maximum pooling and latest techniques to distinguish between animal images using keras

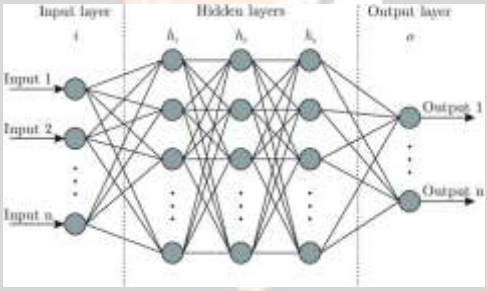
Sunit Fulari ,*Alumni NIT GOA*

Rajesh B Lohani, *Professor ETC Dept GEC*

abstract

It is very overwhelming as to how our brain can recognize all the different images and differentiate and process one image from the other. We can recognize one form of image or suppose an alphabet however twisted and rusted it is as what it means. Similarly in this paper we explore how neural networks work and how we can get datasets as to train the convolutional neural network to recognize one image from the other. More the data set the better will be the accuracy. In this paper we come to the conclusion and make a thorough study of images to say that neural networks is successful in recognizing one set of object from the other. It uses set of matrices in which the features are recognized by 1 and non features are recognized as zeros. We use python keras in getting the libraries such as pooling.

Keyword- *maximum pooling, keras*



This image is taken from google about a neural network just to show how an image from left is used to recognize at the right end. The middle are the hidden layers which get input from the input layers and are in turn optputted to the last which is the output layer. In the hidden layer it is all about which neuron is activated. The activation of the neurons is done by a sigmoid function which are controlled by weights.

Okay going back to the beginning of life when we first developed eyes to see. At around 543 million years ago there were animals floating around but could not see with sensory organs, at around 540 million years ago something extraordinary happened the living animals evolved and they developed sensory organs. It became very competitive and they had to survive very rapidly to survive and so they evolved more rapidly. Sensory organs work by the brains processing, lets see in mammals. There is a visual cortex in the back of the brains, there are simple cells and more complex cells, what has been discovered as that the images travel through these cells in geometric shapes and the ability to recognize and reconstruct what has been found out. It is how can we develop and make something that will recognize and represent objects. This was started in 1970s when computers were very very slow and it was practically impossible with that technology available at that time. Now we have already exceeded the brain of a mouse, ie our computer.

If object recognition is too hard scientists thought lets do image segmentation that we have studied in this convolution neural network. Sunit et al in the paper "A survey on motion models used for object detection in videos" used open cv and optical flow which is a very old technology to identify the feature of faces and moving objects.

Now we have object detection, apart from that in artificial intelligence in writing which is a very good work in which google used the information to read the written works. Which is easily available online and works by various

tech giants. Graphical models, support vector machines whatever is the machine learning models, whatever are they, they are very old and with outdated technology of yesteryears.

So, how do we work on image classification task. There is no clearly intuitive algorithm how it works. People have clearly and definitely made high end efforts to write complex algorithms. One thing is cats have ears,nose,they are highly deformable,eyes and etc and etc. we use test image training image to find out the final pixel.

Going back to the older algorithms like the nearest neighbor algorithm.

The islands and the fingers pushing out of the region are not very good in this algorithm. Now looking at the k-nearest neighbor algorithm which is slightly better. K nearest neighbor is also very slow at times and pixels on distance metrics are not very informative.

Deep learning and computer vision is the new technology advancing very rapidly. We have self driving cars, face id to open electronic devices. With advances and time we may have self driving card with absolute ultimate accuracy running without drivers. Stay tuned!.

In convolutional neural networks we convolve them. That is we take two separate data or two matrices or like two images. We take this each layer though its very confusing, same as feature map or filter and kernel is little different.

We have an input matrix, or matrices that are multiplied by matrices. Feature map,kernel or filter whatever you call it. Heres the great thing about a series of matrices, okay next we talk about pooling used in the coding. Max pooling is we convolve the four sets of data to find the maximum amongst them. We do these for all the sets adjacent to the other.

In our research in this paper we have used test sets and training sets of dogs,cats,horses,lions,and animals to differentiate between them.

The more the data sets the better is the end result.

In this result we find that out of the 25 epochs we find 85% accuracy.

```
8000/8000 [==============================] - 69s - loss: 0.3916 -
acc: 0.8245 - val_loss: 0.4658 - val_acc: 0.8005
Epoch 20/25
8000/8000 [==============================] - 69s - loss: 0.3845 -
acc: 0.8280 - val_loss: 0.4415 - val_acc: 0.8045
Epoch 21/25
8000/8000 [==============================] - 69s - loss: 0.3774 -
acc: 0.8281 - val_loss: 0.4406 - val_acc: 0.8130
Epoch 22/25
8000/8000 [==============================] - 69s - loss: 0.3597 -
acc: 0.8382 - val_loss: 0.4444 - val_acc: 0.8105
Epoch 23/25
8000/8000 [==============================] - 69s - loss: 0.3613 -
acc: 0.8404 - val_loss: 0.4327 - val_acc: 0.8120
Epoch 24/25
8000/8000 [==============================] - 69s - loss: 0.3477 -
acc: 0.8444 - val_loss: 0.4628 - val_acc: 0.8025
Epoch 25/25
8000/8000 [==============================] - 69s - loss: 0.3381 -
acc: 0.8516 - val_loss: 0.4486 - val_acc: 0.8180
```

Result:

We have explored about convolutional neural network and used keras in differentiating between images of different animals with 85% accuracy.

We have also discussed in brief about different concepts in deep learning and computer vision.

References:

[1] Kim, Y., 2014. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.

[2] Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).

[3] Vedaldi, A. and Lenc, K., 2015, October. Matconvnet: Convolutional neural networks for matlab. In *Proceedings of the 23rd ACM international conference on Multimedia* (pp. 689-692). ACM.

[4] Lawrence, S., Giles, C.L., Tsoi, A.C. and Back, A.D., 1997. Face recognition: A convolutional neural-network approach. *IEEE transactions on neural networks*, *8*(1), pp.98-113.

[5] Ji, S., Xu, W., Yang, M. and Yu, K., 2012. 3D convolutional neural networks for human action recognition. *IEEE transactions on pattern analysis and machine intelligence*, *35*(1), pp.221-231.

[6] Gatys, L.A., Ecker, A.S. and Bethge, M., 2016. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2414-2423).

[7] Nagi, J., Ducatelle, F., Di Caro, G.A., Cireşan, D., Meier, U., Giusti, A., Nagi, F., Schmidhuber, J. and Gambardella, L.M., 2011, November. Max-pooling convolutional neural networks for vision-based hand gesture recognition. In *2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)* (pp. 342-347). IEEE.

[8] Fulari, S., 2018, May. A Survey on Motion Models Used for Object Detection in Videos. In *2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI)* (pp. 348-353). IEEE.