



В состав штатной модели входит список анализаторов поведения телеметрического индекса. Под анализатором подразумевается структура, содержащая совокупность правил и ограничений значений конкретного телеметрического индекса в течение определённого промежутка времени. Например, известно, что в течение определённой полётной операции температура в отсеке монотонно возрастает от 17 до 23 градусов. Тогда сформированный анализатор «Монотонно возрастает» будет содержать границы изменяемых значений, аппроксимирующую функцию изменения и возможную погрешность, в рамках которой значения будут трактоваться как допустимые.

Результатом работы модуля будет являться список анализаторов для распознанных индексов, каждый из которых будет содержать атрибуты, описывающие характер изменений их значений со временем, а также привязку к элементу плана в программе полёта.

В дальнейшем эти анализаторы будут извлекаться из базы в процессе мониторинга текущих значений индекса на данной полётной операции.

Заключение

Представлен вариант формирования модели штатного поведения узлов космического корабля для системы анализа и прогнозирования состояния бортовых систем. Он позволяет на основе имеющихся данных о штатном полёте сформировать представление о правилах изменения индексов с целью последующего мониторинга для прогнозирования тенденций к возникновению нештатных ситуаций на борту космического аппарата.

Литература

1. Юрыгина Ю.С., Лахин О.И., Донсков А.В., Мишурова Н.В., Сысоев Д.В., Скорюпина Е.Г. Особенности реализации интеллектуальной системы анализа и прогнозирования состояния космических аппаратов // Материалы VI Международной научно-технической конференции «Открытые семантические технологии проектирования интеллектуальных систем» Open Semantic Technologies for Intelligent Systems (OSTIS-2016), 18 – 20 февраля 2016 г. – Минск: БГУИР, 2016. – С. 487-492.

С.А. Разлацкий, П.Ю. Якимов

ИСПОЛЬЗОВАНИЕ GPU ДЛЯ РАСПОЗНАВАНИЯ ТРЕХМЕРНЫХ ОБЪЕКТОВ ПРИ ПОМОЩИ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ

(Самарский университет)

Аннотация

В настоящей статье описаны современные подходы к распознаванию трехмерных объектов в реальном времени, в частности с использованием трехмерных сверточных нейронных сетей (CNN). Для оценки качества распознавания и быстродействия рассмотренных методов были использованы различные



типы трехмерных данных. Используется реализация CNN из открытой библиотеки VoxNet. Экспериментальные исследования показали эффективность взятой реализации. Для обучения и функционирования используемого метода использован GPU.

Введение

Задача надежного детектирования объектов в трехмерной сцене стала актуальной с появлением и развитием методов получения трехмерных цифровых изображений. Между тем, сенсоры, такие как LiDAR и RGBD камеры, развивались и становились все более распространенным решением для многих автономных робототехнических средств. Настоящая работа является продолжением исследований методов распознавания трехмерных объектов в облаке точек – специализированном формате представления трехмерных данных.

Методы, рассмотренные авторами в [1] и [2], эффективно распознают объекты, и их быстродействия достаточно для функционирования в реальном времени, но главным недостатком описанных методов является требование идентичности искомого объекта заранее заданному эталону. То есть, для распознавания какого-либо зарегистрированного объекта необходимо, чтобы в системе уже имелась трехмерная модель, ассоциированная с определенным классом. Такой подход хорошо подходит для распознавания объектов на модельных данных, но в реальности регистрируемый объект часто отличается от его представления в эталонных моделях.

В последнее время все более широкое распространение получили методы на основе технологии глубинного обучения. В частности, задача детектирования и распознавания трехмерных объектов успешно решается при помощи сверточной нейронной сети (Convolutional Neural Network или CNN), адаптированной для работы с трехмерными объектами [3]. Настоящее исследование посвящено изучению работы библиотеки VoxNet [4], которая содержит эффективную реализацию трехмерной сверточной нейронной сети (3D CNN) для распознавания трехмерных объектов в реальном времени.

Обзор существующих методов

Классические методы распознавания трехмерных объектов

Одним из примеров классического подхода к обнаружению трехмерного объекта является трехмерный метод Хафа [5]. Данный метод стал популярен в применении для двумерных изображений, где в качестве областей интереса использует в основном контуры объектов. При работе в трехмерном формате метод выделяет в трехмерном изображении особые точки для сокращения вычислительных издержек Процедура голосования происходит в аккумуляторном пространстве с учетом только выделенных особых точек. В результате получаются локальные максимумы в тех областях, где потенциально может находиться искомый объект. Еще одна проблема - это вероятность различной ориентации сцены и искомого объекта. Она решается при помощи введения специальных векторов, обеспечивающих инвариантность к вращению и повороту [1]. Немного более проработанным является метод геометрической связности [2]. Основным отличием от трехмерного метода Хафа является иной алгоритм поиска



особых точек, которые объединяются в так называемые особые области и переводятся в формат, описываемый специальным индексом форм. Полученные области в виде значений индексов записываются в двумерные гистограммы, где и происходит процедура голосования по всем локальным окрестностям, содержащимся в тестовом объекте.

Существует большое количество работ по распознаванию объектов в трехмерном облаке точек, получаемых при помощи LiDAR и стереокамеры, использующих комбинацию различных индивидуальных признаков и дескрипторов с классификацией методами машинного обучения [6], [7], [8]. Также широко распространены методы семантической сегментации, где вместо отдельные классификаторы используются структурированные классификаторы. В отличие от указанных выше подходов, исследуемая в настоящей работе архитектура учится извлекать особенности и классифицировать объекты из «сырых» 3D данных.

2.5D CNN

Вдохновившись успешным применением сверточных нейронных сетей для решения задач распознавания на двумерных изображениях, некоторые авторы расширили их использование для стерео данных. Такие подходы обрабатывают канал с «глубиной» как дополнительный канал, наряду с обычными каналами R, G, B. Однако, при этом не в полной мере используется геометрическая информация в трехмерных данных, что затрудняет интеграцию между зрительными точками.

Для LiDAR данных предложены признаки [9], локально полученные на данных с представлением 2.5D, а некоторые работы исследуют данный подход в сочетании с разновидностью так называемого обучения без учителя [10]. В работе [11] предложена кодировка, которая эффективно использует информацию о глубине, но подход все равно двумерно-ориентированный. Получается более точное представление об окружающей среде.

VoxNet. Архитектура 3D CNN

Исходные данные для алгоритма, реализованного в библиотеке VoxNet, представляют собой сегмент облака точек, который может быть получен различными методами сегментации или при помощи алгоритма «скользящего трехмерного окна». Сегмент, как правило, определяется пересечением облака точек с ограничивающим параллелепипедом и может включать в себя фоновые шумы. Задача заключается в определении принадлежности объекта в сегменте к определенному классу. Система решения данной задачи состоит из двух компонентов: «объемной сетки», которая представляет оценку пространственного наполнения, и 3D CNN, которая классифицирует объекты, непосредственно используя объемную сетку. Опишем компоненты более подробно.

На рисунке 1 представлена архитектура VoxNet.

Сетка пространственного наполнения отображает представление состояние среды как 3D решетки случайных величин (каждая соответствует одному вокселу) и формирует вероятностную оценку их размещения, как функцию от входных данных датчика и априорного знания. Также сетка наполнения позво-



ляет эффективно оценить свободные, занятые и неизвестные пространства из измерений дальности, даже для измерений, поступающих с разных точек зрения и в различные моменты времени.

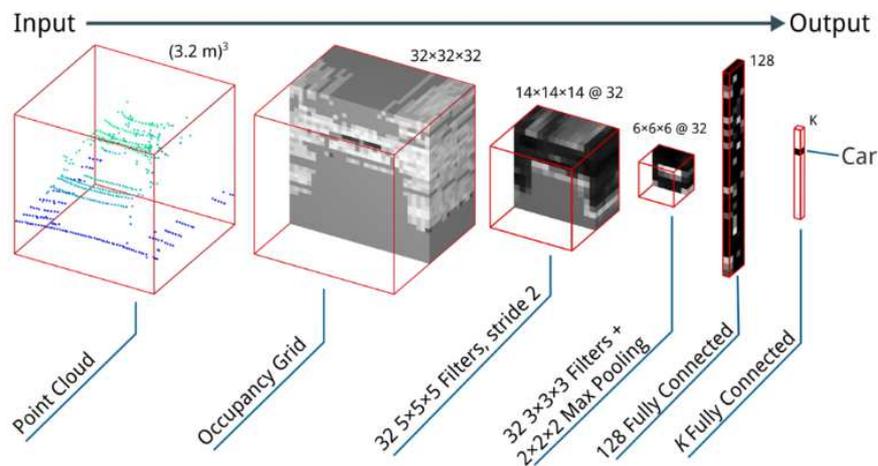


Рис. 1. Первый этап – характеризует облако точек, второй этап - перевод облака точек в пространственную сетку наполнения, третий этап – определение карты признаков при помощи фильтрации блока вокселей с определенным шагом, четвёртый этап – объединение вокселей определенной площадью, пятый этап – вывод слоя с вероятностью нахождения признаков.

Для CNNs есть три основных причины, делающих сверточные нейронные сети привлекательным для решения поставленной задачи. Во-первых, они могут использовать пространственную структуру рассматриваемой задачи. В частности, они могут изучить локальные пространственные фильтры, полезные для задачи классификации. В данном случае, ожидается, что фильтры на уровне входного сигнала будут кодировать пространственные структуры как плоскости и углы при различных ориентациях. Во-вторых, укладывая несколько слоев сети можно построить иерархию из более сложных функций, представляющих большие области пространства, в конечном итоге это приводит к полной классификации для поступающей сетки пространственного наполнения. И наконец, такие операции могут быть выполнены эффективно на графическом аппаратном средстве. В данном исследовании рассматривается CNN, состоящая из нескольких слоев.

Входной слой принимает фиксированного размеры сетки $I*J*K$ вокселей. В выбранной версии используются вариант, где каждый параметр равен 32. В зависимости от выбранной модели пространственного наполнения сетки, значение каждой ячейки находится в диапазоне $(-1,1)$, основываясь на уравнениях или (1), или (2), или (3). Хотя эта работа рассматривает только скалярный вклад, реализация может принимать дополнительные значения в ячейку, такие как LiDAR значения интенсивности или RGB информация от камер.

Сверточные слои $C(f,d,s)$ при помощи фильтра f , размерности исследуемого блока вокселей d и шагом s , создают карту признаков f . Создание карты признаков происходит путем перемножения входных данных с фильтром. При переходе к сверточным слоям следующего уровня, начинается поиск более



сложных признаков. Для повышения эффективности они опираются на результаты карт признаков предыдущего уровня, получая четырехмерный объем данных $d*d*d*f'$, где d является пространственным размером, а f' – карты признаков предыдущего уровня. Свертка также может быть применена при большом пространственном шаге s , что может повысить скорость обучения сети, но также может понизить точность распознавания. Полученный результат пропускается через выпрямленный нелинейный блок (ReLU) с параметром 0,1.

Группировка слоев $P(m)$ подразумевает под собой нелинейное уплотнение карты признаков с шагом m по каждому трехмерному параметру (I,J,K). То есть вместо блока вокселей размера $m*m*m$ мы получаем максимальное значение вокселя, представленного в матрице.

Полносвязанные слои FC(n) имеют n выходных нейронов. Выход каждого нейрона изучает линейная комбинация всех выходов предыдущего слоя, пропуская их через нелинейность. Для этого используется ReLU, которая преобразовывает результаты выходного слоя для получения вероятности нахождения объекта, где число выходов соответствует числу классов.

Стоит понимать, что при первых попытках обнаружения искомым объектов сеть будет мало эффективна, основная причина – это принятия веса карты каждой карты признаков случайным образом. CNN начинает превосходить все ранее исследуемые методы лишь после обучения. Это обучение основывается на стохастическом градиентном спуске. Именно данный метод начинает регулировать вес карт признаков, приводя с каждым тренировочным набором данных данный параметр к эталонному значению.

Эксперименты

Чтобы оценить эффективность VoxNet, используются данные трех разных типов: LiDAR облака точек, RGBD облака точек и CAD модели. Все тренировочные данные были взяты из открытых источников, а сама библиотека также находится в открытом доступе [12]. Рисунок 2 показывает примеры каждого типа.



Рис. 2. LiDAR данные, RGBD данные, CAD данные.

В ходе экспериментов были использованы данные, дополненные вращением. Было рассмотрено 4 различных случая для такого дополнения (аугментации) вращением: в зависимости от того, применяется ли оно или нет при обучении (как дополнение) и тестировании (как голосование) для LiDAR объектов и ModelNet40. Для случаев, в которых голосование не выполняется при тестировании, случайная ориентация применяется на тестовых примерах, и берется среднее время за 4 выполнения. Для случаев, в которых аугментация не выполняется, существует два варианта. В ModelNet40 для обучения выбирается объект в начальной позе. Для LiDAR данных эта информация не доступна, и ис-



пользуется не модифицируемая ориентация для данных. Таблица 1 показывает результаты экспериментов. Они показывают, что тренировочное время аугментации наиболее важно.

Таблица 1. Эффект от аугментации вращением и голосования

Аугментация при обучении	Голосование при тестировании	Сиднейские объекты	ModelNet40
Да	Да	0,72	0,83
Да	Нет	0,71	0,82
Нет	Да	0,69	0,69
Нет	Нет	0,69	0,61

Помимо этого, было произведено сравнение VoxNet с другим методом ShapeNet, предложенным Wu [13] для задачи классификации данных ModelNet40 и ModelNet10. В ShapeNet также используется объемная сверточная архитектура с применением аугментацией вращения для обучения. Однако количество параметров в этой архитектуре более 12 миллионов, как в VoxNet их менее миллиона. В таблице 2 показаны полученные результаты.

Таблица 2. Сравнение с ShapeNet

Набор данных	ShapeNet	VoxNet
ModelNet10	0,84	0,92
ModelNet40	0,77	0,83

Для обучения 3D сверточной нейронной сети и непосредственного выполнения был использован графический процессор Nvidia Tesla K40. Наиболее ресурсозатратная конфигурация VoxNet – мультимасштабность и дополнение поворотами и голосованием – выполняется за 6 мс в случае индивидуальной классификации, и по 1 мс на объект, если обрабатывать сразу 32 объекта. Такая разница объяснима накладными расходами на пересылку данных между GPU и CPU.

Заключение

В настоящей работе был представлен анализ библиотеки VoxNet, реализующей архитектуру трехмерных сверточных нейронных сетей для эффективного и точного обнаружения и распознавания объектов различных типов в облаках точек, а также изучено влияние использования различных конфигураций на производительность системы. Кроме того, было произведено сравнение с аналогичными системы на одинаковых наборах данных, которое показало превосходство исследуемой системы VoxNet. Полученные результаты показали, что данная реализация 3D CNN успешно справилась с поставленными задачи и действует эффективнее других подобных систем.



В дальнейшем планируется применить полученные результаты для исследования распознавания объектов в трехмерных сценах, получаемых при помощи стереокамеры, в реальном времени.

Благодарности

Исследования проводились при поддержке фонда РФФИ (проект 16-37-00362).

Литература

[1] Разлацкий, С.А. Применение метода Хафа для детектирования объектов в трехмерной сцене / С.А. Разлацкий, П.Ю. Якимов // Информационные технологии и нанотехнологии (ИТНТ-2015). -2015.

[2] Разлацкий, С.А. Применение метода геометрической связанности для детектирования объектов в трехмерном облаке точек / С.А. Разлацкий, П.Ю. Якимов // Информационные технологии и нанотехнологии (ИТНТ-2016). -2016.

[3] Nielsen, M. Neural networks and deep learning / M.Nielsen. – 2014.

[4] Maturana, D. VoxNet: A 3D Convolutional Neural Network for Real-Time Object Recognition / D. Maturana, S. Scherer // IEEE/RSJ International Conference on Intelligent Robots and Systems. – 2015.

[5] Goshin, Ye.V. Segmentation of stereo images with the use of the 3D Hough transform / Goshin Ye.V., Loshkareva G.E. // CEUR Workshop Proceedings. – Vol. 1638 – 2016. – P.340-347.

[6] Tombari F., Object recognition in 3D scenes with occlusions and clutter by Hough voting / L. Di Stefano, F. Tombari // Fourth Pacific-Rim Symposium on Image and Video Technology. – 2010. – P. 2-4

[7] Frome, A. Recognizing objects in range data using regional point descriptors / A. Frome, D. Huber, and R. Kolluri // ECCV. - 2004. - Vol. 1. - P. 1–14.

[8] Chen, H. 3D free-form object recognition in range images using local surface patches / H. Chen, B. Bhanu // Pattern Recognition Letters. – vol. 28(10). – 2007. – P. 1252-1262.

[9] Quadros, A. An occlusion-aware feature for range images / A. Quadros, J. Underwood, and B. Douillard // ICRA, May 14-18 2012.

[10] Quadros, A. Unsupervised feature learning for classification of outdoor 3d scans / M. De Deuge, A. Quadros, C. Hung, and B. Douillard // ACRA, 2013.

[11] Gupta, S. Learning rich features from RGB-D images for object detection and segmentation / S. Gupta, R. Girshick, P. Arbelaez, and J. Malik // ECCV, 2014.

[12] GitHub [Электронный ресурс] - URL: <https://github.com/dimatura/voxnet> (дата обращения 12.02.2017)

[13] Wu, Z. 3d shapenets: A deep representation for volumetric shape modeling / Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang and J. Xiao // Proceedings of 28th IEEE Conference on Computer Vision and Pattern Recognition. – 2015.