

Модель автоматической классификации и локализации образов

Л. В. Серебряная, к. т. н., доцент, доцент кафедры ПОИТ

E-mail: l_silver@mail.ru

УО «Белорусский государственный университет информатики и радиоэлектроники», ул. П. Бровки, д. 6, 220013, г. Минск, Республика Беларусь

К. Ю. Бочкарев, магистрант кафедры ИТАС

E-mail: axe777@inbox.ru

УО «Белорусский государственный университет информатики и радиоэлектроники», ул. П. Бровки, д. 6, 220013, г. Минск, Республика Беларусь

А. Я. Попитич, магистр технических наук

E-mail: sasha.popitich@outlook.com

Аннотация. Работа посвящена идентификации образов на изображениях, которая выполняется в результате процедур классификации и локализации. Анализ моделей, методов и алгоритмов показал, что для решения поставленной задачи предпочтительно применять машинное обучение, искусственную нейронную сеть и генетический алгоритм. Предложена архитектура сверточной искусственной нейронной сети, позволяющая решать как задачу классификации, так и задачу локализации образов. Сначала сеть обучается, затем для изображения, подаваемого на ее вход, определяется класс. На заключительном этапе работы сверточной нейронной сети выполняется локализация объектов на изображении. Для этого анализируются выходные значения предпоследнего слоя модели, после чего происходит обход слоев в обратном порядке. Его цель – нахождение на исходном изображении регионов с наибольшим откликом. Комбинированная модель показала приемлемые результаты как по классификации, так и по локализации объектов. Все параметры для работы сети определяются автоматически с помощью генетического алгоритма. Дальнейшее улучшение работы предложенной модели связано с реализацией на ней распределенных вычислений.

Ключевые слова: идентификация; классификация; локализация; модель искусственной нейронной сети; генетический алгоритм

Для цитирования: Серебряная, Л. В. Модель автоматической классификации и локализации образов / Л. В. Серебряная, К. Ю. Бочкарев, А. Я. Попитич // Цифровая трансформация. – 2019. – № 1 (6). – С. 43–48. <https://doi.org/10.38086/2522-9613-2019-1-43-48>



© Цифровая трансформация, 2019

Model of Automatic Classification and Localization of Images

L. V. Serebryanaya, Candidate of Science (Technology), Associate Professor, Associate Professor of the Department "Software of Information Technologies"

E-mail: l_silver@mail.ru

Belarusian State University of Informatics and Radioelectronics, 6 P. Brovka Str., 220013 Minsk, Republic of Belarus

K. Y. Bochkarev, Undergraduate Student of the Department "Information Technologies of Automated Systems"

E-mail: axe777@inbox.ru

Belarusian State University of Informatics and Radioelectronics, 6 P. Brovka Str., 220013 Minsk, Republic of Belarus

A. Y. Popitich, Master of Technical Sciences

E-mail: sasha.popitich@outlook.com

Abstract. The work is devoted to the identification of images in pictures, which is performed as a result of the classification and localization procedures. Analysis of models, methods and algorithms has shown that for solving the set task it is preferable to use machine learning, an artificial neural network and a genetic algorithm. The architecture of a convolutional artificial neural network is proposed. It can solve both the problem of classification and the problem of localizing images. First the network is trained, then a class is determined for the image fed to its input. Objects are localized in the image at the final stage of operations of the convolutional neural network. For this, the output values of the penultimate layer of the model are analyzed, after which the layers are traversed in the reverse order. Its goal is to find the regions with the highest response on the source image. The combined model showed acceptable results both in classification and in localization of objects. All parameters for the network are determined automatically using a genetic algorithm. Further improvement of the proposed model results will be performed by implementing distributed computing on it.

Key words: identification; classification; localization; model of artificial neural network; genetic algorithm

For citation: Serebryanaya L. V., Bochkarev K. Y., Popitich A. Y. Model of Automatic Classification and Localization of Images. *Cifrovaja transformacija* [Digital transformation], 2019, 1 (6), pp. 43–48 (in Russian). <https://doi.org/10.38086/2522-9613-2019-1-43-48>

© Digital Transformation, 2019

Введение. Настоящая работа посвящена решению задачи идентификации заданных образов, которая является результатом классификации объектов с последующей их локализацией на основе визуальной информации.

Решаемая задача напрямую связана с машинным обучением и компьютерным зрением. Они получили широкое применение в различных прикладных областях, чему поспособствовало развитие сети интернет и накопление больших баз данных, на основании которых формируются обучающие выборки. Именно на них построены многие актуальные алгоритмы и приложения.

Основными моделями классификации являются: логистическая регрессия, дерево принятия решений, случайный лес, искусственные нейронные сети (ИНС).

На выходе модели в виде логистической регрессии получается значение статистической вероятности того или иного события, вследствие чего большинство задач логистической регрессии сводится к определению принадлежности образа к одному из двух классов.

Деревья принятия решений, как и случайный лес, способны разделять объекты на большое количество классов и являются достаточно устойчивыми к пропускам признаков. Однако проблемой для таких моделей является процедура переобучения, поскольку деревья решений часто получают очень большой высоты. Для оптимизации и предотвращения переобучения должны применяться достаточно сложные алгоритмы. Сама задача построения оптимального дерева решений является NP-полной задачей.

ИНС, особенно сверточные, являются предпочтительной моделью для классификации. Они способны комбинировать в себе сразу две задачи: классификацию и локализацию образов. Это

подтверждается многочисленными примерами архитектур программных средств, созданных на основе сверточных ИНС. Выбор архитектуры нейронной сети является нетривиальной задачей, связанной с анализом большого объема данных. Часто параметры и архитектура сети подбираются экспериментальным путем, что является трудоемким процессом. Это обусловлено тем, что каждая задача имеет свои уникальные особенности: данные, ожидаемый результат, необходимую обобщающую способность.

Для автоматизации построения сети и поиска оптимального решения на ней могут применяться генетические алгоритмы. Их цель — определить в ходе эволюции архитектуру и параметры ИНС, обеспечивающие успешное решение поставленной задачи.

Модель сверточной нейронной сети для классификации образов. Поиск количества слоев в сети и нейронов в них осуществляется при помощи генетического алгоритма, поэтому на этапе моделирования необходимо лишь определить последовательность и типы слоев в ИНС. Исходя из особенностей сверточной ИНС выбираются следующие типы слоев: сверточный; уменьшения пространства признаков; нормализующий; полносвязный [1].

Архитектура сверточной ИНС приведена на рис. 1. Она состоит из чередующихся сверточных (C-Layer) и субдискретизирующих слоев (S-Layer), а также одного полносвязного выходного слоя. Входной слой часто не учитывается в архитектуре ИНС. Приведенная сеть характеризуется тремя основными принципами: локальным восприятием признаков; разделяемыми весовыми коэффициентами; субдискретизацией.

Локальное восприятие признаков означает, что на вход отдельных нейронов подается не

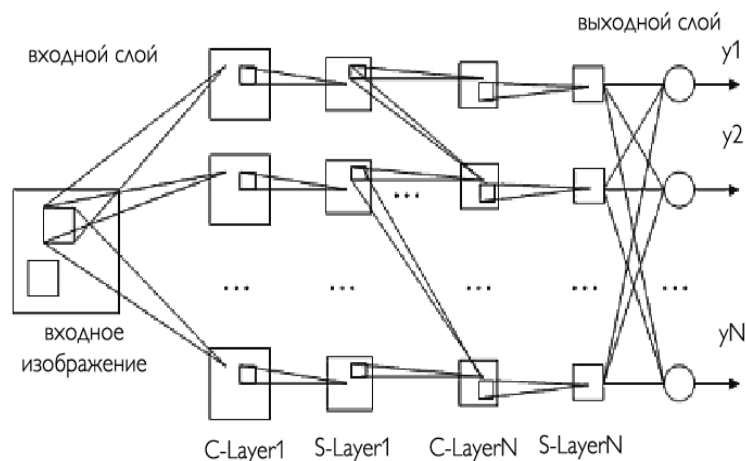


Рис. 1. Архитектура сверточной ИНС
 Fig. 1. Architecture of convolutional ANN

все изображение, а только какая-то его часть. Это позволяет сохранить структуру (топологию) изображения, переходя от слоя к слою, чего нельзя сказать о других видах ИНС, где в большинстве случаев целостность анализируемого изображения теряется вследствие его представления в виде массива данных, поступающих на входной слой.

Принцип разделения весовых коэффициентов предполагает, что один их набор может быть использован для большого количества связей. Тогда, если на вход поступает изображение размера 32×32 пикселя, то каждый из нейронов следующего слоя получит на вход только небольшой участок изображения (например, 5×5 пикселей), причем каждый из фрагментов будет обработан одним и тем же набором весовых коэффициентов. Такие наборы называют ядрами. В результате появляется основное преимущество сверточной ИНС — небольшое количество связей. Например, 10 ядер размера 5×5 пикселей для входного изображения 32×32 образуют примерно 256 000 связей. ИНС в виде многослойного персептрона для таких же исходных данных имеет 10 миллионов связей. Кроме этого, наличие ядер и уменьшение количества связей улучшает обобщающую способность сети, что дает большую эффективность при выделении инвариантов на изображении, игнорируя при этом шум.

Каждый фрагмент изображения умножается на ядро, после чего произведения суммируются. Данную операцию можно рассматривать как своеобразный фильтр, основным свойством которого является прямая зависимость между степенью схожести части изображения и ядра. В результате применения ядра ко всем составляющим изображения будет получена карта признаков.

Под субдискретизацией понимается уменьшение размерности анализируемых признаков путем усреднения построенной карты признаков. В основном субдискретизация необходима для обеспечения инвариантности признаков данных к масштабу ИНС.

Завершающим этапом анализа является преобразование карты признаков в вектор или в число. Для этого признаки обязательно попадают на полносвязный слой [2].

Для определения качества распознавания сети применяется функция среднеквадратичной ошибки:

$$E^p = 1/2 (D^p - O(I^p, W))^2, \quad (1)$$

где E — ошибка распознавания для p -й обучающей пары, D — желаемый выход сети, а O — реальный выход сети, зависящий от входа I и веса W .

Основной задачей обучения сети является настройка весовых коэффициентов W таким образом, чтобы ошибка распознавания E была минимальной. Для ее минимизации применяется градиентный метод.

Анализ рассмотренных сверточных ИНС позволил внести некоторые улучшения в архитектуру создаваемой нейронной сети. Изменения заключаются в уменьшении пространства признаков на каждом слое путем уменьшения количества связей между нейронами. Необходимо увеличить число слоев ИНС, компенсировав это уменьшением количества измерений на каждом слое для каждой карты признаков в сверточной ИНС. Для этого применяется слой, уменьшающий пространство признаков.

Модель локализации объектов на изображениях. Поскольку задачи классификации и ло-

кализации образов тесно связаны между собой, было решено использовать сверточную ИНС, чтобы получать решение этих двух задач в рамках одной модели [3].

Известные сверточные ИНС, представляющие собой как детектор объектов, так и классификатор, теряют способность локализации объектов на изображении при использовании полносвязного слоя для классификации, поэтому в ряде сверточных ИНС полносвязный слой заменен слоем усреднения с уменьшением размерности. В результате существенно уменьшается количество параметров сети на последнем слое, что положительно сказывается на скорости и качестве обучения [4]. Слой усреднения позволяет сохранить детектирующие способности сети вплоть до последнего слоя, целенаправленно не обучая сеть задаче обнаружения объектов. Чтобы определить на изображении регионы, в которых расположен объект, необходимо дополнительно построить активационную карту соответствия.

Процедура генерации активационной карты основана на использовании весовых коэффициентов связей последнего слоя и слоя усреднения. Весовые коэффициенты определяют важность того или иного признака для конкретного класса. Слой усреднения показывает отклик каждого признака.

Рассмотрим построение активационной карты. Пусть $f_k(x, y)$ отвечает за значение активационной функции для модуля k в последнем сверточном слое для пространственных координат (x, y) . Тогда для модуля k результат вычисления глобального усреднения равен $F^k = \sum_k f_k(x, y)$. Следовательно, для класса C этого модуля входное значение для выходного слоя равно $S_c = \sum_k w_k^c F_k$, где w_k^c — весовой коэффициент, соответствующий классу C модуля k . Значение этого веса указывает на важность F^k для класса C .

Подстановка $F_k = \sum_{x,y} f_k(x, y)$ в S_c приводит к следующему выражению:

$$S_c = \sum_k w_k^c \sum_{x,y} f_k(x, y) = \sum_{x,y,k} w_k^c f_k(x, y). \quad (2)$$

Обозначим активационную карту M для класса C следующим выражением:

$$M_c(x, y) = \sum_k w_k^c f_k(x, y). \quad (3)$$

Таким образом, $S_c = \sum_{x,y} M_c(x, y)$ и, следовательно, $M_c(x, y)$ определяют важность конкретного пикселя на изображении для класса C .

Поскольку блоки признаков последнего сверточного слоя ИНС являются своеобразным

фильтром, можно выполнить обратное преобразование, чтобы узнать координаты пикселя на исходном изображении, а затем определить важность данного пикселя на нем.

В ряде работ, посвященных сверточным ИНС, показано, что за последним сверточным слоем следует размещать слой усреднения, а не слой определения максимального значения, так как последний позволяет определять лишь один регион с максимальным откликом. В тех ситуациях, когда требуется локализовать составные области или комбинацию из нескольких похожих объектов, слой усреднения дает лучшие результаты.

Для обоснования применения подобной модели в рамках данной работы были проведены эксперименты на нескольких сверточных ИНС, в которых полносвязные слои были заменены слоем усреднения. Это также позволило на 90 % уменьшить количество параметров, однако ухудшило классификационную способность. При этом резко улучшается качество локализации изображений сетью, если последний сверточный слой имеет высокое разрешение. В ряде ИНС разрешение последнего сверточного слоя увеличивалось при удалении нескольких предыдущих слоев. Тестирование модифицированных сетей показало, что в среднем их классификационная способность снизилась на 1–2 %, но при этом существенно улучшилась способность локализации объектов. Для проверки работы сетей была выбрана обучающая коллекция изображений, для которой ошибка локализации исходной ИНС составляла в среднем 30 %. Измененные ИНС выполнили локализацию образов для этой же выборки с ошибкой 12–15 % [5], поэтому было решено выбрать оптимизированную модель сверточной нейронной сети, способную решать сразу две задачи: классификации и локализации объектов на изображении.

Параметры сети будут определяться генетическим алгоритмом, а способность локализации обеспечиваться специально вычисленной картой активации для последнего слоя ИНС [6–8]. Это позволит сократить время обучения, так как обучать придется только саму нейронную сеть без дополнительных детекторов объектов.

Комбинированная модель классификации и локализации. Работа модели начинается с определения класса изображения, подаваемого на ее вход. Для этого применяется заранее обученная сверточная ИНС. После того как был определен класс, появляется возможность локализации объектов на изображении. Этот процесс также осуществляется с использованием нейронной сети путем

анализа выходных значений предпоследнего слоя усреднения и последующего обхода слоев в обратном порядке для определения регионов с наибольшим откликом на исходном изображении.

С целью автоматизации процесса обучения нейронной сети в модель включен генетический алгоритм, который автоматически подбирает архитектуру ИНС в ходе эволюционного естественного отбора. Эффективность применения генетического алгоритма напрямую зависит от реализации способа кодирования архитектуры нейронной сети в виде хромосомы, эффективности реализации метода кроссовера, мутации и вычисления фитнес-функции.

Анализ существующих архитектур сетей показал, что для ИНС с большим количеством нейронов не рекомендуется выполнять кодирование связей и весовых коэффициентов нейронов, так как получаемая сеть оказывается неэффективной и не соответствующей обучающим данным. В связи с этим кодированию будут подвергаться отдельные слои и их конфигурации. В рамках решаемой задачи для сети были определены четыре типа слоев: активационный, сверточный, полносвязный и слой усреднения.

Активационный слой не имеет параметров, поэтому его можно кодировать числовым значением. Сверточный слой имеет два параметра: первый указывает на количество фильтров для сверточного слоя, а второй — на размер этих фильтров. Кодировать данный слой необходимо тремя значениями. Полносвязный слой имеет один параметр, определяющий количество нейронов в слое, следовательно, его кодирование выполняется двумя значениями. Слой усреднения имеет один параметр — размер блока усреднения, значит его кодирование происходит двумя значениями.

После завершения прямого и обратного проходов по ИНС для определенного обучающего объекта на примере классификации выполняется проверка того, достигнута ли требуемая точность

обучения или нет. В случае положительного ответа алгоритм завершается и происходит переход к процедуре локализации, иначе выбирается новый обучающий объект и для него повторяется процесс обучения сети.

Чтобы определить координаты классифицированного объекта на исходном изображении, необходимо придерживаться нескольких правил.

1. При составлении карты признаков для нейронов первого сверточного слоя необходимо запоминать координаты исходного изображения.

2. Для последующих карт признаков необходимо запоминать родительскую карту признаков, чтобы была возможность восстановить координаты на исходном изображении.

3. Используя предпоследний слой нейронной сети, необходимо вычислить активационную карту, которая будет показывать степень вовлеченности конкретного нейрона в конечный результат классификации. Чем выше степень вовлеченности нейрона, тем больше вероятность того, что нейрон и соответствующие ему пиксели находятся на объекте.

Зависимость карт признаков при построении сети схематически изображена на рис. 2.

На рис. 2 показано, из каких частей состоит активационная карта. Связи активационных карт в дальнейшем используются в алгоритме определения координат заданных объектов на изображении. Первым шагом этого алгоритма является нахождение весовых коэффициентов нейронов последнего слоя, которые необходимы для вычисления активационных карт. Затем для каждого класса с помощью выражения (3) выполняется построение активационной карты. Все они заносятся в список для последующего использования в алгоритме. Далее осуществляется проход по активационным слоям с целью определения множества пикселей, которые больше других участвуют в определении класса объекта. Эти действия повторяются для каждого из классов.

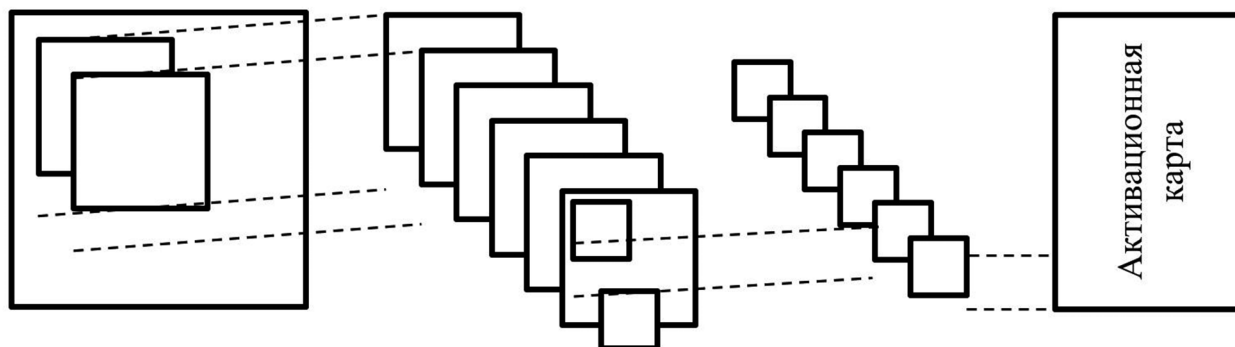


Рис. 2. Зависимость нейронов при построении карты признаков
Fig. 2. The dependence of neurons in the construction of the map of signs

Заключение. Тестирование предложенной модели позволяет сделать следующие выводы. Комбинированная модель показала приемлемые результаты как по классификации, так и по локализации объектов. Все параметры нейронной сети определяются автоматически в результате применения генетических операторов в ходе эволюционного процесса. Становится ненужным подбирать вручную данные для обучения и параметры решаемой задачи.

Главным недостатком модели является необходимость большого количества вычислитель-

ных ресурсов, поскольку каждая особь в популяции представляет собой нейронную сеть, которую требуется обучить. С учетом того, что в ходе работы генетического алгоритма особей необходимо обучать много раз, время выполнения алгоритма и необходимое количество вычислительных ресурсов существенно возрастают. Однако генетические алгоритмы несложно преобразуются для реализации распределенных вычислений. Это позволяет в разы уменьшить время работы и обучения всех сетей, так как в этом случае каждая ИНС обучается независимо от других на отдельном узле.

Список литературы

1. Radcliffe, N. J. Genetic set recombination and its application to neural network topology optimization. Technical Report EPCC-TR-91-21 / N. J. Radcliffe. – Edinburgh: University of Edinburgh, 1991. – 250 p.
2. Stanley, K. O. Evolving Neural Topologies through Augmenting Topologies / K. O. Stanley, R. Miikkulainen // Evolutionary Computation. The MIT Press. – 2002. – Vol. 10 (2). – PP. 99–127.
3. Simonyan, K. Deep inside convolutional networks: Visualising image classification models and saliency maps [Electronic resource] / K. Simonyan, A. Vedaldi // International Conference on Learning Representations Workshop. – 2014. – Mode of access: <https://arxiv.org/pdf/1312.6034.pdf>. – Date of access: 16.03.2019.
4. Perez, S. Apply genetic algorithm to the learning phase of a neural network [Electronic resource] / S. Perez // Department of Mechanical and Aerospace Engineering University of California. – Irvine, 2005. – Mode of access: <https://pdfs.semanticscholar.org/cc48/1cf3f2dfa88fc5fa84cd41d7e9f7f7de4ff2.pdf>. – Date of access: 16.03.2019.
5. Zhou, B. Learning Deep Features for Discriminative Localization [Electronic resource] / B. Zhou, A. Lapedriza // Computer Science and Artificial Intelligence Laboratory. – MIT, 2014. – Mode of access: <https://arxiv.org/pdf/1310.1531.pdf>. – Date of access: 16.03.2019.
6. Donahue, J. Decaf: A deep convolutional activation feature for generic visual recognition [Electronic resource] / J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, T. Darrell // International Conference on Machine Learning. – 2014. – Mode of access: https://web.njit.edu/~usman/courses/cs698_spring18/RCNN.pdf. – Date of access: 16.03.2019.
7. Girshick, R. Rich feature hierarchies for accurate object detection and semantic segmentation [Electronic resource] / R. Girshick, J. Donahue, T. Darrell, J. Malik. – CVPR, 2014. – Mode of access: <https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>. – Date of access: 16.03.2019.
8. Krizhevsky, A. Imagenet classification with deep convolutional neural networks / A. Krizhevsky, I. Sutskever, G. E. Hinton // Advances in Neural Information Processing Systems. – 2012. – PP. 1097–1105.

References

1. Radcliffe, N. J. Genetic set recombination and its application to neural network topology optimization. Technical Report EPCC-TR-91-21. Edinburgh: University of Edinburgh, 1991. 250 p.
2. Stanley K. O. Miikkulainen R. Evolving Neural Topologies through Augmenting Topologies. Evolutionary Computation. The MIT Press, 2002, Vol. 10 (2), pp. 99–127.
3. Simonyan K., Vedaldi A. Deep inside convolutional networks: Visualising image classification models and saliency maps. International Conference on Learning Representations Workshop. Available at: <https://arxiv.org/pdf/1312.6034.pdf> (accessed: 16.03.2019).
4. Perez S. Apply genetic algorithm to the learning phase of a neural network. Department of Mechanical and Aerospace Engineering University of California, Irvine, 2005. Available at: <https://pdfs.semanticscholar.org/cc48/1cf3f2dfa88fc5fa84cd41d7e9f7f7de4ff2.pdf> (accessed: 16.03.2019).
5. Zhou. B., Lapedriza A. Learning Deep Features for Discriminative Localization. Computer Science and Artificial Intelligence Laboratory, MIT, 2014. Available at: <https://arxiv.org/pdf/1310.1531.pdf> (accessed: 16.03.2019).
6. Donahue J., Jia Y., Vinyals O., Hoffman J., Zhang N., Tzeng E., Darrell T. Decaf: A deep convolutional activation feature for generic visual recognition. International Conference on Machine Learning, 2014. Available at: https://web.njit.edu/~usman/courses/cs698_spring18/RCNN.pdf (accessed: 16.03.2019).
7. Girshick R., Donahue J., Darrell T., Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation, CVPR, 2014. Available at: <https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf> (accessed: 16.03.2019).
8. Krizhevsky A., Sutskever I., Hinton G. E. Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.

Received: 26.02.2019

Поступила: 26.02.2019