Современное состояние и ближайшие перспективы развития технологий глубокого обучения и компьютерного зрения

Визильтер Юрий Валентинович, ***viz@gosniias.ru***

начальник подразделения 3000 ФГУП «ГосНИИАС»,
д.ф-м.н., профессор РАН

12-я Конференция ИОИ, Италия, г. Гаэта, 08.10.2018.

Содержание

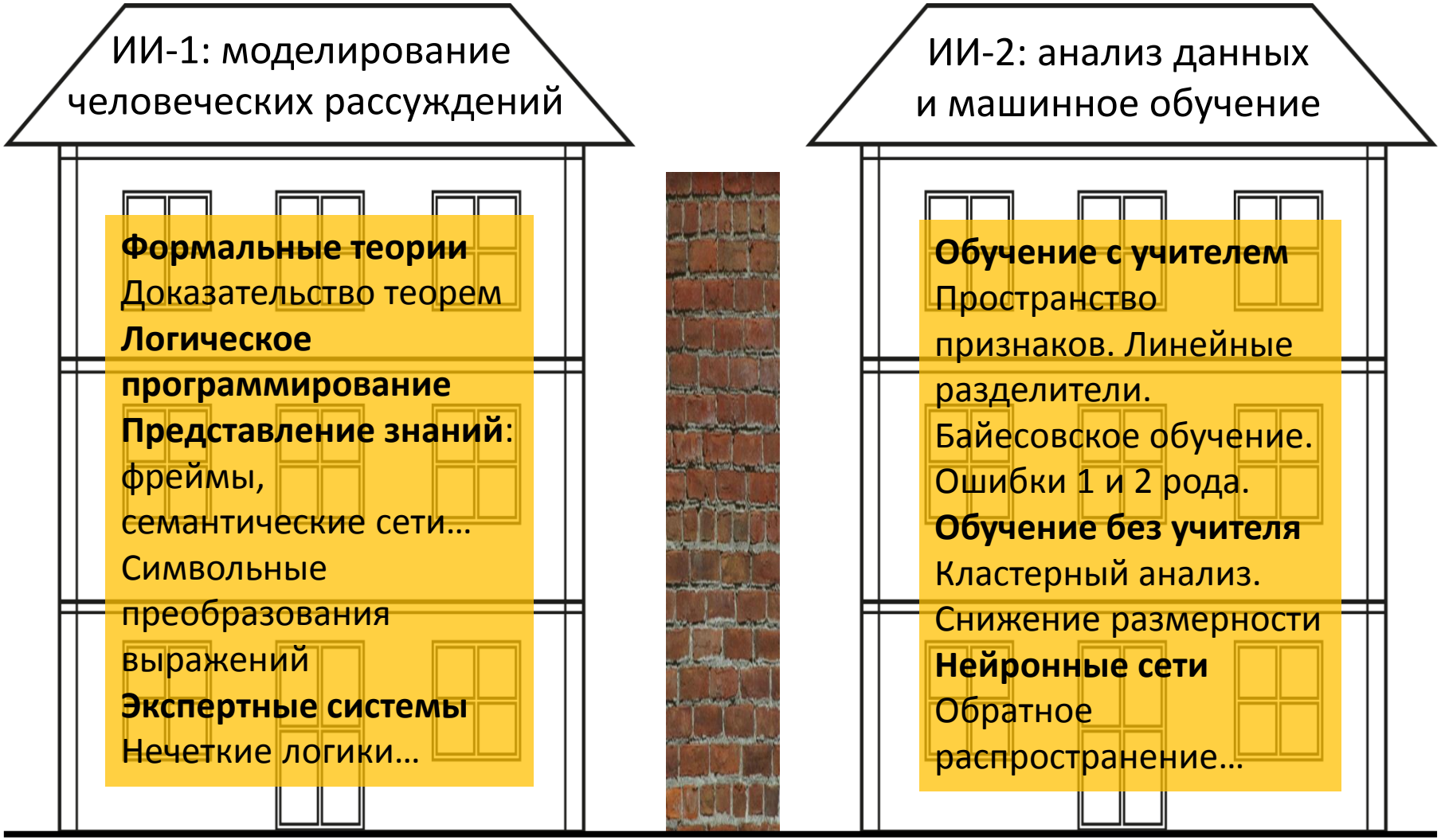
- **Компьютерное зрение и машинное обучение** (*2000-2016 гг. – первая волна технологической революции*)
- **Проекты и результаты ФГУП «ГосНИИАС» (2018)**

- **Перспективные методы и направления глубокого обучения** (*2017+ - второй этап современной революции в машинном обучении и анализе данных*)
- **Проекты и результаты ФГУП «ГосНИИАС» (2018)**

**В докладе можно опознать ряд тем и материалов моих предыдущих обзорных докладов на ИОИ-2014, ММРО-2017, ЭР-2018, ТЗСУ-2018…*

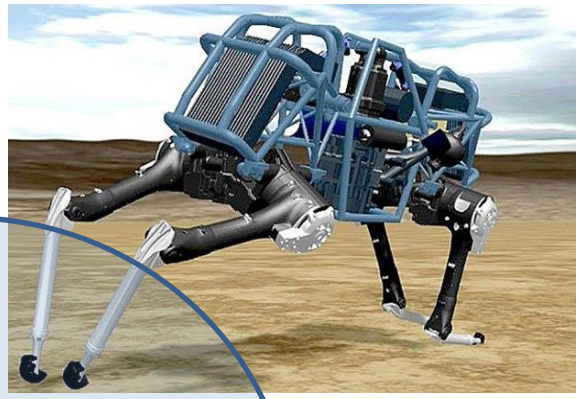
Классический «искусственный интеллект» (до 2000 г.)

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



Насколько пригодны были 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

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

Компьютерное зрение и машинное обучение

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

Компьютерное зрение и машинное обучение

основные задачи, лучшие методы и результаты

(2000-2016 гг. – *первая волна технологической революции*)

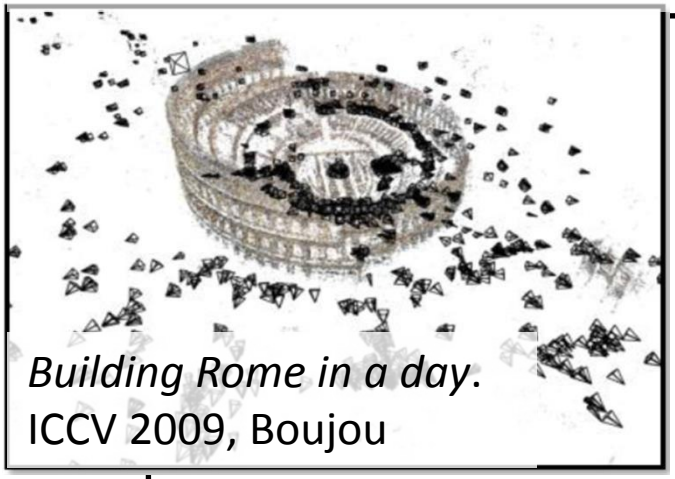
- **Реконструкция 3D сцены и навигация в ней:** Structure-from-Motion, SLAM, Road Scene Understanding and Autonomous Driving
- **Сегментация сцены и понимание видеосюжета:** Saliency maps, Video & 3D Segmentation, 3D-Flow, Multi-Tracking, Human Detection, Human pose estimation, Human Action Detection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

Реконструкция 3D сцены и навигация в ней:

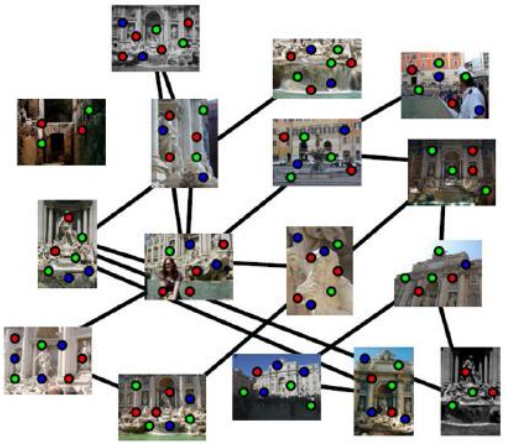
Structure-from-Motion, SLAM, Road Scene Understanding and Autonomous Driving

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

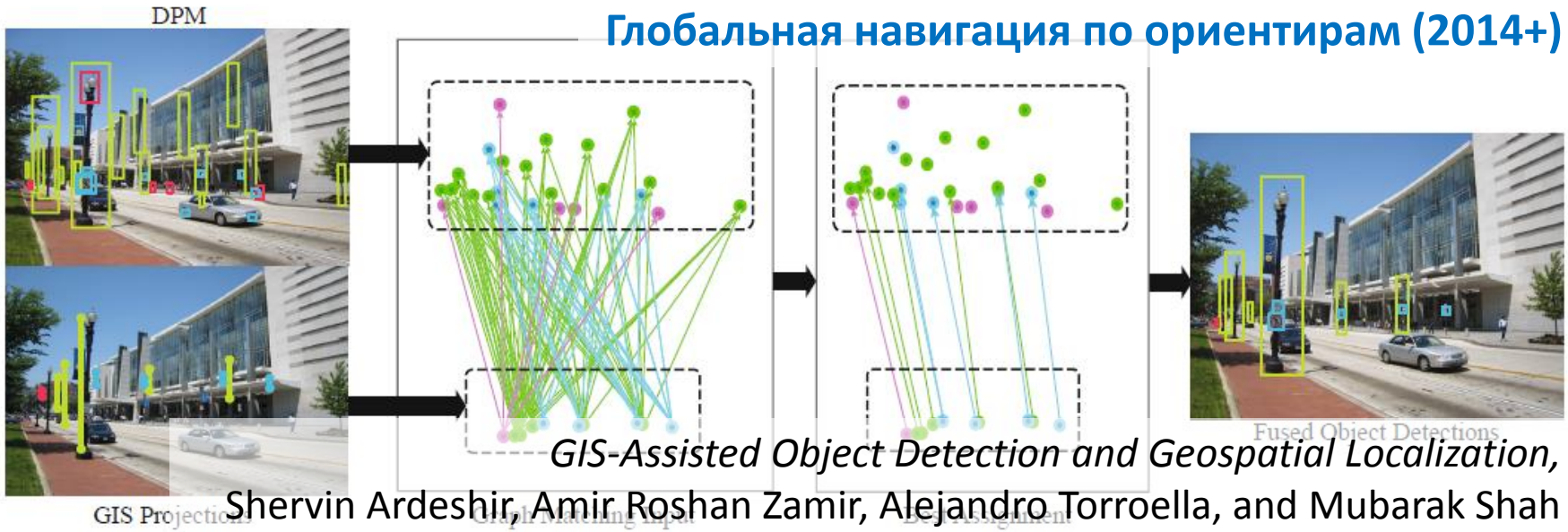


Building Rome in a day. ICCV 2009, Boujou



Scene Chronology. Kevin Matzen, Noah Snavely. ECCV2014

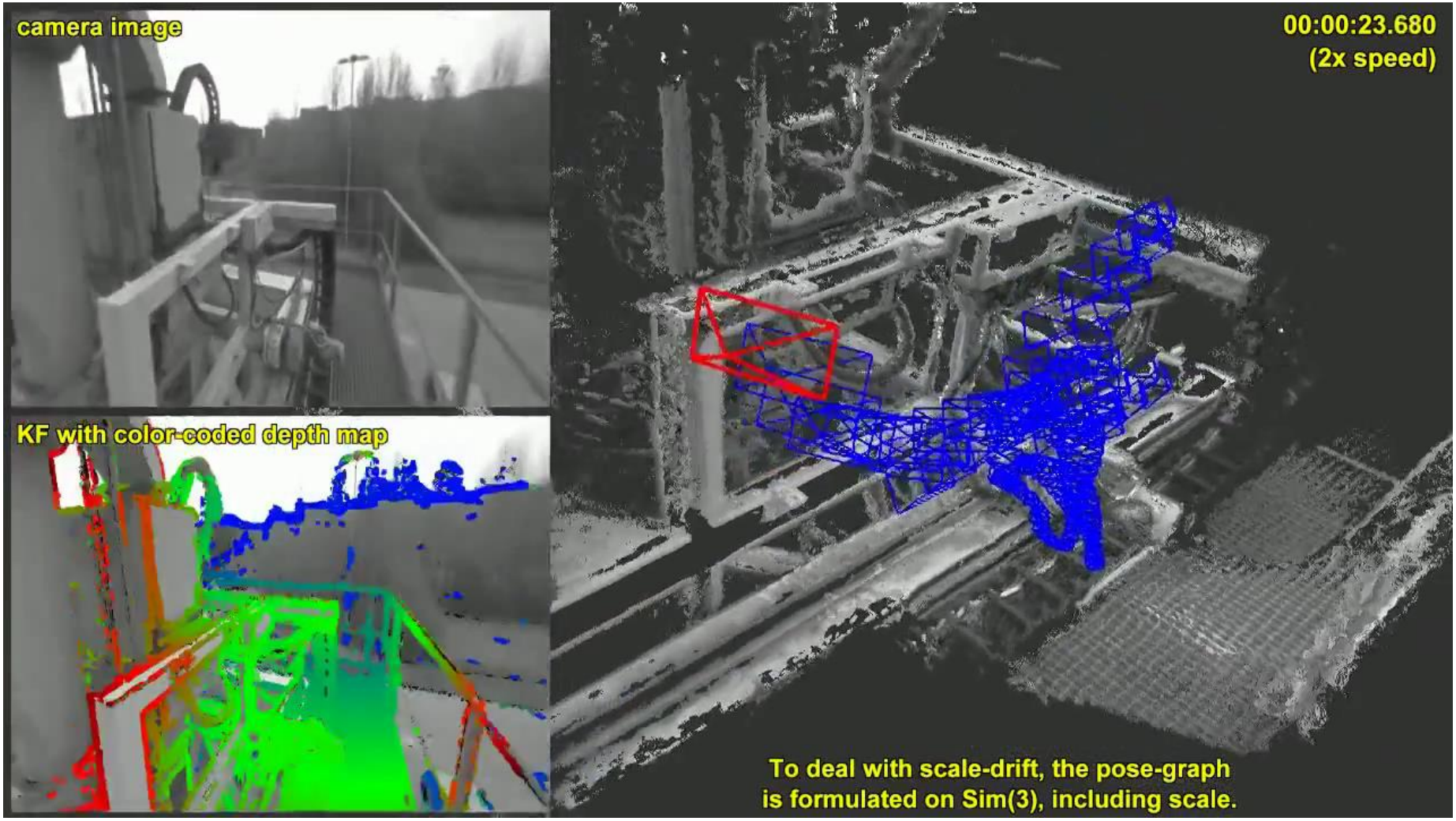
Глобальная навигация по ориентирам (2014+)



GIS-Assisted Object Detection and Geospatial Localization, Shervin Ardeshir, Amir Roshan Zamir, Alejandro Torroella, and Mubarak Shah

Реконструкция 3D сцены и навигация в ней

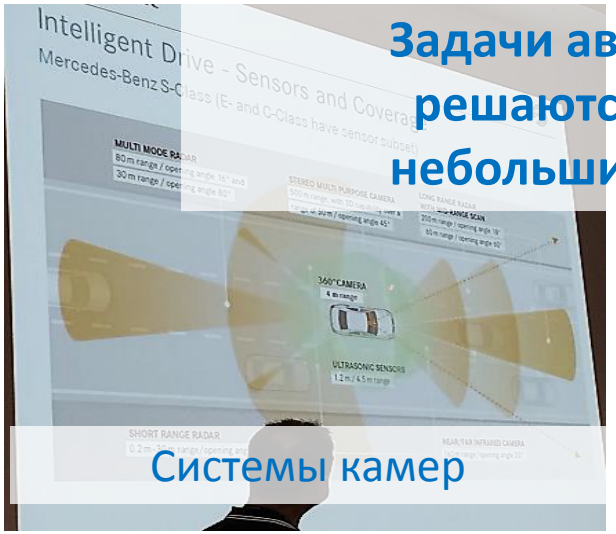
- **SLAM (Simultaneous Localization and Mapping, 2006+, 2014+)** – технология реконструкции 3D сцены и оценки положения/параметров движения камеры



LSD-SLAM: Large-Scale Direct Monocular SLAM, Engel and Schops and Cremers, 2014

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

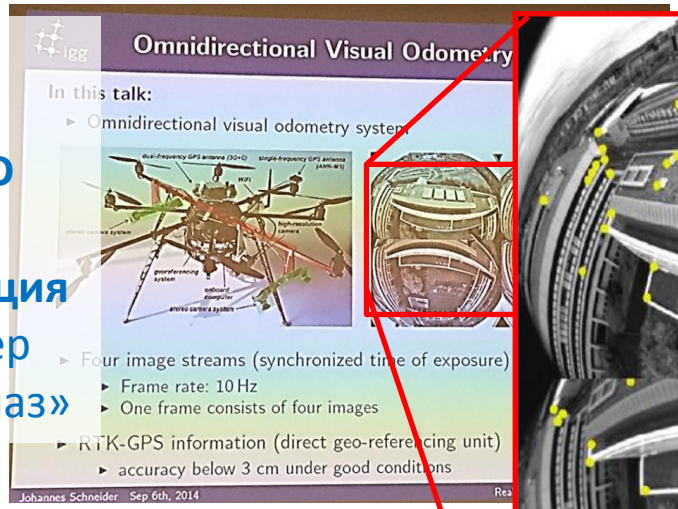
Изображения от камер

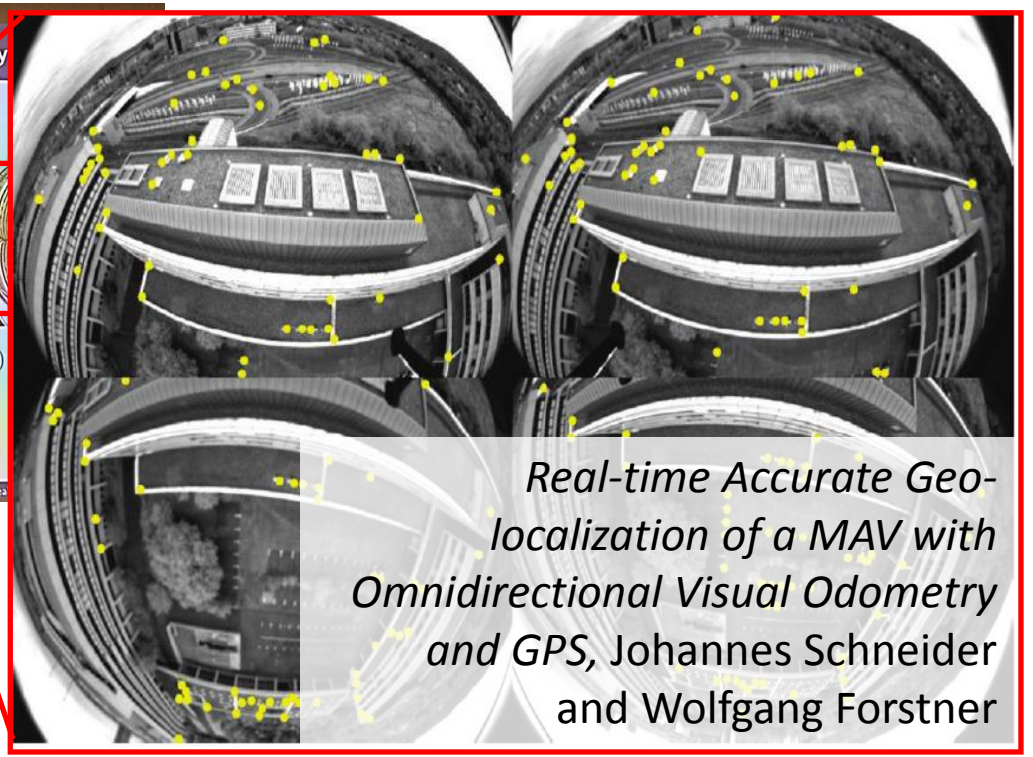


Системы камер

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»





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

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

Сегментация сцены и понимание видеосюжета:

Saliency maps, Video & 3D Segmentation, 3D-Flow, Multi-Tracking, Human Detection, Human pose estimation, Human Action Detection and Prediction, Crowd behavior, Group analysis, Face Detection and Recognition

Saliency Maps (Имитация зрительного внимания)

Примеры карт внимания (фиксаций взгляда) человека
и результатов работы автоматических алгоритмов поиска зон интереса

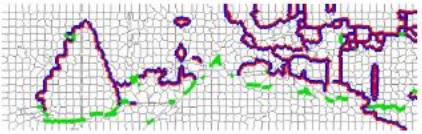
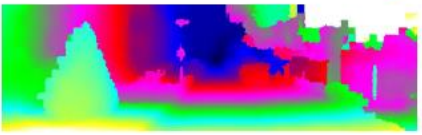
A Closer Look at Context: From Coxels to the Contextual Emergence of Object Saliency Rotem Mairon and Ohad Ben-Shahar, ECCV'14

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

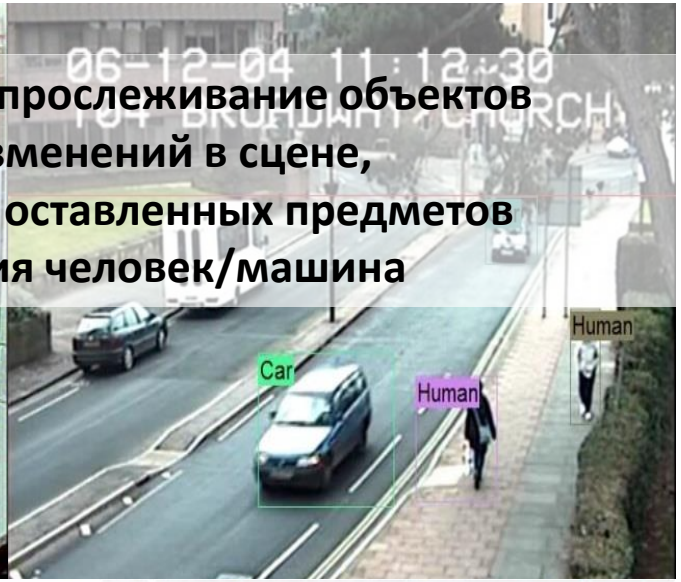
Единый подход к обработке и сегментации данных (2003+)
2D, 2D+T, 3D, 3D+T:
MRF, Energy-based, 3D Flow

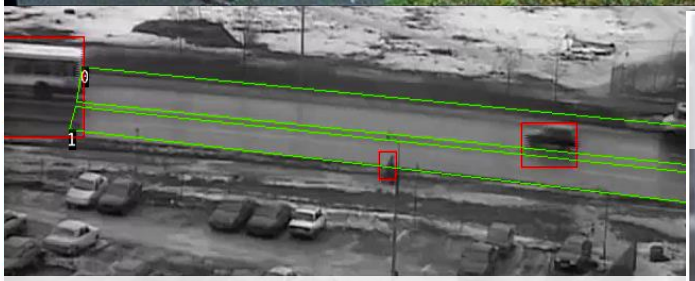




Минимизация целевой функции (энергии)
Модели на основе **Марковских полей**
Использование **карт внимания**

Быстрые методы анализа видеопоследовательностей (до 800 fps)

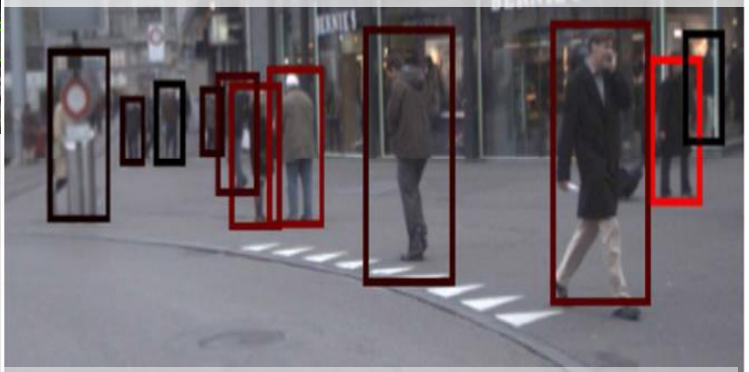
- **Выделение и прослеживание объектов**
- **Выделение изменений в сцене, обнаружение оставленных предметов**
- **Классификация человек/машина**



- 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:

Convolution networks,
Deep learning,
Image Retrieval,
Object Detection

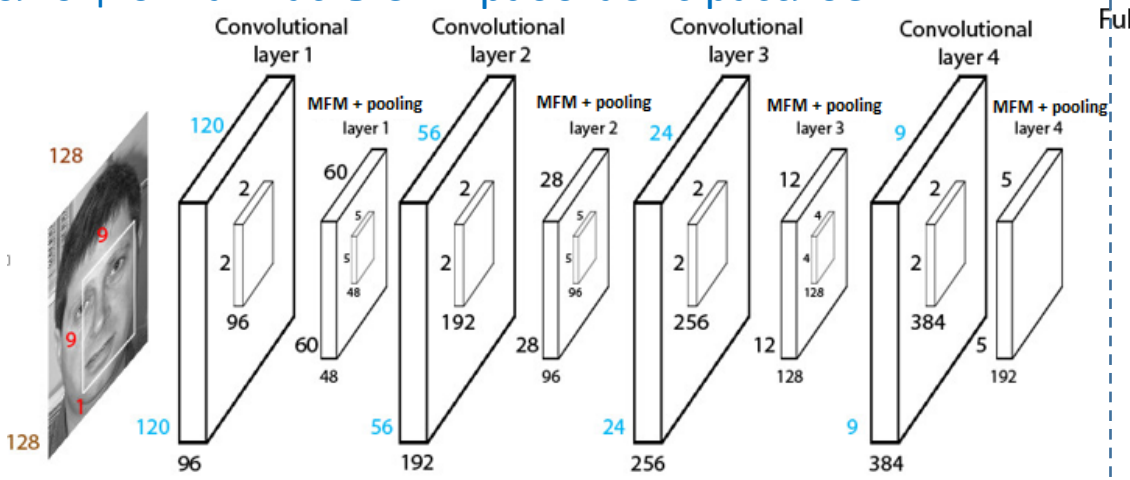
Глубокие конволюционные нейронные сети – новое поколение алгоритмов обнаружения и распознавания объектов на изображениях

2011: First Superhuman Visual Pattern Recognition

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

Интерпретация CNN как процесса эволюции расслоения базового многообразия

Конволюционная часть CNN работает с расслоениями



Вырожденное расслоение = = Базовое многообразие

Углубление расслоения

Компактификация (сжатие) базового многообразия

Глубина вектора признаков = размерности векторного подпространства

ConvID

Вырожденное расслоение = = вектор признаков

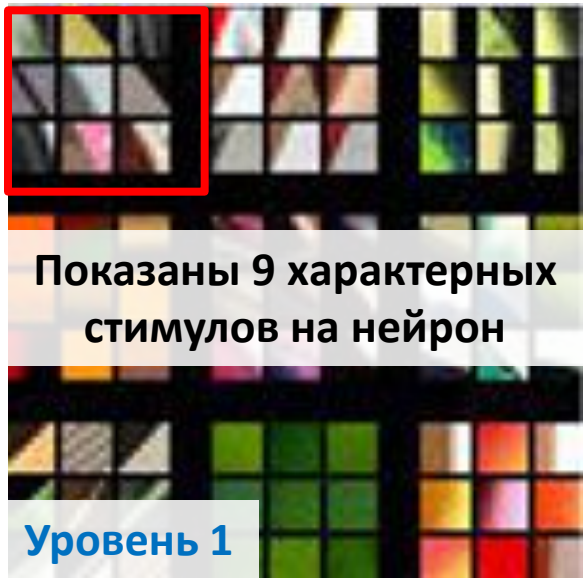
DeepID

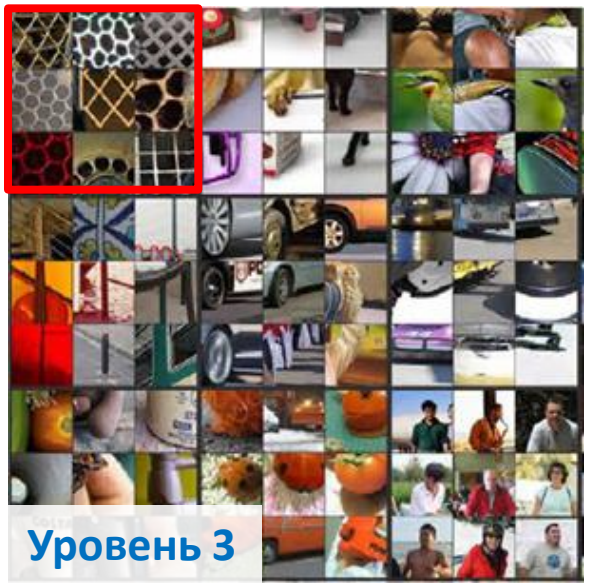
Снижение размерности вектора признаков

А это уже не признаки, а классы

Полносвязная часть CNN (персептрон) работает с векторами

Convolution networks, Deep learning, Image Recognition





На какие элементы изображения реагируют нейроны разных уровней:

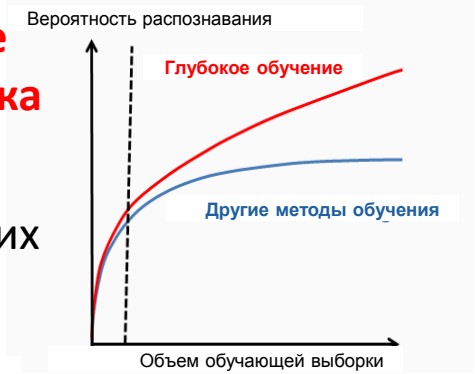
чем выше слой сети, тем выше уровень абстракции

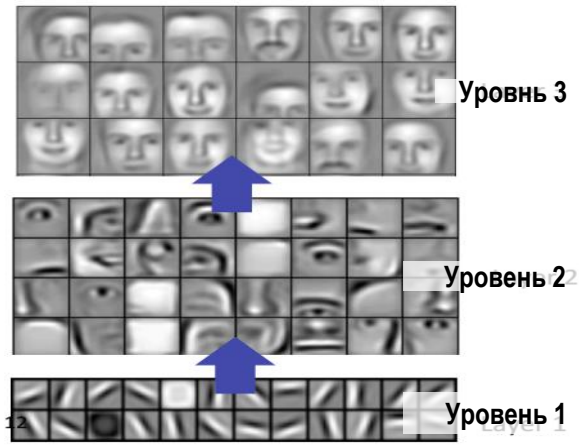
Visualizing and Understanding Convolutional Networks, Matthew Zeiler and Rob Fergus, ECCV'14

Автоматическое обнаружение и распознавание объектов *на базе глубоких конволюционных нейронных сетей (с 2011)*

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

+ Обучение на сверхбольших объемах данных



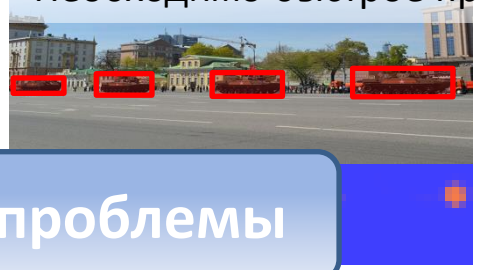


+ Иерархическое обучение с повышением абстракции данных от уровня к уровню

- **Нужны огромные обучающие выборки**
- Длительное моделирование и обучение

Постановка задачи обнаружения и Распознавания объектов заданного типа → Формирование «комплекта обучения» на основе реальных и модельных данных → Обучение алгоритмов обнаружения и глубокой нейронной сети

- **Ресурсоемкость, низкая скорость**
- Необходимо быстрое предобнаружение



Достоинства и **проблемы**

+ Тысячи слоев нейронов

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

Типовая структура глубокой конволюционной сети

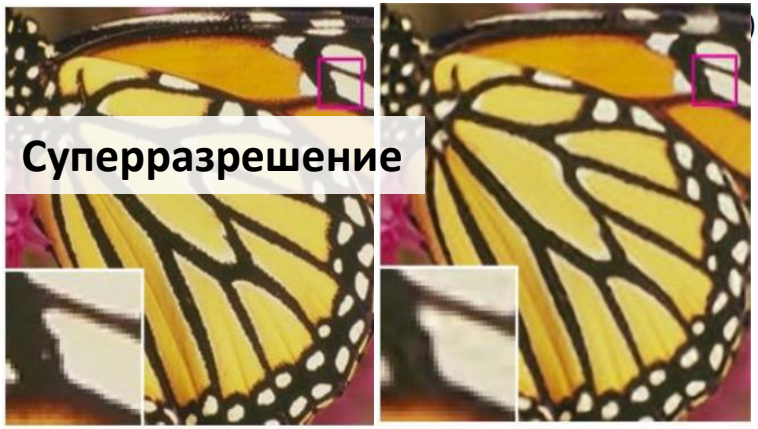
- Необходимость эффективных алгоритмических реализаций
- Необходимость создания нового поколения нейропроцессоров

CNN-based Image Restoration and Analysis

noisy ($\sigma = 25$)PSNR:20.16d ours: PSNR:30.03dB

Фильтрация шумов

Удаление смаза

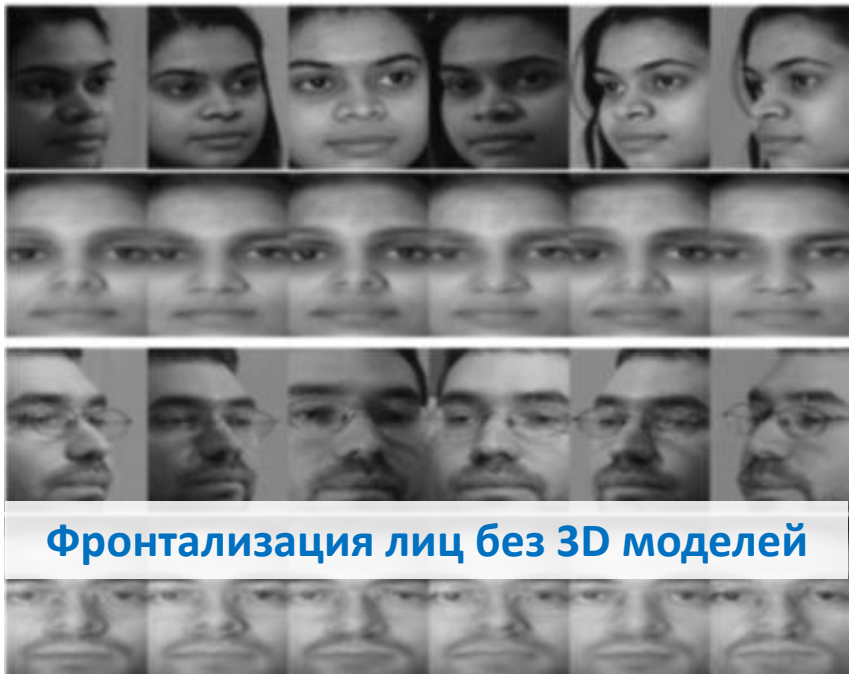


Ours

Суперразрешение

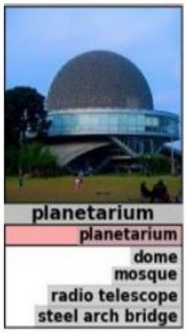
Original / PSNR SRCNN / 27.95 dB

Обнаружение особых точек на лицах

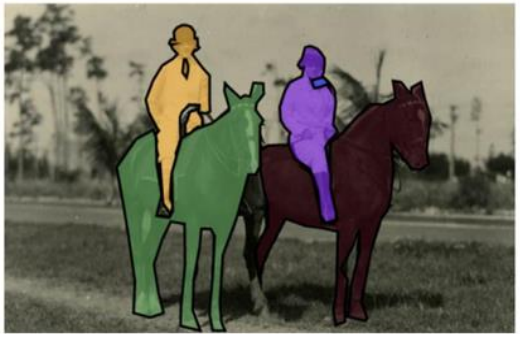


Фронтализация лиц без 3D моделей

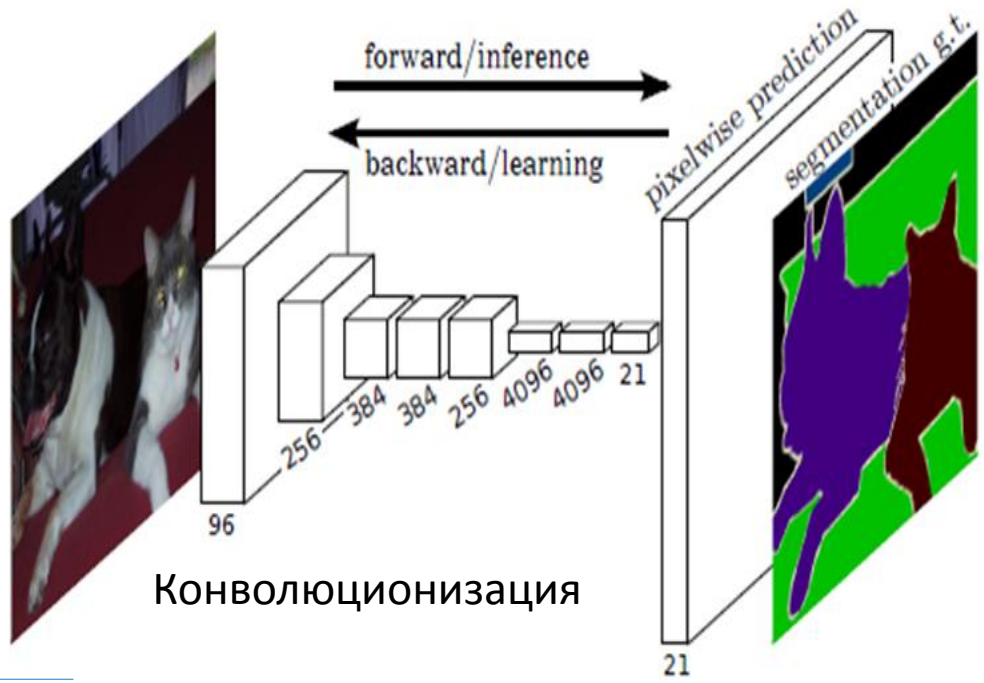
Семантическая сегментация



Задача распознавания: Изображение -> метка класса



Задача семантической сегментации: Изображение -> метка класса + сегментация границ



Конволюционизация

Machine Learning for Aerial Image Labeling - V. Mnih

Effective Semantic Pixel labelling with Convolutional Networks and Conditional Random Fields - Paisitkriangkrai, Sherrah, Janney and Van-Den Hengel

Приложение (ГосНИИАС): автоматическая семантическая сегментация аэроснимков

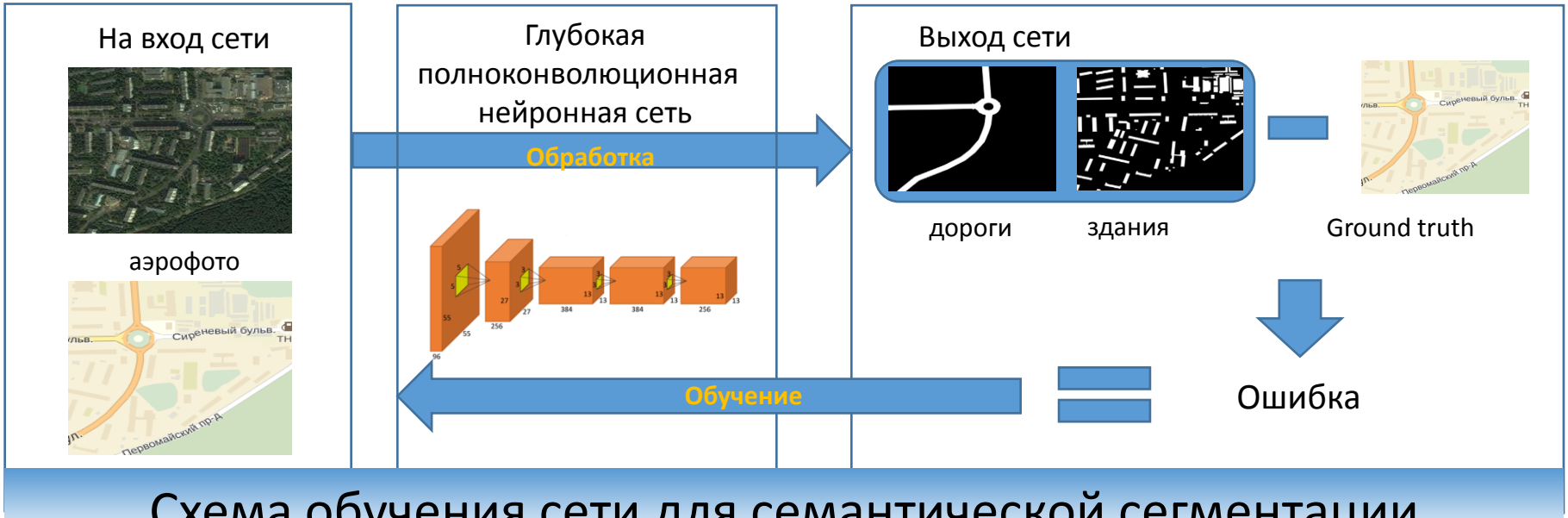
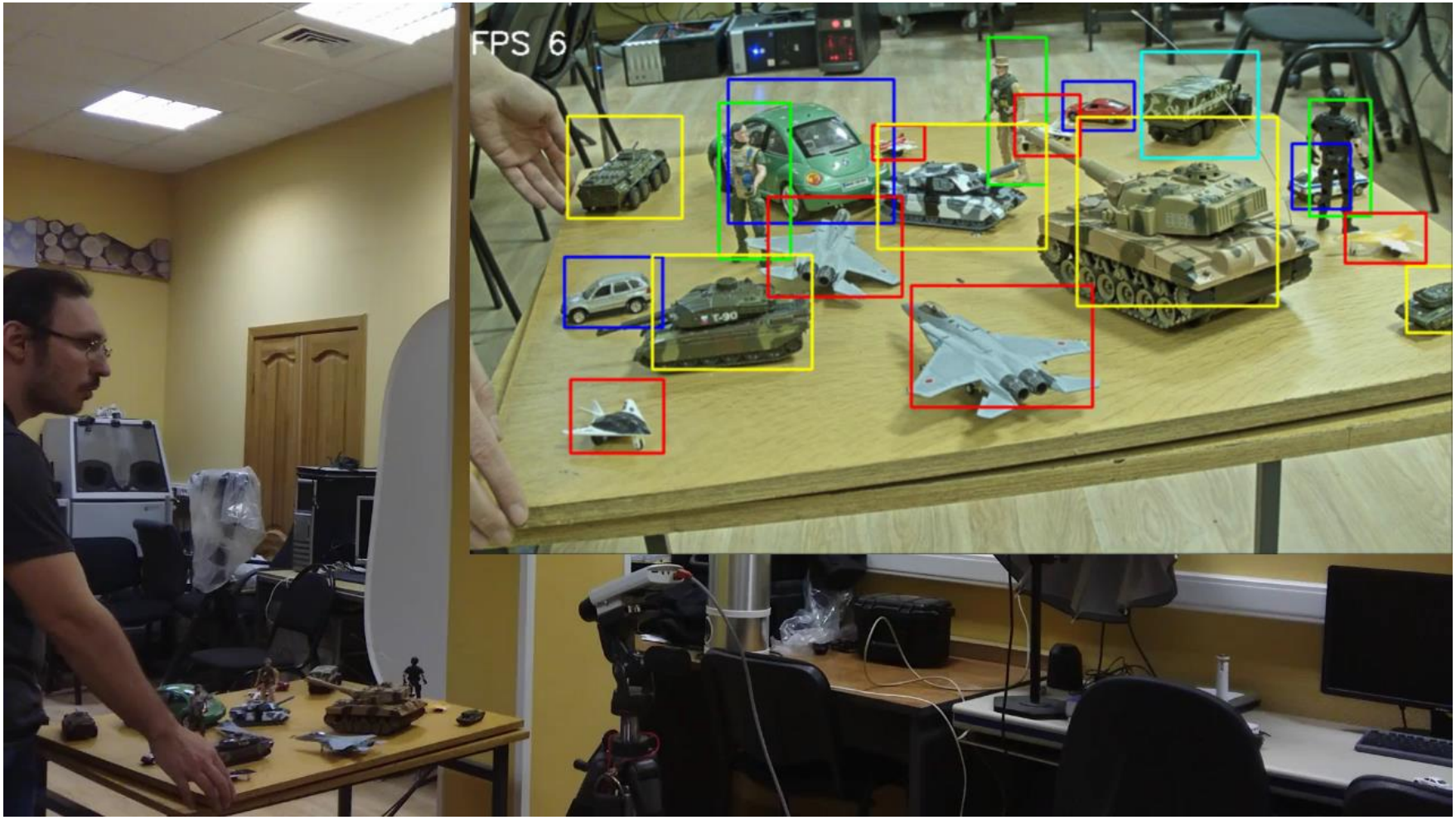


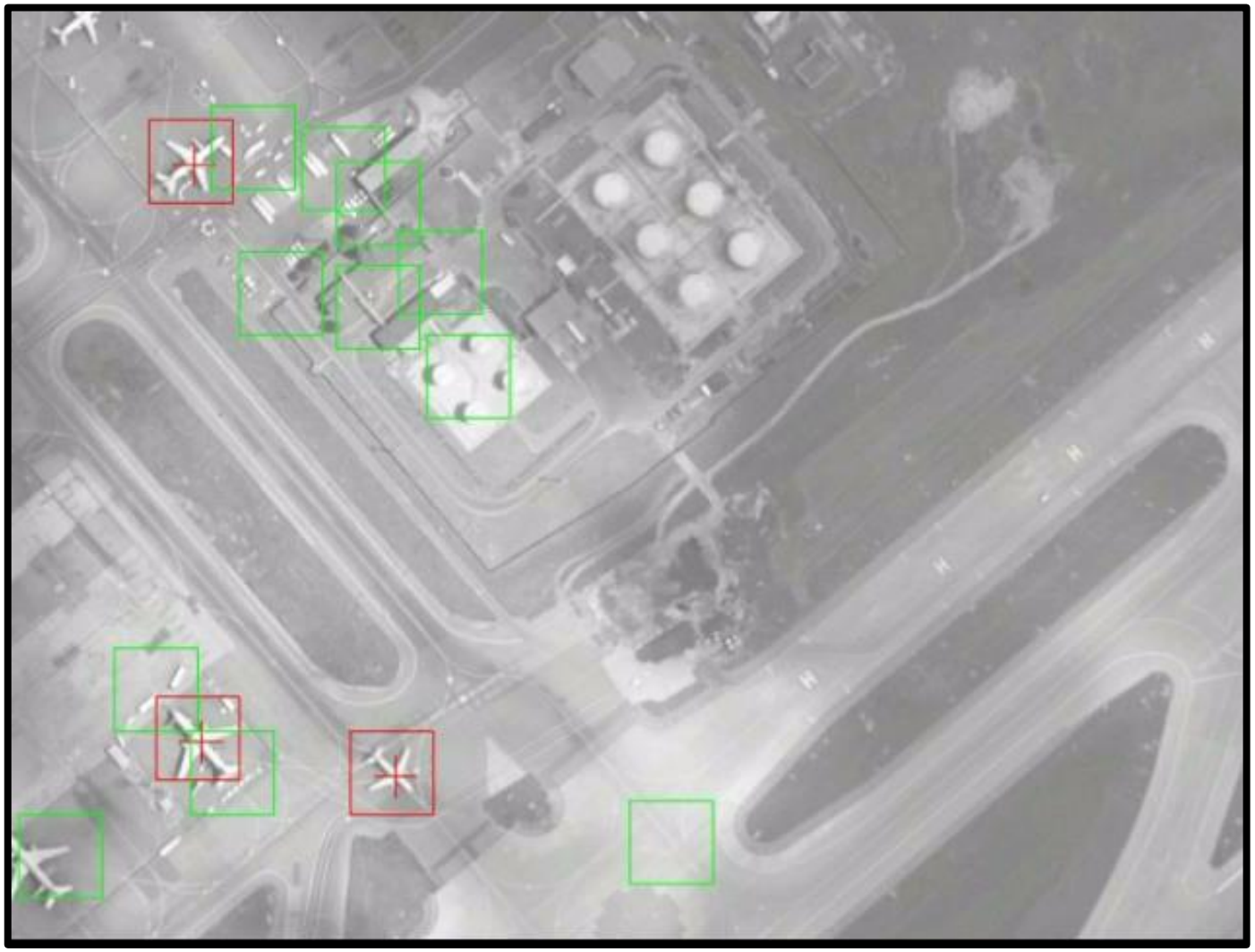
Схема обучения сети для семантической сегментации и пример работы сети для выделения зданий и дорог

Обнаружение и распознавание объектов в реальном времени (ГосНИИАС, 2016)



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

Автоматическое обнаружение самолетов на аэроснимках (ГосНИИАС-2016)



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

ФОНД ПЕРСПЕКТИВНЫХ ИССЛЕДОВАНИЙ

ФГУП «ГосНИИАС» - призер конкурса

ФОНД ПЕРСПЕКТИВНЫХ ИССЛЕДОВАНИЙ

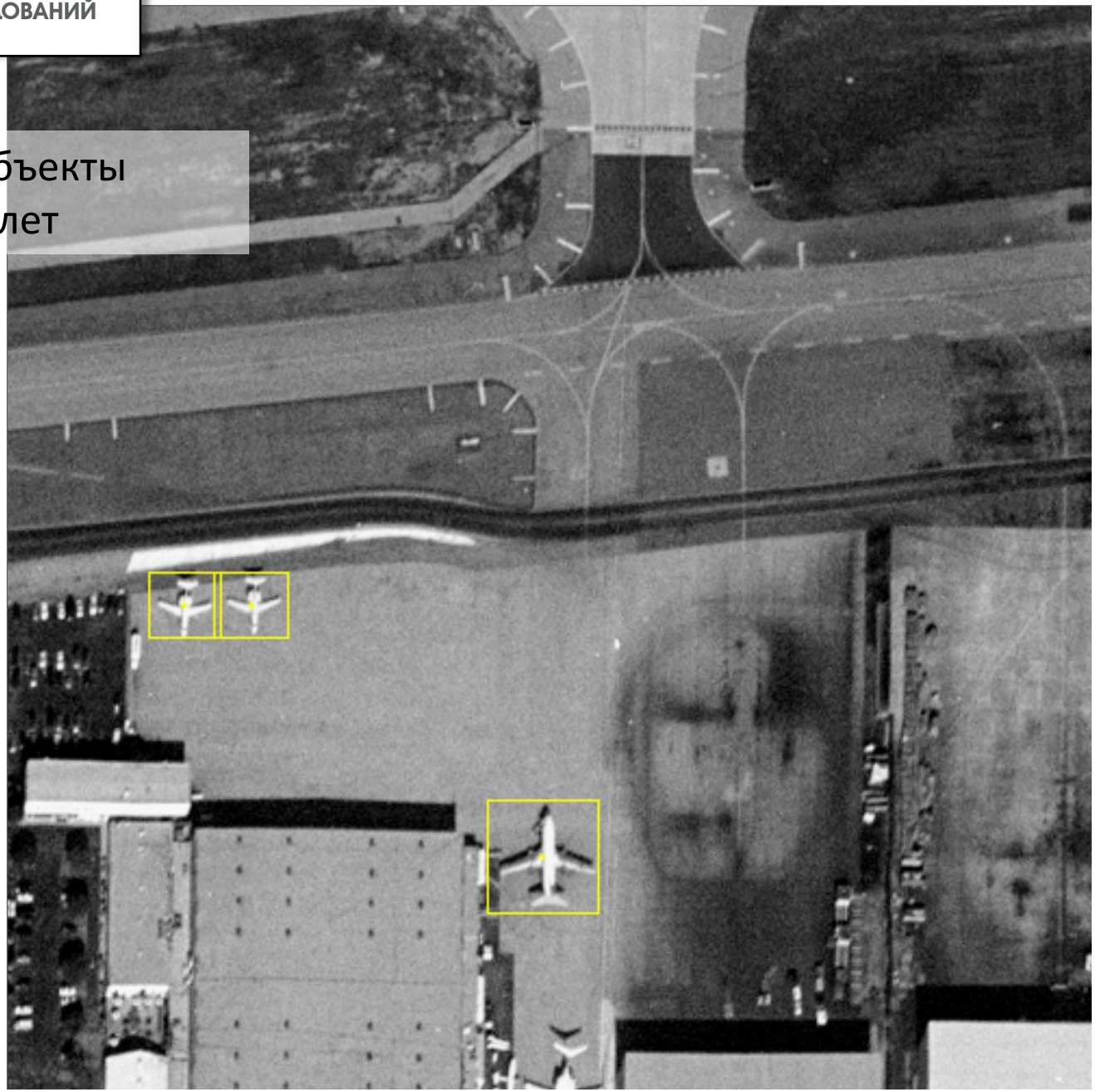
Результаты обнаружения (ГосНИИАС)

Показаны объекты класса самолет

ФОНД ПЕРСПЕКТИВНЫХ ИССЛЕДОВАНИЙ

Результаты обнаружения (ГосНИИАС)

Показаны объекты класса самолет



ФОНД ПЕРСПЕКТИВНЫХ ИССЛЕДОВАНИЙ

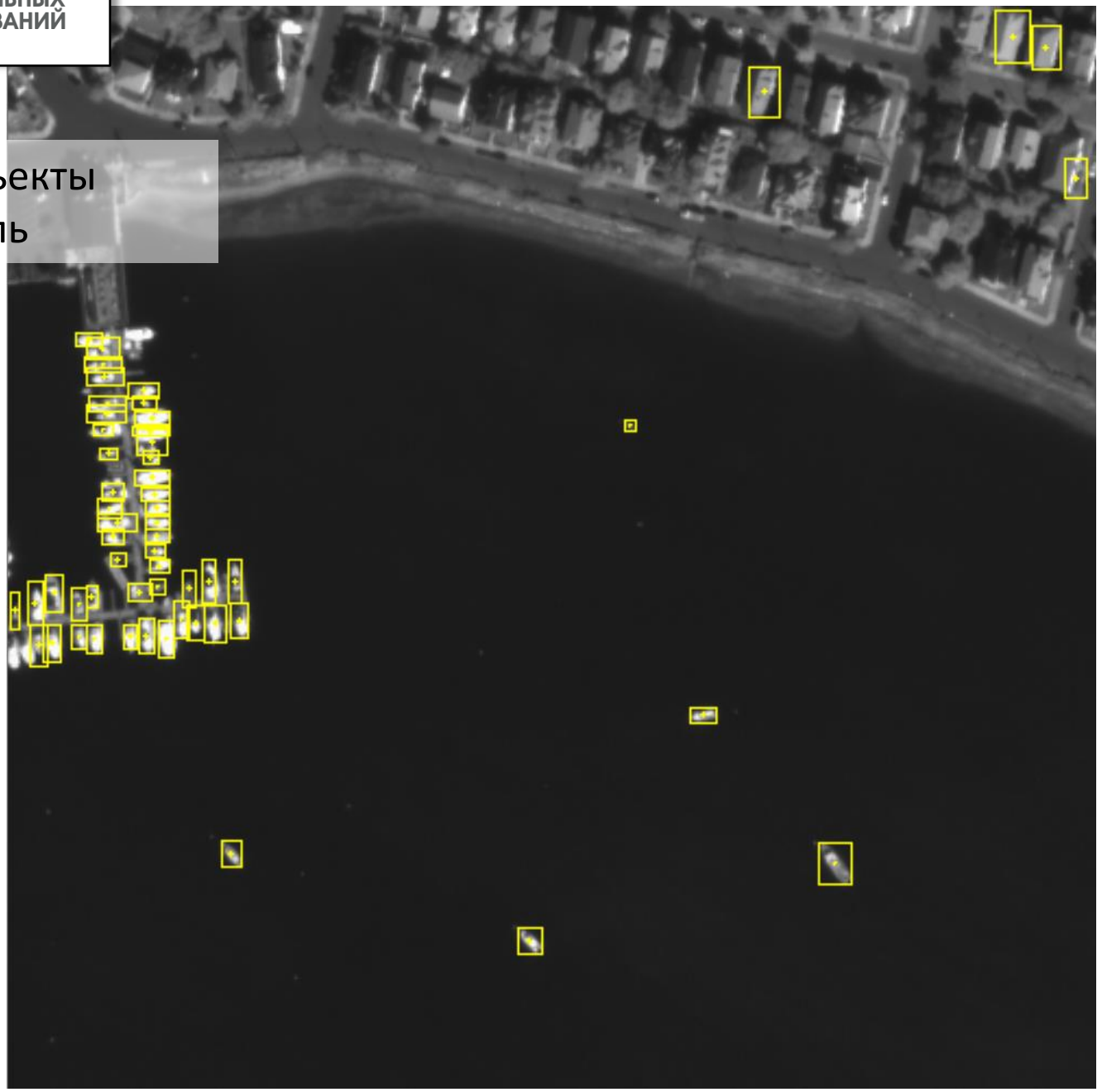
Результаты обнаружения (ГосНИИАС)

Показаны объекты класса самолет

ФОНД ПЕРСПЕКТИВНЫХ ИССЛЕДОВАНИЙ

Результаты обнаружения (ГосНИИАС)

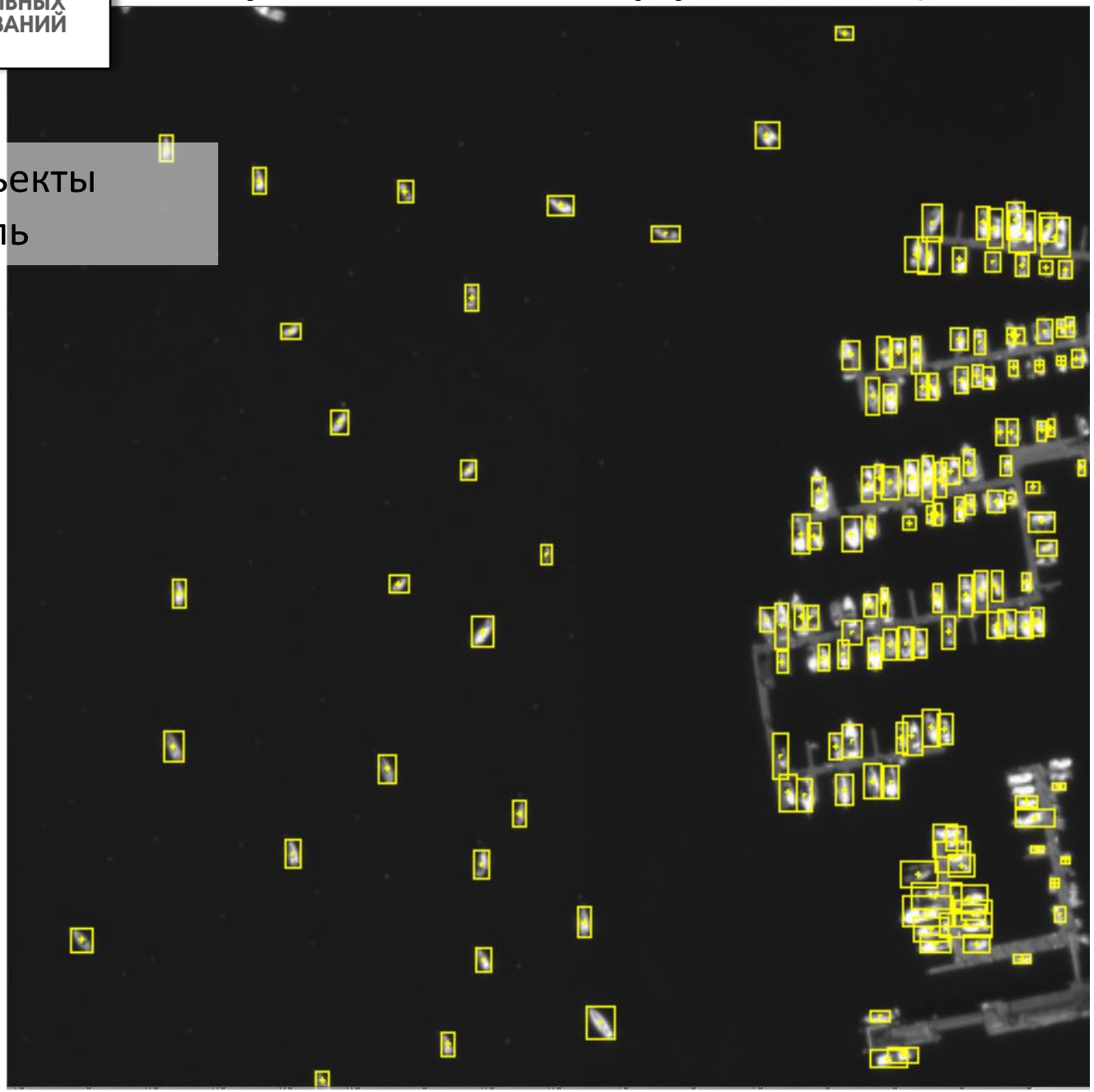
Показаны объекты класса корабль



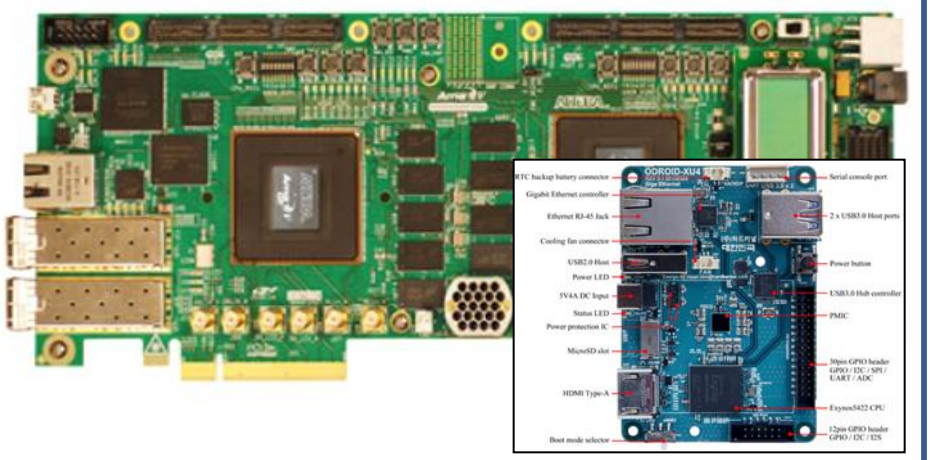
ФОНД ПЕРСПЕКТИВНЫХ ИССЛЕДОВАНИЙ

Результаты обнаружения (ГосНИИАС)

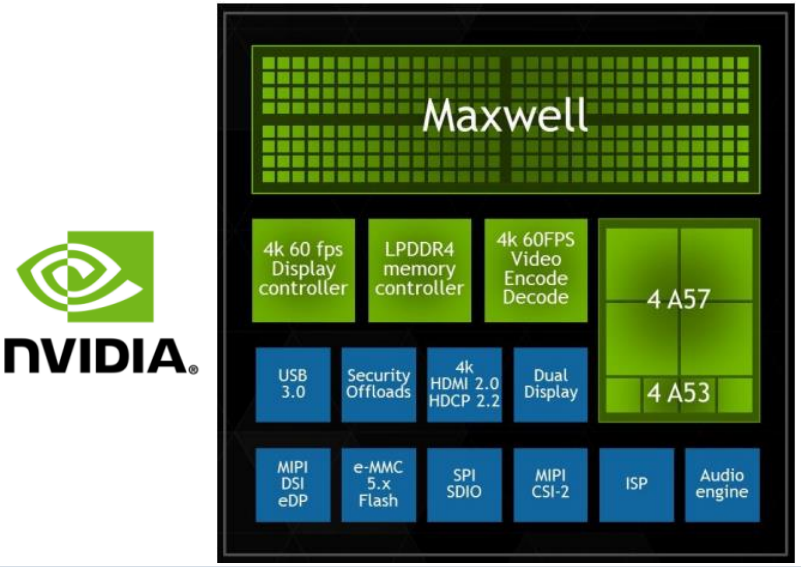
Показаны объекты класса корабль



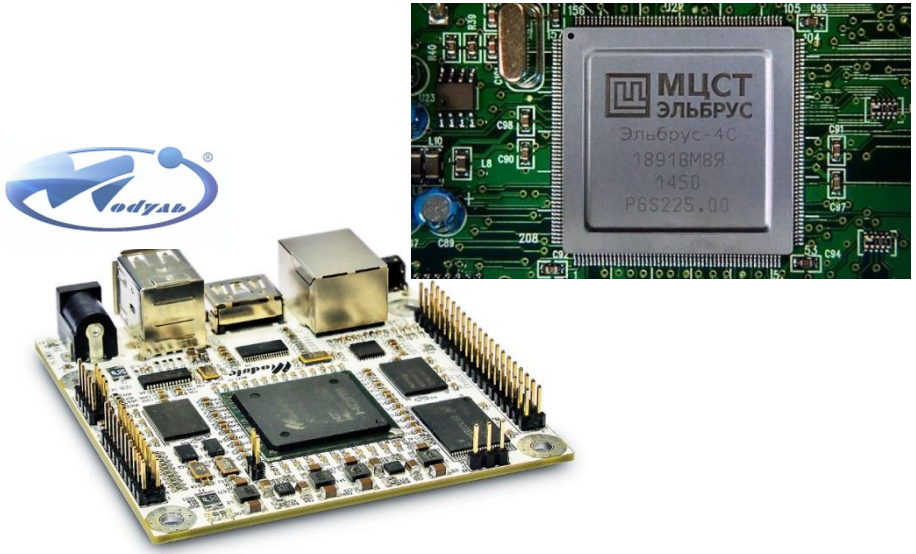
Аппаратно-программные решения



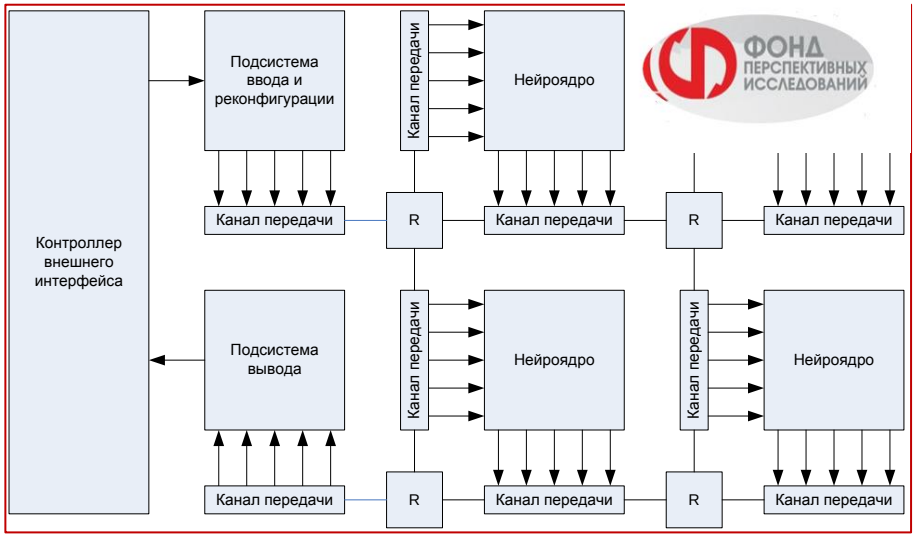
ПЛИС (FPGA), ARM



NVIDIA GPGPU



НТЦ Модуль, МЦСТ Эльбрус



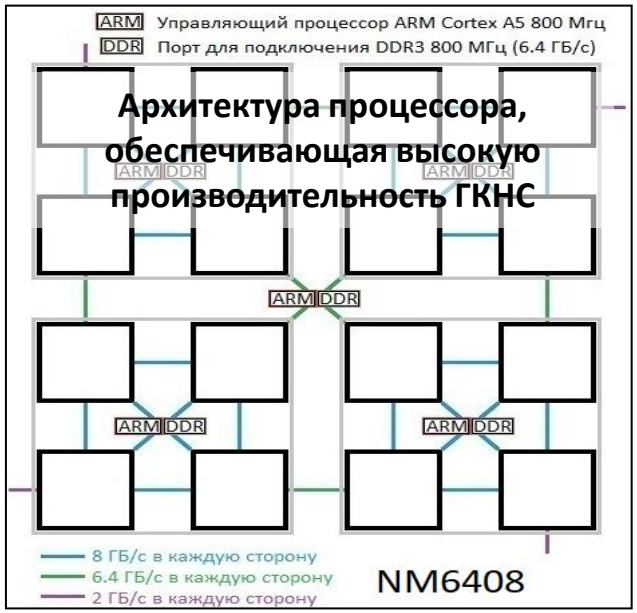
ФПИ «Мемристор», МФТИ

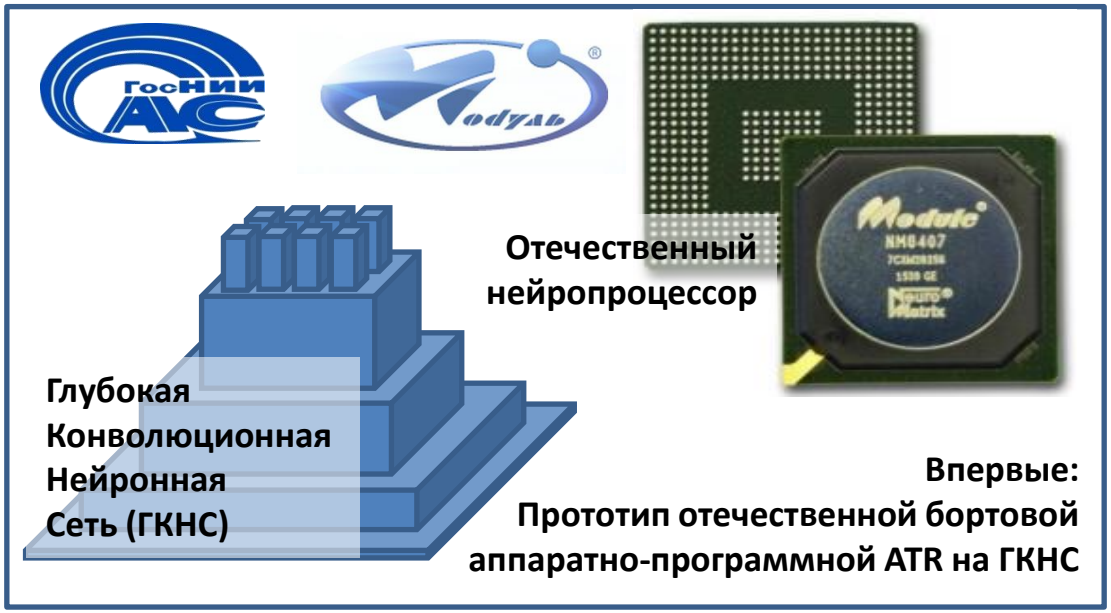
Проекты и результаты ФГУП «ГосНИИАС» 2018

Отработка реализации ГКНС на процессорах NM6407 и NM6408

Прототип системы автоматического обнаружения и распознавания целей на основе глубоких сверточных нейронных сетей. Система на базе платы МС121.01 производства НТЦ «Модуль» с процессором NM6407.

В настоящее время разрабатывается решение для ATR с ГКНС на базе **NM6408**, работающее в 32 раза быстрее.





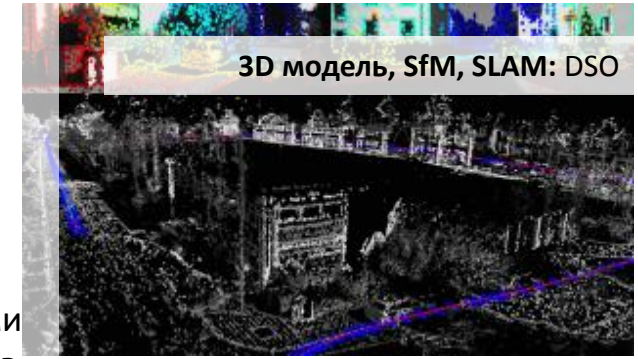
Проект ФПИ по созданию СТЗ для автономных РТК и групп РТК

Создание СТЗ для автономных РТК и групп РТК

При движении РТК по сложной пересеченной местности:

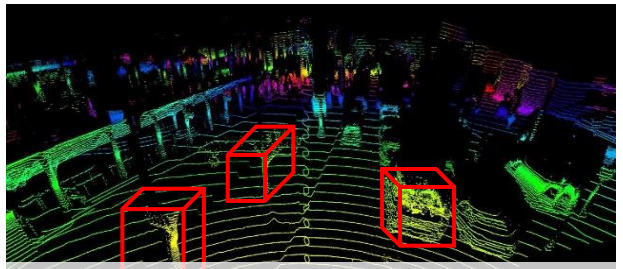
- **Построение в реальном времени информационной трехмерной семантической модели местности** (полная 3D модель окружающей обстановки с распознанными типами поверхностей, ориентиров, объектов, целей и препятствий);

- Автоматическое обнаружение, распознавание и прослеживание множества целей с помощью ГКНС, позволяющие получать результаты, превосходящие качество и скорость работы человека-наводчика.

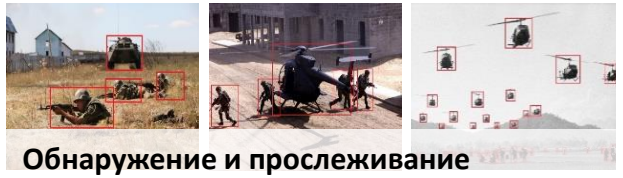


3D модель, SfM, SLAM: DSO

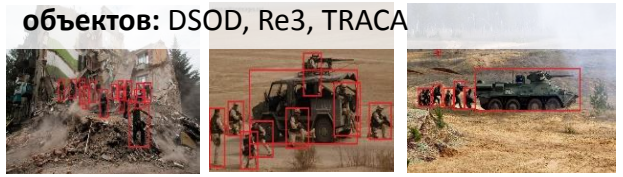
Семантическая сегментация: DeepLab v3+, DeepFeatureFlow



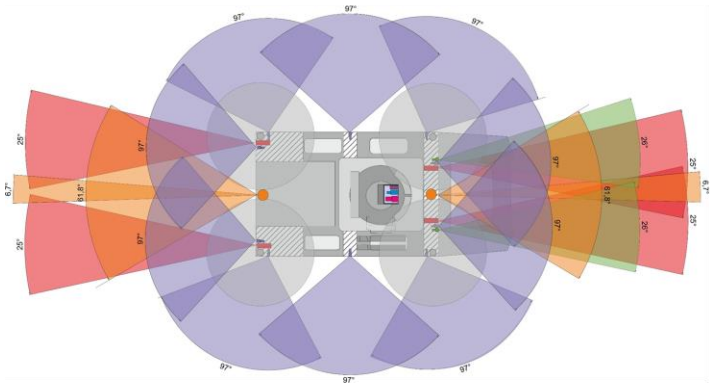
Обнаружение и распознавание препятствий: SqueezeSeg



Обнаружение и прослеживание объектов: DSOD, Re3, TRACA



Гибкая модульная архитектура многоспектральной СТЗ РТП + СТЗ УМПН



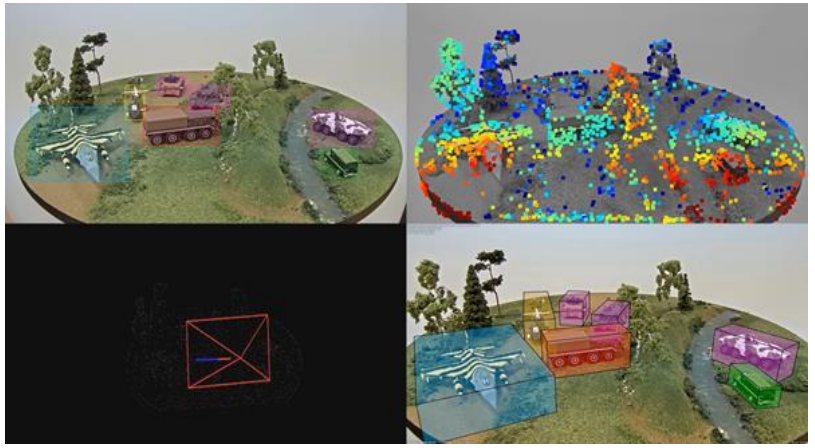
Датчики РТП

- Тепловизор FLIR A615
- Лидар VLP 16
- Цифровая PTZ-камера M5525-E
- PROSILICA GT2050 / Basler acA1920-50gc
- PROSILICA GT2050 / Basler acA1920-50gc
- Радар DELPHI

Датчики УМПН

- Basler acA1920-150uc
- Тепловизор Pergam UM640
- Дальномер Пергам GLR-10

Динамическая трехмерная семантическая модель окружающей обстановки

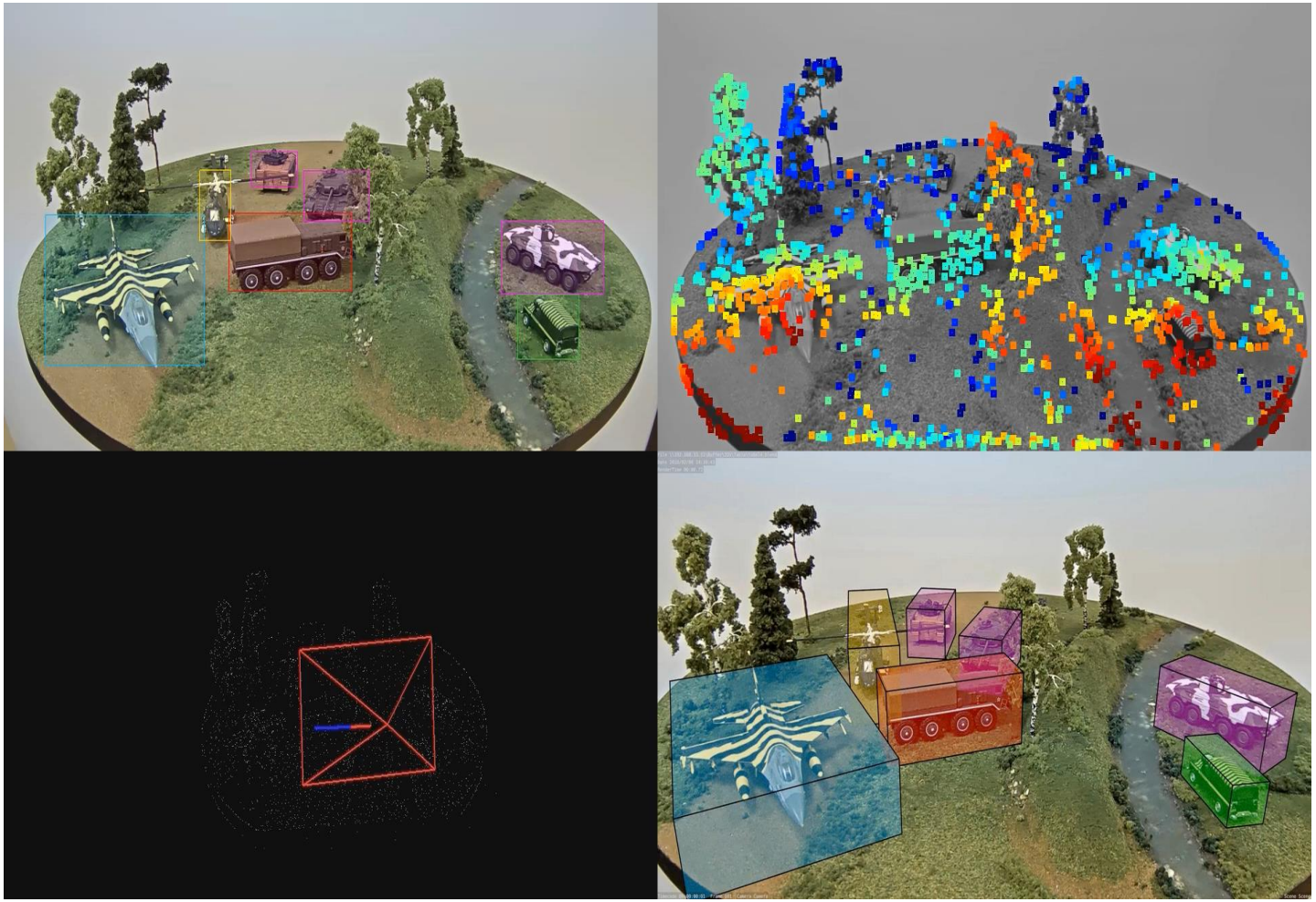


Агрегация всех типов данных в одной информационной 3D модели

Динамическая трехмерная модель мира (камеры + лидары + датчики)

Семантическая сегментация

Обнаружение и прослеживание объектов



Результат: динамическая трехмерная семантическая модель окружающей обстановки

Семантический коррелятор
для обнаружения объектов
по одному эталону
или малому числу эталонов

1. Корреляционное обнаружение объектов на изображениях (Image Matching)

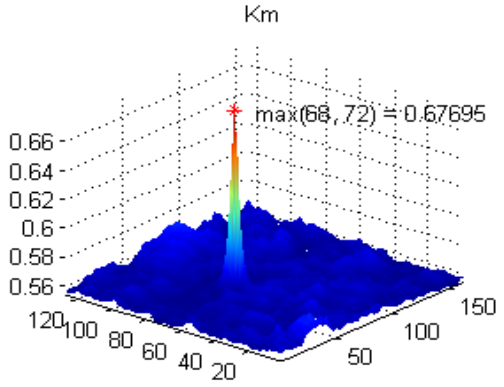
Эталонный фрагмент

Тестовое изображение

Локализация объекта

Корреляционный пик

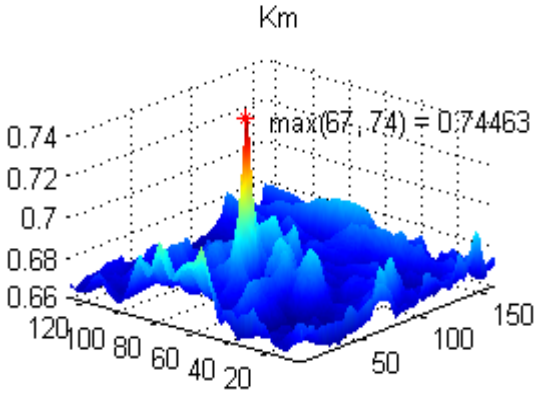
1



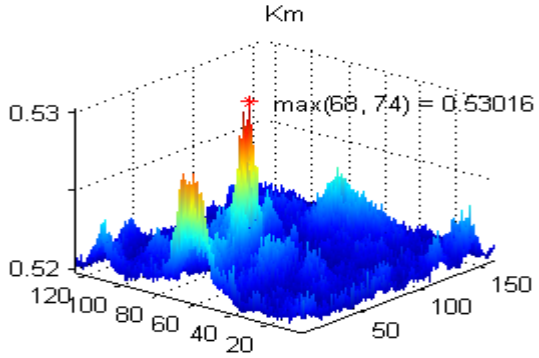
Примеры двумерных корреляционных полей:

1. Высокое отношение сигнал/шум
2. Среднее отношение сигнал/шум
3. Наличие ложного пика
4. Низкое отношение сигнал/шум

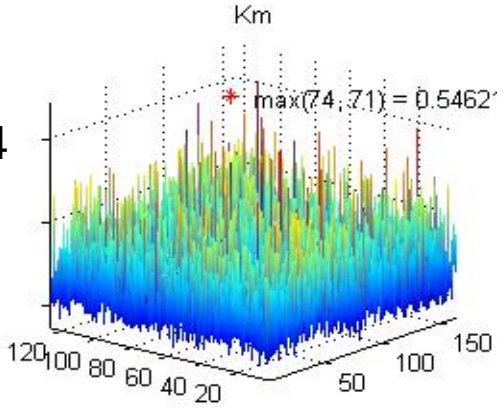
2



3



4



Корреляционное прослеживание объектов на видеопоследовательностях

Коррелятор:

+ дает локализацию на изображении
+ работает по одному эталону
- не инвариантен к масштабу, ракурсам
- не допускает изменчивости формы

захваченный эталон

последовательность кадров

Корреляционное прослеживание объектов на видеопоследовательностях

Коррелятор:

+ дает локализацию на изображении
+ работает по одному эталону
- не инвариантен к масштабу, ракурсам
- не допускает изменчивости формы

захваченный эталон

последовательность кадров

Корреляционное прослеживание объектов на видеопоследовательностях

Коррелятор:

+ дает локализацию на изображении
+ работает по одному эталону
- не инвариантен к масштабу, ракурсам
- не допускает изменчивости формы

Корреляционное прослеживание объектов на видеопоследовательностях

Коррелятор:

+ дает локализацию на изображении
+ работает по одному эталону
- не инвариантен к масштабу, ракурсам
- не допускает изменчивости формы

результат обнаружения

текущий эталон

смена эталона

последовательность кадров

Корреляционное прослеживание объектов на видеопоследовательностях

Коррелятор:

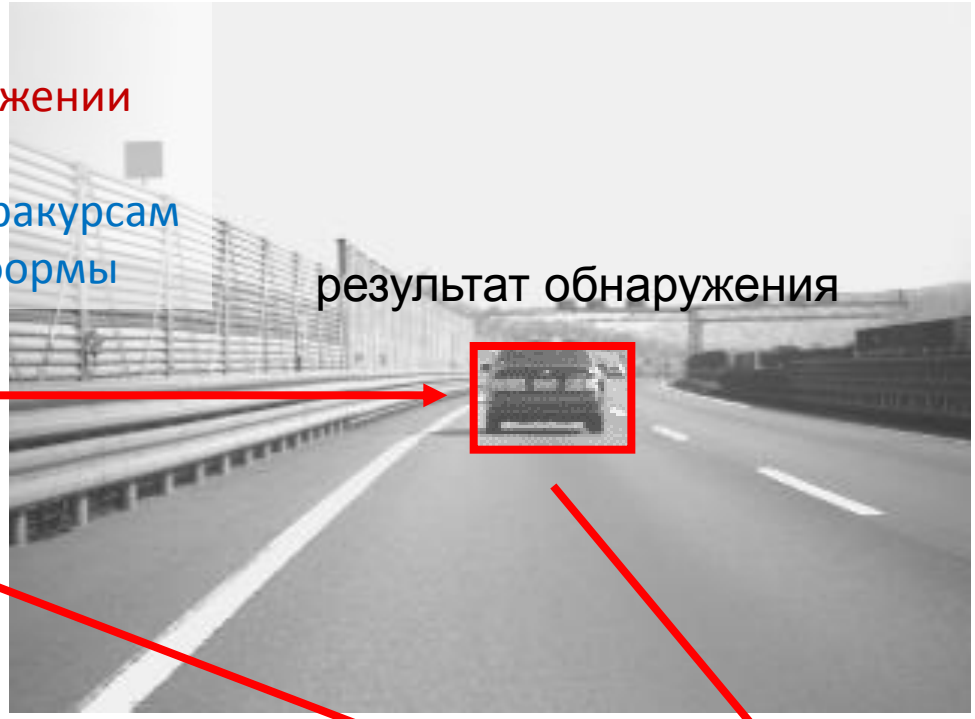
+ дает локализацию на изображении
+ работает по одному эталону
- не инвариантен к масштабу, ракурсам
- не допускает изменчивости формы

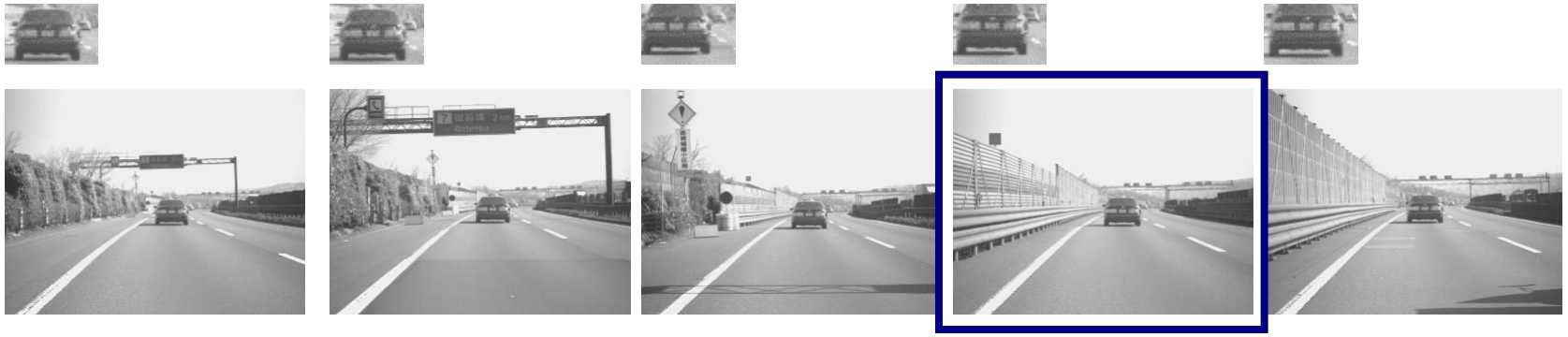
последовательность кадров

Корреляционное прослеживание объектов на видеопоследовательностях

Коррелятор:

+ дает локализацию на изображении
+ работает по одному эталону
- не инвариантен к масштабу, ракурсам
- не допускает изменчивости формы





последовательность кадров

Корреляционное прослеживание объектов на видеопоследовательностях

Основной недостаток корреляторов: проблема изменчивости эталона (потребность в множестве эталонов)

последовательность кадров

2. Семантический поиск (Image Retrieval) в коллекциях изображений при помощи ГКНС

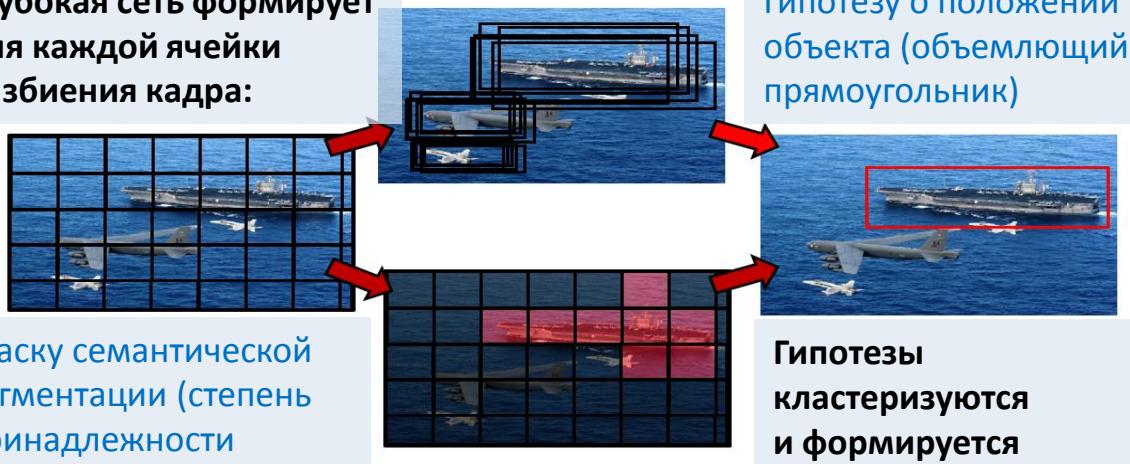
Запрос (эталон) → Список наиболее похожих кандидатов

Семантический поиск:

+ работает по одному эталону
+ инвариантен к масштабу, ракурсам, допускает изменчивость формы
- не дает локализацию на изображении

3. Обнаружение и распознавание объектов в реальном времени при помощи ГКНС

Глубокая сеть формирует для каждой ячейки разбиения кадра:



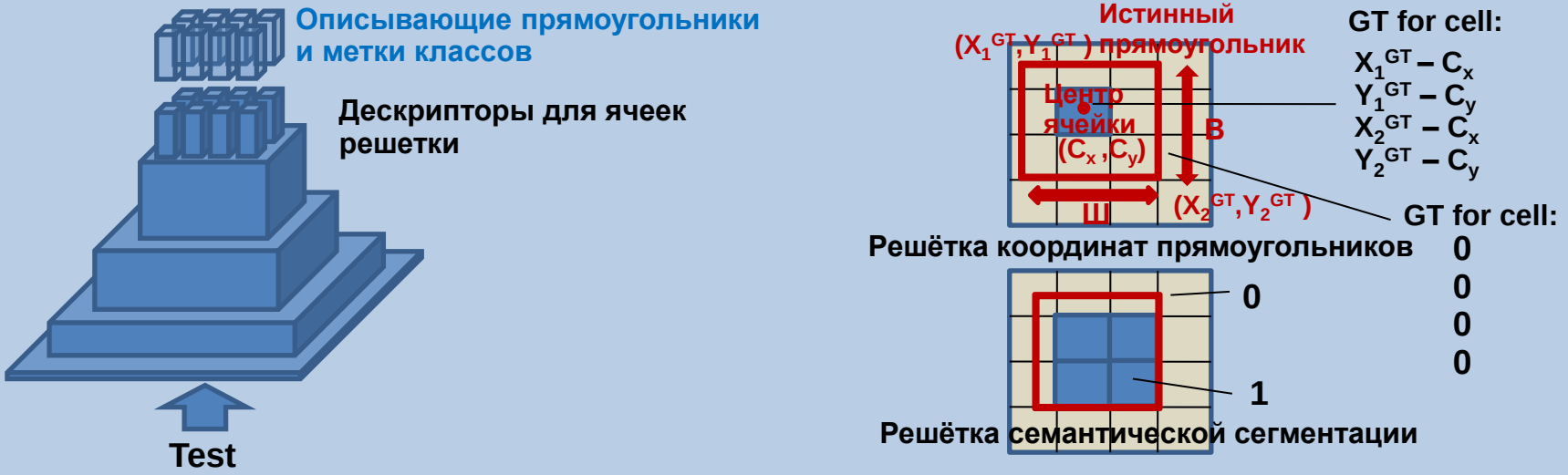
Гипотезу о положении объекта (объемлющий прямоугольник)

Маску семантической сегментации (степень принадлежности объектам классов)

Гипотезы кластеризуются и формируется окончательный список

ГКНС-детектор:

+ инвариантен к масштабу, ракурсам
+ допускает изменчивость формы
+ дает локализацию объекта на изображении
+ не работает по одному эталону, требует для обучения гигантской выборки

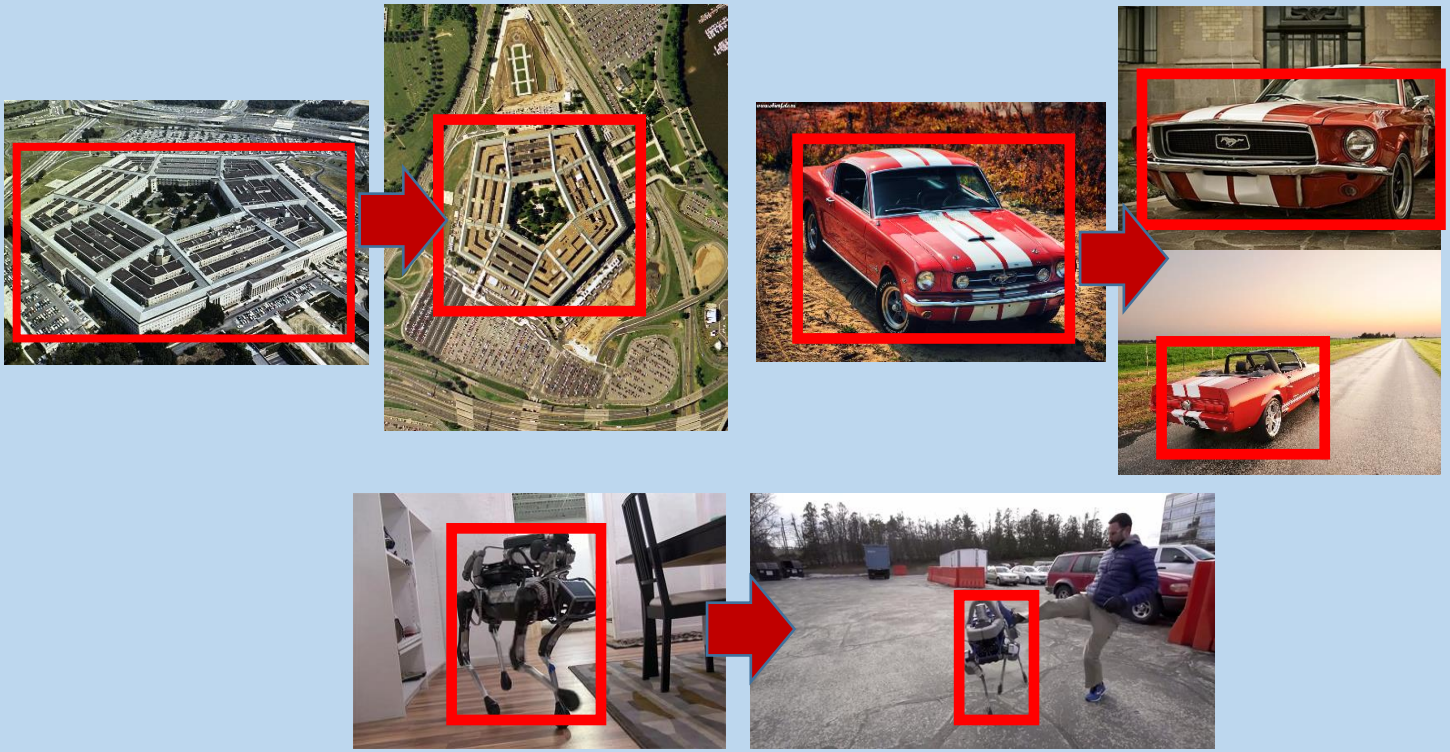


Однопроходный алгоритм обнаружения и распознавания объектов

«3 в 1»: Задача семантического поиска объектов на изображении

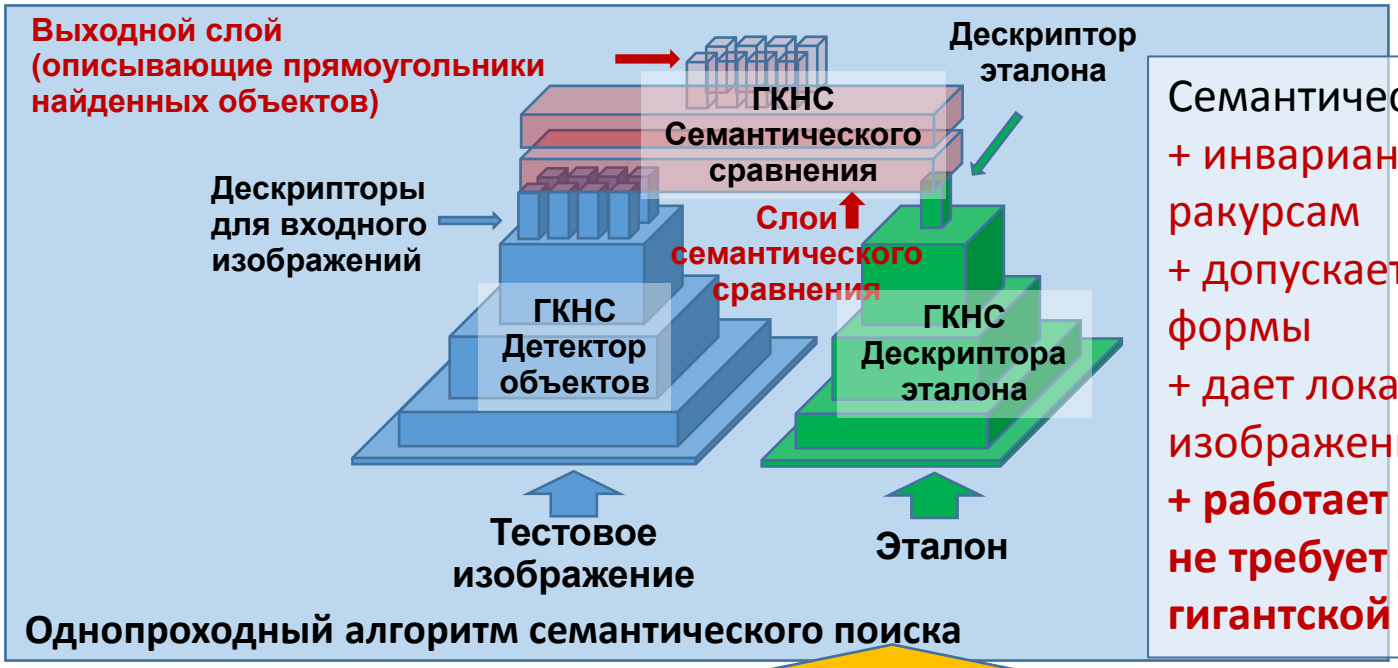
Задача семантического поиска:

Задачей семантического сравнения назовем процедуру обнаружения объекта класса по заданному эталону класса (изображению либо набору изображений).



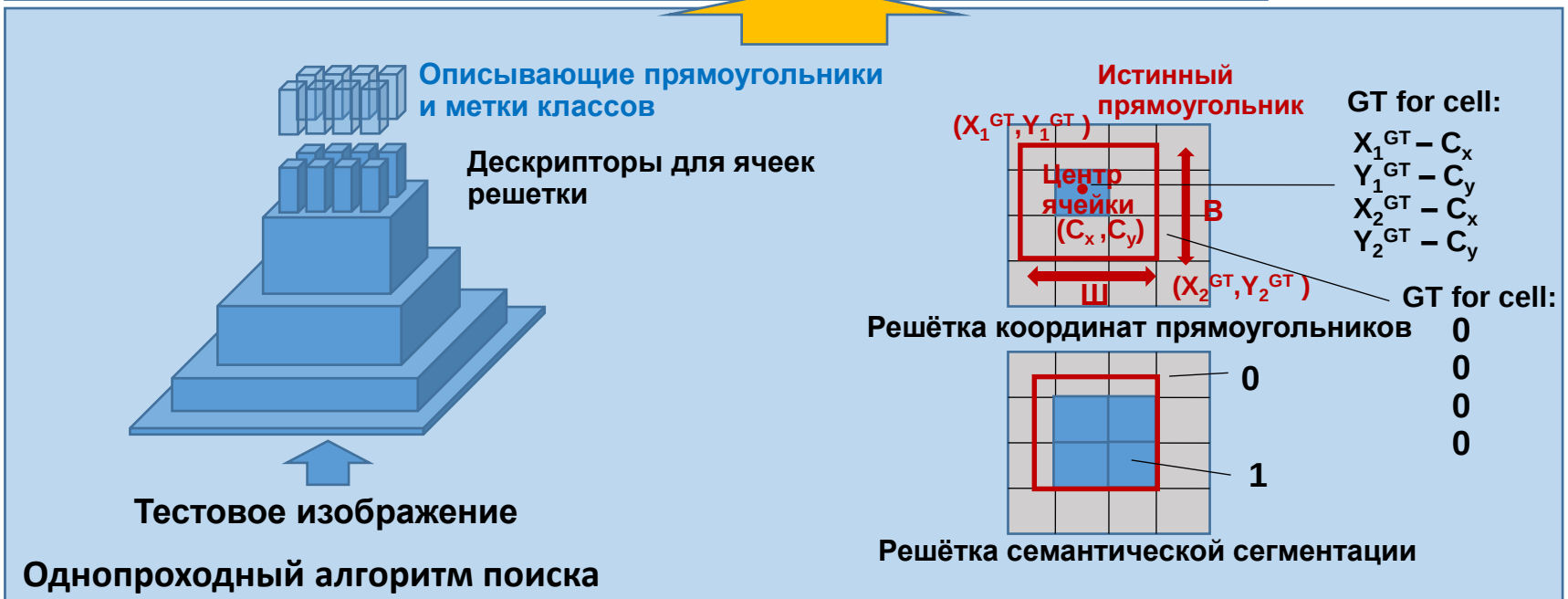
Семантический поиск – поиск объектов того-же класса при условии существенных изменений как в положении и в условиях съемки, так и в структуре самого объекта (например, объект изменил конфигурацию)

«3 в 1»: ГКНС для семантической корреляции



Семантический коррелятор:

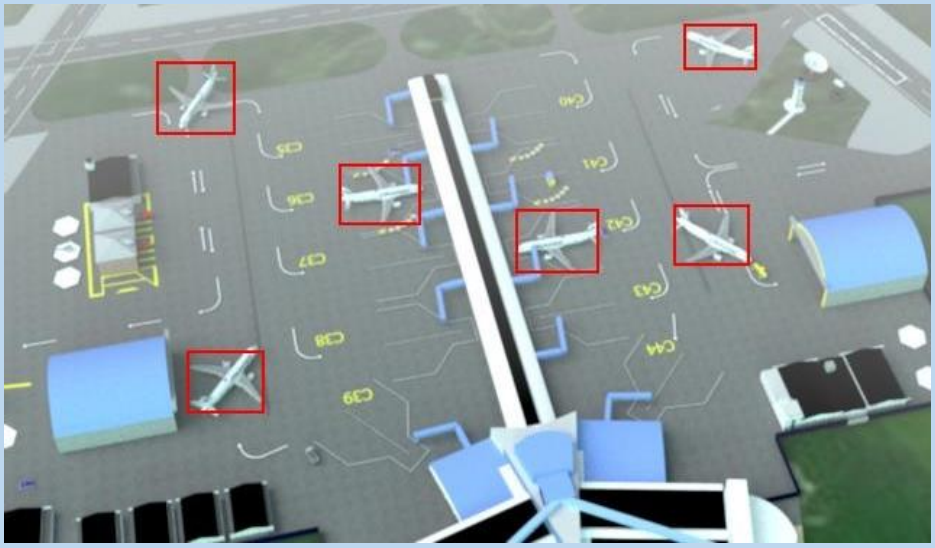
+ инвариантен к масштабу, ракурсам
+ допускает изменчивость формы
+ дает локализацию объекта на изображении
+ работает по одному эталону, не требует для обучения гигантской выборки



Алгоритм семантического поиска изображений

Эталон

Тестовое изображение



Качество работы семантического коррелятора

Обучение:

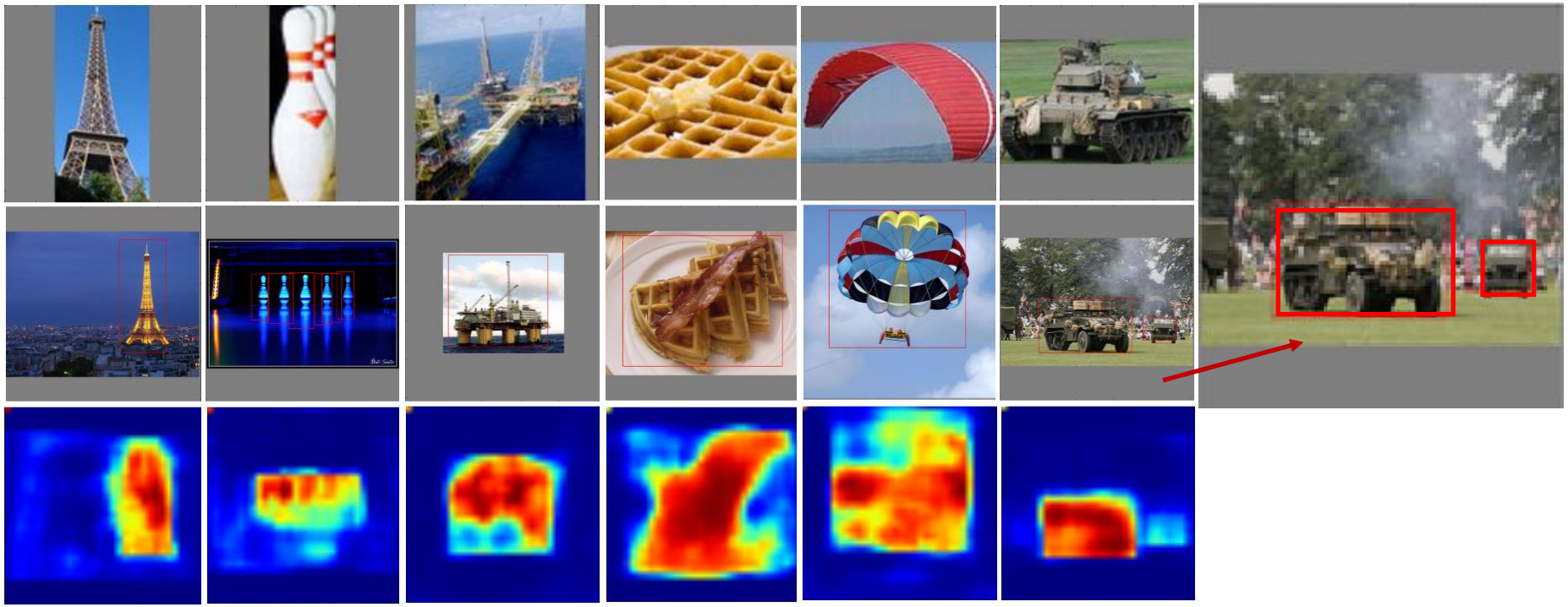
- База изображений ILSVRC 2014 – 456567 изображений, 200 категорий объектов с отмеченными объемлющими прямоугольниками
- 512x512 – размер входного изображения для обнаружения объекта
- 128x128 – размер изображения-запроса (эталона, характеризующего класс)

Тестирование:

- База изображений ILSVRC 2014 – 200 известных and 90 неизвестных классов объектов

Достигнутый результат:

- **Качество обнаружения mAP для неизвестных классов составляет 85% mAP для известных классов!**



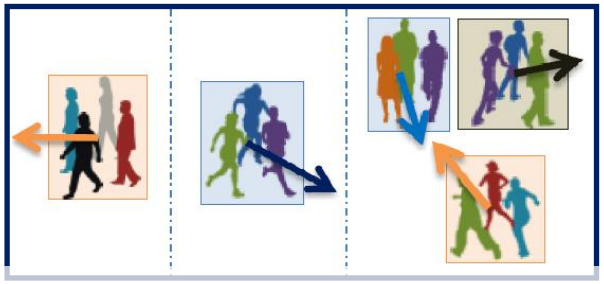
Качественные примеры работы семантического коррелятора: изображение-запрос, тестовое изображение и корреляционное поле

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, ?l), 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, ?l) -> 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, ?l), 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,

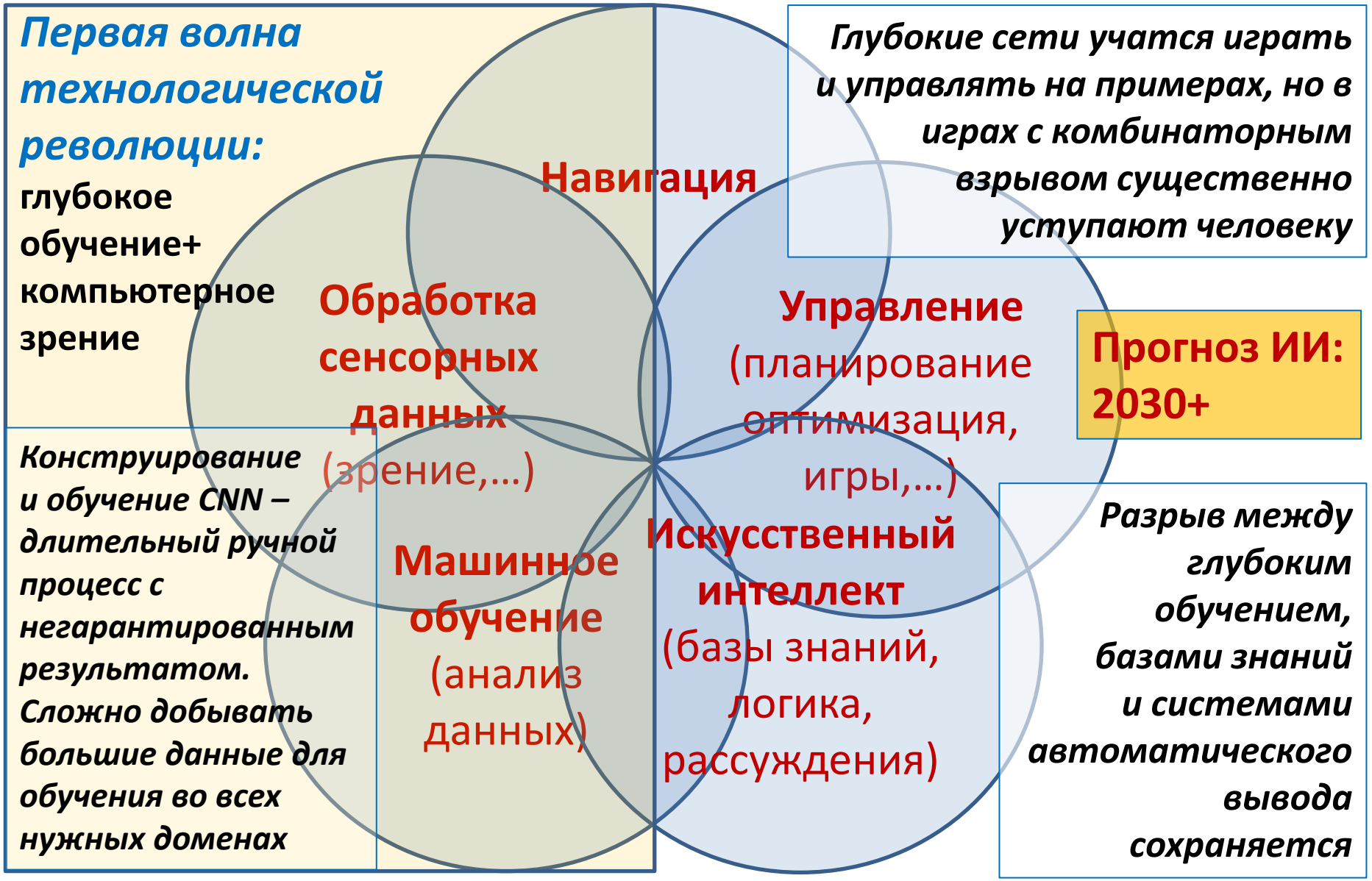
которая была бы способна ответить на следующие основные вопросы:

- **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 ГКНС 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



Перспективные методы и направления глубокого обучения

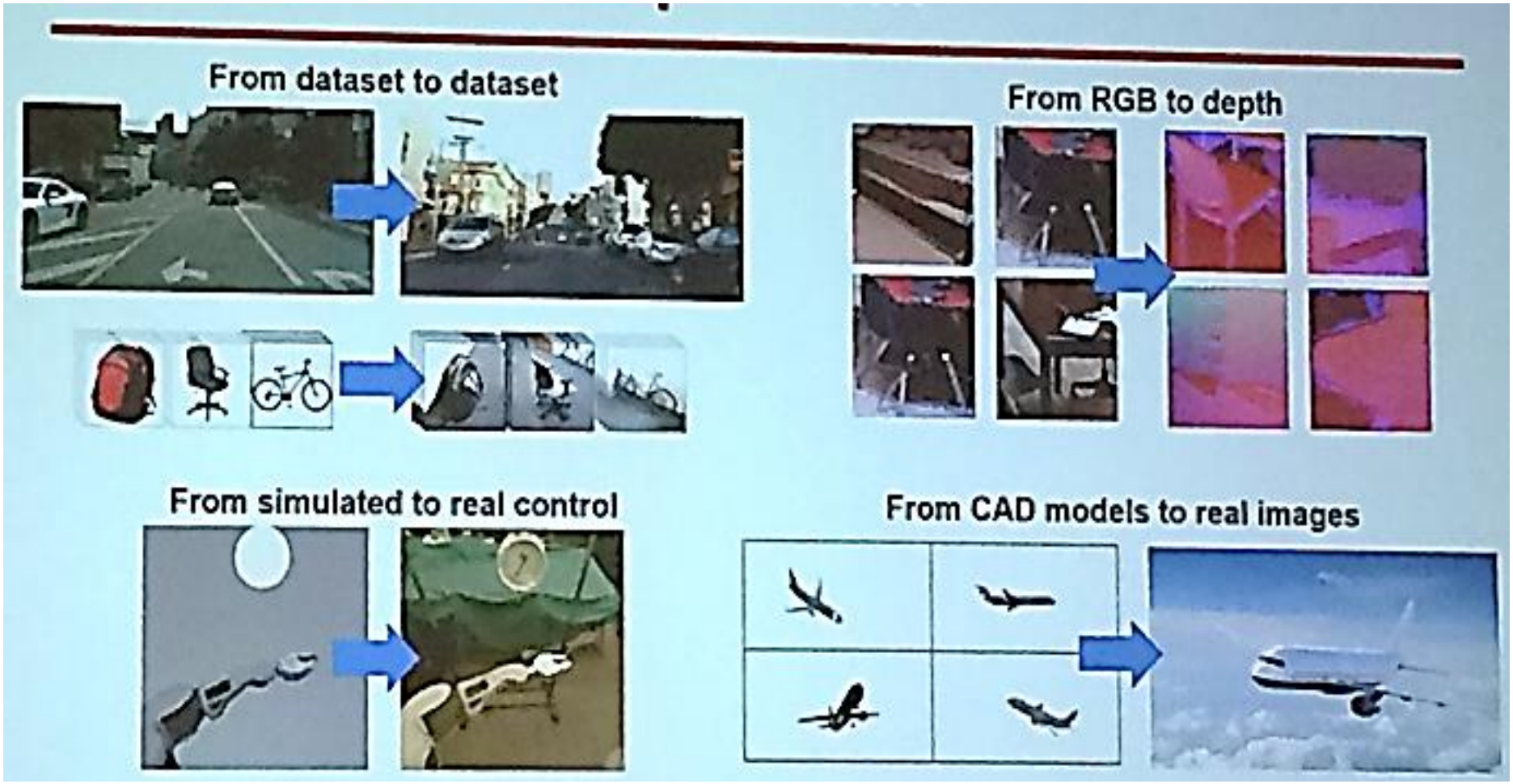
Второй этап современной революции в машинном обучении и анализе данных (2017+)

Компьютерное зрение и машинное 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**
  Reinforsement Learning, Lifelong Learning
- **Глубокое обучение с использованием структурных моделей, баз знаний и программ логического вывода**
  Graph Structured CNN, Deep Visual Reasoning
- **CNN для стратегических игр с комбинаторным взрывом**
- **Автоматическое конструирование и обучение глубоких сетей**

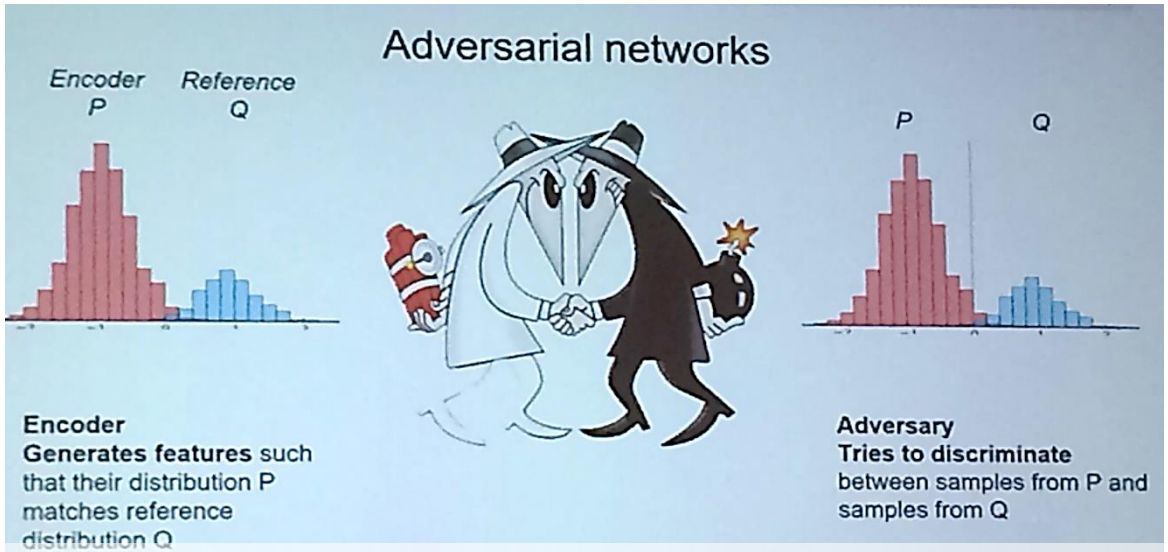
Domain Transfer Learning



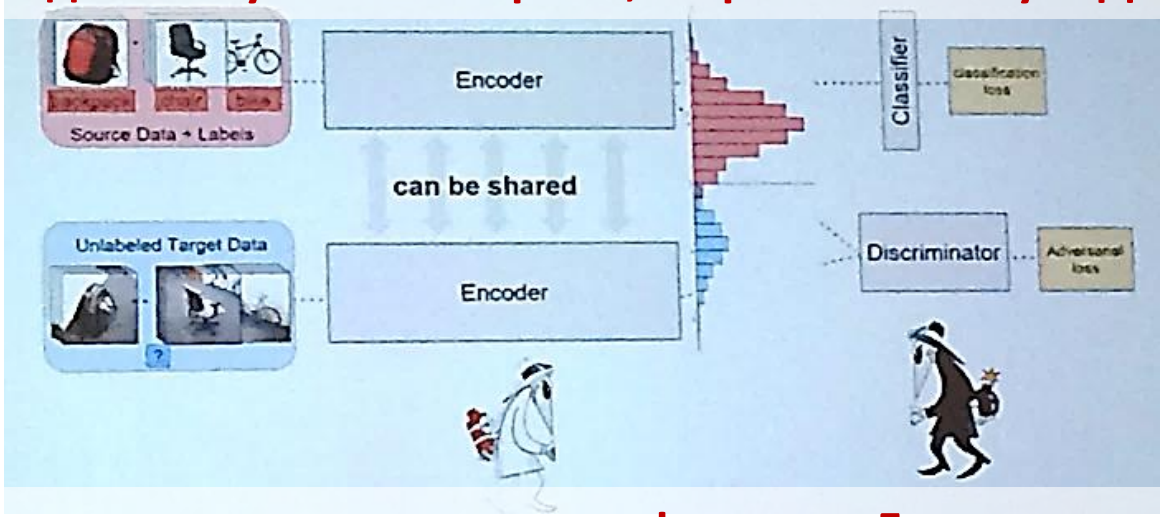
Перенос выученных закономерностей в новую область применения

Adaptive Deep Models across Visual Domains, Kate Saenko, ICCV, 2017

Генеративные конкурирующие сети



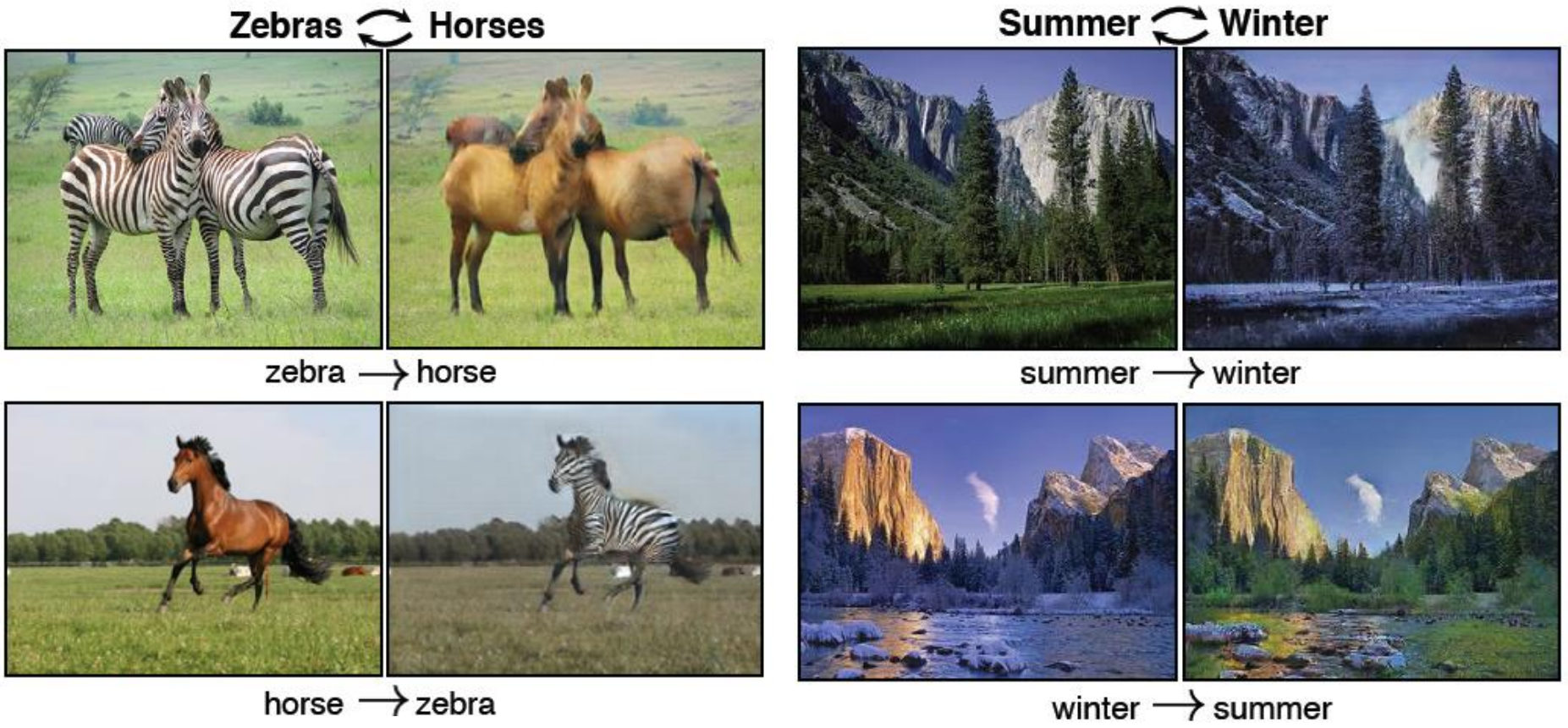
Генератор создает визуальные образы, стараясь обмануть Дискриминатор…



….Дискриминатор старается отличить фантазии Генератора от реальности

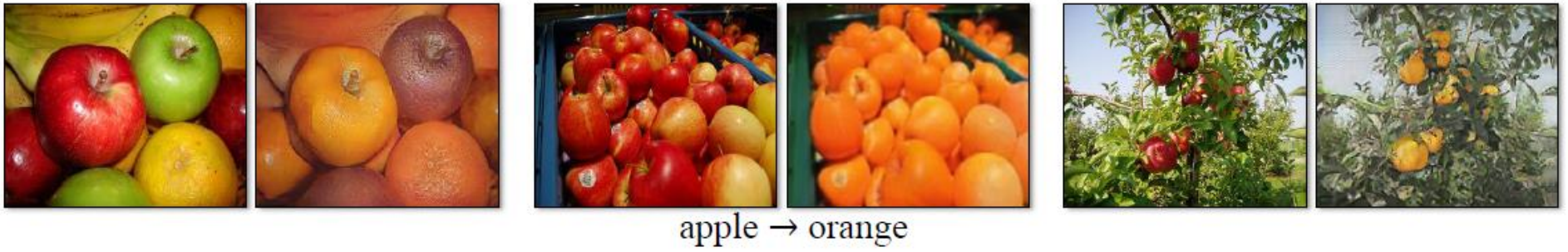
Adaptive Deep Models across Visual Domains, Kate Saenko, ICCV, 2017

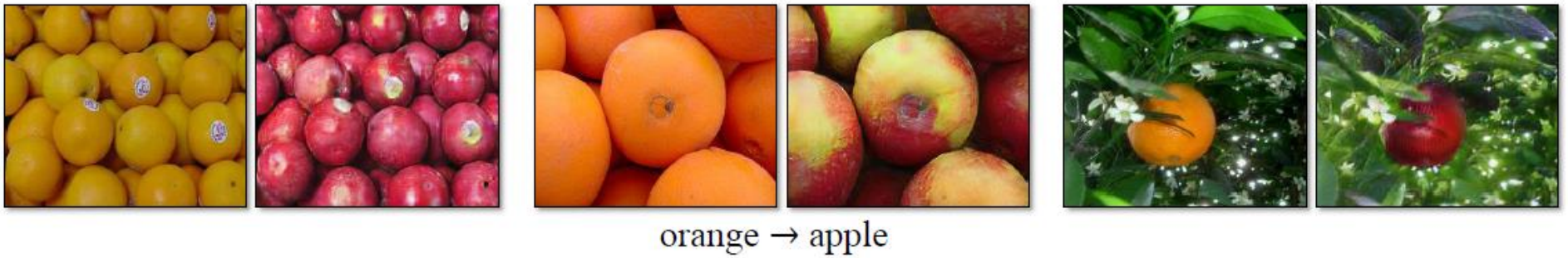
Generative Adversarial Networks (GANs)



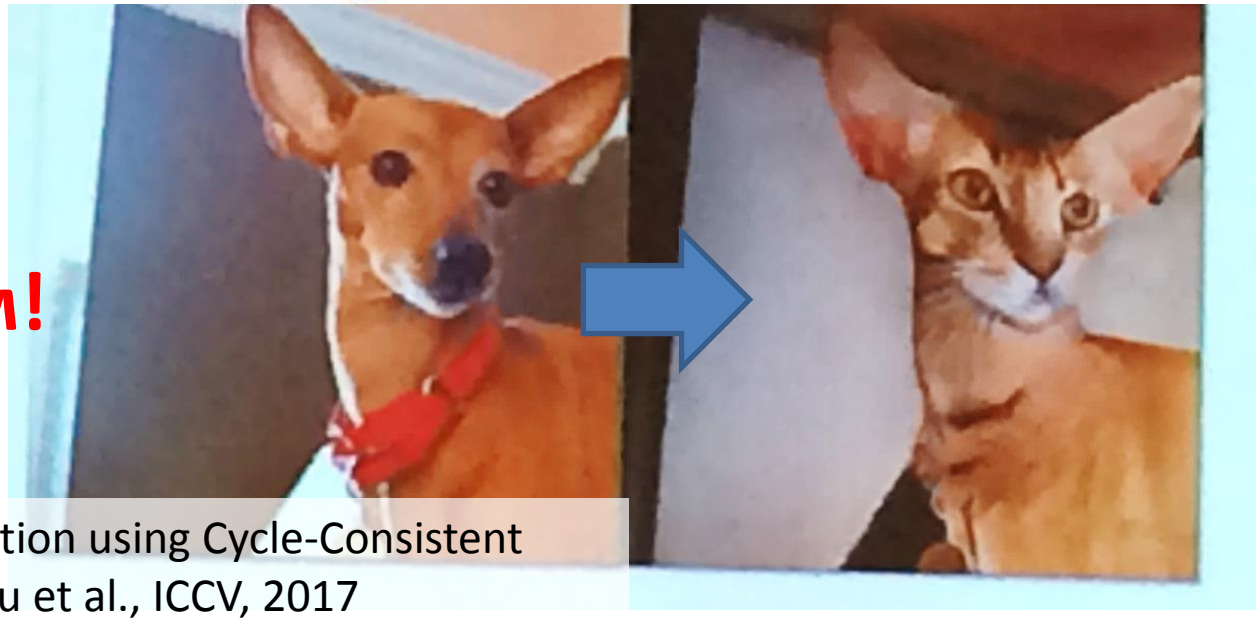
Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, Jun-Yan Zhu et al., ICCV, 2017

Generative Adversarial Networks (GANs)



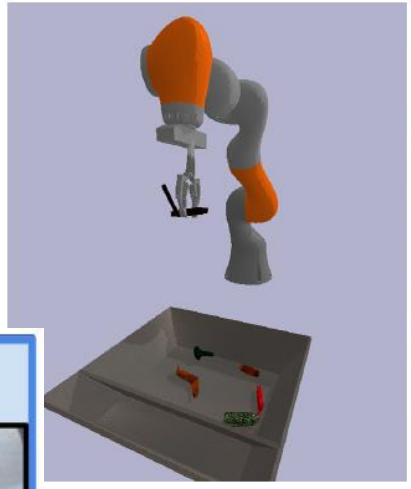


GAN – сеть, обладающая воображением!

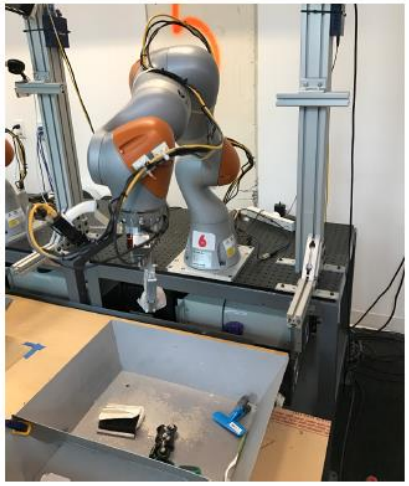


Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, Jun-Yan Zhu et al., ICCV, 2017

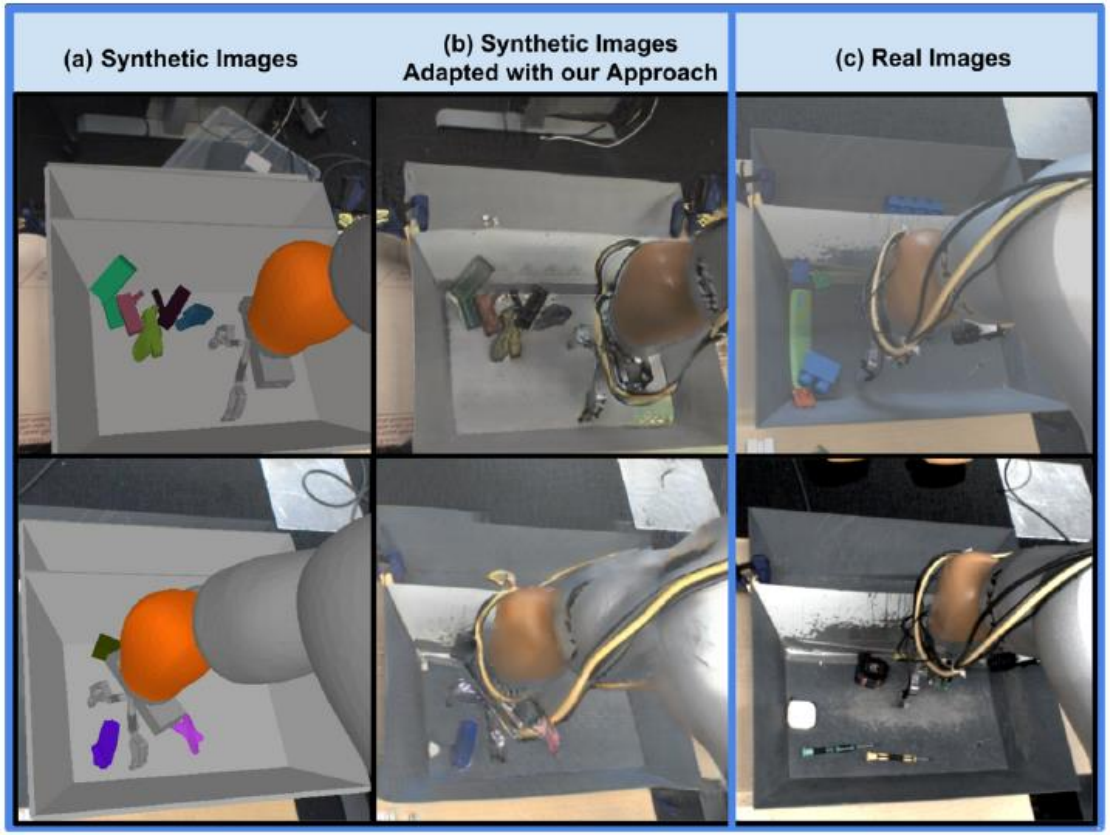
Deep Robot Grasping with GAN



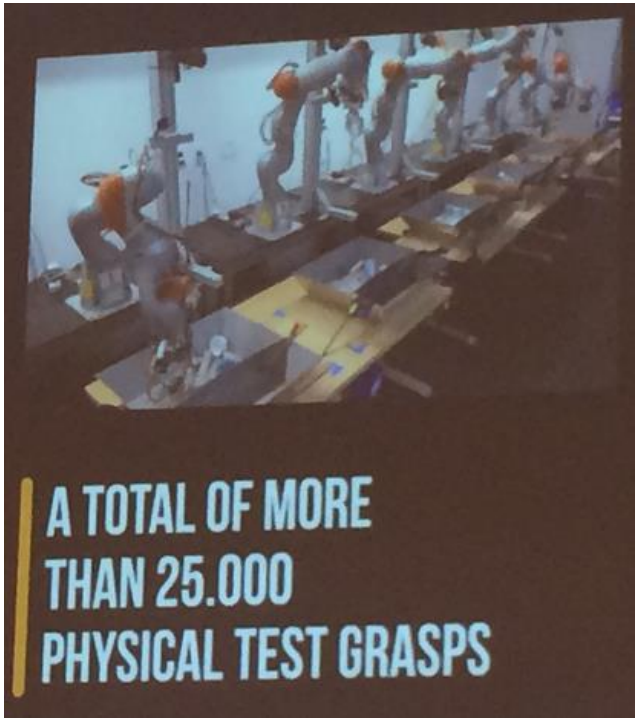
(a) Simulated World



(b) Real World



Using Simulation and Domain Adaptation to Improve Efficiency of Deep Robotic Grasping, Bousmalis et al., 2017



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

реальное ТВ изображение

реальное ИК изображение

ИК изображение, сгенерированное ГКНС

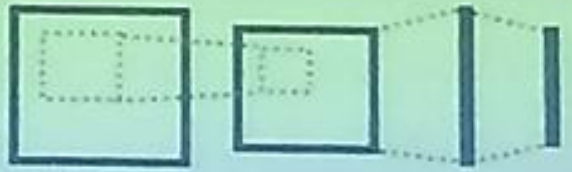
реальное ИК изображение

реальное ТВ изображение

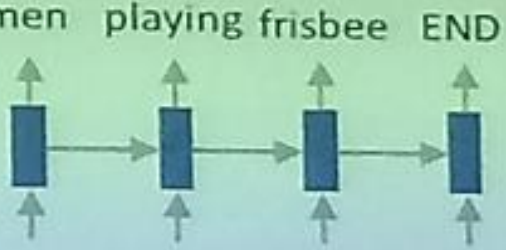
ТВ изображение, сгенерированное ГКНС

Visual Description of Image and Video

Visual Description: Image Captioning



Convolutional Network



Recurrent Network

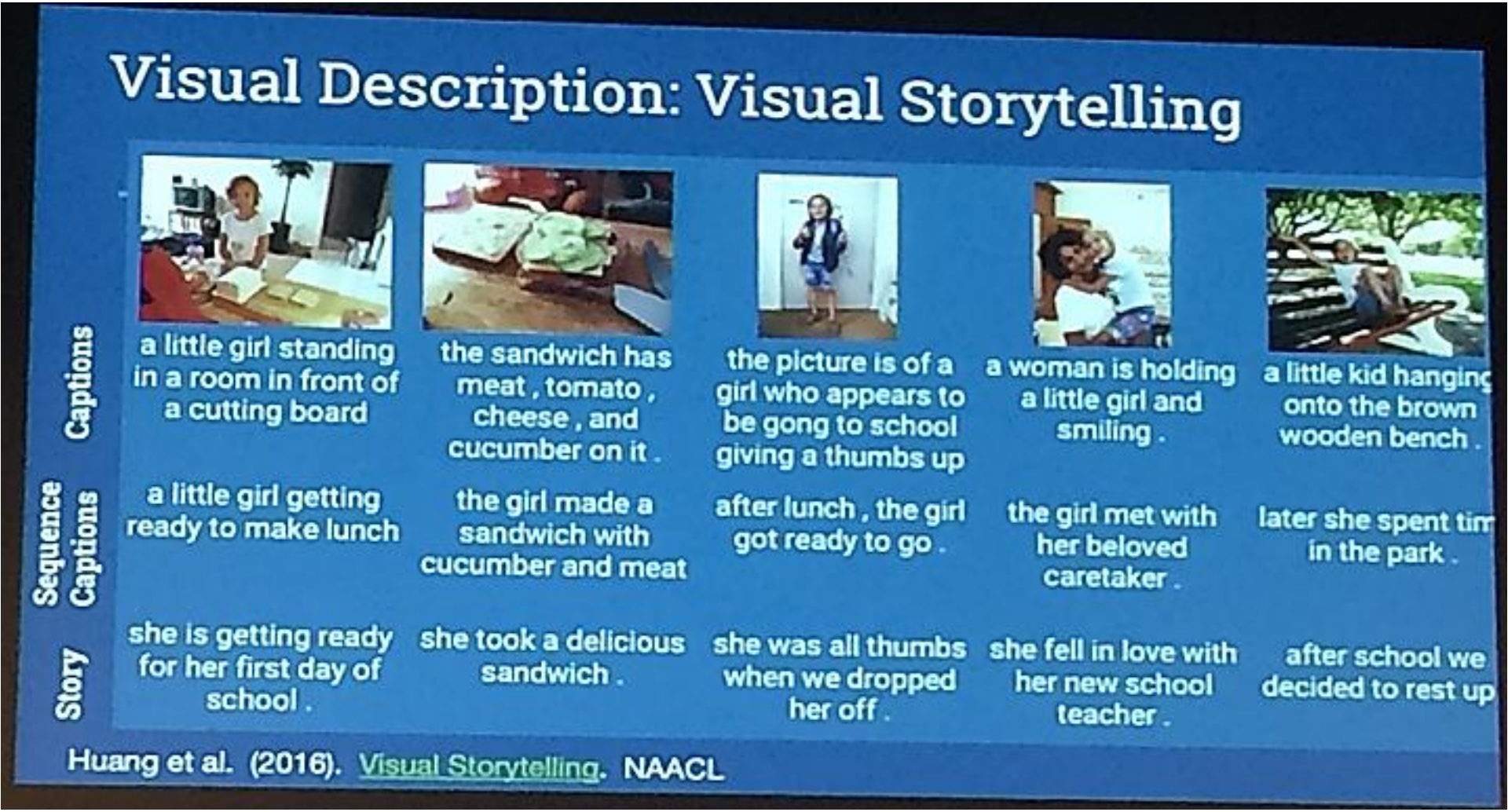
Visual Description: Paragraph Description

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

Closing the Loop Between Video and Language, ICCV Workshop, 2017

Visual Description of Image and Video



Closing the Loop Between Video and Language, ICCV Workshop, 2017

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?

VQA System

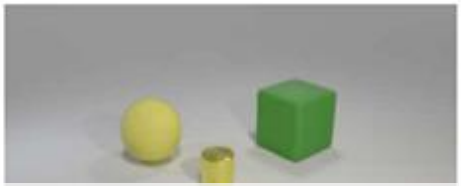
Happy

How does the person in the middle feel?

Вопросы самых различных типов

Вопросы, требующие понимания контекста

CLEVR Dataset



Вопросы, требующие рассуждений

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

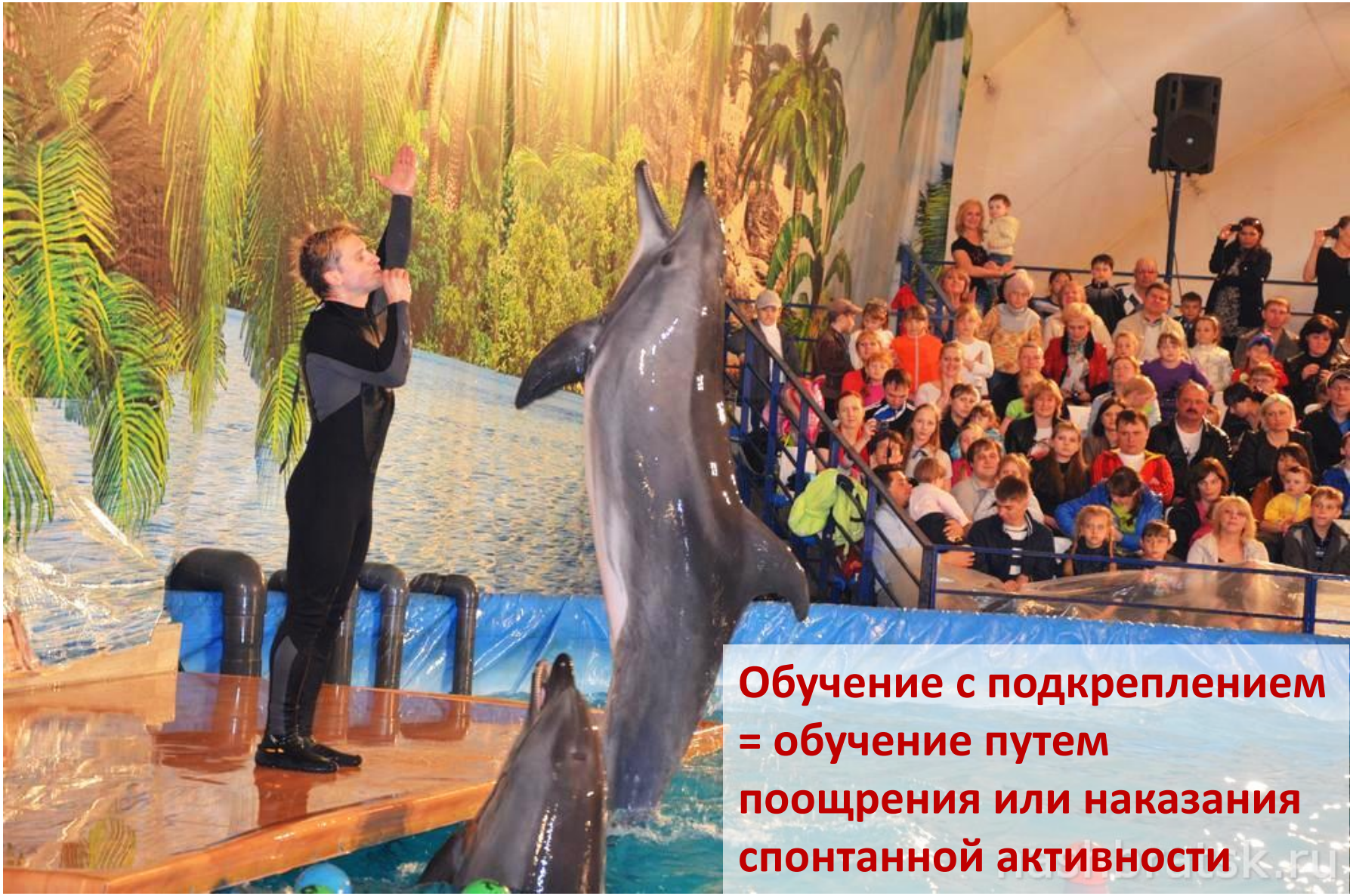
Q: *What shape object is farthest right?* **A:** *cylinder*

Q: *Are all the balls small?* **A:** *no*

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

Closing the Loop Between Video and Language, ICCV Workshop, 2017

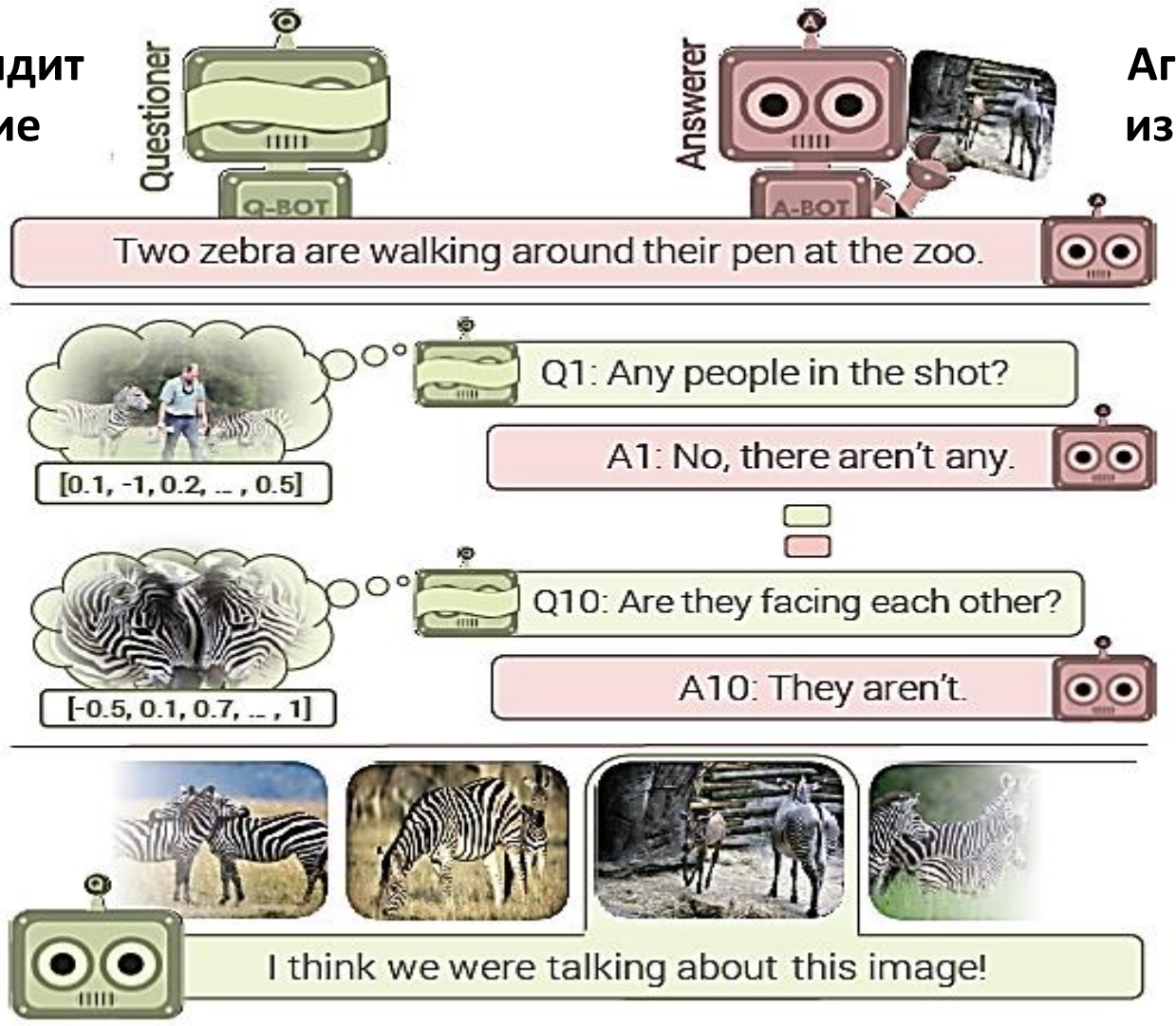
Reinforcement Learning = Оперантное научение роботов



Понимание видеоинформации как кооперативная игра двух агентов, ведущих диалог

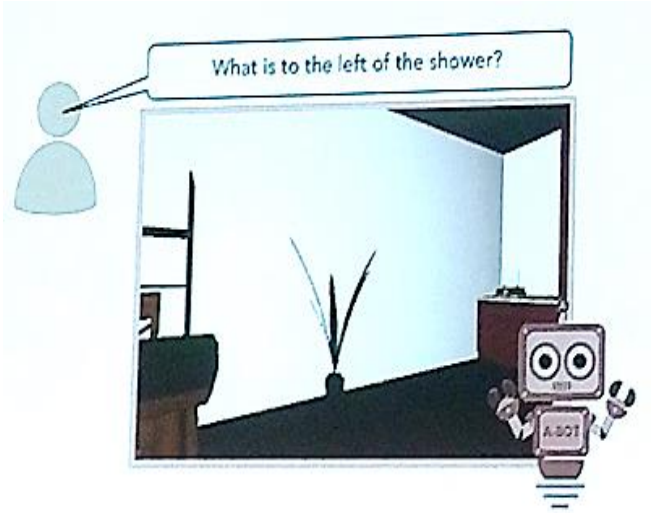
Агент Q не видит изображение

Агент A видит изображение



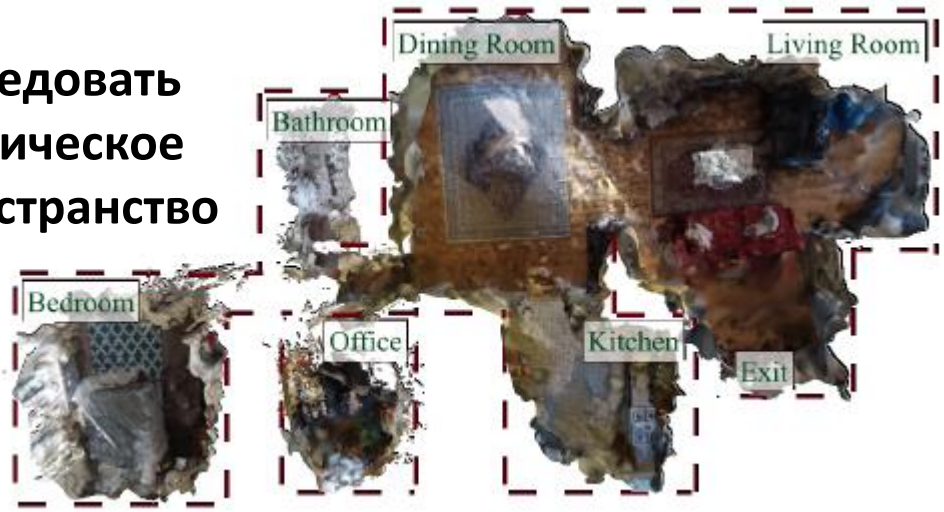
Learning Cooperative Visual Dialog Agents with Deep Reinforcement Learning, Abhishek Das, Satwik Kottur, José M. F. Moura, Stefan Lee, Dhruv Batra, ICCV, 2017

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



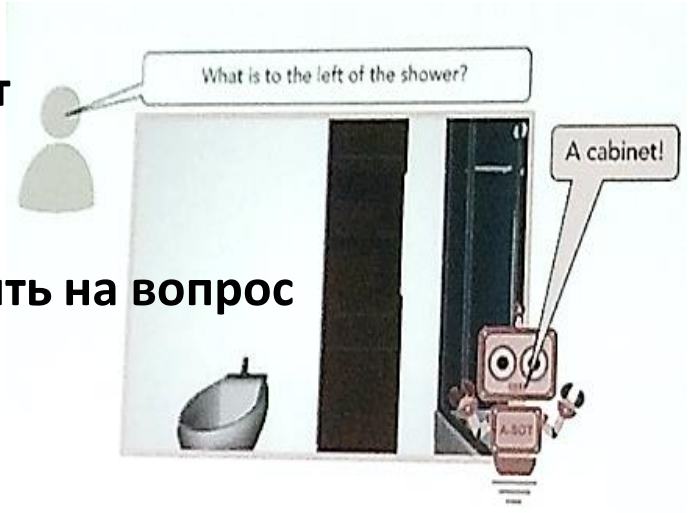
1. Понять вопрос

2. Исследовать семантическое 3D пространство



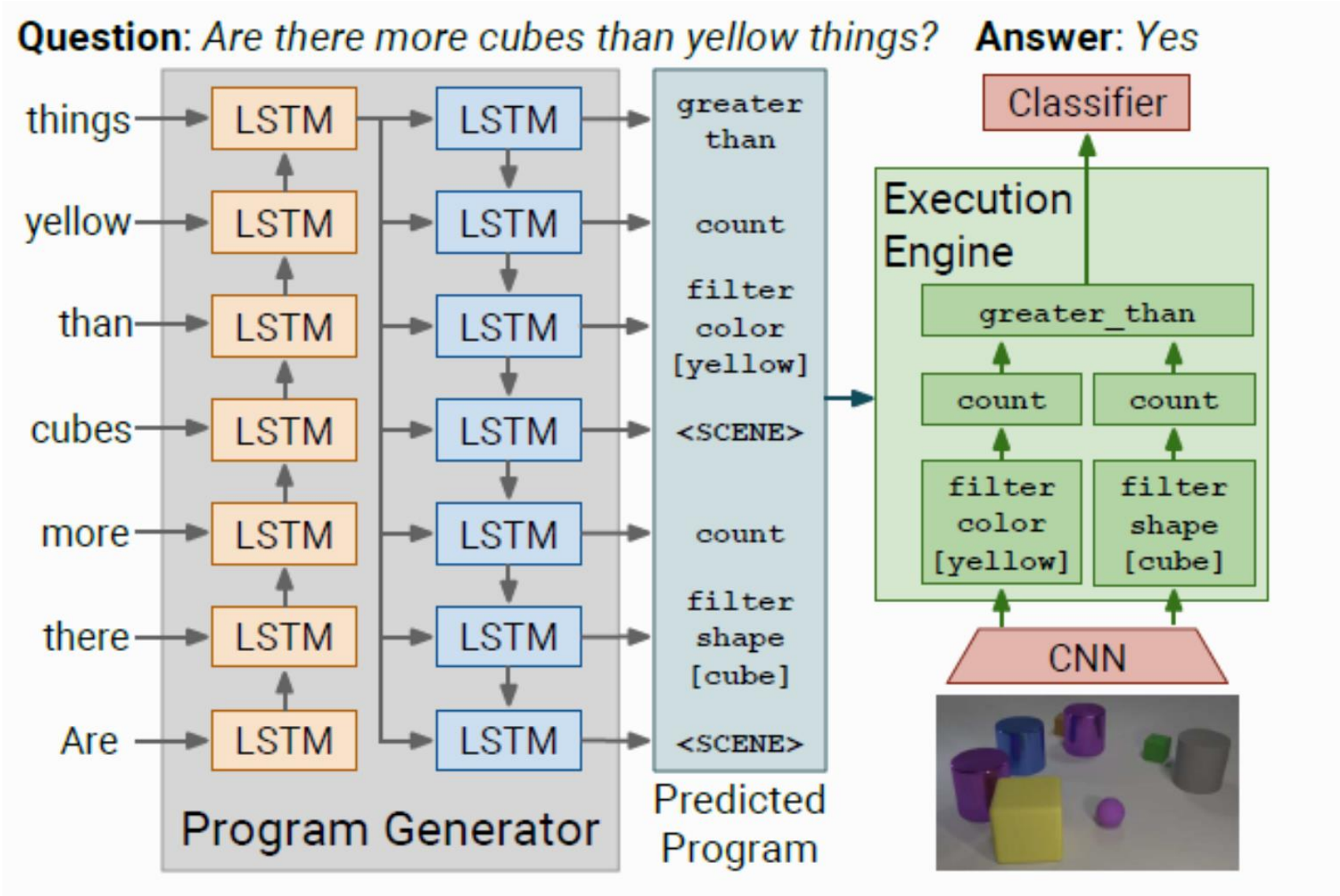
3. Спланировать маршрут

4. Ответить на вопрос



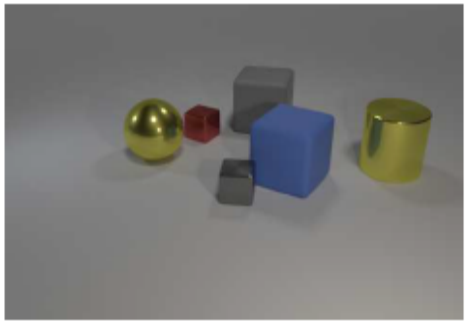
Learning Cooperative Visual Dialog Agents with Deep Reinforcement Learning, Abhishek Das, Satwik Kottur, José M. F. Moura, Stefan Lee, Dhruv Batra, ICCV, 2017

Deep Visual Reasoning for VQA: генератор программ



Inferring and Executing Programs for Visual Reasoning, Johnson et al., ICCV, 2017

Deep Visual Reasoning for VQA: генератор программ

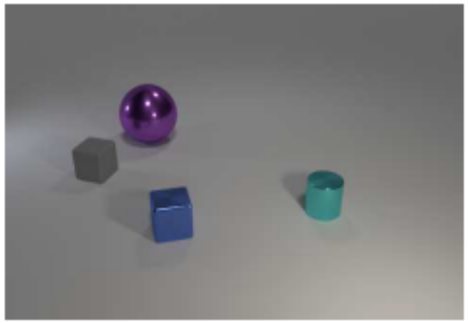


Q: *Is there a blue box in the items?* **A:** *yes*

Predicted Program:

```
exist
filter_shape[cube]
filter_color[blue]
scene
```

Predicted Answer:
✓ *yes*

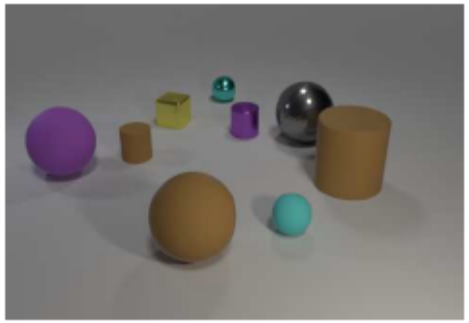


Q: *What shape object is farthest right?* **A:** *cylinder*

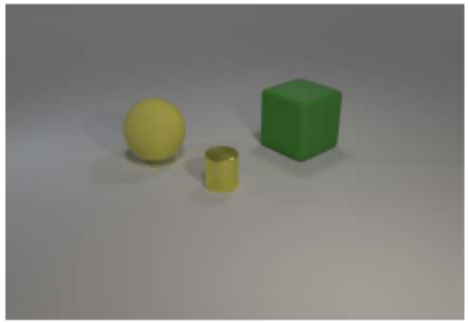
Predicted Program:

```
query_shape
unique
relate[right]
unique
filter_shape[cylinder]
filter_color[blue]
scene
```

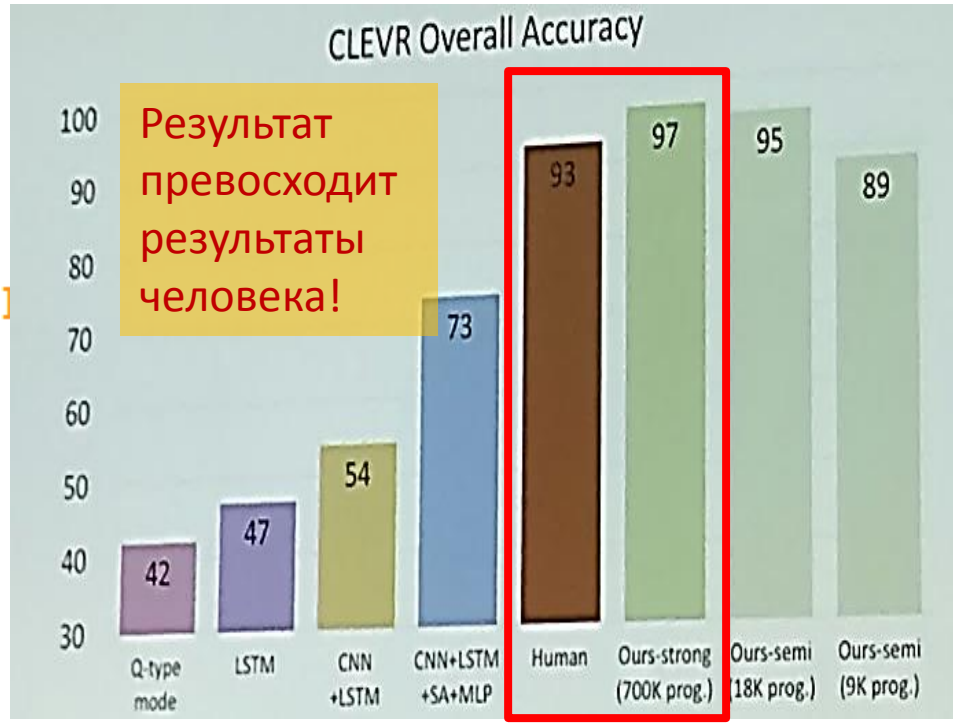
Predicted Answer:
✓ *cylinder*



Q: *Are all the balls small?* **A:** *no*

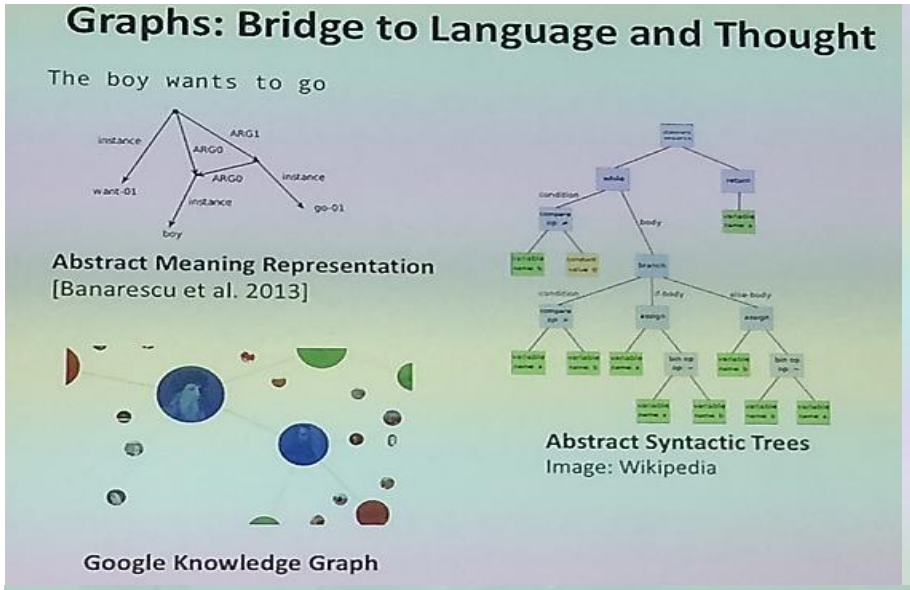


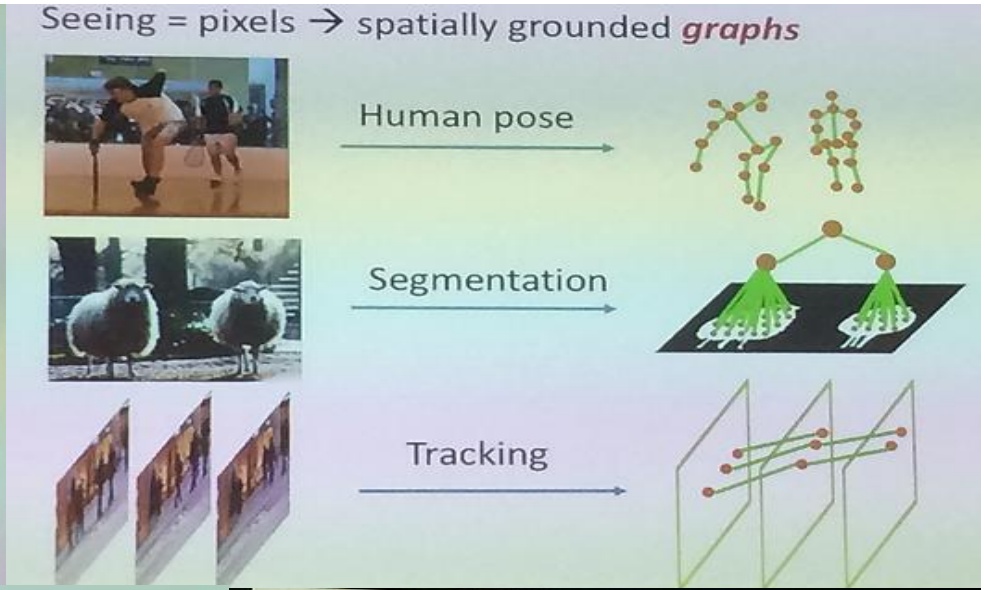
Q: *Is the green block to the right of the yellow sphere?* **A:** *yes*

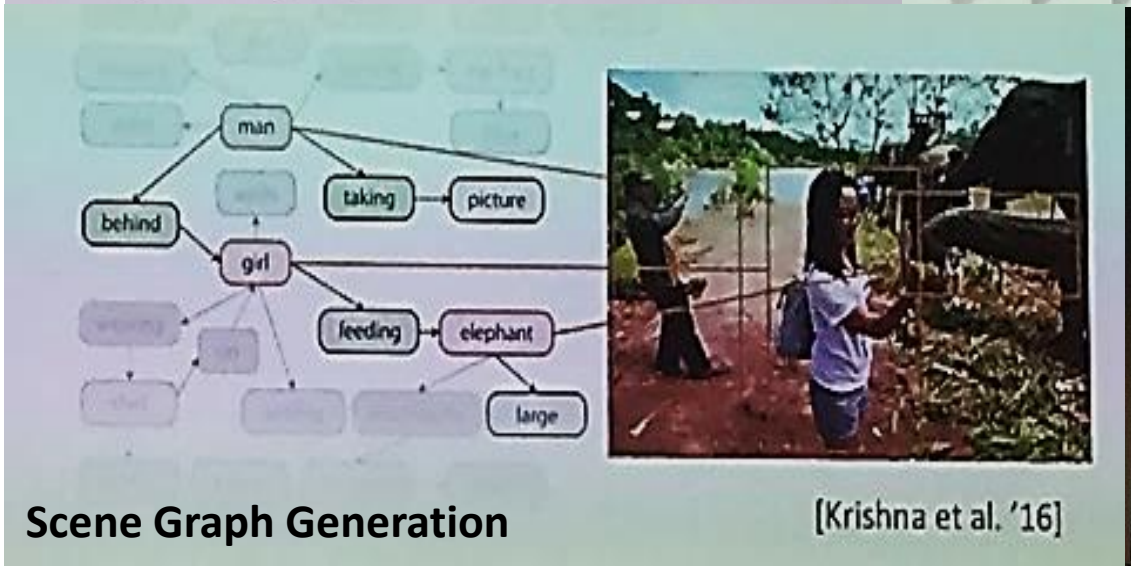


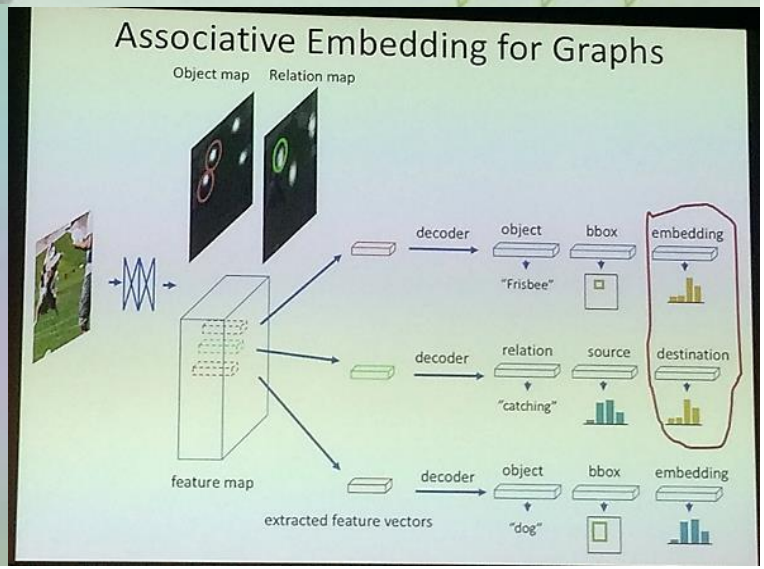
Inferring and Executing Programs for Visual Reasoning, Johnson et al., ICCV, 2017

Visual Reasoning with Graph Structured CNNs



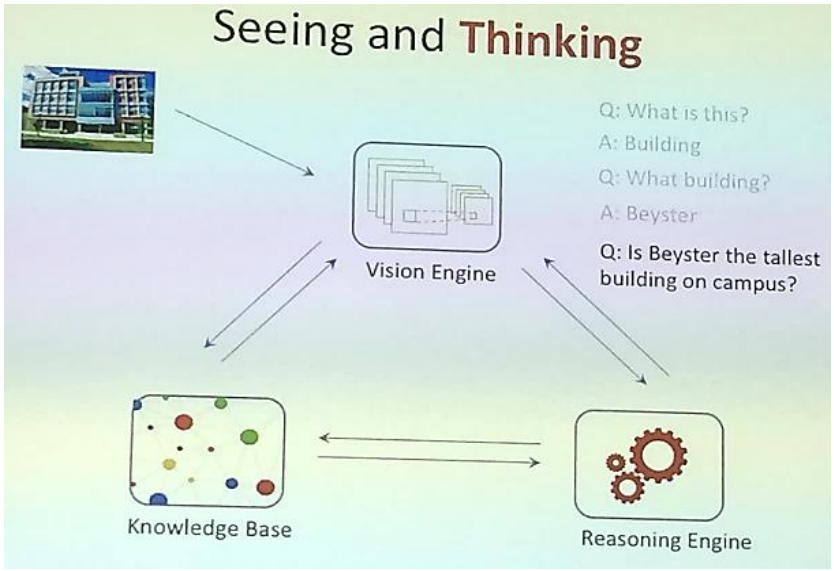


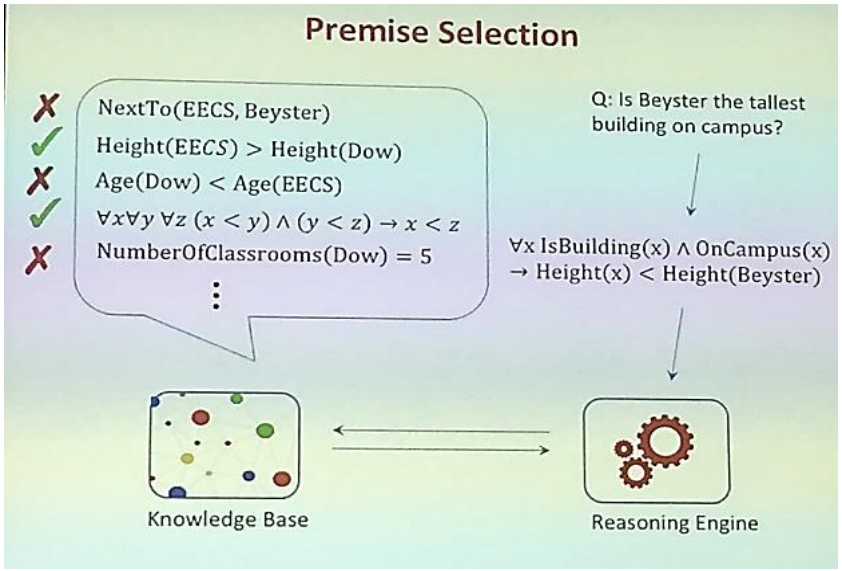


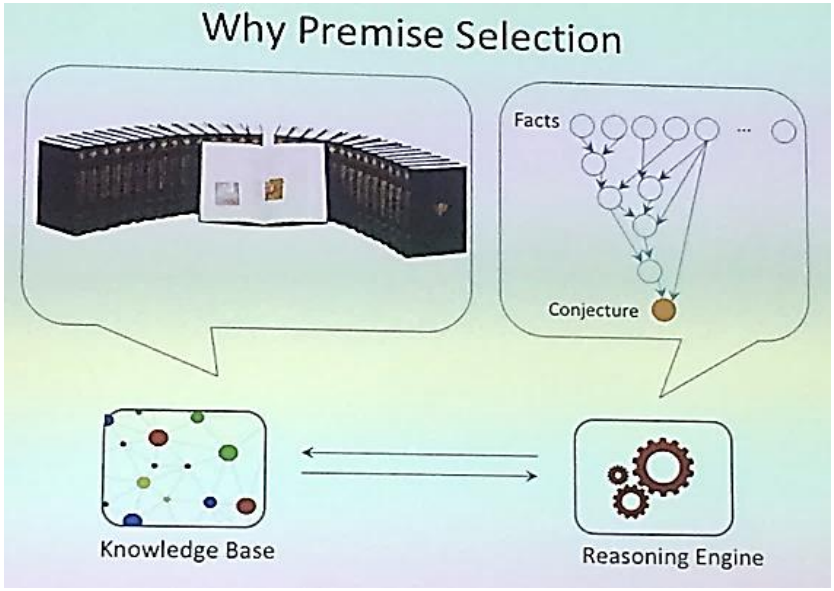


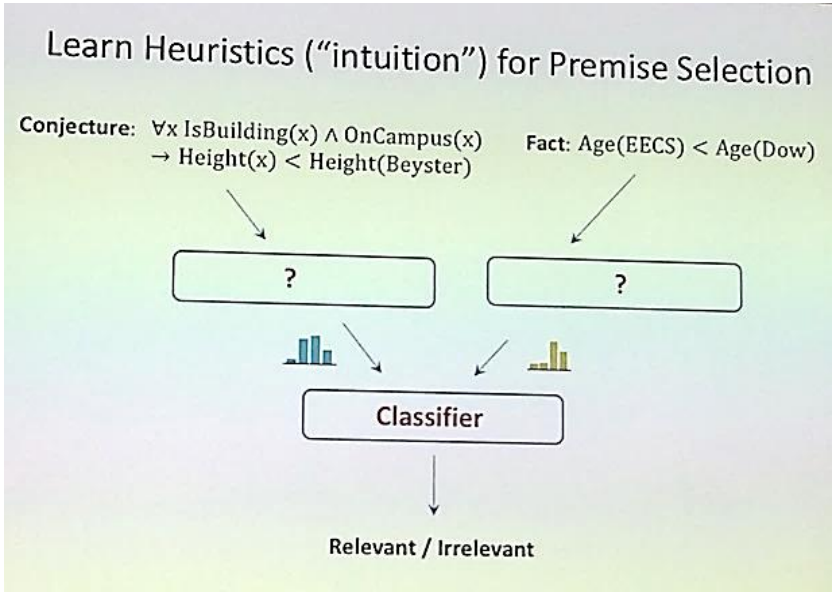
Seen and Thinking, ICCV, 2017

Visual Reasoning with Graph Structured CNNs



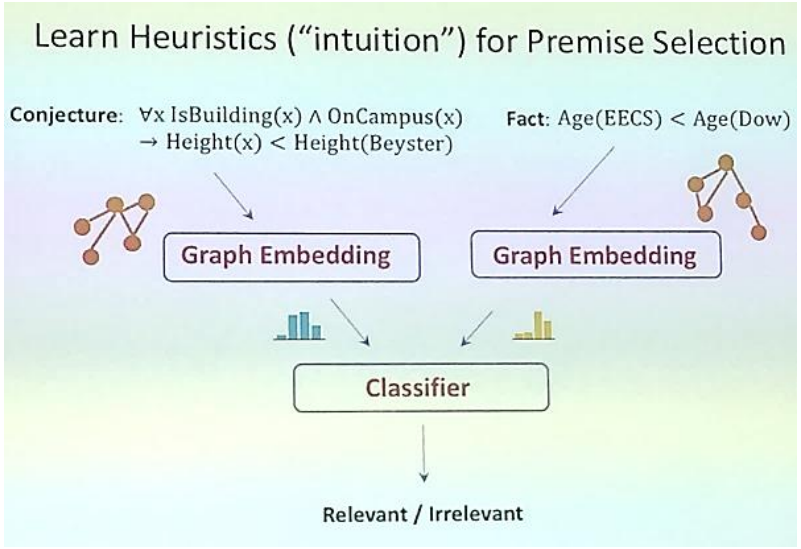


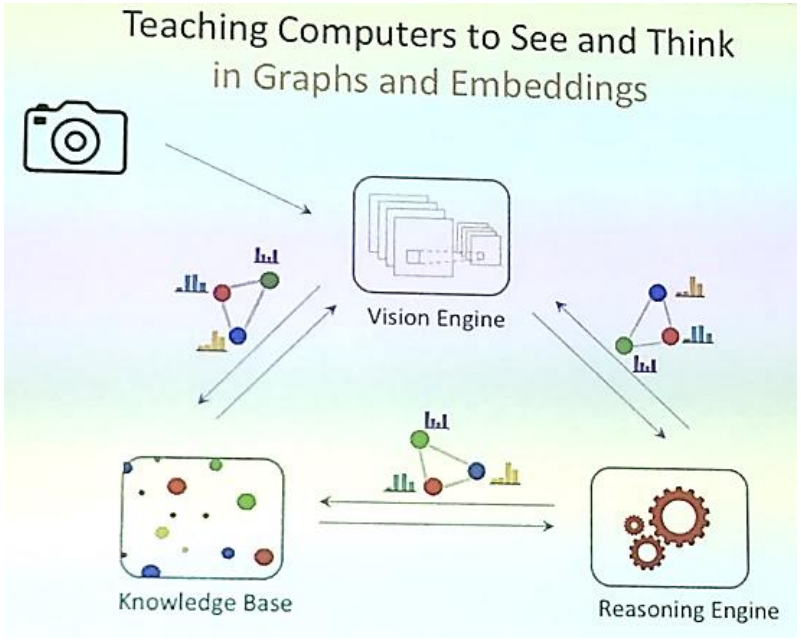




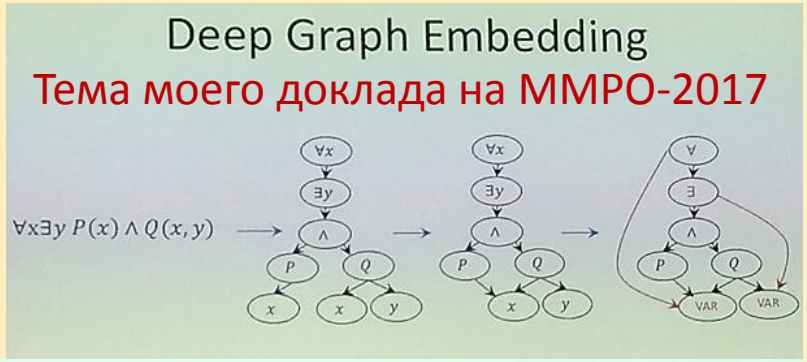
Seen and Thinking, ICCV, 2017

Visual Reasoning with Graph Structured CNNs





Seen and Thinking, ICCV, 2017



HolStep [Kaliszyk et al. 2017]

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

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

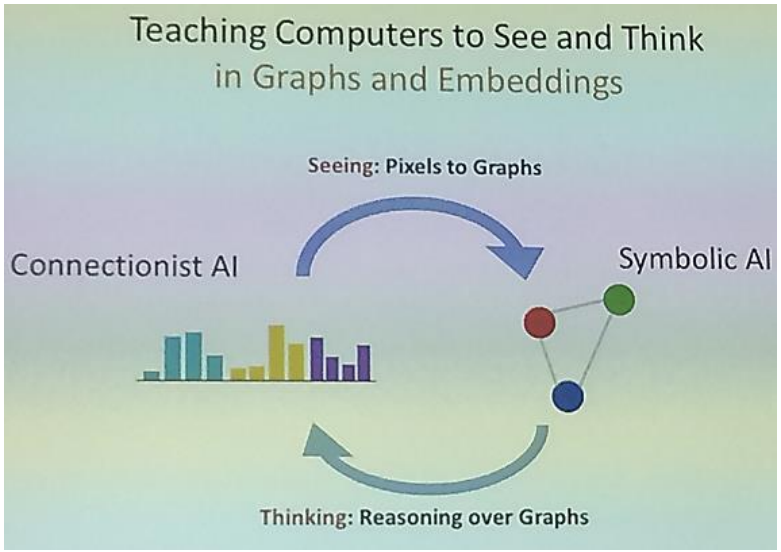
Relevant fact: $\forall\alpha\forall\beta\ sin(\alpha - \beta) = sin(\alpha)cos(\beta) - sin(\beta)cos(\alpha)$

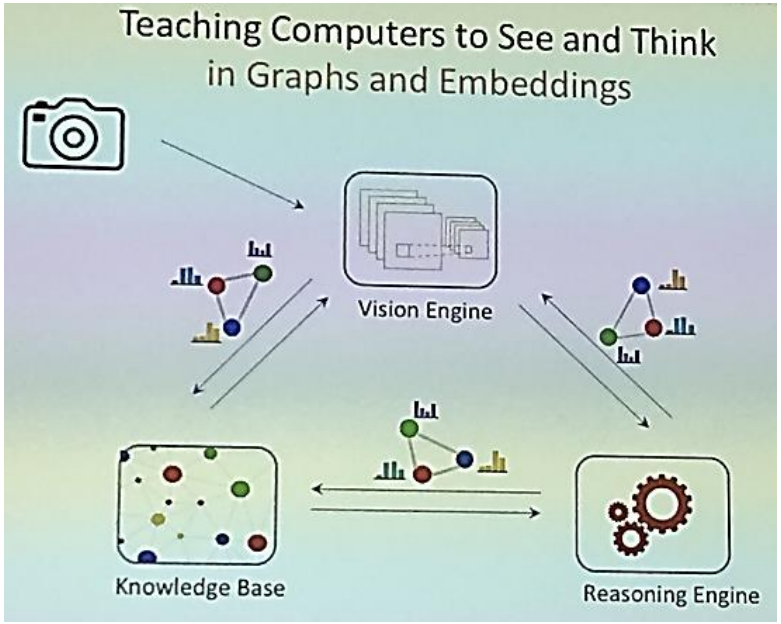
Irrelevant fact: $(x > 0) \wedge (y > 0) \rightarrow (xy > 0)$

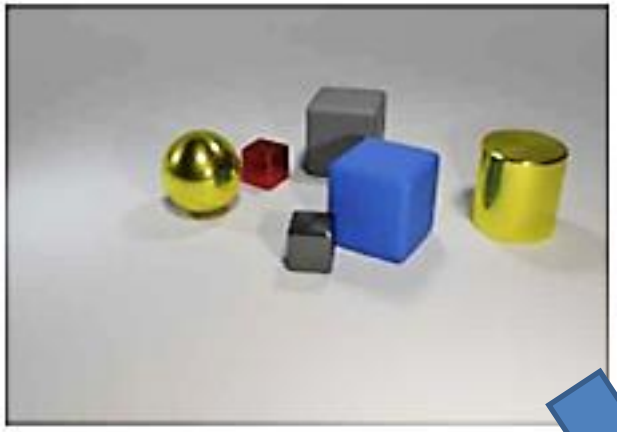
	Sequence embedding		*Graph embedding*
	CNN [Kaliszyk et al. ’17]	CNN-LSTM [Kaliszyk et al. ’17]	Ours
Accuracy	82	83	90.3

Не только визуальные задачи!
CNN показывают лучшие результаты в автоматическом доказательстве теорем

Visual Reasoning with Graph Structured CNNs







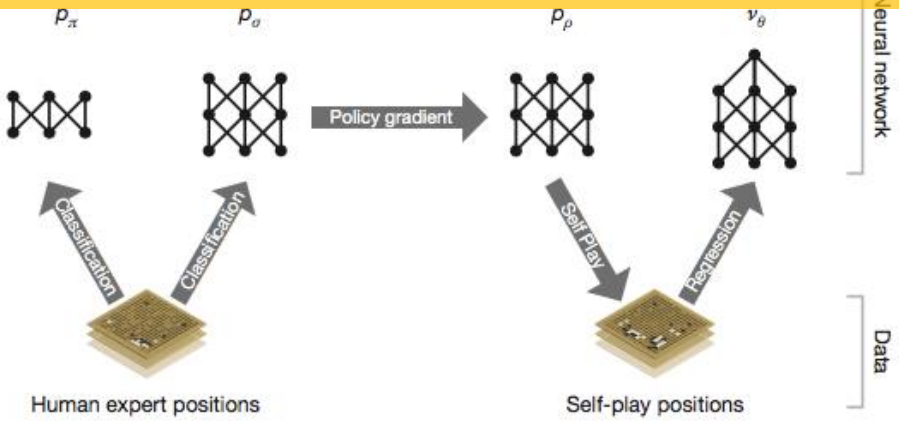
Возможность представления
всей информации о сцене
в виде графа семантических связей
позволяет использовать базы знаний,
логический вывод и эвристики
для анализа видеоданных о реальном мире

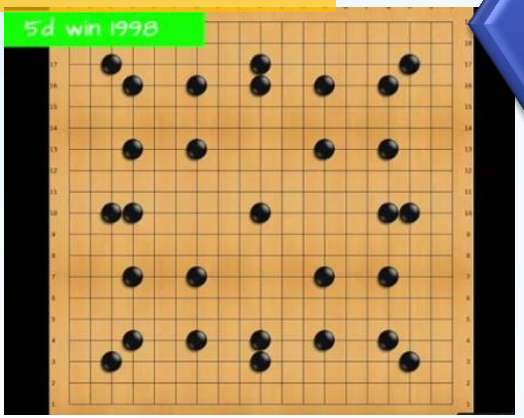
Seen and Thinking, ICCV, 2017

ОТ ALPHAGO К ALPHAZERO

2017: сеть выиграла у чемпиона по игре в ГО

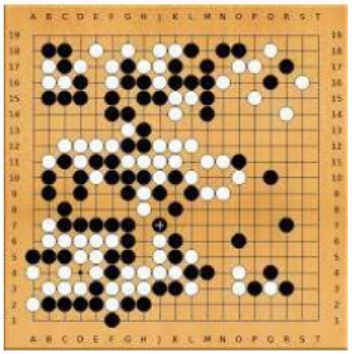
Обучение ГКНС с подкреплением

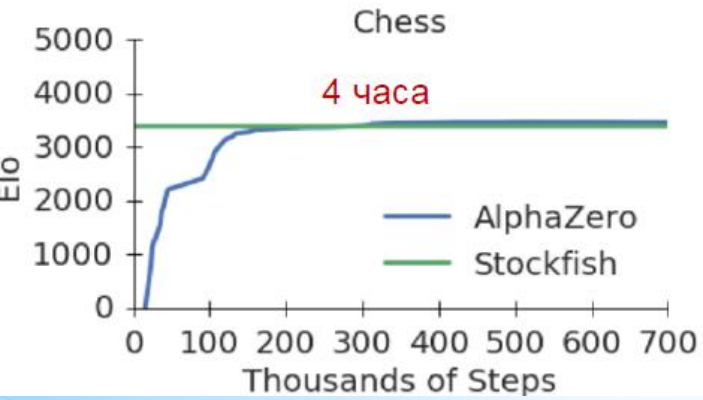


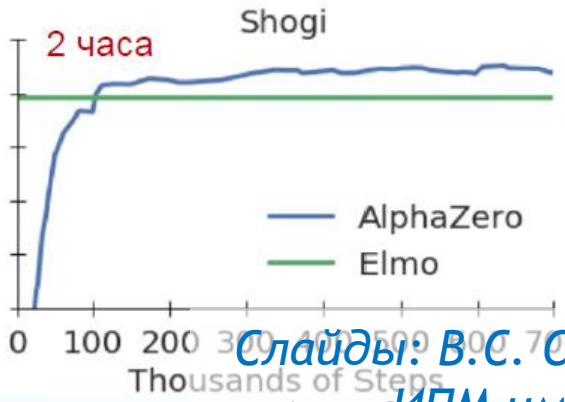


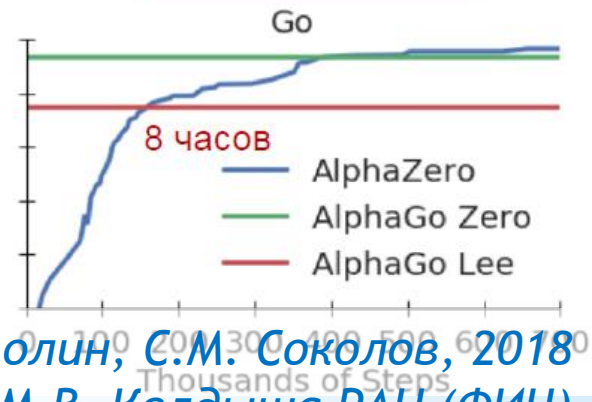
Chess - 10^{47} variants
Go - 10^{171} variants

2017 дек.









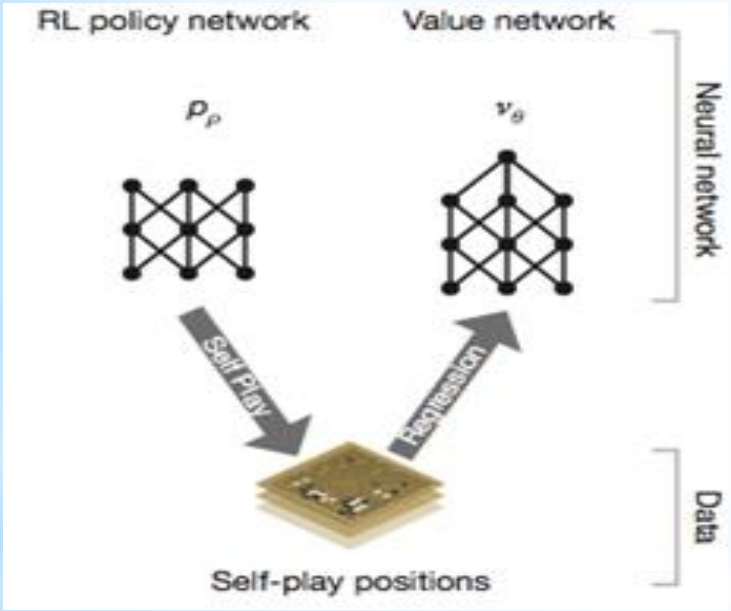
Слайды: В.С. Смолин, С.М. Соколов, 2018
ИПМ им. М.В. Келдыша РАН (ФИЦ)

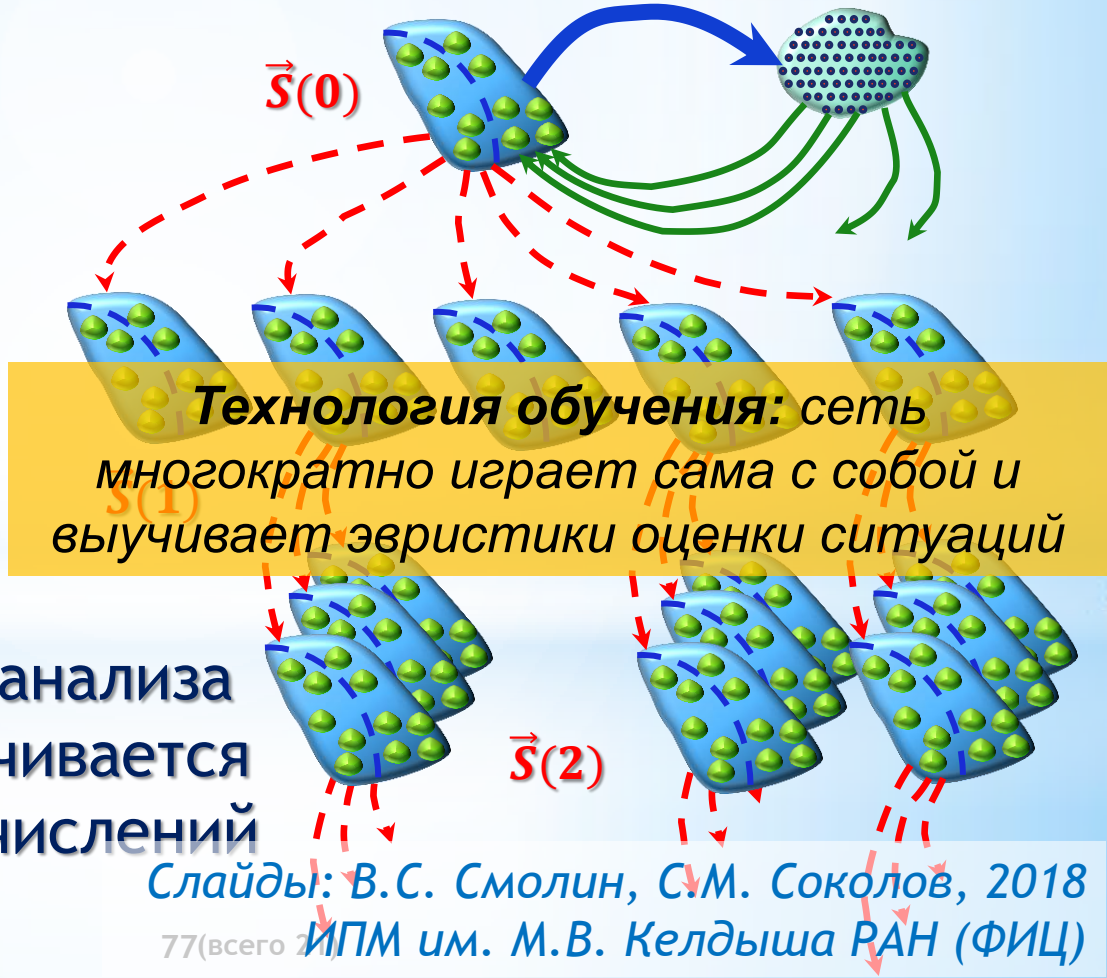
ОБЩАЯ СТРАТЕГИЯ ПОВЕДЕНИЯ

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.

Обучение ГКНС с подкреплением



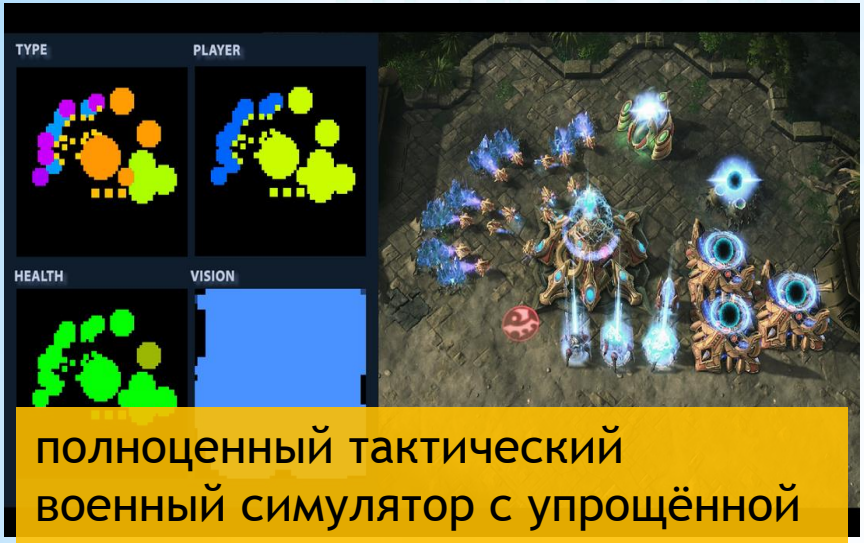


Технология обучения: *сеть многократно играет сама с собой и выучивает эвристики оценки ситуаций*

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

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

ОТ GO К 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

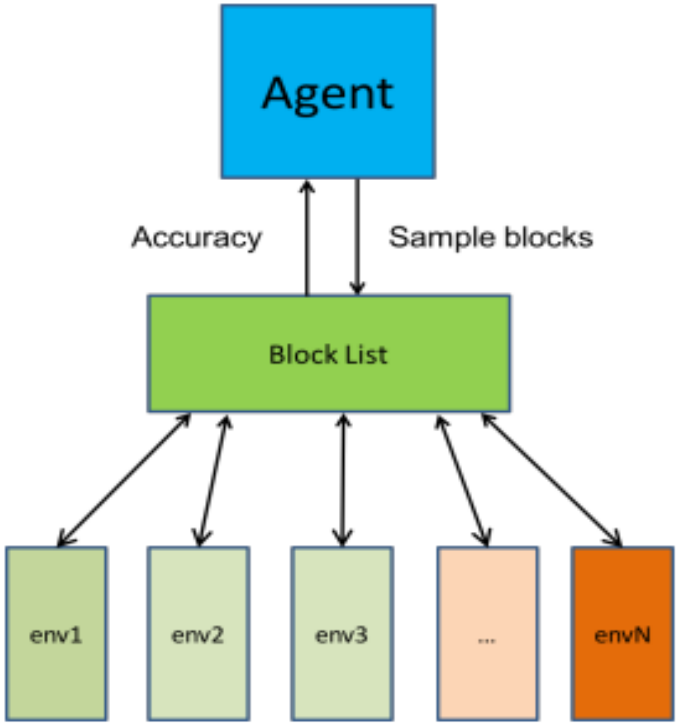
Тактио-стратегические игры

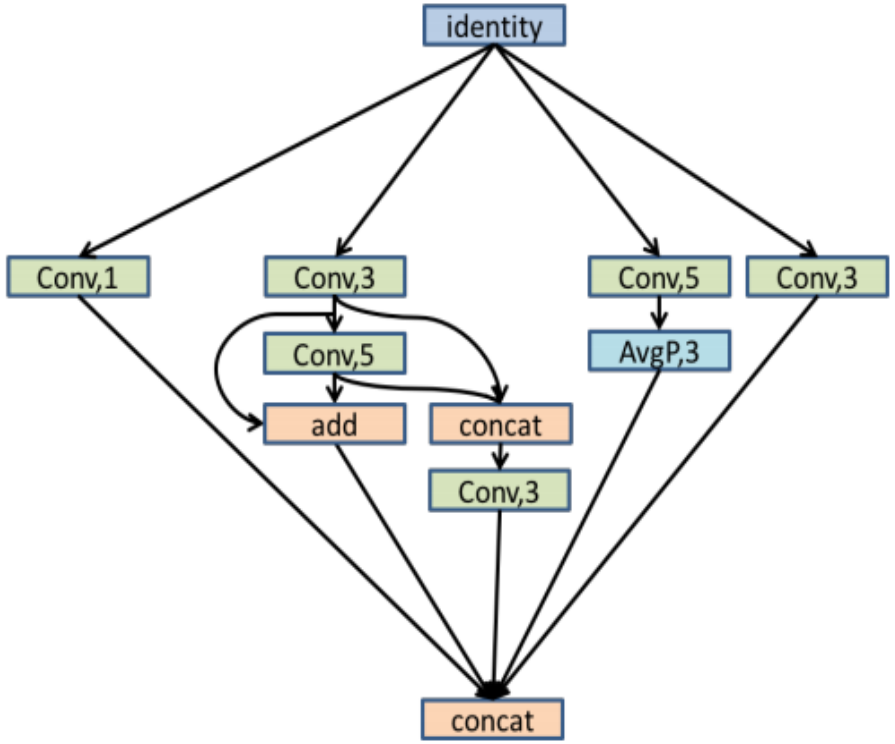
В военно-стратегических играх сети пока проигрывают человеку, но судя по динамике развития, начнут выигрывать ***через 2-3 года (2020+)***

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

Глубокие сети формируют и учат глубокие сети

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

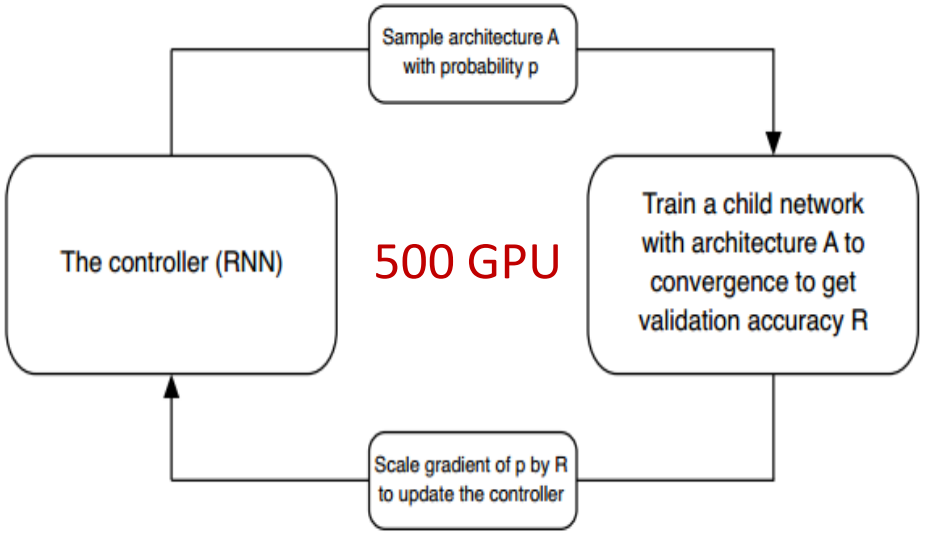




Practical Network Blocks Design with Q-Learning, CVPR-2017

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

Глубокие сети формируют и учат глубокие сети



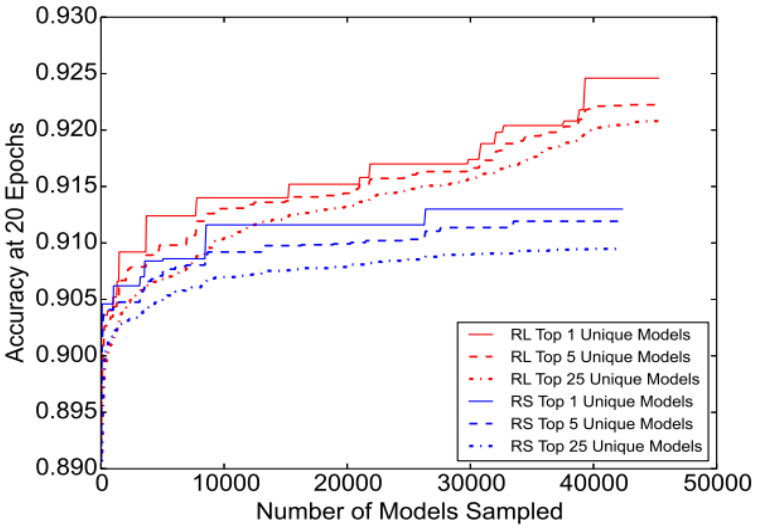


Image Net

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
NASNet-A (5 @ 1538)	**299×299**	**10.9 M**	**2.35 B**	**78.6**	**94.2**
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
NASNet-A (7 @ 1920)	**299×299**	**22.6 M**	**4.93 B**	**80.8**	**95.3**
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
SENet [15]	**320×320**	**145.8 M**	**42.3 B**	**82.7**	**96.2**
NASNet-A (6 @ 4032)	**331×331**	**88.9 M**	**23.8 B**	**82.7**	**96.2**

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+ победа в го+ оперантное научение агентов + сети, обучающие сети

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!

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.

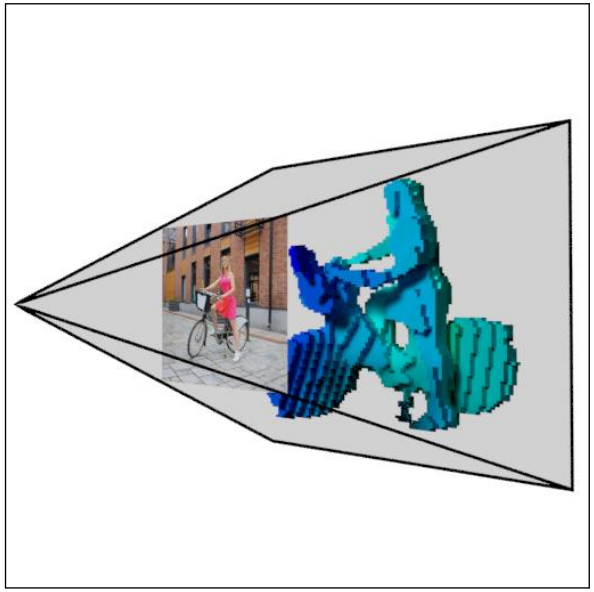
Уровень технологии обучаемого машинного интеллекта (2020+):

Полная готовность к ОКР по созданию интеллектуальных систем!

Проекты и результаты ФГУП «ГосНИИАС» 2018

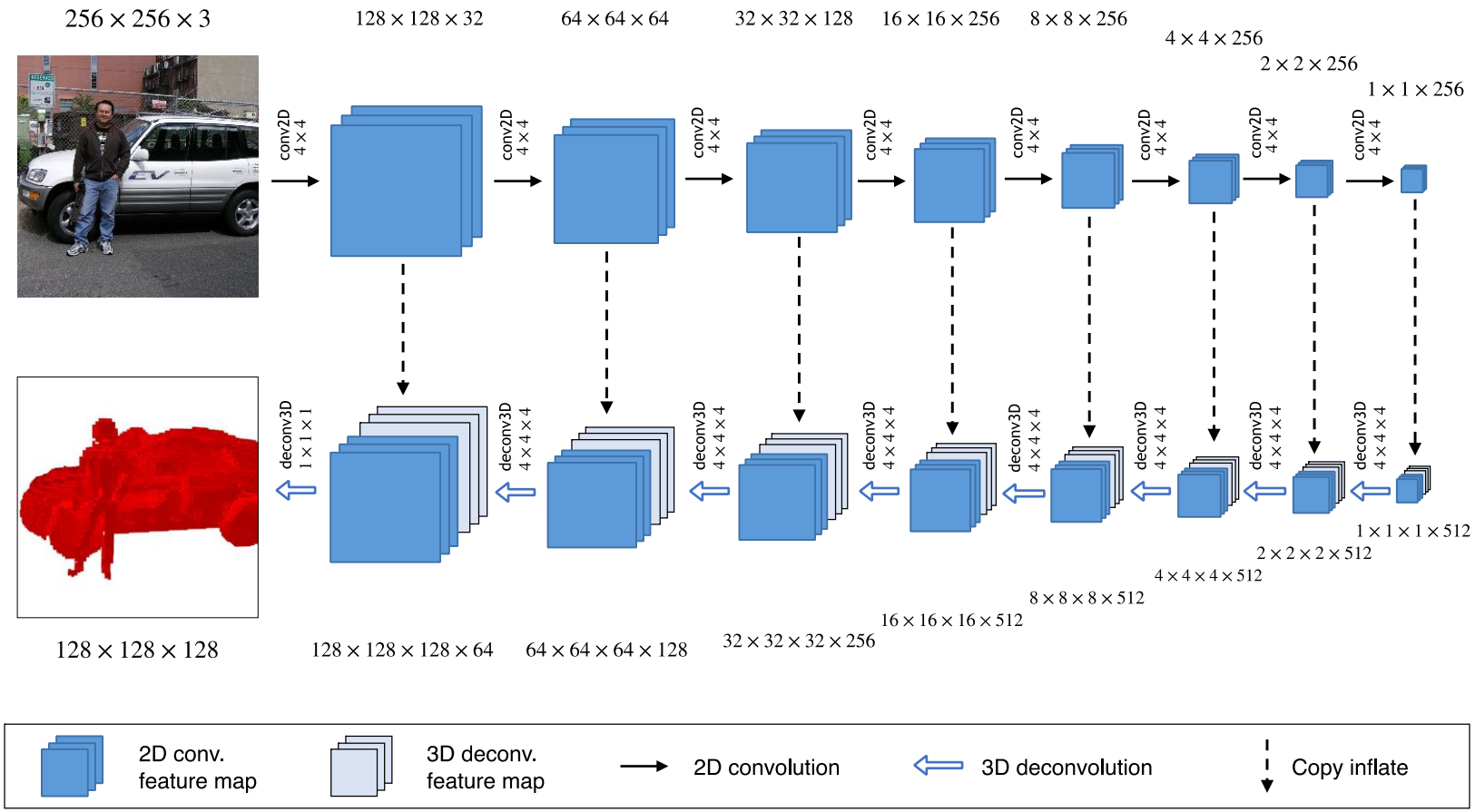
Преобразование 2D изображений в 3D воксельную модель с помощью генеративных конкурирующих сетей

Метод



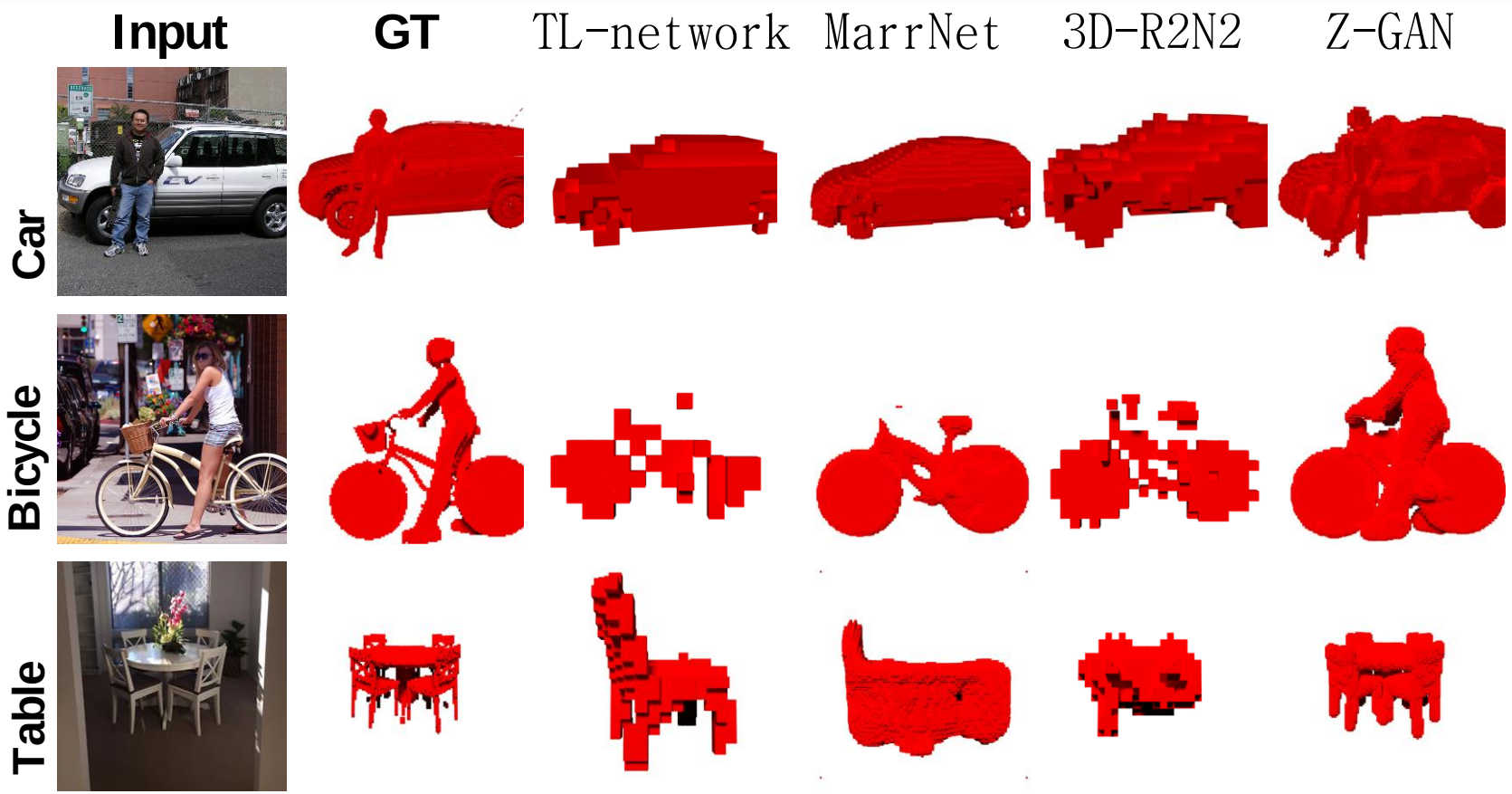
- Для преобразования цветного изображения в воксельную модель разработана генеративно-состязательная сеть Z-GAN
- Для эффективной передачи контурных признаков от изображения в воксельную модель разработан новый вид воксельной модели: «фруксельная модель»
- Фруксельная модель делит пространство сцены на трапециевидные элементы
- При этом каждый элемент контурно соответствует исходному пикселу изображения

Архитектура сети



- Архитектура сети основана на сети `pix2pix`
- Добавлены 3D деконволюционные фильтры
- Модифицированы сквозные связи между слоями (вертикальные линии) для передачи 2D признаков в 3D признаки

Результаты



- Тестирование сети производилось на выборки Pascal3D+
- Сеть сравнивалась с тремя архитектурами «state-of-the-art»
- Z-GAN превосходит аналоги в разрешении воксельной модели и возможности предсказания нескольких объектов в кадре

Преобразование 2D изображений в 3D воксельную модель с помощью генеративных конкурирующих сетей

4rd International Workshop on Recovering 6D Object Pose

BEST PAPER AWARD

Vladimir A. Kniaz, Vladimir V. Kniaz, Fabio Remondino

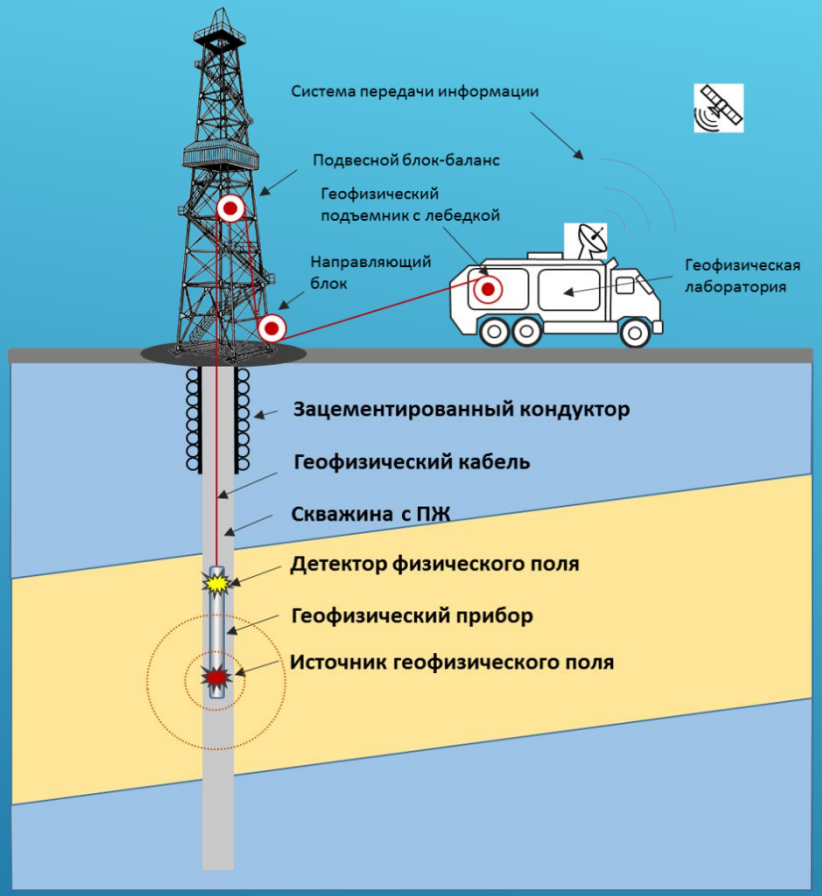
for their work

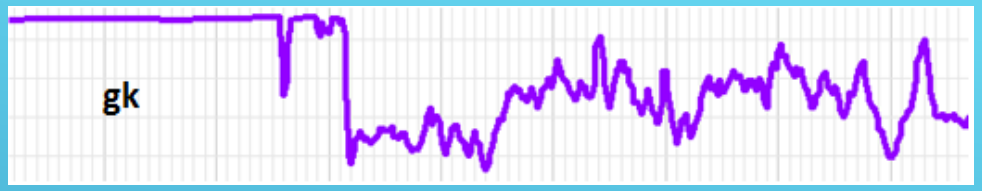
Image-to-Voxel Model Translation with Conditional Adversarial Networks

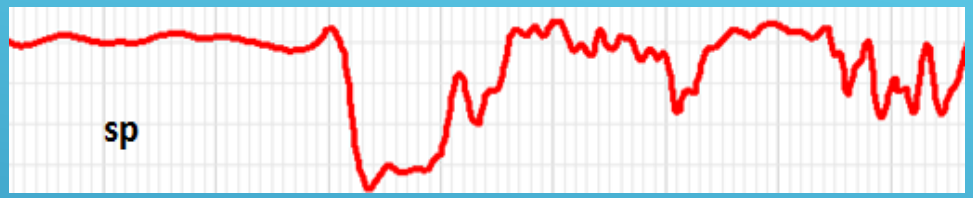
9th September 2018, Munich

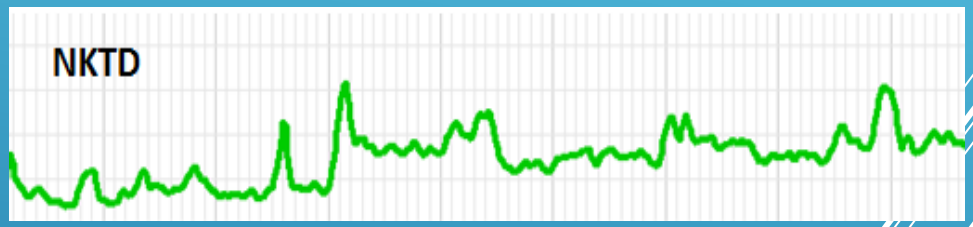
Проект ООО «Газпромнефть НТЦ»: Анализ каротажных диаграмм для автоматической корреляции разрезов скважин

Геофизическое исследование скважин (ГИС)









Анализ кривых ГИС проводится вручную экспертами-геологами

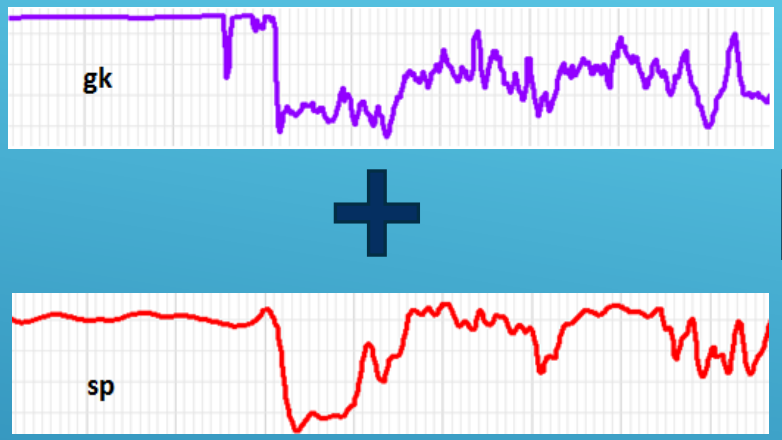
Кривая ГИС представляет собой одномерный сигнал, получаемый в результате измерения (при помощи датчика) значений некоторого физического поля на разных глубинах вдоль ствола скважины

Задачи исследовательского проекта по применению ГКНС к анализу данных ГИС

1. Синтез (восстановление) сигнала ГИС
2. Привязка ГИС-ГИС
3. Поиск (корреляция) участка скважины на другой скважине

Синтез(восстановление) значений сигнала ГИС

Синтез кривой по другим исследованиям

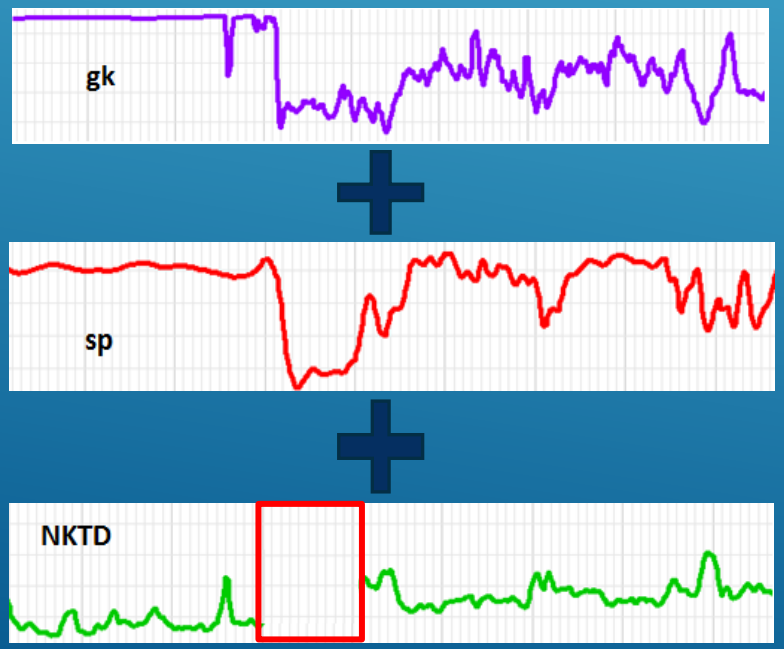


СИНТЕЗ КРИВЫХ ГИС С ИСПОЛЬЗОВАНИЕМ ГЕНЕРАТИВНЫХ КОНКУРИРУЮЩИХ ГЛУБОКИХ КОНВОЛЮЦИОННЫХ НЕЙРОННЫХ СЕТЕЙ

В. Горбацевич, Ю. Визильтер (ФГУП «ГосНИИАС»), А. Хайдаров, А. Яковлев (ООО «Газпромнефть НТЦ»)

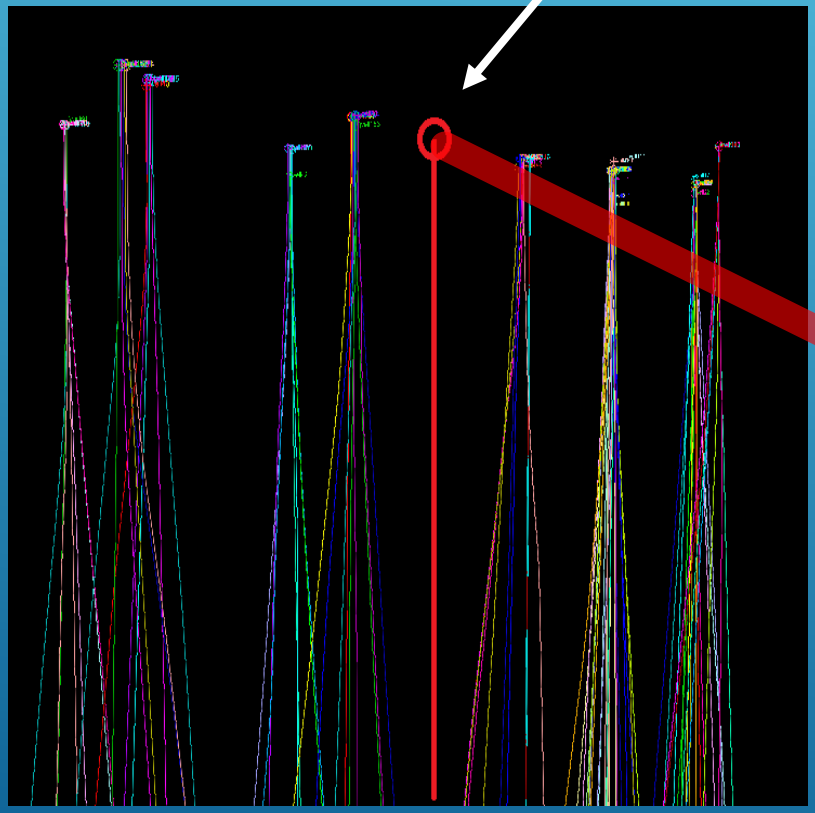
Отдельный доклад!

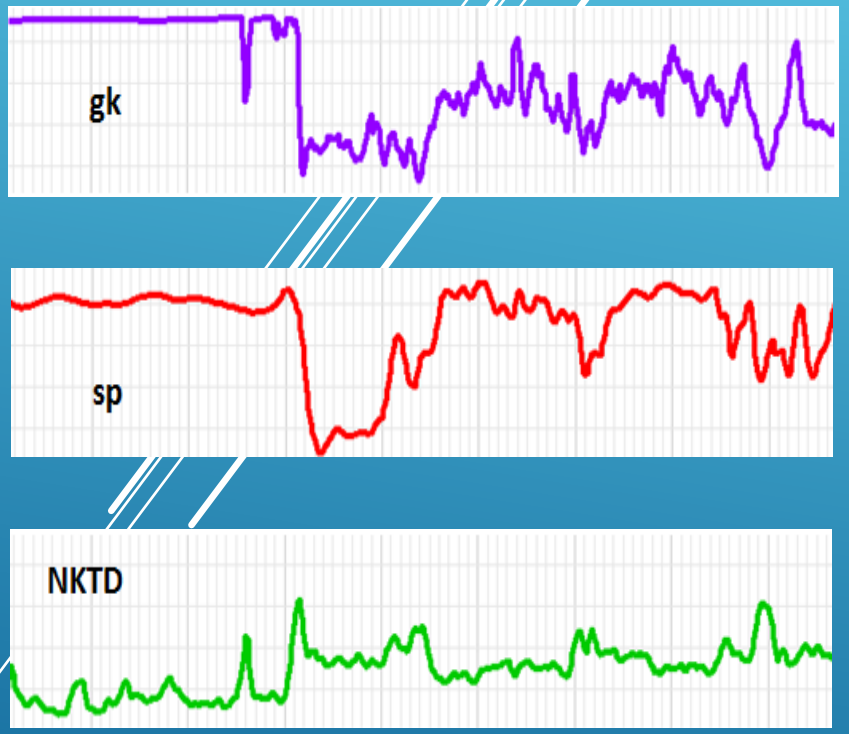
Синтез участка кривой



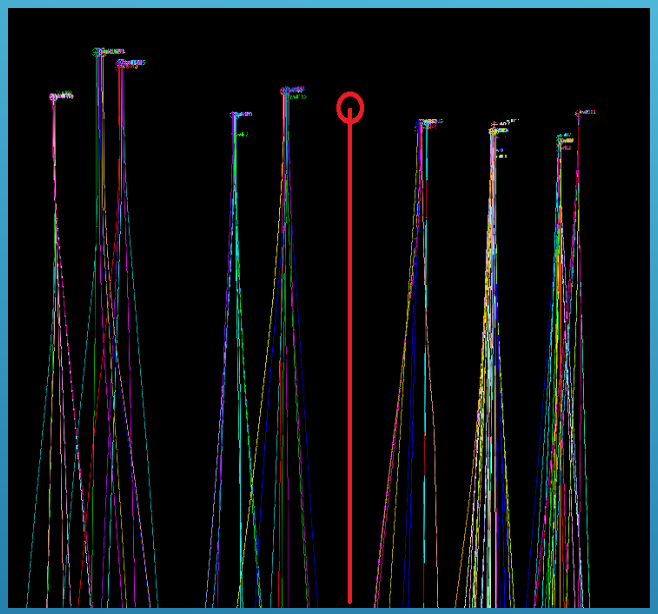
Синтез (восстановление) значений скважины

Указанное местоположение интересующей скважины

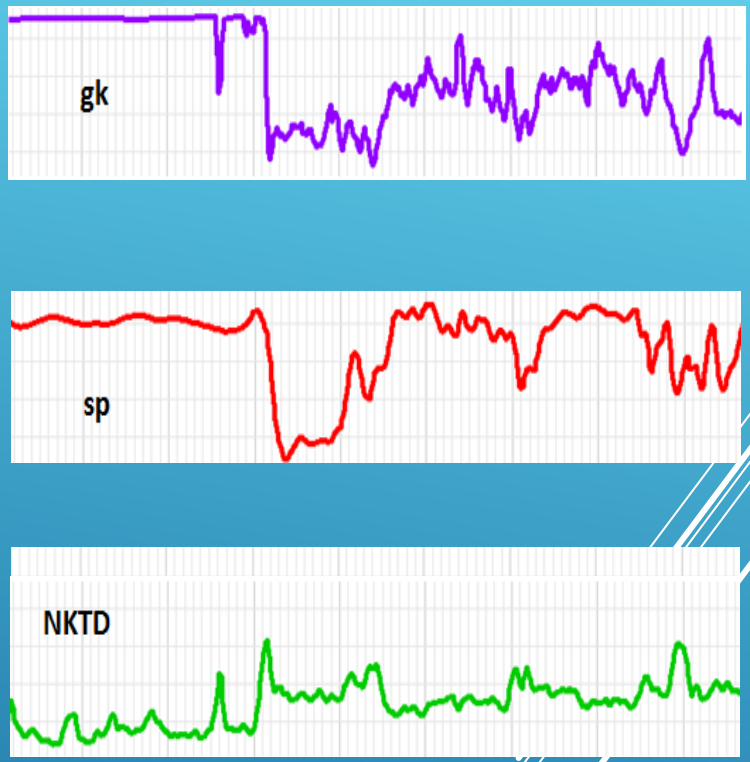




СИНТЕЗ (ВОССТАНОВЛЕНИЕ) ЗНАЧЕНИЙ СКВАЖИНЫ



CNN

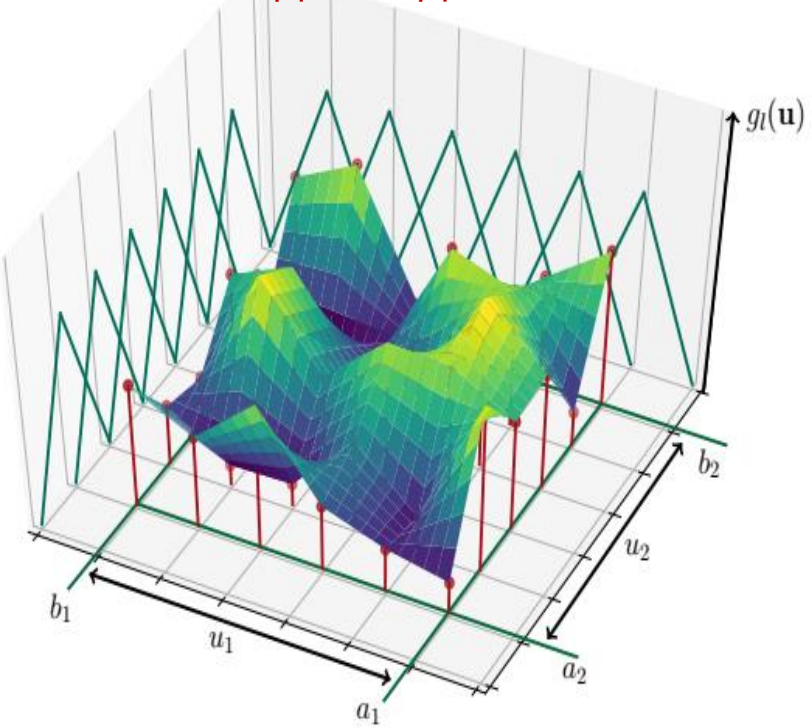


Основная проблема: как представить нерегулярную структуру поля скважин и сложную траекторию каждой скважины для обработки данных в CNN

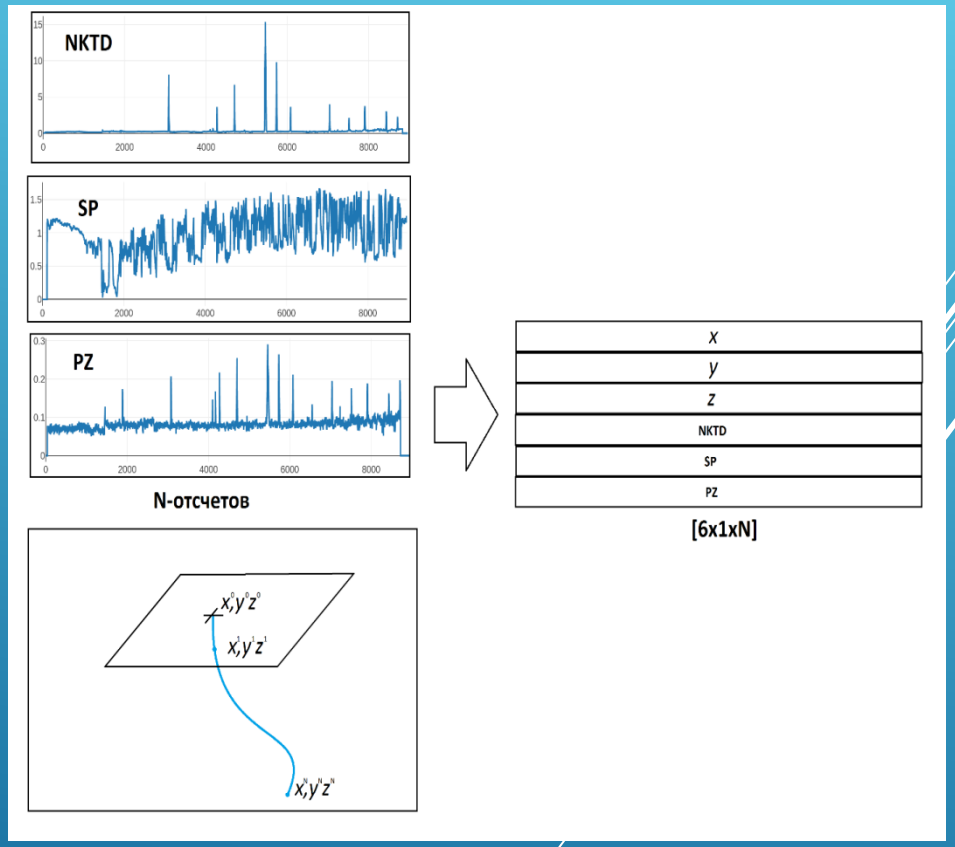
СИНТЕЗ(ВОССТАНОВЛЕНИЕ) ЗНАЧЕНИЙ СКВАЖИНЫ

СФАС и непрерывные ядра для учета нерегулярной структуры поля скважин с данными ГИС

Тема моего доклада на ММРО-2017

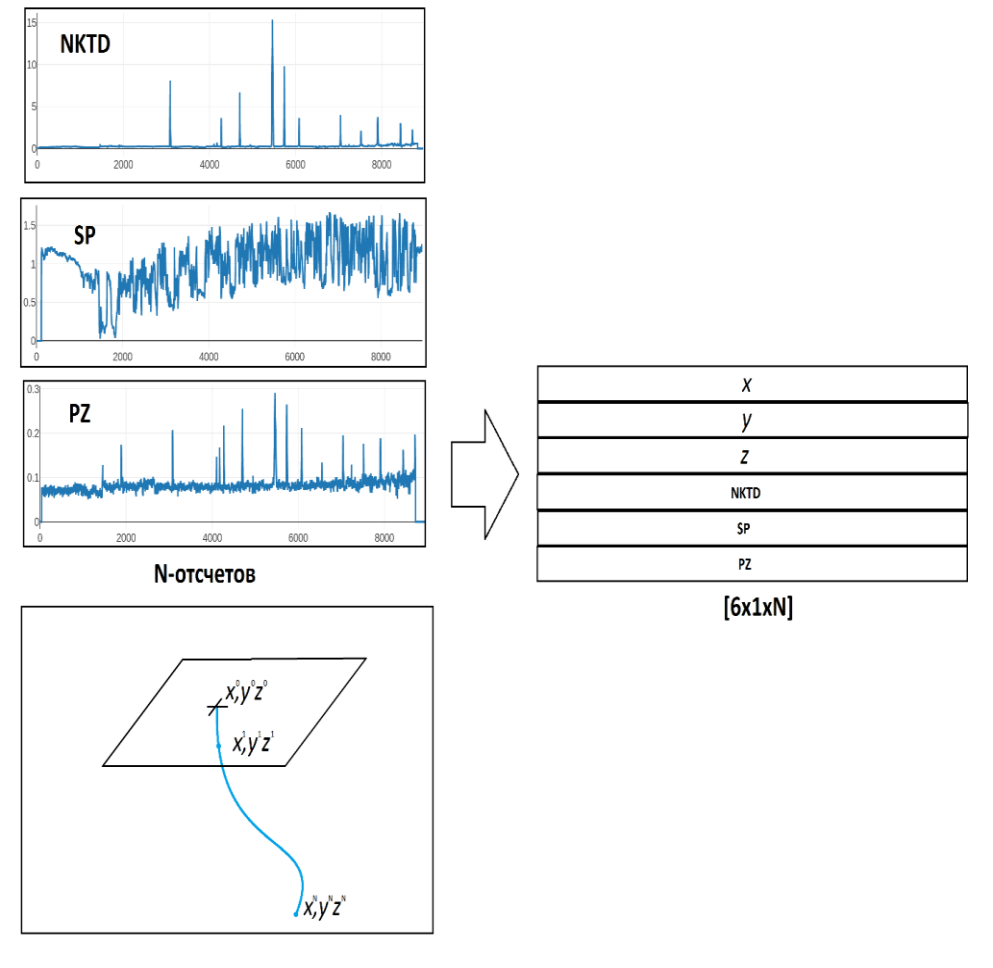


Представление ГИС скважин сложной формы одинаковыми тензорами



Учёт формы поля скважин
и траектории каждой скважины

СИНТЕЗ (ВОССТАНОВЛЕНИЕ) ЗНАЧЕНИЙ СКВАЖИНЫ



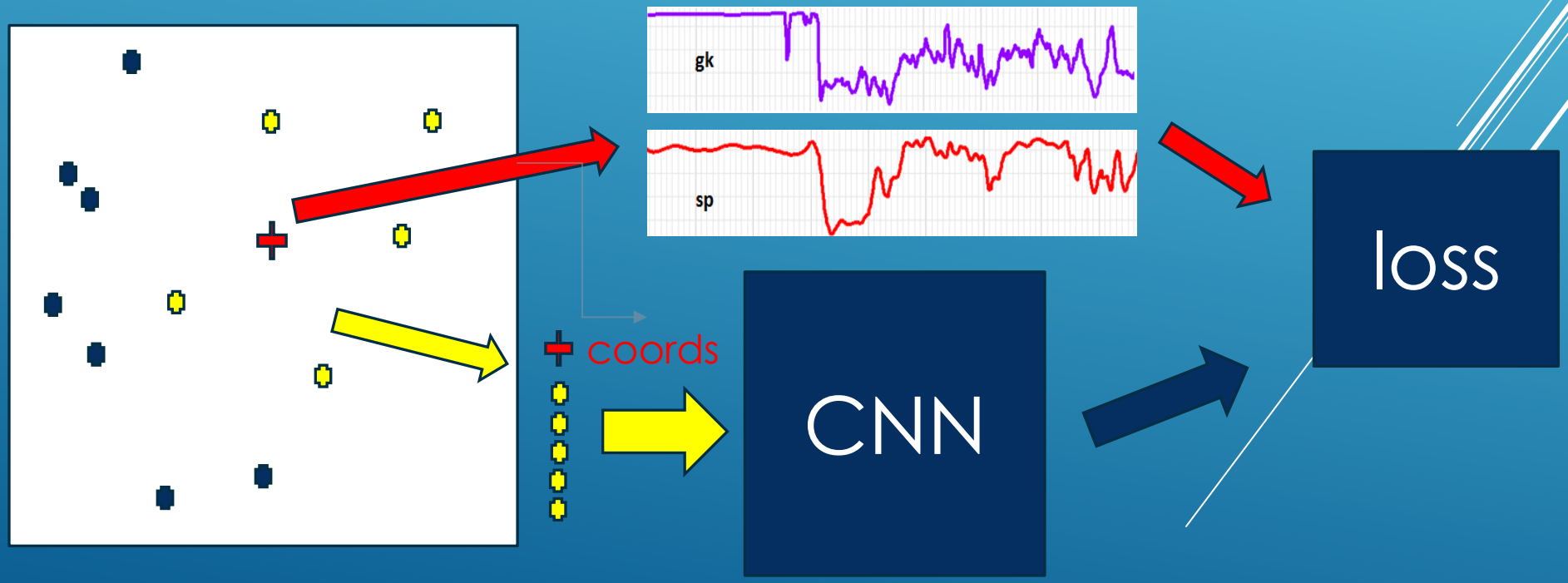
Учёт формы траектории скважины

- Предсказание производится по пяти ближайшим скважинам
- Координаты относительно точки x0,y0 опорной скважины
- Предсказание проводится для исследований gk и sp
- Исследования масштабируются до 1024 отсчётов

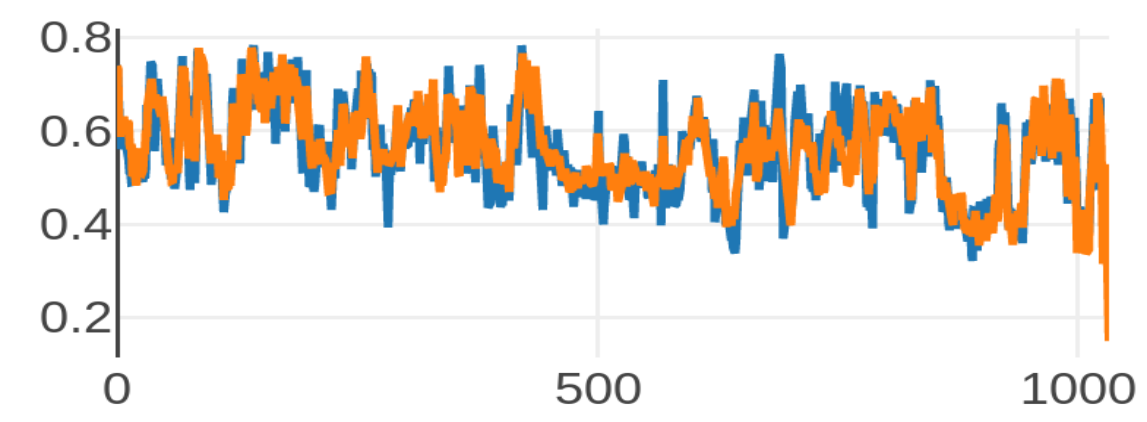
СИНТЕЗ (ВОССТАНОВЛЕНИЕ) ЗНАЧЕНИЙ СКВАЖИНЫ

Алгоритм аугментации данных:

1. Из обучающей выборки случайно выбирается b скважин.
2. Для каждой скважины находится 50 ближайших скважин.
3. Из пятидесяти скважин случайным образом отбирается 5 текущих ближайших скважин, упорядоченных по расстоянию до искомой.



СИНТЕЗ (ВОССТАНОВЛЕНИЕ) ЗНАЧЕНИЙ СКВАЖИНЫ



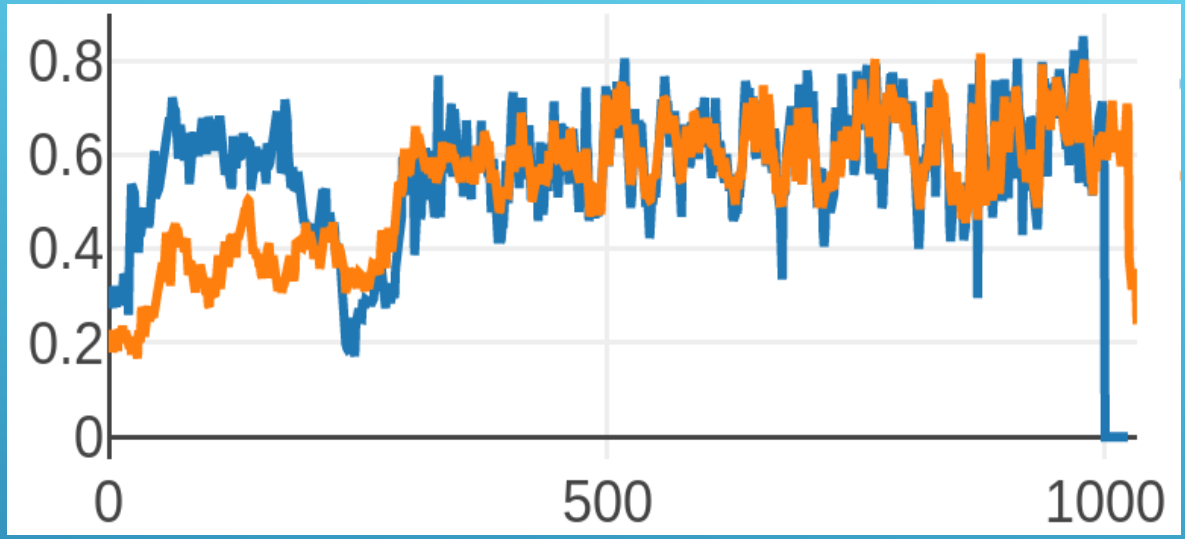
	1 скважина	32 скважины
GPU (GTX 1080)	14 мс	46 мс
CPU (Core i7 5930K 1 поток)	213 мс	6782 мс

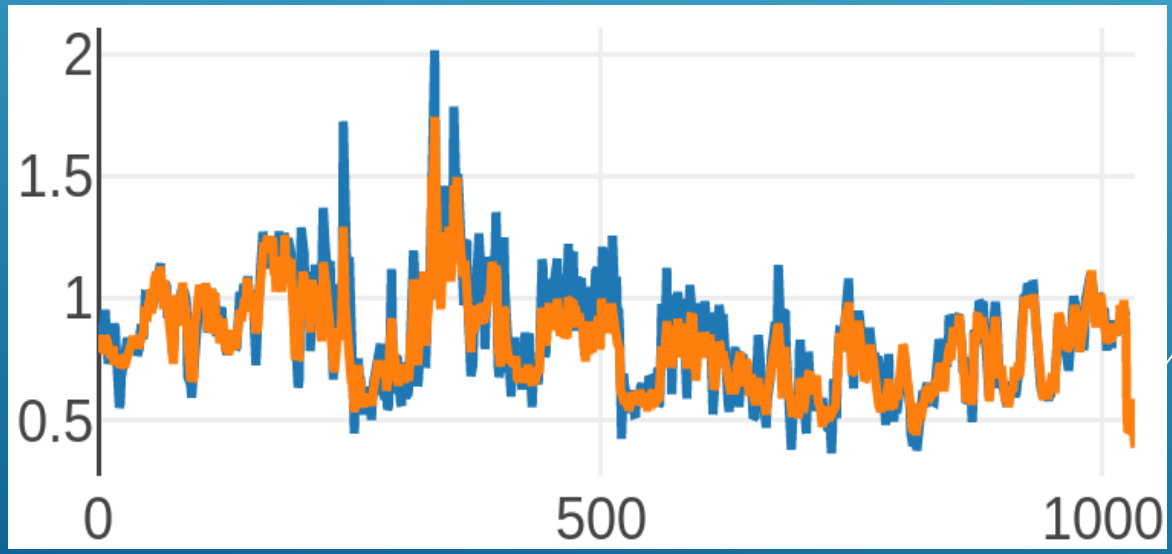
Сеть	Sp	gk
ResNet50-ResNet18	0,91	0,85

Коэффициент корреляции Пирсона

Моделирование решётки 256x256x1024 занимает порядка 3000 сек. (375 сек. при 8GPU) при использовании GPU

СИНТЕЗ (ВОССТАНОВЛЕНИЕ) ЗНАЧЕНИЙ СКВАЖИНЫ





ЗАКЛЮЧЕНИЕ: Ключевые приложения и технологии ИИ (2020+)

SLAM, зрение и интеллект автономных РТК и БЛА, Reinforcement Learning, GAN **РТК, БЛА, Автономный транспорт**	Обнаружение объектов, бортовые CNN, ATR, комплексы обучения ATR **РТК, БЛА, ВТО, АСП**
Биометрия, Интеллектуальная видеоаналитика **Системы безопасности**	Семантическая сегментация, Автоматическое дешифрирование **Разведка, мониторинг**
Стратегический интеллект, Reinforcement Learning, Deep Reasoning, Игры/операции **Военная стратегия, бизнес**	GAN, распознавание сигналов, структурные данные, структурные сети **Добыча и дизайн материалов**
Анализ больших данных о самолете / корабле / предприятии **Транспорт, промышленность**	Нейроинтерфейсы, чтение мыслей, киборгизация **Будущее человечества**
Сети конструируют и учат сети	**Будущее человечества**

Современное состояние и ближайшие перспективы развития технологий глубокого обучения и компьютерного зрения

Ю.В. Визильтер, д.ф.-м.н., проф. РАН, viz@gosniias.ru

12-я Конференция ИОИ, Италия, г. Гаэта, 08.10.2018

Спасибо за внимание!

12-14 марта 2019 г. в Москве, в ИКИ РАН состоится научно-техническая конференция **“Техническое зрение в системах управления - 2019” (ТЗСУ’19)**. http://tvcs2018.technicalvision.ru/

ISPRS International Workshop “Photogrammetric and computer vision techniques for video Surveillance, Biometrics and Biomedicine” – **PSBB19, May 13-15, 2019, Moscow, Russia (ВМК МГУ)** http://technicalvision.ru/ISPRS/PSBB19/