

УДК 51-37

## КЛАССИФИКАЦИЯ ИЗОБРАЖЕНИЙ ПО ИХ СОДЕРЖИМОМУ С ПРИМЕНЕНИЕМ СВЁРТОЧНОЙ НЕЙРОННОЙ СЕТИ: ПОШАГОВАЯ ИНСТРУКЦИЯ

**Нагорный Максим Алексеевич,**

Уральский государственный экономический университет, студент

Электронная почта: maxym.nagorny.61@gmail.com

Россия, Екатеринбург

**Назаров Дмитрий Михайлович,**

Уральский государственный экономический университет, доктор экономических наук,

доцент, заведующий кафедрой бизнес-информатики

Электронная почта: slup2005@mail.ru

Россия, Екатеринбург

### Аннотация

В статье рассмотрены проблемы решения задачи классификации изображений по их содержанию. Описаны процессы создания модели нейронной сети и сбора данных для её обучения и проведения тестов для оценки её качества, и также дано полное описание полной структуры модели и её гиперпараметров. Также составлен краткий алгоритм создания модели свёрточной нейронной сети и её обучения.

**Ключевые слова:** анализ данных, классификация изображений, нейронная сеть, обучение нейронной сети, сбор данных, свёрточная нейронная сеть, форматы изображений.

## IMAGE CLASSIFICATION BASED ON THEIR CONTENT USING CONVOLUTIONAL NEURAL NETWORK: A STEP-BY-STEP GUIDE

**Nagorny Maxim Alekseevich,**

Ural State University of Economics, student

Email: maxym.nagorny.61@gmail.com

Russia, Yekaterinburg

**Nazarov Dmitry Mikhailovich,**

Ural State Mining University, Doctor of Economics, Associate Professor, Head of the Department of Business Informatics

Email: slup2005@mail.ru

Russia, Yekaterinburg

ABSTRACT

This article considers the difficulties of solving the problem of image classification based on their content. The processes of creating a neural network model and collecting data for its training and testing to evaluate its quality are described, and a full description of the complete structure of the model and its hyperparameters is also given. A short algorithm for creating a convolutional neural network model and training it is also presented.

**Keywords:** data analysis, image classification, neural network, neural network training, data mining, convolutional neural network, image file formats.

В последнее время активно ведутся разработки оптимальных универсальных методов решения задачи машинного зрения и классификации изображений по их содержанию. Эти решения нужны для автоматизации многих задач, например: фильтрация контента в медиа файлах, распознавание лиц, объектов или жестов. При этом очень важно качество результата процесса применения этих методов, чтобы гарантировать чёткое разделение содержимого без участия человека, что позволит сэкономить много человеческих ресурсов, которые можно направить на более важные задачи, участие человека в которых необходимо. Для решения вышеописанной задачи обычно используют нейронную сеть.

Существует множество вариаций моделей нейронной сети, среди которых есть:

- многослойный перцептрон (MLP);
- рекуррентная нейронная сеть (RNN);
- свёрточная нейронная сеть (CNN);
- глубокая нейронная сеть (DNN);
- автокодировщик (autoencoder);
- генеративная сеть [1].

Из вышеперечисленных вариантов для решения задачи классификации изображений используют свёрточную нейронную сеть.

Свёрточные нейронные сети (CNN) используют трехмерные данные для классификации изображений и распознавания объектов. Они отличаются от других нейронных сетей превосходной производительностью при работе с изображениями, речью или аудиосигналами. Её структура, указанная на рисунке 1, формируется из трёх основных типов слоев, а именно:

- Свёрточный слой (convolutional layer);
- Пулинг-слой, слой подвыборки, слой уплотнения (pooling layer);
- Полносвязный слой (FC layer).

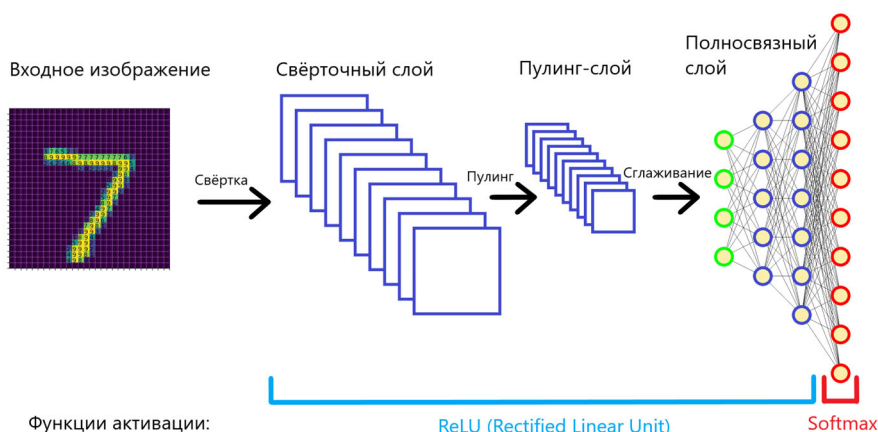


Рисунок 1 – структура свёрточной нейронной сети.

Рассмотрим их подробнее. Свёрточный слой – это основной строительный блок этой модели, в котором происходит большая часть вычислений. Он состоит из нескольких компонентов: входных данных, фильтра и карты признаков. Предположим, что на входе будет цветное изображение, состоящее из матрицы пикселей в 3D. Это означает, что входные данные будут иметь три измерения - высоту, ширину и глубину, которые соответствуют цветовой модели RGB. В слое также есть детектор признаков, также известный как ядро или фильтр (kernel), который перемещается по рецептивным полям изображения, проверяя, присутствует ли признак (признаки модель определяет для себя сама). Этот процесс известен как свёртка.

Пулинг-слои осуществляют снижение размерности, уменьшая количество параметров на входе. Как и в случае со свёрточным слоем, операция объединения обходит фильтр по всему входу, но разница в том, что этот фильтр не имеет весов. Вместо этого ядро применяет агрегирующую функцию к значениям в рецептивном поле, заполняя выходной массив. Существует два основных типа объединения (пулинга):

- Максимальный пулинг (max pooling): по мере перемещения фильтра по входу он выбирает пиксель с максимальным значением для отправки в выходной массив. Этот подход, как правило, используется чаще по сравнению со средним пулингом.
- Средний пулинг (average pooling): по мере перемещения фильтра по входу, он вычисляет среднее значение в пределах рецептивного поля для отправки в выходной массив.

Хоть на этом слое и теряется много информации, он также имеет ряд преимуществ для CNN. Они помогают снизить сложность, повысить эффективность и ограничить риск переобучения.

Структура полносвязного слоя представляет собой обычный многослойный перцептрон. Этот слой выполняет задачу классификации на основе признаков, извлеченных с помощью предыдущих слоев и их различных фильтров. В то время как свёрточные и пулинг слои обычно используют функции активации ReLu, полносвязные слои обычно используют функцию активации Softmax, чтобы классифицировать входные данные, создавая на выходе вероятности принадлежности к каждому классу от 0 до 1 [6].

Это описание структуры классической архитектуры свёрточной нейронной сети, но на её основе созданы архитектуры под различные задачи. Они предлагают определённые настройки, такие как: количество слоёв и нейронов, функции активации, параметры слоёв свёртки и другие. Поэтому перед разработкой модели важно выбрать подходящую архитектуру. Для этого было проведено множество исследований, в которых сравнивались модели с разной структурой и сделаны выводы по их качеству. В качестве примера можно привести исследование по классификации снимков опухолей головного мозга. Самую наибольшую точность показала модель с архитектурой «Multi-Path Convolutional Neural Network» со значением 96% [3]. А при классификации изображений с раком кожи лучше всего справилась модель с архитектурой «SLC-CNN» с применением алгоритма XGBoost, показав точность 99% [7]. Также следует упомянуть, что от сложности архитектуры также зависит и время машинного обучения и соответствующие вычислительные мощности, поэтому при выборе структуры очень важно учесть и продумать этот момент.

Общая архитектура описана, теперь рассмотрим процесс подготовки данных из изображений, которые в дальнейшем будут в неё загружены.

Для файлового представления изображений используют множество форматов: JPEG (далее JPG), PNG, GIF, WebP, TIFF, BMP, HEIF, SVG, EPS, PDF, PSD, AI, XCF, INDD и другие, наиболее популярные из которых JPG, PNG и PDF. Рассмотрим подробнее формат JPG.

JPG – это формат файлов растровых изображений, использующий сжатие с потерями, что позволяет уменьшить количество хранимой информации в файле, сохраняя при этом возможность комфортного восприятия изображения человеческим глазом. Он использует в себе цветовая модель RGB (red green blue) – один из самых распространенных методов представления цвета в компьютерной графике. В ней используется система цветовых координат с тремя основными цветами: красный, зелёный и синий.

Каждый основной цвет (их называют каналами) может принимать значение интенсивности от 0 (наименьшее) до 1 (наибольшее), если говорить о дробном представлении, в целочисленном представлении значения каналов принимают значения от 0 до 255 включительно. Смешивание этих трех каналов с различными уровнями интенсивности позволяет получать множество цветов [5]. И все эти цвета изображения в виде наборов чисел из трёх каналов подаются на первый трёхмерный свёрточный слой модели, после чего результат вычисления последовательно передаётся на следующие слои. Также для того, чтобы модель выявляла для себя больше признаков для каждого класса изображения, важно обеспечить большие наборы изображений в плане их количество и дополнять их различными шумами и искажениями, чтобы было больше разнообразия при обучении. Поскольку модель имеет фиксированные гиперпараметры: количество слоёв модели и узлов в них, необходимо менять размеры всех изображений под размер входного свёрточного слоя.

Говоря про гиперпараметры: они подбираются вслепую, поскольку модель нейронной сети представляет из себя «чёрный ящик» – нельзя заранее точно предугадать наиболее оптимальные значения этих гиперпараметров. Осталось поговорить о функции потерь, оптимизаторе и метриках, которые будут использоваться в модели.

Функция потерь: Категориальная кросс-энтропия (`categorical_crossentropy`). Она широко применяется в задачах классификации [4]. При обучении модель всегда стремится к минимизации значения функции потерь.

Оптимизатор: Среднеквадратичное распространение корня (`RMSprop`), помимо него также можно использовать оптимизатор Adam, оба хорошо себя показывают для решения многих задач [2]. Оптимизаторы при обучении подбирают наиболее оптимальные значения смещений и весов, что упрощает их регулировку и вследствие чего сокращает время для обучения.

Метрика: точность (`Accuracy`). Показывает, насколько точно прогнозы модели соответствуют правильному результату. Для получения точности в процентах нужно умножить получившееся значение на 100. На основе метрик можно с лёгкостью оценить качество модели и определить случаи переобучения на ранних этапах.

При обучении модели нужно отслеживать изменение значений функции потерь и метрики у наборов для обучения и проверок на каждой эпохе: точность должна увеличиваться, а значение функции потерь – уменьшаться.

После обучения модели нужно пропустить наборы изображений, которых не было в наборах для обучения и внутренних проверок, и на основе процентного соотношения правильных распределений к неправильным можно сделать оценку качества модели.

Всю вышеописанную информацию можно вкратце уместить в виде пошагового алгоритма:

1. Собрать большое количество изображений с разными классами их содержимого;

2. Выбрать наиболее оптимальную существующую архитектуру свёрточной нейронной сети на основе исследований по решению задач классификации изображений.
3. Выбрать язык и среду разработки, позволяющие создавать свёрточные нейронные сети (например: необходимый инструментарий есть в библиотеке Keras, которая поддерживается языками R и Python);
4. Описать структуру модели со всеми слоями, узлами и ядрами, которые используются в операциях свёртки и пулинга, на основе выбранной архитектуры.
5. Применить функцию потерь «категориальная кросс-энтропия», метрику «точность» («accuracy») и оптимизатор «Среднеквадратичное распространение корня» («RMSprop») или «Adam».
6. Изменить размеры всех изображений под размеры входного свёрточного слоя.
7. Подготовить на основе изображений наборы данных для обучения, проверок и тестов (данные для проверок используются для получения значения функции потерь и точности на каждой эпохи обучения, а тестовые данные используются после полного завершения обучения). Данные из себя представляют наборы цветов, представленных в формате RGB, при этом значения интенсивности задаются в промежутке от 0 до 1.
8. Дополнить эти наборы теми же изображениями, но с шумами и искажениями (некоторые библиотеки содержат встроенные методы для этого).
9. Обучить модель с помощью данных для обучения и проверок, предварительно выбрав количество эпох обучения. Желательно при обучении отслеживать значения метрики и функции потерь после каждой эпохи.
10. Применить тестовый набор для оценки качества модели, загружая данные на входной слой модели. Качество можно оценить по значению точности и составлению матрицы ошибок.

При наличии качественного набора изображений, данный алгоритм позволит создать модель свёрточной нейронной сети, которая сможет решать задачу классификации изображений по их содержанию.

#### **Список литературы:**

1. Институт информационных систем Государственного университета управления // Виды нейронных сетей - [2023]: сайт. - URL: <https://iis.guu.ru/blog/vidy-neironnih-setey/> / (дата обращения: 18.03.2025).
2. Хабр // Реализуем и сравниваем оптимизаторы моделей в глубоком обучении - [2020]: сайт. - URL: <https://habr.com/ru/companies/skillfactory/articles/525214/> / (дата обращения: 18.03.2025).
3. Amreen Batool, Yung-Cheol Byun // A lightweight multi-path convolutional neural network architecture using optimal features selection for multiclass classification of brain tumor using magnetic resonance images - [2025]: науч. ст. - URL: <https://www.sciencedirect.com/science/article/pii/S2590123025004086> / (дата обращения: 18.03.2025).

4. Balkrishna Pandey // categorical\_crossentropy what is? - [2020]: сайт. - URL: [https://www.goglides.dev/bkpandey/categorical\\_crossentropy-what-is-398f](https://www.goglides.dev/bkpandey/categorical_crossentropy-what-is-398f) / (дата обращения: 18.03.2025).
5. GeeksforGeeks // Computer Graphics | The RGB color model - [2022]: сайт. - URL: <https://www.geeksforgeeks.org/computer-graphics-the-rgb-color-model/> / (дата обращения: 18.03.2025).
6. IBM // What are convolutional neural networks? - сайт. - URL: <https://www.ibm.com/topics/convolutional-neural-networks> / (дата обращения: 18.03.2025).
7. Umme Sara, Utpol Kanti Das // Automated Skin Cancer Classification and Detection Using Convolutional Neural Networks and Dermoscopy Images - [2025]: науч. ст. - URL: <https://www.sciencedirect.com/science/article/pii/S1877050924034446> / (дата обращения: 18.03.2025).

**References:**

1. Institute of Information Systems of the State University of Management // Types of neural networks - [2023]: website. - URL: <https://iis.guu.ru/blog/vidy-neironnih-setey/> / (date of access: 03/18/2025).
2. Habr // Implementing and comparing model optimizers in deep learning - [2020]: website. - URL: <https://habr.com/ru/companies/skillfactory/articles/525214/> / (date of access: 03/18/2025).
3. Amreen Batool, Yung-Cheol Byun // A lightweight multi-path convolutional neural network architecture using optimal features selection for multiclass classification of brain tumor using magnetic resonance images - [2025]: scientific article. - URL: <https://www.sciencedirect.com/science/article/pii/S2590123025004086> / (accessed: 18.03.2025).
4. Balkrishna Pandey // categorical\_crossentropy what is? - [2020]: site. - URL: [https://www.goglides.dev/bkpandey/categorical\\_crossentropy-what-is-398f](https://www.goglides.dev/bkpandey/categorical_crossentropy-what-is-398f) / (accessed: 18.03.2025).
5. GeeksforGeeks // Computer Graphics | The RGB color model - [2022]: site. - URL: <https://www.geeksforgeeks.org/computer-graphics-the-rgb-color-model/> / (accessed: 18.03.2025).
6. IBM // What are convolutional neural networks? - site. - URL: <https://www.ibm.com/topics/convolutional-neural-networks> / (accessed: 18.03.2025).
7. Umme Sara, Utpol Kanti Das // Automated Skin Cancer Classification and Detection Using Convolutional Neural Networks and Dermoscopy Images - [2025]: scientific article - URL: <https://www.sciencedirect.com/science/article/pii/S1877050924034446> / (accessed: 18.03.2025).