

УДК 004.896

Применение методов глубокого обучения для моделирования проходимости ровера и захватываемости предметов

Карташев Всеволод Владимирович^{1(*)}

vsevolod.kartashev@gmail.com

Костин Артем Абдулхакимович²

kostin11artem@gmail.com

Бадртдинов Максим Рашидович²

badrtdinov.maksim@mail.ru

¹ ИПМ им. М.В. Келдыша РАН, Москва, Россия² ФКИ МГУ имени М.В. Ломоносова, Москва, Россия

Аннотация. В работе рассматривается применение методов глубокого обучения для задач роботизированного зрения, включая захват объектов и автономную навигацию. Анализируются архитектуры SSD и Mask R-CNN для детекции и сегментации промышленных деталей, а также U-Net++ с энкодером EfficientNet для навигации планетохода по данным семантической сегментации. Сравниваются их точность, скорость и приводятся рекомендации по внедрению в реальные системы, от производства до исследования космоса.

Ключевые слова: робототехника, глубокое обучение, Single Shot Detector, Mask R-CNN, сегментация изображений

Введение. Современная робототехника немыслима без продвинутых систем технического зрения, которые служат «глазами» для роботов. Традиционные методы компьютерного зрения часто не справляются со сложностью реальных сцен, поэтому необходимо применять подходы на основе глубокого обучения. В данной работе исследуются нейросетевые архитектуры для двух задач: детекции объектов для роботизированного захвата и семантической сегментации для навигации планетохода.

Методы и материалы; результаты. В задаче детекции и сегментации для роботизированного захвата для задач, где роботу-манипулятору нужно точно захватить объект, используются два основных подхода:

Single Shot Detector (SSD): Архитектура, нацеленная на высокую скорость. Она обнаруживает объекты за один проход, что идеально для систем реального времени, где быстрота реакции важнее миллиметровой точности.

Mask R-CNN: Более сложная и точная архитектура. Она не только находит объект, но и создает его попиксельную маску (сегментацию). Это критически важно для задач точного захвата, но требует больше времени на обработку.

В экспериментах Mask R-CNN достиг точности сегментации 90 % по метрике IoU, в то время как SSD показал точность детекции 85 %.

В задаче семантической сегментации для автономной навигации для навигации планетохода необходимо понимать всю сцену, разделяя безопасную поверхность и препятствия. Здесь применяется семантическая сегментация, классифицирующая каждый пиксель изображения (например, «грунт»,

«камень»). Была выбрана связка U-Net++ с энкодером EfficientNet. U-Net++ — это улучшенная версия U-Net, которая эффективно извлекает и восстанавливает признаки на разных масштабах, повышая точность. EfficientNet в качестве энкодера обеспечивает уникальный баланс между точностью и вычислительной эффективностью, что критично для бортовых систем.

Для корректной работы с несбалансированными данными (когда препятствия занимают малую часть изображения) применялась комбинированная функция потерь TverskyFocalLoss.

В эксперименте на синтетических изображениях лунной поверхности модель достигла высокой точности (метрика Жаккара 0.92) и была успешно интегрирована в программный комплекс для обхода препятствий в реальном времени.

Заключение. Результаты подтверждают высокую эффективность глубокого обучения в робототехнике. Для точного захвата рекомендуется Mask R-CNN, для задач реального времени — SSD. Для автономной навигации в сложных средах отлично зарекомендовала себя архитектура U-Net++ с энкодером EfficientNet. Эти подходы значительно расширяют возможности современных роботизированных систем.

Список источников

- [1] Liu W. et al. SSD: Single shot multibox detector. *Proceedings of Computer Vision – ECCV 2016: 14th European Conference*, Amsterdam, The Netherlands, Springer International Publishing, 2016, part I, pp. 21-37. <https://doi.org/10.48550/arXiv.1512.02325>
- [2] Lin T.Y. et al. Feature pyramid networks for object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2117-2125. <https://doi.org/10.48550/arXiv.1612.03144>
- [3] Qassim H., Feinzimer D., Verma A. Residual Squeeze VGG16. *Preprint arXiv*, 2017, <https://doi.org/10.48550/arXiv.1705.03004>
- [4] Murphy A.H. The Finley affair: A signal event in the history of forecast verification. *Weather and forecasting*, 1996, vol. 11, no. 1, pp. 3–20. [https://doi.org/10.1175/1520-0434\(1996\)011%3C0003:TFAASE%3E2.0.CO;2](https://doi.org/10.1175/1520-0434(1996)011%3C0003:TFAASE%3E2.0.CO;2)
- [5] *SSD: Single Shot MultiBox Detector — a PyTorch Tutorial to Object Detection*. URL: <https://github.com/sgrvinod/a-PyTorch-Tutorial-to-Object-Detection> (accessed 15.11.2024).
- [6] Lin T.Y. et al. Microsoft COCO: Common Objects in Context. *Computer Vision – ECCV 2014. Proceedings of 13th European Conference*, Zurich, Switzerland, Springer International Publishing, 2014, part V, pp. 740–755.
- [7] *COCO: Common Objects in Context*. URL: <https://cocodataset.org/#home> (accessed 09.11.2023).
- [8] Панина В.С., Амеличев Г.Э., Белов Ю.С. Интеллектуальная парковочная система как часть интеллектуальной транспортной системы. *E-Scio*, 2022, № 1 (64), с. 445–452.