1. What is the COVARIATE SHIFT Issue, and how does it affect you?

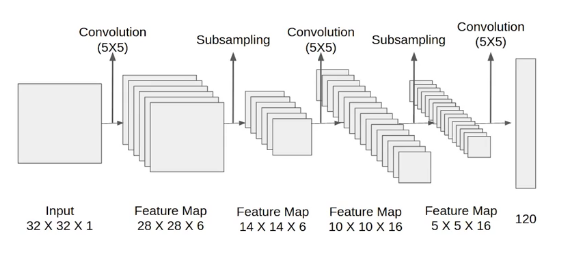
**Ans : Covariate shift occurs when the distribution of variables in the training data is different to real-world or testing data. This means that the model may make the wrong predictions once it is deployed, and its accuracy will be significantly lower.**

2. What is the process of BATCH NORMALIZATION?

**Ans : Batch normalization is a technique for training very deep neural networks that standardizes the inputs to a layer for each mini-batch. This has the effect of stabilizing the learning process and dramatically reducing the number of training epochs required to train deep networks.**

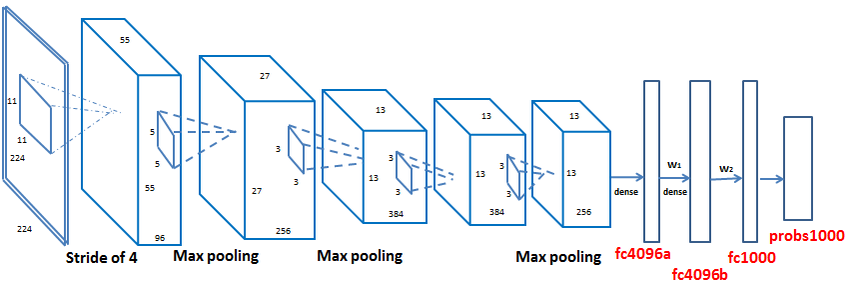
3. Using our own terms and diagrams, explain LENET ARCHITECTURE.

**Ans : LeNet is a convolutional neural network structure proposed by Yann LeCun. onvolutional neural networks are a kind of feed-forward neural network whose artificial neurons can respond to a part of the surrounding cells in the coverage range and perform well in large-scale image processing.**



4. Using our own terms and diagrams, explain ALEXNET ARCHITECTURE.

**Ans : AlexNet architecture consists of 5 convolutional layers, 3 max-pooling layers, 2 normalization layers, 2 fully connected layers, and 1 softmax layer. 2. Each convolutional layer consists of convolutional filters and a nonlinear activation function ReLU. 3. The pooling layers are used to perform max pooling.**



5. Describe the vanishing gradient problem.

**Ans : In machine learning, the vanishing gradient problem is encountered when training artificial neural networks with gradient-based learning methods and backpropagation. The problem is that in some cases, the gradient will be vanishingly small, effectively preventing the weight from changing its value.**

6. What is NORMALIZATION OF LOCAL RESPONSE?

**Ans : Local Response Normalization (LRN) was first introduced in AlexNet architecture where the activation function used was ReLU as opposed to the more common tanh and sigmoid at that time. Apart from the reason mentioned above, the reason for using LRN was to encourage lateral inhibition.**

7. In AlexNet, what WEIGHT REGULARIZATION was used?

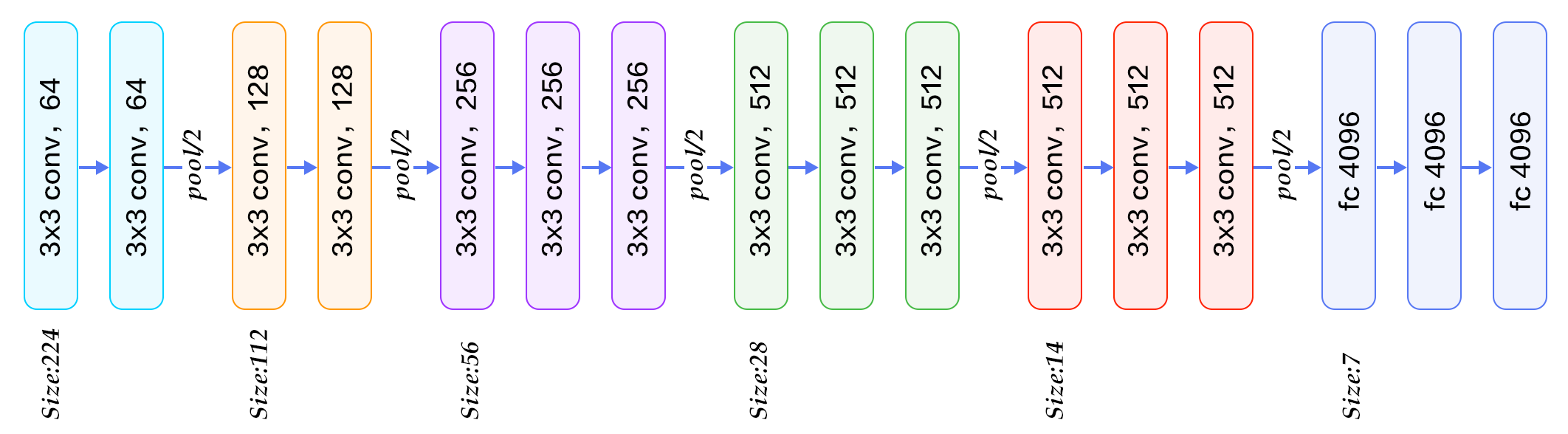
**Ans : Regularization refers to the act of modifying a learning algorithm to favor “simpler” prediction rules to avoid overfitting. Most commonly, regularization refers to modifying the loss function to penalize certain values of the weights you are learning. Specifically, penalize weights that are large.**

8. Using our own terms and diagrams, explain VGGNET ARCHITECTURE.

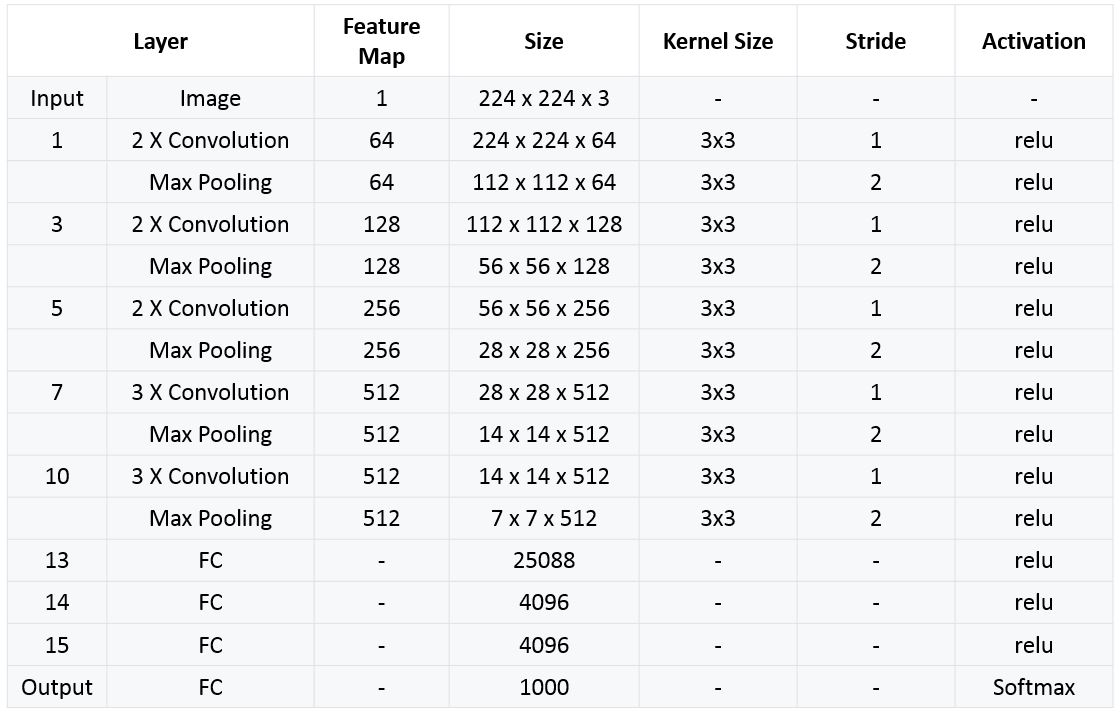
## Ans : The network structure

* **The input of VGG is set to an RGB image of 224x244 size. The average RGB value is calculated for all images on the training set image, and then the image is input as an input to the VGG convolution network. A 3x3 or 1x1 filter is used, and the convolution step is fixed. . There are 3 VGG fully connected layers, which can vary from VGG11 to VGG19 according to the total number of convolutional layers + fully connected layers. The minimum VGG11 has 8 convolutional layers and 3 fully connected layers. The maximum VGG19 has 16 convolutional layers. +3 fully connected layers. In addition, the VGG network is not followed by a pooling layer behind each convolutional layer, or a total of 5 pooling layers distributed under different convolutional layers. The following figure is VGG Structure diagram:**

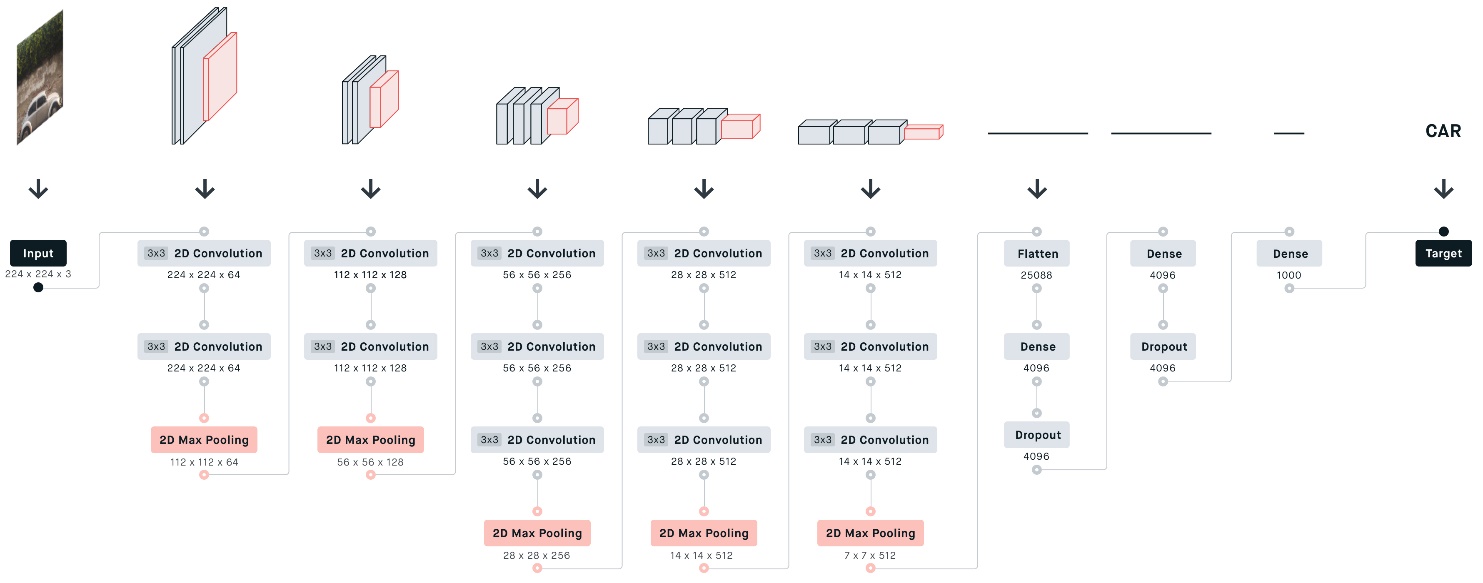
****

****

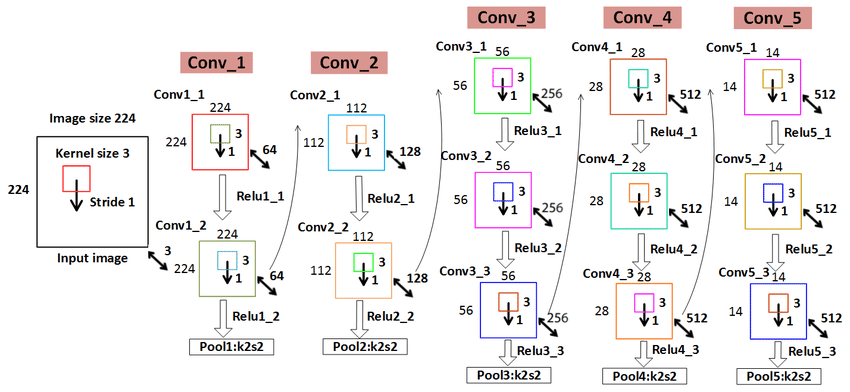
* **VGG16 contains 16 layers and VGG19 contains 19 layers. A series of VGGs are exactly the same in the last three fully connected layers. The overall structure includes 5 sets of convolutional layers, followed by a MaxPool. The difference is that more and more cascaded convolutional layers are included in the five sets of convolutional layers .**

****

* **Each convolutional layer in AlexNet contains only one convolution, and the size of the convolution kernel is 7 *7 ,. In VGGNet, each convolution layer contains 2 to 4 convolution operations. The size of the convolution kernel is 3*3, the convolution step size is 1, the pooling kernel is 2 \* 2, and the step size is 2. The most obvious improvement of VGGNet is to reduce the size of the convolution kernel and increase the number of convolution layers.**

****

* **Using multiple convolution layers with smaller convolution kernels instead of a larger convolution layer with convolution kernels can reduce parameters on the one hand, and the author believes that it is equivalent to more non-linear mapping, which increases the Fit expression ability.**

****

* **Two consecutive 3 *3 convolutions are equivalent to a 5*5 receptive field, and three are equivalent to 7 *7. The advantages of using three 3*3 convolutions instead of one 7 *7 convolution are twofold : one, including three ReLu layers instead of one , makes the decision function more discriminative; and two, reducing parameters . For example, the input and output are all C channels. 3 convolutional layers using 3*3 require 3 (3 *3*C *C) = 27*C *C, and 1 convolutional layer using 7*7 requires 7 *7*C *C = 49C*C. This can be seen as applying a kind of regularization to the 7 *7 convolution, so that it is decomposed into three 3*3 convolutions.**
* **The 1 *1 convolution layer is mainly to increase the non-linearity of the decision function without affecting the receptive field of the convolution layer. Although the 1*1 convolution operation is linear, ReLu adds non-linearity.**

9. Describe VGGNET CONFIGURATIONS.

**Ans : Network A: First mention a shallow network, this network can easily converge on ImageNet. And then?**

* **Network A-LRN: Add something that someone else (AlexNet) has experimented to say is effective (LRN), but it seems useless. And then?**
* **Network B: Then try adding 2 layers? Seems to be effective. And then?**
* **Network C: Add two more layers of 1 1 convolution, and it will definitely converge. The effect seems to be better. A little excited. And then?**
* **Network D: Change the 1 1 convolution kernel to 3 \* 3. Try it. The effect has improved again. Seems to be the best.**

10. What regularization methods are used in VGGNET to prevent overfitting?

**Ans : Use Dropouts. Dropout is a regularization technique that prevents neural networks from overfitting. Regularization methods like L1 and L2 reduce overfitting by modifying the cost function.**