ИНФОРМАТИКА И ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ / INFORMATICS AND INFORMATION PROCESSES

DOI: https://doi.org/10.23670/IRJ.2024.141.12

АВТОМАТИЧЕСКАЯ ГЕНЕРАЦИЯ ОБУЧАЮЩИХ НАБОРОВ ДАННЫХ С ИСПОЛЬЗОВАНИЕМ 3D-МОДЕЛЕЙ ДЛЯ ЗАДАЧИ ДЕТЕКТИРОВАНИЯ ОБЪЕКТОВ

Научная статья

Чичиков С.А.[1, *], Куликов В.Р.[2], Спирин Е.А.[3], Сарамуд М.В.[4]
[1] ORCID : 0000-0002-9223-1935;
[4] ORCID : 0000-0003-0344-9842;
[1, 2, 3, 4] Сибирский государственный университет науки и технологии имени академика М.Ф. Решетнева, Красноярск, Российская Федерация

* Корреспондирующий автор (chis1980[at]mail.ru)

Аннотация

В рамках данного исследования предлагается создавать набор данных путем визуализации изображений из 3D сцен в среде моделирования Blender. Основной особенностью данного подхода является возможность внесения случайных вариаций различных аспектов сцены: масштаб объектов, их положение в пространстве, характеристики освещения, параметры камеры, текстуры, фоновых изображений для получения более разнообразных и вариативных данных с помощью разработанной программы. В работе реализовано автоматическое формирование и сохранение аннотации для сгенерированных изображений в формате YOLO. Экспериментальная оценка показывает, что данный подход дает возможность эффективно обучать детекторы на синтетических данных.

Ключевые слова: нейронные сети, техническое зрение, обучение нейронной сети, synthetic data generation, blender, object detection, data synthesis.

AUTOMATIC GENERATION OF TRAINING DATA SETS USING 3D MODELS FOR OBJECT DETECTION PROBLEM

Research article

Chichikov S.A.[1, *], Kulikov V.R.[2], Spirin E.A.[3], Saramud M.V.[4]
[1] ORCID : 0000-0002-9223-1935;
[4] ORCID : 0000-0003-0344-9842;
[1, 2, 3, 4] Reshetnev Siberian State University of Science and Technology, Krasnoyarsk, Russian Federation

* Corresponding author (chis1980[at]mail.ru)

Abstract

This research proposes to create a dataset by rendering images from 3D scenes in the Blender modelling environment. The main characteristic of this approach is the possibility of introducing random variations of different aspects of the scene: scale of objects, their position in space, lighting characteristics, camera parameters, textures, background images to obtain more diverse and variable data using the developed programme. The work implements automatic generation and saving of annotation for the generated images in YOLO format. Experimental evaluation shows that this approach provides an efficient way to train detectors on synthetic data.

Keywords: neural networks, technical vision, neural network training, synthetic data generation, blender, object detection, data synthesis.

Введение

Детектирование объектов – это процесс, при котором компьютерная система автоматически находит и выделяет объекты определенного класса на визуальных данных [1], и позволяет компьютерам анализировать и интерпретировать визуальную информацию.

Глубокое обучение и методы машинного обучения сегодня позволяют получать хорошие результаты в задаче детектирования, но продолжаются исследования в направлении повышения точности, скорости и надежности детектирования в разнообразных сценариях [2].

Процесс формирования обучающей выборки и его аннотации является трудозатратным, требующим времени, ресурсов и усилий. Каждое изображение должно быть проанализировано и размечено человеком, для дальнейшего обучения нейросетевых моделей [3]. В процессе ручной аннотации всегда присутствует вероятность человеческих ошибок: неточности в определении границ, ошибки в классификации, которые снижают качество обучения моделей. В некоторых случаях реальных данных может быть недостаточно (или они не существуют) для эффективного обучения моделей.

Цель нашей работы – повышение скорости и вариативности создания обучающего набора данных, для задач детектирования объектов на изображениях, исследование возможностей использования 3D-моделей для создания синтетических обучающих данных и оценка предложенного метода

Рисунок 1 - Использование синтетических данных в обучении моделей
DOI: https://doi.org/10.23670/IRJ.2024.141.12.1

Синтетические данные значительно увеличивают объем обучающих данных, что улучшает обобщающую способности моделей. Они могут быть получены с разнообразными параметрами и условиями, что охватывает широкий спектр возможных сценариев и условий, в которых модель может столкнуться в реальном мире. Используя синтетические данные, можно точно контролировать параметры и характеристики сцен, что упрощает проведение экспериментов для анализа влияния различных факторов на работу модели.

В реальных данных редкие случаи могут быть плохо представлены, что усложняет обучение модели. Используя возможности моделирования сцены можно генерировать больше примеры редких случаев, что экономит время и ресурсы.

Для генерации синтетических обучающих выборок мы использовали бесплатный пакет Blender с открытым исходным кодом [4]. В Blender мы создаем трехмерные модели или импортируем готовые 3D модели из CAD систем, которые в дальнейшем используем для создания синтетического набора данных. В нем мы формируем сцену с различными параметрами освещения, фонами, камерами, материалами и текстурами. Мы можем настраивать цвета, отражения, прозрачность и другие атрибуты материалов. Для создания фотореалистичных синтетических изображений, используется встроенный инструмент визуализации.

Для управления генерацией синтетического набора используется встроенный в Blender язык программирования Python. Программный интерфейс приложения взаимодействует с объектами сцены и материалами. Используя эти возможности после создания изображений, мы автоматически вычисляем аннотации для объектов на изображениях, чтобы указать их положение, класс и другие характеристики и полученного изображения, и экспортируем в нужный формат.

Альтернативным вариантом генерации синтетических обучающих выборок на основе 3d моделей является использование игровых движков, например, Unreal Engine [5] и других. Но при использовании игровых движков мы отказываемся от фотореализма при создании синтетического набора данных. Показано, что после увеличения набора данных определенным количеством изображений происходит быстрое переобучение модели нейронной сети [6], что не позволит получить уверенное распознавание на реальных данных.

В работе [7] показано, что моделирование условий освещения и расположения объектов в синтетическом наборе данных существенно улучшает способность нейросети к обобщению и адаптации к разнообразным сценариям в реальном мире.

При использовании синтетического обучения важно качество графики и реализма, особенно если модель должна работать в реальных условиях. Это может потребовать улучшения методов генерации синтетических данных, чтобы обеспечить более высокую степень реализма [8], [9].

В работах [10], [11] экспериментально продемонстрировано, что модели, обученные в синтетической области, конкурируют с моделями, обученными с помощью изображений, составленных из смеси синтетических и реальных данных и обученными исключительно на реальных изображениях.

Создание виртуальной среды и сцен

Сбор изображений для набора данных с большой вариабельностью данных может быть затратным и трудоемким. При решении задач детектирования в системах технического зрения, например, при автоматизации сборочных процессов на производстве, часто встречается ситуация, когда различные детали и компоненты могут иметь множество вариаций и альтернатив, и в процессе производства детали могут заменяться другими версиями, что осложняет поддержку в актуальном состоянии набора данных. Синтетически сгенерированные изображения из 3D сцен могут существенно уменьшить трудозатраты на обучение модели нейронной сети, предоставив более разнообразные и контролируемые сценарии. Но если модель нейронной сети обучается на ограниченных и однотипных данных, она может переобучиться на этих данных, не обобщаясь на новые случаи [6]. Например, если модель нейросети видела объекты только с одного ракурса съемки, или на одном фоне, она может плохо справляться с задачей детекции, либо находить ложные объекты в фоне. Нейронная сеть начинает «запоминать» их особенности, не учитывая общие закономерности или вариации.

Добавление случайных изменений в сцену позволяет избежать этой проблемы, и адаптироваться к различным условиям и вариациям, что повышает способность нейронной сети детектировать и классифицировать в реальных сценах и обнаруживать объекты на основе общих признаков, а не конкретных деталей.

Мы вносим случайные изменения ко всем элементам сцены для каждого изображения синтетического набора данных, который используется во время процесса обучения нейронной сети. Случайным изменениям подвергаются следующие характеристики (рисунок 2):

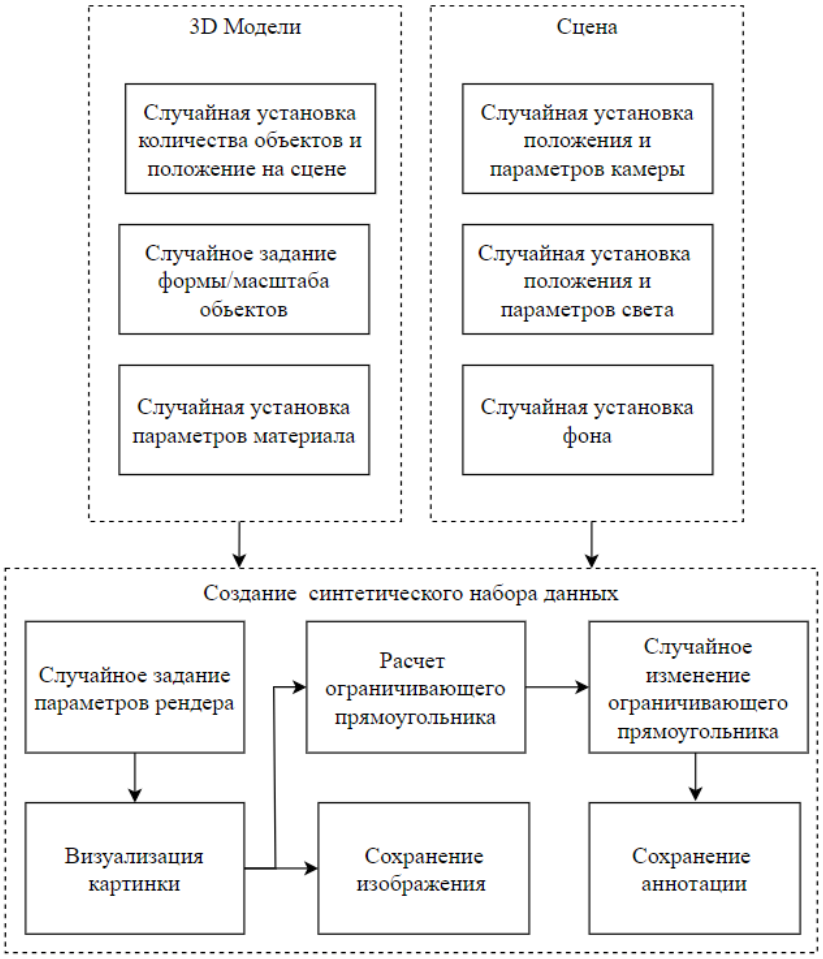

Рисунок 2 - Схема генерации синтетических данных
DOI: https://doi.org/10.23670/IRJ.2024.141.12.2

Мы варьировали количество, положение и ориентацию объектов относительно камеры. Это представляет в наборе данных различные комбинации для того, чтобы нейросеть могла обучаться на различных ситуациях, включая случаи, когда объекты перекрывают друг друга или их тени падают на другие элементы сцены. Количество объектов в каждой сцене в 50 процентах случаев на сцене генерировалось от 2 до 6, а в остальных случаях присутствовал только один объект. Такой подход позволяет нейросети учиться распознавать как отдельные объекты, так их комбинации, и взаимодействие между ними. Положение и ориентация объектов случайным образом менялись относительно камеры, включая перемещение по сцене в различных положениях и их вращение. Разнообразие в сцене позволяет модели нейросети учиться на более широком спектре сценариев и ситуаций. Нейросеть может научиться распознавать объекты как в устойчивых и изолированных условиях, так и в более сложных, когда они взаимодействуют друг с другом, что повышает обобщающую способность модели устойчиво детектировать объекты в реальном мире.

В процессе генерации изображений синтетической сцены применяются изменения в масштабе объектов, что обеспечивает разнообразие в размерах и помогает нейросети обучаться на разных масштабах представления. Размер каждого объекта в сцене варьируется на основе случайного числа, выбранного из диапазона от 0,5 до 3. Оно отображает масштабирования объекта по каждой оси. Затем к этому масштабному коэффициенту добавляем случайное изменение в пределах 10% от начального размера по каждой оси. Таким образом, форма немного меняется и варьируется. Внесение небольших изменений в форму производит больше вариативности в наборе данных.

Для внесения случайных вариаций в освещение сцены в процессе обучения нейронной сети было использовано 3 источника света. Положение каждого источника света генерировалось случайным образом над сценой. Для каждого источника света также изменялась интенсивность света случайным образом с диапазоном от 10 до 100 процентов. Такое изменение интенсивности формируют набор данных как в ярком, так и в тусклом освещении, как может быть в реальных условиях. Кроме того, направление света менялось в пределах ±45 градусов поворотом на случайный угол вокруг своей оси.

Апробация методики выполнялась на макетах реальных объектов, полученных на 3D принтере. Для фотореалистичности и учета технологии FDM-печати, в синтетических сценах был создан материал, имитирующий фактуру поверхности. В материале, который применяется при визуализации каждой сцены, варьируются различные характеристики: изменяется толщина и ширина слоя печати, чтобы передать детали и особенности процесса печати, что добавляет текстурные эффекты и создает эффект визуального объема. Также применяются случайные коррекции цвета, насыщенности, яркости, блеска и отражения материала, что создает вариативность в отображении материала и его внешних характеристик.

При реализации изменения фонового изображения и добавления разных окружений в сцену, был использован метод генерации фона с учетом максимизации фонового беспорядка. Целью этого подхода является использование непредсказуемого фонового изображения, чтобы модель нейросети избегала простых признаков и могла лучше учиться на более сложных аспектах и их взаимодействиях с окружением. Для максимального разнообразия и реализма фоновых изображений, в синтетическом наборе данных было подготовлено 500 изображений, представляющих текстуры дерева. Эти изображения представляют собой различные вариации текстур дерева с разными размерами и структурами. Во время генерации сцены каждый раз, случайным образом, выбирается одно из фоновых изображений из этого набора. Фоновые изображениям могут быть повернуты на угол от 0 до 90 градусов, масштабированы по размеру сцены, или изменена их яркость и контрастность до 20%.

Для разнообразия сцен в синтетических наборах данных мы проводили вариацию параметров камеры: фокусное расстояние и точку обзора. Мы варьировали фокусное расстояние на значения, отличающиеся на ±20% от начального значения, что формирует разные планы фокуса и глубину резкости в изображениях. Перемещая точку обзора, мы создаем вариацию в угле, под которым модель видит объект.

При генерации изображений было использовано разное количество сэмплов в настройках визуализации каждого изображения, изменяемое случайным образом в диапазоне от 15 до 100 сэмплов. Количество сэмплов оказывает влияние на качество и «шумность» результирующего изображения. При использовании меньшего количества сэмплов, например, 15, изображение может содержать больше «шума» и артефактов. В результате получаются менее четкие и более «шумные» изображения. При использовании большего количества сэмплов, например, 100, мы получаем более точную и менее «шумную» картинку.

Для дополнительного разнообразия в синтетических сценах мы также вводили случайное изменение размеров прямоугольника, охватывающего объект на изображении (bound box), изменяя размеры границ на величину от -5% до +5% от их исходного размера.

Каждое изображение в генерируемом синтетическом наборе формируется с помощью объединения двух компонентов: фонового слоя и объектов переднего плана, которые формируются из трехмерных моделей.

На рисунке 3 представлены некоторые изображения, которые созданы с использованием нашего алгоритма.

Рисунок 3 - Примеры визуализации сцен с аннотацией в обучающем наборе данных

DOI: https://doi.org/10.23670/IRJ.2024.141.12.3

Каждое изображение автоматически аннотируется набором ограничивающих рамок, состоящих из числовых координат, которые указывают на положение и размеры прямоугольника, охватывающего каждый объект на изображении. Для определения границ в системе координат камеры применяется процедура перспективной проекции сетки вершин на камеру. Путем нахождения минимальных и максимальных значений среди этих координат, мы определяем угловые точки ограничивающего прямоугольника. Этот прямоугольник визуально охватывает объект на изображении, предоставляя точные границы, определяя верхний левый угол прямоугольника (x, y) и его ширину и высоту (w, h). Для каждого объекта, который мы аннотируем, мы определяем координаты его границ на визуализированном изображении.

Аннотации сохраняются вместе с соответствующими сгенерированными изображениями в формате YOLO в текстовых файлах.

Для построения обучающего и проверочного набора данных было сгенерировано 10 0000 аннотированных изображений, каждое из которых представляет собой уникальную сцену.

Обучение модели детектирования на синтетических данных

Мы использовали архитектуру YOLOv8n для обучения нейронной сети на синтетических данных. YOLO (You Only Look Once) – это архитектура нейронной сети, разработанная для решения задачи детектирования объектов на изображениях и видео. Эта архитектура представляет собой значительное улучшение в сравнении с более ранними методами детектирования R-CNN и Faster R-CNN, и получила широкую популярность благодаря своей эффективности и скорости работы [12].

Нейросеть была обучена на протяжении 800 эпох, что обеспечивает полный цикл прохода через весь тренировочный набор данных. В процессе обучения, система достигла показателя метрики mAP50 99% (рисунок 4а). Этот говорит об эффективности обучения модели по синтетически сгенерированным тренировочным данным. При этом, метрика dfl_loss на всем протяжении обучения уменьшается, что означает, что модель становится лучше в прогнозировании границ объектов (рисунок 4б).

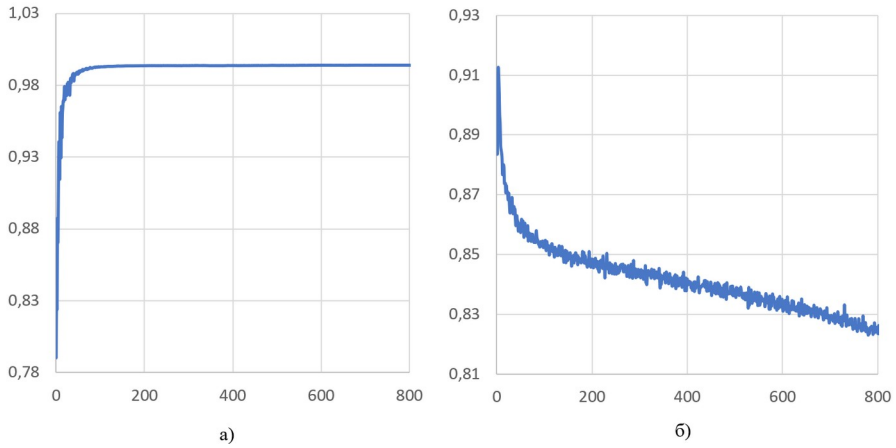

Рисунок 4 - Динамика метрики mAP50 (*а)* и dfl_loss (*б)* на обучающем наборе данных
DOI: https://doi.org/10.23670/IRJ.2024.141.12.4

Тестирование на реальных данных

Для валидации и оценки обученной нейронной сети на синтетических данных был сформирован набор данных, включающий в себя 800 реальных фотографий деталей, изготовленных с использованием 3D-принтера методом FDM печати. Фотографии были сделаны на поверхности с текстурой, имитирующей структуру дерева. Набор данных включает в себя разнообразные снимки деталей в различных контекстах, освещении и ракурсах. На рисунке 5 показан результат детектирования на реальном изображении. Используя ограничительные рамки, полученные в результате обработки алгоритмом детектирования, обученным на синтетических данных нейросети, каждый объект выделен и отмечен соответствующим уровнем уверенности распознавания.

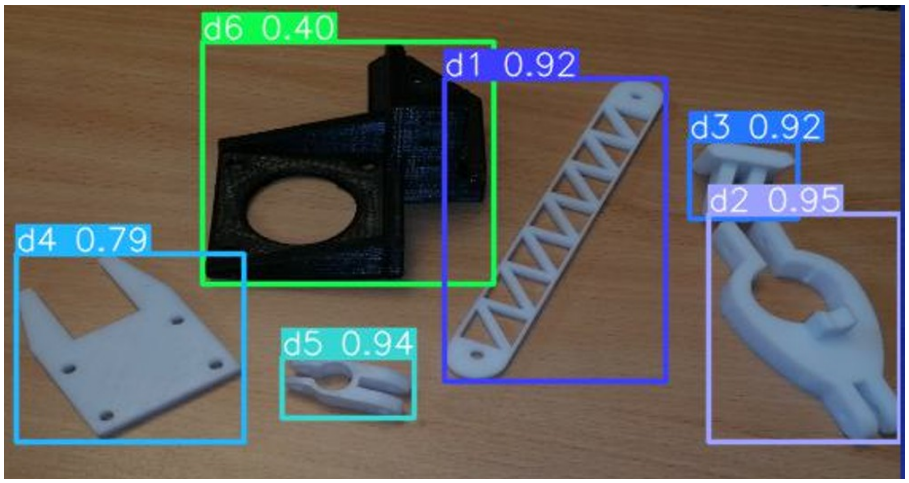

Рисунок 5 - Пример детектирования объектов с границами на реальном изображении
DOI: https://doi.org/10.23670/IRJ.2024.141.12.5

На рисунке 6 отражена нормализованная матрица ошибок, где каждый элемент тестового набора получен по предсказанию модели машинного обучения. Эта матрица дает информацию о классификации каждого элемента в виде «истинно положительного», «истинно отрицательного», «ложноположительного» или «ложноотрицательного» результата.

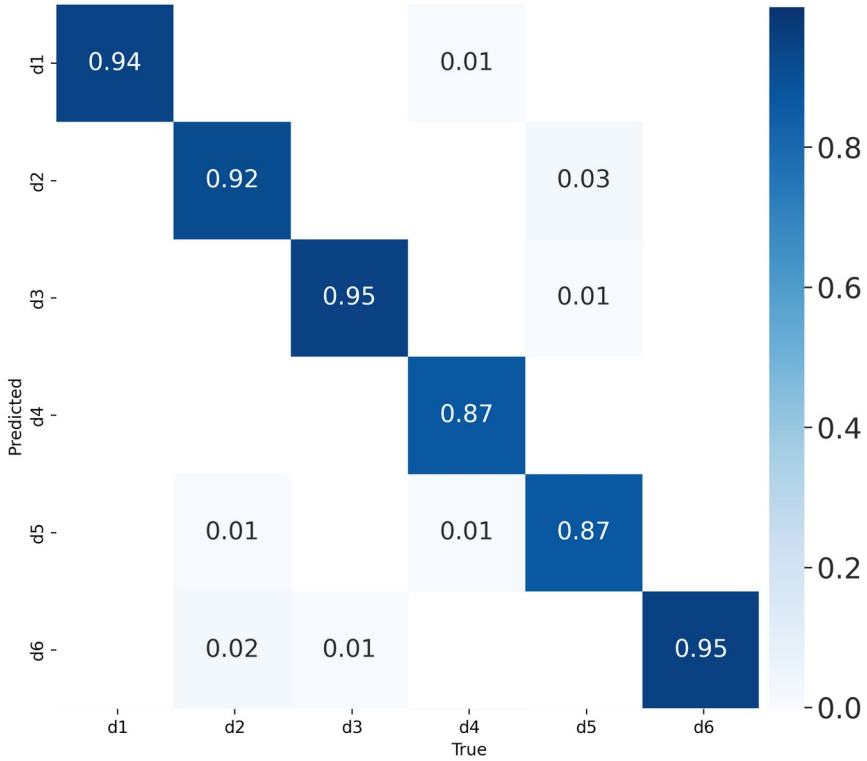

Рисунок 6 - Матрица ошибок для тестового набора данных
DOI: https://doi.org/10.23670/IRJ.2024.141.12.6

На рисунке 7 представлена визуализация тепловых карт активации нейронных сетей, полученная с помощью фреймворка [13], которая демонстрируют области изображения, вызывающие максимальную активацию определенного набора нейронов в сети. Чем более «теплым» является цвет в определенной области, тем сильнее активированы соответствующие нейроны, и мы можем точно определить, где находятся объекты, подлежащие детектированию. Области с выраженной яркостью указывают на те части изображения, которые сильно влияют на принятие решений моделью.

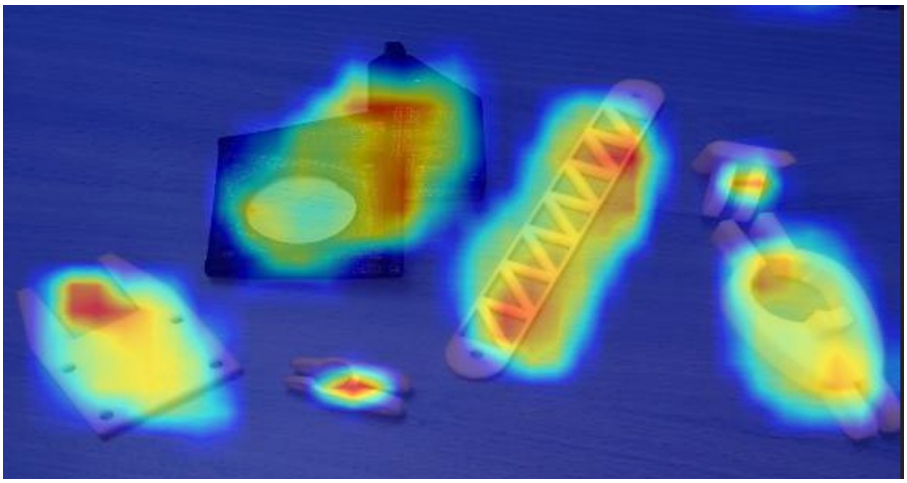

Рисунок 7 - Тепловая карта активаций нейронной сети
DOI: https://doi.org/10.23670/IRJ.2024.141.12.7

Результаты/обсуждение

Синтетические методы генерации обучающей серии изображений обуславливают преимущества в задачах технического зрения, например в системах детектирования деталей узлов сборки, развернутой на автоматизированном производстве, где необходимо вести учет сотен и тысяч сборочных деталей, подверженных частым изменениям. Аннотирование такого огромного количества изображений уже само по себе является очень дорогостоящей задачей. Постоянное обновление обучающих данных из-за изменений в номенклатуре усложняет эту задачу и масштабирование её становится практически невозможным.

Вместе с этим, 3D-модели деталей доступны на этапе проектирования изделий. То есть имея только модель детали, мы уже можем обучить нейронную сеть. Именно по этим причинам мы считаем, что методы генерации полностью синтетических данных необходимы для обеспечения гибкости при развертывании и обновлении систем детектированияобъектов в быстро меняющихся реальных сценариях для производственных процессов. Наше тестирование показывает, что можно качественно обучить модель нейронной сети YOLOv8n, используя только синтетические данные.

Мы предлагаем метод генерации набора данных из синтетических изображений, в котором мы избегаем необходимости ручного аннотирования.

Путем включения большего количества трехмерных моделей, можно увеличить разнообразие генерируемых данных в синтетических сценах, что дополнительно обогатит синтетический набор данных и обеспечит модель машинного обучения большим количеством данных для обучения. Добавление фоновых элементов формирует более разнообразные и реалистичные сцены.

Заключение

Предложенный метод автоматической генерации обучающих наборов данных с использованием 3D-моделей для задачи детектирования объектов показал свою эффективность и применимость. В ходе эксперимента, где модель обучалась на синтетических данных и затем была протестирована на реальных изображениях, отмечается высокий уровень истинно положительной классификации объектов, располагающийся в диапазоне от 87% до 95%. Обычно процесс создания датасета включает съемку изображений или видео с исходными объектами, а также разметку данных с выделением и описанием объектов обучения. В среднем на подготовку одного кадра может уходить от 1 до 2 минуты. Использование синтетических данных, при наличии 3D моделей, существенно сокращает время, затрачиваемые на подготовку набора данных. Для создания сцены и фоновых изображений требуется не более 30 минут, а дальнейшая генерация изображений с аннотациями происходит автоматически, за 15-60 секунд на кадр. Например, для разметки набора данных, аналогичного созданному в работе, с 10 000 изображений потребуется примерно 300 человеко-часов, в то время для синтетической генерации необходимо порядка 60 часов машинного времени.

Финансирование

Работа выполнена при поддержке Министерства науки и высшего образования Российской Федерации (Госконтракт № FEFE-2020-0017).

Конфликт интересов

Не указан.

Рецензия

Артамонов В.А., Международная академия информационных технологий (МНОО "МАИТ), Минск, Беларусь
DOI: https://doi.org/10.23670/IRJ.2024.141.12.8

Funding

This work was supported by the Ministry of Science and Higher Education of the Russian Federation (State Contract No. FEFE-2020-0017).

Conflict of Interest

None declared.

Review

Артамонов V.A., International Academy of information technologies, Minsk, Belarus
DOI: https://doi.org/10.23670/IRJ.2024.141.12.8

Список литературы на английском языке / References in English

1. Zhao Z. Q. Object Detection With Deep Learning: A Review / Z. Q. Zhao, P. Zheng, S. Xu et al. // IEEE Transactions on Neural Networks and Learning Systems. — 2019. — Vol. 30. — Iss. 11. — P. 3212–3232. DOI: 10.1109/TNNLS.2018.2876865.

2. Zou Z. Object Detection in 20 years: A survey / Z. Zou, K. Chen, Z. Shi et al. // Proceedings of the IEEE. — 2023. — Vol. 111. — Iss. 3. — P. 257–276. DOI: 10.1109/JPROC.2023.3238524.

3. Kaur J. Tools, Techniques, Datasets and Application Areas for Object Detection in an Image: a review / J. Kaur, W. Singh // Tools, Techniques, Datasets and Application Areas for Object Detection in an Image. — 2022. — 81. — P. 38297–38351. DOI: 10.1007/s11042-022-13153-y.

4. Blender. — 2023. — URL: https://www.blender.org/ (accessed: 30.11.2023).

5. Tremblay J. Falling Things: A Synthetic Dataset for 3d Object Detection and Pose Estimation / J. Tremblay, T. To, S. Birchfield // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. — 2018. — P. 2038–2041.

6. Tremblay J. Training Deep Networks with Synthetic Data: Bridging the Reality Gap by Domain Randomization / J. Tremblay, A. Prakash, D. Acuna et al. // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. — 2018.

7. Vanherle B. Analysis of Training Object Detection Models with Synthetic Data / B. Vanherle, S. Moonen, F. Van Reeth // 33rd British Machine Vision. — 2022.

8. Kiefer B. Leveraging Synthetic Data in Object Detection on Unmanned Aerial Vehicles / B. Kiefer, D. Ott, A. Zell // 26th International Conference on Pattern Recognition (ICPR). — 2022. DOI: 110.1109/ICPR56361.2022.9956710.

9. Rozantsev A. On Rendering Synthetic Images for Training an Object Detector / A. Rozantsev, V. Lepetit, P. Fua // Computer Vision and Image Understanding. — 2015. — 137. — P. 24–37. DOI: 10.1016/j.cviu.2014.12.006.

10. Hinterstoisser S. An Annotation Saved Is an Annotation Earned: Using Fully Synthetic Training for Object Detection / S. Hinterstoisser, O. Pauly, H. Heibel et al. // Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). — 2019.

11. Cicco M. Di. Automatic Model Based Dataset Generation for Fast and Accurate Crop and Weeds Detection / M. Di. Cicco, C. Potena, G. Grisetti [et al.] // IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). — 2017. DOI: 10.1109/IROS.2017.8206408.

12. Jocher G. Ultralytics YOLOv8 / G. Jocher, A. Chaurasia, J. Qiu. — 2023. — URL: https://github.com/ultralytics/ultralytic (accessed: 30.11.2023).

13. Gildenblat J. PyTorch library for CAM methods / J Gildenblat. — 2021. — URL: https://github.com/jacobgil/pytorchgrad-cam (accessed: 30.11.2023).