

НЕЙРОСЕТЕВЫЕ МЕТОДЫ ДЕТЕКЦИИ И СЕГМЕНТАЦИИ ДЛЯ УПРАВЛЕНИЯ СТРОЕМ ГРУППЫ МОБИЛЬНЫХ РОБОТОВ НА ОСНОВЕ BEV-ПРЕДСТАВЛЕНИЯ ОТ БПЛА

Беляев С.И., Рабцевич Г.С., Базаргармаев Н.Д.

Московский финансово-промышленный университет «Синергия», Москва

Ключевые слова: нейросетевой детектор, YOLOv8, семантическая сегментация, U-Net, Bird's-Eye-View, управление строем, мобильные роботы, БПЛА, карта занятости, компьютерное зрение.

Аннотация. В статье рассматривается применение нейросетевых методов детекции и семантической сегментации изображений для координированного управления строем группы наземных мобильных роботов. Предлагается метод, основанный на представлении сцены «вид сверху», формируемом беспилотным летательным аппаратом. Обнаружение роботов на карте «вид сверху» осуществляется нейросетевым детектором YOLOv8, а классификация проходимости поверхности – с помощью архитектуры U-Net. Описаны функции потерь, метрики качества и архитектурные особенности применяемых моделей. Результаты моделирования подтверждают эффективность предложенного подхода по сравнению с методами, использующими только бортовые сенсоры.

NEURAL NETWORK DETECTION AND SEGMENTATION METHODS FOR FORMATION CONTROL OF A MOBILE ROBOT GROUP BASED ON UAV BEV REPRESENTATION

Beliaev S.I., Rabtsevich G.S., Bazargarmaev N.D.

Moscow University for Industry and Finance "Synergy", Moscow

Keywords: neural network detector, YOLOv8, semantic segmentation, U-Net, Bird's-Eye-View, formation control, mobile robots, UAV, occupancy grid, computer vision.

Abstract. This paper addresses the application of neural network-based object detection and semantic segmentation methods for coordinated formation control of a group of ground mobile robots. The proposed approach utilizes a Bird's-Eye-View representation of the scene generated by an accompanying unmanned aerial vehicle. Robot detection on the Bird's-Eye-View map is performed using the YOLOv8 neural network detector, while surface traversability classification is carried out using the U-Net architecture. The paper describes loss functions, quality metrics, and architectural features of the employed models. Simulation results confirm the effectiveness of the proposed approach compared to methods relying solely on onboard sensors.

Введение

Задача координированного управления группой мобильных роботов является одной из ключевых в современной робототехнике, находя применение в логистике, поисково-спасательных операциях и мониторинге территорий [1, 2]. Системы из нескольких роботов обладают повышенной гибкостью, адаптивностью и отказоустойчивостью по сравнению с одиночными роботами, однако создание и поддержание строя в динамически изменяющейся среде остаётся сложной задачей. Традиционные подходы к управлению строем – «ведущий-ведомый», виртуальные структуры и поведенческие методы – как правило, опираются на данные бортовых сенсоров отдельных роботов, что ограничивает область обзора и затрудняет глобальную координацию [2, 3].

В данной работе предлагается метод управления строем группы наземных мобильных роботов на основе представления сцены «вид сверху» (Bird's-Eye-View, BEV), формируемого сопровождающим беспилотным летательным аппаратом (БПЛА). Ключевой акцент сделан на нейросетевых компонентах системы: детекторе YOLOv8 для обнаружения роботов и архитектуре U-Net для семантической сегментации проходимости поверхности.

Формирование BEV-представления

БПЛА, оснащённый камерой и направленной вертикально вниз, выступает в роли внешнего наблюдателя рабочей области. Изображение с камеры дрона преобразуется в BEV-карту плоскости земли с помощью гомографического преобразования [4]. Матрица гомографии H размерностью 3×3 устанавливает соответствие между точками изображения и координатами на плоскости земли:

$$\mathbf{x}' = H\mathbf{x}, \quad (1)$$

где $\mathbf{x}' = (X, Y, 1)^T$ – однородные координаты точки на изображении, $\mathbf{x} = (u, v, 1)^T$ – однородные координаты соответствующей точки на плоскости земли, H – матрица гомографии.

Матрица H вычисляется методом прямого линейного преобразования (DLT) с последующей фильтрацией выбросов алгоритмом RANSAC на основе не менее четырёх пар соответствий [4]. Это позволяет перейти от перспективной проекции камеры к метрическому представлению сцены, пригодному для задач управления.

Нейросетевая детекция роботов на основе YOLOv8

На полученном BEV-изображении выполняется обнаружение наземных роботов с помощью нейросетевого детектора семейства YOLO [1, 5]. Архитектура YOLOv8 реализует одностадийную (one-stage) детекцию объектов, объединяя в единой модели извлечение признаков, генерацию предложений и классификацию. Модель YOLOv8m (Medium) обеспечивает обнаружение объектов в реальном времени с точностью $mAP@0.5 = 50,2\%$ при скорости вывода менее 2 мс на GPU [5].

Обучение детектора осуществляется посредством минимизации составной функции потерь:

$$L_{\text{total}} = \lambda_{\text{cls}} L_{\text{cls}} + \lambda_{\text{box}} L_{\text{box}} + \lambda_{\text{dfl}} L_{\text{dfl}}, \quad (2)$$

где L_{cls} – потери классификации, вычисляемые с помощью бинарной кросс-энтропии (BCE); L_{box} – потери локализации ограничивающих рамок на основе метрики CIoU (Complete Intersection over Union); L_{dfl} – потери распределённой фокальной функции (Distribution Focal Loss), обеспечивающие уточнение границ рамок; $\lambda_{\text{cls}}, \lambda_{\text{box}}, \lambda_{\text{dfl}}$ – весовые коэффициенты, балансирующие вклад каждой компоненты [1, 5].

Метрика CIoU учитывает не только площадь пересечения предсказанной и истинной рамок, но и расстояние между их центрами, а также соотношение сторон:

$$L_{\text{CIoU}} = 1 - \text{IoU} + \frac{\rho^2(\mathbf{b}, \mathbf{b}_{\text{gt}})}{c^2} + \alpha v, \quad (3)$$

где IoU – Intersection over Union предсказанной и истинной рамок; $\rho(b, b_{\text{gt}})$ – евклидово расстояние между центрами рамок b и b_{gt} ; c – диагональ минимального охватывающего прямоугольника; v – параметр, отражающий согласованность соотношений сторон; α – адаптивный весовой коэффициент [1].

Координаты обнаруженных роботов в пикселях BEV-карты пересчитываются в мировые координаты с помощью матрицы гомографии (1), что позволяет определить текущую конфигурацию строя.

Семантическая сегментация поверхности на основе U-Net

Дополнительно к задаче детекции решается задача классификации проходимости поверхности. Для обеспечения семантической сегментации BEV-изображений используют архитектуру нейронных сетей U-Net [6]. Ее особенность заключается в том, что она делит изображение на два класса: проходимая и непроходимая области. В результате разделения происходит построение карты занятости (occupancy grid) рабочего пространства [7].

В состав архитектуры моделей семейства U-Net входят следующие компоненты:

1. Кодировщик, отвечающий за последовательное снижение пространственного разрешения и извлечение признаков.
2. Декодировщик, восстанавливающий исходное разрешение.

Для обучения модели используется комбинированная функция потерь:

$$L_{\text{seg}} = L_{\text{BCE}} + (1 - \text{Dice}), \quad (4)$$

где L_{BCE} – бинарная кросс-энтропия; Dice – коэффициент Дайса определяемый как:

$$\text{Dice} = \frac{2 |P \cap G|}{|P| + |G|}, \quad (5)$$

где P – множество пикселей, предсказанных как целевой класс; G – множество пикселей истинной разметки.

Коэффициент Дайса принимает значения от 0 до 1, где 1 это идеальное совпадение предсказания. Использование комбинации BCE и Dice позволяет учитывать как попиксельную точность, так и общее перекрытие областей, что особенно важно при дисбалансе классов [5].

Качество сегментации оценивается метрикой IoU (Intersection over Union):

$$\text{IoU} = \frac{|P \cap G|}{|P \cup G|}. \quad (6)$$

Карта занятости и управление строем

Карта занятости представляется в виде вероятностной сетки, где каждая ячейка хранит оценку вероятности занятости, обновляемую по формуле байесовского логарифмического обновления [7]:

$$l(m_i | z_{1:t}) = l(m_i | z_{1:t-1}) + l(m_i | z_t) - l_0, \quad (7)$$

где $l(m_i | z_{1:t})$ – логарифм отношения шансов занятости ячейки m_i после наблюдений $z_{1:t}$; $l(m_i | z_t)$ – обновление на основании текущего наблюдения z_t ; l_0 – априорное значение.

Использование логарифмического представления обеспечивает численную стабильность и эффективность вычислений [7].

На основе BEV-представления и карты занятости формируется целевая конфигурация строя с учётом препятствий. Для управления строем применяется алгоритм на основе распределённого консенсуса [8, 9]:

$$u_i = - \sum_j a_{ij} (x_i - x_j) - (d_i - d_j), \quad (8)$$

где u_i – управляющее воздействие i -го робота; a_{ij} – коэффициент связности между

роботами i и j в графе взаимодействия; x_i, x_j – текущие позиции роботов; d_i, d_j – желаемые позиции в строю.

В отличие от классических распределённых подходов, предлагаемый метод использует централизованное BEV-наблюдение от БПЛА, что снимает проблему ограниченного поля зрения бортовых камер [3, 10].

Результаты моделирования

Предложенный подход апробирован в среде моделирования с использованием набора из пяти виртуальных роботов. В качестве планировщика опорных траекторий использован алгоритм A* на карте занятости, а локальное управление каждым роботом осуществлялось контроллером на основе метода динамического окна (DWA). Предварительные результаты моделирования показали, что применение BEV-представления от дрона позволяет обеспечить формирование и удержание строя с меньшей средней ошибкой позиционирования по сравнению с подходами, использующими только бортовые сенсоры роботов. Результаты моделирования работы системы представлены на рисунке 1.

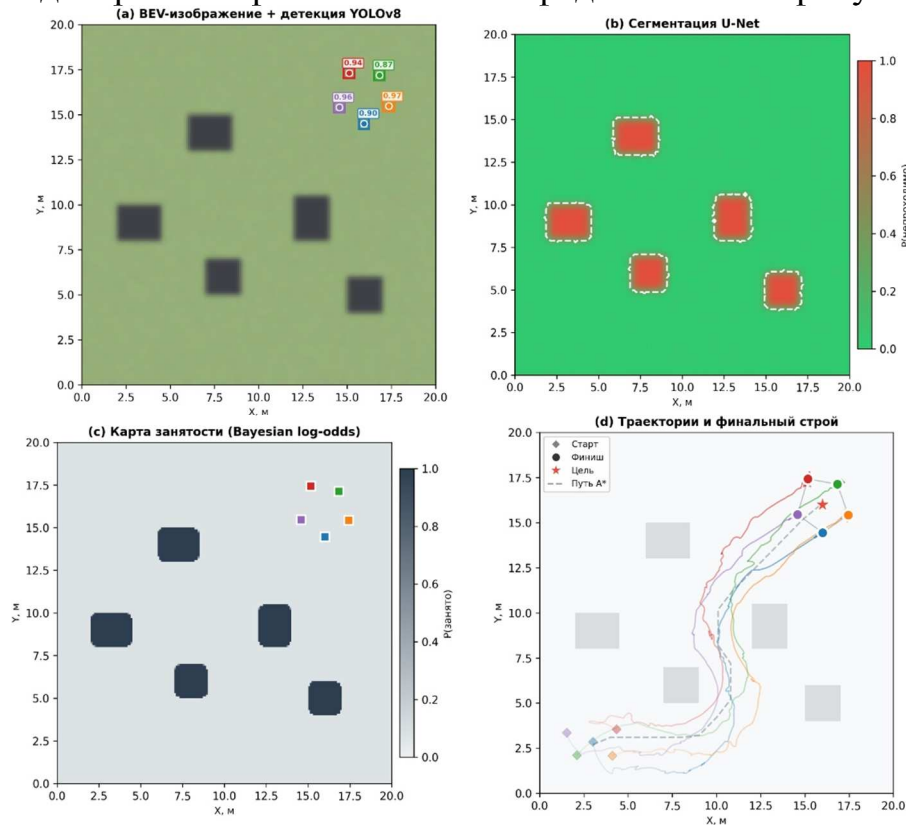


Рис. 1. Результаты моделирования работы системы

Для компенсации временной потери детекции отдельных роботов предусмотрена экстраполяция их положения на основе фильтра Калмана по предшествующим наблюдениям. Модель YOLOv8m обеспечила стабильное обнаружение роботов при различных условиях освещения и масштабах съёмки, а сегментация U-Net продемонстрировала устойчивую классификацию поверхности с $\text{IoU} > 0,85$ для класса «проходимая область» на валидационной выборке. Тем самым показано, что использование внешнего BEV-наблюдателя даёт выигрыш не только в качестве восприятия, но и в итоговом качестве формирования строя, измеряемом по средней позиционной ошибке и устойчивости к пропаданию отдельных детекций.

Заключение

Предложенный метод объединяет нейросетевую детекцию объектов (YOLOv8), семантическую сегментацию (U-Net), геометрическое преобразование изображений и консенсусное управление строем, обеспечивая эффективное решение задачи группового управления на основе визуальной информации от внешнего БПЛА-наблюдателя. В отличие от большинства работ по vision-based formation control, использующих локальные бортовые камеры роботов, предложенный подход опирается на централизованное BEV-представление от БПЛА, совмещённое с нейросетевой детекцией и семантической сегментацией, что обеспечивает глобальное восприятие сцены и явную карту занятости. Это позволяет интегрировать современные методы компьютерного зрения непосредственно в классические схемы консенсусного управления строем и получить количественное снижение ошибки позиционирования по сравнению с системами, основанными только на бортовых сенсорах. Введение формального описания функций потерь детектора и сегментационной модели, а также метрик качества позволяет обеспечить воспроизводимость результатов и обоснованность выбора архитектур. Дальнейшие исследования могут быть направлены на применение более современных архитектур детекции (RT-DETR) и интеграцию модулей оценки глубины сцены из BEV-представления.

Список литературы

1. Redmon J., Divvala S., Girshick R., Farhadi A. You Only Look Once: Unified, Real-Time Object Detection // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016, pp. 779-788.
2. Oh K.-K., Park M.-C., Ahn H.-S. A Survey of Multi-Agent Formation Control // Automatica. 2015, vol. 53, pp. 424-440.
3. Guan R. Vision-Based Distributed Formation Tracking Control and Coordination of Multi-Mobile Robot Systems: PhD Thesis. – Nanyang Technological University, Singapore, 2024.
4. Hartley R., Zisserman A. Multiple View Geometry in Computer Vision. – 2nd ed. – Cambridge University Press, 2003.
5. Jocher G. et al. Ultralytics YOLOv8 [Electronic resource]. – Access mode: <https://docs.ultralytics.com/models/yolov8/>.
6. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation // Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer, 2015, pp. 234-241.
7. Elfes A. Using Occupancy Grids for Mobile Robot Perception and Navigation // Computer. 1989, vol. 22, no. 6, pp. 46-57.
8. Ren W., Beard R.W. Distributed Consensus in Multi-vehicle Cooperative Control: Theory and Applications. – Springer-Verlag London, 2008. – 319 p.
9. Alonso-Mora J., Baker S., Rus D. Distributed Multi-Robot Formation Control in Dynamic Environments // Autonomous Robots. 2019, vol. 43, pp. 1079-1100.
10. Montijano E., Cristofalo E., Zhou D., Schwager M., Sagüés C. Vision-Based Distributed Formation Control without an External Positioning System // Transactions on Robotics. 2016, vol. 32, no. 2, pp. 339-351.

Сведения об авторах:

Беляев Семен Игоревич – аспирант;

Рабцевич Герман Сергеевич – аспирант;

Базаргармаев Нима Дондокович – аспирант.