Умный магазин: автоматическое детектирование пустот в товарной выкладке на основе данных системы видеонаблюдения

Ю. А. Талалаева[1], М. Г. Бабенко[2], В. А. Кучуков[3]
Северо-Кавказский федеральный университет
Ставрополь, Россия
[1]iutalalaeva@ncfu.ru, [2]mgbabenko@ncfu.ru,
[3]vkuchukov@ncfu.ru

А. Н. Черных[4], Г. И. Радченко[5]
Центр научных исследований и высшего образования
Энсенада, Мексика
Южно-Уральский государственный университет
Челябинск, Россия
[4]chernykh@cicese.mx, [5]gleb.radchenko@susu.ru

***Аннотация*. Работа посвящена одному из аспектов создания целостной системы автоматизации для умного магазина – контролю выкладки товаров на полках и своевременному заполнению образовавшихся пустот. Предлагаемая модель решения основана на каскаде нейронных сетей, выполняющих роль сегментатора и классификатора, дополненном алгоритмическим решением выделения областей потенциальных пустот. Для обучения классификатора, с помощью созданного алгоритма была сформирована обучающая выборка из 30000 изображений. И выборка из 3000 изображений для валидации. После обучения на 10000 выборках была достигнута точность в 96.73%.**

Ключевые слова: умный магазин, нейронные сети, семантическая сегментация, системы видеонаблюдения, контроль товарной выкладк

I. Введение

В современном мире развитие технологий искусственного интеллекта и, в частности, компьютерного зрения, дает широкие возможности создания комплексных систем автоматизированного контроля объектов помощью инструментов, полноценно пользоваться которыми раньше мог только человек. Это относится к системам видеоаналитики, которые постепенно превращаются из пассивного «глупого» наблюдателя в центральную нервную систему контролируемого объекта и позволяют говорить о зарождающейся концепции «самоконтроля».

С практической точки зрения парадигма умного магазина преследует две глобальные цели: анализ поведения клиента; управление клиентским опытом [1].

Использование интеллектуальной видеоаналитики позволяет создавать системы, минимизирующие негативный клиентский опыт за счет снижения количества отрицательных факторов, влияющих на него. Благодаря активной аналитике записей с камер видеонаблюдения возможно, как предсказание появления негативных явлений, так и своевременное уведомление персонала о их наличии. В частности, одним из значимых факторов, влияющих на клиентский опыт, является наличие на прилавке товара, необходимого покупателю

Столкнувшись пару раз с отсутствием нужного ассортимента в точке продаж, даже самый лояльный клиент может уйти к конкуренту.

В данной работе описывается возможная модель системы компьютерного зрения для решения задачи своевременного автоматического детектирования пустот в товарной выкладке путем анализа данных с камер системы видеонаблюдения. При выявлении целевого события (отсутствие товара) формируется статистика для дальнейшей бизнес-аналитики и оптимизации закупок, а также рассылаются уведомления ответственному персоналу.

II. Обзор работ

Система компьютерного зрения, которой посвящена данная работа, основана на каскадном применении двух нейронных сетей. Первая выполняет семантическую сегментацию кадра – поиск областей торговых прилавков и отделения всех остальных составляющих как фона, не воспринимаемого последующим алгоритмом. А вторая выполняет классификацию потенциальных пустот, выделяемых авторским алгоритмом внутри области полок.

Традиционно задачи сегментации изображений в компьютерном зрении решались (и продолжают решаться) с помощью аппарата случайных полей и алгоритмов минимизации энергии (см., например, [11, 12]). Начиная с 2012 года для семантической сегментации изображений начинают использоваться варианты сверточных нейронных сетей.

В этом разделе мы опишем основные сверточные архитектуры для компьютерного зрения. Отметим, что качественная семантическая сегментация изображений

Работа выполнена при финансовой поддержке РФФИ, проекты №18-07-01224, 18-07-00109, и стипендий Президента Российской Федерации для молодых ученых и аспирантов № СП-1215.2016.5, СП-2236.2018.5, и ООО «Магазин будущего».

является первым, необходимым, но не достаточным шагом для решения итоговой задачи определения пустот на полках. Для решения итоговой задачи потребуется выделение регионов, достаточно больших либо обладающих определенной геометрической формой, которые требуется считать пустотами.

Модель *SegNet* [14] была разработана в 2015 г. и представляет собой архитектуру кодирования-декодирования (encoder-decoder), позволяющую преобразовывать входное изображение банком фильтров в некоторое внутреннее представление, из которого затем операциями увеличения пространственного разрешения (unpooling, upsampling, deconvolution и т.п. в различных вариантах) вычисляется итоговая сегментация изображения. Кодировщик SegNet основан на нейросетевой архитектуре VGG-16 (последовательностях слоев вида conv-conv-pool) и может быть инициализирован уже настроенной сетью, например, предобученной на базе ImageNet. Точность на VOC датасете: 59.9%.

Одной из важных проблем при сегментации определенных типов изображений, например, результатов сканирования в биомедицине, является необходимость получения максимально точных сегментаций, причем количество обучающих данных для решения этой задачи может быть сравнительно невелико. Поэтому требуется модификация архитектуры «кодировщик-декодировщик», которая бы позволила ограничить число настраиваемых параметров и получить качественную сегментацию

Архитектура *U-Net* [15] (получившая название по форме нарисованной сети) решает эту задачу, вводя прямые соединения между сверточными и «разверточными» слоями, оперирующими с данными одного и того же размера (под «разверточными» понимаются сверточные слои подсети-декодировщика).

DeepLab использует в своей основе архитектуру свёрточной сети Resnet-101 для семантической сегментации выполненной как «atrous convolution» (или тёмная свёртка), разномасштабных входных данных и макс-пулингом для слияния всех масштабов и тёмной разреженной пирамидой. Точность такой модели на VOC датасете составляет 79.7%

Использование архитектуры DeepLab в данной работе обусловлено лучшими показателями точности. Необходимость разработки собственного алгоритма выделения потенциальных пустот внутри сегментированных областей торговых прилавков обусловлена тем, что семейства R-CNN моделей, SSD, YOLO являются детекторами общего назначения. Детектирование однородных областей, по сути, не являющихся отдельными объектами, для них проблематично.

Две различных концепции, лежащих в основе R-CNN и YOLO, обладают различными преимуществами и недостатками. В то время как варианты R-CNN показывают наилучшие показатели точности, YOLO даёт наилучшую производительность.

В данной работе, авторы предлагают алгоритм, достигающий лучших показателей точности и скорости, чем оба названных алгоритма

III. Описание модели

На первом шаге изображение сегментируется алгоритмом DeepLab [2].

Включение сегментатора в пайплайн решения данной задачи обусловлено не только использованием его для сужения области действия алгоритмического детектора, но и функциональной нагрузкой на него. Результат работы сегментатора дает возможность ответить на вопрос, товаров какого рода не хватает в обнаруженной пустоте – это могут быть фрукты, кисломолочная или мясная продукция, бакалея. Эти данные косвенно вычислимы из классов, присвоенных сегментационным маскам, внутри которых находится пустая область.

Обучающая выборка для сегментатора формируется из изображений, состоящих из 4 основных сегментов и фона (рис. 1):

- вертикальные открытые полки – выделены красным;
- вертикальные закрытые полки (чаще холодильники) – выделены зеленым;
- горизонтальные полки – выделены желтым;
- овощной развал – выделены синим;
- фон (пол, люди, прилавки и т.п.) – выделены черным.

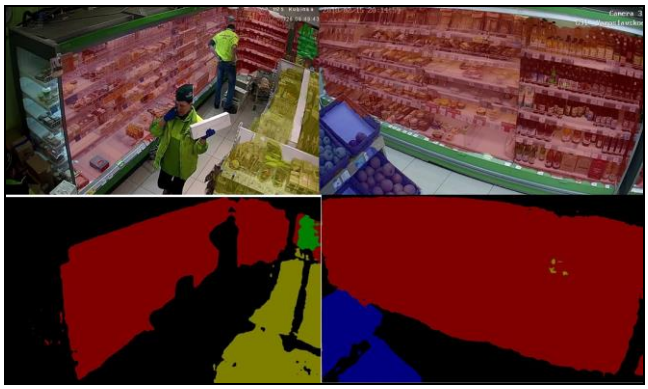

Рис. 1. Сегментированное изображение и его маска

Следующим шагом является алгоритмический поиск потенциальных пустых областей внутри сегментированных масок. С помощью детектора границ Канни на изображении, переведенном в оттенки серого, выделяются контуры однородных областей. Краткое поэтапное описание действий можно свести к следующему:

- удаление шума – сглаживание;
- поиск градиентов (используются четыре фильтра для выделения вертикальных, горизонтальных и диагональных ребер-границ);

- выделение локальных максимумов в качестве границ.
- пороговая фильтрация (параметры threshold1 и threshold2) – выделение только значимых границ.

Значения параметров для вычисления гистерезиса [4] были подобраны эмпирически: threshold1 = 50, threshold2 = 100.

В результате обработки изображение принимает вид представленный на рис. 2.

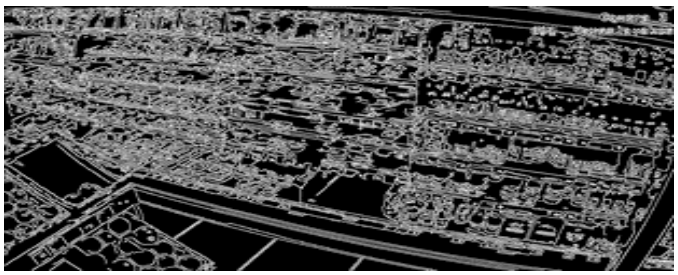

Рис. 2. Результат работы алгоритма Канни

Из приведенного изображения хорошо видно, что пустоты на полках являются однородными областями большой площади. В том числе это касается и корзинок с фруктами или овощами, поскольку они имеют однотонное дно. Однако, границы, выделенные алгоритмом Канни, не во всех случаях являются замкнутыми. Анализ оригинальных изображений показал, что пустотой является однородная область, которая претерпевает изменения при переходе от света к тени при неизменной цветовой составляющей.

При этом вычисление евклидового расстояния для определения однородности области в обычном RGB пространстве становится неприменимо [4]. Для работы с цветовыми составляющими изображение транслируется в цветовое пространство CIELAB [5].

Это позволяет оперировать линейными изменениями цвета с точки зрения человеческого восприятия – одинаковое изменение значений координат цвета в разных областях цветового пространства производит одинаковое ощущение изменения цвета. В качестве метрики цветоразности используется цветовое расстояние Delta E:

$$\Delta E^*_{ab} = \sqrt{\left(L^*_2 - L^*_1\right)^2 + \left(a^*_2 - a^*_1\right)^2 + \left(b^*_2 - b^*_1\right)^2},$$

где $\left(L^*_1, a^*_1, b^*_1\right)$ и $\left(L^*_2, a^*_2, b^*_2\right)$– координаты в цветовом пространстве CIELAB.

В результате получаем набор потенциальных пустот (рис. 3).

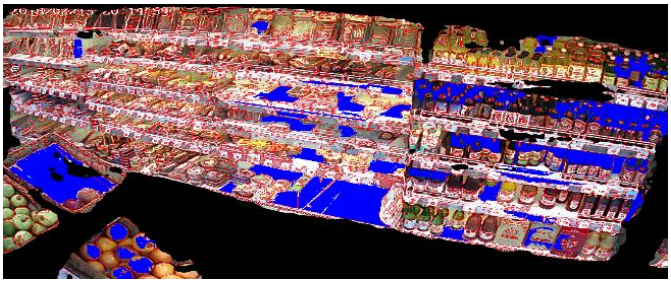

Рис. 3. Потенциально пустые области

Использование маски сегментации позволяет отделить области полок, от других однородных областей (пол, стены). Далее, полученные потенциально пустые области ограничиваются минимально возможной прямоугольной рамкой, с отступом по 10 пикселей по краям.

Для того избавится от областей многократно пересекающихся друг с другом используется алгоритм non-maxima supression [6]. На выходе получаем множественный набор изображений небольшого размера, являющихся потенциальными пустотами (рис. 4).

Рис. 4. Потенциально пустые области

Полученные изображения подаются на вход в классификационную модель, построенную на архитектуре Renset18 [7], отвечающую на простой вопрос, является ли изображение пустотой или нет.

IV. Экспериментальные результаты

В качестве обучающей выборки для сегментатора использовался набор из 6900 размеченных изображений с четырьмя классами сегментации. Ввиду граничащих друг с другом, похожих по формальным признакам классам (например, вертикальные открытые и вертикальные закрытые прозрачной дверцей полки), на данном этапе удалось достигнуть точности в 63%.

Основная ошибка перекрытия IoU наблюдается на краях областей (рис. 1). Поскольку эта метрика не является доминантной в отношении определения точности всего решения, такой показатель является приемлемым.

Для обучения классификатора, с помощью созданного алгоритма была сформирована обучающая выборка из 30000 изображений. И выборка из 3000 изображений для валидации. После обучения на 10000 выборках была достигнута точность в 96.73%.

В общем, по всему решению получилось достигнуть следующих показателей точности детекции пустот на тестовой выборке в 700 изображений: 92,34% пустот были обнаружены; 2% ложных срабатываний; 85% avg IoU.

Следует отметить, что описанный в работе алгоритм нахождения однородных пустот в терминах CIE delta E, позволяет конкурировать с Region Proposal System [7] в семействе R-CNN детекторов. При использовании самой оптимальной реализации Faster R-CNN VGG-16 [8] на сформированном датасете, она формирует от 220 до 315 предлагаемых пустот, в то время как предложенный алгоритм – не более 49. Модель Faster R-CNN выполняет обработку одного изображения в среднем за 0.12 секунды (в зависимости от количества пустот на одном кадре), в то время как наш алгоритм выполняет детекцию в течение 0.07 сек, включая этап классификации.

Однако это не единственное преимущество созданного алгоритма перед семейством R-CNN-подобных моделей. При размере обучающей выборки в 6900 изображений, точность R-CNN при детекции пустот в товарной выкладке достигает лишь 84.93% при 65% avg IoU.

Использование YOLO v.2 в качестве основного детектора также не дало существенных улучшений в качестве, показав результат в 80.96% при 67% avg IoU [10]. Однако скорость работы данного алгоритма значительно опережает авторскую – на обработку одного изображения Yolo потребовалось 0.03 секунды в среднем.

Проводя сравнения с универсальными детекторами, стоит отметить, что предложенный авторами алгоритм не являет собой детектор общего назначения. Это узкоспециализированное решение для конкретной задачи. Поэтому его нельзя сравнивать с решениями общего назначения.

V. Заключение

В рамках данной работы был разработан алгоритм детекции пустот в товарной выкладке магазинных прилавков. Полученные результаты превзошли по скорости и точности использование детекторов общего назначения. Однако, данный метод направлен лишь на решение узкоспециализированной задачи. Поэтому нельзя провести равноправное сравнение представленного алгоритма с архитектурами YOLO или семейством R-CNN-подобных моделей.

На данный момент с помощью семантической сегментации, используемой в решении, косвенно можно определить только группу товаров, которая должна занять пустоту на полке. Подобных выводов не достаточно для полноценного использования предложенного решения в бизнес-аналитике. Поэтому дальнейшая работа будет направлена на более точную локализацию вида товара, который должен занять вакантное место.

Список литературы

[1] Hwangbo H., Kim Y.S., Cha K.J. Use of the smart store for persuasive marketing and immersive customer experiences: A case study of korean apparel enterprise, Mobile Information Systems, 2017.

[2] Chen L.C. et al, Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, arXiv preprint arXiv:1606.00915, 2016.

[3] Canny J. A computational approach to edge detection, Readings in Computer Vision, 1987, pp. 184-203.

[4] Sharma G., Wu W., Dalal E.N. The CIEDE2000 color difference formula: Implementation notes, supplementary test data, and mathematical observations, Color Research & Application, 2005, vol. 30, no.1, pp. 21-30.

[5] Tomasi C., Manduchi R. Bilateral filtering for gray and color images, Proc. in IEEE Sixth International Conference on Computer Vision, 1998, pp. 839-846.

[6] Neubeck A., Van Gool L. Efficient non-maximum suppression, Proc. of the IEEE 18th International Conference on Pattern Recognition, 2006, vol.3, pp. 850-855.

[7] He K. et al, Deep residual learning for image recognition, Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp.770-778.

[8] Girshick R. et al, Rich feature hierarchies for accurate object detection and semantic segmentation, Proceedings of the IEEE conference on computer vision and pattern recognition, 2014,pp.580-587.

[9] Ren S. et al, Faster r-cnn: Towards real-time object detection with region proposal networks, Advances in neural information processing systems, 2015, pp.91-99.

[10] Redmon J., Farhadi A. YOLO9000: better, faster, stronger, arXiv preprint, 2016, vol. 1612.

[11] Pal N.R., Pal S.K. A review on image segmentation techniques, Pattern recognition, 1993, vol.26, no 9, pp.1277-1294.

[12] He X., Zemel R.S., Carreira-Perpiñán M.Á. Multiscale conditional random fields for image labeling, Proceedings of the IEEE computer society conference on Computer vision and pattern recognition, 2004, vol.2.

[13] Long J., Shelhamer E., Darrell T. Fully convolutional networks for semantic segmentation, Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, p.3431-3440.

[14] Badrinarayanan V., Kendall A., Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation, IEEE transactions on pattern analysis and machine intelligence, 2017, vol. 39, no.12, pp.2481-2495.

[15] Ronneberger O., Fischer P., Brox T. U-net: Convolutional networks for biomedical image segmentation, Proc. of the International Conference on Medical image computing and computer-assisted intervention, Springer, Cham, 2015, pp.234-241.