Современное состояние технологий искусственного интеллекта.
Отечественная нейросетевая платформа Plat

Визильтер Юрий Валентинович, viz@gosniias.ru

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

Контакт по Платформе Plat:

support.plat@gosniias.ru

Семинар МГУ

Москва, 13.10.2020

Первая часть доклада

Современное состояние технологий ИИ

- **Глубокие нейронные сети** *совершили революцию в области компьютерного зрения и анализа больших данных*
- **Функциональный искусственный интеллект** *скоро будет создан в ходе второй волны «революции ИИ», которая происходит прямо сейчас*
- **Глубокая оптимизация** *уже в ближайшие годы распространит эту революцию практически на все области техники, технологии, экономики…*
- **Актуальные результаты и открытые проблемы** *угрозы, вызовы, надежды (на примерах из области компьютерного зрения и не только…)*

2011-2016, 2017-2020+: две волны технологической революции в ИИ

Ранний период развития ИИ: созданы основные методы и подходы…

Завышенные ожидания

Разочарование

Рост интереса

Алгоритмы ИИ сильно проигрывают людям

«Кривая хайпа» для ИИ

2000
«Зима ИИ»
Прогноз: 2040+

2016-17

- **Глубокие соревнующиеся сети**
- **ГО с подкреплением**
- **ГО с использованием моделей, баз знаний и логического вывода**
- **Автоматическое обучение ГНС**

Вторая волна революции ГНС

Все задачи ИИ (НИР)

Первая волна революции ГНС

Распознавание (НИР->ОКР)

Функциональный («слабый») ИИ

Прогноз: 2020+

- **Появление глубоких нейронных сетей (ГНС)**
- **Решение задач компьютерного зрения, обработки сигналов и анализа больших данных на уровне человека и выше (superhuman)**

2011-12

…2011+: в области ИИ началась технологическая революция ГНС

Глубокие нейронные сети

(первая волна современной технологической революции ИИ)

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

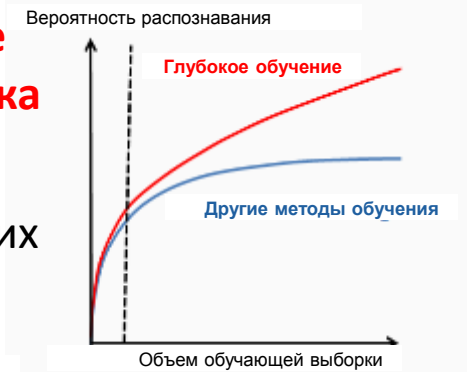
2011: First Superhuman Visual Pattern Recognition

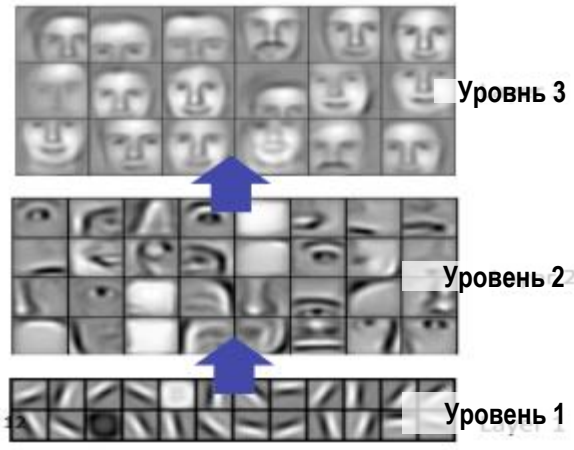
http://people.idsia.ch/~juergen/deeplearning.html

2011: Автоматическое обнаружение и распознавание объектов на базе *глубоких нейронных сетей* (ГНС)

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

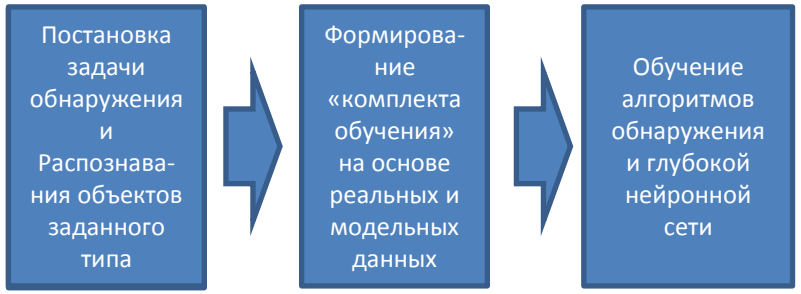
+ Обучение на сверхбольших объемах данных



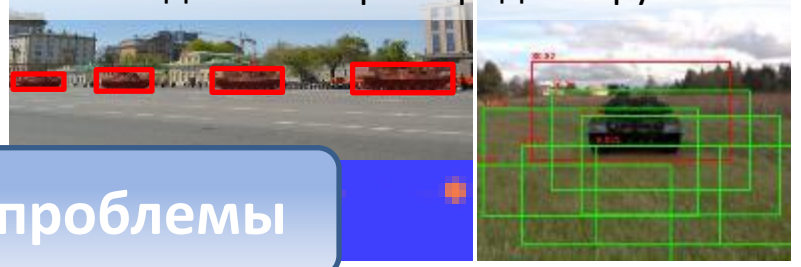


+ Иерархическое обучение с повышением абстракции данных от уровня к уровню

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



- 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

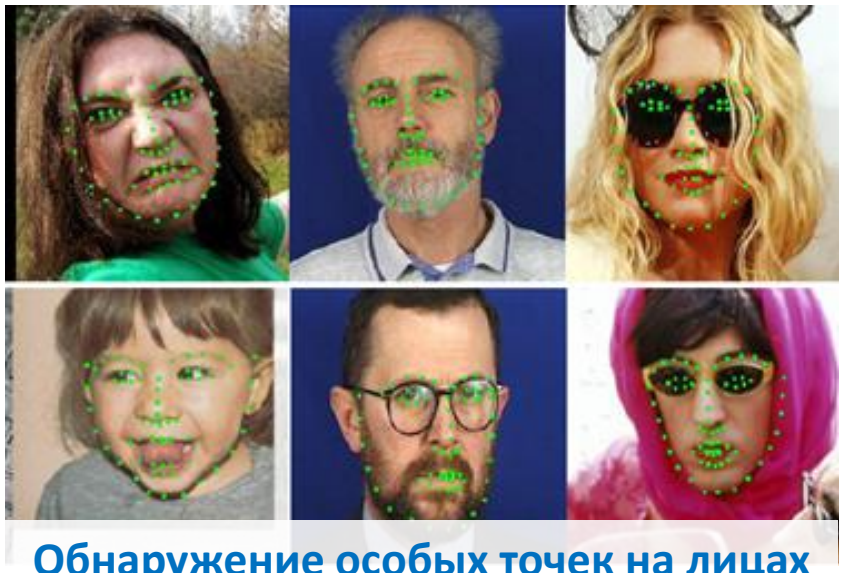
- **Необходимость эффективных алгоритмических реализаций и нового поколения бортовых нейропроцессоров**

2015-16: ГНС решают все задачи компьютерного зрения

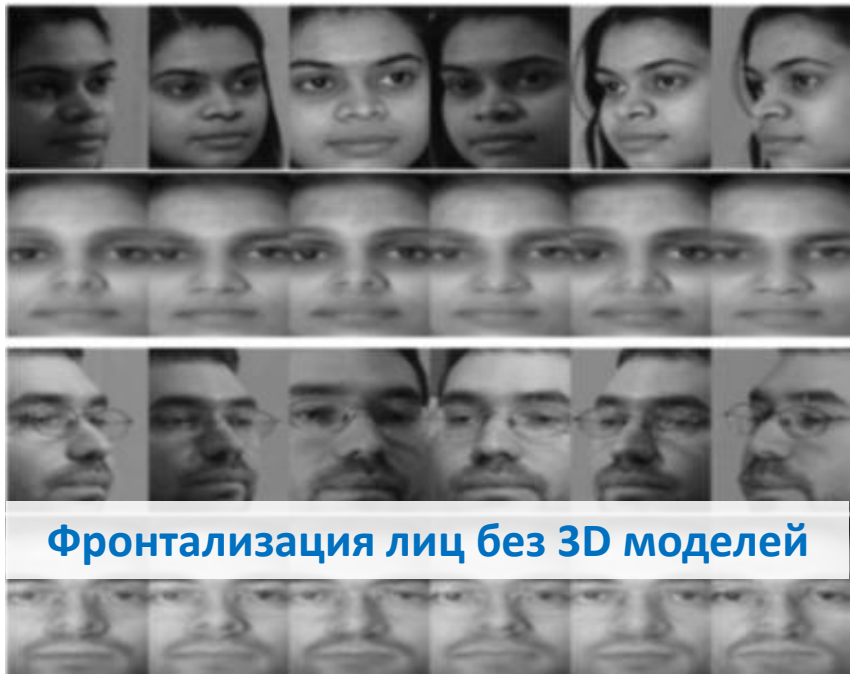
Фильтрация шумов

Удаление смаза

Суперразрешение



Обнаружение особых точек на лицах



Фронтализация лиц без 3D моделей

Распознавание лиц в сложных условиях (ГосНИИАС-2015)

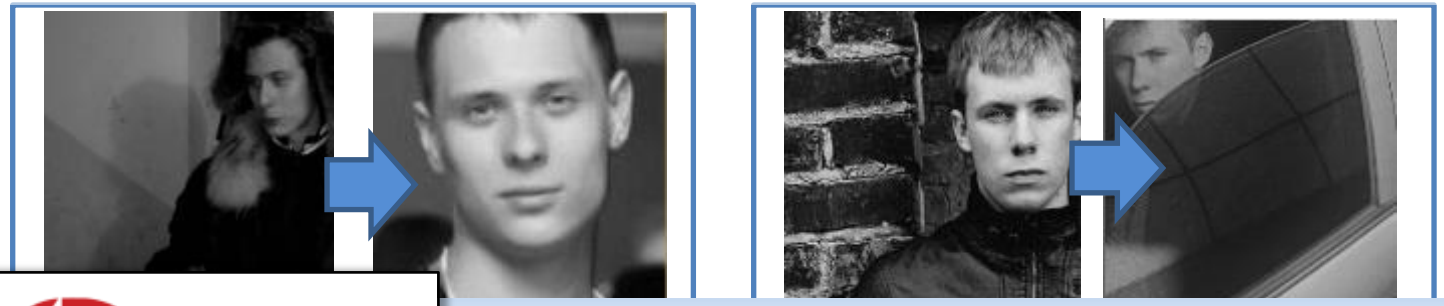
Применение методов глубокого обучения к задаче распознавания лиц

~30 млн сравнений в секунду для 2000 битного шаблона

Сверхкомпактные шаблоны, высокая скорость поиска

~ 200 млн сравнений в секунду для 200 битного шаблона

Качество, сравнимое с результатами человека-оператора



Распознавание LFW:

200 бит ~0.95

2000 бит ~0.987

ФОНД ПЕРСПЕКТИВНЫХ ИССЛЕДОВАНИЙ

Решение победило в конкурсе ФПИ-2015 на лучшую технологию распознавания лиц в России

Что дают нам технологии глубокого обучения, распознавания образов и анализа данных?

СИСТЕМЫ ВИДЕОНАБЛЮДЕНИЯ

БИОМЕТРИЧЕСКИЕ СИСТЕМЫ

ПРОМЫШЛЕННАЯ БЕЗОПАСНОСТЬ И АВТОМАТИЗАЦИЯ

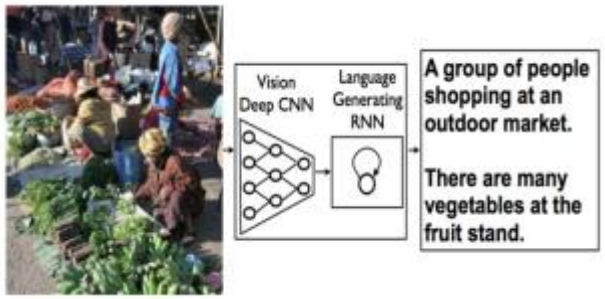
АНАЛИЗ БОЛЬШИХ ДАННЫХ

ПРОШЛОЕ

НАСТОЯЩЕЕ

БУДУЩЕЕ

СИСТЕМЫ ВИДЕОНАБЛЮДЕНИЯ



ПРОШЛОЕ

- **Обнаружение движущихся объектов в несложных условиях**
- Классификация движущихся объектов
- Обнаружение принесенных (оставленных) и унесенных предметов
- Расчет скорости и плотности потока (людей или машин)
- Распознавание объектов

НАСТОЯЩЕЕ

- Распознавание типа оставленного предмета как подтверждение детектора
- Классификация движущихся объектов на основе глубоких нейронных сетей
- **Реидентификация персоны с разных камер по одежде с использованием CNN**
- **Детектирование объектов (например, людей, авто) на отдельных кадрах**
- Распознавание номеров на основе нейросетей

БУДУЩЕЕ

- **Анализ и распознавание поведения людей**
- Высоконадежное обнаружение и прослеживание объектов, в том числе с PTZ-камер
- Появление «умных» камер с алгоритмами детектирования на основе нейронных сетей на борту
- **ИИ: система видеонаблюдения самостоятельно генерирует сообщения оператору с подробным описанием событий**
- **ИИ: система видеонаблюдения самостоятельно принимает решения**

БИОМЕТРИЧЕСКИЕ СИСТЕМЫ

ПРОШЛОЕ

- Обнаружение лиц с помощью группы методов Виолы-Джонса
- Распознавание лиц с использованием обученных признаков
- Обнаружение и распознавание лиц в сложных условиях

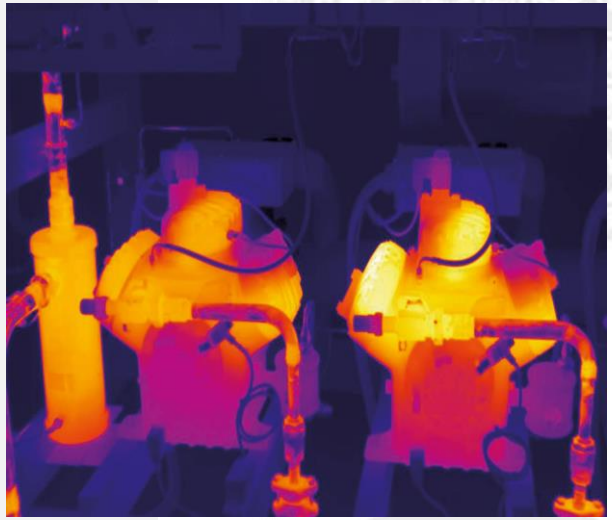
НАСТОЯЩЕЕ

- Детектирование лиц продвинутыми методами предыдущего поколения
- Детектирование лиц в сложных условиях с помощью нейросетей (не real-time)
- **Распознавание лиц по изображениям высокого качества с очень высокими вероятностями**
- **Распознавание лиц в сложных условиях с достаточно высокими вероятностями**

БУДУЩЕЕ

- Повсеместная замена электронных ключей на распознавание лиц
- Высокие точности определения пола, возраста, национальности, эмоций в задачах маркетинга
- Детектирование и распознавание людей в толпе по большим базам с высокой вероятностью -единая биометрическая глобальная система в рамках городов
- Коммерческое внедрение оплаты по лицу

ПРОМЫШЛЕННАЯ БЕЗОПАСНОСТЬ И АВТОМАТИЗАЦИЯ



ПРОШЛОЕ

- **Обнаружение движения или движущихся объектов в запрещенных зонах (sterile zone)**
- Контроль температуры установок с помощью тепловизионных камер
- Замена человека в простейших действиях автоматическими системами (задачи подсчета продукции, считывания маркировок/штрихкодов и пр.)

НАСТОЯЩЕЕ

- Распознавание средств индивидуальной защиты (чувствительно к марке)
- **Задачи технологического контроля:** определение качества продукции, правильности течения технологических процессов и пр.
- Поиск полезных ископаемых по большим данным – набору показаний группы датчиков
- Автоматизированные заводы с минимальным количеством персонала

БУДУЩЕЕ

- Автоматические системы контроля за производством
- **Автоматические интеллектуальные системы управления заводами:** анализ показаний датчиков системы, предсказание чрезвычайных ситуаций, принятие решений
- **Создание автономных подвижных роботов** нового поколения с системой технического зрения на борту

АНАЛИЗ БОЛЬШИХ ДАННЫХ

Объем данных, собранных только с двигателей коммерческих реактивных самолетов США в течение года

= 1 041 600 500 ТВ

НАСТОЯЩЕЕ - БУДУЩЕЕ

✅ **Анализ состояния бортового авиационного оборудования** в масштабе реального времени с возможностью **прогнозирования различных типов отказов**.

✅ Совершенствование средств **прогнозирования времени прибытия воздушных судов** в составе систем управления воздушным движением.

НАСТОЯЩЕЕ - БУДУЩЕЕ

✅ **Глубокие нейронные сети в инвестировании**

✅ **Предсказание эффективности компаний** в терминах выручки, операционных доходов и чистой прибыли на основе финансовой информации и патентах компаний

✅ **Предсказание поведения фондового рынка**

и еще множество приложений…

Функциональный искусственный интеллект

(вторая волна современной технологической революции ИИ)

2000: Что такое функциональный ИИ, из чего он состоит?

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

Два «дома», разделенных прочной стеной: ИИ-1 не учится, ИИ-2 не рассуждает…

2000: Прогноз создания функционального ИИ (на примере робототехники)

Везде алгоритмы сильно проигрывают людям

Навигация

Обработка сенсорных данных (зрение,…)

Управление (планирование оптимизация, игры,…)

Машинное обучение (ИИ-II) (анализ данных)

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

Прогноз ИИ: 2040+

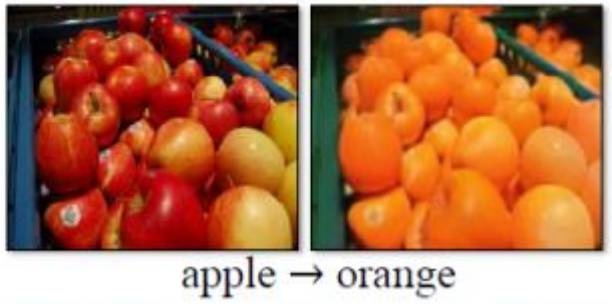
2015: Прогноз создания функционального ИИ (на примере робототехники)

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 с подкреплением 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**

Generative Adversarial Networks (GANs)

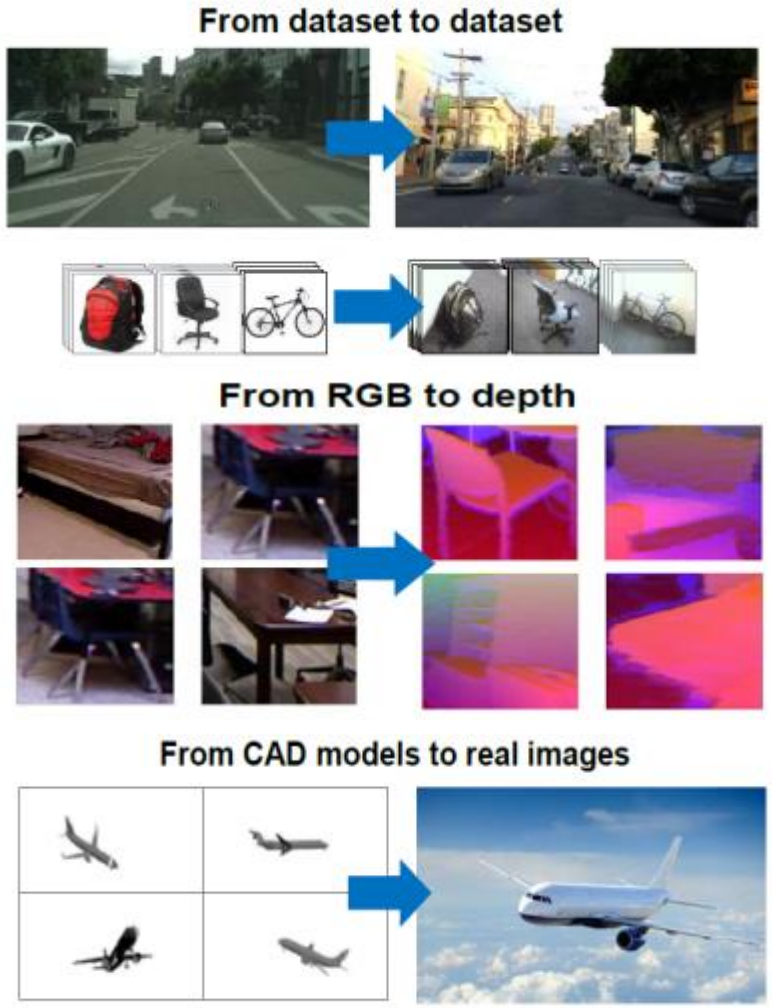


GAN – сеть, обладающая воображением!

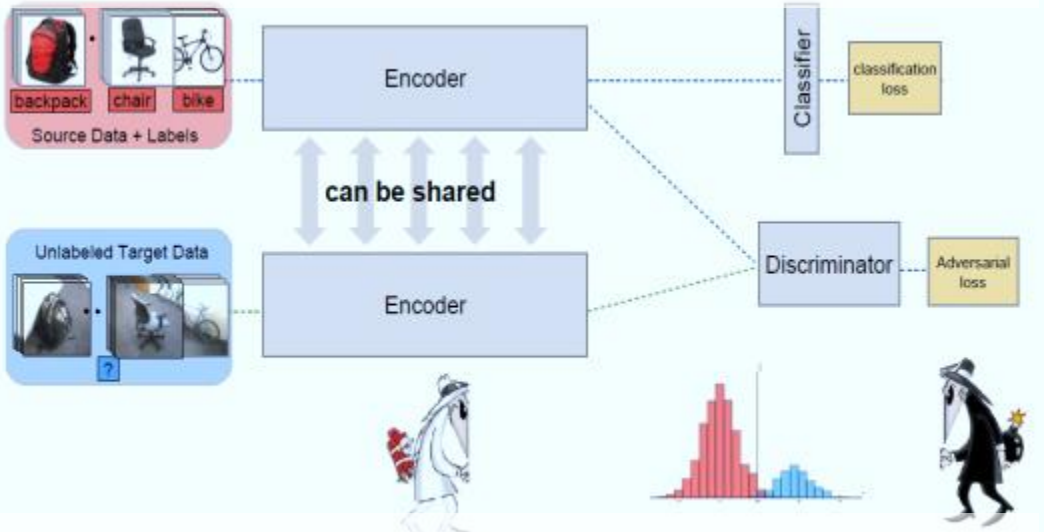
Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, Jun-Yan Zhu et al., ICCV, 2017

Генеративные конкурирующие сети

Перенос обучения в новую область применения (Domain Transfer Learning)



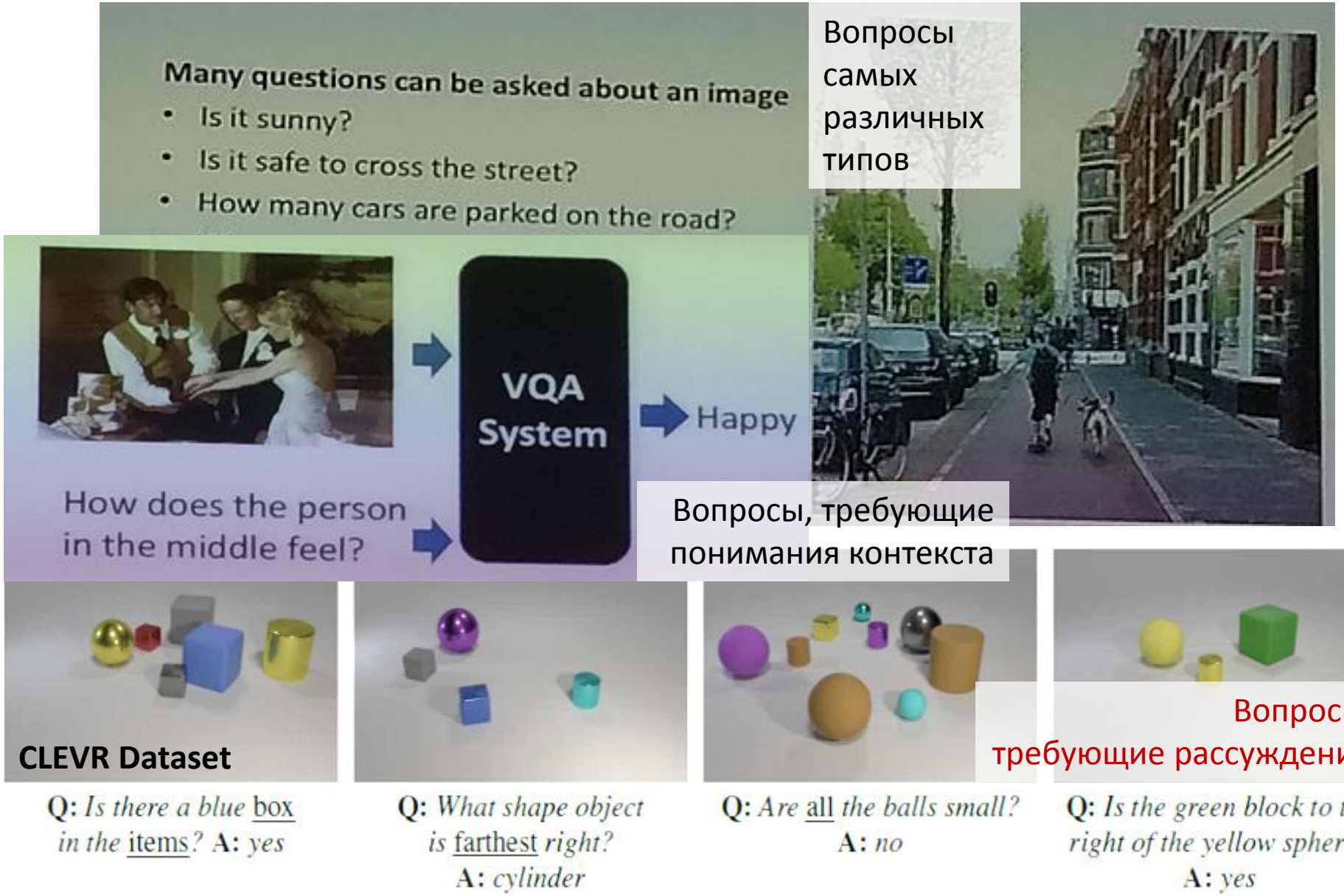
Генератор создает визуальные образы, стараясь обмануть Дискриминатор…



….Дискриминатор старается отличить фантазии Генератора от реальности

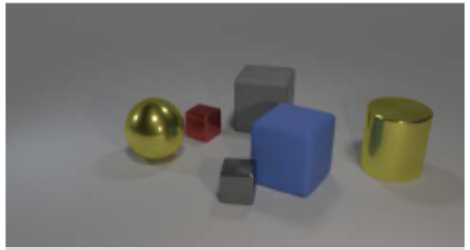
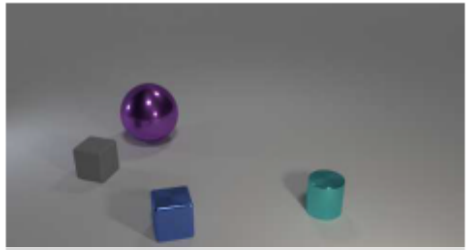
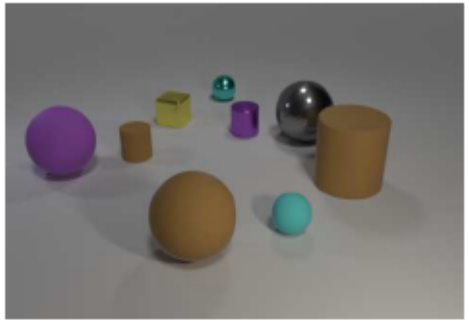
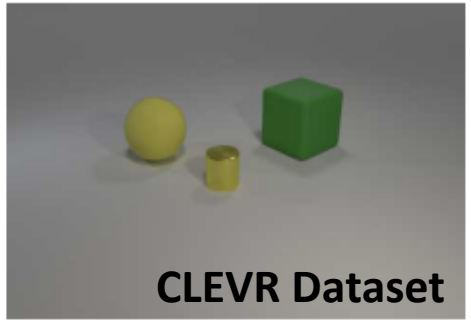
Adaptive Deep Models across Visual Domains, Kate Saenko, ICCV, 2017

Понимание сцены и языка: Visual Question Answering



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

Deep Visual Reasoning for VQA: генератор программ

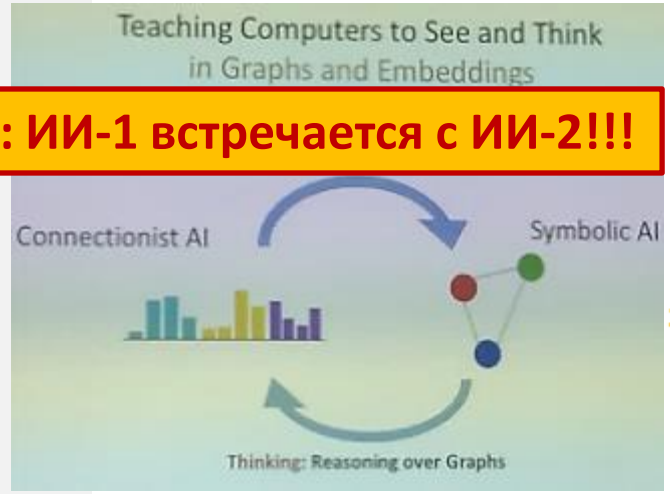
Вопросы, требующие рассуждений:

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*



2017: ИИ-1 встречается с ИИ-2!!!

Predicted Program:

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

Predicted Program:

```
equal_size
query_size
unique
filter_shape[spher
scene
query_size
unique
filter_shape[spher
filter_size[small
```

2017+: Применение ГНС к символическим данным позволяет использовать базы знаний и логический вывод для анализа данных о реальном мире

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

2019+: Прогноз создания функционального ИИ (на примере робототехники)

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 и Starcraft II + обучение с подкреплением + автоматическое обучение

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!

Глубокая оптимизация

(на пути к революции
в технике, технологии и
управлении бизнес-процессами)

Глубокая оптимизация: взгляд-2019+ на революцию ГНС и ее перспективы

2020+: *при помощи ГНС может быть создан «слабый» бортовой ИИ… но это не главное -* ***ГНС могут дать науке и технике гораздо больше!***

На самом деле нет никакого прорыва в методах ИИ - есть прорыв в методах локальной оптимизации, связанный с использованием ГНС.

Мы имеем дело не с новыми методами и подходами в ИИ, а с новой группой мощных инженерных методов: ***«глубокими» методами моделирования, управления и оптимизации***.

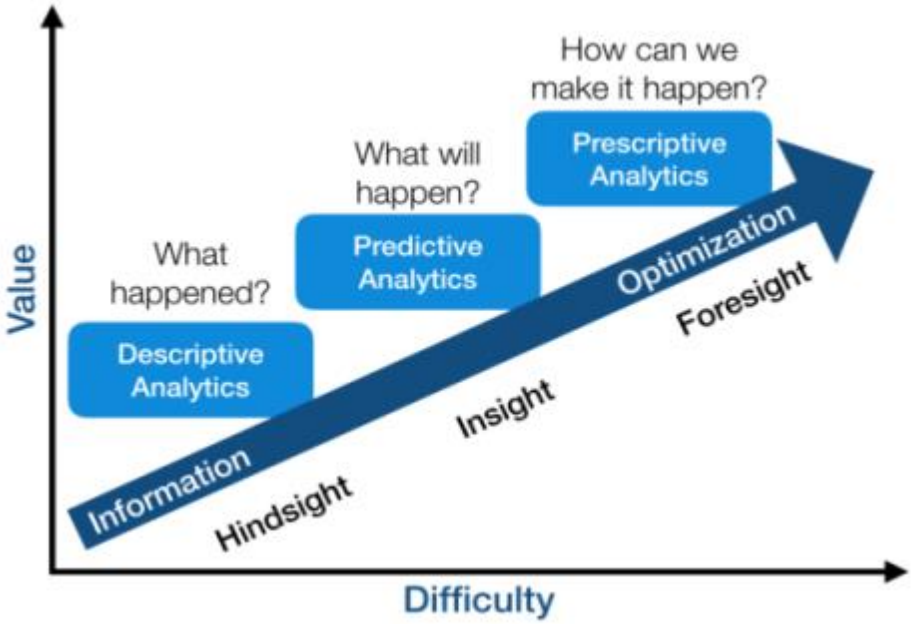
Логика развития этих методов и технологий в последние годы состоит в **переходе от задач обработки и анализа информации к задачам управления и оптимизации.**

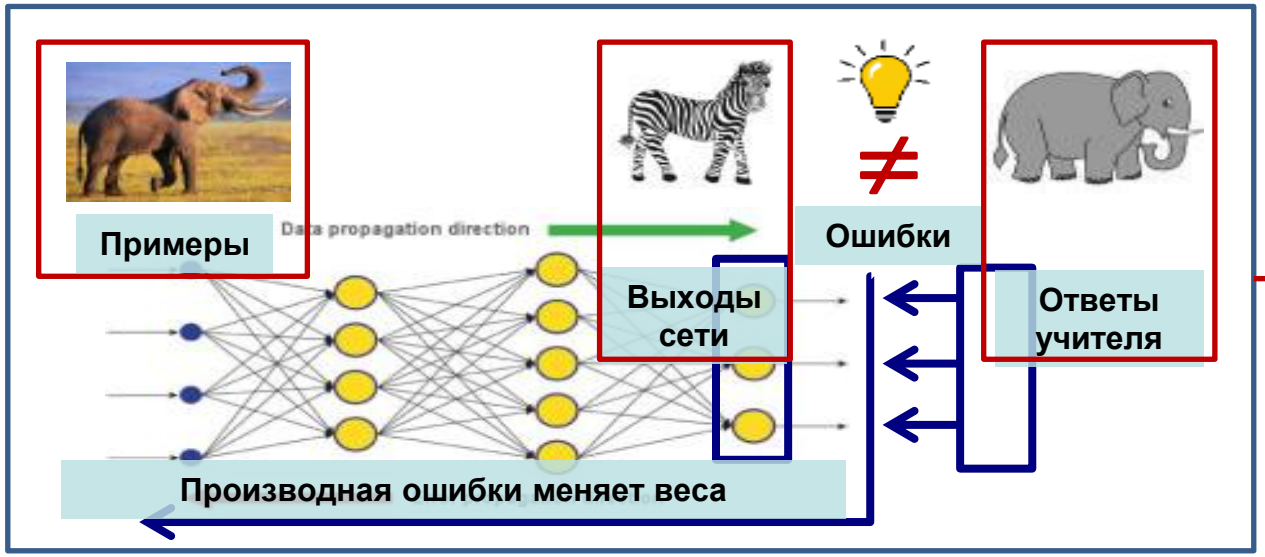
ГНС могут решать любые задачи, которые формулируются как задачи оптимизации.

Перспективные интеллектуальные системы на основе ГНС это уже не только СТЗ и даже не только «бортовой интеллект»! Это про все задачи создания и применения (от облика и алгоритмов до оптимизации производства и эксплуатации) **изделий**.

Переход от глубокого обучения к глубокой оптимизации

Поскольку применение ГНС в оптимизационных задачах машинного обучения оказалось чрезвычайно эффективным,
их начали активно применять и для решения других задач оптимизации…





Обучение = Оптимизация: минимизация ошибок на выборке

Технологии глубокого обучения в классических задачах оптимизации

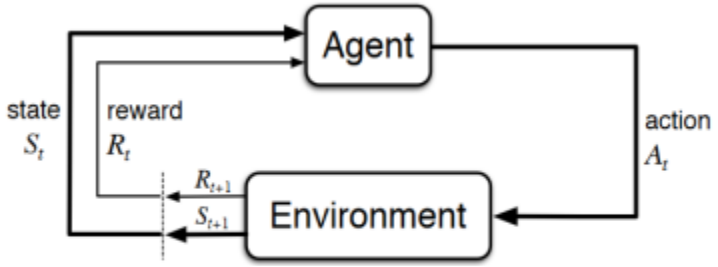
2015+: Глубокие нейросети умеют работать на графах
2015+: ГНС на графах могут выучивать эффективные эвристики решения задач оптимизации.

Deep Reinforcement Learning: обучение с подкреплением

- **Reward $R(t)$:** score you earned at current step
  Вознаграждение (выигрыш после хода)
- **State S:** current screen
  Состояние (что видим на экране)
- **Action i:** move your board left / right
  Действие (что делаем)
- **Action value function $\widehat{Q}(S, i)$:**
  your predicted future total rewards
  Стоимость (выигрыш в будущем)
- **Policy $\pi(s)$:** How to choose your action
  Решение (какой ход выбрать)

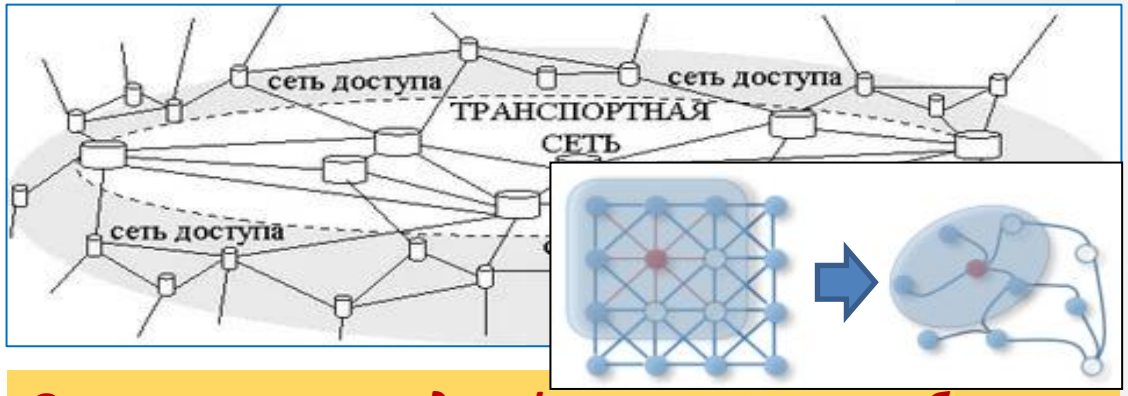
Главный ключ к решению задач глубокой оптимизации: ***глубокое обучение с подкреплением для выучивания эвристик!***

Как научить глубокую сеть (ИИ - агента) выучивать эвристики



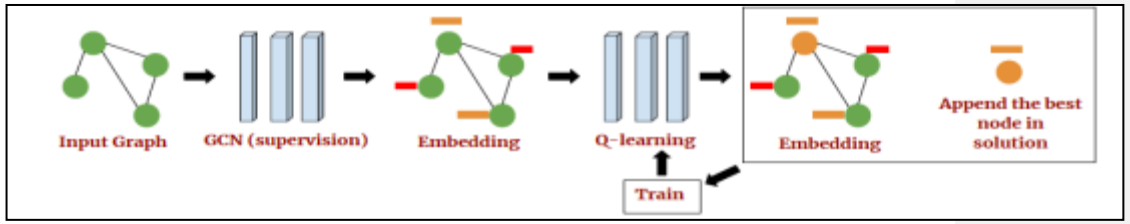
Human-level control through deep reinforcement learning (Minh, et al. Nature, 2015)

Deep Graph Embedding: глубокие сети на графах



Опишем малые подграфы признаками, соберем из них признаки больших подграфов, и так – пока не опишем вектором признаков весь граф

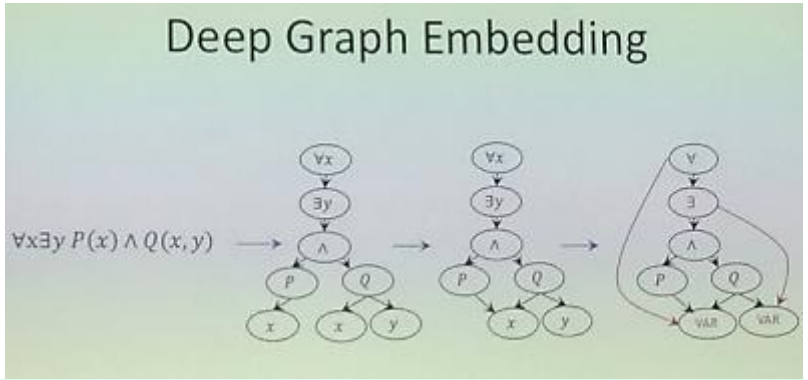
Deep Graph Embedding + Deep Reinforcement Learning

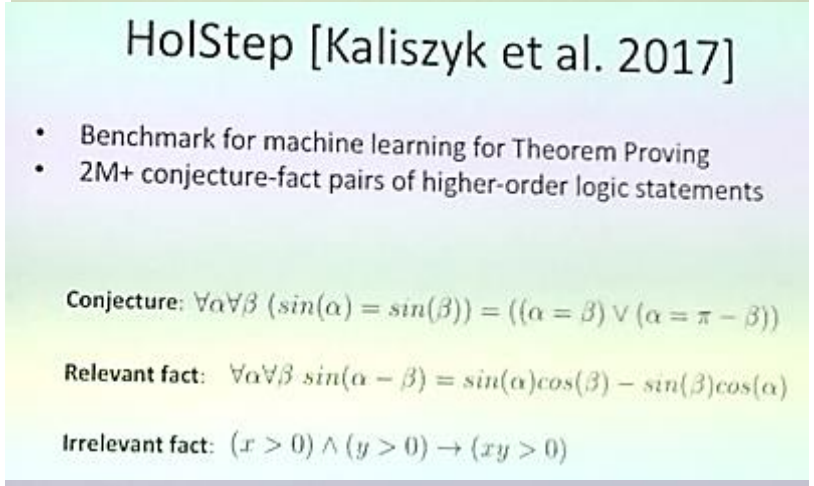


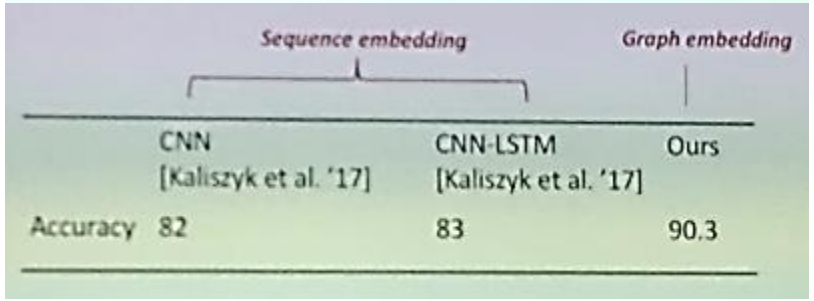
Сегодня (2019+) задачи глубокой оптимизации решаются на графах с миллионами вершин, что позволяет уже переходить к практическому внедрению в самых масштабных приложениях

Learning Heuristics Over Large Graphs Via Deep Reinforcement Learning (Akash Mittal et.al., 2019)

Deep Graph Embedding + Deep Reinforcement Learning
в автоматическом доказательстве теорем (2017-2019)







Seen and Thinking, ICCV, 2017

Более современные работы:

- Urban, J., Kaliszyk, C., Michalewski, H., and Olšák, M. (2018). Reinforcement learning of theorem proving. In *NIPS*. https://arxiv.org/abs/1805.07563
- Automated Theorem Proving in Intuitionistic Propositional Logic by Deep Reinforcement Learning (2108) https://arxiv.org/abs/1811.00796
- HOList: An Environment for Machine Learning of Higher-Order Theorem Proving (extended version) (2019) https://arxiv.org/abs/1904.03241
- https://github.com/tensorflow/deepmath

На самом деле именно глубокая оптимизация и позволила объединить обучаемость ИИ-2 с интеллектуальностью ИИ-1

Технологии глубокого обучения в исследовании операций и управлении

AI Beats a Fighter Pilot in a Virtual Dogfight (2016)

AI ALPHA, built by Psibernetix, Inc. with Air Force Research Laboratory

Оперативно-тактическое и групповое управление

Показатель зрелости: игровые задачи

StarCraft II - полноценный тактический военный симулятор с упрощённой моделью ведения боя.

Прогноз 2018: *ГНС смогут выиграть у людей* ***лет через 5***

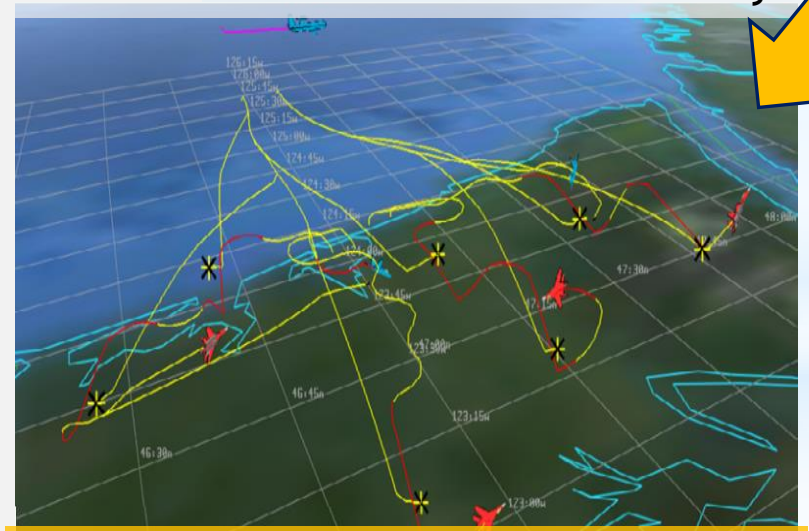
Автономное управление и оптимизация облика изделий

Оптимизация бизнес-процессов и производства

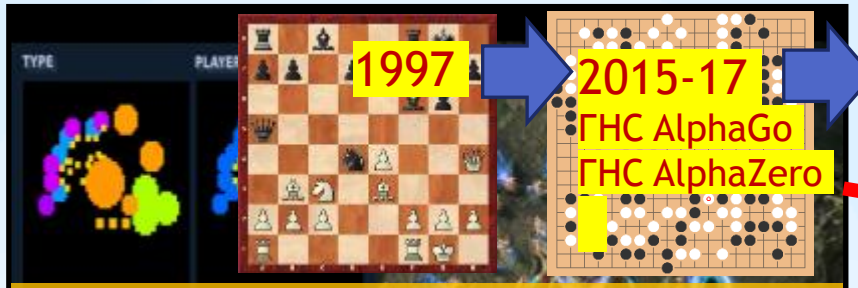
Технологии глубокого обучения в исследовании операций и управлении

AI Beats a Fighter Pilot in a Virtual Dogfight (2016)

AI ALPHA, built by Psibernetix, Inc. with Air Force Research Laboratory



Оперативно-тактическое и групповое управление



Показатель зрелости: игровые задачи

Сочетание технологий ГНС, RL и генетического отбора

2019+: Дорога открыта!

Автономное управление и оптимизация облика изделий

StarCraft II - полноценный тактический военный симулятор с упрощённой моделью ведения боя.

~~***Прогноз 2018: ГНС смогут выиграть у людей лет через 5***~~

25.01.2019: DeepMind AlphaStar со счетом 11:1 победила ведущих профессиональных игроков в StarCraft II

ХИТ ИИ 2019!

Оптимизация бизнес-процессов и производства

«Глубокая оптимизация» как второй этап «цифровизации»

После того, как для некоторого объекта или системы создана достаточно точная действующая модель (**«цифровой двойник»**), эта модель сразу может быть использована для оптимизации этого объекта, либо формирования наилучших алгоритмов взаимодействия с ним. Это в равной степени относится к изделиям, их группам, войсковым соединениям или же промышленным предприятиям.

Логика развития и внедрения «глубоких» технологий: от обработки информации к управлению и оптимизации

Таким образом, **этап создания и применения интеллектуальных средств «глубокой оптимизации» изделий и процессов на основе ГНС является логическим продолжением этапа создания «цифровых двойников»**

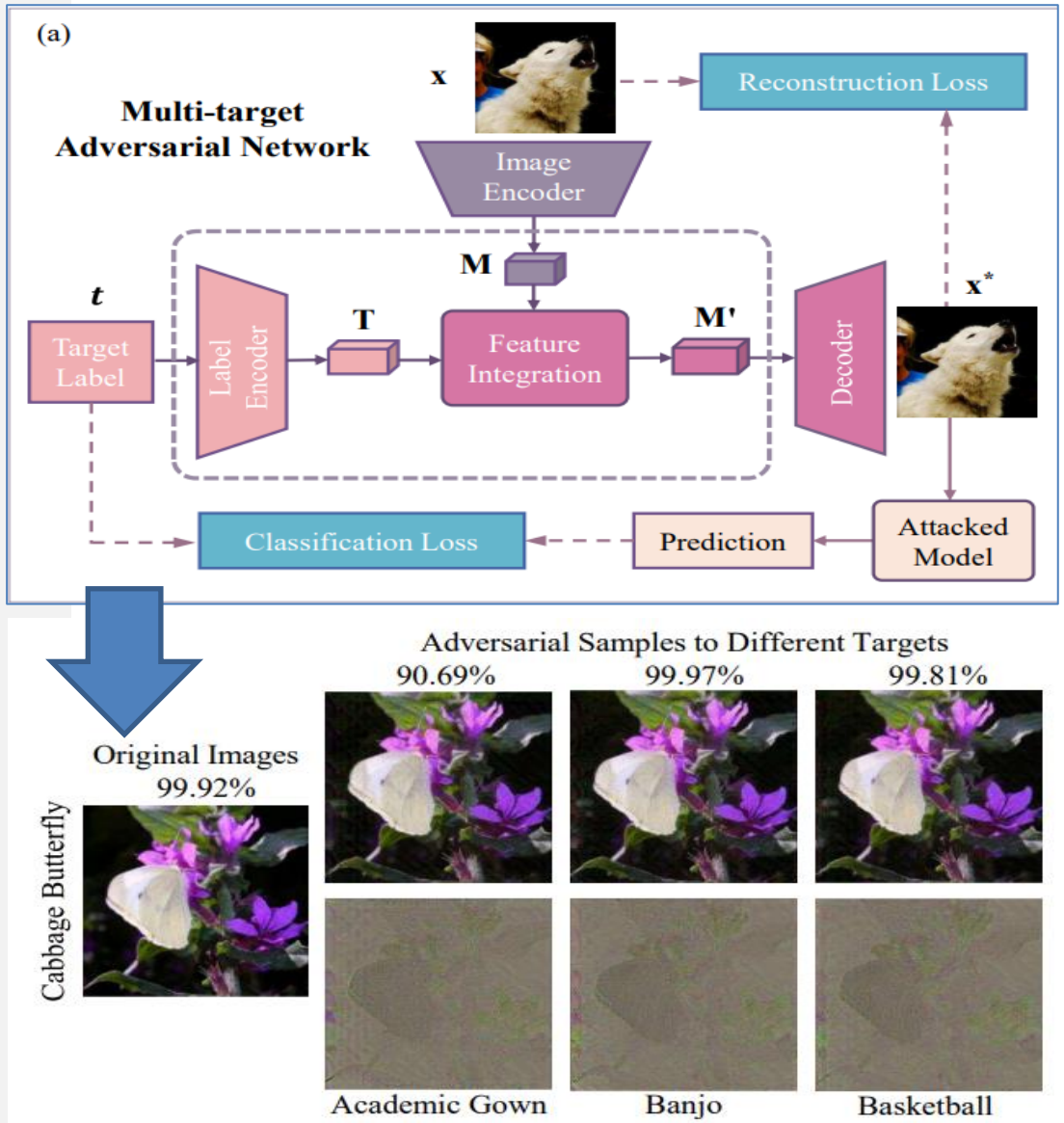
Актуальные результаты и открытые проблемы

(угрозы, вызовы, надежды)

На примерах из области компьютерного зрения и не только…

Атаки на нейросети (Adversarial Attack)

Атаки на распознающие нейронные сети

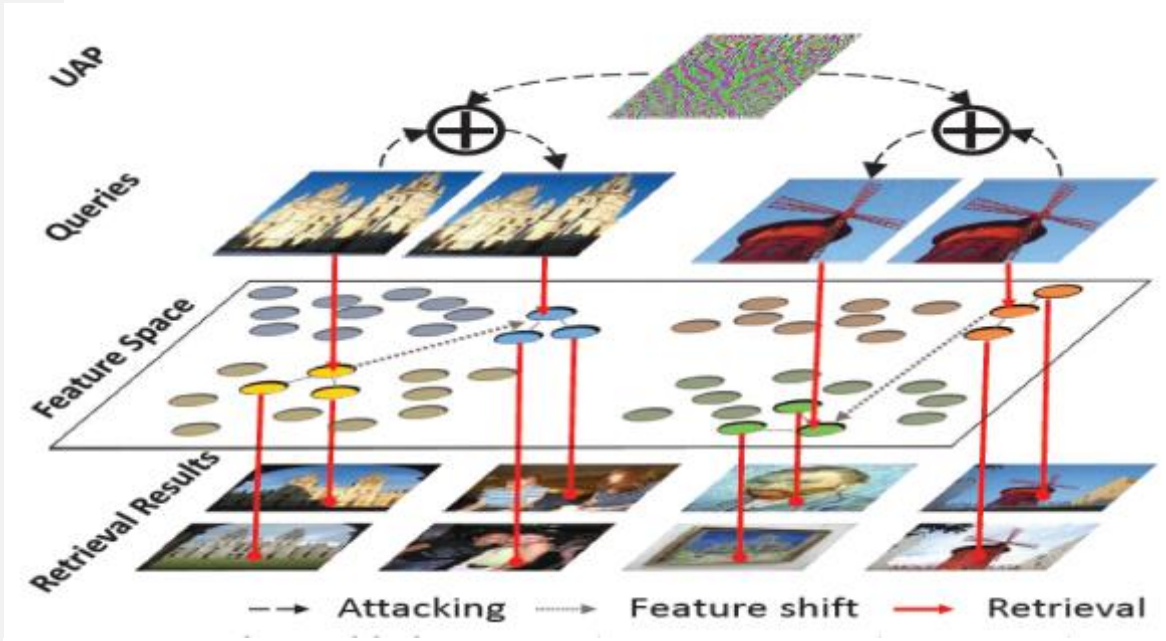


Once a MAN: Towards Multi-Target Attack via Learning Multi-Target Adversarial Network Once, ICCV-2019, Jiangfan Han, Xiaoyi Dong, Ruimao Zhang, Dongdong Chen, Weiming Zhang, Nenghai Yu, Ping Luo, Xiaogang Wang

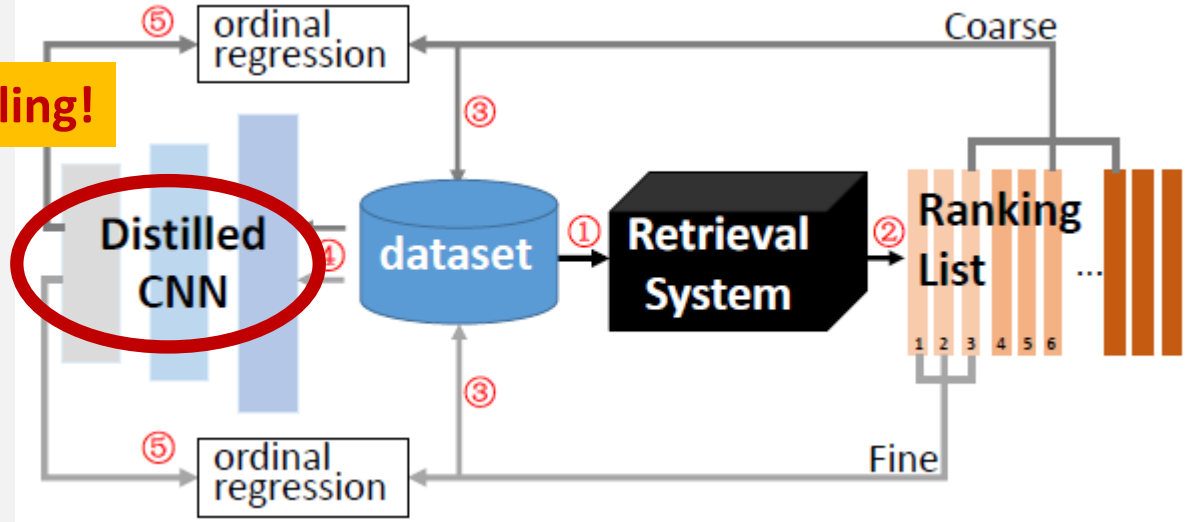
Атака: ***намеренное искажение данных или обучающих выборок с целью вызвать неверное срабатывание ГНС***

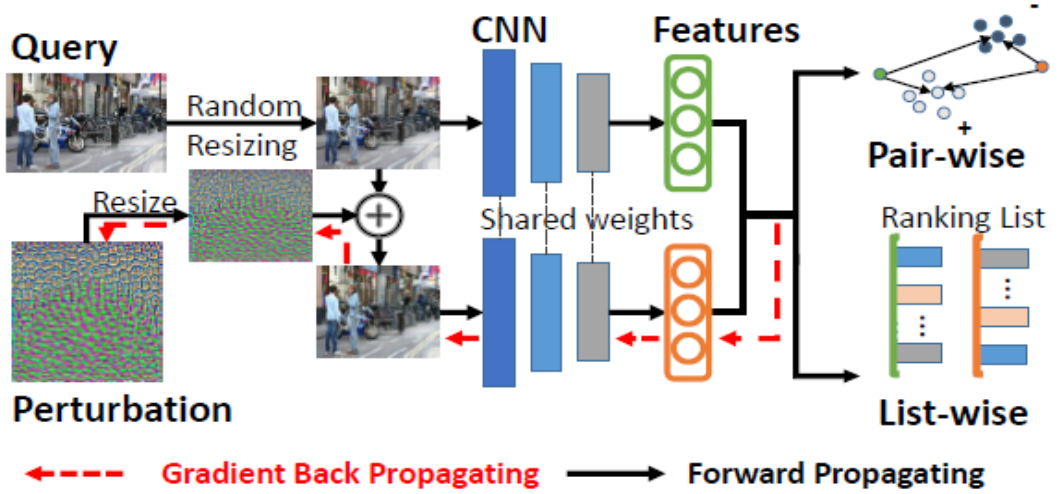
Sparse and Imperceivable Adversarial Attacks
Francesco Croce, Matthias Hein
University of Tubingen

Атаки на поисковые нейронные сети



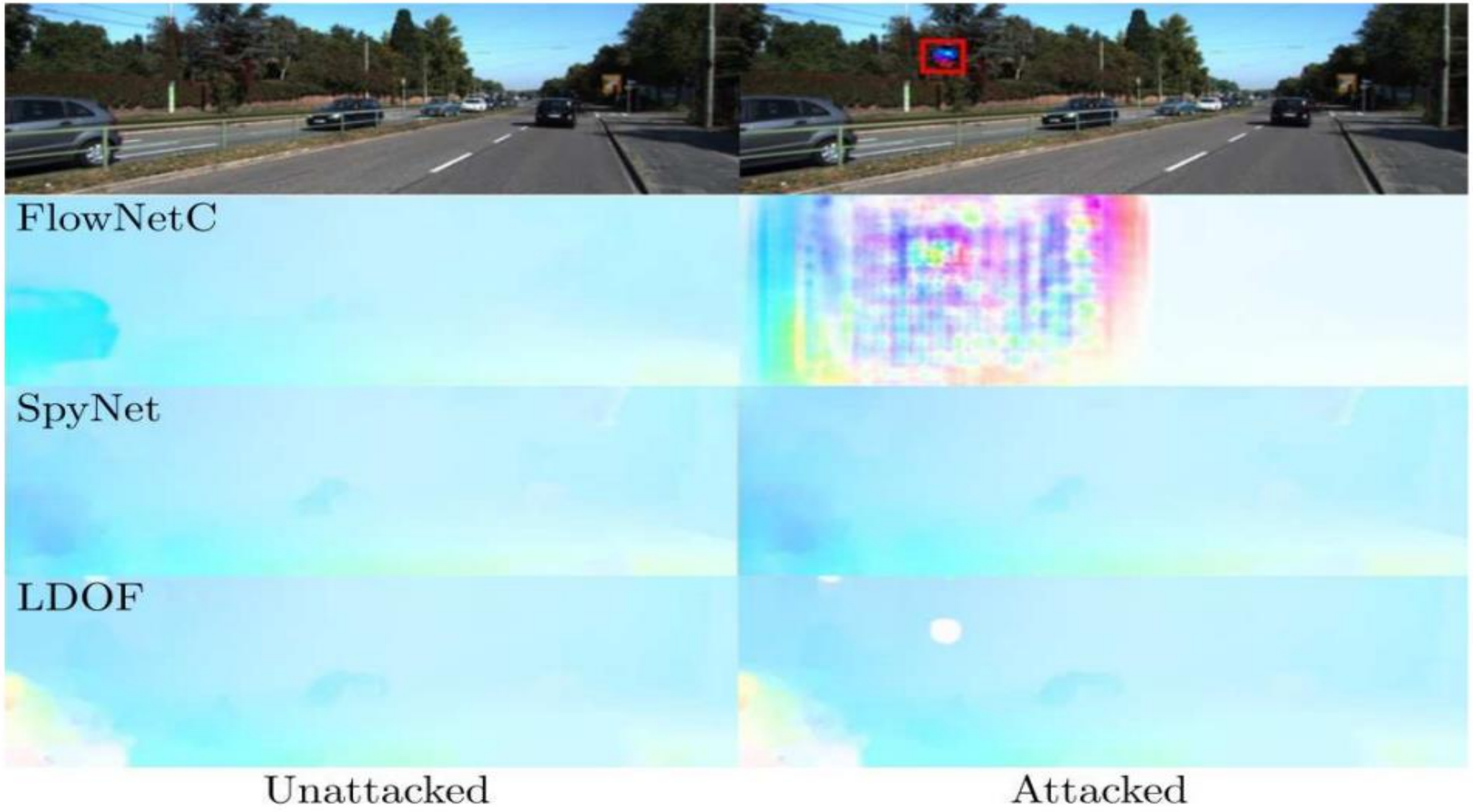
Distilling!





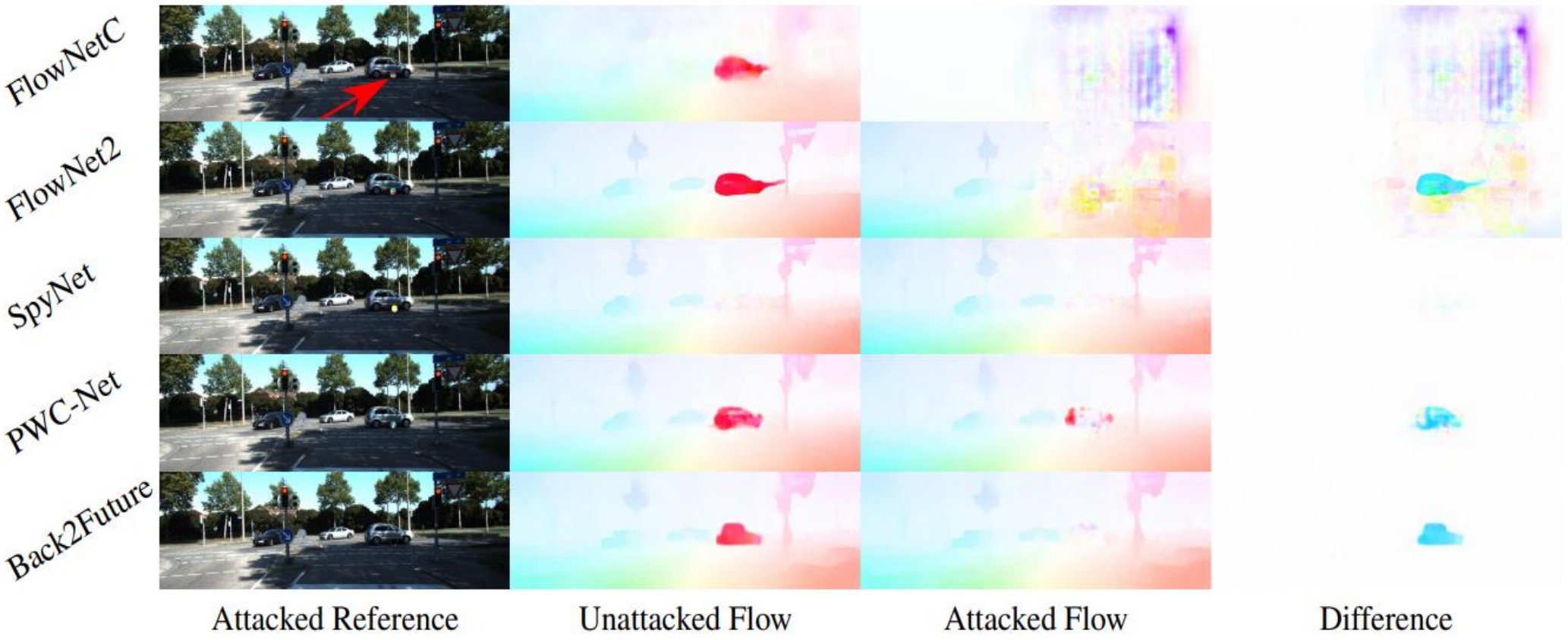
Universal Perturbation Attack Against Image Retrieval
Jie Li, Rongrong Ji, Hong Liu, Xiaopeng Hong, Yue Gao, Qi Tian

Атаки на оптический поток (в автономном вождении)



Attacking Optical Flow, ICCV-2019
Anurag Ranjan, Joel Janai, Andreas Geiger, Michael J. Black

Атаки на оптический поток (в автономном вождении)



Результатом такой атаки в реальном мире может стать авария с человеческими жертвами…

Attacking Optical Flow, ICCV-2019
Anurag Ranjan, Joel Janai, Andreas Geiger, Michael J. Black

Исследование уязвимостей и поиск защиты ГНС (ГосНИИАС, 2018-19)

Предложена методика извлечения адаптивных корреляционных эталонов из распознающих ГКНС для оценки причин возникновения и разработки методов борьбы с пиксельными и другими «атаками» на распознающие ГНС

Для распознавания атак была обучена отдельная сеть AttackFinder архитектуры LeNet-5 (5 слоев), которая получает на вход ядро, восстановленное первой сетью, и по виду ядра определяет, была ли исходная картинка атакована, т.е. соответствует ли полученное ядро "типичным представителям" этого класса. Результаты экспериментов показали, что **предложенный подход может не только детектировать атаки на ГНС, но и обнаруживать собственные ошибки распознавания анализируемых ГНС.**

Linear Kernel Auto Generation for Adversarial Attack Detection, Elena Entina, Vladimir Gorbatsevich, Yury Vizilter

Основная идея: встроить в ГНС средства самоконтроля. Контроль ошибок основан на форме ядра CNN

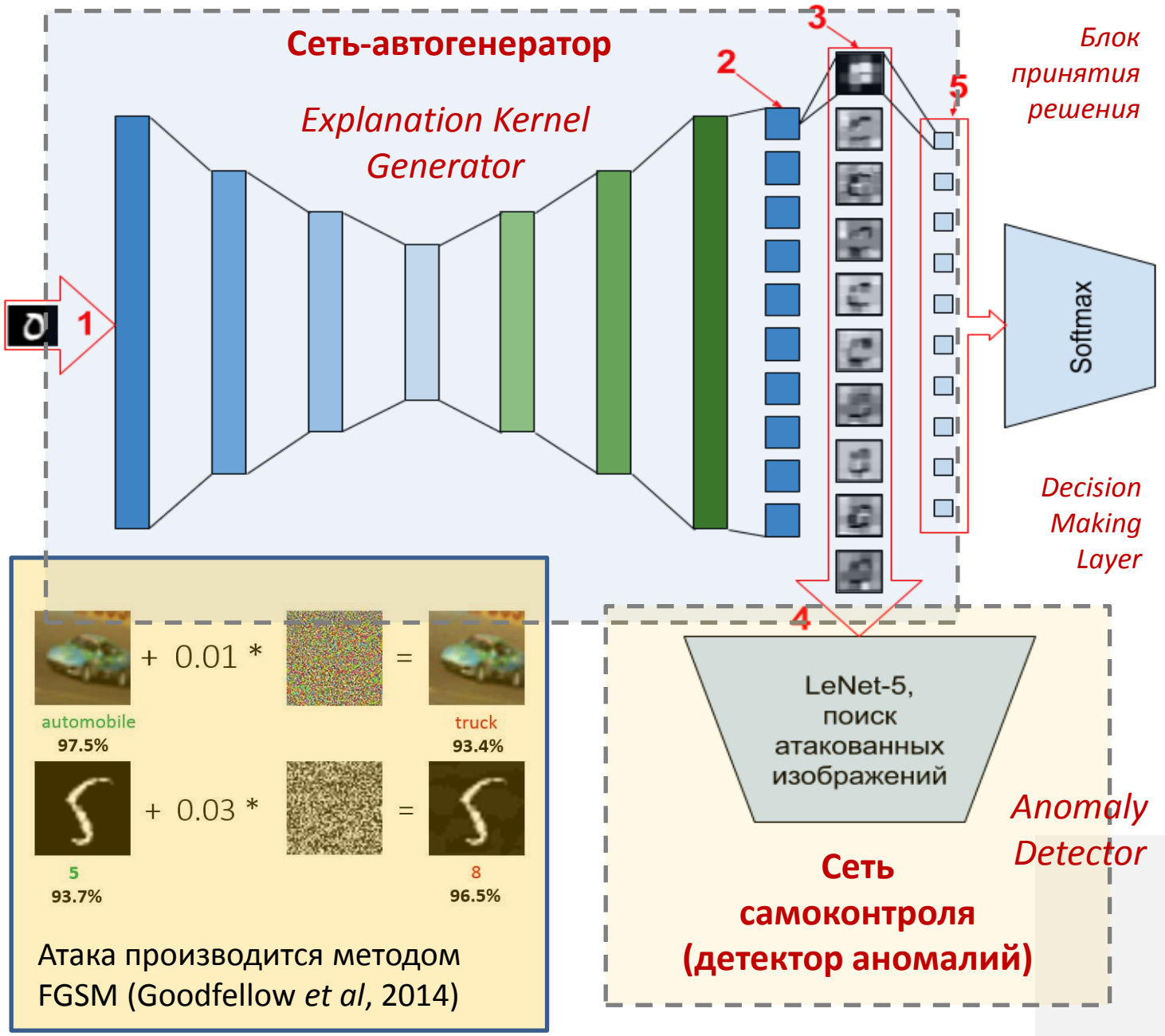
Исследование уязвимостей и поиск защиты ГНС (ГосНИИАС, 2018-19)

	MNIST	CIFAR-10
исходная точность	97,5%	82,3%
% найденных атак	89,9%	94,6%
% ложных срабатываний	8,8%	34,6%
точность на прошедших	99,7%	97,0%

1. Самоконтроль атак возможен, можно находить до 99% атак;

2. Точность классификатора с самоконтролем драматически возрастает;

3. Сеть с самоконтролем может **обнаруживать собственные ошибки распознавания как «real-world» атаки**.



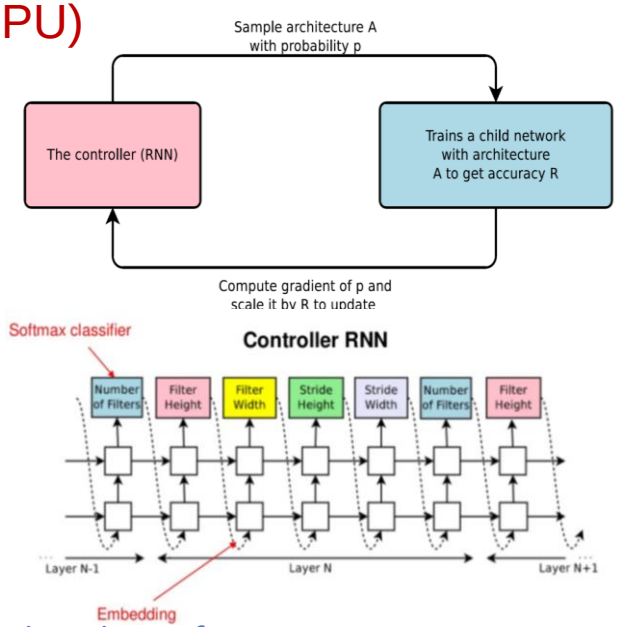
Автоматическое формирование и обучение нейросетей (AutoML, Neural Architecture Search, NAS)

AutoML: автоматическое обучение глубоких сетей

Первое поколение AutoML (2016-2018)

- Лучшие результаты в задаче **классификации** на CIFAR-10/ImageNet (выше придуманных человеком архитектур)
- *Вычислительные требования*: Сотни серверов с GPU/TPU (~500-800 GPU)

2017: Автоматически сформированные глубокие сети впервые превзошли показатели глубоких сетей, сформированных вручную

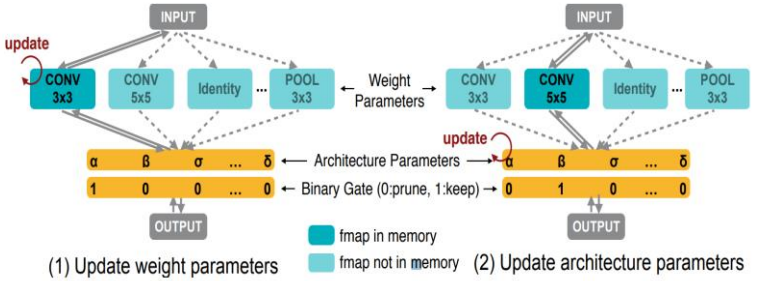


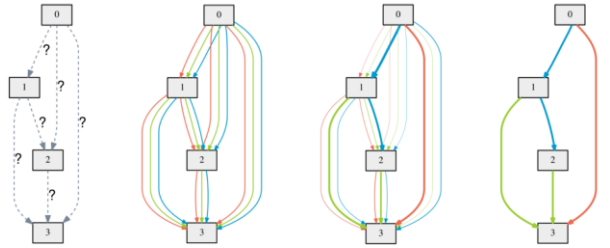
*Neural Architecture Search with Reinforcement Learning
Barret Zoph∗ , Quoc V. Le Google Brain ICCV 2017
*Learning Transferable Architectures for Scalable Image Recognition
Barret Zoph,Vijay Vasudevan, Jonathon Shlens, Quoc V. Le CVPR 2018
*Regularized Evolution for Image Classifier Architecture Search
Esteban Real, Alok Aggarwal, Yanping Huang, Quoc V Le 2018

Второе поколение AutoML (2018-2019)

- Лучшие результаты для задач **обнаружения** и распознавания
- Учет специфики задачи и архитектуры конечного вычислителя
- *Вычислительные требования*: от 200 GPU/часов – сравнимо с обычным обучением

Процесс оптимизации: подбор фильтров, блоков, слоев и параметров в заданном словаре

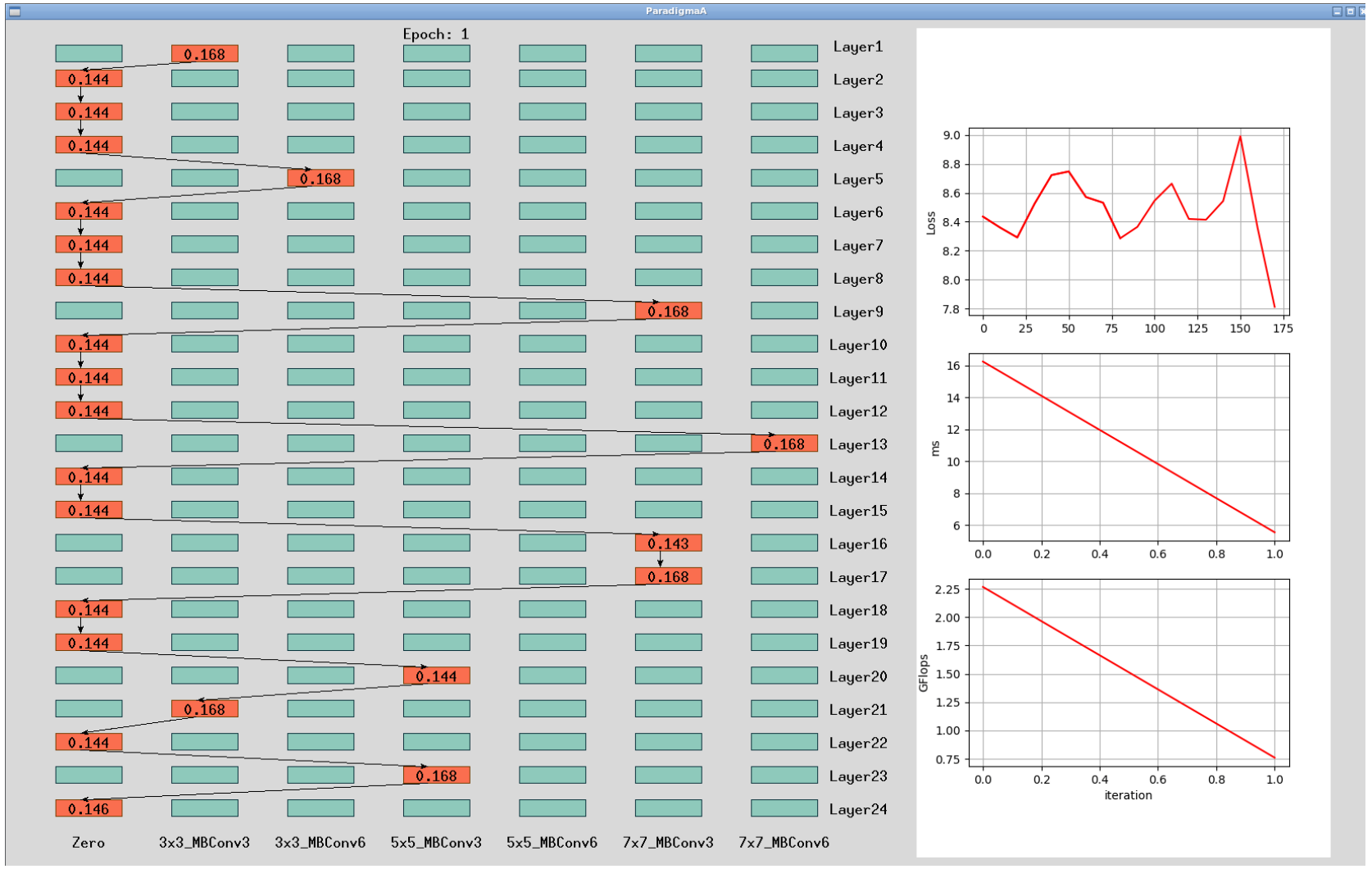




2025+: Перспектива полной автоматизации процессов обучения

*PROXYLESSNAS: DIRECT NEURAL ARCHITECTURE SEARCH ON TARGET TASK AND HARDWARE
Han Cai, Ligeng Zhu, Song Han arxiv 2019
*DARTS: DIFFERENTIABLE ARCHITECTURE SEARCH
Hanxiao Liu, Karen Simonyan, Yiming Yang ICLR 2019

Алгоритм ProxyLessSSD *(ГосНИИАС, 2020)*



Процесс поиска архитектуры

Функция потерь

Время работы сети

Вычислительная сложность сети

Результаты *(ГосНИИАС, 2020)*

ФОНД ПЕРСПЕКТИВНЫХ ИССЛЕДОВАНИЙ

Результаты тестов на БД VisDrone:

Название ГКНС	mAP, mean average precision	mAR, mean average recall	FLOPs (640x480)	число парам-ов (640x480)	Время, мс на GPU 1080Ti (640x480) batch=1
SSD ProxylessNAS	34.2(+8%)	0.264(+6%)	1.9G(+1%)	1.48M(-26%)	9.24(-22%)
SSD MobileNet v2 (manual) 2.0	31.5	0.249	1.88G	2.02M	11.33
SSD MobileNet v1 (manual) 1.0	19.9	0.158	4,56G	5,16M	11.5
SSD MobileNet v3 small x1.5 (AutoML)	23.8	0.189	1.57G	1.69M	15
SSD ShuffleNet v2 (manual)	22.2	0.19	0.88G	1.53M	13.36

Базовые алгоритмы VisDrone 2019

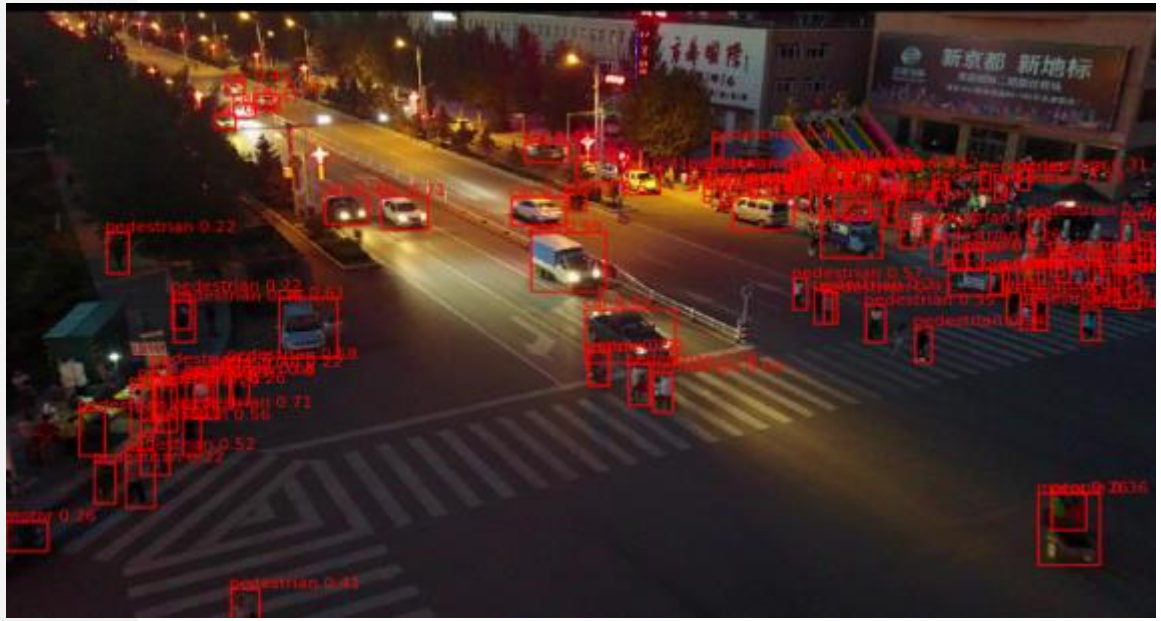
Название ГКНС	mAP	mAR
CornerNet	34.12	24,37
Light-RCNN	32.78	23,09
FPN	32.20	20,72
Cascade R-CNN	31.91	21,37
DetNet59	29.23	20,87

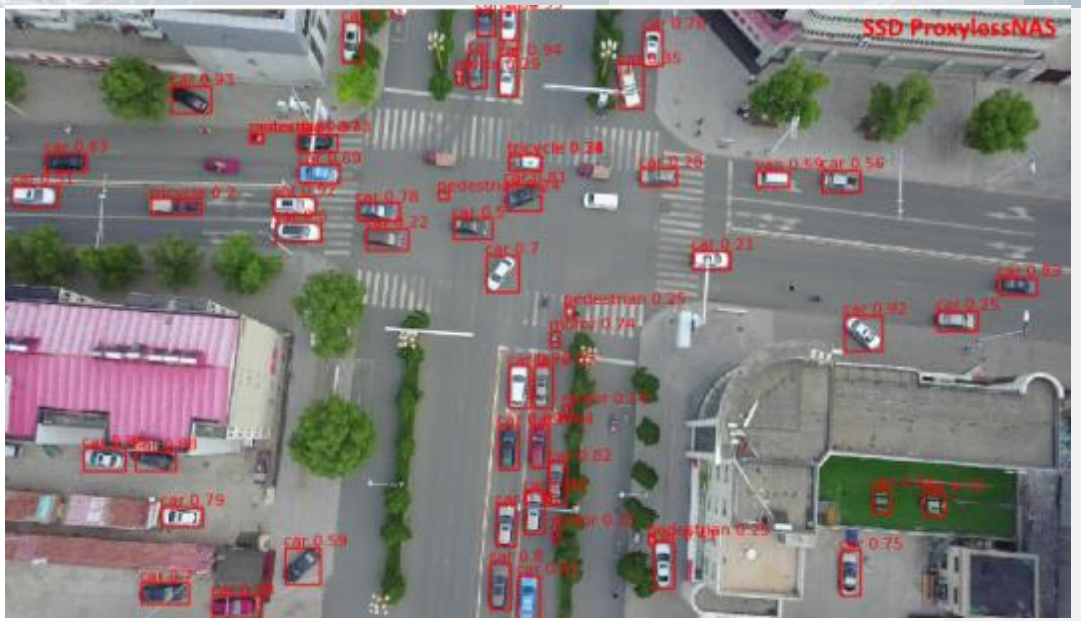
Самый быстрый алгоритм (visDrone 2019):
3 кадра в сек.
Самый медленный:
10 сек. на кадр.

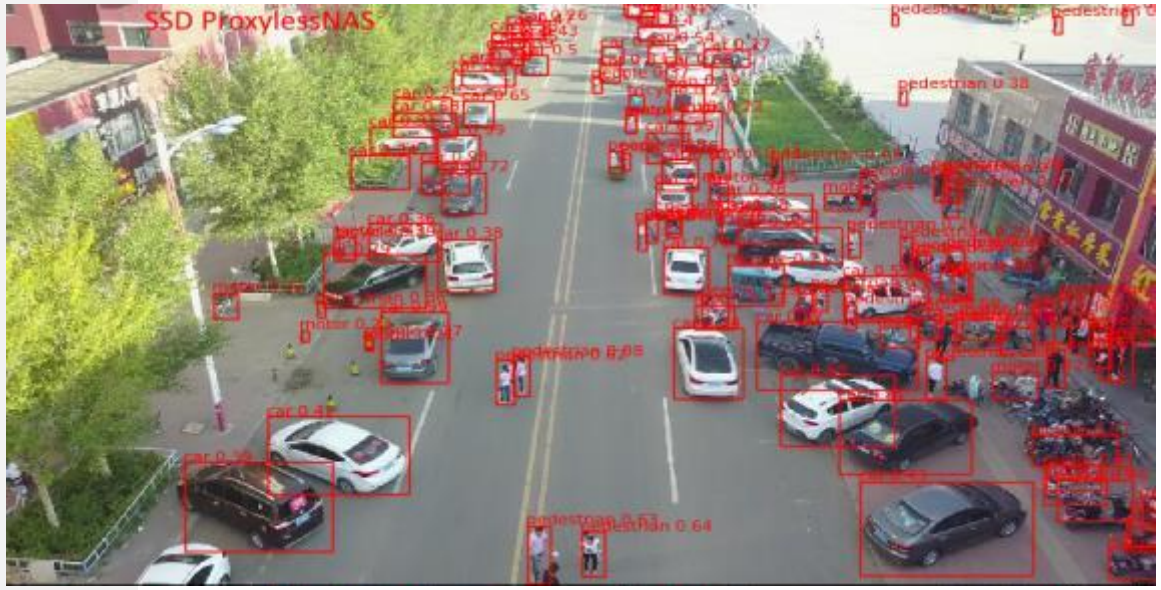
Полное время обучения:
432 GPU/часа

Примеры работы *(ГосНИИАС, 2020)*

ФОНД ПЕРСПЕКТИВНЫХ ИССЛЕДОВАНИЙ







Exploring Randomly Wired Neural Networks for Image Recognition

Saining Xie, Alexander Kirillov, Ross Girshick, Kaiming He

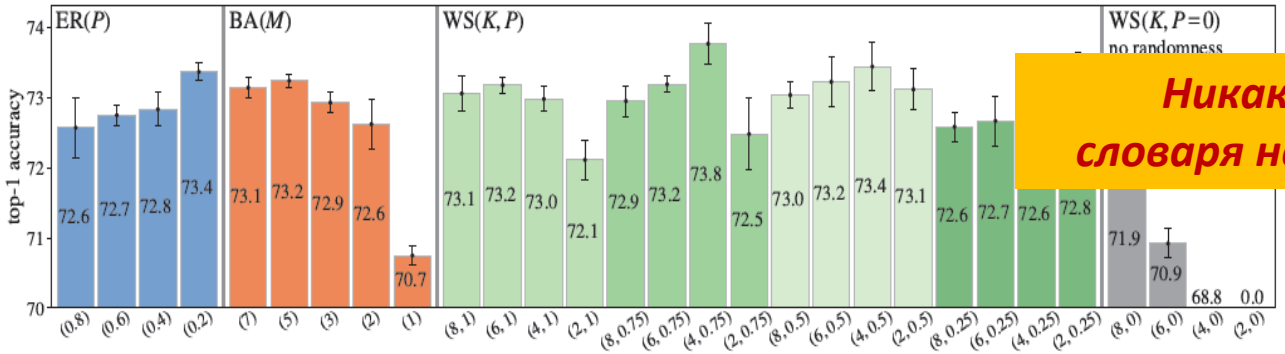
Facebook AI Research (FAIR)

Три случайные архитектуры превзошли ResNet-50!

Figure 1. **Randomly wired neural networks** generated by the classical Watts-Strogatz (WS) [51] model: these three instances of random networks achieve (left-to-right) 79.1%, 79.1%, 79.0% classification accuracy on ImageNet under a similar computational budget to ResNet-50, which has 77.1% accuracy.

1) порождаем случайный граф
2) определяем вход и выход
3) строим и обучаем GCN

Эта работа взрывает все наше понимание NAS!

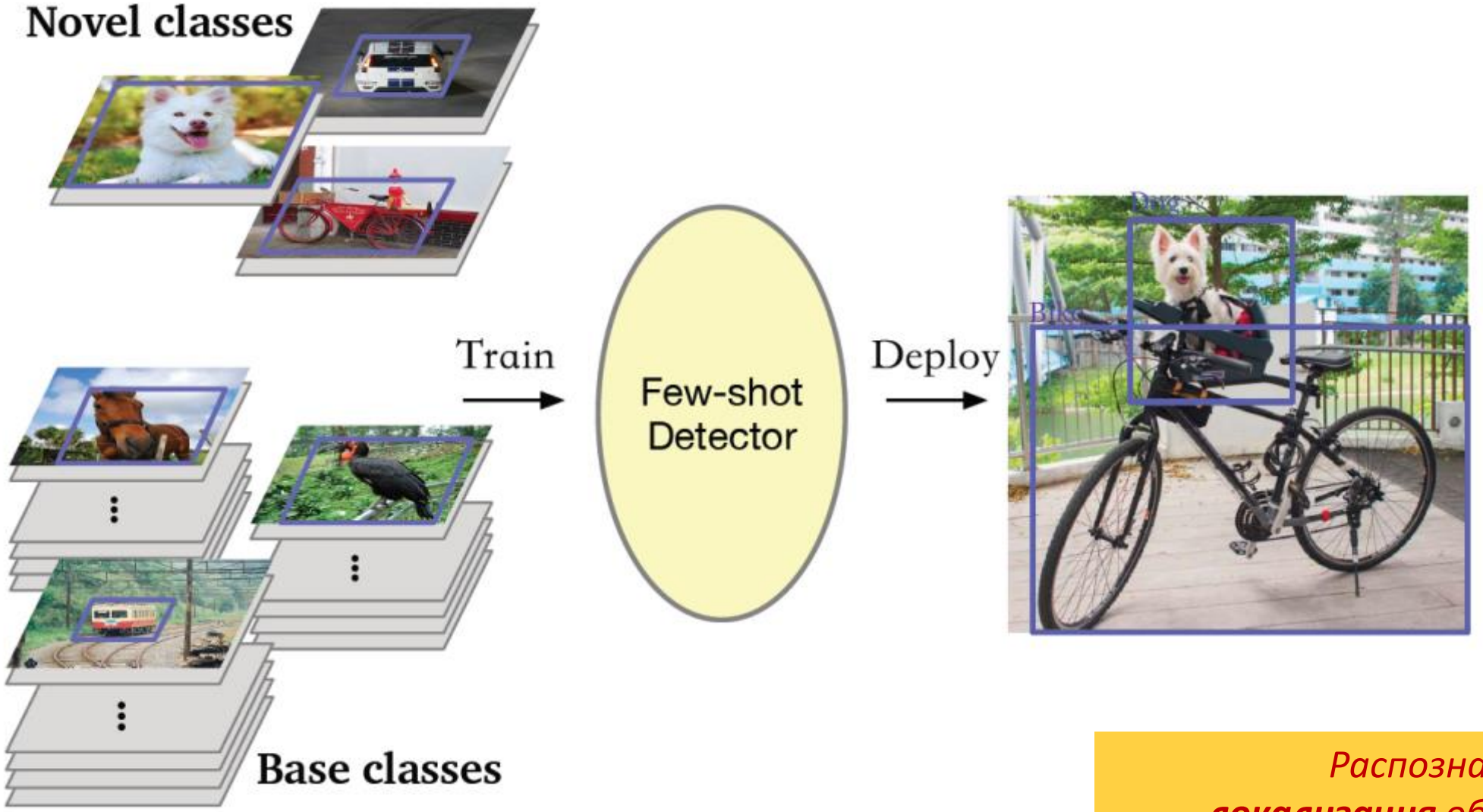


Никакого словаря нет!

Figure 3. **Comparison on random graph generators: ER, BA, and WS** in the small computation regime. Each bar represents the results of a generator under a parameter setting for P, M, or (K, P) (tagged in x-axis). The results are ImageNet top-1 accuracy, shown as mean and standard deviation (std) over 5 random network instances sampled by a generator. At the rightmost, WS$(K, P{=}0)$ has no randomness.

Обучение на малом числе примеров (Few-Shot Learning / Detection / Segmentation)

Few-Shot Learning / **Detection** / Segmentation



Few-Shot Object Detection via Feature Reweighting
Bingyi Kang, Zhuang Liu, Xin Wang, Fisher Yu, Jiashi Feng, Trevor Darrell

Распознавание и ***локализация*** *объектов по одному или малому числу эталонов*

Few-Shot Learning / **Detection** / Segmentation

Результаты на выборке PASCAL VOC 2007 ***(ГосНИИАС, 2019)***

Результаты работы на классах, участвующих в обучении (AP)

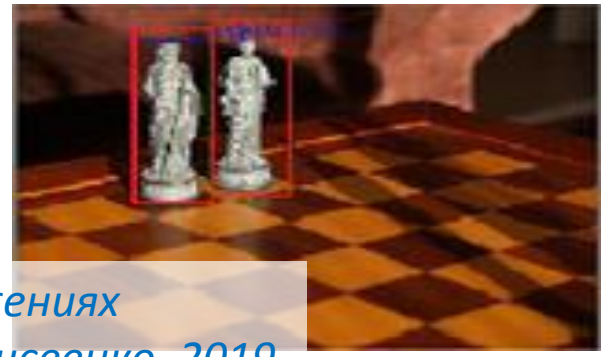
bike	bird	bottle	bus	car	chair	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv	mAP(%)
0.72	0.53	0.51	0.73	0.8	0.44	0.59	0.69	0.76	0.71	0.72	0.35	0.68	0.62	0.69	0.59	0.63

Результаты работы на классах, не участвующих в обучении (AP)

aero	boat	cat	cow	mAP(%)
0.25	0.11	0.66	0.73	0.44

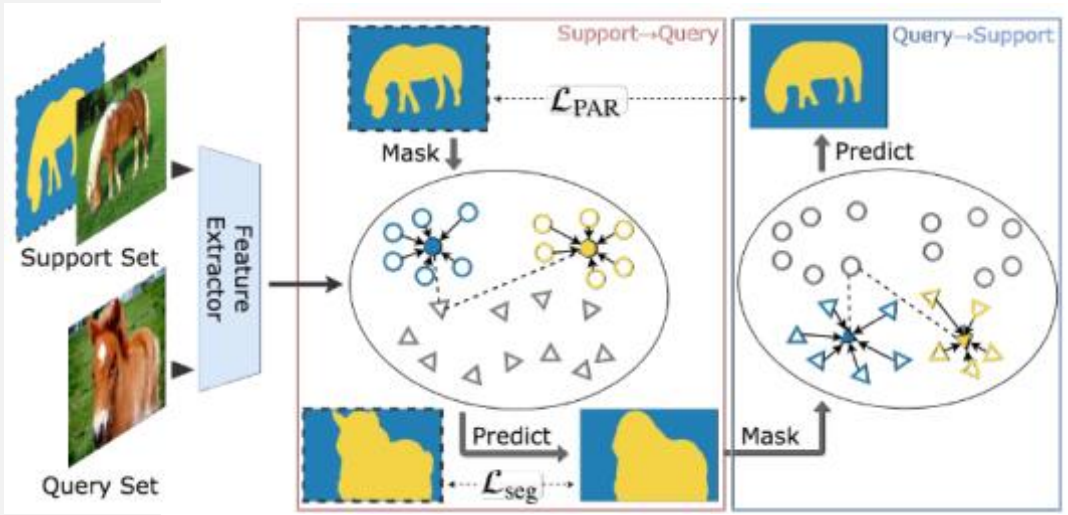
128x128 эталонное изображение

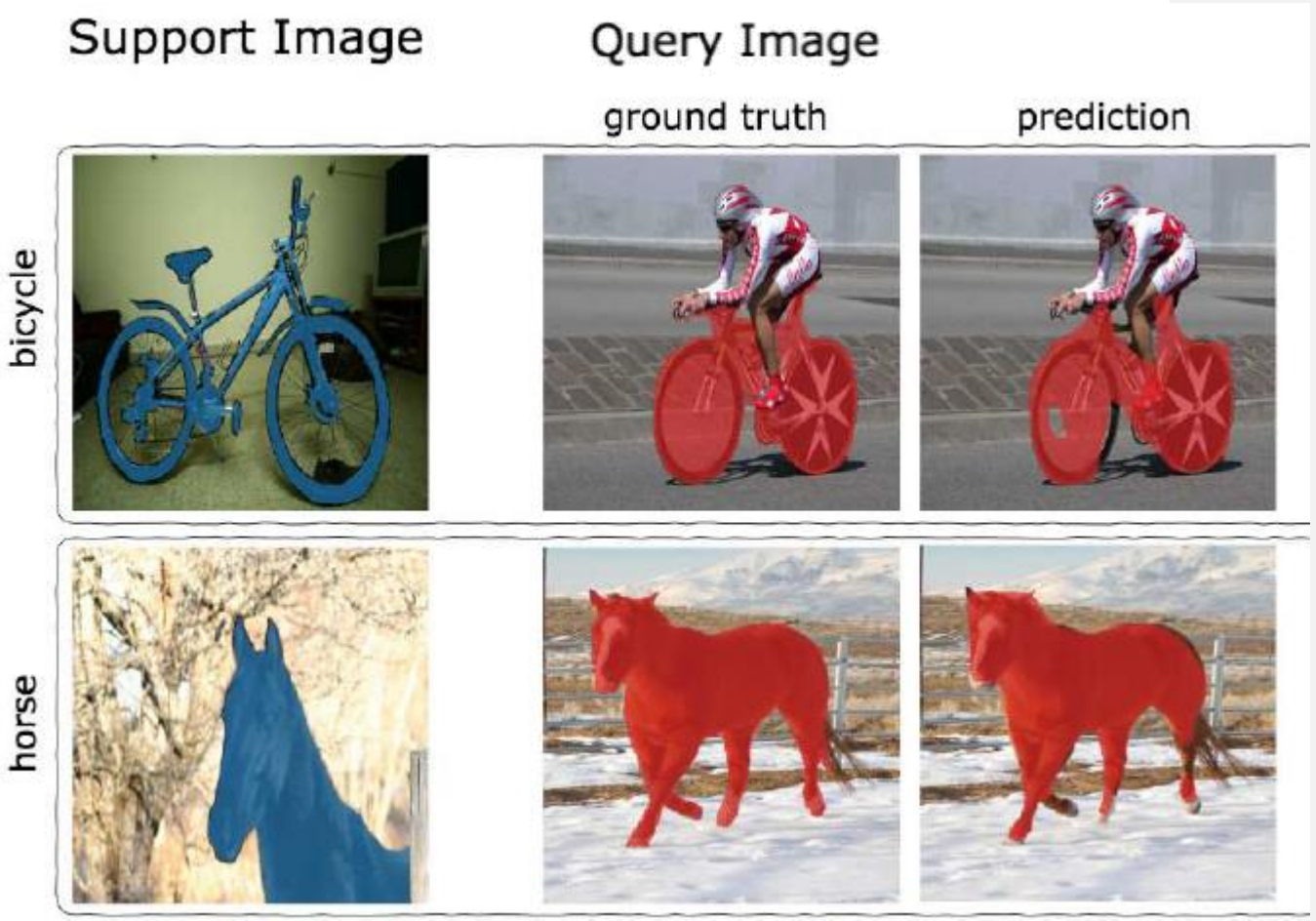
Тестовое изображение



Семантическое обнаружение объектов на изображениях с использованием ГКНС, Горбацевич, Визильтер, Моисеенко, 2019

Few-Shot Learning / Detection / **Segmentation**



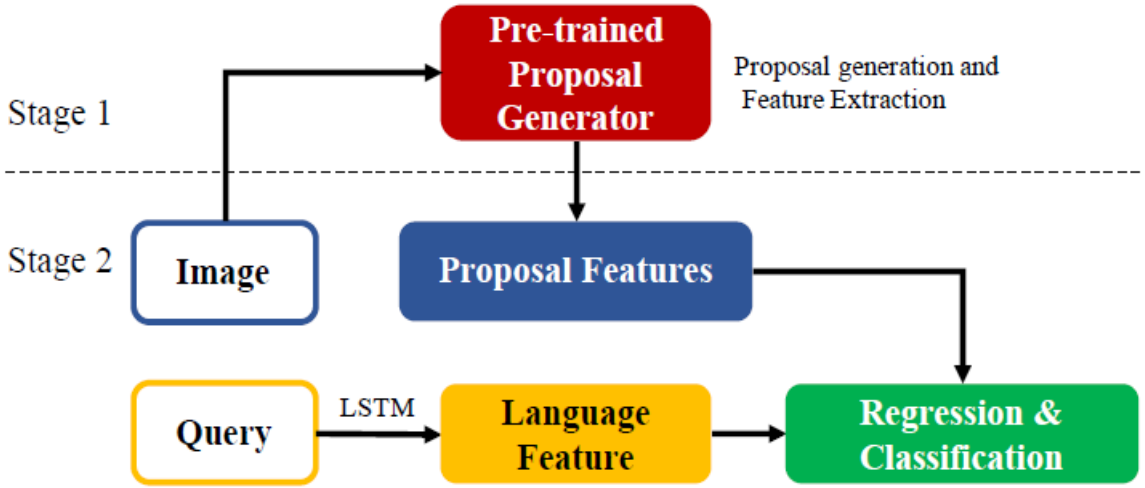


PANet: **Few-Shot** Image **Semantic Segmentation** With Prototype Alignment
Kaixin Wang, Jun Hao Liew, Yingtian Zou, Daquan Zhou, Jiashi Feng

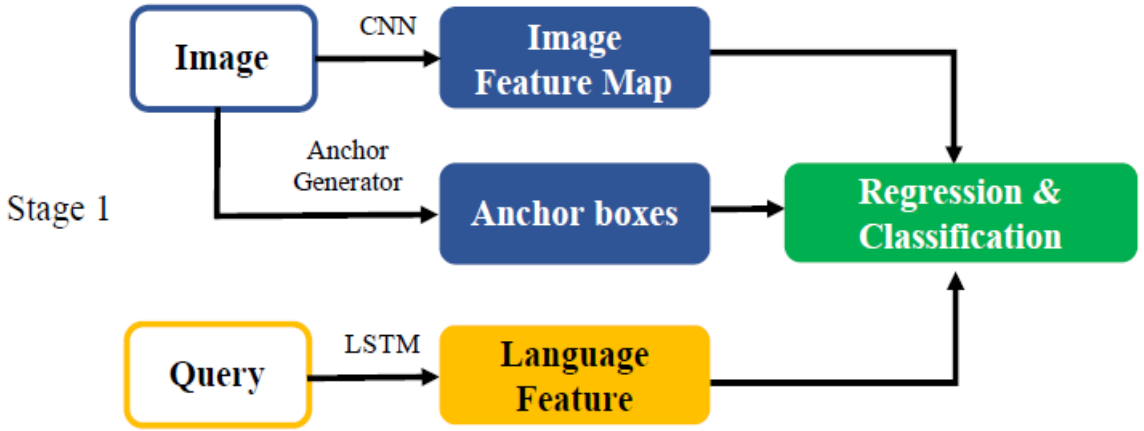
*Распознавание, локализация и **сегментация** объектов по одному или малому числу эталонов*

Обучение без примеров (Zero-Shot Learning, Grounding)

Zero-Shot Grounding



(a) Vanilla 2-stage phrase grounding system



(b) Our 1-stage phrase grounding system

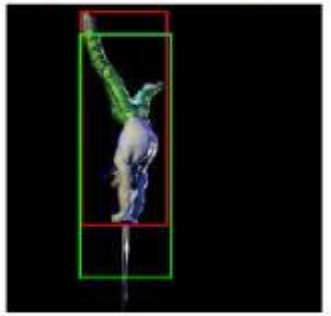

a group of older men


a red beanie cap

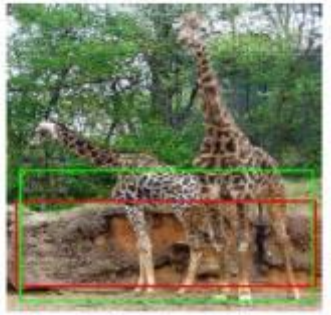

rightmost animal


a cigar


a two-seat kayak


a handstand


a rocky *cliff* (*hill*)

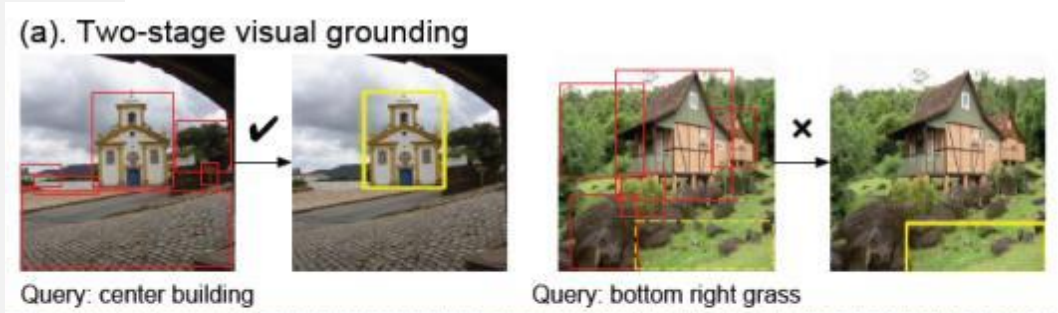

large *boulders* (*rock*)


stairway (*wall*)

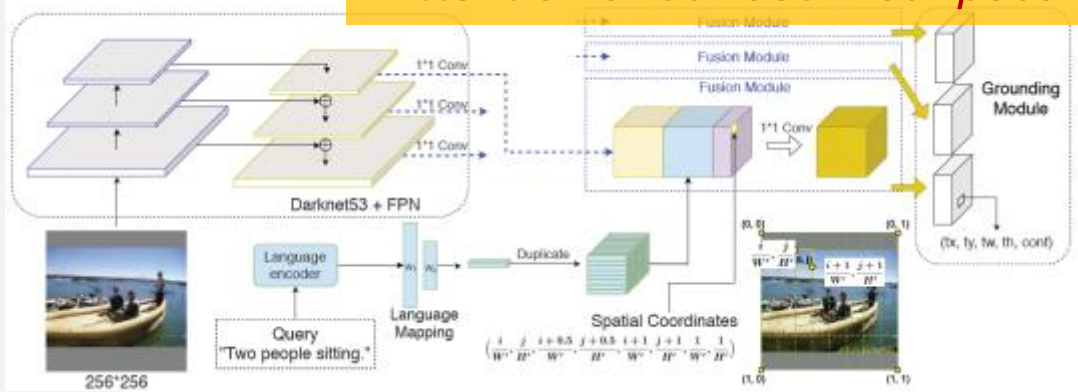
Zero-Shot Grounding of Objects From Natural Language Queries
Arka Sadhu, Kan Chen, Ram Nevatia

Поиск и локализация объектов без эталона (по описанию)

Zero-Shot Grounding



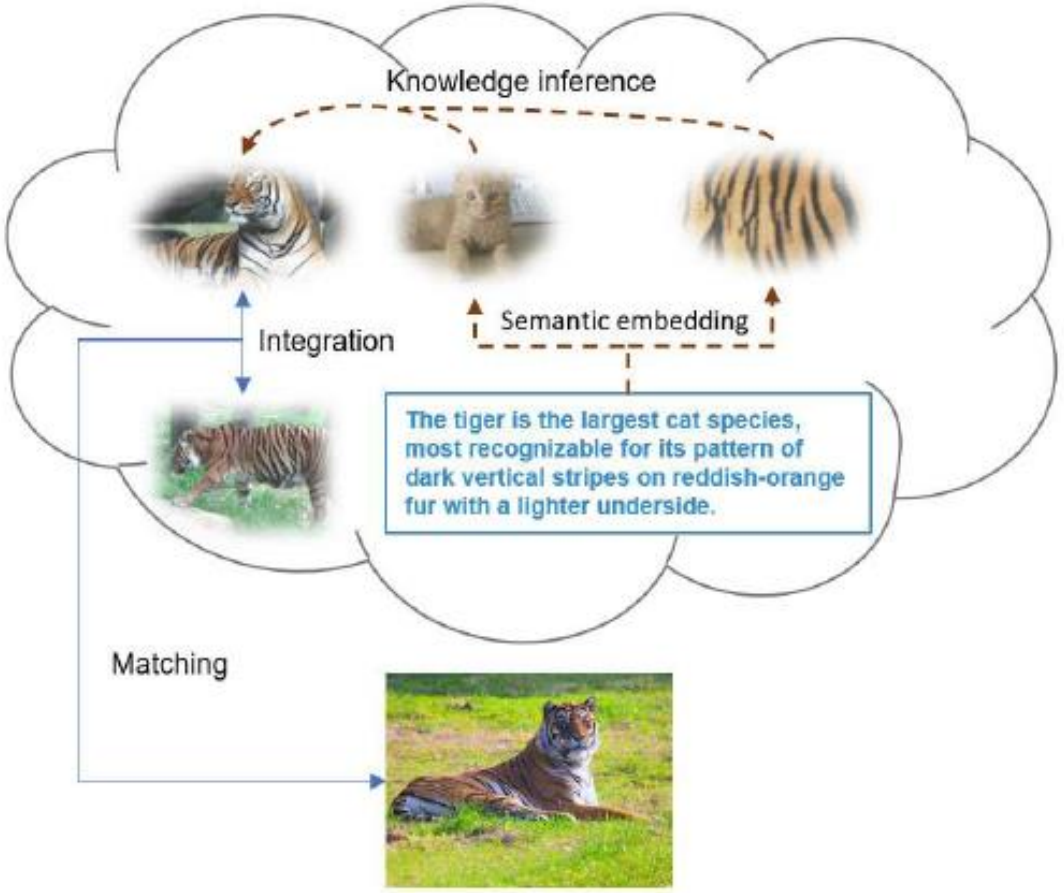
Одноэтапный детектор + Attention с языковым запросом



A Fast and Accurate One-Stage Approach to **Visual Grounding**
Zhengyuan Yang, Boqing Gong, Liwei Wang, Wenbing Huang, Dong Yu, Jiebo Luo

Поиск и локализация объектов без эталона (по сложному описанию)

Few-Shot via Zero-Shot!



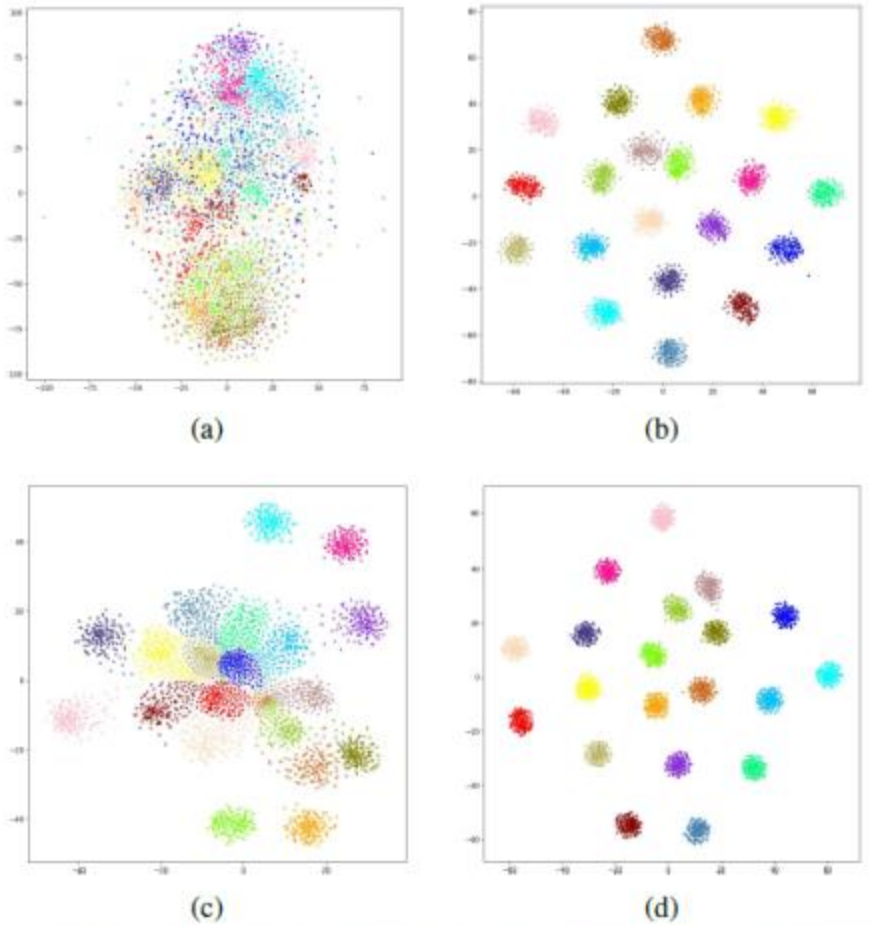


Figure 3. T-SNE visualization results for all novel categories in the Mini-imagenet set on the 1-shot and 5-shot tasks. Each scatter plot contains 20 colored classifier parameter clusters and each color represents a novel category. (a): 1-shot vision-based classifier. (b): 1-shot vision-knowledge classifier. (c): 5-shot vision-based classifier. (d): 5-shot vision-knowledge classifier.

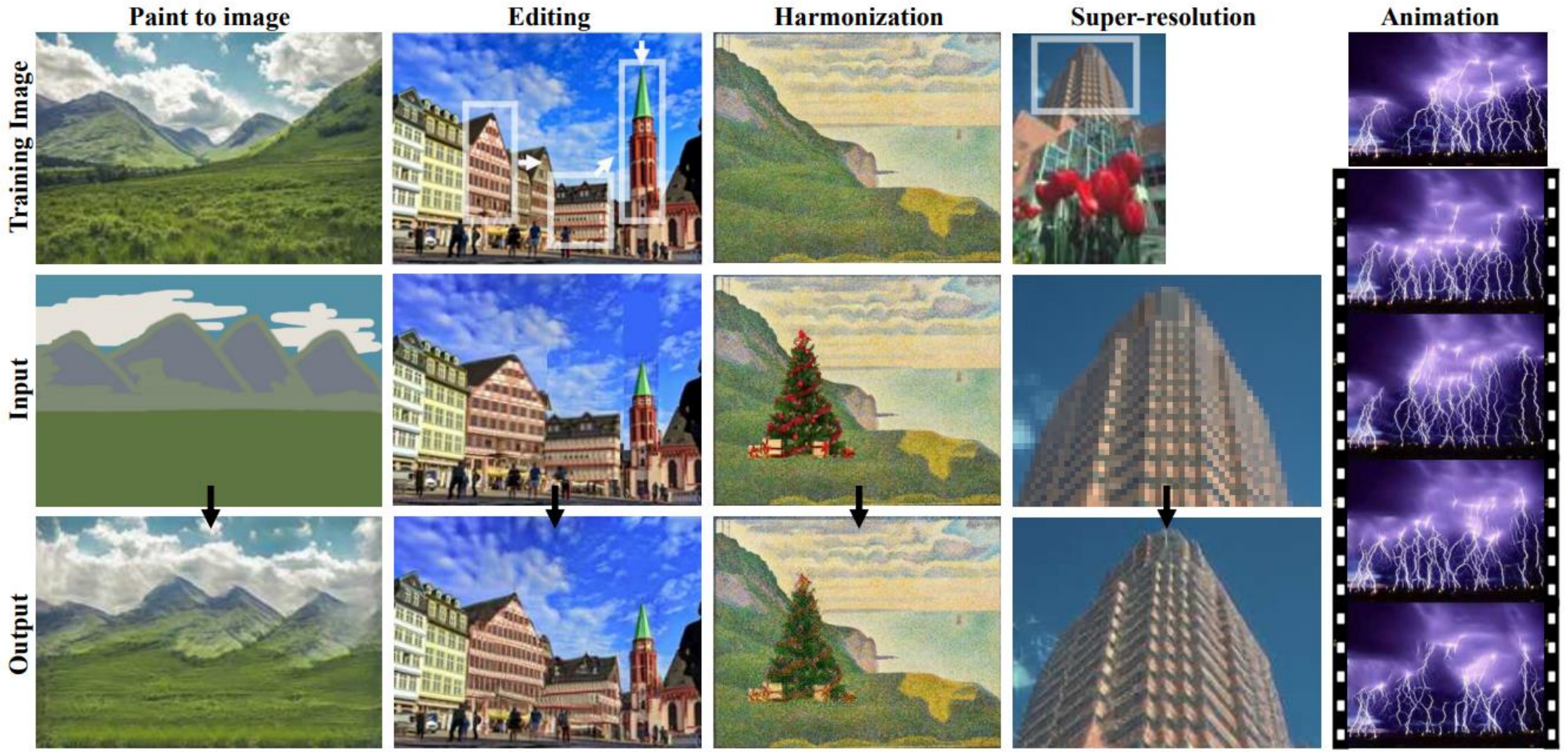
Поиск и локализация объектов ***по эталону класса путем генерализации через словесное описание****!*

Few-Shot Image Recognition With Knowledge Transfer
Zhimao Peng, Zechao Li, Junge Zhang, Yan Li, Guo-Jun Qi, Jinhui Tang

Классы намного более компактные

Генерация реалистичных данных (Domain Adaptation, Generative Adversarial Networks (GAN), Realistic Data Synthesis…)

SinGAN: Learning a Generative Model from a Single Natural Image

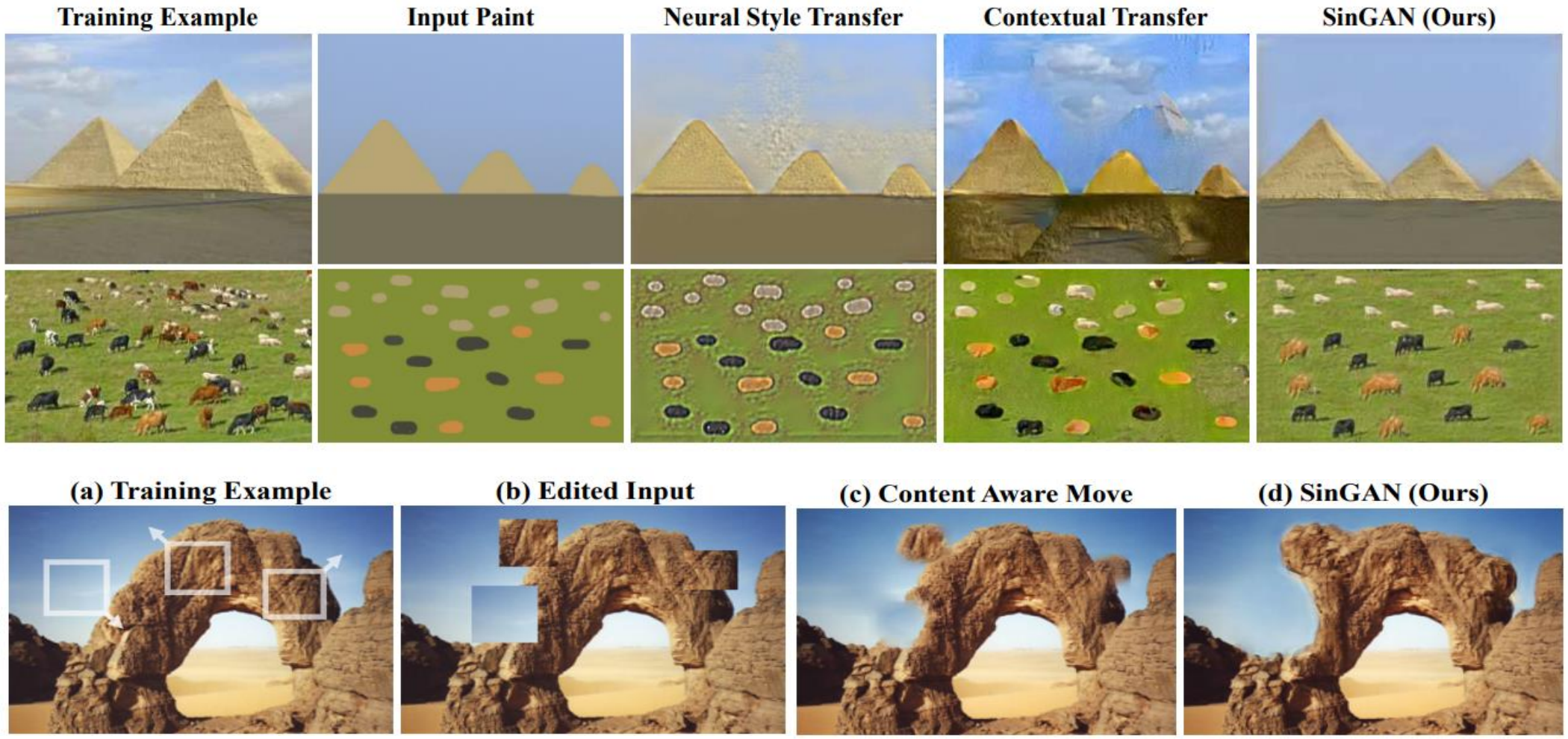


Для обучения реалистичного генератора изображений больше не нужны большие базы примеров!

SinGAN: Learning a Generative Model From a Single Natural Image, ICCV-2019,
Tamar Rott Shaham, Tali Dekel, Tomer Michaeli

ICCV 2019
Seoul, Korea

SinGAN: Learning a Generative Model from a Single Natural Image



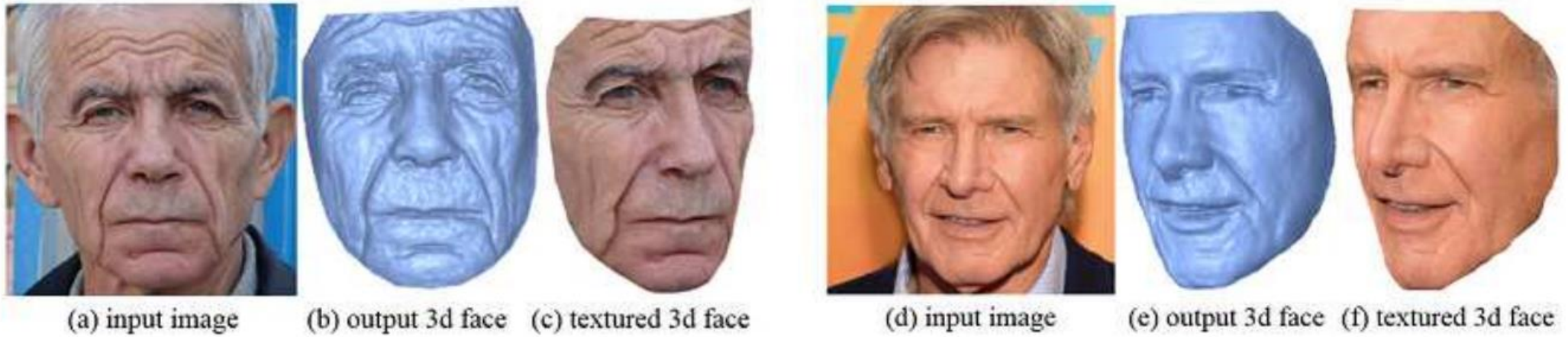
Практически неограниченные возможности реалистичного манипулирования данными…

SinGAN: Learning a Generative Model From a Single Natural Image, ICCV-2019,
Tamar Rott Shaham, Tali Dekel, Tomer Michaeli

Детальная генерация 3D лица по одному изображению

Real Faces

Swapped Faces

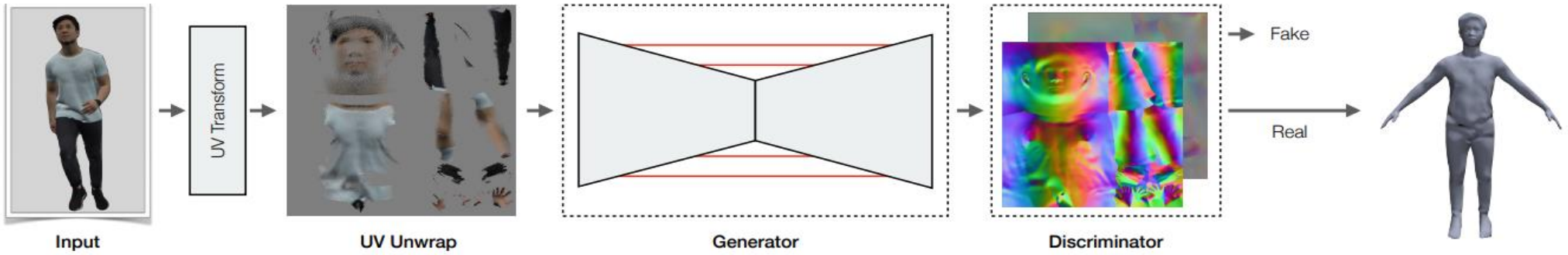


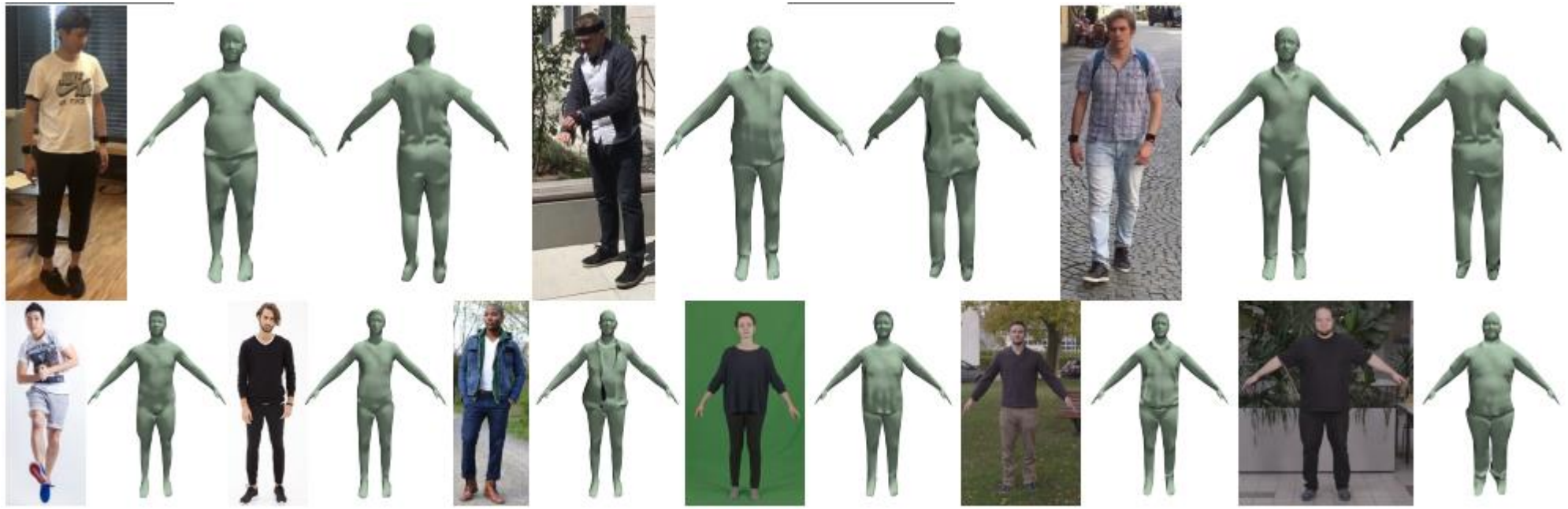
(a) input image (b) output 3d face (c) textured 3d face (d) input image (e) output 3d face (f) textured 3d face

Практически неограниченные возможности реалистичного манипулирования данными…

DF2Net: A Dense-Fine-Finer Network for Detailed 3D Face Reconstruction, ICCV-2019
Xiaoxing Zeng, Xiaojiang Peng, Yu Qiao

Детальная генерация 3D модели тела и позы

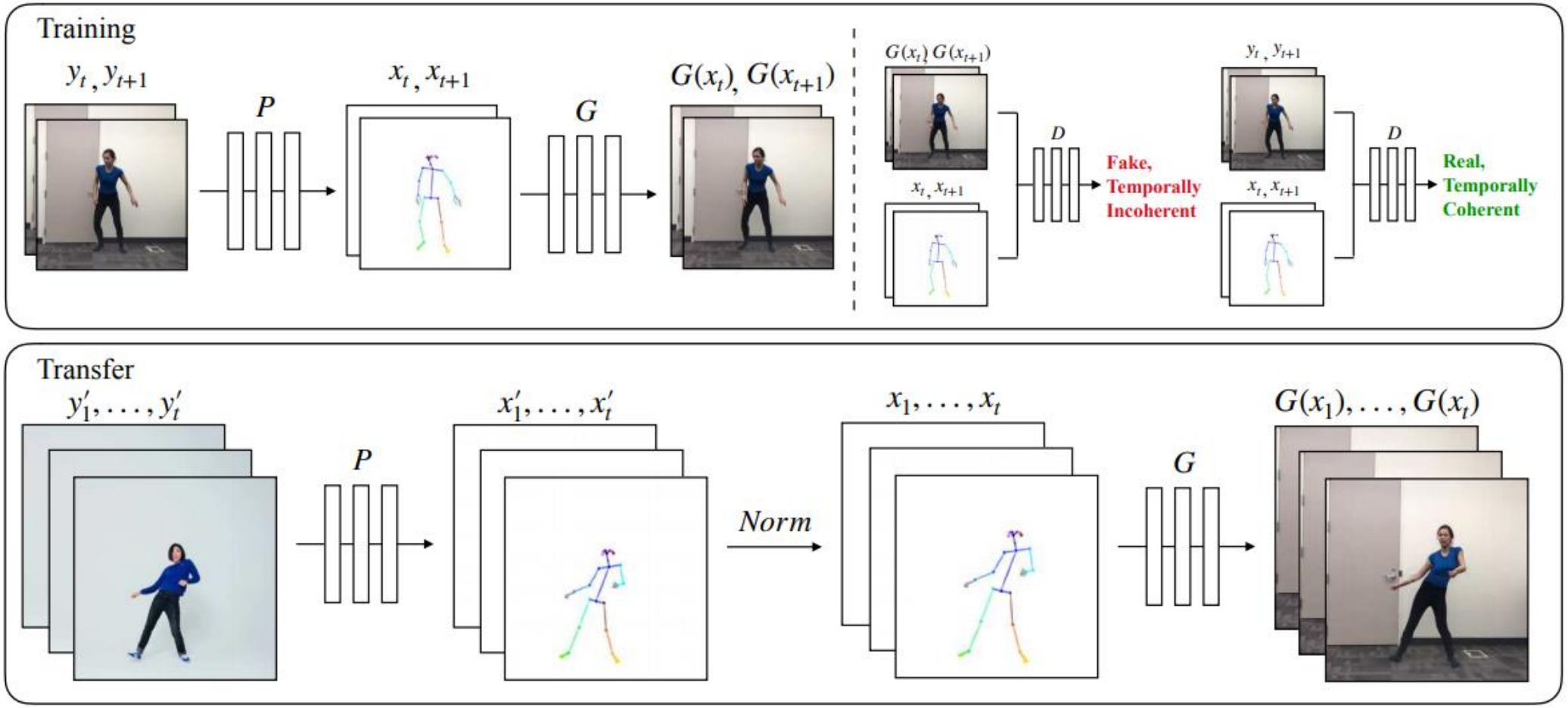




Практически неограниченные возможности реалистичного манипулирования данными…

Tex2Shape: Detailed Full Human Body Geometry From a Single Image, ICCV-2019
Thiemo Alldieck, Gerard Pons-Moll, Christian Theobalt, Marcus Magnor

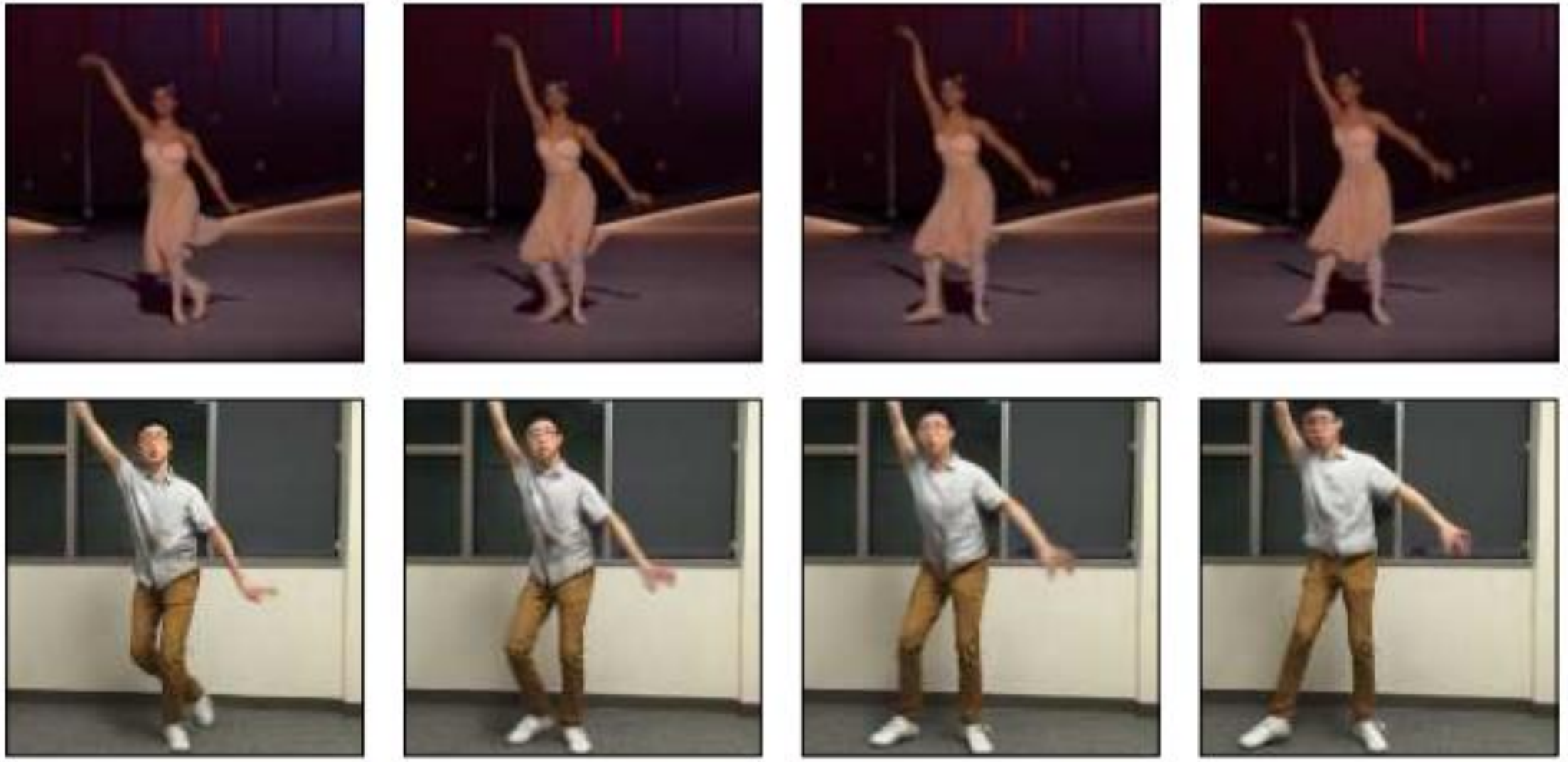
Перенос движения: EveryBody Dance Now



Практически неограниченные возможности реалистичного манипулирования данными…

Everybody Dance Now, ICCV-2109,
Caroline Chan, Shiry Ginosar, Tinghui Zhou, Alexei A. Efros

EveryBody Dance Now



Практически неограниченные возможности реалистичного манипулирования данными…

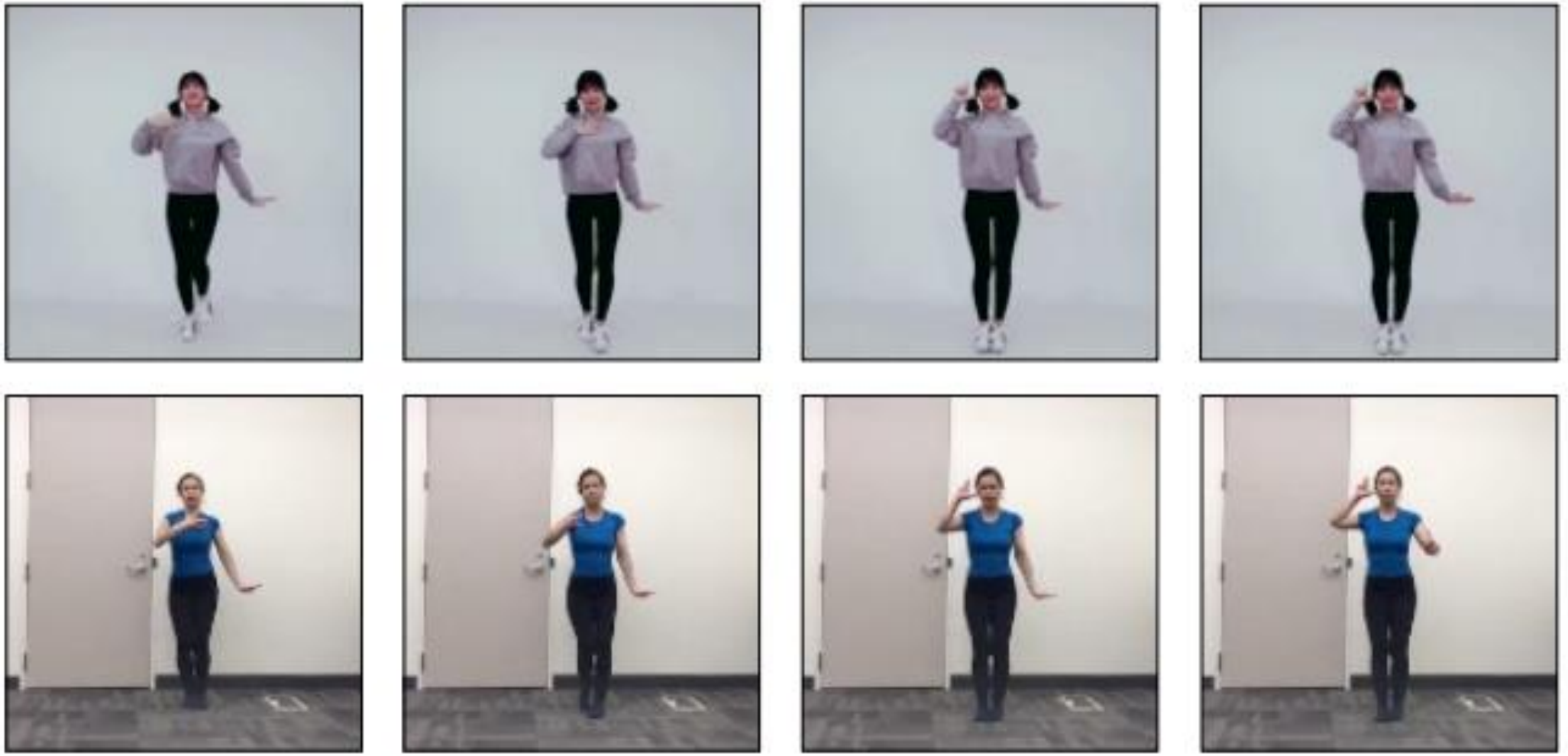
Everybody Dance Now, ICCV-2109,
Caroline Chan, Shiry Ginosar, Tinghui Zhou, Alexei A. Efros

EveryBody Dance Now

Практически неограниченные возможности реалистичного манипулирования данными…

Everybody Dance Now, ICCV-2109,
Caroline Chan, Shiry Ginosar, Tinghui Zhou, Alexei A. Efros

EveryBody Dance Now



Практически неограниченные возможности реалистичного манипулирования данными…

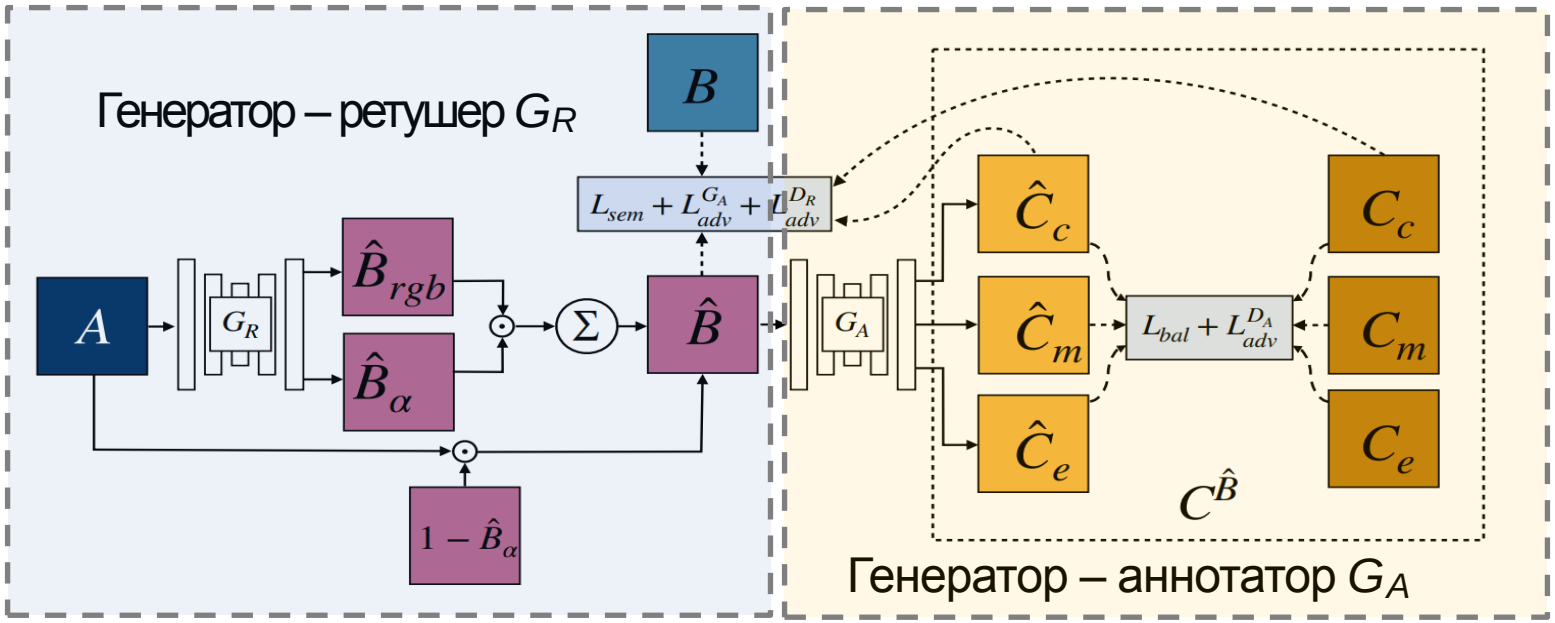
Everybody Dance Now, ICCV-2109,
Caroline Chan, Shiry Ginosar, Tinghui Zhou, Alexei A. Efros

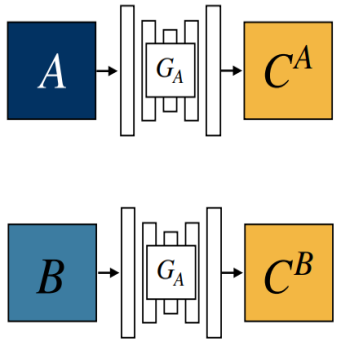
Как разоблачать фейки? Также нейросетями…

Использование смешанных генеративно–состязательных нейронных сетей для обнаружения фотомонтажей

Mixed Adversarial Generators

Генератор – ретушер G_R переводит изображения из домена монтажей A в домен реальных фото B.

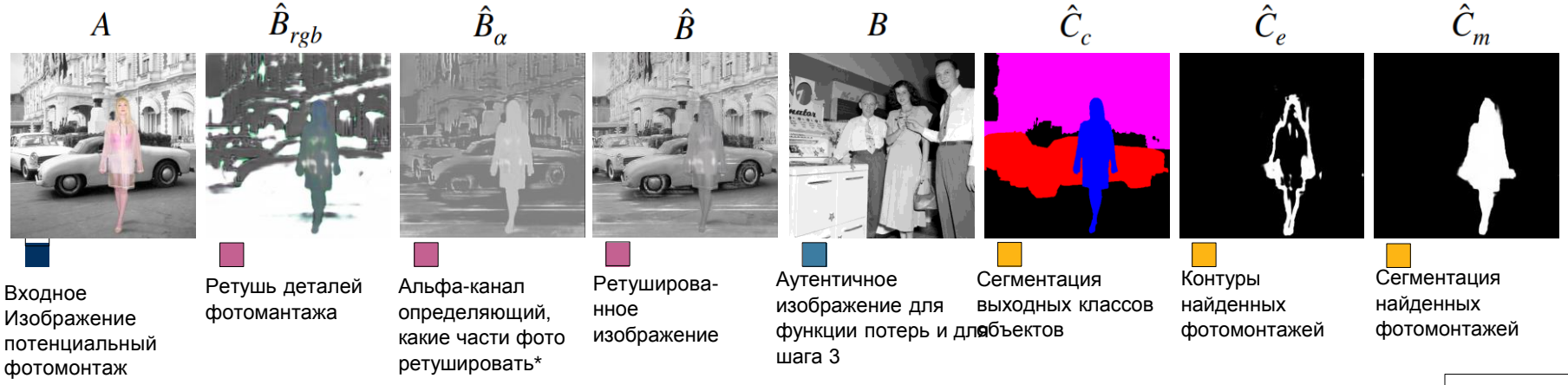




Генератор-аннотатор G_A обучается прогнозированию маски монтажа, маски краёв монтажа и карты сегментации классов объектов

Retoucher Generator G_R

Annotator Generator G_A

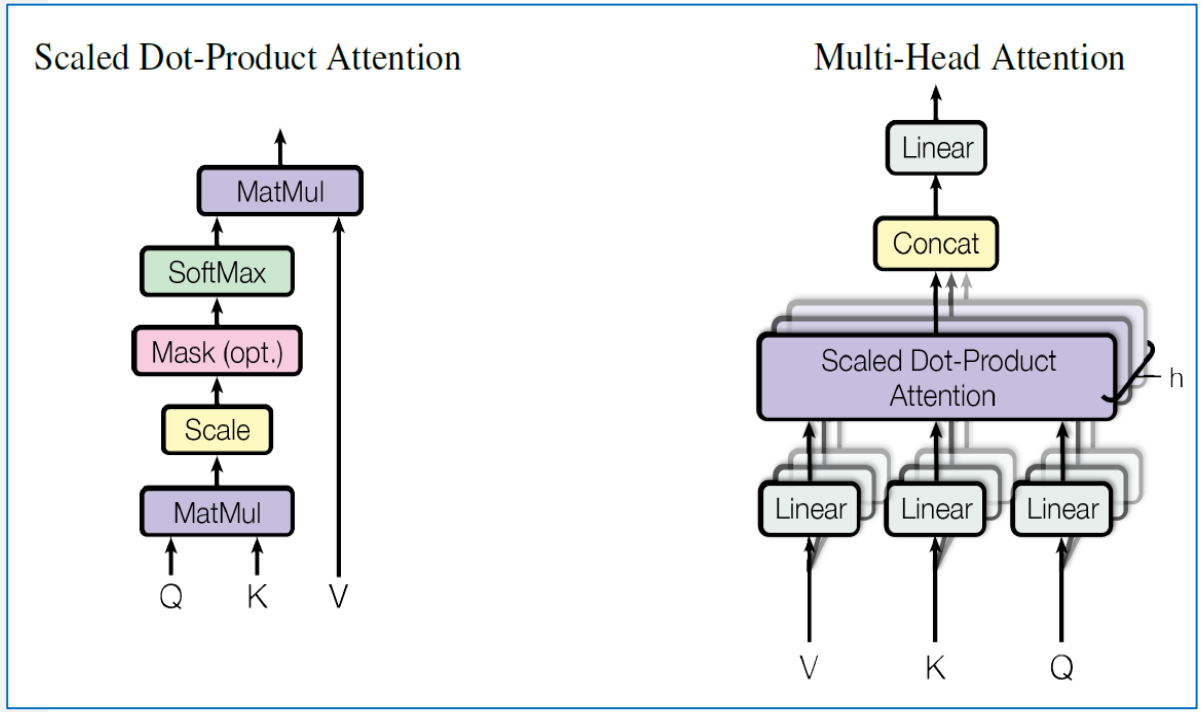


The Point Where Reality Meets Fantasy: Mixed Adversarial Generators for Image Splice Detection (NeurIPS 2019) Vladimir V. Kniaz, Vladimir A. Knyaz, Fabio Remondino

Attention, NLP: BERT, GPT-3,…

(Работа с текстами и неожиданный путь к универсальному «слабому» ИИ)

Graph Attention Networks



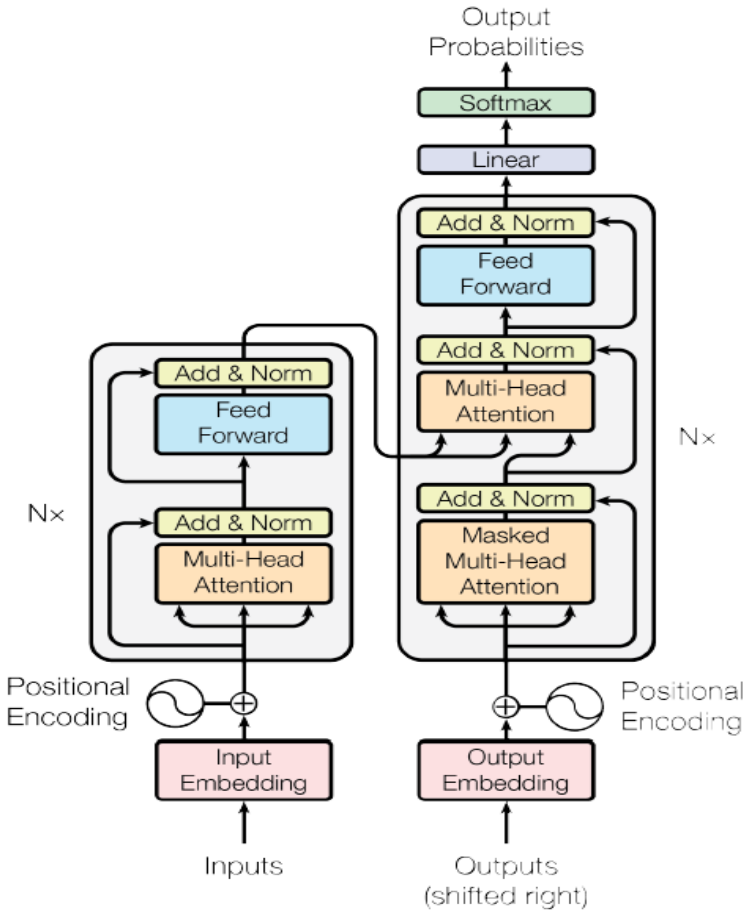


Figure 1: The Transformer - model architecture.

Управление вниманием (адаптивные парные веса) *в GCN оказались ключом к решению самых сложных задач!*

An attention function can be described as **mapping a Query and a set of Key-Value pairs to an output**, where the query, keys, values, and output are all vectors.

In the area of Natural Language Processing where **Transformer** style models have become state of the art on many tasks**. Motivation: Learning long-range dependencies is a key challenge for conv layers, which is important for symbolic sequences.**

Attention Is All You Need (Ashish Vaswani et.al., 2017)

Graph Attention Networks

$$e_{ij} = a(\mathbf{W}\vec{h}_i, \mathbf{W}\vec{h}_j) \qquad \alpha_{ij} = \mathrm{softmax}_j(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k\in\mathcal{N}_i}\exp(e_{ik})}.$$

Управление вниманием:
1. Вычисление адаптивных весов

2. Применение адаптивных весов

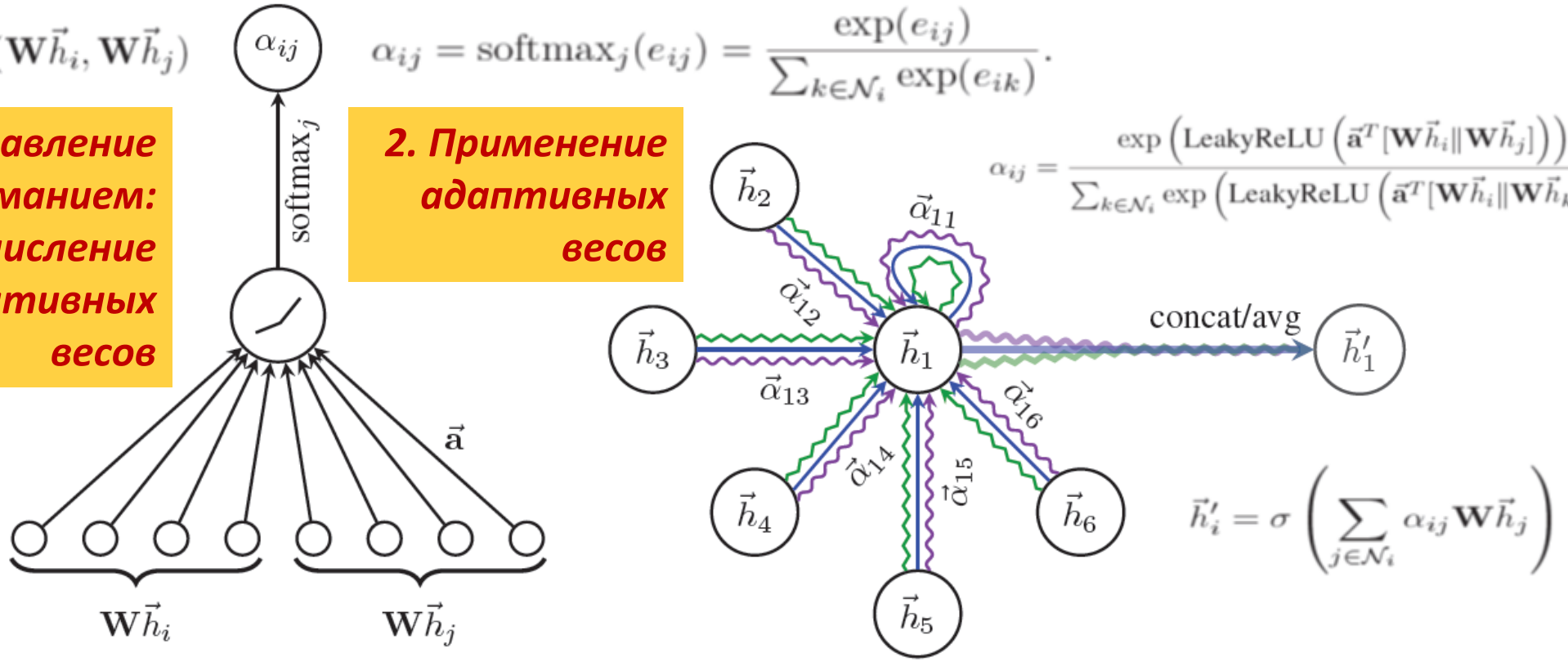


Figure 1: **Left:** The attention mechanism $a(\mathbf{W}\vec{h}_i, \mathbf{W}\vec{h}_j)$ employed by our model, parametrized by a weight vector $\vec{\mathbf{a}} \in \mathbb{R}^{2F'}$, applying a LeakyReLU activation. **Right:** An illustration of multi-head attention (with $K = 3$ heads) by node 1 on its neighborhood. Different arrow styles and colors denote independent attention computations. The aggregated features from each head are concatenated or averaged to obtain $\vec{h}'_1$.

Graph Attention Networks (Petar Velickovic et.al., 2018)

OpenAI GPT-3: “may be the biggest thing since bitcoin”

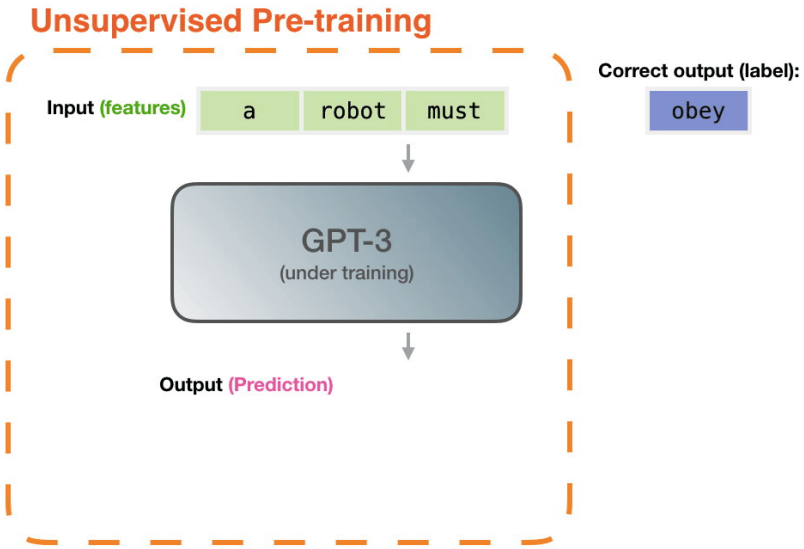
Параметры сети:

Размер сети 175 млрд. параметров(350 Гб fp16)
Контекстное окно: 2048 токена
96 трасформеров
96 self attention head
Размер батча: 3.2 млн

База данных:

The CommonCrawl data - 45 Тб
Прочие БД – 40%

Результаты сравнимые со state-of-the-art в 42 задачах NLP



Ключевая идея: обучение на огромной выборке на продолжение фразы (без учителя)

Суперкомпьютер:

10000 GPU Tesla V100
400 Гбит/сек
285 тыс. CPU ядер
Pytorch

Требует 355 лет и 5 млн. $ при аренде GPU Cloud

Архитектуры:

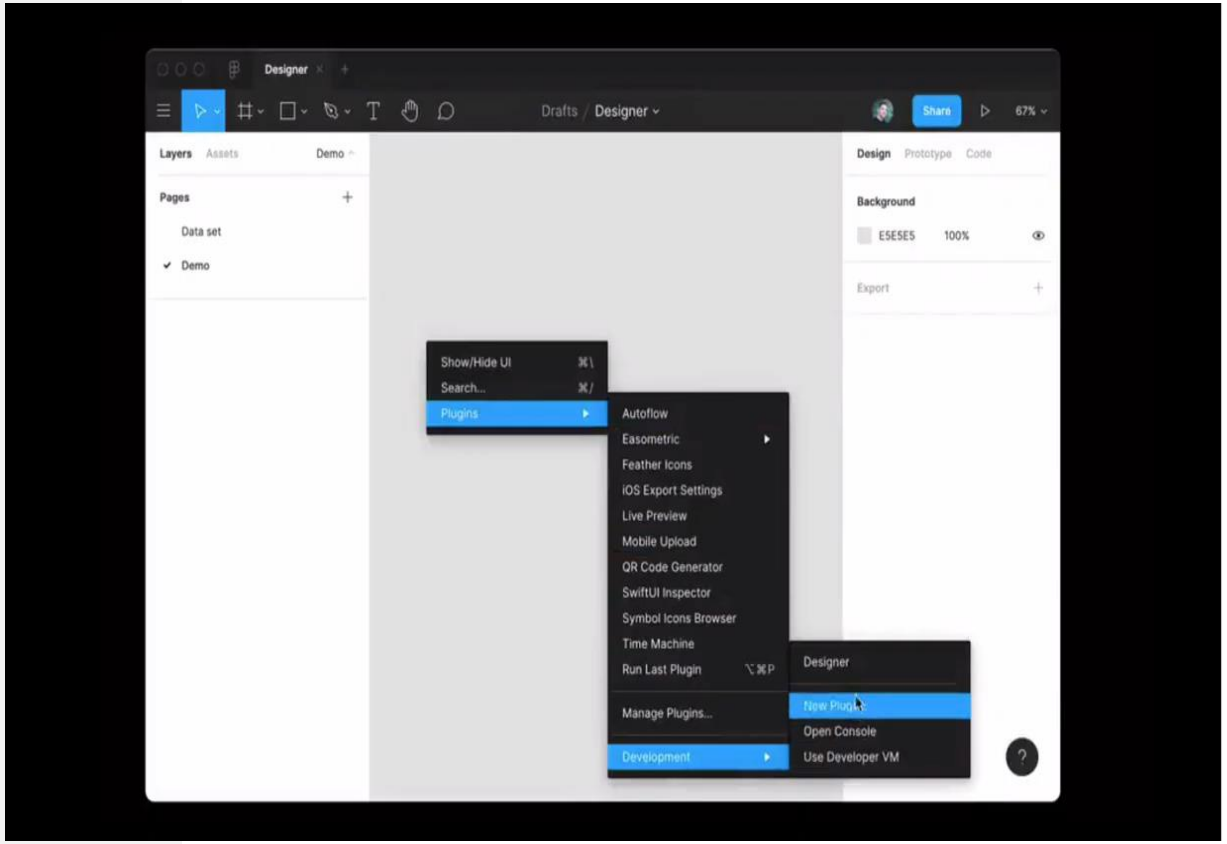
Модификация GPT-2
BERT
Transformer

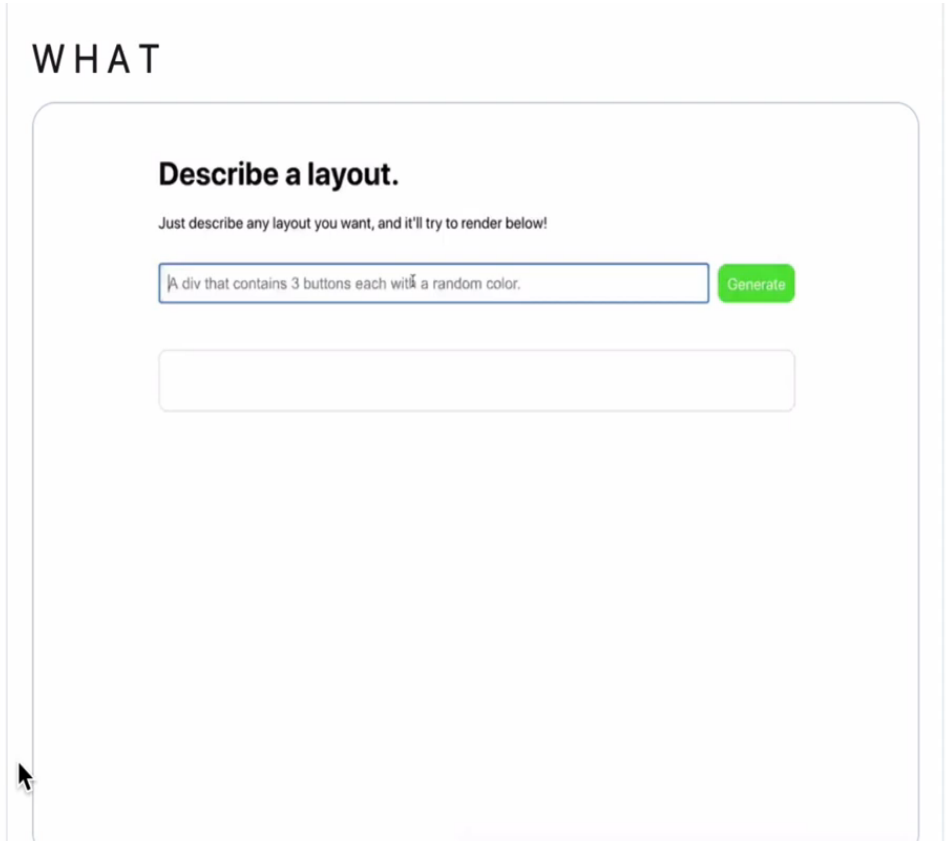
- GPT-3 может отвечать на вопросы на естественном языке
- GPT-3 умеет переводить с любого языка, отвечать на вопросы, писать статьи и т.д. вплоть до написания исходного кода программ

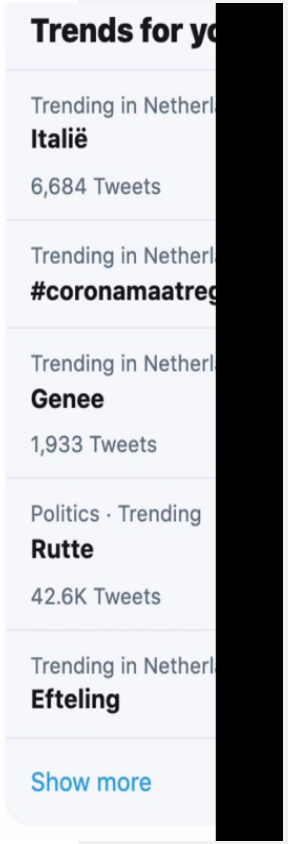
ХИТ ИИ 2020!

OpenAI GPT-3: “may be the biggest thing since bitcoin”

Примеры приложений:

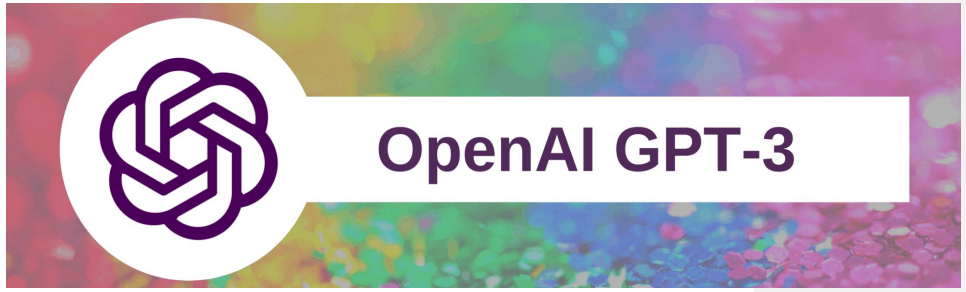






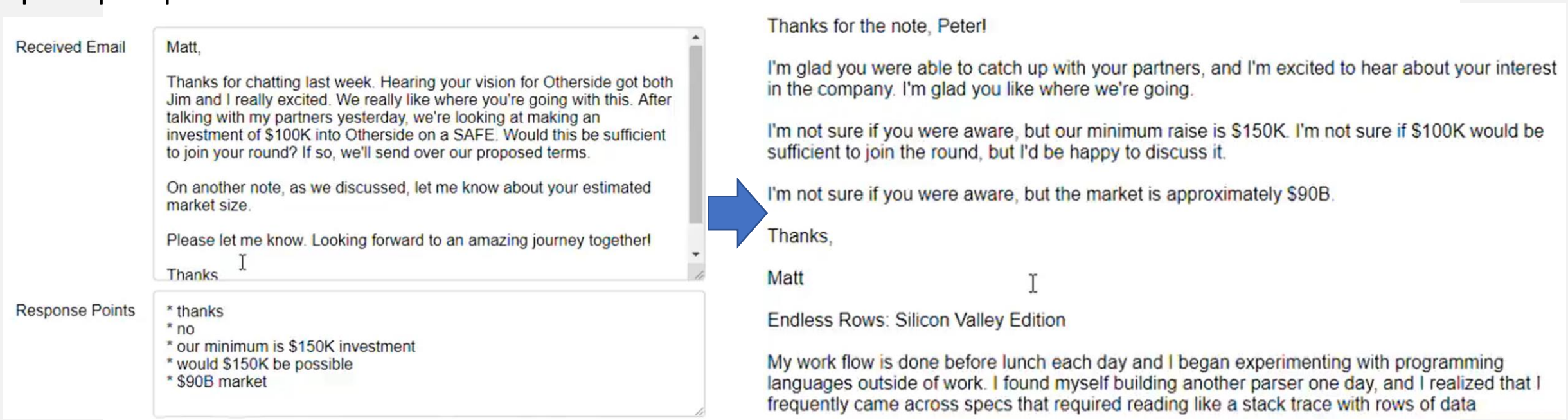
Генерация визуальных форм(кода) по текстовому описанию

https://machinelearningtokyo.com/2020/07/26/10-cool-gpt-3-demos/



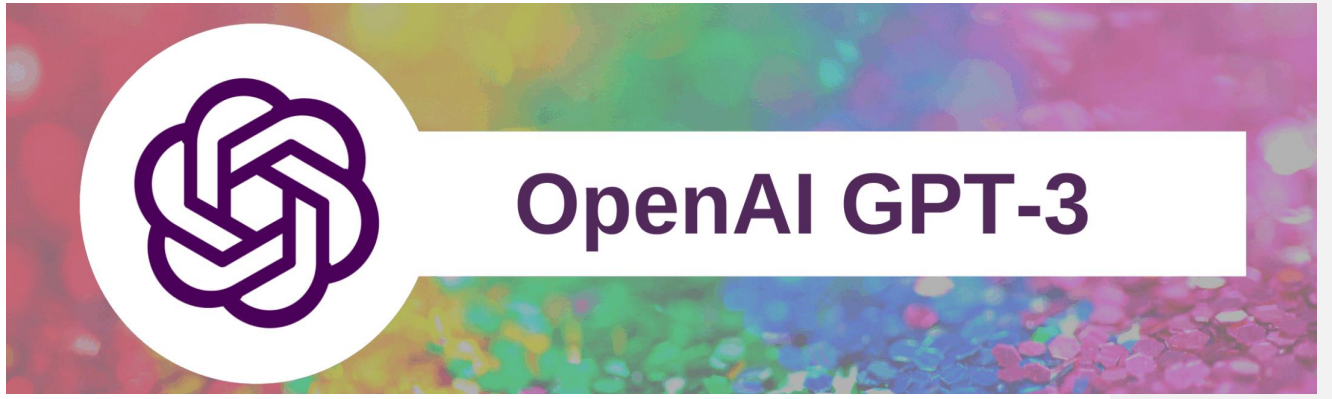
OpenAI GPT-3: “may be the biggest thing since bitcoin”

Примеры приложений:



Автоматическое написание писем по нескольким тезисам

- Общение с ИИ на естественном языке
- Универсальная модель
- Команды для РТК/ИИ
- Построение моделей по уставу/правилам



Открытые проблемы: угрозы, вызовы, надежды (2020)

Проблемы:

- Как справиться с атаками?

- Как эффективно переносить обучение в реальном мире?

- Как разоблачать фейковые данные?

- Реальных данных для практических приложений катастрофически не хватает!

- Перспективные методы обучения требуют слишком больших ресурсов!

- Мостик между зрением и языком/пониманием давно перекинут, но массовый переход пока не случился

Надежды:

- Новые датчики технического зрения

- Прогресс в AutoML/NAS

- Прогресс в Few-Shot/Zero-Shot

- GCN, Attention => GPT-3

- Прогресс в объяснении нейросетевых рассуждений

- Массовый перевод нейростевых рассуждений с уровня отдельных объектов на уровень семантических конструкций (онтологий)

- Прогресс в методах RL, глубокой оптимизации и глубокого управления

- Соединение универсальности GPT-3 с анализом данных и глубокой оптимизацией

ВЫВОДЫ ПО СОСТОЯНИЮ ТЕХНОЛОГИЙ ИИ

1. **Первая волна** современной революции в ИИ породила технологии «**глубокого распознавания**», обеспечивающие решение задач **компьютерного зрения, анализа сигналов и больших данных**.

2. В настоящее время технологическая революция в ИИ переживает **вторую волну**, которая ведет нас прямо к созданию **функционального ИИ**.

3. **Нет никакого специфического прорыва в методах «искусственного интеллекта»**, но **наблюдается технологический прорыв, связанный с ГНС**.

4. **Методы на основе глубоких нейронных сетей (ГНС) не только про СТЗ.** **«Глубокие» технологии развиваются**, фокус их применения смещается **от обработки и анализа данных к семантике, управлению и оптимизации**

5. **Революция в искусственном интеллекте продолжается! (см. GPT-3)**
Скорость и направления дальнейшего продвижения будут зависеть от того, **как и когда удастся справиться с возникшими проблемами**, а также от того, **когда и какие сбудутся надежды**.

Вторая часть доклада

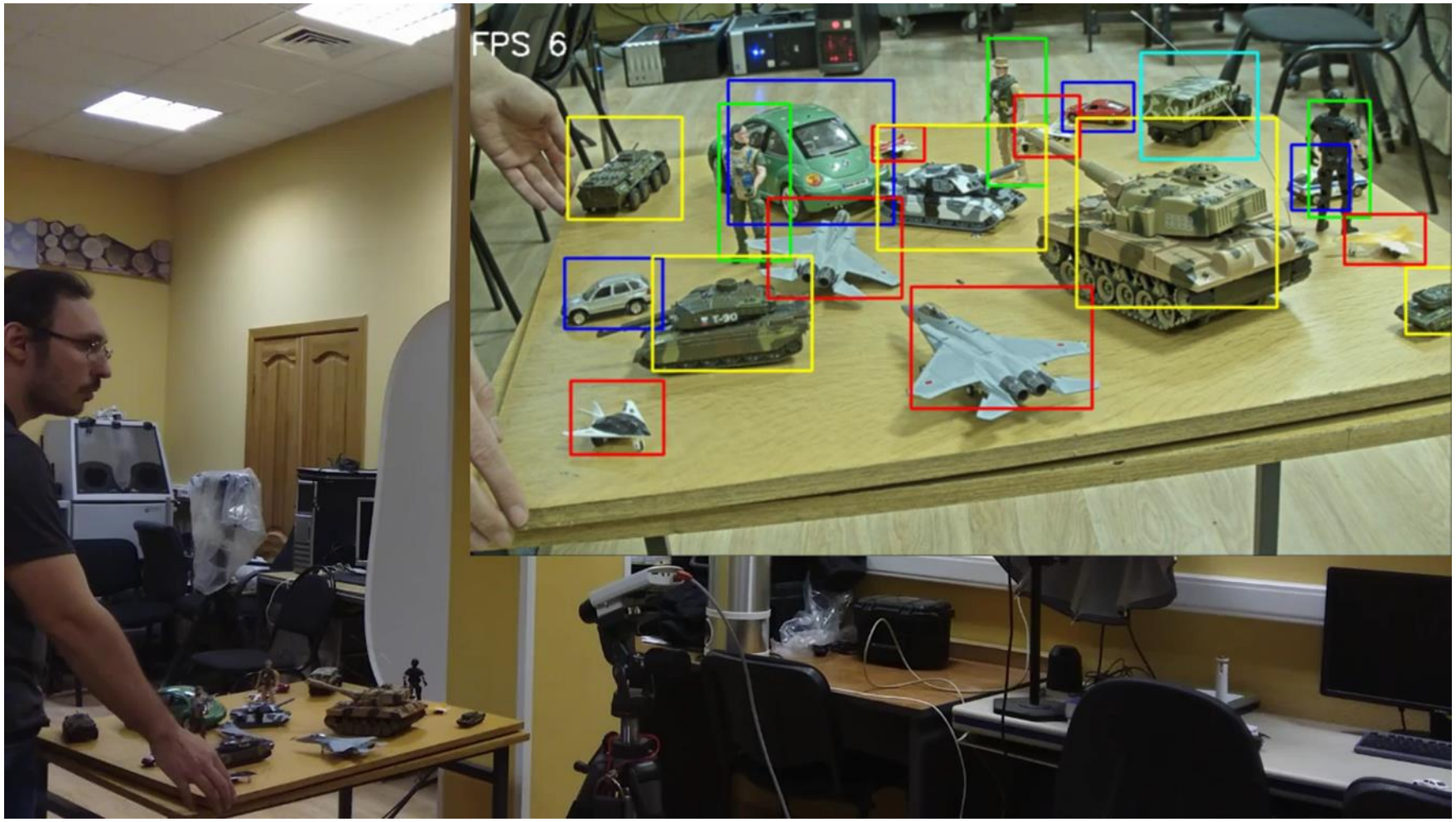
Унифицированная программная платформа для разработки конечно ориентированных программных комплексов на основе нейросетевых подходов

(ПЛАТФОРМА-ГНС, Экосистема Plat)

+ Начальник лаборатории 3050 ФГУП “ГосНИИАС”
Горбацевич Владимир Сергеевич, *gvs@gosniias.ru*

Готовность отечественных технологий

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



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

Где отечественные процессоры?

Где отечественное ПО для обучения?

Готовность отечественных технологий

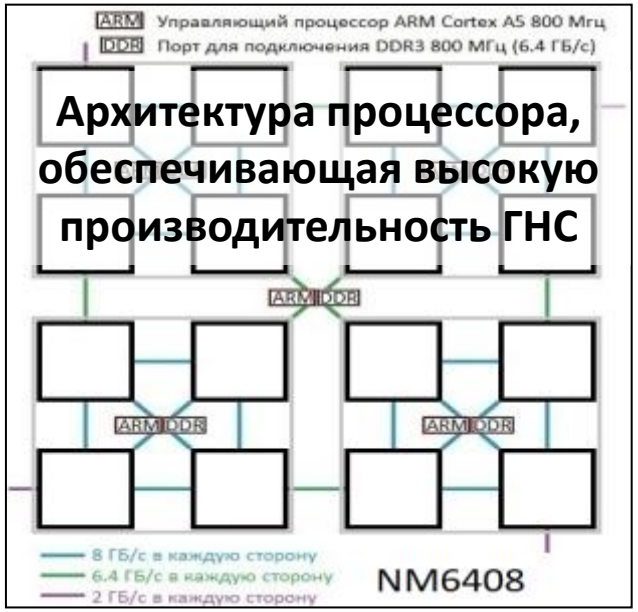
2018-19: ГНС на отечественных процессорах

Стенд ГосНИИАС на выставке «АРМИЯ-2018»

2018: 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 (5 кадров/сек).

2020: создано бортовое решение с ГНС на базе NM6408, работающего в 32 раза быстрее (*скорость обработки - 60 кадров/сек и выше*).

Перспектива-2020+: ожидается выпуск еще трех нейропроцессоров отечественных производителей



2019+: отечественные процессоры уже есть

Готовность отечественных технологий

ПО обучения глубоких нейросетей: проблемы и решение

Задача: бортовая реализация ГНС и использование глубокого обучения в отечественных АПК для критических приложений.

Проблемы:

- **Существуют только зарубежные средства обучения ГНС** (системы Caffe, Caffe2, Pytorch, TensorFlow, Theano и др.) => **техническая зависимость, невозможность сертификации ПО на основе ГНС на НДВ**;
- Отсутствие поддержки и учета особенностей отечественных датчиков и бортовых вычислителей;
- Несовместимость различных средств и систем разработки на базе ГНС…

Решение: создание отечественной унифицированной программной Платформы для разработки конечно ориентированных программных комплексов, которая должна стать основным средством автоматизированного проектирования (формирования и обучения) алгоритмов и бортового нейросетевого ПО всех будущих отечественных АПК.

Нужен не просто отечественный аналог фреймворка, а единая интегрированная среда разработки ГНС.
Почему?

Годится ли нам традиционный путь R&D?

На нашем пути есть объективные барьеры, которые нужно учитывать:
- Нас мало!
- Мы разрознены!
- У нас меньше денег и других (в т.ч. вычислительных) ресурсов!...

Пирамида человеческих ресурсов

ICCV-2019: 300+ статей от США и Китая / менее 10 из России

Прогноз: **В области разработки и внедрения ИИ в конечные изделия наши разрозненные и малочисленные разработчики всегда будут выступать как "малый бизнес", который из-за эффекта масштаба проигрывает крупным иностранным разработчикам.**

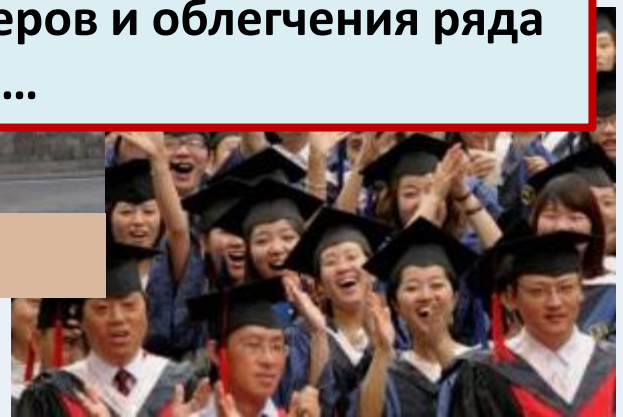
Возможное решение: **создать специальную интегрированную среду поддержки разработок в области ИИ (ГНС), которая позволила бы ускорить внедрение ГНС в России путем преодоления ряда барьеров и облегчения ряда переходов...**

ессор /

анты, П

Студенты, сотрудники

Университет или институт



Платформа для ускорения переходов

Платформа для ускорения переходов

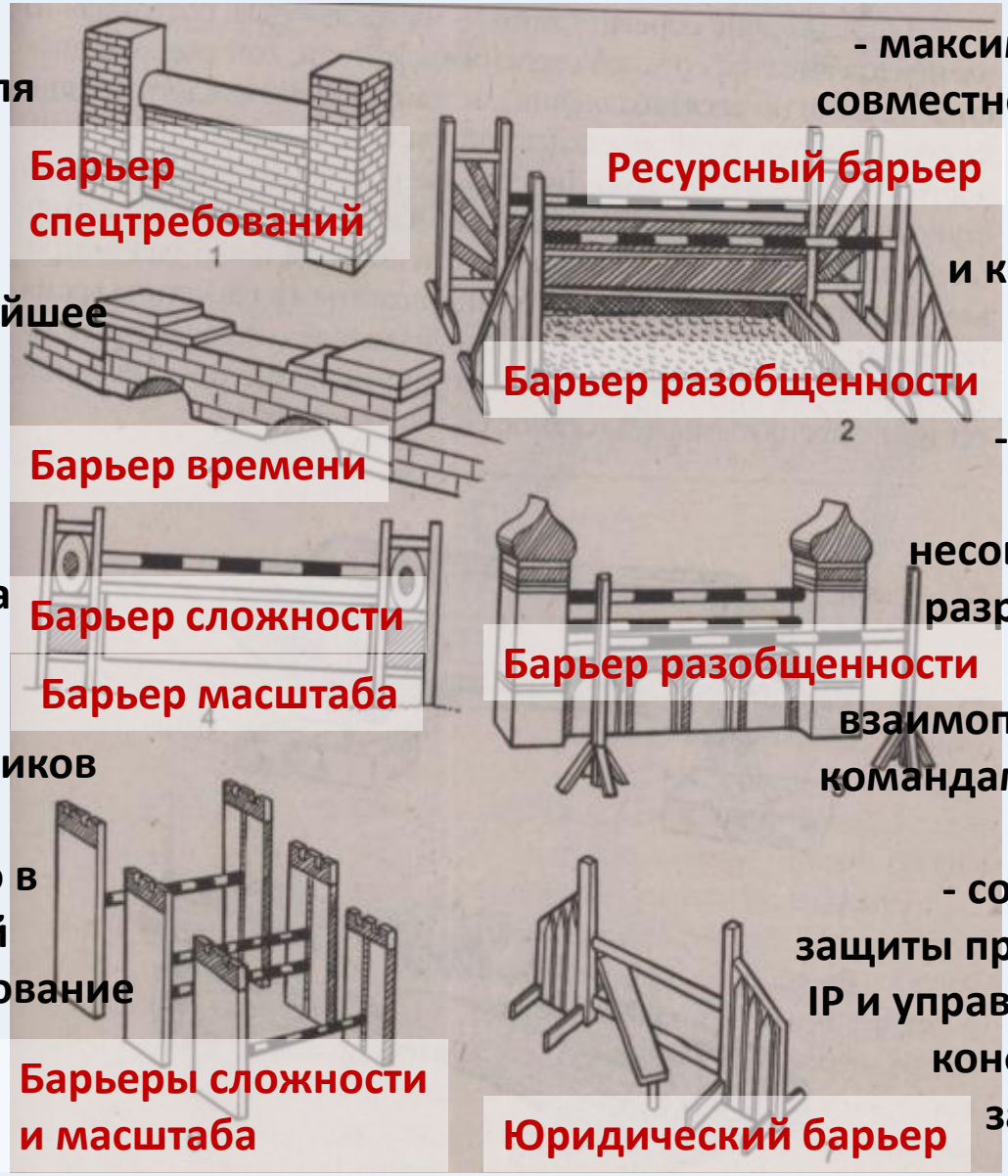
Платформа для преодоления барьеров

- снять ограничения для применения ГНС в области ОПК

- обеспечить наискорейшее внедрение самых передовых ГНС в конечные изделия

- понизить порог входа (потому что нас мало) и резко увеличить сообщество разработчиков

- упростить навигацию в пространстве решений и совместное использование знаний и данных



- максимально облегчить совместное использование ресурсов СКТ, проектную и командную работу

- минимизировать проблемы из-за несовместимости сред разработки, повысить унификацию и взаимопонимание между командами разработчиков

- создать механизмы защиты прав собственников IP и управления доступом к конфиденциальным/ закрытым данным

Готовность отечественных технологий

Проект «Платформа-ГНС» (ГосНИИАС, 2018-20)

2020+: *задача распознавания образов не только алгоритмически решена, но доведена до стадии технологической готовности к ОКР*

Унифицированная платформа

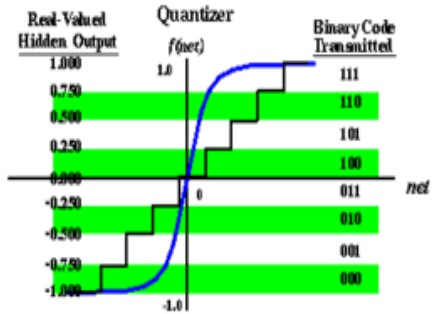
- **Полностью сертифицированный на НДВ исходный код**
- Возможность использования библиотек NVIDIA и т.д.
- Единая экосистема
- Унифицированный формат хранения БД и моделей
- Импорт/экспорт из основных фреймворков и onnx
- **Наличие типовых решений для основных задач**
- Поддержка отечественных аппаратных платформ
- Поддержка зарубежных аппаратных платформ
- Поддержка отечественных ОС
- Контроль доступа к данным и проектам
- Множество возможностей для командной разработки
- **Низкие требования к квалификации ИТР**

Снижение порога входа => резкое увеличение числа разработчиков => мы надеемся на прорыв в массовом внедрении ГНС в ОПК

Единая интегрированная среда

Обучение ГНС

Преобразование ГНС



Аппаратно-ориентированная реализация ГНС

OpenVINO™

TensorRT

2020: создано отечественное ПО

Готовность отечественных технологий

Отечественный фреймворк PlatLib…

Обучение:

- Динамические графы
- Распределенное обучение
- Поддержка GPU AMD/NVIDIA
- Поддержка CPU x86-64/Elbrus
- Поддержка БД Платформы

```
class LeNetPlat(plat.module.Module):
    def __init__(self):
        super(LeNetPlat, self).__init__()
        self.conv1 = conv2D(20, 1, 5)
        self.relu1 = relu()
        self.pool1 = poolmax2d()
        self.conv2 = conv2D(50, 20, 5)
        self.relu2 = relu()
        self.pool2 = poolmax2d()
        self.fc1 = conv2D(500, 50, 4)
        self.relu3 = relu()
        self.fc2 = conv2D(10, 500, 1)
```

```
res_plt = LeNetPlat(inputT)
plt_res = SM(res_plt, labelsT)
plt_res.backward()
optimizer_plat.iter()
```

Обучение:

- Аналогично pytorch
- Отладка на python
- Нет multiGPU
- Распределенное обучение через nccl/rccl/mpi

+ интегрированная среда… + целая экосистема Plat…

(полный стек ипортонезависимых технологий)

Характеристики Платформы

Клиентская часть

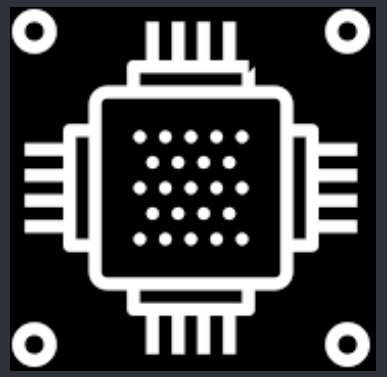
- CentOS/AltLinux
- Аппаратная платформа Intel x86 + Nvidia
- Аппаратная платформа Intel x86 + AMD
- Аппаратная платформа Эльбрус
- Аппаратная платформа Эльбрус + AMD
- Сертификация на НДВ

- ОС Ubuntu/CentOS/AltLinux / AstraLinux
- Аппаратная платформа Intel x86
- Аппаратная платформа Эльбрус
- Кросс-платформенный код
- Сертификация на НДВ

Серверная часть

Вычислительный узел

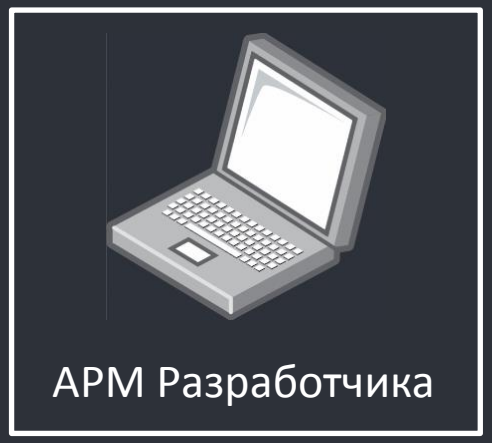
SDK Платформы



- Аппаратная платформа Intel x86 + Nvidia
- Аппаратная платформа Intel x86 + AMD
- Аппаратная платформа ARM (Байкал-М)
- Аппаратная платформа(ы) Эльбрус
- Аппаратная платформа Эльбрус + AMD
- Аппаратная платформа(ы) НТЦ Модуль
- Аппаратная платформа(ы) НПО Элвис
- Аппаратная платформа ПЛИС UltraScale+ (Xilinx)
- Аппаратная платформа(ы) IVA technologies

Схема работы системы

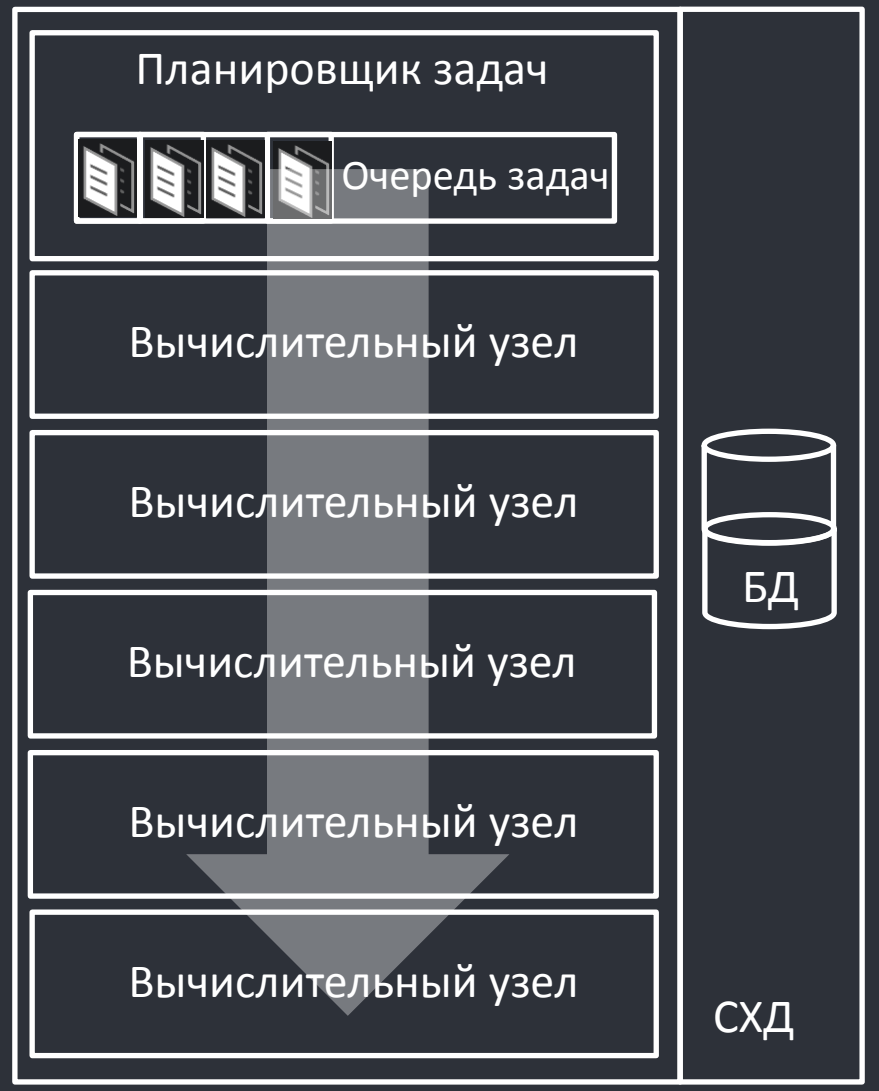
Клиентская часть



Задание

Результаты

Серверная часть



Особенности:

- ✓ Управление вычислительными ресурсами
- ✓ Автоматический выбор узлов
- ✓ Распределенное обучение
- ✓ Доступ к проектам и файлам
- ✓ Совместная работа над проектами
- ✓ Резервное копирование
- ✓ Логирование процесса разработки
- ✓ Удаленная работа с БД

Клиентская часть Платформы

АРМ Разработчика

Особенности:

- ✓ Поддержка ОС AltLinux/ ОС AstraLinux
- ✓ Поддержка архитектуры Эльбрус/x86
- ✓ Кроссплатформенная версия

АРМ Администратора

Особенности:

- ✓ Поддержка ОС AltLinux/ ОС AstraLinux
- ✓ Поддержка архитектуры Эльбрус/x86
- ✓ Кроссплатформенная версия(в разработке)
- ✓ Web-интерфейс

Проект: основные понятия

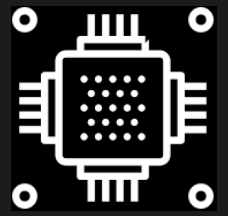
АРМ Разработчика

АРМ Администратора

Проект –включает архитектуру ГСНС, базу аннотированных данных для обучения, а также алгоритмы и настройки параметров для обучения

БД

Алгоритм



Апп. реализация

Задание

Попытка #1

Архитектура
Параметры
Метод
Ноды
Выборки
Аугментация

Задание

Попытка #2

Архитектура
Параметры
Метод
Ноды
Выборки
Аугментация

Задание

Попытка #3

Архитектура
Параметры
Метод
Ноды
Выборки
Аугментация

Задание

Попытка #4

Архитектура
Параметры
Метод
Ноды
Выборки
Аугментация

Проект: уровни разработки

АРМ Разработчика

АРМ Администратора

Три уровня разработки:

Июнь 2020

Начальный уровень (типовые решения):

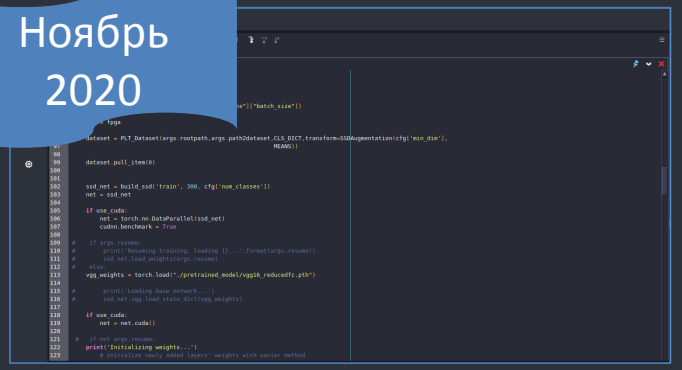
+ готовые state of the art решения
+ все промежуточные операции
+ набор типовых практических задач
+ распределенное обучение
+ оценка производительности

Ноябрь 2020

Средний уровень:

+ визуальное проектирование архитектуры
+ добавление пользовательских слоев
+ удобное и наглядное представление ГКНС

Ноябрь 2020



Продвинутый уровень:

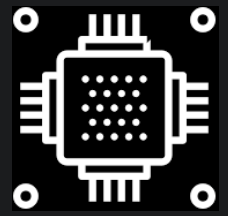
+ написание кода на языке python
+ гибкая разработка в единой среде
+ использование сторонних библиотек
+ возможность написание типовых решений
+ средства визуализации

Типовые решения

Типовое решение - включает архитектуру ГСНС, базу аннотированных данных для обучения, а также алгоритмы и настройки параметров для обучения

 **БД**

 **Алгоритм**

 **Апп. реализация**

 **Защита от изменений**

Список типовых задач:

- обнаружение объектов по изображениям и видеопоследовательностям;
- обнаружение объектов по многоспектральным данным;
- дешифрирование;
- классификация;
- семантическая сегментация;
- распознавание типов объекта;
- сопровождение объектов на видеопоследовательностях;
- обработка и комплексирование изображений различных диапазонов;
- устранение шумов и помех на изображениях (денойсинг);
- устранение смаза и расфокусировки изображений (деблур).

Особенности:

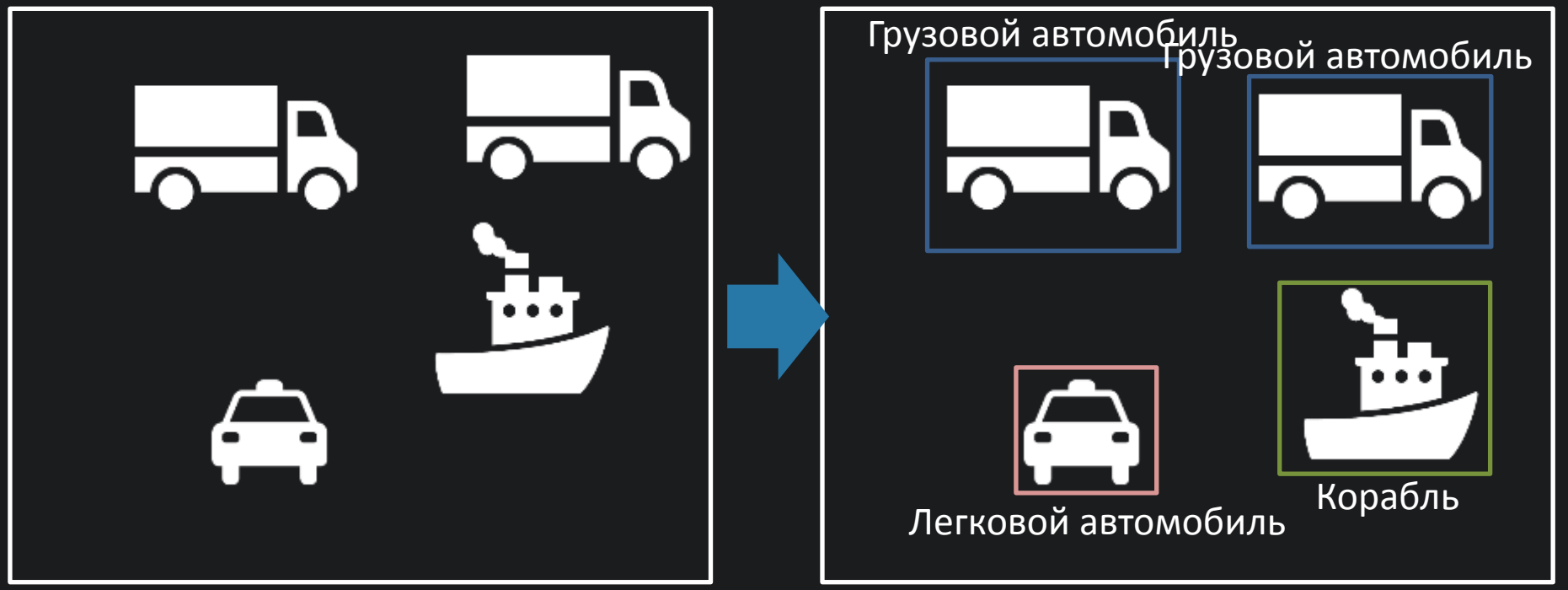
- ***Гарантия работы***
- ***Поддержка***
- ***State of the art***
- ***Аппаратная реализация***

Защита кода:

Целостность кода проверяется при старте платформы

Типовые задачи

Обнаружение объектов по изображениям и видеопоследовательностям:



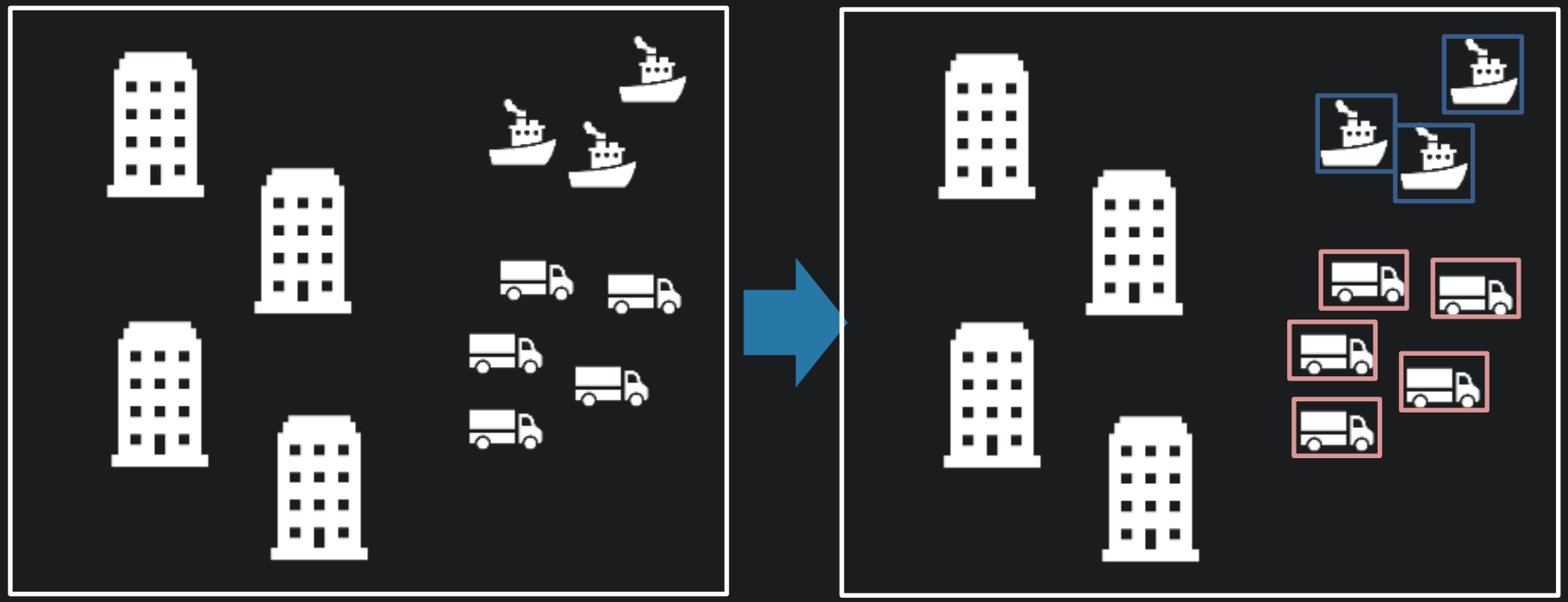
Входное изображение

Результат обнаружения

Выделение описывающих прямоугольников в экранных координатах для заданного набора классов

Типовые задачи

Дешифрирование :



Входное изображение высокого разрешения

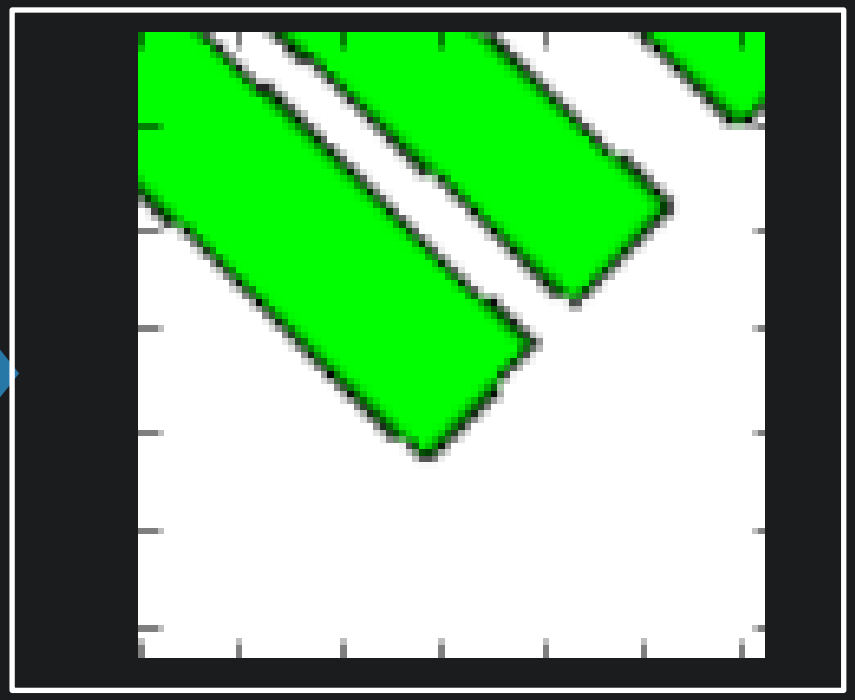
Результат обнаружения

Выделение описывающих прямоугольников в экранных координатах для заданного набора классов

Типовые задачи

Семантическая сегментация :

Входное изображение



Результат сегментации

Выделение точных границ объектов целевых классов

Типовые задачи

Распознавание объектов :

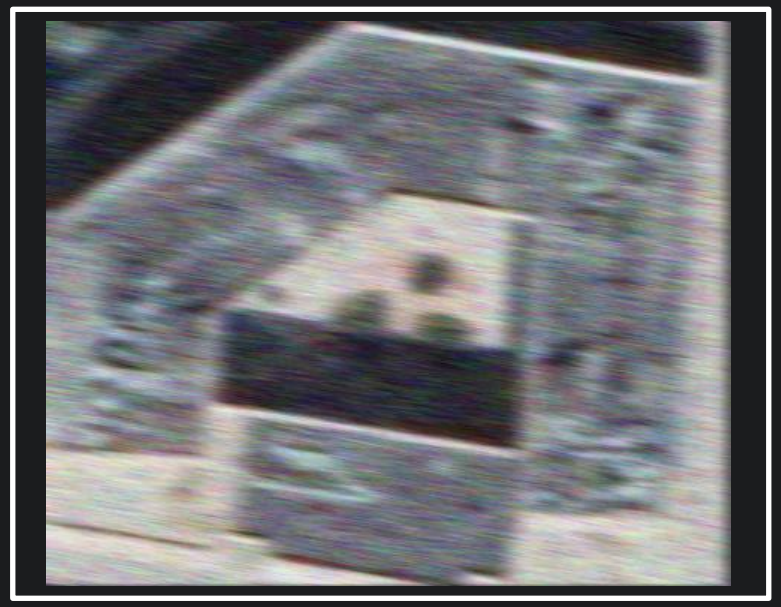
Входное изображение

Класс грузовой автомобиль

Распознавание класса(типа) объекта на изображении

Типовые задачи

Обработка и комплексирование изображений различных диапазонов,
Устранение шумов и помех на изображениях
Устранение смаза и расфокусировки

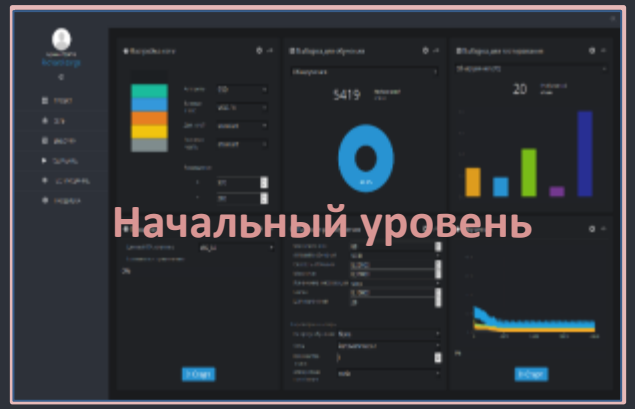


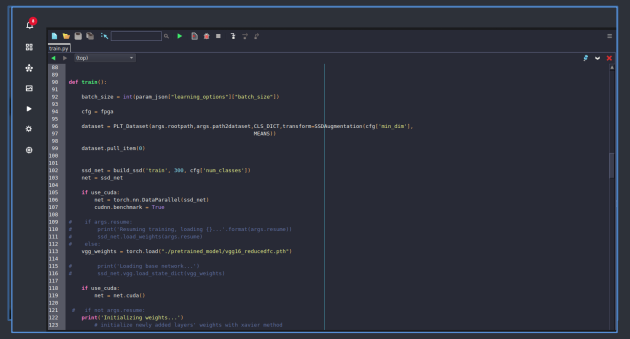
Входное изображение
(один или несколько диапазонов)

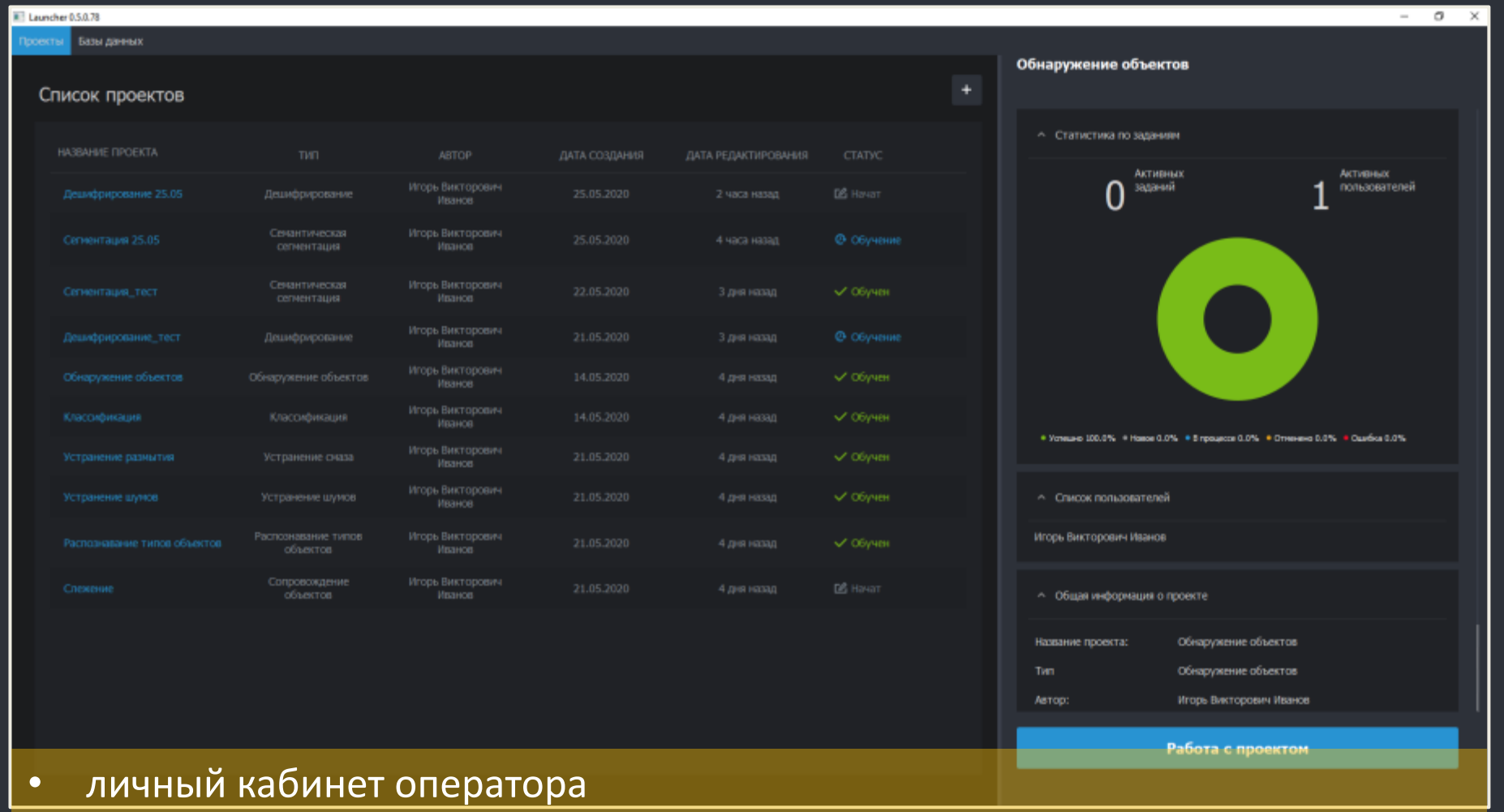
Восстановление максимально информативного изображения

Проект: СПО АРМ-Р: Загрузчик

АРМ Разработчика



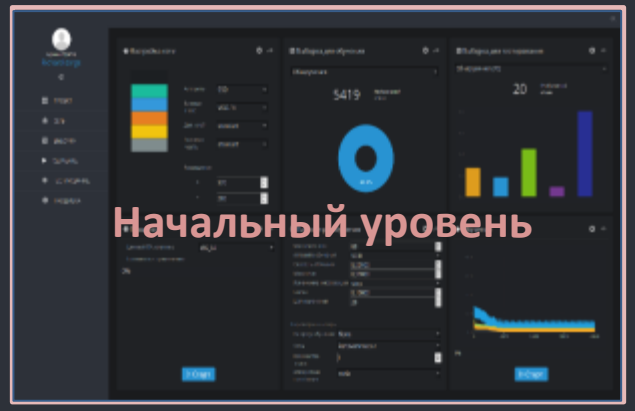


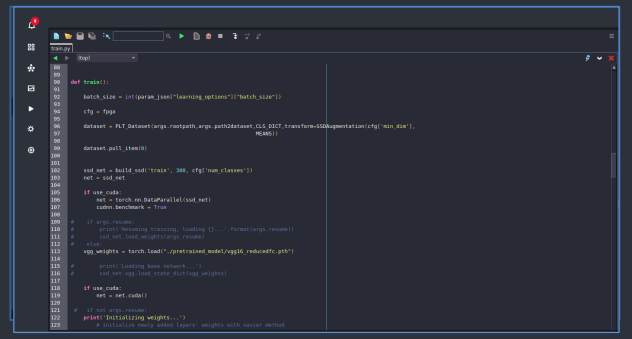


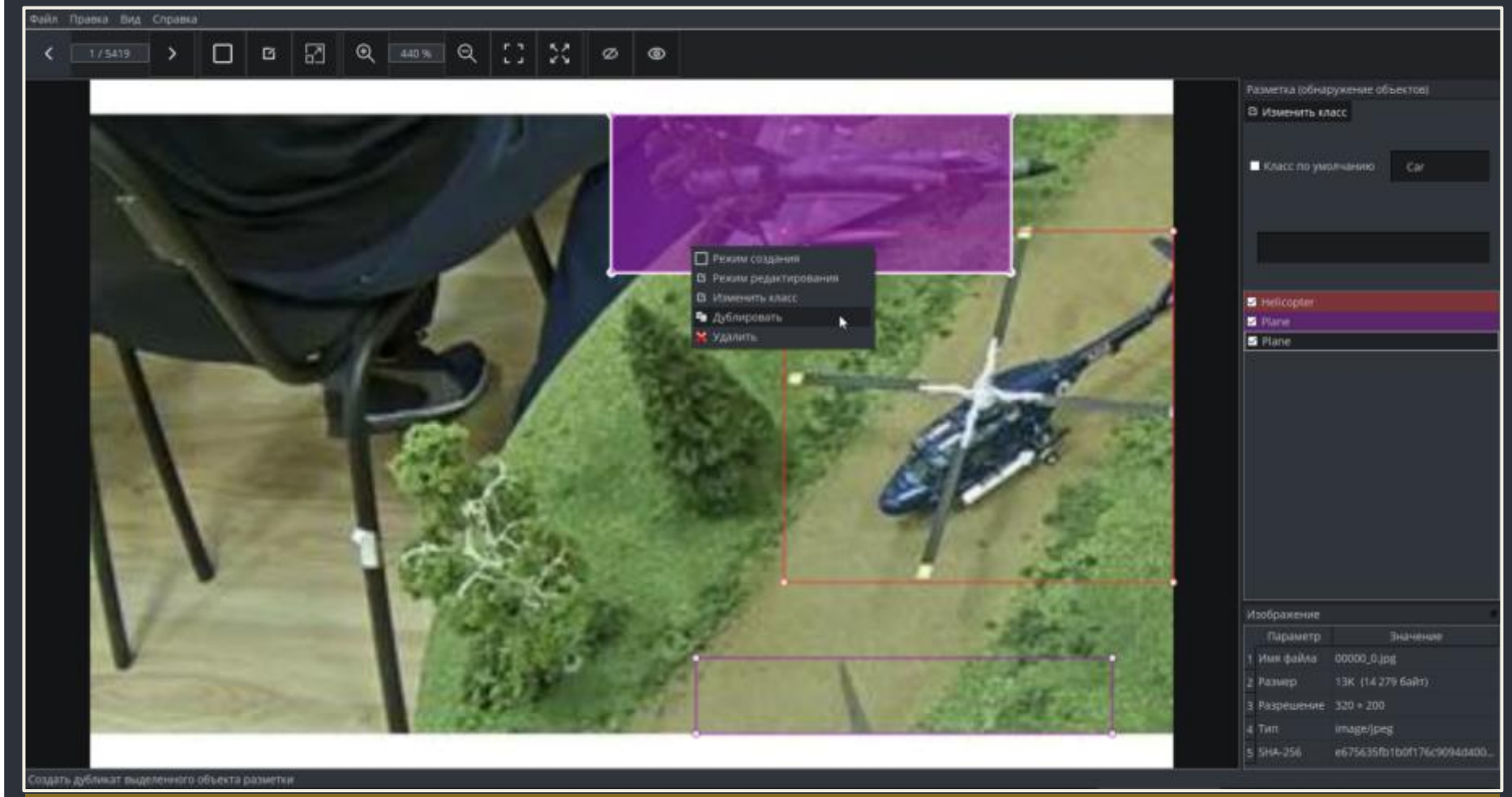
- личный кабинет оператора
- отображение списка всех доступных проектов/баз данных
- отображение свойств и информации о каждом проекте
- статистика по статусам выполнения заданий проекта
- создание нового проекта

Проект: подготовка данных. СПО АРМ-Р: Разметчик

АРМ Разработчика







Режимы работы разметчика изображений:

- предварительный просмотр изображения и существующей разметки с возможностью ее редактировать; - создание новой разметки;
- создание разметки слежением от одного изображения к другому.

Проект: подготовка данных

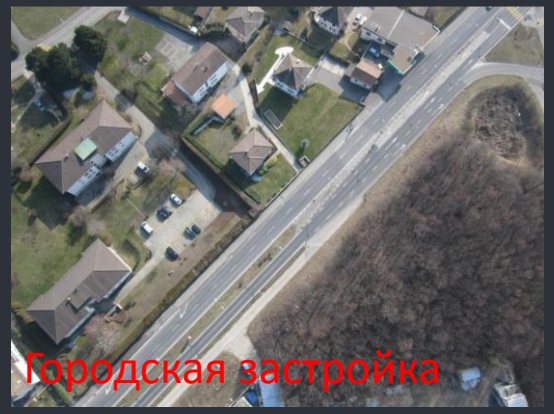
АРМ Разработчика

АРМ Администратора

Разметка изображений:

Разметка описывающими прямоугольниками

Установка атрибутов



Сегментация

Проект: автоматизация подготовки данных

АРМ Разработчика

АРМ Администратора

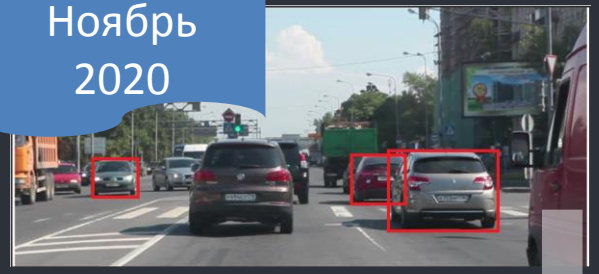
Июнь 2020

Межкадровая интерполяция
Алгоритмы слежения

Ноябрь 2020

Итеративное дообучение/разметка алгоритмом

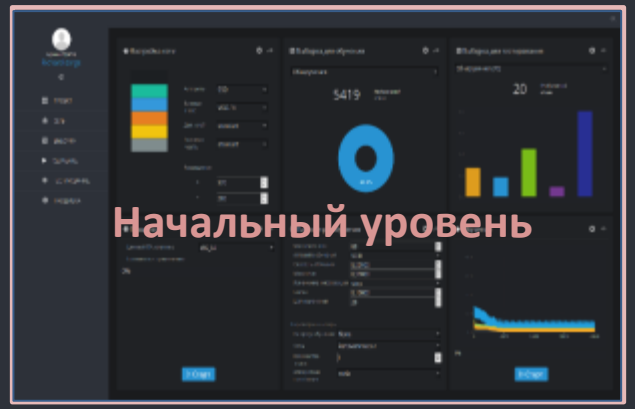
Ноябрь 2020

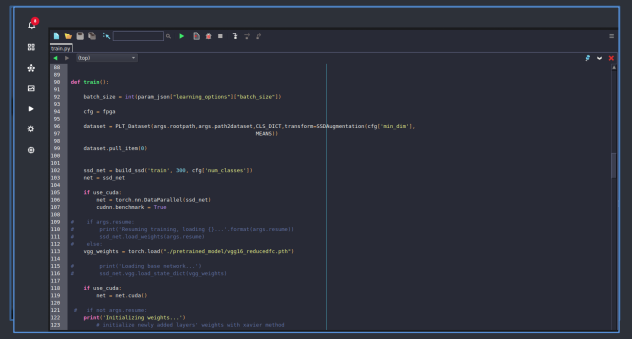


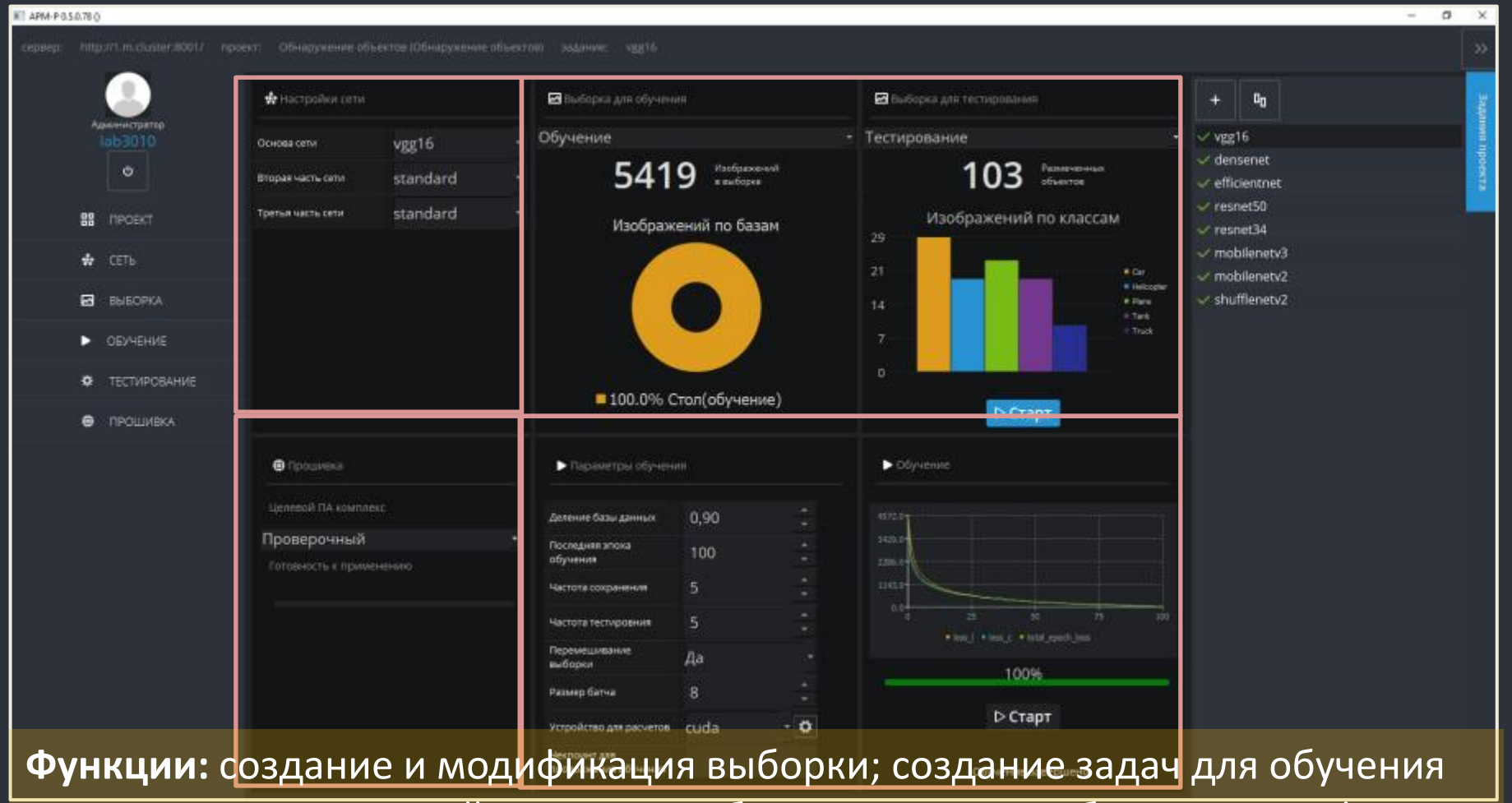
ИИ сегментация объектов

Проект: начальный уровень. СПО АРМ-Р: Работа с проектами и заданиями

АРМ Разработчика



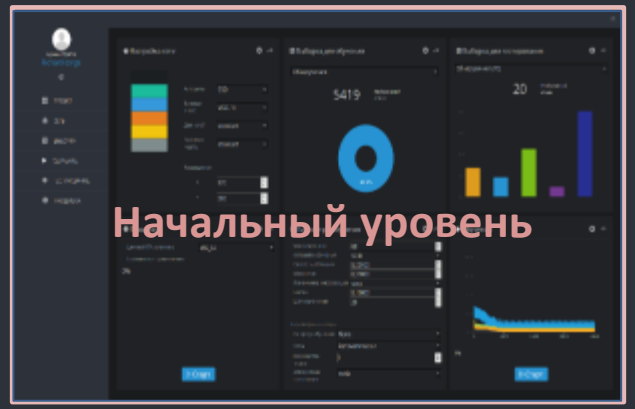


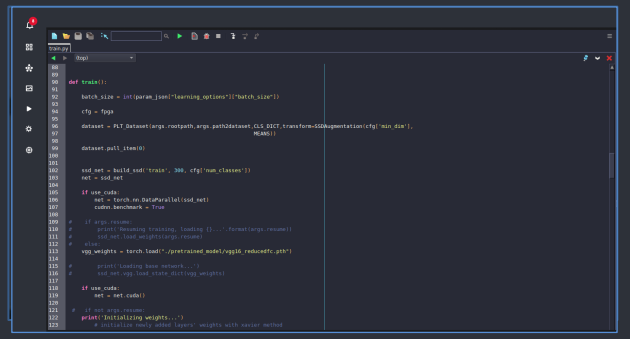


Функции: создание и модификация выборки; создание задач для обучения и тестирования; настройка и запуск обучения задач; отображение графика обучения и статуса задачи; задание и просмотр настроек сети; показ статуса и статистики по базе данных; запуск тестирования уже обученной задачи; визуальный контроль результатов тестирования; запуск генерации прошивки для обученной задачи и выбранной архитектуры.

Проект: формирование выборок

АРМ Разработчика





Создание выборки с использованием мастера выборок:

Доступные БД

БД летных экспериментов

Фильтр

1. Выбрать по тегу “ночь”
2. Выбрать класс “машина”
3. Выбрать класс “самолет”

БД синтез

Фильтр

1. Выбрать 10000 изображений
2. Выбрать класс “грузовик”

БД съемок 1

Фильтр

1. Выбрать данные за апрель
2. Выбрать по тегу “ошибка”
3. Сделать черно-белыми
4. Изменить размер до 320x200

Обучающая

Тестовая

Проект: Аппаратная реализация

АРМ Разработчика

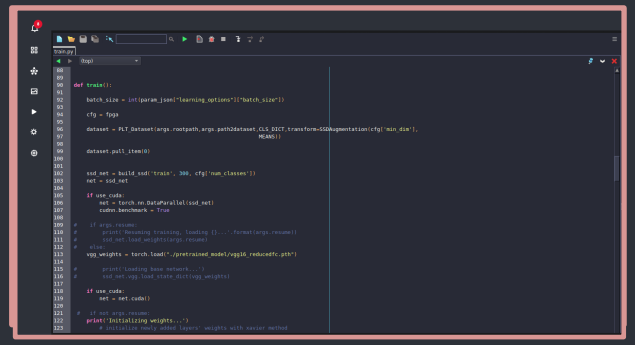
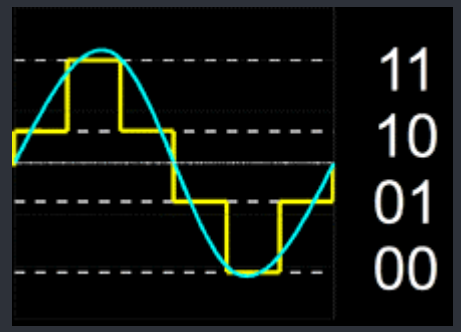


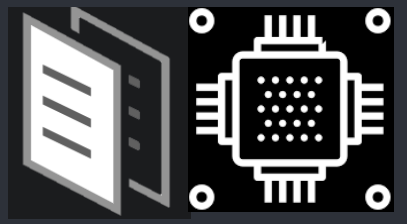
Схема работы SDK Платформы

Планировщик



Квантизация

Трансформация архитектуры



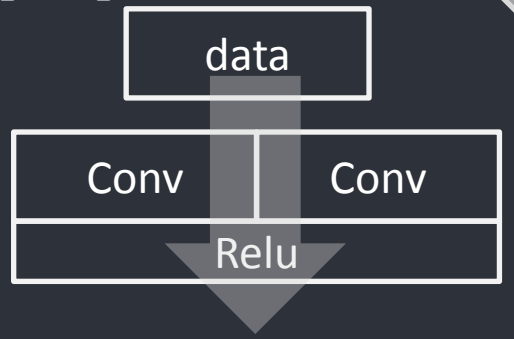
Аппаратно-ориентированное описание

Библиотека примитивов

Conv	ReLU	Deconv
Sigmoid	TanH	SoftMax
Pooling	InnerProduct	

Оптимизированная аппаратная реализация

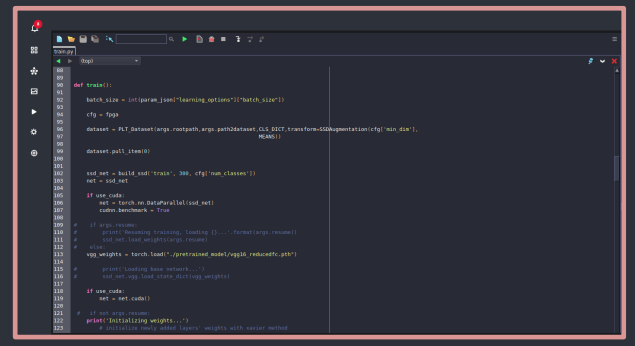
Оператор



Выполнение прямого прохода

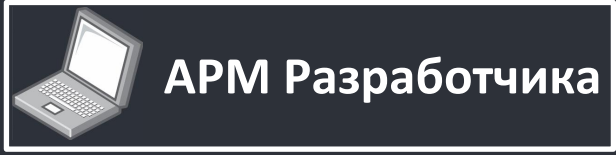
Проект: Аппаратная реализация

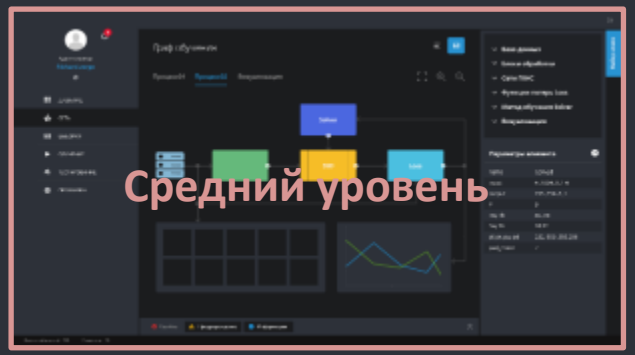
АРМ Разработчика

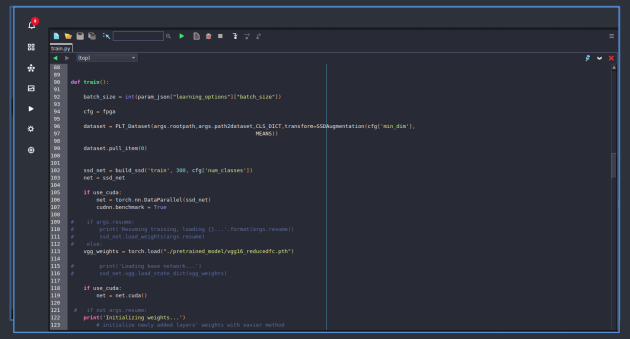


Направление	Поддержка к настоящему моменту	Интегрирование существующих решений или разработка	Состояние Работ	Осуществлено практическое тестирование
CPU x86-64	Да	Разработка	Завершены	Да
CPU ARM (Байкал-М)	Да	Разработка	Завершаются	Да
CPU + NVidia GPU (CC 6.0+)	Да	Интегрирование	Завершаются	Да
CPU + AMD GPU (gfx9)	Да	Интегрирование	Завершаются	Да
NeuroMatrix (НТЦ «Модуль»)	Да	Разработка	Завершаются	Да
ПЛИС UltraScale+ (Xilinx)	Да	Разработка	Завершаются	Да
ELcore (НПЦ «Элвис»)	Да	Интегрирование	Завершаются	Ожидается выпуск процессора
CPU Эльбрус (АО «МЦСТ»)	Да	Разработка	Завершаются	Да
СнК IVA (IVA technologies)	Да	Интегрирование + разработка	Ведутся	Ожидается выпуск процессора

Проект: средний уровень. Конструктор процессов (ноябрь 2020)



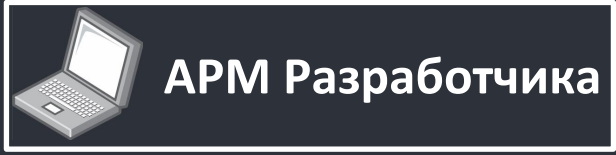


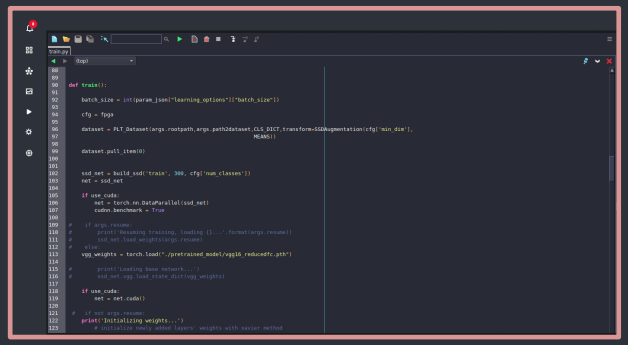


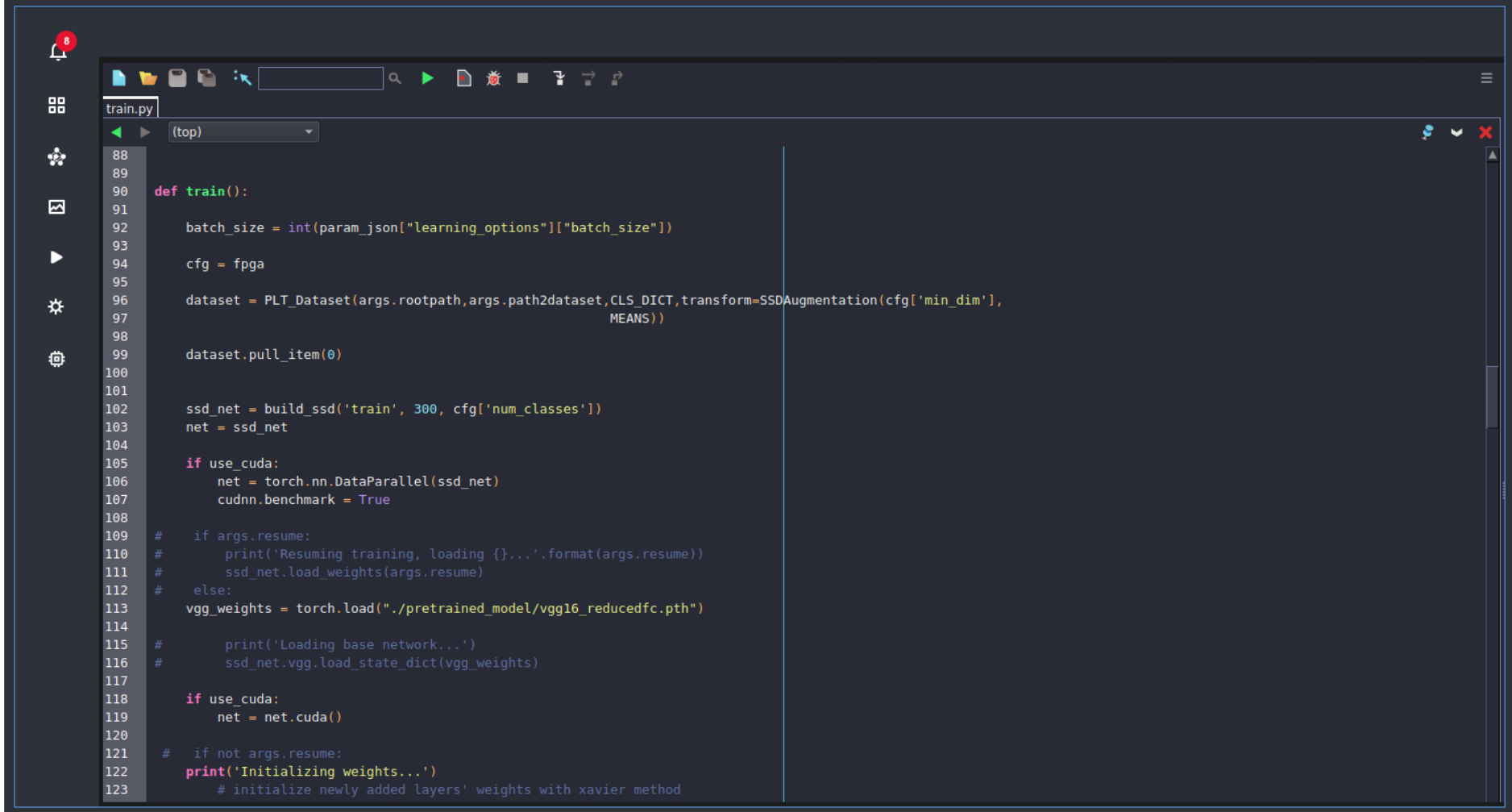
Возможности:

+ Формирование процесса обучения/тестирования
+ Визуализация в реальном времени
+ Задание функций потерь и методов обучения
+ Математические операции

Проект: Продвинутый уровень. Редактирование/отладка кода (ноябрь 2020)







Администрирование. СПО АРМ-А

АРМ Разработчика

Администратор проектов

- ✓ Создание проектов
- ✓ Добавление пользователей в проект
- ✓ Доступ пользователей к БД
- ✓ Отслеживание прогресса по проекту

АРМ Администратора

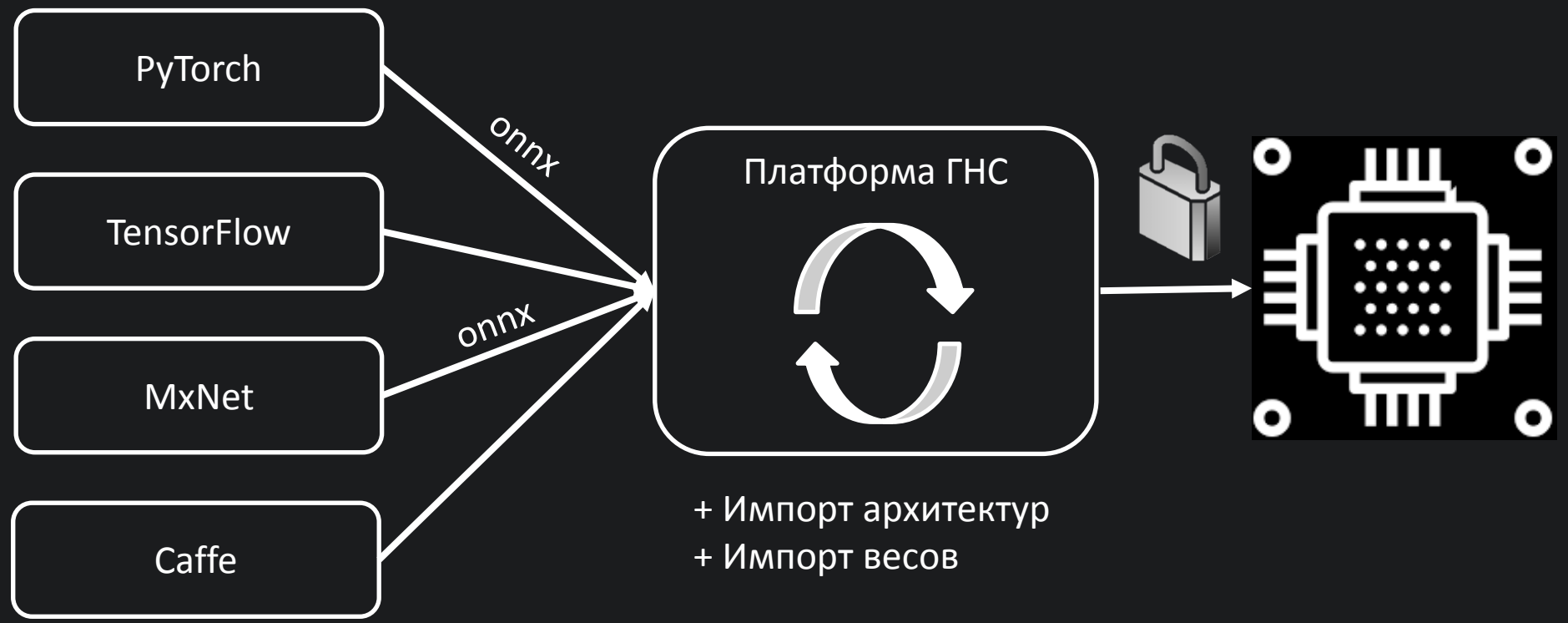
Администратор ресурсов

- ✓ Выделение вычислительных ресурсов
- ✓ Остановка/отмена заданий
- ✓ Резервное копирование
- ✓ Обновление системы
- ✓ Контроль состояния кластера/сервера

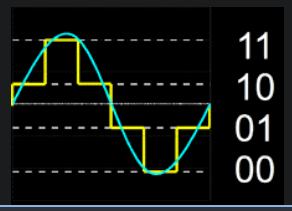
Импорт решений

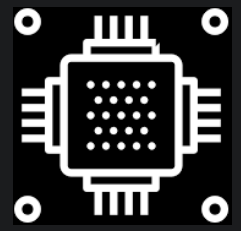
Инфраструктура
- Расп. обучение
- Контроль доступа
- БД
- Дообучение

Импорт решений из других сред



Аппаратная реализация





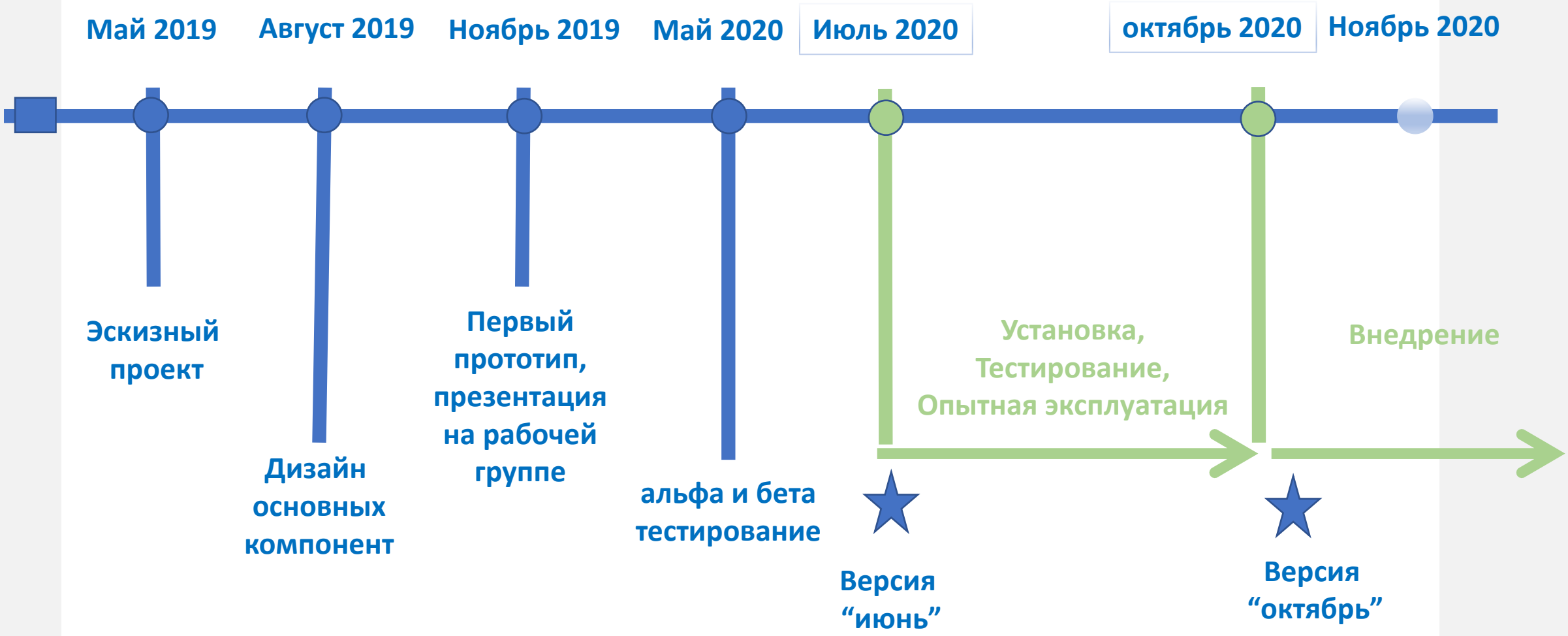
Сертифицированное ПО

Единый формат

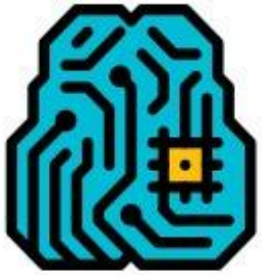
Перспективы развития Платформы (2021+)

Технологическое развитие инфраструктуры	✓ Создание виртуальных кластеров ✓ Управление и мониторинг	**Использование нескольких вычислительных кластеров**
Развитие технологии обучения	**Автоматическое обучение(AutoML)**	✓ Формирование архитектуры под задачу ✓ Подбор гиперпараметров обучения ✓ AutoML как сервис
Задачи управления и ИИ	✓ Решение задач управления ✓ Решение задач планирования ✓ Эффективная работа на кластере	**Обучение с подкреплением**
Задачи на произвольных типах данных	**Графовые сети**	✓ Задачи оптимизации ✓ Задачи большой размерности

Основные этапы проекта «Платформа-ГНС»



Будущая экосистема Платформы-ГНС. Переход к этапу тестирования и внедрения



Разработчики аппаратных платформ

Разработчики алгоритмов

Оператор: Центр технологий искусственного интеллекта *НИЦ «Институт им. Н.Е. Жуковского» (ЦТИИ) создан в 2020*

Невозможно вести такую разработку без учета мнения всех участников рабочего процесса!

Платформа - ГНС

Экспертная группа *проекта «Платформа-ГНС» существует с 2018 г.*

AITECHNOLOGY.RU

ЦТИИ

ЦТИИ.РФ

Информационные ресурсы *экосистемы Plat будут запущены в ноябре 2020*

Разработчики программных платформ

Конечные потребители

Контакт для участия в пилотном тестировании и опытной эксплуатации:

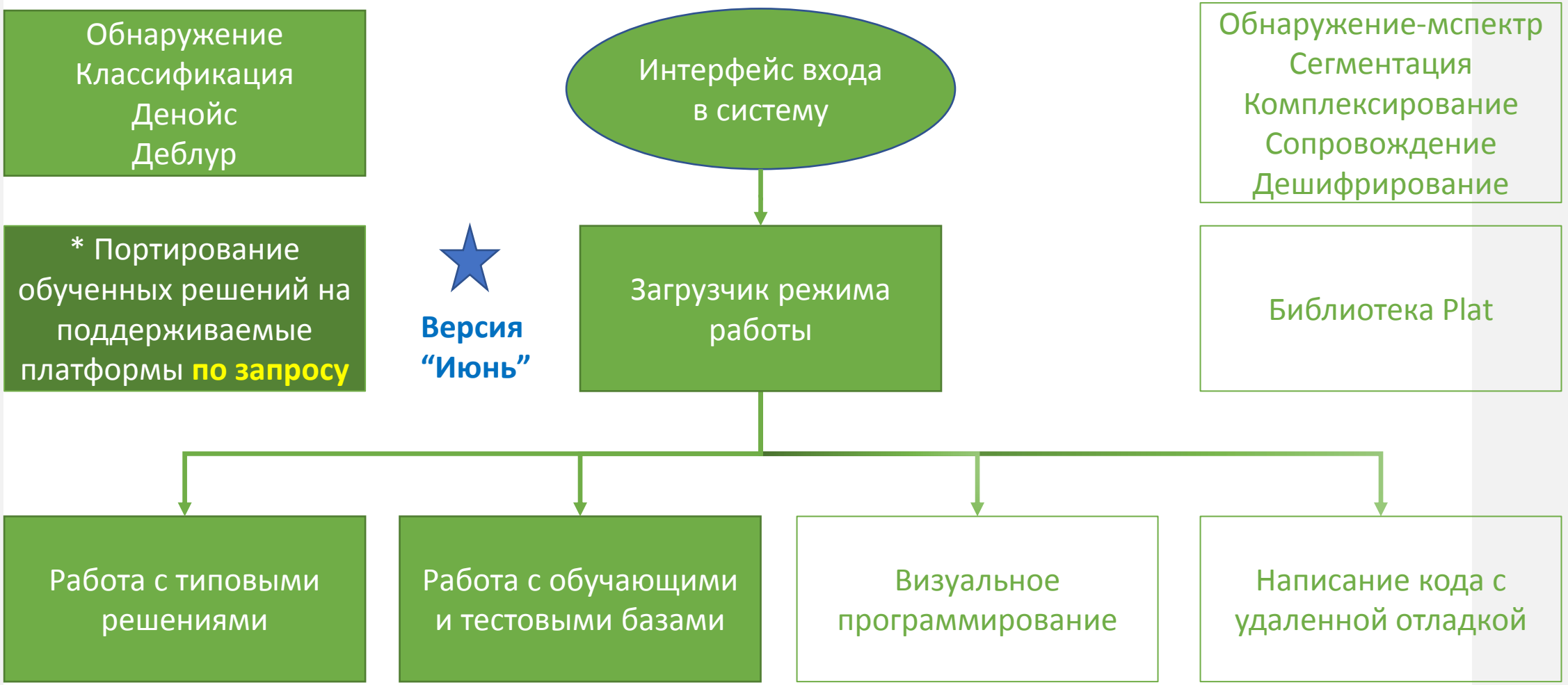
support.plat@gosniias.ru

ЗАЯВКА

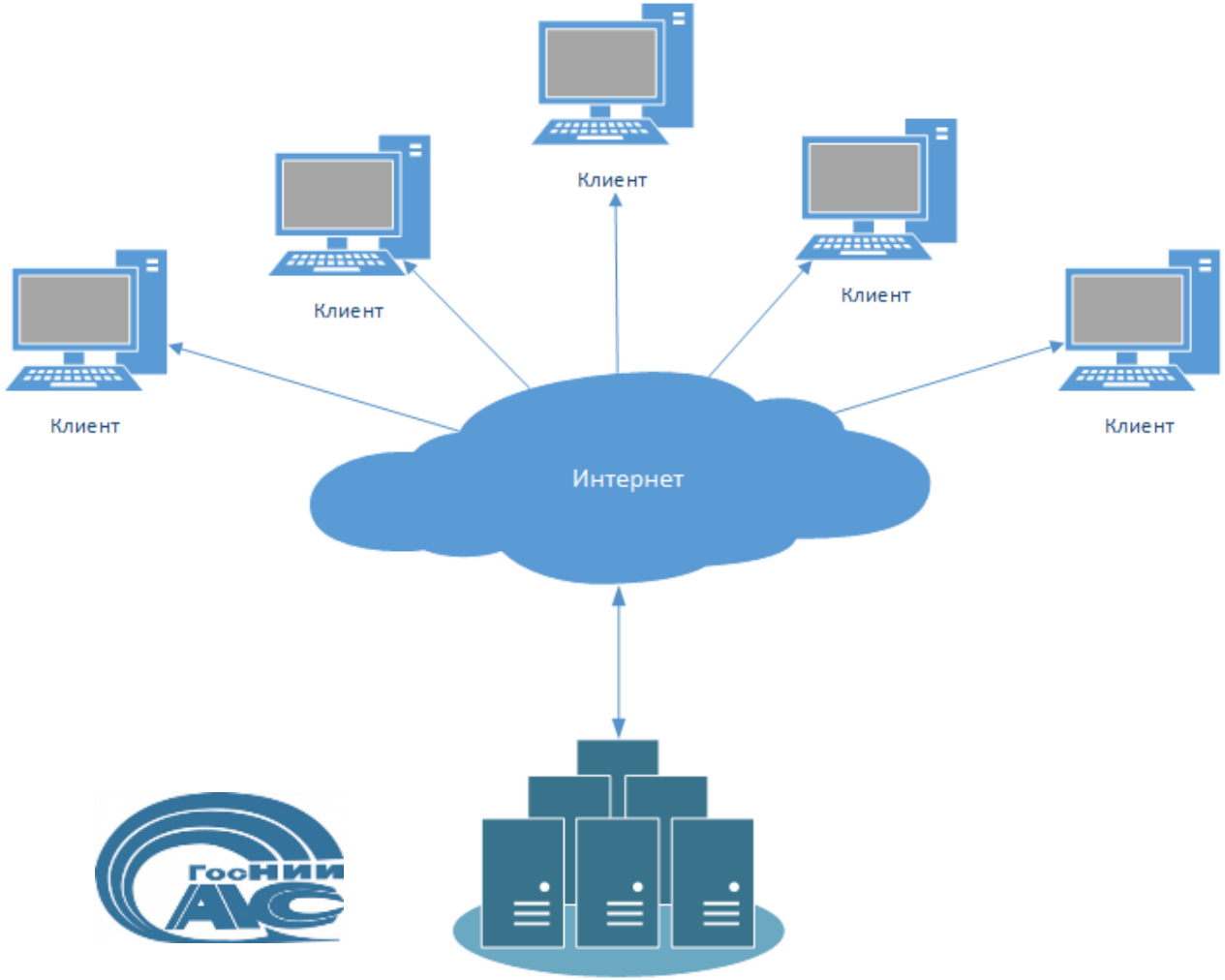
ДОГОВОР

УСТАНОВКА

АРМ-Р: возможности версии “июнь”



Технологическая схема участия в пилотном тестировании и опытной эксплуатации

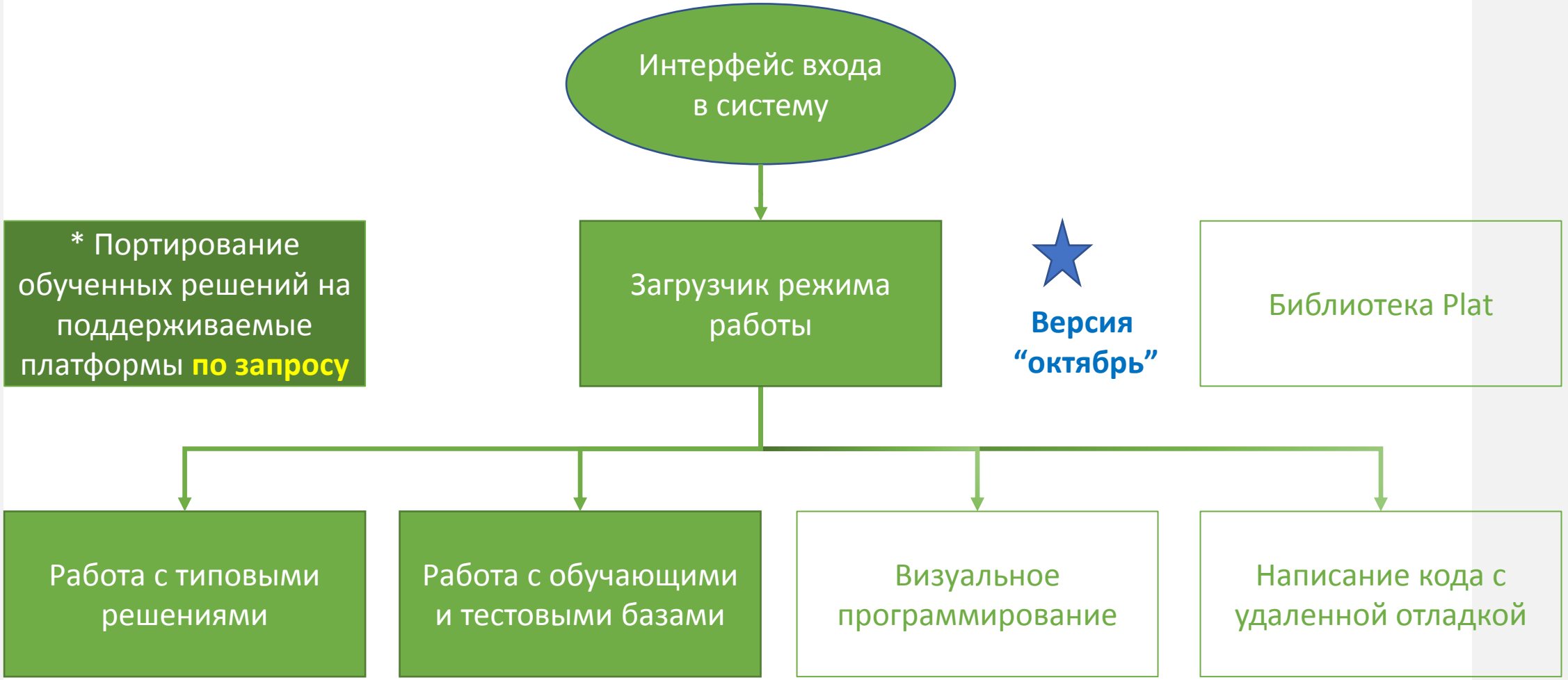


- Кластер ГосНИИАС
- Клиент (кросс-платформенный) – на стороне Заказчика
- Доступные типовые решения:
  - Классификация
  - Обнаружения объектов
  - Устранение шумов
  - Устранение смаза
- Установка на собственный кластер: по запросу

Требования к рабочему месту

1. Процессор: 64-битный Intel или AMD с частотой не менее 2ГГц;
2. ОЗУ не менее 4ГБ;
3. Видеокарта с поддержкой OpenGL 4.0;
4. Экран с разрешением не менее 1920x1080;
5. Свободное место на жестком диске не менее 500МБ;
6. Подключение к серверу со скоростью не менее 100Мбит/с;
7. Операционная система (Windows 7/8/10 или Ubuntu Linux 18.04 и новее или Astra Linux 1.6 релиз «Смоленск»).

АРМ-Р: возможности версии “октябрь”



Графическое программирование

Версия “октябрь”

Конструктор процессов:

Возможности:

+ Формирование процесса обучения/тестирования
+ Визуализация в реальном времени
+ Задание функций потерь и методов обучения

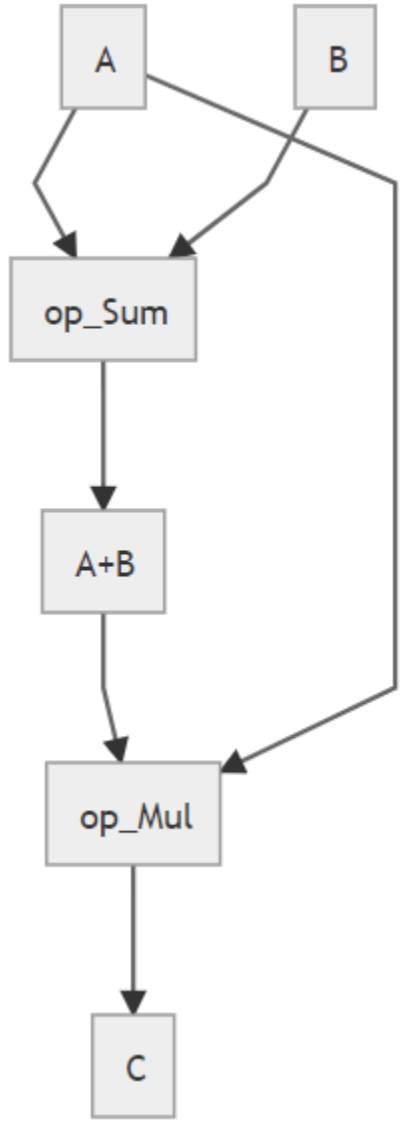
PlatLib

Версия “октябрь”

Обучение:

- Динамические графы
- Распределенное обучение
- Поддержка GPU AMD/NVIDIA
- Поддержка CPU x86-64/Elbrus
- Поддержка БД Платформы

```
A = PTensor(5,5,5,5)
B = PTensor(5,5,5,5)
C = (A+B)*A
```



PlatLib

Обучение:

- Аналогично pytorch
- Отладка на python
- Нет multiGPU
- Распределенное обучение через nccl/rccl/mpi

```
class LeNetPlat(plat.module.Module):
    def __init__(self):
        super(LeNetPlat, self).__init__()
        self.conv1 = conv2D(20, 1, 5)
        self.relu1 = relu()
        self.pool1 = poolmax2d()
        self.conv2 = conv2D(50, 20, 5)
        self.relu2 = relu()
        self.pool2 = poolmax2d()
        self.fc1 = conv2D(500, 50, 4)
        self.relu3 = relu()
        self.fc2 = conv2D(10, 500, 1)
```

```
res_plt = LeNetPlat(inputT)
plt_res = SM(res_plt, labelsT)
plt_res.backward()
optimizer_plat.iter()
```

Программирование в стиле plat: набор примитивов Proj

Обучение:

- Создание собственного типового решения Платформы
- Разделение по стадиям
- Хорошая читаемость кода
- Синхронные аугментации
- Именованные типы тензоров
- Автоматическая генерация

параметров в формате Платформы

для отображения и передачи в APM

```
class LeNetProject(plat.proj.Network):
        def __init__(self):
             super(plat.proj.Network, self).__init__()
            self.net=plat.proj.zoo.LeNet((28,28),10)
            self.loss=plat.nn.softmaxwloss()

        def stepTrain(self,data,labels):
            return({‘loss’,self.loss(self.net(data))})

plat.Trainer.fit(LeNetProject(),plat.db.MNIST())
```

Современное состояние технологий искусственного интеллекта и отечественная нейросетевая платформа Plat

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

Спасибо за внимание!

Контакт по Платформе Plat:

support.plat@gosniias.ru

Семинар МГУ

Москва, 13.10.2020