

МИНОБРНАУКИ РОССИИ
САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
«ЛЭТИ» ИМ. В.И. УЛЬЯНОВА (ЛЕНИНА)
Кафедра МОЕВМ

ОТЧЕТ
по научно-исследовательской работе
Тема: «Супер-разрешение изображений и видео»

Студент гр. 7381

_____ Чигалейчик А.С.

Руководитель

_____ Черниченко Д.А.

Санкт-Петербург

2022

**ЗАДАНИЕ
НА НАУЧНО-ИССЛЕДОВАТЕЛЬСКУЮ РАБОТУ**

Студент Чигалейчик А.С.

Группа 7381

Тема НИР: Суперразрешение изображений и видео

Задание на НИР:

Провести анализ существующих метрик оценок качества суперразрешения,

Провести обзор различных моделей глубокого обучения с точки зрения
скорость-эффективность

Выбрать конкретные модели для использования в дальнейшей разработке

Сроки выполнения НИР: 01.09.2022 – 20.12.2022

Дата сдачи отчета: 01.10.2022

Дата защиты отчета: 27.12.2022

Студент _____ Чигалейчик А.С.

Руководитель _____ Черниченко Д.А.

АННОТАЦИЯ

Данная работа более глубоко рассматривает методы супер-разрешения основанные на глубоких свёрточных сетях. Большое внимание уделяется метрикам супер-разрешения, как для изображений, так и для повышения качества видео. Рассматриваются особенности архитектур генеративно-сопоставительных сетей, а также использование субпиксельных сверток. В конце работы приводятся практические модели и их реализации и формируется общий план разработки.

SUMMARY

This paper takes a deeper look at super-resolution methods based on deep convolutional networks. Great attention is paid to super-resolution metrics, both for images and for video enhancement. The features of the architectures of generative adversarial networks, as well as the use of sub-pixel convolutions, are considered. At the end of the work, practical models and their implementation are given, and a general development plan is formed.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	5
2. ПОСТАНОВКА ЗАДАЧИ	7
3. РЕЗУЛЬТАТЫ РАБОТЫ В ВЕСЕННЕМ СЕМЕСТРЕ	8
3.1. КОМПРОМИСС СКОРОСТИ И ЭФФЕКТИВНОСТИ	8
3.2. МЕТРИКИ СУПЕР-РАЗРЕШЕНИЯ	9
3.3. МОДЕЛИ СУПЕР-РАЗРЕШЕНИЯ	15
4. ОПИСАНИЕ ПРЕДПОЛОГАЕМОГО МЕТОДА РЕШЕНИЯ	23
5. ПЛАН НА ВЕСЕННИЙ СЕМЕСТР	24
ЗАКЛЮЧЕНИЕ	25
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	26

ВВЕДЕНИЕ

Супер-разрешение — это результат восстановления изображения с высоким разрешением (HR) из данного изображения с низким разрешением (LR). Изображение может иметь «более низкое разрешение» из-за меньшего пространственного разрешения (то есть размера) или из-за ухудшения качества (такого как размытие).

Задача повышения разрешения изображений является одной из важнейших задач цифровой обработки изображений. Несмотря на рост разрешения сенсоров камер, эта задача остаётся актуальной, например, для изображений, снятых на старые камеры низкого разрешения. Особенно важна задача повышения разрешения в видеонаблюдении. При увеличении одного изображения может использоваться априорная информация об изображении, например, самоподобие изображения при разных разрешениях [1]. Такой подход делает изображение визуально более качественным, однако результат может отличаться от реального изображения. Более качественного результата можно добиться, используя несколько изображений объекта с субпиксельными сдвигами для построения изображения высокого разрешения. Пиксели камеры, регистрирующей изображение, имеют ненулевой размер, поэтому наблюдаемое значение пикселя является усреднением по некоторой окрестности точки. Если движение объекта

и функция усреднения известны, то можно использовать информацию со всех кадров для построения одного изображения высокого разрешения. К недостаткам методов суперразрешения можно отнести необходимость точного вычисления векторов движения между кадрами. В работе предлагается изучить различные методы суперразрешения и реализовать программу применяющая рассмотренные методы к видеофайлам и изображениям с целью получить более качественную картинку.

Целью данной работы является разработка программы, реализующей методы получения картинки высокого качества из картинки низкого качества

Проблема: восстановление и/или повышение качества картинки низкого разрешения

Объект исследования: алгоритмы по работе с изображениями

Предмет исследования: алгоритмы повышения качества изображений

Практическая ценность работы: разработанная система позволит улучшать получаемые изображение, что может использоваться как для личных целей, так и в промышленных: улучшение качества изображения может сильно повысить качество визуального анализ, такой как распознавание объектов, которое активно применяется во многих промышленных сферах.

1. ПОСТАНОВКА ЗАДАЧИ

Глобально, цель данной НИР – провести анализ существующих способов суперразрешения и выбрать подходящие для будущей программы. В весеннем семестре были разобраны простейшие алгоритмы глубокого обучения, в данном семестре рассматривались более сложные алгоритмы.

- 1) Обзор метрик качества для суперразрешения изображений и видео
- 2) Более подробный обзор и выбор моделей суперразрешения, основанные на глубоком обучении для дальнейшего использования
- 3) Переход от изображений к видео
- 4) Разработка архитектуры будущей программы

2. РЕЗУЛЬТАТЫ РАБОТЫ В ВЕСЕННЕМ СЕМЕСТРЕ

В данной главе приводится обзор архитектур супер-разрешения с целью выбора конкретной модели для работы с приложением.

2.1. Компромисс скорости и эффективности

Компромисс между скоростью и эффективностью является важным фактором в задаче супер-разрешения, которая включает увеличение разрешения изображения или видео в два или более раза. Как правило, более эффективные в вычислительном отношении алгоритмы имеют тенденцию быть быстрее, но не обязательно могут достигать того же уровня производительности, что и алгоритмы с более высокими вычислительными затратами. С другой стороны, алгоритмы с более высокими вычислительными затратами могут обеспечить лучшую производительность, но могут быть медленнее с точки зрения скорости.

Одной из ключевых проблем в задаче сверхвысокого разрешения является поиск баланса между вычислительной эффективностью и производительностью. Это особенно важно в приложениях реального времени, где скорость алгоритма является критическим фактором. В этих случаях может потребоваться снижение производительности, чтобы сократить время обработки.

Так, например, существует целое семейство генеративных (GAN) нейронных сетей для восстановления качества, такие как *ESRGAN* [2] и *TacoGAN* [1]. Данные алгоритмы демонстрируют отличные результаты с точки зрения эффективности, но ввиду сложности сетей требуют гораздо больше вычислительных ресурсов для выполнения задачи.

Однако, использования глубокого обучения не лишает возможности создания быстрых моделей с небольшим количеством вычислений. Так, например, используя субпиксельные свёртки, авторам работы модели *ESPCN* [3]

В целом, компромисс между скоростью и эффективностью является важным фактором в задаче сверхвысокого разрешения, и пользователь должен выбрать подходящий баланс, исходя из своих конкретных потребностей и требований.

В рамках данной НИР предлагается выбрать несколько алгоритмов, чтобы предоставить пользователю выбор между скоростью и производительностью.

Для этого следует выбрать модель для приложения реального времени, а также, модель для нетребуемого быстрого времени выполнения задач.

2.2. Метрики супер-разрешения

Прежде чем переходить к оценке моделей, прежде всего стоит определиться как именно их оценивать. Для этого существуют специальные метрики качества. Рассмотрим самые популярные из них.

2.2.1 Метрики для работы с изображениями

1) *Среднеквадратическая ошибка (MSE – Mean Squared Error)* — это широко используемый показатель для оценки производительности алгоритмов обработки изображений, в том числе используемых для сверхвысокого разрешения изображения (SR). MSE считается *попиксельно*, т.е. вычисляется средний квадрат разности между пикселями оригинального изображения и восстановленного. Данная метрика отлично себя зарекомендовала как метрика для функции потерь в мире анализа данных ввиду своей идейной простоты, дифференцируемости и лёгкости вычислений, и поэтому используется в большинстве работ без GAN, таких как [3].

В общем случае метрика имеет следующую формулу:

$$MSE = \frac{(\sum_{M,N} [I_{mn} - HR_{mn}]^2)}{M * N}$$

Где I — исходное изображение, HR — восстановленное, M, N — размеры изображения

Однако, она обладает плохой интерпретируемостью, без знания о параметрах изначального изображения: метрика MSE, например, со значением

100 может быть как очень хорошим, так и очень плохим результатом, в зависимости от битности изображения, поэтому следующая метрика прямо вытекает из MSE, но учитывает глубину изображения.

2) *PSNR (Peak Signal to Noise Ration -- Пиковое отношение сигнала к шуму)* – это метрика напрямую вытекающая из MSE, однако в каком-то смысле нормирующая её на максимальный сигнал. PSNR рассчитывается как отношение максимально возможной мощности сигнала (в данном случае исходного изображения высокого разрешения) к мощности шума, присутствующего в сигнале (в данном случае разности между исходным изображением высокого разрешения и восстановленного изображение). Обычно он выражается в децибелах (дБ). Более высокое значение PSNR указывает на лучшее качество, а более низкое значение указывает на худшее качество.

Максимальная мощность сигнала будет битность изображения – так, для n -битных изображений максимальный сигнал будет $2^n - 1$

Общая формула выглядит следующим образом:

$$PSNR = 10 \log_{10} \frac{R^2}{MSE}$$

Где R – максимальная мощность сигнала, в данном случае – максимальное значение пикселя. Очевидно, это зависит от количества бит в изображении – так, для n -битных изображений максимальный сигнал будет $2^n - 1$

PSNR является широко используемой метрикой, поскольку ее относительно просто рассчитать и легко интерпретировать. Она также относительно чувствительна к большим различиям между исходным изображением с высоким разрешением и изображением со сверхвысоким разрешением, что делает его полезным для оценки производительности алгоритмов SR.

Однако PSNR имеет некоторые ограничения в качестве показателя SR. Он основан на квадрате разницы между исходным изображением с высоким разрешением и изображением со сверхвысоким разрешением, что означает, что

он уделяет больше внимания большим различиям между двумя изображениями. Это может сделать его менее чувствительным к небольшим различиям и может неточно отражать общее качество изображения со сверхвысоким разрешением. Кроме того, PSNR — это чисто количественная метрика, и она не учитывает субъективное визуальное качество сверхразрешенного изображения. Это означает, что он может неточно отражать воспринимаемое качество изображения по оценке человека-наблюдателя.

В целом, хотя PSNR является широко используемой и полезной метрикой для оценки производительности алгоритмов SR, важно учитывать ее ограничения и использовать ее в сочетании с другими метриками, которые могут лучше отражать субъективное визуальное качество изображения со сверхвысоким разрешением.

3) *SSIM* -- Индекс структурного сходства (Structural Similarity Index Measure) — это показатель, который измеряет структурное сходство между двумя изображениями. Он был разработан, чтобы быть более точным и надежным показателем качества изображения по сравнению с традиционными показателями, такими как пиковое отношение сигнал/шум (PSNR). В отличие от них SSIM пытается больше учитывать «человеческое восприятие» Идея заключается в том, что пиксели имеют сильную взаимосвязь, особенно когда они близки пространственно. Данные зависимости несут важную информацию о структуре объектов и о сцене в целом. SSIM метрика рассчитана на различные размеры окна. Разница между двумя окнами x и y имеющими одинаковый размер $N \times N$ рассчитывается по следующей формуле:

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

Где x, y — одинаковые окна, μ_x и μ_y — средние значения изображений x и y , σ_x^2 и σ_y^2 — дисперсии изображений x и y , σ_{xy} — ковариация изображений x и y , а c_1 и c_2 — константы, которые используются для стабилизации отношения.

SSIM учитывает яркость, контрастность и структуру изображений и выдает значение от -1 до 1, причем более высокие значения указывают на большее сходство. Значение 1 указывает, что два изображения идентичны, а значение -1 указывает, что два изображения совершенно не похожи.

Одним из ключевых преимуществ SSIM является то, что он более устойчив к шуму и другим искажениям по сравнению с традиционными показателями, такими как PSNR. Он также более чувствителен к небольшим различиям между изображениями и лучше способен фиксировать субъективное визуальное качество изображений, оцениваемое людьми-наблюдателями.

SSIM широко используется в различных приложениях для обработки изображений, включая суперразрешение изображений, сжатие изображений и оценку качества изображений. Обычно считается более надежным и точным.

4) Оценка визуального качества (VQA) — это процесс оценки субъективного визуального качества изображения или видео. VQA обычно проводится посредством исследований пользователей, когда людям-наблюдателям показывают исходное изображение или видео с высоким разрешением и модифицированную (например, супер-разрешением, сжатую) версию и просят оценить визуальное качество модифицированной версии.

Существует множество различных подходов к проведению исследований VQA, и конкретный используемый метод будет зависеть от конкретного приложения и желаемых критериев эффективности. Некоторые распространенные методы проведения исследований VQA включают:

1. Оценка абсолютной категории (ACR): В этом методе наблюдателям показывают исходные и измененные изображения или видео и просят оценить визуальное качество модифицированной версии по шкале (например, от 1 до 5).
2. Непрерывная оценка качества одного стимула (SSCQE): в этом методе наблюдателям показывают серию изображений или видео, включая исходную и модифицированную версии, и просят оценить визуальное

качество модифицированной версии по непрерывной шкале (например, от 0 до 100).

3. Непрерывная оценка качества с двойным стимулом (DSCQE): в этом методе наблюдателям показывают пары изображений или видео, включая исходную и измененную версии, и просят сравнить визуальное качество двух версий по непрерывной шкале (например, от 0 до 100).

Исследования VQA могут занимать много времени и средств, но они являются важным способом оценки субъективного визуального качества изображения или видео. Результаты VQA можно использовать для уточнения и улучшения алгоритмов обработки изображений и видео, а также для сравнения производительности различных алгоритмов.

В целом, VQA является важным компонентом исследований в области обработки изображений и видео, поскольку позволяет исследователям понять, насколько хорошо алгоритм работает с точки зрения субъективного визуального качества обработанных изображений или видео.

2.2.2 Метрики для работы с видео

Во время обучения сетей, сети для работы с видео используют те же метрики. Самыми популярными метриками так же остаётся PSNR и SSIM, они считаются базовыми и используются для оценки большинства моделей.

Не смотря на их распространённость, метрики относятся к покадровым сравнениям, и в них теряется фактор «единства» в видео, чтобы видео казалось натуральным, без резких скачков кадров и прочие визуальные оценки именно видео контента. Поэтому в дополнение к указанным метрикам формируются новые оценки качества. Существует много разных способов оценить «натуральность» видео, далее приведены некоторые из них

- 1) Temporal Smoothness (TS) – Временная плавность — это показатель, который измеряет плавность видеопоследовательности во времени. высококачественные результаты, визуально приятные для зрителей.

Плавность может быть оценена несколькими способами, в зависимости от конкретного приложения и желаемых критериев производительности. Некоторые распространенные методы оценки TS включают:

Частота кадров. Более высокая частота кадров обычно приводит к более плавному видео, поскольку имеется больше кадров для сглаживания любых временных артефактов.

Оценка движения: еще один способ оценить TS — оценить движение между последовательными кадрами видеопоследовательности. Это можно сделать с помощью таких методов, как сопоставление блоков или оптический поток. Видео с хорошей временной плавностью будет иметь последовательное и плавное движение между кадрами.

Субъективная оценка: Временная плавность также может субъективно оцениваться людьми-наблюдателями, которые могут смотреть видео и оценивать плавность движения по шкале (например, от 1 до 5). По сути, является VQA упомянутым ранее.

2) Индекс MOVIE — это модель для прогнозирования качества восприятия движущегося изображения или видео (возможно, сжатого или иным образом искаженного) по сравнению с исходным эталонным видео. Модель состоит из двух частей – пространственный индекс, и временной.

Spatial MOVIE (пространственный) работает путем обработки пространственной и временной информации о движущихся изображениях примерно разделимым образом. Прогноз пространственного (кадрового) качества видео определяется путем вычисления пространственно-временной частотной декомпозиции как эталонного, так и тестового (искаженного) видео с использованием фильтров Габора. После процесса разделительной нормализации, обработанные эталонные и тестовые видео объединяются во

взвешенную разницу для получения прогноза качества пространственного изображения.

Temporal MOVIE -- прогноз временного (изменяющегося во времени или межкадрового) качества движущегося изображения рассчитывается с использованием откликов одной и той же пространственно-временной частотной декомпозиции Габора эталонного и тестового видео, но по-разному. Временной MOVIE взвешивает эти ответы, используя взвешивание возбуждающих и тормозных ответов Габора, чтобы настроить их движение в соответствии с локальным измерением движения видео. Измерения движения также выполняются с использованием банка пространственно-временных фильтров с использованием релевантного для восприятия измерения фазового оптического потока. Эти измерения эталонного и тестового видео затем дифференциально объединяются и нормализуются с разделением для получения прогноза временного качества изображения.

Затем общий индекс MOVIE определяется как простое произведение пространственных и временных индексов MOVIE, объединенных во времени (кадры).

Непосредственные формулы вычисления индекса довольно сложны и подробно расписаны в оригинальной работе [4]

2.3. Модели супер-разрешения

2.3.1 Генеративные модели

GAN -- Generative adversarial networks – Генеративно состязательные сети — это класс моделей глубокого обучения, которые используются для создания новых выборок данных, похожих на обучающий набор данных. Они состоят из двух нейронных сетей, которые обучаются вместе состязательным образом: сеть генератора и сеть дискриминатора.

Сеть генератора отвечает за создание новых выборок данных, а сеть дискриминатора отвечает за определение того, является ли данная выборка данных реальной или поддельной. Две сети обучаются вместе состязательным образом: генератор пытается создать выборки, которые достаточно реалистичны, чтобы обмануть дискриминатор, а дискриминатор пытается правильно определить, является ли выборка реальной или поддельной.

Схема классической сети “GAN” представлена на Рисунке 1.

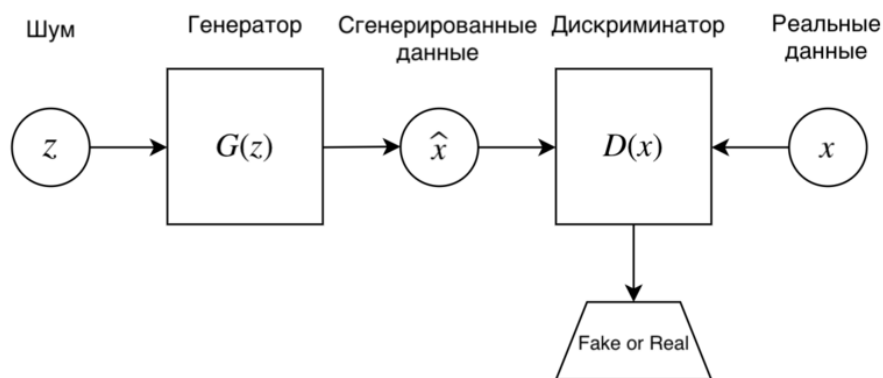


Рисунок 1 -- Классическая схема GAN

Для задачи супер-разрешения данный алгоритм отлично подходит, т.к. генератор восстановит нам утерянную от сжатия информацию.

В случае задачи SR, вместо случайного шума, который обычно подаётся на вход генератору, будет подаваться изображение низкой размерности (LR).

Задача генератора, будет из входа LR превратить в SR (super resolution) – изображение высокого качества. Задача же дискриминатора сводится к задаче определения что именно находится перед ним: восстановленное изображение, или оригинал. Одной из самых популярных сетей использующую GAN является SRGAN (Super Resolution Generative adversarial network) представленной в работе [5]. Схема представлена на Рисунке 2.

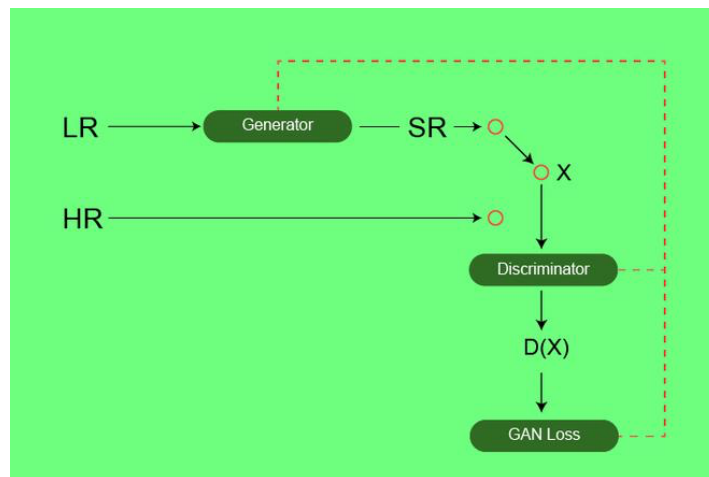


Рисунок 2 -- Схема SRGAN

Одним из основных преимуществ GAN сетей является добавление новой функции потерь помимо стандартных числовых метрик MSE – GAN Loss, т.е. ошибка которую выдаёт дискриминатор. Что приводит к тому, что изображения становятся выглядеть более реалистично. Особенно эффект генеративных моделей становится заметен при больших увеличениях размерности (таких как $\times 4$). Пример такого увеличения представлен на Рисунке 3

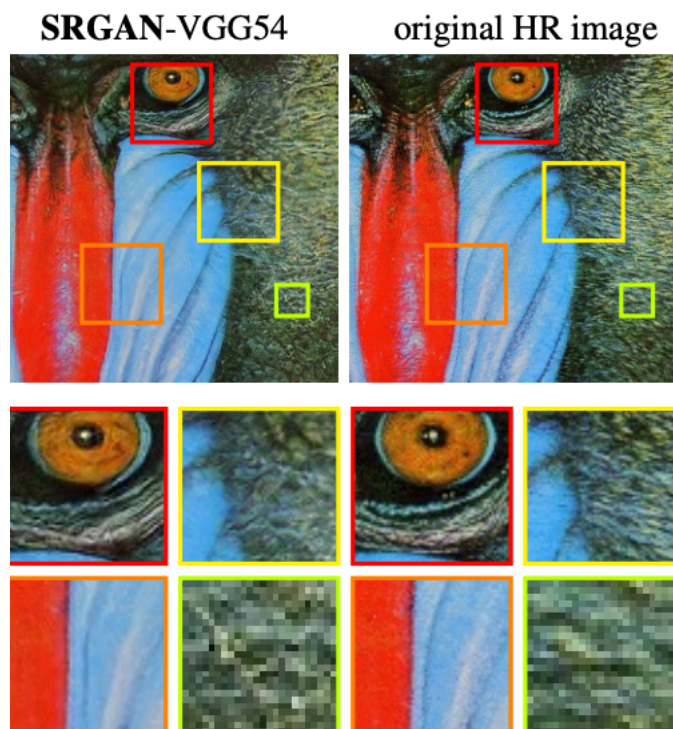


Рисунок 3 -- Большое увеличение размерности

Одним из ключевых преимуществ GAN для SR является их способность генерировать высококачественные изображения со сверхвысоким разрешением, которые визуально неотличимы от исходных изображений с высоким разрешением. Они также могут генерировать изображения, согласованные во времени для видео SR, что важно для сохранения плавности и непрерывности движения в видео.

Однако GAN также имеют некоторые ограничения в контексте SR. Одним из ограничений является то, что их может быть сложно обучить, и для них могут потребоваться большие объемы обучающих данных и вычислительных ресурсов. Они также имеют тенденцию быть чувствительными к качеству обучающих данных и могут давать субоптимальные результаты, если обучающие данные зашумлены или содержат ошибки. Кроме того, GAN не всегда могут давать визуально приятные результаты и могут вносить артефакты или искажения в сгенерированные изображения.

Так, например, если в обучающей выборке было мало лиц, генеративные сети будут очень плохо их «восстанавливать». Однако, можно натренировать такую модель только на определённый тип контента – те же портреты, и использовать модель только для восстановления лиц. Данные подходы с ограничением «универсальности» применения модели уже давно и широко применяются на практике.

Метрики качества сети приведены на следующем Рисунке

Set5	nearest	bicubic	SRCNN	SelfExSR	DRCN	ESPCN	SRResNet	SRGAN	HR
PSNR	26.26	28.43	30.07	30.33	31.52	30.76	32.05	29.40	∞
SSIM	0.7552	0.8211	0.8627	0.872	0.8938	0.8784	0.9019	0.8472	1
MOS	1.28	1.97	2.57	2.65	3.26	2.89	3.37	3.58	4.32

Использовались описанные ранее метрики PSNR, SSIM, а MOS является подвидом визуальной оценки – Mean Opinions Score – средняя оценка людей по шкале 1-5.

Как видно, модель SRGAN при примерном равенстве числовых метрик показывает наилучшие результаты MOS, что в конечной цели и является задачей супер-разрешения.

При большом качестве визуальной оценки, одним важнейшим недостатком GAN является временная неэффективность и требовательность к вычислениям, что делает их в чистом виде непригодными к системам реального времени.

2.3.2 Модель для работы в режим реального времени

Поскольку одной из важных задач суперразрешения является улучшения картинки в «режиме реального времени» -- видео с камер, или использование такого улучшения в условиях слабой связи, где трафик не позволяет передать картинку высокого качества, описанные выше методы могут не подходить. Одной из первых работ в области супер-разрешения видео стала работа ESPCN - Efficient Sub-Pixel CNN [3].

ESPCN – модель глубокого обучения на основе свёрток.

Практически все сверточные модели, упомянутые ранее требуют большого количества вычислений. Данная модель позволяет на порядки повысить скорость вычисления, используя субпиксельный свёрточный слой.

Одной из наиболее важных концепций, предложенных авторами статьи, является свертка субпикселей, также известная как Pixel Shuffle (перетасовка пикселей). Прежде чем понять субпиксельную свертку, необходимо ознакомиться с понятием субпикселя. В цифровом изображении, которое мы видели, пиксели и пиксели связаны друг с другом, в то время как в микроскопическом мире между двумя физическими пикселями есть множество крошечных пикселей. Эти крошечные пиксели называются субпикселями. Пример субпикселей представлен на Рисунке 4.

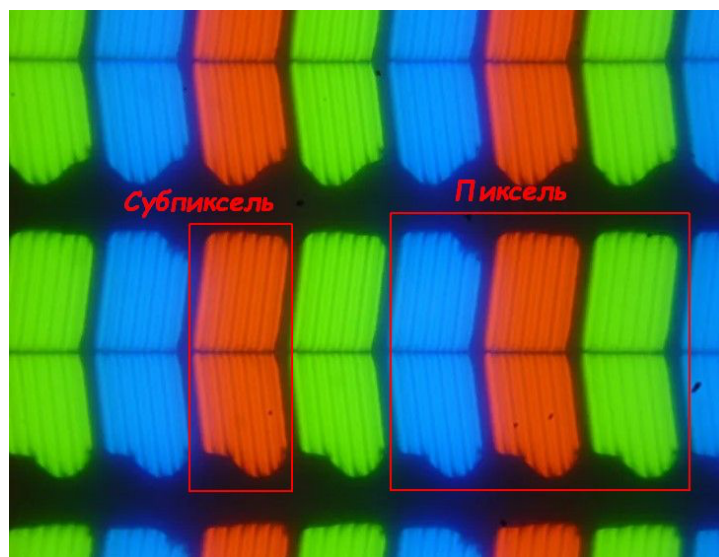


Рисунок 4 -- Субпиксель и пиксель

Метод субпиксельной свертки может быть использован в модели SR для получения изображений высокого разрешения. В общей операции деконволюции мы дополняем изображения нулями, а затем выполняем свертку, что может плохо сказаться на результате. При выполнении субпиксельной свёртки на последнем уровне сети для восстановления изображения LR не требуется операция заполнения нулями. Как показано на рисунке 5, объединение каждого пикселя на многоканальных картах признаков в одну квадратную область размером $r \times r$ на выходном изображении. Таким образом, каждый пиксель на картах объектов эквивалентен субпикселю сгенерированного выходного изображения.

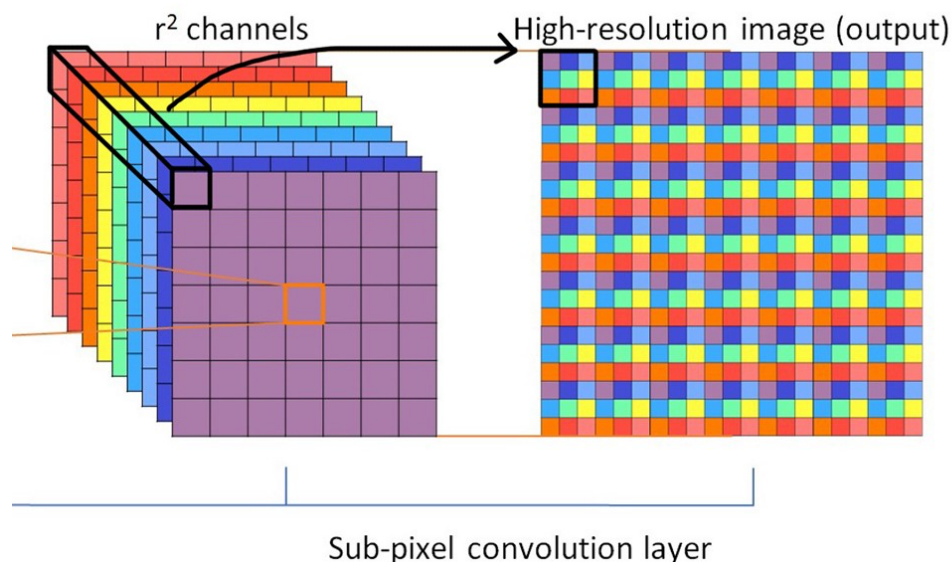


Рисунок 5 -- субпиксельная свёртка

Эффективность данной модели как раз таки и заключается в этом, последнем слое. Так как, по сути, он просто перемещает пиксели с карт признаков на получившееся изображение. Это намного эффективнее других методов повышения размерности, таких как интерполяция

Общая архитектура представлена на рисунке 6

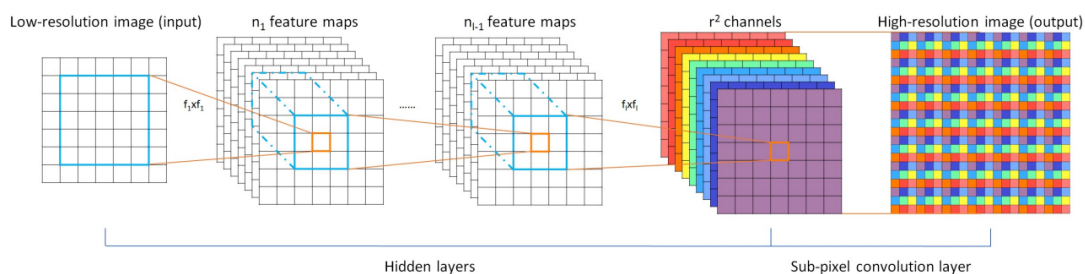


Рисунок 6 -- Архитектура сети ESPCN

Благодаря простой операции «деконволюции» на последнем слое и получается выигрывать в производительности.

На рисунке 7 приведено сравнение PSNR метрики для классической архитектуры, и архитектуры с субпиксельным слоем на различных наборах данных.

Dataset	Scale	SRCNN (91)	ESPCN (91 <i>relu</i>)
Set5	3	32.39	32.39
Set14	3	29.00	28.97
BSD300	3	28.21	28.20
BSD500	3	28.28	28.27
SuperTexture	3	26.37	26.38
Average	3	27.76	27.76

Рисунок 7 -- Сравнение архитектур

Как видно, результаты примерно одинаковые, однако при этом при увеличении в 3 раза, архитектура SRCNN требовала более чем в 10 раз больше времени [5].

3. ОПИСАНИЕ ПРЕДПОЛОГАЕМОГО МЕТОДА РЕШЕНИЯ

В качестве фреймворка для работы с сетями глубокого обучения будет использован используется Pytorch. В предыдущих главах были рассмотрены две разные архитектуры моделей – GAN и ESPCN.

Ввиду того, что обе модели являются моделями глубокого обучения, их требуется обучать. Поскольку обучение (особенно сетей GAN) является долгим и трудоёмким с точки зрения вычислений процессом, обучение самостоятельно не представляется возможным. Поэтому для дальнейшей работы целесообразно используются следующие уже обученные модели

В качестве реализации модели с архитектурой GAN показывающей высокое качество используется работа TesoGan [6] и её обученная модель [7]

Для модели реального времени, используется разобранный ESPCN и её реализация на PyTorch [8].

Для реализации приложения, которое будет использовать пользователь будет написано десктопное приложение. Поскольку от приложения будет требоваться только пользовательский ввод режима работы (реального времени или постобработка), а также доступ к камере, то в качестве фреймворка будет использован простой в использовании фреймворк для создания ПО на C++ -- Qt.

Чтобы упростить разработку и взаимодействие приложения и использования моделей, вместо оригинального фреймворка будет использована PyQt, что позволит ограничиться одним языком для разработки.

В качестве работы с видеопотоком (вебкамера или записанное видео) используется библиотека OpenCV и её реализация на языке Python – cv2

4. ПЛАН НА ВЕСЕННИЙ СЕМЕСТР

В текущем и предыдущем семестре был проведен подробный обзор моделей супер-разрешения и выбраны конкретные реализации, показывающие отличные результаты на соревнованиях по супер-разрешению.

В следующем семестре следует приступить непосредственно к разработке ПО для использования данных моделей.

- 1) Оценить выбранные модели как на изображениях, так и на видео не из тестовых датасетах, а снятых самостоятельно. По полученным данным установить не требуется ли обратиться к другим моделям для конкретных случаев (например, использовать разные модели для изображений и видео)
В идеале – провести собственную визуальную оценку качества с помощью коллег
- 2) Поставить требования к программе – в том числе оценить требуемую производительность (использование моделей требует вычислений)
- 3) Спроектировать пользовательский интерфейс исходя из требуемых функций (работа с видеопотоком реального времени, работа с предзаписанными видео)
- 4) Протестировать работу на разных устройствах

ЗАКЛЮЧЕНИЕ

Данная работа была посвящена углублению в глубокие модели суперразрешения, их подвиды и метрики оценивания. В результате работы были проанализированы популярные архитектуры различных моделей, изучены их особенности, рассмотрены их результаты на тестовом датасете. Результат данной НИР позволяет перейти непосредственно к разработке ПО, обращая к теоретической части в данной работе. Поставленные цели были достигнуты – выбраны модели, метрики, подходящие как и для изображений, так и для видео.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Chu, Mengyu, et al. "Temporally coherent gans for video super-resolution (tecogan)." *arXiv preprint arXiv:1811.09393* 1.2 (2018): 3.
2. Wang, Xintao, et al. "Esrgan: Enhanced super-resolution generative adversarial networks." Proceedings of the European conference on computer vision (ECCV) workshops. 2018.
3. Shi, Wenzhe, et al. "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
4. Seshadrinathan, Kalpana, and Alan Conrad Bovik. "Motion tuned spatio-temporal quality assessment of natural videos." *IEEE transactions on image processing* 19.2 (2009): 335-350.
5. Ledig, Christian, et al. "Photo-realistic single image super-resolution using a generative adversarial network." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
6. Chu, Mengyu, et al. "Learning temporal coherence via self-supervision for GAN-based video generation." *ACM Transactions on Graphics (TOG)* 39.4 (2020): 75-1.
7. <https://github.com/thunil/TecoGAN> -- Официальный репозиторий TecoGan
8. <https://github.com/leftthomas/ESPCN> -- реализация модели ESPCN на PyTorch