

УДК 004

ОБЗОР АЛГОРИМТОВ СЕМАНТИЧЕСКОЙ СЕГМЕНТАЦИИ

Григорьев Д.С.

Научный руководитель: Спицын В.Г., д.т.н., профессор

Национальный Исследовательский Томский политехнический университет,
634050, Россия, г. Томск, пр. Ленина, 30
E-mail: trygx@tpu.ru

The survey describes different techniques for semantic segmentation.

Key words: *Semantic segmentation, sift-descriptors, histogram of oriented gradients, convolution neural network.*

Ключевые слова: *Семантическая сегментация, SIFT-дескриптор, гистограмма направленных градиентов, сверточная нейронная сеть.*

Введение

Семантическая сегментация изображений – это задача разделения частей изображения на подгруппы пикселей, принадлежащих к соответствующим объектам, с его классификацией. В отличие от задач локализации (рис. 1) и детектирования объектов на изображении, задача семантической сегментации является более сложной. Это обусловлено необходимостью не только верно определить принадлежность объекта на изображении к определенному классу, но и максимально четко определить границы и структуру объекта на изображении (рис. 2).

В данной работе рассматриваются «традиционные» алгоритмы применяющиеся для решения задач сегментации/семантической сегментации изображений.

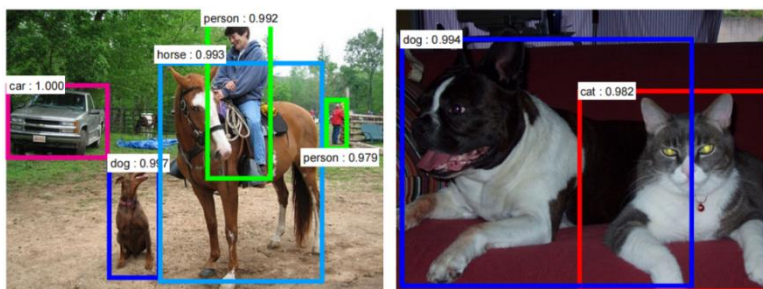


Рис. 1. Пример локализации объекта [1]



Рис. 2. Семантическая сегментация изображения [2]

Традиционные методы

Сегментация изображений при помощи «традиционных» методов отличается тем, что в их основе их работы лежит информация о пространственных признаках объекта, а также отсутствием нейросетевой составляющей. Относительная простота реализации и теоретической базы делают их широко распространенными.

SIFT (Scale Invariant Feature Transform) Этот дескриптор представляет собой локальную гистограмму направлений градиентов изображения. Он строится следующим образом. Окрестность характерной точки делится на четыре квадратных сектора. В каждом пикселе внутри каждого сектора вычисляется градиент изображения, его направление и модуль. Затем модули градиентов умножаются на вес, экспоненциально убывающий с удалением от точки интереса. Смысл применения веса заключается в том, чтобы избежать резких изменений значения дескриптора при небольших изменениях положения окна, а также в том, чтобы градиенты, удаленные от центра дескриптора, вносили меньший вклад в его значение, поскольку градиенты на периферии окрестности точки интереса наименее устойчивы при геометрических преобразованиях изображения. По каждому сектору собирается гистограмма направлений градиентов, причем каждое вхождение взвешивается модулем градиента. Дескриптор SIFT представляет собой вектор, полученный из значений всех элементов гистограмм направлений, и состоит из 128 компонент. Дескриптор нормируется, чтобы повысить его устойчивость к изменениям яркости.

Дескриптор SIFT стал де-факто стандартом в компьютерном зрении. На идее использования локальных гистограмм градиентов яркости изображения основывается большинство современных дескрипторов [3].

HOG (Histogram of oriented gradients) – дескрипторы особых точек, которые используются в компьютерном зрении и обработке изображений с целью распознавания объектов. Данная техника основана на подсчете количества направлений градиента в локальных областях изображения. Этот метод похож на гистограммы направления края, дескрипторы SIFT и контексты формы, но отличается тем, что вычисляется на плотной сетке равномерно распределенных ячеек и использует нормализацию перекрывающегося локального контраста для увеличения точности.

Основной идеей алгоритма является допущение, что внешний вид и форма объекта на участке изображения могут быть описаны распределением градиентов интенсивности или направлением краев. Реализация этих дескрипторов может быть произведена путём разделения изображения на маленькие связанные области, именуемые ячейками, и расчетом для каждой ячейки гистограммы направлений градиентов или направлений краев для пикселей, находящихся внутри ячейки. Комбинация этих гистограмм и является дескриптором. Для увеличения точности локальные гистограммы подвергаются нормализации по контрасту. С этой целью вычисляется мера интенсивности на большем фрагменте изображения, который называется блоком, и полученное значение используется для нормализации. Нормализованные дескрипторы обладают лучшей инвариантностью по отношению к освещению [4].

SVM – Метод опорных векторов (support vector machine) – набор схожих алгоритмов обучения с учителем, использующихся для задач классификации и регрессионного анализа. Принадлежит к семейству линейных классификаторов. Основная идея метода – перевод исходных векторов в пространство более высокой размерности и поиск разделяющей гиперплоскости с максимальным зазором в этом пространстве. Две параллельных гиперплоскости строятся по обеим сторонам гиперплоскости, разделяющей наши классы. Разделяющей гиперплоскостью будет гиперплоскость, максимизирующая расстояние до двух параллельных гиперплоскостей. Алгоритм работает в предположении, что чем больше разница или рассто-

яние между этими параллельными гиперплоскостями, тем меньше будет средняя ошибка классификатора [5].

Нейросетевые алгоритмы

С увеличением производительности вычислительных машин все более широкое распространение получают алгоритмы, основанные на использовании глубоких нейронных сетей. Большинство популярных моделей нейронных сетей, используемых для сегментации изображений, использует в своей архитектуре сверточные слои. Это позволяет существенно сократить количество параметров сети, а также достичь относительной устойчивости к переносу, масштабированию и незначительным искажениям на изображении.

Заключение

На сегодняшний день существует достаточно большое количество алгоритмов, решающих задачи семантической сегментации, однако все более широкое распространение получают алгоритмы основанные на применении нейросетей. Удачными решениями также могут являться и комбинации методов, например CNN/FCN (Convolution/Fully convolution network)+CRF(Conditional random field) что подтверждается результатами различных контекстов.

Список литературы

1. Ren S., He K., Girshick R., Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. [Электронный ресурс] URL: <https://arxiv.org/abs/1506.01497.html>
2. Dai J., He K., Sun J. BoxSup: Exploiting Bounding Boxes to Supervise Convolutional Networks for Semantic Segmentation. [Электронный ресурс] URL: <https://arxiv.org/abs/1503.01640.html>
3. David G. Lowe Distinctive Image Features from Scale-Invariant Keypoints. [Электронный ресурс] URL: <http://link.springer.com/article/10.1023/B:VISI.0000029664.99615.94>
4. Dalal N., Triggs B. Histograms of Oriented Gradients for Human Detection. Cordelia Schmid and Stefano Soatto and Carlo Tomasi. International Conference on Computer Vision & Pattern Recognition (CVPR '05), Jun 2005, San Diego, United States. IEEE Computer Society, 1, pp. 886–893, 2005.
5. Cortes C., Vapnik V. Mach Learn (1995) 20: 273 Machine Learning September 1995, Vol. 20, Iss. 3, pp. 273–297.