22 June 2020

Machine Learning for Image Analysis

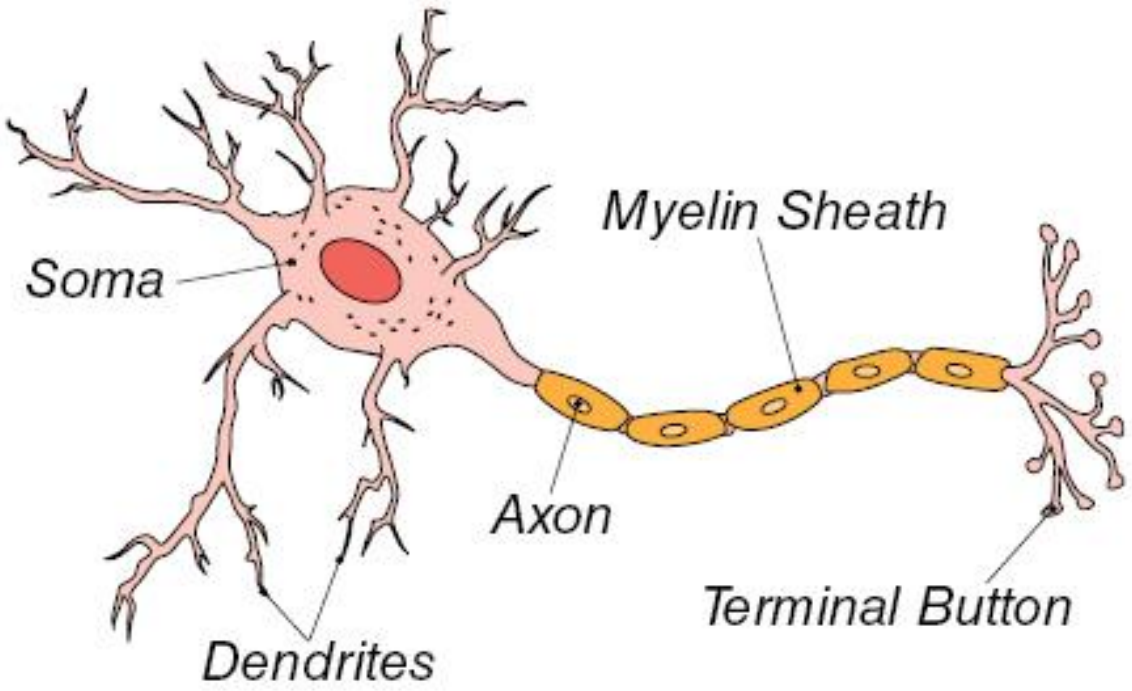
Niclas Danielsson, Axis Communications, Lund

Deep Learning for Video and Audio

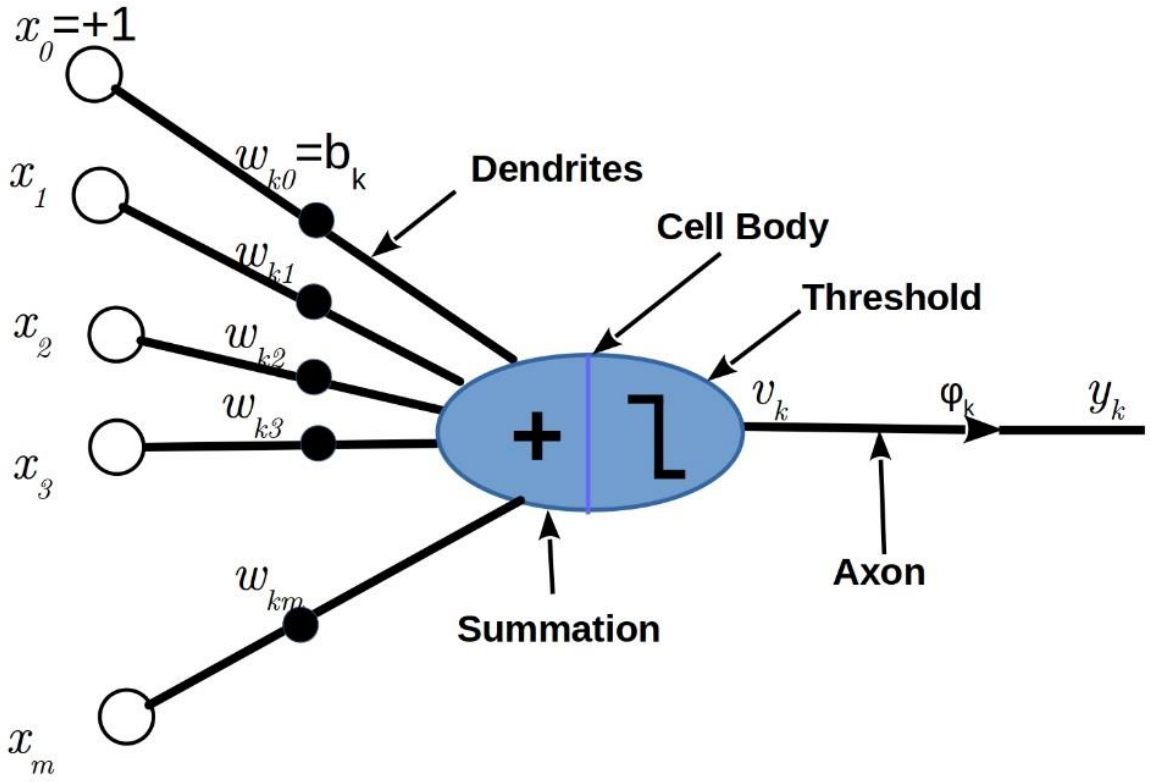
Real-time Object Detection with Deep Learning

https://www.youtube.com/watch?v=VOC3huqHrss

Brain neuron

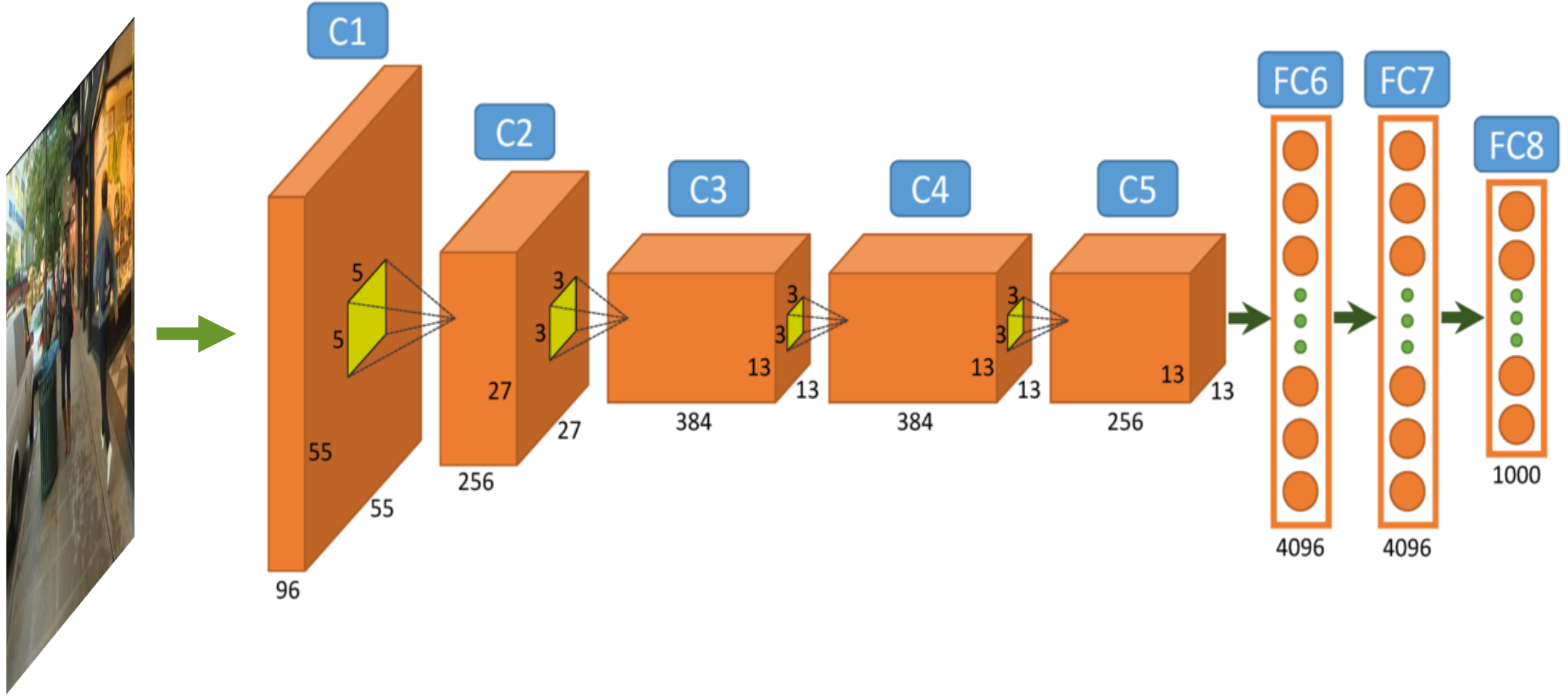


Artificial neuron



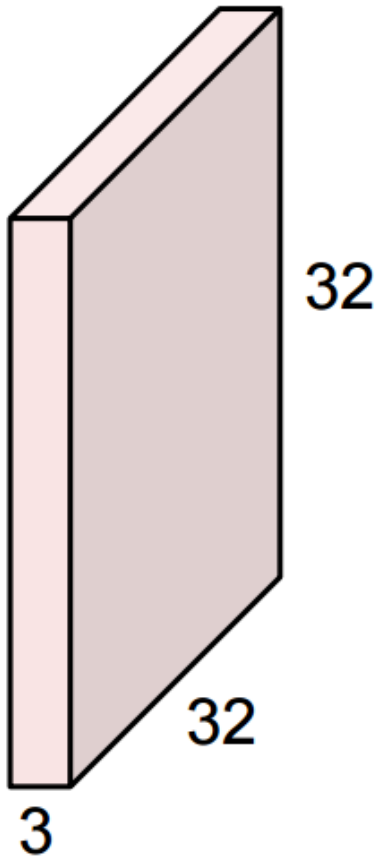
Brief recap of Neural Networks

Overview of a typical architecture

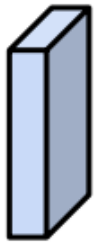


Filters always extend the full depth of the input volume

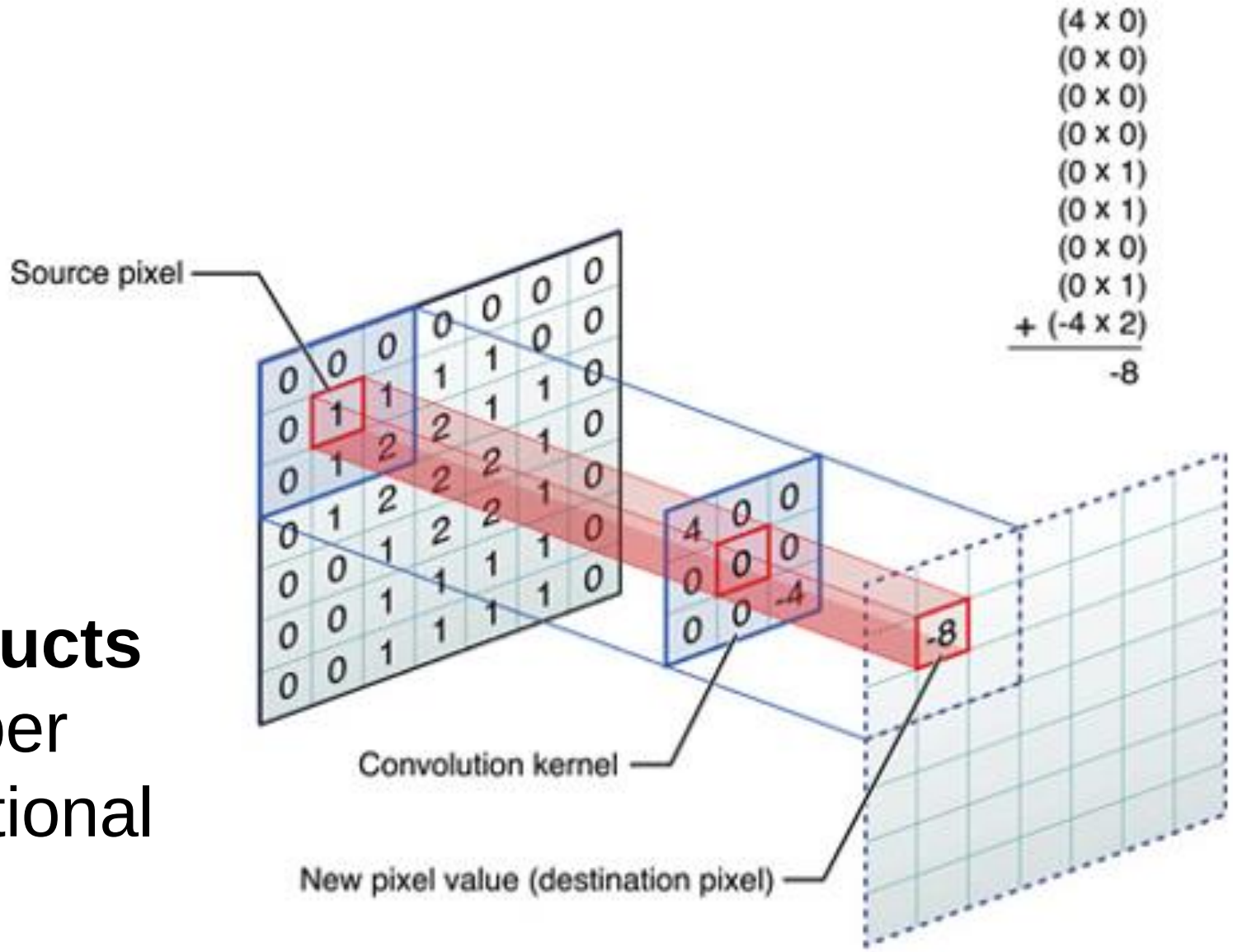
32x32x3 image



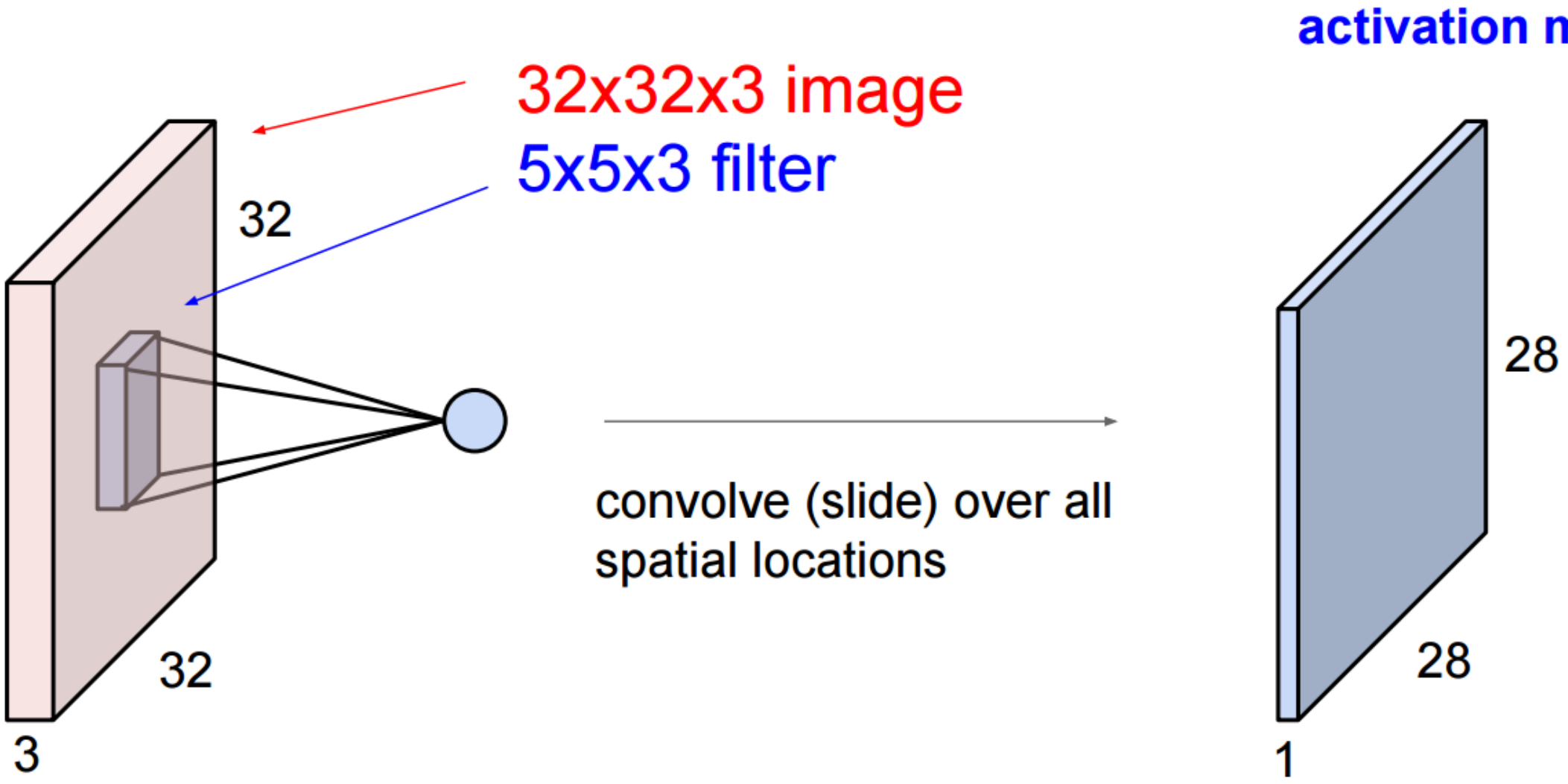
5x5x3 filter



Calculating **dot products** is the essential number crunching in convolutional networks!

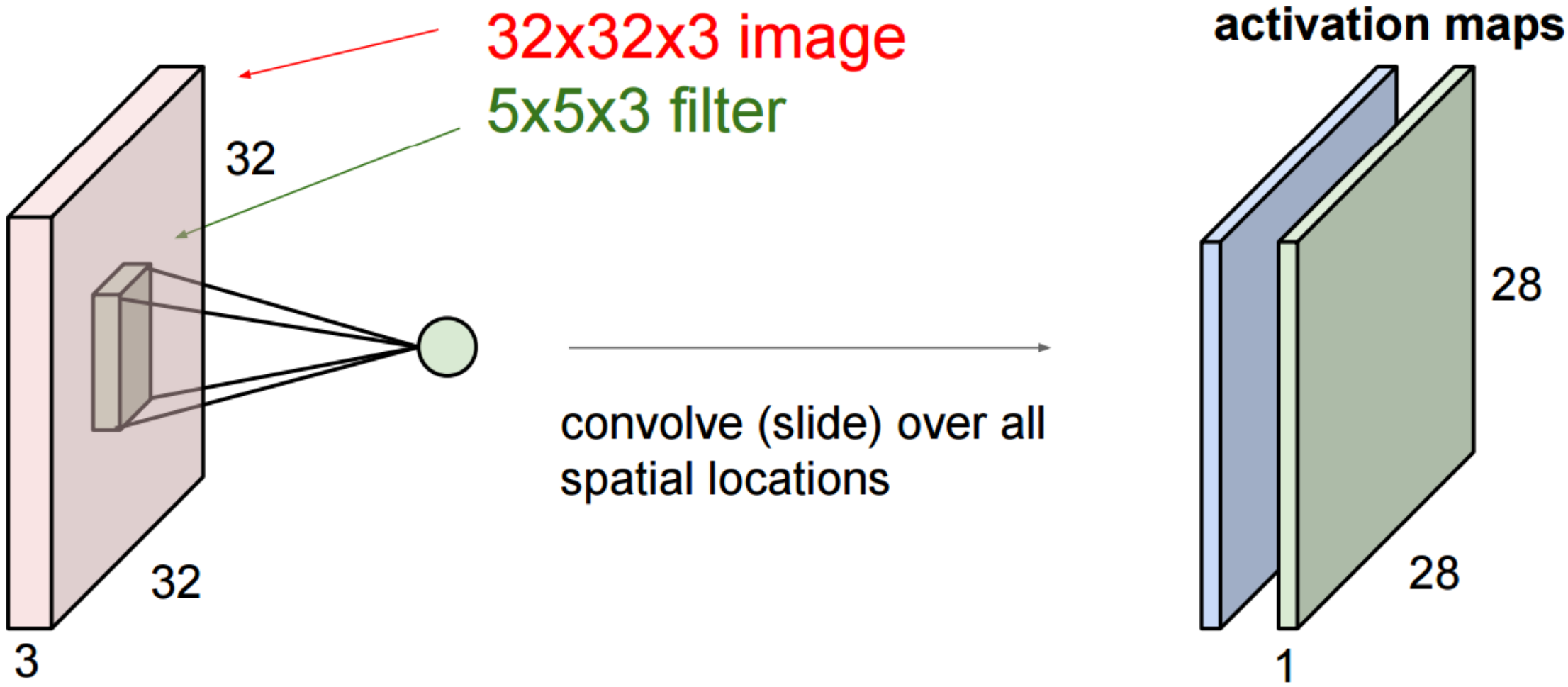


Convolution Layer

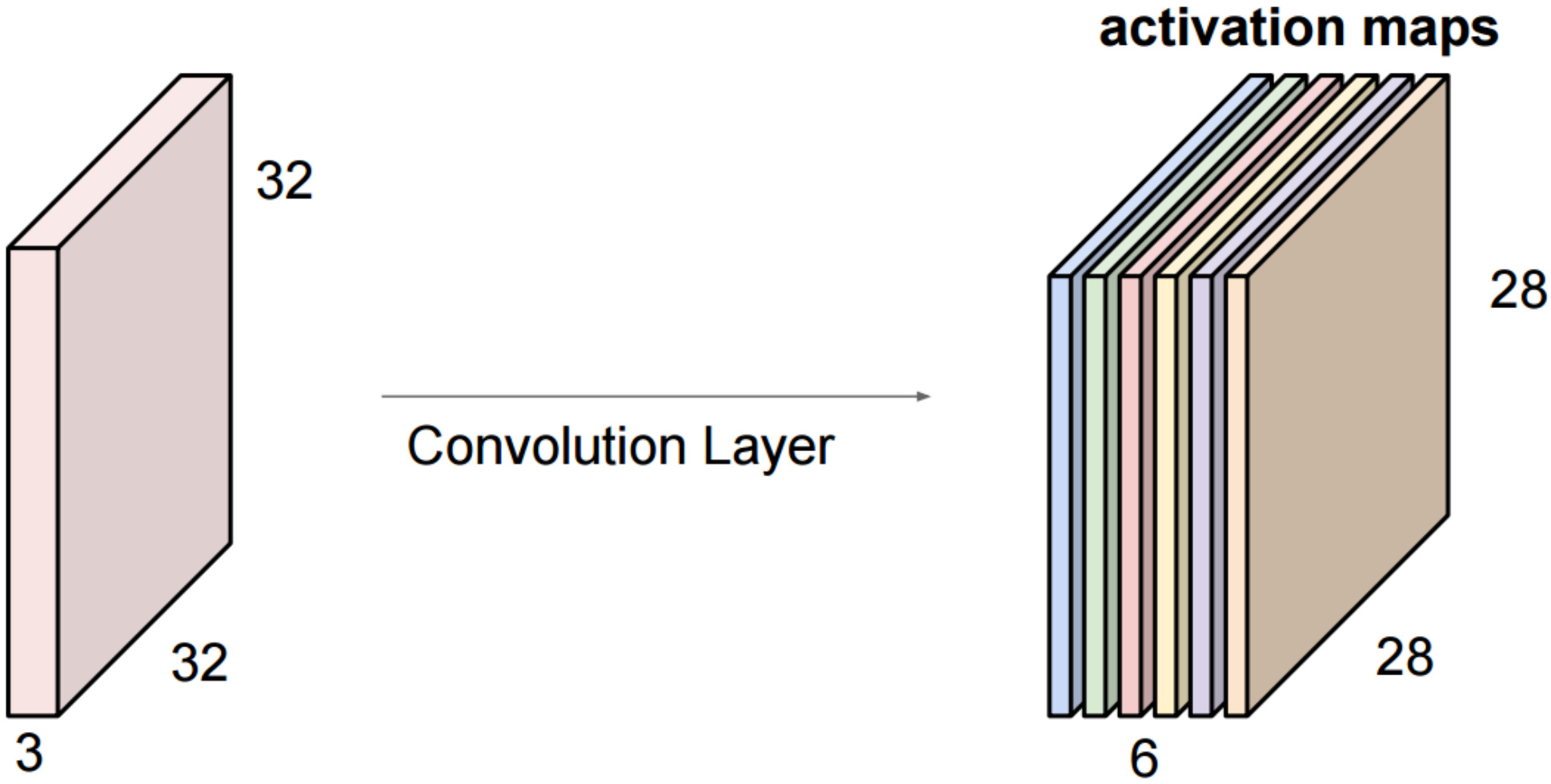


Convolution Layer

consider a second, green filter

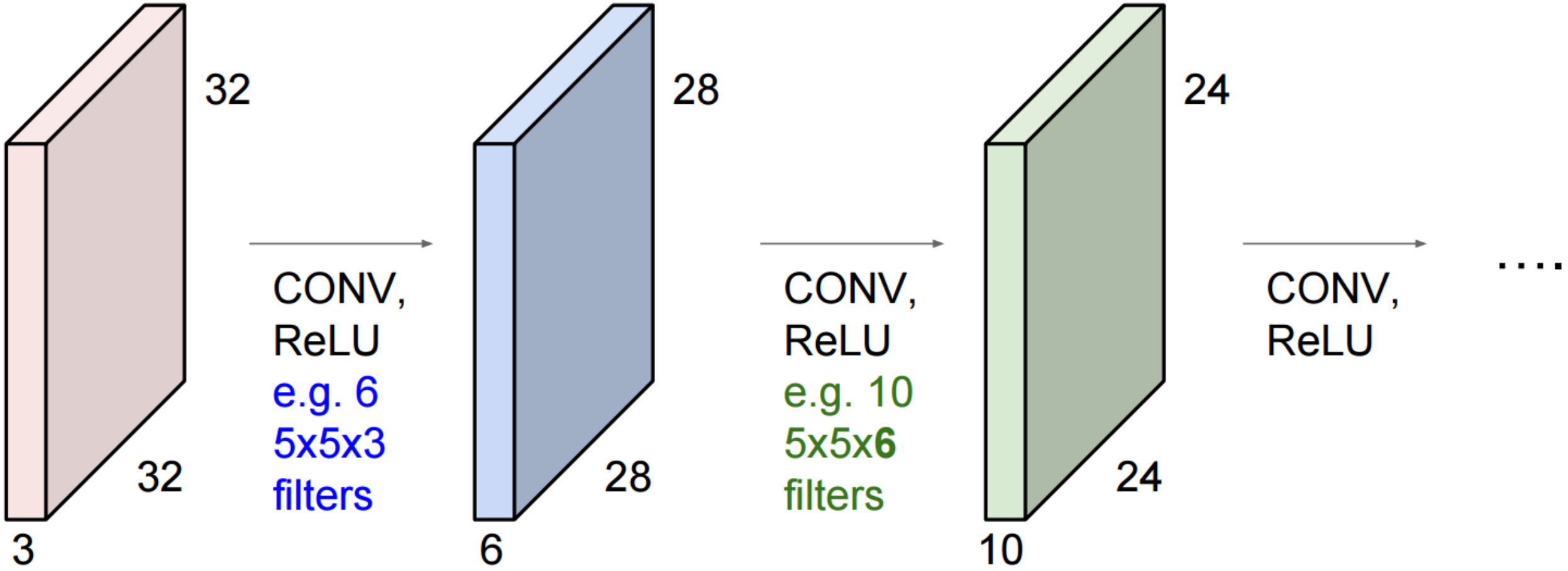


For example, if we had 6 5x5 filters, we'll get 6 separate activation maps:



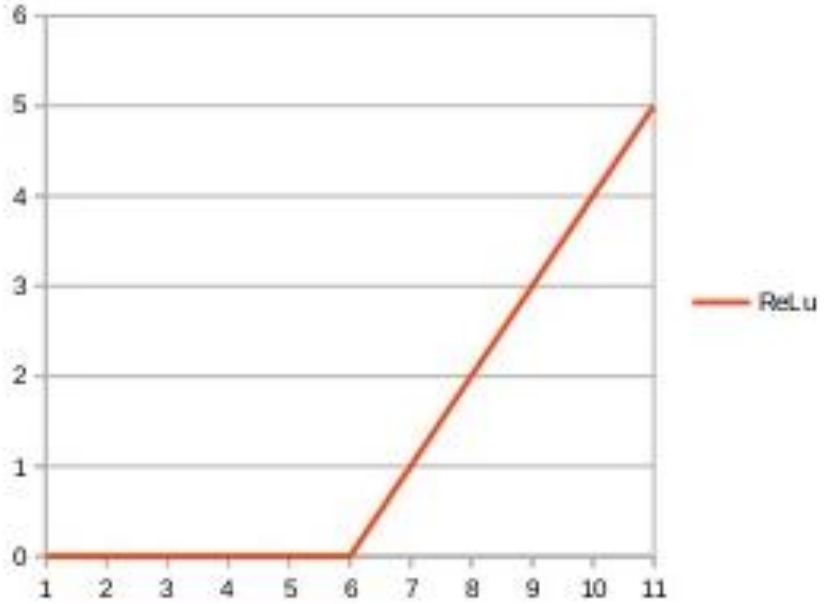
We stack these up to get a "new image" of size 28x28x6!

Preview: ConvNet is a sequence of Convolutional Layers, interspersed with activation functions



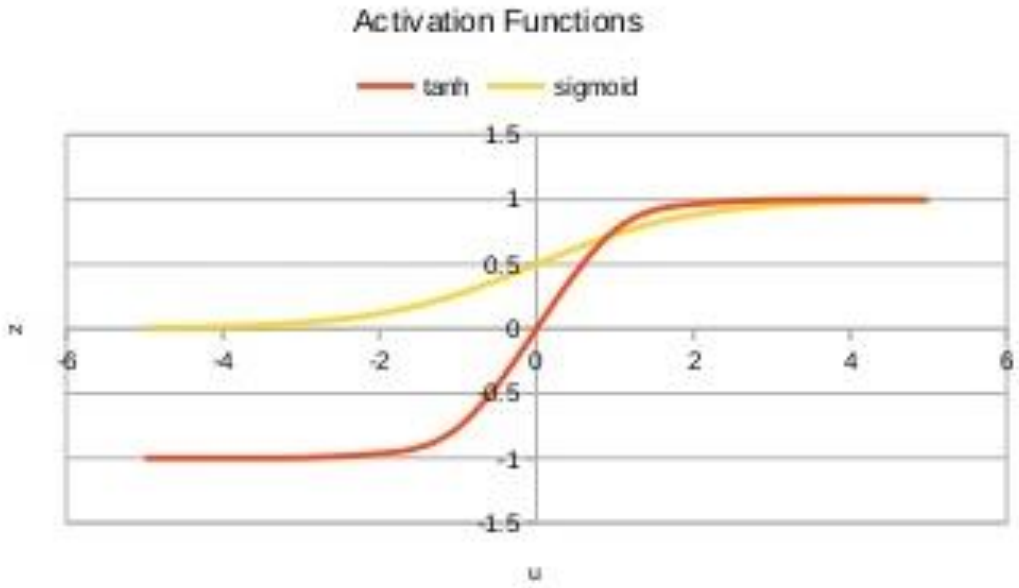
Activation Function Examples

Relu



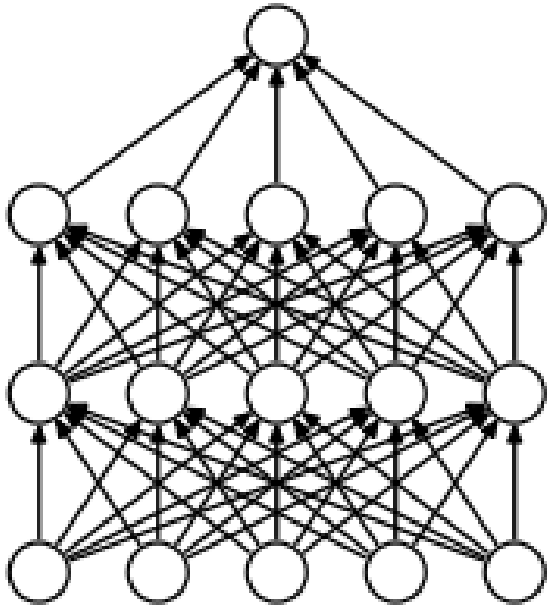
$$f(x) = max(0, x)$$

tanh sigmoid

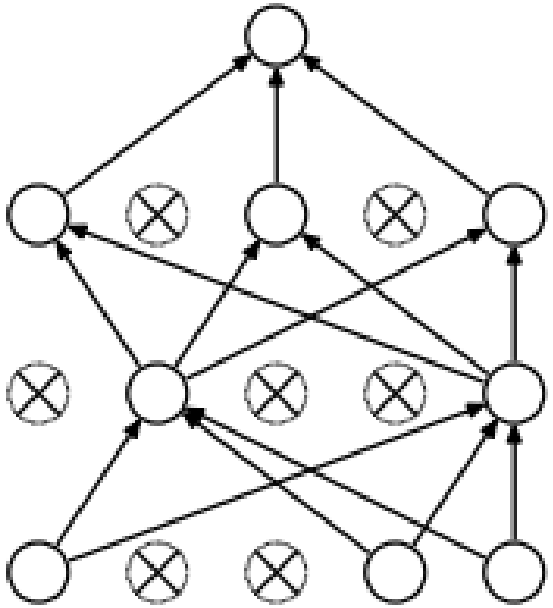


Other layers…

Fully connected



Dropout

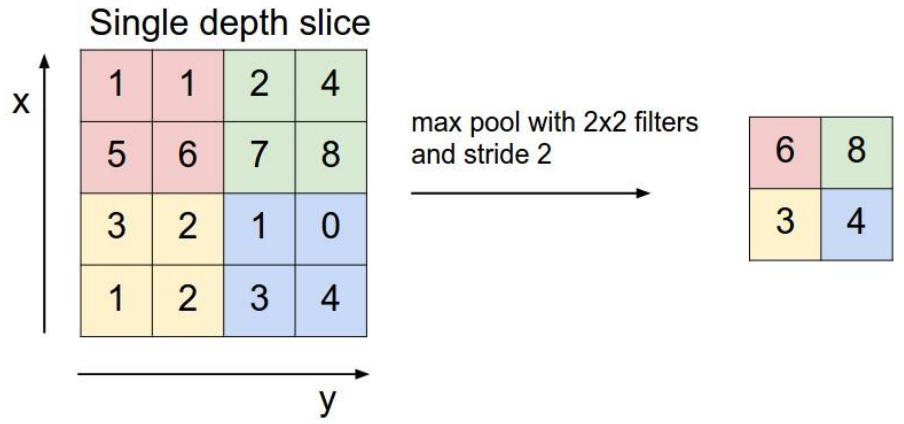


Softmax normalization

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{k} e^{z_k}}$$

$$for\ j = 1, \ldots, k$$

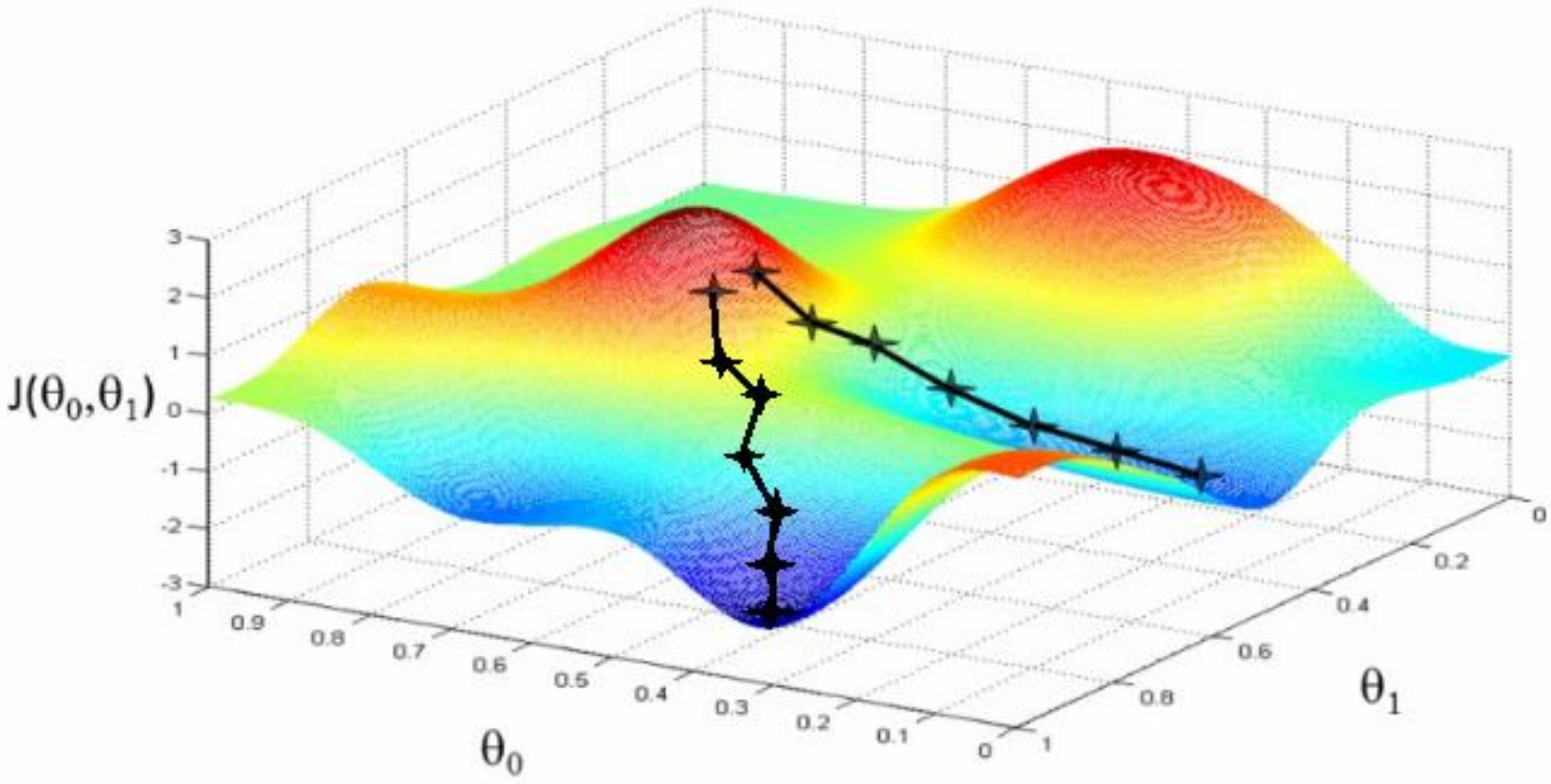
Pooling



- 1957: The Perceptron (Rosenblatt)
- The AI winter…
- 1998: First modern CNN, handwritten digits (LeCun)
- Second AI winter…
- 2012: AlexNet and Imagenet, (A. Krizhevsky)
- 2020: ~~New AI winter~~…or the AI Singularity?

…but why now?

Better training techniques...




MUCH Bigger
Datasets
https://medium.com/abraia/first-steps-with-transfer-learning-for-custom-image-classification-with-keras-b941601fcad5

MUCH better hardware...

https://tech4gamers.com/nvidia-dgx-1-the-supercomputer-for-deep-learning/

Deep Learning in the Industry

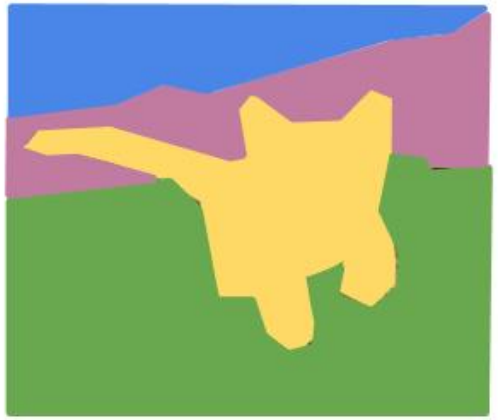
Common Types of Computer Vision Tasks

Common Types of Computer Vision Tasks

IMAGE CLASSIFICATION

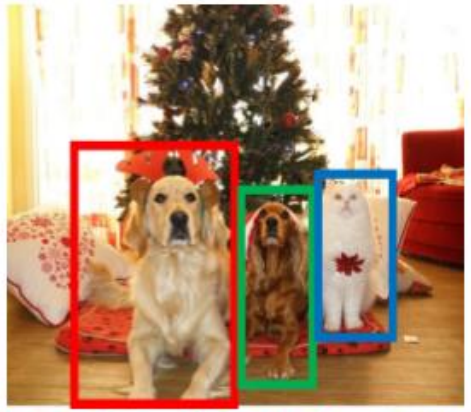
CAT

SEMANTIC SEGMENTATION



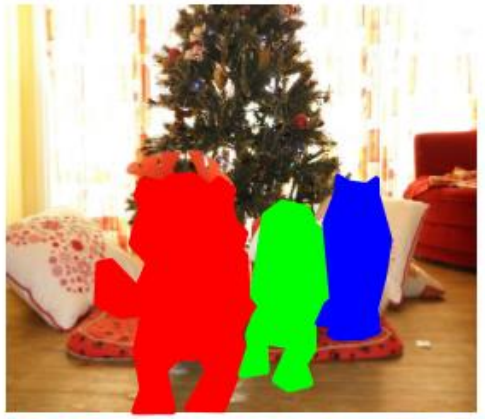
GRASS, CAT, TREE, SKY

OBJECT DETECTION



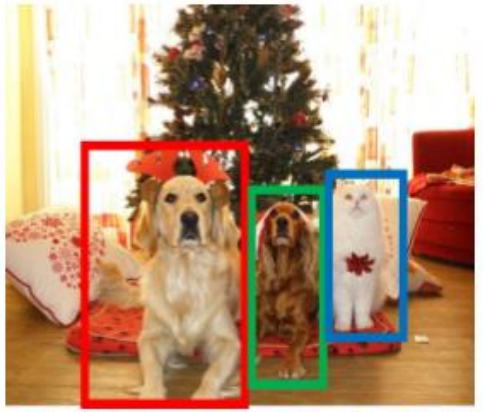
DOG, DOG, CAT

INSTANCE SEGMENTATION



DOG, DOG, CAT

OBJECT RECOGNITION



SAM, PEG, POE

Image classification

IMAGE CLASSIFICATION

CAT

SEMANTIC SEGMENTATION

GRASS, CAT, TREE, SKY

OBJECT DETECTION

DOG, DOG, CAT

INSTANCE SEGMENTATION

DOG, DOG, CAT

OBJECT RECOGNITION

SAM, PEG, POE

Image classification

Common problems – e.g. traffic incident detection

Shadows

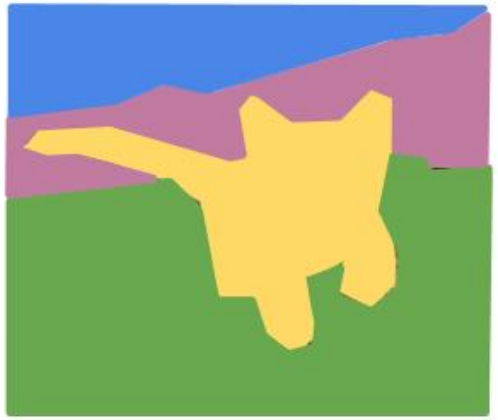
Irrelevant objects

SEMANTIC SEGMENTATION

IMAGE CLASSIFICATION

CAT

SEMANTIC SEGMENTATION



GRASS, CAT, TREE, SKY

OBJECT DETECTION

DOG, DOG, CAT

INSTANCE SEGMENTATION

DOG, DOG, CAT

OBJECT RECOGNITION

SAM, PEG, POE

http://sznur.com/semantic-segmentation-using-fully-convolutional-network

Semantic Segmentation Demo

https://www.youtube.com/watch?v=ATlcEDSPWXY

OBJECT DETECTION

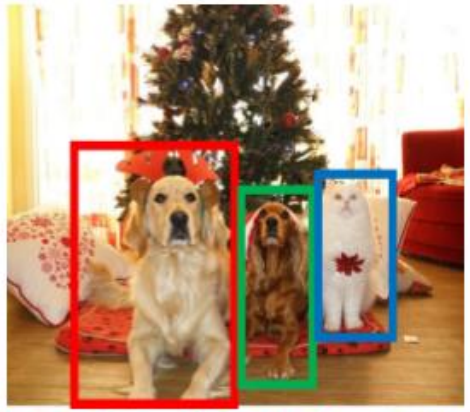
IMAGE CLASSIFICATION

CAT

SEMANTIC SEGMENTATION

GRASS, CAT, TREE, SKY

OBJECT DETECTION



DOG,DOG,CAT

INSTANCE SEGMENTATION

DOG,DOG,CAT

OBJECT RECOGNITION

SAM,PEG,POE

Object Detection Demo

https://www.youtube.com/watch?v=F-lWyJ5Trk4

INSTANCE SEGMENTATION

IMAGE CLASSIFICATION

CAT

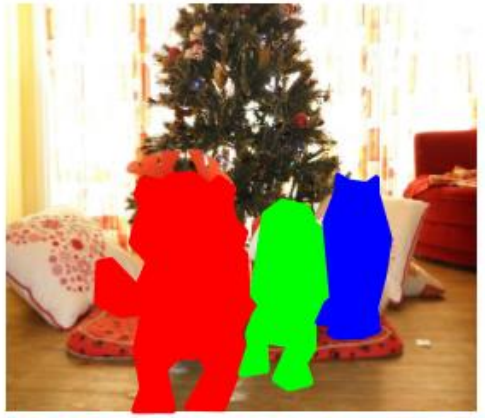
SEMANTIC SEGMENTATION

GRASS, CAT, TREE, SKY

OBJECT DETECTION

DOG,DOG,CAT

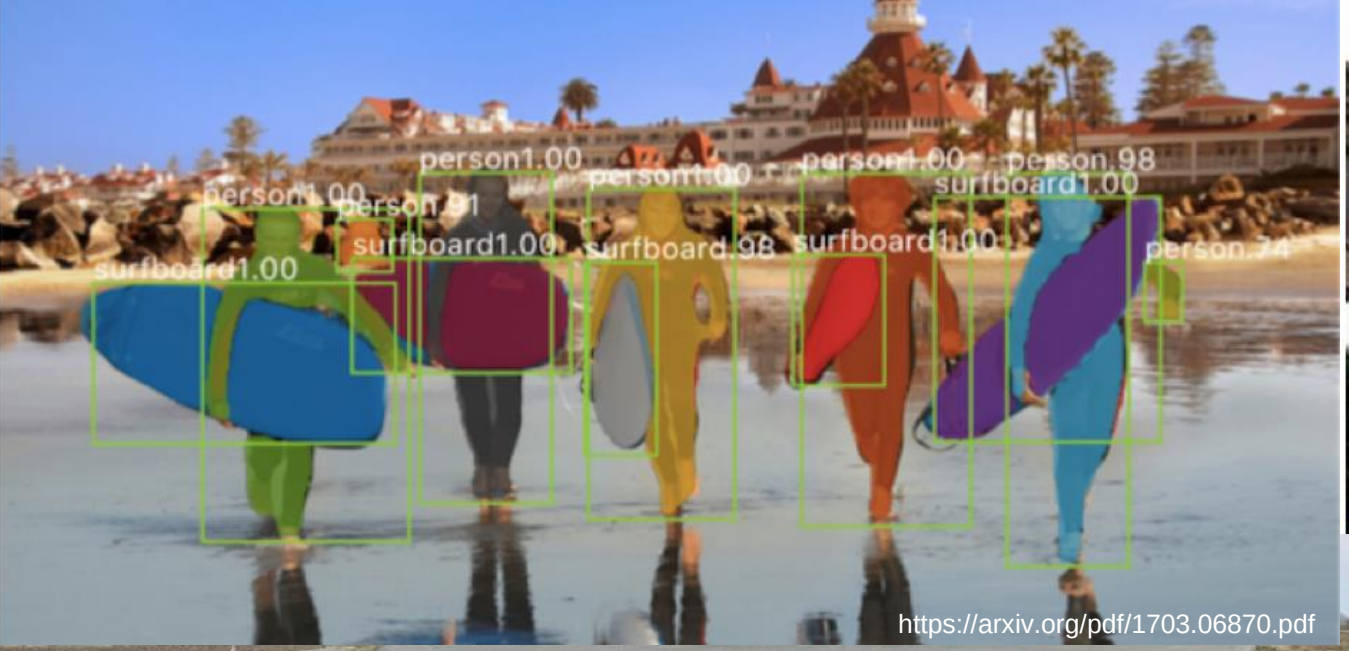
INSTANCE SEGMENTATION



DOG,DOG,CAT

OBJECT RECOGNITION

SAM,PEG,POE



https://arxiv.org/pdf/1703.06870.pdf

https://arxiv.org/pdf/1703.06870.pdf

https://www.freecodecamp.org/news/mask-r-cnn-explained-7f82bec890e3/

https://devmesh.intel.com/projects/detection-and-classification-for-leopards

Instance Segmentation Demo

https://www.youtube.com/watch?v=0pMfmo8qfpQ

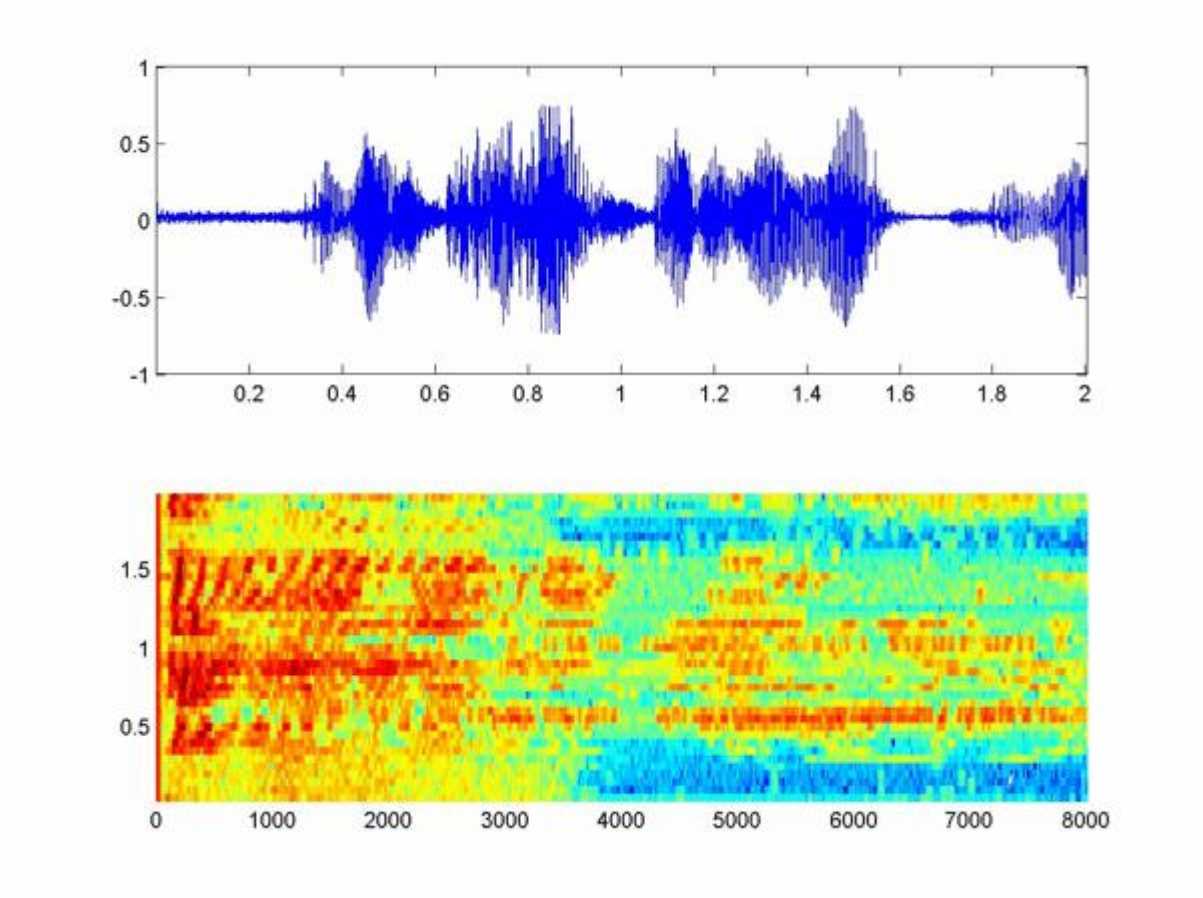
Pose estimation

Pose Estimation Demo

https://www.youtube.com/watch?v=KYNDzIcQMWA

Convolutions not only for video

- Audio can be transformed into ”image-like” format using FFT
- Time-domain methods are of course important, but requires other types of architectures
- Out of scope for this intro.



https://www.mathworks.com/matlabcentral/fileexchange/19933-generate-animated-gif-files-for-plotting-audio-data

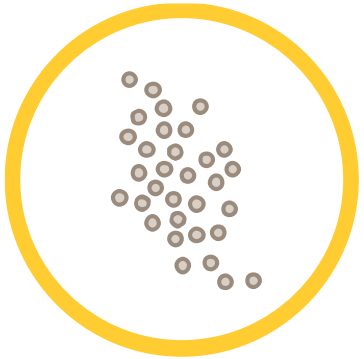
Deployment Platforms

Typical Deployment Alternatives

Cloud

Server

Embedded



”Fog”

Embedded Deployment of Deep Learning

- Embedded compute resources are scarce
- Development of dedicated compute resources for Deep Learning on-edge is changing the relevance of the above statement
  - Fast
  - Support for lower bitwidth processing -> even faster
  - Low flexibility,
  - Low power consumption/op
  - Today mainly inference (poor support for training)
- CNNs can also be made smaller after/during training (e.g. layer pruning)

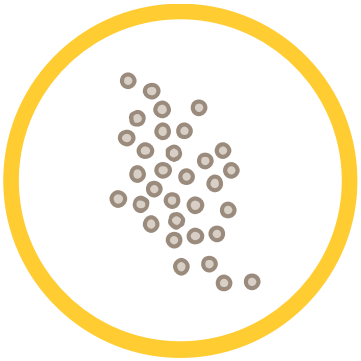
Server Deployment of Deep Learning

- From the DL processing perspective, simply a PC with GPU processing capabilities
- Greater flexibility than edge (GPUs are very programmable)
- New trend is lower bitwidth processing also for server deployment
  - Faster inference (or can run larger CNNs)
  - Can run higher number of parallel video streams
- Many emerging ”standards”, e.g.
  - TensorRT (Nividia)
  - Tensorflow Serving
- Common pattern is to stream on-edge processed data from many nodes to the server side for further processing/analysis.

Cloud Deployment of Deep Learning

- Similar to server but at remote location
  - Security concerns possibly an issue
- Often hosted by 3:rd party
- On-demand scalability

”Fog” Deployment of Deep Learning



- New term coined by Cisco in 2014
- Basically, fog is closer to the end-user
- The nodes are physically much closer to devices, compared to centralized data centers
- Fog can also include *cloudlets* —
  - Small-scale data centers located at the edge of the network.
  - Purpose is to support resource-intensive IoT apps that require low latency.
  - Compare with “on-prem cloud” where focus is on the security aspect (not on latency).


Megapixel
Quality
Ease of use
Network video
Reliability
open
Innovation
Easy installation
Protect
support
integration
Secure
Thank you
Image usability
Service all the way
Thermal
Convergence
HDTV
Access control
World Leader
Quality
Intelligent
Partners
Competence
Focus
Cameras
Safe
Outdoor
Video encoder
Global