

РАЗРАБОТКА НЕЙРОСЕТЕВЫХ МЕТОДОВ РАСПРОСТРАНЕНИЯ ЦВЕТА В ВИДЕО

Лаврушкин С.В., аспирант Московского государственного университета имени М.В. Ломоносова, e-mail: sergey.lavrushkin@graphics.cs.msu.ru

Ватолин Д.С., к.ф.-м.н., старший научный сотрудник Московского государственного университета имени М.В. Ломоносова, e-mail: dmitriy@graphics.cs.msu.ru

DEVELOPMENT OF NEURAL NETWORKS FOR COLOR PROPAGATION IN VIDEO

Lavrushkin S.V., Vatolin D.S.

The task of creating color versions of existing black and white images and videos originated immediately after color images and videos had emerged. The process of color prediction and propagation for grayscale images and videos is called colorization. While colorization has been gradually automated with the development of digital image and video processing, there is a lot of manual work. In this paper, we consider a problem of color propagation in video using a single colored frame, which can be obtained either by manual colorization or by applying automatic image colorization methods. To solve this problem we propose two types of neural network methods: the former are based on the prediction of separable filters, the latter use optical flow to propagate color between frames. The majority of proposed methods have shown high visual quality on test datasets with preservation of colorization quality for a large number of frames forward in the videos.

Key words: deep learning, convolutional neural networks, colorization.

Ключевые слова: глубокое обучение, сверточные нейронные сети, колоризация.

Введение

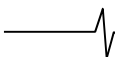
Появление первых черно-белых изображений и фильмов было важным шагом человечества к документированию визуальной информации. Однако, как только воплотился в жизнь захват цветных изображений, возникла задача создания цветных версий уже существующих черно-белых материалов. Процесс добавления и распространения цвета в изображениях и видео в оттенках серого называется колоризацией. Прежде всего, колоризация изображений применяется для улучшения визуального качества черно-белых изображений и видео, цветные версии которых не существуют или находятся в плохом состоянии, но также колоризация применима в медицинских изображениях для разметки областей интереса, матировании изображений, дизайне интерьера, средствах моделирования, например, грима и макияжа и в других задачах.

Процесс колоризации фотографий, осуществляемый вручную, возник ненамного позже появления самой фотографии. Существуют примеры ручной колоризации изображений, датируемые 1842 годом, хотя возможны и более ранние примеры. Ручной процесс колоризации фильмов также появился достаточно давно. Так, в начале 1900-х годов это практиковала французская компания Pathe, а также ручной подход колоризации использовался в случае диафильмов в 1930-х годах. Полностью ручным был, например, процесс колоризации кадров космической программы Аполлон в 1960-х, что потребовало колоссального объема работы. Компьютерный процесс для колоризации впервые был предложен в 1970 году для добавления цвета в черно-

Задача создания цветных версий уже существующих черно-белых изображений и видео возникла сразу же после того, как научились снимать и показывать цветные изображения и видео. Процесс добавления и распространения цвета в изображениях и видео в оттенках серого называется колоризацией. И хотя наряду с развитием методов цифровой обработки изображений и видео колоризация также постепенно автоматизируется, для ее осуществления до сих пор используется много ручного труда. В данной работе рассматривается задача распространения цвета в видео по одному цветному кадру, который может быть получен путем ручной колоризации либо путем применения методов автоматической колоризации изображений. Для решения этой задачи рассмотрено два типа нейросетевых методов: первые основаны на предсказании сепарабельных фильтров, вторые используют оптический поток для распространения цвета между кадрами. Большинство из рассмотренных методов показали высокое качество работы на тестовых выборках с сохранением качества при колоризации на большое количество кадров вперед.

белые фильмы, тогда же и был сформулирован термин колоризация [1].

Наряду с развитием методов цифровой обработки изображений и видео колоризация изображений и видео также постепенно автоматизируется для упрощения ручного труда. Колоризация привлекает большое внимание со стороны исследователей в области компьютерного зрения, и новые технологии, направленные на сокращение необходимого числа человеческих усилий для колоризации изображений и видео, все еще продолжают разрабатываться. В настоящее время колоризация по-прежнему считается достаточно сложной проблемой, требующей произвольных высокоуровневых знаний относительно содержания изображений и видео для успешного решения задачи. Поэтому для колоризации до сих пор используется много ручного труда, а сам процесс осуществ-



ляется посредством использования плагинов к ПО по обработке изображений и видео общего назначения, таких как Adobe Photoshop, Nuke и Adobe After Effects.

В данной работе рассматривается задача распространения цвета с одного цветного кадра видеопоследовательности на все остальные черно-белые кадры. В работе предложены новые нейросетевые методы распространения цвета в видео. В дальнейшем входные данные о цвете можно дополнительно сократить за счет использования методов по распространению цвета внутри кадра и сначала распространять введенные пользователем «мазки», либо не использовать входные данные о цвете вообще и раскрашивать исходный кадр с помощью методов автоматической колоризации изображений.

Обзор методов колоризации

В целом постановка задачи колоризации изображений или видео является некорректной, так как эта задача не обладает единственным решением, отвечающим условию высокого визуального качества (например, без уменьшения визуального качества в разные цвета может быть окрашена одежда). Для получения удовлетворительных результатов колоризации было предложено две категории методов: распространение цвета, управляемое пользователем, и автоматическая колоризация.

Методы распространения цвета, управляемого пользователем [2-5], требуют ввода цветных «мазков» поверх изображения в оттенках серого и распространяют полученные цвета по всему изображению, решая глобальную задачу оптимизации. Эти методы могут достичь впечатляющих результатов колоризации, но часто требуют интенсивного взаимодействия с пользователем (иногда число взаимодействий достигает 50 для изображений небольшого размера), так как каждая область изображения с различными цветами должна быть явно отмечена пользователем. Пример такого изображения с входными данными пользователя представлен на рис. 1. Поскольку данные методы полностью полагаются на ввод от пользователей, возникает необходимость указывать цвета даже для областей, в которых цветовая неопределенность не возникает, например, зеленая растительность или синее небо. И хотя методы распространения цвета, управляемого пользователем, могут быть легко адаптированы для колоризации видео путем определения соседних пикселей в оптимизационной функции не только в пространственных координатах, но и во времени за счет использования алгоритмов оптического потока, данная модификация потребует от пользователя колоссальный объем входных данных, так как для каждого кадра придется вручную определить большое число цветных областей. В противном случае, при малом объеме входных данных, результат колоризации будет деградировать со временем, как показано на рис. 2.

Чтобы избавиться от проблемы использования пользовательских данных при решении данной задачи, ранние методы автоматической колоризации раскрашивали изображения путем обучения цветовых признаков по одному или нескольким шаблонным цветным изобра-

жениям с похожим семантическим содержанием [6-8]. К сожалению, иногда может потребоваться очень много времени для поиска подходящего шаблонного цветного изображения, либо такого изображения может попросту не быть, поэтому данные методы не применялись на практике. Однако за последние несколько лет большую популярность и повсеместное использование в области компьютерного зрения обрели сверточные нейронные сети, существенно улучшив качество работы методов в таких задачах, как классификация изображений, обнаружение объектов, семантическая сегментация и других. Поскольку сверточные нейронные сети обладают способностью обучаться высокоуровневым признакам в изображениях, они также были использованы и для автоматической колоризации [9-11], значительно улучшив показатели данных методов. Нейросетевые методы автоматической колоризации способны генерировать правдоподобные цветные изображения в большинстве случаев (пример удачной и неудачной колоризации с помощью метода [11] представлен на рис. 3).



Рис. 1. Слева представлен пример изображения в оттенках серого с введенными пользователем цветами. Справа – результат колоризации с помощью метода [2]. Как можно видеть, для получения качественной колоризации одного изображения небольшого размера с помощью оптимизационного метода необходимо большое количество взаимодействий с пользователем.

Источник: <https://www.cs.huji.ac.il/~yweiss/Colorization>



Рис. 2. Пример работы алгоритма [2] для колоризации видео при предоставлении цвета для первого кадра. Результат работы отчетливо показывает, что использование такого алгоритма без дополнительных данных для следующих кадров приводит к потере качества колоризации для этих кадров

Однако эти методы обладают рядом существенных недостатков. Ввиду того, что различные объекты могут обладать несколькими подходящими цветами, результаты работы данных методов могут содержать цвета или стили, которые пользователи не ожидали либо хотели бы изменить. Например, пользователям может понадобиться раскрасить гору в зеленый цвет под весенний период времени, но алгоритм представит желтую ко-

лоризацию горы под осенний период времени. Также методы [9-11] решают только задачу колоризации изображений, и для их применения для видео необходимо существенно изменить нейросетевую архитектуру, а также процесс обучения. При использовании сверточной нейронной сети, обученной для колоризации изображений, на видео возможно возникновение сильной нестабильности результатов во времени.



Рис. 3. Пример удачной (слева)

и неудачной (справа) колоризации с помощью метода [11]. В большинстве случаев неудачная колоризация связана с отсутствием хороших аналогов в обучающей выборке

Но недавно начали появляться нейросетевые методы для распространения данных в видео, в том числе цветowych [12, 13]. Данные методы позволяют распространить цвет с одного предоставленного цветного кадра, который может быть получен путем колоризации изображения с помощью описанных выше алгоритмов либо ручной раскраски.

Предложенные нейросетевые методы Описание входных данных

Видеопоследовательность рассматривается в цветовом пространстве YUV. В качестве входных данных подаются изображения в оттенках серого, соответствующие каналу Y, а также цветочные каналы (U и V) для одного кадра сцены для дальнейшего распространения цвета (данных каналов U и V) с этого кадра на остальные. Распространение цвета на следующие кадры осуществляется по-разному для разных предложенных нейросетевых методов – по-разному учитывается движение в кадрах: с помощью предсказания сепарабельных фильтров непосредственно в сети либо с использованием оптического потока. Далее приведены описания разработанных нейросетевых архитектур.

Нейросетевые методы, предсказывающие сепарабельные фильтры для распространения цвета

В данных методах в качестве базовой нейросетевой архитектуры была выбрана сеть, предсказывающая сепарабельные фильтры для распространения цвета с одного кадра на другой. Аналогичный подход впервые был использован в задаче интерполяции кадров видео, в котором для получения промежуточных кадров сетью предсказываются сепарабельные фильтры как для первого, так и для второго кадра, осуществляется свертка

соответствующих кадров с предсказанными свертками и результаты свертки складываются [14]. Данный подход был адаптирован на задачу распространения цвета с одного кадра на другой: сеть предсказывает сепарабельный фильтр для первого кадра, а для получения цветочных каналов для второго кадра осуществляется свертка цветочных каналов первого кадра с предсказанным фильтром. Общая схема метода представлена на рис. 4.

В качестве базовой сети была выбрана сеть типа кодировщик-декодировщик, состоящая из набора сверточных блоков. Детали выбранной архитектуры представлены в табл. 1.

Для улучшения качества работы базовой сети в нее был добавлен учет границ за счет использования дифференцируемых фильтров. Для этого были использованы фильтры domain transform [15] и fast global smoother [16]. Дифференцируемость фильтра domain transform была продемонстрирована в работе [17], в которой он был использован для постобработки результатов в нейронной сети для семантической сегментации изображений. Аналогичным свойством обладает и фильтр fast global smoother. Для доказательства рассмотрим этот метод. Данный фильтр решает следующую оптимизационную задачу для 1D сигнала (то есть для каждой строки и каждого столбца изображения):

$$\arg \min_u \sum_x \left((u_x - f_x)^2 + \lambda \sum_{i \in N(x)} (u_x - u_i) \right)$$

где u – отфильтрованный сигнал, f – фильтруемый сигнал, $N(x) = \{x-1, x+1\}$, $\omega_{x,x-1} = e^{-\sqrt{2} \frac{g_x}{\sigma_c}}$, $\omega_{x,x+1} = \omega_{x+1,x}$, g_x – граница изображения в пикселе x , $\lambda > 0, \sigma_c > 0$ – параметры, контролирующие степень фильтрации.

Решение данной задачи оптимизации осуществляется в два прохода:

1. Прямой проход:

$$\tilde{c}_x = \frac{c_x}{b_x - \tilde{c}_{x-1} a_x}, \tilde{c}_0 = \frac{c_0}{b_0}, \tilde{f}_x = \frac{f_x - \tilde{f}_{x-1} a_x}{b_x - \tilde{c}_{x-1} a_x}, \tilde{f}_0 = \frac{f_0}{b_0},$$

$$a_x = -\lambda \omega_{x,x+1}, b_x = 1 + \lambda (\omega_{x,x-1} + \omega_{x+1,x}), c_x = -\lambda \omega_{x+1,x},$$

$$a_0 = 0, c_{w-1} = 0,$$

где w – размерность 1D сигнала.

2. Обратный проход:

$$u_x = f_x - \tilde{c}_x u_{x+1}, u_{w-1} = f_{w-1}.$$

Покажем дифференцируемость обратного прохода. При обратном распространении ошибки с последующих

слоев нейронной сети приходит $\frac{\partial L}{\partial u_x}$. Тогда для обратного прохода получаем следующие производные:

$$\frac{\partial L}{\partial \tilde{f}_x} \leftarrow \frac{\partial L}{\partial u_x}, \frac{\partial L}{\partial \tilde{c}_x} \leftarrow \frac{\partial L}{\partial u_x} \cdot u_{x+1}, \frac{\partial L}{\partial u_{x+1}} \leftarrow \frac{\partial L}{\partial u_{x+1}} - \frac{\partial L}{\partial u_x} \cdot \tilde{c}_x.$$

Аналогично для прямого прохода. С обратного прохода

приходят $\frac{\partial L}{\partial \tilde{f}_x}, \frac{\partial L}{\partial \tilde{c}_x}$. Тогда получаем следующие производные для прямого прохода:

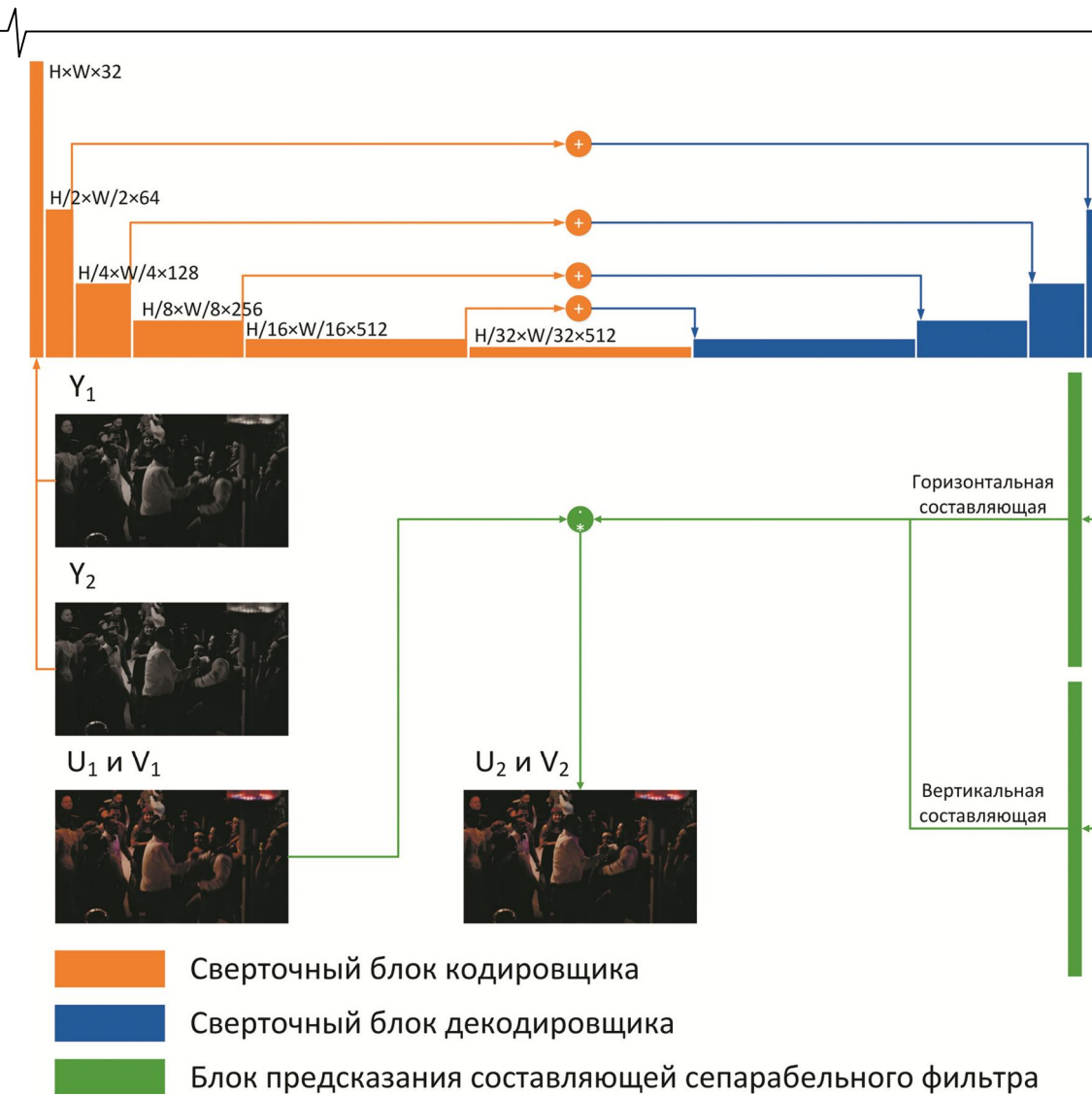


Рис. 4. Схема архитектуры сети, предсказывающей сепарабельные фильтры для распространения цвета с одного кадра на другой

Таблица 1. Конфигурация блоков сверточной нейронной сети для распространения цвета между кадрами видео посредством предсказания сепарабельных фильтров

Слой	Размерность	Шаг	Коэффициент дилатации	Функция активации
Сверточный блок кодировщика				
Сверточный	3 x 3	1	1	ReLU
Сверточный	3 x 3	1	1	ReLU
Сверточный	3 x 3	1	1	ReLU
Сверточный	3 x 3	1	1	ReLU
Субдискретизации с выбором среднего	2 x 2	2	1	—
Сверточный блок декодировщика				
Сверточный	3 x 3	1	1	ReLU
Сверточный	3 x 3	1	1	ReLU
Сверточный	3 x 3	1	1	ReLU
Билинейной интерполяции	—	—	—	—
Сверточный	3 x 3	1	1	ReLU
Блок предсказания составляющей сепарабельного фильтра				
Сверточный	3 x 3	1	1	ReLU
Сверточный	3 x 3	1	1	ReLU
Сверточный	3 x 3	1	1	ReLU
Билинейной интерполяции	—	—	—	—
Сверточный	3 x 3	1	1	ReLU
Softmax	—	—	—	—

$$\frac{\partial L}{\partial f_x} \leftarrow \frac{\partial L}{\partial \tilde{f}_x} \cdot \frac{1}{b_x - \tilde{c}_{x-1} a_x},$$

$$\frac{\partial L}{\partial a_x} \leftarrow \frac{\partial L}{\partial \tilde{f}_x} \cdot \frac{f_x \tilde{c}_{x-1} - \tilde{f}_{x-1} b_x}{(b_x - \tilde{c}_{x-1} a_x)^2} + \frac{\partial L}{\partial \tilde{c}_x} \cdot \frac{c_x \tilde{c}_{x-1}}{(b_x - \tilde{c}_{x-1} a_x)^2},$$

$$\frac{\partial L}{\partial c_x} \leftarrow \frac{\partial L}{\partial \tilde{c}_x} \cdot \frac{1}{b_x - \tilde{c}_{x-1} a_x},$$

$$\frac{\partial L}{\partial b_x} \leftarrow \frac{\partial L}{\partial \tilde{f}_x} \cdot \frac{\tilde{f}_x a_x - f_x}{(b_x - \tilde{c}_{x-1} a_x)^2} + \frac{\partial L}{\partial \tilde{c}_x} \cdot \frac{c_x}{(b_x - \tilde{c}_{x-1} a_x)^2},$$

$$\frac{\partial L}{\partial \omega_{x,x-1}} \leftarrow \left(\frac{\partial L}{\partial b_x} - \frac{\partial L}{\partial a_x} \right) \cdot \lambda,$$

$$\frac{\partial L}{\partial \omega_{x+1,x}} \leftarrow \frac{\partial L}{\partial \omega_{x+1,x}} + \left(\frac{\partial L}{\partial b_x} - \frac{\partial L}{\partial c_x} \right) \cdot \lambda,$$

$$\frac{\partial L}{\partial g_x} \leftarrow \frac{\partial L}{\partial \omega_{x,x-1}} \cdot \omega_{x,x-1} \cdot \frac{\sqrt{2}}{\sigma_c},$$

$$\frac{\partial L}{\partial \tilde{f}_{x-1}} \leftarrow \frac{\partial L}{\partial \tilde{f}_{x-1}} - \frac{\partial L}{\partial \tilde{f}_x} \cdot \frac{a_x}{b_x - \tilde{c}_{x-1} a_x},$$

$$\frac{\partial L}{\partial \tilde{c}_{x-1}} \leftarrow \frac{\partial L}{\partial \tilde{c}_{x-1}} + \frac{\partial L}{\partial \tilde{f}_x} \cdot \frac{a_x (f_x - \tilde{f}_x a_x)}{(b_x - \tilde{c}_{x-1} a_x)^2} + \frac{\partial L}{\partial \tilde{c}_x} \cdot \frac{a_x c_x}{(b_x - \tilde{c}_{x-1} a_x)^2}.$$

Таким образом, данный фильтр дифференцируем и может быть использован в качестве постобработки для учета границ непосредственно в нейронной сети.

При применении описанных выше фильтров базовый метод модифицируется за счет добавления сверточной сети для предсказания весов фильтров и добавления слоя, реализующего фильтрацию с помощью одного из фильтров. Для предсказания весов фильтров была использована сверточная нейронная сеть для предсказания границ, аналогичная по архитектуре сети [17]. Общая схема модифицированного метода представлена на рис. 5.

Также базовый метод был модифицирован для учета во время обучения нескольких подряд идущих кадров



Рис. 5. Схема модифицированного метода, предсказывающего сепарабельные фильтры для распространения цвета с фильтрацией результата с помощью дифференцируемых фильтров

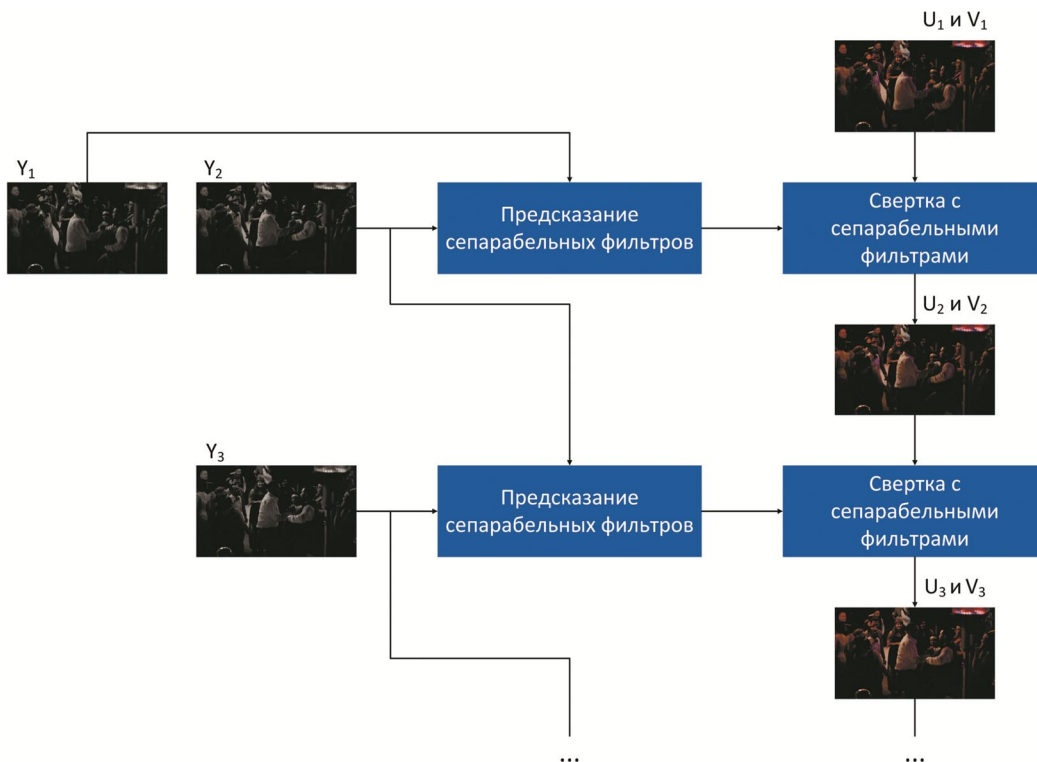


Рис. 6. Схема модифицированного метода, предсказывающего сепарабельные фильтры для распространения цвета с учетом нескольких подряд идущих кадров

в обучающей выборке для повышения итоговой временной стабильности метода и потенциального увеличения качества работы колоризации. Для этого сверточная нейронная сеть, используемая в базовом методе, во время обучения применялась к нескольким подряд идущим кадрам, распространяя цвет с предыдущего кадра на текущий вместо распространения с одного цветного кадра, предоставленного на вход: для каждого текущего кадра для распространения цвета с помощью предсказанных сепарабельных фильтров используется результат колоризации предыдущего кадра. При этом на каждом этапе колоризации очередного кадра используется одна и та же сверточная нейронная сеть. Общая схема модифицированного метода представлена на рис. 6.

Нейросетевые методы, использующие оптический поток

В данных методах для учета движения между кадрами используется оптический поток. Распространение цвета в видео происходит в два этапа.

1. Распространение цвета на соседние кадры осуществляется с помощью оптического потока. В качестве используемого оптического потока выбран метод PWC-Net [18]. Для получения цветов для еще нераскрашенного кадра цвет с входного цветного кадра компенсируется по найденным векторам движения к нераскрашенному кадру с помощью билинейной интерполяции. Но так как найденный оптический поток не везде точен, использование только одного оптического потока для колоризации видео приведет к существенным визуальным артефактам. Для поиска областей, содержащих ошибки оптического потока (например, в областях открытия/закрытия – областях, видимых в одном кадре и невидимых в другом), используется критерий, оценивающий консистентность прямого и обратного оптических потоков путем построения карт разности расстояний между исходными пикселями одного изображения и их отображениями, полученными с помощью перехода сначала по прямому оптическому потоку, а затем по обратному оптическому потоку.

2. Распространение цвета внутри кадра и исправление ошибок компенсации осуществляется с помощью сверточной нейронной сети, на вход которой подается изображение в оттенках серого, соответствующее каналу Y, а также цвет (каналы U и V), распространенный на предыдущем шаге с другого кадра, и карта консистентности оптического потока.

Для распространения цвета внутри кадра и исправления ошибок компенсации в первом варианте была использована нейросетевая архитектура GridNet [19], представляющая из себя модификацию архитектуры кодировщик-декодировщик и ранее применявшаяся для задачи семантической сегментации.

Данный метод распространения цвета был также модифицирован для учета нескольких подряд идущих кадров в видео для потенциального улучшения качества работы метода. В отличие от исходного метода модифицированная сеть принимает на вход не только яркость раскрашиваемого изображения (канал Y), цветовые каналы (U и V), полученные распространением цве-

та исходного цветного кадра с помощью оптического потока, и соответствующую карту консистентности оптического потока, но также цветовые каналы (U и V), полученные распространением цвета с предыдущего раскрашенного кадра с помощью оптического потока, и соответствующую этому оптическому потоку карту консистентности. Таким образом, во время обучения рассматриваются не отдельные пары кадров, а несколько подряд идущих кадров. При этом на каждом этапе колоризации очередного кадра используется одна и та же сверточная нейронная сеть. Общая схема модифицированного метода представлена на рис. 7.

Также для задачи распространения цвета были опробованы нейросетевые архитектуры, по-разному учитывающие карту консистентности оптического потока вместо обычного ее подмешивания к входным признакам, как это было сделано в предыдущем методе. Так, для распространения цвета внутри кадра и исправления ошибок компенсации была использована нейросетевая архитектура кодировщик-декодировщик, где вместо обычных сверточных слоев применялись стробированные свертки [20]. Конфигурация данной сети представлена в табл. 2. Аналогично методу, использующему архитектуру GridNet, метод на основе стробированных сверток был модифицирован для учета нескольких подряд идущих кадров в видео. Схема модифицированного метода совпадает со схемой метода модифицированного GridNet (рис. 7) и отличается только используемой для распространения цвета сетью.

Обучение разработанных нейросетевых методов

Для обучения предложенных методов была использована обучающая часть одной из версий набора данных Vimeo-90K [21], в которой каждый элемент состоит из семи кадров видеопоследовательности, что позволяет обучать модели, учитывающие несколько подряд идущих кадров. Всего набор данных включает в себя 91701 пример из семи кадров с разрешением 448×256.

В качестве базовой функции стоимости, оптимизируемой во время обучения всех моделей, было использовано среднеквадратическое отклонение предсказанных цветовых каналов C от истинных значений цветовых каналов \hat{C} :

$$L(\hat{C}, C)_{base} = \sum_{k=1}^D (\hat{C}_k - C_k)^2,$$

где D – размерность цветовых каналов. При этом в методах, использующих дифференцируемые фильтры для учета границ, данный функционал вычисляется как для результата работы базовой сети, так и для финального результата, полученного после фильтрации. Пусть \hat{C}' – промежуточный результат колоризации с помощью базовой сети, предсказывающей сепарабельные фильтры. Тогда для данных методов оптимизируется функционал:

$$L(\hat{C}, \hat{C}', C)_{filtered} = L(\hat{C}', C)_{base} + L(\hat{C}, C)_{base}.$$

Также используемая функция стоимости модифицируется для методов, учитывающих несколько подряд идущих кадров, – вычисляется среднеквадратическое отклонение предсказанных цветовых каналов $\hat{C}^{(i)}$ от

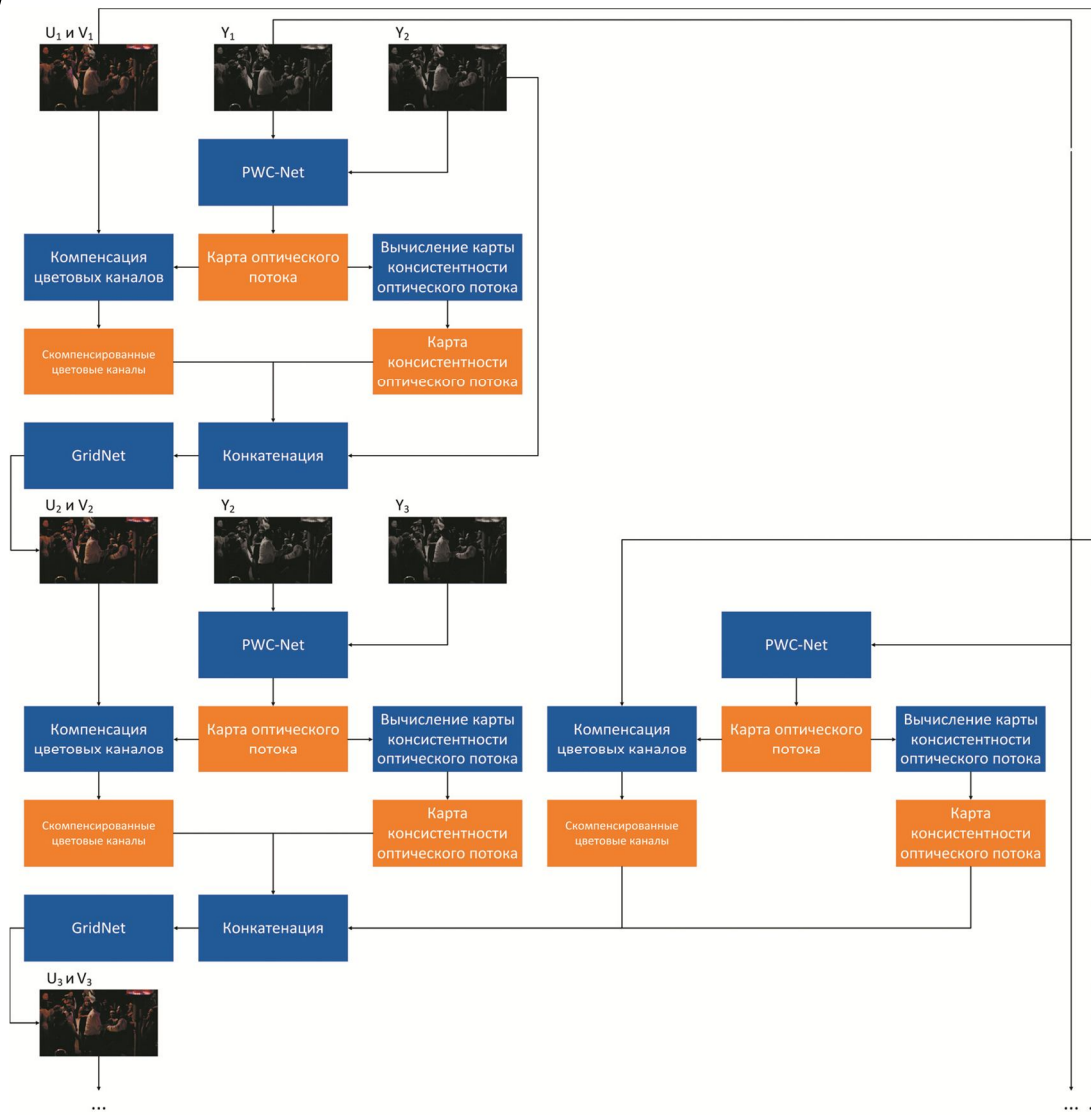
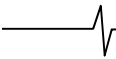


Рис. 7. Схема модифицированного метода для распространения цвета в видео на основе GridNet

Таблица 2. Конфигурация нейронной сети с использованием стробированных сверток для распространения цвета внутри кадра по результатам компенсации по оптическому потоку

Слой	Размерность	Количество вы- ходных каналов	Шаг	Коэффициент дилатации	Функция активации
Кодировщик					
Стробированная свертка	3 x 3	32	1	1	Leaky ReLU
Стробированная свертка	3 x 3	64	2	1	Leaky ReLU
Стробированная свертка	3 x 3	64	1	1	Leaky ReLU
Стробированная свертка	3 x 3	128	2	1	Leaky ReLU
Стробированная свертка	3 x 3	128	1	1	Leaky ReLU
Стробированная свертка	3 x 3	128	1	1	Leaky ReLU
Стробированная свертка	3 x 3	128	1	2	Leaky ReLU
Стробированная свертка	3 x 3	128	1	4	Leaky ReLU
Стробированная свертка	3 x 3	128	1	8	Leaky ReLU
Стробированная свертка	3 x 3	128	1	16	Leaky ReLU
Декодировщик					
Стробированная свертка	3 x 3	128	1	1	Leaky ReLU
Стробированная свертка	3 x 3	64	1	1	Leaky ReLU
Билинейной интерполяции	—	—	—	—	—
Стробированная свертка	3 x 3	64	1	—	—
Стробированная свертка	3 x 3	32	1	1	Leaky ReLU
Билинейной интерполяции	—	—	—	—	—
Стробированная свертка	3 x 3	16	1	1	Leaky ReLU
Стробированная свертка	3 x 3	2	1	1	Линейная



истинных значений цветовых каналов $C^{(i)}$ для каждого рассматриваемого кадра i :

$$L(\hat{C}, C)_{sed} = \sum_{i=1}^n \sum_{k=1}^D (\hat{C}_k^{(i)} - C_k^{(i)})^2,$$

где n – количество учитываемых кадров в каждом примере, которое было выбрано равным 4.

Перед обучением веса нейронных сетей инициализировались с помощью метода Xavier [22]. Для оптимизации функции стоимости был использован метод Adam [23]. В зависимости от количества параметров в сети и сложности модели использовались следующие параметры обучения: число примеров на каждой итерации обучения – 3-8, количество эпох обучения модели – 20-60, коэффициент скорости обучения – 10^{-4} , с уменьшением коэффициента в 10 раз каждую треть от общего количества шагов обучения.

Экспериментальная оценка

Тестирование разработанных методов проводилось на тестовой части набора данных Vimeo-90K [21], в которой каждый элемент состоит из семи кадров видеопоследовательности, а также на последовательностях из набора данных ACT [24], состоящего из 11234 видео, принадлежащих 43 классам действий, с разрешением 448×256 . На данных видеопоследовательностях измерялось качество распространения цвета на большое количество кадров, а не только на ближайшие кадры. Также для сравнения были реализованы и обучены аналоги по распространению данных видео – методы TPN [12] и STPN [13]. Во время тестирования вычислялись следующие объективные метрики качества: пико-

вое отношение сигнала к шуму – PSNR, индекс структурного сходства – SSIM.

Результаты тестирования на тестовой части набора данных Vimeo-90K представлены в табл. 3. Как видно из таблицы, лучший результат как по показателю PSNR, так и по показателю SSIM был получен с помощью метода на основе оптического потока и модели GridNet с учетом последовательности кадров. При этом добавление учета последовательности кадров в модель, предсказывающую сепарабельные фильтры, значительно ухудшает качество работы модели и поэтому в дальнейшем не используется.

Результаты тестирования на видеопоследовательностях из набора данных ACT представлены в табл. 4. В данном случае измерялся средний показатель PSNR для 10, 20, 30 и 40 кадров последовательностей для того, чтобы посмотреть качество колоризации в зависимости от расстояния между раскрашиваемыми кадрами. При тестировании использовались модели, обученные на обучающей части набора данных Vimeo-90K, и дообучения моделей под набор данных ACT не проводилось. Снова лучшие результаты показала модель на основе оптического потока и модели GridNet с учетом последовательности кадров.

Таким образом, наилучшим качеством обладает модель на основе оптического потока и модели GridNet с учетом последовательности кадров как по сравнению с недавними аналогами, так и с другими разработанными в данной работе методами. Примеры более качественной работы предложенного метода по сравнению с аналогами представлены на рис. 8.

Таблица 3. Результаты тестирования методов распространения цвета в видео на тестовой части набора данных Vimeo-90K

Метод	PSNR	SSIM
TPN [12]	39,2919	0,9760
STPN [13]	39,3702	0,9760
Предсказание сепарабельных фильтров	41,4124	0,9812
Предсказание сепарабельных фильтров + domain transform	41,9002	0,9824
Предсказание сепарабельных фильтров + fast global smoother	41,6647	0,9811
Предсказание сепарабельных фильтров + учет последовательности кадров	38,7839	0,9715
Оптический поток + GridNet	39,0103	0,9737
Оптический поток + GridNet + учет последовательности кадров	41,9163	0,9847
Оптический поток + стробированные свертки	38,9978	0,9749
Оптический поток + стробированные свертки + учет последовательности кадров	39,9714	0,9799

Таблица 4. Результаты тестирования методов распространения цвета в видео на наборе данных ACT

Метод	PSNR-10	PSNR-20	PSNR-30	PSNR-40
TPN [12]	35,2908	33,4738	32,4591	31,9222
STPN [13]	35,3388	33,4690	32,4487	31,9014
Предсказание сепарабельных фильтров	35,4241	33,0028	31,7690	31,0834
Предсказание сепарабельных фильтров + domain transform	35,8828	33,4696	32,2324	31,5522
Предсказание сепарабельных фильтров + fast global smoother	36,0216	33,4983	32,2237	31,5199
Оптический поток + GridNet	34,3889	32,1770	31,1329	30,6210
Оптический поток + GridNet + учет последовательности кадров	36,2783	34,1846	32,9208	32,0460
Оптический поток + стробированные свертки	34,3291	32,1950	31,1769	30,6642
Оптический поток + стробированные свертки + учет последовательности кадров	35,1742	32,7899	31,4909	30,6594

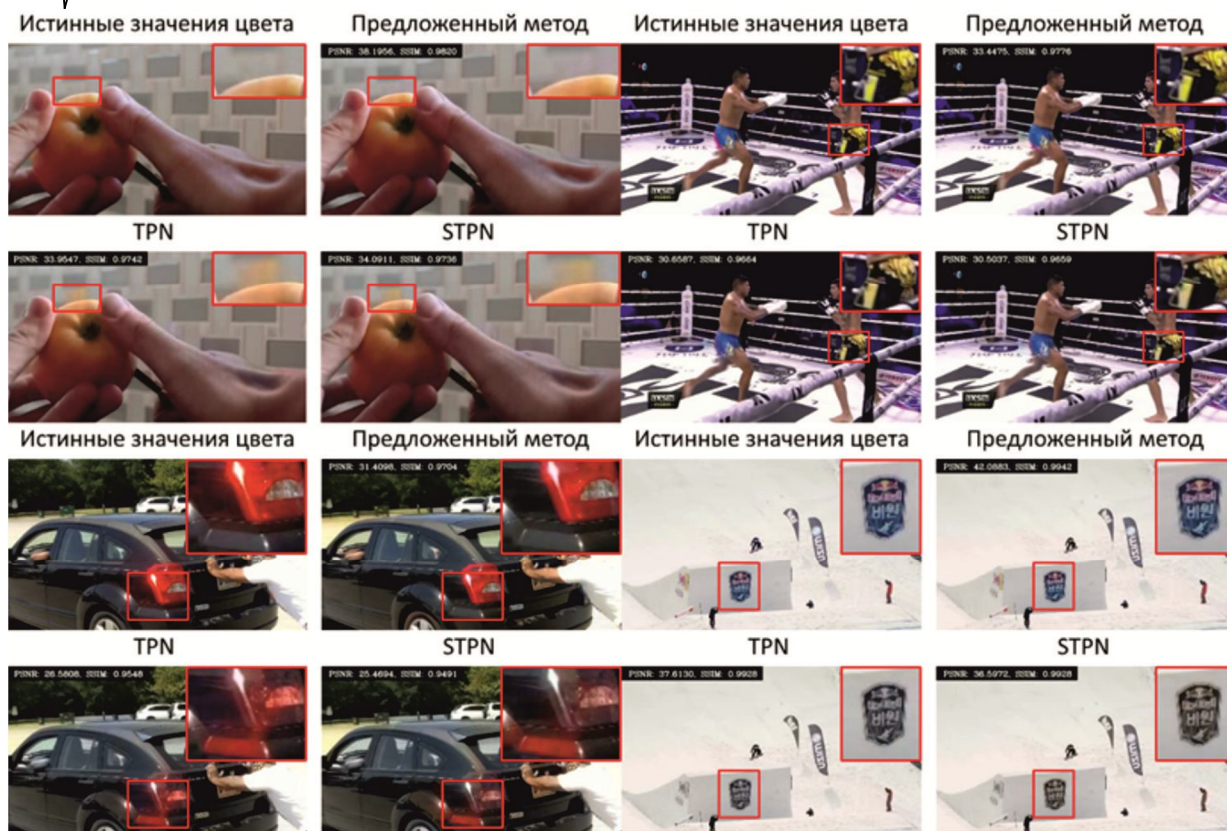


Рис. 8. Примеры высокого качества колоризации с помощью предложенного метода по сравнению с аналогами

Заключение

В данной работе предложен ряд методов колоризации видео на основе предсказания сепарабельных фильтров и использования оптического потока для учета движения в кадре. Первый тип методов был также модифицирован путем добавления фильтрации результатов на основе дифференцируемых фильтров, учитывающих границы, дифференцируемость одного из которых была показана в работе. Во втором типе методов были рассмотрены архитектуры GridNet и кодировщик-декодировщик на основе стробированных сверток. Оба типа методов были модифицированы для учета нескольких подряд идущих кадров. Для обучения предложенных методов была использована обучающая часть набора данных Vimeo-90K. Для тестирования была использована тестовая часть набора данных Vimeo-90K, а также сохранение качества результатов со временем было проверено на наборе данных АСТ. Наилучший результат как на тестовой части Vimeo-90K, так и на наборе данных АСТ, по сравнению со всеми предложенными методами и аналогами показал метод на основе оптического потока и модели GridNet с учетом последовательности кадров.

Исследования выполнены при финансовой поддержке РФФИ в рамках научного проекта 19-01-00785 а.

Обучение моделей производилось на вычислительном комплексе IBM Polus факультета вычислительной математики и кибернетики МГУ: <http://hpc.cmc.msu.ru/polus>.

Литература

1. Burns G. Colorization // Museum of Broadcast Communication: Encyclopedia of Television.

2. Levin A., Lischinski D., Weiss Y. Colorization using optimization // ACM Transactions on Graphics (ToG). – 2004, vol. 23, no. 3, pp. 689-694.

3. Horiuchi T., Kotera H. Colorization for monochrome image with texture // Color and Imaging Conference. 2005, no. 1, pp. 245-250.

4. Kawulok M., Smolka B. Competitive image colorization // IEEE International Conference on Image Processing (ICIP). 2010, pp. 405-408.

5. Kawulok M., Kawulok J., Smolka B. Textural features for scribble-based image colorization // Computer Recognition Systems 4. 2011, pp 269-278.

6. Chang Y. et al. Example-based color stylization of images // ACM Transactions on Applied Perception. 2006 vol. 2, no. 3, pp. 322-345.

7. Chia A. Y. S. et al. Semantic colorization with internet images // ACM Transactions on Graphics (TOG). – ACM, 2011, vol. 30, no. 6, pp. 156.

3. Gupta R. K. et al. Image colorization using similar images // Proceedings of the 20th ACM international conference on Multimedia. 2012, pp. 369-378.

9. Cheng Z., Yang Q., Sheng B. Deep colorization // Proceedings of the IEEE International Conference on Computer Vision. 2015, pp. 415-423.

10. Larsson G., Maire M., Shakhnarovich G. Learning representations for automatic colorization // European Conference on Computer Vision. 2016, pp. 577-593.

11. Zhang R., Isola P., Efros A. A. Colorful image colorization // European Conference on Computer Vision. 2016, pp. 649-666.

12. Liu S. et al. Learning affinity via spatial propagation networks // Advances in Neural Information Processing Sys-

tems. 2017, pp. 1520-1530.

13. Liu S. et al. Switchable temporal propagation network // Proceedings of the European Conference on Computer Vision (ECCV). 2018, pp. 87-102.

14. Niklaus S., Mai L., Liu F. Video frame interpolation via adaptive separable convolution // Proceedings of the IEEE International Conference on Computer Vision. 2017, pp. 261-270.

15. Gastal E. S. L., Oliveira M. M. Domain transform for edge-aware image and video processing // ACM Transactions on Graphics (ToG). ACM, 2011, vol. 30, no. 4, 69 p.

16. Min D. et al. Fast global image smoothing based on weighted least squares // IEEE Transactions on Image Processing. 2014, vol. 23, no. 12, pp. 5638-5653.

17. Liu Y. et al. Richer convolutional features for edge detection // Proceedings of the IEEE conference on computer vision and pattern recognition. 2017, pp. 3000-3009.

18. Sun D. et al. PWC-Net: CNNs for optical flow using pyramid, warping, and cost volume // Proceedings of the

IEEE Conference on Computer Vision and Pattern Recognition. 2018, pp. 8934-8943.

19. Fourure D. et al. Residual conv-deconv grid network for semantic segmentation // arXiv preprint arXiv:1707.07958. 2017.

20. Yu J. et al. Free-form image inpainting with gated convolution // Proceedings of the IEEE International Conference on Computer Vision. 2019, pp. 4471-4480.

21. Xue T. et al. Video enhancement with task-oriented flow // International Journal of Computer Vision. 2017, pp. 1-20.

22. Glorot X., Bengio Y. Understanding the difficulty of training deep feedforward neural networks // Proceedings of the thirteenth international conference on artificial intelligence and statistics. 2010, pp. 249-256.

23. Kingma D. P., Ba J. Adam: A method for stochastic optimization // arXiv preprint arXiv:1412.6980. 2014.

24. Wang X., Farhadi A., Gupta A. Actions~ transformations // Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. 2016, pp. 2658-2667

НОВЫЕ КНИГИ

Ремизов С.Л., Рясный Ю.В., Дежина Е.В., Черных Ю.С.

Основы цифровой обработки сигналов / Учебное пособие для вузов – М.: Изд-во «Горячая линия-Телеком», 2019 г. – 492 с.: ил.

Рекомендовано Научно-методическим советом федерального государственного бюджетного образовательного учреждения высшего образования «Сибирский государственный университет телекоммуникаций и информатики» (СибГУТИ) в качестве учебного пособия для студентов, обучающихся по направлениям 11.00.00 – «Электроника, радиотехника и системы связи», 10.00.00 – «Информационная безопасность».

Изложены основы теории дискретных сигналов и цепей, линейной дискретной обработки сигналов, приведены краткие сведения из области аналоговых цепей и сигналов, необходимые для более глубокого понимания изложенного материала. Рассмотрены цифровые избирательные фильтры и методы их проектирования, значительное внимание уделено КИХ-фильтрам с линейной фазой, поскольку они являются основой разработки избирательных фильтров с конеч-

ной импульсной характеристикой. Даны основы теории адаптивной обработки сигналов и рассмотрены вопросы её применения для моделирования и идентификации систем, для коррекции передаточных и частотных характеристик.

Для студентов, обучающихся по направлениям 11.00.00 – «Электроника, радиотехника и системы связи», 10.00.00 – «Информационная безопасность» при изучении курса «Основы цифровой обработки сигналов». Будет полезно читателям, самостоятельно изучающим курс цифровой обработки сигналов.



Витязев В.В.

Многоскоростная обработка сигналов – М.: Изд-во «Горячая линия-Телеком», 2017 г. – 336 с.: ил.

Рассмотрена эволюция теории и технологий многоскоростной обработки сигналов в период с начала 70-х гг. прошлого столетия до наших дней с позиции вклада, который внесли в их развитие работы российских ученых и специалистов в области цифровых информационных технологий реального времени. Описаны методы и алгоритмы многоступенчатой и многокаскадной реализации цифровых узкополосных фильтров и банков цифровых фильтров на основе эффектов прореживания по времени (децимация во временной области) и по частоте (децимация в частотной области).

Приведена методика оптимизации параметров многоступенчатых и многокаскадных структур цифровых полосовых фильтров. Построение оптимальных структур и расчет параметров фильтров частотной селекции иллюстрируется многочисленными примерами.

Для специалистов, научных работников, преподавателей вузов, аспирантов; будет полезна студентам информационных и инфокоммуникационных направлений подготовки.

