Dog Breed Classification

Sim Kae Wanq

Abstract—There are around 340 breeds of dog recognized by the Federation Cynologique Internationale (FCI), which is a big amount of data and information that cannot be carry by one for classification. However, With the potential of Deep Learning, this difficult task sound more possible. Stanford Dogs Dataset by Stanford University contains images of 120 breeds of dogs from around the world. This dataset was trained using DIGITS, but the accuracy wasn't as good as expected in classifying dogs.

Index Terms—Robot, IEEEtran, Udacity, LaTeX, deep learning.

1 Introduction

There are around 340 breeds of dog recognized by the Federation Cynologique Internationale (FCI). However, the standards for breed recognition vary from country to country. Most people can't classify one dog to another easily due to limited knowledge in mind. With potential of Deep Learning, classification of hundreds of dog breeds sound more reliable and possible. As country like China, using technology like video analytic, image processing, Artificial Intelligence (AI), etc. to improve and enhance the surveillance system, and increase advantage in domain like financial, health care, etc. Making robotic inference growing stronger. Ability to classifying dog breeds or any flora and fauna with power of AI, will benefit people from different industry and domain.

Fig. 1. Can you differentiate i.Alaskan Malamute and Siberian Husky ii.Cairn Terrier and Norwich Terrier iii.Lhasa Apso and Shih Tzu iv.Boston Terrier and French Bulldog

1.1 Robotic inference vs human inference

When it comes to identifying faces, humans are better at subjective judgments like whether someone looks happy, sad, angry or just a bit shifty. Thus, Most humans have trouble in identifying faces. However, computers beat human on the quantitative side of things such as measuring the relative dimensions of the features on a face. [1] In classifying dog breeds, the difference between dogs could be much ambiguous, e.g. the difference between Alaskan Malamute and Siberian Husky is the pattern of tail, the difference between Cairn Terrier and Norwich Terrier is the length of leg and color. These tiny features detail decide what a dog kind is, but they are too difficult to be recognized by human perceptions. By providing sufficient data and sample to train a deep learning machine, a model could be designed and trained for solve classifying dog breed problem, or even study the historical, biological relationship between dogs.

2 Background / Formulation

In this experiment, dataset Stanford Dogs Dataset provided by Stanford University was used. [2] [3] [4] NVIDIA Deep Learning GPU Training System (DIGITS) was used to conducting the experiment. As the dataset provide 256 X 256 color images, Alexnet convolutional neural network was used for training and generating the classifying model. Alexnet has similar performance compare to GoogleNet but better runtime. [5]

3 Data Acquisition

Due to the limitation of capturing large amount of data for each breed of dog, dataset - Stanford Dogs Dataset, used in this experiment, was provided by Stanford University. Stanford Dogs Dataset contains 120 breeds of dogs from around the world, not more than 150 images per class, and total images of 20580. Dataset images were resized into standard 256 x 256 x 3 RGB images, and 25 percent of dataset was kept for validation process.

To train and test the model, following parameter was used:

- Training epochs : 5 / 10 / 20
- Snapshot interval (in epochs) : 1
- Validation interval (in epochs) : 1
- Random seed : [none]
- Batch size : [network defaults]
- Base Learning Rate : 0.01

Noted that different training epochs - 5/10/20 was used to create 3 different model, when other parameter was remained unchanged. For each created model, same batch of random images were used to verified the accuracy.

Fig. 2. Partial dataset used in this experiment

4 Results

The result shows that model of 20 training epochs has the highest accuracy, compared to 5 and 10 epochs model. Table below shows the accuracy and validation loss of each model:

TABLE 1
Accuracy and validation loss of each model

No. of Epochs	Accuracy	Valid Loss
5 epochs	$\approx 5\%$	$\approx 83\%$
10 epochs	$\approx 9\%$	$\approx 77\%$
20 epochs	$\approx 17\%$	$\approx 68\%$

In figure 3-5, the accuracy line and validation loss line show logarithmic growth, meaning the accuracy will not increase over certain point and validation loss will not drop over certian point even if the training epochs increased.

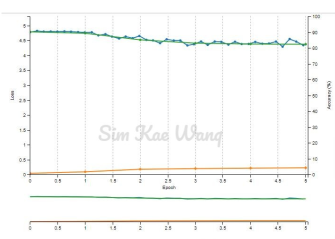

Fig. 3. 5 epochs model

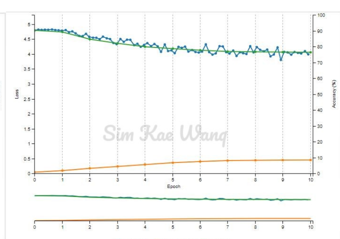

Fig. 4. 10 epochs model

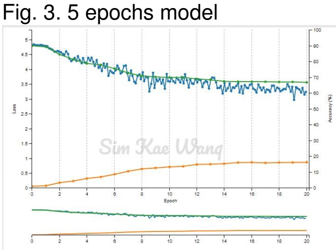

Fig. 5. 20 epochs model

Once models were created, samples images was used to determine the accuracy of inference. Figure 6-8 below show the inference result of each model. Model with 20 training epochs give the correct answer, Shih Tzu, with 11.39% of confidence level.

Fig. 6. 5 epochs model

Fig. 7. 10 epochs model

Fig. 8. 20 epochs model

5 Discussion

Even though model with 20 training epochs get inference correctly the most compared to model with 5/10 training epochs. However, the confidence level was not ideal that suitable to be deployed in actual application. Several factor can be the reason causing result above. In this experiment, dataset images were resized before input into network. The process of resize made some images distorted and lost important feature and information for classification. Besides, some images in the dataset contain irrelevant object or complicated background, which might become the noise input into the network. Classifying dog breeds is far more difficult and complicated than human facial recognition, because entire body feature and information are need to decide which dog breed it is, instead of just part of body, like face.

6 Conclusion / Future work

In this paper, Stanford dog dataset was used as input of Alexnet to create a classification model that able to recognize different dog breeds. However, the accuracy of that trained model was immature and bad for application. Dataset for training the model was not appropriate, caused noise and error in this experiment.For future research and experiment, dataset could be pre-processed and annotated as per the network input requirement, before fed into network. Collecting images that provide more feature and information of dog breed could improve the accuracy.

References

[1] K. W. B. R. W. V. B. P. J. G. G. W. Q. M. P. P. Jonathon Phillips, Patrick J. Flynn, “Distinguishing identical twins by face recognition,” tech. rep., 2011.
[2] B. Y. L. F.-F. Aditya Khosla, Nityananda Jayadevaprakash, “Stanford dogs dataset.”
[3] 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.
[4] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A Large-Scale Hierarchical Image Database,” in *CVPR09*, 2009.
[5] Z. I.-N. S. Nur Azida Muhammad, Amelina Ab Nasir, “Evaluation of cnn, alexnet and googlenet for fruit recognition,” tech. rep., 2018.