

*Декабрьские чтения  
в Томске*



# **Современное состояние и перспективы развития методов компьютерного зрения и глубокого обучения**

Ю.В. Визильтер, д.ф.-м.н., проф. РАН, [viz@gosniias.ru](mailto:viz@gosniias.ru)

*ФГУП «Государственный научно-исследовательский  
институт авиационных систем» (ГосНИИАС), г. Москва*

**Томск, 13.12.2018**

# Содержание

- **Современное состояние и перспективы развития методов компьютерного зрения и глубокого обучения**  
*(две волны современной технологической революции «ИИ»)*
- **Структурно-функциональный анализ и синтез глубоких нейронных сетей**  
*(некоторые математические аспекты происходящей технологической революции)*

# Классический «искусственный интеллект» (до 2000 г.)

Функциональный «ИИ» = АО/ПО, способные автоматически выполнять полезные функции, которые ранее могли быть выполнены только человеком.

ИИ-1: моделирование  
человеческих рассуждений

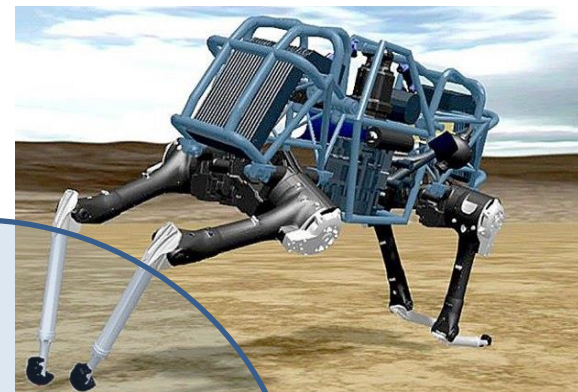
**Формальные теории**  
Доказательство теорем  
**Логическое**  
**программирование**  
**Представление знаний:**  
фреймы,  
семантические сети...  
Символьные  
преобразования  
выражений  
**Экспертные системы**  
Нечеткие логики...

ИИ-2: анализ данных  
и машинное обучение

**Обучение с учителем**  
Пространство  
признаков. Линейные  
разделители.  
Байесовское обучение.  
Ошибки 1 и 2 рода.  
**Обучение без учителя**  
Кластерный анализ.  
Снижение размерности  
**Нейронные сети**  
Обратное  
распространение...

*Насколько пригодны эти технологии к практическому внедрению?*

# Алгоритмическое обеспечение, необходимое для автономных интеллектуальных систем



**Навигация**

**Обработка  
сенсорных  
данных**  
(зрение,...)

**Управление**  
(планирование  
оптимизация,  
игры,...)

**Машинное  
обучение  
(ИИ-II)**  
(анализ  
данных)

**Искусственный  
интеллект  
(ИИ-I)** (базы  
знаний, логика,  
рассуждения)



# До 2000: Алгоритмическое обеспечение, необходимое для автономных интеллектуальных систем

Нерешенные задачи: 3D реконструкция и визуальная навигация, Сегментация и понимание сцены, Обнаружение объектов и распознавание изображений

Математическое программирование + экспертные знания.  
Нет возможности обучения на примерах и опыте действий



**Обработка сенсорных данных**  
(зрение,...)

**Управление**  
(планирование, оптимизация, игры,...)

**Прогноз ИИ: 2040+**

Обучаемые классификаторы уступают человеку  
Нет возможности обучения на больших данных

Экспертные системы уступают человеку  
Нет возможности автоматического обучения систем, основанных на ИИ-I

**Современное состояние  
и перспективы развития  
методов компьютерного зрения  
и глубокого обучения**  
*(две волны современной  
технологической революции ИИ)*

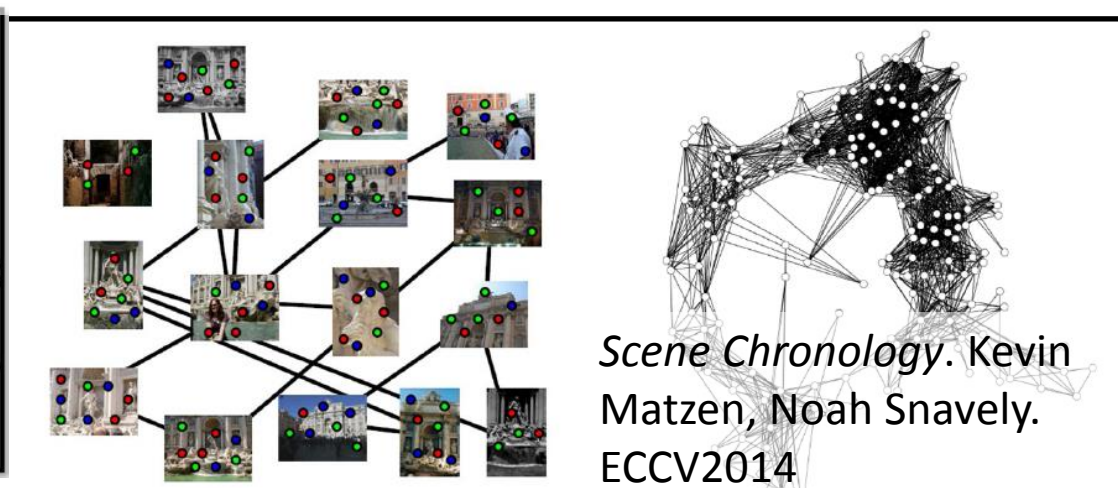
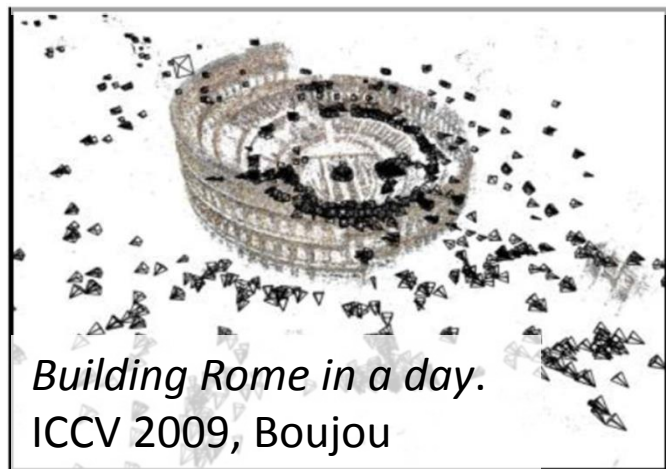
# Компьютерное зрение и глубокое обучение для мобильных технических систем

(2000-2016 гг. — первая волна технологической революции)

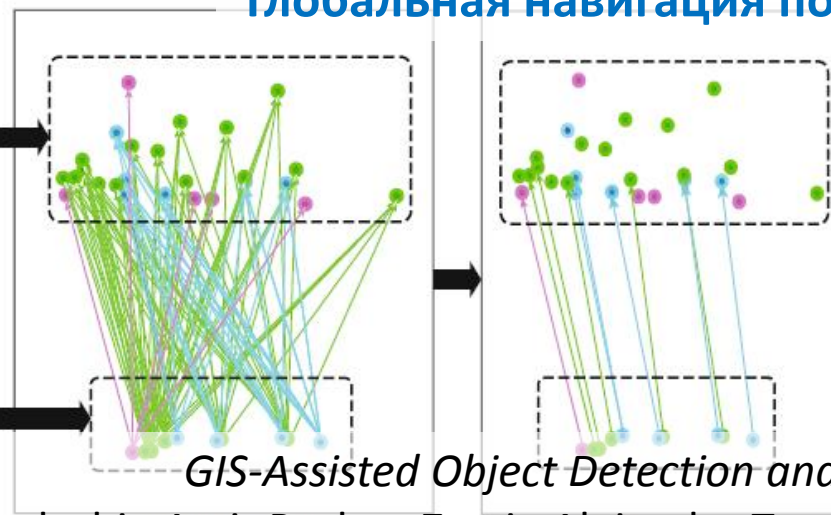
- **Реконструкция 3D сцены и навигация в ней:**  
Structure-from-Motion, SLAM, Road Scene  
Understanding and Autonomous Driving
- **Сегментация сцены и понимание видеосюжета:**  
Saliency maps, Video & 3D Segmentation, 3D-Flow,  
Multi-Tracking, Human Detection, Human pose  
estimation, Face Detection and Recognition
- **Распознавание изображений:**  
Convolution networks, Deep learning, Object  
Detection, Image Recognition

# Реконструкция 3D сцены и навигация в ней

- **Structure-from-Motion** (2000+) – технология реконструкции 3D сцены на основе множества разноракурсных снимков и оценки положения/параметров относительной ориентации снимков



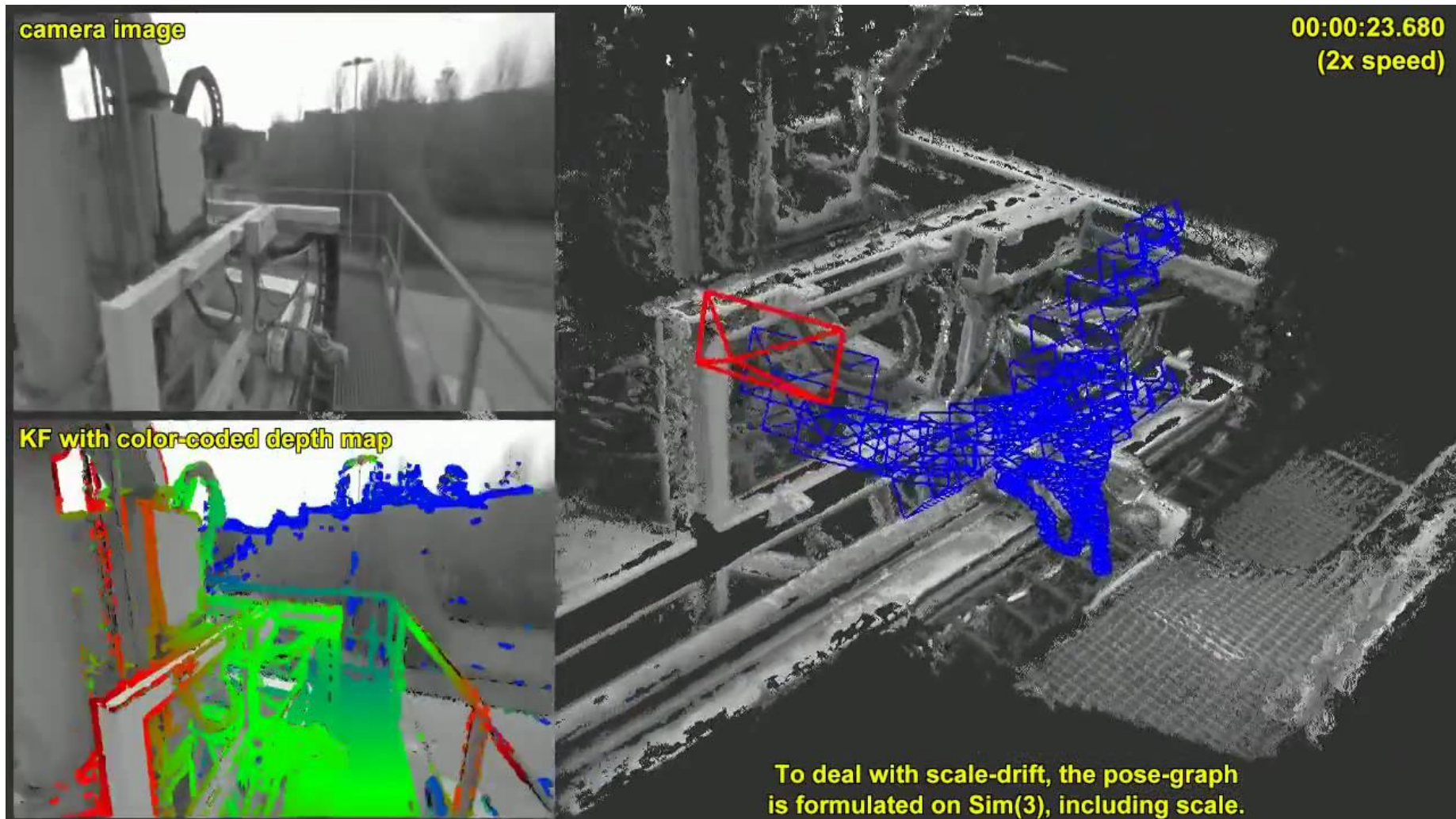
## Глобальная навигация по ориентирам (2014+)





# Реконструкция 3D сцены и навигация в ней

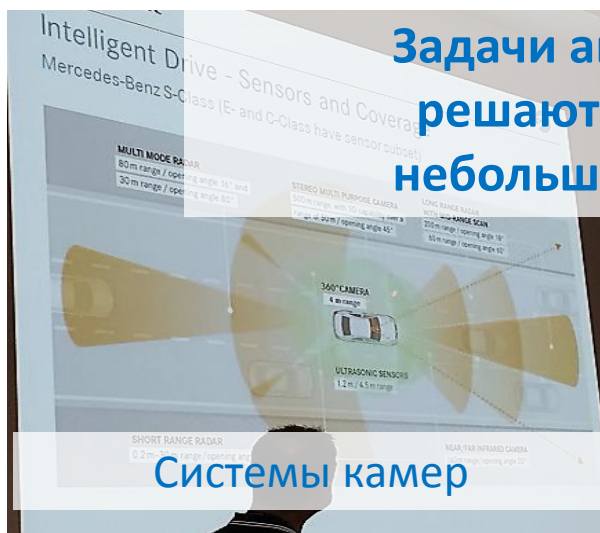
- SLAM (Simultaneous Localization and Mapping, 2006+, 2014+) – технология реконструкции 3D сцены и оценки положения/параметров движения камеры



# Автономное вождение, полет: SLAM, UAV, Fish-eye Cameras



Изображения от камер

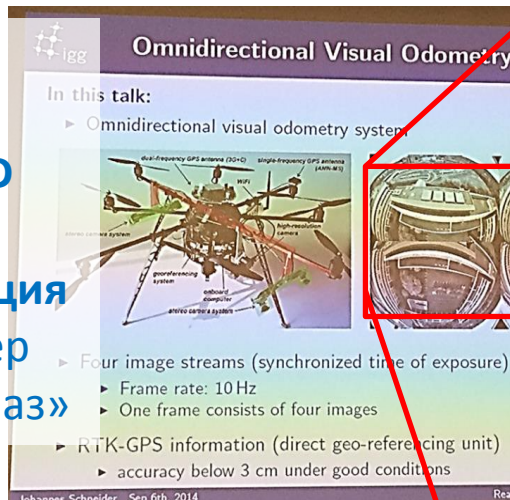


Системы камер

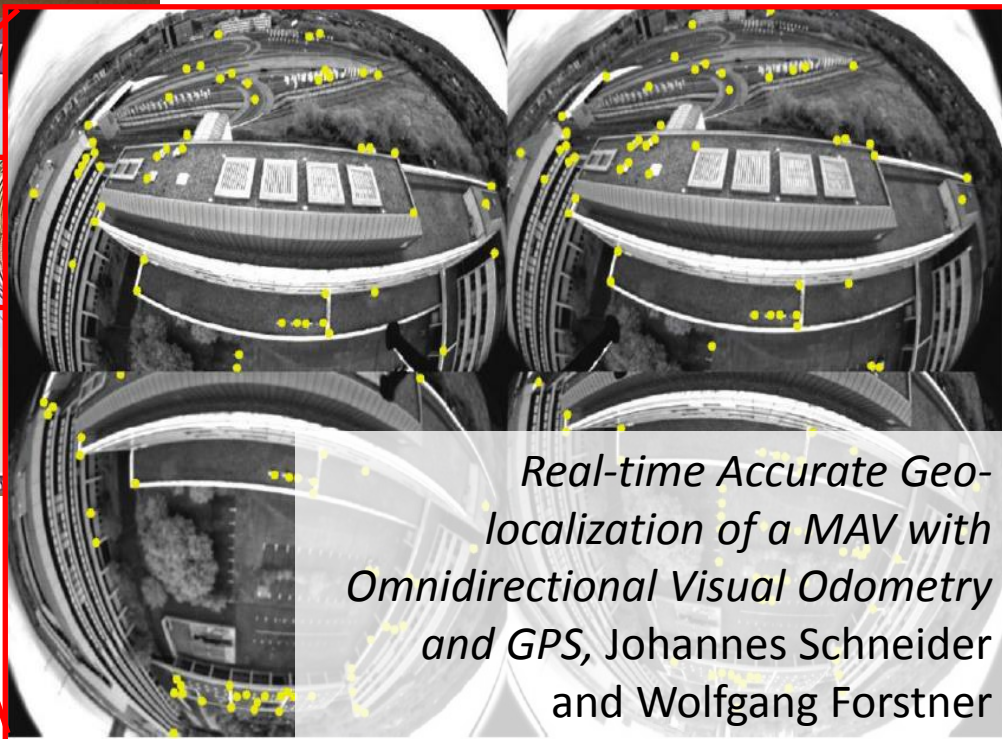
Задачи автономной навигации в 3D решаются при помощи множества небольших широкоугольных камер



2012-15 –  
высокоточные  
калибровка, 2D  
одометрия и  
3D реконструкция  
на основе камер  
типа «рыбий глаз»



*Multi-Camera Systems in the V-Charge Project: Fundamental Algorithms, Self Calibration, and Long-Term Localization, Paul Furgale, ECCV'14, W15*



*Real-time Accurate Geo-localization of a MAV with Omnidirectional Visual Odometry and GPS, Johannes Schneider and Wolfgang Forstner*



# Saliency Maps (Имитация зрительного внимания)



*A Closer Look at Context: From Coxels to the Contextual Emergence of Object Saliency* Rotem Mairon and Ohad Ben-Shahar, ECCV'14

*RGBD Salient Object Detection: A Benchmark and Algorithms,* Houwen Peng, Bing Li, Weihua Xiong, Weiming Hu, and Rongrong Ji

*Saliency in Crowd,* Ming Jiang, Juan Xu, and Qi Zhao

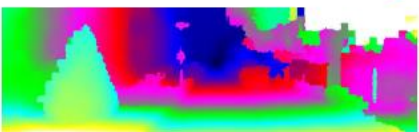
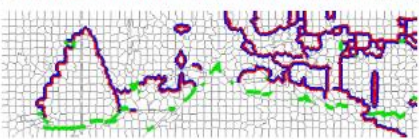
# Сегментация сцены и понимание видеосюжета

Единый подход к обработке и

сегментации данных (2003+)

2D, 2D+T, 3D, 3D+T:

MRF, Energy-based, 3D Flow



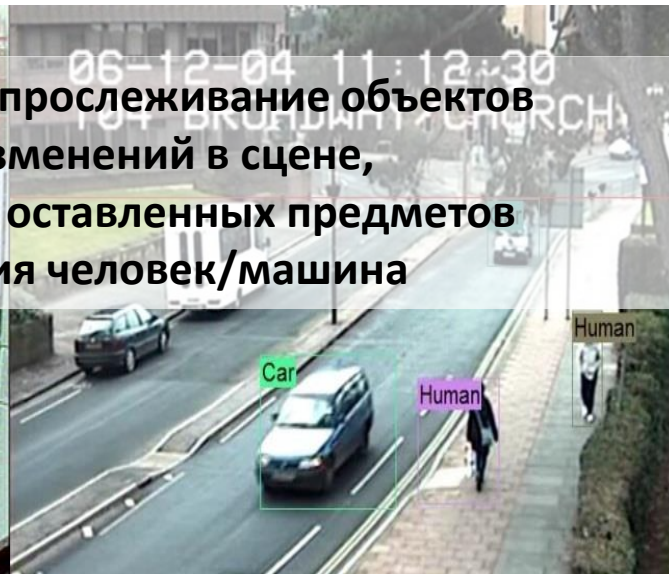
Минимизация  
целевой функции  
(энергии)

Модели на основе  
Марковских  
полей

Использование  
карт внимания

Быстрые методы анализа видеопоследовательностей  
(до 800 fps)

- Выделение и прослеживание объектов
- Выделение изменений в сцене, обнаружение оставленных предметов
- Классификация человек/машина



- Использование правил анализа динамической сцены для генерации событий и сообщений в системах видеонаблюдения

Надежное обнаружение людей  
на видео



Dollár, Wojek, Schiele, Perona, 2014

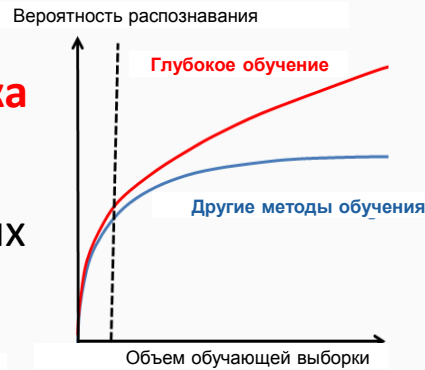
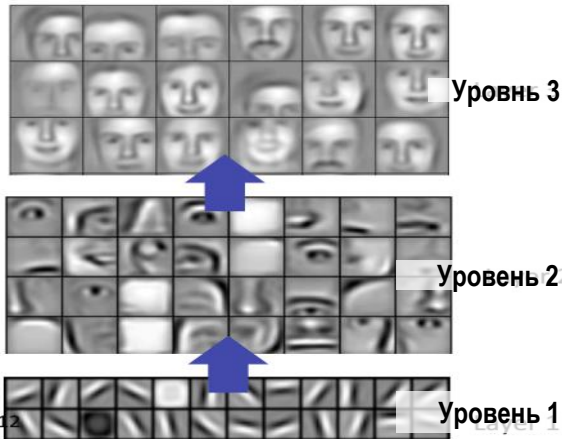


# Распознавание и обработка изображений

на базе глубоких конволюционных нейронных сетей (с 2011)

+ С 2011 г. - **распознавание образов на уровне человека или выше** (superhuman)

+ Обучение на сверхбольших объемах данных



+ Иерархическое обучение с повышением абстракции данных от уровня к уровню

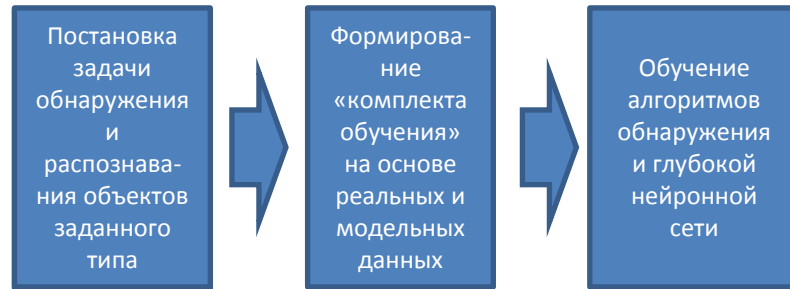
## Достоинства и проблемы

+ Тысячи слоев нейронов  
+ Учет специфики изображений как объекта распознавания (локальность, инвариантность к сдвигу, нечеткая локализация)

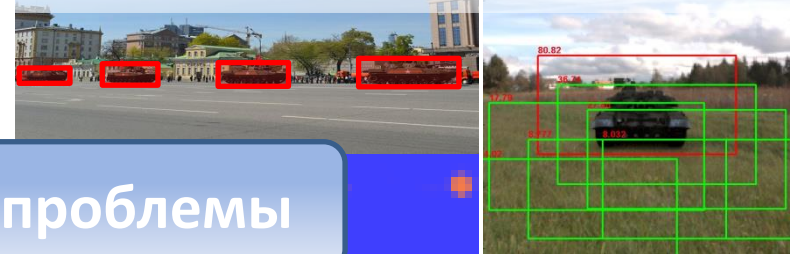


Типовая структура глубокой конволюционной сети

- Нужны огромные обучающие выборки
- Длительное моделирование и обучение



- Ресурсоемкость, низкая скорость
- Необходимо быстрое предобнаружение



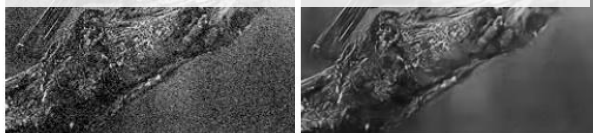
- Необходимость эффективных алгоритмических реализаций
- Необходимость создания нового поколения нейропроцессоров



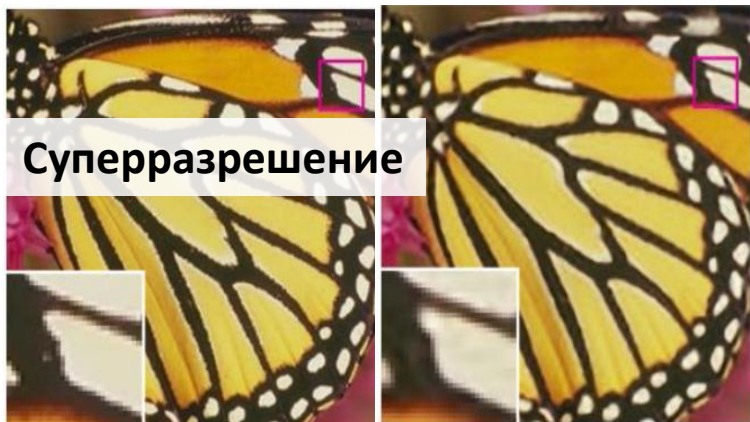
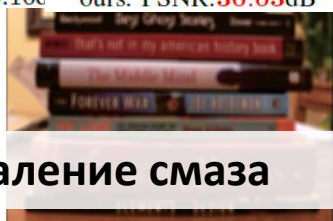
# CNN-based Image Restoration and Analysis



Фильтрация шумов



Удаление смаза



Суперразрешение

Ours

Original / PSNR

SRCNN / 27.95 dB



Обнаружение особых точек на лицах

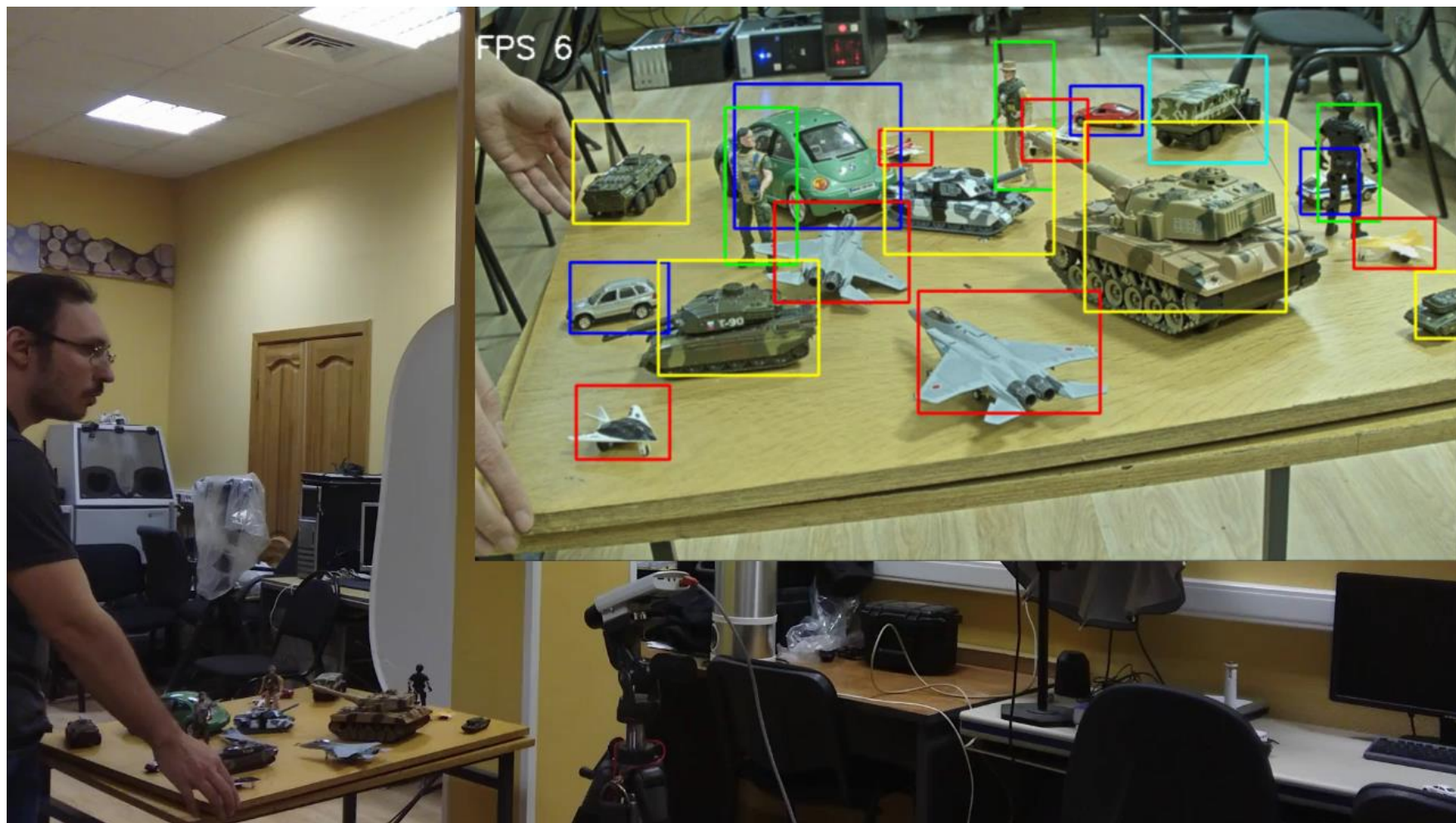


Фронтализация лиц без 3D моделей





# Обнаружение и распознавание объектов в реальном времени (2016: CNN)



Цветом указаны классы объектов: желтый – бронетехника, красный – самолеты, голубой – грузовики, синий – автомашины, зеленый - люди

# Обнаружение и распознавание объектов в реальном времени (2017: CNN)



Цветом указаны классы объектов: сиреневый – бронетехника, голубой – самолеты, красный – грузовики, зеленый – автомашины



# Автоматическая семантическая сегментация аэроснимков (ГосНИИАС-2015)

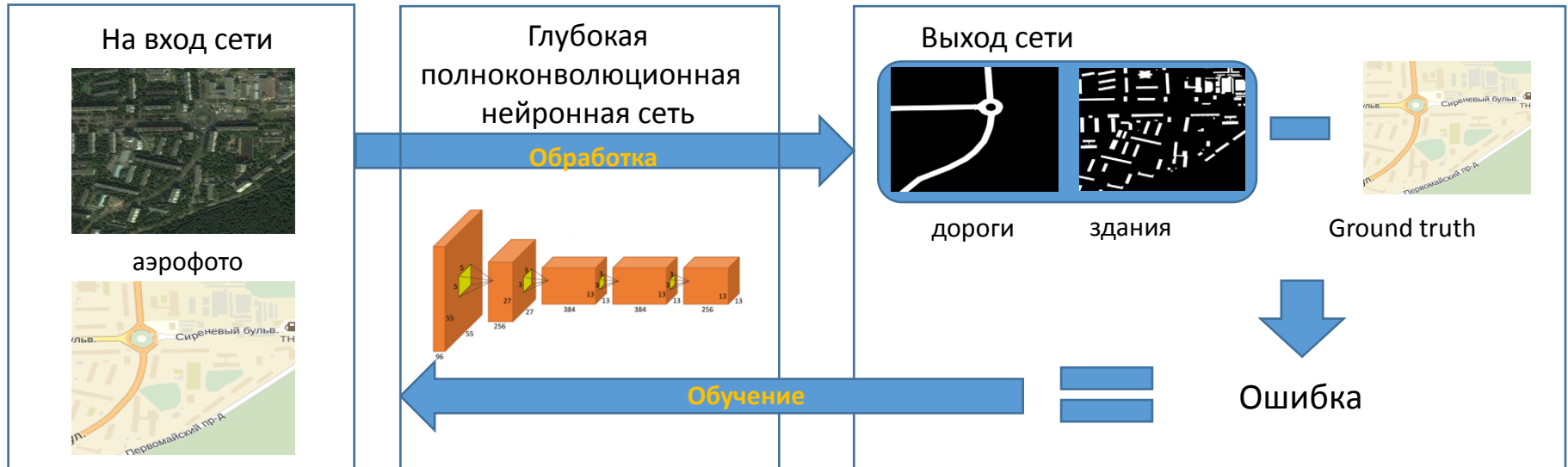
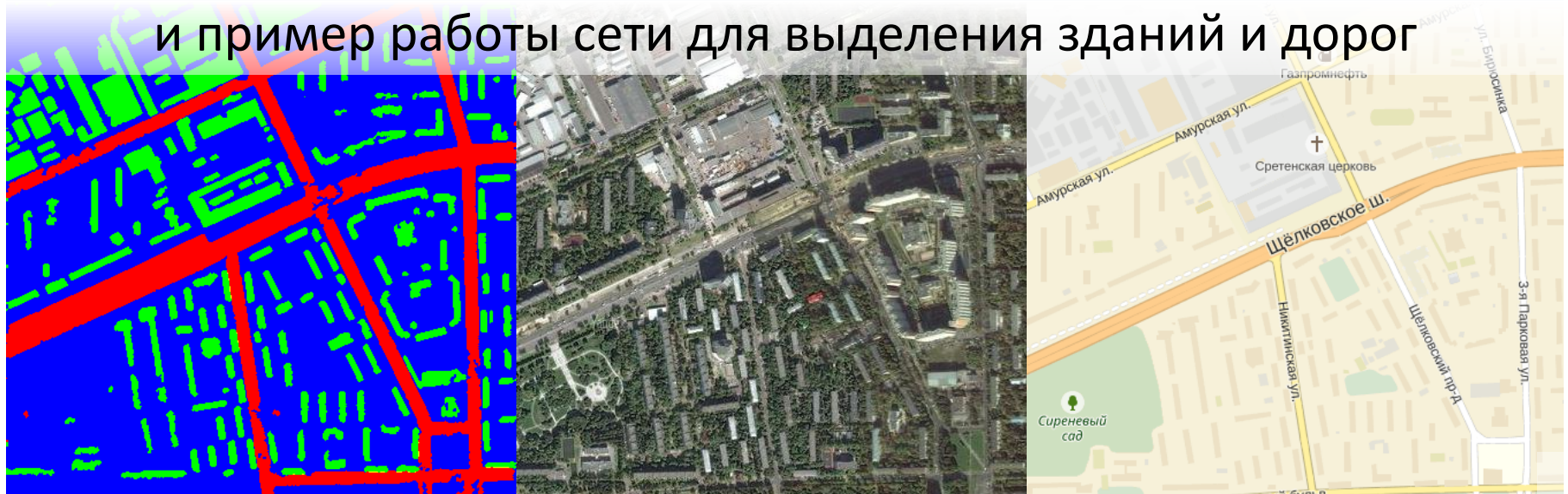
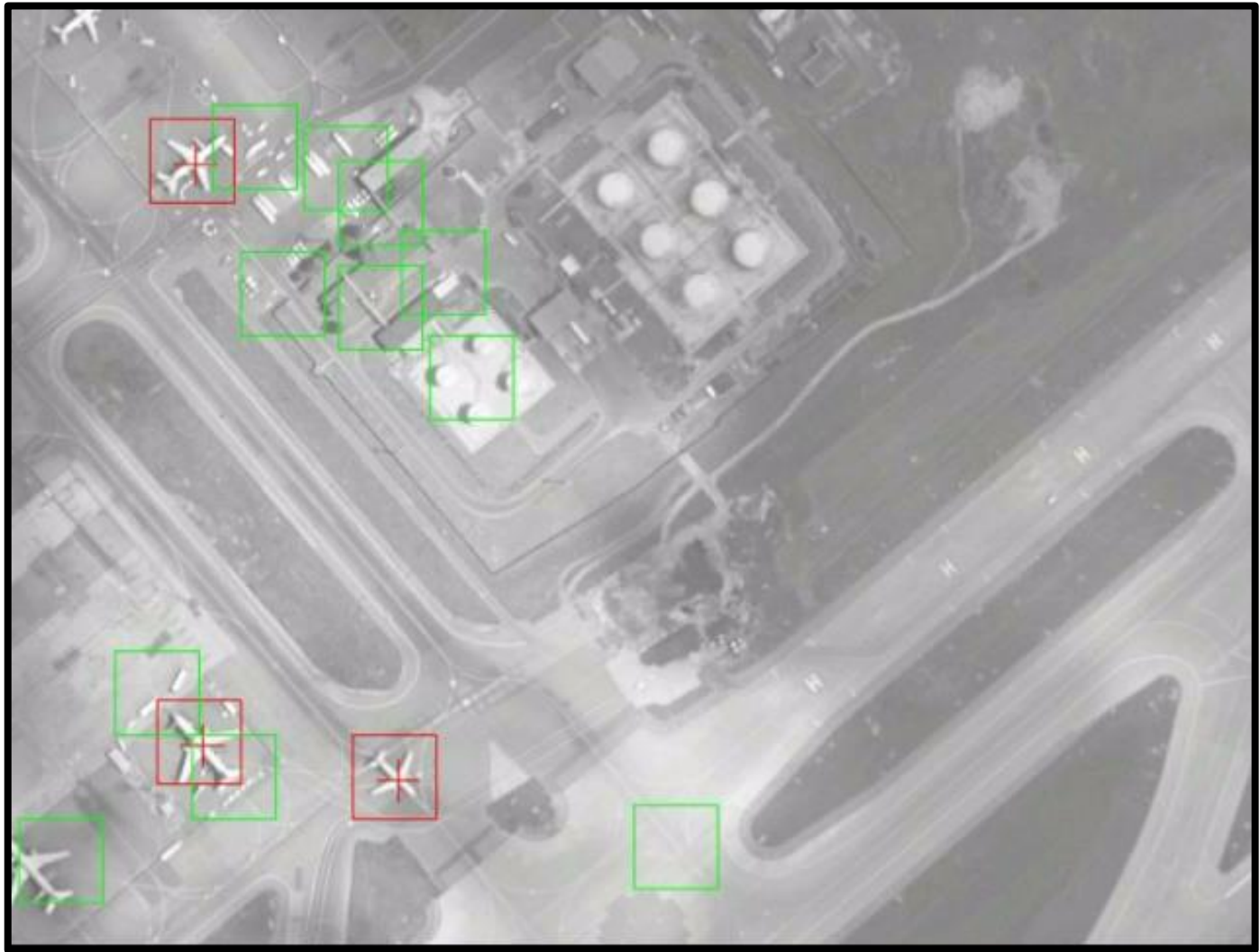


Схема обучения сети для семантической сегментации и пример работы сети для выделения зданий и дорог



# Автоматическое обнаружение самолетов на аэроснимках (ГосНИИАС-2016)





# Открытый конкурс на лучшее решение в области создания интеллектуальных технологий дешифрирования видовой аэрокосмической информации (2017)



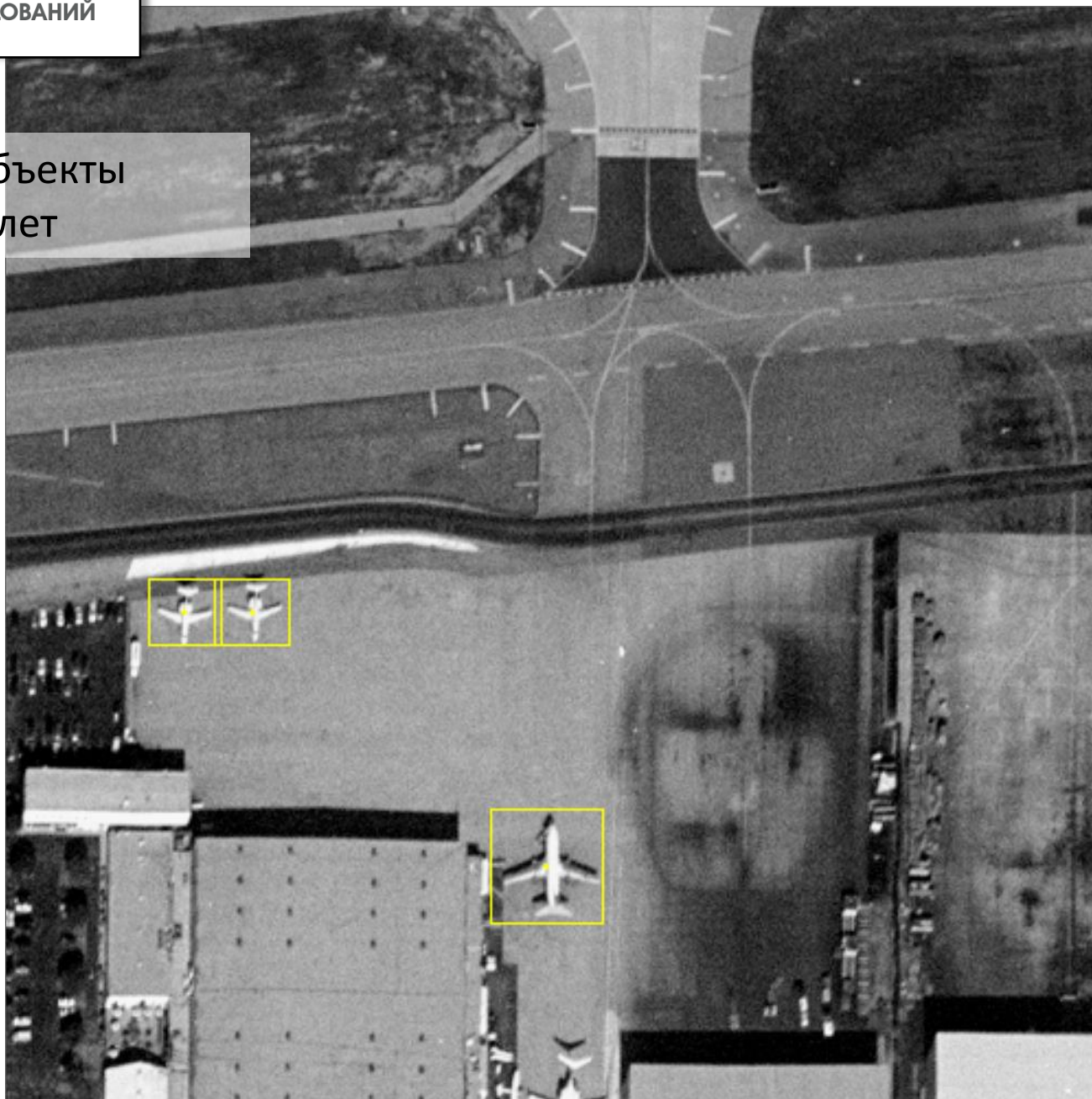
*ФГУП «ГосНИИАС» - призер конкурса*

Показаны объекты  
класса самолет





Показаны объекты  
класса самолет

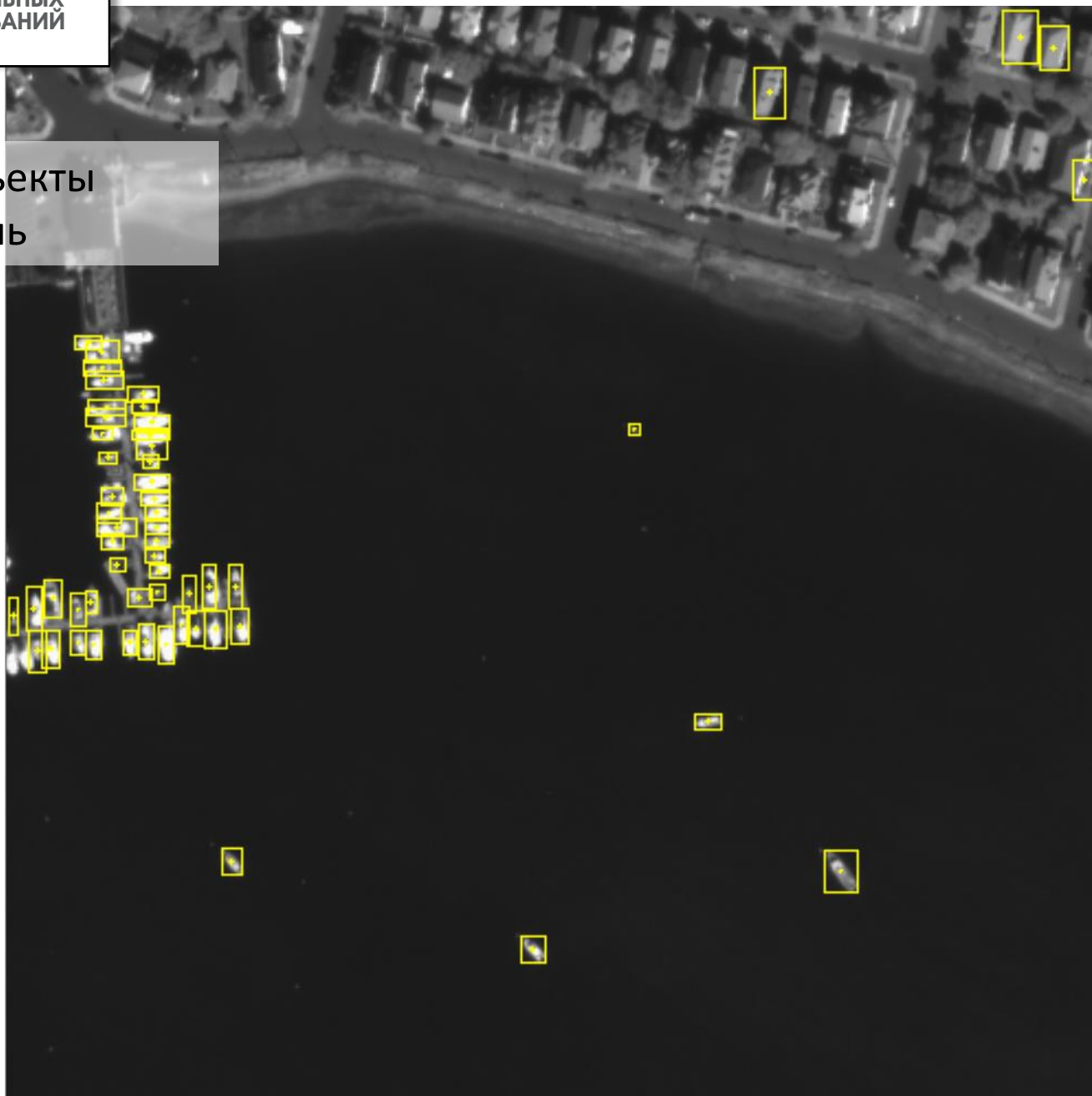


Показаны объекты  
класса самолет

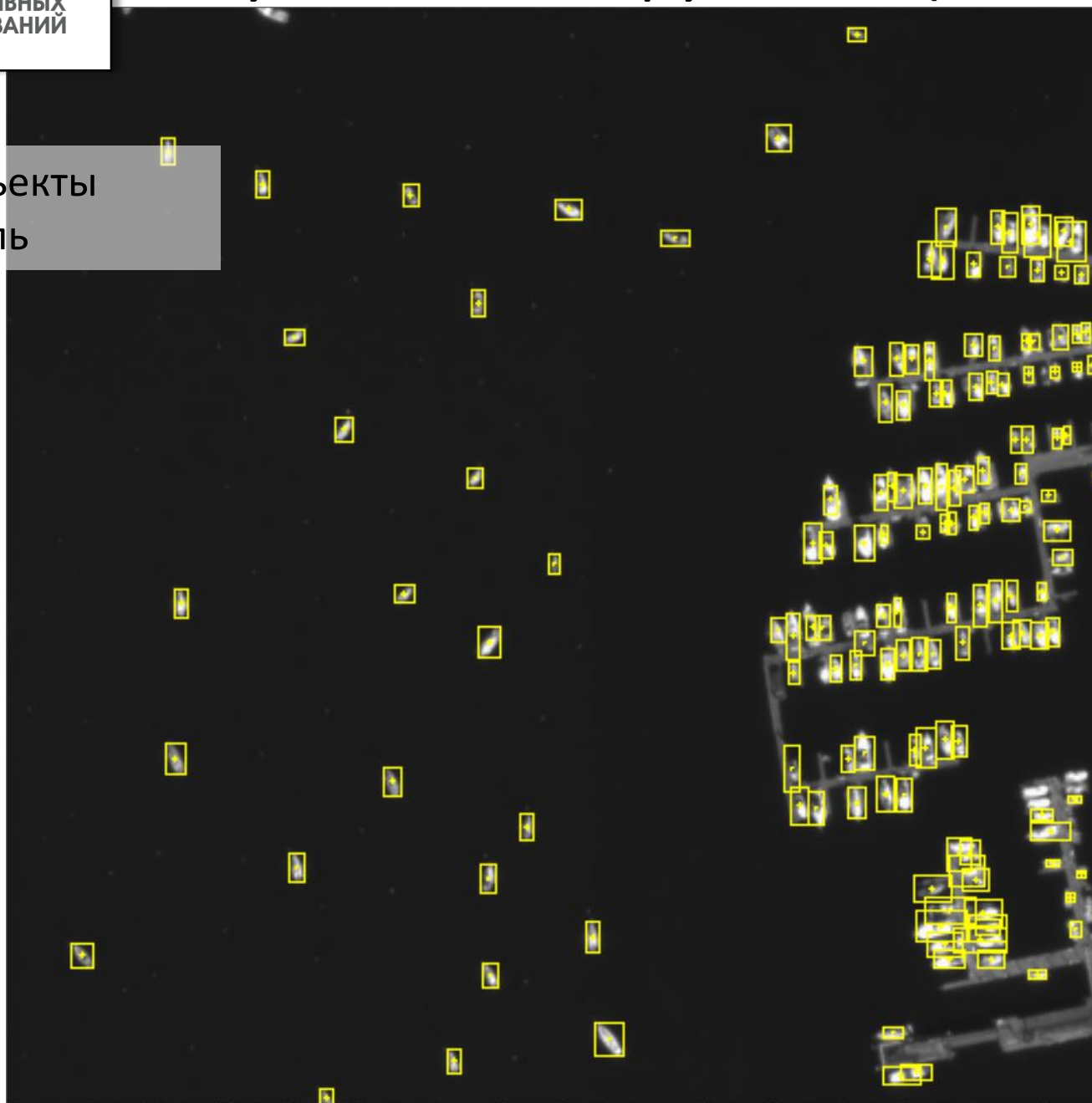




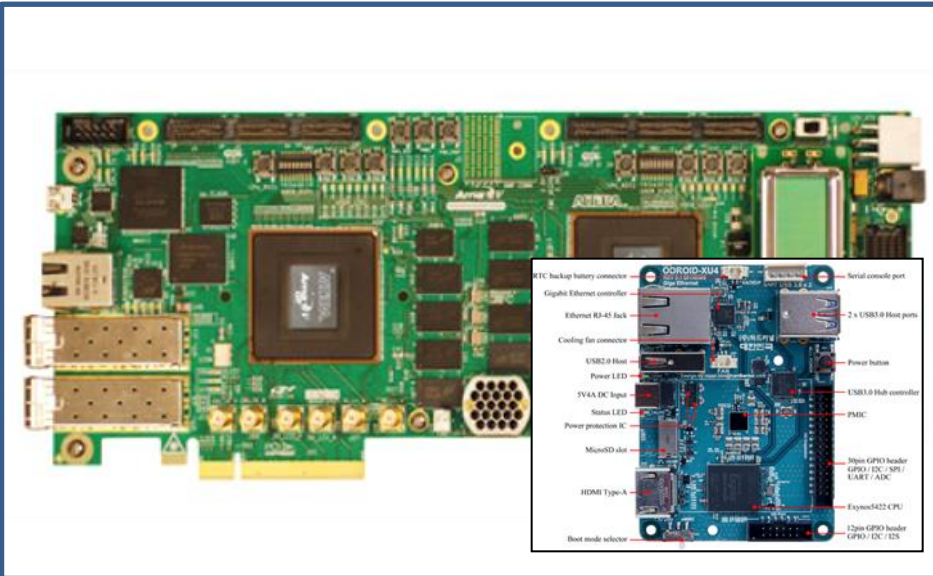
Показаны объекты  
класса корабль



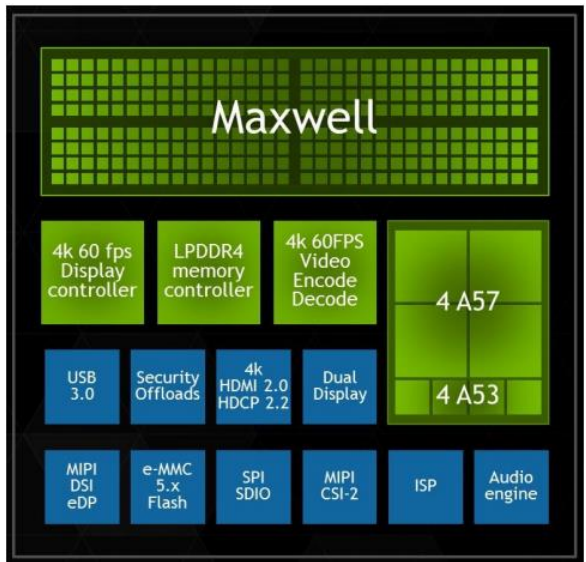
Показаны объекты  
класса корабль



# Аппаратно-программные решения



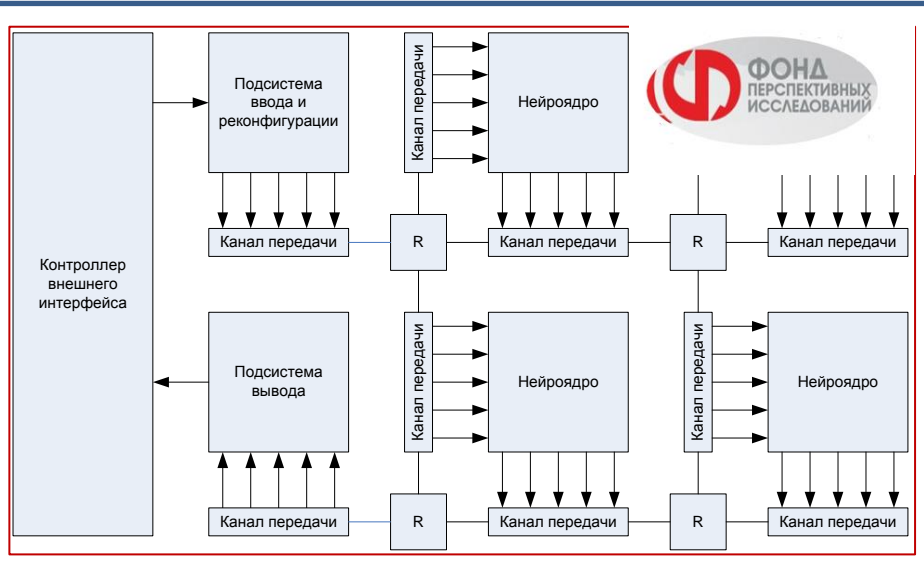
**ПЛИС (FPGA), ARM**



**NVIDIA GPGPU**



**ИТЦ Модуль, МЦСТ Эльбрус**

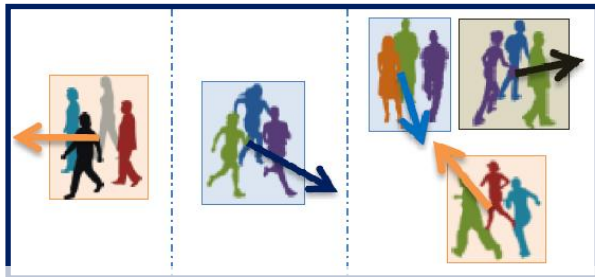


**ФПИ «Мемристор», МФТИ**

**Ограничения  
и нерешенные проблемы**  
*(в рамках первой волны  
технологической революции)*



# Основное функциональное ограничение классических CNN: невозможность построения и использования структурных моделей, пространственно-временных логик и онтологий для анализа сложных объектов и динамических сцен



Модель событий и процессов

Group walking, Group running, Group merging and Group splitting.

Без обучения получаются  
слишком сложные  
описания очевидных для  
человека ситуаций



Модель сцены

```
BB(?BBx), BB(?BBy), Frame(?F1), MBB(?MBB1), MBB(?MBB2), BB_Detected_In_Frame(?BBx, ?F1),  
BB_Detected_In_Frame(?BBy, ?F1), BB_Bottom_Left_Point_Y(?BBx, ?h), BB_Bottom_Right_Point_Y(?BBy,  
?d), BB_Number(?BBx, ?n4), BB_Number(?BBy, ?n5), BB_Top_Left_Point_X(?BBx, ?a),  
BB_Top_Left_Point_X(?BBy, ?f), BB_Top_Left_Point_Y(?BBx, ?e), BB_Top_Left_Point_Y(?BBy, ?i),  
BB_Top_Right_Point_X(?BBx, ?j), BB_Top_Right_Point_X(?BBy, ?b), BB_Top_Right_Point_Y(?BBy, ?c),  
MBB_ID(?MBB1, ?n1), MBB_ID(?MBB2, ?n1), Number_BB_In_Frame(?F1, 2), Number_Frame(?F1, ?n1),  
Number_MBB(?MBB1, ?n2), Number_MBB(?MBB2, ?n3), add(?x2, ?b, 20), greaterThan(?a, ?b),  
greaterThan(?h, ?d), greaterThan(?n3, ?n2), greaterThanOrEqual(?b, ?x1), greaterThanOrEqual(?e, ?c),  
lessThanOrEqual(?a, ?x2), lessThanOrEqual(?e, ?d), subtract(?x1, ?a, 20), subtract(?z1, ?j, ?f), subtract(?z2,  
?h, ?i) -> BB_Represent_MBB(?BBx, ?MBB1), BB_Represent_MBB(?BBy, ?MBB1),  
MBB_Detected_In_Frame(?MBB1, ?F1), MBB_H(?MBB1, ?z1), MBB_Top_Left_Point_X(?MBB1, ?f),  
MBB_Top_Left_Point_Y(?MBB1, ?i), MBB_W(?MBB1, ?z2)
```

*Events detection using a video-surveillance Ontology and a rule-based approach,*  
Yassine Kazi Tani, Adel Lablack, Abdelghani Ghomari, and Ioan Marius Bilasco, 2014

# Технологическая проблема: конструирование и обучение CNN – длительный ручной процесс с негарантированным результатом

Нет и не предвидится теории построения и обучения CNN, которая была бы способна ответить на следующие основные вопросы:

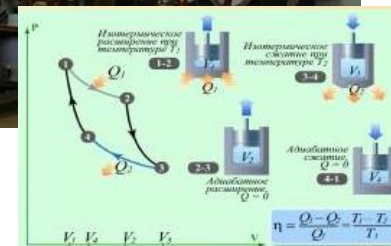
- Какова должна быть сложность CNN для решения задач определенного типа на данных определенной сложности при обучении на выборках определенного размера?
- Как оптимально формировать структуру глубокой сети для конкретных типов задач и конкретных типов данных?
- Как оптимально формировать стратегию обучения глубокой сети для конкретных типов задач и конкретных типов данных?
- От чего зависит скорость обучения глубокой сети, и как на нее влиять в процессе обучения?
- Как заранее оценить достижимые результаты некоторой заданной ГКС при обучении?
- Как предсказать или хотя бы семантически интерпретировать структурные описания, которые порождает глубокая сеть?



Обучение CNN скорее искусство, чем наука



В области CNN практика серьезно опережает теорию





# 2015: Алгоритмическое обеспечение, необходимое для автономных интеллектуальных систем

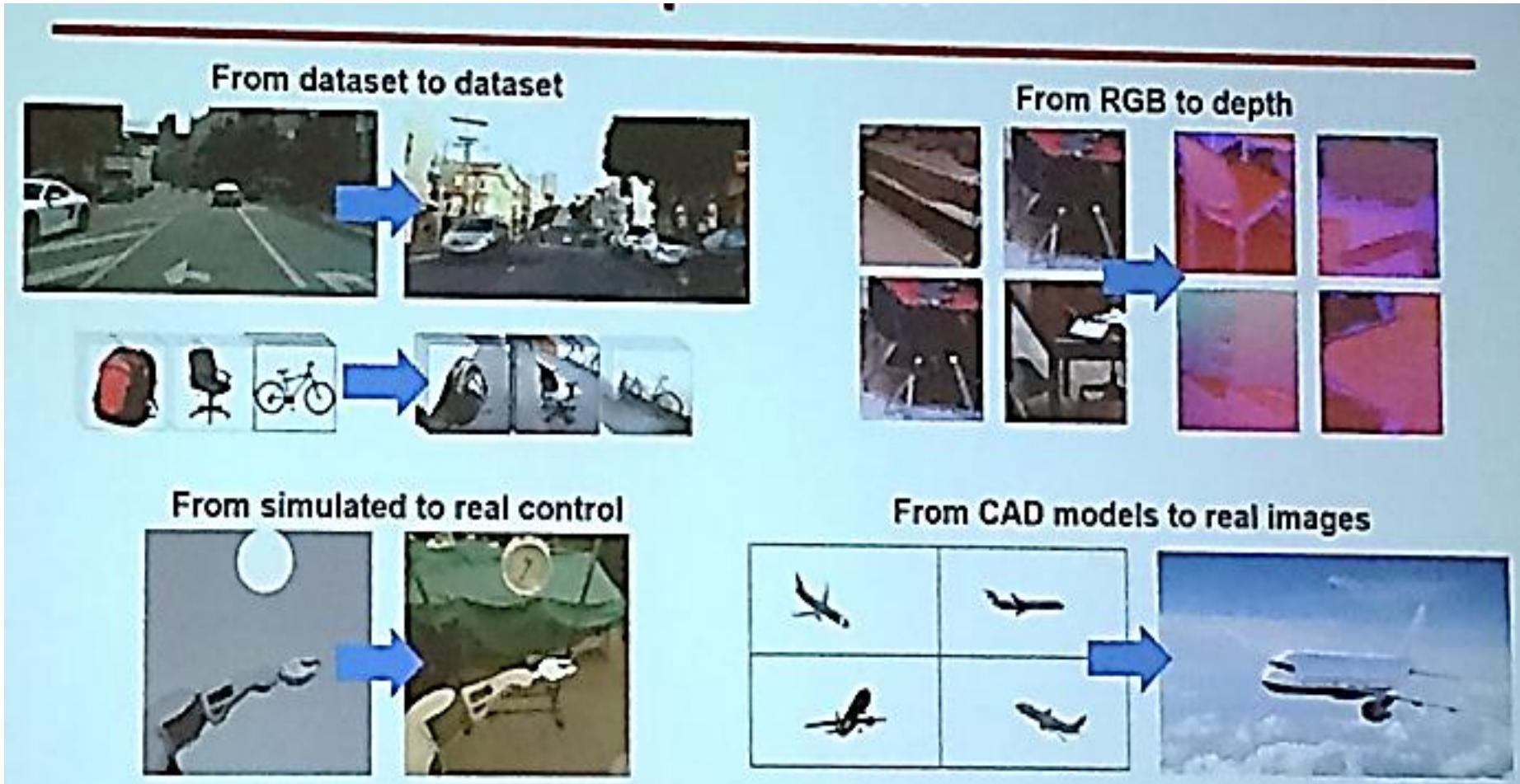


# Компьютерное зрение и глубокое обучение для интеллектуальных систем

**(2017+, вторая волна технологической революции)**

- **Глубокие соревнующиеся сети для имитации данных**  
GAN, Domain Transfer Learning, Zero-Shot Learning
- **Интерпретация динамической визуальной информации на естественном языке**  
Action Detection and Prediction, Video Annotation, Video and Language Understanding, Text-to-Video, VQA
- **Обучение глубоких сетей как активных агентов**  
Reinforcement Learning, Lifelong Learning
- **Глубокое обучение с использованием структурных моделей, баз знаний и программ логического вывода**  
Graph Structured CNN, Deep Visual Reasoning
- **Автоматическое конструирование и обучение глубоких сетей**

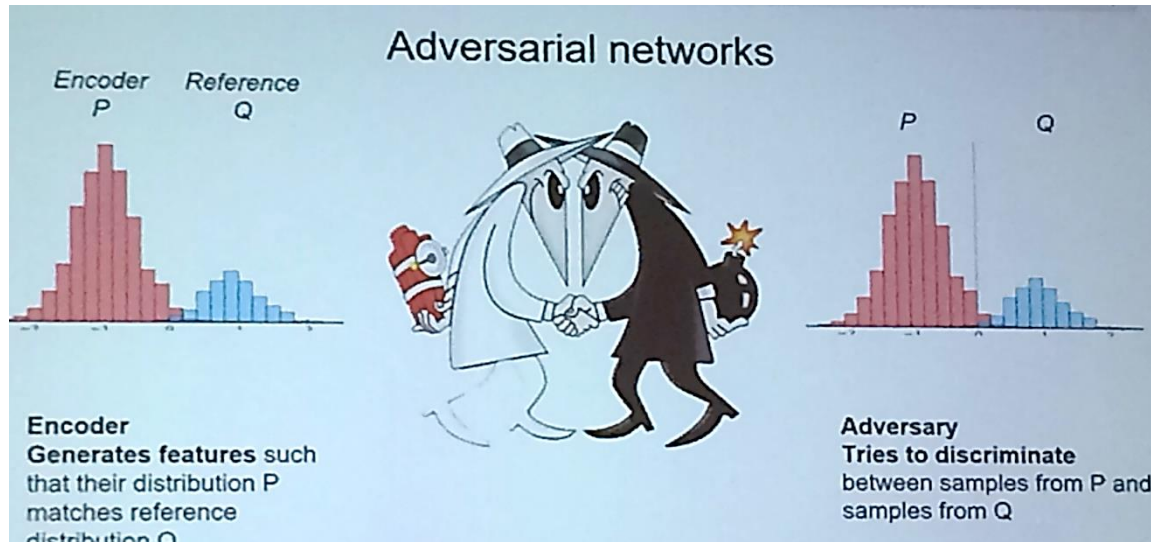
# Domain Transfer Learning



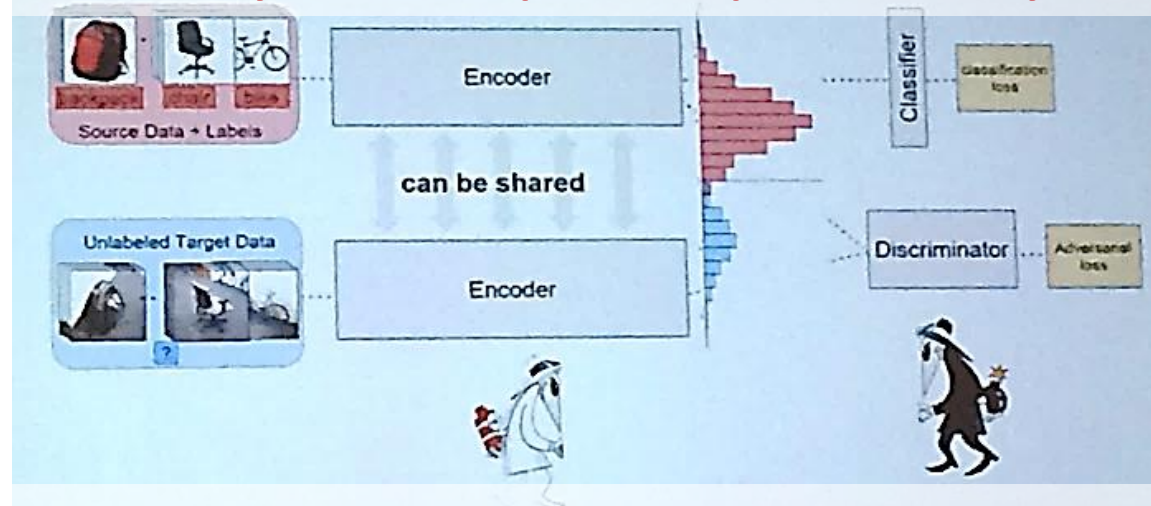
Перенос выученных закономерностей в новую область применения



# Генеративные конкурирующие сети



**Генератор создает визуальные образы, стараясь обмануть Дискриминатор...**



**....Дискриминатор старается отличить фантазии Генератора от реальности**

# Generative Adversarial Networks (GANs)

Zebras ↔ Horses



zebra → horse



horse → zebra

Summer ↔ Winter



summer → winter



winter → summer



# Generative Adversarial Networks (GANs)



apple → orange



orange → apple

**GAN – сеть,  
обладающая  
воображением!**

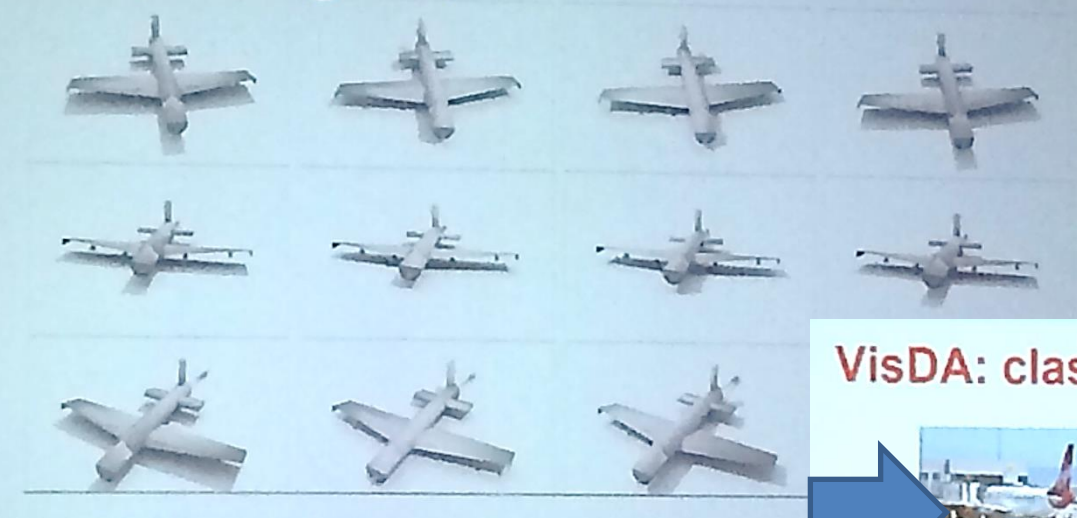


Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, Jun-Yan Zhu et al., ICCV, 2017



# Generative Adversarial Networks (GANs)

VisDA: classification source domain (CAD 2D imgs)  
12 object categories,



VisDA: classification target domain (real imgs)



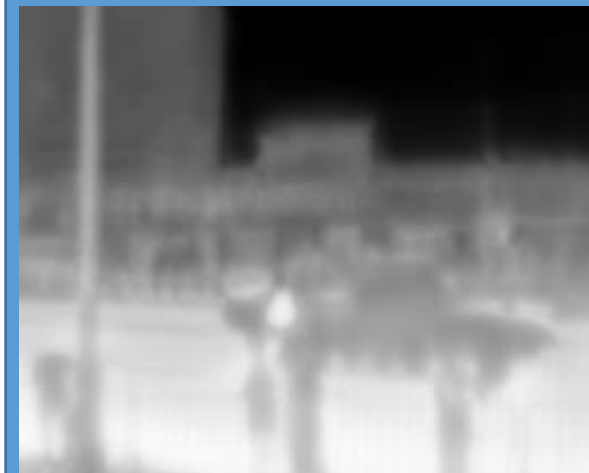
# Применение конкурентных ГНС для синтеза фотореалистичных изображений различного диапазона (ГосНИИАС-2017)



реальное ТВ изображение



реальное ИК изображение



ИК изображение, сгенерированное ГНС



реальное ИК изображение



реальное ТВ изображение

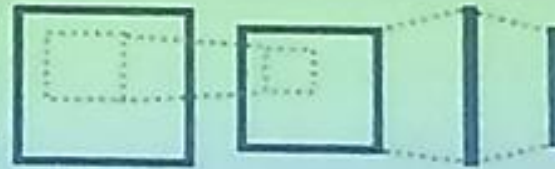


ТВ изображение, сгенерированное ГНС



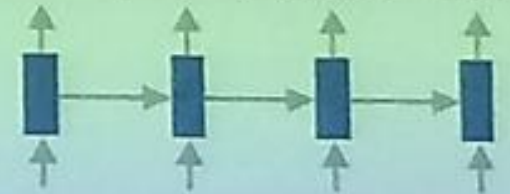
# Visual Description of Image and Video

## Visual Description: Image Captioning



Convolutional Network

men playing frisbee END



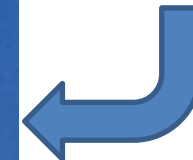
Recurrent Network

## Visual Description: Paragraph Description

*This is an image of a baseball game. The batter is wearing a white uniform with black lettering and a red helmet. The batter is wearing a white uniform with black lettering and a red helmet. The catcher is wearing a red helmet and red shirt and black pants. The catcher is wearing a red shirt and gray pants. The field is brown dirt and the grass is green.*



A man in black shirt is playing a guitar





# Visual Question Answering

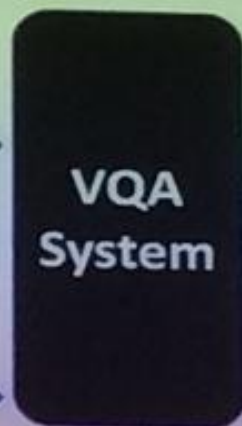
Many questions can be asked about an image

- Is it sunny?
- Is it safe to cross the street?
- How many cars are parked on the road?

Вопросы  
самых  
различных  
типов



How does the person  
in the middle feel?



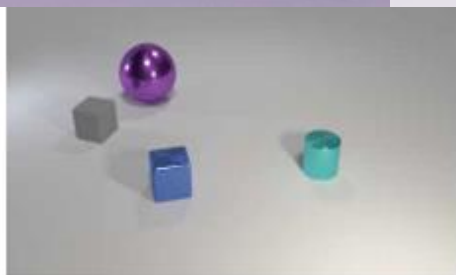
Happy

Вопросы, требующие  
понимания контекста



**CLEVR Dataset**

Q: Is there a blue box  
in the items? A: yes



Q: What shape object  
is farthest right?  
A: cylinder



Q: Are all the balls small?  
A: no

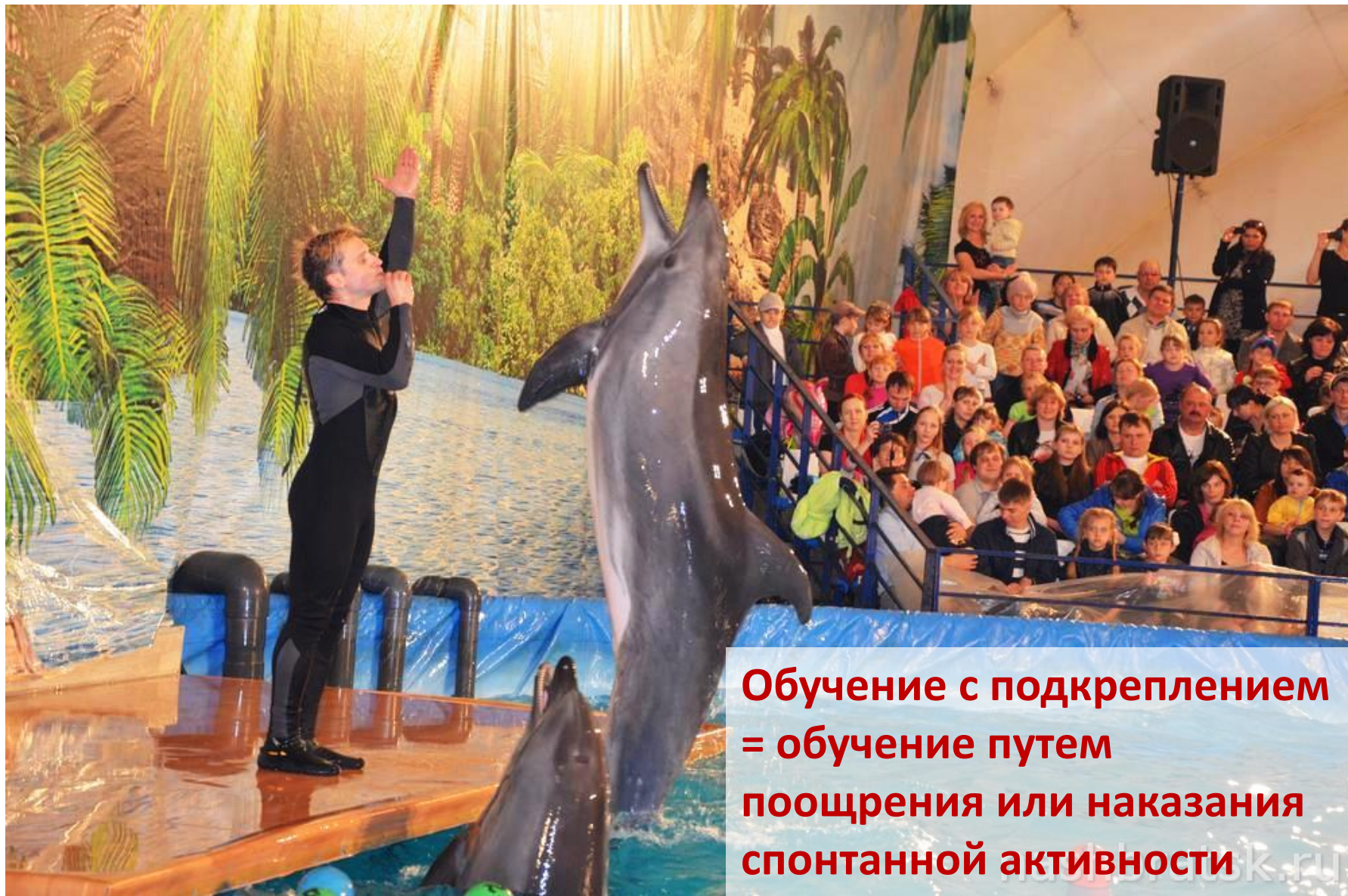


Вопросы, требующие  
рассуждений

Q: Is the green block to the  
right of the yellow sphere?  
A: yes



# Reinforcement Learning = Оперантное научение роботов



**Обучение с подкреплением  
= обучение путем  
поощрения или наказания  
спонтанной активности**

# Понимание видеoinформации как кооперативная игра двух агентов, ведущих диалог

Агент Q не видит изображение



Агент A видит изображение

Two zebra are walking around their pen at the zoo.



[0.1, -1, 0.2, ..., 0.5]



Q1: Any people in the shot?

A1: No, there aren't any.



[-0.5, 0.1, 0.7, ..., 1]



Q10: Are they facing each other?

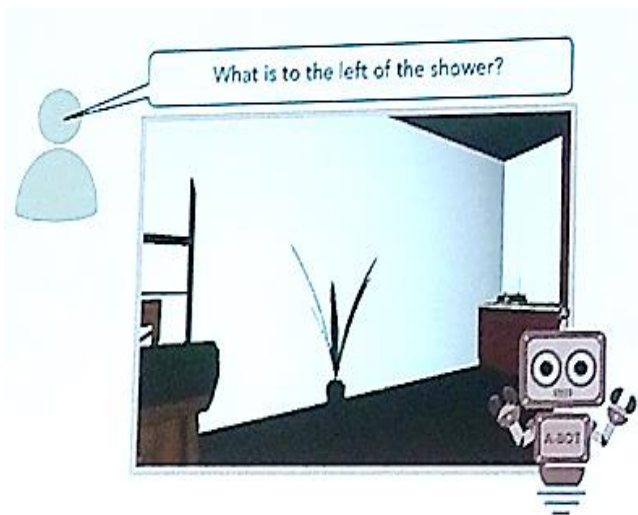
A10: They aren't.

I think we were talking about this image!

A green robot-like agent labeled 'Q' is pointing to four small images of zebras in different poses and settings.

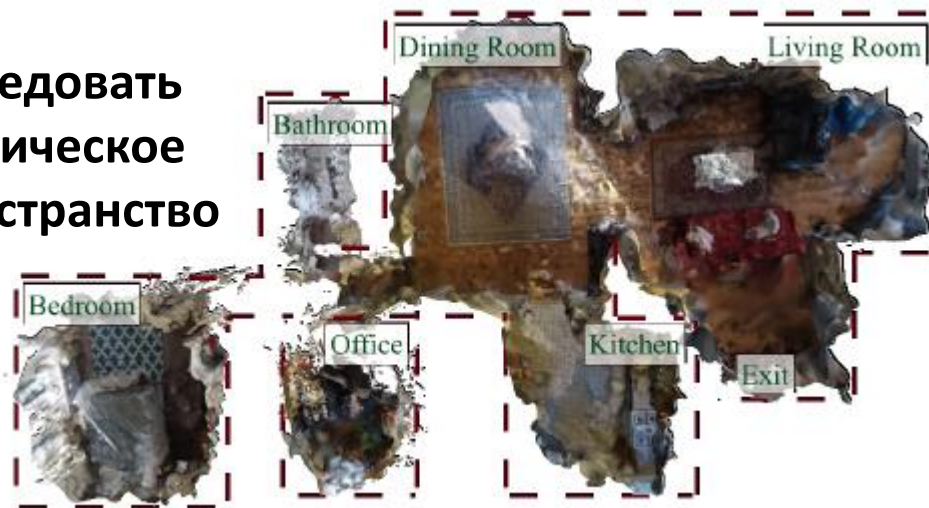


# Обучение с подкреплением автономных роботов в неизвестной виртуальной 3D сцене (challenge)

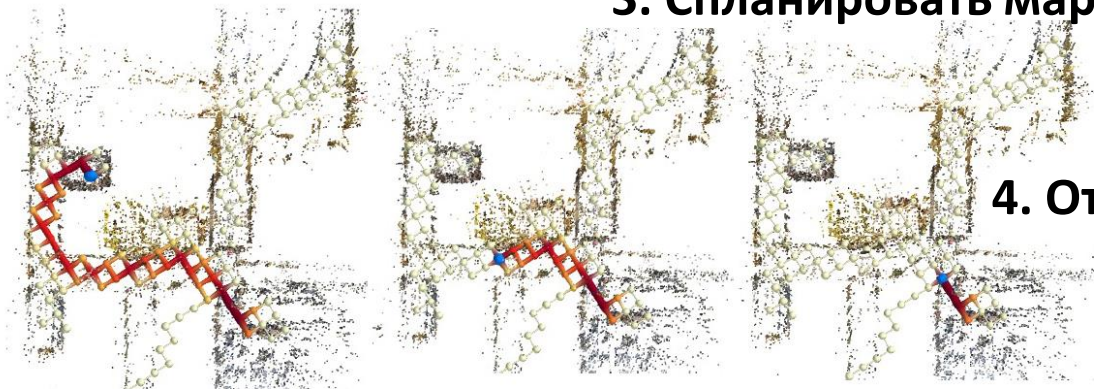


1. Понять вопрос

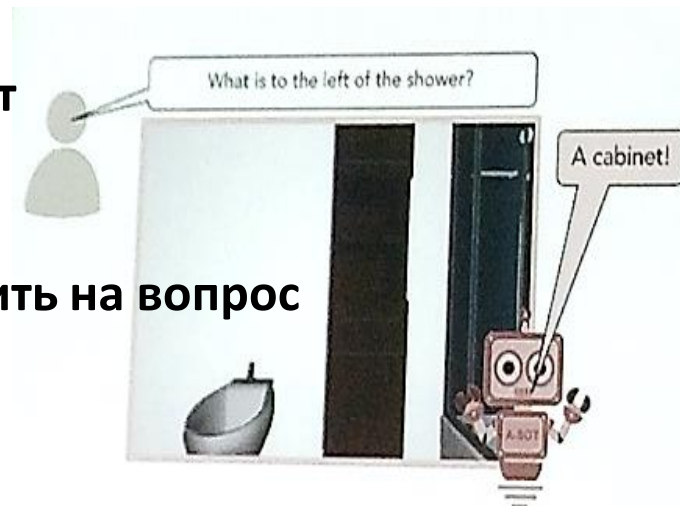
2. Исследовать семантическое 3D пространство



3. Спланировать маршрут

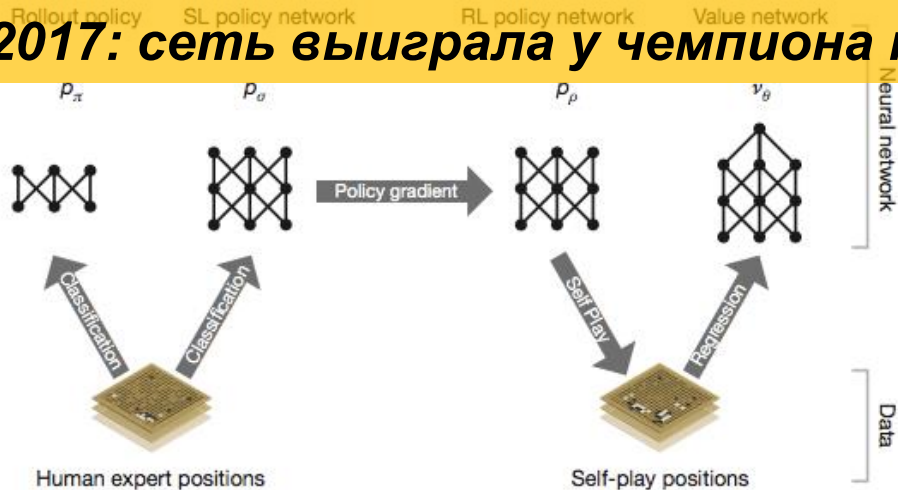


4. Ответить на вопрос



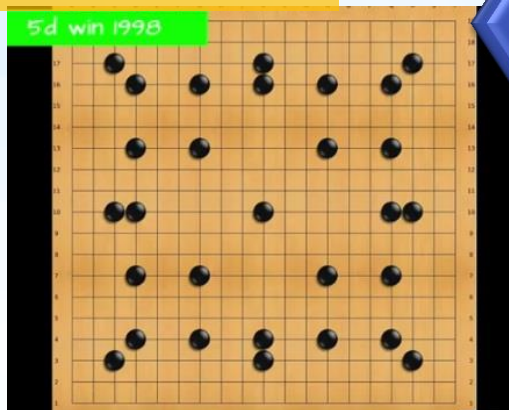
# ОТ ALPHAGO К ALPHAZERO

2017: сеть выиграла у чемпиона по игре в ГО



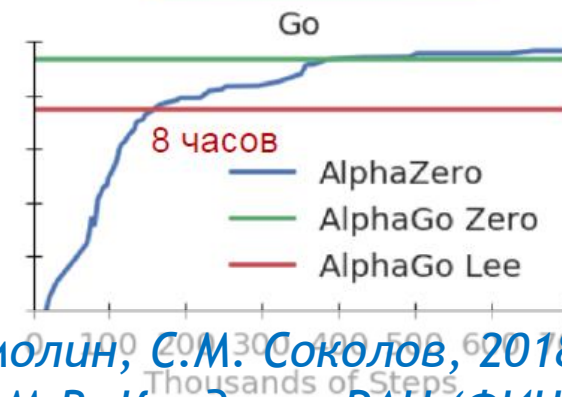
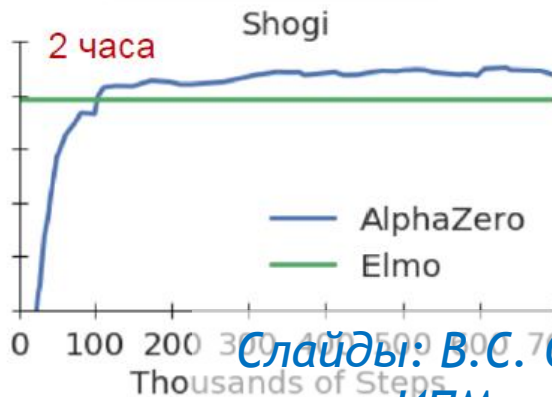
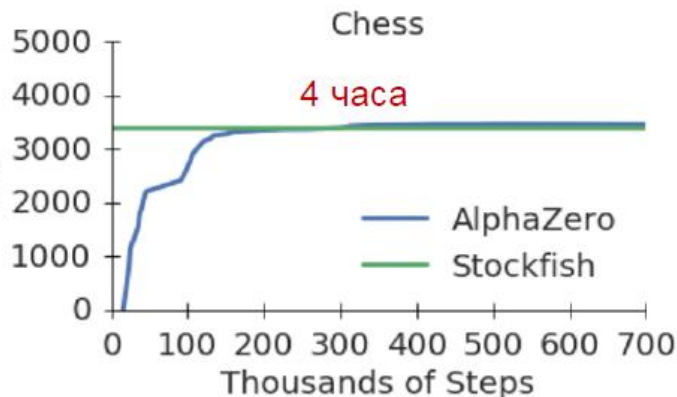
Обучение ГНС с подкреплением

Chess -  $10^{47}$  variants  
Go -  $10^{171}$  variants



2017  
дек.

Обучение с подкреплением: стратегические игры и решение оперативно-тактических задач



Слайды: В.С. Смолин, С.М. Соколов, 2018  
ИПМ им. М.В. Келдыша РАН (ФИЦ)

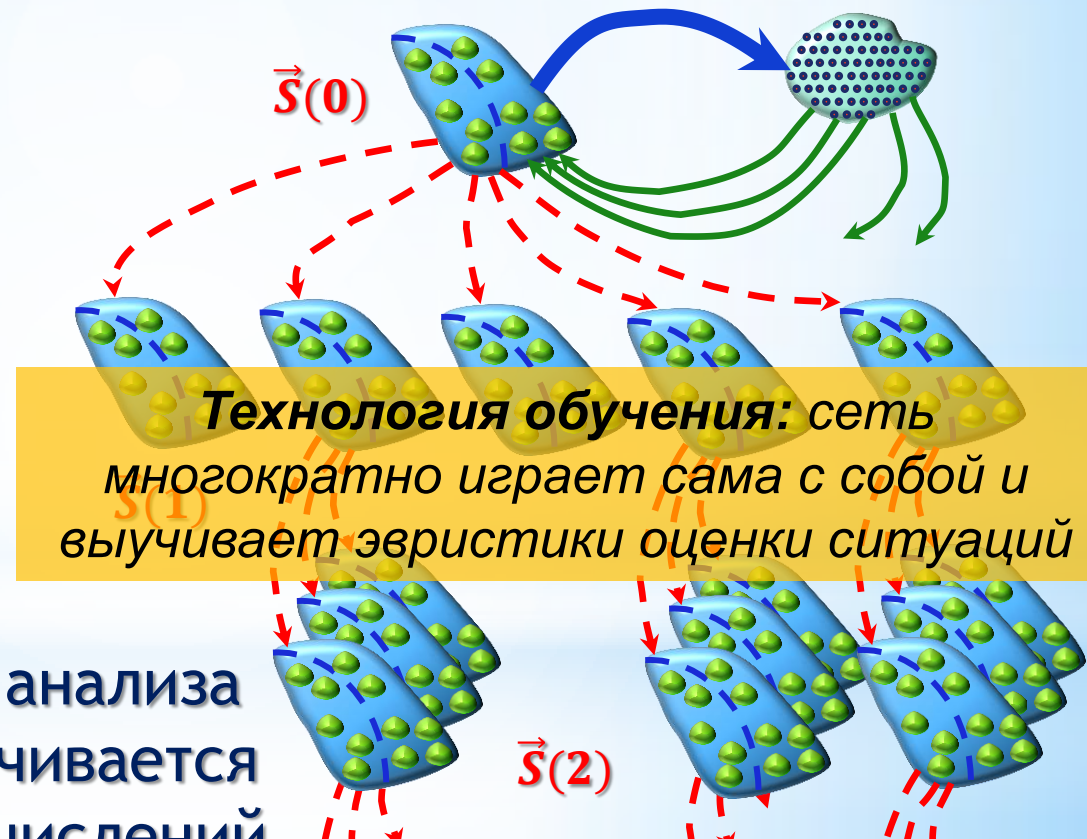
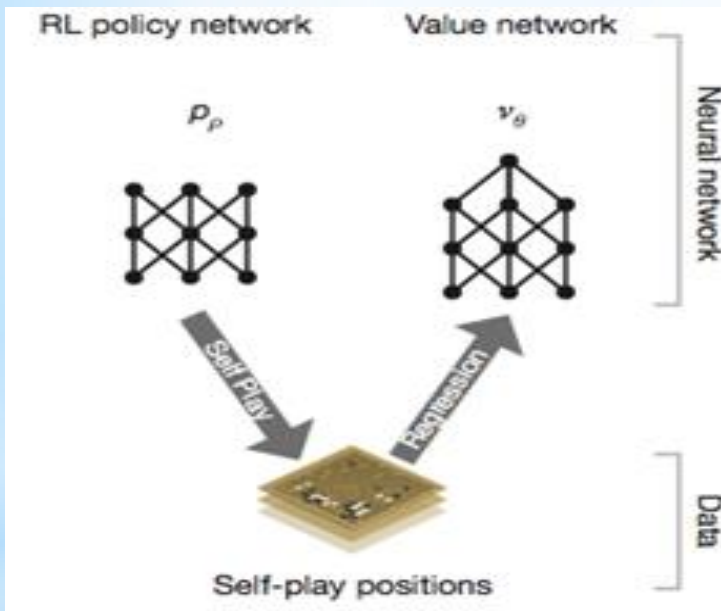


# ОБЩАЯ СТРАТЕГИЯ ПОВЕДЕНИЯ

Обучение  
ГКНС с  
подкреплением

**Комбинаторный взрыв** приводит к невозможности формирования и запоминания предварительной оценки всех допустимых действий.

Правила «игры» могут быть **как дискретными, так и непрерывными**, важно наличие выбора альтернативных действий.



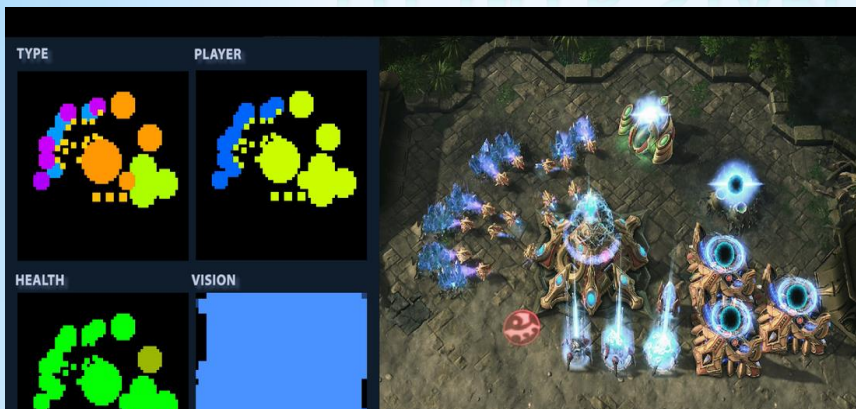
Глубина и ширина дерева анализа развития ситуации ограничивается доступными временем вычислений и объёмом памяти.

Слайды: В.С. Смолин, С.М. Соколов, 2018

43(всего 2) ИГМ им. М.В. Келдыша РАН (ФИЦ)



# OT GO K STARCRAFT II



полноценный тактический военный симулятор с упрощённой моделью ведения боя.

Протокол mini games :

**MoveToBeacon:** движение по маршруту

**CollectMineralShards:** построение оптимального маршрута через n точек

**FindAndDefeatZerglings:** Поиск и уничтожение одиночных вражеских единиц

**DefeatRoaches:** Поиск и уничтожение однотипной вражеской группы.

**DefeatZerglingsAndBanelings:** Поиск и уничтожение вражеской группы, состоящей из двух типов противников.

**CollectMineralsAndGas:** Сбор ресурсов.

**BuildMarines:** Постройка зданий, сбор ресурсов, производство единиц.

	GO	Starcraft
Размер поля	19x19	64x64, 256x256 и выше
Тип игры	пошаговая с полной информацией	реального времени с неполной информацией

## Особенности Starcraft II:

- более 20 различных типов боевых единиц и более 15 типов строений для каждой из трех сторон конфликта
- для каждой стороны конфликта боевые единицы и строения имеют уникальные характеристики, влияющие на тактику применения
- игра с неполной информацией (необходимость разведки и т.д.)
- наличие упрощённой экономической модели
- сложные игровые правила (наличие авиации, ПВО, артиллерии, проходимых/непроходимых участков местности)
- правила конкурса обязывают использовать стандартный интерфейс пользователя

# ИИ-ИНФОРМАЦИОННАЯ СИСТЕМА

Перспективные глубокие нейросетевые модели должны позволить оперативно строить модели тактических ситуаций и проводить анализ оптимальных путей решения стратегических задач

Тактико-стратегические игры

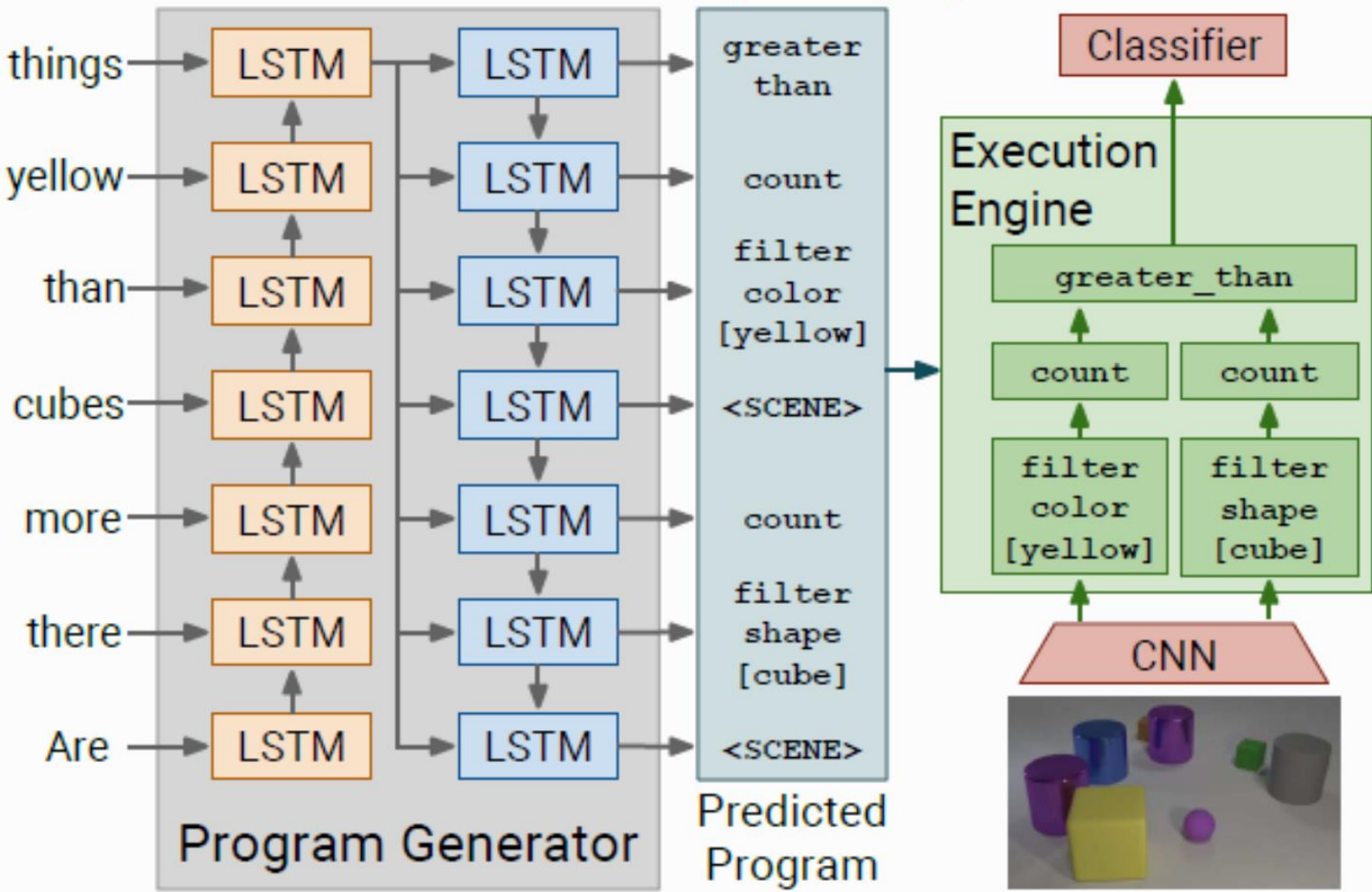
*В военно-стратегических играх сети пока проигрывают человеку, но судя по динамике развития, начнут выигрывать через 2-3 года (2020+)*



Слайды: В.С. Смолин, С.М. Соколов, 2018  
ИПМ им. М.В. Келдыша РАН (ФИЦ)

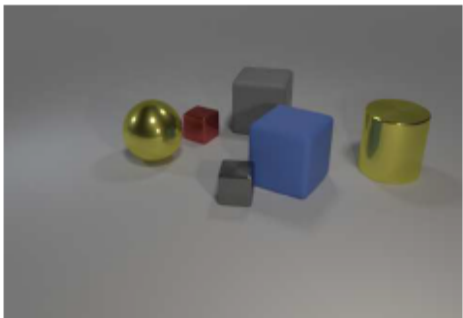
# Deep Visual Reasoning for VQA: генератор программ

**Question:** Are there more cubes than yellow things? **Answer:** Yes

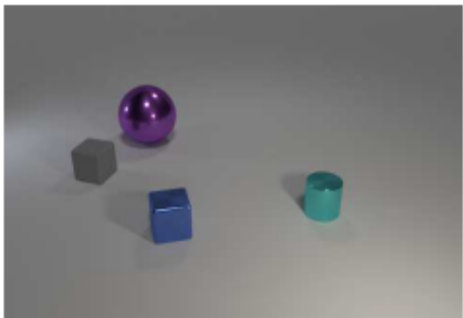




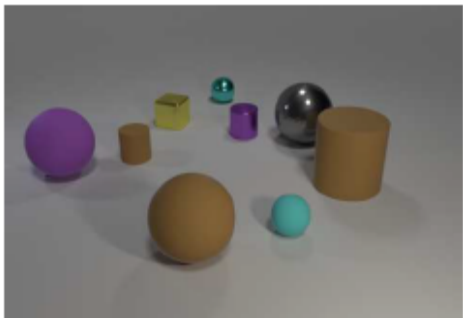
# Deep Visual Reasoning for VQA: генератор программ



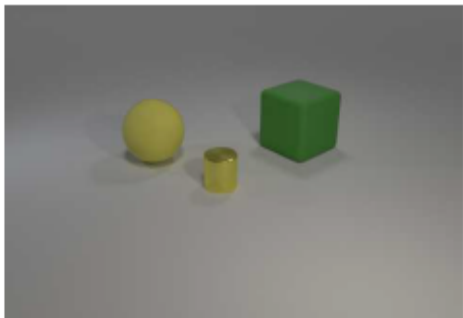
Q: Is there a blue box in the items? A: yes



Q: What shape object is farthest right? A: cylinder



Q: Are all the balls small? A: no



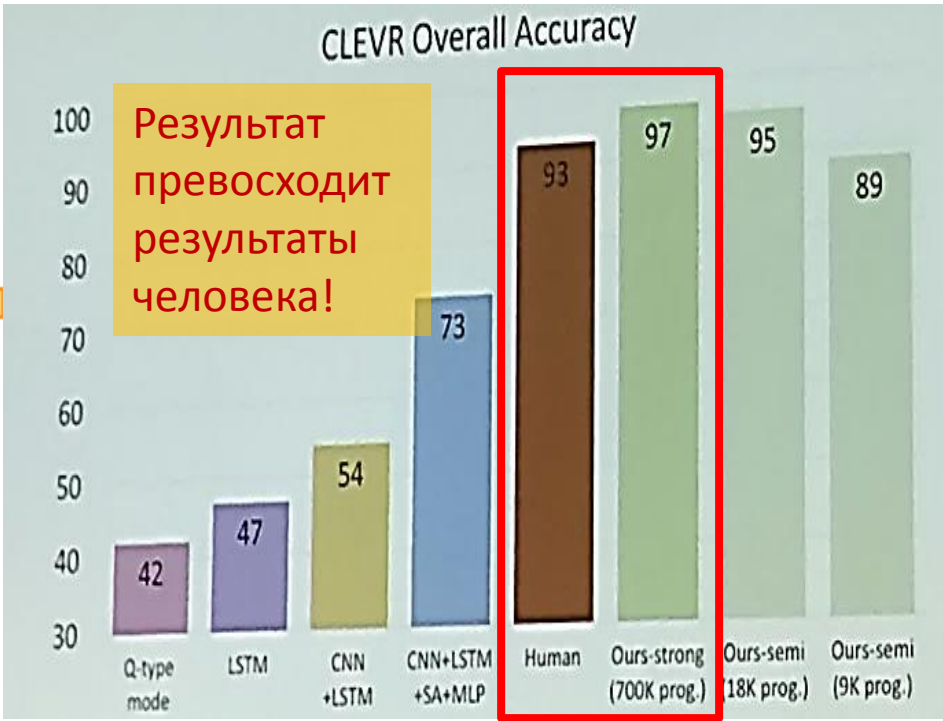
Q: Is the green block to the right of the yellow sphere? A: yes

**Predicted Program:**  
 exist  
 filter\_shape [cube]  
 filter\_color [blue]  
 scene

**Predicted Answer:**  
 ✓ yes

**Predicted Program:**  
 query\_shape  
 unique  
 relate [right]  
 unique  
 filter\_shape [cylinder]  
 filter\_color [blue]  
 scene

**Predicted Answer:**  
 ✓ cylinder



# Visual Reasoning with Graph Structured CNNs

## Graphs: Bridge to Language and Thought

The boy wants to go



Abstract Meaning Representation  
[Banarescu et al. 2013]



Google Knowledge Graph



Abstract Syntactic Trees  
Image: Wikipedia

Seeing = pixels  $\rightarrow$  spatially grounded *graphs*



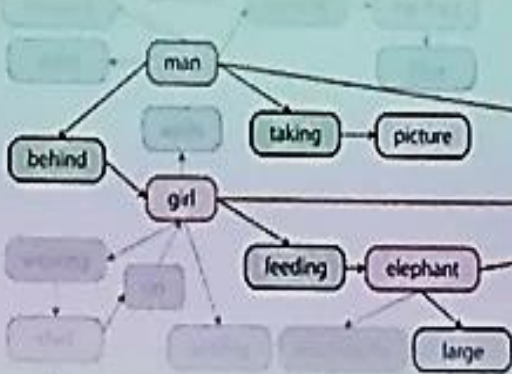
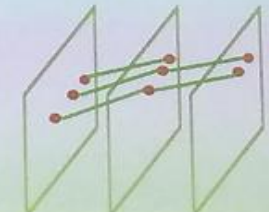
Human pose



Segmentation



Tracking

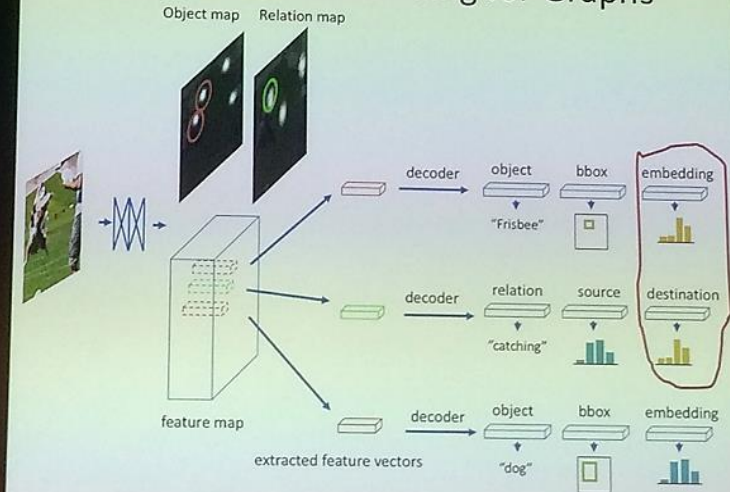


Scene Graph Generation



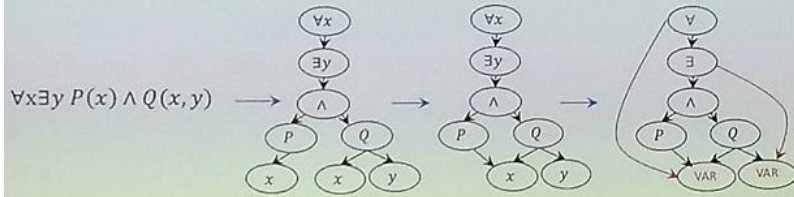
[Krishna et al. '16]

## Associative Embedding for Graphs



# Глубокие сети улучшают результаты ИИ-1 в области автоматического доказательства теорем (2017)

## Deep Graph Embedding



## HolStep [Kaliszyk et al. 2017]

- Benchmark for machine learning for Theorem Proving
- 2M+ conjecture-fact pairs of higher-order logic statements

Conjecture:  $\forall \alpha \forall \beta (\sin(\alpha) = \sin(\beta)) = ((\alpha = \beta) \vee (\alpha = \pi - \beta))$

Relevant fact:  $\forall \alpha \forall \beta \sin(\alpha - \beta) = \sin(\alpha)\cos(\beta) - \sin(\beta)\cos(\alpha)$

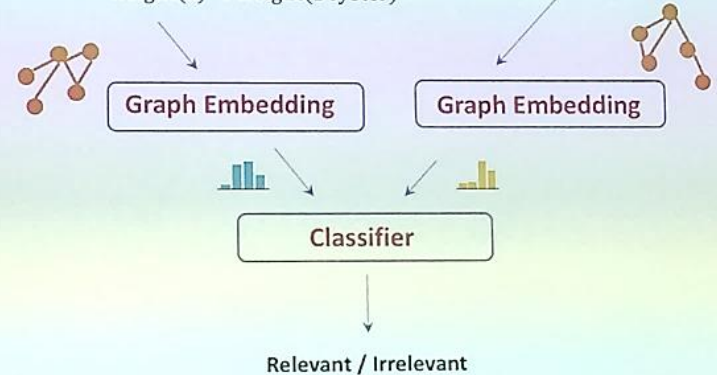
Irrelevant fact:  $(x > 0) \wedge (y > 0) \rightarrow (xy > 0)$

	Sequence embedding	Graph embedding
	CNN [Kaliszyk et al. '17]	CNN-LSTM [Kaliszyk et al. '17]
Accuracy	82	83
		Ours
		90.3

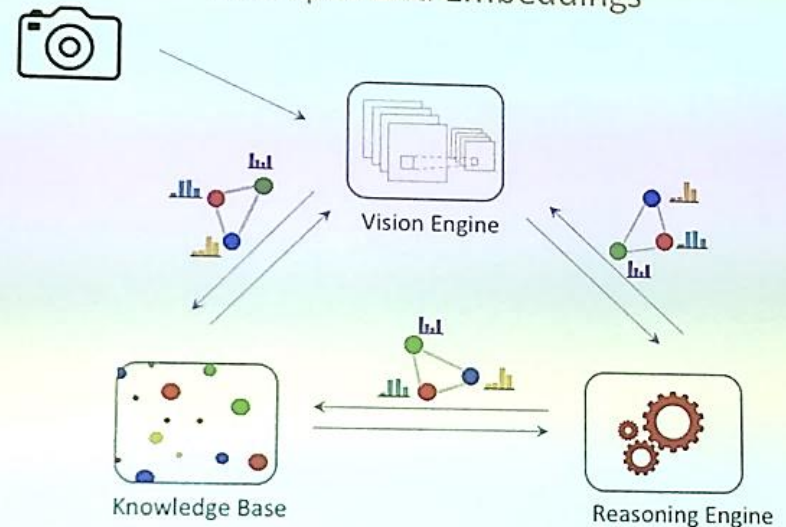
## Learn Heuristics ("intuition") for Premise Selection

Conjecture:  $\forall x \text{IsBuilding}(x) \wedge \text{OnCampus}(x) \rightarrow \text{Height}(x) < \text{Height}(\text{Beyster})$

Fact:  $\text{Age}(\text{EECS}) < \text{Age}(\text{Dow})$

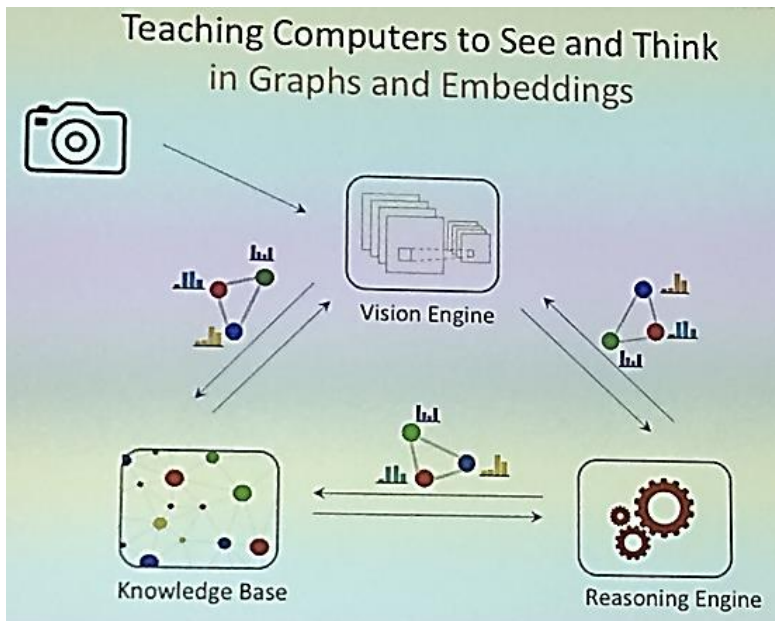
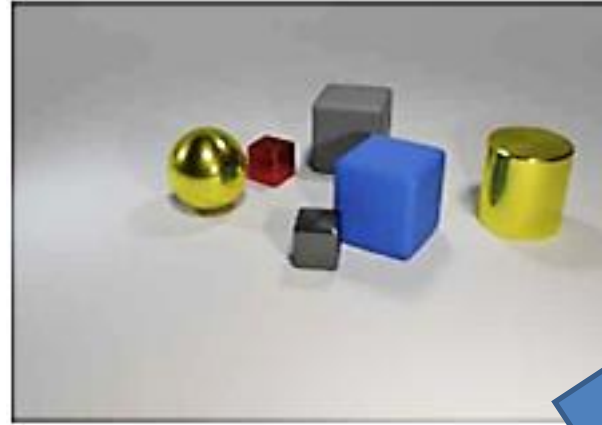
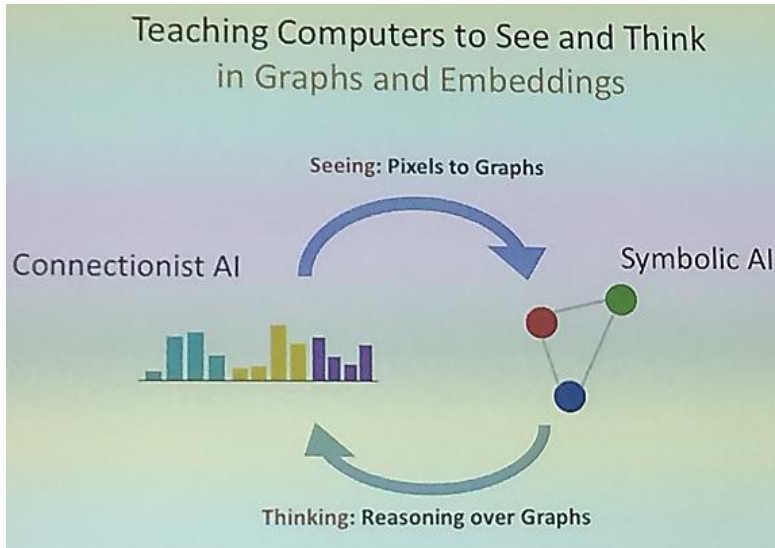


## Teaching Computers to See and Think in Graphs and Embeddings





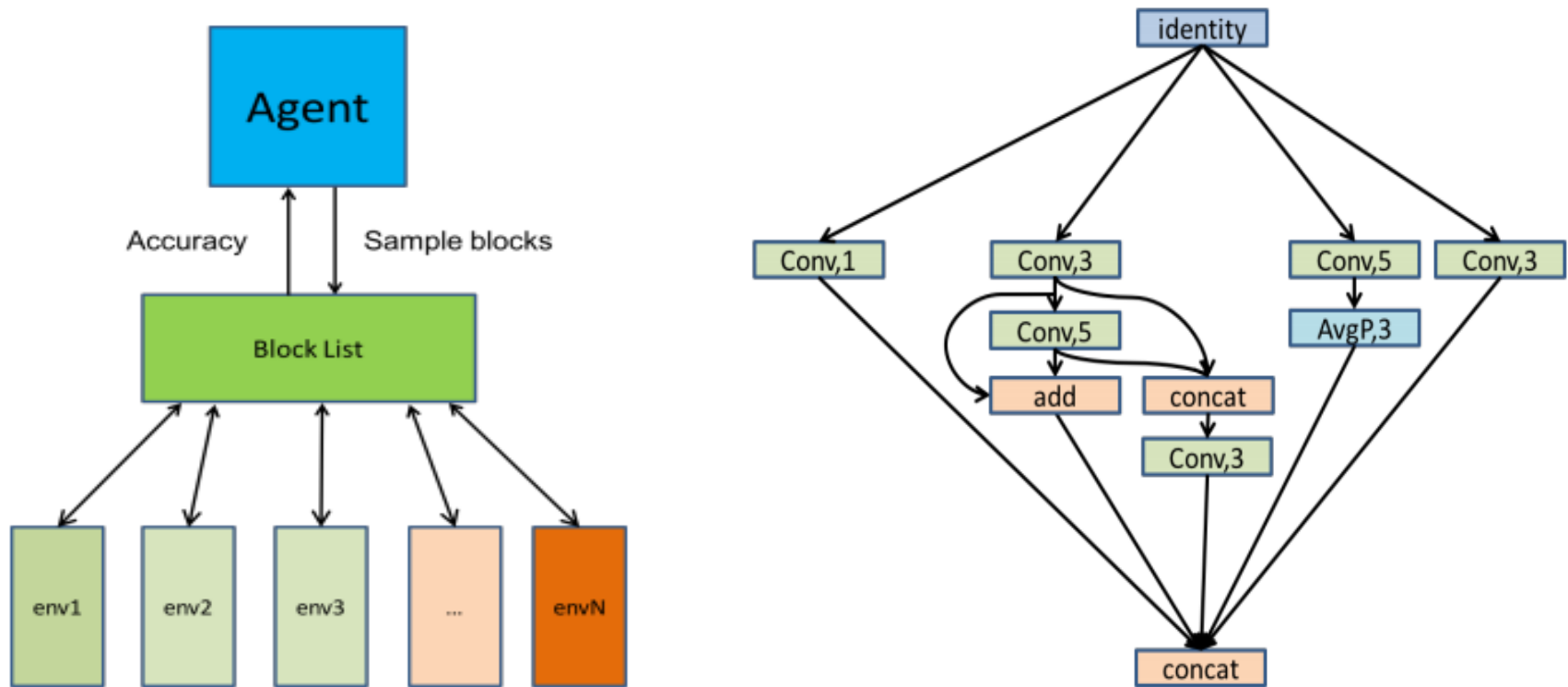
# Visual Reasoning with Graph Structured CNNs



*Возможность представления всей информации о сцене в виде графа семантических связей позволяет использовать базы знаний, логический вывод и эвристики для анализа видеоданных о реальном мире*

# Глубокие сети формируют и учат глубокие сети

Обучение с подкреплением. Q-Learning.  
32 GPU (Предыдущий вариант 800 GPU)



Practical Network Blocks Design with Q-Learning, CVPR-2017

<https://arxiv.org/pdf/1708.05552.pdf>

# Глубокие сети формируют и учат глубокие сети

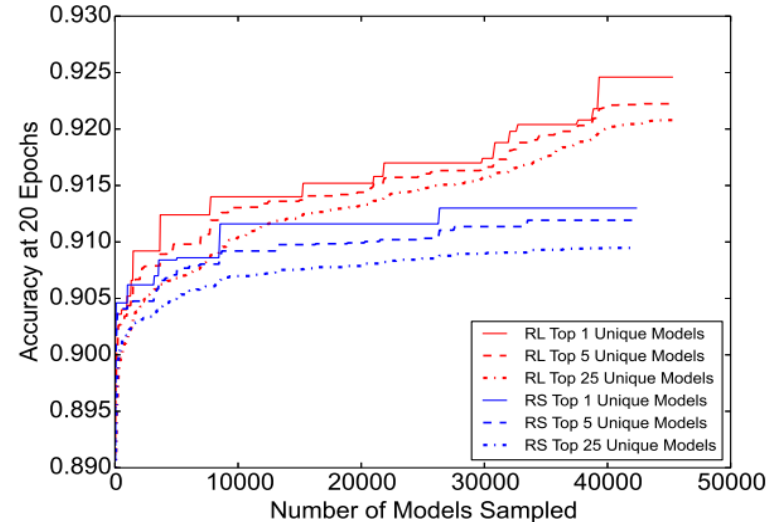
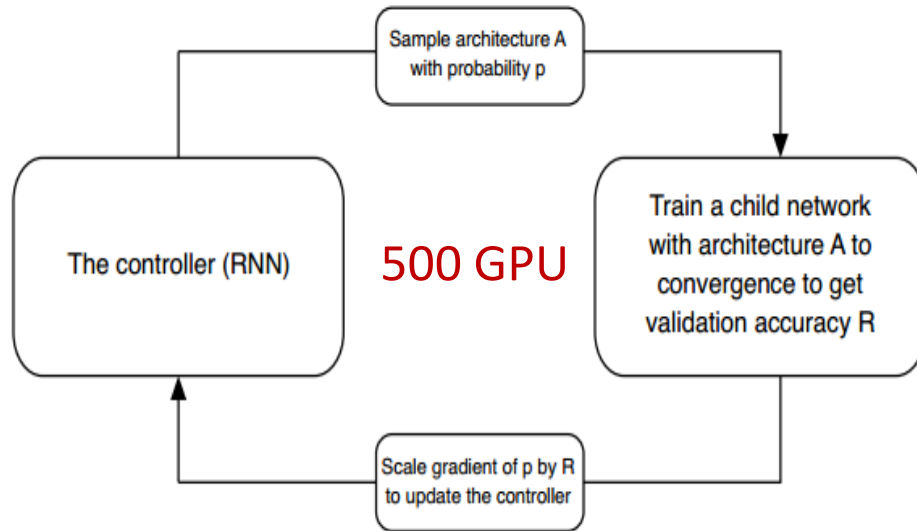


Image Net	Model	image size	# parameters	Mult-Adds	Top 1 Acc. (%)	Top 5 Acc. (%)
	Inception V2 [19]	224×224	11.2 M	1.94 B	74.8	92.2
	<b>NASNet-A (5 @ 1538)</b>	<b>299×299</b>	<b>10.9 M</b>	<b>2.35 B</b>	<b>78.6</b>	<b>94.2</b>
	Inception V3 [36]	299×299	23.8 M	5.72 B	78.0	93.9
	Xception [5]	299×299	22.8 M	8.38 B	79.0	94.5
	Inception ResNet V2 [34]	299×299	55.8 M	13.2 B	80.4	95.3
	<b>NASNet-A (7 @ 1920)</b>	<b>299×299</b>	<b>22.6 M</b>	<b>4.93 B</b>	<b>80.8</b>	<b>95.3</b>
	ResNeXt-101 (64 x 4d) [41]	320×320	83.6 M	31.5 B	80.9	95.6
	PolyNet [42]	331×331	92 M	34.7 B	81.3	95.8
	DPN-131 [4]	320×320	79.5 M	32.0 B	81.5	95.8
	<b>SENet [15]</b>	<b>320×320</b>	<b>145.8 M</b>	<b>42.3 B</b>	<b>82.7</b>	<b>96.2</b>
	<b>NASNet-A (6 @ 4032)</b>	<b>331×331</b>	<b>88.9 M</b>	<b>23.8 B</b>	<b>82.7</b>	<b>96.2</b>

*Автоматически сформированные глубокие сети впервые превзошли показатели глубоких сетей, сформированных вручную (2017)*

Learning Transferable Architectures for Scalable Image Recognition, CVPR-2017

<https://arxiv.org/pdf/1707.07012.pdf>



# 2018: Алгоритмическое обеспечение, необходимое для автономных и интеллектуальных систем

## Вторая волна технологической революции:

глубокое обучение+  
компьютерное зрение+  
базы знаний+  
семантические модели+  
системы логического вывода+  
автоматическое программирование+  
общение с человеком на естественном языке+  
оперантное обучение агентов +

Навигация



Обработка сенсорных данных  
(зрение,...)

Управление  
(планирование, оптимизация, игры,...)

Машинное обучение  
(анализ данных)

Искусственный интеллект  
(базы знаний, логика, рассуждения)

Прогноз ИИ: 2020+



Все необходимое для автономных систем!

победа в го+  
сети, обучающие сети

## Прогноз (2020+): возникнет единая технология обучаемого машинного интеллекта, которая обеспечит:

- Ведение информативного двустороннего диалога ИИ-ИИ и ИИ-оператор на естественном (русском) языке в процессе постановки и выполнения оперативных задач;
- Обучение ГКНС решению задач обработки сенсорной информации с учетом задач управления;
- Обучение ГКНС пониманию сложной наблюдаемой динамической сцены с использованием структурных моделей, баз знаний и логического вывода;
- Обучение ГКНС автономных систем, действующих в заранее неизвестной виртуальной динамической 3D сцене;
- Непрерывное самообучение ГКНС ИИ на протяжении всего цикла их функционирования;
- Автоматизированный процесс конструирования, обучения и оперативного дообучения ГКНС ИИ с использованием обучающих глубоких сетей.

***Уровень технологии обучаемого машинного интеллекта (2020+):  
Полная готовность к созданию практических интеллектуальных систем!***

# **Структурно-функциональный анализ и синтез глубоких нейронных сетей**

*(некоторые математические аспекты  
происходящей технологической  
революции)*



# Как все мы знаем, в компьютерном зрении и машинном обучении произошла революция...

В задачах распознавания визуальных образов (и не только)

SUPERHUMAN VISUAL  
PATTERN RECOGNITION

с 2011 г. результаты автоматического распознавания изображений на уровне или лучше, чем у человека

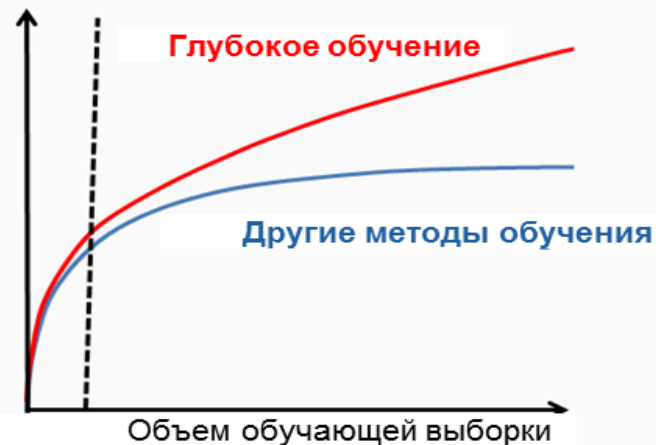
JÜRGEN SCHMIDHUBER 2013

**2017: Общей теории построения и обучения CNN как не было, так и нет...**

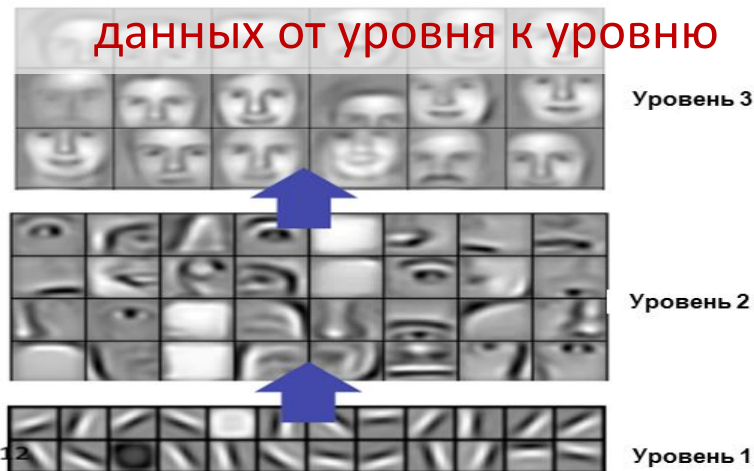
<http://people.idsia.ch/~juergen/deeplearning.html>

- обучение на сверхбольших объемах данных

Вероятность распознавания



- иерархическое обучение с повышением абстракции данных от уровня к уровню

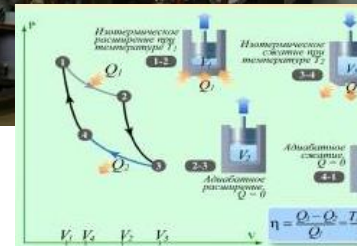
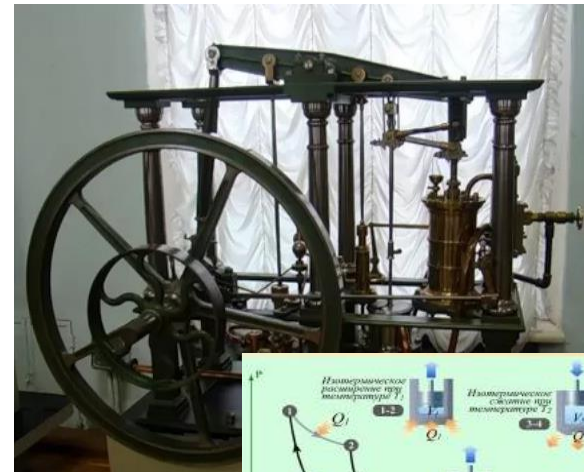


# Какой теории у нас нет и в ближайшее время не будет

## Хотелось бы иметь общую теорию построения и обучения CNN,

которая была бы способна ответить на следующие основные вопросы:

- Какова должна быть сложность CNN для решения задач определенного типа на данных определенной сложности при обучении на выборках определенного размера?
- Как оптимально формировать структуру глубокой сети для конкретных типов задач и конкретных типов данных?
- Как оптимально формировать стратегию обучения глубокой сети для конкретных типов задач и конкретных типов данных?
- От чего зависит скорость обучения глубокой сети, и как на нее влиять в процессе обучения?
- Как заранее оценить достижимые результаты некоторой заданной ГКС при обучении?
- Как предсказать или хотя бы семантически интерпретировать структурные описания, которые порождает глубокая сеть?



# Какую теорию мы можем попытаться построить

## Теория Структурно-Функционального Анализа и Синтеза ГКНС

**Задача СФАС:** Не моделировать и предсказывать характеристики и поведение CNN, а систематизировать и обобщать известные архитектуры и «рецепты» их построения. **Методика СФАС** должна ответить на следующие вопросы:

- Из каких типовых структурно-функциональных элементов (СФЭ) могут строиться глубокие сети?
- Каковы необходимые математические свойства этих элементов, и соответственно, каковы могут быть их различные алгоритмические реализации?
- Какие комбинации СФЭ являются допустимыми и эффективными с точки зрения построения и обучения глубоких сетей?
- Каковы возможные пути построения и обучения глубоких сетей для анализа и распознавания нерегулярных, неоднородных или сложно структурированных данных?



ПЕРИОДИЧЕСКАЯ СИСТЕМА ЭЛЕМЕНТОВ Д.И. МЕНДЕЛЕЕВА																			
I										VII								VIII	
II										(H)									
1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
1	H	He	Li	Be	B	C	N	O	F	Ne	Na	Mg	Al	Si	P	S	Cl	Ar	
2	Li	Be	B	C	N	O	F	Ne	Na	Mg	Al	Si	P	S	Cl	Ar	K	Ca	
3	Na	Mg	Al	Si	P	S	Cl	Ar	K	Ca	Sc	Ti	V	Cr	Mn	Fe	Cu	Zn	
4	K	Ca	Sc	Ti	V	Cr	Mn	Fe	Cu	Zn	Ga	Ge	As	Se	Br	Kr	Rb	Sr	
5	Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe	
6	Cs	Ba	La	Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu	Hf	
7	Ra	Ac	Th	Pa	U	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No	Lr	Rf	Os	
8																			
9																			
10																			
ЛАНТАНОИДЫ																			
58	59	60	61	62	63	64	65	66	67	68	69	70	71						
Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu						
АКТИНОИДЫ																			
88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103				
Ra	Ac	Th	Pa	U	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No	Lr				



# *Какие данные мы хотели бы обрабатывать и распознавать при помощи CNN*

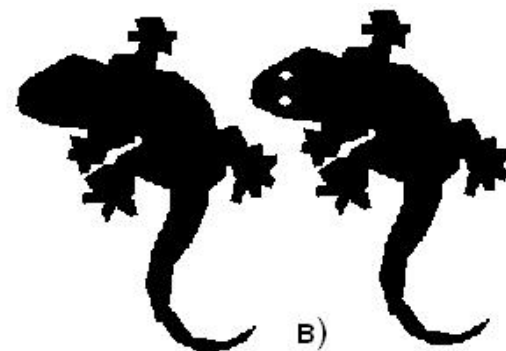
*Двумерные фигуры...*



а)

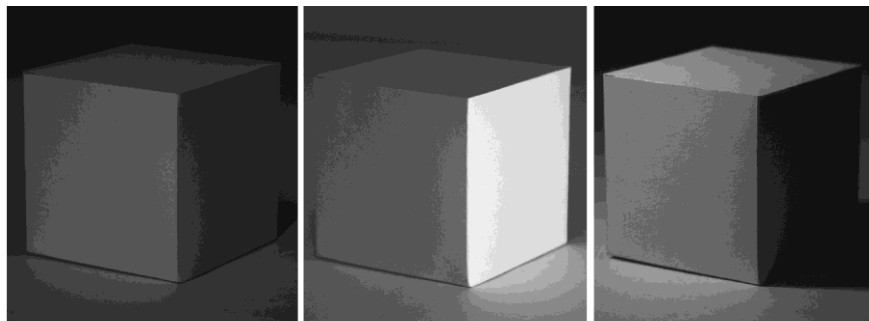


б)



в)

*Двумерные изображения...*



г)

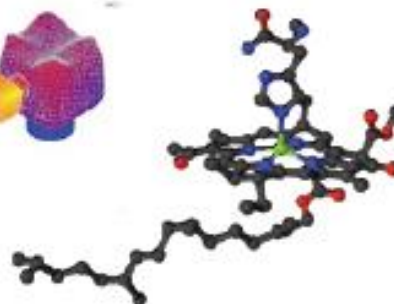
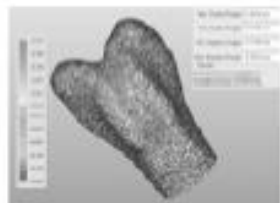


д)

*Иллюстрации из работ Л.М.Местецкого, Ю.П. Пытьева, А.И.Чуличкова*

# Какие данные мы хотели бы обрабатывать и распознавать при помощи CNN

Трехмерные объекты, поверхности, облака точек, ДНК...



Многомерные данные, сигналы, произвольные графы...

**Раз уж CNN победили другие методы анализа данных на изображениях, пусть проявят себя и на всех остальных типах данных**

...Любые структурно распределенные данные

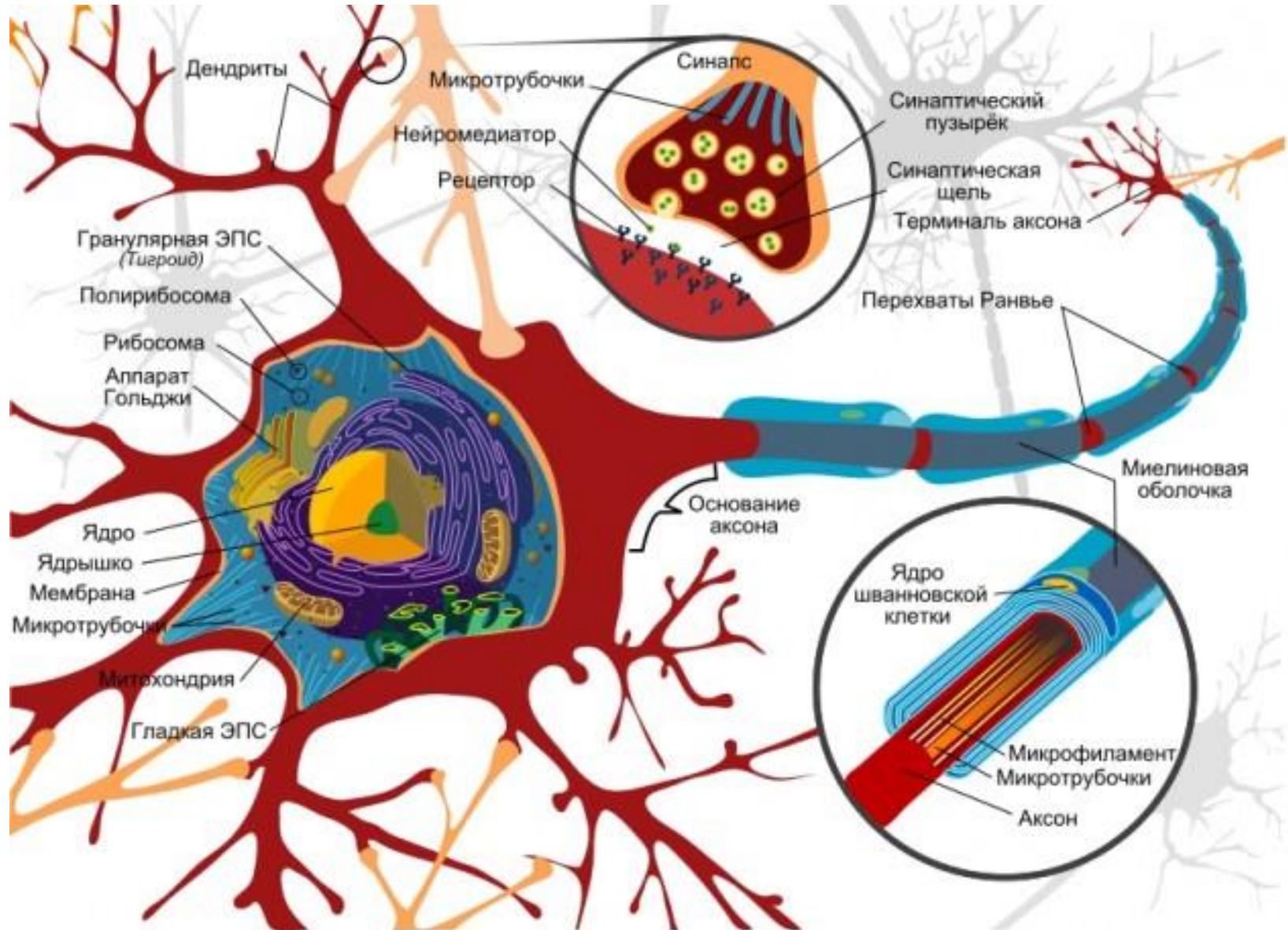
*Иллюстрации из работ С. Лафона, Ю. Визильтера, В. Горбацевича*

*Глубокие конволюционные  
нейронные сети  
(краткое напоминание)*



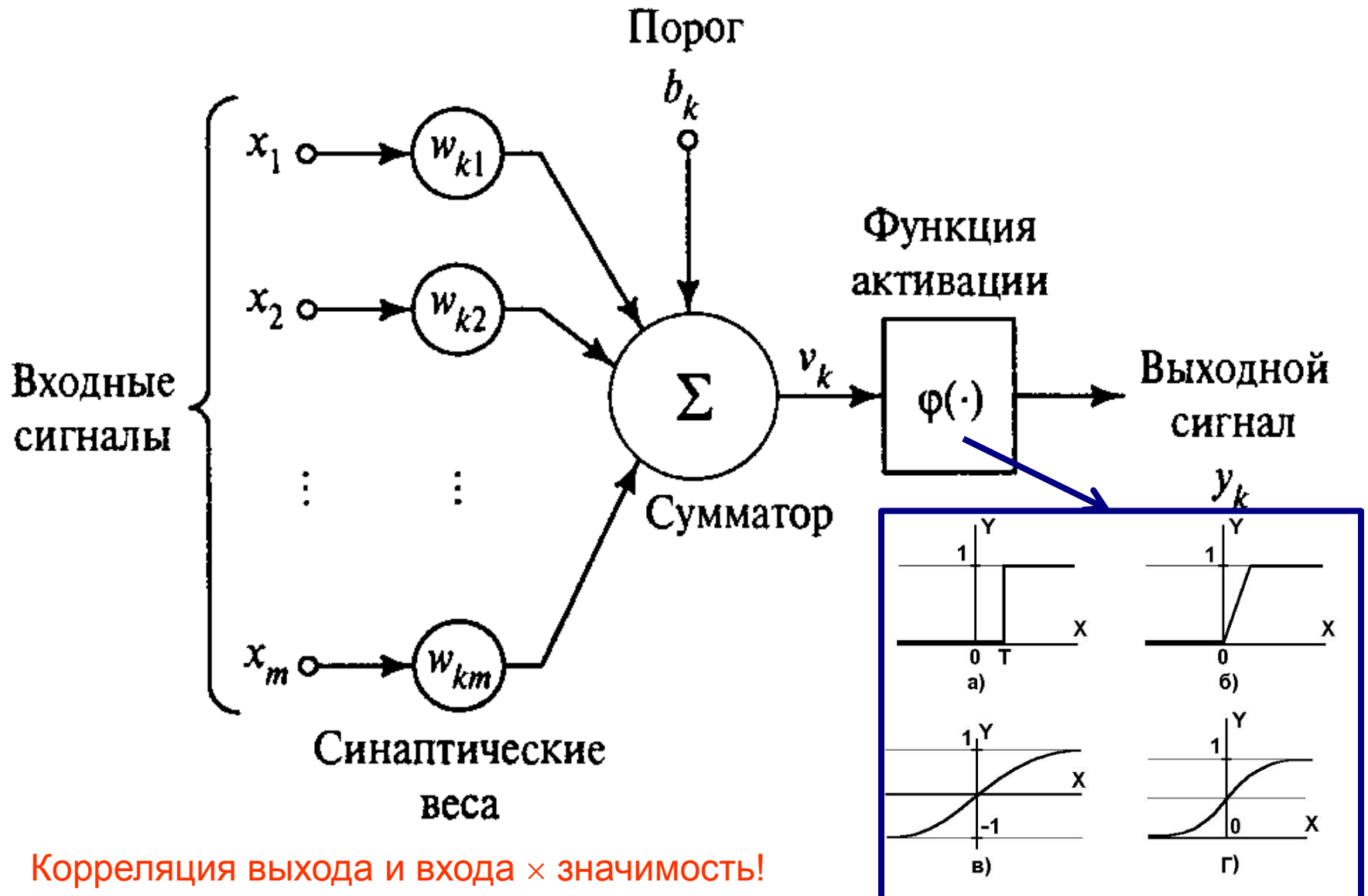
# Нейроны: идея из биологии

Настоящие нейроны и нейронные сети в нервной системе



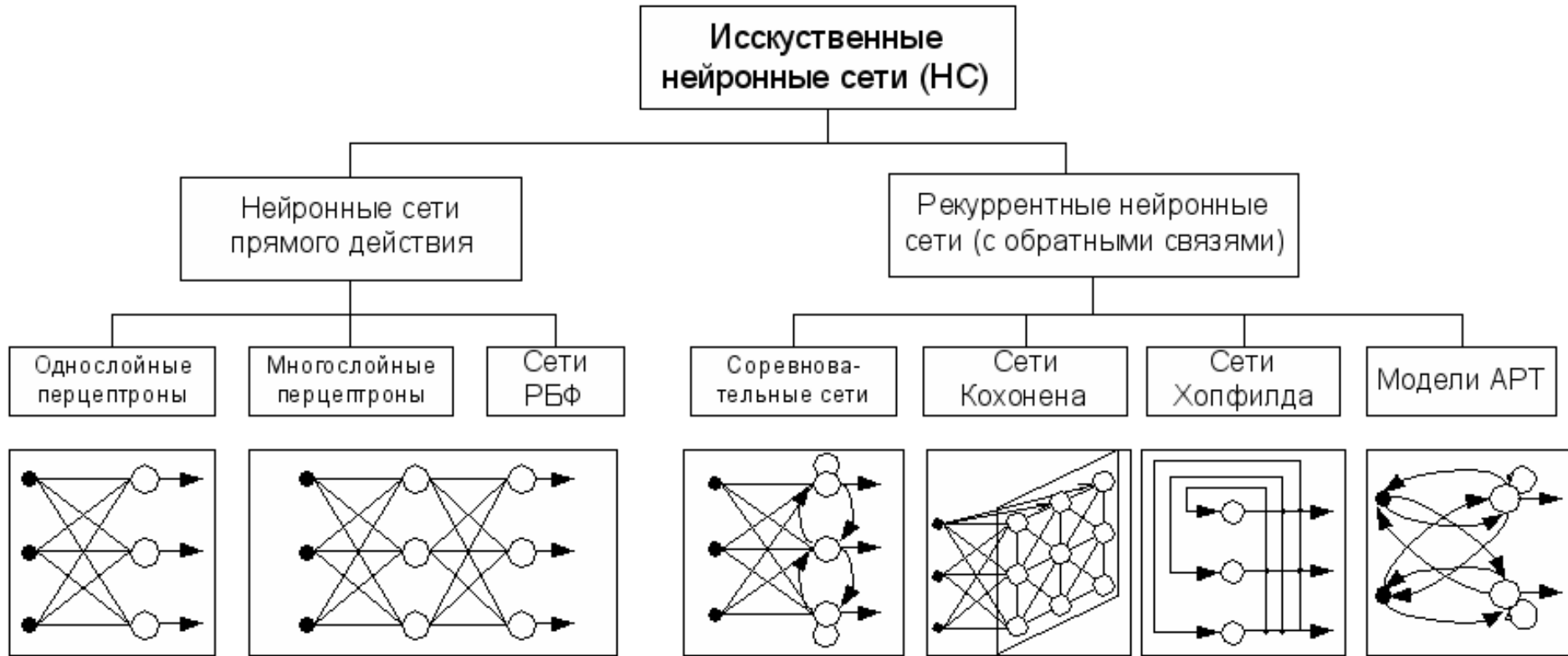
# Нейроны и нейронные сети

Модель искусственного нейрона – сумматор с нелинейностью



# Нейроны и нейронные сети

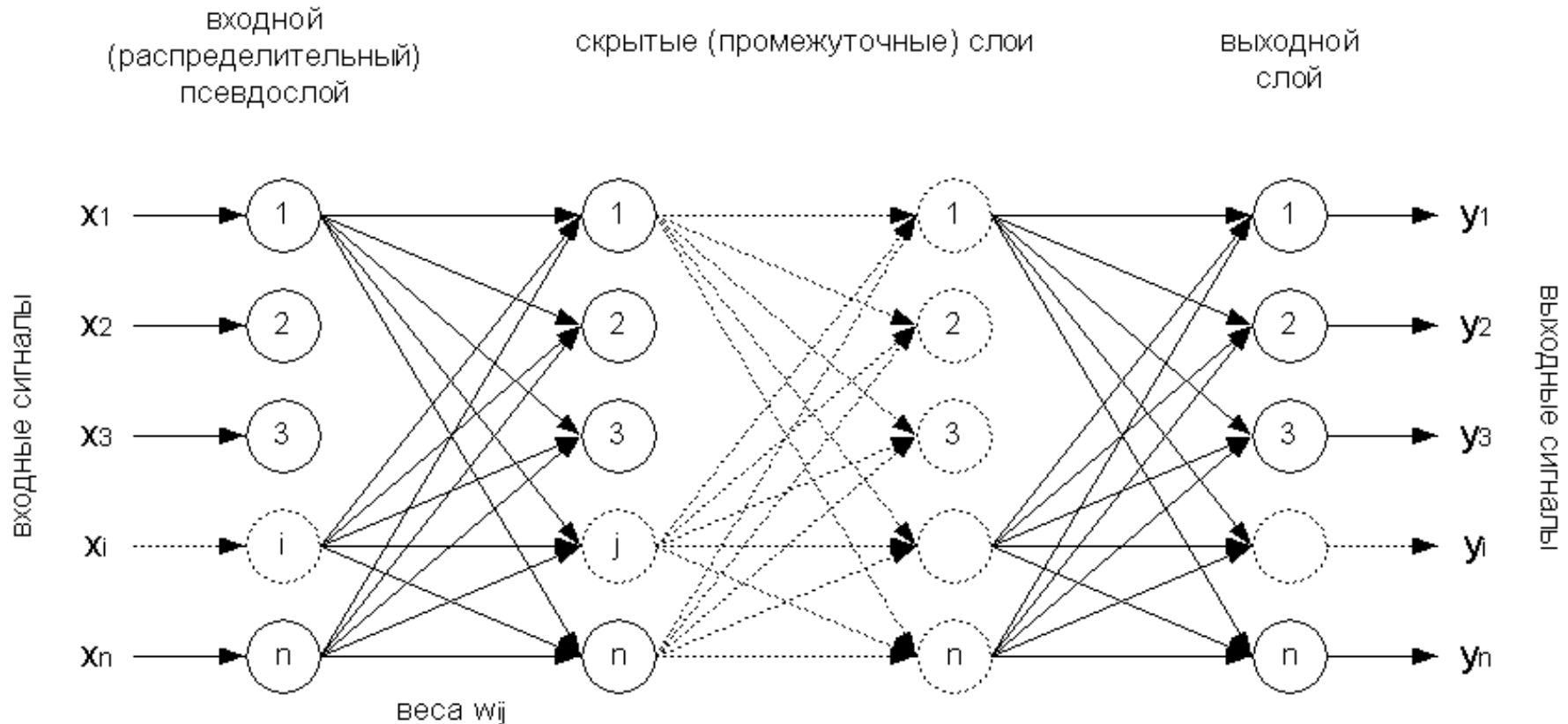
Нейронные сети могут иметь различную архитектуру





# Нейроны и нейронные сети

## Многослойный персептрон

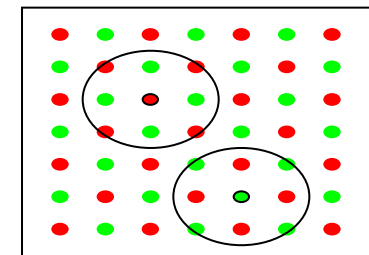
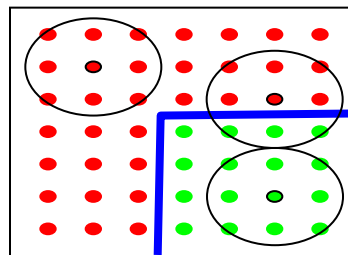
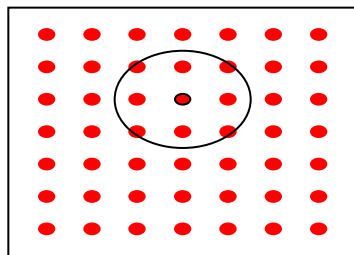
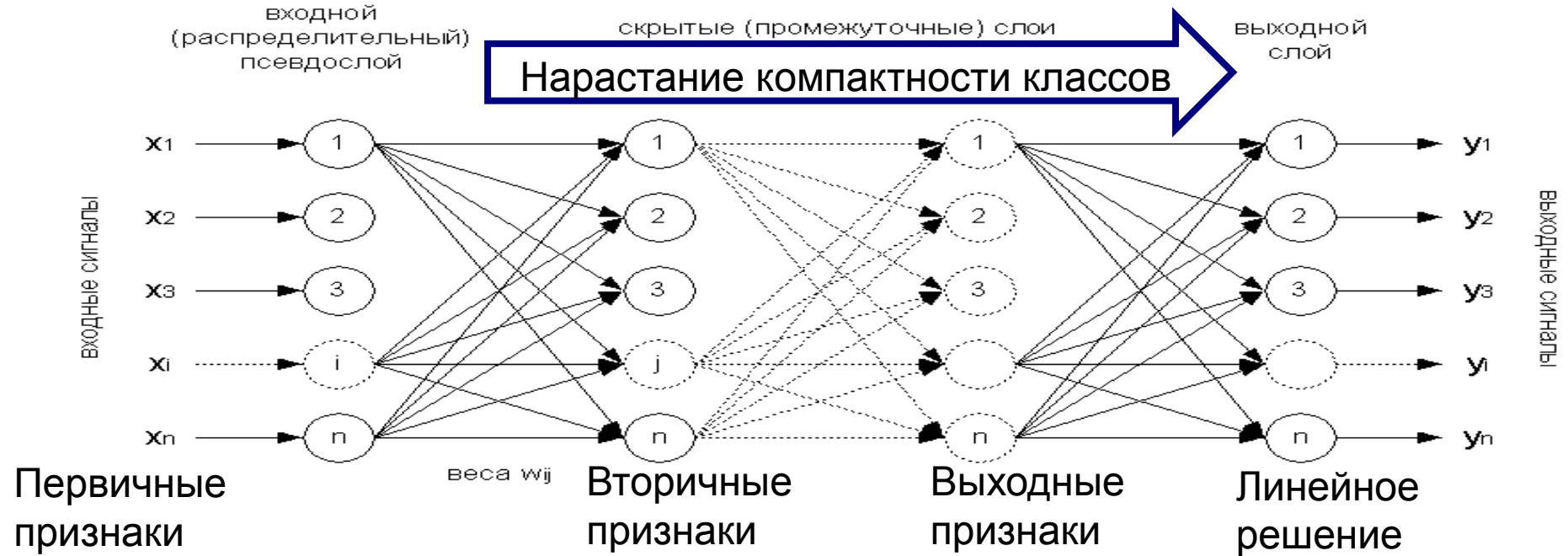


### Зачем нам много слоев нейронов?

**Зачем нелинейности между слоями?** Затем, чтобы многослойный персептрон нельзя было заменить эквивалентным однослойным, т.е. чтобы с ростом слоев росла сложность классификатора. **А зачем увеличивать сложность?....**

# Нейроны и нейронные сети

## Многослойный персептрон



компактный класс

локально компактные классы

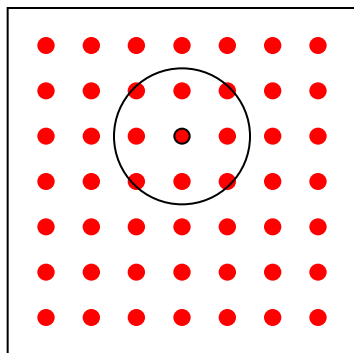
некомпактные классы

От слоя к слою происходит постепенное построение хорошего для данной задачи распознавания пространства признаков. Персептрон сам формирует такие выходные признаки, по которым классы легко разделяются линейными классификаторами.

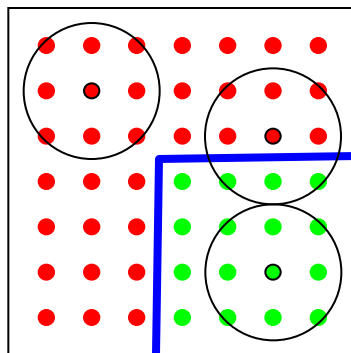
# Почему могут работать линейные классификаторы на выходе?

**Принцип компактности:** более близкие объекты должны с большей вероятностью принадлежать к одному классу.

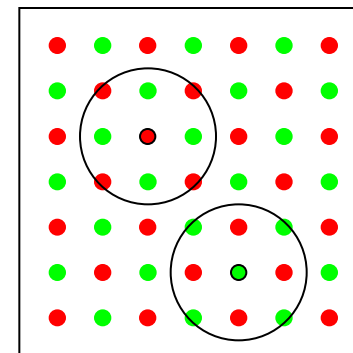
- Айзерман М. А., Браверман Э. М., Розоноэр Л. И. *Метод потенциальных функций в теории обучения машин*. М.: Наука, 1970. 320 pp.
- Хачай М. Ю. *Топологический подход к выводу условий равномерной по классу событий сходимости частот к вероятностям*. // Интеллектуализация обработки информации: 8-я международная конференция (ИОИ-8), Кипр, г.Пафос, 2010 г.: Сборник докладов. – М.: МАКС Пресс, 2010, с.91-94.



компактный класс



локально компактные классы



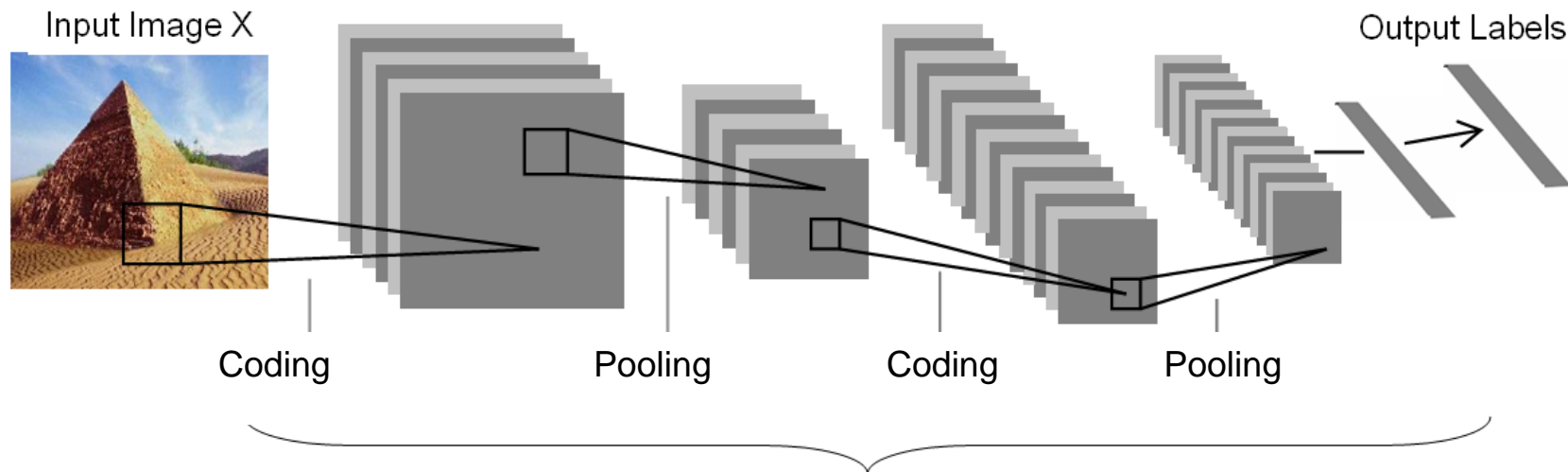
некомпактные классы

**Свойством компактности обладает хорошее для данной задачи распознавания пространство признаков. Это не свойство конечного классификатора, а свойство пространства признаков. Если признаки выбраны (сформированы) удачно, то классы будут легко разделяться простыми классификаторами. Отсюда возникает задача формирования пространства признаков. Раньше это делал человек, теперь - сама CNN.**



# Convolution networks, Deep learning, Image Recognition

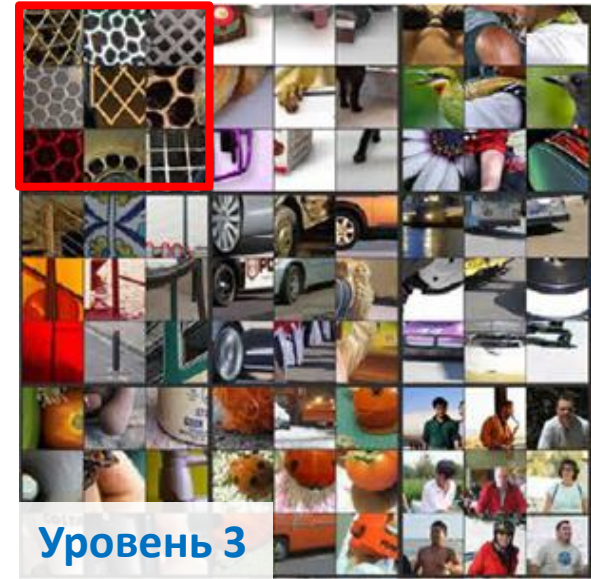
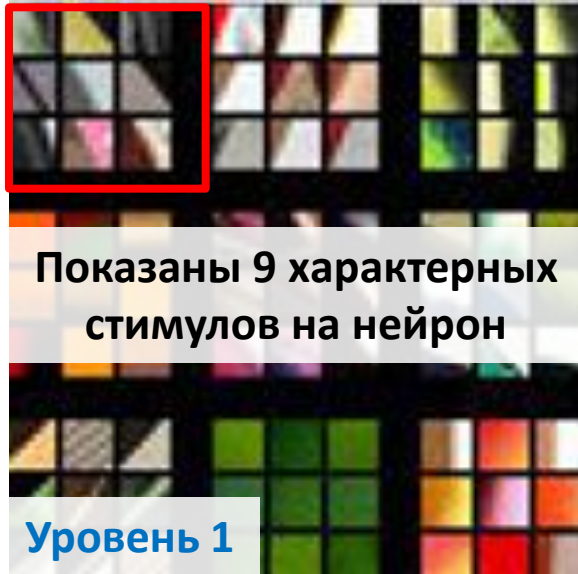
## Сверточные сети, глубокое обучение – современная реинкарнация перцептронов



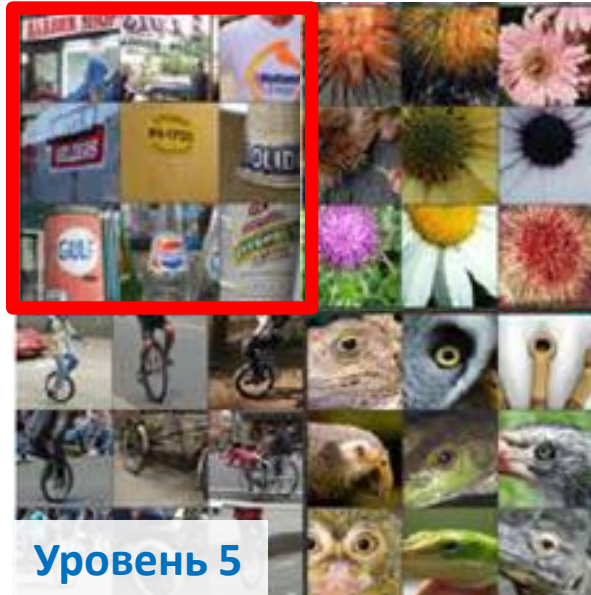
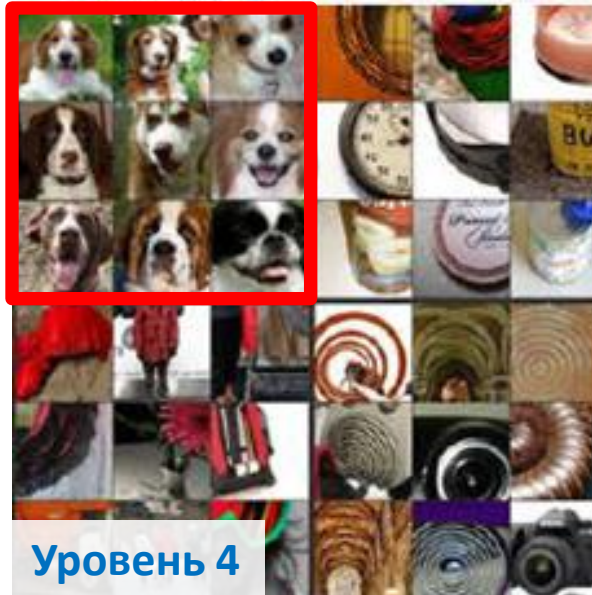
Архитектура “Coding + Pooling” = “Кодирование + Объединение”

- Иерархическое выделение **локальных** признаков при обучении и/или самообучении на сверхбольших выборках изображений
- Много десятков уровней (кодирование + объединение данных)
- **Инвариантность к сдвигу** и масштабу изображения (**свертка**)
- На верхних уровнях – собственно классификатор (полносвязный)

# Convolution networks, Deep learning, Image Recognition



На какие элементы изображения реагируют нейроны разных уровней: **чем выше слой сети, тем выше уровень абстракции**

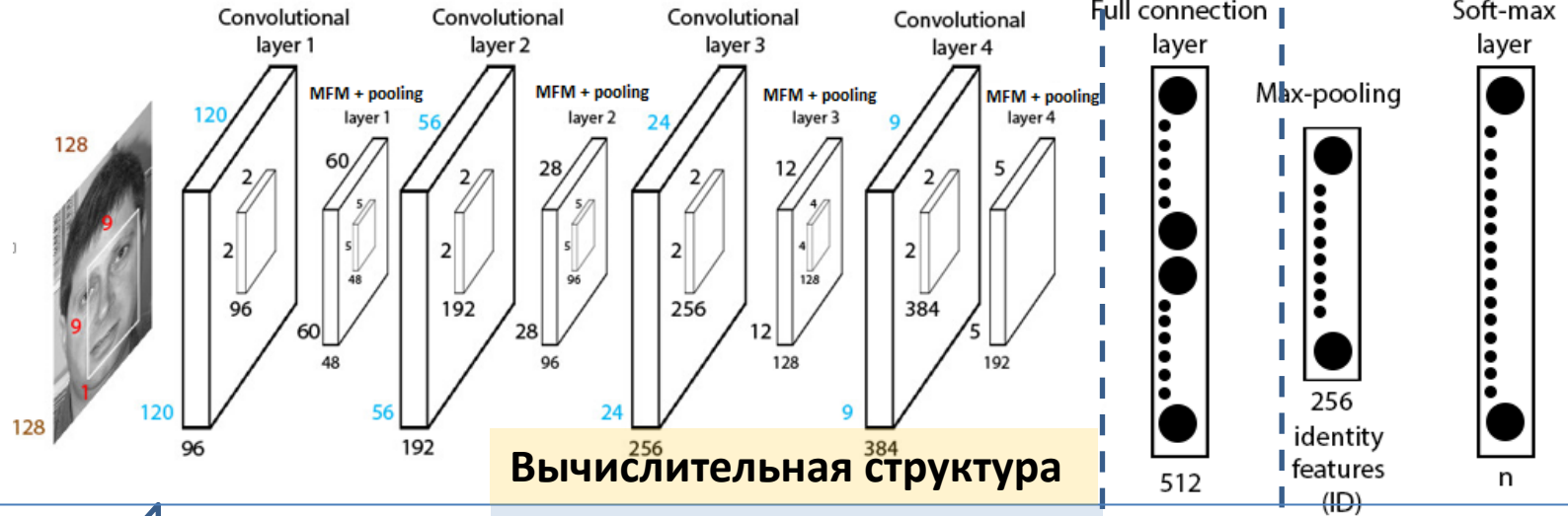


*Структурно-  
функциональный анализ  
ГКНС*



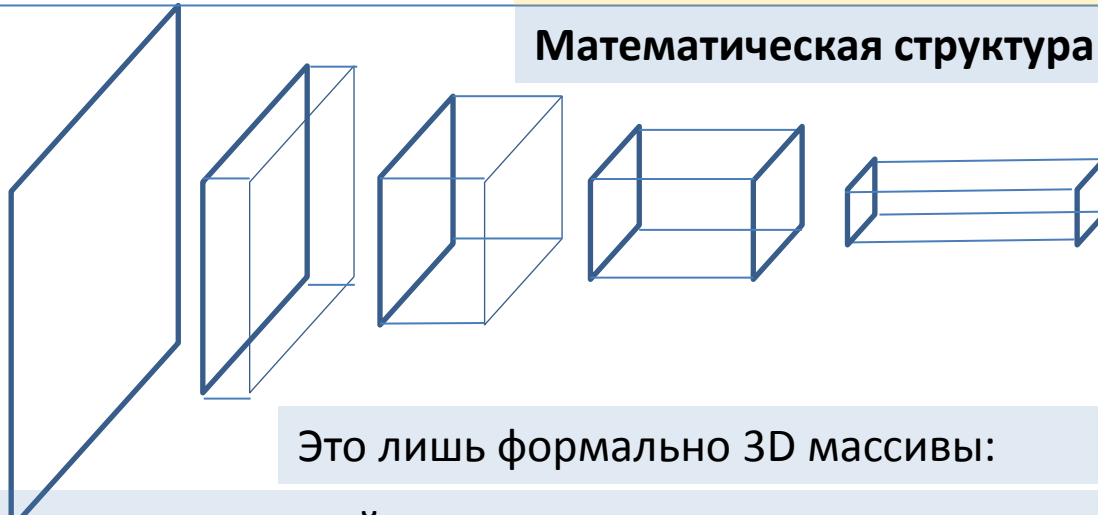
# Интерпретация CNN как процесса обработки массивов (тензоров) – модель TensorFlow

Конволюционная часть CNN работает с 3D массивами (тензорами)



Вычислительная структура

Математическая структура

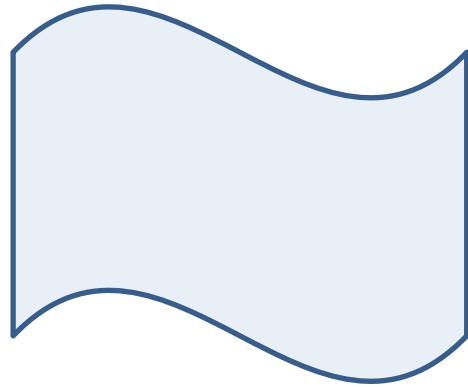


Это лишь формально 3D массивы:

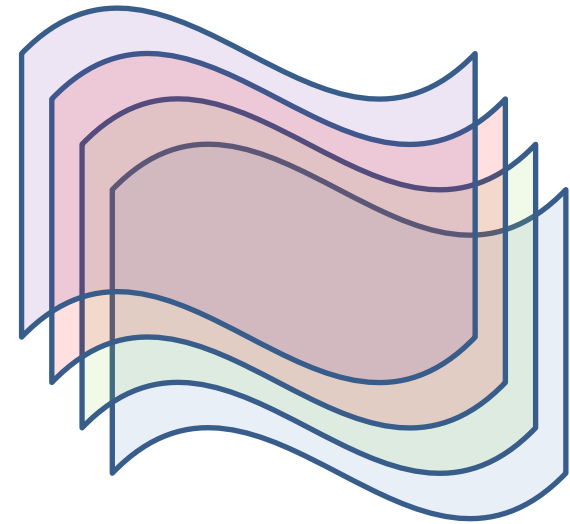
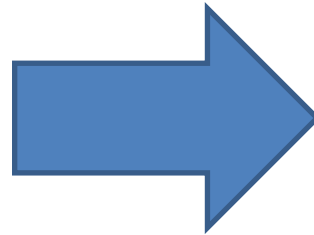
У структурных измерений - геометрия и топология данных,  
У признаковых измерений – евклидово пространство

Полносвязная часть CNN (перцептрон) работает с векторами

# Многообразия, расслоения и векторные поля



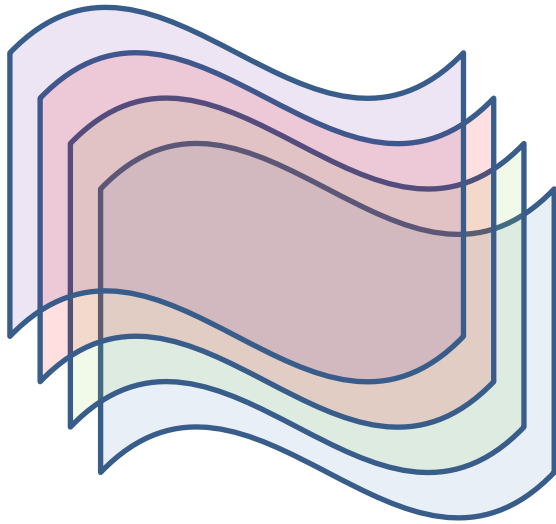
Многообразие  
со своей структурой  
(внутренней  
геометрией и  
топологией)



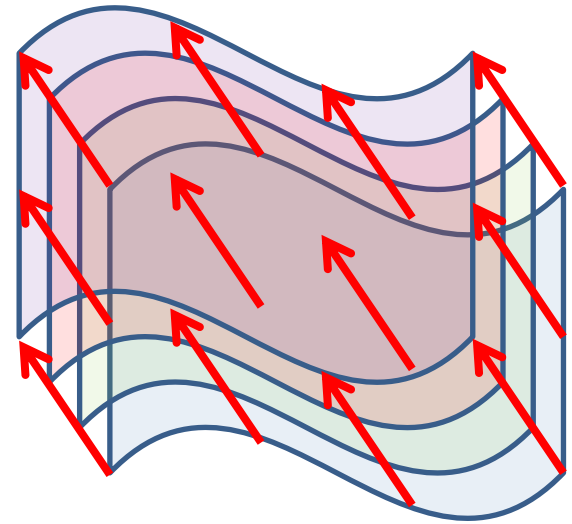
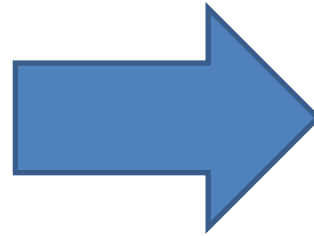
Расслоение  
многообразия как пачка  
одинаковых  
многообразий  
с заданными на них  
различными  
скалярными функциями

Хотя на практике мы имеем дело с дискретными изображениями, в теории ничто не мешает описывать их как непрерывные, как это делается, например, в математической морфологии...

# Многообразия, расслоения и векторные поля



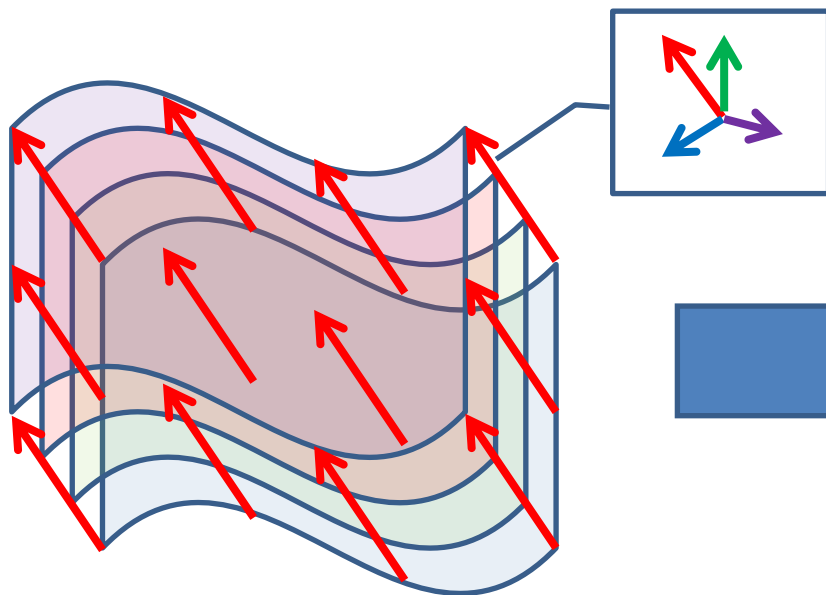
Расслоение  
многообразия как пачка  
одинаковых  
многообразий  
с заданными на них  
различными  
скалярными функциями



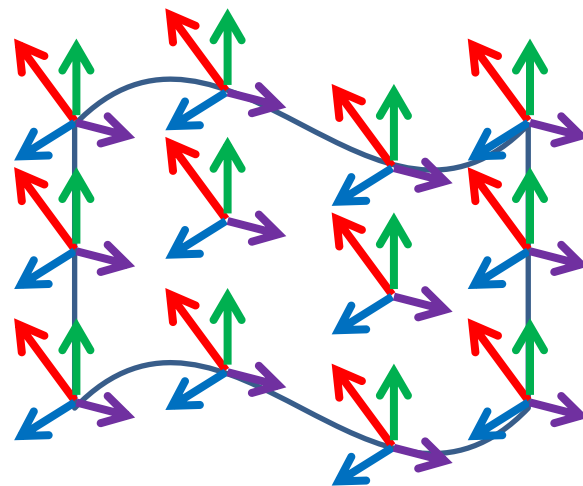
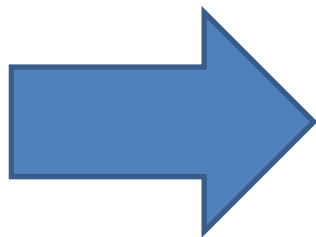
Расслоение  
многообразия как  
многообразие с  
заданной на нем  
векторной функцией  
(аналог векторного поля  
в физике)



# Многообразия, расслоения и векторные поля

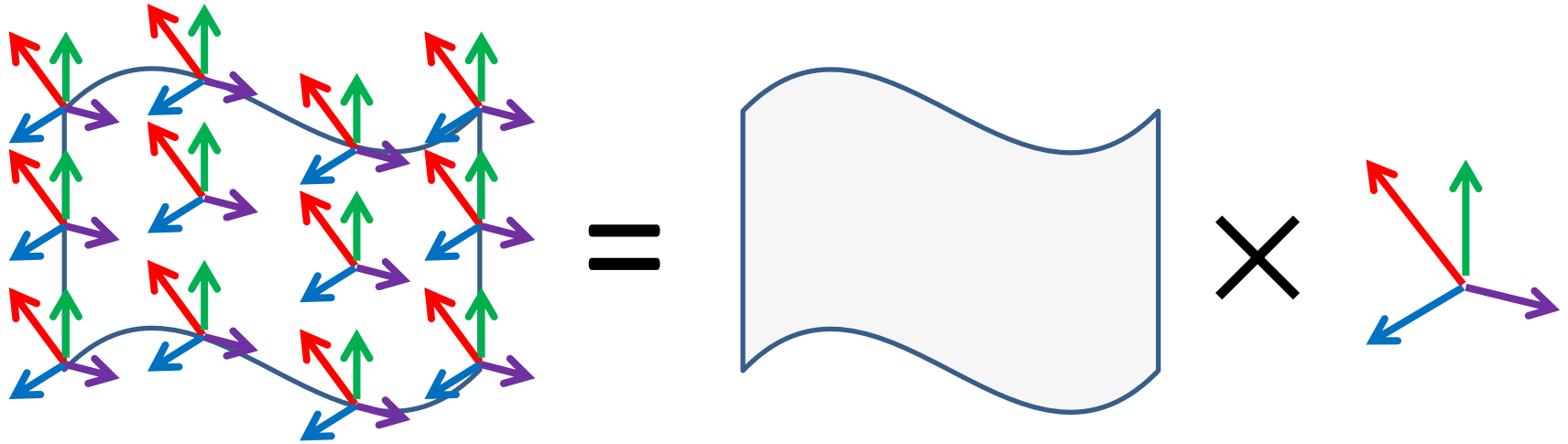


Расслоение  
многообразия как  
многообразиие с  
заданной на нем  
векторной функцией  
(аналог векторного поля  
в физике)



Расслоение  
многообразия как  
многообразиие,  
в каждой точке которого  
привязано векторное  
пространство

# Многообразия, расслоения и векторные поля

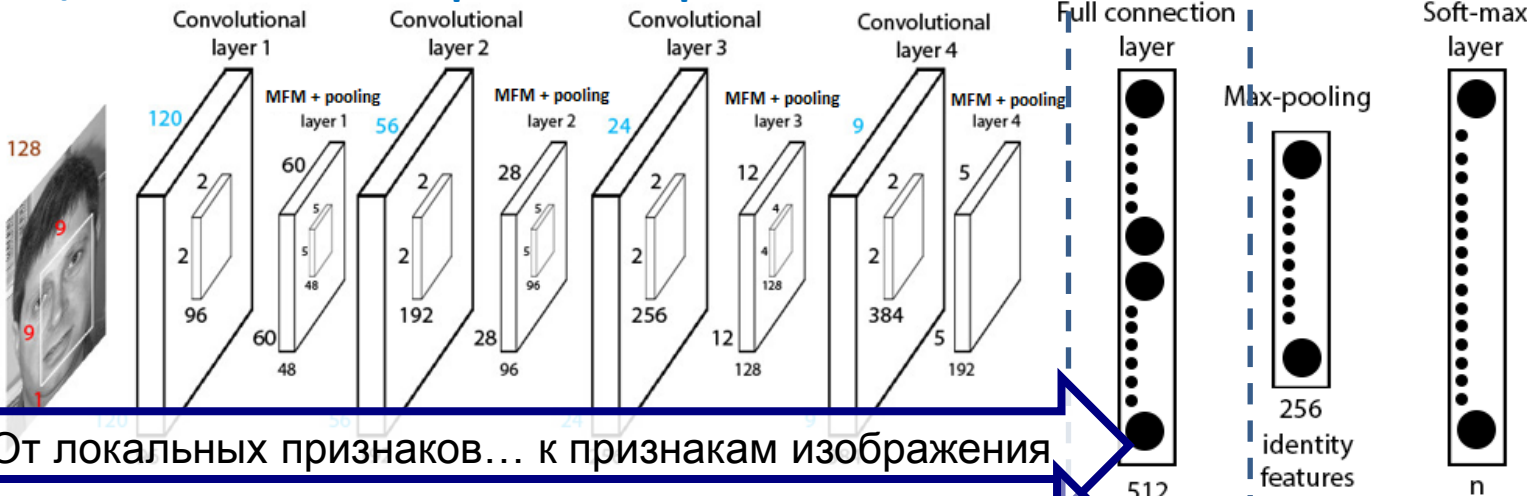


Расслоение  
многообразия как  
многообразиие,  
в каждой точке которого  
привязано векторное  
пространство

Расслоение  
многообразия  
как прямое  
произведение  
многообразия на  
векторное пространство

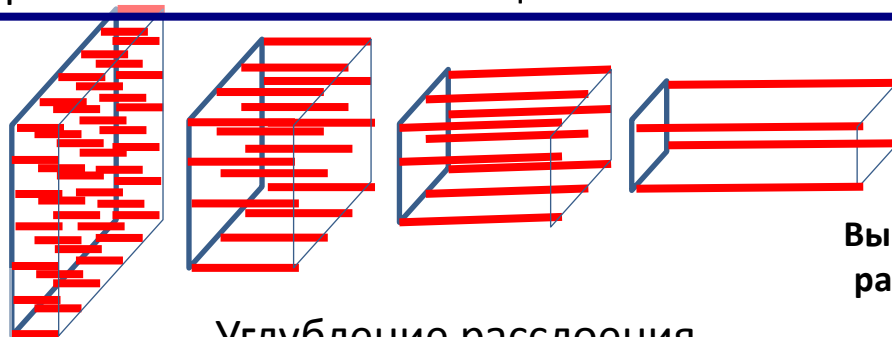
# Интерпретация CNN как процесса эволюции расслоения базового многообразия

Конволюционная часть CNN работает с расслоениями!



От локальных признаков... к признакам изображения

Нарастание компактности целевых классов



Вырожденное расслоение =  
= Базовое многообразие

Углубление расслоения

Компактификация (сжатие)  
базового многообразия

Вырожденное расслоение =  
= вектор признаков

DeepID

Снижение размерности  
вектора признаков

А это уже не  
признаки,  
а классы

Полносвязная часть CNN  
(перцептрон) работает  
с векторами

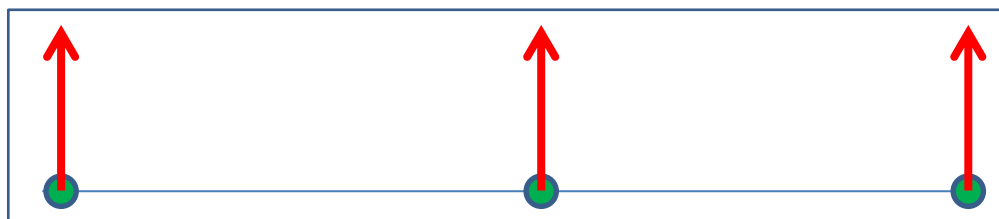
Глубина вектора признаков = размерности векторного подпространства



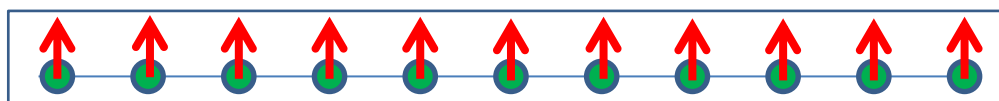
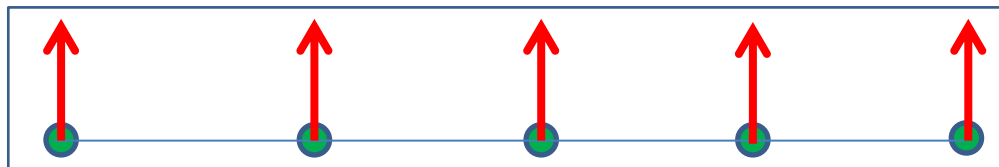
# Компактификация входного многообразия

Размер (точнее мощность или норма) многообразия  $X$  равен количеству его элементов  $\| X \| = \#X$ .

Процесс последовательной «перегонки» информации о геометрии и топологии объекта в глобальный вектор признаков = Процесс компактификации многообразия  $X^{(k)}$  до тех пор, пока на шаге  $k^*$  оно не окажется сжато в один элемент (точку), который уже не имеет структуры



$$X^{(k^*-1)} \times F^{(k^*-1)}$$



$$X^{(1)} \times F^{(1)}$$



$$X^{(0)} \times F^{(0)}$$

$$\| X^{(0)} \| \gg 1, \dim(F^{(0)}) = 1;$$

$$\exists k^*: \| X^{(k^*)} \| = 1, \dim(F^{(k^*)}) \gg 1;$$

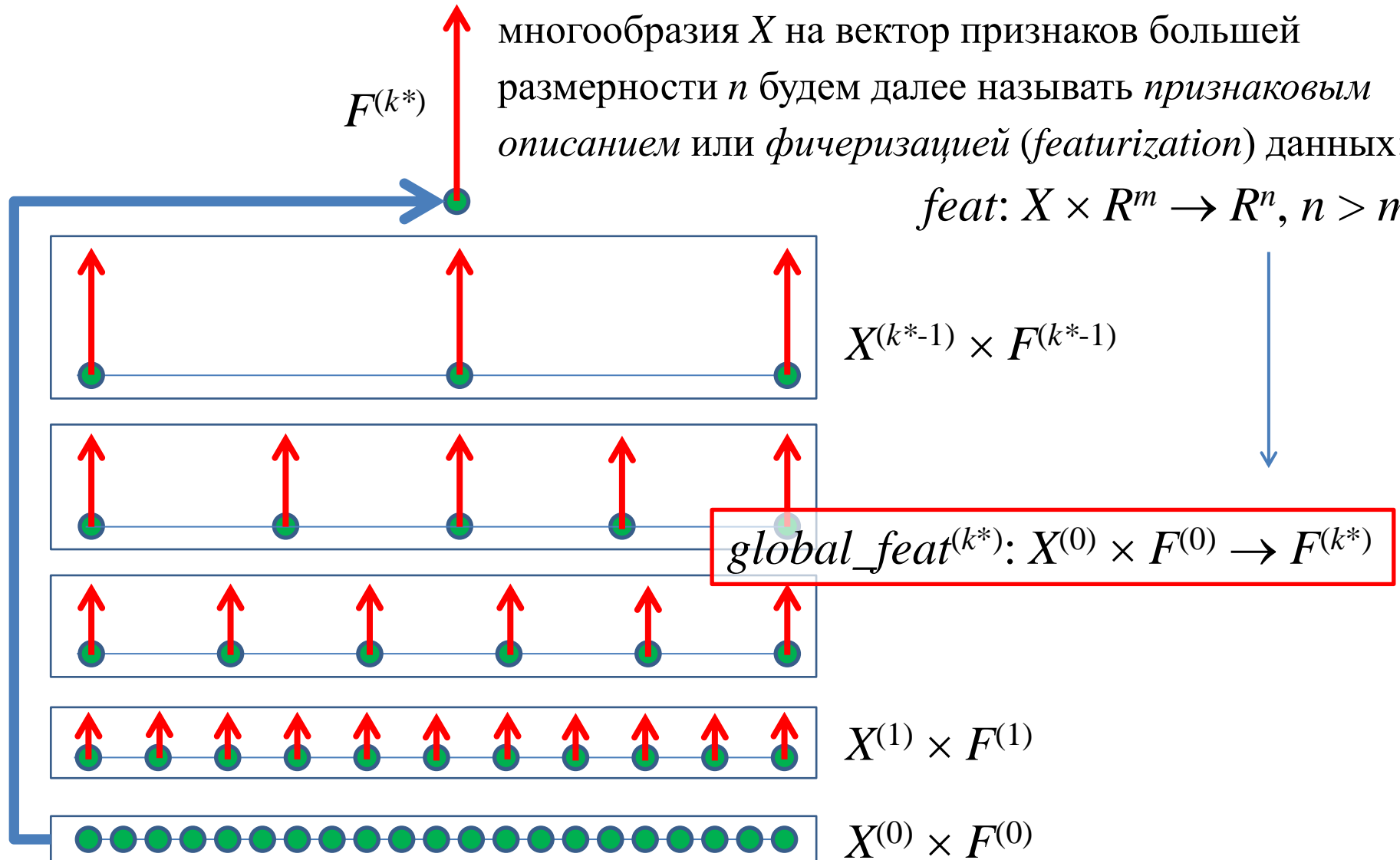
$$\forall k < k^*: \| X^{(k)} \| > \| X^{(k+1)} \|,$$

$$\dim(F^{(k)}) < \dim(F^{(k+1)}).$$

# Глобальная фичеризация входных данных

Отображение  $m$ -расслоения некоторого многообразия  $X$  на вектор признаков большей размерности  $n$  будем далее называть *признаковым описанием* или *фичеризацией (featurization)* данных:

$$\text{feat}: X \times \mathbb{R}^m \rightarrow \mathbb{R}^n, n > m.$$

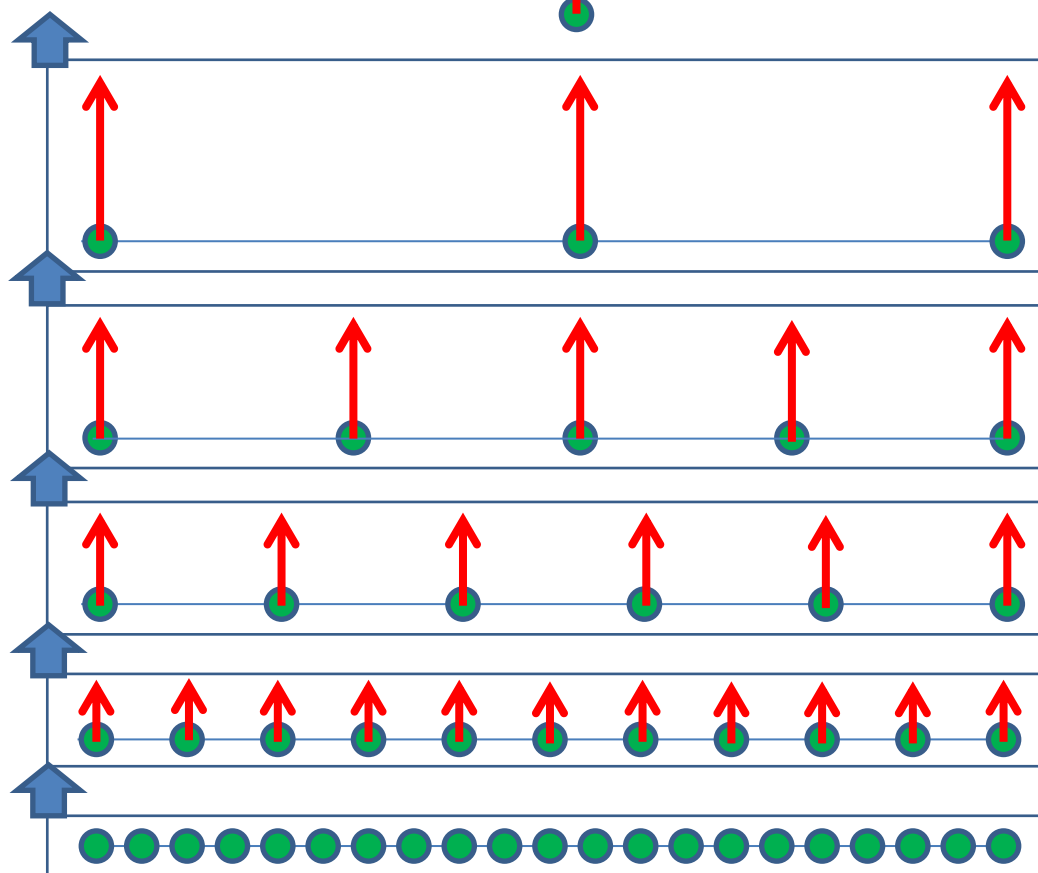


# Итеративное сжатие с частичной фичеризацией

$$part\_shrink\_feat^{(k)}: X^{(k)} \times F^{(k)} \rightarrow X^{(k+1)} \times F^{(k+1)},$$

$$\| X^{(k)} \| > \| X^{(k+1)} \|, \dim(F^{(k)}) < \dim(F^{(k+1)}).$$

$F^{(k^*)}$



$$global\_feat^{(k^*)}: X^{(0)} \times F^{(0)} \rightarrow F^{(k^*)},$$

$$gf^{(k^*)} = psf^{(0)} \circ psf^{(1)} \circ \dots \circ psf^{(k^*-1)}:$$

$$X^{(0)} \times F^{(0)} \rightarrow X^{(1)} \times F^{(1)} \rightarrow \dots$$

$$\rightarrow X^{(k^*-1)} \times F^{(k^*-1)} \rightarrow F^{(k^*)}.$$

$$X^{(k^*-1)} \times F^{(k^*-1)}$$

Шаги сжатия (частичной  
компактификации)  
с частичной фичеризацией

$$X^{(1)} \times F^{(1)}$$

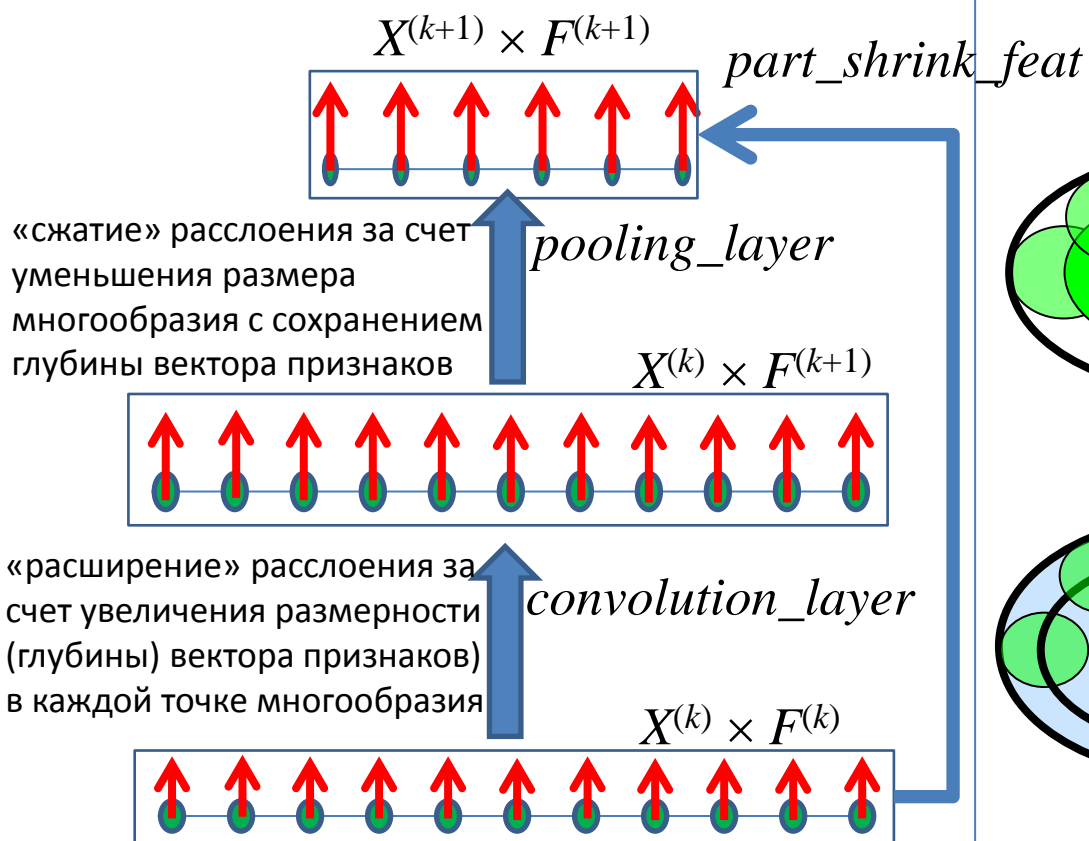
$$X^{(0)} \times F^{(0)}$$



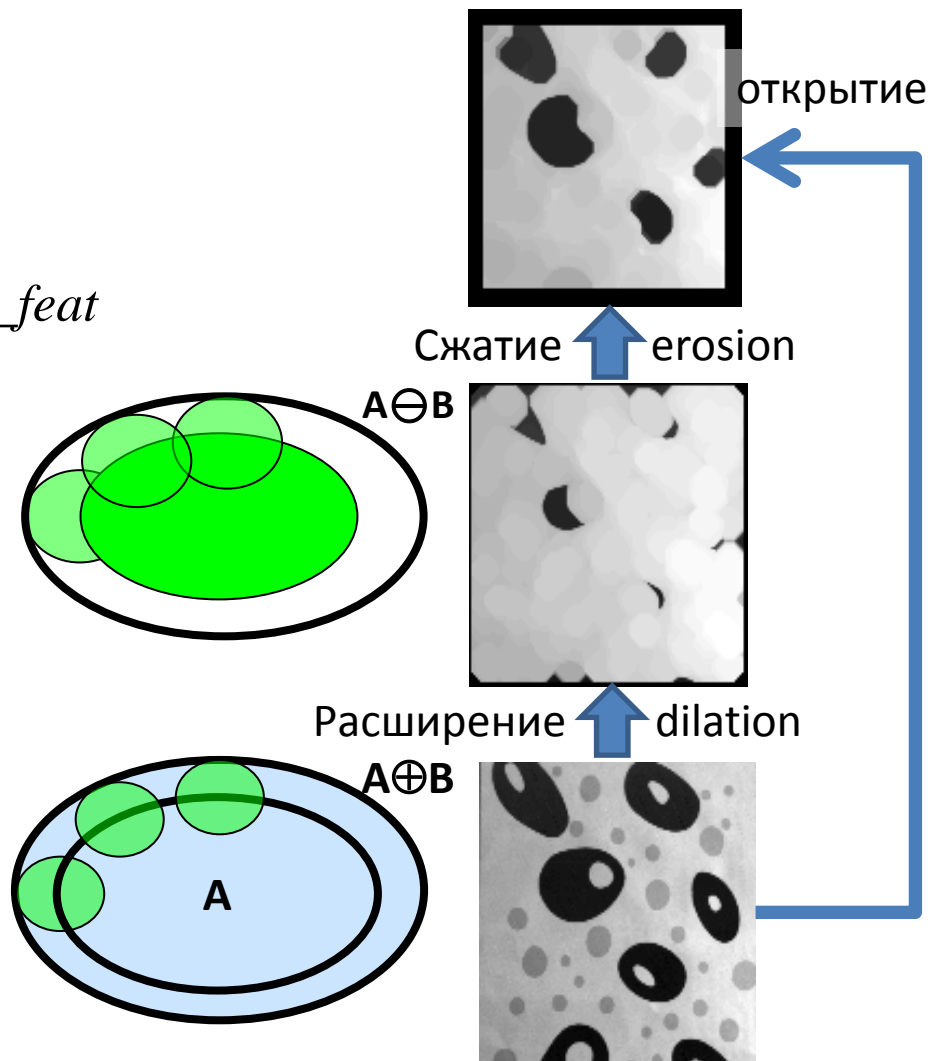
# Сжатие с фичеризацией как модульный оператор (свертка + пулинг)

$$part\_shrink\_feat^{(k)} = convolution\_layer^{(k)} \circ pooling\_layer^{(k)}$$

комбинация операций «расширения» и «сжатия», при которых объем дескриптора данных сначала увеличивается, а затем уменьшается



Аналогия из Математической Морфологии Серра



# Сжатие с фичеризацией как объединение локальных фичеризаций

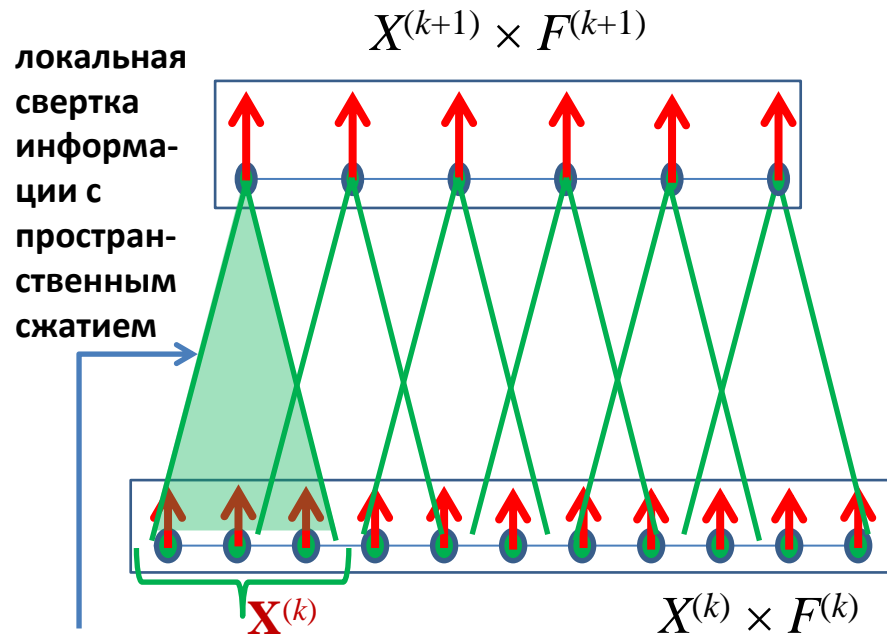
$$part\_shrink\_feat^{(k)} = \cup local\_shrink\_feat^{(k)}(\mathbf{X}^{(k)} \times F^{(k)}),$$

$$local\_shrink\_feat^{(k)}: \mathbf{X}^{(k)} \times F^{(k)} \rightarrow x^{(k+1)} \times F^{(k+1)},$$

$$\mathbf{X}^{(k)} \times F^{(k)} \subseteq X^{(k)} \times F^{(k)}, x^{(k+1)} \in X^{(k+1)},$$

$$\dim(F^{(k)}) < \dim(F^{(k+1)}).$$

комбинация (объединение) операций локальной фичеризации по всем подмножествам заданной формы и размера  $\mathbf{X} \subseteq X^{(k)}$ ,  $\|\mathbf{X}\| = sub\_size(k)$ :

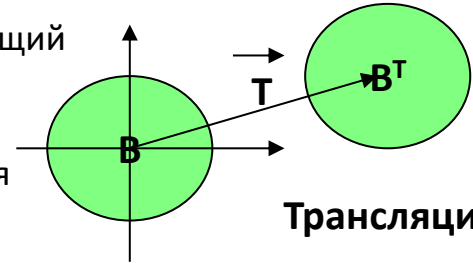


$$local\_shrink\_feat^{(k)}: \mathbf{X}^{(k)} \times F^{(k)} \rightarrow x^{(k+1)} \times F^{(k+1)},$$

$$\mathbf{X}^{(k)} \times F^{(k)} \subseteq X^{(k)} \times F^{(k)}, x^{(k+1)} \in X^{(k+1)}, \dim(F^{(k)}) < \dim(F^{(k+1)}).$$

Аналогия из Математической Морфологии Серра

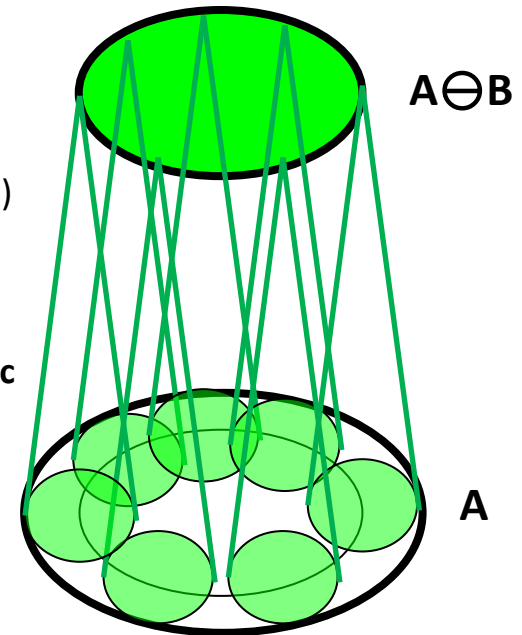
Структурирующий элемент – область агрегирования информации



Вычитание Минковского (эрозия):

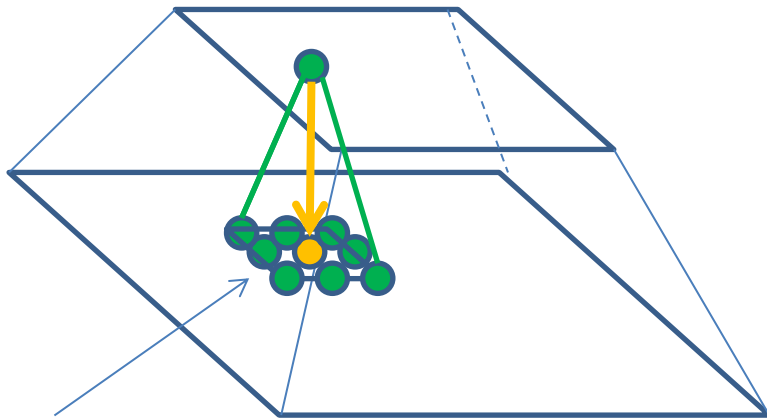
$$A \ominus B = \{z \mid B_z \subseteq A\} = \cup \{A_z\}$$

Все структурирующие элементы, получаемые трансляцией (сдвигом) обрабатываются одинаково (локальная min-свертка информации с пространственным сжатием)



# Выборка локального многообразия (окрестности)

Для регулярных многообразий с геометрией и топологией прямоугольной решетки



Прямоугольное окно заданного размера с центром в заданной точке

Для нерегулярных многообразий с произвольной геометрией и топологией

Результат операции при заданных центре и размере подмногообразия должен однозначно определяться локальной внутренней структурой (геометрией и топологией) входного многообразия

- выборка локального подмногообразия для локального признакового описания

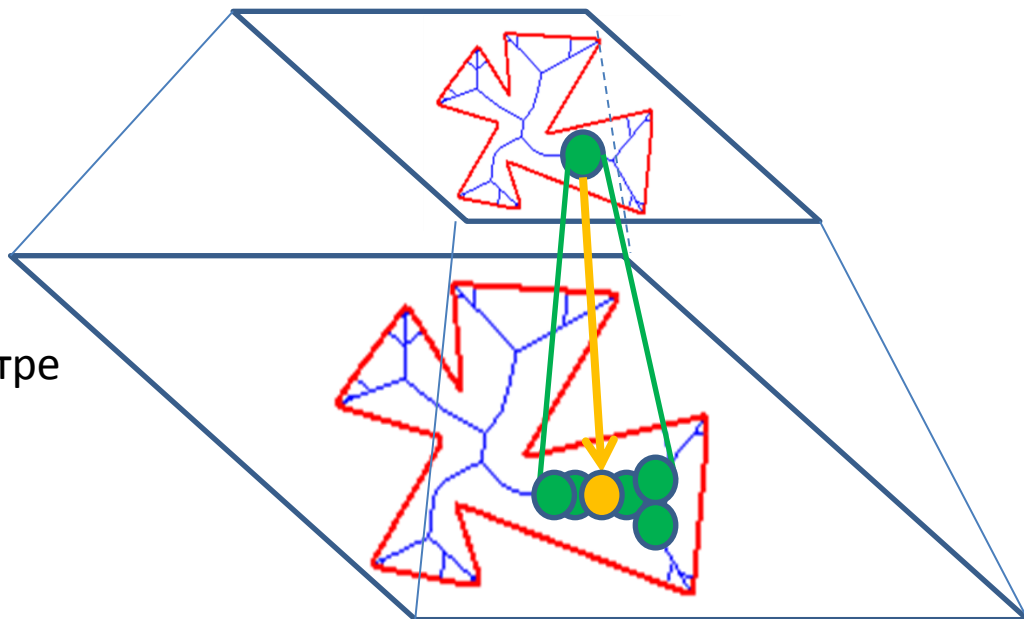
$get\_local\_conv\_submanifold^{(k)}: x \in X^{(k+1)} \rightarrow \mathbf{X}_x^{(k)} \subseteq X^{(k)}, \|\mathbf{X}_x^{(k)}\| = conv\_size(k) = const(x);$

- выборка локального подмногообразия для локального пулинга

$get\_local\_pool\_submanifold^{(k)}:$

$x \in X^{(k+1)} \rightarrow \mathbf{P}_x^{(k)} \subseteq X^{(k)},$

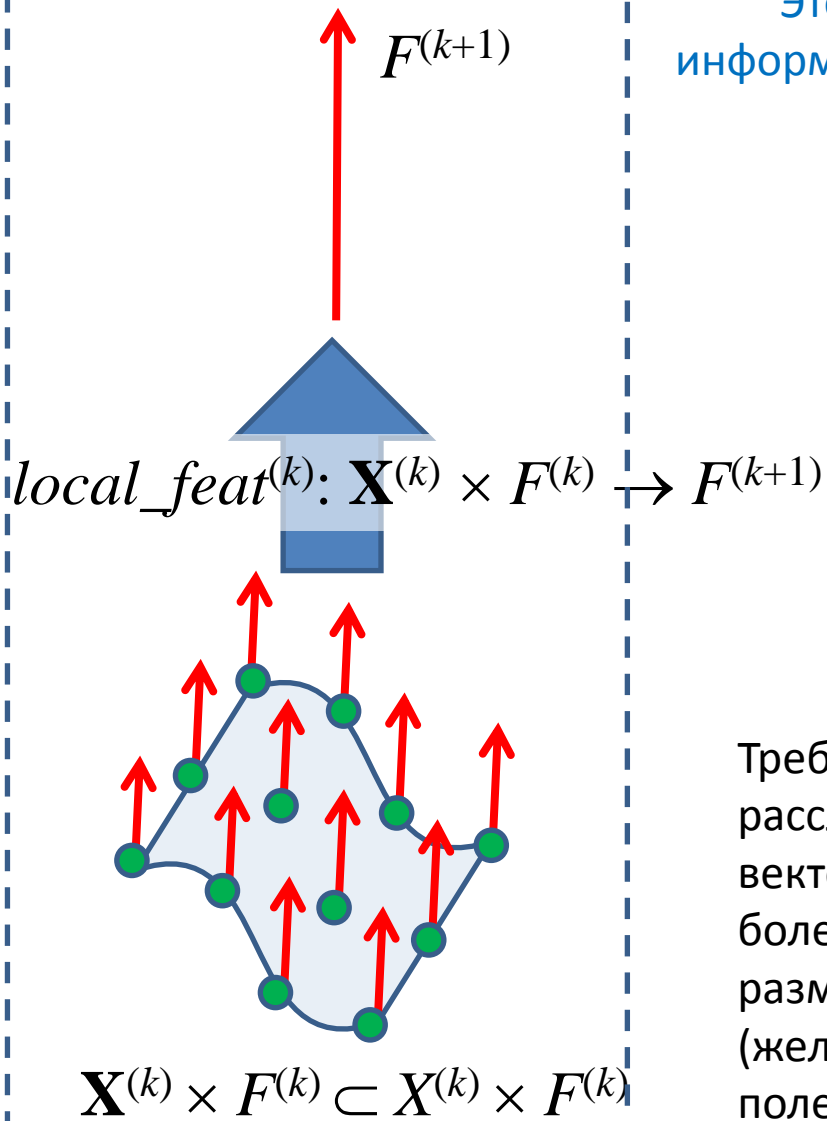
$\|\mathbf{P}_x^{(k)}\| = pool\_size(k) = const(x);$





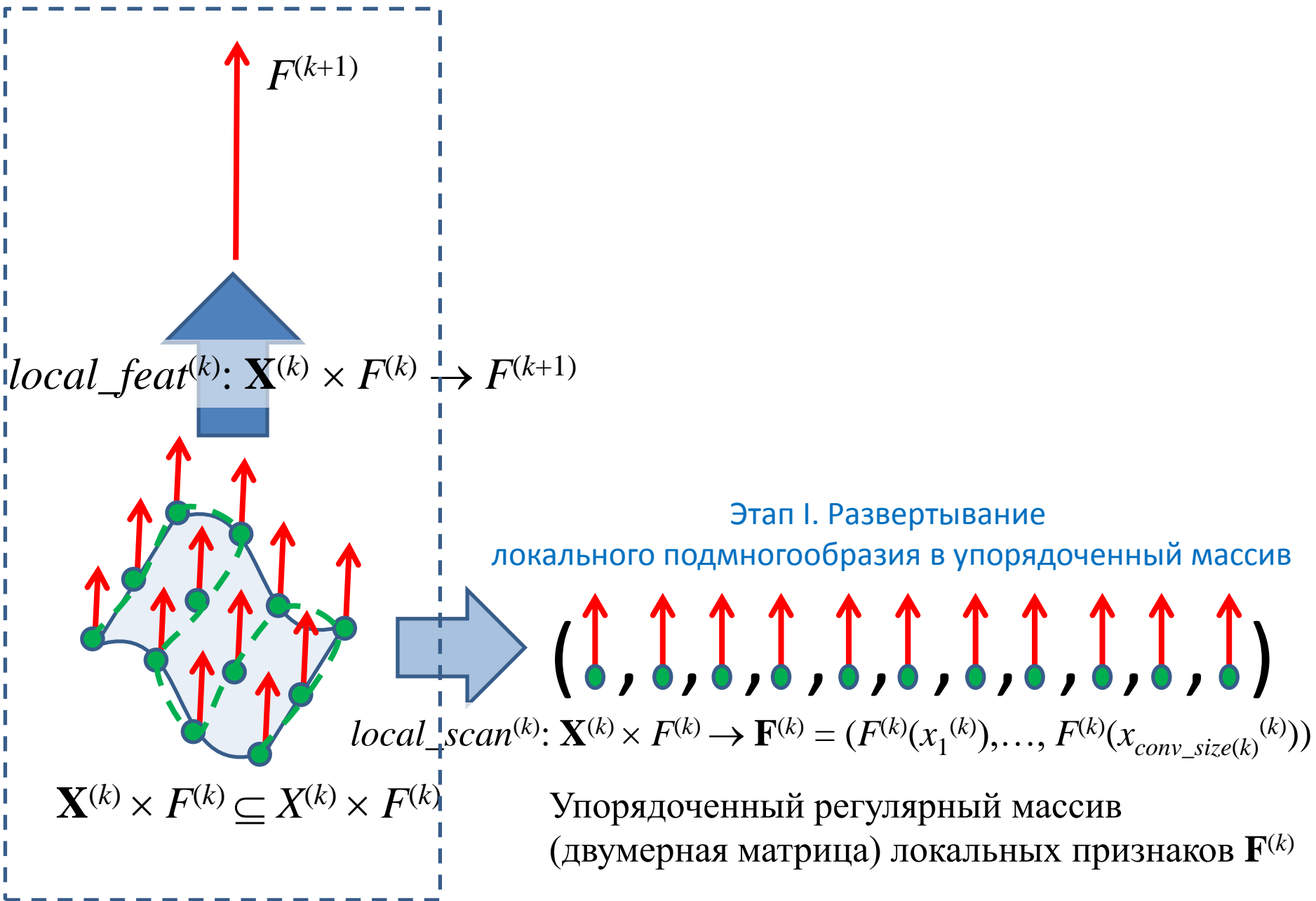
# Модульное представление локальной фичеризации

Это такой способ векторного описания входной информации, при котором результат описания **зависит** от порядка и количества входов

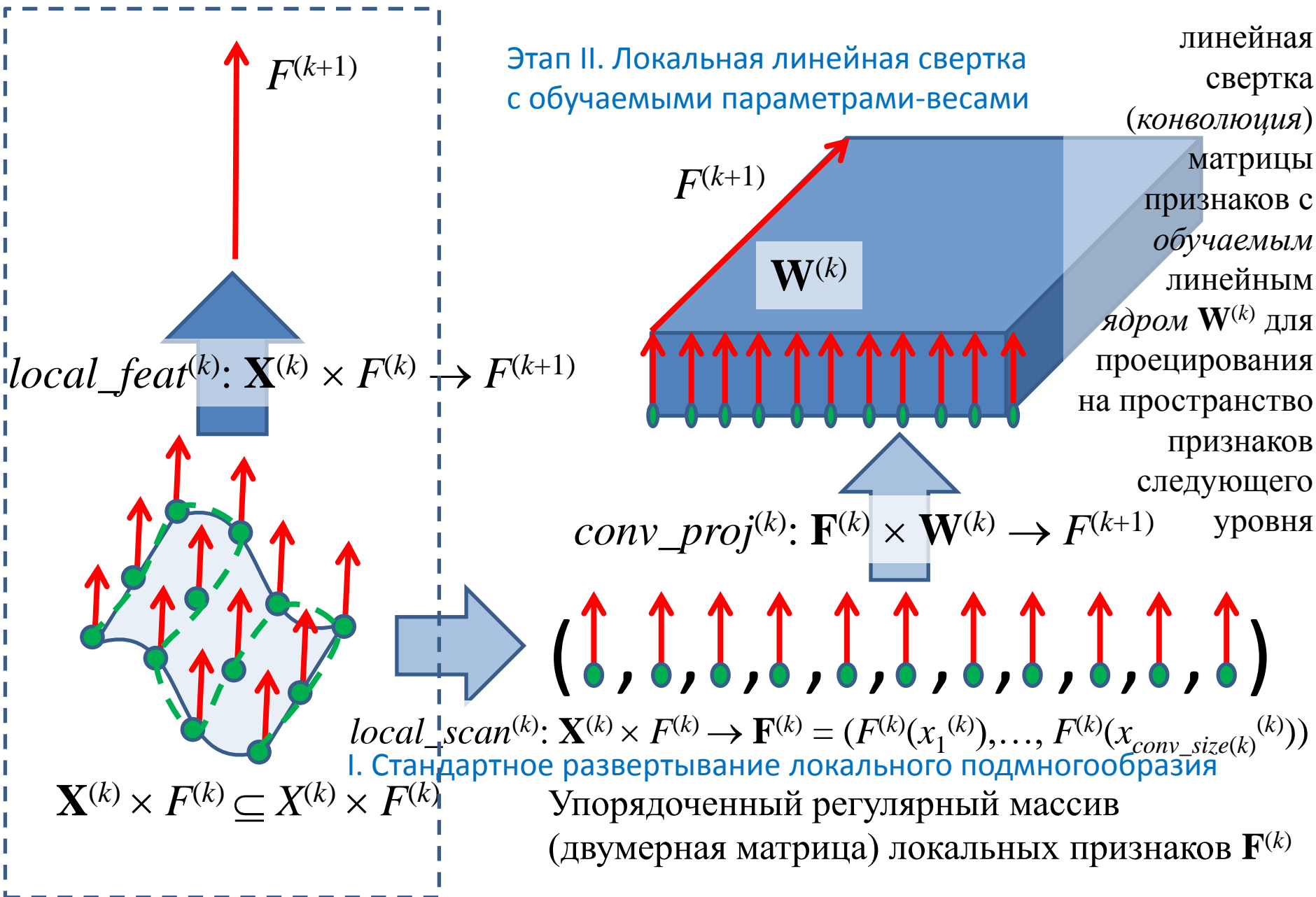


Требуется описать это расслоение единым вектором признаков более высокой размерности (желательно без потери полезной информации)

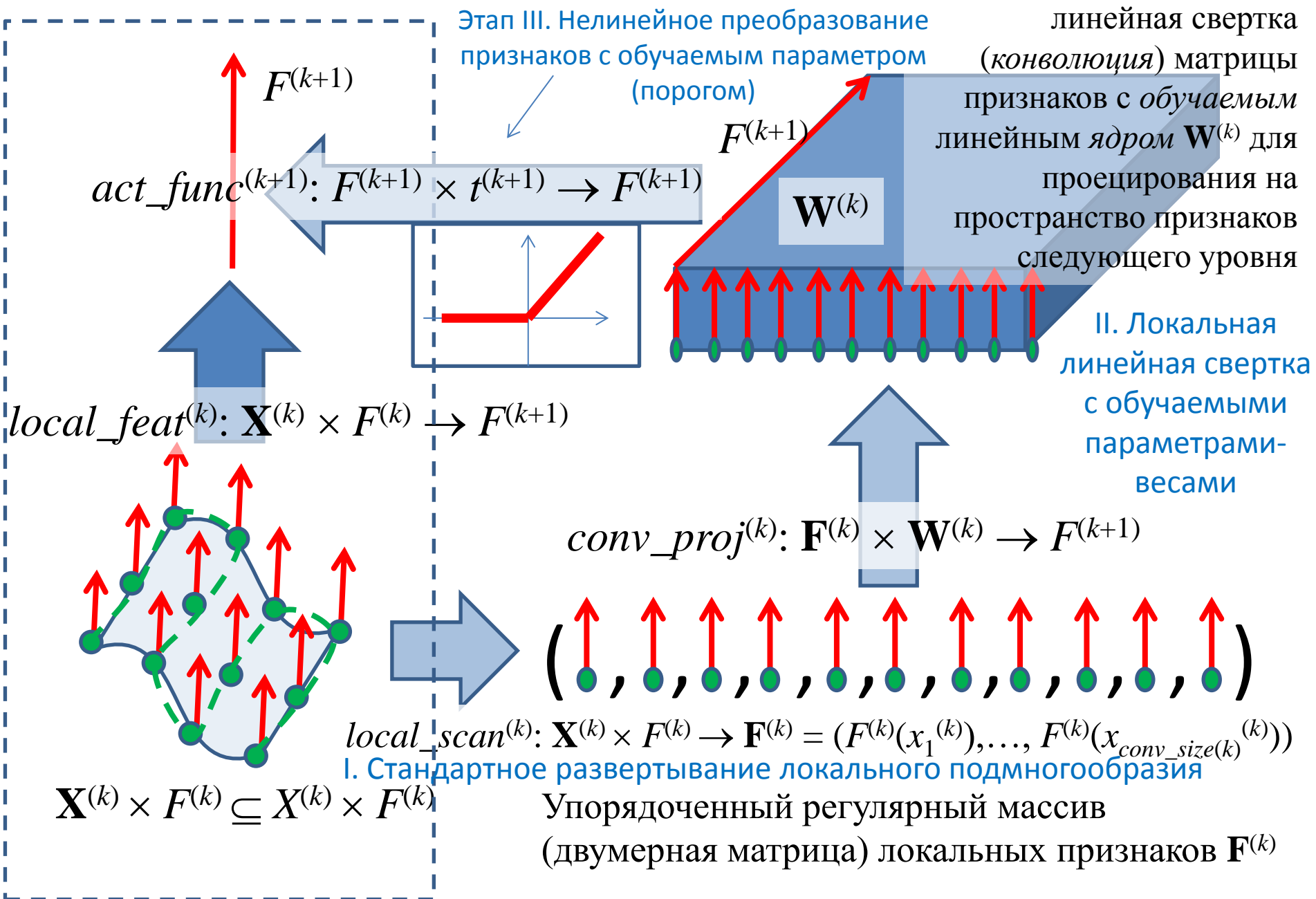
# Модульное представление локальной фичеризации



# Модульное представление локальной фичеризации



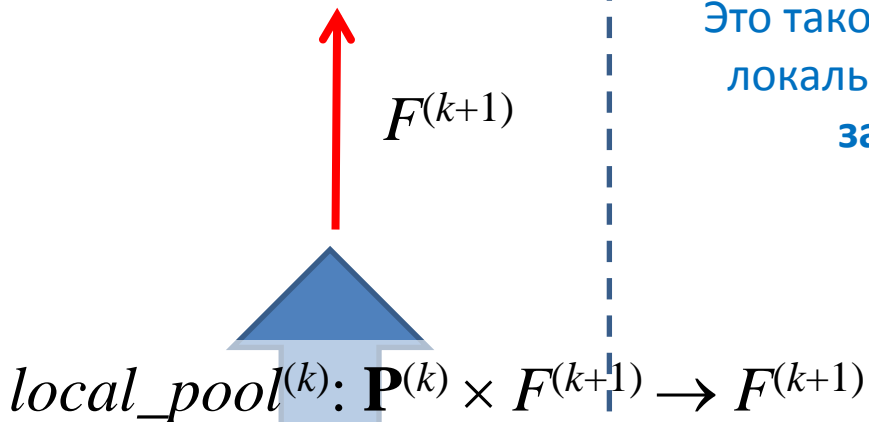
# Модульное представление локальной фичеризации





# Модульное представление локального пулинга

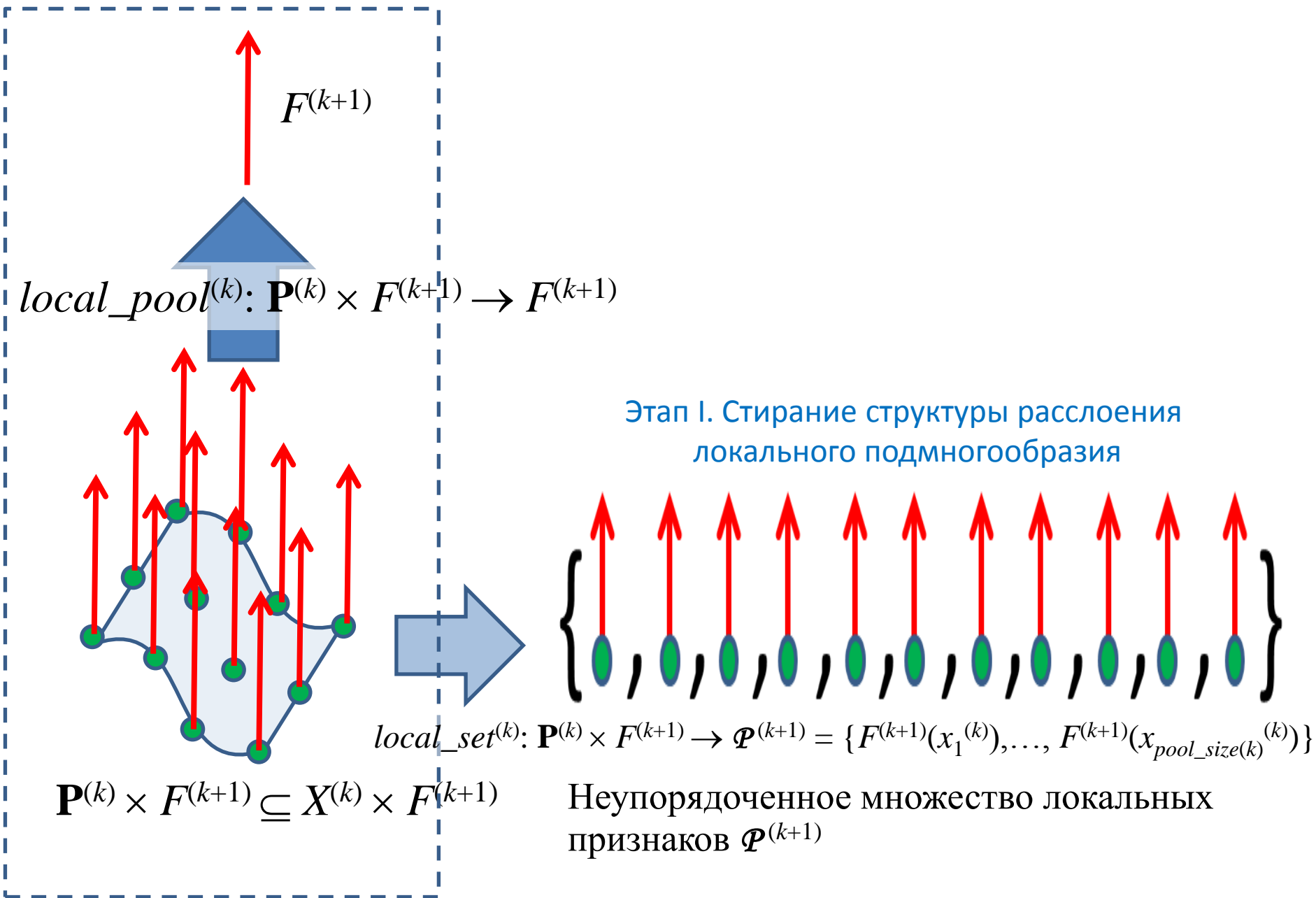
Это такой способ комплексирования (агрегирования) локальной информации, при котором результат **не зависит** от порядка и количества входов


$$local\_pool^{(k)}: \mathbf{P}^{(k)} \times F^{(k+1)} \rightarrow F^{(k+1)}$$

Требуется описать это расслоение единым вектором признаков той же размерности (здесь происходит сжатие информации!)

$$\mathbf{P}^{(k)} \times F^{(k+1)} \subseteq X^{(k)} \times F^{(k+1)}$$

# Модульное представление локального пулинга

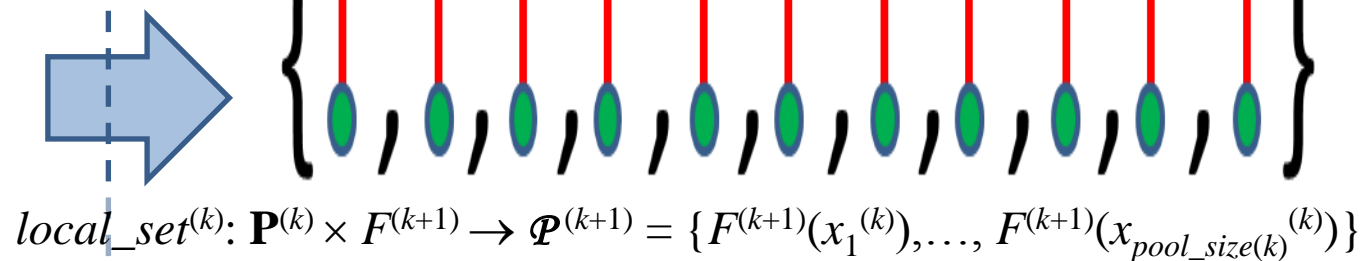
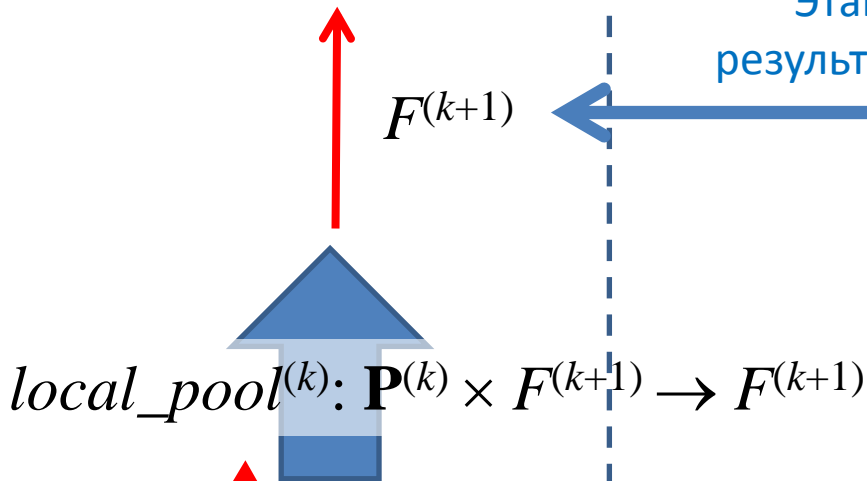


# Модульное представление локального пулинга

Этап II. Агрегирование входной информации:  
результат не зависит от порядка и количества входов

*Агрегирование признаков* входного  
неупорядоченного множества  
(mean, max, min и т.п.)  
 $pool\_feat\_set^{(k+1)}: \mathcal{P}^{(k+1)} \rightarrow F^{(k+1)}$

Этап I. Стирание структуры расслоения  
локального подмногобразия



$$\mathbf{P}^{(k)} \times F^{(k+1)} \subseteq X^{(k)} \times F^{(k+1)}$$

Неупорядоченное множество локальных признаков  $\mathcal{P}^{(k+1)}$

# Структурно-функциональные элементы глубоких CNN

Каждый шаг пространственного сжатия (частичной компактификации) с частичной фичеризацией представлен как комбинация трех функций:

- **функция сжатия** (частичной компактификации) структурированного многообразия;
- **функция выборки** локального подмногообразия для локальной фичеризации;
- **функция фичеризации** (описания вектором признаков) локального подмногообразия.

**Функция локальной фичеризации** представлена в виде комбинации функций:

- **функции выборки** локального подмногообразия для локального признакового описания;
- **функции развертывания** расслоения подмногообразия в упорядоченный регулярный массив локальных признаков;
- **функции линейной свертки** (конволюции) массива признаков с обучаемым линейным ядром для проецирования на пространство признаков следующего уровня;
- **функции нелинейного преобразования** признаков (активации нейрона) с порогом.

**Функция локального пулинга** описана как комбинация одинаковых операций локального пулинга, состоящих из:

- **функции выборки** локального подмногообразия для локального пулинга;
- **функции стирания структуры** расслоения подмногообразия путем его преобразования в неупорядоченное множество локальных признаков;
- **функции агрегирования признаков** входного неупорядоченного множества.



*Принципы и методика  
структурно-  
функционального  
синтеза ГКНС*

# Систематическая таблица СФАС («Таблица Менделеева» для глубоких сетей)

Для практического решения задач структурно-функционального анализа и синтеза CNN на основе СФЭ и правил их комбинации предлагается построить и использовать т.н. **систематическую таблицу СФАС**.

Задача определения общей структуры такой таблицы соответствует задаче **структурно-функционального анализа** глубоких сетей.

Задача целенаправленного поиска нужных ячеек для решения конкретных прикладных задач распознавания различных типов сложных нерегулярных данных является задачей **структурно-функционального синтеза CNN**.

# Систематическая таблица СФАС («Таблица Менделеева» для глубоких сетей)

СФЭ								
part_shrink								
get_local_conv_submanifold								
local_scan								
conv_proj								
act_func								
get_local_pool_submanifold								
local_set								
pool_feat_set								
concat_features								

**Строками систематической таблицы СФАС являются основные СФЭ.**

# Систематическая таблица СФАС («Таблица Менделеева» для глубоких сетей)

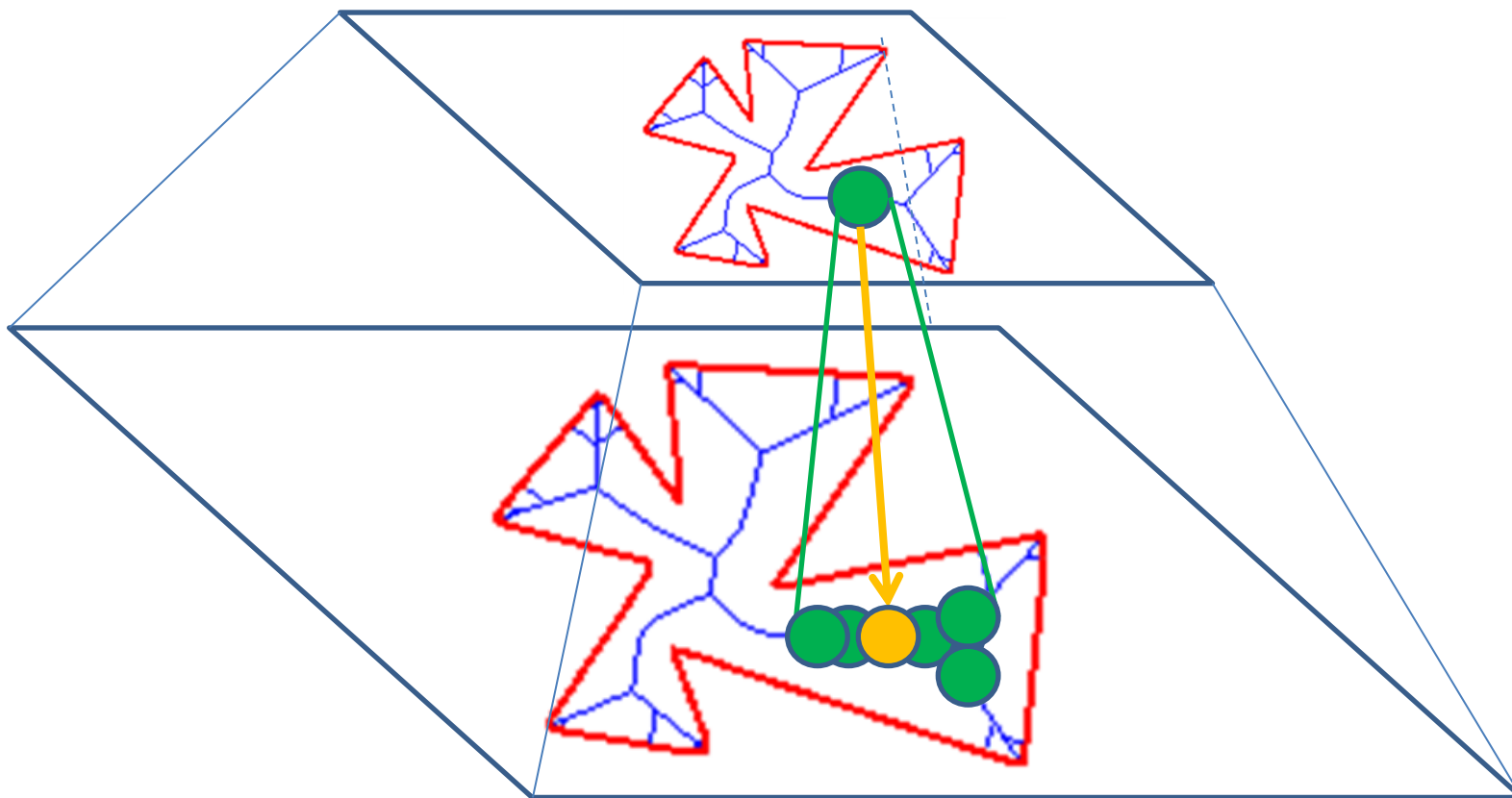
СФЭ	1. 1D, 2D и 3D регулярные массивы			
part_shrink	<ul style="list-style-type: none"><li>• Прямоугольные решетки (классический случай)</li><li>• Шестиугольные и др. регулярные решетки</li></ul>			
get_local_conv_submanifold				
local_scan	2. Скелеты и обогащенные скелеты <i>(интересный частный случай с нетривиальной геометрией и топологией)</i>			
conv_proj				
act_func				
get_local_pool_submanifold	3. Деревья (иерархические деревья и просто графы без циклов)			
local_set	4. Графы и гиперграфы <ul style="list-style-type: none"><li>• Сегментированные изображения</li><li>• Триангулированные 2D и 3D данные</li><li>• Гиперграфы, описывающие марковские модели</li></ul>			
pool_feat_set				
concat_features				

**Строками систематической таблицы СФАС** являются основные СФЭ. **Столбцами систематической таблицы СФАС** являются различные по семантике, геометрии и топологии типы входных данных, для которых необходимо реализовать каждый из СФЭ для того, чтобы построить ГКС и осуществить ее обучение. На пересечении строк и столбцов должны быть описаны способы (алгоритмы) реализации соответствующих СФЭ для соответствующих типов входных данных.



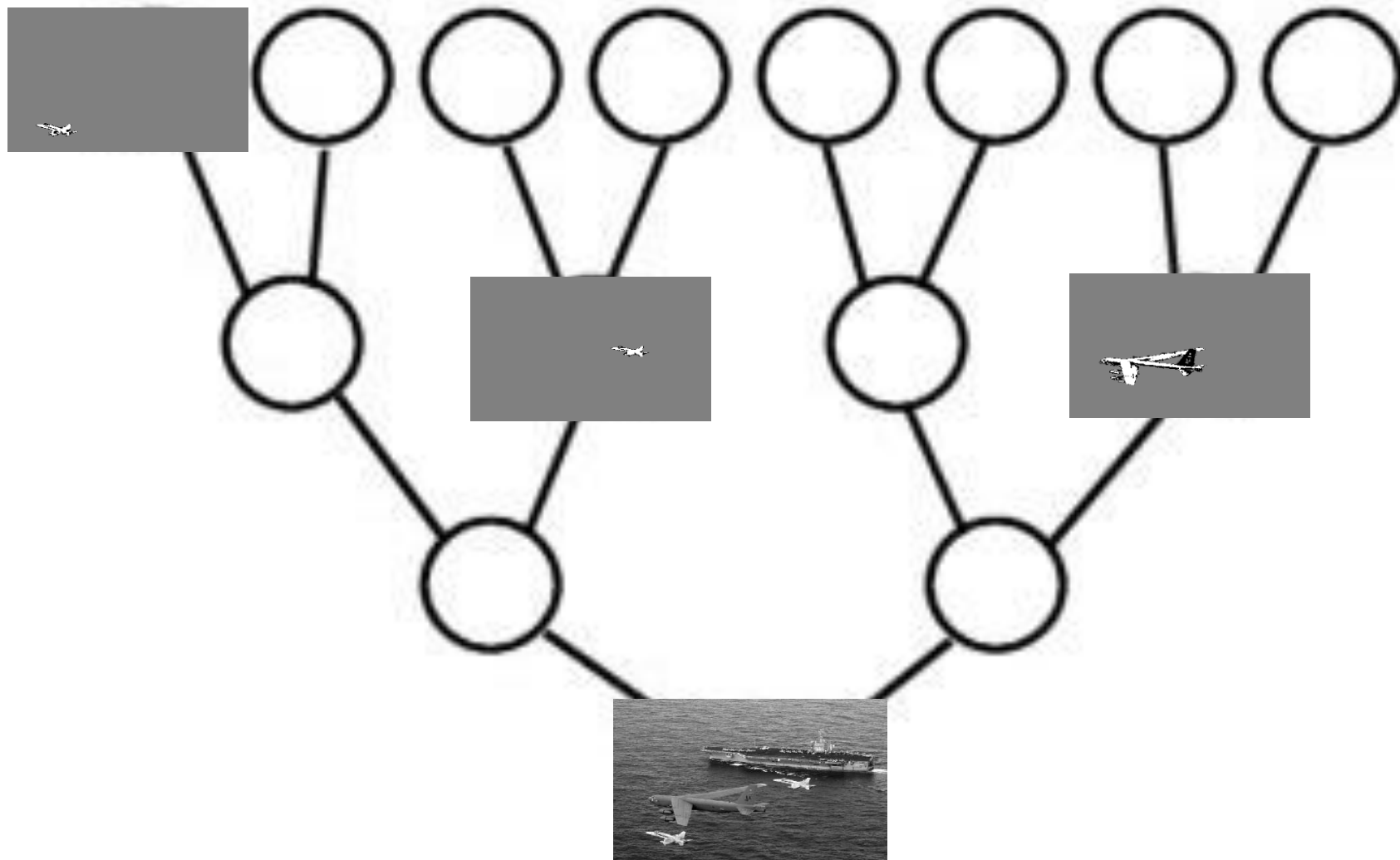
# Пример произведения многообразий:

Скелет, где к каждой точке привязан локальный дескриптор (ГСЭ)



# Пример произведения многообразий:

Дерево морфлетов, где к каждому узлу привязано 2D изображение



*Related works*  
*и приложения*

# Глубокие сети сложной структуры

- **Network in network (NIN)** [1]. В конволюционных слоях вместо конволюционных фильтров используются простые классические нейронные сети состоящие из нескольких слоёв.
- **Модульные сети** (GoogleNet [2], SqueezeNet[3], Inception v4 [4]). В частности, Inception module представляет собой набор из параллельно выполняемых конволюционных слоев и слоев пулинга с различными параметрами и слоя конкатенации объединяющего их результаты, а ГКНС состоит из последовательности таких модулей.
- **Фрактальные сети** [5]. ГКНС формируются как последовательность блоков, каждый из которых построен по принципу фрактальной структуры из классических конволюционных слоев.

---

[1] M. Lin, Q. Chen, and S. Yan. Network in network. arXiv preprint arXiv:1312.4400, 2013.

[2] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1–9, 2015

[3] Forrest N. Iandola and Song Han and Matthew W. Moskewicz and Khalid Ashraf and William J. Dally and Kurt Keutzer SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and  $\leq$  \$0.5MB model size. arXiv preprint arXiv:1602.07360, 2016

[4] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. arXiv preprint arXiv:1602.07261, 2016

[5] Gustav Larsson, Michael Maire, Gregory Shakhnarovich FRACTALNET: ULTRA-DEEP NEURAL NETWORKS WITHOUT RESIDUALS. arXiv preprint arXiv:1605.07648, 2017



# CNN для анализа нерегулярных данных

Для анализа структур **нерегулярных пространственно распределенных входных данных** в настоящее время, как правило, используется **приведение данных к регулярной структуре** с их дальнейшим анализом средствами классических регулярных ГНС, например:

- **3D-данные:** Для анализа облаков точек их представляют в воксельном виде с последующей их обработкой ГНС с трехмерными конволюционными фильтрами [6]
- **2D+T-данные:** Аналогичный подход используется и при обработке треков объектов в видеопоследовательностях [7].

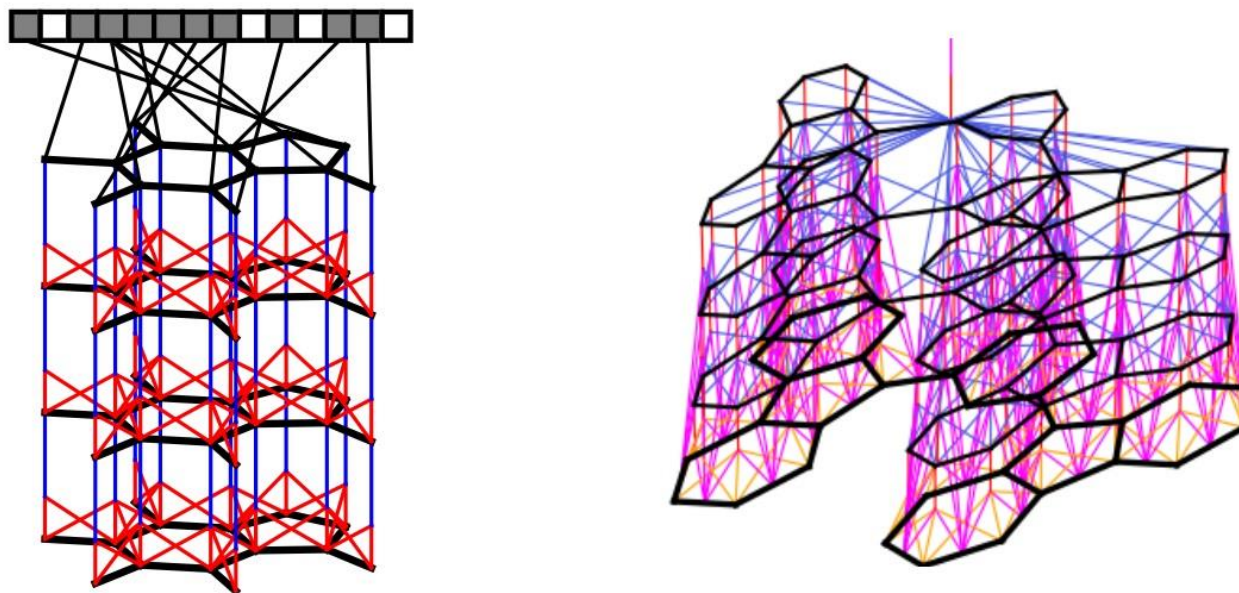
**Однако оснаждающиеся примеры использования сетей с нерегулярной структурой имеются в химии и биологии!**

---

[6] Daniel Maturana, Sebastian Scherer VoxNet: A 3D Convolutional Neural Network for real-time object recognition., 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).

[7] Du Tran, Lubomir Bourdev, Rob Fergus Learning Spatiotemporal Features with 3D Convolutional Networks. , 2015 IEEE International Conference on Computer Vision (ICCV).

# Глубокие нейронные сети в дизайне материалов

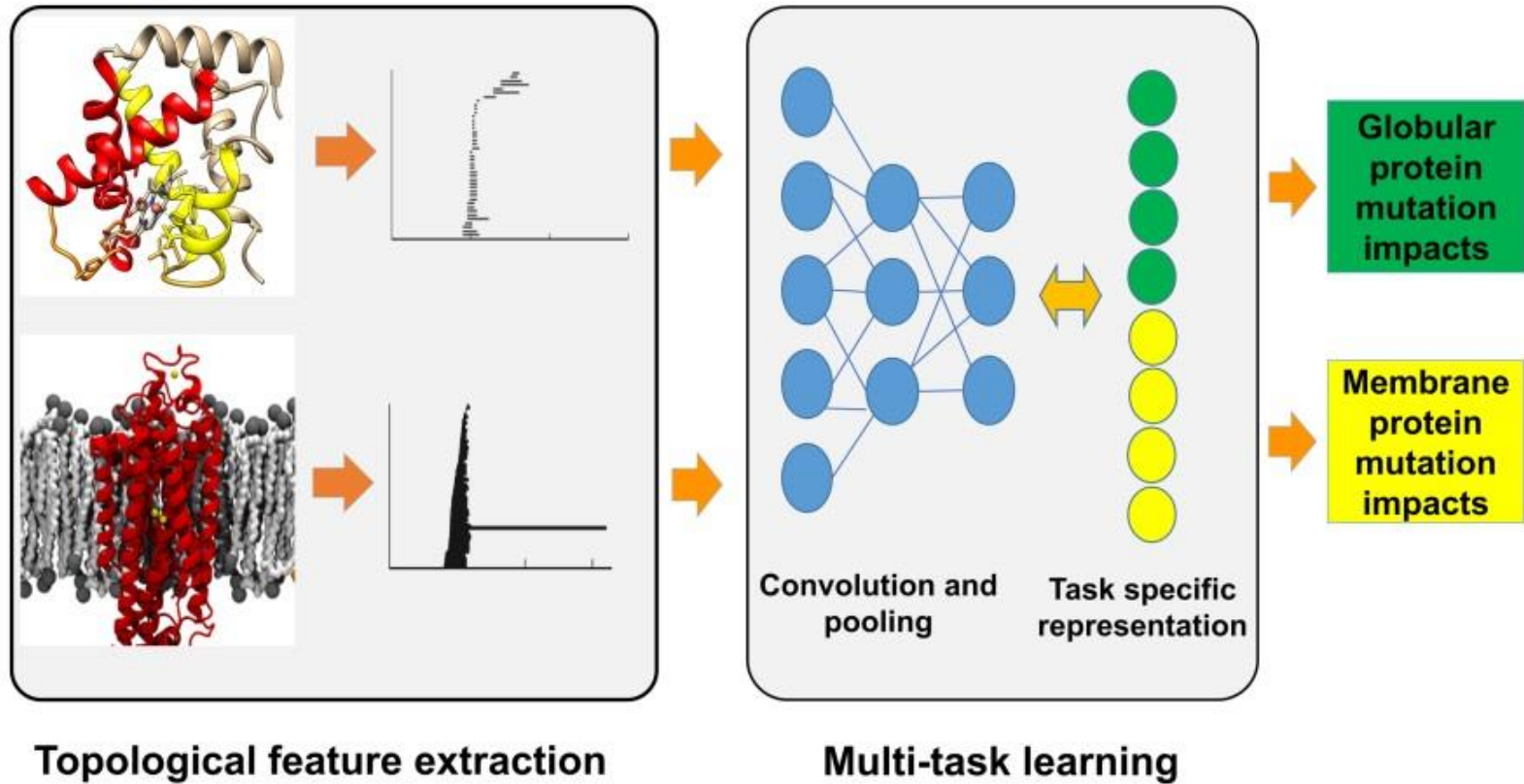


- Методы извлечения молекулярных признаков, с помощью которых выводятся хорошо интерпретируемые молекулярные особенности и достигается увеличение предсказательной способности в области дизайна материалов (например в [8]).
- Исследование и разработка материалов (например в [9]).

[8] Convolutional Networks on Graphs for Learning Molecular Fingerprints. David Duvenaud et al., 2015.

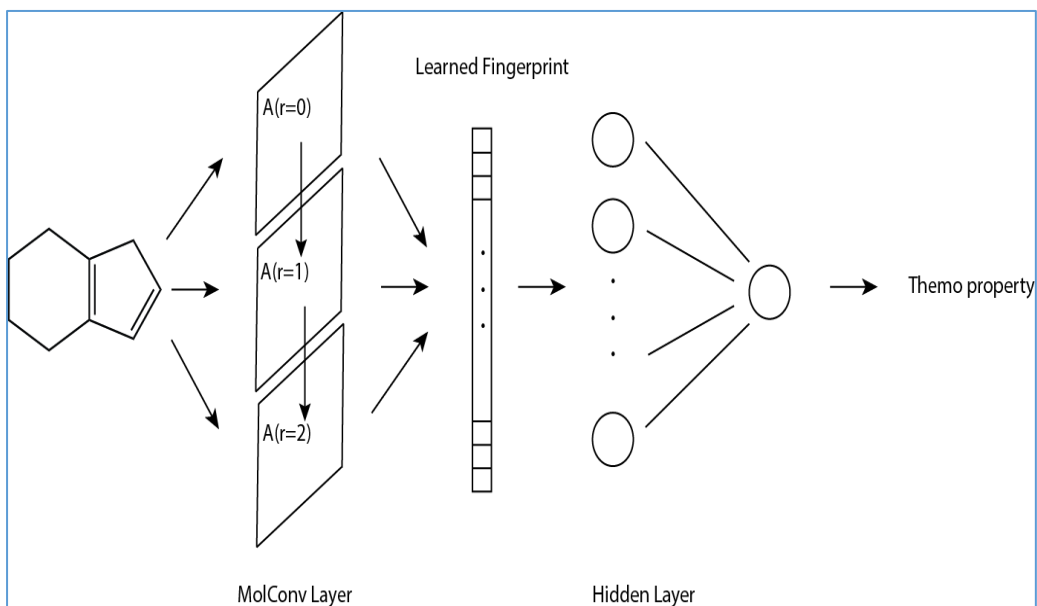
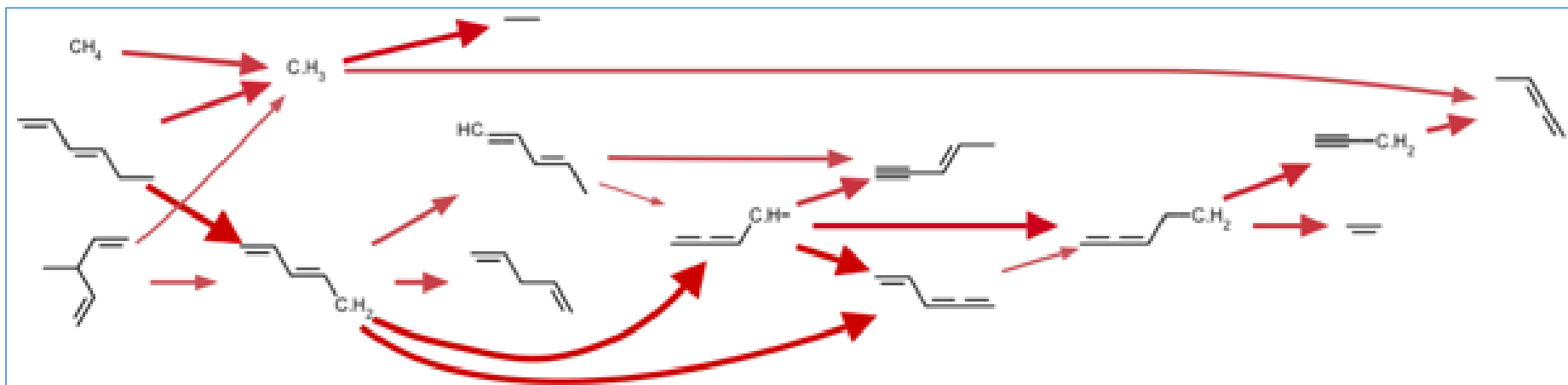
[9] Deep Learning for Chemical Compound Stability Prediction. Ruoqian Liu et al., 2016.

# Молекулярные исследования



- Предсказание биомолекулярных функций и свойств, путем анализа данных биомолекулярных структур [10].

# Молекулярные исследования



- Предсказание термодинамических свойств для неизвестных молекулярных структур (enthalpy, entropy, heat capacities)
- Фреймворк Reaction Mechanism Generator (RMG) [11]



# *Заключение*

**1. Предложен общий подход к Структурно-Функциональному Анализу и Синтезу (СФАС) глубоких конволюционных нейронных сетей (CNN), который позволяет определить:**

- из каких типовых структурно-функциональных элементов (СФЭ) могут строиться CNN;
- каковы необходимые математические свойства СФЭ;
- какие комбинации СФЭ являются допустимыми;
- каковы возможные пути построения и обучения глубоких сетей для анализа и распознавания нерегулярных, неоднородных или сложно структурированных данных.

**2. Определен базовый состав СФЭ, описаны их математические свойства.**

**3. Предложена методика синтеза CNN для заданной структуры входных данных на основе структурной таблицы СФЭ.**

# Заключение

**4. Для практической реализации предлагаемого подхода потребуются разработка специализированного программного обеспечения для формирования и обучения ГКНС на основе СФАС, которое дополнило бы ныне существующие библиотеки и среды разработки и обучения CNN (Caffe, Caffe2, TensorFlow, PyTorch и др.).**

При этом необходимо будет разработать алгоритмы обучения и выполнения SCNN, которые позволят эффективно реализовывать быстрые массовые вычисления на основе удачной «трансляции» данных из произвольных структурных расслоений в регулярные массивы.

Разработка таких алгоритмов и соответствующего программного обеспечения также должны явиться предметом дальнейшей работы в области реализации предложенного подхода.

**Современное состояние и перспективы развития  
методов компьютерного зрения  
и глубокого обучения**

Ю.В. Визильтер, д.ф.-м.н., проф. РАН, [viz@gosniias.ru](mailto:viz@gosniias.ru)

***Спасибо за внимание!***

*Декабрьские чтения  
в Томске*



**Декабрьские чтения**

**Томск, 13.12.2018**