Evolution of CNN Architectures: Interview Summary

1. AlexNet (2012)

AlexNet was the first deep CNN that significantly outperformed traditional machine learning methods on the ImageNet challenge.

- 8 layers (5 Conv + 3 FC)
- Introduced ReLU, Dropout, and overlapping pooling
- Used 2 GPUs for training

Key contribution: Deep learning scalability with better regularization and faster training.

2. ZFNet (2013)

ZFNet improved upon AlexNet by tuning filter sizes and strides using feature map visualizations.

- Reduced first filter size (11x11 -> 7x7)
- Decreased stride for better feature capture
- Introduced deconv visualization for interpretability

Key contribution: Fine-tuning and interpretability of CNNs.

3. VGGNet (2014)

VGG used a very consistent architecture based only on 3x3 convolutions.

- VGG-16 and VGG-19 are the most popular variants
- Demonstrated that deeper models improve performance

Key contribution: Uniform architecture with small filters and deep stacks.

4. GoogLeNet / Inception v1 (2014)

GoogLeNet introduced the Inception module that applies multiple filters in parallel.

- Combines 1x1, 3x3, 5x5 convs + pooling
- Uses 1x1 convs to reduce dimensionality
- Removes fully connected layers (uses Global Average Pooling)

Key contribution: Multi-scale feature extraction and efficient parameter usage.

5. ResNet (2015)

Evolution of CNN Architectures: Interview Summary

ResNet solved the vanishing gradient problem in deep networks using residual (skip) connections.

- Residual Block: F(x) + x
- Enabled networks as deep as 152+ layers

Key contribution: Training of very deep networks reliably with skip connections.

Summary

Each architecture built upon the last:

- AlexNet proved depth worked
- ZFNet refined feature extraction
- VGGNet deepened and standardized the structure
- GoogLeNet optimized width and computation
- ResNet solved training for very deep networks

Together, they shaped modern deep learning in computer vision.