

УДК 004.04

DOI: 10.46548/21vek-2020-0952-0004

**МЕТОД ПРОГРАММНОЙ ОБРАБОТКИ НАБОРА ДАННЫХ, ОПТИМИЗИРУЮЩИЙ  
КАЧЕСТВО ОБУЧЕНИЯ СВЕРТОЧНЫХ СЕТЕЙ, РЕШАЮЩИХ ЗАДАЧУ  
ПОИСКА И ИДЕНТИФИКАЦИИ ОБЪЕКТОВ**

© 2020

**Малинин Дмитрий Денисович**, ведущий программист

ООО «Ай-Текс ВЦ»

(440026, Россия, Пенза, ул. Пушкина, д. 3., e-mail: dmt.malin@gmail.com)

**Аннотация.** Данная статья посвящена описанию метода программной подготовки набора данных для обучения сверточных нейронных сетей, которые решают задачу поиска и идентификации объектов. Описанный в статье метод был применен к набору данных для обучения сверточной нейронной сети, который представляет собой множество изображений с размеченными областями интересов (Region Of Interest). Характерной особенностью данного набора является то, что области интересов охватывают не только объекты идентификации, но и другие информационные признаки изображения, которые не характеризуют искомые объекты. Метод предлагает провести программную подготовку данных, путем структурного анализа границ двоичного изображения, с последующим построением маски по найденным контурам с целью наложения на исходное изображение, что в конечном итоге позволяет оставить в области интересов только объект идентификации. Применение данного метода на целевом наборе данных позволило существенно улучшить качество обучения сверточной сети Fast-RCNN для задачи идентификации объекта на изображении. Качественная оценка проводилась путем сравнения на каждой эпохе обучения приближения кросс-энтропии к нулевому значению. На наборе, обработанном предлагаемым методом, скорость приближения кросс-энтропии к нулевому значению более чем на 16% быстрее относительно исходного набора без обработки.

**Ключевые слова:** Нейронная сеть, сверточная сеть, подготовка данных, компьютерное зрение, фильтр Гаусса, метод Оцу, алгоритм Рамера - Дугласа - Пекера.

**DATA SET PROGRAM PROCESSING METHOD OPTIMIZING QUALITY OF TRAINING OF  
CONVOLUTION NETWORKS SOLVING THE OBJECT OF SEARCH AND IDENTIFICATION OF  
OBJECTS**

© 2020

**Malinin Dmitriy Denisovich**, lead programmer

Limited liability company «I-teco Computer Center»

(440026, Russia, Penza, Pushkina St., 3, e-mail: dmt.malin@gmail.com)

**Abstract.** This article is devoted to the description of the method of software preparation of a data set for training convolutional neural networks that solve the problem of searching and identifying objects. The method described in the article was applied to a dataset for training a convolutional neural network, which is a set of images with marked regions of interest (Region Of Interest). A characteristic feature of this set is that the areas of interest cover not only identification objects, but also other information signs of the image that do not characterize the desired objects. The method suggests that the data be prepared programmatically by structural analysis of the boundaries of the binary image, followed by the construction of a mask according to the found contours in order to overlay the original image, which ultimately allows you to leave only the identification object in the area of interest. The application of this method on the target data set made it possible to significantly improve the quality of training of the Fast-RCNN convolution network for the task of identifying an object in the image. A qualitative assessment was carried out by comparing, at each training epoch, the approximation of cross-entropy to a zero value. The rate of cross-entropy approaching to zero on a set processed by the proposed method is more than 16% faster than the initial set without processing.

**Keywords:** Neural network, convolutional network, data preparation, computer vision, Gaussian filter, Otsu method, Ramer-Douglas-Pecker algorithm.

**Введение.** В задачах машинного обучения качество моделей сильно зависит от качества данных, на которых производится процесс обучения. Под качеством данных следует понимать максимально возможный набор информационных характеристик, из которых модель сможет наиболее эффективно построить нелинейные зависимости для решения поставленной перед ней задачи [1,2]. Наиболее распространенной задачей в машинном обучении является задача поиска и идентификации объектов, практическое применение которой встречается в различных технических сферах.

Решения данной задачи используются, например, в системах контроля доступа для идентификации человека и в промышленных системах на производственных линиях для подсчета продукции в товарных упаковках. В сверточных сетях типа *Fast-RCNN*, *YOLO*, которые решают проблему поиска и идентификации объекта, обучающим набором данных являются изображения с размеченными классификаторами области интересов. Классификаторы области интересов представляют собой множество прямоугольников  $P = \{p_1, \dots, p_n\}$ . Каждый прямоугольник  $p_i \in P$  описывается точ-

кой  $A_i(x_i, y_i)$  и парой положительных действительных чисел  $(w_i, h_i)$ , которые соответствуют длине и ширине прямоугольника. Таким образом область интересов ограничивает объекты прямоугольником  $p_i \in P$  от остального изображения [3,4].

Однако, исходя из определения, в область интереса могут попадать данные, которые не содержат полезных характеристик, что может отрицательно влиять на скорость и качество обучения в сверточных сетях. Для оценки качества обучения используется функция потерь, которая характеризует насколько далеко прогнозируемые значения отклоняются от фактических значений в наборе данных. Существует несколько видов оценки функции потерь. В данном исследовании отклонение от фактических значений характеризуется кросс-энтропийной функцией (*Cross-entropy loss*), которая интерпретирует выходной сигнал в интервале значений от 0 до 1. Другими словами, идеально обученная модель будет иметь нулевое значение кросс-энтропии [5].

**Целью** данной статьи является разработка метода,

который позволяет удалить из области интересов информацию, не содержащую характеристик идентифицируемого объекта.

В процессе исследования, был подготовлен набор данных для обучения сети. В последствии набор данных был обработан предлагаемым методом, на котором проводилось повторное обучение. В результате произведено сравнение двух функций отклонений от процесса обучения с данными без обработки и с обработкой соответственно. Функция потерь от процесса с обработкой показала лучшее приближение кросс-энтропии к нулевому значению.

**Материалы исследования.** Метод заключается в структурном анализе границ двоичного изображения (контурный анализ), с последующим построением маски по найденным контурам с целью наложения на исходное изображение, что в конечном итоге позволяет оставить в области интересов только объект идентификации [6-8]. Предложенный метод обработки изображения можно представить с помощью алгоритма на рисунке 1.

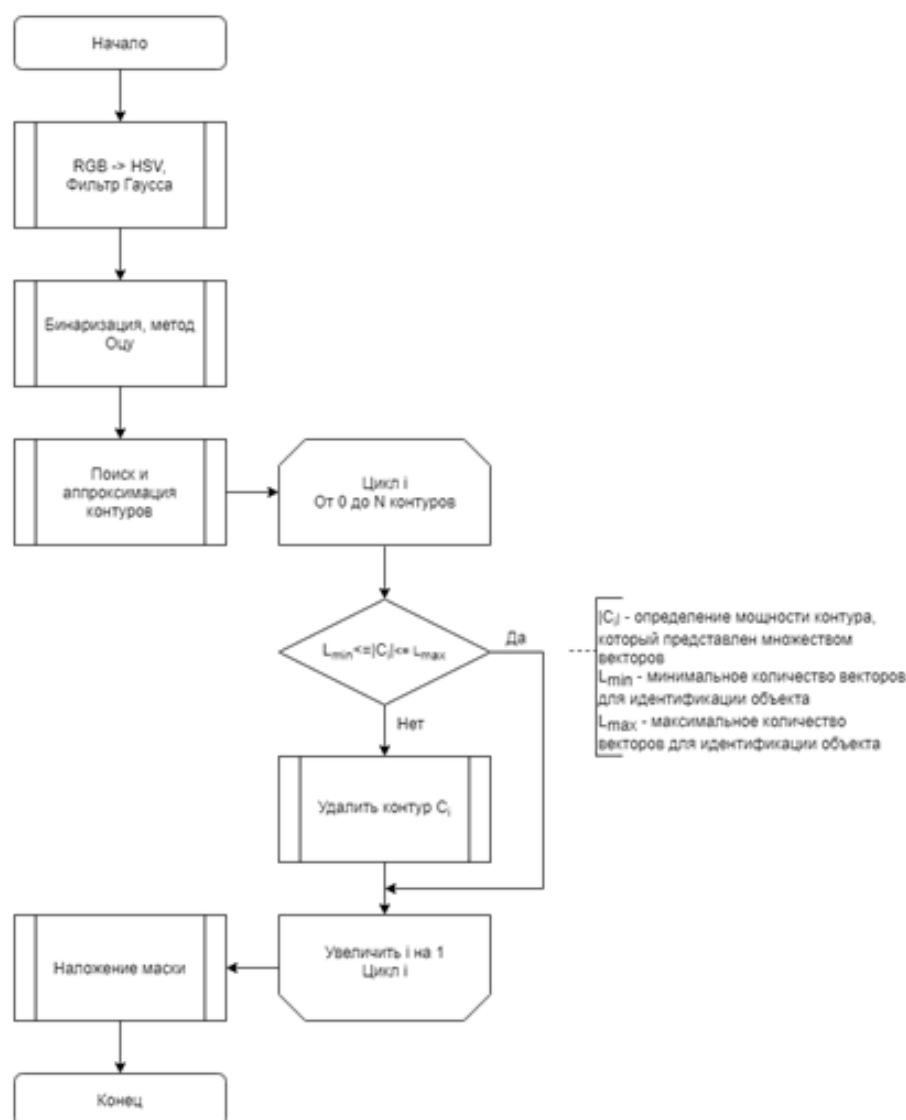


Рисунок 1 – Алгоритм обработки изображения

На вход алгоритма поступает изображение в цветовом пространстве  $RGB$ . Пространство  $RGB$  необходимо преобразовать к  $HSV$ . Целью этого преобразования является то, чтобы получить более выраженные детали изображения, независимо от условий освещения [9]. После преобразования обработка производится на канале насыщенности  $S$ . Преобразование из цветового пространства  $RGB$  в  $HSV$  и расчет  $S$ -канала приведено в уравнении 1 [10].

$$\begin{aligned} C_{max} &= \max(R, G, B) \\ C_{min} &= \min(R, G, B) \\ S &= \begin{cases} 0, & C_{max} = 0 \\ 1 - \frac{C_{min}}{C_{max}}, & C_{max} \neq 0 \end{cases} \end{aligned} \quad (1)$$

После выделения канала насыщенности, на изображение накладывается гауссовский фильтр (*Gaussian Blur*) для устранения помех [11]. Общее математическое выражение гауссовского фильтра приведено в уравнении 2.

$$G(r) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{r^2}{2\sigma^2}} \quad (2)$$

Для поиска границ объектов, подготовленный канал насыщенности необходимо привести к бинарному виду, путем нахождения пороговых значений. Порого-

вые значения на изображении определяются методом Оцу (*Otsu's Method*) [12]. Метод Оцу использует гистограмму изображения для расчета порога бинаризации.

Суть метода заключается в разделении пикселей изображения на два класса: «объектный» и «фоновый», путем минимизации внутриклассовой дисперсии, которая определяется как взвешенная сумма дисперсий двух классов, что выражается в уравнении 3.

$$\sigma_w^2 = \omega_1 \sigma_1^2 + \omega_2 \sigma_2^2 \quad (3)$$

$\omega_1$  и  $\omega_2$  – вероятности первого и второго классов соответственно.

В свою очередь, как показано в исследовании Оцу, минимизация внутриклассовой дисперсии эквивалентна максимизации межклассовой дисперсии, что показано в уравнении 4.

$$\sigma_b^2 = \omega_1 \omega_2 (a_1 - a_2)^2 \quad (4)$$

$a_1$  и  $a_2$  – средние арифметические значения для каждого из классов.

Применяя метод Оцу на канал насыщенности  $S$ , получим бинарное изображение. На рисунке 2 изображен канал насыщенности  $S$ , результат применения метода Оцу и гистограмма распределения канала  $S$  для определения минимизации межклассовой дисперсии соответственно.



Рисунок 2 – Канал  $S$ , бинарное изображение, гистограмма распределения канала  $S$

Далее по бинарному изображению производится топологический структурный анализ для определения контуров с помощью алгоритма, предложенным Сатоши Сузуки (*Satoshi Suzuki*) [13,14]. Алгоритм предполагает нахождение контуров с учетом вложенности, то есть способен определить, когда в контур одного объекта вложен другой. Результатом применения алгоритма является множество подмножеств векторов (контуров), которые описывают, находящиеся в области интересов, границы объектов.

Но как уже было отмечено, в области интересов могут присутствовать объекты, которые не являются объектами идентификации. Исключение таких объектов, производится путем нахождения и анализа аппроксимированных кривых, построенных по найденным контурам. Для нахождения аппроксимирующей кривой используется алгоритм Рамера — Дугласа — Пекера [15,16]. Алгоритму задается исходная кривая и максимальное расстояние ( $\epsilon$ ), которое может быть между исходной и упрощенной кривой. Максимальное расстояние, в данном исследовании, вычисляется как произведение коэффициента аппроксимации на длину кривой. Коэффициент аппроксимации устанавли-

вается эмпирически в интервале от 0,01 до 0,1. Формула расчета  $\epsilon$  представлена в уравнении 5.

$$\epsilon = k * L, \quad (5)$$

$k$  - коэффициент аппроксимации,  $L$  — длина кривой.

В результате применения алгоритма Рамера — Дугласа — Пекера, объект на изображении определяется подмножеством упрощенных векторов (контур). Другими словами, тип объекта можно определить исходя из количества, описывающих его, векторов или минимальным и максимальным пороговым значением. Тем самым, на изображении исключаются объекты, у которых количество аппроксимированных кривых не эквивалентно количеству кривых идентифицируемого объекта или количество кривых не находится в интервале пороговых значений [17]. Так же дополнительным фильтром для исключения может являться длина кривой, которая должна быть больше установленного порогового значения.

Последним шагом предлагаемого метода, является построение маски с последующим наложением на исходное изображение, для того, чтобы исключить цветовую информацию, которая находится вне най-

денных границах [18]. Маска строится путем создания нулевой матрицы  $\Theta_{IKW}$  с последующим заполнением по границам значениями, которые соответствуют представлению белого цвета.

Далее наложение маски на исходное изображение производится путем побитового соединения, формула которого представлена в уравнении 5.

$$dst(I)=src1(I)Asrc2(I) \quad (5)$$

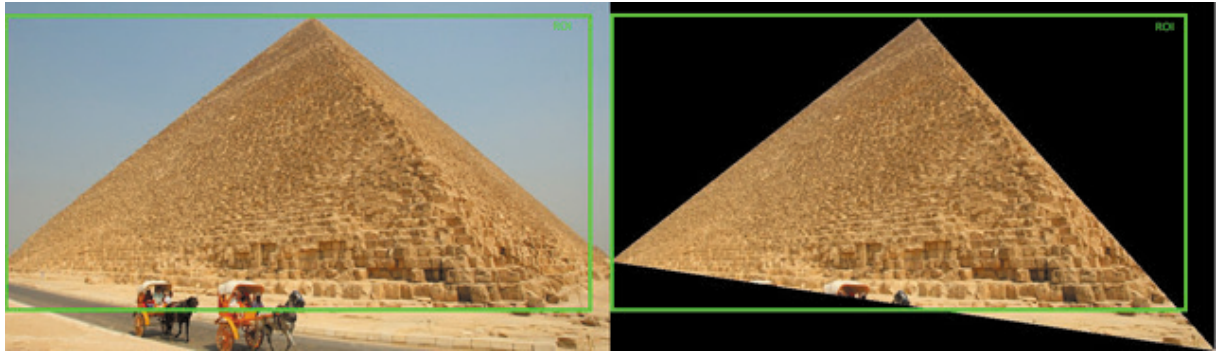


Рисунок 3 – Исходное и обработанное алгоритмом изображение. Объект идентификации – пирамида

На рисунке 4 представлен график кросс-энтропии [19] потерь на каждой эпохе обучения для исходного набора и для набора с обработкой. Как видно из рисунка 3 приближение кросс-энтропии к нулевому значению происходит быстрее на наборе данных, который был обработан предложенным методом.

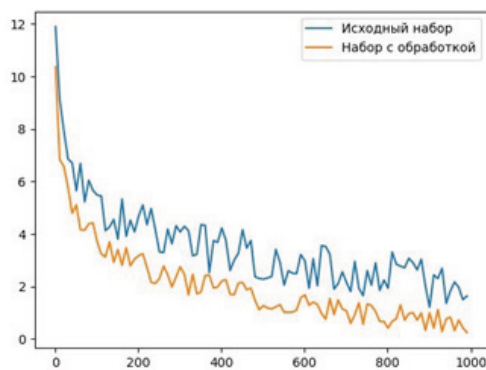


Рисунок 4 – график функций потерь обучения

Разница скорости приближения к нулевому значению кросс-энтропии составила 16.46% в пользу набора данных обработанным предложенным методом. Формула расчета процентной разности для каждой эпохи обучения представлена в уравнении 6 [20].

$$diff = \frac{\sum_{i=1}^n \left| (\eta_{исх\ i} - \eta_{обр\ i}) / \left( \frac{\eta_{исх\ i} + \eta_{обр\ i}}{2} \right) \right|}{n} \cdot 100, \quad (6)$$

$\eta_{исх}$  и  $\eta_{обр}$  – значение кросс-энтропии на каждой эпохе обучения для исходного и обработанного набора соответственно.

**Заключение.** Таким образом, применяя данный метод при подготовки данных для обучения свер-

**Результаты исследования.** В результате применения данного метода область интересов содержит только информацию, характеризующую объект идентификации. На рисунке 3 представлены исходное изображение, обработанное алгоритмом и размеченная область интересов (ROI), которая описывает объект идентификации. Как видно из рисунка 3, алгоритм успешно удалил информацию, которая не относится к объекту идентификации.

точных сетей, типа *Fast-RCNN*, *YOLO*, возможно существенно улучшить качество и скорость обучения. Также к достоинствам данного метода можно отнести универсальность, так как метод может применяться практически на любых наборах данных, подготовленных для поиска и идентификации объектов.

#### СПИСОК ЛИТЕРАТУРЫ:

1. Умберто М. Прикладное глубокое обучение. Подход к пониманию глубоких нейронных сетей на основе метода кейсов, БХВ-Петербург // - 2020. - С. 368.
2. Горлач Б. Тензорная алгебра и тензорный анализ, Лань // - 2015. - С. 160.
3. Rafael C. Gonzalez, Richard E. Woods, Digital Image Processing (4th Edition) // - 2017, - С. 1192
4. Клетте Р. Компьютерное зрение. Теория и алгоритмы, ДМК Пресс // - 2019. - С. 506.
5. Паттанаяк С. Глубокое обучение и TensorFlow для профессионалов. Математический подход к построению систем искусственного интеллекта на Python, Вильямс // - 2019. - С. 480.
6. Скиена С. Алгоритмы. Руководство по разработке, BHV // - 2017. - С. 720.
7. Трутнев Д. Р. Архитектуры информационных систем. Основы проектирования. // СПб.: НИУ ИТМО, 2012. – 66 с.
8. Кнут Д. Искусство программирования т.1. Основные алгоритмы // Вильямс Издательский дом 2000. - С. 27.
9. T. Pavlidis, Algorithms for Graphics and Image Processing, Computer Science Press, Rockville, Maryland, 1982.
10. Гарсия Г. Обработка изображений с помощью OpenCV, ЛитРес // - 2017. - С. 212.
11. J. Howse. Learning OpenCV 3 Computer Vision with Python - Second Edition, Packt Publishing // - 2015. – С. 266.
12. N. Otsu. A threshold selection method from gray level histograms. IEEE Trans. systems. Man. and Cybernetics, 9:62–66, 1979.
13. Suzuki, S. and Abe, K., Topological Structural Analysis of Digitized Binary Images by Border Following. CVGIP 30 1, pp 32–46, 1985.
14. Viola P., Jones M. Rapid object detection using a boosted

cascade of simple features // Proc. IEEE Conference on Computer Vision and Pattern Recognition. – Washington DC: IEEE Computer Society, 2001. – P. 511–518.

15. Ramer Urs. An iterative procedure for the polygonal approximation of plane curves // Computer Graphics and Image Processing. 1972. № 1(3). P. 244-256.

16. Douglas David, Peucker Thomas. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature // The Canadian Cartographer. 1973. № 10(2). P. 112-122.

17. Farell A. James. From Pixels to Animation: An Introduction to Graphics Programming, Academic Press // - 2014. - С. 692.

18. Raja Y., McKenna S., Gong S. Colour model selection and adaptation in dynamic scenes // Proceedings of the 5th European Conference on Computer Vision (ECCV '98). – London: Springer-Verlag, 1998. – V. 1. – P. 460-475.

19. J. de Vries. Object Recognition: A ShapeBased Approach using Artificial Neural Networks //-2006

20. P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan, Object Detection with Discriminatively Trained Part Based Models, IEEE Trans. Pattern Anal. and Mach. Intell. 32(9), 1627–1645 (2010).

*Статья поступила в редакцию 17.08.2020*

*Статья принята к публикации 14.09.2020*