Lecture 3. Convolutional Neural Network

Young, Shen

August 11, 2024

1 Image Identification

Considering image identification, which can be seen as a simple classification problem. With the input being a flatten vector containing the RGB tensor of the picture, and the output being a One-Hot vector indicating what is in this picture. This means the lenth of the output vector shows the number of max classes the network can distinguish input into. With this said, we can apply our network that does the classification job to solving the problem.

However, sometimes the relation between input layer and hidden layers can be very complex when dealing with an RGB picture. For example, to read a picture with 100*100 pixels, there are 30,000(100*100*3) neurons in input layer. Supposing we have 1,000 neurons in our first hidden layer, the weights between them is 30,000,000 with 1,000 bias, Which is a very large number as parameter, and this is just a small picture with 100*100 pixels. To deal with this problem, we need **Convolutional Neural Network (CNN).**

2 Convolutional Neural Network

Considering the process when human being identify images:

1. We look for features. When we see beak, craw, wings appear in the image, we can be 90% certain that there is a bird in the image.
2. We don't actually mind where in the image the feature appears: beaks can appear at the corner or in the middle or anywhere else,
3. If we take the pixels of the image at certain intervals, it nearly doesn't influence us identify the image.

Based on these methods when human being identify images, we can make some simplification to the network:

1. The neurons don't have to be fully connected. Each neuron next layer has a **receptive field** of previous layer.

   Note: For a certain receptive field, there may be multiple neurons that is receptive.
2. Neurons have different receptive fields may **Share Parameters**, so that a feature can be detected all over the image. The shared parameter sets are also called **Filters**. A filter can be understood as a finder of a certain feature.
3. **Pooling layers** can be used to reduce some insignificant information of the image, namely, sum up some zones of the layer.

   Note: pooling is optional, its mainly use is to reduce computing complexity. ALphaGO uses CNN but it doesn't use pooling!

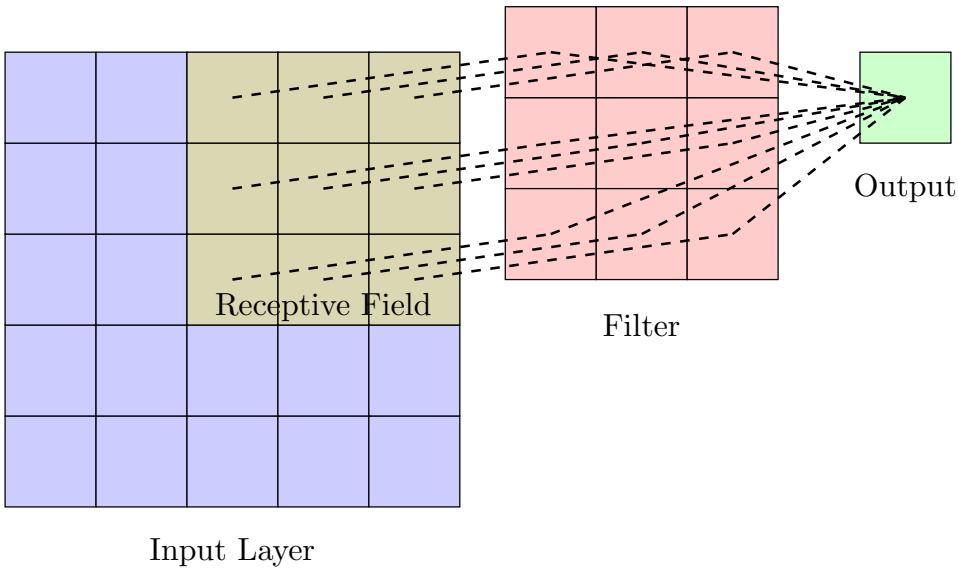

Fig. 1. The Function of a Filter

The 1 and 2 points are altogether called **Convolutional Layer**. The function of a filter can be shwon below:

The figure(1) only shows **ONE** step of **ONE** filter. During convolution, the filter goes over the whole image, with **Stride** as a hyperparameter. When the filter moves over the image, a 2-D matrix is generated accordingly. When the receptive field goes out of the image, **Padding** is then used, namely, adding numbers to the undefined area. Also, there are usually more than one filters. The matrix they generated are then piled together, making a 3-D tensor. Each layer of the tensor is called **Channel**, each channel contains information of the image based on corresponding filter.

The behavior of filters is consistent with traditional neural network with 1,2 simplification mentioned before.

- The movement of each filter shows the idea of **Parameter Sharing**
- Filters themselves shows the idea of **receptive field**,
- That there are usually more than one filters moving all over the image shows "For a certain receptive field, there may be multiple neurons that is receptive".
- The matrix-like output can be transferred into an array by flattening them.

The whole structure of CNN can be shown as below:

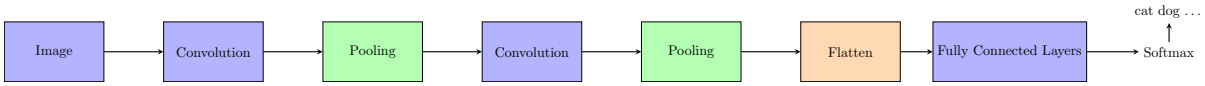

Fig. 2. The Structure of a CNN

3 When to use CNN

CNN is mainly used for image identification. Meanwhile, it can be used in other fields that have similar features with image identification. Generally speaking, CNN can be used to process data with **grid structure**. For example, audio processing, natural language process(NLP), video

analysis, medical image analysis and more. When using CNN, carefully consider whether pooling is needed(remember it is not used in AlphaGO!!!). Moreover, when doing image identification, CNNs can't deal with:

- rotated images
- zoomed images

To deal with this, we need **Spatial Transformation Network (STN)**[1].

References

[1] Max Jaderberg et al. *Spatial Transformer Networks*. 2016. arXiv: 1506.02025 [cs.CV]. URL: https://arxiv.org/abs/1506.02025.