

## РАЗРАБОТКА НЕЙРОСЕТЕВОГО МЕТОДА ДЛЯ СТАБИЛЬНОГО ВО ВРЕМЕНИ МАТИРОВАНИЯ ВИДЕОПОСЛЕДОВАТЕЛЬНОСТЕЙ С ЛЮДЬМИ

*Молодецких Иван Андреевич*

*Студент*

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

*E-mail: ivan.molodetskikh@graphics.cs.msu.ru*

*Научный руководитель — Ерофеев Михаил Викторович*

Матирование, или нахождение карты прозрачности объекта переднего плана, является одной из основных операций обработки изображений и видеопоследовательностей. Результатом матирования является «вырезанное» изображение объекта, которое можно вставить на другой, произвольный фон при помощи композиции. В общем случае задачи матирования неизвестно, какие из объектов являются объектами переднего плана, в связи с чем необходимы дополнительные входные данные.

Целью настоящей работы является создание стабильного во времени нейросетевого алгоритма для матирования видеопоследовательностей с людьми. Стабильность во времени означает, что на результирующих картах прозрачности отсутствует резкое изменение значений прозрачности между кадрами. Явное определение человеческих фигур как объектов переднего плана позволяет обойтись без дополнительных входных данных.

Для нахождения и сегментации человеческих фигур на входных кадрах видео используется нейросетевая архитектура DeepLab v3 [1]. На выходе данной сети — бинарная маска человеческой фигуры. Следующим шагом является нахождение значений прозрачности на полупрозрачных областях (границы, волосы, полупрозрачная одежда) и обеспечение стабильности во времени.

Здесь рассматриваются два подхода. Первый подход заключается в использовании нейросетевой архитектуры матирования изображений и последующей стабилизации полученных карт прозрачности во времени некоторым алгоритмом. Второй подход — построение нейросетевой архитектуры, выполняющей оба этих шага одновременно.

Обучение нейросети матирования изображений для первого подхода производится на нескольких наборах данных с картами прозрачности фотографий людей. Для сглаживания во времени предлагается подход с использованием карт расстояний.

Нейросетевая архитектура, рассматриваемая для второго подхода, обрабатывает одновременно несколько последовательных кадров видео. Это позволяет явно отобразить в функции потерь требование к стабильности результирующих карт прозрачности во времени. Для передачи информации между кадрами видео в архитектуре используются свёрточные LSTM-модули.

Главные сложности в реализации второго подхода — большой объём входных данных и отсутствие готового набора данных для обучения. Создание видеопоследовательностей для обучения происходит из наборов фоновых фотографий и видео и набора фотографий людей с картами прозрачности. К фотографиям людей и исходным картам прозрачности применяется случайная деформация, симулирующая движение. Деформация генерируется на нескольких масштабах для получения как движения человека в целом, так и локальных деформаций граничных областей. Последовательность кадров получается путём многократного применения деформации и последующей композиции на кадры заднего плана.

Тестирование алгоритмов производится как при помощи объективных метрик [2], так субъективно. На рис. 1 приведён пример карты прозрачности, полученной одним из рассматриваемых подходов.

### Иллюстрации



Рис. 1. Пример полученной карты прозрачности.

### Литература

1. Chen L. C. Encoder-decoder with atrous separable convolution for semantic image segmentation // In Proceedings of the European conference on computer vision (ECCV), 2018, С. 801–818.
2. Erofeev M. Perceptually Motivated Benchmark for Video Matting // In Proceedings of the British Machine Vision Conference (BMVC), 2015, С. 99.1–99.12