

УДК 004

АВТОМАТИЧЕСКАЯ РАЗМЕТКА ОБУЧАЮЩИХ ВЫБОРОК В КОМПЬЮТЕРНОМ ЗРЕНИИ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

А. К. Журавлёв¹ [0009-0001-1265-9916], **К. А. Григорян**² [0000-0001-6470-1832]

^{1, 2}*Казанский (Приволжский) федеральный университет, ул. Кремлевская,
35, г. Казань, 420008*

¹UnMelow@yandex.ru, ²karigri@yandex.ru

Аннотация

Рассмотрена проблема автоматической разметки обучающих выборок в области компьютерного зрения с использованием методов машинного обучения.

Разметка данных является ключевым этапом в разработке и обучении моделей глубокого обучения, однако процесс создания размеченных данных зачастую требует значительных временных и трудовых затрат. В статье предложен механизм автоматической разметки, основанный на использовании сверточных нейронных сетей и методов активного обучения.

Предложенная методология включает анализ и оценку существующих подходов к автоматической разметке. Эффективность предложенных решений оценена на общедоступных наборах данных. Результаты показали, что предложенный метод в значительной мере сокращает время, необходимое для разметки данных, но в любом случае требует вмешательства оператора-разметчика.

Обзор литературы включает анализ современных методов разметки и существующих автоматических систем, что позволяет лучше понять контекст и преимущества предлагаемого подхода. В заключении обсуждены достижения, ограничения и возможные направления для будущих исследований в данной области.

Ключевые слова: компьютерное зрение, машинное обучение, автоматическая разметка данных, обучающая выборка, сегментация изображений.

ВВЕДЕНИЕ

С быстрым развитием технологий в области искусственного интеллекта и машинного обучения компьютерное зрение стало одной из самых активных и быстрорастущих областей исследований. Эта область включает в себя методы и технологии для автоматического извлечения, анализа и интерпретации значимой информации из изображений и видео. Основными задачами компьютерного зрения являются классификация, сегментация и обнаружение объектов.

Одним из ключевых аспектов успешного обучения моделей глубокого обучения является наличие большого объема качественно размеченных данных. Разметка данных – критически важный этап в создании высококачественных обучающих выборок. Однако процесс разметки является трудоемким и требует значительных ресурсов, что служит основным препятствием для быстрой разработки новых моделей и приложений. Согласно исследованию, на разметку крупных наборов данных уходит до 80% времени всего проекта по разработке модели машинного обучения [1].

Для облегчения процесса разметки используют различные инструменты и библиотеки, такие как LabelImg и VGG Image Annotator (VIA). LabelImg привлекает своей простотой и доступностью [2, 3], поддерживает популярные форматы аннотаций, такие как Pascal VOC, и обеспечивает интуитивно понятный интерфейс для пользователей [4]. Однако, несмотря на его достоинства, LabelImg не предоставляет продвинутых возможностей автоматизации и интеграции. VIA – инструмент с открытым исходным кодом и поддержкой различных типов аннотаций (для аннотирования доступны ограничивающие рамки, ломаные линии, многоугольники, круги, овалы, точки и эллипсы). Работать с этим инструментом можно через веб-браузер, что делает его удобным для удаленной работы и совместного использования.

В последние годы были разработаны различные методы автоматической разметки данных, которые могут значительно сократить время и усилия, необходимые для этого процесса. Одним из таких методов является использование сверточных нейронных сетей (Convolutional Neural Network, CNN), которые показали высокую эффективность в задачах классификации и сегментации изображений.

Активное обучение, другой важный метод, также используется для автоматической разметки данных. Активное обучение позволяет модели выбирать наиболее информативные примеры для разметки, что значительно улучшает эффективность процесса разметки и снижает затраты на разметку больших объемов данных. Установлено, что этот метод при разметке данных на видео снижает рабочую нагрузку на человека-оператора примерно на 78%, что значительно превосходит полную ручную разметку [5].

В настоящей работе проведены результаты исследования методов автоматической разметки обучающих выборок для задач компьютерного зрения с использованием передовых методов машинного обучения. Рассмотрены основные подходы к автоматической разметке данных, а также представлены результаты экспериментов по оценке эффективности предложенных методов.

ОБЗОР ЛИТЕРАТУРЫ

Современные методы автоматической разметки данных в задачах компьютерного зрения демонстрируют значительный прогресс благодаря использованию сверточных нейронных сетей (CNN) и активного обучения. В работе [6] рассмотрены различные инструменты и методы разметки изображений, систематизированы существующие программные решения и оценены по таким параметрам, как степень автоматизации и удобство пользовательского интерфейса. Большинство современных инструментов стремится минимизировать ручной труд и повысить точность разметки за счет внедрения методов машинного обучения и активного обучения.

В исследовании [7] описан один из методов автоматической аннотации изображений, основанный на сверточных нейронных сетях и алгоритме оптимизации порога (threshold optimization, THOP), который значительно улучшает

точность разметки по сравнению с традиционными методами и демонстрирует на 11% более высокую точность на нескольких наборах данных. В названном методе CNN превосходит традиционные методы по извлечению признаков и точности аннотации.

Модель автоматической аннотации изображений на основе модели внимания и равновесия данных, описанная в обзоре [8], использует глубокие нейронные сети для извлечения признаков и автоэнкодер для балансировки словаря аннотированных ключевых слов. Эксперименты показали, что предложенная модель превосходит существующие модели по критериям производительности на двух контрольных наборах данных.

Обзор [9] современных методов сегментации изображений на основе глубокого обучения подчеркивает важность сверточных нейронных сетей, таких как U-Net и Fully Convolutional Networks (FCN). Эти архитектуры показали высокую производительность в задачах сегментации естественных и медицинских изображений.

Исследование [10] также представляет интерактивные системы самоаннотации для детекции объектов на видео, сочетающие автоматические и интерактивные процессы, что позволяет ускорить аннотацию и снизить затраты на ручную работу. Эксперименты показали, что такие системы могут эффективно аннотировать видео и снижать затраты времени на аннотацию без потери качества.

Кроме того, статьи [11, 12] описывают модель Segment Anything (SAM), которая использует подход "promptable segmentation", позволяя модели эффективно справляться с задачами сегментации при различных входных данных, таких как точки, боксы или текст. Это значительно улучшает гибкость и применимость модели для автоматической аннотации данных в различных сценариях.

МЕТОДОЛОГИЯ

Для оценки различных методов автоматической разметки данных мы рассмотрели несколько современных подходов, включая использование мо-

делей сверточных нейронных сетей, активного обучения, интерактивной аннотации, а также методов, реализованных в модели YOLOv8 и в возможностях платформы Roboflow [13]. Основное внимание было уделено следующим аспектам: эффективность, точность на различных наборах данных и применимость в различных ситуациях.

Эффективность. Оценено время, необходимое для разметки одного изображения. Оценка включает измерение средней продолжительности аннотации для каждого метода (рис. 1).

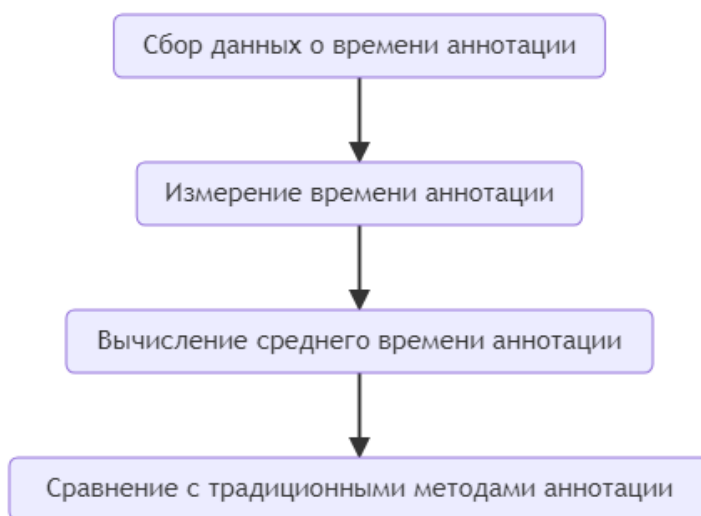


Рис. 1. Алгоритм оценки эффективности отдельно взятого метода

Точность. Определяется средней точностью (mean average precision, mAP) на стандартных наборах данных. Оценивалась на различных уровнях точности, таких как AP50 и AP75 (формулы (1) и (2) ниже), средняя точность при уровне уверенности 50% и 75% соответственно (рис. 2).

$$AP_{50} = 1/N \sum_{i=1}^N 1(IoU_i \geq 0.50), (1)$$

где (AP_{50}) – средняя точность при пороге IoU 50%; (N) – общее количество предсказаний; $(1(\cdot))$ – индикаторная функция, принимающая значение 1, если условие выполняется, и 0 – если нет; (IoU_i) – метрика степени пересечения между двумя ограничивающими рамками для (i) -го предсказания.

$$AP_{75} = 1/N \sum_{i=1}^N 1(IoU_i \geq 0.75), (2)$$

где (AP_{75}) – средняя точность при пороге IoU 75%; (N) – общее количество предсказаний; $(1(\cdot))$ – индикаторная функция, принимающая значение 1, если условие выполняется, и 0 – если нет; (IoU_i) – метрика степени пересечения между двумя ограничивающими рамками для (i) -го предсказания.

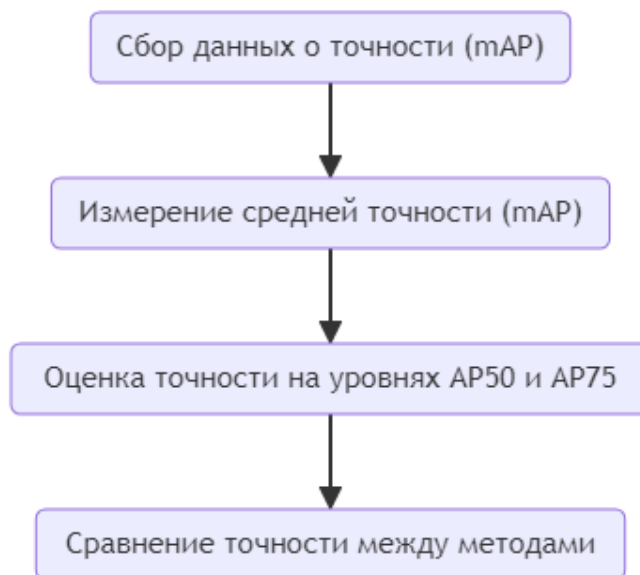


Рис. 2. Алгоритм оценки точности отдельно взятого метода

Применимость. Оценивалась способность метода адаптироваться к различным задачам и сценариям, включая аннотацию в реальном времени, медицинскую аннотацию и сложные сцены (рис. 3).

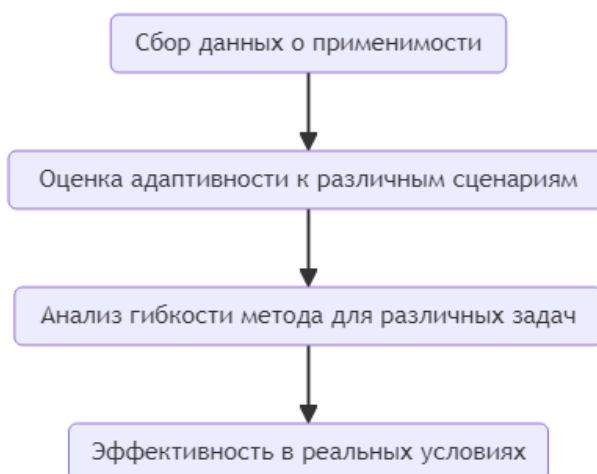


Рис. 3. Алгоритм оценки применимости отдельно взятого метода

РЕЗУЛЬТАТЫ И ВЫВОДЫ

Анализ показал, что различные методы автоматической разметки данных имеют свои уникальные преимущества и ограничения (таблица 1). Для оценки были использованы стандартные наборы данных COCO [14] и Cityscapes [15], которые позволяют объективно сравнить производительность методов в различных сценариях. Основные метрики включают эффективность (время аннотации одного изображения), точность (средняя точность, mAP) и применимость в реальных задачах.

Таблица 1. Результаты сравнения подходов к автоматической разметке

Метод	Эффективность (время/изобр.)	Точность (mAP, COCO)	Точность (mAP, Cityscapes)	Применимость
Mask RCNN	1–2 секунды	37,1%	57,58%	Городские сцены, медицинские изображения
Интер-активная аннотация	Варьируется (в среднем 5 секунд)	45,0%	60,0%	Сложные задачи, требующие ручной доработки
YOLOv8	<1 секунды	50,2%	63,0%	Обнаружение объектов в реальном времени
Roboflow Auto Label [16]	0,5–1 секунд	55,6%	64,2%	Медицинская аннотация, автономные транспортные средства

В результате анализа были получены следующие результаты. Mask RCNN отличается высокой точностью в сегментации сложных сцен и медицинских изображений, однако время аннотации может быть выше по сравнению с другими методами. Показатель mAP для COCO составляет 37,1%, а для Cityscapes – 57,58%.

Интерактивная аннотация, в свою очередь, эффективно снижает затраты на ручную разметку и достигает высокой точности благодаря ручной корректировке. Среднее время аннотации повышается в связи с разметкой вручную и составляет около 5 секунд, точность (mAP) для COCO – 45,0%, для Cityscapes – 60,0%. Подходит для сложных задач, требующих ручной доработки.

Наилучшие результаты по скорости и точности для задач реального времени дает YOLOv8. Показатель mAP для COCO составляет 50,2%, а для Cityscapes – 63,0%. Время обработки одного изображения занимает менее 1 секунды.

Инструмент Roboflow Auto Label демонстрирует наибольшую гибкость, обеспечивая быстрое и точное аннотирование с возможностью ручной корректировки. Показатель mAP для COCO составляет 55,6%, а для Cityscapes – 64,2%. Время разметки одного изображения занимает 0,5–1 секунды. Указанный инструмент подходит для множества сценариев, включая медицинскую аннотацию и аннотацию для автономных транспортных средств.

ЗАКЛЮЧЕНИЕ

Использование современных методов автоматической разметки данных позволяет значительно сократить время и усилия, затрачиваемые на аннотацию изображений, обеспечивая при этом высокую точность. Наиболее эффективными и универсальными методами являются те, которые используют современные модели глубокого обучения и активное обучение. Метод Roboflow Auto Label выделяется как лучший вариант, предлагая высокую точность и гибкость благодаря использованию текстовых подсказок и предобученных моделей. Этот метод обеспечивает оптимальное сочетание автоматической разметки и возможности ручной корректировки, что делает его пригодным для широкого спектра задач в области компьютерного зрения.

СПИСОК ЛИТЕРАТУРЫ

1. *Council J.* Data challenges are halting AI projects, IBM executive says // The Wall Street Journal. 2019. Vol. 28.
2. *LabelImg for Image Annotation.*
URL: <https://visio.ai/computer-vision/labelimg-for-image-annotation/>.
3. *VGG Image Annotator.*
URL: https://www.robots.ox.ac.uk/~vgg/software/via/via_demo.html.
4. *Everingham M. et al.* The pascal visual object classes challenge: A retrospective // International Journal of Computer Vision. 2015. Vol. 111. P. 98–136.
5. *Berg A. et al.* Semi-automatic annotation of objects in visual-thermal video // Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019. <https://doi.org/10.1109/ICCVW.2019.00277>
6. *Sager C., Janiesch C., Zschech P.* A survey of image labelling for computer vision applications // Journal of Business Analytics. 2021. Vol. 4, No. 2. P. 91–110.
7. *Cao J., Zhao A., Zhang Z.* Automatic image annotation method based on a convolutional neural network with threshold optimization // Plos one. 2020. V. 15, No. 9. e0238956. <https://doi.org/10.1371/journal.pone.0238956>
8. *Vatani A., Ahvanooy M.T., Rahimi M.* An effective automatic image annotation model via attention model and data equilibrium // arXiv preprint arXiv:2001.10590. 2020.
9. *Gu Y. et al.* Automatic lung nodule detection using a 3D deep convolutional neural network combined with a multi-scale prediction strategy in chest CTs // Computers in Biology and Medicine. 2018. Vol. 103. P. 220–231.
10. *Levine S. et al.* Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection // The International Journal of Robotics Research. 2018. Vpl. 37, No. 4-5. P. 421–436.
11. *Kirillov A. et al.* Segment anything // Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023. P. 4015–4026.

12. Zou X. et al. Segment everything everywhere all at once // Advances in Neural Information Processing Systems. NIPS '23: Proceedings of the 37th International Conference on Neural Information Processing Systems. Article No. 868. P.19769–19782. <https://dl.acm.org/doi/10.5555/3666122.3666990>
 13. Ultralytics YOLOv8 Docs. URL: <https://docs.ultralytics.com/ru>.
 14. COCO Dataset. URL: <https://cocodataset.org/#home>.
 15. Cityscapes Dataset. URL: <https://www.cityscapes-dataset.com/>.
 16. Auto-Label. URL: <https://roboflow.com/auto-label>.
-

AUTOMATIC ANNOTATION OF TRAINING DATASETS IN COMPUTER VISION USING MACHINE LEARNING METHODS

A.K. Zhuravlyov¹ [0009-0001-1265-9916], **K.A. Grigorian**² [0000-0001-6470-1832]

^{1, 2}*Kazan (Volga Region) Federal University, 35 Kremlevskaya str., Kazan, 420008*

¹UnMelow@yandex.ru, ²karigri@yandex.ru

Abstract

This paper addresses the issue of automatic annotation of training datasets in the field of computer vision using machine learning methods. Data annotation is a key stage in the development and training of deep learning models, yet the process of creating labeled data often requires significant time and labor. This paper proposes a mechanism for automatic annotation based on the use of convolutional neural networks (CNN) and active learning methods.

The proposed methodology includes the analysis and evaluation of existing approaches to automatic annotation. The effectiveness of the proposed solutions is assessed on publicly available datasets. The results demonstrate that the proposed method significantly reduces the time required for data annotation, although operator intervention is still necessary.

The literature review includes an analysis of modern annotation methods and existing automatic systems, providing a better understanding of the context and advantages of the proposed approach. The conclusion discusses achievements, limitations, and possible directions for future research in this field.

Keywords: *computer vision, machine learning, automatic data annotation, training datasets, image segmentation.*

REFERENCES

1. Council J. Data challenges are halting AI projects, IBM executive says // The Wall Street Journal. 2019. Vol. 28.
2. Labelling for Image Annotation.
URL: <https://viso.ai/computer-vision/labelimg-for-image-annotation/>.
3. VGG Image Annotator.
URL: https://www.robots.ox.ac.uk/~vgg/software/via/via_demo.html.
4. Everingham M. et al. The pascal visual object classes challenge: A retrospective // International Journal of Computer Vision. 2015. Vol. 111. P. 98–136.
5. Berg A. et al. Semi-automatic annotation of objects in visual-thermal video // Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019. <https://doi.org/10.1109/ICCVW.2019.00277>
6. Sager C., Janiesch C., Zschech P. A survey of image labelling for computer vision applications // Journal of Business Analytics. 2021. Vol. 4, No. 2. P. 91–110.
7. Cao J., Zhao A., Zhang Z. Automatic image annotation method based on a convolutional neural network with threshold optimization // Plos one. 2020. V. 15, No. 9. e0238956. <https://doi.org/10.1371/journal.pone.0238956>
8. Vatani A., Ahvanooy M.T., Rahimi M. An effective automatic image annotation model via attention model and data equilibrium // arXiv preprint arXiv:2001.10590. 2020.
9. Gu Y. et al. Automatic lung nodule detection using a 3D deep convolutional neural network combined with a multi-scale prediction strategy in chest CTs // Computers in Biology and Medicine. 2018. Vol. 103. P. 220–231.
10. Levine S. et al. Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection // The International Journal of Robotics Research. 2018. Vpl. 37, No. 4-5. P. 421–436.
11. Kirillov A. et al. Segment anything // Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023. P. 4015–4026.

12. Zou X. et al. Segment everything everywhere all at once // Advances in Neural Information Processing Systems. NIPS '23: Proceedings of the 37th International Conference on Neural Information Processing Systems. Article No. 868. P.19769–19782. <https://dl.acm.org/doi/10.5555/3666122.3666990>
 13. Ultralytics YOLOv8 Docs. URL: <https://docs.ultralytics.com/ru>.
 14. COCO Dataset. URL: <https://cocodataset.org/#home>.
 15. Cityscapes Dataset. URL: <https://www.cityscapes-dataset.com/>.
 16. Auto-Label. URL: <https://roboflow.com/auto-label>.
-

СВЕДЕНИЯ ОБ АВТОРАХ



ЖУРАВЛЁВ Алексей Константинович – выпускник бакалавриата Института информационных технологий и интеллектуальных систем Казанского (Приволжского) федерального университета, г. Казань.

Aleksey Konstantinovich ZHURAVLEV – graduate student of the Institute of Information Technologies and Intelligent Systems, Kazan (Volga region) Federal University, Kazan.

Email: UnMelow@yandex.ru

ORCID: 0009-0001-1265-9916



ГРИГОРЯН Карен Альбертович – кандидат экономических наук, доцент, Казанский (Приволжский) федеральный университет, г. Казань.

Karen Albertovich GRIGORIAN – Candidate of Economics, Associate Professor, Kazan (Volga region) Federal University, Kazan.

Email: karigri@yandex.ru

ORCID: 0000-0001-6470-1832

Материал поступил в редакцию 25 сентября 2024 года
