

СТАТЬЯ

УДК 004.93

**МОДЕЛЬ ПОГРАНИЧНЫХ БЛОКОВ ДЛЯ РАСПОЗНАВАНИЯ ОБЪЕКТОВ**

**Титова А.А., Козина А.В., Белов Ю.С.**

*Московский государственный технический университет имени Н.Э. Баумана, Калужский филиал,  
Калуга, e-mail: fn1-kf@mail.ru*

Использование модели пограничных блоков для распознавания объектов является эффективным новым подходом к увеличению вычислительной эффективности их обнаружения. В данной статье описывается метод подбора ограничивающего окна объекта с вычислением границ краев окна для того, чтобы найти и выделить объект. Края обеспечивают разреженное, но информативное представление для изображения, выделяя не сам объект, но заключая его в прямоугольную область. Главное наблюдение состоит в том, что количество контуров, которые полностью содержатся в ограничивающем прямоугольнике, указывает на вероятность того, что прямоугольник содержит объект. Используя эффективные структуры данных, миллионы блоков кандидатов могут быть оценены за доли секунды, возвращая ранжированный набор из нескольких тысяч предложений с наивысшими оценками. Цель данной статьи – описать возможности модели пограничных блоков и дать более углубленное знание процесса распознавания и поиска объектов на изображении, особенностей алгоритма данной модели. В этой работе описаны различия между классическими моделями поиска объекта на изображении, а также показаны особенности метода пограничных блоков, его сильные и слабые стороны.

**Ключевые слова:** распознавание объектов, компьютерное зрение, глубокое обучение, сверточные нейронные сети

**THE EDGE BLOCK MODEL TO RECOGNIZE OBJECTS**

**Titova A.A., Kozina A.V., Belov Yu.S.**

*Bauman Moscow State Technical University, Kaluga branch, Kaluga, e-mail: fn1-kf@mail.ru*

Using the boundary block model for object recognition is an effective new approach to increasing the computational efficiency of object detection. This article describes a method for fitting the bounding box of an object and calculating the bounding box edges in order to find and select an object. Edges provide a sparse but informative representation of the image, not highlighting the object itself, but enclosing it in a rectangular area. The main observation is that the number of paths that are entirely contained in the bounding box indicates the likelihood that the box contains an object. Using efficient data structures, millions of candidate blocks can be evaluated in fractions of a second, returning a ranked set of several thousand highest-scored proposals. The purpose of this article is to describe the capabilities of the boundary block model and give a more in-depth knowledge of the process of recognizing and searching for objects in the image, the features of the algorithm of this model. This paper describes the differences between the classical models of searching for an object in the image, as well as shows the features of the boundary block method, its strengths and weaknesses.

**Keywords:** object recognition, computer vision, deep learning, convolutional neural networks

Цель обнаружения объекта – определить, существует ли объект на изображении, и если да, то в каком месте изображения он встречается. Доминирующим подходом к этой проблеме в последнее десятилетие была парадигма скользящих окон, в которой классификация объектов выполняется в каждом месте и в любом масштабе изображения [1]. Недавно была предложена альтернативная структура для обнаружения объектов. Вместо поиска объекта в каждом местоположении и масштабе изображения сначала генерируется набор предложений ограничивающей рамки объекта с целью сокращения набора позиций, которые необходимо дополнительно проанализировать [2]. Замечательные открытия, сделанные вышеперечисленными подходами, состоит в том, что предложения объектов могут быть точно созданы способом, который не зависит от типа обнаруживаемого объекта [3].

Высокая отзывчивость и эффективность – важнейшие свойства генератора

предложений объекта [4]. Если предложение не генерируется в непосредственной близости от объекта, этот объект не может быть обнаружен. Эффективный генератор способен обеспечить высокую степень отклика с использованием относительно небольшого количества ограничивающих рамок кандидатов, обычно от сотен до нескольких тысяч на изображение.

Цель исследования: описать модель пограничных блоков, используемый алгоритм для процесса поиска объекта на изображении. Описать различия между классическими моделями поиска объекта на изображении, а также показать особенности метода пограничных блоков, его сильные и слабые стороны.

*Разбор модели пограничных блоков.* Предложения по объектам ранжируются на основе единой оценки, рассчитанной на основе контуров, полностью заключенных в ограничивающую рамку кандидата [5]. Все начинается с описания структу-

ры данных, основанной на группах ребер, которая позволяет эффективно отделить контуры, полностью заключенные в рамку, от контуров, которые выходят за ее пределы. Затем определяется функция оценки на основе ребер. Наконец, детализируется подход к поиску предложений по объектам с самым высоким рейтингом, в котором используется структура скользящего окна, оцениваемая по положению, масштабу и соотношению сторон, с последующим уточнением с использованием простого грубого поиска [6].

Учитывая изображение, сначала вычисляется краевой отклик для каждого пикселя. Отклики краев обнаруживаются с помощью детектора структурированных краев, который показал хорошую производительность при прогнозировании границ объекта и в то же время очень эффективен [7]. Используется одномасштабный вариант с улучшением резкости, чтобы сократить время выполнения. Учитывая плотные отклики краев, выполняется не максимальное подавление (NMS), ортогональное отклику края на пики края, рис. 1. Результатом является карта разреженных краев, где каждый пиксель  $p$  имеет величину края  $m_p$  и ориентацию  $\theta_p$ . Края определяются как пиксели с  $m_p > 0.1$ . Контур определяется как набор

ребер, образующих когерентную границу, кривую или линию.

*Пограничные группы и индекс соответствия.* Как показано на рис. 1, цель – идентифицировать контуры, которые перекрывают границу ограничивающего прямоугольника и поэтому вряд ли принадлежат объекту, содержащемуся в ограничивающем прямоугольнике. Для данного блока  $b$  мы идентифицируем эти ребра, вычисляя для каждого  $p \in b$  с  $m_p > 0.1$  его максимальное число с ребром на границе блока. Интуитивно понятно, что ребра, соединенные прямыми контурами, должны иметь большее значение, тогда как те, которые не соединены или не соединены контуром с высокой кривизной, должны иметь меньшее значение. Для вычислительной эффективности было обнаружено, что выгодно группировать ребра, которые имеют большее значение, и вычислять только единицы между группами ребер. Далее формируются группы ребер, используя простой жадный алгоритм, который объединяет 8-соединенные ребра до суммы их ориентации, где разница превышает пороговое значение ( $\pi/2$ ). Небольшие группы объединяются с соседними группами. Иллюстрация групп кромок показана на рис. 1 в строке 3.



Рис. 1. Примеры, показывающие (сверху вниз) первоначальное изображение, структурированные части, сгруппированные части, правильные окна ограничения и маркировка краев и неправильные

Учитывая набор групп ребер  $s_i \in S$ , вычисляется число между каждой парой соседних групп. Для пары групп  $s_i$  и  $s_j$  индекс соответствия вычисляется на основе их средних положений  $x_i$  и  $x_j$  и средних ориентаций  $i$  и  $j$ . Интуитивно понятно, что группы ребер имеют большее число, если угол между средними значениями групп аналогичен ориентации групп. В частности, вычисляется индекс соответствия  $a(s_i, s_j)$ , используя

$$a(s_i, s_j) = |\cos(\theta_i - \theta_{ij}) \cos(\theta_j - \theta_{ij})|^\gamma, \quad (1)$$

где  $\theta_{ij}$  – угол между  $x_i$  и  $x_j$ . Значение  $\gamma$  может использоваться для настройки чувствительности индекса соответствия к изменению ориентации, на практике  $\gamma = 2$ . Если две группы краев разделены более чем на 2 пикселя, их значение равно нулю. Для повышения вычислительной эффективности сохраняются только значения выше небольшого порога (0,05), а остальные принимаются равными нулю [8].

*Оценка ограничивающей рамки.* Учитывая набор групп ребер  $S$  и их индекс соответствия, можно вычислить оценку предложения объекта для любого ограничивающего прямоугольника  $b$ -кандидата. Чтобы определить оценку, сначала вычисляется сумма величин  $m_p$  для всех ребер  $p$  в группе  $s_i$ . Также выбирается произвольная позиция  $x_i$  некоторого пикселя  $p$  в каждой группе  $s_i$ . Далее станет ясно, что точный выбор  $p \in s_i$  не имеет значения.

Для каждой группы  $s_i$  мы вычисляем непрерывное значение  $\omega_b(s_i) \in [0, 1]$ , которое указывает, содержится ли  $s_i$  полностью в  $b$ ,  $\omega_b(s_i) = 1$  или нет,  $\omega_b(s_i) = 0$ . Пусть  $S_b$  будет набором групп ребер, которые перекрывают границу бокса  $b$ . Необходимо найти  $S_b$ , используя эффективную структуру данных, представленную на рис. 2. Для всех  $s_i \in S_b$ ,  $\omega_b(s_i)$  устанавливается значение в 0. Аналогично  $\omega_b(s_i) = 0$  для всех  $s_i$  для которых  $\bar{x}_i \notin b$ , так как все его пиксели должны либо находиться вне  $b$ , либо  $s_i \in S_b$ . Для остальных групп ребер, для которых

$\bar{x}_i \in b$  и  $s_i \notin S_b$ , вычисляется  $\omega_b(s_i)$  следующим образом:

$$\omega_b(s_i) = 1 - \max_T \prod_j^{|T|-1} a(t_j, t_{j+1}), \quad (2)$$

где  $T$  – упорядоченный путь групп ребер длиной  $|T|$ , который начинается с некоторого  $t_1 \in S_b$  и заканчивается на  $t_{|T|} \in s_i$ . Если такого пути не существует, определяется  $\omega_b(s_i) = 1$ . Таким образом, уравнение (2) находит путь с наибольшим индексом соответствия между группой ребер  $s_i$  и группой ребер, которая перекрывает границу окна. Используя вычисленные значения  $\omega_b$ , определяется нужная оценка следующим образом:

$$h_b = \frac{\sum_i \omega_b(s_i) m_i}{2(b_w + b_h)^k}, \quad (3)$$

где  $b_w$  и  $b_h$  – ширина и высота бокса. Также следует обратить внимание, что деление происходит по периметру окна, а не по его площади, так как края имеют ширину в один пиксель независимо от масштаба. Тем не менее значение  $k = 1.5$  используется для компенсации смещения больших окон, имеющих в среднем больше краев.

На практике используется интегральное изображение, чтобы ускорить вычисление числителя в уравнении (3). Интегральное изображение используется для вычисления суммы всех  $m_p$  для которых  $\bar{x}_i \in b$ . Затем для всех  $s_i$  с  $\bar{x}_i \in b$  и  $\omega_b(s_i) < 1$  из этой суммы вычитается  $(1 - \omega_b(s_i)) m_i$ .

*Поиск пересекающихся групп ребер.* В предыдущем разделе предполагалось, что известен набор групп ребер  $S_b$ , которые перекрывают границу блока  $b$ . Поскольку оценивается огромное количество ограничивающих прямоугольников, эффективный метод поиска  $S_b$  имеет решающее значение. Простые подходы, такие как исчерпывающий поиск всех пикселей на границе поля, были бы непомерно дорогими, особенно для больших блоков [9, 10].

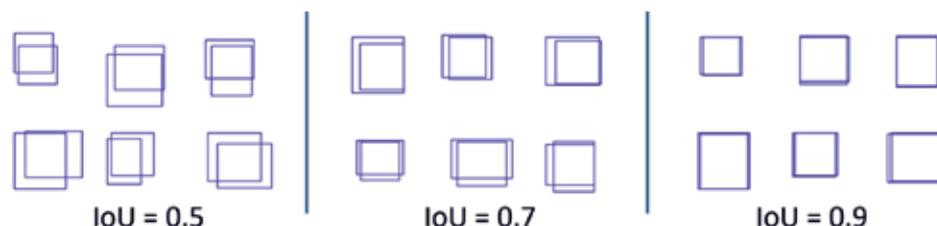


Рис. 2. Примеры случайных ограничивающих прямоугольников

Существует эффективный метод поиска пересекающихся групп ребер для каждой стороны ограничивающего прямоугольника, основанный на двух дополнительных структурах данных [11]. Ниже описывается процесс поиска пересечений вдоль горизонтальной границы от пикселя  $(c_0, r)$  до  $(c_1, r)$ . С вертикальными границами можно работать аналогичным образом. Для горизонтальных границ создаются две структуры данных для каждой строки изображения. В первой структуре данных хранится упорядоченный список  $L_r$  индексов групп ребер для строки  $r$ . Список создается путем сохранения порядка, в котором группы ребер располагаются вдоль строки  $r$ . Индекс добавляется к  $L_r$  только в том случае, если индекс группы краев изменяется от одного пикселя к другому. В результате размер  $L_r$  намного меньше ширины изображения. Если между группами краев есть пиксели, которые не являются краями, в список добавляется ноль. Создается вторая структура данных  $K_r$  с таким же размером, как ширина изображения, которая сохраняет соответствующий индекс в  $L_r$  для каждого столбца  $c$  в строке  $r$ . Таким образом, если пиксель  $p$  в местоположении  $(c, r)$  является членом группы краев  $s_r$ ,  $L_r(K_r(c)) = i$ . Поскольку большинство пикселей не принадлежат группе краев, используя эти две структуры данных, можно эффективно найти список перекрывающихся групп краев, выполняя поиск  $L_r$  от индекса  $K_r(c_0)$  до  $K_r(c_1)$ .

*Поиск стратегии.* При поиске предложений по объектам следует учитывать алгоритм обнаружения объекта. Некоторые алгоритмы обнаружения могут требовать предложения объектов с высокой точностью, в то время как другие более терпимы к ошибкам при размещении ограничивающей рамки. Точность ограничивающего прямоугольника обычно измеряется с помощью метрики «Пересечение по объему» (IoU). IoU вычисляет пересечение прямоугольника-кандидата и основного прямоугольника истинности, деленное

на площадь их объединения. При оценке алгоритмов обнаружения объектов порог IoU, равный 0,5, обычно используется для определения правильности обнаружения. Однако, как показано на рис. 2, оценка IoU 0,5 является довольно слабой. Даже если предложение объекта сгенерировано с IoU 0,5 с достоверностью, алгоритм обнаружения может дать низкую оценку. В результате обычно желательны оценки IoU более 0,5.

Необходимо описать стратегию поиска предложения объекта на основе желаемого IoU,  $\delta$ , для детектора. Для больших значений  $\delta$  создается более концентрированный набор ограничивающих рамок с более высокой плотностью. Для мелких значений  $\delta$  окна могут иметь большее разнообразие, поскольку предполагается, что детектор объектов может учитывать умеренные ошибки в расположении окон. Важно, что предыдущие методы имеют неявные предубеждения, для которых  $\delta$  предназначены, например Objectness и Randomized Prim настроены на низкий и высокий показатель  $\delta$  соответственно, тогда как в данном случае обеспечивается явный контроль над разнообразием по сравнению с точностью.

Начинается поиск возможных ограничивающих рамок с помощью скользящего окна поиска по положению, масштабу и соотношению сторон. Размер шага для каждого определяется с использованием одного параметра  $\alpha$  указывающего IoU для соседних блоков. То есть размеры шага при перемещении, масштабе и соотношении сторон определяются таким образом, что один шаг приводит к тому, что соседние блоки имеют IoU, равное  $\alpha$ . Значения масштаба варьируются от минимальной области блока  $\delta = 1000$  пикселей до полного изображения. Соотношение сторон от  $1/\tau$  до  $\tau$ , где  $\tau = 3$  используется на практике. Значение  $\alpha = 0.65$  идеально подходит для большинства значений  $\delta$ . Однако, если требуется высокая точность  $\delta > 0.9$ , можно увеличить до 0,85.

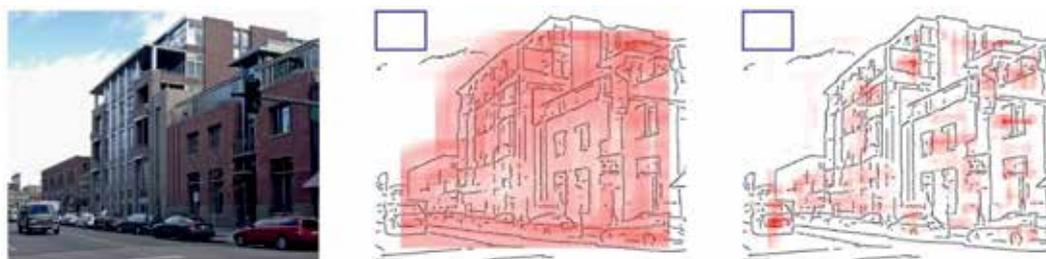


Рис. 3. Пример (оригинальный рисунок слева) оценки с использованием (в центре) и удалением (справа) контуров, перекрывающих границу ограничивающего окна

После выполнения поиска в скользящем окне все местоположения ограничивающего прямоугольника с оценкой  $h_b^{in}$  выше небольшого порога уточняются. Уточнение выполняется с помощью жадного итеративного поиска, чтобы максимизировать  $h_b^{in}$  по положению, масштабу и соотношению сторон. После каждой итерации шаг поиска сокращается вдвое. Поиск останавливается, когда размер шага перевода меньше 2 пикселей.

После того как ограничивающие прямоугольники-кандидаты улучшены, их максимальные оценки записываются и сортируются. Заключительный этап выполняет не максимальное подавление (NMS) отсортированных окон. Окно удаляется, если его IoU больше, чем у окна с большим количеством значений. Было обнаружено, что на практике установка  $\beta = \delta + 0.05$  обеспечивает высокую точность по всем значениям  $\delta$ .

### Заключение

В данной статье была рассмотрена классическая модель работы поиска объектов с помощью скользящих окон, описана модель пограничных блоков для распознавания объектов на изображении и используемый для этого алгоритм поиска. Также были описаны различия между классическими моделями поиска объектов на изображении и показаны плюсы и минусы описываемого в статье метода.

### Список литературы

1. Клетте Р. Компьютерное зрение. Теория и алгоритмы: учебник / Пер. с англ. А.А. Слинкина. М.: ДМК Пресс, 2019. 506 с.
2. Гизатуллин З.М., Ляшева С.А., Морозов О.Г. Метод обнаружения контуров на основе весовой модели изображения // Компьютерная оптика. 2020. № 3. С. 393–400.
3. Захаров А.А., Баринев А.Е., Жизняков А.Л. Поиск объектов на изображениях с использованием структурного дескриптора на основе графов // Компьютерная оптика. 2018. № 2. С. 283–292.
4. Боков П.А. Экспериментальный анализ точности и производительности разновидностей архитектур уоу для задач компьютерного зрения // Программные продукты и системы. 2020. № 4. С. 635–640.
5. Корсунов Н.И. Метод классификации изображений на основе кластеризации сложных объектов // Научные ведомости Белгородского государственного университета. Серия: Экономика. Информатика. 2016. № 23. С. 100–103.
6. Гудфеллоу Я., Бенджио И., Курвилль А. Глубокое обучение / Пер. с англ. А.А. Слинкина. 2-е изд. М.: ДМК Пресс, 2018. 652 с.
7. Шепелев К.В. Детектирование и классификация движущихся объектов в видеопоследовательности // Известия высших учебных заведений. Поволжский регион. Технические науки. 2017. № 4. С. 45–54.
8. Петрин Д.А., Белов Ю.С. Улучшение качества моделей машинного обучения в задачах классификации изображений на основе подходов извлечения признаков и точной настройки модели // Электронный журнал: наука, техника и образование. 2020. № 1 (28). С. 104–111.
9. Манокова Н.В. Компьютерное зрение как средство извлечения информации из видеоряда // Математические структуры и моделирование. 2015. № 4. С. 123–128.
10. Шапиро Л., Стокман Д. Компьютерное зрение: учебное пособие / Под ред. С.М. Соколова; пер. с англ. А.А. Богуславского. 4-е изд. М.: Лаборатория знаний, 2020. 763 с.
11. Смольянинов В.А., Белов Ю.С. Сравнительный анализ алгоритмов поиска дорожных знаков // Высокие технологии и инновации в науке: сборник избранных статей Международной научной конференции. 2020. С. 186–190.