NPTEL ONLINE CERTIFICATION COURSES

Course Name: Deep Learning

Faculty Name: Prof. P. K. Biswas

Department : E & ECE, IIT Kharagpur

Topic

Lecture 38: Popular CNN Models II

CONCEPTS COVERED

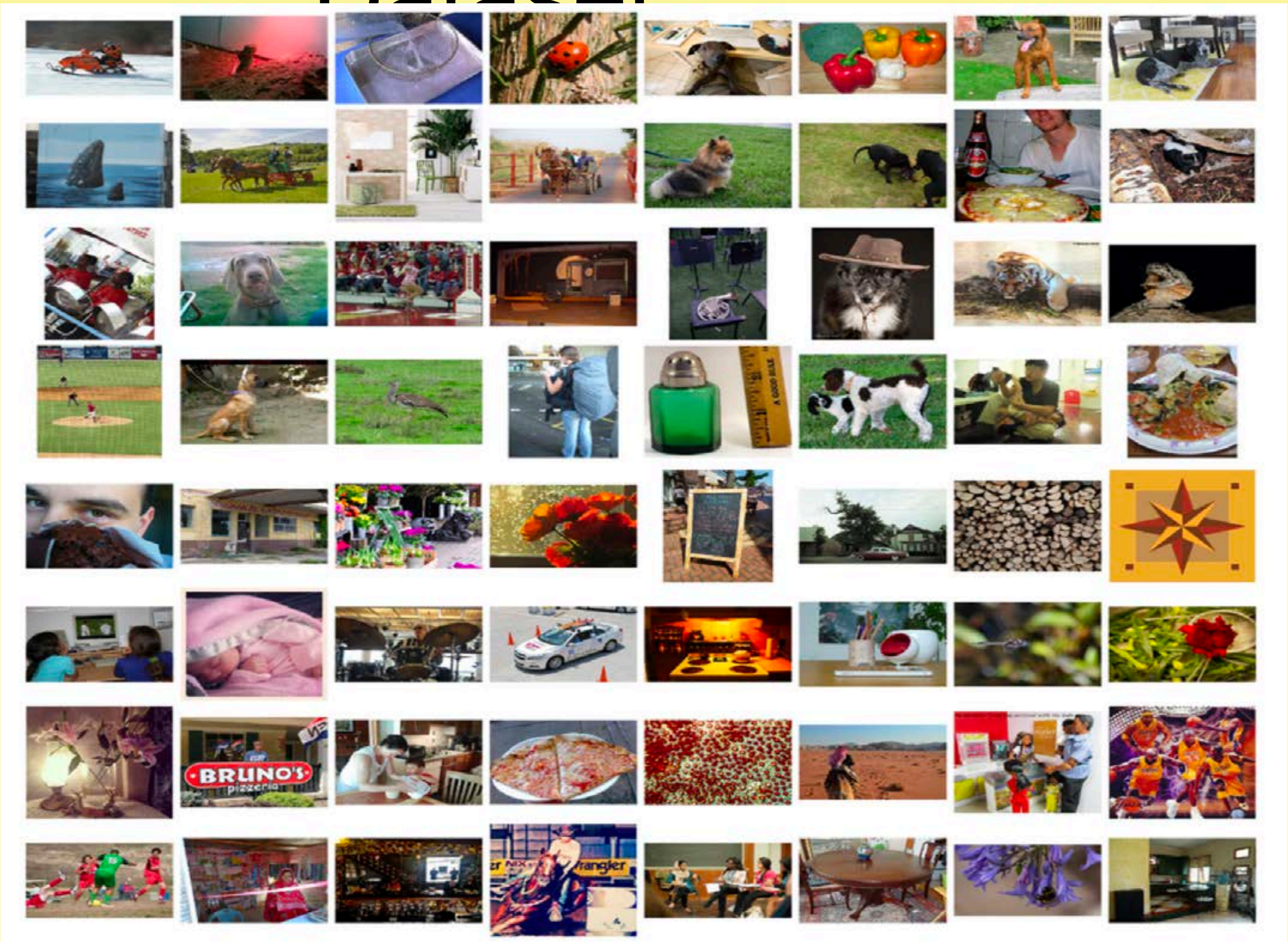
Concepts Covered:

- CNN
  - LeNet
  - ILSVRC
  - AlexNet
  - VGG Net
  - GoogLeNet
  - etc.

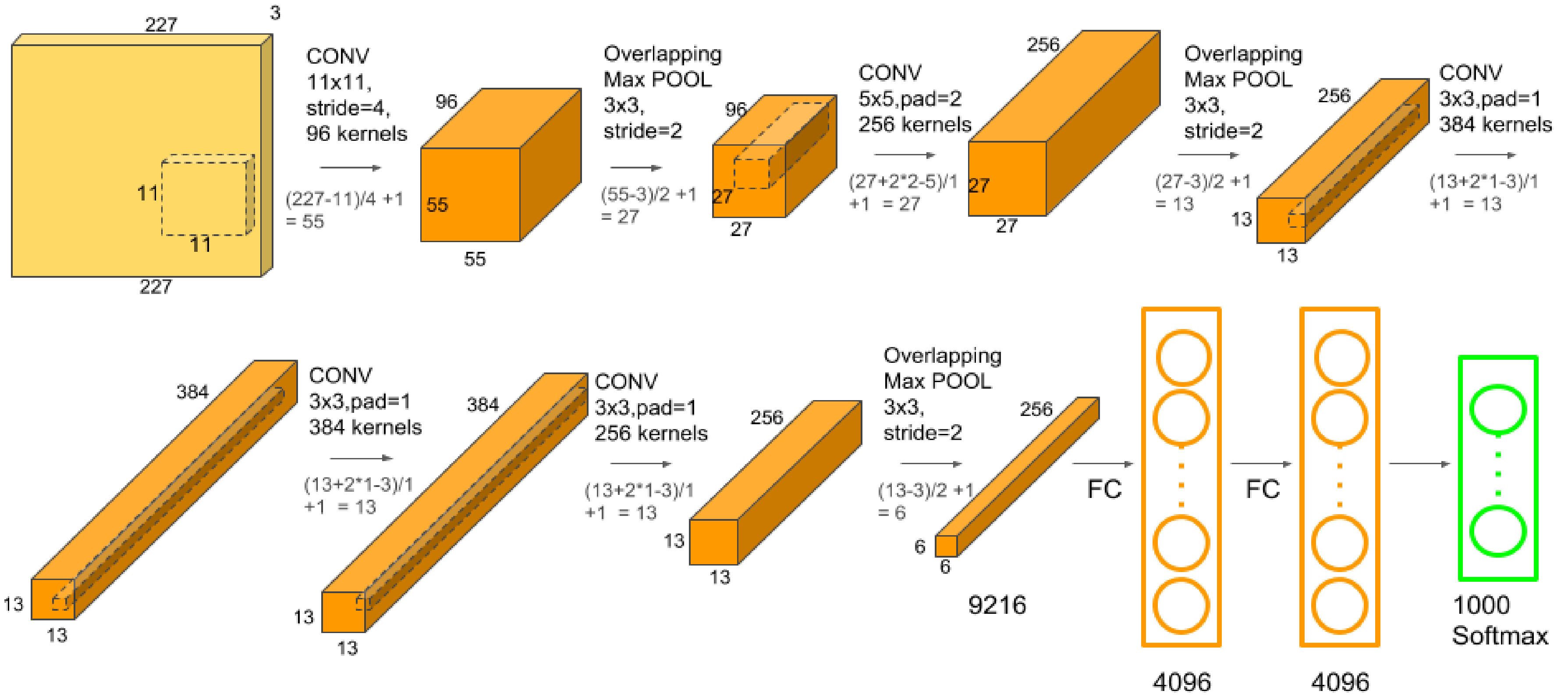
AlexNet
ILSVRC 2012 Winer

Krizhevsky Alex, Ilya Sutskever and Geoffrey E. Hilton, “Imagenet Classification with deep convolutional neural networks”, Advances in Neural Information Processing Systems, 2012

Sample Images from ImageNet Dataset



AlexNet

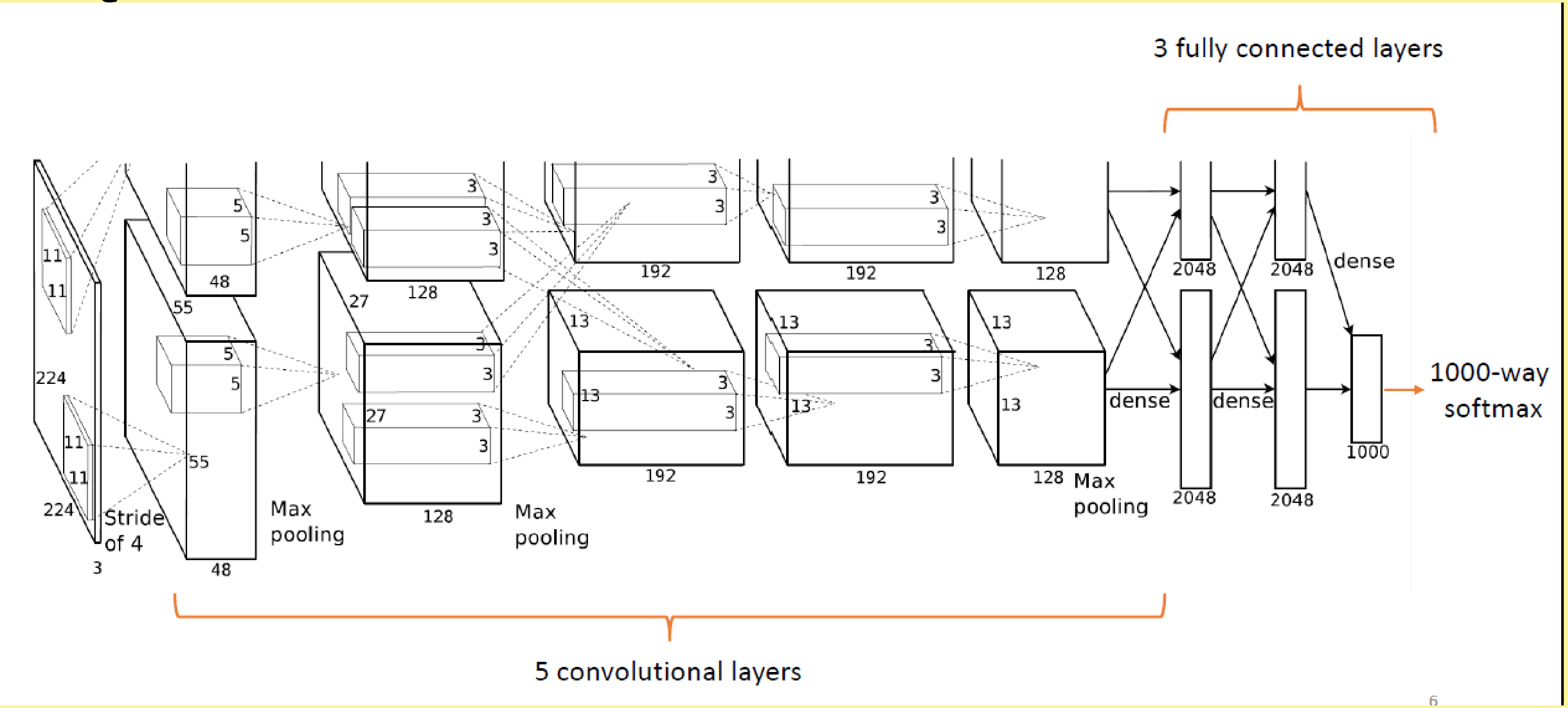


https://www.learnopencv.com/understanding-alexnet/

ILSVRC 2012
Winner

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

AlexNet



Krizhevsky Alex, Ilya Sutskever and Geoffrey E. Hilton, “Imagenet Classification with deep convolutional neural networks”, Advances in Neural Information Processing Systems, 2012

AlexNet

- 60 Million parameters and 650000 neurons.
- The network is split into two pipelines and was trained on two GPU.
- Input Image size 256 x 256 RGB.
- Grey scale images to be replicated to obtain 3-Channel RGB
- Random crops of size 227 x 227 are fed to the input layer of AlexNet.
- Stochastic Gradient Descent with Momentum Optimizer.
- Top-5 error rate 15.3%.

Krizhevsky Alex, Ilya Sutskever and Geoffrey E. Hilton, “Imagenet Classification with deep convolutional neural networks”, Advances in Neural Information Processing Systems, 2012

Vanishing Gradient Problem

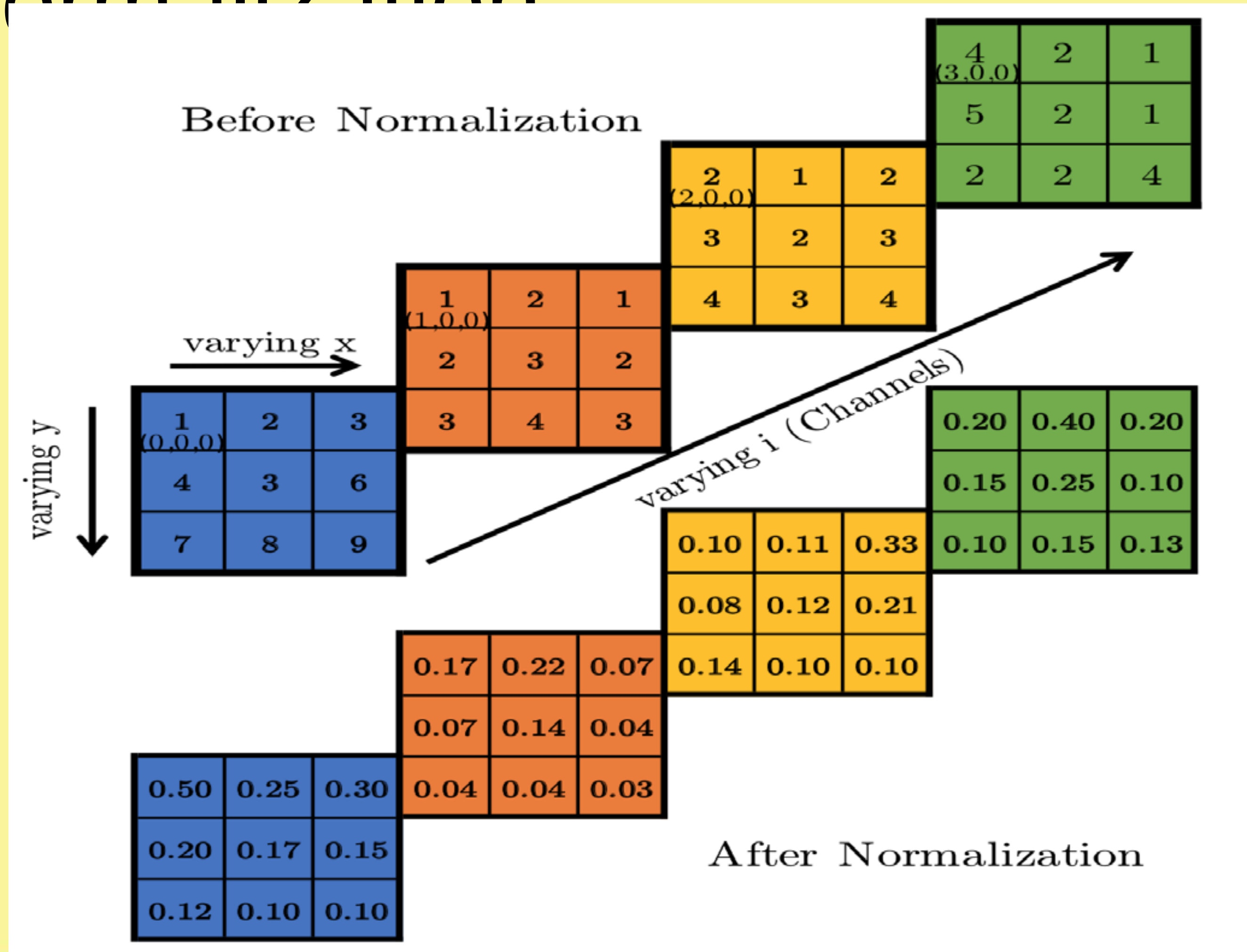
- Uses ReLU activation instead of sigmoidal function.
- ReLU output is unbounded- uses Local Response Normalization (LRN).
- LRN carries out a normalization amplifying the excited neuron while dampening the surrounding neurons at the same time in a local neighbourhood.
- Encourage *Lateral Inhibition*: concept in neuro biology that indicates capacity of a neuron to reduce activity of its neighbours.

Krizhevsky Alex, Ilya Sutskever and Geoffrey E. Hilton, "Imagenet Classification with deep convolutional neural networks", Advances in Neural Information Processing Systems, 2012

Local Response Normalization (Inter-Channel)

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

Local Response Normalization

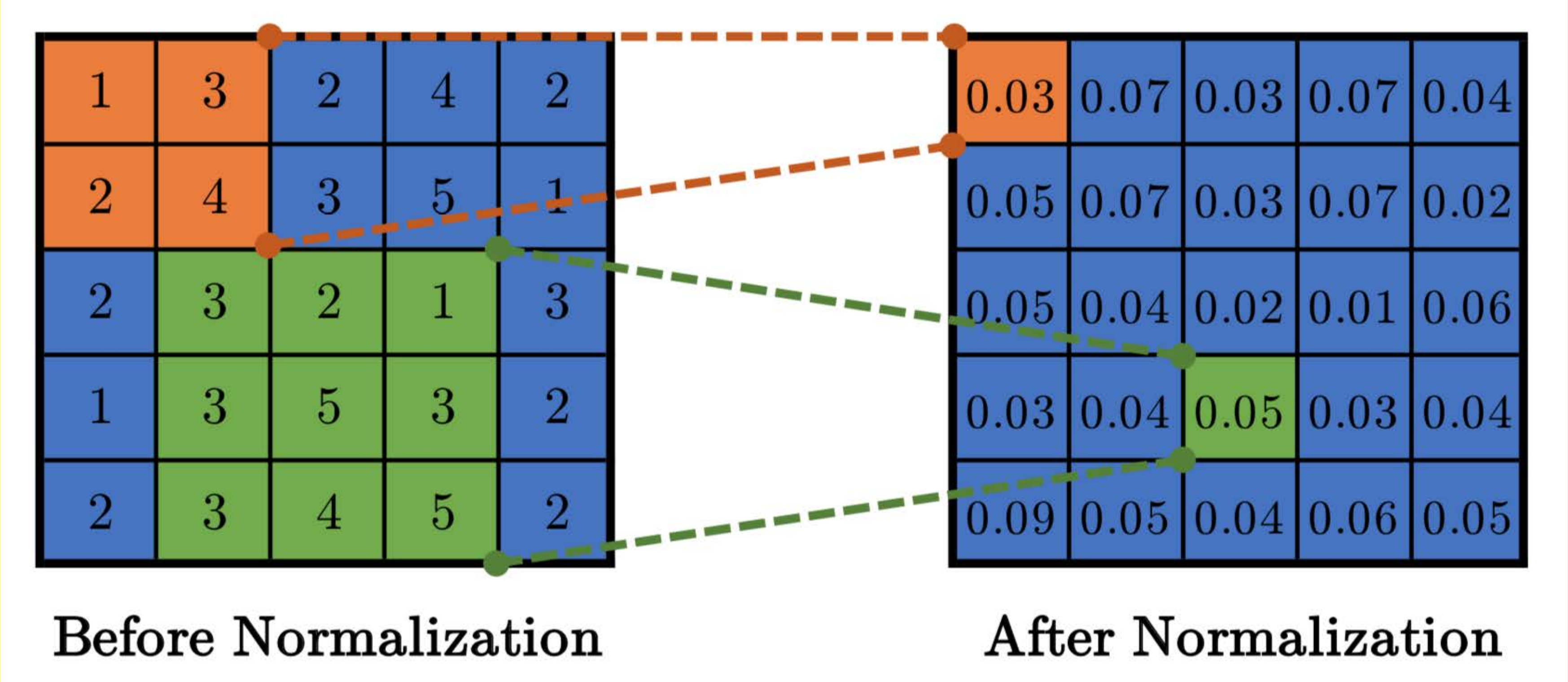


https://towardsdatascience.com/difference-between-local-response-normalization-and-batch-normalization-272308c034ac

Local Response Normalization (Intra-Channel)

$$b_{x,y}^{i} = \frac{a_{x,y}^{i}}{\left(k + \alpha \sum_{p=\max(0,x-n/2)}^{\max(W,x+n/2)} \sum_{q=\max(0,y-n/2)}^{\min(H,y+n/2)} \left(a_{p,q}^{i}\right)^2\right)^{\beta}}$$

Local Response Normalization



https://towardsdatascience.com/difference-between-local-response-normalization-and-batch-normalization-272308c034ac

Reducing Overfitting

- Train the network with different variants of the same image helps avoiding overfitting.
  - Generate additional data from existing data (Augmentation).
  - Data augmentation by mirroring.
  - Data Augmentation by random crops.
- Dropout Regularization.

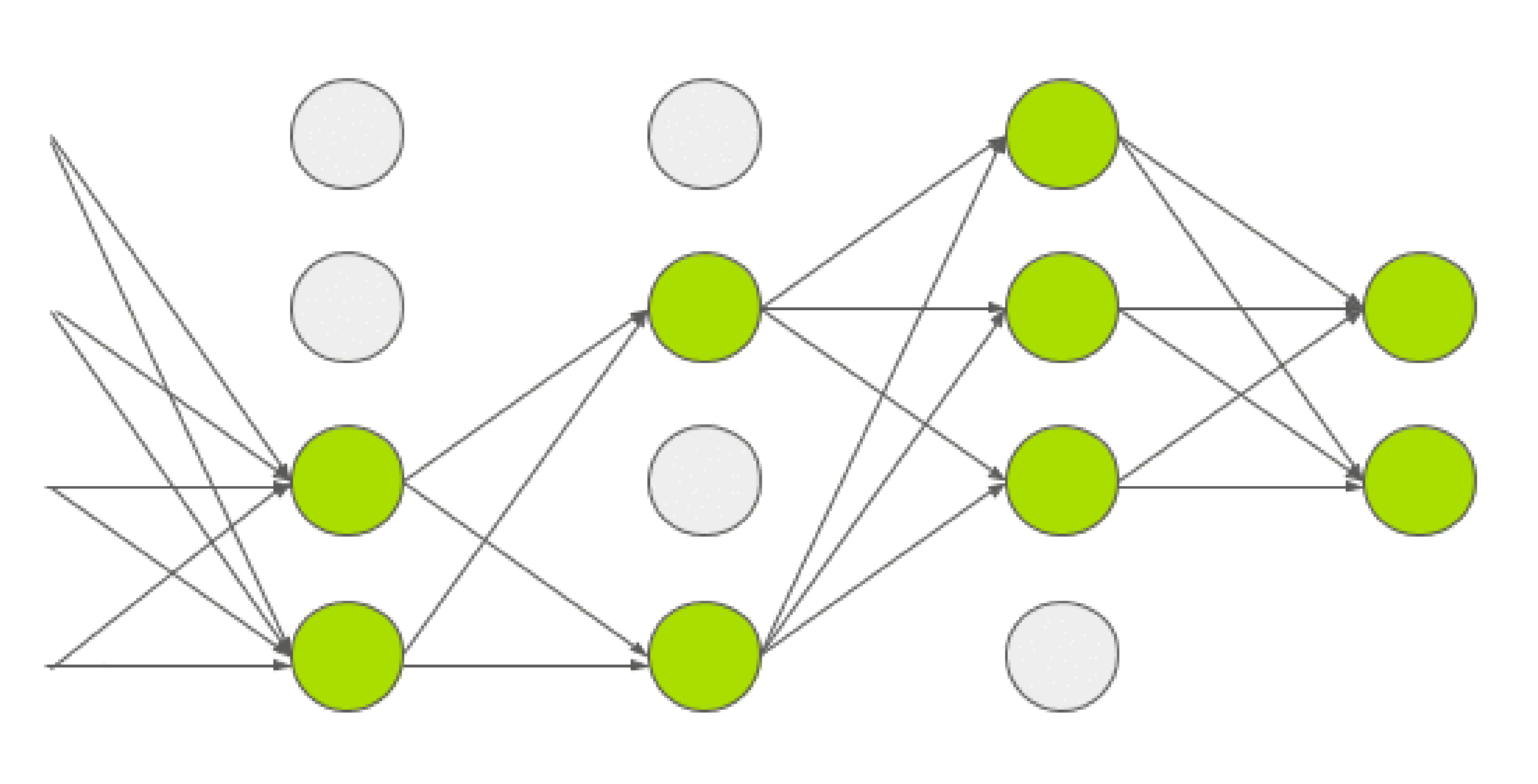
Krizhevsky Alex, Ilya Sutskever and Geoffrey E. Hilton, “Imagenet Classification with deep convolutional neural networks”, Advances in Neural Information Processing Systems, 2012

Dropout

- Regularization Technique proposed by Srivastava et. al. in 2014.
- During training randomly selected neurons are dropped from the network (with probability 0.5) temporarily .
- Their activations are not passed to the downstream neurons in the forward pass.
- In the backward pass weight updates are not applied to theses neurons.

Srivastava Nitish et. al. “Dropout: A Simple Way to Prevent Neural Networks from Overfitting” Journal of Machine Learning Research 15 (2014), 1929-1958

Dropout

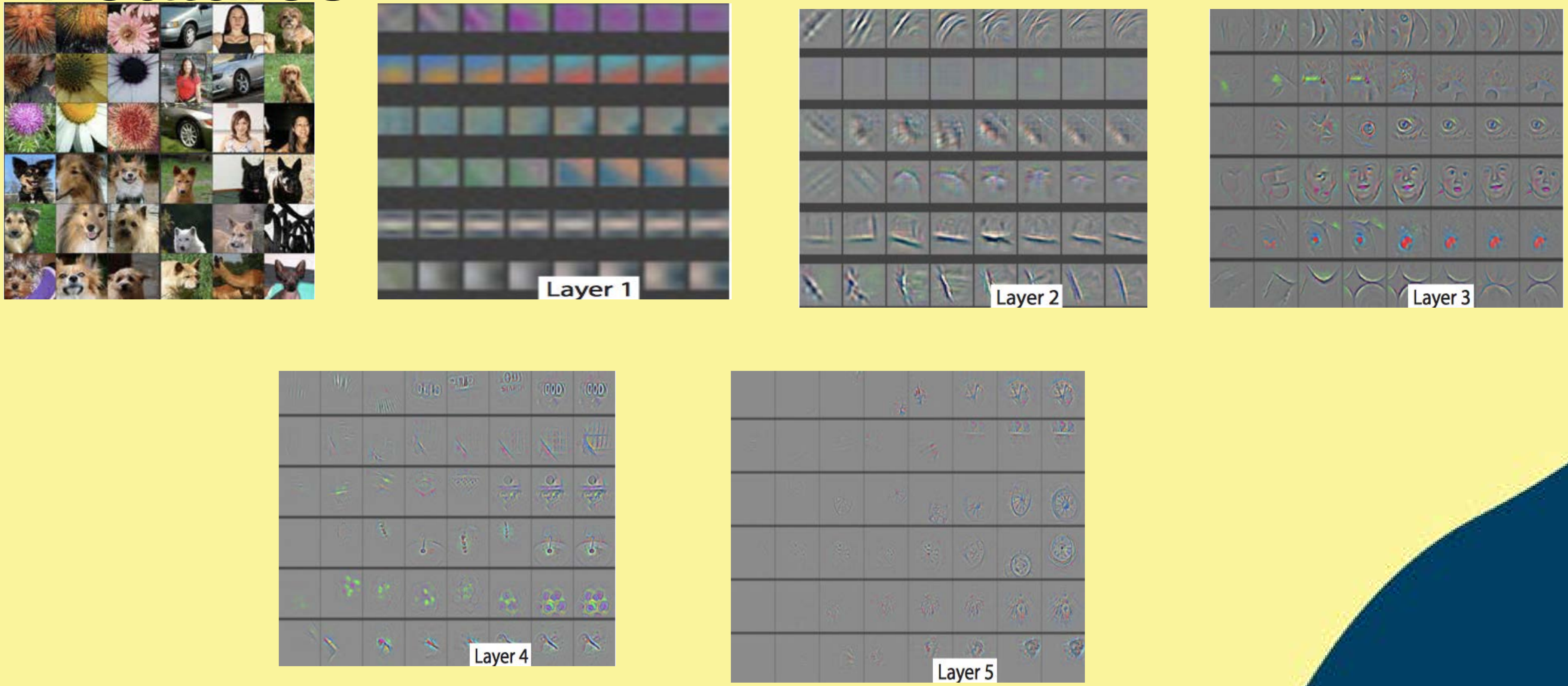


https://www.learnopencv.com/understanding-alexnet/

How does it help?

- While training weights of neurons are tuned for specific features that provides some sort of specialization.
- Neighbouring neurons starts relying on these specializations (co-adaptation).
- This leads to a neural network model too specialized to the training data.
- As neurons are randomly dropped other neurons have to step in to compensate.
- Thus the network learns multiple independent representations

Learned Features



How does it help?

- This makes the network less sensitive to specific weights.
- Enhances the generalization capability of the network
- Less vulnerable to overfitting.
- The whole network is used during testing – there is no dropout.
- Dropout increases number of iterations for the network to converge.
- But helps avoid overfitting.

Srivastava Nitish et. al. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting" Journal of Machine Learning Research 15 (2014), 1929-1958

NPTEL ONLINE CERTIFICATION COURSES

Thank you