

О подходах к идентификации экземпляров мусора на цифровых изображениях

Макаров А.А., СПбГУ, Санкт-Петербург a.a.makarov@spbu.ru,

Макарова С.В., СПбГУ, Санкт-Петербург sdrobot@mail.ru,

Шабунин А.Н., СПбГУ, Санкт-Петербург shandr52@gmail.com

Аннотация

В работе анализируются современные методы идентификации экземпляров мусора на цифровых изображениях с использованием технологий искусственного интеллекта. Рассмотрены подходы компьютерного зрения, глубокого обучения и гибридные системы, а также их применение для классификации, детекции и сегментации отходов. Выделены ключевые проблемы: вариативность объектов, дисбаланс данных и требования к инфраструктуре.

Введение

Рост объемов отходов и ужесточение экологических стандартов требуют принципиально новых методов обработки мусора. Ключевым этапом в цепочке переработки является точная идентификация каждого экземпляра отходов – от пластиковой бутылки до электронного лома. Автоматизация сортировки отходов становится критически важной задачей в контексте глобальных экологических вызовов. Рост объемов мусора, ограниченность ресурсов и необходимость сокращения полигонных отходов требуют внедрения инновационных технологий. Современные подходы на базе искусственного интеллекта, в частности методы компьютерного зрения, позволяют не только точно распознавать и классифицировать экземпляры вторичного сырья на цифровых изображениях, но и определять их границы, материал и степень загрязнения, что критично для автоматизации сортировки. Это не только повышает эффективность переработки, но и снижает затраты, заменяя трудоемкую ручную сортировку автоматизированными системами. Экономический аспект здесь очевиден: автоматизация сокращает операционные расходы и минимизирует человеческие ошибки. Кроме того, стремительное развитие глубокого обучения и увеличение вычислительных мощностей открывают новые горизонты для создания адаптивных и высокоточных моделей.

Глобальный объем твердых отходов превышает 2 млрд. тонн в год, при этом эффективность их переработки остается крайне низкой [1]. Основным

барьером является несовершенство методов сортировки: ручной труд демонстрирует до 30% ошибок неправильной классификации. Важной задачей для нашей страны является обеспечение перехода к созданию замкнутого цикла вторичной переработки отходов, в том числе через внедрение технологий искусственного интеллекта и развитие перерабатывающих мощностей, для возврата ресурсов в производство. Как показывает опыт внедрения системы AMP Robotics, автоматизация на базе технологий искусственного интеллекта, включая компьютерное зрение и глубокое обучение, снижает операционные расходы и увеличивает как скорость обработки, так и точность сортировки [2, 3]. Развитие нейросетевых архитектур (трансформеры, сверточные сети) и появление специализированных датасетов (TACO, WasteNet, TrashCan и т.п.) сделали возможной идентификацию даже мелких или деформированных объектов. Актуальность работы обусловлена экологическими (сокращение полигонных отходов) и экономическими (оптимизация переработки) факторами.

Цель работы – систематизировать современные подходы, оценить их эффективность и определить направления развития. В работе анализируются современные методы идентификации экземпляров мусора на цифровых изображениях с использованием технологий искусственного интеллекта. Рассмотрены подходы компьютерного зрения, глубокого обучения и гибридные системы, а также их применение для классификации, детекции и сегментации отходов. Приведены примеры внедрения в промышленности и бытовом секторе. Обозначены перспективы интеграции с мультимодальными сенсорами и открытыми стандартами. Проведенное исследование подтверждает, что автоматизация идентификации мусора является критическим элементом перехода к циркулярной экономике.

Методы идентификации

Классические алгоритмы компьютерного зрения

Ранние методы основывались на анализе низкоуровневых признаков. Например, пороговые методы, опираясь на цветовые модели (RGB, HSV), использовались для отделения стекла (высокая яркость) от пластика. Однако такие методы демонстрировали низкую точность при сложном освещении [4]. Другой подход заключался в анализе текстуры и формы – детекция на основе HOG (Histogram of Oriented Gradients) и SIFT (Scale-Invariant Feature Transform) – и позволял выявлять ключевые точки для идентификации формы объектов, но уступал нейросетевым методам в скорости и точности [5].

Глубокое обучение

- **Детекция объектов.** Современные архитектуры, такие как YOLO (You Only Look Once), обрабатывают изображения в реальном времени (до 150 FPS) с точностью mAP 70% для датасета TACO [6]. Модель Faster R-CNN, использующая прогноз регионов кандидатов (Region Proposal Network), достигает mAP 85%, но требует значительных вычислительных ресурсов [7].
- **Сегментация.** Для сегментации объектов применяются архитектура U-Net, изначально разработанная для медицинских изображений, позволяющая выделять контуры мусора на сложном фоне [8], и архитектура Mask R-CNN, совмещающая детекцию и выделение контуров, эффективная для работы с перекрывающимися объектами, такими как смятая упаковка, смятая бутылка [9].
- **Трансформеры.** Трансформеры, такие как Vision Transformers (ViT), анализируют глобальные зависимости в изображениях, улучшая идентификацию фрагментированных объектов (осколки стекла) [10]. Иерархическая структура трансформера Swin Transformer снижает вычислительные затраты при работе с высоким разрешением [11].

Гибридные подходы

Комбинация сверточных нейронных сетей с данными от лидаров или спектрометров (multispectral imaging) повышает точность распознавания (см., например, [12, 13]). Применение тепловизоров (thermal imaging) может использоваться для идентификации органических отходов по тепловому следу.

Данные и вычислительная инфраструктура

Наборы данных и аугментация

Публичные датасеты, такие как TACO [14], WasteNet [15], TrashCan [16] и т.п., могут предоставить базовую основу для обучения. Однако они недостаточно покрывают редкие классы (батарейки, электроника). Для решения этой проблемы применяется генерация синтетических данных, например, через GAN (StyleGAN3) или 3D-рендеринг (Blender), моделируя деформированные объекты, такие как смятые бутылки с вариациями освещения [17]. Разметка

данных выполняется с помощью полуавтоматических инструментов (CVAT, Supervisely), а аугментация включает геометрические преобразования (поворот, масштабирование) и наложение артефактов (блики, пыль, капли воды) через библиотеку Albumentations [18]. Микс-аугментация (MixUp) улучшает обобщающую способность моделей, особенно для наборов несбалансированных данных [19].

Проблемы идентификации

Вариативность объектов внутри одного класса (например, ПЭТ бутылки разного цвета, объема, некоторые из которых плохо перерабатываются) и их динамические деформации (смятие, разрывы) усложняют обучение моделей. Фоновый шум, включая наложение объектов друг на друга, блики и погодные условия, требует дополнительной обработки изображений.

Вычислительные мощности

Модели, обученные на синтетических данных, могут некорректно работать в реальных условиях. Обучение больших моделей требует значительных ресурсов. Для работы современных моделей необходимы промышленные системы GPU с поддержкой TensorRT. Мобильные решения (с применением дронов и роботов) могут использовать компактные архитектуры типа MobileNet [20].

Практическое применение

- **Роботизированные линии (AmpRobotics, ZenRobotics).** Интеграция технологий искусственного интеллекта с роботами-манипуляторами позволяет повысить скорость сортировки [2, 3].
- **Мобильные приложения.** On-device модели, например на базе TensorFlow Lite, помогают идентифицировать отходы, используя мобильные устройства [21].
- **Умные контейнеры (Bin-e).** Датчики и камеры анализируют наполнение контейнеров и оптимизируют логистику вывоза [22].

Перспективные технологические направления

Совершенствование методов идентификации мусора требует внедрения инновационных технологий и интеграции смежных подходов. Одним из подходов является использование NeRF (Neural Radiance Fields) [23], позволяющего синтезировать объемные структуры объектов из изображений с известными положениями камеры. Эта технология применима для идентификации деформированных объектов, таких как смятые банки и бутылки, за счет анализа их объемной геометрии. Дополняют этот подход диффузионные модели (Stable Diffusion, DALL-E), которые могут генерировать синтетические изображения мусора в сложных условиях, например, с имитацией загрязнений или частичного перекрытия. Такие модели помогают расширять наборы данных, компенсируя нехватку реальных данных. Для оптимизации процесса разметки можно использовать активное обучение (Active Learning), позволяющее сократить затраты на подготовку данных.

Комбинация мультимодальных сенсоров (RGB-камеры, гиперспектральные датчики, тактильные сенсоры и т. п.) повышает точность распознавания. Технология блокчейна внедряется для прозрачного трекинга жизненного цикла отходов [24]. Пилотные проекты, такие как IBM Waste Tracker, уже демонстрируют эффективность этой методики для отслеживания переработки пластика.

Заключение

Идентификация экземпляров мусора на цифровых изображениях – комплексная задача, требующая сочетания методов компьютерного зрения, глубокого обучения и мультимодальных данных. Современные подходы демонстрируют высокую эффективность, однако их внедрение требует учета вариативности объектов, ограничений данных и инфраструктуры. Перспективы связаны с гибридными системами, сочетающими компьютерное зрение, спектроскопию и робототехнику.

Список литературы

- [1] Kaza, S., et al. (2018). *What a Waste 2.0: A Global Snapshot of Solid Waste Management to 2050*. Urban Development. DOI: 10.1596/978-1-4648-1329-0.

- [2] *AMP Robotics: Case Studies*. URL: <https://ampsortation.com/resources/case-studies>.
- [3] *ZenRobotics: Case Studies*. URL: <https://www.terex.com/zenrobotics/case-studies>.
- [4] Krizhevsky, A., et al. (2012). "ImageNet Classification with Deep Convolutional Neural Networks." *Advances in Neural Information Processing Systems*, 25. URL: NIPS 2012.
- [5] Lowe, D. G. (2004). "Distinctive Image Features from Scale-Invariant Keypoints." *International Journal of Computer Vision*, 60(2), 91-110. DOI: 10.1023/B:VISI.0000029664.99615.94.
- [6] Redmon, J., et al. (2016). "You Only Look Once: Unified, Real-Time Object Detection." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. DOI: 10.1109/CVPR.2016.91.
- [7] Ren, S., et al. (2015). "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." *Advances in Neural Information Processing Systems*, 28. URL: NeurIPS 2015.
- [8] Ronneberger, O., et al. (2015). "U-Net: Convolutional Networks for Biomedical Image Segmentation." *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. DOI: 10.1007/978-3-319-24574-4_28.
- [9] He, K., et al. (2017). "Mask R-CNN." *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. DOI: 10.1109/ICCV.2017.322.
- [10] Dosovitskiy, A., et al. (2020). "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale." *International Conference on Learning Representations (ICLR)*. URL: ICLR 2020.
- [11] Liu, Z., et al. (2021). "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows." *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. DOI: 10.1109/ICCV48922.2021.00986.
- [12] Lu, G., et al. (2022). "Deep multimodal learning for municipal solid waste sorting." *Sci. China Technol. Sci.*, 65, 324–335. DOI: <https://link.springer.com/article/10.1007/s11431-021-1927-9>.
- [13] Shiddiq, M. et al. (2023). "Plastic and organic waste identification using multispectral imaging."

- Materials Today: Proceedings*, 87 (2), 338-344. DOI: <https://www.sciencedirect.com/science/article/pii/S2214785323014724>.
- [14] Proença, P. F., et al. (2020). *TACO: Trash Annotations in Context Dataset*. arXiv:2003.06975. URL: arXiv:2003.06975.
- [15] *WasteNet: The world's largest dataset for waste*. URL: <https://recycleye.com/wastenet/>.
- [16] Hong, J., et al. (2020). *Trashcan: a semantically-segmented dataset towards visual detection of marine debris*. arXiv:2007.08097. URL: arXiv:2007.08097.
- [17] Karras, T., et al. (2020). "Analyzing and Improving the Image Quality of StyleGAN." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. DOI: 10.1109/CVPR42600.2020.00813.
- [18] Buslaev, A., et al. (2020). "Albumentations: Fast and Flexible Image Augmentations." *Information*, 11(2), 125. DOI: 10.3390/info11020125.
- [19] Zhang, H., et al. (2017). "Mixup: Beyond Empirical Risk Minimization." arXiv:1710.09412. URL: arXiv:1710.09412.
- [20] Howard, A. G., et al. (2017). "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications." arXiv:1704.04861. URL: arXiv:1704.04861.
- [21] Handhayani, T., Hendryli, J. (2023). "Leboh: An Android Mobile Application for Waste Classification Using TensorFlow Lite." *Lecture Notes in Networks and Systems*, 544. DOI: 10.1007/978-3-031-16075-2_4
- [22] *Bin-e: recycling solutions*. URL: <https://www.bine.world/solutions>.
- [23] Mildenhall, B., et al. (2020). "NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis." *European Conference on Computer Vision (ECCV)*. DOI: 10.1007/978-3-030-58452-8_24.
- [24] Nakamoto, S. (2008). *Bitcoin: A Peer-to-Peer Electronic Cash System*. URL: <https://bitcoin.org/bitcoin.pdf>.