

Convolutional Neural Network based Inception v3 Model for Animal Classification

Jyotsna Bankar¹, Nitin R Gavai²

Shri Venutai Chavan Polytechnic Pune¹

Sinhgad College of Engineering Pune²

Abstract: Identification of similar types of objects in image processing now become regular task, but when dissimilar objects come into picture it becomes quite complex. Even it become more complicated when not just we need to identify objects but also categorizes them into their specific classes. We are using machine learning technique to classify the animal and put them into specified classes. Inception-v3 is the open source algorithm made by Google and used for objects classification. In this paper, based on Inception-v3 model in TensorFlow platform, we use the transfer learning technology to retrain the animal category datasets, which can greatly improve the accuracy of animal classification.

Keywords: Classification, Inception-v3, image processing, TensorFlow.

I. INTRODUCTION

Zoology is the study of animals. Currently there are over 66 thousand (less than 5% of all animals) vertebrate species, and over 1.3 million (over 95% of all animals) invertebrate species in existence. species identification and its classification is a fundamental research work in the zoology [18]. When human eyes see the mammals it can easily identify what exactly it is but when it comes to embedded vision devices it become complex task. Therefore, having a fast and accurate animal classifier will bring a lot of eagerness in peoples' lives. There are some challenges in mammal classification like the similarity between the different species of animals, the complex background of animal image. We cannot just rely on a single feature, such as colour, texture or shape to distinguish them. The same animal will be different because of the shades of colours, shape, scale, view point etc.

In 2012 ImageNet Large Scale Visual Recognition Competition (ILSVRC) [12] Convolutional Neural Networks (CNN) have become more popular than AlexNet in computer vision after winning the competition [13] [11]. To achieve higher accuracy in image classification development and usage of deeper and complex CNN become trend in research [14] [17] [15] [9]. However, the complex structure of models does not to not necessarily making networks more efficient with respect to speed and size to increase the accuracy. On computationally limited platform object detection and recognition has to be done with time critical manner, for example online games, self-driving car, robotics and automation, augmented reality [8].

Convolutional neural network is an efficient recognition method which has been developed in recent years. This network avoids the complex pre-processing of the image, and people can input the original image directly. It uses local receptive field, weights sharing and pooling technology and makes the training parameters greatly reduced compared to the neural network [2] [6]. It also has a certain degree of translation, rotation and distortion invariance of image. It has made great progress in the field of image classification.

TensorFlow [1] is the second generation of artificialintelligence learning system developed by Google, which supports the convolutional neural network (CNN), recurrentneural network (RNN) and other depth of the neural networkmodel, which can be used in speech recognition, imagerecognition and so on many machines deep learning field. In this paper we have shown a mammals' classification using machine learning pertained model Inception v-3. In mammals we are using classes as Lion, Tiger, Elephant, Giraffe, Monkey. Section II describes existing work with TensorFlow in constructing small models. Section III describes main theme, the Inception-v3 architecture. Section IV gives building of classification model; Section V refer to experimentation on five different mammals' datasets. Section VI concludes with summary.

II. EXISTING WORK

Compared with the traditional image classification methods, convolutional neural network use multilayer convolution to extract features and combine the features automatically. It has a higher recognition rate and a wider range of applications.

TensorFlow [1] as the second generation of Google artificial intelligence learning system has got much attention and affirmation in the field of the machine learning in all over the world. TensorFlow has ranked first in all machine learning and deep learning programs so far. It has the advantages of high availability and high flexibility, and with the support of TensorFlow researchers, the efficiency of TensorFlow is improved. Today, Google has released a number of pretrained models on the TensorFlow's official website, to facilitate the use of researchers in different fields.

Inception-v3 [3] is one of the pretrained models on the TensorFlow. It is a rethinking for the initial structure of computer vision after Inception-v1 [5], Inception-v2 [4] in 2015. The Inception-v3 [3] model is trained on the ImageNet datasets, containing the information that can identify 1000 classes [12] [14]. In ImageNet, the error rate of top-5 is 3.5%, the error rate of top-1 dropped to 17.3%. Tensorflow [1] also provides detailed tutorials for us to retrain Inception's final Layer for new categories using transfer learning.

Transfer learning is a new machine learning method which can use the existing knowledge learned from one environment and solve the other new problem which is different but has some relation with the old problem. For example, we can apply the knowledge learned from the motorcycle problem to the study of bike problem. Compared with the traditional neural network, it only needs to use a small amount of data to train the model, and achieve high accuracy with a short training time.

III. CONSTRUCTION OF ANIMAL CLASSIFICATION MODEL

This section focuses on experimental setup for mammals' classification model using Inception-v3 on Tensor Flow framework. Here classification model is separated into following four stages: image preprocessing, training, verification followed by testing.

A. Image Preprocessing

Image pre-processing is the name for operations on images at the lowest level of abstraction whose aim is an improvement of the image data that suppress undesired distortions or enhances some image features important for further processing. It does not increase image information content. Its methods use the considerable redundancy in images. Neighbouring pixels corresponding to one object in real images have the same or similar brightness value and if a distorted pixel can be picked out from the image, it can be restored as an average value of neighbouring pixels.

B. Training Process

While training the model we use approximately 2000 image dataset, around 400 images per mammal, every image is used multiple times through training process. Computing the layers behind the layer just before the final output layer which performs the grouping for each image takes a substantial time. As the lower layers of the network are not being changed their outputs can be stored and used again.

C. Verification and testing process

By *testing*, we mean evaluating the system in several conditions and observing its behavior, as stated above we are not just providing single image as input to the inception model instead multiple images multiple times watching for defects. By *verification*, we mean producing a compelling argument that the system will not misbehave under a very broad range of circumstances so the accuracy of model will not be varied.

IV. EXPERIMENT

To complete this work, we used python 3.5, Tensorflow 1.2 and the hardware platform is Dell N series laptop: processor 2.5GHz Intel i5, memory 12GB.

In the proposed system, the mammals' image dataset is collected from google and class wise batches are created. This dataset is used as input to the machine learning model in which retraining will be provided based upon pretrained inception v3 algorithm [16]. TensorBoard is a monitoring and inspection tool included with tensorflow. We use it to monitor the training progress. Once the training is done sample input image is used for testing and validation on which the accurate recognition will be done [18].

A. Procedure of Inception-v3

In Inception-v3, transfer learning method is used. We should keep the parameters of the previous layer, then remove the last layer and input the mammal dataset to retrain the new last layer, the number of output nodes will be changed to 5 because we used only 5 different mammals. The last layer of the model is trained by back propagation algorithm, and the cross entropy cost function is used to adjust the weight parameter by calculating the error between the output of the softmax layer and the label vector of the given sample category [3] [16].

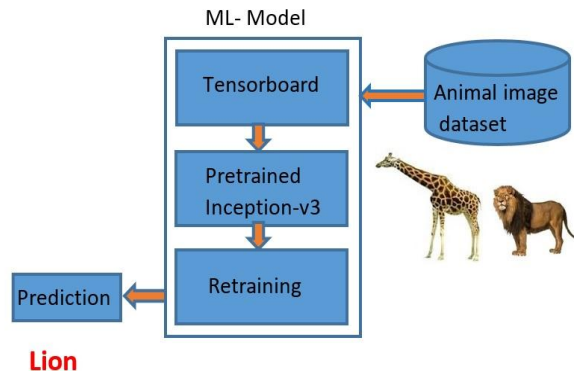


Figure1: System Architecture

The namesake of Inception v3 is the Inception modules it uses, which are basically mini models inside the bigger model. The same Inception architecture was used in the GoogLeNet model [6] which was a state of the art image recognition net in 2014. The idea is that you don't need to know ahead of time if it was better to do, for example, a 3x3 then a 5x5. Instead, just do all the convolutions and let the model pick what's best. Additionally, this architecture allows the model to recover both local feature via smaller convolutions and high abstracted features with larger convolutions.

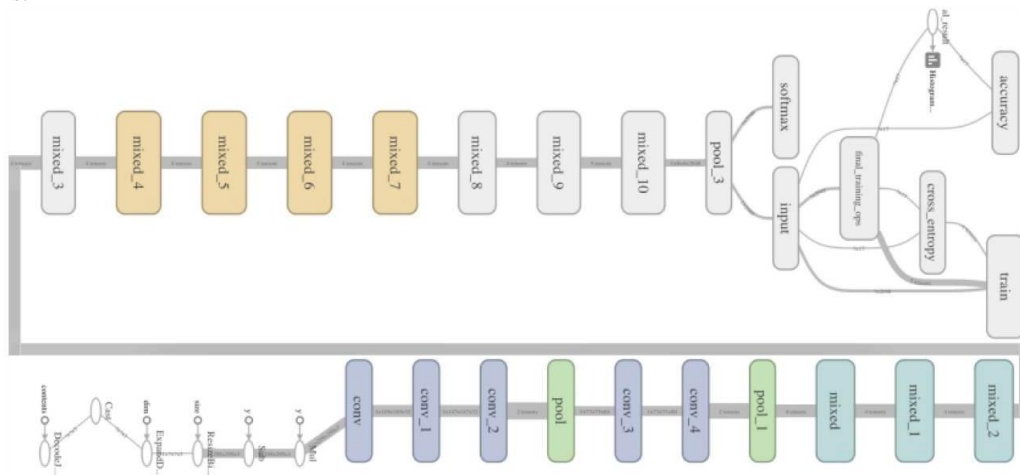


Figure2: Main graph of Inception-v3 model



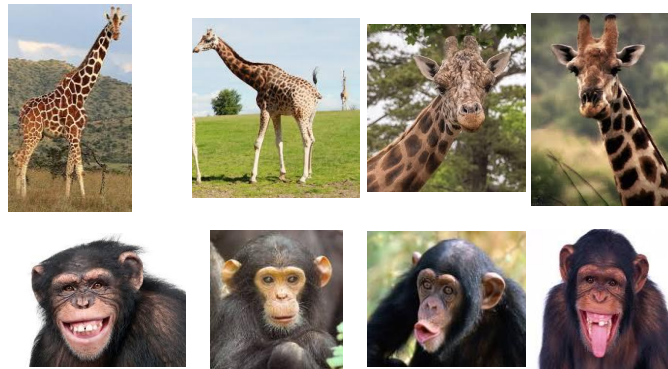


Figure 3: Sample image dataset

B. Results

Figure 4 and figure 5 show the variation of accuracy and cross entropy based on mammals' dataset. The orange line represents the training set, and the green line represents the validation set.

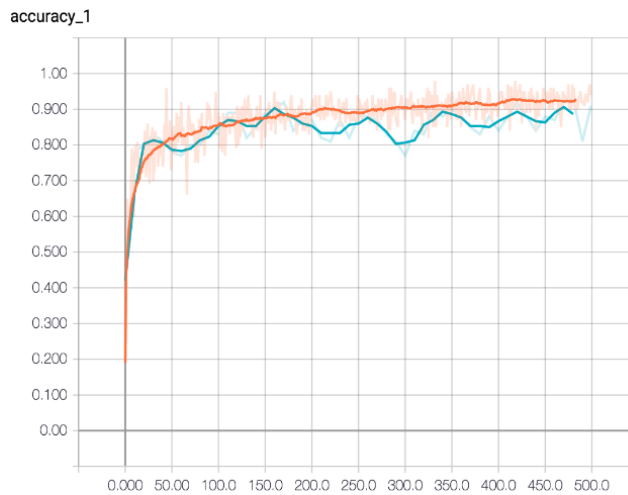


Figure 4. The variation of accuracy on mammals' dataset

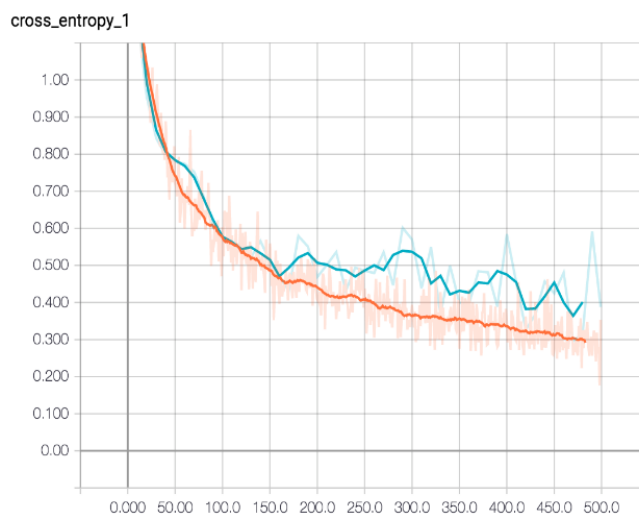


Figure 5. The variation of cross entropy on mammals' dataset

The training accuracy shows the percentage of the images used in the current training batch that were labeled with the correct class. Validation accuracy: The validation accuracy is the precision (percentage of correctly-labelled images) on a randomly-selected group of images from a different set. Cross entropy is a loss function that gives a glimpse into how well the learning process is progressing (lower numbers are better here)

Table 1: Description of the four figures

Index	Accuracy
accuracy of training set	99%-100%
accuracy of validation set	95%
cross entropy of training set	0.10
cross entropy of validation set	0.24

After giving the random image input of tiger following is the output given by the Inception v3 model. Evaluation time for same input image is 2.160 sec and accuracy of model is 97.92 %.

```
Evaluation time (1-image): 2.160s
tiger 0.9792021
lion 0.020207532
giraffe 0.000312877
monkey 0.00022765477
elephant 4.9853978e-05
```

Few Comparisons

MobileNets[7] are utilise depth-wise separable convolutions, very similar to inception towers in inception v3 [3]. These also reduce the number of a parameters and hence latency [7]. MobileNets also have useful model-shrinking parameters than we can call before training to make it exact size we want. The other models are very deep, large models. The reduced number of parameters or style of convolution is not used for low latency but just for the ability to train very deep models, essentially. ResNet introduced residual connections between layers which were originally believed to be key in training very deep models [20]. These aren't seen in the previously mentioned low latency models.

V. CONCLUSION

We proposed the system based on the Inception-v3 model of TensorFlow platform, in which we use the transfer learning technology to train an animal classification model on mammals' dataset. The classification accuracy of the model is approximately 95% on given dataset, which is higher than other method available for classification. The future work is to study and develop a more effective and accurate model for image classification.

REFERENCES

- [1] Peter Goldsborough Fakultät für Informatik, "A Tour of TensorFlow arXiv:1610.01178v1 [cs.LG] 1 Oct 2016.
- [2] Abhinav Saxena, "Convolutional neural networks: an illustration in TensorFlow". ACM Crossroads 22(4): 56-58, 20 16.
- [3] Xiaoling Xia, Cui Xu, Bing Nan, "Inception-v3 for Flower Classification," 2nd International Conference on Image, Vision and Computing, 2017.
- [4] J. Krause, B. Sapp, A. Howard, H. Zhou, A. Toshev, T. Duerig, J. Philbin, and L. Fei-Fei. "The unreasonable effectiveness of noisy data for fine-grained recognition". arXiv preprint arXiv:1511.06789, 2015.
- [5] Sergey Ioffe, Christian Szegedy, et al. "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift". ICML 2015: 448-456, 20 15.
- [6] J. Christian Szegedy, Wei Liu, et al, "Going Deeper with Convolutions". arXiv:1409.4842, 2014.
- [7] Nitin R. Gavai; Yashashree A. Jakhade; Seema A. Tribhuvan; Rashmi Bhattad "MobileNets for flower classification using TensorFlow" International Conference on Big Data, IoT and Data Science (BIG) 154 – 158, 2017
- [8] Angelova A, Zhu S, Efficient Object Detection and Segmentation for Fine-Grained Recognition[C]. Computer Vision and Pattern Recognition: 8 1 1-8 18, IEEE, 2013.
- [9] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. arXiv preprint arXiv:1512.03385, 2015.
- [10] Maria-Elena Nilsback, Andrew Zisserman: Automated Flower Classification over a Large Number of Classes. ICVGIP 2008: 722- 729, 2008.
- [11] F. N. Iandola, M. W. Moskewicz, K. Ashraf, S. Han, W. J. Dally, and K. Keutzer. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and; 1mb model size. arXiv preprint arXiv:1602.07360, 2016
- [12] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pages 1097–1105, 2012
- [13] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al. Imagenet large scale visual recognition challenge. International Journal of Computer Vision, 115(3):211–252, 2015
- [14] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.
- [15] C. Szegedy, S. Ioffe, and V. Vanhoucke. Inception-v4, inception-resnet and the impact of residual connections on learning. arXiv preprint arXiv:1602.07261, 2016
- [16] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1–9, 2015.
- [17] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. "Rethinking the inception architecture for computer vision". arXiv preprint arXiv:1512.00567, 2015.
- [18] K. He, X. Zhang, S. Ren, and J. Sun. "Deep residual learning for image recognition". arXiv preprint arXiv:1512.03385, 2015.
- [19] Merriam-Webster's, "Animals". Retrieved 16 May 2010. 2 a: one of the lower animals as distinguished from human beings b: mammal; broadly: vertebrate"
- [20] Kaiming He Xiangyu Zhang Shaoqing Ren Jian Sun. "Deep Residual Learning for Image Recognition" Microsoft Research, Dec 2015.