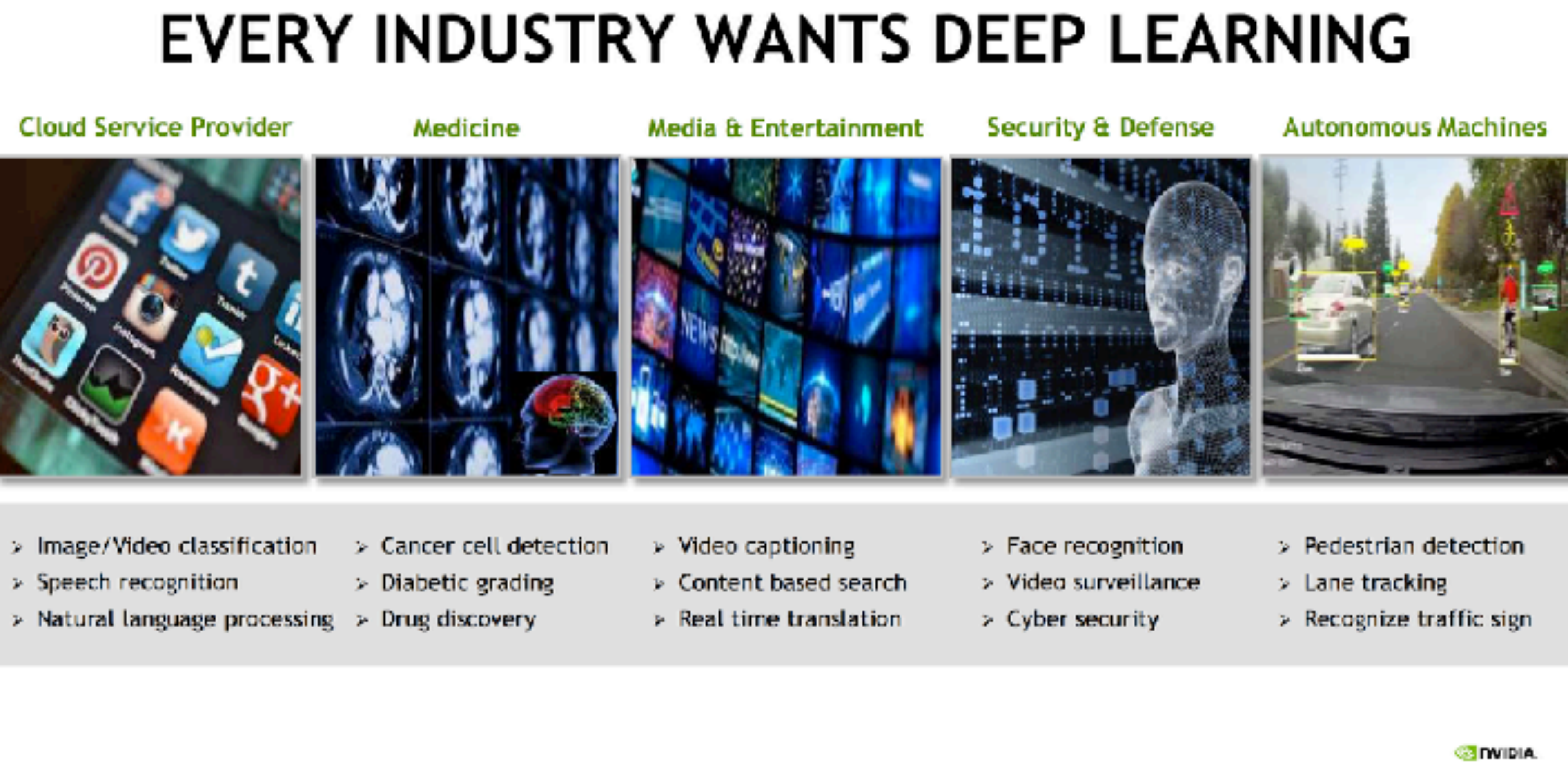
Data Mining

Lecture 10 - Deep Learning

The Deep Learning Revolution



Video: https://www.youtube.com/watch?v=Dy0hJWltsyE

Why now?

KEY DRIVERS FOR DEEP LEARNING

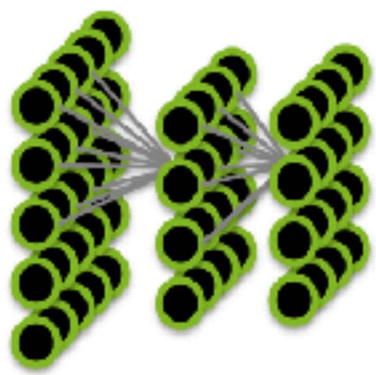
Big Data

facebook — 350 millions images uploaded per day

Walmart — 2.5 Petabytes of customer data hourly

YouTube — 300 hours of video uploaded every minute

Better Algorithms

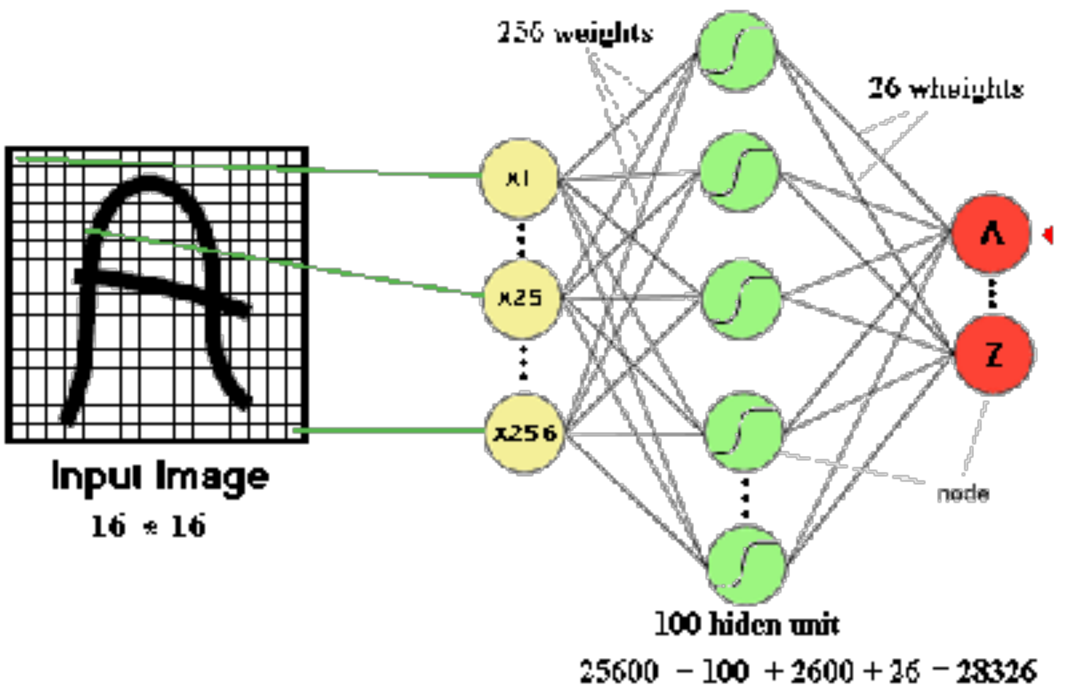


GPU Acceleration

"The Three Breakthroughs that have Finally Unleashed A.I. on the World" WIRED

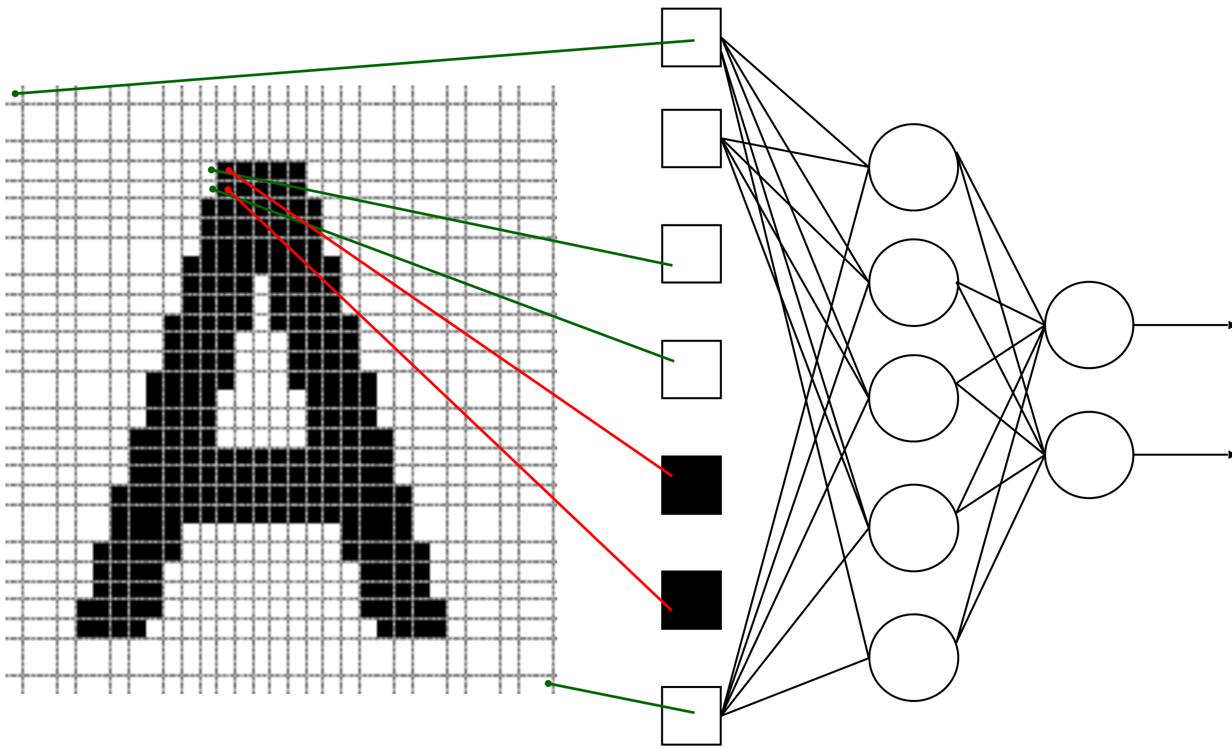
Drawbacks of previous neural networks

The number of trainable parameters becomes extremely large



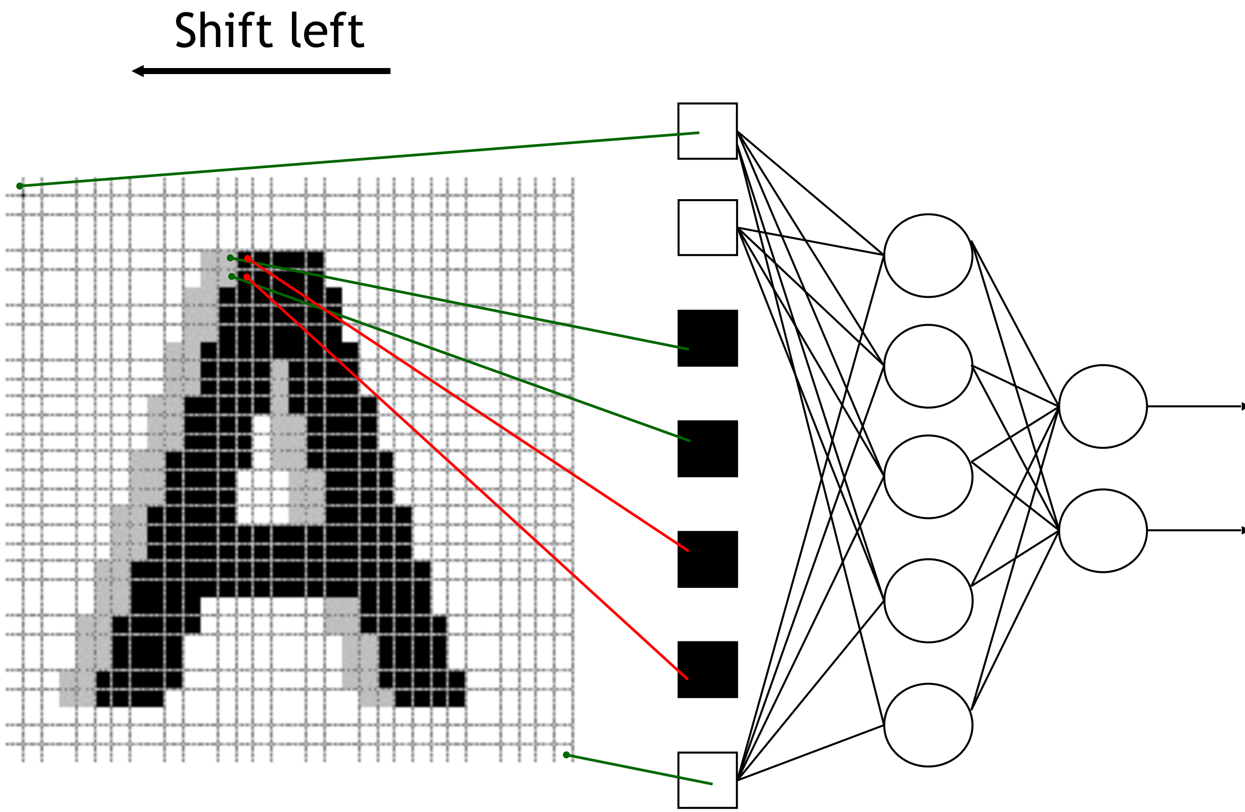
Drawbacks of previous neural networks

Little or no invariance to shifting, scaling, and other forms of distortion

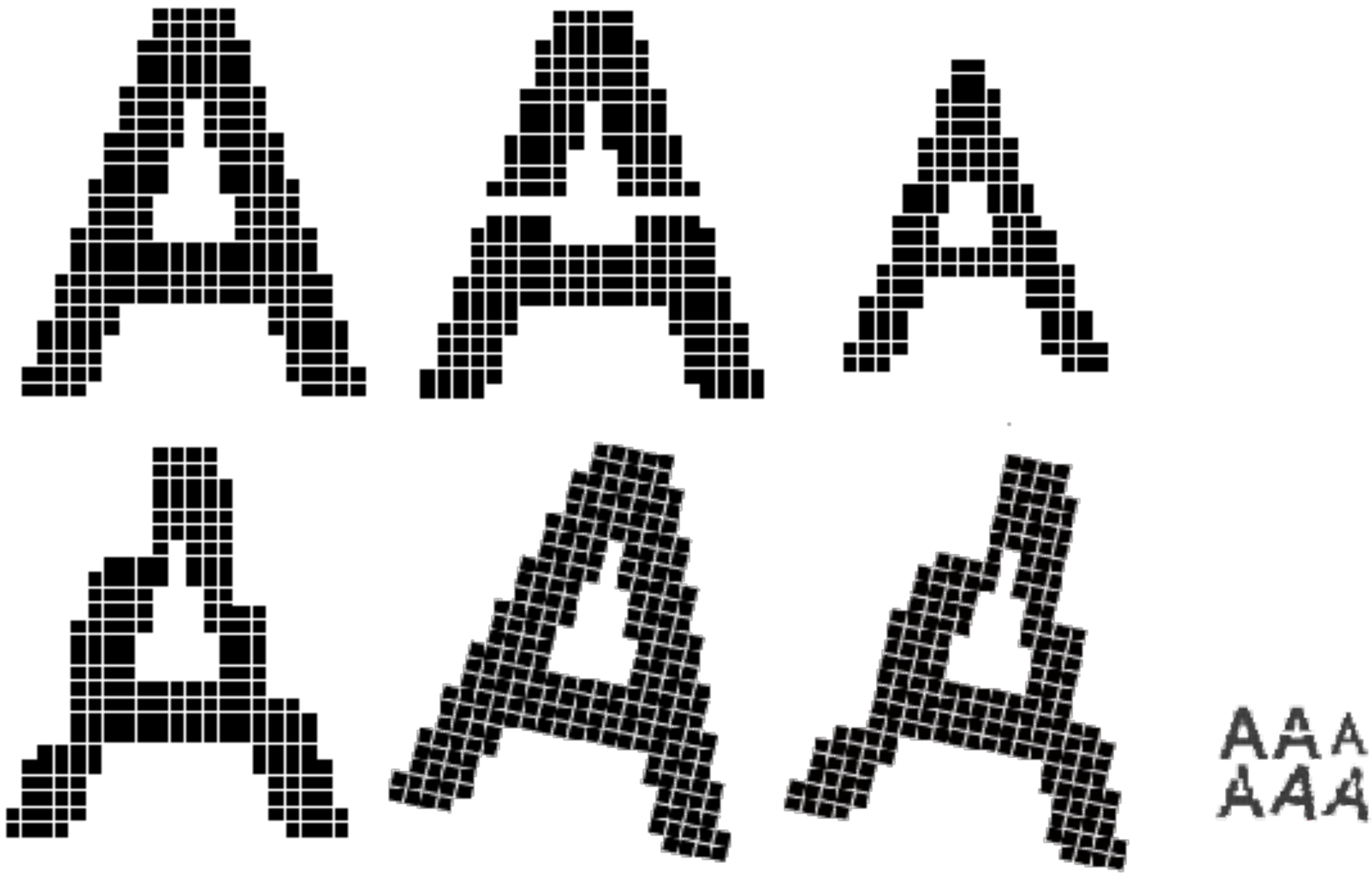


Drawbacks of previous neural networks

Little or no invariance to shifting, scaling, and other forms of distortion

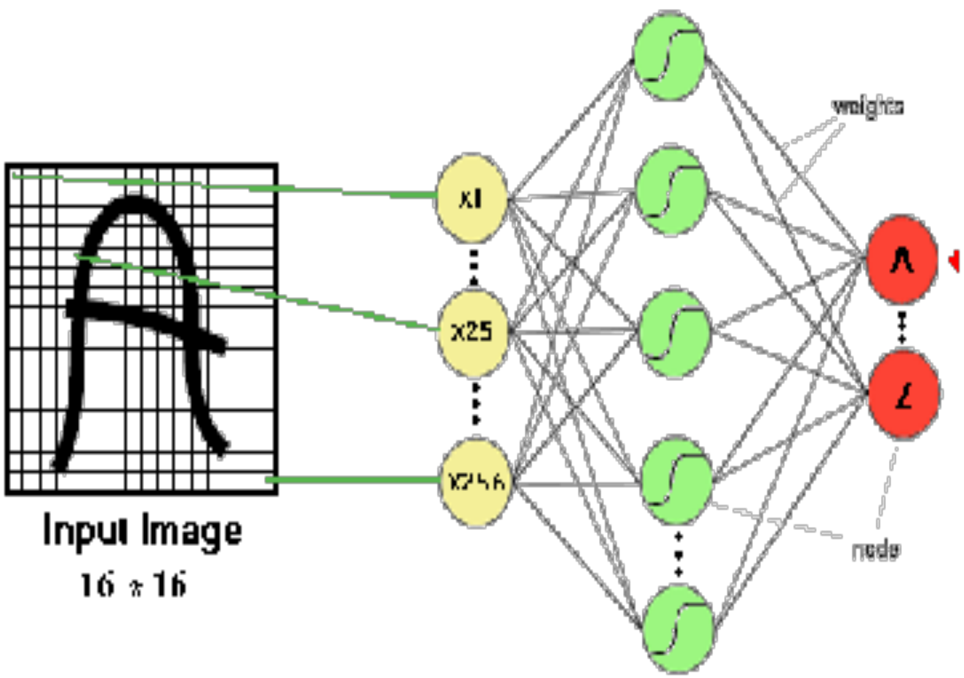


Drawbacks of previous neural networks



Drawbacks of previous neural networks

The topology of the input data is completely ignored



Convolutional neural networks (CNNs)

About CNN's

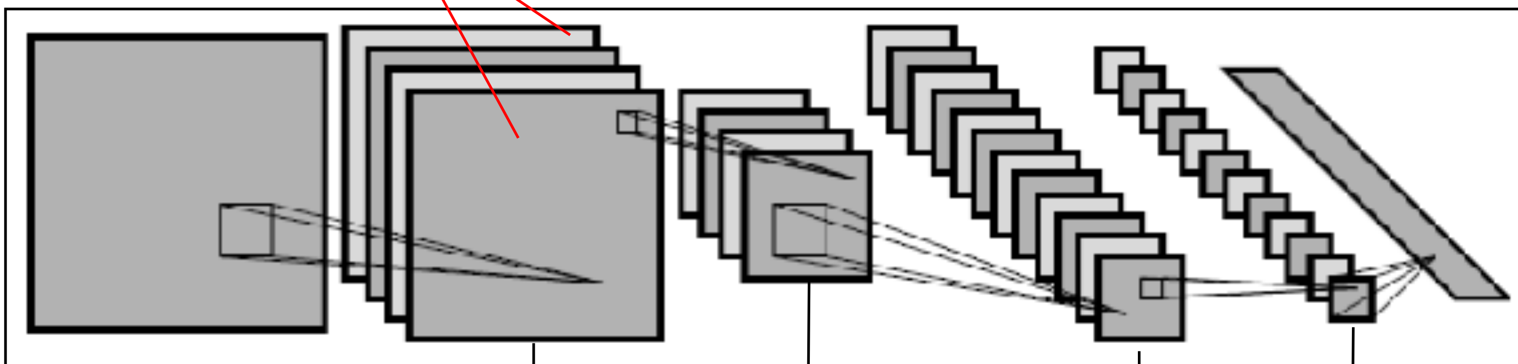
- CNN's were neurobiologically motivated by the findings of locally sensitive and orientation-selective nerve cells in the visual cortex.

- They designed a network structure that implicitly extracts relevant features.

- Convolutional Neural Networks are a special kind of multi-layer neural networks.

CNN’s Topology

Feature maps



Feature extraction layer
Convolution layer

C

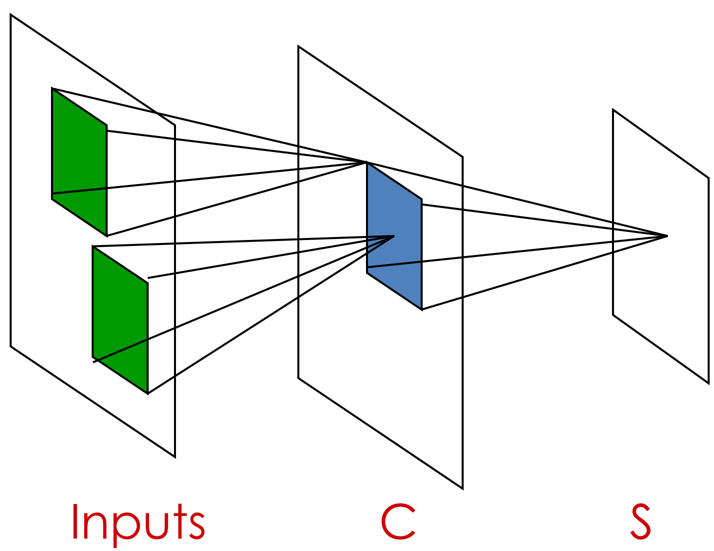
Shift and distortion invariance or
Subsampling layer

S

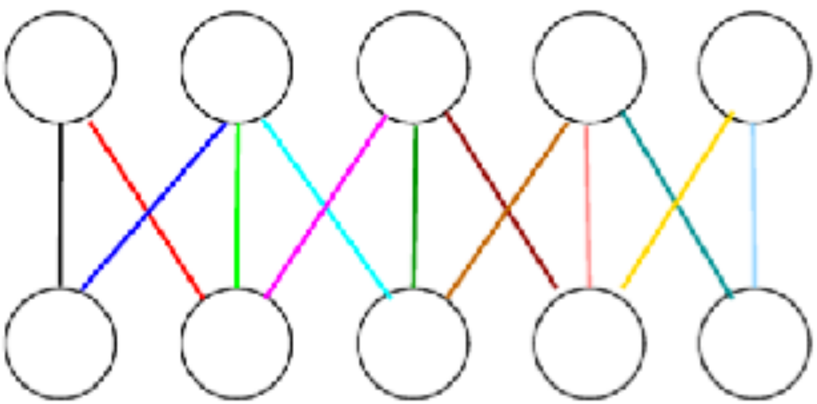
Feature extraction

- Shared weights: all neurons in a feature share the same weights
- In this way all neurons detect the same feature at different positions in the input image.

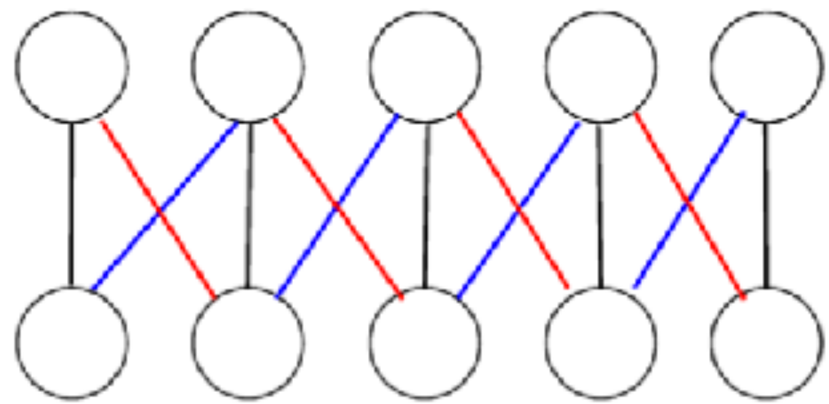
- Reduce the number of free parameters.



Local connectivity

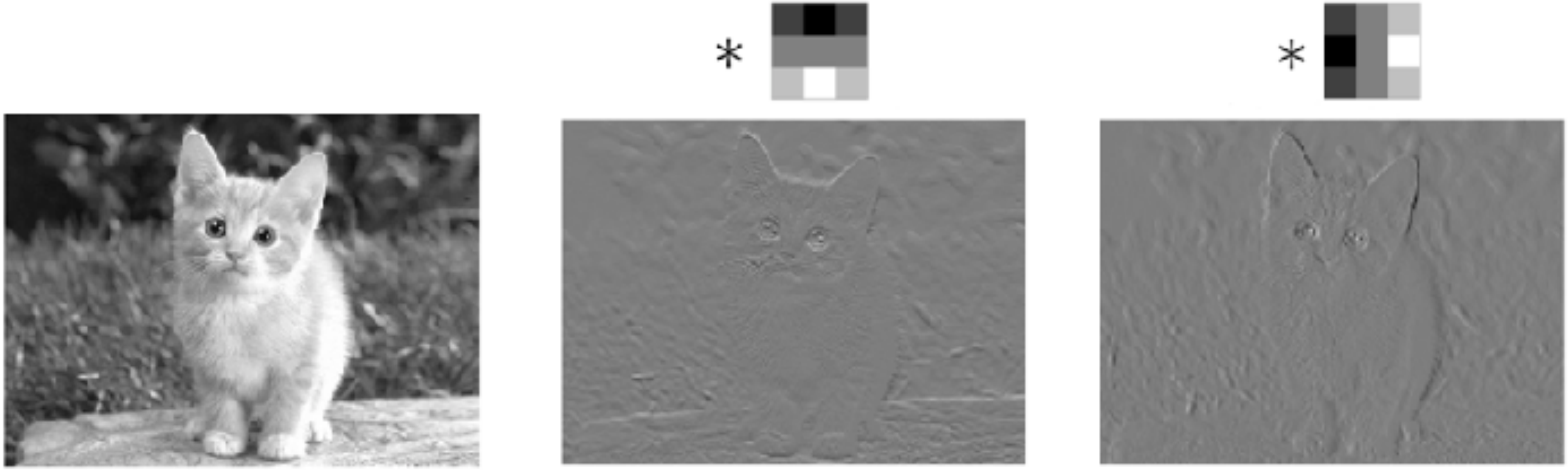


13 parameters



3 parameters

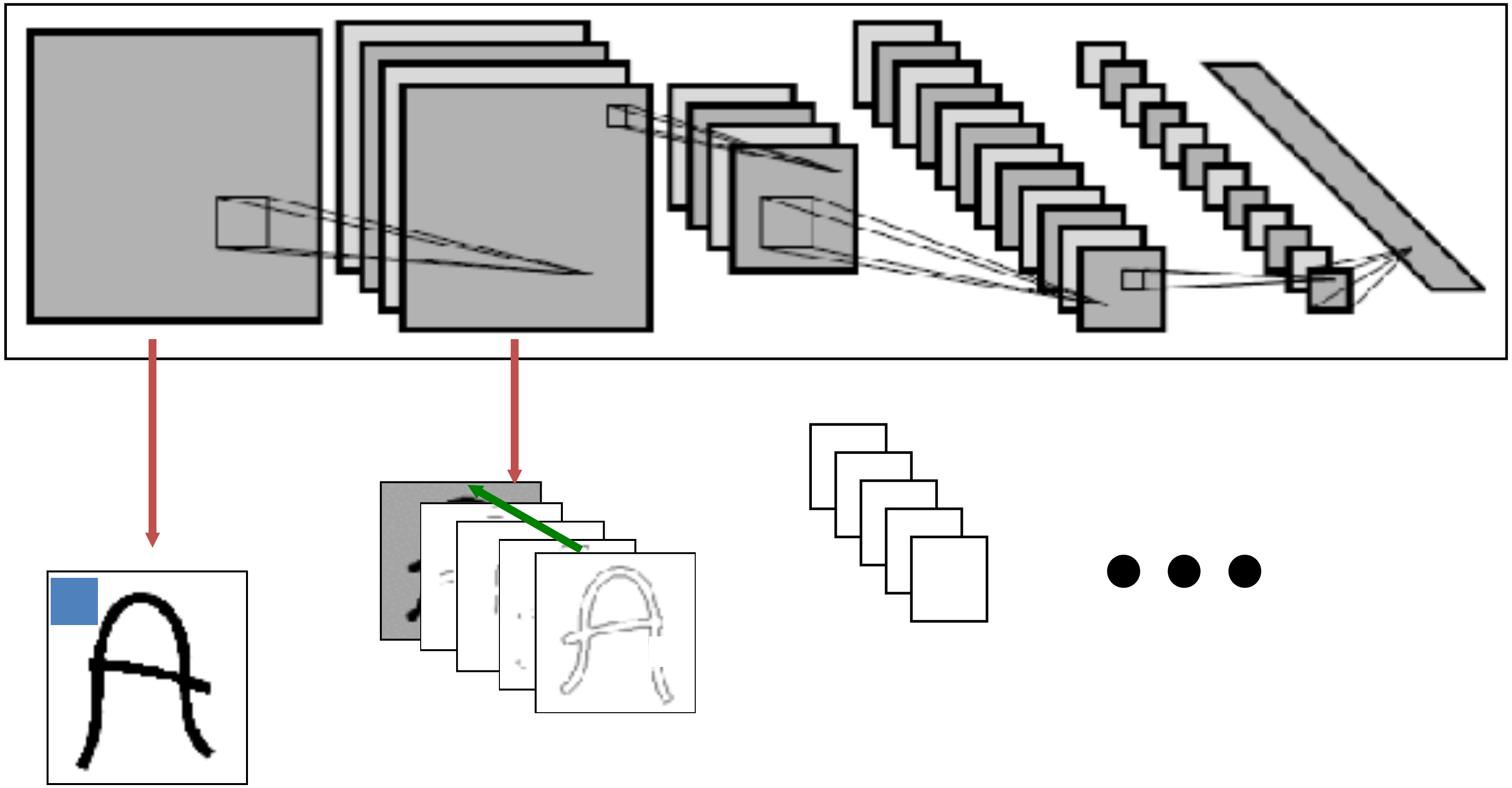
Convolution Details



Video at: http://cs231n.github.io/convolutional-networks/

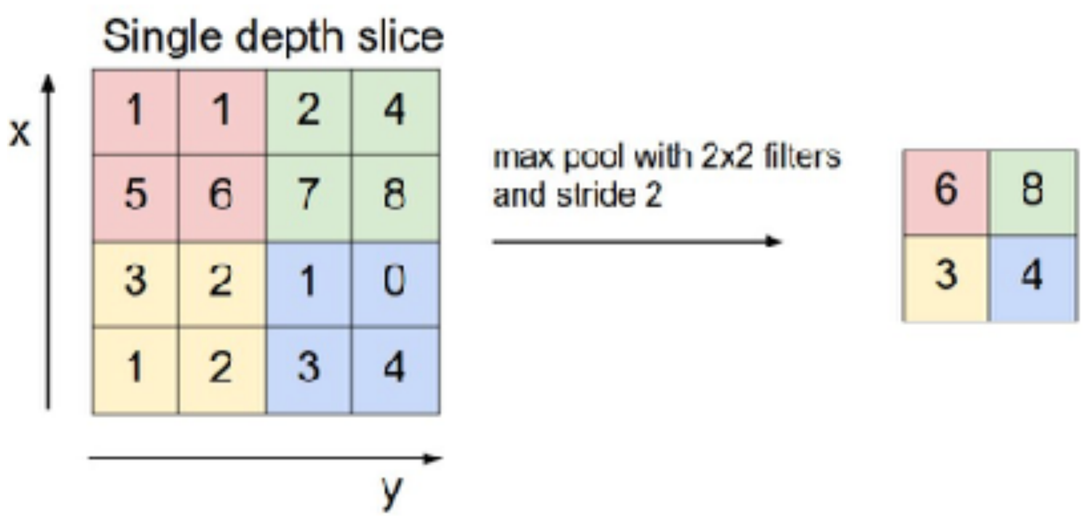
Feature extraction

If a neuron in the feature map fires, this corresponds to a match with the template.

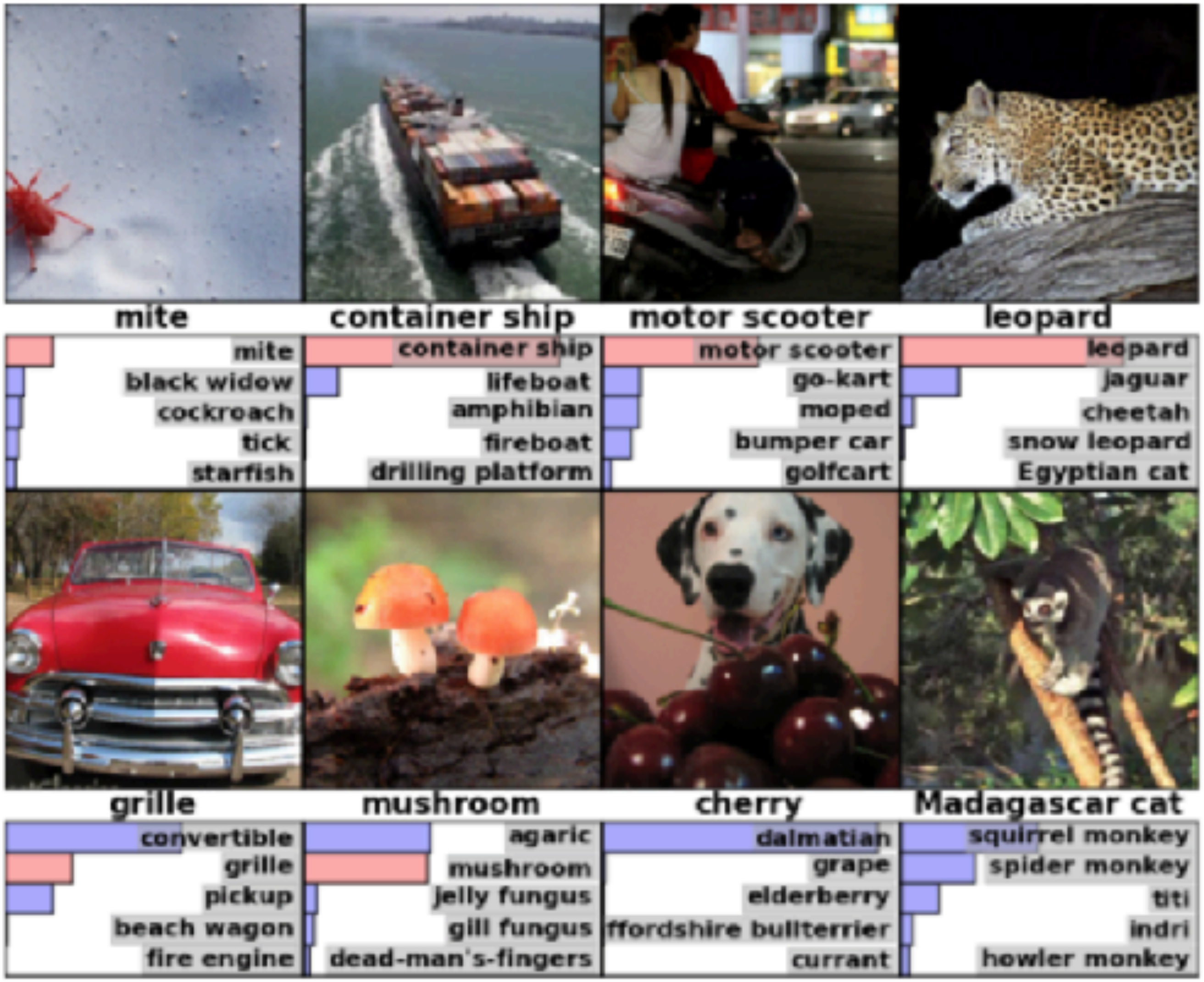


Subsampling layer

- The subsampling layers reduce the spatial resolution of each feature map (also called pooling)
- By reducing the spatial resolution of the feature map, a certain degree of shift and distortion invariance is achieved.



ImageNet - image classification



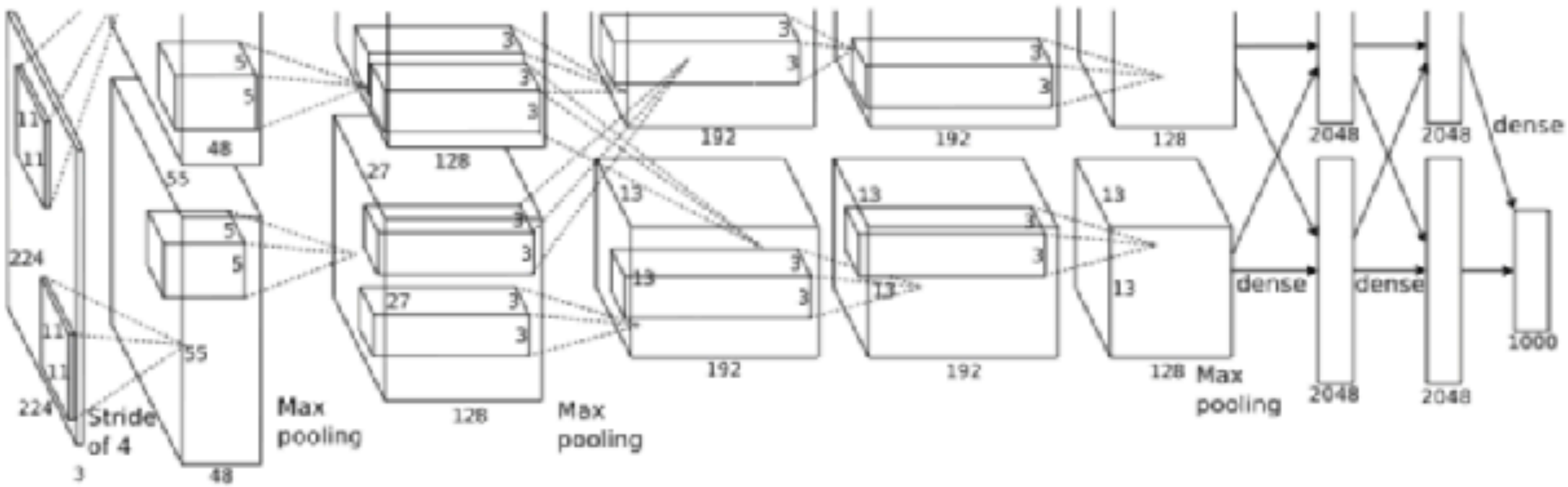
- 1.000 different classes
- 1.000.000 training images

ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky
University of Toronto
kriz@cs.utoronto.ca

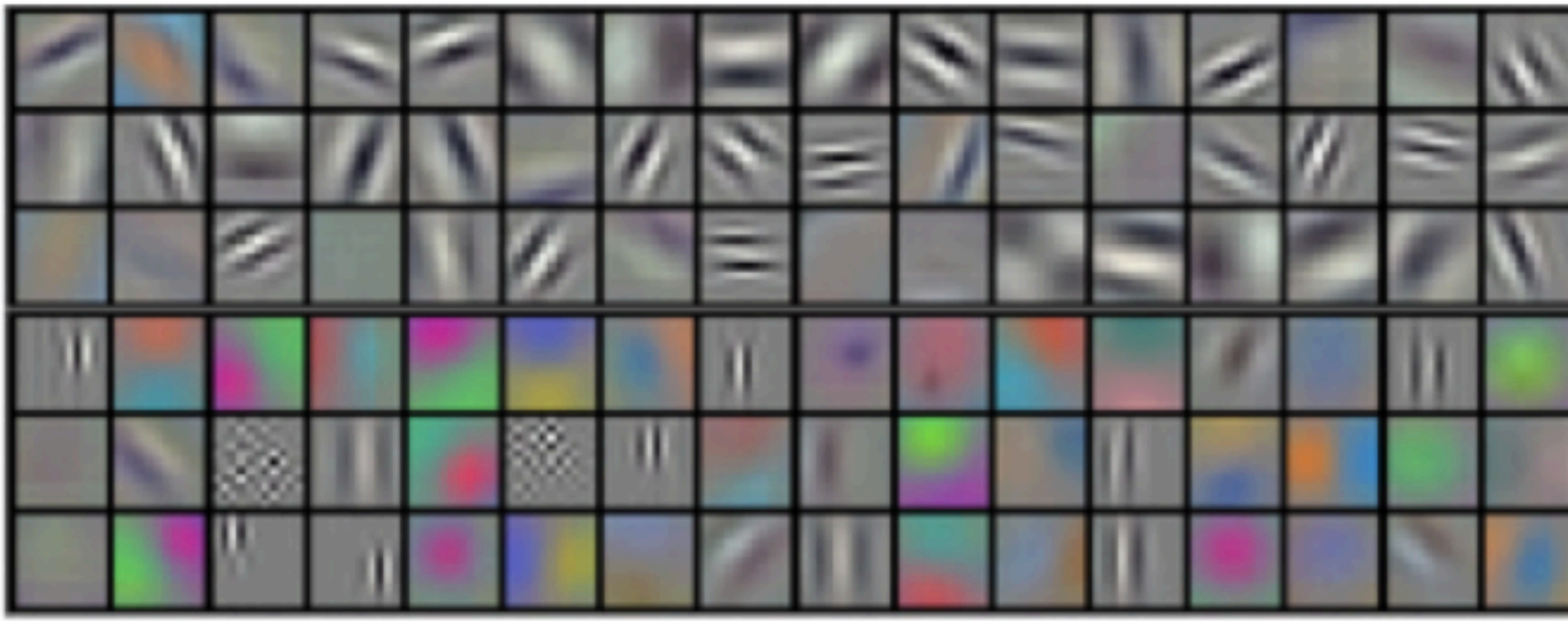
Ilya Sutskever
University of Toronto
ilya@cs.utoronto.ca

Geoffrey E. Hinton
University of Toronto
hinton@cs.utoronto.ca



www.cs.toronto.edu/˜fritz/absps/imagenet.pdf

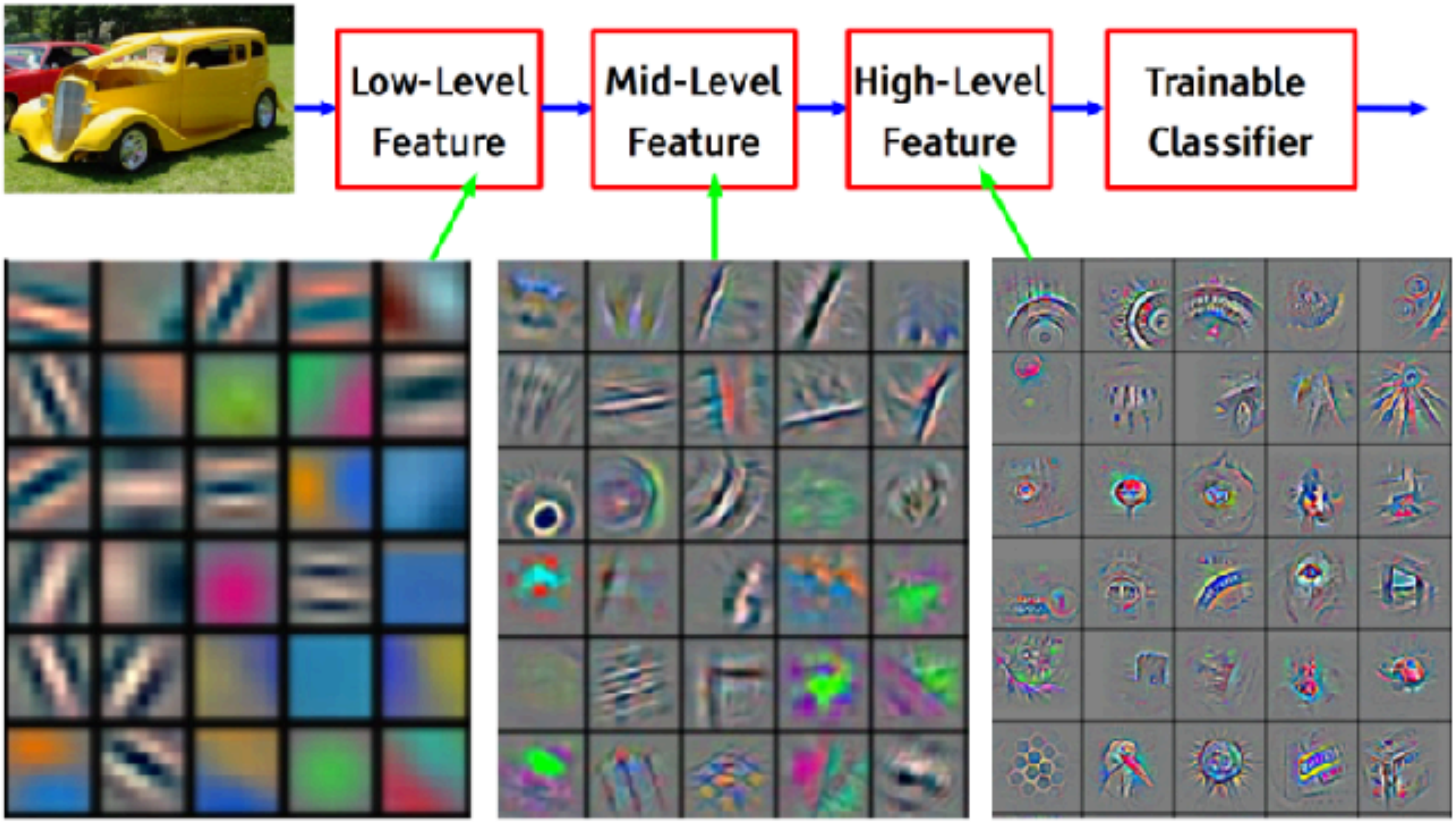
Learned features of first layer



Deepvis: https://www.youtube.com/watch?v=AgkfIQ4IGaM&feature=youtu.be

Deep Learning = Learning Hierarchical Representations

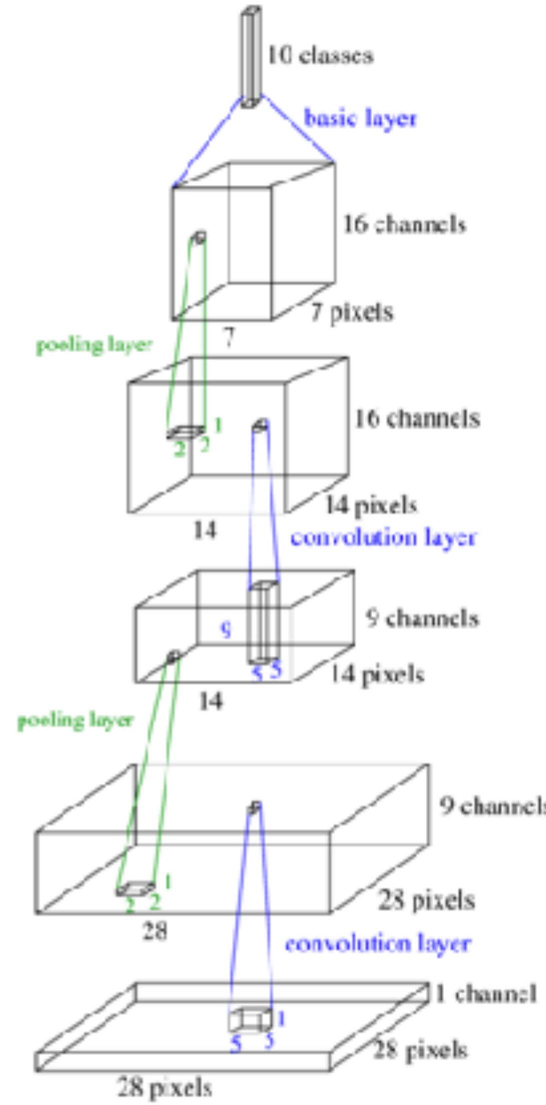
- It's deep if it has more than one stage of non-linear feature transformation



Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

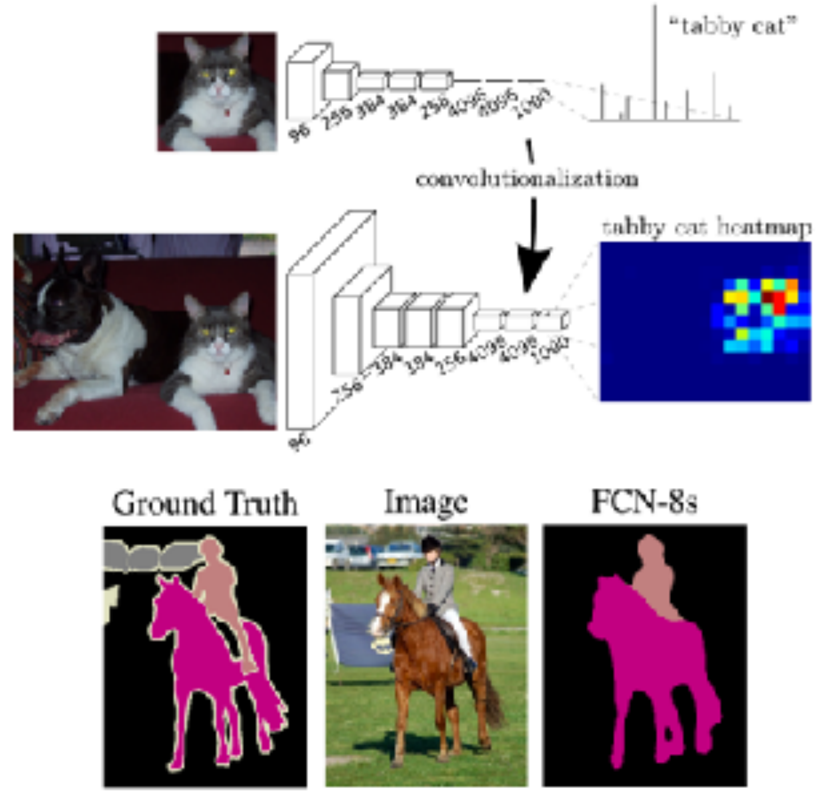
Network Example

- Often resolution is decreased and channels increased going up
- Important: Each unit looks at all channels of the previous layer
- Recent trends go towards small filters (3 × 3), sometimes even without pooling



Can also be used for Semantic Segmentation

- Basic idea: Slide a CNN to classify each location of the image

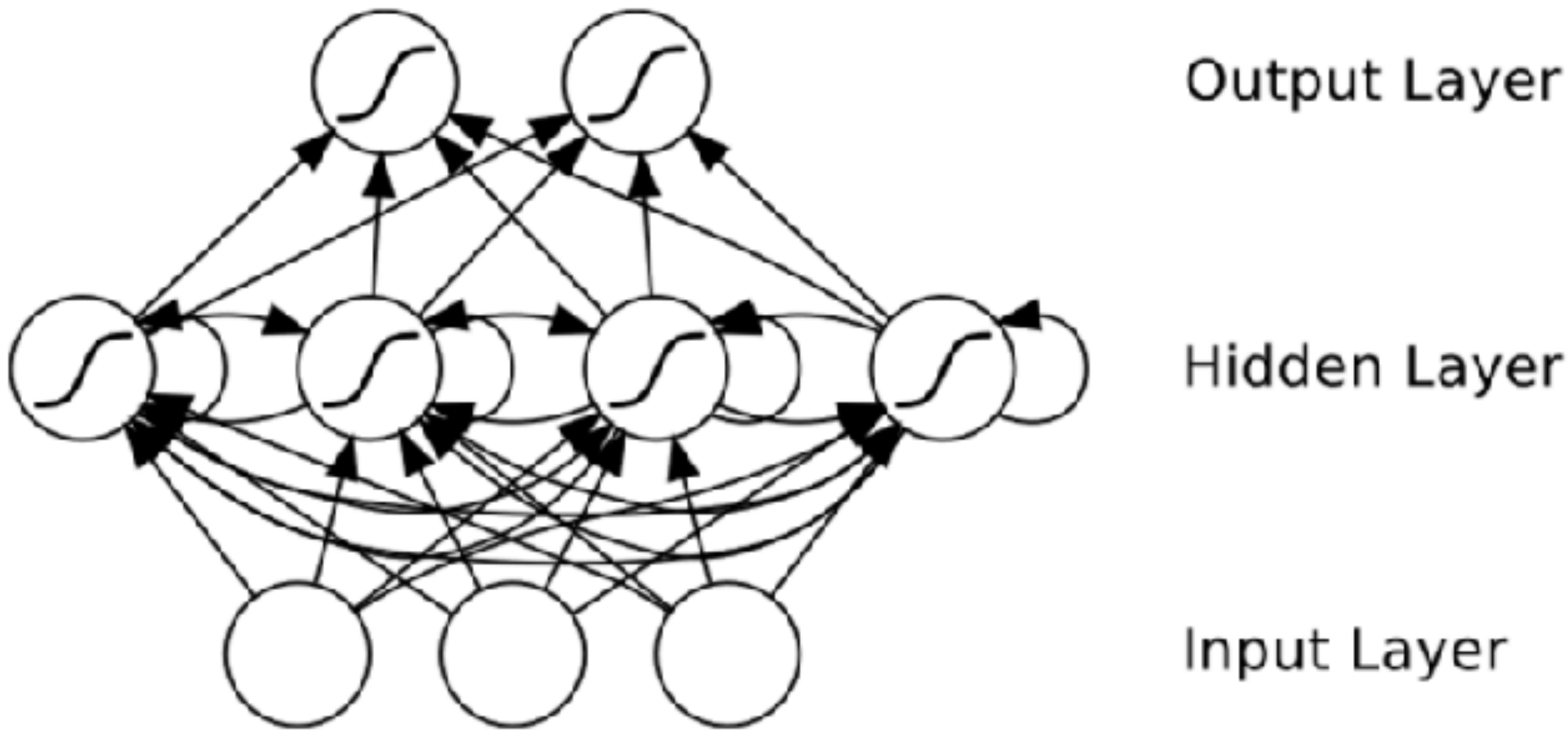


Long et al., 2015

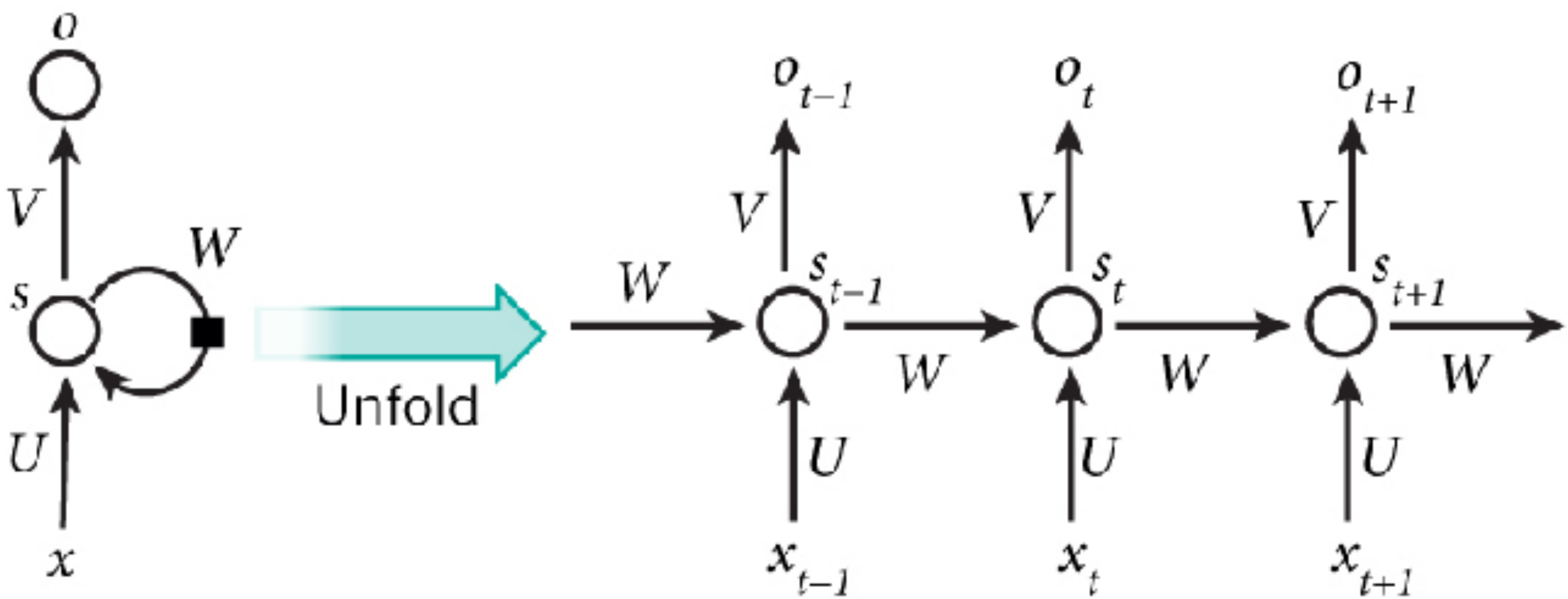
MNIST Example Code (Tensorflow+Keras)

```
model = Sequential()
model.add(Conv2D(32, kernel_size=(3, 3),
                 activation='relu',
                 input_shape=input_shape))
model.add(Conv2D(64, (3, 3), activation='relu'))
model.add(MaxPooling2D(pool_size=(2, 2)))
model.add(Dropout(0.25))
model.add(Flatten())
model.add(Dense(128, activation='relu'))
model.add(Dropout(0.5))
model.add(Dense(num_classes, activation='softmax'))

model.compile(loss=keras.losses.categorical_crossentropy,
              optimizer=keras.optimizers.Adadelta(),
              metrics=['accuracy'])

model.fit(x_train, y_train,
          batch_size=batch_size,
          epochs=epochs,
          verbose=1,
          validation_data=(x_test, y_test))
score = model.evaluate(x_test, y_test, verbose=0)
print('Test loss:', score[0])
print('Test accuracy:', score[1])
```

Recurrent Networks can learn sequential tasks



Recurrent Network Unrolled

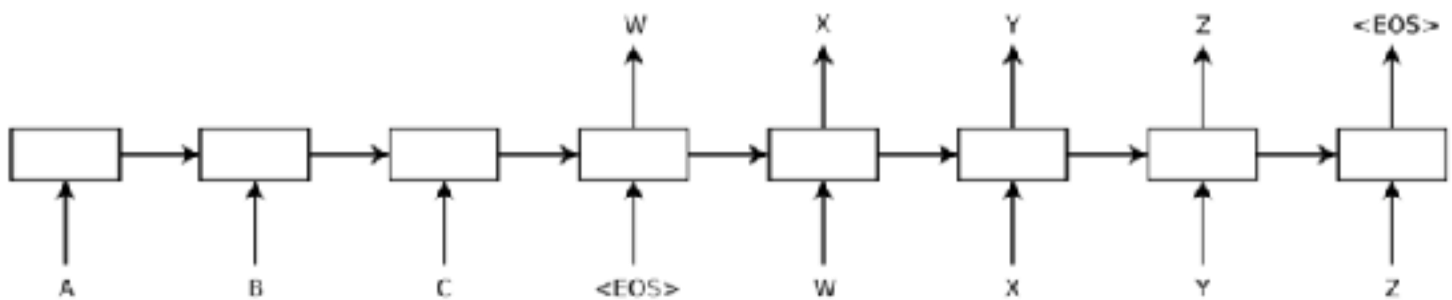


Sequence to Sequence Learning with Neural Networks

Ilya Sutskever
Google
ilyasu@google.com

Oriol Vinyals
Google
vinyals@google.com

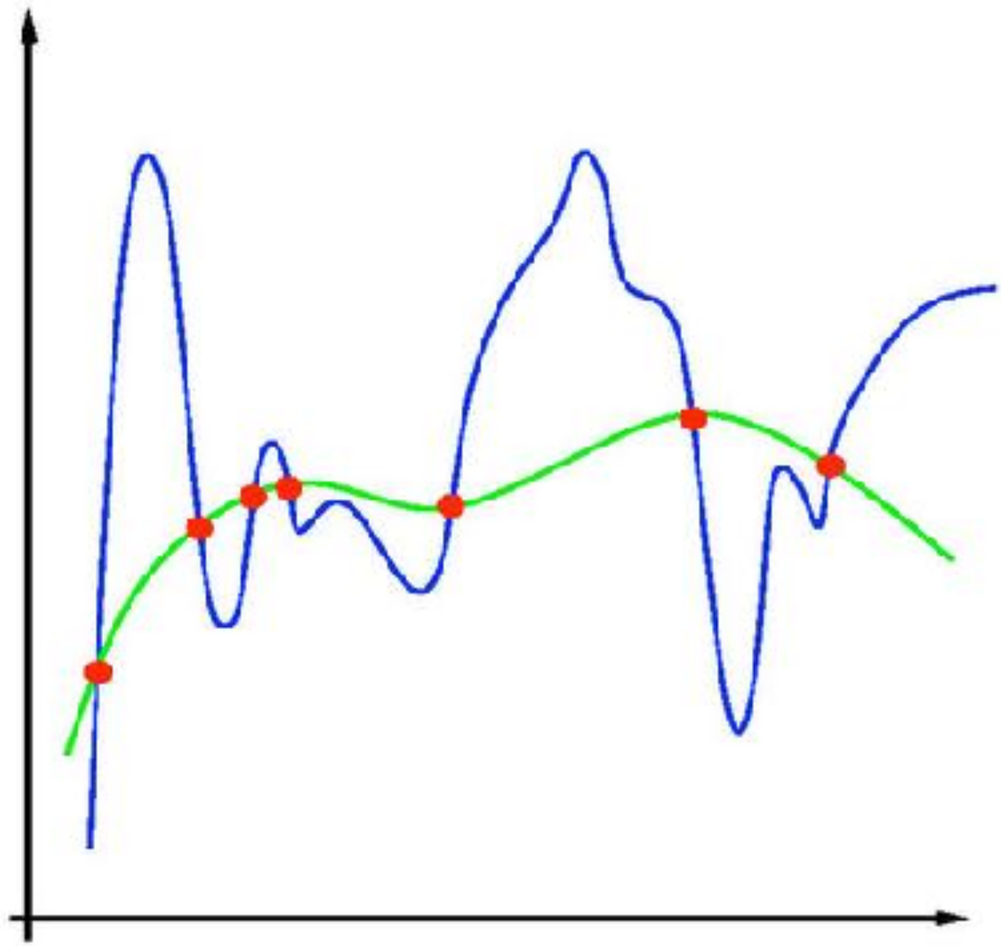
Quoc V. Le
Google
qvl@google.com



Neural network attention mechanisms

A woman is throwing a frisbee in a park.

The more parameters, the higher the chances of overfitting

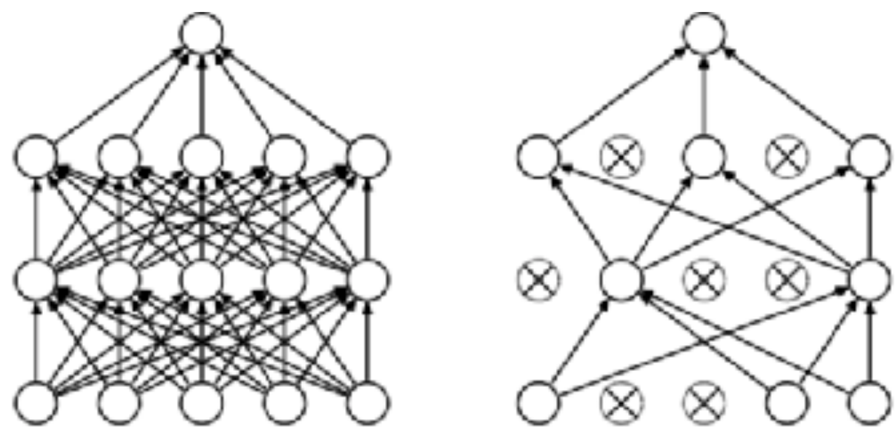


Tricks of the Trade: Early stopping

- Divide data into three sets:
  - Training
  - Testing
  - Validation
- Train model on training set
- Stop when error on validation set increases
- Evaluate accuracy on test set

Tricks of the Trade cont.

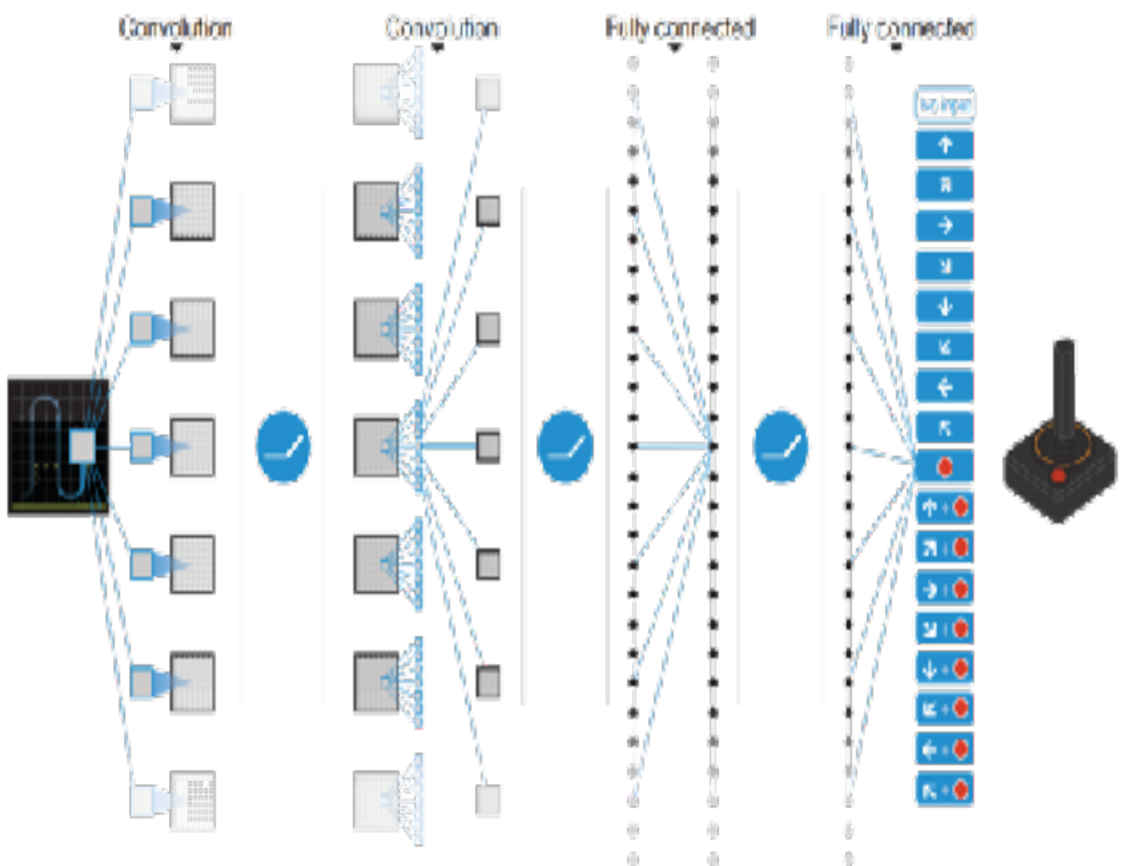
- Careful initialisation
- Dropout
- Regularization (L1, L2)
- And many more



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

Srivastava et al., 2014

Deep Q-Learning: Mnih et al. “Human-level control through deep reinforcement learning”

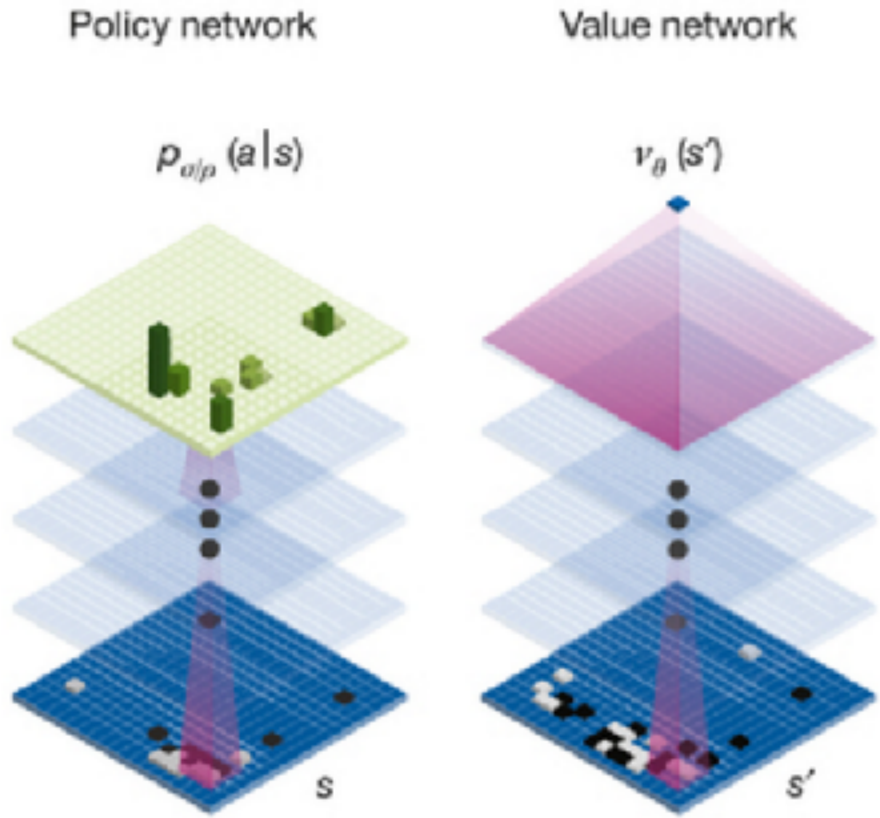


Also works for Doom!

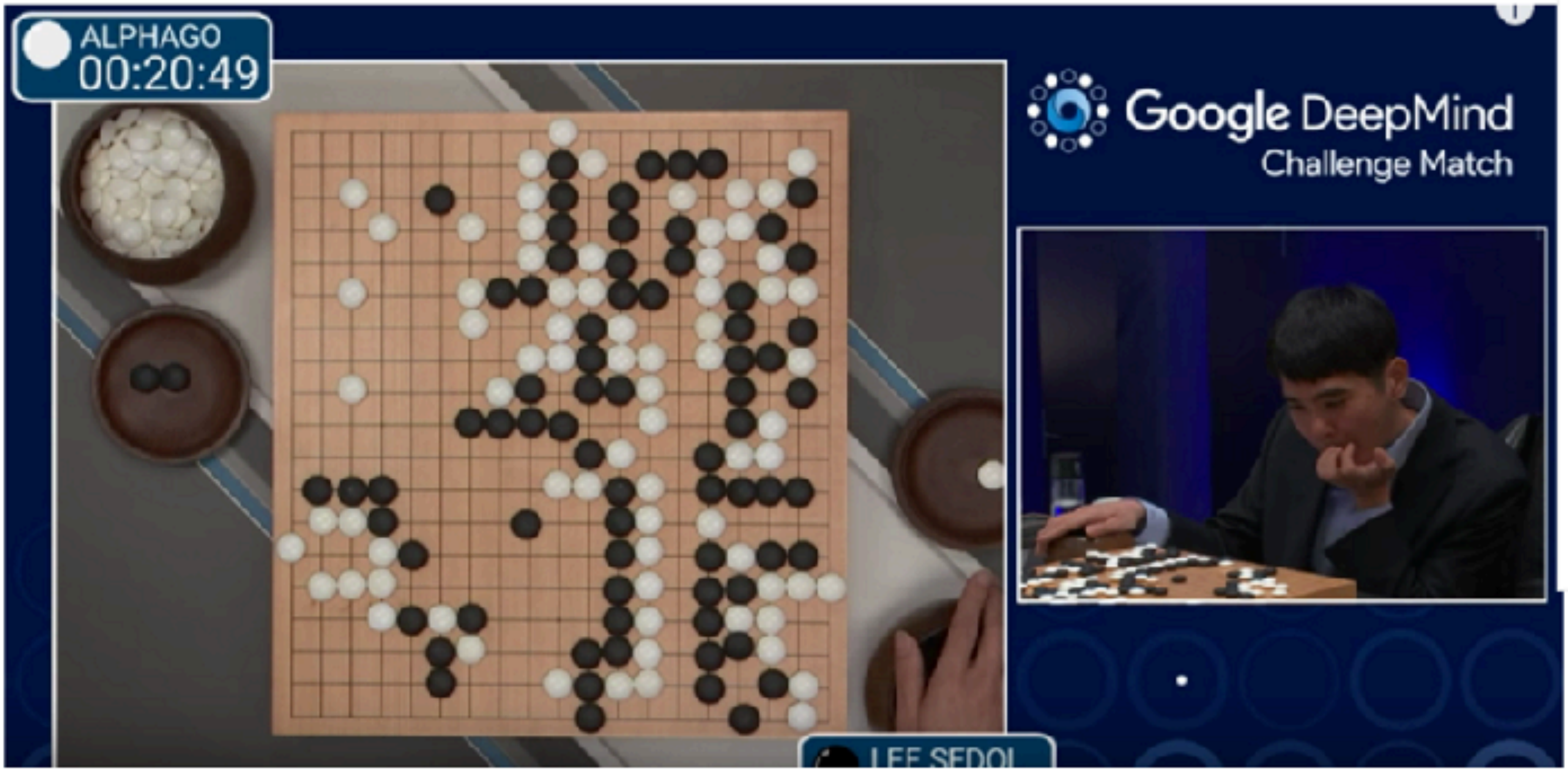
Videos: https://www.youtube.com/playlist?list=PLduGZax9wmiHg-XPFSgqGg8PEAV51q1FT
Paper: http://arxiv.org/pdf/1609.05521v1.pdf

AlphaGo

- Combining Deep Neural Networks with Tree Search



http://www.nature.com/nature/journal/v529/n7587/full/nature16961.html


ALPHAGO
00:20:49
Google DeepMind
Challenge Match