ZKG INTERNATIONAL ISSN: 2366-1313

Deep Learning Models-Based Dog Breed Recognition System

[1] Patnala Yuva Naga DurgaVasantha Kishore, [2] Dr. B. Gohin,

[1] MCA Student, Dept. Of MCA, Swarnandhra College of Engineering and Technology, Seetharampuram, Narsapur, Andhra Pradesh 534280,

kishorepatnala2001@gmail.com

[2,] Associate Professor, Dept. Of MCA, Swarnandhra College of Engineering and Technology, Seetharampuram, Narsapur, Andhra Pradesh 534280,

Abstract: *The cutting-edge paper presents a pleasant-grained photo reputation trouble, one in all multi-class categories, namely determining the breed of a canine in a given photo. The offered system employs revolutionary techniques in deep getting to know, which includes convolution neural networks. Two one-of-a-kind networks are skilled and evaluated on the Stanford Dog's dataset. The usage/assessment of convolution neural networks is supplied via a software gadget. It includes a primary server and a cellular customer, which incorporates components and libraries for comparing on a neural network in each on-line and offline environments.*

Keywords— convolution neural networks, dog breed identification, fine-grained image recognition, image classification, inception resent v2, mobile trained model, Stanford dog dataset

I. INTRODUCTION

Nowadays, convolution neural networks (CNN) [1] are famous in exceptional subjects of deep studying: picture recognition [1], detection [2], speech popularity [3], data generation [4], and so on.

Several conventional image popularity techniques are acknowledged: Scale-Invariant Feature Transform (SIFT) [5], Histogram of Oriented Gradients (HoG) [6], attribute classification with classifiers: Support Vector Machine (SVM), Surtax and Cross Entropy loss. However, CNNs have also won significant traction in this subject in latest years, more often than not because of widespread reoccurring architectures being viable for fixing many different problems.

The present day paper gives the technique and outcomes of excellent-tuning CNNs for 2 specific architectures, using the Stanford Dogs dataset [7]. This constitutes a class trouble, however also one of fine-grained photo reputation, where there are few and minute variations setting apart

ISSN: 2366-1313

training.

Convolution neural networks are very just like Artificial Neural Networks [8], which have learnable weights and biases. The distinction is the filters, which method over the whole image and is effective in image recognition and type issues. Deep CNNs are possible on large dataset [9] and are even correct in massive-scale video classification [10]. Fine-tuning techniques and learning outcomes for the Inception-Resnet V2 [11] and NAS Net-A cell [12] architectures are supplied. Furthermore, using the trained convolution neural networks is visualized via a separate software program system using modern technologies. This machine is able to decide the breed of a canine in a picture provided by the user, and also presentations detailed statistics approximately each diagnosed breed. It consists of two predominant components: a cell client and a centralized net server.

The rest of the file is structured as follows: Section II provides an outline of similar tactics within the literature, while Section III affords the used and pre-processed Stanford Dogs dataset. Section IV info the gaining knowledge of two unique CNNs, with Section V encapsulating the results thereof. Providing a practical side to these CNNs, the accompanying software program device is described in Section VI. Conclusions are drawn and in addition development plans are proposed in Section VII.

II RELATED WORKS

The contemporary section offers preceding tries at addressing the issues tackled by way of the modern-day studies. Abdel-Hamid et al. [3] clear up a comparable hassle, one among speech reputation, the use of traditional strategies, making use of the dimensions and function of each local element, and PCA, at the same time as Sermanet et al. [13] and Simon et al. [14] mention convolution neural networks with special architectures.

Liu et al. [15] present alternative studying methods in 2016 the use of interest localization, whilst Howard et al. In 2017 [16] gift a mastering of a CNN using the Mobile Net architecture and the Stanford Dogs dataset prolonged with noisy statistics.

Similar pleasant-grained photo reputation troubles are solved by detection. For example, Zhang et al. [17] generalize R-CNN to detect exceptional parts of a picture, whilst Duan et al. [18] discover localized attributes. Angelova et al. [19] use segmentation and item detection to tackle the same problem. Chen et al. [20] use selective pooling vector for fine-grained photo popularity.

The present day studies is based on fine-

ISSN: 2366-1313

tuning CNNs and the consequences thereof are reproducible at the original Stanford Dogs dataset using the provided methods.

III DATASET AND PREPROCESSING

The supplied CNN getting to know technique revolves across the Stanford Dogs [7] dataset. It incorporates a hundred and twenty one-of-a-kind canine breeds and is a subset of the extra fashionable Image Net [21]. It is separated into education and test dataset. Both units comprise pix of various sizes and each picture is given a label representing the embodied canine breed. The schooling dataset incorporates 12.000 pictures with kind of a hundred consistent with breed; the take a look at information consists of eight.580 unevenly dispensed photos.

The first step of the pre-processing is to split the schooling information into teach folds and validation fold for experimental tuning of the studying hyper parameters. Before splitting the data, the canine photographs are resized to 256x256 pixels (for NAS Net-A cellular) and 299x299 pixels (Inception-Resnet V2 input).

For experimenting with the hyper parameters of the CNNs, fine-tuning and five-fold cross-validation is used, which ultimately produces 5 specific schooling and validation subsets. Each of those datasets consists of 9.600 training images and a pair of. Four hundred validation records. After getting the quality exceptional-tuned hyper parameters, the entire Stanford Dogs training dataset is used for training, at the same time as the test data is exclusively used for evaluation.

IV EXPERIMENTS

After obtaining the essential formatted and resized information, the following step is great-tuning the convolution neural networks. This segment affords the applied technology, CNN architectures, methods, hyper parameters and the use of the trained fashions as frozen graph.

The supplied hassle suits into the category of first-rate-grained photograph recognition, because the variations linking any sample photograph to a sure magnificence are few and minuscule; the CNN ought to don't forget small key features to dissolve ambiguity. For examinational, the husky and the malamute breeds gift small enough differences amongst them to make differentiating hard even for trained eyes.

A. CNN Architectures

Transfer getting to know [22] offers a overall performance improve by no longer requiring a complete education from scratch. Instead, it makes use of pre-

ISSN: 2366-1313

educated models which can be taught standard reoccurring functions. These fashions are often skilled at the Image Net [21] dataset, which has a competition every year and some pre-skilled models are posted. The getting to know of those fashions represents pleasant-tuning the given dataset with the learned weights and biases. The current research incorporates two extraordinary public pre- educated convolution neural networks, which are fine-tuned: NAS Net-A cell [12] and Inception-Resnet V2 [11].

The NAS Net-A structure is created based at the approach of the Neural Architecture Search (NAS) body- paintings [23], by means of the Google Auto ML [24].

The Inception-Resnet V2 is a totally deep structure containing over three hundred layers, which is created by using the Google developer's team.

B. **Data augmentation**

The most not unusual method to lessen over fitting on education information is to use specific changes before the feed forward bypass at some stage in the education; this is referred to as records augmentation [9]. During the learning of the CNN models, another pre-processing approach, augmentation, is carried out. The Inception-Resnet V2 and the NAS Net-A mobile structure makes use of Inception Pre-processing, this is the following feature:

$$f(x) = \left(\frac{x}{255.0} - 0.5\right) * 2.0$$

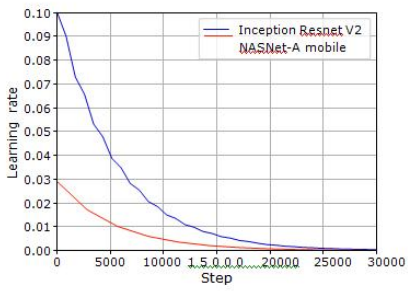


Fig. 1. The adaptive learning rate during the training sessions decay exponentially by 10% every 3 epochs. The blue is the learning rate of the Inception-Resent V2 with an initial value of 0.1, while the red is the NAS Net-A learning rate with an initial value of 0.029.

Wherein x the picture. Before the Inception pre-processing, the picture is randomly reflected and cropped by way of the Tensor Flow's distorted bound box algorithm.

For the assessment of the validation or the test dataset, the applied augmentation steps encompass an 87.Five% valuable crop and the Inception pre-processing.

C. **Learning and hyper parameters**

The nice-tuning experiments of the convolution neural internet- works are carried out on a personal laptop with a Ge Force GTX 1080 GPU, an Intel Core i5-6400 CPU and sixty four GB RAM. During the studying, a Soft max Cross-Entropy loss feature and Nester momentum [25] optimizers are used for the fine- tuning of the NAS Net-A cellular

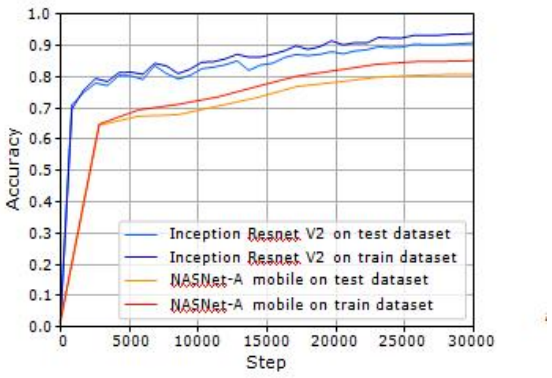
ISSN: 2366-1313

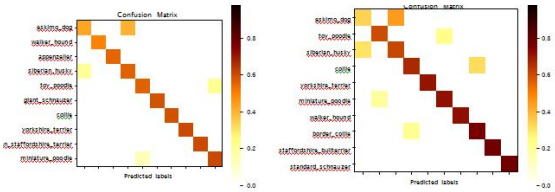
model and the Inception- Resnet V2 model. During the training, the closing absolutely-connected layer (log its) is unfrozen and pleasant-tuned, whilst the alternative layers are unchanged and frozen. This uses the advantages of the pre-trained fashions.

The first components of the CNN training contain hyper parameter tuning using go-validation. During the experiment the hyper parameters are chosen empirically. After ensuing in the correct parameters, every other section of gaining knowledge of begins on the entire Stanford Dogs training dataset, with the model evaluated on the check dataset in this example.

Both convolution neural networks are educated with the subsequent commonplace hyper parameters: a batch length of sixty four, an ex- potentially reducing studying rate with exclusive preliminary value, in which the charge decays 10% each 3 epochs (about 563 steps), a 0.0001 weight decay, and training for 30.000 steps. The NAS Net-A cell structure is pleasant-tuned with a gaining knowledge of fee with an initial value of 0.029 (see Figure 1). The Inception-Resnet V2 is high-quality-tuned with a studying charge with an initial fee of 0.1 (see Figure 1). Training the mobile version takes three times much less than the Inception Resnet V2 model.

Further experimented hyper parameters include: fixed analyzing fees (0.01, zero.001), exponentially adaptive learning quotes (with initial values of 0.031, zero.1/2), default weight decay (0.0004), different numbers of steps for training (15.000, 20.000, and 20.500) and distinct optimizer (RMS prop).





(a) NASNet-A mobile bottom 10 classes (b) Inception Resnet V2 bottom 10 classes

V RESULTS

The trained fashions are evaluated each 10 mins at the schooling and test dataset. In this phase, we present the applicable metrics for assessment: accuracy (see Figure IV-C), precision, consider (see Table I) and confusion matrices (see Figure three). The assessment factors are linearly dispensed in time, now not in variety of steps, as a result the uneven step distribution.

ISSN: 2366-1313

The accuracy is monitored at the schooling and takes a look at datasets; this metric represents the mean percent of successfully classified classes on a dataset. The NAS Net-A cell architecture achieves eighty five.06% accuracy at the train dataset and 8.72% at the take a look at dataset. The accuracy with a deeper CNN indicates better outcomes: the Inception Resnet V2 community achieves 93.Sixty six% accuracy at the teach dataset and ninety. Sixty nine% at the take a look at dataset (see Figure IV-C). The advanced performance of the latter is not sudden, due to the fact that it is a much deeper community.

Precision and don't forget also are measured for the duration of the assessment of the education and test dataset; outcomes are presented in Table I. The precision and they do not forget for the skilled Inception Resnet V2 model is higher, the deeper version extracts more features from a photograph and classifies higher than the educated NAS Net-A cell model.

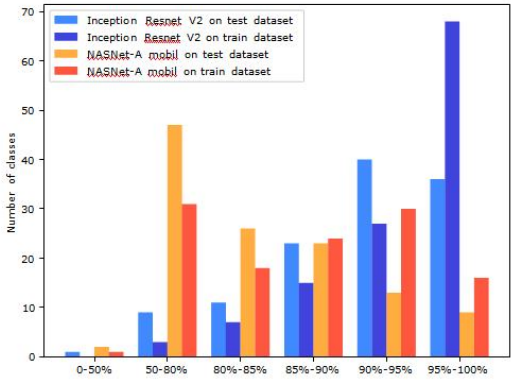


Fig. 4. Histogram of classified class percentages for both CNN architectures on both datasets. A bar represents the count of classes within a range of accuracies.

TABLE II
NASNET-A MOBILE ACCURACY WITH DIFFERENT OPTIMIZERS

Optimizer	*Training accuracy*	*Test accuracy*
Nester momentum	85.06%	80.72%
RMSprop	84.94%	80.57%

TABLE III
PERFORMANCE OF SPECIES CATEGORIZATION USING STANFORD DOGS

Method	*Accuracy*
Chen et al. [20]	52.00%
Simon et al. [14]	68.61%
Google LeNet [13]	75.00%
Krause et al. [26]	82.6%
Liu et al. [15] (ResNet-50)	88.9%
Ours, NASNet-A mobile	80.72%
Ours, Inception Resnet V2	90.69%

Model, most classes from the train dataset are classified in the (95%, 100%] interval, while from the test dataset the most classes are classified in the (90%, 95%] range. The test dataset is classified well with some inaccuracies remaining. The classified classes from train and test dataset for the NAS Net-A model have a wider spread, with a majority of the classes falling in the (50%, 80%] range. The large accuracy difference between the two models is understandably in correlation with the size and complexity of the architectures.

ISSN: 2366-1313

The mentioned accuracies using the Nas Net-A architecture are achieved using the Nester momentum optimizer. For a comparison, an alternative optimizer, RM Sprop, is also tested;the results are similar (see Table II).

Comparison to related work on Stanford Dogs dataset is given in the Table III.

After the evaluation of each trained convolution neural network, Grad-CAM [27] heat map visualization is made for the NAS Net-A mobile trained model (see Figure 5). The last convolution block"pays attention "mostly to the heads of each dog on an image.

Examining the heat map images, the NAS Net-A model mostly focuses on the head of the dogs. In the second image, which shows a German shepherd in a different position, the CNN pays attention also to the body. If an image contains more than one dog, the network is interested in all of the recognized dogs in varied percentages, and evaluates the image taking each appeared breed into consideration. For an accurate classification it must consider the evaluate an image, which contains different parts of the dog including the head. The displayed images are evaluated correctly by the NAS Net-A mobile trained model.

Using different data augmentation: random image rotation between 0 and 15 degrees, random zooming or 87.5% central cropping does not help to improve the accuracy of the trained models.

The alternative hyper parameters presented in Section IV-C (fixed vs. adaptive learning rates, weight decay and step counts) prove to not improve the accuracy or other valuable metrics after training the CNNs.

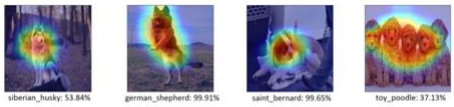


Fig. 5. Grad-CAM visualization with heat maps for the last layers of the NASNet-A mobile trained model. The last convolution block "pays attention" to the warm colored parts of the images (mostly to the heads of each dogs), while the cold colors represents the less interested parts of the image. These pictures are not part of Stanford Dogs dataset.

VI THE SOFTWARE SYSTEM

The usage of the trained convolution neural networks is provided via a software program machine, referred to as Sniff!. Its related cell application offers the possibility for users to take a image or choose an current one from the gallery in a cellular utility, which not handiest classifies the photo, but additionally displays exact records about every evaluated breed. The displayed statistics serves educational and informative functions.

The software program system consists of two issues: a central server written in the Go programming language, and a cellular consumer found out in React Native. The additives communicate via HTTP requests/responses.

ISSN: 2366-1313

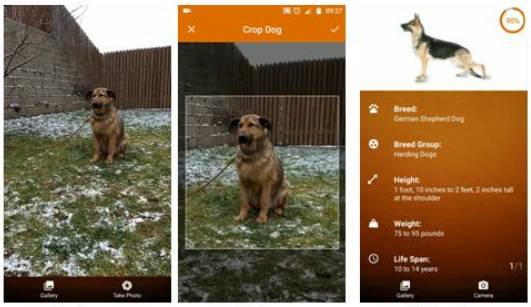
The server includes a classifying module the use of the Ten- workflow Go library; it masses the trained convolution neural network, pre-processing and evaluating photographs. The Inception Resnet V2 and NAS Net-A version are both unable at the server, because a desktop machine can make use of more CPU/GPU resources for a quicker evaluation. By default, the Resnet model is used, since it reaches higher accuracies.

The outcomes of the assessment for a picture are for each instruction; the maximum instructions are evaluated with zero% percentage. To avoid the wrong class with low percentage there is set a threshold on the server and additionally on the cell consumer. The cellular patron can take an image with the digital camera of the cell phone or import one from its gallery, and publish it for class. The system can show up on-line the usage of the valuable server for a faster category, or offline the use of the telephone sources in case of a lacking network connection. Offline assessment is facilitated via React Native Tensor flow [28] wrapper library. Figure VI shows the usage of the Sniff!

Utility.

The app uses the NAS Net-A mobile educated version, which hundreds each evaluation into the reminiscence of the tool; this depends on the resources of the smart phone. The reminiscence is freed after the evaluation.

The app shows designated facts approximately the detected breeds, with records net scraped from A-Z Animals2 and dog- time.Com3.



(a) Camera (b) Crop (c) Result

Fig. 6. Main components in the Sniff! application.

VII Conclusion

Two different convolution neural network architectures had been supplied: the NAS Net-A cell structure and the Inception Resnet V2 deep structure.

The architectures have been tested on a niche photo class problem: that of spotting canine breeds. The pre-educated networks are excellent-tuned the use of the Stanford Dogs dataset.

Results are promising even for the smaller, mobile-pleasant CNN, achieving simplest 10% much less accuracy than the deep Inception Resnet V2 model.

We have also presented a usage of the nice-tuned convolution neural networks

via a software device, called Sniff!: a cellular application, that could determine the breed of a dog from an image (even without an Internet connection).

The convolution neural networks may be similarly developed by using: Generative Adversarial Nets (GAN) [4] to increase the education dataset, using other loss function like centre loss [29], schooling other convolution neural network architectures, increasing the dataset with different famous dog breeds, the usage of detectors for locating a couple of puppies on an image and optimizing the server and mobile class.

REFERENCES

1. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Image Net Classification with Deep Convolution Neural Networks," in *Advances in Neural Information Processing Systems 25*. Curran Associates, Inc., 2012,pp. 1097–1105.

2. S. Ren, K. He, R. B. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *CoRR*, vol. abs/1506.01497, 2015.

3. O. Abdel-Hamid, A. r. Mohamed, H. Jiang, L. Deng, G. Penn, and D. Yu, "Convolution neural networks for speech recognition," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22,no. 10, pp. 1533–1545, Oct 2014.

4. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley,

S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672– 2680.

5. T. Lindeberg, "Scale Invariant Feature Transform," *Scholarpedia*, vol. 7,no. 5, p. 10491, 2012, revision #153939.

6. N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1. IEEE, 2005, pp. 886–893.

7. A. Khosla, N. Jayadevaprakash, B. Yao, and L. Fei-Fei, "Novel dataset for fine-grained image categorization," in *First Workshop on Fine- Grained Visual Categorization, IEEE Conference on Computer Vision and Pattern Recognition*, Colorado Springs, CO, June 2011.

8. R. J. Schalkoff, *Artificial neural networks*. McGraw-Hill New York, 1997, vol. 1.

9. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Image net classification with deep convolution neural networks."

10. A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and

L. Fei-Fei, "Large-scale video classification with convolution neural

networks," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2014, pp. 1725–1732.

11. C. Szegedy, S. Ioffe, and V. Vanhoucke, "Inception-v4, inception- resnet and the impact of residual connections on learning," *CoRR*, vol. abs/1602.07261, 2016.

12. B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Teaching trans- ferable architectures for scalable image recognition," *CoRR*, vol. abs/1707.07012, 2017.

13. P. Sermanet, A. Frome, and E. Real, "Attention for fine-grained categorization," *CoRR*, vol. abs/1412.7054, 2014.

14. Prasadu Peddi (2019), "Data Pull out and facts unearthing in biological Databases", International Journal of Techno-Engineering, Vol. 11, issue 1, pp: 25-32.

15. X. Liu, T. Xia, J. Wang, and Y. Lin, "Fully convolution attention localization networks: Efficient attention localization for fine-grained recognition," *CoRR*, vol. abs/1603.06765, 2016.

16. Prasadu Peddi (2015) "A machine learning method intended to predict a student's academic achievement", ISSN: 2366-1313, Vol 1, issue 2, pp:23-37.