ImageNet Classification with Deep Convolutional Neural Networks

NIPS 2012

GIST EECS
윤준영

Paper Info

Paper : ImageNet Classification with Deep Convolutional Neural Networks

Authors : Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton

Journal : Neural Information Processing Systems (NIPS)

Citations : 77883

Main problem

- DNN(Deep Neural Network) – good performance only on small dataset

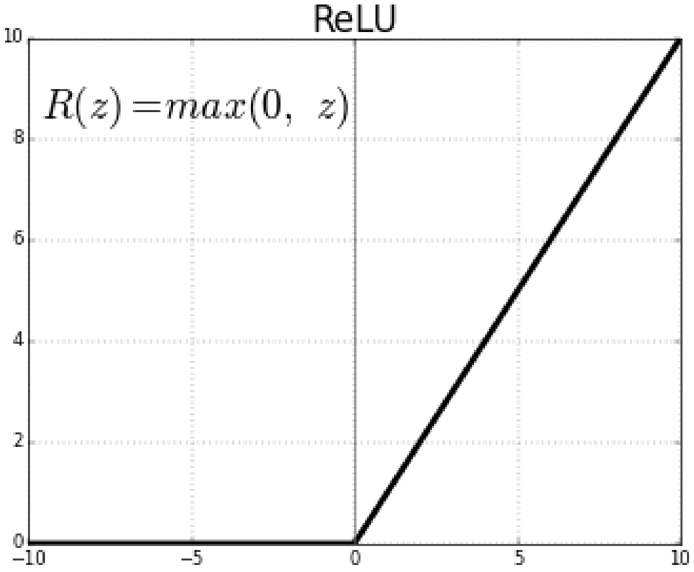
Try deep convolutional neural network to classify Large image set

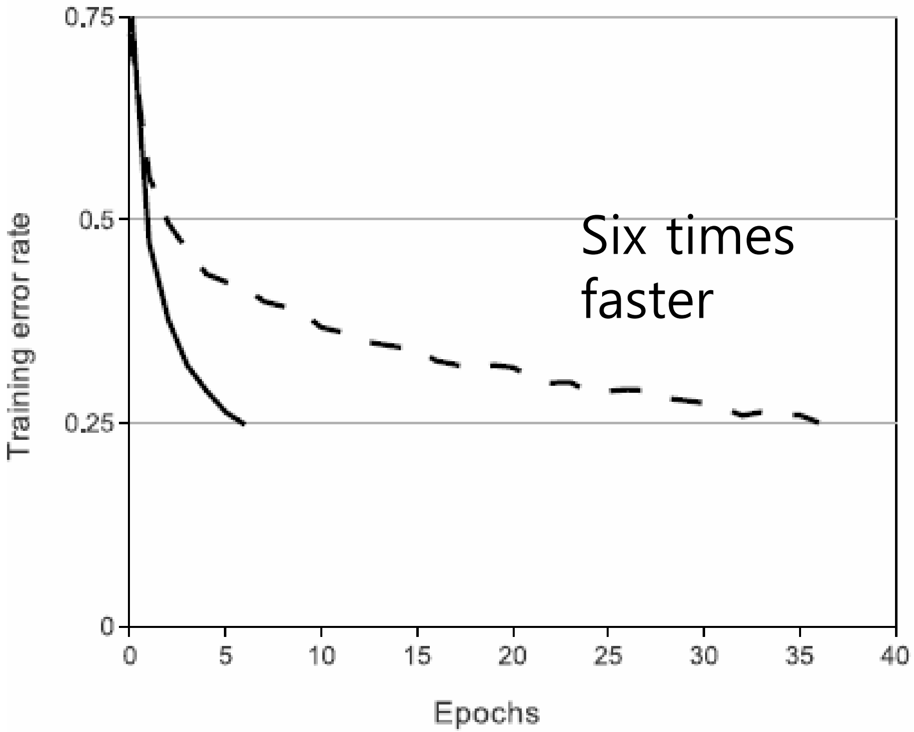
Dataset Input

Down-sampled the images to a fixed resolution of 256X256

AlexNet : Architecture

- ReLU
  - Train faster than tanh or sigmoid units
  - Non-saturation
  - Used in all layers as activation function





Traing error rate on CIFAR-10

ReLU (**solid line**)
tanh (**dashed line**)

AlexNet : Architecture

- LRN(Local Response Normalization)
  - Relu – do not require input normalization (non-saturation)
  - To prevent effect similar to lateral inhibition
  - Reduced error rates 1.4% and 1.2%

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

where

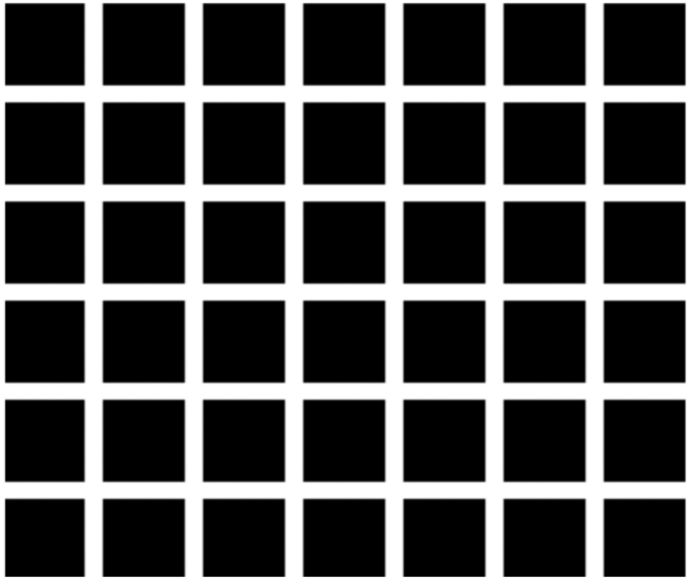
$b^i_{x,y}$ – regularized output for kernel i at position x, y

$a^i_{x,y}$ – source ouput of kernel i applied at position x, y

N – total number of kernels

n – size of the normalization neigbourhood

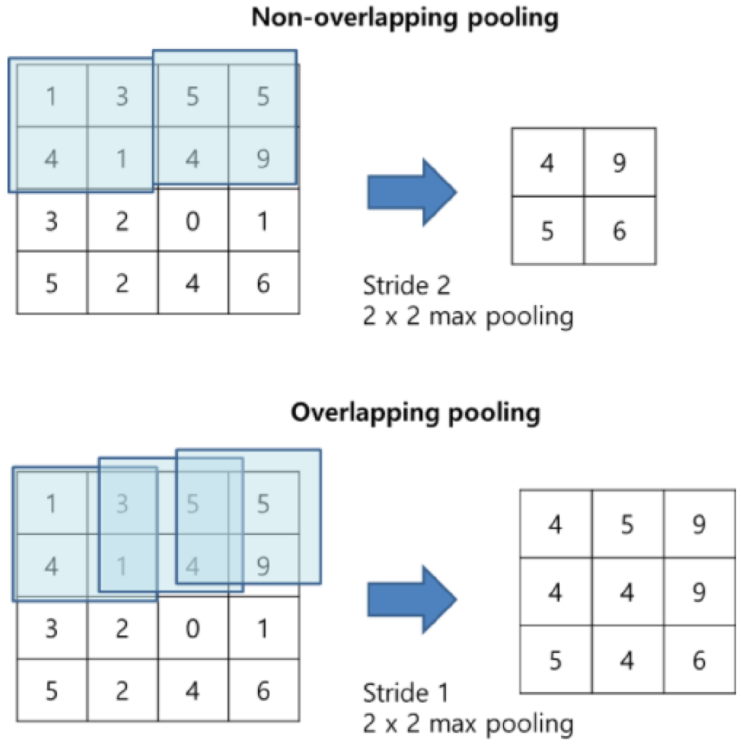
$\alpha, \beta, k, (n)$ – hyperparameters



hermann-grid

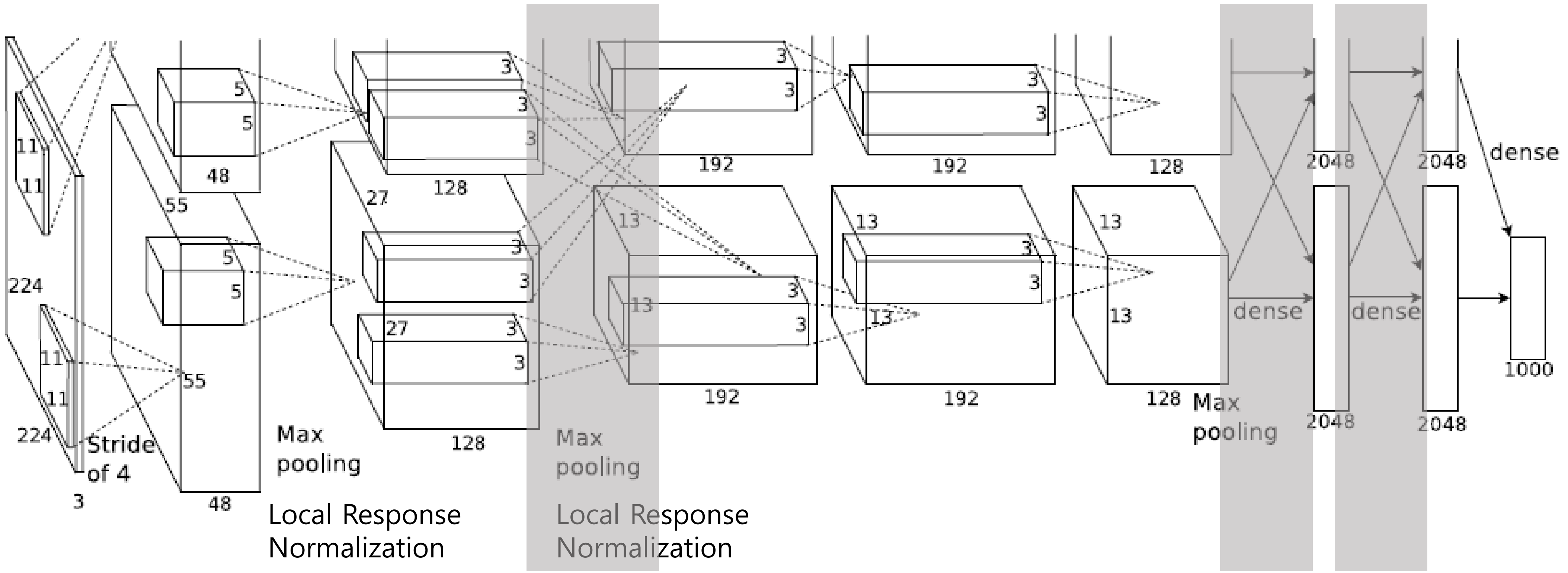
AlexNet : Architecture

- Overlapping pooling
  - Set stride smaller than feature map size
  - Overlapping max pooling was used
  - Error rate reduced 0.4%



AlexNet : Architecture

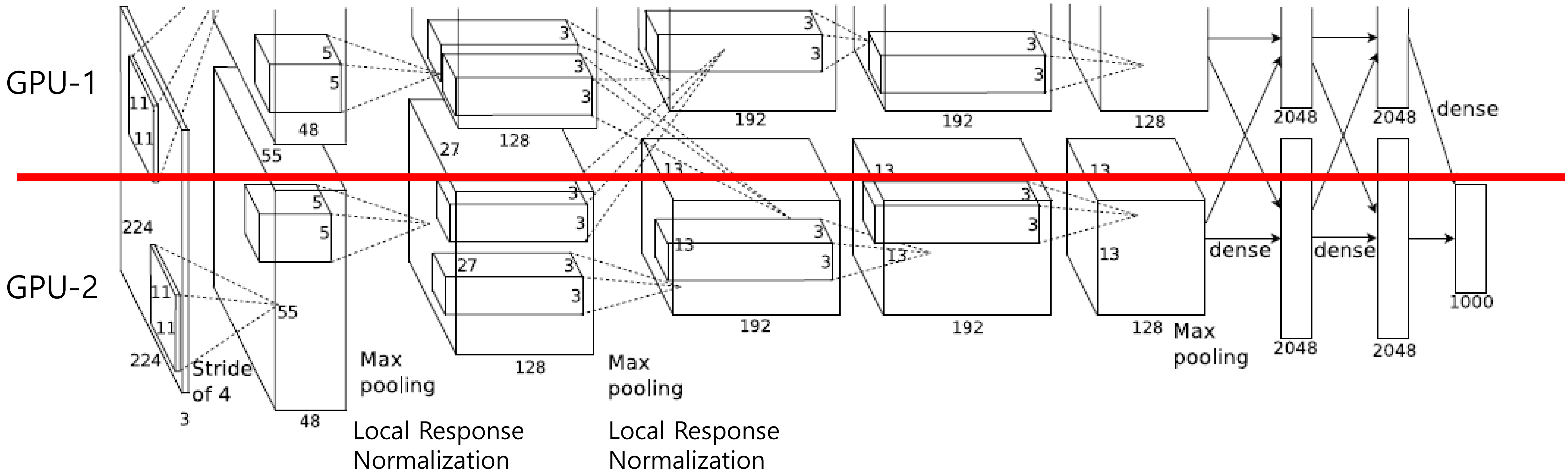
Multiple GPUs



Communications in few layers, reduced error rates, faster than one GPU

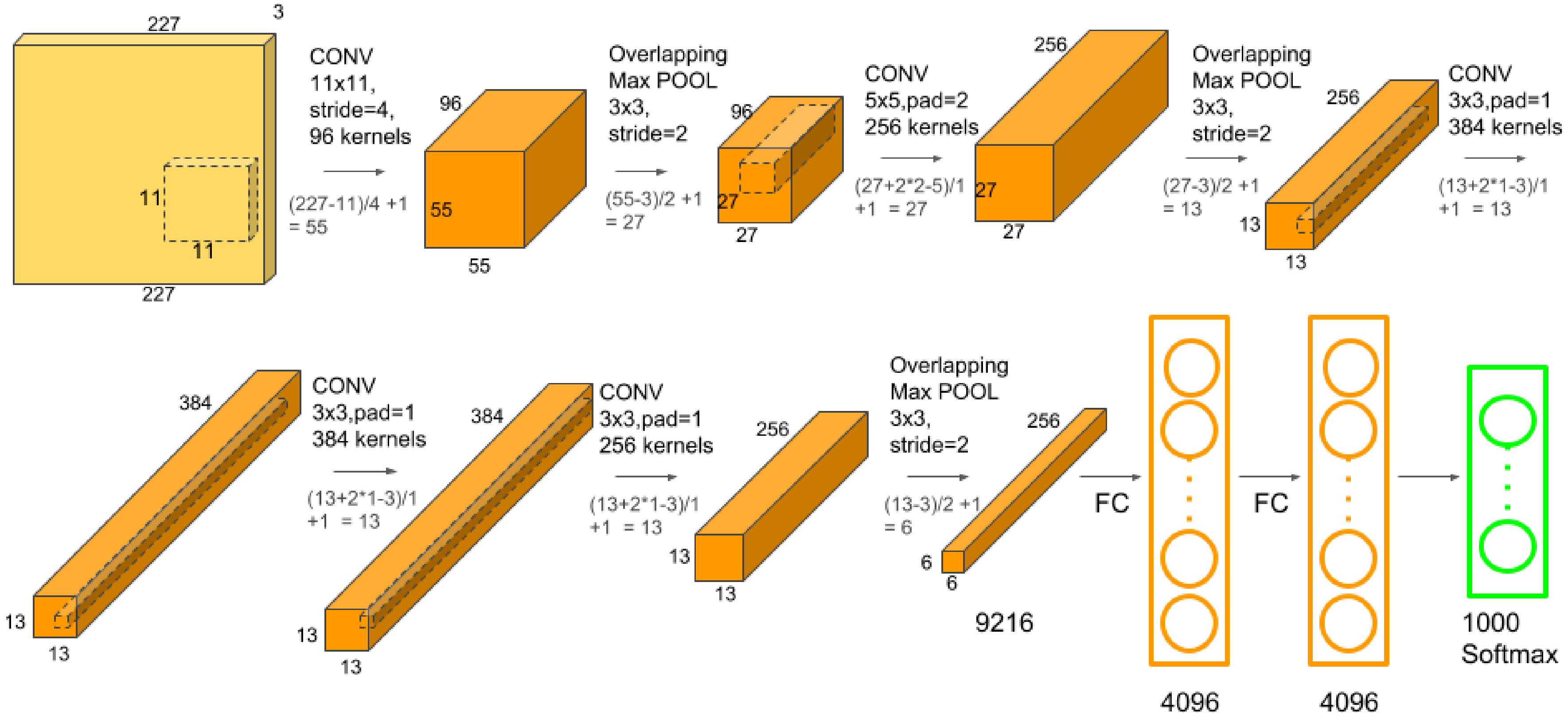
AlexNet : Architecture

GPU-1 : data irrelevant with color



GPU-2 : data relevant with color

AlexNet : Architecture



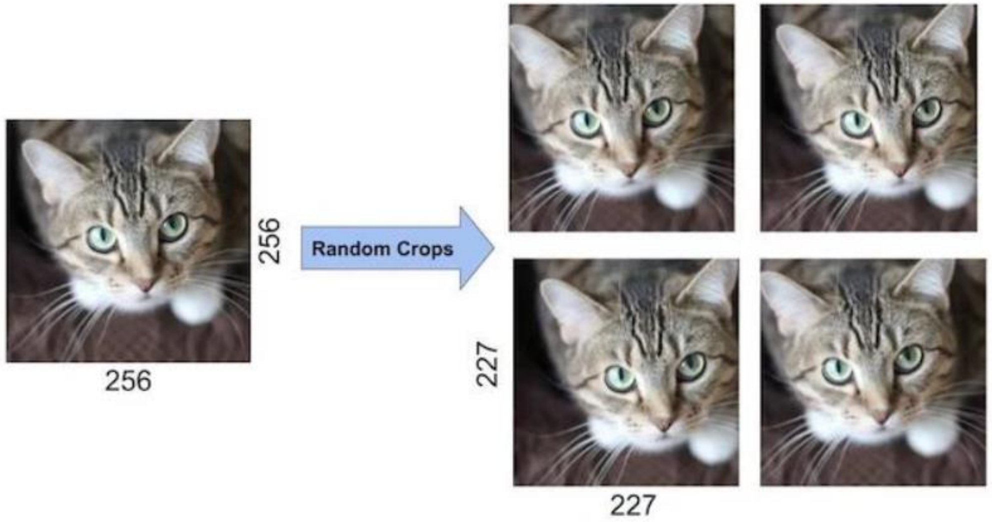
5 Convolution layers + 3 Fully-connected layers

Overfitting

- Data augmentation
  - Mirroring
  - Random crops
  - PCA on RGB pixel values
  - top-1 error rate reduce 1%

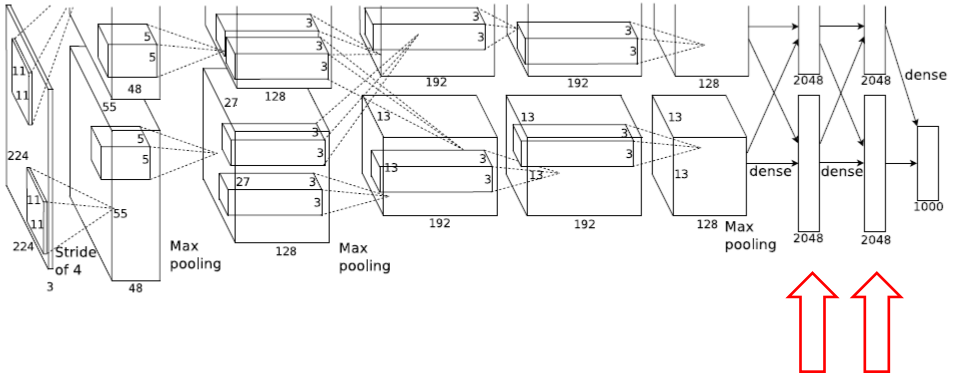
$$I_{xy} = [I^R_{xy},\ I^G_{xy},\ I^B_{xy}]^T + [\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3][\alpha_1\lambda_1, \alpha_2\lambda_2, \alpha_3\lambda_3]^T$$

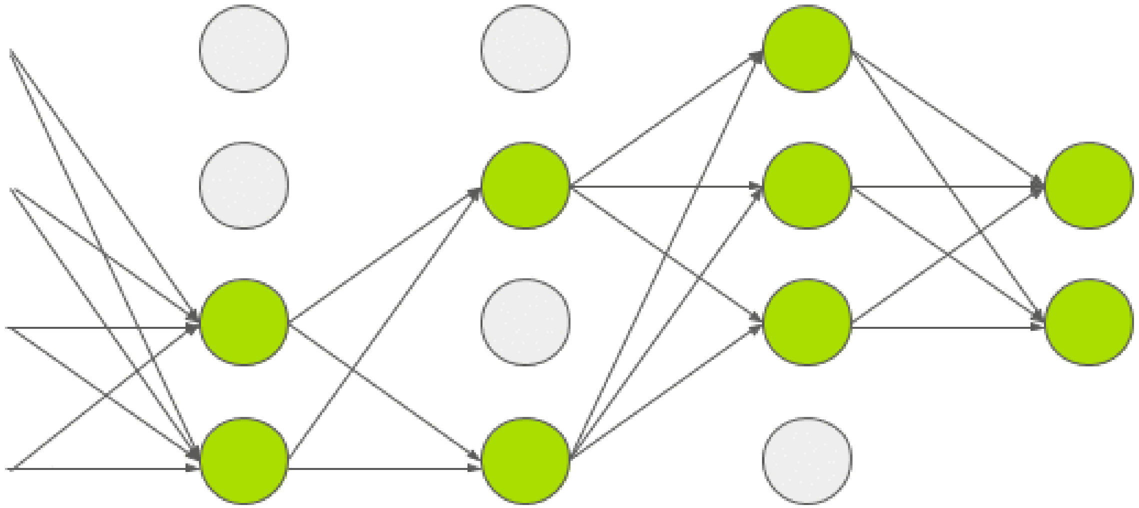
$$\alpha_i \sim N(0,\ 0.1)$$



Overfitting

- Dropout
  - Used in two fully-connected layers





Results

- Best results were achived in ILSVRC-2010

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.

Conclusion

- Successful GPU implementation of the convolution operation
- Efficient result by CNN
- Developed GPU and good architecture will present better performance

Reference

[1] https://blog.naver.com/laonple/220662317927

[2] https://laonple.blog.me/220654387455

[3] https://curaai00.tistory.com/4

[4] https://learnopencv.com/understanding-alexnet/

[5] https://bskyvision.com/421