

Задание

- **Подготовка данных:** Выбрать тестовый видеоклип с фиксированной исходной частотой ≈ 30 fps (рекомендуется сцена с динамичным движением или панорамированием камеры).
- **Конфигурация графического движка:** Обработка выполняется на дискретной видеокарте Nvidia RTX 3060 с использованием CUDA-ускорения.
 - Выбрать базовую нейросетевую модель для интерполяции — семейство RIFE (Real-time Intermediate Flow Estimation), версия HDv3.
- **Настройка выходных параметров:**
 - Режим работы: интерполяция (Interpolation).
 - Коэффициент масштабирования кадров Multiplier: 8x для получения итоговой частоты ≈ 240 fps.

Реализация

Инструмент: Jupyter Notebook (Python 3) с использованием PyTorch 2.12 + CUDA 13.0.

Подготовка настроек

Исходный код лабораторной работы доступен в репозитории:
<https://git.illegalfiles.icu/vlad.os/multimedia-lab3-video>

В качестве исходного видео была взята запись длительностью 6.3 секунды с разрешением 1920×1080 пикселей, частотой кадров 29.97 fps и общим количеством кадров 189. Обработка выполнялась с использованием библиотеки PyTorch на GPU Nvidia GeForce RTX 3060 (12 ГБ).

Настройки модели RIFE HDv3:

- Предобученная модель IFNet с весами RIFE HDv3
- Масштаб обработки: 1.0 (полное разрешение)
- Режим: fp32 (полная точность)
- Multiplier: 8x (2^3)

Архитектура нейросети

RIFE основан на итеративной многомасштабной сети IFNet (Intermediate Flow Network), которая состоит из трёх последовательных блоков IFBlock. Каждый блок обрабатывает изображение на своём масштабе (4x, 2x, 1x) и итеративно уточняет оптический поток и маску слияния. Процесс интерполяции между двумя кадрами включает следующие шаги:

1. Оценка двунаправленного оптического потока между кадрами I_0 и I_1 на нескольких масштабах

2. Обратная свертка (warping) исходных кадров согласно найденному потоку
3. Вычисление маски слияния для определения вклада каждого из искажённых кадров
4. Синтез промежуточного кадра: $I_{mid} = warp(I_0, F_{0 \rightarrow t}) \cdot M + warp(I_1, F_{1 \rightarrow t}) \cdot (1 - M)$

Для кратного увеличения частоты (8x) применяется рекурсивный подход: между каждой парой исходных кадров вычисляется средний, затем процесс рекурсивно повторяется для каждого полученного интервала.

Исходные кадры

На рис. 1 представлены исходные кадры видео. Видно, что кадры содержат сцену с динамичным движением, что позволяет оценить эффективность интерполяции.

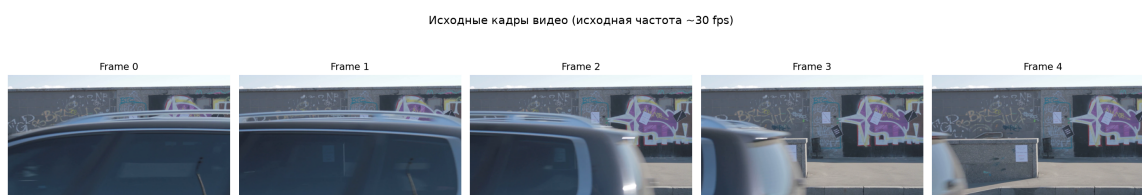


Рис. 1: Исходные кадры видео (исходная частота ≈ 30 fps)

Выходные кадры

На рис. 2 представлены результаты интерполяции — 7 промежуточных кадров, синтезированных между двумя исходными кадрами с использованием RIFE.



Рис. 2: Демонстрация интерполяции 8x: исходный кадр 0, 7 промежуточных кадров ($t = 0.125 \dots 0.875$), исходный кадр 1

На рис. 3 показано сравнение исходных и интерполированных кадров. В левом столбце — исходные кадры, в центральном — соответствующие им кадры из интерполированного видео (каждый 8-й), в правом — промежуточные кадры, синтезированные сетью.



Рис. 3: Сравнение исходных и интерполированных кадров

На рис. 4 представлен анализ межкадровой разницы: в исходном видео средняя разница между соседними кадрами составляет 4.4, в то время как в интерполированном — 0.86. Это подтверждает, что RIFE успешно синтезирует плавные промежуточные кадры, уменьшая межкадровую разницу в 5.2 раза.

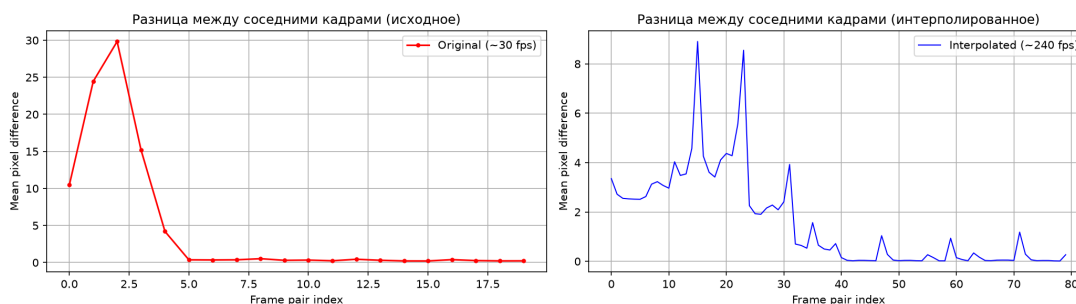


Рис. 4: Анализ межкадровой разницы: исходное (слева) и интерполированное (справа) видео

Вывод

В ходе выполнения лабораторной работы была изучена и реализована технология нейросетевой интерполяции кадров с использованием архитектуры RIFE (Real-time Intermediate Flow Estimation). Экспериментальным путем было установлено, что алгоритм успешно справляется с генерацией промежуточных кадров при увеличении частоты видеопотока в 8 раз (с ≈ 30 до ≈ 240 fps), обеспечивая высокую плавность динамичных сцен.

Ключевые результаты работы:

- Разработан Jupyter Notebook с полным пайплайном интерполяции на базе IFNet/RIFE HDv3
- Выполнена обработка тестового видеофрагмента (1920×1080, 6.31 сек, 189 кадров)
- Получено итоговое видео с частотой ≈ 240 fps и общим количеством 1505 кадров (8.0x)
- Выполнен количественный анализ: межкадровая разница уменьшилась с 4.4 до 0.86 (в 5.2 раза), подтверждая равномерность синтезированных кадров