

РАЗРАБОТКА СИСТЕМЫ КОМПЬЮТЕРНОГО ЗРЕНИЯ ДЛЯ СЕЛЬСКОХОЗЯЙСТВЕННОЙ ТЕХНИКИ

К. П. Лошаков

Томский политехнический университет
kpl2@tpru.ru

Введение

Рост сложности задач современных робототехнических систем послужил развитию алгоритмов и методов обнаружения, отслеживания, классификации и сегментации окружающего пространства. Эта тенденция охватила все сферы жизни человека, от промышленности до сельского хозяйства и медицины. Робототехнические комплексы могут обрабатывать окружающее пространство в двухмерном и трехмерном виде. В качестве систем 3D могут выступать стереокамеры - строящие карту глубины, RGB-D-датчиков, а также LIDAR-датчики, строящие окружающее пространство в виде облака точек.

LIDAR-датчики используются компанией “Waymo” для создания беспилотного автомобиля. Преимуществами является работоспособность в любое время суток, высокая точность. Вместе с тем эффективность снижается при работе в плохих погодных условиях (дождь, туман, снег), а также при загрязнении камеры.

Одним из представителей RGB-D-датчика является датчик Kinect, который был использован Rosell-Polo et al в проектах точного сельского хозяйства [1]. Эксперимент проводился с датчиками в разных сезонных условиях. Было обнаружено, что Kinect неэффективен при обнаружении малых или сложных целей при дневном свете. Дальность обзора 4 м. Обработка до 30 кадров в секунду.

Камера ZED является одним из представителей стереокамер. Камера позволяет обрабатывать до 100 кадров в секунду и строить карту глубины до 20 м. Однако в задачах распознавания объектов существенное значение имеют шумы [2].

В системах 2D зрения используются спектральные, промышленные или цифровые камеры.

Применение спектральных камер является также актуальным по решению ряда задач, несмотря на высокую стоимость. Их эффективность обуславливается анализом изображения в определенном диапазоне длин волн электромагнитного спектра, причем границы этого диапазона выходят за границы диапазона видимого света, что приносит дополнительную информацию, которая помогает решить множество различных проблем, в том числе мониторинг состояния растительности [3].

Для получения информации могут использоваться также промышленные или обычные цифровые камеры. Для дальнейшей обработки информации используются два подхода. Один из вариантов таких подходов – создание алгоритмов компьютерного зрения, недостатком которого является сложность разработки алгоритмов. Второй вариант основывается на использовании нейронных сетей.

Недостатком является приобретение дополнительных вычислительных мощностей.

Рисунок 1 отображает сравнительную характеристику каждого из устройств, используемого в компьютерном зрении.

Устройство	Стоимость	Скорость	Работа вне помещения	Стоимость доп. оборудования
LIDAR-датчик	Худший выбор	Средний выбор	Лучший выбор	Средний выбор
Сtereoкамера	Средний выбор	Средний выбор	Средний выбор	Средний выбор
RGB-D-датчик	Средний выбор	Средний выбор	Худший выбор	Средний выбор
Цифровая камера	Средний выбор	Средний выбор	Средний выбор	Средний выбор
Спектральная камера	Худший выбор	Худший выбор	Средний выбор	Средний выбор

Худший выбор Средний выбор Лучший выбор

Рис. 1. Сравнение устройств в системах компьютерного зрения

Исходя из данной таблицы, можно сделать вывод, что LIDAR-датчик имеет высокую стоимость и невысокую скорость обработки изображений, что ведет к необходимости дополнительных вычислительных мощностей. RGB-D-датчик рассчитан на малую дальность обзора и не работоспособен в уличных условиях. Спектральные камеры же имеют возможность решить задачу, но достаточно дорогие и обладают низкой скоростью обработки изображений. Именно поэтому для задачи детектирования и сегментирования необработанных участков сельскохозяйственных культур в поле применена нейронная сеть с архитектурой mask R-CNN на основе информации с камеры [4].

Экспериментальная часть

В ходе работы использована одна из предобученных моделей репозитория Tensorflow Object Detection API [5] – mask_rcnn_inception_v2_coco_2018_01_28 – для детектирования и сегментирования необработанных комбайном участков поля. Для составления обучающей выборки было получено 65 изображений с разрешением 640x480 пикселей в цветовой модели RGB в разных окружающих средах. К каждому элементу выборки прикреплялись два файла. Один из них содержал данные о картинке в том числе и об координатах ограничивающей объект рамки, полученной при выделении объекта. Другой содержал бинарную маску изображения, в котором каждому пикселю принадлежащему объекту присваивалось значение 1, другим 0. Данные из этих файлов перевели в стандартный формат для библиотеки, с которой работаем, то есть в формат TFRecords. После

чего настроен конфигурационный файл для обучения модели с нужным количеством обучающихся изображений за эпоху.

Визуализация данных обучения нейронной сети представлена на рисунке 2, график на котором отображает зависимость общей ошибки в ходе обучения, складывающейся из ошибок в определении класса, координат рамки и вероятности присвоения данного объекта к классу от номера эпохи.

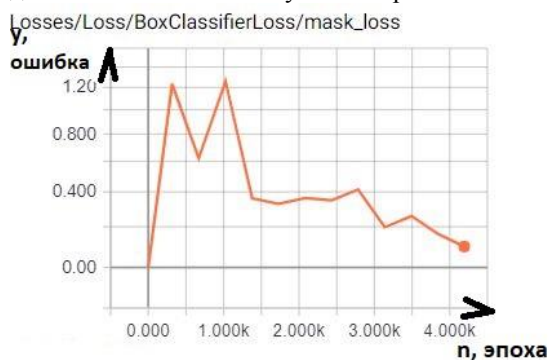


Рис. 2. Общая ошибка модели

Можно заметить, что в первой четверти процесса обучения ошибка увеличивается в связи с настройкой весов нейронной сети и такой продолжительный монотонный рост можно объяснить тем, что во время обучения за одну эпоху обрабатывалось одно изображение в связи с нехваткой технических ресурсов для более продуктивного обучения модели. В последующее время наблюдается монотонное уменьшение ошибки. Ошибка при выполнении задачи сегментации меньше 1.000 на протяжении всей времени обучения с тысяча двусотой эпохи, что может означать эффективную реализацию нейронной сети для задачи сегментации. На рисунке 3 представлены результаты обработки изображений из тестовой выборки обученной нейронной сетью.

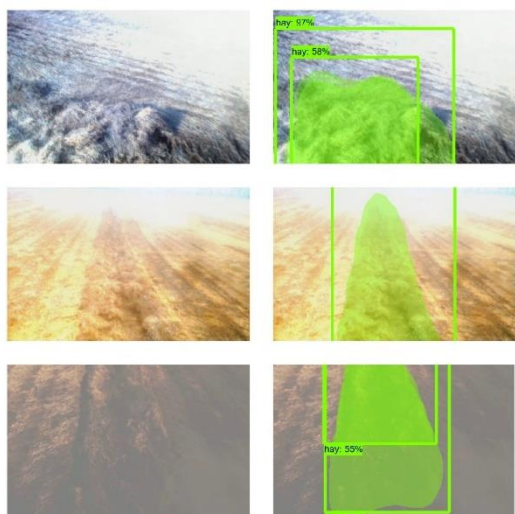


Рис. 3. Результаты тестирования модели

Тестовая выборка, часть изображений которых представлена на рисунке 3, состояла не только из объектов, располагающихся в центральной части камеры. Точность результатов сегментации на тестовой выборке, состоящей из 10 изображений, составила 90.00%.

Заключение

Один из главных факторов, влияющих на эффективность нейронной сети является выделение каждого пикселя исследуемого объекта на изображении при подготовке данных для обучающей выборки. Факт того, что изображения на обучающей выборке содержали исследуемый объект по центру изображения, не повлиял на способность обученной нейронной сети определять необработанные участки поля, располагающиеся по краям камеры (рисунк 3).

Список использованных источников

1. J. R. Rosell-Polo, F. A. Cheein, E. Gregorio, D. Andujar, L. Puigdomenech, J. Masip, and A. Escol ` a, “Advances in structured light ` sensors applications in precision agriculture and livestock farming,” *Advances in Agronomy*, vol. 133, pp. 71–112, 2015.
2. R. Xiang, H. Jiang, and Y. Ying, “Recognition of clustered tomatoes based on binocular stereo vision,” *Computers and Electronics in Agriculture*, vol. 106, pp. 75–90, 2014.
3. Геоскан. Сельское хозяйство. – URL: <https://www.geoscan.aero/ru/application/agriculture> (дата обращения 10.10.2018).
4. Mask R-CNN. – URL: <https://arxiv.org/abs/1703.06870> (дата обращения 10.10.2018).
5. Tensorflow Object Detection API. – URL: https://github.com/tensorflow/models/tree/master/research/object_detection (дата обращения 8.10.2018).