International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-4, November 2019

Distracted Driver Detection with Deep Convolutional Neural Network

O. G. Basubeit, D. N. T. How, Y. C. Hou, K. S. M. Sahari

***Abstract:** According to the World Health Organization (WHO), over 1.3 million deaths occur worldwide each year due to traffic accidents alone. This figure elevates traffic mishaps to be the eight leading cause of death. According to another study the United States National Highway Traffic Safety Administration (NHTSA), the major cause of road deaths and injury is distracted drivers. Motivated by recent advancement of deep learning and computer vision in predicting drivers' behaviour, this paper attempts to investigate the optimal deep learning network architecture to accurately detect distracted drivers over visual feed. Specifically, a thorough evaluation and detailed benchmark comparisons of pretrained deep convolutional neural network is carried out. Results indicate that the proposed VGG16network architecture is capable of achieving 96% accuracy on the test dataset images.*

***Keywords:** Deep convolutional neural network, transfer learning, pretrained model, online dataset.*

I. INTRODUCTION

Distracted driving refers to any activity that turns the driver away from the safe driving task. Examples of activities that may cause distraction during driving are texting or talking on mobile phone, adjust the radio, drinking, talking to a passenger and hair or makeup [1]. In the past few decades, the standard of living has improved that caused the families to possess their own cars. Thus, people usually drive their cars for business or trip. However, long distance driving usually makes drivers bored. However, some drivers do some other activities rather than focusing on driving which causes a lot of accidents. Since the last few years the number of car accidents has increased because of distracted driving. National Highway Traffic Safety Administration of United State (NHTSA) has been announced 3,477 deaths and 391,000 injuries cases in 2015 due to distracted driving [2]. In US, distracted driving considered as a major killed reason by a daily rate of 9 cases and 1000 injured cases [3].

Manuscript published on November 30, 2019.
* Correspondence Author
O. G. Basubeit, College of Engineering, Universiti Tenaga Nasional, Jalan IKRAM-UNITEN, 43000 Kajang, Selangor, Malaysia
D. N. T. How, College of Engineering, Universiti Tenaga Nasional, Jalan IKRAM-UNITEN, 43000 Kajang, Selangor, Malaysia
Y. C. Hou, Institute of Informatics and Computing in Energy, Universiti Tenaga Nasional, Jalan IKRAM-UNITEN, 43000 Kajang, Selangor, Malaysia
K. S. M. Sahari, Institute of Informatics and Computing in Energy, Universiti Tenaga Nasional, Jalan IKRAM-UNITEN, 43000 Kajang, Selangor, Malaysia

At present, major companies are starting to work with new systems called Advanced Driver Assistance Systems (ADAS) by creating techniques to alert the driver when detection activities are occur. From this deep learning takes part of ADAS techniques.

Deep learning (DL) is a study of artificial neural networks and other Machine learning algorithms that hold more than one hidden layer.

Deep learning has various structures such as Convolutional Neural Networks (CNNs), Deep Neural Networks (DNNs) and Recurrent Neural Networks (RNNs). Deep learning has been applied to many fields such as computer vision, speech recognition and natural language processing [4]. According to the Centre for Disease Control and Prevention (CDC) [5] among types of driver distractions are visual, manual and cognitive. Visual distraction refers to any activity that takes the drivers eyes on the road. Example of visual distraction includes checking mobile phone for notifications, looking for items in the car, and observing billboards. Cognitive distraction refers to any activity that causes the mind to lose focus on the driving. Examples of cognitive distractions are listening to the radio, talking to passengers in the car, or being lost in thought. Manual distraction refers to activities that cause the driver to take their hands off the steering wheel. Examples of manual distractions include adjusting the stereo player, eating while driving, or smoking. As shown in Figure 2, multiple types of distractions can coexist together in a task. For example, talking on a mobile phone while drive entails manual and cognitive distractions. To date, there are three primary types of modality used to recognize distracted drivers [5]:

a) Physiological data such as electrocardiogram (ECG) and electroencephalogram (EEG).

b) Vehicle control data such as pedal positions and steering wheel movements.

c) Visual data such as eye movements, body movements and images or videos of the driver's facial expressions.

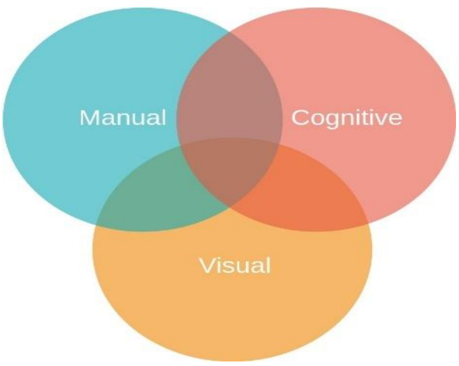

Fig. 1 Types of driving distraction

This study proceeds to focus on the detection of distracted drivers via visual data. There are various methods used to detect distracted drivers with computer vision. Among the most notable are [5]:

a) Threshold: the simplest way, which the value of a feature is compared to a predetermination threshold.

b) Traditional machine learning: such as Support Vector Machine (SVM).

c) Deep learning: deep learning has gained more attention recently in distraction detection.

There are many recent studies that addresses the problem of detecting distracted drivers. Jin et al. designed a system to detect cognitive distraction using only vehicle control data [6].The authors claimed that distracted drivers usually move the steering wheel and apply the pedals differently than normal drivers. Tabriz et al. used a percentage of eye closure (PERCLOS) to detect driver drowsiness [7].The frequency and duration of a driver's eye glimpse for a minor task are used to produce a total measurement of eyes off the road [8].Pohl et al. proposed a distracted driving detection system based on gaze direction and head position [9].The instantaneous distraction was determined and the distraction level was classified. Cray et al. introduced a hidden Markov model (HMM) based method to detect distracted driving activities. Their method requires the detection of the driver's face and right arm [10]. Abouelnag et al. provided a dataset and introduced a real-time distraction detection approach using a combination of five Alex Net and five Google Net models with hand, face, and skin features [11]. In this study, main focus is placed on detecting driver's behaviour via a webcam installed in the car cabin. The webcam field-of-view covers the upper body of the driver sitting on the driver's seat. Deep learning is employed to classify the images taken by the webcam and determine whether or not the driver is distracted.

II. DATASET DESCRIPTION

The dataset used in the study is taken from a public Kaggle challenge by State Farm [12].The dataset consists of 22,400 training and 79,727 validation labelled images (640x480 full colour) of driver behaviours. There are a total of ten classes of behaviours provided in the dataset. Table 1 tabulates the ten distinct behaviours. Figure 2 illustrate a sample image for each of the ten classes.

Table. 1 List of distracted driving activities

Class	Behaviour
C0	Safe driving
C1	Texting using right hand
C2	Talking on the phone using right hand
C3	Texting using left hand
C4	Talking on the phone using left hand
C5	Operating (Adjusting) the radio
C6	Drinking
C7	Reaching behind
C8	Hair and makeup
C9	Talking to passenger

Fig. 2 Sample images from the Kaggle State Farm dataset [12]

III. PROPOSED DEEP LEARNING MODEL

Convolutional Neural Network

Convolutional Neural Network (CNN) is a category of deep learning model that has proven to be very effective in areas such as image recognition and classification. CNN have been successful in identifying faces [13],objects and traffic signs apart from powering vision in robots and self-driving cars [14]. Figure 3 illustrates the basic architecture of the CNN. The CNN architecture involves multiple layers of operations performed on the input image.

A typical CNN model includes multiple convolutions followed by a pooling operation one after another. These layers of convolution and pooling in succession if also known as the feature extraction layers. The resulting vectors from the feature extraction layers are then flattened and stacked to the classifier layers.

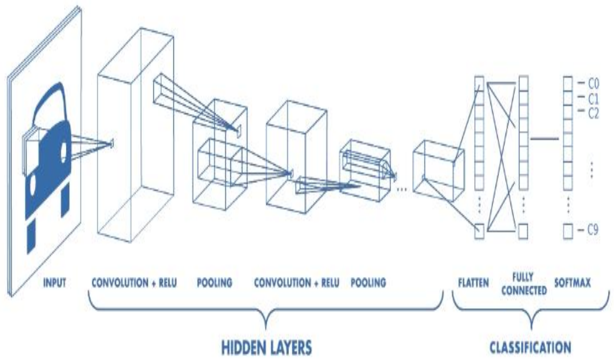

Fig. 3 Architecture of a typical CNN for image classification [15]

Transfer Learning

Transfer learning is a technique in deep learning where one model that is trained on a task is re-purposed to fit another task. For example a model that is trained on images of household objects can be re-purposed to classify the types of furniture. All the models used in this study is originally trained on the Image net object classification dataset which Comprised of 1,000 classes of objects. In order to apply transfer learning, the weights on the top layers of the pretrained models are retrained with the State Farm dataset. These are three general strategies to retrain the pretrained models using the State Farm dataset of distracted drivers. The first strategy is to retrain only the last classifier layer of the pretrained model. This is the most straight forward strategy using the least amount of time and computational power in exchange for model accuracy. The second strategy is to retrain the last few layers of the model including the classifier layer. This strategy may result in improved accuracy as well as more intensive computational power and takes longer time to run. The last strategy is to retrain the entire model from scratch. This is the least desirable method because it takes the most time and computational cost. Figure 4 illustrates the differences among the three elaborated strategies. In this study Strategy 1 (S1) and Strategy 2 (S2) is thoroughly evaluated on each of the pretrained model.

Table. 2 List of pre-trained models evaluated [16]

Model	Input size (W x H)	Depth	License
Xception [17]	299 x 299	126	MIT
VGG19 [18]	224 x 224	26	Oxford
VGG16 [18]	224 x 224	23	Oxford
ResNet50 [19]	224 x 224	168	MIT
InceptionV3 [20]	299 x 299	159	Apache
InceptionResNetV2 [21]	299 x 299	572	Apache
MobileNet [22]	224 x 224	88	Apache
DenseNet121 [23]	224 x 224	121	BSD 3
DenseNet169 [23]	224 x 224	169	BSD 3
DenseNet201 [23]	224 x 224	201	BSD 3
NASNetLarge [24]	331 x 331	-	Apache
NASNetMobile [24]	224 x 224	-	Apache

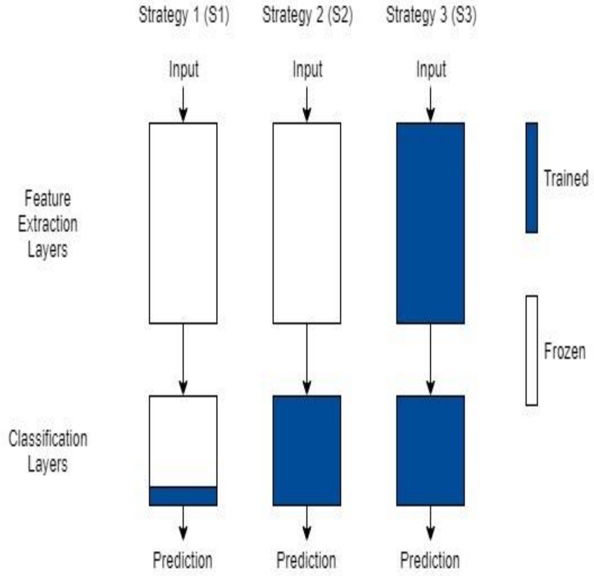

Fig. 4 Blocks in blue indicate the layers that are retrained during transfer learning while the blocks in white are left intact during transfer learning

Throughout the study, a high-level deep learning package known as Keras [25] is extensively used for rapid experimentation of pretrained CNN models. Keras is written in Python language and is capable to run on top of Tensor Flow, CNTK, or Theano. There are numerous pretrained models and weights made available by the community. These models can be used for prediction, feature extraction, and transfer learning [25]. Table 2tabulates the pretrained models used in the study.

IV. RESULTS AND DISCUSSIONS

This section highlights the results (accuracy and loss values) of all pretrained models on the training and validation sets. Due to time constraints and computational limitations, all pretrained models are retrained using S1 and S2 with only 20 epochs. Figure 5 and Figure 6, shows the accuracy and loss chart for all pretrained models. As observed in Figure 5, retraining with S2 always results in a better accuracy compared to S1. This comes at the cost of increased training time and computation cost. Comparing over all models, VGG16 with S2 yields the highest accuracy and the lowest loss rate after running for 20 epochs. In order to further evaluate the full capacity of the VGG16, the model is trained further for 100 epochs. As a result, theVGG16 is able to reach a top accuracy of 99.84% of train set and 99.11% of validation set. The loss values of the VGG16 are 0.0052 and 0.0338 for train and validation sets respectively. Figure 6 illustrates the learning curve of theVGG16.

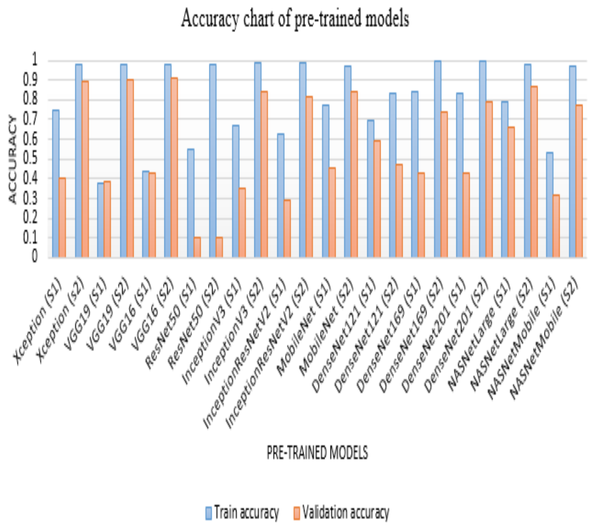

Fig. 5 Accuracy chart of 12 Keras pretrained models with two strategies

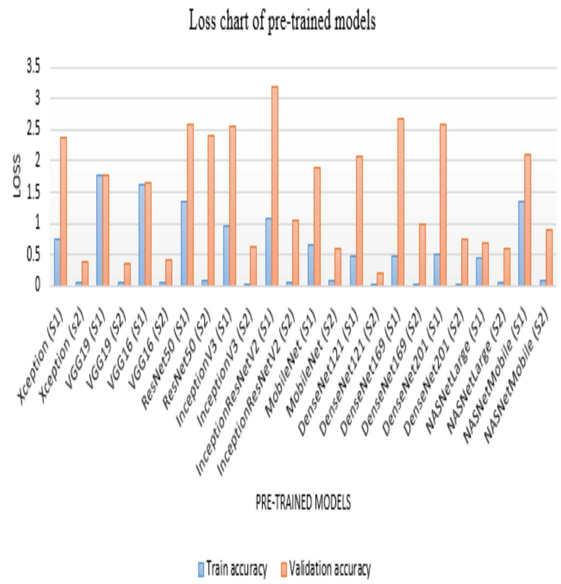

Fig. 6 Loss chart of 12 Keras pretrained models with two strategies

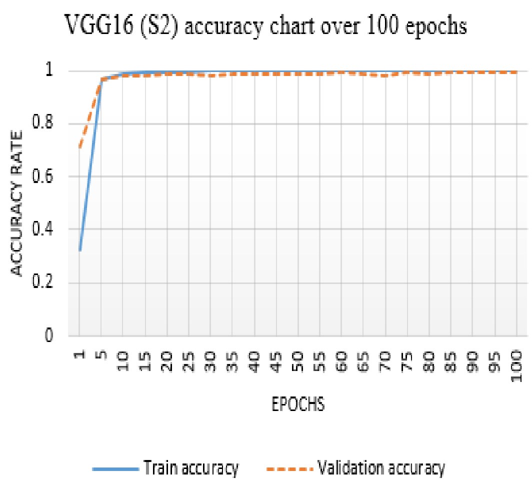

Fig. 7 Accuracy chart of VGG16 strategy 2 (S2)

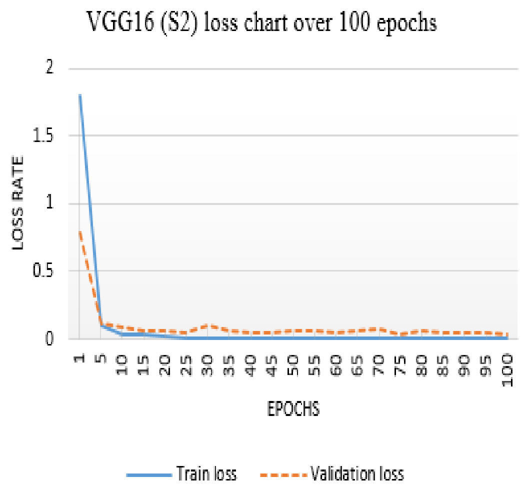

Fig. 8 Loss charts of VGG16 strategy 2 (S2)

In order to evaluate the robustness of the VGG16 model after training with 100 epochs, the model is bench marked against the test dataset. The test dataset contains 200 unseen images (20 images per class) of during the training. On the test set, the VGG16 scored 96% as overall accuracy. The model successfully predicted 196 out of the 200 novel images. Figure 9 shows the confusion matrix of the test set prediction. Table 3 tabulates the precision, accuracy, recall and the F1 score of the VGG16 on test dataset. As observed in the test dataset classification results, the VGG16 model tends to misclassify C0 (Safe driving) into C9(Talking to passenger) and C8 (Hair and makeup).

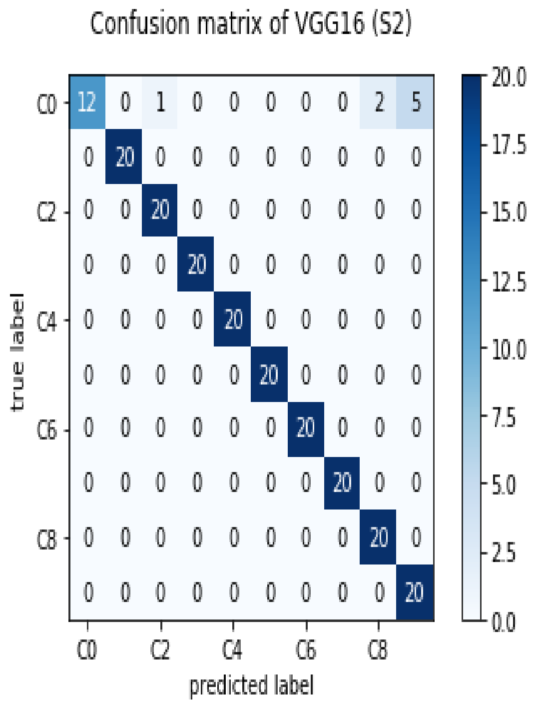

Fig. 9 Confusion matrix of the VGG16 model on the test dataset

Retrieval Number: D5131118419/2019©BEIESP
DOI:10.35940/ijrte.D5131.118419
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Table. 3 Class-wise accuracy of VGG16 on the test dataset

Class	Samples	Correct	Incorrect	Precision (%)	Truth	Accuracy (%)	Recall	F1 Score
C0 Safe driving	20	12	8	60	12	96	1.0	0.75
C1 Phone text - right	20	20	0	100	20	100	1.0	1.0
C2 Phone talk - right	20	20	0	100	20	100	1.0	1.0
C3 Phone text - left	20	20	0	100	21	99.5	0.95	0.98
C4 Phone talk - left	20	20	0	100	20	100	1.0	1.0
C5 Operating the radio	20	20	0	100	20	100	1.0	1.0
C6 Drinking	20	20	0	100	20	100	1.0	1.0
C7 Reaching behind	20	20	0	100	20	100	1.0	1.0
C8 Hair and makeup	20	20	0	100	22	99	0.91	0.95
C9 Talking to passenger	20	20	0	100	25	97.5	0.80	0.89

V. CONCLUSION

This work has demonstrated that deep learning models pretrained on the Image net dataset can be fine tuned to classify distracted driver images with good accuracy. The study evaluated twelve pretrained deep convolutional neural network models and retrained them on the State Farm dataset and evaluated its performance on the test dataset. The results indicate a positive outlook on re-purposing the VGG16 model to classify distracted drivers with up to 96% accuracy on unseen images. To summarize the findings in this study, the contribution of this paper is stated as follows:

a) Comparing over twelve pretrained CNN models the VGG16 yields the best performance despite having lesser computation layers compared to other models.

b) Training strategy S2 always yields better results on all pretrained models at the cost of time and computational resources.

In upcoming works, this study will extend the scope to involve an automated architecture search for CNN instead of relying on the predefined architectures. Recent development has shown that this may result in a more efficient and tailored model to the task at hand.

REFERENCES

1. Vasanti S, Neha P and Anniruddha H. Distracted driver detection using CNN and data augmentation Techniques. In: International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), 2018, 7(4):130-135.
2. National Highway Traffic Safety Administration (NHTSA).Distracted driving 2015. Retrieved from https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812381; March 2017.
3. Center for Disease Control and Prevention (CDC). Distracted driving. Retrieved from https://www.cdc.gov/motorvehiclesafety/distracted_driving/index.html; 30 March 2019.
4. Pariwat O. Artificial intelligence, machine learning and deep learning. In: 15th International Conference on ICT and Knowledge Engineering (ICT&KE), 2017.
5. Duy T, Ha MD, Weihua S, He B and Girish C. Real-time detection of distracted driving based on deep learning. IET Institution of Engineering and Technology, 2018, 12(10):1210-1219.
6. Lisheng J, Qingning N, Haijing H, Huacai X, Yali W and Dongdong S. Driver cognitive distractiondetection using driving performance measures. Discrete Dynamics in Nature and Society, 2012.
7. Pooneh. RTand Reza. AZ. Drowsiness detection based on brightness and numeral features of eyeimage. In: 2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IEEE, 12-14 Sep 2009.
8. Farber E, FoleyJ, ScottS. Visual attention design limits for ITS in-vehicle systems: the society of automotiveengineers standard for limiting visual distraction while driving.Transportation Research Board Annual General Meeting, Washington DC, 2000.
9. Pohl J, Birk W and Westervall L. A driver-distraction-based lane-keeping assistance system. In: Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering, 2007, 221(4):541-52.
10. Craye C and Karray F. Driver distraction detection and recognition using RGB-D sensor. arXiv preprint arXiv:1502.00250. 2015.
11. Abouelnaga Y, Eraqi HM and Moustafa MN. Real-time distracted driver posture classification. arXiv preprint arXiv:1706.09498, 2017.
12. Kaggle: State farm distracted driver detection. Retrieved fromhttps://www.kaggle.com/c/state-farm-distracted-driver-detection/data; 31 March 2019.
13. ParkhiOM, Vedaldi A, and Zisserman A. Deep face recognition. In: bmvc. 1(3):6, 2015.
14. Redmon J and Farhadi A. Yolov3: an incremental improvement.arXiv preprint arXiv:1804.02767, 2018.
15. Cornelisse D. An intuitive guide to convolutional neural networks. Retrieved from https://www.freecodecamp.org/news/an-intuitive-guide-to-convolutional-neural-networks-260c2de0a050; 24 April 2018.
16. Keras applications. Retrieved from https://keras.io/applications/; 31 March 2019.
17. Chollet F. Xception: deep learning with depthwise separable convolutions.In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1251-1258, 2017.
18. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.
19. He K, Zhang X,Ren S and Sun J. Deep residual learning for image recognition.In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778, 2016.
20. Szegedy C, Vanhoucke V, Ioffe S, Shlens J andWojna Z. Rethinking the inception architecture for computer vision.In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2818-2826, 2016.
21. Szegedy C, Ioffe S, Vanhoucke V and Alemi AA. Inception-v4, inception-resnet and the impactof residual connections on learning.In: Thirty-First AAAI Conference on Artificial Intelligence, 2017.
22. Howard AG, Zhu M, Chen B,Kalenichenko D,Wang W,Weyand T, Andreetto M andAdam H. Mobilenets: Efficient convolutional neural networks for mobile vision applications.arXiv preprint arXiv:1704.04861, 2017.
23. Huang G, Liu Z, Van Der Maaten L and Weinberger KQ. Densely connected convolutionalnetworks.In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700-4708, 2017.
24. Zoph B, Vasudevan V, Shlens J and Le QV. Learning transferable architectures for scalable image recognition.In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 8697-8710, 2018.
25. Keras: The python deep learning library. Retrieved from https://keras.io/; 31 March 2019.