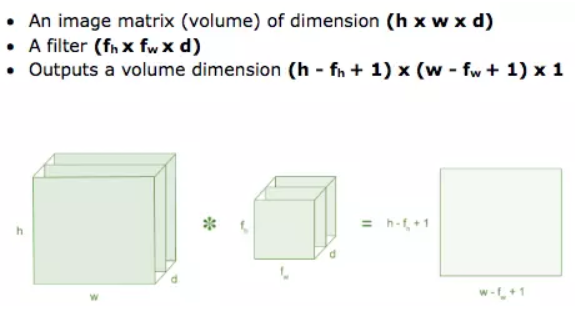
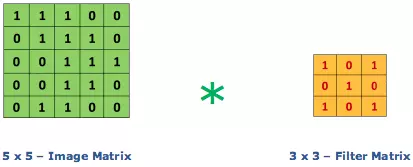
**Convolutional Neural Networks**

**521H0489 – Hồ Hữu An**

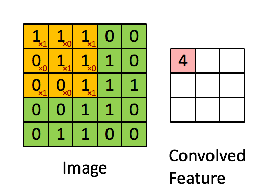
**1. The basics of the Convolutional Neural Networks model**

*Convolution Layer:*

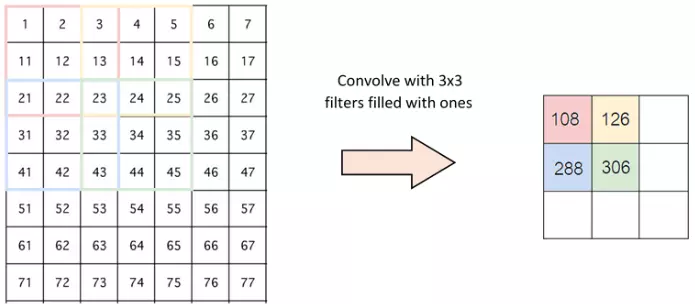
Convolution is the first layer to extract features from the input image. Convolution maintains the relationship between pixels by learning image features using small squares of input data. It is a mathematical operation that takes two inputs such as an image matrix and a filter or kernel.

Consider a 5 x 5 matrix with pixel values ​​of 0 and 1. A 3 x 3 filter matrix as shown below.

Then the convolution layer of the 5 x 5 image matrix is ​​multiplied by the 3 x 3 filter matrix called 'Feature Map' as shown below

The combination of 1 image with different filters can perform operations such as edge detection, blurring, and sharpening by applying the filters. The example below shows different convolution images after applying different Kernels.

*Stride:*

Stride is the number of pixels that change on the input matrix. When stride is 1, we move the kernels by 1 pixel. When stride is 2, we move the kernels by 2 pixels and so on. The image below is the convolution layer operating with a stride of 2.

*Padding:*

Sometimes the kernel doesn't match the input image. We have 2 options:

+ Insert additional zeros into the 4 borders of the image (padding).

+ Crop the image at points that are inconsistent with the kernel.

*ReLU:*

ReLU stands for Rectified Linear Unit, which is a nonlinear function. With the output being: ƒ (x) = max (0, x).

Why ReLU is important: ReLU introduces nonlinearity in ConvNet. Because the data in the world we learn about are non-negative linear values.

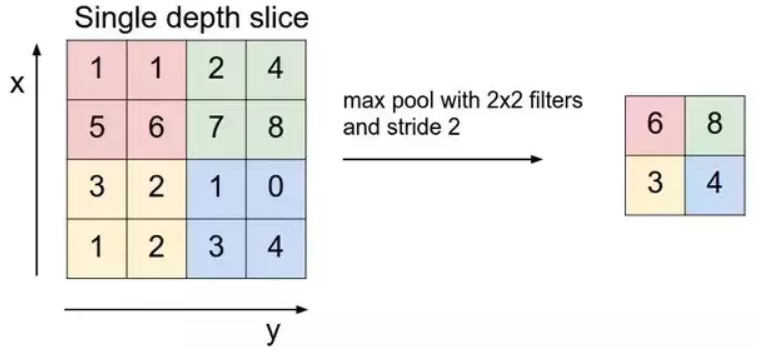
A diagram of a function

Description automatically generated

There are some other nonlinear functions such as tanh, sigmoid that can also be used instead of ReLU. Most people often use ReLU because it has good performance.

*Pooling Layer:*

* The pooling layer will reduce the number of parameters when the image is too large. Spatial pooling also known as subsampling or downsampling reduces the size of each map but still retains important information. Pooling can be of different types:
* Max Pooling
* Average Pooling
* Sum Pooling
* Max pooling takes the largest element from the object matrix, or takes the average sum. The sum of all elements in the map is called sum pooling

****

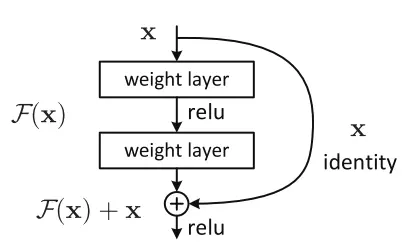
**2. The architecture of the Residual Networks model and its variations.**

*Why did the ResNet network appear?*

The ResNet(R) network is a CNN network designed to work with hundreds of layers. A problem that occurs when building a CNN network with many convolutional layers is the Vanishing Gradient phenomenon, leading to a bad learning process.

*ResNet network architecture*

The solution that ResNet offers is to use uniform "off" connections to penetrate one or more layers. Such a block is called a Residual Block, as shown in the following figure:



ResNet is almost similar to networks including convolution, pooling, activation and fully-connected layers. The image above shows the residual blocks used in the network. A curved arrow appears starting from the beginning and ending at the end of the residual block. In other words, it will add Input x) is the predicted value, F(x) is the true value (label), we want H(x) to be equal to or approximately F(x). F(x) is obtained from x as follows:

X 🡪 weight1 🡪 ReLU 🡪 weight2

The value H(x) is obtained by:

F(x) + x 🡪 ReLU

Stacking layers will not reduce network performance. We can simply overlay identical mappings onto the existing network and the performance of the architecture remains unchanged. This makes deep architectures at least no worse than shallow architectures. Furthermore, with this architecture, the layers above get more direct information from the layers below, so they adjust the weights more effectively.

A diagram of a diagram

Description automatically generatedA diagram of different types of blocks

Description automatically generated"ID BLOCK" in the image above stands for Identity block and ID BLOCK x3 means there are 3 Identity blocks overlapping each other. The content of the image above is as follows:

Zero-padding : Input with (3,3)

Stage 1: Convolution (Conv1) with 64 filters with shape(7,7), using stride(2,2). BatchNorm, MaxPooling (3,3).

Stage 2: Convolutional block uses 3 filters with size 64x64x256, f=3, s=1. There are 2 Identity blocks with filter size 64x64x256, f=3.

Stage 3: Convolutional uses 3 filter sizes 128x128x512, f=3,s=2. There are 3 Identity blocks with filter size 128x128x512, f=3.

Stage 4: Convolutional uses 3 filter sizes 256x256x1024, f=3,s=2. There are 5 Identity blocks with filter size 256x256x1024, f=3.

Stage 5 :Convolutional uses 3 filter sizes 512x512x2048, f=3,s=2. There are 2 Identity blocks with filter size 512x512x2048, f=3.

The 2D Average Pooling : used with size (2,2).

The Flatten.

Fully Connected (Dense): uses softmax activation.

*Variant of resnet*

ResNet-50, ResNet-101 and ResNet-152: These are the original versions of ResNet with different numbers of convolutional layers. ResNet-50 has 50 layers, ResNet-101 has 101 layers and ResNet-152 has 152 layers. These versions are commonly used for image recognition and classification tasks.

ResNeXt: ResNeXt is a variant of ResNet with the idea of ​​expanding the information space in each block. Instead of using a set of convolutional layers, ResNeXt uses modules called "cardinality" to expand the information space. This extension improves the network's learning ability and achieves better results on multiple data sets.

Wide ResNet: Wide ResNet increases the number of convolution channels in each layer. Instead of using 64 or 128 channels as in the original versions, Wide ResNet uses 256 or 512 channels. This helps the network learn more complex features and have better representation capabilities.

Pre-activated ResNet (ResNet with pre-activation): This is a variation of ResNet in which the building blocks are executed in a different order. Instead of applying batch normalization and activating ReLU before executing the convolutional layer, pre-activated ResNet performs batch normalization and activating ReLU after executing the convolutional layer. This architecture has shown better performance in several tasks.