

ISSN 1816-0301 (Print)
ISSN 2617-6963 (Online)

ОБРАБОТКА СИГНАЛОВ, ИЗОБРАЖЕНИЙ И РАСПОЗНАВАНИЕ ОБРАЗОВ
SIGNAL AND IMAGE PROCESSING AND PATTERN RECOGNITION

УДК 004.9
<https://doi.org/10.37661/1816-0301-2020-17-2-7-16>

Поступила в редакцию 25.02.2020
Received 25.02.2020

Принята к публикации 04.04.2020
Accepted 04.04.2020

Обнаружение объектов на изображениях с большим разрешением на основе их пирамидально-блочной обработки

Р. П. Богуш¹✉, И. Ю. Захарова¹, С. В. Абламейко^{2,3}

¹Полоцкий государственный университет, Новополоцк, Беларусь
✉E-mail: bogushr@mail.ru

²Белорусский государственный университет, Минск, Беларусь

³Объединенный институт проблем информатики
Национальной академии наук Беларуси, Минск, Беларусь

Аннотация. Предлагается алгоритм для обнаружения объектов на изображениях с большим разрешением, основанный на многомасштабном представлении изображения, пирамидально-блочной обработке с перекрытием, применении сверточной нейронной сети для каждого блока и объединении обнаруженных областей. Количество слоев пирамиды определяется размерами изображения и входного слоя используемой сверточной нейронной сети. На всех уровнях, кроме самого верхнего, выполняется блочное разбиение, а применение при этом перекрытия позволяет улучшить правильную классификацию объектов, которые разделяются на фрагменты и расположены в соседних блоках. Решение об объединении таких областей принимается на основе анализа метрики пересечения над объединением для них и принадлежности к одному классу. Представленные результаты тестирования алгоритма подтверждают, что рассмотренный подход позволяет повысить точность обнаружения объектов небольших размеров на изображениях с большим разрешением.

Ключевые слова: сверточная нейронная сеть, блочная обработка, разрешение 4К, обнаружение объектов, многомасштабное представление изображения

Для цитирования. Богуш, Р. П. Обнаружение объектов на изображениях с большим разрешением на основе их пирамидально-блочной обработки / Р. П. Богуш, И. Ю. Захарова, С. В. Абламейко // Информатика. – 2020. – Т. 17, № 2. – С. 7–16. <https://doi.org/10.37661/1816-0301-2020-17-2-7-16>

Object detection in high resolution images based on multiscale and block processing

Rykhard P. Bohush¹✉, Iryna Yu. Zakharava¹, Sergey V. Ablameyko^{2,3}

¹Polotsk State University, Novopolotsk, Belarus
✉E-mail: bogushr@mail.ru

²Belarusian State University, Minsk, Belarus

³The United Institute of Informatics Problems of the National Academy of Sciences of Belarus, Minsk, Belarus

Abstract. In the paper the algorithm for object detection in high resolution images is proposed. The approach uses multiscale image representation followed by block processing with the overlapping value. For each block the object detection with convolutional neural network was performed. Number of pyramid layers is limited by

the Convolutional Neural Network layer size and input image resolution. Overlapping blocks splitting to improve the classification and detection accuracy is performed on each layer of pyramid except the highest one. Detected areas are merged into one if they have high overlapping value and the same class. Experimental results for the algorithm are presented in the paper.

Keywords: convolutional neural network, block processing, 4K resolution, object detection, multiscale image representation

For citation. Bohush R. P., Zakharava I. Yu., Ablameyko S. V. Object detection in high resolution images based on multiscale and block processing. *Informatics*, 2020, vol. 17, no. 2, pp. 7–16 (in Russian). <https://doi.org/10.37661/1816-0301-2020-17-2-7-16>

Введение. Развитие цифровых видеокамер привело к тому, что на практике все шире используются устройства, позволяющие формировать изображения и видео с большим разрешением формата 4K×2K [1]. Большая сторона изображений при этом может принимать размеры от 3600 до 4500 пикселей для разрешения 4K [2]. На таких изображениях значительно увеличивается качество отображения объектов, в том числе небольших размеров либо удаленных при съемке на значительное расстояние от видеокамеры, и, соответственно, предоставляется возможность автоматического обнаружения объектов, которые при меньшем разрешении не обладают достаточной информативностью для решения этой задачи.

Можно выделить два подхода к обнаружению объектов: на основе сопоставления с эталонным набором признаков и на основе использования сверточных нейронных сетей (СНС). Первый подход предполагает сравнение входного изображения с признаками заранее подготовленных искомым изображений, в качестве которых могут выступать как уровни яркостей каждого пиксела, так и признаки, сформированные по определенным правилам, например гистограммы ориентированных градиентов, признаки Хаара и др. Поскольку выбор оптимального набора информативных признаков для точного описания произвольных изображений является не решенной в полной мере задачей, то при таком подходе требуются разработка и тестирование алгоритмов формирования дескрипторов объектов, ориентированных на заданные изображения. Поэтому в последнее время благодаря стремительному развитию вычислительной мощности компьютерной техники все шире используется другой подход, который основан на применении СНС, предварительно обученных на больших базах данных изображений. Наличие операций свертки с различными типами фильтров и множества слоев в их архитектуре позволяет формировать эффективные наборы признаков, размер которых уменьшается от начала к выходу СНС, для описания объектов, отличающихся нечеткой структурой и множеством вариаций в пределах класса. Однако на первом этапе при использовании СНС выполняется масштабирование входного изображения к размерам входного слоя, что обуславливает некоторое уменьшение информативности объектов, а при небольших их исходных размерах приводит к невозможности обнаружения на изображении. Поэтому актуальной задачей является разработка алгоритмов, позволяющих детектировать объекты малых размеров на изображениях с большим разрешением.

Существующие методы обнаружения объектов на изображениях с большим разрешением. В статье [3] предложен подход для локализации объектов на изображениях с большим разрешением, который основан на блочной обработке с перекрытием и применении СНС для выделения областей объектов в каждом блоке. Основным недостатком этого метода является достаточно высокая вероятность объединения нескольких различных объектов, расположенных на границах обрабатываемых блоков, в одну область и пропуска объекта за счет его фрагментации по блокам.

В работе [4] представлен алгоритм, на вход которого поступают исходное и масштабированное до размера [227×227] изображения. На уменьшенной копии с помощью СНС AlexNet детектируются объекты. Для обнаружения объектов с меньшими размерами исходное изображение делится на непересекающиеся блоки, размеры которых изменялись от 600 до 75 пикселей и обрабатывались с использованием СНС SCN (Spatial Correlation Network). Перекрытие при этом варьировалось от 25 до 50 % относительно размеров блоков. Для алгоритма [4] авторами получено значение полноты обнаружения (recall), составившее 80 % на базе данных

SUN2012 [5]. Известно, что метрика recall не позволяет оценить ложные срабатывания, а применяемая авторами база данных практически не содержит изображения с разрешением 4К.

В статье [6] рассмотрен подход для обнаружения людей, в котором входное изображение разделяется на пересекающиеся блоки большого размера. Если человек обнаружен, то выполняется разбиение на блоки меньшего размера для уточнения ограничивающих рамок обнаруженных областей интереса. Объединение найденных областей основано на вертикально вытянутой (прямоугольной) форме человека. Это означает, что для других классов объектов алгоритм не адаптирован. Оценка точности алгоритма выполнена с использованием базы данных PEVID-UHD [7] и закрытой базы, содержащей несколько большее количество объектов. Для первого набора данных значение средней точности равно 90,7 %, для второго – 75,4 %. При этом использование перекрытия блоков в диапазоне от 20 до 50 пикселей не является достаточным в ряде случаев, кроме того, оно не адаптивно к размеру входного изображения.

В работе [8] предложен алгоритм для обнаружения небольших объектов классов «транспортное средство» и «человек» на изображениях с разрешением 2К и размерами [1920×1080], полученных с малых беспилотных летательных аппаратов для реализации на графических процессорах мобильных устройств. При этом используется блочная обработка исходного изображения СНС PeleeNet со взаимным перекрытием 25 %. Полученные результаты тестирования на базе данных VisDrone2018 свидетельствуют о достижении значения метрики mAP (mean Average Precision) для данного алгоритма 36,67 %. Однако в работе не представлена оценка влияния величины перекрытия блоков на точность алгоритма. Величина пропуска объектов класса «человек» значительно выше по сравнению с классом «транспортное средство». Следовательно, объекты небольших размеров данный алгоритм обнаруживает значительно хуже даже на изображениях с разрешением ниже 4К.

Среди известных моделей СНС, отличающихся архитектурой, точностью и быстродействием, для решения поставленной задачи следует выделить модификации R-CNN (Region-based Convolutional Neural Networks) [9], ResNet (Residual Networks) [10] и YOLO (You Only Look Once) [11]. Первая версия модели R-CNN основана на предварительном выделении регионов на изображении, вычислении признаков с использованием СНС (например, AlexNet) и применении классификатора для идентификации объектов. Данная модель получила развитие в версиях Fast R-CNN [12] и Faster R-CNN [13]. Архитектура Fast R-CNN направлена на уменьшение использования числа регионов на изображении за счет применения слоя ROI, который действует по принципу слоев субдискретизации, позволяя получать области уменьшенного размера [14]. В Faster R-CNN предполагается выполнение двух этапов. На первом из них определяются области на изображении, в которых предположительно могут быть расположены объекты, с применением глубокой полносвязной сети (Region Proposal Network). На втором этапе используется детектор Fast R-CNN, который выполняет поиск объектов в предложенных регионах. Модель Faster R-CNN неустойчива к зашумленным изображениям и требует значительных вычислительных затрат.

Архитектура СНС ResNet использует пропускающие соединения, которые позволяют минимизировать ухудшение качества работы при увеличении количества слоев СНС, если на некотором слое сети достигнут предел точности. Коэффициент ошибок составляет 3,57 % в метрике top 5. В работе [15] представлена модель Inception-ResNet как развитие модели Inception путем ввода замыкающих соединений (shortcut connection), которые при необходимости позволяют пропускать слой и, соответственно, обнулять его влияние на результат работы детектора. Это дает возможность изменять архитектуру сети так, чтобы конечное количество слоев определялось для конкретной задачи в процессе обучения, что позволяет уменьшить коэффициент ошибки до 3,1 % в метрике top 5. Вместе с тем представленные модели характеризуются высокими временными затратами. Поэтому для решения многих задач используются усеченные варианты архитектур с уменьшенным количеством слоев (например, ResNet-34, ResNet-50 и др.), которые, однако, не позволяют достичь точности базовой модели.

Семейство СНС YOLO относится к однопроходным и дает возможность осуществлять локализацию и идентификацию объектов одной моделью. Среди существующих версий данных СНС по результатам тестирования в метрике top 5 наибольшей точностью (93,8 %) отличается YOLO третьей версии (URL: <https://arxiv.org/pdf/1804.02767.pdf>), которая для выделения при-

знаков применяет улучшенную архитектуру Darknet-53, содержащую 53 слоя и использующую 23 замыкающих соединения. Так, в качестве СНС для детектирования объектов в алгоритме их обнаружения на изображениях с разрешением 4К целесообразно применять СНС YOLO v3, поскольку данная версия характеризуется хорошей точностью обнаружения и удовлетворительным временем обработки.

Алгоритм обнаружения объектов на основе многомасштабного представления и блочной обработки сверточной нейронной сетью. Предлагаемый алгоритм требует пирамидального представления исходного изображения в виде набора его копий уменьшающегося масштаба. По мере движения вверх по пирамиде масштаб (размеры и разрешение) оригинала уменьшается. На каждом уровне пирамиды изображения разделяются на блоки, в которых выполняется детектирование с применением СНС. Для обеспечения высокой результативности обнаружения объектов малого размера необходимо перед обработкой минимизировать их разбиение на части, т. е. важно стремиться, чтобы ни на одном из блоков, подаваемых на вход СНС, объект не был фрагментирован. Поэтому используется взаимное перекрытие блоков, и чем выше величина перекрытия, тем больших размеров объект может быть обнаружен без вероятности его разделения на части, что, однако, потребует дополнительных вычислительных затрат. После обработки с применением СНС необходима процедура объединения найденных областей на всех уровнях.

Таким образом, алгоритм обнаружения объектов на изображениях с большим разрешением состоит из следующих шагов:

1. Определение количества уровней пирамиды изображений с учетом того, что размеры верхнего уровня должны быть приближены к размерам входного слоя используемой СНС по формуле

$$P = \lceil \log_2(\max(W, H) / l) \rceil + 1,$$

где W, H – ширина и высота входного изображения; $l \times l$ – размер входного слоя используемой СНС; $\lceil * \rceil$ – ближайшее целое.

Данный подход исключает фрагментацию объектов на минимальном масштабе и позволяет увеличить их правильную классификацию в целом.

2. Разбиение изображения на слое p на блоки с перекрытием. При этом их количество определяется равенством

$$B_p = \left(\frac{W / p - l}{k} + 1 \right) \cdot \left(\frac{H / p - l}{k} + 1 \right) = [B_{W_p}] \cdot [B_{H_p}],$$

где p – номер уровня пирамиды, $p = 1, \dots, P$; k – величина сдвига для блока (в пикселах). При округлении B_{W_p} либо B_{H_p} в сторону уменьшения последний неполный блок объединяется с предыдущим и масштабируется к размеру входного слоя СНС, а при округлении в сторону увеличения дополняется нулями.

Величина перекрытия блоков может быть рассчитана по формуле

$$\alpha = \frac{k}{l} \cdot 100 \% .$$

3. Обнаружение областей-кандидатов, которые могут содержать объект или фрагмент объекта. Для этого каждый блок обрабатывается с использованием СНС, а обнаруженная область-кандидат описывается набором признаков:

$$F = (x_1, y_1, x_2, y_2, E, Cl),$$

где x_1, y_1 – координаты верхнего левого угла найденной области на исходном изображении; x_2, y_2 – координаты нижнего правого угла найденной области на исходном изображении; Cl – класс выделенного объекта; E – уверенность СНС в правильной классификации.

4. Объединение областей-кандидатов.

4.1. Для объединения областей-кандидатов O_i и O_j проводится анализ величины их перекрытия на всех уровнях и блоках, а также принадлежности к одному классу:

$$IoU(O_i, O_j) > 0,5, Cl_i = Cl_j,$$

где

$$IoU(O_i, O_j) = \frac{O_i \cap O_j}{O_i \cup O_j}, \quad i, j \in 1, N | i \neq j.$$

Дескриптор объединенной области $F(O')$ формируется следующим образом:

$$F(O') = \left(\min(x_1^{O_i}, x_1^{O_j}), \min(y_1^{O_i}, y_1^{O_j}), \max(x_2^{O_i}, x_2^{O_j}), \max(y_2^{O_i}, y_2^{O_j}), \max(E^{O_i}, E^{O_j}), Cl(\max(E^{O_i}, E^{O_j})) \right).$$

4.2. При блочной обработке результатов детектирования возможно, что один и тот же объект или его фрагменты будут обнаружены со смещением координат на разных масштабах или в соседних ячейках. Кроме того, признаки фрагмента объекта меньшего размера могут отличаться от признаков исходного объекта и даже быть ближе к дескрипторам другого класса. Поэтому при объединении дополнительно применяется постобработка. Для этого предлагается использовать правило, по которому анализируются и объединяются все области, полученные на предыдущем шаге:

$$IoU(O'_i, O'_j) > 0,2, Cl_i = Cl_j, d_x > t_x, d_y > t_y,$$

где d_x и d_y – значения разности между размерами областей-кандидатов по осям x и y соответственно; t_x и t_y – пороговые уровни.

При уменьшении порогового значения на шаге 4.1 увеличится вероятность ложных объединений объектов одного класса, которые расположены на незначительном расстоянии друг от друга.

Результаты исследования предложенного алгоритма. Для тестирования результативности обнаружения объектов применялась выборка из 780 изображений с разрешением 4К, полученных при различных погодных условиях, а также с разными высотой установки и вертикальным углом наклона видеокамеры. Подготовка аннотированной базы данных изображений выполнена с использованием инструмента labelImg (URL: <https://github.com/tzutalin/labelImg>). Всего на изображениях аннотировано 6049 объектов двух классов («человек» и «транспортное средство») с размерами от [24×14] до [308×763]. При этом класс «транспортное средство» является составным, так как в него включены классы «автобус», «машина» и «грузовик».

На рис. 1, а всего размечено 19 объектов, из них 4 принадлежат классу «человек», а 15 относятся к классу «транспортное средство». На рис. 1, б показан пример с тремя объектами, которые расположены в затемненной зоне на близком расстоянии друг от друга.

Для проведения экспериментов предложенный алгоритм был реализован на языке программирования Python с применением фреймворка машинного обучения Pytorch, с помощью которого синтезировались основные элементы архитектуры СНС. Базовые операции над изображениями реализованы средствами библиотеки компьютерного зрения OpenCV. Для повышения скорости обнаружения применена пакетная обработка блоков изображений для каждого уровня пирамиды с использованием технологии CUDA, которая позволяет выполнять параллельную обработку данных на базе ресурсов графического процессора.



Рис. 1. Примеры аннотирования изображений для тестовой базы: а) для двух классов объектов; б) для одного класса объектов

Для оценки качества работы алгоритма вычислялась метрика mAP, предполагающая усреднение значений известной метрики AP (Average Precision) по классам. При этом на первом этапе рассчитывалось количество верных T_p и ложных F_p обнаружений на основе применения метрики объединения над пересечением для найденных объектов и аннотированных на изображениях базы данных с пороговым уровнем $T = 50\%$ для принятия решения. Затем вычислялись точность p и полнота r :

$$p = \frac{T_p}{T_p + F_p}, \quad r = \frac{T_p}{T_p + F_n},$$

где F_n – количество пропусков.

На основе полученных значений рассчитывалась метрика AP для каждого класса для 11 пороговых уровней в диапазоне от 0 до 1 с шагом (0,1) по формуле [16]

$$AP = \frac{1}{11} \cdot \sum_{i=0}^{10} p_{\text{int}}(r_i),$$

где $p_{\text{int}}(r_i) = \max p(\tilde{r}), \tilde{r} \geq r_i$.

На последнем этапе при вычислении метрики mAP полученные на предыдущем шаге значения AP усреднялись для двух классов объектов.

Для того чтобы более оптимально использовать возможности СНС при детектировании и идентификации небольших размеров объектов, необходимо определить пороговую степень уверенности СНС и размер входного слоя, обеспечивающие наибольшую результативность работы. Результаты экспериментов подтвердили, что для обеспечения высокой точности при обнаружении объектов небольших размеров на изображениях с большим разрешением необходимо минимизировать их масштабирование. Так, при размере входного слоя $[608 \times 608]$ значение метрики mAP составляет 21 %, а при размере $[1024 \times 1024]$ (использовался компьютер с центральным процессором Intel Core i5-8600 3,6 ГГц, ОЗУ 16 ГБ, NVIDIA GTX 1060) наибольшее значение для mAP 47 % достигается на тестовом наборе изображений с разрешением 4К, при этом пороговая степень уверенности $T = 50\%$. С учетом этого и в соответствии с шагом 1 предложенного алгоритма изображение с разрешением 4К для обработки необходимо представить в виде пирамиды из трех уровней.

На следующем этапе экспериментов выполнена оценка точности работы алгоритма при различных величинах перекрытия блоков и пороговых уровнях СНС с шагом 5 %. Следует отметить, что для минимального масштаба представления исходного изображения пороговое значение соответствует наиболее результативному для используемой СНС, в данном случае $T = 50\%$. Результаты экспериментов представлены в таблице. Их анализ показывает, что при любых значениях α и T обеспечивается превышение метрики mAP рассмотренного алгоритма по сравнению с mAP используемой СНС. Наибольший выигрыш обеспечивается при $\alpha = 60\%$ и $T = 75\%$, при этом mAP = 75,9 %.

Результаты исследования точности предложенного алгоритма, mAP, %

$T \backslash \alpha$	15	20	25	30	35	40	45	50	55	60	65
65	71,2	71,3	70,3	65,4	73,6	72,3	71,7	71,1	72,8	73,9	72,8
70	71,6	71,8	71,2	66,4	72,9	74,4	72,9	72,3	73,0	74,2	74,9
75	70,2	68,6	69,9	65,4	73,5	73,1	71,1	71,4	74,7	75,9	73,3
80	67,8	68,2	67,4	64,5	73,2	71,1	70,9	70,5	74,0	73,3	72,7
85	64,1	66,8	65,7	64,5	69,6	68,6	68,7	69,3	70,7	70,0	72,2
90	61,3	63,9	63,8	60,5	64,6	64,3	65,5	64,6	64,0	69,9	68,4
95	54,5	56,7	57,2	54,0	56,8	56,7	58,2	58,1	59,5	59,7	61,2

На рис. 2 изображены фрагменты обработанных изображений с разрешением 4K из подготовленной базы данных. С помощью предложенного алгоритма детектировано девять объектов (рис. 2, а), минимальный размер одного из них [33×13], а применение СНС YOLO v3 позволило обнаружить объект размером [149×54] (рис. 2, б).



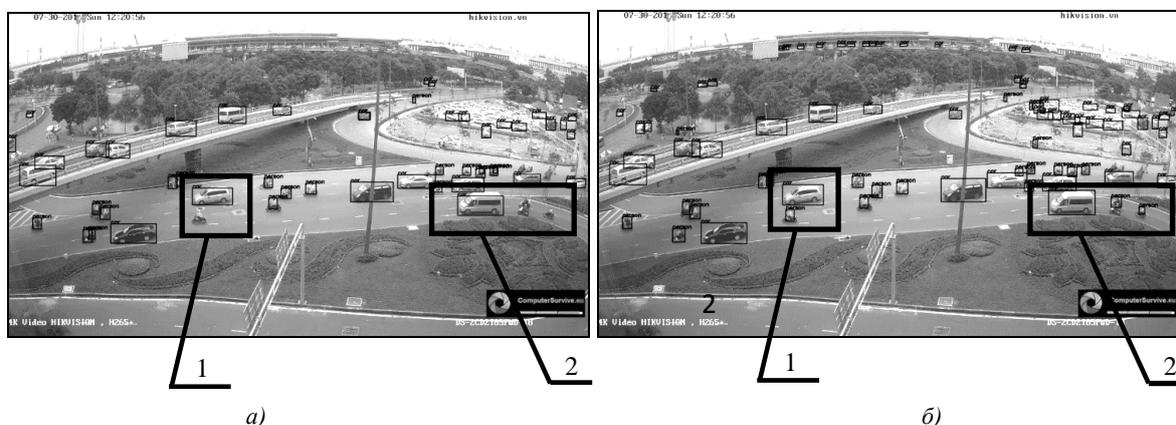
а)



б)

Рис. 2. Примеры обнаружения объектов на изображениях с разрешением 4K: а) с помощью предложенного алгоритма; б) с помощью алгоритма СНС YOLO v3

Тестирование результативности работы предложенного алгоритма также выполнено и на других 4K-изображениях, которые не были включены в аннотированную базу данных. На рис. 3, а показан пример обнаружения объектов на видеокadre (URL: https://www.youtube.com/watch?v=nG_wGUi-Ozc) с разрешением 4K, применением СНС YOLO v3, размером входного слоя [1024×1024] и $T = 50\%$. При этом обнаружено 50 объектов, минимальный размер детектированного объекта составляет [41×20] пикселей.



а)

б)

Рис. 3. Примеры обнаружения объектов на видеокadre с разрешением 4K: а) с помощью алгоритма СНС YOLO v3; б) с помощью предложенного алгоритма

На рис. 3, б представлен результат детектирования объектов предложенным алгоритмом с определенными в результате экспериментов параметрами: размером входного слоя СНС [1024×1024], количеством уровней пирамиды $P=3$, $\alpha=60\%$ и $T=75\%$. При этом обнаружено 80 объектов, их минимальный размер составляет [33×11] пикселей. На рис. 3 видно, что наибольший выигрыш в детектировании достигается именно для объектов небольших размеров, которые на данном изображении размещены в верхней и правой частях изображения.

Анализ результатов детектирования на различных масштабах изображения (рис. 4, а–в) и объединения областей (рис. 4, г) для увеличенного фрагмента 2 из рис. 3, б с тремя объектами подтверждает, что применение пирамидальной обработки позволяет обнаружить объекты различного размера на разных уровнях и повышает точность алгоритма в целом. СНС YOLO v3 не детектировала человека из выделенного фрагмента (см. рис. 3, а).



Рис. 4. Примеры обнаружения объектов на разных уровнях пирамиды для изображения размером [3840×216]: а) на исходном изображении; б) на втором уровне [1920×1080]; в) на третьем уровне [960×540]; г) итоговый результат

На рис. 5 показан пример объединения областей-кандидатов на четвертом шаге алгоритма для фрагмента 1 из рис. 3, б, который свидетельствует об эффективности этого этапа. После третьего шага алгоритма обнаружено шесть областей-кандидатов (рис. 5, а), применение процедуры объединения позволило правильно локализовать человека и улучшить результат для автомобиля (рис. 5, б), а дальнейшая постобработка дала возможность отфильтровать ошибочную локализацию его передней части. Модель СНС YOLO v3 не обнаружила человека в данной области видеокadra (см. рис. 3, а).

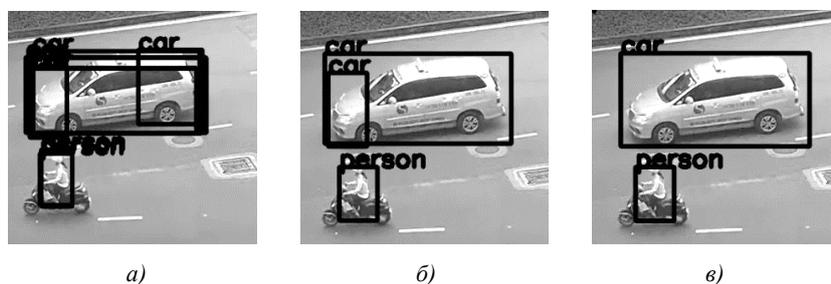


Рис. 5. Примеры объединения областей-кандидатов: а) результат обнаружения областей-кандидатов на всех уровнях пирамиды; б) результат применения объединения (см. шаг 4.1 алгоритма); в) результат постобработки (см. шаг 4.2 алгоритма)

Заключение. В статье предложен алгоритм для обнаружения объектов на изображениях с большим разрешением (4К и выше), включающий: формирование пирамиды изображений, пока размеры верхнего уровня не будут приближены к размерам входного слоя используемой СНС; блочное разбиение с перекрытием для всех полученных слоев; обнаружение объектов с применением СНС в каждом блоке; постобработку для полученных результатов на предыдущем шаге. Для детектирования объектов была использована СНС YOLO v3 с размерами входного слоя [1024×1024], который обеспечивает высокую результативность обнаружения небольших объектов на изображениях с разрешением 4К. Программная реализация разработанного алгоритма выполнена на языке Python с применением библиотеки компьютерного зрения OpenCV, фреймворка машинного обучения Pytorch и технологии CUDA. Для проведения экспериментов по оценке эффективности предложенного алгоритма подготовлена база данных изображений с разрешением 4К с размеченными объектами небольших и средних размеров классов «человек» и «транспортное средство».

Проведенные эксперименты показали, что для используемых данных значение метрики mAP может достигать 75,9%. Соответственно, рассмотренный подход является перспективным при

решении прикладных задач обнаружения объектов малых размеров на изображениях с большим разрешением.

Список литературы

1. Дворкович, А. В. Метрологическое обеспечение видеoinформационных систем / А. В. Дворкович, В. П. Дворкович. – М. : Техносфера, 2015. – 784 с.
2. Goulekas, K. *Visual Effects in a Digital World: a Comprehensive Glossary of over 7,000 Visual Effects Terms* / K. Goulekas. – San Francisco : Morgan Kaufmann, 2001. – 600 p.
3. An effective object detection algorithm for high resolution video by using convolutional neural network / D. Vorobjov [et al.] // *Advances in Neural Networks-ISNN2018. Lecture Notes in Computer Science*. – 2018. – Vol. 10878. – P. 503–510. https://doi.org/10.1007/978-3-319-92537-0_58
4. Yongxi, L. Efficient object detection for high resolution images / L. Yongxi, T. Javidi // *Proc. of 53rd Annual Allerton Conf. on Communication, Control, and Computing (Allerton), Monticello, IL, USA, 30 Sept. – 2 Oct. 2015. – Monticello, 2015. – P. 1091–1098. <https://doi.org/10.1109/ALLERTON.2015.7447130>*
5. Sun database: large-scale scene recognition from abbey to zoo / J. Xiao [et al.] // *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), San Francisco, CA, USA, 13–18 June 2010. – San Francisco, 2010. – P. 3485–3492. <https://doi.org/10.1109/CVPR.2010.5539970>*
6. Ruzicka, V. Fast and accurate object detection in high resolution 4K and 8K video using GPUs / V. Ruzicka, F. Franchetti // *Proc. of 2018 IEEE High Performance Extreme Computing Conf. (HPEC), Waltham, MA, USA, 25–27 Sept. 2018. – Waltham, 2018. – P. 1–7. <https://doi.org/10.1109/HPEC.2018.8547574>*
7. Korshunov, P. UHD video dataset for evaluation of privacy / P. Korshunov, T. Ebrahimi // *Proc. of Sixth Intern. Workshop on Quality of Multimedia Experience (QoMEX), Singapore, 18–20 Sept. 2014. – Singapore, 2014. – P. 232–237. <https://doi.org/10.1109/QoMEX.2014.6982324>*
8. Unel, F. O. The power of tiling for small object detection / F. O. Unel, B. Ozkalayci, C. Çigla // *CVPR Workshops [Electronic resource]. – 2019. – Mode of access: http://openaccess.thecvf.com/content_CVPRW_2019/papers/UAVision/Unel_The_Power_of_Tiling_for_Small_Object_Detection_CVPRW_2019_paper.pdf. – Date of access: 18.01.2020.*
9. Region-based convolutional networks for accurate object detection and segmentation / R. Girshick [et al.] // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. – 2016. – Vol. 38. – P. 142–158. <https://doi.org/10.1109/TPAMI.2015.2437384>
10. Deep residual learning for image recognition / K. He [et al.] // *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2016), Las Vegas, NV, USA, 27–30 June 2016. – Las Vegas, 2016. – P. 770–778. <https://doi.org/10.1109/CVPR.2016.90>*
11. You only look once: unified, real-time object detection / J. Redmon [et al.] // *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2016), Las Vegas, NV, USA, 27–30 June 2016. – Las Vegas, 2016. – P. 779–788. <https://doi.org/10.1109/CVPR.2016.91>*
12. Girshick, R. Fast R-CNN / R. Girshick // *Proc. of IEEE Intern. Conf. on Computer Vision (ICCV), Santiago, Chile, 11–18 Dec. 2015. – Santiago, 2015. – P. 1440–1448. <https://doi.org/10.1109/ICCV.2015.169>*
13. Faster R-CNN: towards real-time object detection with region proposal networks / S. Ren [et al.] // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. – 2015. – Vol. 39, no. 6. – P. 1137–1149. <https://doi.org/10.1109/TPAMI.2016.2577031>
14. Глубокое обучение для детектирования объектов на изображениях документов / А. А. Крощенко и др. // *Вестник БрГТУ. Физика, математика, информатика*. – 2017. – № 5(107). – С. 2–9.
15. Inception-v4, inception-ResNet and the impact of residual connections on learning / C. Szegedy [et al.] // *Proc. of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17), San Francisco, California, USA, 4–9 Febr. 2017. – San Francisco, 2017. – P. 4278–4284.*
16. The pascal Visual Object Classes (VOC) challenge / M. Everingham [et al.] // *Intern. J. of Computer Vision*. – 2010. – Vol. 88. – P. 303–338. <https://doi.org/10.1007/s11263-009-0275-4>

References

1. Dvorkovich A. V., Dvorkovich V. P. *Metrologicheskoe obespechenie videoinformatsionnykh system. Metrological Support of Video Information Systems*, Moscow, Technosfera, 2015, 784 p. (in Russian).
2. Goulekas K. *Visual Effects in a Digital World: a Comprehensive Glossary of over 7,000 Visual Effects Terms*. San Francisco, Morgan Kaufmann, 2001, 600 p.
3. Vorobjov D., Zakharova I., Bohush R., Ablameyko S. An effective object detection algorithm for high resolution video by using convolutional neural network. *Advances in Neural Networks-ISNN2018. Lecture Notes in Computer Science*, 2018, vol. 10878, pp. 503–510. https://doi.org/10.1007/978-3-319-92537-0_58

4. Yongxi L., Javidi T. Efficient object detection for high resolution images. *Proceedings of 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton), Monticello, IL, USA, 30 September – 2 October 2015*. Monticello, 2015, pp. 1091–1098. <https://doi.org/10.1109/ALLERTON.2015.7447130>
5. Xiao J., Hays J., Ehinger K., Oliva A., Torralba A. Sun database: large-scale scene recognition from abbey to zoo. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, CA, USA, 13–18 June 2010*. San Francisco, 2010, pp. 3485–3492. <https://doi.org/10.1109/CVPR.2010.5539970>
6. Ruzicka V., Franchetti F. Fast and accurate object detection in high resolution 4K and 8K video using GPUs. *Proceedings of 2018 IEEE High Performance Extreme Computing Conference (HPEC), Waltham, MA, USA, 25–27 September 2018*. Waltham, 2018, pp. 1–7. <https://doi.org/10.1109/HPEC.2018.8547574>
7. Korshunov P., Ebrahimi T. UHD video dataset for evaluation of privacy. *Proceedings of Sixth International Workshop on Quality of Multimedia Experience (QoMEX), Singapore, 18–20 September 2014*. Singapore, 2014, pp. 232–237. <https://doi.org/10.1109/QoMEX.2014.6982324>
8. Unel F. O., Ozkalayci B., Çigla C. The power of tiling for small object detection. *CVPR Workshops, 2019*. Available at: http://openaccess.thecvf.com/content_CVPRW_2019/papers/UAVision/Unel_The_Power_of_Tiling_for_Small_Object_Detection_CVPRW_2019_paper.pdf. (accessed 18.01.2020).
9. Girshick R., Donahue J., Darrell T., Malik J. Region-based convolutional networks for accurate object detection and segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, vol. 38, pp. 142–158. <https://doi.org/10.1109/TPAMI.2015.2437384>
10. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016), Las Vegas, NV, USA, 27–30 June 2016*. Las Vegas, 2016, pp. 770–778. <https://doi.org/10.1109/CVPR.2016.90>
11. Redmon J., Divvala S. K., Girshick R. B., Farhadi A. You only look once: unified, real-time object detection. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016), Las Vegas, NV, USA, 27–30 June 2016*. Las Vegas, 2016, pp. 779–788. <https://doi.org/10.1109/CVPR.2016.91>
12. Girshick R. Fast R-CNN. *Proceedings of IEEE Intern. Conf. on Computer Vision (ICCV), Santiago, Chile, 11–18 December 2015*. Santiago, 2015, pp. 1440–1448. <https://doi.org/10.1109/ICCV.2015.169>
13. Ren S., He K., Girshick R., Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, vol. 39, no. 6, pp. 1137–1149. <https://doi.org/10.1109/TPAMI.2016.2577031>
14. Kroshchenko A., Golovko V., Bezobrazov S., Mikhno E., Khatskevich M, ..., Brich A. Glubokoe obuchenie dlia detektirovaniia obiektov na izobrazheniakh dokumentov [Deep training for detecting of objects at images of documents]. *Vestnik Brestskogo gosudarstvennogo tekhnicheskogo universiteta. Fizika, matematika, informatika [Bulletin of the Brest State Technical University. Physics, mathematics, Computer Science]*, 2017, vol. 5 (107), pp. 2–9 (in Russian).
15. Szegedy C., Ioffe S., Vanhoucke V. Inception-v4, inception-ResNet and the impact of residual connections on learning. *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17), San Francisco, California, USA, 4–9 February 2017*. San Francisco, 2017, pp. 4278–4284.
16. Everingham M., Van Gool L., Williams C., Winn J., Zisserman A. The pascal Visual Object Classes (VOC) challenge. *International Journal of Computer Vision*, 2010, vol. 88, pp. 303–338. <https://doi.org/10.1007/s11263-009-0275-4>

Информация об авторах

Богуш Рихард Петрович, кандидат технических наук, доцент, заведующий кафедрой вычислительных систем и сетей, факультет информационных технологий, Полоцкий государственный университет, Полоцк, Беларусь.
E-mail: bogushr@mail.ru

Захарова Ирина Юрьевна, магистр технических наук, аспирант кафедры вычислительных систем и сетей, факультет информационных технологий, Полоцкий государственный университет, Полоцк, Беларусь.
E-mail: ira9992011@yandex.ru

Абламейко Сергей Владимирович, академик Национальной академии наук Беларуси, доктор технических наук, профессор, профессор механико-математического факультета, Белорусский государственный университет, Минск, Беларусь.
E-mail: ablameyko@bsu.by

Information about the authors

Rykhard P. Bohush, Cand. Sci. (Eng.), Associate Professor, Head of the Department of Computer Systems and Networks, Polotsk State University, Novopolotsk, Belarus.
E-mail: bogushr@mail.ru

Iryna Yu. Zakharava, M. Sci. (Eng.), Postgraduate Student at the Department of Computer Systems and Networks, Polotsk State University, Novopolotsk, Belarus.
E-mail: ira9992011@yandex.ru

Sergey V. Ablameyko, Academician of the National Academy of Sciences of Belarus, Dr. Sci. (Eng.), Professor, Professor of the Faculty of Mechanics and Mathematics, Belarusian State University, Minsk, Belarus.
E-mail: ablameyko@bsu.by