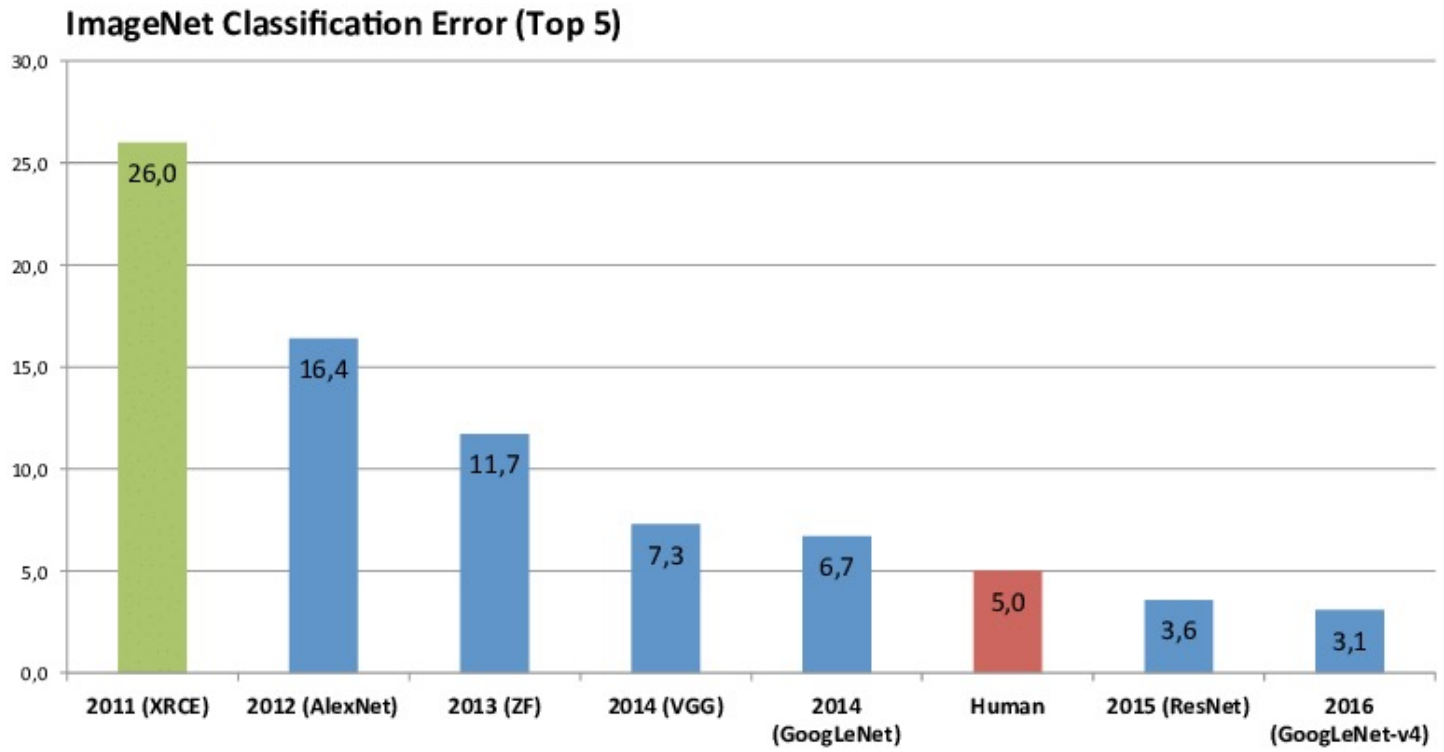
AlexNet

에이아이스쿨(AISchool) 대표
양진호 (솔라리스)
http://aischool.ai
http://solarisailab.com

연도별 ILSVRC 대회 우승 모델들

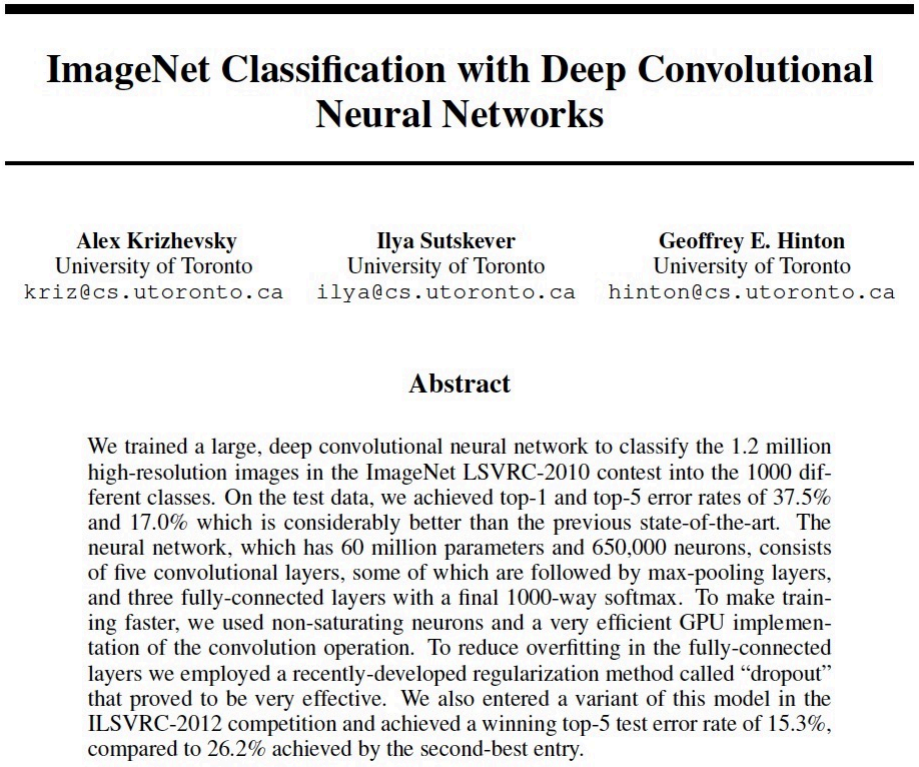
- 2012년 이전에는 CNN 외 기법들이 우승을 차지함
- 2012년 AlexNet이 큰 성능 gap을 만들면서 우승을 차지하면서 이후 대부분의 참가자들이 CNN 모델을 사용함
- 표준 CNN 모델들은 ILSVRC 대회에서 우승을 하거나 준우승을 차지한 모델들을 지칭함



Reference : https://devopedia.org/imagenet

AlexNet

- Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems 25 (2012): 1097-1105.
- https://papers.nips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf



ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky
University of Toronto
kriz@cs.utoronto.ca

Ilya Sutskever
University of Toronto
ilya@cs.utoronto.ca

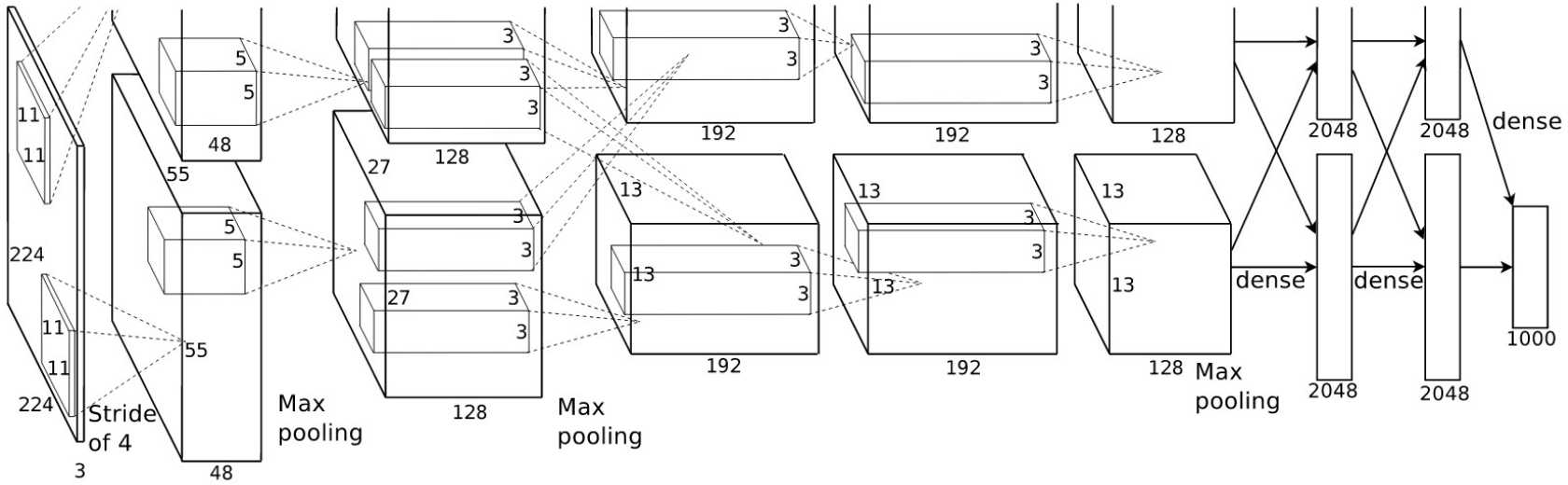
Geoffrey E. Hinton
University of Toronto
hinton@cs.utoronto.ca

Abstract

We trained a large, deep convolutional neural network to classify the 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into the 1000 different classes. On the test data, we achieved top-1 and top-5 error rates of 37.5% and 17.0% which is considerably better than the previous state-of-the-art. The neural network, which has 60 million parameters and 650,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers, and three fully-connected layers with a final 1000-way softmax. To make training faster, we used non-saturating neurons and a very efficient GPU implementation of the convolution operation. To reduce overfitting in the fully-connected layers we employed a recently-developed regularization method called "dropout" that proved to be very effective. We also entered a variant of this model in the ILSVRC-2012 competition and achieved a winning top-5 test error rate of 15.3%, compared to 26.2% achieved by the second-best entry.

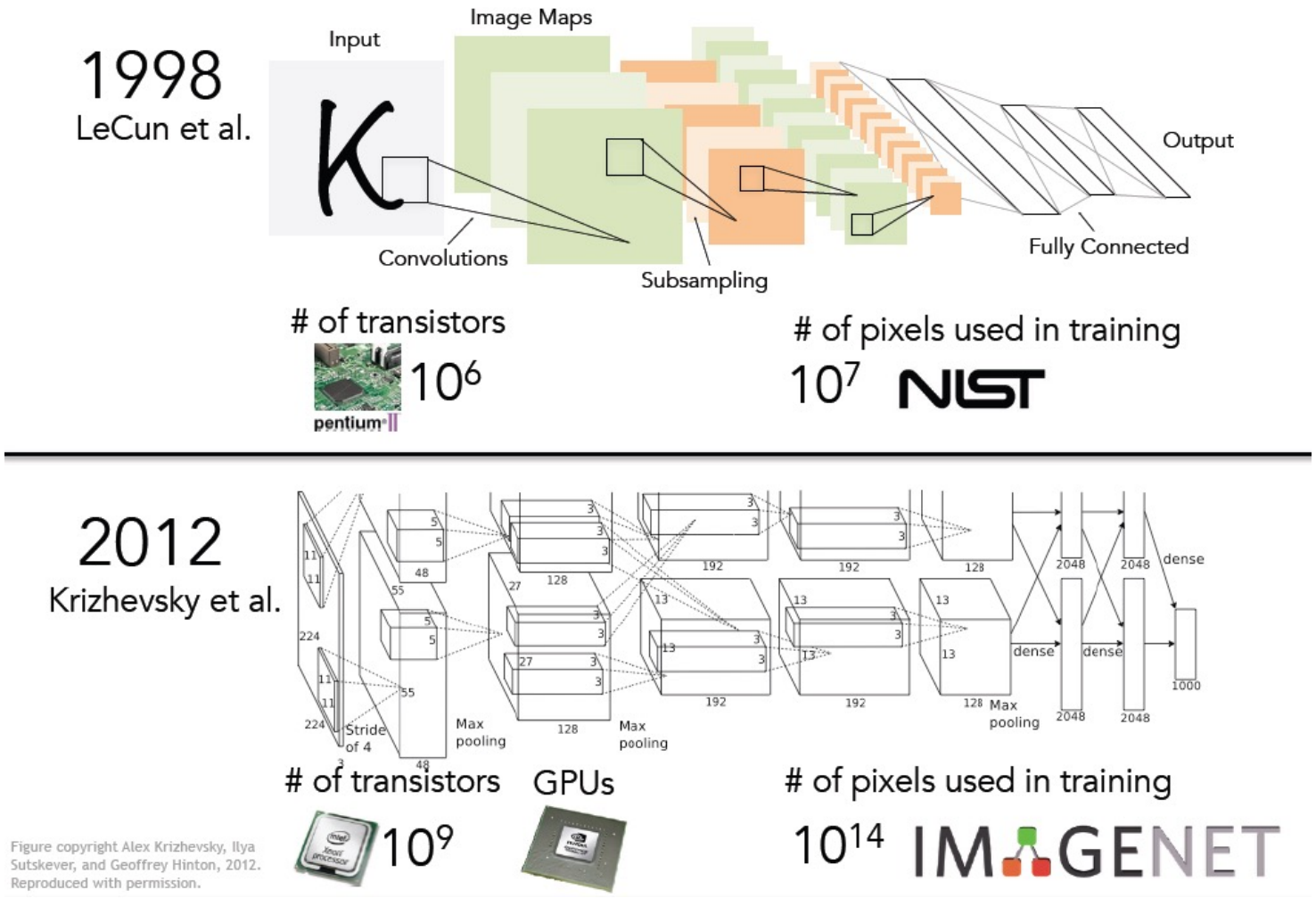
논문 리뷰 - ImageNet Classification with Deep Convolutional Neural Networks

- **핵심 아이디어** : GPU를 이용한 Deep Convolutional Neural Networks(AlexNet)를 이용해서 Image Classification을 수행



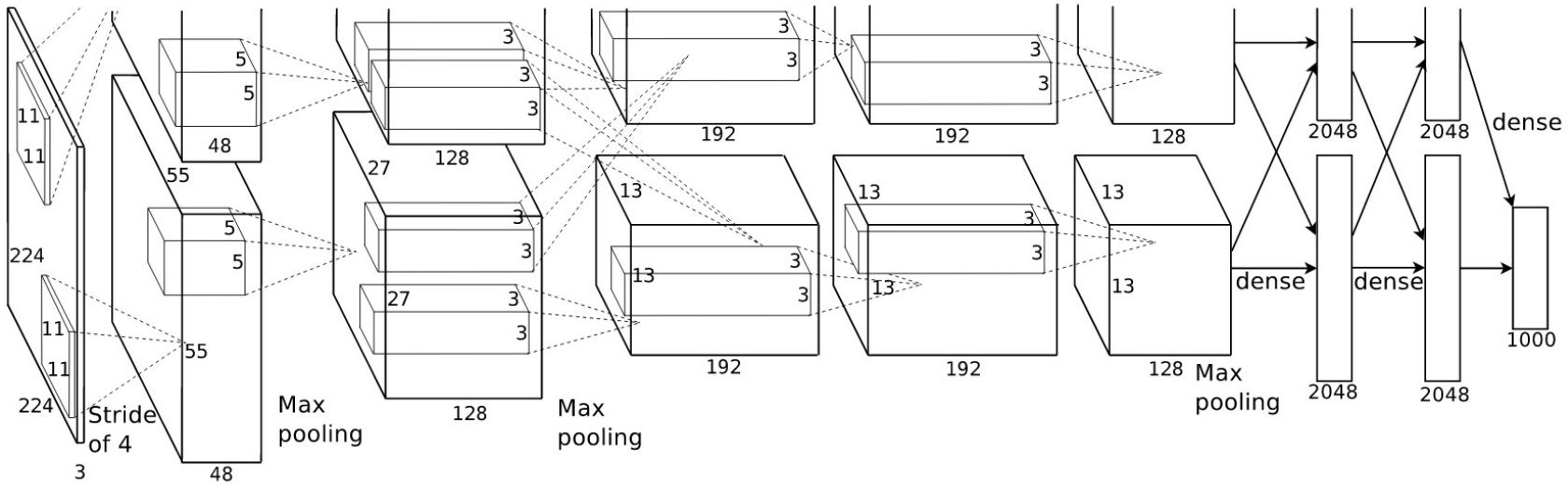
Deep Convolutional Neural Networks(AlexNet)

- 아이디어 자체는 1998년 LeNet과 크게 다르지 않음



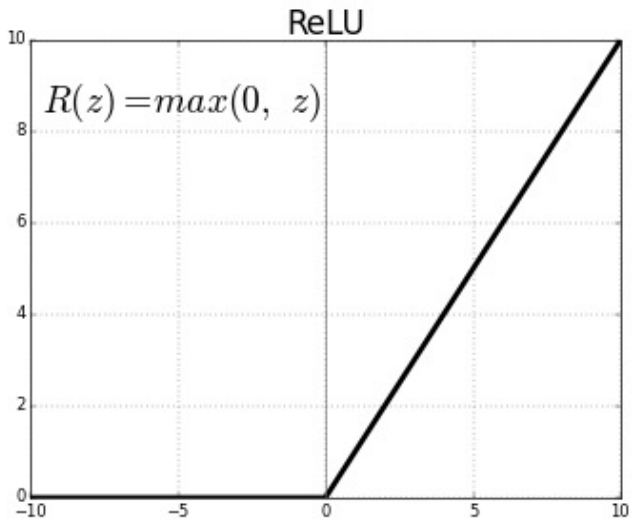
Model Architecture

- **Input** : 평균값을 뺀 227x227 크기의 RGB 이미지
- **Output** : 1000개의 Label (e.g. cat, tiger, hen, …)에 대한 확률



새로운 기법 1,2 : ReLU Non-Linearity, Multiple-GPU

- **새로운 기법 1,2** : ReLU Non-Linearity, Multiple-GPU(2개의 GPU를 사용)
- **ReLU NonLinearity** : Overfitting을 방지하고 학습속도가 훨씬 빨라짐(아래 실험의 경우 약 6배가 더 빠름)



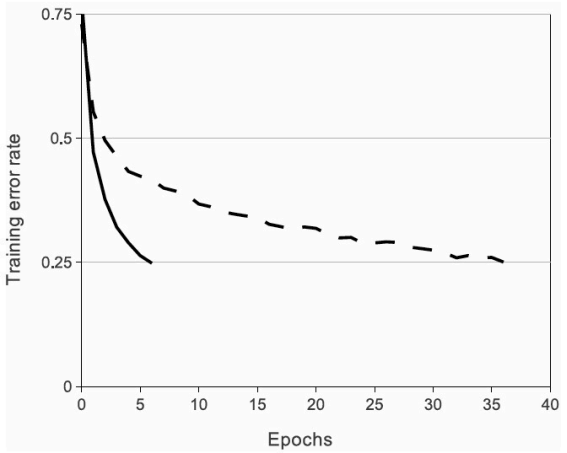
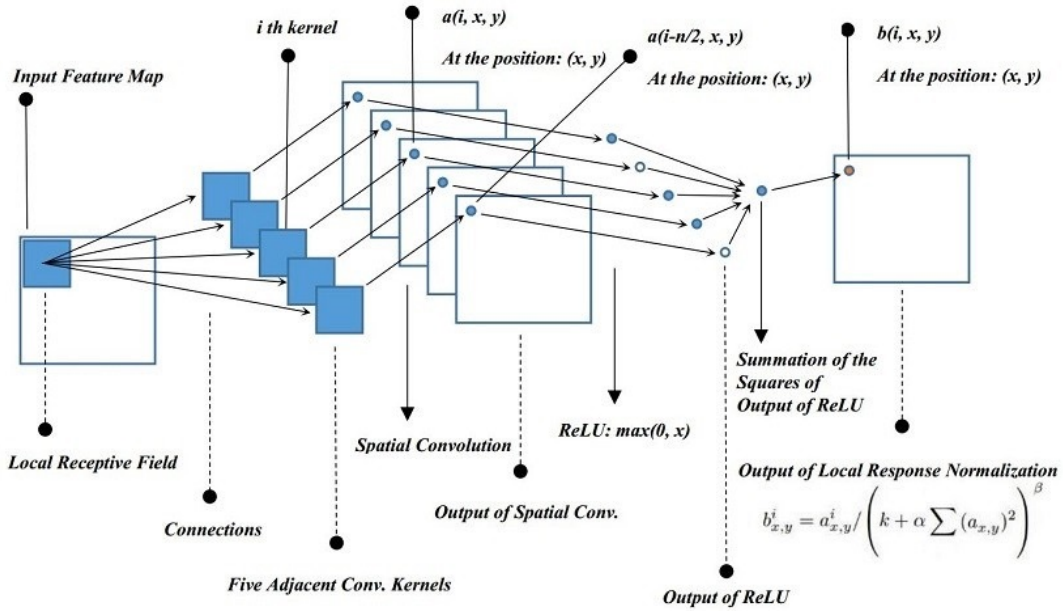


Figure 1: A four-layer convolutional neural network with ReLUs **(solid line)** reaches a 25% training error rate on CIFAR-10 six times faster than an equivalent network with tanh neurons **(dashed line)**. The learning rates for each network were chosen independently to make training as fast as possible. No regularization of any kind was employed. The magnitude of the effect demonstrated here varies with network architecture, but networks with ReLUs consistently learn several times faster than equivalents with saturating neurons.

새로운 기법 3 : Local Response Normalization

- 새로운 기법 3 : Local Response Normalization (generalization을 도와줌-강한 뉴런이 약한 뉴런의 값을 막는 현상을 억제함)
- Local Response Normalization : top-1과 top-5 에러율을 각각 1.4%, 1.2% 줄임.

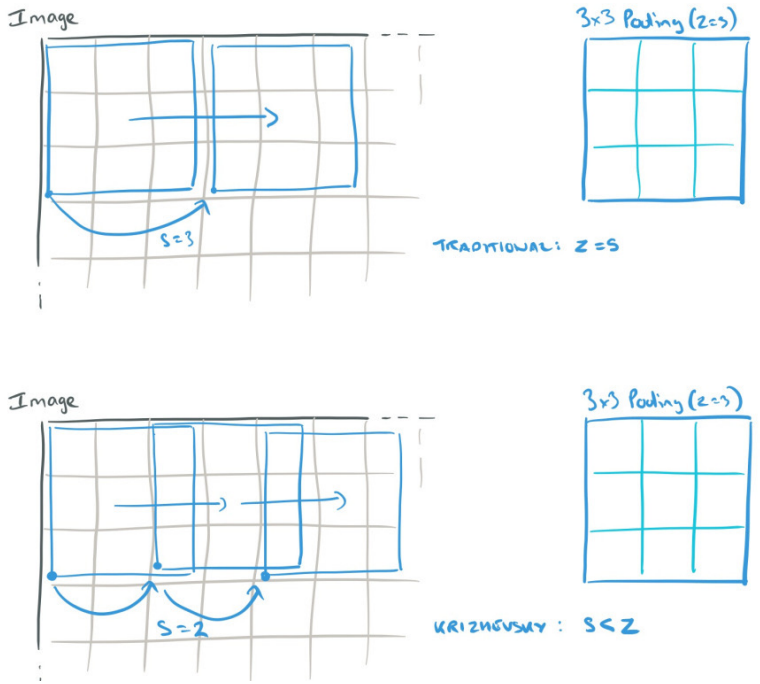
$$b^i_{x,y} = a^i_{x,y} / \left(k + \alpha \sum_{j=\max(0,i-n/2)}^{\min(N-1,i+n/2)} (a^j_{x,y})^2 \right)^{\beta}$$



Reference : https://www.quora.com/What-is-local-response-normalization

새로운 기법 4 : Overlapping Pooling

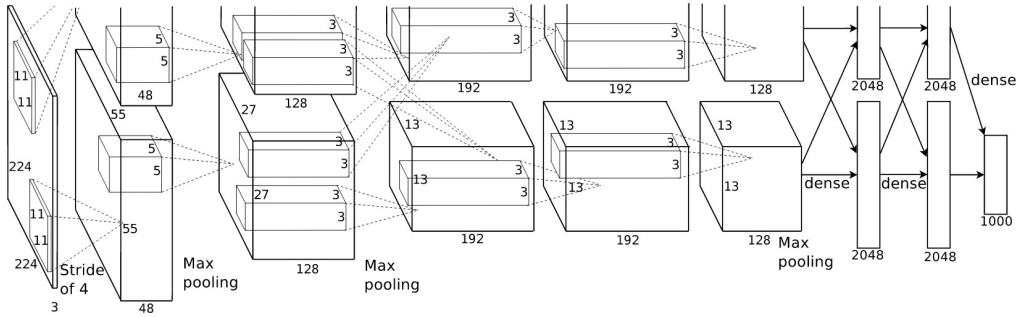
- 새로운 기법 4 : Overlapping Pooling
- **Overlapping Pooling** : 기존에는 stride와 filter size의 크기를 같게 하여(s=z) pooling할 때 overlapping이 일어나지 않게 사용함. 하지만, 이 논문에서는 s=2(stride), z=3(Filter size)을 사용하여 pooling을 overlapping함
- top-1과 top-5 에러율을 각각 0.4%, 0.3% 줄임.



Reference : https://blog.acolyer.org/2017/03/20/convolutional-neural-networks-part-1/

Overall Architecture Summary

- 2번째, 4번째, 5번재 컨볼루션 레이어와 각각의 다음 레이어는 같은 GPU에서 연결되어 있다.
- 3번째 컨볼루션 레이어에서는 모든 커널이 4번째 레이어와 연결되어 있다.풀리-커넥티드 레이어는 이전 레이어와 모두 연결되어 있다. ReLU 비선형 레이어는 모든 convolutional 레이어와 풀리-커넥티드 레이어의 output에 적용한다.
- [227x227x3] INPUT (논문에서는 224 x224로 잘못표현 됨-Zero-Padding 3이 미리 추가된걸로 추측-)
  [55x55x96] CONV1 : 96@ 11x11, s = 4, p = 0
  [27x27x96] MAX POOL1 : 3x3, s = 2
  [27x27x96] NORM1 :
  [27x27x256] CONV2 : 256@ 5x5, s = 1, p = 2
  [13x13x256] MAX POOL2 : 3x3, s = 2
  [13x13x256] NORM2 :
  [13x13x384] CONV3 : 384@ 3x3, s = 1, p = 1
  [13x13x384] CONV4 : 384@ 3x3, s = 1, p = 1
  [13x13x256] CONV5 : 256@ 3x3, s = 1, p = 1
  [6x6x256] MAX POOL3 : 3x3, s = 2
  [4096] FC6 : 4096 neurons
  [4096] FC7 : 4096 neurons
  [1000] FC8 : 1000 neurons



Overfitting 방지 기법 1 – Data Augmentation

- 1. 이미지 개수 증가 - 원본 이미지 크기 256x256에서 224x224 사이즈의 패치를 랜덤하게 추출한다. 상하 대칭으로도 이미지를 똑같은 방법으로 추출한다. 이렇게 추출하면 한 개의 원본 이미지로 2048-(256-224)*(256-224)* 2=2048-가지의 경우의 수가 나온다.
- 테스트 단계에서는 원본, 상하반전 이미지에서 5개씩 224x224 패치를 추출하여 softmax의 평균을 내어 추측한다. 이 때 5개의 패치는 4개의 코너와 중앙에서 추출한다.

Reference : https://learnopencv.com/understanding-alexnet/

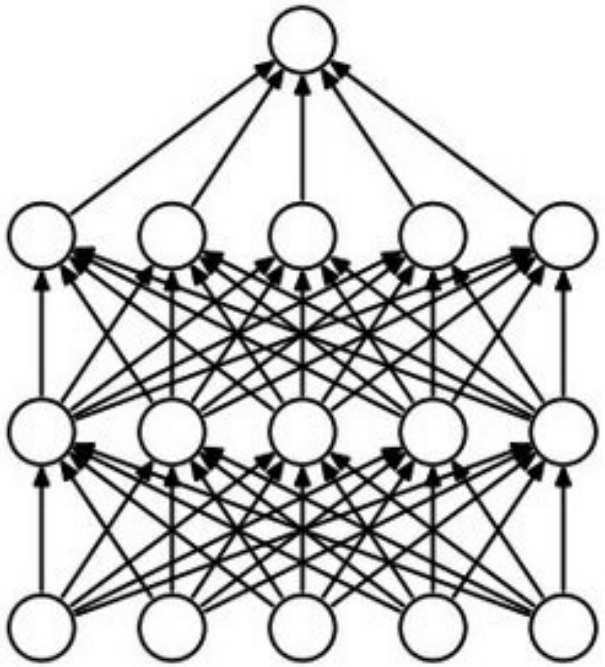
Overfitting 방지 기법 1 – Data Augmentation

- 2. 이미지 RGB 값 변화 - PCA를 통해 테스트 이미지의 **RGB 채널 강도를 변화시키는 augmentation**을 진행한다.

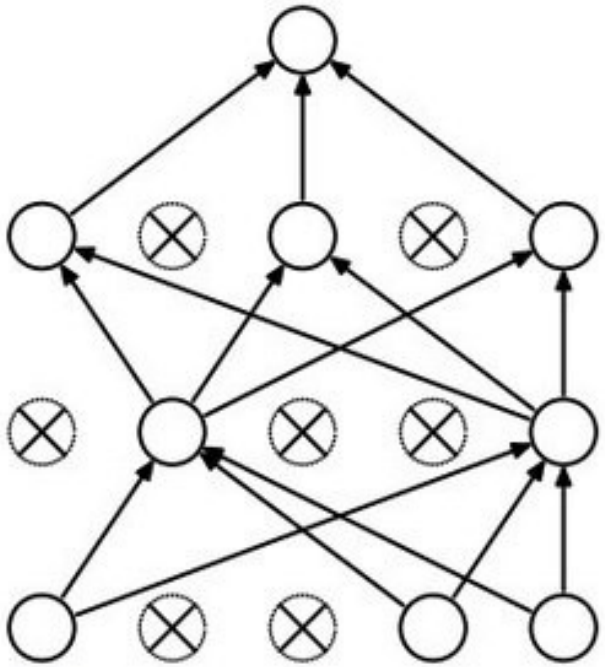
Reference : https://pixelatedbrian.github.io/2018-04-29-fancy_pca/

Overfitting 방지 기법 2 – Dropout

- Dropout을 통해 Overfitting을 방지.



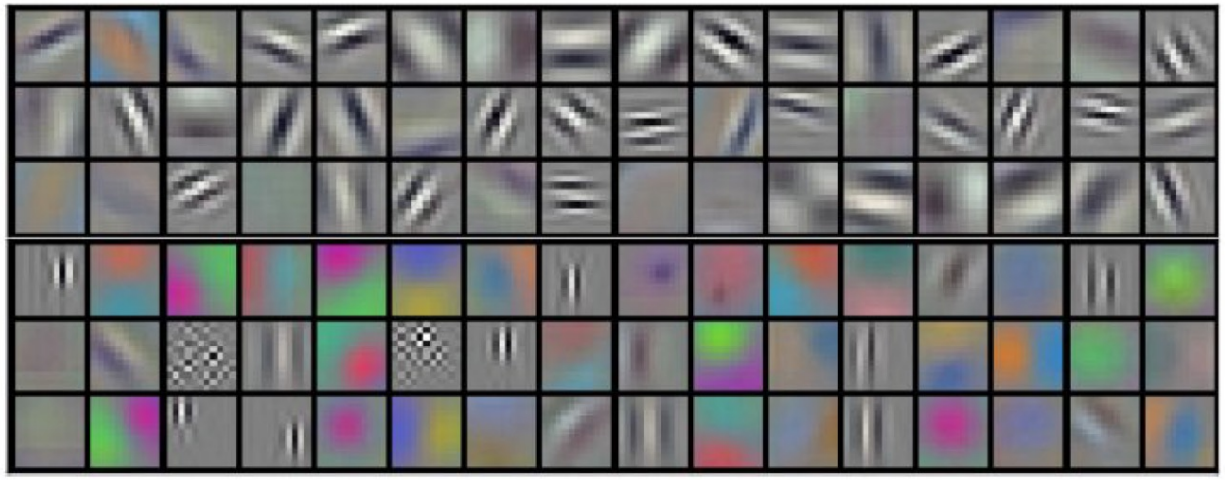
(a) Standard Neural Net



(b) After applying dropout.

학습된 필터들

- 트레이닝 결과로 학습된 필터들
- GPU1의 커널은 대부분 컬러와 상관없으며, GPU2의 커널은 대부분 컬러와 관련이 있다.이러한 특수성은 모든 동작에서 발생하며, 랜덤하게 초기 가중치를 설정하는 것과는 독립적이다.



Experiment Result 1

- 다른 방법들과의 성능 비교
- ILSVRC-2011 데이터로 pre-training한 뒤에 ILSVRC-2012 데이터로 Fine-Tuning한 경우 더 좋은 성능을 얻을 수 있었다.

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

Table 1: Comparison of results on ILSVRC-2010 test set. In *italics* are best results achieved by others.

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%**

Table 2: Comparison of error rates on ILSVRC-2012 validation and test sets. In *italics* are best results achieved by others. Models with an asterisk* were "pre-trained" to classify the entire ImageNet 2011 Fall release. See Section 6 for details.

Experiment Result 2

- 실험결과

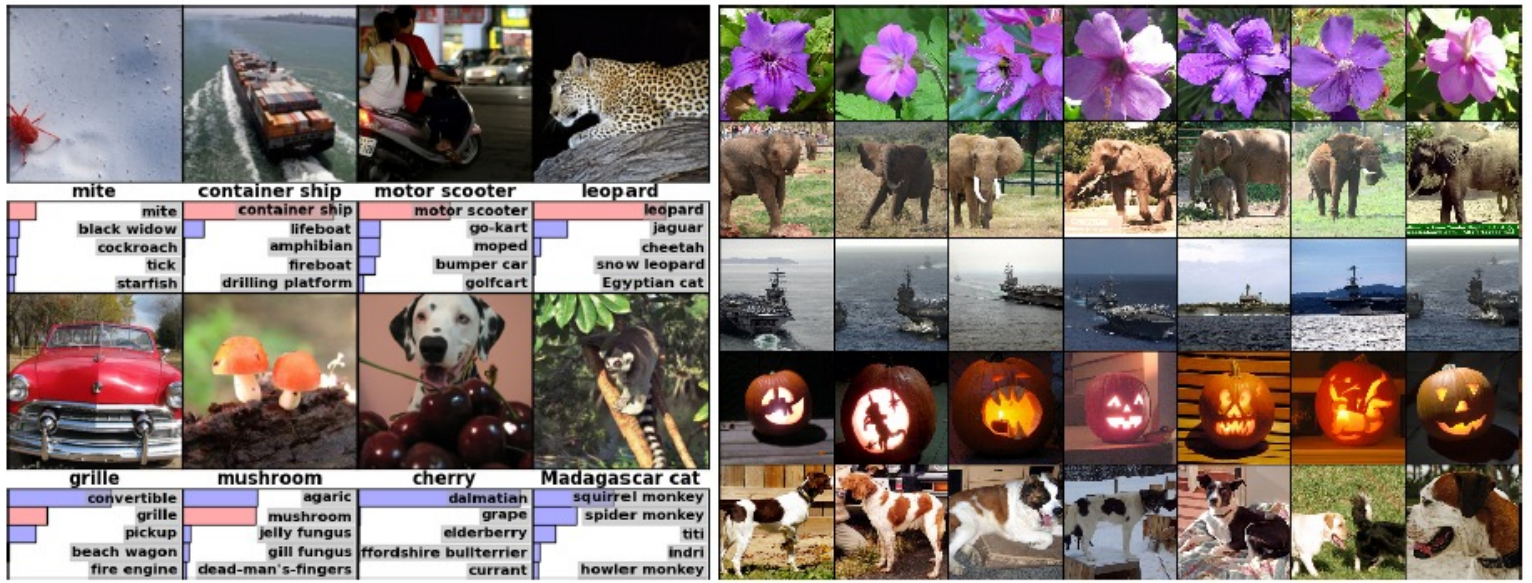


Figure 4: **(Left)** Eight ILSVRC-2010 test images and the five labels considered most probable by our model. The correct label is written under each image, and the probability assigned to the correct label is also shown with a red bar (if it happens to be in the top 5). **(Right)** Five ILSVRC-2010 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.

Experiment Result 2

- 실험결과

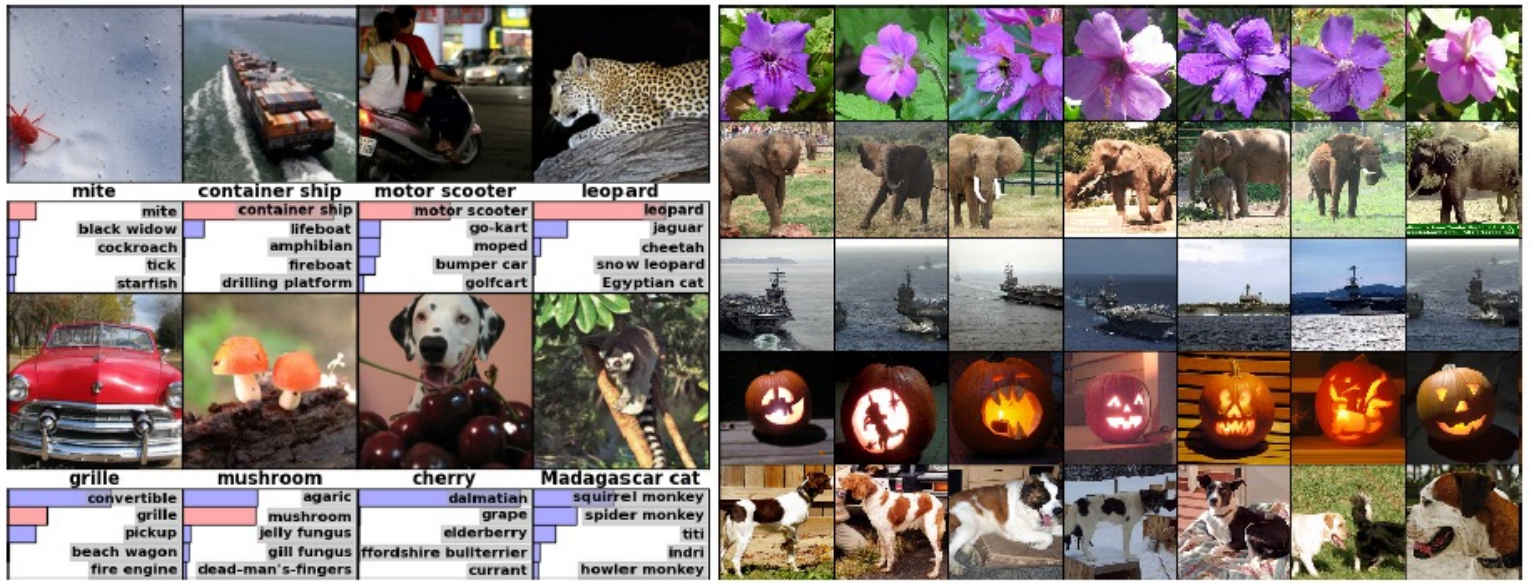


Figure 4: **(Left)** Eight ILSVRC-2010 test images and the five labels considered most probable by our model. The correct label is written under each image, and the probability assigned to the correct label is also shown with a red bar (if it happens to be in the top 5). **(Right)** Five ILSVRC-2010 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.

AlexNet의 의의

- 딥러닝 기반 컴퓨터 비전 기법의 시작과 딥러닝의 붐을 일으킨 기념비적인 모델
- 초기 모델이다 보니 개선할 부분은 많이 존재

Thank you!