

**Автоматический поиск гигантских галактик низкой поверхностной яркости**

Е. Б. Самородова\*

*Московский государственный университет имени М. В. Ломоносова, факультет космических исследований  
Россия, 119991, Москва, Ленинские горы, д. 1, стр. 52*

(Поступила в редакцию 19.05.2022; подписана в печать 13.06.2022)

Гигантские галактики низкой поверхностной яркости (gLSB) сложно обнаружить вследствие наличия слабой протяженной периферии, которая слабо заметна на фоне неба. При этом исследование подобных галактик необходимо для уточнения механизмов формирования и эволюции гигантских дисковых галактик. Для увеличения количества известных гигантских галактик низкой поверхностной яркости планируется использовать модели машинного обучения для решения задачи бинарной классификации. Для снижения размерности задачи в качестве обучающих данных предполагается использовать радиальные профили яркости галактик, полученные методом обработки фотометрических fits-изображений галактик обзора HSC2. Для этого была разработана система поточной обработки изображений галактик и проведен изофотный анализ 26008 галактик в визуально проинспектированном ранее квадранте неба, включая 27 gLSB и 13 гигантских дисковых галактик. Все полученные профили яркости визуализированы на одном графике, вследствие чего был сформулирован критерий для отбора галактик: у потенциально интересных объектов отношение сигнал/шум на расстоянии 30 кпк от центра должно быть не менее 2 в фильтрах  $g$  и  $r$  для того, чтобы протяженная периферия явно детектировалась на фоне неба.

PACS: 98.52.Nr

УДК: 524.7-33.

Ключевые слова: гигантские дисковые галактики, формирование галактик, изофотный анализ.

**ВВЕДЕНИЕ**

Гигантские дисковые галактики, обладающие слабо светящимися протяженными отдаленными областями на расстоянии более 30 кпк от центра галактики (giant Low Surface Brightness, gLSB) — редкие, но крайне интересные астрономические объекты. Механизмы формирования подобных галактик с протяженными дисками недостаточно полно описываются существующей иерархической моделью эволюции галактик, предполагающей множество случаев взаимодействия галактик и увеличение размеров галактики в результате слияния галактик, т.к. в процессе столкновений и слияний с другими галактиками диск должен был разрушиться. На данный момент существуют два сценария формирования gLSB галактик — катастрофический и некатастрофический [1], и развитие методов автоматического поиска и детектирования гигантских галактик низкой поверхностной яркости необходимо для уточнения процессов эволюции галактик. На данный момент известно всего около сотни gLSB галактик, тогда как увеличение числа известных гигантских галактик низкой поверхностной яркости позволит перейти к статистическому подходу в исследованиях. Задача поиска новых gLSB галактик особенно актуальна сейчас в связи с появлением глубоких фотометрических обзоров (The DECam Legacy Survey, Hyper Suprime-Cam Subaru Strategic Program). Ручная инспекция таких обзоров практически невозможна, поэтому необходимо разрабатывать автоматический поиск исследуемых галактик, в частности на основе методов машинного обучения.

В данной работе разрабатывается метод автоматического поиска gLSB галактик. Главная идея подхода заключается в использовании радиальных профилей поверхностной яркости галактик вместо двумерных изображений галактик, что является физически мотивированным понижением размерности задачи для упрощения обучения классификатора.

Таким образом, была разработана система поточной обработки (pipeline) изображений для вычисления радиальных профилей галактик. Полученный pipeline был применен к глубоким фотометрическим данным обзора Hyper Suprime-Cam Subaru Strategic Program DR2 [2], доступным через cutout-сервис проекта Legacy Survey, для небольшой области неба ( $29 < \alpha < 40$ ,  $-7 < \delta < 3$ ), в которой ранее вручную было задетектировано 27 gLSB галактик (Saburova et al. in prep).

Для того, чтобы иметь физические размеры дисков галактик, рассматриваются галактики только с известными красными смещениями в диапазоне  $0.03 < z \leq 0.3$ . В результате проведен изофотный анализ изображений 33748 галактик, отобранных по указанным критериям.

Этот массив данных будет использован для обучения и тестирования различных методов классификации, включая статистические методы и алгоритмы машинного обучения (Random Forest Classifier, SVM). Методы машинного обучения, в т.ч. сверточные нейронные сети, уже применялись для анализа поверхностной яркости галактик [3] и для поиска галактик низкой поверхностной яркости по оптическим изображениям [4], однако потребуются существенная модификация подходов для поиска гигантских галактик протяженностью свыше 30 кпк при обучающей выборке, состоящей всего из нескольких десятков объектов.

\* katsam@passap.ru

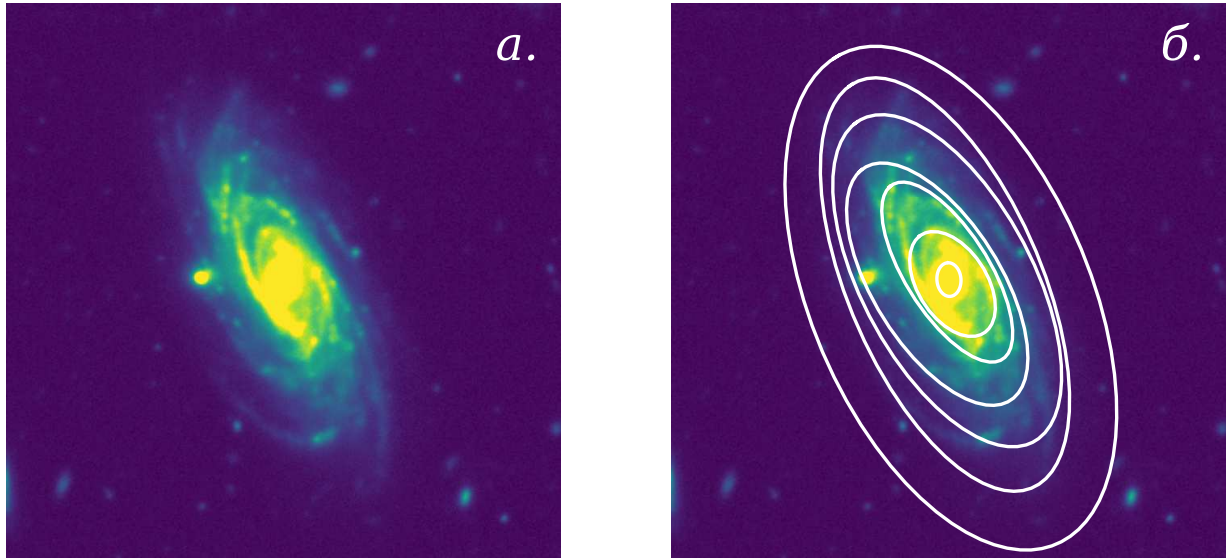


Рис. 1. Исходное изображение галактики (а) и полученный по нему набор изофот (б)

## 1. ДАННЫЕ

Для тестирования алгоритма обработки изображений галактик был выбран квадрант неба ( $29 < \alpha < 40$ ,  $-7 < \delta < 3$ ), который уже был визуально проинспектирован ранее. Для этого были отобраны все галактики в указанной области неба из каталога RCSED [5] с красным смещением в пределах  $0.03 < z \leq 0.3$ . Фильтр  $z > 0.03$  был применен для того, чтобы исключить слишком близкие галактики, изображение которых занимает большой участок неба, а  $z \leq 0.3$  — для того, чтобы исключить слишком далекие галактики с небольшими угловыми размерами, у которых периферия низкой поверхностной яркости детектируется сложнее, чем у более близких галактик. Из отобранных таким образом галактик 27 являются gLSB, а ещё 13 галактик — гигантские спиральные. Информация обо всех галактиках хранится в формате fits-таблицы, из которой в дальнейшем используются значения следующих колонок:

1. *obj\_id* — уникальный идентификатор галактики
2. *ra\_obj* — прямое склонение галактики
3. *dec\_obj* — наклонение галактики
4. *redshift* — красное смещение галактики
5. *amaj* — большая полуось эллипса, описывающего галактику
6. *ell* — степень эллиптичности эллипса, описывающего галактику
7. *pa* — угол поворота эллипса, описывающего галактику

8. (*kcorr\_g*, *kcorr\_r*, *kcorr\_z*) — *k*-поправки для (*g*, *r*, *z*) фильтров

После первичного отбора 33748 галактик были загружены fits-изображения каждой галактики обзора Hyper Suprime-Cam Subaru Strategic Program DR2 размером  $100 \times 100$  кпк при помощи сервиса Legacy Survey с масштабом 0.25 угловых секунд на пиксель.

## 2. ИЗОФОТНЫЙ АНАЛИЗ

Для каждого загруженного fits-изображения галактики проводится многоступенчатая процедура обработки полученных фотометрических данных, реализованная на языке программирования Python с использованием пакета *photutils.isophote*. Трехканальные (*g*, *r*, *z*) изображения галактик в формате fits рассматриваются как массивы чисел размера  $(3 \times N \times N)$ , где *N* вычисляется по следующей формуле с использованием пакета *astropy.cosmology*:

$$N = (\text{cosmo.arcsecperkpc} \cdot \text{comoving}(z) / 0.25 \cdot 100),$$

где 0.25 — масштаб изображения (в угловых секундах на пиксель), а 100 — сторона изображения в кпк.

Использованный метод обработки, основанный на итеративном построении изофот [6], состоит из следующих этапов:

1. Построение начальных приближений эллипсов, описывающих галактику, для каждого из трех каналов изображения, исходя из известных приближений *amaj*, *ell*, *pa*. После этого выбирается медианное значение из трех полученных центров эллипсов и фиксируется в качестве центра эллипсов *center* для всех трех каналов.

- Аналогичным образом уточняется приближение размера большой полуоси эллипсов, описывающих галактику. Для каждого канала берется значение  $ama_j$  из таблицы, если оно не превышает  $N/3$ , иначе вычисляется минимальный радиус, на котором яркость изображения составляет не более 90% от максимальной яркости галактики. Три полученных значения радиусов округляются до целочисленных килопарсеков, после чего медианное значение фиксируется в качестве большой полуоси эллипсов  $sta$  для всех трех каналов.
- Заново строится уточненное начальное приближение эллипса, описывающего галактики, с фиксированными параметрами для всех трех фильтров.
- На основе заданного начального приближения эллипса строится набор изофот с шагом 1 кпк на радиусах 0...50 кпк. При этом если некоторые изофоты не обучились вследствие плохого качества изображений или других факторов — в дальнейшем данная галактика исключается из набора обрабатываемых данных. Пример изофот, описывающих галактику, приведен на рис. 1.
- Для всех изофот, полученных в результате обработки изображения галактики, сохраняются значения средней интенсивности вдоль эллипса  $intens$ , средней ошибки  $err$  и соответствующий интервал ( $intens-err$ ,  $intens+err$ ). При этом интенсивность переводится из потоков в звездные величины по следующей логарифмической формуле, поэтому интервал получается несимметричный:

$$intens = 22.5 - 2.5 \cdot \log_{10}(flux)$$

- После этого сохраняется график полученной зависимости яркости галактики от радиуса с выделением средней величины разброса полученных значений, пример которого представлен на рис. 2.
- В конце процедуры обработки изображения галактики строится изображение эллиптической модели галактики, полученное по набору эллипсов, и разница между реальным изображением и моделью для каждого из трех фильтров. Пример такого изображения приведен на рис. 3.

### 3. ОТБОР ГАЛАКТИК

В результате проведения изофотного анализа 33748 изображений галактик было получено 26008 радиальных профилей яркости — для остальных галактик процедура построения изофот не завершилась успешно по разным причинам. После это все полученные профили яркости были изображены на одном графике (рис. 4)

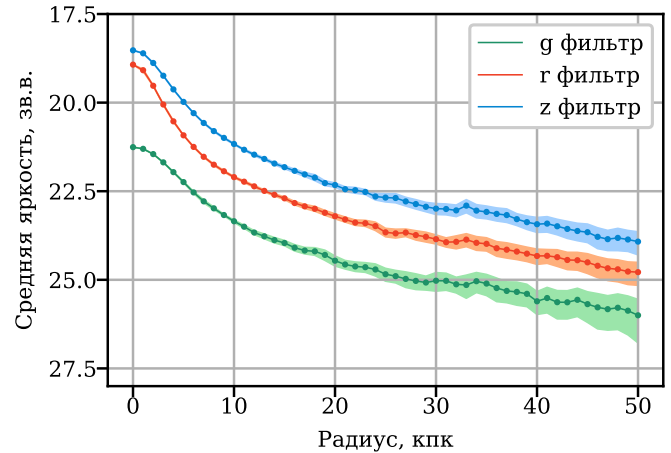


Рис. 2. Радиальный профиль поверхностной яркости галактики

с выделением гигантских галактик низкой поверхностной яркости и гигантских спиральных галактик.

По данным графикам видно, что полученные радиальные профили яркости gLSB и гигантских спиральных галактик занимают определенную область среди профилей остальных галактик. Поэтому был определен критерий, отделяющий верхнюю область профилей яркости, содержащую нужные профили, от большей части остальных профилей галактик. Согласно этому критерию, отношение значения сигнала к шуму на расстоянии 30 кпк должно быть не меньше 2 — это нужно для того, чтобы на радиусе 30 кпк на изображении была явно детектируемая периферия галактики, а не зашумленный фон неба. По указанному критерию ( $snr \geq 2$  в  $g$ ,  $r$  фильтрах) отобрано 4456 галактик, включая 27 gLSB галактик и 13 гигантских спиральных галактик.

Для дальнейшего выделения гигантских галактик низкой поверхностной яркости среди остальных объектов предполагается использовать различные критерии, в том числе уровень голубого цвета, задающегося разницей между интенсивностью потоков излучения в  $g$  и  $r$  фильтрах. Далекая слабая периферия, характерная для gLSB галактик, обычно бывает голубоватого цвета, поэтому значение  $g-r$  на 30...50 кпк у подобных галактик должно быть в среднем меньше, чем у остальных галактик, т.к. большие значения  $g-r$  задают более красные области изображения. Эти и другие аналитических признаки в дальнейшем предполагается использовать для обучения модели бинарного классификатора для детектирования галактик низкой поверхностной яркости по радиальным профилям яркости и интегральным численным признакам, вычисленных по профилям яркости.

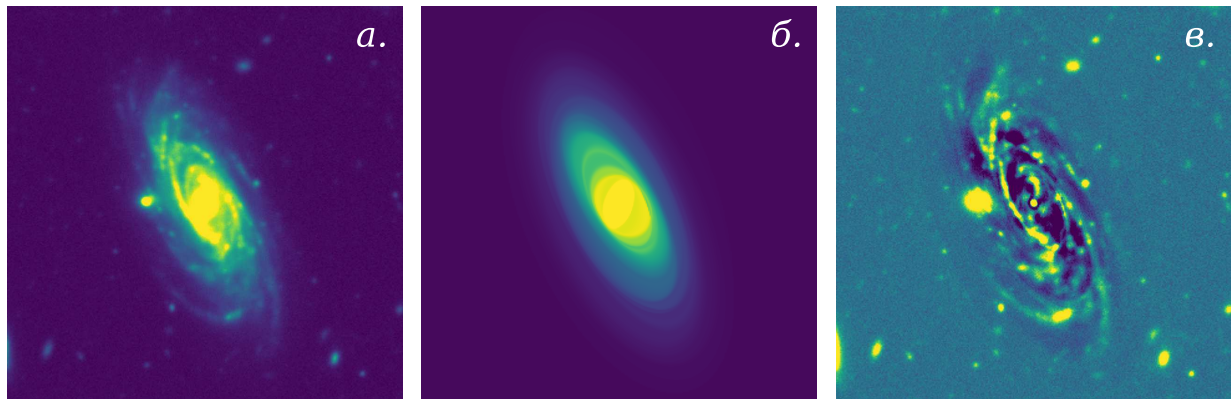


Рис. 3. Исходное изображение галактики (а), изофотная модель галактики (б), разница между изофотной моделью и исходным изображением (в)

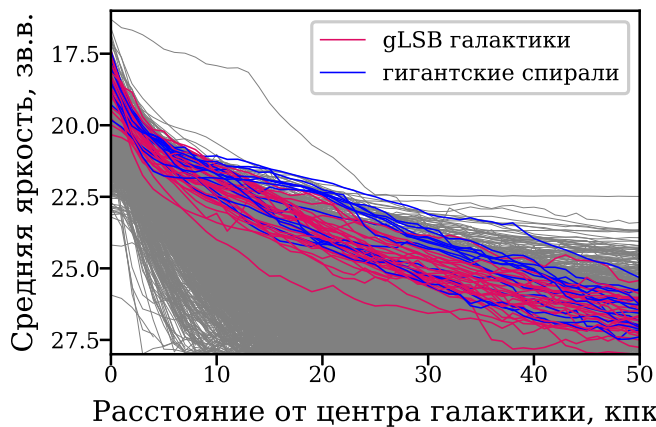


Рис. 4. Радиальные профили поерхностной яркости галактик (красные линии — гигантские галактики низкой поерхностной яркости, синии линии — гигантские спиральные галактики, серые линии — остальные объекты)

### ЗАКЛЮЧЕНИЕ

Детектирование гигантских галактик низкой поерхностной яркости является одной из важных задач для изучения механизмов формирования и эволю-

ции гигантских дисковых галактик. Однако вследствие недостаточного количества известных на данный момент примеров таких галактик решение задачи бинарной классификации напрямую по оптическим изображениям галактик представляется трудоемким процессом, поэтому требуется применение различных методов решения задачи несбалансированной классификации. Для этого был проведен изофотный анализ 33748 галактик в визуально проинспектированном раннее квадранте неба, в результате чего было получено 28006 радиальных профилей яркости, из которых было отобрано 4456 потенциально интересных объектов с отношением сигнал/шум не меньше 2 в фильтрах  $g$  и  $r$ . Среди этих объектов с детектируемой периферией на расстоянии 30 кпк — 27 гигантских галактик низкой поерхностной яркости и 13 гигантских спиральных галактик. Сравнение разницы между излучениями в  $g$  и  $r$  фильтрах показало, что далекие области gLSB галактик действительно немного более голубые, чем у остальных объектов, однако для решения задачи детектирования gLSB галактик требуется дальнейший анализ и генерирования новых аналитических признаков, указывающих на различие между профилями яркости разных типов галактик.

[1] Saburova A.S. et al. Monthly Notices of the Royal Astronomical Society. 2021. **503**, N 1. P. 803.  
[2] Hiroaki Aihara et al. // Publications of the Astronomical Society of Japan. 2019. **71**, N 6.  
[3] Li R. et al. Galaxy Light profile convolutional neural NETworks (GaLNeTs). I. Fast and accurate structural parameters for billion galaxy samples. 2021.

[4] Tanoglidis D., Ćlpijanović A., Driica-Wagner A. DeepShadows: Separating Low Surface Brightness Galaxies from Artifacts using Deep Learning, 2020.  
[5] Reference Catalog of galaxy SEDs. 2020. <https://rcsed2.voxastro.org/>  
[6] Jedrzejewski R.I. // Monthly Notices of the Royal Astronomical Society. 1987. **226**, N 4. P. 747.



## Automatic Search for Giant Low–Surface–Brightness Galaxies

E. B. Samorodova

<sup>1</sup>*Faculty of Space Research, Lomonosov Moscow State University. Moscow 119991, Russia E-mail: katsam@passap.ru*

Giant galaxies of low surface brightness (gLSB) are difficult to detect due to the presence of a faint extended periphery, which is barely visible against the sky. At the same time, the study of such galaxies is necessary to clarify the mechanisms of formation and evolution of giant disk galaxies. To increase the number of known low surface brightness giant galaxies, it is planned to use machine learning models to solve the binary classification problem. To reduce the dimension of the problem, it is proposed to use as training data the radial brightness profiles of galaxies obtained by processing photometric fits images of galaxies in the HSC2 survey. For this purpose, a system for in-line processing of galaxy images was developed and an isophotic analysis of 26008 galaxies in the visually inspected early quadrant of the sky, including 27 gLSB and 13 giant disk galaxies, was carried out. All obtained brightness profiles were visualized on one graph, as a result of which a criterion for selecting galaxies was formulated: for potentially interesting objects, the signal-to-noise ratio at a distance of 30 kpc from the center must be at least 2 in the  $g$  and  $r$  filters in order for the extended periphery to be clearly detected against the sky.

PACS: 98.52.Nr

*Keywords:* giant disk galaxies, galaxy formation, isophote analysis.

*Received 19 May 2022.*

### Сведения об авторе

Самородова Екатерина Борисовна — студентка магистратуры; e-mail: katsam@passap.ru.

---