УДК 004.93
МРНТИ 50.05.19

NEURAL NETWORKS TO CONVOLUTIONAL NEURAL NETWORKS: EXPANSION AND DETAILED EXPLANATION

YAKUFUJIANG AZATI[1], F. MALIKOVA[1,2], A. TEMIRBEKOV[1], S. KENZHEGULOVA[3]

[1]Al-Farabi Kazakh National University
[2]Almaty University of Power Engineering and Telecommunications
[3]University of International Business

Abstract: *In the last century, scientists discovered several visual neurological features. The optic nerve has a local receptive field. The recognition of a whole picture is composed of multiple local recognition points. Different neurons have the ability to recognize different shapes, and the optic nerve has superposition ability. The pattern can be composed of low-level simple lines. Later, people found that after the operation of the concatenation, the process of optic nerve processing calculation can be well reflected. The LeNet-5, which was invented by LeCun in 1998[1], can greatly enhance the recognition effect. This article mainly focuses on the neural network evaluation, from neural networks to convolutional neural networks, convolutional layer, the pooling layer, and the overall CNN structure.*

Keywords: *NN, CNN, Python, Computer Vision, Machine Learning*

НЕЙРОНДЫҚ ТОРАПТАРҒА АРНАЛҒАН НЕЙРОНДЫҚ ЖЕЛІЛЕР: КЕҢЕЙТУ ЖӘНЕ ТОЛЫҚ ТҮСІНДІРУ

Аңдатпа: *Бұл мақала негізінен нейрондық желілерді, нейрондық желілерден бастап ұюға дейінгі нейрондық желілерді, ұю қабатын, біріктіру қабатын бағалауға арналған. Өткен ғасырда ғалымдар бірнеше визуалды неврологиялық функцияларды тапты. Көру нервінің жергілікті рецептивті өрісі бар. Тұтас суретті тану бірнеше жергілікті тану нүктелерінен тұрады. Әртүрлі нейрондар әрқилы формаларды тануға қабілетті, ал көру нерві суперпозицияның қабілетіне ие екені байқалды.*

Түйінді сөздер: *NN, CNN, Python, компьютерлік көру, машиналық оқыту*

НЕЙРОННЫЕ СЕТИ ДЛЯ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ: РАСШИРЕНИЕ И ПОДРОБНОЕ ОБЪЯСНЕНИЕ

Аннотация: *Эта статья в основном посвящена оценке нейронных сетей, от нейронных сетей до сверточных нейронных сетей, сверточного слоя, слоя объединения. В прошлом веке ученые обнаружили несколько визуальных неврологических функций. Зрительный нерв имеет локальное рецептивное поле. Распознавание целого изображения состоит из нескольких локальных точек распознавания. Различные нейроны обладают способностью распознавать различные формы, а зрительный нерв обладает способностью суперпозиции.*

Ключевые слова: *NN, CNN, Python, компьютерное зрение, машинное обучение*

INTRODUCTION

We know that the structure of a general neural network is like this:

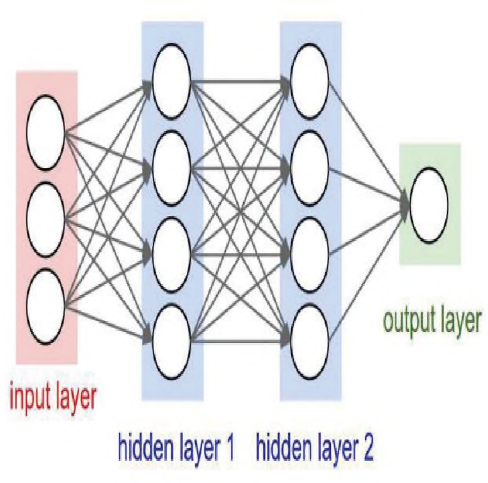


Figure 1.1 - structure of general neural networks

Then, what's the relationship between a general neural network and a convolutional neural network?

In fact, the convolutional neural network is still a hierarchical network, but the function and form of the layer have changed. It can be said that it is an improvement of the general neural network [2]. For example, in the figure below (fig 1.2), there are many levels that are not found in general neural networks.

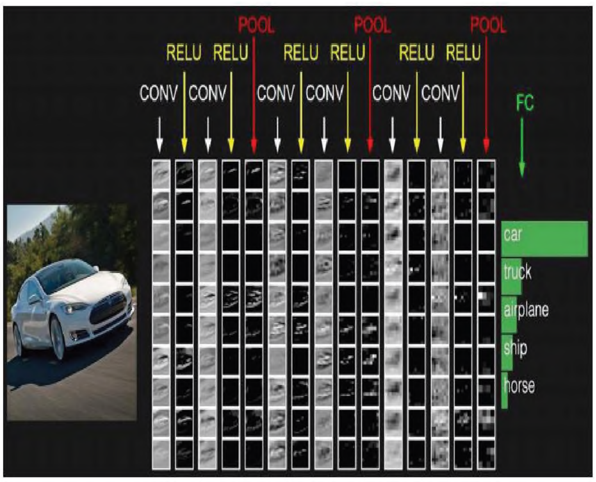


Figure 1.2 - A convolutional neural network for image recognition

1.1 CONVOLUTIONAL NEURAL NETWORK HIERARCHICAL STRUCTURE OF CONVOLUTIONAL NEURAL NETWORKS

A standard convolutional neural network generally includes the following hierarchical network structures [3].

- Data input layer / Input layer
- Convolution calculation layer / CONV layer
- ReLU excitation layer / ReLU layer
- Pooling layer / Pooling layer
- Fully connected layer / FC layer

1.2 DATA INPUT LAYER

The processing to be performed by the input layer is mainly to preprocess the original image data, including:

- De-average: Centers each dimension of the input data to 0, as shown in the following figure. The purpose is to pull the center of the sample back to the origin of the coordinate system.
- Normalization: The amplitude is normalized to the same range, as shown below, which reduces the interference caused by the difference in the range of values of each dimension. For example, we have two dimensions of features A and B, and the range of A is 0. To 10, and the range of B is 0 to 10000. If the two features are directly used, it is a good practice to normalize, that is, the data of both A and B becomes 0 to 1.
- PCA/whitening: using PCA to reduce dimensionality. Whitening is normalized to the amplitude of each characteristic axis of the data [4].

De-average and normalization effect (fig 2.1):

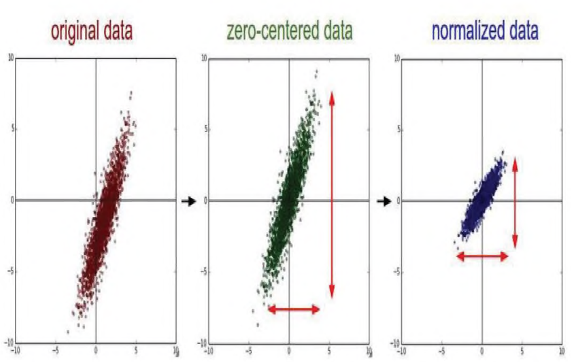


Figure 2.1 - De-average and normalization effect

De-correlation and whitening effect (fig 2.2):

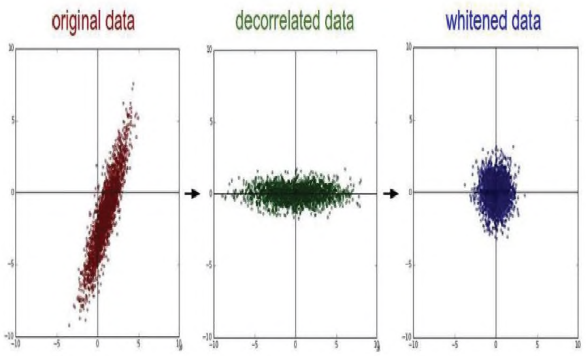


Figure 2.2 - De-correlation and whitening effect.

1.3 CONVOLUTION CALCULATION LAYER

This layer is the most important level of convolutional neural networks and the source of the name of the Convolutional Neural Network. At this convolution level, there are two key operations:

- Local association. Each neuron is seen as a filter
- Window (receptive field) sliding, filter for local data calculation

The relationship between depth, stride (the length of the window sliding), and the padding value (zero-padding) is as follows (Figure 2.3 -):

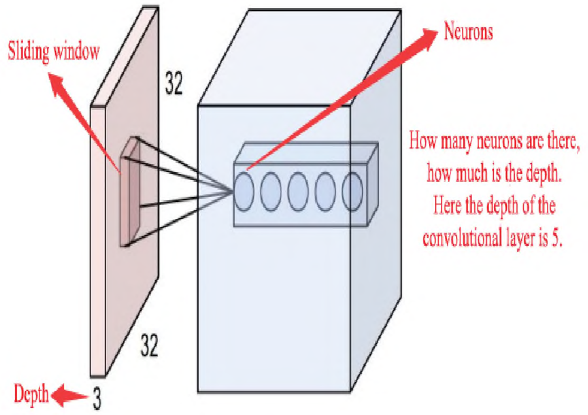


Figure 2.3 - Stride, Depth and Neurons

What is the zero-padding? The following picture (Figure 2.4 -) is an example. For example, if there is such a 5*5 picture (one grid and one pixel), our sliding window takes 2*2 and the step size takes 2, then we find that there is still one pixel left to slide, so how?

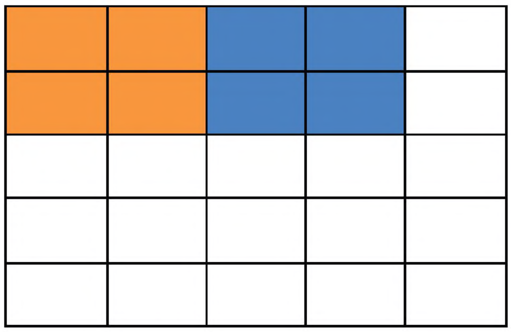


Figure 2.4 - Sliding

Then we add a zero layer to the original matrix (Figure 2.5 -), so that it becomes a 6*6 matrix, then the window can just traverse all the pixels. This is what the zero-padding does.

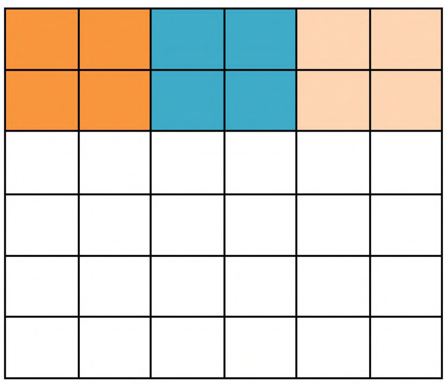


Figure 2.5 - Zero-padding

Convolution calculation (note that there is a circle of gray boxes around the blue matrix below, those are the zero-padding mentioned above (Figure 2.6 -))

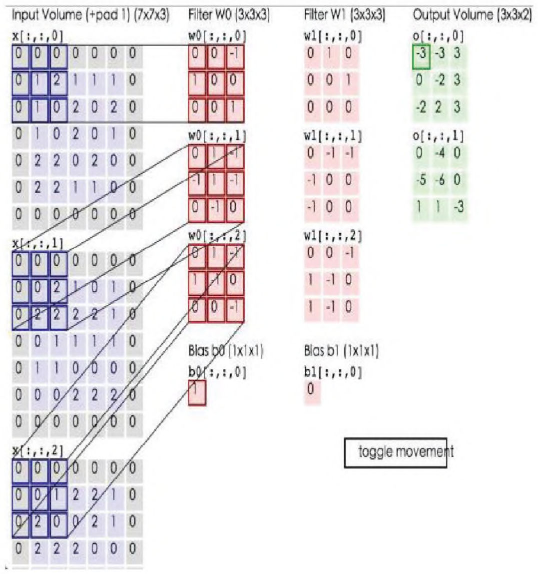


Figure 2.6 - Convolutional calculation (with zero-padding)

The blue matrix here is the input image, and the pink matrix is the convolutional layer of neurons, which shows that there are two neurons (w0, w1). The green matrix is the output matrix after the convolution operation, here the step size is set to 2.

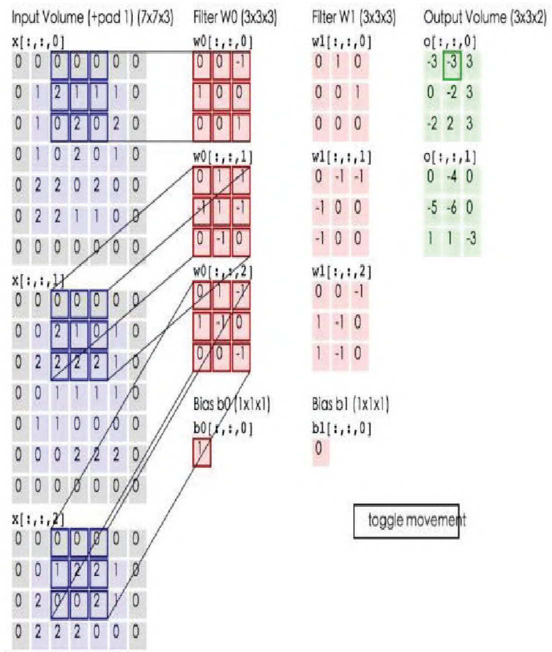
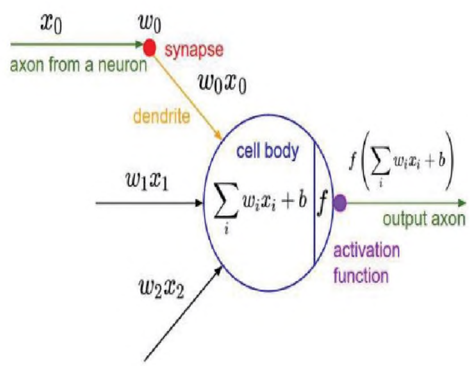


Figure 2.7 - Convolutional layer calculation

The blue matrix (input image) performs a matrix inner product calculation on the pink matrix and adds the result of the three inner product operations to the offset value b (such as the calculation of the above figure: 2+(-2+1) -2) +(1-2-2) + 1= 2 - 3 - 3 + 1 = -3), the calculated value is an element of the green box matrix.

1.4 INCENTIVE LAYER

The excitation layer acts to nonlinearly map the convolutional layer output. The excitation function used by CNN is generally ReLU[5][6] (The Rectified Linear Unit), which is characterized by fast convergence and simple gradient, but is weak, and the image is as follows.



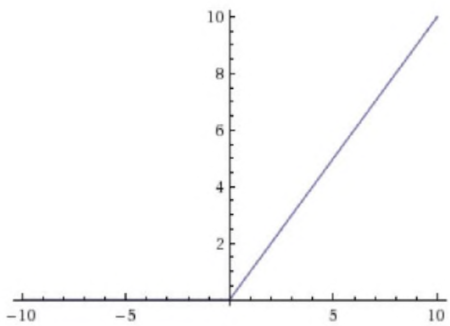


Figure 2.8 - ReLU layer

1.5 POOLING LAYER

The pooling layer is sandwiched between successive convolutional layers to compress the amount of data and parameters, reducing overfitting. In short, if the input is an image, then the primary role of the pooling layer is to compress the image. Here We again expand the specific role of the pooling layer:

1. Feature invariance, which is the scale invariance of the features we often mention in

image processing. The pooling operation is the resize of the image. Usually, the image of a dog is doubled. We can recognize that this is a dog. The photo shows that the most important feature of the dog remains in this image. We can judge that the image is a dog in the image. The information removed during image compression is just some insignificant information, leaving Information is a feature with scale invariance and is the feature that best expresses images.

2. Feature dimension reduction, we know that an image contains a lot of information, there are many features, but some information does not have much use or duplication for us to do image tasks, we can remove such redundant information, put the most The extraction of important features is also a major part of the pooling operation.

3. To some extent, to prevent over-fitting, it is more convenient to optimize.

1.6 FULLY CONNECTED LAYER

All neurons between the two layers have the right to reconnect, usually the fully connected layer is at the end of the convolutional neural network. That is, the connection with traditional neural network neurons is the same [7] (fig 1.1)

2. CONCLUSION

A convolutional network is essentially an input-to-output mapping that learns a large number of mappings between input and output without the need for any precise mathematical expression between input and output, as long as it is known. The mode trains the convolutional network, and the network has the ability to map between input and output pairs.

A very important feature of CNN is that it is top-heavy (the smaller the input weight, the more the output weight is, the more the output weight is), which presents an inverted triangle shape, which avoids the backpropagation in BP neural network. The gradient is lost too fast [8].

Convolutional neural network CNN is mainly used to identify two-dimensional graphics of displacement, scaling and other forms of distortion invariance. Since CNN's feature detection layer learns through training data, when CNN is used, explicit feature extraction is avoided, and learning is implicitly learned from training data; and further, due to neuron weights on the same feature mapping surface The same, so the network can learn in parallel, which is also a big advantage of convolutional networks relative to the network of neurons connected to each other. The convolutional neural network has unique advantages in speech recognition and image processing with its special structure of local weight sharing. Its layout is closer to the actual biological neural network, and weight sharing reduces the complexity of the network, especially multidimensional. The feature that the input vector image can be directly inputted into the network avoids the complexity of data reconstruction during feature extraction and classification.

REFERENCES

1. LeCun, Y., & Bengio, Y. (1998). Convolution Networks for Images, Speech, and Time-Series. Igarss 2014. https://doi.org/10.1007/s13398-014-0173-7.2
2. He, T., Zhang, Z., Zhang, H., Zhang, Z., Xie, J., & Li, M. (n.d.). Bag of Tricks for Image Classification with Convolutional Neural Networks.
3. Cheng, J., Wang, P., Li, G., Hu, Q., & Lu, H. (2018). A Survey on Acceleration of Deep Convolutional Neural Networks. Retrieved from https://arxiv.org/pdf/1802.00939.pdf
4. Benton, T., Staab, J., & Evans, D. L. (2007). Medical Co-Morbidity in Depressive Disorders. Annals of Clinical Psychiatry, 19, 289–303. https://doi.org/10.1080/10401230701653542
5. Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2016). Image-to-Image Translation with Conditional Adversarial Networks. Retrieved from http://arxiv.org/abs/1611.07004
6. Nielsen, M. A. (2015). Neural Networks and Deep Learning. Retrieved from http://neuralnetworksanddeeplearning.com/

7. He, T., Zhang, Z., Zhang, H., Zhang, Z., Xie, J., & Li, M. (n.d.). Bag of Tricks for Image Classification with Convolutional Neural Networks. Retrieved from https://github.com/dmlc/gluon-cv
8. Bishop, C. M. (2006). Pattern Recognition and Machine Learning. In M. Jordan, J. Kleinberg, & B. Schölkopf (Eds.), Pattern Recognition. https://doi.org/10.1117/1.2819119