International Journal of Engineering & Technology, 7 (3.1) (2018) 13-16

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

The Role and Utilization of CNN in Automatic Logo Based Document Image Retrieval Methods

Raveendra K[1]*, R Vinoth Kanna[2]

[1]Research Scholar at K L University, Vijayawada, and
Associate Professor at SV Engineering College for Women, Tirupati
ORCID ID:0000-0001-9695-1908
[2]Professor at K L University, Vijayawada
**Corresponding author E-mail: raveendrasvew@gmail.com*

Abstract

Automatic logo based document image retrieval process is an essential and mostly used method in the feature extraction applications. In this paper the architecture of Convolutional Neural Network (CNN) was elaborately explained with pictorial representations in order to understand the complex Convolutional Neural Networks process in a simplified way. The main objective of this paper is to effectively utilize the CNN in the process of automatic logo based document image retrieval methods.

Keywords: *CNN; Computer Vision; Deep Learning; Image Classification; Machine Learning; Pattern Recognition.*

1. Introduction

The feature extraction is playing a vital role in the automatic logo based document image retrieval methods (Logo is a primary key feature for the image or picture or document identity) in the various primary and essential society needs like Semantic Segmentation, Image Patches, Hand Written Characters recognition, Automatic Scene Text Recognition, Sentiment Analysis of Short Texts, Generic Visual Recognition, Speech Recognition, Image Style Transfer, No-Reference Image Quality Assessment, Large Scale Video Classification, Face Detection, Person Re-Identification, Heterogeneous Multi-task Learning for Human Pose Estimation, Depth estimation from a Single Image, Learning Transferable Features, Online Detection and Classification of Dynamic Hand Gestures, Visual Tracking, Mid-Level Image Representations, Scene Labelling, Image Question Answering, Real-Time Single Image and Video Super-Resolution, Visual Document Analysis, Dropout, 3D Shape Recognition, Fine-grained Image Classification, Learning to Compare Image Patches and Computing the Stereo Matching Cost Closed Circuit Tele Vision (CCTV), Drones, Traffic Surveillance, and Video analysis. The research is going on very rapidly in the invention of new powerful methods for the feature extraction process. The Convolutional Neural Network (CNN) is one of such new powerful method useful for the feature extraction. Even though the CNN was introduced in 1990s, due to its deep learning characteristics it is compatible to use in latest technologies of Object recognition, Computer Vision, Pattern Recognition, Image Classification, Deep Learning, Machine Learning, and Artificial Intelligence applications. The CNN is itself consists of a deep learning architecture which directly deals with the very fundamental particles called neurons. In this Paper we have explained the architecture of CNN in the process of automatic logo based document image retrieval. This paper is having the following structure. Chapter 2 will give the complete architecture of the Convolutional Neural Networks with an example included with detailed pictures (images) where ever required. Chapter 3 gives the processing steps in CNN, Chapter 4 gives the related work done so far using CNN, Chapter 5 gives the future scope of the research work that can be extended in the field of automatic logo based document image retrieval methods. And finally Chapter 6 gives the conclusion.

2. Architecture of CNN

The Convolutional Neural Networks architecture suitable for image processing applications was actually came from the basic neural networks architecture. The neural network is a system consists of both hardware and software arranged or patterned after the operation of neurons in the human brain. The artificial neural networks are very much useful in deep learning technologies to solve the pattern or object recognition problems. Neural networks may use some principles or rules to obtain the required output from the given raw inputs. These rules or principles of the neural networks can be considered as a set of training process to accomplish the given task. These training levels can be considered as hidden layers between the input and output layers of an artificial neural network system. The number of these hidden layers and nodes will depend on the complexity of the problem. The architecture of convolutional neural networks to address for the object or logo recognition in the image analysis can be defined effectively by considering how the machine can recognize a scene or image and how we train the neural network to identify the objects or logos in the image.

In our day-to-day life every second we are seeing large number of images and unknowingly we are processing those images also with the help of our human brain which can be considered as a powerful machine. The natural machines cannot function like our human brain in the process of image or object or logo recognition. The convolutional neural networks architecture will solve this problem. The neural networks should be trained to understand the

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

images and to store them. The Neural network will consider the image as an array of pixels or dots as features to predict the image and to store. This way of training the neural network is a very complex task. In this paper we are giving the easiest way of recognition of image by the machines. Let us assume that there is an image with a number 9 written on it as shown in figure 1. In order to recognize this image by machine the image has to be broken into a pattern (matrix) of dots or pixels and also to recognize the color of each such pixel.

63	63	63	63
45	45	45	45
63	18	18	63
45	18	18	45
63	18	18	63
45	18	18	45
63	63	63	63
45	45	45	45
18	18	18	63
18	18	18	45
18	18	18	63
18	18	18	45
63	63	63	63
45	45	45	45

Fig.1: An Image with Number 9 on it.

For the simplicity of understanding only one black and white colors are considered in this example and it was assumed that the number 256 is the darkest shade of black color and 1 as the white color. The machine will recognize these color values (in terms of pixel values) in simple series fashion as:
63,45,63,45,63,45,63,45,18,18,18,18,63,45,63,45,18,18,18,18,63,4518,18,18,18,63,45,63,45,18,18,18,18,63,45,18,18,18,63,45,63,45,63,45,63,45,63,45,63,45,63,and 45(These pixel values are are considered from top to bottom and then from left to right)
Consisting of totally 56 pixel values.(14 rows x 04 columns. In this example we considered only three different pixel values 63, 45 and 18 only for easy understanding) by seeing this series arrangement of pixel values no one can understand that this is an image written a number 9 on it. This was happened by the machine recognition due to the loss in the spatial (actual) arrangement of pixels in the original image. In order to retain the spatial arrangement of the pixels in actual order we need to take the both the horizontal and vertical directions also in the process of recognition by the machine. This can be done easily by taking a weight in the hidden layer of training algorithm of the machine. The selected weight value will enhance or decrease the pixel color values. If the weight is higher value then the color of the corresponding pixel will enhance and will give a better indication for the machine to recognize the color of the pixel. Similarly if the weight value is a low value(less than one) will decrease (diminish) the pixel color value.
The figure 2 will give the enhancement details in pixel values by taking the weight as 3.

63	63	63	63			189	189	189	189
45	45	45	45			135	135	135	135
63	18	18	63			189	54	54	189
45	18	18	45			135	54	54	135
63	18	18	63			189	54	54	189
45	18	18	45			135	54	54	135
63	63	63	63	3		189	189	189	189
45	45	45	45	Weight		135	135	135	135
18	18	18	63			54	54	54	189
18	18	18	45			54	54	54	135
18	18	18	63			54	54	54	189
18	18	18	45			54	54	54	135
63	63	63	63			189	189	189	189
45	45	45	45			135	135	135	135

Fig. 2: Image with Enhancement in Pixel values for Number 9

The machine will recognize these enhanced color values in simple series fashion as 189,135,189,135, … 189,135,189,135(Totally 56 elements). The figure 3 will give the enhancement details in pixel values by taking the weight as 0.5. In this case The entire was diminished and there may be a wrong recognition that there is no number 9 written on the image due to the loss in pixel color values by 50% as the considered weight value of 0.5. Anyway in this case also the machine will recognize the diminished color pixel values as 31.5,22.5,31.5,22.5 … 31.5,22.5,31.5,22.5. Higher the weight value higher is the pixel color value and very easy for visual perception. Lower the weight value lower is the pixel color value and difficult for visual perception.

63	63	63	63			31.5	31.5	31.5	31.5
45	45	45	45			22.5	22.5	22.5	22.5
63	18	18	63			31.5	9	9	31.5
45	18	18	45			22.5	9	9	22.5
63	18	18	63			31.5	9	9	31.5
45	18	18	45			22.5	9	9	22.5
63	63	63	63	0.5		31.5	31.5	31.5	31.5
45	45	45	45			22.5	22.5	22.5	22.5
18	18	18	63			9	9	9	31.5
18	18	18	45			9	9	9	22.5
18	18	18	63			9	9	9	31.5
18	18	18	45			9	9	9	22.5
63	63	63	63			31.5	31.5	31.5	31.5
45	45	45	45			22.5	22.5	22.5	22.5

Fig. 3: Image with Decrement in Pixel values for Number 9

But in both the cases the machine has not preserved the spatial arrangement of pixels. In order to preserve the spatial arrangement the machine has to consider the both horizontal and vertical directions at least with weight matrix size of 2x2 (2 rows and two columns totally four weights). By taking the window having the total four weights as 1,1,0.3 and 2 the color pixel values of all the elements are shown in figure 4.

63	63	63	63			198	198	198
45	45	45	45			122.4	108.9	198.9
63	18	18	63			121.5	68.4	144.9
45	18	18	45			108.9	68.4	171.9
63	18	18	63			121.5	68.4	144.9
45	18	18	45			198.9	171.9	185.4
63	63	63	63	1	1	198	198	198
45	45	45	45	0.3	2	108.9	108.9	198.9
18	18	18	63			68.4	68.4	144.9
18	18	18	45			68.4	68.4	171.9
18	18	18	63			68.4	68.4	144.9
18	18	18	45			171.9	171.9	185.4
63	63	63	63			198	198	198
45	45	45	45					

Fig. 4: Image with four weight values1,1,0.3 and 2 for Number 9

Now the machine recognizes all the pixel values as 198,122.4,121.5,108.9,121.5,198.9,198,108.9,68.4,68.4,68.4,171.9,198,198,108.9,68.4,68.4,68.4,171.9,198,108.9,68.4,68.4,68.4,171.9,198,198,198.9,144.9,171.9,144.9,185.4,198,198.9,144.9,171.9,144.9,171.9,144.9,185.4,198(totally 39 elements only). The reduction in the total number of elements from 56 to 39 is due to the movement of 2x2 sized window on original image with sliding of one element in both horizontal and vertical directions. Here the spatial arrangement of pixels are preserved. This method of preserving the spatial arrangement along with the image recognition by the machine is exactly similar that a convolutional neural networks does. By taking an input image first then defining weight matrix size and then the input is convolved to the required specific features without losing the spatial arrangement also. Therefore in order to train the machine using artificial neural network the Neural network system should consists of at least three fundamental or basic components: The Convolutional Layer, The Pooling Layer and the output layer.
The convolutional layer will take care of the recognition and spatial arrangement preservation as shown in the above figure4.

3. Processing Steps in CNN

Let us assume that there is an image of size 6x6. And also assume that the certain feature we want to extract from this 6x6 image is of size 3x3. Now the below figure 5 will explain this process of feature extraction.

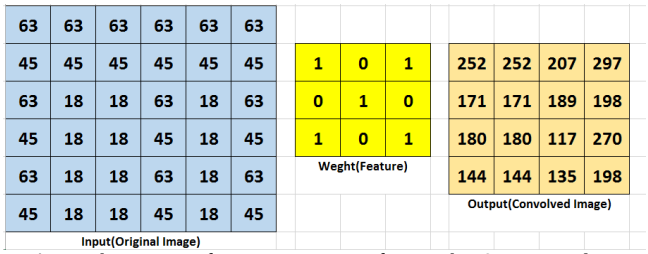


Fig. 5: The process of Feature extraction for Number 9 using strides:1

In the convolved output image the pixel values are obtained as: For the first element its corresponding convolved output is 252=(63*1)+(63*0)+(63*1)+(45*0)+(45*1)+(45*0)+(63*1)+(18*0)+(18*1)=63+0+63+0+45+0+63+0+18=252.

Similarly last element

198=(45*1)+(18*0)+(45*1)+(63*0)+(18*1)+(63*0)+(45*1)+(18*0)+(45*1)=45+0+45+0+18+0+45+0+45=198.

The final convolved output image is of size 4x4 only for the input image size of 6x6 and weight matrix size of 3x3. Here the weight actually plays the main role of feature extraction from the input image. The feature may be an edge or color or unwanted noise or some other specific character of the image.The weight matrix can be cosidered as a hyperparameter or simply as a filter which moves (strides) across the entire image one pixel at a time. Moving the hyperparameter (some required feature) by one pixel at a time is time consuming or slow process. Now we will see what happens if we move the the weight matrix filter by two pixels at a time for the same input image as shown in figure6.

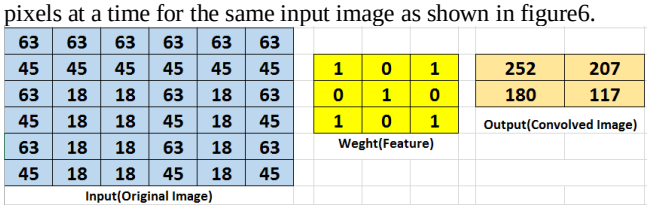


Fig. 6: The process of Feature extraction for Number 9 using strides:2

Now The final convolved output image is of size 2x2 only for the input image size of 6x6 and weight matrix size of 3x3 (with pixel values of 252,180,207 and 117). In the similar way if we move or stride the weght matrix by three pixels at a time there remains only one element in the output matrix of value 252. So finally as the stride value increases the corresponding decrement happens in the size of output image matrix. The decrement in size of the output matrix indicates that feature mappings are low. Therefore we need a compromise between the speed (more stride value) and accuracy (Big size of Convolved Outpot Image). To have this compromise between speed and accuracy we can apply method called the padding.Padding is nothing but adding the zeros around the input image inorder to increase the size of the output image matrix without decresing the stride value and hence the speed. More than one layer of zeros can also be added around the image for more stride values by retaing the original input image features without any change. The one layer of padding to the input image is shown in the figure7 below:

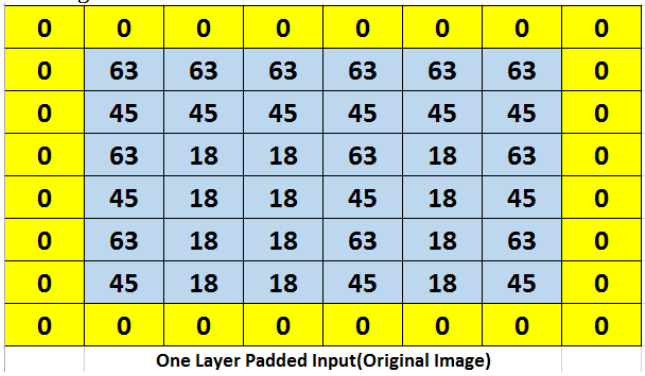


Fig.7: One layer padding

Similarly the multiple layer padding can also be done for the applications where the quality of the image dominates than the accuracy in speed. Therefore the three processing steps: convolution, striding and padding are the primary processing steps in Automatic logo based document image retrieval methods.

4. Related Work

With the high range performance of CNN it was impressed so many researchers to work and get succeed in the fields of Semantic Segmentation [1], Image Patches[2], Hand Written Characters recognition[3], Automatic Scene Text Recognition[4], Sentiment Analysis of Short Texts[5], Generic Visual Recognition[6], Speech Recognition[7], Image Style Transfer[8], No-Reference Image Quality Assessment[9], Large Scale Video Classification[10], Face Detection[11], Person Re-Identification[12], Heterogeneous Multi-task Learning for Human Pose Estimation[13], Depth estimation from a Single Image[14], Learning Transferable Features[15], Online Detection and Classification of Dynamic Hand Gestures[16], Visual Tracking[17], Mid-Level Image Representations[18], Scene Labelling[19], Image Question Answering[20], Real-Time Single Image and Video Super-Resolution[21], Visual Document Analysis[22], Dropout[23], 3D Shape Recognition[24], Fine-grained Image Classification[25], Learning to Compare Image Patches[26] and Computing the Stereo Matching Cost[27].

5. Future Scope

Due to its high performance, The CNN architecture can be utilized in all feature extraction applications like Computer vision, Pattern Recognition, Deep and Machine Learning applications and can also be extended to the Motional Applications also to have better performance in the image analysis. I am promising here that the CNN architecture will be utilized in my research problem "Automatic logo based document image retrieval methods" in future.

6. Conclusion

In this paper we have explained the architecture of CNN in a simplified manner to understand easily in terms of the basic fundamental pixels or dots language of image or picture. We hope this paper will definitely create a platform for the researchers in understanding the architecture of CNN and how it can be utilized for image analysis problems.

Acknowledgement

I heart fully thank my supervisor Dr. R.Vinoth Kanna sir for his continuous supportive encouragement in my research work done so far including this paper.

References

[1] Hyeonwoo Noh, Seunghoon Hong, Bohyung Han, "Learning Deconvolution Network for Semantic Segmentation", http://doi.ieeecomputersociety.org/10.1109/ICCV.2015.178 pp.1520-1528

[2] Jure Zbontar, Yann LeCun, "Stereo Matching by training a convolutional neural network to compare image patches", *Journal of Machine Learning Research 17(2016) 1-32*

[3] Yoshua Bengio, Yan Le Cun, Donnie Henderson, "Globally Trained Handwritten Word Recognizer using Spatial Representation, Convolutional Neural Networks and Hidden Markov Models"

[4] Zohra Saidane, Christophe Garcia, "Automatic Scene Text Recognition using a Convolutional Neural Network"

[5] Cicero Nogueira dos Santos, Maira Gatti, "Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts", proceedings of COLING 2014, THE 25th International Conference on Computational Linguistics: Technical Papers, pages 69-78, Dublin, Ireland, August 23-29 2014.

[6] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, Trevor Darrell "DeCAF: A Deep convolutional Activation Feature for Generic Visual Recognition" proceedings of the 31st International Conference on Machine Learning, Beijing, China 2014. JMLR:W&CP volume32.

[7] Ossama Abdel-Hamid, Li Deng, Dong Yu "Exploring Convolutional Neural Network Structures and Optimization Techniques for Speech Recognition INTERSPEECH 2013 25-29 August 2013, Lyon, France

[8] Leon A.Gatys, Alexander S.Ecker, Matthias Bethge "Image Style Trnasfer Using Convolutional Neural Networks" http://doi.ieeecomputersociety.org/10.1109/CVPR.2016.265 pp:2414-2423

[9] Le Kang, Peng Ye, Yi Li, David Doermann "Convolutional Neural Networks for No-Reference Image Quality Assessment"

[10] Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, Li Fei-Fei " Large Scale Video Classification with Convolutional Neural Networks" http://doi.ieeecomputersociety.org/10.1109/CVPR.2014.223 pp:1725-1732

[11] Haoxiang Li, Zhe Lin, Xiohui Shen, Jonathan Brandt, Gang Hua "A Convolutional Neural Network Cascade for Face Detection" in CVPR, 2015, pp. 5325–5334.

[12] Wei Li, Rui Zhao, Tong Xiao, Xiaogang Wang " DeepReID: Deep Filter Pairing Neural Network for Person Re-Identification" http://doi.ieeecomputersociety.org/10.1109/CVPR.2014.27pp:152-159

[13] Sijin Li, Zhi-Qiang Liu, Antoni B. Chan "Hetrogeneous Multi-task Learning for Human Pose Estimation with Deep Convolutional Neural Network" *in CVPR, 2014,pp.488-495*

[14] Fayao Liu, Chunhua Shen, Guosheng Lin "Deep Convolutional Neural Fields for Depth Estimation from a Single Image" *in CVPR, 2015,pp.5162-5170*

[15] Mingsheng Long, Yue Cao, Jianmin Wang, Michael I.Jordan " Learning Transeferable Features with Deep Adaptation Networks" proceedings of the 32nd International Conference on Machine Learning, Lille, France, 2015. JMLR: W&CP volume 37.

[16] Pavlo Molchanov, Xiaodong Yang, Shalini Gupta, Kihwan Kim, Stephen Tyree, Jan Kautz "Online Detection and Classification of Dynamic Hand Gestures with Recurrent 3D Convolutional Neural Networks" 2016 IEEE Conference on Computer Vision and Pattern Recognition ISBN: 978-1-4673-8851-1

[17] Hyeonseob Nam, Bohyung Han " Learning Multi-Domain Convolutional Neural Networks for Visual Tracking" *Computer vision foundation in CVPR, 2016,pp.4293-4302*

[18] Maxime Oquab, Leon Bottou, Ivan Laptev, Josef Sivic "Learning and Transferring Mid-Level Image Representations using Convolutional Neural Networks' *CVPRW, 2015,pp.91-96*

[19] Pedro O.Pinheiro, Ronan Collobert "Recurrent Convolutional Neural Networks for Scene Labelling" proceedings of the 31st International Conference on Machine Learning, Beijing, China,2014. JMLR: W&CP volume 32.

[20] Hyeonwoo Noh, Paul Hongsuck Seo, Bohyung Han " Image Question Answering using Convolutional Neural Network with Dynamic Parameter Prediction" *in CVPR,2016,pp.30-38*

[21] Wenzhe Shi, Jose Caballero, Ferenc Huszar, Johannes Totz, Andrew P.Aitken,Rob Bishop, Daniel Rueckert, Zehan Wang " Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network" *in CVPR,2016, ISBN: 978-1-4673-8851-1*

[22] Patrice Y.Simard, Dave Steinkraus, John C. Platt " Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis"

[23] Nitish Srivastava, Geoffry Hinton, Alex Krizhevsky, Ilya Sutskever, Ruslan Salakhutdinov "Dropout: A Simple way to Prevent Neural Networks from Overfitting" Journal of Machine Learning Research 15 (2014) 1929-1958

[24] Hang Su, Subharansu Maji, Evangelos Kalogerakis, Erik Learned-Miller " " *Computer vision foundation*

[25] Tianjun Xiao, Yichong Xu, Kuiyuan Yang, Jiaxing Zhang, Yuxin Peng, Zheng Zhang " The Application of Two-level Attention Models in Deep Convolutional Neural Network for Fine-grained Image Classification" *in CVPR, 2015,pp.1063-6919*

[26] Sergey Zagoruyko, Nikos Komodakis "Learning to Compare Image Patches via Convolutional Neural Networks" *in CVPR, 2015,pp.4353-4361*

[27] Jure Zbontar, Yann LeCun "Computing the Stereo Matching Cost with a Convolutional Neural Network" *Computer vision foundation*