

## ПРИМЕНЕНИЕ МЕТОДА ГЛУБОКОГО ОБУЧЕНИЯ В ЗАДАЧЕ ИДЕНТИФИКАЦИИ ДОСТОПРИМЕЧАТЕЛЬНОСТЕЙ НА ИЗОБРАЖЕНИИ

Нгуен Куй Тхань

thanhngq@lqdtu.edu.vn

МГТУ им. Н.Э. Баумана, Москва, Российская Федерация

---

### Аннотация

Рассмотрен подход к идентификации достопримечательностей на изображениях с использованием сверточных нейронных сетей. Представлен обзор основных этапов алгоритма: подготовка данных, выбор архитектуры CNN, обучение модели, оценка результатов. Описаны такие методы предобработки изображений, как удаление шума и объектов заданного плана. Приведены примеры популярных архитектур CNN. Этап обучения включает инициализацию весов, прямое и обратное распространение, обновление весов с целью минимизации потерь. Для оценки точности классификации применены метрики типа accuracy, precision, recall. Рассмотрены подходы к улучшению результатов, такие как увеличение данных, изменение структуры сети, улучшение качества изображений. Подчеркнута важность распознавания достопримечательностей в сферах туризма, архитектуры, наследия. Выполнен обзорный анализ применения глубокого обучения для автоматизации данной задачи.

### Ключевые слова

Компьютерное зрение, сверточная нейронная сеть, распознавание изображений, обнаружение объектов, обработка изображений, глубокое обучение, методы искусственного интеллекта, DELF

Поступила в редакцию 05.10.2023

© МГТУ им. Н.Э. Баумана, 2023

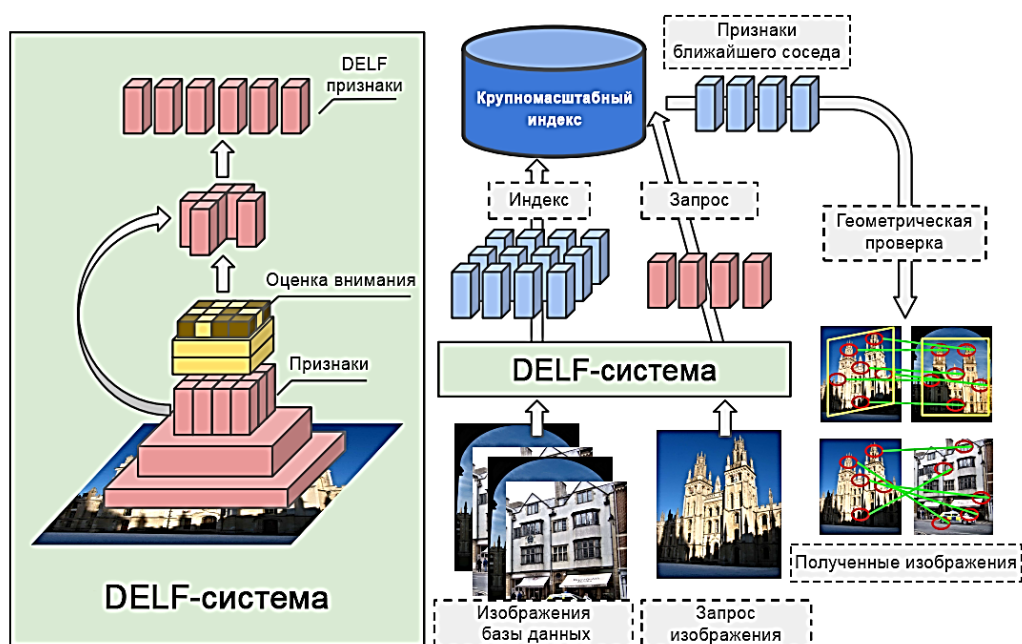
---

**Введение.** В последние десятилетия, характеризующиеся развитием компьютерного зрения и искусственного интеллекта, глубокое обучение стало одной из наиболее эффективных и мощных методик в области анализа и обработки изображений. Одно из новых направлений применения глубокого обучения — идентификация изображений.

Идентификация достопримечательностей на изображениях имеет огромное значение в различных областях: в туристической отрасли, различных исследованиях в архитектуре и строительстве, при изучении и сохранении культурного наследия [1]. Традиционные методы распознавания изображений обычно основаны на ручном создании признаков и алгоритмах классификации, что требует большого объема работы и экспертных знаний. Однако с возникновением глубокого обучения и сверточных нейронных сетей (CNN) появились новые возможности для автоматизации и повышения точности распознавания и идентификации достопримечательностей.

### Основы глубокого обучения в задачах распознавания изображений.

В настоящее время наиболее эффективным методом является DELF (DEep Local Features — глубокие локальные признаки) (см. рисунок), разработанный компанией Google. Этот метод объединяет сравнение локальных признаков с применением глубокого обучения. Набор DELF-признаков получают путем передачи входного изображения через сверточную нейронную сеть.



Архитектура системы поиска изображений в целом, основанная на использовании глубоких локальных признаков (DELf) [2]

Сверточные нейронные сети обладают способностью автоматически изучать иерархические признаки изображений. В сверточных слоях нейронные сети применяют фильтры, которые скользят по изображению, выполняя операцию свертки для извлечения различных признаков, таких как границы, текстуры и формы. Последующие слои объединяют признаки с целью уменьшения размерности и сохранения наиболее значимых характеристик изображения. Наконец, полносвязные слои выполняют классификацию на основе извлеченных признаков.

Одним из ключевых преимуществ глубокого обучения для распознавания изображений служит его способность к эффективному изучению иерархических признаков. Вместо того чтобы вручную определять и описывать признаки, CNN автоматически находят и извлекают различные уровни абстракции, что позволяет модели быть более гибкой и адаптивной к различным типам изображений.

**Алгоритм распознавания достопримечательностей на изображении с применением CNN.** Алгоритм распознавания достопримечательностей с применением сверточных нейронных сетей (CNN) представляет собой итеративный процесс, который состоит из нескольких этапов.

**Этап 1. Подготовка данных.** На первом этапе осуществляется подготовка данных, включающая сбор и предварительную обработку обучающего набора изображений достопримечательностей с соответствующими классовыми метками. Затем набор данных разделяется на обучающую и тестовую выборки для оценки производительности модели [3, 4].

**Этап 2. Создание сверточной нейронной сети.** На втором этапе создается сверточная нейронная сеть, в которой определяется ее архитектура, включающая сверточные слои, слои объединения и полносвязные слои [5]. Также настраиваются гиперпараметры модели, такие как размер фильтров, количество нейронов, функции активации и скорость обучения.

Выбор подходящей архитектуры сверточной нейронной сети — один из важных факторов, влияющих на эффективность распознавания достопримечательностей. На сегодняшний день в литературе представлено множество архитектур, включая AlexNet [6], VGGNet [7] и ResNet [8], которые различаются по глубине и сложности. Каждая из этих архитектур имеет свои особенности и преимущества, которые необходимо учитывать.

При выборе архитектуры сверточной нейронной сети необходимо принять во внимание целый ряд факторов: доступные вычислительные ресурсы, размер обучающего набора и требуемая точность распознавания. Некоторые архитектуры, например, ResNet, имеют глубокую структуру с большим числом слоев, что может потребовать значительных вычислительных ресурсов при обучении и дальнейшем использовании. В то же время более простые архитектуры, такие как AlexNet, могут быть менее требовательны к вычислительным ресурсам, но могут иметь ограниченную способность извлекать сложные признаки.

**Этап 3. Обучение модели.** Далее происходит обучение модели, которое включает несколько итераций с целью достижения сходимости. В начале процесса инициализируются веса модели. Затем изображения обучающей выборки проходят через сверточные и полносвязные слои модели в процессе прямого распространения. В каждом слое происходит вычисление активации нейронов с использованием соответствующих весов и функции активации.

После прямого распространения вычисляется функция потерь, которая позволяет измерить разницу между предсказанными классами и фактическими метками обучающих изображений. В задачах классификации в качестве функции потерь часто применяется функция кросс-энтропии.

Затем происходит обновление весов модели с использованием алгоритма обратного распространения ошибки. Этот алгоритм распространяет ошибку от выходного слоя к начальным слоям модели пропорционально влиянию каждого веса на общую ошибку. Веса обновляются в направлении, противоположном

ном градиенту функции потерь с учетом скорости обучения, которая определяет изменение весов на каждой итерации.

Этот процесс повторяется на нескольких эпохах, где каждая эпоха представляет собой полный проход через всю обучающую выборку. Цель обучения состоит в том, чтобы уменьшить функцию потерь и настроить веса модели таким образом, чтобы она максимально точно предсказывала классы достопримечательностей на изображениях.

**Этап 4. Оценка производительности модели.** После обучения модель оценивается на тестовой выборке для определения ее производительности [9]. Для оценки способности модели к правильной классификации достопримечательностей на изображениях используют различные метрики, включая меткость (accuracy), точность (precision), полноту (recall) и F1-меру (F1-score).

**Этап 5. Применение модели для распознавания достопримечательностей.** После успешного обучения и оценки модели ее можно применять для распознавания достопримечательностей на новых изображениях. Модель получает изображение на вход и с помощью прямого распространения выдает предсказание класса достопримечательности.

**Шаг 6. Оценка и улучшение модели.** Оценка результатов распознавания и анализ ошибок играют важную роль для оптимизации модели идентификации достопримечательностей [10]. С целью улучшения производительности модели и достижения более высокой точности и надежности применяют различные подходы.

Один из подходов заключается в увеличении обучающего набора данных. Путем добавления большего числа разнообразных изображений достопримечательностей в обучающую выборку модель получает больше информации для обучения и может лучше обобщать признаки достопримечательностей.

Изменение архитектуры модели также служит важным аспектом оптимизации. Варьирование числа слоев, размеров фильтров и добавление дополнительных сверточных или полносвязных слоев может привести к улучшению способности модели распознавать достопримечательности.

Другим подходом является улучшение качества входного изображения. Поскольку входные данные представляют собой естественную сцену, в них присутствует множество нежелательных элементов, которые могут негативно сказаться на результате и качестве работы нейронной сети. Поэтому перед подачей данных на вход сети для обучения, на этапе предварительной обработки изображения, необходимо удалить объекты, не связанные с достопримечательностью, такие как деревья, люди на переднем плане и так далее.

**Заключение.** Применение глубокого обучения для распознавания достопримечательностей на изображениях представляет собой активно развивающуюся область исследований с большим потенциалом для применения в различных сферах, таких как туризм, архитектура и культурное наследие.

Описанный алгоритм распознавания достопримечательностей с применением CNN включает несколько этапов. На первом этапе происходит предварительная обработка входных изображений, включающая удаление шумовых элементов и нежелательных объектов, не связанных с ландшафтом или достопримечательностями. Затем происходит этап обучения модели, включающий инициализацию весов, прямое распространение изображений через сверточные и полносвязные слои, вычисление функции потерь и обновление весов с использованием алгоритма обратного распространения ошибки. Этот процесс повторяется на нескольких эпохах для достижения сходимости модели. Оценка результатов распознавания и анализ ошибок являются важными этапами для оптимизации модели.

## Литература

- [1] Распознавание сцен на изображениях с помощью глубоких сверточных нейронных сетей. URL: <https://habr.com/ru/companies/vk/articles/359214/> (дата обращения 19.05.2023).
- [2] Noh H., Araujo A., Sim J., Weyand T., Han B. Large-Scale Image Retrieval with Attentive Deep Local Features. *ICCV*, 2017, pp. 3456–3465.
- [3] Weyand T., Araujo A., Cao B., Sim J. Google Landmarks Dataset v2 – A Large-Scale Benchmark for Instance-Level Recognition and Retrieval. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 2575–2584.
- [4] Google Landmark Recognition 2021. URL: <https://www.kaggle.com/c/landmark-recognition-2021/overview/iccv-2021> (accessed May 19, 2023).
- [5] Xu C., Wang W., Liu S., Wang Y., Tang Y., Bian T., Yan Y., She Q., Yang C. 3rd Place Solution to Google Landmark Recognition Competition 2021. *arXiv preprint arXiv:2110.02794*, 2021.
- [6] AlexNet — Сверточная нейронная сеть для классификации изображений. URL: <https://neurohive.io/ru/vidy-nejrosetej/alexnet-svjortochnaja-nejronnaja-set-dlja-raspoznavanija-izobrazhenij/> (дата обращения 19.05.2023).
- [7] VGG Very Deep Convolutional Networks (VGGNet) — What you need to know. URL: <https://viso.ai/deep-learning/vgg-very-deep-convolutional-networks/> (accessed May 19, 2023).
- [8] ResNet (34, 50, 101): «остаточные» CNN для классификации. URL: <https://neurohive.io/ru/vidy-nejrosetej/resnet-34-50-101/> (дата обращения 19.05.2023).
- [9] Нгуен Т.К., Сырямкин В.И., Нгуен Ч.Х.Т. Модель метода распознавания объектов на изображениях с использованием «сверточной нейронной сети — CNN». *Современные наукоемкие технологии*, 2020, vol. 12–2, с. 269–280. <http://doi.org/10.17513/snt.38445>
- [10] Шарибаев Р.Н., Джураев Ш.С., Тохиржонова М.Р. Улучшение классификации по сортам коконов с использованием сверточных нейронных сетей. *Теория и практика современной науки*, 2023, № 6 (96), с. 212–214.

**Нгуен Куй Тхань** — студент кафедры «Информационные системы и телекоммуникации», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

**Научный руководитель** — Боровик Ирина Геннадьевна, старший преподаватель кафедры «Информационные системы и телекоммуникации», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

**Ссылку на эту статью просим оформлять следующим образом:**

Нгуен Куй Тхань. Применение метода глубокого обучения в задаче идентификации достопримечательностей на изображении. *Политехнический молодежный журнал*, 2023, № 11 (88). <http://dx.doi.org/10.18698/2541-8009-2023-11-950>

## APPLICATION OF THE DEEP LEARNING METHOD IN SOLVING THE PROBLEM OF IDENTIFYING LANDMARKS IN THE IMAGE

Nguyen Quy Thanh

thanhnq@lqdtu.edu.vn

Bauman Moscow State Technical University, Moscow, Russian Federation

---

### Abstract

The paper considers an approach to identifying landmarks in the images using the convolutional neural networks. It presents an overview of the algorithm main phases including data preparation, selecting the CNN architecture, model learning, and results evaluation. Image preprocessing methods such as removing noise and background objects are described. Examples of the popular CNN architectures are provided. The learning phase includes weight initialization, forward and backward propagation and weights updating to minimize losses. To assess the classification accuracy, the following metrics are used: accuracy, precision, and recall. Approaches to improving the results are considered: data increase, network structure alteration, and image quality improvement. Importance of recognizing landmarks in tourism, architecture, and heritage is emphasized. A review analysis of introducing deep learning in automation of this task is presented.

### Keywords

Computer vision, convolutional neural network, image recognition, object detection, image processing, deep learning, artificial intelligence methods, DELF

Received 05.10.2023

© Bauman Moscow State Technical University, 2023

---

### References

- [1] *Raspoznavanie stsen na izobrazheniyakh s pomoshch'yu glubokikh svertochnykh neyronnykh setey* [Scene recognition in images using deep convolutional neural networks]. URL: <https://habr.com/ru/companies/vk/articles/359214/> (accessed May 19, 2023).
- [2] Noh H., Araujo A., Sim J., Weyand T., Han B. Large-Scale Image Retrieval with Attentive Deep Local Features. *ICCV*, 2017, pp. 3456–3465.
- [3] Weyand T., Araujo A., Cao B., Sim J. Google Landmarks Dataset v2 – A Large-Scale Benchmark for Instance-Level Recognition and Retrieval. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 2575–2584.
- [4] *Google Landmark Recognition 2021*. URL: <https://www.kaggle.com/c/landmark-recognition-2021/overview/iccv-2021> (accessed May 19, 2023).
- [5] Xu C., Wang W., Liu S., Wang Y., Tang Y., Bian T., Yan Y., She Q., Yang C. 3rd Place Solution to Google Landmark Recognition Competition 2021. *arXiv preprint arXiv: 2110.02794*, 2021.
- [6] *AlexNet — Svertochnaya neyronnaya set' dlya klassifikatsii izobrazheniy* [AlexNet — Convolutional neural network for image classification]. URL: <https://neurohive.io/ru/vidy-nejrosetej/alexnet-svjortchnaja-nejronnaja-set-dlja-raspoznavanija-izobrazhenij/> (accessed May 19, 2023).

- [7] VGG Very Deep Convolutional Networks (VGGNet) — *what you need to know*. URL: <https://viso.ai/deep-learning/vgg-very-deep-convolutional-networks/> (accessed May 19, 2023).
- [8] ResNet (34, 50, 101): *ostatochnye CNN dlya klassifikatsii* [ResNet (34, 50, 101): residual CNNs for classification]. URL: <https://neurohive.io/ru/vidy-nejrosetej/resnet-34-50-101/> (accessed May 19, 2023).
- [9] Nguyen T.C., Syryamkin V.I., Nguyen T.H.T. Model for the method of object recognition in images using convolution neuron network — CNN. *Modern high technology*, 2020, vol. 12–2, pp. 269–280. (In Russ.). <http://doi.org/10.17513/snt.38445>
- [10] Sharibaev R.N., Juraev Sh.S., Tokhirzhonova M.R. Improving classification by cocoon varieties using convolutionary neural networks. *Theory and practice of modern science*, 2023, no. 6 (96), pp. 212–214. (In Russ.).

**Nguyen Quy Thanh** — Student, Department of Information Systems and Telecommunications, Bauman Moscow State Technical University, Moscow, Russian Federation.

**Scientific advisor** — Borovik I.G., Senior Lecturer, Department of Information Systems and Telecommunications, Bauman Moscow State Technical University, Moscow, Russian Federation.

**Please cite this article in English as:**

Nguyen Quy Thanh. Application of the deep learning method in solving the problem of identifying landmarks in the image. *Politekhicheskiy molodezhnyy zhurnal*, 2023, no. 11 (88). (In Russ.). <http://dx.doi.org/10.18698/2541-8009-2023-11-950>