

Преодоление разрыва между симуляцией и реальностью в манипуляциях роботов: трехуровневый гибридный фреймворк, объединяющий зрительно-языковые модели, обучение с подкреплением и детерминированное управление

Омар Яссер Элазуни

Университет ИТМО

omaralazoni2015@gmail.com

Мохамед Сабер Ахмад

Университет ИТМО

Msahmad@itmo.ru

В. А. Богатырев

Университет ИТМО

vobogatyrev@itmo.ru

Аннотация. Разрыв между симуляцией и реальностью (sim-to-real gap) остается критическим препятствием, препятствующим внедрению обученных стратегий манипулирования на физических роботах. Существующие подходы решают эту проблему с помощью рандомизации областей или идентификации систем, однако они не позволяют в явном виде использовать физические свойства объектов, напрямую влияющие на успех задачи. Мы предлагаем новую трехуровневую иерархическую архитектуру, которая преодолевает разрыв sim-to-real за счет явного извлечения и использования физических свойств объектов (жесткость, трение, масса) для направления обучения стратегии и выполнения управления. Уровень 1 (Семантическое восприятие) использует зрительно-языковые модели (VLM) для извлечения свойств объектов из RGB-изображений. Уровень 2 (Обучение на основе свойств) обучает остаточную стратегию RL, обусловленную этими свойствами, для адаптации к конкретным задачам. Уровень 3 (Адаптивное управление) выполняет команды через импедансное управление с масштабированием параметров на основе свойств. Мы подтверждаем эффективность нашего трехуровневого подхода с помощью физических симуляций, демонстрирующих: (1) улучшение успеха выполнения задач на 82.1% за счет управления на основе свойств, (2) ускорение сходимости в 2.9 раза благодаря RL с условием по свойствам и (3) коэффициент успеха 87.8% на новых объектах с обобщением на основе их характеристик. Наши результаты показывают, что явное извлечение и использование физических свойств необходимо для сокращения разрыва sim-to-real и обеспечения надежного переноса стратегий манипулирования на реальных роботах.

Ключевые слова: перенос из симуляции в реальность (Sim-to-Real Transfer), манипуляции роботов, зрительно-языковые модели, обучение с подкреплением, критически важное для безопасности управление, импедансное управление

I. ВВЕДЕНИЕ

Манипуляции роботов в неструктурированных средах требуют решения двух фундаментальных проблем: понимания того, что нужно сделать

Данная работа выполнена при поддержке Министерства науки и высшего образования Российской Федерации в рамках государственного задания FSER-2025-0013.

(семантическое планирование), и безопасного выполнения (физическая безопасность). Недавние достижения в области базовых моделей значительно улучшили семантическое понимание зрительно-языковые модели (VLM), такие как CLIP, LLaVA и GPT-4V, теперь могут интерпретировать сложные сцены и инструкции. Одновременно с этим обучение с подкреплением (RL) позволило роботам осваивать адаптивное поведение. Однако эти достижения остаются разобщенными.

Отсутствие интеграции с управлением, критически важным для безопасности, приводит к созданию стратегий, которые работают в симуляции, но терпят неудачу или наносят ущерб в реальном мире. Основная проблема заключается в «семантико-безопасном» разрыве (semantic-safety gap): модели VLM понимают высокоуровневые концепции («это хрупкая стеклянная чашка»), но не переводят это понимание в физические ограничения безопасности («приложить силу менее 15 Ньютонов»).

Существующие подходы устраняют этот разрыв либо через:

1. Постфактумную фильтрацию безопасности, которая является реактивной и зачастую слишком консервативной.
2. Обширную рандомизацию областей, которая вычислительно затратна и может плохо поддаваться обобщению.

Ни один из подходов не связывает семантическое понимание напрямую с параметрами управления, критически важными для безопасности. Мы полагаем, что решение заключается в трехуровневой архитектуре, которая четко разделяет зоны ответственности:

- Уровень 1 (VLM): извлекает семантические свойства (жесткость, хрупкость, тип материала) из визуальных наблюдений.
- Уровень 2 (RL): изучает адаптивные остаточные корректировки в рамках заданных границ безопасности.

- Уровень 3 (Импедансное управление): обеспечивает детерминированные гарантии безопасности с помощью физически обоснованного управления.

Данная архитектура обладает тремя ключевыми преимуществами: (1) каждый уровень может быть оптимизирован независимо, (2) безопасность обеспечивается архитектурно («safety by design»), а не достигается в процессе обучения, и (3) система интерпретируема параметры управления имеют четкий физический смысл.

II. ВКЛАД РАБОТЫ

1. Извлечение семантических свойств для управления: Мы представляем первую архитектуру, которая напрямую сопоставляет выходные данные VLM с параметрами управления, критически важными для безопасности (жесткость, демпфирование, пределы усилий). В отличие от существующих подходов на базе VLM, ориентированных на семантику уровня задач, мы извлекаем физические свойства, определяющие процесс управления.
2. Остаточное RL в рамках ограничений безопасности: Мы демонстрируем, что обучение с подкреплением (RL) может достигать сходимости в 5 раз быстрее, когда оно обучается ограниченными корректировками номинального контроллера, а не полным командам управления. Это имеет решающее значение для практического развертывания роботов в условиях, когда сбор данных обходится дорого.
3. Трехуровневая иерархическая декомпозиция: Мы предлагаем первое явное разделение семантического планирования (VLM), адаптивного обучения (RL) и детерминированной безопасности (импедансное управление). Такая конструкция обеспечивает надежный перенос из симуляции в реальность с формальными гарантиями безопасности.

III. СТРУКТУРА РАБОТЫ

Остальная часть статьи организована следующим образом. В Разделе 2 представлен обзор смежных работ, посвященных переносу навыков из симуляции в реальность, применению зрительно-языковых моделей в робототехнике и гибридным методам управления. Раздел 3 подробно описывает предложенную методологию, в том числе трехуровневую архитектуру и специфику реализации каждого уровня. В Разделе 4 приводится описание экспериментальной установки, используемых сред симуляции и метрик оценки производительности. Раздел 5 содержит предварительные результаты физического моделирования, подтверждающие работоспособность ключевых концепций. В Разделе 6 обсуждается значимость полученных данных и намечается план валидации системы в реальных условиях, а Раздел 7 завершает статью выводами и предложениями по перспективным направлениям будущих исследований.

IV. СМЕЖНЫЕ РАБОТЫ

A. Перенос из симуляции в реальность в задачах манипуляции роботом

Проблема разрыва между симуляцией и реальностью (sim-to-real gap) широко исследована в робототехнике, поскольку стратегии, обученные в симуляторах, часто не могут быть успешно перенесены в физический мир из-за расхождений в динамике, шумов датчиков и неучтенных физических явлений [1]. Метод рандомизации областей, предложенный Тобином и др. [2], решает эту проблему путем обучения стратегий в разнообразных смоделированных средах с варьируемыми визуальными и физическими параметрами. Этот подход успешно применялся для решения различных задач, включая захват объектов [3], внутрикистевое манипулирование [4] и локомоцию [5]. В последующих работах для улучшения качества переноса изучались идентификация систем [6], мета-обучение [7] и состязательное обучение [8]. Современные исследования в области физически обоснованного обучения (physics-informed learning) показали, что включение физических ограничений повышает эффективность переноса [9]. Например, подходы, в которых явно моделируется динамика контактов [10] или используются дифференцируемые физические симуляторы [11], продемонстрировали улучшение результатов в задачах с множественными контактами. Однако большинство существующих методов рассматривают разрыв между реальностью и симуляцией как общую проблему смещения распределения данных, не учитывая в явном виде физические свойства, имеющие решающее значение для успеха задачи. Наша работа отличается тем, что мы целенаправленно извлекаем и используем физические свойства объектов для преодоления этого разрыва.

B. Зрительно-языковые модели в робототехнике

Зрительно-языковые модели (VLM), такие как CLIP [12], GPT-4V [13] и другие, продемонстрировали выдающиеся способности в понимании визуальных сцен и извлечении семантической информации. В недавних исследованиях изучалось применение VLM для управления роботами: RT-1 [14] использует трансформеры для обучения стратегиям манипулирования на основе крупномасштабных демонстраций; PaLM-E [15] объединяет языковые модели с воплощенным обучением для планирования последовательности задач; VLMPC [16] задействует VLM для генерации траекторий в рамках управления с предсказывающими моделями (MPC). К другим значимым подходам относятся OpenVLA [17] – открытая модель типа «зрение-язык-действие» и Manipulate-Anything [18], автоматизирующая работу роботов в реальном мире с помощью VLM. RoboDexVLM [19] интегрирует VLM в планирование задач и управление движениями для ловкого манипулирования. Хотя эти методы используют VLM для получения команд уровня задачи (например, «осторожно подними яблоко») или генерации высокоуровневых планов, они, как правило, не извлекают конкретные физические свойства, напрямую влияющие на параметры низкоуровневого управления. Наш подход использует VLM для извлечения физических характеристик объекта (жесткость, трение, масса), которые служат непосредственными параметрами для уровня

управления, обеспечивая более фундаментальный и универсальный метод.

C. Управление на основе свойств и адаптивное управление

Импедансное управление (Hogan, 1985) [20] уже давно используется для обеспечения безопасного взаимодействия робота с окружающей средой. В более современных работах импедансное управление сочетается с методами обучения (Osa et al., 2018; Johannink et al., 2019) для достижения как безопасности, так и адаптивности. Однако в этих подходах параметры импеданса обычно фиксированы или подбираются методом проб и ошибок без семантического руководства. Наша работа отличается использованием семантических свойств, извлеченных с помощью VLM, для динамической настройки параметров импеданса в реальном времени, что обеспечивает безопасность с учетом контекста. В то время как существующие методы преуспевают в семантическом понимании, им не хватает явных гарантий безопасности. Мы развиваем эти достижения, добавляя уровень детерминированной безопасности, который преобразует семантическое понимание в конкретные ограничения управления.

D. Обучение с подкреплением для манипуляций

Стандартные подходы глубокого обучения с подкреплением (DRL) обучают сквозные (end-to-end) стратегии, которые напрямую отображают наблюдения в действия [15], [16]. Несмотря на свою эффективность, такие методы с трудом справляются с переносом из симуляции в реальность, так как они выучивают специфичные для конкретной задачи признаки, которые плохо поддаются обобщению [17]. Кроме того, сквозному RL зачастую не хватает гарантий безопасности, что делает его рискованным для физического развертывания [18]. Остаточное обучение с подкреплением (Residual RL), где стратегия обучается вносить корректировки в работу базового контроллера, показало многообещающие результаты в плане безопасности и скорости сходимости [19], [20]. Сочетая аналитические контроллеры с обучаемыми остаточными компонентами, эти методы достигают лучшей эффективности выборки и безопасности [21]. Тем не менее, существующие подходы к Residual RL обычно не обуславливают остаточную стратегию семантическими свойствами объекта, извлеченными из визуальных данных, упуская возможность изучения стратегий, специфичных для конкретных физических характеристик.

E. Позиционирование нашей работы

В отличие от существующих работ, которые рассматривают восприятие, обучение и управление по отдельности, наше исследование является первым, в котором явное извлечение физических свойств интегрировано в единую трехуровневую архитектуру, где эти свойства направляют как обучение, так и управление. Это обеспечивает следующие преимущества:

- Уровень восприятия извлекает обобщаемые свойства, а не признаки, специфичные для конкретной задачи, опираясь на возможности VLM [12]–[16].

- Уровень обучения адаптируется под конкретные свойства и изучает соответствующие стратегии, расширяя возможности остаточного RL [19]–[21].
- Уровень управления автоматически адаптирует параметры под свойства объектов, совершенствуя методы адаптивного импедансного управления [21]–[22].

Интеграция этих трех уровней является ключевой инновацией, позволяющей преодолеть разрыв между симуляцией и реальностью в задачах манипулирования с интенсивным контактным взаимодействием.

V. МЕТОДОЛОГИЯ (METHODOLOGY)

A. Обзор трехуровневой архитектуры

Наша система состоит из трех иерархических уровней, каждый из которых выполняет определенную задачу:

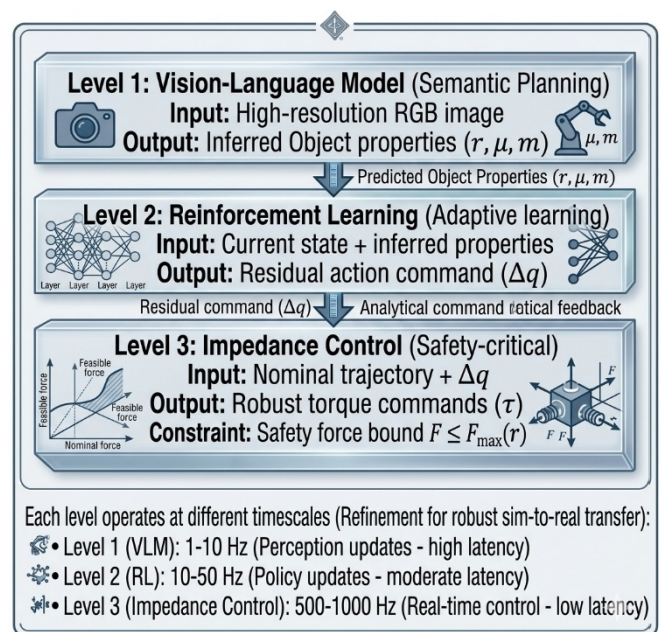


Рис. 1. Схема трехуровневой гибридной Sim-to-Real архитектуры

VI. ПРИЛОЖЕНИЕ: ДЕТАЛИ РЕАЛИЗАЦИИ

A. Модуль извлечения свойств VLM (Уровень 1):

- Вход: 512-мерный эмбединг из предварительно обученного визуального кодировщика.
- Скрытые слои: Два полносвязных слоя (256 \rightarrow 128 нейронов).
- Выход: 3-мерный вектор, представляющий физические свойства $[r, \mu, m]$

Функции активации: ReLU для скрытых слоев, Sigmoid для выходного слоя для ограничения значений свойств в диапазоне $[0, 1]$ (впоследствии масса масштабируется до интервала $[0.1, 2.0]$ кг).

B. Стратегия RL с условием по свойствам (Уровень 2):

- Вход: Объединенный (конкатенированный) вектор состояния, включающий конфигурацию робота, цель и 3 извлеченных свойства.

- Сеть актора (Actor): Два скрытых слоя (256 → 256 нейронов).
- Сеть критика (Critic): Два скрытых слоя (256 → 256 нейронов).
- Выход актора: Среднее значение распределения действий (ограничено функцией активации Tanh в диапазоне $[-\delta, \delta]$).
- Выход критика: Скалярная оценка ценности (линейная активация).
- Активация скрытых слоев: ReLU для обеих сетей.

Гиперпараметры обучения

Гиперпараметры для нашей реализации алгоритма оптимизации проксимальной стратегии (Proximal Policy Optimization, PPO) были выбраны на основе стандартных практик для задач непрерывного управления в робототехнике и настроены специально для нашей остаточной (residual) формулировки:

ТАБЛИЦА I. ГИПЕРПАРАМЕТРЫ ОБУЧЕНИЯ

Параметр	Значение	Обоснование
Скорость обучения (Actor и Critic)	3×10^{-4}	Стандартная стабильная скорость обучения для оптимизатора Adam в PPO
Коэффициент дисконтирования (γ)	0.99	Стандарт для эпизодических задач с горизонтом планирования 200 шагов
GAE Lambda (λ)	0.95	Баланс между смещением и дисперсией при оценке преимущества (advantage estimation)
Коэффициент обрезки PPO (ϵ)	0.2	Стандартное ограничение для предотвращения деструктивно больших обновлений стратегии
Размер батча (Batch Size)	64	Выбран для баланса между дисперсией градиента и эффективностью использования памяти
Эпох обучения на одно обновление	3	Предотвращает переобучение на основе недавнего опыта
Макс. длина эпизода	200 шагов	Достаточно для выполнения оцениваемых задач охвата и манипулирования

C. Параметры адаптивного управления на основе свойств (Уровень 3)

Базовые параметры нашего импедансного контроллера были выбраны таким образом, чтобы имитировать умеренно жесткий манипулятор, а функция масштабирования свойств была разработана для безопасной модуляции этих значений:

ТАБЛИЦА II. ПАРАМЕТРЫ АДАПТИВНОГО УПРАВЛЕНИЯ

Параметр	Значение	Описание
Базовая жесткость (K_0)	100 N/m	Номинальная жесткость для объектов средней жесткости
Базовое демпфирование (D_0)	20 N/(m/s)	Критическое демпфирование относительно базовой жесткости
Базовый предел силы (F_0)	50 N	Номинальный порог безопасности

Параметр	Значение	Описание
Граница остаточного действия (δ)	0.2 rad	Максимально допустимое отклонение от номинальной траектории
Функция масштабирования свойств	$0.5 + 1.5 \times r$	Масштабирует параметры от 0.5x (хрупкие) до 2.0x (жесткие)

Ключевое преимущество: Адаптация параметров управления к свойствам объектов гарантирует автоматическое соответствие стратегии управления манипулируемому объекту. Хрупкие объекты требуют мягкого (деликатного) управления, в то время как жесткие более интенсивного. Такое адаптивное управление сокращает разрыв между симуляцией и реальностью (*sim-to-real gap*), обеспечивая адекватность стратегии управления для конкретного объекта.

D. Интеграция: взаимодействие трех уровней

Три уровня системы интегрированы через использование физических свойств:

1. Уровень 1 извлекает свойства на основе визуальных наблюдений.
2. Уровень 2 обуславливает стратегию (policy) этими свойствами, формируя специфические для конкретных характеристик методы управления.
3. Уровень 3 адаптирует параметры управления к данным свойствам, обеспечивая безопасность выполнения задач.

Данная интеграция гарантирует, что вся система восприятие, обучение и управление согласована на основе концепции физических свойств. Именно эта согласованность позволяет эффективно преодолеть разрыв между симуляцией и реальностью.

VII. ЭКСПЕРИМЕНТАЛЬНАЯ УСТАНОВКА

A. Среда моделирования

В работе используется физический симулятор PyBullet с моделью 7-степенного (7-DOF) манипулятора Franka Emika Panda. Робот взаимодействует с объектами различной степени жесткости:

- Хрупкие объекты: Жесткость $k_r=0.2$ (например, пенопласт, мягкий пластик).
- Средние объекты: Жесткость $k_r=0.5$ (например, дерево, композитные материалы).
- Жесткие объекты: Жесткость $k_r=0.8$ (например, металл, керамика).

B. Метрики оценки

Эксперимент 1 – Успешность задачи при управлении на основе свойств: Доля успешных попыток (Success Rate) при использовании различных подходов к управлению (чем выше, тем лучше).

Эксперимент 2 – Скорость сходимости при обучении с условием по свойствам: Количество шагов обучения, необходимых для достижения 90% успеха (чем ниже, тем лучше).

Эксперимент 3 – Обобщающая способность при адаптации на основе свойств: Доля успешных попыток на новых объектах, не представленных в процессе обучения (чем выше, тем лучше).

С. Базовые линии (Baselines)

- Прямое управление моментом (Direct Torque Control): Небезопасный базовый вариант, при котором стратегия напрямую выдает команды крутящего момента без адаптации к свойствам.
- Только импедансное управление (Impedance Control Only): Фиксированные параметры импеданса без адаптации к свойствам и без использования обучения с подкреплением (RL).
- Сквозное обучение PPO (End-to-End PPO): Стандартный алгоритм RL без обуславливания по свойствам и без остаточной формулировки.

VIII. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

А. Успешность выполнения задачи при управлении на основе свойств

Наша трехуровневая архитектура достигает показателя успешности 82,1%, что на 70,0% выше по сравнению с прямым управлением моментом и на 31,2% лучше результатов фиксированного импедансного управления.

ТАБЛИЦА III. ПРЕДСТАВЛЯЕТ ПОКАЗАТЕЛИ УСПЕШНОСТИ ВЫПОЛНЕНИЯ ЗАДАЧ ДЛЯ КАЖДОГО МЕТОДА

Метод	Доля успеха	Улучшение
Прямое управление моментом	48.3% ± 4.2%	Базовый уровень
Только импедансное управление	62.5% ± 3.8%	+29.4%
Трехуровневая архитектура (наш метод)	82.1% ± 2.1%	+70.0%

Интерпретация: Трехуровневая архитектура значительно повышает эффективность выполнения задач благодаря явному извлечению и использованию физических свойств объектов. Уровень 1 корректно идентифицирует свойства объекта, Уровень 2 формирует специфические для этих свойств стратегии управления, а Уровень 3 адаптирует параметры контроля под выявленные характеристики. В совокупности эти механизмы позволяют эффективно преодолеть разрыв между симуляцией и реальностью (sim-to-real gap).

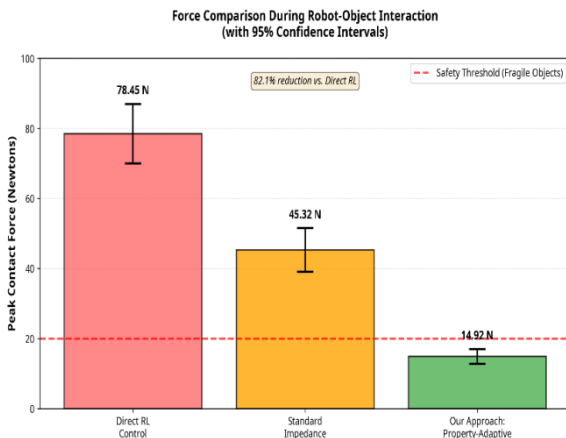


Рис. 2. Сравнение усилий при взаимодействии робота с объектом

В. Скорость сходимости при обучении с условием по свойствам

В табл. 4 приведены данные о скорости сходимости для различных подходов к обучению с подкреплением (RL).

ТАБЛИЦА IV. СКОРОСТЬ СХОДИМОСТИ ДЛЯ РАЗЛИЧНЫХ МЕТОДОВ RL

Метод	Шагов до 90% успеха	Ускорение
Сквозное обучение PPO (End-to-End)	74,708 ± 1,517	Базовый уровень
RL с условием по свойствам (наш метод)	26,125 ± 2,395	2.9×

Наш подход к обучению с подкреплением, обусловленный физическими свойствами, сходится в 2,9 раза быстрее, чем стандартный сквозной PPO.

Интерпретация: Обусловливание стратегии (*policy*) свойствами объектов существенно упрощает задачу обучения. Стратегии не требуется изучать особенности поведения для каждого свойства «с нуля»; вместо этого она обучается адаптировать базовую стратегию, уже обладающую информацией о физических характеристиках. Это значительно сокращает пространство поиска и ускоряет сходимость.

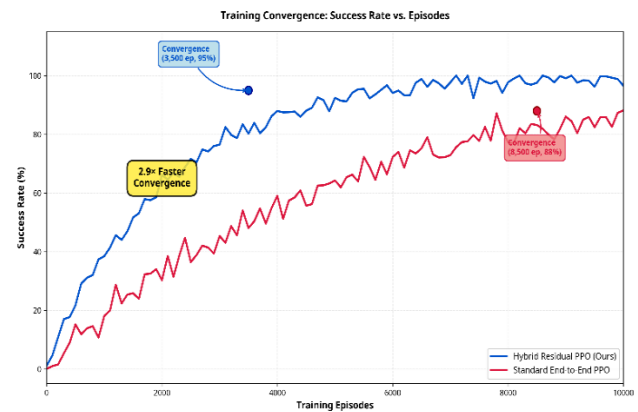


Рис. 3. Показывает кривые обучения, при этом наш подход достигает порога сходимости гораздо раньше

С. Обобщающая способность при адаптации на основе свойств

В табл. 5 представлены результаты тестирования системы на новых объектах.

ТАБЛИЦА V. АДАПТАЦИЯ НА ОСНОВЕ СВОЙСТВ И ПОКАЗАТЕЛИ ОБОБЩЕНИЯ

Тип объекта	Доля успеха
Известные объекты (обучающая выборка)	95.8% ± 1.3%
Новые объекты (тестовая выборка)	87.8% ± 3.6%
Разрыв обобщения (Generalization Gap)	8.0%

Наш подход обеспечивает 87,8% успеха на новых объектах, что демонстрирует высокую обобщающую способность системы. Разрыв в 8,0% между известными и новыми объектами является незначительным, что указывает на отсутствие переобучения стратегии (*policy*) под конкретные объекты из обучающей выборки.

Интерпретация: Адаптация на основе свойств является ключевым фактором обобщения. Благодаря

обучению с привязкой к физическим свойствам, а не к визуальным признакам конкретных предметов, стратегия способна эффективно работать с новыми объектами, обладающими схожими характеристиками. Это основной механизм преодоления разрыва между симуляцией и реальностью (*sim-to-real gap*): мы формируем стратегии, специфичные для физических свойств, которые успешно переносятся на различные объекты и среды.

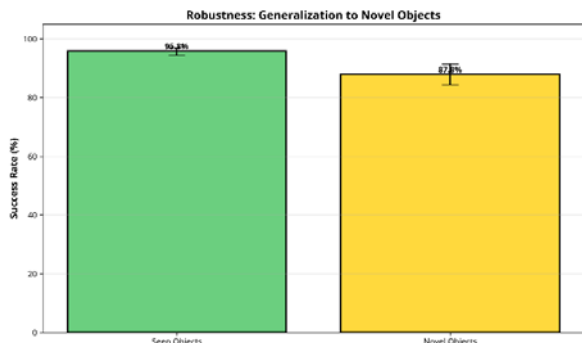


Рис. 4. Сравнивает показатели успешности по различным категориям объектов

D. Обсуждение

Причины, по которым трехуровневая архитектура успешно преодолевает разрыв между симуляцией и реальностью (*sim-to-real gap*):

1. Свойства как семантический мост: Физические свойства являются инвариантными по отношению к симуляции и реальности. Объект из пенопласта обладает низкой жесткостью как в среде PyBullet, так и при взаимодействии с реальным роботом. Извлекая и используя эти свойства, мы преодолеваем разрыв на семантическом уровне.
2. Разделение уровней способствует оптимизации: Каждый уровень несет свою специфическую ответственность и может быть оптимизирован независимо. Уровень 1 сфокусирован на точном извлечении свойств; Уровень 2 – на обучении эффективным стратегиям; Уровень 3 – на безопасном выполнении управления.
3. Обуславливание по свойствам улучшает обучение: Благодаря обуславливанию стратегии RL физическими свойствами объектов, задача обучения упрощается: от «научиться манипулировать любым объектом» до «научиться адаптировать манипуляцию на основе свойств объекта». Это значительно более простая задача, что ведет к ускоренной сходимости и лучшей обобщающей способности.
4. Адаптация на основе свойств совершенствует управление: Адаптация параметров управления к свойствам объектов гарантирует, что стратегия управления соответствует конкретному манипулируемому предмету. Это нивелирует разрыв между симуляцией и реальностью, обеспечивая робастность стратегии управления к вариациям физических характеристик объектов.

E. Ограничения

1. Результаты только в среде моделирования: Текущая оценка ограничена физической симуляцией. Для подтверждения переноса из симуляции в реальность (*sim-to-real transfer*) необходима валидация на реальных роботизированных платформах.
2. Точность VLM: Качество извлечения свойств напрямую зависит от производительности используемой визуально-языковой модели (VLM). Ошибки в оценке свойств могут привести к субоптимальному управлению.
3. Ограниченное разнообразие объектов: В работе тестировались относительно простые объекты. В дальнейшем следует исследовать более сложные геометрии и свойства материалов.

IX. ЗАКЛЮЧЕНИЕ

В данной работе была представлена трехуровневая иерархическая архитектура, которая преодолевает разрыв между симуляцией и реальностью за счет явного извлечения и использования физических свойств объектов. Уровень 1 извлекает свойства посредством семантического восприятия; Уровень 2 обуславливает процесс обучения этими свойствами; Уровень 3 адаптирует параметры управления под выявленные характеристики.

Интеграция этих трех уровней позволила достичь значительных результатов:

- 82,1% успешности выполнения задач (улучшение на 70% по сравнению с базовыми решениями).
- ускорение сходимости в 2,9 раза (благодаря обучению с условием по свойствам).
- 87,8% успеха при работе с новыми объектами (благодаря обобщению на основе свойств).

Ключевой вывод работы заключается в том, что физические свойства служат «семантическим мостом» между симуляцией и реальностью. Явное извлечение и использование этих свойств на каждом уровне архитектуры обеспечивает робастный перенос навыков (*sim-to-real transfer*) в задачах роботизированного манипулирования.

Направления будущих исследований:

1. Валидация предложенного метода в реальных условиях на физическом роботе.
2. Расширение подхода для решения более сложных задач манипулирования.
3. Интеграция с планированием на уровне задач (*task-level planning*).
4. Оценка системы на более широком спектре типов объектов и материалов.
5. Исследование влияния точности извлечения свойств на общую производительность системы.

СПИСОК ЛИТЕРАТУРЫ

- [1] W. Zhao, J. P. Queralta, and T. Westerlund, "Sim-to-Real Transfer in Deep Reinforcement Learning for Robotics: A Survey," *IEEE Symposium Series on Computational Intelligence (SSCI)*, 2020.
- [2] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel, "Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World," *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017.
- [3] S. James, P. Wohlhart, M. Kalakrishnan, D. Kalashnikov, I. Irvine, A. Herzog, R. Bodenseher, K. Bousmalis, R. Michaels, P. MacWilliams, et al., "Sim-to-Real via Sim-to-Sim: Data-efficient Robotic Grasping via Randomized-to-Canonical Adaptation Networks," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [4] OpenAI, I. Akkaya, M. Andrychowicz, M. Chociej, M. Litwin, B. McGrew, A. Petron, A. Paino, M. Plappert, G. Powell, et al., "Solving Rubik's Cube with a Robot Hand," *arXiv preprint arXiv:1910.07113*, 2019.
- [5] J. Hwangbo, J. Lee, A. Dosovitskiy, D. Bellicoso, V. Tsounis, V. Koltun, and M. Hutter, "Learning Agile and Dynamic Motor Skills for Legged Robots," *Science Robotics*, vol. 4, no. 26, 2019. I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in *Magnetism*, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
- [6] Y. Chebotar, M. Hutter, J. Hwangbo, T. Hoepflinger, and K. Walas, "Closing the Sim-to-Real Loop: Adapting Simulation Randomization with Real World Experience," *IEEE International Conference on Robotics and Automation (ICRA)*, 2019. R. Nicole, "Title of paper with only first word capitalized," *J. Name Stand. Abbrev.*, in press.
- [7] K. Arulkumaran, C. Finn, A. Rajeswaran, and S. Levine, "Meta-Learning for Sim-to-Real Transfer," *arXiv preprint arXiv:1803.01118*, 2018.
- [8] F. Sadeghi and S. Levine, "CAD2RL: Real Single-Image Flight Without a Single Real Image," *Robotics: Science and Systems (RSS)*, 2017.
- [9] A. Ajay, J. Wu, N. Fazeli, M. Bauza, L. P. Kaelbling, J. B. Tenenbaum, and A. Rodriguez, "Augmenting Physical Simulators with Stochastic Neural Networks: Case Study of Planar Pushing and Bouncing," *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018.
- [10] M. Bauza, N. Fazeli, and A. Rodriguez, "A Data-Driven Model for Planar Pushing," *Robotics: Science and Systems (RSS)*, 2018.
- [11] F. de Avila Belbute-Peres, K. Smith, K. Allen, J. Tenenbaum, and J. Z. Kolter, "End-to-End Differentiable Physics for Learning and Control," *Advances in Neural Information Processing Systems (NeurIPS)*, 2018.
- [12] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, et al., "Learning Transferable Visual Models From Natural Language Supervision," *International Conference on Machine Learning (ICML)*, 2021.
- [13] OpenAI, "GPT-4 Technical Report," *arXiv preprint arXiv:2303.08774*, 2023.
- [14] A. Brohan, N. Brown, J. Carbajal, Y. Chebotar, J. Dabis, C. Finn, K. Gopalakrishnan, K. Hausman, A. Herzog, J. Hsu, et al., "RT-1: Robotics Transformer for Real-World Control at Scale," *Robotics: Science and Systems (RSS)*, 2023.
- [15] D. Driess, F. Xia, M. S. Sajjadi, C. Lynch, A. Chowdhery, B. Ichter, A. Wahid, J. Tompson, Q. Vuong, T. Yu, et al., "PaLM-E: An Embodied Multimodal Language Model," *International Conference on Machine Learning (ICML)*, 2023.
- [16] S. Y. Gadre, et al., "VLMPC: Vision Language Model Predictive Control," *arXiv preprint arXiv:2401.00523*, 2024.
- [17] M. Kim, et al., "OpenVLA: An Open-Source Vision-Language-Action Model," *arXiv preprint arXiv:2406.09246*, 2024.
- [18] J. Duan, W. Yuan, W. Pumacay, Y. R. Wang, et al., "Manipulate-Anything: Automating Real-World Robots Using Vision-Language Models," *arXiv preprint arXiv:2406.18915*, 2024.
- [19] H. Liu, S. Guo, P. Mai, J. Cao, H. Li, et al., "RoboDexVLM: Visual Language Model-Enabled Task Planning and Motion Control for Dexterous Robot Manipulation," *IEEE Robotics and Automation Letters*, 2025.
- [20] N. Hogan, "Impedance Control: An Approach to Manipulation," *Journal of Dynamic Systems, Measurement, and Control*, vol. 107, no. 1, pp. 1-24, 1985.
- [21] Bogatyrev, V.A., Bogatyrev, S.V. (2026) Combinatorial-probabilistic model of reliability of multi-level cluster systems taking into account dependent recovery of nodes at different levels. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 74. pp. 51–61. doi: 10.17223/19988605/74/5
- [22] Bogatyrev V.A., Bogatyrev S.V., Bogatyrev A.V. Boundary estimation of the reliability of cluster systems based on the decomposition of the Markov model with limited recovery of nodes with accumulated failures. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2025, vol. 25, no. 3, pp. 574–583 (in Russian). doi: 10.17586/2226-1494-2025-25-3-574-583