AlexNet (2012) — Report

Paper: ImageNet Classification with Deep Convolutional Neural Networks — Alex Krizhevsky, Ilya Sutskever, Geoffrey Hinton (NIPS 2012).

1. Executive summary

AlexNet demonstrated that a deep convolutional neural network (CNN), trained on large labeled data (ImageNet) and accelerated by GPUs, could dramatically outperform prior methods on large-scale object recognition. The model's combination of architectural scale, GPU implementation, and several training/regularization innovations (ReLU, dropout, data augmentation, local response normalization, overlapping pooling) produced a large improvement in ImageNet performance and triggered the deep-learning revolution in computer vision.

2. Dataset & goal

Trained on the ImageNet ILSVRC subset: ~1.2 million training images, 50k validation, 150k test images across 1000 classes. The paper focused experiments on the ILSVRC-2010 and also reported results for the 2012 competition.

3. Architecture (high level)

- 8 learned layers — 5 convolutional + 3 fully connected, with ReLU after every layer.
- Model split across two GPUs to fit memory and speed training.
- Input images: 224x224 crops from resized 256x256 images.

4. Key innovations & practical techniques

1. ReLU activations (faster training than tanh/sigmoid).
2. GPU implementation with multi-GPU parallelism.
3. Local Response Normalization (LRN).
4. Overlapping max-pooling.
5. Dropout in fully connected layers.
6. Data augmentation: random crops, flips, PCA color augmentation.

5. Training setup and hyperparameters

- Hardware: Two NVIDIA GTX 580 GPUs, training time ≈ 5–6 days.
- Optimization: SGD with momentum (0.9), weight decay 0.0005, batch size 128.
- Learning rate: 0.01, reduced ×10 when validation error plateaued.
- Initialization: Gaussian weights (mean 0, std 0.01); some biases set to 1.

6. Size / capacity

AlexNet has ~60 million parameters. Regularization and data augmentation were essential to prevent overfitting.

7. Results reported

- ILSVRC-2010: top-1 error 37.5%, top-5 error ~17–18.9%.
- ILSVRC-2012: ensemble reduced top-5 error to ~15.3%, a huge margin over prior state of the art.

8. Impact & legacy

AlexNet proved deep CNNs trained with GPUs and large datasets could outperform hand-engineered features. It triggered the modern deep learning revolution, inspiring VGG, GoogLeNet, ResNet, etc.

9. Limitations and criticisms

- Some components (LRN, GPU-splitting) were pragmatic and later dropped.
- Competition-winning system relied on ensembles, not just a single model.
- Hardware constraints shaped unusual design choices.
- Today, AlexNet is mainly educational/historical.

10. Practical notes for reproducing

- Use 224x224 random crops, flips, PCA color augmentation.
- Dropout in FC layers, ReLU everywhere, SGD with momentum.
- Many modern frameworks (PyTorch, TensorFlow) have AlexNet variants (sometimes with 227x227 input).

11. Short annotated reading list

- Original paper (NIPS 2012): Krizhevsky, Sutskever & Hinton.
- ACM reprint (2017).
- Wikipedia: AlexNet overview & impact.

12. Model summary

AlexNet (2012) — 5 conv + 3 FC layers, ~60M parameters, introduced ReLU, dropout, data augmentation, LRN, overlapping pooling; trained on ~1.2M ImageNet images on 2 GPUs in ~6 days; reduced ImageNet error dramatically and started deep learning revolution in vision.