НЕЙРОСЕТЕВЫЕ МЕТОДЫ МАТТИНГА ИЗОБРАЖЕНИЙ НА ОСНОВЕ ИНТЕРАКТИВНОЙ СЕГМЕНТАЦИИ

Эмиров Артур Станиславович

Студент

Факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия

E-mail: s02220582@gse.cs.msu.ru

Научный руководитель — Конушин Антон Сергеевич

Маттинг – фундаментальная задача компьютерного зрения, суть которой заключается в выделении переднего плана на изображении путем вычисления его альфа-канала. Маттинг актуален для различных задач, таких как редактирование изображений и видео, замена фона в онлайн-конференциях, виртуальная реальность, и часто используется для выделения людей и животных.

Одной из проблем в задаче маттинга является сложность разметки эталонных изображений, из-за чего существует крайне мало качественных наборов данных, которые можно было бы использовать для обучения алгоритмов. В то же время, существует задача сегментации (классификация каждого пикселя изображения на принадлежность целевому объекту) является глубоко проработанной, и для нее существует множество хорошо исследованных алгоритмов и качественных наборов данных различной природы.

Отдельного внимания заслуживают алгоритмы интерактивной сегментации. Их отличие от обычной сегментации в том, что они требуют участия пользователя на протяжении всего процесса. В частности, пользователь может выделять нужные объекты и исправлять ошибки алгоритма с помощью кликов. Интерактивная сегментация может использоваться в различных приложениях и для редактирования изображений. Например, в Paint 3D на Windows 10 есть функция выделения областей изображения кликом. Такой подход позволяет значительно ускорить разметку человеком изображений для сегментации или упростить ретушь фотографий. В работе предложен подход, объединяющий методы интерактивной сегментации и адаптивного маттинга для повышения гибкости и эффективности алгоритма маттинга.

Основой метода стала архитектура RITM, изначально разработанная для интерактивной сегментации. Для её адаптации к задаче маттинга применён параметрически эффективный метод LoRA, позволяющий дообучить модель с минимальными изменениями исход-

ных параметров. Интерактивность реализована через интеграцию пользовательских кликов, которые кодируются как пространственные маски и подаются на вход сети вместе с изображением. Это позволяет итеративно уточнять альфа-канал, корректируя ошибки предсказания в реальном времени.

Эксперименты проведены на комбинированном датасете. Результаты показали, что предложенный метод сохраняет скорость базового алгоритма, при этом демонстрируя высокое качество маттинга. Проведено сравнение с подходами, не использующими интерактивность. Интеграция LoRA позволила сократить вычислительные затраты на адаптацию модели, что делает метод применимым для устройств с ограниченными ресурсами.

Преимущества подхода: интерактивность (пользовательские клики позволяют уточнять маску маттинга) и эффективность (параметрическая адаптация снижает требования к вычислительным ресурсам).

Перспективы работы связаны с расширением функционала взаимодействия (штрихи, прямоугольники) и интеграцией трансформерных архитектур для повышения точности.

Литература

1. Edward J Hu and Yelong Shen and Phillip Wallis and Zeyuan Allen-Zhu and Yuanzhi Li and Shean Wang and Lu Wang and Weizhu Chen. LoRA: Low-Rank Adaptation of Large Language Models. ICLR, 2022
2. Jiachen Li and Jitesh Jain and Humphrey Shi. Matting Anything. Arxiv
3. K. Sofiiuk, I. A. Petrov and A. Konushin, "Reviving Iterative Training with Mask Guidance for Interactive Segmentation," 2022 IEEE International Conference on Image Processing (ICIP), Bordeaux, France, 2022, pp. 3141-3145, doi: 10.1109/ICIP46576.2022.9897365