

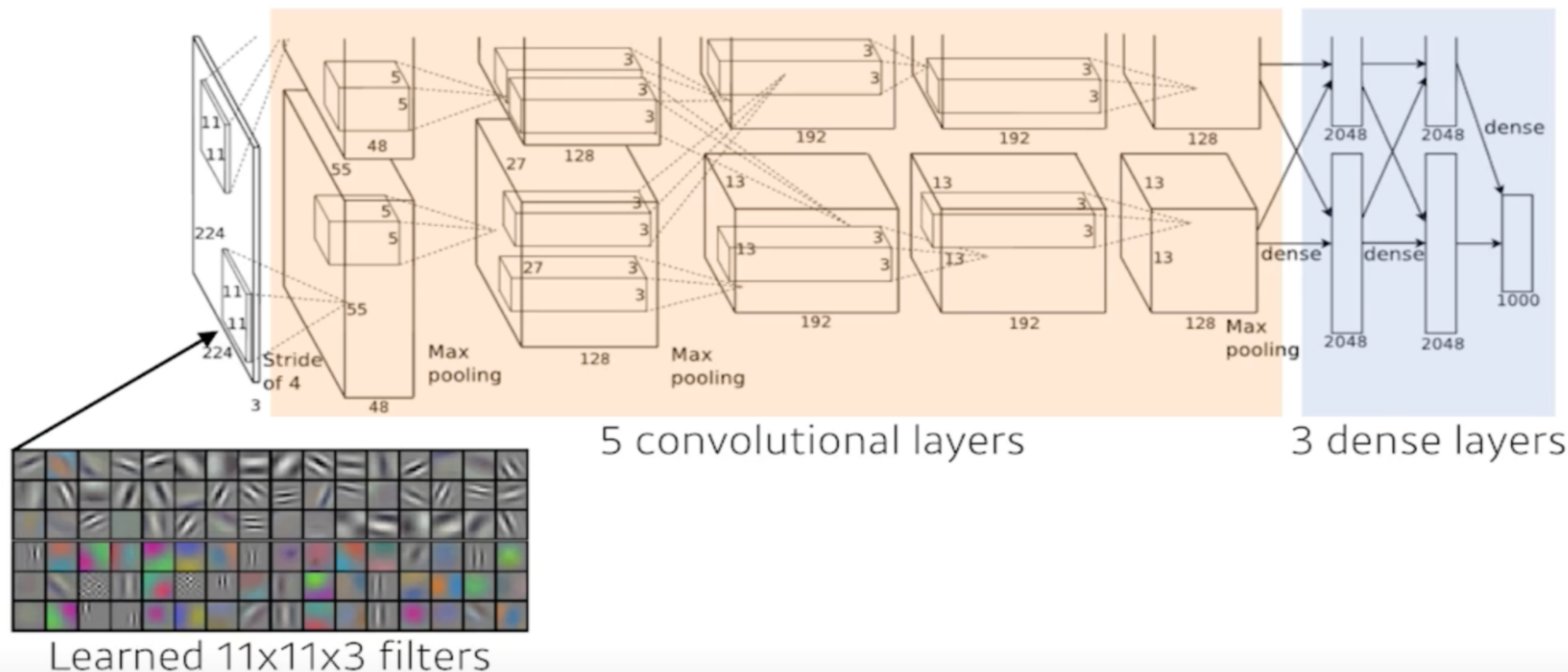
ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton

Abstract

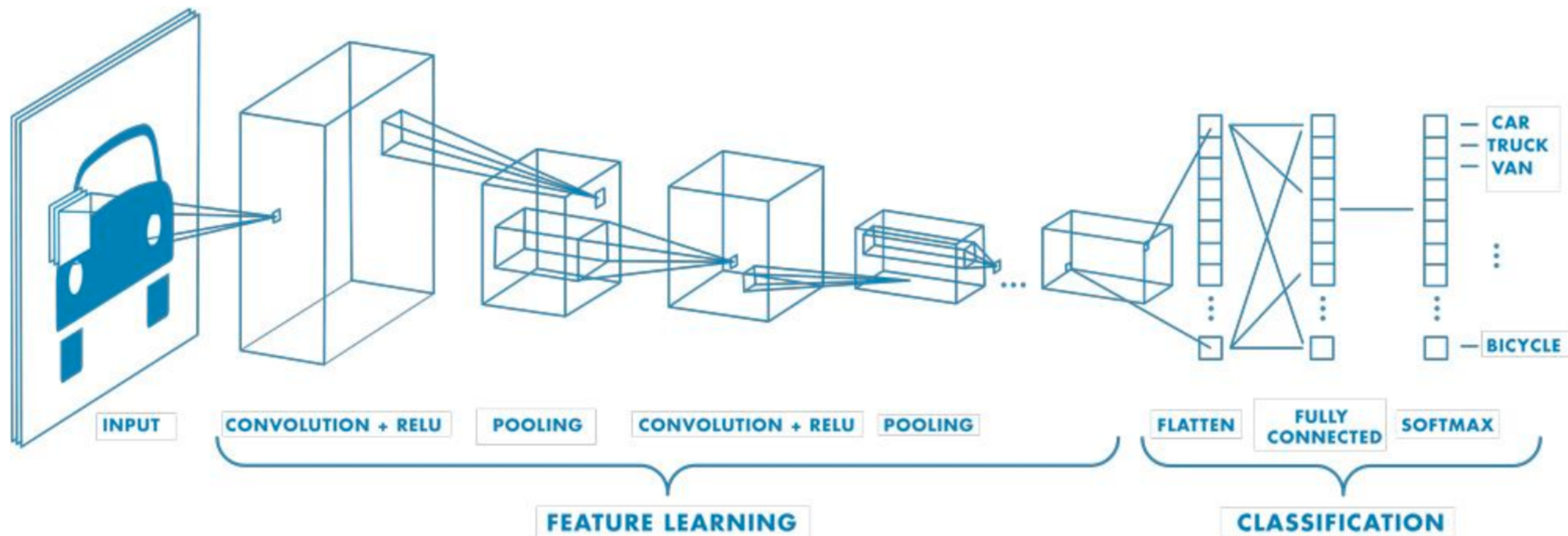
AlexNet을 한 줄로 요약하면 CNN에 여러가지 방법을 사용하여 층을 깊게 쌓을 수 있게 만듦으로써 딥러닝을 대중화 시킨 모델이라고 할 수 있다.

AlexNet 이후로 여러 딥러닝 학습 방법들이 제시 되면서 이미지 분류 분야뿐만 아니라 여러 분야에서 딥러닝 모델이 대중화되었다.



CNN

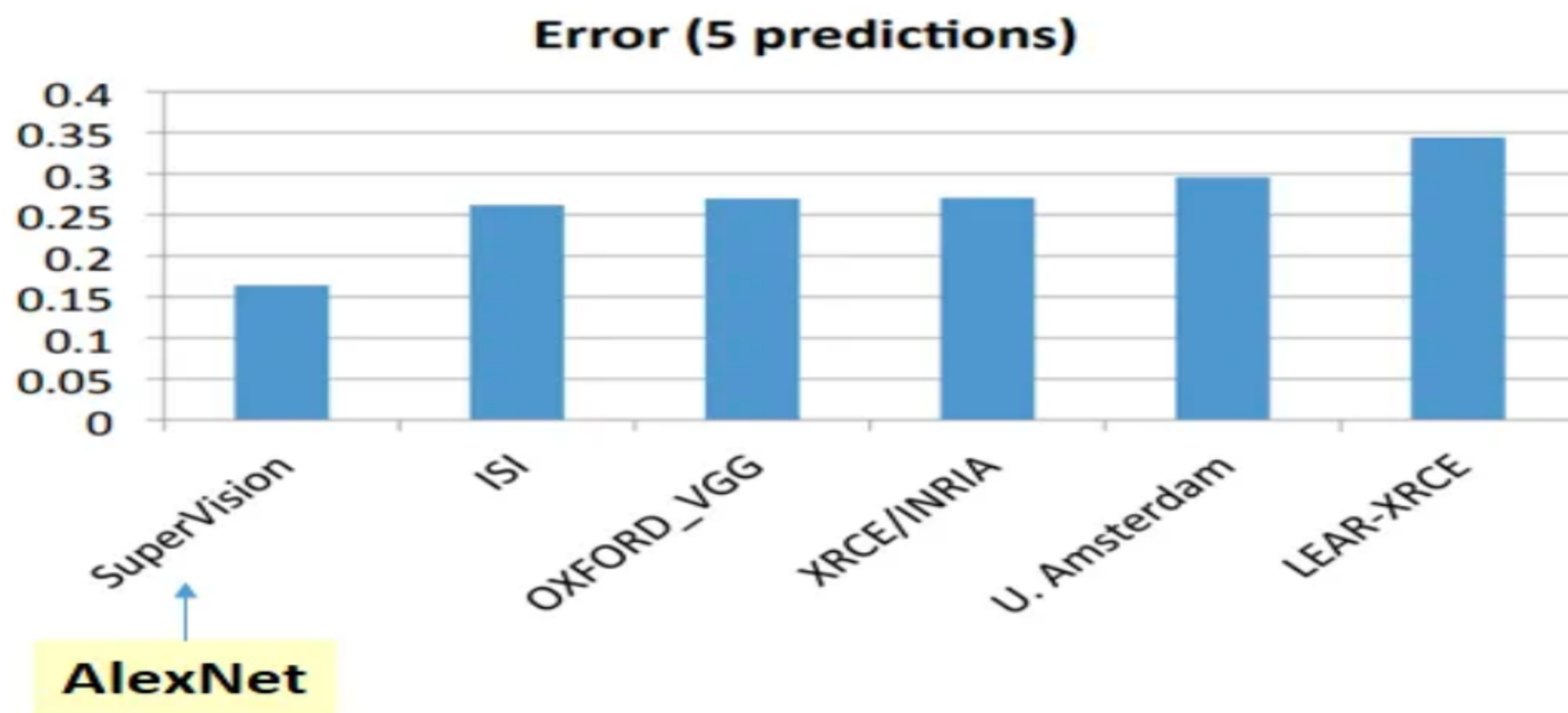
데이터의 공간적인 정보를 잃지 않고 데이터의 패턴을 분석하는 신경망. 1989년에 얀 르쿤이 우편물의 우편 번호를 분류하기 위해 최초로 고안하였다. 하지만 CNN은 해상도가 높은 데이터에 필요한 깊은 신경망에서 성능이 떨어지고 학습이 안 되는 문제점이 있었다.



AlexNet

AlexNet은 ILSVRC(Imagenet Large-Scale Visual Recognition Challenge)-2012 대회에서 2위와 에러율 10% 이상의 차이로 1위를 했다. 이 이후로 SVM과 같은 고전적인 머신러닝 방법이 아닌 딥러닝 방법만 대회에서 우승했다. 약 6000만개의 parameter를 가지고 있다. 당시에는 일반적이지 않은 방법을 사용하여 이러한 깊은 신경망을 학습시켰다. 논문에서는 성능을 높이는 방법 4가지를 저자들이 평가한 중요도 순서대로 나열하고 있다.

Ranking of the best results from each team

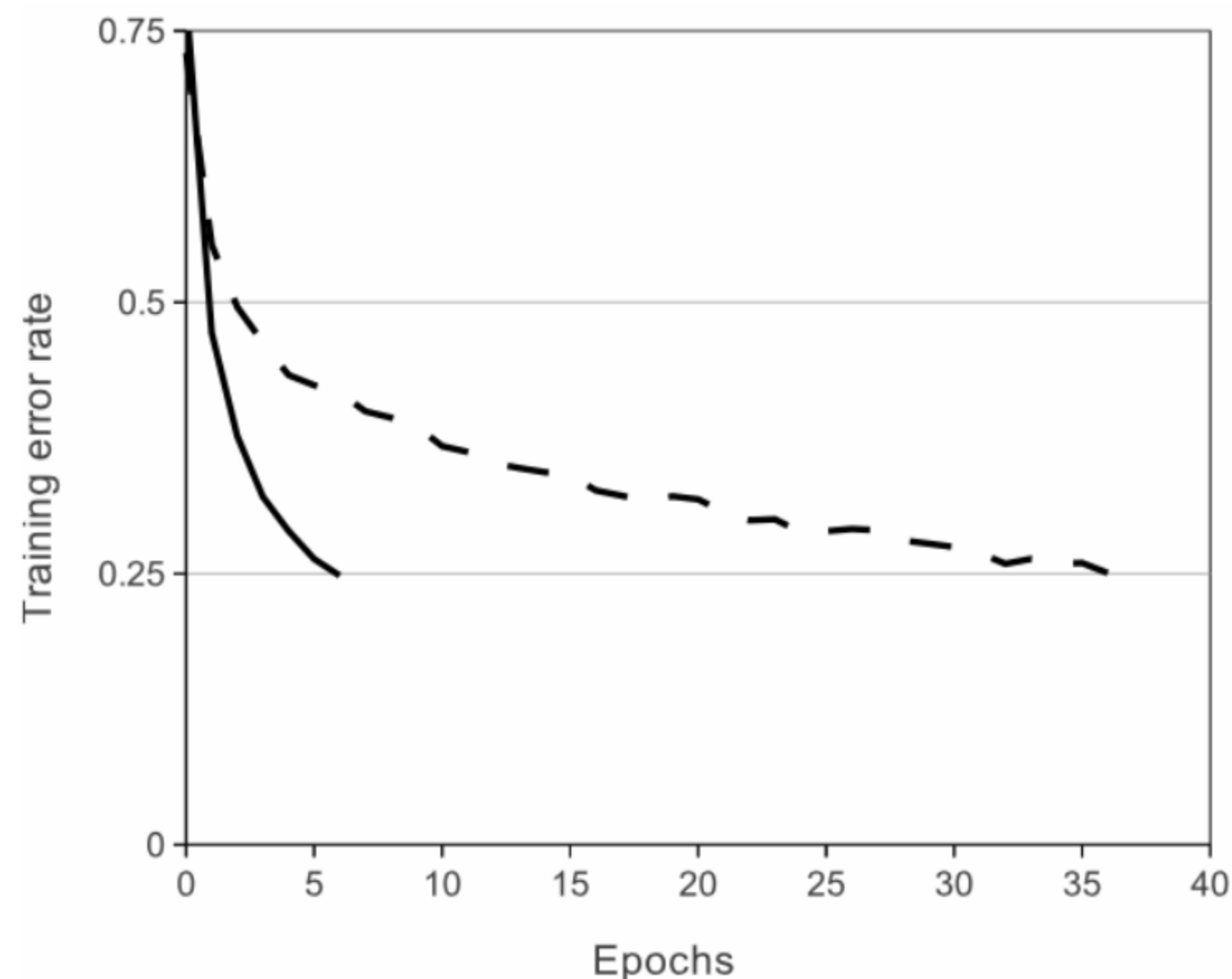


ReLU(Rectified Linear Unit)

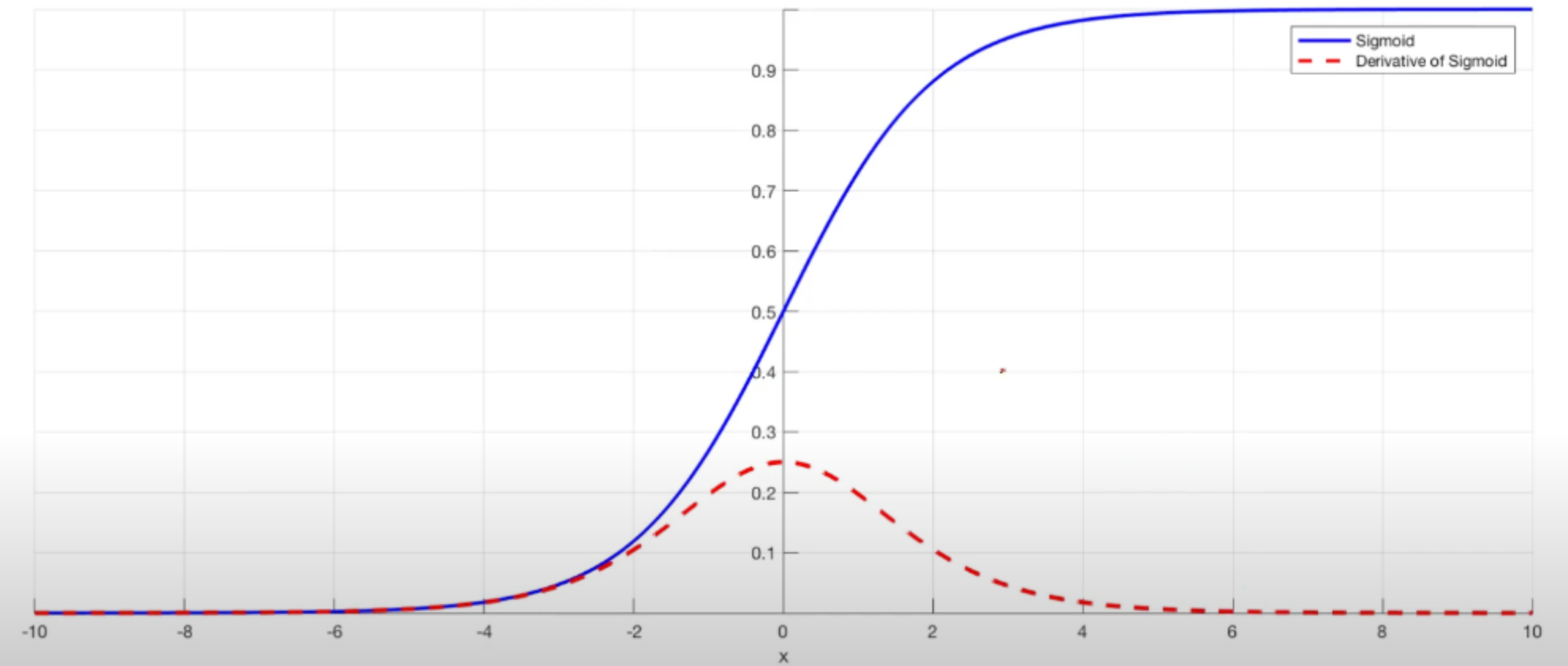
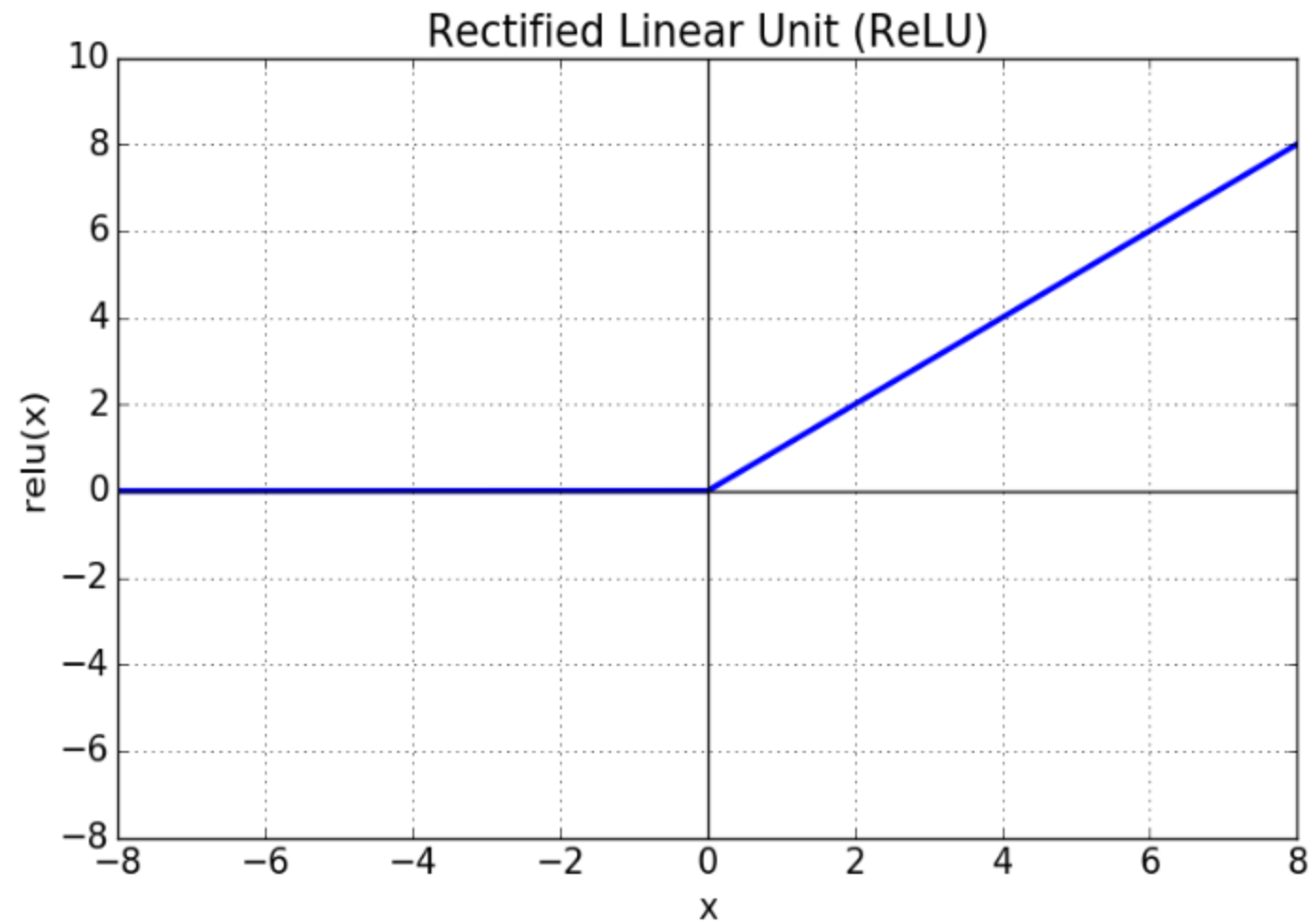
기존에 사용하던 활성화함수인 sigmoid나 하이퍼볼릭탄젠트는 신경망이 깊어질수록 기울기가 소멸되는 문제가 있었으나 AlexNet에서는 이러한 문제를 각 레이어에 ReLU 함수를 적용하여 해결했다. 논문에서는 이러한 Relu 함수가 없었다면 깊은 신경망을 학습할 수 없었을 것이라고 말한다.

ReLU함수의 장점

- 미분 계산이 쉬워 학습 속도가 빠름
- non-saturation 함수: 학습이 안 되는 문제점 개선
- vanishing gradient 문제 완화

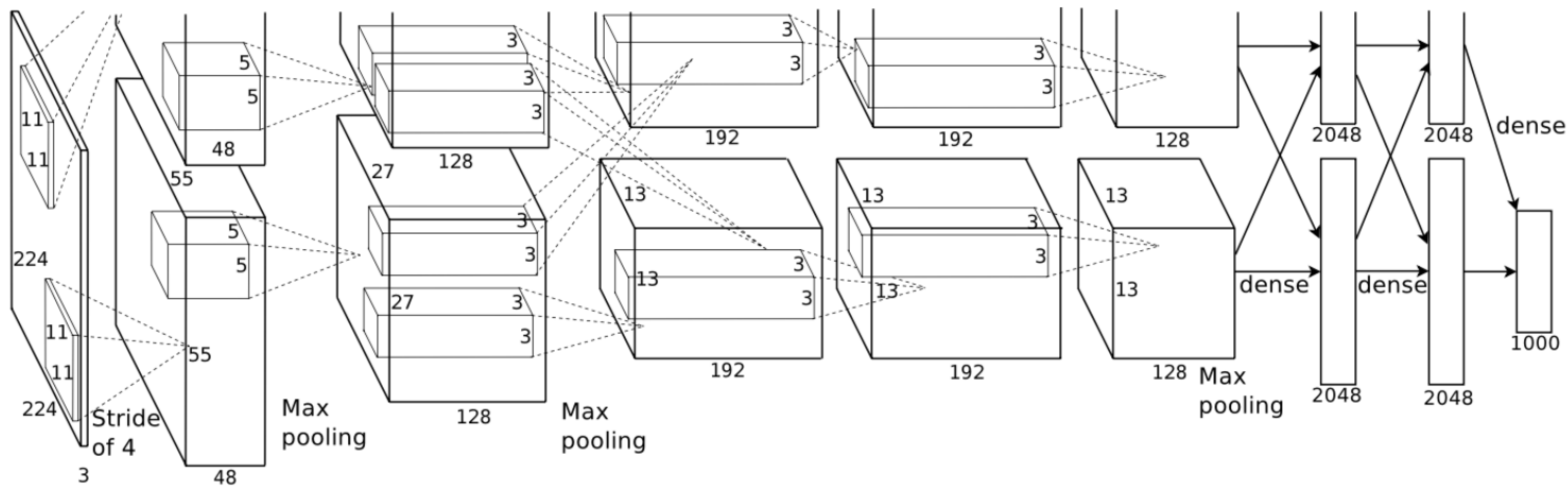


ReLU (Rectified Linear Unit)



Training on GPU

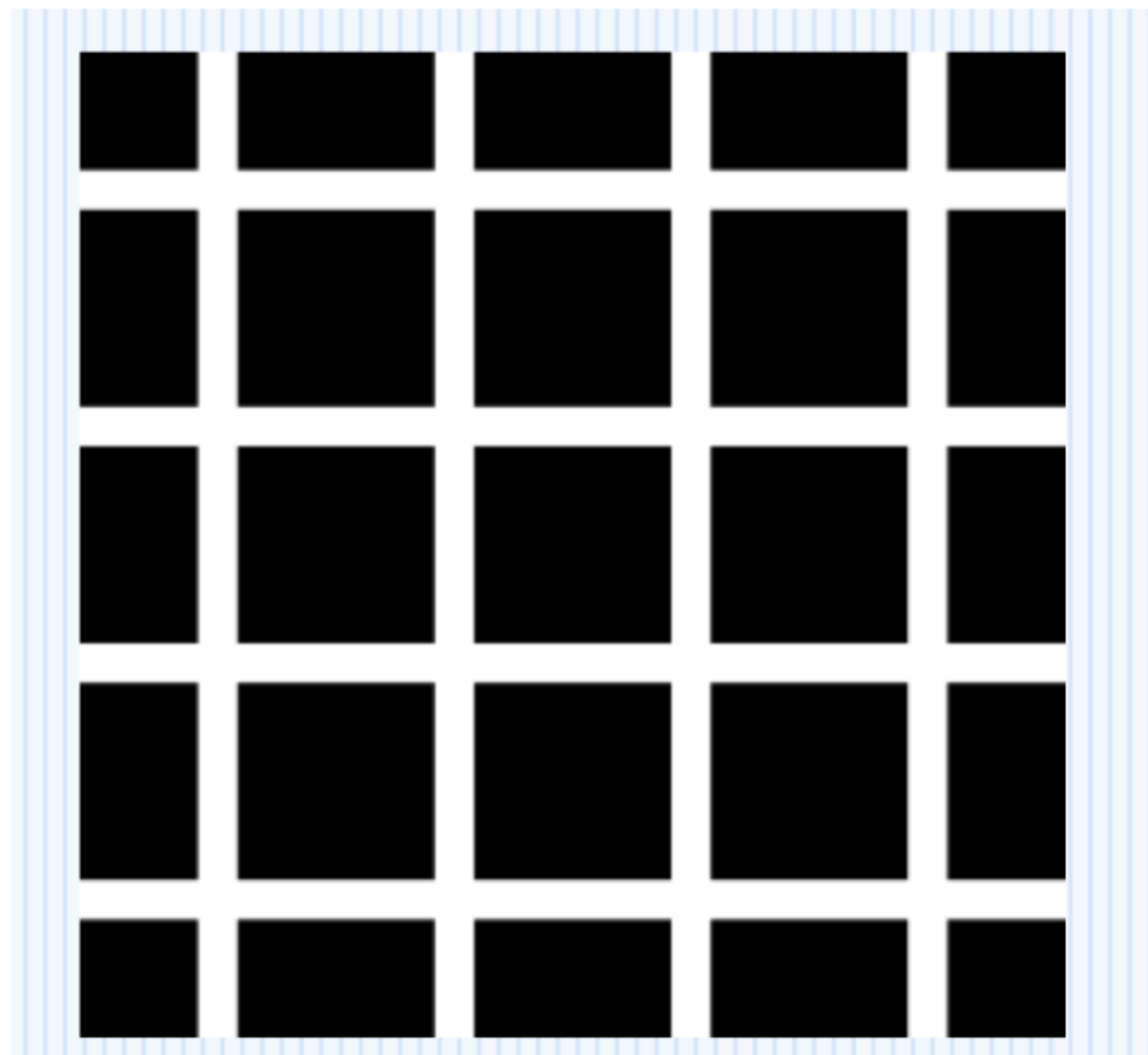
AlexNet은 3GB의 GTX 580 GPU 2개를 사용하여 학습을 했다. 당시에는 GPU의 성능이 좋지 않았기 때문에 이러한 방법을 사용했다. 논문에서는 GPU의 성능이 좋아짐에 따라 AlexNet의 성능도 좋아질 것이라고 말한다.



Local Response Normalization

신경망에서 어떤 가중치가 주변 가중치에 비해 상당히 클 때 그 값을 주변 가중치와 비슷하게 줄여주는 정규화 방법이다. ReLU 함수는 출력 값이 매우 클 수 있기 때문에 LRN을 통해 출력 값이 너무 커지지 않게 조절한다. 현재는 batch normalization을 많이 사용한다.

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



Overlapping Pooling

Non-overlapping pooling

1	3	5	5
4	1	4	9
3	2	0	1
5	2	4	6



4	9
5	6

Stride 2
2 x 2 max pooling

Overlapping pooling

1	3	5	5
4	1	4	9
3	2	0	1
5	2	4	6

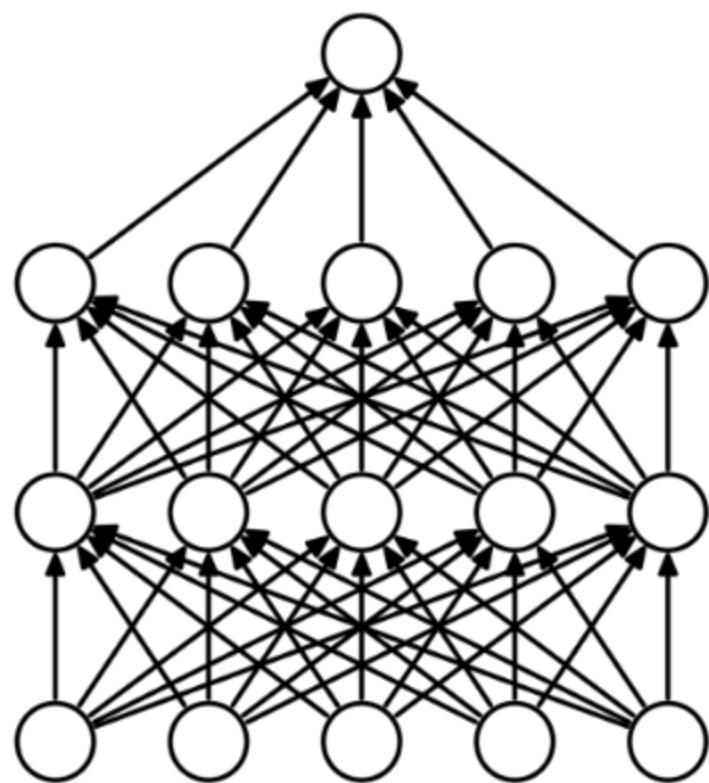


4	5	9
4	4	9
5	4	6

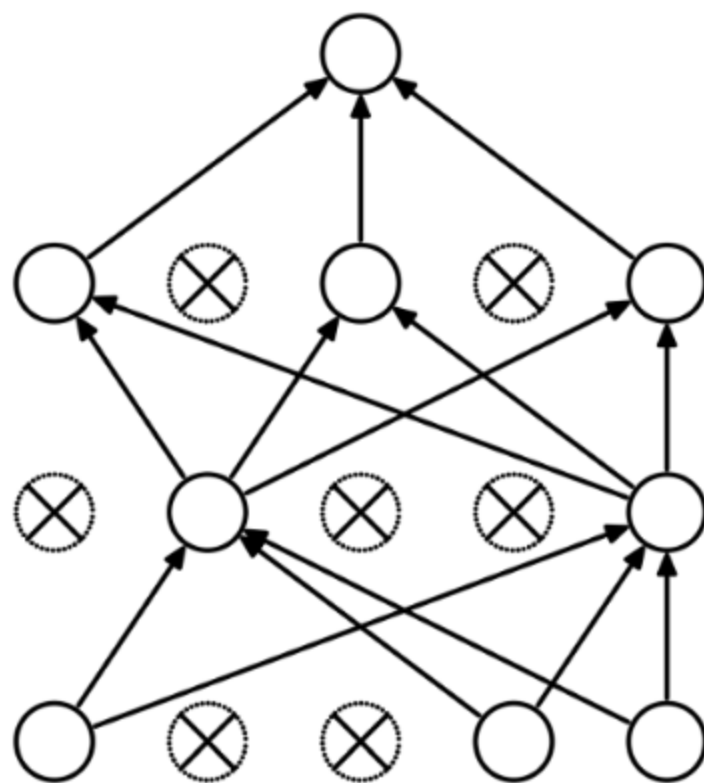
Stride 1
2 x 2 max pooling

Reducing Overfitting

신경망에서는 층이 깊어질수록 overfitting의 가능성이 커진다. 따라서 논문에서는 overfitting을 줄이는 두 가지 방법으로 Data argumentation과 Dropout을 제시했다.



(a) Standard Neural Net



(b) After applying dropout.

Discussion

- 지금은 당연하게 쓰이는 방법들이지만 해당 논문이 나온 2012년 당시에는 당연하지 않은 방법들을 사용했다.
- 기존에 많이 쓰이던 머신러닝 방법들이 아닌 딥러닝 방법을 사용하였다.
- 하나의 층이 제거 될 때마다 성능이 상당히 줄어들었다. -> 앞으로는 층을 깊게 쌓는 것이 딥러닝에 중요한 요소가 될 것이다.
- 컴퓨팅 파워의 발전이 곧 deep neural network의 발전이 될 것이다. -> GPU의 발전에 따라 deep neural network의 성능이 발전할 것이라고 예측된다.