

Хомутов Денис Александрович, бакалавр
Уфимский университет науки и технологий
Khomutov Denis Aleksandrovich
Ufa University of Science and Technology

СЕМАНТИЧЕСКАЯ СЕГМЕНТАЦИЯ ИЗОБРАЖЕНИЙ ДЛЯ АНАЛИЗА ОКРУЖАЮЩЕГО ПРОСТРАНСТВА МОБИЛЬНОГО РОБОТА SEMANTIC IMAGE SEGMENTATION FOR ANALYSIS OF A MOBILE ROBOT'S ENVIRONMENT

Аннотация. Рассмотрено применение семантической сегментации для анализа окружающего пространства мобильного робота в помещениях и на открытых территориях. Описаны основные архитектурные подходы и способы выделения проходимых областей на основе пиксельной классификации. Показано, какие навигационные признаки могут быть получены из сегментационной маски, а также обозначены ограничения метода и требования к обработке видеопотока в реальном времени.

Abstract. The use of semantic segmentation for analyzing the environment of a mobile robot indoors and in outdoor areas is considered. The main architectural approaches and methods for identifying traversable regions based on pixel-wise classification are described. The navigation cues that can be obtained from a segmentation mask are presented; the limitations of the method and the requirements for real-time video-stream processing are also outlined.

Ключевые слова: Семантическая сегментация, мобильная робототехника, компьютерное зрение, свободное пространство, нейронные сети, автономная навигация.

Keywords: Semantic segmentation, mobile robotics, computer vision, free space, neural networks, autonomous navigation.

Введение

Автономное движение мобильного робота требует понимания того, какие области наблюдаемой сцены доступны для перемещения. Камера предоставляет плотную визуальную информацию, однако исходное изображение не содержит явного разделения на проходимые поверхности, границы движения, препятствия и другие семантические классы. Семантическая сегментация решает эту задачу посредством классификации каждого пикселя и формирует пространственно согласованное представление сцены [1, 2].

В отличие от детекции объектов, сегментация сохраняет форму областей и границ между ними. Для мобильной робототехники это позволяет как в помещениях, так и на открытых территориях выделять проходимые поверхности и их границы, элементы инфраструктуры, растительность, людей, транспортные средства и иные препятствия. Полученная маска может использоваться для оценки свободного пространства и локального направления движения. Цель статьи – систематизировать методы семантической сегментации, применимые для анализа окружающей среды мобильного робота.

Постановка задачи и роль семантической сегментации

Пусть входным сигналом является изображение I размером $H \times W$, а множество допустимых классов содержит C категорий. Для каждой позиции (u, v) модель оценивает вероятности принадлежности пикселя классам и выбирает класс с максимальной вероятностью: $c^*(u, v) = \arg \max_p (p(c | I, u, v), c \in C$, где $c^*(u, v)$ – итоговая метка пикселя. Результатом становится маска того же пространственного размера, что и исходный кадр.

Для навигации полный набор классов не всегда необходим. Практически значимо разделение сцены на проходное пространство и препятствия; более детальная схема может включать дорожное или напольное покрытие, границы проходимой области, элементы зданий и инфраструктуры, растительность, транспортные средства и людей. После сегментации классы объединяются в функциональные группы, поэтому модель предоставляет не управляющую команду, а структурированное описание локальной сцены.



Особое значение имеет точность границ. Ошибка на границе между проходимой поверхностью и препятствием опаснее ошибки внутри крупной однородной области. Поэтому важны устойчивость к освещению, сохранение небольших объектов и корректная пиксельная разметка. Значимость плотного семантического описания для машинного восприятия демонстрируют специализированные наборы данных, включая Cityscapes [7].

Основные архитектурные подходы

Современная семантическая сегментация получила развитие с появлением полностью свёрточных сетей. Fully Convolutional Network заменила полносвязные слои свёрточными операциями и позволила выполнять сквозное плотное предсказание для изображения произвольного размера [1]. В дальнейшем широкое распространение получили энкодер-декодерные схемы.

Энкодер уменьшает пространственное разрешение и формирует признаки высокого уровня, а декодер восстанавливает карту классов. В SegNet при повышении разрешения используются индексы операций максимального объединения [2]. В U-Net связи между соответствующими уровнями энкодера и декодера передают локальные признаки и помогают уточнять границы сегментов [4].

Многомасштабный контекст используется в DeepLabv3+, где разреженные свёртки сочетаются с декодером для уточнения границ [3]. Трансформерные модели, например SegFormer, формируют иерархические признаки и объединяют их лёгким декодером [6]. Для бортовых систем важны архитектуры реального времени: BiSeNet разделяет сохранение пространственных деталей и извлечение контекста, обеспечивая компромисс между качеством и скоростью [5].

Формирование свободного пространства и навигационных признаков

Для использования сегментации в навигации многоклассовую маску преобразуют в представление свободного пространства. Пиксели проходимых классов объединяются в бинарную маску M_{free} , а препятствия исключаются. Затем могут удаляться малые изолированные компоненты и выбираться связная область, примыкающая к нижней части кадра, то есть к ближайшей наблюдаемой роботом части проходимой поверхности.

Простейший навигационный признак определяется по положению свободной области. Для горизонтального сечения вычисляются левая и правая границы, а затем центр x_c . Нормированное боковое отклонение можно представить как $e = (x_c - W/2) / (W/2)$. Его знак показывает направление смещения центра проходимой области относительно оптического центра камеры. Для большей устойчивости анализируют несколько сечений или среднюю линию всей маски.

При наличии калибровки камеры и модели опорной поверхности координаты изображения могут быть преобразованы в приближённые метрические координаты. Однако такая оценка зависит от положения камеры и геометрии поверхности. На широкой открытой площадке или в просторном помещении свободная область может занимать почти весь кадр, поэтому её центр уже не задаёт требуемый маршрут. На развилке несколько направлений также могут быть одинаково проходимыми.

Следовательно, семантическая сегментация отвечает прежде всего за локальное восприятие. Она показывает, где находится свободное пространство и какие области заняты препятствиями, но сама по себе не определяет глобальный маршрут и требуемое направление на развилках. Для этого требуется отдельная информация о положении робота и цели движения.

Устойчивость и требования реального времени

Ошибки сегментации вызываются изменением освещения, тенями, бликами, осадками, размытием при движении, необычными ракурсами и недостаточно представленными в обучающих данных объектами. Для навигации особенно опасны ложные пиксели свободного пространства внутри препятствия. Поэтому полезны контроль связности маски, удаление малых областей и временная проверка стабильности результата.



В видеопотоке соседние кадры обычно изменяются плавно, поэтому резкий скачок границы свободной зоны может рассматриваться как нестабильное предсказание. Сглаживание навигационных признаков уменьшает влияние единичных ошибок, но не должно скрывать появление нового препятствия.

На мобильной платформе критична предсказуемая задержка. Разрешение входного изображения, размер модели и структура энкодера определяют вычислительную стоимость. Архитектуры реального времени специально балансируют пространственную детализацию и контекст [5], тогда как более тяжёлые модели требуют больших вычислительных ресурсов.

В полноценной системе сегментация должна дополнять локализацию. Локализация определяет положение робота и направление маршрута, но не гарантирует актуальность карты относительно временных препятствий. Сегментация, напротив, анализирует текущую видимую сцену. Их совместное применение разделяет задачи глобальной привязки и локального безопасного движения.

Заключение

Семантическая сегментация обеспечивает плотный анализ окружающего пространства мобильного робота в помещениях и на открытых территориях и позволяет выделять проходимые области и препятствия на уровне пикселей. Современные методы включают полностью свёрточные, энкодер-декодерные, многомасштабные и трансформерные архитектуры; для бортового применения используются облегчённые решения, ориентированные на реальное время.

Результат сегментации наиболее полезен как источник локальных навигационных признаков: границ свободной зоны, её центра и относительного бокового отклонения. При этом выбор маршрута в широких пространствах, на открытых площадках и на развилках требует отдельной локализации и планирования. Поэтому семантический анализ целесообразно использовать как компонент комбинированной навигационной системы

Список литературы:

1. Long J., Shelhamer E., Darrell T. Fully Convolutional Networks for Semantic Segmentation // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015. P. 3431–3440.
2. Badrinarayanan V., Kendall A., Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2017. Vol. 39, No. 12. P. 2481–2495.
3. Chen L.-C., Zhu Y., Papandreou G., Schroff F., Adam H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation // Proceedings of the European Conference on Computer Vision (ECCV). 2018. P. 801–818.
4. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation // Medical Image Computing and Computer-Assisted Intervention (MICCAI). 2015. Vol. 9351. P. 234–241.
5. Yu C., Wang J., Peng C., Gao C., Yu G., Sang N. BiSeNet: Bilateral Segmentation Network for Real-time Semantic Segmentation // Proceedings of the European Conference on Computer Vision (ECCV). 2018. P. 325–341.
6. Xie E., Wang W., Yu Z., Anandkumar A., Alvarez J. M., Luo P. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers // Advances in Neural Information Processing Systems. 2021. Vol. 34. P. 12077–12090.
7. Cordts M., Omran M., Ramos S., Rehfeld T., Enzweiler M., Benenson R., Franke U., Roth S., Schiele B. The Cityscapes Dataset for Semantic Urban Scene Understanding // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. P. 3213–3223

