Research on Vehicle Multidimensional Feature Recognition Technology Based on Cascade Convolutional Neural Network

Helin Zhu[a], Hua Pan[b]

Guangxi transportation science and technology group co.,LTD. Nanning 530007, China

[a]244059419@qq.com, [b]597310014@qq.com

Keywords: ETC, Vehicle face recognition, Deep learning, Convolutional neural network

Abstract: After the cancellation of provincial toll stations, a large amount of abnormal charging data appeared in actual applications. These data have more or less OBU information missing or even lack of image information, which cannot be accurately deducted. In view of this situation, this paper proposes a vehicle multi-dimensional feature recognition method based on cascaded convolutional neural network, which constructs the vehicle face feature information, and can subsequently restore the trajectory of abnormal vehicles based on these information. Finally, the experimental results show that the method can quickly and accurately find vehicles with abnormal OBU or suspected evasion of expenses, and identify their trajectories, thereby achieving more accurate billing.

1. Introduction

After the gradual implementation of the cancellation of provincial toll stations, revolutionary changes have taken place in China's highway toll collection methods. On the one hand, it has greatly facilitated vehicle owners, and on the other, it has brought new challenges to the toll management department. Due to the inherent shortcomings of ETC technology when the vehicle passes through the gantry at high speed, and the increase of one-way mileage, the increase of the single charge amount, and the simplification of the OBU issuance process, the owner ’s willingness and motivation to evade fees have increased. The superposition of factors makes a large amount of abnormal charging data appear in actual applications, which cannot accurately identify vehicle inbound and outbound information, vehicle trajectory information, and then accurate deductions. These data have more or less OBU information missing or even lack of image information. How to quickly and accurately realize the trajectory restoration of tolled vehicles based on these abnormal data will have great practical application value for vehicles at the entrance and exit.

Deep neural network technology, also known as deep learning (DL) technology, is a new research direction in the field of machine learning (ML), and is the core foundation of the third wave of artificial intelligence that has developed rapidly in recent years. Deep learning combines low-level features to form more abstract high-level representation attribute categories or features to discover distributed feature representations of data. The motivation for studying deep learning is to build a neural network that simulates the human brain for analysis and learning. It mimics the mechanisms of the human brain to interpret data, such as images, sounds, and text [2]. After nearly 10 years of rapid development, deep learning has brought revolutionary changes to the field of computer vision. Among them, the most widely used is face recognition [3]. Vehicle recognition in traffic scenarios has always been another widely used field of computer vision recognition, including vehicle license plate recognition, vehicle detection, vehicle tracking, and traffic abnormal behavior analysis. There is no doubt that the application of deep learning technology in the field of vehicle recognition will also make a very big breakthrough.

Convolutional neural network (CNN) is a type of feedforward neural network that includes convolution calculations and has a deep structure. It is one of the representative algorithms for deep learning [4, 5]. CNN has achieved unprecedented great success in the field of computer vision [6]. It includes one-dimensional convolutional neural network, two-dimensional convolutional neural network and three-dimensional convolutional neural network. After nearly ten years of development, many variants of convolutional neural networks have appeared, mainly used for image classification, target detection, and semantic segmentation. For image classification, it mainly includes the LeNet mentioned above, and later AlexNet, VGG [7], GoogleNet, ResNet [8], Inception series, etc. The target detection includes Faster-RCNN, SSD, Yolo-v1, Yolo-v2, Yolo-v3, etc.

This paper proposes a vehicle multi-dimensional feature recognition method based on a cascaded convolutional neural network. It uses the most advanced deep convolutional neural network technology to achieve “vehicles face recognition" for vehicles with up to 4,000-dimensional features. It can accurately identify features such as vehicle toll models, vehicle brand models, vehicle color models, vehicle interior and exterior trims, etc., and based on big data technology and GIS technology to achieve multi-point image comparison analysis of vehicle entrances and exits and gantry. Thereby, it can realize the analysis and restoration of vehicle running trajectory, consistency check of OBU information and image identification information, etc. The method provided in this article can effectively prevent vehicle evasion, correct and restore abnormal data, so as to achieve more accurate billing and recover economic losses.

2. Identification requirements of highway vehicles

At present, due to the inaccurate identification of vehicles on the expressway, and the driver's evasion of tolls has also caused great economic losses [9]. Because of the needs of traffic management, accurate identification of vehicles has always been a key part of intelligent traffic management. In the past, a large number of license plate recognition systems were applied on the roads in China to identify the license plates of passing vehicles. Also, there are many studies on license plate recognition in various academic papers [10, 11, and 12].

Although academic research has shown that in a laboratory environment, license plate recognition can reach 100% accuracy. However, in actual scenes due to environmental lighting, shooting angles, excessive camera explosion, etc., the accuracy of license plate recognition often does not achieve satisfactory results. In practice, there are often abnormalities such as defaced license plates, unsuspended license plates, and blocked license plates. This makes it impossible for a single license plate to express the unique identity of a vehicle in the process of highway toll collection. After the provincial border toll station is cancelled, ETC will become the main vehicle identity, and the license plate will serve as a secondary identity. However, after a period of operation, we found in the actual system that a large amount of abnormal data still exists in the system operation, mainly in the following categories:

(1) There is OBU data, but the vehicle model or license plate does not match the information recorded by OBU. This situation is mainly caused by the owner's initiative or negligence, causing the vehicle to hang with inconsistent tags. For example, in order to reduce the toll, the large car uses a small car tag, or the owner changes the license plate without changing the OBU.

(2) There is no OBU data, but there are license plate recognition results. This situation may be caused by excessive vehicle speed or the failure of the gantry system ETC antenna, the vehicle OBU information cannot be read, or the owner intentionally destroys or shields the vehicle OBU in order to evade expenses.

(3) There is no OBU data and no license plate recognition results, but there are actually vehicles passing through. This may be because the vehicle speed is too fast, and the gantry system ETC antenna does not read the vehicle information, or the owner evades the fee, intentionally destroys or shields the vehicle OBU, and blocks the number plate.

The above situations will cause the toll system to undercharge, mischarge or owe accounts, causing huge losses to highway owners. Therefore, finding a feature other than a license plate is particularly necessary to identify the identity of the vehicle. Combining the progress of deep learning technology mentioned in recent years, this paper proposes a method for vehicle face recognition based on convolutional neural networks. Based on this method, vehicle detection can be accurately performed. In addition, it can also recognize the multi-dimensional characteristics of the vehicle. It can not only accurately identify the license plate, but also the characteristics of the vehicle model, year, brand, color, and interior and exterior of the vehicle. Based on these characteristics, the identity of the vehicle can be described more accurately, and the trajectory of the vehicle can be retrieved, which can effectively counteract the above several abnormal situations.

3. Vehicle Multi-dimensional Feature Recognition Scheme Based on Convolutional Neural Network

3.1 Establishment of Sample Library

At present, deep learning is still mainly a supervised machine learning model, which requires a large number of samples for training. Considering the peculiarities of highway toll data we selected 1 million pieces of vehicle picture data detected on the highway gantry and highway toll capture data. At the same time, in order to enhance the adaptability of the data, we crawled 500,000 vehicle data from low perspectives of the vehicle through the Internet crawler. The sample data set will cover three perspectives: gantry bayonet, toll gate capture and handheld capture. We found through experiments that training by mixing data from multiple perspectives has better generalization ability than training using data from a single perspective.

3.2 Marking of samples

There are mainly four types of data tags to be marked here: license plate number, vehicle brand year, vehicle type, and vehicle interior and exterior characteristics. Considering that the image data reaches 2 million pieces, if it is manually labeled one by one, it will be very labor-intensive, so we use an iterative labeling method. For example, for the license plate number, we first use traditional license plate recognition software to recognize the license plate of the picture, and then manually correct and verify the recognition result. Because the traditional license plate recognition algorithm can basically achieve a recognition rate of more than 95%, in practice, only less than 5% of license plates need to be corrected and manually marked, which can save a lot of labor. Figure 1 below shows our vehicle identification method.

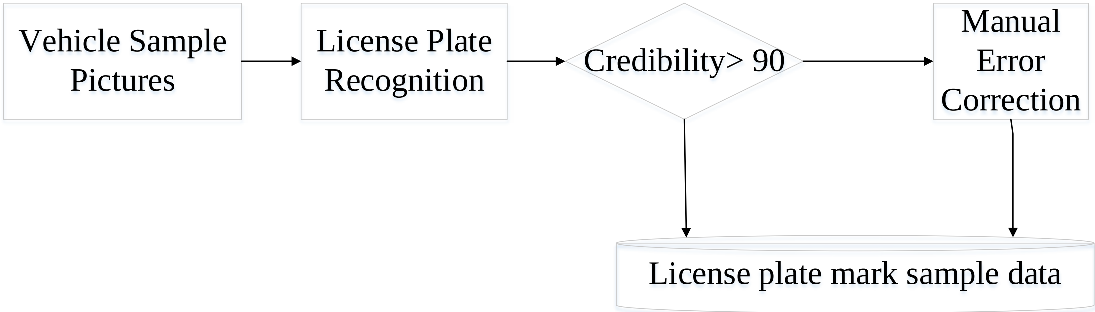

Figure 1. Vehicle identification mark

The vehicle brand year mark is the most challenging in practice, because we humans can only have the cognitive ability to very limited common brands, such as Mercedes-Benz, BMW and so on. But for vehicle sub-brands and even models, it is difficult for others to accurately identify unless they are professionals who are very familiar with vehicles. The workload would be unbearable if manual marking were used entirely. We still use the iterative labeling method of human-machine coupling to complete. Specific steps are as follows:

Step 1: build a small-scale training set V1 based on a small number of open source datasets currently on the Internet and labeled data crawled by web crawlers (these data are basically classified according to model).

Step 2: Based on the training set V1, use deep learning for vehicle detection and training to obtain the classification model M1.

Step 3: Based on the training model M1, the gantry bayonet bayonet data A1, tolled entrance and exit shots B1 are automatically identified and classified, and then the classification results are manually screened. Images with obvious errors or unconfirmable data are called out as the data set C1. Finally get data sets A2, B2.

Step 4: Combine the three data sets V1, A2, and B2, train to obtain model M2, and then re-identify C1 based on the model M2, remove the obviously wrong pictures again, and classify the correct data set as C2. Furthermore, the data sets V1, A2, B2, and C2 are obtained.

Step 5: Repeat step 4 to get relatively accurate data sets V1, A2, B2, C2, A3, B3, C3 ... Through this method of automatic machine classification and manual selection, the workload of data marking is greatly reduced, and the impossible data marking work becomes possible.

Vehicle types are classified based on the newly released standard "Vehicle classification of the toll for highway" issued by the Ministry of Transport, which classifies trucks into 6 types and passenger cars into 4 types.

The vehicle interior and exterior decoration features are mainly used to search for maps. That is, without a vehicle license plate, the vehicle can be retrieved by relying only on the picture of the vehicle. The establishment of such samples is different from the identification of the above vehicle characteristics. In the system, 5000 vehicles are taken, and each vehicle has 100 vehicle pictures in different positions. A total of 500,000 images are used as training samples. In order to simulate the actual situation to the greatest extent, we artificially obliterate the vehicle picture, as shown in Figure 2 below:

Figure 2. Picture of vehicle without license plate

3.3 Network Structure Design and Training

In the process of highway tolling, the accuracy rate of vehicle identification is very important to the accuracy of toll collection. Therefore, how to design a good framework and process is very critical. In comparison, the speed of vehicle recognition is less important, and in practice, the performance of recognition can be improved through distributed architecture, edge computing, and other methods. Therefore, the network design in this paper is mainly based on accuracy.

Because of the lot of background images in the image, there are six steps in the vehicle detection and recognition process: first, target detection, then vehicle model, brand, and color recognition, followed by license plate recognition, and finally vehicle feature extraction. Different from the multitask training method mentioned in paper [13], the method in this paper has better guarantee in recognition accuracy.

(1) Vehicle Detection Module

Object detection algorithms based on convolutional neural networks have developed very rapidly in recent years. Among them, the more famous frameworks are Faster-RCNN, SSD, and Yolo. By far, the most commonly used framework is the Yolo-v3 framework [14]. Yolo-v3 changed the softmax to logistic when performing object classification, instead of using softmax when predicting object categories, it changed to using logistic output for prediction, which can support multi-label objects. Yolo-v3 draws on the residual network structure to form a deeper network layer and multi-scale detection, which improves the detection effect of mAP and small objects. When using COCO mAP50 as the evaluation index, Yolo-v3 performed very well. This paper uses Yolo-v3 to detect vehicles.

(2) Vehicle Identification Module

Vehicle recognition is mainly based on the boundary position of the target detection to obtain the sub-picture of the vehicle, and then classify and identify the sub-picture of the vehicle based on the convolutional neural network. Residual Network (ResNet for short) is arguably the most groundbreaking work in the field of computer vision and deep learning in the past few years. ResNet makes it possible to train hundreds or even thousands of layers and still show superior performance in this case. Because of its powerful representation capabilities, in addition to image classification, many computer vision applications, including object detection and face recognition, have achieved performance improvements. This article chooses the Resnet18 method.

Triplet Loss, originally proposed in *Facenet: A unified embedding for face recognition and clustering* [15]. Considering that we need to output vehicle types, brands, models, and colors at the same time, this article builds a Resnet network based on multi-tasking. Compared with the multi-tasking framework, the multi-tasking cascading network can achieve a better recognition rate. Of course, it loses a certain performance. Fortunately, this application does not require high delay, and the accuracy is more important.

(3) License Plate Recognition Module

When the convolutional neural network refreshes higher accuracy in various visual recognition tasks, it can also be significantly improved when applied to license plate recognition. This paper adopts LPRNet [16] for license plate recognition, and presents an end - to - end automatic license plate recognition method, which is not necessary to segment Initial character before recognition. The method uses a deep neural network, capable of real-time computation at a speed of 3ms per license plate on nVIDIA GeForce GTX 1080 graphics CARDS and 1.3ms per license plate on Intel core i7-6700k. LPRNet consists of a lightweight convolutional neural network, so it can be trained end-to-end.

LPRNet's basic network building Blocks were inspired by SqueezeNet Fire Blocks [17] and Inception Blocks [18]. Unfortunately, LPRNet only solves the problem of recognition. Lbp-cascade is used for license plate detection, which makes this method less than perfect. In conclusion, we adopted Yolo-v3 to realize the detection of the vehicle and license plate position in the picture in the aforementioned vehicle detection process, and then input the license plate position to LPRNet, which not only combined with the detection advantages of Yolo-v3, but also combined with the recognition advantages of LPRNet.

(4) Vehicle Feature Extraction

In this paper, the 512-dimensional feature vector output by the last avgpool of Resnet18 is adopted as the vehicle feature. In the feature retrieval, we use the cosine distance (cosine similarity) as a measure of similarity between two vehicle images. In simple terms, cosine similarity is the cosine of the Angle between two vectors. The cosine distance is 1 minus the cosine similarity. Assume the two vehicles output characteristic vector for A, B, A, B direct cosine similarity is:

$$\cos(\theta) = \frac{\sum_{i=1}^{N} A_i B_i}{\sqrt{\sum_{i=1}^{N} A_i} \sqrt{\sum_{i=1}^{N} B_i}}$$

The corresponding cosine distance is:

$$D(A,B) = 1 - \cos(\theta) = 1 - \frac{\sum_{i=1}^{N} A_i B_i}{\sqrt{\sum_{i=1}^{N} A_i} \sqrt{\sum_{i=1}^{N} B_i}}$$

In feature retrieval, it is only necessary to find the corresponding distance between the feature vectors of the vehicles to be retrieved and the feature vectors of all the vehicles in the picture set, and then sort them. The vehicles with the smallest distance are the corresponding vehicles.

3.4 Analysis of Experimental Results

The experimental algorithm in this paper is written by the programming language Python and C++, based on the deep learning framework Caffe and the image processing library OpenCV, and trained by the data of 2 million pictures. The ratio of training set and verification set is 9:1, that is, 1.8 million training images and 200,000 verification sets. During the training, we used the model based on ImageNet pre-training to initialize the parameters and fine-tune the parameters to prevent the network from falling into the local optimal solution.

Based on the algorithm in this paper, the accurate recognition of vehicles, vehicle model recognition, vehicle brand recognition, vehicle feature vector extraction can be realized.

The accuracy of the vehicle feature vector is calculated by the first hit in the map search. Through the random sampling of 10,000 vehicles among the passing vehicles of 10 million vehicles every day in GuangXi province, the results are shown in the following table 1.

Table 1. Accuracy of vehicle eigenvector

	Detection rate	Vehicle model accuracy	Brand accuracy	Car model accuracy	Color accuracy	License plate	Feature accuracy
Our	99.87%	99.2%	96.3%	86%	97.2%	98.6%	97%
Current	98%	85%	—	—	84%	91.7%	—

'Our' in table 1 refers to the statistical results obtained based on the method identified in this paper, 'Current' in table 1 refers to the output results of the front-end camera in the original system. We can see that we have significantly exceeded the original system not only in terms of recognition items, but also in terms of recognition accuracy, which provides a strong data basis for subsequent vehicle trajectory analysis and abnormal data, and provides a guarantee of searching by image for the unlicensed car, blocking the number plate of the vehicle.

4. Application of Car Face Recognition in Highway Toll Collection

Based on the vehicle face recognition results in this paper, the vehicle license plate number, brand, model and vehicle characteristic value can be accurately identified. Each vehicle can be described by the following structure:

$$CarInfo\colon \{PlateNo, Brand, Type, Color, Feature\}$$

In order to realize the trajectory analysis of vehicles, we added the location longitude, latitude, shooting time and other information for each vehicle

$$Car\colon \{CarInfo, SnapshotTime, SnapshotLocation\}$$

After recording these information into the database, the system is based on big data analysis and GIS spatial comparison analysis. Specifically, for several abnormal data mentioned earlier in this paper:

(1) OBU data is available, but the vehicle model or license plate does not match the information recorded by OBU.By adding the car face recognition mentioned in this paper, the system can accurately identify the vehicle model, brand and license plate, and then make a comparison with OBU data. If the car model or license plate is found to be inconsistent, it will be directly written to the blacklist.

(2) No OBU data, but license plate recognition results existed. In view of this situation, the system can identify the track based on the vehicle license plate recognition results and the capture position, and then define the vehicle fare standard based on the vehicle recognition results. After manual audit, the vehicle cost can be calculated and then synchronized to the charging system.

(3) The vehicle pass through the camera without OBU data license plate recognition result. In view of this situation, the system carries out vehicle retrieval based on vehicle characteristics to obtain similar vehicles, and then combines the vehicle snapshot time and position to artificially confirm the authenticity of the vehicle, and finally fits the vehicle trajectory.

The following figure shows several methods of handling abnormal data.

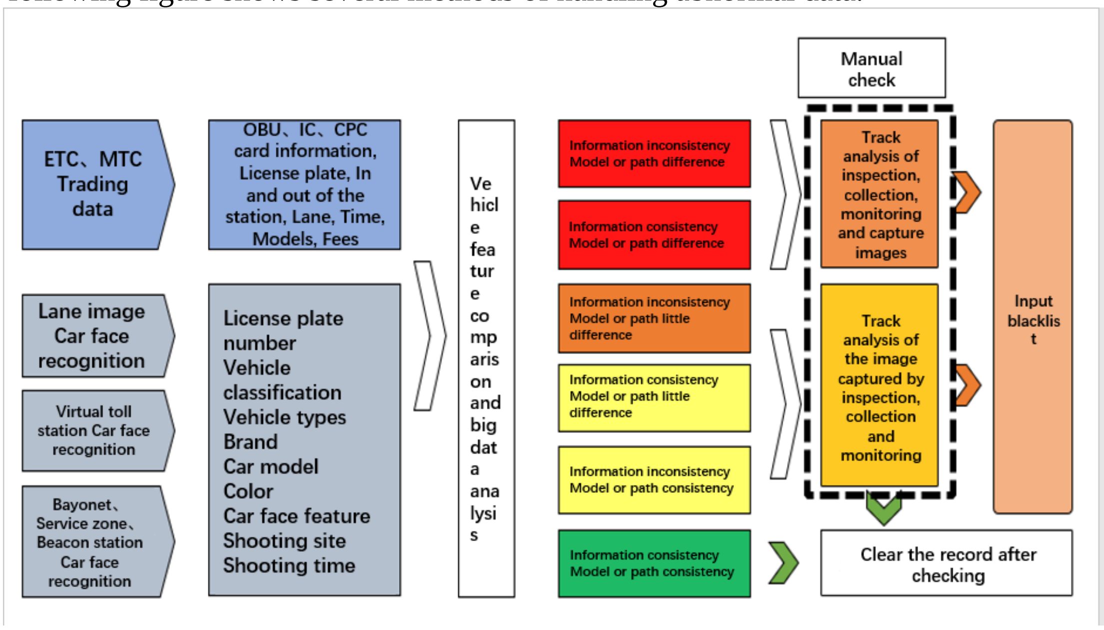

Figure 3. Abnormal data processing method

5. Conclusion

After the gradual implementation of cancelling the toll stations in provincial boundaries, combined with the characteristics of expressway toll collection after cancelling the toll stations in provincial boundaries, this paper provides two methods for the various vehicle evasive charges happened on the expressway, and the abnormal vehicle data caused by technical reasons. One is the data marking method for training convolutional neural network. The other is multi-dimensional feature recognition of vehicles based on cascade convolutional neural network is proposed, and the feature information of vehicle's vehicle face is constructed by training a large amount of actual data, which can be used to realize trajectory restoration of abnormal vehicles. Measured results show that the proposed method is not only be able to achieve far more than the existing system of recognition accuracy, but also implement the OBU exception or suspected escape detection and track of the vehicle identification by multidimensional characteristics, and then achieve more accurate charging, effectively combat evasion, and reduce amount of abnormal data produced by the technical reasons, to provides a good technical support for our province highway fee.

References

[1] MA Ji. 2019. The risk of cancelling the toll stations in provincial boundaries its countermeasures. *Communications Science and Technology Heilongjiang*, Vol. 10, No. 308(214-216).

[2] WEI Jian, LIU Aijuan, TANG Jianwen. 2017. Research on alarm model of digital TV monitoring platform based on deep learning neural network technology. *Cable Television Technology*, Vol. 7(78-82).

[3] YANG Jucheng, LIU Na, Fang Shanshan, XIE Ying. 2016. Review of Face Recognition Methods Based on Deep Learning, *Journal of Tianjin University of Science & Technology*, Vol. 31. No. 6(1-10).

[4] Goodfellow I, Bengio Y, Courville A. 2016. *Deep learning* (Vol. 1). Cambridge: MIT press (326-366).

[5] GU Jiuxiang, WANG Zhenhua, et al. 2015. Recent advances in convolutional neural networks. *Computer Science*.

[6] Hadji I, Wildes R P. 2018. What Do We Understand About Convolutional Networks? [J].

[7] Simonyan K, Zisserman A. 2014. Very deep convolutional networks for large-scale image recognition [J]. *arXiv preprintarXiv*:1409.1556, 2014.

[8] He K, Zhang X, Ren S, et al. 2016. Deep residual learning for image recognition[C]. *Proceedings of the IEEE conference on computer vision and pattern recognition*, (770-778).

[9] FAN Yongpan. 2019. The method that manages highway evades charge after cancelling the toll stations in provincial boundaries. *Highways & Automotive Applications [J]*, Vol. 1, No. 160(168-170).

[10] DING Jinchao, ZHANG Weiwei, WU Xuncheng. 2018. License plate recognition algorithm based on bidirectional long - term memory network [J]. *Journal of Electronic Measurement and Instrument*, Vol. 6.

[11] ZHAO Hanli, LIU Junru, JIANG Lei, et al. 2019. A two-line license plate segmentation algorithm based on convolutional neural network [J]. *Journal of Computer-Aided Design & Computer Graphics*, Vol. 8.

[12] LI Xiangpeng, MIN Weidong, HAN Qing, et al. 2019. License plate location and recognition method based on deep learning [J]. *Journal of Computer-Aided Design & Computer Graphics*, Vol. 6.

[13] WANG Yaowei, TANG Lun, LIU Yunlong, Chen Qianbin, et al. 2018. Vehicle multi-attribute recognition based on multi-task convolutional neural network. *Computer Engineering and Applications*, Vol. 54, No. 8(21-27).

[14] Redmon J, Farhadi A. 2018. YOLOv3: An Incremental Improvement [J].

[15] Schroff F, Kalenichenko D, Philbin J. Facenet. 2015. A unified embedding for face recognition and clustering[C] //CVPR. 2015: 815-823.

[16] Zherzdev S, Gruzdev A. 2018. LPRNet: License Plate Recognition via Deep Neural Networks [J]. arXiv: 1806.10447.

[17] SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5mb model size, arXiv: 1602.07360.

[18] Inception-v4, Inception-ResNet and the Impact of Residual Connections on learning arXiv: 1602.07261.