

Как машинное обучение меняет подход к познанию?

**Новые задачи для
квантово-статистической
теории анализа данных**

**Павловский Евгений Николаевич,
к.ф.-м.н.**

**Лаборатория аналитики
поточковых данных и
машинного обучения НГУ**



План выступления

1. О машинном обучении в науках
2. Классы задач машинного обучения
3. Современные методы машинного обучения
4. Новые задачи для совместного исследования



План выступления

1. О машинном обучении в науках
2. Классы задач машинного обучения
3. Современные методы машинного обучения
4. Новые задачи для совместного исследования



Определения

- Машинное обучение – метод создания программы без непосредственного кодирования программы, а посредством обучения на примерах
- *Проведение функции через заданные точки в сложно устроенных пространствах (К.В. Воронцов)*



Машинное обучение как наука

- около 1 00 000 научных публикаций в год
- тысячи алгоритмов
- наиболее успешное направление искусственного интеллекта, вытеснившее экспертные системы и инженерию знаний
- более инженерия, нежели наука

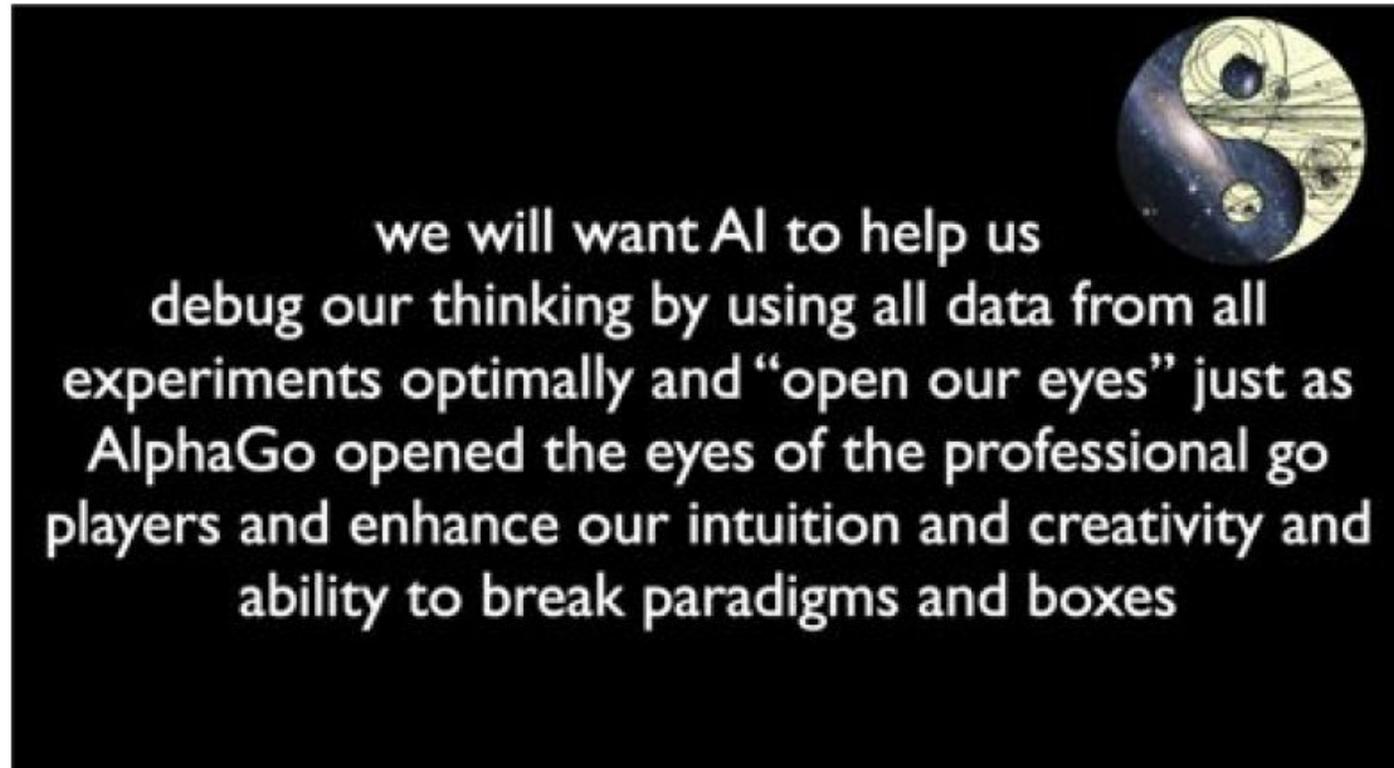


Примеры машинного обучения в физике

- Машинное обучение и квантовый отжиг нашли распады бозона Хиггса, <https://nplus1.ru/news/2017/10/19/higgs-learning>, 19.10.2017
- ФВЭ: Learning Particle Physics by Example: Location-Aware Generative Adversarial Networks for Physics Synthesis, <https://arxiv.org/abs/1701.05927>, 13.06.2017
- Астро: Generative Adversarial Networks recover features in astrophysical images of galaxies beyond the deconvolution limit, <https://arxiv.org/abs/1702.00403>, 1.02.2017



Надежды на ИИ и МО



Maria Spiropulu (Caltech)



План выступления

1. О машинном обучении в науках
2. Классы задач машинного обучения
3. Современные методы машинного обучения
4. Новые задачи для совместного исследования



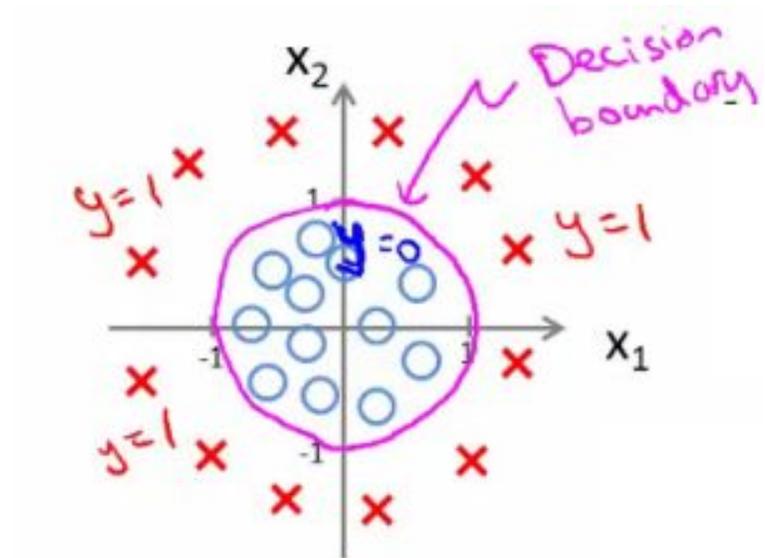
Классы задач машинного обучения

- Кластеризация
- Классификация
- Регрессия
- Цензурирование
- Снижение размерности



Классификация

	F1	F2	F3	...	F _M	Class
X1	1,1	Red	01.11.17	...	A	C1
X2	2,32	Red	07.11.17	...	B	C2
...
X _N	4,56	Black	27.11.17	...	AAB	?





Измерительные шкалы

- Номинальные (categorical): только равенство. **Green** \neq **Blue**
 - `bool`, `list`
- Порядковые: сравнение. «ВСС» > «ВСВ» (lexicographic)
 - `char`
- Интервальные: порядковые + расстояние. **30**°C больше **15**°C как
 - `date` **-5**°C больше **-20**°C
- Отношений: интервальные + ноль лекция в **2** раза дольше чем вчерашняя
 - ... вес 70кг в **1.4** раза больше чем 50кг
- Абсолютная: отношений + сложение \mathbb{N} , \mathbb{Q} , \mathbb{R}
 - `double`, `int`



Методология DIKW

Данные



Информация



Знания



Мудрость

– зафиксированные факты (данность)

– уменьшает неопределённость
(всегда есть источник и приёмник)

– дают предсказуемый результат
(рецепты)

– понимание условий использования
(«знания» в платоновском смысле)



План выступления

1. О машинном обучении в науках
2. Классы задач машинного обучения
3. Современные методы машинного обучения
4. Новые задачи для совместного исследования

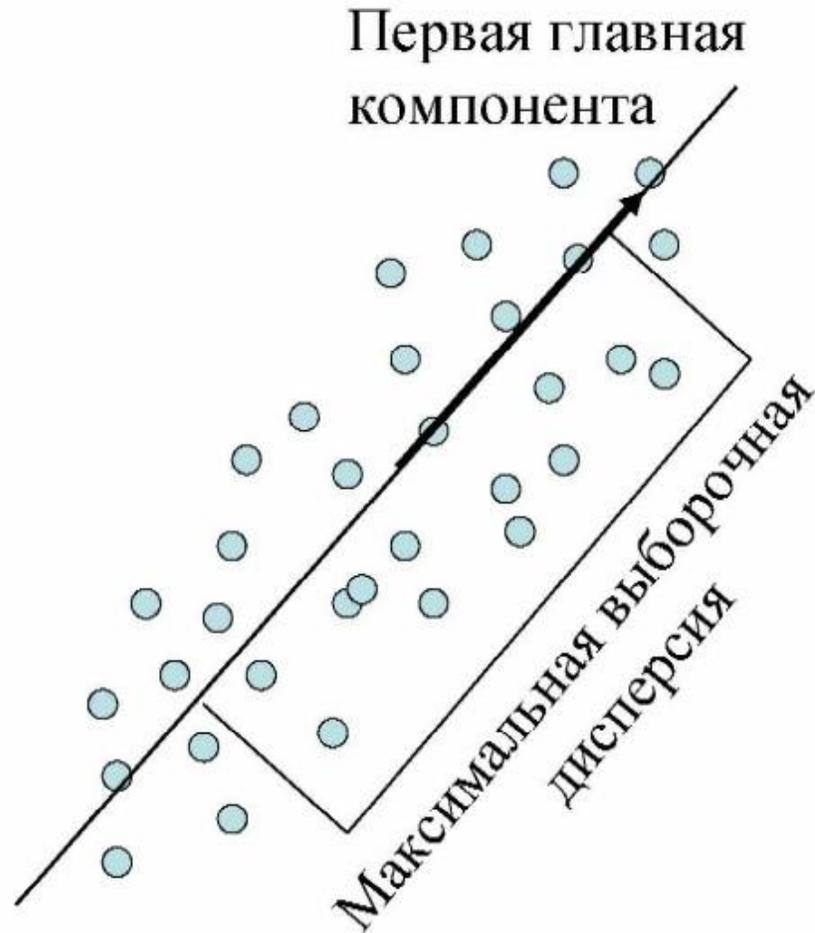


Снижение размерности

PCA, t-SNE, AE

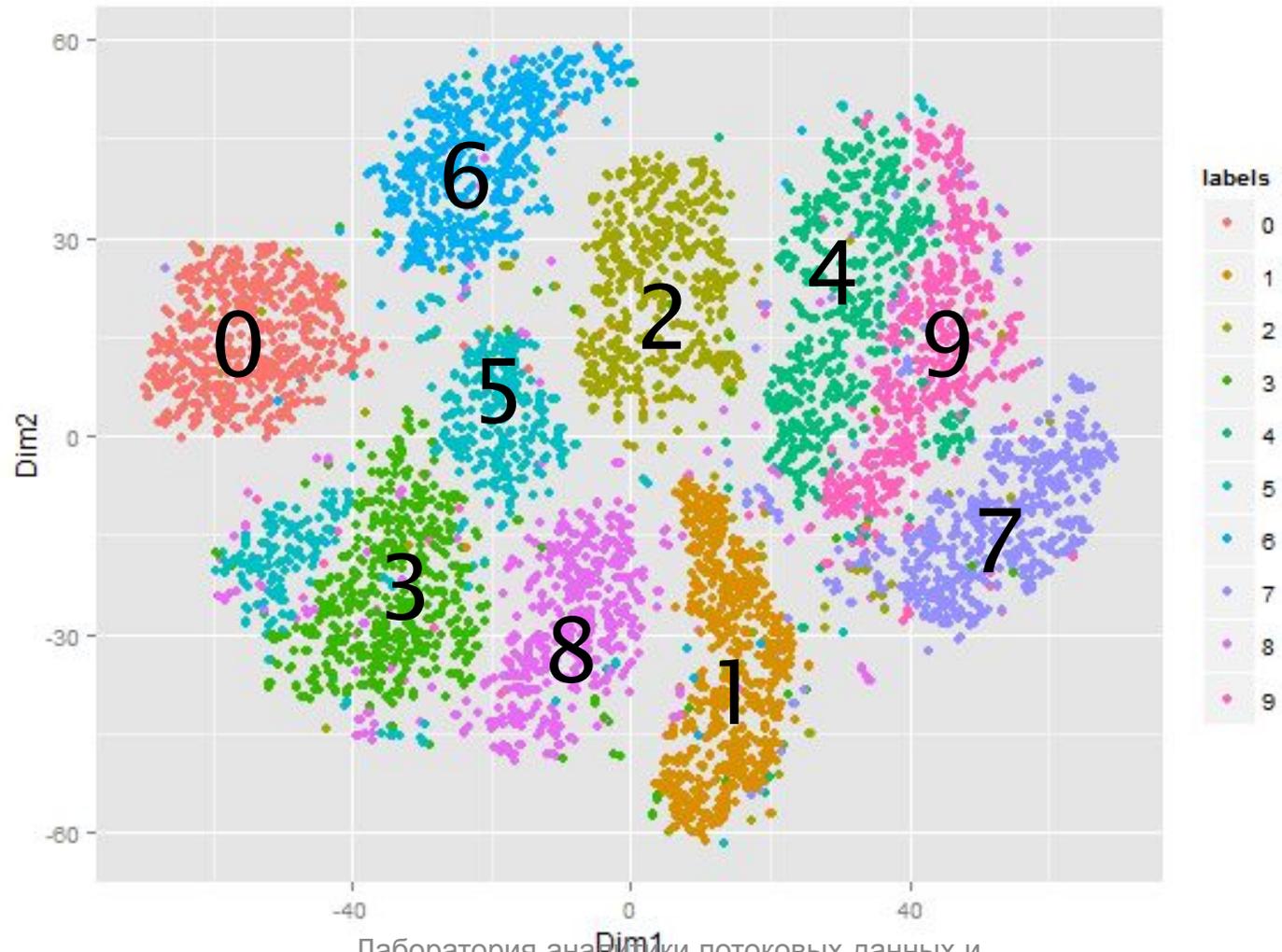


Метод главных компонент (РСА)





t-SNE





1. Маленький корпус
1000 документов (EHR) -
Кардиохирургия
2. Обработан конвейером
3. Word2vec 200 dims
4. Проекция T-SNE для визуализации





t-SNE



Кластер беременности, здесь кроме однокоренных отнесены "неделя" и аббревиатура "нед"

Кластер гипертонии: семантически близки все возможные "гипер *тонии *тензии" и "риск". Термин "эссенциальный" вызвал вопросы, но оказалось есть "эссенциальная гипертония" (95 % всех случаев).





t-SNE



Кластер протезирования.
Из интересного тут: "Тромбэктомия"
- операция по удалению
тромботических масс из ранее
установленного протеза. "Карбоникс"
- марка протезов.



t-SNE в астрофизике

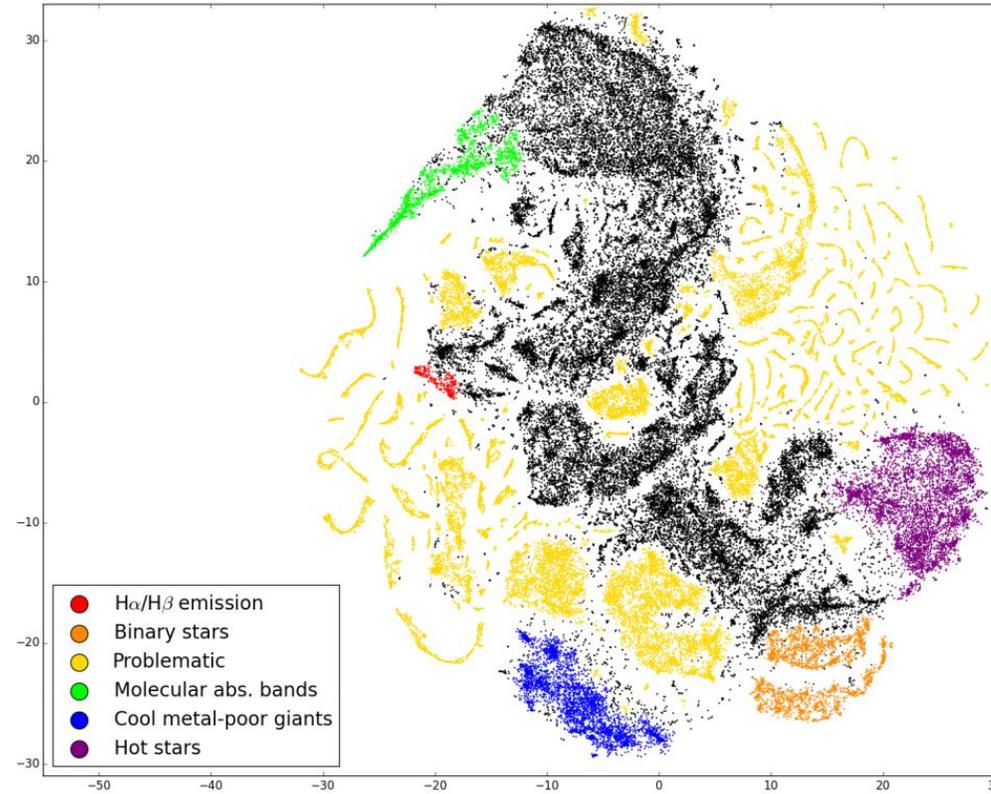


Figure 3. The result of the classification procedure, based on the projection map in Figure 2. Collections of spectra assigned to distinct categories are flagged (colour coded), the rest are black. The axes of the panels have no physical meaning, they merely span the low-dimensional projection space.

The Galah Survey: Classification and diagnostics with t-SNE reduction of spectral information

<https://arxiv.org/pdf/1612.02242.pdf>, 09.12.2016
лаборатория анализа потоковых данных и
машинного обучения НГУ (С)

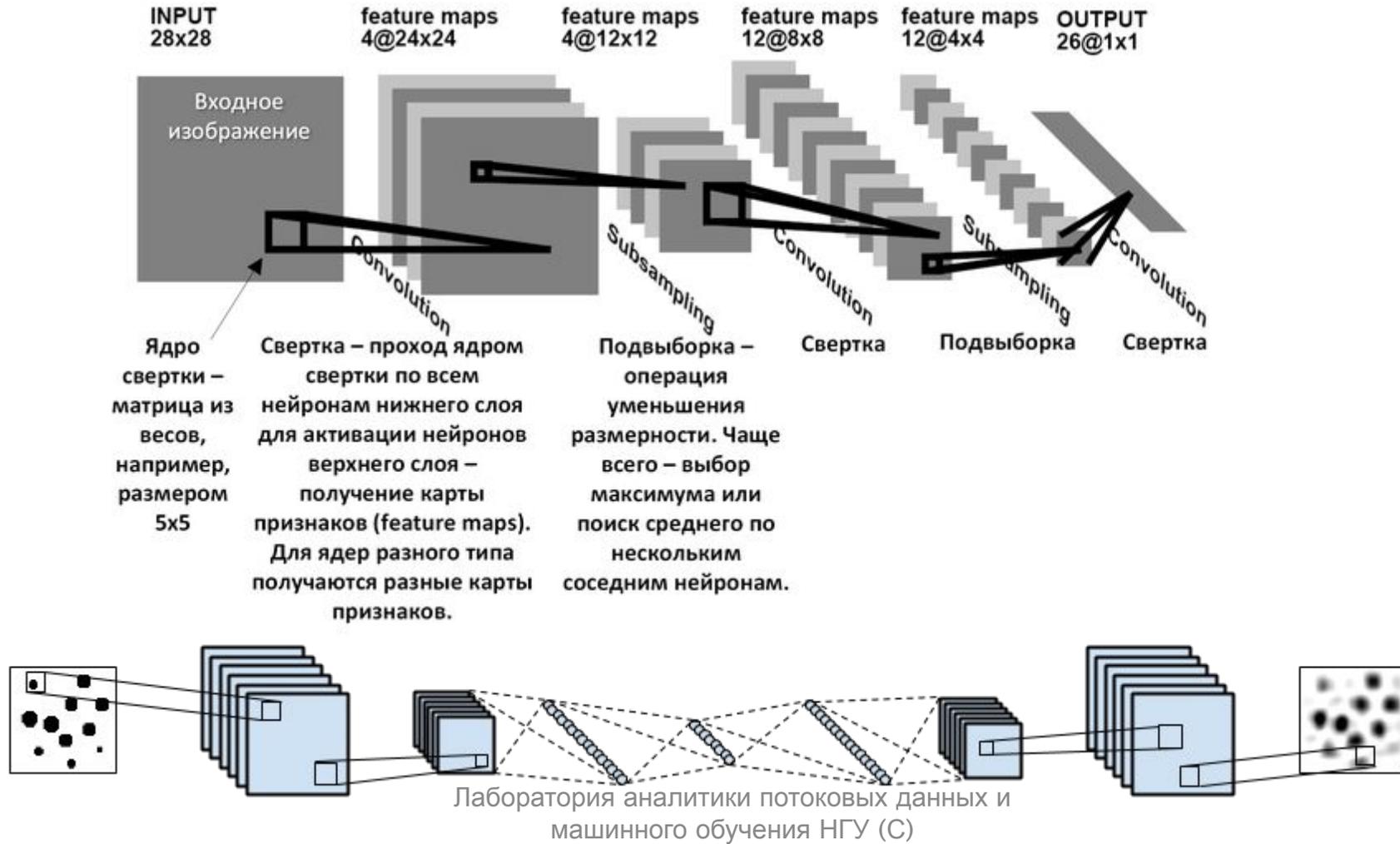


t-SNE – материалы курса МФТИ

- <https://www.coursera.org/learn/unsupervised-learning/lecture/Bn22S/mietod-t-sne>

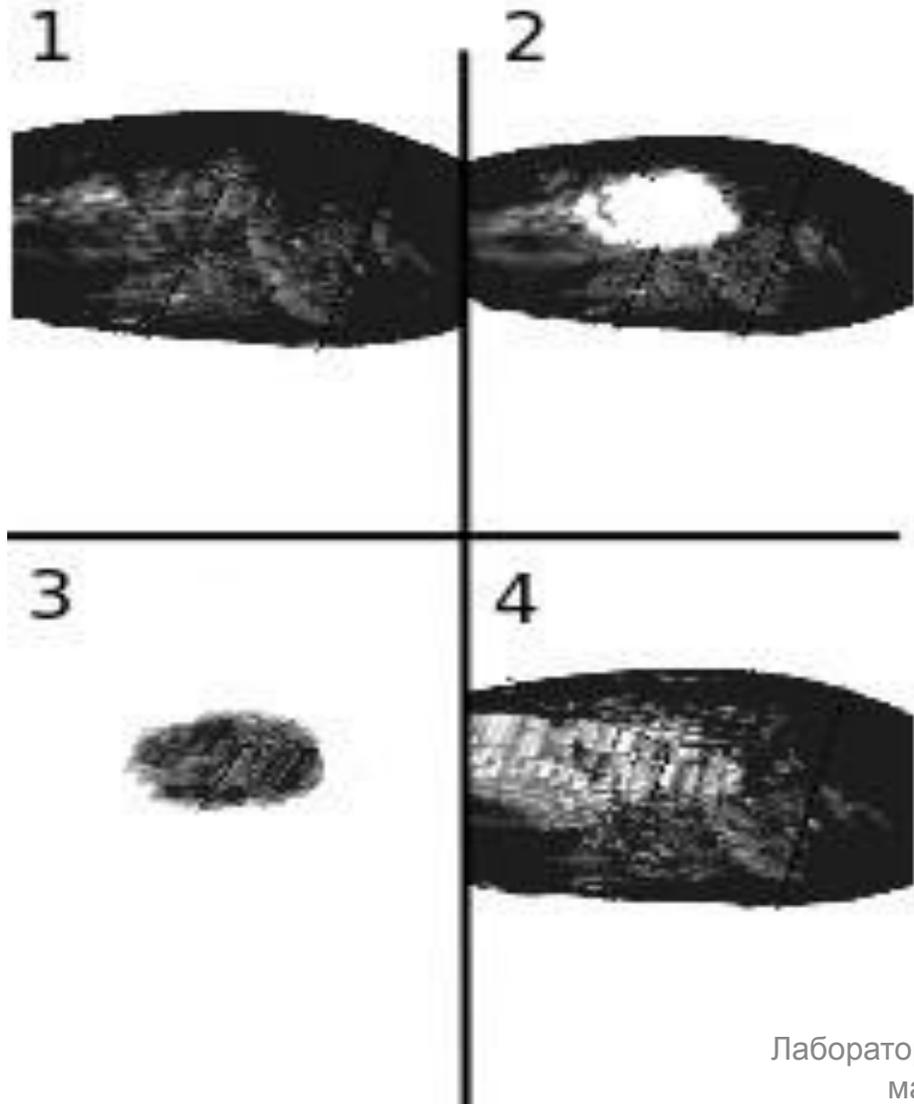


Автоэнкодер (Сверточный)





Восстановление поверхности



1 - череп без повреждения;
2 - череп со смоделированным повреждением;
3 - имплантат для поврежденной модели;
4 - восстановленная модель черепа.

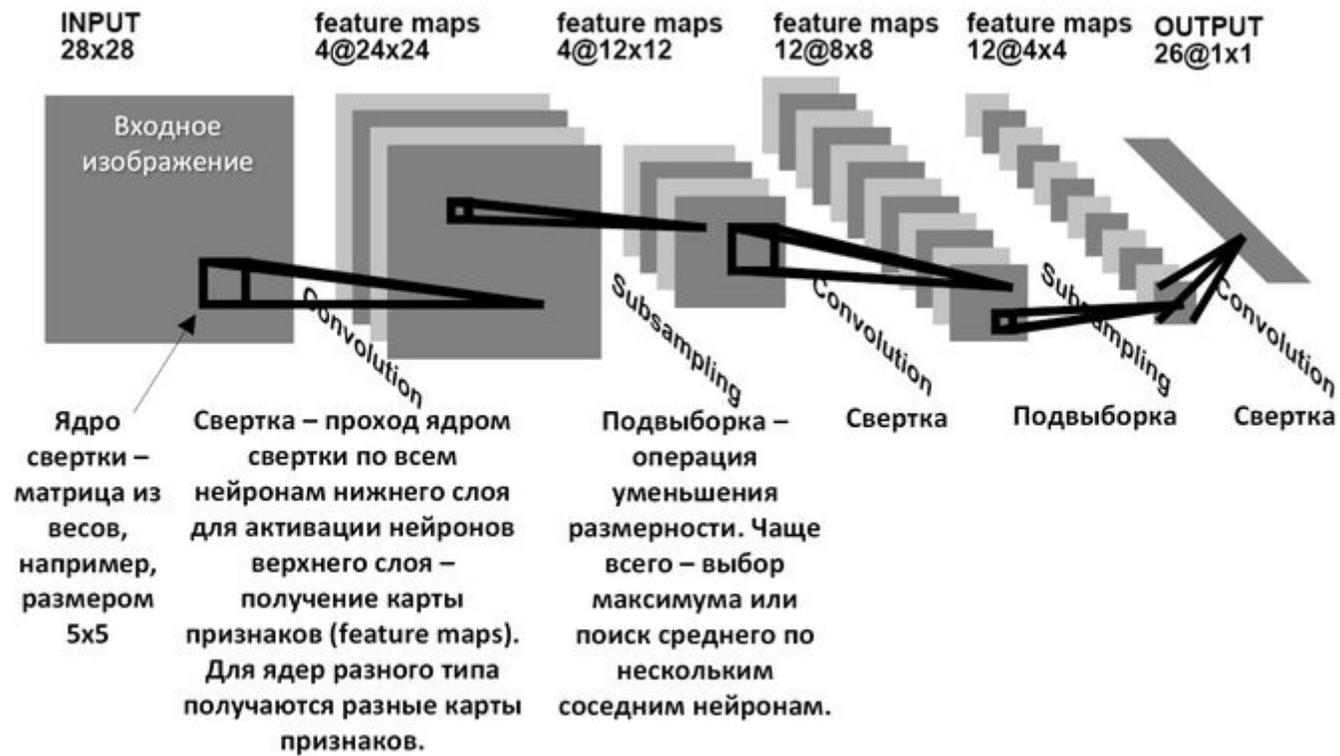


Глубокие нейронные сети

CNN, RNN, LTSM, GAN, ...

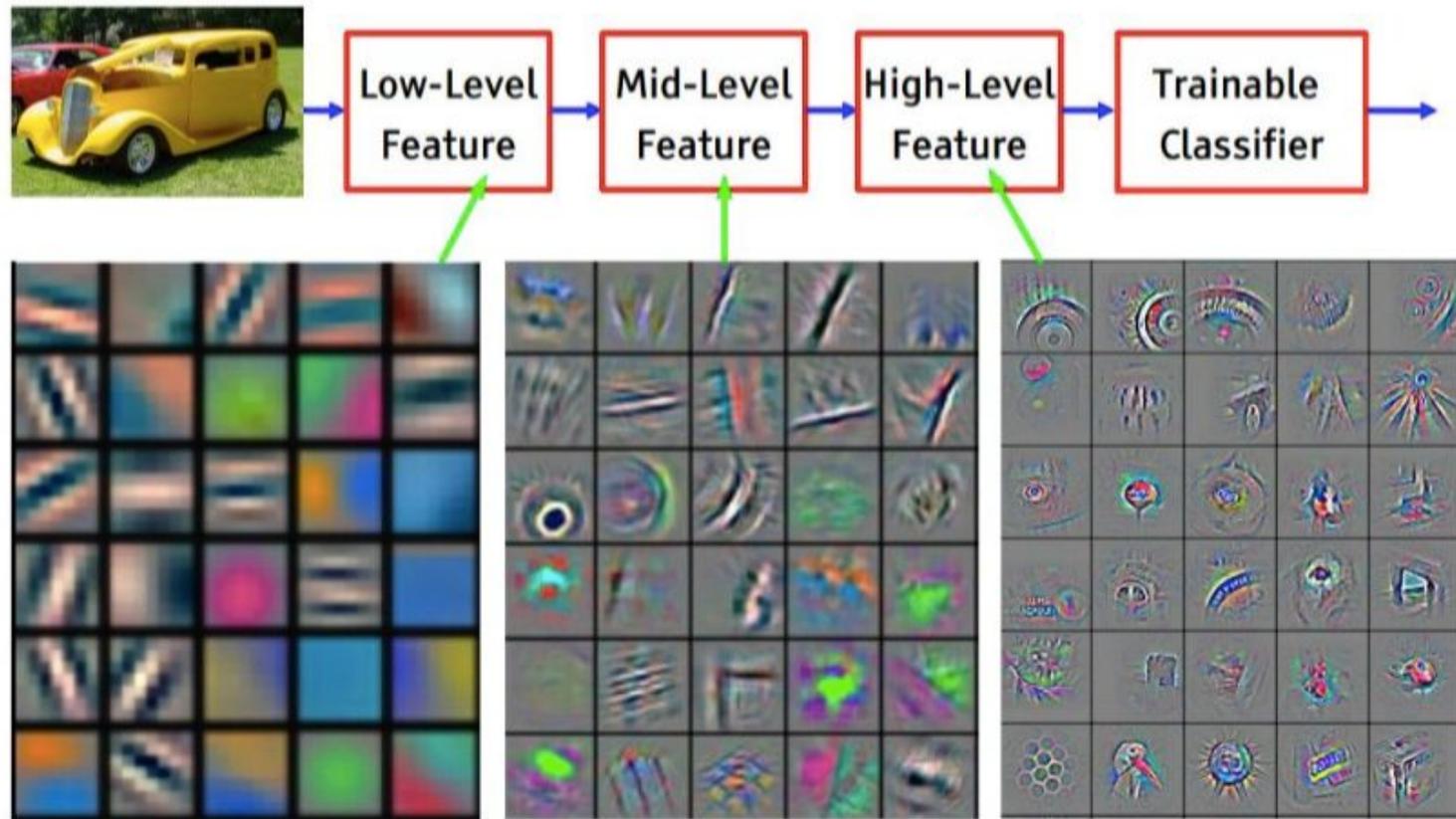


CNN





CNN – извлечение признаков



Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]



CNN

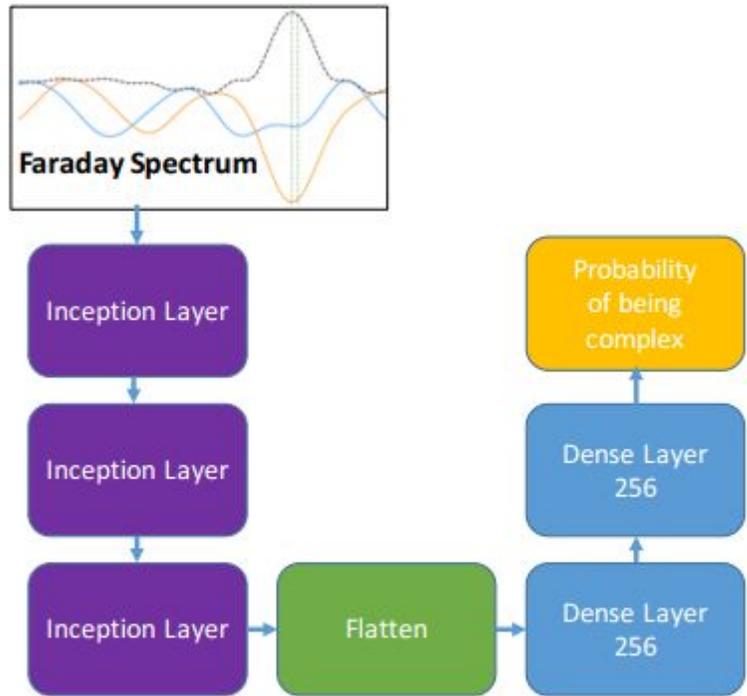


Figure 1. Three inception layer convolutional neural network (CNN) classifier. Each dense layer has an additional dropout (0.5) and activation (ReLU) layer within it.

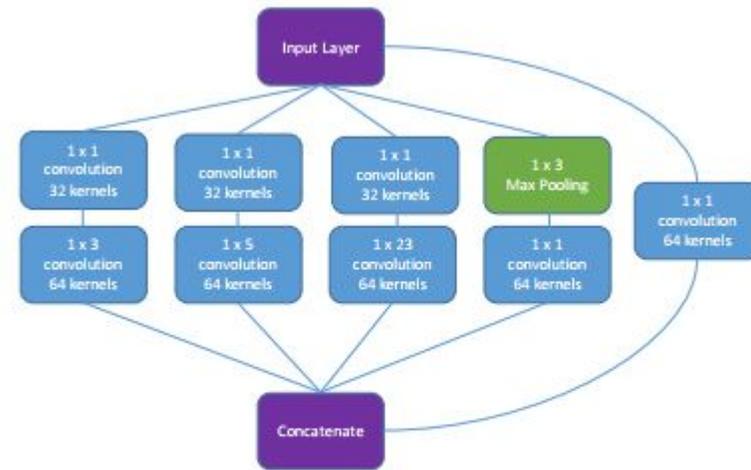


Figure 2. A zoom in of an inception layer. Each of the convolutional layers (in blue) has additional batch-normalization and activation (ReLU) layers within it.

Classifying Complex Faraday Spectra with Convolutional Neural Networks

<https://arxiv.org/abs/1711.03252>, 9.11.2017



CNN для поиска гравитационных линз

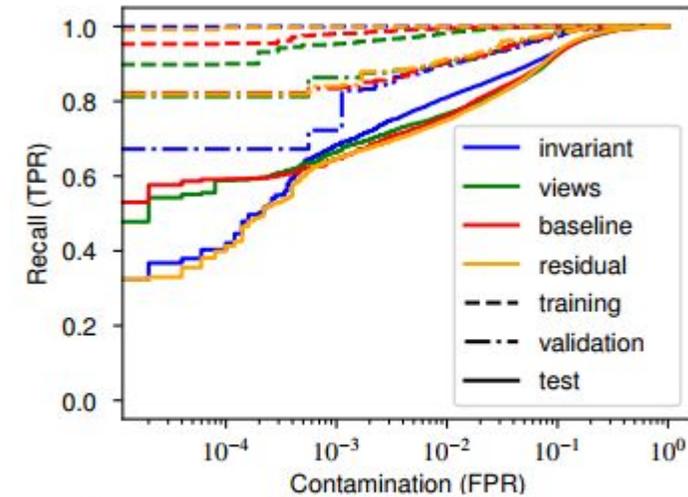


Fig. 10: Logarithmic ROC curves on **ground-based** data. Training (dotted line), validation (half dotted line) and test (solid line) score of all four architectures. Data comes from the best of five runs in terms of validation set score.

Deep Convolutional Neural Networks as strong gravitational lens detectors

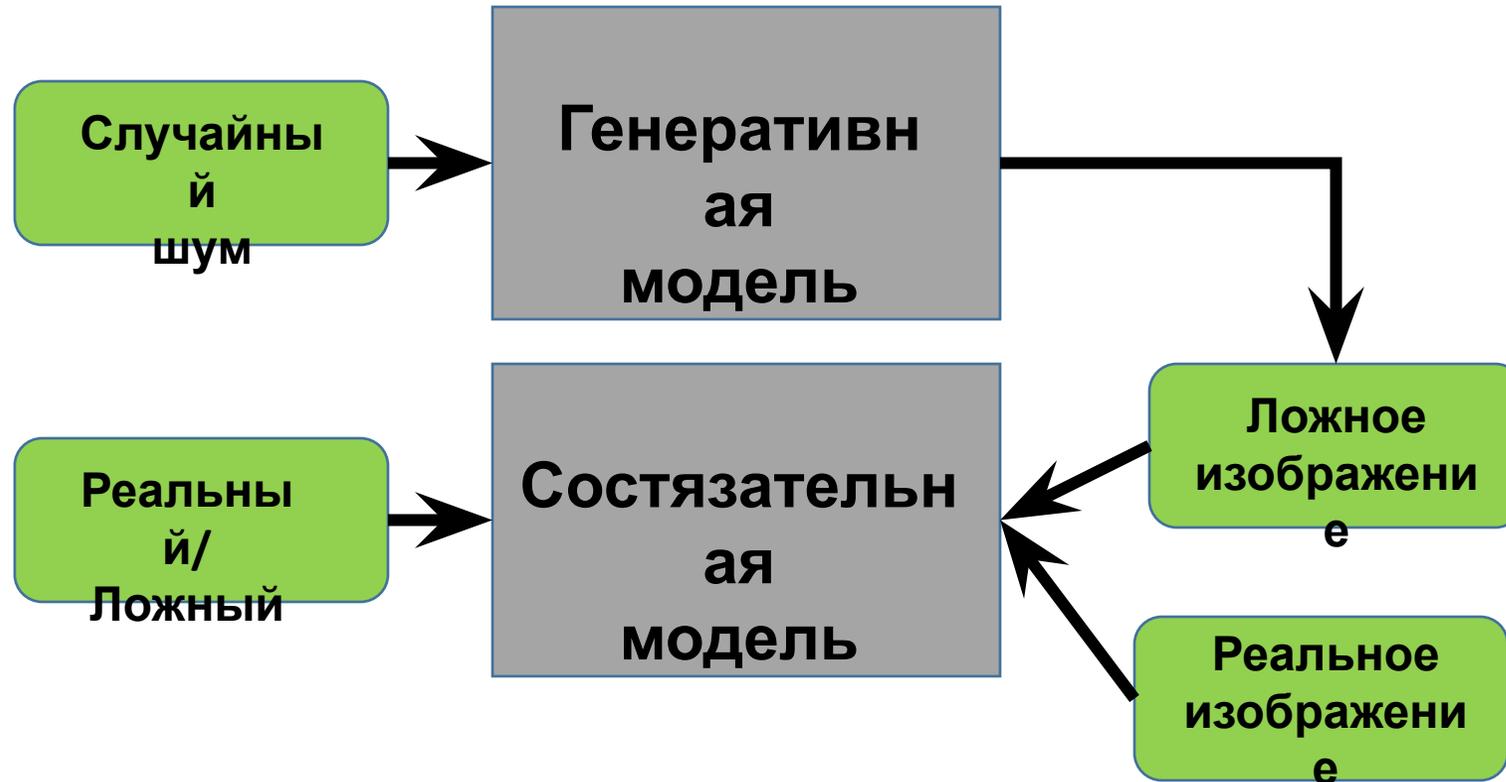
<https://arxiv.org/pdf/1705.07132.pdf>, 19.05.2017

Лаборатория аналитики потоковых данных и машинного обучения НГУ (С)



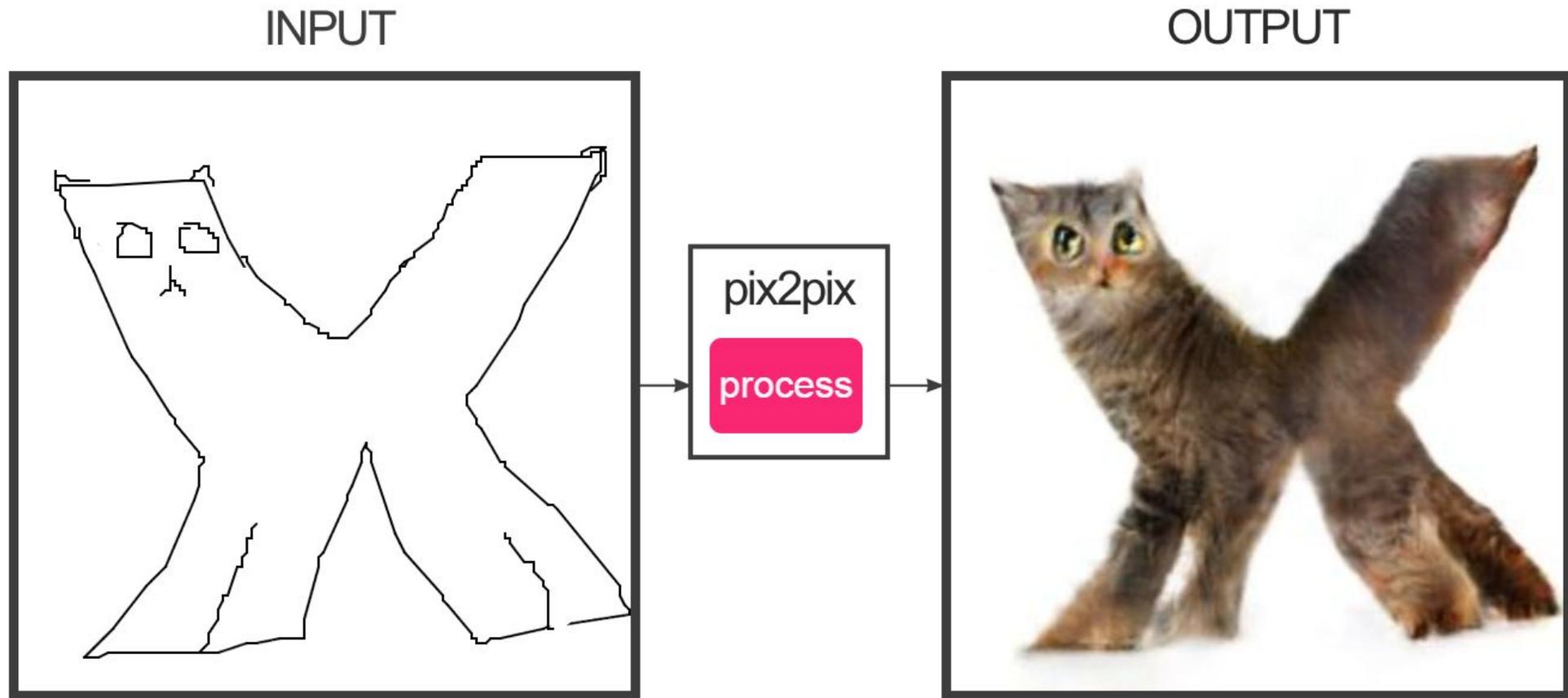
GAN (Архитектура)

Генеративная состязательная сеть





GAN (pix2pix)



Local Awareness GAN

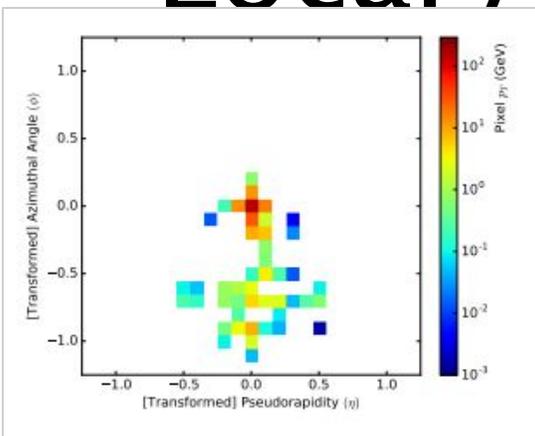
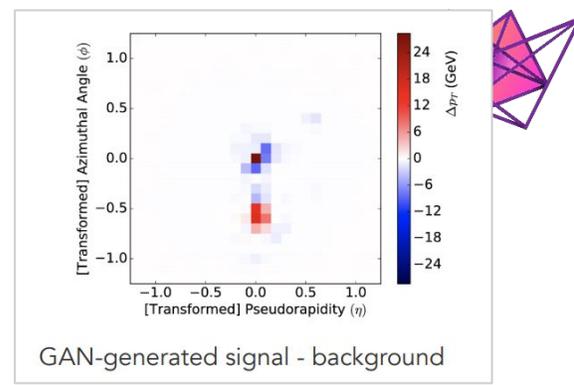
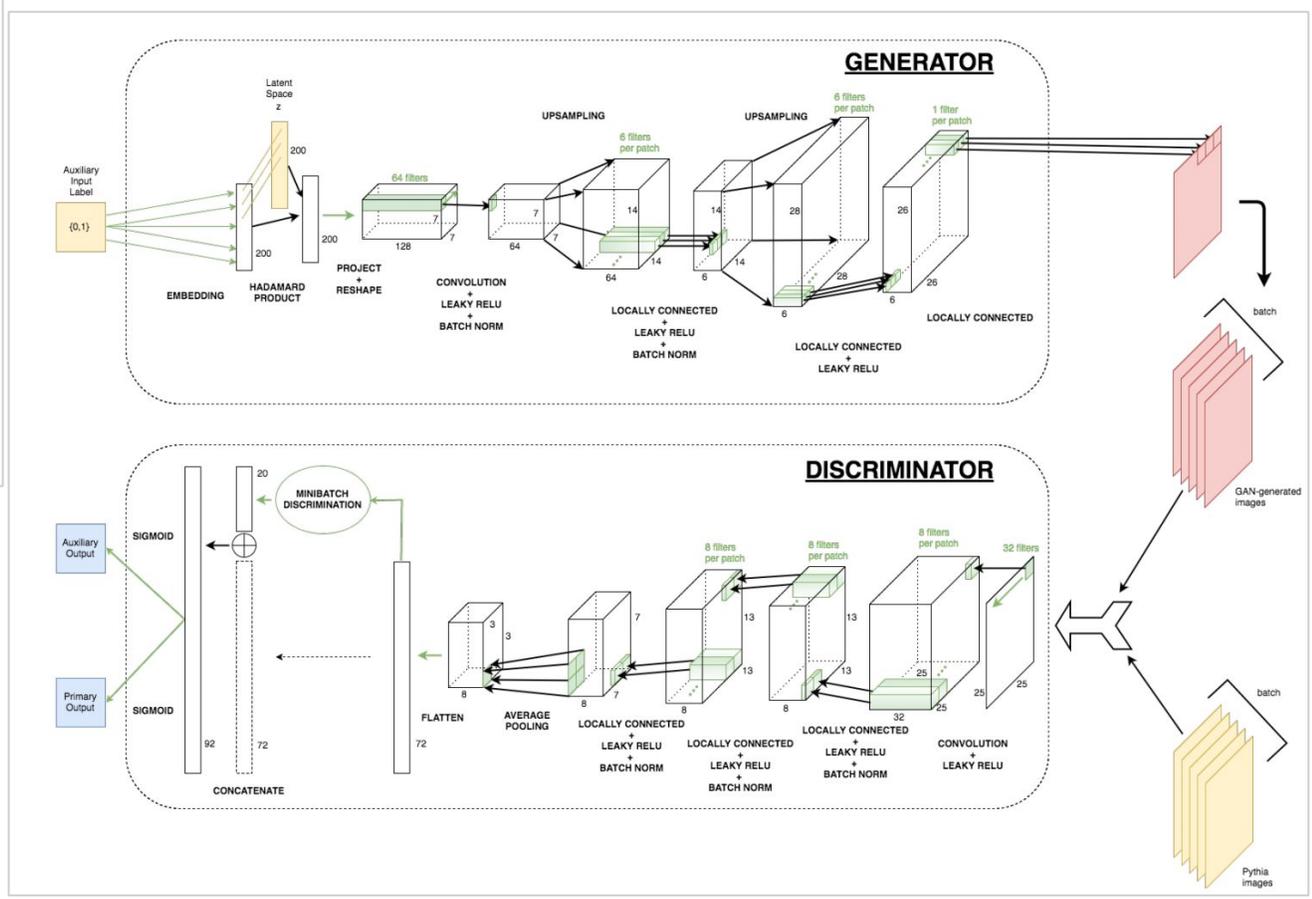
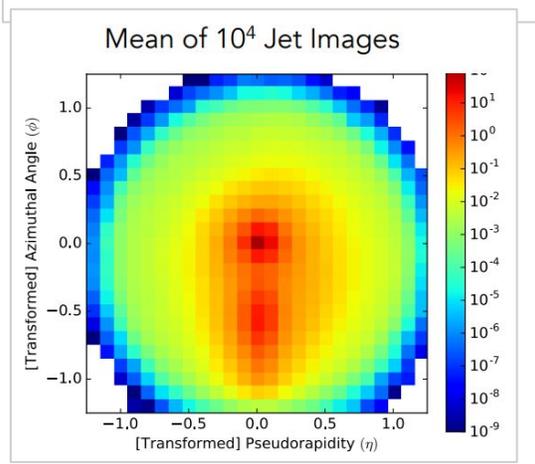
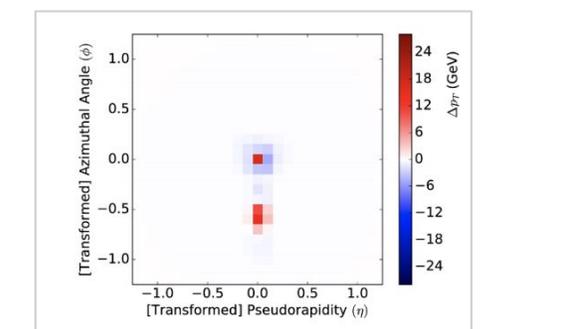


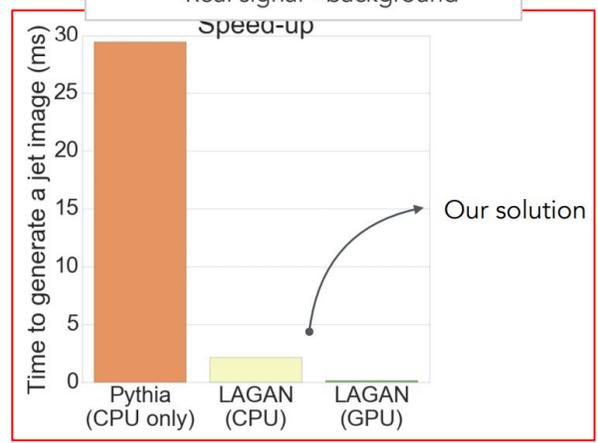
Figure 1: A typical jet image.



GAN-generated signal - background



Real signal - background





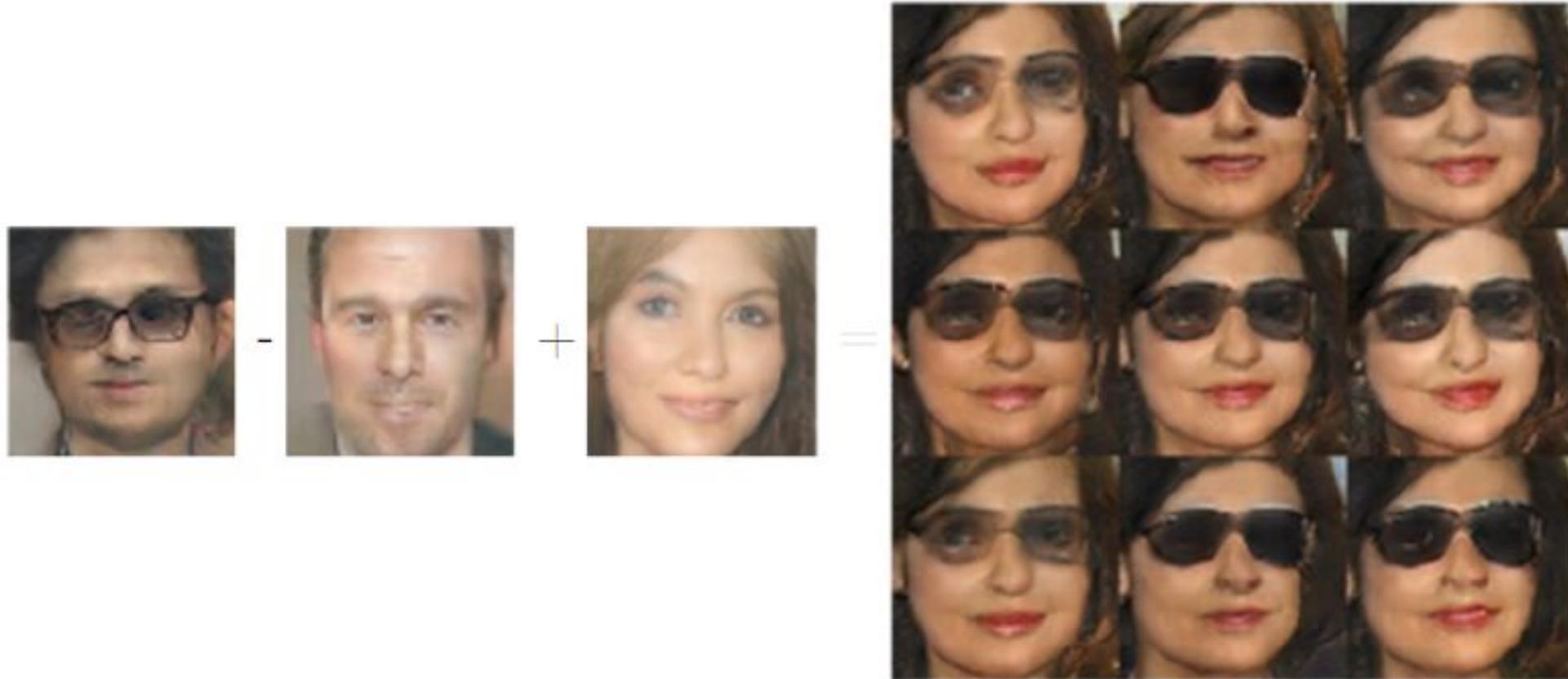
Distributed Representation

Слово представляется вектором в многомерном пространстве





Distributed Representation





План выступления

1. О машинном обучении в науках
2. Классы задач машинного обучения
3. Современные методы машинного обучения
4. Новые задачи для совместного исследования



Новые задачи для квантово-статистической теории анализа данных



Смешанные состояния

- Задача:
 - Учесть влияние наблюдателя и прибора на измерения
 - Метод: при измерении восстановить исходную матрицу плотности макросистемы «измеряемое-прибор-наблюдатель»
- Зачем?
 - реализовать запутанные системы и смешанные состояния
 - использовать матрицу плотности вместо исходных данных



Нецифровые шкалы

- Множество подмножеств: $X \subset N$
- Какими алгоритмами работать с такими измерениями?
 - Инвариантность относительно выбранной шкалы
 - Проследить все операции с этим типом данных



ИСТОЧНИКИ

- <https://nplus1.ru/news/2017/10/19/higgs-learning>
 - Машинное обучение и квантовый отжиг нашли распады бозона Хиггса
- <https://arxiv.org/abs/1701.05927>
 - Learning Particle Physics by Example: Location-Aware Generative Adversarial Networks for Physics Synthesis
- <https://arxiv.org/abs/1711.03252>, 9.11.2017
 - Classifying Complex Faraday Spectra with Convolutional Neural Networks
- <http://www.nsu.ru/xmlui/bitstream/handle/nsu/13448/08.pdf>
 - Павловский Е. Н., Пакулич Д. В., Поспелов С. О. Восстановление 3D-модели дефекта черепа на основе глубоких нейронных сетей // Вестн. НГУ. Серия: Информационные технологии. 2017. Т. 15, № 3. С. 74–78. DOI 10.25205/1818-7900-2017-15-3-74-78. ISSN 1818-7900.
- <https://arxiv.org/pdf/1705.07132.pdf>, 19.05.2017
 - Deep Convolutional Neural Networks as strong gravitational lens detectors



Контакты

к.ф.-м.н.
Павловский Евгений Николаевич
зав. лабораторией аналитики
потоковых данных и машинного
обучения НГУ
+79139117907
pavlovskiy@post.nsu.ru

