A

Synopsis Report

on

Building partially understandable convolutional neural networks by differentiating class-related neural nodes.

submitted in partial fulfillment of the requirements
for completion of AI LAB
of

TY COMP

in

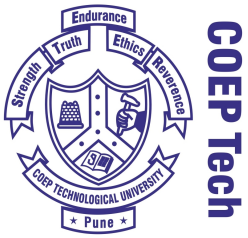
Computer Engineering

by

Om Khare - 112003066
Harshmohan Kulkarni- 112003075
Mihika Sanghvi - 112003084

Under the guidance of

Suraj Sawant
Professor
Department of Computer Engineering



Department of Computer Engineering,
COEP Technological University (COEP Tech)
(A Unitary Public University of Govt. of Maharashtra)
Shivajinagar, Pune-411005,Maharashtra, INDIA

November, 2022

1 Introduction

In recent years, CNN (convolutional neural network) has been used for multiple applications and has created a great breakthrough in the field of image processing. The most common use for CNNs is image classification, for example identifying satellite images that contain roads or classifying hand written letters and digits. There are other quite mainstream tasks such as image segmentation and signal processing, for which CNNs is used.

CNN is a category of deep learning neutral networks where the model itself also means knowledge. It is a kind of multi-layer neural network with artificial neurons. The multi-layer consists of the input layer, hidden layer and output layer. The kernel which exists in the first hidden layer can gain certain features from input images and further it will be transmitted to the next layer. During the process, the feature obtained becomes more complex as the layer is deeper. With different image classes, CNN can extract some prominent features from the images fed to it and proceed with classification.

The Dogs vs. Cats image classification is a common classification problem tested with many baseline machine learning models. It is easy for humans to distinguish the two animals, but certain images create difficulties even for humans pertaining that it will be a significantly difficult task for a computer to do this. Hence training a computer to do so requires intricate layers for identifying the image. Many people have worked or are working on constructing machine learning classifiers to address this problem. A classifier based on color features got 56.9 percent accuracy on the Asirra dataset. An accuracy of 82.7 percent was achieved from a SVM classifier based on a combination of color and texture features. In our project we are going to build a convolutional neural network which can better represent the decision making of the model. Another addition we are including in this project is that instead of only using the Kaggle data set comprising of total 25000 images.

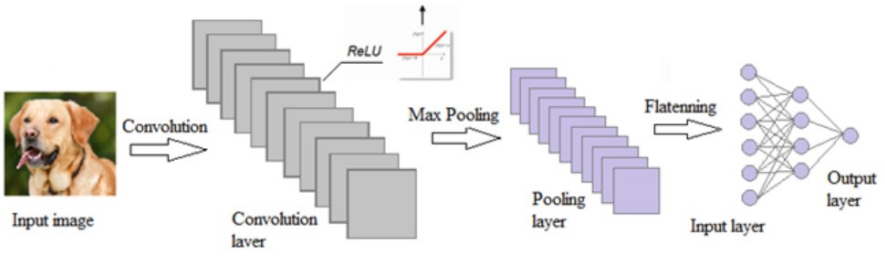


Figure 1: General structure of a CNN architecture

2 Motivation

While exploring Convolutional Neural Networks, the CNN model that we build that itself is the knowledge which is learnt while training. It is beneficial for the development of the architecture that we can express the knowledge learnt by it in a human understandable way rather than considering it as a black box. Three fundamental issues can be noted underlying this problem as given below:

Q1) What type of knowledge a neural network learns in a task?

Q2) The knowledge learned by a neural network is said to be unpredictable. We generally just passively accepted what the model has learned and not understand why the model has learned this. Hence the next question arises why the model learned that particular knowledge?

Q3) The decision made by a neural model was generally not evaluable. So When did the agent succeed and when did it fail? When can the results be trusted?

These 3 questions became our motivation. We needed to have an in depth understanding of the functioning and decision making of the model. Our purpose is to build a modified neural model to clarify information coding in some specific high conv-layers of CNNs, and to evaluate further the decisions made by a neural model, which can still maintain the original learning and generalization abilities.In fact, in many fields, concerns exist regarding the application of deep CNN models, not only because the model itself cannot provide sufficient information, but also in terms of security. For example, application systems constructed around Artificial Intelligence (AI) will often be involved in and affect many fields, such as autonomous transportation, face recognition systems, AI-assisted medical systems, and many

other fields. However, considering the challenges highlighted above, the usefulness and security of these AI systems will be limited by our ability to understand, explain and control them.

3 Literature Review

The project is regarding distinguishing the characteristics between a dog and a cat is done using convolutional neural networks. It examines the features and details of the image provided and accordingly computations are made at every layer in the CNN which helps give the desired output.

A large amount of our research was from the research paper [1] whose authors are Dawei Dai, Chengfu Tang, Guoyin Wang, Shuyin Xia. The paper highlighted the advantages of convolutional neural network in image processing and implementation of the same with the help of models. ResNet50 model has 50 layers which increases the accuracy of correct output as more layer implies more accuracy. We implemented the project/idea presented in the above paper. Along with this resource, we referred to other websites and research papers [2] [3] [4]

The model used is ResNet50 which is a Convolutional Neural Network model which includes layers like convolutional layer, pooling layer (max pool and average pool) and relu layer. Padding of image also takes place to reduce the lose of information after passing the picture into these layers. These techniques and various layers help identify the image as cat or dog. More about these techniques will be understood later on in this paper.

The data set we have used has been taken from Cats vs Dogs Classifier kaggle which has around 25000 images. This data set is quite large and sufficient to find out the accuracy of our model.

4 Research Gaps

Work has been done in identifying how exactly neural network models work. For example visualization methods have been developed to understand what concepts or features were learned by a neural model by transforming the layers into images that can be understood by humans. However this was restricted to deep neural networks.

Another method of proxy model was built using interpretable rules to simulate the input and output of the black-box model, and the new models were used to interpret the black-box models. However these models fail to answer the three questions raised. We still cannot understand the features learned by a neural model, we still cannot expect what neural models learned, and we still cannot interpret why models perform as they do.

5 Problem Statement and Objectives

We have focused on understanding and knowing the type of knowledge learned by a neural network and evaluate the decision making of the network.

Objectives:

1. To understand the type of knowledge learned by the neural network.
2. Predict the knowledge learned by the neural network.
3. Evaluate the decision made by the neural network.

6 Methodology

For neural information coding, we can assume that a neuron in the particular neural network learns or encodes information pertaining to only one class or multiple class. As this happens, neurons tend to participate in classification of multiple classes making it difficult for us to understand it.

Thus we aim at modifying the baseline model for some of the developed Neural Network Models to have a clear understanding of the information learned by

neural layers. We plan to preassigned some of the nodes in the model to encode information of a particular class which is to be classified. We will be able to regard the weights in the preassigned neurons as the Standard Template for a particular class.

When a particular sample will be evaluated, then the information coding generated by that particular sample can be matched with the Standard Template and an estimate of why the model performed in a particular way can be judged.

For the implementation of this model we plan to split the feature maps of a particular layer into two parts and each encode information related to a particular class. For the splitting of these features, we have two options with us. Either we can go with some objective functions which will decide the set of the feature maps used for a particular sample while training as well as tesing, or we'll connect a Dense Layer before the splitting and on the classification result of the layer will be used for the selecting the feature maps.

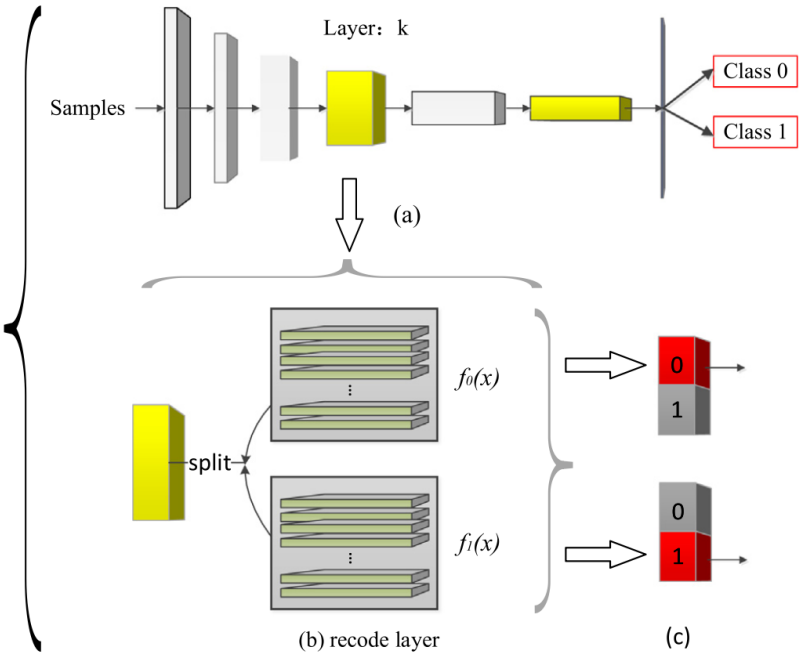


Figure 2: Splitting of the Feature Maps

We will be implementing this on the ResNet50 CNN Model. It is considered as it is a deep network and usually deeper networks have stronger expression ability.

ResNet50 eliminates the degradation problem to some extent by means of shortcut structure. As depth of the neural network structure increases, the accuracy over test and training data sets remains constant at a particular value and then starts decreasing. We can reduce this problem with shortcut structure which will add the skip some layers in the architecture and we will feed the input of previous layer to the next layer directly or by adding it with the output of the previous layer.

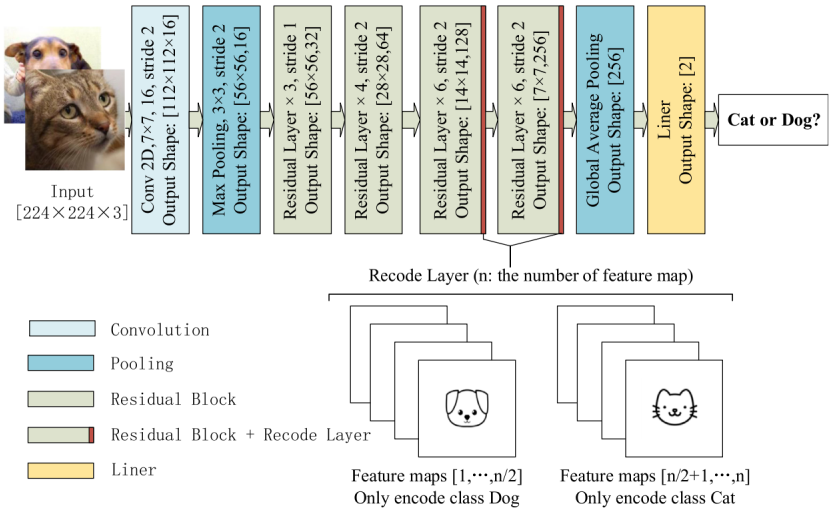


Figure 3: Splitting of the Feature Maps

The modified structure of the ResNet50 architecture is given in Fig. 3, the separate feature maps which we call as ”Recode Layers” will be implemented in the last convolutional layer of the third and the fouth residual blocks. These residual blocks majorly contain multiple sets of Convolutional Blocks and Identity Blocks. These contain the convulational layers implemented with shortcut structure.

Some initial image preprocessing will be done on the test and the train data which includes cropping the images to 244 X 244, normalising the images by dividing it by 255 per pixel. We have adopted the batch normalisation after every convulational layer and we have trained the model from scratch.

7 Hardware and Software Requirements

For the working of the modified architecture we will need a high end system with GPU cores present for enhanced speed during training of the huge data set consisting of 25000 images.

We majorly need the following things in the software side as a prerequisite for being able to train and test the model.

1. Python 3.8 or greater
2. Keras - An API interface for tensorflow liberaries
3. Tensorflow - Free library for training and testing Neural Networks
4. Scikit-learn - Free Library for Machine Learning
5. SciPy - Free Library for scientific computations.
6. Matplotlib - Plotting Library for Python.
7. NumPy - Library for calculations of multidimensional arrays and matrices.
8. Pandas - Library for data manipulations and analysis.
9. Test Data of Cats and Dogs.We are also working on collection of drone data.

Figure 4: Cats and Dogs Data Set

8 Conclusions

In recent years, neural networks have a developed to a great extent but there can be greater progress if they become better understood. Our model focuses on the representation and generalization of the feature maps learned by the model compared to the current baseline models. Our model provides interpretability to the decisions taken by CNN model which is very helpful to judge a particular decision of the model. The expected information coding can now be interpreted using the feature maps and comparing them with the ones of a particular sample helps in the judgement. But still this model struggles as the clear information coding is only possible at a category level and only at some particular layers. The representation however is not state-of-the-art result. We are exploring some methods to generalise our attempt.

9 Timeline Chart

Timeline	
Date	Work Done
15/8/22	Shortlisting research papers for project implementation
22/8/22	Selecting Project and going through research paper
1/9/22	Group discussion on research paper and planning regarding project implementation
15/9/22	Research on topics like CNN, ResNet50
17/9/22	Distribution of work and Writing synopsis
22/9/22	Data preparation
26/9/22	Layer implementation (ResNet50)
	Obtaining result

References

[1] D. Dai, C. Tang, G. Wang, and S. Xia, "Building partially understandable convolutional neural networks by differentiating class-related neural nodes," *Neurocomputing*, vol. 452, pp. 169–181, 2021.

[2] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions," *Journal of big Data*, vol. 8, no. 1, pp. 1–74, 2021.

[3] R. Poojary and A. Pai, "Comparative study of model optimization techniques in fine-tuned cnn models," in *2019 International Conference on Electrical and Computing Technologies and Applications (ICECTA)*. IEEE, 2019, pp. 1–4.

[4] N. Sharma, V. Jain, and A. Mishra, "An analysis of convolutional neural networks for image classification," *Procedia computer science*, vol. 132, pp. 377–384, 2018.