

Использование
нейронного
автокодировщика для
генерации показаний
поверхностных детекторов
Telescope Array

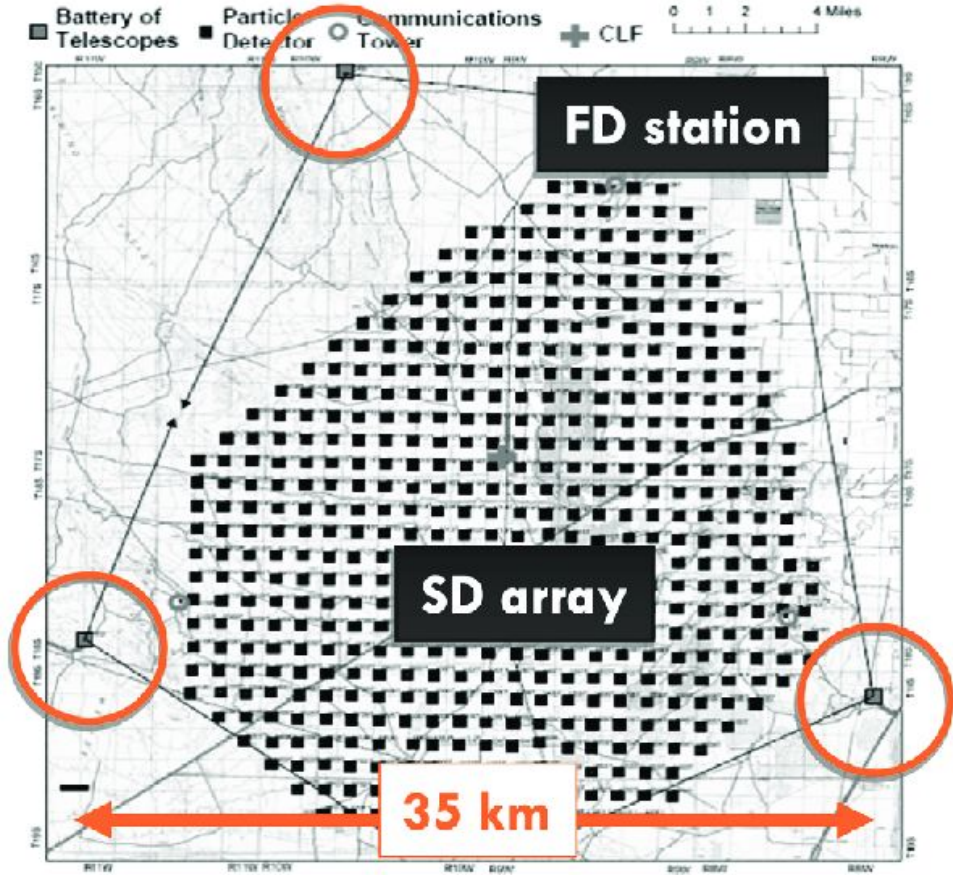
Фитагдинов Робер
ИЯИ РАН, МФТИ
DLCP 2025

Telescope Array

Эксперимент содержит из 507
сцинтилляционного детектора.

Расстояние между ними
составляет 1,2 км.

Детекторы измеряют время срабатывания и сигнал (MIP) полученный в ходе прилета ШАЛ.



Цели и мотивация

- Латентное представление может помочь в поиске аномальных данных в эксперименте (Unsupervised anomaly detection)
- С помощью латентного представления можно сделать модельно независимое представление данных (МК данные не согласуются с реальными)
- Сжатое представление может помочь улучшить алгоритмы реконструкции параметров ШАЛ.

Данные

данные содержат 6 каналов:

- координаты детектора (x,y,z)
- интегральный сигнал
- время реконструированного плоского фронта
- время прилета ШАЛ

Сортировка по возрастанию времени плоского фронта

Нормировка каждого канала со средним и дисперсией 0 и 1

Данные разбиваются на нормальные (на которых модель училась - протонные) и на аномальные (модель в обучении их не видит - аномальные)

AutoEncoder

Декодер и Энкодер базируются на LSTM

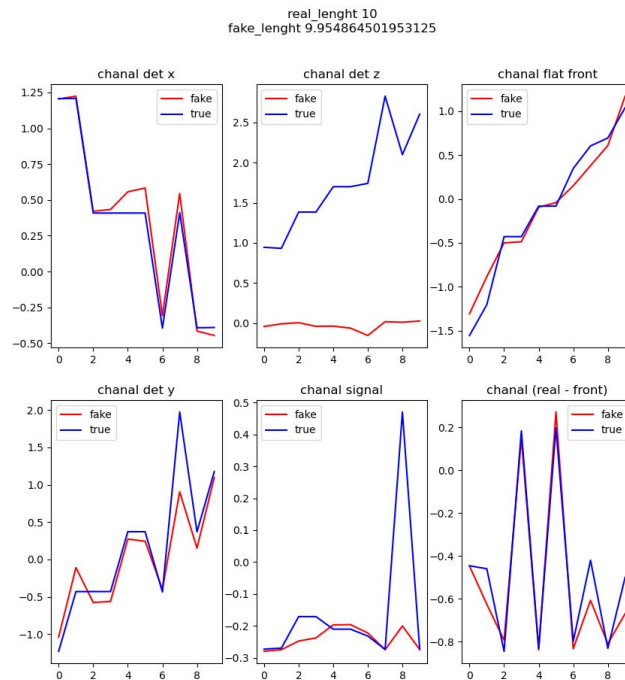
Размерность латентного пространства: 32

Размерность скрытого представления: 512

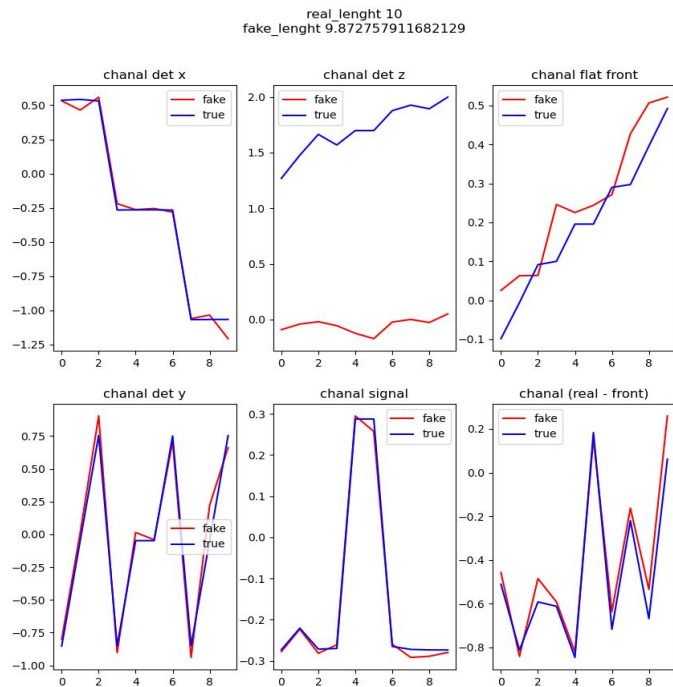
Выход декодера:

- предсказанная длина последовательности
- реконструированные данные (все **кроме z** координаты)

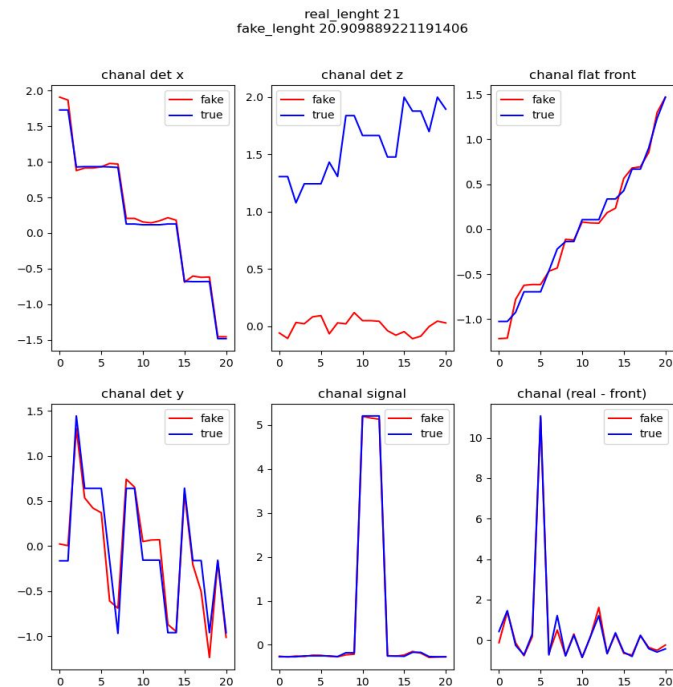
Трансформеры в качестве энкодера работают заметно хуже



Результат реконструкции

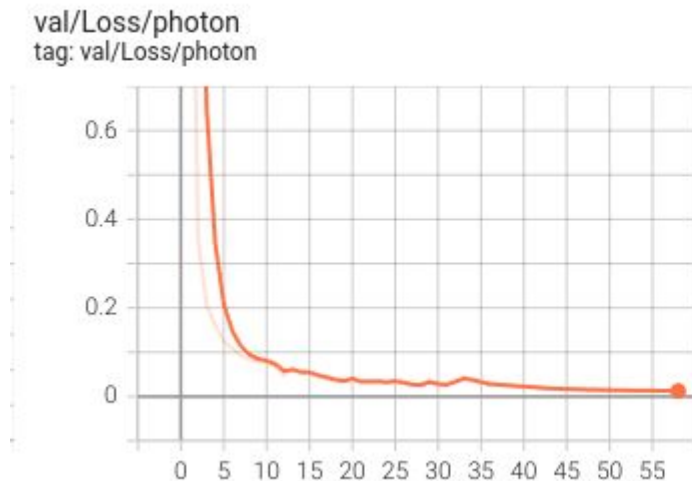


Фотон



Протон

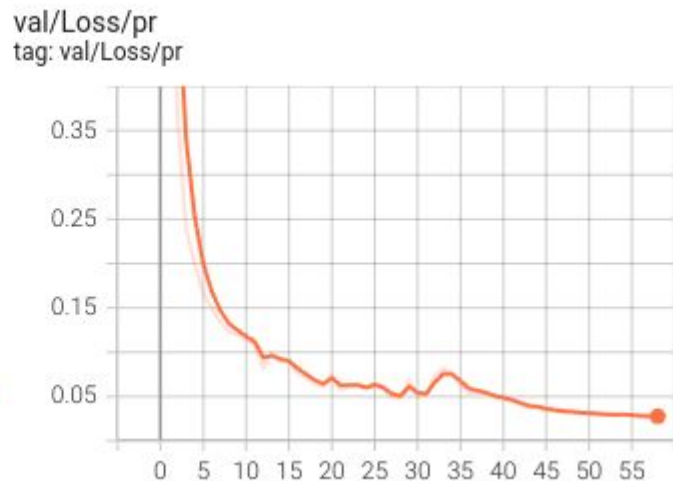
Ошибка реконструкции по частицам



Фотон:

ошибка кол-ва активных детекторов: **8e-6**

ошибка реконструкции: **0.013**



Протон:

ошибка кол-ва активных детекторов: **2e-6**

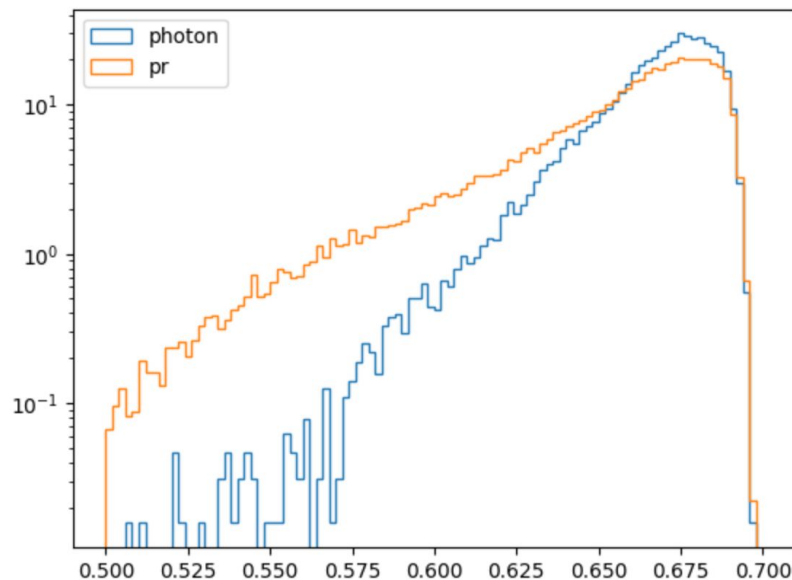
ошибка реконструкции: **0.027**

Анализ кластеризации по латентному пространству

Все применяемые методы анализа латентного пространства говорят одно:

Фотон - это наиболее типичный протон

Проблема в запоминании данных, а не в понимании физики процесса!



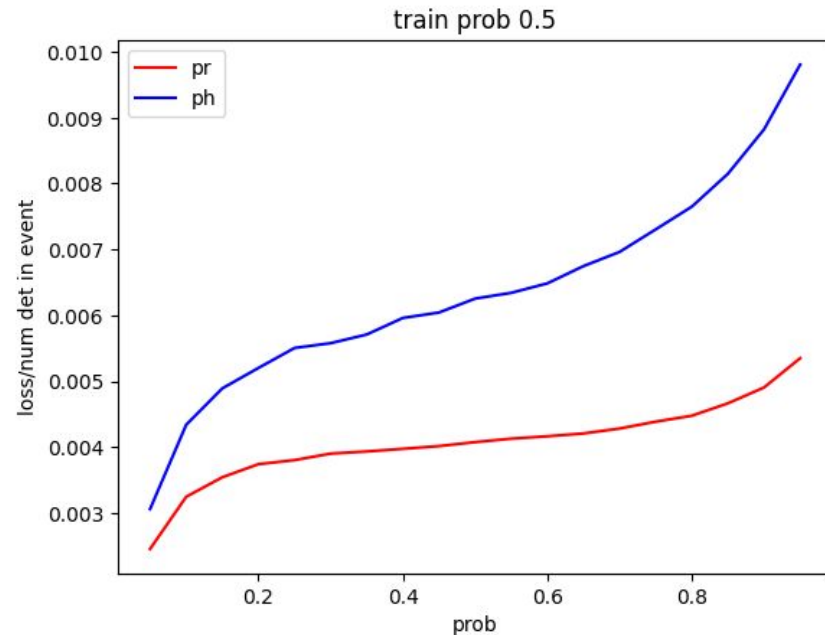
Пример работы IsolationForest

Что делать?

Уходить от автокодировщиков или модернизировать их!

Потенциально возможный вариант:

- Mask autoencoder (MAE)
- Mask autoencoder + AE



Ошибка MAE для протонов и фотонов от вероятности маскирования

Спасибо за внимание

[github](#)

[email](#)

