ImageNet Classification with Deep Convolutional Neural Networks

T11 PangCloud
Tianyang Liu, Su Pu, Yicheng Wang

Datasets of Images

NORB: **194.4k** images

Caltech-101/256: **30.6k** images

CIFAR-10/100: **60k** images

MNIST: **70k** images

LabelMe: **249.4k** images

ImageNet: **14197.1k** images

q1. Which is the largest dataset of images

A. LabelMe

B. CIFAR-10/100

C. ImageNet

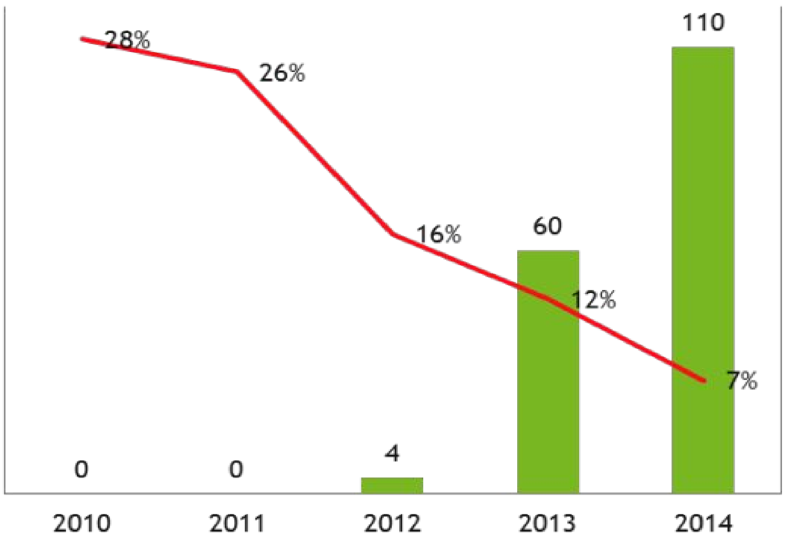
D. Caltech-101/256

ImageNet Dataset

Over 15 million labeled high-resolution images belonging to 22k categories.

ImageNet **L**arge-**S**cale **V**isual **R**ecognition **C**hallenge, an annual competition using a 1.2 million subset of ImageNet of 1000 classes.

This paper in 2012 achieved **15.3%** compared to the second best of **26.2%**.



q2: Which is the competition using ImageNet

A. ACM

B. ILSVRC

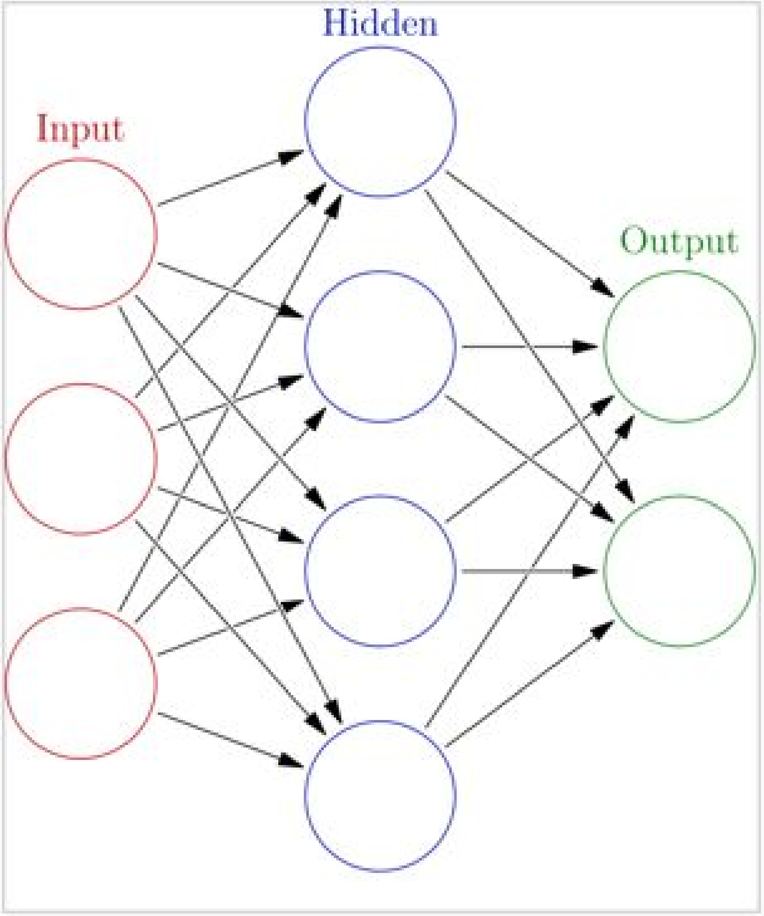
C. MCM

D. TI

Neural Networks

Simulate how brain solves problems.

- Stretch pictures of e.g. 32x32 into 1024x1 as inputs.
- Go through parameters of the hidden neurons.
- Minimize the difference between outputs and expected results, through optimizing hidden parameters.



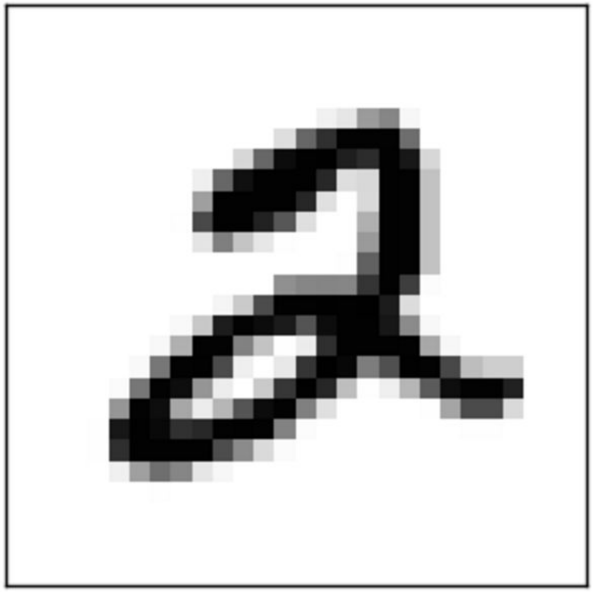
q3: What will be lost if stretch a picture of number

A. Continuity

B. Greylevel

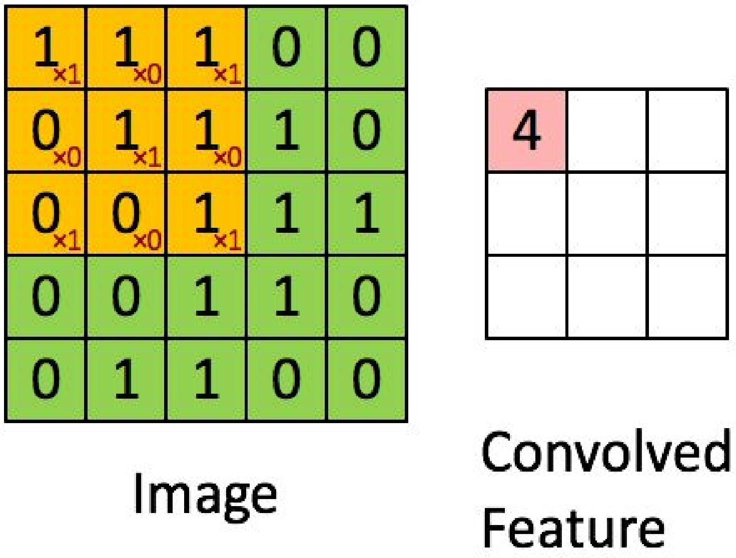
C. Pixels

D. None of above

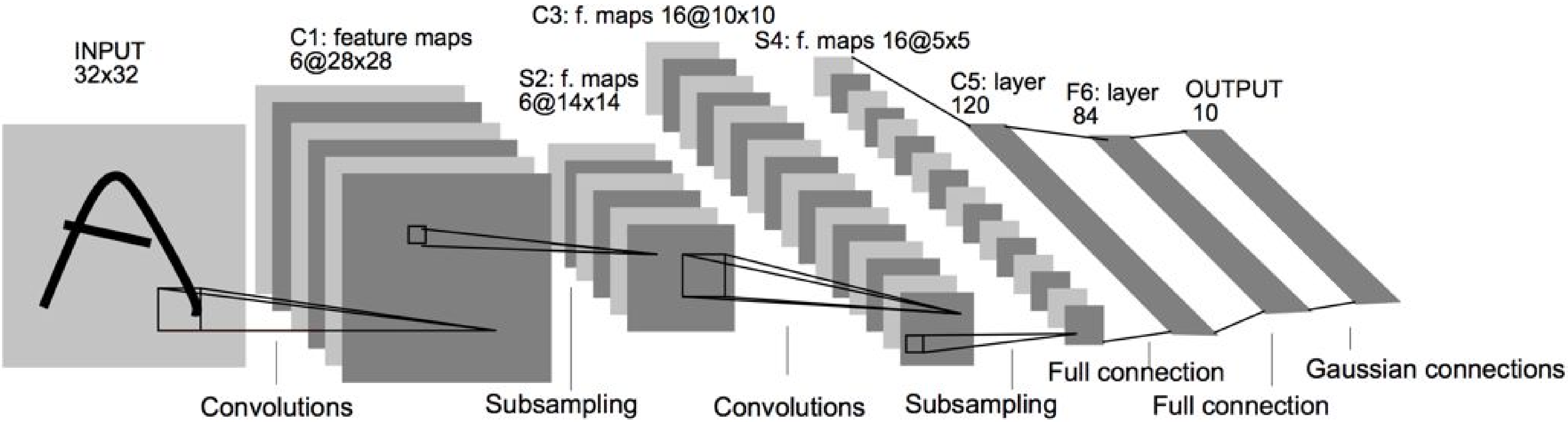


Neural Networks

Convolutional **N**eural **N**etworks extract more information from a picture.



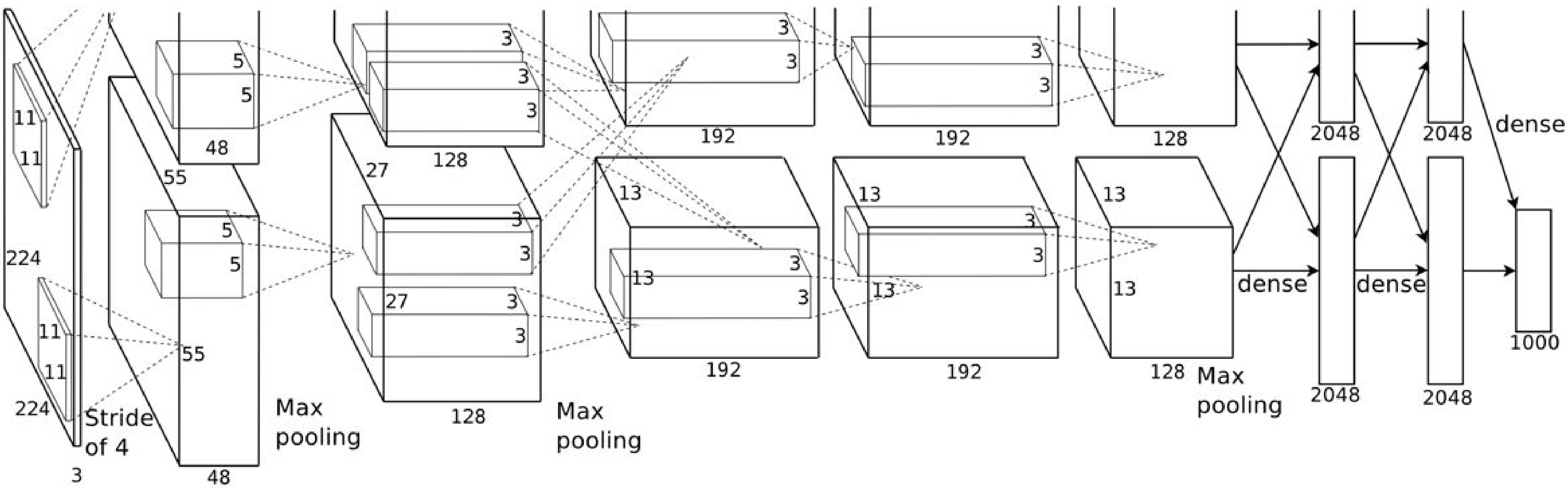
Architecture: Background



LeNet

LeCun, Yann, et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE 86.11 (1998): 2278-2324.

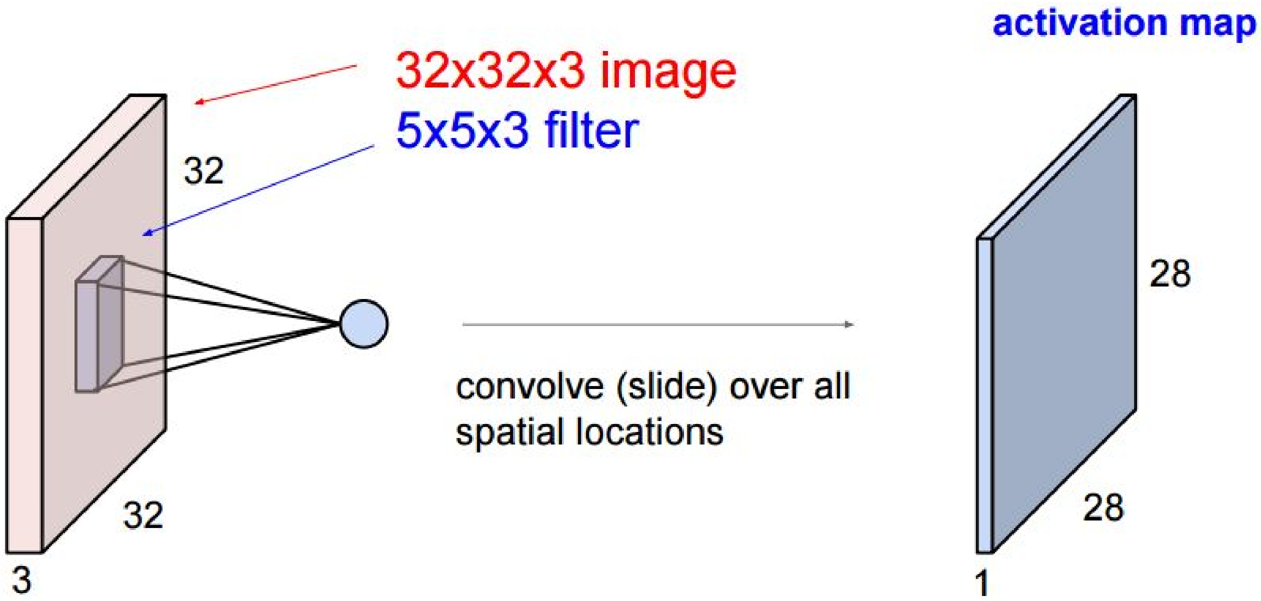
Architecture: Overview

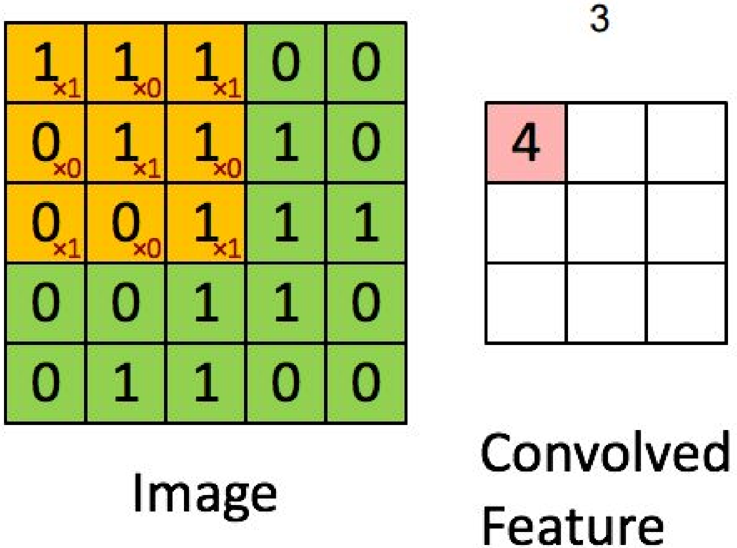


Five convolutional layers, three pooling layers, three fully-connected layers.

Convolution makes the pictures thick to extract more information, while pooling makes the pictures small to save computation.

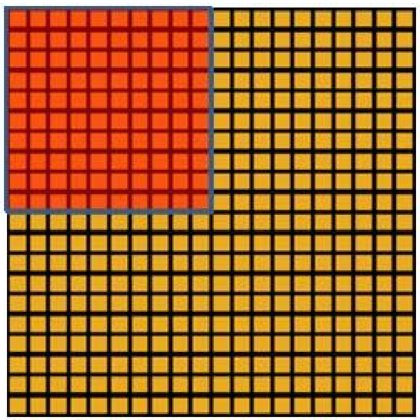
Convolutional layer



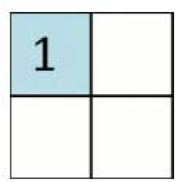


Pooling layer

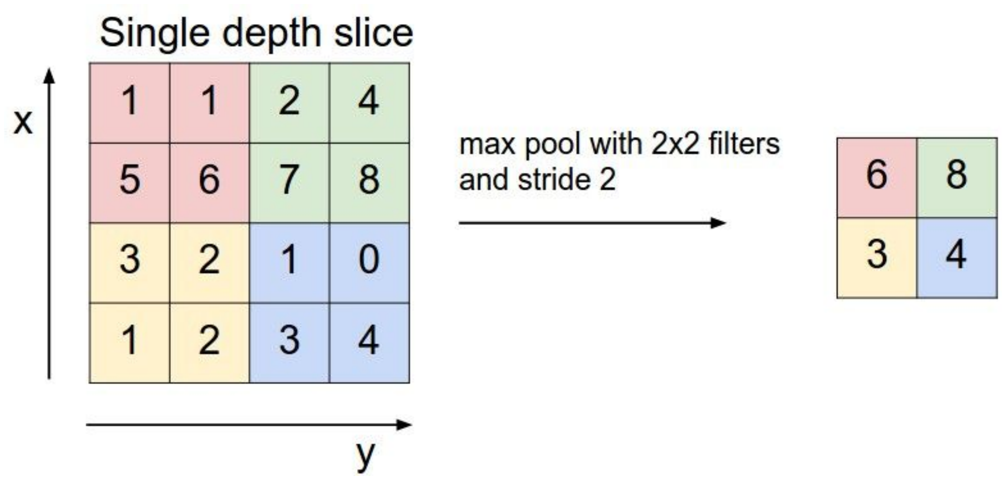
Down sample to make the representations smaller and more manageable



Convolved feature



Pooled feature



Max Pooling

q4: What do convolutional layers do in CNN

A. Extract more information

B. Save computation

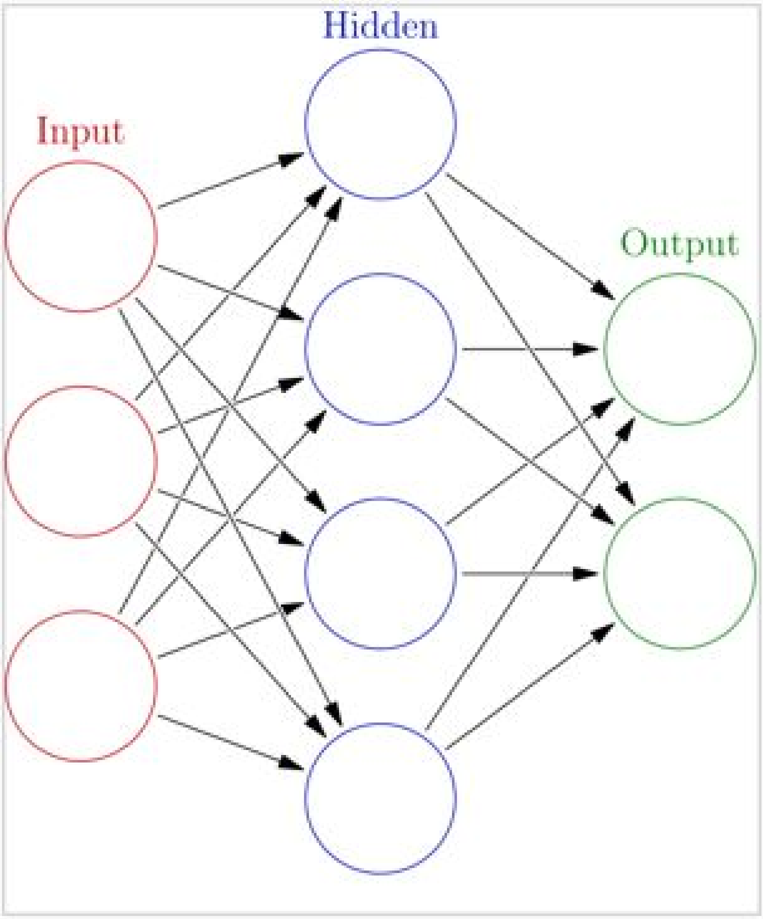
C. Both A and B

D. None of above

Architecture: ReLU

Activation function:

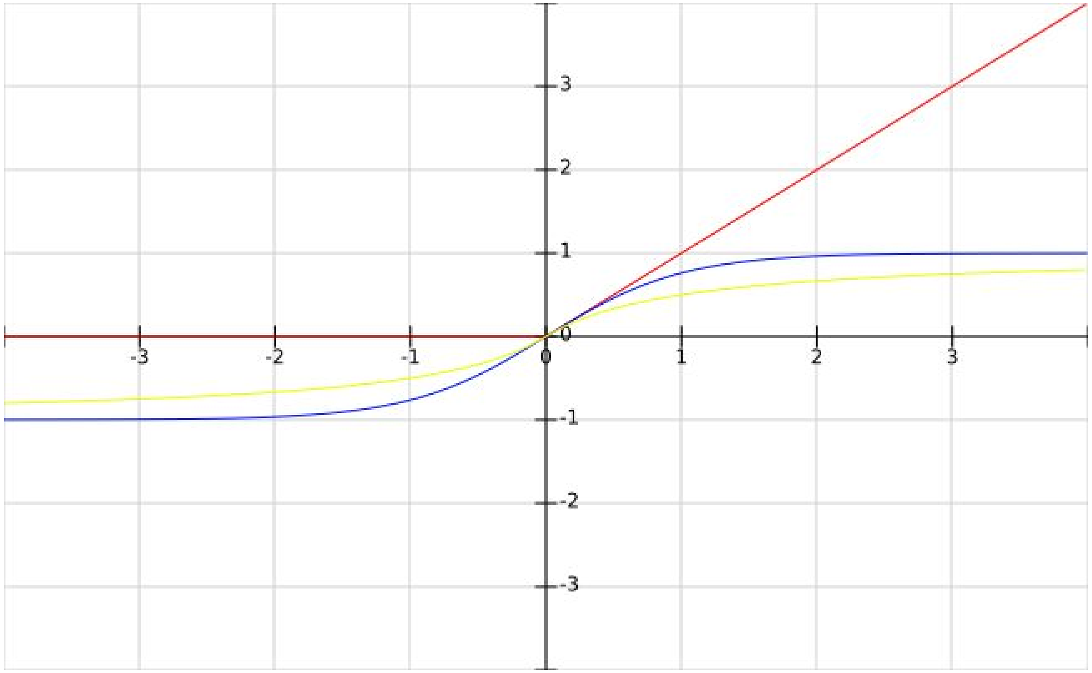
- TanH: tanh(x)
- Softsign: x/(1+|x|)
- **Re**ctified **L**inear **U**nit: Max(0, x)
- and more...



Architecture: ReLU

Advantages of ReLU:

- Fast
- Gradient does not vanish
- Sparse representation



q5: Which is not the advantage of ReLU

A. Sparse representation

B. Fast

C. Gradient does not vanish

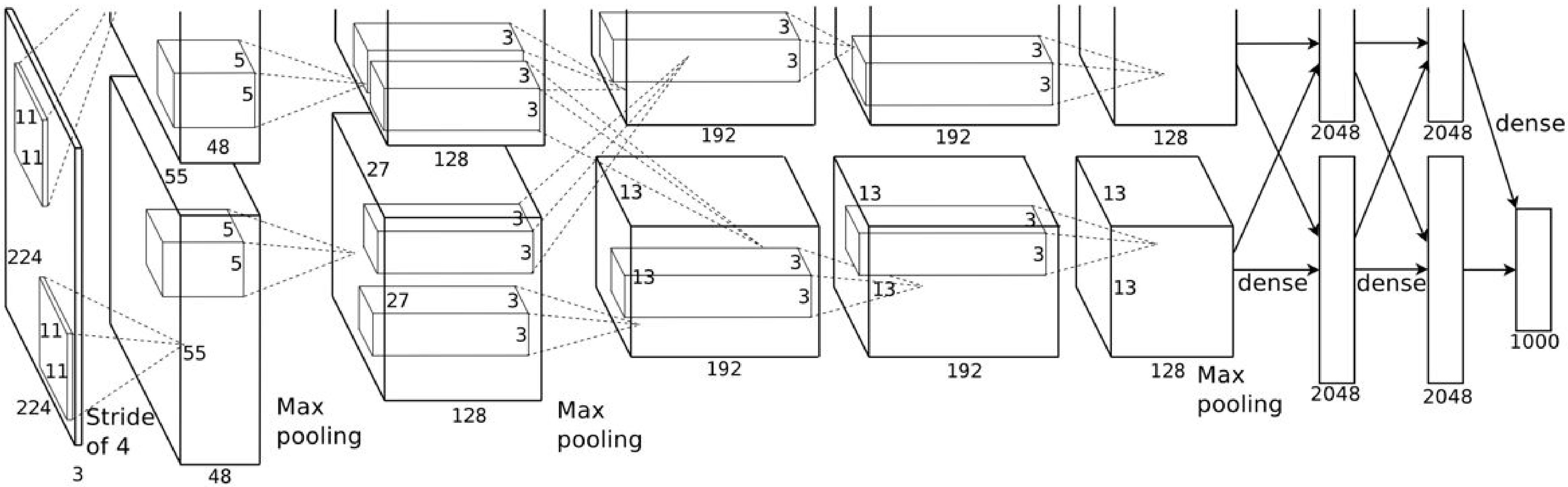
D. Avoid overfitting

Architecture: Multiple GPUs

A single GTX 580 GPU cannot support so large a CNN for 1.2 million pictures.

Spread the CNN across two GTX 580 GPUs, communicate in C3 and F1.

Direct read from and write to another GPU without going through the host.



Architecture: Local Normalization

ReLU does not require input normalization to prevent them from saturating.

Still find the local normalization scheme aids generalization.

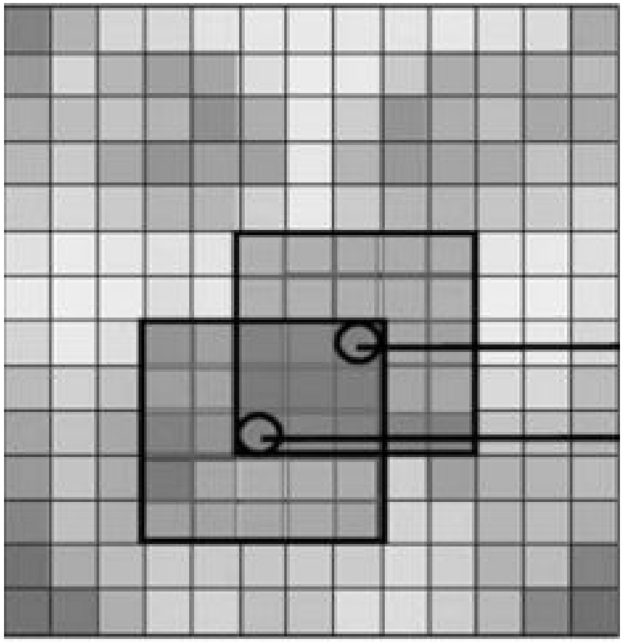
$$b^i_{x,y} = a^i_{x,y} / (t + a \sum_{j=max(0,i-n/2)}^{min(N-1,i+n/2)} (a^j_{x,y})(a^j_{x,y}))^\beta$$

Architecture: Overlapping Pooling

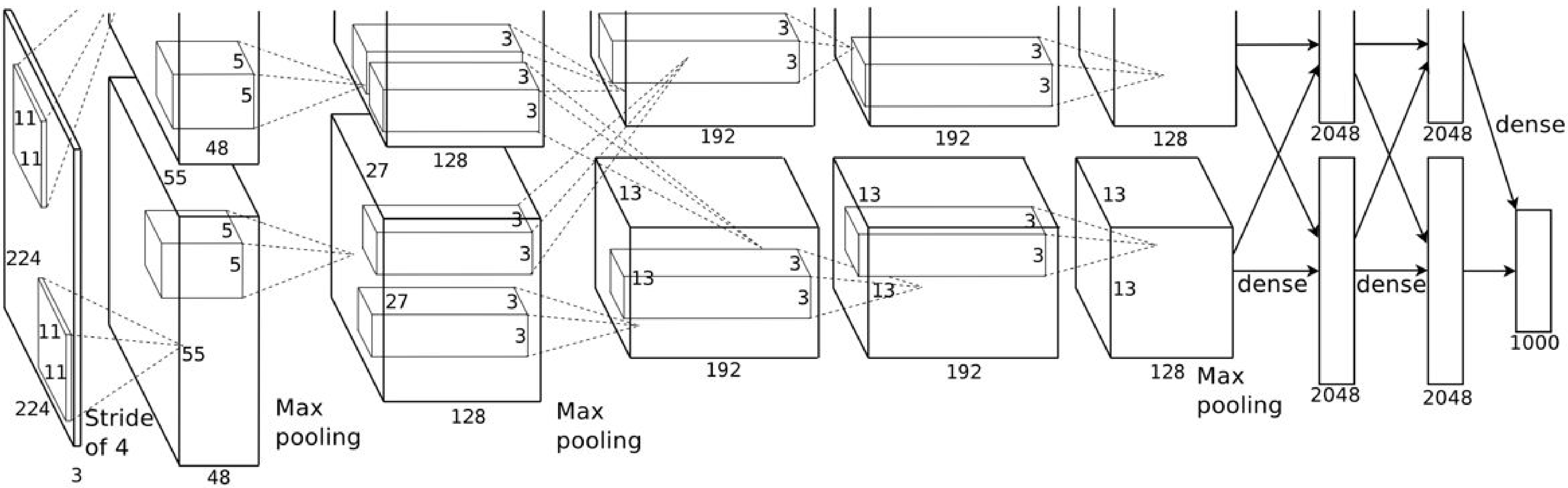
Does pooling on a picture of 6x6 using a 3x3 pooling block.

If non-overlapping it results in a (6/3)x(6/3) i.e. 2x2 picture.

If overlapping it results in a (6-3+1)x(6-3+1) i.e. 4x4 picture.



Architecture: Overview Again



Multiple GPUs, ReLU, local normalization, overlapping pooling.

q6: What is not included in the CNN

A. Sigmoid

B. Fully-connected layer

C. Normalization

D. Pooling layer

Overfitting

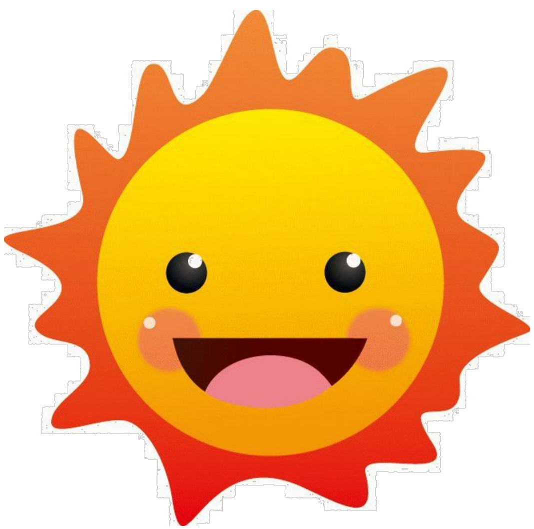
60 million parameters to train but not enough data.

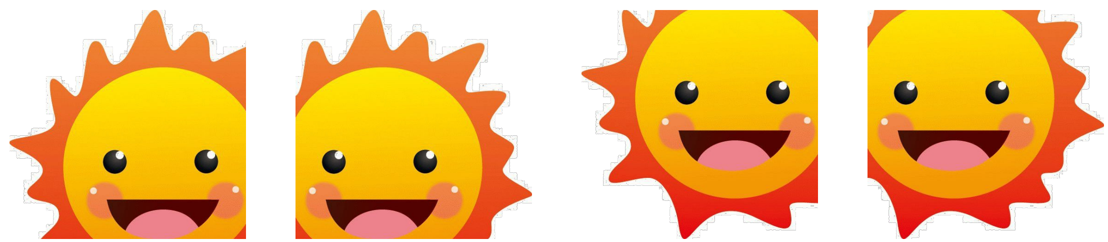
- Data augmentation
- Dropout

Overfitting: Data Augmentation

Method 1:

Generating image translations and horizontal reflections. E.g. extracting random 224x224 patches from a 256x256 image, at most increase with a factor of (256-224)x(256-224)x2=2048.





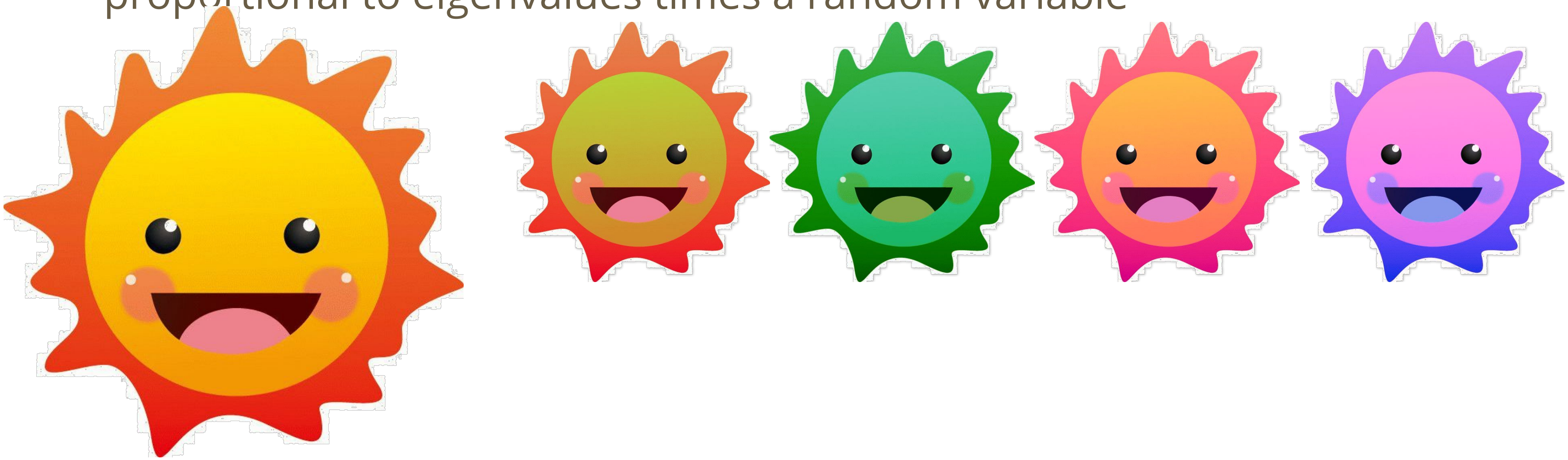
up-left up-right down-left down-right

Overfitting: Data Augmentation

Method 2:

Altering the intensities of the RGB channels in training images.

- Perform PCA on the set of RGB pixel values throughout the training set
- To each image add multiples of principle components, magnitudes proportional to eigenvalues times a random variable



q7: Target of data augmentation is

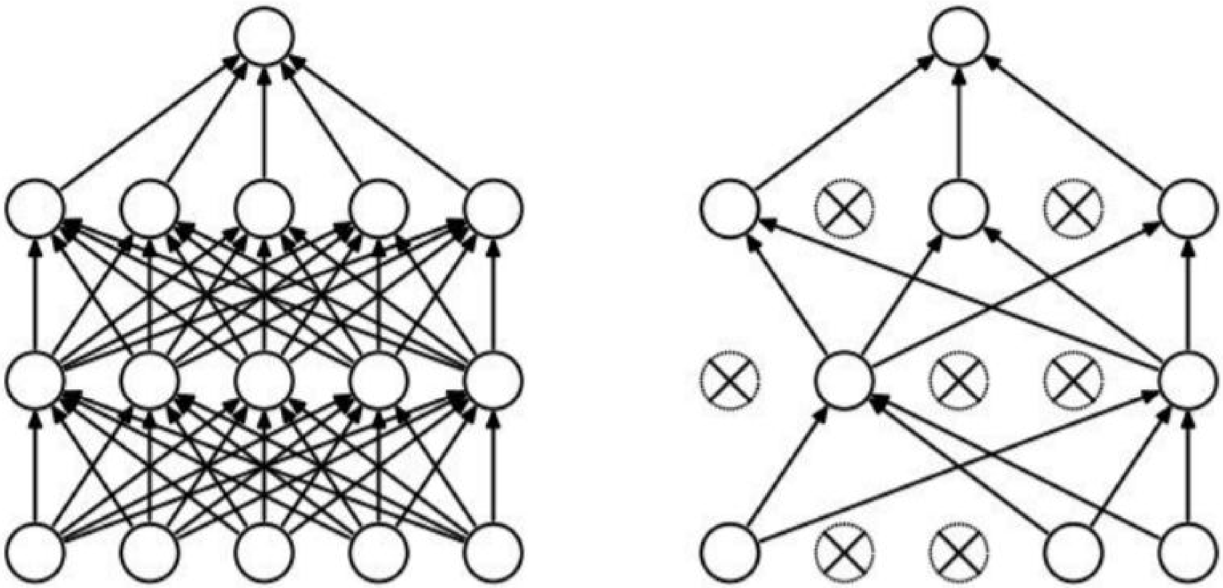
A. Increase overfitting

B. Reduce overfitting

C. Increase computation

D. Decrease computation

Overfitting: Dropout



(a) Standard Neural Net (b) After applying dropout.

Setting to zero the output of each hidden neuron with a 50% chance.

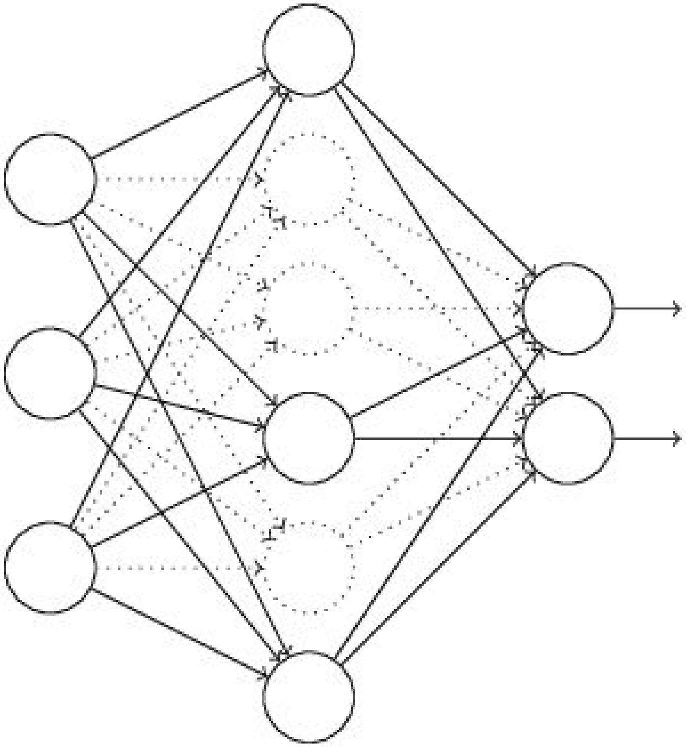
The dropped out neurons do not contribute to the forward pass and do not participate in back-propagation.

Overfitting: Dropout

When training without dropout, the common practice is to use same training data to train several seperate neural network, then take the average values as the final structure.

which is similar to training with dropout.

Dropout almost doubles the number of iterations required to converge.



q8: Which requires more iterations to converge

A. With dropout

B. Without dropout

C. Not sure

D. Same iterations

Demo!

Thank you!

T11 PangCloud