

ОБЗОР СОВРЕМЕННЫХ МЕТОДОВ СЕГМЕНТАЦИИ ИЗОБРАЖЕНИЙ

Панасик А. А., Кукин Д. П., Купчина Е. В.
Кафедра вычислительных методов и программирования,
Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: {a.panasik, kudin, e.kupchina}@bsuir.by

Статья представляет собой обзор современных методов сегментации изображений, от классических алгоритмов до архитектур глубокого обучения. Проводится систематический анализ различных подходов.

ВВЕДЕНИЕ

Сегментация изображений – это процесс разделения цифрового изображения на множественные сегменты (множества пикселей) с целью упрощения или изменения представления изображения для более простого анализа [1]. По сути, это задача классификации каждого пикселя изображения. В отличие от простого детектирования объектов, сегментация предоставляет точные границы и маски объектов, что необходимо для детального понимания сцены.

Актуальность технологии подтверждается широким спектром применений:

1. В автономных транспортных системах для идентификации элементов дорожной сцены [2].
2. В медицинской диагностике для точного выделения анатомических структур и патологий [3].
3. В дистанционном зондировании для классификации земных покровов и мониторинга изменений.

Современные подходы к сегментации разделяются на классические алгоритмы, основанные на анализе низкоуровневых признаков, и методы глубокого обучения с использованием CNN для автоматического выделения признаков.

I. КЛАССИЧЕСКИЕ МЕТОДЫ СЕГМЕНТАЦИИ

Методы, которые опираются на низкоуровневые признаки изображения, такие как яркость, цвет, текстура или градиент.

Пороговые методы:

Пороговая сегментация является одним из простейших методов. Ее идея заключается в разделии пикселей на классы на основе их интенсивности путем выбора одного или нескольких пороговых значений [4].

Метод Оцу: Алгоритм автоматически вычисляет оптимальный порог, минимизирующий внутриклассовую дисперсию или, что эквивалентно, максимизирующий межклассовую дисперсию. Метод эффективен для изображений с бимодальной гистограммой.

Методы, основанные на кластеризации:

Методы рассматривают пиксели как точки в пространстве признаков и группируют их в кластеры.

K-means: Алгоритм итеративно назначает пиксели k кластерам, центроиды которых постоянно пересчитываются. Главным недостатком является необходимость априорного задания числа кластеров k .

Mean Shift: Алгоритм не требует задания числа кластеров. Он итеративно смещает точки к режиму плотности распределения данных, что позволяет находить кластеры произвольной формы.

Методы, основанные на областях (Region-based):

Подходы этой группы основываются на поиске связанных областей с однородными свойствами.

Region Growing: Алгоритм начинается с «затравки» (seed point) и последовательно присоединяет соседние пиксели, если они удовлетворяют заданному критерию однородности. Эффективность сильно зависит от выбора начальной точки и критерия.

Split and Merge: Метод начинается с всего изображения как одного региона, который рекурсивно разбивается на части, если регион неоднороден. Затем соседние однородные регионы объединяются.

Методы на основе активных контуров (Active Contours):

Модель «змей» (snakes) представляет собой энергетически-минимизирующую сплайну, которая эволюционирует под действием внутренних и внешних сил [5].

Внутренние силы отвечают за гладкость контура.

Внешние силы «притягивают» контур к границам объекта (например, вычисляются через градиент изображения).

Модель способна давать гладкие и замкнутые контуры, но требует аккуратной инициализации и может «застрять» в локальных минимумах. Основная цена за это — требование к большему размеченным датасетам и значительные вычислительные затраты на обучение и инференс.

II. Методы на основе глубокого обучения

С появлением больших размеченных датасетов и мощных GPU, глубокое обучение произвело революцию в сегментации изображений.

Семантическая сегментация (Semantic Segmentation):

Задача присвоения каждой пиксельной метки класса без разделения отдельных экземпляров объектов.

Fully Convolutional Networks (FCN) [6]: Замена полносвязных слоев на сверточные позволила обрабатывать изображения произвольного размера и генерировать карты сегментации. Для восстановления разрешения используются транспонированные свертки.

U-Net [7]: U-образная архитектура, ставшая стандартом в биомедицинской сегментации. Состоит из: энкодера (сжатие и извлечение признаков); декодера (восстановление разрешения); Skip-connections для передачи пространственной информации, утраченной при пулинге.

DeepLab Family - семейство архитектур, характеризующееся тремя ключевыми усовершенствованиями: Atrous (Dilated) Convolution, который увеличивает поле обзора без потери разрешения, Atrous Spatial Pyramid Pooling (ASPP), который обеспечивает многомасштабную обработку объектов, Encoder-Decoder with Depthwise Separable Convolution в котором используется декодер для уточнения границ объектов.

Сегментация экземпляров (Instance Segmentation):

Более сложная задача, которая требует не только классификации пикселей, но и разделения отдельных объектов одного класса.

Mask R-CNN: Является расширением Faster R-CNN [9], дополненное третьей ветвью для генерации бинарных масок объектов с помощью небольшой FCN. Ключевое улучшение - замена RoIPooling на RoIAlign - обеспечило существенное повышение точности пиксельного соответствия масок.

Классические методы эффективны в условиях ограниченных вычислительных ресурсов, для задач с сильной дифференциацией по яркости/цвету или когда обучение модели невозможно.

Методы на основе глубокого обучения демонстрируют точность на сложных, зашумленных данных с неочевидными границами объектов. Они способны учитывать высокоуровневый контекст сцены.

ЗАКЛЮЧЕНИЕ

Проведен системный обзор методов сегментации изображений: от простых пиксельных и

границных подходов к сложным контекстно-зависимым моделям глубокого обучения. Доминирование архитектур на основе CNN, является установившимся фактом для задачи семантической сегментации, в то время как Mask R-CNN и его производные лидируют в Instance Segmentation. Текущие исследования в области направлены на решение следующих проблем:

Эффективность: Разработка более легких и быстрых моделей для работы на мобильных и встраиваемых устройствах.

Новые архитектуры: Интеграция трансформеров (Vision Transformers, SegFormer) в задачи сегментации, что позволяет лучше моделировать глобальные зависимости в изображении.

Интерпретируемость: Создание моделей, не только дающих точный результат, но и позволяющих понять, на основе каких признаков было принято решение.

Таким образом, область сегментации изображений продолжает активно развиваться, двигаясь в сторону создания более умных, эффективных и универсальных моделей, способных работать в условиях реального мира.

СПИСОК ЛИТЕРАТУРЫ

1. Szeliski, R. Computer Vision: Algorithms and Applications / R. Szeliski. – Springer, 2010.
2. Chen, L. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation / L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A. L. Yuille // Computer Vision – ECCV 2018. – 2018. – P. 801-818.
3. Ronneberger, O. U-Net: Convolutional Networks for Biomedical Image Segmentation / O. Ronneberger, P. Fischer, T. Brox // Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. – 2015. – P. 234-241.
4. Otsu, N. A Threshold Selection Method from Gray-Level Histograms / N. Otsu // IEEE Transactions on Systems, Man, and Cybernetics. – 1979. – Vol. 9(1). – P. 62-66.
5. Kass, M. Snakes: Active contour models / M. Kass, A. Witkin, D. Terzopoulos // International Journal of Computer Vision. – 1988. – Vol. 1(4). – P. 321-331.
6. Long, J. Fully convolutional networks for semantic segmentation / J. Long, E. Shelhamer, T. Darrell // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2015. – P. 3431-3440.
7. He, K. Mask R-CNN / K. He, G. Gkioxari, P. Dollár, R. Girshick // Proceedings of the IEEE International Conference on Computer Vision (ICCV). – 2017. – P. 2961-2969.
8. Vaswani, A. Attention Is All You Need / A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin // Advances in Neural Information Processing Systems (NeurIPS). – 2017. – Vol. 30.
9. Xie, E. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers / E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, P. Luo // Advances in Neural Information Processing Systems (NeurIPS). – 2021. – Vol. 34. – P. 12077-12090.