

Отчет по научно-исследовательской работе

по теме:

«Сопровождение объектов на видеопоследовательности»

Научный руководитель: Павлов В. А.

Работу выполнил: Эхтибаров О. О.

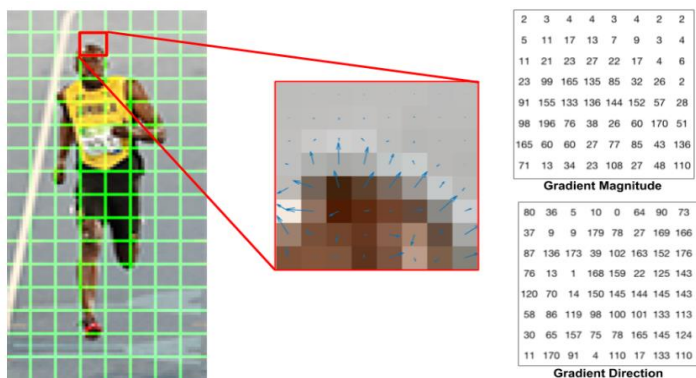
Санкт-Петербург

2021

Целью данной работы было написать программу, которая будет считать количество вошедших и вышедших из автобуса пассажиров. Соответственно работа была разделена на две части: первая – детектирование и трекинг людей, то есть определение движения в автобусе, символизирующего вход и выход из автобуса, вторая – инкрементирование счетчика вошедших и вышедших пассажиров.

Детектирование и трекинг

Существует немало алгоритмов детектирования объектов и их сопровождения. Одним из самых распространенных, опираясь на небольшую выборку статей на эту тему, которые я просмотрел, является алгоритм HOG-детектора (гистограмма ориентированных градиентов). В основе данного метода лежит предположение о том, что вид распределения градиентов изображения (элемента видеопоследовательности) позволяет достаточно точно определить наличие и форму присутствующих на нем объектов. При описании изображения мы разбиваем его на несколько ячеек, в которых вычисляем гистограммы направленных градиентов точек внутри ячейки, после чего эти гистограммы объединяются в одну и нормализуются по яркости.



Center : The RGB patch and gradients represented using arrows. Right : The gradients in the same patch represented as numbers

Нормализация блоков [\[править \]](#) [\[править код \]](#)

Далал и Триггс исследовали четыре метода нормализации блоков. Пусть v — ненормированный вектор, содержащий все гистограммы данного блока, $\|v\|_k$ — его k -норма при $k = 1, 2$ и e — некая малая константа (точное значение не так важно). Тогда нормировочный множитель можно получить одним из следующих способов:

$$\text{L2-норма: } f = \frac{v}{\sqrt{\|v\|_2^2 + e^2}}$$

L2-hys: L2-норма ограничивается сверху (значения v , большие 0,2, полагаются равными 0,2) и перенормируется, как в^[5]

$$\text{L1-норма: } f = \frac{v}{(\|v\|_1 + e)}$$

$$\text{корень из L1-нормы: } f = \sqrt{\frac{v}{(\|v\|_1 + e)}}$$

Далал и Триггс установили, что L1-норма дает менее надежные результаты, чем остальные три, которые работают приблизительно одинаково хорошо, однако все четыре метода значительно улучшают результаты по сравнению с ненормализованными.^[4]

Нормализация по яркости делается для предотвращения потери участка детектирования при изменении яркости. После чего применяется метод опорных векторов (SVM). В своей работе я не стал использовать этот метод, поскольку он показался мне немного устаревшим. Я стал искать более современные методы детектирования объектов (совсем недавние статьи с того же ICCV я не стал просматривать, поскольку мне показалось, что в них я практически ничего не пойму) и наткнулся на метод, который называется «Single Shot Multibox Detector» (в конце отчета есть все ссылки на используемые ресурсы и статьи). Суть этого метода следующая: на вход подается изображение 300x300x3 (квадрат 300x300 пикселей в RGB формате), после чего к изображению применяется некоторое количество сверточных слоев

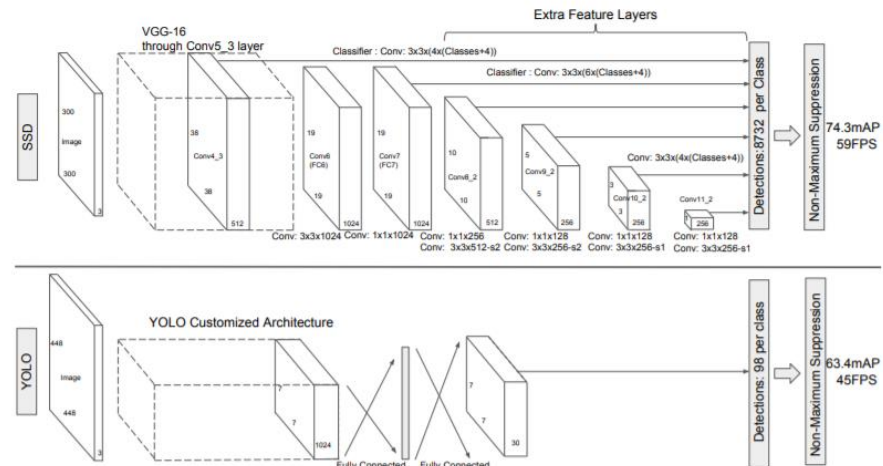


Fig. 2: A comparison between two single shot detection models: SSD and YOLO [5]. Our SSD model adds several feature layers to the end of a base network, which predict the offsets to default boxes of different scales and aspect ratios and their associated confidences. SSD with a 300×300 input size significantly outperforms its 448×448 YOLO counterpart in accuracy on VOC2007 test while also improving the speed.

Актив:
Чтобы а

На рисунке есть упоминание еще и об YOLO алгоритме, но затрагивать этот алгоритм я не буду. После детектирования объектов нужно было следить за их трекингом. Для этого я использовал алгоритм центр-трекинга: отмечал центры объектов в кадре и по мере изменения кадров центры тоже меняли свои координаты (вызвано изменением положения объекта в пространстве или тем, что объект просто вышел из кадра, а в кадре появился новый и т.д.). Основная проблема этого метода заключается в том, чтобы определить, является ли новый центр обновленной координатой того же объекта или же это уже новый объект и его центр принимает лишь первое значение (раньше объекта не было в кадре). Для этого надо вычислить Евклидово расстояние между старыми и новыми (обновленными) центрами. Дальше мы делаем предположение о том, что расстояние между центрами объекта в кадре n и в кадре $n + 1$ будет меньше, чем все остальные расстояния. Таким образом мы трекаем конкретный объект x в разных кадрах. А если в новом кадре центров больше, чем было в предыдущем, и расстояния от одного центра до старых центров больше, чем эти же расстояния от других новых центров, то этот центр является новым в конкретном кадре, что означает, что в кадре появился новый объект. Мы рассмотрели варианты изменения положения объекта в видеопоследовательности, появление нового объекта в видеопоследовательности, но что же делать, если существующий объект выходит из кадра. На этот счет тоже существует несколько способов определения этого момента, в данной работе я использовал следующий способ: мы будем deregистривать старый объект (центр), когда этот центр не может быть сопоставлен ни с одним центром на последующих k кадрах. Что касается определение направления движения объекта, то определяется оно следующим образом: мы просто считываем координаты в настоящем кадре и в предыдущем

кадре. Этих данных нам достаточно, чтобы определить направление движения объекта. На этом алгоритм детектирования объекта и его движения заканчивается.

Обновление счетчика

Что касается трекинга, то тут все попроще. Я просто изобразил на видеопоследовательности линию примерно на уровне двери автобуса, чтобы заранее не был заинкрементирован счетчик вошедших или вышедших из автобуса пассажиров. Таким образом, когда дверь открывается срабатывает трекинг объектов (людей), и если центр определенного объекта находился сначала по одну сторону от этой линии, а потом по другую, то инкрементируется счетчик вошедших (вышедших) людей.

Вывод

В данной работе была разработана мини-программа для подсчета вошедших и вышедших из автобуса людей. Для детектирования и трекинга объектов использовались SDD алгоритм и алгоритм центр-трекинга, а для счета людей метод виртуальной линии. Результаты получились неточными. Мне кажется, что это вызвано тем, что эти алгоритмы все же больше рассчитаны на менее динамичные видеопоследовательности с меньшим количеством объектов, которые нужно детектировать и трекать. В автобусе же много объектов, за которыми нужно следить, более того, эти объекты расположены очень близко относительно друг друга, а после открытия двери автобуса происходит быстрое движение объектов, из-за чего детектор и трекер не успевают переключаться между объектами. Возможно, если использовать еще более современные методы детектирования и трекинга объектов, результаты значительно улучшатся.

Источники

1. <https://arxiv.org/pdf/1512.02325.pdf>
2. <https://www.machinelearningmastery.ru/understanding-ssd-multibox-real-time-object-detection-in-deep-learning-495ef744fab/>
3. https://ru.wikipedia.org/wiki/Гистограмма_направленных_градиентов
4. google.com