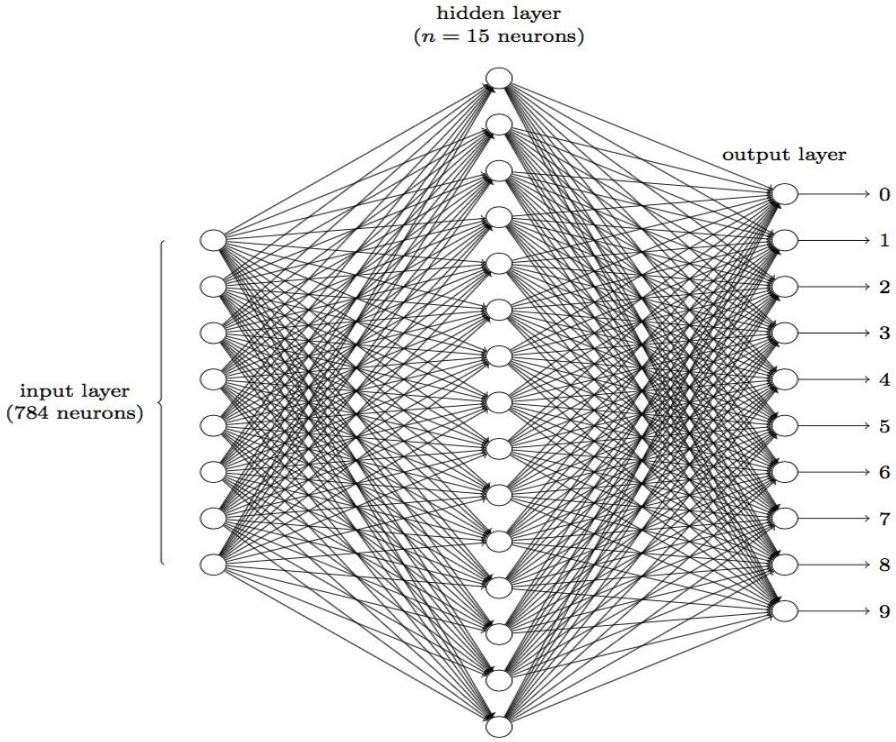
Convolutional Neural Networks, CNN

1. Why do we use neural networks?
    - High efficiency of feature extraction: To ensure the accuracy of classification, we need to define features. Each feature is a dimension, and if the number of features is too small, we may not be able to accurately classify it, that is, we call it underfitting. If the number of features is too large, it may cause us to pay too much attention to a certain feature in the classification process, leading to classification error, namely overfitting. However, the emergence of neural network makes us do not need to do a lot of feature engineering, such as designing the content of features in advance or the number of features, etc., we can directly put the data into it, let it train itself, self-correction, can get a better effect.
    - Simplicity of data format：In a traditional machine learning classification problem, we "fill" in data cannot be directly into, need to do some processing data, such as dimensional normalization, format conversion, and so on, but we don't need extra in the neural network to do too much processing data, specific reasons we can see the back of the is deduced in detail.
2. Why we use convolutional neural networks?

Because of the shortcoming of traditional neural networks. As we know, the image is composed of pixels, and each pixel has three channels, respectively representing RGB color. Then, if the size of an image is (28, 28, 1), it means that the image is represented by an image with a length and width of 28 and a channel of 1 (channel is also called depth, where 1 represents gray image).If use full connection network structure, namely, the neural network with the adjacent layers on each neuron is connected, it means that our network have 28 * 28 = 784 neurons, adopted 15 hidden layer neurons, so simple to calculate, we need the number of parameters (w and b) a: 784 * 15 * 15 + 10 = 10 + 117625, this parameter is too much, just a back propagation is a huge amount of calculation, from the Angle of the computing resources and adjustable parameter is not in the traditional neural network. (Charlotte77, 2017)

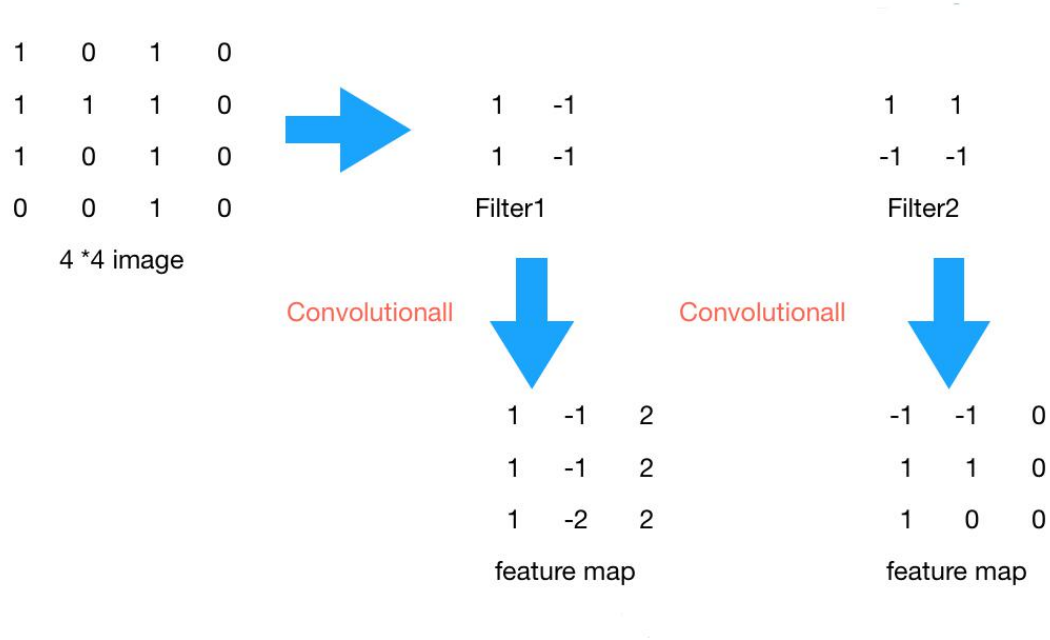


3. What’ s convolutional neural networks, CNN?

   3.1. Three basic layers:

      3.1.1. **Convolutional Layer:** As we know, images are inherently local. For example, if we look at a picture of a cat, we may know that it is a picture of a cat by looking at its glasses or mouth, but we do not need to say that we have seen all the parts to know, ah, this is a cat. So, if we can somehow identify a typical feature of an image, then we know the category of the image. And that's where the idea of convolution comes

      in.



Picture above shows a simple example of convolution operation. There is no padding, the stride is 1, original picture is 4*4 and filter is 2*2. In addition, we use the feature map size calculation formula: $w_{out} = \frac{w_{in} + 2 * padding - F}{stride} + 1$ so we get the final size which is 3*3. After that, we add a bias and use the activation function Relu. So, the convolution process is complete. From the above calculation, we can see that neurons in the same layer can share the convolution kernel, so the processing of high order data will be very simple. Moreover, the size of the image after using the convolution kernel becomes smaller, which is convenient for subsequent calculation.

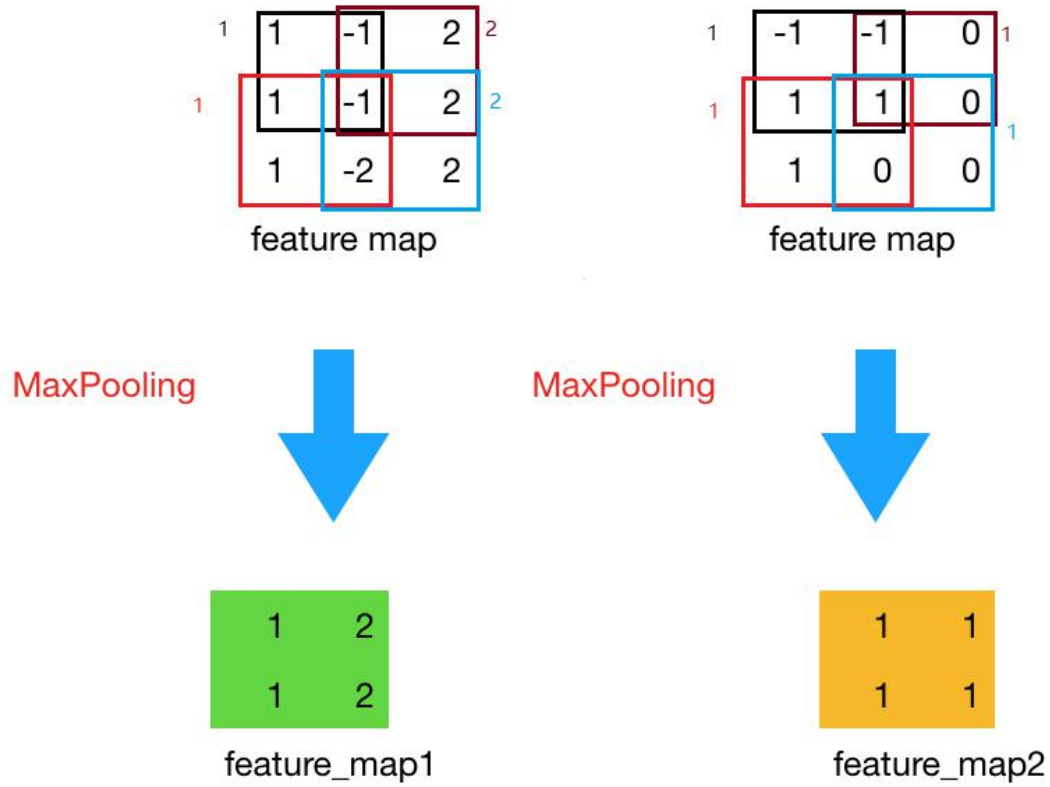
3.1.2. Why CNN is better than traditional neural networks?

Let's take a closer look at the calculation above. The feature map calculated by the first convolution kernel is a 3d data, and the absolute value of the third column is the largest, which indicates that there is a vertical feature in the corresponding place of the original image, that is, the pixel value changes greatly. The two convolution kernels can respectively extract or detect specific features of the original image. At this time, we can understand the convolution kernel as a feature extractor. We only need to input the image data, and design the size, number and sliding step of the convolution kernel to automatically extract some features of the image, to achieve the classification effect.

3.2. **Pooling Layer:** After the operation of the convolution kernel of the previous layer 2*2, we changed the size of the original image from 4*4 to a new image of 3*3. The main purpose of the pooling layer is to compress the image and reduce the parameters without affecting the image quality by reducing sampling. To put it simply, suppose that the size of the pool layer is 2*2 and the step size is 1. The maximum value of each window is taken and then the size of the picture will be changed from 3*3 to 2*2 :(3-2) +1=2.From the above example,

there will be the following

transformation:



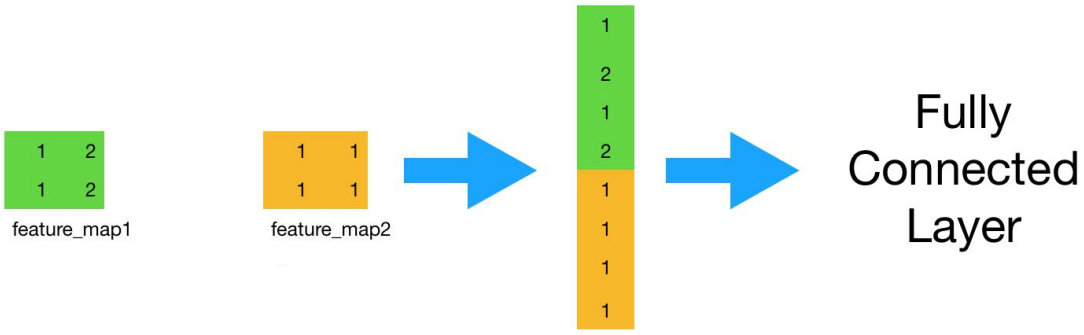
3.2.1. There are two methods of pooling:

- MaxPooling: take the maximum value in the sliding window.
- AveragePooling: take the average value of all values in the sliding window.

3.3. **Flatten Layer & Fully Connected Layer:** At this point, we have completed a complete "convolution part". If we want to superimpose the number of layers, it is usually superimposed with "conv-maxpooing". By constantly designing the size and

number of convolution cores, we can extract more features and finally identify different types of objects. After Max Pooling, the data will be "Flatten" and dropped into the Flatten Layer. Then the output of the Flatten Layer will be placed in the full connected Layer and softmax will be used to classify it. Usually, CNN network will relate to several fully connected layers after the convolutional layer, and the feature map generated by the convolutional layer will be mapped into an eigenvector of fixed

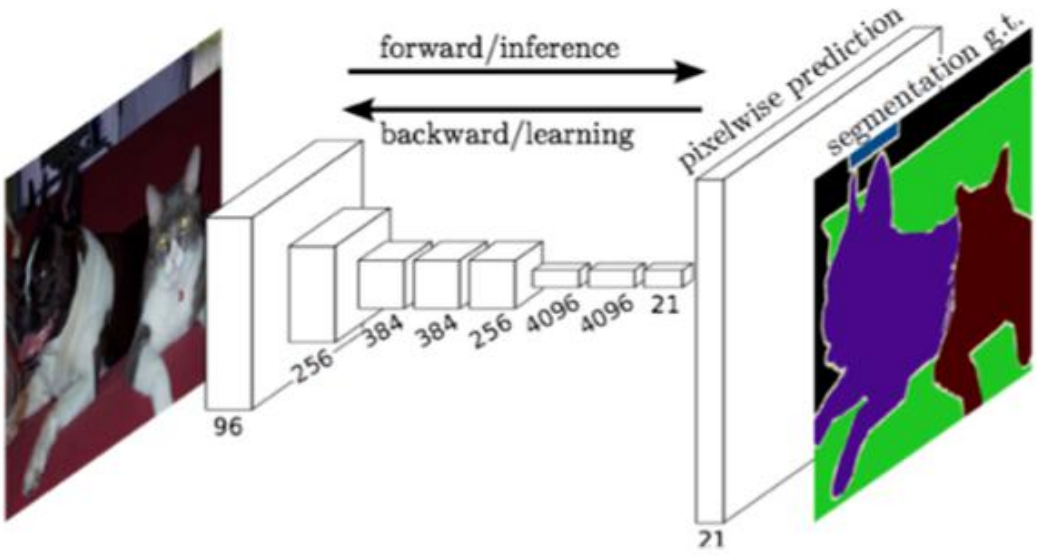
length.



4. What' s fully convolutional networks?

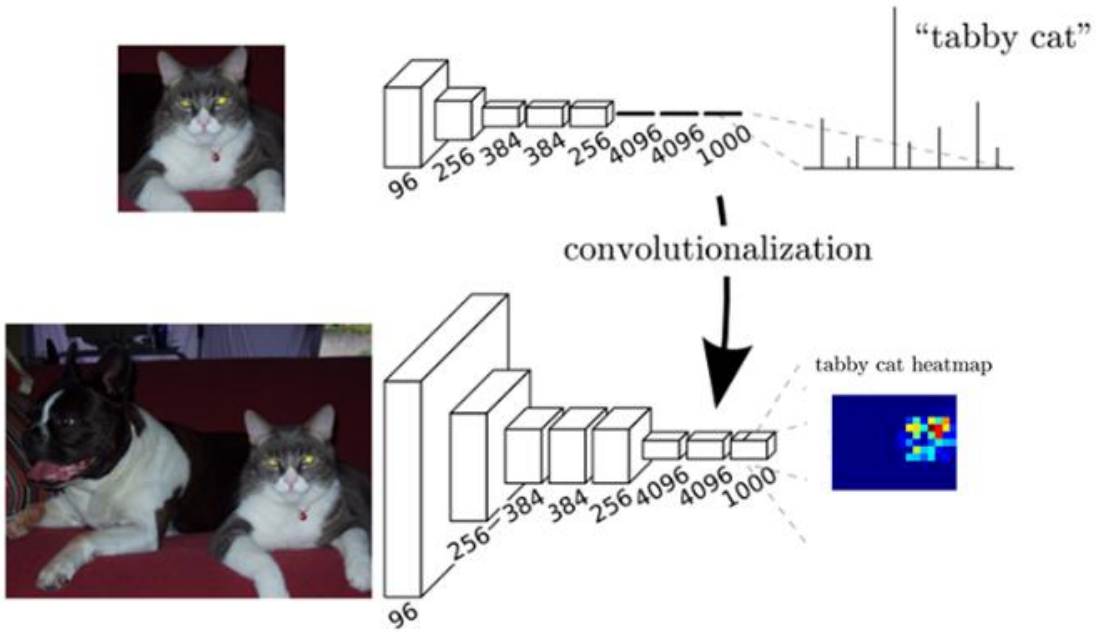
With classic CNN in convolution using all connections get fixed length layer feature vector to different classification, FCN can accept any size of the input image, the deconvolution of last convolution layer on the feature map of sampling, make it back to the input image of the same size, which can produces a forecast for each pixel, while retaining the original input space information of images, the last in the sampling

per-pixel classification on the characteristics of the diagram.



Finally, the loss of softmax classification is calculated pixel by pixel, which is equivalent to one training sample for each pixel.

(SunnyFish-ty, 2018)



Rreferences

Charlotte77. (2017, 11 1). *Basic principles of CNN convolutional neural network*. Retrieved from cnblogs: https://www.cnblogs.com/charlotte77/p/7759802.html

SunnyFish-ty. (2018, 6 5). *The difference between full convolution network FCN and convolutional neural network CNN*. Retrieved from blog.csdn: https://blog.csdn.net/qq_33485434/article/details/80580415