

ЗАМЕНА ФОНА НА ИЗОБРАЖЕНИИ ВЫСОКОГО РАЗРЕШЕНИЯ В РЕАЛЬНОМ ВРЕМЕНИ

Исмаил А-К.Х.

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Перцев Д.Ю. – канд. техн. наук, доцент

Представлен алгоритм матирования изображения высокого разрешения на основе рекуррентной нейронной сети с использованием алгоритмов кодирования-декодирования изображения. Также представлено приложение на платформе Android, реализующее разработанный алгоритм. Приведены результаты анализа производительности.

Матирование изображения I является задачей из области компьютерного зрения и основано на разделении заднего плана B (от англ. background) и переднего плана F (от англ. foreground):

$$I = \alpha F + (1 - \alpha)B, \quad (1)$$

где F – изображение переднего плана;

B – изображение заднего плана;

α – карта прозрачности переднего плана, элементы в которой принимают значение в диапазоне $[0, 1]$ (значение 0 – пиксель с координатой (x, y) принадлежит заднему плану, значение 1 – переднему плану).

Алгоритм, представленный в данной работе, имеет существенное преимущество перед классическими алгоритмами матирования [1,2]: в то время, как более старые алгоритмы требуют наличия заранее подготовленного фона [3], тримапа [4] или набросков [5], данный алгоритм лишён этого недостатка. Это достигается за счёт рекуррентной архитектуры нейронной сети, и позволяет использовать его в более широком классе задач, так как отсутствует необходимость прямого пользовательского вмешательства. Таким образом, данный алгоритм наилучшим образом подходит для использования в мобильных приложениях.

На рисунке 1 представлена архитектура нейронной сети, осуществляющая матирование. Она состоит из кодера, рекуррентного декодера, выходного блока, а также блока повышения разрешения. Перед попаданием в первый блок, изображение высокого разрешения уменьшается. Данный шаг позволяет ускорить обработку кадра, при этом из-за применения фильтра повышения разрешения на выходе сети, существенной потери качества на выходе не происходит. Блок кодирования извлекает полезные признаки на изображении, уменьшает полученный кадр в 2 раза и передаёт на следующий кодер. Данный этап повторяется 4 раза. Таким образом, кодеры извлекают признаки в масштабе 12, 14, 18 и 116. Затем рекуррентный декодер поэтапно обновляет состояния сети и декодирует признаки, понижая количество каналов исходного изображения до одного – альфа. Выходной блок выполняет финальное извлечение признаков и генерирует коэффициенты альфа, а также формирует передний план, который затем, вместе с исходным изображением в высоком разрешении, передаются фильтру для повышения разрешения результирующих изображений.

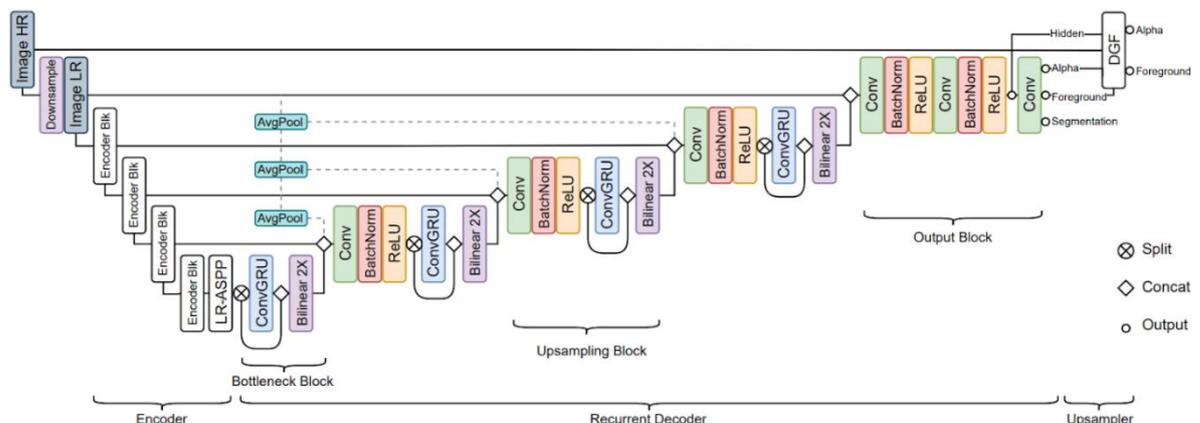


Рисунок 1 – Архитектура нейронной сети

Одной из ключевых проблем представленной реализации алгоритма является то, что на данный момент не существует мобильного устройства способного выполнить такой объём

вычислений за малый промежуток времени. Кроме того, существуют ограничения, затрудняющие реализацию подобных комплексных алгоритмов напрямую на устройствах на платформе Android.

Таким образом, было принято решение, что наиболее оптимальным и высокопроизводительным методом реализации такого приложения является клиент-серверная архитектура, в которой мобильное приложение является клиентом, отправляющим изображения на удалённый сервер, который, в свою очередь, обрабатывает принятый кадр и возвращает результат. Далее, задачей клиента остаётся только вывести полученное от сервера изображение. При использовании такой архитектуры приложения, сложные вычисления выполняются на полноценных, более мощных, графических процессорах. В дополнение к этому, такой подход даёт возможность создания кроссплатформенного приложения без необходимости изменения серверной части, что сильно ускоряет процесс его разработки.

Данный подход также имеет и свои недостатки. Для работы приложению необходимо постоянное подключение к Интернету. Более того, скорость работы приложения напрямую зависит от скорости обмена информацией между клиентом и сервером.

В таблице 1 представлены параметры тестовых систем. В таблицах 2 и 3 представлены временные метрики для клиента и сервера, соответственно.

Таблица 1 – Параметры тестовых систем

	Смартфон	Laptop-клиент	Сервер
CPU	HiSilicon Kirin 970	Intel Core i7-8550U	4x Intel Xeon Platinum 8180M
GPU	Mali-G72 MP12	Nvidia MX250	Nvidia Tesla K80
NIC	WI-FI	Intel Dual Band Wireless-AC 8265	?

Таблица 2 – Временные метрики клиента

	Разрешение камеры	Кодирование изображения, мс			Отправка запроса на сервер (и ожидание ответа), мс			Декодирование ответа от сервера, мс			Средняя частота кадров, к/с
		Мин.	Ср.	Макс.	Мин.	Ср.	Макс.	Мин.	Ср.	Макс.	
Смартфон	1280x720	38	40	42	470	500	1500	63	75	92	1,08
	1920x1080	77	80	86	650	900	1700	75	105	124	0,7
	2816x1584	140	166	168	843	1100	2100	165	180	193	0,55
Laptop-клиент	1920x1080	13	17	27	490	550	820	13	30	49	1,15

Таблица 3 – Временные метрики сервера

Разрешение изображения	Декодирование изображения от клиента, мс	Предсказание нейросети, мс	Постобработка изображения, мс
1280x720	40	5	166
1920x1080	40	6	170
2816x1584	80	6,5	250

Исходя из результатов тестирования можно судить о том, что «бутылочным горлышком» данной реализации является код, ответственный за кодирование-декодирование изображений.

Список использованных источников:

1. Levin A. A Closed Form Solution to Natural Image Matting / A. Levin, [et al.] // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2008. – Volume 30. – Issue 2. – P. 228-242. – DOI: 10.1109/TPAMI.2007.1177.
2. Xu N. Deep Image Matting [Electronic Resource] / arXiv. – Mode of access: <https://arxiv.org/abs/1703.03872>. – Date of access: 06.04.2022.

3. *Real-Time High-Resolution Background Matting [Electronic Resource] / University of Washington. – Mode of access: <https://grail.cs.washington.edu/projects/background-matting-v2>. – Date of access: 06.04.2022.*

4. *Chuang Y. A Bayesian Approach to Digital Matting / Y.-Y. Chuang, [et al.] // Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001. – 2001. – Kauai, HI, USA. – DOI: 10.1109/CVPR.2001.990970.*

5. *Zheng Y. Learning Based Digital Matting / Y. Zheng, C. Kambhamettu // 2009 IEEE 12th International Conference on Computer Vision. – 2009. – Kyoto, Japan. – DOI: 10.1109/ICCV.2009.5459326*