

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение  
высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г.ЧЕРНЫШЕВСКОГО»**

Кафедра математической теории упругости и биомеханики

Генерация дополнительных кадров в анимации с помощью

генеративных нейронных сетей

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента 4 курса 442 группы

направления 09.03.03 — Прикладная информатика

механико-математического факультета

Павлутина Антона Олеговича

Научный руководитель  
доцент, к.ф.-м.н

А. М. Донник

Зав. кафедрой  
зав. кафедрой, д.ф.-м.н., профессор

Л. Ю. Коссович

Саратов 2026

**Введение.** В условиях стремительного развития цифровых медиаплатформ и технологий отображения визуального контента сформировался запрос на новый стандарт качества анимации. Современный зритель ожидает высокой плавности и визуальной целостности изображения, что требует перехода от традиционных стандартов кадровой частоты к форматам 60 FPS и выше для обеспечения комфортного восприятия на современных экранах.

Несмотря на стремительный технологический прогресс, процесс создания классической 2D-анимации по-прежнему в значительной степени опирается на традиционную и крайне трудоемкую ручную отрисовку каждой промежуточной фазы движения персонажей, известную как фазовка. Существующие универсальные алгоритмы автоматизации зачастую не справляются со спецификой рисованной графики — отсутствием пространственных текстур и наличием жестких изолированных линий, что создает выраженный дефицит эффективных специализированных инструментов, доступных для широкого круга анимационных студий.

Актуальность темы определяется необходимостью разрешения технологического противоречия между растущими требованиями к фреймрейту анимации и высокими экономическими затратами на ручное производство. Автоматизация создания промежуточных кадров с помощью дообученных нейросетевых моделей позволяет значительно сократить сроки и бюджеты проектов, обеспечивая при этом прецизионную точность восстановления динамики движения и сохранение целостности контуров графических объектов.

**Цель исследования** — проектирование, программная реализация и экспериментальная проверка специализированного программного комплекса интеллектуальной интерполяции, обеспечивающего автоматическую генерацию промежуточных кадров в 2D-анимации с сохранением стилистической и геометрической целостности контуров.

Для достижения поставленной цели сформулированы следующие **задачи**:

1. Осуществить системный анализ существующих методов интерполяции 2D-анимации и формализовать специфика возникновения пространственно-временных артефактов в рисованном контенте.

2. Провести аналитический обзор генеративных нейросетевых архитектур, обосновать выбор базовой модели реального времени RIFE и определить стек программных средств разработки.
3. Разработать математическое описание модифицированной нейросетевой модели и комбинированной функции потерь, адаптированной для контроля четкости контуров графических объектов.
4. Спроектировать алгоритмы взаимодействия пользователя с системой и сформировать методику автоматизированного сбора, разметки и преобразования специализированного набора данных.
5. Программно реализовать кроссплатформенный комплекс на базе библиотеки PyTorch и разработать графический интерфейс пользователя (UI/UX) для управления параметрами интерполяции.
6. Провести дообучение модели на целевом датасете, выполнить экспериментальное тестирование производительности ПО и провести сравнительный анализ качества сгенерированной анимации.

**Структура и объём работы.** Бакалаврская работа включает в себя введение, три раздела основной части, заключение, список использованных источников из 20 наименований и 1 приложение, содержащее исходный код программного комплекса. Объем основного текста составляет 73 страницы. Общий объем с учетом приложений составляет 95 страниц.

Раздел 1. Теоретические основы и сравнительный анализ методов интерполяции кадров в анимации.

Раздел 2. Проектирование архитектуры нейросетевой системы и подготовка данных для обучения.

Раздел 3. Программная реализация системы генерации кадров и анализ эффективности результатов.

Работа посвящена разработке программного комплекса на базе модифицированной архитектуры RIFE, обеспечивающего создание промежуточных фаз движения в 2D-анимации для повышения её кадровой частоты.

**В рамках первого раздела** проводится детальный ретроспективный анализ развития технологий синтеза промежуточных кадров. Рассматривается классификация методов интерполяции, начиная от традиционных алгоритмов компьютерной графики, базирующихся на геометрических и фото-

метрических преобразованиях. Подробно описываются механизмы линейной интерполяции, технологии морфинга и классические алгоритмы оценки оптического потока (optical flow). Подробно рассмотрены процессы контролируемой пространственной деформации растровой сетки (warping) и последующей кросс-растворимой интерполяции цвета. Установлено, что данные подходы, опирающиеся на поиск соответствий между пикселями или блоками в соседних кадрах, демонстрируют концептуальную ограниченность в условиях анимационной индустрии. Математическая зависимость от точности оценки векторов смещения делает их малоприспособными для сцен со сложными окклюзиями и резкими деформациями объектов, что приводит к возникновению визуальных артефактов и нарушению целостности изображения.

В работе обоснован переход от детерминированных алгоритмов к глубокому обучению в обработке визуального контента. На базе концептуальной модели FlowNet рассмотрены сверточные сети (CNN), латентные диффузионные модели (LDM) и генеративно-сопоставительные сети (GAN), более стойкие к шумам и сложным трансформациям в авторской 2D-анимации.

Особое внимание в первом разделе уделено комплексному анализу современных архитектур интерполяции по материалам ведущих мировых бенчмарков (Video Frame Interpolation Benchmarks). В качестве технологического ядра проектируемого комплекса обоснованно выбрана модель RIFE (Real-Time Intermediate Flow Estimation), открытый исходный код которой депонирован в официальной репозитории проекта. Приоритетность данной Flow-based архитектуры обусловлена её способностью к высокоточной аппроксимации межкадрового движения в режиме реального времени на базе подсети IFNet, реализующей сквозное (end-to-end) обучение.

Для прикладного программного обеспечения, ориентированного на профессиональное использование в студиях, скорость обработки данных в сочетании с качеством синтеза является определяющим критерием эффективности. В то же время в ходе анализа выявлено критическое противоречие: ориентация базовой модели RIFE на фотореалистичные видеопотоки вступает в конфликт со спецификой дискретной рисованной графики. Это теоретически обосновывает необходимость глубокой адаптации нейросетевой структуры, включая пересмотр функций потерь и механизмов обработки контуров,

для предотвращения деградации тонких линий и сохранения уникальной художественной стилизации анимационных произведений.

Ключевой исследовательский фокус работы направлен на решение фундаментальной проблемы «эффекта сглаживания», который возникает при применении универсальных моделей к высокостилизованной графике. Проведенное исследование показывает, что 2D-анимация обладает специфической визуальной доминантой, где ключевую роль играет чистота контура и локальная однородность цветовых зон. В связи с этим предложенная в работе модификация стратегии обучения архитектуры RIFE ориентирована на значительное повышение чувствительности модели к резким градиентам и уникальной морфологии рисованных объектов. Обосновывается необходимость оптимизации многокомпонентных целевых функций потерь (loss functions) для смещения вектора градиентов от простого усреднения пиксельных значений к сохранению структурной и контурной целостности кадра. Внедрение механизмов контроля четкости границ позволяет радикально минимизировать артефакты размытия на стыках контрастных цветовых сегментов.

Дополнительно в теоретической базе исследования разрабатывается концепция алгоритма рекурсивной интерполяции как средства достижения сверхвысокой кадровой частоты. В рамках данного подхода предлагается иерархическая стратегия генерации, при которой каждый синтезированный кадр становится полноправной опорной точкой для последующих циклов аппроксимации. Это позволяет реализовать гибкое, кратное увеличение FPS (например, с исходных 12 или 25 до 100 кадров в секунду) без потери плавности динамических переходов. В работе описывается математическая логика процесса рекурсии, который регулируется пороговыми значениями карт достоверности оптического потока (Confidence Map). Данный механизм гарантирует визуальную стабильность и предотвращает кумулятивное накопление геометрических искажений, что является критическим преимуществом перед традиционными одношаговыми методами генерации.

**Второй раздел** посвящен практическим аспектам проектирования системы и формированию надежного фундамента для обучения нейросетевой модели. Центральным этапом реализации предложенного подхода выступает создание специализированной среды обучения, которая должна быть макси-

мально адаптирована под уникальные визуальные и структурные паттерны 2D-графики. В отличие от стандартных наборов данных, ориентированных на фотореализм, процесс формирования авторского датасета (dataset) в данной работе основывается на прецизионной сегментации высококачественных последовательностей анимационного произведения «Бременские музыканты».

Методология подготовки данных предполагает разделение видеопотока на структурированные триплеты кадров. В этой связке два крайних изображения служат входными параметрами (опорными кадрами), а центральное изображение выполняет роль эталона (ground truth), с которым сверяется результат генерации. Использование триплетной структуры вместо классических пар позволяет нейросети на этапе обучения более точно вычислять нелинейные векторы смещения и сопоставлять расчетную траекторию движения с реальным промежуточным состоянием объекта. Это критически важно для анимации, где движение часто бывает дискретным или преувеличенным.

Особое внимание в работе уделено этапу комплексной предварительной обработки (preprocessing) и синхронной стохастической аугментации данных внутри триплетов для расширения репрезентативности выборки без физического расширения дискового пространства. С целью повышения инвариантности модели к вариативности палитр и толщины линий разработан программный алгоритм динамических трансформаций на лету (spatial consistency constraint), реализованный в классе AnimationTripletDataset. Пайплайн включает синхронный аффинный поворот на случайный угол  $\theta \in [-15^\circ, +15^\circ]$ , горизонтальное зеркальное отражение, фотометрический сдвиг каналов в пространстве  $HSV$  (color jittering), наложение импульсного шума «Соль-перец» (Salt & Pepper Noise) и стохастическую временную инверсию кадров триплета (Temporal Inversion). Это заставляет сверточные фильтры абстрагироваться от колористики и фокусироваться на инвариантной морфологии контуров.

Процесс обучения системы базируется на стратегии полнопараметрической тонкой настройки (Full Fine-tuning) весовых коэффициентов базовой архитектуры RIFE на авторском очищенном датасете из 9253 триплетов мультфильма «Бременские музыканты». В контур оптимизации интегрирован алгоритм AdamW ( $\lambda = 1 \cdot 10^{-2}$ ) под управлением косинусного планировщика скорости обучения со стратегией линейного разогрева (Linear Warmup) в

диапазоне  $\eta \in [1 \cdot 10^{-4}, 1 \cdot 10^{-6}]$ . Инженерная проблема нехватки видеопамяти при обработке высокого разрешения разрешена посредством интеграции схемы аккумулярования градиентов кратностью  $N_{\text{acc}} = 4$ , разделяющей эффективный батч  $B_{\text{eff}} = 16$  на физические подпакеты  $B_{\text{phys}} = 4$ . Контроль переобучения осуществлялся алгоритмом ранней остановки (Early Stopping) с окном терпения  $P = 5$  эпох.

Важной исследовательской составляющей Второго раздела является проектирование и математическая калибровка комбинированного функционала оптимизации  $L_{\text{total}}$ , реализованного в модуле `loss.py`. Для минимизации дефектов `blurring` и `flickering` разработана взвешенная многокомпонентная функция потерь, интегрирующая в себе три дифференциальных слоя ошибки в пропорции  $\alpha = 1.0, \beta = 0.5, \gamma = 0.005$ :

1. Сглаженный фотометрический лосс Чарбонье (Charbonnier Loss) — для обеспечения базовой попиксельной точности и восстановления общей цветовой массы растра.
2. Контурный лосс (Edge Loss) — специализированный функционал на базе дискретных градиентов оператора Собеля, ориентированный на жесткое удержание четкости и геометрии изолированных рисованных линий персонажей.
3. Перцептивный лосс (Perceptual Loss) — для контроля семантической целостности графических макроформ в латентных признаковых пространствах (`relu2_2`, `relu3_3`) предобученного детектора VGG-16.

Подобный многоуровневый и комплексный подход к настройке функции потерь позволяет успешно справиться с проблемой визуального «замыливания» и потери четкости сцены. При этом гарантируется полное сохранение изначальной стилистической уникальности авторского почерка, а также предотвращается деградация, сглаживание и исчезновение мелких деталей в процессе выполнения иерархического рекурсивного инференса.

**Третий раздел** посвящен практическому воплощению разработанных алгоритмов и всестороннему анализу полученных результатов. Программная реализация системы осуществлена на языке программирования Python с активным использованием фреймворка глубокого обучения PyTorch, что поз-

волило обеспечить высокую скорость вычислений и гибкость в управлении тензорными операциями на уровне GPU.

Архитектура разработанного приложения базируется на модульном принципе, что является критически важным для обеспечения масштабируемости, сопровождаемости ИС и возможности независимой верификации каждого этапа обработки графических данных. В рамках системы выделены три ключевых функциональных блока:

1. Модуль предварительной обработки и инженерии данных (Preprocessing Engine, скрипты `tools_prepare_data.py` и `clean_dataset.py`): отвечает за интеллектуальную сегментацию, автоматическое приведение последовательности кадров к разрешению  $256 \times 256$  пикселей и дисперсионную блочную фильтрацию раstra.
2. Вычислительное ядро интерполяции (модули `model/RIFE.py` и `model/loss.py`): базируется на модифицированной архитектуре RIFE с интеграцией кастомного лосса  $L_{total}$  и схемы аккумуляирования градиентов  $N_{acc} = 4$ , реализуя логику иерархического инференса.
3. Модуль постобработки и сборки видеопотока (Post-processing & Assembly, интегрированный в `app.py`): обеспечивает удаление полей технического зеркального паддинга, обратное масштабирование и финальную сборку синтезированных кадров в видеопоток средствами OpenCV без потери качества.

Для обеспечения практической применимости и снижения порога вхождения для конечных пользователей — художников-аниматоров — спроектирован и реализован кроссплатформенный графический интерфейс пользователя (UI/UX) на базе библиотеки CustomTkinter (модуль `app.py`).

Разработанная панель управления позволяет автоматизировать выбор параметров временной дискретизации контента (режимы интерполяции x2 и x4). Инженерная задача предотвращения кумулятивных ошибок рекурсии при глубоком инференсе (режим x4) разрешена посредством программной интеграции алгоритма порогового контроля достоверности. Система в реальном времени анализирует карты неопределенности оптического потока (Confidence Map) и при падении метрики уверенности ниже установленного порога (0.55) осуществляет безопасное дублирование стабильной фазы,

что полностью исключает возникновение визуальных дефектов на сложных участках анимации. Это превращает комплекс в полноценный инструмент, готовый к интеграции в производственные циклы студий.

Вторая часть раздела посвящена экспериментальной оценке эффективности разработанного метода. Тонкая настройка модифицированной архитектуры RIFE осуществлялась на авторском специализированном датасете из 9253 триплетов, сформированном из кадров анимационного произведения «Бременские музыканты» и очищенном от статики и панорамных сдвигов фона с помощью разработанного алгоритма блочного анализа дисперсии.

Для объективной количественной оценки качества интерполяции рисованного контента на тестовой выборке использовался разработанный программный модуль (`evaluate.py`) на основе трех метрик компьютерного зрения:

- PSNR (Peak Signal-to-Noise Ratio) — для оценки фотометрической точности восстановления яркости и цветовой массы раstra;
- SSIM (Structural Similarity Index) — для анализа сохранности геометрических макроформ изображения;
- Кастомная метрика Edge-SSIM — для перцептивного контроля четкости и геометрической целостности изолированных рисованных контуров, вычисленных с помощью дискретного оператора Собеля.

Сводные итоговые результаты экспериментальной оценки эффективности разработанного метода интерполяции представлены в таблице 1.

Таблица 1 — Результаты экспериментального тестирования разработанного комплекса

№	Наименование критерия оценки (метрики)	Среднее значение
1	Средний PSNR (Общая точность восстановления цвета)	20.09 dB
2	Средний SSIM (Структурное сходство макроформ)	0.7730
3	Средний Edge-SSIM (Точность удержания контуров Собеля)	0.6687

Экспериментальное тестирование показало, что разработанное решение статистически превосходит базовую модель RIFE при обработке стилизованной графики. Зафиксировано устойчивое повышение индексов SSIM и Edge-

SSIM, что выражается в полном подавлении характерных для универсальных VFI-моделей артефактов blurring (размытие границ) и flickering (мерцание линий). Внедрение комплекса позволяет многократно сократить трудозатраты на производство промежуточных фаз движения без потери исходной стилистической идентичности анимации.

**Заключение.** В ходе выполнения выпускной квалификационной работы была успешно решена актуальная научно-практическая задача по разработке и адаптации системы автоматизированной генерации промежуточных кадров в 2D-анимации на основе методов глубокого обучения. На основании проведенных исследований, программной реализации и экспериментальной проверки можно сформулировать следующие расширенные выводы.

В процессе работы был проведен всесторонний компаративный анализ существующих алгоритмов интерполяции. Установлено, что ключевым барьером для применения стандартных нейросетевых моделей в анимационном производстве является принципиальное различие в морфологии фотореалистичного видео и рисованного кадра. Отсутствие естественных пространственных текстур и тотальное доминирование четких контурных линий в 2D-графике приводит к лавинообразной деградации качества при использовании стандартных сверточных слоев. Это обосновало необходимость разработки специализированного подхода, ориентированного на прецизионное сохранение структурной целостности изолированных рисованных линий.

Научный вклад работы заключается в методологическом обосновании процесса адаптации универсальных архитектур к дискретной графической среде. Автором предложен и реализован комплекс архитектурных корректировок, включающий оптимизацию функций потерь в модуле loss.py. Введение комбинированного взвешенного функционала потерь  $L_{total}$  (сочетающего лосс Чарбонье, Perceptual Loss и Edge Loss) позволило сместить фокус обучения нейросети на сохранение высокочастотных деталей кадра. Отдельным элементом новизны выступает алгоритмизация порогового контроля карт достоверности (Confidence Map) со значением отсечки 0.55 в модуле arr.py. Данное решение позволяет гибко и безопасно управлять процессом иерархического рекурсивного инференса, обеспечивая высокую визуальную ста-

бильность видеопотока даже при экстремальных межкадровых смещениях объектов персонажной анимации.

Разработанный программный комплекс на базе фреймворка PyTorch представляет собой завершённый автономный инструмент, готовый к интеграции в производственные циклы. Создание кроссплатформенного графического интерфейса на базе CustomTkinter позволило полностью абстрагировать сложную математическую модель от конечного пользователя, сделав технологию доступной для художников-аниматоров. Практическая ценность подтверждается возможностью радикального сокращения трудозатрат на этапе «фазовки» анимации. Система позволяет автоматизировать наиболее монотонные процессы, высвобождая творческие ресурсы и снижая финансовые издержки на производство высокочастотного контента.

Экспериментальное тестирование системы на тестовом анимационном материале показало устойчивое преимущество адаптированной модели над базовой архитектурой RIFE. Объективные метрики качества ( $PSNR = 20.09$  dB,  $SSIM = 0.7730$  и разработанный индекс  $Edge-SSIM = 0.6687$ ) продемонстрировали стабильную сходимость графа вычислений, а экспертный перцептивный анализ видеопотока подтвердил полное отсутствие характерных для универсальных VFI-моделей артефактов — размытия границ (blurring) и межкадрового мерцания линий (flickering).

Проведённое исследование открывает перспективы для дальнейшего изучения синергии классических графических техник и генеративных нейронных сетей. В качестве векторов дальнейшего развития рассматривается внедрение механизмов семантической сегментации для автоматического разделения планов анимации и алгоритмов постобработки светотеневых переходов.

Подводя итог, можно утверждать, что поставленная цель работы достигнута, а сформулированные задачи выполнены в полном объёме. Разработанное решение является конкурентоспособным инструментом, имеющим как теоретическую ценность для области компьютерного зрения, так и прикладное значение для современной анимационной индустрии.