

DOI: 10.15593/2499-9873/2018.3.06

УДК 004.852

И.В. Каракулов, А.В. Ключев

Пермский национальный исследовательский
политехнический университет, Пермь, Россия

ВЛИЯНИЕ РАЗМЕРОВ ОБЪЕКТА ОБУЧАЮЩЕЙ ВЫБОРКИ НА КАЧЕСТВО СЕГМЕНТАЦИИ МЕТОДАМИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Описывается опыт работы со сверточными нейронными сетями. Проводится детектирование объектов определенного размера при помощи сверточных нейронных сетей, с использованием обучающих множеств, составленных определенным образом. Результатом проведенного исследования являются рекомендации для построения обучающего множества, которые учитывают размеры объектов, детектируемых нейронной сетью.

Ключевые слова: нейронная сеть, детектирование объектов, обучающее множество, точность детектирования объектов.

I.V. Karakulov, A.V. Kliuev

Perm National Research Polytechnic University, Perm, Russian Federation

TRAINING SAMPLE OBJECT SIZE INFLUENCE UPON THE ARTIFICIAL INTELLIGENCE METHODS SEGMENTATION QUALITY

The experience with convolutional neural networks is described. Detection of objects of a certain size using convolutional neural networks is carried out using training sets compiled in a certain way. The result of the work is recommendations for constructing a learning set that will take into account the sizes of objects detected by the neural network.

Keywords: neural network, object detection, training set, accuracy of object detection.

Введение

Нейронные сети повсеместно используются в жизни человека, решая задачи детектирования и классификации объектов, прогнозирования событий [1]. Для каждой задачи требуется выбрать свою конфигурацию нейронной сети, которая зависит от различных факторов, таких как время работы сети, максимальная точность, объем используемой памяти при обучении и прямом прогоне сети, детектирование

объектов определенного размера. Но также существуют параметры, которые подбираются в процессе обучения сети, такие как количество итерации при обучении, параметры для изменения функции обратного распространения ошибки, поиск и разметка данных для обучения нейронной сети. Если с выбором нейронной сети и подбором параметров для обучения можно определиться, изучив накопленный опыт в данной предметной области, то с выбором обучающего множества все гораздо сложнее. Четких рекомендаций, в которых содержится описание обучающего множества для определенного типа сетей, нет. Обучающее множество влияет как на точность классификации объекта, так и на размеры детектируемых объектов.

Детектирующие сети SSD и Faster R-CNN лучше детектируют объекты большого размера, при уменьшении размеров объектов точность детектирования падает [2]. Например, на рис. 1 объекты (люди), имеющие больший размер, детектируются лучше, чем объекты меньшего масштаба.



Рис. 1. Пример изображения для детектирования людей

Цель работы заключается в составлении рекомендаций для построения обучающего множества, которое будет учитывать размеры объектов, детектируемых сетью. Для задачи сегментации объектов

были проанализированы работы, в которых проанализированы структуры нейронных сетей Faster R-CNN, SSD 300, SSD 500 [3, 4]. В этих работах рассматривают только структуру нейронных сетей, их основные параметры и точность обучения на различных обучающих множествах, что не позволяет сделать вывод о качестве детектирования объектов определенного размера.

Во многих работах отмечается, что обучающее множество можно формировать различными способами. Например, в статье [5] рассмотрено формирование нечетких обучающих множеств для обучения нейронных сетей, а также исследованы особенности обучения нейронных сетей при решении задачи сжатия данных без потерь. В статье [6] говорится о способе подбора обучающего множества для конкретной задачи и описан метод выбора обучающего множества только для одной нейронной сети.

Во многих работах подчеркивается, что повышение точности детектирования объектов на изображении остается одной из основных задач, поставленных перед сегментирующими нейронными сетями. Например, в статье [7] проводится анализ модели детектирования визуальных объектов в сверточных нейронных сетях. Применяется способ настройки модели, который определяет расположение объекта в пространстве для конкретной обучающей базы, используя метод кластеризации k-средних. Показывается, что данный метод позволяет повысить точность детектирования объектов на изображении.

Как показал анализ работ, при составлении обучающих множеств для каждой задачи подбирают специальный набор данных, что не позволяет использовать такой же алгоритм для составления обучающих множеств в другой предметной области. Поэтому исследования по влиянию размеров объектов обучающей выборки на качество детектирования объектов являются актуальными и позволяют решить проблему повышения качества построения обучающих множеств в достаточно общем случае.

1. Выбор критерия качества обучения нейронной сети

Требуется выбрать критерий качества детектирования объектов. Существуют три критерия качества [8]: точность (precision), полнота (recall) и достоверность (accuracy). Каждый из критериев вычисляется при помощи следующих оценок:

* True positive rate (tpr) – детектор находит искомый объект, и он там на самом деле присутствует.

* False positive rate (fpr) – детектор находит объект, но его там на самом деле нет.

* True negative rate (tnr) – детектор не находит искомый объект, и он там на самом деле отсутствует.

* False negative rate (fnr) – детектор не находит искомый объект, но он там есть.

Оценку fnr называют ошибкой 1-го рода, а fpr – ошибкой 2-го рода. Точность, полнота и достоверность определяются следующим образом:

$$precision = \frac{tpr}{tpr + fpr}, \quad (1)$$

$$recall = \frac{tpr}{tpr + fnr}, \quad (2)$$

$$accuracy = \frac{tpr + tnr}{tpr + fnr + tnr + fnr}. \quad (3)$$

Для определения качества детектирования объектов требуется использовать несколько критериев. Это связано с тем, что при детектировании важно найти все объекты на изображении и отнести их к определенному классу. В связи с этим было принято решение использовать комбинированный критерий качества, а именно – F -меру [8], которая вычисляется по формуле

$$F = (\beta^2 + 1) \frac{precision \cdot recall}{\beta^2 \cdot precision + recall} 100 \%, \quad (4)$$

где β принимает значения в диапазоне $0 < \beta < 1$, если требуется точность, и $\beta > 1$, если требуется полнота. Если оба параметра важны, то $\beta = 1$.

2. Математическая постановка задачи

При детектировании объектов требуется решить две задачи. Первая задача (сегментация объекта) служит для поиска объекта на изображении, а вторая задача (распознавание) – для классификации найденных объектов. Рассмотрим эти задачи более подробно.

Задача распознавания

Для решения задачи распознавания требуется найти такую решающую функцию $f : X \rightarrow R$, для которой будут построены решаю-

щие правила. Здесь X – множество изображений; R – множество классов, используемых для обучения.

Для построения решающей функции требуется решить задачу классификации объектов для множества X . Для решения задачи классификации требуется наличие эксперта. Если мы отнесем все объекты множества X к классам, имеющимся во множестве R , то получим обучающее множество. Оно позволит нам настроить нейронную сеть таким образом, чтобы каждый элемент множества X можно было отнести к классу, имеющемуся во множестве R . В этом случае нейронная сеть играет роль функции f , которая будет хранить в себе правила распознавания объектов. В случае с нейронной сетью правила – это веса обученной сети.

Задача сегментации

Для поиска объектов на изображении требуется решить задачу сегментации объектов. Для решения задачи требуется выделить все области, в которых находятся объекты искомого класса. В конфигурации рассматриваемых нейронных сетей используется метод, который предлагает области-кандидаты. Каждая область-кандидат обрабатывается предобученной сверточной нейронной сетью, которая определяет точность принадлежности области к классам объектов, детектируемых сетью. На следующем этапе выбирается наибольшее значение принадлежности области-кандидата к искомым классам и сравнивается с пороговым значением. Если значение области-кандидата больше, чем пороговое значение, то область относится к классу с наибольшим значением принадлежности.

Математически задачу сегментации можно представить в следующем виде. Рассматривается система уравнений:

$$\begin{cases} y = F_y(x); \\ \gamma = F_c(y), \end{cases} \quad (5)$$

где x – исходное изображение; F_y – функция областей-кандидатов; γ – степень уверенности для области, в которой есть искомый объект; y – область-кандидат; F_c – функция, вычисляющая степень уверенности, с которой искомый объект находится в области-кандидате. Поиск функции областей-кандидатов осуществляется с помощью нейронных сетей. Для сети SSD поиск функции осуществляется с помощью алгоритма, выделяющего специальные регионы для поиска объектов, а сеть

Faster R-CNN для поиска областей использует метод скользящего окна. Функция F_c определяется также при помощи нейронной сети.

Для изображений x при помощи функции F_y выберем все области y , которые предположительно относятся к искомому классу. После определения всех областей-кандидатов для каждой из них рассчитаем степень уверенности, с которой они относятся к искомому классу.

Для сети задается параметр γ^* , который определяет степень уверенности для отнесения области-кандидата к выбранному классу. Требуется для всех областей-кандидатов определить отличие от параметра γ^* . Если $\gamma > \gamma^*$, то в области-кандидате находится детектируемый объект. Если $\gamma < \gamma^*$, то область-кандидат не детектирует искомый класс объекта.

3. Описание вычислительных экспериментов

Для проведения эксперимента были выбраны три сети: SSD 300, SSD 500, Faster R-CNN. Для каждой из сетей формируются два обучающих и одно валидационное множество, с помощью которого определяется точность решения поставленной задачи. Перед формированием обучающих множеств зададим константу S – площадь объекта, размеры которого требуется детектировать при помощи нейронной сети. По мощности первое и второе обучающее множество одинаковы. В первом обучающем множестве содержатся объекты, площадь которых меньше, чем S . Во втором обучающем множестве, наоборот, содержатся объекты, площадь которых больше, чем S . Валидационное множество состоит из объектов, размеры которых приблизительно равны S . Максимальное отклонение от S в валидационном множестве равно 30 пикселям. Площадь объектов, которые мы хотим детектировать, равна 10 000 пикселей. Иными словами, примерный размер объекта составляет 100×100 пикселей.

4. Обучение сверточной нейронной сети

Вычислительный эксперимент заключался в проверке обучающего множества, состоящего из 5 классов. Обучающее множество состояло из 5000 объектов для обучения, по 1000 элементов на один класс. Тестовое множество состояло из 1500 объектов, по 300 объектов на класс. Валидационное множество также содержало 1500 объектов, по 300 объектов для каждого класса. Для обучения сети использовался метод обучения с учителем [1]. Для каждого класса объектов было составлено обучающее множество, элементы которого относились к одному

из детектируемых классов. Начальные значения весовых коэффициентов сети были выбраны случайными величинами с нормальным распределением. Для обучения требовалось размеченное изображение в формате jpg и xml-файл с размеченными на изображении объектами. Сеть обучалась на следующих классах: человек, кот, собака, машина, велосипед.

Обучение производилось на следующем аппаратном обеспечении: процессор Intel(R) Core(TM) i7-7700K, оперативная память CPU 16GB DIMM Synchronous 2133MHz (0,5 ns.) 2 шт. 4.20GHz, жесткие диски 240GB KINGSTON, 1TB WDC, видеокарта NVIDIA GeForce GTX 1080 12GB 1 шт.

При проведении эксперимента сеть SSD 300 обучалась 20 ч, SSD 500 – 24 ч, Faster R-CNN – 26 ч. Все обучения проводились на видеокарте. Проверка точности на валидационном множестве составляла 3 мин. Прогон сети на валидационном множестве проводился на видеокарте.

5. Результаты

На рис. 2 показана зависимость точности детектирования объектов от числа итераций при обучении сети SSD300. По оси абсцисс указаны номера итераций, а по оси ординат – точность в процентах. Точность высчитывалась с использованием F -меры (см. формулу (4)). Для первого обучающего множества точность составила 73 %, для второго – 95 %. Параметр $\beta = 1$, так как считалось, что одинаково важна точность и полнота при детектировании объектов.

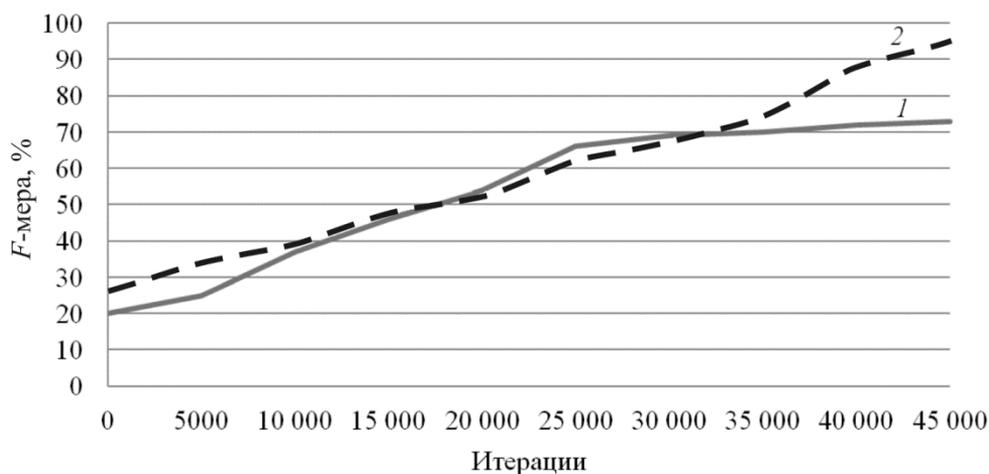


Рис. 2. Зависимость точности от числа итераций для сети SSD300: 1 – первое обучающее множество; 2 – второе обучающее множество

На рис. 3 показана зависимость точности детектирования объектов от числа итераций при обучении сети SSD 500. Для первого обучающего множества точность составила 73 %, для второго – 98 %.

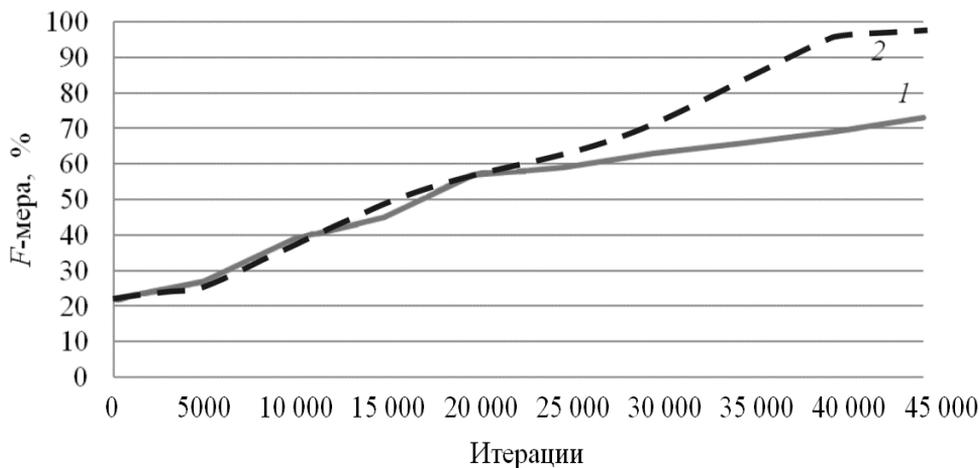


Рис. 3. Зависимость точности от числа итераций для сети SSD500: 1 – первое обучающее множество; 2 – второе обучающее множество

На рис. 4 показана зависимость точности детектирования объектов от числа итераций при обучении сети Faster R-CNN. Для первого обучающего множества точность составила 81 %, для второго – 98 %.

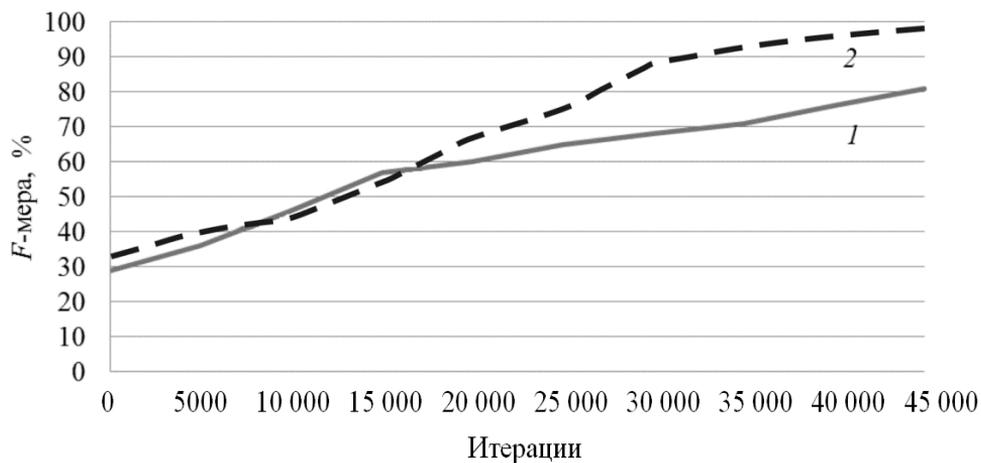


Рис. 4. Зависимость точности от числа итераций для сети Faster R-CNN: 1 – первое обучающее множество; 2 – второе обучающее множество

Из представленных результатов видно, что все исследованные нейронные сети после обучения позволяют с достаточно высокой точностью решать задачу детектирования. Однако второе обучающее множество позволяет добиться более точных результатов.

Заключение

В результате работы каждая из конфигураций нейронных сетей была обучена на двух обучающих множествах, которые состояли из пяти классов объектов. Данный вычислительный эксперимент показал, что точность нейронной сети, обученной на втором множестве (с более крупными объектами), на 20 % больше по сравнению с сетью, которая была обучена на множестве, состоящем из малых объектов.

На основании выполненной работы и проведенных экспериментов можно дать следующую рекомендацию: при детектировании объектов определенного размера необходимо составлять обучающее множество таким образом, чтобы размеры объектов были больше, чем детектируемый объект.

Список литературы

1. Ясницкий Л.Н. Интеллектуальные системы. – М.: Лаборатория знаний, 2016. – 221 с.
2. Speed/accuracy trade-offs for modern convolutional object detectors / J. Huang, V. Rathod, Ch. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Ya. Song, S. Guadarrama, K. Murphy // ArXiv preprint arXiv. – 2016. – 1611.10012.
3. Faster R-CNN: Towards Real-Time Object / Sh. Ren, K. He, R. Girshick, J. Sun // ArXiv preprint arXiv. – 2015. – 1506.01497.
4. SSD: Single Shot MultiBox Detector / W. Liu, D. Anguelov, D. Erhan, Ch. Szegedy, S. Reed, Cheng-Yang Fu, A.C. Berg // ArXiv preprint arXiv. – 2015. – 1512.02325.
5. Иваськив Ю.Л. Формирование нечетких обучающих множеств для нейронных сетей в задачах сжатия данных без потерь // Математические машины и системы. – 2009. – № 2. – С. 53–60.
6. Беляков И.А. Метод генерации множества обучающих примеров для нейронной сети, выполняющей задачу верификации // Из-

вестия Петербургского университета путей и сообщения. – 2012. – № 3. – С. 19–28.

7. Гагарин О.И. Способ настройки многомасштабных моделей детектирования визуальных объектов в сверточной нейронной сети // *Нейрокомпьютеры: разработка, применение.* – 2018. – № 2. – С. 50–56.

8. David M.W. Powers Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation // *School of Informatics and Engineering Flinders University, Technical Report SIE-07-001.* – December 2007.

References

1. Yasnitsky L.N. *Intellectual systems.* M.: Laboratory of Knowledge, 2016. 221p.

2. Jonathan Huang, Vivek Rathod, Chen Sun, Menglong Zhu, Anoop Korattikara, Alireza Fathi, Ian Fischer, Zbigniew Wojna, Yang Song, Sergio Guadarrama, Kevin Murphy. Speed/accuracy trade-offs for modern convolutional object detectors. // *ArXiv preprint arXiv:1611.10012* - 2016.

3. Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun. Faster R-CNN: Towards Real-Time Object. // *ArXiv preprint arXiv:1506.01497* - 2015.

4. Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg. SSD: Single Shot MultiBox Detector. // *ArXiv preprint arXiv:1512.02325.* 2015.

5. Ivaskiv. Yu.L. Formation of fuzzy learning sets for neural networks in problems of data compression without losses // *mathematical machines and systems.*, 2009. 53p.

6. Belyakov I.A. A method for generating a set of training examples for a neural network performing the verification task. // *Proceedings of St. Petersburg University of Railways and Communications.*, 2012. 19p.

7. Gagarin O.I. A method for tuning multiscale models for detecting visual objects in a convolutional neural network. // *Neurocomputers: development, application.*, 2018.- 50p.

8. David M. W. Powers Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation // *School of Informatics and Engineering Flinders University, Technical Report SIE-07-001- December 2007.*

Получено 25.04.2018

Об авторах

Каракулов Игорь Владимирович (Пермь, Россия) – магистрант Пермского национального исследовательского политехнического университета (614990, г. Пермь, Комсомольский пр., 29, e-mail: ikarakulov@techinfocom.com).

Клюев Андрей Владимирович (Пермь, Россия) – кандидат физико-математических наук, доцент кафедры «Вычислительная математика и механика» Пермского национального исследовательского политехнического университета (614990, г. Пермь, Комсомольский пр., 29, e-mail: kav@gelicon.biz).

About the authors

Igor V. Karakulov (Perm, Russian Federation) – Master Student, Perm National Research Polytechnic University (29, Komsomolsky av., Perm, 614990, Russian Federation, e-mail: KarakulovIV@yandex.ru).

Andrei V. Kliuev (Perm, Russian Federation) – Ph.D. in Physics and Mathematics, Associate Professor, Department of Computational Mathematics and Mechanics, Perm National Research Polytechnic University (29, Komsomolsky av., Perm, 614990, Russian Federation, e-mail: kav@gelicon.biz).