

Компьютерное зрение и глубокое обучение: актуальные результаты и направления развития

Визильтер Юрий Валентинович, viz@gosniias.ru

начальник подразделения 3000 ФГУП «ГосНИИАС»,

д.ф-м.н., профессор РАН

Москва, ММРО-2019, 26.11.2019

Содержание:

тенденции/перспективы и актуальные результаты в ИИ

- Компьютерное зрение, глубокое обучение и **искусственный интеллект**

- краткий дайджест предыдущих докладов

- что мы делаем для внедрения ГНС (2018-2019)

**Взгляд-2017:
Революция ИИ**

- Глубокие нейронные сети, **глубокая оптимизация** и глубокое управление

- от анализа данных к оптимизации и управлению

- глубокое обучение с подкреплением на графах

**Взгляд-2019:
Революция ГНС**

- Актуальные результаты в **компьютерном зрении** и глубоком обучении:

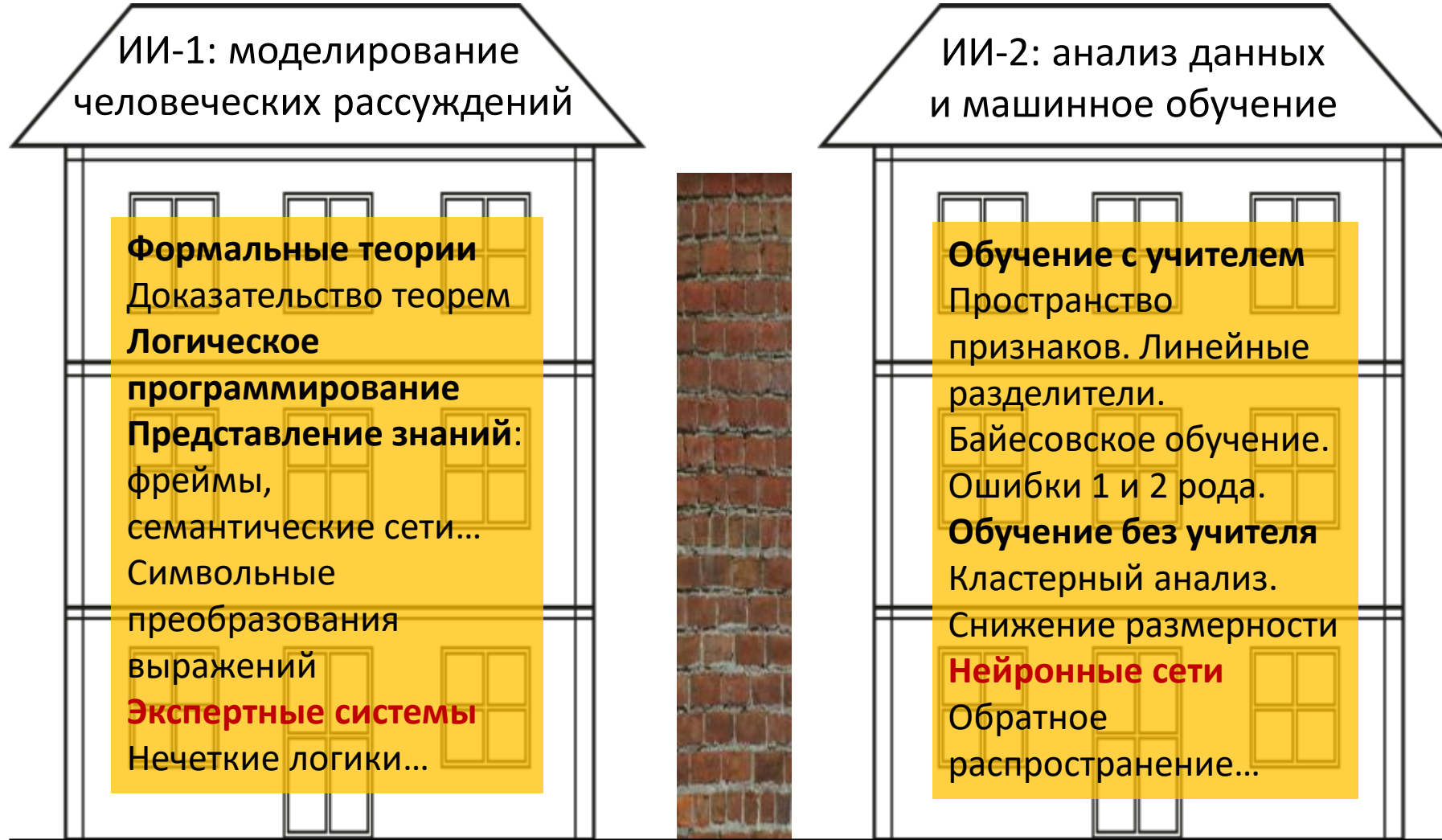
- атаки и защиты, внимание, графы сцен, синтез данных, объяснения, обучение без примеров...

**ICCV-2019:
Революция
продолжается,
но замедляется?
(успехи и вызовы)**

**Компьютерное зрение,
глубокое обучение
и искусственный интеллект**

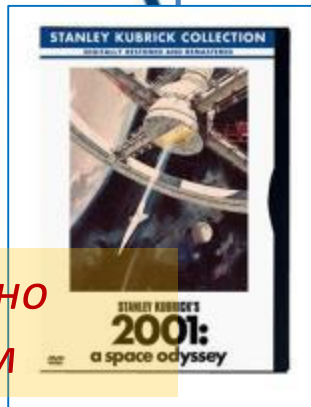
Технологии «искусственного интеллекта» (ИИ)

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



Как развивались технологии «ИИ», чего достигли, чего ждать? (Взгляд-2017)

Ранний период развития ИИ: созданы основные методы и подходы...



Завышенные ожидания

2000
«Зима ИИ»
Прогноз: 2040+

Разочарование

Рост интереса

Алгоритмы ИИ сильно проигрывают людям

«Кривая хайпа» для ИИ

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) | **Прогноз: 2020+**

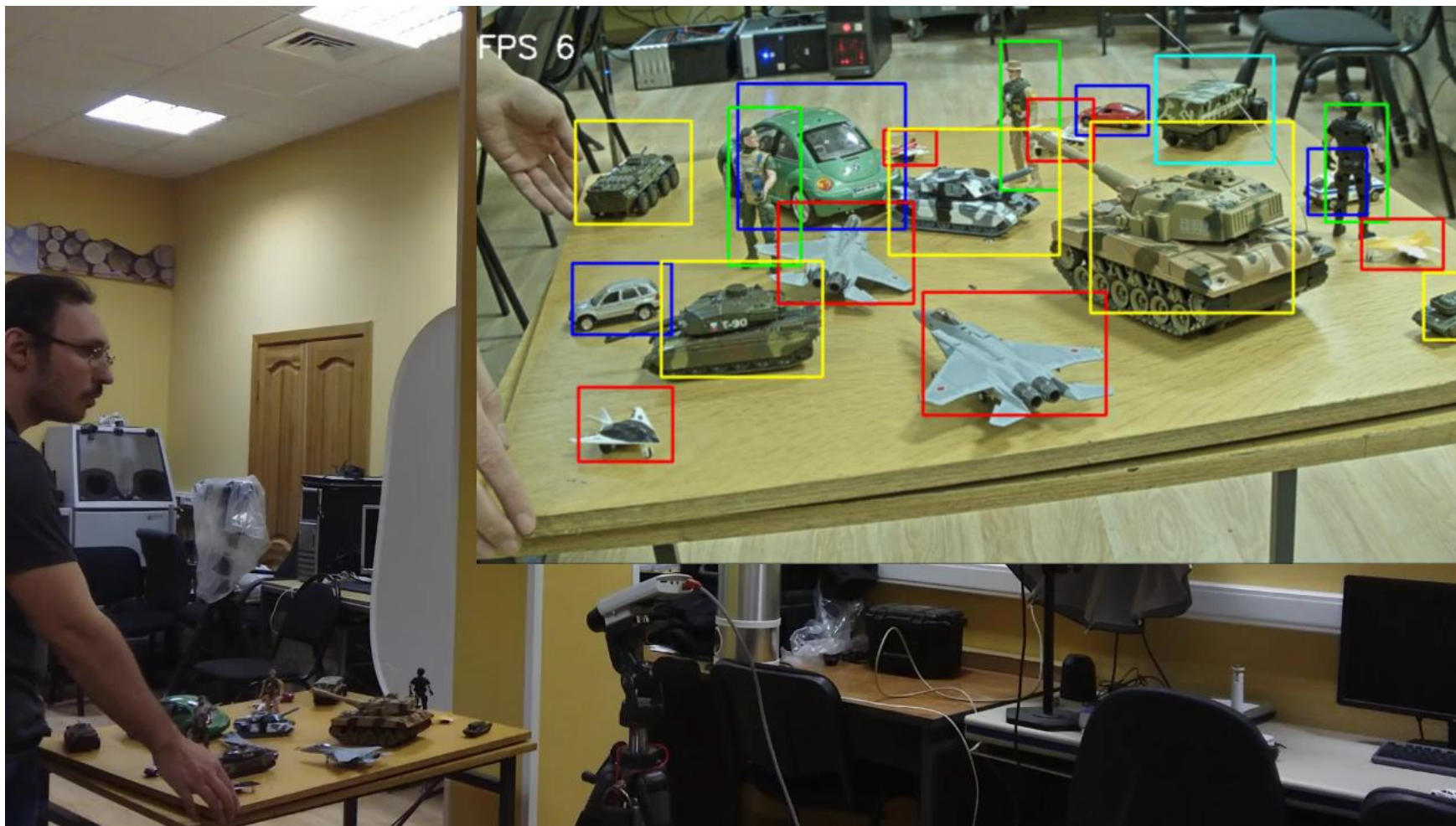
- Появление глубоких нейронных сетей (ГНС)
- Решение задач компьютерного зрения, обработки сигналов и анализа больших данных на уровне человека и выше (superhuman)

2011-12

Функциональный («слабый») ИИ

...2011+: две волны технологической революции, «слабый» ИИ в 2020

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?

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 MC121.01 производства НТЦ «Модуль» с процессором NM6407 (5 кадров/сек).

2019: В настоящее время завершается разработка решения для АТР с ГНС на базе **NM6408**, работающего в 32 раза быстрее (скорость обработки - 60 кадров/сек и выше).

Перспектива-2020+: Проект ФПИ «Мемристор» (МФТИ, НТЦ «Модуль»)



2019: отечественные процессоры уже есть

ОКР «Платформа-ГНС» (2018-2020)

Проблема: используются зарубежные средства обучения ГНС (системы Caffe, Caffe2, Pytorch, TensorFlow, Theano и др.):

- Невозможна сертификация на НДС
- Нет поддержки отечественных датчиков и вычислителей
- Несовместимость различных средств и систем разработки

Решение: создание отечественной программной Платформы для обучения и бортовой реализации ГНС

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

Унифицированная платформа

- **Сертифицированный на НДС исх. код**
- Импорт/экспорт из основных фреймворков и опnx
- **Наличие типовых решений для основных задач**
- Поддержка отечественных аппаратных платформ и отечественных ОС
- Контроль доступа к данным и проектам
- **Низкие требования к квалификации ИТР**

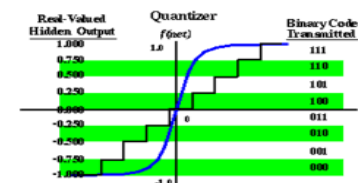
Снижение порога входа => резкое увеличение числа разработчиков => возможен прорыв в массовом внедрении ГНС в изделия ВВСТ

Единая интегрированная среда

Обучение ГНС

Преобразование ГНС

Аппаратная реализация ГНС



2020: будет отечественное ПО

**Глубокие нейронные сети,
глубокая оптимизация
и глубокое управление**

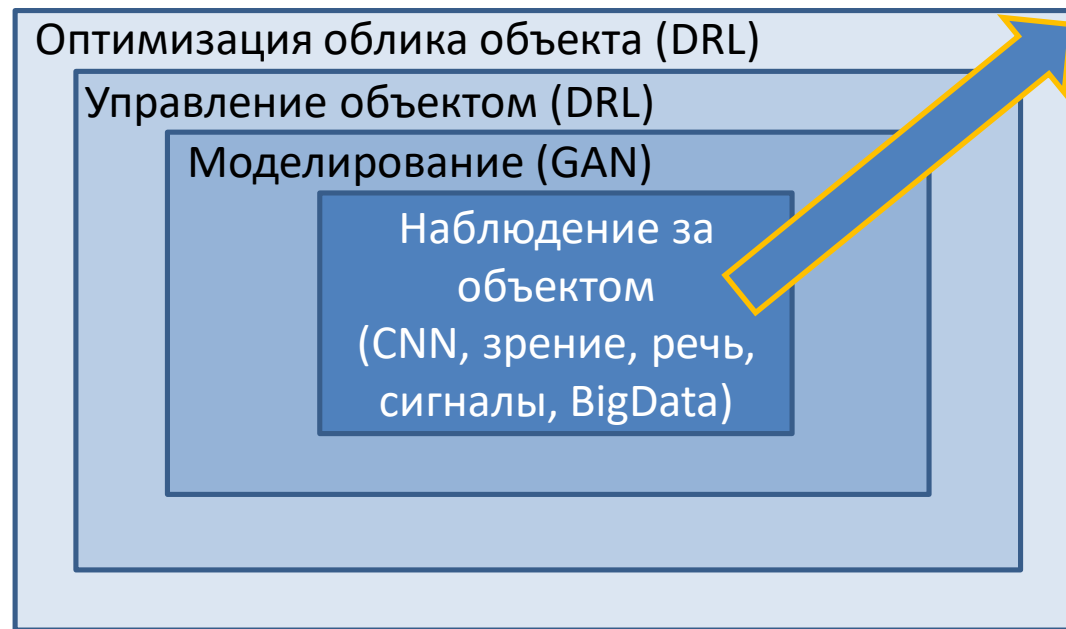
Глубокая оптимизация: **взгляд-2019** на революцию ГНС и ее перспективы

*2020+:
при
помощи
ГНС
может
быть
создан
«слабый»
бортовой
ИИ...
но это не
главное -
ГНС
могут
дать
науче 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.**



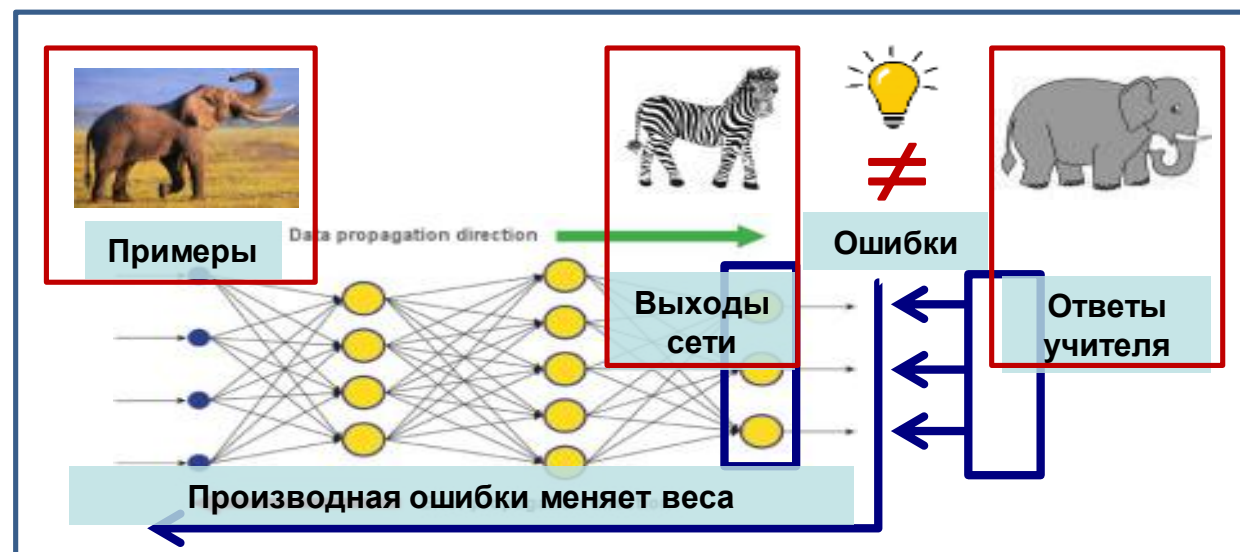
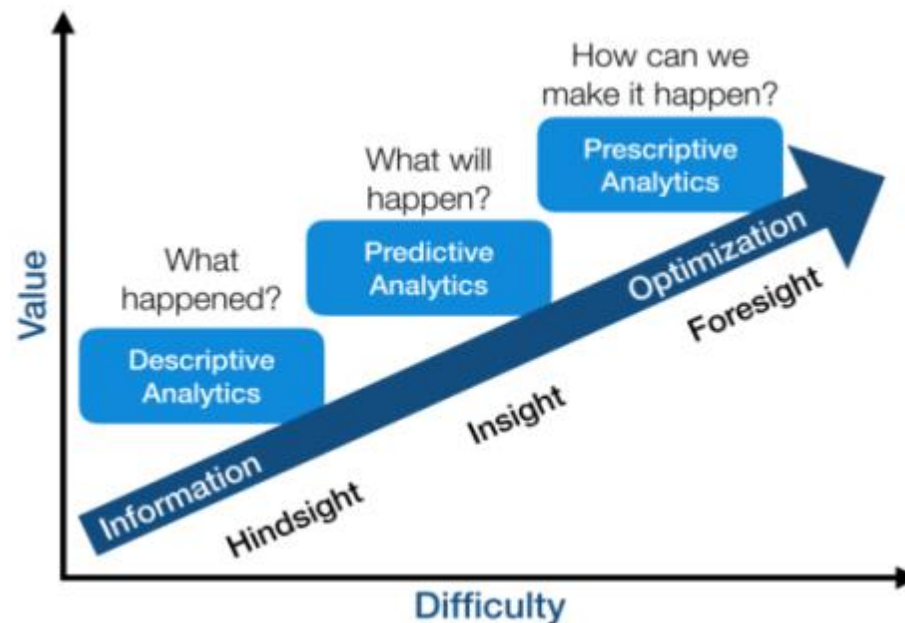
ГНС могут решать любые задачи, которые формулируются как задачи оптимизации.



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

Глубокая оптимизация: как и почему это работает

Мы исходим из двух фактов:

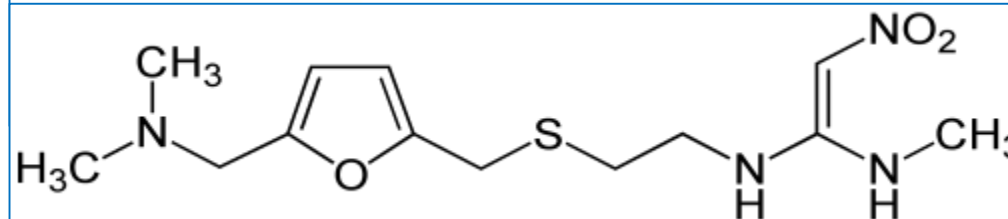
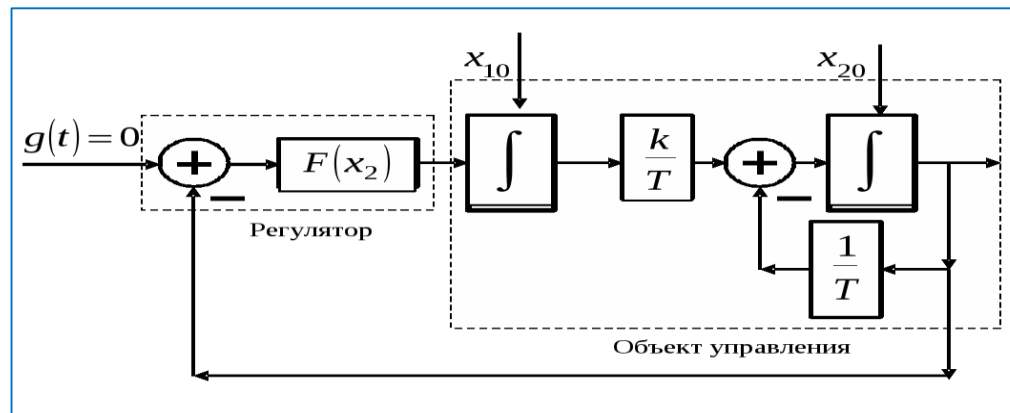
1. Все математические модели описываются графами
2. Все задачи прикладной математики решаются как задачи оптимизации на графах

Какое отношение это имеет к нейросетям?

Вот какое:

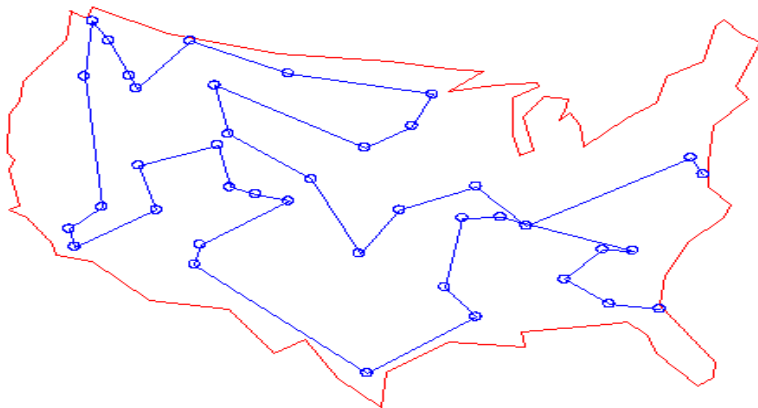
1. Глубокие Нейросети теперь умеют обрабатывать графы (Deep Graph Embedding)
2. ГНС на графах могут помочь значительно лучше решать задачи оптимизации.

Почему? Как это работает?



Глубокая оптимизация в классических задачах

ПРИМЕР: Математически доказано, что найти оптимальный (кратчайший) путь с возвратом, посещая каждый город один раз, можно только полным перебором (NP-трудная задача)

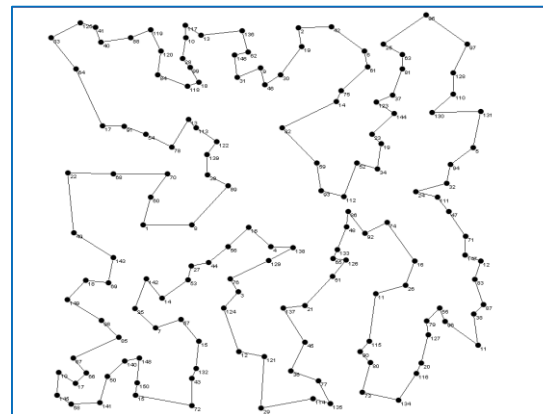


ГНС могут эффективно выучивать эвристики работы с графами в этих задачах.

Как правило, реальные задачи в частном случае намного проще, чем в общем.

ГНС могут выучить неявные особенности практической частной задачи (неизвестное распределение случайных элементов) и выразить их в эвристиках ее эффективного решения.

Задача коммивояжера



Реальная
практическая
задача



Задача официанта

NP-трудна
(решается
полным
перебором)

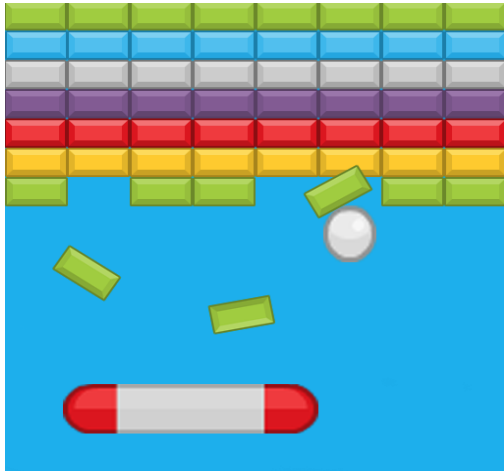
??????

Эффективная
эвристика
неочевидна для
человека, но
может быть
распознана ГНС

Оче-видна
(эвристика
решения легко
распознается
по образу
задачи)

СЛОЖНОСТЬ ЗАДАЧИ

Deep Reinforcement Learning: обучение с подкреплением



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

Каким способом учить такие сети? Способ оказался давно известен, и он связан с задачами управления!

- **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 $\hat{Q}(S, i)$** : your predicted future total rewards
Стоимость (выигрыш в будущем)
- **Policy $\pi(s)$** : How to choose your action
Решение (какой ход выбрать)

Функция Беллмана!

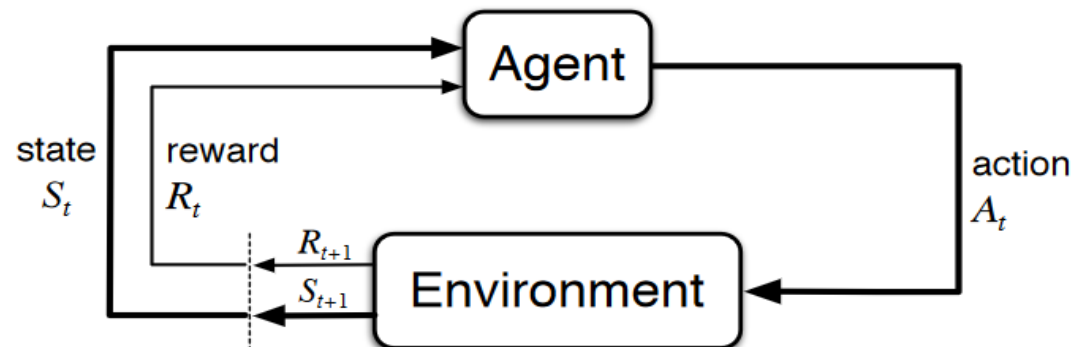
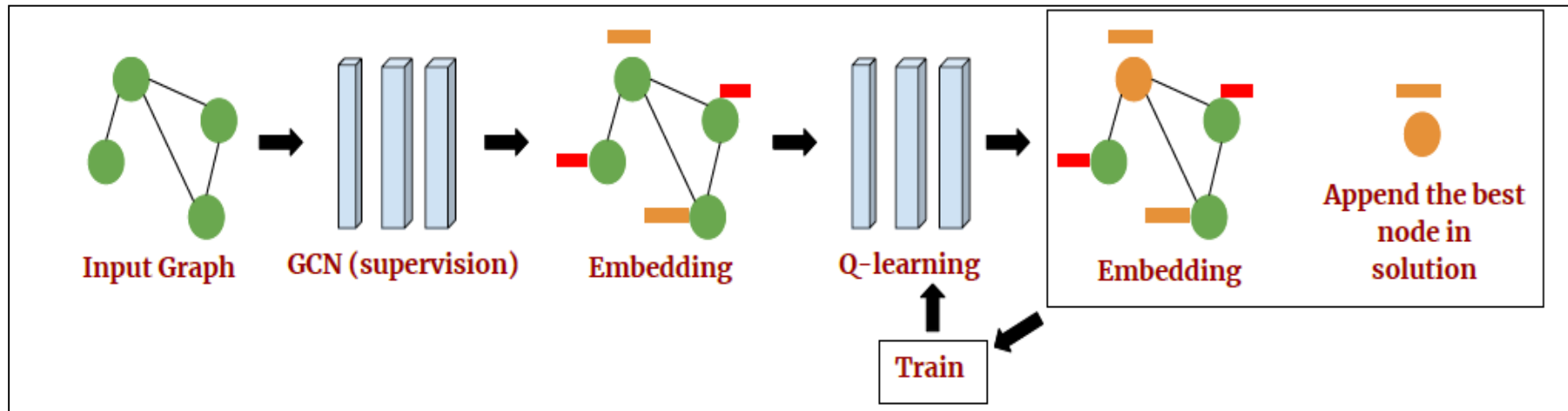


Схема знакома всем специалистам по теории управления

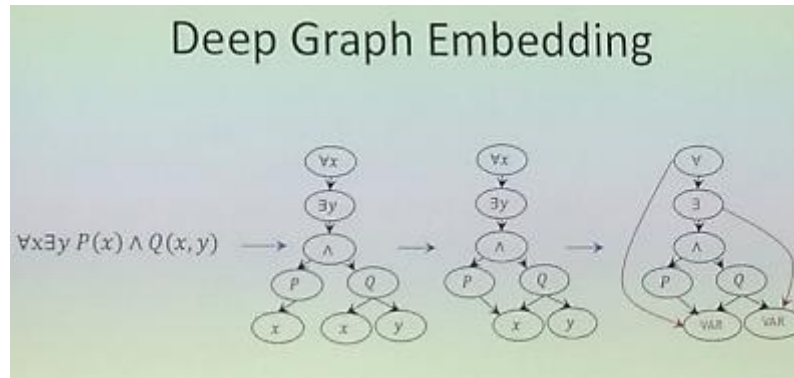
Deep Graph Embedding + Deep Reinforcement Learning

Соединяем глубокие сети на графах и обучение с подкреплением: учимся «играть» в классические задачи оптимизации, которые традиционно используются в бизнесе, логистике, транспорте...



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

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



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

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>

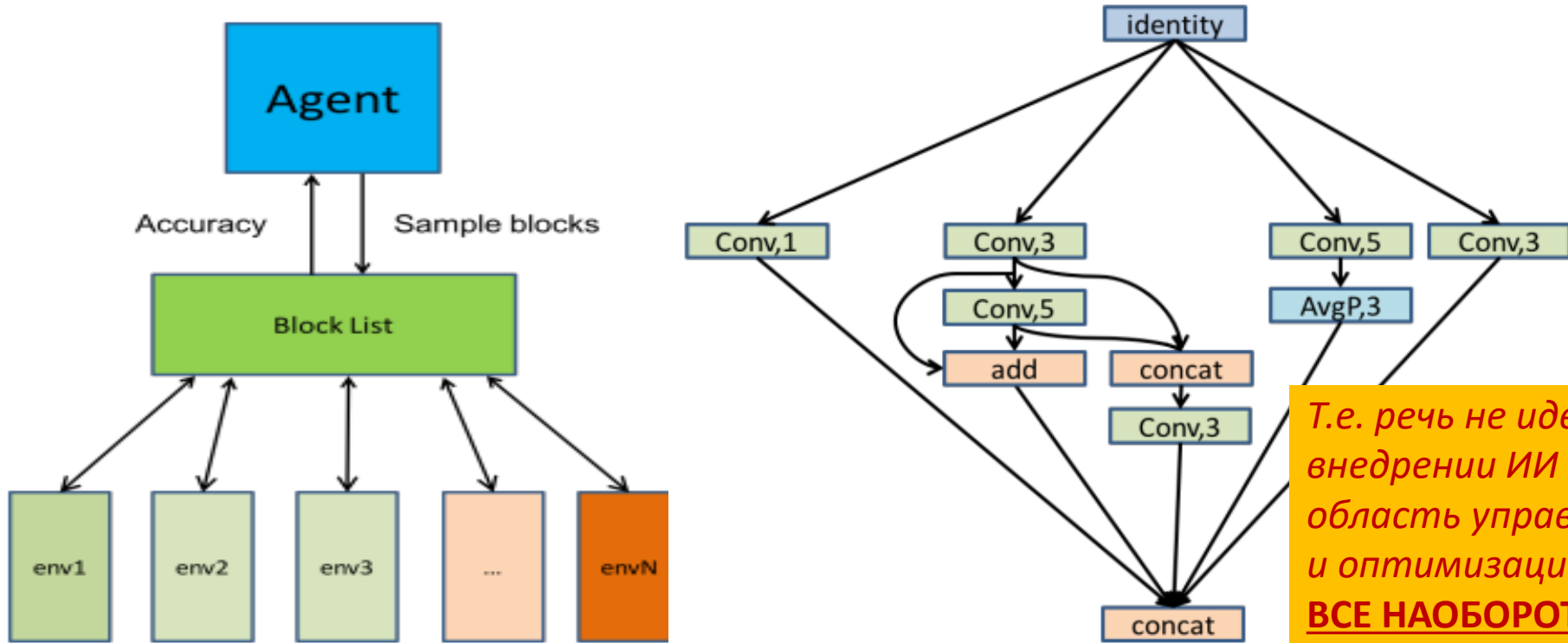
**Глубокая оптимизация и
позволила объединить
обучаемость ИИ-2
с интеллектуальностью ИИ-1**

Deep Graph Embedding + Deep Reinforcement Learning

Глубокие сети формируют и учат глубокие сети

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

Глубокая оптимизация
– основа AutoML



Т.е. речь не идет о внедрении ИИ в область управления и оптимизации. **ВСЕ НАОБОРОТ:** Методы глубокого управления и оптимизации внедряются в области ИИ!

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

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

Глубокая оптимизация в исследовании операций и управлении

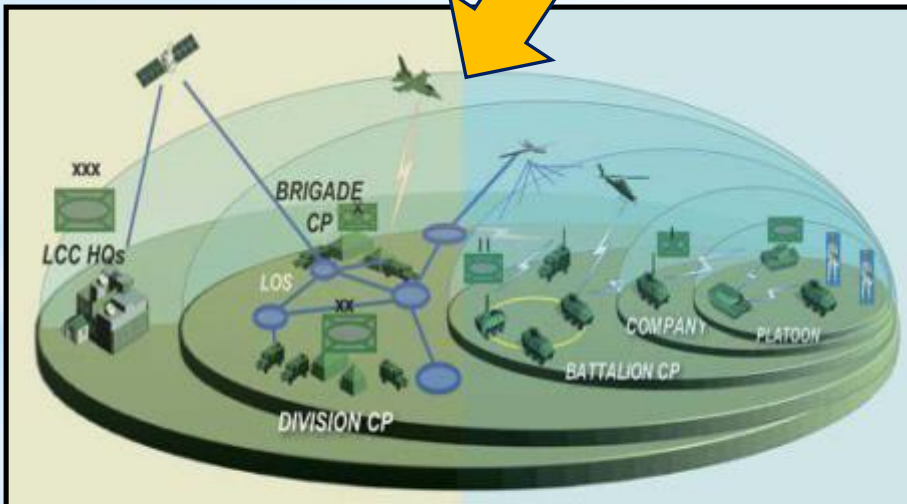
2019: ГНС могут решать любые задачи, которые формулируются как задачи оптимизации

Технология созрела, необходимо ставить серьезные НИОКР

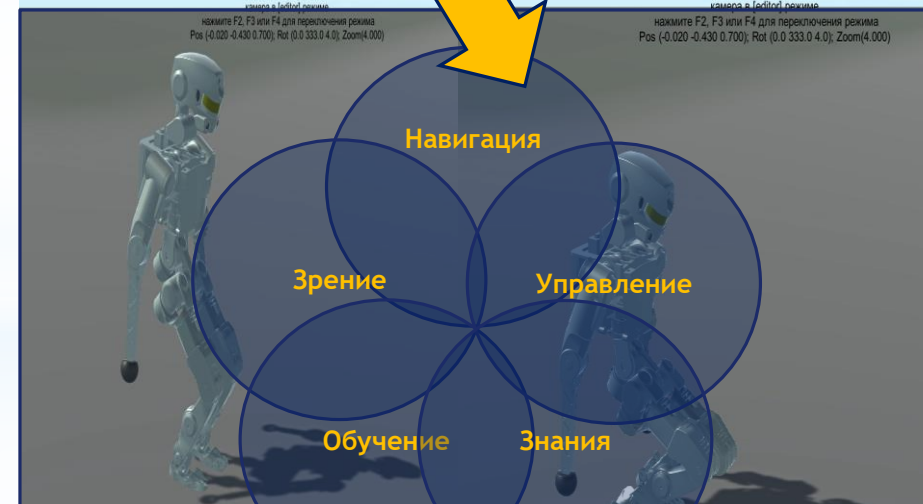


StarCraft II - тактический симулятор с упрощённой моделью ведения боя

25 января 2019 г. компания DeepMind опубликовала релиз AlphaStar: Mastering the Real-Time Strategy Game StarCraft II - глубокая сеть AlphaStar счетом 11:1 победила ведущих профессиональных игроков в StarCraft II



Оперативно-тактическое и групповое управление



Автономное управление и оптимизация облика изделий

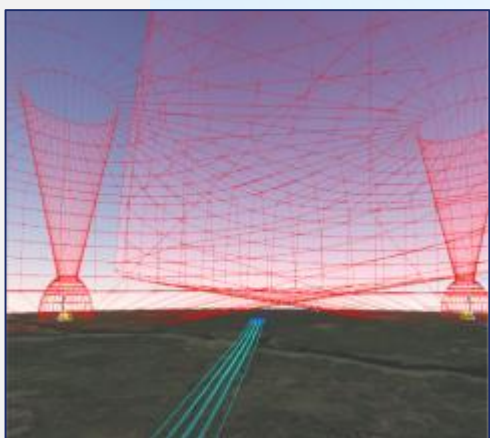
ТЕКУЩИЕ ПРОЕКТЫ (ГОСНИИАС, 2019)

Решение задачи планирования/отражения атаки групп БЛА на наземные объекты

Компьютерная модель ПВО и БЛА ГосНИИАС



- Моделирование всех современных средств-поражения ПВО и ВВС
- Моделирование БЛА и их носителей
- Моделирование особенностей работы средств РЭБ



- Моделирование рельефа и работы РЛС
- Моделирование структуры ВС и подразделений
- Возможность возврата и по времени

Алгоритмы:

multi-agent RL

Human-level performance in first-person multiplayer games with population-based deep reinforcement learning. 2017

population based RL

Human-level performance in first-person multiplayer games with population-based deep reinforcement learning 2018

actor-critic

IMPALA: Scalable Distributed Deep-RL with Importance Weighted Actor-Learner Architectures. 2018



Апрель 2019: U.S. Army Research, Development and Engineering Command (RDECOM) Research Laboratory и U.S. Army Research Laboratory (ARL) сообщили о решении **в 2019 г. создать Армейский Инновационный Институт Искусственного Интеллекта (A2I2)** для координации и ускорения фундаментальных исследований специфичных для армии проблем **развития ИИ в направлении автономного управления межвидовыми операциями**

Ключевая технология:
обучение
с подкреплением
на графах

ТЕКУЩИЕ ПРОЕКТЫ (ГОСНИИАС, 2019)

Обучение движению антропоморфного робота (совместно с АО “Андроидная техника”)



Алгоритмы: **multi-agent RL**

Human-level performance in first-person multiplayer games with population-based deep reinforcement learning. 2017

population based RL

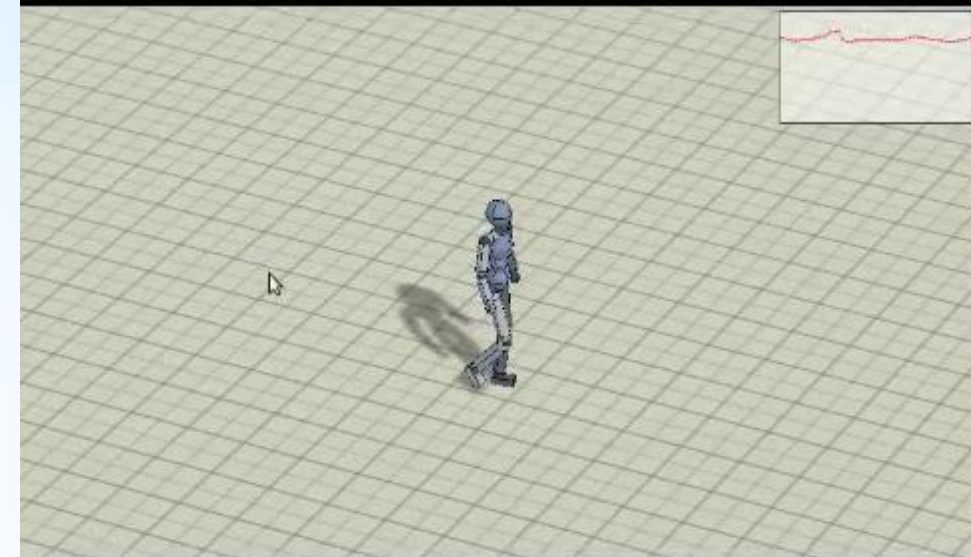
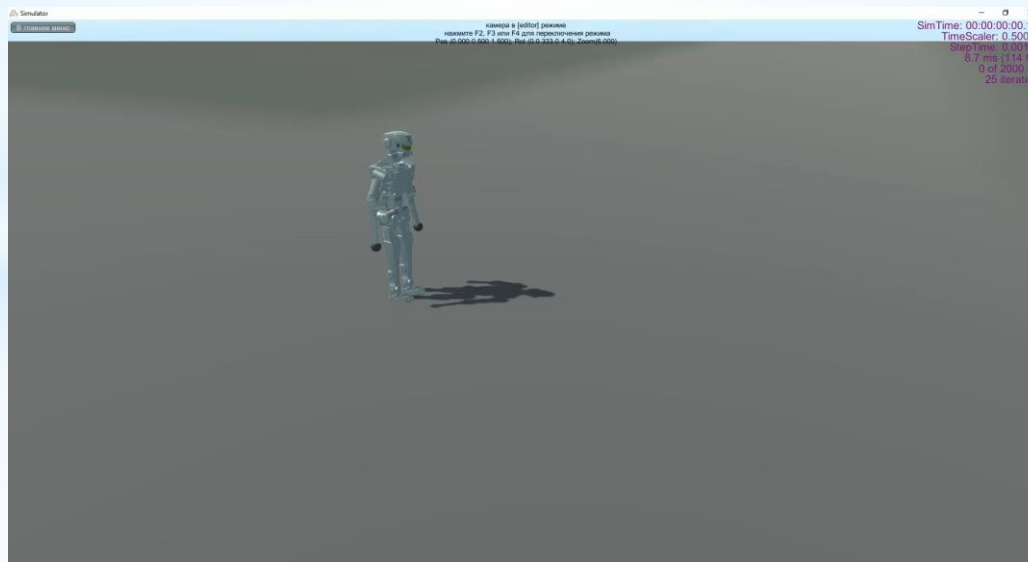
Human-level performance in first-person multiplayer games with population-based deep reinforcement learning 2018

actor-critic

IMPALA: Scalable Distributed Deep-RL with Importance Weighted Actor-Learner Architectures. 2018

Основные проблемы:

Медленная мат. модель



Глубокая
оптимизация
в игровых
задачах
и задачах
управления

**Актуальные результаты
в компьютерном зрении
и глубоком обучении**

По материалам конференции ICCV-2019



ICCV 2019
Seoul, Korea

Ключевые темы ICCV-2019

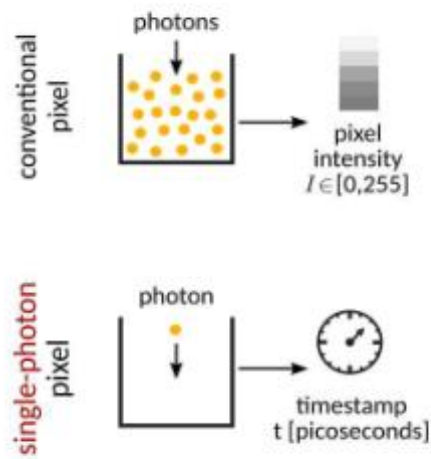
- Sensors
- Adversarial Attack
- AutoML, Neural Architecture Search (NAS)
- Graph Convolutional Networks (GCN), Attention
- Domain Adaptation, Generative Adversarial Networks (GAN)
- Knowledge Distilling
- Few-Shot Learning / Detection / Segmentation
- Zero-Shot, Grounding
- MemoryNet
- Scene Graph, VQA
- Explanation

Sensors

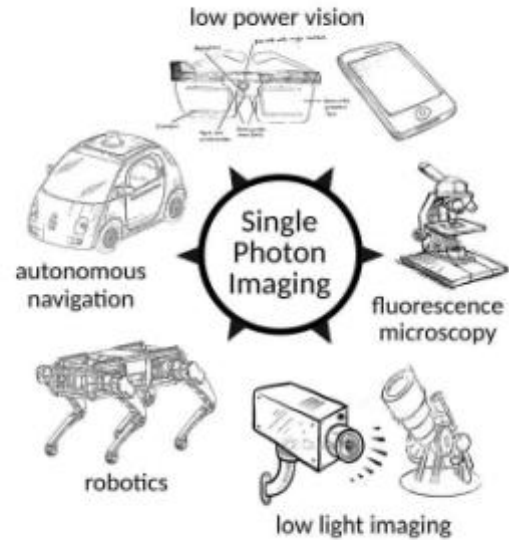
Asynchronous Single-Photon 3D Imaging

Anant Gupt, Atul Ingle, Mohit Gupta
University of Wisconsin-Madison

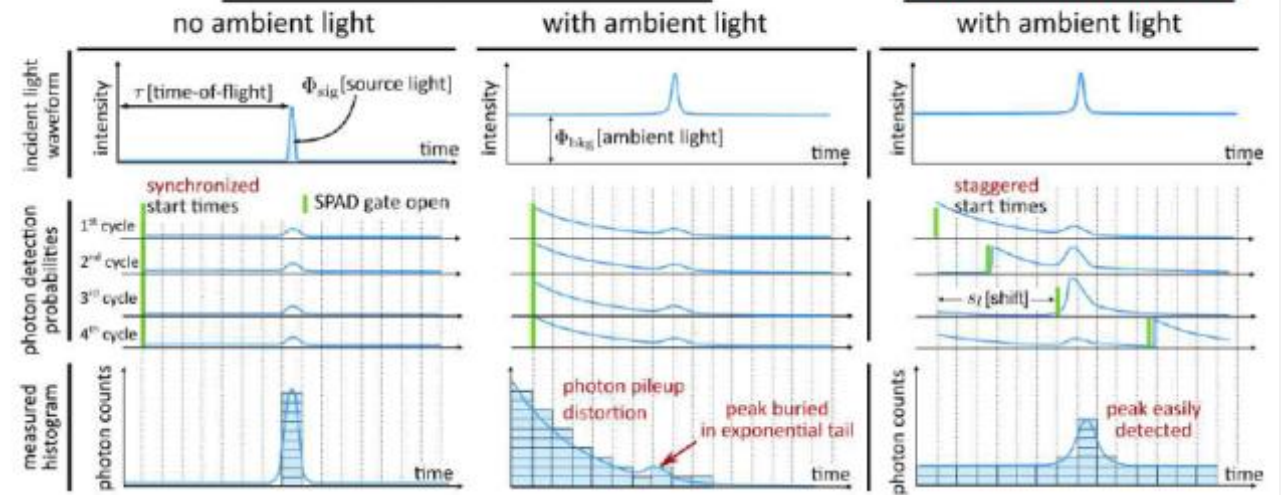
(a) Single-Photon Cameras



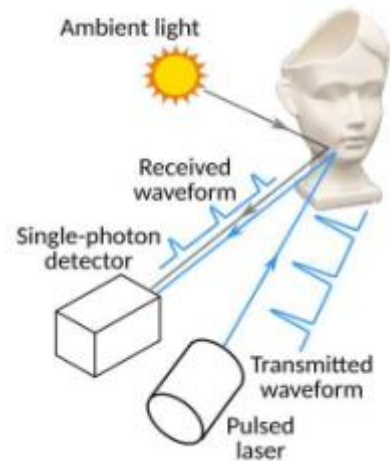
(b) Potential Applications



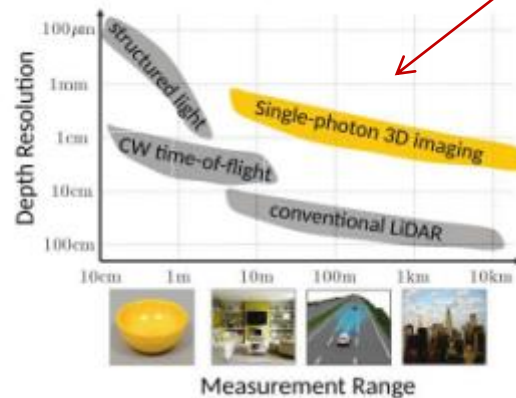
Synchronous Acquisition [Conventional]



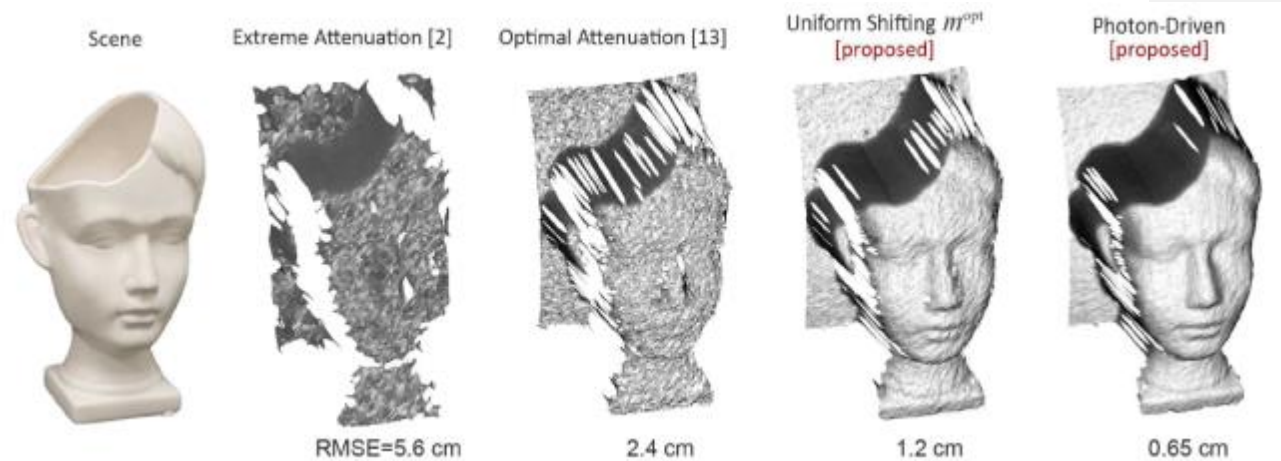
(c) Single-Photon 3D Imaging



(d) Landscape of Active 3D Imaging



Резкое увеличение дальности и качества 3D зрения



Learning Single Camera Depth Estimation using Dual-Pixels

Rahul Garg Neal Wadhwa Sameer Ansari Jonathan T. Barron
Google Research

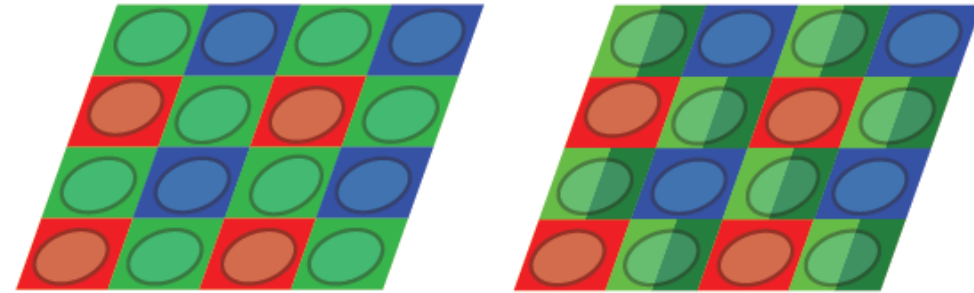
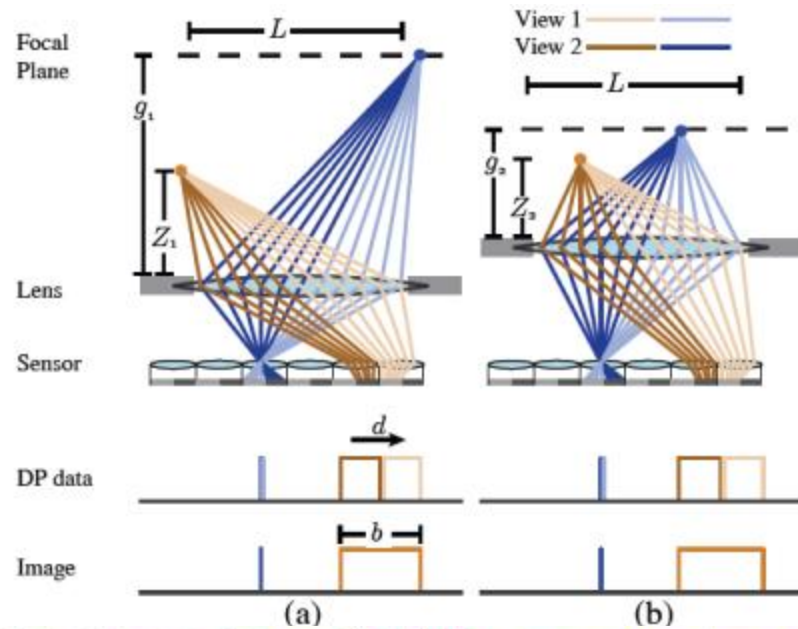
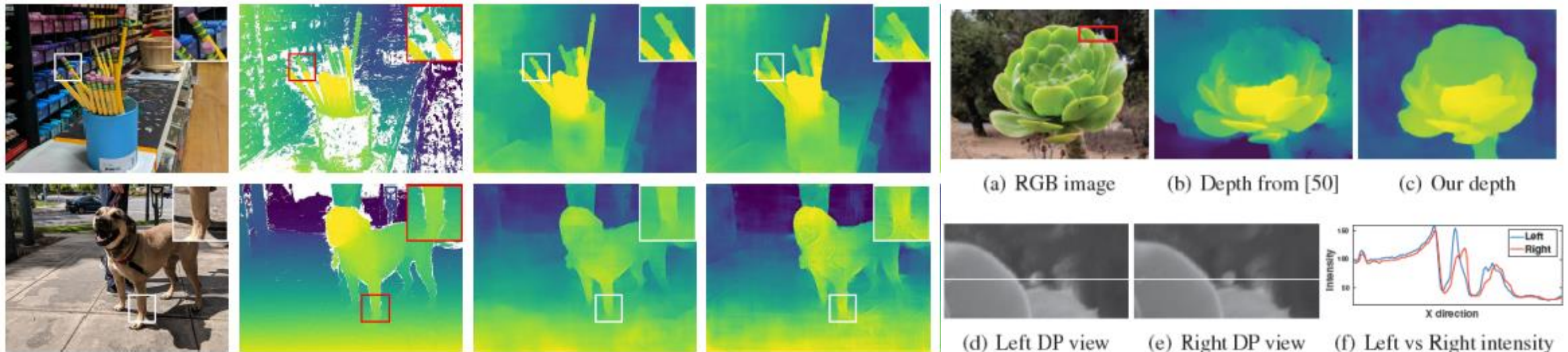


Figure 2. A modern Bayer sensor consists of interleaved red, green, and blue pixels underneath a microlens array. (a). In dual-pixel sensors, the green pixel under each microlens is split in half (b), resulting in two green images that act as a narrow-baseline stereo camera, much like a reduced light field camera.



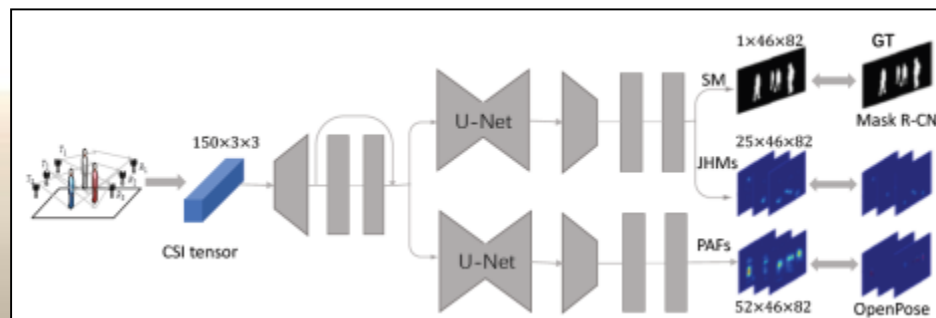
Person-in-WiFi: Fine-grained Person Perception using WiFi

Fei Wang, Sanping Zhou, Stanislav Panev, Jinsong Han, Dong Huang

The Robotics Institute, Carnegie Mellon University



3 Transmitter Antennas



3 Receiver Antennas
Camera for Annotation

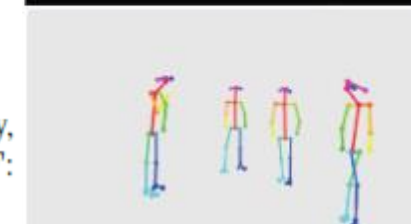
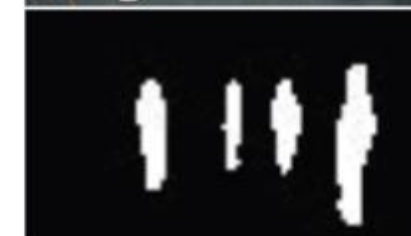
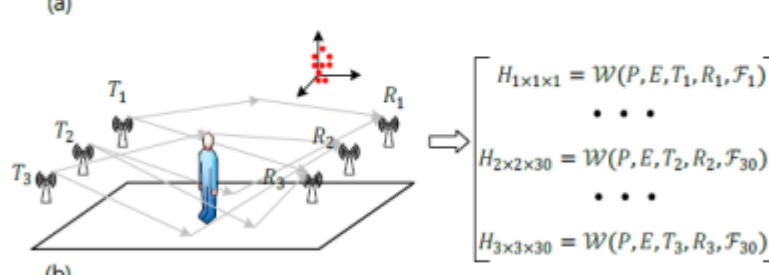
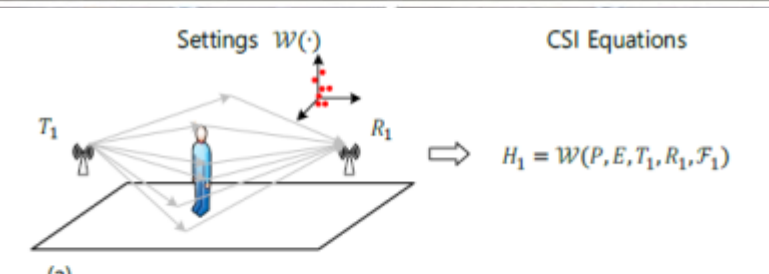
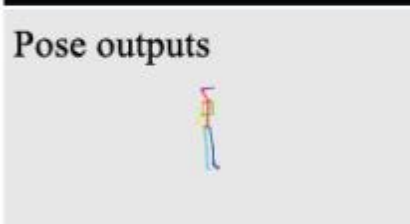


Figure 3. WiFi sensing system. H : CSI sample, P : person body, E : environment, T : transmitter antenna, R : receiver antenna, F : EM frequency.

Making the Invisible Visible: Action Recognition Through Walls and Occlusions

Tianhong Li, Lijie Fan, Mingmin Zhao, Yingcheng Liu, Dina Katabi

MIT CSAIL

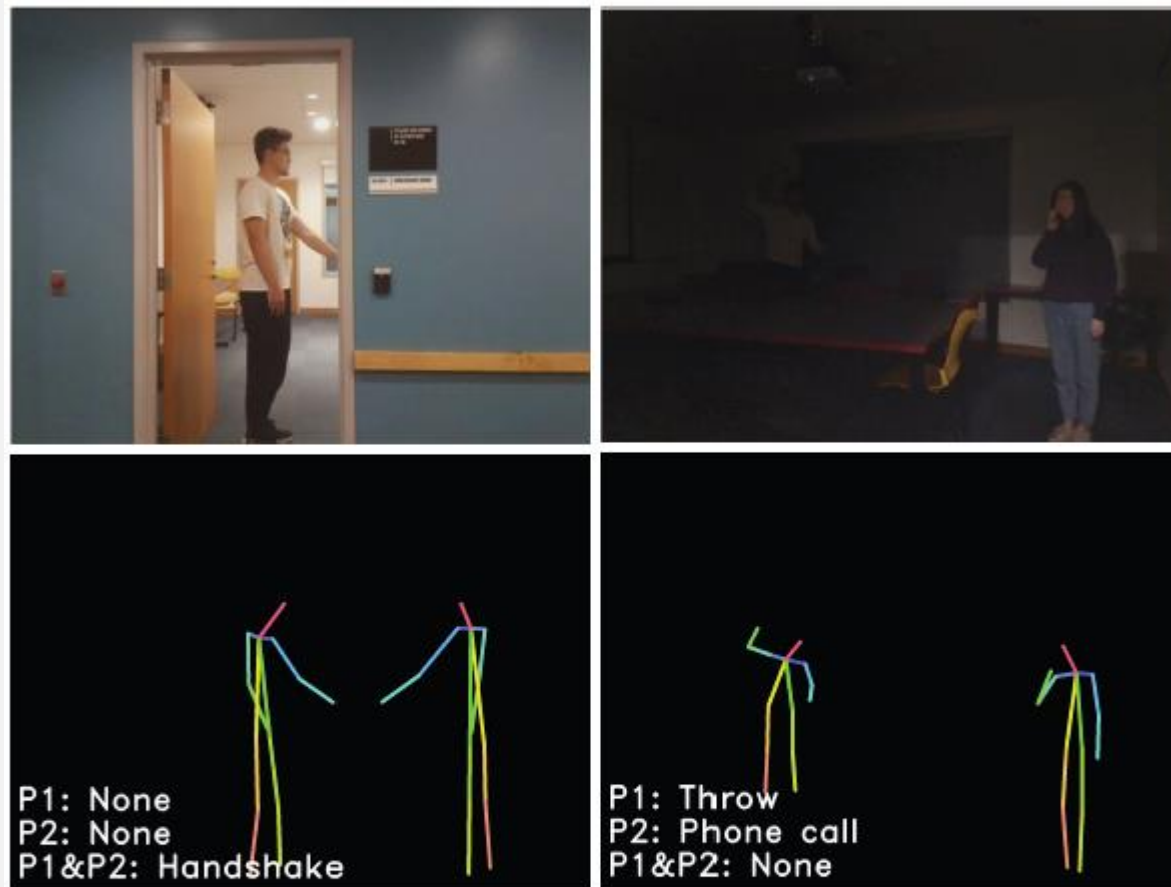
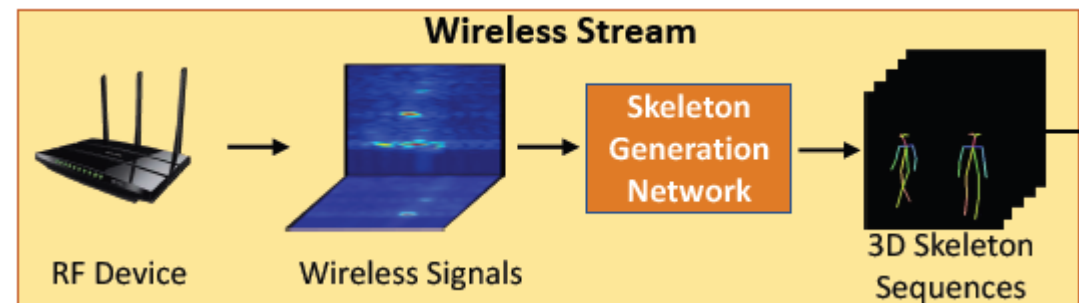
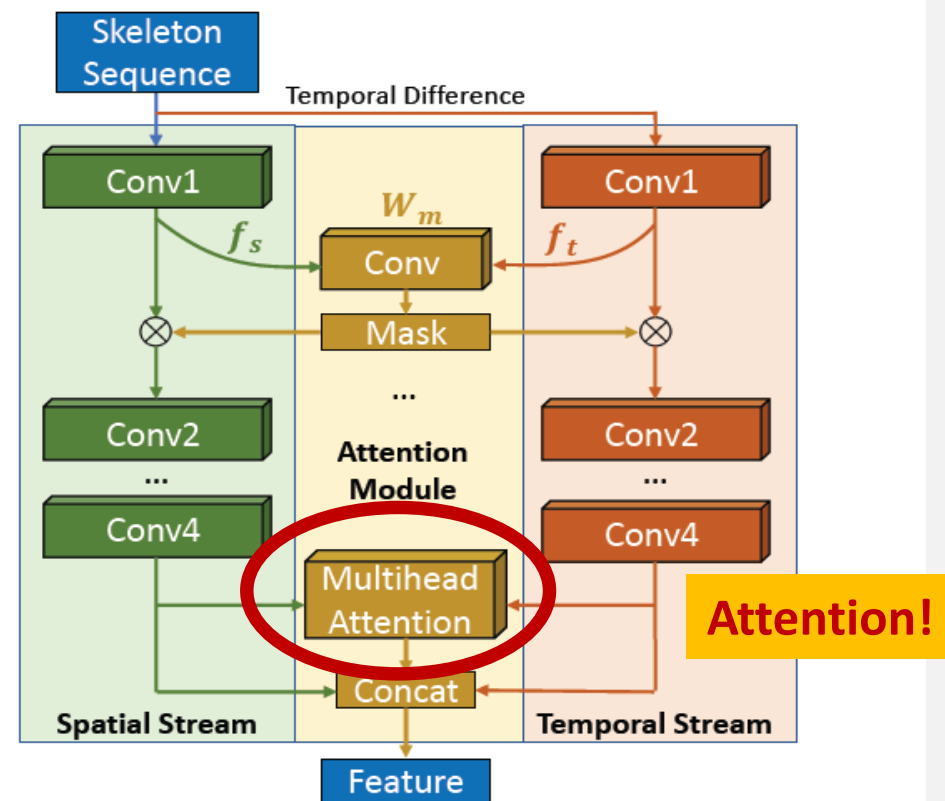
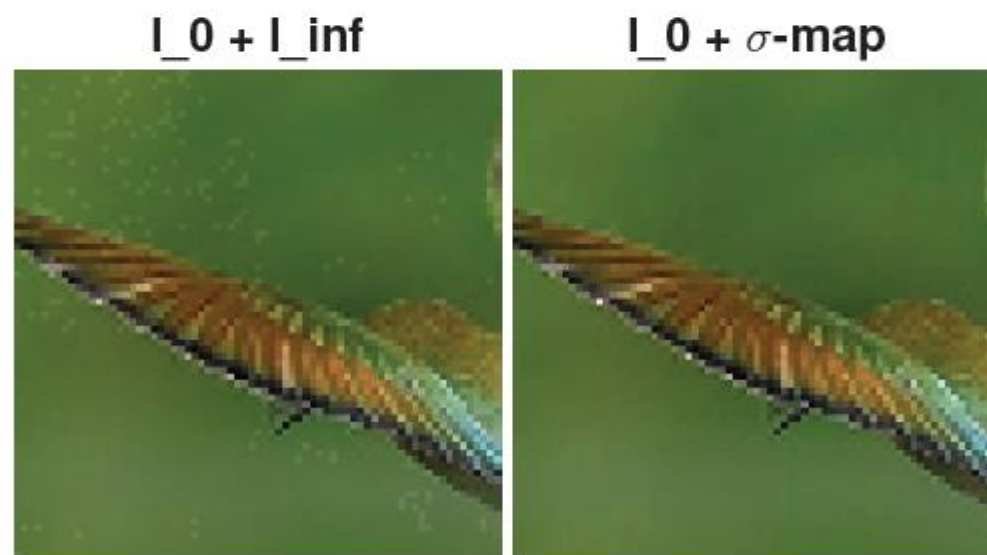
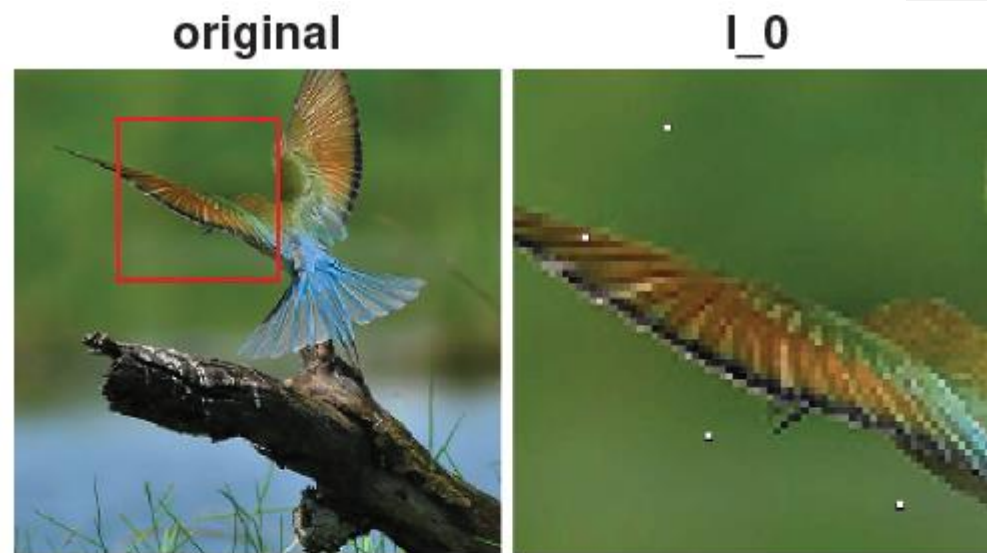
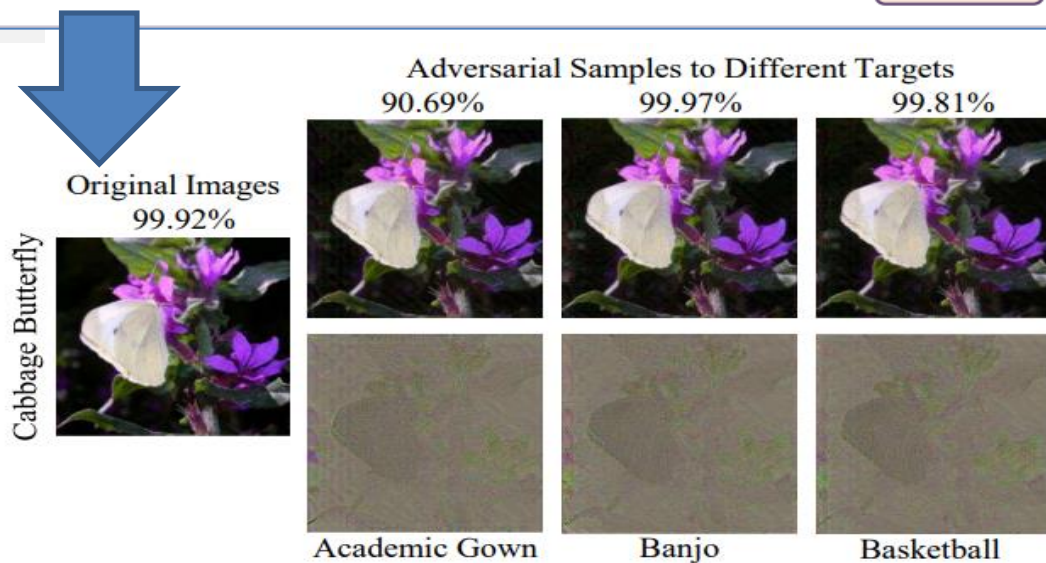
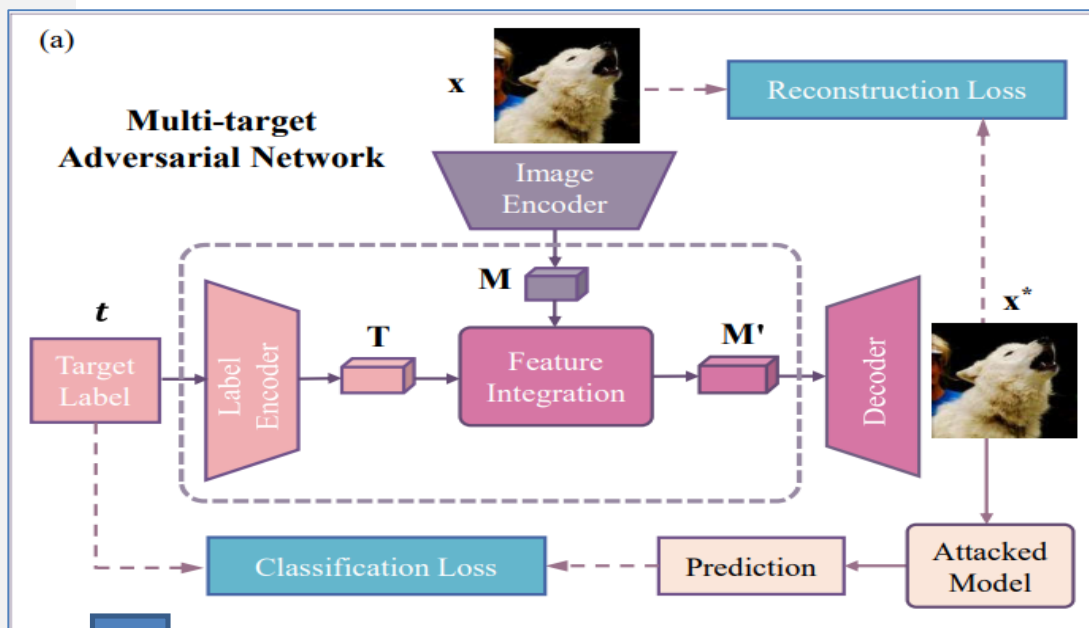


Figure 1: The figure shows two test cases of our system. On the left, two people are shaking hands, while one of them is behind the wall. On the right, a person is hiding behind the wall and throwing an object at another person who is making a phone call. The bottom row shows both the skeletal representation generated by our model and the action prediction.



Adversarial Attack

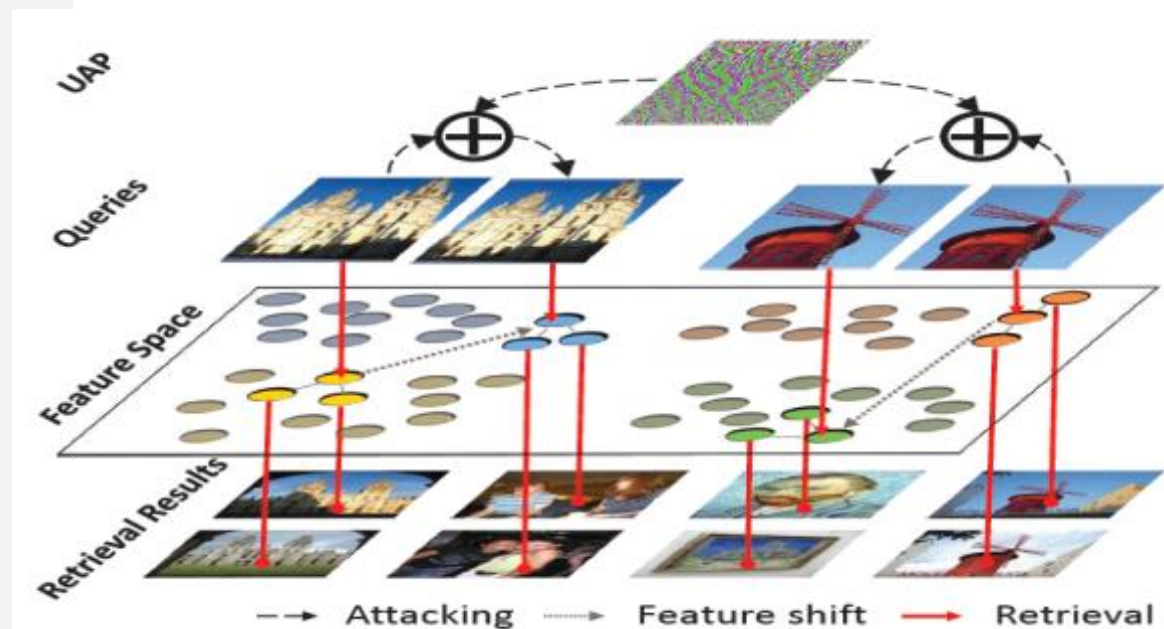
Атаки на распознающие нейронные сети



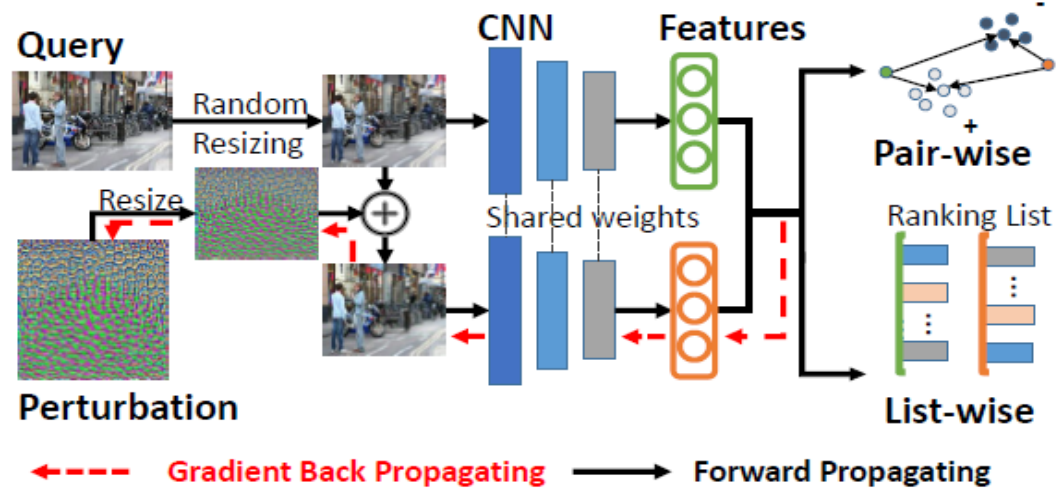
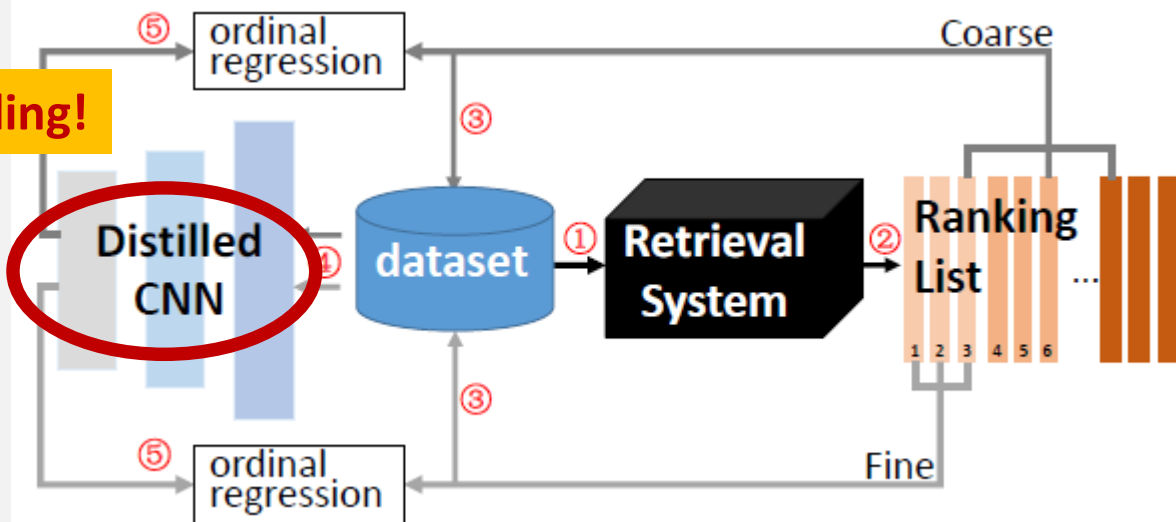
Once a MAN: Towards Multi-Target Attack via Learning Multi-Target Adversarial Network Once, ICCV-2019, Jiangfan Han, Xiaoyi Dong, Ruimao Zhang, Dongdong Chen, Weiming Zhang, Nenghai Yu, Ping Luo, Xiaogang Wang

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

Атаки на поисковые нейронные сети

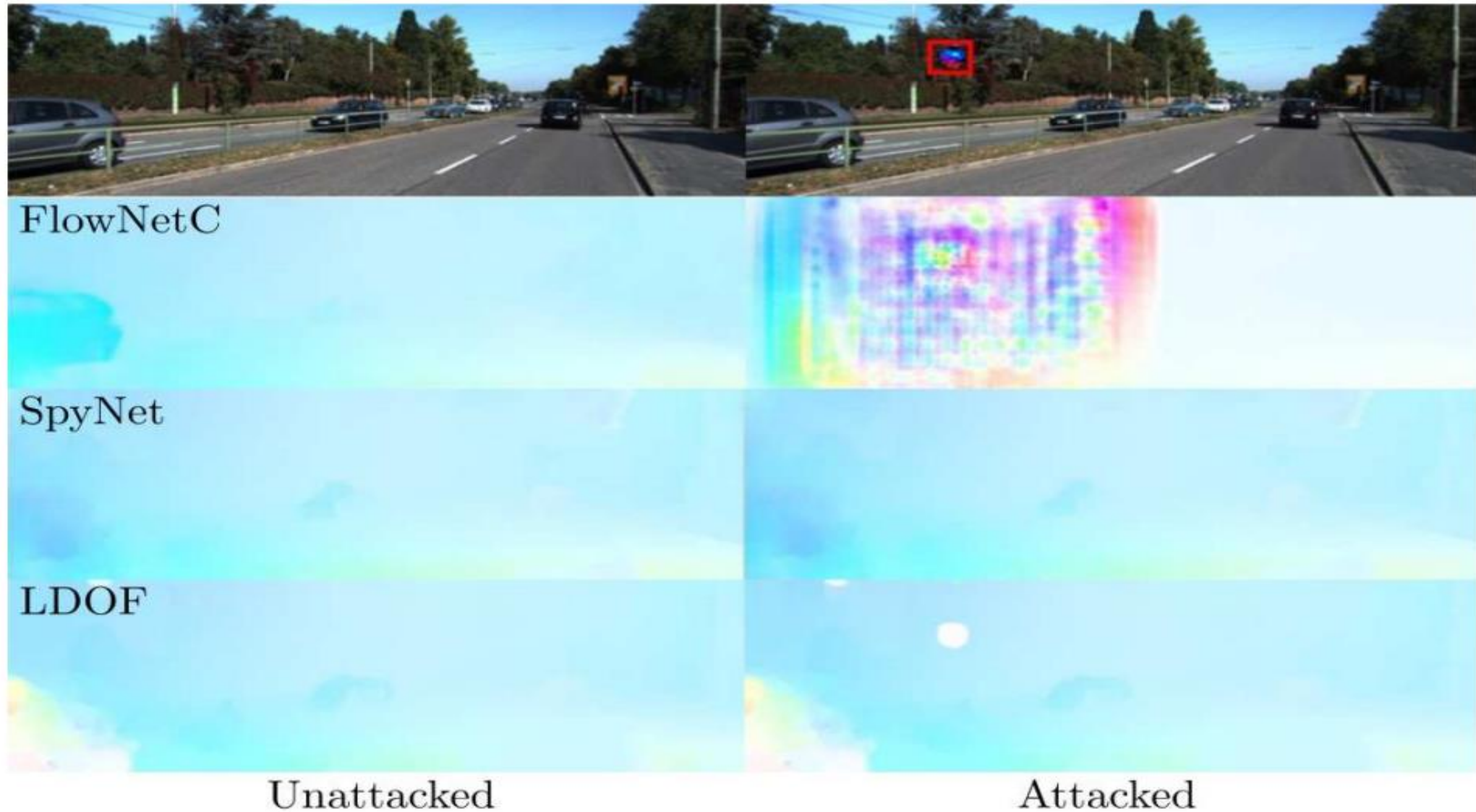


Distilling!

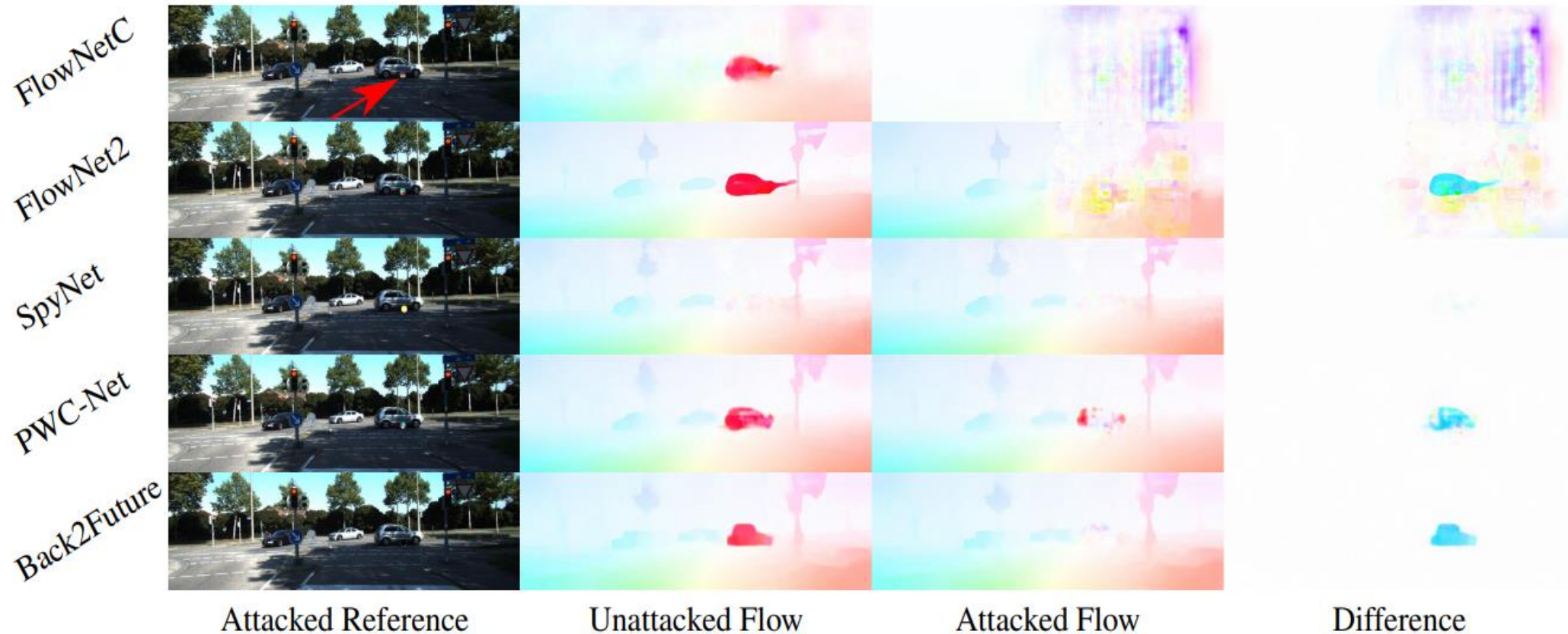


Universal Perturbation Attack Against Image Retrieval
 Jie Li, Rongrong Ji, Hong Liu, Xiaopeng Hong, Yue Gao, Qi Tian

Атаки на оптический поток (в автономном вождении)



Атаки на оптический поток (в автономном вождении)



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

Геометрия атак и основные идеи защиты от атак

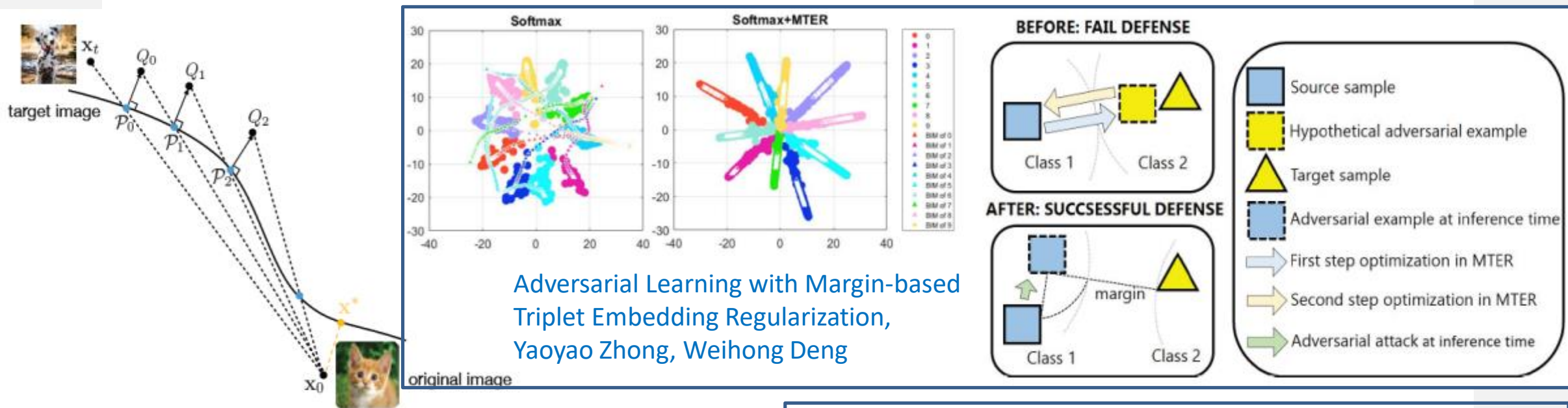









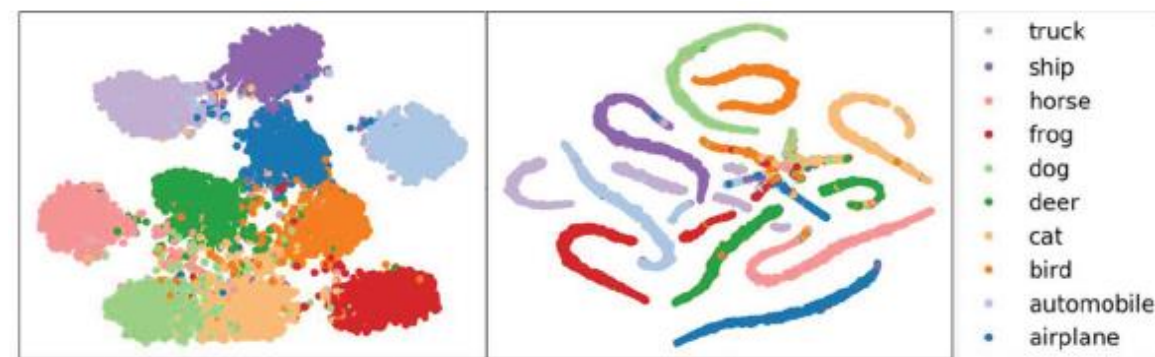
Figure 2: An illustration of qFool in targeted attacks.

original class	target class					
“school bus”	“French loaf”			“golden retriever”		
						
original distance	0.159	0.180	0.182	0.122	0.192	0.211
#query	99993	99127	99443	99514	99998	99135
adversarial distance	5.27e-4	7.90e-4	5.79e-4	3.09e-4	7.07e-4	3.08e-4

A Geometry-Inspired Decision-Based Attack
Yujia Liu, Seyed-Mohsen Moosavi-Dezfooli, Pascal Frossard

Improving Adversarial Robustness via **Guided Complement Entropy**, Hao-Yun Chen, Jhao-Hong Liang, Shih-Chieh Chang, Jia-Yu Pan, Yu-Ting Chen, WeiWei, and Da-Cheng Juan

$$-\frac{1}{N} \sum_{i=1}^N y_{ig}^{\alpha} \cdot \frac{1}{\log(K-1)} \sum_{j=1, j \neq g}^K \left(\frac{\hat{y}_{ij}}{1 - \hat{y}_{ig}} \right) \log \left(\frac{\hat{y}_{ij}}{1 - \hat{y}_{ig}} \right)$$



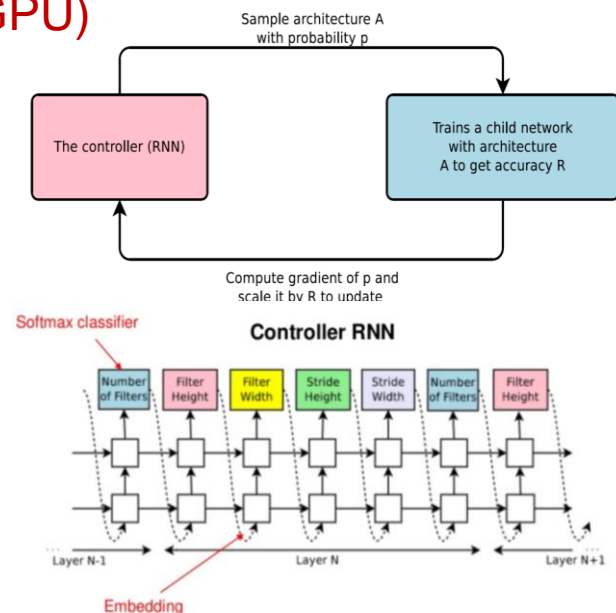
**AutoML,
Neural Architecture Search (NAS)**

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

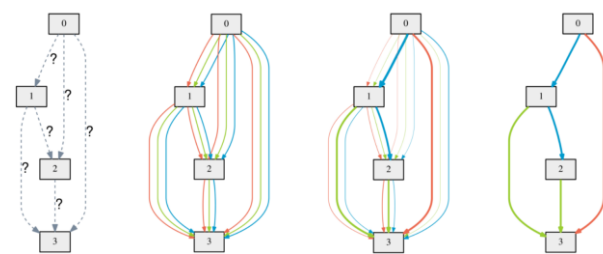
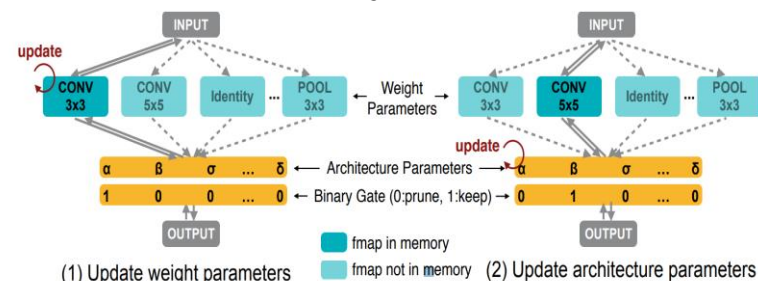
Barret Zoph, Vijay Vasudevan, Jonathon Shlens, Quoc V. Le CVPR 2018

*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



*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

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

2022+:
Перспектива полной автоматизации процессов обучения

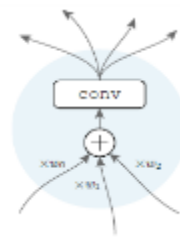
Exploring Randomly Wired Neural Networks for Image Recognition

Saining Xie, Alexander Kirillov, Ross Girshick, Kaiming He

Facebook AI Research (FAIR)



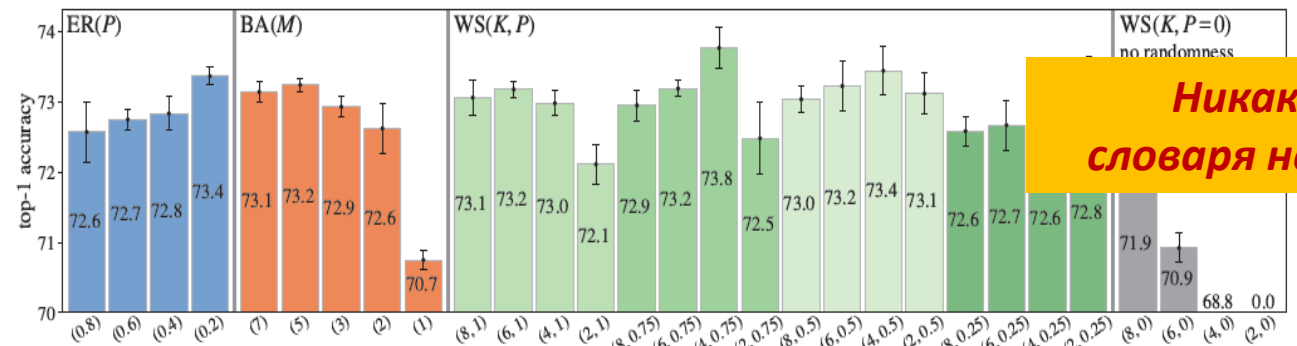
Три случайные архитектуры превзошли ResNet-50!



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

GCN!

Эта работа взрывает все наше понимание NAS!



Никакого словаря нет!

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.

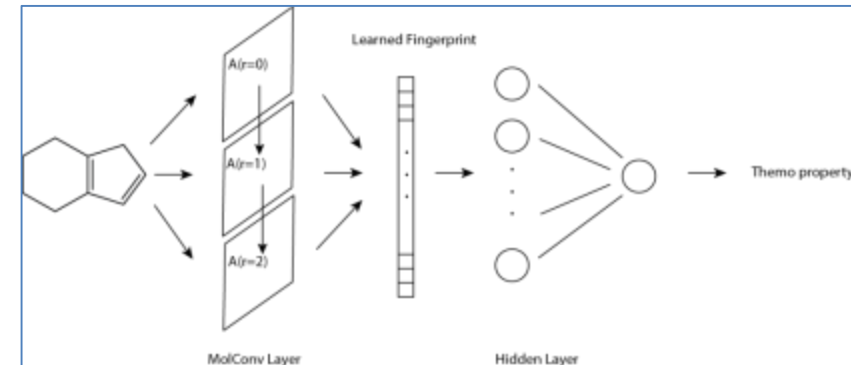
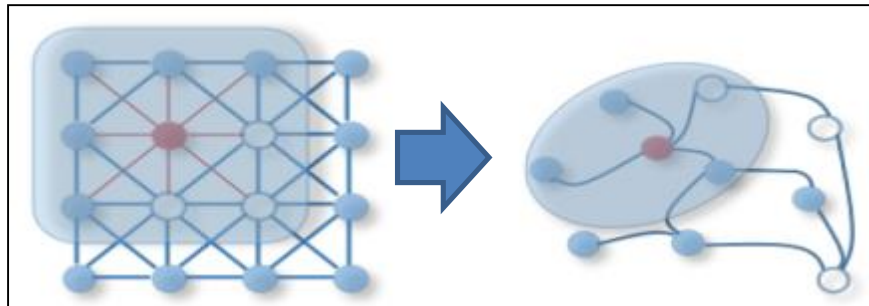
Figure 3. Comparison on random graph generators: ER, BA, and WS in the small computation regime. Each bar represents the results of a generator under a parameter setting for P , M , or (K, P) (tagged in x-axis). The results are ImageNet top-1 accuracy, shown as mean and standard deviation (std) over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

**Graph Convolutional
Networks (GCN), Attention,
Matching Network,
Relation Network**

Deep Graph Embedding: глубокие сети на графах



На изображениях глубокие сети «снизу вверх» вычисляют структурные признаки окрестностей разного масштаба



Значит, и на графах нам нужна система «окошек» разного масштаба

Опишем кусочки графа векторами признаков, потом соберем из них признаки больших кусочков, и так – пока не опишем вектором весь граф

Deep Graph Embedding: глубокие сети на графах

Сегодня сети на графах уже не экзотика. Имеется множество доступных реализаций...

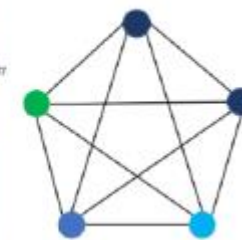
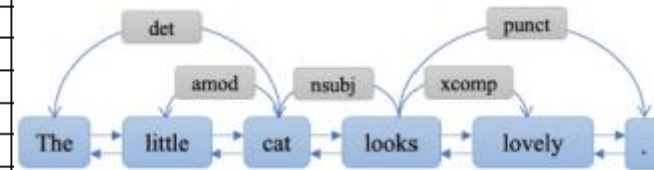
Model	Framework	Github Link
ChebNet (2016) [12]	tensorflow	https://github.com/mdeff/cnn_graph
1stChebNet (2017) [14]	tensorflow	https://github.com/tkipf/gcn
GGNNs (2015) [19]	lua	https://github.com/yujiali/ggnn
SSE (2018) [20]	c	https://github.com/Hanjun-Dai/steady_state_embedding
GraphSage (2017) [25]	tensorflow	https://github.com/williamleif/GraphSAGE
LGCN (2018) [28]	tensorflow	https://github.com/divelab/lgcn/
SplineCNN (2018) [89]	pytorch	https://github.com/rusty1s/pytorch_geometric
GAT (2017) [15]	tensorflow	https://github.com/PetarV-/GAT
GAE (2016) [62]	tensorflow	https://github.com/limaosen0/Variational-Graph-Auto-Encoders
ARGA (2018) [64]	tensorflow	https://github.com/Ruiqi-Hu/ARGA
DNGR (2016) [42]	matlab	https://github.com/ShelsonCao/DNGR
SDNE (2016) [43]	python	https://github.com/suanrong/SDNE
DRNE (2016) [66]	tensorflow	https://github.com/tadpole/DRNE
GraphRNN (2018) [67]	tensorflow	https://github.com/snap-stanford/GraphRNN
DCRNN (2018) [73]	tensorflow	https://github.com/liyaguang/DCRNN
CNN-GCN (2017) [74]	tensorflow	https://github.com/VeritasYin/STGCN_IJCAI-18
ST-GCN (2018) [75]	pytorch	https://github.com/yysijie/st-gcn
Structural RNN (2016) [76]	theano	https://github.com/asheshjain399/RNNexp

A Summary of Open-source Implementations

Deep Graph Embedding: глубокие сети на графах

Area	Application	Algorithm	Deep Learning Model
Text	Text classification	GCN	Graph Convolutional Network
		GAT	Graph Attention Network
		DGCNN	Graph Convolutional Network
		Text GCN	Graph Convolutional Network
		Sentence LSTM	Graph LSTM
	Sequence Labeling (POS, NER)	Sentence LSTM	Graph LSTM
	Sentiment classification	Tree LSTM	Graph LSTM
	Semantic role labeling	Syntactic GCN	Graph Convolutional Network
	Neural machine translation	Syntactic GCN	Graph Convolutional Network
		GGNN	Gated Graph Neural Network
	Relation extraction	Tree LSTM	Graph LSTM
		Graph LSTM	Graph LSTM
		GCN	Graph Convolutional Network
	Event extraction	Syntactic GCN	Graph Convolutional Network
	AMR to text generation	Sentence LSTM	Graph LSTM
Multi-hop reading comprehension	GGNN	Gated Graph Neural Network	
	Sentence LSTM	Graph LSTM	
	RN	MLP	
Relational reasoning	Recurrent RN	Recurrent Neural Network	
	IN	Graph Neural Network	
Image	Social Relationship Understanding	GRM	Gated Graph Neural Network
	Image classification	GCN	Graph Convolutional Network
		GGNN	Gated Graph Neural Network
		ADGPM	Graph Convolutional Network
		GSNN	Gated Graph Neural Network
	Visual Question Answering	GGNN	Gated Graph Neural Network
	Object Detection	RN	Graph Attention Network
	Interaction Detection	GPNN	Graph Neural Network
		Structural-RNN	Graph Neural Network
	Region Classification	GCNN	Graph CNN
Semantic Segmentation	Graph LSTM	Graph LSTM	
	GGNN	Gated Graph Neural Network	
	DGCNN	Graph CNN	
	3DGNN	Graph Neural Network	

Сегодня сети на графах уже не экзотика. Имеется множество доступных реализаций и практических приложений

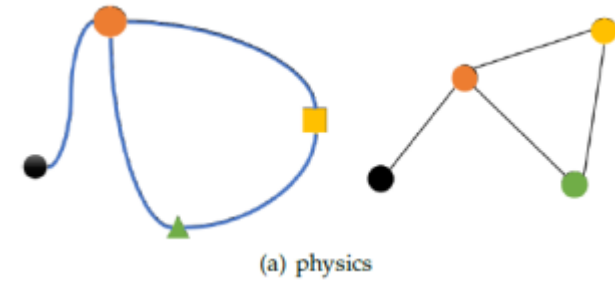


Applications of graph neural networks

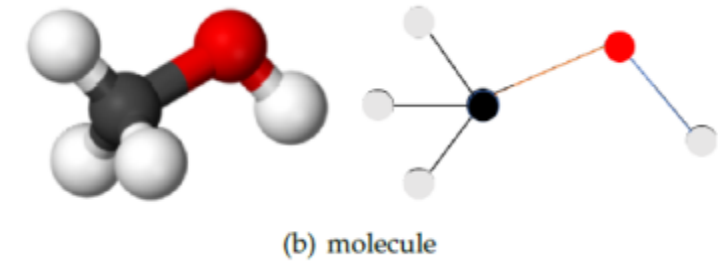
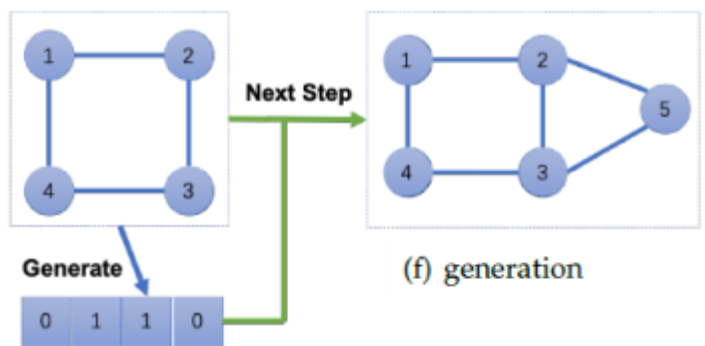
Graph Neural Networks: A Review of Methods and Applications (Jie Zhou et.al., 2019)

Deep Graph Embedding: глубокие сети на графах

Area	Application	Algorithm	Deep Learning Model	
Science	Physics Systems	IN	Graph Neural Network	
		VIN	Graph Neural Network	
		GN	Graph Networks	
	Molecular Fingerprints	NGF	Graph Convolutional Network	
		GCN	Graph Convolutional Network	
		Protein Interface Prediction	GCN	Graph Convolutional Network
		Side Effects Prediction	Decagon	Graph Convolutional Network
Disease Classification	PPIN	Graph Convolutional Network		
Knowledge Graph	KB Completion	GNN	Graph Neural Network	
	KG Alignment	GCN	Graph Convolutional Network	
Combinatorial Optimization		structure2vec	Graph Convolutional Network	
		GNN	Graph Neural Network	
		GCN	Graph Convolutional Network	
		AM	Graph Attention Network	
Graph Generation		NetGAN	Long short-term memory	
		GraphRNN	Recurrent Neural Network	
		Regularizing VAE	Variational Autoencoder	
		GCPN	Graph Convolutional Network	
		MolGAN	Relational-GCN	



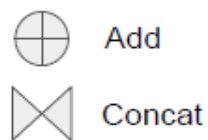
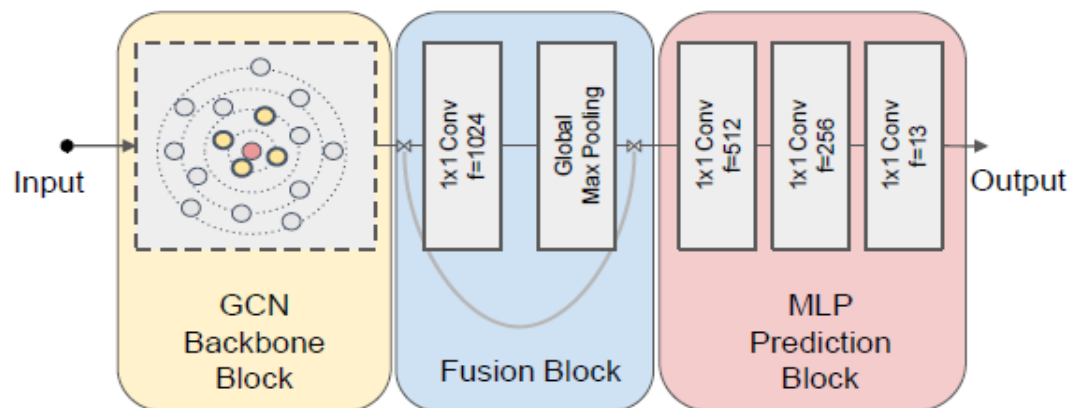
Сегодня сети на графах уже не экзотика. Имеется множество доступных реализаций и практических приложений



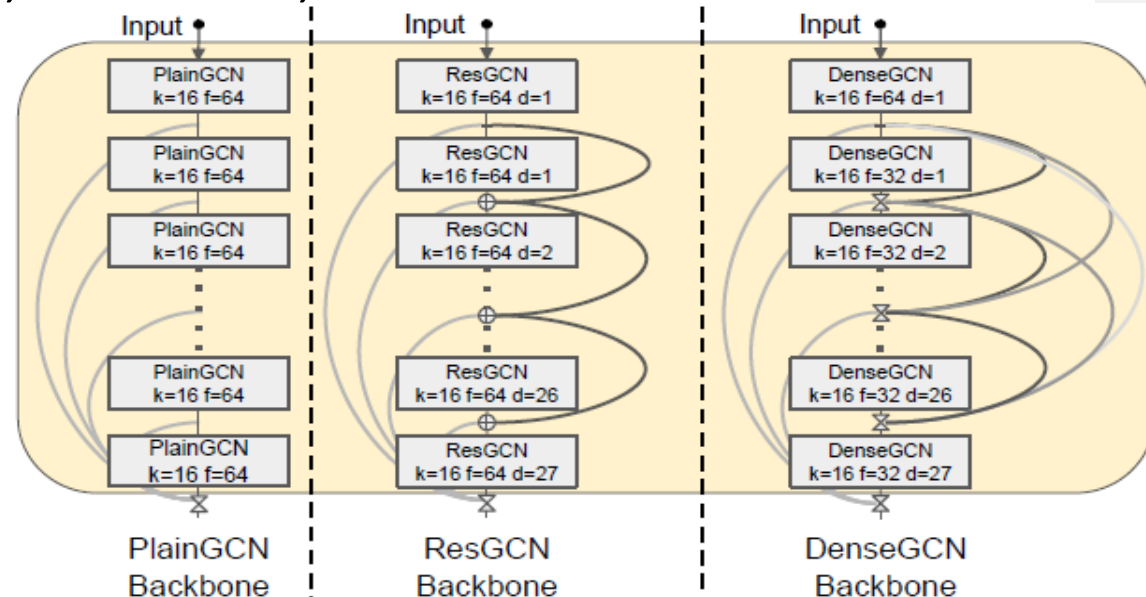
Applications of graph neural networks

DeepGCNs: Can GCNs Go as Deep as CNNs?

Guohao Li, Matthias Muller, Ali Thabet, Bernard Ghanem

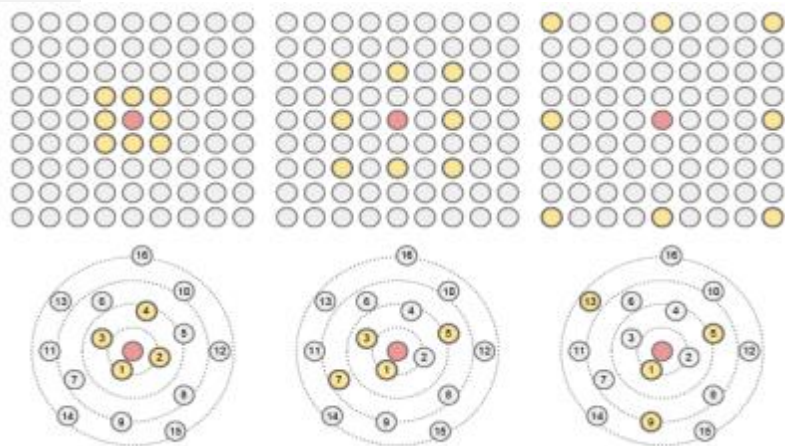


k = # of nearest neighbors
 f = # of filters or hidden units
 d = dilation rate

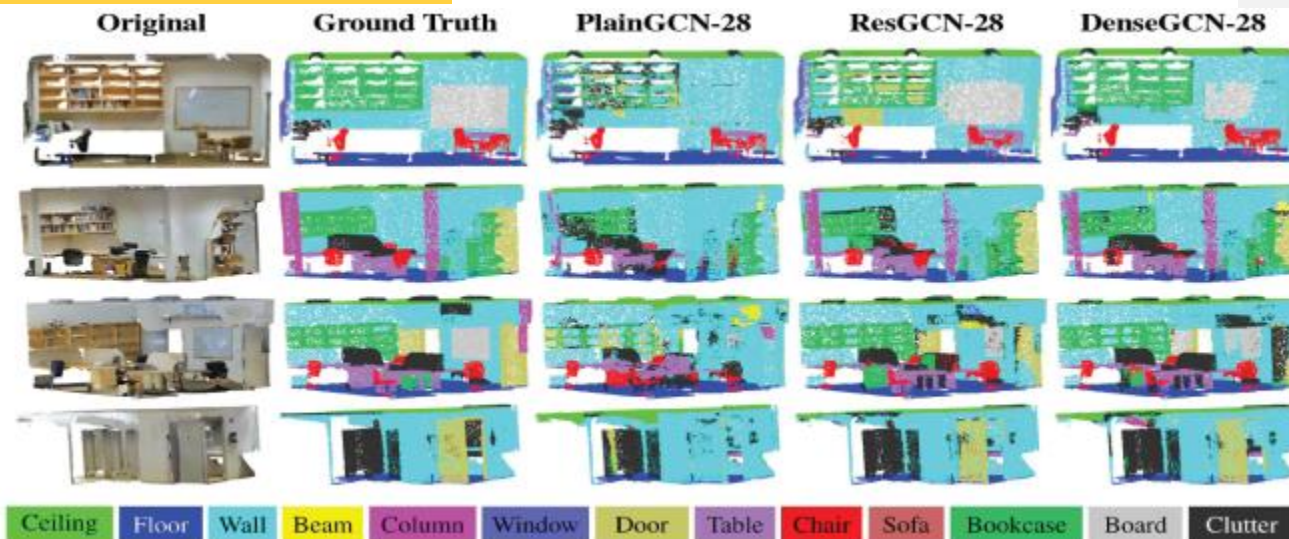


Архитектуры CNN переносимы и эффективны в GCN

Резидуальные связи

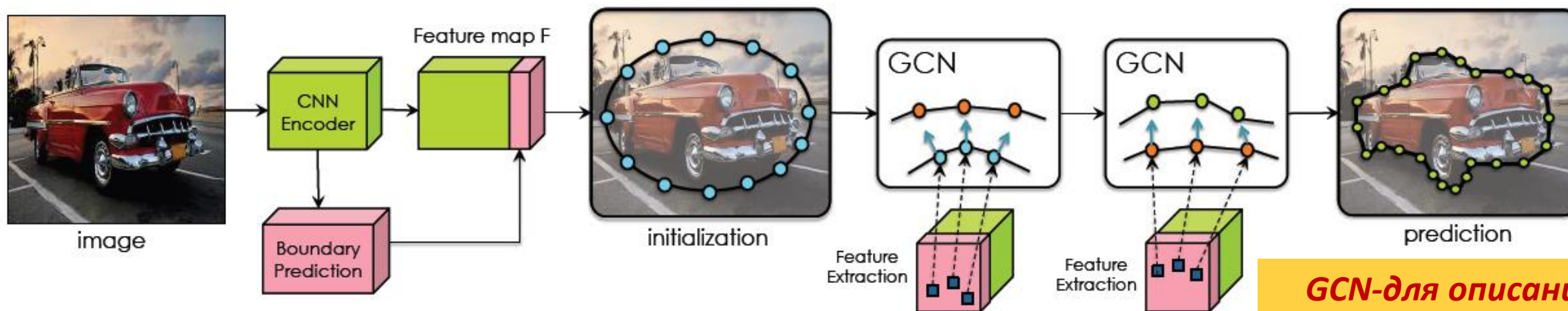


Дилатация ядер свертки



2D Shape Representation: Fast Interactive Object Annotation with Curve-GCN

Huan Ling, Jun Gao, Amlan Kar, Wenzheng Chen, Sanja Fidler



*GCN-для описания 2D форм
идея «активного контура»*



Figure 5: Automatic Mode on Cityscapes. The input to our model are bounding boxes for objects.

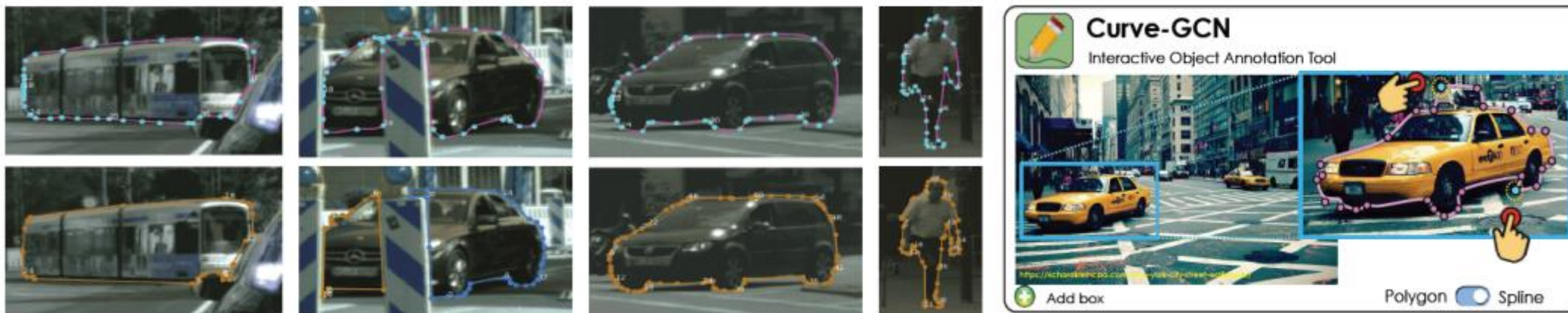


Figure 6: Automatic mode on Cityscapes. We show results for individual instances. (top) Spline-GCN, (bottom) ground-truth. We can observe that our model fits object boundaries accurately, and surprisingly finds a way to “cheat” in order to annotate multi-component instances.

3D Shape: Learning to Predict 3D Objects with an Interpolation-based Differentiable Renderer

Wenzheng Chen, Jun Gao, Huan Ling, Edward J. Smith, Jaakko Lehtinen, Alec Jacobson, Sanja Fidler

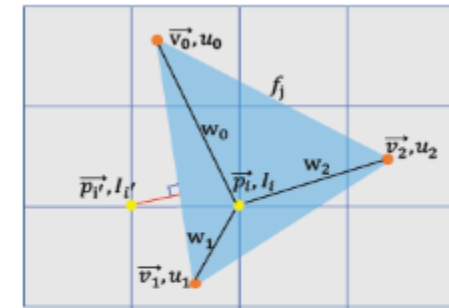
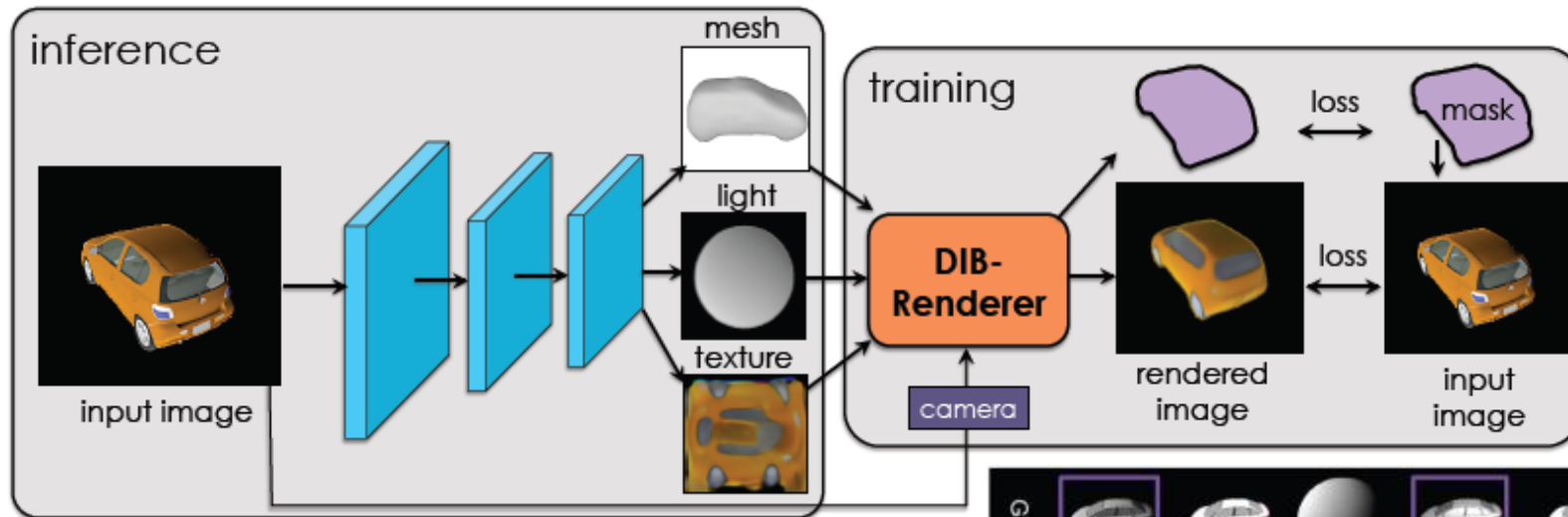
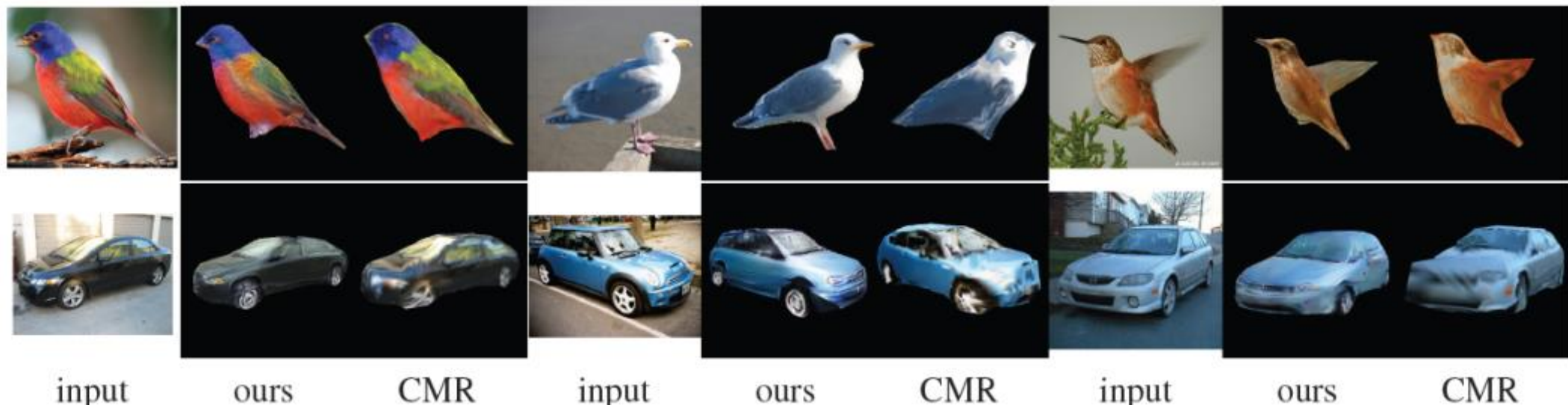
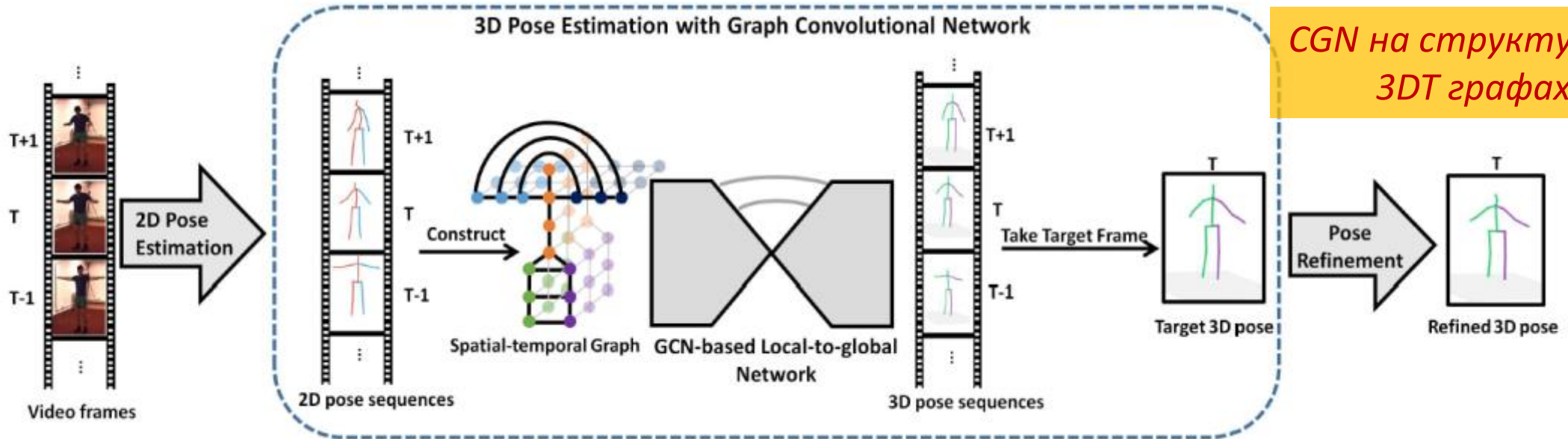
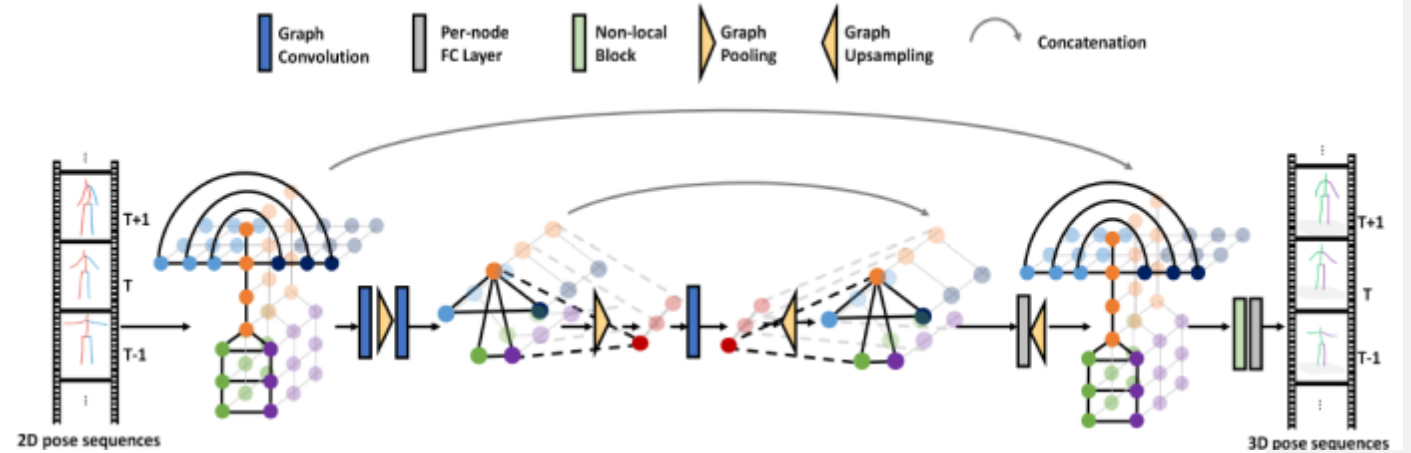
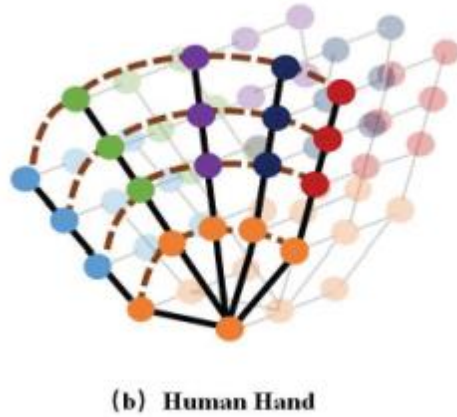
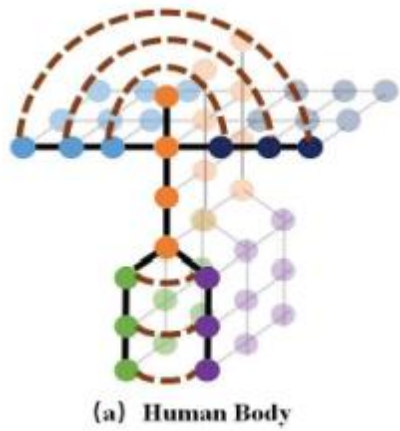


Figure 1: Illustration of our Differentiable Rasterization.

Те же GCN-«активные контуры» только в 3D (деформация сетки-сферы вместо окружности)

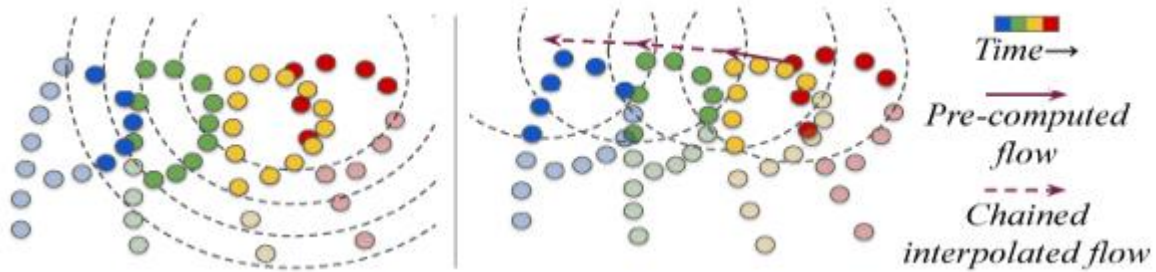
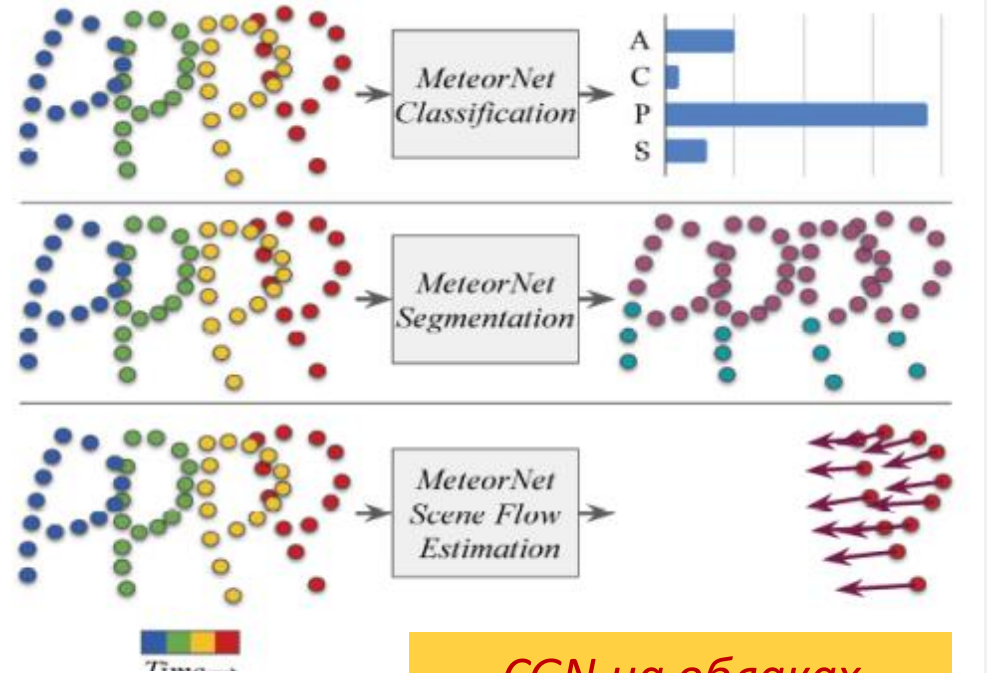
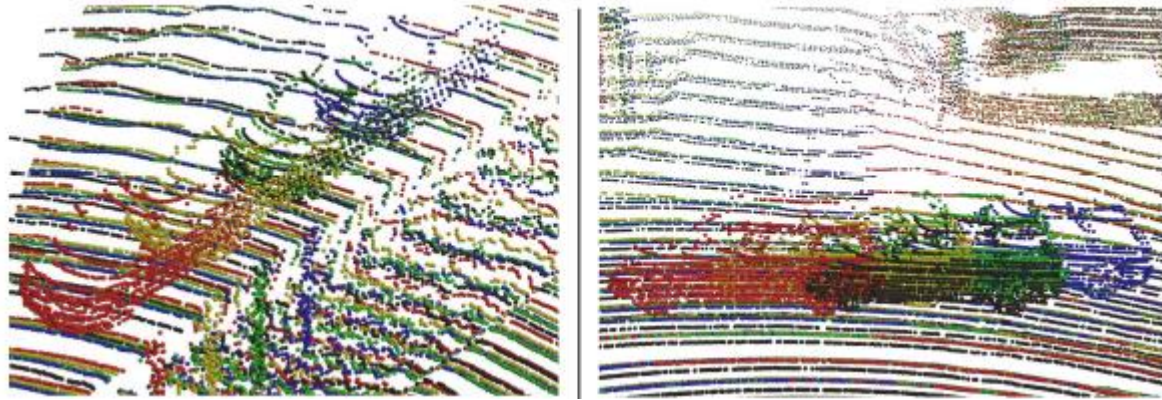


Сети на графах для описания **трехмерных динамических сцен**

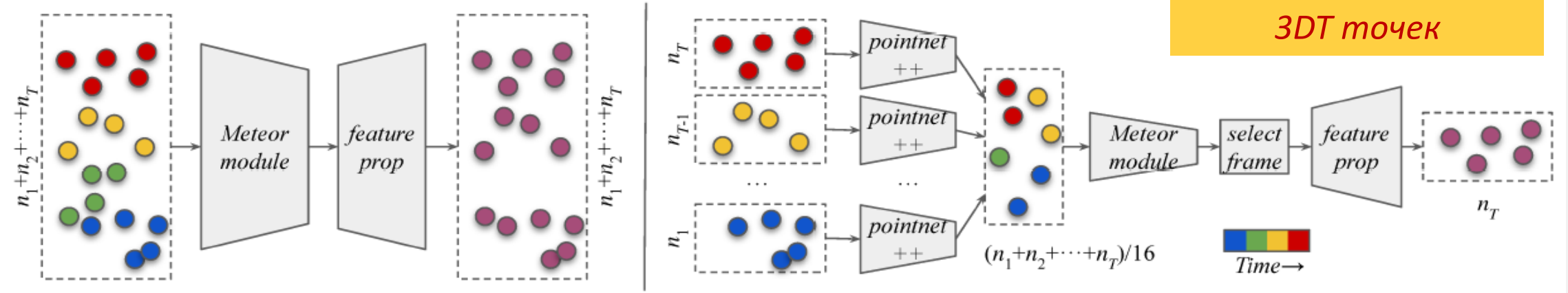


CGN на структурных 3DT графах

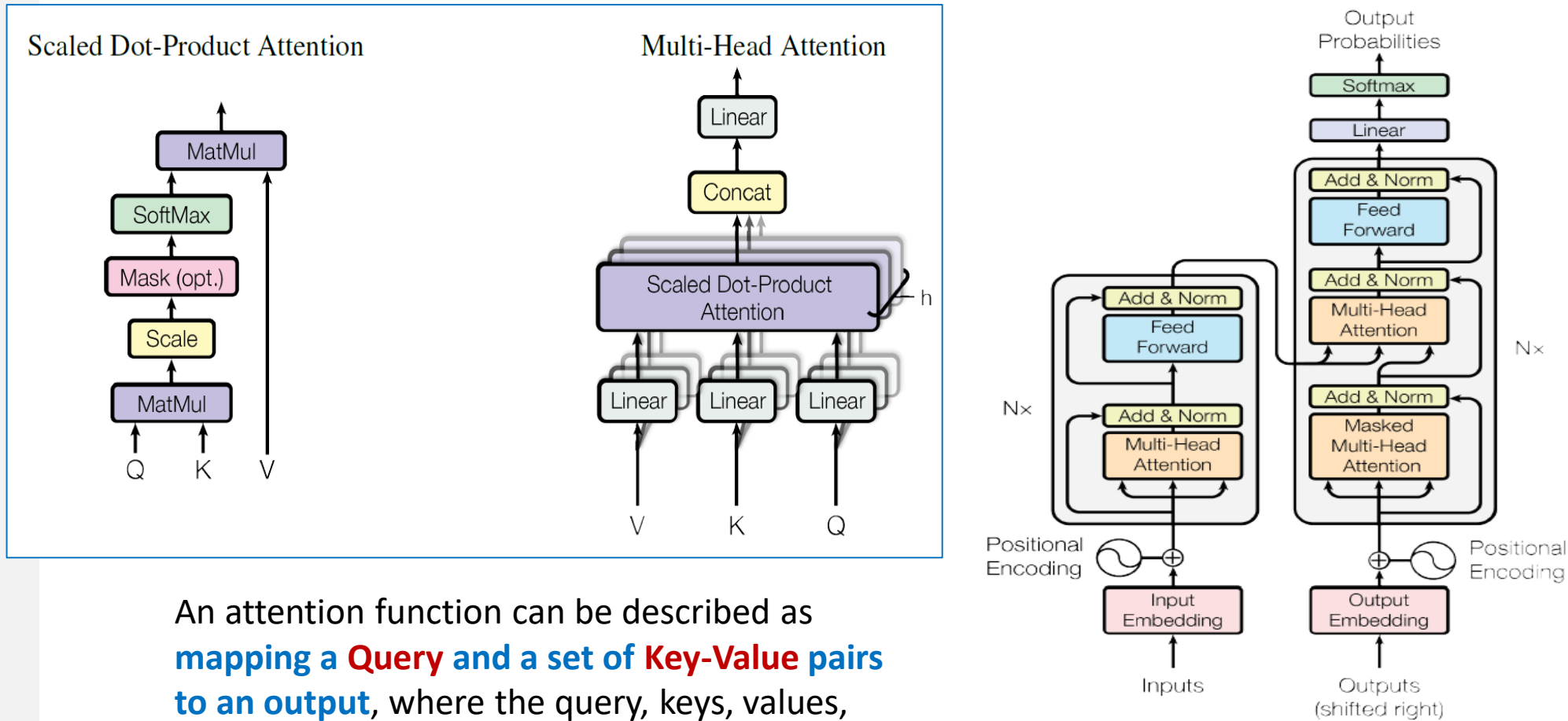
Сети на графах для описания **трехмерных динамических сцен**



**CGN на облаках
3DT точек**



Graph Attention Networks



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.**

Figure 1: The Transformer - model architecture.

Graph Attention Networks

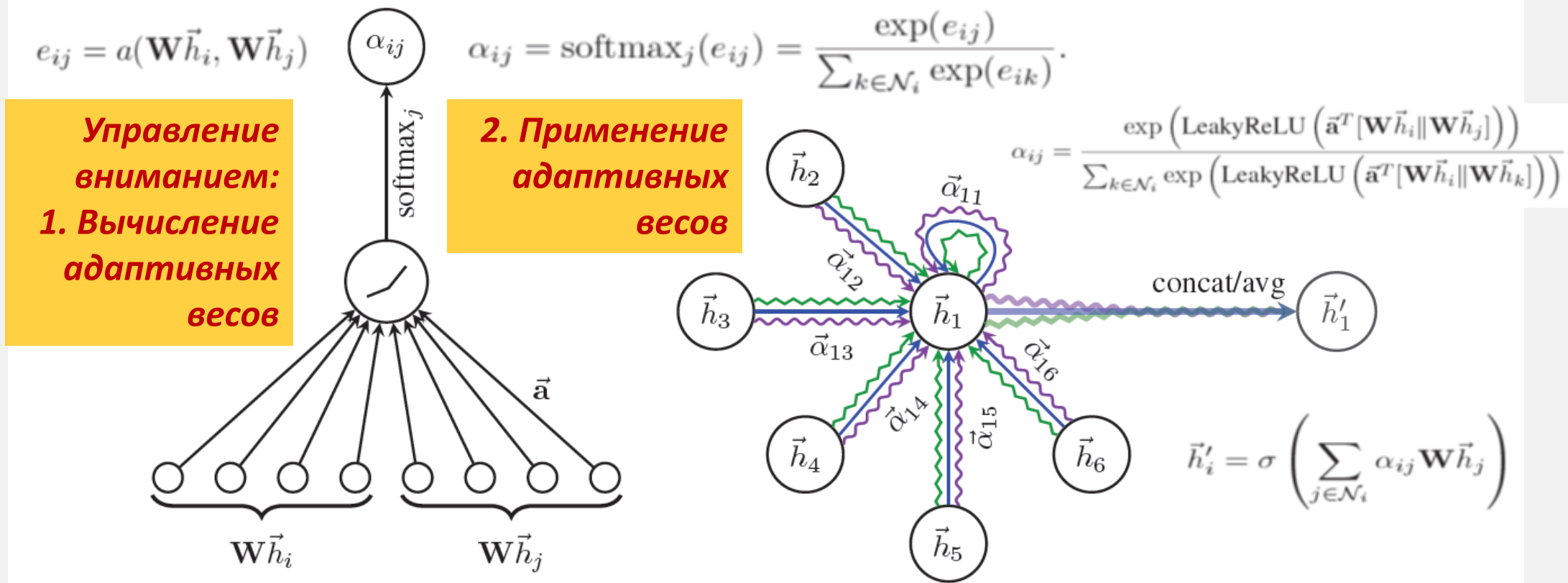


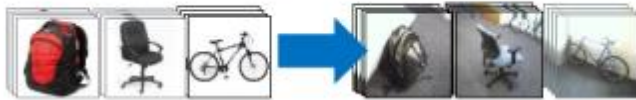
Figure 1: **Left:** The attention mechanism $a(\mathbf{W}\vec{h}_i, \mathbf{W}\vec{h}_j)$ employed by our model, parametrized by a weight vector $\vec{\mathbf{a}} \in \mathbb{R}^{2F'}$, applying a LeakyReLU activation. **Right:** An illustration of multi-head attention (with $K = 3$ heads) by node 1 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 .

**Domain Adaptation, Generative
Adversarial Networks (GAN),
Realistic Data Synthesis...**

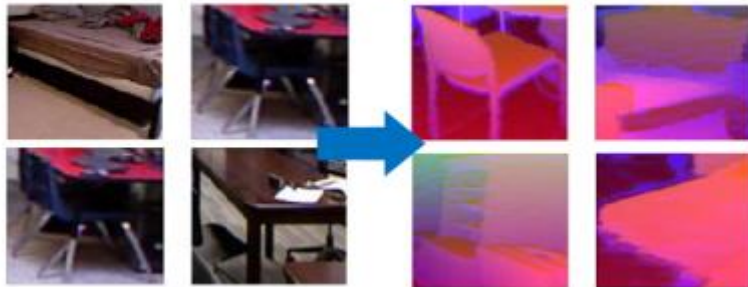
Генеративные конкурирующие сети

Перенос обучения в новую область применения
(Domain Transfer Learning)

From dataset to dataset



From RGB to depth



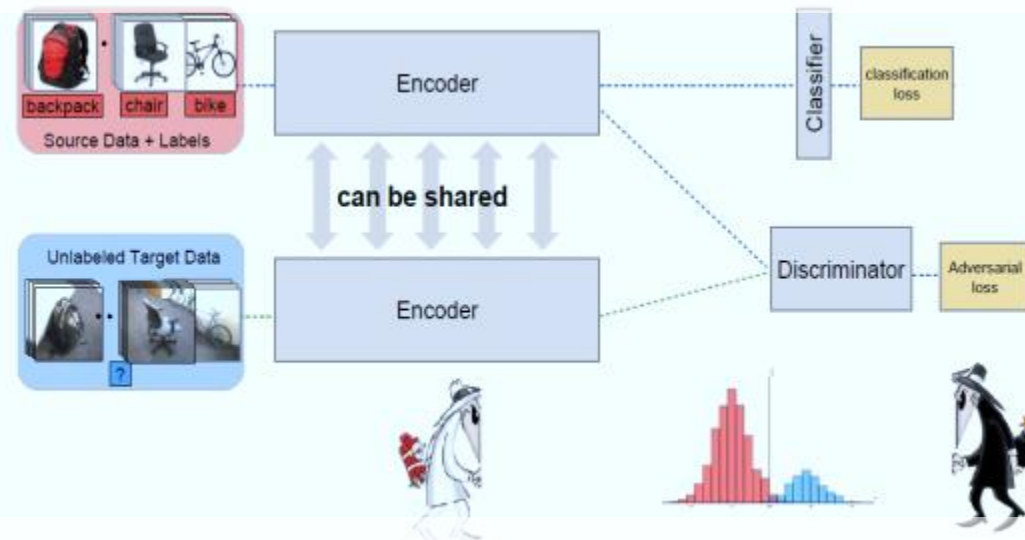
From CAD models to real images



Adversarial networks
Борющиеся интеллектуальные агенты (ГНС)



Генератор создает визуальные образы,
стараясь обмануть Дискриминатор...



....Дискриминатор старается отличить
фантазии Генератора от реальности

Generative Adversarial Networks (GANs)



zebra → horse



apple → orange



summer → winter



horse → zebra



orange → apple



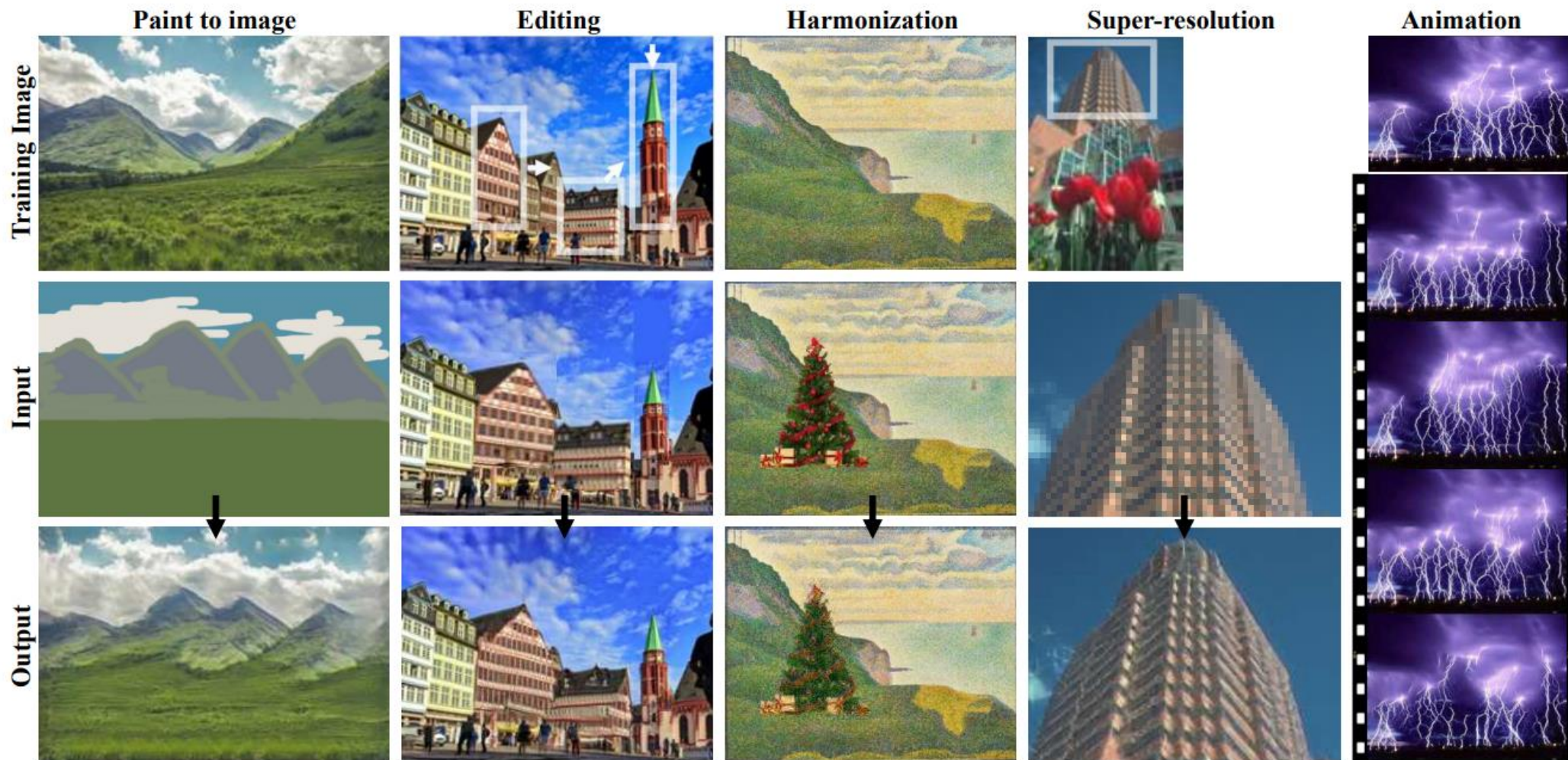
winter → summer

**GAN – сеть,
обладающая
воображением!**



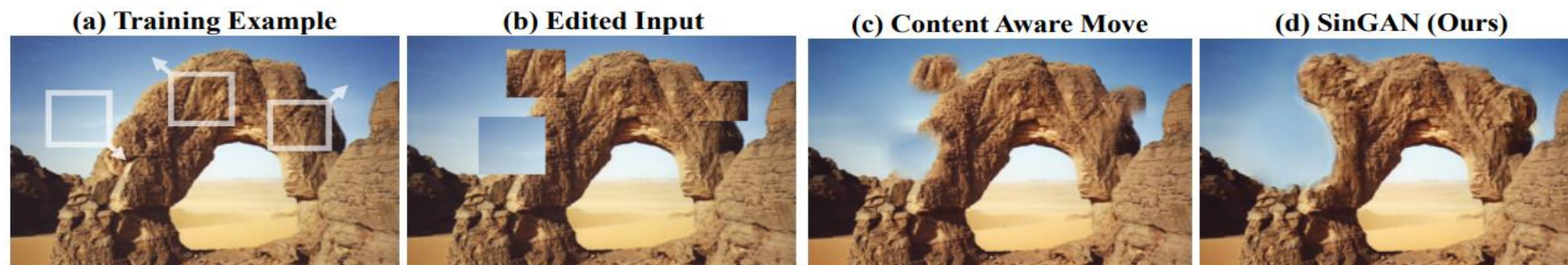
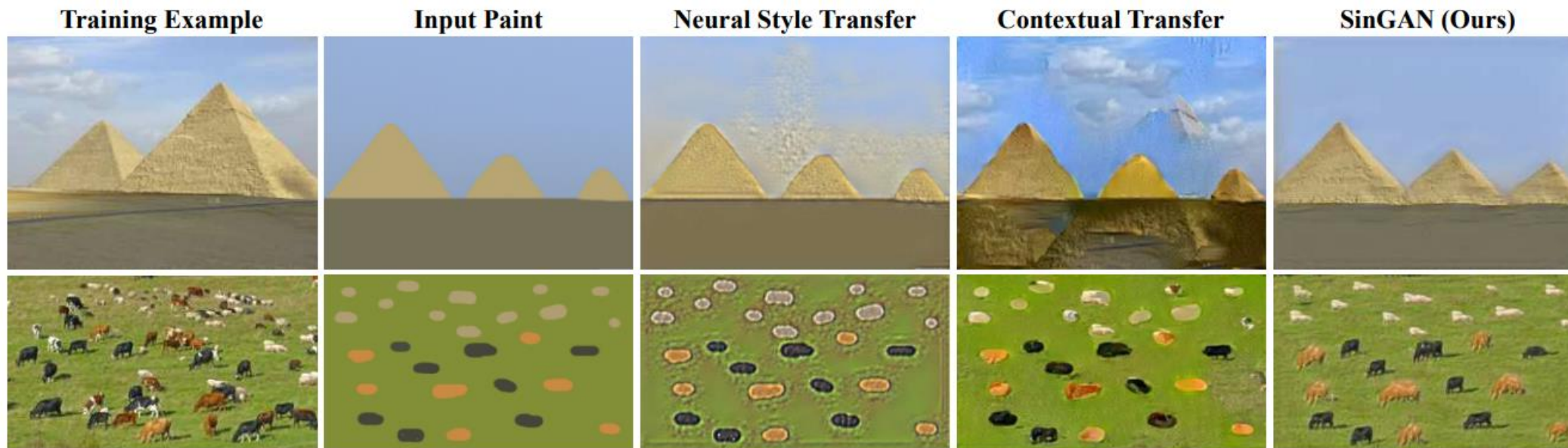
Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, Jun-Yan Zhu et al., ICCV, 2017

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

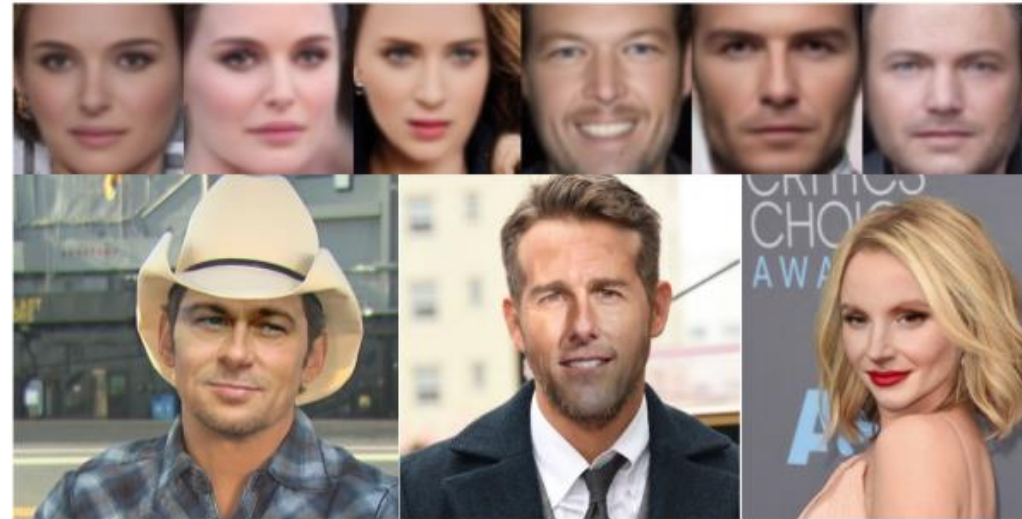


Практически неограниченные возможности реалистичного манипулирования данными...

Детальная генерация 3D лица по одному изображению



Real Faces



Swapped Faces



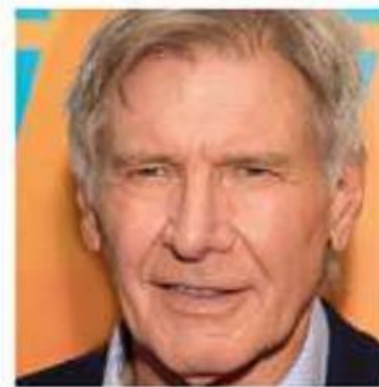
(a) input image



(b) output 3d face



(c) textured 3d face



(d) input image



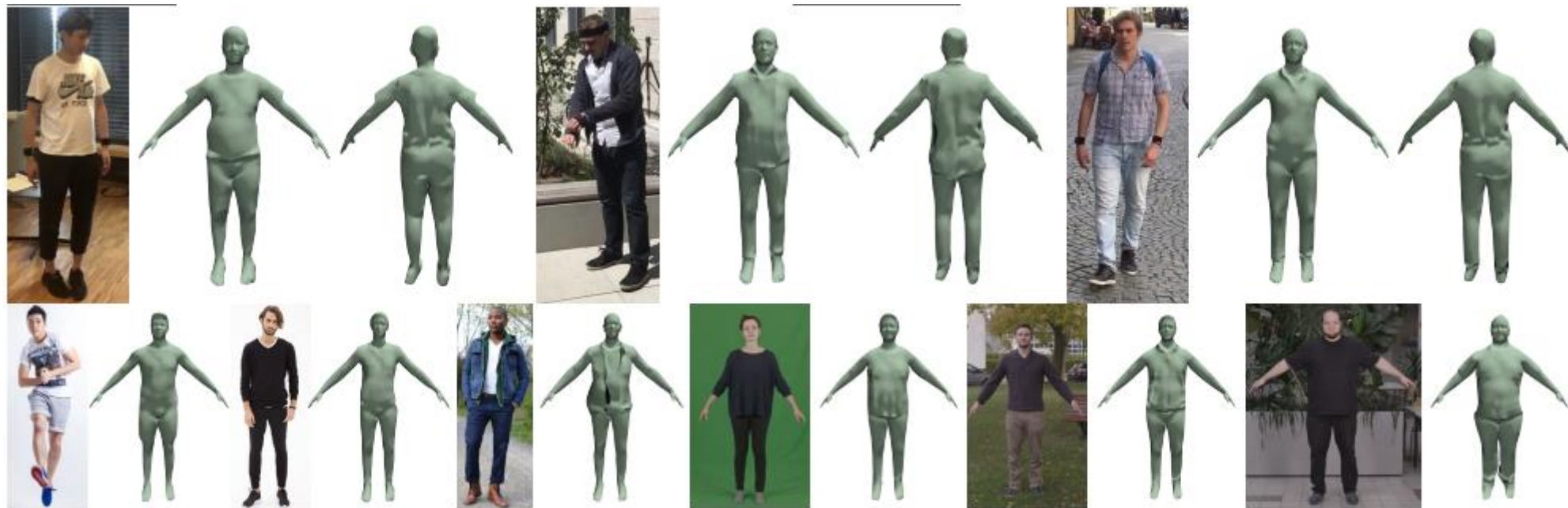
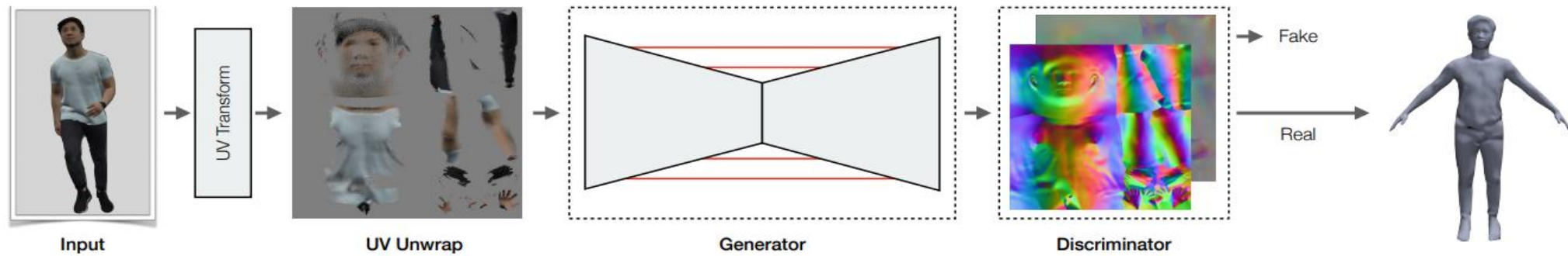
(e) output 3d face



(f) textured 3d face

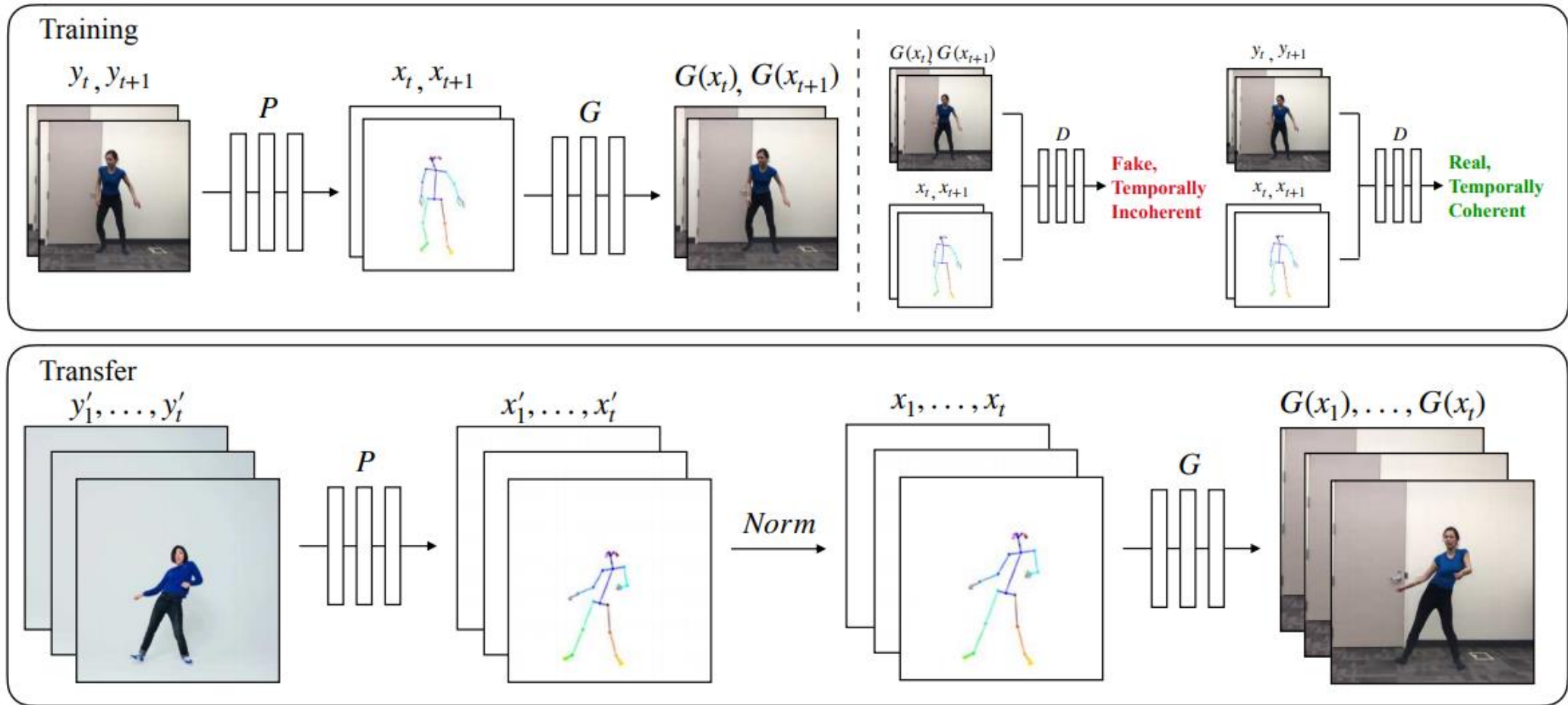
Практически неограниченные возможности реалистичного манипулирования данными...

Детальная генерация 3D модели тела и позы



Практически неограниченные возможности реалистичного манипулирования данными...

Перенос движения: Everybody Dance Now



Практически неограниченные возможности реалистичного манипулирования данными...

Everybody Dance Now



Практически неограниченные возможности реалистичного манипулирования данными...

Everybody Dance Now



Практически неограниченные возможности реалистичного манипулирования данными...

Everybody Dance Now



Практически неограниченные возможности реалистичного манипулирования данными...

Rethinking ImageNet Pre-training

Kaiming He, Ross Girshick, Piotr Dollár

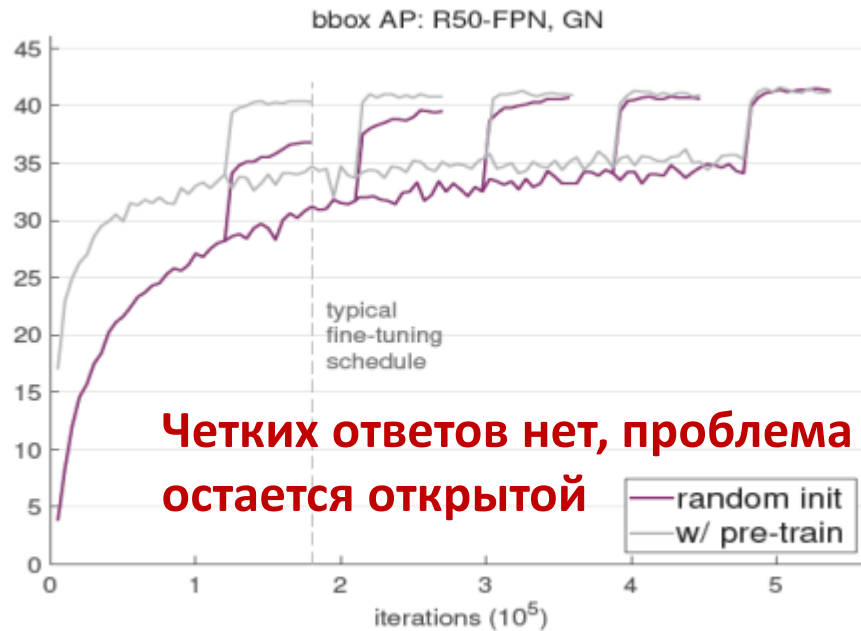


Figure 1. We train Mask R-CNN [13] with a ResNet-50 FPN [26] and GroupNorm [48] backbone on the COCO train2017 set

Перенос обучения в целевой домен путем предобучения на больших базах из другого домена не обязателен! Можно учить и с нуля.

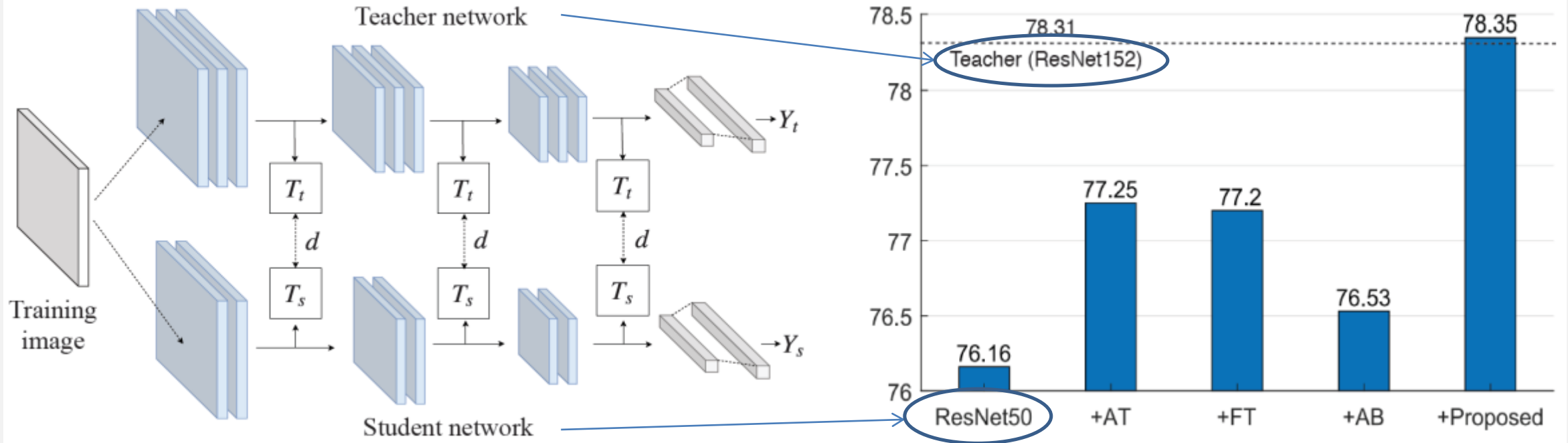
Предобучение полезно для ускорения обучения и компенсации недостатков выборки в целевом домене.

Большие базы из других доменов могут быть полезны, но полезнее были бы большие базы из целевого домена!

- Training **from scratch** on target tasks is **possible without architectural changes**.
- Training **from scratch** requires **more iterations** to sufficiently converge.
- Training **from scratch can be no worse** than its ImageNet pre-training counterparts under many circumstances, down to as few as 10k COCO images.
- ImageNet **pre-training speeds up convergence** on the target task.
- ImageNet **pre-training does not necessarily help reduce overfitting** unless we enter a very small data.
- ImageNet **pre-training helps less** if the target task is more sensitive to **localization than classification**.

Knowledge Distilling

Knowledge Distilling (Teacher/Student CNN Learning)

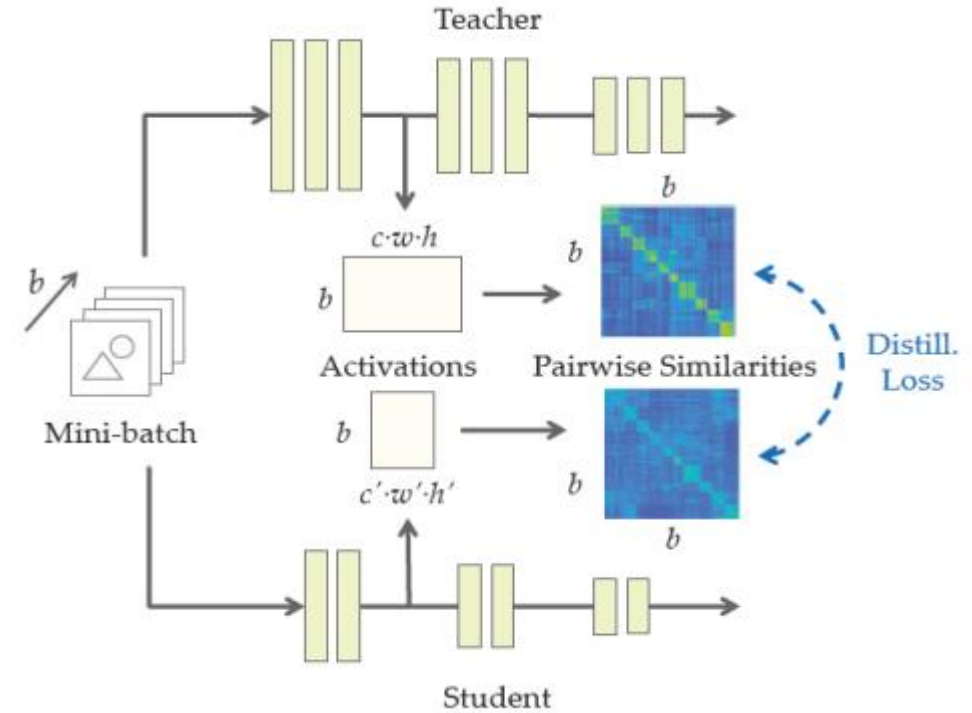
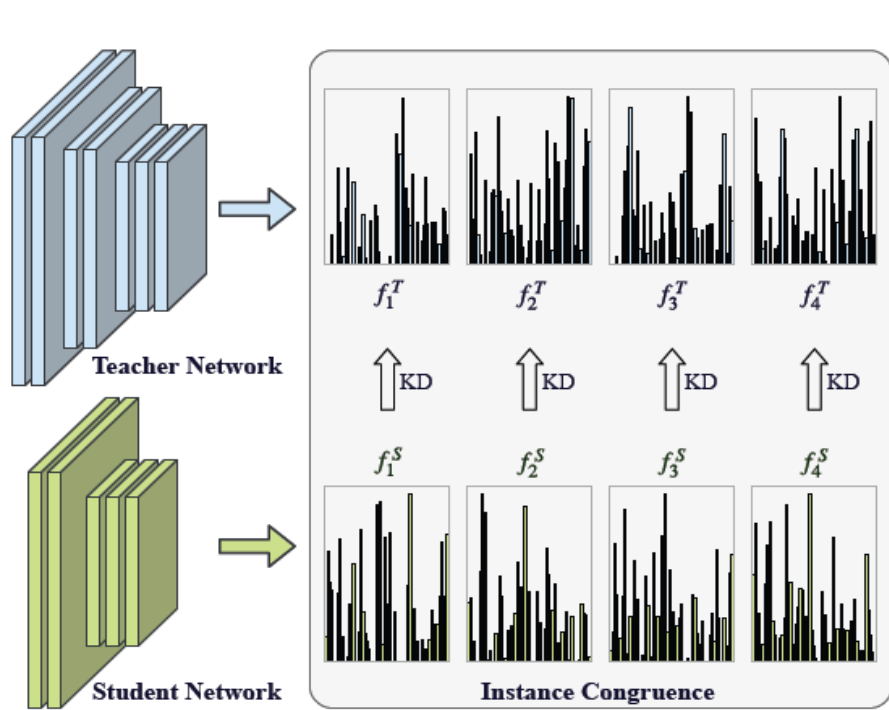


A Comprehensive Overhaul of **Feature Distillation**

Byeongho Heo, Jeesoo Kim, Sangdoon Yun, Hyojin Park, Nojun Kwak, Jin Young Choi

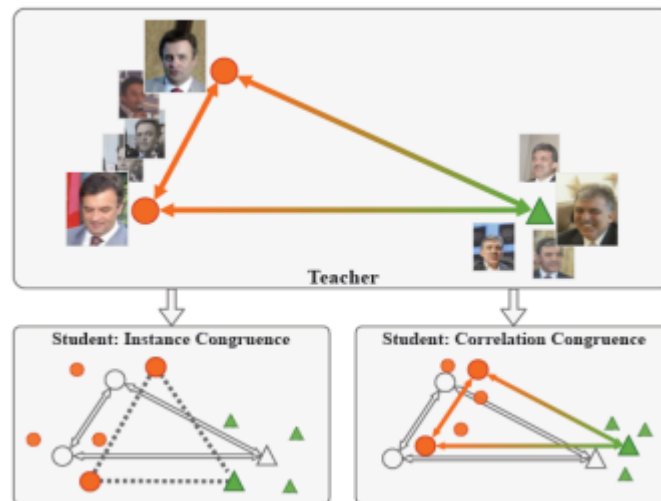
Distilling = извлечение знаний из другой сети путем имитации ее выходов

Knowledge Distilling (Teacher/Student CNN Learning)



Correlation Congruence for Knowledge Distillation

Baoyun Peng, Xiao Jin, Jiaheng Liu, Dongsheng Li, Yichao Wu, Yu Liu, Shunfeng Zhou, Zhaoning Zhang

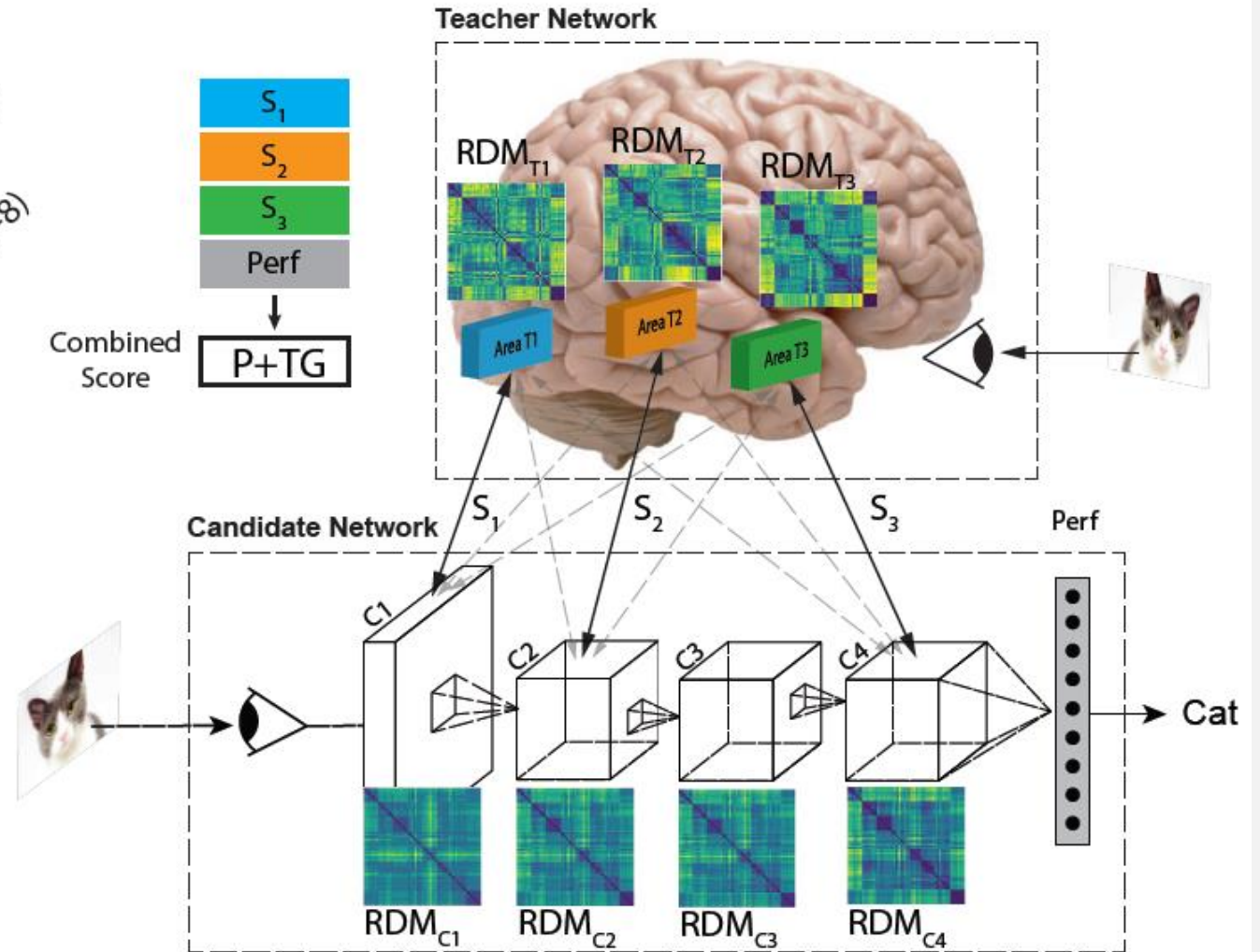
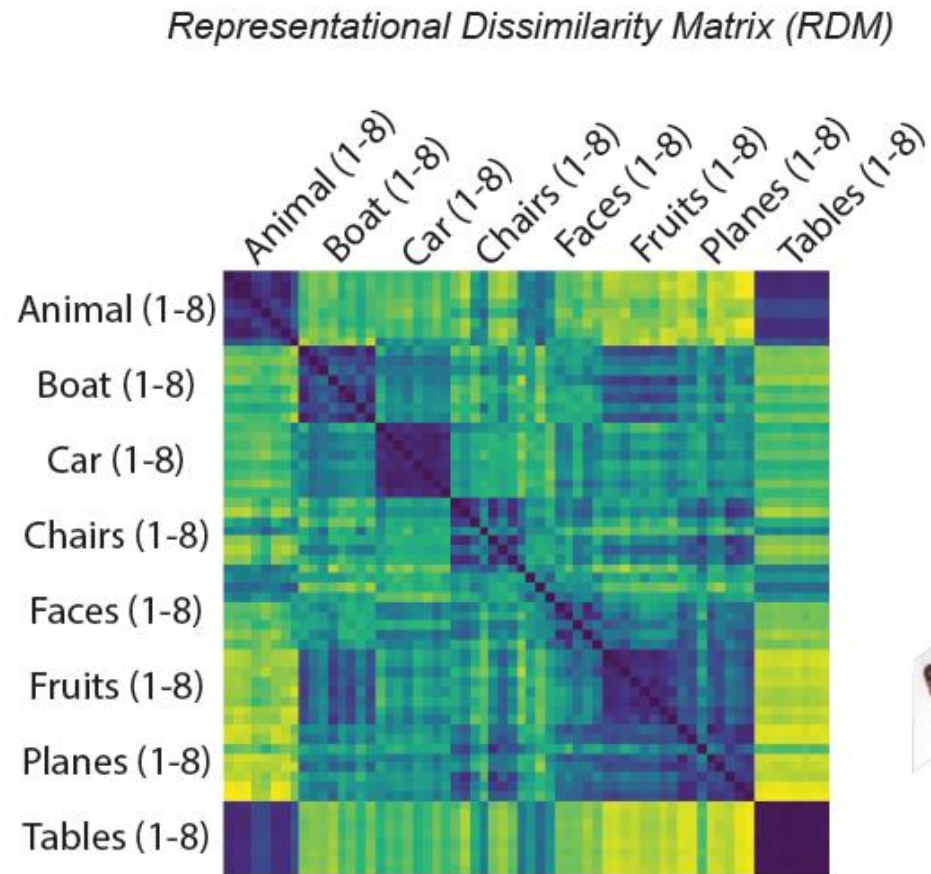


Similarity-Preserving Knowledge Distillation

Frederick Tung and Greg Mori

Distilling = извлечение знаний из другой сети путем имитации ее выходов

Knowledge Distilling for NAS

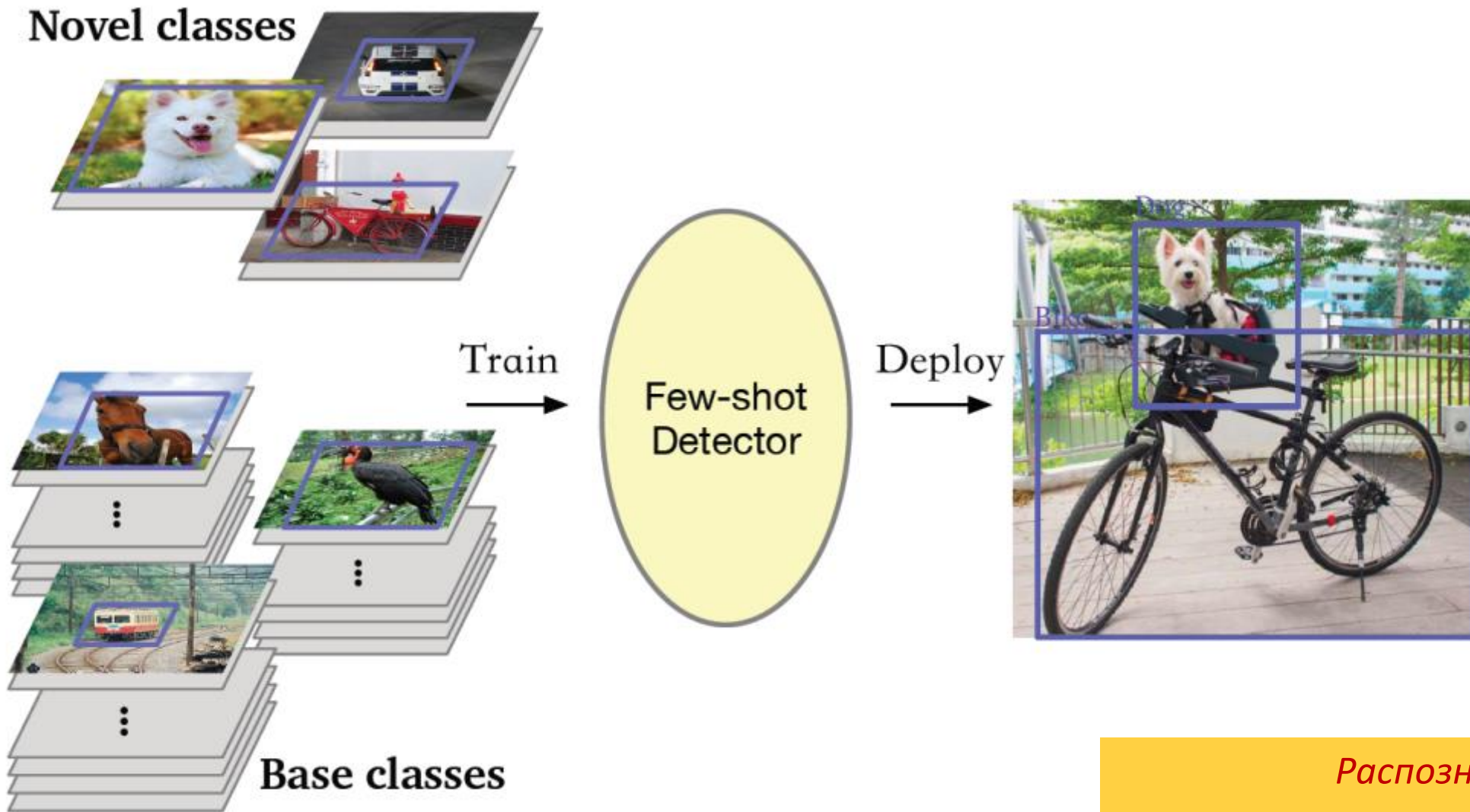


Teacher Guided Architecture Search

Pouya Bashivan, Mark Tensen, James J. DiCarlo

Few-Shot Learning / Detection / Segmentation

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

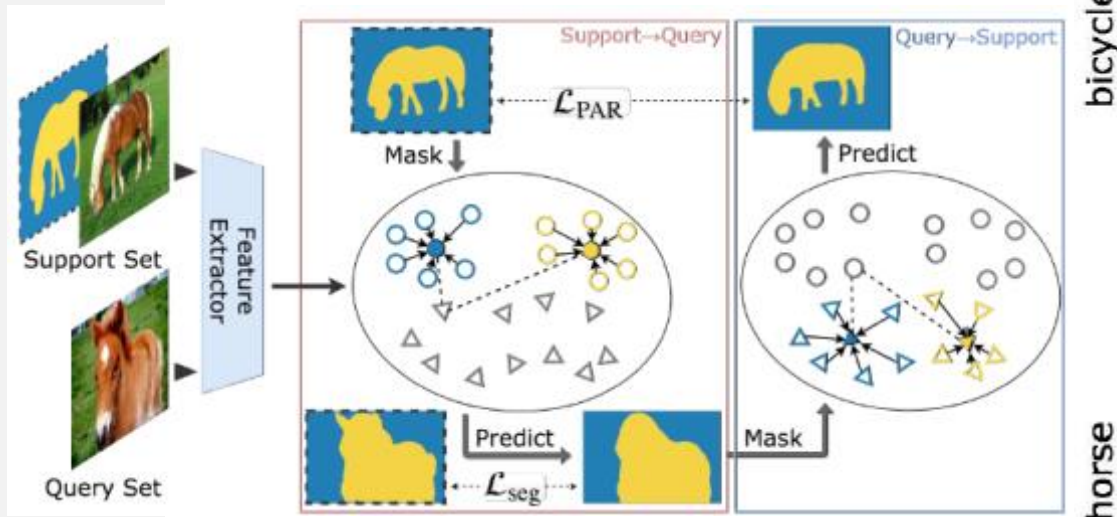


Few-Shot Object Detection via Feature Reweighting

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



Support Image

Query Image

ground truth

prediction

bicycle



horse

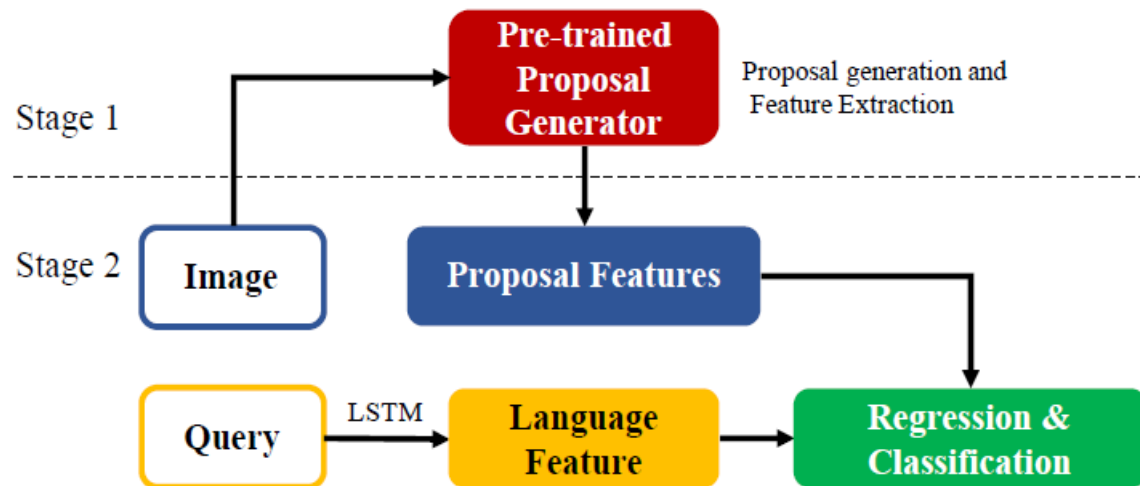


PANet: **Few-Shot** Image **Semantic Segmentation** With Prototype Alignment
Kaixin Wang, Jun Hao Liew, Yingtian Zou, Daquan Zhou, Jiashi Feng

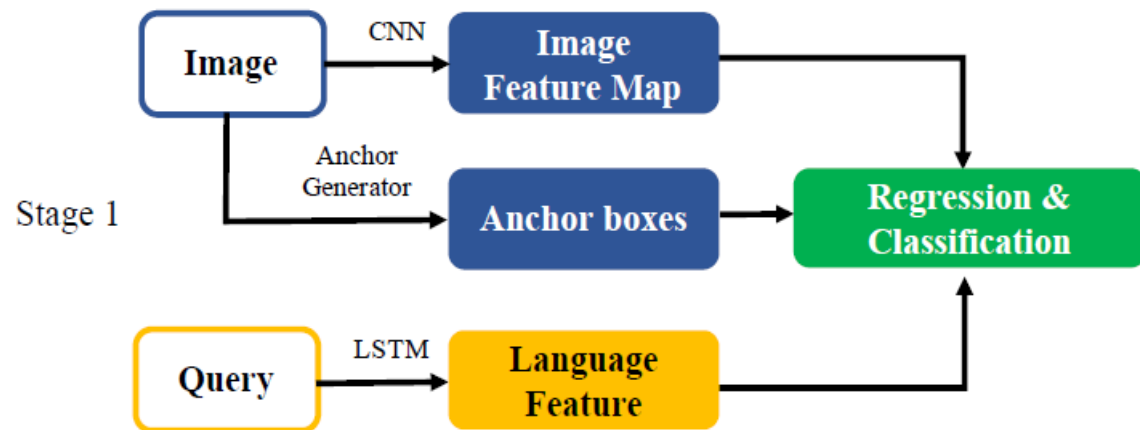
*Распознавание, локализация
и сегментация объектов по
одному или малому числу
эталонов*

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



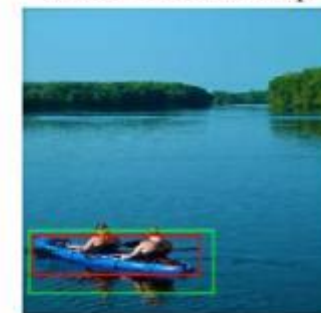
a red beanie cap



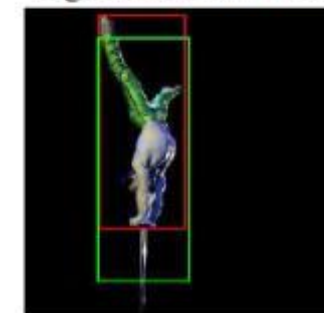
rightmost animal



a cigar



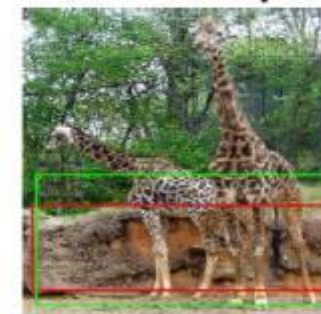
a two-seat kayak



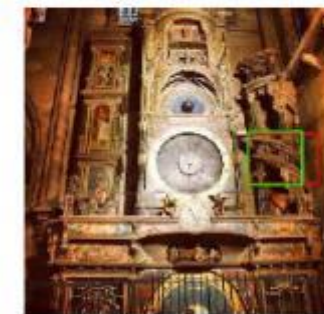
a handstand



a rocky cliff (hill)



large boulders (rock)



stairway (wall)

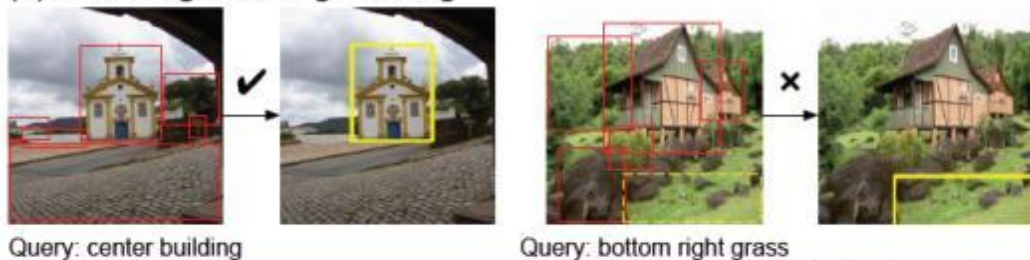
Zero-Shot Grounding of Objects From Natural Language Queries

Arka Sadhu, Kan Chen, Ram Nevatia

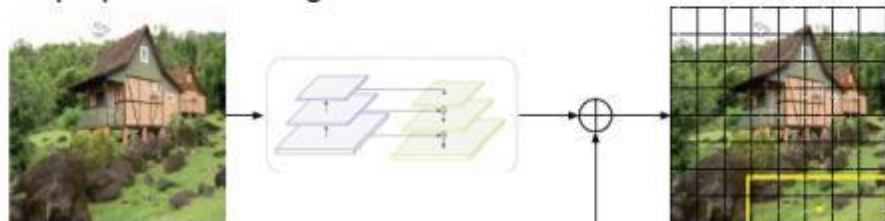
Поиск и локализация объектов без эталона (по описанию)

Zero-Shot Grounding

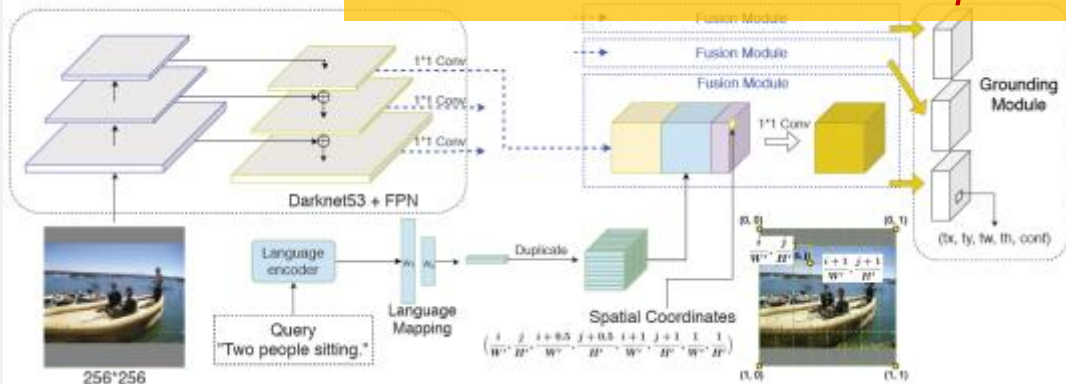
(a). Two-stage visual grounding



(b). The proposed one-stage method



Одноэтапный детектор + Attention с языковым запросом



(a). Query: bike of blue pant lady



(b). Query: the bowl of bean on the bottom



(c). Query: person on the right



(g). Query: man in blue



(h). Query: kid left



(i). Query: window above colonial

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!

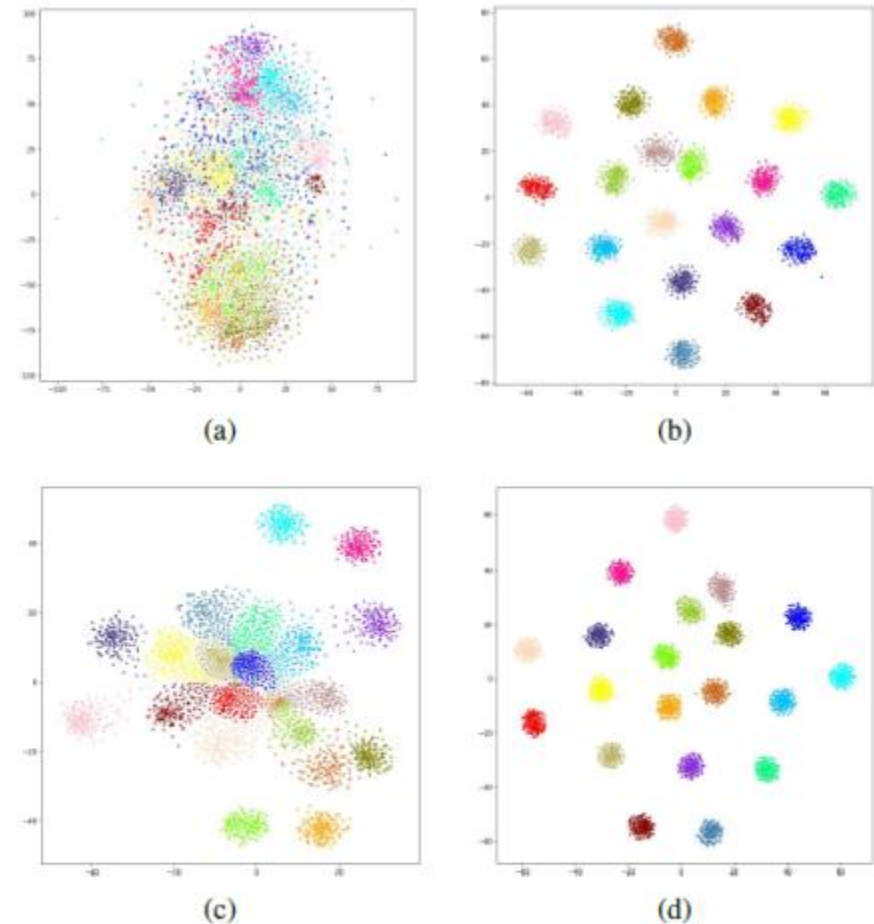
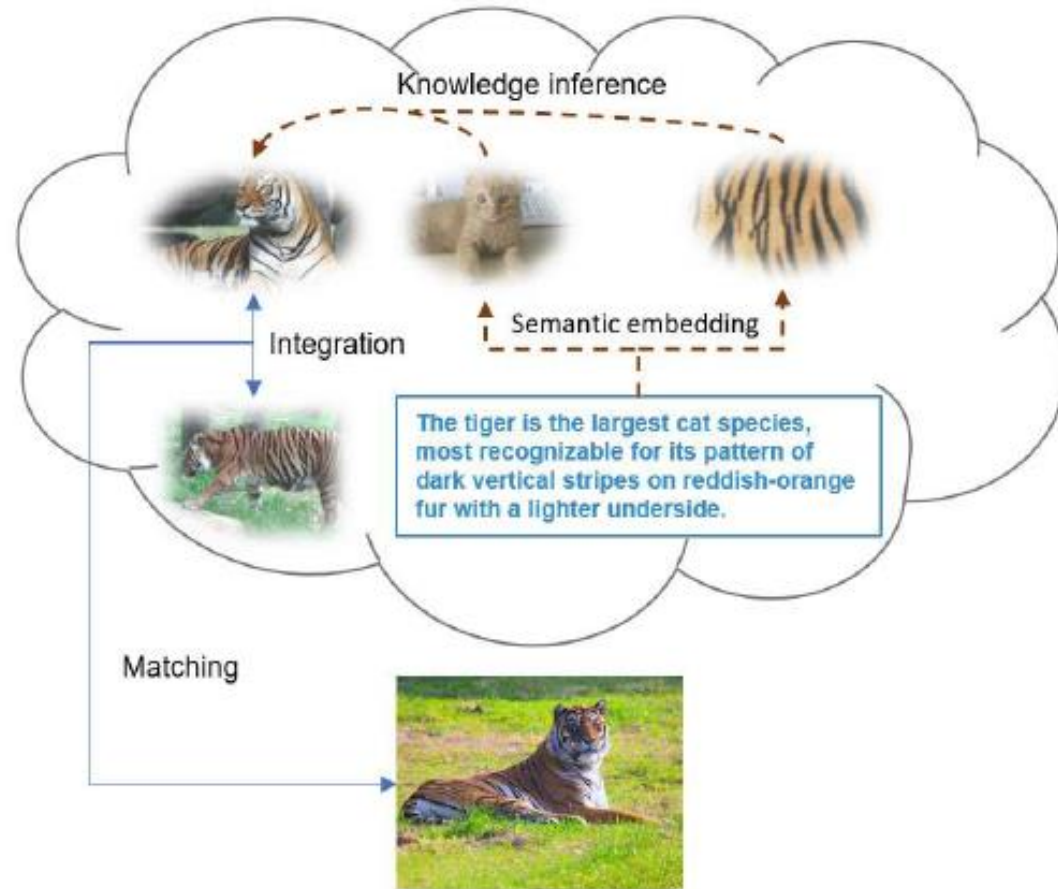


Figure 3. T-SNE visualization results for all novel categories in the Mini-imagenet set on the 1-shot and 5-shot tasks. Each scatter plot contains 20 colored classifier parameter clusters and each color represents a novel category. (a): 1-shot vision-based classifier. (b): 1-shot vision-knowledge classifier.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

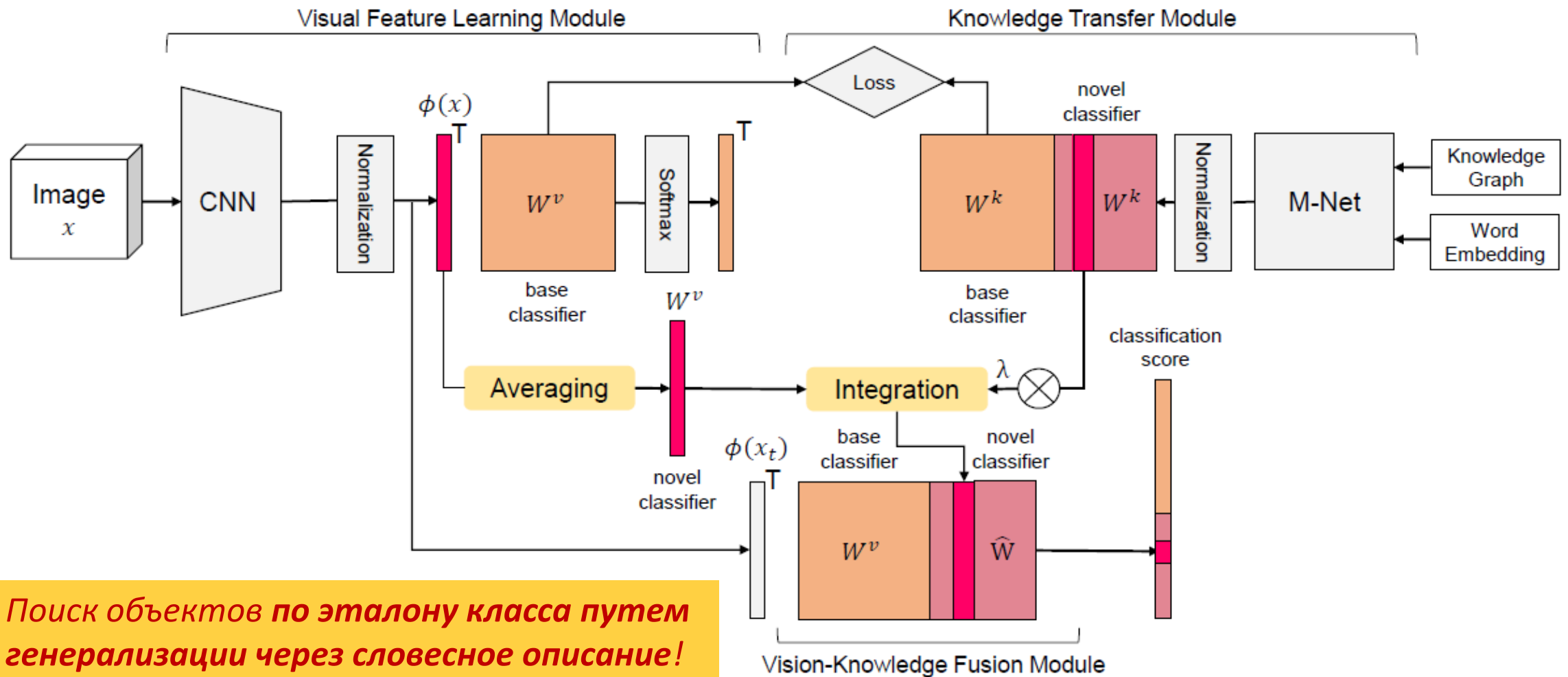
Поиск и локализация объектов по эталону класса путем генерализации через словесное описание!

Few-Shot Image Recognition With Knowledge Transfer

Zhimao Peng, Zechao Li, Junge Zhang, Yan Li, Guo-Jun Qi, Jinhui Tang

Классы намного более компактные

Few-Shot via Zero-Shot!



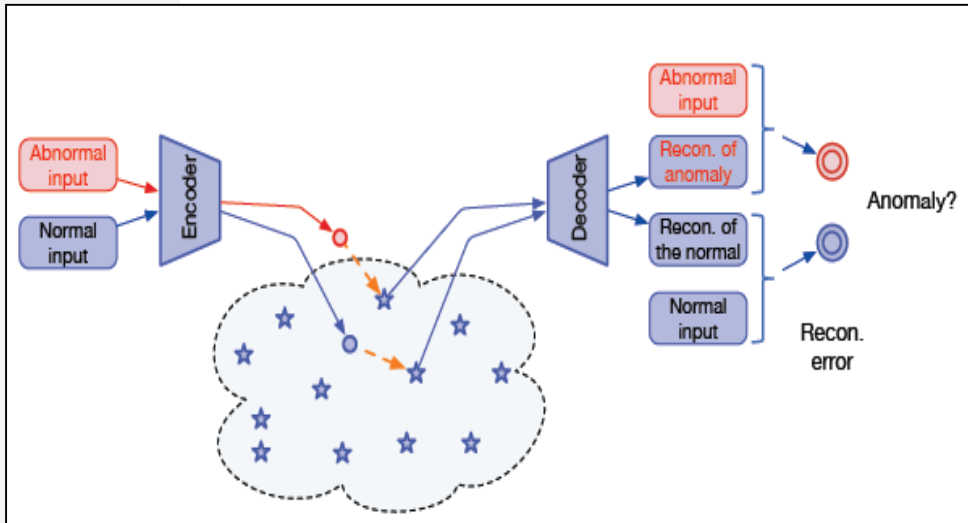
Поиск объектов по эталону класса путем генерализации через словесное описание!

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

Поиск и локализация объектов по эталону

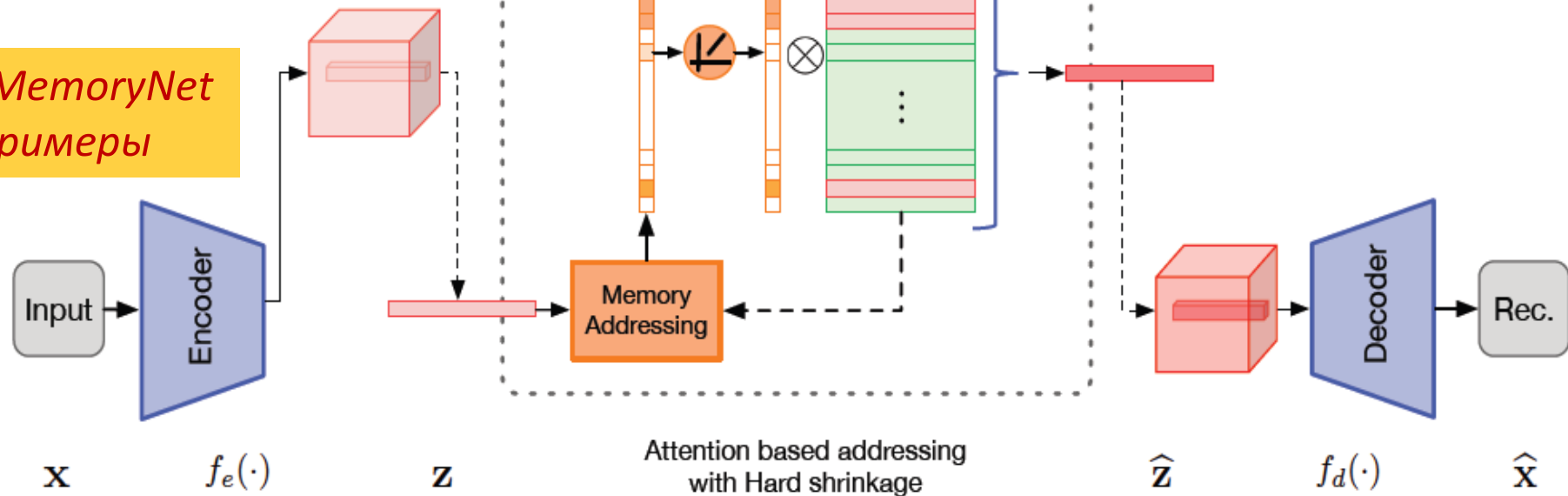
MemoryNet

MemoryNet for Anomaly Detection

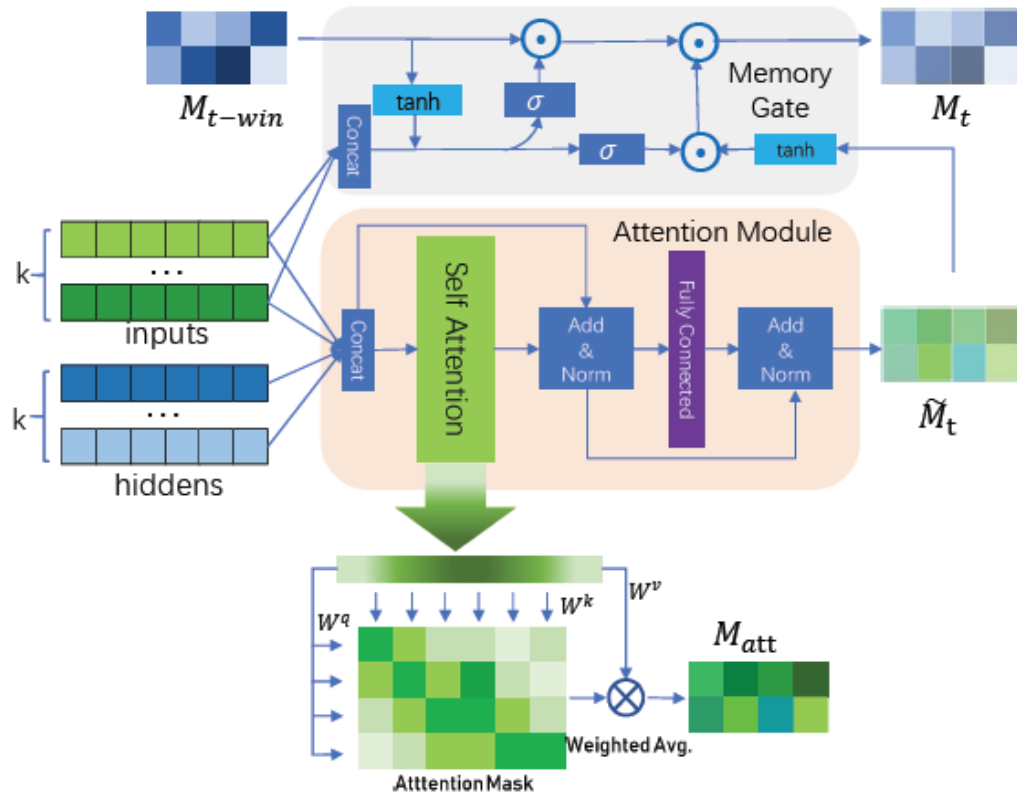


MemoryNet напоминает архитектуру фон Неймана тем, что имеет в явном виде блок памяти

В данном случае MemoryNet запоминает примеры

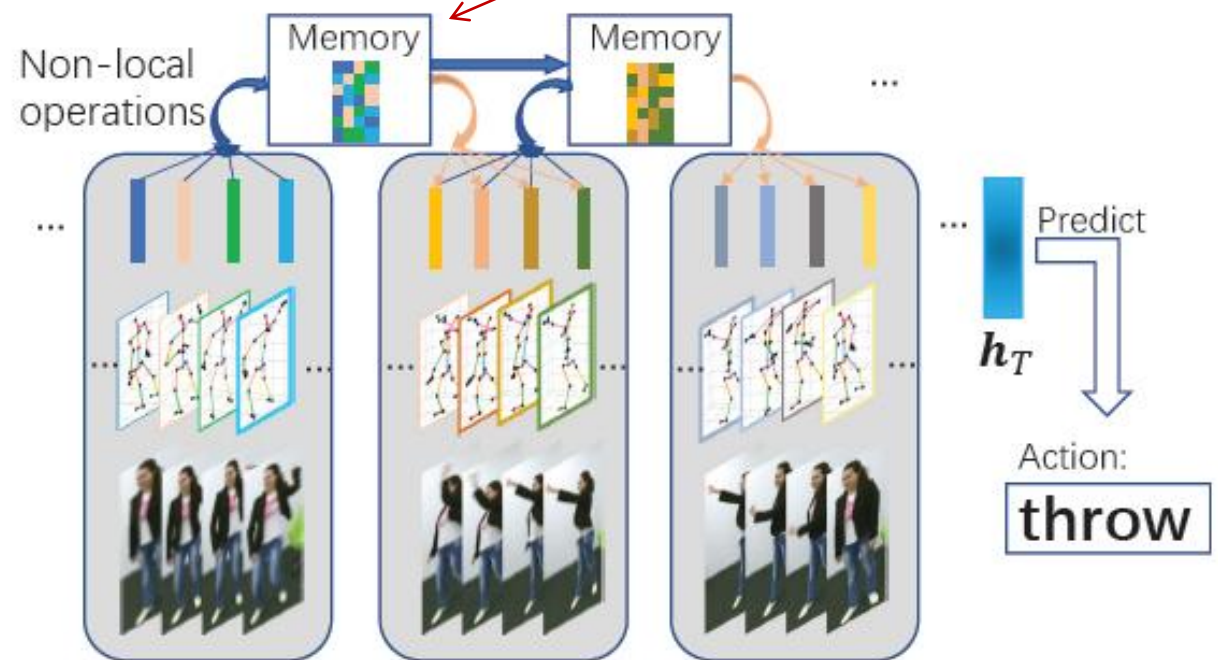


MemoryNet for Temporal Information Storage



В данном случае MemoryNet запоминает информацию о прошлых событиях и образах

MemoryNet напоминает архитектуру фон Неймана тем, что имеет в явном виде блок памяти



Non-Local Recurrent **Neural Memory** for Supervised **Sequence Modeling**

Canmiao Fu, Wenjie Pei, Qiong Cao, Chaopeng Zhang, Yong Zhao, Xiaoyong Shen, Yu-Wing Tai

Scene Graph, VQA,
Image-to-Text

Specifying Object Attributes and Relations in Interactive Scene Generation

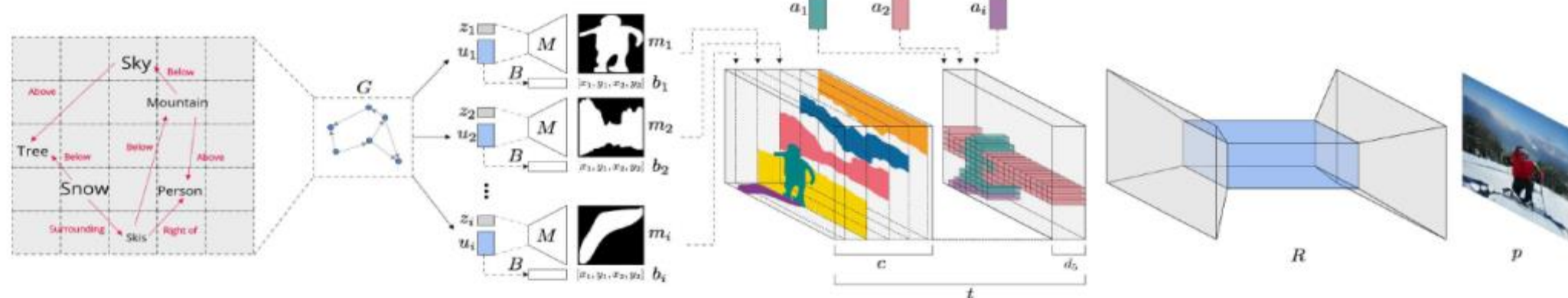
Oron Ashual, Lior Wolf

Tel Aviv University and Facebook AI Research



Семантический граф сцены извлекается из реального изображения...

...и используется для генерации модельного



Meta-Sim: Learning to Generate Synthetic Datasets

Amlan Kar, Aayush Prakash, Ming-Yu Liu, Eric Cameracci, Justin Yuan, Matt Rusiniak, David Acuna, Antonio Torralba, Sanja Fidler

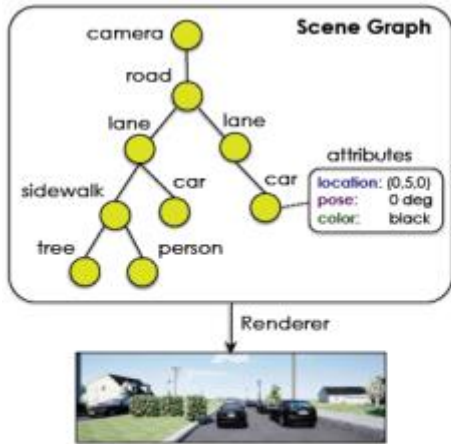


Figure 3. Simple scene graph example for a driving scene.

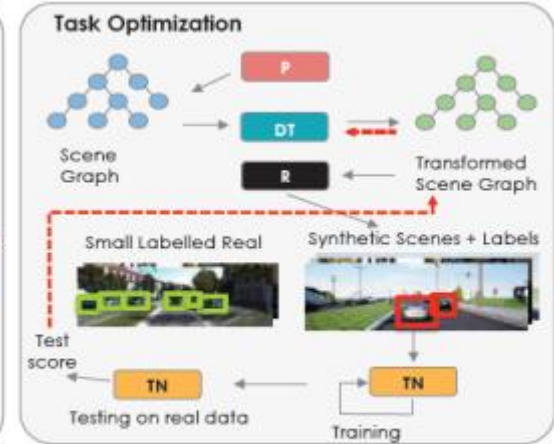
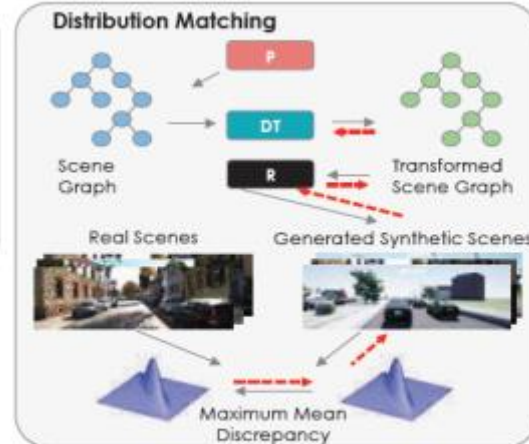
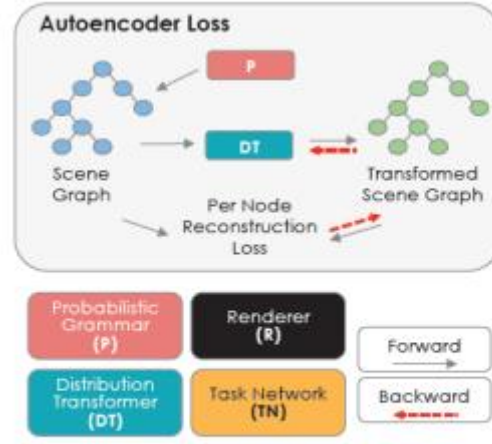
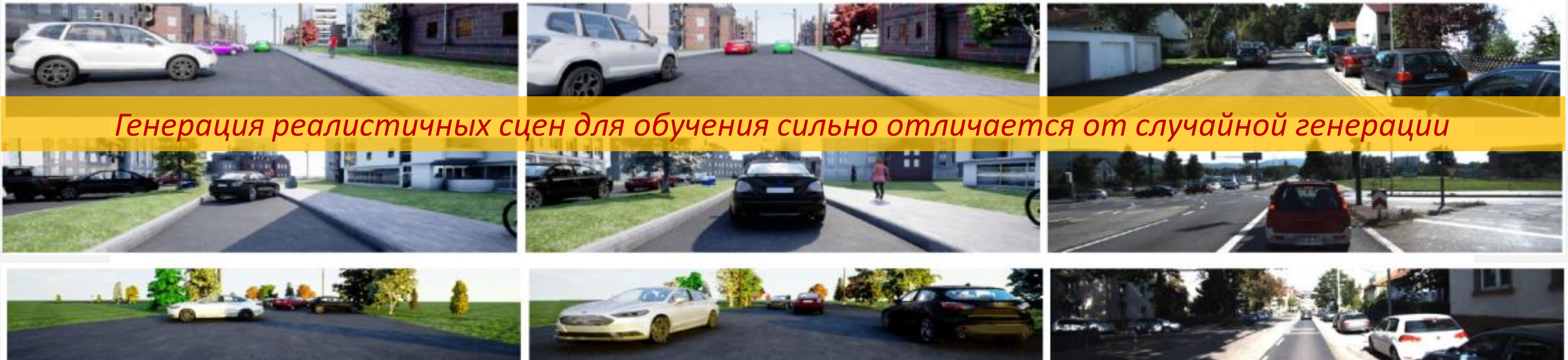


Figure 4. Illustration of different losses used in *Meta-Sim*, including forward and backward pass control flow for each step. We indicate transformed attributes of a scene graph by changing colors of the nodes.



Генерация реалистичных сцен для обучения сильно отличается от случайной генерации

По случайному графу

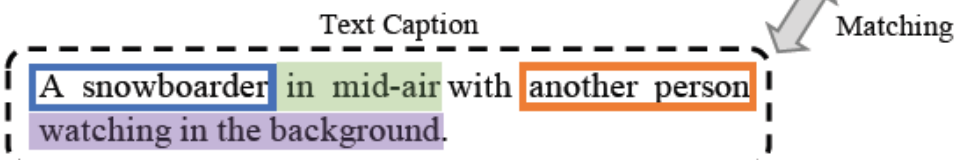
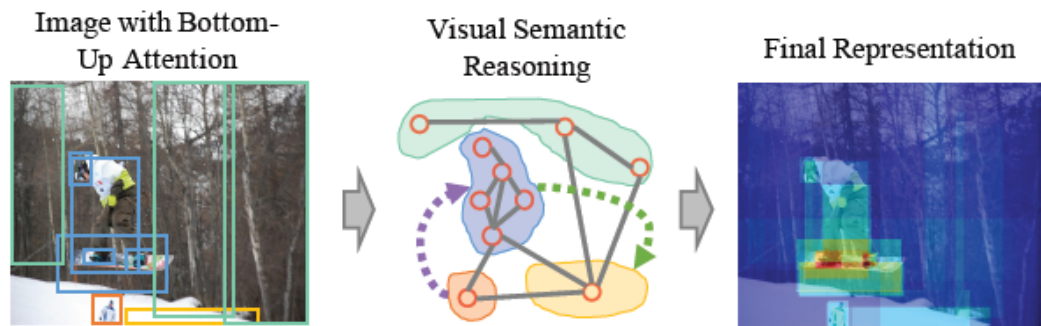
По реалистичному графу

Реальное изображение

Figure 12: (left) samples from our prob. grammar, (middle) *Meta-Sim*'s corresponding samples, (right) random samples from KITTI

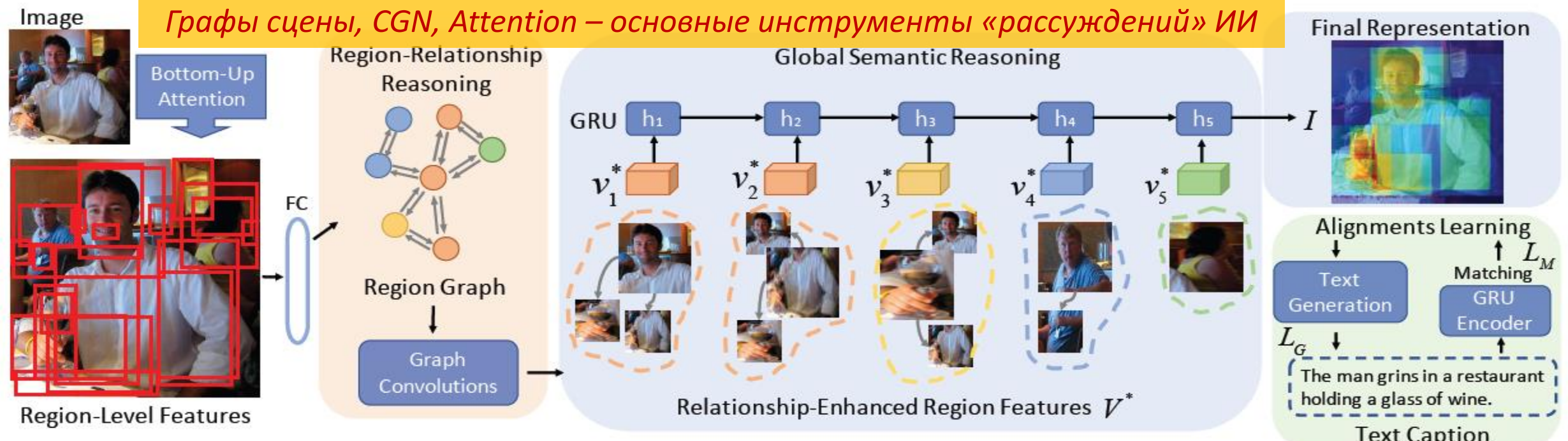
Visual Semantic Reasoning for Image-Text Matching

Kunpeng Li, Yulun Zhang, Kai Li, Yuanyuan Li and Yun Fu



- (f) Query Image Attention
1. Traffic light hanging near power lines with trees in back ✓
 2. Two traffic signals on a pole arm at an intersection. ✓
 3. A street sign and two traffic lights hang over US Route 1 ✗

Графы сцены, CGN, Attention – основные инструменты «рассуждений» ИИ

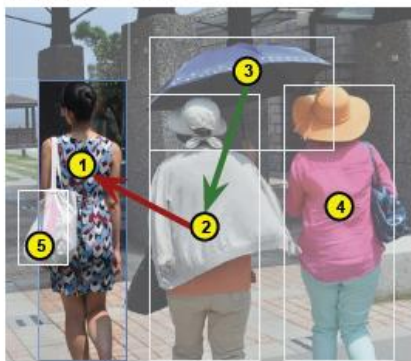


Language-Conditioned Graph Networks for Relational Reasoning

Ronghang Hu, Anna Rohrbach, Trevor Darrell, Kate Saenko

Question: Is there a person to the left of the woman holding a blue umbrella?

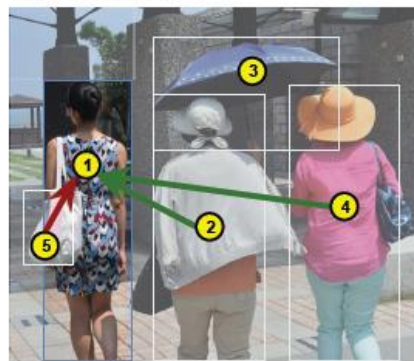
Answer: Yes



input image

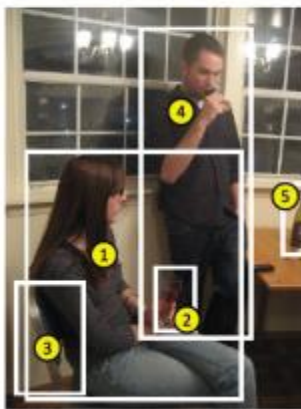
Question: Is the left-most person holding a red bag?

Answer: No

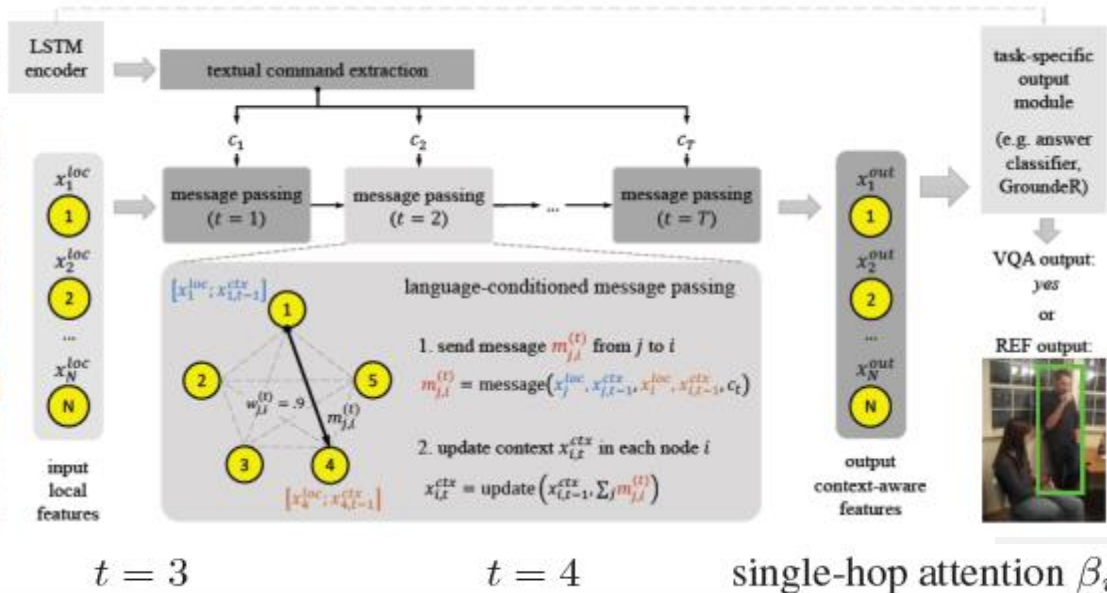


$t = 1$

Is there a man on the right of a person sitting on a chair holding a wine glass?



$t = 2$



question: *what color is the matte ball that is the same size as the gray metal thing?* prediction: yellow ground-truth: yellow



question: *how many other things are the same size as the yellow rubber ball?* prediction: 3 ground-truth: 3

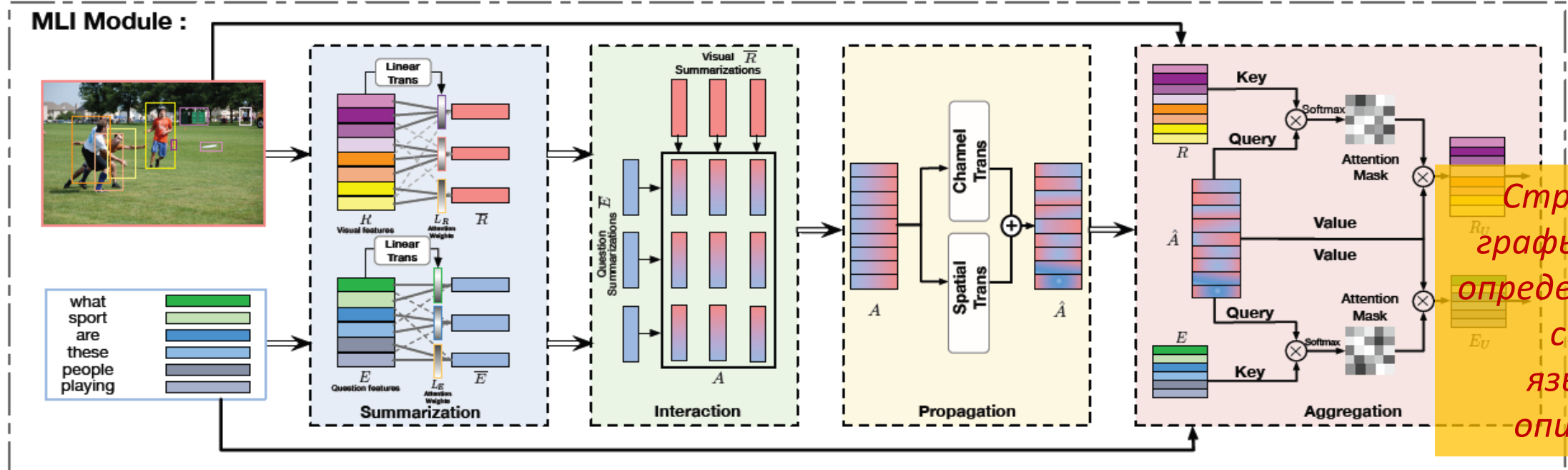
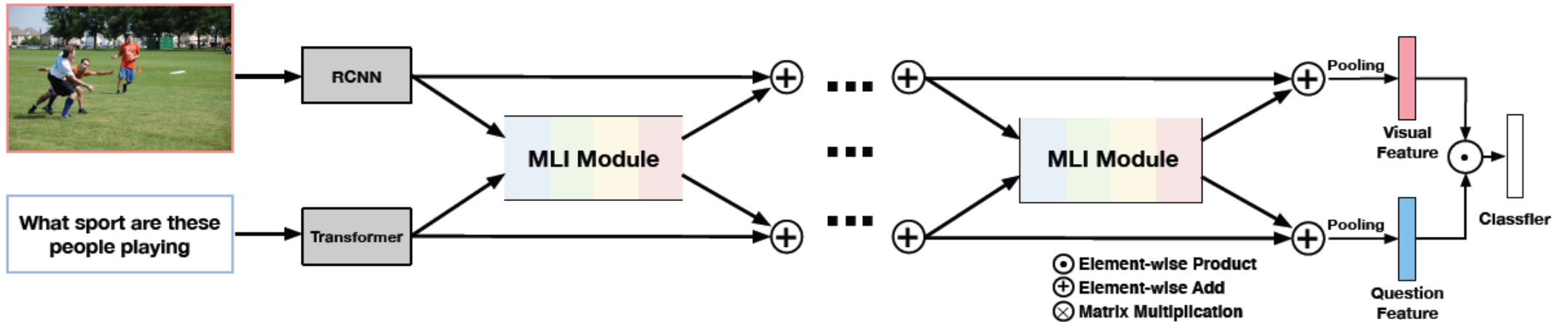


Структура графы сцены определяется языковым описанием

Семантические графы, CGN, Attention – основные инструменты «рассуждений» ИИ

Multi-modality Latent Interaction Network for **Visual Question Answering**

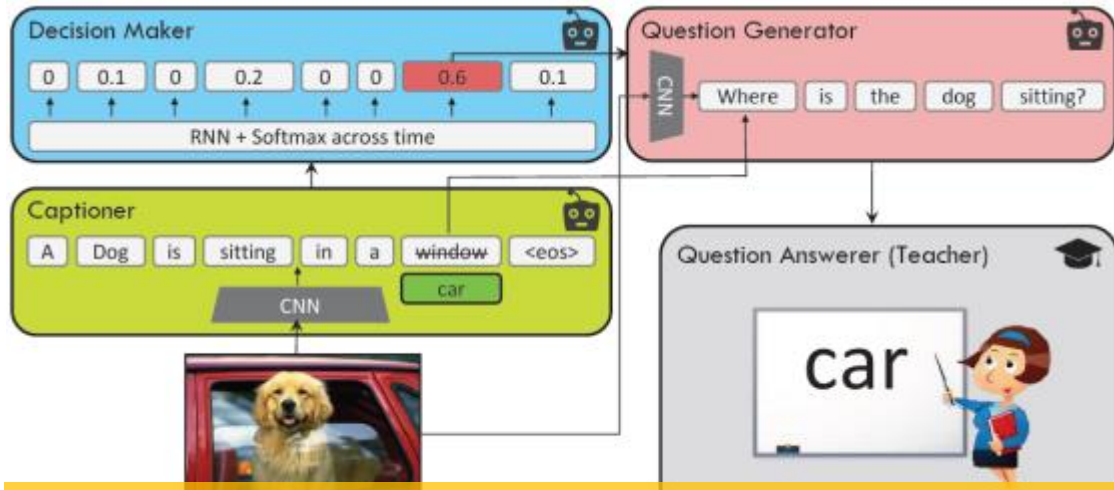
Peng Gao, Haoxuan You, Zhanpeng Zhang, Xiaogang Wang, Hongsheng Li



Семантические графы, CGN, Attention – основные инструменты «рассуждений» ИИ

Learning to Caption Images Through a Lifetime by Asking Questions

Tingke Shen, Amlan Kar, Sanja Fidler

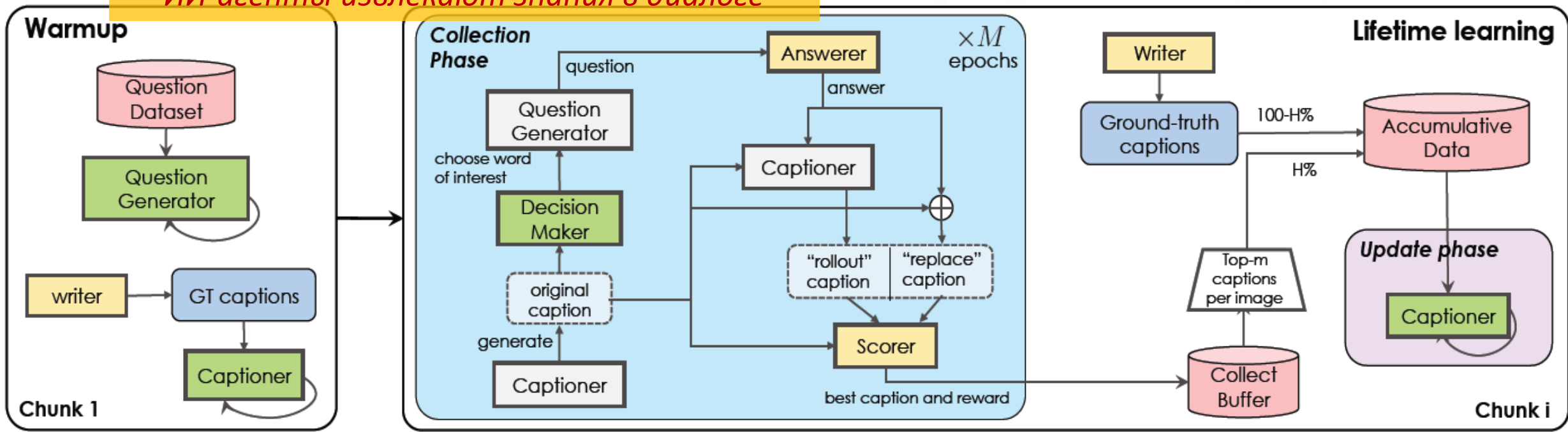


C: Two cats sit in a room with a cat.
Q1: What animal is in the photo? **A:** cat
Q2: What are the cats doing? **A:** looking out window
Q3: Are these cats sitting or outside? **A:** inside
Q4: What are the cats looking at? **A:** window
GTQ: What animals are shown?
GTQ: How many cats are there?



C: A train sitting on the tracks.
Q1: What is the yellow object? **A:** train
Q2: Is this train moving or coming? **A:** going
Q3: Is the train in or outside? **A:** outside
Q4: Where is the train? **A:** station
GTQ: What color are the train doors on the right?
GTQ: What shape are the windows?

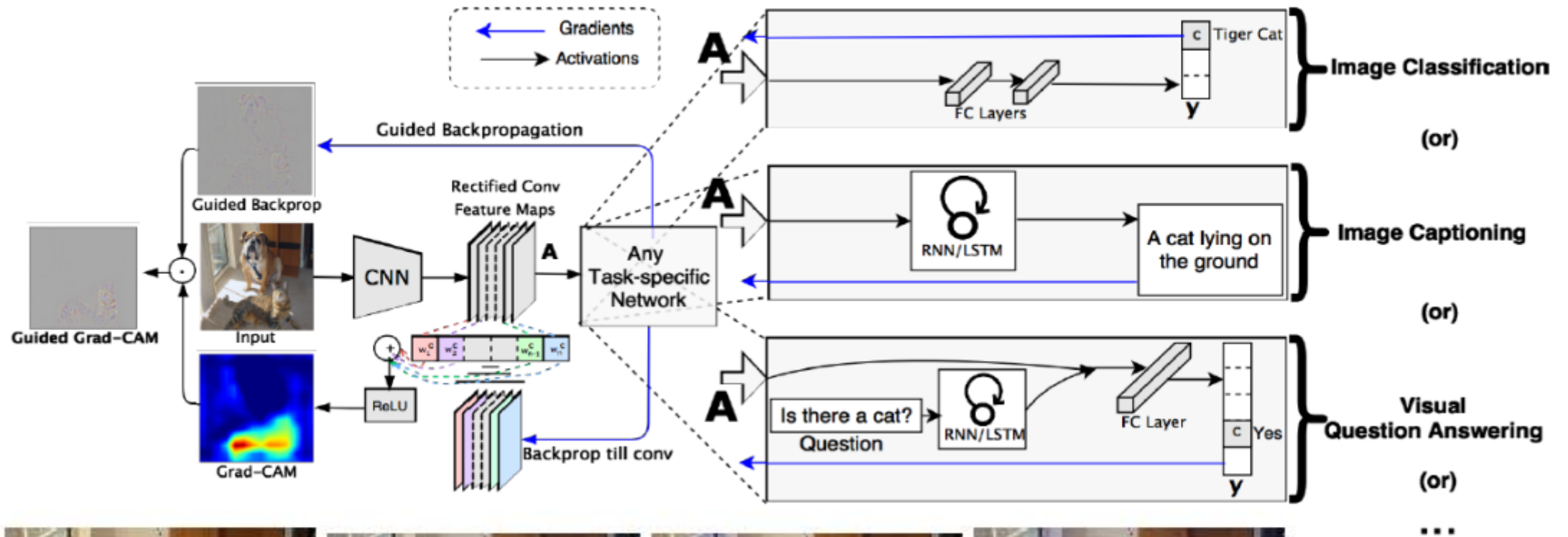
ИИ-агенты извлекают знания в диалоге



Explanation

Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization (2016)

Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam



Boxer: 0.40 Tiger Cat: 0.18

(a) Original image



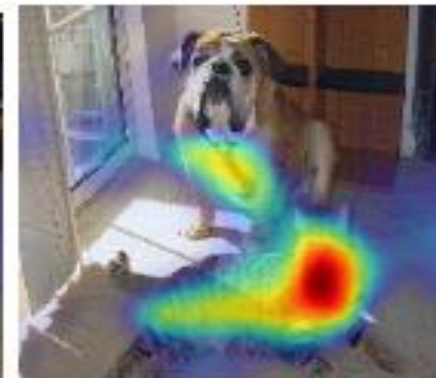
Airliner: 0.9999

(b) Adversarial image



Boxer: $1.1e-20$

(c) Grad-CAM "Dog"



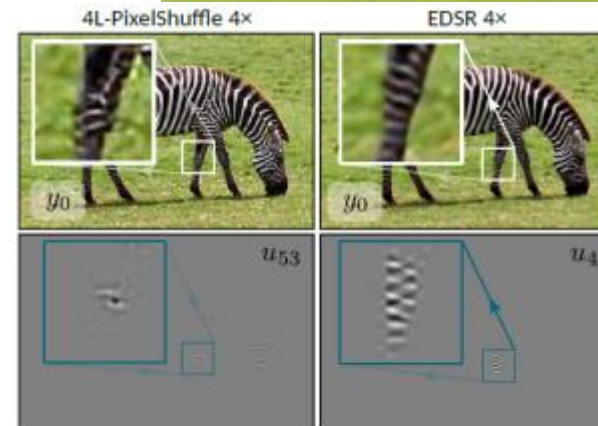
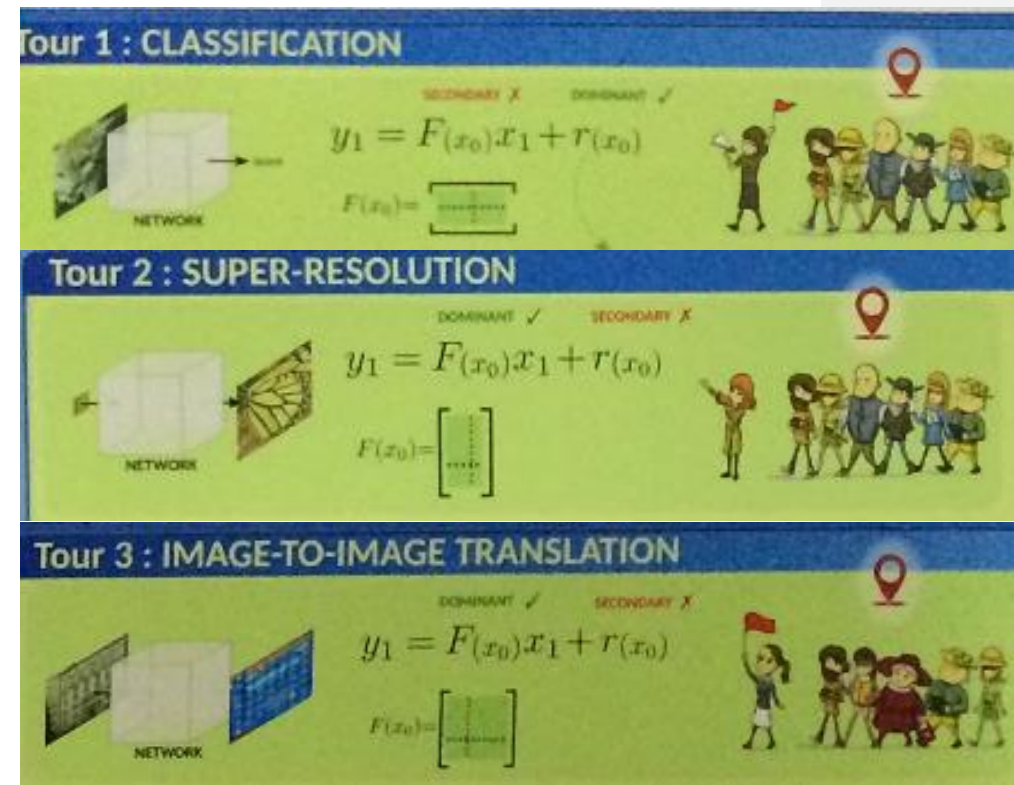
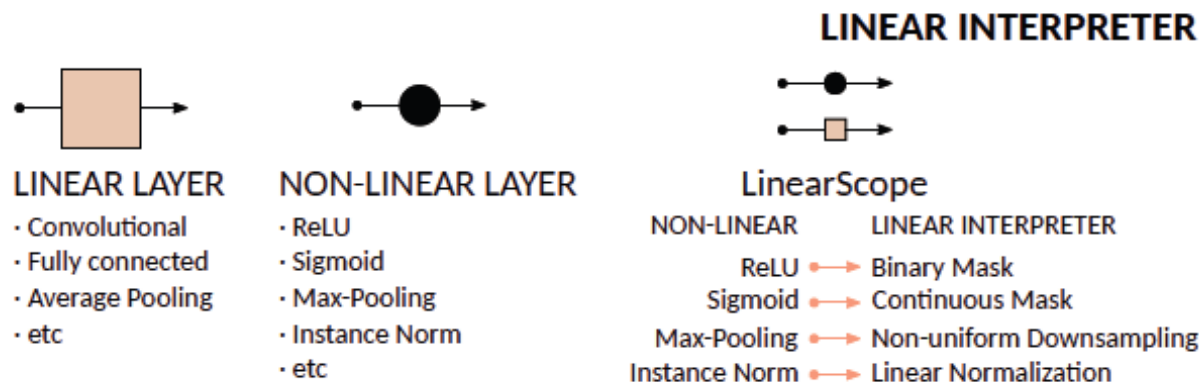
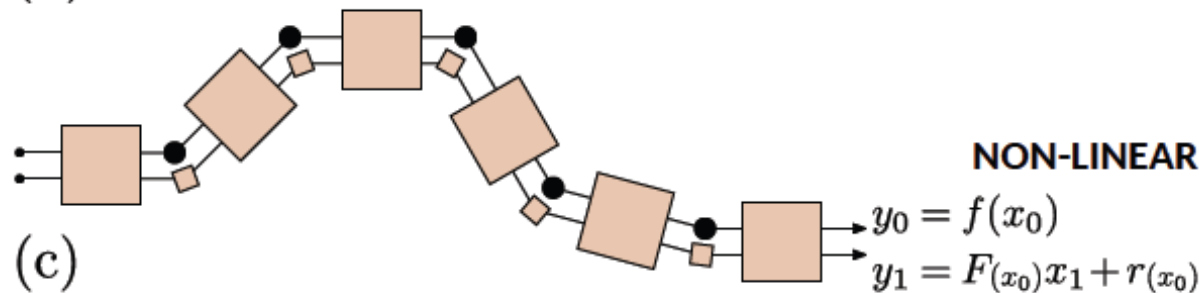
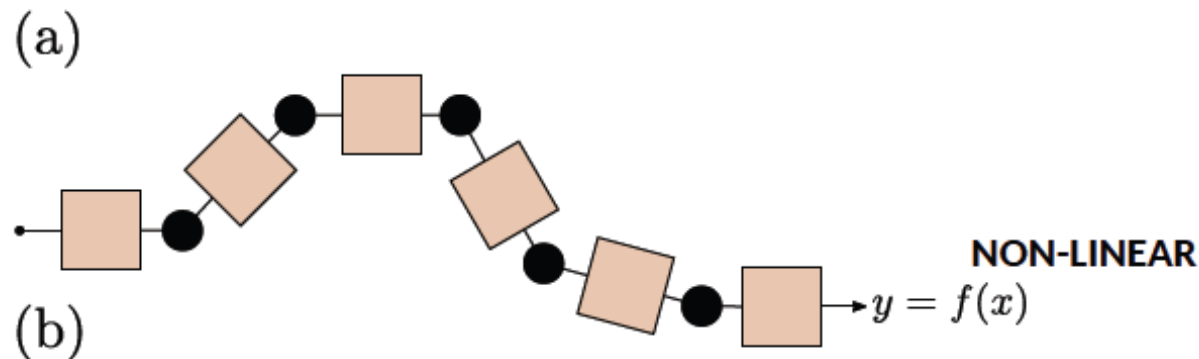
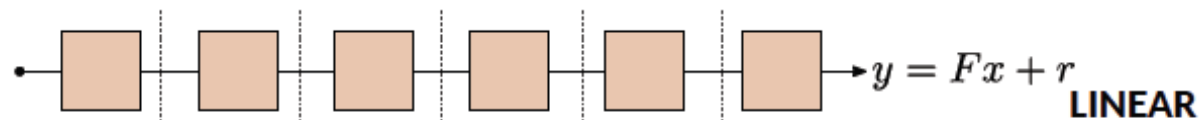
Tiger Cat: $6.5e-17$

(d) Grad-CAM "Cat"

Объяснение, на что «смотрит» сеть, когда ошибается, через анализ градиентов

A Tour of Convolutional Networks Guided by Linear Interpreters

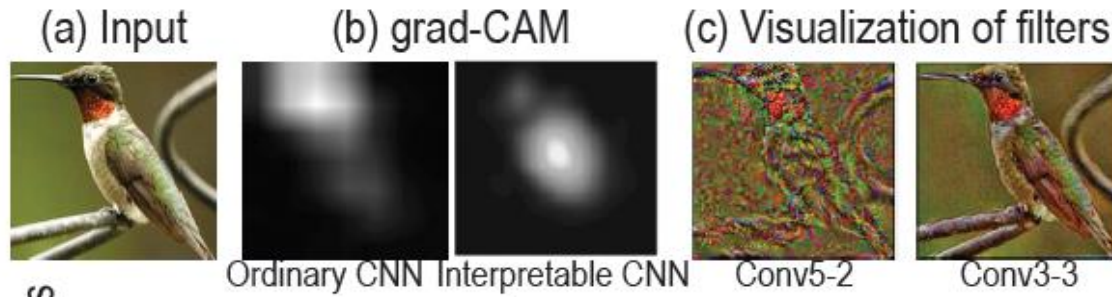
Pablo Navarrete Michelini, Hanwen Liu, Yunhua Lu, Xingqun Jiang



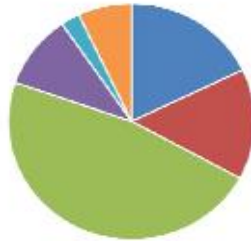
Объяснение, на что «смотрит» сеть, через адаптивный линейный фильтр

Explaining Neural Networks Semantically and Quantitatively

Runjin Chen, Hao Chen, Jie Ren, Ge Huang, and Quanshi Zhang

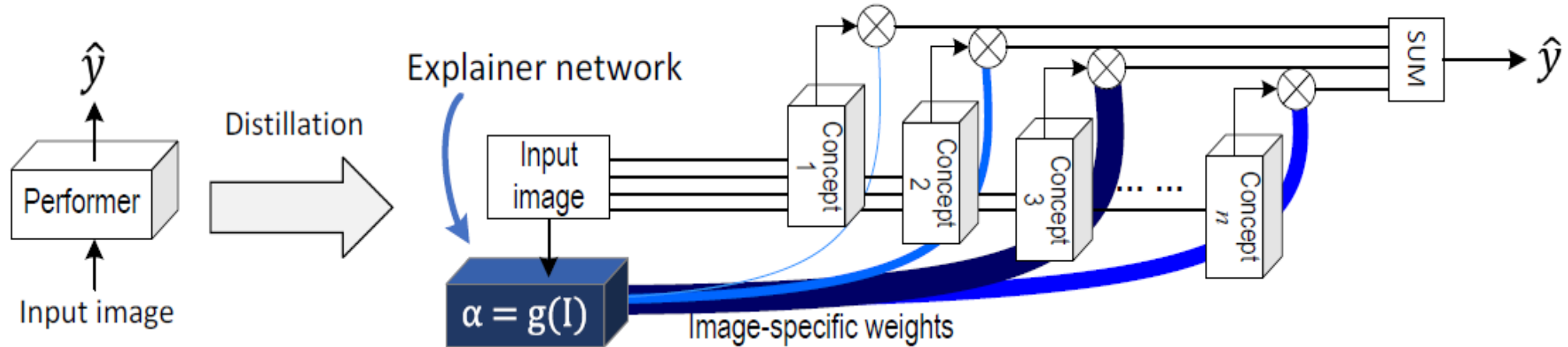


1) Our explanations



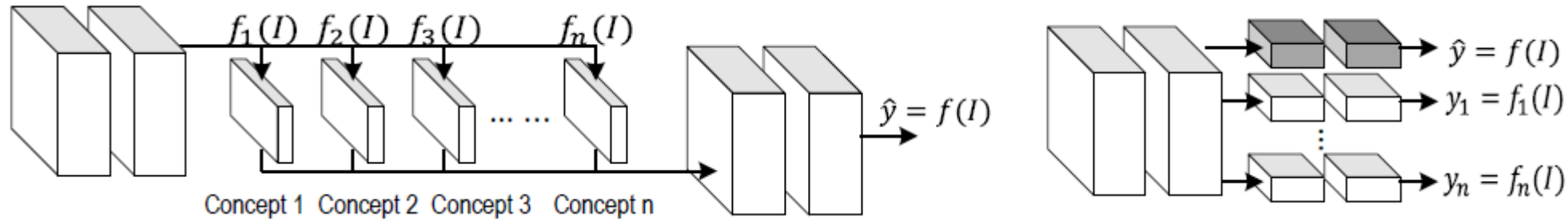
- **Head** contributes 17.8%
- **Joint of head & torso** contribute 15.1%
- **Torso** contributes 47.4%
- **Joint of body & tail** contribute 10.1%
- **Joint of legs & feet** contribute 2.5%
- **Wing** contributes 7.1%

Объяснение, на что «смотрит» сеть, через дистиллированный адаптивный линейный фильтр

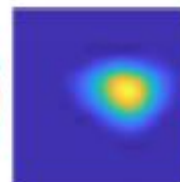
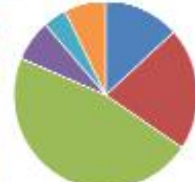
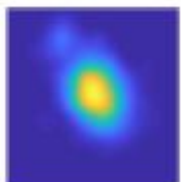
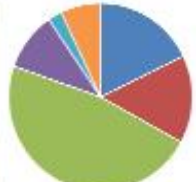
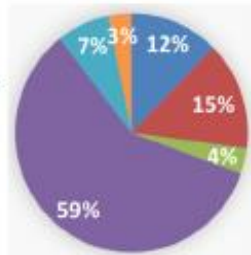
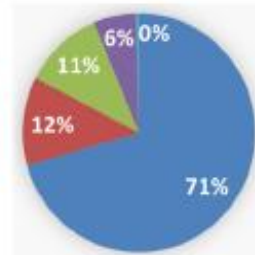


Explaining Neural Networks Semantically and Quantitatively

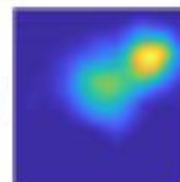
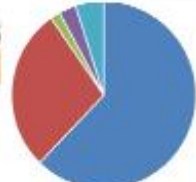
Runjin Chen, Hao Chen, Jie Ren, Ge Huang, and Quanshi Zhang



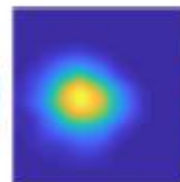
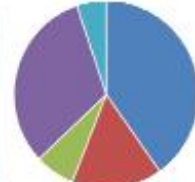
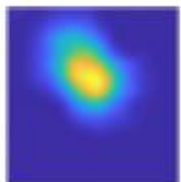
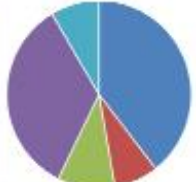
Объяснение, на что «смотрит» сеть, через дистиллированный адаптивный линейный фильтр с семантическими признаками



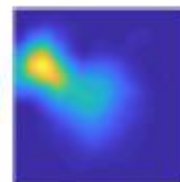
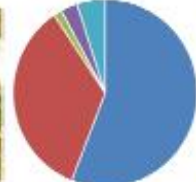
- Head
- Head & torso
- Torso
- Body & tail
- Legs & feet
- Wing



- Eye
- Mouth & nose
- Ear
- Torso
- Legs & feet

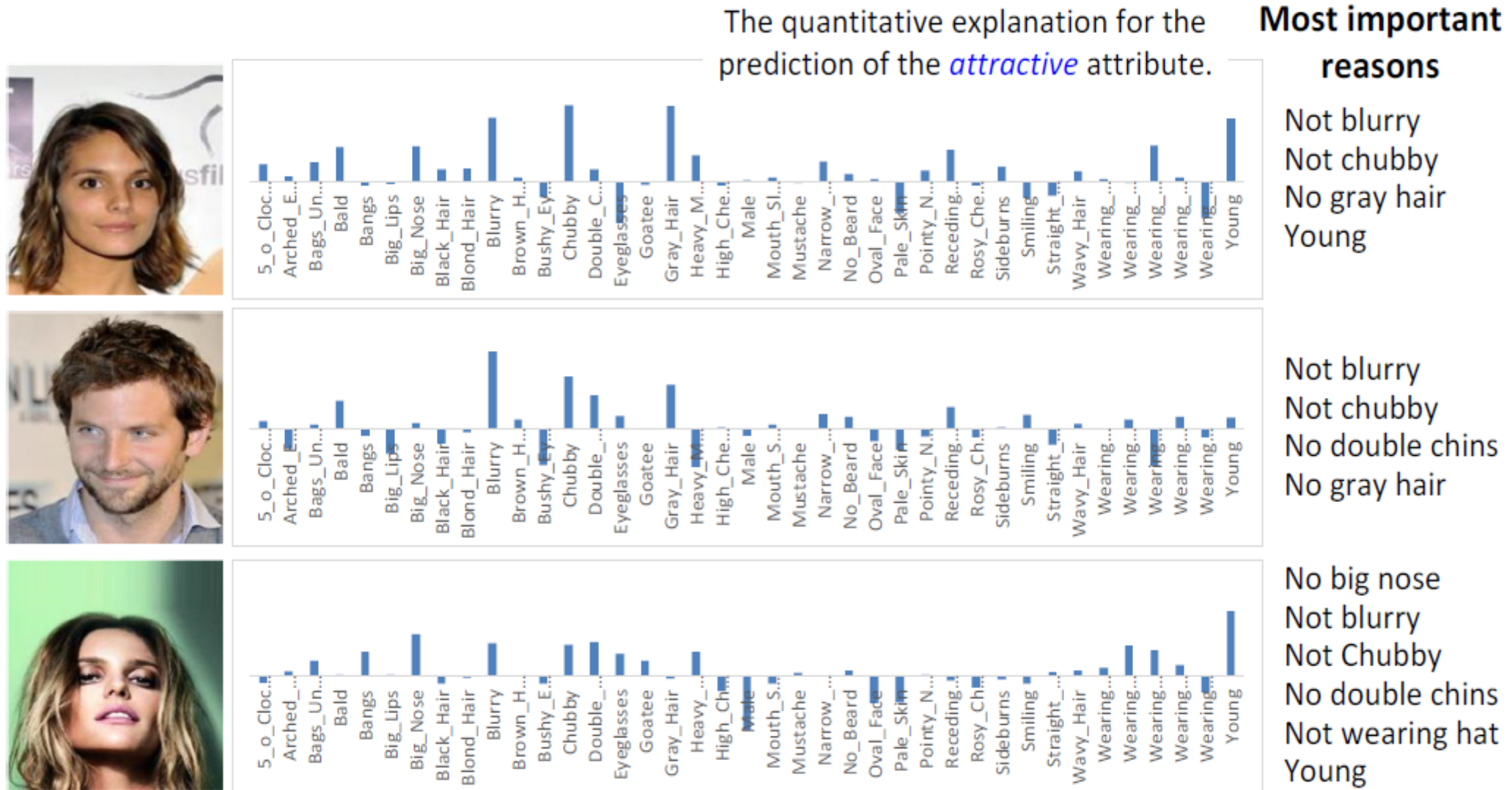


- Eye
- Mouth & nose
- Entire face
- Torso
- Legs



Explaining Neural Networks Semantically and Quantitatively

Runjin Chen, Hao Chen, Jie Ren, Ge Huang, and Quanshi Zhang



Объяснение, на что «смотрит» сеть, через дистиллированный адаптивный линейный фильтр с семантическими признаками

Открытые проблемы компьютерного зрения (2019)

Проблемы:

- Как справиться с атаками?
- Как эффективно переносить обучение в реальном мире?
- Реальных данных для практических приложений катастрофически не хватает
- Перспективные методы обучения требуют слишком больших ресурсов
- Мостик через пропасть между зрением и языком/пониманием давно перекинут, но массовый переход по нему пока не случился

Надежды:

- Новые датчики технического зрения
- Прогресс в AutoML/NAS
- Прогресс в Few-Shot/Zero-Shot
- GCN+Attention
- Прогресс в объяснении нейросетевых рассуждений
- Массовый переход нейросетевых рассуждений с уровня видимых объектов на уровень семантических конструкций (онтологий)
- Прогресс в методах RL, глубокой оптимизации и глубокого управления
- Переход к совместному решению задач зрения и управления (disentangled)

ЗАКЛЮЧЕНИЕ

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

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

3. Технологической основой для отечественных «глубоких» инженерных решений может стать создаваемая в ГосНИИАС **унифицированная программная платформа нейросетевой разработки «Платформа-ГНС».**

4. Революция в компьютерном зрении продолжается, но замедляется. 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.**

В любом случае, **ближайшие годы будут не менее интересными для исследователей, чем для инженеров-прикладников!**

Компьютерное зрение и глубокое обучение: актуальные результаты и направления развития

*Визильтер Юрий Валентинович, viz@gosniias.ru
начальник подразделения 3000 ФГУП «ГосНИИАС»,
д.ф-м.н., профессор РАН*

Спасибо за внимание!

Москва, ММРО-2019, 26.11.2019