Автоматическое создание и разметка RGB-D изображений для обучения систем машинного зрения

А.Д. Жданов [1], Д.Д. Жданов [1], Е.Д. Хилик [1]

[1] *Университет ИТМО, Кронверкский проспект, д.49, литер А., г. Санкт-Петербург, 197101, Россия*

Аннотация

В связи с активным развитием технологий искусственного интеллекта, машинного зрения и глубокого обучения, а также появлением RGB-D камер, позволяющих получить объемное изображение сцены, все большее внимание уделяется различным задачам обработки трехмерных данных. Одной из таких задач является задача сегментации облака точек, которая находит применение в различных областях, от робототехники до архитектуры и решается методами машинного зрения. Для обучения систем машинного зрения требуется создание и аннотирование датасетов, которое занимает значительную часть времени проектирования и разработки. В данной работе предлагается автоматизировать процесс создания датасета с помощью компьютерных систем интерпретатора сценариев и реалистичного рендеринга, которые могут существенно сократить время, необходимое для создания датасета. Приводится пример создания датасета, обучения нейронной сети на этом наборе данных и использование сети, обученной на этом наборе данных, для классификации объектов на изображении сцены.

Ключевые слова

Машинное зрение, нейронные сети, датасет, глубокое обучение, реалистичный рендеринг.

Automatic Creation and Annotation of RGB-D Images for Training Machine Vision Systems

A.D. Zhdanov [1], D.D. Zhdanov [1], E.D. Khilik [1]

[1] *ITMO University, Kronverksky Pr. 49, bldg. A, St. Petersburg, 197101, Russia*

Abstract

Due to the active development of artificial intelligence technologies, machine vision, and deep learning, as well as the emergence of RGB-D cameras that allow you to get a three-dimensional image of the scene, more and more attention is paid to various tasks of processing three-dimensional data. One of these problems is the problem of point cloud segmentation, which is used in various fields, from robotics to architecture, and is solved by machine vision methods. The training of machine vision systems requires the creation and annotation of datasets, which takes up a significant part of the design and development time. In this paper, it is proposed to automate the process of creating a dataset using a scripting interpreter and realistic rendering computer systems, which can significantly reduce the time required to create a dataset. An example of creating a dataset, training a neural network on this dataset, and using a network trained on this dataset to classify objects in a scene image is given.

Keywords

Machine vision, neural networks, dataset, deep learning, realistic rendering.

ГрафиКон 2023: 33-я Международная конференция по компьютерной графике и машинному зрению, 19-21 сентября 2023 г., Институт проблем управления им. В.А. Трапезникова Российской академии наук, г. Москва, Россия
EMAIL: andrew.gtx@gmail.com (А.Д. Жданов); ddzhdanov@mail.ru (Д.Д. Жданов); khilik.egor@gmail.com (Е.Д. Хилик)
ORCID: 0000-0002-2569-1982 (А.Д. Жданов); 0000-0001-7346-8155 (Д.Д. Жданов); 0009-0004-7058-4432 (Е.Д. Хилик)

© 2023 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

1. Введение

В связи с активным развитием технологий искусственного интеллекта, машинного зрения и глубокого обучения, а также появлением RGB-D камер, позволяющих получить объемное изображение сцены, все большее внимание уделяется различным задачам обработки трехмерных данных. Одной из таких задач является задача сегментации облака точек, которая находит применение в различных областях, от робототехники до архитектуры. Однако, для эффективного решения этой задачи необходимо иметь доступ к качественной базе данных RGB-D изображений сцены [1]. Создание и разметка таких баз данных вручную требует много времени. Кроме того, человечески фактор в может привести к ошибкам при разметке данных. Поэтому задача упрощения и ускорения процесса создания и автоматической разметки датасетов для обучения нейронных сетей [2, 3] является актуальной.

Новизна предложенного подхода заключается в том что, система проводит симуляцию движения камеры в виртуальной среде и автоматически создает размеченные изображения. Автоматический подход к построению и разметке баз данных RGB-D изображений позволит исследователям быстро и эффективно создавать датасеты, что поможет улучшить качество результатов обучения систем машинного зрения и ускорить развитие робототехники.

2. Методы автоматической сегментации облака точек

Сегментация изображения – это процесс разделения цифрового изображения на отдельные группы или сегменты на основе их геометрических, структурных и семантических характеристик. Таким образом, обработка и анализ больших изображений становится проще и легче. При работе с трехмерными данными, например, полученными с использованием современных RGB-D камер, возникает аналогичная задача сегментации. На рисунке 1 показан пример сегментации облака точек, полученного при сканировании стула и определении его составных частей, таких как спинка, сиденье и ножки.

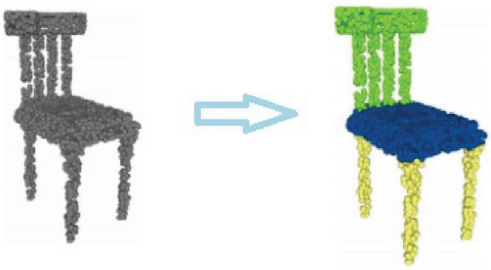

Рисунок 1 – Сегментация облака точек объекта

Существует несколько методов сегментации облака точек [4]. Их можно разделить на следующие основные группы: методы на основе нахождения границ объектов, методы на основе обнаружения поверхностей, методы на основе графов и методы на основе машинного обучения.

Методы сегментации облака точек на основе границ являются одним из популярных и изученных подходов, который основан на обнаружении и использовании локальных геометрических особенностей облака точек для определения границ объектов и их сегментации. Методы, основанные на обнаружении областей, используют информацию об окрестности для объединения близлежащих точек, имеющих сходные свойства, для получения изолированных областей и, следовательно, для поиска различий между областями сцены. Данные группы методов дают высокое качество сегментации изображения и обладают высокой скоростью работы, что позволяет обрабатывать большие объемы данных за короткие промежутки времени. Однако, они чувствителен к шуму и к неравномерной плотности точек. Таким образом, в случае если на изображении присутствует много шумовых элементов или если точки на изображении расположены неравномерно, то результаты сегментации могут быть недостаточно точными.

Методы сегментации изображений на основе теории графов используют графическое представление облака точек чтобы эффективно разделить изображение на несколько сегментов.

Каждая точка в облаке точек представляет собой узел в графе, а ребра соединяют близлежащие точки. Данные методы также чувствительны к качеству входных данных и могут иметь ограниченную производительность в реальном времени. Это связано с тем, что построение и сегментация графа может быть достаточно трудоемким процессом. Кроме того, такие методы могут быть менее точными, если облако точек содержит шумы или пропущенные точки.

Последняя группа методов основывается на методах машинного обучения [5, 6] и может автоматически сегментировать облако точек на основе различных признаков, таких как геометрические характеристики, цвет, текстуры и т.д. Эти признаки позволяют определять различные объекты и фон на изображении. Данные методы могут быть более точными и эффективными, чем рассмотренные ранее, а также учитывать более сложные признаки изображения и обрабатывать больший объем данных. Однако, для обучения моделей требуется предварительно создать обучающий набор данных.

Все из перечисленных методов не могут гарантировать полную корректность автоматической сегментации изображений и требуют ручной проверки и коррекции при формировании датасета. Кроме того, требуется ручное аннотирование сегментированных данных, что является времезатратным процессом и приводит существенному ограничению размеров датасетов.

3. Автоматическая сегментация и разметка данных

В рамках данной работы предлагается использовать программную сегментацию и синтезированные изображения для автоматического создания и разметки датасета. Преимуществом использования синтезированных изображений является дополнительная информация, которую можно получить от системы рендеринга, такая как индексы объектов, с которыми произошло первое пересечение луча для заданной точки изображения и после ряда зеркальных отражений на трассе данного луча. Полученная информация служит основой для сегментации и разметки изображения. Кроме того, для каждой точки изображения можно вычислить расстояние до объектов сцены и сформировать карту глубин изображения сцены, что позволяет создать облако точек объектов. При изменении положения камеры можно создать ряд облаков точек, соответствующих наблюдению сцены с различных направлений без необходимости проведения трехмерного сканирования реальных объектов. Кроме того, модель камеры может быть построена с учетом реальной оптики, используемой при наблюдении, что добавляет «физичность» в процесс виртуального 3D сканирования. На рисунке 2 показан пример изображения и дополнительной информации, достаточной для его автоматической сегментации. На правом изображении можно заметить что каждый объект сцены имеет свой цвет, благодаря этому легко выделить и идентифицировать координаты различных объектов в сцене. Стоит отдельно отметить, что для использования данного метода требуется физическая корректность и реалистичность синтезированных изображений. В противном случае датасет, созданный с использованием полученных изображений, не может быть использован для обучения систем машинного зрения, работающих в реальном мире.

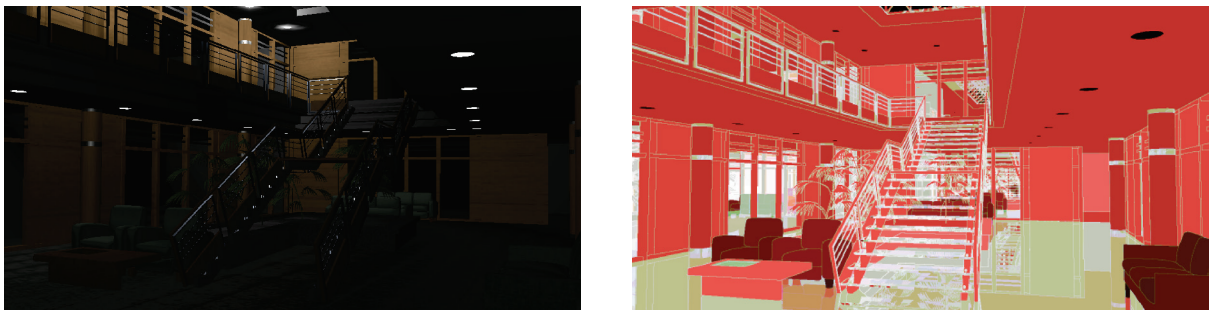

Рисунок 2 – Автоматическая сегментация сцены системой реалистичного рендеринга

Задача по сбору данных и формированию сцены состоит из нескольких этапов, показанных на рисунке 3.

Первый этап – сбор данных. На этом этапе осуществляется симуляция передвижения наблюдателя по сцене в виртуальной среде. Камера записывает последовательность шагов,

сохраняя изображения сегментированных участков сцены и карту глубины. Каждый шаг включает изменение положения и угла камеры, чтобы получить различные ракурсы объектов сцены. Наблюдатель перемещается по сцене полу-случайно, выполняя движения на фиксированный шаг в направлении различных объектов сцены и поворачивает, в случае столкновения с препятствием. Кроме того, на каждом шаге происходит анализ «корректности» положения камеры, а именно, если камера смотрит в упор на объекты сцены или сегментируемые объекты отсутствуют или они расположены глубже допустимой дистанции, то кадр исключается из обработки. После выполнения определенного количества шагов движение камеры прекращается и запускается обработка накопленных данных.

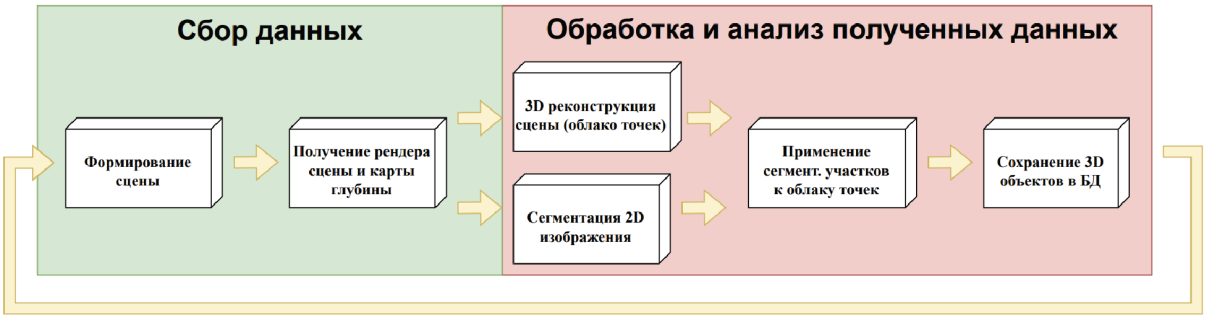

Рисунок 3 – Блок-схема алгоритма автоматического синтеза, сегментации и разметки RGB-D изображений сцены и создания датасета

Второй этап – обработка и анализ полученных данных. На этом этапе производится анализ карты глубин и сегментация изображений и облаков точек. Сначала, производится реконструкция трехмерной сцены из карты глубин в облако точек. Далее, производится сегментация 2D изображений, где каждому пикселю присваивается значение, соответствующее объекту на сцены, полученному от системы фотореалистичного рендеринга в процессе синтеза изображения. Аналогично, сегментированные участки применяются к трехмерному облаку точек, чтобы каждый объект имел соответствующую аннотацию.

Для корректного аннотирования изображений и облаков точек необходимо провести предварительную обработку моделей трехмерных сцен, используемых для автоматического создания датасета. Стоит отметить, что большинство качественных трехмерных сцен имеют иерархическую структуру с корректными и понятными названиями объектов. Поэтому достаточно создать таблицу соответствия объектов сцены с требуемыми классами объектов. Процесс автоматической разметки изображения и облака точек из результатов реалистичного моделирования показан на рисунке 4

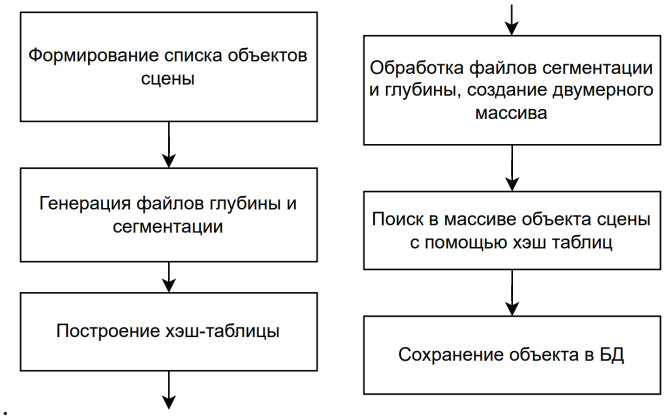

Рисунок 4 – Автоматическая разметка результатов реалистичного моделирования для изображения (слева) и облака точек (справа)

Полученные 3D облака точек сохраняются в базе данных для дальнейшего использования. Файлы сохраняются в формате PLY (Polygon File Format), который позволяет хранить точечные облака данных. Это позволяет сохранить трехмерные модели объектов и использовать их в дальнейших приложениях, анализе сцен и в машинном обучении.

4. Результаты прототипирования

Целью прототипирования была проверка возможности обучить нейронную сеть используя датасет, созданный автоматически на основе изображений, сформированных системой реалистичного рендеринга. Требовалось оценить корректность созданного датасета, корректность его аннотирования и результат работы нейронной сети, обученной с использованием созданного автоматически датасета.

Для прототипирования метода автоматического создания и разметки RGB-D изображений были использованы четыре сцены, представленные на рисунке 5. Каждая из этих сцен представляет различные интерьерные среды с различными объектами. Такой подход позволил создать датасет и обучить на нем модель нейронной сети для классификации деталей интерьера.

Описанный в предыдущей главе метод был реализован на языке Python в системе реалистичного рендеринга Lumicept. Данная система была выбрана по причине использования языка Pyhton в качестве интерпретаторе сценариев, посредством которого предоставляется полный доступ к данным сцены, синтезу изображений, формированию карты глубин и сегментации синтезированного изображения. Кроме того, использования языка Python позволило полностью реализовать весь процесс создания датасета в виде единого программного кода без необходимости использовать внешние программные средства.

Рисунок 5 – Изображения сцен, использованных для создания и разметки RGB-D изображений

4.1. Создание датасета

На рисунках 6 и 7 показаны результаты автоматического создания облаков точек для объектов «камин» и «кресло», которые попали в кадр во время перемещения виртуального

наблюдателя по сцене. Поскольку в данном случае создавался датасет облаков точек, то изображение сцены было сформировано с низким качеством исключительно для визуальной оценки сегментированного объекта. Облако точек при этом создавалось на основе данных из карты глубин и автоматической сегментации сцены системой реалистичного рендеринга.

При анализе рисунка 6 можно заметить, что сегментация объектов изображения визуализируется однотонной структурой, что означает присутствие одного объекта. Такая ситуация возникла из-за того, что камера в данном случае оказалась направлена исключительно на камин, который представляет собой отдельный объект сцены. Это привело к тому, что в итоговом изображении большая часть пикселей принадлежит только к одному объекту (камину), что делает изображение однородным и требует ручной проверки на пригодность для дальнейшего использования в датасете. С другой стороны, количество таких изображений в сформированном датасете незначительно, поэтому они могут быть либо обработаны вручную или удалены автоматически. В рамках данного прототипирования они были удалены. На рисунке 6 применяется градация цветов от синего к красному. Красные цвета отображают значения близкие к нулевой координате, в то время как синие цвета представляют значения, далекие от неё. Этот подход визуализации данных повышает наглядность и облегчает их понимание.

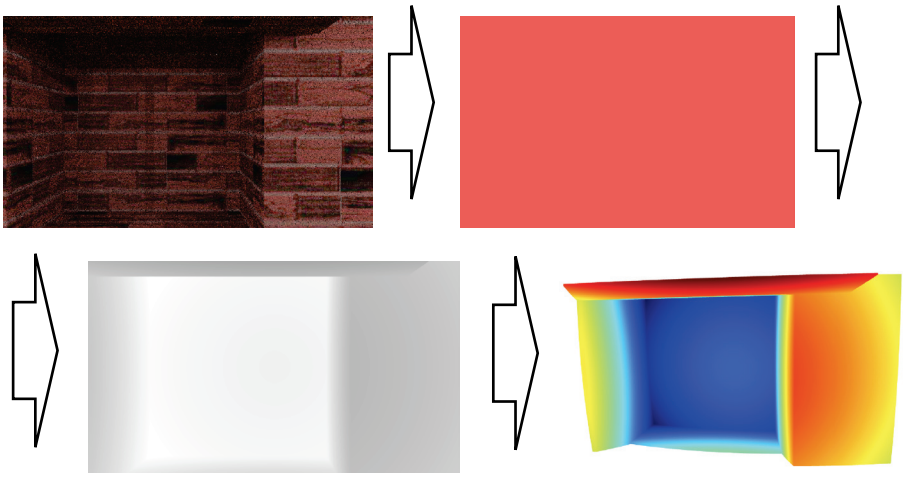

Рисунок 6 – Процесс создания и разметки сегментированного облака точек камина: синтезированное изображение, сегментация изображения, карта глубин, облако точек

Большая часть изображений имеет характер, схожий с объектами типа «кресло», изображенными на рисунке 7, где объекты четко различимы на фоне других объектов и могут быть использованы для автоматической разметки и создания датасета без дополнительных проверок.

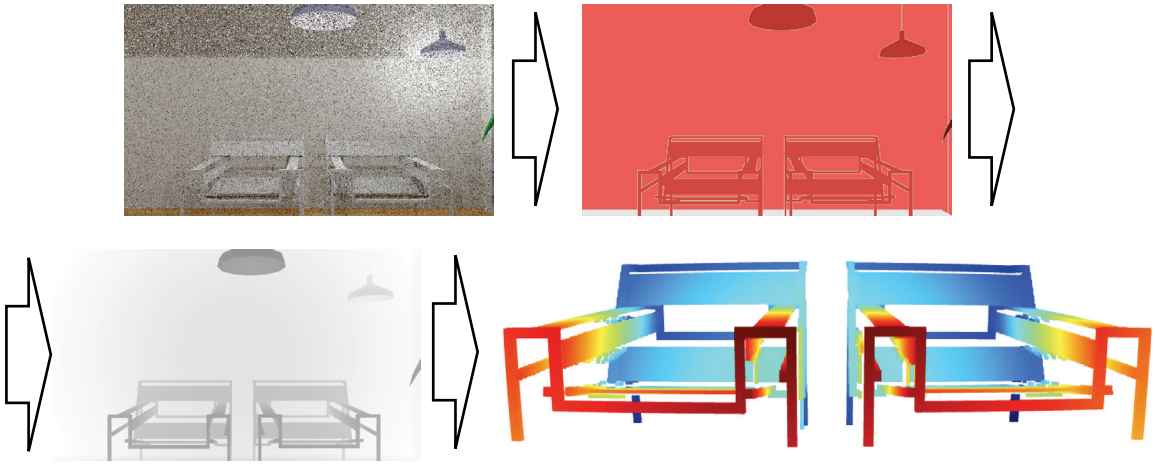

Рисунок 7 – Процесс создания и разметки сегментированного облака точек кресел: синтезированное изображение, сегментация изображения, карта глубин, два облака точек

4.2. Обучение нейронной сети

Для обучения тестовой модели классификации картинок была выбрана сверточная нейронная сеть (Convolutional Neural Network, CNN) [7], которая является одной из самых распространенных архитектур для анализа изображений и показывает хорошие результаты, что делает ее подходящим выбором для задачи тестирования автоматически созданного датасета. Сверточная нейронная сеть состоит из нескольких сверточных слоев, которые извлекают различные признаки из изображений, и пулинговых слоев, которые уменьшают размерность данных. Затем следуют полносвязные слои, которые служат для классификации или регрессии. Для обучения нейронной сети использовались библиотеки TensorFlow [8] и Keras [9]. Обучение проводилось на 211 изображениях, которые содержали информацию об объектах, классифицированных как «цветы», «столы», «стулья» и «лампы». Модель обучалась на протяжении 10 эпох. Для увеличения датасета использовались методы аугментации.

Далее, было проведено тестирование обученной модели. Для тестирования модели был использован отдельный набор данных, который не использовался в процессе обучения. Этот набор данных содержал изображения объектов из классов «цветы», «столы», «стулья» и «лампы». Результаты тестирования модели показали, что ее точность составляет 62.5%.

Результат тестирования показал, что использование автоматически созданного датасета позволяет успешно обучить нейронную сеть и решать задачи классификации. Для повышения качества классификации необходимо, во-первых, увеличить размер формируемого датасета, что не составляет сложности, и, во-вторых, использовать более сложную модель нейронной сети в процессе тестирования.

5. Заключение

В данной работе был представлен алгоритм автоматического создания датаетов для последующего обучения нейронных сетей, который позволяет автоматически собирать данные о глубине и сегментации сцены, имитируя движение камеры в виртуальной среде. Применение разработанного алгоритма должно повысить качество датасетов и, как следствие, качество обучения нейронных сетей в различных областях, таких как машинное зрение и глубокое обучение, виртуальная реальность, автоматический анализ сцен и системы автоматизации.

В дальнейшем планируется повысить качество модели с помощью оптимизации ее параметров и внедрения дополнительных признаков. Для дополнительного повышения эффективности использования и обучаемости нейронных сетей на основе автоматически созданного набора данных требуется увеличение базы данных сцен и применения методов аугментации данных.

6. Благодарности

Работа выполнена при финансовой поддержке Российского Научного Фонда, проект № 22-11-00145.

7. Список источников

[1] Lai, K. A Large-Scale Hierarchical Multi-View RGB-D Object Dataset / K. Lai, L. Bo, X. Ren, D. Fox // IEEE International Conference on Robotics and Automation, 2011.

[2] Chang, Angel & Funkhouser, Thomas & Guibas, Leonidas & Hanrahan, Pat & Huang, Qixing & Li, Zimo & Savarese, Silvio & Savva, Manolis & Song, Shuran & Su, Hao & Xiao, Jianxiong & Yi, Li & Yu, Fisher. (2015). ShapeNet: An Information-Rich 3D Model Repository

[3] Charles, R. & Su, Hao & Mo, Kaichun & Guibas, Leonidas. (2017). PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. 77-85. 10.1109/CVPR.2017.16.

[4] 23.3D ML. Часть 6: Обзор алгоритмов семантической сегментации облака точек [Электронный ресурс] //URL:https://habr.com/ru/companies/itmai/articles/534036/

[5] Vinodkumar, Prasoon Kumar, Dogus Karabulut, Egils Avots, Cagri Ozcinar, and Gholamreza Anbarjafari. 2023. "A Survey on Deep Learning Based Segmentation, Detection and Classification for 3D Point Clouds" Entropy 25, no. 4: 635
[6] Wang, Yuan & Shi, Tianyue & Yun, Peng & Tai, Lei & Liu, Ming. (2018). PointSeg: Real-Time Semantic Segmentation Based on 3D LiDAR Point Cloud.
[7] Maturana, Daniel & Scherer, Sebastian. (2015). VoxNet: A 3D Convolutional Neural Network for real-time object recognition. 922-928. 10.1109/IROS.2015.7353481.
[8] Abadi, M. (2016, September). TensorFlow: learning functions at scale. In Proceedings of the 21st ACM SIGPLAN International Conference on Functional Programming.
[9] Manaswi, N. K., & Manaswi, N. K. (2018). Understanding and working with Keras. Deep learning with applications using Python: Chatbots and face, object, and speech recognition with TensorFlow and Keras, 31-43.