AlexNet: Revolutionizing Deep Learning in Computer Vision

Prepared by: Mohit Raj Aryal
AI Intern, Amnil Technologies
October 2025

Introduction to CNNs

CNNs are deep learning models for visual data.

Before AlexNet, models were shallow and limited by compute.

Background and Motivation

Developed to tackle the ImageNet Challenge (2012).

Needed deeper architectures and GPU support.

Overview of AlexNet

Created by Alex Krizhevsky, Ilya Sutskever, Geoffrey Hinton (2012).

Achieved top-5 error rate of ~15%, a huge improvement.

Architecture Overview

8 layers: 5 convolutional, 3 fully connected.

Flow: Conv → ReLU → Pool → Conv → ReLU → Pool → FC → Softmax

Detailed Layer Breakdown

Conv layers extract features using kernels.

ReLU introduces non-linearity.

Pooling reduces spatial dimensions.

FC layers perform classification.

Key Innovations

ReLU activation for faster training.

Dropout to prevent overfitting.

Data augmentation to expand training set.

GPU parallelism for large-scale training.

Training Details

Dataset: ImageNet (1.2M images, 1000 classes).

Optimizer: SGD with momentum.

Batch size ≈128, learning rate 0.01, 90 epochs.

Performance & Results

Reduced Top-5 error rate from 26% to 15% on ImageNet 2012.

Huge improvement over all prior models.

Impact on Computer Vision

Popularized deep learning for image tasks.

Enabled breakthroughs in object detection and recognition.

Inspired modern architectures like VGG and ResNet.

Limitations

Large model and high computational cost.

Overfitting on smaller datasets.

Later replaced by more efficient models.

Legacy & Influence

Foundation for modern CNN architectures.

Marked the start of the deep learning revolution.

Conclusion

AlexNet was a turning point in AI history.

It proved that deep learning could outperform traditional methods.