run:
ai   Platform   Solutions   Resources   About   Get a Demo

Neural Networks

A Guide

Guide Categories   ⌄

RELATED ARTICLES

Deep Learning for
Computer Vision

Deep Convolutional
Neural Networks

Python Parallel
Processing

PyTorch CNN

PyTorch GAN:
Understanding GAN
and Coding it in
PyTorch

PyTorch ResNet

TensorFlow CNN

What are Deep Convolutional Neural Networks?

Deep learning is a machine learning technique used to build artificial intelligence (AI) systems. It is based on the idea of artificial neural networks (ANN), designed to perform complex analysis of large amounts of data by passing it through multiple layers of neurons.

There is a wide variety of deep neural networks (DNN). Deep convolutional neural networks (CNN or DCNN) are the type most commonly used to identify patterns in images and video. DCNNs have evolved from traditional artificial neural networks, using a three-dimensional neural pattern inspired by the visual cortex of animals.

Deep convolutional neural networks are mainly focused on applications like object detection, image classification, recommendation systems, and are also sometimes used for natural language processing.

In this article, you will learn:

- Deep Convolutional Neural Networks Explained
  - Convolutional Layer
  - ReLU Activation Layer
  - Pooling Layer
  - Fully Connected Layer
- Types of Deep Convolutional Neural Networks
  - R-CNN
  - Fast R-CNN
  - GoogleNet (2014)
  - VGGNet (2014)
  - ResNet (2015)
- Business Applications of Convolutional Neural Networks
  - Image Classification

- Medical Image Analysis
- Optical Character Recognition

This is part of our series of articles about Deep Learning for Computer Vision

Deep Convolutional Neural Networks Explained

The strength of DCNNs is in their layering. A DCNN uses a three-dimensional neural network to process the Red, Green, and Blue elements of the image at the same time. This considerably reduces the number of artificial neurons required to process an image, compared to traditional feed forward neural networks.

Deep convolutional neural networks receive images as an input and use them to train a classifier. The network employs a special mathematical operation called a "convolution" instead of matrix multiplication.

The architecture of a convolutional network typically consists of four types of layers: convolution, pooling, activation, and fully connected.
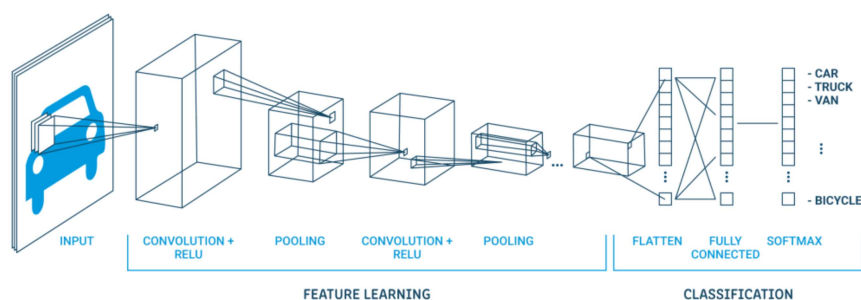
Convolutional Layer

Applies a convolution filter to the image to detect features of the image. Here is how this process works:

- **A convolution**—takes a set of weights and multiplies them with inputs from the neural network.
- **Kernels or filters**—during the multiplication process, a kernel (applied for 2D arrays of weights) or a filter (applied for 3D structures) passes over an image multiple times. To cover the entire image, the filter is applied from right to left and from top to bottom.
- **Dot or scalar product**—a mathematical process performed during the convolution. Each filter multiplies the weights with different input values. The total inputs are summed, providing a unique value for each filter position.

3_0	3_1	2_2	1	0
0_2	0_2	1_0	3	1
3_0	1_1	2_2	2	3
2	0	0	2	2
2	0	0	0	1

12.0	12.0	17.0
10.0	17.0	19.0
9.0	6.0	14.0
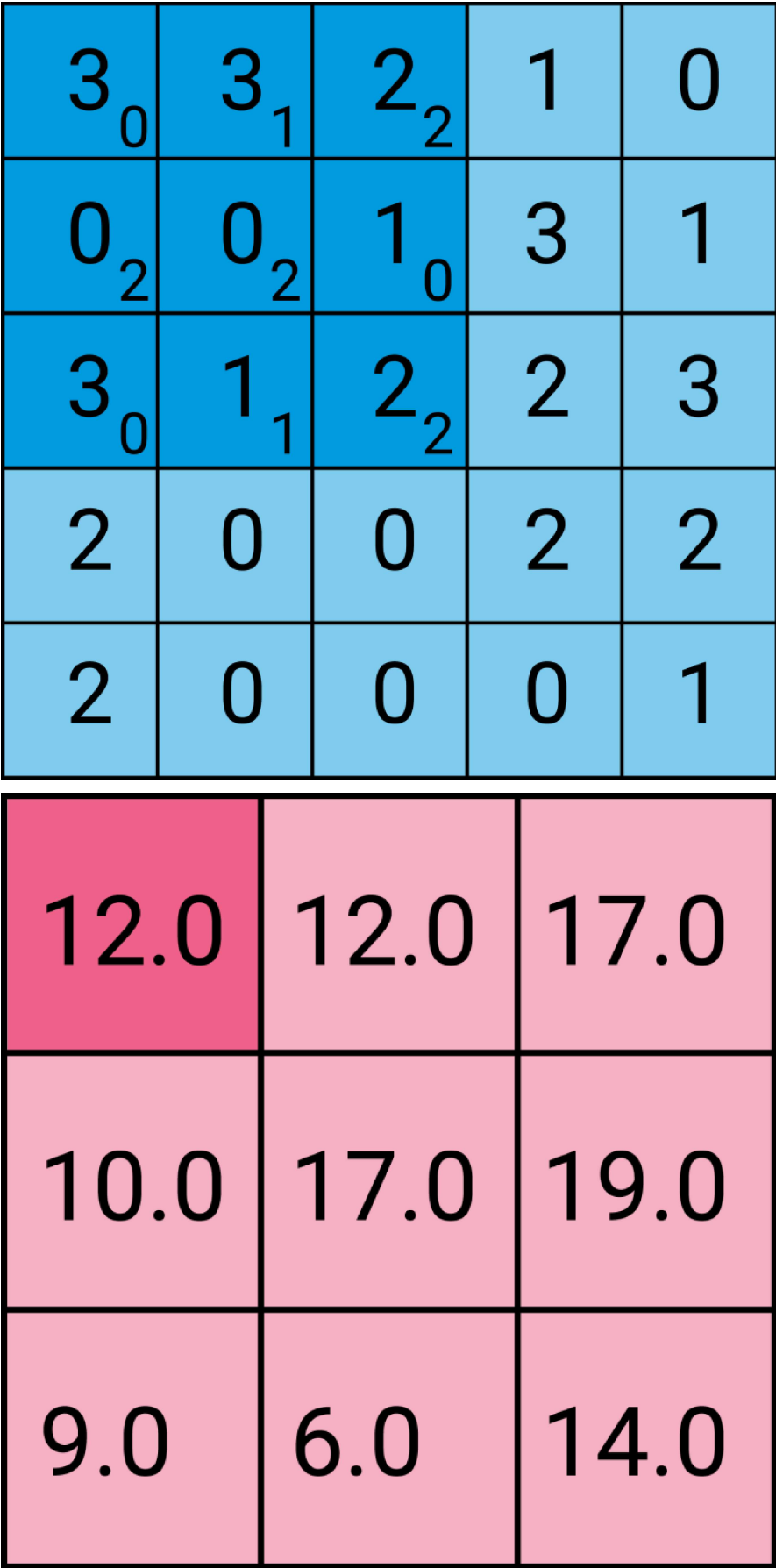
ReLU Activation Layer

The convolution maps are passed through a nonlinear activation layer, such as Rectified Linear Unit (ReLu), which replaces negative numbers of the filtered images with zeros.

Pooling Layer

The pooling layers gradually reduce the size of the image, keeping only the most important information. For example, for each group of 4 pixels, the pixel having the maximum value is retained (this is called max pooling), or only the average is retained (average pooling).

Pooling layers help control overfitting by reducing the number of calculations and parameters in the network.

After several iterations of convolution and pooling layers (in some deep convolutional neural network architectures this may happen thousands of times), at the end of the network there is a traditional multi layer perceptron or "fully connected" neural network.

Fully Connected Layer

In many CNN architectures, there are multiple fully connected layers, with activation and pooling layers in between them. Fully connected layers receive an input vector containing the flattened pixels of the image, which have been filtered, corrected and reduced by convolution and pooling layers. The softmax function is applied at the end to the outputs of the fully connected layers, giving the probability of a class the image belongs to – for example, is it a car, a boat or an airplane.

Related content: read our guide to deep learning for computer vision.

What are the Types of Deep Convolutional Neural Networks?

Below are five deep convolutional neural network architectures commonly used to perform object detection and image classification.

R-CNN

Region-based Convolutional Neural Network (R-CNN), is a network capable of accurately extracting objects to be identified in the image. However, it is very slow in the scanning phase and in the identification of regions.

The poor performance of this architecture is due to its use of the selective search algorithm, which extracts approximately 2000 regions of the starting image. Afterwards it executes N CNNs on top of each region, whose outputs are fed to a support vector machine (SVM) to classify the region.

Fast R-CNN

Fast R-CNN is a simplified R-CNN architecture, which can also identify regions of interest in an image but runs a lot faster. It improves performance by extracting features before it identifies regions of interest. It uses only one CNN for the entire image, instead of 2000 CNN networks on each superimposed region. Instead of the SVM which is computationally intensive, a softmax function returns the identification probability. The downside is that Fast R-CNN has lower accuracy than R-CNN in terms recognition of the bounding boxes of objects in the image.

GoogleNet (2014)

GoogleNet, also called Inception v1, is a large-scale CNN architecture which won the ImageNet Challenge in 2014. It achieved an error rate of less than 7%, close to the level of human performance. The architecture consists of a

22-layer deep CNN based on small convolutions, called "inceptions", batch normalization, and other techniques to decrease the number of parameters from tens of millions in previous architectures to four million.

VGGNet (2014)

A deep convolutional neural network architecture with 16 convolutional layers. It uses 3x3 convolutions, and trained on 4 GPUs for more than two weeks to achieve its performance. The downside of VGGNet is that unlike GoogleNet, it has 138 million parameters, making it difficult to run in the inference stage.

ResNet (2015)

The Residual Neural Network (ResNet) is a CNN with up to 152 layers. ResNet uses "gated units", to skip some convolutional layers. Like GoogleNet, it uses heavy batch normalization. ResNet uses an innovative design which lets it run many more convolutional layers without increasing complexity. It participated in the ImageNet Challenge 2015, achieving an impressive error rate of 3.57%, while beating human-level performance on the trained dataset.

Business Applications of Convolutional Neural Networks

Image Classification

Deep convolutional neural networks are the state of the art mechanism for classifying images. For example, they are used to:

- **Tag images**—an image tag is a word or combination of words that describes an image and makes it easier to find. Google, Facebook and Amazon use this technology. Labeling includes identifying objects and even analyzing the sentiment of the image.

- **Visual search**—matching the input image with an available database. Visual search analyzes the image and searches for an existing image with the identified information. For example, Google search uses this technique to find different sizes or colors of the same product.

- **Recommendation engines**—using CNN image recognition to provide product recommendations, for example in websites like Amazon. The engine analyzes user preferences and returns products whose images match previous products they viewed or bought, for example, a red dress or red shoes with red lipstick.

Medical Image Analysis

CNN classification on medical images is more accurate than the human eye and can detect abnormalities in X-ray or MRI images. Such systems can analyze sequences of images (for examples, tests taken over a long period of time) and identify subtle differences that human analysts might miss. This also makes it possible to perform predictive analysis.

Classification models for medical images are trained on large public health databases. The resulting models can be used on patient test results, to identify medical conditions and automatically generate a prognosis.

Optical Character Recognition

Optical character recognition (OCR) is used to identify symbols such as text or numbers in images. Traditionally OCR was performed using statistical or early machine learning techniques, but today many OCR engines use deep convolutional neural networks.

OCR powered by CNNs can be used to improve search within rich media content, and identify text in written documents, even those with poor quality or hard to recognize handwriting. This is especially important in the banking and insurance industries. Another application of deep learning OCR is for automated signature recognition.

Deep Convolutional Neural Networks with Run:AI

Run:AI automates resource management and orchestration for machine learning infrastructure. With Run:AI, you can automatically run as many compute intensive experiments as needed.

Here are some of the capabilities you gain when using Run:AI:

- **Advanced visibility**—create an efficient pipeline of resource sharing by pooling GPU compute resources.
- **No more bottlenecks**—you can set up guaranteed quotas of GPU resources, to avoid bottlenecks and optimize billing.
- **A higher level of control**—Run:AI enables you to dynamically change resource allocation, ensuring each job gets the resources it needs at any given time.

Run:AI simplifies machine learning infrastructure pipelines, helping data scientists accelerate their productivity and the quality of their models.

Learn more about the Run:AI GPU virtualization platform.

Learn More About Deep Learning for Computer Vision

Read more in our series of guides about deep learning for computer vision.