

УДК 004.852

Применение генеративных моделей для создания 3D-сред в задачах мультиагентного обучения

Арутин Виктор Валерьевич

arutinvv@student.bmstu.ru

SPIN-код: 6159-4618

Алфимцев Александр Николаевич

alfim@bmstu.ru

SPIN-код: 7802-7020

МГТУ им. Н.Э. Баумана, Москва, Россия

Аннотация. Рассмотрены современные подходы к генерации 3D-сред для задач мультиагентного обучения. Выполнен сравнительный анализ моделей Genie 3 и Hunyuan3D 2.0 с оценкой их архитектуры, возможностей и ограничений. Показано, что Hunyuan3D 2.0 эффективно создает структурированные 3D-объекты по 2D-изображениям, при этом разрешение октодерев и масштаб направления определяют четкость границ, необходимую для навигации агентов. Сделаны выводы, что сгенерированные вариации сред пригодны для расширения обучающих выборок в MARL-симуляциях, а открытые модели рекомендованы для первого этапа внедрения процедурно генерируемого контента.

Ключевые слова: мультиагентное обучение, генеративные модели, Genie 3, Hunyuan 3D 2.0, разрешение октодерев, масштаб направления

Application of generative models to create 3D environments in multi-agent learning problems

Arutin Victor Valerievich

arutinvv@student.bmstu.ru

SPIN-code: 6159-4618

Alfimtsev Alexander Nicolaevich

alfim@bmstu.ru

SPIN-code: 7802-7020

BMSTU, Moscow, Russia

Abstract. Modern approaches to generating 3D environments for multi-agent learning tasks are considered. A comparative analysis of the Genie 3 and Hunyuan3D 2.0 models is performed, evaluating their architecture, capabilities, and limitations. It is shown that Hunyuan3D 2.0 effectively creates structured 3D objects from 2D images, with the octree resolution and direction scale determining the boundary clarity required for agent navigation. It is concluded that the generated environment variations are suitable for expanding training samples in MARL simulations, and the open models are recommended for the first stage of implementing procedurally generated content.

Keywords: multi-agent learning, generative models, Genie 3, Hunyuan3D 2.0, octree resolution, guidance scale

Введение. Мультиагентное обучение (MARL) — динамично развивающееся направление ИИ, где ключевым фактором успеха выступает среда взаимодействия агентов. Сложные и динамичные среды способствуют выработке адаптивных стратегий, тогда как статичные приводят к переобучению и снижению обобщающей способности. Традиционные детерминированные модели не позволяют агентам адаптироваться к вариативности реального мира.

Необходимость преодоления этих ограничений обуславливает переход к динамическим окружениям. Эволюция подходов прошла путь от ручного проектирования к процедурной генерации и применению генеративных нейросетевых моделей (GAN, VAE, диффузионные модели) [1]. Данный подход обеспечивает автоматический синтез сложных трехмерных пространств, а его новизна заключается в создании бесконечного множества уникальных контролируемых окружений, позволяющих агентам вырабатывать универсальные стратегии поведения, эффективные за пределами обучающей выборки.

Материалы и методы. Исследования Google DeepMind (проект SIMA) демонстрируют системный подход к интеграции 3D-сред в мультиагентное обучение [2, 3]. Ключевые преимущества трехмерных окружений: богатый контекст восприятия (оценка глубины и перспективы), универсальность интерфейса (единый набор действий для разных сред) и безопасность экспериментов (многократный перезапуск без последствий), что необходимо для задач мультиагентного обучения.

В задачах генерации 3D-сред выделяются две модели. Genie 3 (Google DeepMind, 2025) — фундаментальная модель мира на основе авторегрессионной диффузии, генерирующая интерактивные окружения с физикой и анимацией [4]. Ограничения: малое время жизни сред, отсутствие моделирования взаимодействий агентов и закрытый код. Hunyuan3D 2.0 (Tencent, 2025) — двухэтапный конвейер (Hunyuan3D-DiT для геометрии, Hunyuan3D-Paint для текстур), ориентированный на синтез детализированных 3D-объектов с открытым кодом [5].

Для эксперимента выбрана модель Hunyuan3D 2.0 благодаря доступности (открытый код на GitHub/Hugging Face), пригодности для MARL (генерация статичных структурированных объектов) и возможности варьирования параметров [6]. Развертывание выполнено локально через клонирование репозитория и запуск установочных скриптов.

Ключевые параметры генерации: `steps` (итерации диффузии), `octree_resolution` (детализация геометрии), `guidance_scale` (соответствие изображению) и `seed` (воспроизводимость). Цель эксперимента — оценка их влияния на качество 3D-моделей для использования в качестве стационарных MARL-сред.

Входное изображение выбрано из карт StarCraft II, традиционно применяемых для оценки алгоритмов MARL благодаря сложной архитектуре [7]. Библиотека SMAC подтверждает их пригодность [8]. Использована карта Eastwatch, содержащая коридоры, перекрестки, тупики и открытые пространства, что репрезентативно для отработки навигационных навыков [9].

Результаты. На основе изображения карты Eastwatch были сгенерированы несколько 3D-объектов, каждый из которых был определен уникальным набором параметров, как показано в таблице. Примеры сгенерированных объектов представлены на рис. 1 и 2.

Параметры экспериментальных объектов

Версия объекта	Параметры			
	steps	octree_resolution	guidance_scale	seed
Объект 1	30	256	5.5	1234
Объект 2	40	256	5.5	1234
Объект 3	30	256	11	1234
Объект 4	30	512	5.5	1234
Объект 5	40	512	11	1234



Рис. 1. Результат генерации объекта 1

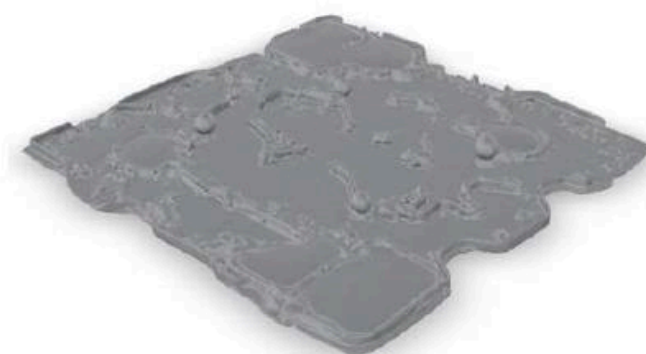


Рис. 2. Результат генерации объекта 5

Обсуждение. Анализ зависимостей, полученных в ходе экспериментального тестирования модели Hunyuan3D 2.0, позволяет выявить закономерности, имеющие прямое значение для задач мультиагентного обучения. Увеличение количества шагов диффузии с 30 до 40 привело к более точной проработке мелких деталей топологии лабиринта, однако критическими факторами выступают разрешение октодерева и масштаб направления. Повышение данных параметров с 256 до 512 и с 5.5 до 11 соответственно обеспечивает качественное улучшение геометрической четкости границ объектов —

коридоры приобретают выраженные стенки, перекрестки сохраняют ортогональность, а тупики не «заплывают» в результате сглаживания. Для задач MARL, особенно на этапе навигационного обучения, именно четкость границ является определяющим фактором: агенты должны однозначно идентифицировать проходимые и непроходимые области, что при низком разрешении становится проблематичным из-за размытия геометрии [10]. Полученные результаты полностью коррелируют с теоретическими положениями, изложенными в оригинальной статье разработчиков Hunyuan3D 2.0, где авторы указывают на прямую зависимость качества реконструкции геометрии от вычислительных ресурсов, выделяемых на этапе построения октодеревя [11]. Экспериментально подтверждено, что заявленная разработчиками способность модели к высокодетализированной генерации реализуется в полном объеме при достижении пороговых значений.

Сгенерированные модели объектов демонстрируют полное соответствие структурным требованиям, предъявляемым к мультиагентным средам навигационного типа. В полученных 3D-моделях отчетливо идентифицируются все ключевые элементы, необходимые для постановки обучающих задач: разветвленная сеть коридоров различной ширины, перекрестки с множеством направлений выбора, тупиковые зоны, требующие от агентов способности к развороту и перестроению маршрута, а также открытые пространства, пригодные для отработки групповых взаимодействий. Для первого этапа интеграции генеративных технологий в конвейеры мультиагентного обучения именно открытые модели с предсказуемым поведением представляются предпочтительным инструментом, позволяющим выстроить надежную систему экспериментов с дальнейшим улучшением до моделирования сложных динамических сред.

Заключение. В результате проведенного исследования, включавшего анализ современных подходов к генерации 3D-сред и экспериментальное тестирование модели Hunyuan3D 2.0, могут быть сформулированы следующие основные выводы. Во-первых, экспериментально показано, что генеративные модели на основе диффузионных архитектур (на примере Hunyuan3D 2.0) позволяют эффективно создавать сложные структурированные 3D-объекты по двумерным изображениям-референсам, сохраняя ключевые топологические характеристики исходных сред. Во-вторых, установлено, что среди набора параметров генерации наибольшее влияние на пригодность получаемых моделей для задач навигации агентов оказывают разрешение октодеревя и масштаб направления: достижение пороговых значений обеспечивает четкость геометрических границ, необходимую для однозначной идентификации проходимых и непроходимых областей агентами. В-третьих, подтверждено, что сгенерированные вариации карты Eastwatch полностью сохраняют структурные элементы, требуемые для постановки мультиагентных задач (коридоры, перекрестки, тупики, открытые пространства), и могут быть непосредственно использованы для расширения обучающей выборки в MARL-симуляциях, таких как SMAC.

Полученные результаты подтверждают исходную гипотезу исследования: интеграция генеративных моделей в процесс создания обучающих окружений позволяет существенно повысить вариативность сред, доступных для мультиагентного обучения, преодолевая ограничения статических и детерминированных подходов. Несмотря на то, что новейшие динамические модели, такие как Genie 3, открывают перспективы генерации интерактивных сред с физическим взаимодействием, их закрытый статус ограничивает воспроизводимость исследований. В этом контексте открытые модели класса Hunyuan3D 2.0 представляются оптимальным инструментом для первого этапа внедрения генеративных технологий в исследовательскую практику. Дальнейшие перспективы работы включают в себя как интеграцию сгенерированных 3D-объектов в физические симуляторы для проверки переносимости стратегий агентов в среды с реалистичной физикой, так и систематическое тестирование политик агентов, обученных исключительно на процедурно сгенерированном контенте, с целью оценки их обобщающей способности при переносе в неизвестные ранее окружения.

Работа выполнена в рамках НИР Госзадание FSFN-2024-0059.

Список источников

- [1] Каптерев А.И., Ромашкова О.Н., Чискидов С.В., Ермакова Т.Н. Современное состояние и перспективы моделирования цифровых профессиональных пространств в бизнесе и образовании. *Информатизация образования*, 2024, т. 20, № 4, с. 358–372.
- [2] *Официальный сайт лаборатории Google DeepMind*. URL: <https://deepmind.google/> (дата обращения 20.01.2026).
- [3] Abi Raad M., Ahuja A., Barros C., Besse F., Bolt A. Scaling Instructable Agents Across Many Simulated Worlds. URL: <https://deepmind.google/discover/blog/sima-generalist-ai-agent-for-3d-virtual-environments/> (accessed 20.01.2026).
- [4] *Официальная документация Google DeepMind Genie 3*. URL: <https://deepmind.google/blog/genie-3-a-new-frontier-for-world-models/> (accessed 21.01.2026).
- [5] *Официальная документация Tencent Hunyuan3D 2.0*. URL: <https://github.com/Tencent-Hunyuan/Hunyuan3D-2> (accessed 22.01.2026).
- [6] *Официальное хранилище исходных данных для Tencent Hunyuan3D 2.0*. URL: <https://huggingface.co/tencent/Hunyuan3D-2> (accessed 22.01.2026).
- [7] Young Joon Park, Yoon Sang Cho, Seoung Bum Kim. Multi-agent reinforcement learning with approximate model learning for competitive games. *PLoS ONE*, 2019, vol. 14(9), art. no. e0222215. <https://doi.org/10.1371/journal.pone.0222215>
- [8] *Официальная документация библиотеки SMAC*. URL: <https://github.com/oxwhirl/smac/blob/master/docs/smac.md> (accessed 24.01.2026).
- [9] *Исходные данные карты Eastwatch, StarCraft II Maps wiki*. URL: <https://starcraft.fandom.com/wiki/Eastwatch> (accessed 25.01.2026).
- [10] Bruce J., Dennis M., Edwards A. Genie: Generative Interactive Environments. DeepMind, arXiv:2402.15391v1 [cs.LG], 23 Feb 2024. <https://doi.org/10.48550/arXiv.2402.15391>
- [11] Chunchao Guo, Jingwei Huang, Zibo Zhao. Hunyuan3D 2.0: Scaling Diffusion Models for High Resolution Textured 3D Assets Generation. Tencent, arXiv:2501.12202v3 [cs.CV], 26 Feb 2025. <https://doi.org/10.48550/arXiv.2501.12202>