

Quiz for Convolutional Network

Sunday, May 5, 2019

3:11 PM

1

point

1. Question 1

What do you think applying this filter to a grayscale image will do?

$\begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 1 & 3 & 3 & 1 & -1 & -3 & -3 & -1 & 0 & -1 & -1 & 0 \end{bmatrix}$



Detect vertical edges



Detect image contrast



Detect 45 degree edges



Detect horizontal edges

Question 2

1

point

2. Question 2

Suppose your input is a 300 by 300 color (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)?



9,000,001



9,000,100



27,000,001 ✗



27,000,100

Question 3

1

point

3. Question 3

Suppose your input is a 300 by 300 color (RGB) image, and you use a convolutional layer with 100 filters that are each 5x5. How many parameters does this hidden layer have (including the bias parameters)?



2501 ✗



2600 ✗



7500

☐
☒

7500

7600

Question 4

1

point

4. Question 4

You have an input volume that is $63 \times 63 \times 16$, and convolve it with 32 filters that are each 7×7 , using a stride of 2 and no padding. What is the output volume?

☒
☐
☐
☐

29x29x32

16x16x16

29x29x16

16x16x32

Question 5

1

point

5. Question 5

You have an input volume that is $15 \times 15 \times 8$, and pad it using "pad=2." What is the dimension of the resulting volume (after padding)?

☐
☐
☐
☒

19x19x12

17x17x10

17x17x8

19x19x8

Question 6

1

point

6. Question 6

You have an input volume that is $63 \times 63 \times 16$, and convolve it with 32 filters that are each 7×7 , and stride of 1. You want to use a "same" convolution. What is the padding?

☐
☐
☒
☐

1

2

3

7

Question 7

1

point

7. Question 7

You have an input volume that is $32 \times 32 \times 16$, and apply max pooling with a stride of 2 and a filter size of 2. What is the output volume?

☐
☐

16x16x8

15x15x16

☐
☒

32x32x8
16x16x16

Question 8

1

point

8. Question 8

Because pooling layers do not have parameters, they do not affect the backpropagation (derivatives) calculation.

☐
☒

True
False

Question 9

1

point

9. Question 9

In lecture we talked about “parameter sharing” as a benefit of using convolutional networks. Which of the following statements about parameter sharing in ConvNets are true? (Check all that apply.)

It allows parameters learned for one task to be shared even for a different task (transfer learning).

☐

It allows gradient descent to set many of the parameters to zero, thus making the connections sparse.

☐

It allows a feature detector to be used in multiple locations throughout the whole input image/input volume.

☒

It reduces the total number of parameters, thus reducing overfitting.

☒

Question 10

1

point

10. Question 10

In lecture we talked about “sparsity of connections” as a benefit of using convolutional layers. What does this mean?

☒

Each activation in the next layer depends on only a small number of activations from the previous layer.

☐

Regularization causes gradient descent to set many of the parameters to zero.✗

☐

Each layer in a convolutional network is connected only to two other layers

☐

Each filter is connected to every channel in the previous layer.

1

point

1. Question 1

Which of the following do you typically see as you move to deeper layers in a ConvNet?

- ☐ n_{HnH} and n_{WnW} decreases, while n_{CnC} also decreases
- ☒ n_{HnH} and n_{WnW} decrease, while n_{CnC} increases
- ☐ n_{HnH} and n_{WnW} increases, while n_{CnC} decreases
- ☐ n_{HnH} and n_{WnW} increases, while n_{CnC} also increases

Question 2

1

point

2. Question 2

Which of the following do you typically see in a ConvNet? (Check all that apply.)

- ☒ Multiple CONV layers followed by a POOL layer
- ☐ Multiple POOL layers followed by a CONV layer
- ☒ FC layers in the last few layers
- ☐ FC layers in the first few layers

Question 3

1

point

3. Question 3

In order to be able to build very deep networks, we usually only use pooling layers to downsize the height/width of the activation volumes while convolutions are used with “valid” padding. Otherwise, we would downsize the input of the model too quickly.

- ☐ True
- ☒ False

Question 4

1

point

4. Question 4

Training a deeper network (for example, adding additional layers to the network) allows the network to fit more complex functions and thus almost always results in lower training error. For this question, assume we’re referring to “plain” networks.

- ☐ True
- ☒ False

Question 5

1

point

point

5. Question 5

The following equation captures the computation in a ResNet block. What goes into the two blanks above?

$$a^{[l+2]} = g(W^{[l+2]}g(W^{[l+1]}a^{[l]} + b^{[l+1]}) + b^{[l+2]} + \underline{\hspace{2cm}}) + \underline{\hspace{2cm}}$$

- ☐ 0 and $z^{[l+1]}z^{[l+1]}$, respectively
- ☐ 0 and $a^{[l]}a^{[l]}$, respectively
- ☒ $a^{[l]}a^{[l]}$ and 0, respectively
- ☐ $z^{[l]}z^{[l]}$ and $a^{[l]}a^{[l]}$, respectively

Question 6

1

point

6. Question 6

Which ones of the following statements on Residual Networks are true? (Check all that apply.)

- ☒ Using a skip-connection helps the gradient to backpropagate and thus helps you to train deeper networks
- ☐ A ResNet with L layers would have on the order of L^2L^2 skip connections in total.
- ☐ The skip-connections compute a complex non-linear function of the input to pass to a deeper layer in the network.
- ☒ The skip-connection makes it easy for the network to learn an identity mapping between the input and the output within the ResNet block.

Question 7

1

point

7. Question 7

Suppose you have an input volume of dimension $64 \times 64 \times 16$. How many parameters would a single 1×1 convolutional filter have (including the bias)?

- ☐ 1
- ☐ 4097
- ☒ 17
- ☐ 2

Question 8

1

point

8. Question 8

Suppose you have an input volume of dimension $n_H \times n_W \times n_C$. Which of the following statements you agree with? (Assume that “ 1×1 convolutional layer” below

always uses a stride of 1 and no padding.)

☐ You can use a 1x1 convolutional layer to reduce n_{HnH} , n_{WnW} , and n_{CnC} .

☒ You can use a pooling layer to reduce n_{HnH} , n_{WnW} , but not n_{CnC} .

☒ You can use a 1x1 convolutional layer to reduce n_{CnC} but not n_{HnH} , n_{WnW} .

☐ You can use a pooling layer to reduce n_{HnH} , n_{WnW} , and n_{CnC} .

Question 9

1

point

9. Question 9

Which ones of the following statements on Inception Networks are true? (Check all that apply.)

Making an inception network deeper (by stacking more inception blocks together) should not hurt training set performance.	<input type="checkbox"/>
Inception blocks usually use 1x1 convolutions to reduce the input data volume's size before applying 3x3 and 5x5 convolutions.	<input checked="" type="checkbox"/>
A single inception block allows the network to use a combination of 1x1, 3x3, 5x5 convolutions and pooling.	<input checked="" type="checkbox"/>
Inception networks incorporate a variety of network architectures (similar to dropout, which randomly chooses a network architecture on each step) and thus has a similar regularizing effect as dropout.	<input type="checkbox"/>

Question 10

1

point

10. Question 10

Which of the following are common reasons for using open-source implementations of ConvNets (both the model and/or weights)? Check all that apply.

☐ The same techniques for winning computer vision competitions, such as using multiple crops at test time, are widely used in practical deployments (or production system deployments) of ConvNets.

☒ Parameters trained for one computer vision task are often useful as pretraining for other computer vision tasks.

☒ It is a convenient way to get working an implementation of a complex ConvNet architecture.

☐ A model trained for one computer vision task can usually be used to perform data augmentation even for a different computer vision task.

1. Question 1

You are building a 3-class object classification and localization algorithm. The classes are: pedestrian (c=1), car (c=2), motorcycle (c=3). What would be the label for the following image? Recall $y = [p_c, b_x, b_y, b_h, b_w, c_1, c_2, c_3]$
 $y=[pc,bx,by,bh,bw,c1,c2,c3]$



- ☒ $y = [1, 0.3, 0.7, 0.3, 0.3, 0, 1, 0]$ $y=[1,0.3,0.7,0.3,0.3,0,1,0]$
- ☐ $y = [1, 0.7, 0.5, 0.3, 0.3, 0, 1, 0]$ $y=[1,0.7,0.5,0.3,0.3,0,1,0]$
- ☐ $y = [1, 0.3, 0.7, 0.5, 0.5, 0, 1, 0]$ $y=[1,0.3,0.7,0.5,0.5,0,1,0]$
- ☐ $y = [1, 0.3, 0.7, 0.5, 0.5, 1, 0, 0]$ $y=[1,0.3,0.7,0.5,0.5,1,0,0]$
- ☐ $y = [0, 0.2, 0.4, 0.5, 0.5, 0, 1, 0]$ $y=[0,0.2,0.4,0.5,0.5,0,1,0]$

Question 2

1

point

2. Question 2

Continuing from the previous problem, what should y be for the image below?

Remember that “?” means “don’t care”, which means that the neural network loss function won’t care what the neural network gives for that component of the output.

As before, $y = [p_c, b_x, b_y, b_h, b_w, c_1, c_2, c_3]$ $y=[pc,bx,by,bh,bw,c1,c2,c3]$.





$y = [0, ?, ?, ?, ?, ?, ?, ?]$ $y = [0, ?, ?, ?, ?, ?, ?, ?]$



$y = [1, ?, ?, ?, ?, 0, 0, 0]$ $y = [1, ?, ?, ?, ?, 0, 0, 0]$



$y = [1, ?, ?, ?, ?, ?, ?, ?]$ $y = [1, ?, ?, ?, ?, ?, ?, ?]$



$y = [0, ?, ?, ?, ?, 0, 0, 0]$ $y = [0, ?, ?, ?, ?, 0, 0, 0]$



$y = [?, ?, ?, ?, ?, ?, ?, ?]$ $y = [?, ?, ?, ?, ?, ?, ?, ?]$

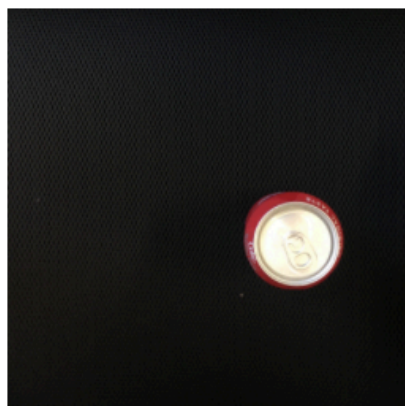
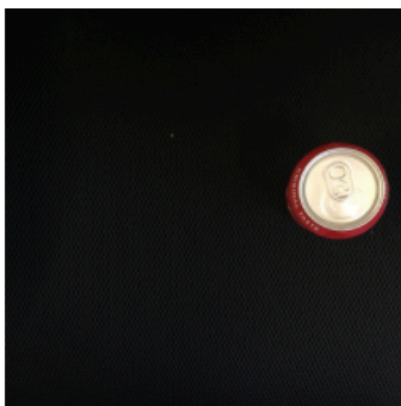
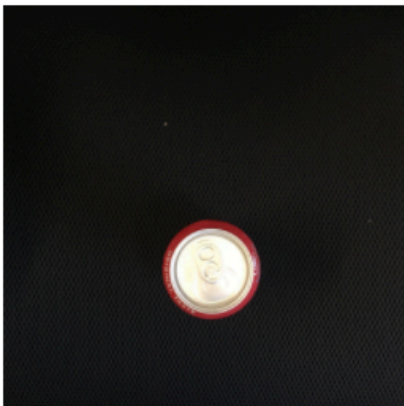
Question 3

1

point

3. Question 3

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 appears as 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:



What is the most appropriate set of output units for your neural network?

- ☐ Logistic unit (for classifying if there is a soft-drink can in the image)
- ☒ Logistic unit, b_{xbx} and b_{yby}
- ☐ Logistic unit, b_{xbx} , b_{yby} , b_{hbh} (since $b_{wbw} = b_{hbh}$)
- ☐ Logistic unit, b_{xbx} , b_{yby} , b_{hbh} , b_{wbw}

Question 4

1

point

4. Question 4

If you build a neural network that inputs a picture of a person's face and outputs N landmarks on the face (assume the input image always contains exactly one face), how many output units will the network have?

- ☐ N
- ☒ $2N$
- ☐ $3N$
- ☐ N^2

Question 5

1

point

5. Question 5

When training one of the object detection systems described in lecture, 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.

- ☐ True
- ☒ False

Question 6

1

point

6. Question 6

Suppose you are applying a sliding windows classifier (non-convolutional implementation). Increasing the stride would tend to increase accuracy, but decrease computational cost.

- ☐ True
- ☒ False

Question 7

1

point

7. Question 7

In the YOLO algorithm, at training time, only one cell ---the one containing the center/midpoint of an object--- is responsible for detecting this object

center/midpoint of an object--- is responsible for detecting this object.

☒
☐

True

False

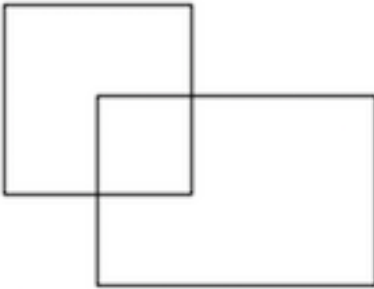
Question 8

1

point

8. Question 8

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.

☐
☒
☐
☐

1/6

1/9

1/10

None of the above

Question 9

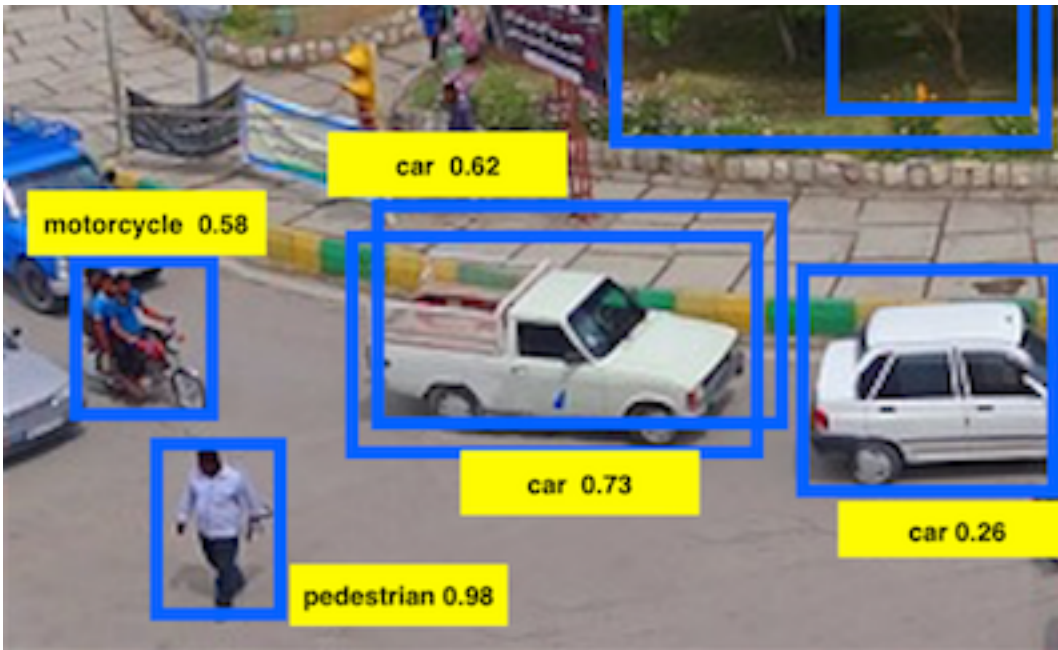
1

point

9. Question 9

Suppose you run non-max suppression on the predicted boxes above. The parameters you use for non-max suppression are that boxes with probability ≤ 0.4 are discarded, and the IoU threshold for deciding if two boxes overlap is 0.5. How many boxes will remain after non-max suppression?





- ☐
- ☐
- ☒
- ☐
- ☐

3
4
5
6
7

Question 10

1

point

10. Question 10

Suppose you are using YOLO on a 19×19 grid, on a detection problem with 20 classes, and with 5 anchor boxes. During training, for each image you will need to construct an output volume y as the target value for the neural network; this corresponds to the last layer of the neural network. (y may include some "?", or "don't cares"). What is the dimension of this output volume?

- ☐
- ☐
- ☒
- ☐

$19 \times 19 \times (20 \times 25)$
 $19 \times 19 \times (5 \times 20)$
 $19 \times 19 \times (5 \times 25)$
 $19 \times 19 \times (25 \times 20)$

1. Question 1

Face verification requires comparing a new picture against one person's face, whereas face recognition requires comparing a new picture against K person's faces.

- ☒
- ☐

True
False

Question 2

1

point

2. Question 2

Why do we learn a function $d(\text{img1}, \text{img2})$ for face verification? (Select all that apply.)

- ☐ This allows us to learn to predict a person's identity using a softmax output unit, where the number of classes equals the number of persons in the database plus 1 (for the final "not in database" class).
- ☒ We need to solve a one-shot learning problem.
- ☒ This allows us to learn to recognize a new person given just a single image of that person.
- ☐ Given how few images we have per person, we need to apply transfer learning.

Question 3

1

point

3. Question 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.

- ☐ True
- ☒ False

Question 4

1

point

4. Question 4

Which of the following is a correct definition of the triplet loss? Consider that $\alpha > 0$. (We encourage you to figure out the answer from first principles, rather than just refer to the lecture.)

- ☐ $\max(\|f(A)-f(N)\|^2 - \|f(A)-f(P)\|^2 - \alpha, 0)$
- ☐ $\max(\|f(A)-f(N)\|^2 - \|f(A)-f(P)\|^2 + \alpha, 0)$
- ☐ $\max(\|f(A)-f(P)\|^2 - \|f(A)-f(N)\|^2 - \alpha, 0)$
- ☒ $\max(\|f(A)-f(P)\|^2 - \|f(A)-f(N)\|^2 + \alpha, 0)$

Question 5

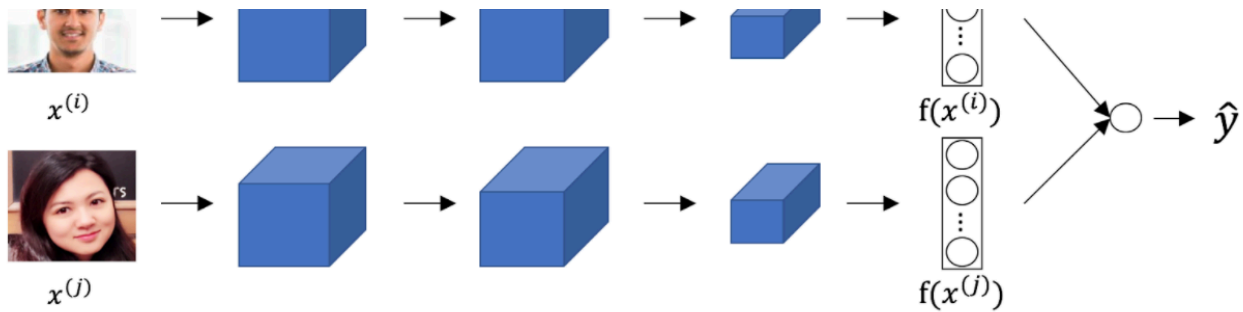
1

point

5. Question 5

Consider the following Siamese network architecture:





The upper and lower neural networks have different input images, but have exactly the same parameters.



Question 6

1

point

6. Question 6

You train a ConvNet on a dataset with 100 different classes. You wonder if you can find a hidden unit which responds strongly to pictures of cats. (I.e., a neuron so that, of all the input/training images that strongly activate that neuron, the majority are cat pictures.) You are more likely to find this unit in layer 4 of the network than in layer 1.



Question 7

1

point

7. Question 7

Neural style transfer is trained as a supervised learning task in which the goal is to input two images (x), and train a network to output a new, synthesized image (y).



Question 8

1

point

8. Question 8

In the deeper layers of a ConvNet, each channel corresponds to a different feature detector. The style matrix $G^l G^l$ measures the degree to which the activations of different feature detectors in layer l vary (or correlate) together with each other.



Question 9

1

+

point

9. Question 9

In neural style transfer, what is updated in each iteration of the optimization algorithm?

☐

The regularization parameters

☐

The neural network parameters

☒

The pixel values of the generated image GG

☐

The pixel values of the content image CC

Question 10

1

point

10. Question 10

You are working with 3D data. You are building a network layer whose input volume has size $32 \times 32 \times 32 \times 16$ (this volume has 16 channels), and applies convolutions with 32 filters of dimension $3 \times 3 \times 3$ (no padding, stride 1). What is the resulting output volume?

☒

$30 \times 30 \times 30 \times 32$

☐

Undefined: This convolution step is impossible and cannot be performed because the dimensions specified don't match up.

☐

$30 \times 30 \times 30 \times 16$