

РАЗРАБОТКА НЕЙРОСЕТЕВОГО МЕТОДА ИНТЕРАКТИВНОГО МАТИРОВАНИЯ ИЗОБРАЖЕНИЙ

Москаленко Андрей Викторович

Студент

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

E-mail: andrey.moskalenko@graphics.cs.msu.ru

Научный руководитель — Ватолин Дмитрий Сергеевич

Матирование – нахождение карты прозрачности объекта переднего плана, является важной задачей компьютерного зрения. Цель состоит в том, чтобы определить значения прозрачности, а также истинные значения цвета пикселей переднего плана и фона. Результатом матирования является карта прозрачности, позволяющая отделить от фона объект, который затем можно наложить на произвольный фон с помощью уравнения композиции.

Существующие методы автоматического матирования, как правило, требуют качественно размеченной тернарной тримап-маски, где закодированы передний и задний план, а также неизвестная область, которую нужно восстановить методу. В данной работе предлагается метод, который не требует на вход качественно размеченной тримап-маски, а также позволяет интерактивно менять свои предсказания в зависимости от пожеланий пользователя при помощи простых кликов мышью.

В качестве базовой архитектуры был использован DeepLabv3 [1] с следующими изменениями – размер ядра первой свертки был увеличен на 3 канала для кодирования положительных и отрицательных взаимодействий, а также для маски прошлого раунда взаимодействия (либо нулевой тензор на первом раунде), а последний слой вместо классификации возвращает один канал – альфа карту. Для обучения использовалась конкатенация из нескольких доступных датасетов с эталонными картами прозрачности. Для обучения использовалось 80% от каждого датасета, для валидации оставшиеся 20%. В качестве функции потерь использовался L1Loss.

Для генерации пользовательских вводов была разработана другая стратегия, подразумевающая генерацию вводов на лету. Из-за сложности генерации штрихов, было решено использовать только клики фиксированного радиуса. В качестве маски прошлого раунда нейросеть в этой стратегии получает свои же настоящие предсказания. Общая схема предложенного метода приведена на Рис. 1.

Для сравнений был использован общепринятый датасет компо-

зиций [3]. Он состоит из 493 объектов в обучающей выборке и 50 в тестовой. Для каждого объекта определены синтетически накладываемые фоны. Для сравнения использовалась метрика SAD. Полученный метод получил значение ошибки в 75 единиц, при этом метод [3] имеет ошибку 50, а лучший метод [2] ошибку в 25 единиц. Отметим, что все указанные методы, кроме предложенного требуют на вход качественно размеченную тримаш-маску для работы и не имеют возможности интерактивно менять свои предсказания в зависимости от пожеланий пользователя. Предлагаемый подход хоть и уступает в качестве последним методам, основанным на архитектуре трансформера, не обладает упомянутыми недостатками и является более эффективным.

Иллюстрации

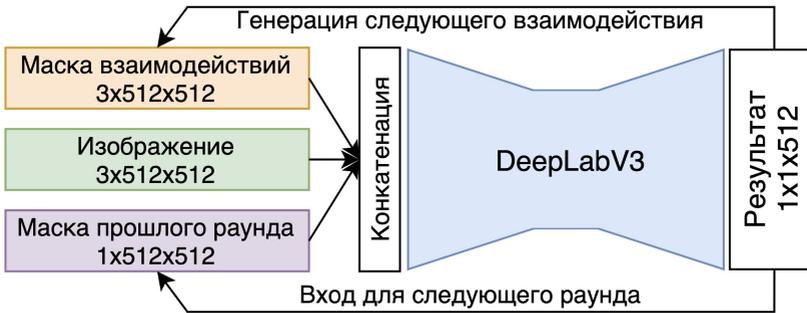


Рис 1. Общая схема предложенного метода.

Литература

1. Chen L., Papandreou G., Schroff F., Adam H. Rethinking atrous convolution for semantic image segmentation. // In arXiv preprint, 2017, P. 1–14.
2. Park G., Son S., Yoo J., Kim S., Kwak N. Matteformer: Transformer-based image matting via prior-tokens. // In Proceedings of the IEEE conference on computer vision and pattern recognition, 2022, P. 11696–11706.
3. Xu N., Price B., Cohen S., Huang T. Deep image matting // In Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, P. 2970–2979.