

Автоматическое детектирование и распознавание объектов на изображениях*

*Дудоров Н. А., Томинин Я. Д., Томинин В. Д., Ерлыгин Л. О.,
Демидова Ю. О., Разумов И. О.*

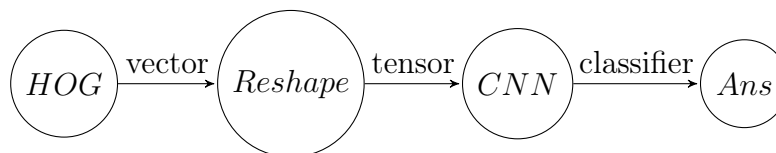
МФТИ(ГУ)

Автоматическое детектирование и распознавание объектов на изображениях является одной из основных задач компьютерного зрения. Характерные свойства изображения объекта - это набор признаков, приближенно описывающий интересующий объект. Эти признаки делятся на два класса: локальные и интегральные. Одной из проблем компьютерного зрения является то, что интегральные признаки, характеризующие изображения объекта в целом, не устойчивы к изменению структуры объекта и сложным условиям освещения. На сегодняшний день мы имеем универсальный метод получения признаков изображения - CNN. Однако нейронные сети требуют больших вычислительных мощностей, таким образом, наша задача состоит в создании признакового описания объекта на основе уже известных алгоритмов (таких как HOG, LBP) в условиях, трудных для работы вышеперечисленных алгоритмов.

Введение

HOG неплохо зарекомендовал себя в качестве дескриптора по выделению интегральных признаков в изображении. Однако, по нашему мнению, этот дескриптор мог бы работать лучше, если модернизировать его поведение. Проблема заключается в том, что в оригинальном алгоритме выделенные в виде векторов признаки конкатенируются в вектор размерности 1. Тем самым мы теряем информацию относительно расположения выделенных признаков. Наша задача - улучшить работу HOG за счет удаления указанного недостатка.

Остановимся на двух методах выделения признаков - CNN и HOG. На момент написания статьи эти способы считаются противоположными подходами к обработке изображений. Возможное решение задачи - объединение этих двух методов. Наша цель - оценить качество работы алгоритма, модель которого представлена ниже. Оценка качества работы алгоритмов будет производиться с помощью построения DET-кривых для оригинального и модифицированного алгоритмов. Для обучения и тестирования полученных в работе алгоритмов будут использованы базы данных «CITYSCAPES» и «INRIA Person».



Постановка задачи Пусть $X = \{x\}^l$ - набор изображений из обучающей выборки. $Y = \{y\}^l : y \in \{1, -1\}$ - ответы на обучающей выборке (два класса изображений: содержащие людей и не содержащие). Обозначим за $H : x \rightarrow z$ - дескриптор HOG, а $F : z \rightarrow y$ - модель, которую мы применяем к выходу алгоритма HOG (в нашем случае этой моделью будет CNN).

Наша задача состоит в том, чтобы добиться лучшей DET-кривой $MR(FP)$, где :

$$MR = \frac{FN}{TP + FN}$$

- FN - ложно-положительные ответы
- TP - верные положительные ответы
- FP - ложно-положительные ответы