Свёрточные нейронные сети

Часть 1: Основные понятия, архитектура

План

- Туториалы и домашнее задание
- Мотивация
- Свёртки, Stride, Padding
- Свёрточный слой
- CNN
- AlexNet

TensorFlow tutorials

Hvass-Labs / TensorFlow-Tutorials

Watch 329 | Star 3,829 | Fork 1,785

Code | Issues 0 | Pull requests 0 | Insights

Branch: master | TensorFlow-Tutorials / 02_Convolutional_Neural_Network.ipynb | Find file | Copy path

Hvass-Labs Minor fixes. 92d3b3c on 14 Dec 2017

1 contributor

2337 lines (2336 sloc) | 225 KB

Raw | Blame | History

TensorFlow Tutorial #02

Convolutional Neural Network

by Magnus Erik Hvass Pedersen / GitHub / Videos on YouTube

Introduction

The previous tutorial showed that a simple linear model had about 91% classification accuracy for recognizing hand-written digits in the MNIST data-set.

Kaggle

Tracks

Machine Learning

Machine learning is the hottest field in data science, and this track will get you started quickly.

Pandas

Short hands-on challenges to perfect your data manipulation skills

Data Visualisation

Visualisation is one of the most versatile skills in data science. Make insightful and beautiful graphics to see what's happening in any dataset.

SQL

Learn SQL for working with databases, using Google BigQuery to scale to massive datasets.

R

Learn the language designed for data analysis. This track includes data set-up, machine learning and data visualization.

Deep Learning

Learn TensorFlow to take Machine Learning to the next level. Your new skills will amaze you.

Домашка

Convolutional Neural Networks: Step by Step

Welcome to Course 4's first assignment! In this assignment, you will implement convolutional (CONV) and pooling (POOL) layers in numpy, including both forward propagation and (optionally) backward propagation.

Notation:

- Superscript $[l]$ denotes an object of the lth layer.
  - Example: $a^{[4]}$ is the 4th layer activation. $W^{[5]}$ and $b^{[5]}$ are the 5th layer parameters.
- Superscript (i) denotes an object from the ith example.
  - Example: $x^{(i)}$ is the ith training example input.
- Lowerscript i denotes the ith entry of a vector.
  - Example: $a_i^{[l]}$ denotes the ith entry of the activations in layer l, assuming this is a fully connected (FC) layer.
- n_H, n_W and n_C denote respectively the height, width and number of channels of a given layer. If you want to reference a specific layer l, you can also write $n_H^{[l]}$, $n_W^{[l]}$, $n_C^{[l]}$.
- $n_{H_{prev}}$, $n_{W_{prev}}$ and $n_{C_{prev}}$ denote respectively the height, width and number of channels of the previous layer. If referencing a specific layer l, this could also be denoted $n_H^{[l-1]}$, $n_W^{[l-1]}$, $n_C^{[l-1]}$.

We assume that you are already familiar with `numpy` and/or have completed the previous courses of the specialization. Let's get started!

Мотивация

Успехи кибернетики поставили в порядок дня создание эффективно действующих устройств, способных выполнять некоторые высшие, «психические» функции, считавшиеся до сих пор доступными лишь человеческому мозгу. Это задача первостепенной важности для ряда областей науки, техники, народного хозяйства, космонавтики, для успешного развития которых требуется сделать следующий шаг в развитии автоматики.

Из книги
Ф. Розенблата
«Основы
нейродинамики»

Архитектура(зрительной коры)

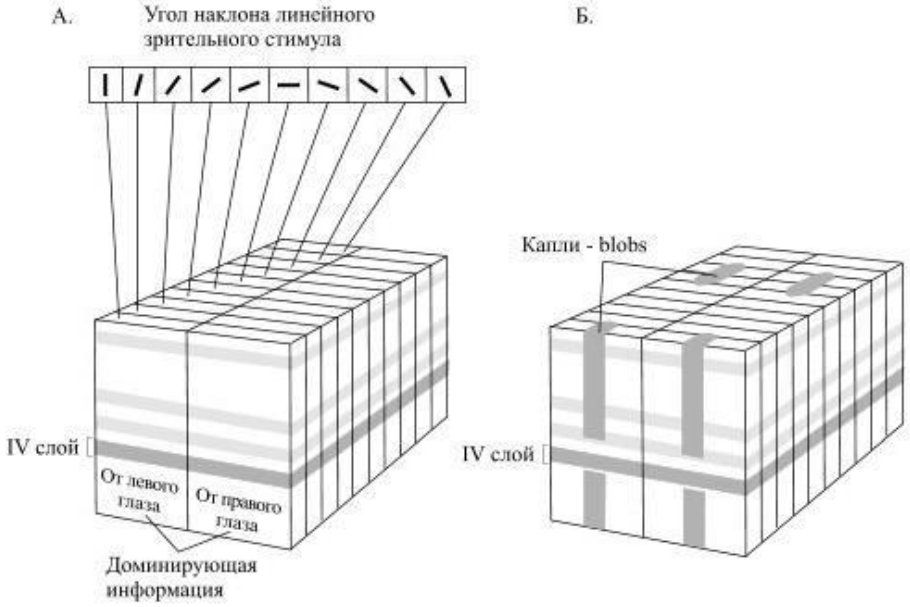
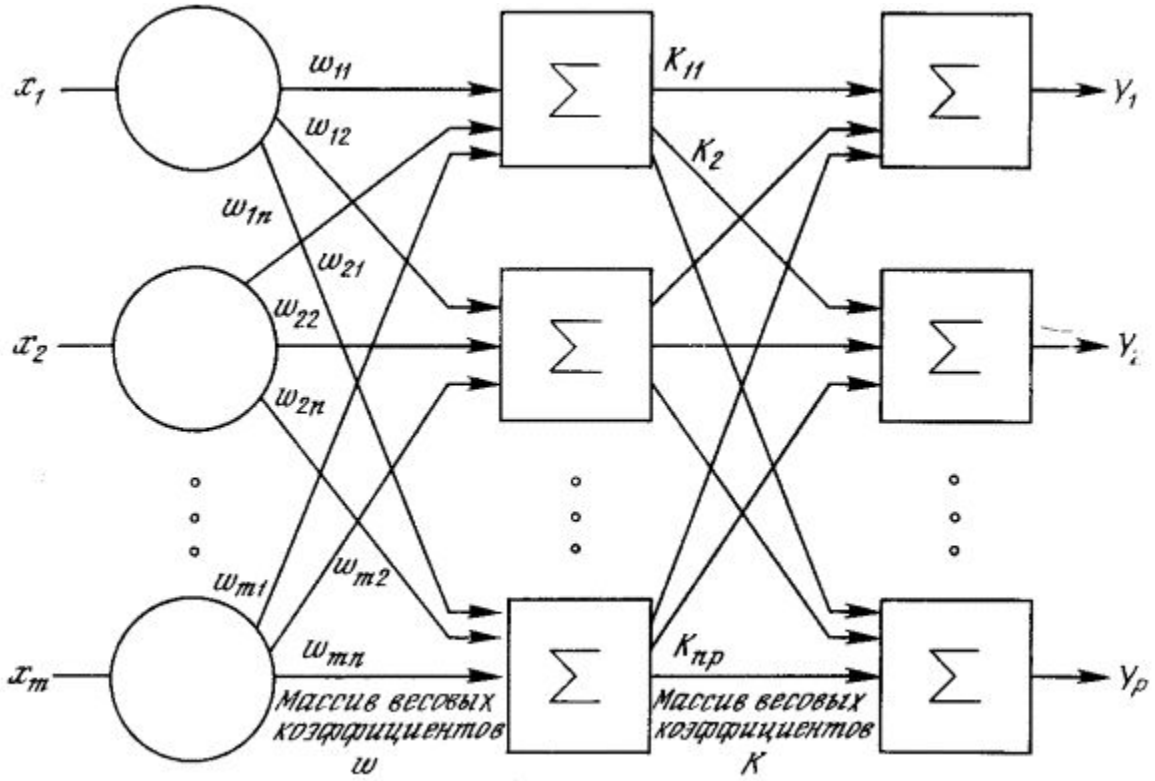
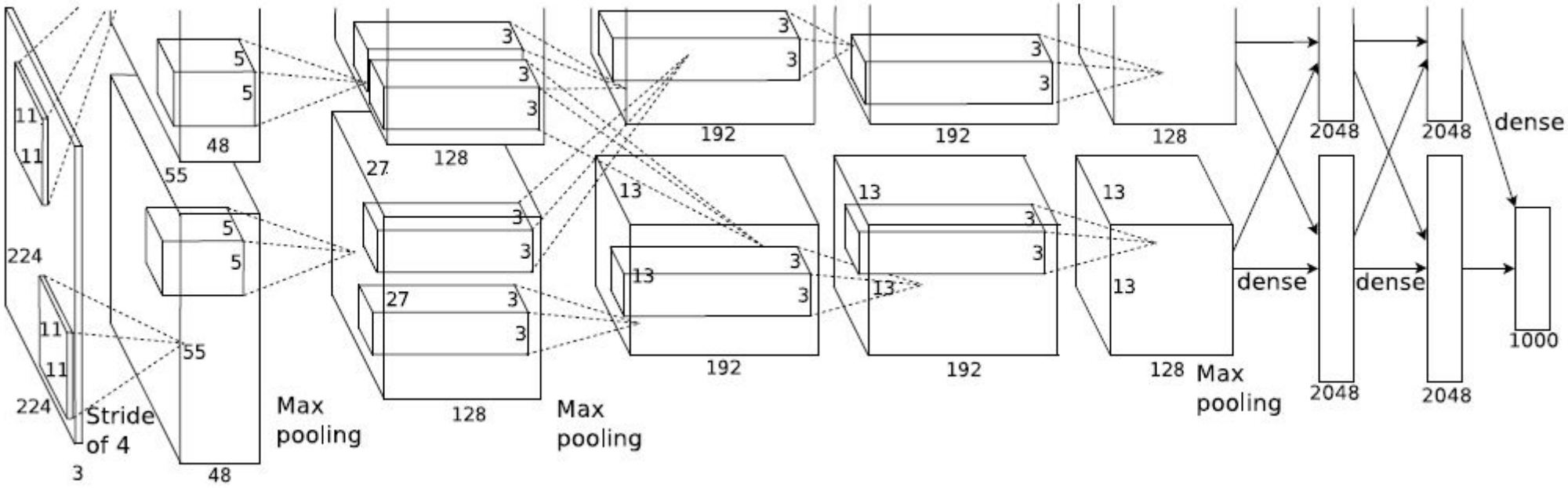


Рис. 9.6. Зрительные кортикальные колонки
А. Каждая колонка ориентирована на определенный наклон линейного стимула; у большинства бинокулярных клеток обнаружена глазодоминантность.
Б. Скопления нейронов, специализирующихся на переработке информации о цветовых характеристиках стимула, отсутствуют в IV слое и носят название капель - blobs (по Livingstone M.S. и Hubel D.H., 1984).

Архитектура(многослойный перцептрон)



AlexNet

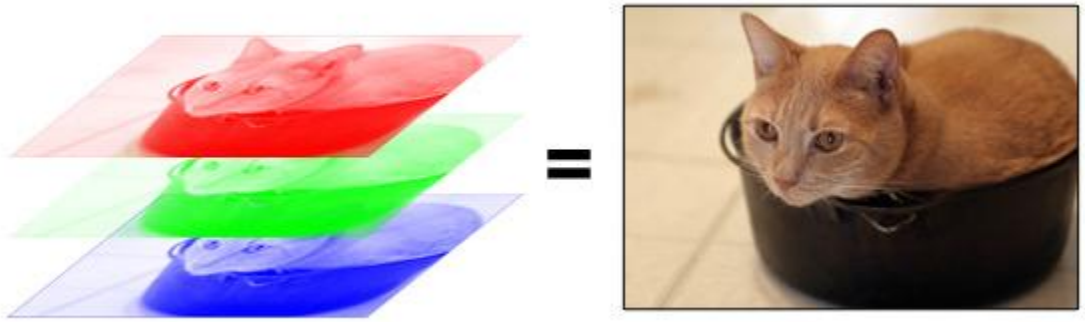

224
224
3
11
11
11
11
Stride
of 4
55
55
48
48
5
5
5
5
Max
pooling
27
27
128
128
3
3
3
3
3
3
3
3
Max
pooling
13
13
13
13
192
192
3
3
3
3
192
192
3
3
3
3
128
128
13
13
13
13
Max
pooling
dense
dense
dense
2048
2048
2048
2048
1000

Как выглядят наши данные?

- Существует много разных типов данных, но все они, вообще говоря, представляют собой некоторый многомерный массив «тензор»
- Данные, как правило, имеют определённую структуру. Направления по осям неравнозначные (напр. шкала времени)

Как выглядят наши данные?

Главный представитель для нас – изображения. В данном случае изображение
цветное, поэтому каждый пиксель это три числа, соответствующие своим каналам RGB, т.е. цветное изображение RGB это трёхмерные «тензоры»



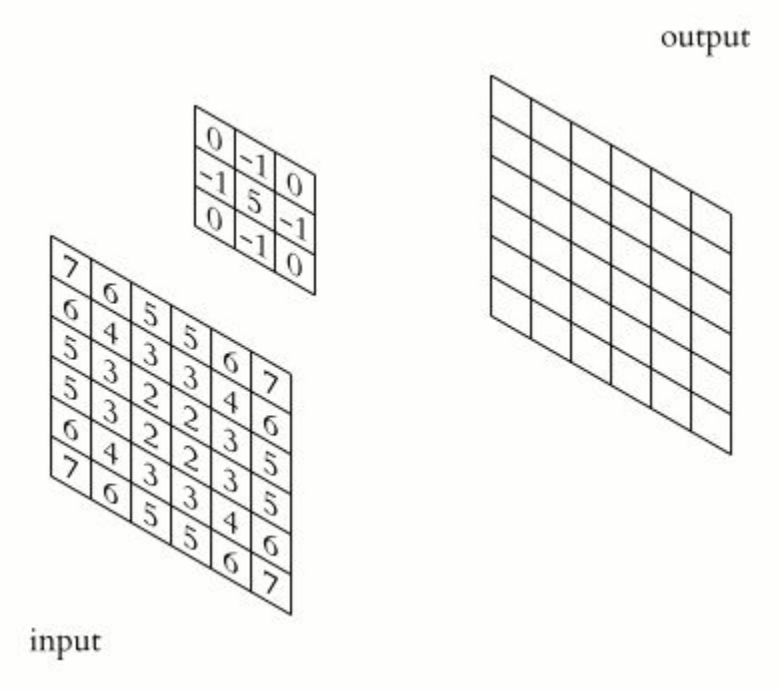
Как выглядят наши данные?

В данном случае изображения чёрно-белые и задаются матрицами интенсивностей, т. е. представляют собой двухмерные «тензоры»

Свёртка(интуиция)

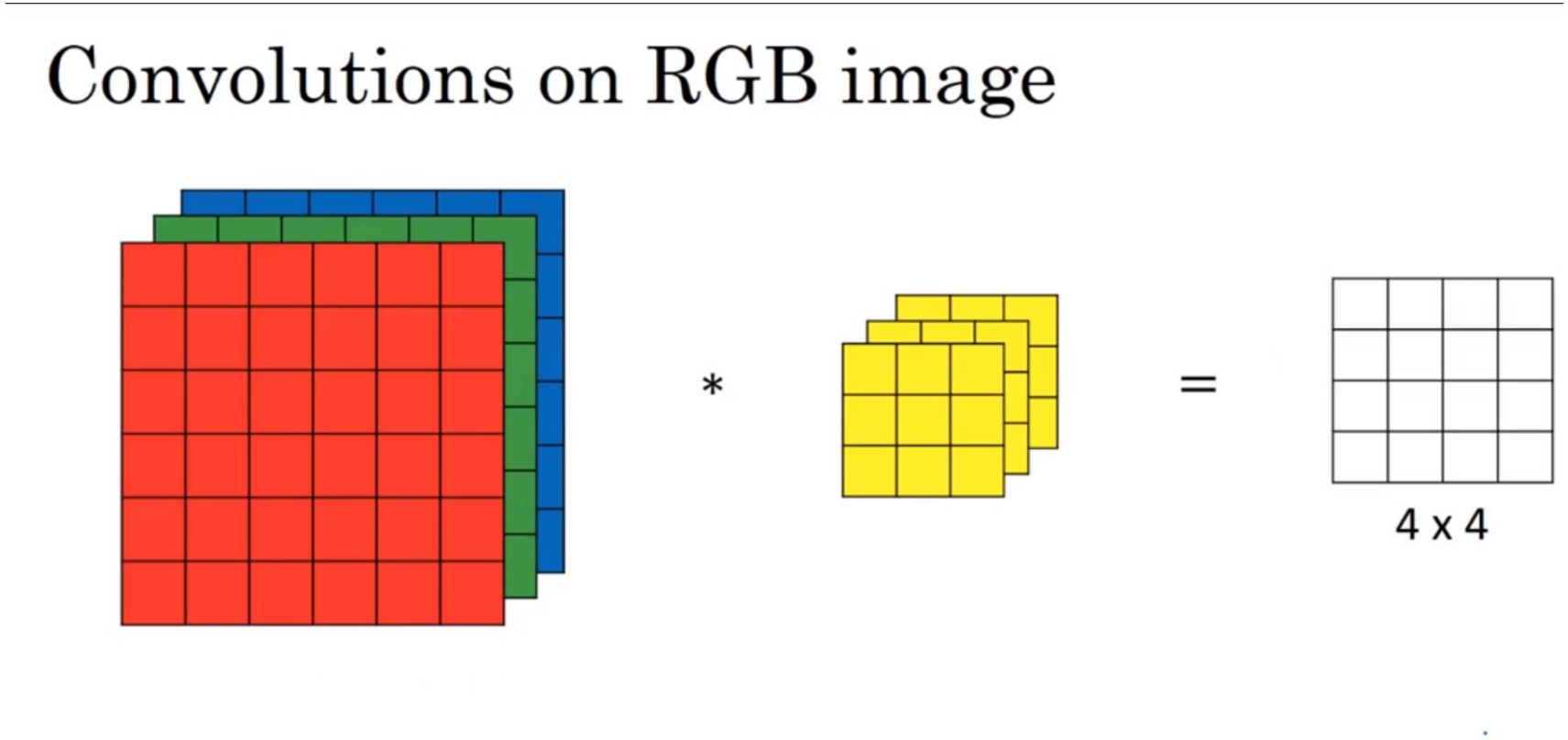
- Распознавание объектов на изображении должно осуществляться безотносительно их положения
- Нужно использовать наши знания о структуре и топологии данных(в обычных полносвязных сетях этого нет)

Свёртка 2D



gif was stolen from Wiki

Свёртка 3D

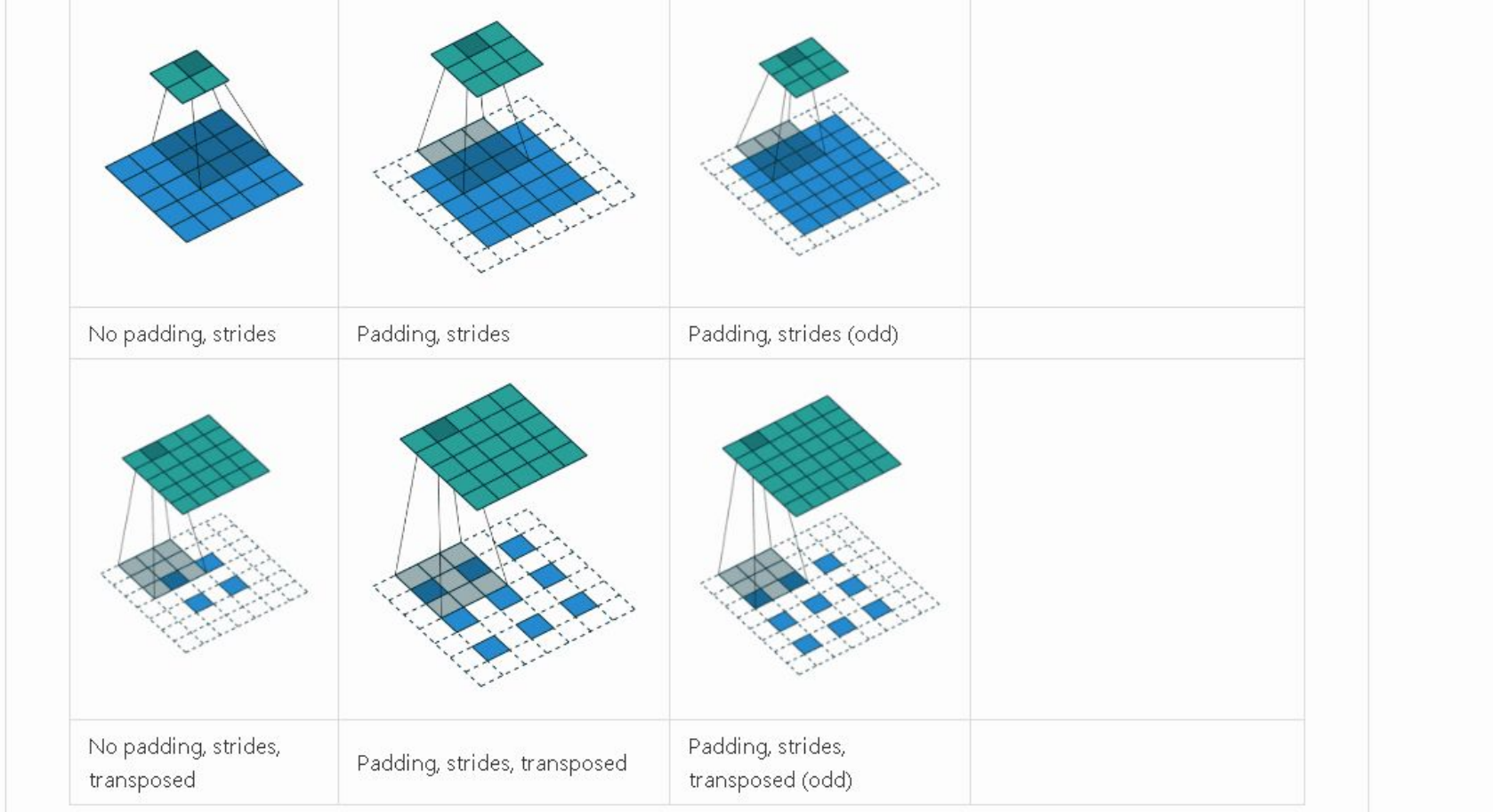


Слайд из презентации
Andrew Ng

Padding

0	0	0	0	0	0
0	35	19	25	6	0
0	13	22	16	53	0
0	4	3	7	10	0
0	9	8	1	3	0
0	0	0	0	0	0

Stride


No padding, strides
Padding, strides
Padding, strides (odd)
No padding, strides, transposed
Padding, strides, transposed
Padding, strides, transposed (odd)

Summary of convolutions

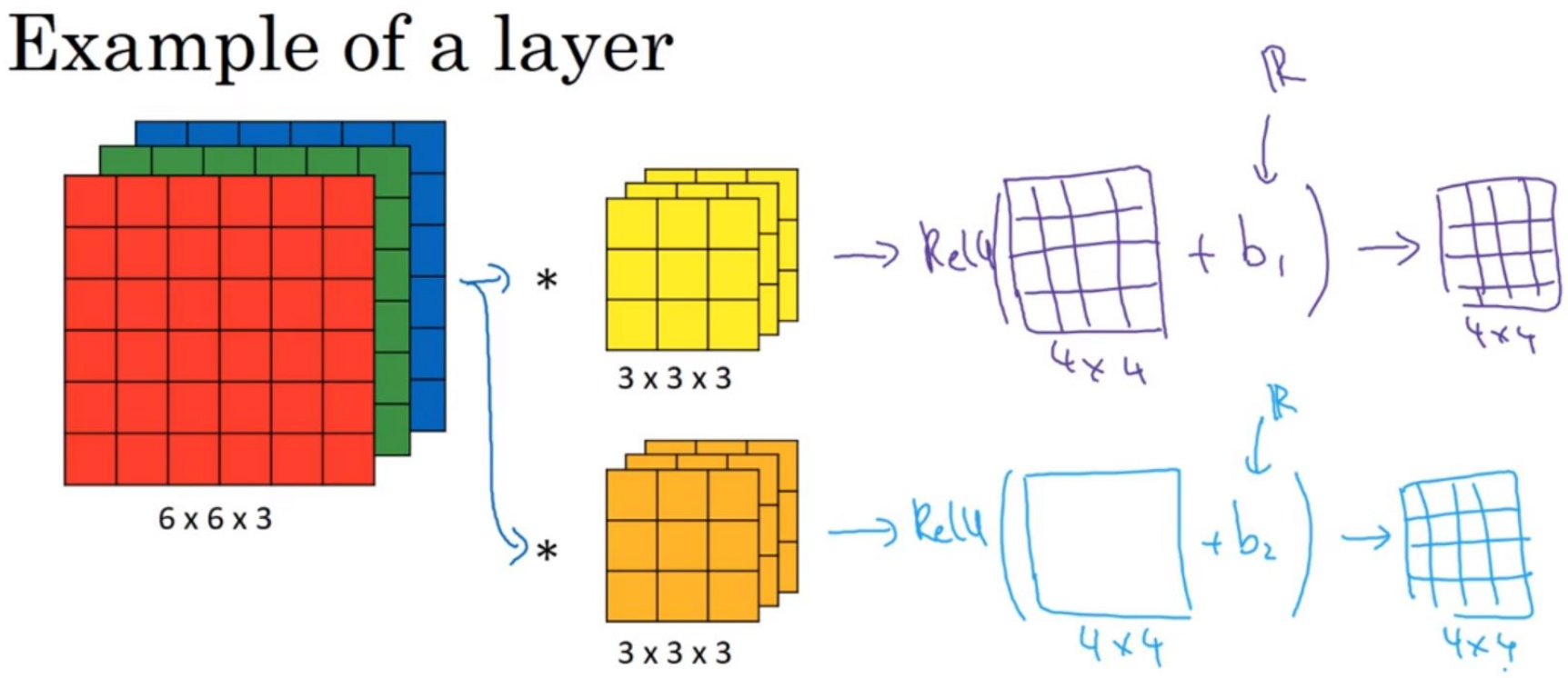
$n \times n$ image $\quad f \times f$ filter

padding p $\quad$ stride s

$$\left\lfloor \frac{n+2p-f}{s} + 1 \right\rfloor \quad \times \quad \left\lfloor \frac{n+2p-f}{s} + 1 \right\rfloor$$

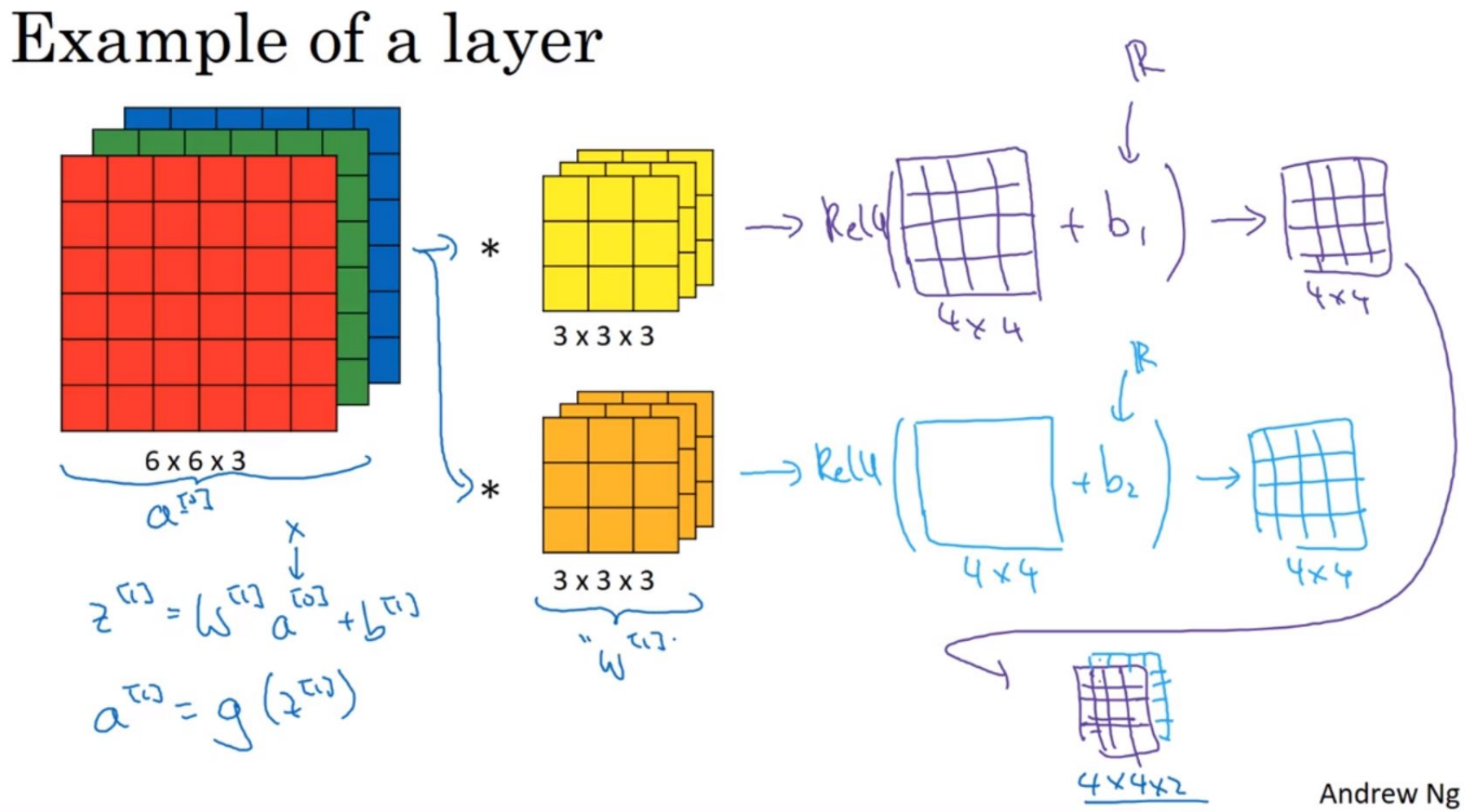
Andrew Ng

Свёрточный слой



Andrew Ng

Свёрточный слой



Свёрточный слой

Summary of notation

If layer l is a convolution layer:

$f^{[l]}$ = filter size

$p^{[l]}$ = padding

$s^{[l]}$ = stride

$\boxed{n_c^{[l]}}$ = number of filters

Each filter is: $f^{[l]} \times f^{[l]} \times \boxed{n_c^{[l-1]}}$

Activations: $a^{[l]} \rightarrow n_H^{[l]} \times n_W^{[l]} \times n_c^{[l]}$

Weights: $f^{[l]} \times f^{[l]} \times n_c^{[l-1]} \times n_c^{[l]}$

bias: $n_c^{[l]} - (1,1,1,n_c^{[l]})$ ← #filters in layer l.

Input: $n_H^{[l-1]} \times n_W^{[l-1]} \times \boxed{n_c^{[l-1]}}$

Output: $n_H^{[l]} \times n_W^{[l]} \times \boxed{n_c^{[l]}}$

$$n_{H,W}^{[l]} = \left\lfloor \frac{n_{H,W}^{[l-1]} + 2p^{[l]} - f^{[l]}}{s^{[l]}} + 1 \right\rfloor$$

$$A^{[l]} \rightarrow m \times \underbrace{n_H^{[l]} \times n_W^{[l]} \times n_c^{[l]}}_{n_c \times n_H \times n_W}$$

Andrew Ng

Свёрточный слой

- Сколько параметров в свёрточном слое с 10 фильтрами размером 3 x 3 x 3?
- Ответ 280

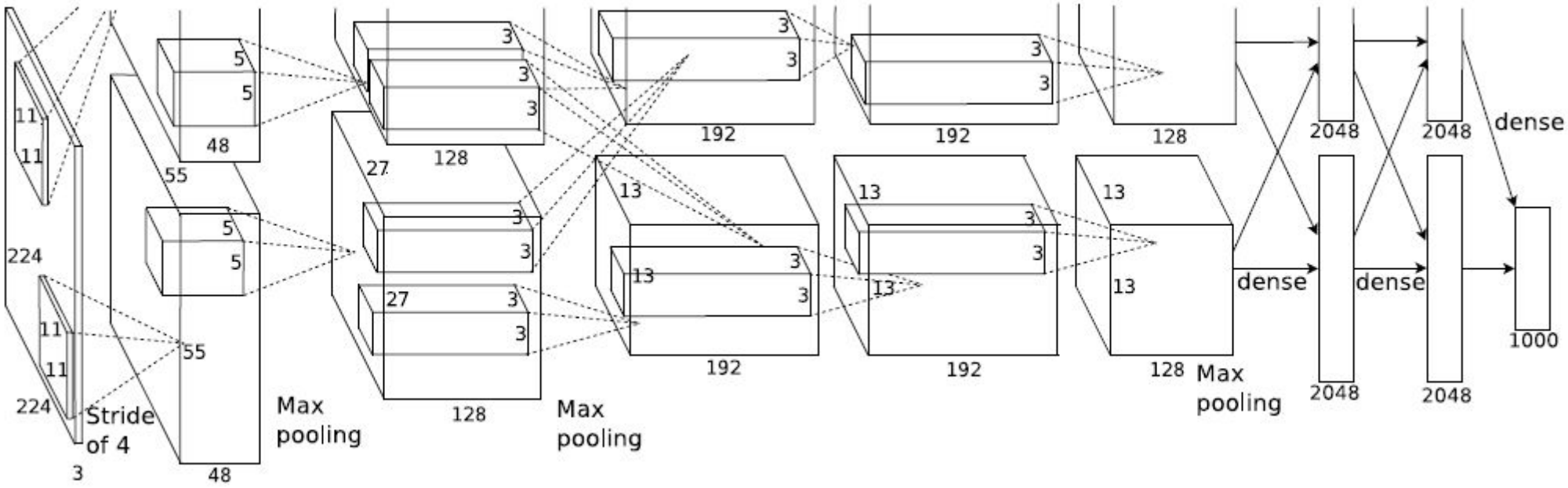
CNN это

- Свёрточные слои
- FC-слои
- Pooling-слои

Pooling layer

- Max Pooling
- Average Pooling

AlexNet


224
224
3
11
11
11
11
Stride
of 4
55
55
48
48
5
5
5
5
Max
pooling
27
27
128
128
3
3
3
3
3
3
3
3
Max
pooling
13
13
13
13
192
192
3
3
3
3
192
192
3
3
3
3
128
128
13
13
13
13
Max
pooling
dense
dense
dense
2048
2048
2048
2048
1000

AlexNet

Abstract

We trained a large, deep convolutional neural network to classify the 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into the 1000 different classes. On the test data, we achieved top-1 and top-5 error rates of 37.5% and 17.0% which is considerably better than the previous state-of-the-art. The neural network, which has 60 million parameters and 650,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers, and three fully-connected layers with a final 1000-way softmax. To make training faster, we used non-saturating neurons and a very efficient GPU implementation of the convolution operation. To reduce overfitting in the fully-connected layers we employed a recently-developed regularization method called "dropout" that proved to be very effective. We also entered a variant of this model in the ILSVRC-2012 competition and achieved a winning top-5 test error rate of 15.3%, compared to 26.2% achieved by the second-best entry.

AlexNet

Abstract

We trained a large, deep convolutional neural network to classify the 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into the 1000 different classes. On the test data, we achieved top-1 and top-5 error rates of 37.5% and 17.0% which is considerably better than the previous state-of-the-art. The neural network, which has 60 million parameters and 650,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers, and three fully-connected layers with a final 1000-way softmax. To make training faster, we used non-saturating neurons and a very efficient GPU implementation of the convolution operation. To reduce overfitting in the fully-connected layers we employed a recently-developed regularization method called "dropout" that proved to be very effective. We also entered a variant of this model in the ILSVRC-2012 competition and achieved a winning top-5 test error rate of 15.3%, compared to 26.2% achieved by the second-best entry.

Интересные факты

Действительно, язык современной математики (как, впрочем, и всякий другой язык) представляет собой исторически сложившуюся форму выражения, возникшую из совершенно других задач, в частности из задач физики, в применении к которым он может с успехом выполнять свои функции. Естественно, что этот же аппарат не может без существенных изменений быть применен к биологическим и физиологическим проблемам. По-видимому, здесь нужно создавать новый математический язык, более приспособленный к исследованиям физиологических процессов, а не привязывать уже существующий к задачам физиологии. Конечно, до тех пор, пока такого нового языка нет, можно пользоваться только тем, который есть, имея, однако, в виду необходимые изменения и в аппарате по мере накопления результатов. Наиболее перспективным здесь представляется создание