Consider an input image of size $1000 \times 1000 \times 10$ where 10 refers to the number of channels (Such images do exist!). Suppose we want to apply a convolution operation on the entire image by sliding a kernel of size $1 \times 1 \times d$ What should be the depth d of the kernel?

Convolution with a Kernel: Summary

1. **Image Dimensions:**
   - The image size is $1000 \times 1000 \times 10$, where 10 represents the number of channels.
2. **Kernel Dimensions:**
   - The kernel size is $1 \times 1 \times d$.
3. **Depth Matching:**
   - The kernel's depth d must match the number of channels in the image for the convolution to work.
   - Since the image has 10 channels, **the depth d of the kernel must be 10.**

2) For the same input image in Q1, suppose that we apply the following kernels of differing sizes.

$K_1 : 3 \times 3$
$K_2 : 7 \times 7$
$K_3 : 17 \times 17$
$K_4 : 41 \times 41$

Assume that stride $s = 1$ and no zero padding. Among all these kernels which one shrinks the output dimensions the most?

- ○ K_1
- ○ K_2
- ○ K_3
- ◉ K_4

To determine which kernel shrinks the output dimensions the most for an input image of size $1000 \times 1000 \times 10$ when applying kernels of sizes $3 \times 3, 7 \times 7, 17 \times 17$, and 41×41 with a stride of $s = 1$ and no zero padding, we can use the formula for the output dimensions after applying a convolution:

Formula for Output Dimensions

Given an input of size $H \times W$ (height and width), kernel size $K \times K$, stride s, and no padding, the output dimensions $H' \times W'$ can be calculated as follows:

$$H' = \left(\frac{H - K}{s}\right) + 1$$

$$W' = \left(\frac{W - K}{s}\right) + 1$$

Since $s = 1$ and no padding, this simplifies to:

$$H' = W' = H - K + 1$$

Calculation for Each Kernel:

1. **For K1 (3×3):**

$$H' = 1000 - 3 + 1 = 998$$

2. **For K2 (7×7):**

$$H' = 1000 - 7 + 1 = 994$$

3. **For K3 (17×17):**

$$H' = 1000 - 17 + 1 = 984$$

4. **For K4 (41×41):**

$$H' = 1000 - 41 + 1 = 960$$

Comparing Output Dimensions:

- **K1**: 998×998
- **K2**: 994×994
- **K3**: 984×984
- **K4**: 960×960

Conclusion:

Among all these kernels, **K4 (41×41)** shrinks the output dimensions the most, with an output size of 960×960.

The term "shrink" indicates how much smaller the output dimensions (height and width) become after the convolution operation is performed. Larger kernels typically result in greater shrinkage of the output dimensions.

3) Which of the following statements about CNN is (are) true?

- [x] CNN is a feed-forward network
- [x] Weight sharing helps CNN layers to reduce the number of parameters
- [ ] CNN is suitable only for natural images
- [ ] The shape of the input to the CNN network should be square

Yes, the answer is correct.
Score: 1

Accepted Answers:
CNN is a feed-forward network

Weight sharing helps CNN layers to reduce the number of parameters

Consider an input image of size $100 \times 100 \times 3$. Suppose that we used kernel of size 3×3, zero padding $P=1$and stride value $S=2$. What will be the output dimension?

Given Parameters:

- **Input Size**: $100 \times 100 \times 3$ (height x width x channels)
- **Kernel Size**: 3×3
- **Padding** P: 1
- **Stride** S: 3

Output Size Calculation

Using the output size formula:

$$\text{Output Size} = \frac{\text{Input Size} - \text{Kernel Size} + 2 \times \text{Padding}}{\text{Stride}} + 1$$

1. **Calculate the output height:**

$$\begin{aligned}\text{Output Height} &= \frac{100 - 3 + 2 \times 1}{3} + 1 \\ &= \frac{100 - 3 + 2}{3} + 1 \\ &= \frac{99}{3} + 1 \\ &= 33 + 1 = 34\end{aligned}$$

2. **Calculate the output width** (same calculation as height):

$$\begin{aligned}\text{Output Width} &= \frac{100 - 3 + 2 \times 1}{3} + 1 \\ &= \frac{100 - 3 + 2}{3} + 1 \\ &= \frac{99}{3} + 1 \\ &= 33 + 1 = 34\end{aligned}$$

3. **Output depth:** Assuming we are using one filter, the output depth will remain the same as the input depth, which is 3.

Final Output Dimensions

Thus, the output dimensions after the convolution operation will be:

$$\text{Output Dimensions} = (34, 34, 3)$$

So, the final output size of the image after applying the convolution will be $34 \times 34 \times 3$.

2. Final Formula for Output Dimensions

Including width, height, and depth, the formulas are as follows:

- Width:

$$W_2 = \frac{W_1 - F + 2P}{S} + 1$$

- Height:

$$H_2 = \frac{H_1 - F + 2P}{S} + 1$$

- Depth:

$$D_2 = K$$

5) Consider an input image of size $100 \times 100 \times 3$. Suppose that we use 10 kernels (filters) each of size 1×1, zero padding $P = 1$ and stride value $S = 2$. How many parameters are there? (assume no bias terms)

- 5
- 10
- 15
- 30

Yes, the answer is correct.
Score: 1

Accepted Answers:
30

$$\text{Number of Parameters} = (\text{Kernel Height} \times \text{Kernel Width} \times \text{Input Depth}) \times \text{Number of Kernels}$$

Given Parameters:

- **Kernel Size:** 1×1
- **Input Depth:** 3 (since the input image is of size $100 \times 100 \times 3$)
- **Number of Kernels:** 10
- **Padding:** 1 (not relevant for parameter calculation)
- **Stride:** 2 (not relevant for parameter calculation)

Calculating the Number of Parameters:

1. **Calculate the parameters for one kernel:**
   - Kernel height = 1
   - Kernel width = 1
   - Input depth = 3

$$\text{Parameters per Kernel} = 1 \times 1 \times 3 = 3$$

2. **Total Parameters for all Kernels:**

$$\text{Total Parameters} = 3 \times 10 = 30$$

Conclusion

Therefore, the total number of parameters when using 10 kernels of size 1×1 with no bias terms is **30.**

The correct answer is **30.**

6) Which statement is true about the size of filters in CNNs?

- [ ] The size of the filter does not affect the features it captures.
- [ ] The size of the filter only affects the computation time.
- [x] Larger filters capture more global features.
- [x] Smaller filters capture more local features.

Yes, the answer is correct.
Score: 1

Accepted Answers:

Larger filters capture more global features.

Smaller filters capture more local features.

1. **Larger filters capture more global features.**
   - Larger filters have a wider receptive field, allowing them to capture broader patterns and features in the input data, making them more suitable for identifying global structures.
2. **Smaller filters capture more local features.**
   - Smaller filters focus on a limited area of the input, which enables them to detect finer details and local patterns, such as edges or textures.

7) What is the motivation behind using multiple filters in one Convolution layer?

- ○ Reduced complexity of the network
- ○ Reduced size of the convolved image
- ○ Insufficient information
- ◉ Each filter captures some feature of the image separately

Yes, the answer is correct.
Score: 1

Accepted Answers:

Each filter captures some feature of the image separately

8) Which of the following architectures has the highest no of layers?

- ○ AlexNet
- ○ GoogleNet
- ◉ ResNet
- ○ VGG

Yes, the answer is correct.
Score: 1

Accepted Answers:
ResNet

1. **AlexNet**:
   - Typically has **8 layers** (5 convolutional layers followed by 3 fully connected layers).
2. **VGG**:
   - Commonly refers to VGG16 or VGG19, which have **16 and 19 layers**, respectively. Most layers are convolutional.
3. **GoogleNet (Inception)**:
   - Has **22 layers** when considering its architecture with auxiliary classifiers and various convolutional layers.
4. **ResNet**:
   - The ResNet architecture comes in various configurations, with the ResNet-50 having **50 layers**, ResNet-101 having **101 layers**, and even deeper versions like ResNet-152 with **152 layers**.

9) What is the purpose of guided backpropagation in CNNs?

- ○ To train the CNN to improve its accuracy on a given task.
- ○ To reduce the size of the input images in order to speed up computation.
- ◉ To visualize which pixels in an image are most important for a particular class prediction.
- ○ None of the above.

Yes, the answer is correct.
Score: 1

Accepted Answers:
To visualize which pixels in an image are most important for a particular class prediction.

10) Which of the following statements is true regarding the occlusion experiment in a CNN?

- [ ] It is a technique used to prevent overfitting in deep learning models.
- [ ] It is used to increase the number of filters in a convolutional layer.
- [ ] It is used to determine the importance of each feature map in the output of the network.
- [x] It involves masking a portion of the input image with a patch of zeroes.

Partially Correct.
Score: 0.5

Accepted Answers:

It is used to determine the importance of each feature map in the output of the network.

It involves masking a portion of the input image with a patch of zeroes.

Occlusion Experiment in CNNs:

- **Method**: It involves masking a portion of the input image with a patch of zeroes.
- **Purpose**: It helps determine the importance of different regions of the input image, which can indirectly indicate the relevance of the associated feature maps in the output of the network.

Consider the following:

- Weight array, W = [0.2, 0.7, 0.05, 0.75, 0.86, 0.21]
- Input X = [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9]

What is the next revised estimate S that is obtained by sliding the filter W over the input X?

Note: Include two digits after the decimal.

Example:

For a weight array W of size $k = 6$ and an input X of size n:

- The first revised estimate S_0 corresponds to the first 6 elements of X:

$$S_0 = W_0 \cdot X_0 + W_1 \cdot X_1 + W_2 \cdot X_2 + W_3 \cdot X_3 + W_4 \cdot X_4 + W_5 \cdot X_5$$

As the filter slides, you would calculate S_1 using $X[1]$ to $X[6]$, and so forth until the end of the input array.

Here only asked next st means S1

The first revised estimate S_1 using the weight array W and the input X can be calculated as follows:

$$S_1 = (0.2 \times 0.1) + (0.7 \times 0.2) + (0.05 \times 0.3) + (0.75 \times 0.4) + (0.86 \times 0.5) + (0.21 \times 0.6)$$

Calculating this gives:

$$S_1 = 0.02 + 0.14 + 0.015 + 0.3 + 0.43 + 0.126 = 1.031$$

Rounded to two decimal places:
$S_1 = 1.03.$

2) Given the following Input matrix and kernel, What is the (0,0) element in the output matrix of the convolution operation if the stride rate is 1?
Input Matrix:

1	2	3	4
2	1	4	3
1	3	2	4
4	1	2	3

Kernel Matrix:

1	2
2	1

2. **Perform element-wise multiplication of the overlapping elements and sum the results.** The calculation will be as follows:

$$\text{(0,0) Element} = (1 \times 1) + (2 \times 2) + (2 \times 2) + (1 \times 1)$$

Breaking it down:

- $1 \times 1 = 1$
- $2 \times 2 = 4$
- $2 \times 2 = 4$
- $1 \times 1 = 1$

3. **Sum the results:**

$$1 + 4 + 4 + 1 = 10$$

Conclusion:

The (0,0) element in the output matrix of the convolution operation is **10**.

3) Predict the output image for the convolution operation on the given input image and filter.

Input:

Kernel:

1	1	1
1	1	1
1	1	1

○ 

4) Identify the kernel that can be used to sharpen an image.

Consider a 2D filter convolution over a 3D input. What would be the dimension of the output?

Given the input image with dimensions 125 x 49, filter of size 5 x 5, padding P = 2 and stride S = 2, what is the dimension of the output image of this convolution operation?

Options:

- 62 x 24
- 63 x 25
- 64 x 26
- None of the above

To calculate the dimensions of the output image after applying a convolution operation, you can use the following formula:

$$\text{Output Dimension} = \left\lfloor \frac{\text{Input Dimension} + 2P - F}{S} \right\rfloor + 1$$

where:

- P = padding
- F = filter size
- S = stride
- The floor function $\lfloor \cdot \rfloor$ indicates that we take the whole number part of the calculation.

Given:

- Input dimensions: 125×49
- Filter size: 5×5 (so $F = 5$)
- Padding: $P = 2$
- Stride: $S = 2$

Let's calculate the output dimensions separately for each dimension:

1. **For the height (125):**

$$\text{Output Height} = \left\lfloor \frac{125 + 2 \cdot 2 - 5}{2} \right\rfloor + 1$$

$$= \left\lfloor \frac{125 + 4 - 5}{2} \right\rfloor + 1$$

$$= \left\lfloor \frac{124}{2} \right\rfloor + 1$$

$$= \lfloor 62 \rfloor + 1 = 62 + 1 = 63$$

2. **For the width (49):**

$$\text{Output Width} = \left\lfloor \frac{49 + 2 \cdot 2 - 5}{2} \right\rfloor + 1$$

$$= \left\lfloor \frac{49 + 4 - 5}{2} \right\rfloor + 1$$

$$= \left\lfloor \frac{48}{2} \right\rfloor + 1$$

$$= \lfloor 24 \rfloor + 1 = 24 + 1 = 25$$

Now we combine the output dimensions:

- Output height = 63
- Output width = 25

7. Pick out the technique(s) that involve static feature extraction using handcrafted kernels to learn the weights of the classifier.

- SIFT
- HOG
- Both SIFT and HOG
- Neither SIFT nor HOG

The correct answer is **Both SIFT and HOG**.

Both SIFT (Scale-Invariant Feature Transform) and HOG (Histogram of Oriented Gradients) are techniques that involve static feature extraction using handcrafted kernels. They do not learn features from data in the way that deep learning methods do; instead, they extract specific features based on predefined algorithms. These features can then be used with classifiers that may learn weights but do so based on the handcrafted features extracted by SIFT and HOG.

9) What is the number of parameters to be computed for the convolution operation given the following notations?

Input image dimension = $a \times a \times d$

Number of filters = n

Filter dimension = $f \times f$

Stride = s

Padding = p

○ $(f \times f \times d) \times n$

○ $(n \times n \times d) \times f$

○ $(n \times f \times n) \times d$

○ $(d \times d \times f) \times n$

8. Given the stride rate 1, compute the output for the Average pooling operation on the given input matrix.

3 7 6 4

2 6 4 8

1 8 5 2

1 9 7 5

Average Pooling Calculation (2x2 window, stride 1)

1. **(1st window)** `3, 7, 2, 6` → Average = $\frac{3+7+2+6}{4} = 4.5$
2. **(2nd window)** `7, 6, 6, 4` → Average = $\frac{7+6+6+4}{4} = 5.75$
3. **(3rd window)** `6, 4, 4, 8` → Average = $\frac{6+4+4+8}{4} = 5.5$
4. **(4th window)** `2, 6, 1, 8` → Average = $\frac{2+6+1+8}{4} = 4.25$
5. **(5th window)** `6, 4, 8, 5` → Average = $\frac{6+4+8+5}{4} = 5.75$
6. **(6th window)** `4, 8, 5, 2` → Average = $\frac{4+8+5+2}{4} = 4.75$
7. **(7th window)** `1, 8, 1, 9` → Average = $\frac{1+8+1+9}{4} = 4.75$
8. **(8th window)** `8, 5, 9, 7` → Average = $\frac{8+5+9+7}{4} = 7.25$
9. **(9th window)** `5, 2, 7, 5` → Average = $\frac{5+2+7+5}{4} = 4.75$

```
4.5   5.75  5.5
4.25  5.75  4.75
4.75  7.25  4.75
```

The question is:

9. What is the number of parameters to be computed for the convolution operation given the following notations?

Notations:

- Input image dimension = a x a x d
- Number of filters = n
- Filter dimension = f x f
- Stride = s
- Padding = p

Options:

- (f x f x d) x n
- (n x n x d) x f

- (n x f x n) x d
- (d x d x f) x n

- **Input image dimension** = $a \times a \times d$ (where a is the height and width of the image, and d is the number of channels, e.g., RGB channels)
- **Number of filters** = n
- **Filter dimension** = $f \times f$
- **Stride** = s
- **Padding** = p

Parameters in Convolution

For each filter, the parameters consist of the weights and biases that the filter learns. The weights required for each filter depend on the size of the filter and the depth of the input:

1. **Weights for each filter:** A filter of size $f \times f$ that operates across d input channels will require $f \times f \times d$ weights.
2. **Bias for each filter:** Each filter typically has one additional bias term.

Total Parameters Calculation

Thus, for n filters, the total number of parameters can be calculated as:

$$\text{Total parameters} = n \times (f \times f \times d + 1)$$

However, since the options seem to not include the bias term, we will focus only on the weights part:

$$\text{Total weights} = n \times (f \times f \times d)$$

10) We have a trained CNN. We have the picture on the left which when fed into the network as input is given the label 'HUMAN' with high probability. The picture on the right is the same image with some added noise. If we feed the right image as input to the CNN then which of the following statements is True? *1 poin*

Left Image Right Image

- (selected) CNN will detect the image as 'HUMAN'
- CNN will not detect the image as 'HUMAN' since noise is added to the image.
- CNN will detect the image as 'HUMAN' but with a lower probability than the left image.

2) Consider a convolution operation with an input image of size 100x100x3 and a filter of size 8x8x3, using a stride of 1 and a padding of 1. What is the output size? *1 point*

- 100x100x3
- 98x98x1
- 102x102x3
- (selected) 95x95x1

3) Consider a convolution operation with an input image of size 256x256x3 and 40 filters of size 11x11x3, using a stride of 4 and a padding of 2. What is the height of the output size? *1 point*

- ( ) 63
- (•) 64
- ( ) 40
- ( ) 3

4) Which statement is true about the number of filters in CNNs? *1 point*

- (•) More filters lead to better accuracy.
- ( ) Fewer filters lead to better accuracy.
- ( ) The number of filters has no effect on accuracy.
- ( ) The number of filters only affects the computation time.

6) Which of the following is an innovation introduced in GoogleNet architecture? *1 point*

- [x] 1x1 convolutions to reduce the dimension
- [ ] ReLU activation function
- [ ] Dropout regularization
- [x] use of different-sized filters for the same input

7) What is the purpose of guided backpropagation in CNNs? *1 point*

- (•) To visualize which pixels in an image are most important for a particular class prediction.
- ( ) To train the CNN to improve its accuracy on a given task.
- ( ) To reduce the size of the input images in order to speed up computation.
- ( ) None of the above.

7) What is the purpose of guided backpropagation in CNNs? *1 point*

- (•) To visualize which pixels in an image are most important for a particular class prediction.
- ( ) To train the CNN to improve its accuracy on a given task.
- ( ) To reduce the size of the input images in order to speed up computation.
- ( ) None of the above.

8) Which layer in a CNN is used for guided backpropagation? *1 point*

- ( ) Input layer
- ( ) Convolutional layer
- (•) Activation layer
- ( ) Pooling layer

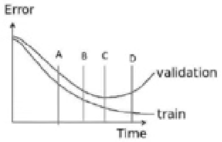
9) Which of the following is a technique used to fool CNNs in Deep Learning? *1 pc*

- (•) Adversarial examples
- ( ) Transfer learning
- ( ) Dropout
- ( ) Batch normalization

7) What is the purpose of guided backpropagation in CNNs? *1 point*

- (•) To visualize which pixels in an image are most important for a particular class prediction.
- ( ) To train the CNN to improve its accuracy on a given task.
- ( ) To reduce the size of the input images in order to speed up computation.
- ( ) None of the above.

While training a neural network for image recognition task, we plot the graph of training error and validation error. Which is the best for early stopping?



a. A
b. B
c. C
d. D

- a.
- b.
- c.
- d.

4)

Which among the following is NOT a data augmentation technique?

a. Random horizontal and vertical flip of image
b. Random shuffle all the pixels of an image
c. Random color jittering
d. All the above are data augmentation techniques

- a.
- b.
- c.
- d.

6) Which of the following is an innovation introduced in GoogleNet architecture?

- 1x1 convolutions to reduce the dimension
- ReLU activation function
- Dropout regularization
- use of different-sized filters for the same input

8) Which layer in a CNN is used for guided backpropagation?

- Input layer
- Convolutional layer
- Activation layer
- Pooling layer

9) Which of the following is a technique used to fool CNNs in Deep Learning?

- Adversarial examples
- Transfer learning
- Dropout
- Batch normalization

10) We have a trained CNN. We have the picture on the left which when fed into the network as input is given the label 'HUMAN' with high probabili
The picture on the right is the same image with some added noise. If we feed the right image as input to the CNN then which of the following stateme
True?

Left Image Right Image

- ○ CNN will detect the image as 'HUMAN'
- ○ CNN will not detect the image as 'HUMAN' since noise is added to the image.
- ○ CNN will detect the image as 'HUMAN' but with a lower probability than the left image.
- ○ Insufficient information to say anything

Which among the following is NOT a data augmentation technique?

a. Random horizontal and vertical flip of image
b. Random shuffle all the pixels of an image
c. Random color jittering
d. All the above are data augmentation techniques

○ a.
◉ b.
○ c.
○ d.

Batch Normalization is helpful because

a. It normalizes all the input before sending it to the next layer
b. It returns back the normalized mean and standard deviation of wei
c. It is a very efficient back-propagation technique
d. None of these

◉ a.
○ b.
○ c.
○ d.

1) Which of the following 3x3 kernels works best for the edge detection task in an image?

- $\begin{bmatrix} 0 & -1 & 4 \\ -1 & 7 & -1 \\ 0 & -1 & 0 \end{bmatrix}$
- $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$
- $\begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix}$
- $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$

3) We have the following two images. We observe that a neuron is in the 3rd convolutional layer of a CNN. Which of the following statements is always true?

Image 1 Image 2

- ○ The neuron will fire again because both images contain the word 'OPEN'
- ○ The neuron will not fire again because the color of the word 'OPEN' is different in both images
- ○ The neuron will fire again since the background for images is dark.
- ◉ None of these.

4) What is the main innovation proposed in Resnet architecture?

- Use of 1x1 convolutions
- Residual connections
- Use of different-sized filters at a single layer
- Increasing the width of the network.

5) Consider a convolution operation with an input image of size 256x256x3 and 63 filters of size 11x11x3, using a stride of 4 and a padding of 2. What is the depth (h in w x b x h) of the output image?

- 63
- 64
- 40
- 3

6) What is the result of convolving a 3x3 matrix with a 2x2 filter with stride 1 and no padding?

- 2 × 2 matrix
- 1 × 1 matrix
- 4 × 4 matrix
- 3 × 3 matrix

7) What is the purpose of using a stride larger than 1 in convolutional neural networks?

- To increase the number of convolution operation in a layer
- To reduce the size of the output feature map
- To increase the number of trainable parameters
- To increase the receptive field of the convolutional layer

8) Which of the following layers doesn't contain learnable parameters in CNN?

- Max-pooling layers
- Convolutional layers
- Feedforward Network's hidden layers
- Feedforward Network's output layer

9) Which statement is true about the size of filters in CNNs?

- Larger filters capture more global features.
- Smaller filters capture more local features.
- The size of the filter does not affect the features it captures.
- The size of the filter only affects the computation time.

10) What is the purpose of guided backpropagation in CNNs?

- To visualize which pixels in an image are most important for a particular class prediction.
- To train the CNN to improve its accuracy on a given task.
- To reduce the size of the input images in order to speed up computation.
- None of the above.

Here are the questions with options, followed by the answers:

Question 1:

What is the purpose of guided backpropagation in CNNs?

- **Options:**
  1. To visualize which pixels in an image are most important for a prediction.
  2. To train a CNN to be more accurate on specific types of images.
  3. To speed up the training process of a CNN.
  4. To reduce the number of parameters in a CNN.
  5. None of the above.

- **Answer:** 1. **To visualize which pixels in an image are most important for a prediction.**

Question 2:

Which layer in a CNN is used for guided backpropagation?

- **Options:**
  1. Input layer
  2. Convolutional layer
  3. Pooling layer
  4. Fully connected layer
  5. All of the above

- **Answer:** 2. **Convolutional layer**

Question 3:
Which of the following is a technique used to fool CNNs in Deep Learning?

- **Options:**
  1. Adversarial examples
  2. Transfer learning
  3. Dropout
  4. Batch normalization
  5. None of the above

- **Answer:** 1. **Adversarial examples**

Question 4:
Which of the following is an innovation introduced in GoogleNet architecture?

- **Options:**
  1. 1x1 convolutions to reduce the dimension
  2. ReLU activation function
  3. Dropout regularization
  4. Use of different-sized filters for the same input
  5. None of the above

- **Answer:** 1. **1x1 convolutions to reduce the dimension**

Question 5:

Which of the following is NOT a data augmentation technique?

- **Options:**
  1. Random cropping
  2. Random flipping
  3. Random rotation
  4. Random color jittering
  5. All of the above are data augmentation techniques

- **Answer:** 5. **All of the above are data augmentation techniques**

Question 6:

Which of the following is the most important factor to consider when choosing a loss function?

- **Options:**
  1. The type of data being used
  2. The complexity of the model
  3. The desired output format
  4. The computational resources available
  5. None of the above

- **Answer:** 1. **The type of data being used**

Question 1:
Which of the following 3x3 kernels works best for the edge detection task in an image?

- **Options:**
  1. [0,−1,4],[−1,0,−1],[4,−1,0][0, -1, 4], [-1, 0, -1], [4, -1, 0][0,−1,4],[−1,0,−1],[4,−1,0]
  2. [0,0,0],[0,7,0],[0,0,0][0, 0, 0], [0, 7, 0], [0, 0, 0][0,0,0],[0,7,0],[0,0,0]
  3. [0,−1,0],[−1,8,−1],[0,−1,0][0, -1, 0], [-1, 8, -1], [0, -1, 0][0,−1,0],[−1,8,−1],[0,−1,0]
  4. [1,1,1],[1,1,1],[1,1,1][1, 1, 1], [1, 1, 1], [1, 1, 1][1,1,1],[1,1,1],[1,1,1]
  5. [1,1,1],[1,−8,1],[1,1,1][1, 1, 1], [1, -8, 1], [1, 1, 1][1,1,1],[1,−8,1],[1,1,1]
- **Answer:** 3. **[0,−1,0],[−1,8,−1],[0,−1,0][0, -1, 0], [-1, 8, -1], [0, -1, 0][0,−1,0],[−1,8,−1],[0,−1,0]**

Question 2:
We have the same CNN. We have the picture on the left, which, when fed into the network as input, is given the label "HUMAN" with high probability. The picture on the right is the same image with some noise added to it. When fed as input to the CNN, which of the following statements is true?

- **Options:**
  1. CNN will detect the image as "HUMAN"
  2. CNN will detect the image as "NOT HUMAN"
  3. CNN will detect the image as "HUMAN" but with a lower probability than the left image

    4. CNN will not say anything

- **Answer:** 3. **CNN will detect the image as "HUMAN" but with a lower probability than the left image**

Question 3:
Which among the following is NOT a data augmentation technique?

- **Options:**
    1. Random horizontal and vertical flip of the image
    2. Random shuffle all the pixels of an image
    3. Random color jittering
    4. All the above are data augmentation techniques
- **Answer:** 2. **Random shuffle all the pixels of an image**

Question 4:
We have the following two images. We have trained a CNN for the 3rd convolutional layer of a CNN. Which of the following statements is always true?

- **Options:**
    1. The neuron will fire again because both images contain the word "OPEN"
    2. The neuron will not fire again because the color of the word "OPEN" is different in both images
    3. The neuron will fire again since the background for both images is different
    4. None of these
- **Answer:** 4. **None of these**