

# Neural Networks

Declan Groves

June 30 2016

# Outline

## Neural Networks

Declan Groves

What is a  
Neural  
Network?

Training:  
Gradient  
Descent

Activations

Convolutional  
Neural  
Networks

- 1 What is a Neural Network?
- 2 Training: Gradient Descent
- 3 Activations
- 4 Convolutional Neural Networks

# Overview

## Neural Networks

Declan Groves

What is a  
Neural  
Network?

Training:  
Gradient  
Descent

Activations

Convolutional  
Neural  
Networks

- Hypest ML
- Good at unstructured problems
- Suboptimal at structured problems

# History

## Neural Networks

Declan Groves

What is a  
Neural  
Network?

Training:  
Gradient  
Descent

Activations

Convolutional  
Neural  
Networks

- Around since the 1950s
- Resurgence in 1970s
- Resurgence in late 2000s

# A graphical linear model

Neural  
Networks

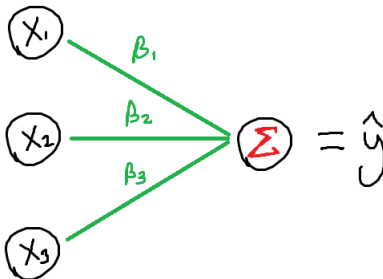
Declan Groves

What is a  
Neural  
Network?

Training:  
Gradient  
Descent

Activations

Convolutional  
Neural  
Networks



# A graphical linear model

Neural  
Networks

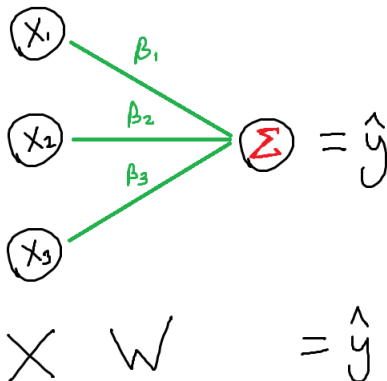
Declan Groves

What is a  
Neural  
Network?

Training:  
Gradient  
Descent

Activations

Convolutional  
Neural  
Networks



# A graphical linear model

Neural  
Networks

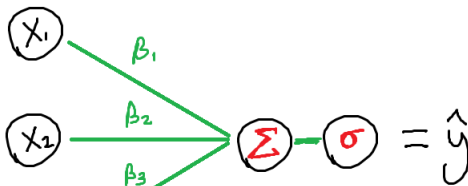
Declan Groves

What is a  
Neural  
Network?

Training:  
Gradient  
Descent

Activations

Convolutional  
Neural  
Networks



$$\sigma(XW) = \hat{y}$$

# A graphical linear model

Neural  
Networks

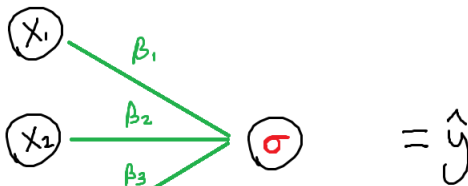
Declan Groves

What is a  
Neural  
Network?

Training:  
Gradient  
Descent

Activations

Convolutional  
Neural  
Networks



$$\sigma(XW) = \hat{y}$$



# A simple neural network

Neural  
Networks

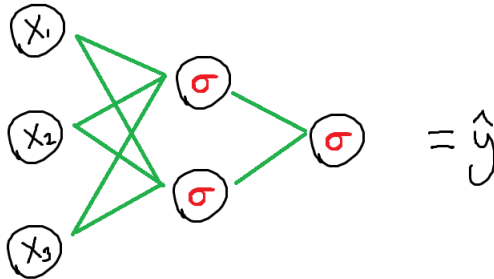
Declan Groves

What is a  
Neural  
Network?

Training:  
Gradient  
Descent

Activations

Convolutional  
Neural  
Networks



# A simple neural network

Neural  
Networks

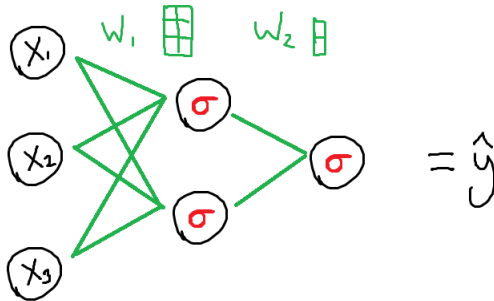
Declan Groves

What is a  
Neural  
Network?

Training:  
Gradient  
Descent

Activations

Convolutional  
Neural  
Networks



# A simple neural network

Neural  
Networks

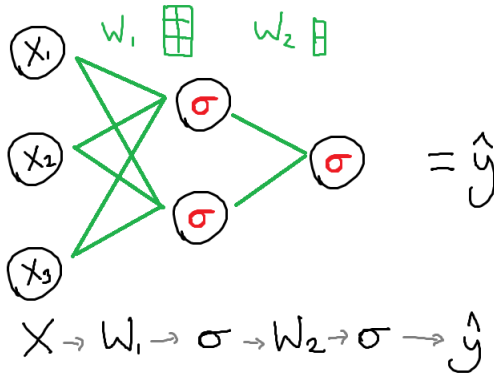
Declan Groves

What is a  
Neural  
Network?

Training:  
Gradient  
Descent

Activations

Convolutional  
Neural  
Networks



# A simple neural network

## Neural Networks

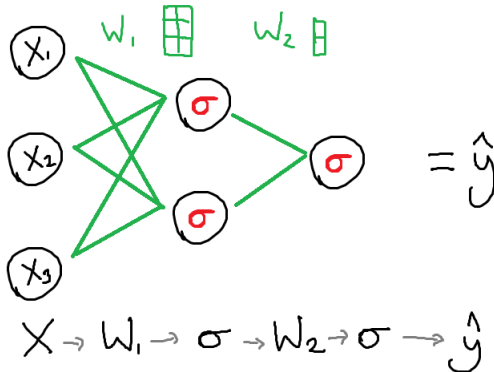
Declan Groves

What is a Neural Network?

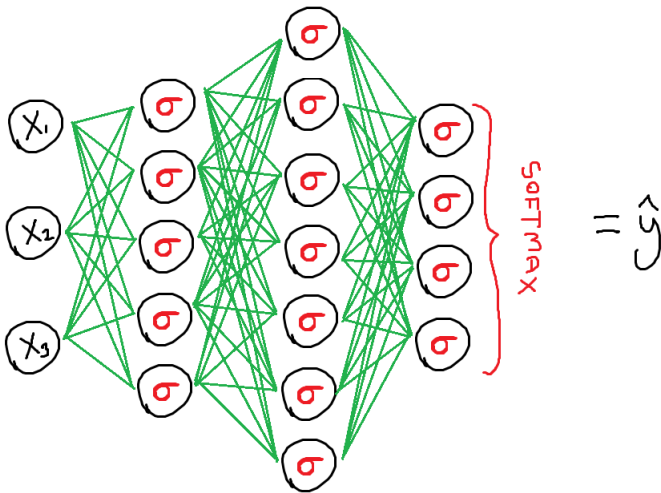
Training: Gradient Descent

Activations

Convolutional Neural Networks



# A multilayer multinomial classifier



Neural  
Networks

Declan Groves

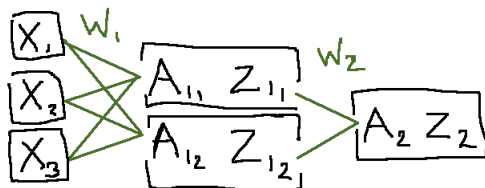
What is a  
Neural  
Network?

Training:  
Gradient  
Descent

Activations

Convolutional  
Neural  
Networks

# Forward pass



Neural  
Networks

Declan Groves

What is a  
Neural  
Network?

Training:  
Gradient  
Descent

Activations

Convolutional  
Neural  
Networks

# Forward pass

Neural  
Networks

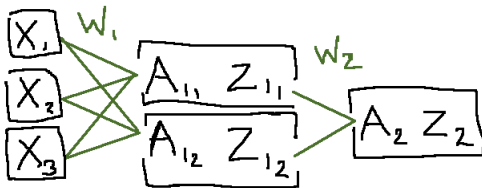
Declan Groves

What is a  
Neural  
Network?

Training:  
Gradient  
Descent

Activations

Convolutional  
Neural  
Networks



$$\left. \begin{array}{l} A_1 = XW_1 \\ Z_1 = \sigma(A_1) \\ A_2 = Z_1W_2 \\ Z_2 = \sigma(A_2) = \hat{y} \end{array} \right\} \text{Forward pass}$$

# Backwards pass

## Neural Networks

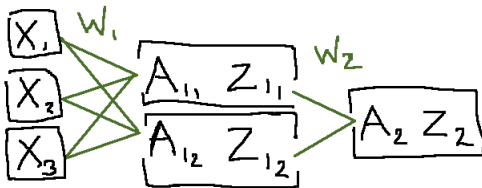
Declan Groves

What is a Neural Network?

Training: Gradient Descent

Activations

Convolutional Neural Networks



$$J = \text{err}(y, \hat{y}); J \leftarrow \frac{1}{2}(y - \hat{y})^2$$



# Backwards pass

Neural  
Networks

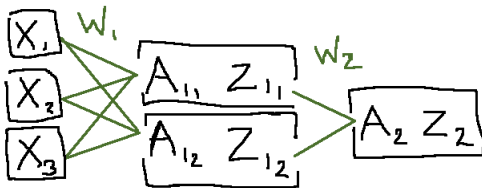
Declan Groves

What is a  
Neural  
Network?

Training:  
Gradient  
Descent

Activations

Convolutional  
Neural  
Networks



$$J = \text{err}(y, \hat{y}); J \leftarrow \frac{1}{2}(y - \hat{y})^2$$

$$\partial J / \partial W_2 =$$

# Backwards pass

## Neural Networks

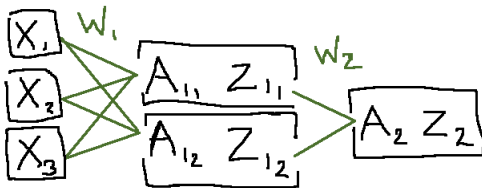
Declan Groves

What is a Neural Network?

Training: Gradient Descent

Activations

Convolutional Neural Networks



$$J = \text{err}(y, \hat{y}); J \leftarrow \frac{1}{2}(y - \hat{y})^2$$

$$\frac{\partial J}{\partial w_2} = \frac{\partial J}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial w_2}$$

# Backwards pass

## Neural Networks

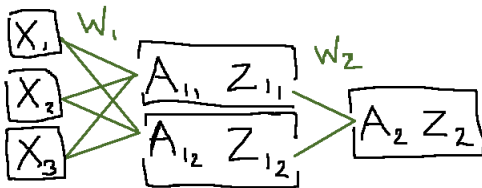
Declan Groves

What is a Neural Network?

Training: Gradient Descent

Activations

Convolutional Neural Networks



$$J = \text{err}(y, \hat{y}); J \leftarrow \frac{1}{2}(y - \hat{y})^2$$

$$\frac{\partial J}{\partial w_2} = \frac{\partial J}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial A_2} \frac{\partial A_2}{\partial w_2}$$

# Backwards pass

Neural  
Networks

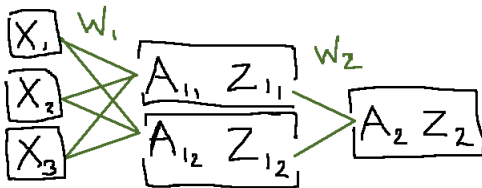
Declan Groves

What is a  
Neural  
Network?

Training:  
Gradient  
Descent

Activations

Convolutional  
Neural  
Networks



$$J = \text{err}(y, \hat{y}); J \leftarrow \frac{1}{2}(y - \hat{y})^2$$

$$\begin{aligned} \frac{\partial J}{\partial w_2} &= \frac{\partial J}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial A_2} \frac{\partial A_2}{\partial w_2} \\ &= \underbrace{(y - \hat{y}) \sigma'(A_2)}_{\delta} Z_1 \end{aligned}$$

# Backwards pass

## Neural Networks

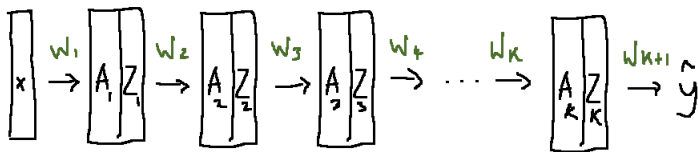
Declan Groves

What is a Neural Network?

Training: Gradient Descent

Activations

Convolutional Neural Networks



$$\partial J / \partial w_i =$$

$$\left( \partial J / \partial \hat{y} \right) \sigma'(A_k) w_k \sigma'(A_{k-1}) w_{k-1} \dots \\ \dots w_{i+1} \sigma'(A_i) Z_{i-1}$$

# Backwards pass

Neural  
Networks

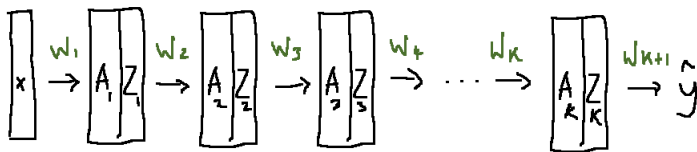
Declan Groves

What is a  
Neural  
Network?

Training:  
Gradient  
Descent

Activations

Convolutional  
Neural  
Networks



$$\partial J / \partial w_i =$$

$$\left( \partial J / \partial \hat{y} \right) \underline{\sigma'}(A_k) w_k \underline{\sigma'}(A_{k-1}) w_{k-1} \dots$$

$$\dots w_{i+1} \underline{\sigma'}(A_i) Z_{i-1} \quad !!$$

# Backwards pass

## Neural Networks

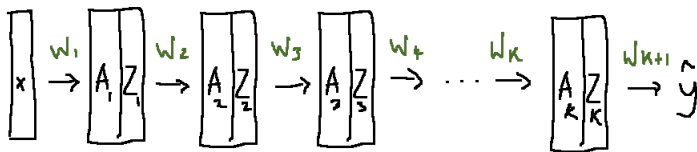
Declan Groves

What is a Neural Network?

Training: Gradient Descent

Activations

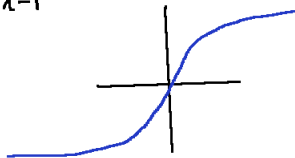
Convolutional Neural Networks



$$\partial J / \partial w_{i+1} =$$

$$\left( \partial J / \partial \hat{y} \right) \sigma'(A_k) w_k \sigma'(A_{k-1}) w_{k-1} \dots$$

$$\dots w_{i+1} \sigma'(A_i) Z_{i-1}$$



# Resources

Neural  
Networks

Declan Groves

What is a  
Neural  
Network?

Training:  
Gradient  
Descent

Activations

Convolutional  
Neural  
Networks

- Cool youtube series:  
<https://www.youtube.com/watch?v=bxe2T-V8XR8>
- Pain and toil: Elements of Statistical Learning ch11



# Linear

## Neural Networks

Declan Groves

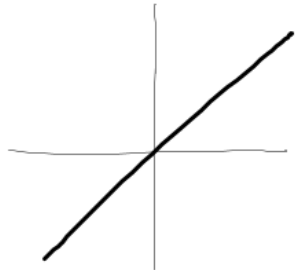
What is a  
Neural  
Network?

Training:  
Gradient  
Descent

## Activations

Convolutional  
Neural  
Networks

## ■ Useless



# Threshold

## Neural Networks

Declan Groves

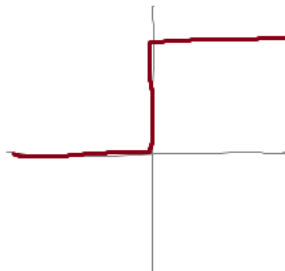
What is a Neural Network?

Training: Gradient Descent

## Activations

Convolutional Neural Networks

- Similar to biological neuron
- No gradient



# Sigmoid

## Neural Networks

Declan Groves

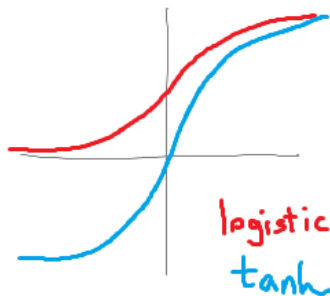
What is a  
Neural  
Network?

Training:  
Gradient  
Descent

## Activations

Convolutional  
Neural  
Networks

- *tanh* preferred
- Gradients can vanish



# ReLU

## Neural Networks

Declan Groves

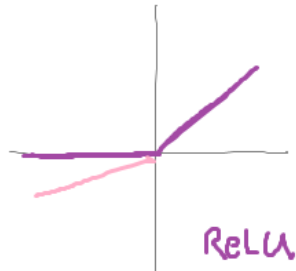
What is a Neural Network?

Training: Gradient Descent

## Activations

Convolutional Neural Networks

- No vanishing gradient
- Cheap to compute
- Can explode and die
- Popular with CNNs



ReLU

Leaky ReLU

# Radial basis functions

## Neural Networks

Declan Groves

What is a Neural Network?

Training: Gradient Descent

## Activations

Convolutional Neural Networks

- Gaussian + others
- Train very quickly
- Good at interpolation



# Convolutional Neural Networks

Neural  
Networks

Declan Groves

What is a  
Neural  
Network?

Training:  
Gradient  
Descent

Activations

Convolutional  
Neural  
Networks

- Image recognition killer

# Conceptual structure

## Neural Networks

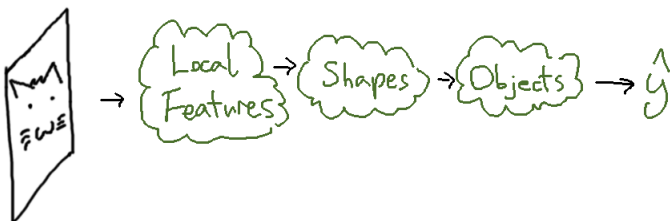
Declan Groves

What is a Neural Network?

Training: Gradient Descent

Activations

Convolutional Neural Networks



# MLP $\rightarrow$ too many weights!

## Neural Networks

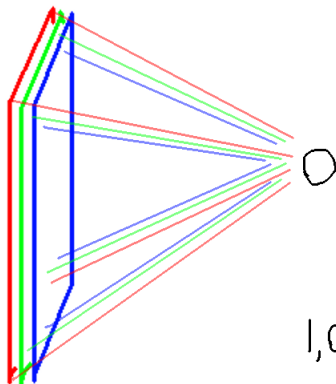
Declan Groves

What is a Neural Network?

Training: Gradient Descent

Activations

Convolutional Neural Networks



$$= 3,072 \text{ weights } N^{-1}$$

1,024 Neurons  
 $\rightarrow$   $> 3M$  weights



# Local connectivity

## Neural Networks

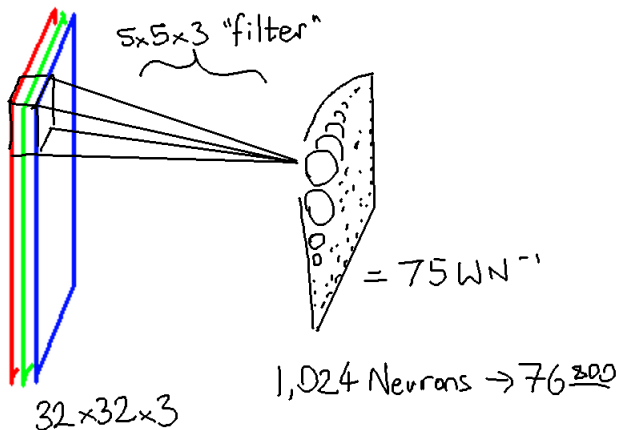
Declan Groves

What is a  
Neural  
Network?

Training:  
Gradient  
Descent

Activations

Convolutional  
Neural  
Networks



# Local connectivity + convolution

## Neural Networks

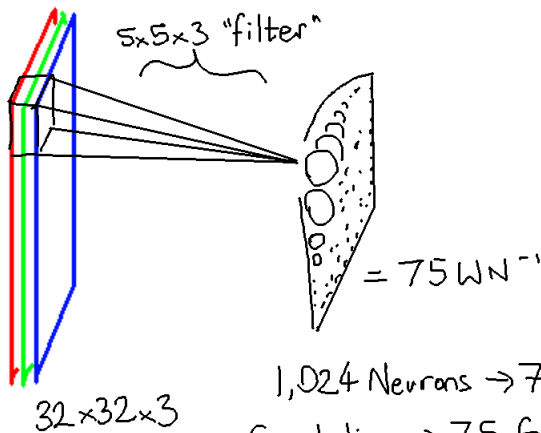
Declan Groves

What is a Neural Network?

Training: Gradient Descent

Activations

Convolutional Neural Networks



1,024 Neurons  $\rightarrow$  76 800  
Convolution  $\rightarrow$  75 for Layer

# Local connectivity + convolution

## Neural Networks

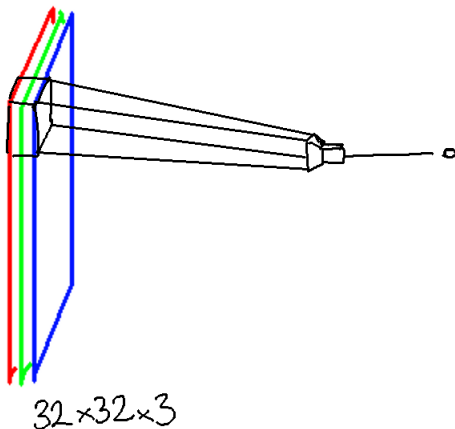
Declan Groves

What is a Neural Network?

Training: Gradient Descent

Activations

Convolutional Neural Networks



# Local connectivity + convolution

## Neural Networks

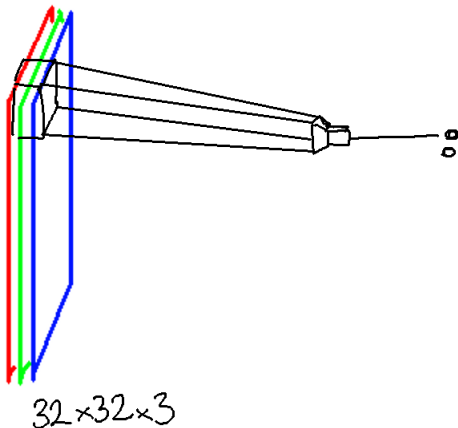
Declan Groves

What is a  
Neural  
Network?

Training:  
Gradient  
Descent

Activations

Convolutional  
Neural  
Networks



# Local connectivity + convolution

## Neural Networks

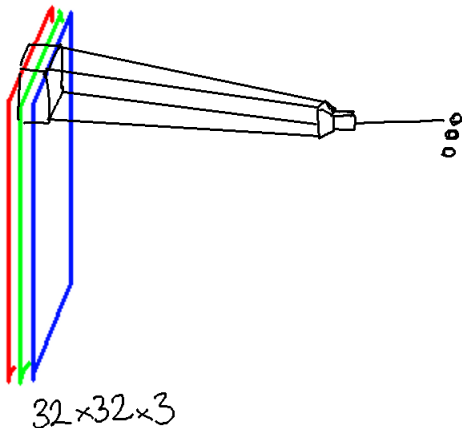
Declan Groves

What is a  
Neural  
Network?

Training:  
Gradient  
Descent

Activations

Convolutional  
Neural  
Networks



# Local connectivity + convolution

## Neural Networks

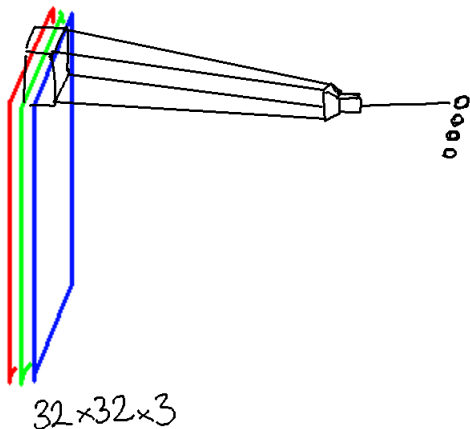
Declan Groves

What is a  
Neural  
Network?

Training:  
Gradient  
Descent

Activations

Convolutional  
Neural  
Networks



# Local connectivity + convolution

Neural  
Networks

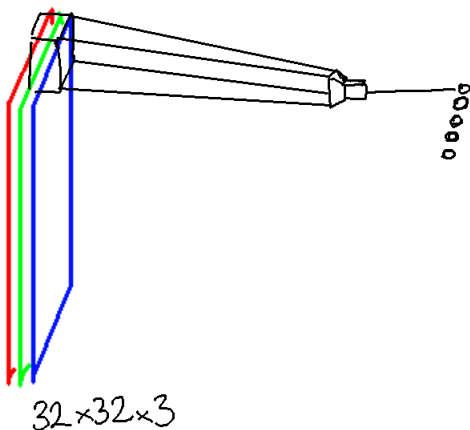
Declan Groves

What is a  
Neural  
Network?

Training:  
Gradient  
Descent

Activations

Convolutional  
Neural  
Networks



# One filter forms a surface

## Neural Networks

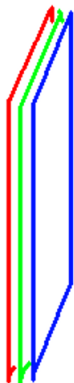
Declan Groves

What is a Neural Network?

Training: Gradient Descent

Activations

Convolutional Neural Networks



$32 \times 32 \times 3$



$32 \times 32 \times 1$



# Many filters form a volume

## Neural Networks

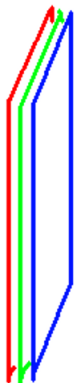
Declan Groves

What is a Neural Network?

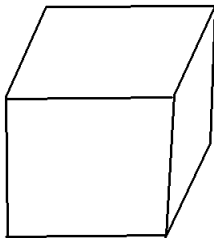
Training: Gradient Descent

Activations

Convolutional Neural Networks



$32 \times 32 \times 3$



$32 \times 32 \times N_F$

# Typical processing structure

## Neural Networks

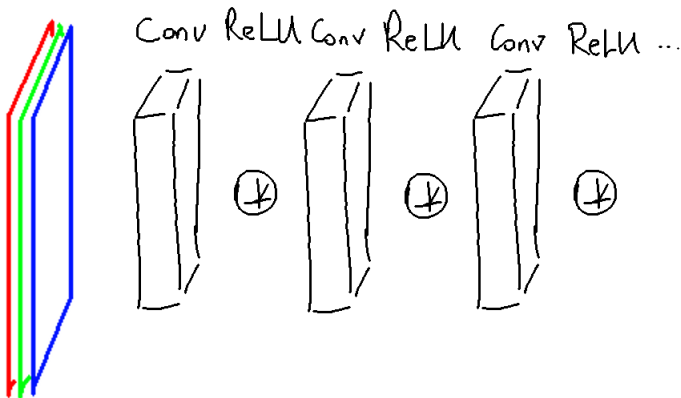
Declan Groves

What is a Neural Network?

Training: Gradient Descent

Activations

Convolutional Neural Networks



# AlexNet's filters

Neural  
Networks

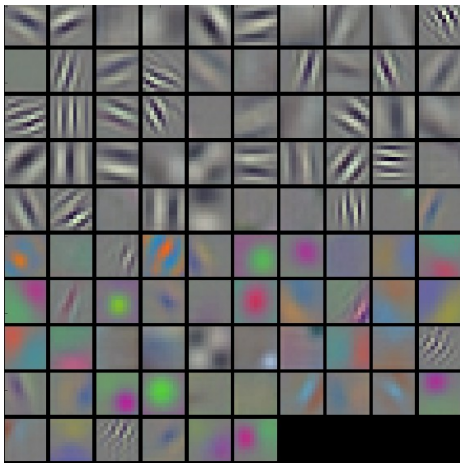
Declan Groves

What is a  
Neural  
Network?

Training:  
Gradient  
Descent

Activations

Convolutional  
Neural  
Networks



# Pooling

## Neural Networks

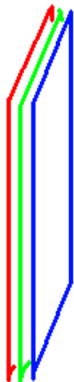
Declan Groves

What is a Neural Network?

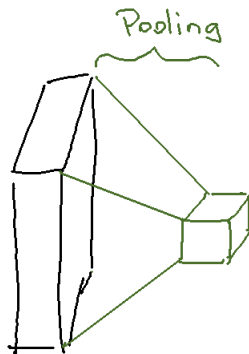
Training: Gradient Descent

Activations

Convolutional Neural Networks



$\{\text{Conv}, \text{ReLU}\}_m$



$\{x, y, 3\} \rightarrow \{x, y, k\} \rightarrow \{\ll x, \ll y, k\}$

# Pooling

## Neural Networks

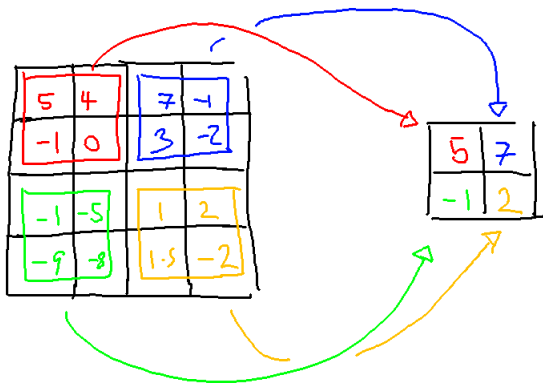
Declan Groves

What is a Neural Network?

Training: Gradient Descent

Activations

Convolutional Neural Networks



# Typical structure of a CNN

Neural  
Networks

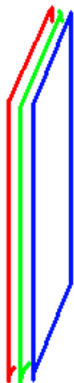
Declan Groves

What is a  
Neural  
Network?

Training:  
Gradient  
Descent

Activations

Convolutional  
Neural  
Networks



Standard Convnet

$$\{\{Conv, ReLU\}_m Pool\}_N \quad \{FC\}_K \rightarrow \hat{y}$$