

Программные системы и вычислительные методы

Правильная ссылка на статью:

Захаров А.А. Метод обнаружения объектов на изображениях на основе нейронных сетей на графах и небольшого количества обучающих примеров // Программные системы и вычислительные методы. 2024. № 4. DOI: 10.7256/2454-0714.2024.4.72558 EDN: UTTFCH URL: https://nbpublish.com/library_read_article.php?id=72558

Метод обнаружения объектов на изображениях на основе нейронных сетей на графах и небольшого количества обучающих примеров

Захаров Алексей Александрович

кандидат технических наук

доцент; кафедра программной инженерии; Муромский институт (филиал) федерального государственного бюджетного образовательного учреждения высшего образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых»

ведущий научный сотрудник; Муромский институт (филиал) федерального государственного бюджетного образовательного учреждения высшего образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых»

602264, Россия, Владимирская область, г. Муром, ул. Орловская, 23, ауд. 402

✉ aa-zaharov@ya.ru



[Статья из рубрики "Компьютерная графика, обработка изображений и распознавание образов"](#)

DOI:

10.7256/2454-0714.2024.4.72558

EDN:

UTTFCH

Дата направления статьи в редакцию:

03-12-2024

Дата публикации:

11-12-2024

Аннотация: В представленной работе объектом исследования являются системы компьютерного зрения. Предмет исследования – метод обнаружения объектов на изображениях на основе нейронных сетей на графах и небольшого количества обучающих примеров. Подробно рассматриваются такие аспекты темы как использование структурного представления сцены для повышения точности обнаружения объектов. Предлагается совместное использование информации о структуре сцены на основе нейронных сетей на графах и обучения с «нескольких

выстрелов» для повышения точности обнаружения объектов. Устанавливаются отношения между классами с помощью внешних семантических связей. Для этого предварительно создаётся граф знаний. Метод содержит два этапа. На первом этапе выполняется обнаружение объектов на основе обучения с «нескольких выстрелов». На втором этапе выполняется повышение точности обнаружения с использованием нейронной сети на графах. Основой разрабатываемого метода является использование свёртки на основе спектральной теории графов. Каждая вершина представляет собой категорию в графе знаний, а вес ребра графа рассчитывается на основе условной вероятности. На основе свёртки объединяется информация из соседних вершин и рёбер для обновления значений вершин. Научная новизна разработанного метода заключается в совместном использовании свёрточных сетей на графах и обучения с «нескольких выстрелов» для увеличения точности обнаружения объектов. Особым вкладом автора в исследование темы является применение свёрточной сети на основе графа знаний для улучшения результатов работы метода обнаружения объектов при использовании малого количества обучающих примеров. Метод исследовался на тестовых наборах изображений из области компьютерного зрения. Используя наборы данных PASCAL VOC и MS COCO продемонстрировано, что предлагаемый метод увеличивает точность обнаружения объектов за счет анализа структурных взаимосвязей. Средняя точность обнаружения объектов при использовании разработанного метода увеличивается на 1-5% по сравнению с методом обучения с «нескольких выстрелов» без использования структурного представления.

Ключевые слова:

компьютерное зрение, обнаружение объектов, свёрточные сети, небольшой набор данных, глубокое обучение, ограниченная аннотация, граф, распознавание образов, искусственный интеллект, структурное представление сцен

Исследование выполнено за счет гранта Российского научного фонда № 23-21-00486, <https://rscf.ru/project/23-21-00486/>

Введение

Обнаружение объектов – это важная задача компьютерного зрения, которая заключается в том, чтобы найти интересующий объект на входном изображении, а затем точно отнести его к определенному классу. Наиболее значимыми характеристиками результатов обнаружения объектов являются точность локализации и классификации, а также скорость обнаружения. Обнаружение объектов служит основой для многих других областей: автономная навигация, человеко-машинные интерфейсы, контроль технологических процессов, дистанционное зондирование Земли, медицинская диагностика, биометрия, видеонаблюдение и т.д.

В последние годы быстрое развитие методов глубокого обучения значительно способствовало прогрессу в области обнаружения объектов. В настоящее время разработано большое количество методов обнаружения объектов с использованием нейронных сетей. Методы обнаружения объектов на основе глубокого обучения принято делить на одноэтапные и двухэтапные [1]. К одноэтапным методам относятся YOLO [2], SSD [3], RetinaNet [4] и др. К двухэтапным методам принадлежат Fast R-CNN [5], Faster R-CNN [6], Mask R-CNN [7] и др.

Следует отметить, что методы обнаружения объектов на основе глубокого обучения сталкиваются со следующими критическими проблемами:

- многообразие сцен при реальных наблюдениях. Большое количество методов обнаружения объектов достаточно хорошо работает в лабораторных условиях: равномерное освещение, однородный статичный фон, отсутствие движения камеры и т.д. Однако в реальных условиях наблюдения точность обнаружения объектов значительно уменьшается. Это связано с наличием сложного текстурированного фона и низкого контраста изображений, присутствием посторонних подвижных объектов, наличием взаимных перекрытий, затененных областей и шумов.
- необходимость разметки больших наборов изображений ручным способом. Ключевым компонентом революции в области глубокого обучения была доступность больших аннотированных наборов данных. Несмотря на то, что большинство наборов данных компьютерного зрения маркируются с помощью краудсорсинга, этот процесс по-прежнему является дорогостоящим и требует много времени, что становится узким местом при развертывании систем глубокого обучения. Многие существующие методы обнаружения, основанные на глубоком обучении, показывают в лабораторных условиях хорошие результаты с использованием больших наборов данных. Однако эти методы труднореализуемы в реальных условиях из-за невозможности создания больших аннотированных наборов данных.
- необходимость обнаружения объектов по категориям, у которых количество экземпляров в обучающем наборе очень мало. Часто возникает проблема обнаружения объектов по категориям, у которых нет экземпляров в обучающем наборе или их количество ограничено. Если количество обучающих примеров слишком мало, по сравнению со всеми возможными вариациями, то возникает проблема обучения с использованием небольшой выборки.

Таким образом, одной из основных проблем при реализации методов обнаружения объектов на основе глубокого обучения является необходимость создания большого объема аннотированных данных, что не всегда возможно по экономическим и техническим причинам.

В последние годы активно разрабатываются методы обнаружения объектов на основе обучения с «нескольких выстрелов» (few-shot learning), которые пытаются решить задачу с использованием небольшого количества примеров [\[8\]](#). Часто бывает, что данные базового класса ограничены несколькими примерами для обучения. В этом случае модель предварительно обучается на крупномасштабном наборе данных из другой области. Основное же назначение нескольких примеров – это адаптация представления к предметной области [\[9\]](#). Таким образом, для обучения модели требуется значительно меньше маркированных данных.

Известны следующие методы обнаружения объектов, основанные на обучении с «малым количеством выстрелов»: Multi-Scale Positive Sample Refinement for Few-Shot Object Detection (MPSR) [\[10\]](#), Frustratingly Simple Few-Shot Object Detection (TFA) [\[11\]](#), Few-Shot Object Detection via Feature Reweighting (MetaYOLO) [\[12\]](#) и др. Однако точность методов обнаружения с использованием небольшого количества обучающих данных остается низкой.

Предлагается использовать структурное представление сцены для повышения точности обнаружения объектов. Для описания структуры будут использоваться графы.

Особенности графа позволяют оценить связи между элементами изображения. Графы используются в различных областях компьютерного зрения для сегментации изображений [13], обнаружения значимых областей [14], кластеризации [15] и др. Возможности графов позволяют анализировать структурные отношения между объектами сцен, что даёт возможность собирать больше информации по сравнению с локальным анализом данных.

В последние годы изучение графов быстро продвинулось вперед благодаря наличию больших наборов данных, мощных вычислительных ресурсов, а также достижениям в области машинного обучения и искусственного интеллекта [16]. Методы глубокого обучения могут эффективно кодировать и представлять данные о графах в виде векторов. Эти векторы затем можно использовать в различных задачах с высокой производительностью. Нейронная сеть на графах (Graph Neural Network, GNN) – это архитектура глубокого обучения, специально разработанная для данных, описанных при помощи графов [17-20]. В отличие от традиционных алгоритмов глубокого обучения, которые в первую очередь были разработаны для текста и изображений, GNN созданы специально для обработки и анализа структурированных наборов данных.

Целью исследования является разработка метода, повышающего точность обнаружения объектов. Предполагается, что при использовании структурного представления точность обнаружения объектов на основе небольшого количества обучающих примеров повысится.

Научная новизна разработанного метода заключается в совместном использовании нейронных сетей на графах и обучения с «нескольких выстрелов» для увеличения точности обнаружения объектов.

Разработка метода обнаружения объектов на изображениях на основе нейронных сетей на графах и небольшого количества обучающих примеров

Разработанный метод включает два этапа.

Этап 1. Обнаружение объектов с использованием обучения с «нескольких выстрелов» [12].

Этап 2. Повышение точности обнаружения с использованием нейронной сети на графах. Структурная схема метода представлена на рисунке 1.

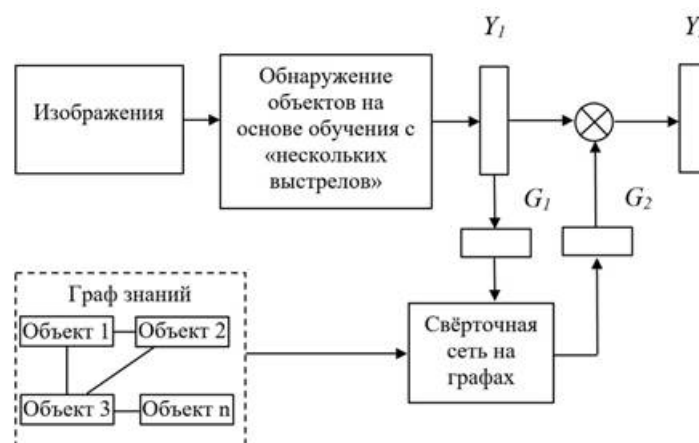


Рис. 1. Структурная схема метода обнаружения объектов на изображениях на основе нейронных сетей на графах и небольшого количества обучающих примеров

На этапе 1 происходит обнаружение объектов с использованием обучения с «нескольких выстрелов». Далее на основе обнаруженных ограничивающих прямоугольников объектов строится матрица вероятностей Y_1 :

$$Y_1 = B \times C,$$

где B – количество обнаруженных ограничивающих областей, C – количество классов, Y_1 – вероятность c -ого класса b -ой ограничивающей области.

На этапе 2 осуществляется повышение точности обнаружения с использованием нейронной сети на графах. Связи между объектами задаются с использованием графа знаний. Граф знаний представляет собой семантическую сеть, которая хранит информацию о различных классах объектов и связях между ними. Граф знаний, описывающий связи между классами объектов, показан на рис. 2.

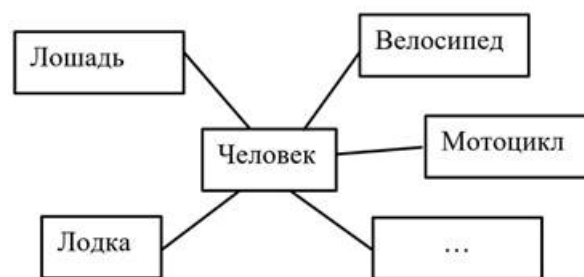


Рис.2. Граф знаний, описывающий связи между классами объектов

Предполагается, что совместное присутствие на рисунке таких объектов, как человек и велосипед, человек и мотоцикл, человек и лодка и т.д. поможет повысить точность обнаружения с использованием имеющихся связей (рис. 3).

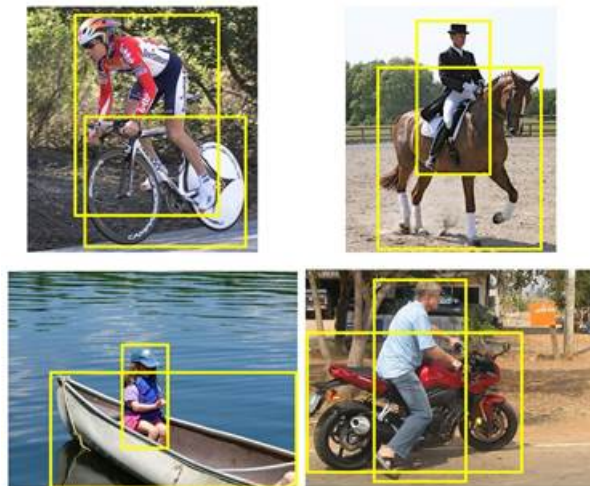


Рис. 3. Совместное присутствие на изображениях объектов различных классов

В графе знаний каждая вершина представляет некоторый класс, а ребро графа от вершины V_1 до вершины V_2 является условной вероятностью [21]

$$P(V_2 / V_1).$$

Например, если человек и велосипед появляются вместе в наборе данных 10 раз, а всего человек появляется в наборе данных 20 раз, то ребро от вершины класса «человек» до вершины класса «велосипед» будет иметь значение

$$P (\text{человек/ велосипед}) = 0,5.$$

Граф знаний описывается матрицей смежности $A = \mathbb{R}^{C \times C}$, где C — количество классов, представленных графом.

На вход свёрточной сети на графах подается следующий вектор:

$$G_{1,c} = \max_{b=1, 2, \dots, B} (Y_{1,bc}), \quad c=1, 2, C,$$

где $G_{1,c}$ — значение максимальной вероятности c -ого класса среди всех обнаруженных ограничивающих прямоугольников.

В задаче классификации узлов GNN использует информацию для создания векторного представления каждого узла в графе. Такое представление включает в себя не только исходные характеристики вершины, но также информацию о связях между вершинами. Вместо того, чтобы ограничиваться исходными атрибутами, GNN добавляет к свойствам исходных вершин атрибуты из соседних вершин и ребер, что делает представление гораздо более полным и значимым. Новые представления вершин затем используются для выполнения конкретных задач, таких как классификация вершин, регрессия или прогнозирование связей. В частности, GNN определяет операцию свёртки графа, которая объединяет информацию из соседних вершин и рёбер для обновления представлений. Эта операция выполняется итеративно, что позволяет модели изучать более сложные взаимосвязи между вершинами по мере увеличения количества итераций (рис. 4). В работе сеть содержит четыре свёрточных слоя.

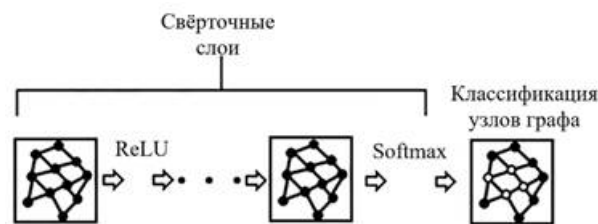


Рис. 4. Классификация узлов графа на основе свёрточной сети

Свёрточная сеть на графах описывается следующим правилом потока информации между слоями [17]:

$$H_{l+1} = \text{ReLU}((aD^{-1} + I)H_l W_l) + B_l,$$

где I — единичная матрица, H_l — активационная матрица слоя l , H_{l+1} — активационная матрица слоя $l+1$, D — степеньная матрица графа, W_l — матрица весов, B_l — переменная для управления средним значением выходного сигнала, a — настраиваемый параметр, который используется для определения влияния связанных узлов. По умолчанию $a=0.5$.

Результат работы нейронной сети на графах представлен в виде матрицы корректирующих весов:

$$G_2 = \text{GNN}(G_1).$$

Окончательный результат представляет собой поэлементное произведение исходной матрицы Y_1 и матрицы корректирующих весов G_2 :

$Y_2=Y_1G_2.$

Исследование метода

Для проведения исследования использовались изображения из набора данных PASCAL VOC. Для каждой категории отбирались *K* экземпляров: *K* = 1, 3, 5, 10.

Была рассчитана средняя точность (Mean average precision):

$mAP=(SUM_{k=1...n}(AP_k))/n,$

..

где *n* – количество классов, *AP_k*– средняя точность класса *k*

Результаты экспериментов были сведены в таблицу 1.

Таблица 1. Расчет показателя *mAP* для разного количества экземпляров

Метод обнаружения объектов	<i>mAP</i> (%)			
	1 экземпляр	3 экземпляра	5 экземпляров	10 экземпляров
Meta-RCNN (few-shot)	17.38	25.92	32.63	48.35
Meta-RCNN (few-shot) + нейронная сеть на графах	18.89	28.78	37.51	53.78

В процессе обучения нейронной сети на графах были рассчитаны потери. Нейронная сеть на графах обучалась в течение 500 эпох (рис. 5).

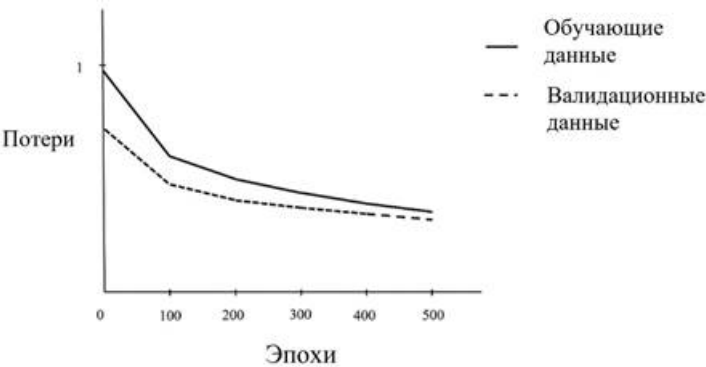


Рис. 5. График функции потерь при обучении нейронной сети на графах

Для оптимизации применялся алгоритм ADAM. ADAM позволяет настраивать скорость обучения для каждого параметра. Для обучения модели использовался графический процессор GeForce RTX 3060. Таким образом, разработанный метод позволяет повысить среднюю точность на 1-5% за счет анализа структурных связей между объектами.

Заключение

В статье предложен метод, который частично компенсирует недостатки метода обучения

с «малым количеством выстрелов» при обнаружении объектов. Метод основан на использовании нейронной сети на графах для описания структуры анализируемой сцены. Эксперименты показали эффективность разработанного метода. Средний показатель точности обнаружения объектов увеличился до пяти процентов. Также было показано, что предложенный метод позволяет достичь еще большей точности при увеличении объема обучающих данных.

Библиография

1. Zou Z., Chen K., Shi Z., Guo Y., Ye J. Object Detection in 20 Years: A Survey // Proceedings of the IEEE. 2023. Vol. 111 (3). Pp. 257-276.
2. Redmon J., Divvala S., Girshick R., Farhadi A. You only look once: Unified, real-time object detection // IEEE Conference on Computer Vision and Pattern Recognition. 2016. Pp. 779-788.
3. Liu W., Anguelov D., Erhan D., Szegedy C., Reed S., Fu C.Y., Berg A. C. Ssd: Single shot multibox detector // European Conference on Computer Vision. 2016. Pp. 21-37.
4. Lin T.Y., Goyal P., Girshick R., He K., Dollar P. Focal loss for dense object detection // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2018. Vol. 42(2). Pp. 318-327.
5. Girshick P. Fast R-CNN // 2015 IEEE International Conference on Computer Vision (ICCV). 2015. Pp. 1440-1448.
6. Ren S., He K., Girshick R., Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks // Advances in Neural Information Processing System. 2015. Pp. 91-99.
7. He K., Gkioxari G., Dollar P., Girshick R. Mask R-CNN // Proceedings of the IEEE International Conference on Computer Vision. 2017. Pp. 2961- 2969.
8. Köhler M., Eisenbach M., Gross H. M. Few-Shot Object Detection: A Survey // IEEE Transactions on Neural Networks and Learning Systems. 2024. Vol. 35 (9). Pp. 11958 – 11978.
9. Huang G., Laradji I., Vazquez D., Lacoste-Julien S., Rodriguez P. A Survey of Self-Supervised and Few-Shot Object Detection // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2023. Vol. 45(4). Pp. 4071-4089.
10. Wu J., Liu S., Huang D., Wang Y. Multi-scale positive sample refinement for few-shot object detection // European Conference on Computer Vision. 2020. Pp. 456-472.
11. Wang X., Huang T. E., Gonzalez J., Darrell T., Yu F. Frustratingly simple few-shot object detection // Proceedings of the 37th International Conference on Machine Learning (ICML). 2020. Pp. 9919-9928.
12. Kang B., Liu Z., Wang X., Yu F., Feng J., Darrell T. Few-shot object detection via feature reweighting // 2019 IEEE/CVF International Conference on Computer Vision. 2019.
13. Захаров А.А., Тужилкин А.Ю. Сегментация спутниковых изображений на основе суперпикселей и разрезов на графах // Программные системы и вычислительные методы. 2018. № 1. С.7-17. DOI: 10.7256/2454-0714.2018.1.25629 URL: https://e-notabene.ru/itmag/article_25629.html
14. Захаров. А.А., Титов Д.В., Жизняков А.Л., Титов В.С. Метод визуального внимания на основе ранжирования вершин графа по разнородным признакам изображений // Компьютерная оптика. 2020. Т. 44, № 3. С. 427-435.
15. Barinov A.E., Zakharov A.A. Clustering using a random walk on graph for head pose estimation // International Conference on Mechanical Engineering, Automation and Control Systems, MEACS. 2015.
16. Cao P., Zhu Z., Wang Z., Zhu Y., Niu Q. Applications of graph convolutional networks in computer vision // Neural Computing and Applications. 2022. № 34. Pp. 13387-13405.

17. Kipf T.N. Deep Learning with Graph-Structured Representations, Universiteit van Amsterdam, 2020.
18. Li W., Liu X., Yuan Y. SIGMA++: Improved Semantic-Complete Graph Matching for Domain Adaptive Object Detection // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2023. Vol.45 (7). Pp. 9022-9040.
19. Chen C., Li J., Zhou H.Y., Han X., Huang Y., Ding X., Yu Y. Relation matters: Foreground-aware graph-based relational reasoning for domain adaptive object detection // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2023. Vol. 45 (3). Pp. 3677-3694.
20. Chen T., Lin L., Chen R., Hui X., Wu X. Knowledge-Guided Multi-Label Few-Shot Learning for General Image Recognition // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2022. Vol.44 (3). Pp.1371-1384.
21. Liu Z., Jiang Z., Feng W., Feng H. OD-GCN: Object Detection Boosted by Knowledge GCN // 2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW). 2020

Результаты процедуры рецензирования статьи

В связи с политикой двойного слепого рецензирования личность рецензента не раскрывается.

Со списком рецензентов издательства можно ознакомиться [здесь](#).

Представленная статья на тему «Метод обнаружения объектов на изображениях на основе нейронных сетей на графах и небольшого количества обучающих примеров» соответствует тематике журнала «Программные системы и вычислительные методы» и посвящена вопросу обнаружения объектов, что является важной задачей компьютерного зрения, которая заключается в том, чтобы найти интересующий объект на входном изображении, а затем точно отнести его к определенному классу. Наиболее значимыми характеристиками результатов обнаружения объектов являются точность локализации и классификации, а также скорость обнаружения. Обнаружение объектов служит основой для многих других областей: автономная навигация, человеко-машинные интерфейсы, контроль технологических процессов, дистанционное зондирование Земли, медицинская диагностика, биометрия, видеонаблюдение и т.д.

В статье представлен широкий анализ литературных российских и зарубежных источников, посвященных методам обнаружения объектов с использованием нейронных сетей, в том числе одноэтапных (YOLO, SSD, RetinaNet и др) и двухэтапных (Fast R-CNN, Faster R-CNN, Mask R-CNN и др.); методам обнаружения объектов на основе обучения с «нескольких выстрелов» (few-shot learning), которые пытаются решить задачу с использованием небольшого количества примеров; методам обнаружения объектов, основанным на обучении с «малым количеством выстрелов» (Multi-Scale Positive Sample Refinement for Few-Shot Object Detection (MPSR), Frustratingly Simple Few-Shot Object Detection (TFA).

Авторами статьи предложен метод, основанный на использовании нейронной сети на графах для описания структуры анализируемой сцены, проведен эксперимент по совместному использованию нейронных сетей на графах и обучения с «нескольких выстрелов» для увеличения точности обнаружения объектов. По заявлению авторов эксперимент показал эффективность разработанного метода. Средний показатель точности обнаружения объектов увеличился до пяти процентов. Также было показано, что предложенный метод позволяет достичь еще большей точности при увеличении объема обучающих данных.

Стиль и язык изложения материала является достаточно доступным для широкого круга читателей. Статья по объему соответствует рекомендуемому объему от 12 000 знаков.

Сформулирована цель исследования и научная новизна, которая заключается в совместном использовании нейронных сетей на графах и обучения с «нескольких выстрелов» для увеличения точности обнаружения объектов. Практическая значимость прослеживается из содержания статьи и заключается в повышении точности обнаружения объектов на изображениях.

Статья достаточно структурирована - в наличии введение, заключение, внутреннее членение основной части. Было бы целесообразным добавить о перспективах дальнейшего исследования.

Статья «Метод обнаружения объектов на изображениях на основе нейронных сетей на графах и небольшого количества обучающих примеров» рекомендована к публикации в журнале «Программные системы и вычислительные методы».