

**Оценка эффективности детектора дорожной обстановки при использовании
мультимодального смещения данных сенсоров машинного зрения**

Кучков А.В. (ИТМО)

**Научный руководитель – кандидат технических наук, доцент Кашевник А.М.
(ИТМО)**

Введение. Мультисенсорные системы компьютерного зрения представляют собой устройства, воспринимающие и обрабатывающие данные с двух и более сенсоров. Совместная обработка данных нескольких сенсоров компьютерного зрения представляет собой нетривиальную задачу, требующую индивидуального подхода к сенсорам различных модальностей. Для обеспечения полноты восприятия окружающей сцены был разработан метод на основе использования механизма внимания, позволяющий смешивать данные сенсоров различных модальностей и извлекать признаки на объединенных данных. Метод был апробирован на общедоступном наборе данных KITTI [1] и было исследовано как механизм внимания влияет на метрики работы детектора. Изучение способов представления смешанных данных особенно актуально при построении мультисенсорных систем компьютерного зрения, требующих генерации предсказаний в широком диапазоне мете- и дорожных условий.

Основная часть. В исследовании проводится сравнительный анализ нескольких методов раннего смещения данных лидара и камеры. А именно, исследование содержит следующие этапы:

- 1) Исследуются методы совместного представления лидарных и растровых данных в виде объединенной карты признаков с учетом различных масштабов;
- 2) Исследуется влияние механизмов внимания на качество смещения данных в моделях основанных на PointNet++ [2].

Этап 1 направлен на исследование методов слияния на разных масштабах для наиболее полного извлечения признаков в объединенных данных. Использование различных карт признаков на разных масштабах необходимо для учета разнообразия размеров предметов на сцене.

Этап 2 ставит перед собой задачу исследования следующих механизмов внимания на процесс извлечения признаков и точность работы детектора:

- 1) Пространственное внимание;
- 2) Поканальное внимание;
- 3) Внимание к веткам обработки.

В качестве референсной архитектуры в работе используется модель двухстадийного детектора PointRCNN [3], предобученного на наборе данных KITTI.

Выводы: В настоящей работе проведен анализ методов раннего смещения данных со сравнительным анализом нескольких механизмов внимания и учетом вариативности размеров детектируемых объектов на сцене. Было определено, что использование комбинации методов поканального и пространственного внимания дает прирост метрик качества детектора.

Список использованных источников:

1. Geiger A. et al. Vision meets robotics: The kitti dataset //The International Journal of Robotics Research. – 2013. – Т. 32. – №. 11. – С. 1231-1237.
2. Qi C. R. et al. Pointnet++: Deep hierarchical feature learning on point sets in a metric space //Advances in neural information processing systems. – 2017. – Т. 30.

3. Shi S., Wang X., Li H. Pointrcnn: 3d object proposal generation and detection from point cloud //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. – 2019. – C. 770-779.