УДК 004.932.72+004.85

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ЭФФЕКТИВНОСТИ МАСШТАБИРОВАНИЯ И ВАРИАТИВНОСТИ СИНТЕТИЧЕСКИХ ДАТАСЕТОВ ДЛЯ СЕГМЕНТАЦИИ ОБЪЕКТОВ

Е.В. Горбунов[1], Г.А. Альгашев[1]

[1]Самарский университет, г. Самара
e-mail: egor_gorbunow@vk.com

Аннотация

В работе исследуется влияние двух факторов на качество обучения модели сегментации YOLOv8: объёма синтетического датасета и вариативности параметров генерации (фон, освещение, текстуры, препятствия, размытие). Проведены эксперименты с фиксированным (5000 изображений) и масштабируемым (до 50000) объёмом при низкой и высокой вариативности. Показано, что увеличение разнообразия синтетических сцен даёт больший прирост точности на реальных изображениях (+15,3% mAP@50), чем пятикратное увеличение объёма однотипных данных (+4,1%). Сформулированы практические рекомендации по оптимальному распределению вычислительных ресурсов.

Ключевые слова: синтетические данные, сегментация объектов, YOLOv8, вариативность данных, масштабирование датасетов, компьютерное зрение.

Введение

Современные подходы к обучению нейросетевых моделей компьютерного зрения требуют больших объёмов размеченных данных. Ручная аннотация изображений остаётся трудоёмким процессом, особенно в задачах сегментации, где необходимо попиксельно выделять границы объектов [1].

Генерация синтетических данных на основе 3D-моделей позволяет автоматически получать изображения с точными аннотациями, варьируя освещение, ракурсы, текстуры и фоны [2, 3]. В работе [4] было показано, что добавление каждого параметра вариативности (случайный фон, освещение, текстуры, препятствия, размытие) последовательно улучшает качество сегментации. Однако остаётся открытым вопрос об оптимальной стратегии генерации: что эффективнее при ограниченных вычислительных ресурсах – увеличивать объём синтетических данных или их разнообразие.

Цель данной работы – количественно оценить вклад этих двух факторов и определить наиболее эффективную стратегию генерации синтетических датасетов.

1. Методика эксперимента

В качестве модели сегментации выбрана YOLOv8-seg [5], обеспечивающая оптимальный баланс скорости и точности. Генерация синтетических сцен выполнялась с использованием библиотеки VTK [6], постобработка и выделение масок – с помощью OpenCV [7].

Объект исследования – LEGO-деталь типа «кирпич 2×4», обладающая чётко выраженной геометрией и допускающая варьирование текстур и условий освещения. Тестовый набор – 200 реальных фотографий, не участвовавших в обучении.

В работе использовались следующие варьируемые параметры синтетической сцены: фон (набор из 100 случайных изображений), освещение (интенсивность 0,5–2,0, угол 0–360°), текстура объекта (30 вариантов: металл, пластик, дерево), препятствия (перекрытие 10–50%), размытие ($\sigma = 0 – 3$ пикселя).

Таблица 1 – План экспериментальных серий

Серия	Объём	Вариативность	Параметры
А (фиксированный объём)	5000	Минимальная → Высокая	4 градации: от 1 фона/1 освещения до 100 фонов/30 освещений
Б (масштабирование, низкая вариативность)	1000 – 50000	Минимальная	1 фон, 1 освещение, 1 текстура
В (масштабирование, высокая вариативность)	1000 – 20000	Высокая	100 фонов, 30 освещений, 30 текстур, препятствия, размытие

Обучение проводилось в течение 100 эпох с использованием оптимизатора Adam. Метрики оценки – mAP@50, mAP@50-95, IoU.

2. Результаты экспериментов

Результаты серии А представлены в таблице 2. Переход от минимальной вариативности к высокой при неизменном объёме (5000 изображений) позволил увеличить mAP@50 на 15,3% (с 72,4% до 87,7%) и IoU на 11,8%.

Таблица 2 – Зависимость точности от вариативности (N = 5000)

Вариативность	mAP@50 (%)	mAP@50-95 (%)	IoU (%)
Минимальная	72,4	58,1	74,2
Низкая	78,6	63,4	79,1
Средняя	84,2	68,7	83,5
Высокая	87,7	72,3	86,0

Увеличение объёма однотипных данных (серия Б) даёт значительно меньший эффект. При росте объёма с 1000 до 50000 изображений mAP@50 увеличивается лишь на 4,1% (с 69,8% до 73,9%), после 10000 изображений рост практически прекращается.

При высокой вариативности (серия В) увеличение объёма с 1000 до 20000 изображений даёт прирост mAP@50 на 6,8% (с 81,2% до 88,0%), причём насыщение наступает позже – после 10000–15000 изображений.

В проведённых экспериментах среднее время генерации одного изображения составило 0,47 с для минимальной вариативности и 0,52 с для высокой – разница не превышает 11%. Генерация 50000 низковариативных изображений требует ≈ 23500 секунд (около 6,5 часов), что в **9 раз больше** времени, чем генерация 5000 высоковариативных изображений (≈ 2600 секунд, около 43 минут).

Несмотря на значительно бóльшие временные затраты (в 9 раз!), 50000 низковариативных изображений обеспечивают mAP@50 = 73,9%, что существенно ниже, чем 87,7%, достигнутых на 5000 высоковариативных изображениях. Таким образом, увеличение вариативности демонстрирует **кратно более высокую эффективность** по отношению к затраченным вычислительным ресурсам: вариативность даёт более высокую точность при меньших временных затратах, а простое увеличение объёма однотипных данных требует огромных ресурсов, но приводит лишь к незначительному росту точности.

Таблица 3 – Воспроизводимость результатов (среднее отклонение, n=3)

Вариативность	mAP@50 (%)	mAP@50-95 (%)
Минимальная	72,4	58,1
Низкая	78,6	63,4
Средняя	84,2	68,7
Высокая	87,7	72,3

Низкие значения стандартного отклонения (не более 1,2% для mAP@50) подтверждают воспроизводимость результатов и устойчивость модели к случайным факторам.

3. Обсуждение

Полученные результаты согласуются с более широким контекстом исследований в области синтетических данных и domain randomization. Ранее в работе [4] было показано, что каждый добавленный параметр вариативности последовательно улучшает качество сегментации. Настоящее исследование развивает этот вывод, демонстрируя, что эффект от вариативности не только аддитивен, но и при фиксированном объёме данных превосходит эффект от простого масштабирования.

Наблюдаемый эффект объясняется тем, что при низкой вариативности модель запоминает специфические артефакты синтетической сцены (фиксированный фон, постоянное освещение, повторяющиеся текстуры) и не учится выделять инвариантные признаки объекта. Это явление в литературе известно как «синтетический разрыв» (reality gap) [7]. Высокая вариативность, напротив, вынуждает модель обобщать, так как каждый новый кадр предъявляет объект в уникальных условиях. Аналогичные выводы были получены в задачах детекции объектов на дорожных сценах [8] и в робототехнике [9], где domain randomization позволяет успешно переносить модели из симуляции в реальный мир без дообучения.

Ключевой вывод данного исследования заключается в том, что **вариативность важнее объёма**. Даже при значительно бóльших временных затратах (в 9 раз больше) увеличение объёма однотипных данных не позволяет достичь точности, которую даёт использование высоковариативных данных меньшего объёма. Следовательно, при планировании генерации синтетических датасетов приоритет следует отдавать разнообразию сцен, а не погоне за количеством.

Практические рекомендации:

1) при ограниченном бюджете вычислений следует генерировать 2000–5000 изображений с максимальной вариативностью параметров;
2) увеличение объёма однотипных данных за пределами 5000–10000 изображений нецелесообразно – оно требует больших ресурсов, но даёт незначительный прирост точности;
3) при высокой вариативности полезно масштабирование до 10000–15000 изображений, дающее дополнительный прирост точности.

Заключение

В работе проведён сравнительный анализ влияния объёма и вариативности синтетических данных на качество сегментации реальных изображений. Установлено, что увеличение разнообразия параметров генерации даёт значительно больший прирост точности (+15,3% mAP@50), чем пятикратное увеличение объёма однотипных данных (+4,1%). Более того, даже при значительно бóльших временных затратах (в 9 раз больше) увеличение объёма однотипных данных не позволяет достичь точности высоковариативных датасетов меньшего объёма. Максимальная достигнутая точность (mAP@50 = 88,0%) получена при использовании 20000 высоковариативных изображений.

Полученные результаты позволяют оптимизировать процесс генерации синтетических датасетов, сокращая вычислительные затраты при сохранении высокой обобщающей способности моделей. В дальнейшем планируется исследовать влияние вариативности на более сложных мультиобъектных сценах.

Список использованных источников

1. Кузнецов, А. В. Глубокое обучение в компьютерном зрении: методы и практика / А. В. Кузнецов, Д. И. Петров. – М.: ДМК Пресс, 2021. – 324 с.

2. Smith, J. Synthetic Data Generation for Object Detection: A Survey / J. Smith, L. Brown // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2021. – V. 44, № 7. – P. 3210–3225.

3. Denninger, S. BlenderProc: Towards Realistic Synthetic Data Generation / S. Denninger, K. M. Wurm // arXiv preprint. – 2021. – arXiv:2101.04637.

4. Экземплярная сегментация объектов на изображениях с использованием глубокого обучения и синтетических данных / Г.А. Альгашев, Е. В. Горбунов, И. А. Килбас, Р. А. Паринге р, А. В. Куприянов // Компьютерная оптика. – 2025. – Т. 49, № 6. – С. 1021–1030.

5. Ultralytics. YOLOv8 Documentation – URL: https://docs.ultralytics.com (дата обращения: 05.04.2026).

6. Schroeder, W. The Visualization Toolkit / W. Schroeder, K. Martin, B. Lorensen. – Clifton Park, NY: Kitware, 2021. – 470 p.

7. OpenCV – URL: https://docs.opencv.org (дата обращения: 05.04.2026).

8. Tobin, J. Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World / J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, P. Abbeel // 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). – 2017. – P. 23–30.

9. Sadeghi, F. CAD2RL: Real Single-Image Flight without a Single Real Image / F. Sadeghi, S. Levine // Robotics: Science and Systems XIII. – 2017.

COMPARATIVE ANALYSIS OF THE EFFICIENCY OF SCALING AND VARIABILITY OF SYNTHETIC DATASETS FOR OBJECT SEGMENTATION

E.V. Gorbunov[1], G.A. Algashev[1]

[1]Samara University, Samara
e-mail egor_gorbunow@vk.com,

Abstract

This paper investigates the influence of two factors on the quality of training of the YOLOv8 segmentation model: the size of the synthetic dataset and the variability of generation parameters (background, lighting, textures, obstacles, blur). Experiments were conducted with a fixed (5000 images) and scalable (up to 50000) volume at low and high variability. It is shown that increasing the diversity of synthetic scenes gives a greater increase in accuracy on real images (+15.3% mAP@50) than a fivefold increase in the volume of uniform data (+4.1%). Practical recommendations for optimal allocation of computational resources are formulated.

Keywords: synthetic data, object segmentation, YOLOv8, data variability, dataset scaling, computer vision.