

ПОВЫШЕНИЕ ТОЧНОСТИ РАСПОЗНАВАНИЯ БОЛЕЗНЕЙ РАСТЕНИЙ КОМПАКТНОЙ VLM С ПОМОЩЬЮ БАЗЫ ЗНАНИЙ

Шиян А. С.¹, Ларин А.¹

Научный руководитель – док. техн. наук, доцент Жукова Н. А.¹

¹Университет ИТМО

an.sshiyang@gmail.com

Работа выполнена в рамках проекта №623106 «Автономные интеллектуальные системы».

Введение

В условиях ограниченных вычислительных ресурсов использование компактных мультимодальных моделей (например, Qwen2.5-VL-3B) сопровождается снижением точности при решении узкоспециализированных задач, таких как диагностика заболеваний растений.

Предобученные мультимодальные модели глубокого обучения, такие как CLIP, показали высокую эффективность в задачах сопоставления изображений и текстовых описаний за счёт обучения в общем семантическом пространстве [1]. В области диагностики заболеваний растений исследуются подходы, основанные на zero-shot и fine-tuning стратегиях для мультимодальных моделей [2].

Таким образом, научная проблема заключается в разработке метода повышения точности малой визуально-языковой модели при диагностике заболеваний растений за счёт интеграции внешних знаний и мультимодальных признаков без увеличения числа параметров модели.

Основная часть

В работе рассматривается повышение точности визуально-языковой модели на примере Qwen2.5-VL-3B на наборе данных PlantDoc [3] за счёт интеграции структурированной базы знаний заболеваний растений и их симптомов и предобученной модели CLIP.

Предлагаемый подход включает следующие этапы:

1. Извлечение описания растения и наблюдаемых симптомов с использованием визуально-языковой модели Qwen2.5-VL-3B;
2. Вычисление сходства между симптомами и описаниями растений из базы знаний и описанием изображения с использованием предобученной модели CLIP;
3. Извлечение из базы знаний классов растения и болезни по выбранным CLIP описаниям.

Структурированная база знаний формируется в виде связей между культурами, заболеваниями и их характерными симптомами. Это позволяет формализовать экспертные знания и использовать их как внешний источник информации.

Для оценки вклада интеграции базы знаний и модуля CLIP было выполнено сравнительное исследование базовой и расширенной архитектур:

1. Использование только визуально-языковой модели Qwen2.5-VL-3B. На вход подавалось изображение растения, на выходе определялись культура и заболевание. Точность составила 14.28%.

2. Использование визуально-языковой модели совместно с базой знаний и CLIP. Модель Qwen2.5-VL-3B генерировала текстовое описание растения и симптомов. Из описания удалялись артефакты генерации (повторяющиеся фразы, нерелевантные символы). Далее CLIP сопоставлял полученное описание с текстовыми представлениями симптомов из базы знаний. Итоговое решение выбиралось по максимальному сходству. Точность составила 46.4%.

Полученные результаты демонстрируют, что интеграция структурированной базы знаний и сопоставления описаний предобученной моделью CLIP существенно снижает влияние галлюцинаций и повышает устойчивость предсказаний. Предлагаемый подход является экономичным, так как не требует дообучения крупной модели и не увеличивает количество её параметров.

Выводы

Интеграция структурированной базы знаний болезней растений и предобученной мультимодальной модели CLIP позволяет существенно повысить точность небольшой визуально-языковой модели на примере Qwen2.5-VL-3B при решении задач диагностики заболеваний растений без значительного увеличения числа параметров и вычислительных затрат.

Литература

1. Radford A. et al. Learning transferable visual models from natural language supervision //International conference on machine learning. – PmLR, 2021. – С. 8748-8763.
2. Campos-Mocholí M., Chacón-Albero O., Julian V. CLIP in the Field: A Study of Plant Disease Detection via Zero-Shot and Fine-Tuning //International Conference on Practical Applications of Agents and Multi-Agent Systems. – Cham : Springer Nature Switzerland, 2025. – С. 17-28.
3. Singh D. et al. PlantDoc: A dataset for visual plant disease detection //Proceedings of the 7th ACM IKDD CoDS and 25th COMAD. – 2020. – С. 249-253.