

# Анализ данных **и** Машинное обучение в гидрологии

...

Неделя 8

# Последняя неделя

...

# План

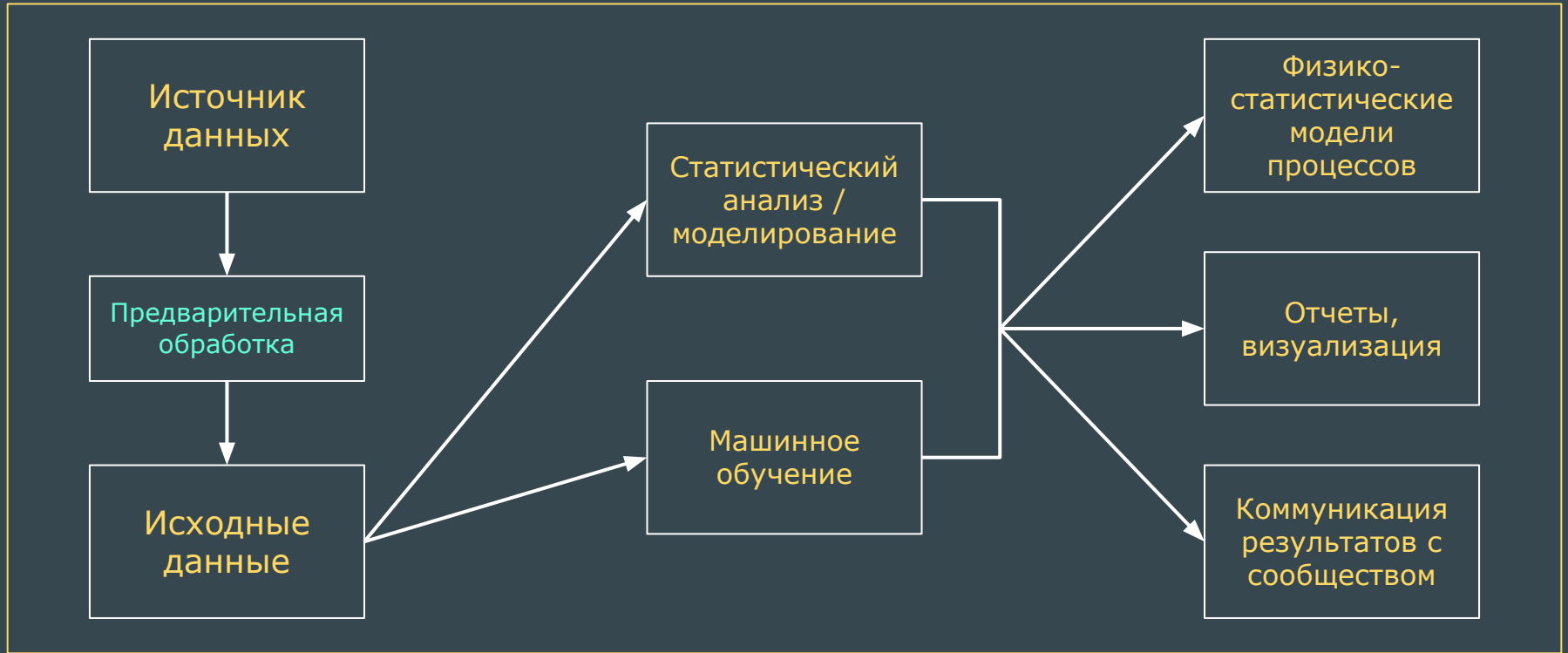
## Лекция

- глубинное или глубокое обучение?
- предпосылки развития
- основные успехи
- обобщающая способность,  
интерпретируемость
- направления применения в  
геонауках

## Практикум

- прогнозирование временного ряда
- подготовка данных
- статистическое исследование
- конструирование, настройка и  
обучение рекуррентной нейросети
- анализ эффективности

# Research workflow (practice)



# Deep learning

- + глубокое vs глубинное обучение
- + узнаем через несколько лет
- + русскоязычная терминология формируется медленно:
  - + малая аудитория
  - + материалы на английском
  - + среда на английском
- + Мой выбор: **глубокое обучение**

# Глубокое обучение

- методы машинного обучения, основанные изучении множества уровней представления / абстракции
- возможность определять сложные объекты на основании успешно изученных уровней представления / абстракции

(Yoshua Bengio, CIFAR)

# Предпосылки развития

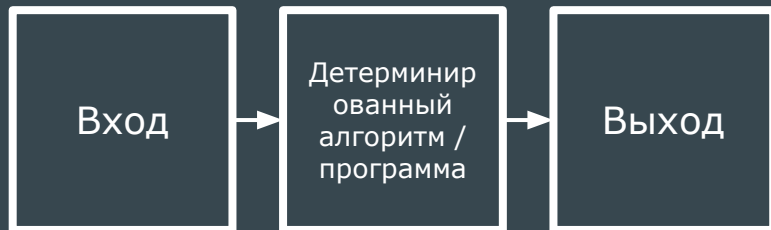
- + большие данные
- + суперкомпьютеры
- + деньги

# Основные игроки

- + Alphabet (Google Brain)
- + Amazon
- + Facebook
- + Baidu
- + Uber



Детерминированные алгоритмы / программы



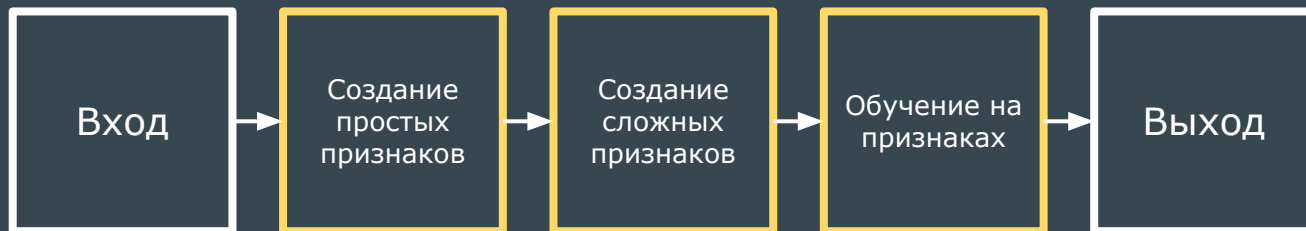
Классическое машинное обучение



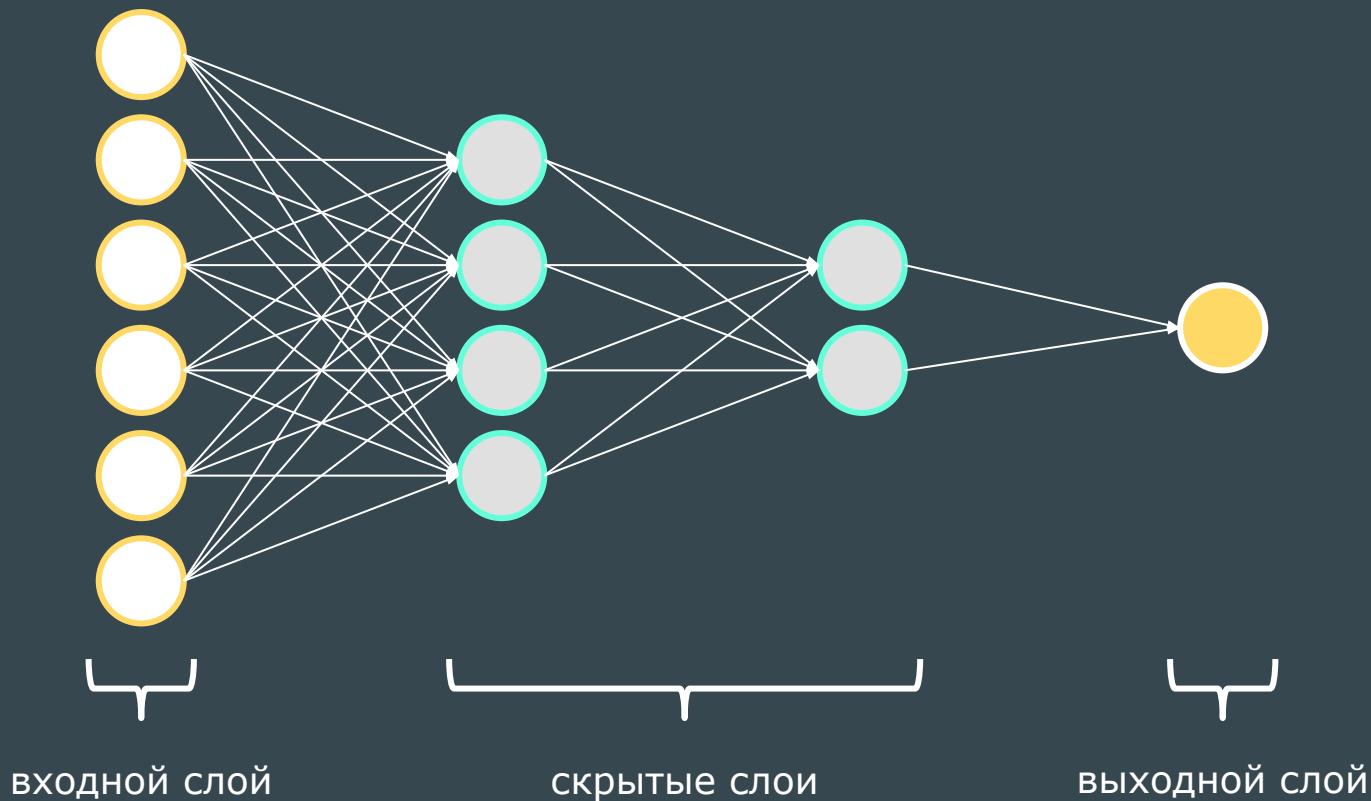
Признаковое обучение



Глубокое обучение

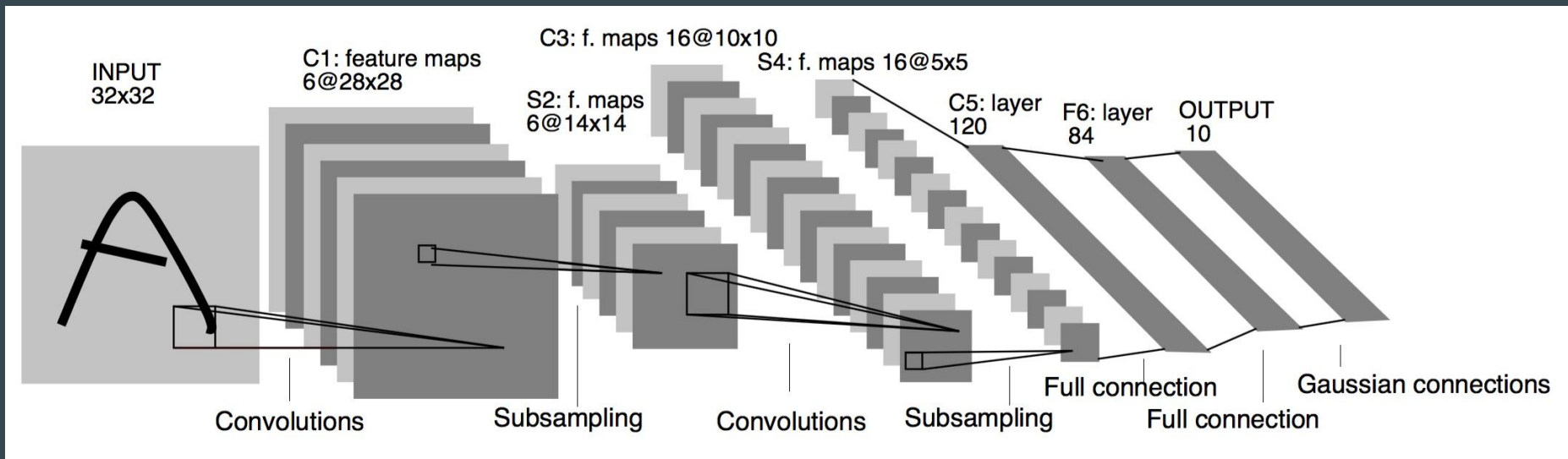


# Простая нейронная сеть



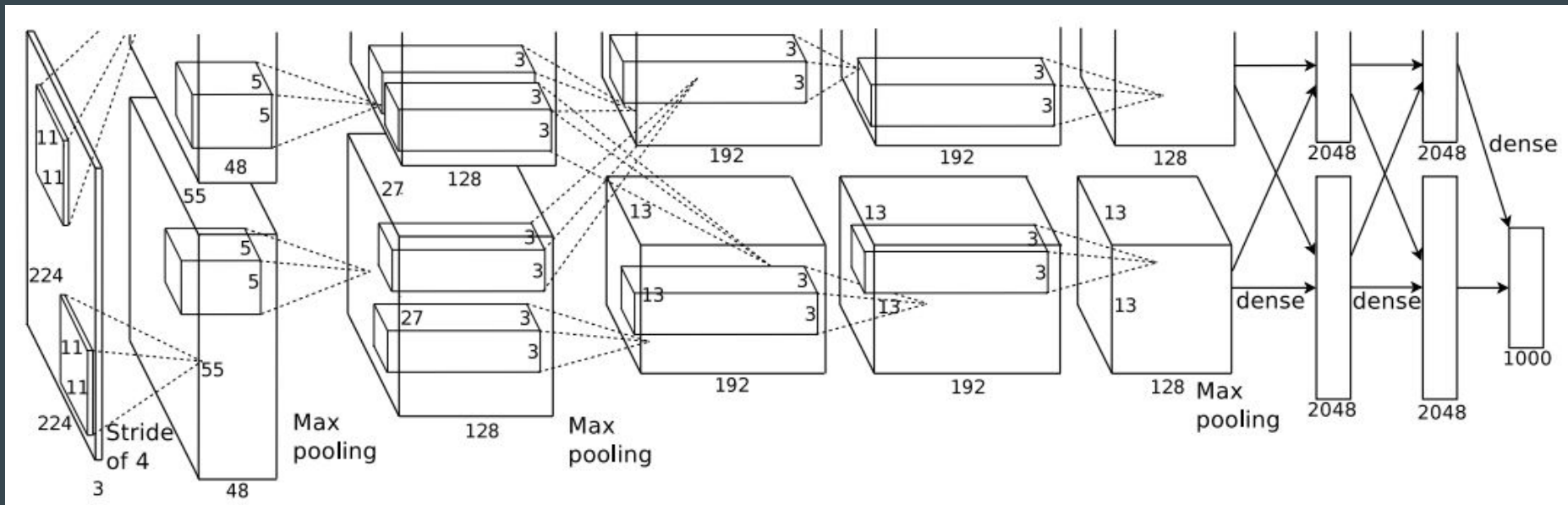


# LeNet, 1998



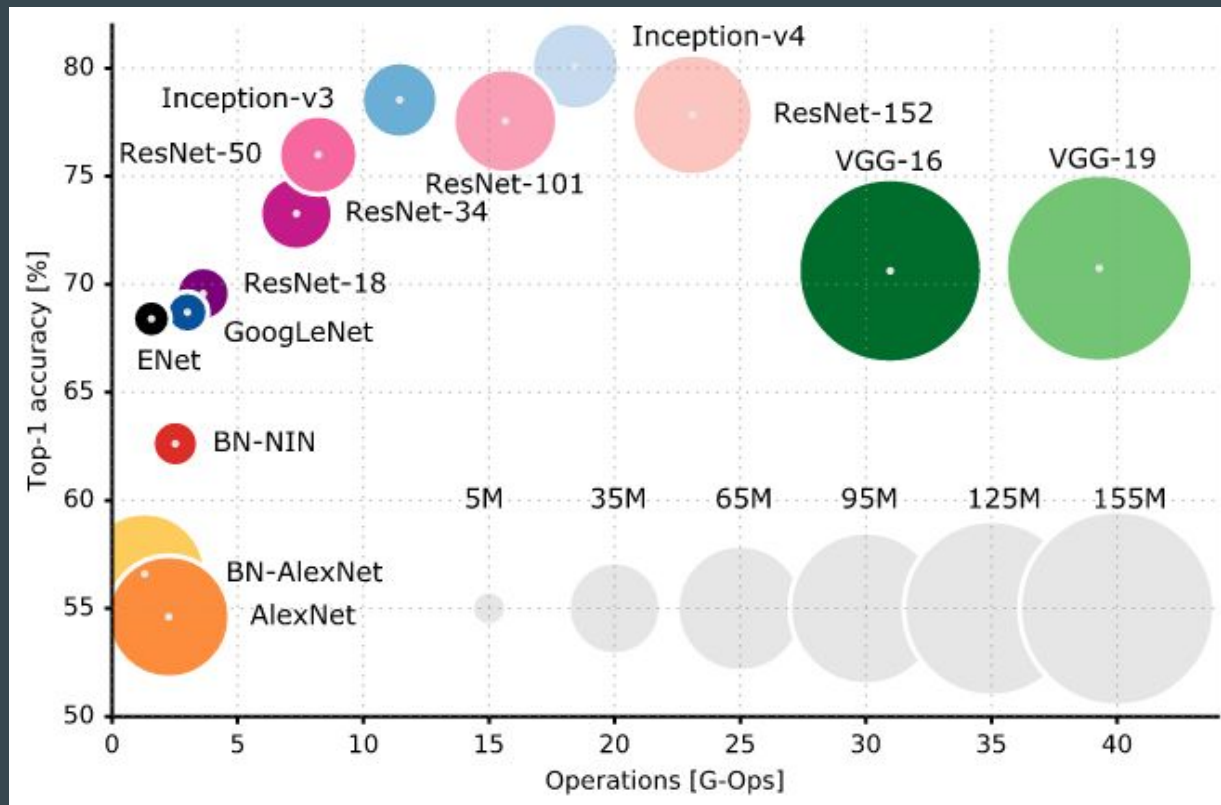
[статья](#)

# AlexNet, 2012



[статья](#)

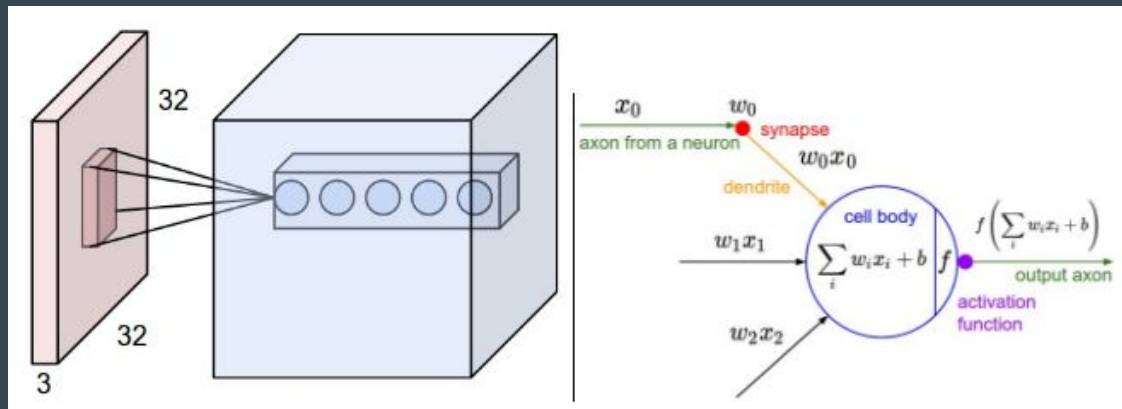
# Зоопарк глубоких архитектур



[статья](#)



# Сверточные нейронные сети



- + сильный алгоритм
- + распознавание образов
- + основа лучших решений

Input Volume (+pad 1) (7x7x3)	Filter W0 (3x3x3)	Filter W1 (3x3x3)	Output Volume (3x3x2)
$x[:::,0]$	$w0[:::,0]$	$w1[:::,0]$	$o[:::,0]$
0 0 0 0 0 0 0	1 0 1	-1 -1 1	7 0 2
0 0 1 2 0 2 0	-1 -1 1	-1 0 0	4 7 7
0 1 2 2 0 2 0	-1 -1 0	0 0 -1	3 5 2
0 0 0 0 1 0 0			
0 1 2 0 1 0 0	$w0[:::,1]$	$w1[:::,1]$	$o[:::,1]$
0 2 2 1 0 1 0	-1 0 0	-1 1 1	-2 -2 1
0 0 0 0 0 0 0	0 0 0	1 -1 -1	0 -8 -3
	1 1 -1	-1 1 0	4 -5 0
$x[:::,1]$	$w0[:::,2]$	$w1[:::,2]$	
0 0 0 0 0 0 0	1 0 -1	-1 1 -1	
0 2 0 2 1 1 0	1 0 0	0 0 0	
0 2 0 1 0 1 0	0 0 1	0 1 1	
0 1 0 2 0 1 0	Bus b0 (1x1x1)	Bus b1 (1x1x1)	
0 1 2 0 1 1 0	$b0[:::,0]$	$b1[:::,0]$	
0 0 1 1 2 2 0	1	0	
0 0 0 0 0 0 0			
$x[:::,2]$			
0 0 0 0 0 0 0			
0 0 0 0 2 2 0			
0 2 2 2 1 2 0			
0 1 0 1 2 0 0			
0 1 0 1 0 1 0			
0 0 2 2 2 2 0			
0 0 0 0 0 0 0			

toggle movement

Input Volume (+pad 1) (7x7x3)	Filter W0 (3x3x3)	Filter W1 (3x3x3)	Output Volume (3x3x2)
$x[:::,0]$	$w0[:::,0]$	$w1[:::,0]$	$o[:::,0]$
0 0 0 0 0 0 0	1 0 1	-1 -1 1	7 0 2
0 0 1 2 0 2 0	-1 -1 1	-1 0 0	4 7 7
0 1 2 2 0 2 0	-1 -1 0	0 0 -1	3 5 2
0 0 0 0 1 0 0			
0 1 2 0 1 0 0	$w0[:::,1]$	$w1[:::,1]$	$o[:::,1]$
0 2 2 1 0 1 0	-1 0 0	-1 1 1	-2 -2 1
0 0 0 0 0 0 0	0 0 0	1 -1 -1	0 -8 -3
	1 1 -1	-1 1 0	4 -5 0
$x[:::,1]$	$w0[:::,2]$	$w1[:::,2]$	
0 0 0 0 0 0 0	1 0 -1	-1 1 -1	
0 2 0 2 1 1 0	1 0 0	0 0 0	
0 2 0 1 0 1 0	0 0 1	0 1 1	
0 1 0 2 0 1 0	Bus b0 (1x1x1)	Bus b1 (1x1x1)	
0 1 2 0 1 1 0	$b0[:::,0]$	$b1[:::,0]$	
0 0 1 1 2 2 0	1	0	
0 0 0 0 0 0 0			
$x[:::,2]$			
0 0 0 0 0 0 0			
0 0 0 0 2 2 0			
0 2 2 2 1 2 0			
0 1 0 1 2 0 0			
0 1 0 1 0 1 0			
0 0 2 2 2 2 0			
0 0 0 0 0 0 0			

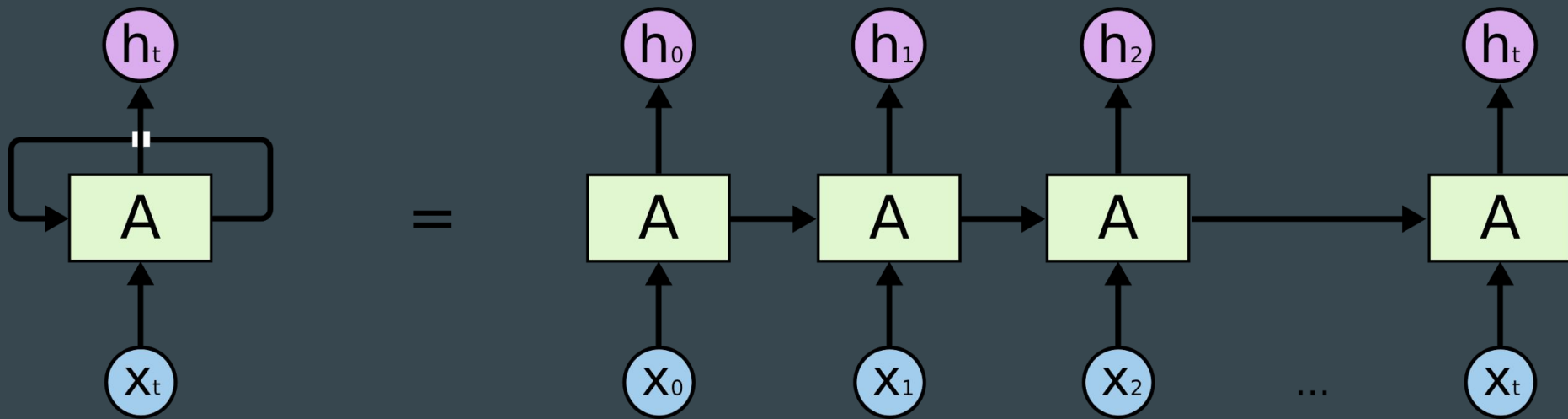
toggle movement

Input Volume (+pad 1) (7x7x3)	Filter W0 (3x3x3)	Filter W1 (3x3x3)	Output Volume (3x3x2)
$x[:::,0]$	$w0[:::,0]$	$w1[:::,0]$	$o[:::,0]$
0 0 0 0 0 0 0	1 0 1	-1 -1 1	7 0 2
0 0 1 2 0 2 0	-1 -1 1	-1 0 0	4 7 7
0 1 2 2 0 2 0	-1 -1 0	0 0 -1	3 5 2
0 0 0 0 1 0 0			
0 1 2 0 1 0 0	$w0[:::,1]$	$w1[:::,1]$	$o[:::,1]$
0 2 2 1 0 1 0	-1 0 0	-1 1 1	-2 -2 1
0 0 0 0 0 0 0	0 0 0	1 -1 -1	0 -8 -3
	1 1 -1	-1 1 0	4 -5 0
$x[:::,1]$	$w0[:::,2]$	$w1[:::,2]$	
0 0 0 0 0 0 0	1 0 -1	-1 1 -1	
0 2 0 2 1 1 0	1 0 0	0 0 0	
0 2 0 1 0 1 0	0 0 1	0 1 1	
0 1 0 2 0 1 0	Bus b0 (1x1x1)	Bus b1 (1x1x1)	
0 1 2 0 1 1 0	$b0[:::,0]$	$b1[:::,0]$	
0 0 1 1 2 2 0	1	0	
0 0 0 0 0 0 0			
$x[:::,2]$			
0 0 0 0 0 0 0			
0 0 0 0 2 2 0			
0 2 2 2 1 2 0			
0 1 0 1 2 0 0			
0 1 0 1 0 1 0			
0 0 2 2 2 2 0			
0 0 0 0 0 0 0			

toggle movement

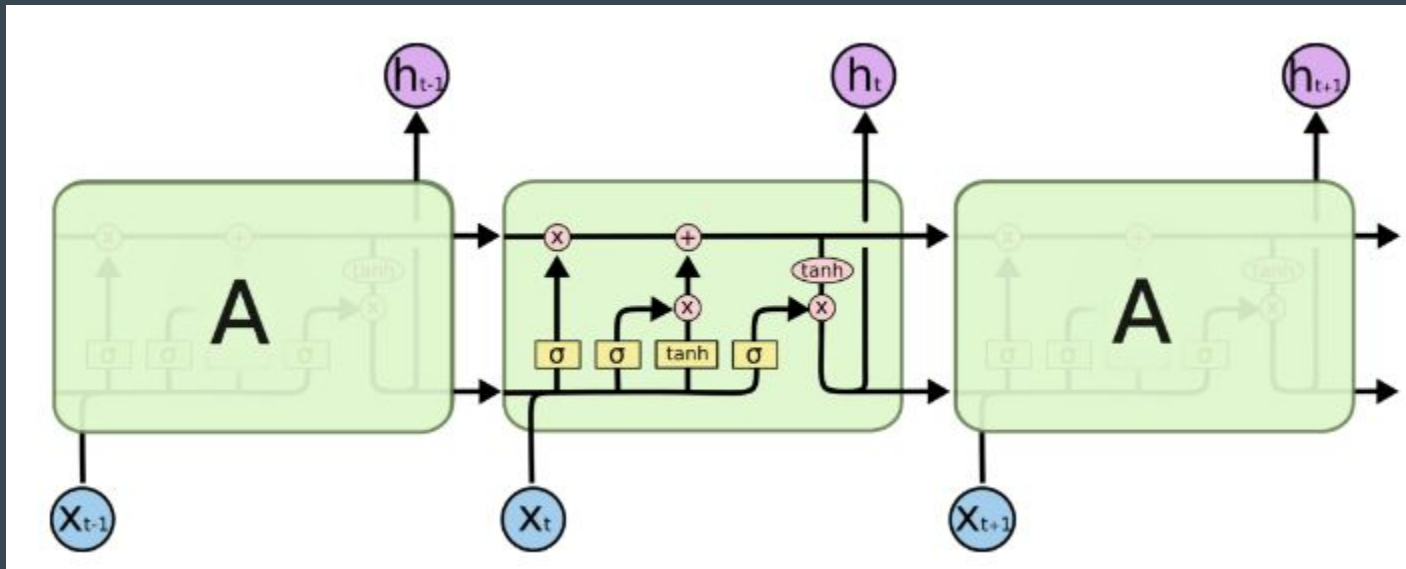


# Рекуррентные нейронные сети



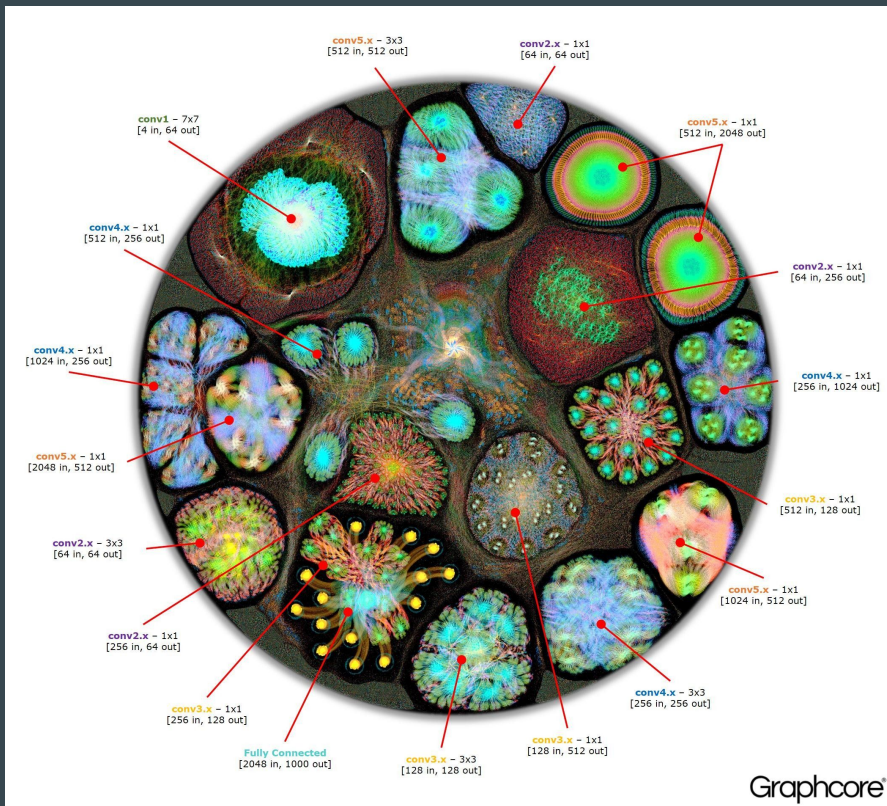
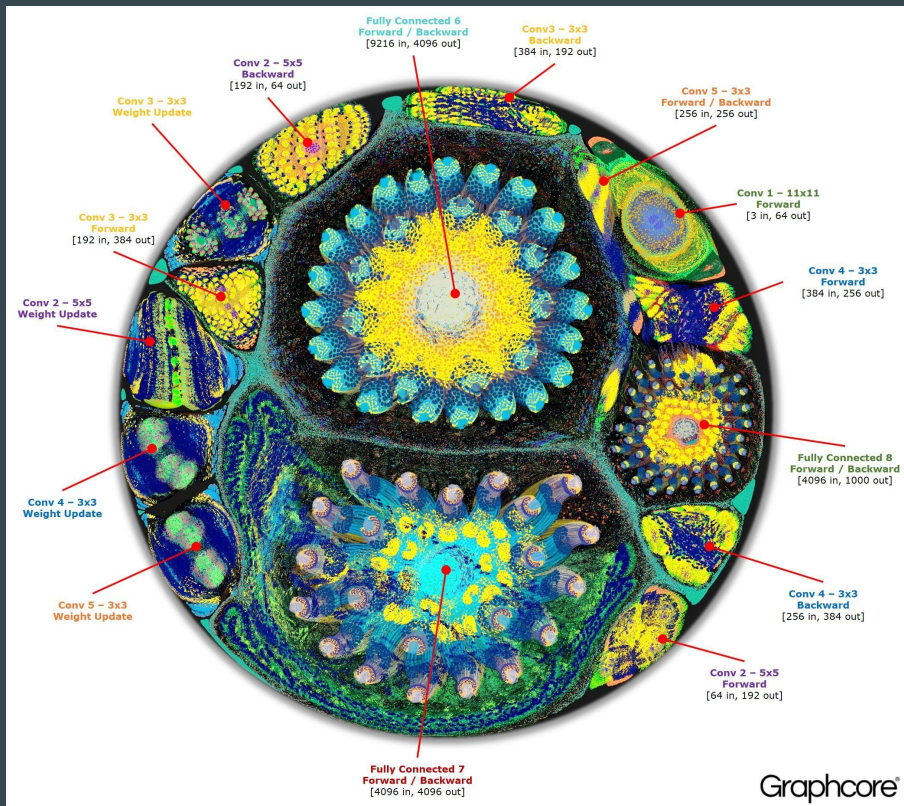
- + концепция ассоциативной памяти
- + моделирование / прогнозирование временных последовательностей
- + машинный перевод, генерация звука / текста

# LSTM сети (Long Short-Term Memory)



- + не теряют информацию
- + фильтрация шума и слабого контекста

# Интерпретируемость (AlexNet / ResNet)



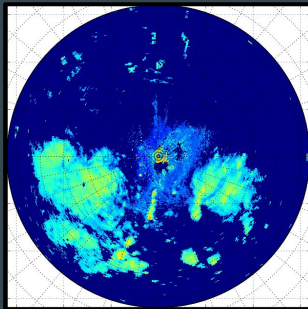
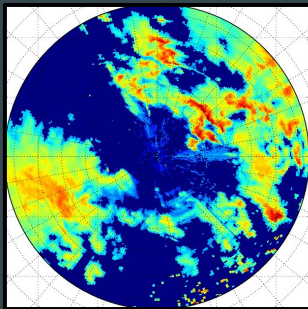
# Успехи глубокого обучения

- + распознавание образов
- + машинный перевод
- + генерирование сложных последовательностей (музыка / тексты)
- + автопилот
- + детекция редких сигналов (ЦЕРН)

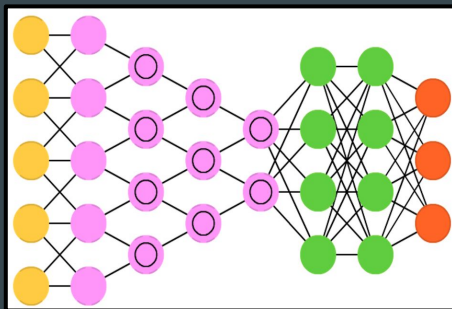
# Перспективные области в геонауках

## Наукастинг полей осадков

Данные: радарные снимки

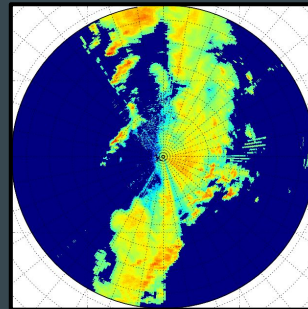


Модель: сверточные нейросети



asimovinsitute.org/neural-network-zoo/

Прогноз: 3 ч.

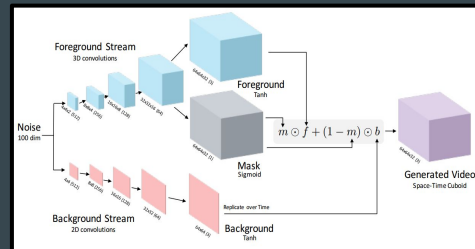


wradlib.org/wradlib-docs/

- + Большие данные
- + Современный подход
- + Неинтерпретируемость

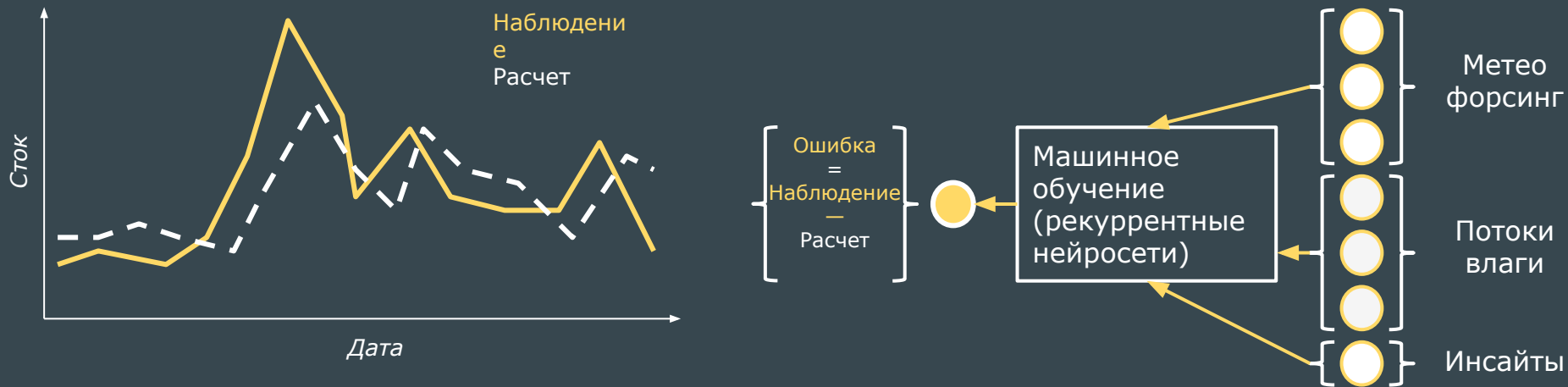
Удачный пример из области обработки изображений:

Vondrick et al. *Generating Videos with Scene Dynamics*. NIPS 2016.  
<http://web.mit.edu/vondrick/tiny-video/paper.pdf>



# Перспективные области в геонауках

## Постпроцессинг расчетов речного стока



- + Сопряжение геомodelей со статистическими
- + Выявление структурных неопределенностей
- + Онлайн-обучение, обучение с подкреплением

- + Мультидисциплинарность
- + Значимость для общества



Не хочу теорию, хочу крутой фреймворк  
и фыр-фыр-фыр



# TensorFlow

- + библиотека машинного обучения от Google
- + открытый исходный код
- + TF 1.0 (16 февраля)
- + огромное сообщество
- + python
- + GPU

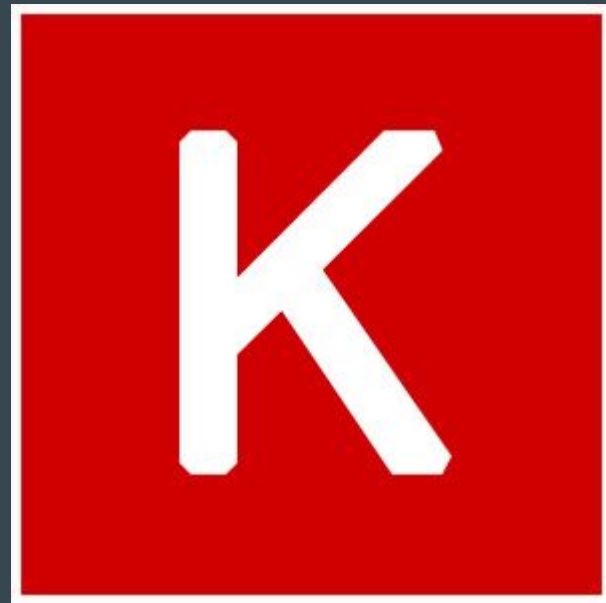


<https://www.tensorflow.org/>



# Keras

- + библиотека машинного обучения
- + бэкенд Theano, TensorFlow
- + простая структура API
- + CPU, GPU



<https://keras.io/>

# Подведем итоги

- + глубокое обучение -- область, которая будет на волне технологического развития ближайшие годы
- + вопрос применения глубоких архитектур нейросетей для решения задач в области наук о Земле неисследован
- + без детального знания теории невозможно добиться великолепного результата
- + зная python и базовые принципы машинного обучения можно получить хорошие результаты за короткий срок

# Материалы по теме:

1. Yoshua Bengio, [Deep learning theoretical motivations](#)
2. Eugenio Culurciello, [Neural Network Architectures](#)
3. Matthew Rubashkin, [Getting Started with Deep Learning](#)
4. Andrej Karpathy, [Convolutional Neural Networks](#)
5. Andrej Karpathy, [The Unreasonable Effectiveness of Recurrent Neural Networks](#)
6. N+1, Валентин Малых, [Азбука ИИ: «Рекуррентные нейросети»](#)

# Чем занять себя две недели?

Подготовка к #OpenDataHack от ECMWF

<https://ecmwf-opendatahack.devpost.com/>

- + "GET OUT"
- + "GET CREATIVE"
- + "GET GEEKY"
- + "GET FUN"



## 6 марта

Итоги участия команды

ИБП РАН в хакатоне

#OpenDataHack от ECMWF

## 20 марта

Официальная дата зачета  
(для аспирантов)

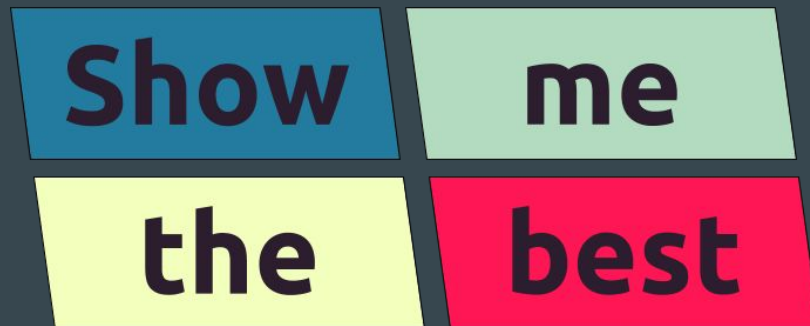
# На правах рекламы

Plovcast



новости науки глазами молодых  
ученых ИВП и ИО РАН

Show me the best



все яркие события из мира машинного  
обучения, python и науки

# Важно

Вы можете помочь существенно улучшить этот курс!

- [ayzelgv@gmail.com](mailto:ayzelgv@gmail.com), [hydrogo@yandex.ru](mailto:hydrogo@yandex.ru)
- [vk.com/ayzelgv](https://vk.com/ayzelgv), [facebook.com/ayzelgv](https://facebook.com/ayzelgv)
- ИВП РАН, кабинет 617