification results and adopts automated counting to implement the automated non-contact FEQS. Under this framework, a partially loaded truck will be detected by the deep learning as it is not visually similar to full-load trucks and will be considered as empty as partially loaded does not qualify for counting as one instance of earthmoving work.

As vision-based judgement of the truck load condition is the core work of the proposed FEQS framework, this paper thus seeks to first verify the feasibility of CNN and CNN-TL models in solving full/empty-load-truck classification, and then evaluate the efficiency and ability of different models and identify the well performing learning models among the tested models for application suggestions. These are the premises of feasible and better implementation of the proposed FEQS. Hence, a comparison study is performed wherein multiple open-source CNN models and their TL forms are evaluated in a suitable counting FEQS scenario. The comparison study is the important research part of this paper, which provides a reference and support for the FEQS framework application. Three main works are performed in the comparison study: (i) Collecting empty-load and full-load truck images from a surveillance video source to form the training, validation, and testing data sets for deep learning; (ii) adopting 4 classical CNN models and 2 TL methods to construct 12 deep learning models for the comparison study; and (iii) testing the full/empty-load truck classification effect of each deep learning model, and further discussing the results.

4. Methodology

The methodology section consists of three phases, including the introduction of the CNN and the choice of four typical CNN models, the introduction of the TL and two main TL methods, and the determination of the deep learning models of the CNN prototypes or CNN TL forms to be tested.

4.1. Convolutional Neural Network

Convolutional neural network (CNN) is a type of feedforward neural network inspired by the biological visual cognitive mechanism [51]. It is one of the most popular deep learning approaches in the field of graphic processing as CNN performs well in image processing and directly deals with raw images. CNN extracts image features and compresses the data volume by operations, such as convolution, pooling, etc. The model is trained through gradient descent and back propagation algorithms [23], so that it can achieve functions like image classification. Generally, five layers constitute the main architecture of CNN: The input layer, convolution layer, activation layer, pooling layer, and fully connected layer. The operation procedure of CNNs is shown in Figure 3, and related descriptions are as follows:

1. **Input Layer.** This is the entrance for raw image data. In this layer, images can be preprocessed using operations, including normalization, principal component analysis, and whitening. Preprocessing makes images normative, which helps to speed up the training of network models and thus elevates model performance.
2. **Convolution Layer.** This is the main layer of a CNN, which performs convolution on inputted images to extract image features. Generally, a convolution layer contains multiple convolution kernels as filters so that it can obtain multiple image feature results.
3. **Activation Layer.** This layer is used for the nonlinear mapping of convolution results so that the multi-layer network can be nonlinear and has a better expression ability. Commonly used activation functions are the Relu function and the Sigmoid function.
4. **Pooling Layer.** This is also known as the down-sampling layer, and is the part that conducts dimensionality reduction for extracted feature and data compression, so that overfitting can be reduced and fault tolerance of the model can be improved. Pooling methods include MaxPooling and AveragePooling, and MaxPooling is commonly used now.

5. **Fully Connected Layer.** This is the result output layer that achieves the object classification function. This layer integrates the feature information from every neuron in the upper layer and classifies images according to the objective. There are generally two kinds of classification functions, the Sigmoid function for binary classification and the Softmax function for multiple classification.

Advantages of CNN in image recognition are revealed on a yearly basis in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [52]. The ImageNet dataset contains more than 13 million pictures from 20,000 categories, and ILSVRC randomly draws a subset with 1000 image categories from ImageNet for recognition contests (top-five error rate as the elevation index). In view of the guiding significance of ILSVRC in CNN-based image recognition, this paper selected four classical CNN models that relate to ILSVRC as the research basis, including:

- **VGGNet** [53] is a deep learning CNN model developed by Oxford University and Google DeepMind. It is based on the AlexNet [54], which increases the number of neural network layers from 8 to 16 or 19, and uses the small convolution kernel (3×3) to replace the big convolution kernel (11×11). In ILSVRC 2014, VGGNet won second place in the overall event and was champion in the positioning event; its top-five error rate is 7.50%.
- **Inception** [55] was developed by Szegedy et al., who changed the straight up and straight down “serial” network into the “parallel” sparse connection network with multiple convolution modes, and used the global average pooling layer to replace the full connected layer. InceptionV1 won champion of ILSVRC 2014, and its top-five error rate is 6.67%.
- **Xception** [56] is the improvement of Inception in 2017, which is a kind of extreme inception and uses the depthwise separable convolution layer to replace the convolution layer within InceptionV3. Its ILSVRC top-five error rate is 5.50%.
- **ResNet** [57] can solve the problem of over-fitting caused by too many neural network layers by introducing a residual network, and its network layers are deepened into 152 layers. ResNet won the champion of ILSVRC 2015, and its top-five error rate is 3.57%, which is better than the average ability of humans (5%).

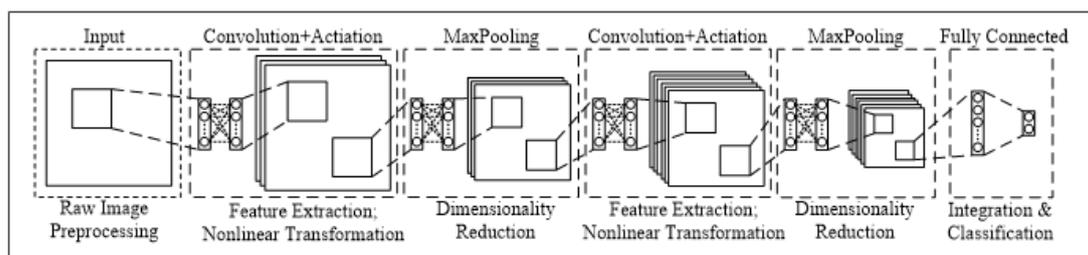


Figure 3. Operation procedure of Convolutional Neural Networks.

In this paper, all four CNN models adopted their classical mode, which are VGG16, InceptionV3, Xception, and Resnet50.

4.2. Transfer Learning Methods

Transfer learning (TL) is an improvement for CNN, which transfers the pre-trained experience from the source domain to the target domain so that the CNN model can possess a better image recognition ability or deal with a new objective that has few labeled images [28,58,59]. It has been proven that TL forms of CNN have good generalization, and compared to the result of prototypes, CNN-TL models have a stronger image feature extraction ability outside the range of the training data [53,60,61]. However, as TL needs pre-trained experience transfer, CNN-TL models have a different working process and different efficiencies in training and testing compared to their prototypes. Neither CNN nor

CNN-TL have been tested in full/empty-load truck classification before. Thus, it is necessary to compare CNN-TL and CNN models to determine their relation within the context of the comparison study.

Currently, for TL, the source domain directly adopts the ImageNet data set, and the main TL methods include the bottleneck feature (BF) and fine tune (FT) [26,62,63]. Hence, in this paper, CNN-TL models refer to the classical CNN models that adopt TL methods of BF or FT. The schematics of the two TL methods are shown in Figure 4 and the related explanations are as follows. Here, the expression of the convolution block (CB) refers to the combination of one convolution layer, one activation layer, and one pooling layer; the abbreviation of FC refers to the fully connected layer.

- Bottleneck Feature:** When a CNN model adopts the TL method of BF, all its CBs should be frozen, while all its FCs are kept as trainable and can be customized according to the classification demand. The operational process of BF is shown in (a): (i) Transferring the pretrained network weights from the source domain (ImageNet in this case) to the whole CNN model; (ii) obtaining the bottleneck feature with training data and through frozen BCs according to the recognition objective; (iii)–(v) applying the bottleneck feature to the trainable FCs and training all FCs, so that BF weights, including weights for both BCs and FCs, are obtained; (vi) applying the BF weights to the original CNN model, then its TL form of the CNN-BF model can be generated. Notably, during the BF training process, the weights of all BCs are kept unchanged as the pretrained weights from the source domain, because all BCs are frozen, while the weights of all FCs are trained and updated.
- Fine Tune:** Fine tune (FT) is based on BF, i.e., the secondary TL after obtaining the BF weights. Before the processing of FT, part of the rear BCs should be unfrozen, i.e., allowing some BCs to participate in the model training. The operational process of FT is shown in (b): (i) Applying the BF weights to the whole CNN model as the pretrained weights; (ii)–(iii) training the trainable part with training data according to the recognition objective, thus the FT weights are obtained; (iv) applying the FT weights to the original CNN model, then its TL form of the CNN-FT model can be generated. Notably, during the training process of FT, the weights of frozen BCs are kept unchanged while the weights of unfrozen BCs and FCs are trained and updated.

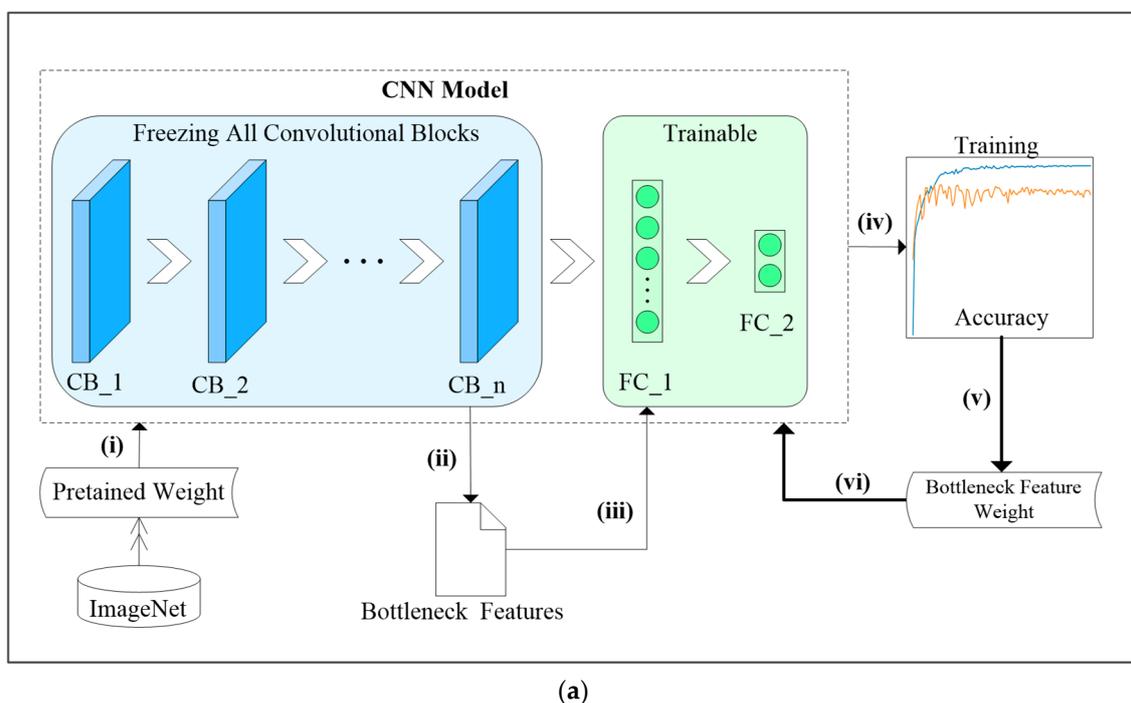


Figure 4. Cont.

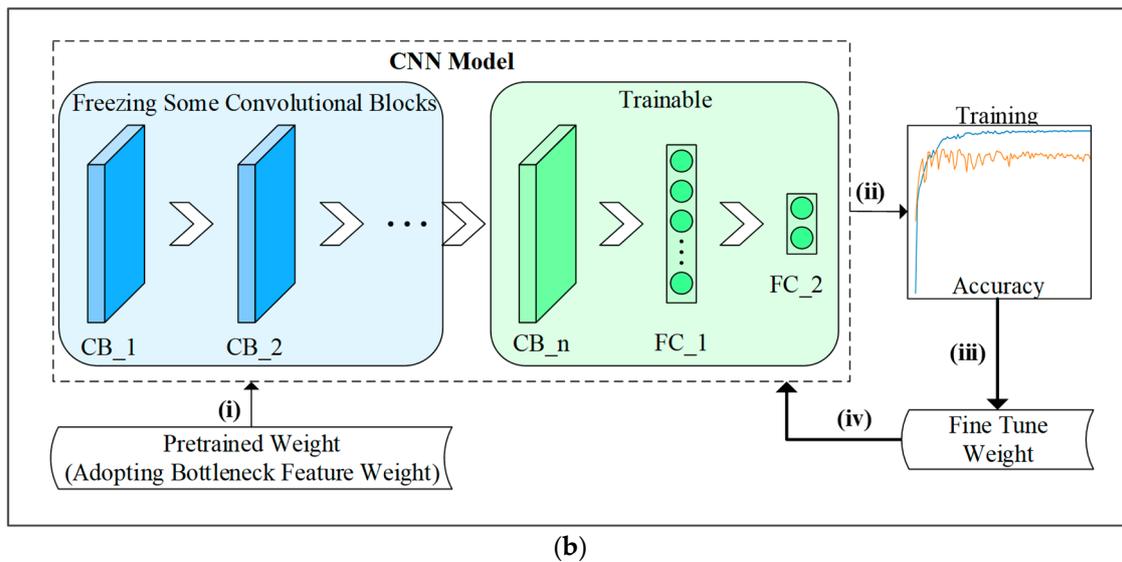


Figure 4. Schematics of Transfer Learning methods: (a) Bottleneck feature; (b) fine tune.

4.3. Models to Be Tested

In this paper, the deep learning models to be tested included four classical CNN models and their eight TL forms that apply BF or FT. Hence, 12 models in total were involved (Table 1).

Table 1. Deep learning models to be tested.

CNNs	TL Methods	Models to be Tested
VGG16	None	VGG16
	BF	VGG16-BF
	FT	VGG16-FT
InceptionV3	None	InceptionV3
	BF	InceptionV3-BF
	FT	InceptionV3-FT
Xception	None	Xception
	BF	Xception-BF
	FT	Xception-FT
Resnet50	None	Resnet50
	BF	Resnet50-BF
	FT	Resnet50-FT

5. Results

The results section involves the preliminary assessment and results of the comparison study. Both the training results (time; convergence; accuracy) and testing results (speed; accuracy) of models are compared in this section. The testing results of speed and accuracy are the primary model selection reference, because testing results correspond to the working effectiveness in an actual application.

Details about the deep learning model comparison study are shown in Figure 5.

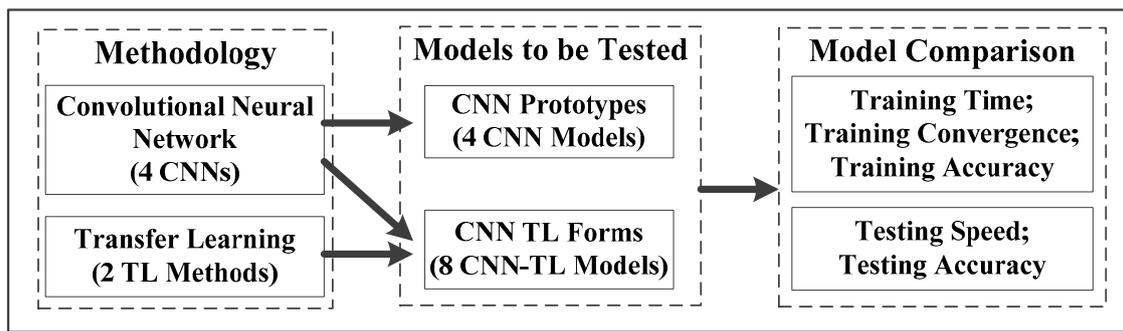


Figure 5. Deep learning model comparison study.

5.1. Preliminary

The comparison study was based on a large-scale landscape transformation project, which adopts the counting FEQS logic, and thus requires the full/empty-load judgment for earthmoving trucks. The project site was located in an open field that is quarantined from the public, and thus no interference from external vehicles existed. All earthmoving trucks were uniform in models and loading capacity, and truck buckets were uncovered. The surveillance camera system was deployed on the route of truck transportation. Hence, the surveillance video was used as the data source.

Before the model comparison, enough truck images were collected to form the data sets for training, validation, and testing. The principles of image collecting in this study were:

- **Number:** Over 500 images for each full-load or empty-load truck state should be collected.
- **Labeling:** Truck images are taken from the surveillance video by a screenshot, and are manually labeled to guarantee their correctness.
- **Size:** Truck images should be uniform in size; in this case, the size should be around 350×250 pixels, and the truck should be in the middle of the image and occupying about one third to one half of the frame.
- **Visibility:** The truck bucket should be visible so that the full-load and empty-load condition can be distinguished, and the truck in the image should be distinguishable from the background, i.e., the image should be collected and become usable data only if its truck loading condition can be distinguished by human eyes.

In practice, 2454 images were taken and included as the data for the deep learning study following the above principles, thus they were uniform in size, manually labeled, and have good distinguishability by visual sight. The image collecting spanned three months (July to September) under different lighting or weather conditions. Among 2454 images, 1588 were full-load truck images and 866 were empty-load truck images. On the basis of fully utilizing the image data, all collected images were randomly divided into three sets according to the ratio of training set:validation set:testing set = 6:3:1. For the three sets: (i) The training set included labeled images of full/empty-load trucks and was used to train deep learning models; (ii) the validation set included labeled images of full/empty-load trucks, which did not participate in the model training but was used to test the training performance of models at the end of each training epoch; (iii) the testing set included unlabeled images of trucks, which was used to test the generalization ability of trained models, i.e., the actual working performance. Details of the three data sets are shown in Table A1 (in Appendix A).

Based on the three data sets, the 12 deep learning models were trained, validated, and tested. The training set was first uploaded to train deep learning models. During the training, images in the training or validation sets were used to test the model training performance. Finally, after finishing all training epochs, the trained model was tested by the testing set in full/empty-load trucks classification and working performance of models can be revealed. Notably, classification errors in the test results were manually obtained. Training/testing results of the 12 models can be seen in Table A2 (in the Appendix A).

Since the data sets of training and validation here were disproportional in image numbers of the full-load and empty-load state, a contrast that adopts proportional data sets was attached to show the possible effect of disproportions. Hence, the number of full-load truck images in the original training and validation sets were reduced to the equal number of empty-load truck images, and the 12 deep learning models were trained and tested again under the reduced but proportional data sets. Details of these proportional data sets and training/testing results of this new round are shown in Tables A3 and A4 (in Appendix A). The contrast between Tables A2 and A4 shows that disproportions can cause value changes in the training/testing results of deep learning models, whereas these changes are slight and do not affect the ranking among deep learning models. Hence, the effect of disproportions can be considered as acceptable, and this paper adopted the original disproportional data sets for the comparison study as they can fully utilized the collected images.

The hardware environment used in this study included the following: Intel Core i7-8700 CPU, NVIDIA GTX 1070 Ti GPU, and 32 GB RAM. The software environment includes: Windows10 OS, Python3.6, Keras2.2.4, and Tensorflow-gpu1.12.0. The training setting included the epoch was set as 100, batch size was set as 10, training images were resized into 224×224 pixels, optimizer adopted the SGD, learning rate was set as 0.0001, and momentum was set as 0.9. Furthermore, all CNN-FT models only unfroze the last CB.

5.2. Study Results

Comparison study results were determined as:

- Training time costing: The time costing of 100 epochs of each model are shown in Figure 6.

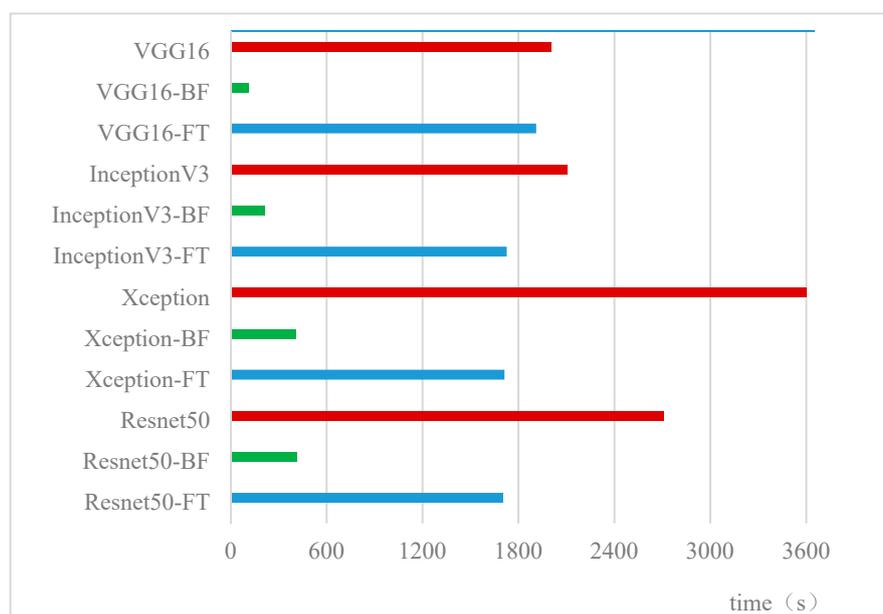


Figure 6. Training time costing results.

It can be seen that among 12 models, the VGG16-BF was the fastest (113 s) and the Xception was the slowest (3604 s).

- Accuracy change during training: Generally, in the training process of deep learning models, with the constant updating of neuron parameters, the accuracy of the model in truck classification is elevated. To reflect the accuracy changes during training, two accuracy curves based on the training set itself and the validation set, respectively, were adopted for each model, and by analyzing the two accuracy curves, the model training performance can be evaluated. Generally, models

with a smaller curve fluctuation during accuracy elevation have better training convergence, and models with two closer accuracy curves of the training set and the validation set and higher accuracy have a better training performance. The accuracy curves of each model in the training process are shown in Figure 7. Here, the front mark of NTL refers to not adopting TL, while BF or FT refers to the corresponding TL; the rear mark of _Train refers to the accuracy curve basing on the training set, while _Validation refers to the accuracy curve based on the validation set.

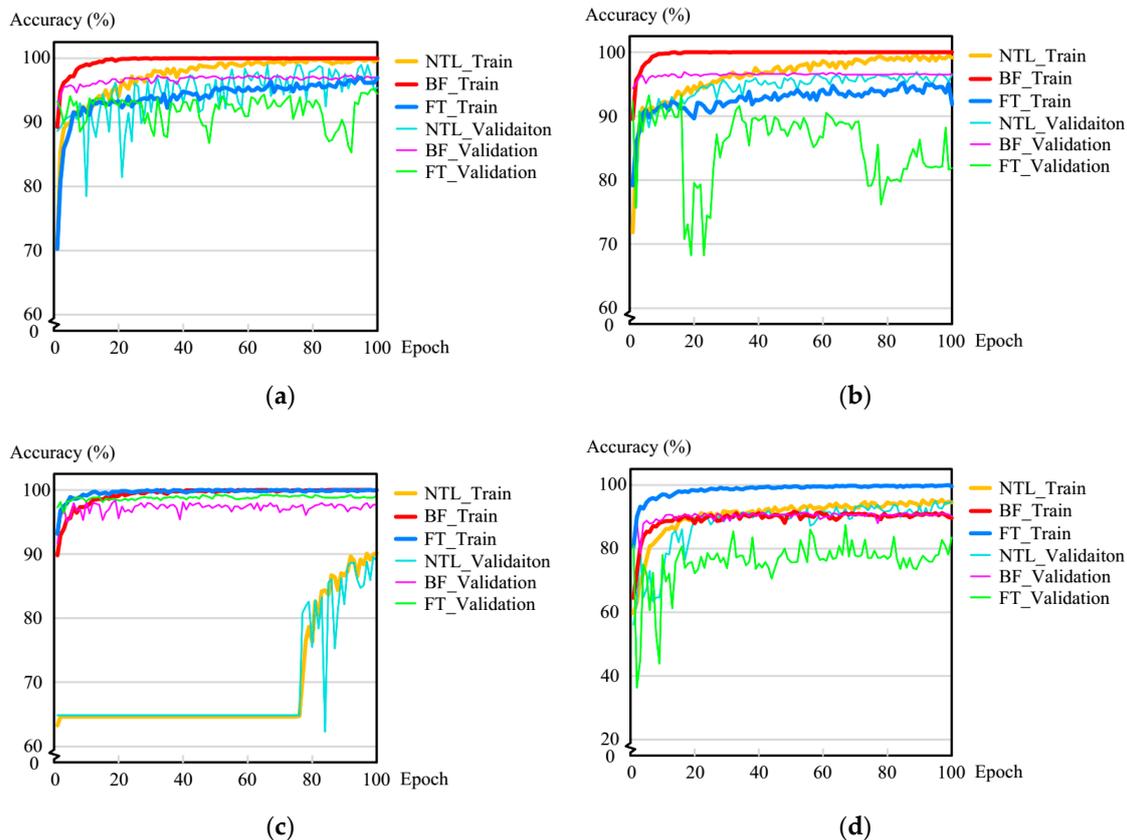


Figure 7. Training accuracy changing of deep learning model prototypes and TL forms: (a) VGG16 and VGG16-TL; (b) InceptionV3 and InceptionV3-TL; (c) Xception and Xception-TL; (d) Resnet50 and Resnet50-TL.

It can be seen that, among the 12 models, the VGG16-FT in (a) had the best convergence and training performance. For VGG16-FT, both its accuracy curves based on the training set and validation set had small fluctuation (only about 2% after the 20th epoch), thus its convergence can be considered as good. Meanwhile, its two accuracy curves were the closest by contrast, and at the later stage of training, its accuracy based on the training set approached 100% while its accuracy based on the validation set approached 98% (the highest among the 12 models), thus its training performance can also be considered as the best. Other relatively good deep learning models include the VGG16-BF, the InceptionV3-BF, and the Xception-BF in (a), (b), and (c). The VGG16, in (a), was the worst in training outcomes compared to the other models, as both its accuracy curves had a low accuracy of 68% before the 78th epoch, and did not break through the 90% accuracy at the end of the training. The InceptionV3-FT, in (b), and the Resnet50-FT, in (d), both had poor training convergence, as their accuracy curves based on the validation set had large fluctuation (over 20% and 40%, respectively), and the differences between their two curves were also large.

- Testing results of trained models: As deep learning models are trained, the testing set can be used to test the models' usability and working performance. Here, testing accuracy and speed

are the main indicators for evaluating a model, and to reach a satisfactory application, the truck classification accuracy should be over 95% to be seen as qualified, and the working speed should be as fast as possible. The testing accuracy and testing speed results are shown in Figures 8 and 9.

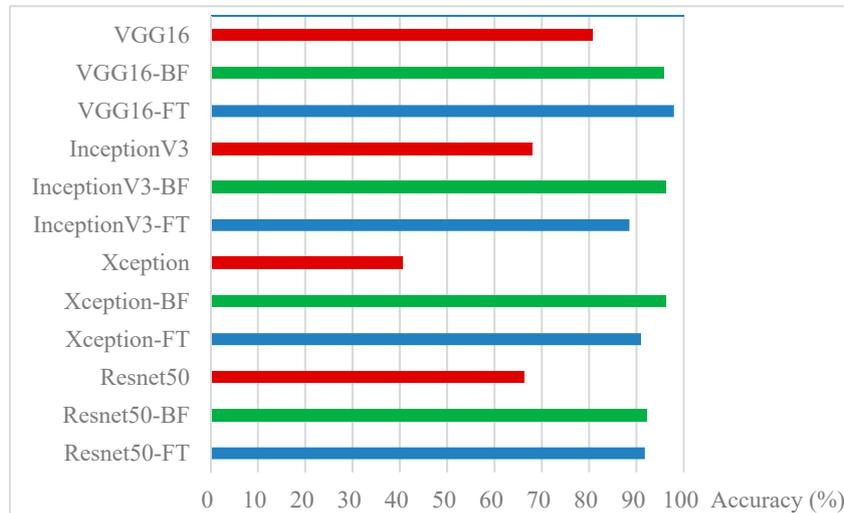


Figure 8. Testing accuracy result.

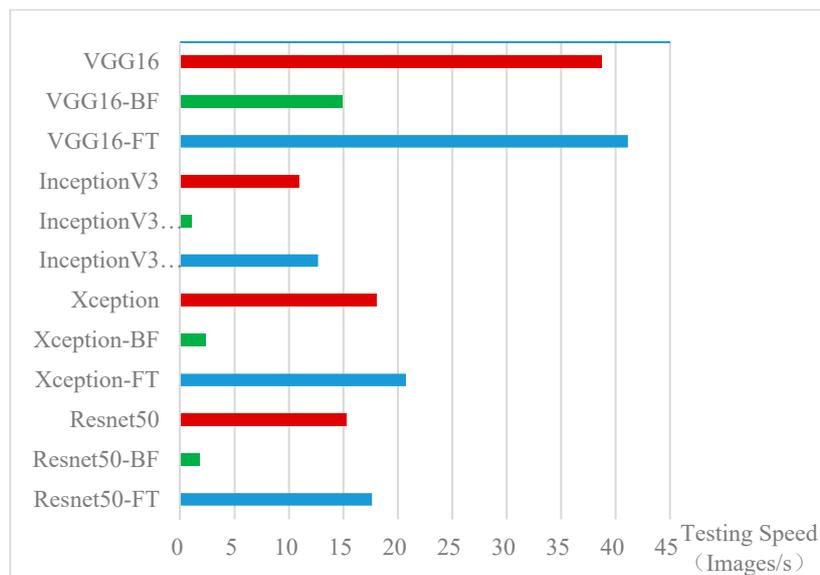


Figure 9. Testing speed result.

It can be seen that, among 12 models, the VGG16-FT had the highest accuracy of full/empty-load truck classification (98%), and its test speed was also the fastest (41.1 images/s). VGG16-BF, InceptionV3-BF, and Xception-BF all reached the qualified level of testing accuracy (over 95%) but had slower testing speeds than the VGG16-FT. The Xception had the lowest testing accuracy of the full/empty-load truck classification (40.6%). The InceptionV3-BF had the slowest testing speed (1.1 images/s).

Evidently, the vision-based deep learning was able to replace human eyes for full/empty-load truck classification in counting FEQS as the VGG16-FT showed a good performance that exceeded the accuracy goal, and three CNN-BF models just reached the accuracy goal. Hence, the core work feasibility of the proposed FEQS framework was proven.

6. Discussion

Based on the results of the comparison study, further discussion is provided to reveal more useful information for the implementation of proposed FEQS.

6.1. In the Aspect of Model Training Time

As Figure 6 shows, the training time costings of different forms of the four classical CNNs showed a similar tendency, i.e., for a CNN, its prototype model had the longest training time, its TL form of the CNN-FT model had the second longest training time, while its TL form of the CNN-BF model was significantly faster than the other two forms. The reasons for this tendency can be concluded as: (i) The CNN prototypes have the most trainable parameters as the whole neural network participates in the model training, hence their training time is the longest; (ii) TL forms freeze some CBs, hence they have a shorter training time than prototypes; and (iii) CNN-BF models freeze all CBs while CNN-FT models only freeze part of CBs, hence CNN-BF models have the shortest time costing. In summing up, adopting CNN-BF can evidently shorten the model training time.

Since deep learning models are generally trained in high-performance workstations and then migrate to terminal devices for field practice, the training time advantage only reduces deployment time but does not provide a better application effect. Hence, in this paper, the training time costing is not considered as a main indicator for identifying the optimal model but a model selection reference.

6.2. In the Aspect of Model Training Performance

As Figure 7 shows, during the training process, TL forms of CNN generally have advantages to their prototypes in the training accuracy, but this advantage was not absolute. Detailed speaking, for model training accuracy after 100 epochs: VGG16-BF and VGG16-FT were all better than the VGG16, in (a); the InceptionV3-BF was better than the InceptionV3, and the Xception-BF was better than the Xception, in (b) and (c); and, however, for the Resnet50, its prototype was better than all its TL forms, in (d). For training convergence, though the VGG16-FT had the best convergence among the 12 models, other CNN-FT models just had worse convergence than their prototypes, and CNN-BF models generally had better convergence than their prototypes. Since two accuracy curves of CNN-BF models were generally good in fluctuation control and closeness, CNN-FT models, other than the VGG16-FT, had very drastic fluctuation in the validation set-based accuracy curves and large differences between the two accuracy curves.

To summarize, besides the VGG16-FT, CNN-BF models generally had better overall training performance than the prototypes and CNN-FT models, and special attention should be paid to the Resnet50 as its prototype had better training accuracy than its TL forms. The training performance reflects the stability of the deep learning model and should be considered in model selection.

6.3. In the Aspect of Model Testing Performance

As Figures 8 and 9 show, all CNN prototypes had poor testing accuracy that was lower than their TL forms, and the highest accuracy among prototypes was just 81% for the VGG16. The VGG16-FT achieved the best accuracy of 98%, and the other three models of the VGG16-BF, InceptionV3-BF, and Xception-BF all had good accuracy of over 95%. For testing speed comparison, CNN-FT models were slightly faster than the prototypes while CNN-BF models were evidently slower than both the prototypes and CNN-FT models. The VGG16-FT also had the fastest testing speed of 41.1 images/s, thus just besides the prototype of VGG16, the VGG16-FT had a significant speed advantage (nearly double the speed of other models). Summing up, this comparison study showed that TL forms of CNN have advantages in testing accuracy, and CNN-FT models have advantages in testing speed.

Summing up, it can be concluded that the VGG16-FT was optimal on the whole, as it had the highest accuracy, fastest operation speed, and best training convergence. Hence, based on this comparison study, the VGG16-FT is recommended as the most suitable model for the proposed

vision-based FEQS method. Meanwhile, it can be seen from the discussion of results that for the full/empty-load truck binary classification problems, TL forms of CNN have advantages to CNN prototypes because CNN-BF models are generally better in model training and both CNN-BF and CNN-FT models are better in model testing than prototypes. As the testing accuracy results of CNN prototypes all are lower than 85%, adopting CNN prototypes in actual application needs consideration, because poor accuracy fails to reach the working goal. For TL forms of CNN, CNN-BF models have advantages in both training and testing accuracy, CNN-FT models have advantages in testing speed and are also good in testing accuracy.

Hence, model choice suggestion in applications can be provided: The adopted model should be chosen according to the working demands and conditions, i.e., when the working accuracy is of first priority, CNN-BF models can be recommended as they are better in accuracy; when real-time working is required, CNN-FT models can be recommended as their fast operation speed can reduce the working time delay.

7. Conclusions

This paper presented the framework of a novel, automated field earthmoving quantity statistics (FEQS) that mainly applies vision-based deep learning for full/empty-load truck classification as the core work and counts full-load trucks (Figure 2). The proposed FEQS contributes to relieving current problems in FEQS of manual, laborious, high cost, truck traffic interference, continuous maintenance demands, and application limitations, because it utilizes the field-equipped surveillance video system and deep learning CNN-related image recognition models to achieve unmanned and non-contact truck load condition judgement.

As deep learning CNN-related models (prototypes and TL forms) are numerous, the authors introduced a comparison study to test and evaluate CNN-related models' performance in full/empty-load earthmoving truck classification. Thus, the core work of the proposed FEQS framework can be assessed in terms of feasibility, and well-performed models can be identified for model choice suggestions in future FEQS implementation.

The comparison study involved 12 deep learning models constructed by four classical CNNs of VGG16, InceptionV3, Xception, and Resnet50 and two popular TL methods of BF and FT. Based on a proper earthmoving project scenario, the training and testing results of the 12 models were obtained. The study results showed that, on the whole, the VGG16-FT was the optimal model among the 12 models, as it had the highest working accuracy of 98% and the fastest truck image classification speed of 41.1 images/s. In addition, the VGG16-BF, InceptionV3-BF, and Xception-BF all reached the satisfactory goal of a working accuracy of over 95%, and showed advantages in model training, but their working speeds were not as fast. Hence, the VGG16-FT was able to achieve full/empty truck classification in the application level, and other three CNN-BF models also had further application potential. It can be concluded that, through the comparison study, the core work of the proposed vision-based FEQS framework was proven as theoretically feasible.

Further discussion showed that compared to CNN prototypes, their TL forms generally have a better working accuracy and training performance in full/empty truck classification, and the four classical CNN prototypes all have relatively lower working accuracy than their TL forms. For CNN-TL models, generally, CNN-BF models have the advantage in working accuracy, while CNN-FT models have the advantage in working speed. Hence, in industrial applications, TL forms of CNN are recommended to replace their prototypes, and the specific choice of CNN-TL models depends on the working demand. Model choice suggestions are CNN-BF models are more suitable for tasks with a high accuracy demand, while CNN-FT models are more suitable for real-time tasks.

This paper provides a reference and support for the application of vision-based deep learning CNN and CNN-TL models in earthmoving operations, civil engineering management, and intelligent engineering.

Limitations about the current study lie on the fact that the proposed vision-based FEQS requires quarantined projects in an open field that use uniform and uncovered trucks. However, the requirement

for non-occluded scenes is a common limitation of vision-based methods, and the other constraints are fulfilled in a large number of infrastructure projects as indicated in the literature review. Nevertheless, future work will involve applying vehicle recognition before performing full/empty classification to exclude unwanted vehicles and to extract truck information in advance. Moreover, using machine vision to recognize the load weight of covered trucks will also be studied in the future to replace current weighing methods with a non-contact method.

Author Contributions: Conceptualization, Q.L., Z.S. and C.F.; methodology, Q.L., C.F. and Z.S.; software, C.F.; validation, Q.L., Z.S. and J.L.; formal analysis, Q.L. and J.L.; resources, J.Z.; Writing—Original draft preparation, Z.S. and C.F.; Writing—Review and editing, Q.L. and J.L.; visualization, Z.S. and C.F.; supervision, Q.L.; funding acquisition, Q.L.

Funding: This study was funded by the National Natural Science Foundation of China, grant number 51379164.

Acknowledgments: This study was supported by the Changjiang Survey, Planning, Design and Research Co.,Ltd, which are gratefully acknowledged.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Classification of applied data sets.

Data Sets	Training Set	Validation Set	Testing Set
Full-Load Trucks	950 images	480 images	244 images
Empty-Load Trucks	520 images	260 images	
Objective	For model training	For model validation during training	For model performance testing

Table A2. Results of the deep learning model comparison.

Models	Training Time (s)	Testing Speed (Images/s)	Testing Accuracy (%)
VGG16	2004	38.7	80.7
VGG16-BF	113	14.9	95.9
VGG16-FT	1911	41.1	98.0
InceptionV3	2103	10.9	68.0
InceptionV3-BF	214	1.1	96.3
InceptionV3-FT	1727	12.7	88.5
Xception	3604	18.0	40.6
Xception-BF	410	2.3	96.3
Xception-FT	1713	20.7	91.0
Resnet50	2706	15.3	66.4
Resnet50-BF	412	1.8	92.2
Resnet50-FT	1705	17.6	91.8

Table A3. Classification of reduced but proportional data sets.

Data Sets	Training Set	Validation Set	Testing Set
Full-Load Trucks	520 images	260 images	244 images
Empty-Load Trucks	520 images	260 images	
Objective	For model training	For model validation during training	For model performance testing

Table A4. Results of the deep learning model using proportional data sets.

Models	Training Time (s)	Testing Speed (Images/s)	Testing Accuracy (%)
VGG16	1703	40.1	90.2
VGG16-BF	110	14.8	98.8
VGG16-FT	1169	42.5	99.6
InceptionV3	1509	10.5	35.7
InceptionV3-BF	113	1.0	97.1
InceptionV3-FT	1109	12.6	91.0
Xception	2431	18.9	35.0
Xception-BF	213	2.3	97.5
Xception-FT	1104	21.2	94.7
Resnet50	1806	16.0	64.3
Resnet50-BF	213	1.8	92.6
Resnet50-FT	1104	19.3	91.8

References

- Gomes Correia, A.; Winter, M.G.; Puppala, A.J. A review of sustainable approaches in transport infrastructure geotechnics. *Transp. Geotech.* **2016**, *7*, 21–28. [CrossRef]
- Smith, S. Earthmoving Productivity Estimation Using Linear Regression Techniques. *J. Constr. Eng. Manag.* **1999**, *125*, 133–141. [CrossRef]
- Pan, Z.; Zhou, Y.; Zhao, C.; Hu, C.; Zhou, H.; Fan, Y. Assessment Method of Slope Excavation Quality based on Point Cloud Data. *KSCE J. Civ. Eng.* **2019**, *23*, 935–946. [CrossRef]
- Jabri, A.; Zayed, T. Agent-based modeling and simulation of earthmoving operations. *Autom. Constr.* **2017**, *81*, 210–223. [CrossRef]
- Moselhi, O.; Alshibani, A. Optimization of Earthmoving Operations in Heavy Civil Engineering Projects. *J. Constr. Eng. Manag.* **2009**, *135*, 948–954. [CrossRef]
- You, S.I.; Ritchie, S.G. A GPS Data Processing Framework for Analysis of Drayage Truck Tours. *KSCE J. Civ. Eng.* **2017**, *22*, 1454–1465. [CrossRef]
- Lee, S.S.; Park, S.-I.; Seo, J. Utilization analysis methodology for fleet telematics of heavy earthwork equipment. *Autom. Constr.* **2018**, *92*, 59–67. [CrossRef]
- Bell, K.E.; Figliozzi, M.A. Ancillary Functions for Smartphone Weight–Mile Tax Truck Data. *Transp. Res. Rec. J. Transp. Res. Board* **2013**, *2378*, 22–31. [CrossRef]
- Hannan, M.A.; Arebey, M.; Begum, R.A.; Basri, H. Radio Frequency Identification (RFID) and communication technologies for solid waste bin and truck monitoring system. *Waste Manag.* **2011**, *31*, 2406–2413. [CrossRef]
- Yi, W.; Chan, A.P.C. Effects of Heat Stress on Construction Labor Productivity in Hong Kong: A Case Study of Rebar Workers. *Int. J. Environ. Res. Public Health* **2017**, *14*, 1055. [CrossRef]
- Guo, C.; Xu, J.; Wang, M.; Yan, T.; Yang, L.; Sun, Z. Study on Oxygen Supply Standard for Physical Health of Construction Personnel of High-Altitude Tunnels. *Int. J. Environ. Res. Public Health* **2015**, *13*, 64. [CrossRef]
- Chao, P.C.; Juang, Y.J.; Chen, C.J.; Dai, Y.T.; Yeh, C.Y.; Hu, C.Y. Combined effects of noise, vibration, and low temperature on the physiological parameters of labor employees. *Kaohsiung J. Med. Sci.* **2013**, *29*, 560–567. [CrossRef]
- Lin, H.; Xiang, H.; Wang, L.; Yang, J. Weighing method for truck scale based on neural network with weight-smoothing constraint. *Measurement* **2017**, *106*, 128–136. [CrossRef]
- Lee, J.B.; Chow, G. Operation Analysis of the Electronic Screening System at a Commercial Vehicle Weigh Station. *J. Intell. Transp. Syst.* **2011**, *15*, 91–103. [CrossRef]
- Fekpe, E.S.K.; Clayton, A.; Alfa, A.S. Aspects of performance of truck weigh stations. *Can. J. Civ. Eng.* **1993**, *20*, 380–385. [CrossRef]
- Buck, K. How Much Does A Truck Scale Cost. Available online: <https://www.carltonscale.com/much-truck-scale-cost/> (accessed on 10 September 2019).
- Samandar, M.S.; Williams, B.M.; Ahmed, I. Weigh Station Impact on Truck Travel Time Reliability: Results and Findings from a Field Study and a Simulation Experiment. *Transp. Res. Rec.* **2018**, *2672*, 120–129. [CrossRef]

18. Bajwa, R.; Coleri, E.; Rajagopal, R.; Varaiya, P.; Flores, C. Development of a Cost-Effective Wireless Vibration Weigh-In-Motion System to Estimate Axle Weights of Trucks. *Comput.-Aided Civ. Infrastruct. Eng.* **2017**, *32*, 443–457. [[CrossRef](#)]
19. Zhang, W.; Suo, C.; Wang, Q. A Novel Sensor System for Measuring Wheel Loads of Vehicles on Highways. *Sensors* **2008**, *8*, 7671–7689. [[CrossRef](#)]
20. Jain, R.; Kasturi, R.; Schunck, B.G. *Machine Vision*; McGraw-Hill New York: New York, NY, USA, 1995; Volume 5.
21. Turing, A.M. Computing Machinery and Intelligence. In *Parsing the Turing Test*; Springer: Dordrecht, The Netherlands, 2009.
22. Voulodimos, A.; Doulamis, N.D.; Doulamis, A.D.; Protopapadakis, E. Deep Learning for Computer Vision: A Brief Review. *Comput. Intell. Neurosci.* **2018**, *2018*, 1–13. [[CrossRef](#)]
23. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)]
24. Raina, R.; Battle, A.; Lee, H.; Packer, B.; Ng, A.Y. Self-taught learning: Transfer learning from unlabeled data. In Proceedings of the 24th International Conference on Machine Learning, Corvallis, OR, USA, 20–24 June 2007; pp. 759–766.
25. Kolar, Z.; Chen, H.; Luo, X. Transfer learning and deep convolutional neural networks for safety guardrail detection in 2D images. *Autom. Constr.* **2018**, *89*, 58–70. [[CrossRef](#)]
26. Gao, Y.; Mosalam, K.M. Deep Transfer Learning for Image-Based Structural Damage Recognition. *Comput.-Aided Civ. Infrastruct. Eng.* **2018**, *33*, 748–768. [[CrossRef](#)]
27. Kim, H.; Kim, H.; Hong, Y.W.; Byun, H. Detecting Construction Equipment Using a Region-Based Fully Convolutional Network and Transfer Learning. *J. Comput. Civ. Eng.* **2018**, *32*, 04017082. [[CrossRef](#)]
28. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [[CrossRef](#)]
29. Durr, O.; Sick, B. Single-Cell Phenotype Classification Using Deep Convolutional Neural Networks. *J. Biomol. Screen.* **2016**, *21*, 998–1003. [[CrossRef](#)]
30. Patrício, D.I.; Rieder, R. Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. *Comput. Electron. Agric.* **2018**, *153*, 69–81. [[CrossRef](#)]
31. Li, N.; Hao, H.; Gu, Q.; Wang, D.; Hu, X. A transfer learning method for automatic identification of sandstone microscopic images. *Comput. Geosci.* **2017**, *103*, 111–121. [[CrossRef](#)]
32. Reis, J.; Goncalves, G.M. Laser Seam Welding optimization using Inductive Transfer Learning with Artificial Neural Networks. In Proceedings of the Emerging Technologies and Factory Automation, Turin, Italy, 4–7 September 2018; pp. 646–653.
33. Zhang, A.; Wang, K.C.P.; Li, B.; Yang, E.; Dai, X.; Peng, Y.; Fei, Y.; Liu, Y.; Li, J.Q.; Chen, C. Automated Pixel-Level Pavement Crack Detection on 3D Asphalt Surfaces Using a Deep-Learning Network. *Comput.-Aided Civ. Infrastruct. Eng.* **2017**, *32*, 805–819. [[CrossRef](#)]
34. Xue, Y.; Li, Y. A Fast Detection Method via Region-Based Fully Convolutional Neural Networks for Shield Tunnel Lining Defects. *Comput.-Aided Civ. Infrastruct. Eng.* **2018**, *33*, 638–654. [[CrossRef](#)]
35. Ding, L.; Fang, W.; Luo, H.; Love, P.E.D.; Zhong, B.; Ouyang, X. A deep hybrid learning model to detect unsafe behavior: Integrating convolution neural networks and long short-term memory. *Autom. Constr.* **2018**, *86*, 118–124. [[CrossRef](#)]
36. Fang, W.; Zhong, B.; Zhao, N.; Love, P.E.D.; Luo, H.; Xue, J.; Xu, S. A deep learning-based approach for mitigating falls from height with computer vision: Convolutional neural network. *Adv. Eng. Inform.* **2019**, *39*, 170–177. [[CrossRef](#)]
37. Awang, S.; Azmi, N.M.A.N. Vehicle Counting System Based on Vehicle Type Classification Using Deep Learning Method. In *IT Convergence and Security 2017*; Springer: Singapore, 2018; pp. 52–59.
38. Xu, Y.; Yu, G.; Wang, Y.; Wu, X.; Ma, Y. Car Detection from Low-Altitude UAV Imagery with the Faster R-CNN. *J. Adv. Transp.* **2017**, *2017*, 1–10. [[CrossRef](#)]
39. Biswas, D.; Su, H.; Wang, C.; Blankenship, J.; Stevanovic, A. An Automatic Car Counting System Using OverFeat Framework. *Sensors* **2017**, *17*, 1535. [[CrossRef](#)]
40. Lee, W.-J.; Kim, D.; Kang, T.-K.; Lim, M.-T. Convolution Neural Network with Selective Multi-Stage Feature Fusion: Case Study on Vehicle Rear Detection. *Appl. Sci.* **2018**, *8*, 2468. [[CrossRef](#)]
41. Liu, W.; Zhang, M.; Luo, Z.; Cai, Y. An Ensemble Deep Learning Method for Vehicle Type Classification on Visual Traffic Surveillance Sensors. *IEEE Access* **2017**, *5*, 24417–24425. [[CrossRef](#)]

42. Tsai, C.; Tseng, C.; Tang, H.; Guo, J. Vehicle Detection and Classification based on Deep Neural Network for Intelligent Transportation Applications. In Proceedings of the Asia Pacific Signal and Information Processing Association Annual Summit and Conference, Honolulu, HI, USA, 12–15 November 2018; pp. 1605–1608.
43. Wang, X.; Zhang, W.; Wu, X.; Xiao, L.; Qian, Y.; Fang, Z. Real-time vehicle type classification with deep convolutional neural networks. *J. Real-Time Image Process.* **2019**, *16*, 5–14. [[CrossRef](#)]
44. Xiang, X.; Lv, N.; Guo, X.; Wang, S.; El Saddik, A. Engineering Vehicles Detection Based on Modified Faster R-CNN for Power Grid Surveillance. *Sensors* **2018**, *18*, 2258. [[CrossRef](#)]
45. Kim, J.; Chi, S.; Seo, J. Interaction analysis for vision-based activity identification of earthmoving excavators and dump trucks. *Autom. Constr.* **2018**, *87*, 297–308. [[CrossRef](#)]
46. Golparvar-Fard, M.; Heydarian, A.; Niebles, J.C. Vision-based action recognition of earthmoving equipment using spatio-temporal features and support vector machine classifiers. *Adv. Eng. Inform.* **2013**, *27*, 652–663. [[CrossRef](#)]
47. Nguyen, B.; Brilakis, I.; Vela, P.A. Optimized Parameters for Over-Height Vehicle Detection under Variable Weather Conditions. *J. Comput. Civ. Eng.* **2017**, *31*, 04017039. [[CrossRef](#)]
48. Ho, G.T.S.; Tsang, Y.P.; Wu, C.H.; Wong, W.H.; Choy, K.L. A Computer Vision-Based Roadside Occupation Surveillance System for Intelligent Transport in Smart Cities. *Sensors* **2019**, *19*, 1796. [[CrossRef](#)]
49. Memarzadeh, M.; Golparvar-Fard, M.; Niebles, J.C. Automated 2D detection of construction equipment and workers from site video streams using histograms of oriented gradients and colors. *Autom. Constr.* **2013**, *32*, 24–37. [[CrossRef](#)]
50. Rezaazadeh Azar, E.; McCabe, B. Automated Visual Recognition of Dump Trucks in Construction Videos. *J. Comput. Civ. Eng.* **2012**, *26*, 769–781. [[CrossRef](#)]
51. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
52. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Li, F.F. ImageNet: A Large-Scale Hierarchical Image Database. In Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition, Miami, FL, USA, 20–25 June 2009.
53. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2015**, arXiv:1409.1556.
54. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012.
55. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. *arXiv* **2014**, arXiv:1409.4842.
56. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1800–1807.
57. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
58. Pu, Y.; Apel, D.B.; Szmigielski, A.; Chen, J. Image Recognition of Coal and Coal Gangue Using a Convolutional Neural Network and Transfer Learning. *Energies* **2019**, *12*, 1735. [[CrossRef](#)]
59. Sun, C.; Yang, Y.; Wen, C.; Xie, K.; Wen, F. Voiceprint Identification for Limited Dataset Using the Deep Migration Hybrid Model Based on Transfer Learning. *Sensors* **2018**, *18*, 2399. [[CrossRef](#)]
60. Hasan, M.J.; Kim, J.-M. Bearing Fault Diagnosis under Variable Rotational Speeds Using Stockwell Transform-Based Vibration Imaging and Transfer Learning. *Appl. Sci.* **2018**, *8*, 2357. [[CrossRef](#)]
61. Izadpanahkakhk, M.; Razavi, S.; Taghipour-Gorjilaie, M.; Zahiri, S.; Uncini, A. Deep Region of Interest and Feature Extraction Models for Palmprint Verification Using Convolutional Neural Networks Transfer Learning. *Appl. Sci.* **2018**, *8*, 1210. [[CrossRef](#)]

62. Liu, S.; Deng, W. Very deep convolutional neural network based image classification using small training sample size. In Proceedings of the Asian Conference on Pattern Recognition, Kuala Lumpur, Malaysia, 3–6 November 2015; pp. 730–734.
63. Zhang, Y.; Wang, G.; Li, M.; Han, S. Automated Classification Analysis of Geological Structures Based on Images Data and Deep Learning Model. *Appl. Sci.* **2018**, *8*, 2493. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).