

УДК 004.93

Ю. А. Каспаров, В. М. Зуев

Федеральное государственное бюджетное научное учреждение
«Институт проблем искусственного интеллекта», г. Донецк
283048, г. Донецк, ул. Артема, 118 б

ОБУЧЕНИЕ МОДЕЛЕЙ КОМПЬЮТЕРНОГО ЗРЕНИЯ ДЛЯ ЗАДАЧИ ДЕТЕКЦИИ И СЕГМЕНТАЦИИ СТВОЛОВЫХ КЛЕТОК

U. A. Kasparov V. M. Zuev

Federal State Budgetary Scientific Institution «Institute of Artificial Intelligence Problems»
283048, Donetsk, Artem str., 118 b

TRAINING COMPUTER VISION MODELS FOR STEM CELL DETECTION AND SEGMENTATION

U. A. Kasparov, V. M. Zuev

Budgetary Foederati Publica Institutione Scientific Institutum Intelligentiae Artificialis Problems,
Donetsk 118 b Artyom str., Donetsk, 283048

EXERCITATIO MODELLORUM VISIONIS COMPUTATRICIS AD DETECTIONEM ET SEGMENTATIONEM CELLULARUM STEMICARUMB

В статье рассматриваются основные подходы обучения моделей компьютерного зрения для задачи детекции и сегментации стволовых клеток. Разработана и апробирована методология работы с изображениями высокого разрешения для задач сегментации изображения.

Ключевые слова: детекция, изображение, сегментация, высокое разрешение

This article examines key approaches to training computer vision models for stem cell detection and segmentation. A methodology for working with high-resolution images for image segmentation tasks is developed and tested.

Keywords: location: detection, image, segmentation, high resolution

Hic articulus modos clavis ad exempla visionis computatralis ad detectionem et segmentationem cellularum primordialium exercenda examinat. Methodologia ad imagines altae resolutionis ad segmentationem imaginum tractandam explicatur et probatur.

Affatus clavis: detectio, imago, segmentation, alta resolutio

Введение

В настоящее время актуальны работы по обнаружению и классификации объектов на больших изображениях с высоким разрешением, как например на изображении 4000×4000 пикселей или больше, и при этом объекты могут занимать 20×20 пикселей или менее. Такие задачи возникают при детекции объектов на снимках из космоса, при охране территорий, в медицине и в других приложениях.

Ранее мы имели успешный опыт работы с детекцией объектов на изображениях стандартных размеров с использованием YOLOv8 [1], [2], SSD [3], [4] и других методик [5-7]. В предыдущих проектах [8], [9] были достигнуты стабильные результаты детекции стандартных объектов с использованием прямоугольных ограничивающих рамок в формате YOLO [10-13]. Этот опыт показал эффективность современных архитектур компьютерного зрения для многих задач реального времени. Однако, в рамках разработки новых систем, в которых требовалось обнаружение мелких объектов на больших кадрах, как например мониторинг роста стволовых клеток, известные методы [14-18] показали отрицательный результат.

Актуальность работы. Основная цель работы – автоматизация процесса анализа динамики роста клеток, имеет важное значение для медицинских исследований. В отличие от предыдущих работ [19], [20], где работа велась с детекцией стандартных объектов, новая задача потребовала внесения коррекций в известные детекторы объектов [20,21] из-за имеющими специфические особенности исходных данных.

Постановка задачи. Ставилась задача детекции мелких объектов на больших изображениях с высоким разрешением.

Анализ исходных данных и выявление особенностей. На начальном этапе был проанализирован предоставленный набор данных. Он имел следующие ключевые особенности: Изображения высокого разрешения: 4000×4000 пикселей и более (в отличие от стандартных 640×640), клетки как объекты детекции большей частью мелкие и могут занимать 20×20 пикселей или менее, но четко различимые на исходном разрешении. При этом клетки имеют неправильную, сложную форму с высокой плотностью расположения и часто переплетаются фрагментами. Часто имеется наличие групп клеток, размеченных как единые объекты. В некоторых случаях невозможно точно определить границы между соседними клетками. В связи с наличием групповых объектов, классический подсчет количества клеток был признан некорректным подходом. В качестве основного метода анализа была выбрана оценка общей площади, занимаемой клетками, что требовало точного определения границ каждого объекта.

Основные метрики оценки качества моделей компьютерного зрения, используемые в данной работе, следующие [1-3].

1. IoU (Intersection over Union) – показатель степени пересечения над степенью объединения, определяет степень перекрытия между предсказанной областью и истинной разметкой, вычисляется как отношение площади пересечения к площади объединения, принимает значение от 0 до 1, где 1 – означает идеальное совпадение

2. mAP (mean Average Precision) – средняя точность, основная метрика для задач детекции и сегментации

Основные варианты mAP:

- mAP@0.5: AP при IoU=0.5 (стандартный порог)
- mAP@0.5:0.95: mAP для IoU от 0.5 до 0.95 с шагом 0.05 (более строгая метрика)
- mAP@0.75: AP при IoU=0.75 (высокий порог качества)

Предварительные эксперименты и выявление проблем

Изначально была предпринята попытка использовать YOLO [1], [2], хорошо зарекомендовавшую себя ранее. Однако сразу были выявлены проблемы: объекты мелких размеров относительно всего изображения плохо детектировались. Наблюдалась потеря деталей при стандартном уменьшении изображений, как это требует YOLO, а также неудовлетворительное качество сегментации сложных форм. Использовался разнообразный формат разметки: применялся как полигональный, так и рамками. Получив не удовлетворительный результат, мы вернулись к более древнему методу, а именно Mask R-CNN [16].

Для улучшения качества сегментации было принято решение использовать Mask R-CNN с фреймворком detectron2 [22], специально разработанным для задач сегментации.

В процессе обучения Mask R-CNN были выявлены следующие проблемы в выборе метрик обучения, а именно: показатели детекции при ограничивающих рамках стабильно росли: $mAP@0.5:0.95$ (bbox), $mAP@0.5$ (bbox), в то время, как показатели сегментации (segm) оставались нулевыми: $mAP@0.5:0.95$ (segm) = 0, $mAP@0.5$ (segm) = 0, $mAP@0.75$ (segm) = 0. Нулевые значения метрик сегментации указывали на то, что модель не могла научиться правильно определять границы объектов, хотя ограничивающие рамки детектировались удовлетворительно.

Поиск решений и последовательные улучшения

Были испробованы аугментация, предобработка изображений, валидация и коррекция аннотаций. Так, были опробованы различные методы аугментации данных: изменение яркости, контрастности, насыщенности, повороты, масштабирование, кадрирование, добавление шумов и артефактов. К сожалению, эти меры не привели к значительному улучшению показателей. Так, была опробована предобработка изображений: устранение зашумленности с помощью билатерального фильтра [23], увеличение контрастности методом CLAHE [24], усиление границ с использованием фильтра Собеля [25], нормализация яркости. Несмотря на значительное улучшение видимости и четкости клеток, это не привело к улучшению параметров классификации. В конце была проведена повторно валидация и коррекция аннотаций. Она заключалась в тщательной проверке набора данных на предмет незамкнутых полигонов, самопересекающихся контуров, неправильной ориентации точек разметки, некорректных масок сегментации. Разработан скрипт автоматической коррекции найденных ошибок.

Анализ причины проблем

Было установлено, что основная проблема заключается в принудительном уменьшении размера изображения, которая подаётся на нейросеть Mask R-CNN [16] или YOLO в процессе обучения нейросетевых моделей. Стандартными размерами этих способов является приведение исходного размера к входному размеру 640×640 или 1024×1024. При больших исходных входных размерах это приводит к потере детализации мелких объектов, невозможности корректной сегментации сложных форм, снижению информативности изображений

Ключевое решение – метод разбиения исходного изображения на фрагменты.

Был разработан способ разделения исходного изображения на фрагменты на фрагменты, который включает в себя:

- автоматический подбор размера целевого фрагмента: 1024×1024 или 640×640 пикселей;
- подбор перекрытия между патчами в районе 20% от целевого фрагмента;
- установка минимального количества объектов в фрагменте, равном 3;
- автоматическая адаптация под особенности каждого изображения.

Техническая реализация способа

В результате работы создан комплекс скриптов для автоматического разделения изображений на фрагменты, трансформации аннотаций в соответствии с новыми координатами, конвертации из COCO формата в формат YOLO, разделения на тренировочную и валидационную выборки (80/20)

В результате внедрения способа получены следующие результаты:

Количественные результаты: исходные изображения: 20 шт. (4000×4000) разделены на 180 фрагментов по (1024×1024) пикселей. Качественно улучшение: сохранена детализация всех объектов, исключена необходимость деструктивного автоматического уменьшения размера, увеличена эффективность использования памяти при обучении, достигнуты рабочие показатели детекции и сегментации

Выводы и перспективы

Разработана и апробирована методология работы с изображениями высокого разрешения для задач сегментации изображения. Ключевые элементы предложенного способа: обязательная предобработка изображений, тщательная валидация и коррекция аннотаций и главное - использование разбиения исходного изображения на фрагменты для сохранения детализации.

Практическая значимость способа в том, что: подготовленный набор данных и отработанная методология позволяют перейти к этапу промышленного использования системы мониторинга роста клеток.

Заключение

Известные способы детекции и сегментации малых объектов на больших изображениях часто могут оказаться неработоспособными. Предложенный способ обучения заключается в разбиении большого изображения на фрагменты. В результате работы создан комплекс скриптов для автоматического разделения изображений на фрагменты, трансформации аннотаций в соответствии с новыми координатами, конвертации из COCO формата в формат YOLO, и в конечном итоге детекцию мелких объектов, которая обычными методами была неудовлетворительной.

Работа выполнена в рамках работы по разработке теоретических основ обнаружения и оценки объектов интеллектуальными робототехническими системами в условиях быстро меняющейся обстановки.

RESUME

U.A Kasparov. V.M. Zuev

Training computer vision models for stem cell detection and segmentation

Known methods for detecting and segmenting small objects in large images, can often prove ineffective. The proposed training method involves splitting a large image into fragments. As a result, a set of scripts was created for automatically splitting images into fragments, transforming annotations according to new coordinates, converting from COCO to YOLO format, ultimately, reliable detection of small objects, which was unsatisfactory using conventional methods, was achieved.

РЕЗЮМЕ

Ю.А. Каспаров В.М. Зувев

Обучение моделей компьютерного зрения для задачи детекции и сегментации стволовых клеток

Известные способы детекции, такие как, и сегментации малых объектов на больших изображениях часто могут оказаться неработоспособными. Предложенный способ обучения заключается в разбиении большого изображения на фрагменты. В результате работы создан комплекс скриптов для автоматического разделения изображений на фрагменты, трансформации аннотаций в соответствии с новыми координатами, конвертации из COCO формата в формат YOLO, и в конечном итоге достигнута уверенная детекция мелких объектов, которая обычными методами была неудовлетворительной.

Список литературы

1. Terven, Juan; Córdova-Esparza, Diana-Margarita; Romero-González, Julio-Alejandro (2023-11-20). "A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS". Machine Learning and Knowledge Extraction. 5 (4): 1680–1716. [arXiv:2304.00501](#). doi:10.3390/make5040083. ISSN 2504-4990.
2. "YOLOv2: Real-Time Object Detection". pjdreddie.com. Retrieved 2024-09-12.
3. Menke, Maximilian; Wenzel, Thomas; Schwung, Andreas (October 2022). "Improving GAN-based Domain Adaptation for Object Detection". 2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC). pp. 3880–3885. doi:10.1109/ITSC55140.2022.9922138. ISBN 978-1-6654-6880-0. S2CID 253251380.
4. Menke, Maximilian; Wenzel, Thomas; Schwung, Andreas (2022-08-31). "AWADA: Attention-Weighted Adversarial Domain Adaptation for Object Detection". [arXiv:2208.14662](#).
5. Oza, Poojan; Sindagi, Vishwanath A.; VS, Vibashan; Patel, Vishal M. (2021-07-04). "Unsupervised Domain Adaptation of Object Detectors: A Survey". [arXiv:2105.13502](#).
6. Khodabandeh, Mehran; Vahdat, Arash; Ranjbar, Mani; Macready, William G. (2019-11-18). "A Robust Learning Approach to Domain Adaptive Object Detection". [arXiv:1904.02361](#).
7. Soviany, Petru; Ionescu, Radu Tudor; Rota, Paolo; Sebe, Nicu (2021-03-01). "Curriculum self-paced learning for cross-domain object detection". Computer Vision and Image Understanding. [arXiv:1911.06849](#). doi:10.1016/j.cviu.2021.103166. ISSN 1077-3142. S2CID 208138033.
8. Зувев В.М. Сравнение обнаружения объектов средствами искусственного интеллекта в сравнении с классическими методами *Проблемы искусственного интеллекта*. 2024. № 3(34). С. 30-35. ISSN 2413-7383. DOI 10.24412/2413-7383-2024-3-30-35.
9. Зувев В. М. Ортогонализация весовых коэффициентов нейросети *Журнал «Проблемы искусственного интеллекта»*. 2025. №1 (36). С. 135–140. ISSN 2413-7383. DOI 10.24412/2413-7383-135-140.
10. Rémy, Philippe (2024-09-05), [//github.com/philipperemy/yolo-9000](#), retrieved 2024-09-12
11. Bochkovski, Alexey; Wang, Chien-Yao; Liao, Hong-Yuan Mark (2020-04-22). "YOLOv4: Optimal Speed and Accuracy of Object Detection". [arXiv:2004.10934](#).

12. Wang, Chien-Yao; Bochkovskiy, Alexey; Liao, Hong-Yuan Mark (2021-02-21). "Scaled-YOLOv4: Scaling Cross Stage Partial Network". [arXiv:2011.08036 \[cs.CV\]](#).
13. Li, Chuyi; Li, Lulu; Jiang, Hongliang; Weng, Kaiheng; Geng, Yifei; Li, Liang; Ke, Zaidan; Li, Qingyuan; Cheng, Meng; Nie, Weiqiang; Li, Yiduo; Zhang, Bo; Liang, Yufei; Zhou, Linyuan; Xu, Xiaoming (2022-09-07). "YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications". [arXiv:2209.02976 \[cs.CV\]](#).
14. Zhu, Jun-Yan; Park, Taesung; Isola, Phillip; Efros, Alexei A. (2020-08-24). "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks". [arXiv:1703.10593 \[cs.CV\]](#).
15. Dalal, Navneet (2005). [//Hal.inria.fr/file/index/docid/548512/filename/hog_cvpr2005.pdf/](#). "Histograms of oriented gradients for human detection" (PDF). Computer Vision and Pattern Recognition. 1.
16. Pang, Jiangmiao; Chen, Kai; Shi, Jianping; Feng, Huajun; Ouyang, Wanli; Lin, Dahua (2019-04-04). "Libra R-CNN: Towards Balanced Learning for Object Detection". [arXiv:1904.02701v1](#).
17. Zhang, Shifeng (2018). "Single-Shot Refinement Neural Network for Object Detection". Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4203–4212. [arXiv:1711.06897](#).
18. Lin, Tsung-Yi (2020). "Focal Loss for Dense Object Detection". IEEE Transactions on Pattern Analysis and Machine Intelligence. 42 (2): 318–327. [arXiv:1708.02002](#). Bibcode:2020ITPAM..42..318L. doi:10.1109/TPAMI.2018.2858826. PMID 30040631. S2CID 47252984.
19. Zhu, Xizhou. "Deformable ConvNets v2: More Deformable, Better Results". [arXiv:1811.11168](#).
20. Dai, Jifeng "Deformable Convolutional Networks". [arXiv:1703.06211](#).
21. Alsanabani, Ala; Ahmed, Mohammed; AL Smadi, Ahmad (2020). "Vehicle Counting Using Detecting-Tracking Combinations: A Comparative Analysis". 2020 the 4th International Conference on Video and Image Processing. pp. 4854. doi:10.1145/3447450.3447458. ISBN 9781450389075. S2CID 233194604.
22. M. M. Rahman and T. Finin, "Deep understanding of a documents structure," in 4th IEEE/ACM Int. Conf. on Big Data Computing, Applications and Technologies, December 2017.
23. R. Kimmel, R. Malladi, and N. Sochen. Images as embedding maps and minimal surfaces: Movies, color, and volumetric medical images. IEEE CVPR'97, pp. 350-355, Puerto Rico, June 17–19, 1997. [https://www.cs.technion.ac.il/~ron/PAPERS/cvpr97.pdf](#)
24. Коротков В. А, Коротков К. В., Новичихин Е. П. Модификация метода CLANE для компенсации влияния гидрометеоров *Журнал радиоэлектроники*, ISSN 1684-1719, №10, 2017 с1-21.
25. Ватутин Э.И., Мирошниченко С.Ю., Титов В.С. Программная оптимизация оператора Собела с использованием SIMD-расширений процессоров семейства x86. *Телекоммуникации*. 2006. № 6. С. 12—16. (2006)

References

1. Terven, Juan; Córdova-Esparza, Diana-Margarita; Romero-González, Julio-Alejandro (2023-11-20). "A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS". Machine Learning and Knowledge Extraction. 5 (4): 1680–1716. [arXiv:2304.00501](#). doi:10.3390/make5040083. ISSN 2504-4990.
2. "YOLOv2: Real-Time Object Detection". [pjreddie.com](#). Retrieved 2024-09-12.
3. Menke, Maximilian; Wenzel, Thomas; Schwung, Andreas (October 2022). "Improving GAN-based Domain Adaptation for Object Detection". 2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC). pp. 3880–3885. doi:10.1109/ITSC55140.2022.9922138. ISBN 978-1-6654-6880-0. S2CID 253251380.
4. Menke, Maximilian; Wenzel, Thomas; Schwung, Andreas (2022-08-31). "AWADA: Attention-Weighted Adversarial Domain Adaptation for Object Detection". [arXiv:2208.14662](#).
5. Oza, Poojan; Sindagi, Vishwanath A.; VS, Vibashan; Patel, Vishal M. (2021-07-04). "Unsupervised Domain Adaptation of Object Detectors: A Survey". [arXiv:2105.13502](#).
6. Khodabandeh, Mehran; Vahdat, Arash; Ranjbar, Mani; Macready, William G. (2019-11-18). "A Robust Learning Approach to Domain Adaptive Object Detection". [arXiv:1904.02361](#).
7. Soviany, Petru; Ionescu, Radu Tudor; Rota, Paolo; Sebe, Nicu (2021-03-01). "Curriculum self-paced learning for cross-domain object detection". Computer Vision and Image Understanding. [arXiv:1911.06849](#). doi:10.1016/j.cviu.2021.103166. ISSN 1077-3142. S2CID 208138033.
8. Zuev V.M Comparison of object detection by means of artificial intelligence in comparison with classical methods [Text]//V.M. Zuev. Journal "Problems of Artificial Intelligence". – 2024. – № 3(34). – С. 30-35. – ISSN 2413-7383.-DOI 10.24412/2413-7383-2024-3-30-35.

9. Zuev V.M. Orthogonalization of neural network weight coefficients [Text] // Journal "Problems of Artificial Intelligence". – 2025. – №1 (36). – С. 135–140. – ISSN 2413-7383. – DOI 10.24412/2413-7383-135-140.
10. Rémy, Philippe (2024-09-05), <https://github.com/philipperemy/yolo-9000>, retrieved 2024-09-12
11. Bochkovskiy, Alexey; Wang, Chien-Yao; Liao, Hong-Yuan Mark (2020-04-22). "YOLOv4: Optimal Speed and Accuracy of Object Detection". arXiv:2004.10934.
12. Wang, Chien-Yao; Bochkovskiy, Alexey; Liao, Hong-Yuan Mark (2021-02-21). "Scaled-YOLOv4: Scaling Cross Stage Partial Network". arXiv:2011.08036 [cs.CV].
13. Li, Chuyi; Li, Lulu; Jiang, Hongliang; Weng, Kaiheng; Geng, Yifei; Li, Liang; Ke, Zaidan; Li, Qingyuan; Cheng, Meng; Nie, Weiqiang; Li, Yiduo; Zhang, Bo; Liang, Yufei; Zhou, Linyuan; Xu, Xiaoming (2022-09-07). "YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications". arXiv:2209.02976 [cs.CV].
14. Zhu, Jun-Yan; Park, Taesung; Isola, Phillip; Efros, Alexei A. (2020-08-24). "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks". arXiv:1703.10593 [cs.CV].
15. Dalal, Navneet (2005). http://hal.inria.fr/file/index/docid/548512/filename/hog_cvpr2005.pdf. "Histograms of oriented gradients for human detection" (PDF). Computer Vision and Pattern Recognition. 1.
16. Pang, Jiangmiao; Chen, Kai; Shi, Jianping; Feng, Huajun; Ouyang, Wanli; Lin, Dahua (2019-04-04). "Libra R-CNN: Towards Balanced Learning for Object Detection". arXiv:1904.02701v1.
17. Zhang, Shifeng (2018). "Single-Shot Refinement Neural Network for Object Detection". Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4203–4212. arXiv:1711.06897.
18. Lin, Tsung-Yi (2020). "Focal Loss for Dense Object Detection". IEEE Transactions on Pattern Analysis and Machine Intelligence. 42 (2): 318–327. arXiv:1708.02002. Bibcode:2020ITPAM..42..318L. doi:10.1109/TPAMI.2018.2858826. PMID 30040631. S2CID 47252984.
19. Zhu, Xizhou. "Deformable ConvNets v2: More Deformable, Better Results". arXiv:1811.11168.
20. Dai, Jifeng "Deformable Convolutional Networks". arXiv:1703.06211.
21. Alsanabani, Ala; Ahmed, Mohammed; AL Smadi, Ahmad (2020). "Vehicle Counting Using Detecting-Tracking Combinations: A Comparative Analysis". 2020 the 4th International Conference on Video and Image Processing. pp. 4854. doi:10.1145/3447450.3447458. ISBN 9781450389075. S2CID 233194604.
22. M. M. Rahman and T. Finin, "Deep understanding of a documents structure," in 4th IEEE/ACM Int. Conf. on Big Data Computing, Applications and Technologies, December 2017.
23. R. Kimmel, R. Malladi, and N. Sochen. Images as embedding maps and minimal surfaces: Movies, color, and volumetric medical images. IEEE CVPR'97, pp. 350-355, Puerto Rico, June 17–19, 1997. <https://www.cs.technion.ac.il/~ron/PAPERS/cvpr97.pdf>
24. V.A. Korotkov, K.V. K. B. Korotkov, E.P. Novochichin Модификация метода CLANE для компенсации влияния гидрометеоров Журнал радиоэлектроники, ISSN 1684-1719, №10, 2017 с1-21.
25. E.I. Vatutin ?S.U. Miroshnichenko, V.S. Titov. Программная оптимизация оператора Собела с использованием SIMD-расширений процессоров семейства x86. Телекоммуникации. 2006. № 6. С. 12—16. (2006)

Каспаров Юрий Ашотович – младший научный сотрудник, Федеральное государственное бюджетное научное учреждение «Институт проблем искусственного интеллекта», г. Донецк. *Область научных интересов:* интеллектуальные робототехнические системы. эл. почта kua.ipai@mail.ru, адрес: 283048, г. Донецк, ул. Артема, д. 118 б, телефон: +7949 4048702

Зуев Владимир Михайлович -Научный сотрудник, Зав.отделом, Федеральное государственное бюджетное научное учреждение «Институт проблем искусственного интеллекта», г. Донецк. *Область научных интересов:* интеллектуальные робототехнические системы. эл. почта zvm.ipai@mail.ru, адрес: 283048, г. Донецк, ул. Артема, д. 118 б, телефон: +7949 4048702

Статья поступила в редакцию 02.10.2025.