Экземплярная сегментация объектов на изображениях с использованием глубокого обучения и синтетических данных

Г.А. Альгашев[1], Е.В. Горбунов[1], И.А. Килбас[1], Р.А. Парингер[1], А.В. Куприянов [1]
[1] Самарский национальный исследовательский университет имени академика С.П. Королёва, 443086, Россия, г. Самара, Московское шоссе, д. 34

Аннотация

В работе рассматривается задача экземплярной сегментации объектов на изображениях с использованием современных моделей глубокого обучения и синтетических данных. Основное внимание уделено исследованию эффективности синтетических данных, созданных на основе 3D-моделей, для предварительного обучения моделей сегментации. Рассматриваются такие архитектуры, как U-Net, DeepLabV3+, Mask R-CNN и YOLOv8. Для улучшения качества синтетических данных использовались различные параметры автоматической генерации данных, включая случайное позиционирование объектов, добавление фонов, изменение освещения, изменение текстуры объекта, добавление размытия и добавление препятствий. Проведённые эксперименты показали, что каждый из этих шагов вносит значительный вклад в точность моделей, а их сочетание обеспечивает наилучшие результаты (mAP 92,1 %). Результаты подтверждают, что комбинированное использование синтетических и реальных данных позволяет преодолеть разрыв между синтетической и реальной средой. Наилучшая производительность была достигнута моделью YOLOv8, которая продемонстрировала высокую точность и скорость обработки. Полученные выводы подчёркивают важность тщательной настройки параметров генерации синтетических данных для улучшения сегментации в условиях реальных приложений.

Ключевые слова: экземплярная сегментация объектов, сегментация объектов, глубокое обучение, свёрточные нейронные сети, синтетические данные, нейросетевые модели, компьютерное зрение, обучение без ручной разметки.

Цитирование: **Альгашев, Г.А.** Экземплярная сегментация объектов на изображениях с использованием глубокого обучения и синтетических данных / Г.А. Альгашев, Е.В. Горбунов, И.А. Килбас, Р.А. Парингер, А.В. Куприянов // Компьютерная оптика. – 2025. – Т. 49, № 6. – С. 1037-1046. – DOI: 10.18287/2412-6179-CO-1656.

Citation: Algashev GA, Gorbunov EV, Kilbas IA, Paringer RA, Kupriyanov AV. Instance segmentation of objects in images using deep learning and synthetic data. Computer Optics 2025; 49(6): 1037-1046. DOI: 10.18287/2412-6179-CO-1656.

Введение

Экземплярная (инстансная) сегментация объектов на изображении – одна из ключевых задач компьютерного зрения, направленная на выделение пиксельных областей, соответствующих положениям объектов интереса (объектов определённых классов) на изображении. Цель этого процесса заключается в создании точной пиксельной маски, которая определяет границы каждого объекта интереса и отделяет его от фона.

В отличие от задач классификации, которая предоставляет информацию о типе объекта, или детектирования, определяющего местоположение объекта с помощью ограничивающего прямоугольника (bounding box), экземплярная сегментация обеспечивает более детализированное представление сцены. Она позволяет идентифицировать объекты конкретных классов и точно определить их форму, что особенно важно для анализа сложных визуальных данных.

Задача экземплярной сегментации объектов на изображениях является фундаментальной для многих современных предметных областей:

1) Робототехника. Сегментация позволяет роботам точно определять положение и форму объектов, что необходимо для манипуляций, например, захвата предметов, работы на сборочных линиях или автономной навигации.
2) Системы контроля качества. Автоматизированные системы с использованием сегментации могут проверять продукцию на наличие дефектов, например, обнаруживать трещины или отклонения от заданной формы.
3) Дополненная, виртуальная и интегрированная реальность. Сегментация играет ключевую роль в создании правдоподобных и интерактивных сцен, где виртуальные элементы интегрируются с реальным миром. В дополненной реальности сегментация используется для корректного наложения цифровых объектов на реальную сцену. В виртуальной реальности сегментация объектов в реальном времени позволяет точно отслеживать взаимодействия пользователя с виртуальными объектами. Интегрированная реальность, объединяющая элементы

AR и VR, требует сегментации для построения гибридных сцен, где физические и виртуальные объекты могут взаимодействовать в реальном времени.

Рис. 1. Пример экземплярной сегментации объектов на изображениях

Особый интерес представляют приложения в области интерактивного обучения, медицины и архитектуры, где точная сегментация позволяет пользователям более глубоко погружаться в виртуальные миры, анализировать сложные данные и принимать информированные решения.

Несмотря на значительный прогресс в области экземплярной сегментации, основным препятствием для её внедрения остается необходимость подготовки больших объемов размеченных данных. Ручная разметка областей для каждого объекта интереса требует значительных затрат времени и ресурсов, так как необходимо точно определить границы каждого объекта определённого класса на изображении.

Основные трудности ручной разметки включают:

1) Высокую стоимость. Для качественной разметки требуется участие экспертов, способных учитывать особенности объектов и сцены.

2) Человеческий фактор. Процесс разметки подвержен ошибкам из-за усталости разметчиков или неоднозначности границ объектов.

3) Ограниченность данных. Сбор реальных изображений часто сопряжен с трудностями, такими как сложные условия освещения, ракурсы или необходимость фиксировать редкие события.

Эти проблемы особенно остро стоят для систем, где необходимо учитывать вариативность объектов и условий съемки, таких как системы дополненной и интегрированной реальности. Использование реальных изображений ограничивает масштабируемость и замедляет процесс внедрения новых решений.

Для преодоления этих ограничений все больше исследователей обращаются к использованию синтетических данных, генерируемых на основе 3D-моделей. Такой подход позволяет создавать разнообразные и качественные датасеты без необходимости ручной разметки, что делает сегментацию более доступной и масштабируемой.

1. Современные алгоритмы на основе глубокого обучения, решающие задачу экземплярной сегментации объекта на изображении

Современные подходы к экземплярной сегментации объектов на изображениях с использованием глубоких сверточных нейронных сетей значительно повысили точность и адаптивность сегментации в разнообразных сценариях. Этот переход стал возможным благодаря увеличению объема данных, доступных для обучения, и развитию вычислительных мощностей, что позволило разрабатывать более сложные архитектуры нейронных сетей.

Классические методы, такие как пороговая сегментация [1], методы активных контуров [2] и графовые подходы [3], имеют ограниченные возможности адаптации к сложным сценам и зависят от ручной настройки параметров. Глубокие нейронные сети, напротив, способны извлекать сложные признаки из данных, что делает их более универсальными.

Ключевые факторы, способствовавшие переходу к глубокому обучению:

1) Данные. Большие объемы размеченных данных стали доступны благодаря созданию открытых датасетов, таких как COCO [4], Pascal VOC [5] и Cityscapes [6]. Эти датасеты содержат высококачественные аннотации, которые служат основой для обучения моделей сегментации.

2) Вычислительные ресурсы. Развитие графических процессоров (GPU), тензорных процессоров (TPU) и распределенных вычислений значительно ускорило процесс обучения глубоких

нейронных сетей, делая возможным использование сложных архитектур.

3) Общедоступные библиотеки. Такие фреймворки, как TensorFlow, PyTorch и Keras, упростили разработку и тестирование нейросетевых моделей, ускорив их внедрение в прикладные задачи [7].

Современные архитектуры глубокого обучения для сегментации строятся на основе Fully Convolutional Networks (FCN), представляющих класс нейронных сетей, где все слои являются сверточными [8]. Это позволяет моделям обрабатывать изображения любого размера и формировать выходные маски без необходимости использования полносвязных слоев. Такие подходы являются основой большинства современных решений в сегментации.

Одной из первых значимых реализаций FCN является сама одноименная архитектура FCN, предложенная как базовый метод для задач семантической сегментации. Её особенностью стало использование слоев деконволюции (upsampling) для восстановления пространственного разрешения, что сделало её применимой для обработки изображений различного размера. Однако базовая FCN имеет ограничения в точности, связанные с недостаточным использованием контекстной информации, что стимулировало развитие более сложных архитектур.

На основе идей FCN были разработаны более специализированные модели, такие как U-Net, DeepLab, Mask R-CNN и YOLO.

U-Net изначально была разработана для биомедицинских изображений, но быстро нашла применение в других областях [9]. Её архитектура включает симметричное построение: блок энкодера для выделения признаков и декодера для восстановления изображения. Характерной особенностью является использование скип-соединений, которые передают информацию с энкодера на декодер, сохраняя пространственные детали. Вариации U-Net, такие как Attention U-Net и ResUNet, расширяют базовую архитектуру, добавляя механизмы внимания и остаточные блоки для повышения точности.

DeepLab применяет пространственные пирамиды (Atrous Spatial Pyramid Pooling, ASPP) для захвата контекста на разных масштабах [10]. Версии v3 и v3+ включают улучшенные механизмы декодирования и агрегации признаков, что делает DeepLab подходящим для сегментации объектов в сложных сценах. DeepLab v3+ дополняет модель специальным декодером для восстановления мелких деталей, что повышает точность сегментации на границах объектов.

Mask R-CNN представляет собой расширение Faster R-CNN, предназначенное для задач instance-сегментации [11]. Она сочетает задачи выделения объектов (bounding box) и масок объектов. Модель добавляет модуль для прогнозирования маски, что делает её мощным инструментом для задач, требующих точного выделения объектов на изображении.

Модели серии YOLO (You Only Look Once) изначально были разработаны для задач детектирования, но их последние версии, такие как YOLOv5 и YOLOv8, включают модули для instance-сегментации [12]. YOLO отличается высокой скоростью обработки изображений, что делает её подходящей для приложений реального времени. Однако точность сегментации может уступать более специализированным моделям, таким как Mask R-CNN.

К преимуществам архитектур глубокого обучения можно отнести:

1) Адаптивность. Глубокие нейронные сети способны обучаться на данных с разнообразными характеристиками, что позволяет их использовать в динамичных и сложных условиях.

2) Автоматическое выделение признаков. Модели глубокого обучения минимизируют необходимость ручной настройки, что упрощает процесс применения.

3) Высокая точность. Современные архитектуры обеспечивают точную сегментацию, включая сложные случаи с неоднородным фоном и мелкими деталями.

Однако архитектуры глубокого обучения имеют следующие ограничения:

1) Требовательность к данным. Обучение глубоких моделей требует больших объемов размеченных данных, что может быть затруднительно в ряде случаев.

2) Вычислительная сложность. Обучение и применение моделей глубокого обучения требует значительных вычислительных ресурсов, что увеличивает затраты.

3) Генерализация. Модели, обученные на одном датасете, могут демонстрировать низкую точность на изображениях, имеющих иную природу.

2. Преимущества и недостатки применения синтетического датасета для обучения моделей глубокого обучения

В данном исследовании мы фокусируемся на преодолении ключевого ограничения экземплярной сегментации – зависимости от ручной разметки. Исследуется, как комбинация синтетических данных, созданных на основе 3D-моделей с вариативными условиями (освещение, текстуры, препятствия), и последующее дообучение на реальных изображениях позволяют достичь точности, сопоставимой с обучением на полностью размеченных реальных датасетах. Это может открыть путь к масштабированию методов сегментации в условиях ограниченных ресурсов.

Подготовка качественного набора данных является важной и сложной задачей при разработке моделей сегментации на основе глубокого обучения. В реальных условиях сбор и разметка данных требуют значительных временных и финансовых ресурсов, особенно если для успешного обучения модели необходимы

разнообразные условия съемки, такие как различное освещение и разнообразные углы обзора. Эти проблемы стимулируют интерес к синтетическим данным, которые могут использоваться для создания больших и вариативных наборов данных без необходимости их реальной съемки и разметки. Однако у синтетических данных есть свои ограничения, включая разрыв между синтетической и реальной средой (reality gap), что усложняет их применение в реальных задачах [13].

При этом реальный датасет остаётся эталонным источником данных для обучения моделей сегментации, так как он отражает действительные условия и особенности среды, где модель будет использоваться. Однако создание и поддержание такого датасета сопряжено с рядом сложностей:

1) Создание и поддержание реального датасета для сегментации требует решения ряда задач, включая организацию процесса съёмки, обработку изображений и их разметку, а также обеспечение разнообразия условий.

2) Сборка реального датасета часто оказывается длительным и трудоемким процессом, требующим участия большого числа людей. В случае сегментации необходимо вручную размечать контуры объектов на каждом изображении, что повышает сложность и стоимость разметки. Особенно это касается промышленных приложений, где требования к точности разметки высоки и ошибки могут приводить к ухудшению качества сегментации и необходимости повторного обучения моделей.

3) Для повышения обобщающей способности модели важно собирать изображения объектов в различных условиях, таких как разные уровни освещённости, различные углы обзора и фоны. Это позволяет моделям лучше справляться с изменчивыми условиями в реальных приложениях. Однако создание такого вариативного датасета в реальной среде требует дополнительных усилий, так как каждое условие съемки может потребовать новой сессии, повышая временные и финансовые затраты на сбор данных.

В связи с трудностями создания реальных наборов данных внимание исследователей все чаще сосредоточено на синтетических данных. Основное преимущество синтетических данных заключается в возможности автоматического создания большого количества изображений с разнообразными характеристиками, что позволяет значительно сократить временные и финансовые затраты на подготовку датасета. Синтетические данные создаются с использованием программного обеспечения для 3D-моделирования и рендеринга, что позволяет генерировать изображения объектов в любых ракурсах и условиях освещения.

Процесс создания синтетических данных обычно включает следующие ключевые этапы:

1) Рендеринг 3D-моделей. С помощью компьютерной графики создаются изображения, основанные на 3D-моделях, что позволяет гибко изменять ракурсы, масштаб и прочие параметры объектов.

2) Добавление процедурно сгенерированных фонов и освещения. Используются алгоритмы, которые случайным образом создают фоны, текстуры и условия освещения, повышая разнообразие синтетических данных и улучшая способность моделей обобщать информацию.

Синтетические данные востребованы в тех областях, где сбор и разметка реальных изображений представляют сложность, например, в робототехнике, медицинской визуализации и системах дополненной реальности. Использование синтетических данных также позволяет адаптировать модели к условиям, которые невозможно полностью воссоздать при сборе реальных данных.

Несмотря на значительные преимущества синтетических данных, существует ряд проблем, связанных с их применением, главная из которых – это разрыв между синтетической и реальной средой, также известный как reality gap. Reality gap проявляется в том, что модель, обученная на синтетических данных, часто демонстрирует пониженную точность при применении к реальным изображениям.

Этот разрыв обусловлен следующими факторами:

1) Ограниченная фотореалистичность. Синтетические изображения, созданные с использованием 3D-моделирования, часто имеют различия с реальными в деталях текстур, освещения и теней, что затрудняет их восприятие нейросетями.

2) Недостаток шумов и артефактов. Реальные изображения обычно содержат различные искажения и шумы, возникающие из-за особенностей камеры, освещения и других факторов, тогда как синтетические изображения зачастую «чисты» и лишены подобных искажений.

3) Разница в контексте. В реальных изображениях присутствуют объекты, которые могут находиться в случайных местах и сочетаться в непредсказуемых композициях, тогда как синтетические изображения часто ограничены заранее заданными сценариями.

Решение этой проблемы требует применения различных методов, таких как адаптация доменов, когда модель обучается учитывать различия между синтетическими и реальными изображениями [14]. Другой подход – это дополнение синтетических данных реальными, что позволяет модели более эффективно адаптироваться к условиям реальной среды.

3. Предлагаемый подход подготовки синтетического датасета

Предлагаемый метод использует трехмерные модели объектов, загружаемые в сцену и преобразуемые с использованием ряда случайных параметров, что позволяет добиться разнообразия позиций, условий освещения и фонов. В качестве платформы для создания сцен используется библиотека VTK (Visualization Toolkit) на

языке Python, что обеспечивает гибкость при работе с трехмерными объектами и текстурами [15].

Шаги генерации синтетических данных:

1) Загрузка 3D-объекта. Для подготовки данных загружается 3D-модель объекта в формате STL, который широко используется в приложениях трехмерного моделирования и содержит точное представление геометрии объекта.

2) Случайное позиционирование и ориентация объекта. Для обеспечения разнообразия визуальных данных положение и ориентация объекта в сцене задаются случайно. Это включает установку трёхмерных координат в случайные значения и поворот объекта вокруг осей трёх осей вращения, что позволяет охватить различные ракурсы и позиции объекта на изображении.

3) Создание маски объекта. Для определения точных границ объекта и создания маски используется окрашивание объекта в черный цвет на белом фоне сцены. Изображение затем анализируется для выделения точных границ объекта, что позволяет автоматически генерировать маску для обучения.

4) Восстановление оригинального цвета объекта и добавление случайного фона. После генерации маски объект восстанавливает свои цвета, а на фон сцены накладывается случайное изображение из подготовленного набора. Этот шаг позволяет моделировать разнообразие реальных фонов, повышая устойчивость обучаемой модели к разным условиям.

5) Случайная настройка цвета и текстуры объекта. Для дальнейшего повышения вариативности данных случайным образом задаются цвет объекта, гладкость его текстуры и отражательные свойства. Эти параметры позволяют имитировать различные материалы (например, металл, пластик) и создавать изображения, более приближенные к реальным сценам.

6) Случайное расположение источников освещения и настройка их параметров. В сцену добавляются один или несколько источников света со случайными параметрами (интенсивность, угол и наличие теней), что позволяет моделировать разнообразные условия освещения, встречающиеся в реальной жизни.

7) Добавление препятствий для частичного перекрытия объекта. Для симуляции сложных условий съемки в сцену добавляется случайное препятствие, которое частично перекрывает объект. Это позволяет моделировать условия, когда объект частично закрыт другими элементами на изображении.

8) Применение размытия. В качестве финального шага применяется эффект размытия, который имитирует особенности реальной съемки, такие как несфокусированность изображения или движение камеры. Это позволяет модели адаптироваться к шуму и другим искажениям, присутствующим в реальных данных.

Таким образом, предложенный метод позволяет эффективно создавать синтетические изображения с заданными масками, охватывающие широкое разнообразие условий, что делает полученные данные ценными для обучения современных моделей сегментации.

4. Основные метрики в задаче экземплярного сегментирования

В задаче сегментирования объектов на изображении метрики играют ключевую роль в оценке качества работы моделей. Существует несколько распространённых метрик, которые помогают оценить как точность, так и полноту сегментации. В этом параграфе рассмотрены наиболее используемые метрики для оценки моделей сегментации: Average Precision, Mean Average Precision, Intersection over Union и FPS [16].

Average Precision (AP) – это метрика, используемая для оценки качества сегментации, особенно в задачах, где важно учитывать как точность (precision), так и полноту (recall). AP рассчитывается как площадь под кривой Precision-Recall (PR), которая отображает зависимость между точностью и полнотой при изменении порогов уверенности модели:

$$\mathrm{AP} = \int_0^1 P\left(R\right) dR\,. \tag{1}$$

На практике AP вычисляется как сумма произведений разницы между соседними значениями ($R_n - R_{n-1}$) на соответствующую точность (P_n):

$$\mathrm{AP} = \sum_n \left(R_n - R_{n-1}\right) \cdot P_n\,. \tag{2}$$

Высокое значение AP указывает на то, что модель эффективно сегментирует объекты, минимизируя ошибки пропуска и ложные срабатывания.

Mean Average Precision (mAP) – это среднее значение AP, усреднённое по всем классам или объектам в задаче сегментации:

$$\mathrm{mAP} = \frac{1}{C}\sum_{c=1}^{C} \mathrm{AP}_c\,. \tag{3}$$

Здесь C – количество классов, AP_C – значение AP для каждого класса. mAP позволяет оценить общую производительность модели, обеспечивая сводную характеристику для задач с несколькими классами.

Intersection over Union (IoU) — метрика для оценки совпадения между предсказанной маской объекта и истинной разметкой. IoU вычисляется как отношение площади пересечения ($A \cap B$) к площади объединения ($A \cup B$):

$$\mathrm{IoU} = \frac{\left|A \cap B\right|}{\left|A \cup B\right|}\,. \tag{4}$$

Значения IoU варьируются от 0 до 1, где 1 соответствует полному совпадению масок. Обычно значение IoU выше 0,5 считается успешным предсказанием.

FPS (Frames Per Second) – это метрика, которая измеряет скорость работы модели, показывая количество обработанных изображений (или кадров) в секунду. FPS является важной метрикой для оценки производительности модели в реальных приложениях, таких как робототехника или дополненная реальность, где требуется высокая скорость обработки изображений для быстрого принятия решений.

5. Проведение эксперимента

Цель эксперимента – оценить влияние синтетических данных на качество сегментации реальных изображений, включая анализ эффективности предварительного обучения на синтетических данных с последующим дообучением на реальных данных. Это позволит определить, как комбинирование синтетических и реальных данных способствует преодолению разрыва между искусственной и реальной средой и повышает точность моделей. В качестве классов объектов выбраны четыре типа LEGO-деталей: кубик 2×2, кубик 2×4, кубик 1×4 и элемент арка 1×3×2.

В качестве обучающих и тестовых данных использовались изображения с разрешением 512×512 пикселей, на которых отображены LEGO-детали, размещённые в различных положениях и с разнообразными фонами. Для проведения экспериментов было сгенерировано 20000 изображений (5000 изображений на каждый класс), из которых 90 % использовались для обучения и 10 % для промежуточной проверки (рис. 2*а*). Для реальных данных было подготовлено 4000 изображений (1000 изображений на каждый класс), которые тоже были разделены в пропорциях 80-10-10 для обучения, промежуточной проверки и итогового тестирования (рис. 2*б*).

Эксперименты были проведены с использованием различных архитектур моделей для задачи сегментации, включая U-Net, DeepLabV3+, Mask R-CNN и YOLOv8. Все модели были предобучены на наборе данных COCO [17], который содержит большой объем разнообразных изображений с объектами из различных категорий. Для всех моделей использовались стандартные параметры оптимизатора Adam ($\beta_1 = 0,9, \beta_2 = 0,999, \varepsilon = 1e-8$) и функция потерь Cross-Entropy Loss [18]. Обучение проводилось в течение 50 эпох, что является стандартной практикой для большинства исследуемых моделей.

Такой подход позволил сократить общее время обучения и повысить общую производительность, так как предобученные веса уже содержат базовые представления об объектах и их характеристиках, что минимизирует необходимость обучения «с нуля» на синтетических или реальных данных [19].

Для всех выбранных моделей было проведено три серии экспериментов:

1) Обучение на синтетических данных. Все модели были обучены на синтетически сгенерированных данных. Важно отметить, что данные были сбалансированы по классам, что обеспечивало равномерное распределение примеров для каждого из четырёх классов LEGO-деталей.

2) Дообучение на реальных данных. После первоначального обучения на синтетических данных модели дообучались на реальных данных в течение того же количества эпох (50 эпох). Такой подход позволил оценить влияние синтетических данных на общее качество модели в реальных условиях.

3) Обучение только на реальных данных. Для оценки влияния синтетических данных был проведён дополнительный эксперимент, в котором модели обучались исключительно на реальных данных. В этом случае использовались те же параметры обучения, однако результаты служат для сравнения с результатами, полученными при обучении на синтетических данных.

Рис. 2. Пример подготовленных данных для обучения и тестирования: (а) синтетические изображения, (б) реальные изображения

Все эксперименты проводились с использованием графического процессора NVIDIA RTX A6000, что обеспечивало высокую производительность моделей и возможность замера количества кадров в секунду (FPS). Важным аспектом является то, что обучение моделей на синтетических данных использовало преимущества быстрого вычисления благодаря GPU, что также позволяет ускорить процесс дообучения на реальных данных.

В табл. 1 приведены результаты проведения экспериментов. Для каждой из моделей были получены следующие результаты, отображающие точность сегментации на реальных данных. Основными метриками, использованными для оценки, были AP, mAP, IoU и FPS.

Результаты экспериментов позволяют сделать несколько важных выводов.

Во-первых, модели, обученные исключительно на синтетических данных, показали хорошие результаты на тестовом наборе реальных данных. Например, YOLOv8 достигла mAP 87,1 %, что подтверждает возможность использования синтетических данных для обучения моделей, особенно в случаях, когда сбор и аннотация реальных данных затруднительны. Это подчеркивает значимость синтетических данных как инструмента для начального обучения.

Во-вторых, дообучение моделей на реальных данных после предварительного обучения на синтетике привело к заметному увеличению точности. YOLOv8, дообученная на реальных данных, показала прирост mAP с 87,1 % до 94,5 %, а IoU увеличился до 94,6 %. Эти улучшения демонстрируют важность комбинации синтетических и реальных данных для достижения высокой производительности.

Третье наблюдение касается моделей, обученных только на реальных данных. Хотя они продемонстрировали приемлемые результаты, их производительность была ниже по сравнению с моделями, прошедшими смешанное обучение. Например, YOLOv8, обученная только на реальных данных, достигла mAP 83,2 %, что на 11,3 % ниже, чем у модели, дообученной на реальных данных. Этот результат можно объяснить ограниченным объемом реального набора данных, что не позволило моделям полностью раскрыть свой потенциал.

Среди рассмотренных моделей YOLOv8 продемонстрировала наивысшие результаты по точности (mAP и IoU) и скорости обработки (FPS), что делает её наиболее перспективной для задач сегментации. Это превосходство объясняется несколькими факторами. Во-первых, YOLOv8 обладает оптимизированной архитектурой, которая эффективно использует предобученные веса, обученные на датасете MS COCO, для извлечения признаков. Это позволяет модели лучше обобщать информацию и адаптироваться к разным типам данных. Во-вторых, YOLOv8 сочетает высокую скорость обработки (до 121 FPS) с сохранением высокой точности, что делает её особенно полезной для приложений в реальном времени. Наконец, модель демонстрирует устойчивость к разрыву между синтетическими и реальными данными, что подтверждается её результатами как на синтетических, так и на реальных тестовых наборах. Эти преимущества делают YOLOv8 лидером среди рассмотренных архитектур для задач сегментации.

Табл. 1. Результаты проведения экспериментов

Модель	Тип обучения	IoU (%)	AP (%)	mAP (%)	FPS
U-Net	Только синтетические данные	84,2	82,2	81,2	10
DeepLabV3+		87,4	85,8	84,3	6
Mask R-CNN		88,2	86,3	85,4	5
YOLOv8		92,1	86,6	90,4	121
U-Net	Синтетические данные + дообучение на реальных данных	87,1	86,2	85,6	10
DeepLabV3+		89,8	87,7	89,7	6
Mask R-CNN		91,4	88,4	91,3	5
YOLOv8		94,6	92,7	94,5	121
U-Net	Только реальные данные	77,4	78,8	76,5	10
DeepLabV3+		78,6	79,1	76,3	6
Mask R-CNN		80,9	78,3	79,7	5
YOLOv8		83,2	82,5	83,2	121

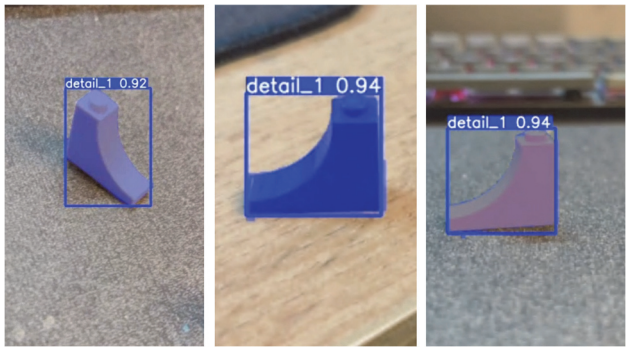

Рис. 3. Демонстрация сегментирования объектов моделью YOLOv8

Модели, обученные исключительно на синтетических данных, демонстрируют конкурентный уровень точности на реальных данных, что подтверждает применимость синтетических данных для начального этапа обучения. Однако из-за разрыва между синтетической и реальной средой (reality gap) их производительность уступает моделям, дообученным на реальных данных. Это подчёркивает важность комбинированного подхода, который позволяет значительно повысить точность сегментации в реальных условиях.

6. Влияние параметров генерации синтетических данных на точность сегментации

Для оценки влияния различных параметров генерации синтетических данных на точность сегментации были проведены эксперименты с использованием последовательного добавления следующих шагов при генерации изображений:

1) Добавление случайного фона.
2) Настройка случайного цвета и текстуры объекта.
3) Случайное расположение источников освещения.
4) Добавление препятствий.
5) Добавление размытия.

Для эксперимента были сгенерированы следующие варианты синтетических датасетов:

1) По одному шагу за раз. Датасеты, в которых применяется только один из шагов.

2) Все шаги, кроме одного. Датасеты, в которых исключён один из параметров.

3) Все шаги применены. Полный набор параметров генерации данных.

Важно отметить, что ко всем вариантам датасета применялась случайная генерация положения и ориентации объекта. Каждый датасет был сгенерирован аналогично, как в предыдущей серии экспериментов, 20000 изображений (по 5000 изображений на каждый класс), из которых 90 % использовались для обучения и 10 % для промежуточной проверки.

Для обучения было решено взять модель YOLOv8, так как она показала лучший результата производительности в предыдущей серии экспериментов. С YOLOv8 проводилось обучение на каждом варианте синтетического датасета. Тестирование осуществлялось на одном и том же реальном тестовом наборе данных.

Результаты из табл. 2 показывают, что каждый из шагов генерации синтетических данных делает важный вклад в повышение качества сегментации.

Модель, обученная на базовом датасете без каких-либо улучшений, показала самые низкие результаты (mAP 78,0 %). Это подчеркивает недостаточность простого подхода к генерации синтетических данных.

Влияние отдельных параметров:

1) Случайный фон добавил ещё 2,2 %, улучшая адаптацию модели к реальным изображениям с разнообразными фонами.

2) Случайный цвет и текстура повысили mAP на 6,5 % относительно базового уровня, подтверждая значимость этих параметров для повышения устойчивости модели к визуальным вариациям.

3) Случайное освещение обеспечило прирост mAP на 7,9 %, моделируя условия, характерные для реальных сцен с переменной яркостью.

4) Препятствия увеличили mAP на 9,8 %, так как модели научились лучше справляться с объектами, частично закрытыми другими элементами сцены.

5) Размытие дало прирост mAP на 10,6 %, подчеркивая его роль в повышении устойчивости к размытым изображениям.

Исключение отдельных шагов:

1) Исключение случайного цвета и текстуры или размытия приводит к значительным потерям качества (снижение mAP на 2,7 % и 1,8 % соответственно). Это подтверждает их важность для повышения точности сегментации.

2) Исключение других параметров также снижает mAP, но в меньшей степени (от 1,8 % до 3,2 %), что показывает их поддержку общей устойчивости модели.

Датасет с применением всех шагов генерации показал наивысшие значения метрик (mAP 92,1 %, IoU 84,0 %), что демонстрирует синергетический эффект от объединения всех параметров.

Табл. 2. Результаты экспериментов с различными параметрами генерации синтетических данных

Модель	Вариант датасета	IoU (%)	AP (%)	mAP (%)
YOLOv8	Базовый датасет без добавления шагов	78,0	70,5	76,8
YOLOv8	Добавление случайного фона	84,7	76,1	82,3
YOLOv8	Добавление случайного цвета и текстуры объекта	86,5	78,4	84,1
YOLOv8	Добавление случайного освещения	85,9	77,6	83,5
YOLOv8	Добавление препятствия	87,8	79,3	85,7
YOLOv8	Добавление случайного размытия	88,6	80,1	86,9
YOLOv8	Все шаги, кроме случайного фона	88,9	80,5	86,9
YOLOv8	Все шаги, кроме случайного освещения	89,1	80,7	87,5
YOLOv8	Все шаги, кроме случайного цвета и текстуры	89,4	81,0	87,5
YOLOv8	Все шаги, кроме размытия	90,3	81,9	88,2
YOLOv8	Все шаги, кроме препятствий	89,6	81,3	87,9
YOLOv8	Все шаги применены	92,1	86,6	90,4

Эксперименты подтверждают, что каждый шаг в процессе генерации синтетических данных вносит значительный вклад в качество обучения модели сегментации. Однако максимальное улучшение достигается при их комбинировании. Такой подход позволяет создавать высокоэффективные синтетические датасеты, что особенно важно для задач, где доступ к большим наборам реальных данных ограничен.

Заключение

В данной работе исследована задача экземплярной сегментации объектов на изображениях с использованием современных моделей глубокого обучения и синтетических данных. Проведённые эксперименты подтвердили, что использование синтетических данных, созданных с учётом параметров, таких как случайное позиционирование объектов, фоны, освещение, текстуры, размытие и препятствия, существенно улучшает качество моделей сегментации. Каждый шаг генерации данных вносит свой вклад в общую точность, а их комбинация приводит к максимальной производительности моделей.

Результаты экспериментов подтвердили, что современные архитектуры глубокого обучения, такие как U-Net, DeepLabV3+, Mask R-CNN и YOLOv8, демонстрируют высокую производительность на задаче сегментации. Среди рассмотренных моделей YOLOv8 показала наилучшие результаты, достигая

наибольших значений mAP и IoU, сохраняя при этом высокую скорость обработки. Это делает её наиболее перспективной для практического применения, где требуется баланс между точностью и производительностью.

Одной из ключевых целей исследования было изучение потенциала синтетических данных для обучения моделей сегментации. Проведённые эксперименты продемонстрировали, что модели, обученные исключительно на синтетических данных, показывают конкурентные результаты на реальных изображениях. Однако дообучение моделей на реальных данных после предварительного обучения на синтетике привело к значительному увеличению точности. Например, YOLOv8 улучшила mAP с 87,1 % до 94,5 %, а IoU – с 89,2 % до 94,6 %. Это подтверждает, что сочетание синтетических и реальных данных является наиболее эффективной стратегией.

Следует отметить, что модели, обученные только на реальных данных, показали более низкие результаты, чем те, которые использовали комбинацию синтетических и реальных данных. Это связано с ограниченным объёмом реальных данных, что ограничивало возможности моделей для обобщения.

Таким образом, использование синтетических данных в комбинации с небольшими наборами реальных данных решает проблему дефицита размеченных данных и позволяет улучшить качество сегментации. Полученные результаты показывают, что синтетические данные могут быть эффективно использованы для предварительного обучения, предоставляя модели возможность адаптироваться к реальным условиям на этапе дообучения. Это делает предложенный подход полезным для задач, где сбор и разметка реальных данных являются дорогостоящими и трудозатратными.

Благодарности

Работа выполнена в рамках государственного задания по теме FSSS-2023-0006.

References

[1] Turajlić E. Multilevel Thresholding Image Segmentation Based on Multi-swarm Particle Swarm optimization with a Dynamic Learning Strategy and Kapur's entropy. 31st Telecommunications Forum (TELFOR) 2023. 1–4. DOI: 10.1109/TELFOR59449.2023.10372741.

[2] Iqbal E, Niaz A, Munir A, Choi KN. Hybrid Active Contour Model for Segmentation of Synthetic and Real Images. 2021 Int Symposium on Intelligent Signal Processing and Communication Systems (ISPACS) 2021: 1-2. DOI: 10.1109/ISPACS51563.2021.9651047.

[3] Zhang L, Zhang H, Wang J, Yang Q. GrabCut: Interactive foreground extraction using graph cuts. ACM Trans Graph 2004; 23(3): 309-314. DOI: 10.1145/1015706.1015720.

[4] Puri D. COCO Dataset stuff segmentation challenge. 2019 5th Int Conf on Computing, Communication, Control and Automation (ICCUBEA) 2019: 1-5. DOI: 10.1109/ICCUBEA47591.2019.9129255.

[5] Everingham D, Van Gool L, Williams C, Winn J, Zisserman A. The Pascal Visual Object Classes (VOC) challenge. Int J Comput Vis 2010; 88(2): 303-338. DOI: 10.1007/s11263-009-0275-4.

[6] Cordts M, Omran M, Ramos S. The cityscapes dataset for semantic urban scene understanding. 2016 IEEE Conf on Computer Vision and Pattern Recognition (CVPR) 2016: 3213-3223. DOI: 10.1109/CVPR.2016.350.

[7] Bovshik PP. Analysis of frameworks for neural networks [In Russian]. Science, technology and education. 2021; 3: 20–23.

[8] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. 2015 IEEE Conf on Computer Vision and Pattern Recognition (CVPR) 2015: 3431-3440. DOI: 10.1109/CVPR.2015.7298965.

[9] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In Book: Navab N, Hornegger J, Wells WM, Frangi AF, eds. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III. Dordrecht: Springer International Publishing Switzerland; 2015: 234-241. DOI: 10.1007/978-3-319-24574-4_28.

[10] Chen L-C, Papandreou G, Kokkinos I, Dollár P, Zhang LY. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans Pattern Anal Machine Intell 2018; 40(4): 834-848. DOI: 10.1109/TPAMI.2017.2699184.

[11] He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. 2017 IEEE Int Conf on Computer Vision (ICCV) 2017: 2961-2969. DOI: 10.1109/ICCV.2017.322.

[12] Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. 2016 IEEE Conf on Computer Vision and Pattern Recognition (CVPR) 2016: 779-788. DOI: 10.1109/CVPR.2016.91.

[13] Konushin AS, Faizov BV, Shakhuro VI. Road images augmentation with synthetic traffic signs using neural networks. Computer Optics 2021; 45(5): 736-748. DOI: 10.18287/2412-6179-CO-859.

[14] Imbusch B, Schwarz M, Behnke S. Synthetic-to-Real domain adaptation using contrastive unpaired translation. arXiv Preprint. 2022. Source: <https://arxiv.org/abs/2203.09454>. DOI: 10.48550/arXiv.2203.09454.

[15] Makarov SN, Verhoglyad AG, Stupak MF, Ovchinnikov DA, Oberemok JA. Mathematical simulation of a 3D scanner for controlling the mirror system of the Millimetron Observatory. Computer Optics 2021; 45(4): 541-550. DOI: 10.18287/2412-6179-CO-833.

[16] Bochkovskiy A, Wang C-Y, Liao H-YM. YOLOv4: Optimal speed and accuracy of object detection. arXiv Pre-print. 2020. Source: <https://arxiv.org/abs/2004.10934>. DOI: 10.48550/arXiv.2004.10934.

[17] Lin T-Y, Maire M, Belongie S, et al. Microsoft COCO: Common objects in context. In Book: Fleet D, Pajdla T, Schiele B, Tuytelaars T, eds. Computer Vision – ECCV 2014. 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V. Cham: Springer International Publishing Switzerland; 2014: 740-755. DOI: 10.1007/978-3-319-10602-1_48.

[18] Kingma DP, Ba J. Adam: A method for stochastic optimization. Int Conf on Learning Representations (ICLR) 2015.

[19] Krasnov DI. Attention modules in convolutional neural networks for small object recognition. Computer Optics 2024; 48(6): 963-968. DOI: 10.18287/2412-6179-CO-1468.

Сведения об авторах

Альгашев Геннадий Андреевич, 1996 года рождения, в 2020 году окончил факультет информатики Самарского национального исследовательского университета имени академика С.П. Королёва, ассистент кафедры информационных систем и технологий Самарского национального исследовательского университета имени академика С.П. Королёва. Область научных интересов: нейронные сети, цифровая обработка изображений, дополненная и виртуальная реальность, компьютерное зрение, интеллектуальный анализ данных.
E-mail: *algashev.ga@ssau.ru*

Горбунов Егор Вадимович, 2004 года рождения, студент Института информатики и кибернетики Самарского национального исследовательского университета имени академика С.П. Королева (Самарский университет). Область научных интересов: нейронные сети, цифровая обработка изображений, компьютерное зрение.
E-mail: *egorgorbunovclashroyale@gmail.com*

Килбас Игорь Александрович, 2000 года рождения, в 2023 году окончил магистратуру Института информатики и кибернетики Самарского национального исследовательского университета имени академика С.П. Королёва, старший лаборант научно-исследовательской лаборатории автоматизированных систем научных исследований. Круг научных интересов включает интеллектуальный анализ данных, распознавание образов и искусственные нейронные сети. E-mail: *kilbas.ia@ssau.ru*

Парингер Рустам Александрович, 1990 года рождения, доцент кафедры технической кибернетики Самарского национального исследовательского университета имени академика С.П. Королева (Самарский университет). В 2013 году окончил факультет информатики СГАУ. Кандидат технических наук с 2017 года. Круг научных интересов включает интеллектуальный анализ данных, распознавание образов и искусственные нейронные сети.
E-mail: *rusparinger@ssau.ru*

Куприянов Александр Викторович, профессор кафедры технической кибернетики Самарского национального исследовательского университета имени академика С.П. Королёва; старший научный сотрудник лаборатории математических методов обработки изображений Института систем обработки изображений РАН, НИЦ «Курчатовский институт» – филиала ФНИЦ «Кристаллография и фотоника» РАН. Сфера научных интересов: цифровая обработка сигналов и изображений; распознавание образов и искусственный интеллект; анализ и интерпретация биомедицинских сигналов и изображений. E-mail: *akupr@ssau.ru*

ГРНТИ: 20.53.19
Поступила в редакцию 06 декабря 2024 г. Окончательный вариант – 11 марта 2025 г.

Instance segmentation of objects in images using deep learning and synthetic data

G.A. Algashev [1], E.V. Gorbunov [1], I.A. Kilbas [1], R.A. Paringer [1], A.V. Kupriyanov [1]
[1] Samara National Research University, 443086, Samara, Russia, Moskovskoye Shosse 34

Abstract

The paper considers the problem of instance segmentation of objects in images using modern deep learning models and synthetic data. The main attention is paid to the study of the effectiveness of synthetic data created on the basis of 3D models for pre-training segmentation models. Such architectures as U-Net, DeepLabV3+, Mask R-CNN and YOLOv8 are considered. To improve the quality of synthetic data, various parameters of automatic data generation were used, including random positioning of objects, adding backgrounds, changing lighting, changing object texture, adding blur and adding obstacles. The experiments showed that each of these steps significantly contributes to the accuracy of the models, and their combination provides the best results (mAP 92.1%). The results confirm that the combined use of synthetic and real data allows bridging the gap between the synthetic and real environment. The best performance was achieved by the YOLOv8 model, which demonstrated high accuracy and processing speed. The obtained findings highlight the importance of carefully tuning the parameters of synthetic data generation to improve segmentation in real-world applications.

Keywords: instance segmentation of objects, object segmentation, deep learning, convolutional neural networks, synthetic data, neural network models, computer vision, learning without manual labeling.

Citation: Algashev GA, Gorbunov EV, Kilbas IA, Paringer RA, Kupriyanov AV. Instance segmentation of objects in images using deep learning and synthetic data. Computer Optics 2025; 49(6): 1037-1046. DOI: 10.18287/2412-6179-CO-1656.

Acknowledgements: The research was carried out within the state assignment theme FSSS-2023-0006.

Author's information

Gennady Andreevich Algashev, (b. 1996), graduate of the master's degree program of Informatics faculty at Samara National Research University, assistant of the Department of Information Systems and Technologies of Samara National Research University. Research interests: neural networks, digital image processing, augmented and virtual reality, computer vision, data mining. E-mail: *algashev.ga@ssau.ru*

Egor Vadimovich Gorbunov, (b. 2004), student of the Institute of Informatics and Cybernetics of the Samara National Research University. Research interests: neural networks, digital image processing, computer vision. E-mail: *egorgorbunovclashroyale@gmail.com*

Igor Alexandrovich Kilbas, (b. 2000), graduate of the master's degree program of Informatics faculty at Samara National Research University, senior lab assistant of the research laboratory "Photonics for a Smart Home and Smart City". Research interests include data mining, artificial neural networks and language models. E-mail: *kilbas.ia@ssau.ru*

Rustam Alexandrovich Paringer, (born 1990), received Master's degree in Applied Mathematics and Informatics from Samara State Aerospace University (2013). He received his PhD in 2017. Associate professor of the Technical Cybernetics department of Samara National Research University. Research interests: data mining, machine learning and artificial neural networks. E-mail: *rusparinger@ssau.ru*

Alexandr Victorovich Kupriyanov, (b. 1978), graduated (2001) from the S.P. Korolyov Samara State Aerospace University (SSAU). He received his PhD in Technical Sciences (2004). At present he is a senior researcher at the Image Processing Systems Institute, NRC "Kurchatov Institute", and holding a part-time position of Associate Professor at Technical Cybernetics department of Samara University. The area of interests includes digital signals and image processing, pattern recognition and artificial intelligence, biomedical imaging and analysis. His list of publications contains more than 80 scientific papers, including 35 articles and 1 monograph published. E-mail: *akupr@ssau.ru*

Received December 06, 2024. The final version – March 11, 2025.