CS172 Computer Vision I

Quiz 2

Reminders: 1) This is a **closed-book** quiz. 2) Please answer in **English**.

Name (in Chinese):

Student ID:

1. When training a model, we sometimes find that as the number of training iterations increases, the model's performance on the training set continues to improve, but its performance on the test set starts to decrease. This phenomenon is known as overfitting. Please describe the causes of overfitting and provide at least three methods to solve this problem.

2. Suppose we have a batch of input images and we would like to design a 2D convolutional neural network to recognize the image patterns. Assume the size of input images are $32 \times 224 \times 224 \times 3$, where $32, 224, 224$ and 3 represent the batch size, height, width, and channel number of the images.

(a) Let's consider the first convolutional layer. Assume we use 5 convolution filters of size $3 \times 3 \times 3$, with stride 1 and zero padding 1.
   - What is the size of the output feature map?
   - What is the number of parameters in the filters? (Note: make sure to consider the bias.)

(b) Let's move on to the pooling layers, which are often placed after convolution filters in practice.

- What is the purpose of designing pooling layers? Explain it briefly.
- Assume the input size of a feature map is $32 \times 128 \times 128 \times 64$, and we apply 2×2 filters with stride 2 for max pooling, what is the size of the output feature map? And what is the number of parameters in max pooling?

3. LSTM (Long Short-Term Memory) is a type of deep learning model specifically designed for handling sequence data. Unlike traditional RNNs (Recurrent Neural Networks), LSTMs incorporate three distinct gates and one cell state to effectively manage long-term dependencies within the sequence. Please sketch a basic unit of an LSTM and explain the functions of the three gates and the cell state.