Comparative Study of CNN and LSTM for Fish image Classification

Nepomscene Nduwarugira, Bong-Kee Sin
Pukyong National University, Department of IT Convergence and Application Engineering
neponduwarugira8484@gmail.com,bkshin@pknu.ac.kr

Abstract

Deep neural network (DNN) has become popular in image classification in recent years due to high performance and easy access to powerful softwares. Convolutional neural network (CNN) and long short-term memory (LSTM) are two common types of DNN. In this paper, the two models are applied for studying and comparing their performances. The results show that they return varying results on fish image classification. CNN and LSTM achieved respectively 81% and 47% accuracy with a dropout regularization to avoid overfitting and multiple answer candidates instead of one.

Keywords: Deep neural network, LSTM, CNN, fish classification.

I. Introduction

Machine learning is an area in artificial intelligence where the relationship between the input and output data of a system is modeled by extracting attributes from the input data. Artificial neural networks are a powerful tool machine learning which is inspired by the biological brain, having layers of neurons. Deep learning is a recent approach of neural network where a neural network consists of many and often very many layers. The goal of this paper is to compare the performance of two popular types of neural networks on fish identification using the image in QUT dataset [4]. The CNN is a type of feed-forward neural network which is generally used for image classification [1] while the LSTM works on the principle of saving the output of a layer and feed it back to the input layer [2].

II. Proposed methods

2.1. Convolutional neural network

Convolutional neural networks (CNNs) generally based on three types of layers: convolutional layer, pooling layer (max pooling/mean pooling) and fully connected layer. See fig 2.

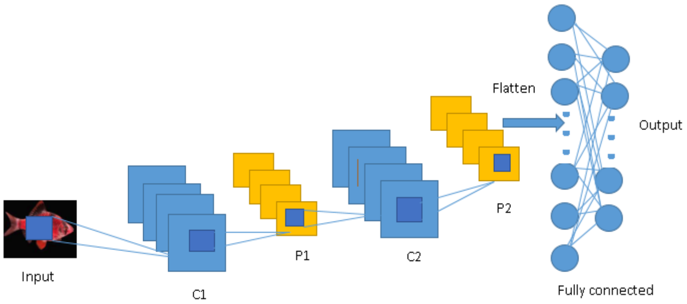
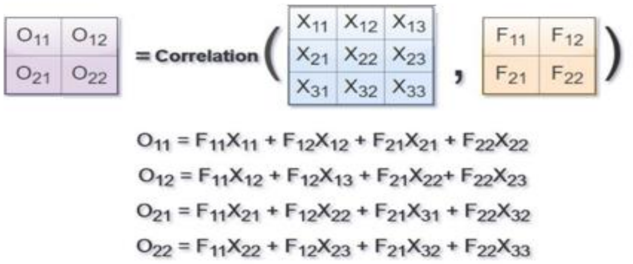


Figure 1: CNN structure Ci: convolutional layer
Pi: pooling layer

In this research, an input RGB image is resized to 112 by 112. It is followed by a convolutional layer to extract a number of features. The convolutional layers extract the local features in each neuron. A standard convolutional layer takes image as an input and applies randomly weighted filters, also called as kernels, which extract a different feature from the input.

Normally, convolutional layer takes input data and utilizes weighted filters for extracting different feature maps of input. The convolutional layer performs element-wise products and their sum, after a certain nonlinear transformation, goes to the output. The same procedure applied for every region.



$$O_{11} = F_{11}X_{11} + F_{12}X_{12} + F_{21}X_{21} + F_{22}X_{22}$$
$$O_{12} = F_{11}X_{12} + F_{12}X_{13} + F_{21}X_{22} + F_{22}X_{23}$$
$$O_{21} = F_{11}X_{21} + F_{12}X_{22} + F_{21}X_{31} + F_{22}X_{32}$$
$$O_{22} = F_{11}X_{22} + F_{12}X_{23} + F_{21}X_{32} + F_{22}X_{33}$$

Figure 2

A Kernel can be defined to be of any size as needed. In convolutional layers, the number of filters is not known a priori and can be chosen according to the task. We can increase the complexity of features extracted via additional layers. Selecting the right number of convolutional layer is not easy, it must be chosen carefully depending on the problem. Generally, each convolutional layer is followed by a pooling layer often max pooling or average pooling, which is important in reducing the size of the input representation to help decrease the amount of parameters. Most common pooling operation is carried out on 2x2 boxes with maximum operation.

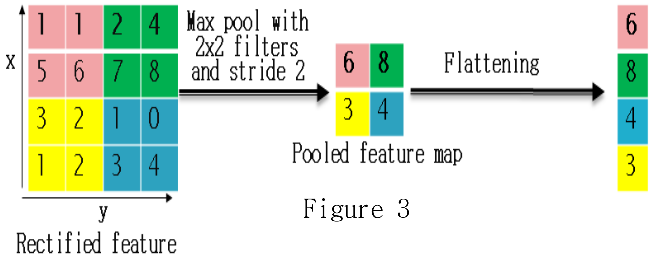
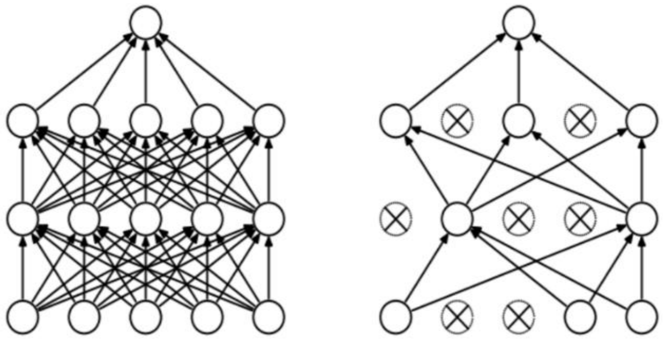


Figure 3

Usually,the last pooling layer is flattened into one-dimenssional array data before making the final classification.see fig.3.

2.2. Dropout

The output of the last pooling layer acts as an input to the so called fully connected layer. Usually there are many nodes and edges that may lead to overfitting. Fortunately, during training, regularization technique called ' dropout' [3] can intervene to overcome this issue by removing some nodes and edges which are not very active in classification stage. See fig.4.



Fully connected layer

After applying Dropout

Figure 4

2.3. Top k candidates

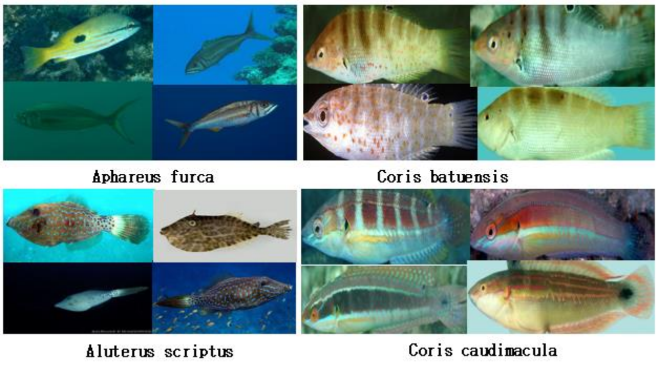
Due to difficulties in many classification problems, creating or returning top likely candidates is often a preferred method for future post-processing. It is applied during testing to return multiple answers instead of a single best.

2.4. Long Short-Term Memory (LSTM)

LSTM is a type of artificial recurrent neural network (RNN) that can be considered as an alternative to CNN in deep learning. More commonly, it is very useful for text classification.

III.QUT Fish Dataset

In this research we take QUT fish image dataset. This is used to compare two different deep neural networks structures. This dataset has been also used in [4] for object classification. It is a collection of 1140 fish images from 76 different categories.



Aphareus furca

Coris batuensis

Aluterus scriptus

Coris caudimacula

Figure 5: Some samples from QUT dataset

IV. Experiments

Deep neural networks with various numbers of layers and filters have been tested for comparison. Two to four layers have been applied in the design. However, due to 2 large number of connections, training is often difficult. Several methods such as dropout and top k candidates were adopted to overcome this issue. With three convolutional layers, we observed a performance of 53 % with single best without dropout but the figure rose to 59% with 0.5 dropout. Indeed, the performance increases gradually with increasing numbers of candidates. From single best to top 5, when it reaches 81%.For comparison, the cases of two layers and four layers didn’ t work as well, respectively 80% and 64%.In CNN the number of filters played an import role in the fish classification task. See Table 1, the third low. A comparison of CNN with three layers and the LSTM with 500 layers gives 62% and 47%.

Table 1: Convolutional layered structures.

Number of layers	Filters size and Filters numbers	Dropout rate	Top candidates (Top k)	Performance (%)
2	5x5x3x128 4x4x128x512 28*28*512	0.5	5	80%
3	5x5x3x64 4x4x64x128 3x3x128x256 14*14*256	0.5	1 5	53% 75%
3	5x5x3x64 4x4x64x256 3x3x256x512 14*14*512	0	1 5	53% 62%
3	5x5x3x64 4x4x64x256 3x3x256x512 14*14*512	0.5	1 5	59% 81%
4	5x5x3x64 4x4x64x256 3x3x256x512 3x3x256x512 7*7*512	0.5	5	65%

Table 2: CNN compared with LSTM.

Methods	Top candidates (top k)	Number of hidden layer	Performance (%)
CNN	5	3	62%
LSTM	5	112 400 500	35% 44% 47%

V. Conclusion

This paper aimed at a comparative study of CNN and LSTM for fish classification. Through a set of experiments, it is found that CNN is the preferred method recording 81% by applying 3 convolutional layers [64, 256, 512 filters], 0.5 dropout and top 5.The future work would be comparing this work with GANs, still another method in DNN.

References

[1] M.Sarigül and M. Avci, “Comparison of Different Deep Structures for Fish Classification,” *International Journal of Computer Theory and Engineering*, Vol.9, no.5, pp.362-366, October 2017.

[2] B., F. Ren, Y. Bao, “Investigating Lstm With k-Max Pooling For Text Classification, ” *IEEE 11th International Conference on Intelligent Computation Technology and Automation*,pp. 31-34, 2018.

[3] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from over fitting,” *Journal of Machine Learning Research,* pp. 1929-1958, June 2014.

[4] K. Anantharajah, Z. Y. Ge, C. McCool, S. Denman, C. Fookes, P. Corke, etal, “Local Inter-Session Variability Modelling for Object Classification,” *IEEE Winter Conference on Applications of Computer Vision,* pp.309-316, June 2014.