Figure 5. Interplay of recognition and comparison layers

During training, vectors from the training set are sequentially fed to the input of the network and the network weights are adjusted so that similar vectors activate the same neuron [4].

Thus, the ART network can recognize input vectors that are very similar to the reference images, adjust the reference images for modifications of the input vectors or track slow changes in the reference, as well as increase the number of reference images.

Of course, when modeling ART networks on traditional computers, it is difficult to achieve high performance, in such cases it is especially advantageous to use parallel computers.

1. F. Wasserman. Neurocomputer technology. Theory and Practice, 1992. 184 p.
2. Grossberg S. Competitive learning: From interactive activation to adaptive resonance // Cognitive Science N11/1987, p.23-63
3. Adaptive Resonance Theory. [Elektronnyy resurs]. Rezhim dostupa: <https://www.geeksforgeeks.org/adaptive-resonance-theory-art/>
4. Neural networks and neurocontrollers. M.V. Burakov. Tutorial. St.Pb, 2013. - 284 p.

Mitrofanova A.S., Komlev G.V.
Convolution Networks

*Reshetnev Siberian State University of Science and Technology
 (Russia, Krasnoyarsk)*

doi: 10.18411/lj-11-2019-58

idsp: lj-11-2019-58

Аннотация

Сегодня искусственная нейронная сеть широко используется для решения различных задач в реальной жизни. Такие проблемы, как распознавание речи или оптическое распознавание символов, теперь решают с использованием нейронной сети с высокой точностью. В работе рассматривается одна специальная архитектура нейронной сети – сверточная нейронная сеть, а также её структура и применение для классификации различных типов данных.

Ключевые слова: нейронные сети, сверточная нейронная сеть

Abstract

Today, an artificial neural network is widely used to solve various problems in real life. Problems such as speech recognition or optical character recognition can now be solved using a neural network with high accuracy. The paper considers one special architecture of the neural network - the convolutional neural network, as well as its structure and application for the classification of various types of data.

Keywords: neural networks, convolutional neural network

One of the main directions in pattern recognition is the use of artificial neural networks. And although this approach, like any other, has its weaknesses and strengths, it can definitely be said that to date, the potential of neural networks is far from fully revealed.

In a broad sense, recognition of objects can also be defined as the ability to distinguish between physical properties (for example, shape, color, structure), as a result of which various semantic attributes can be applied to this object, which determine the understanding of the possibilities and previous experience of its use, taking into account relationships with other objects.

Based on this, the recognition process can be divided into four stages:

- 1) Processing the basic properties of the object, such as shape, color, contrast, etc.
- 2) Grouping of the above basic properties on the basis of similarity.
- 3) Description of the visual representation of the object using logical structures in memory.
- 4) Comparison of semantic attributes with visual representation, providing a common sense, and, therefore, recognition of an object.

All the above recognition processes can be resolved using convolutional neural networks (CNN). CN are some of the most popular types of artificial neural networks. So they proved their effectiveness in the recognition of visual images (video and images), recommendation systems and language processing.

CNN differ to other forms of Artificial Neural Network in that instead of focusing on the entirety of the problem domain, knowledge about the specific type of input is exploited. This in turn allows for a much simpler network architecture to be set up.

This paper has outlined the basic concepts of CNN, explaining the layers required to build one and detailing how best to structure the network in most image analysis tasks. Research in the field of image analysis using neural networks has somewhat slowed in recent times. This is partly due to the incorrect belief surrounding the level of complexity and knowledge required to begin modelling these superbly powerful machine learning algorithms.

Recent advances in the architecture of deep-learning CNN have brought the field of image classification and computer vision to a new level. Very deep CNN emerged in 2014 and have achieved new levels of accuracy in several artificial intelligence classification problems.

Current CNN models are able to differentiate the image of a leopard from that of a scooter, but moreover can differentiate images of leopards from their biological cousins – cheetahs and snow leopards. Deep CNN architecture achieved a 3.5-percent top-5 error (how frequent the model fails to predict the correct class as one of the top 5 guesses) and 17.3-percent top-1 error in a visual recognition challenge in 2015. The current benchmark in object category classification and detection consists of hundreds of mixed-object categories and millions of images [1].

CNN are a specific type of neural network that has revolutionized computer vision and pattern recognition.

It is also used for speech recognition, audio processing, time series processing, for analyzing the meaning of texts, and recently even for playing Go, for beating the masters of this game. We can say that at the moment this is the most successful model, the most successful innovation in the framework of what is called deep or deep learning.

The peculiarity of the CNN is that in it the neurons of the first levels are arranged in a special structure, namely: on the first layers, neurons are divided into images of a certain size (they are also sometimes called cards), and different cards inside one layer correspond to different types of neurons, which respond to different features of images. And calculations of the activation of the next layer in CNN are of two main types. In the first type of calculations, the activation of neurons of the next level is calculated as a linear combination of activations of neurons of the previous level, and the weights of these linear activations depend only on

the relative positions of neurons, types of neurons, but do not depend on the position of a given neuron inside the map [2].

The CN is a multilayer perceptron specially created for recognition of two-dimensional surfaces with a high degree of invariance to transformations, scaling, distortions, and other types of deformation. Learning to solve this complex problem is carried out with the teacher, while using networks whose architecture satisfies the following restrictions.

Feature extraction. Each neuron receives an input signal from a local receptive field in the previous layer, thus extracting its local signs. Once a feature has been retrieved, its exact location does not matter, since it is approximately located relative to other features.

Feature mapping each computing layer of the network consists of many feature maps, each of which has the shape of a plane on which all neurons must share the same set of synaptic weights. This form of structural constraint has the following advantages.

Shift invariance implemented by means of feature maps using convolution with a small kernel that performs the «flatten» function.

Reduction in the number of free parameters realized by sharing synoptic scales.

Subsampling each convolution layer is followed by a computational layer that performs local averaging and subsampling. By this, a reduction in resolution for feature maps is achieved. This operation leads to a decrease in the sensitivity of the output signal of the feature display operator, to displacement and other forms of deformation.

This construction of convolution networks has a neurobiological justification, described in innovative works on the locally sensitive and orientation-selective neuron of the cat's visual apparatus [3].

Benefits of the CNN:

- 1) Reducing the number of learning parameters and increasing the speed of learning compared to a fully connected neural network.
- 2) The ability to parallelize computations and implement training algorithms for the network on graphic processors (GPUs).
- 3) Resistance to shifting the position of the object in the input data. During training, the convolutional neural network shifts in parts of the object. Therefore, the traits taught do not depend on the position of the «important parts».

Thus, this property of the convolutional neural network helps to improve the quality of classification. [4].

With all three layers, the convolutional image classifier looks like this:

Figure 1. Convolutional classifier

The convolutional layer (CONV) uses filters that perform convolution operations by scanning the input image. Its hyperparameters include a filter size, which can be 2x2, 3x3, 4x4, 5x5 (but not limited to this) and step S. The result O is called a feature map or an activation map in which all features are calculated using input layers and filters. Below, in fig. 2 shows the generation of feature maps when applying convolution.

Figure 2. CONS convolutional layer

The merge layer (POOL) is used to compact the features commonly used after the convolution layer. There are two types of union operations - this is the maximum and average union, where the maximum and average values of the characteristics are taken, respectively. In fig. 3 illustrates the operation of merge operations.

Figure 3. POOL merge layer

Fully connected layers (FCs) operate with a flat input, where each input is connected to all neurons. Usually they are used at the end of the network to connect hidden layers to the output layer, which helps to optimize class estimates (Fig. 4).

Figure 4. Fully connected FC layers

Figure 5. The principle of processing convolutional networks

It is in this way that convolutional neural networks, layer by layer, transform the original image, starting with the initial pixel values and ending with the definition of the class (Fig. 5). Note that optional layers must contain parameters. In particular, the convolution layer and the fully connected layer perform transformations, which are not only a function of the input activation volume, but also parameters (weight and displacement of neurons). On the other hand, the linear rectification unit and the pooling layer realize a fixed function. The parameters in the CONV and FC layers will be trained using gradient descent, so the class definition of a CNN depends on the labels in the training set for each image.

Eventually:

1. In the simplest case, the CNN architecture is a set of layers that transform an image of an image into an output image (for example, a class definition).
2. Each layer is responsible for a certain stage of the image processing process (the convolution, linear rectification, pooling and fully connected layer are the most popular today).
3. Each layer receives 3D information at the input and transforms it while preserving 3D volume using a differentiable function.
4. The layer may not have parameters (CONV and FC have, RELU and POOL - no).
5. The layer may not have additional hyperparameters (for example, CONV, FC and POOL have, RELU - no) [5].

So, a convolutional neural network can be used to solve various types of problems: image classification, speech recognition, text classification, etc. When using a convolutional neural network, you can significantly reduce the number of trained parameters and get a high quality classification.

Convolutional neural networks allow computer vision to work with simple tasks as well as with complex products and services, from facial recognition to improving medical diagnoses. They may be a key method in computer vision in the future, although a new breakthrough may already be around the corner. One thing is certain: convolutional neural networks are the amazing foundation of many modern innovative applications, and certainly deserve deep study [6].

1. Convolutional Neural Networks. [Elektronnyy resurs]. Rezhim dostupa: <https://explorer.aapg.org/story/articleid/49527/convolutional-neural-networks>.
2. Convolutional Neural Networks. Victor Lempitsky. [Elektronnyy resurs]. Rezhim dostupa: <https://forany.xyz/a-392>.
3. X. Simon. Neural networks: full course, 2006. – 1104 c.
4. Le Mans Ha. Convolutional neural network for solving the classification problem. WORKS MFTI. 2016. Tom 8, no.3, c. 91-97.
5. Convolutional neural networks: an inside look. [Elektronnyy resurs]. Rezhim dostupa: <http://ru.datasides.com/code/cnn-convolutional-neural-networks/>.
6. How does the convolutional neural network work: architecture, examples, features. [Elektronnyy resurs]. Rezhim dostupa: <https://neurohive.io/ru/osnovy-data-science/glubokaya-svertochnaja-nejronnaja-set/>.