

РАСШИРЕННЫЙ КАТАЛОГ ОБЪЕКТОВ СЮНЯЕВА–ЗЕЛЬДОВИЧА ПО ДАННЫМ СПУТНИКА PLANCK С ИСПОЛЬЗОВАНИЕМ ГЛУБОКОГО ОБУЧЕНИЯ

© 2022 г. А. В. Мещеряков^{1,2*}, А. Немешаева^{1,2},
Р. А. Буренин¹, М. Р. Гильфанов^{1,3}, Р. А. Сюняев^{1,3}

¹Институт космических исследований РАН, Москва, Россия

²Московский государственный университет им. М.В. Ломоносова, Москва, Россия

³Институт астрофизики Общества им. Макса Планка, Гархинг, Германия

Поступила в редакцию 02.08.2022 г.

После доработки 03.08.2022 г.; принята к публикации 03.08.2022 г.

Представлен расширенный каталог источников Сюняева–Зельдовича (СЗ) по данным спутника PLANCK, основанный на соединении выборок, полученных двумя независимыми подходами к детектированию объектов в микроволновых данных: сегментация источников моделью глубокого обучения (семейства U-Net) на картах интенсивности в HFI-каналах 100–850 ГГц дополнена поиском СЗ-источников на готовых картах комптоновского u -параметра (NILC). Путем сравнения полученной выборки СЗ-объектов PLANCK с наиболее полными каталогами скоплений галактик, полученных по микроволновым (PSZ2, ACT) и рентгеновским (SRG/ePOZITA предварительный каталог скоплений 1 года на восточном галактическом небе; MCXC по наблюдениям спутника ROSAT) обзорам неба, мы проанализировали функцию отбора скоплений галактик от массы (M_{500}), красного смещения (z) и рентгеновского потока ($F_{X,0.5-2}$). Мы видим два сценария использования полученных нами расширенных каталогов СЗ-источников по данным PLANCK: (а) выборка только надежных СЗ-объектов из каталога кандидатов может использоваться непосредственно, например, для поиска в ней уникальных объектов или исследования популяции скоплений, (б) весь каталог СЗ-объектов содержит значительное число ложных источников, но при этом располагает практически всеми возможными детектированными скоплениями галактик в данных PLANCK (он может использоваться совместно с другими каталогами кандидатов в скопления галактик, полученных независимо из данных в других диапазонах или другими инструментами).

Ключевые слова: активные галактики, галактики с большим красным смещением, ядра галактик, рентгеновские лучи.

DOI: 10.31857/S0320010822090066

1. ВВЕДЕНИЕ

Уже в 30-х гг. XX в. стало понятно, что галактики на небесной сфере на угловых масштабах \sim мегапарсек не образуют равномерного распределения на небе и группируются в скопления галактик, состоящие главным образом из темного вещества (Цвикки, 1937). В наблюдаемой Вселенной крупномасштабная структура материи образует ячеистую структуру, напоминающую паутину, рост которой происходит от меньших объектов к большему (Бонд и др., 1996; де Лаппарент и др., 1986): темное и барионное вещество по мере эволюции Вселенной вытекает из областей низкой плотности (пустоты, англ. voids) и формирует стены

и волокна. Скопления и группы галактик расположены на волокнах крупномасштабной структуры и представляют собой самые крупные динамически связанные объекты во Вселенной. История формирования таких объектов, а также их физические характеристики (в первую очередь масса) сильно зависят от космологических параметров (в частности, от доли темной энергии в плотности Вселенной в современную эпоху). Изучение скоплений галактик представляет огромный интерес как с точки зрения физики формирования и эволюции галактик в различном окружении, так и для измерения параметров космологической модели Вселенной.

Видимая в электромагнитном излучении (барионная) компонента скоплений — горячий межгалактический газ и звезды — дает возможность ре-

*Электронный адрес: mesch@cosmos.ru

гистрировать скопления галактик в трех основных спектральных диапазонах: рентгеновском, микроволновом и оптическом. Скопления детектируются благодаря различным физическим эффектам: в рентгеновском диапазоне — по излучению горячего межгалактического газа, разогретого до вириальной температуры в потенциальной яме скопления; в микроволновом диапазоне — по обратному комптоновскому рассеянию фотонов реликтового излучения на горячих тепловых электронах скопления (тепловой эффект Сюняева–Зельдовича); в оптическом ИК-диапазоне — по излучению звездной компоненты в галактиках и межгалактической среде.

К настоящему времени по наблюдениям в различных диапазонах на всем небе были обнаружены тысячи скоплений галактик. Первым из крупных каталогов скоплений галактик стал оптический каталог Эйбела, содержащий данные о богатых скоплениях галактик на красных смещениях $z < 0.2$ (первоначально найденных вручную в Паломарском обзоре северного неба). В последней версии каталога Эйбела (Эйбл и др., 1989) содержатся данные по 4073 скоплениям на всем внегалактическом небе. Современным продолжением оптических обзоров скоплений галактик является каталог redMaPPer (Рыкофф и др., 2014), который (v.6.3) содержит 26 111 скоплений и групп галактик на красном смещении $z \lesssim 0.6$, отобранных в полях фотометрического обзора SDSS (DR8) на основе автоматического метода идентификации галактик «красной последовательности». Необходимо отметить, что звездная компонента, видимая в оптическом излучении галактик, отражает лишь малую долю общего бюджета массы гало (скопления) внутри вириального радиуса ($\approx 1\text{--}4\%$, см. Кравцов и др., 2018). Поэтому наблюдения в рентгеновском и микроволновом диапазонах (напрямую связанные с излучением газовой компоненты, которая доминирует в барионной массе скоплений) являются более эффективным инструментом для составления полной выборки скоплений вплоть до определенной массы (англ. mass-limited sample), необходимой для космологии. В данной работе мы будем сравнивать наши результаты с каталогами скоплений, полученных по данным рентгеновского и микроволнового диапазонов.

Поиск скоплений в рентгеновском диапазоне в настоящее время осуществляется по данным трех рентгеновских обсерваторий: ROSAT, XMM-Newton, SRG/eROZITA (Сюняев и др., 2021; Предель и др., 2021). Рентгеновский каталог 1646 скоплений галактик X-CLASS (Кулуридис и др., 2021) получен по данным наблюдений обсерватории XMM-Newton и покрывает лишь незначительную часть небесной сферы (269 кв. град). Космические обсерватории ROSAT и SRG/eROZITA

в режиме обзора покрывают наблюдениями все небо. Компиляция имеющихся рентгеновских выборок скоплений галактик, полученных различными группами по архивным данным обзоров спутниками ROSAT и Einstein, собраны в мета-каталоге МСХС (Пиффаретти и др., 2011), содержащем 1743 объекта с надежными оптическими отождествлениями. В настоящее время на орбите находится рентгеновская обсерватория SRG (Сюняев и др., 2021) с телескопом eROZITA (Предель и др., 2021), закончившим четыре полных скана неба (двухлетний обзор) в рентгеновском диапазоне 0.53–8 кэВ. В случае выполнения всей программы обзорной части миссии SRG/eROZITA планируется обнаружить рекордные 100 000 рентгеновских скоплений галактик (Мерлони и др., 2012) на всем небе. После выполнения восьми сканов всего неба финальный каталог рентгеновских скоплений SRG/eROZITA будет обладать рекордной полнотой отбора массивных скоплений $M_{500} \gtrsim 3 \times 10^{14} M_{\odot}$ во всей видимой Вселенной. В настоящее время, по данным годового обзора SRG/eROZITA, на восточном внегалактическом небе (входящем в область ответственности Российского Консорциума SRG/eROZITA) сформирован предварительный каталог протяженных рентгеновских источников, насчитывающий более 25 тыс. объектов в восточной галактической полусфере, который будет служить нам в данной работе основной тестовой выборкой скоплений галактик.

Поиск скоплений галактик по эффекту Сюняева–Зельдовича в микроволновом диапазоне в настоящее время осуществляется на основе данных трех обсерваторий: PLANCK, ACT (англ. Atacama Cosmology Telescope) и SPT (англ. South Pole Telescope). Опубликован ряд каталогов — PSZ2 (Коллаборация Планк и др., 2016b), ACT (Хилтон и др., 2021), SPT-SZ (Боскуэт и др., 2019), SPTpol (Блим и др., 2020). Важным преимуществом данных обзора PLANCK является то, что он покрывает наблюдениями всю небесную сферу. Каталог PSZ2 (Коллаборация Планк и др., 2016) содержит 1653 СЗ-объектов, из них 1203 являются оптически подтвержденными скоплениями галактик (на 86% площади неба вне Млечного Пути). В то же время наибольшее число СЗ-объектов найдено по данным ACT, который обладает более хорошим пространственным разрешением, чем PLANCK, и позволяет находить далекие маломассивные скопления (но покрывает наблюдениями лишь $\sim 1/3$ площади неба). Каталог ACT DR5 SZ (Коллаборация Планк и др., 2016b) содержит только оптически подтвержденные скопления галактик: 4195 объектов на площади 13 211 кв. град. Все указанные выше каталоги содержат СЗ-объекты, отобранные на высоком уровне значимости. Тем не

менее, если рассматривать только микроволновые данные без идентификации скоплений в оптическом диапазоне, то число ложных СЗ-источников в данных микроволновых обсерваторий оказывается довольно значительным и не вполне соответствует выбранному высокому уровню значимости при отборе детектируемых источников. Так, в каталоге PSZ2 содержатся объекты, детектированные выше порога $S/N > 4$, при этом чистота выборки составляет 86.2% (Агуадо-Барохона и др., 2019), появление ложных детектирований СЗ-источников коррелирует с областями теплового излучения пыли в Галактике.

Увеличение чистоты каталогов скоплений галактик может помочь комбинация независимых выборок скоплений, найденных в различных спектральных диапазонах. Так, выборка ComPRASS (Таррио и др., 2019) была получена на основе одновременного поиска скоплений в рентгеновских (ROSAT) и микроволновых (PLANCK) данных на всем небе и обладает высокой чистотой (>84.5%). Также делаются попытки объединения данных, полученных различными микроволновыми инструментами (в частности, каталог скоплений PSZSPT получен на основе комбинации данных PLANCK и SPT, Мелин и др., 2021), и попытки поиска скоплений галактик путем комбинации данных в рентгеновском и оптическом диапазонах (например, каталог MARD-Y3 (Кляйн и др., 2019) получен на основе одновременного использования рентгеновских (2RXS ROSAT) и оптических (DES-Y3) измерений в области 5000 кв. градусов). Отметим, что в случае одновременного поиска скоплений, по данным в различных спектральных диапазонах, требования к значимости детектирования объектов в каждом отдельном обзоре можно снижать, сохраняя высокую чистоту детектирования скоплений галактик в объединенном каталоге.

Из сказанного выше следует важность формирования как можно более полных каталогов скоплений галактик по имеющимся данным обзорам неба в разных спектральных диапазонах, вплоть до объектов с низкой значимостью. Такие каталоги в дальнейшем могут быть эффективно скомбинированы по данным из разных диапазонов. Целью настоящей работы является создание расширенного каталога СЗ-объектов по данным спутника PLANCK. Микроволновые данные PLANCK хорошо дополняют возможности поиска скоплений по рентгеновским данным SRG/ePOZITA — по обоим спутникам доступны данные по всему небу.

Статья организована следующим образом. В разделе 2 описаны используемые нами данные спутника PLANCK (карты интенсивности HFI и карты y -параметра комптонизации) и каталоги скоплений галактик, полученные по данным крупнейших обзоров неба в рентгеновском (ROSAT,

SRG/ePOZITA) и микроволновом (PLANCK, ACT) диапазонах. В разделе 3 описаны метод детектирования СЗ-источников на y -картах и нейросетевая модель сегментации и детектирования скоплений по картам интенсивностей в каналах HFI. В разделе 4 обсуждаются полученные нами результаты. В последнем разделе приведены наши выводы.

2. ДАННЫЕ

2.1. Данные PLANCK для сегментации СЗ-объектов

Космическая миссия PLANCK работала на орбите в 2009–2013 гг., с ее помощью была получена карта всего неба в нескольких диапазонах микроволнового излучения двумя приборами: низкочастотным (LFI, Low Frequency Instrument, 30–70 ГГц, Коллаборация Планк и др., 2020b) и высокочастотным HFI (High Frequency Instrument, 100–857 ГГц, Коллаборация Планк и др., 2020a) приемниками. Последний обладает более хорошим пространственным разрешением и лучше приспособлен для обнаружения скоплений галактик по эффекту Сюняева–Зельдовича, что является целью настоящей работы. Далее мы будем использовать только данные HFI.

В архиве данных PLANCK¹ доступны разные версии карт всего неба, мы будем использовать данные Planck Public Data Release 3 в формате HEALPix (Горски и др., 2005). Средний размер пиксела на HEALPix-картах HFI составляет $1.72'$ ($N_{\text{side}} = 2048$).

Для поиска СЗ-источников базовым алгоритмом (Буренин, 2017) и сравнения с картами сегментации СЗ-объектов, полученными нейросетевым алгоритмом, мы будем использовать карты параметра Комптона миссии PLANCK (далее y -карты), полученные с помощью алгоритма разделения компонент NILC (см. Коллаборация Планк и др., 2020a).

2.2. Каталоги скоплений галактик

Для обучения моделей детектирования СЗ-источников, а также их тестирования мы будем использовать каталоги галактик, обнаруженных в двух основных спектральных диапазонах: микроволновом (PSZ2, ACT) и рентгеновском (MCXC, SRG/ePOZITA). В табл. 1 представлена сводная статистика по рассматриваемым каталогам скоплений — общее число скоплений, количество объектов в тестовых областях, а также число объектов в восточной внегалактической области

¹<https://lambda.gsfc.nasa.gov/product/planck/curr/>

Таблица 1. Каталоги скоплений галактик по данным в микроволновом и рентгеновском диапазонах

Каталоги	PSZ2/PSZ2(z)	ACT	MCXC	ePOZITA
Спектральный диапазон	микроволновой		рентгеновский	
Объекты	1653/1094	4195	1743	$\sim 2.5 \times 10^4$
Тестовые области				
9 ($b = 41.81^\circ, l = 247.5^\circ$)	42/29	143	38	
41 ($b = -41.81^\circ, l = 247.5^\circ$)	58/53	328	68	
38 ($b = -41.81^\circ, l = 112.5^\circ$)	60/50	131	60	$\sim 10^3$
6 ($b = 41.81^\circ, l = 112.5^\circ$)	58/40		70	$\sim 2 \times 10^3$
ePOZITA (1-й год)	692/511	1591	900	$\sim 2 \times 10^4$
$ b > 20^\circ, 0 < l < 180^\circ$				

СРГ/ePOZITA. Так как каталог ePOZITa еще не опубликован, мы приводим эти числа с точностью до первой значащей цифры. В каталоге PSZ2 только часть СЗ-объектов имеет надежные оптические отождествления, мы обозначаем эту подвыборку PSZ2(z) и отдельно приводим ее характеристики в таблице, рядом с данными по всему каталогу PSZ2. На рис. 1 показано распределение объектов из каталогов скоплений PSZ2, ACT, MCXC на небе. Распределение на небе скоплений СРГ/ePOZITA будет приведено в работе по каталогу скоплений ePOZITa.

На рис. 2 показано распределение объектов из каталогов скоплений по красному смещению (левая панель), полной массе внутри радиуса R_{500} (правая панель). На графиках показано распределение для выборки PSZ2(z) скоплений PLANCK с надежными красными смещениями и для компилятивного каталога скоплений при объединении выборок PSZ2(z), ACT и MCXC. Как можно видеть из графиков, данные ACT серьезно дополняют выборку скоплений PLANCK в сторону объектов меньших масс и больших красных смещений (при этом нет гарантии, что все эти дополнительные объекты могут быть обнаружены в данных PLANCK).

Данные обзора СРГ/ePOZITA позволяют регистрировать скопления галактик на существенно меньших рентгеновских потоках, чем это мог делать ROSAT. Благодаря уникальным данным СРГ/ePOZITA, число известных рентгеновских скоплений возросло более чем на порядок.

В следующих частях раздела каждый используемый здесь каталог скоплений будет описан более подробно.

2.2.1. PSZ2. PSZ2 (Коллаборация Планк и др., 2016b) представляет собой каталог источников Сюняева–Зельдовича (СЗ), обнаруженных на всем небе по 29-месячным данным миссии PLANCK. Для обнаружения СЗ-объектов в каталоге PSZ2 использовались алгоритмы MMF1, MMF3 (Мелин и др., 2006), PowellSnakes (Капвальйо и др., 2009, 2012). Каталог построен как объединение результатов поиска СЗ-объектов, найденных MMF1/MMF3/PowellSnakes на высоком уровне значимости $S/N > 4.5$, и содержит в общей сложности 1653 источника. Отметим, что PSZ2 содержит несколько сотен ложных СЗ-объектов, чистота каталога PSZ2 (по данным Коллаборация Планк и др., 2020b) составляет не менее 83%. Для выборки всех подтвержденных по оптическим данным 1094 скоплений из PSZ2 с измеренными спектральными красными смещениями (далее PSZ2(z)) были сделаны оценки масс M_{500} (Коллаборация Планк и др., 2014). Мы будем использовать PSZ2(z) в качестве основы для создания тренировочной выборки при обучении нейросетевых моделей в настоящей работе.

2.2.2. Каталог скоплений ACT DR5 SZ. Это крупнейший на сегодня каталог СЗ-объектов (Хилтон и др., 2021) состоит из 4195 оптически подтвержденных скоплений галактик в области 13168 кв. град (1649 скоплений каталога имеют спектроскопические измерения красных смещений) и получен по данным Атакамского космологического телескопа (ACT, Atacama Cosmology Telescope). Источники Сюняева–Зельдовича с высоким уровнем значимости ($S/N > 4$) были отобраны на картах ACT по наблюдениям 2008–2018 гг. с использованием метода многочастотного согласованного фильтра (MMF, multi-frequency

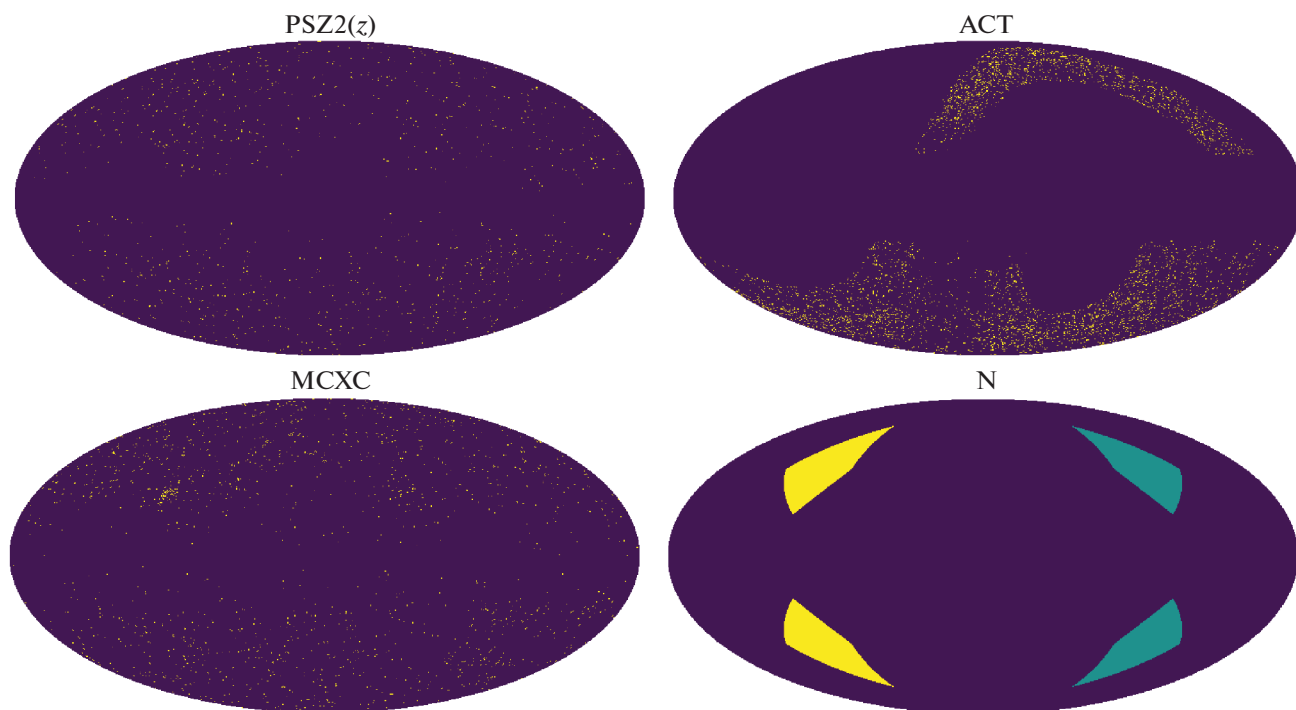


Рис. 1. Распределение скоплений галактик по небу (север вверху, восток слева) в различных каталогах по данным рентгеновского спутника ROSAT и экспериментов в микроволновом диапазоне (PLANCK, ACT). На правой нижней панели показано расположение восточных (слева) и западных (справа) тестовых областей. Номера областей соответствуют индексу разбиения HEALPix с $N_{\text{side}} = 2 : 6$ (северо-восточная), 38 (юго-восточная), 9 (северо-западная), 41 (юго-западная).

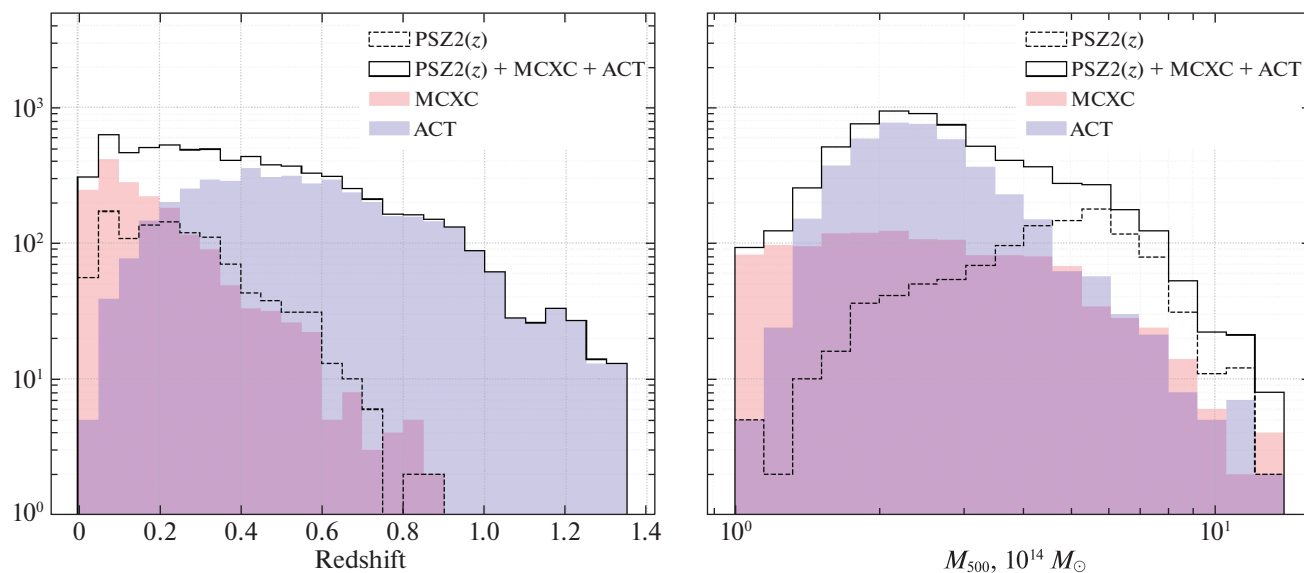


Рис. 2. Распределение по красному смещению z (слева) и массе M_{500} (справа) для каталогов скоплений галактик PSZ2(z), ACT, MCXC, а также для компилятивной выборки скоплений PSZ2(z) + MCXC + ACT.

matched filter, см. Мелин и др., 2006) для двух рабочих диапазонов инструмента (98 и 150 ГГц). Скопления в каталоге ACT имеют красное смещение в диапазоне $0.04 < z < 1.91$ (со средним значением $z = 0.52$ и 221 скопление на $z > 1$).

2.2.3. MCXC. MCXC представляет собой большой составной каталог рентгеновских скоплений галактик, обнаруженных по данным обсерватории ROSAT. Выборка включает скопления, найденные в ходе проведения обзора неба ROSAT All

Sky Survey (NORAS, REFLEX, BCS, SGP, NEP, MACS и CIZA) и программ точечных наведений спутников ROSAT (160SD, 400SD, SHARC, WARPS) и EINSTEIN (EMSS). В общей сложности каталог МСХС содержит 1743 рентгеновских скопления галактик с надежной оптической идентификацией.

2.2.4. eРОЗИТА. Каталог скоплений галактик телескопа СРГ/eРОЗИТА произведен в рабочей научной группе по каталогу рентгеновских источников российского консорциума eРОЗИТА по данным первого года сканирования неба. При построении каталога использовались данные восточной галактической полусферы $0 < l < 180$, за обработку которой отвечает российский консорциум телескопа СРГ/eРОЗИТА. Детали наблюдений, обработки данных, процедуры детектирования источников и верификации полученного каталога, так же как и сам каталог, не являются предметом данной работы и будут опубликованы в отдельной статье.

В данной работе использована только часть каталога во внегалактической области с галактической широтой $|b| > 20^\circ$.

3. МОДЕЛИ ДЕТЕКТИРОВАНИЯ СЗ-ОБЪЕКТОВ

В настоящее время в литературе используются несколько подходов к детектированию СЗ-источников: алгоритмы детектирования, основанные на классическом физическом моделировании (MMF), байесовском моделировании (PowellSnakes) и нейросетевых моделях. Появляются также гибридные подходы (см. Хариер и др., 2021), где нейронные сети используются для уточнения карты сегментации СЗ-источников, полученной классическими подходами.

Классический алгоритм MMF (Мелин и др., 2006) опирается на аппроксимацию сигнала на картах интенсивности микроволнового излучения при помощи низкопараметрических моделей для СЗ-сигнала и других спектральных и пространственных компонент излучения, присутствующих в данных, с последующим построением оптимальных фильтров для получения карт сегментации и детектированием на них СЗ-объектов выше определенного порога по значимости. В PowellSnakes (Карвальйо и др., 2009) используется методика байесовского моделирования. С помощью подходов MMF и PowellSnakes были получены каталоги СЗ-источников из официальных выпусков данных обсерваторий PLANCK и ACT, содержащие объекты с высоким уровнем значимости. Более слабые СЗ-источники (не попавшие в опубликованные каталоги) скопления PLANCK можно искать на

картах y -параметра комптонизации, которые также публиковались в рамках выпусков данных обсерватории PLANCK (Коллаборация Планк и др., 2016с). Для реконструкции y -карт применялись подходы, базирующиеся на ILC (Internal Linear Combination): MILCA (Modified Internal Linear Combination Algorithm, Хариер и др., 2013), NILC (Needlet Independent Linear Combination, Ремазеиллес и др., 2013). В работе (Буренин, 2017) был выполнен поиск СЗ-источников на реконструированных y -картах PLANCK (NILC) и получен расширенный каталог СЗ-источников вплоть до самых низких порогов по значимости. В настоящей работе мы публикуем обновленную версию расширенного каталога СЗ-источников PLANCK, полученную в рамках данного подхода (см. п. 3.3).

В последние годы, по мере бурного развития нейросетевых технологий и расширения доступных тренировочных выборок астрономических объектов, открылись принципиально новые возможности по созданию обучаемых нейросетевых алгоритмов сегментации и детектирования, работающих непосредственно с изображениями неба, полученными в рамках обзоров неба в выбранных спектральных диапазонах (в данной работе — с картами интенсивности по данным микроволновых наблюдений PLANCK). Нейросетевая модель сегментации выполняет тесную аппроксимацию сложной функции $X \rightarrow y$ для преобразования изображений размера $n \times n$ в шести каналах HFI ($X \subset \mathbb{R}^{6 \times n \times n}$) в карту сегментации СЗ-источников ($y \subset \mathbb{R}^{n \times n}$). Нейросетевые алгоритмы обучения с учителем способны автоматически распознавать имеющиеся закономерности в данных и заметно превосходить по точности стандартные подходы физического моделирования (часто имеющие те или иные априорные ограничения) при использовании достаточно большой и репрезентативной тренировочной выборки. В последнее время нейросетевые модели показали высокую эффективность при поиске СЗ-объектов (Лин и др., 2021), в том числе для данных PLANCK (Бонжеан, 2020). Однако каталоги кандидатов в скопления галактик, полученные нейросетевыми алгоритмами, до настоящего времени не публиковались. В данной работе мы представляем расширенный каталог СЗ-источников PLANCK, полученный нейросетью семейства U-Net (см. п. 3.4).

Ниже в п. 3.1 мы рассмотрим разбиение данных, а в п. 3.2 — метрики качества, используемые для обучения и тестирования моделей детектирования СЗ-источников на данных PLANCK. Будет описана (см. п. 3.3) базовая модель поиска СЗ-источников на картах комптоновского параметра и нейросетевые модели сегментации (см. п. 3.4). В последней части данной главы мы опишем создание расширенного каталога источников PLANCK,

полученного путем объединения прогнозов моделей.

3.1. Разбиение данных для обучения и тестирования методов поиска СЗ-источников

Все небо было разделено на несколько частей (см. рис. 1, правая нижняя панель), представляющих собой тренировочные и тестовые области. Соответственно, данные из тренировочной области использовались при создании обучающей выборки для нейросетевых моделей, восточные тестовые зоны использовались для выбора оптимальной эпохи на кривых обучения (см. ниже рис. 4) и независимого тестирования нейросетевых моделей сегментации, а также точности и полноты полученных каталогов СЗ-объектов. Тренировочные и тестовые области на небе были выбраны через представление данных HEALPix по схеме NESTED с параметром разбиения $N_{\text{side}} = 2$ (всего 48 областей). В качестве тестовых были выбраны пиксели 9, 41 в западной галактической полусфере и 6, 38 в восточной галактической полусфере (все остальные 44 области разбиения неба могли использоваться для обучения моделей). Отметим, что выбор тестовых областей совпадает с работой Бонжеан (2021), что кроме прочего облегчает сравнение наших моделей с указанной работой. В табл. 1 приведено количество объектов в разных каталогах галактик для различных тестовых областей. Немного подробнее о тестовых областях:

- Юго-западная тестовая область (индекс 41 в разбиении неба HEALPix с центром в направлении $b = 41.81^\circ$, $l = 247.5^\circ$) содержит (табл. 1) наибольшее число скоплений галактик из каталогов АСТ, МСХС, PSZ2(z) с практически равномерным их распределением в этой области. В данной области мы будем сравнивать полноту и точность различных каталогов СЗ-объектов относительно скоплений из объединения трех каталогов АСТ, МСХС, PSZ2(z).
- Восточные тестовые области (индексы 6 и 38) полностью покрываются имеющимися в нашем распоряжении наблюдениями СРГ/еРОЗИТА (принадлежащими российскому консорциуму). Северо-восточная тестовая область расположена вблизи северного полюса эклиптики, поэтому чувствительность рентгеновского обзора неба СРГ/еРОЗИТА (и соответственно число скоплений галактик в каталоге для этой зоны) заметно выше, чем для юго-восточной тестовой области (табл. 1). В данных областях мы будем сравнивать полноту и

точность различных каталогов СЗ-объектов относительно каталога рентгеновских скоплений СРГ/еРОЗИТА.

- Дополнительно для тестирования моделей поиска СЗ-источников в данных PLANCK мы будем использовать каталог скоплений СРГ/еРОЗИТА на всем восточном внегалактическом небе.

3.2. Метрики оценки качества моделей

В данном разделе мы рассмотрим метрики, которые будут использованы для обучения нейросетевых моделей (QE), метрики для тестирования качества полученных нейросетевых карт сегментации СЗ-объектов (IoU , $DICE$) и, наконец, метрики для оценки качества полученных каталогов скоплений ($Recall$, $Precision$).

Бинарная кросс-энтропия (QE) является стандартной метрикой для оценки качества модели машинного обучения в задачах классификации и семантической сегментации. В задаче сегментации на изображении с N элементами изображения (пикселями) она определяется следующим образом:

$$QE = -\frac{1}{N} \sum_{i=1}^N (y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)), \quad (1)$$

где y_i и \hat{y}_i — значение истинной (англ. ground true) карты сегментации (0,1) и вероятностный прогноз модели для i -го пикселя. Как видно из формулы (1), QE рассчитывается независимо для каждого пикселя изображения, а затем усредняется по всем элементам изображения и далее всем изображениям в выборке. Мы выбрали QE в качестве функции ошибок для оптимизации весов нейросетевых моделей.

Для оценки качества нейросетевой карты сегментации СЗ-источников мы использовали стандартные метрики IoU и $DICE$, которые широко применяются в подобных задачах компьютерного зрения. IoU — это площадь области перекрытия между прогнозируемой картой сегментации (\hat{S}) и истинной картой сегментацией (S) объектов, нормированная на площадь их объединения.

$$IoU = \frac{|\hat{S} \cap S|}{|\hat{S} \cup S|}. \quad (2)$$

Коэффициент $DICE$ равен удвоенной площади перекрытия прогнозируемой и истинной карт сегментации, деленной на сумму этих площадей:

$$DICE = \frac{2|\hat{S} \cap S|}{|\hat{S}| + |S|}. \quad (3)$$

В результате детектирования объектов на нейросетевых картах сегментации (или на картах параметра комптонизации в случае базовой модели) были получены каталоги СЗ-источников. Для оценки их качества мы воспользовались метриками полноты (*Recall*) и точности (*Precision*) относительно выбранных каталогов скоплений (табл. 1). Полнота рассматриваемого каталога СЗ-объектов определяется как

$$Recall = \frac{N_{true}}{N_{cat}}, \quad (4)$$

где N_{cat} — общее число объектов в выбранном для сравнения каталоге скоплений, N_{true} — число СЗ-объектов, у которых в заданном радиусе сопоставления (мы использовали $R_{match} = 400''$) имеется скопление. Точность считалась как:

$$Precision = \frac{N_{true}}{N_{det}}, \quad (5)$$

где N_{det} — общее число СЗ-объектов в рассматриваемом каталоге.

В последующих пунктах описываются модели поиска СЗ-источников.

3.3. Базовый метод поиска СЗ-источников на картах параметра комптонизации

Эффект Сюняева–Зельдовича (Сюняев, Зельдович, 1970, 1972) состоит в обратном комптоновском рассеянии фотонов реликтового излучения на горячих электронах в направлении на скопления галактик. Величина эффекта задается параметром комптонизации (y), который зависит от плотности (n_e) и температуры (T_e) тепловых электронов скопления на луче зрения:

$$y = \frac{\sigma_T}{m_e c^2} \int n_e(l) k_B T_e(l) dl, \quad (6)$$

где σ_T — сечение когерентного (томпсоновского) рассеяния, k_B — константа Больцмана, c — скорость света, m_e — масса электрона.

Для поиска источников сигнала Сюняева–Зельдовича нами были использованы карты параметра комптонизации, взятые из архива обсерватории им. Планка². Была использована версия этих карт, опубликованная в выпуске данных 2015 г. (Коллаборация Планк и др., 2016с). Поиск источников проводился по картам *NILC*, поскольку они имеют несколько меньше шума на малых угловых масштабах (см. подробнее в работе Сообщества Планка, 2016с). Из карты параметра комптонизации был вычтен сигнал на угловом масштабе больше $20'$. Это должно в некоторой степени увеличить чувствительность поиска

далеких скоплений, которые имеют характерный угловой размер меньше углового разрешения карт параметра комптонизации обзора обсерватории им. Планка, $10'$. Карта соотношения сигнал–шум была получена делением на карту стандартного отклонения, сглаженную круглым окном радиусом $20'$. Максимумы на этой карте отношения сигнал–шум рассматривались в качестве источников сигнала СЗ.

Полученный таким образом список содержит 6901 СЗ-источник на российской половине неба обзора СРГ/еРОЗИТА (со значимостью $>5\sigma$). Сравнение этого списка с каталогом скоплений обзора *400d* (Буренин и др. 2007), а также отождествление подобного списка объектов в оптическом и ИК-диапазонах (Буренин, 2017), показывают, что более половины из этих объектов являются ложными, однако, подавляющее большинство скоплений с массами выше $\approx 5 \times 10^{14} M_\odot$ должны содержаться в этом списке. Сравнение этого каталога с каталогом протяженных рентгеновских источников из обзора СРГ/еРОЗИТА обсуждается также в работе Буренина и др. (2022).

3.4. Нейросетевой метод поиска СЗ-объектов

Общая схема нейросетевой модели поиска СЗ-источников представлена на рис. 3. Закрашенными прямоугольниками показаны блоки модели, открытыми прямоугольниками изображены данные. Для получения каталога СЗ-объектов мы используем двухэтапный подход: (i) получение карты сегментации с помощью U-Net архитектуры и (ii) детектирование СЗ-источников на карте сегментации и создание каталога. В случае использования дообучения, этапы (i) и (ii) повторяются с обновленной тренировочной выборкой, которая формируется на основе скоплений из каталогов PSZ2, ACT, MCXC, успешно сопоставленных объектами из каталога СЗ-объектов, полученного моделью на предыдущем этапе. Ниже мы подробно рассмотрим построенные нами нейросетевые модели поиска СЗ-источников.

3.4.1. Базовая нейросетевая модель сегментации. Модели сегментации в статье основаны на классической архитектуре U-Net (Роннебергер и др. 2015), представляющей собой автокодировщик с дополнительными связями (англ. skip-connections) между симметричными слоями кодировщика и декодировщика. Архитектура U-Net позволяет формировать сжатое представление входных данных в латентном слое и затем с помощью декодировщика разворачивать его в маску сегментации. При этом использование дополнительных связей между симметричными слоями кодировщика и декодировщика помогает получать

²<http://pla.esac.esa.int/>

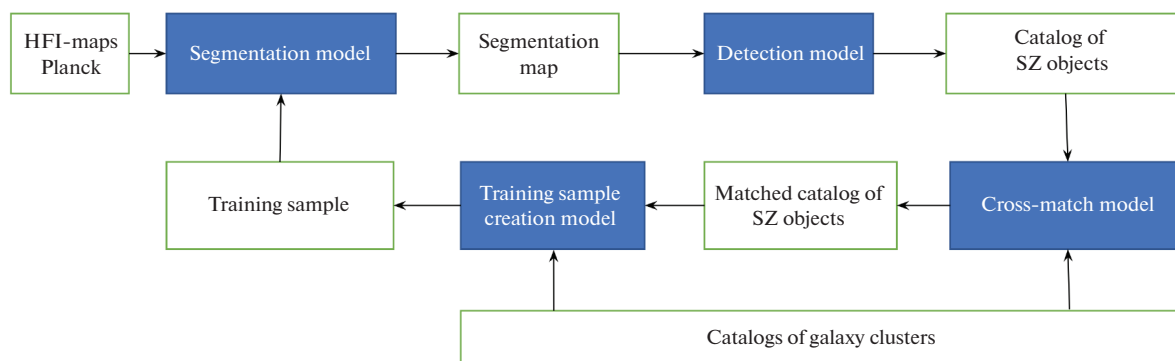


Рис. 3. Схема обучения нейросетевых моделей поиска СЗ-источников.

карту сегментации с разрешением, соответствующим исходным данным.

Отметим, что в работе (Бонжеан, 2020) ранее было продемонстрировано удачное применение архитектуры U-Net в задаче сегментации и поиска СЗ-источников PLANCK в данных HFI. Однако автор Бонжеан (2020) не выложил код построенной им нейросетевой модели (а также полученный им каталог СЗ-объектов) в открытый доступ, поэтому в настоящей работе мы построили свою U-Net архитектуру для сегментирования СЗ-источников в данных PLANCK HFI с нуля.

Перед попаданием в нейросетевую модель сегментации данные PLANCK были нормализованы (аналогично работе Бонжеан, 2020). Архитектура U-net для базовой нейросетевой сегментации СЗ-источников представляла собой:

- Кодировщик, состоящий из пяти блоков; в каждом блоке — по две свертки ядром 3×3 , функцией активации ReLU и слоем max-pooling. После каждого блока кодировщика количество фильтров возрастает в 2 раза (вплоть до 128 в латентном слое). Последний слой кодировщика дополняется сверткой 1×1 .
- Декодировщик, состоящий из пяти блоков; в каждом блоке — 2×2 свертка с увеличением разрешения (англ. upsampling), вектор признаков конкатенировался с вектором на выходе соответствующего симметричного блока кодировщика, и две свертки 3×3 с функцией активации ReLU. После каждого блока декодировщика количество фильтров уменьшается в 2 раза.

После каждой свертки (в качестве регуляризатора для увеличения эффективности обучения) был добавлен слой случайного выбивания нейронов (англ.

Dropout) с параметром вероятности равным $p = 0.2$. Перед выходом нейросети, в качестве последней функции активации использовалась сигмоида для того, чтобы преобразовать карту выходных признаков модели в карту вероятностей сегментации. После экспериментов с разными вариантами моделей мы остановились на нейросети семейства U-Net с асимметрией (несколько большим числом параметров) в крыле декодировщика. Общее число параметров выбранной нейросетевой архитектуры составило 2 161 489. Подробнее архитектуру нейросетевой модели можно посмотреть в репозитории статьи.

Для обучения нейросетей мы использовали тренировочные изображения (англ. patches) размером $1.83^\circ \times 1.83^\circ$ (для шести HFI-каналов размерность матрицы изображения составила $64 \times 64 \times 6$), нарезанные из данных PLANCK. Для каждого тренировочного изображения была создана “целевая” маска сегментации скоплений, в которой все пиксели в радиусе $2.5'$ от центров скопленных тренировочного каталога положили равными единице, все остальные пиксели маски были приравнены нулю. Области для обучения нейросети (100 тыс. участков изображений PLANCK размера 64×64) выбирались случайно по небесной сфере с условием, чтобы в каждое изображение попало хотя бы одно скопление из тренировочного каталога. Все архитектуры нейросетей в данной работе были реализованы с помощью библиотеки tensorflow.keras³. Веса нейросетевых моделей обучались методом стохастического градиентного спуска с оптимизатором Adam (с параметром $lr = 10^{-4}$) вплоть до 50 эпох. Число изображений в пакете (англ. batch size) при обучении нейросети было выбрано равным 20. В качестве функции ошибок в задаче семантической сегментации использовалась бинарная кросс-энтропия QE (см.

³https://www.tensorflow.org/api_docs/python/tf/keras

формулу (1)). По результатам каждой эпохи обучения вычислялись метрики качества сегментации IoU , $DICE$ на валидационной выборке, состоящей из случайно выбранных 12 000 тестовых изображений 64×64 из тестовых областей (6, 38, 41 элементы разбиения неба, см. п. 3.1). При обучении нейросетей использовалась стандартная техника аугментации входных изображений, состоящая из случайных поворотов картинки (и маски) на угол, кратный $\pi/2$.

Базовая модель нейросетевой сегментации обучалась на выборке скоплений PSZ2(z) из каталога PSZ2, для которых имеются спектроскопические измерения красных смещений (см. табл. 1), полностью аналогично нейросетевой модели в работе (Бонжеан, 2020). На рис. 4 представлена зависимость функции ошибок (QE) и метрик сегментации (IoU , $Dice$) от числа эпох обучения на тренировочной и валидационной выборках для базовой нейросетевой модели сегментации. Как можно видеть на рисунке, на первых эпохах обучения функция ошибок (QE) убывает как на тренировочной, так и на валидационной выборке. Затем QE на валидационной выборке достигает минимума и далее следует ее постепенный рост. В то же время QE на тренировочной выборке монотонно убывает с увеличением количества эпох обучения. Такое поведение кривых обучения (рост ошибки модели на валидационной выборке с одновременным убыванием ошибки модели на тренировочных данных) соответствует режиму переобучения нейросетевого алгоритма, при котором ошибка обобщения моделей будет расти. Чтобы избежать подобной ситуации, мы останавливаем обучение моделей на эпохе начала роста ошибки QE на валидационной выборке (см. вертикальные линии на рис. 4).

Обученная нейросетевая модель позволяет построить карту сегментации СЗ-источников на участке изображения PLANCK размером 64×64 (из HEALPix-разбиения неба с $n_{\text{side}} = 8$). Для того чтобы получить карту сегментации по всему небу, нам необходимо было “просканировать” окном сегментации 64×64 все области, на которые было разделено небо по схеме HEALPix. По каждой из 48 областей размера 1024×1024 мы перемещали “окно сегментации” с заданным шагом и затем усредняли прогнозы нейросети для каждого пикселя изображения. Мы попробовали сканировать данные PLANCK с различным выбором шага (от 8 до 64). При уменьшении размера шага сканирования с 64 до 8 пикселей качество карты сегментации улучшалось за счет усреднения по окну сканирования. Таким образом, для построения нейросетевой карты сегментации мы выбрали шаг сканирования 8 пикселей, при котором через каждый пиксель карты PLANCK

(если он не лежит рядом с границей изображения) окно сканирования проходит $(64/8)^2 = 64$ раза.

Как упоминалось ранее, базовая нейросетевая модель сегментации (обозначим ее SZcat(base) близка к архитектуре нейросетевой модели сегментации из работы Бонжеан (2020). Мы построили каталог СЗ-источников (выделяя области выше порога вероятности $p_i > 0.1$ на картах сегментации) и задались вопросом, насколько результирующий каталог в работе Бонжеан (2020) близок к выборке SZcat(base). К сожалению, Бонжеан (2020) не опубликовал полученные им каталоги СЗ-объектов, поэтому мы имеем возможность только провести сравнение с метриками, приведенными автором данной работы. Так, автор статьи приводит значение метрики Recall на всем небе относительно данных каталогов Planck_z (наш PSZ2(z)), Planck_noz (PSZ2 исключая объекты PSZ2(z)) и MCXCwp (MCXC исключая объекты PSZ2). Радиус сопоставления СЗ-источника и скопления в работе Бонжеан (2020) был равным $R_{\text{match}} = 5'$. Сравнение результатов нашего каталога СЗ-объектов (также для $R_{\text{match}} = 5'$) с результатами статьи Бонжеан (2020) приведено в табл. 2. Как можно видеть, метрики для модели SZcat(base) хорошо согласуются с результатами Бонжеан (2020). Таким образом, мы можем сделать вывод, что в целом SZcat(base) достаточно близко воспроизводит модель сегментации СЗ-объектов из указанной работы.

На следующем этапе мы построили более точные модели сегментации СЗ-объектов, с применением техники дообучения на каталогах скоплений галактик АСТ и MCXC.

3.4.2. Дообучение модели сегментации на каталогах скоплений PSZ2, АСТ и MCXC. Идея дообучения нейросетевой модели сегментации СЗ-источников PLANCK на каталогах скоплений галактик, полученных другими микроволновыми и рентгеновскими обсерваториями, довольно проста.

Таблица 2. Полнота найденных скоплений галактик (Recall) по каталогам скоплений на всем небе (PSZ2, MCXC) в сравнении с результатами из работы (Бонжеан, 2020)

Выборки	PSZ2(z)	PSZ2*	MCXC*
Бонжеан (2020)	0.98	0.76	0.21
SZcat(base)	0.98	0.80	0.19

Примечание. PSZ2*, MCXC* — подвыборки PSZ2 и MCXC соответственно после удаления объектов PSZ2(z). Радиус сопоставления СЗ-объекта и скопления здесь был выбран (как и в работе Бонжеан, 2020) равным $R_{\text{match}} = 5'$.

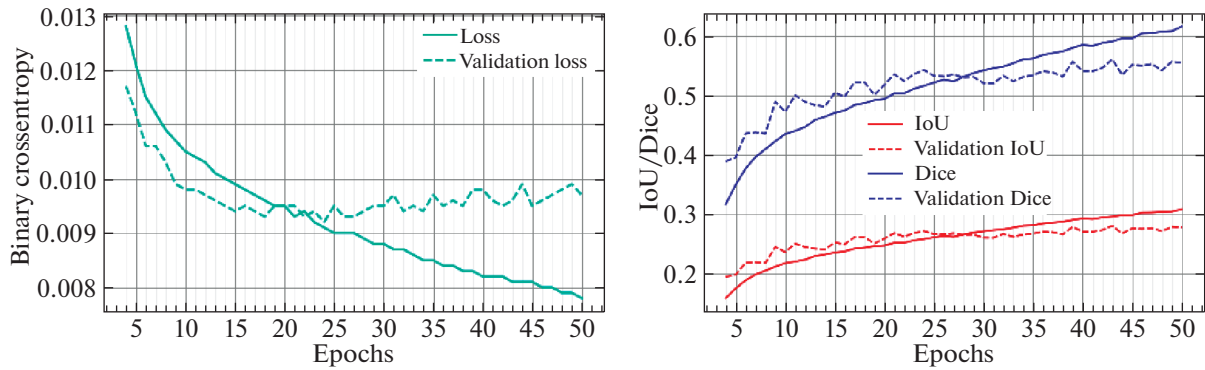


Рис. 4. Зависимость функции ошибок (бинарной кросс-энтропии QE), а также метрик сегментации (IoU , $DICE$) от эпохи обучения на валидационных и тренировочных данных для лучшей нейросетевой модели.

Выборка PSZ2(z) скоплений PSZ2, имеющих оптическую спектроскопию, на которой была обучена базовая нейросетевая модель сегментации (аналогичная работе Бонжеан, 2020), содержит только СЗ-объекты с высоким уровнем значимости детектирования в данных PLANCK. Менее массивные или более далекие скопления галактик (тем не менее видимые в данных PLANCK на более низком уровне значимости) могут находиться в каталогах скоплений галактик, полученных по данным других обсерваторий.

Основной проблемой для обучения нейросетевой модели на всех имеющихся каталогах скоплений является вероятное наличие в них (в частности, в каталоге скоплений АСТ) большого числа скоплений, которые в данных PLANCK находятся глубоко под уровнем шума. Включение их в обучение снизит качество нейросетевой модели. Чтобы решить данную проблему, мы использовали технику дообучения, в рамках которой в процесс тренировки моделей включаются не все скопления из каталогов, а только объекты, которые нейросетевая модель обнаружила на предыдущих этапах детектирования (рис. 3).

Обучив базовую нейросетевую модель на данных PSZ2(z) и построив каталог детектирования, мы сопоставили его с каталогами скоплений PSZ2, MCXC, АСТ. Сопоставление каталогов проводилось в радиусе $R = 400''$, внутрь которого попадает большая часть верных сопоставлений при минимальном уровне случайных объектов. Все СЗ-детектирования, которые оказались соотносены со скоплениями галактик в каталогах, были добавлены в обучающую выборку, и далее на ней было проведено дообучение нейросетевой модели (стартуя с весов базовой модели). Области для дообучения 64×64 выбирались случайно по небесной сфере с условием, чтобы в каждое такое изображение обязательно попадало хотя бы одно скопление каталога АСТ из новой обучающей выборки.

Мы построили каталог СЗ-источников путем выделения на карте областей с вероятностью сегментации выше порога ($p_i > 0.1$). Центром источника считался барицентр выбранной области. Для каждого источника также определялся параметр p_{\max} (наибольшее значение параметра сегментации в выбранной области).

3.5. Объединенный каталог СЗ-объектов

Для каждого СЗ-объекта в каталогах детектирования определялся параметр значимости источника. Для базового каталога, полученного по картам параметра комптонизации, значимость источника имела физический смысл — максимальное отношение сигнал/шум в области источника на картах y -параметра. Для нейросетевого каталога параметр значимости p_{\max} определяется как максимальное значение вероятности сегментации внутри области источника. Отметим, что параметр значимости нейросетевой модели детектирования не имеет четкого физического обоснования в терминах сигнал/шум. Фактически отбор объектов нейросетью ведется согласно некоторому внутреннему критерию (который определяется архитектурой нейросетевой модели и доступной обучающей выборкой) и требует калибровки относительно стандартных величин статистической значимости источника. Данное свойство присуще нейросетевым методам детектирования в целом (см., например, обсуждение нейросетевых моделей детектирования радиисточников SKA SDC2, Хартли и др., 2022).

На рис. 5 приведена зависимость метрики точности детектирования от порога по значимости СЗ-источника в нейросетевых каталогах SZcat(base), SZcat(AL) (график слева), в каталоге на основе карт параметра комптонизации SZcat(y -map) (посередине). Черными линиями показана точность детектирования, рассчитанная на объединенной выборке скоплений галактик из каталогов

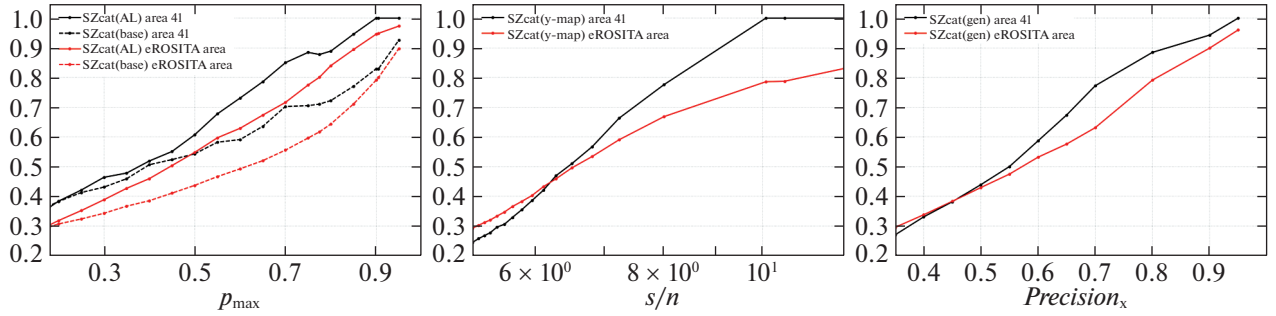


Рис. 5. Зависимость метрики точности детектирования ($Precision$) от порога по значимости СЗ-источника в нейросетевых каталогах SZcat(AL), SZcat(AL) (слева), в каталоге на основе карт параметра комптонизации SZcat(y-map) (посередине) и в объединенном каталоге SZcat(AL) (справа). Черными линиями показана точность детектирования, рассчитанная на объединенной выборке скоплений галактик из каталогов PSZ2(z), MCXC, АСТ в юго-западной тестовой области. Красные линии — оценка точности детектирования на тестовой выборке скоплений из каталога протяженных источников eРОЗИТА (1 год) в восточной внегалактической области ($b > 20^\circ$, $0 < l < 180^\circ$).

PSZ2(z), MCXC, АСТ в юго-западной тестовой области. Красные линии ($Precision_X$) — оценка точности детектирования на тестовой выборке скоплений из каталога скоплений eРОЗИТА в восточной внегалактической области ($b > 20^\circ$, $0 < l < 180^\circ$). На основе последней зависимости, для каждого СЗ-детектирования из каталогов SZcat(AL) и SZcat(y-map), мы определили универсальный параметр достоверности СЗ-источника следующим образом:

$$\begin{cases} Precision_X(p_{\max}), & obj \in SZcat(AL), \\ Precision_X(S/N), & obj \in SZcat(y-map). \end{cases} \quad (7)$$

Имея определенный параметр достоверности $Precision_X$ для всех СЗ-детектирований, мы объединили каталоги, полученные нейросетью и с использованием карт параметра комптонизации в единый каталог SZcat(gen). На рис. 5 (слева) приведена зависимость метрики точности детектирования от порога по достоверности СЗ-источников в объединенном каталоге.

Подробнее о публикуемом объединенном каталоге СЗ-детектирований PLANCK (см. Приложение).

4. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

С помощью моделей сегментации и детекции объектов, описанных в предыдущих разделах, нами были созданы карты сегментации и каталоги СЗ-объектов по данным PLANCK. Каталоги SZcat(base) и SZcat(AL) были получены на основе карт сегментации, произведенных алгоритмами глубокого обучения на базе архитектуры U-Net. При этом модель сегментации SZcat(base) близко (по архитектуре модели, разбиению неба на тренировочные/тестовые области, схеме аугментации и пр.) воспроизводит модель сегментации

СЗ-источников из работы Бонжеан (2020) и была обучена на выборке PSZ2(z) объектов каталога PSZ2 со спектральными красными смещениями. Более точная модель сегментации SZcat(AL) отличается от варианта SZcat(base) использованием процедуры дообучения на каталогах скоплений АСТ, MCXC, PSZ2. Наконец, каталог SZcat(y-map) был получен без использования глубокого обучения путем поиска объектов на картах параметра комптонизации (см. п. 3.3). Объединенный каталог SZcat(gen) содержит СЗ-детектирования из лучшего нейросетевого SZcat(AL) и базового SZcat(y-map) каталогов.

Ниже представлены основные результаты работы. В п. 4.1 обсуждается нейросетевая карта сегментации СЗ-источников. Далее, в п. 4.2 проанализированы построенные нами каталоги СЗ-объектов.

4.1. Нейросетевая карта сегментации СЗ-сигнала

На рис. 6 для случайно отобранного скопления из каталога PSZ2 показан пример нейросетевой карты сегментации СЗ-сигнала, полученной моделью SZcat(AL) (нижняя левая панель), в сравнении с картой параметра комптонизации (нижний средний рисунок) той же области и исходными картами интенсивности в каналах PLANCK HFI (верхние и нижние панели). Поле центрировано на скопление PSZ2 G104.98-54.79 ($\alpha = 2.558396^\circ$, $\delta = 6.633860^\circ$, центр скопления отмечен крестиком) на красном смещении $z = 0.2912$ с массой $M_{500} = 4.80 \times 10^{14} M_\odot$. Достаточно близко к выбранному объекту (внизу на границе изображения) расположено еще одно скопление галактик PSZ2 G104.71-54.54 на красном смещении $z = 0.2441$ с массой $M_{500} = 4.90 \times 10^{14} M_\odot$ (также отмечено крестом на рисунках). На нижней правой панели показана маска сегментации в этой области.

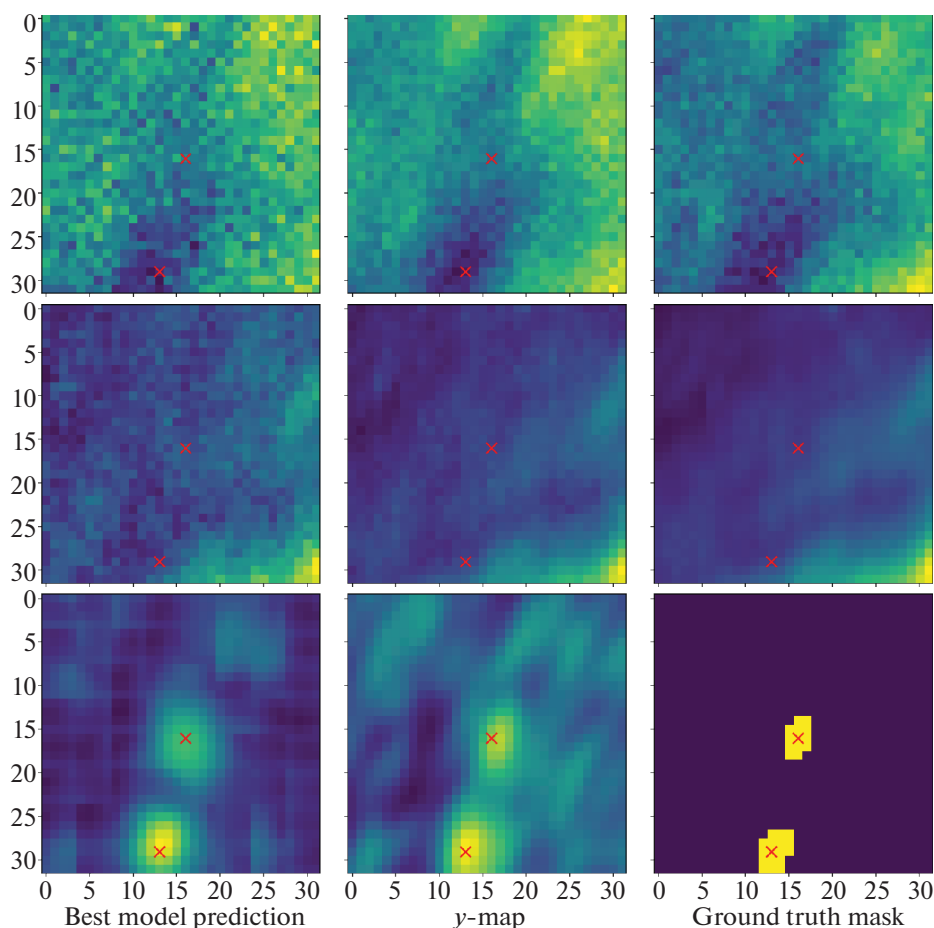


Рис. 6. Пример данных PLANCK — нормированных карт интенсивности в шести HFI-каналах (верхняя и средняя панели) и соответствующая карта сегментации нейросетевой модели (слева внизу), карта параметра комптонизации (посередине внизу) и эталонная маска сегментации в данной области (правый нижний рисунок). В выбранной области находятся два скопления каталога PSZ2 (отмечены крестиками): PSZ2 G104.98-54.79 (на красном смещении $z = 0.2912$ с массой $M_{500} = 4.80 \times 10^{14} M_{\odot}$) по центру картинки и PSZ2 G104.71-54.54 ($z = 0.2441$, $M_{500} = 4.90 \times 10^{14} M_{\odot}$) в нижней части картинки.

Сравнивая нейросетевую карту сегментации SZcat(AL) и карту параметра комптонизации, можно видеть на первой заметное подавление шума в окрестности скоплений, что, в свою очередь, позволит найти меньше ложных СЗ-объектов в нейросетевом каталоге источников, полученном на ее основе. С другой стороны, морфология объектов, по-видимому, лучше видна на y -картах. На карте SZcat(AL) также заметен остаточный эффект от выбранной сетки сканирования при построении карты сегментации на всем небе (см. подпункт 3.4.1).

В дальнейшем мы планируем провести более детальное исследование морфологии скоплений на картах нейросетевой сегментации и выяснить, насколько она может отражать реальную картину диффузного СЗ-излучения и темного вещества в скоплениях (см. Бонжеан, 2020, рис. 8 и обсужде-

ние к нему). В данной работе мы сконцентрируемся на анализе каталогов СЗ-источников.

4.2. Сравнение каталогов СЗ-источников

На рис. 7 показано распределение на небе источников из каталогов СЗ-детектирования, полученных нейросетевой моделью SZcat(AL) (слева) и базовой моделью на картах параметра комптонизации SZcat(y -map) (справа), для внегалактической области $|b| > 20^\circ$.

В табл. 3 представлено сравнение полученных каталогов СЗ-источников по метрикам точности (*Precision*) и полноты (*Recall*) детектирования объектов (см. п. 3.2). Сравнение приводится для юго-восточной тестовой области и для восточной внегалактической области ($|b| > 20^\circ$, $0^\circ < l < 180^\circ$). Метрики были рассчитаны относительно нескольких “эталонных” каталогов

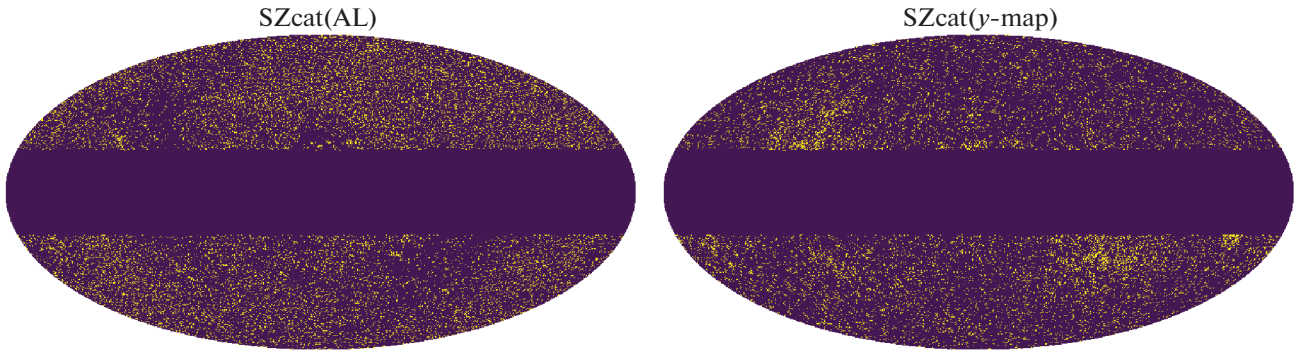


Рис. 7. Распределение объектов в расширенных каталогах СЗ-источников, построенных в данной работе.

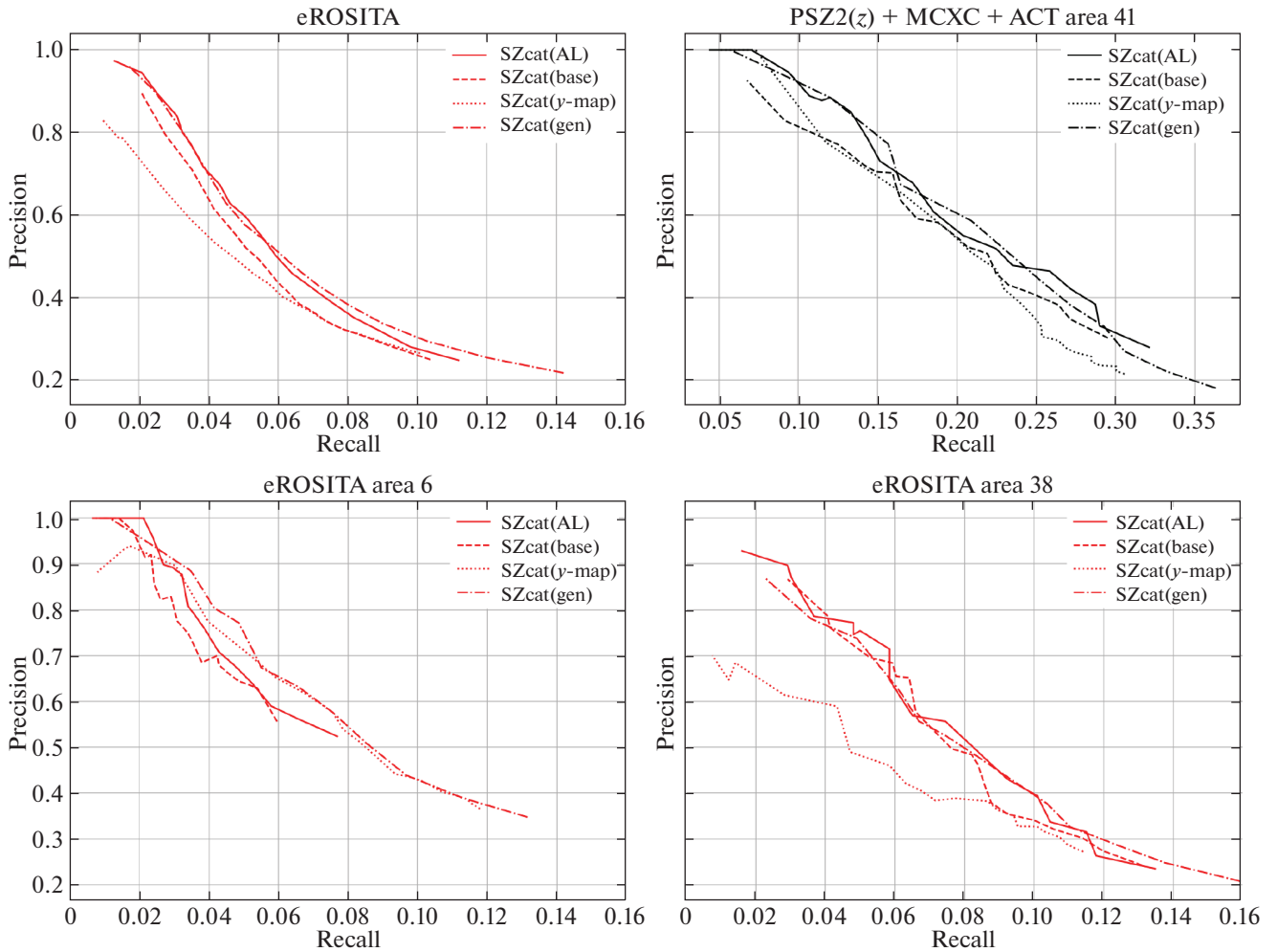


Рис. 8. Зависимости точности от полноты каталогов СЗ-источников для юго-западной (слева сверху) и восточной внегалактической $|b| > 20^\circ$, $0^\circ < l < 180^\circ$ (справа сверху), северо-восточной (слева внизу) и юго-восточной (справа внизу) тестовых областей. Северо-восточная тестовая область расположена вблизи северного полюса эклиптики, где чувствительность СРГ/еРОЗИТА и соответственно плотность рентгеновских скоплений максимальна.

скопления галактик: мы использовали каталоги PSZ2(z), MCXC, АСТ (по отдельности и в комбинации) и каталог протяженных рентгеновских источников еРОЗИТА 1 года (в восточной галакти-

ческой полусфере). Для соотнесения СЗ-объекта и скопления использовался ранее выбранный нами радиус сопоставления $R_{\text{match}} = 400''$. Для сравнения в таблице приведены также метрики

Precision и *Recall* для опубликованного второго каталога СЗ-источников миссии PLANCK–PSZ2.

Как можно видеть из табл. 3, 30–35% скоплений АСТ, $\approx 50\%$ рентгеновских скоплений ROSAT и $\approx 14\%$ рентгеновских скоплений ePOZITA (1 год) обнаруживаются в новых каталогах СЗ-детектирования PLANCK. Расширенные каталоги СЗ-детектирования PLANCK содержат несколько раз больше скоплений галактик, чем каталог PSZ2. В то же время новые СЗ-каталоги имеют значительную долю ложных объектов (более 70%). Долю ложных источников среди СЗ-детектирования можно значительно уменьшить, если отфильтровать объекты каталогов по параметру надежности детектирования.

На рис. 8 (верхняя панель графиков) приведены зависимости точности от полноты детектирования СЗ-источников — *Precision(Recall)*. Каждая точка на таком графике соответствует значению метрик, рассчитанных на выборке СЗ-детектирования, которые были отобраны выше некоторого порога по параметру надежности детектирования объекта. Изменяя величину порога от минимального до максимального значения надежности детектирования, мы получаем кривые *Precision(Recall)*, которые характеризуют построенные нами СЗ-каталоги для различных возможных режимов их использования (как в режиме высокой точности, так и в режиме высокой полноты отбираемых объектов). Чем выше лежит кривая каталога на графике *Precision(Recall)*, тем лучше соответствующий каталог СЗ-детектирования относительно выбранного класса объектов (в нашем случае — скоплений галактик).

На рис. 8 (нижняя панель графиков) приведены зависимости *Precision(Recall)* отдельно для северо-восточной (слева) и юго-восточной (справа) тестовых областей. В качестве эталонного каталога скоплений здесь использовались данные 1 года СРГ/ePOZITA. Северо-восточная тестовая область расположена вблизи северного полюса эклиптики, где чувствительность СРГ/ePOZITA и соответственно плотность рентгеновских скоплений значительно выше, чем в экваториальной эклиптической плоскости (например, в районе юго-восточной тестовой области).

Одной из проблем при использовании эмпирических алгоритмов глубокого обучения для сегментации СЗ-объектов стала неравномерность чувствительности на радио-картах по небесной сфере (увеличение чувствительности в области полюсов эклиптики), вызванная геометрией сканирования неба спутником PLANCK. Как можно видеть на нижней левой панели рис. 8, в северо-восточной тестовой области (примыкающей к северному полюсу эклиптики), в силу недостатка данных о

скоплениях, нейросетевой каталог СЗ-объектов оказался менее чувствительным, чем полученный стандартными методами по y -картам. В будущем мы планируем адаптировать наш алгоритм сегментации для эффективного поиска объектов на радиокартах разной чувствительности. Также представляется перспективным использование рентгеновского каталога СРГ/ePOZITA непосредственно для обучения модели (вместе или вместо каталога скоплений АСТ). В настоящей работе для получения максимально полного каталога СЗ-объектов по всему небу мы объединили каталоги SZcat(AL) и SZcat(y -map), получив каталог SZcat(gen) (см. п. 3.5). Как можно видеть из сравнения кривых *Precision(Recall)* на рис. 8, объединенный каталог SZcat(gen) является наиболее точным и полным по данным во всех тестовых областях.

На рис. 9 показано распределение по красному смещению z (левая панель) и массе M_{500} скопления (правая панель) для СЗ-источников объединенного каталога SZcat(gen), ассоциированных (в радиусе $R_{\text{match}} = 400''$) со скоплениями галактик по объединенным данным каталогов PSZ2(z), MCXC, АСТ. Для сравнения также приведено распределение объектов из выборки PSZ2(z). Из рисунка можно видеть, как дополнительные скопления галактик, содержащиеся в SZcat(gen), распределены по красному смещению и массе.

Разделив число объектов SZcat(gen), сопоставленных со скоплениями галактик (рис. 9), на полное число объектов в соответствующих каталогах скоплений PSZ2(z) + MCXC + АСТ (рис. 2), мы можем оценить функцию отбора скоплений галактик в каталоге SZcat(gen) по красному смещению и массе. На рис. 10 приведена функция отбора каталогов SZcat(AL), SZcat(gen) по красному смещению z (левая панель) и массе M_{500} (правая панель) относительно компиляции трех выборок скоплений галактик (PSZ2(z) + MCXC + АСТ) во внегалактической области $|b| > 20^\circ$. Сплошные и штриховые линии на графиках соответствуют каталогам SZcat(AL) и SZcat(gen). Для сравнения отдельной (штрихпунктирной) кривой показана зависимость функции отбора каталога PSZ2 от z и M_{500} .

Как видно из рис. 10, полнота отбора скоплений в представленных каталогах СЗ-детектирования убывает с ростом красного смещения, достигая $\approx 40\%$ на $z = 0.4$ и $\approx 20\%$ на $z = 0.6$. Полнота отбора скоплений также падает с уменьшением массы, достигая $\approx 40\%$ на массе $M_{500} = 3 \times 10^{14} M_\odot$. Отметим, что поскольку компилятивный каталог PSZ2(z) + MCXC + АСТ не является полной выборкой скоплений по массе и красному смещению, то оценки функции отбора SZcat(gen) на его

Таблица 3. Полнота и точность детектирования скоплений для каталогов СЗ-объектов на юго-западном тестовом пикселе (область 41) и в восточной внегалактической области ($|b| > 20^\circ$, $0^\circ < l < 180^\circ$) в сравнении с каталогами скоплений PSZ2(z), MCXC, ACT, а также в сравнении с каталогом рентгеновских скоплений из годового обзора неба eРОЗИТА. Значение метрики Precision приведено относительно каталога рентгеновских скоплений СРГ/eРОЗИТА

SZ-catalog	Recall	Recall	Recall	Recall	Precision
Юго-западная тестовая область					
	PSZ2(z)	MCXC	ACT	PSZ2(z) + MCXC + ACT	
SZ-catalog	Recall	Recall	Recall	Recall	Precision
SZcat(y-map)	0.962	0.500	0.284	0.305	0.219
SZcat(AL)	0.943	0.515	0.293	0.321	0.279
SZcat(gen)	0.943	0.529	0.345	0.363	0.181
PSZ2	1	0.382	0.113	0.144	1
Восточная внегалактическая область $ b > 20^\circ$, $0^\circ < l < 180^\circ$					
	PSZ2(z)	MCXC	ACT	eРОЗИТА: 1 год	
SZcat(y-map)	0.965	0.427	0.220	0.101	0.266
SZcat(AL)	0.988	0.421	0.274	0.112	0.248
SZcat(gen)	0.994	0.472	0.302	0.142	0.217
PSZ2	1	0.301	0.102	0.033	0.812

основе, приведенные выше, следует рассматривать как верхние пределы.

Наконец, используя доступный нам каталог eРОЗИТА, мы аналогичным образом нашли, что полнота отбора скоплений в SZcat(gen) падает с уменьшением рентгеновского потока скопления, достигая $\approx 40\%$ на $F_{X,0.5-2} = 8 \times 10^{-13}$ эрг/с/см².

5. ВЫВОДЫ

Результатом настоящей работы стал первый в своем роде расширенный каталог объектов Сюняева–Зельдовича, построенный по высоко-частотным (HFI) данным миссии PLANCK с помощью техники глубокого обучения (модели семейства U-Net), дополненный источниками, детектируемыми на картах параметра комптонизации. Каталог также включает в себя источники с низким уровнем значимости и оценку достоверности детектирования.

Для полученного каталога СЗ-объектов мы проанализировали функцию отбора скоплений

галактик от массы (M_{500}), красного смещения (z) и рентгеновского потока ($F_{X,0.5-2}$).

Мы видим два сценария использования полученных нами расширенных каталогов СЗ-источников по данным PLANCK: (а) весь каталог СЗ-объектов имеет большое число ложных источников, но при этом содержит практически все возможные детектирования скоплений в данных PLANCK (он может использоваться совместно с другими каталогами кандидатов в скопления галактик, полученных независимо на других данных) и (б) выборка только надежных СЗ-объектов из каталога кандидатов может использоваться непосредственно, например, для поиска в ней уникальных объектов или исследования популяции скоплений.

Настоящее исследование основано на наблюдениях телескопа eРозита на борту обсерватории СРГ. Обсерватория СРГ изготовлена Роскосмосом в интересах Российской академии наук в лице Института космических исследований (ИКИ) в рамках Российской федеральной научной программы с участием Германского центра авиации

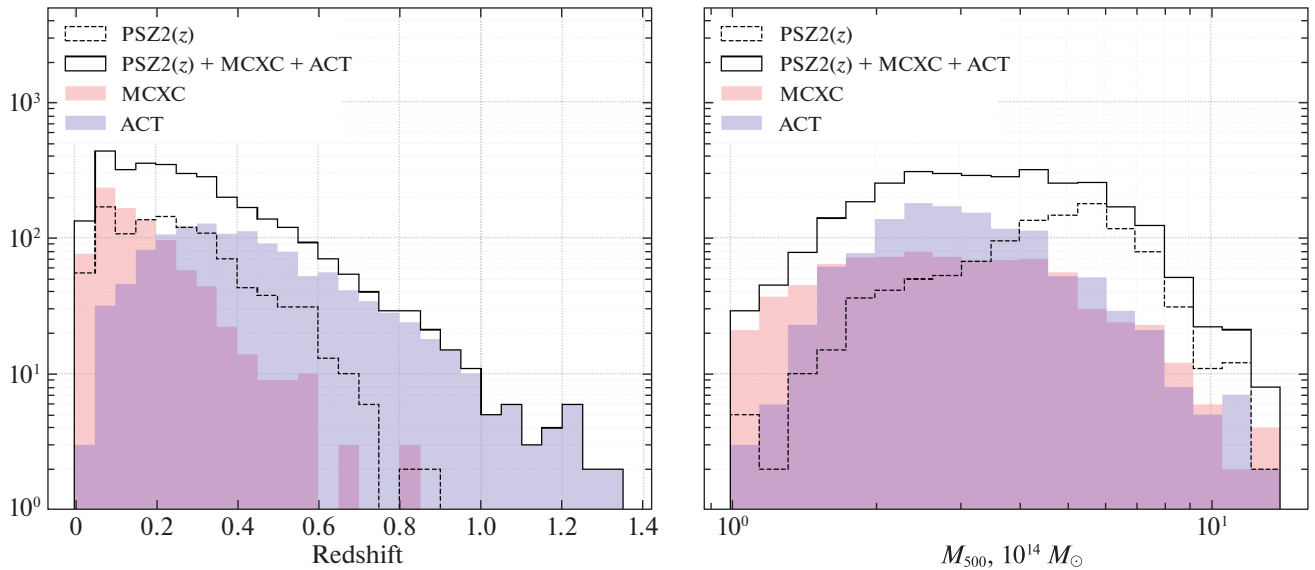


Рис. 9. Распределение числа скопления по красному смещению z (панель слева) и массе M_{500} (панель справа) для СЗ-источников объединенного каталога SZcat(gen) ассоциированных со скоплениями галактик, по данным PSZ2(z), MCXC, ACT и компилятивного каталога PSZ2(z) + MCXC + ACT.

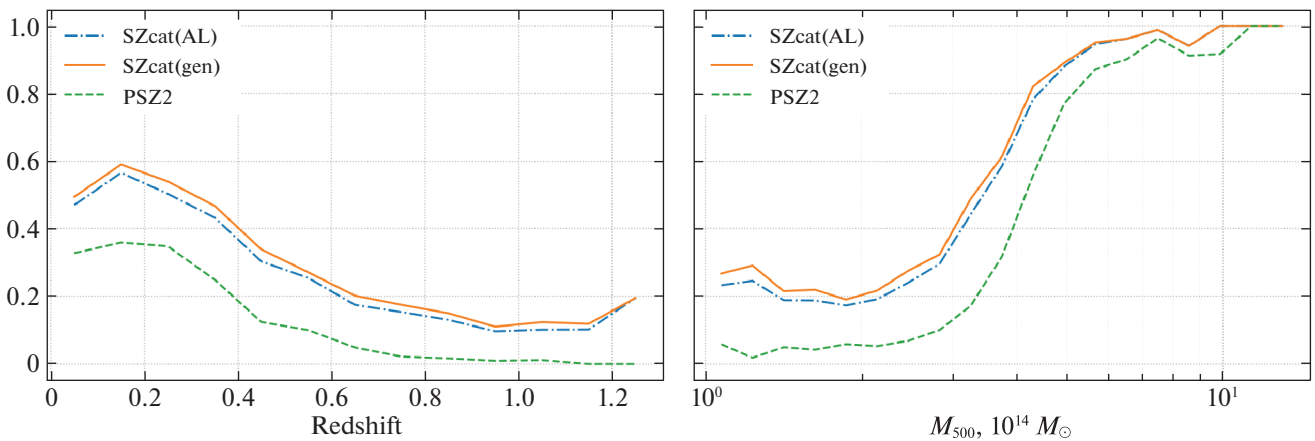


Рис. 10. Функция отбора каталогов SZcat(AL), SZcat(gen) и PSZ2 (для сравнения) по красному смещению z (левая панель) и массе M_{500} (правая панель) относительно компиляции трех выборок скопления галактик (PSZ2(z) + MCXC + ACT) во внегалактической области $|b| > 20^\circ$. Сплошные и штриховые линии на графиках соответствуют каталогам SZcat(AL) и SZcat(gen).

и космонавтики (DLR). Рентгеновский телескоп SRG/eРозита изготовлен консорциумом германских институтов во главе с Институтом внеземной астрофизики Общества им. Макса Планка (МРЕ) при поддержке DLR. Космический аппарат SRG спроектирован, изготовлен, запущен и управляется НПО им. Лавочкина и его субподрядчиками. Прием научных данных осуществляется комплексом антенн дальней космической связи в Медвежьих озерах, Усурийске и Байконуре и финансируется Роскосмосом. Используемые в настоящей работе данные телескопа eРозита обработаны с по-

мощью программного обеспечения eSASS, разработанного германским консорциумом eРОЗИТА и программного обеспечения для обработки и анализа данных, разработанного российским консорциумом телескопа eРОЗИТА.

При анализ данных и построении нейросетевых моделей в данной работе использовались публично доступные пакеты программного обеспечения с открытым исходным кодом (open source): python3, NumPy, Pandas, Tensorflow, HealPy, AstroPy, Matplotlib, Scikit-Image.

Таблица 4. Первые несколько строчек из расширенного каталога СЗ-детектирований PLANCK

id	RA	DEC	p_{\max}	S/N	$Precision_X$
SZ_G203.64-20.20	80.739063	-1.298711	0.100002	9.563881	0.312993
SZ_G160.26+48.65	149.798243	54.081141	0.100004	-1.000000	0.312993
SZ_G224.47-43.95	64.810863	-26.402030	0.100010	-1.000000	0.312993
SZ_G123.16-22.95	13.129393	39.918397	0.100023	-1.000000	0.312995
SZ_G132.06+17.45	55.778215	77.164844	0.100039	-1.000000	0.312996
...

ПРИЛОЖЕНИЕ

А. РАСШИРЕННЫЙ НЕЙРОСЕТЕВОЙ КАТАЛОГ СЗ-ИСТОЧНИКОВ PLANCK

Расширенный каталог СЗ-детектирований PLANCK объединяет в себе источники, найденные на высоких галактических широтах $|b| > 20^\circ$ нейросетевым алгоритмом (SZcat(AL)) на картах интенсивностей в каналах HFI, и источники на картах параметра комптонизации (SZcat(y -map)). Для каждого СЗ-объекта в каталоге мы записываем следующие характеристики:

- id — идентификатор источника,
- RA — прямое восхождение,
- DEC — склонение,
- p_{\max} — максимальное значение вероятности сегментации по области объекта для источников SZcat(AL),
- S/N — максимальное значение величины сигнал/шум в области источника на карте y -параметра для объектов SZcat(y -map),
- $Precision_X$ — единая величина достоверности детектирования СЗ-объекта, определяемая через точность детектирования скопленных галактик в рентгеновском каталоге протяженных источников СРГ/еРОЗИТА (см. п. 3.5).

В табл. 4 приведены первые несколько строчек из каталога. Общее количество объектов в каталоге — 30 917, из них 23 352 найдены с помощью нейросетевой модели (SZcat(AL)), а 13 777 обнаружены на картах параметра комптонизации (SZcat(y -map)). Количество общих объектов, найденных

обоими подходами, составляет 6212. Полную версию каталога можно найти в github-репозитории по ссылке https://github.com/astromining/planck_szcat/blob/main/SZcatgen.csv. Следуя лучшим практикам воспроизводимости методов глубокого обучения, мы также выкладываем в репозитории исходный код для построенных нами моделей.

СПИСОК ЛИТЕРАТУРЫ

1. Агуадо-Барохона и др. (A. Aguado-Barahona, R. Barrena, A. Streblyanska, A. Ferragamo, J.A. Rubiño-Martín, D. Tramonter, et al.), *Astron. Astrophys.* **631**, A148 (2019).
2. Блим и др. (L.E. Bleem, S. Bocquet, B. Stalder, M.D. Gladders, P.A. R. Ade, S.W. Allen, et al.), *Astrophys. J. Suppl. Ser.* **247**, 25 (2020).
3. Бонд и др. (J.R. Bond, L. Kofman, and D. Pogosyan), *Nature* **380**, 603 (1996).
4. Бонжеан (V. Bonjean), *Astron. Astrophys.* **634**, A81 (2020).
5. Боскуэт и др. (S. Bocquet, J.P. Dietrich, T. Schrabback, L.E. Bleem, M. Klein, S.W. Allen, et al.), *Astrophys. J.* **878**, 55 (2019).
6. Буренин и др. (R.A. Burenin, A. Vikhlinin, Hornstrup, H. Ebeling, H. Quintana, and A. Mescheryakov), *Astrophys. J. Suppl. Ser.* **172**, 561 (2007).
7. Буренин Р.А., и др., готовится к публикации (2022).
8. Буренин Р.А., Письма в Астрон. журн. **43**, 559 (2017) [R.A. Burenin, *Astron. Lett.* **43**, 507 (2017)].
9. Горски и др. (K.M. Górski, E. Hivon, A.J. Banday, B.D. Wandelt, F.K. Hansen, M. Reinecke, et al.), *Astrophys. J.* **622**, 759 (2005).
10. де Лаппарент и др. (V. de Lapparent, M.J. Geller, and J.P. Huchra), *Astrophys. J. (Lett.)* **302**, L1 (1986).
11. Карвальйо и др. (P. Carvalho, G. Rocha, and M.P. Hobson), *MNRAS* **393**, 681 (2009).
12. Кляйн и др. (M. Klein, S. Grandis, J.J. Mohr, M. Paulus, T.M. C. Abbott, J. Annis, et al.), *MNRAS* **488**, 739 (2019).

13. Коллаборация Планк и др. (Planck Collaboration, P.A.R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, et al.), *Astron. Astrophys.* **571**, A29 (2014).
14. Коллаборация Планк и др. (Planck Collaboration, R. Adam, P.A.R. Ade, N. Aghanim, Y. Akrami, M.I.R. Alves, et al.), *Astron. Astrophys.* **594**, A1 (2016a).
15. Коллаборация Планк и др. (Planck Collaboration, P.A.R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, et al.), *Astron. Astrophys.* **594**, A27 (2016b).
16. Коллаборация Планк и др. (Planck Collaboration, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Vaccigalupi, et al.), *Astron. Astrophys.* **594**, A22 (2016c).
17. Коллаборация Планк и др. (Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Vaccigalupi, et al.), *Astron. Astrophys.* **641**, A3 (2020a).
18. Коллаборация Планк и др. (Planck Collaboration, Y. Akrami, F. Argüeso, M. Ashdown, J. Aumont, C. Vaccigalupi, et al.), *Astron. Astrophys.* **641**, A2 (2020b).
19. Кравцов А.В., Вихлинин А.А., Мещеряков А.В., Письма в Астрон. журн. **44**, 4 (2018) [A.V. Kravtsov, A.A. Vikhlinin, and A.V. Meshcheryakov, *Astron. Lett.* **44**, 8 (2018)].
20. Кулуридис и др. (E. Koulouridis, N. Clerc, T. Sadibekova, M. Chira, E. Drigga, L. Faccioli, et al.), arXiv e-prints, p. arXiv:2104.06617 (2021).
21. Лин и др. (Z. Lin, N. Huang, C. Avestruz, W.L. Kimmy Wu, S. Trivedi, J. Caldeira, et al.), arXiv e-prints, p. arXiv:2102.13123 (2021).
22. Мелин и др. (J.B. Melin, J.G. Bartlett, and J. Delabrouille), *Astron. Astrophys.* **459**, 341 (2006).
23. Мелин и др. (J.B. Melin, J.G. Bartlett, P. Tarrío, and G.W. Pratt), *Astron. Astrophys.* **647**, A106 (2021).
24. Мерлони и др. (A. Merloni, P. Predehl, W. Becker, H. Böhringer, T. Boller, H. Brunner, et al.), *eROSITA Science Book: Mapping the Structure of the Energetic Universe* (2012).
25. Пиффаретти и др. (R. Piffaretti, M. Arnaud, G.W. Pratt, E. Pointecouteau, and J.B. Melin), *Astron. Astrophys.* **534**, A109 (2011).
26. Предель и др. (P. Predehl, R. Andritschke, V. Arefiev, V. Babyshkin, O. Batanov, W. Becker, et al.), *Astron. Astrophys.* **647**, A1 (2021).
27. Ремазеиллес и др. (M. Remazeilles, N. Aghanim, and M. Douspis), *MNRAS* **430**, 370 (2013).
28. Роннебергер и др. (O. Ronneberger, P. Fischer, and T. Brox), *U-Net: Convolutional Networks for Biomedical Image Segmentation* (2015).
29. Рыкофф и др. (E.S. Rykoff, E. Rozo, M.T. Busha, C.E. Cunha, A. Finoguenov, A. Evrard, et al.), *Astrophys. J.* **785**, 104 (2014).
30. Сюняев, Зельдович (R.A. Sunyaev and Y.B. Zeldovich), *Astrophys. Space Sci.* **7**, 20 (1970).
31. Сюняев, Зельдович (R.A. Sunyaev and Y.B. Zeldovich), *Comm. Astrophys. Sp. Phys.* **4**, 173 (1972).
32. Сюняев и др. (R. Sunyaev, V. Arefiev, V. Babyshkin, A. Bogomolov, K. Borisov, M. Buntov, et al.), arXiv e-prints, p. arXiv:2104.13267 (2021).
33. Таррио и др. (P. Tarrío, J.B. Melin, and M. Arnaud), *Astron. Astrophys.* **626**, A7 (2019).
34. Хариер и др. (G. Hurier, J.F. Macías-Pérez, and S. Hildebrandt), *Astron. Astrophys.* **558**, A118 (2013).
35. Хариер и др. (G. Hurier, N. Aghanim, and M. Douspis), *Astron. Astrophys.* **653**, A106 (2021).
36. Хартли и др. (R.A. Hartley et al.), готовится к публикации (2022).
37. Хилтон и др. (M. Hilton, C. Sifón, S. Naess, M. Madhavacheril, M. Oguri, E. Rozo, et al.), *Astrophys. J. Suppl. Ser.* **253**, 3 (2021).
38. Цвикки (F. Zwicky), *Astrophys. J.* **86**, 217 (1937).
39. Эйбелл и др. (G.O. Abell, Corwin, and H.G. Jr.), *A Catalog of Rich Clusters of Galaxies*, *Astrophys. J. Suppl. Ser.* (1989).