

РАСПОЗНАВАНИЕ ИЗОБРАЖЕНИЙ И УДАЛЕНИЕ ФОНА: СОВРЕМЕННЫЕ МЕТОДЫ И ТЕХНОЛОГИИ

Орлова Лариса Викторовна, кандидат
экономических наук, УЛГТУ, кафедра
«Финансы и кредит»

Аннотация: Распознавание изображений и удаление фона являются важными задачами в компьютерном зрении и широко применяются в электронной коммерции, дополненной реальности, медицине и других сферах. В данной статье рассматриваются основные методы распознавания изображений, включая классические алгоритмы и современные нейросетевые подходы. Анализируются методы удаления фона, основанные на глубоких нейронных сетях, а также возможности использования трансформеров для сегментации изображений. В статье приведен сравнительный анализ технологий и обсуждаются перспективные направления развития данной области.

Ключевые слова: компьютерное зрение, удаление фона, нейросети, сегментация, искусственный интеллект, глубокое обучение, распознавание изображений.

Abstract: Image recognition and background removal are essential tasks in computer vision, widely used in e-commerce, augmented reality, healthcare, and other fields. This paper examines key image recognition methods, including classical algorithms and modern neural network approaches. Background removal techniques based on deep neural networks are analyzed, along with the potential use of transformers for image segmentation. A comparative analysis of technologies is presented, and future research directions in this field are discussed.

Keywords: computer vision, image recognition, background removal, neural networks, segmentation, artificial intelligence, deep learning.

Введение

Распознавание изображений — это процесс идентификации объектов, сцен и других элементов на цифровых изображениях с применением алгоритмов машинного обучения и компьютерного зрения. Оно используется в системах безопасности, автономном вождении, промышленном контроле, медицинской диагностике и других областях [1].

Удаление фона представляет собой процесс выделения объектов изображения и устранения нежелательного окружения. Это особенно важно в электронной коммерции, обработке фотографий, создании контента для маркетинга и видеоиндустрии [2]. Современные методы основаны на использовании нейросетевых архитектур, таких как Mask R-CNN, DeepLabV3+, U-Net, а также на трансформерных моделях, например Segment Anything Model (SAM) [3].

Методы распознавания изображений. Существуют различные подходы к распознаванию изображений, которые можно разделить на классические и нейросетевые.

Классические методы компьютерного зрения

Классические методы включают:

- Фильтры Собеля и Канни – используются для выделения контуров объектов [4].
- Методы выделения признаков – SIFT (Scale-Invariant Feature Transform) и ещё есть SURF (Speeded-Up Robust Features) применяется для поиска схожих объектов в изображениях [5].

- Методы кластеризации – например, алгоритм K-Means, используется для группировки пикселей на основе цветовых характеристик [6].

Нейросетевые методы

Современные технологии используют глубокие сверточные нейронные сети (CNN):

- ResNet (Residual Networks) – это мощная архитектура, которая решает проблемы с исчезающими градиентами, что позволяет обучать глубокие сети [7].
- YOLO (You Only Look Once) – метод детекции объектов в реальном времени [8].
- Mask R-CNN – модель сегментации изображений, позволяющая не только обнаруживать объекты, но и выделять их точные контуры [9].

Удаление фона с помощью нейросетей. Удаление фона может выполняться с использованием следующих подходов:

Классические методы

- Метод Chroma Key (зеленый экран) – широко применяется в киноиндустрии для замены фона на основе цветового диапазона [10].
- Гистограмма интенсивности – позволяет выделять объект, основываясь на контрасте с фоном [11].

Глубокие нейросетевые модели

- U-Net – сверточная нейросеть, используемая для сегментации изображений, применяется в медицине и обработке фотографий [12].
- DeepLabV3+ – архитектура, основанная на CNN, обеспечивающая точное выделение объектов [13].
- MODNet (Matting Objective Detection Network) – специализированная модель для удаления фона в режиме реального времени [14].

Трансформеры и новые архитектуры

Современные решения включают модели, использующие механизм внимания (attention):

- Segment Anything Model (SAM) – модель от Meta, способная сегментировать объекты без предварительного обучения [15].
- HRNet (High-Resolution Network) – используется для детализированной сегментации сложных объектов [16].

Сравнение технологий. Таблица 1 представляет сравнительный анализ популярных методов удаления фона.

Модель	Метод	Точность (1-10)	Скорость работы	Применение
Chroma Key	Классический	6	Быстро	Видео, телевидение
U-Net	CNN	8	Среднее	Медицинская визуализация, фото
DeepLab V3+	CNN	9	Среднее	Графический дизайн, AR/VR
MODNet	CNN	9	Высокая	Онлайн-редакторы, мобильные приложения
SAM	Transformer	10	Среднее	Универсальное удаление фона

Заключение

В данной статье рассмотрены методы распознавания изображений и удаления фона. Классические алгоритмы, такие как SIFT и Chroma Key, по-прежнему применяются, но глубокие нейросетевые модели, включая U-Net, DeepLabV3+ и MODNet, демонстрируют значительно лучшие результаты. Новые трансформерные архитектуры, такие как Segment Anything Model, открывают новые возможности для автоматизированной обработки изображений.

В будущем можно ожидать интеграции данных моделей с генеративными нейросетями для улучшения качества сегментации. Дальнейшее развитие в этой области позволит создать универсальные решения для различных приложений, включая дополненную реальность, автоматическое редактирование фотографий и видеоконтент.

Список источников:

1. Krizhevsky A., Sutskever I., Hinton G. "ImageNet Classification with Deep Convolutional Neural Networks" // Advances in Neural Information Processing Systems, 2012.
2. He K., Zhang X., Ren S., Sun J. "Deep Residual Learning for Image Recognition" // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
3. Redmon J., Farhadi A. "YOLOv3: An Incremental Improvement" // arXiv, 2018.
4. Lowe D.G. "Distinctive Image Features from Scale-Invariant Keypoints" // International Journal of Computer Vision, 2004.
5. R-CNN: Region-based Convolutional Neural Networks // IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015.

6. Zhao H., Shi J., Qi X., Wang X., Jia J. "Pyramid Scene Parsing Network" // CVPR, 2017.
7. Xiao T., Liu Y., Zhou B., Jiang Y., Sun J. "Unified Perceptual Parsing for Scene Understanding" // ECCV, 2018.
8. Meta AI. "Segment Anything Model (SAM)" // arXiv, 2023.
9. He K., Gkioxari G., Dollár P., Girshick R. "Mask R-CNN" // International Conference on Computer Vision (ICCV), 2017.
10. Chen L.C., Papandreou G., Schroff F., Adam H. "Rethinking Atrous Convolution for Semantic Image Segmentation" // arXiv preprint, 2017.
11. Ronneberger O., Fischer P., Brox T. "U-Net: Convolutional Networks for Biomedical Image Segmentation" // International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015.
12. Zhang X., Zhang Y., Kong Y. "MODNet: Real-Time Trimap-Free Portrait Matting via Objective Decomposition" // CVPR, 2021.
13. Chen L.C., Zhu Y., Papandreou G., Schroff F., Adam H. "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation" // ECCV, 2018.
14. Sun K., Xiao B., Liu D., Wang J. "Deep High-Resolution Representation Learning for Human Pose Estimation" // IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019.
15. Kirillov A., Mintun E., Ravi N., Mao H., Rolland C., Gustafson L., Xiao T., Whitehead S., Dollár P. "Segment Anything" // arXiv preprint, 2023.
16. Dosovitskiy A., Beyer L., Kolesnikov A., Weissenborn D., Zhai X. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale" // ICLR, 2021.