Санкт-Петербургский государственный университет

Системное программирование

Обзор решений распознавания препятствий по видео с камеры кругового обзора

Ри Александра Ильинична, группа 23.Б10-мм

4 января 2025

Научный руководитель: старший преподаватель кафедры системного программирования, к. т. н., Литвинов Ю. В.
Консультант: инженер-программист, АО “Кама”, Осечкина М. С.

Санкт-Петербург
2025

Введение

- Для кругового обзора часто используются камеры типа “рыбий глаз”, которые создают радиальные искажения
- Существует множество решений распознавания объектов на неискаженных изображениях, но нет данных о их применимости к искаженным изображениям
- Реализаций существующих решений детекции на изображениях типа “рыбий глаз” нет в открытом доступе, доступно лишь описание метода
- Необходимо решение с работой в режиме реального времени с ограничением по вычислительным ресурсам для внедрения в системы помощи водителю (дальше ADAS)

Постановка задачи

Цель: анализ и сравнение методов распознавания, их точность обнаружения препятствий на искаженных изображениях, применимость в режиме реального времени с ограничением по вычислительным ресурсам

Задачи:

- Выполнить обзор существующих решений
- Протестировать методы на имеющихся наборах данных
- Проанализировать результаты тестирования

Критерии отбора материала

- Были рассмотрены первые несколько ссылок в поисковой системе Google, на сайтах researchgate.net и arxiv.org по ключевым словам: “fisheye”, “vehicle detection”, “pedestrian detection”
- После рассматривались модели, используемые в этих методах, а также модели, с которыми они сравнивались

Методика тестирования

Для оценки точности обнаружения объектов используется общепринятая метрика mAP (Mean Average Precision)

- Считается Presicion$=\frac{TP}{TP+FP}$ и Recall$=\frac{TP}{TP+FN}$
- Вычисляется AP для каждого класса, как площадь под графиком Recall-Presicion
- mAP — среднее AP по всем классам

Свёрточные нейронные сети (далее CNN)

- Свёрточные слои: операции свёртки — каждый фрагмент изображения поэлементно умножается на матрицу свёртки, результат суммируется и записывается в соответствующую позицию
- Скалярный результат каждой свёртки подаётся на нелинейную функцию активации, которая обнуляет отрицательные значения
- Происходит нелинейное уплотнение карты признаков: группа пикселей (обычно 2×2) уплотняется до одного пикселя

Семейство архитектур R-CNN (Region-based Convolutional Neural Networks)

Двухэтапные модели:

- Генерация предложений RoI (Regions of Interest)
- Классификация и уточнение координат обнаруженного объекта

Faster R-CNN:

- Изображение пропускается через CNN (VGG-16)
- С помощью дополнительных свёрточных слоев генерирует RoI
- Полученные признаки для каждого RoI подаются различным предобученным классификаторам, которые определяют только один класс объектов

Считаются одними из самых точных моделей, но потребляют много памяти и имеют достаточно низкую скорость работы (Faster R-CNN: 7 FPS)

SSD (Single Shot MultiBox Detector)

- Пропускает изображение через CNN (VGG-16)
- Модель независимо обрабатывает все карты признаков, полученные со всех свёрточных слоев, с помощью дополнительных свёрточных слоев
- Ядра дополнительных свёрточных слоёв выдают два типа результата:
  - Вероятность принадлежности области к классу
  - Необходимое смещение координат рамки по умолчанию, для лучшего охвата объекта

Семейство моделей YOLO (You Only Look Once)

YOLOv4:

- Пропускает изображение через CNN (CSPDarknet53), получим карту признаков $S \times S$
- Каждая ячейка отвечает за обнаружение объекта, центр которого находится в её границах, а также уже имеет рамки по умолчанию в разных масштабах
- Анализируя и комбинируя значения, полученные с разных слоев, определяется вероятность принадлежности к классу и уточнение координат рамок

Экспериментальное исследование

Постановка эксперимента:

- Какой mAP имеют модели, обученные распознавать объекты на неискаженных изображениях, на наборе изображений, полученных с камер типа “рыбий глаз”
- Насколько хуже модели определяют объекты на искаженных изображениях, чем на неискаженных

Условия эксперимента:

- Набор данных без искажений — PASCAL VOC 2007[1]
- Набор данных с искажениями типа “рыбий глаз”— WoodScape[2]
- Оборудование: GPU NVIDIA Tesla T4 (на Google Colab)
- Тестируемые модели: Faster R-CNN, SSD, YOLOv4

[1] http://host.robots.ox.ac.uk/pascal/VOC/voc2007/
[2] https://github.com/valeoai/WoodScape

Результаты экспериментального исследования

- Исследование показало, что модели ожидаемо не сильно отличаются по значению mAP, из-за достаточно точного определения объектов ближе к центру изображения, где искажения меньше

Модель	**PASCAL VOC 2007**		**WoodScape**	
	mAP	FPS	mAP	FPS
Faster R-CNN	73,2%	7,0	31,0%	7,0
SSD	77,2%	46,0	35,1%	8,3
YOLOv4	82,0%	45,0	39,2%	9,2

Table: Результаты работы моделей

- Модели показывают существенно более низкую точность на изображениях типа “рыбий глаз”, чем на неискаженных изображениях, что говорит о невозможности использования моделей без модификаций для определения объектов в системах ADAS

Результаты

- Выполнен обзор существующих решений распознавания объектов на неискаженных изображениях
- Выбраны модели распознавания объектов на изображениях и протестированы на наборе изображений с искажениями типа “рыбий глаз”[3]
- Было выяснено, что выбранные модели не подходят для распознавания препятствий на изображениях с радиальными искажениями в режиме реального времени с ограничением по вычислительным ресурсам

[3] Реализация в GitHub репозитории URL:`https://github.com/xImoZA/detection_models_on_fisheye` (дата обращения: 4.01.2025)