

# Метод автоматического расцветчивания рисованной мультипликации

В. Ф. Коновалов  
Самарский национальный  
исследовательский университет им.  
академика С.П. Королева  
Самара, Россия  
vitfvk@gmail.com

**Аннотация.** В статье рассматривается задача расцветчивания черно-белой рисованной мультипликации с использованием нейронных сетей. Исследуется эффективность модификаций существующего алгоритма-прототипа при различных комбинациях функций потерь, рассматриваются различные модификации алгоритма-прототипа. Предлагается новая функция потерь для нейронной сети — с использованием сегментирования изображения по цветам. Исследуется эффективность модифицированного алгоритма с предложенной функцией потерь.

**Ключевые слова** — сегментация по цвету, машинное обучение, расцветчивание изображений, нейронные сети, оптический поток, watershed алгоритм.

## 1. ВВЕДЕНИЕ

Задача расцветчивания заключается в получении полноцветного изображения по изображению в оттенках серого или наброску. Алгоритмы компьютерного зрения применяются для расцветчивания фильмов уже достаточно давно [1-6]. В то же время, не так много работ обращают внимание на задачу расцветчивания рисованной мультипликации.

В ходе работы была выдвинута гипотеза о том, что четкие границы объектов и монотонность цветов в рисованной мультипликации могут быть использованы для улучшения качества обучения нейросети путем введения новой функции потерь.

В разделе 2 производится обзор существующих архитектур и способов расцветчивания изображений и видео. В разделе 3 описывается выбранный для проведения исследования алгоритм-прототип. В разделе 4 рассматриваются предлагаемые дополнения к алгоритму-прототипу, направленные на улучшение качества расцветчивания. Раздел 5 посвящен исследованию алгоритма при различных комбинациях функций потерь.

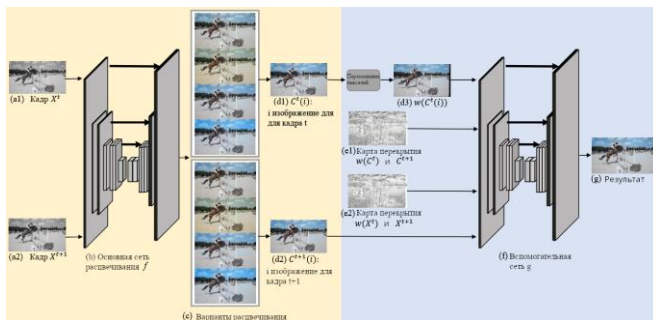


Рис. 1. Архитектура исследуемого алгоритма[6]

## 2. ОБЗОР СУЩЕСТВУЮЩИХ МЕТОДОВ

Существующие алгоритмы расцветчивания делятся на две большие группы — расцветчивание с участием пользователя и полностью автоматические. Участие пользователя может включать текстовый ввод [1], грубые мазки правильных цветов [2], палитру цветов изображения [4], или вспомогательное изображение [3]. Однако большинство этих методов плохо адаптируются к видео. Даже полностью автоматические алгоритмы расцветчивания изображений, такие как алгоритм [5] примененный к данной задаче, рассматривают каждый кадр видео отдельно, что приводит к меняющимся от кадра к кадру цветам соответствующих пикселей, нестабильности.

Задача расцветчивания является задачей один ко многим — некоторые объекты могут быть корректно расцветчены несколькими способами. Таким образом, хороший алгоритм расцветчивания должен предлагать на выбор пользователя несколько результатов. При расцветчивании последовательности кадров - видео - алгоритм расцветчивания должен учитывать взаимосвязь между последовательными кадрами. Алгоритм удовлетворяющий данным требованиям описан в [6], именно он будет использован как алгоритм-прототип в данной работе.

## 3. ОПИСАНИЕ ИССЛЕДУЕМОГО АЛГОРИТМА

Рассматриваемый алгоритм-прототип состоит из двух сетей — основной и вспомогательной (refinement - улучшающей). Обе сети представляют собой модифицированные сети U-Net[7]. Общая схема алгоритма прототипа приведена на рисунке 1.

Для обучения используются несколько функций потерь:

- *RankDiv* представляет собой перцептуальную функцию потерь. Значением функции потерь является средняя разница между откликами соответствующих слоев сети VGG-19;

- функция потерь *Bilateral* для каждого пикселя  $p$  исходного цветного изображения ищет несколько ближайших по координате и цвету соседей  $q$  и минимизирует расстояние между ними для предсказанного изображения;

- функция потерь *Temporal* использует полученный по оптической потоку из двух последовательных кадров маску перекрытия и искажения кадров для получения разницы между неперекрытыми пикселями двух последовательных предсказанных кадров.

#### 4. ПРЕДЛАГАЕМЫЕ ДОПОЛНЕНИЯ

Первой предлагаемой модификацией алгоритма-прототипа является применение новых функций потерь, обозначенных ниже. Второй модификацией выступает замена оптического потока, склонного к ошибкам при расчете по рисованным или полутоновым изображениям, на совместный анализ последовательных кадров.

Предлагаемая функция потерь исходит из того, что в рисованной мультипликации цвета меняются с гораздо меньшей частотой, чем в естественных изображениях. Из-за этого функция потерь Bilateral может приводить к нежелательным эффектам, в частности, крупные сегменты одного цвета могут быть раскрашены разными цветами.

Предлагаемая функция потерь основывается на сегментации входных изображений алгоритмом watershed. Далее для  $n$  сегментов исходного изображения  $S$  рассчитывается средний цвет, а так же находятся ближайшие по среднему цвету сегменты. Полученные данные используются для вычисления функции потерь:

$$\sum_{t=1}^n \sum_q \|M[f_t(x)] - M[f_q(x)]\|_1 + \sum_{t=1}^n M[f_t(x)] - M[C_t] \quad (1)$$

где:  $X$  – исходное изображение;  
 $f$  – выход нейросети расцвечивания;  
 $C$  — истинное цветное изображение;  
 $t$  – сегменты изображения;  
 $q$  – ближайшие по цвету к  $t$  сегменты.

Замена оптического потока выполняется путем подачи на вход основной сети двух последовательных кадров в оттенках серого. При этом вспомогательная сеть  $g$  исключается из обучения, поскольку необходимость ее использования определяется наличием оптического потока.

#### 5. ПРОВЕДЕННЫЕ ИССЛЕДОВАНИЯ

В рамках работы было проведено исследование алгоритма, описанного в разделе 2 при разных комбинациях функций потерь:

В качестве набора данных был выбран мультфильм Tom&Jerry.

Тренировка сети проводилась согласно указаниям из работы [6], количество эпох было уменьшено до 500. Для эксперимента с предложенной функцией потерь сеть обучалась в течении 500 эпох, первые 200 — с уменьшенным размером изображений 116x160, далее — 240x320. Параметр DivNum был выбран равным 4.

С использованием обученных сетей было проведено расцвечивание кадров тестовой выборки. Для расцвеченных и истинных кадров было проведено сравнение с использованием метрик PSNR, LPIPS и MSE в пространстве Lab. В таблицу внесены средние значения по полученным различным вариантам расцвечивания.

Таблица I. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

	PSNR	LPIPS	MSE Lab
Оттенки серого	20,54358	0,254359	0,003924
Bilateral + RankDiv + Temporal, основная	20,915756	0,207897	0,003493

Bilateral + RankDiv + Temporal, основная + вспомогательная	19,404139	0,249332	0,005149
RankDiv + Segment, основная	20,893824	0,215922	0,003830
RankDiv + Bilateral + Segment, основная	21,040308	0,207812	0,003641
RankDiv + Bilateral, основная	20,819466	0,211376	0,003793

По результатам исследования можно сделать вывод, что сеть с предложенной функцией потерь показывает наилучший результат. Использование вспомогательной сети сильно ухудшает результат.

В дальнейшем планируется введение дополнительных модификаций предложенного решения: использовать сегментированные кадры, либо уже расцвеченные ранее кадры в качестве дополнительных входных данных. Также представляет интерес исследовать для расцвечивания архитектуру нейросети, подобной описанной в работе [5].

#### ЛИТЕРАТУРА

- [1] Manjunatha, V. Learning to Color from Language / V. Manjunatha, M. Iyyer, J. Boyd-Graber, L. Davis // Proceedings of the 2018 Conference of the North American. — 2018. — Vol. 2. — P. 764–769.
- [2] Zhang, R. Real-Time User-Guided Image Colorization with Learned Deep Priors / R. Zhang, J.-Y. Zhu, P. Isola, X. Geng, A. S. Lin, T. Yu, A. A. Efros // ACM Transactions on Graphics. — 2017. — Vol. 36(4). — P. 1–11.
- [3] Xu, Z. Stylization-Based Architecture for Fast Deep Exemplar Colorization / Z. Xu, T. Wang, F. Fang, Y. Sheng, G. Zhang // IEEE/CVF Conference on Computer Vision. — 2020. — P. 9360–9369. DOI: 10.1109/CVPR42600.2020.00938.
- [4] Bahng, H. Coloring with Words: Guiding Image Colorization Through Text-based Palette Generation / H. Bahng, S. Yoo, W. Cho, D. K. Park, Z. Wu, X. Ma, J. Choo // In book: Computer Vision – ECCV — 2018. — P. 443–459.
- [5] Isola, P. Image-to-Image Translation with Conditional Adversarial Networks / P. Isola, J.-Y. Zhu, T. Zhou, A. A. Efros // IEEE Conference on Computer Vision and Pattern Recognition. — 2017. — P. 5967–5976. DOI: 10.1109/CVPR.2017.632.
- [6] Lei, C. Fully Automatic Video Colorization with Self-Regularization and Diversity / C. Lei, Q. Chen // 2019 IEEE/CVF Conference. — 2019. DOI: 10.48550/arXiv.1908.01311.
- [7] Ronneberger, O. U-Net: Convolutional Networks for Biomedical Image Segmentation / O. Ronneberger, P. Fischer, T. Brox // Lecture Notes in Computer Science. — 2015. — Vol. 9351. DOI: 10.48550/arXiv.1505.04597.
- [8] Sun, D. PWC-Net: CNNs for Optical Flow Using Pyramid, Warping, and Cost Volume / D. Sun, X. Yang, M.-Y. Liu, J. Kautz // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. — 2018. — P. 8934–8943. DOI: 10.48550/arXiv.1709.02371.