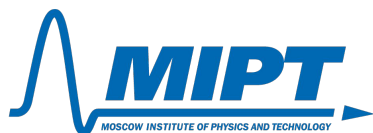




Генерация показаний наземного детектора эксперимента Telescope Array и поиск аномалий с помощью нейронных сетей

Фитагдинов Роберт
robertfitagdinov@gmail.com

2023



Актуальность

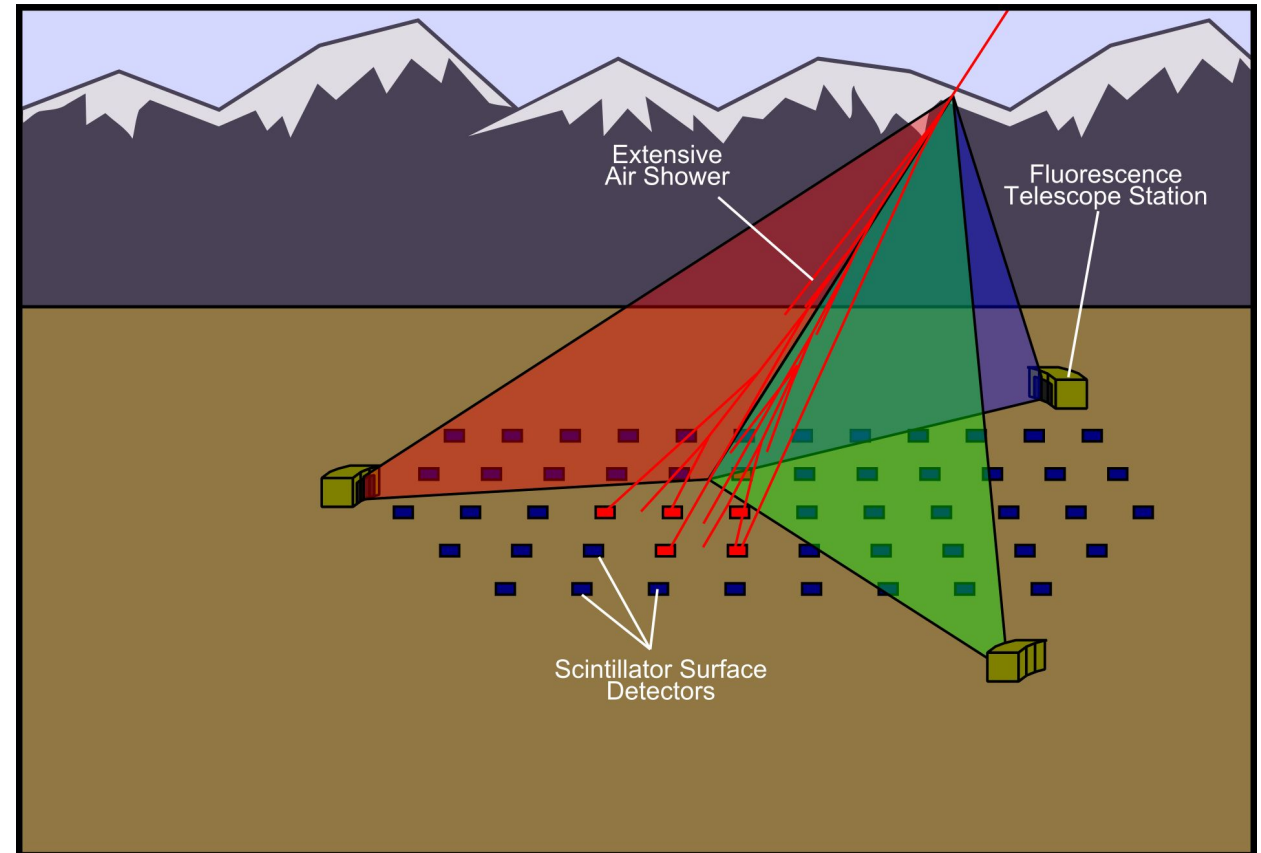
Космические лучи высоких энергий представляют уникальную информацию из далеких галактик

- Энергии превышающие 10^{18} эВ, что на порядки большей максимальной достигнутой энергии на ускорителях
- Генерация данных с помощью генеративно состязательных сетей на порядки быстрее, чем генерация с помощью метода Монте-Карло
- Алгоритм поиска аномалий может быть применим для анализа моделирования другими методами и поисками интересной физики в эксперименте

Эксперимент Telescope Array

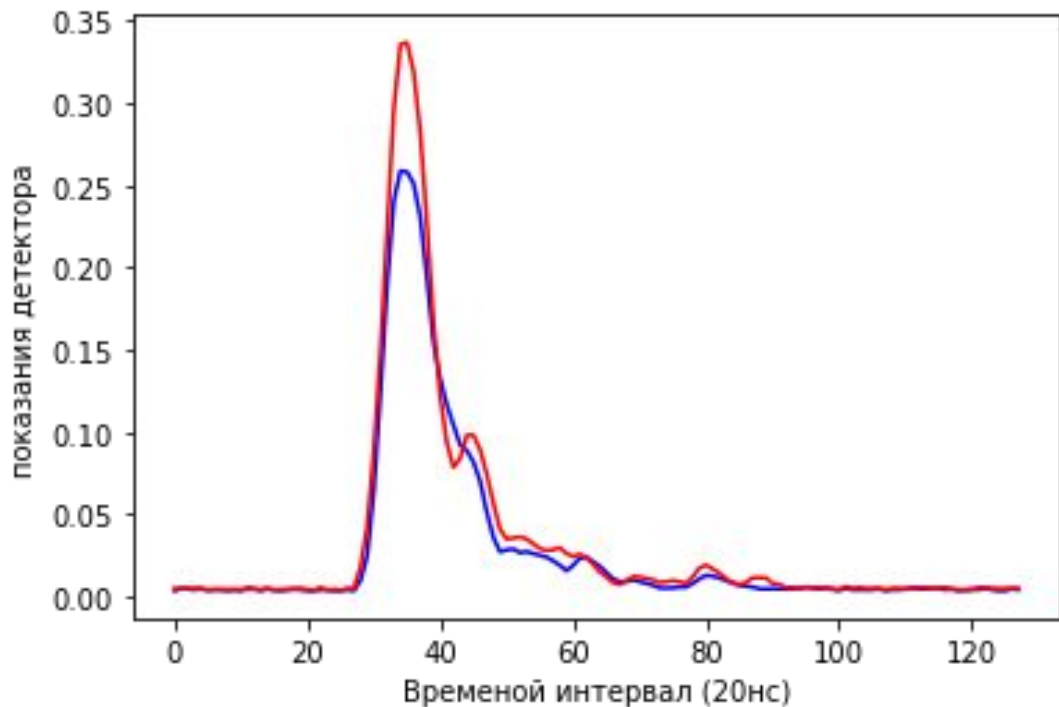
Обсерватория представляет собой гибридную систему детекторов, состоящую из массива из 507 сцинтилляционных поверхностных детекторов, которые измеряют распределение заряженных частиц на поверхности Земли.

Поверхностные детекторы равномерно распределены по решетке 762 км^2 с расстоянием 1,2 км между каждым устройством.



https://en.wikipedia.org/wiki/Telescope_Array_Project#/media/File:TelescopeArray.svg

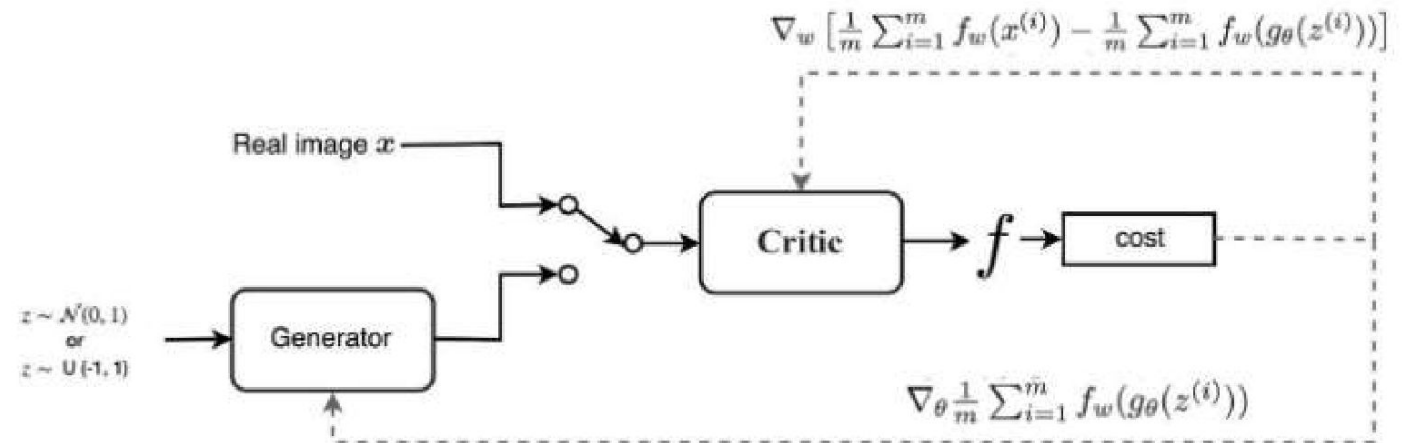
Данные



- Данные имеют размер (472351, 128, 2)
2 - показания двух каналов детектора
128- количество бинов, на которые разбивается интервал в 20нс
472351- количество смоделированных данных методом МК, используемых для обучения модели

Генеративно-состязательная сеть вассерштейна с градиентным штрафом (WGAN-GP)

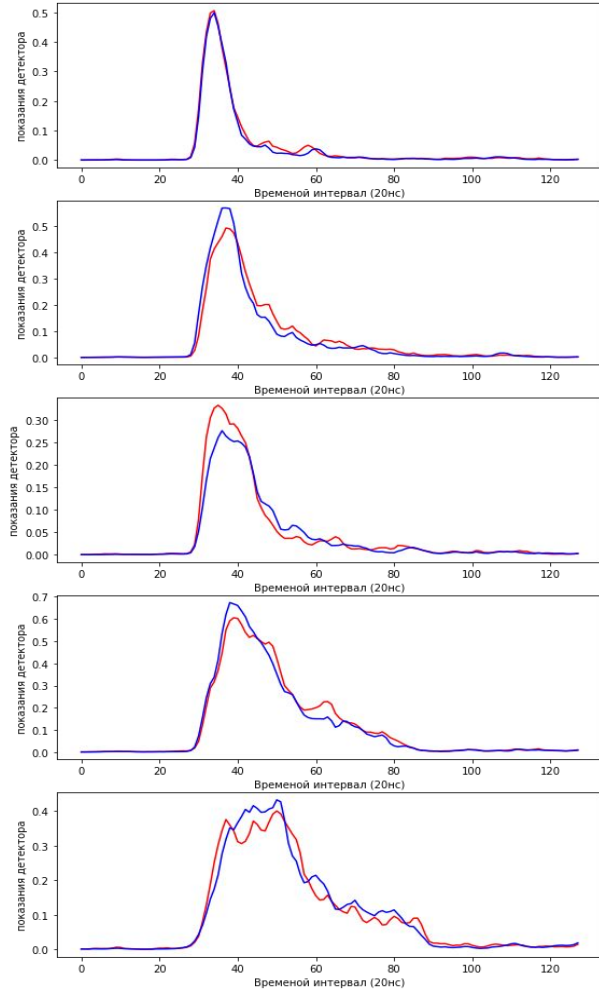
- WGAN состоит из двух сетей: генератор и критик. Суть обучения GAN состоит в конкуренции этих двух моделей. Генератор производит реалистичные данные, критик пытается отличить реальные от сгенерированных.



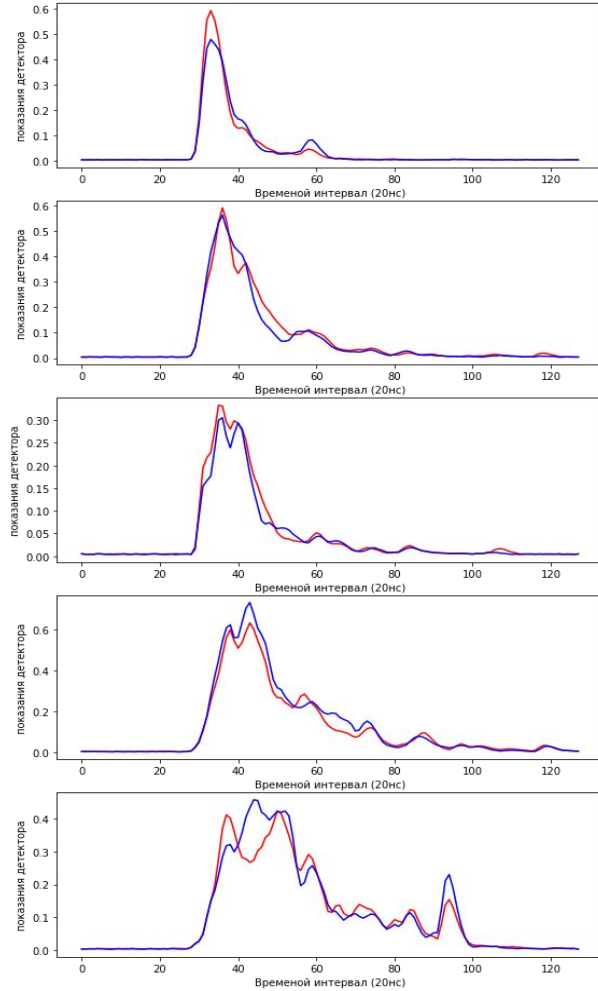
$$L = \underbrace{\mathbb{E}_{\hat{x} \sim P_g} [D(\hat{x})] - \mathbb{E}_{x \sim P_r} [D(x)]}_{\text{Original critic loss}} + \lambda \underbrace{\mathbb{E}_{\hat{x} \sim P_g} \left[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2 \right]}_{\text{Our gradient penalty}}.$$

Результаты

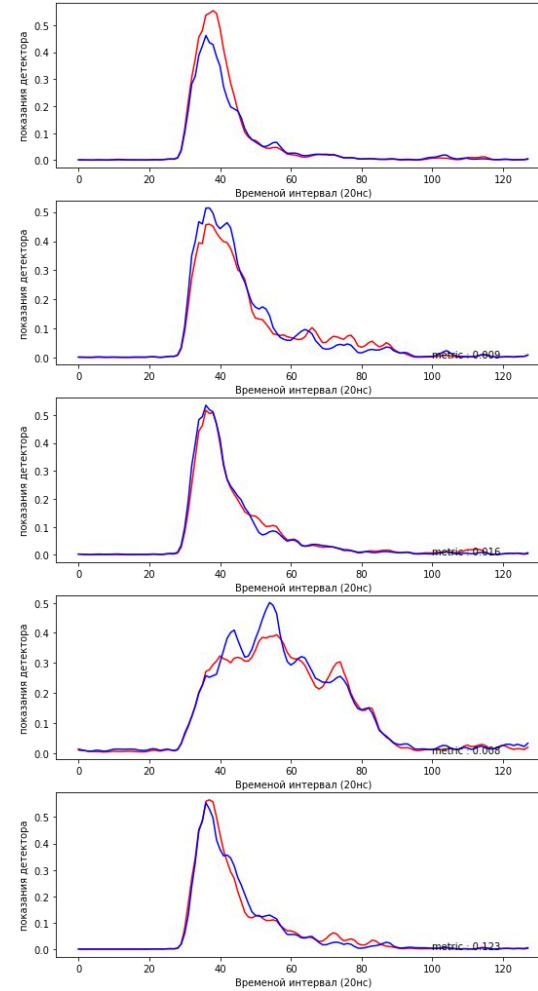
сгенерированные



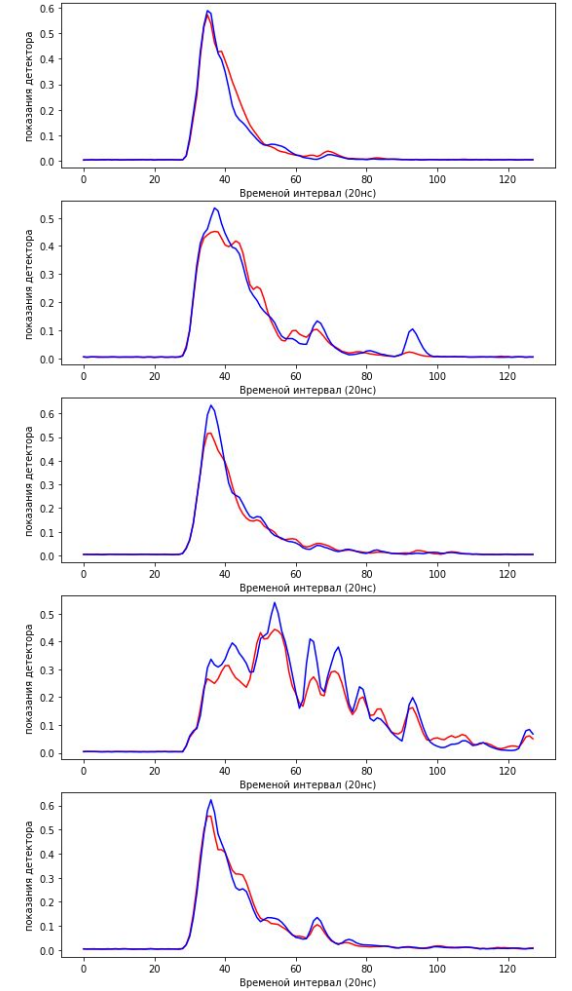
реальные



сгенерированные

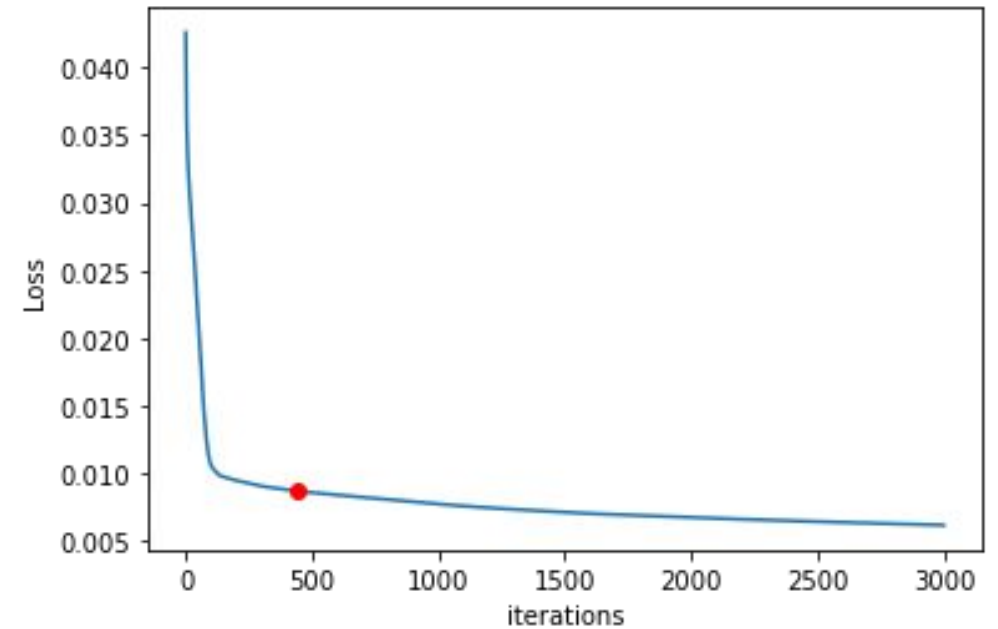


реальные



Поиск аномалий

- Для поиска аномалий был написан алгоритм, получающий на вход изображение и подбирающий шум для генератора посредством градиентного спуска с ранней остановкой. Ошибка состояла из 2х частей:
 - 1 Сумма по всем элементам разницы реальных и сгенерированных данных
 - 2 разница показаний детектора. Эта ошибка имела вес 0,001.



Функция ошибок для поиска оптимального шума

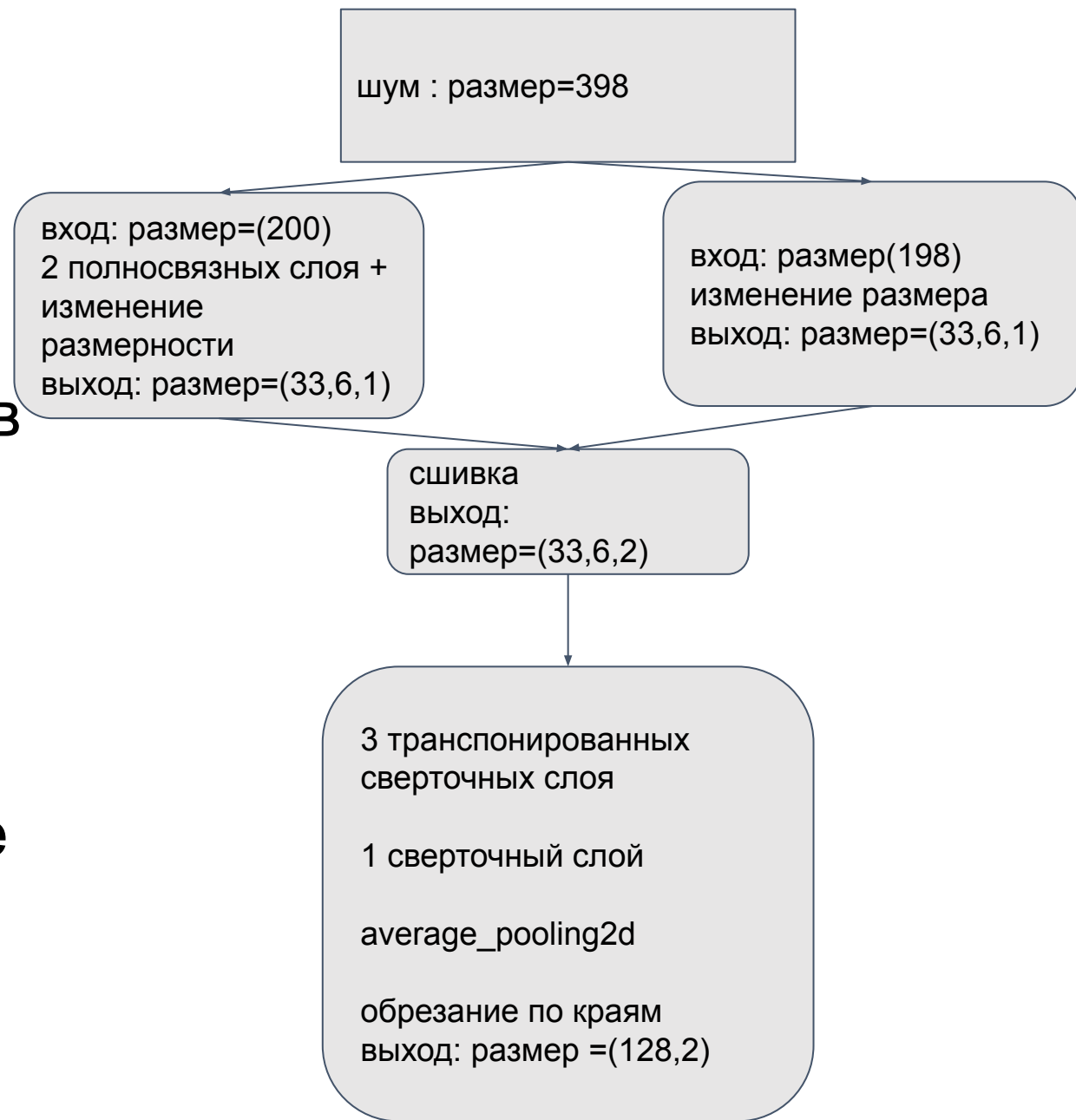
Мера оценки наилучшей модели

- В связи с улучшением качества генерации данных, визуально стало тяжело различать качество моделей. В связи с этим была взята в качестве меры ошибка описанная на прошлом слайде. Таким образом наиболее лучшая модель имела наименьшую меру т.е. выдавала наиболее реалистичные данные.

	model 1	model 2	model 3	model 4	model 5
epochs					
1	0.009665	0.009894	0.009614	0.009702	0.010054
5	0.009689	0.009829	0.009884	0.010173	0.010288
10	0.009714	0.010362	0.009777	0.010420	0.010251
15	0.009724	0.010088	0.010100	0.010400	0.009825
20	0.009710	0.010132	0.010504	0.010376	0.009756
30	0.009950	0.010674	0.010495	0.009801	0.009754

Модели

- Модель критика состоит из 5 сверточных двухмерных слоев с ядром (5,2), после которого идет слой MaxPooling(выделение максимального элемента из области) с ядром (2,1).
- Модель генератора на рисунке справа.
- Каждая модель имеет около 150 тысяч обучаемых параметров.



Итоги и дальнейшая работа

В настоящий момент мы можем генерировать визуально и качественно реалистичные данные одного детектора эксперимента Telescope Array. Так же написан алгоритм поиска аномалий, который помогает сравнивать качественно модели между собой и искать различия между реальными данными и сгенерированными методом МК.

В Настоящий момент идет работа над генерацией интегрированных показаний квадрата из 36 детекторов эксперимента.

СПАСИБО ЗА ВНИМАНИЕ