Lecture 4 – ResNets

Provide short answers (max. 25 words each) to the following questions.

- What is the main limitation of very deep feedforward networks?
- Why does adding more layers sometimes reduce performance?
- What is the shattered gradients phenomenon?
- Why do shattered gradients occur?
- What is a residual connection?
- What is the purpose of residual connections?
- How do residual connections affect network structure?
- Why do residual connections help with shattered gradients?
- What is a residual block?
- How many paths exist between input and output in a network with residual blocks?
- Which initialization is recommended for ResNets?
- Why is batch normalization used in ResNets? Provide three reasons.
- What does batch normalization do?
- What parameters does batch normalization learn?
- How is batch normalization applied during testing?
- What is the basic structure of a ResNet block?
- What is a bottleneck residual block?
- Why are bottleneck blocks used in very deep networks?
- What is ResNet200?
- How does DenseNet differ from ResNet?
- What is the advantage of DenseNet over ResNet?
- What is U-Net used for?
- How does U-Net use skip connections?
- Why do residual connections allow deeper networks?
- What is the main benefit of skip connections for gradient flow?
- How do residual networks act like ensembles?
- What role does normalization play in residual networks?
- Why are residual architectures considered a breakthrough?

Multiple-choice questions

What is the main limitation of very deep feedforward networks?

A) They always improve performance
B) They eliminate activation functions
C) They suffer from shattered gradients and optimization issues
D) They require fewer parameters

What happens when too many layers are added to a sequential network?

A) Performance always improves
B) Performance may degrade due to unstable gradients
C) Training becomes faster
D) Loss becomes convex

What is the shattered gradients phenomenon?

A) Gradients vanish completely
B) Gradients explode uncontrollably
C) Gradients change unpredictably in early layers
D) Gradients remain constant

What is a residual connection?

A) A connection that multiplies input by output
B) A skip connection that adds input to output
C) A concatenation of layers
D) A normalization layer

Why are residual connections introduced?

A) To reduce dataset size
B) To ease optimization and improve gradient flow
C) To remove activation functions
D) To increase regularization only

What is a residual block?

A) A block without activation functions
B) A normalization-only layer
C) A fully connected layer
D) A set of layers with a skip connection

How do residual connections affect network structure?

A) They create a single path
B) They remove convolutional layers
C) They create multiple computational paths
D) They eliminate normalization

Why do residual connections help with shattered gradients?

A) They provide direct paths for gradients to flow
B) They increase learning rate
C) They remove batch normalization
D) They reduce dataset size

Which initialization is recommended for ResNets?

A) Xavier initialization
B) Random uniform initialization
C) He initialization
D) Zero initialization

Why is batch normalization used in ResNets?

A) To eliminate skip connections
B) To stabilize training and allow higher learning rates
C) To reduce dataset size
D) To remove activation functions

What does batch normalization do?

A) Normalizes weights
B) Normalizes activations within a batch
C) Normalizes gradients
D) Normalizes loss values

Which parameters does batch normalization learn?

A) Mean and variance only
B) Activation function parameters
C) Learning rate and momentum
D) Scale (γ) and shift (δ)

How is batch normalization applied during testing?

A) Uses batch statistics
B) Removes normalization
C) Uses global mean and variance from training
D) Applies dropout

What is the basic structure of a ResNet block?

A) Convolution → Activation → Skip connection
B) Convolution → BatchNorm → Activation + skip connection
C) Fully connected → Activation → Dropout
D) Activation → Pooling → Skip connection

What is a bottleneck residual block?

A) A block with no skip connections
B) A block with only pooling layers
C) A block using 1×1 convolutions to reduce dimensions

D) A block without normalization

Why are bottleneck blocks used in very deep networks?

A) To increase parameter count
B) To eliminate skip connections
C) To remove activation functions
D) To reduce parameter count and computational cost

What is ResNet200?

A) A network with 200 neurons
B) A residual network with 200 layers
C) A DenseNet variant
D) A U-Net variant

How does DenseNet differ from ResNet?

A) DenseNet adds inputs and outputs
B) DenseNet concatenates inputs and outputs
C) DenseNet removes skip connections
D) DenseNet uses pooling instead of convolution

What is the advantage of DenseNet over ResNet?

A) Better parameter efficiency and feature reuse
B) Fewer skip connections
C) No normalization required
D) Faster training without regularization

What is U-Net used for?

A) Image classification
B) Image segmentation and image-to-image tasks
C) Text generation
D) Speech recognition

How does U-Net use skip connections?

A) Adds encoder features to decoder features
B) Removes encoder features
C) Concatenates encoder features to decoder features
D) Uses pooling layers instead

Why do residual connections allow deeper networks?

A) They increase dataset size
B) They reduce learning rate

C) They remove activation functions
D) They mitigate vanishing gradients and improve optimization

What is the main benefit of skip connections for gradient flow?

A) They slow down training
B) They provide shorter paths for backpropagation
C) They eliminate normalization
D) They increase parameter count

How do residual networks act like ensembles?

A) Each path behaves like a smaller sub-network
B) They use multiple datasets
C) They train multiple models separately
D) They remove convolutional layers

What role does normalization play in residual networks?

A) Improves convergence and acts as implicit regularization
B) Removes skip connections
C) Reduces dataset size
D) Eliminates activation functions

Why are residual architectures considered a breakthrough?

A) They allow training of networks with hundreds or thousands of layers
B) They eliminate the need for GPUs
C) They remove activation functions
D) They reduce dataset size

Which architecture uses concatenation instead of addition for skip connections?

A) ResNet
B) DenseNet
C) U-Net
D) Hourglass