

ImageNet Classification with Deep Convolutional Neural Networks

Ha-larm

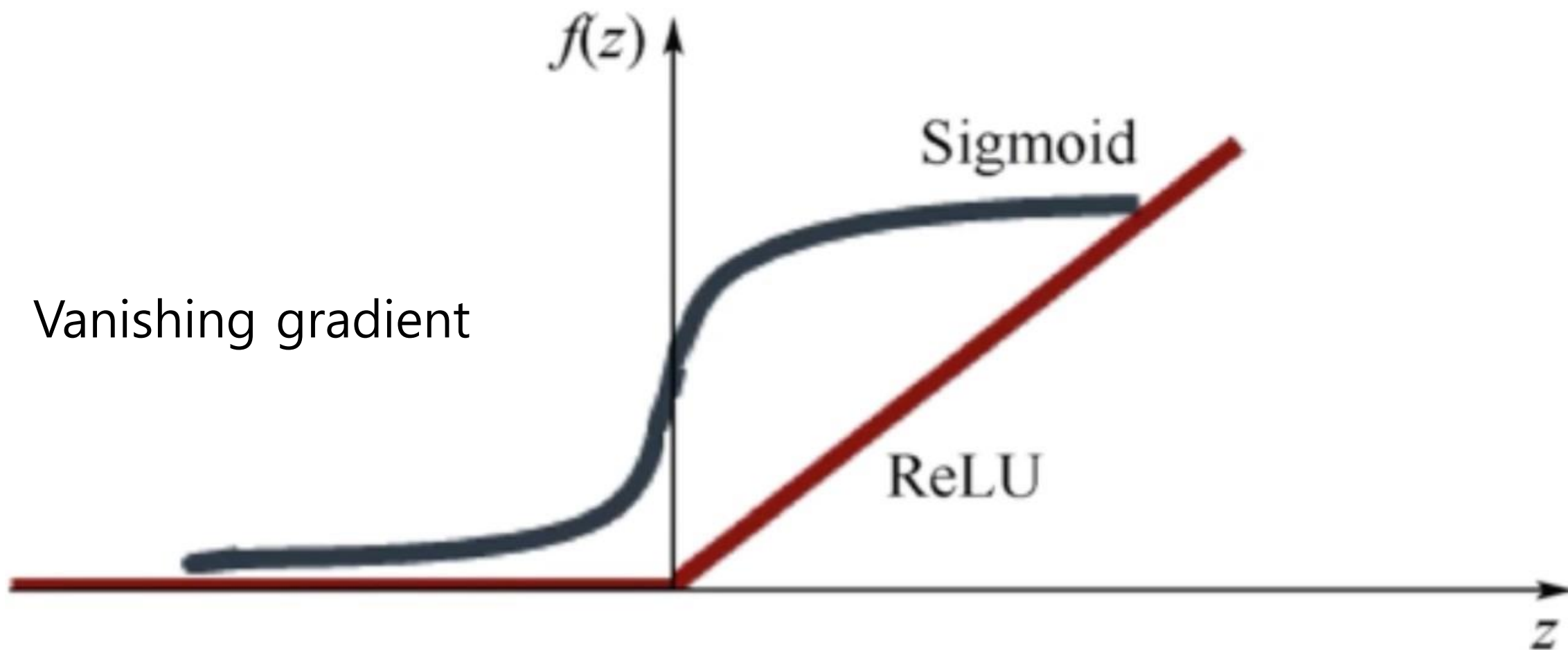
Abstract

- Conv5 + FC3 = 7 Layer Neural Network
- ImageNet이라는 대규모 데이터 셋에 대해 학습하여 성능향상
- 더 빠른 학습을 위하여 비 포화 뉴런과 GPU사용
- Overfitting 현상을 막기위해 'dropout'기법 사용
- 압도적인 정확도

Introduction

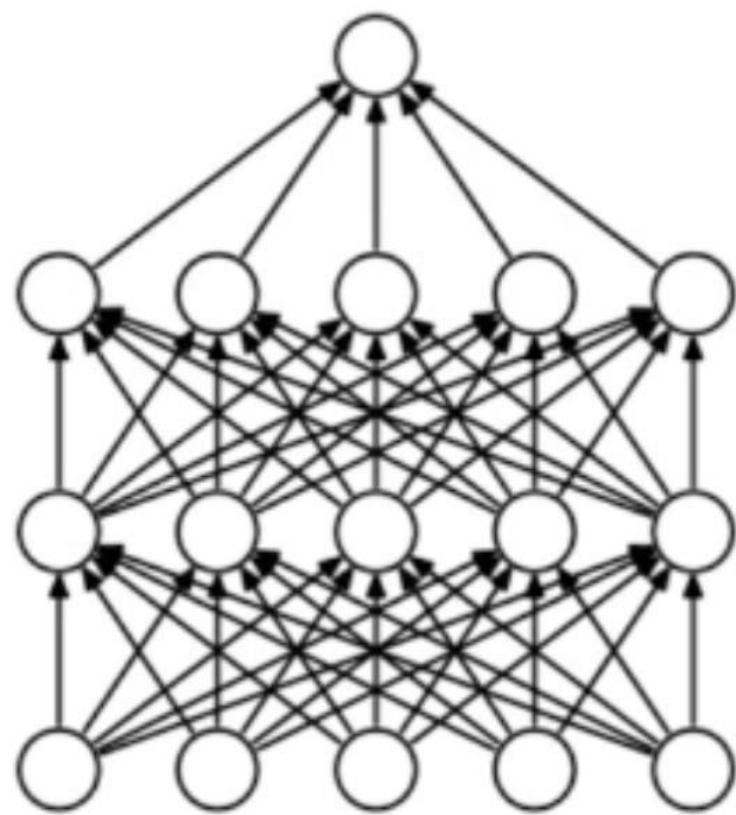
- Relu Activation을 사용
- FC1, FC2에는 0.5 비율의 Dropout을 추가
- LRN(Locally Response-Normalization)를 1,2,5 Conv Layer의 Pooling 직전에 삽입->Generalization에 도움 0.2%정도의 성능 향상
- 1,2,5 Conv Layer의 뒤에 Overlapping Pooling을 사용
->Overfitting 방지

Sigmoid!

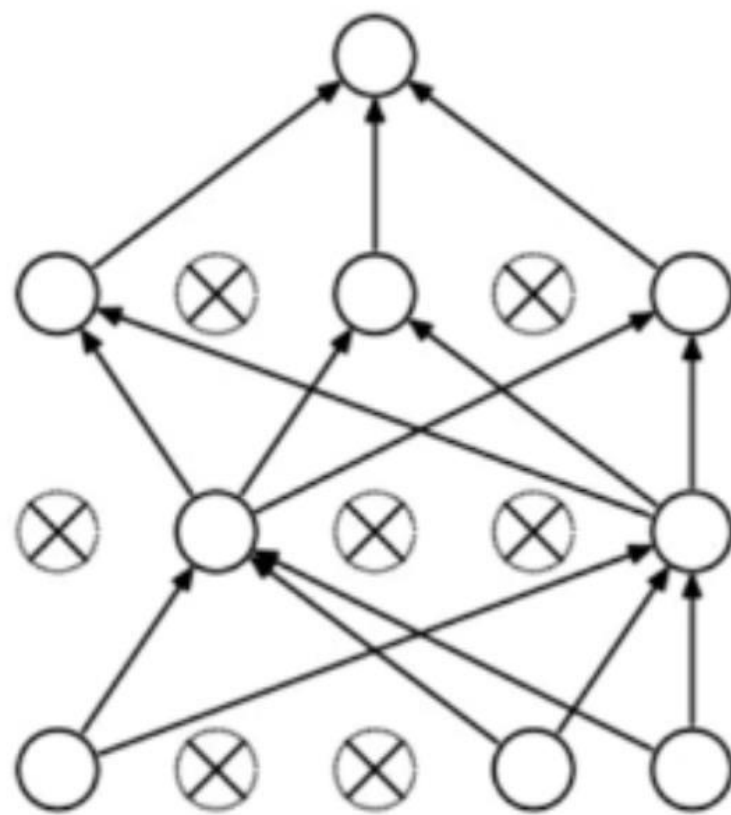


Regularization: **Dropout**

“randomly set some neurons to zero in the forward pass”



(a) Standard Neural Net



(b) After applying dropout.

[Srivastava et al., 2014]

Overfitting의 위험성



LRN(Locally Response-Normalization)

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

Network Architecture

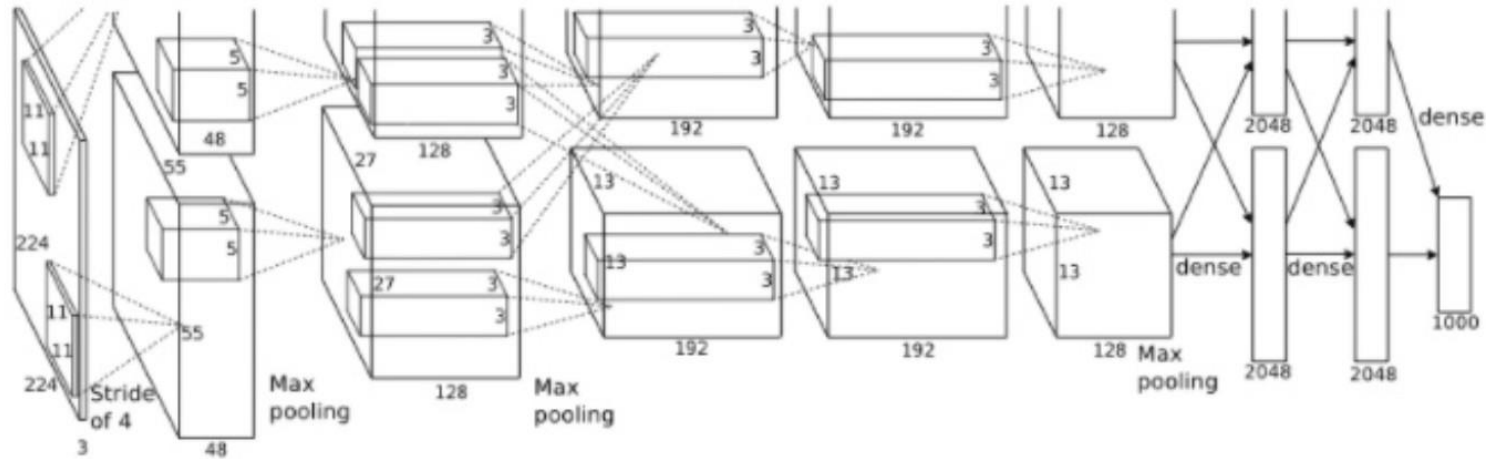


Figure 2: An illustration of the architecture of our CNN, explicitly showing the delineation of responsibilities between the two GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts at the bottom. The GPUs communicate only at certain layers. The network's input is 150,528-dimensional, and the number of neurons in the network's remaining layers is given by 253,440–186,624–64,896–64,896–43,264–4096–4096–1000.

Input Layer : 224 Width x 224 Height x 3 Channel Image
Convolution 1 : 96 Channel x 11x11 Filter + LRN + Pooling
Convolution 2 : 256 Channel x 5x5 Filter + LRN + Pooling
Convolution 3 : 384 Channel x 3x3 Filter
Convolution 4 : 384 Channel x 3x3 Filter
Convolution 5 : 192 Channel x 3x3 Filter + Pooling
Fully-Connected 1 : 4096
Fully-Connected 2 : 4096
Output Layer : 1000 Class

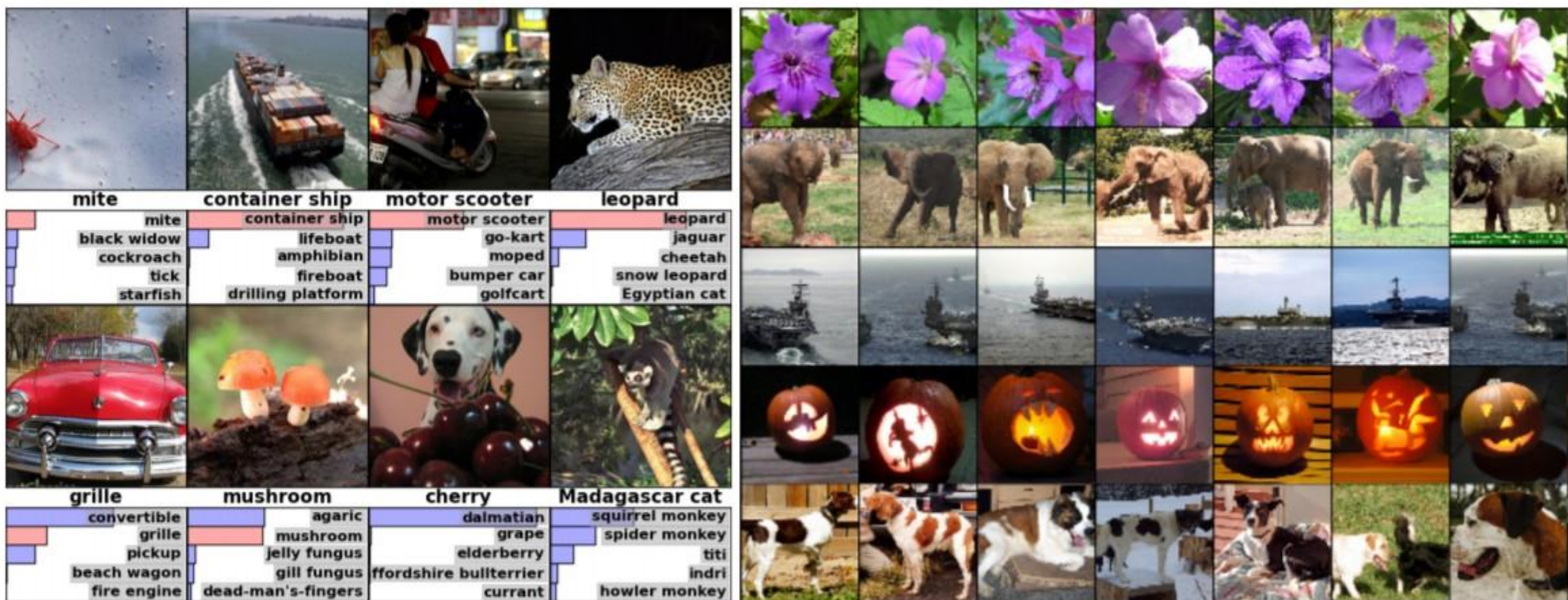


Figure 4: **(Left)** Eight ILSVRC-2010 test images and the five labels considered most probable by our model. The correct label is written under each image, and the probability assigned to the correct label is also shown with a red bar (if it happens to be in the top 5). **(Right)** Five ILSVRC-2010 test images in the first column. The remaining columns show the six training images that produce feature vectors in the last hidden layer with the smallest Euclidean distance from the feature vector for the test image.

Q & A

Thank you!