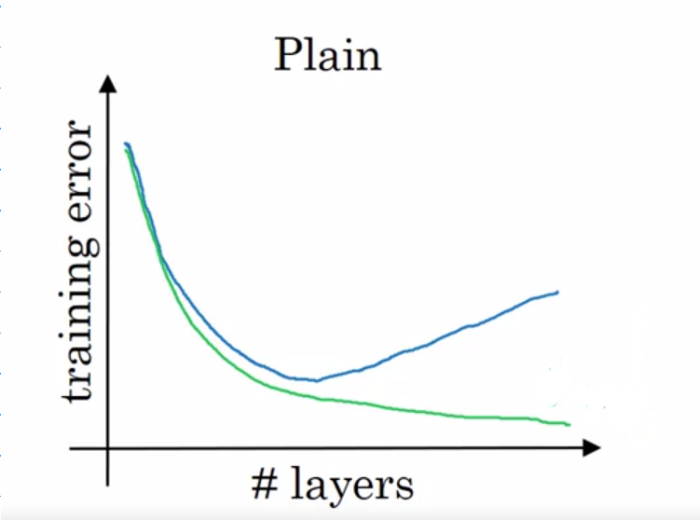
## Week 1 – The Basics of Conv Nets

1. What do you think applying this filter to a grayscale image will do?
2. Detect horizontal edges. (There is a high difference between the values in the top part from those in the bottom part of the matrix. When convolving this filter on a grayscale image, the horizontal edges will be detected.)
3. Suppose your input is a 300 by 300 colour (RGB) image, and you are not using a convolutional network. If the first hidden layer has 100 neurons, each one fully connected to the input, how many parameters does this hidden layer have (including the bias parameters)?
   1. 27,000,100 (The number of weights is . When you add the bias terms (one per neuron), you get .)
4. Suppose your input is a 256 by 256 grayscale image, and you use a convolutional layer with 128 filters that are each . How many parameters does this hidden layer have (including the bias parameters)?
   1. 1,280 (Since the input volume has only one channel, each filter has weights including the bias, which means the total is .)
5. You have an input volume that is , and convolve it with 32 filters of , using a stride of 2 and no padding. What is the output volume?
   1. (Using the formula , with , , and , we get 62. (the number of filters).)
6. You have an input volume that is , and pad it using “pad=1”. What is the dimension of the resulting volume (after padding)?
   1. (If the padding is 1, you add 2 to the height and width dimensions.)
7. You have an input volume that is , and convolve it with 32 filters that are each , and stride of 1. You want to use a “same” convolution. What is the padding?
   1. 3. (You need to satisfy the following equation: as you want to keep the dimensions between the input and output volumes. In order words,  
      .)
8. You have an input volume that is , and apply max pooling with a stride of 4 and a filter size of 4. What is the output volume?
   1. (Using the formula , with , , and , we get 32. .)
9. Because pooling layers do not have parameters, they do not affect the backpropagation (derivatives) calculation.
   1. False. (Everything that influences the loss should appear in the backpropagation because we are computing derivatives. In fact, pooling layers modify the input by choosing one value out of several values in their input volume. Also, to compute derivatives for the layers that have parameters (Convolutions, Fully-Connected), we still need to backpropagate the gradient through the Pooling layers.)
10. Which of the following are the benefits of using convolutional layers?
    1. It reduces the total number of parameters, thus reducing overfitting through parameter sharing. (A convolutional layer uses parameter sharing and usually has a lot fewer parameters than a fully-connected layer.)
    2. Convolutional layers are good at capturing translation invariance.
11. The sparsity of connections and weight sharing are mechanisms that allow us to use fewer parameters in a convolutional layer making it possible to train a network with smaller training sets. True/False?
    1. True. (Weight sharing reduces significantly the number of parameters in a neural network, and sparsity of connections allows us to use a smaller number of inputs thus reducing even further the number of parameters.)

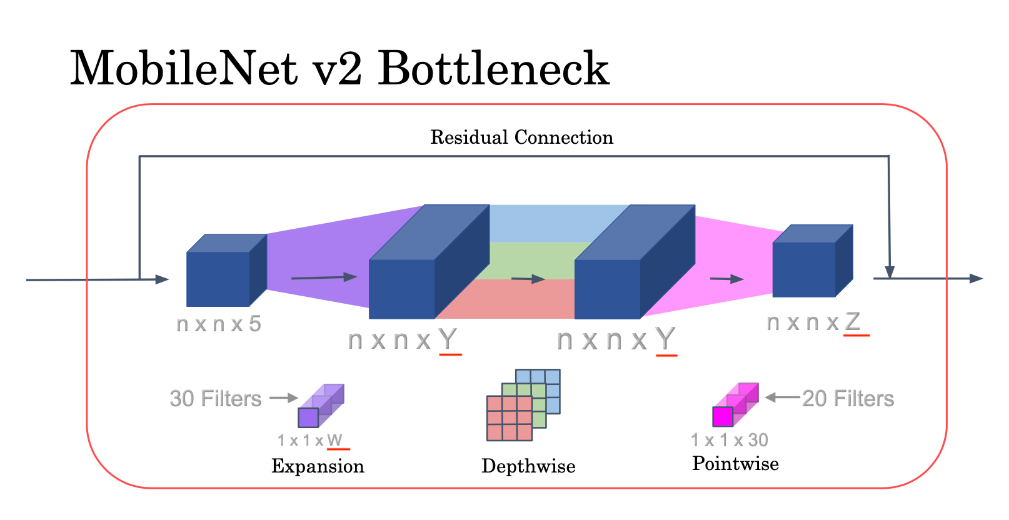
## Week 2 – Deep Convolutional Models

1. Which of the following do you typically see in a CNN?
   1. Use of fully-connected layers after flattening the volume to output classes. (FC layers are typically used in the last few layers to generate the output in classification.)
2. LeNet-5 made extensive use of padding to create valid convolutions, to avoid increasing the number of channels after every convolutional layer. True/False?
   1. False. (Padding wasn’t used back in 1998 when the LeNet paper was written.)
3. When using plain neural networks, which curve corresponds to the expected behaviour in theory and which curve corresponds to the behaviour seen in reality?



* 1. The green curve depicts the theoretical behaviour and the blue curve depicts the actual behaviour. (In theory, we expect the training error to decrease as the number of layers increases, but in reality, the error starts to increase after a certain number of layers.)

1. The following equation captures the computation in a ResNet block. What goes into the two blanks?
   1. and 0 respectively.
2. Adding a ResNet block to the end of a network makes it deeper. Which of the following is true?
   1. The performance of the network isn’t hurt as the ResNet block can easily approximate the identity function. (In a ResNet block, the computations are given by: Thus if and are zero then we get the identity function.)
3. Suppose you have an input volume of dimension . Which of the following statements do you agree with assuming that the convolutional layer has a stride of 1 and no padding.)
   1. You can use a convolutional layer to reduce but not or
   2. You can use a 2D pooling layer to reduce and but not
4. Which of the following are true about the Inception network?
   1. Inception blocks allow the use of a combination of , , and convolutions and pooling layers by stacking up all the activations resulting from each type of layer. (The use of several different types of layers and stacking up the results to get a single volume is at the heart of the Inception network.)
   2. One problem with simply stacking up several layers is the computational cost of it. (That is why the bottleneck layer is used to reduce the computational cost.)
5. When having a small training set to construct a classification model, which of the following is a strategy of transfer learning that you would use to build the model?
   1. Use an open-source network trained on a larger dataset, freeze the layers and re-train the softmax layer.
6. Which of the following are true about depthwise-separable convolutions?
   1. The pointwise convolution convolves the output volume with filters.
   2. The depthwise convolution convolves each channel in the input volume with a separate filter. (The output of this kind of convolution is the same as the input.)
   3. Depthwise-separable convolutions are composed of two different types of convolutions. (It is composed of a depthwise convolution followed by a pointwise convolution.)
7. Fill in the missing dimensions shown in the image below (marked W, Y, Z).



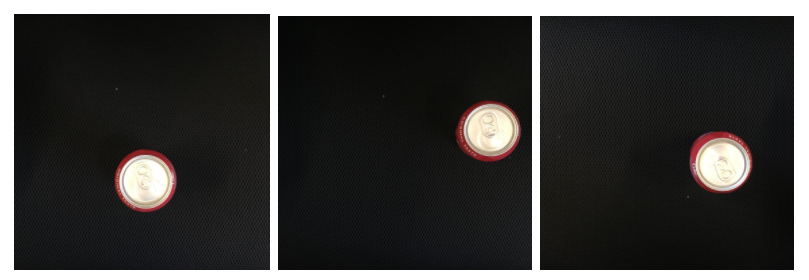
## Week 3 – Detection Algorithms

1. You are building a 3-class object classification and localisation algorithm. The classes are pedestrian (c = 1), car (c = 2) and motorcycle (c = 3). What should be for the image below? Recall that “?” means “don’t care”, which means that the neural network loss function won’t care what the neural network produces for that component of the output, and that .



* 1. . ( since there is a pedestrian in the picture. We can see that as percentages of the image are approximately correct as well , and the value of for a pedestrian.

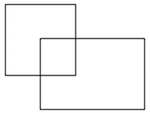
1. You are working on a factory automation task. Your system will see a can of soft-drink coming down a conveyor belt, and you want it to take a picture and decide whether (i) there is a soft-drink can in the image, and if so (ii) its bounding box. Since the soft-drink can is round, the bounding box is always square, and the soft drink can always appear the same size in the image. There is at most one soft drink can in each image. Here’re some typical images in your training set:



To solve this task, it is necessary to divide the task into two: 1. Construct a system to detect if a can is present or not. 2. Construct a system that calculates the bounding box of the can when present. Which of the following do you agree with the most?

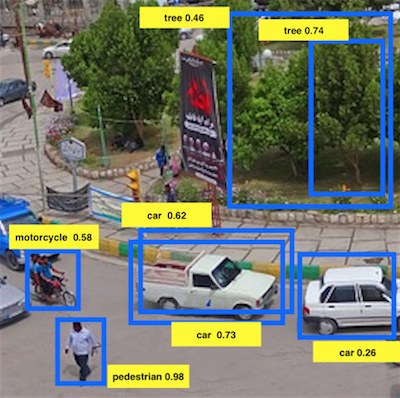
* 1. We can approach the task as an image classification with a localisation problem. (We can use a network to combine the two tasks.)

1. When building a neural network that inputs a picture of a person's face and outputs N landmarks on the face (assume that the input image contains exactly one face), we need two coordinates for each landmark, thus we need 2N output units. True/False?
   1. True. (Recall that each landmark is a specific position in the face’s image, thus we need to specify two coordinates for each landmark.)
2. When training one of the object detection systems described in the lectures, you need a training set that contains many pictures of the object(s) you wish to detect. However, bounding boxes do not need to be provided in the training set, since the algorithm can learn to detect the objects by itself.
   1. False. (You need bounding boxes in the training set. Your loss function should try to match the predictions for the bounding boxes to the true bounding boxes from the training set.)
3. What is the IoU between these two boxes? The upper-left box is 2x2, and the lower-right box is 2x3. The overlapping region is 1x1.



* 1. . (The left box’s area is 4 and the right box’s area is 6. Their intersection’s area is 1, and their union’s area is , which results in an intersection over union value of .)

1. Suppose you run non-max suppression on the predicted boxes below. The parameters you use for non-max suppression are that boxes with probability 4 are discarded, and the IoU threshold for deciding if two boxes overlap is 0.5.



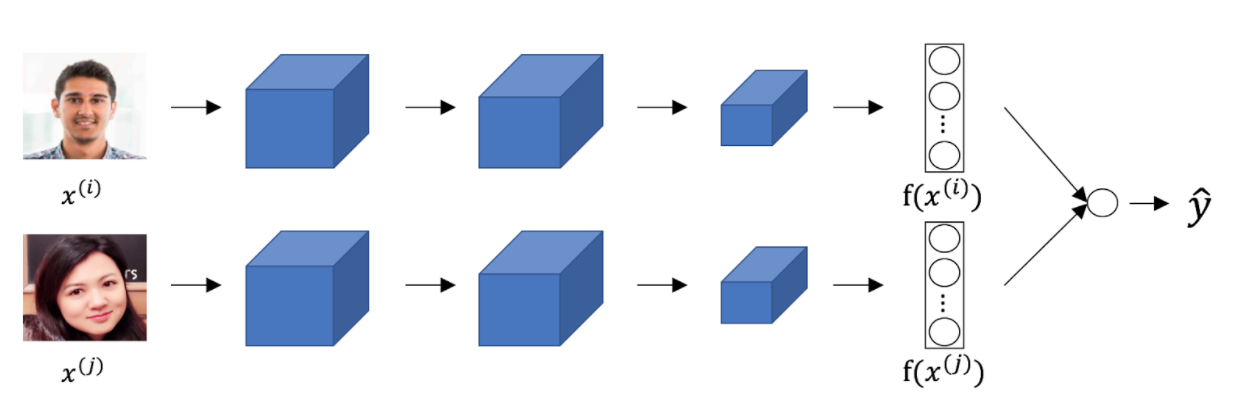
Notice that there are three bounding boxes for cars. After running non-max suppression, only the bounding box of the car with 0.73 is kept from the three bounding boxes for cars. True/False?

* 1. True. The non-maximum suppression eliminates the bounding boxes with scores lower than the ones of the maximum. (The bounding box for the car on the right is eliminated because its probability is less than 0.4. Of the two bounding boxes in the middle, one is eliminated because their IoU is higher than 0.5. So, only one bounding box remains.)

1. Which of the following do you agree with about the use of anchor boxes in YOLO?
   1. Each object is assigned to an anchor box with the highest IoU inside the assigned cell. (This is the way the corresponding anchor box is chosen.)
   2. Each object is assigned to the grid cell that contains that object’s midpoint. (This is the way the corresponding cell is chosen.)
2. What is semantic segmentation?
   1. Locating objects in an image by predicting the class of each pixel.
3. Using the concept of transpose convolution, fill in the values of , , and . Assume and .
   1. , ,
4. When using the U-Net architecture with an input , where denotes the number of channels, the output will always have the shape . True/False?
   1. False. (The output of the U-Net architecture can be , where is the number of classes. The number of channels doesn’t have to match between input and output.)

## Week 4 – Special Applications: Face Recognitions and Neural Style Transfer

1. Face verification requires comparing a new picture against one person’s face, whereas face recognition requires comparing a new picture against K persons’ faces.
   1. True
2. You want to build a system that receives a person's face picture and determines if the person is inside a workgroup. You have pictures of all the faces of the people currently in the workgroup, but some members might leave, and some new members might be added. Which of the following do you agree with?
   1. This can be considered a one-shot learning task. (Since we might have only one example of the person we want to recognise.)
   2. It will be more efficient to learn a function for this task. (Since this is a one-shot learning task, this function will allow us to compare two images to verify identity.)
3. In order to train the parameters of a face recognition system, it would be reasonable to use a training set comprising 100,000 pictures of 100,000 different persons.
   1. False. (To train a network using the triplet loss, you need several images of the same person.)
4. Which of the following is a correct definition of the triplet loss? Consider that .
5. Consider the following Siamese network architecture:



The upper and lower networks share parameters to have a consistent encoding for both images. True/False?

* 1. True. (Part of the idea behind the Siamese network is to compare the encoding of the images, thus they must be consistent.)

1. Units that respond more to complex features are more likely to be in deeper layers. True/False?
   1. True. (Neurons that understand more complex shapes are more likely to be in deeper layers of a neural network.)
2. Neural style transfer is trained as a supervised learning task in which the goal is to input two images , and train a network to output a new, synthesized image .
   1. False. (Neural style transfer is about training the pixels of an image to make it look artistic, it is not learning any parameters.)
3. In neural style transfer, we define style as:
   1. The correlation between activations across channels of an image. (This correlation is represented by for the image .
4. In neural style transfer, which of the following better express the gradients used?
   1. . (We use the gradient of the cost function over the value of the pixels of the generated image.)
5. You are working with 3D data. The input "image" has size , if you apply a convolutional layer with 16 filters of size , zero padding and stride 2. What is the size of the output volume?
   1. . (We can use the formula to the three first dimensions.)