

Нейросетевые методы поиска скоплений галактик в микроволновом диапазоне по данным спутника Planck

Кафедра интеллектуальных информационных технологий
к.ф.-м.н. Мещеряков А. В., Герасимов С. В., Немешаева Алиса,
Москва, 2021

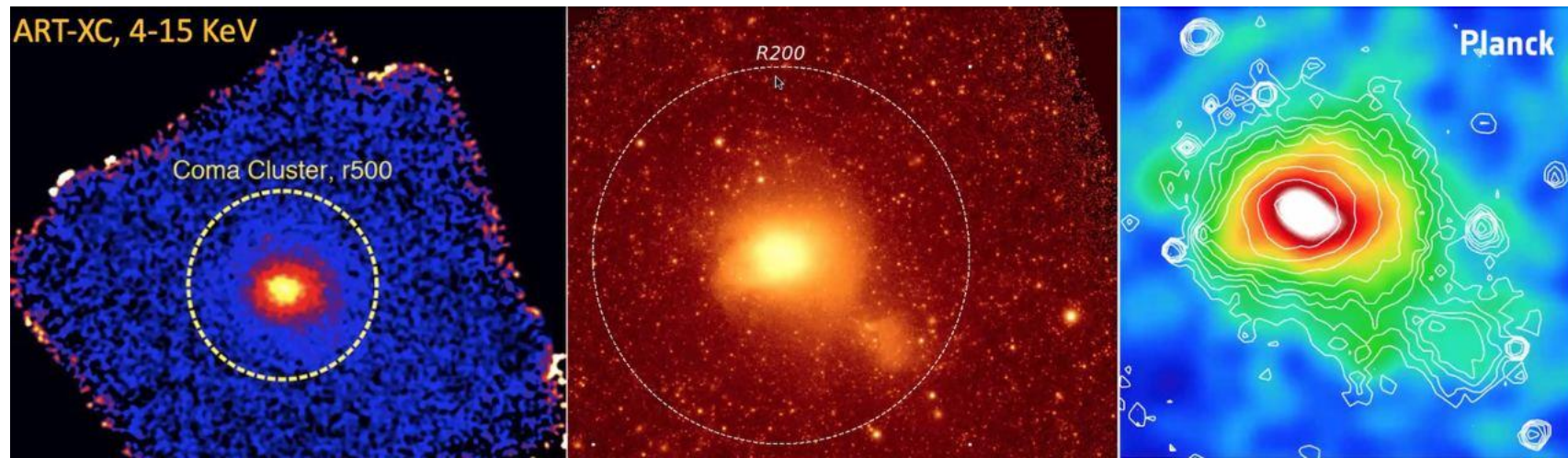
Рентгеновский телескоп eROSITA

- Рентгеновский телескоп eROSITA снимет всё небо и составит 8 карт неба в мягком рентгеновском излучении.
- Ожидается, что eROSITA обнаружит 100 000 скоплений галактик.
- Это будет самый полный каталог скоплений.
- Требуется сравнение с каталогами других диапазонов.



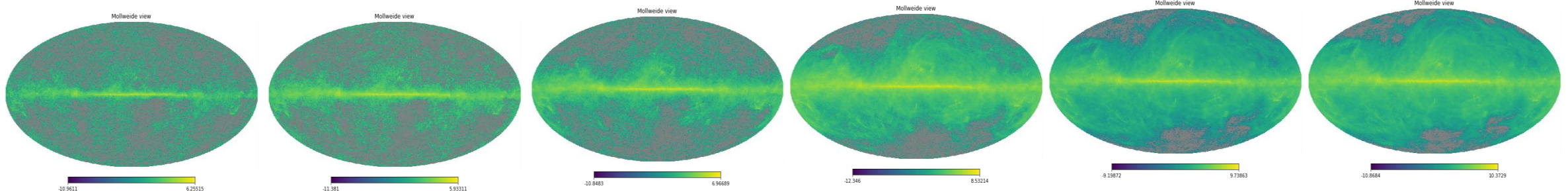
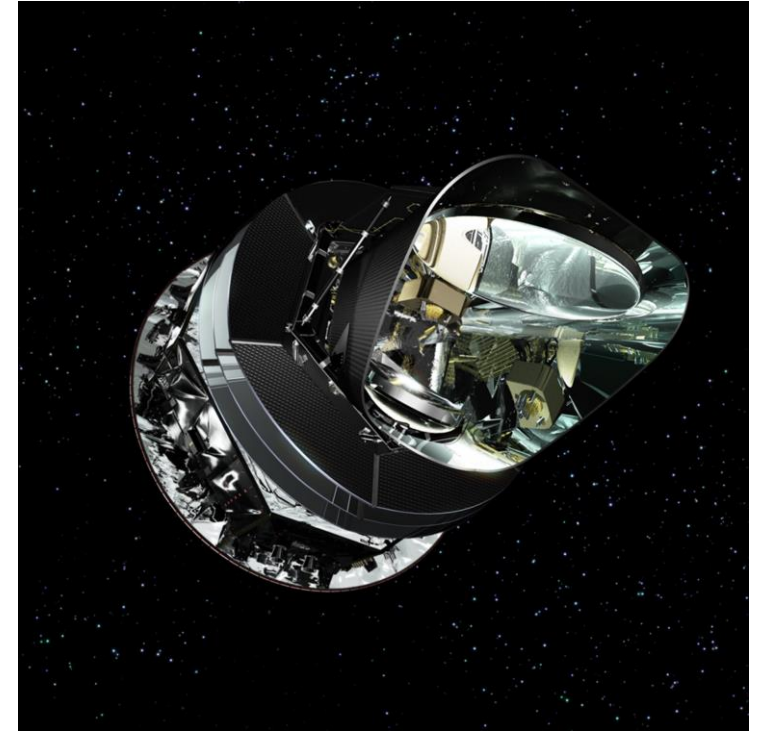
Скопления галактик

- Скопления галактик играют важную роль в задачах определения параметров Вселенной.
- Скопления видимы в **рентгеновском** диапазоне из-за наличия в их составе горячего газа, излучающего энергию в рентгеновском диапазоне.
- В **микроволновом** диапазоне скопления галактик проявляются из-за эффекта Сюняева-Зельдовича.
- Скопления состоят из галактик, которые излучают энергию в видимом диапазоне, поэтому для их поиска можно использовать и **оптический** диапазон.
- Кроме того, скопления важны для изучения эволюции галактик.



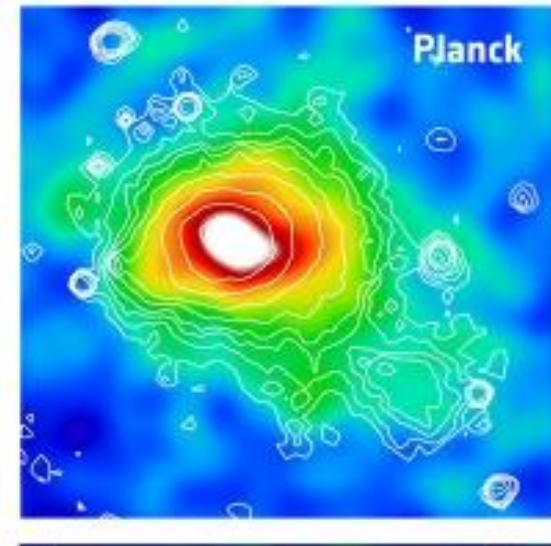
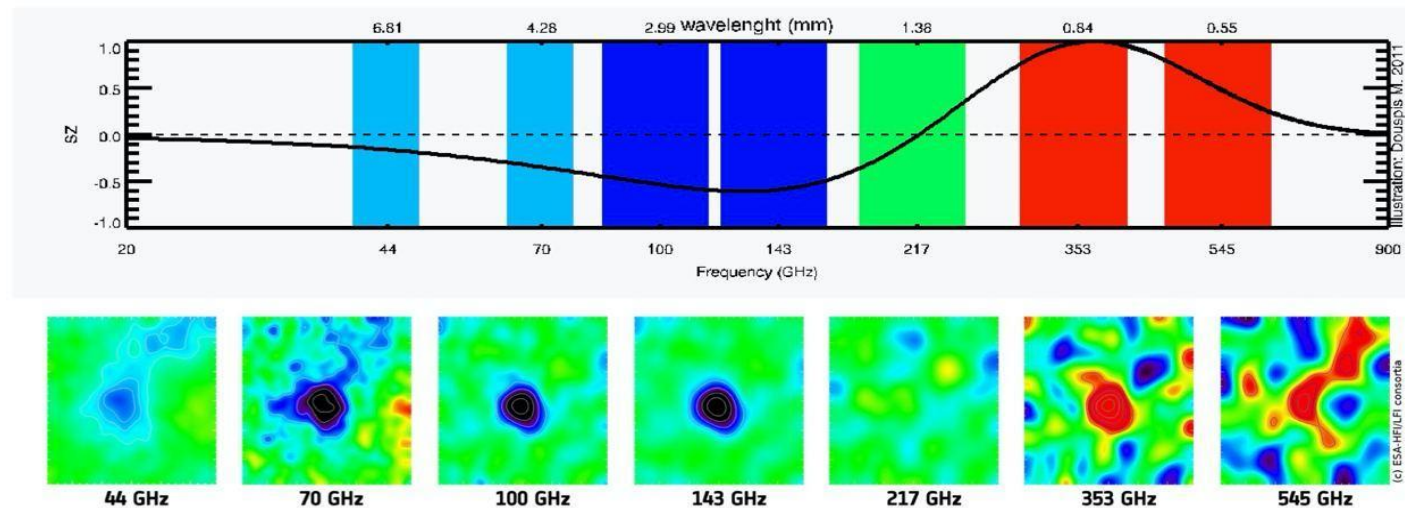
Микроволновой телескоп Planck

- **Planck** — астрономический спутник Европейского космического агентства, созданный для изучения вариаций космического микроволнового фона — реликтового излучения. Запущен 14 мая 2009 года. В период с сентября 2009 по ноябрь 2010 года «Планк» успешно закончил основную часть своей исследовательской миссии, перейдя к дополнительной, завершившейся 23 октября 2013 года.



Эффект Сюняева-Зельдовича

- Эффект Сюняева-Зельдовича – изменение интенсивности микроволнового излучения реликтового фона на горячих электронах межзвёздного и межгалактического газа.
- Микроволновые данные позволяют изучать морфологию распределения тёмной материи в скоплениях, а также их взаимодействие.



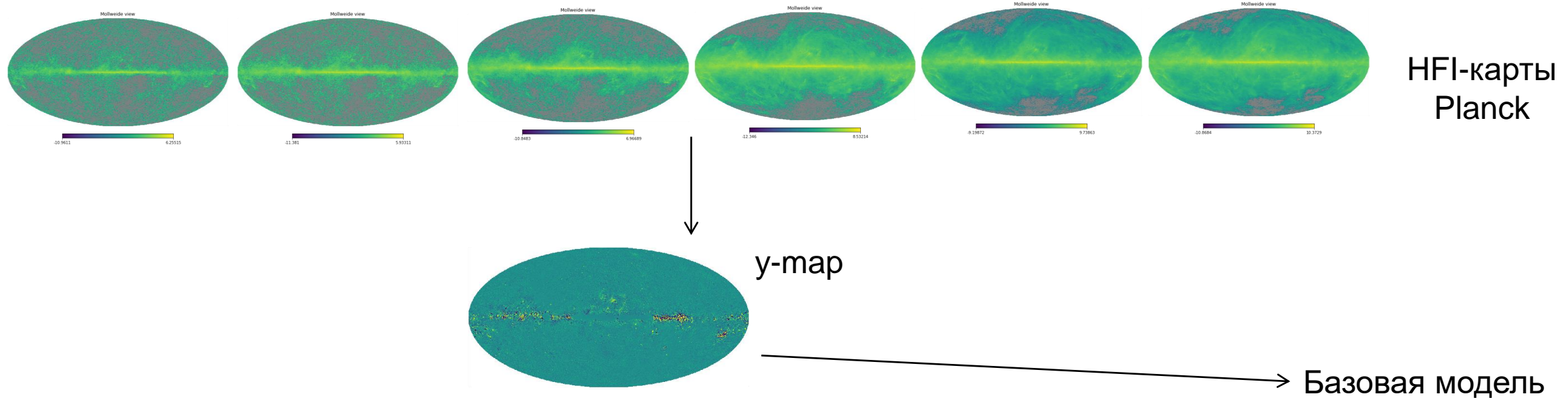
Coma Cluster, Planck data

Каталоги скоплений галактик

Каталог/ Свойство	PSZ2	ACT	SPT	MCXC	eRosita	RedMaPPer	Abell
Диапазон	Микроволн.	Микроволн.	Микроволн.	Рентген.	Рентген.	Оптич.	Оптич.
По всему небу	Да	Нет	Нет	Да	Да (но доступна только половина)	Да	Да
Количество объектов	1653	4195	343	1743	~100 000	25325	4073
Количество скоплений	1425						
Методы	MMF, PwS	MMF	MMF				

Обзор существующих решений

MMF	PwS	Базовая модель	CNN	Unet
Фильтр в пространстве Фурье, для которого нужно определить условия шума и условия сигнала	Быстрый байесовский метод, позволяющий детектировать объекты скрытые в шуме	Обрабатывает у-карты параметра Комптона и находит максимумы в разных слоях	Свёрточная сеть для классификации (в статье DeepSZ используется архитектура ResNet)	Свёрточная сеть для сегментации со skip-connection связями



Постановка задачи

Исследование и разработка нейросетевых методов сегментации и детекции источников Сюняева-Зельдовича в данных Planck, а также построение наиболее полного каталога скоплений галактик.

- Создание модели сегментации по данным Planck
- Исследование морфологии объектов на картах сегментации
- Создание каталога источников Сюняева-Зельдовича
- Анализ функции отбора каталога и его чистоты
- Сравнение с eROSITA

Формальная постановка задачи

Для данных карт Planck в виде матриц

$$P_{n \times m \times 6} \in R_{n \times m \times 6}$$

найти преобразование в карты сегментации в виде матриц вида

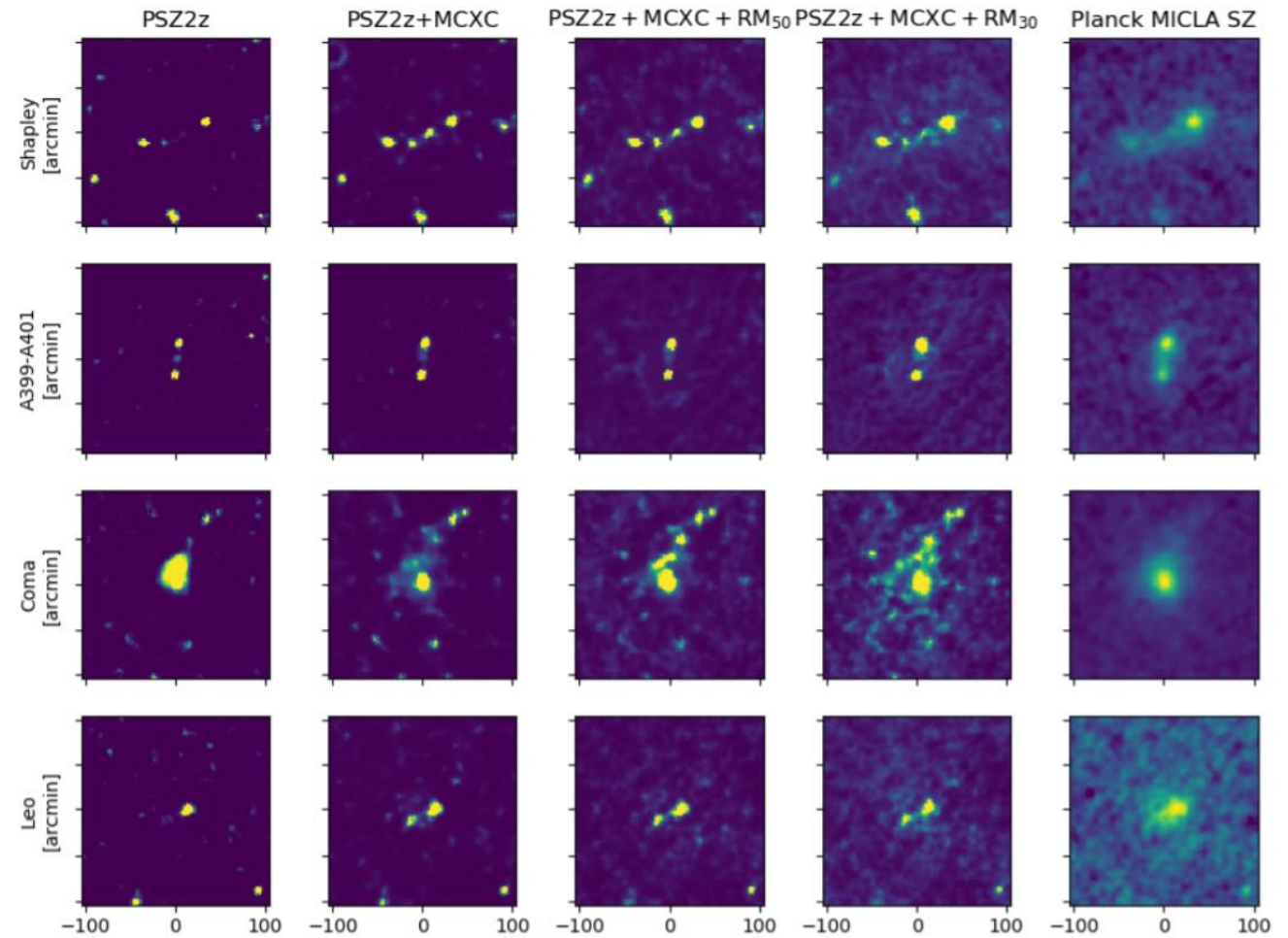
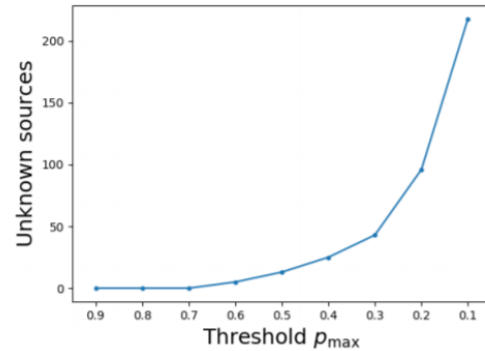
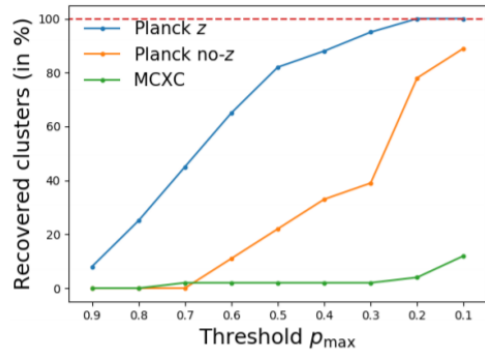
$$S_{n \times m \times 1} \in R_{n \times m \times 1},$$

и из этих карт в список детектированных скоплений, который будет наиболее полным по сравнению с выбранными каталогами скоплений галактик.

План доклада

- Обзор
 - Deep Learning for Sunyaev-Zel'dovich Detection in Planck», V. Bonjean, 2019
 - U-net
 - HEALPix
- Построение решения
- Результаты

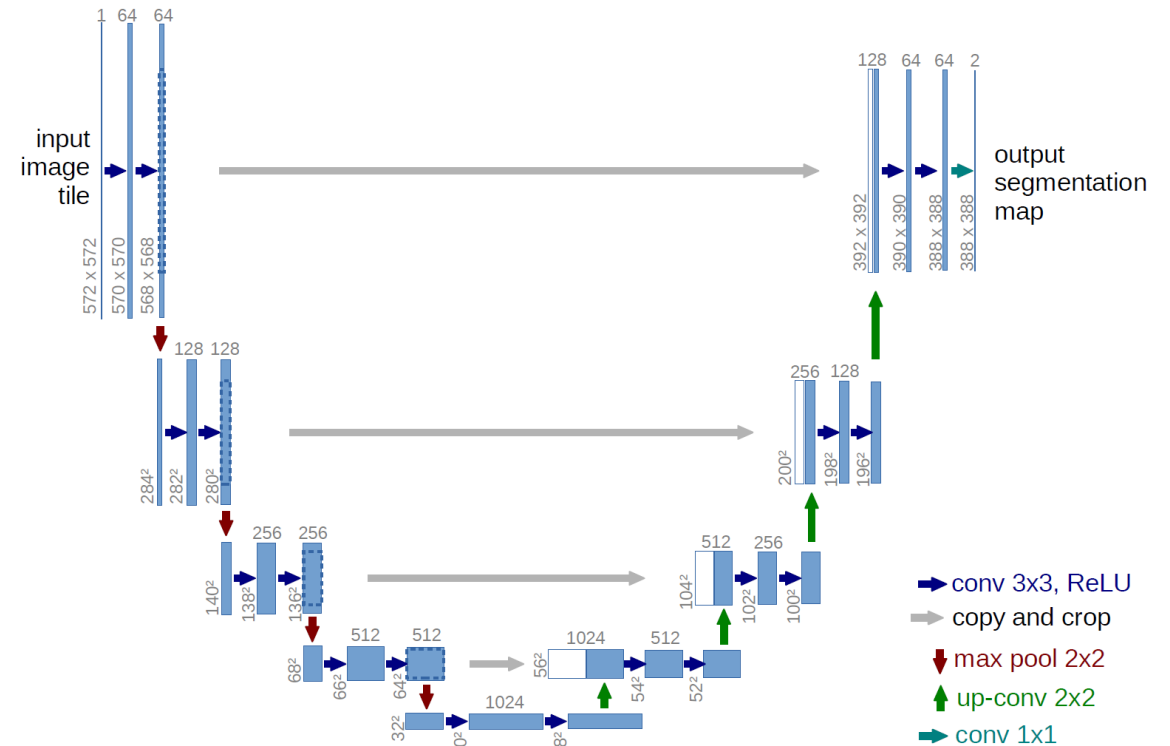
«Deep Learning for Sunyaev-Zel'dovich Detection in Planck», V. Bonjean, 2019



Автор не опубликовал каталоги и модель, нужно повторить эксперимент

U-net

- Одной из лучших нейросетевых архитектур для сегментации изображений является **U-net**. Её ключевой особенностью является наличие skip-connection слоёв, таким образом часть слоёв энкодера конкатенируется со слоями декодера, и при сжатии признаков нейросеть теряет меньше информации об изображении, на котором она обучалась.
- Olaf Ronneberger, Philipp Fischer, Thomas Brox, 2015



Метрики

Метрики сегментации:

- Бинарная кросс-энтропия, IoU, Dice

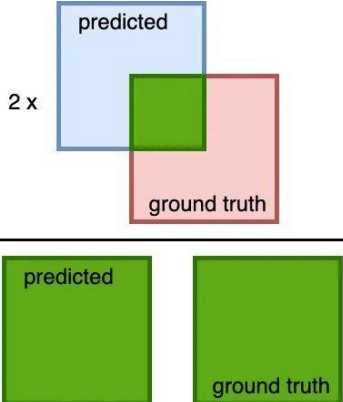
Метрики детекции:


- Recall, Precision

$$QE = -\frac{1}{N} \sum_{i=1}^N y_i * \log(p(y_i)) + (1 - y_i) * \log(1 - p(y_i))$$

$$Recall = \frac{N_{true}}{N_{cat}}$$

$$Precision = 1 - \frac{N_{false}}{N_{det}}$$

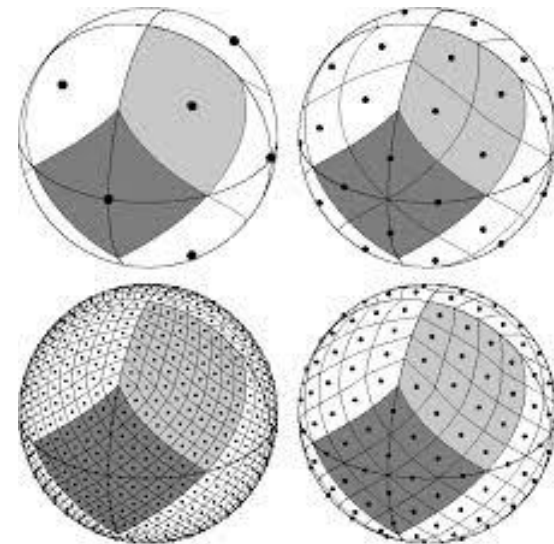
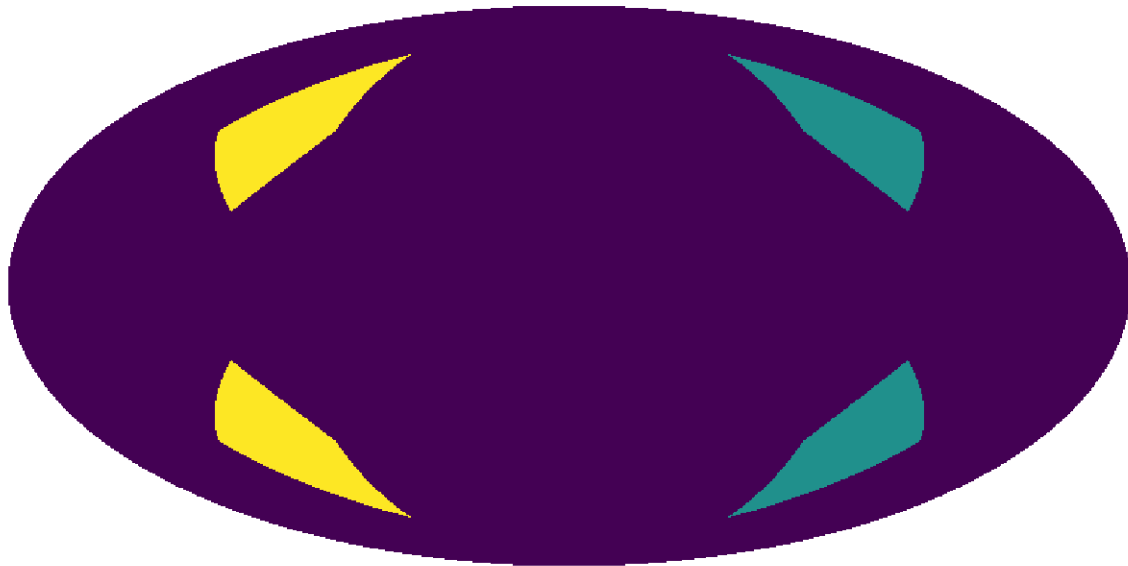
$$\text{Dice coefficient} = \frac{2 \times \text{area of overlapped (green)}}{\text{total area (green)}} = \frac{2 \times \text{area of overlapped (green)}}{\text{area of predicted} + \text{area of ground truth}}$$


$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$


HEALPix

- HEALPix – алгоритм иерархического разбиения сферы.
- Позволяет спроектировать на сферу данные и выбрать для них подходящее разрешение.
- Данные Planck хранятся в качестве изображения сферы, проиндексированного согласно HEALPix.
- HEALPix не искажает площадь объекта, но искажает форму.

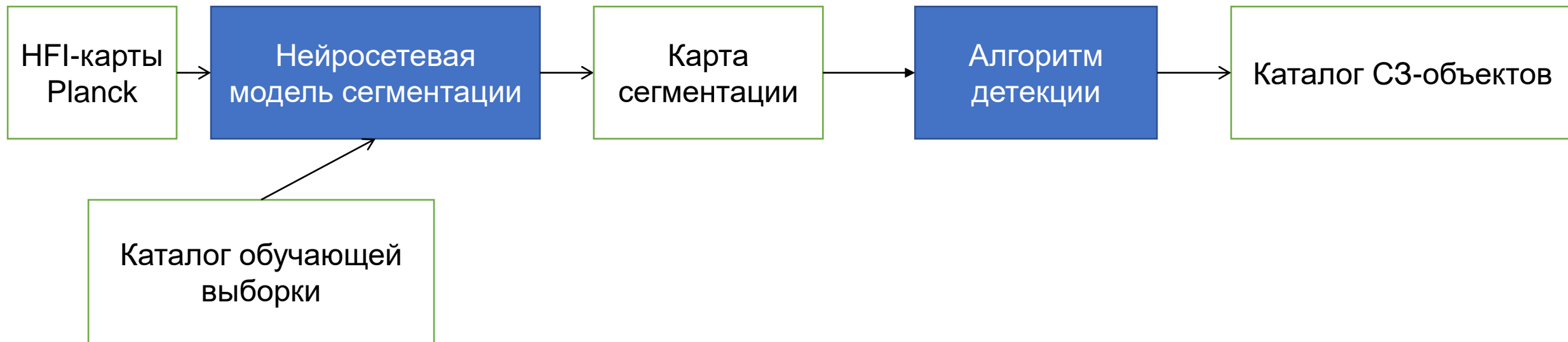
east and west validation pixels



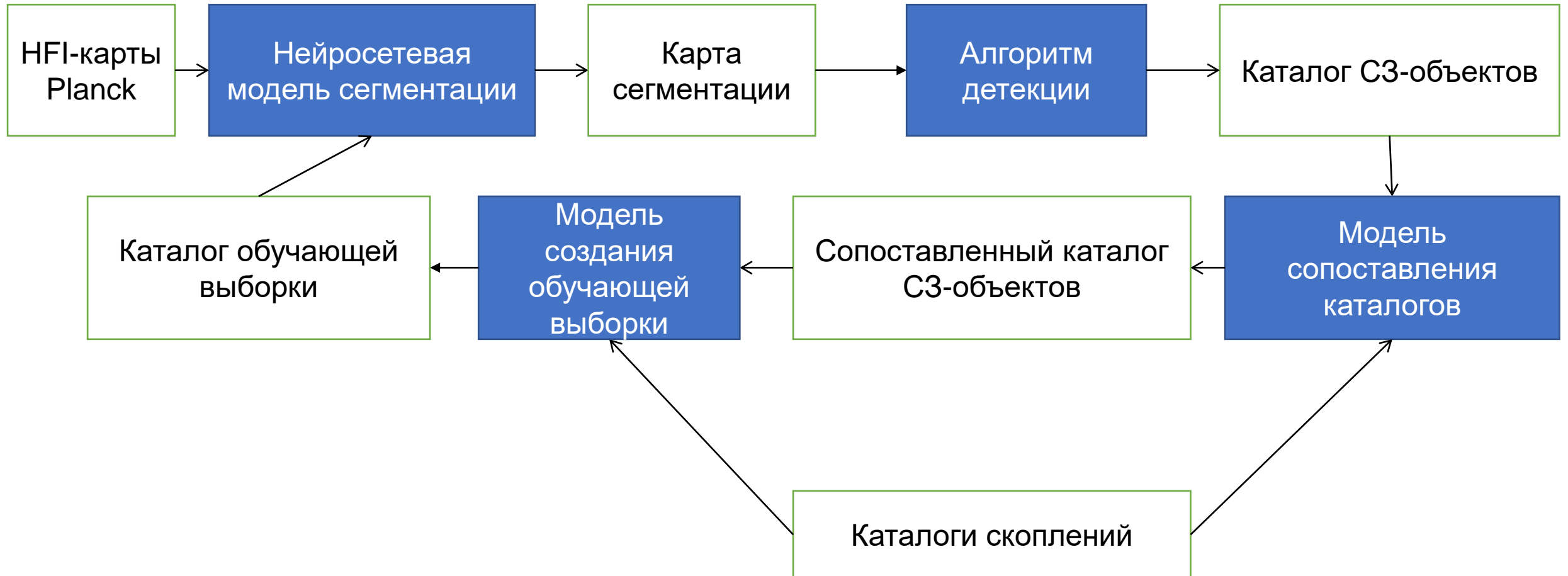
План доклада

- Обзор
- Построение решения
 - Обучение модели
 - Активное обучение
 - Детекция
- Результаты

Обучение модели

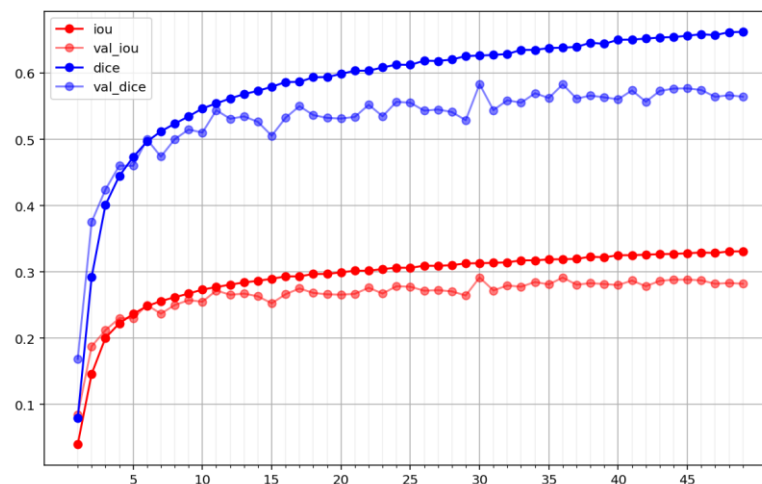
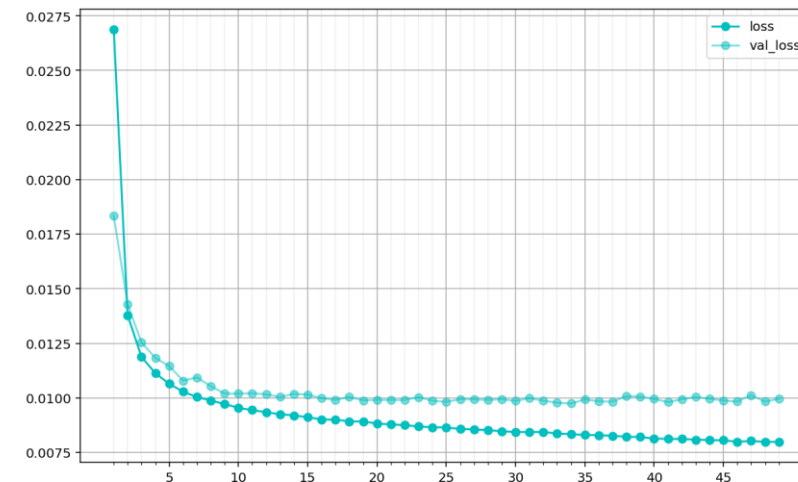


Активное обучение

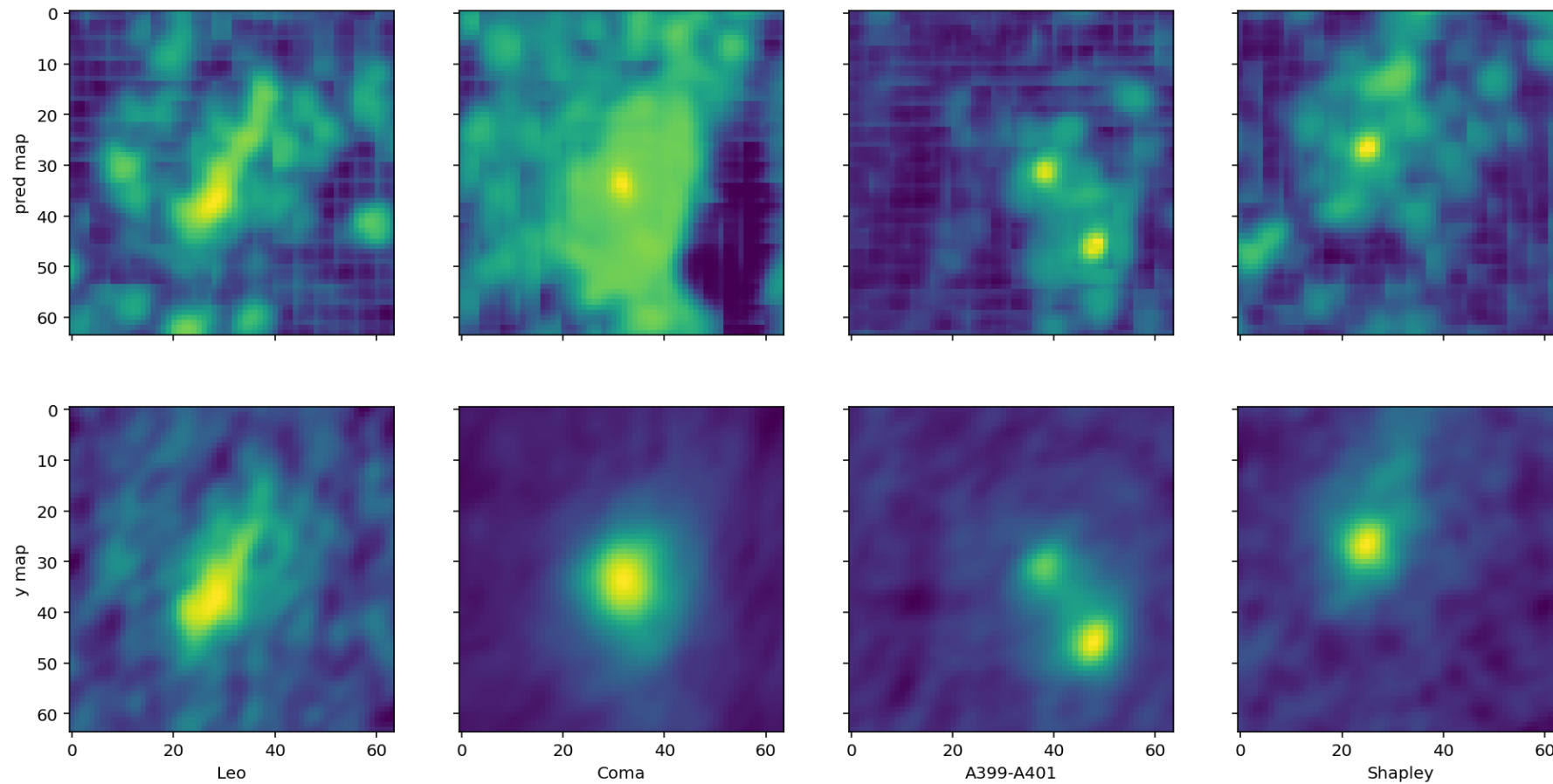


Процесс обучения нейросетевых моделей

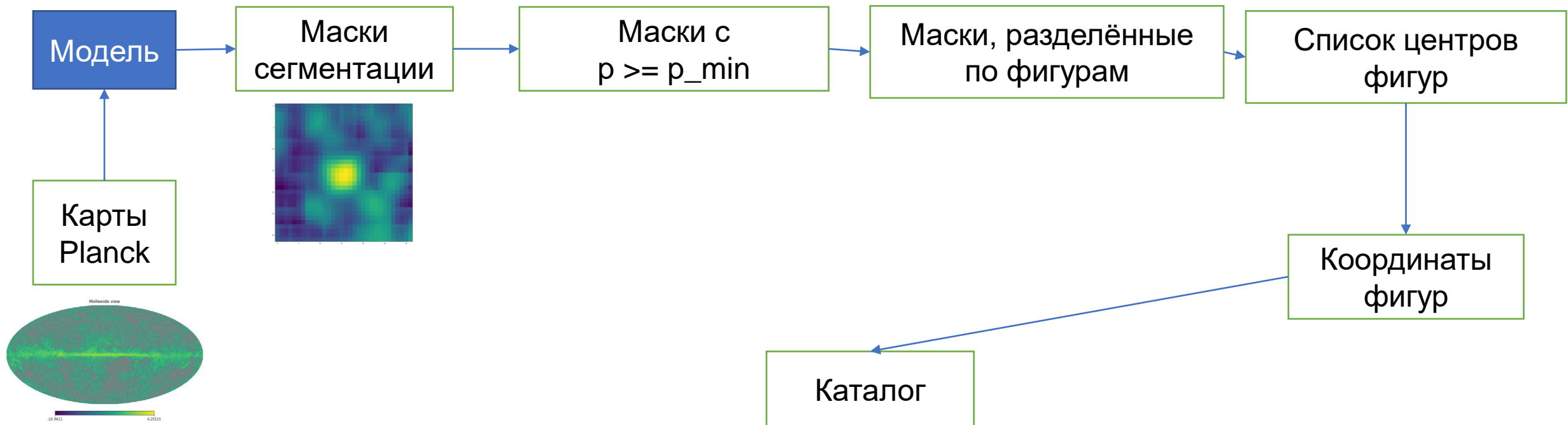
- Каждая модель обучалась по 50 эпох.
- Время, нужное для завершения одной эпохи: ~4.5 мин.
- Learning rate: 0.0001
- Google Colab Tesla K80



Морфология объектов на картах сегментации



Алгоритм детекции



RA	DEC	area	min_rad	max_rad	mean_rad	min_pred	max_pred	tRA	tDEC	status	catalog
259.732221	5.094132	4.0	0.701176	2.119449	1.320277	0.901549	0.933698	NaN	NaN	fp	NaN
283.542185	26.460609	4.0	0.700159	2.110640	1.317352	0.901048	0.957279	283.519410	26.482998	tp	planck_no_z
280.308927	30.912179	7.0	0.700308	2.503611	1.673052	0.900285	0.988940	280.278406	30.932247	tp	planck_z
276.363142	30.419430	7.0	0.292191	2.616653	1.576871	0.906883	0.985809	276.336046	30.431996	tp	planck_z
266.093042	32.971296	5.0	0.447466	2.275328	1.422385	0.903614	0.968407	266.067814	32.998643	tp	planck_z

План доклада

- Обзор
- Построение решения
- Результаты
 - Полученные модели
 - Выбор лучшей модели
 - Сравнение с базовой моделью
 - Сравнение с eROSITA

Полученные модели

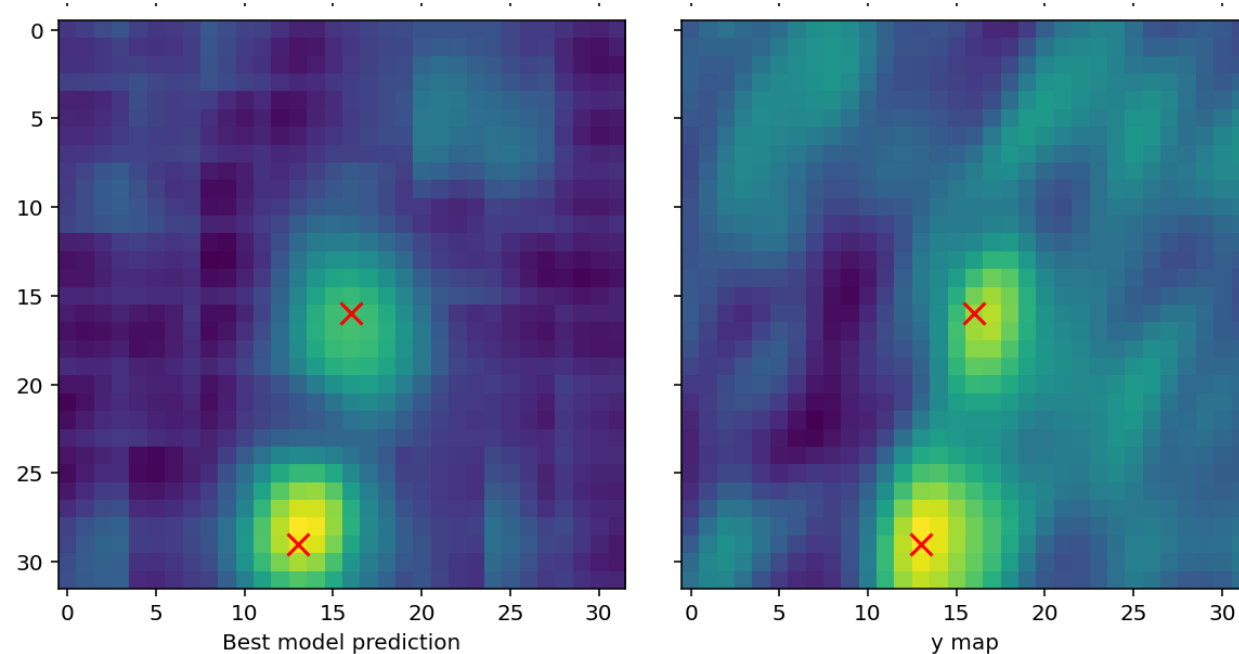
№	Патчи	Тренировочные данные	Dropout rate	Аугмент.
1	PSZ2	PSZ2	0.2	Нет
2	PSZ2		0.2	Да
3	PSZ2 + АСТ	PSZ2 + АСТ	0.2	Нет
4	PSZ2 + АСТ (отсеченный по z/M500)	PSZ2 + АСТ (отсеченный по z/M500)	0.1	Да
5	PSZ2		0.1	Да
6	PSZ2	PSZ2 + АСТ (отсеченный по активному обучению)	0.2	Да
7	PSZ2 + АСТ (отсеченный по активному обучению)		0.2	Да
8	PSZ2 + АСТ (отсеченный по активному обучению)	PSZ2 + АСТ + MCXC (отсеченный по активному обучению)	0.2	Да
9	PSZ2 + АСТ + MCXC (отсеченный по активному обучению)		0.2	Да
...

Выбор лучшей модели

Для анализа моделей считался recall (отношение найденных/общее количество) для различных ground-truth каталогов.

	PSZ2	MCXC	RM	ACT	Abell	fp	all
1	0.90	0.42	0.05	0.20	0.18	13775	16729
2	0.92	0.43	0.06	0.22	0.20	20036	23331
3	0.74	0.35	0.05	0.45	0.15	8966	12239
4	0.90	0.41	0.04	0.19	0.18	14437	17210
5	0.92	0.43	0.06	0.23	0.20	20387	23801
6	0.92	0.44	0.06	0.22	0.20	19654	22985
7	0.93	0.43	0.06	0.23	0.20	18577	21946
8	0.92	0.44	0.06	0.24	0.21	19849	23352

Сравнение с базовой моделью



Базовая модель для поиска скоплений в микроволновых данных: оценка пиков в картах у-параметра (величина эффекта Сюняева-Зельдовича).

	PSZ2	MCXC	RM	ACT	Abell	fp	all
Модель 8	0.92	0.44	0.06	0.24	0.21	21018	23352
Базовая модель	0.74	0.38	0.04	0.19	0.18	11891	13689

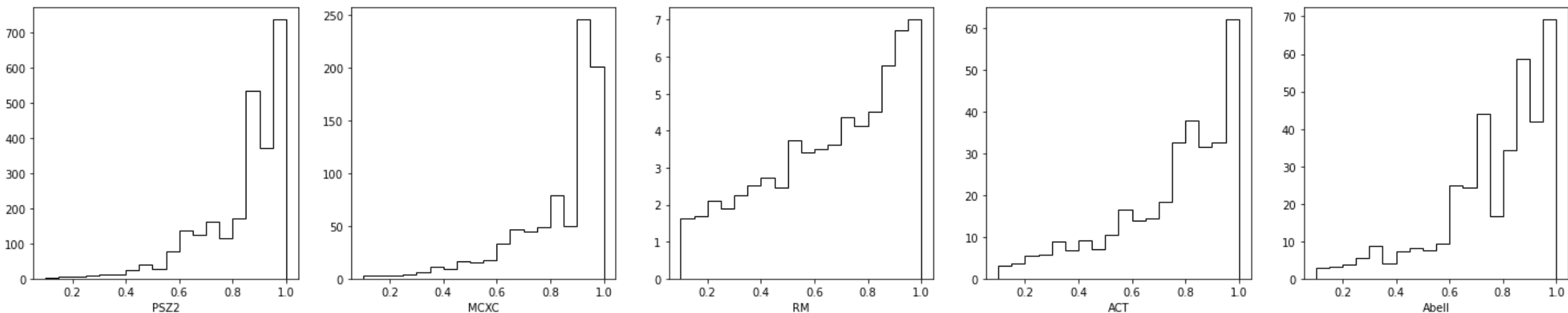
Параметры каталога

- `area` – площадь пятна
- `min_rad`, `max_rad`, `mean` – минимальное/максимальное/среднее расстояние от центра до внешних пикселей.
- `min_pred`, `max_pred` – минимальное/максимальное значение маски.

[illegible]

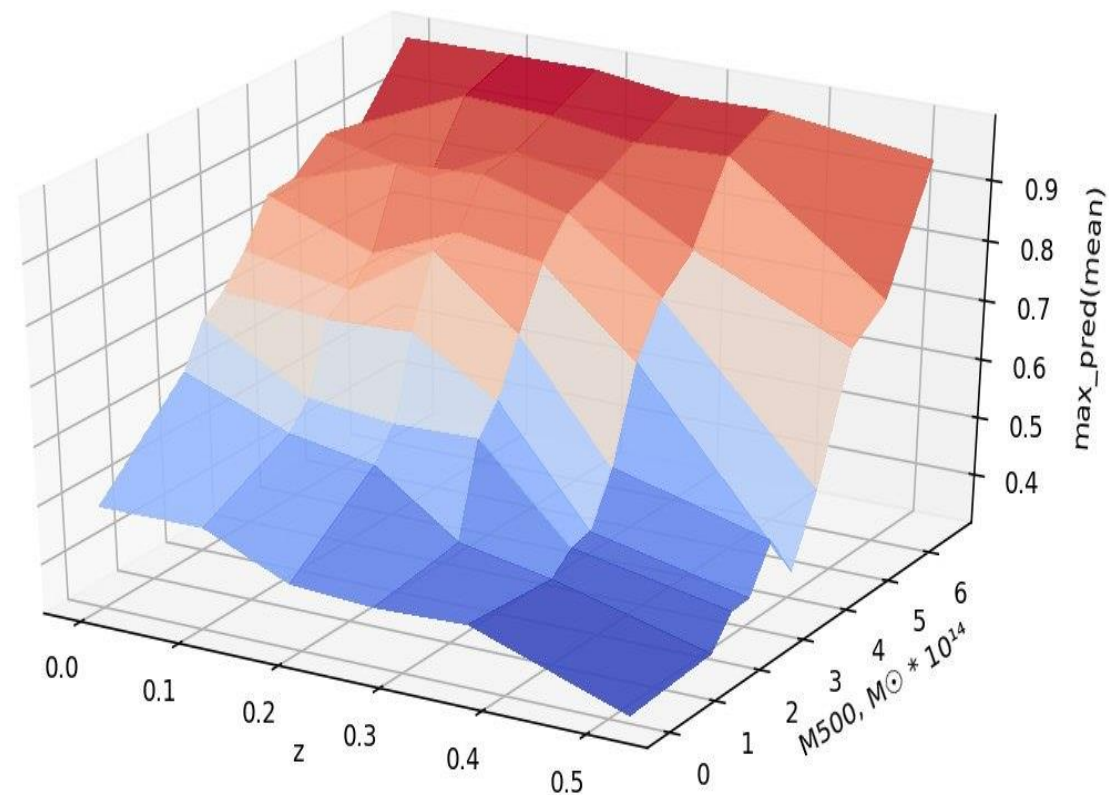
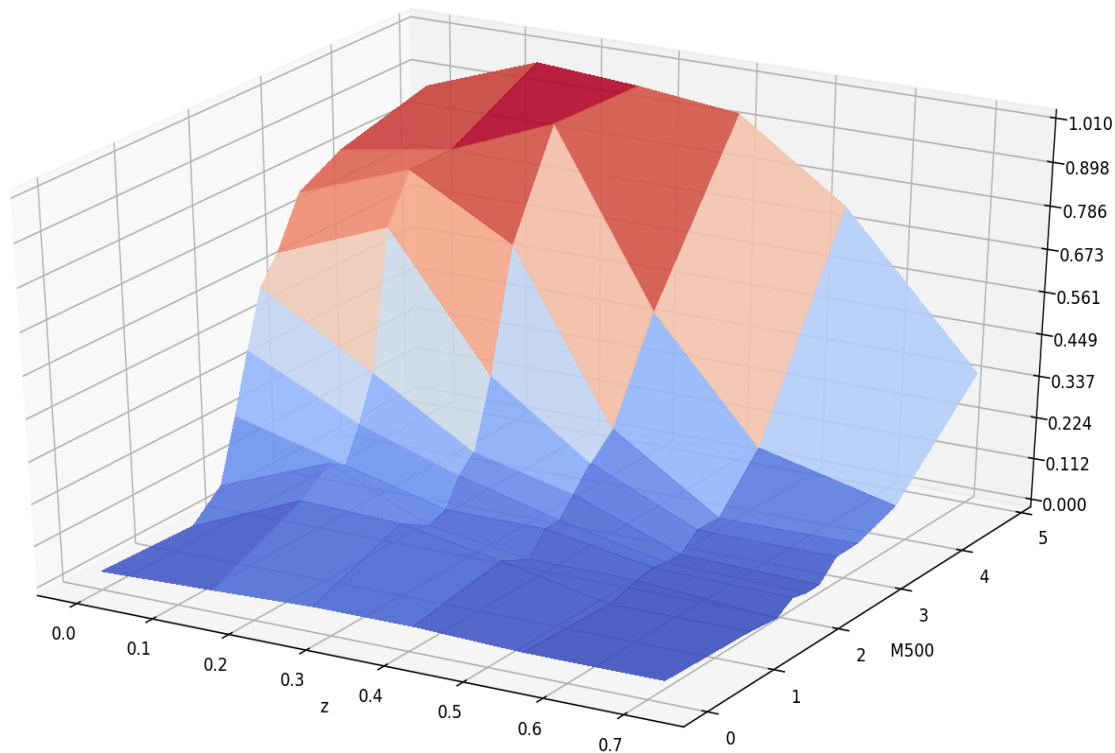
Чистота выборки

Чтобы определять чистоту полученного каталога, можно использовать один из параметров маски сегментации.



Отношение найденных объектов к ошибочным и их распределение по параметру `max_pred`.

Исследование функции отбора



Маска сегментации модели позволяет более явно выделить скопления по сравнению с картами у-параметра.

Сравнение с eROSITA

При сравнении различных каталогов с eROSITA были получены следующие результаты:

Количество пересечений с:

- Модель 8 – 2214
- Базовая модель – 1828
- PSZ2 – 603
- redMaPPer - 4938

Результаты

- Обучены модели для сегментации данных Planck.
- Создан каталог на основе масок сегментации, исследованы его свойства. Recall на лучшем каталоге превосходит на 18% recall базового каталога по PSZ2 и на 5% по ACT. Среди объектов eROSITA найдено на 386 объектов больше. Этот каталог важен для сравнения с другими каталогами кандидатов скоплений.
- Проведён анализ чистоты каталога.
- Проведено сравнение с eROSITA.
- Проведён анализ функции отбора каталога. Более 70% лучшего каталога имеют $M500 > 4 * 10^{14} \odot$. Исследовано распределение prediction index полученных моделей и чистоты выборки на различных каталогах из разных диапазонов.

Дальнейшие планы

- Статья о полученных каталогах.
- Перейти на детекцию скоплений в других диапазонах излучения (рентгеновский диапазон).