Research Article

IMPLEMENTATION OF PRE-TRAINED DEEP LEARNING MODEL FOR DOG BREED CLASSIFICATION

[*]Dr. DipeshG.Kamdar[a], Dr. Nirali A. Kotak[b], Dr. Bhavin S. Sedani[c] and Dr. Komal R. Borisagar[d]

[a]Associate Professor, Electronics and Communication Department, VVP Engineering College, Rajkot
[b]Assistant Professor, Electronics and Communication Department, L.D. College of Engineering, Ahmedabad
[c]Professor, Electronics and Communication Department, L.D. College of Engineering, Ahmedabad
[d]Associate Professor, Graduate School of Engineering and Technology, Gujarat Technological University, Ahmedabad

Article History: Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 16 April 2021

Abstract: Image classification made tremendous advancement with improved techniques and its accuracy continuously improves. However when comes to fine-grained classification, huge scope of improvement. It is easy to identify the various animals from the image, but not easy to identify the animal breed. In this paper, effort is made to improve the classification of animal breed. Standard Stanford dog breed dataset is used to train and test various pre-trained deep learning models. The effort is made to fine-tune pre-train network and compare the result.

Keywords: Deep Learning Model, Fine Tuning of Pre-trained Model, Convolutional Neural Network

1. INTRODUCTION

Conventionally, Computer vision problems have been solved with the help feature extraction and comparison. However, trend is shifted towards Convolutional Neural Network (CNN) since the success of AlexNet [1] at 2012 and VGG [2] atImageNet ILSVRC-2014. Afterwards, CNN based models has been improved and many deep learned modes such as GoogleNet, ResNet [3] and DenseNet [4]. Along with the technology development, the various filters are learned and included by training a convolutional neural network on the datasets and numbers of layers are increased. GoogleNet introduces 22- layer Architecture in 2014, ResNet introduces 152 layer architecture, and that leads to many more layers to come, which will not in benefit. In order to over this issue, single fully connected layer at the end likewise FC1000. Till the date, many updated pre-trained networks are available likewise, and few of them are enlisted in the table 1. Although, these network are good, but not fine tune enough to identify the breeds of animal, as it needs little bit of up gradation and fine tuning of this network. The Stanford Dog's breed dataset is available, and is used to train and test this pre-trained deep learning model.

Dataset Description:

The dataset used in this study is Stanford Dogs dataset [5], which contains 120 dog breeds and 20580 images for training and testing. This images are split into training and test set, the proportion this split are 70 and 30 percent for training and test set respectively. Figure 1 shows some random selected image from data set.

Network	Depth	Size (MB)	Para-meters (Million)	Image Input Size
AlexNet	8	227	61.0	227x227
VGG16	16	515	138	224x224
VGG19	19	535	144	224x224
Squeeze Net	18	4.6	1.24	227x227
GoogleNet	22	27	7.0	224x224
InceptionV3	48	89	23.9	299x299
DenseNet201	201	77	20.0	224x224
MobileNetV2	53	13	3.5	224x224
ResNet18	18	44	11.7	224x224
ResNet50	50	96	25.6	224x224
ResNet101	101	167	44.6	224x224
Exception	71	85	22.9	299x299

Table 1. Various pre-trained network

2. PROPOSED METHOD

The pre-trained network models AlexNet, Google Net, DenseNet201, andResNet50, are feed with this new dataset and new features are extracted from last 3 layers to fine-tune the model. The basic architecture of deep learning model is shown in the figure 2. Convolutional Layer (CONV), Rectified Linear Unit(RELU) and Polling layer are grouped, and few layers of this group are there to process. At Final stage, Flatten Layer, Fully Connected Layer and SoftMax as a classifier are used.

Figure 1. Species of Stanford dog breeds datasets

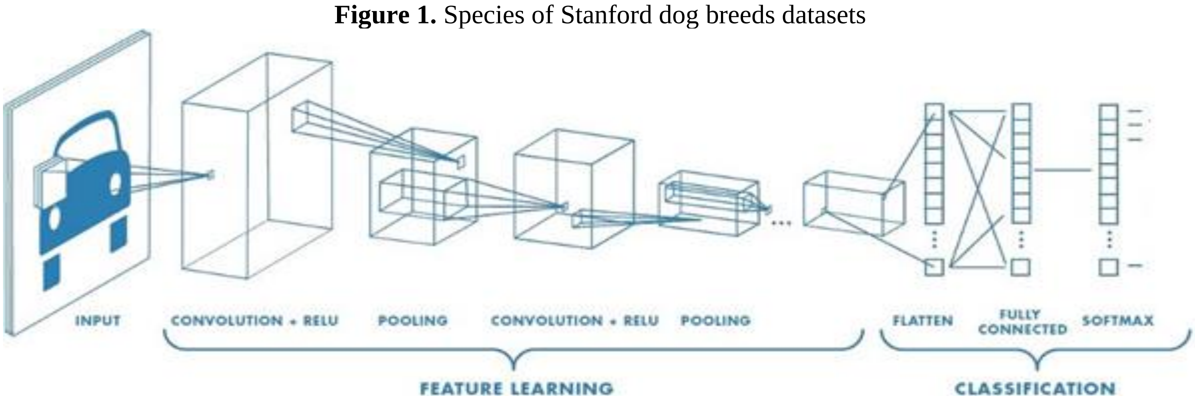

Figure 2. Basic architecture of deep learning model

The CONV is the first layer to extract features from image, and it preserves the relation between pixels by small filter mask also known as kernel. ReLU introduce non-linearity, with non-negative linear values. Pooling layers section would decrease the various parameters when the images are very large. The features which are extracted will be flattened and stored to fully connected layer. The fine tuning of the networks is done on each network, tested the network on test Stanford Dog Breeds dataset, confusion matrix is constructed and check the classifier for its correctness.

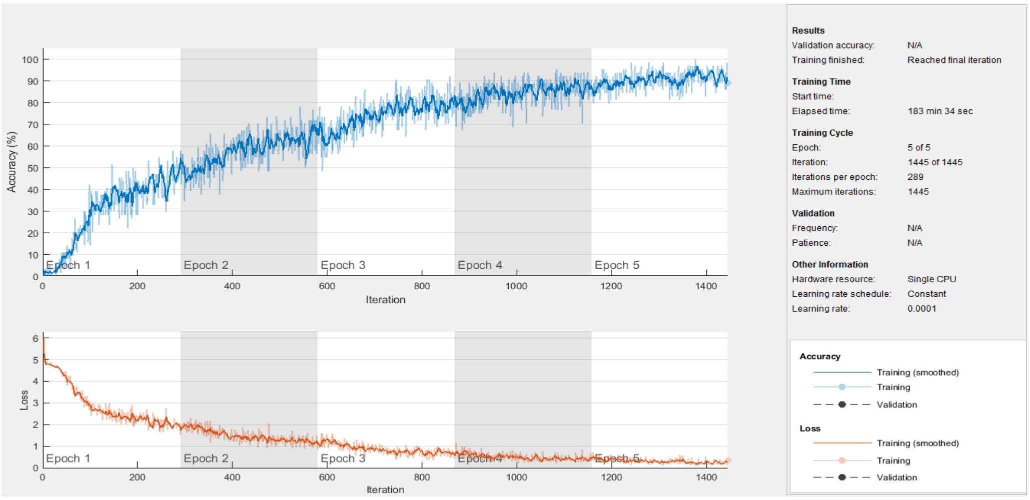

Figure 3. Training progress for AlexNet

Epoch	Iteration	Time Elapsed (hh:mm:ss)	Mini-batch Accuracy	Mini-batch Loss	Base Learning Rate
1	1	0:00:14	1.56%	5.987	1.00E-04
1	50	0:09:18	7.81%	3.9918	1.00E-04
1	100	0:15:38	14.06%	3.0596	1.00E-04
1	150	0:21:45	29.69%	2.6781	1.00E-04
1	200	0:27:21	28.13%	2.4075	1.00E-04
1	250	0:32:53	51.56%	1.8038	1.00E-04
2	300	0:38:19	51.56%	1.792	1.00E-04
2	350	0:44:05	45.31%	1.6631	1.00E-04
2	400	0:53:08	60.94%	1.5354	1.00E-04
2	450	0:59:19	54.69%	1.3747	1.00E-04
2	500	1:04:50	67.19%	1.3117	1.00E-04
2	550	1:10:35	67.19%	1.1921	1.00E-04
3	600	1:15:57	57.81%	1.4489	1.00E-04
3	650	1:21:44	73.44%	1.0162	1.00E-04
3	700	1:27:44	78.13%	0.716	1.00E-04
3	750	1:35:37	81.25%	0.6839	1.00E-04
3	800	1:41:52	78.13%	0.6911	1.00E-04
3	850	1:47:46	70.31%	0.7062	1.00E-04
4	900	1:53:21	89.06%	0.3209	1.00E-04
4	950	1:58:48	89.06%	0.3225	1.00E-04
4	1000	2:04:10	81.25%	0.6261	1.00E-04
4	1050	2:09:35	90.63%	0.3485	1.00E-04
4	1100	2:14:54	82.81%	0.4559	1.00E-04
4	1150	2:20:16	85.94%	0.4069	1.00E-04
5	1200	2:25:33	89.06%	0.2861	1.00E-04
5	1250	2:30:49	92.19%	0.2584	1.00E-04
5	1300	2:47:45	93.75%	0.2921	1.00E-04
5	1350	2:53:11	82.81%	0.597	1.00E-04
5	1400	2:58:38	90.63%	0.3338	1.00E-04
5	1445	3:03:34	89.06%	0.3497	1.00E-04

Table 2. Training result for AlexNet

3. EXPERIMENT RESULT

As shown in table 2, the Stanford Dogs breed dataset, which contains 120 classes of dog breeds, with 100+ images in each class sums upto 20580 images. The small dataset can lead to over fitting, especially with this pre training network used here. The dataset is randomly split into a training set approximately 70%, and testing set of 30%. The training started with max 5 epochs over each network. The learning rate is set to 1e-4. Validation is also conducted following every 50 iteration using validation sets. As shown in figure 3, training progress plot is plotted for each model. After training completed, testing is done on the test data and results obtained.

4. CONCLUSION

The training process data while fine tune training the AlexNet model is recorded in tabular form and graph are shown in Table 2 and figure 3 respectively. From the recorded data of each network, comparative analysis is done and shown in table 3. AlexNet gives testing accuracy of 84.35%, GoogleNet gives testing accuracy of 81.53%, DenseNet201 gives testing accuracy of 87.15 % and ResNet50gives testing accuracy of 90.12%. These results are obtained by modifying last 3 layers of each network. The research can be extended by taking more models in consideration and can go for more number of layers modifications.

Model Name	Test Accuracy
AlexNet + FT	84.35%
GoogleNet + FT	81.53%
DenseNet201 + FT	87.15%
ResNet50 + FT	90.12%

Table 3. Comparison of fine-tuned models

5 REFERENCES

1. Krizhesvsky, i. Sutskever and G. E Hinton (2012). ImageNet Classification with Deep Convolutional Neural Networks, Adv. NeuralInf. Process. Syst., 19-24.
2. K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman (2014). Return of the Devil in the Details: Delving Deep into Convolutional Nets, arXivPrepr. arXiv, 111-115.

3. K. He, X. Zhang, S. Ren, and J. Sun (2015). Deep Residual Learning for Image Recognition, Arxiv.Org, vol. 7, no. 3, 171-180.
4. Szegedy et al. (2015). Going deeper with convolutions, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 07, 19-25.
5. Khosla, Aditya, et al. (2011). Novel dataset for fine-grained image categorization:Stanford dogs, Proc. CVPR Workshop on Fine-Grained Visual Categorization (FGVC).Vol. 2.
6. Higa, Xavier S. (2019). Dog Breed Classification Using Convolutional Neural Networks: Interpreted Through a LockeanPerspective.