# Сравнение методов реконструкции промежуточных кадров видеопоследовательностей с динамической сценой

Ю.Х. Ганеева
*Самарский национальный исследовательский университет им. акад. С.П. Королева*
Самара, Россия
jganeeva99@gmail.com

***Аннотация*—Задача реконструкции промежуточных кадров видеопоследовательностей направлена на увеличение частоты кадров за счет синтеза новых кадров на основе информации о соседних кадрах видеопоследовательности. Более высокая частота кадров позволяет улучшить качество визуального восприятия. Также методы реконструкции промежуточных кадров получили распространение в задаче 3D-реконструкции объектов и сцен. В данной работе представлен результат сравнения качества работы современных (англ. state of the art) методов реконструкции промежуточных кадров на видеопоследовательностях с динамической сценой.**

***Ключевые слова— реконструкция кадров, динамическая сцена, XVFI, RRIN. CDFI, RIFE, AdaCof.***

## 1. Введение

Реконструкция кадров видеопоследовательности — это процесс синтеза новых кадров видеопоследовательности на основе информации о соседних кадрах. Результат процедуры реконструкции кадров позволяет повышать частоту кадров видеопоследовательности, что в результате приводит к улучшению визуального восприятия. Важную роль процедура реконструкции кадров может играть при решении задачи 3D-реконструкции объектов, когда исходная видеопоследовательность, описывающая внешний вид объекта или сцены, имеет низкую частоту съемки. Процедура реконструкции кадров позволяет получить изображение объекта с ракурса, которого не было в исходной видеопоследовательности, что в результате приводит к более точной и реалистичной реконструкции объекта.

Целью данного исследования является сравнение качества работы state of the art методов реконструкции кадров видеопоследовательностей XVFI [1], RRIN [2], CDFI [3], RIFE [4], AdaCof [5] на наборе данных, состоящем из видеопоследовательностей с динамической сценой.

Структура работы следующая. Раздел 2 посвящен краткому описанию рассматриваемых методов. В разделе 3 описывается используемый набор данных, используемые метрики качества, порядок проведения экспериментальных исследований и результаты. Работа заканчивается заключением, в котором делаются выводы на основе результатов, полученных в ходе экспериментальных исследований.

## 2. Описание рассматриваемых подходов реконструкции кадров видеопоследовательностей

XVFI [1] – это метод реконструкции кадров, основанный на использовании архитектуры нейронной сети XVFI-Net, предложенной авторами работы. XVFI-Net имеет рекурсивную многомасштабную структуру и состоит из двух каскадных модулей BIOF-I и BIOF-T, которые обучаются предсказывать двунаправленные оптические потоки между двумя входными кадрами и от истинного кадра к входными кадрам, соответственно. Результаты экспериментальных исследований, представленных в работе, показали, что предложенный авторами метод позволяет успешно фиксировать информацию об объектах, имеющих высокую динамику между соседними кадрами.

RRIN [2] – это нейросетевой подход реконструкции промежуточных кадров видеопоследовательностей, в котором используется остаточное уточнение и адаптивные веса для синтеза промежуточных кадров. Метод остаточного уточнения используется при работе с оптическим потоком и при генерации новых кадров для повышения точности и улучшении визуального вида. Адаптивная карта весов же объединяет кадры с прямой и обратной деформацией для уменьшения артефактов. Все подмодули в методе реализованы с архитектурой U-Net меньшей глубины (в отличие от оригинальной архитектуры), что обеспечивает эффективность работы метода.

CDFI [3] – это нейросетевой подход реконструкции промежуточных кадров. Авторы предлагают управляемую сжатием архитектуру сети, которая использует упрощение модели за счет оптимизации, вызывающей разреженность для значительного уменьшения размера модели при достижении хорошей производительности. Сначала авторы упрощают модель AdaCof и показывают, что 10-кратно сжатая модель AdaCof работает так же, как и оригинальная реализация, затем авторы дополнительно улучшают модель, вводя модуль деформации с несколькими разрешениями, который повышает визуальную согласованность на многослойных объектах.

RIFE [4] – это нейросетевой метод реконструкции промежуточных кадров видеопоследовательностей. Учитывая пару последовательно идущих RGB-кадров $I_0$ и $I_1$ и целевой шаг t ($0 \leqslant t \leqslant 1$), цель метода – синтезировать промежуточный кадр $\hat{I}_t$. Авторы оценивают

промежуточные оптические потоки $F_{(t\to 0)}$, $F_{(t\to 1)}$ и карту слияния M путем подачи входных кадров и t как дополнительный канал в предложенную для оценки промежуточного оптического потока нейронную сеть IFNet. В результате используют сверточный энкодер RefineNet [6] для детализации высокочастотной области $\hat{I}_t$ и снижения артефактов в результатах модели «ученика». Вычислительная стоимость аналогична использованию IFNet. RefineNet выдает остаточный результат реконструкции промежуточного кадра Δ ($-1 \leqslant \Delta \leqslant 1$). В результате получается реконструированное изображение $\hat{I}_{t+\Delta}$.

AdaCof [5] – это нейросетевой метод реконструкции промежуточных кадров видеопоследовательностей. Авторы предлагают использовать полностью свёрточную нейронную сеть, которая оценивает веса $W_{k,I}$, векторы смещения ($\alpha_{k,I}$,$\beta_{k,I}$) и карту окклюзии V. Первая часть нейронной сети представляет собой архитектуру U-Net [7], которая состоит из кодера и декодера, между слоями которых выполняется операция конкатенации карт признаков. За архитектурой U-Net следует 7 подсетей, которые окончательно оценивают выходные данные $W_{k,I}$, $\alpha_{k,I}$, $\beta_{k,I}$ для каждого кадра, а также карту окклюзии V.

## 3. Экпериментальные исследования

Для проведения экспериментальных исследований был собран набор данных, состоящий из видеопоследовательностей с динамической сценой. Характер динамики в видеопоследовательностях следующий: несколько действий, одно медленное действие, одно быстрое действие. Видеопоследовательности были сняты на камеры GoPro Hero 9 и имеют частоту 60 FPS и 240 FPS. Для создания проверочного набора, по которому производилась оценка качества реконструкции кадров, из видеопоследовательностей с частотой 60 FPS посредством удаления кадров были сформированы видеопоследовательности с частотой 30 FPS; из видеопоследовательностей с частотой 240 FPS аналогичным образом были сформированы видеопоследовательности с частотой 60 FPS и 120 FPS.

На вход методов, указанных во введении, подавались пары кадров сформированных видеопоследовательностей. Результатом работы методов являлись восстановленные на местах пропусков кадры. Качество работы методов оценивалось с помощью метрик PSNR и SSIM.

$$PSNR = 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right), \qquad (1)$$

где $MSE = \frac{1}{m \cdot n}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2$.

$$SSIM = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \qquad (2)$$

где $\mu_x$, $\mu_y$ – математическое ожидание $x, y$, $\sigma_x^2$, $\sigma_y^2$ – дисперсия $x, y$, $\sigma_x\sigma_y$ – ковариация $x$ и $y$,$c_1 = (k_1 L)^2, c_2 = (k_2 L)^2, L$ – динамический диапазон пикселей, $k_1 = 0{,}01, k_2 = 0{,}03$.

Следует отметить, что в методах [3] и [5] отсутствует возможность восстановления нескольких подряд идущих кадров (более двух).

В таблице 1 представлены лучшие результаты для каждого типа действий.

Таблица III. Лучшие результаты для каждого типа действия

| **Тип действий** | ***Входное количество кадров в секунду*** | ***Выходное количество кадров в секунду*** | ***Метод*** | **Качество работы метода** | |
|---|---|---|---|---|---|
| | | | | ***PSNR*** | ***SSIM*** |
| Одно медленное действие | 120 | 240 | CDFI [3] | 38.0691 | 0.9709 |
| Одно быстрое действие | 120 | 240 | AdaCof [5] | 37.0799 | 0.9684 |
| Несколько действий | 120 | 240 | AdaCof [5] | 37.9498 | 0.9712 |

## 4. Заключение

В результате проведенного исследования было выполнено сравнение качества работы state-of-the-art методов реконструкции кадров видеопоследовательностей, а именно XVFI [1], RRIN [2], CDFI [3], RIFE [4], AdaCof [5], на собранном наборе данных, состоящем из видеопоследовательностей с динамической сценой. Полученные результаты показывают довольно хорошее качество реконструкции промежуточных кадров, однако с появлением сильной динамики визуальное и количественное качество методов снижается. В связи с этим в следующих работах планируется разработать метод, устойчивый к сильной динамике.

## Литература

[1] Sim, H. XVFI: eXtreme Video Frame Interpolation / H. Sim, J. Oh, M. Kim // arXiv: 2103.16206, 2021.

[2] Li, H. Video Frame Interpolation Via Residue Refinement / H. Li, Y. Yuan, Q. Wang // ICASSP – IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). – 2020. – P. 2613-2617. DOI: 10.1109/ICASSP40776.2020.9053987.

[3] Ding, T. CDFI: Compression-Driven Network Design for Frame Interpolation / T. Ding, L. Liang, Z. Zhu, I. Zharkov // arXiv: 2103.10559, 2021.

[4] Huang, Z. RIFE: Real-Time Intermediate Flow Estimation for Video Frame Interpolation / Z. Huang, T. Zhang, W. Heng, B. Shi, S. Zhou // arXiv: 2011.06294, 2021.

[5] Lee, H. AdaCoF: Adaptive Collaboration of Flows for Video Frame Interpolation / H. Lee, T. Kim, T. Chung, D. Pak, Y. Ban, S. Lee // arXiv: 1907.10244, 2020.

[6] Lin, G. RefineNet: Multi-Path Refinement Networks for High-Resolution Semantic Segmentation / G. Lin, A. Milan, C. Shen, I. Reid // arXiv: 1611.06612, 2016.

[7] Ronneberger, O. U-Net: Convolutional Networks for Biomedical Image Segmentation / O. Ronneberger, P. Fischer, T. Brox // arXiv: 1505.04597, 2015.