Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 December 2018 doi:10.20944/preprints201812.0232.v1

Modified Deep Neural Networks for Dog Breeds Identification

Aydin Ayanzadeh
Department of Applied Informatics,
Informatics Institute, Istanbul Technical
University, Istanbul, Turkey
ayanzadeh17@itu.edu.tr

Sahand Vahidnia
Istanbul Technical University
Istanbul, Turkey
vahidnia17@itu.edu.tr

***Abstract*—In this paper, we leverage state of the art models on Imagenet data-sets. We use the pre-trained model and learned weighs to extract the feature from the Dog breeds identification data-set. Afterwards, we applied fine-tuning and data-augmentation to increase the performance of our test accuracy in classification of dog breeds datasets. The performance of the proposed approaches are compared with the state of the art models of Image-Net datasets such as ResNet-50, DenseNet-121, DenseNet-169 and GoogleNet. we achieved 89.66% , 85.37% 84.01% and 82.08% test accuracy respectively which shows the superior performance of proposed method to the previous works on Stanford dog breeds datasets.**

***Index Terms*—Computer vision, Data Augmentation, Fine-Tuning, Imagenet**

I. Introduction

Conventionally, computer vision problems have been solved using hand-engineered features like HOG, SURF and various type of features. However, features extraction task has been mostly shifted toward Convolutional Neural Networks(CNN) since the success of AlexNet[3] at 2012 and VGG[4] at ImageNet ILSVRC-2014 . Afterwards, most of the competitors in ImageNet started to use CNN based models and state of the art has been improved every following year by a deeper model such as GoogleNet, ResNet[6] and DenseNet [5]. For instance,ResNet reduces the top-1 error to 3.5 percent at Imagenet ILSVRC-2015 from the results of GoogleNet and VGG that were 6.7 and 7.3 percent respectively at ImageNet ILSVRC-2014. In following approaches, filters are learned by training a convolutional neural network on the data-set. Discussing about details of Convolutional neural networks are out-of scope in this report, thus, it should suffice to state that CNN is a type of artificial neural networks that takes an image as input and dot product the randomly initialized filters (3×3, 5×5, etc.) across all image and outputs a class probabilities of object or objects in the image using softmax function. Back-propagation is used to learn the optimal parameters of filters that minimize the loss function. As mentioned earlier, GoogleNet was introduced at 2014 and has 22 layers. GoogleNet utilizes inception modules with parallel filter operations and concatenate outputs depth-wise. Inputs of the inception modules are 28x28xN. The problem with GoogleNet is its complexity and high number of parameters. To solve this, GoogleNet implements "bottleneck" layers to project input feature maps to lower dimension before convolutional operations. In comparison to previous techniques, GoogleNet is more complex. Full architecture of DenseNet has shown in Fig1. ResNet is another high-performing, state of the art CNN architecture with 34, 50, 101 and 152 layers. ResNet has significantly more layers in contrast to GoogleNet and its predecessors like VGG and Alexnet. Stacking up this number of plain layers and going deep will not have benefits. Also, it is important to note that deeper models are harder to optimize. Therefore,in order to overcome this issue, residual blocks have been implemented. Moreover, similar to GoogleNet, ResNet also deploys bottlenecks, except ResNet-34 to decrease complexity. Additionally, ResNet does not have Fully Connected layers at the end, but only a single FC1000. Fig1 shows architecture of ResNet-34 architecture.
mi

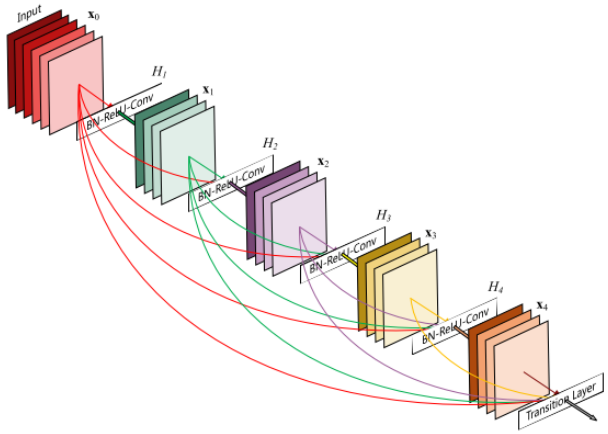

Fig. 1: Architecture of 5-layer dense block with of DenseNet [9].

DenseNet is one the most recent and deep CNN architectures, with 121, 161, 169 and 201 layers. DenseNet utilizes Dense-blocks in these architectures. DenseNet architectures for ImageNet has four DenseBlocks and three transition layers. According to the literature, DenseNet potentially can achieve much lower validation error with respect to number of pa-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 December 2018 doi:10.20944/preprints201812.0232.v1

rameters, in contrast to ResNet; but the final result in our paper remains to be seen in next sections. Beside of ImageNet challenging datasets,in the field of dog breed classification, there are some studies that have done for specific parts of ImageNet or free access dataset in the internet datasets;

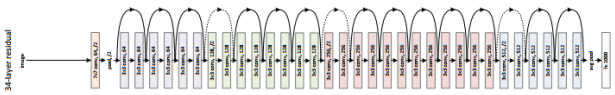

Fig. 2: Architecture of ResNet with 34 layers [6].

A. Dataset Description

The dataset used in this study is Stanford Dogs dataset [7], which contains 120 dog breeds and 20580 images for training and testing. In this paper, we split the dataset into training and test set, the proportion this split are 70 and 30 percents for training and test set respectively. This datasets is the small part of ImageNet challenging datasets.

Fig. 3: Species of Stanford dog breeds datasets

II. Proposed Method

According to the previous sections, we [8]realize that learned filters on such as large dataset are transferable into other image classification domains. Thus, in this work, we use the DenseNet-121,ResNet50,DenseNet169 and GoogleNet [1] model pre-trained on ImageNet ILSVRC [2], feed-forward images into the model and extract the features from the last layers (last two,four and six layers) which called fine-tuning.In fine-tuning we fix the pre-train weight of the all layers expect of the last layers, number of the tuned layer depend on the size of image dataset, If we have a very small image datasets, we just update the last layer and fix the weight of remaining layers. Hence,In the next section, we discuss about results of our experiments.

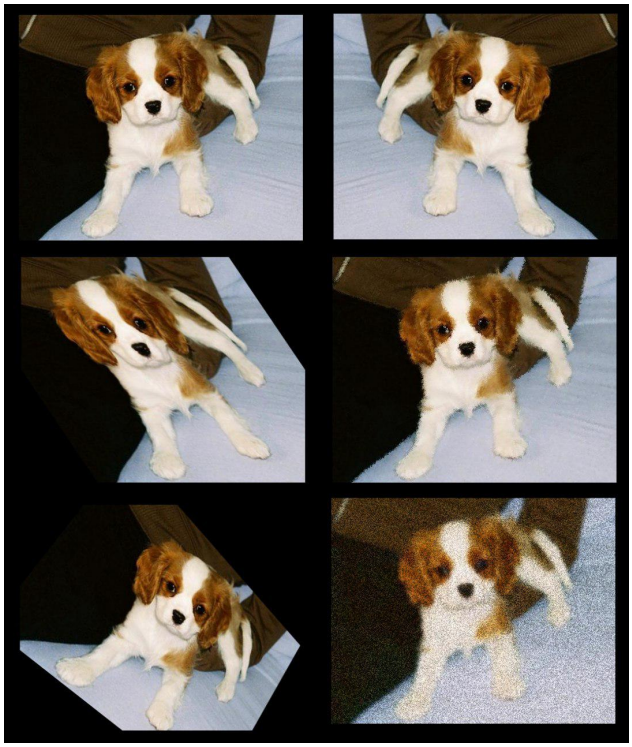

Fig. 4: Data-Augmentation of datasets that contain flipping, projection and noisy images.

A. Fine-Tuning

when datasets just concentrated on specific domain and have little number of images for each classes of datasets, the routine models of ImageNet can not classifying efficiently due to their high complexity. Therefore, they will be going to be over-fitted in the steps of test set in the specified datasets. To solve this, Fine-Tuning(FT) is good technique for the datasets which has small number of image which can not support the complexity of the state of the art models of imagenet datasets.

As previous work of this technique, H.Chen [13], utilized fine-tuned pre-trained CNN for localization of standard planes in ultrasound images. M.Gao [14] fine-tuned the whole of the layers of a pre-trained CNN for auto-classification of interstitial lung diseases. H.C.Shin [15] used fine-tuned pre-trained CNN to auto-mapping of the medical images to document-level of topics, document-level of sub-topics and sentence-level of topics.

B. Data Augmentation

As a matter of fact, Deep Learning models are very data hungry. Thus, providing enough training data can be a challenging task and sometimes impossible. Lack of data may result in over-fitting and sub-par scores. Hence, there are methods to mitigate this drawback of deep learning models models. A popular method for overcoming this issue is utilizing data augmentation. Basically, in this method it is required to produce new data, using prior data. In computer vision, this is done by transforming original data with preserving the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 December 2018 doi:10.20944/preprints201812.0232.v1

labels.
The transformations may vary according to different models. For instance, in [11] a random scale augmentation and cropping are performed. Also a random per-pixel mean subtraction is also conducted in this model. Other traditional augmentation techniques may include rotation, swirl, vertical and horizontal flipping and applying noise. Additionally, augmentation can be both balanced and unbalanced.
In the[12], the effect of augmentation has been studied in deep neural networks. The study claims to have achieved 100 percent accuracy on 30 times of augmentation, while getting very low accuracy on original data-set. This is important to note that the initial data in these experiments were relatively small. This study has concluded that balanced augmented method has an edge over unbalanced augmentation, and it is possible to achieve better results with smaller augmentation size in in balanced augmentation. Although the test model is not a very deep network, we can assume that the result may also apply to deeper networks. Thus, it is foreseeable to have major improvements on accuracy of our tested methods.

III. Experiment Results

In this section, we extract features of each image in train and test set using pre-trained [1] networks of the state of the art models that we have discussed about their architectures in previous sections.The feature vectors is tuned with splitting train set to the validation set 70 and 30 percent respectively. Afterwards, for predicting the probability of each class our base estimator is calibrated to output probability distribution for classes. Finally, we used Fine-tuning technique to update the last layers of the models and fixing the weight of the prior layers (The method was tested for last two, four and six layers). Experiment different number of components to generate test set results. We have evaluate the performance of different state of the art architecture in which the best result belongs to the DenseNet-169 with the fine-tuning of the last 3-layers and with the 30 epoch. we can reaching 89.66 % test accuracy of classification in ResNet50 model with Fine-Tuning(FA) of the last three layers and mixing this by data-augmentation to avoid from over-fitting of models. Moreover, we have compared the accuracy of testset in each models on 30 epochs with and without fine tuning and data augmentation in the tabel1 .

IV. Conclusion and Future Works

We have demonstrated modified approach of the state of the art networks such as ResNet, DenseNet and GoogleNet in Stanford dog breeds datasets. Due to the small number of the training datasets, we implemented the Fine-tuning and data augmentation to increase the accuracy of experiments in the test set. Actually, as it has been discussed in [8] last layers of CNN mainly learn filters to detect edges and simple shapes in the images which are mutual between different objects, however, last layers, are mostly like a template related to objects in the base dataset. Thus, performance of our proposed approach has increased relatively with respect to pure models of the state of the arts. As further work, due to increasing the size of datasets using some optimization methods to find the best hyper-parameters for learning rate to reduce the computational time of testset besides the classification accuracy in the test set. cyclical learning rate[10] can be efficient competitor to the adaptive learning rate. This technique can significantly reduce the time complexity and provide great opportunity to have higher level of augmenting in dataset to increase the accuracy of classification on the Stanford dog breeds datasets.

[1] https://keras.io/applications

TABLE I: Comparison of state of the art models with our modified models, DA and FT is the abbreviation of the Data Augmentation and Fine-Tuning respectively.

Model Name	Test Accuracy
DenseNet-121	74.28%
DenseNet-169	76.23%
GoogleNet	72.11%
ResNet-50	73.28%
DenseNet-121+FT+DA	84.01%
DenseNet-169+FT+DA	85.37%
ResNet-50+FT+DA	**89.66%**
GoogleNet+FT+DA	82.08%

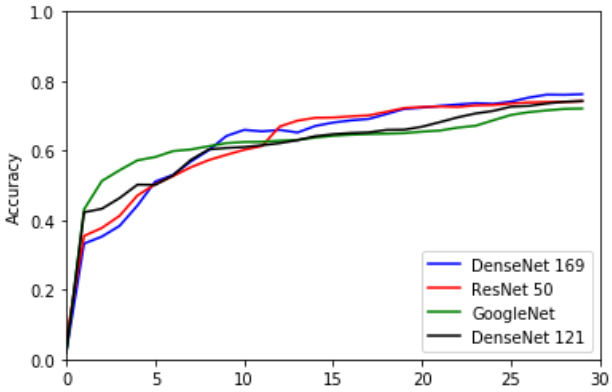

Fig. 5: Comparison of state of the arts models on the Stanford dog breeds datasets.

References

[1] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, Rethinking the Inception Architecture for Computer Vision, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 28182826, 2016.
[2] O. Russakovsky et al., ImageNet Large Scale Visual Recognition Challenge.
[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, Adv. Neural Inf. Process. Syst., pp. 19, 2012.
[4] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, Return of the Devil in the Details: Delving Deep into Convolutional Nets, arXiv Prepr. arXiv , pp. 111, 2014.
[5] C. Szegedy et al., Going deeper with convolutions, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 0712June, pp. 19, 2015.
[6] K. He, X. Zhang, S. Ren, and J. Sun, Deep Residual Learning for Image Recognition, Arxiv.Org, vol. 7, no. 3, pp. 171180, 2015.

 doi:10.20944/preprints201812.0232.v1

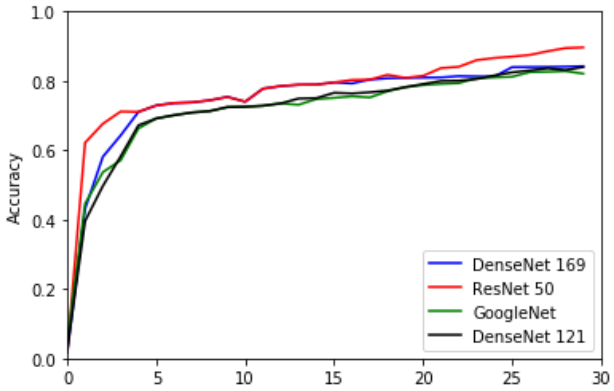

Fig. 6: Comparison of our modified approaches which is reinforced with data augmentation and Fine-tuning techniques.

[7] Khosla, Aditya, et al. "Novel dataset for fine-grained image categorization: Stanford dogs." Proc. CVPR Workshop on Fine-Grained Visual Categorization (FGVC). Vol. 2. 2011.

[8] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, How transferable are features in deep neural networks?, pp. 19.

[9] Huang, Gao, et al. "Densely connected convolutional networks." arXiv preprint arXiv:1608.06993 (2016).

[10] Smith, Leslie N. "Cyclical learning rates for training neural networks." Applications of Computer Vision (WACV), 2017 IEEE Winter Conference on. IEEE, 2017.

[11] J. Salamon and J. P. Bello, Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification, no. November, pp. 15, 2016.

[12] N. M. R. Aquino, M. Gutoski, L. T. Hattori, and H. S. Lopes, The Effect of Data Augmentation on the Performance of Convolutional Neural Networks, in Brazilian Society of Computational Intelligence, 2017.

[13] H. Chen, D. Ni, J. Qin, S. Li, X. Yang, T. Wang, and P. A. Heng. Standard plane localization in fetal ultrasound via domain transferred deep neural networks. Biomedical and Health Informatics, IEEE Journal of , 19(5):16271636, Sept 2015.

[14] M. Gao, U. Bagci, L. Lu, A. Wu, M. Buty, H.-C. Shin, H. Roth, G. Z. Papadakis, A. Depeursinge, R. M. Summer- s, et al. Holistic classification of ct attenuation patterns for interstitial lung diseases via deep convolutional neural net- works. the 1st Workshop on Deep Learning in Medical Image Analysis, International Conference on Medical Image Computing and Computer Assisted Intervention, at MICCAI- DLMIA15 , 2015

[15] H.C. Shin, L. Lu, L. Kim, A. Seff, J. Yao, and R. M. Sum- mers. Interleaved text/image deep mining on a very large- scale radiology database. In Proceedings of the IEEE Con- ference on Computer Vision and Pattern Recognition , pages 10901099, 2015

[16] Liu, Jiongxin, et al. "Dog breed classification using part localization." European Conference on Computer Vision. Springer, Berlin, Heidelberg, 2012.