

УДК 004.032.26 (004.89)

Igor Olenchenko¹, Yuliya Sobol²

¹PhD student National University “Zaporizhzhia Polytechnic”

²PhD (Philology), associate professor National University “Zaporizhzhia Polytechnic”

CONVOLUTIONAL NEURAL NETWORK FOR IMAGE RECOGNITION

Convolutional neural network(CNN) is a special architecture of neural networks designed for classification and recognition of objects.

The main purpose of classifying images is to take the initial image and output its class or group of probable classes, which best describe the image.

Convolutional neural networks operate on the basis of filters that recognize specific characteristics of an image (for example, straight lines). A filter is a collection of kernels (sometimes a single kernel is used in the filter). Kernel is an ordinary matrix of numbers called weights, which are trained to search for specific characteristics on images. The filter moves along the image and determines whether some desired characteristic is present in a particular part of it. To obtain this response, convolution operation is performed, which is the sum of the products of the filter elements and the input signal matrix, forming the pixel. Convolution is

a method to get an approximate version of the input data; and this contributes to a significant reduction in the number of calculations.

Kernel repeats described operation for each position it occupies, turning the original two-dimensional matrix of characteristics into some other, also two-dimensional, matrix. Obtained characteristics are actually the sums of the "weighted" input characteristics, where elements of the kernel matrix are weights, the initial data - elements of the matrix that are under the kernel at its given position.

If some desired characteristic is present in the image fragment, the output convolution operation will produce a number with a relatively large value. If the characteristic is absent, the output number will be small. The result of moving this filter along the entire image is a matrix consisting of the results of unit convolutions.

There are two basic techniques used to build convolutional layers: padding and stride. In the process of kernel movement, edges of the original matrix are cut off: edge pixels never enter center of the kernel, which is due to the accepted rules of movement. This is not always acceptable, especially when the sizes of the source and resulting matrices should be the same.

This problem is solved using padding, which contains false pixels, the value of which is usually zero (this padding is usually called "zero"). Due to additional space around original matrix, kernel gets opportunity to move around the edge pixels. In this case, the sizes of the initial and resulting matrices are equal.

Quite often, the purpose of a convolutional layer is to reduce the dimension of input signals matrix. This is the basis of convolutional neural networks in which the pulses entering the system are reduced at a speed directly proportional to the number of channels. One method to achieve this result is to use the stride.

The principle of stride is based on the idea that a kernel skips a certain number of steps when it is moving. Stride with a value of 1 indicates that the kernel is shifted by one pixel, which leads us to the standard convolution. Stride with a value of 2, indicates that the kernel is shifted by two pixels. It reduces the size of the resulting matrix by about half compared to the previous solution, etc.

Images usually have three channels (RGB: Red, Green, Blue). Numbers of neural network layers define how many channels can be processed. In this case, filter is a collection of kernels: kernel for each input channel, each kernel from the collection is unique.

Each filter in the convolutional layer generates one output channel. Each kernel that is part of the filter moves along its appropriating input channel, producing a processed version of the signals. Some kernel may have a stronger weight than others, giving preference to specific input channels. Further, the results of each kernel are summarized in order to obtain a single channel. Kernel generates its own version of the channel, which creates resulting output channel.

In the end of this process bias must be applied to the result. In convolutional neural networks, bias is applied to each output filter - this operation is usual addition of bias to each signal from resulting (output) channel. It is necessary to bring out the neural network of the deadlocks that have purely mathematical reasons.

Subsequently, these responses are summarized to a global response. The number of channels from this response are equal to the number of filters. The global response can be passed through a non-linear converter and transferred to the next convolutional layer, where the process repeats.

Despite the fact that convolutional neural networks give the best result in the field of pattern recognition, there are conditions in which they can make banal errors. Also, the use and training of convolutional neural networks involves a large number of computing resources. This type of neural network is demanding teaching methods and training samples. But, in spite of these disadvantages, convolutional neural networks allow the field of computer vision to enter a qualitatively new stage of development. This neural network architecture is used in complex programs, starting with personal identification by the face and ending with medical diagnoses.