

Instructions

You have 45 minutes to attempt the quiz.

This quiz was locked May 17 at 6am.

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	15 minutes	200 out of 200

Score for this quiz: 200 out of 200
Submitted May 16 at 8:10pm
This attempt took 15 minutes.

Question 1

20 / 20 pts

According to the universal approximation theorem, given enough capacity, we know that a feedforward network with a single layer is sufficient to represent any function. However, the layer might be massive and the network is prone to overfitting the data. Therefore, there is a common trend in the research community that our network architecture needs to go deeper.

Deeper layers help in:

Correct!

☒ Creating large number of correlation templates for same objects

Correct!

☒ Break down large objects in smaller components

Correct!

☒ Achieving higher receptive fields

Correct!

☒ Easier training when compared to only 1 layer

Question 2

20 / 20 pts

The receptive field of a network is 128, but the images it is trained on is only 32. What is going on?

Correct!

☒ Large receptive field is making a large template with many orientations/object-variants for better correlation/convolution

Correct!

☒ Large receptive field is making templates which include background for the objects as well

☐ Large Receptive fields add many parameters because of which network can learn better

☐

Question 3

10 / 10 pts

The receptive field of a network is 128, but the images it is trained on is only 32. Can we now train on a similar amount of data or we need to increase it?

☐ Doesn't matter - apunko sab chalta hai

☐ Reduce it

Correct!

☒ Increase it

☐ Same is fine

Question 4

10 / 10 pts

What is residue in Residual Network?

☐ The output just before ReLU

☐ The output

Correct!

☒ he Convolutions we add

☐ X

Question 5

10 / 10 pts

If we add 16 Identity Layers to VGG16, what would happen to the performance?

☐ Will reduce

☐ Will reduce, but can be trained to increase

☐ Will increase

Correct!

☒ Will remain same

Question 6

10 / 10 pts

Select which all are true:

☐ In the Projection Connections in ResNet, we use 3x3 kernels.

☐ (In ResNet) ReLU applied to $H(x)$, where $H(x) = F(x) + x$, can lead to gradient explosion.

☐ If we remove the Addition step in ResNet by Concatenation, accuracy "will" increase.

Correct!

☒ When we do not apply ReLU to $H(x)$, where $H(x) = F(x) + x$ in ResNet, then rarely ever the gradients flowing through the network would be zero.

☐ (In ResNet) BN applied to $H(x)$, where $H(x) = F(x) + x$, can lead to gradient vanishing

Question 7

20 / 20 pts

ResNet 34 has how many Convolutional Layers?

☐ 32

Correct!

☒ 36

☐ 33

☐ 34

Question 8

10 / 10 pts

Does ResNet suffers from Checkerboard issue?

Correct!

☒ True

☐ False

Question 9

10 / 10 pts

You now know the core difference between Inception and ResNet.

Can we add just a skip connection to Inception and convert it to ResNet?

☐ True

Correct!

☒ False

Question 10

10 / 10 pts

Assuming we are comparing ResNet and Inception with same number of layers. Which one will need compute operations?

Correct!

☒ Inception

☐ ResNet

Question 11

10 / 10 pts

When are train ResNet (say 34). If we randomly drop a subset of layers during each training pass (basically bypass with identity function), what would happen?

☐ Would increase Vanishing gradient problem

☐ Will increase the training time

Correct!

☒ Will reduce the training time

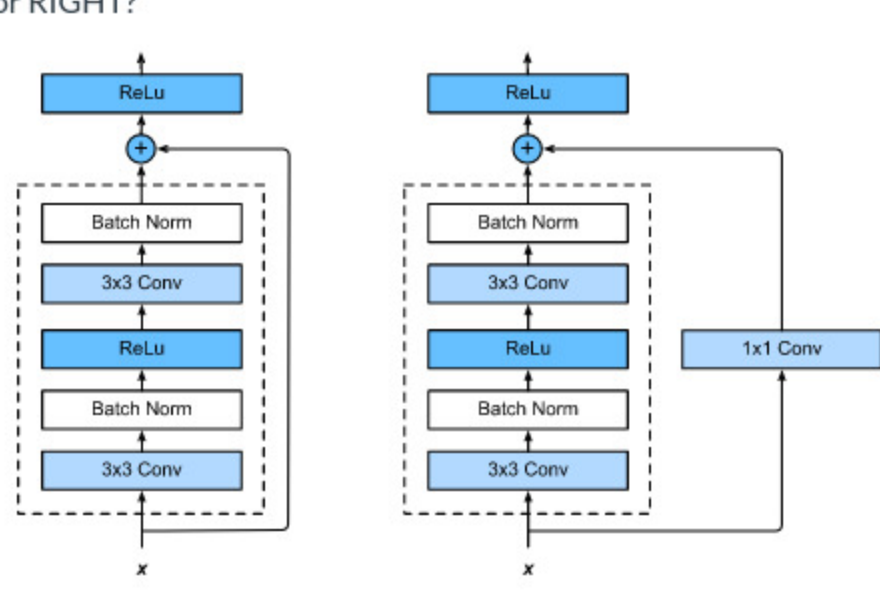
Correct!

☒ Would reduce Vanishing gradient problem further

Question 12

10 / 10 pts

Assuming we have taken care of the channels while addition in the image below, which block would work better, LEFT or RIGHT?



☐ RIGHT

Correct!

☒ LEFT

Question 13

10 / 10 pts

If every Identity mappings in the ResNet50 architecture is multiplied by 0.5, what will happen?

☐ Exploding Gradients

Correct!

☒ Vanishing Gradient

☐ Lesser accurate network, since only half the actual amplitudes of channels can pass through

☐ Nothing

Question 14

10 / 10 pts

Every Identity mapping in the ResNet50 architecture is multiplied by sigmoid (network A) or tanH(network B), which one will definitely show vanishing gradient problems?

☐ Both will show gradient explosion, not vanishing gradients.

☐ Both can manifest vanishing gradient issues

Correct!

☒ Network A

☐ Network B

Question 15

10 / 10 pts

If all the 3x3 stride 2 kernels in ResNet architecture are replaced by 3x3 stride followed by Maxpooling, do you think the accuracy will increase or reduce?

☐ Reduce

☐ It is a secret and I don't want to share it with anyone

Correct!

☒ Increase

☐ Can't Say

Question 16

20 / 20 pts

Select which all are true:

Correct!

☒ increases cardinality is more effective than going deeper or wider when we increase capacity

☐ A 3x3 convolution step with 64 Kernels is replaced by a "root" module with 8 roots. If earlier each kernel was 3x3x32x64, then now it is 3x3x32x8

Correct!

☒ A 3x3 convolution step with 64 Kernels is replaced by a "root" module with 8 roots. If earlier each kernel was 3x3x32x64, then now it is 3x3x4x8

☐ ResNet34 is a special case of ResNeXt

Correct!

☒ A transformation done by a parallel paths of 3x3 and 1x1 (and the concatenated) is a subspace of 3x3 and another 3x3 parallel paths (and the concatenated)

Quiz Score: 200 out of 200