

УДК 681.7.068
EDN: JXCXBA

PACS: 42.30.Va, 42.79.Bh, 07.05.Mh

**Определение оптимального объема синтетических данных
для обучения нейросетей детектирования объектов (на примере YOLOv8)**

© А. А. Берилло

АО «НПО «Орион», Москва, 111538 Россия
E-mail: berillo.andrey@yandex.ru*Статья поступила в редакцию 25.06.2026; после доработки 9.07.2026; принята к публикации 06.08.2026
Шифр научной специальности: 2.2.6*

Рассматривается задача определения минимального достаточного объема синтетических данных для обучения нейросетей детектирования объектов в формате YOLO. Предложен подход к генерации размеченных синтетических изображений с использованием разработанного программного модуля Yolo Synthetic Data Generator. Проведено обучение трёх вариантов модели YOLOv8 (nano, small, medium) на датасетах различных размеров (от 100 до 10000 кадров). Оценка качества выполнена на пяти тестовых видеороликах по числу обнаруженных объектов и кадров с детекциями. Показано, что для моделей nano и small оптимальным является датасет объемом ~1000 кадров, для модели medium – ~500 кадров. Дальнейшее увеличение датасета не приводит к значимому росту качества, но увеличивает вычислительные затраты. Полученные результаты могут быть использованы при разработке систем технического зрения, в том числе для автофокусировки на основе ROI.

Ключевые слова: синтетические данные, YOLO, детекция объектов, обучение нейронных сетей, объем датасета, Yolo Synthetic Data Generator, автофокусировка.

DOI: 10.51368/1996-0948-2026-4-98-102

Введение

Современные системы компьютерного зрения всё чаще используют глубокие нейронные сети для детектирования объектов. Одним из ключевых факторов, определяющих качество работы таких систем, является наличие размеченного обучающего датасета. Однако ручная разметка изображений трудоёмка и дорогостояща, особенно для специализированных задач, где объекты могут быть редкими или труднодоступными. Альтернативой выступает использование синтетических данных, генерируемых программными средствами с автоматической аннотацией [1].

Ранее автором была разработана программа Yolo Synthetic Data Generator, позво-

ляющая создавать размеченные синтетические изображения с трёхмерными моделями объектов на статических фонах, с контролируемой рандомизацией положения, поворота, экспозиции и гаммы. Программа зарегистрирована в Реестре программ для ЭВМ (свидетельство № 2026667324). В предыдущей работе [2] рассматривалось применение нейросетевой детекции для выделения области интереса (ROI) при автофокусировке оптико-электронной системы. В настоящей статье ставится задача определить оптимальный объем синтетических данных, необходимый для обучения модели детектирования заданного объекта (губки) на различных фонах, с целью обеспечения высокой точности при минимальных вычислительных затратах.

1. Метод генерации синтетических данных

Программа *Yolo Synthetic Data Generator* реализована на C++14 с использованием графического API Vulkan и библиотек *tinyglTF*, *GLM*, *ImGui*. Она позволяет загружать трёхмерные модели в форматах *glTF/GLB*, накладывать их на фоновые изображения (JPG) с сохранением перспективы и освещения (PBR), а также автоматически формировать аннотации в формате YOLO – нормализованные координаты ограничивающих прямоугольников (bounding boxes).

В процессе генерации датасета предусмотрена рандомизация следующих параметров:

- положение объекта в пределах видимой области сцены;
- поворот вокруг трёх осей (диапазон 360°);
- экспозиция (0,1–10,0) и гамма-коррекция (0,1–4,0).

Для обучения были подготовлены семь датасетов, состоящих из 75 различных фоновых изображений, на каждый из которых генерировалось 100, 500, 1000, 1500, 3000, 5000 или 10000 кадров, соответственно. Таким образом, общий объём каждого датасета составлял 7500, 37500, 75000, 112500, 225000, 375000 и 750000 изображений. Во всех случаях использовалась одна и та же трёхмерная модель объекта – губка (размеры 160×55 мм).

Генерация выполнялась с разрешением 1280×1280 пикселей с последующим приведением к формату, принятому для обучения YOLOv8 (640×640). Аннотации сохранялись в текстовых файлах с нормализованными координатами.

2. Экспериментальная методика

Для обучения выбраны три варианта архитектуры YOLOv8: **nano** (наименьшая, 3,2 млн параметров), **small** (11,2 млн) и **medium** (25,9 млн). Каждая модель обучалась на каждом из семи датасетов с одинаковыми гиперпараметрами: количество эпох – 100, размер батча – 16, оптимизатор – SGD с моментом 0,937, начальная скорость обучения – 0,01 с косинусным затуханием. Обучение про-

водилось на графическом процессоре NVIDIA GeForce RTX 4080 (16 ГБ).

Оценка качества выполнялась на пяти тестовых видеороликах длительностью ~2883 кадра каждый (всего около 14415 кадров). Видеоролики содержали ту же модель губки, размещённую в различных условиях освещения и ракурсах, не совпадающих с обучающими фонами. Для каждой обученной модели на каждом видео вычислялись:

- общее количество обнаруженных объектов (*total_detections*);
- число кадров, в которых было хотя бы одно обнаружение (*det_frames*);
- доля кадров с детекциями относительно общего числа кадров;
- время обработки одного видео (*time_sec*).

В качестве порога уверенности использовалось значение 0,7, а порог IOU при подавлении немаксимумов – 0,35. Все замеры проводились на одном и том же вычислительном устройстве.

3. Результаты и обсуждение

В таблице приведены усреднённые по пяти видеороликам показатели для каждой модели и каждого размера датасета. Для наглядности также представлены данные для видео № 4, на котором были зафиксированы максимальные значения детекций.

Анализ таблицы показывает, что для моделей **nano** и **small** максимальное количество обнаружений достигается при размере датасета **1000 кадров**. При дальнейшем увеличении объёма данных наблюдается снижение числа детекций, что может быть связано с переобучением на синтетических данных или с недостаточной вариативностью фонов. Для модели **medium** пик также приходится на 1000 кадров, однако значения при 500 кадрах близки к максимальным, а при 3000 кадрах наблюдается вторичный подъём (возможно, за счёт случайных факторов). Учитывая, что модель **medium** требует больше вычислительных ресурсов, оптимальным для неё можно считать датасет в 500 кадров, обеспечивающий хорошее качество при меньших затратах на обучение.

Таблица

Средние значения по пяти видео (в скобках – значения для видео № 4)

Размер датасета (кадров)	Модель	Total Detections	Det Frames	Время, с
100	nano	442 (596)	442 (596)	48,6
	small	485 (653)	485 (653)	59,1
	medium	602 (771)	602 (771)	67,5
500	nano	476 (641)	476 (641)	48,3
	small	549 (763)	549 (763)	59,0
	medium	695 (866)	695 (866)	67,4
1000	nano	532 (760)	529 (759)	48,0
	small	561 (793)	560 (793)	58,9
	medium	618 (776)	617 (776)	67,1
1500	nano	511 (644)	506 (643)	47,9
	small	533 (658)	531 (658)	58,8
	medium	626 (744)	624 (744)	67,2
3000	nano	399 (480)	395 (480)	48,1
	small	469 (570)	467 (570)	59,0
	medium	620 (790)	617 (790)	67,2
5000	nano	306 (281)	304 (281)	48,0
	small	395 (460)	393 (460)	58,9
	medium	532 (621)	530 (621)	67,3
10000	nano	358 (441)	355 (441)	48,1
	small	448 (547)	446 (546)	59,2
	medium	550 (584)	548 (584)	66,9

Время обработки одного видео (инференс) практически не зависит от размера датасета и определяется только архитектурой модели: nano – ~48 с, small – ~59 с, medium – ~67 с. Это означает, что при выборе модели следует руководствоваться не только качеством, но и требованиями к быстродействию. Так, модель nano обеспечивает приемлемое качество (760

детекций на видео № 4) при минимальном времени, что делает её предпочтительной для систем реального времени, например, для автофокусировки по ROI.

На рисунке представлена зависимость среднего числа обнаружений от размера датасета для каждой модели.

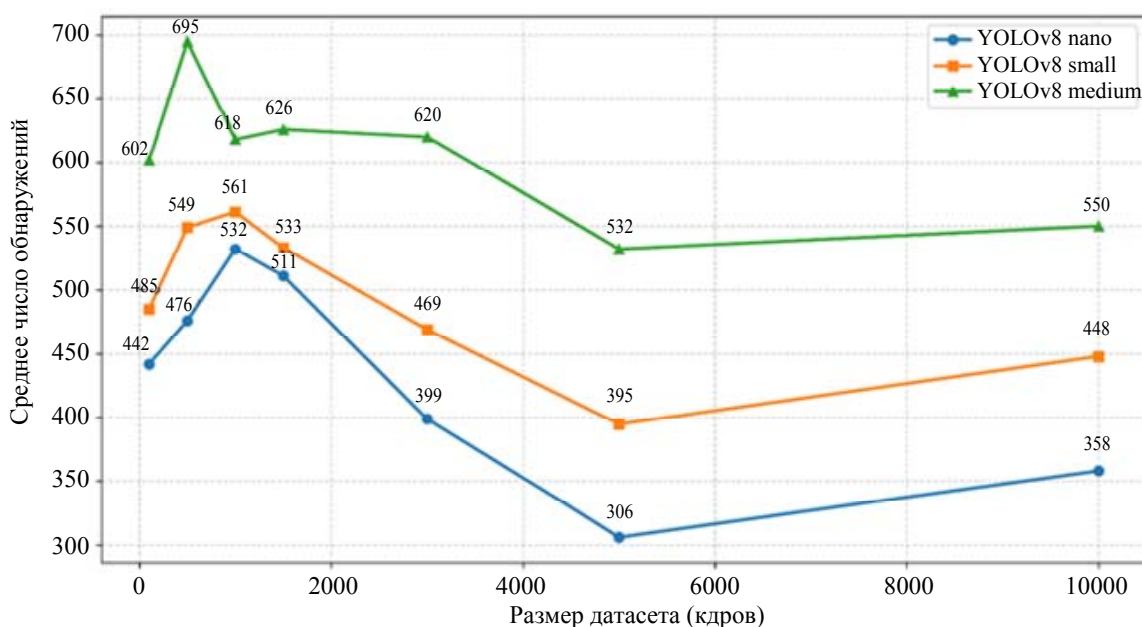


Рис. Зависимость среднего числа обнаружений от объёма датасета (усреднено по пяти видео)

Из рисунка видно, что кривые для всех моделей имеют схожий характер: быстрый рост на начальном участке (до 500–1000 кадров) и последующее насыщение или даже спад. Это подтверждает гипотезу о существовании оптимального объёма данных, превышение которого не улучшает, а иногда ухудшает обобщающую способность модели.

Обсуждение

Полученные результаты согласуются с известными работами по использованию синтетических данных [3, 4], где также отмечается, что качество модели перестаёт расти после некоторого порога, а дальнейшее увеличение датасета может приводить к переобучению на артефактах синтеза. В нашем случае спад качества при больших датасетах (3000–10000 кадров) вероятно связан с тем, что модель начинает запоминать специфические особенности фонов и ракурсов, которые не встречаются в тестовых видеороликах, что снижает её способность к обобщению.

Следует отметить, что синтетические данные генерировались с ограниченным разнообразием (75 фонов, один объект). Возможно, при увеличении числа фонов или вариаций объекта оптимальный объём данных мог бы возрасти. Тем не менее, для практических задач, подобных нашей (детекция одного объекта на ограниченном наборе фонов), полученные рекомендации являются обоснованными.

Применение результатов данной работы в системах автофокусировки [2] позволяет сократить время на ручную разметку и ускорить процесс разработки. Используя синтетический датасет объёмом 1000 кадров и модель YOLOv8 nano, можно добиться высокой точности обнаружения объекта в реальном времени, что обеспечивает корректное выделение ROI для наведения фокуса.

Заключение

В работе исследовано влияние объёма синтетических данных на качество обучения

нейросетей детектирования объектов. На основе экспериментов с моделями YOLOv8 nano, small и medium на датасетах различного размера установлено:

- для моделей nano и small оптимальный размер датасета составляет ~1000 кадров (при 75 фонах), что обеспечивает максимальное число обнаружений на тестовых видео;
- модель medium показывает близкие результаты уже при 500 кадрах, однако её использование оправдано только при необходимости повышенной точности и наличии достаточных вычислительных ресурсов;
- дальнейшее увеличение объёма данных (свыше 1000–1500 кадров) не приводит к улучшению качества, а в некоторых случаях ухудшает его, что свидетельствует о переобучении.

Практическая значимость работы заключается в формировании рекомендаций по выбору объёма синтетического датасета для задач детекции одного объекта на статических фонах. В сочетании с разработанным генератором синтетических данных это позволяет существенно сократить трудозатраты на подготовку размеченных выборок и ускорить разработку систем технического зрения, включая системы автофокусировки на основе ROI.

Дальнейшие исследования планируется направить на расширение вариативности синтетических данных (увеличение числа фонов, добавление различных объектов, изменение условий освещения) и изучение влияния этих факторов на оптимальный объём датасета.

СПИСОК ЛИТЕРАТУРЫ

1. Николенко С. И., Кадури А. А., Архангельская Е. О. Глубокое обучение. – СПб.: Питер, 2018.
2. Берилло А. А. / Прикладная физика. 2026. № 2. С. 82–85.
3. Tremblay J., Prakash A., Acuna D., et al. Training deep networks with synthetic data: Bridging the reality gap by domain randomization / arXiv:1804.06516, 2018.
4. Dwibedi D., Misra I., Hebert M. / ICCV. 2017. P. 1301–1310.
5. Redmon J., Farhadi A. YOLOv3: An incremental improvement / arXiv:1804.02767, 2018.

PACS: 42.30.Va, 42.79.Bh, 07.05.Mh

Determination of the optimal volume of synthetic data for training object detection neural networks (using YOLOv8 as an example)

A. A. Berillo

RD&P Center ORION, JSC, Moscow, 111538 Russia

E-mail: berillo.andrey@yandex.ru

Received 25.06.2026; revised 9.07.2026; accepted 06.08.2026

The problem of determining the minimum sufficient volume of synthetic data for training YOLO-format object detection neural networks is considered. An approach to generating labeled synthetic images using the developed Yolo Synthetic Data Generator software module is proposed. Three variants of the YOLOv8 model (nano, small, medium) were trained on datasets of various sizes (from 100 to 10,000 frames). Quality assessment was performed on five test videos based on the number of detected objects and frames with detections. It is shown that for the nano and small models, a dataset of ~1,000 frames is optimal, while for the medium model, ~500 frames are sufficient. A further increase in the dataset does not lead to a significant improvement in quality but increases computational costs. The obtained results can be used in the development of machine vision systems, including for ROI-based autofocusing.

Keywords: synthetic data, YOLO, object detection, neural network training, dataset size, Yolo Synthetic Data Generator, autofocus.

REFERENCES

1. Nikolenko S. I., Kadurin A. A., and Arkhangelskaya E. O. Deep Learning. St. Petersburg: Piter Publ., 2018 [in Russian].
2. Berillo A. A., Applied Physics, № 2, 82–85 (2026) [in Russian].
3. Tremblay J., Prakash A., Acuna D., et al., Training deep networks with synthetic data: Bridging the reality gap by domain randomization, arXiv:1804.06516, 2018.
4. Dwibedi D., Misra I., and Hebert M., Cut, paste and learn: Surprisingly easy synthesis for instance detection ICCV, 1301–1310 (2017).
5. Redmon J. and Farhadi A., YOLOv3: An incremental improvement, arXiv:1804.02767, 2018.

Об авторе

Берилло Андрей Анатольевич, ведущий инженер-программист, магистр, АО «НПО «Орион» (111538, Россия, Москва, ул. Косинская, 9). E-mail: berillo.andrey@yandex.ru