

УДК 004.85

АНАЛИЗ АЛГОРИТМОВ ПОИСКА ОБЪЕКТОВ НА ИЗОБРАЖЕНИЯХ С ИСПОЛЬЗОВАНИЕМ РАЗЛИЧНЫХ МОДИФИКАЦИЙ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ

А. А. Сирота, Е. Ю. Митрофанова, А. И. Милованова

Воронежский государственный университет

Поступила в редакцию 10.09.2019 г.

Аннотация. Рассматриваются задачи поиска и классификации объектов на изображениях с использованием современных алгоритмов компьютерного зрения. Приведено описание основных типов алгоритмов и методов для поиска объектов, основанных на использовании глубоких нейронных сетей. Проведен сравнительный анализ и моделирование нейросетевых алгоритмов для решения задачи классификации и поиска объектов на изображениях. Приведены результаты тестирования нейросетевых моделей с различными архитектурами на наборах данных VOC2007 и COCO. Проанализированы результаты исследования точности распознавания в зависимости от различных гиперпараметров обучения. Исследовано изменение значения времени определения местоположения объекта в зависимости от различных архитектур нейронной сети.

Ключевые слова: поиск объектов, классификация объектов, нейронные сети, глубокое обучение, R-CNN, Fast R-CNN, Faster R-CNN.

ВВЕДЕНИЕ

Поиск объектов на изображениях является одной из важнейших задач анализа сцен и машинного зрения. Задача поиска объектов на изображениях обычно состоит из двух подзадач: первая из них относится к классу задач обнаружения (детектирования) объекта, принадлежащего одному из заданных классов на изображении; вторая – в классификации найденного объекта по принадлежности одному из заданных классов. Решение указанных задач может осуществляться как в два последовательно выполняемых этапа, так совместно, без разделения на этапы по времени. Сложность рассматриваемой общей задачи определяется высокой степенью изменчивости реальных изображений и представленных на них объектов. Меняется прак-

тически все: положение, ракурс, освещение, цвет, формой и т. д.

До недавнего времени задачи указанного класса задач решались с использованием различных алгоритмов обработки информации, в том числе: алгоритмов адаптивного усиления [1], алгоритмов, основанных на использовании гистограмм градиентов [2, 3] и информации о цвете [4], алгоритмов каскадных классификаторов на основе метода Виолы-Джонса [5, 6], который особенно хорошо показал себя для детектирования лиц людей на изображениях. Также широкое распространение получили алгоритмы, основанные на методах контурного анализа [7].

Известны также решения, основанные на обычных нейронных сетях с предварительным выделением признаков классификации объектов, а также стандартных глубоких сверточных нейронных сетей, применяемых в режиме просмотра последовательно перебираемых в скользящем режиме окон, в кото-

© Сирота А. А., Митрофанова Е. Ю.,
Милованова А. И., 2019

рых проводится анализ [8]. Следует отметить, что такой подход является весьма затратным в вычислительном отношении, несмотря на все преимущества, которые изначально предоставляют нейронные сети с точки зрения точности классификации.

В последнее время для решения данной задачи широкое распространение получили алгоритмы, основанные на применении региональных глубоких сверточных нейронных сетей или Regional Convolutional Neural Networks (R-CNN), которые принципиально ориентированы на решение задачи поиска объектов с одновременной их классификацией [9, 10, 11]. По сравнению с методами, представленными выше, эти алгоритмы принципиально адаптированы для решения задачи поиска объектов на изображениях. Их исходная реализация базируются на использовании специальных алгоритмов предобработки – алгоритмов region-proposal-function, обеспечивающих предложение так называемых областей внимания, в которых потенциально могут находиться интересующие нас объекты. Такой «специализированный» подход предлагает сократить вычислительные затраты, а также позволяют добиться минимального времени определения местоположения объекта и высокой точности его классификация. К настоящему моменту имеется большое количество вариантов реализации подобных алгоритмов, которые достигли хороших показателей по данным критериям [12, 13]. Таким образом, целью данного исследования является проведение сравнительного анализа нейросетевых алгоритмов класса R-CNN, для поиска объектов на изображениях. При этом для реализации одного из рассмотренных вариантов используется оригинальный вариант алгоритма region-proposal-function.

МЕТОДЫ И МАТЕРИАЛЫ ИССЛЕДОВАНИЯ

Рассмотрены следующие модели нейронных сетей для решения задач поиска и классификации объектов на изображениях: R-CNN (region-based convolutional neural networks) с оригинальным алгоритмом region-proposal-

function, Fast R-CNN, Faster R-CNN [9, 10, 11]. При проведении исследований использовались наборы данных VOC 2007 и COCO.

Набор данных VOC 2007 (20 классов, 9963 изображений) [14] состоит из изображений размером 500 × 375, каждое из которых имеет файл аннотации, дающий координаты ограничивающей рамки и метку класса объекта из перечня двадцати классов.

Набор данных COCO (80 классов, 330000 изображений) [15] схож по структуре с VOC2007, при этом размер изображений – 640 × 507. Помимо файлов с изображениями, также содержит аннотации, в которых содержатся координаты объекта, класс, к которому он принадлежит, ссылку на него и дополнительную информацию.

Для обучения и тестирования моделей использовались 5300 изображений из каждого набора данных.

Модель R-CNN. При поиске объектов с использованием модели R-CNN [9] выполняется следующая последовательность шагов, представленных на рис. 1.

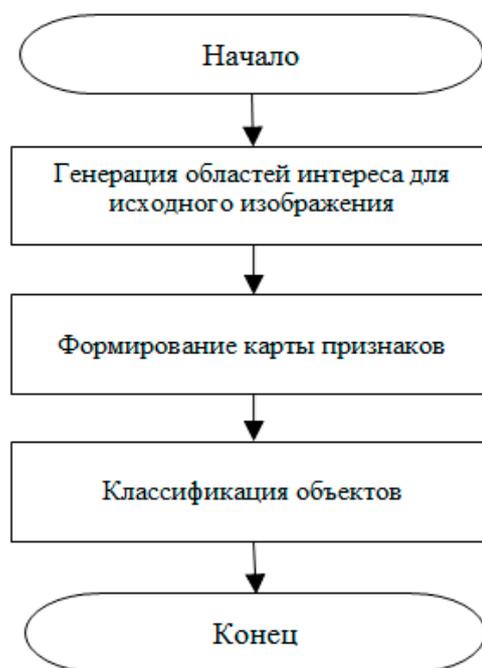


Рис. 1. Алгоритм R-CNN

Шаг 1. Выполняется генерация областей интереса (region proposals), предположительно содержащих в себе искомые объекты (обычно до 2000 возможных областей) с использованием различных алгоритмов, пред-

назначенных для снижения вычислительной сложности обнаружения объектов на изображении (например, алгоритмы Edge Boxes, Selective search).

В стандартной схеме поиска объектов, вначале выполняется проход детектором по исходному изображению (или по набору изображений) в разных масштабах со сдвигом окна детектирования. Таким образом, формируется набор областей интереса прямоугольной формы, возможно содержащих объект.

Алгоритм Edge Boxes [16] предназначен для генерации областей интереса в виде прямоугольной рамки, ограничивающей объект, и основан на выделении контуров объектов. Контурные обеспечивают разреженное, но информативное представление исходного изображения. При этом количество контуров, полностью содержащихся в ограничивающем прямоугольнике, указывает на вероятность того, что данный прямоугольник содержит объект. Оценка каждого блока выполняется путем исключения из множества контуров, которые содержатся в ограничивающем прямоугольнике (блоке), тех контуров, которые пересекаются с границами блока. Тем самым в качестве генерируемых областей интереса для исходного изображения используются области, содержащие замкнутые или почти замкнутые контуры.

В основе алгоритма селективного поиска (selective search) [17] лежит использование метода иерархической группировки похожих регионов на основе соответствия цвета, текстуры, размера или формы и графов. Для этих графов вершиной является интенсивность текущего пикселя, а ребра соединяют пару соседних пикселей. Абсолютная разница интенсивностей пикселей вершин используется в качестве веса ребра. С помощью графа выделяются фрагменты, которые затем группируются, согласно следующему принципу: ребра между двумя вершинами в одной группе должны иметь меньшие веса, а ребра между двумя вершинами в разных группах должны иметь больший вес. После того, как два наиболее похожих региона сгруппированы, между полученным регионом и его соседями вычисляется новое сходство. Процесс

группировки наиболее похожих областей повторяется до тех пор, пока все изображение не станет одной областью.

Таким образом, алгоритм селективного поиска реализуется с помощью детальной сегментации изображения в зависимости от интенсивности пикселей, с использованием метода сегментации на основе графов и выборочного поиска.

Шаг 2. На данном шаге выполняется формирование карты признаков для исходного изображения. Проводится масштабирование сформированных областей интереса в размер, сопоставимый с архитектурой нейронной сети CNN. Для этого выполняется аффинное преобразование, и каждая область интереса преобразуется в квадрат 227×227 , так как используемая архитектура CNN требует входы фиксированного размера 227×227 пикселей. При этом перед выполнением преобразования ограничивающая рамка области интереса расширяется таким образом, чтобы после преобразования получить вокруг области, возможно содержащей объект, кайму шириной 16 пикселей.

Сформированные таким образом данные поступают на вход сверточной нейронной сети (CNN), архитектура которой представлена на рис. 2.

На выходе нейронной сети CNN формируется 4096-мерный вектор признаков для каждой области интереса.

Шаг 3. На данном шаге выполняется классификация объектов для каждой области интереса с использованием сформированного вектора признаков на основе метода опорных векторов (SVM).

Для этого, изначально, перед выполнением классификации, для каждой сгенерированной области применяется подавление немаксимумов (алгоритм Non-Maximum Suppression), на основе которого только локальные максимумы отмечаются как контур объекта. Подавление немаксимумов необходимо для исключения сгенерированных дублирующих областей интереса для каждого объекта на исходном изображении. Пиксель, значение которого выше верхней границы, принимает максимальное значение, т. е. кон-

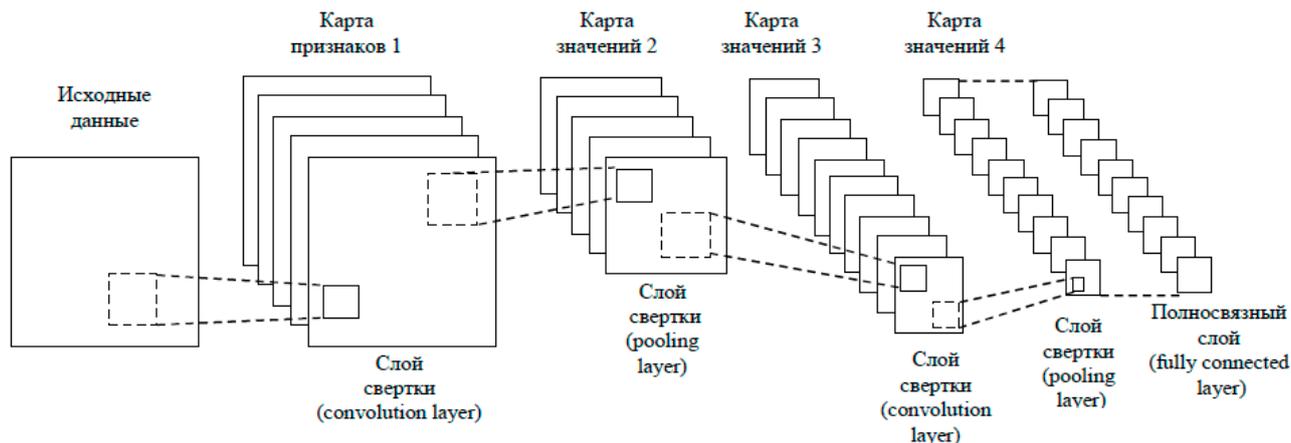


Рис. 2. Архитектура CNN

тур считается достоверным. Если значение пикселя не достигает нижнего порога пиксель подавляется.

Для оценки качества классификации используется показатель, равный отношению площади пересечения прямоугольника (области интереса), полученного в результате детектирования, и прямоугольника из разметки к площади их объединения (*Intersection over Union, IoU*). Таким образом, считается, что объект обнаружен правильно, если данный показатель превышает некоторый порог, в противном случае считается, что объект не обнаружен. Показатель *IoU* рассчитывается по следующей формуле:

$$IoU = \frac{AoO}{AoU}, \quad (1)$$

где *AoO* (*Area of Overlap*) – площадь пересечения истинного ограничивающего прямоугольника и предсказанного; *AoU* (*Area of Union*) – площадь объединения истинного ограничивающего прямоугольника и предсказанного.

На рис. 3 показан процесс поиска объектов на изображении с использованием сверточных нейронных сетей R-CNN.

Алгоритм region-proposal-function RCNN. Для модели R-CNN реализован оригинальный алгоритм определения областей интереса Region-proposal-function. Его применение для генерации этих областей предполагает следующую последовательность шагов, представленную на рис. 4. По своей архитектуре этот алгоритм близок к алгоритму, предло-

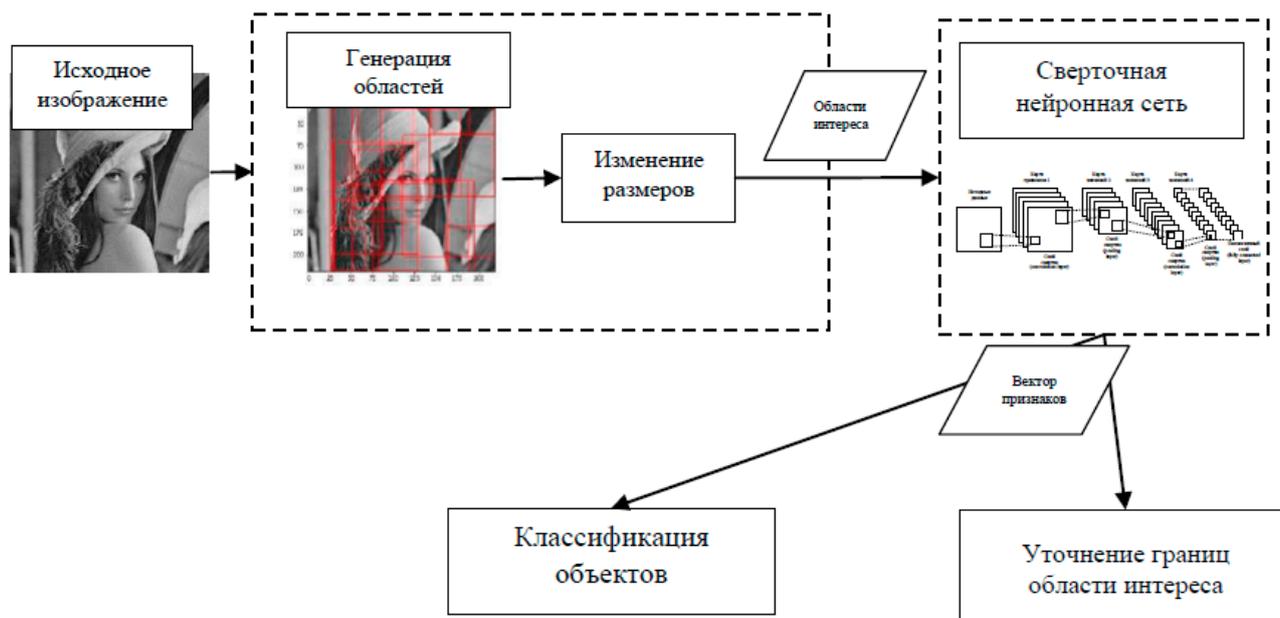


Рис. 3. Поиск объектов на изображении, R-CNN

женному в [9], но несколько проще и лучше по быстродействию.

Шаг 1. Выполняется генерация областей интереса, предположительно содержащих в себе объект. Исходное изображение преобразуется к полутоновому. После этого задаются константы, которые будут использоваться как размер шага для поиска областей интереса, в виде ограничивающих прямоугольников, по масштабу и соотношению сторон, а также параметры α (определяет перекрытие соседних ограничивающих прямоугольников) и β (определяет порог, согласно которому кандидаты будут удаляться). Для генерации областей интереса используется модификация алгоритма Edge Boxes.

1. Выполняется выделение контуров, содержащихся на изображении, используя при этом фильтр Собеля, при этом пиксель изображения считается относящимся к контуру, если соответствующий ему пиксель изображения после операции фильтрации имеет значение, большее заданного порога $por = 0.1$. Таким образом, исключаются пиксели, принадлежащие малым контурам. Затем вычисляется градиент изображения, при этом выполняется операция одномерной дифференцирующей маски в горизонтальном и/или вертикальном направлении с ядрами $[1, 0, -1]$ и $[1; 0; -1]$ соответственно, после чего формируется результирующая область интереса с заданными размерами (вычисляется корень из суммы квадратов полученных горизонтальной и вертикальной составляющих для построения карты контуров на изображении).

2. Выполняется поиск связанных компонентов в исходном изображении, после чего определяется самый большой контур и удаляется, то есть все пиксели данного контура принимают значение 0, если он меньше порогового значения. Оставшиеся контуры также анализируются, при этом, если их количество меньше заданного количества ограничивающих прямоугольников, то повторяется отсечение пикселей, принадлежащих малым контурам, с порогом $por = 0.05$.

3. Формируется матрица меток из связанной структуры контуров, содержащая целочисленные значения, пиксели, отмечен-

ные значением 0, являются фоном, пиксели с меткой 1 составляют один объект, пиксели, помеченные 2, составляют второй объект и так далее. На основе того полученных данных формируются области интереса. Используя обновленные области интереса, вычисляется интегральное изображение, в котором каждый пиксель является совокупной суммой пикселей непосредственно выше его и с левой стороны от него. Полученное изображение затем будет использовано для оценки сгенерированных ограничивающих прямоугольников (областей интереса).

4. Для итогового формирования областей интереса, предположительно содержащих объект, выполняется организация поиска данных областей, с использованием заданных константных значений (размер шага для поиска областей интереса, в виде ограничивающих прямоугольников, по масштабу и соотношению сторон).

Для поиска областей интереса организуется двойной цикл по масштабу, затем по соотношению сторон, в котором происходит расчет длины и ширины, а также шага смещения для ограничивающих прямоугольников, при этом перекрытие прямоугольников будет равно величине α . Полученные смещения используются далее для формирования векторов x и y , которые определяют координаты генерируемых прямоугольников, на основе чего предварительно рассчитывается количество ограничивающих прямоугольников. С помощью полученных конфигураций прямоугольников и их числа формируется окончательный массив конфигураций ограничивающих прямоугольников. Для этого аналогичным способом формируется поиск по масштабу и соотношениям сторон, рассчитываются координаты x и y ограничивающих прямоугольников. Таким образом, итоговый массив содержит в себе координаты x и y , а также длину и ширину ограничивающих прямоугольников.

Шаг 2. На данном шаге выполняется оценка для полученного перечня областей интереса (ограничивающих прямоугольников), для этого вычисляется размер прямоугольника и суммарный вес контуров внутри него. Если

отношение веса контуров к размеру прямоугольника превышает 1, то это означает, что в прямоугольнике есть значительные контуры. Далее на основе карты контуров и матрицы меток определяется число контуров в ограничивающем прямоугольнике.

Для выделенных контуров находятся координаты их пикселей, после чего выполняется проверка, соприкасается ли найденный контур с ограничивающим прямоугольником. Если контур выходит за пределы прямоугольника или соприкасается с его границами, то его веса не включаются в оценку. Полученная оценка нормируется на размер ограничивающего прямоугольника и записывается в массив оценок, который сортируется по убыванию, при этом прямоугольники с минимальным уровнем относительно максимального значения отсекаются.

Шаг 3. Выполняется классификация объектов для каждой области интереса. После этапа оценки для первых 1000 ограничивающих прямоугольников применяется подавление немаксимумов. Для этого организовывается цикл по всем выделенным прямоугольникам и для текущего рассматриваются все прямоугольники, которые с ним пересекаются. Далее вычисляется площадь перекрытия прямоугольников, а также их суммарная площадь, за вычетом перекрытия, на основе чего рассчитывается значение IoU . Если значение IoU больше порогового значения β , то такой ограничивающий прямоугольник удаляется.

Далее полученные ограничивающие прямоугольники подаются на вход классификатора в виде CNN, который определяет, содержит ли данный кандидат объект или он является фоном.

В целом следует отметить, что модель R-CNN является достаточно медленной из-за того, что алгоритмы генерации областей интереса создают много ограничивающих прямоугольников, которые обрабатываются нейронной сетью на анализ которых затрачивается значительное время.

Модель Fast R-CNN. Несмотря на получаемые с ее использованием неплохие результаты, производительность R-CNN не столь высока, особенно для более глубоких,



Рис. 4. Этапы выполнения алгоритма region-proposal-function R-CNN

чем CaffeNet сетей (таких как VGG16). Кроме того, при использовании этой модели необходимо осуществлять хранение большого объема данных. Схема работы алгоритма для модели Fast R-CNN представлена на рис. 5.

Шаг 1. Выполняется генерация областей интереса (region proposals), предположительно содержащих в себе объект (например, алгоритм Edge Boxes).

Шаг 2. Выполняется формирование карты признаков для исходного изображения. Для этого на вход нейронной сети CNN подается полное исходное изображение, но при этом последний слой *max-pool* заменяется на слой *RoI pooling*. Слой *RoI pooling* принимает на вход карту признаков, полученную от последнего сверточного слоя нейронной сети, и сгенерированные области интереса (в координатах изображения). Область интереса преобразуется из координат изображения в координаты на карте признаков и на полученный прямоугольник накладывается сетка $W \times H$ с заданными размерами, область интереса шириной w и высотой h преобразуется в сетку, имеющую $H \times W$ ячеек размером $h/H \times w/W$ (например, для VGG16 используются $W = H = 7$). Таким образом слой *RoI pooling* преобразует вектор признаков произвольно-



Рис. 5. Алгоритм Fast R-CNN

го прямоугольника из исходного изображения в вектор признаков фиксированной размерности. Затем к каждой такой ячейке применяется *Max Pooling* для выбора только одного значения, таким образом, формируется результирующая матрица признаков $H \times W$.

Шаг 3. Выполняется уточнение границ области интереса при помощи регрессионной модели (Bounding Box Regression). Полученные области интереса и полученные вектора признаков подаются на вход двум новым полностью связным слоям. Первый из этих слоев используется для уточнения границ прямоугольника, а второй для классификации объекта, расположенного внутри этого прямоугольника (Шаг 4).

Для уточнения ограничивающего прямоугольника регрессионная модель (Bounding Box Regression) обучается корректировать прогнозируемую ограничивающую рамку с использованием функций CNN. При заданной предсказанной координате ограничительной рамки $p = (p_x, p_y, p_w, p_h)$, где p_x – координата x центра рамки, p_y – координата y центра рамки, p_h – ширина ограничительной рамки, p_w – длина ограничительной рамки, и соответствующих ей координатах пря-

моугольника истинности $g = (g_x, g_y, g_h, g_w)$, где g_x – координата x прямоугольника истинности, g_y – координата y прямоугольника истинности, g_h – ширина прямоугольника истинности, g_w – длина прямоугольника истинности регрессор настроен на масштаб-инвариантное преобразование между двумя центрами и лог-масштабное преобразование между шириной и высотой, отображенное на рис. 6.

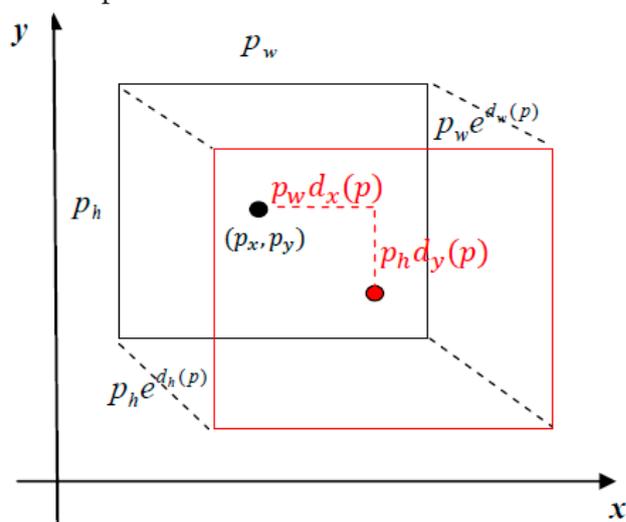


Рис 6. Преобразования между прогнозируемой и истинной ограничивающими рамками

$$g_x = p_w d_x(p) + p_x, \quad g_y = p_h d_y(p) + p_y, \\ g_w = p_w e^{d_w(p)}, \quad g_h = p_h e^{d_h(p)}, \quad (2)$$

где $d_x(p)$ и $d_y(p)$ – функции масштабно-инвариантного преобразования центра с координатами x и y , $d_w(p)$ и $d_h(p)$ – функции лог-масштабного преобразования между шириной w и высотой h .

Шаг 4. Выполняется классификация объектов, содержащихся в предполагаемых областях интереса. Для классификации объектов используется слой softmax с $K+1$ выходами (где $K+1$ – величина, характеризующая число классов объектов, с учетом наличия фона на исходном изображении). Шаг 3 и Шаг 4 выполняются параллельно.

На рис. 7 показан процесс поиска объектов на изображении с использованием сверточных нейронных сетей Fast R-CNN.

Модель Faster R-CNN. По сравнению с Fast R-CNN, в Faster R-CNN вместо использования внешнего алгоритма для генерации областей интереса, предположительно содержащих объект, (например, Edge Boxes) используется специальная сеть Region Proposal Network (RPN). Схема работы алгоритма для модели Faster R-CNN представлена на рис. 8.

Шаг 1. Выполняется формирование карты признаков на основе исходного изображения. Для этого исходное изображение подается на вход сверточной нейронной сети CNN, архитектура которой представлена на рис. 2.

Шаг 2. Выполняется генерация областей интереса, возможно содержащих объект. С использованием нейронной сети Region Proposal Network (RPN) сформированная карта признаков обрабатывается скользящим окном заданного размера 3×3 . Таким обра-

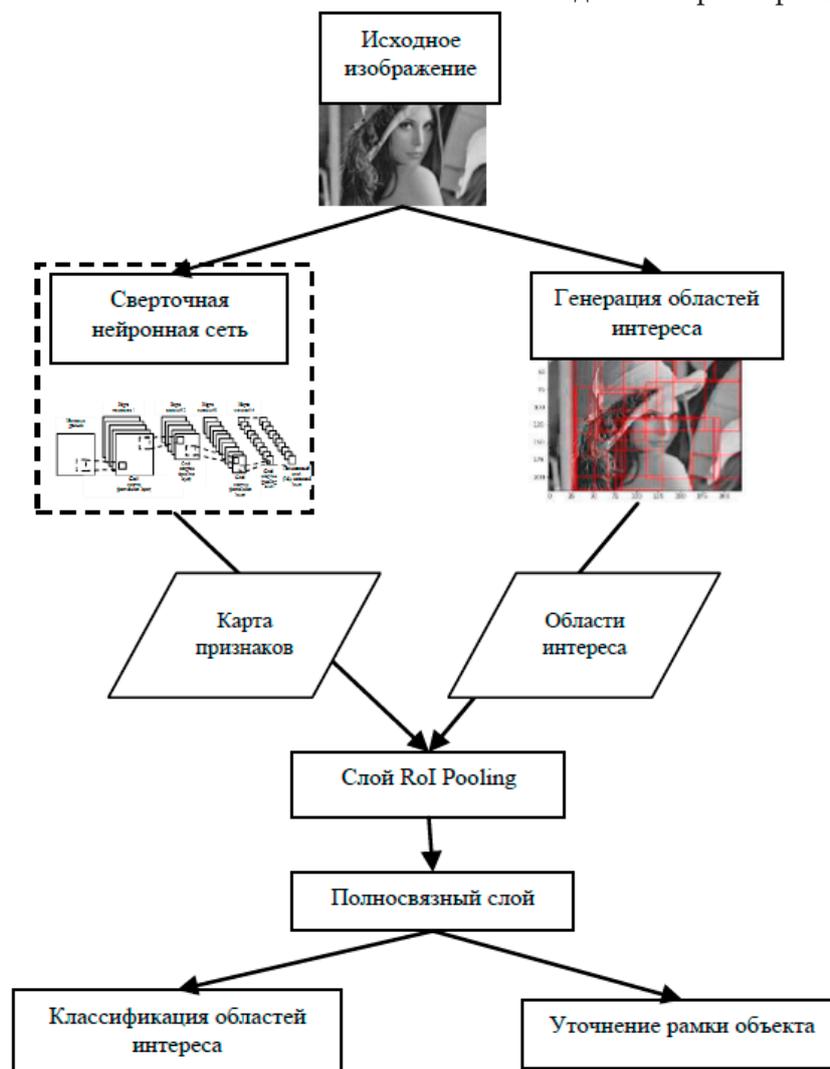


Рис 7. Поиск объектов на изображении, Fast R-CNN

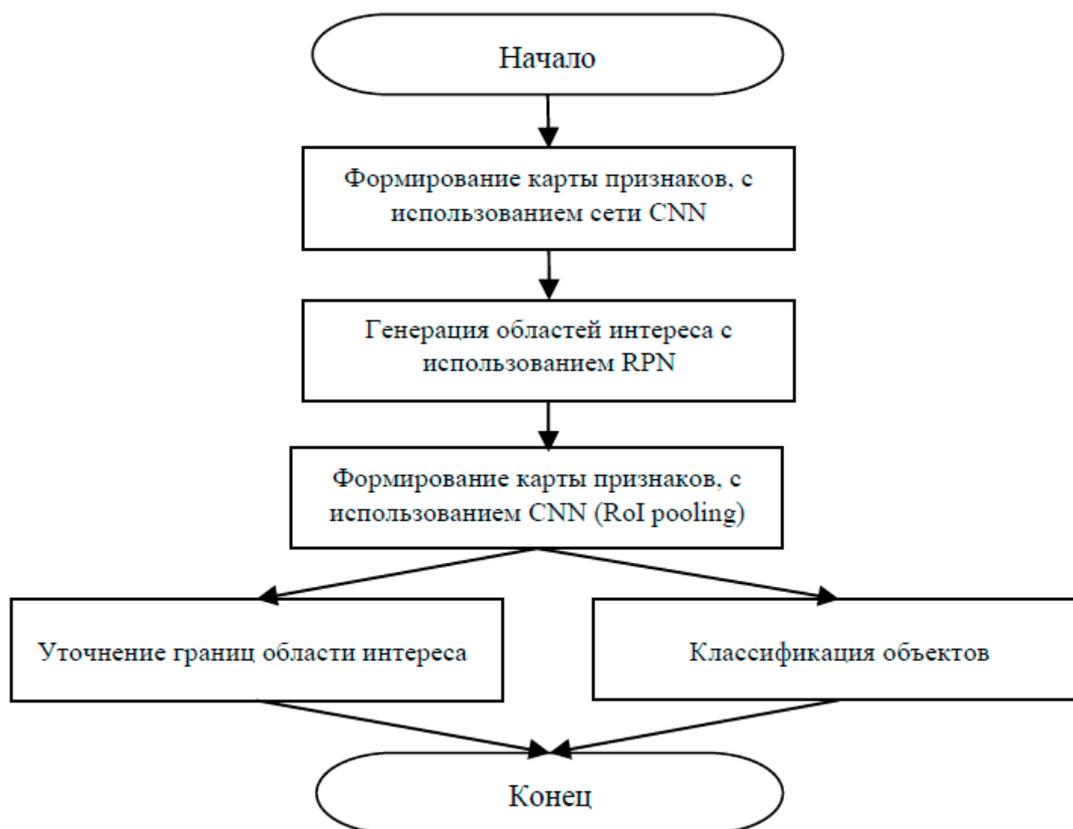


Рис. 8. Алгоритм Faster R-CNN

зом, генерируется набор из 9 областей интереса, которые имеют один и тот же центр, но с различными пропорциями и масштабами. Для каждой позиции окна извлекается вектор признаков малой размерности. Кроме того, для каждой из сформированных областей вычисляется значение IoU , на основе значения которого принимается решение о текущей области интереса.

Шаг 3. Выполняется преобразование вектора признаков области интереса (произвольного прямоугольника) из исходного изображения в вектор признаков фиксированной размерности, с помощью слоя *RoI pooling*.

Шаг 4. Выполняется уточнение границ области интереса при помощи регрессионной модели (Bounding Box Regression). Полученные области интереса и полученные вектора признаков подаются на вход двум новым полностью связным слоям. Первый из этих слоев используется для уточнения границ прямоугольника, а второй для классификации объекта, расположенного внутри этого прямоугольника (Шаг 5).

Шаг 5. Выполняется классификация объектов, содержащихся в предполагаемых областях интереса. Для классификации объектов используется слой softmax с $K+1$ выходами (где $K+1$ – величина, характеризующая число классов объектов, с учетом наличия фона на исходном изображении). Шаг 4 и Шаг 5 выполняются параллельно.

На рис. 9 показан процесс поиска объектов на изображении с использованием сверточных нейронных сетей Faster R-CNN.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЙ И ИХ ОБСУЖДЕНИЕ

В ходе исследований реализованы модели R-CNN, Fast-RCNN, Faster-RCNN с помощью Python 3.6 и использованием библиотеки машинного обучения Keras (версия 2.2.4) с поддержкой GPU. Обучение и тестирование нейронных сетей проводилось на данных без использования аугментации. В качестве показателей эффективности полученных моделей использовались время об-

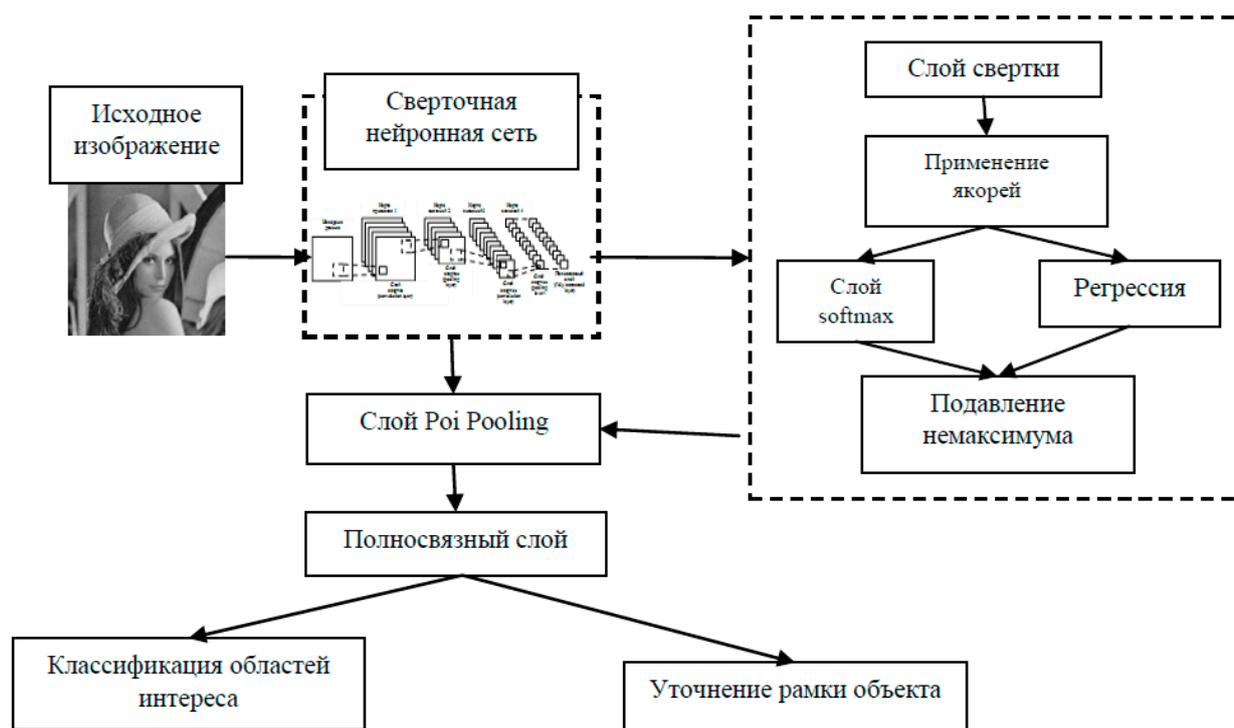


Рис. 9. Поиск объектов на изображении, Faster R-CNN

наружения и точность классификации объекта на изображении.

При обучении сетей была применена технология переноса обучения (transfer learning), позволяющая использовать готовые нейронные сети для решения задач нового типа, для которых сети предварительно не обучались. Для реализации переноса обучения использовались сверточные нейронные сети VGG16 и ResNet50, обрезанные до предпоследнего полносвязного слоя, с которого снимался вектор признаков размерности 4096 для VGG16 и 2048 для ResNet50. Полученный вектор подавался на новый полносвязный слой, отвечающий за классификацию объекта на изображении.

При реализации модели R-CNN с использованием в качестве классификатора нового полносвязного слоя с функцией активации softmax использовалась архитектура нейронной сети без уточнения ограничивающих прямоугольников (без регрессии).

В ходе проведения экспериментов исследовалось влияние двух алгоритмов генерации областей интереса, а именно Selective search и оригинального алгоритма Regional-proposal-function. Результаты проведенных экспери-

ментальных исследований представлены в табл. 1. Полученные результаты показывают, что время обнаружения объекта с использованием Regional-proposal-function меньше, чем с использованием Selective search. Объясняется это тем, что для генерации областей интереса при помощи первого алгоритма затрачивается меньше времени при прочих равных параметрах обучения.

Полученные результаты показывают, что предложенная модификация алгоритма Regional-proposal-function является более быстрой действующей, но в некоторых случаях результирующий результат его применения с нейронными сетями CNN оказывается несколько хуже.

Проводилось также исследование показателей качества сети в зависимости от метода оптимизации. Исследовалось влияние двух методов: SGD (стохастического градиентного спуска) и алгоритма Adam (алгоритм стохастической оптимизации), при этом для метода SGM менялся параметр скорости обучения сети. Результаты, приведенные в табл. 2, показывают, что при увеличении скорости обучения сети для метода SGD значительно снижается точность распознавания объекта на изображении.

Таблица 1

Результаты работы R-CNN (набор данных VOC2007, метод оптимизации Adam)

Алгоритм генерации областей интереса	Модель CNN	Количество эпох обучения	Точность распознавания на тестовой выборке	Время обнаружения объекта, сек
Regional-proposal-function	VGG16	50	0.7475	8.373
		100	0.8846	8.442
		150	0.8936	8.343
	ResNet50	50	0.8362	8.151
		100	0.9034	8.183
		150	0.9452	8.051
Selective search	VGG16	50	0.7962	10.840
		100	0.8699	10.701
		150	0.9198	10.866
	ResNet50	50	0.8279	10.715
		100	0.9358	10.322
		150	0.9681	10.376

Таблица 2

Результаты работы при изменении параметра скорости обучения сети (R-CNN, набор данных VOC2007, количество эпох обучения 150)

Метод стохастического градиентного спуска (SGD)		Точность распознавания на тестовой выборке
Скорость обучения	Импульс Нестерова	
0.01	–	0.9467
0.01	+	0.9639
0.05	+	0.5967

Таблица 3

Исследование влияния методов оптимизации (R-CNN, набор данных VOC2007, количество эпох обучения 150)

Метод SGD	Метод Adam	Точность распознавания на тестовой выборке	Время обнаружения объекта, сек
+	–	0.9467	10.583
–	+	0.9681	10.376

Результаты, представленные в табл. 3, показывают, что точность распознавания объекта при обучении сети методом Adam оказалась выше, чем при использовании метода SGD.

В табл. 4–6 представлены результаты работы моделей Fast R-CNN и Faster R-CNN, обученных методом Adam. Полученные результаты свидетельствуют о том, что модель Faster R-CNN превосходит по скорости работы на

тестовых изображениях другие архитектуры нейронных сетей.

Графики, представленные на рис. 10, иллюстрируют изменение функции потерь и значения точности классификации объекта при использовании Faster R-CNN, обученной на наборе данных COCO. Очевидно, что данная нейронная сеть достигает высоких показателей по точности классификации и низким значениям функции потерь.

Таблица 4

Результаты работы Fast R-CNN (набор данных VOC2007, метод оптимизации Adam)

Модель CNN	Количество эпох обучения	Точность распознавания на тестовой выборке	Время обнаружения объекта, сек
VGG16	50	0.7621	3.212
	100	0.8037	3.454
	150	0.9232	3.294
ResNet50	50	0.8372	3.345
	100	0.9224	3.664
	150	0.9528	3.552

Таблица 5

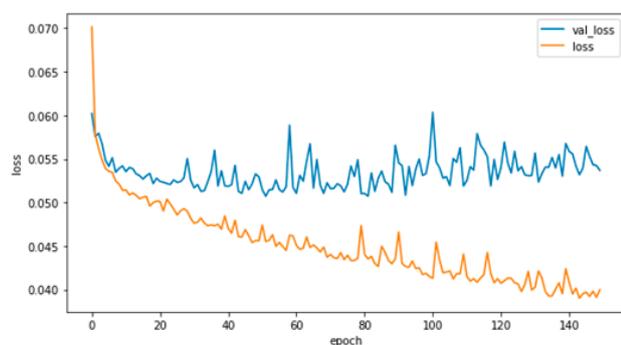
Результаты работы Faster R-CNN (набор данных VOC2007, метод оптимизации Adam)

Модель CNN	Количество эпох обучения	Точность распознавания на тестовой выборке	Время обнаружения объекта, сек
VGG16	50	0.6882	1.283
	100	0.7325	1.364
	150	0.8537	1.298
ResNet50	50	0.7851	1.382
	100	0.8683	1.482
	150	0.9028	1.372

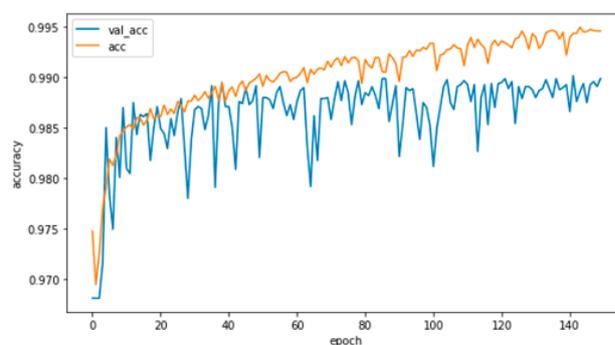
Таблица 6

Результаты работы Faster R-CNN (набор данных COCO, метод оптимизации Adam)

Модель CNN	Количество эпох обучения	Точность распознавания на тестовой выборке	Время обнаружения объекта, сек
VGG16	50	0.8867	1.684
	100	0.9054	1.627
	150	0.9547	1.739
ResNet50	50	0.9234	1.542
	100	0.9648	1.536
	150	0.9832	1.622



(а)



(б)

Рис. 10. Faster R-CNN (а) – график функции потерь
(б) – график точности распознавания объекта

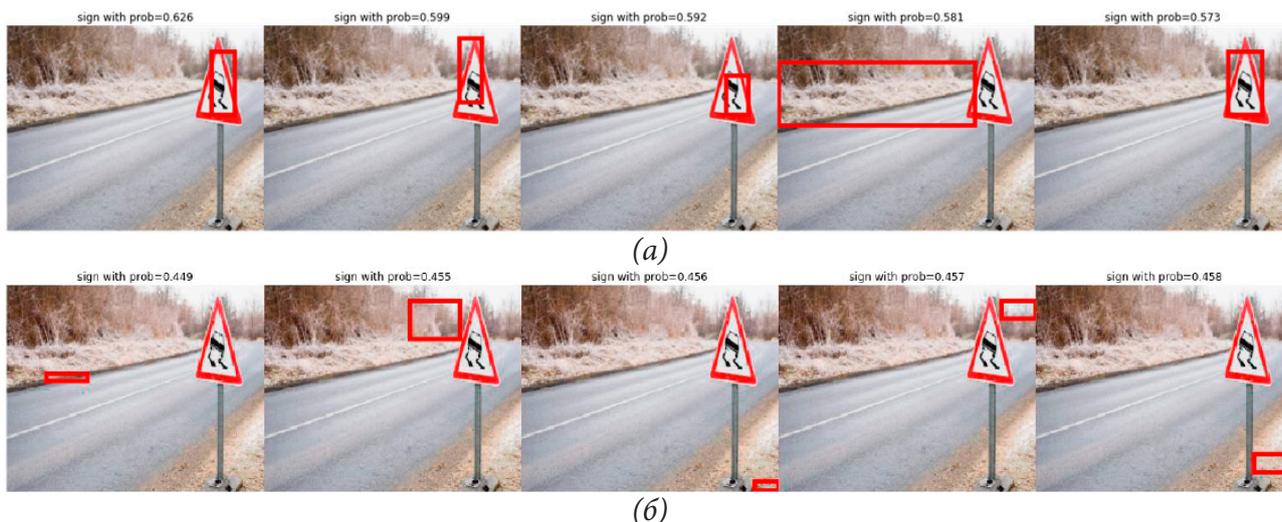


Рис. 11. Пример обнаружения объекта на изображении при помощи R-CNN: (а) – первые пять примеров расположения ограничивающего прямоугольника с наибольшей вероятностью содержащего объект, (б) – последние пять примеров расположения ограничивающего прямоугольника с наименьшей вероятностью содержащей объект

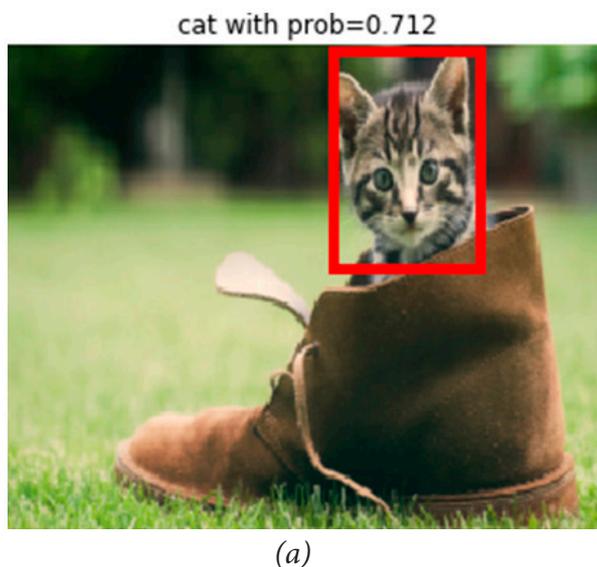


Рис. 12. Пример детектирования объекта на изображении (а) при помощи Fast R-CNN, (б) при помощи Faster R-CNN

На рис. 11, 12 представлены примеры обнаружения объектов (дорожный знак).

ЗАКЛЮЧЕНИЕ

В статье рассмотрены задачи поиска и классификации объектов на изображениях с использованием моделей нейронных сетей R-CNN, Fast R-CNN, Faster R-CNN. В процессе исследования проведен сравнительный анализ реализованного в рамках модели R-CNN алгоритма regional-proposal с алгоритмом селективного поиска. Исходя из представленных

результатов, можно сделать вывод, что использование предложенного подхода к генерации областей интереса ускоряет процесс поиска объекта на изображении. Результаты тестирования работы модификаций различных моделей R-CNN показывают, что модель существенно Faster R-CNN превосходит по скорости другие модели из-за использования в качестве алгоритма генерации областей интереса специальной сети формирования соответствующих предложений. В целом, полученные результаты моделирования и тестирования приведенных моделей свидетель-

ствуют о высокой точности классификации и обнаружения объектов на изображении для всех рассмотренных алгоритмов.

СПИСОК ЛИТЕРАТУРЫ

1. Бутенко, В. В. Поиск объектов на изображении с использованием алгоритма адаптивного усиления / В. В. Бутенко // Молодой ученый. – 2015. – №4. – С. 52–56.
2. Борисов, Е. О. О задаче поиска объекта на изображении. Часть 2: Применение методов машинного обучения. [Электронный ресурс] / Е. О. Борисов // Обработка изображений. – Режим доступа: <http://mechanoid.kiev.ua/cv-image-detector2.html>. – (Дата обращения 15.07.2019)
3. Dalal, N. Histograms of oriented gradients for human detection / N. Dalal, B. Triggs // IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA. – 2005. – № 1. – С. 886–893.
4. Артемов, А. А. Проблема поиска объектов на изображениях с помощью компьютерного зрения на основе информации о цвете / А. А. Артемов, М. В. Кавалеров, Г. С. Кузнецов // Вестник ПНИПУ. Электротехника, информационные технологии, системы управления. – 2011. – № 5.
5. Акимов, А.В. Модели и алгоритмы искусственного размножения данных для обучения алгоритмов распознавания лиц методом Виолы-Джонса / А. В. Акимов, А. А. Сирота // Компьютерная оптика. – 2016. – № 6. – С. 899–906.
6. Viola, P. Robust real time face detection / P. Viola., M. Jones // International Journal of Computer Vision. – 2004. – № 57(2). – С. 137–154.
7. Нгуен, Т. Т. Алгоритмическое и программное обеспечение для распознавания фигур с помощью Фурье – дескрипторов и нейронной сети / Т. Т. Нгуен // Известия Томского политехнического университета. – 2010. – № 5. – С. 122–125.
8. R-CNN, Fast R-CNN, Faster R-CNN, YOLO–Object Detection Algorithms [Электронный ресурс]. – Режим доступа: <https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e>. – (Дата обращения 17.07.2019).
9. Girshick, R. Region-Based Convolutional Networks for Accurate Object Detection and Segmentation / R. Girshick, J. Darrell, T. Malik // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2015. – № 38.
10. Girshick, R. Fast R-CNN / R. Girshick // International Conference on Computer Vision (ICC). – 2015.
11. Girshick, R. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks / R. Girshick, R. Shaoqing, H. Kaiming // Neural Information Processing Systems (NIPS). – 2015.
12. Mask R-CNN: архитектура современной нейронной сети для сегментации объектов на изображениях. – Режим доступа: <https://habr.com/ru/post/421299>. – (Дата обращения 17.07.2019).
13. Wang, Yuanyuan. Automatic Ship Detection Based on RetinaNet Using Multi-Resolution / Y. Wang, C. Wang, H. Zhan, G. Yingbo, S. Wei // Remote Sensing. – 2019. – № 11. – С. 531
14. The PASCAL Visual Object Classes Challenge 2007 (VOC2007). – Режим доступа: <http://host.robots.ox.ac.uk/pascal/VOC/voc2007/>. – (Дата обращения 17.07.2019).
15. Microsoft COCO: Common Objects in Context. – Режим доступа: <http://cocodataset.org/#home>. – (Дата обращения 17.09.2019).
16. Zitnick, C. Edge boxes: Locating object proposals from edges / C. Zitnick, P. Dollár. // Computer Vision. - 2014. - С. 391–405.
17. Sande, J. Selective Search for Object Recognition / J. Sande, K. Gevers, T. Smeulders // International Journal of Computer Vision. – 2013. – №104. – С. 154.

Сирота Александр Анатольевич – профессор, зав. кафедрой технологий обработки и защиты информации Воронежского государственного университета, e-mail: sir@cs.vsu.ru

Митрофанова Елена Юрьевна – доцент кафедры технологий обработки и защиты информации Воронежского государственного университета, e-mail: mitrofanova.e_yu@mail.ru

Милованова Анастасия Ивановна – магистрант 2 года обучения по программе «Безопасность информационных систем» факультета компьютерных наук Воронежского государственного университета, e-mail: anastasi.milovanova@gmail.com

ANALYSIS OF ALGORITHMS FOR SEARCHING OBJECTS IN IMAGES USING VARIOUS MODIFICATIONS OF CONVOLUTIONAL NEURAL NETWORK

A. A. Sirota, E. Yu. Mitrofanova, A. I. Milovanova

Voronezh State University

Annotation. The article deals with the tasks of classifying and detecting objects in images using computer vision algorithms. The description of the main types of algorithms and methods for searching objects based on the use of deep neural networks is given. A comparative analysis and modeling of neural network algorithms for solving the problem of classification and search for objects in images has been carried out. The results of testing neural network models with different architectures on VOC2007 and COCO data sets are given. The results of the study of recognition accuracy are analyzed depending on various learning hyper-parameters. The change in the time value of determining the location of an object depending on various neural network architectures is investigated.

Keywords: object detection, object classification, neural networks, deep learning, RCNN, Fast RCNN, Faster RCNN.

Sirota Alexander Anatolievich – prof., Head of the Chair of Information Processing and Security Technologies at Voronezh State University, e-mail: sir@cs.vsu.ru

Mitrofanova Elena Yurievna – docent at the Chair of Information Processing and Security Technologies at Voronezh State University, e-mail: mitrofanova.e_yu@mail.ru

Milovanova Anastasia Ivanovna – Master of 2 year training program «Information Systems Security», Computer Sciences Faculty, Voronezh State University, e-mail: anastasi.milovanova@gmail.com