

УДК 524.62

## КЛАССИФИКАЦИЯ ГАЛАКТИК С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ ГЛУБОКОГО ОБУЧЕНИЯ



**Д.В. Береснев**

*Магистрант кафедры электронных вычислительных машин факультета компьютерных систем и сетей БГУИР,*



**Д.И. Самаль**

*Доцент кафедры электронных вычислительных машин факультета компьютерных систем и сетей БГУИР, кандидат технических наук, доцент*

*Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь,  
E-mail: beresnev.dima8@gmail.com, samal@bsuir.by*

### **Д.В. Береснев**

*Магистрант кафедры электронных вычислительных машин БГУИР. Специальность «Технологии виртуализации и облачных вычислений». Получил высшее образование в БГУИР по специальности «Вычислительные машины, системы и сети».*

### **Д.И. Самаль**

*Кандидат технических наук, доцент кафедры ЭВМ БГУИР. Окончил Белорусский государственный университет информатики и радиоэлектроники по специальности «Вычислительные машины, системы и сети». Защитил кандидатскую диссертацию по тематике биометрической идентификации. Автор 60 научных работ по обработке изображений, распознаванию лиц, машинной графике, анализу данных, а также четырёх учебно-методических пособий.*

**Аннотация.** В данной статье описываются способ решения задачи классификации галактик с помощью технологии глубокого обучения. Представлен общий конвейер обработки изображений галактик.

**Ключевые слова:** анализ данных, большие объёмы данных, глубокое обучение, галактика, классификация галактик, galaxy zoo.

**Введение.** Прогресс в области высоких технологий, микро- и наноэлектроники, привел к созданию экспериментальных систем, генерирующих большие объёмы данных, достигающие сотен терабайт и петабайт в самых различных сферах науки и техники, в том числе и в астрономических наблюдениях. Большие массивы данных, получаемые в ходе проведения научных исследований и экспериментов необходимо хранить, обрабатывать, передавать и анализировать. По этой причине современные исследования в научных направлениях, ведущиеся на стыке наук, требуют новых математических методов, современных инженерных подходов и технических решений, которые позволили бы получать из больших массивов данных новые знания. Таким образом, построение систем обработки аст-

рономических данных, а также разработка эффективных методов анализа и технологий автоматизированной обработки этих данных являются крайне актуальными задачами в современных исследованиях. Одним из таких эффективных инструментов является глубокое обучение нейронных сетей.

*Глубокое обучение (глубинное обучение; англ. Deep learning)* — совокупность методов машинного обучения (с учителем, с частичным привлечением учителя, без учителя, с подкреплением), основанных на обучении представлениям (англ. *feature/representation learning*), с помощью алгоритмов машинного обучения, которые пытаются моделировать высокоуровневые абстракции в данных, используя архитектуры, состоящие из множества нелинейных преобразований (преобразования, при которых изменяется форма сигнала и происходит обогащение его спектра новыми частотными компонентами). Таким образом, граф, описывающий сложную иерархию высокоуровневых абстракций, будет глубоким — содержащим много уровней. Поэтому такой подход в области исследования искусственного интеллекта называется глубоким обучением.

Галактика — это массивная система, которая гравитационно связана, и состоит из звезд, звездных остатков, а также межзвездного вещества и большого количества неизвестной субстанции, условно называемой темной материей.

Морфологическая классификация галактик — система разделения галактик на группы по визуальным признакам, используемая в астрономии [1].

Последовательность Хаббла — это наиболее известный способ морфологической классификации, который был предложен в 1926 году Эдвином Хабблом. В классификации по Хабблу произведено разделение всех галактик на три больших класса, опираясь на визуальные признаки:

- класс эллиптических галактик — они имеют достаточно гладкую эллиптическую форму (геометрическая форма которых варьируется от сильно вытянутых эллипсов, до почти круглых) без отличительных характеристик с достаточно равномерным убыванием значения яркости от центра к краям.

- класс спиральных галактик — характеризуется тем, что галактики данной группы состоят из крайне плоского диска из звезд и газа, который в центре имеет сферическое уплотнение, в литературе также именуемое как балдж, а также крупное сферическое гало. Яркие спиральные рукава формируются в плоскости диска.

- неправильные или иррегулярные галактики — это галактики, значительного и ярко выраженного ядра и лишены вращательной симметрии.

**Цели и методы.** Галактики — одно из самых ярких космических явлений, и с помощью таких инструментов как «Sloan Digital Sky Survey» и «Hubble» научному сообществу теперь доступны тысячи изображений галактик, которые можно анализировать и исследовать. Однако одна из современных проблем астрономов и астрофизиков заключается в способах и методах классификации этих изображений галактик. Целью настоящей работы являлась разработка программного средства на базе методов глубокого обучения для решения задачи морфологической классификации галактик. В качестве основной модели использовалась свёрточная нейронная сеть. Свёрточные нейронные сети (CNN) уже успешно использовались и применялись в разнообразных исследованиях в этой области, прежде всего благодаря их эффективности именно в распознавании изображений. Как и другие нейронные сети, CNN состоят из слоев нейронов (которые обрабатывают входные данные), связей и их соответствующих весов (см. рис. 1).

При использовании свёрточной нейронной сети возникает проблема — большое количество выделенных признаков. Использование огромного количества признаков для решения задач распознавания изображений оказалось неэффективно. Для уменьшения размера пространства признаков проводится субдискретизация (англ. *pooling*), путём разделения

карты признаков, полученных от свёрточной нейронной сети, на фиксированное количество частей на каждой части вычисляется её максимальное значение (англ. max pooling) или среднее значение (англ. mean pooling).

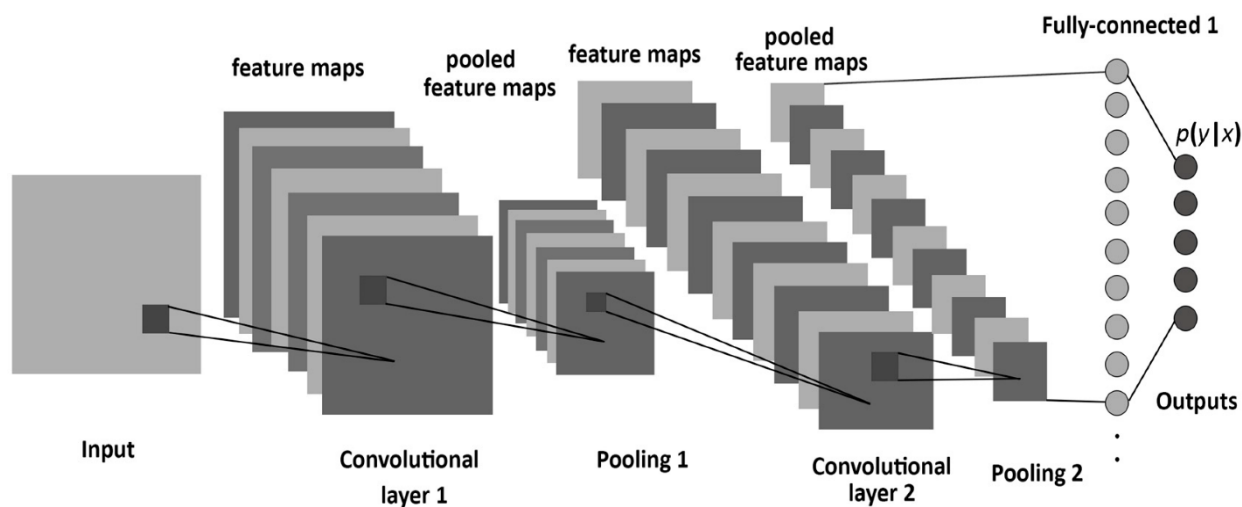


Рисунок 1. Процесс работы CNN

Можно отметить выделить следующие преимущества сверточной нейронной сети:

– уменьшение количества обучаемых параметров и повышение скорости обучения по сравнению с полносвязной нейронной сетью.

– возможность распараллеливания вычислений и реализации алгоритмов обучения сети на графических процессорах (GPU).

– устойчивость к сдвигу позиции объекта во входных данных. При обучении свёрточная нейронная сеть сдвигается по частям объекта. Поэтому обучаемые признаки не зависят от позиции «важных частей», т.е. свёрточная нейронная сеть выделяет одинаковые признаки для двух картинок, хотя позиции на изображениях могут быть разные. Таким образом, это свойство свёрточной нейронной сети помогает повышать качество при решении разнообразных задач распознавания изображений.

Метод, на базе которого производилось построение программного средства для классификации, подробно описан в [2]. Данный подход дополняет и адаптирует последовательность Хаббла для решения практических задач на реальных астрономических данных (см. рис. 2). При проектировании программного обеспечения для морфологической классификации галактик использовались отдельные теоретические и прикладные аспекты из [3]. В ходе реализации программного средства была построена нейронная сеть, архитектура которой представлена на рис. 3.

**Данные.** В ходе экспериментов использовались тренировочный набор – чуть более 60 000 изображений и тестовый набор – 80 000 изображений, каждое из которых представляло собой цветное цифровое фото в формате JPEG размером 424 x 424. После тщательного анализа изображений было выявлено, что полезной областью является только центральная часть изображения. Это позволило обрезать исходные изображения до размеров 212 x 212 относительно их центров, а так же уменьшить исходное разрешение в два раза, чтобы уменьшить количество параметров, которые придется настраивать.

**Система.** Итоговое решение имеет следующие характеристики:

- количество классов: 4;
- размеры картинки: 299 x 299;
- batch size: 8;
- количество эпох для обучения: 40;
- конволюционных слоев: 4 (с ReLu-активацией);
- полносвязных слоев: 4;
- уровень dropout: 0.5;
- выходной слой: softmax activation;
- предобработка: xception preprocess\_input (см. Keras);
- learning rate: 0.001.
- метод оптимизации: стохастический градиентный спуск;

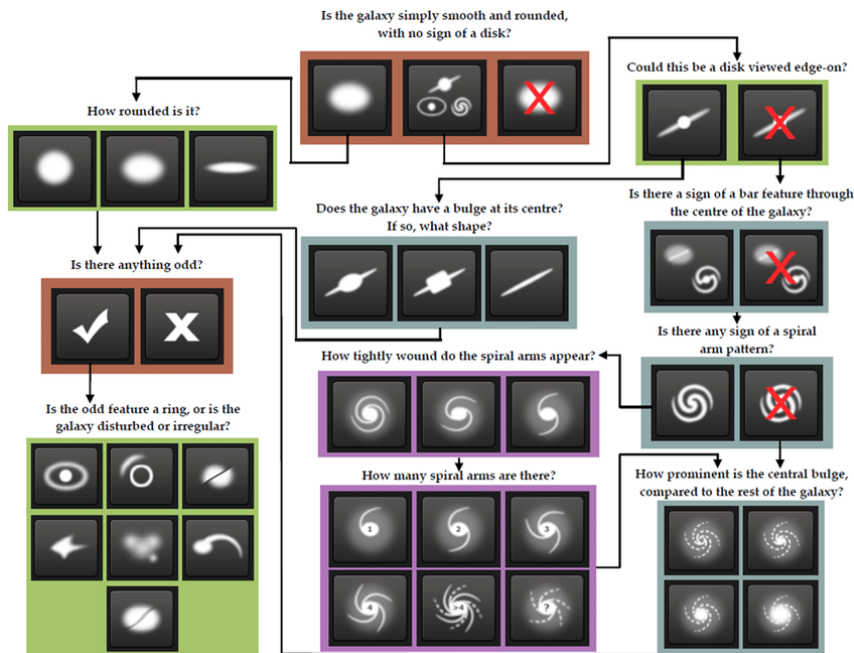


Рисунок 2. Дерево принятия решений, предложенное в проекте GalaxyZoo

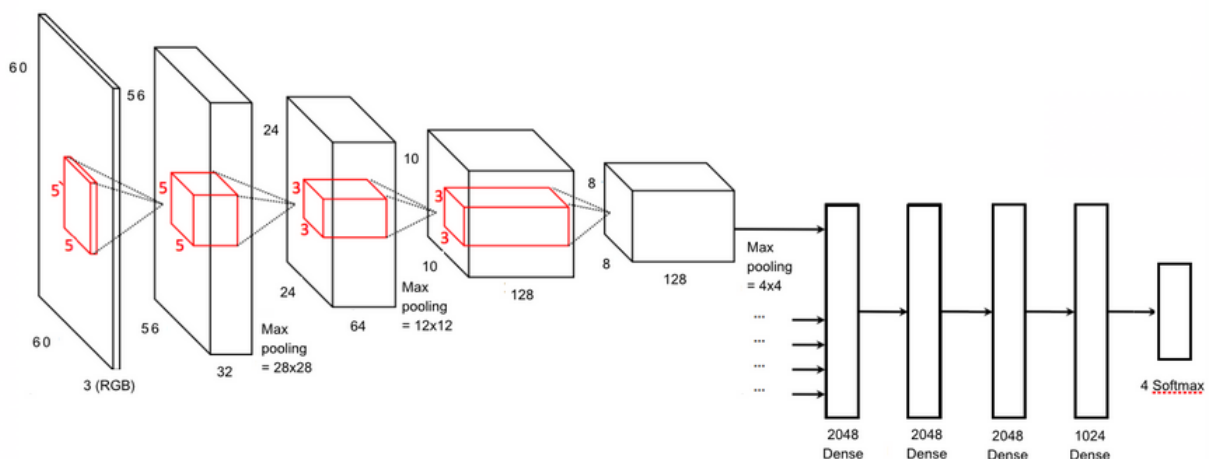


Рисунок 3. Архитектура нейронной сети

**Результаты.** В целом, CNN работал очень хорошо с loss 0,252 и средней точностью 90,9%. Сеть показала наилучшие результаты с классом галактик ellipticals и классом edge-view spirals, а наихудшие – с классом barred spirals (рисунок 4).

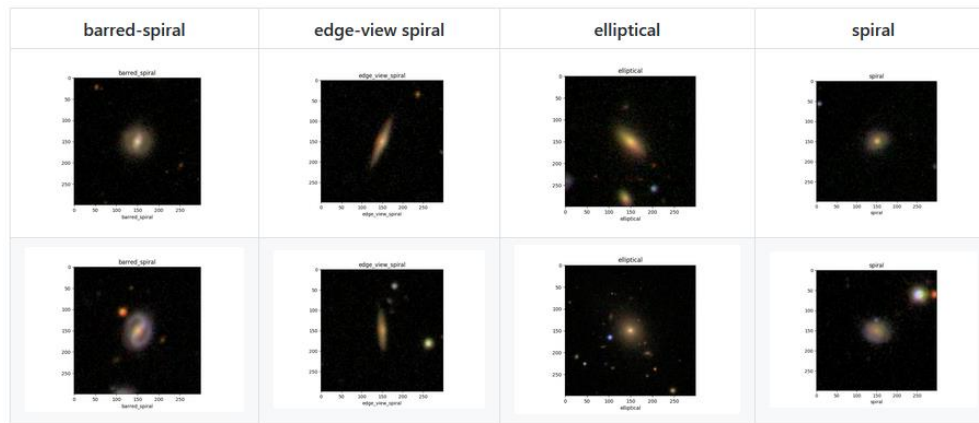


Рисунок 4. Результаты работы нейронной сети

**Заключение.** Таким образом, в результате экспериментов был разработан программный комплекс, который позволяет выполнять классификацию галактик на три основных класса: эллиптические, спиральные и иррегулярные в соответствии с упомянутым выше подходом с достаточно высоким уровнем точности. Стоит отметить, что построение систем обработки астрономических данных, а также разработка эффективных методов анализа являются крайне актуальными задачами в современных исследованиях. Следует подчеркнуть, что разработка интеллектуальных систем позволит автоматизировать многие процессы, что в свою очередь способно ускорить прогресс в астрономии и астрофизике.

#### *Литература*

- [1]. Сурдин, В.Г. Галактики / В.Г. Сурдин. – М.: ФИЗМАТЛИТ, 2013. – 432 с.
- [2]. Willet, K. W. Galaxy Zoo 2: detailed morphological classifications for 304,122 galaxies from the Sloan Digital Sky Survey / K. W. Willet, Chris J. Lintott, Steven P. Bamford. – USA : Coornell University, 2013. – 30 p. – Mode of access: <https://arxiv.org/pdf/1308.3496.pdf>. – Date of access: 09.04.2018.
- [3]. Banerji, M. Galaxy Zoo: Reproducing Galaxy Morphologies Via Machine Learning / Manda Banerji, Ofer Lahav, Chris J. Lintott. – USA : Coornell University, 2010. – 13 p. – Mode of access: <https://arxiv.org/pdf/0908.2033.pdf>. – Date of access: 09.04.2018.

## **GALAXY CLASSIFICATION USING DEEP LEARNING METHODS**

***D.V. BERESNEV***

*Master Student of the Department of  
Electronic Computing Machines  
of the Faculty of Computer Systems and Net-  
works  
of BSUIR*

***D.I. SAMAL***

*PhD (Candidate of Technical Sciences),  
Associate Professor of the Department of Elec-  
tronic Computing Machine of the Faculty of  
Computer Systems and Networks  
of BSUIR*

*Belarusian State University of Informatics and Radioelectronics, Republic of Belarus*

*E-mail: beresnev.dima8@gmail.com, samal@bsuir.by*

**Abstract.** This article describes a method for solving the galaxy classification problem using deep learning technology. A general galaxy images processing pipeline is presented.

**Keywords:** Data analysis, deep learning, galaxy, galaxy classification, galaxy zoo.