

Применение синтетических данных, полученных с помощью генеративной нейросети, для повышения качества моделей детекции.

Andrei Filatov^{*1, 3}, Daniil Dorin^{*2}, Nikita Barinov^{2, 3}, Uliana Izmesteva², Igor Ignashin², Ilya Stepanov², Viacheslav Vasilev^{2, 3}, Maxim Kurkin^{1, 3}, Dmitry Yudin^{2, 3}, Aibek Alanov^{3, 5}, Sergey Zagoruyko^{1, 4}, Denis Dimitrov³, Andrey Kuznetsov³

¹Center for Artificial Intelligence Technology, ²MIPT, ³AIRI, ⁴MTS AI, ⁵HSE University

Correspondence: filatovandrev@gmail.com

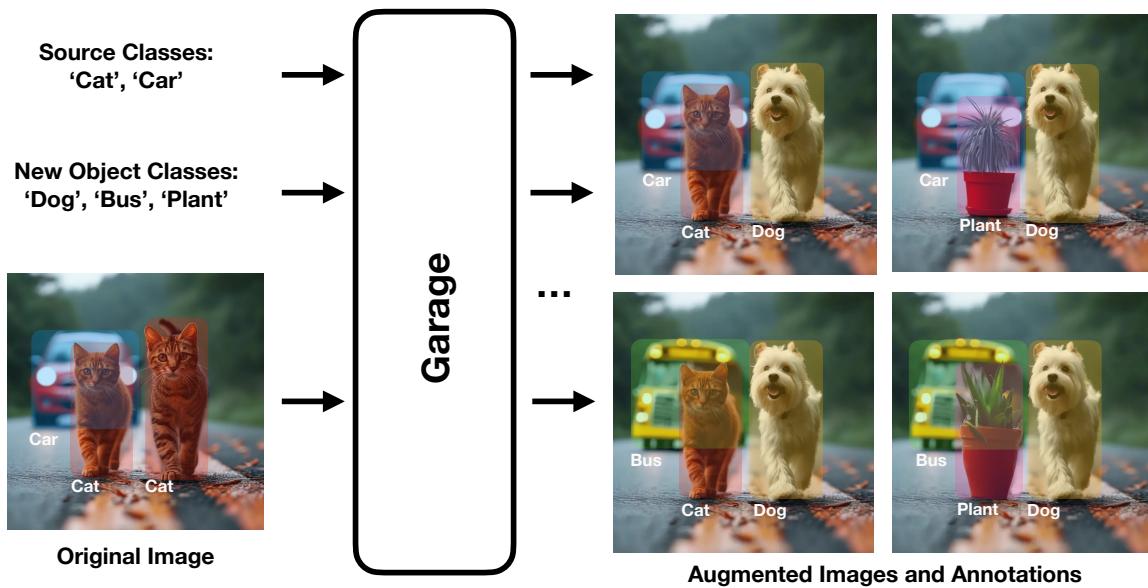


Рис. 1: Исходное изображение (слева) изображает двух кошек, идущих по дороге, с машиной на заднем плане. Вариации включают замену кошки на растение, кошки на собаку и машины на автобус. Данные аугментации показывают, как генеративные модели могут создавать разнообразные примеры для обучения, изменения различные элементы исходного изображения, что способствует повышению устойчивости моделей компьютерного зрения.

Abstract

Аугментация данных играет ключевую роль в улучшении производительности моделей машинного обучения, особенно в области компьютерного зрения. Традиционные методы, такие как повороты, сдвиги и регулировка яркости, ограничены в своей способности обеспечивать значительные семантические вариации, что часто приводит к плохому обобщению на новые данные.

В данной статье мы представляем новый подход, который позволяет заменять объекты на изображениях. Используя Visual-Language Models, мы получаем описания изображений, а затем, используя языковые нейросети, определяем, что можно за-

менить на изображении и составляем расширенный запрос, с помощью которого можем генерировать модифицированные данные. Таким образом, мы увеличиваем количество данных с использованием осмысленных запросов.

Мы демонстрируем эффективность данного подхода, проводя бенчмаркинг моделей, обученных с аугментированными и без аугментированных данных, показывая улучшения в производительности моделей детекции.

1 Введение

Аугментация данных — важный инструмент в арсенале современных исследователей в области машинного обучения и компьютерного

зрения. Она позволяет увеличить объём данных, улучшая общую производительность моделей путём создания различных вариаций. Однако традиционные методы аугментации, такие как повороты, сдвиги и изменения яркости, ограничены в своих возможностях. Они не обеспечивают значительных семантических расширений данных, что могло бы существенно улучшить обучение моделей. Например, при обучении модели детекции стандартные методы аугментации, такие как повороты и масштабирование, не создают достаточного разнообразия объектов на изображениях. Это может привести к тому, что модели не будут хорошо обобщаться на новые данные.

В этой работе мы предлагаем новую модель для аугментации данных, которая позволяет заменять объекты на изображениях. Этот подход семантически увеличивает объём данных, что важно для улучшения способности моделей машинного обучения к обобщению.

Наши вклады следующие:

- Мы представляем модель для аугментации данных, предназначенную для замены объектов, что обеспечивает значительное семантическое обогащение наборов данных.
- Мы демонстрируем экспериментальные результаты, которые показывают улучшенные способности к обобщению моделей, обученных на наших дополненных данных.
- Мы предлагаем открытое исходное решение, позволяя исследовательскому сообществу воспользоваться нашими достижениями в области аугментации данных.

2 Архитектура

2.1 Интерактивная аугментация

Рабочий процесс интерактивной аугментации следующий:

- Пользователь загружает изображение и выбирает объект для замены, используя текстовый запрос для описания объекта, который передается в модель

GroundedSAM (Ren et al., 2024) для извлечения маски объекта.

- Пользователь указывает класс, на который будет заменен исходный объект.
- Наша модель расширяет предоставленный запрос с помощью LLaMA (Dubey et al., 2024) и генерирует дополненное изображение с замененным объектом с использованием PowerPoint (Zhuang et al., 2023).

2.2 Автоматическая аугментация

Помимо ручной аугментации, наша система предлагает полностью автоматический режим.

После загрузки исходного набора данных пользователи могут указать, какие классы нужно добавить или дополнить. Например, если пользователь хочет добавить класс "калибра", он может указать это, и класс будет включен в набор данных. Затем система автоматически генерирует необходимые аугментации, в результате чего набор данных расширяется. Эта комплексная система состоит из трех этапов, описанных ниже.

2.3 Выбор объекта

Первый этап заключается в выборе объекта для аугментации. Для автоматического выбора мы используем LLaVA (Liu et al., 2024) для идентификации объектов, которые могут быть заменены, и предоставления описания изображения. LLaVA извлекает список объектов с изображения, из которого случайным образом выбирается объект для аугментации. Затем мы применяем GroundedSAM (Kirillov et al., 2023) для генерации маски объекта на основе выбранного объекта. Для подробностей о запросах LLaVA, пожалуйста, смотрите Приложение A.

2.4 Аугментация

Когда объект для аугментации выбран, используется LLaMA (Dubey et al., 2024) для генерации подходящей замены. Пользователь может предоставить список потенциальных объектов для замены, из которого модель сделает окончательный выбор. Если список не предоставлен, объект будет выбран автоматически. Кроме того, модель получает

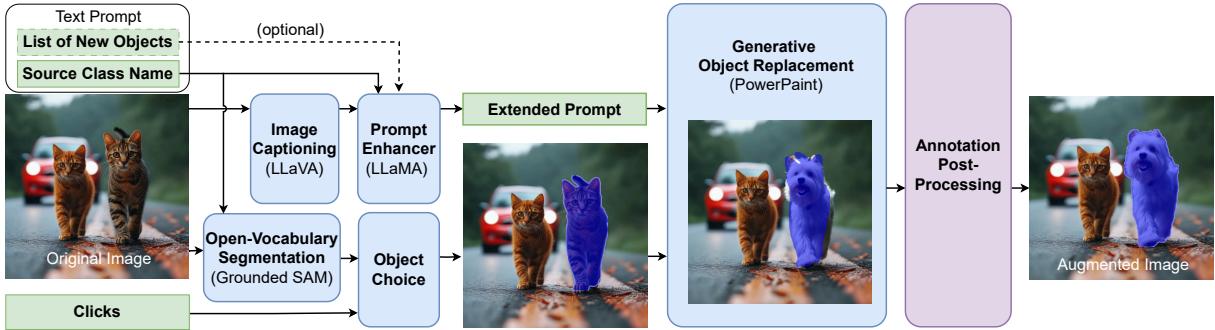


Рис. 2:

Детали архитектуры. На рисунке показаны последовательные этапы работы системы. Он включает в себя: Выбор объекта: Выбор объекта на основе пользовательских запросов. Аугментация: Генерация аугментации с дополнительными пользовательскими запросами или автоматическим выбором. Постобработка: Удаление артефактов генерации и фильтрация некорректных примеров. Каждый этап важен для получения качественных результатов.

полное описание изображения от LLaVA, чтобы гарантировать, что LLaMA полностью понимает контекст.

Для генерации замены мы применяем технологию встраивания объектов PowerPoint (Zhuang et al., 2023) с выбранным запросом. В ходе исследования мы выявили проблему: создание объекта с коротким запросом обычно дает неудовлетворительные результаты. Поэтому мы расширяем запрос с использованием LLaMA для предоставленного запроса. В конце процедуры расширенный запрос передается в PowerPoint для получения аугментированного изображения.

2.5 Последующая обработка

Мы применяем двухэтапную постобработку. На первом этапе мы используем Alpha-CLIP (Sun et al., 2024) для фильтрации некачественной генерации. Alpha-CLIP действует как обычный CLIP (Radford et al., 2021), но принимает маску, что позволяет вычислять сходство с конкретной областью изображения. Если значение CLIP превышает определенный порог, мы принимаем сгенерированную аугментацию. В противном случае мы генерируем изображение заново. После того как изображение прошло оценку CLIP, мы применяем SAM для получения более качественной генерации маски объекта. Это необходимо, потому что встраивание часто генерирует новый объект, не совсем точно соответствующий аннотации, поэтому нужно применить коррекцию для получения пра-

вильной маски. Несоответствие между объектом и маской можно увидеть на Рисунке 2.

3 Эксперименты

Для проверки эффективности нашего подхода к аугментации мы провели эксперименты с аугментациями на наборе данных VOC (Everingham et al., 2010).

3.1 Добавление нового класса

Один из возможных вариантов применения нашего метода – генерация данных для класса, который отсутствует или представлен ограниченно в существующем наборе данных.

Для проверки данного применения мы провели следующий эксперимент. Из набора данных VOC мы удалили данные с определённым классом, например, классом "кошка". Затем мы выполнили аугментацию данных на оставшихся данных, чтобы добавить отсутствующий класс в выборку. Во время процесса аугментации предложенный алгоритм применялся ко всем изображениям в наборе данных, на которых не был представлен этот класс.

(Вот тут вообще не понял что имелось ввиду) Для изображений с несколькими объектами случайным образом выбирался один, чей Bounding Boxes находился в пределах относительной площади не более 0.5, если такие объекты были доступны. Это было сде-

лано для того, чтобы избежать наложения маленьких объектов в процессе генерации.

Затем мы обучили модели детекции (FasterRCNN (Ren et al., 2015), DETR (Carion et al., 2020), YOLOv10-N (Jocher et al., 2023)) на данных с различными аугментациями отсутствующего класса, чтобы исследовать влияние наших аугментированных данных на результаты. Для обучения мы использовали стандартные скрипты из MMDetection (Chen et al., 2019) и Ultralytics (Jocher et al., 2023). Результаты представлены в таблице 1. Из наших экспериментов видно, что использование нашего метода улучшает качество детекции отсутствующего класса при сохранении общего качества модели на том же уровне. Важно отметить, что качество детекции для других классов не изменилось.

3.2 Улучшение качества подсказок

При использовании стандартных подсказок мы заметили снижение визуального качества генерируемых изображений, особенно при более коротких подсказках, таких как простые метки классов. Чтобы решить эту проблему, мы использовали расширенные подсказки, воспользовавшись лингвистическими возможностями модели LLaMA. Например, простая инструкция, такая как cat (кошка), может быть расширена в более описательную фразу, например, The ginger tabby cat has a sleek body, pointy ears, and curious green eyes gazing around from its playful pose (Имбирная полосатая кошка с изящным телом, острыми ушами и любопытными зелёными глазами, оглядывающимися вокруг с игривой позы). Это расширение вводит важную контекстную информацию, эффективно сокращая разрыв между текстовыми и визуальными представлениями.

Чтобы оценить важность расширения подсказок при обучении модели, мы провели эксперимент, создав два набора данных с идентичными изображениями: один с использованием только базовой подсказки, а другой — с расширенной подсказкой, включающей текстовое описание изображения. Этот подход позволил нам оценить, как расширенные подсказки могут улучшить качество генерируемых изображений. Расширенные инструкции предоставили последней модели PowerPoint подробную информацию и

ниюансы, что позволило создавать изображения более высокого качества, разнообразные и визуально последовательные, соответствующие инструкции.

Как уже упоминалось, генерация объекта с использованием короткой инструкции обычно приводит к низкому качеству. Мы провели дополнительный эксперимент, чтобы изучить, как расширение подсказки влияет на обучение модели детекции объектов. Мы обучили модель FasterRCNN на наборе данных, где не было изображений кошек, используя классическую подсказку и расширенную подсказку. Результаты показаны в таблице 3. Из результатов видно, что расширенная подсказка улучшает качество аугментированных данных и, следовательно, модель детекции.

4 Смежные работы

Аугментация данных – это широко используемая техника в области машинного обучения и компьютерного зрения для увеличения разнообразия и объема обучающих наборов данных. Традиционные методы аугментации данных, такие как повороты, обрезка и изменение яркости цветов, стали основой для повышения устойчивости моделей (Buslaev et al., 2020). Например, (Krizhevsky et al., 2012) продемонстрировал эффективность этих техник в своей работе по задаче классификации на наборе данных ImageNet. Однако эти методы часто не способны повысить качество сложных моделей.

Более современные подходы используют генеративные модели в качестве источников для создания аугментаций. (Alimisis et al., 2024) исследует использование передовых генеративных моделей для создания разнообразных и реалистичных аугментаций. (Yin et al., 2023) использовали модели "text-to-text" и "text-to-image" для генерации аугментаций для задач классификации изображений, продемонстрировав значительные улучшения в производительности модели. (Fang et al., 2024) использовали адаптер ControlNet для генерации аугментаций, что повысило вариативность и устойчивость набора данных. (Kipun and Rupprecht, 2024) применили методы InPaint для аугментации данных, сохраняя при этом те же аннотации меток.

Таблица 1: Результаты детекции объектов на наборе данных Pascal VOC. Результаты показывают, что обучение на наших данных значительно улучшает производительность модели при различных процентах данных, что подтверждается более высокими значениями mAP в строках "наши" по сравнению с "оригинальные". Проценты представляют собой долю изображений с кошками, использованных из оригинального набора данных.:

| Датасет | Модель | Данные | 0% | 25% | 50% | 75% | 100% |
|------------|-------------|--------------|------|------|------|------|------|
| Pascal VOC | DETR | оригинальные | 0.0 | 57.5 | 61.5 | 65.4 | 69.1 |
| | | наши | 5.3 | 55.6 | 62.2 | 66.3 | 70.3 |
| | YOLOv10-N | оригинальные | 0.0 | 50.9 | 54.8 | 56.9 | 60.4 |
| | | наши | 31.5 | 51.3 | 53.6 | 57.9 | 61.3 |
| | Faster RCNN | оригинальные | 0.0 | 74.5 | 77.4 | 75.5 | 76.6 |
| | | наши | 66.4 | 77.3 | 80.1 | 83.5 | 84.0 |

| АР для категории "кошка" | |
|--------------------------|-----------------------|
| Подсказка | Расширенная подсказка |
| 64.6 | 66.4 |

Таблица 2: Сравнение качества обучения модели FasterRCNN в зависимости от расширения инструкции. Расширение подсказки улучшает качество модели. Результаты показывают, что использование расширенных подсказок значительно повышает среднюю точность (AP) для категории "кошка" по сравнению с использованием подсказок без расширения.

Альтернативой аугментациям, которые улучшают семантику, является Open Vocabulary Detection (Wu et al., 2024), где модель учится обнаруживать объекты с использованием языковой модели. Хотя этот подход эффективен, он ограничен словарем объектов, на котором модель была обучена, и не позволяет увеличивать разнообразие набора данных через уточнение свойств обнаруживаемых объектов. Более того, создание таких архитектур требует значительных ресурсов, что может стать существенным ограничением при обучении эффективной модели.

(Вот тут ощущение, что вода, не очень понимаю почему остальные модели улучшают существующие а наша типо генерит новые) Наша работа отличается от предыдущих подходов к генеративной аугментации тем, что она позволяет не просто улучшать существующие, а семантически и структурно дополнять набор данных, предлагая более комплексную стратегию аугментации.

5 Вывод / Ограничения

В данной статье мы представили новый подход, предназначенный для решения задач, связанных с аугментацией данных в машинном обучении и компьютерном зрении.

Мы продемонстрировали эффективность нашей нейросети через всесторонние эксперименты и бенчмаркинг. Результаты показали, что модели, обученные на данных, аугментированных с помощью нашего подхода, демонстрируют улучшения.

Одним из ограничений нашего подхода является потребность в вычислительных ресурсах, включая необходимость мощных графических процессоров и времени для генерации каждого аугментированного изображения. Однако это ограничение можно смягчить, используя дистиллированные модели, такие как SD3 Turbo (Sauer et al., 2024), которые являются оптимизированными версиями исходных моделей. Дистиллированные модели могут выполнять те же задачи более эффективно, снижая вычислительную нагрузку и время, необходимое для генерации изображений, что делает платформу Garage более практичной и масштабируемой для реального использования.

В заключение, наша модель представляет собой значительный прорыв в области аугментации данных, устранив ключевые ограничения существующих методов и предоставив комплексное, автоматизированное решение для создания высококачественных, разнообразных и сбалансированных обучающих наборов данных. Мы уверены, что представленная нейросеть будет способствовать разработке более устойчивых и качественных

моделей машинного обучения, что в конечном итоге приведет к прогрессу в данной области.

6 Этические вопросы

Это исследование представляет нейросеть для аугментации данных, поднимая несколько важных этических вопросов. Обеспечение конфиденциальности данных является ключевым, особенно при работе с наборами данных, содержащими личную информацию. Необходимы надежные методы анонимизации и соблюдение регламентов защиты данных, таких как GDPR, чтобы защитить личности. Кроме того, методы аугментации данных должны контролироваться, чтобы предотвратить введение или усиление предвзятости, связанной с расой, полом и социально-экономическим статусом. Следует прилагать усилия для выявления и смягчения этих предвзятостей, чтобы гарантировать разработку справедливых и беспристрастных моделей машинного обучения.

Способность заменять объекты и генерировать новые изображения представляет риск злоупотреблений, таких как создание вводящего в заблуждение или обманчивого контента. Необходимы защитные меры и этические рекомендации для предотвращения вредоносных применений представленной модели. Прозрачность также играет важную роль; исследователи должны предоставлять четкую документацию методов, открыто делиться кодом и наборами данных, а также признавать ограничения и возможные этические проблемы. Важным аспектом является также учет воздействия на окружающую среду вычислительных ресурсов, используемых при обучении моделей, с акцентом на энергоэффективные алгоритмы. Наконец, несмотря на достижения в области автоматизации, человеческий контроль остается важным для поддержания этических стандартов и контроля качества, обеспечивая ответственное использование и развитие нашей модели.

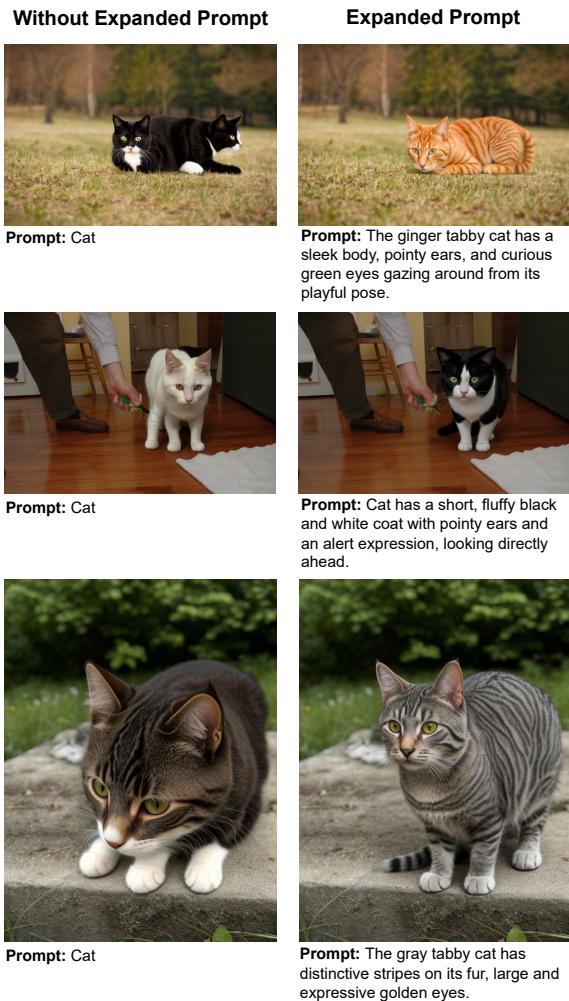


Рис. 3: Сравнение генерации в зависимости от расширения подсказки. Изображения слева, сгенерированные с использованием расширенных подсказок, демонстрируют значительно большее количество деталей и точности по сравнению с изображениями справа, сгенерированными с минимальными подсказками. Это показывает, как наша модель работает лучше с подробными описаниями, более эффективно захватывая нюансы и особенности желаемого результата.

Литература

References

- Panagiotis Alimisis, Ioannis Mademlis, Panagiotis Radoglou-Grammatikis, Panagiotis Sarigiannidis, and Georgios Th Papadopoulos. 2024. Advances in diffusion models for image data augmentation: A review of methods, models, evaluation metrics and future research directions. arXiv preprint arXiv:2407.04103.
- Alexander Buslaev, Vladimir I Iglovikov, Eugene Khvedchenya, Alex Parinov, Mikhail Druzhinin, and Alexandr A Kalinin. 2020. Albumentations: fast and flexible image augmentations. *Information*, 11(2):125.
- Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. 2020. End-to-end object detection with transformers. In European conference on computer vision, pages 213–229. Springer.
- Kai Chen, Jiaqi Wang, Jiangmiao Pang, Yuhang Cao, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng, Ziwei Liu, Jiarui Xu, et al. 2019. Mmdetection: Open mmlab detection toolbox and benchmark. arXiv preprint arXiv:1906.07155.
- Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, et al. 2024. The llama 3 herd of models. arXiv preprint arXiv:2407.21783.
- Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. 2010. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88:303–338.
- Haoyang Fang, Boran Han, Shuai Zhang, Su Zhou, Cuixiong Hu, and Wen-Ming Ye. 2024. Data augmentation for object detection via controllable diffusion models. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pages 1257–1266.
- Glenn Jocher, Ayush Chaurasia, and Jing Qiu. 2023. Yolo by ultralytics.
- Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. 2023. Segment anything. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 4015–4026.
- Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.
- Orest Kupyn and Christian Rupprecht. 2024. Dataset enhancement with instance-level augmentations. arXiv preprint arXiv:2406.08249.
- Haotian Liu, Chunyuan Li, Yuheng Li, Bo Li, Yuanhan Zhang, Sheng Shen, and Yong Jae Lee. 2024. *Llava-next: Improved reasoning, ocr, and world knowledge*.
- Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning transferable visual models from natural language supervision. In Proceedings of the 38th International Conference on Machine Learning, volume 139 of Proceedings of Machine Learning Research, pages 8748–8763.
- Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28.
- Tianhe Ren, Shilong Liu, Ailing Zeng, Jing Lin, Kunchang Li, He Cao, Jiayu Chen, Xinyu Huang, Yukang Chen, Feng Yan, et al. 2024. Grounded sam: Assembling open-world models for diverse visual tasks. arXiv preprint arXiv:2401.14159.
- Axel Sauer, Frederic Boesel, Tim Dockhorn, Andreas Blattmann, Patrick Esser, and Robin Rombach. 2024. Fast high-resolution image synthesis with latent adversarial diffusion distillation. arXiv preprint arXiv:2403.12015.
- Zeyi Sun, Ye Fang, Tong Wu, Pan Zhang, Yuhang Zang, Shu Kong, Yuanjun Xiong, Dahua Lin, and Jiaqi Wang. 2024. Alpha-clip: A clip model focusing on wherever you want. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 13019–13029.
- Jianzong Wu, Xiangtai Li, Shilin Xu, Haobo Yuan, Henghui Ding, Yibo Yang, Xia Li, Jiangning Zhang, Yunhai Tong, Xudong Jiang, et al. 2024. Towards open vocabulary learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Yuwei Yin, Jean Kaddour, Xiang Zhang, Yixin Nie, Zhenguang Liu, Lingpeng Kong, and Qi Liu. 2023. Ttida: Controllable generative data augmentation via text-to-text and text-to-image models. arXiv preprint arXiv:2304.08821.
- Junhao Zhuang, Yanhong Zeng, Wenran Liu, Chun Yuan, and Kai Chen. 2023. A task is worth one word: Learning with task prompts for high-quality versatile image inpainting. arXiv preprint arXiv:2312.03594.

A Пример Appendix

A.1 Подсказки

A.1.1 LLaVA подсказки для описывания изображений

Подсказка предназначена для того, чтобы направить модель LLaVa на создание полного описания изображения. Основные элементы, которые должны быть включены в описание, — это детали объектов на изображении, их относительные позиции и количество. Этот структурированный подход нужен для того, чтобы сгенерированное описание было тщательным и информативным, предоставляя подробное понимание визуального контента.

USER: <image> Provide a detailed caption for this image. Include details about the relative position of objects in the picture and their number.

A.1.2 LLaMA подсказки для выбора нового объекта

Эта подсказка предназначена для того, чтобы помочь модели LLaMa найти подходящий заменяющий объект в заданной сцене. Нейросети предлагается предложить новый объект, который будет явно отличаться от существующего, основываясь только на предоставленном описании сцены и текущем объекте. Это гарантирует, что замена сохранит контекст и целостность сцены, одновременно добавляя разнообразие.

USER: Imagine you are an object replacer. Your task is generating a replacement object instead of the existing object on the scene. It's important that the new object is not the same as the existing one. I will give you a description of the scene and the existing object. You must give me an object which could be depicted instead of the existing object. So, image description: {ImageDescription}, existing object: {CurrentObject}, a list of potential new objects: {NewObjectsList}. You should select and return only the name of the new object from the provided list, which fits into the picture to replace the existing one.

Если подаётся список возможных объектов:

USER: Imagine you are an object replacer. Your task is generating a replacement object instead of the existing object on the scene. It's important that the new object is not the same as the existing one. I will give you a description of the scene, existing object, and a list of potential new objects. You must give me an object from the list of potential new objects which could be depicted instead of the existing object. The new object should fit well into the picture in place of the existing object. The new object should be approximately the same size as the existing object. If no object from the list fits into the picture, return the existing object. The image should remain believable after replacement. So, image description: {ImageDescription}, existing object: {CurrentObject}, a list of potential new objects: {NewObjectsList}. You should select and return only the name of the new object from the provided list, which fits into the picture to replace the existing one.

A.1.3 Подсказка для расширения с использованием LLaMA

USER: Imagine that you want to describe the {NewObject}'s appearance to an artist in one sentence, under 15 words. Mention {NewObject} in the description for clarity. Focus solely on the realistic description of the {NewObject}, ignoring any external elements or surroundings. For example, if the object is an animal, the description should include the animal's color, size, breed, pose, view direction, etc. If the object is a vehicle, the description should include the vehicle's brand or model, color, size, type, etc. If the object is a person, the description should include the person's age, gender, height, weight, hair color, eye color, clothing, pose, etc. Do not add anything extra to the visual description that is not directly related to {NewObject}.

A.2 Примеры генераций

Все 20 классов из набора данных PascalVOC были аугментированы на [Figure 5](#).

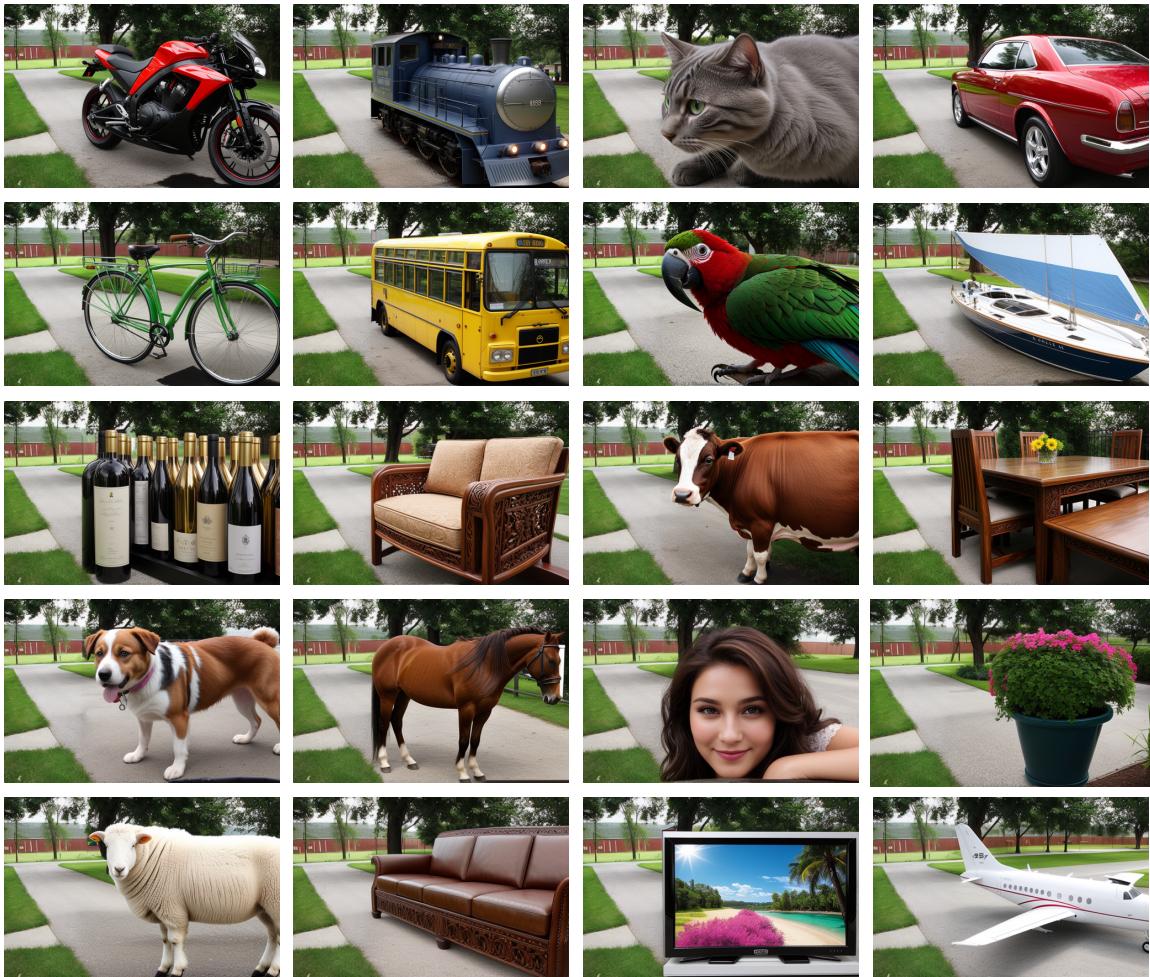


Рис. 4: Аугментации PascalVOC. Этот рисунок иллюстрирует различные классы объектов из набора данных PascalVOC, каждый из которых был аугментирован с использованием нашей платформы. Категории включают: (верхний ряд) мотоцикл, поезд, кошка, автомобиль; (второй ряд) велосипед, автобус, птица, лодка; (третий ряд) бутылка, стул, корова, обеденный стол; (четвертый ряд) собака, лошадь, человек, комнатное растение; (нижний ряд) овца, диван, телевизор и самолет. Каждый класс объектов изображен в реалистичной манере, что демонстрирует потенциал генеративных аугментаций для создания реалистичных и разнообразных экземпляров для обучения моделей машинного обучения.