

Р.В. Дзвинко

О НЕКОТОРЫХ ПОДХОДАХ К МАШИННОМУ ОБУЧЕНИЮ
СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ В ЗАДАЧАХ ОБРАБОТКИ
АЭРОФОТОИЗОБРАЖЕНИЙ РАЗЛИЧНЫХ ФОРМАТОВ
ПРЕДСТАВЛЕНИЯ

Аннотация. В представленной научной работе приведен один из подходов к решению задачи распознавания образов на массиве исходных данных аэрофотоснимков различного разрешения и форматов представления. На основе критического анализа существующего инструментария был выбран механизм сверточных нейронных сетей для решения задач классификации, детектирования, сегментации и анализа произошедших изменений с малой дискретностью изменений. Автором работы предложена программная реализация выбранного способа распознавания, в том числе с решением задач моделирования объектов реального мира на картографической основе.

Ключевые слова: сверточные нейросети, распознавание образов, изображение.

R.V. Dzvinko

ON SOME APPROACHES TO MACHINE LEARNING
OF CONVOLUTIONAL NEURAL NETWORKS IN THE TASKS
OF PROCESSING AERIAL IMAGES OF VARIOUS PRESENTATION FORMATS

Abstract. The presented scientific paper presents one of the approaches to solving the problem of pattern recognition on an array of initial data of aerial photographs of various resolutions and presentation formats. Based on a critical analysis of the existing tools, a convolutional neural network mechanism was chosen to solve the problems of classification, detection, segmentation and analysis of changes with a small discreteness of changes. The author of the paper proposed a software implementation of the selected recognition method, including solving problems of modeling real-world objects on a cartographic basis.

Keywords: convolutional neural networks, pattern recognition, image.

Введение

В настоящий момент как в нашей стране, так и в мире в целом в рамках развития технологий диджитализации стремительно развивается отрасль цифровой идентификации физического мира. Данный тренд вызван необходимостью комплексного формирования взаимодействия в компьютерных аналитических системах на основе процедур с использованием цифровых образов.

Проблема идентификации различных предметов на цифровых описаниях весьма актуальна и востребована, особенно это касается сферы оценки и исследования изображений с аэрофотосъемок, так как до сих пор практически нет целостных машинных методов обработки изображений с применением различных способов деления для отождествления и установления местоположения объектов [1].

Для решения задач обработки изображений был разработан достаточно широкий спектр разнообразных математических алгоритмов – от элементарной граничной бина-

Дзвинко Роман Валерьевич

исполнительный директор научно-производственного центра «БизнесАвтоматика», Москва.
Сфера научных интересов: машинное обучение; распределенные вычисления и облачное хранение больших данных; повышение эффективности функционирования сложных информационных систем.
Электронный адрес: dzvinko@pba.su

ризации до использования алгоритмов Виолы – Джонса и диаграмм направленных градиентов [2; 3; 4].

Необходимо отметить, что сейчас все большую популярность в силу своих функциональных возможностей для решения подобных задач набирают различные типы нейронных сетей (например, многослойные и сверточные), которые созданы на теоретических основах разработанных в прошлом веке. Нейронные сети, реализованные в разных цифровых архитектурах, также могут решать такие задачи как нахождение разнотипных объектов при идентификации изображений, а также осуществлять их распознавание [5, 6].

Общеизвестно, что корнями нейронные сети уходят в такую науку, как биология, и по своей сути являются некими элементарными цифровыми аналогами мозга человека. Основа их функционирования построена на следующих принципах. Искусственная «нервная клетка» (математическая, цифровая) производит изменение направления сигналов со входа в сигнал выходной посредством скалярного умножения направления сигналов со входа на специфическое направление (вектор) называемое «вектором синаптических весов» этой цифровой клетки (нейрона) и дополнением к скалярному произведению показателя именуемого порогом подобного нейрона. После этого к полученному решению применяют своеобразную (по большей части нелинейную) функцию, называемую функцией активации нейрона.

Нейронная сеть многослойного характера

В нейронной сети многослойного характера различают следующие разновидности нейронов:

- функционирующие на входе, назначение которых заключается в фиксации определенных исходных данных с внешней стороны, к примеру, это могут быть характеристики какого-либо объекта;
- функционирующие на выходе, назначение которых заключается в возвращении устанавливаемых переменных, а именно данные об отнесении какого-либо объекта к определенному классу объектов;
- нейроны переходного (промежуточного) характера, назначение которых заключается в исполнении вспомогательных функций в нейронной сети.

По своей архитектуре стандартная многослойная нейронная сеть включает в себя от двух и более слоев различных разновидностей нейронов, упомянутых выше. При этом необходимо сказать, что от второго слоя и далее любой нейрон непосредственно соединен с определенными, либо со всеми нейронами прошлых слоев нейросети [7].

В свою очередь, такая разновидность, как сверточные нейронные сети, представляют собой логическое продолжение и совершенствование элементарного представителя нейронных сетей – многослойного персептрона. Данные сети применяются для обработки

входящей информации (например, цифрового изображения с аэрофотосъемки), обладающей определенной топологией сетчатого характера.

Функционал сверточных нейронных сетей построен на объемной (пространственной) зависимости сигнала. Основными их достоинствами в вопросах решения задач по распознаванию объектов на изображениях, являются:

- пространственное объединение;
- использование пространственно разделяемых весов.

Основной чертой рассматриваемых нейросетей является их большие функциональные возможности в вопросах исследования отличительных черт какой-либо информации, которые независимы (неизменны) от сдвига, то есть они представляют собой своеобразные фильтры необходимые для оптимальной обработки тех или иных изображений.

Их главным преимуществом над многослойными нейросетями является применение целостных общих весов в сверточных слоях, то есть для пикселя изображения в сверточном слое применяется один банк веса (фильтр) [1].

Непосредственно иллюстрация о том, что из себя представляет сверточная сеть, ее типичная архитектура приведена на Рисунке 1.

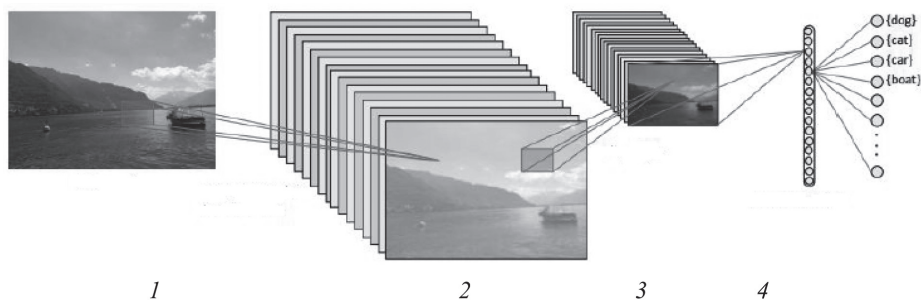


Рисунок 1. Архитектура сверточной сети для идентификации объектов: 1 – входные данные; 2 – сверточный слой + нелинейность; 3 – объединяющий слой; 4 – линейный классификатор

Стандартная сверточная сеть включает в себя определенное количество пунктов (этапов). Каждый из них включает набор 2-мерных цифровых массивов информации, именуемых картами объектов, которые в свою очередь представляют собой результат обработки одного фильтра к целому цифровому образу [8].

В целом же алгоритм установления и распознавания какого-либо объекта сводится к тому, что, во-первых, нужно определить характерные черты объекта, во-вторых, сопоставить эти черты (особенности) с ключевыми характеристиками, полученными из различных пространственных площадей общего изображения.

Состав любого цифрового образа является совокупностью множества пикселей в виде матрицы с определенными величинами цветовой гаммы (R (Red) – красный, G (Green – зеленый), B (Blue) – голубой), следовательно, обработка в данном случае тоже должна быть матричной.

Свертка каждой матрицы цифрового образа с различными матрицами фильтров отличается определенными особенностями. По функциональному алгоритму после каждой свертки создается слой. Затем после получения определенного комплекса функций реализуется процесс категоризации, результатом которого является получение функций, которые будут коррелировать с каким-либо слоем изучаемого объекта [9].

После краткого рассмотрения сути, функционала и возможностей нейронных сетей необходимо перейти непосредственно к реализации главной цели настоящей научной работы.

Комплекс задач обработки аэрофотоизображений, в том числе нахождения объектов на изображении, включает в себя, как правило, задачи классификации, детектирования, семантической сегментации и анализа изменений [10–12].

Для решения задач указанных задач применяются решения, построенные на сверточных нейронных сетях (далее – СНС).

Системы машинного обучения, использующие СНС, обладают хорошей устойчивостью к изменениям масштаба, смещениям, поворотам, смене ракурса и прочим искажениям изображения.

Задача классификации объектов определяет класс объекта на изображении (например, класс «жилой дом», «самолеты», «транспортное средство» и др.)

Задача детекции объектов – это задача, в рамках которой необходимо выделить несколько объектов на изображении посредством нахождения координат ограничивающих рамок объекта и классификации этих ограничивающих рамок из множества заранее известных классов.

Задача семантической сегментации предназначена для точной классификации и сегментации каждого объекта, представленного на изображении. Семантическая сегментация позволяет программе полностью интерпретировать обрабатываемую сцену. Задача семантической является более сложной, чем задача классификации изображений и поиска объектов, что обусловлено не только необходимостью определения классов объектов, но и выявления их структуры, правильного выделения частей объектов на изображении.

Задача анализа изменений предполагает отображение изменений при проведении сравнения двух изображений, сделанных в разные временные промежутки.

Входными данными для обучения являются параметры обучения, задаваемые пользователем, в которые входят:

- набор данных (изображения в разных форматах);
- нейросеть для проведения обучения (выбирается из существующих в программе или создается);
- количество эпох обучения;
- критерии остановки обучения;
- контрольная точка (при необходимости);
- другие дополнительные параметры (выставляются по умолчанию либо настраиваются пользователем в программе).

Выходными данными являются:

- контрольные точки обучения;
- значения метрик обучения и валидации (значение функций потерь и точность, которые доступны к отображению на экране);
- сообщения пользователю (доступны к отображению на экране).

Функция потерь определяет разницу между ожидаемым результатом (target) и результатом, полученным с помощью машинного обучения (prediction).

Для задач классификации могут быть использованы различные функции потерь, которые определяют качество обучения сети.

Функция потерь MAE – средняя абсолютная ошибка. Вычисляется модуль разницы между спрогнозированным значением (predict) и желаемым результатом (target), затем эти значения усредняются:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y'_i - y_i|,$$

где n – количество спрогнозированных значений; y'_i – спрогнозированное значение; y_i – желаемый результат.

Функция потерь MSE рассчитывает среднюю квадратичную ошибку между спрогнозированным значением и желаемым результатом:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y'_i - y_i)^2,$$

где n – количество предсказанных значений; y'_i – спрогнозированное значение; y_i – желаемый результат.

Кросс-энтропия (Cross Entropy) – логарифмическая функция потерь, вычисляющая разницу между двумя вероятностными распределениями.

Кросс-энтропия определяется как

$$H(P, Q) = - \sum P(x) \log Q(x),$$

где P – распределение истинных ответов; Q^x – распределение вероятностей спрогнозированных значений; x – спрогнозированные значения.

При обучении одношаговых СНС для задачи детекции используется функция потерь, состоящая из трех частей: ошибки локализации (detectionloss), ошибки классификации (classifierloss), ошибки уверенности (confidenceloss), и считается как их сумма.

В процессе обучения СНС для задачи семантической сегментации итоговая ошибка складывается из трех функций потерь: Jaccardloss, Diclloss, Cross-entropyloss.

Формула вычисления Jaccardloss

$$Jaccardloss = 1 - J(A, B),$$

где $J(A, B)$ – индекс Жаккара; A – предсказание нейросети; B – истинные значения из разметки.

Формула вычисления Diclloss

$$Jaccardloss = 1 - D(A, B),$$

где $D(A, B)$ – коэффициент Дайса; A – прогноз нейросети; B – истинные значения из разметки.

В процессе обучения СНС для задачи анализа изменений используется та же функция ошибки, что и для задачи семантической сегментации.

Для решения указанных классов задач разрабатываются специализированные программные изделия (далее – ПИ), выполняющие функции машинного обучения [13].

Как правило, такие ПИ предназначены для загрузки, редактирования, обучения и проверки нейронных сетей.

Основными функциями таких ПИ являются:

- создание моделей нейронных сетей;
- загрузка обученных моделей и проверенных архитектур;
- редактирование в графическом представлении архитектуры параметров нейросети;
- обучение нейронной сети с контролем в процессе обучения качества обучения;

О некоторых подходах к машинному обучению сверточных нейронных сетей...

- проверка загруженных и обученных моделей нейронных сетей.

ПИИ в качестве входных данных при редактировании и/или дообучении нейронной сети принимает файлы в форматах YAML, ONNX, PTN и PT. Данные форматы так же являются выходными для сохранения обученных нейросетевых моделей.

Входными данными при обучении являются изображения в форматах JPEG, TIFF, ECW и RSW и файлы разметки в форматах TXT и XML.

В качестве интерфейса взаимодействия «человек-машина» при работе с такими ПИИ используется графический интерфейс, представляющий собой набор информационных и функциональных панелей (страниц, вкладок, окон), сгруппированных по тематическим признакам.

На информационных панелях располагаются сведения, необходимые для управления вычислительными процессами и анализа результатов их выполнения.

На функциональных панелях располагаются функциональные элементы, позволяющие осуществлять настройку условий запуска вычислительных процессов и управлять ходом их выполнения.

Логическая структура ПИИ функционально состоит из следующим модулей и компонент (см. Рисунок 2):

- модуль авторизации;
- модуль создания и отображения архитектуры нейронной сети;
- компонент создания и удаления слоев;
- компонент отображения архитектуры;
- компонент создания и удаления связей между слоями;
- модуль отображения информации:
- компонент вывода графических данных;
- компонент отображения параметров сети;
- компонент отображения процесса обучения;
- модуль проверки и обучения нейронной сети:
- компонент проверки архитектуры;
- компонент проверки параметров;
- компонент обучения;
- модуль загрузки и сохранения данных;
- модуль распределения ресурсов.

Взаимодействие программных модулей осуществляется с использованием интерфейса Web API.

В модуле создания и отображения архитектуры нейронных сетей реализуется пользовательский интерфейс. Модуль обеспечивает выполнение следующих функций:

- поиск по названиям существующих нейронных сетей;
- визуализация описания нейронной сети;
- создание нейронной сети;
- клонирование существующей нейронной сети;
- удаление нейронной сети;
- экспортирование нейронной сети;
- импортирование нейронной сети;
- визуализация пользовательского описания нейронной сети в виде комментария;
- редактирование нейронной сети в графическом представлении;

- установка параметров слоёв и блоков;
- группировка слоев и блоков;
- проверка нейронной сети на возможность обучения;
- сохранение нейронной сети;
- очистка рабочего поля.

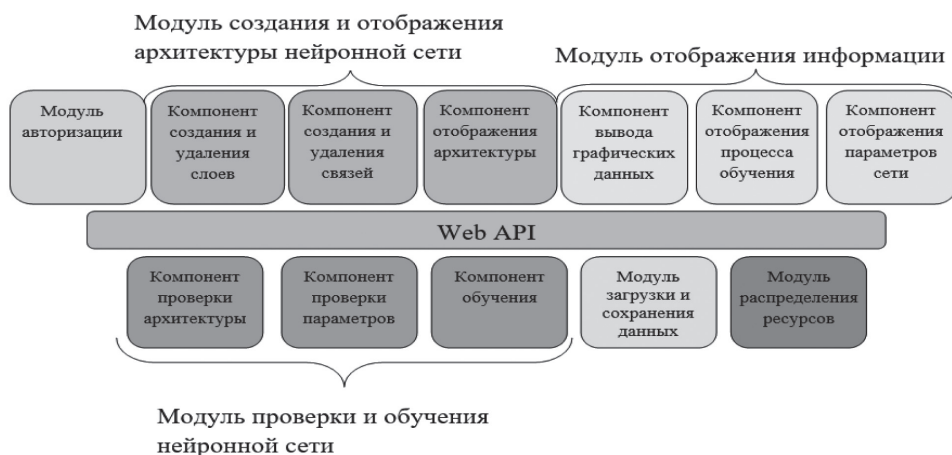


Рисунок 2. Логическая структура программного изделия обучения нейросети

Модуль загрузки и сохранения данных предназначен для выполнения следующих функций:

- поиск набора данных;
- визуализация оценки сбалансированности классов;
- визуализация выбранных изображений из набора данных;
- визуализация параметров выбранного набора данных;
- визуализация типа задачи набора данных;
- импортирование набора данных в виде архива;
- изменение названия набора данных.

В модуле обучения и проверки нейронных сетей реализуется основной ресурсоемкий процесс обучения нейронной сети. Модуль проверки и обучения нейронной сети размещается, как правило, на сервере и выполняет следующие функции:

- выбор набора данных для обучения;
- выбор нейронной сети для обучения;
- настройка параметров обучения;
- выбор контрольной точки (при наличии; необязательный параметр);
- запуск обучения;
- сохранение контрольных точек и обученной модели.
- визуализация процесса обучения в графическом представлении и отправка сообщений пользователю.

Модуль распределения ресурсов выполняет следующие функции:

- визуализация загруженности устройств (графического процессора, оперативного запоминающего устройства, постоянного запоминающего устройства);
- визуализация количества активных пользователей;

- визуализация количества запущенных задач на графическом и центральном процессорах;
- визуализация прогресса процесса обучения;
- визуализация занятых графических ускорителей;
- визуализация статуса занятости центрального процессора.

Заключение

Таким образом, резюмируя результаты настоящей научной работы, необходимо отметить:

- обоснована актуальность направления исследования, приведена суть, особенности, разновидности и возможности различных нейросетей (многослойных и сверточных);
- предложена программная реализация выбранного автором способа распознавания аэрофотографий различных форматов представления, на основе решения задач моделирования объектов реального мира на картографической основе.

Литература

1. *Нуеун В.Ч.* Разработка алгоритмов распознавания объектов воздушной съемки на основе сверточных нейронных сетей с иерархическим классификатором: дис. ... канд. техн. наук: 05.13.01. М., 2022. 134 с.
2. *Qiang Chen, Quan-sen Sun, Pheng Ann Heng, De-shen Xia* (2008) A double-threshold image binarization method based on edge detector. *Pattern Recognition*, 41:4, 1254–1267.
3. *Pitkanen J.* (2001) Individual tree detection in digital aerial images by combining locally adaptive binarization and local maxima methods. *Canadian Journal of forest research*, 31:5, 832–844.
4. *Rosebrock A.* (2015) Histogram of Oriented Gradients and Object Detection (Retrieved August 31, 2015). Available at: <http://www.pyimagesearch.com/2014/11/10/histogram-oriented-gradients-object-detection/> (date of the application: 01.12.2022).
5. *Hinton G.E., Srivastava N.* (2012) Improving neural networks by preventing co-adaptation of feature detectors, 2012. Available at: <https://arxiv.org/abs/1207.0580> (date of the application: 01.12.2022).
6. *Cheng G., Zhou P., Han J.* (2016) Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 54:12, 7405–7415.
7. *Ranka S., Mohan C., Mehrotra K., Menon A.* (1996) Characterization of a class of sigmoid functions with applications to neural networks. *Neural Networks*, vol. 9, pp. 819–835.
8. *Каунг М.С., Абрамов Ю.А., Гинзгеймер С.А.* Сверточные нейронные сети в задачах глубокого обучения // Международный студенческий научный вестник. 2018. № 5. URL: <https://eduherald.ru/ru/article/view?id=18951> (дата обращения: 01.12.2022).
9. *Кыонг Н.Т., Сырямкин В.И., Тхуи Н.Ч.Х.* Модель метода распознавания объектов на изображениях с использованием сверточной нейронной сети CNN // Современные наукоемкие технологии. 2020. № 12. С. 269–280.
10. *Боровская Е.В., Давыдова Н.А.* Основы искусственного интеллекта: учеб. пособие. М.: Лаборатория знаний, 2020. 130 с.
11. *Остроух А.В.* Введение в искусственный интеллект: монография. Красноярск: Научно-инновационный центр, 2020. 250 с.
12. *Родзин С.И.* Системы искусственного интеллекта: учеб. пособие. Таганрог: ИКТИБ ЮФУ, 2015. 177 с.

13. Спири́н И.А. Разработка информационного обеспечения автоматизированной системы обнаружения объектов на изображении // Молодой ученый. 2017. № 4 (138). С. 61–64.

References

1. Nguen V.Ch. (2022) *Razrabotka algoritmov raspoznavaniya obektov vozduшной s#emki na osnove svertochnyh nejronnyh setej s ierarhicheskim klassifikatorom: dis. ... kand. tekhn. nauk* [Development of aerial object recognition algorithms based on convolutional neural networks with a hierarchical classifier]. Moscow, 134 p. (in Russian).
2. Qiang Chen, Quan-sen Sun, Pheng Ann Heng, De-shen Xia (2008) A double-threshold image binarization method based on edge detector. *Pattern Recognition*, 41:4, 1254–1267.
3. Pitkanen J. (2001) Individual tree detection in digital aerial images by combining locally adaptive binarization and local maxima methods. *Canadian Journal of forest research*, 31:5, 832–844.
4. Rosebrock A. (2015) Histogram of Oriented Gradients and Object Detection (Retrieved August 31, 2015). Available at: <http://www.pyimagesearch.com/2014/11/10/histogramoriented-gradients-object-detection> (date of the application: 01.12.2022).
5. Hinton G.E., Srivastava N. (2012) Improving neural networks by preventing co-adaptation of feature detectors, 2012. Available at: <https://arxiv.org/abs/1207.0580> (date of the application: 01.12.2022).
6. Cheng G., Zhou P., Han J. (2016) Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 54:12, 7405–7415.
7. Ranka S., Mohan C., Mehrotra K., Menon A. (1996) Characterization of a class of sigmoid functions with applications to neural networks. *Neural Networks*, vol. 9, pp. 819–835.
8. Kaung M.S., Abramov Yu.A., Ginzgejmer S.A. (2018) *Svertochnye nejronnye seti v zadachah glubokogo obucheniya* [Convolutional Neural Networks in Deep Learning Problems]. *Mezhdunarodnyj Studencheskij Nauchnyj Vestnik*, No. 5. Available at: <https://eduherald.ru/ru/article/view?id=18951> (date of the application: 01.12.2022) (in Russian).
9. Kyong N.T., Syryamkin V.I., Thui N.Ch.h. (2020) *Model'metoda raspoznavaniya obektov na izobrazheniyah s ispol'zovaniem «svertochnoj nejronnojseti – CNN* [Model of the method for recognizing objects in images using a “convolutional neural network - CNN]. *Sovremennye naukoemkie tekhnologii*, No. 12, pp. 269–280 (in Russian).
10. Borovskaya E.V., Davydova N.A. (2020) *Osnovy iskusstvennogo intellekta* [Fundamentals of artificial intelligence]. Moscow: Laboratoriya znaniy Publishing, 130 p. (in Russian).
11. Ostrouh A.V. (2020) *Vvedenie v iskusstvennyjintellekt* [Introduction to artificial intelligence]. Krasnoyarsk: Nauchno-innovacionnyjcentr, 250 p. (in Russian).
12. Rodzin S.I. (2015) *Sistemy iskusstvennogo intellekta* [Artificial intelligence systems]. Taganrog: IK-TIB YUFU, 177 p. (in Russian).
13. Spirin I.A. (2017) *Razrabotka informacionnogo obespecheniya avtomatizirovannoj sistemy obnaruzheniya ob#ektov na izobrazhenii* [Development of information support for an automated system for detecting objects in an image]. *Molodoy Uchenyj*, No. 4 (138), pp. 61–64 (in Russian).