



Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет
имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ ИУ «Информатика и системы управления»

КАФЕДРА ИУ-7 «Программное обеспечение ЭВМ и информационные технологии»

**РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА
К НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ
НА ТЕМУ:**

***«Классификация известных методов увеличения
разрешения видеопотока»***

Студент ИУ7-73Б _____ Марченко В.

Руководитель НИР _____ Тассов К. Л.

Рекомендуемая руководителем НИР оценка _____

РЕФЕРАТ

Отчет 26 с., 10 рис., 1 табл., 8 источн., 1 прил.

ВИДЕО, ВИДЕОПОТОК, ВИДЕОИЗОБРАЖЕНИЕ, РАЗРЕШЕНИЕ, НЕЙРОННЫЕ СЕТИ

Объектом исследования являются методы увеличения разрешения видеопотока.

Цель работы: классификация известных методов увеличения разрешения видеопотока.

В результате исследования было проведено сравнение 4-х методов увеличения разрешения видеопотока по 4-м критериям.

Область применения результатов — выбор метода для увеличения разрешения видеопотока.

Результат работы. Нельзя точно сказать, какой из рассмотренных методов увеличения разрешения видеопотока является наилучшим. Самый высокий показатель пикового отношения сигнала к шуму у метода нелокальной регуляризации и надежной оценки движения, в то время как у метода динамических фильтров увеличения разрешения без явной компенсации движения наилучший показатель индекса структурного сходства. При этом второй метод позволяет увеличивать разрешение видеопотока в 4 раза. Наихудшим из рассмотренных по всем показателям оказался метод бикубической интерполяции.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	6
1 Анализ предметной области	8
1.1 Суперразрешение видеопотока	8
1.2 Понижение разрешения	9
1.3 Подходы к увеличению разрешения видеопотока	9
2 Методы увеличения разрешения видеопотока	11
2.1 Бикубическая интерполяция	11
2.2 Метод, основанный на нелокальной регуляризации и надежной оценке движения	12
2.3 Методы, основанные на использовании нейронных сетей	14
2.3.1 Динамические фильтры увеличения разрешения без яв- ной компенсации движения	15
2.3.2 Остаточная обратимая пространственно-временная ней- ронная сеть	17
3 Классификация методов увеличения разрешения видеопотока	22
3.1 Критерии оценки методов увеличения разрешения видеопотока	22
3.2 Сравнение методов увеличения разрешения видеопотока	23
ЗАКЛЮЧЕНИЕ	24
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	25
ПРИЛОЖЕНИЕ А Презентация	26

ПЕРЕЧЕНЬ СОКРАЩЕНИЙ И ОБОЗНАЧЕНИЙ

В настоящей расчетно-пояснительной записке к научно-исследовательской работе применяют следующие сокращения и обозначения:

VSR	Суперразрешение видео (video super-resolution)
RGB	Красный, зеленый, синий (red, green, blue)
SISR	Суперразрешение фото (single-image super-resolution)
DFT	Дискретное преобразование Фурье (discrete Fourier transform)
DCT	Дискретное косинусное преобразование (discrete cosine transform)
DWT	Дискретное вейвлет-преобразование (discrete wavelet transform)
POCS	Проецирование в выпуклые множества (projections onto convex sets)
RLS	Рекуррентный метод наименьших квадратов (recursive least squares)
MAP	Оценка апостериорного максимума (maximum a posteriori probability)
MLE	Метод максимального правдоподобия (maximum likelihood estimation)
MRF	Марковское случайное поле (Markov random field)
SSIM	Индекс структурного сходства (structure similarity)
PSNR	Пиковое отношение сигнала к шуму (peak signal-to-noise ratio)

ВВЕДЕНИЕ

Суперразрешение — это способ получения видеопотока (изображения) с высоким разрешением из видеопотока (изображения) низкого разрешения [1]. В отличие от суперразрешения одного изображения (SISR), основная цель суперразрешения видео — не только восстановить больше мелких деталей при сохранении крупных, но и сохранить согласованность движения.

Во многих областях, работающих с видео, люди имеют дело с различными типами деградации видео, включая понижение разрешения. Разрешение видео может снизиться из-за несовершенства устройств фото- и видеосъемки. Плохое освещение и погодные условия добавляют шум. Движение объектов и камеры также ухудшает качество видео. Методы суперразрешения помогают восстановить исходное видео. Это полезно в широком спектре приложений, таких как [2]:

- 1) видеонаблюдение (для улучшения качества видео, снятого с камеры, а также распознавания номеров автомобилей и лиц);
- 2) медицинская визуализация (чтобы лучше обнаружить некоторые органы или ткани для клинического анализа и медицинского вмешательства);
- 3) судебно-медицинская экспертиза (для помощи в расследовании в ходе уголовного процесса);
- 4) астрономия (для улучшения качества видео звезд и планет);
- 5) дистанционное зондирование (для облегчения наблюдения за объектом);
- 6) микроскопия (для усиления возможностей микроскопов).

Суперразрешение видео также помогает решить задачу обнаружения объектов, распознавания лиц и символов (в качестве этапа предварительной обработки).

Суперразрешение видео является давней сложной задачей, главным образом по следующим двум причинам: эта задача по своей сути является некорректно поставленной из-за характера отображения «один ко многим» (один кадр низкого разрешения может отображаться в различные кадры высокого разрешения) и на сегодняшний день не существует удовлетворительной

архитектуры, предназначенной для интеграции пространственной и временной информации в единую структуру [3].

Цель научно-исследовательской работы: провести обзор известных методов увеличения разрешения видеопотока и классифицировать их по сформулированным критериям.

Задачи научно-исследовательской работы:

- 1) исследовать предметную область увеличения разрешения видеопотока;
- 2) проанализировать известные методы увеличения разрешения видеопотока;
- 3) сформулировать критерии для сравнения этих методов;
- 4) сравнить методы увеличения разрешения видеопотока по сформулированным критериям.

1 Анализ предметной области

1.1 Суперразрешение видеопотока

Суперразрешение — это набор действий, выполняемых с целью получения изображения (или последовательности изображений) высокого разрешения из последовательности изображений низкого разрешения. Концепция суперразрешения представлена на рисунке 1.1. Суперразрешение позволяет получить изображение или видео повышенного качества с большим количеством деталей на сцене, что важно для точного анализа [2].



Рисунок 1.1 – Концепция суперразрешения [2]

Суперразрешение может быть оптическим и геометрическим. В оптических методах используются характеристики оптики, датчиков и компонентов дисплея устройства визуализации, которые отвечают за ухудшение качества или понижение разрешения изображения. Улучшение пространственного разрешения устройства визуализации может быть достигнуто путем модификации аппаратного обеспечения двумя способами [2]: увеличить количество пикселей (но есть ограничения, т. к. это уменьшает отношение сигнал/шум и увеличивает время получения изображения) и увеличить размер чипа, необходимого для получения изображений высокого разрешения (такие чипы достаточно дорогие) [1].

Хорошей альтернативой обоим подходам является использование метода автономного улучшения разрешения, то есть геометрического суперразрешения. В этом типе суперразрешения для восстановления и реконструкции изображения используются методы цифровой обработки изображений [2].

Благодаря широкой применимости концепции суперразрешения это одно из наиболее быстро развивающихся направлений в области обработки изображений [4].

1.2 Понижение разрешения

На рисунке 1.2 показан процесс понижения разрешения изображения.

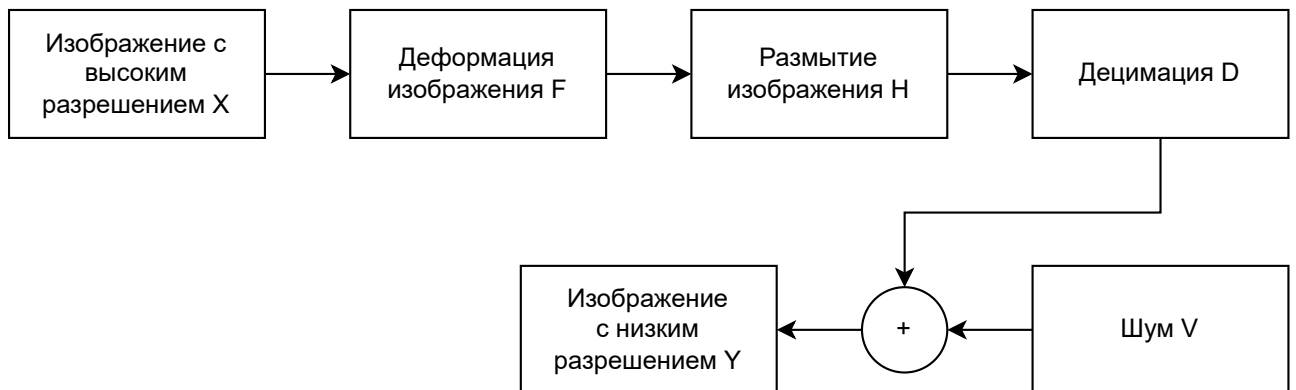


Рисунок 1.2 – Процесс понижения разрешения изображения [2]

Приведенный процесс можно записать с помощью формулы:

$$Y = D * H * F * X + V, \quad (1.1)$$

где Y — изображение с низким разрешением, H — коэффициент размытия (которое появляется из-за особенностей камеры), D — коэффициент децимации, F — деформация, а V — коэффициент шума [2].

В приведенной выше формуле факторами деградации являются F , H , D и V . Если эти коэффициенты известны разработчику, то система называется системой с предварительно известными данными, а изображение с высоким разрешением получается путем решения математического уравнения 1.1 [2].

1.3 Подходы к увеличению разрешения видеопотока

Самый простой способ реализовать суперразрешение видео — покадровый запуск алгоритмов суперразрешения фото. Однако, поскольку методы суперразрешения фото не учитывают временные отношения между кадрами, существует высокая вероятность того, что последовательные кадры не будут соединены естественным образом, что приведет к мерцающим артефактам [5].

Суперразрешение осуществляется или покадрово, или используя сразу несколько кадров. Субпиксельный сдвиг между последовательными кадрами используется для восстановления кадров высокого разрешения в многокадровых методах суперразрешения. Однокадровые методы стремятся улучшить

качество изображения без добавления размытия. Алгоритмы суперразрешения работают в двух областях — частотной и пространственной. На рисунке 1.3 представлены некоторые методы суперразрешения видео [2].

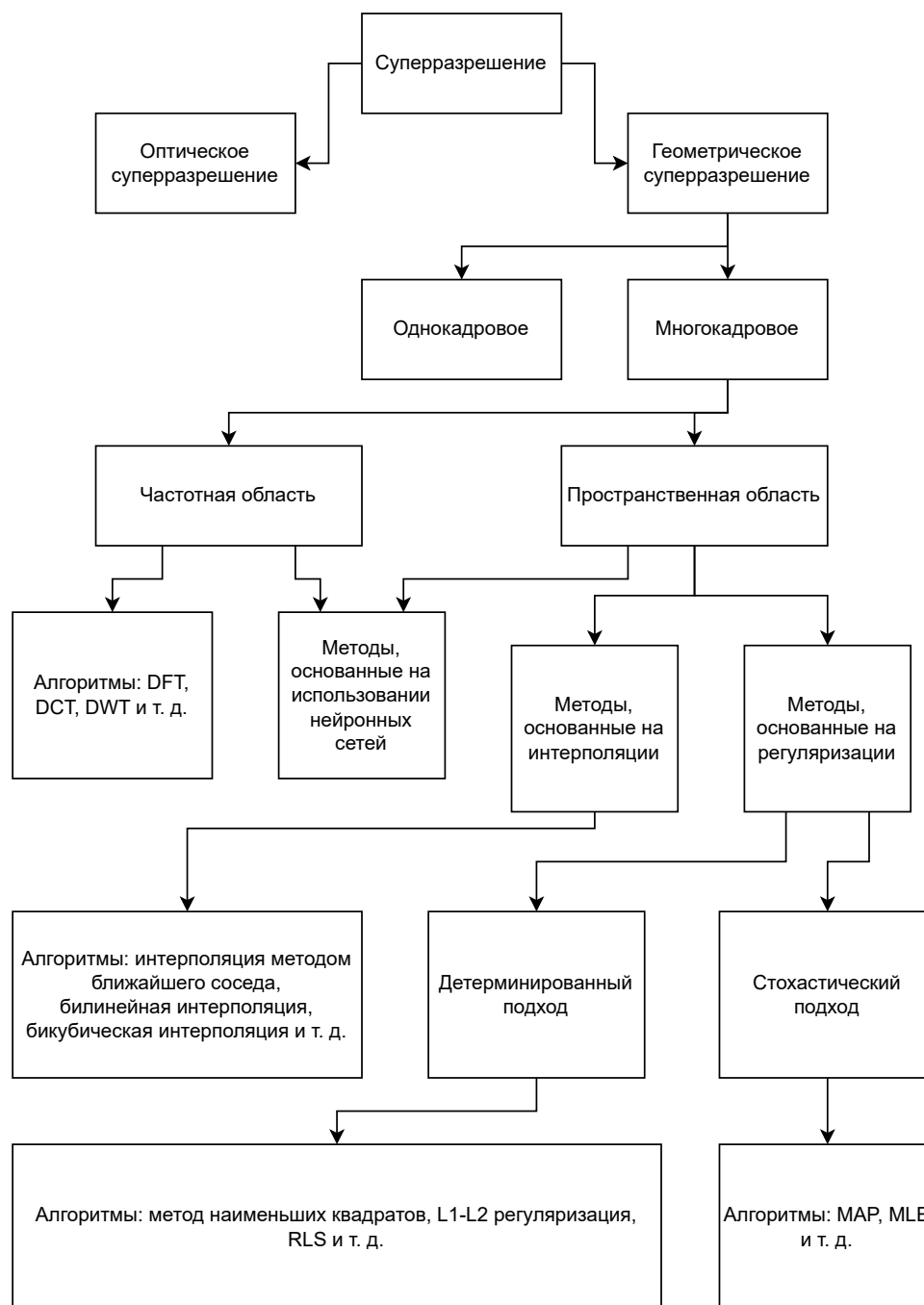


Рисунок 1.3 – Некоторые методы суперразрешения видеопотока [2]

2 Методы увеличения разрешения видеопотока

2.1 Бикубическая интерполяция

В пространственной области процесс восстановления происходит путем обработки на уровне пикселей вместо работы с каким-либо признаком изображения. Алгоритмы, работающие в пространственной области, в основном делятся на интерполяционные и регуляризационные [2].

Самый простой способ повысить разрешение изображения — интерполяция. Процесс интерполяции — это оценка нового пикселя с помощью заданного набора пикселей. Регистрация, интерполяция и восстановление — три основных этапа интерполяционных методов суперразрешения [6]. Геометрическое выравнивание происходит при регистрации изображений, при которой изображения низкого разрешения выравниваются по одному конкретному изображению низкого разрешения, используемому в качестве эталона. Смещения и повороты субпикселей необходимы для точной оценки параметров движения перед их объединением для создания изображения высокого разрешения [2].

Бикубическая интерполяция является более трудоемким методом, чем интерполяция методом «ближайшего соседа» и билинейная интерполяция. В отличие от билинейной интерполяции, коэффициенты бикубической интерполяции в 16-и ближайших пикселях обеспечивают более гладкую поверхность, поскольку связь с известными пикселями имеет полиномиальный, а не линейный характер. Полином третьей степени с коэффициентами a_0, \dots, a_3 записывается как $f(x) = \sum_{i=0}^3 a^i x^i$. Предположим, есть четыре известные точки $f(p_1), f(p_2), f(p_3), f(p_4)$, которые окружают неизвестную точку $f(p)$, где f — некий полином третьей степени. Запишем функцию f и известные точки p_1, \dots, p_4 в матричном представлении [7]:

$$y = \begin{bmatrix} f(p_1) & f(p_2) & f(p_3) & f(p_4) \end{bmatrix}^T, \quad (2.1)$$

$$B = \begin{bmatrix} (p_1)^3 & (p_1)^2 & (p_1)^1 & (p_1)^0 \\ (p_2)^3 & (p_2)^2 & (p_2)^1 & (p_2)^0 \\ (p_3)^3 & (p_3)^2 & (p_3)^1 & (p_3)^0 \\ (p_4)^3 & (p_4)^2 & (p_4)^1 & (p_4)^0 \end{bmatrix}, \quad (2.2)$$

$$a = \begin{bmatrix} a_3 & a_2 & a_1 & a_0 \end{bmatrix}^T, \quad (2.3)$$

$$y = Ba. \quad (2.4)$$

Чтобы посчитать $f(p)$, степени p_1, \dots, p_4 умножаются на коэффициенты a_0, \dots, a_3 . Так как константы a_0, \dots, a_3 неизвестны, они выражаются из предыдущих уравнений, что приводит к уравнениям [7]:

$$b = \begin{bmatrix} p^3 & p^2 & p^1 & p^0 \end{bmatrix}, \quad (2.5)$$

$$f(p) = ba = bB^{-1}y. \quad (2.6)$$

Бикубическая интерполяция — это кубическая интерполяция по двум измерениям для поиска поверхности, а не прямой. На рисунке 2.1 показаны распространенные артефакты, возникающие при использовании методов интерполяции: алиасинг (эффект «ступенчатости» изображения), сглаживание и эффект гало [7].



Рисунок 2.1 – Распространенные артефакты, возникающие при использовании методов интерполяции [7]

2.2 Метод, основанный на нелокальной регуляризации и надежной оценке движения

Данный метод состоит из двух шагов: регистрация, где оценивается движение между изображениями низкого разрешения, и восстановление изображения, при котором изображение высокого разрешения восстанавливается

из изображений низкого разрешения с использованием информации о движении и размытии [8].

Основан метод на многостороннем фильтре, который используется для регуляризации процесса оценки движения. Этот фильтр может адаптивно корректировать оценку движения в соответствии с надежностью оценки, неоднородностью интенсивности изображения и несходством движения [8].

На рисунке 2.2 изображена диаграмма суперразрешения видео. Исходная видеопоследовательность высокого разрешения создается путем деформации опорного кадра (красный прямоугольник) несколькими полями движения. Затем последовательность кадров высокого разрешения сглаживается с помощью ядра размытия, субдискретизируется и загрязняется шумом для создания наблюдаемой последовательности кадров низкого разрешения. Предлагаемый метод суперразрешения видео не только оценивает последовательность кадров высокого разрешения, но также движение, ядро размытия и веса соответствия [8].

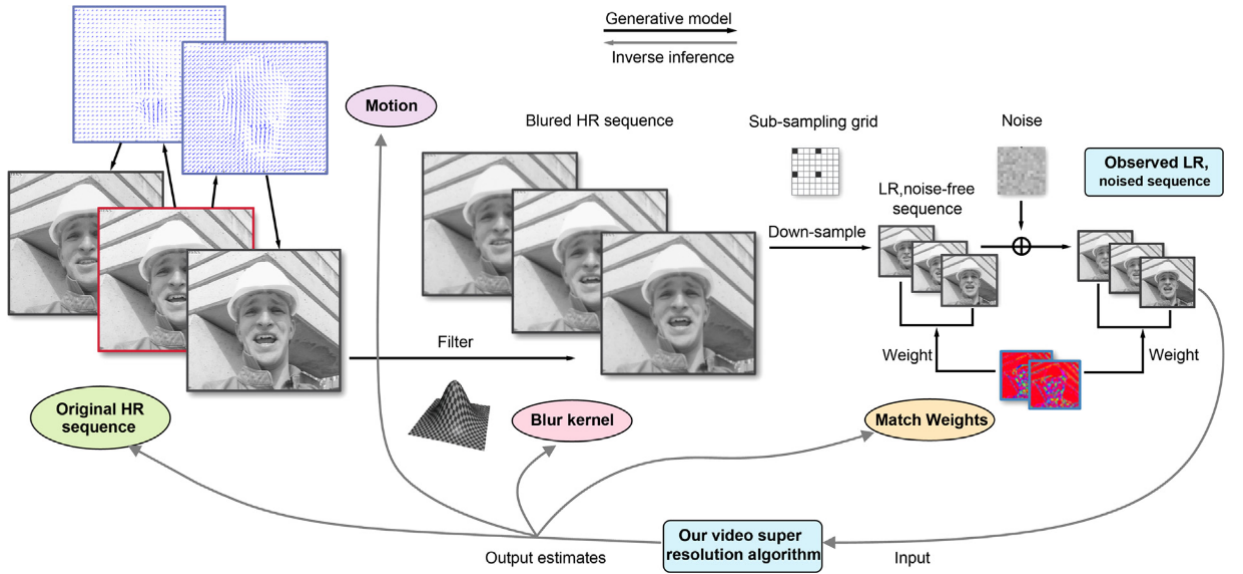


Рисунок 2.2 – Диаграмма суперразрешения видео [8]

Пусть z — «размытое» изображение высокого разрешения, F_t — поле движения и W_t — матрица весов. Для решения задачи регистрации используется метод оценки апостериорного максимума [8]:

$$\{z^*, \{F_t\}^*, \{W_t\}^*\} = \arg \max p(z, \{F_t\}, \{W_t\} | \{y_t\}). \quad (2.7)$$

В этой формуле апостериорная вероятность — произведение априорной веро-

ятности на функцию правдоподобия [8]:

$$p(z, \{F_t\}, \{W_t\}|\{y_t\}) = p(\{y_t\}|z, \{F_t\}, \{W_t\}) \cdot p(z) \cdot p(\{F_t\}, \{W_t\}). \quad (2.8)$$

В данной формуле $p(\{y_t\}|z, \{F_t\}, \{W_t\})$ — функция правдоподобия, а $p(z)$ и $p(\{F_t\}, \{W_t\})$ — априорные значения скрытого изображения и полей движения соответственно [8].

Условное распределение кадра y_t низкого разрешения:

$$p(y_t|z, F_t, W_t) = \exp\left(-\frac{\|DF_t z - y_t\|_{W_t}^2}{2\sigma^2}\right), \quad (2.9)$$

где запись $\|a\|_W^2 = a^T W a$ представляет взвешенную сумму квадратов. Предполагая статистическую независимость гауссовского шума среди полученных изображений низкого разрешения, условное распределение набора этих изображений y_t может быть записано как [8]:

$$p(\{y_t\}|z, \{F_t\}, \{W_t\}) = \prod_{t=1}^T p(y_t|z, F_t, W_t) = \exp\left(-\frac{\|DF_t z - y_t\|_{W_t}^2}{2\sigma^2}\right). \quad (2.10)$$

С помощью приведенной оценки апостериорного максимума изображение высокого разрешения и оценка движения могут постепенно уточняться альтернативным и итеративным способами [8].

2.3 Методы, основанные на использовании нейронных сетей

Традиционные методы суперразрешения видео используют несколько кадров низкого разрешения в качестве входных данных и на выходе выдают кадры высокого разрешения, принимая во внимание субпиксельные движения между соседними кадрами низкого разрешения. Все методы суперразрешения видео, основанные на глубоком обучении, работают именно по этому принципу и состоят из двух этапов: оценки движения и процедуры компенсации, за которой следует процесс увеличения разрешения. Одна из проблем этого двухэтапного подхода заключается в том, что результаты во многом зависят от точной оценки движения. Другая потенциальная проблема заключается в том, что выходной кадр высокого разрешения создается путем смешивания

значений из нескольких входных кадров низкого разрешения с компенсацией движения через сверточные нейронные сети, что может привести к размытому выходному кадру высокого разрешения [5].

2.3.1 Динамические фильтры увеличения разрешения без явной компенсации движения

В этом методе вместо явного вычисления и компенсации движения между входными кадрами, информация о движении неявно используется для генерации динамических фильтров увеличения разрешения. С помощью сгенерированных фильтров кадр высокого разрешения напрямую строится путем локальной фильтрации входного центрального кадра. Поскольку этот метод не полагается на явное вычисление движения и не объединяет напрямую значения из нескольких кадров, можно создавать гораздо более четкие и согласованные по времени видео высокого разрешения [5].

На рисунке 2.3 показан пример масштабирования пикселя $(3, 3)$ центрального входного кадра X_t с помощью коэффициента масштабирования $r = 4$. Шестнадцать сгенерированных фильтров от $F_t^{3,3,0,0}$ до $F_t^{3,3,3,3}$ используются для создания шестнадцати пикселей в области от $(12, 12)$ до $(15, 15)$ кадра \hat{Y}_t высокого разрешения [5].

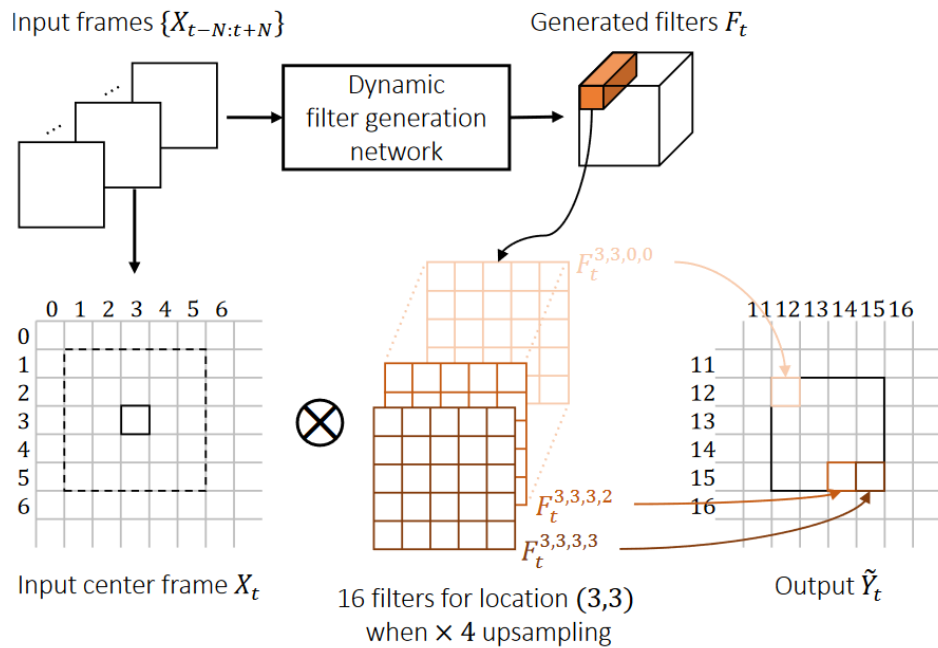


Рисунок 2.3 – Пример масштабирования пикселя [5]

Цель суперразрешения видео — оценить кадры $\{\hat{Y}_t\}$ высокого разре-

ния по последовательности кадров $\{X_t\}$ низкого разрешения. Кадры $\{X_t\}$ низкого разрешения — это субдискретизированные исходные кадры $\{Y_t\}$, где t — шаг по времени. С предложенной нейронной сетью G и параметрами сети θ задача суперразрешения видео определяется как:

$$\hat{Y}_t = G_\theta(X_{t-N:t+N}), \quad (2.11)$$

где N — временной радиус. Форма входного массива данных для G — $T \times H \times W \times C$, где $T = 2N + 1$, H и W — высота и ширина входного кадра низкого разрешения, а C — количество цветовых каналов. Соответствующая форма выходного массива данных — $1 \times rH \times rW \times C$, где r — коэффициент масштабирования [5].

Нейронная сеть G на выходе дает два значения для генерации конечного кадра высокого разрешения \hat{Y}_t из множества кадров низкого разрешения $\{X_{t-N:t+N}\}$: динамические фильтры F_t увеличения разрешения и остаток R_t . Входной центральный кадр X_t сначала локально фильтруется с помощью динамических фильтров F_t увеличения разрешения, а затем остаток R_t добавляется к результату для окончательного вывода \hat{Y}_t .

На рисунке 2.4 показана архитектура нейронной сети.

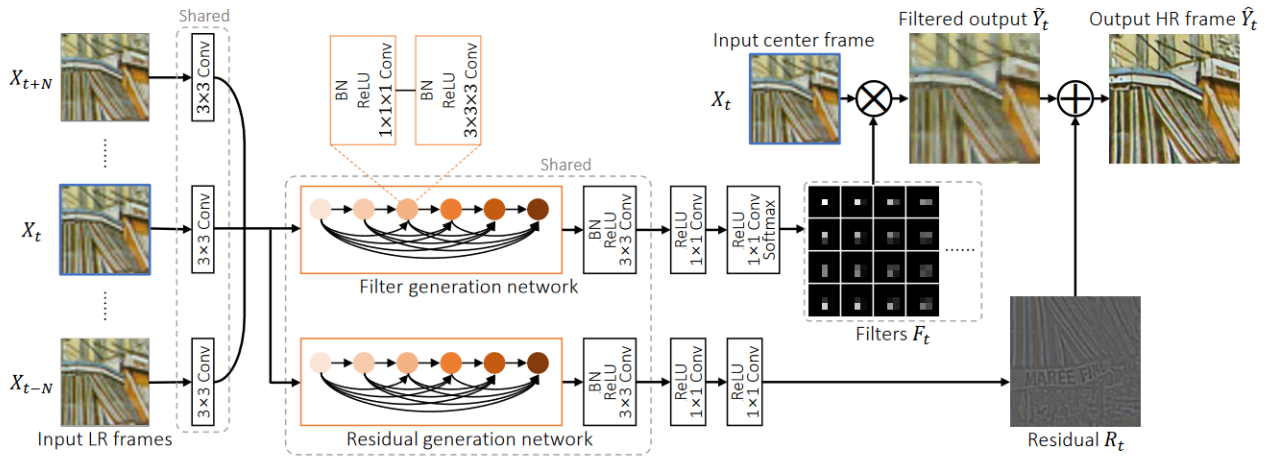


Рисунок 2.4 – Архитектура нейронной сети [5]

Динамические фильтры увеличения разрешения. Сначала множество входных кадров $\{X_{t-N:t+N}\}$ низкого разрешения попадают на вход сети генерации динамических фильтров. Обученная сеть выдает множество $r^2 HW$ фильтров F_t увеличения разрешения определенного размера, которые затем

используются для генерации новых пикселей отфильтрованного кадра \hat{Y}_t . Далее создаются выходные пиксели высокого разрешения с помощью локальной фильтрации входного кадра X_t с помощью соответствующего фильтра:

$$\hat{Y}_t(yr + v, xr + u) = \sum_{j=-2}^2 \sum_{i=-2}^2 F_t^{y,x,v,u}(j + 2, i + 2) X_t(y + j, x + i), \quad (2.12)$$

где y и x — координаты пикселя кадра низкого разрешения, v и u — координаты каждого выходного блока $r \times r$ ($0 \leq v, u \leq r - 1$). Эта операция аналогична обратной свертке, поэтому данную сеть можно обучать сквозным образом, поскольку она допускает обратное распространение ошибки [5].

Добавление остатка. Результату после применения динамических фильтров увеличения разрешения не хватает резкости, поскольку он представляет собой взвешенную сумму входных пикселей. Могут быть детали, которые невозможно восстановить с помощью линейной фильтрации. Чтобы решить эту проблему, дополнительно оценивается остаточное изображение, чтобы увеличить детализацию [5].

2.3.2 Остаточная обратимая пространственно-временная нейронная сеть

В данном методе используется сеть, которая состоит из трех компонентов: пространственная составляющая, временная составляющая и составляющая восстановления (реконструкции). В пространственном компоненте остаточный обратимый блок предназначен для извлечения информативных признаков с помощью пространственной информации. Во временном компоненте используется остаточная плотная сверточная длинная кратковременная память (RDC-LSTM) для изучения последовательного представления признаков. Компонент реконструкции используется для интеграции пространственных и временных характеристик в единую структуру. На рисунке 2.5 показана структура остаточной обратимой пространственно-временной сети [3].

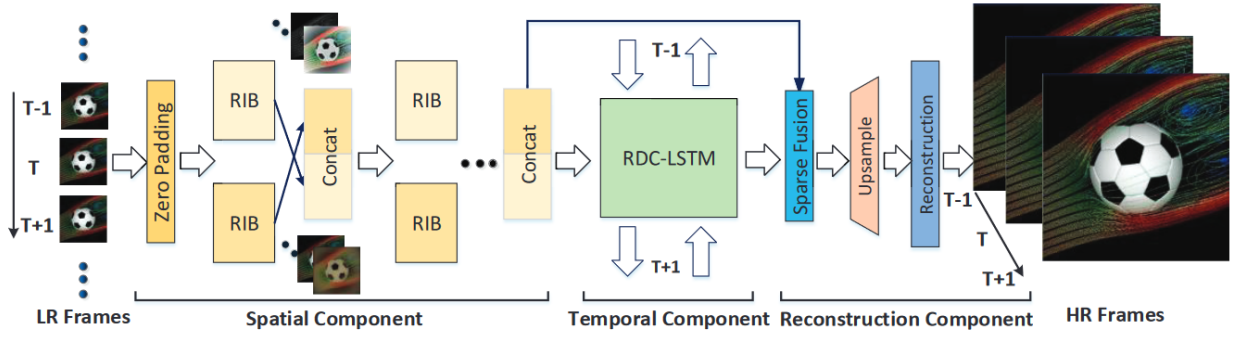


Рисунок 2.5 – Структура остаточной обратимой пространственно-временной сети [3]

В пространственном компоненте последовательные кадры низкого разрешения подаются на слой дополнения, который создает исходные карты признаков путем дополнения нулями в каналах RGB. Два последующих параллельных остаточных обратимых блока имеют разную архитектуру с разным количеством слоев для использования иерархических признаков. Выходные карты признаков предыдущего остаточного обратимого блока будут объединены и затем помещены в следующие параллельные остаточные обратимые блоки. Примечательно, что объединение может эффективно увеличить разнообразие карт признаков. Во временном компоненте предлагается использовать остаточную плотную сверточную сеть с длинной краткосрочной памятью для обработки признаков непрерывных кадров. В компоненте реконструкции используется метод объединения разреженных признаков для интеграции пространственных и временных карт признаков, причем объединенные карты признаков подвергаются увеличению разрешения до целевого размера высокого разрешения. Наконец, слой реконструкции используется для восстановления кадров высокого разрешения RGB-канала [3].

Конечная цель суперразрешения видео — обучить производящую функцию F , которая оценивает кадры высокого разрешения по входным кадрам низкого разрешения. Пусть I_T^{LR} — входные кадры низкого разрешения, I^{HR} — исходные кадры высокого разрешения, тогда задача суперразрешения видео может быть описана следующим образом:

$$I_T^{HR} = F(\{I_T^{LR}, I_{T+i}^{LR}\}), i \in \{\pm 1, \dots, \pm k\}, \quad (2.13)$$

где T — текущая временная метка, i — последовательная i -я временная

метка [3].

Остаточный обратимый блок. Кадры высокого разрешения должны иметь структуру, аналогичную входным кадрам низкого разрешения — это важное свойство называется пространственной информацией. В текущем методе используется остаточный обратимый блок, в котором создается остаточное соединение, а параллельный обратимый блок предназначен для изучения разницы между кадрами низкого и высокого разрешения. На рисунке 2.6 показана архитектура остаточного обратимого блока. Знак \oplus означает поэлементное сложение [3].

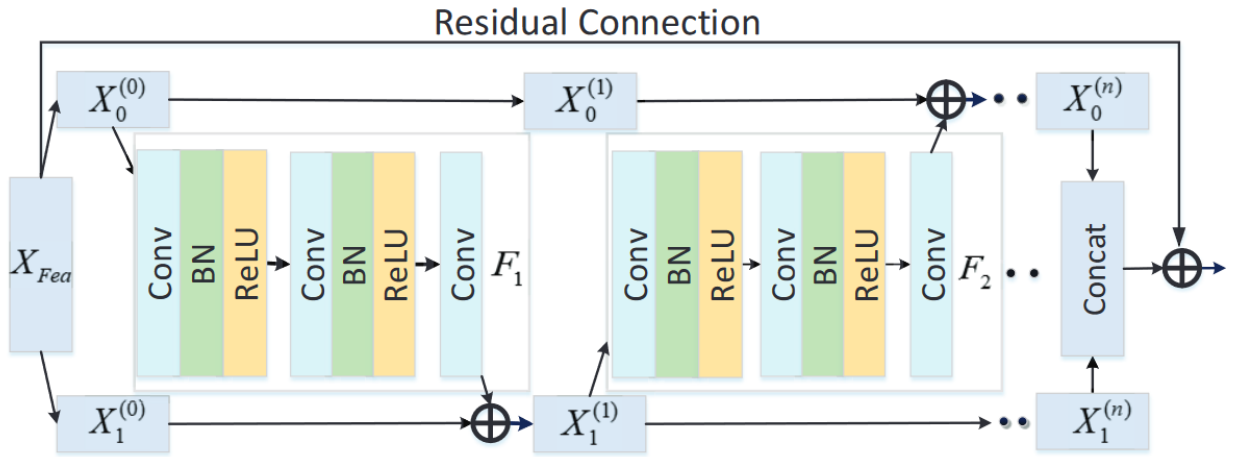


Рисунок 2.6 – Архитектура остаточного обратимого блока [3]

На рисунке показано, что входные признаки F_{fea} делятся на два под-слоя $X_0^{(0)}$ и $X_1^{(0)}$. Далее определяется сверточное бутылочное горлышко F_i , $i \in [1, 2, \dots, n - 1]$. Сверточное бутылочное горлышко состоит из слоев свертки, пакетной нормализации и срезанных линейных узлов. Признаки $X_1^{(i-1)}$ и $X_0^{(i-1)}$ могут быть получены по формулам:

$$X_1^{(i-1)} = X_1^{(i)} - F_i(X_0^{(i-1)}), \quad (2.14)$$

$$X_0^{(i-1)} = X_0^{(i)}. \quad (2.15)$$

Согласно приведенным выше формулам, предыдущие признаки могут быть последовательно выведены из любого $X_1^{(i)}$ и $X_0^{(i)}$. Таким образом, результат работы пространственного компонента можно записать в следующем виде:

$$X_{out} = [X_0^{(n)}, X_1^{(n)}] + X_{fea}, \quad (2.16)$$

где , обозначает объединение карт признаков [3].

Рекуррентная модель с короткими соединениями. Во временном компоненте используется сверточная долгая краткосрочная память для определения информативных признаков последовательных кадров. В отличие от обычного одномерной долгой краткосрочной памяти, сверточная захватывает двумерные признаки из соседних временных меток. Для тщательного использования временной согласованности сверточная долгая краткосрочная память построена как двунаправленная архитектура, в которой выходные данные прямого и обратного хода объединяются и образуют выходные данные одного нейрона. На рисунке 2.7 показаны архитектуры различных модификаций сверточной долгой краткосрочной памяти [3].

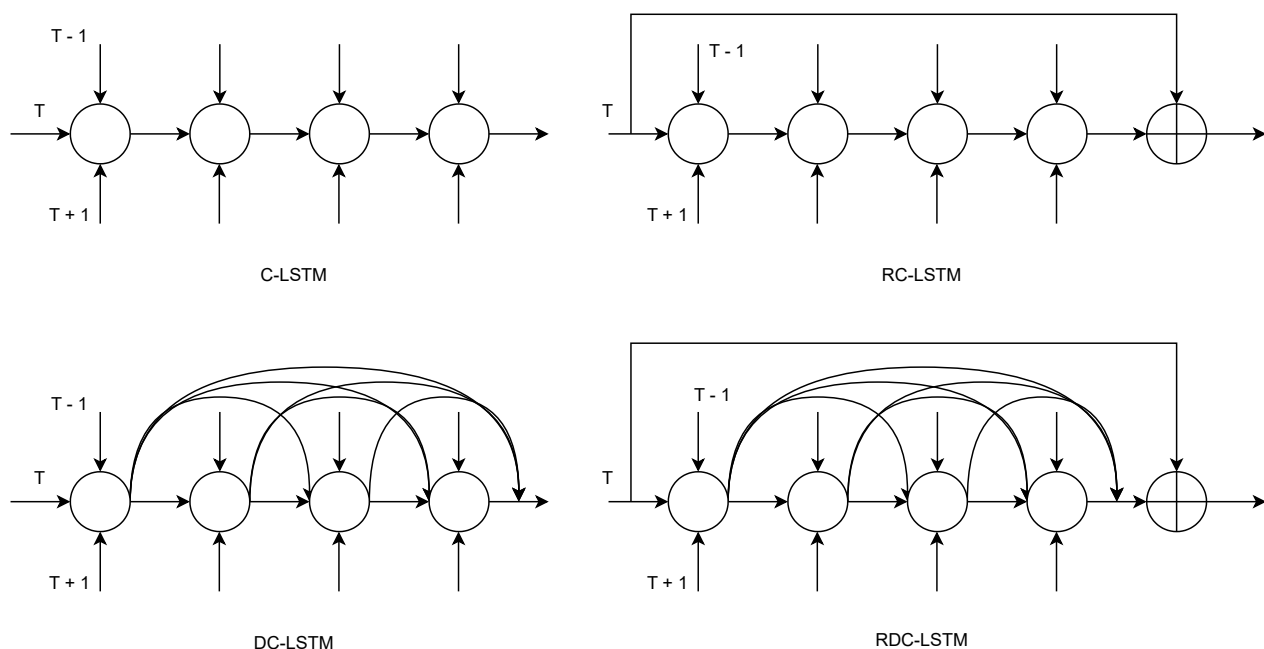


Рисунок 2.7 – Архитектуры различных модификаций сверточной долгой краткосрочной памяти [3]

Результат работы временного компонента можно записать в следующем виде:

$$X_{out} = W_{1 \times 1 \times c \times c'} * X_{in} + [H_0, H_1, \dots, H_{n-1}]_{c'}, \quad (2.17)$$

где $[H_0, H_1, \dots, H_{n-1}]$ — конкатенация карт признаков, полученных на всех предыдущих слоях, X_{in} и X_{out} — входные и выходные данные временного компонента, W — матрица сверточного фильтра размера 1×1 , c — исходное количество цветовых каналов, а $*$ обозначает операцию свертки, которая преобразует c в c' [3].

Слияние разреженных признаков. Временные признаки будут преобразованы в то же пространство, что и пространственные признаки, с использованием слоя отображения. Предположим, что карты пространственных признаков X_s имеют c_1 каналов, а карты временных признаков X_t — c_2 каналов. Пусть $c = 2 \times c_1$, тогда объединенные карты признаков X_{concat} могут быть представлены в виде:

$$X_{concat} = [W_{1 \times 1 \times c_2 \times c_1} * X_t, X_s]_c, \quad (2.18)$$

где W — сверточный фильтр временно-пространственного отображения, c_2 — исходное количество каналов, c_1 — выходное количество каналов, $*$ обозначает операцию свертки, а $,$ — перекрестная конкатенация. Затем используется разреженная матрица $SM \in \mathbb{R}^{c \times c/2}$, предназначенная для выбора карт полезных признаков и адаптивного сжатия каналов признаков. Объединенные карты признаков X_{fused} могут быть посчитаны по формуле:

$$X_{fused} = X_{concat} \times SM, \quad (2.19)$$

где \times означает матричное умножение [3].

Увеличение разрешения во время реконструкции. В компоненте реконструкции создаются деконволюционные слои для увеличения разрешения карт признаков до целевого высокого разрешения. В данном методе используются слои обратной свертки в качестве слоя увеличения разрешения в компоненте реконструкции для того, чтобы преобразованные объекты подвергались увеличению разрешения в конце сети. В отличие от субпиксельной свертки, уровень деконволюции адаптивно допускает в качестве входных данных произвольные номера каналов, а не фиксированные числа. Для увеличения разрешения карт признаков используются два стека слоев деконволюции с небольшими ядрами размером 3×3 и 256-ю картами признаков [3].

3 Классификация методов увеличения разрешения видеопотока

3.1 Критерии оценки методов увеличения разрешения видеопотока

Для сравнения и оценки производительности методов суперразрешения используются такие показатели, как пиковое отношение сигнала к шуму и индекс структурного сходства. Среднеквадратическая ошибка, которая представляет собой среднюю ошибку между исходным и улучшенным изображением, используется для расчета пикового отношения сигнала к шуму. Формула для нахождения среднеквадратической ошибки [2]:

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [X(i, j) - \hat{X}(i, j)]. \quad (3.1)$$

Формула для нахождения пикового отношения сигнала к шуму [2]:

$$PSNR = 20 \log_{10} \left(\frac{1}{\sqrt{MSE}} \right). \quad (3.2)$$

Сходство между исходным изображением и изображением, полученным в результате использования метода суперразрешения, вычисляется по индексу структурного сходства. Параметрами, используемыми для расчета, являются яркость, контрастность и структурные изменения между двумя изображениями. Индекс структурного сходства определяется уравнением [2]:

$$SSIM(X, \hat{X}) = \frac{(2\mu_x\mu_{\hat{x}} + C_1)(2\sigma_{x\hat{x}} + C_2)}{(\mu_x^2 + \mu_{\hat{x}}^2 + C_1)(\sigma_x^2 + \sigma_{\hat{x}}^2 + C_2)}, \quad (3.3)$$

где μ_x и $\mu_{\hat{x}}$ — среднее x и \hat{x} соответственно, σ_x^2 и $\sigma_{\hat{x}}^2$ — дисперсия x и \hat{x} соответственно, $\sigma_{x\hat{x}}$ — ковариация x и \hat{x} , а $C_1 = (k_1L)^2$ и $C_2 = (k_2L)^2$ — две переменные (где L — динамический диапазон пикселей, а $k_1 = 0.01$ и $k_2 = 0.03$ — константы).

3.2 Сравнение методов увеличения разрешения видеопотока

В таблице 3.1 приведены данные, по которым можно сравнить рассмотренные методы увеличения разрешения видеопотока. Четыре критерия: подход, пиковое отношение сигнала к шуму, индекс структурного сходства и максимальное значение коэффициента масштабирования $\max(r)$.

Таблица 3.1 – Сравнение методов увеличения разрешения видеопотока

Метод	Подход	PSNR	SSIM	$\max(r)$
Бикубическая интерполяция	интерполяционный	26.4037	0.7227	2
Нелокальная регуляризация и надежная оценка движения	регуляризационный	33.18	—	2
Динамические фильтры увеличения разрешения без явной компенсации движения	нейронные сети	31.65	0.897	4
Остаточная обратимая пространственно-временная нейронная сеть	нейронные сети	28.90	0.8898	4

Исходя из полученных данных нельзя точно сказать, какой из рассмотренных методов увеличения разрешения видеопотока является наилучшим. Самый высокий показатель пикового отношения сигнала к шуму у метода нелокальной регуляризации и надежной оценки движения. Но недостаток данного метода — максимальный коэффициент масштабирования, равный 2-м. У метода динамических фильтров увеличения разрешения без явной компенсации движения наилучший показатель индекса структурного сходства. Наихудшие показатели у бикубической интерполяции.

ЗАКЛЮЧЕНИЕ

В ходе выполнения научно-исследовательской работы была достигнута поставленная цель, а также решены все задачи:

- 1) исследована предметная область увеличения разрешения видеопотока;
- 2) проанализированы известные методы увеличения разрешения видеопотока;
- 3) сформулированы критерии для сравнения этих методов;
- 4) проведено сравнение методов увеличения разрешения видеопотока по сформулированным критериям.

В результате проведения сравнения по 4-м критериям было выявлено, что самый высокий показатель пикового отношения сигнала к шуму у метода нелокальной регуляризации и надежной оценки движения, в то время как у метода динамических фильтров увеличения разрешения без явной компенсации движения наилучший показатель индекса структурного сходства. Притом второй метод позволяет увеличивать разрешение видеопотока в 4 раза. Наихудшим из рассмотренных по всем показателям оказался метод бикубической интерполяции.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. *Park S. C., Park M. K., Kang M. G.* Super-resolution image reconstruction: a technical overview // IEEE Signal Processing. — 2003. — С. 21—36.
2. *Mrunmayee D. V., Sachin R. D.* Video Super Resolution: A Review // Department of Electronics Engineering, Walchand College of Engineering, Sangli, Maharashtra, India. — 2021. — С. 6.
3. Residual Invertible Spatio-Temporal Network for Video Super-Resolution / Z. Xiaobin [и др.] // Proceedings of the AAAI Conference on Artificial Intelligence. — 2019. — Т. 33. — С. 5981—5988.
4. Image super-resolution: The techniques, applications, and future / L. Yue [и др.] // IEEE Signal Processing. — 2016. — Т. 128. — С. 389—408.
5. Deep Video Super-Resolution Network Using Dynamic Upsampling Filters Without Explicit Motion Compensation / J. Younghyun [и др.] // IEEE/CVF Conference on Computer Vision and Pattern Recognition. — 2018. — С. 3224—3232.
6. A performance comparison among different super-resolution techniques / D. Thapa [и др.] // Computers and Electrical Engineering. — 2016. — Т. 54. — С. 313—329.
7. *Bavenstrand E.* Real-Time Video Super-Resolution // KTH Royal Institute of Technology. — 2021. — С. 92.
8. *Jian L., HongRan Z., Yi S.* Video super resolution based on non-local regularization and reliable motion estimation // Signal Processing: Image Communication. — 2014. — С. 16.

ПРИЛОЖЕНИЕ А

Презентация