РЕЦЕНЗИЯ
на выпускную квалификационную работу

Кичик Маргариты Георгиевны,

на тему «Распознавание физических свойств объектов трехмерных сцен по их RGB-изображениям»

Актуальность исследования

Для робототехники, AR/VR и объектов-ориентированного поиска знание материала предмета критично — оно определяет выбор захвата, силы контакта, термическую и электробезопаспость. При этом существующие датасеты ограничены десятками классов материалов и не отражают богатство реального мира. Автор решает двойную проблему — нехватку качественных данных и отсутствие универсального бенчмарка для многометочной классификации материалов, предлагая масштабное расширение LVIS и новый метод обучения. Таким образом работа отвечает современным запросам vision-робототехники и мультимодального ИИ.

Основные результаты исследования включают: создание датасета MatLVIS с более чем 1.2 млн естественных описаний объектов и около 1600 уникальных категорий материалов, что значительно превосходит существующие датасеты; разработку модифицированной мультимодальной архитектуры на основе CLIP, объединяющей контрастивные функции потерь InfoNCE и MultiSupCon; достижение точности классификации 70% с полным совпаднием меток в 62% случаев, превосходящее все рассмотренные альтернативы.

Положительные стороны работы:

- Масштаб и новизна датасета, существенно расширяющего словарь материалов.
- Чёткая методология сравнения: одинаковый препроцесс, метрики micro/macro, SubsetAcc.
- Открытая перспектива использования MatLVIS для zero-shot-transfer и few-shot-learning.
- Лаконичное, хорошо иллюстрированное изложение (рис. 6-10) и приведённый код экспериментов.

Теоретическая и практическая значимость работы заключается в развитии методологии мультимодального контрастивного обучения для распознавания материалов и создании ресурса для дальнейших исследований в области компьютерного зрения.

Достоверность результатов подтверждается корректной методологией экспериментов, сравнением с множественными базовыми методами и детальным анализом различных архитектурных решений.

Недостатки исследования:

1. **Низкое качество по редким классам.** Macro-F1 остаётся ≤ 0,32 даже у лучшей модели, что указывает на сильный дисбаланс и требует техник re-weighting или focal-loss.
2. **Отсутствие анализа вычислительных ресурсов.** Для контрастивного дообучения использовался батч = 1024 и 10 000 итераций, но не указаны GPU-конфигурация, суммарное время и энергозатраты; это затрудняет оценку применимости на реальных роботах.
3. **Качество автолейблинга.** Аннотации получены без масштабной ручной проверки; не оценена точность (precision) самого датасета, возможны лексические шумы (*fabric ↔ textile*).
4. **Эксперименты лишь на статичных изображениях.** Нет проверки модели в роботизированных pipeline'ах (grasp-selection, WBC), что ограничивает выводы о фактической пользе для embodied-AI.
5. **Сегментационные бенчмарки ограничены DMS.** Не рассмотрены более сложные outdoor-датасеты (GTOS, LAS&T), а mIoU-цифры для Mask2Former не приведены в тексте обзора.

Несмотря на отмеченные недостатки, работа представляет собой качественное исследование с новыми теоретическими и практическими результатами, имеющими значение для развития области компьютерного зрения и робототехники.

Учитывая описанные недостатки, работа заслуживает оценки "отлично" (9 баллов).
Рекомендую присвоение студенту квалификации магистра.

Рецензент:

Алексей Леонидович Постников, PhD,
Исполнительный Директор, Центр робототехники ПАО Сбербанк

_____ / Постников А.Л.
(подпись рецензента, заверенная по месту работы)

_____
(дата)