

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

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

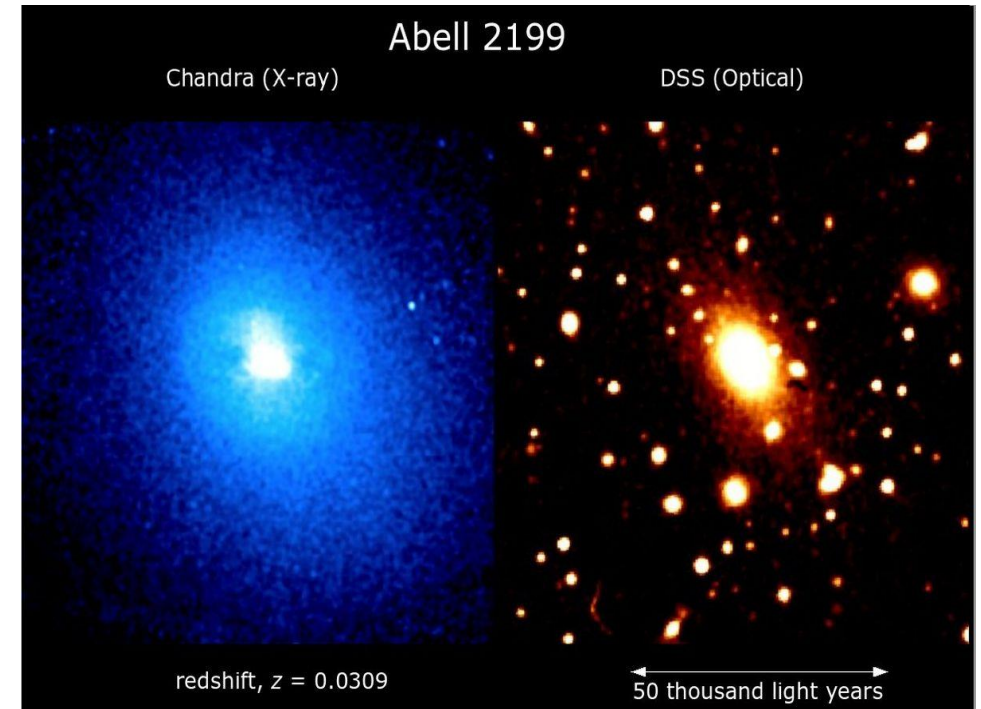
# Рентгеновский телескоп eROSITA

- 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.



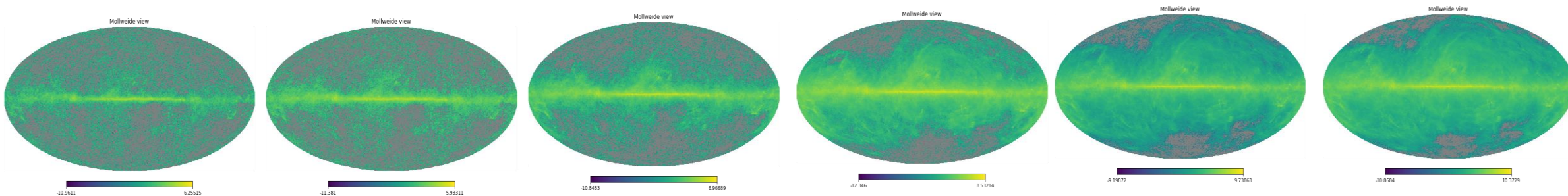
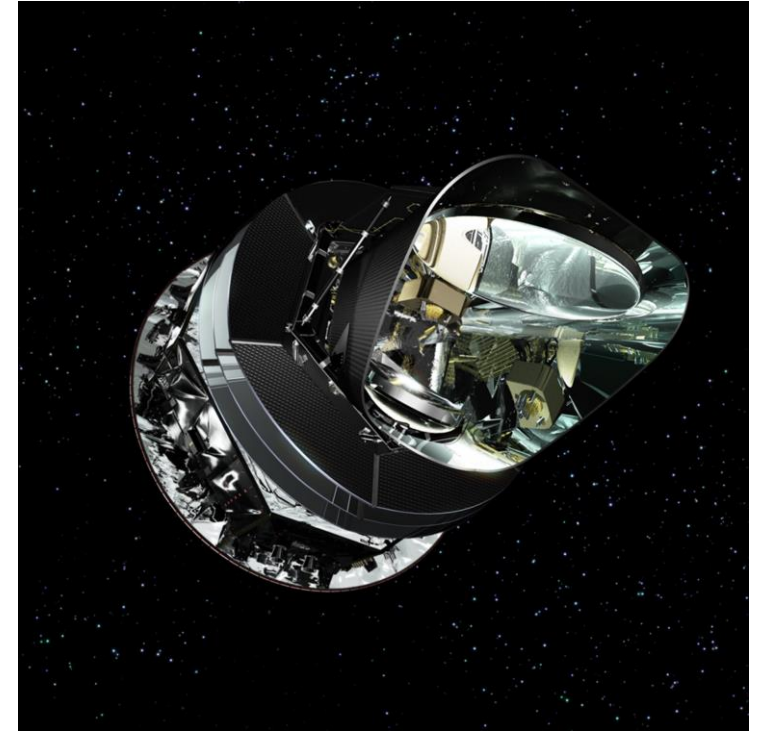
# Скопления галактик

- 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

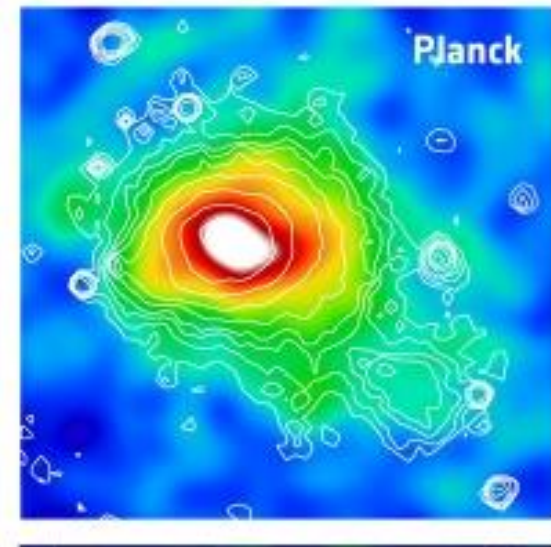
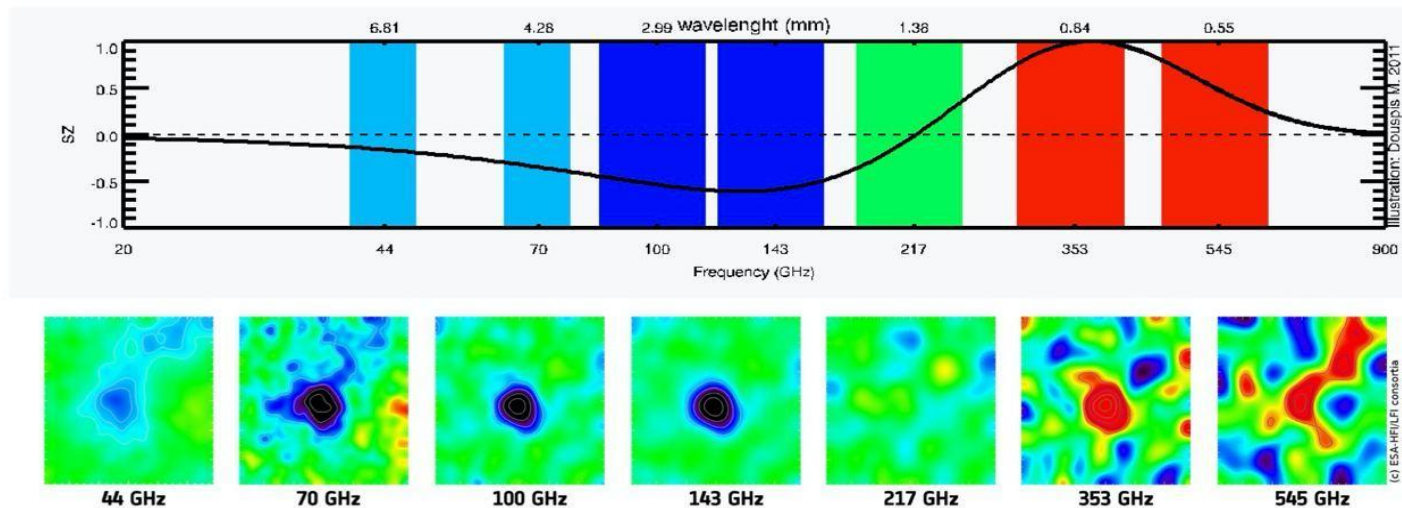
- **Planck** — астрономический спутник Европейского космического агентства, созданный для изучения вариаций космического микроволнового фона — реликтового излучения. Запущен 14 мая 2009 года. В период с сентября 2009 по ноябрь 2010 года «Планк» успешно закончил основную часть своей исследовательской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



Coma Cluster, Planck data

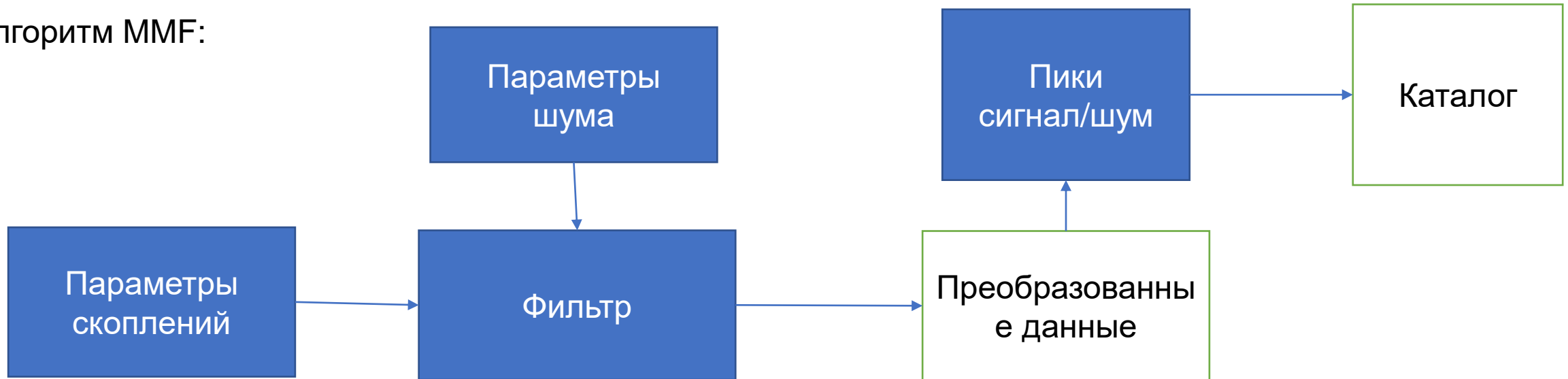
# Каталоги скоплений галактик

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	1094						
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

Алгоритм MMF:



# Постановка задачи

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

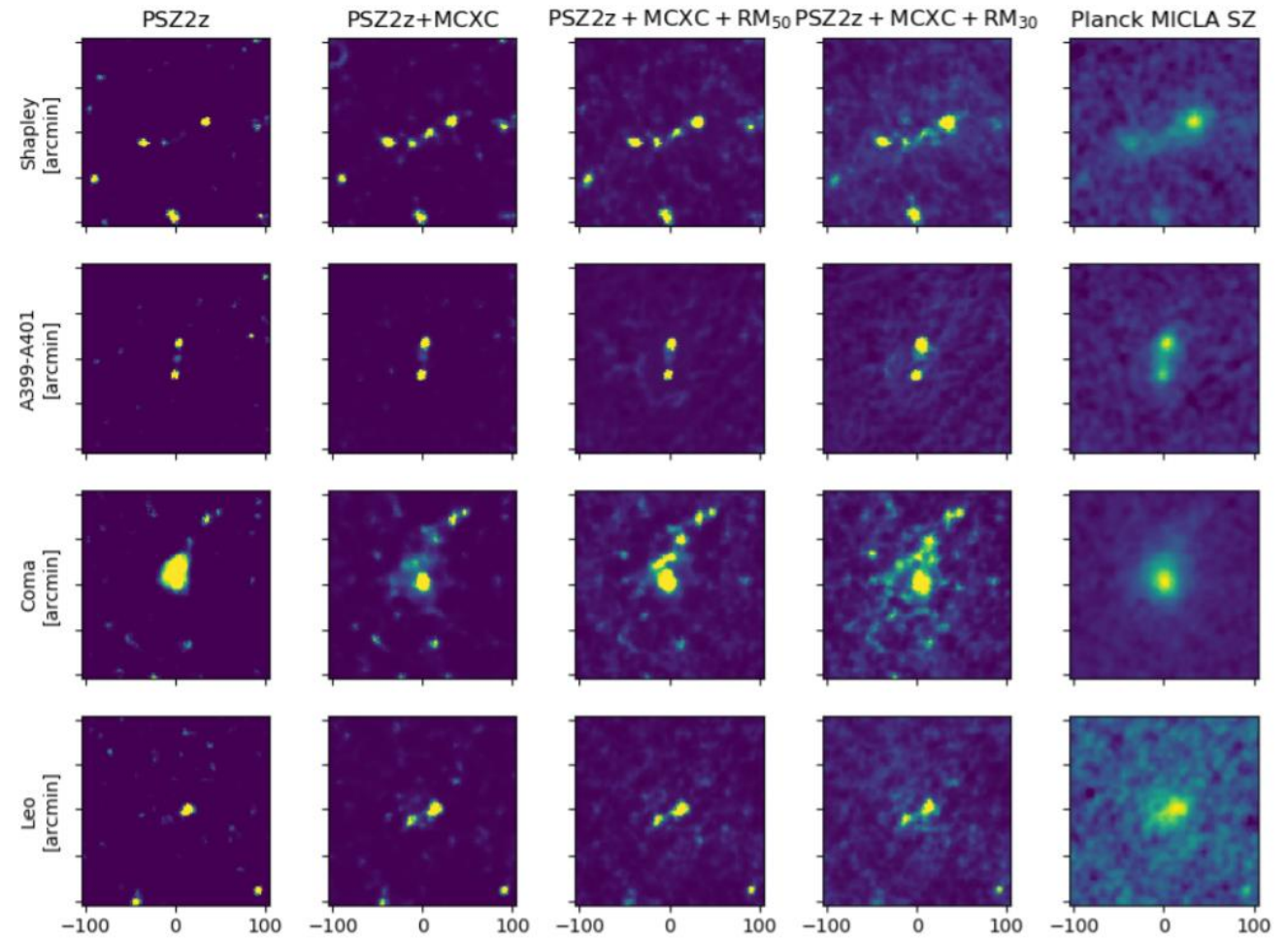
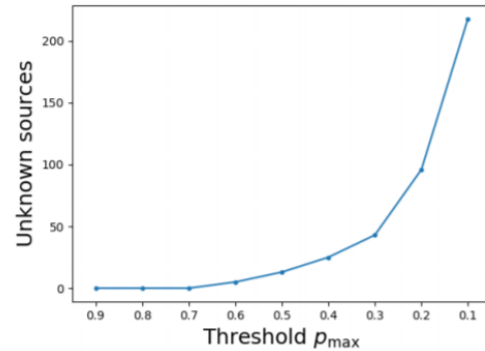
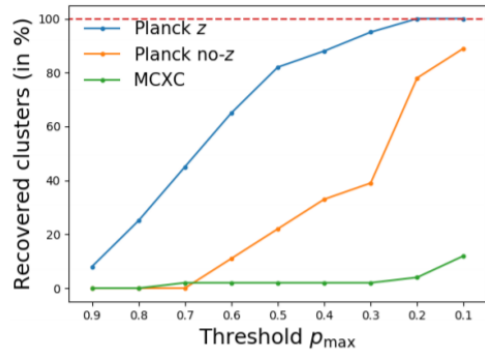
- Создание модели сегментации по данным Planck
- Создание каталога источников СЗ
- Анализ функции отбора каталога ( $z$ , M500)
- Анализ чистоты каталога
- Сравнение с eRosita



# План доклада

- Обзор
  - Deep Learning for Sunyaev-Zel'dovich Detection in Planck», V. Bonjean, 2019
  - U-net
- Построение решения
  - Активное обучение
  - Сканирование неба моделью
  - Детекция
  - Метрики
- Результаты
  - Полученные модели
  - Выбор лучшей модели
  - Сравнение с базовой моделью
  - Сравнение с eROSITA

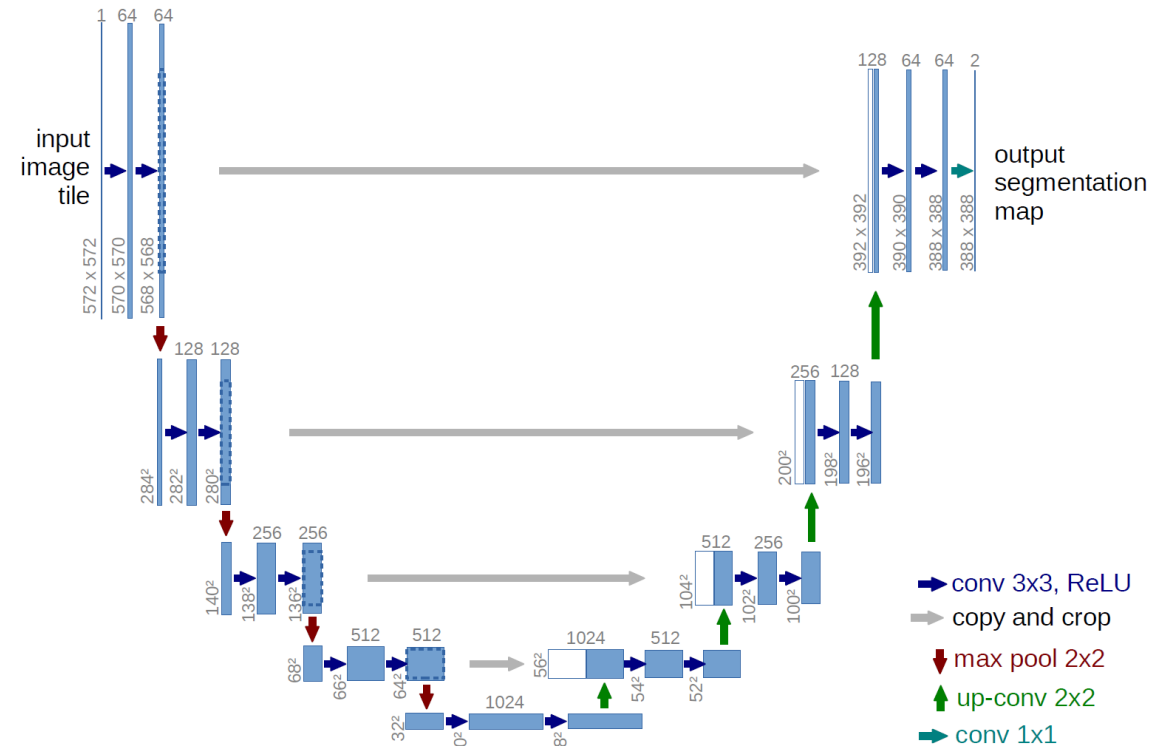
# «Deep Learning for Sunyaev-Zel'dovich Detection in Planck», V. Bonjean, 2019



Автор не опубликовал каталоги и модель, нужно повторить эксперимент

# U-net

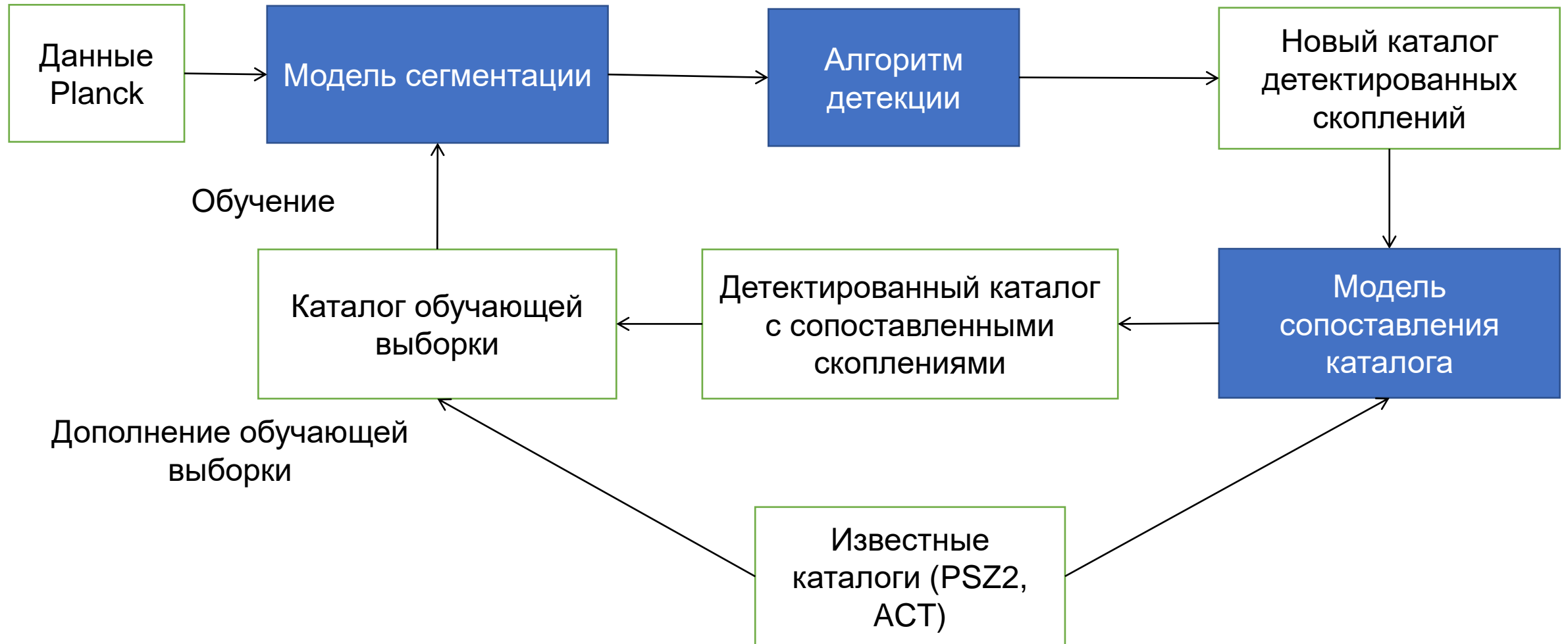
- Одной из лучших нейросетевых архитектур для сегментации изображений является **U-net**. Её ключевой особенностью является наличие skip-connection слоёв, таким образом часть слоёв энкодера конкатенируется со слоями декодера, и при сжатии признаков нейросеть теряет меньше информации об изображении, на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



# План доклада

- Обзор
  - Deep Learning for Sunyaev-Zel'dovich Detection in Planck», V. Bonjean, 2019
  - U-net
- Построение решения
  - Активное обучение
  - Сканирование неба моделью
  - Детекция
  - Метрики
- Результаты
  - Полученные модели
  - Выбор лучшей модели
  - Сравнение с базовой моделью
  - Сравнение с eROSITA

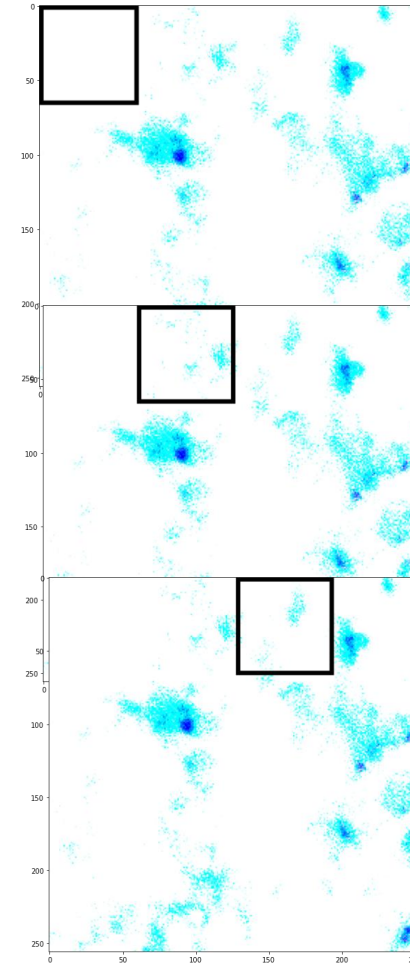
# Активное обучение



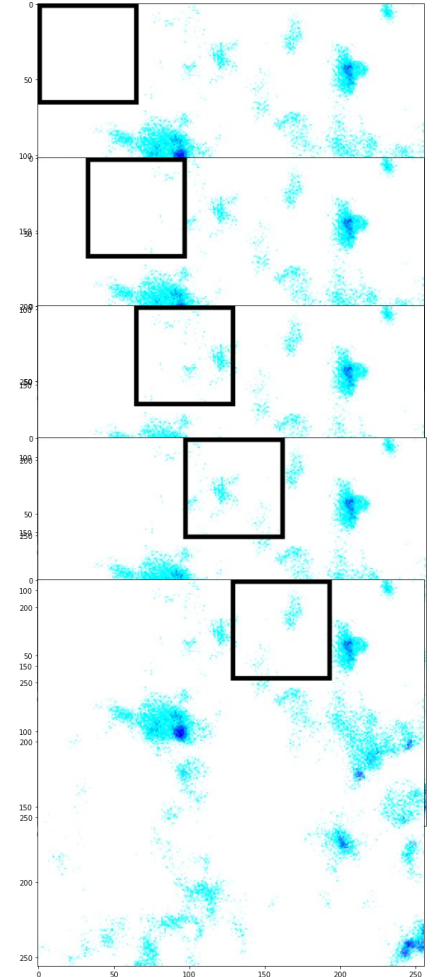
# Сканирование неба моделью

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

Step = 64 = size



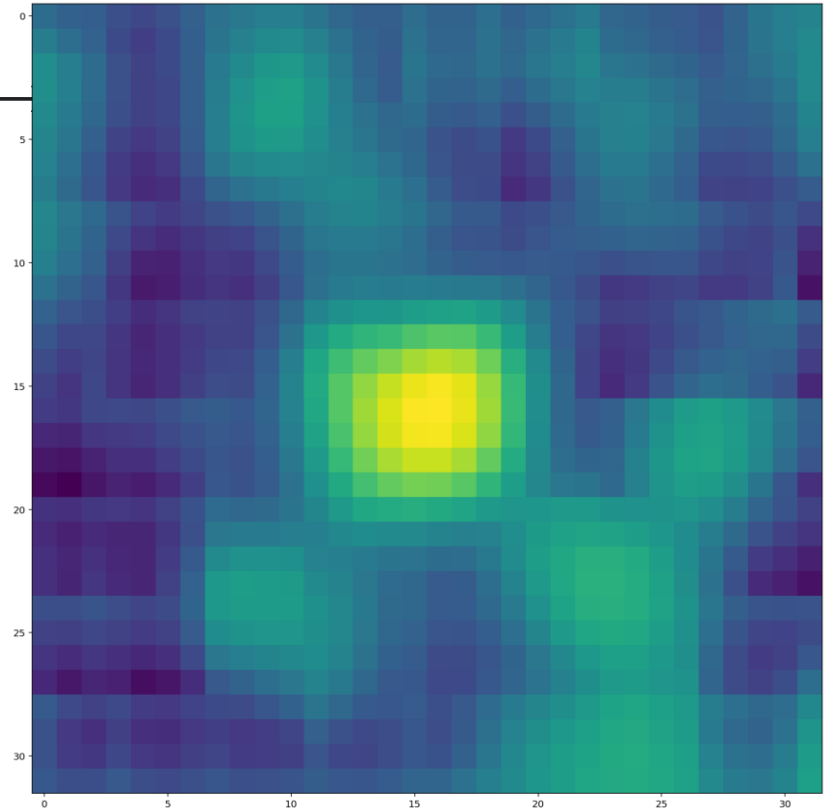
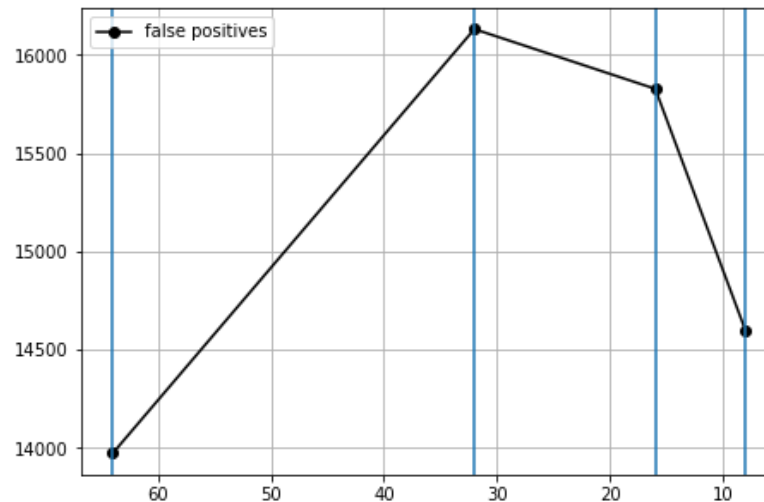
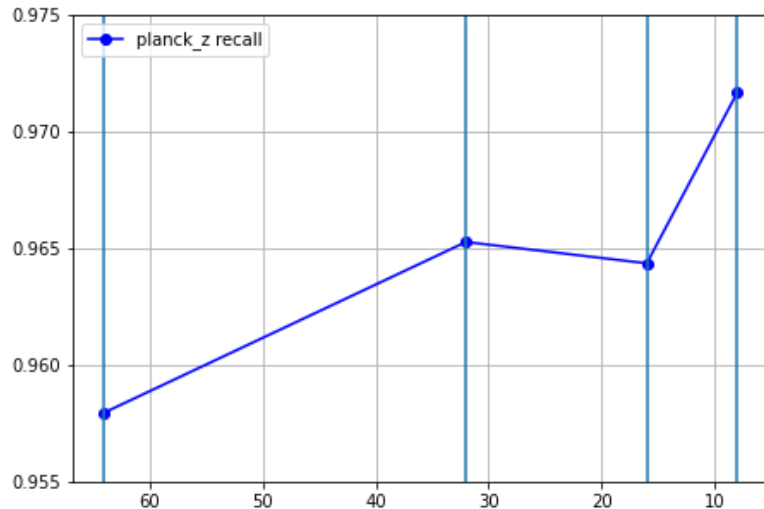
Step = 32 = size / 2



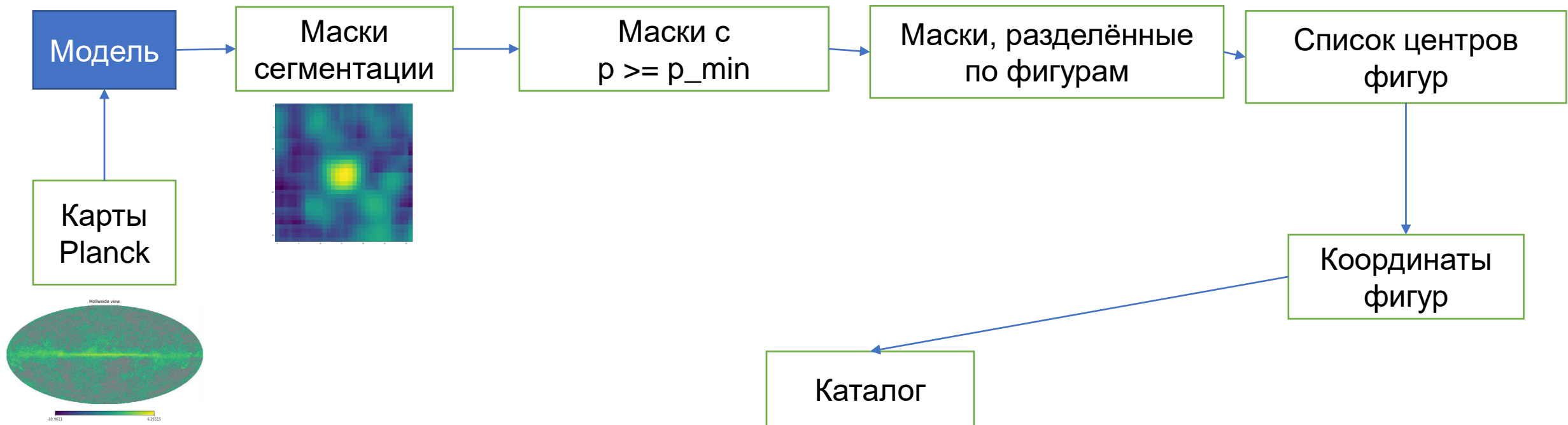


# Сканирование неба моделью

- 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

# Метрики

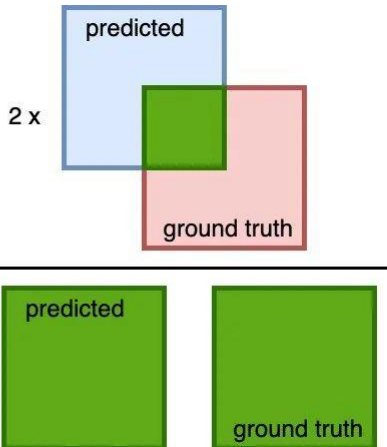
Метрики  
сегментации:


- IoU, Dice

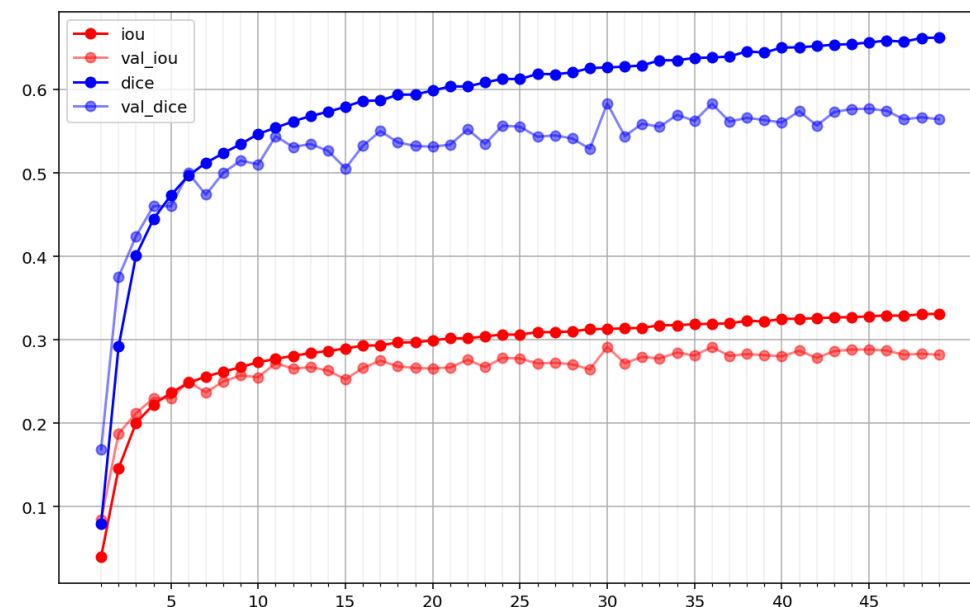
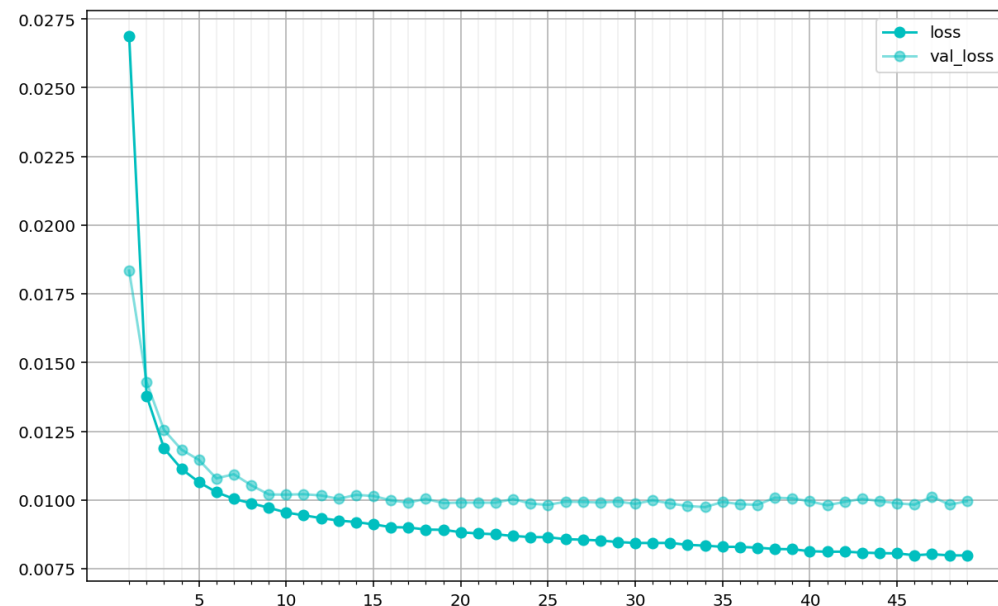
Метрики

детекции:

- Recall,  
Precision

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


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




# План доклада

- Обзор
  - Deep Learning for Sunyaev-Zel'dovich Detection in Planck», V. Bonjean, 2019
  - U-net
- Построение решения
  - Активное обучение
  - Сканирование неба моделью
  - Детекция
  - Метрики
- 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
...	...	...	...	...

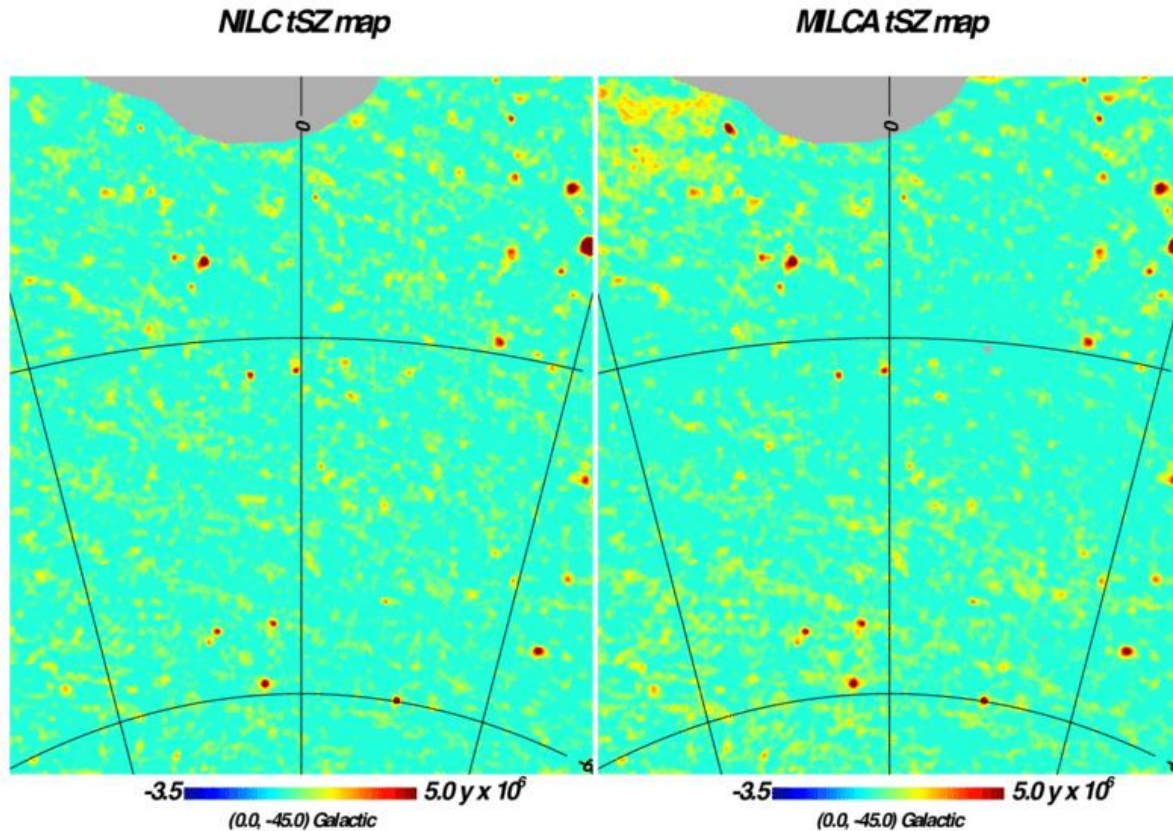
# Выбор лучшей модели

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

	PSZ2	MCXC	RM	ACT	Abell	fp	all
1	0.901391	0.418818	0.048485	0.202145	0.181305	13775	16729
2	0.916515	0.425129	0.056336	0.218832	0.196026	20036	23331
3	0.737447	0.345382	0.047107	0.449106	0.146712	8966	12239
4	0.896552	0.405622	0.042970	0.186889	0.181796	14437	17210
5	0.915306	0.430866	0.060281	0.227175	0.203631	20387	23801
6	0.921960	0.437177	0.057064	0.220977	0.200196	19654	22985
7	0.929825	0.432014	0.056681	0.230751	0.201914	18577	21946
8	0.922565	0.438325	0.061009	0.235757	0.208538	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

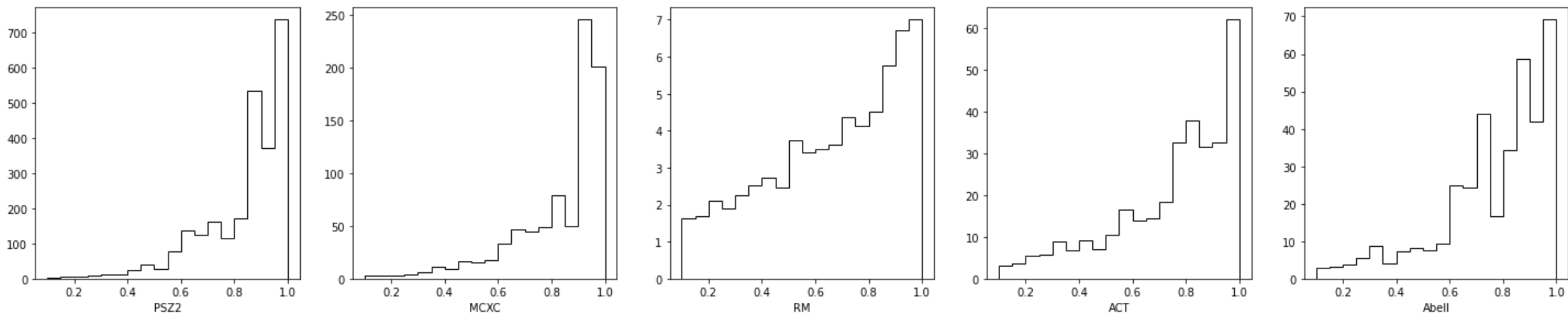
# Параметры каталога

- `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]

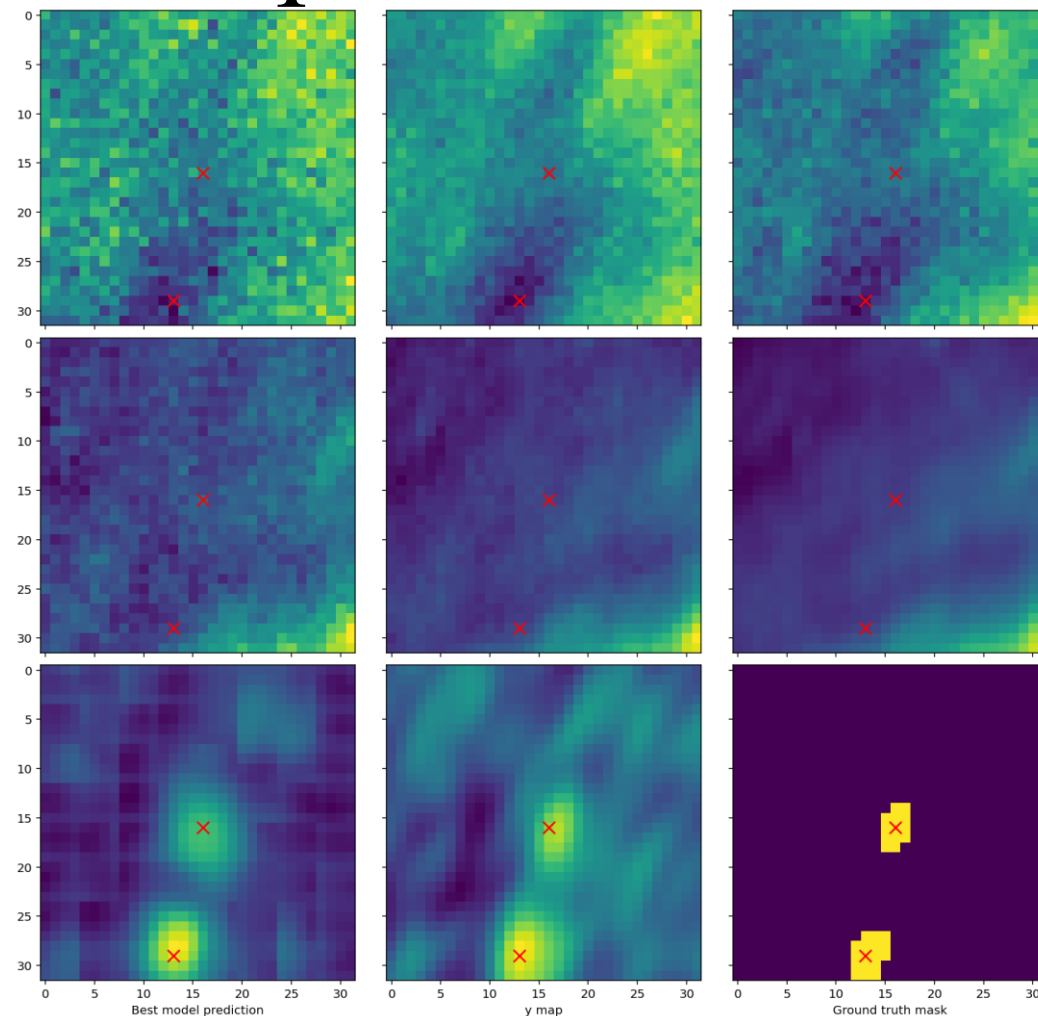
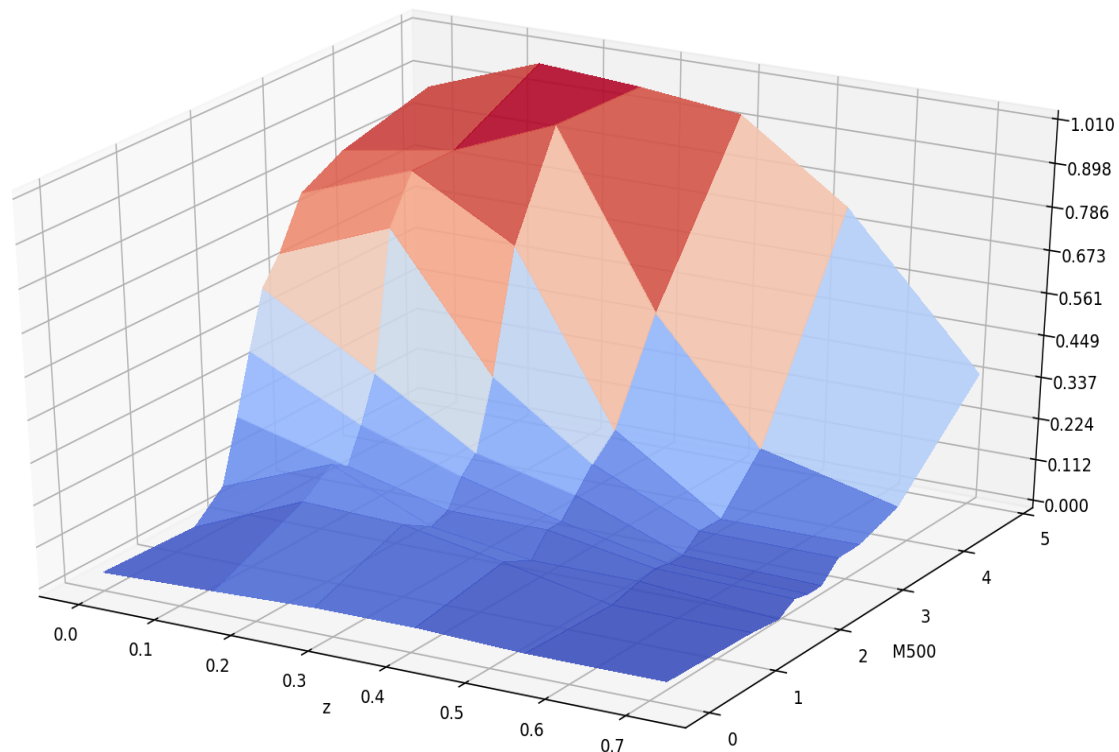
# Чистота выборки

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 13% recall базового каталога по PSZ2 и на 5% по ACT. Этот каталог важен для исследования параметров Вселенной, а также 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  $M500 > 4 \cdot 10^{14} \odot$ . Исследовано распределение prediction index полученных моделей и отклика на различных каталогах из разных диапазонов.