

# Нейросетевые методы поиска и сегментации объектов в данных современных космических обзоров

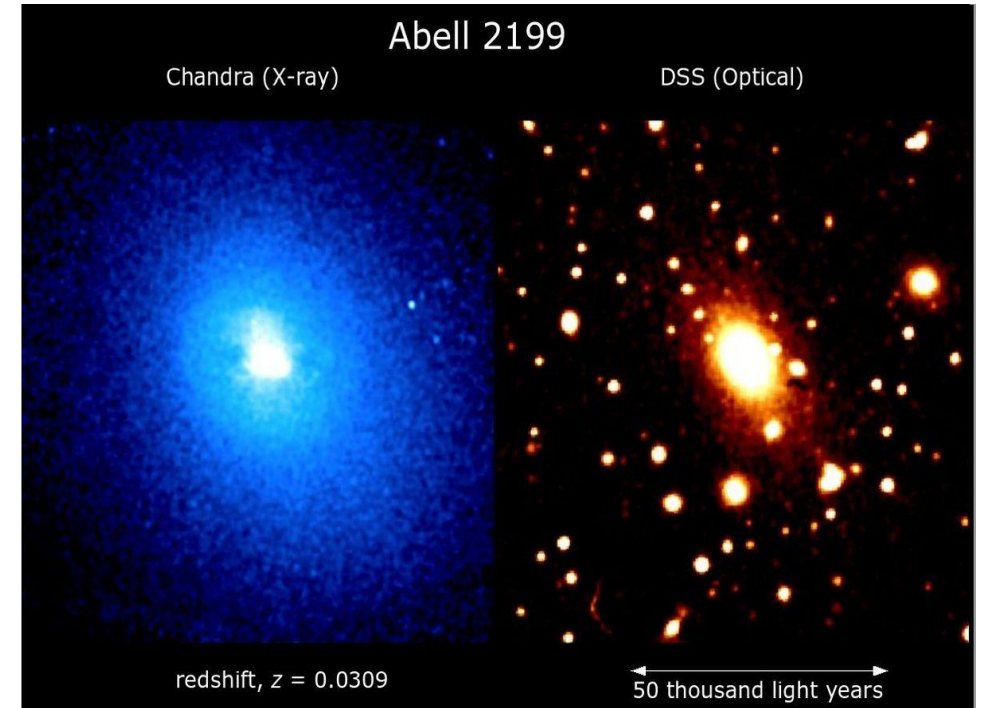
Научные руководители:

Герасимов С. В., к.ф.-м.н. Мещеряков А. В.

Студент: Немешаева Алиса,  
4 курс бакалавриата ВМК МГУ

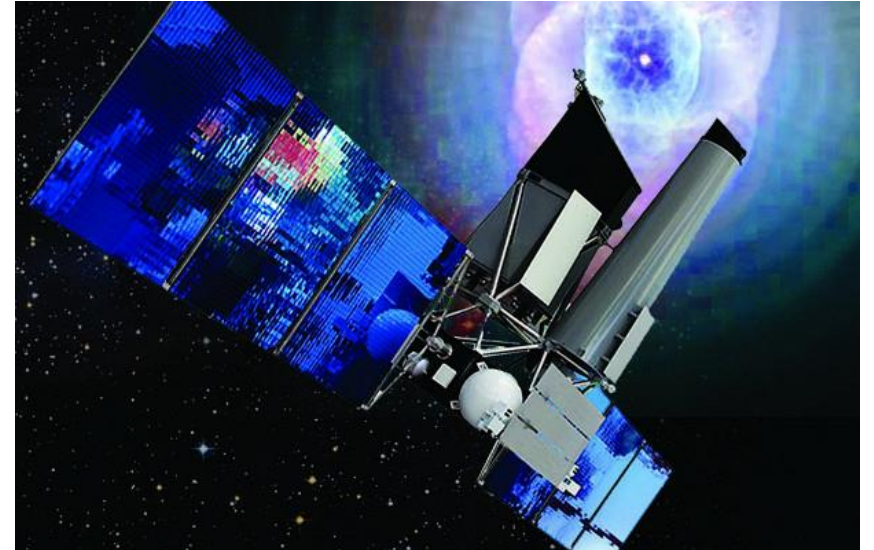
# Скопления галактик

- 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.



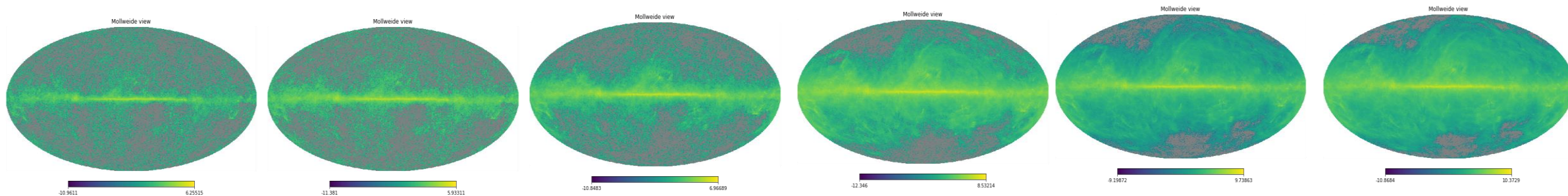
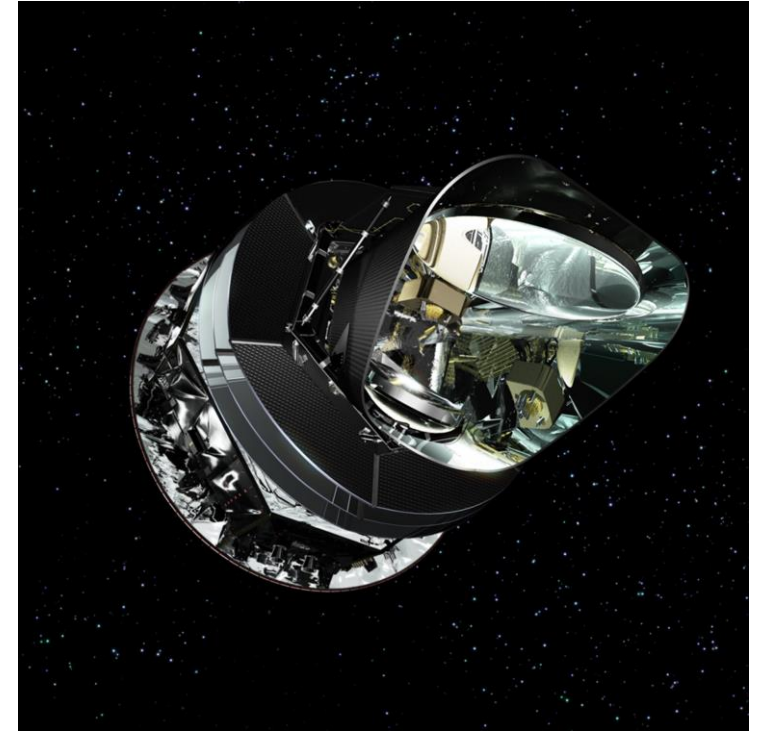
# Рентгеновский телескоп eROSITA

- Рентгеновский телескоп eRosita снимет всё небо и составит 8 карт неба в мягком рентгеновском излучении.
- Ожидается, что eRosita обнаружит 100 000 скоплений галактик.
- Каталоги, составленные по данным этого телескопа понадобятся, чтобы оценить полноту полученных по данным Planck каталогов.



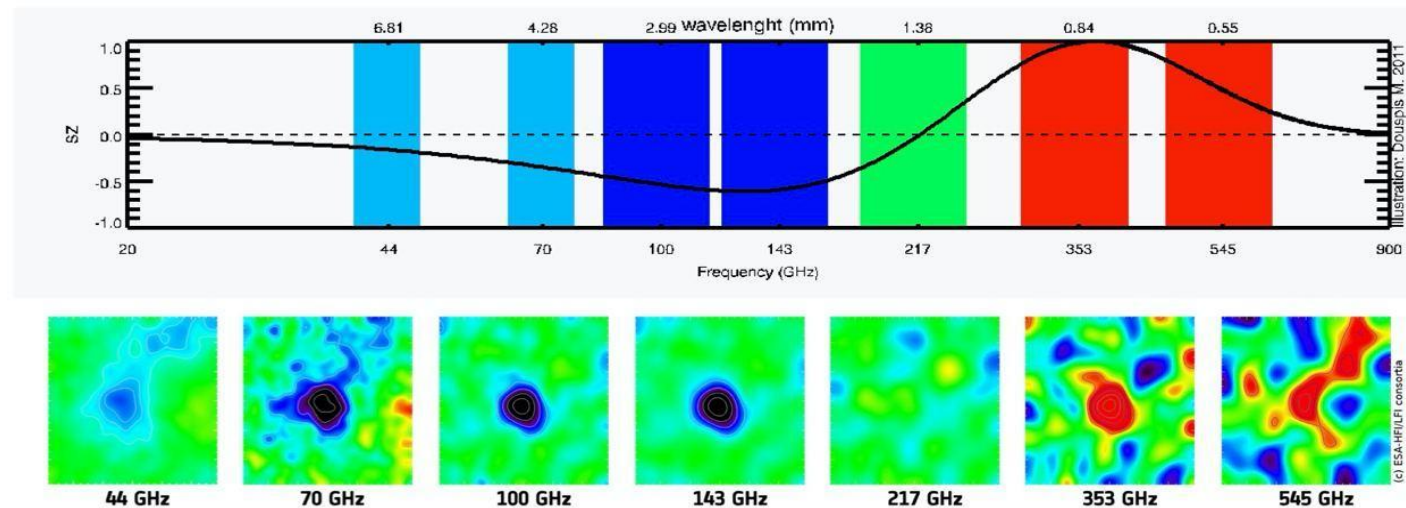
# Микроволновой телескоп Planck

- **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.





# 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	Нет	Да	Да
Количество скоплений	1653	4195	343	1743	~100 000	25325	4073
Методы	MMF, PwS	MMF	MMF				

# Сравнение существующих методов

- Классические методы: MMF, PwS.
- Нейросетевые методы: CNN (ResNet, Unet).

Важные отличия:

- предварительная обработка данных
- применимость к различным диапазонам

# Постановка задачи

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

- Создание модели сегментации по данным Planck
- Создание каталога скоплений
- Анализ функции отбора каталога ( $z$ , M500)



# 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

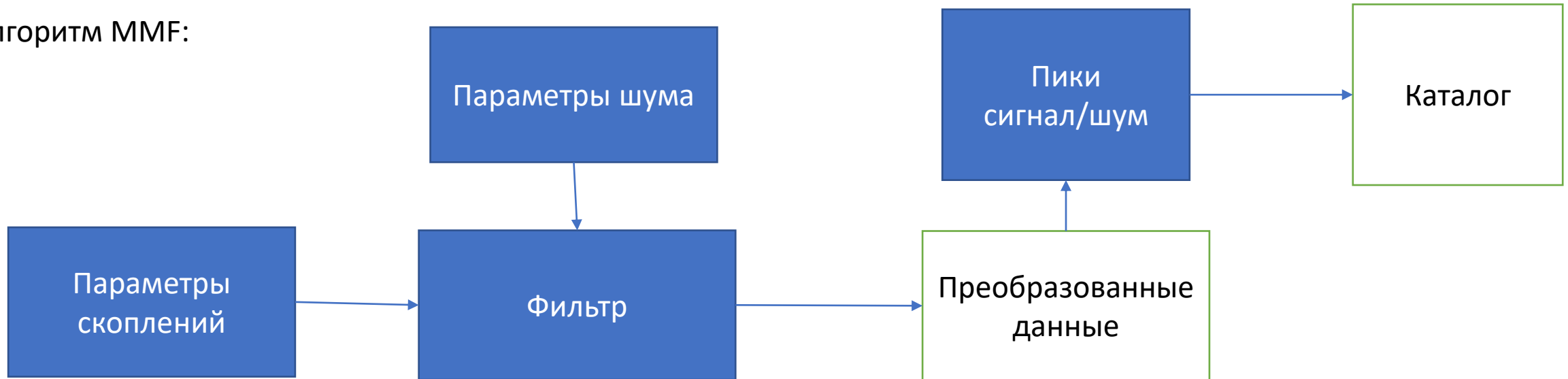
$$S_{n \times m \times 1} \in R_{n \times m \times 1},$$

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

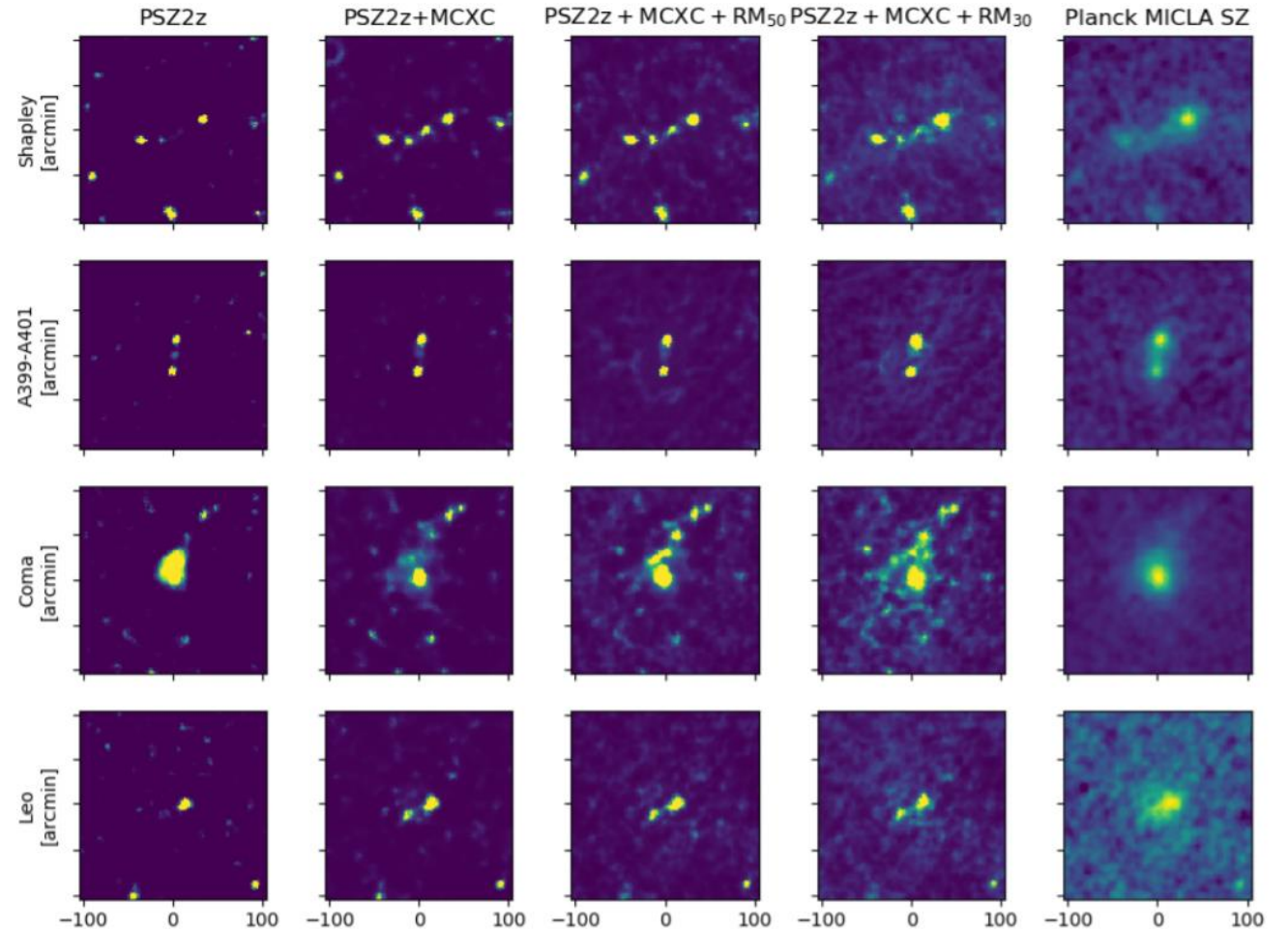
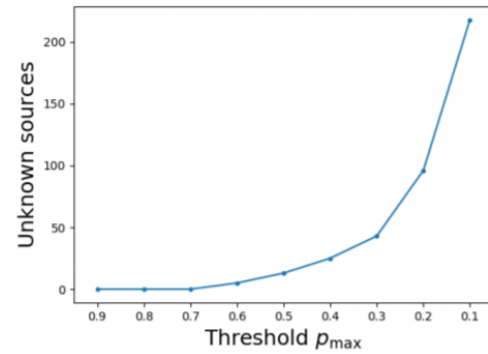
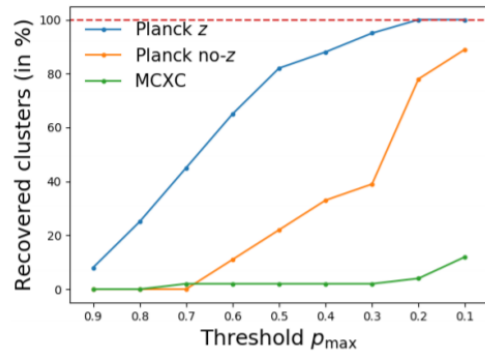
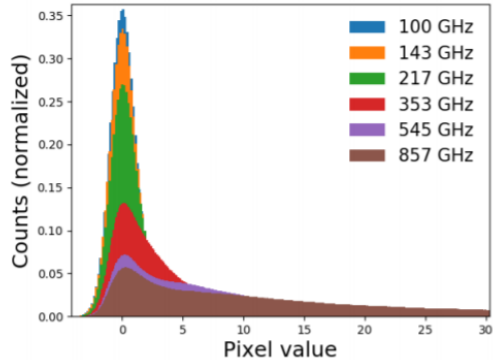
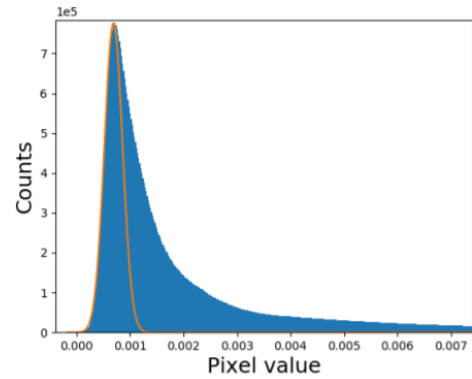
# Обзор существующих решений

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

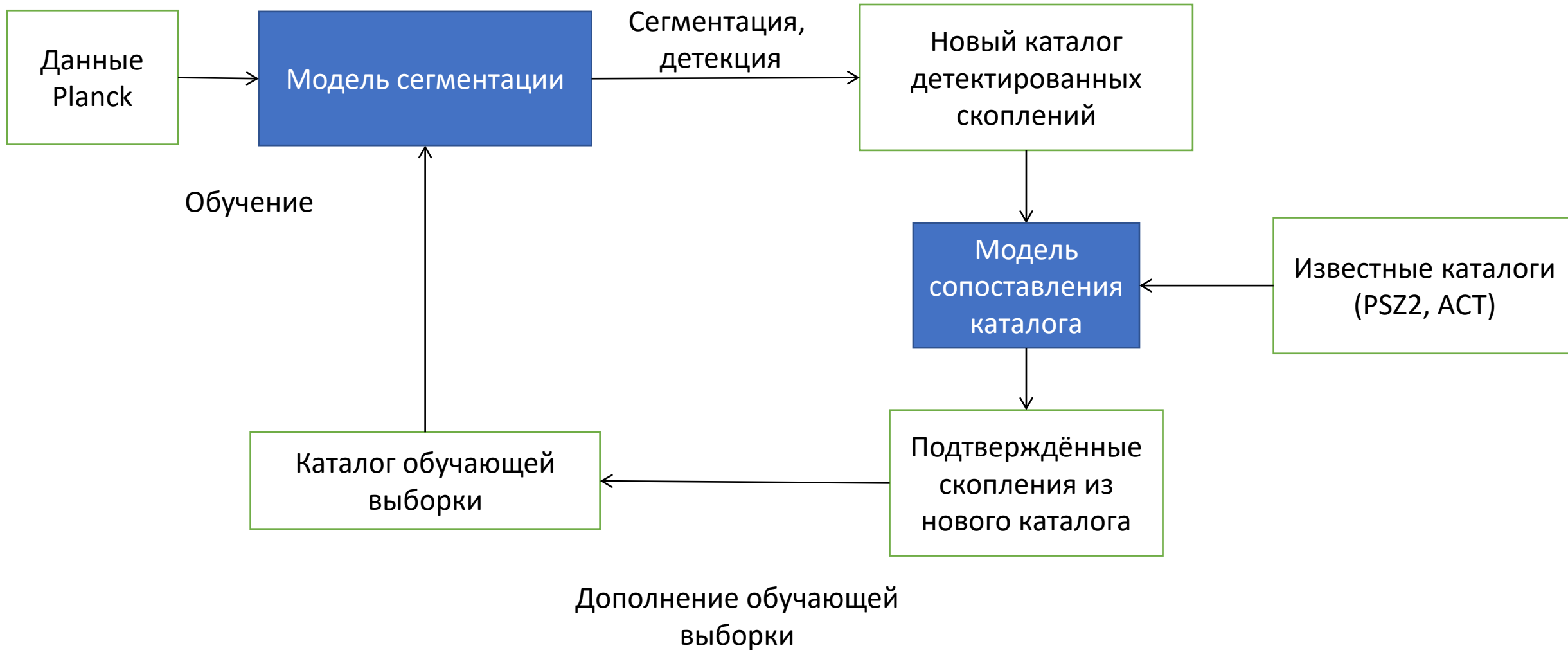
Алгоритм MMF:



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



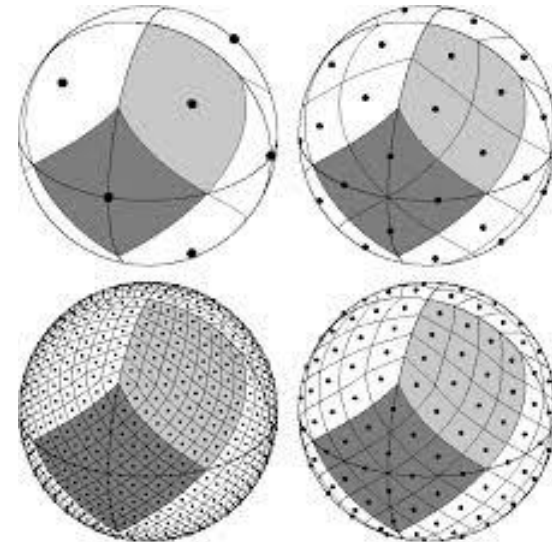
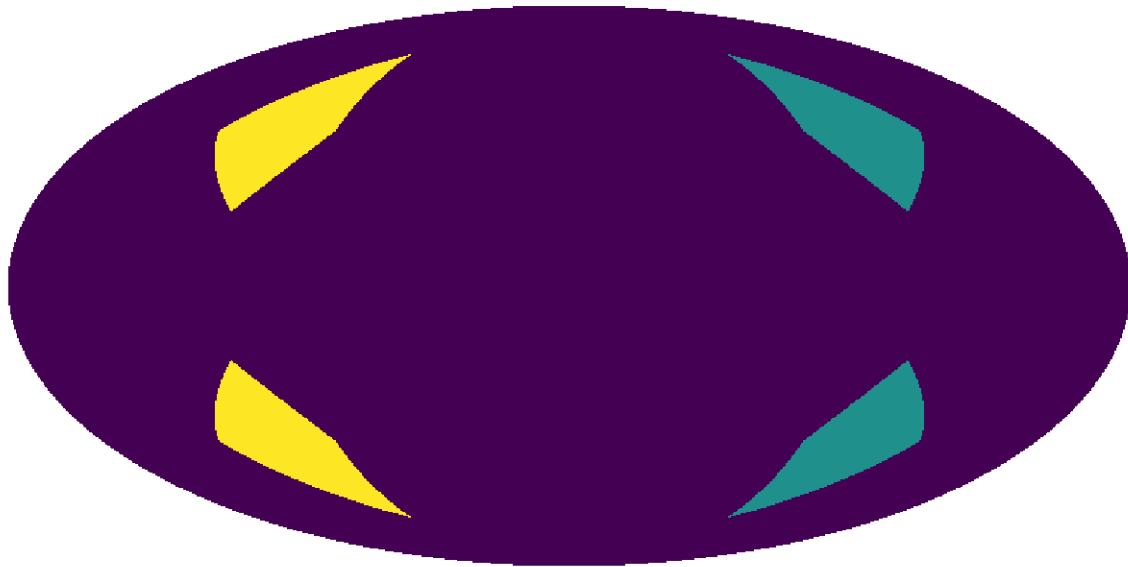
# Активное обучение



# 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 может исказить форму.

east and west validation pixels




# Метрики

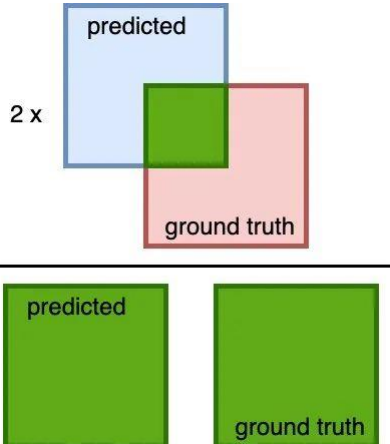
Метрики сегментации:

- IoU, Dice

Метрики детекции:

- Recall, Precision

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


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


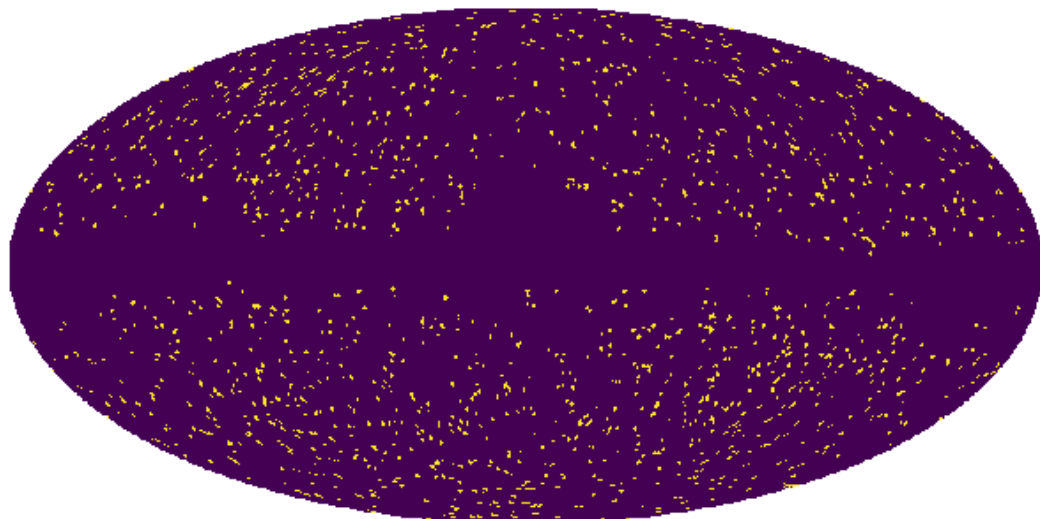
# Построение решения

1. Предварительная обработка.
2. Генерация данных для обучения.
3. Выбор модели и алгоритма детекции.
4. Обучение модели.
5. Применение модели (создание карт сегментации), детекция (создание каталога)
6. Анализ функции отбора каталога.

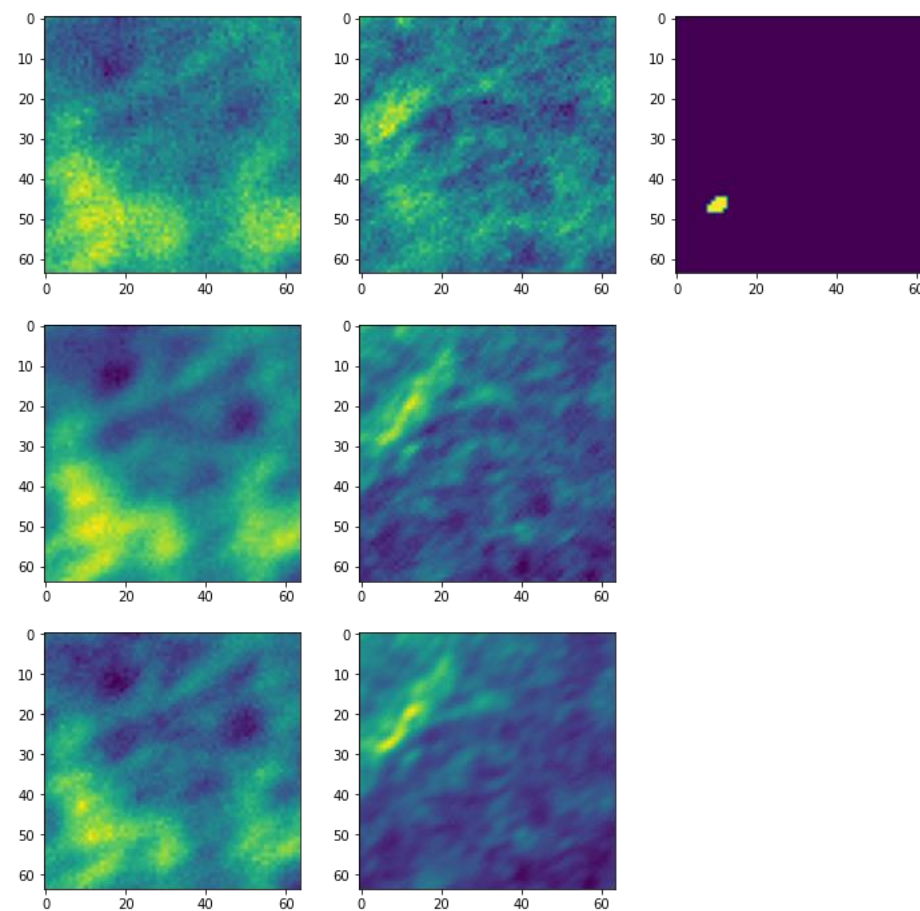
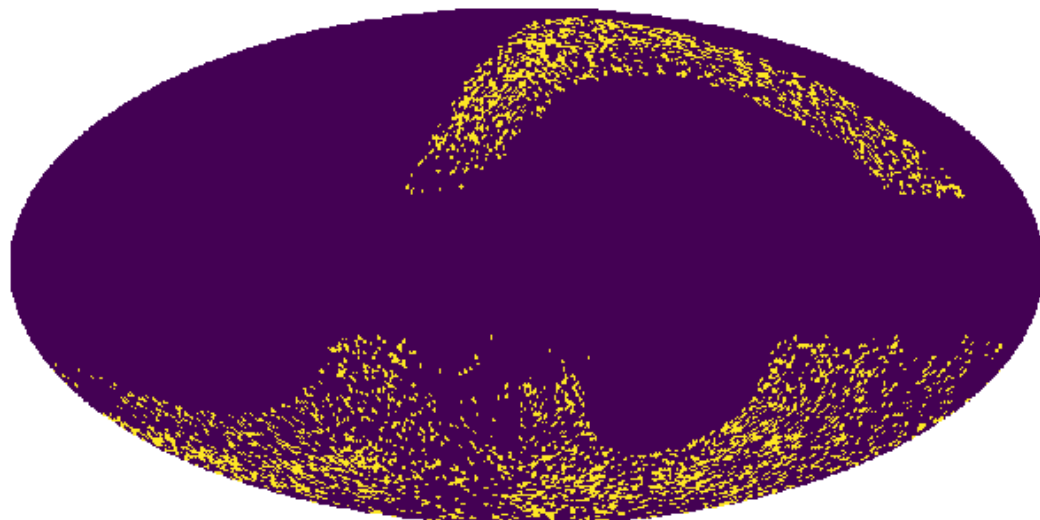


# Генерация патчей

PSZ2 clusters

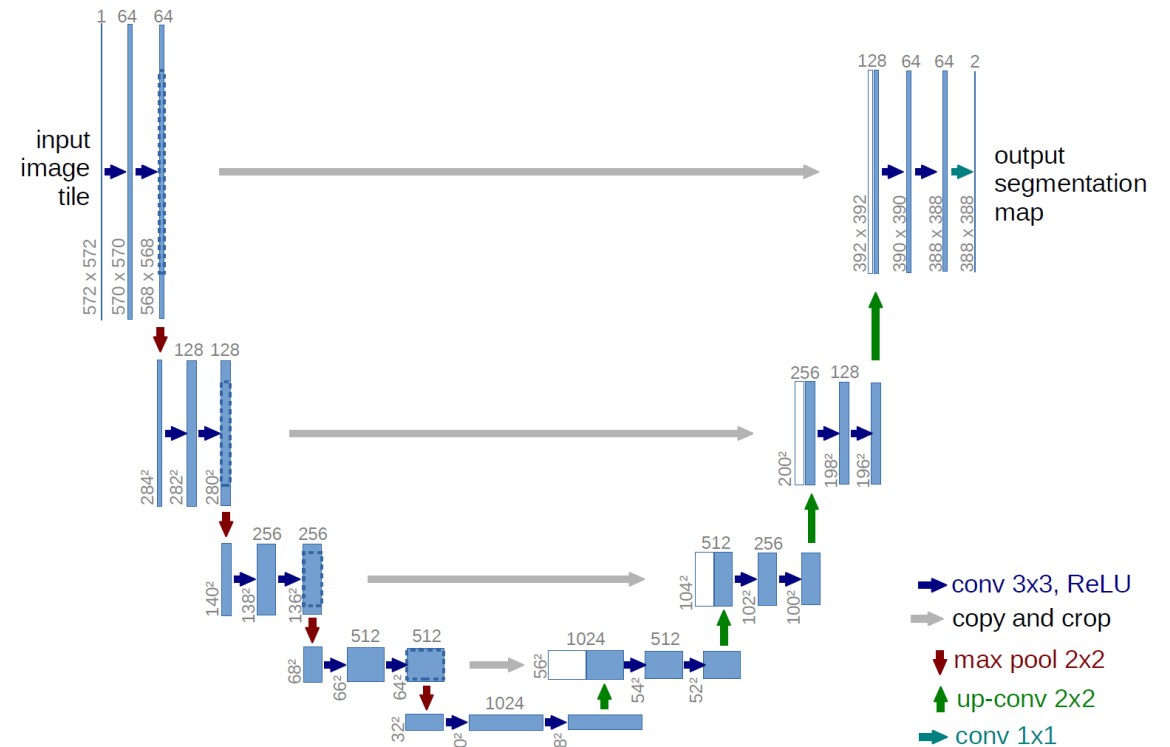


ACT clusters



# Выбор модели

- 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.

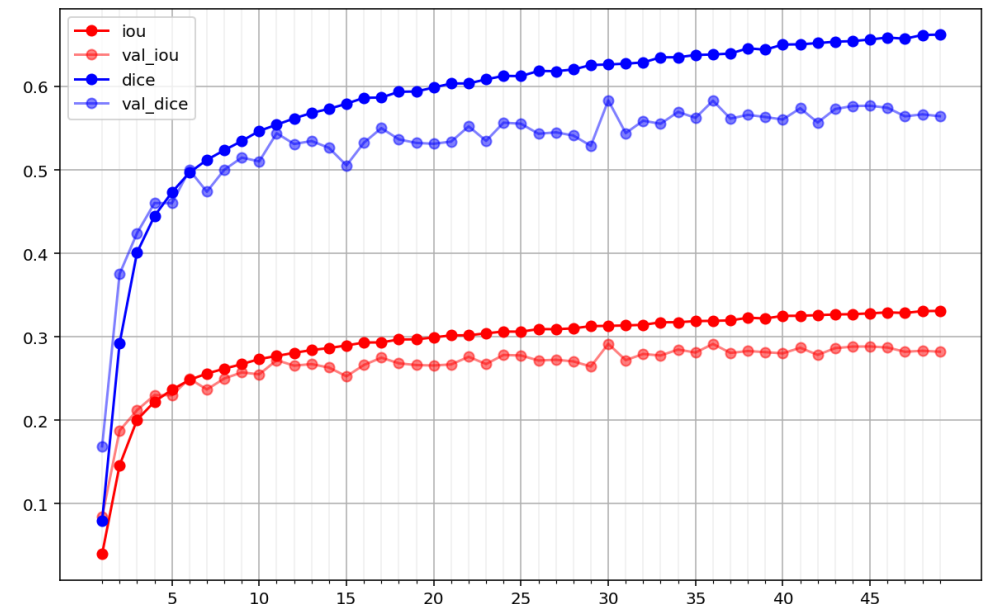
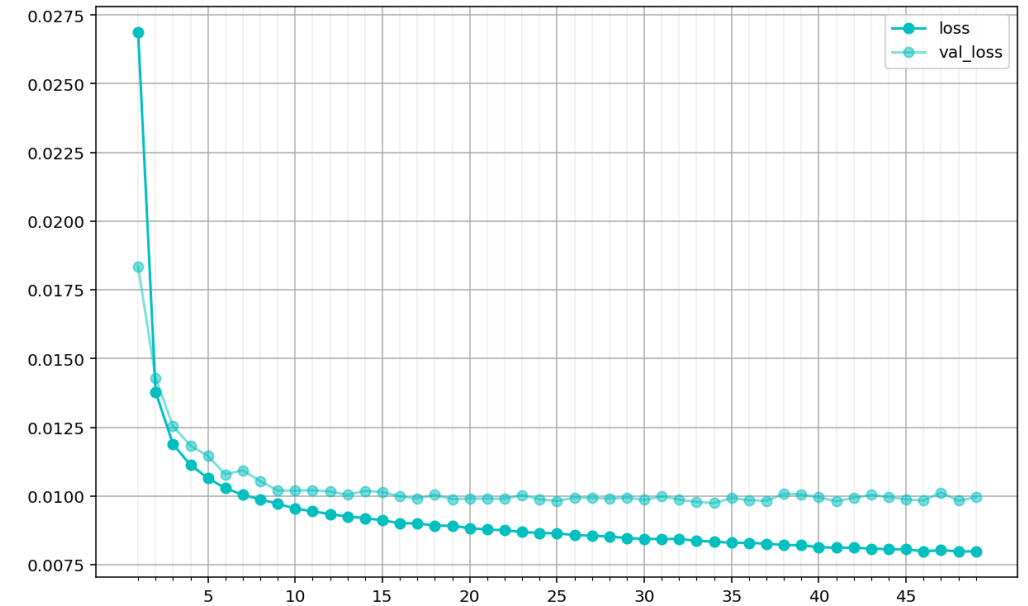


# Полученные модели

Название модели	Координаты патчей	Аугмент.	Маски	Dropout rate
pz	PSZ2	Нет	PSZ2	0.2
pz_rot	PSZ2	Да		0.2
pz_act	PSZ2 + ACT	Нет	PSZ2 + ACT	0.2
pz_act_jan_rot	PSZ2 + ACT (отсеченный по z/M500)	Да	PSZ2 + ACT (отсеченный по z/M500)	0.1
pz_act_feb_rot	PSZ2	Да		0.1
pz_act_found		Да	PSZ2 + ACT (отсеченный по активному обучению)	0.2
pz_act_found2		Да		0.2
pz_all_found	PSZ2 + ACT (отсеченный по активному обучению)	Да	PSZ2 + ACT + MCXC (отсеченный по активному обучению)	0.2
pz_all_found2	PSZ2 + ACT + MCXC (отсеченный по активному обучению)	Да		0.2

# Обучение моделей

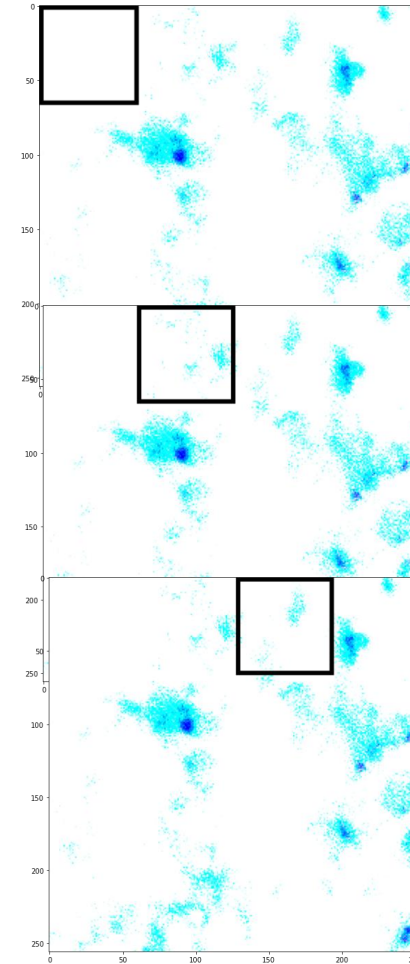
- Обучение длилось 50 эпох.
- Каждую эпоху нейросеть обучается на 100 000 патчах. Размер батча 20.
- Также рассчитываются метрики на патчах из валидации, 20000 патчей.



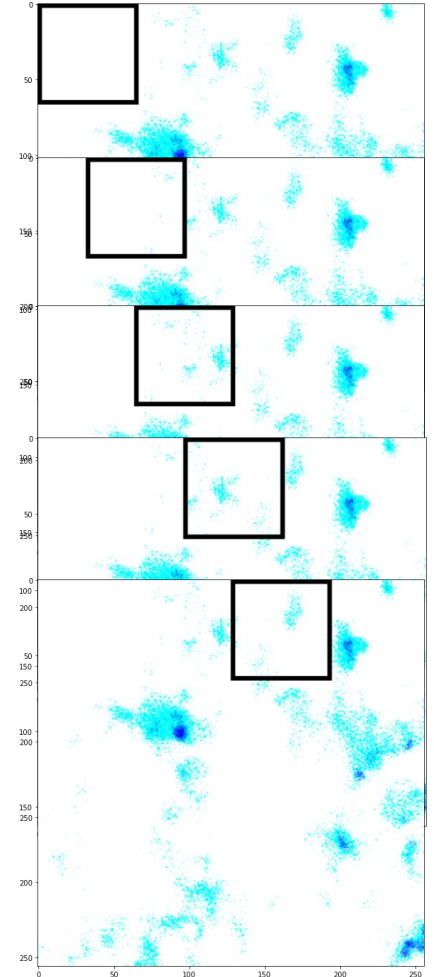
# Сканирование неба моделью

- Для сканирования нужно разбить небо на патчи 64x64.
- Повторное сканирование разных областей.
- Объединение всех масок в общую.
- Новый параметр детекции – «шаг».

Step = 64 = size

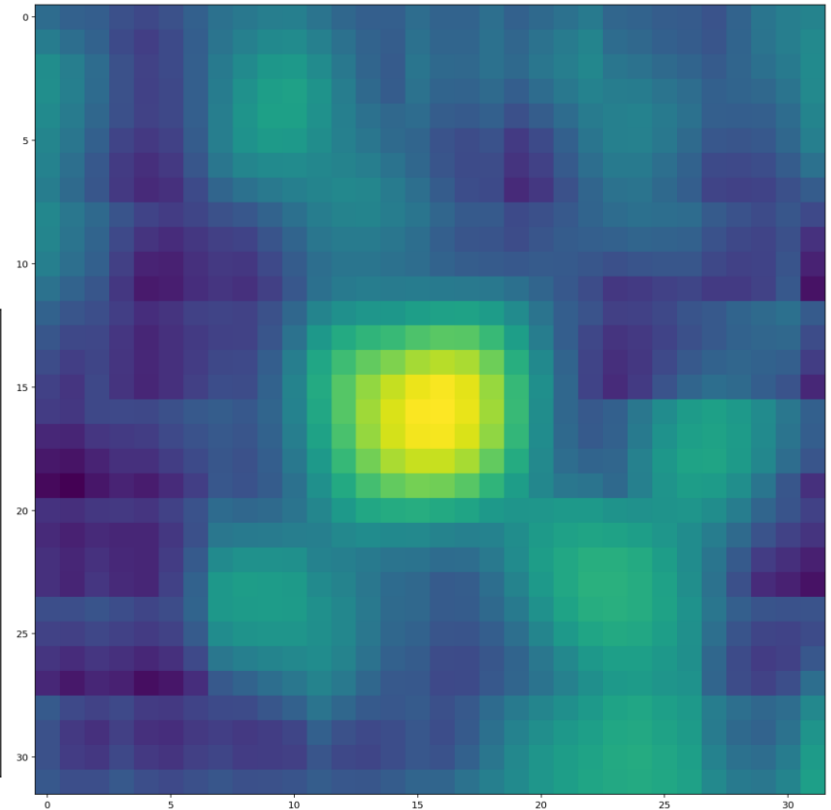
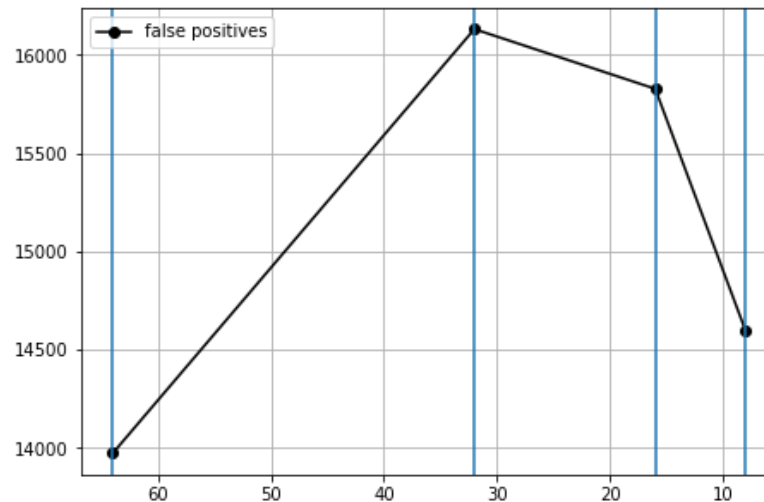
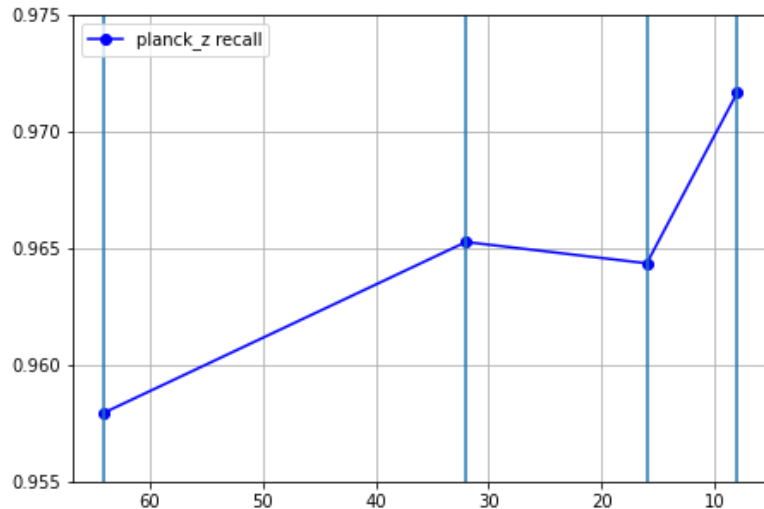


Step = 32 = size / 2

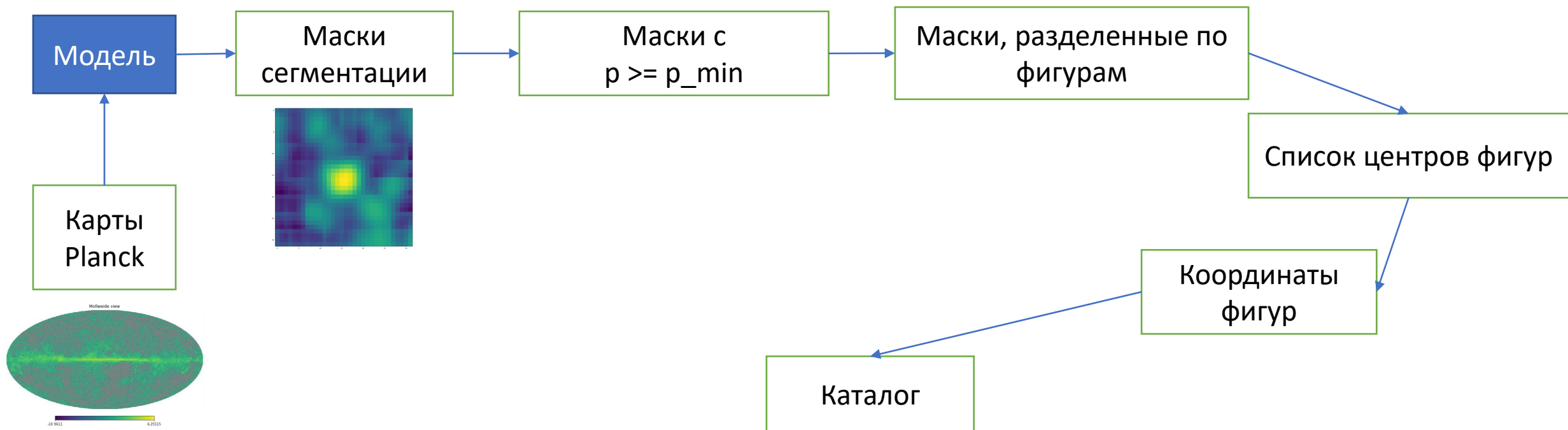


# Сканирование неба моделью

- Маска сегментации с шагом 4 →
- Графики recall и fp ↓



# Детекция

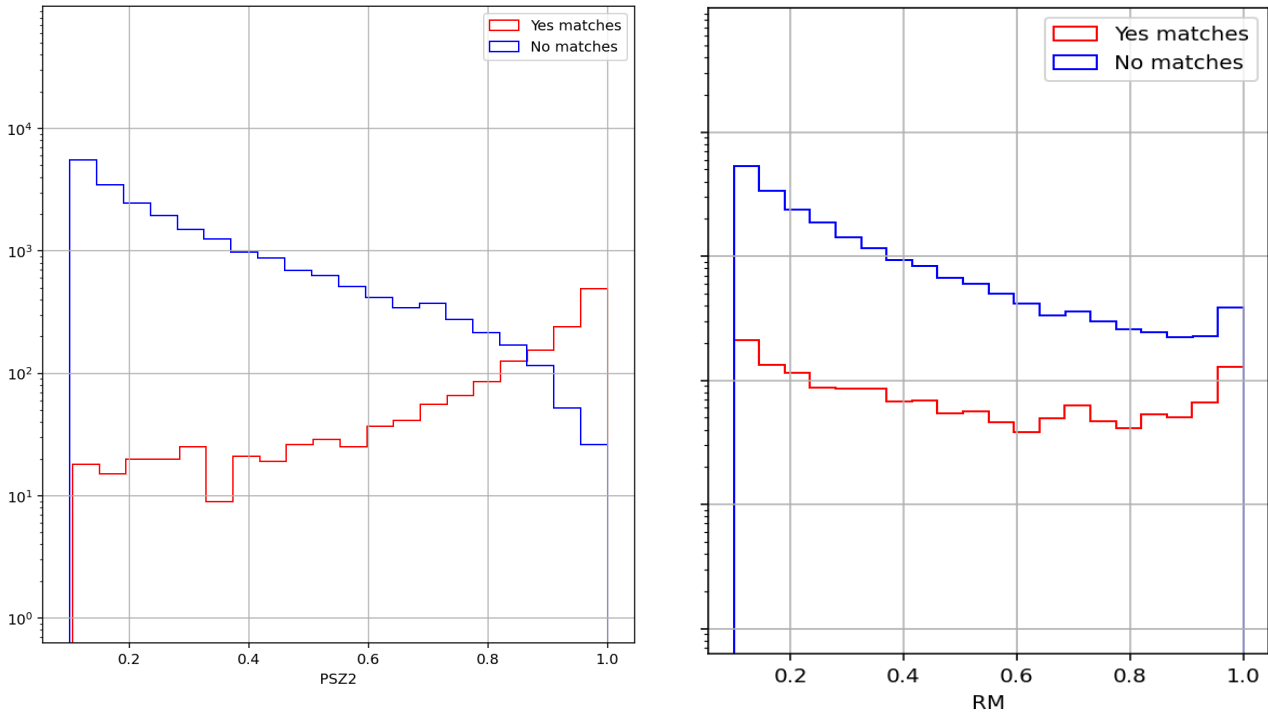


RA	DEC	area	min_rad	max_rad	mean_rad	min_pred	max_pred	tRA	tDEC	status	catalog
259.732221	5.094132	4.0	0.701176	2.119449	1.320277	0.901549	0.933698	NaN	NaN	fp	NaN
283.542185	26.460609	4.0	0.700159	2.110640	1.317352	0.901048	0.957279	283.519410	26.482998	tp	planck_no_z
280.308927	30.912179	7.0	0.700308	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



Результаты

# Выбор лучшей модели

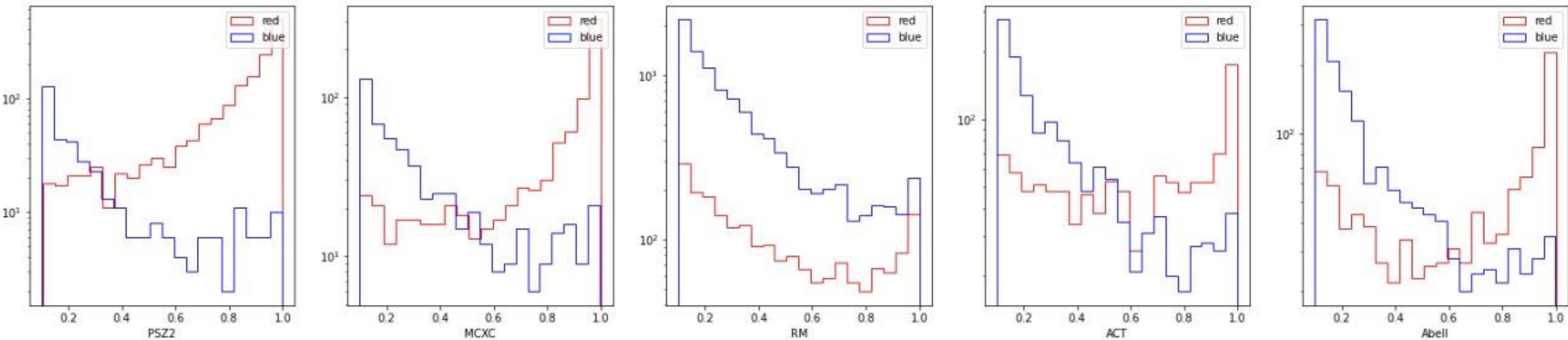


:

	PSZ2	MCXC	RM	ACT
pz14	0.901391	0.418818	0.048485	0.202145
pz40	0.909256	0.424555	0.050362	0.210012
pz_rot_28	0.916515	0.425129	0.056336	0.218832
pz_act10	0.822747	0.379805	0.040519	0.224076
pz_act14	0.737447	0.345382	0.047107	0.449106
pz_act_rot_drop0.1_ep9	0.869328	0.391853	0.041017	0.185697
pz_act_jan_rot_drop0.1_ep6	0.896552	0.405622	0.042970	0.186889
pz_act_feb_rot_drop0.1_ep5	0.915306	0.433161	0.062464	0.228367
pz_act_feb_rot_drop0.2_ep10	0.915306	0.430866	0.060281	0.227175
pz_act_feb_rot_drop0.3_ep14	0.887477	0.426277	0.065566	0.237902
pz_act_found21	0.921960	0.437177	0.057064	0.220977
pz_act_found2_22	0.929825	0.432014	0.056681	0.230751
pz_all_found34	0.922565	0.438325	0.061009	0.235757

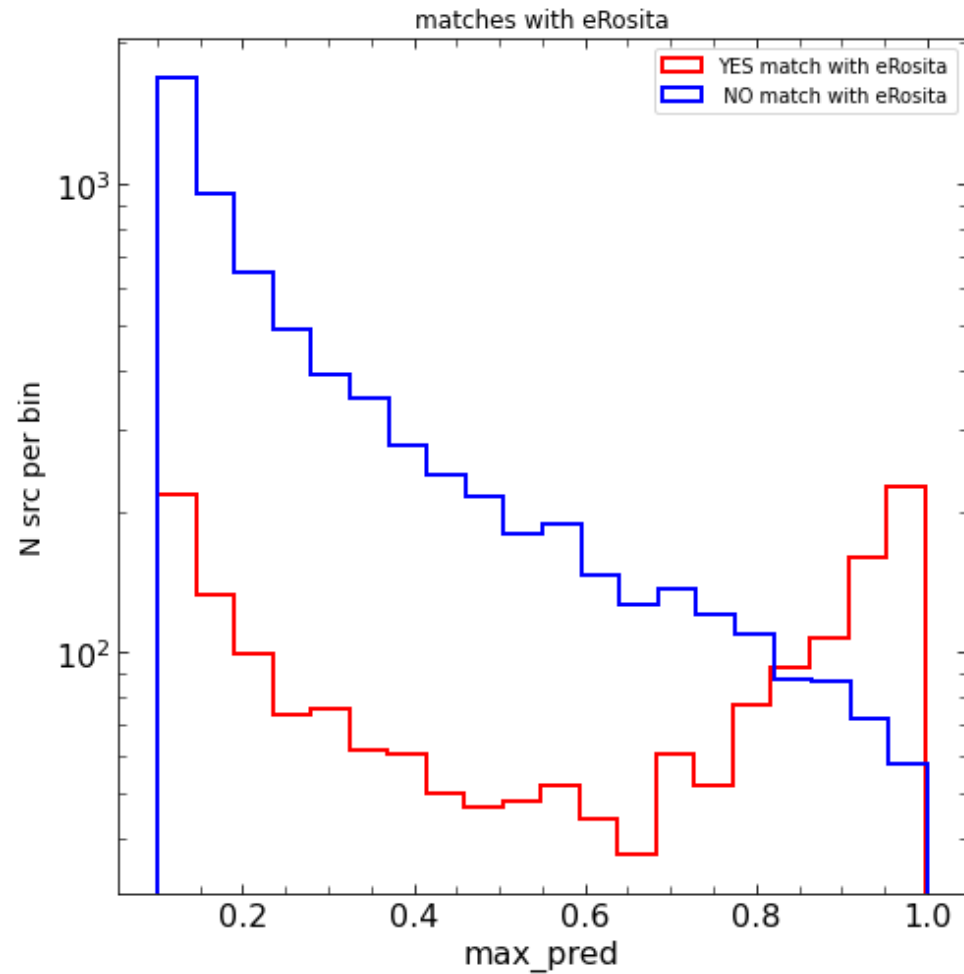
# Сравнение с базовой моделью

Распределение max\_pred pz\_all\_found  
и сопоставление с разными каталогами



	PSZ2	MCXC	RM	ACT	Abell	fp	all
pz_all_found34	0.92	0.44	0.06	0.24	0.21	21018	23352
br_cat	0.74	0.38	0.04	0.19	0.18	11891	13689

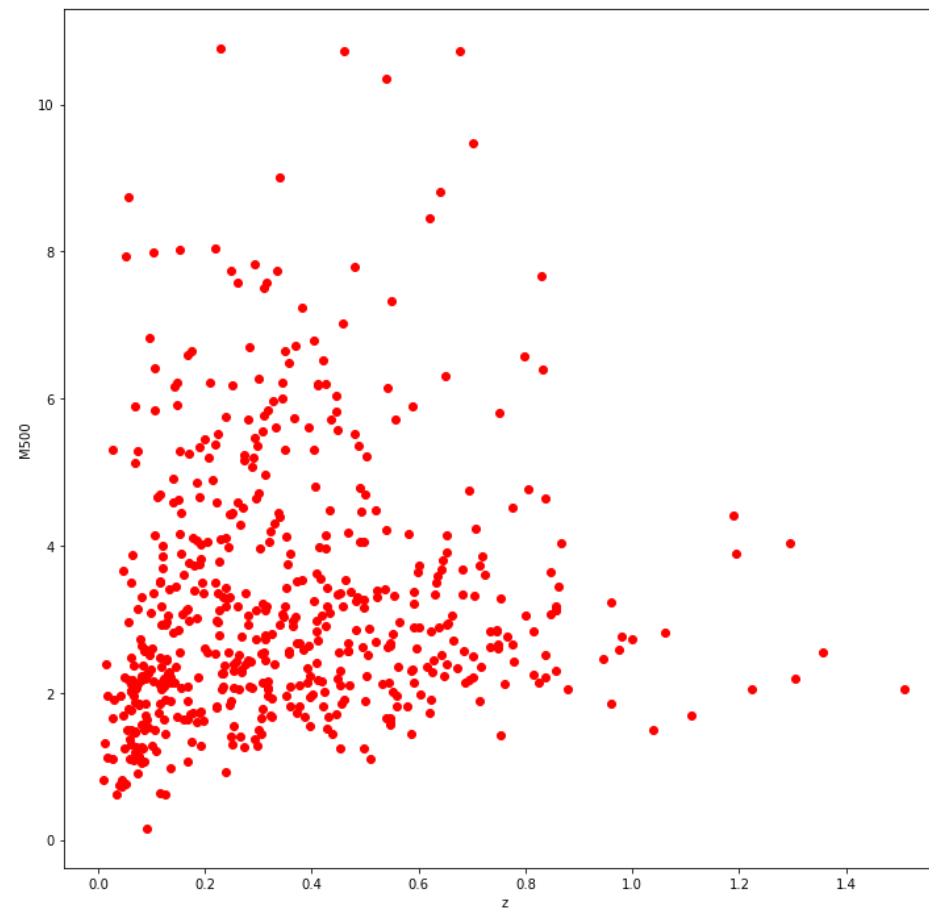
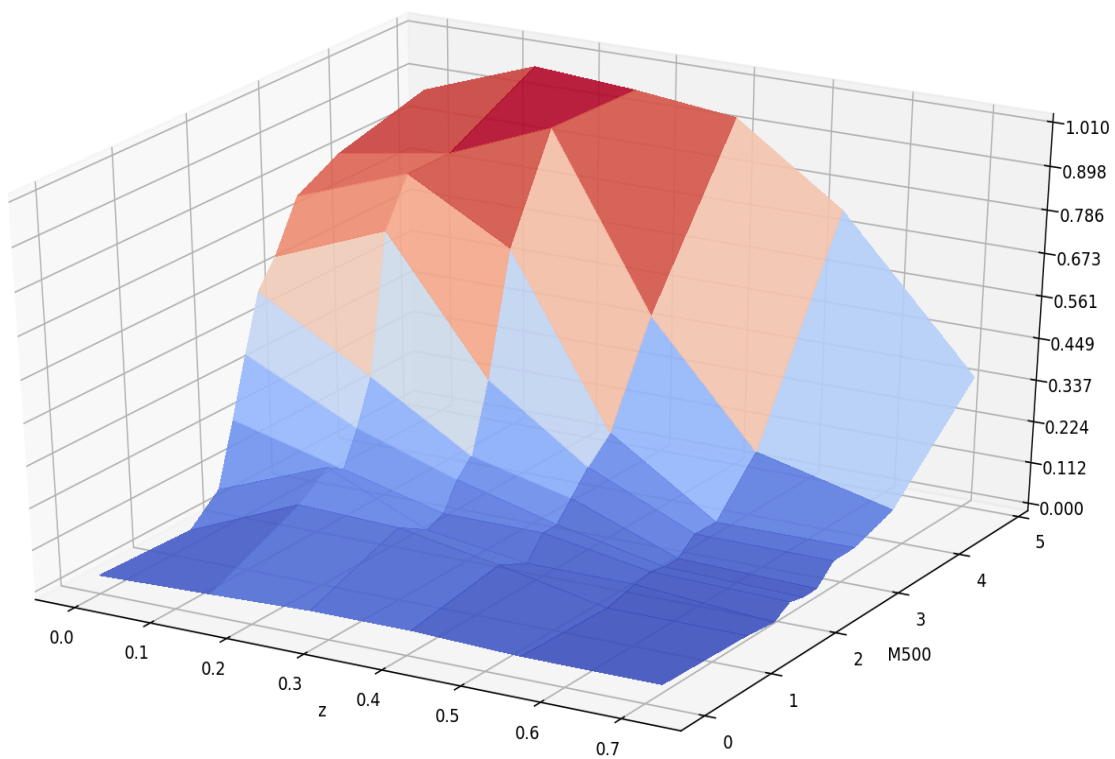
# Сравнение с eROSITA



Matrix of UNIQUE eRosita matches

	ancat	brcat	psz2	redmp
ancat	2355	1022	590	993
brcat	1020	1715	531	641
psz2	589	533	593	227
redmp	994	643	228	4461

# Исследование функции отбора



# Результаты

- Обучены модели для сегментации данных Planck. Используются модели с различными параметрами (Dropout, количество слоёв) и с различными обучающими выборками (в том числе полученными с помощью активного обучения).
- Создан каталог на основе масок сегментации, исследованы его свойства. Recall на лучшем каталоге превосходит на 13% recall базового каталога по PSZ2 и на 5% по АСТ. Этот каталог важен для исследования параметров Вселенной, а также для сравнения с другими каталогами кандидатов скоплений.
- Проведён анализ функции отбора каталога. Более 70% лучшего каталога имеют  $M500 > 4 * 10^{14} \odot$ . Исследовано распределение prediction index полученных моделей и отклика на различных каталогах из разных диапазонов.

# Дальнейшие планы

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