Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 6.017

IJCSMC, Vol. 7, Issue. 12, December 2018, pg.191 – 196

A Survey on the Layers of Convolutional Neural Networks

Ali Fadhil Yaseen

Control and System Engineering Department, University of Technology, Baghdad, Iraq

Email: sydali85@yahoo.com

Abstract— The field of computer vision has seen a significant development by used Convolutional Neural Networks. Recognize for objects are the essential parts of a computer vision. The detection stage and the classification stage can be implemented one step using Convolutional Neural Networks. The detection and classification for objects must be using monitoring this object temporal and spatial transforming in a sequence the video, shape, with presence, size and location etc. This paper provides a brief overview of the various the layers of Convolutional Neural Networks algorithms presented in the literature, containing comparative and analysis study of different techniques used for different stages.

Keywords— Convolutional Neural Networks, CNN layers, Object Detection, Object Classification, video processing

I. INTRODUCTION

A convolutional neural network, symbolled as CNN, or ConvNet, in the world of machine learning, is considered as one of the most popular algorithms for deep learning, feed-forward artificial neural networks, most commonly applied to analyse visual imagery. CNNs use a variation of multilayer perceptrons designed to require minimal pre-processing [1] . Such networks are characterised by being multi-layered, and so they are large and they are consequently complex. Training and evaluating networks with such characteristics require extensive computing systems [2]. Due to their architecture of shared-weights and due to their characteristics of translation invariance, these networks are also called “shift invariant” type of artificial neural networks and symbolled as SIANN [3].

The convolutional network, like other ANNs, is a try of mimicking biological operations. Its pattern of connectivity among neurons mimics the structure of visual cortex found in animals [4]. A specific cortical neuron acts to motivations just in a limited part of the visual area defined as the receptive area. Partially, the receptive areas for the diverse neurons overlap. They overlap in such a way that all of the visual area is covered.

Relative to other algorithms used for classification of image, CNNs need a slight pre-processing. More specifically, unlike traditional algorithms in which filters are engineered by hand, these networks are capable of learning filters. Having such a feature of independency from the human interaction or effort when it comes to the design of feature, is really a vital privilege.

The outcomes of recent breakthroughs in the world of developing convolutional neural networks being multi-layered, were the formation of state of the art advances in the accuracy of tasks of recognition. The classification of images of large-category and recognition of speech automatically are just two sample cases regarding these

advances [1]. Some of the related applications are in recognition of images as well as videos, systems designed for recommending [5] and processors of the natural language [6].

A typical way to establish the state-of-the-art deep CNNs is through forming them into two neural network layers, the first of which is alternating convolutional, and the second of which is the max-pooling. These are to be succeeded by some layers that are dense and that are fully-connected. This is depicted by the topology sample given in figure (1) [1]. Any of the three dimension volumes stands for a layer input, and is converted into another three dimensional volume to feed the next layer. In the example, considered, the number of convolutional layers is five, the number of layers of max-pooling type is three, and that is in addition to three layers being fully-connected.

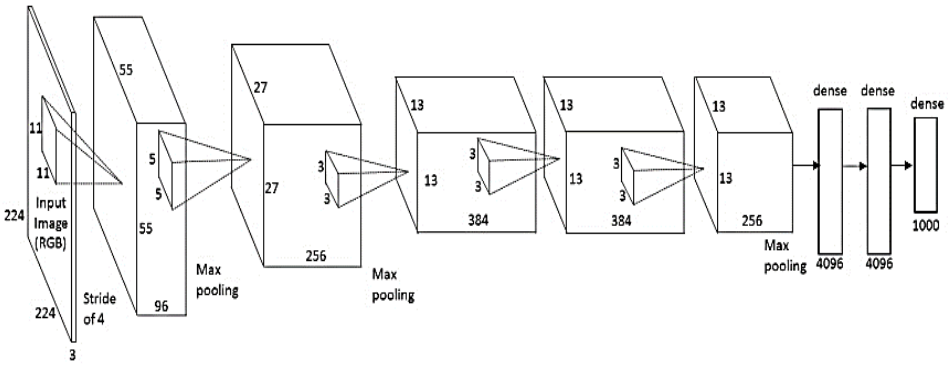

Fig. 1 A deep CNN network used for classification of images [1].

Since the breakthrough in 2012 ImageNet competition achieved by AlexNet which represents the first entry that used a Deep Neural Network (DNN), several other DNNs with increasing complexity have been submitted to the challenge in order to achieve better performance [7].

A convolution process in CNNs is described by convention. When it comes to mathematics, rather than convolution, such a process is a cross-correlation. In what follows, the names of the layers, how these layers work, and some of the concepts needed in the work of CNN are given.

II. LITERATURE SURVEY

Convolutional Neural Networks are class of deep learning a biologically inspired models that replace a single neural network that is trained all three stages form end to end from pixel raw values to outputs classifier[8].

large networks trained features learned by on ImageNet [9] yield state-of-the-art performance have been shown, even with no fine-tuning to across many standard image recognition datasets when classified with an SVM [10]. One of the state of the art is Deep learning considered as subjects In recent years, the great success the deep learning has achieved in many fields, such as natural language processing and computer vision. machine learning methods Compared to traditional, deep learning has can make better use of datasets and a strong learning ability for feature extraction [11]**.** A powerful set of techniques for learning deep the learning provides in neural networks [12]. Deep learning is represents a big step forward and a recent approach to artificial intelligence [13].

III.STUDIES RELATED TO THE LAYER OF CONVOLUTION TYPE

The first layer used to abstract features from an image being input is the convolution layer. The relationship among pixels are conserved by convolution through learning the features of the image depending small squares of the data being input. As a mathematical operation, convolution deals with two inputs one of which is a kernel or filter, and the other is an image matrix. This is depicted in figure (2). In this figure we have:

- An image matrix of dimensions: height (h), width (w), and depth (d).
- A filter of dimensions: filter height (f_h) filter width (f_w), and depth (d).
- An output with volume dimensions: $(h - f_h + 1) * (w - f_w + 1) * 1$.

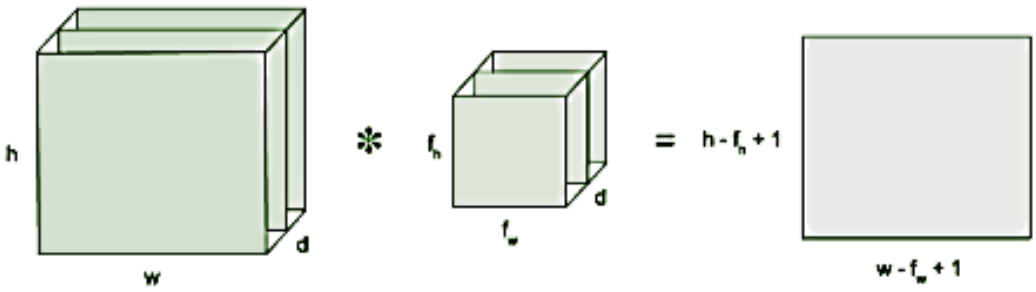

Fig. 2 Image matrix multiplication with the kernel or filter matrix.

IV. STUDIES RELATED TO STRIDES

In CNNs, a stride represents the number of shifts of pixels in the input matrix. If a stride is of value one, then at a time, the filters are moved one pixel. In case the stride value is two, then at a time, the filters are moved two pixels. And other values of stride are considered similarly. Figure (3) explains the way convolution works when stride value is two.

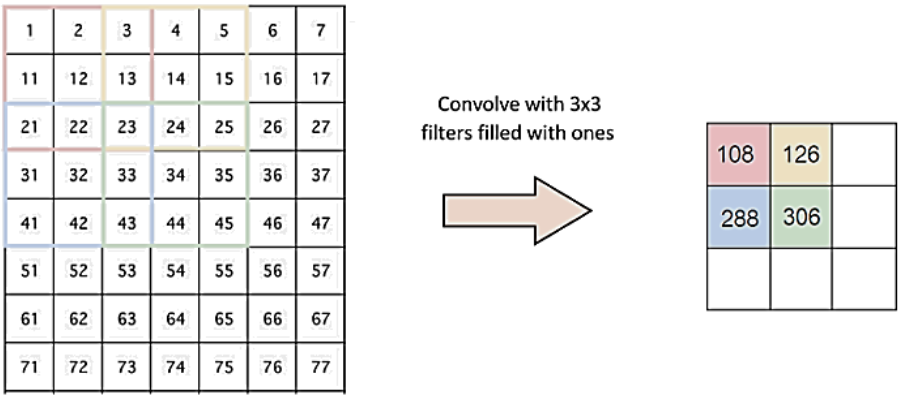

Fig. 3 An example of a stride with a value of two pixels.

V. STUDIES RELATED TO THE PADDING

In some occasions, the filters do not perfectly be a fit for the considered input image. In such a case, two options exist:

- Zero-padding: Simply, pad zeros in the picture to get a fit.
- Remove the image part where the filter do not fit. Such a thing is known as valid padding that reserves just the image valid part.

VI. STUDIES RELATED TO THE (RELU) NON LINEARITY

The abbreviation ReLU means "Rectified Linear Unit" and it's a non-linear operation. ReLU is significant and is intended for inserting non-linearity in the CNN. ReLU allows for faster and more effective training by mapping negative values to zero and maintaining positive values. This is sometimes referred to to as activation, because only the activated features are carried forward into the next layer. ReLU operation is shown in figure (4). Other nonlinearities do exist like the (tanh) or like the (sigmoid) that can replace the ReLU. Most researchers or designers use ReLU for the improved performance achieved with it.

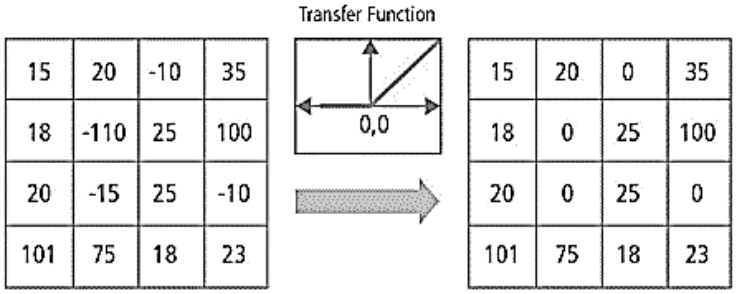

Fig. 4 ReLU operation.

VII. STUDIES RELATED TO THE LAYER OF POOLING

This section contributes in parameters' number reduction in case the images are so big. Spatial type pooling, also known as subsampling or down sampling, leads to dimensionality reduction of each map and at the same time keeps information of importance. Some of the types of spatial type pooling are: the sum pooling, the max pooling and the average pooling. Max pooling assumes the biggest element in the modified feature map. Assuming the biggest element could also assume the average pooling. Summing all feature map elements produces the sum pooling. The Max type pooling is depicted in figure (5).

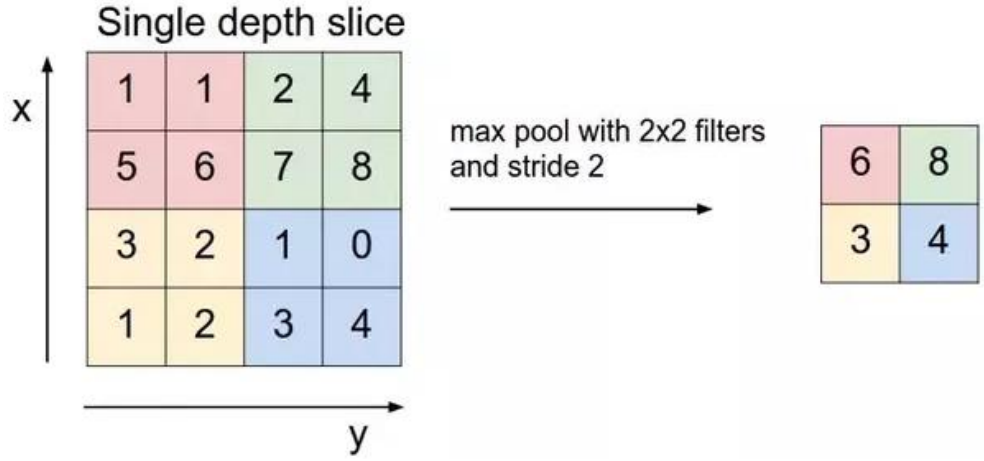

Fig. 5 Max type Pooling.

VIII. STUDIES RELATED TO THE FULLY CONNECTED TYPE LAYER (FC)

In the fully connected (FC) type layer, the matrix of interest is flattened to a vector form and input, like with neural network, into a fully connected type layer. This is shown in figure (6). This FC layer comes after the pooling layer.

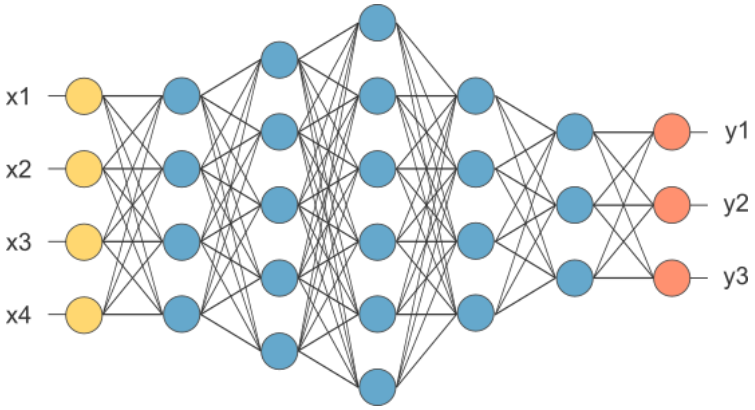

Fig. 6 The step of flatting the matrix as FC layer.

A complete CNN architecture is shown in Figure (7). In this architecture the matrix of the feature map will be turned into vectors (x1, x2, x3, …). The creation of a model is achieved by a combination of features using the FC layers. At last, one of the activation functions, like "softmax" or "sigmoid" , is used for output classifying into categories, like for example, a tree, a house, a cat, a boat, and so on.

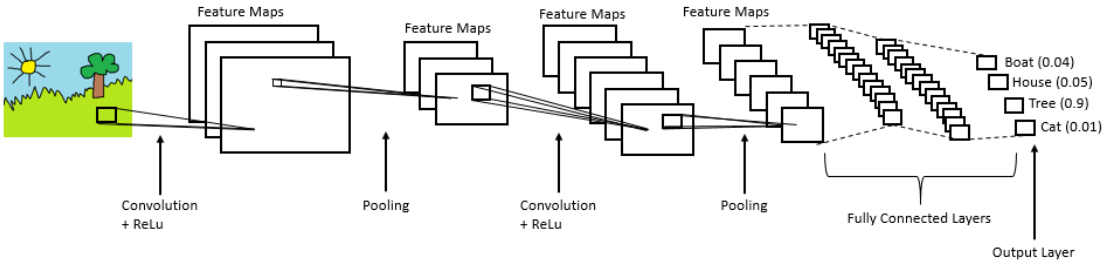

Fig. 7 An overall architecture of a CNN.

IX. CONCLUSIONS

In this paper, Convolutional Neural Networks techniques are dissected in detail by using the layers of the Convolutional Neural Networks techniques such as the detection of an object and the classification of an object. This approach used to increase the detection and the classification in one step of an object with new ideas. Some of the border and interest are also considered for Convolutional Neural Networks in the image sequences. We have observed some methods that have high computational complexity but give precision to results.

REFERENCES

[1] I. Krizhevsk, Alex Sutskever and G. E. Hinton, "ImageNetClassificationWith DeepConvolutionalNeural Networks," *Adv. Neural Inf. Process. Syst.*, 2012.

[2] T. Chilimbi, Y. Suzue, J. Apacible, and K. Kalyanaraman, "ProjectAdam:BuildingAnEfficientAndScalableDeepLearningTrainingSystem," *11th USENIX Symp. Oper. Syst. Des. Implement.*, pp. 571–582, 2014.

[3] W. Zhang, K. Itoh, J. Tanida, and Y. Ichioka, "Parallel distributed processing model with local space-invariant interconnections and its optical architecture," *Appl. Opt.*, 1990.

[4] D. L. K. Yamins, H. Hong, C. F. Cadieu, E. A. Solomon, D. Seibert, and J. J. DiCarlo, "Performance-optimized hierarchical models predict neural responses in higher visual cortex," *Proc. Natl. Acad. Sci.*, vol. 111, no. 23, pp. 8619–8624, 2014.

[5] A. van den Oord, S. Dieleman, and B. Schrauwen, "Deep content-based music recommendation," *Electron. Inf. Syst. Dep.*, 2013.

[6] R. Collobert and J. Weston, "A unified architecture for natural language processing: Deep neural networks with multitask learning," *Proc. 25th Int. Conf. Mach. Learn.*, 2008.
[7] A. Canziani, A. Paszke, and E. Culurciello, "An Analysis of Deep Neural Network Models for Practical Applications," *arXiv:1605.07678v4*, 2017.
[8] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, 1998.
[9] Jia Deng, Wei Dong, R. Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009.
[10] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: An astounding baseline for recognition," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2014.
[11] X. Du, Y. Cai, S. Wang, and L. Zhang, "Overview of deep learning," *Proc. - 2016 31st Youth Acad. Annu. Conf. Chinese Assoc. Autom. YAC 2016*, pp. 159–164, 2017.
[12] M. Nielsen, "Neural networks and deep learning," 2017. [Online]. Available: http://neuralnetworksanddeeplearning.com/.
[13] A. Goodfellow, Ian, Bengio, Yoshua, Courville, "Deep Learning," *MIT Press*, 2016.

