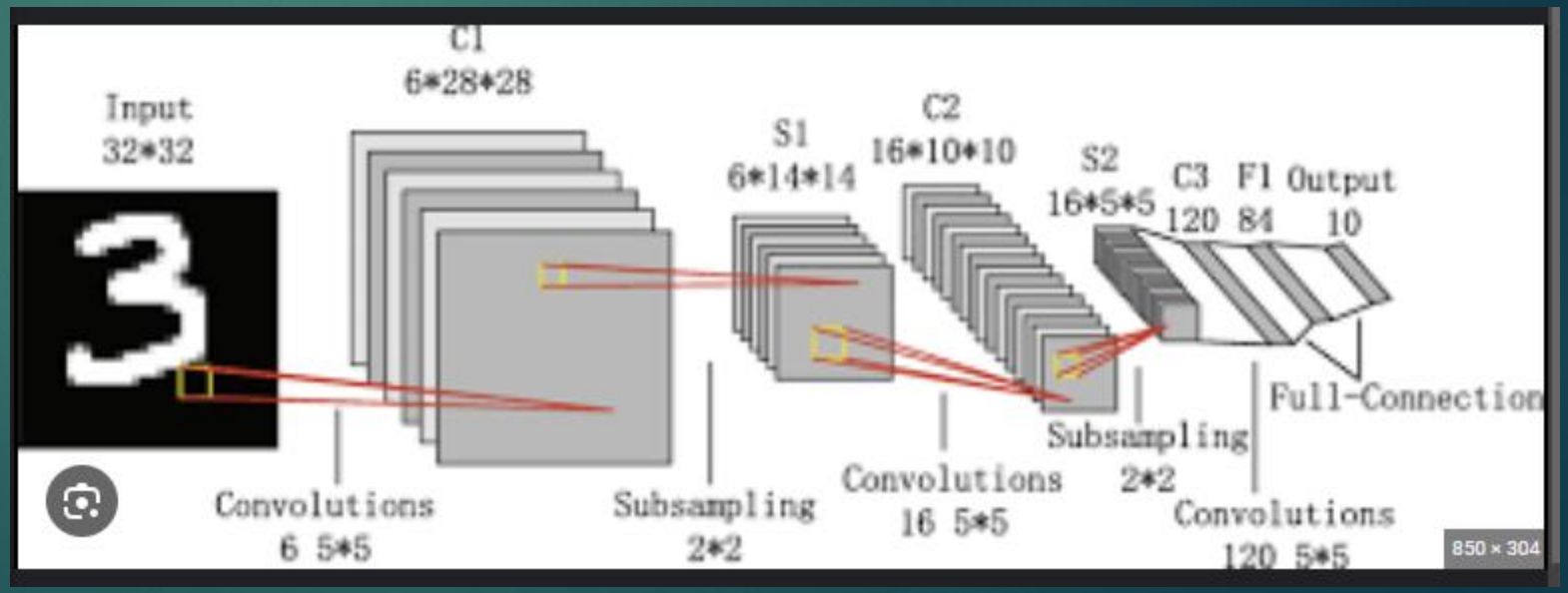
ImageNet Classification with Deep Convolutional Neural Networks

BY VIJAY VENKATESAN

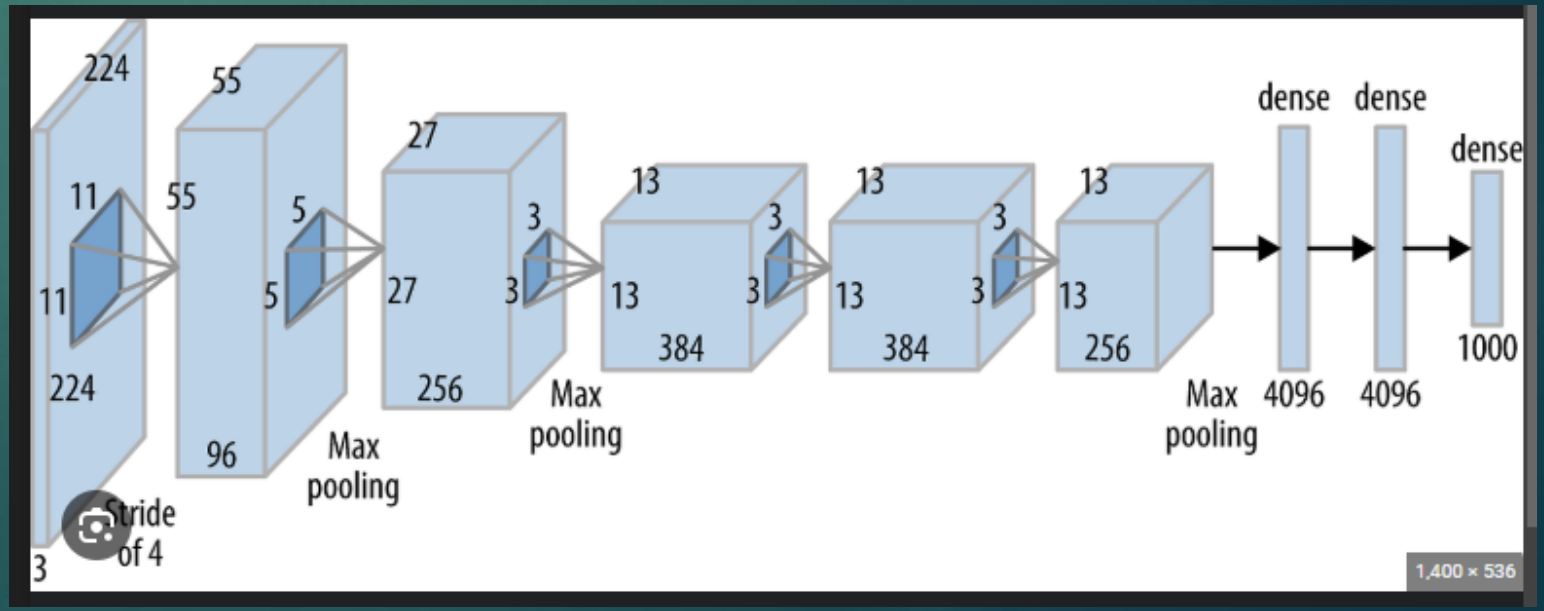
LeNet-5

- Developed by Yann LeCun
- Earliest CNNs
- Handwritten digits
- Disadvantages
  - Shallow architecture
  - Complex images
  - Overfitting
  - Color images
  - tanh



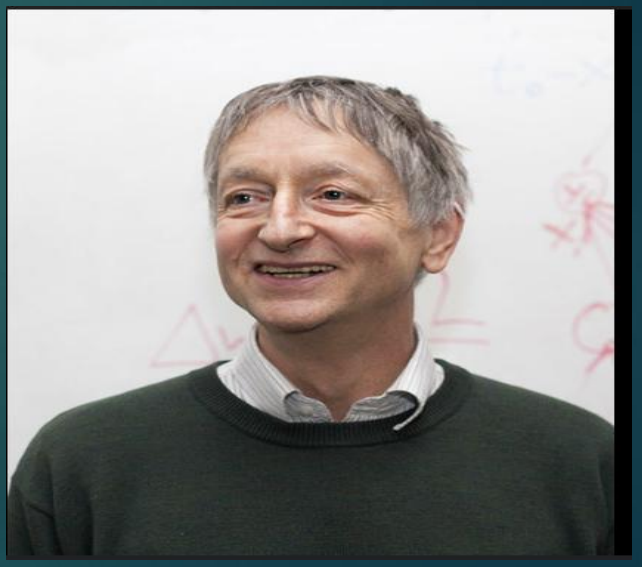
AlexNet

- Deeper Architecture
- Trained on ImageNet
- ILSVRC-2012
- Top-1 and Top-5 Error Rates



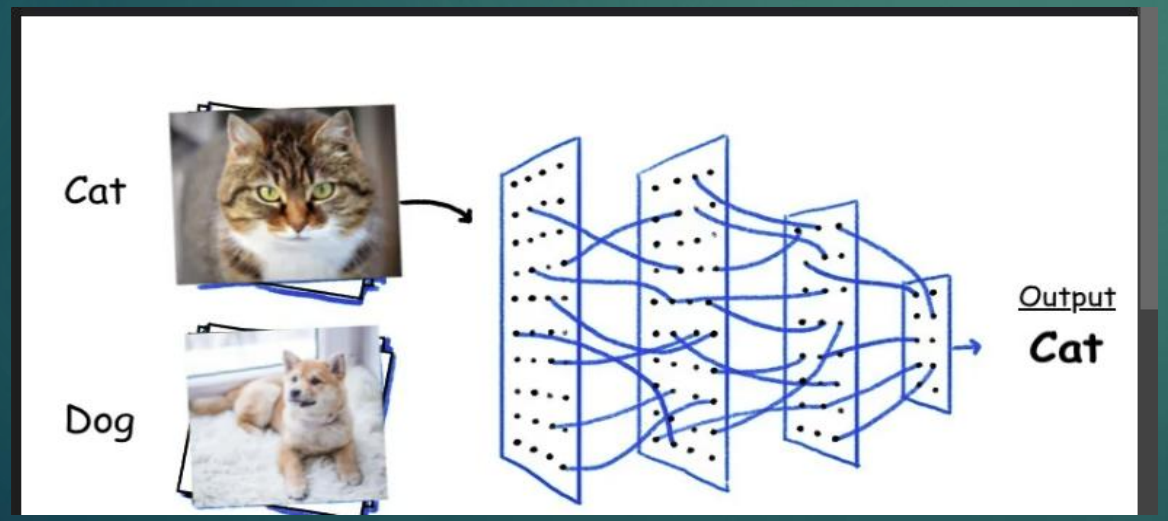
Masterminds behind AlexNet

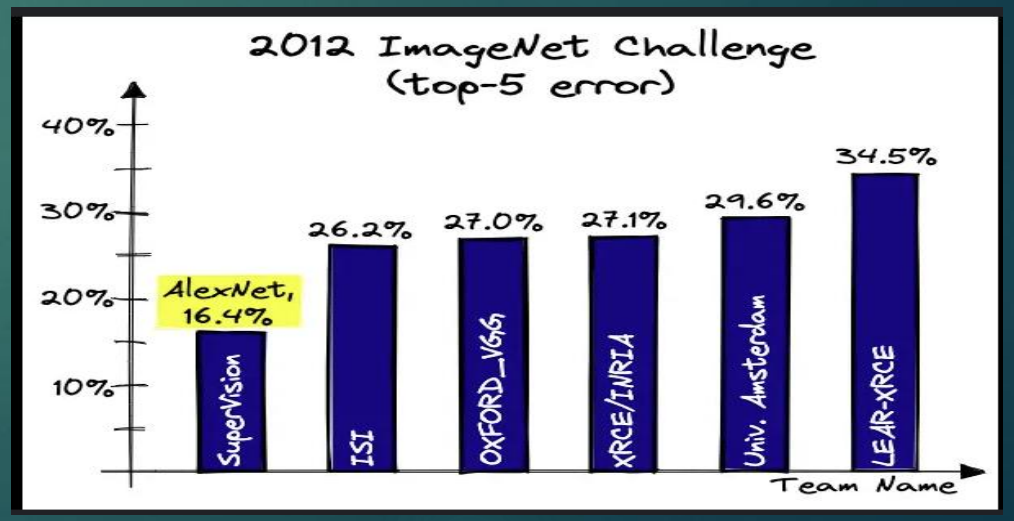
- Alex Krizhevsky
- Ilya Sutskever
- Geoffrey E Hinton (Godfather of A.I.)
- University of Toronto



Abstract

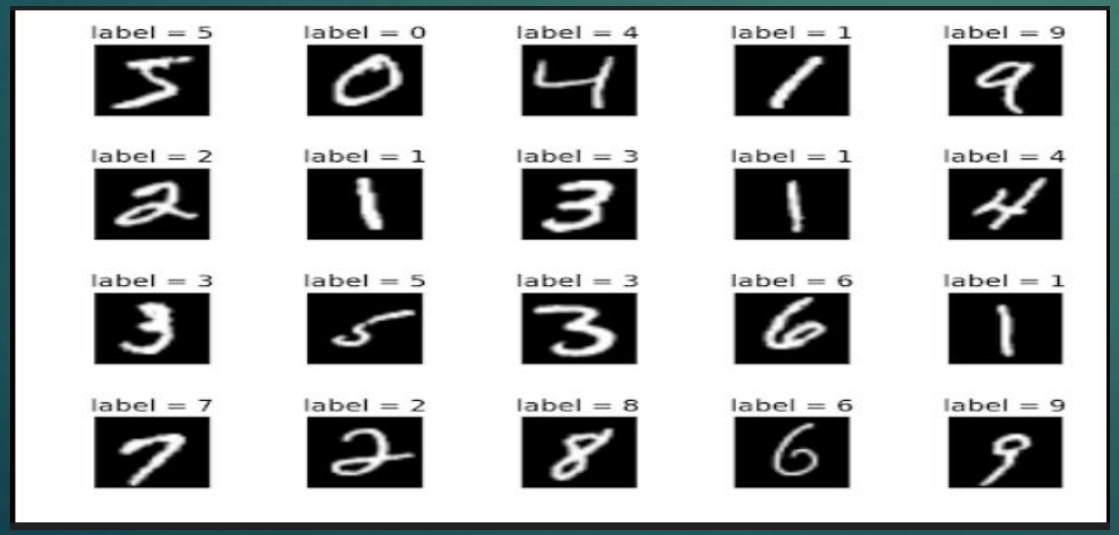
- Image Classification (1.2M)
- Large Parameters (in 2012)
- ILSVRC-2010, ILSVRC-2012
- Top-1 and Top-5 error rates





Introduction

- Smaller datasets
- Simpler Tasks
  - MNIST dataset
- Various Image Types
- Large Learning Capacity

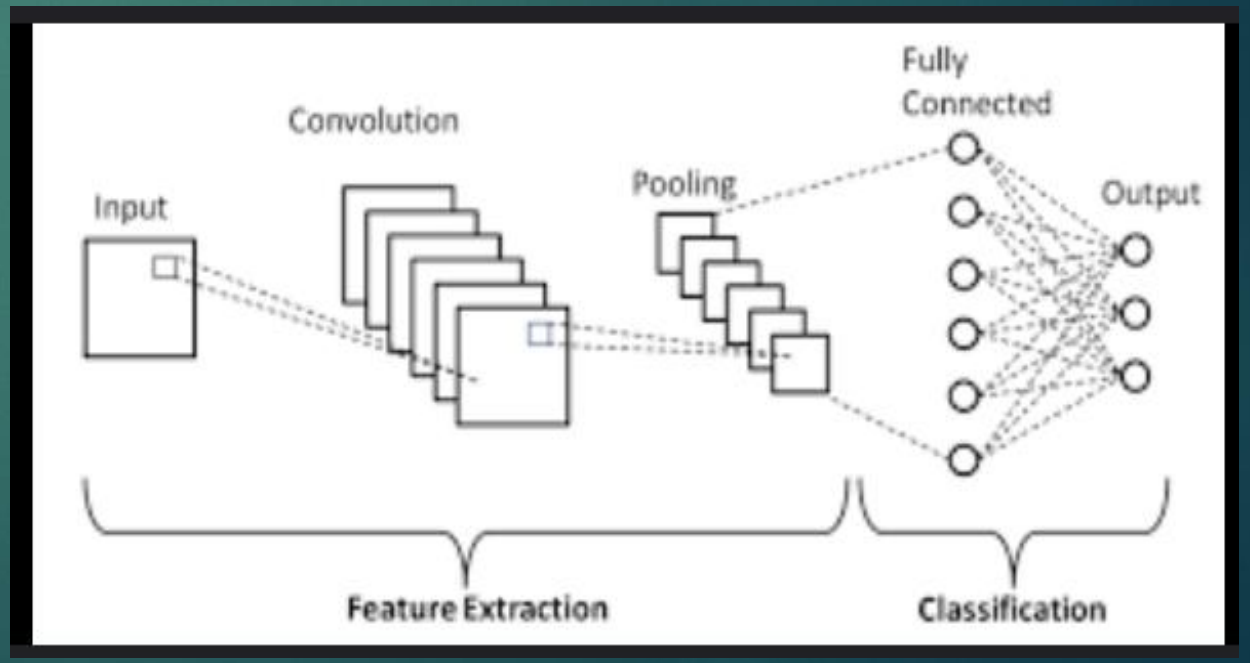


The ImageNet Dataset

- 15M images
- Vast collection of images
- Standard benchmark
- Train/Validation/Test
- ILSVRC

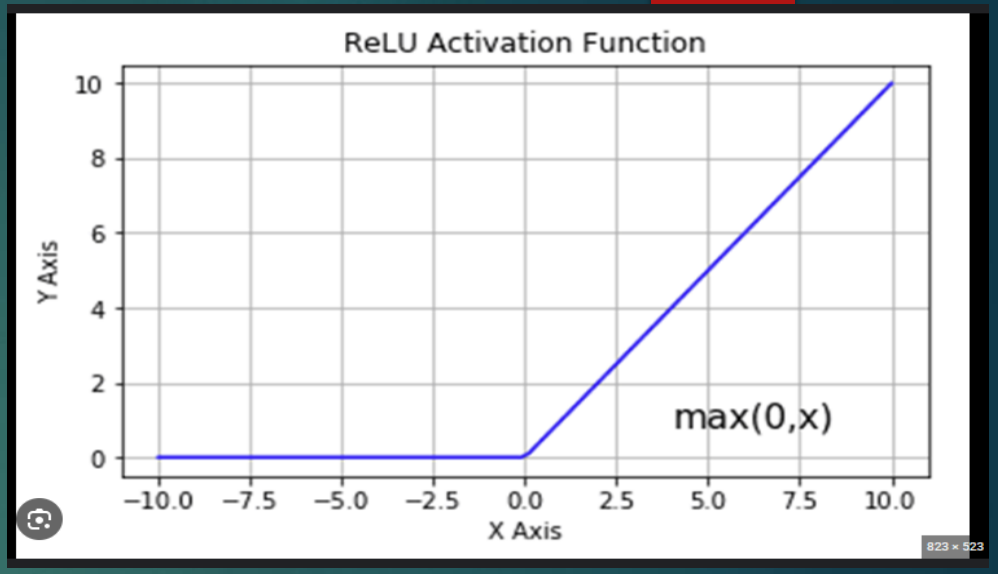
The Architecture

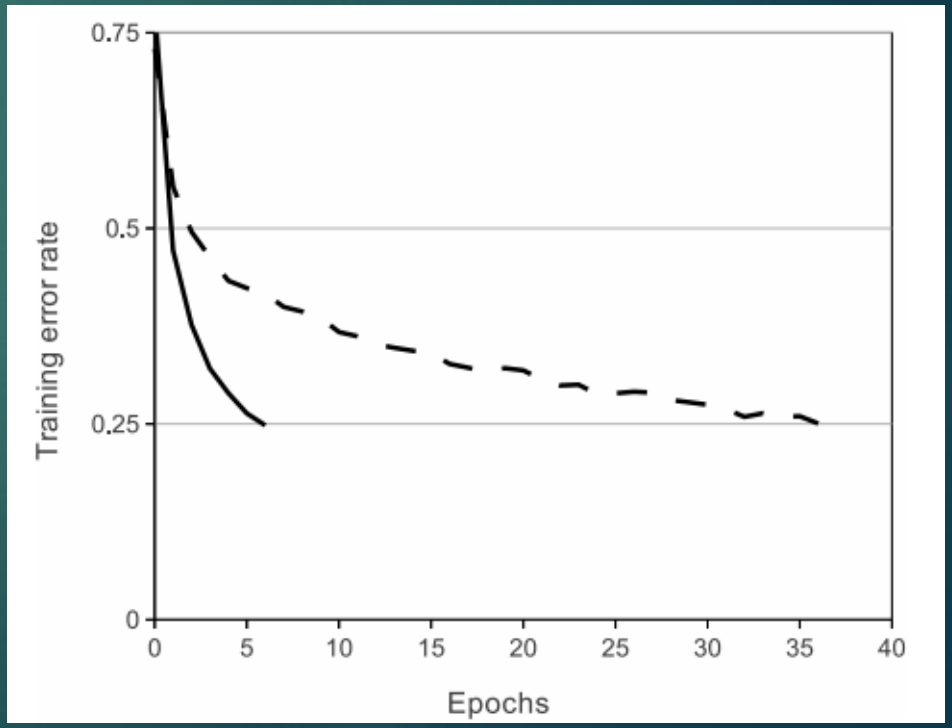
- Known So Far
  - Eight layers
  - 1000 neurons (Output layer)
  - Softmax Activation Function
- Yet To Be Discussed
  - ReLU Nonlinearity
  - Training on Multiple GPUs
  - Local Response Normalization
  - Overlapping Pooling
  - Convolution Optimizations



ReLU Nonlinearity

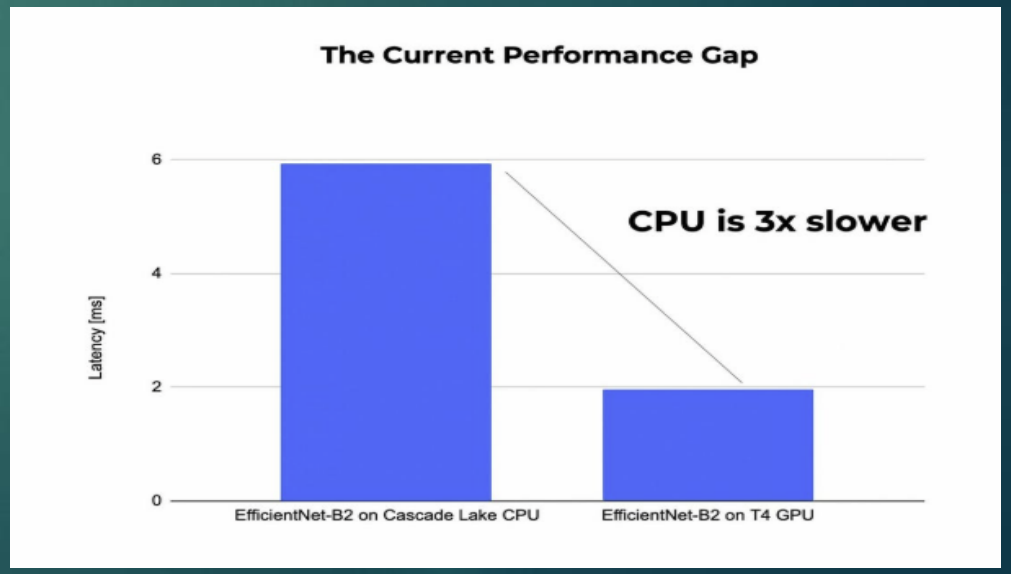
- f(x) = max(0,x)
- Adds nonlinearity
- Faster model training
- Speed Advantage
  - Simpler equation
  - No vanishing gradient issue





Training on Multiple GPUs

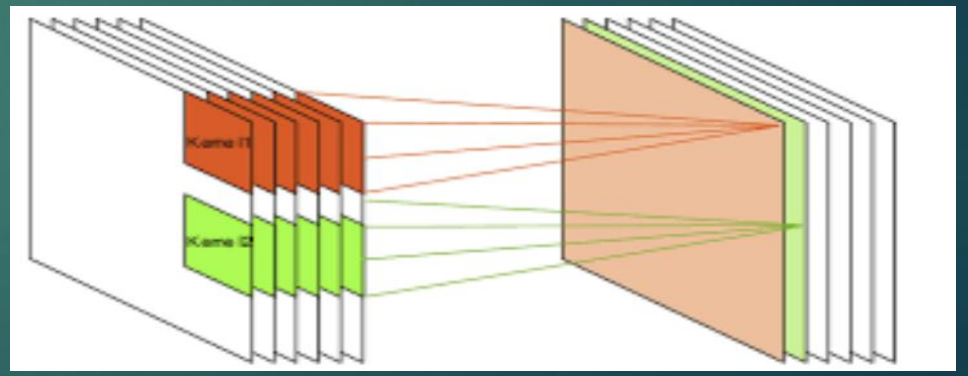
- Two GTX 580 GPUs
- Reduced Training Time
- Cross-GPU Parallelization
- Reduced Error Rates



Local Response Normalization

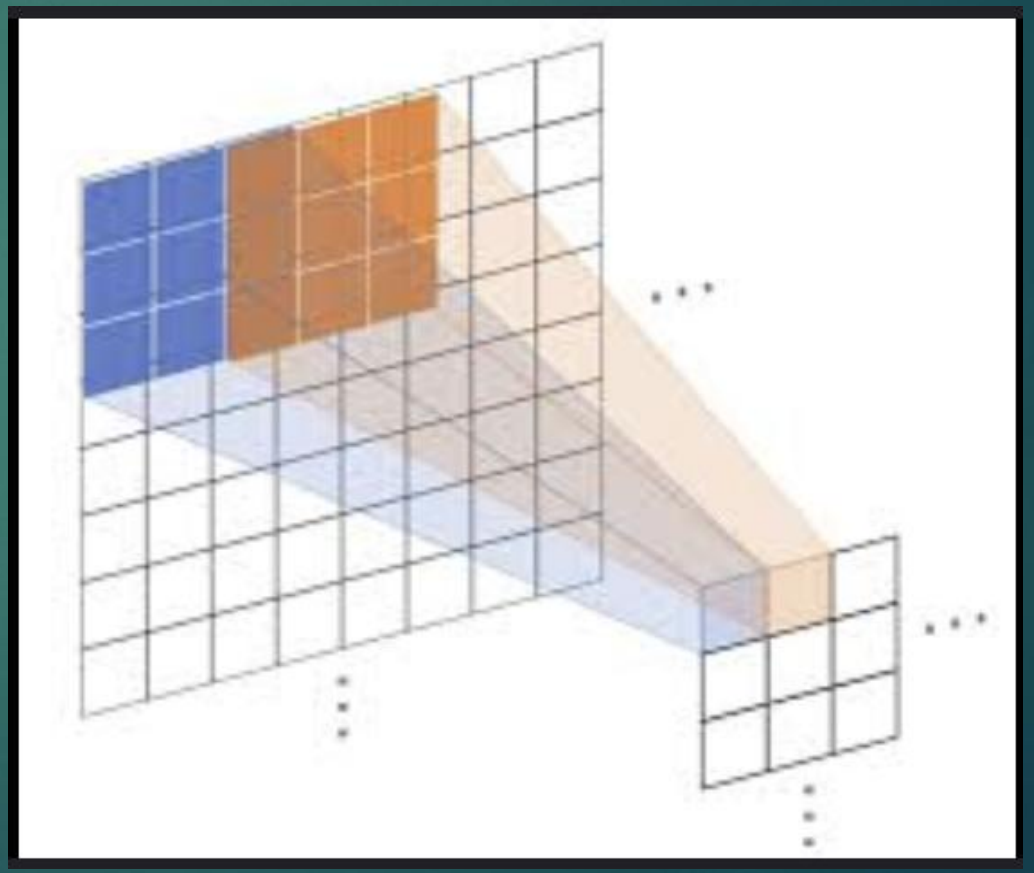
- Scales feature map elements
- Utilizes neighboring feature maps
- Improves generalization and learning
- Reduced error rates

$$b^i_{x,y} = a^i_{x,y} \Big/ \left(k + \alpha \sum_{j=\max(0,i-n/2)}^{\min(N-1,i+n/2)} (a^j_{x,y})^2 \right)^{\beta}$$



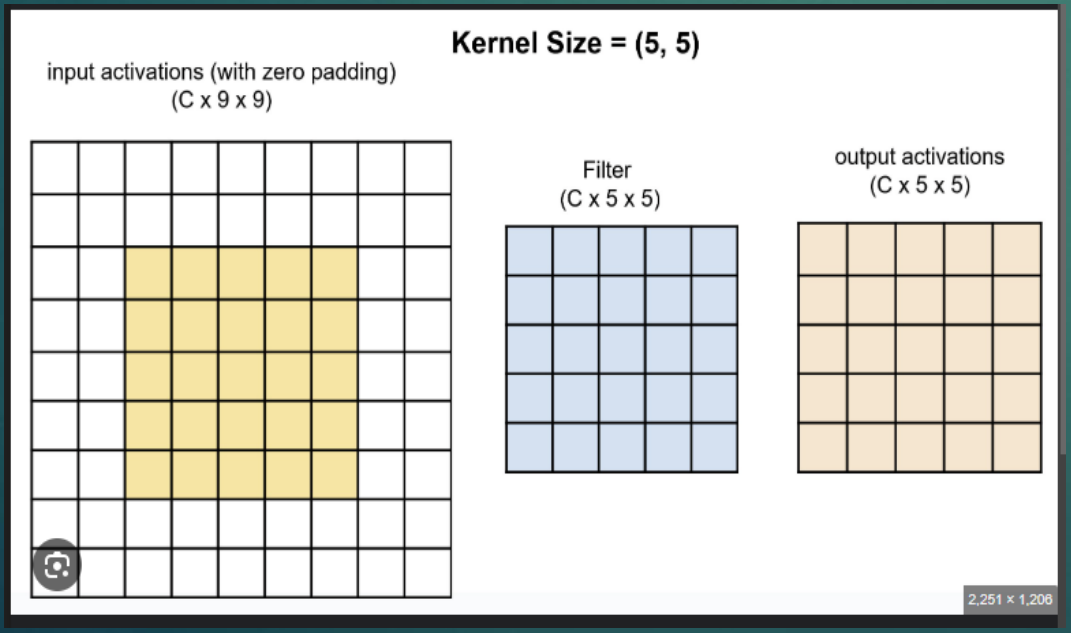
Overlapping Pooling

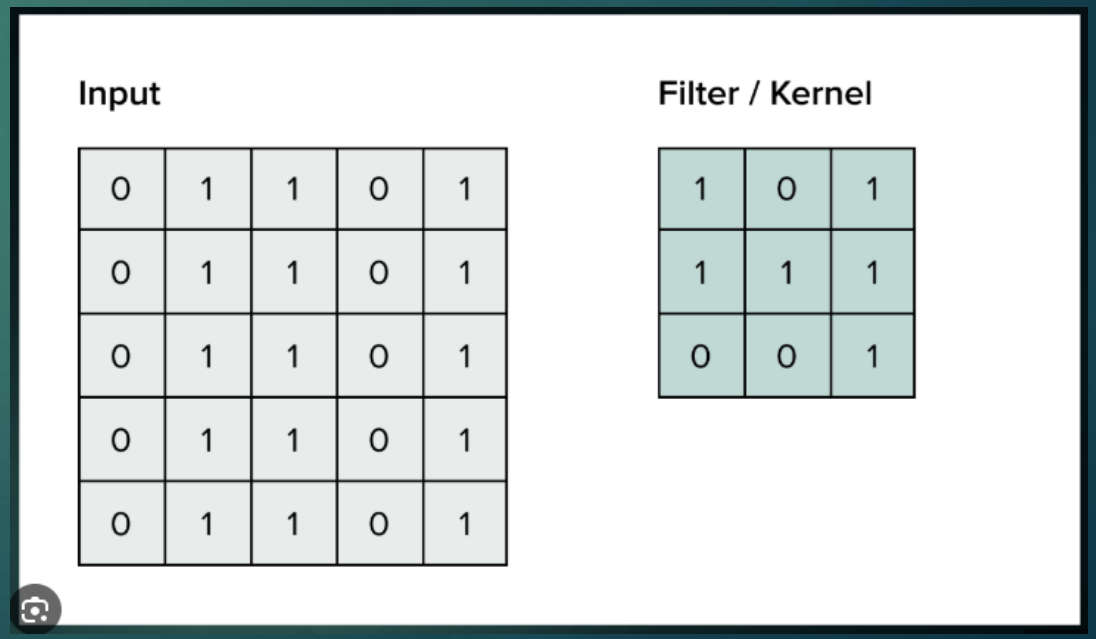
- Overlapping windows
  - s -> window size
  - z -> region size
- Reduced error rates
- Less prone to overfitting



Convolution Optimizations

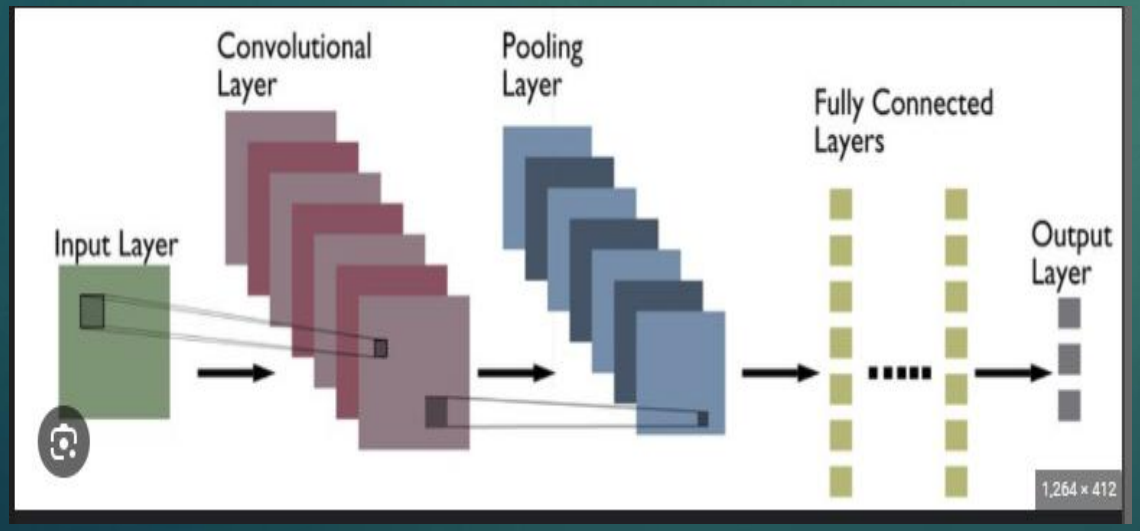
- Larger kernel sizes
  - Broad patterns
- Smaller kernel sizes
  - Fine-grained details





Overall Architecture

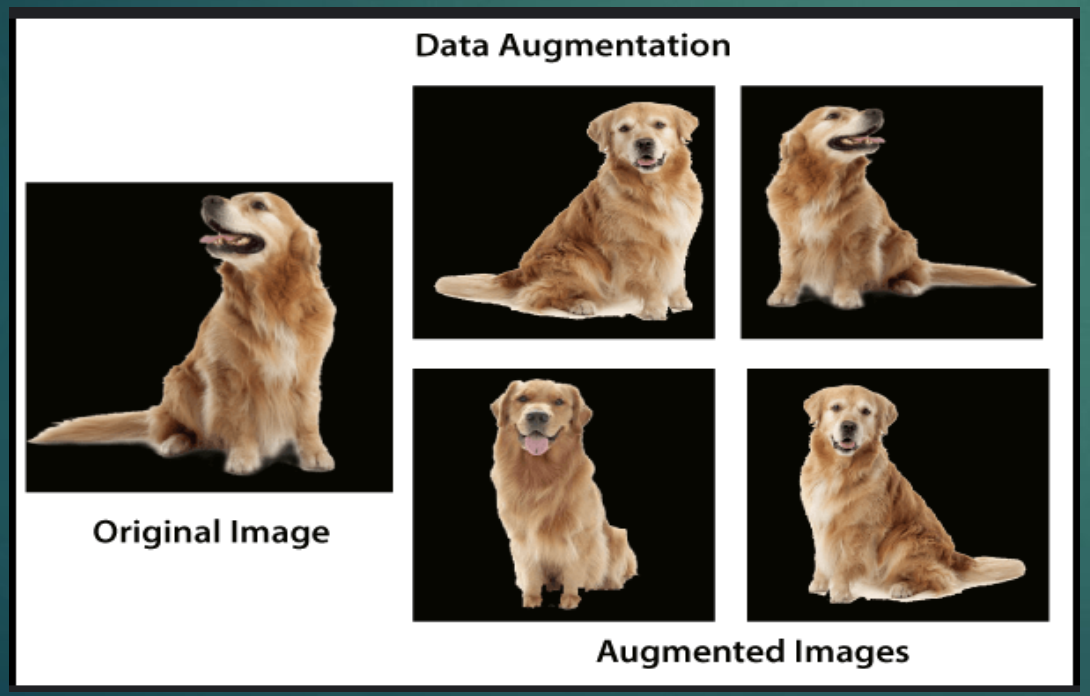
- Eight layers
  - Five convolutional layers
  - Three fully connected layers
- Subsequent convolutional layers
- Softmax activation function

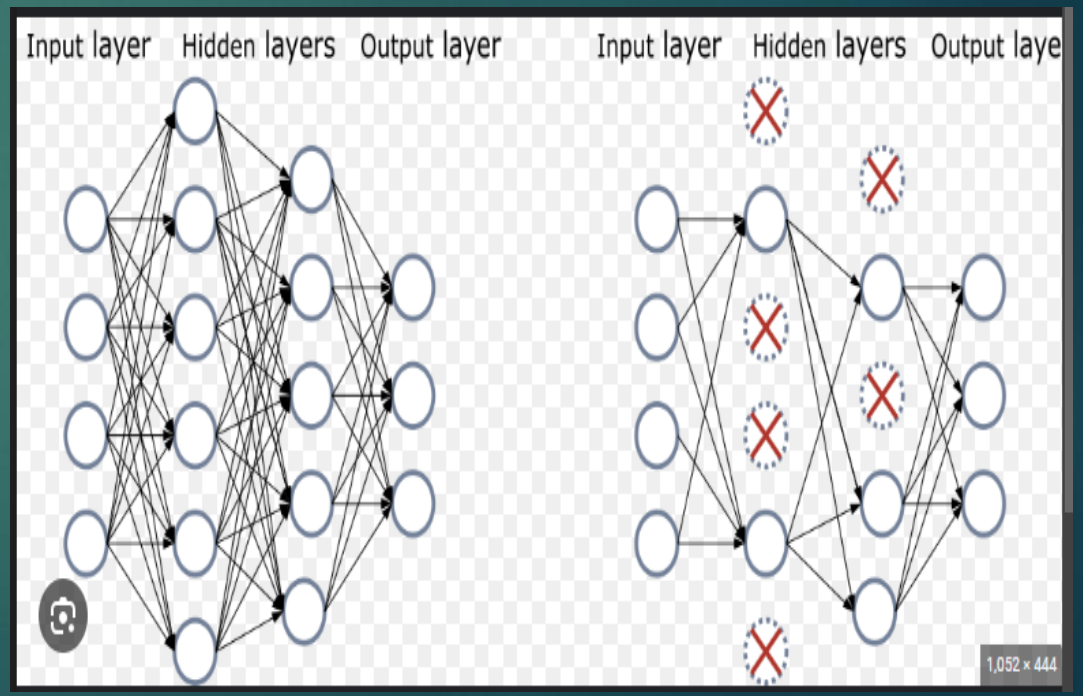


$$s\left(x_i\right) = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}}$$

How to Combat Overfitting

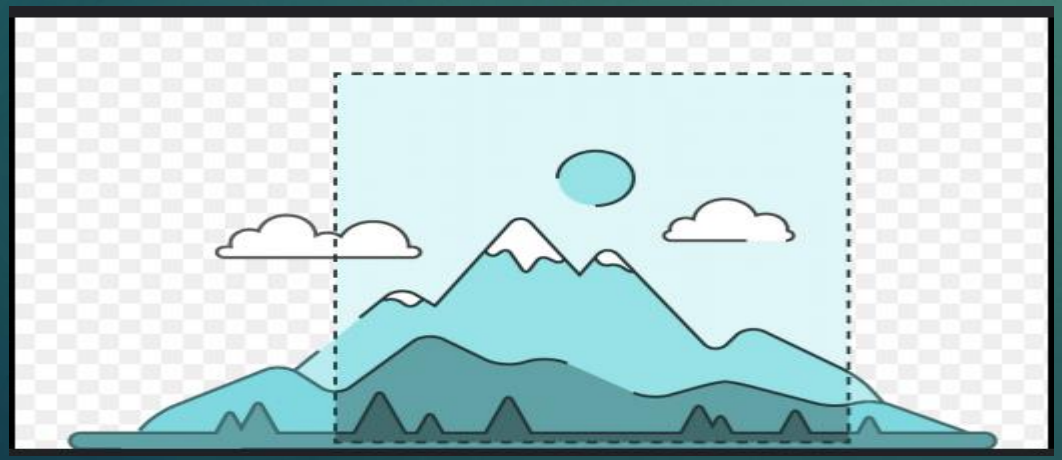
- Data Augmentation
- Dropout

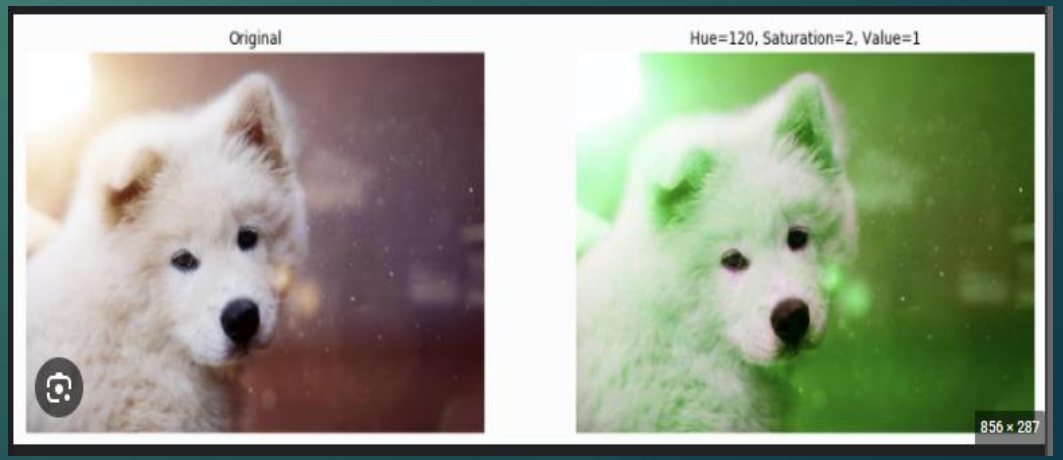




Data Augmentation

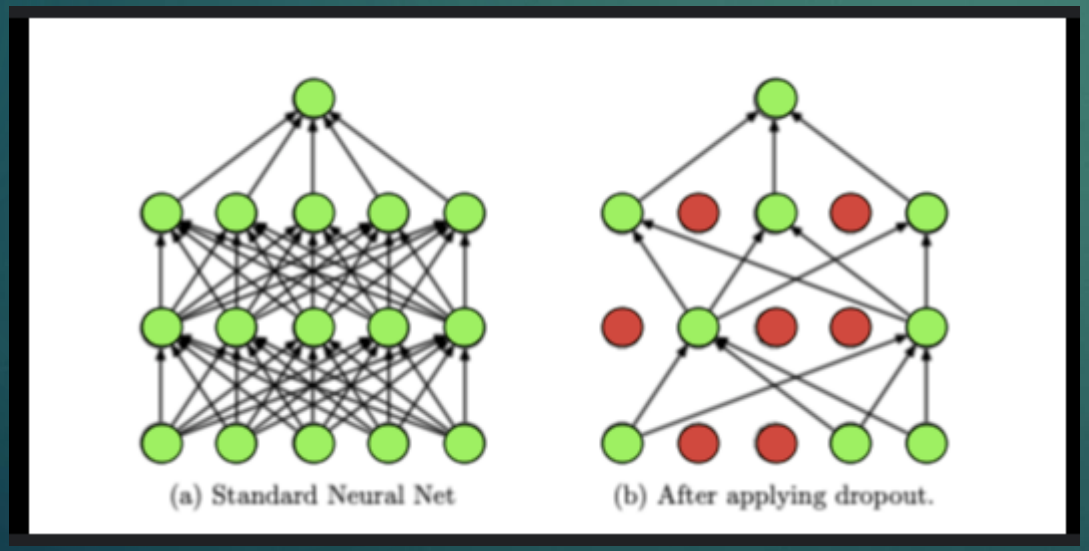
- Random Cropping
  - Translation Invariance
  - Improve Generalization
- PCA-based Color Augmentation
  - Color variations
  - Light variations

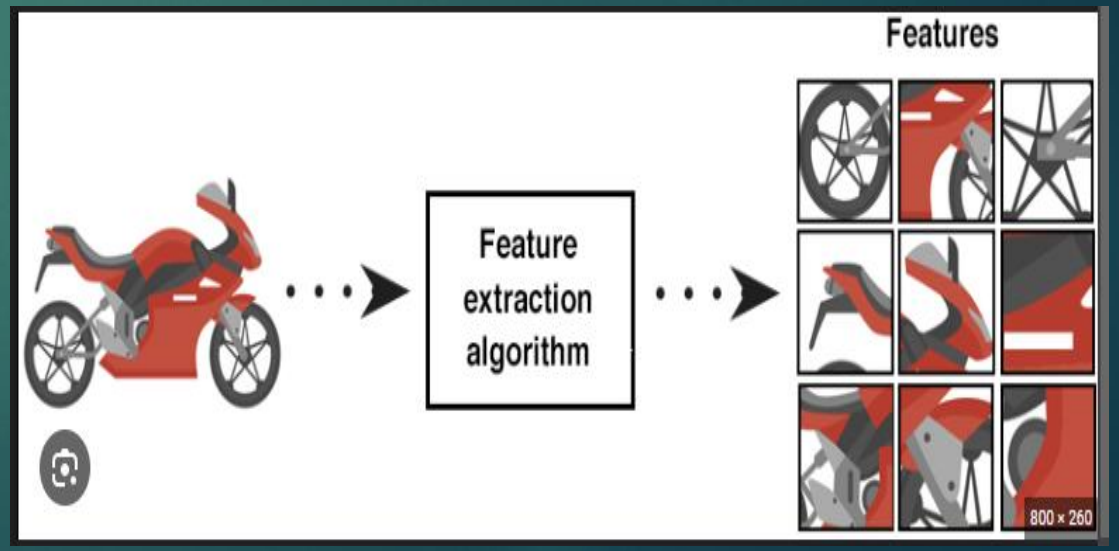




Dropout

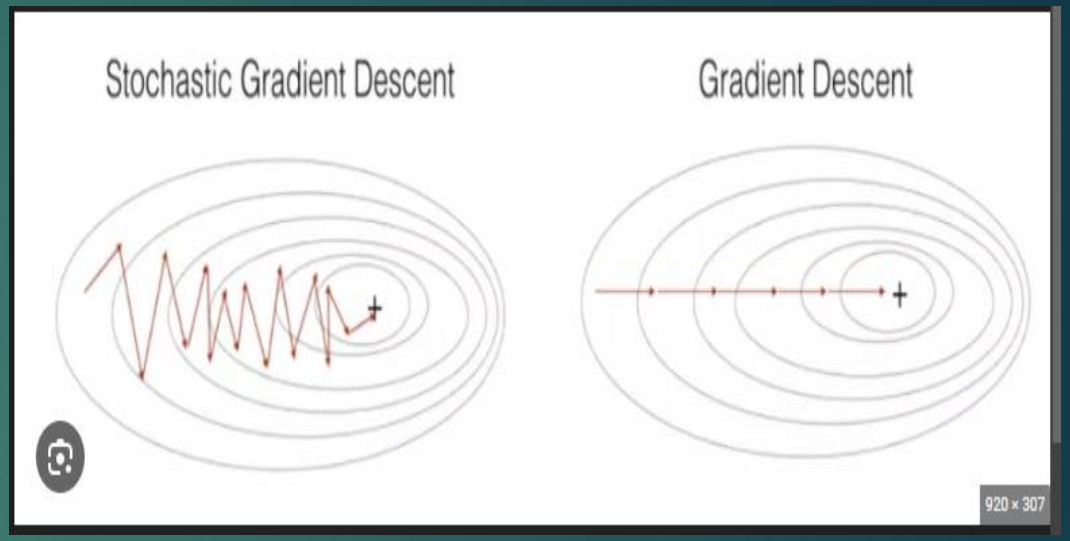
- Set to zero (Neurons)
- Less Dependent
- Model learns robust features
- Invented by Geoffrey Hinton

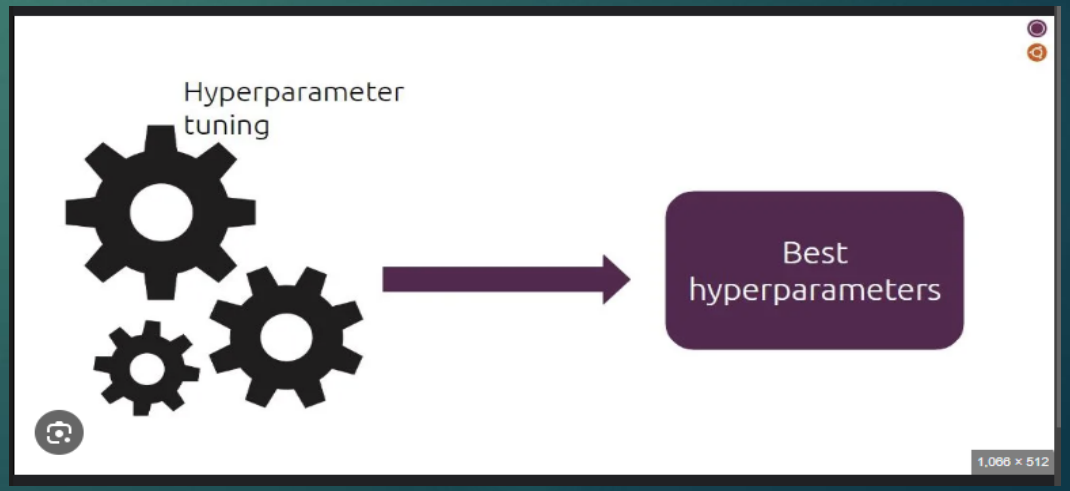




Details of learning

- Stochastic Gradient Descent (SGD)
- Key Hyperparameters
  - Learning Rate = 0.00001
  - Batch Size = 128
  - Momentum = 0.9
  - Weight Decay = 0.0005
  - Epochs = 90





Results

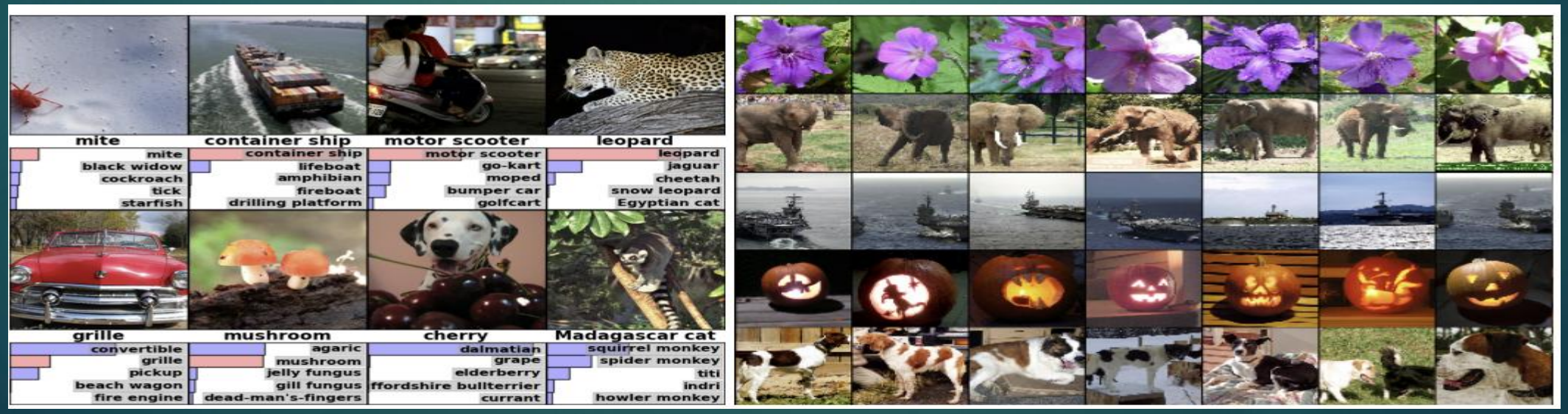
- ILSVRC-2010
  - Top-1, Top-5 Error Rates
- ILSVRC-2012
  - Top-5 Error Rate
  - Large Margin

Model	Top-1	Top-5
Sparse coding [2]	*47.1%*	*28.2%*
SIFT + FVs [24]	*45.7%*	*25.7%*
CNN	**37.5%**	**17.0%**

Model	Top-1 (val)	Top-5 (val)	Top-5 (test)
SIFT + FVs [7]	—	—	*26.2%*
1 CNN	40.7%	18.2%	—
5 CNNs	38.1%	16.4%	**16.4%**
1 CNN*	39.0%	16.6%	—
7 CNNs*	36.7%	15.4%	**15.3%**

Qualitative Evaluations

- LHS
  - Eight Test Images (ILSVRC-2010)
- RHS
  - Five Test Images (ILSVRC-2010)



Discussion and Takeaways

- Discussion
  - Convolutional layers are important
  - Expand into videos
- Takeaways
  - Groundbreaking technology (2012)
  - Deep Learning
  - Limitless Possibilities

References

- R.M. Bell and Y. Koren. Lessons from the Netflix Prize challenge. *ACM SIGKDD Explorations Newsletter*, **9**(2):75–79, 2007.
- A. Berg, J. Deng, and L. Fei-Fei. Large scale visual recognition challenge 2010. www.image-net.org/challenges, 2010.
- L. Breiman. Random forests. *Machine Learning*, **45**(1):5–32, 2001.
- D. Cireşan, U. Meier, and J. Schmidhuber. Multi-column deep neural networks for image classification. *Arxiv preprint arXiv:1202.2745*, 2012.
- D.C. Cireşan, U. Meier, J. Masci, L.M. Gambardella, and J. Schmidhuber. High-performance neural networks for visual object classification. *Arxiv preprint arXiv:1102.0183*, 2011.


Questions


Thanks for
Listening