

МИНОБРНАУКИ РОССИИ
САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
«ЛЭТИ» ИМ. В.И. УЛЬЯНОВА (ЛЕНИНА)
Кафедра математического обеспечения и применения ЭВМ

ОТЧЕТ
по научно-исследовательской работе
Тема: Разработка приложения
для повышения разрешения изображения

Студентка гр. 3303

Саклакова В.А.

Руководители

Шолохов А.В.

Кринкин К.В.

Санкт-Петербург

2018

ЗАДАНИЕ НА НАУЧНО-ИССЛЕДОВАТЕЛЬСКУЮ РАБОТУ

Студентка Саклакова В.А.

Группа 3303

Тема НИР: Разработка приложения для повышения разрешения изображения

Задание на НИР:

В ходе выполнения данной научно-исследовательской работы предстоит проанализировать различные методы и варианты построения архитектур в рамках задачи повышения разрешения изображения и описать выбранную архитектуру. Были определены следующие задачи научной исследовательской работы:

- Описание принципа отбора аналогов.
- Характеристика и перечисление отобранных альтернатив.
- Список критериев с обоснованием выбора каждого.
- Характеристика каждой альтернативы по каждому критерию.
- Изучение нейронных методов повышения разрешения изображения.
- Описание архитектуры предполагаемого решения.

Сроки выполнения НИР: 13.09.2018 – 31.12.2018

Дата сдачи отчета: 31.12.2018

Дата защиты отчета: 31.12.2018

Студентка

Саклакова В.А.

Руководители

Шолохов А.В.

Кринкин К.В.

АННОТАЦИЯ

Целью научно-исследовательской работы является изучение основ машинного обучения, исследование нейронных методов повышения разрешения изображения и описание предполагаемого способа решения поставленной задачи. В ходе данного исследования был произведен обзор предметной области, в частности были рассмотрены такие понятия, как машинное обучение, нейронные сети, были рассмотрены различные нейронные методы повышения разрешения изображения, и описана архитектура предполагаемого способа решения задачи.

SUMMARY

The purpose of the research work is to study of fundamentals of machine learning, to research methods of image super-resolution and description of the proposed solution. During this research the review of data domain was made, in particular such concepts as machine learning, neural networks, were considered, different neural methods of image super-resolution were considered, and description of the architecture of the proposed solution.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	5
1. Описание принципа отбора аналогов.....	6
2. Характеристика и перечисление отобранных альтернатив	6
2.1. Постановка задачи	6
2.2. Методы повышения разрешения изображения	7
3. Список критериев с обоснованием выбора каждого	9
3.2. Характеристика каждой альтернативы по каждому критерию.....	9
3.3. Выводы	11
4. Список используемых терминов	12
5. Нейросетевые методы повышения разрешения изображения	13
5.1. Свёрточная нейронная сеть	13
5.2. Генеративно-состязательная нейронная сеть	14
5.3. Описание архитектуры GAN-сети	16
5.4. Сценарии использования	18
5.5. Выводы	18
ЗАКЛЮЧЕНИЕ	19
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	20

ВВЕДЕНИЕ

В ходе выполнения данной научно-исследовательской работы предстоит подробно изучить предметную область, проанализировать различные методы повышения разрешения изображения и провести их сравнение. Также необходимо отдельно изучить выбранный в результате сравнения метод, проанализировать различные нейронные методы повышения разрешения изображения и описать архитектуру предполагаемого способа решения.

Были определены следующие задачи научной исследовательской работы:

- Описание принципа отбора аналогов.
- Характеристика и перечисление отобранных альтернатив.
- Список критериев с обоснованием выбора каждого.
- Характеристика каждой альтернативы по каждому критерию.
- Изучение нейронных методов повышения разрешения изображения.
- Описание архитектуры предполагаемого решения.

1. Описание принципа отбора аналогов

В рамках поставленной задачи разработки приложения для повышения разрешения изображения исследуется возможность повышения разрешения с помощью программных средств, который ограничивается выбором следующих альтернатив, применимых к данной задаче: линейные, нелинейные и нейросетевые методы повышения разрешения изображения.

2. Характеристика и перечисление отобранных альтернатив

2.1. Постановка задачи

Цифровое изображение имеет конечное количество пикселей. Количество этих пикселей, деленное на единицу площади, называется разрешением цифрового изображения.

В обработке изображений существует класс методов повышения разрешения изображения, которые позволяют качественно увеличить разрешение исходного изображения, при этом происходит преодоление оптического предела объектива и/или физического разрешения цифрового сенсора, который записал изображение [1].

Методы повышения разрешения изображения часто используются для того, чтобы рассмотреть подробности небольшого фрагмента снимка. Их суть заключается в увеличении количества пикселей на некоторую единицу длины (например, на сантиметр или на дюйм). На сегодняшний день существует множество подходов к решению этой задачи, однако наибольшей популярностью пользуются алгоритмы интерполяции (в частности, метод ближайшего соседа, билинейная и бикубическая интерполяция) [1]. Принцип их работы довольно прост: каждый новый пиксель достраивается на основе некоторого набора окружающих пикселей в соответствии с выбранным фильтром. Такие методы позволяют увеличить разрешение изображения довольно быстро, однако они имеют существенный недостаток — небольшие фрагменты получаются слишком гладкими и мелкие детали картинки теряются.

Использование нейросетевых подходов к решению поставленной задачи позволяют обойти эти недостатки.

2.2. Методы повышения разрешения изображения

– Линейные методы;

Простейшим методом повышения разрешения изображений является интерполяция. При интерполяции изображение представляется в виде функции, где пиксели изображения – точки, в которых значения функции известны. Интерполяция – нахождение промежуточных значений по дискретному набору известных значений. Точное восстановление информации путем интерполяции невозможно. Методы интерполяции основаны на использовании свертки изображения и ядер разных типов.

При решении задачи повышения разрешения цифровых изображений требуется не построение непрерывного изображения, а переход с более грубой сетки на более мелкую сетку. Такой процесс называют ресамплингом изображений, а коэффициент отношения шага крупной сетки к шагу мелкой – коэффициентом повышения разрешения изображения.

Самым простым и самым быстрым линейным методом интерполяции является метод «ближайшего соседа», который заключается в растяжении пикселей. В случае ресамплинга с целым коэффициентом увеличения он представляет собой простое повторение пикселей изображения. Недостатком применения данного метода является ярко выраженный эффект ступенчатости. Восстановленная функция становится более гладкой, если вместо ступенек соединять точки прямыми. Так осуществляется линейная интерполяция первого порядка. Билинейной называется линейная интерполяция первого порядка в двумерном случае. В результате применения такого метода интерполяции появляется неярко выраженные эффекты ступенчатости и размытия изображения.

Наиболее распространенные негативные эффекты, возникающие при повышении разрешения изображений, связаны с искажением высокочастотной информации [1]. Такими эффектами являются размытие, алиасинг

(ступенчатость контуров) и эффект Гиббса. В задачах обработки изображений эффект Гиббса проявляется как эффект ложного оконтуривания, проявляющийся в виде ореолов возле резких контуров.

– Нелинейные методы;

Избавиться от недостатков линейных методов интерполяции можно с помощью адаптивных нелинейных методов, для которых ядро задается отдельно для каждого интерполируемого пикселя и зависит от значений пикселей интерполируемого изображения. Примером нелинейных методов является класс градиентных методов. Градиентные алгоритмы позволяют достичь лучших результатов при интерполяции вдоль контуров деталей изображения путем использования разных ядер вдоль и поперек границ (например, использование функции Гаусса с переменными радиусами по разным направлениям) [2]. В целом результат получается близким к результату бикубической интерполяции, но эффект алиасинга оказывается практически полностью подавленным.

– Нейросетевые методы.

Одним из преимуществ использования искусственной нейронной сети является её способность к обучению, которая может реализовываться в виде изменения внутренних свойств нейронной сети для выполнения требуемых функций. Данный подход предпочтителен тем, что мы можем генерировать почти неограниченную выборку низкоразрешенных изображений, причем как шум, так и нелинейная функция, в отличие от предыдущих методов, могут быть практически произвольными. Подобные подходы активно применяются для решения задачи удаления шума с изображений или устранения размытости в изображениях.

В последние годы наилучшие результаты во многих задачах компьютерного зрения были достигнуты с помощью использования глубоких свёрточных нейронных сетей [3].

3. Список критериев с обоснованием выбора каждого

На сегодняшний день существует два выделенных класса методов оценки качества восстановленных изображений: субъективные оценки (экспертные) и математические оценки (метрики) [4]. В рамках данной работы будут выбраны математические оценки, так как они упрощают и ускоряют процесс оценки, снижают затраты на реализацию. Наиболее распространенными метриками для оценки качества восстановленных изображений являются среднеквадратичная ошибка (Mean Square Error (MSE)) и пиковое отношение сигнала к шуму (Peak Signal-to-Noise Ratio (PSNR)) [4].

MSE определяет среднеквадратичное отклонение оригинального и оцениваемого изображения:

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

где m, n – размер сравниваемых изображений I и K .

Данная метрика является самой распространённой [4], так как низкозатратная программная реализация и быстрое вычисление. Из недостатков можно отметить, что не всегда большая разница между значениями пикселей – показатель отличия изображений между собой.

PSNR – мера отношения сигнала к шуму:

$$PSNR = 20 \log_{10} \left(\frac{MAX_I}{\sqrt{MSE}} \right),$$

где MAX_I – максимально возможное значение пикселей на изображении I .

Данная метрика отражает соотношение между максимум возможным значения пикселей изображения и мощностью шума, воздействующего на данное изображение. Единица измерений – децибелы (dB).

3.2. Характеристика каждой альтернативы по каждому критерию

В качестве примера линейного метода была рассмотрена бикубическая интерполяция. Бикубическая интерполяция обладает одним из лучших

сочетаний качества и скорости. Бикубическая интерполяция является обобщением интерполяции кубическими сплайнами на двумерный случай. На рисунке 1 представлен результат работы метода бикубической интерполяции при увеличении масштаба изображения в 4 раза.

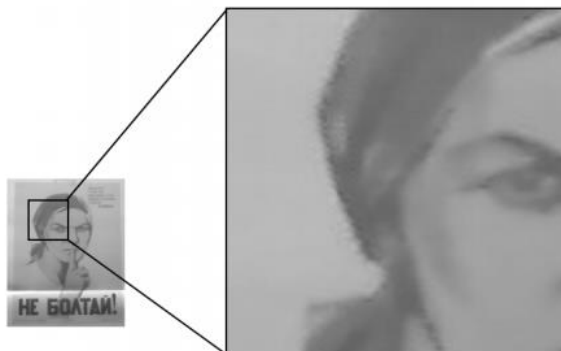


Рисунок 1 – Метод бикубической интерполяции.

В качестве примера нелинейного метода был рассмотрен градиентный метод. Суть градиентного метода в том, что интерполирование производится по краям (контуру) изображения и ядро интерполирования по оси абсцисс отличается от ядра интерполирования по оси ординат. На рисунке 2 представлен результат работы градиентного метода.

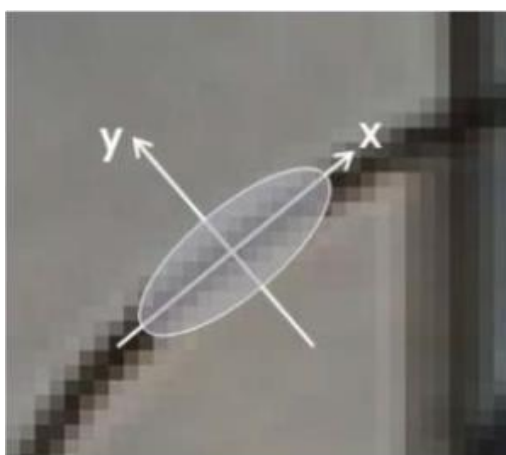


Рисунок 2 – Градиентный метод.

В качестве примера нейросетевого метода была рассмотрена модель GAN сети (подробно описана в п.5.2). На рисунке 3 можно видеть сравнение работы бикубической интерполяции и GAN-сети. Градиентный метод не рассматривался, так как результаты данного метода очень схожи вычислительно и визуально с результатами бикубической интерполяции [5].



Рисунок 3 – Результаты бикубической интерполяции и GAN-сети.

3.3. Выводы

По итогам обзора предложенных альтернатив по выделенным меркам можно сделать вывод, что вычислительные значения данных метрик не всегда сравнимы с разницей человеческого восприятия. Также во многих работах экспериментально это было продемонстрировано ([6], [7], [8]). Также можно видеть, что нейросетевые методы значительно лучше увеличивают разрешение изображения с точки зрения человеческого восприятия (рисунок 3). Также предлагается в качестве дополнительного критерия для более объективной вычислительной оценки взять такую метрику, как Structural Similarity (SSIM). Её отличительной чертой можно назвать тот факт, что она учитывает структурные свойства изображения, и несколько больше соответствует особенностям человеческого восприятия.

4. Список используемых терминов

Алгоритм – набор инструкций, которые описывают порядок действий для достижения некоторого результата.

Изображение – предмет, рисунок, изображающий кого-нибудь, что-нибудь; зрительное воспроизведение чего-нибудь.

Интерполяция – способ нахождения промежуточных значений величины по имеющемуся дискретному набору известных значений.

Искусственные нейронные сети (ИНС) — математическая модель, а также её программное или аппаратное воплощение, построенная по принципу организации и функционирования биологических нейронных сетей.

Математическая модель – это сильно приближенное описание объекта моделирования, выраженное с помощью математических символов.

Машинное обучение (Machine Learning) — обширный подраздел искусственного интеллекта, изучающий методы построения алгоритмов, способных обучаться.

Нейрон - узел искусственной нейронной сети, являющийся упрощённой моделью естественного нейрона.

Обучение – процесс первоначальной настройки нейронной сети.

Пиксель – наименьший логический элемент двумерного цифрового изображения в растровой графике, или элемент матрицы дисплеев, формирующих изображение.

Приложение – программное обеспечение, программа.

Разрешение изображения - величина, определяющая количество точек (элементов растрового изображения) на единицу площади (или единицу длины).

Функция — математическое понятие, отражающее связь между какими-либо значениями.

5. Нейросетевые методы повышения разрешения изображения

Одним из преимуществ использования искусственной нейронной сети является её способность к обучению, которая может реализовываться в виде изменения внутренних свойств нейронной сети для выполнения требуемых функций. Данный подход предпочтителен тем, что мы можем генерировать почти неограниченную выборку низкоразрешенных изображений, причем как шум, так и нелинейная функция, в отличие от предыдущих методов, могут быть практически произвольными. Подобные подходы активно применяются для решения задачи удаления шума с изображений или устранения размытости в изображениях.

В последние годы наилучшие результаты во многих задачах компьютерного зрения были достигнуты с помощью использования глубоких свёрточных нейронных сетей [3].

5.1. Свёрточная нейронная сеть

Свёрточная нейронная сеть – специальная архитектура искусственных нейронных сетей, предложенная Яном Лекуном в 1988 году и нацеленная на эффективное распознавание изображений, входит в состав технологий глубокого обучения. Использует некоторые особенности зрительной коры, в которой были открыты так называемые простые клетки, реагирующие на прямые линии под разными углами, и сложные клетки, реакция которых связана с активацией определённого набора простых клеток. Таким образом, идея свёрточных нейронных сетей заключается в чередовании свёрточных слоев и слоёв подвыборки. Структура сети — однонаправленная (без обратных связей), принципиально многослойная. Для обучения используются стандартные методы, чаще всего метод обратного распространения ошибки. Функция активации нейронов (передаточная функция) — любая, по выбору исследователя.

Название архитектура сети получила из-за наличия операции свёртки, суть которой в том, что каждый фрагмент изображения умножается на матрицу

(ядро) свёртки поэлементно, а результат суммируется и записывается в аналогичную позицию выходного изображения.

На рисунке 4 изображена архитектура свёрточной нейронной сети.

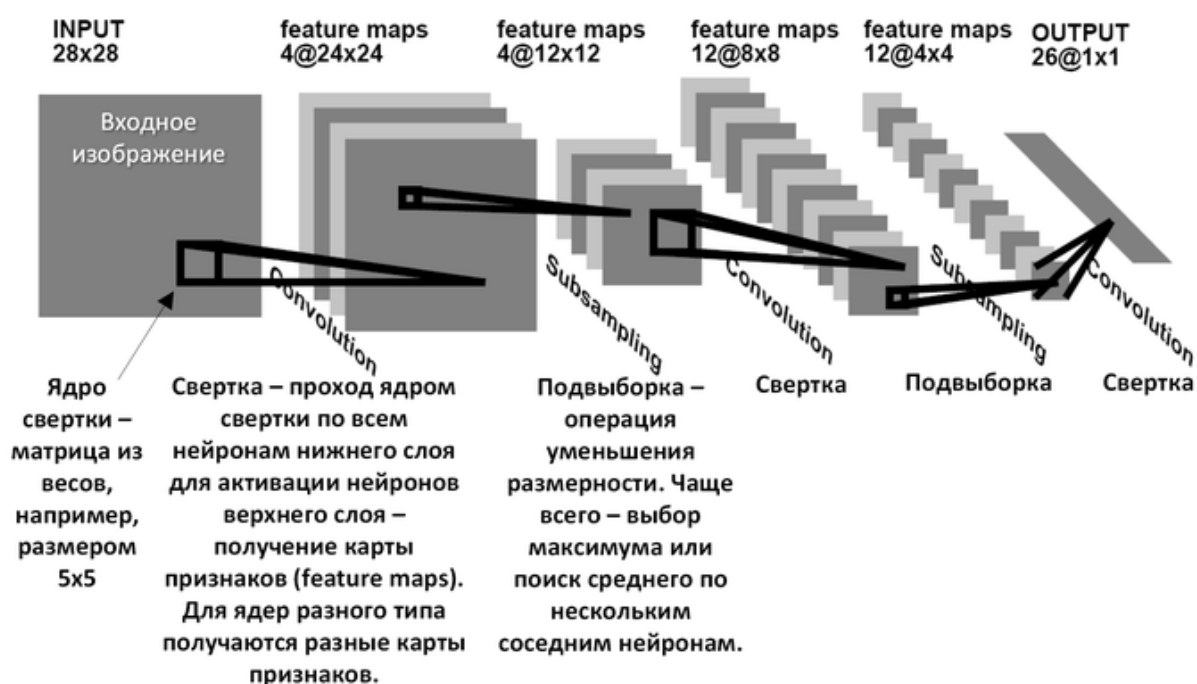


Рисунок 4 – Архитектура свёрточной нейронной сети.

Большинство подходов к решению задачи генерации фотореалистичных изображений используют для задания вероятностного распределения графические модели. Подобные модели зачастую оказываются неэффективными из-за необходимости проведения приближенных процедур вывода, а также использования Марковских цепей для генерации выборки. Недавно предложенная модель конкурирующих сетей [9] позволяет выучивать сложные распределения, а также генерировать выборки произвольного размера используя только вычислительно эффективный алгоритм обратного распространения ошибки для нейронных сетей.

5.2. Генеративно-состязательная нейронная сеть

Генеративно-состязательная сеть (Generative adversarial network) – алгоритм машинного обучения без учителя, построенный на комбинации из двух нейронных сетей, одна из которых (сеть G) генерирует образцы, а другая (сеть D) старается отличить правильные («подлинные») образцы от

неправильных. Так как сети G и D имеют противоположные цели — создать образцы и отбраковать образцы — между ними возникает Антагонистическая игра. Генеративно-состязательную сеть описал Ян Гудфеллоу из компании Google в 2014 г.

Использование этой техники позволяет, в частности, генерировать фотографии, которые человеческим глазом воспринимаются как натуральные изображения. Кроме того, GAN может использоваться для улучшения качества нечётких или частично испорченных фотографий [4].

На рисунке 5 изображена модель GAN сети.

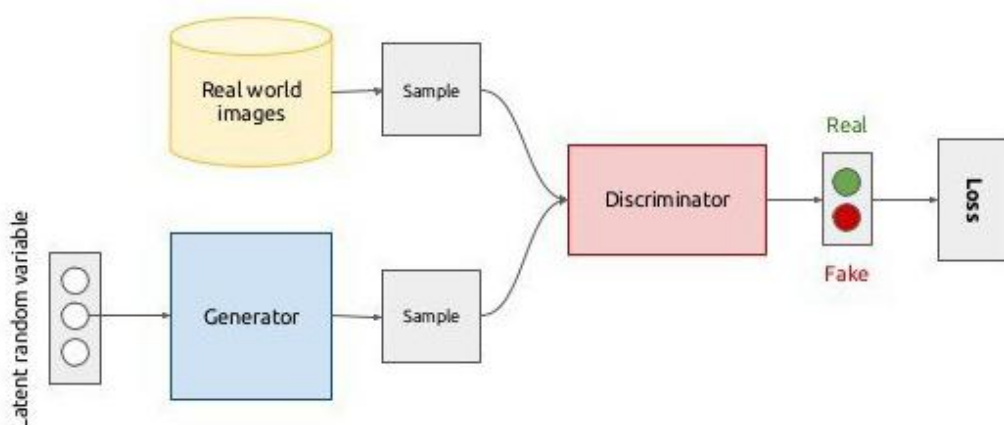


Рисунок 5 – Модель GAN сети.

Принцип состязательности в сети GAN нередко описывается через метафоры. Например, генеративная сеть уподобляется фальшивомонетчику или подделывателю картин, а дискриминативная — эксперту который стремится распознать подделку. Другой пример — образ двух боксёров, один из которых учился у мастера, а второй вынужден подражать ученику. В популярном приложении генерации человеческих лиц в качестве подлинных данных выступают реальные фотографии, а генеративная сеть пытается создать искусственные лица, варьируя комбинации таких латентных параметров, как цвет волос, пропорции лица, разрез глаз, форма носа, размер ушей, наличие бороды и усов и т.д.

Формально подобную модель можно записать в виде следующей минимакс игры:

$$G, D \leftarrow \min_G \max_D (\mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))])$$

Здесь z — это некоторые модельные данные (случайный шум), $p_z(z)$ — это модельное распределение, выборки из которого подаются на вход сети G . x — объекты из распределения данных $p_{data}(x)$, из которого сеть G учится генерировать выборки. В качестве модельного распределения p_z обычно используется равномерное распределение $U[-1,1]$.

5.3. Описание архитектуры GAN-сети

Для сети Дискриминатора и сети Генератора в основу легли архитектуры не обычных нейронных сетей, а свёрточных. Таким образом, были совмещены генеративно-состязательные сети и свёрточные сети.

Генератор G , дискриминатор D . Есть какое-то количество оригинальных изображений X для D , и пусть на выходе у него будет число диапазоном от нуля до единицы, чтобы оно выражало уверенность D в том, что поданные ему на вход изображения оригинальные. Еще — поскольку G нейронная сеть, ей нужны какие-то входные данные, назовем их z . В общем случае, это просто случайный шум, который модель будет стараться превратить в реальное изображение.

Тогда, очевидно, цель G — это максимизировать $D(G(z))$, то есть сделать так, чтобы D был уверен, что подделки — настоящие.

Цель D посложнее — ему нужно одновременно положительно опознавать оригиналы, и отрицательно — подделки. Запишем это как максимизацию $D(x)(1 - D(G(z)))$. Умножение можно превратить в сложение, если взять логарифм, поэтому получается:

Для дискриминатора: максимизировать $\log D(x) + \log(1 - D(G(z)))$

Для генератора: максимизировать $\log D(G(z))$

Так как в поставленной задаче на вход генератору подается не шум, а изображение с низким разрешением, то применяя общую формулу на задачу, получается данная формула:

$$\min_G \max_D V(G, D) = E_{I_{generated} \sim p_{train}(I_{original})} [\log D(I_{original})] + E_{I_{generated} \sim p_G(I_{input})} [\log(1 - D(G(I_{input})))] \quad (1)$$

которая значит, что $I_{original}$ – множество (тренировочное), которое заранее приведено к какому-то распределению. $I_{generated}$ – множество генерируемых изображений Генератором должно быть таким же распределением, как и $I_{original}$

На вход такой сети подаётся изображение 16x16, которое сеть Генератора масштабирует в 4 раза (до 64x64), и пытается восстановить размытое изображение. Сеть Дискриминатора обучается на то, чтобы различать оригинальные высокоразрешенные изображения и изображения, которые являются выходными данными из сети Генератора.

На рисунке 6 представлена архитектура DCGAN сети, которая ляжет в основу разрабатываемого приложения.

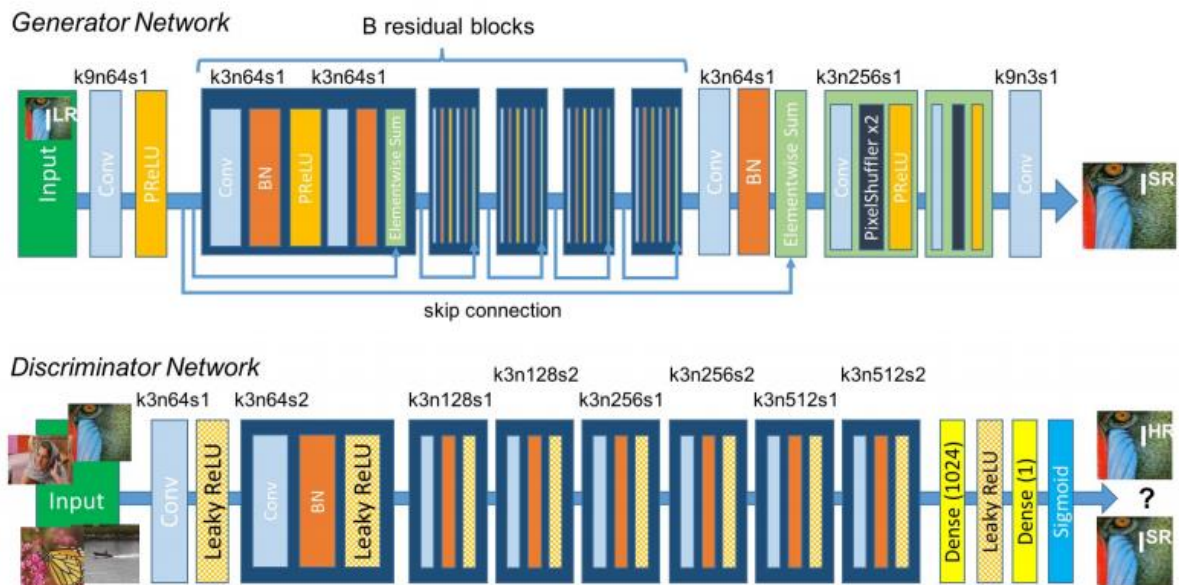


Рисунок 6 – Архитектура DCGAN сети.

5.4. Сценарии использования

В общем предполагается разработка приложения, в основу которого ляжет архитектура GAN-сети. На рисунке 7 представлен сценарий использования приложения.

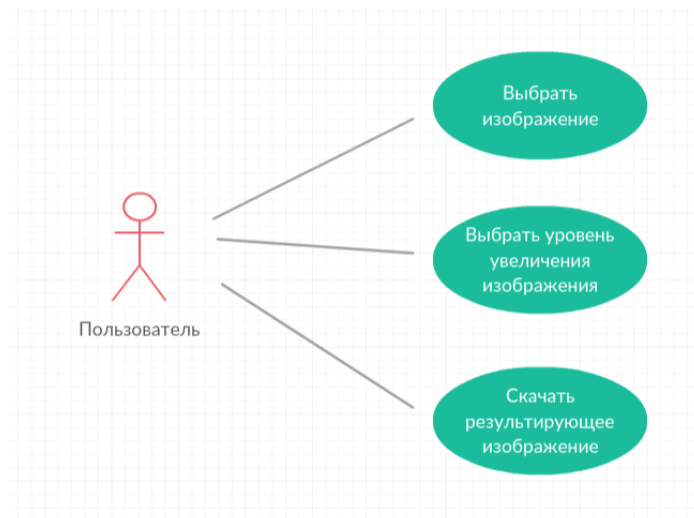


Рисунок 7 – Сценарий использования.

5.5. Выводы

Рассмотрев область нейросетевых методов повышения разрешения изображения, можно сделать вывод, что данная область очень широка, и существует помимо объёмного выбора самих методов, но и множества параметров оценки работы уже выбранной архитектуры, так называемых гиперпараметров. Поэтому предлагается отдельно разработать дополнительный программный комплекс, который будет оценивать и сравнивать различные архитектуры нейронных сетей. По результатам работы, будет выбрана лучшая, которая в итоге ляжет в основу конечного приложения.

ЗАКЛЮЧЕНИЕ

В результате проведения научной исследовательской работы были выполнены следующие задачи:

- Описаны принципа отбора аналогов.
- Выполнена характеристика отобранных альтернатив.
- Приведён список критериев с обоснованием выбора каждого.
- Выполнена характеристика каждой альтернативы по каждому критерию.
- Изучены нейронные методы повышения разрешения изображения.
- Описана архитектура предполагаемого решения.

Также в течение текущего семестра в рамках включённого обучения в Финляндии был пройден курс «Глубокое обучение для когнитивных вычислений». В рамках изучения дисциплины были исследованы на практике методы и алгоритмы, используемые в когнитивных вычислениях с особым упором на технологию глубокого обучения, что также позволяет применить эти знания в рамках работы над магистерской диссертацией.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Маркелов К.С. Модель повышения информативности цифровых изображений на базе метода суперразрешения // Инженерный вестник. 2013, вып. (№3). С. 525-542.
2. Насонов А.В. Регуляризирующие методы повышения изображений: автореф. дисс. канд. техн. наук / МГУ им. М.В. Ломоносова, М., 2011.
3. J. Kim, J. K. Lee, and K. M. Lee, “Accurate image super-resolution using very deep convolutional networks,” arXiv preprint arXiv:1511.04587, 2015.
4. Сидоров Д.В. К вопросу оценки качества множества восстановленных изображений // Прикладная информатика. 2008. №4. URL: <https://cyberleninka.ru/article/n/k-voprosu-otsenki-kachestva-mnozhestva-vosstanovlennyh-izobrazheniy>.
5. Шергин В.С. Увеличение полей цифровых изображений методом полей ориентации: диплом. работа. Саратов. гос. университет, Саратов, 2017.
6. J. Johnson, A. Alahi, and L. Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” arXiv preprint arXiv:1603.08155, 2016.
7. D. Kundu and B. L. Evans, “Full-reference visual quality assessment for synthetic images: A subjective study,” in Image Processing (ICIP), 2015 IEEE International Conference on. IEEE, 2015, pp. 2374—2378.
8. P. Hanhart, P. Korshunov, and T. Ebrahimi, “Benchmarking of quality metrics on ultra-high definition video sequences,” in Digital Signal Processing (DSP), 2013 18th International Conference on. IEEE, 2013, pp. 1—8.
9. L. Xu, J. S. Ren, C. Liu, and J. Jia, “Deep convolutional neural network for image deconvolution,” in Advances in Neural Information Processing Systems, 2014, pp. 1790—1798.