

Выделение неизвестных объектов на произвольном фоне

А.В. Алексеев, Д.Р. Лещинер

Московский физико-технический институт (государственный университет)

В последние годы наблюдается существенный рост интереса к усовершенствованным методам выделения и распознавания объектов на изображениях ([1], [2]), классификации изображений с помощью многослойных нейронных сетей (deep learning), сверточных сетей (convolutional networks) в особенности ([3], [4], [5]). Как правило, применяемые методы работают с выборкой заранее классифицированных тем или иным способом объектов, и строят распознающие правила на ее основе. Предложены также алгоритмы, позволяющие решать задачу выделения объектов на основе изображений, помеченных заданными метками, притом, что само выделение объектов на этих изображениях не входит в обучающие данные [6]. Упомянутые выше технологии основаны на построении в той или иной форме описаний распознаваемых объектов, и выделении объектов с использованием этих описаний. Применение подобных методов к задаче поиска разнородных, трудно классифицируемых объектов на разнообразном, варьирующемся фоне представляет немалые трудности. Пример такой задачи – поиск мусора на поверхности земли (траве, тротуарах, и т.п.). Если бы мы имели алгоритм, выделяющий нужные объекты на изображениях с достаточной детальностью (то есть, удовлетворительно отличающий пиксели, принадлежащие объекту, от пикселей, принадлежащих фону), то дальнейшее применение методов классификации к заранее выделенным объектам было бы существенно упрощено. При этом как полнота, так и точность нахождения нужных объектов не критичны для построения обучающей выборки. В дальнейшем, получаемые таким образом классификации могут помочь и в сегментации изображения, и в выделении объектов.

В настоящей работе предложен достаточно простой алгоритм выделения «заметно отличающихся от фона объектов среднего размера» на основе кластеризации пикселей, обучения и сегментации на основе получающихся кластеров. Алгоритм не требует размеченной выборки, но может далее применяться для построения обучающих выборок объектов как с разметкой, так и без разметки. Работоспособность алгоритма была проверена на задаче выделения бутылок на траве, ожидается применение к выделению и опознанию и других объектов. Ниже следует описание алгоритма (включая и эмпирические параметры).

Исходное пространство признаков для кластеризации пикселей 5-мерно: координаты точки (x, y) (измеренные в числе пикселей) и координаты цвета (H, S, V) (hue, saturation, value – их исходные значения из диапазона $[0,1]$ масштабированы умножением на коэффициенты $4 \cdot 10^8$, 10^5 , 10^4 соответственно), с евклидовой метрикой. Первым делом мы разбиваем точки на 3 кластера потоковой (поступившая точка относится к ближайшему центроиду, центроид пересчитывается, и точка забывается) рандомизированной (точки поступают в случайном порядке) версией метода k-means. Затем из 3-х кластеров выбираем наиболее компактный (с наименьшим средним расстоянием между точками) – в нем мы и будем искать объекты.

Из этого кластера выбираем сравнительно компактные множества точек (соответствующие отдельным объектам) следующим образом: выбираем случайную точку кластера, вычисляем массив квадратов (евклидовых) расстояний в пространстве (x, y) от остальных точек до нее, сортируем его по возрастанию. Ищем в отсортированном массиве первое место, где разность между 2 соседними элементами более 30. Все расстояния до этого места соответствуют одному объекту. Объект удаляем из кластера, и повторяем процедуру выделения. К выделенным объектам применяем вариант метода КЧП [7] для удаления «помех». А именно, если называть пиксели объекта «черными», а фона «белыми», то перекрашиваем в противоположный цвет короткие одноцветные отрезки в одномерных (сначала в горизонтальных, а затем в вертикальных) срезах. Прodelываем это трижды: перекрашиваем сперва «черные» отрезки короче 4 пикселей, затем «белые» короче 15, затем снова «черные» короче 10. Оставшаяся совокупность «черных» есть наш объект.

Далее, находимые объекты зачастую оказывались меньше, чем хотелось бы (но никогда не больше). Для расширения границ объекта применялись два различных метода. В первом случае, строим массив квадратов расстояний (в исходном 5-мерном пространстве) от всех точек вне объекта до ближайшей (в смысле кратчайшего пути по решетке пикселей) точки границы объекта. Сортируем массив по возрастанию и выбираем из него точки объекта – вплоть до первого места, где разность между 2 соседними элементами более 10^{10} . Во втором способе (показавшем лучшие результаты) использован следующий алгоритм: вписываем найденный объект в минимальный по площади прямоугольник. Строим 2 подобных ему прямоугольника с тем же центром и ориентацией – с вдвое большей диагональю (**x2**), и с диагональю, большей в полтора раза (**x1.5**). Пространство между ними считаем выборкой пикселей фона. Конструируем обучающую выборку из пикселей фона и пикселей объекта с признаками для каждого пикселя: его цвет, расстояние (в двух направлениях) до сторон прямоугольника (**x2**), количество пикселей «похожего» цвета (сумма модулей разностей по R, G, B – меньше 100) в квадрате 7x7 вокруг данного, перепад "яркости" (R+G+B) в квадрате 7x7. На выборке обучаем классификатор, отличающий наш объект от фона, с помощью случайного леса решающих деревьев (sklearn.ensemble.RandomForestClassifier, параметры по умолчанию, кроме n_estimators=50). Полученным классификатором добавляем точки объекта внутри прямоугольника (**x1.5**). Затем с заново выделенным объектом всю процедуру (начиная с вписывания в минимальный прямоугольник) можно повторить. Процесс повторения сходится быстро, и дает осмысленное представление границ объекта.

Литература

1. *Stewart R., Andriluka M.* End-to-end people detection in crowded scenes // arXiv:1506.04878.
2. *Ba J.L., Mnih V., Kavukcuoglu K.* Multiple Object Recognition with Visual Attention // International Conference on Learning Representations (ICLR). 2015.
3. *Krizhevsky A., Sutskever I., Hinton G.E.* ImageNet Classification with Deep Convolutional Neural Networks // Conference on Neural Information Processing Systems (NIPS). 2012.
4. *Ren S., He K., Girshick R., Sun J.* Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks // NIPS. 2015.
5. *Long J., Shelhamer E., Darrell T.* Fully Convolutional Networks for Semantic Segmentation // Conference on Computer Vision and Pattern Recognition (CVPR). 2015.

6. *Oquab M., Bottou L., Laptev I., Sivic J.* Is object localization for free? Weakly-supervised learning with convolutional neural networks // CVPR. 2015.
7. *Губерман Ш.А.* Выделение объектов (Системный подход) // Неформальный анализ данных в геологии и геофизике. - М.: Недра, 1987.