

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: Описание предполагаемого способа решения
на тему «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: Описание предполагаемого способа решения на тему «Разработка приложения для повышения разрешения изображения»

Задание на НИР:

В ходе выполнения данной научной исследовательской работы предстоит описать предполагаемый способ решения в рамках задачи повышения разрешения изображения. Были определены следующие задачи научной исследовательской работы:

- Изучение основ машинного обучения.
- 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. Обзор предметной области	6
1.1. Машинное обучение	6
1.2. Нейронные сети	7
2. Описание предполагаемого способа решения задачи	9
2.1. Постановка задачи	9
2.2. Нейросетевые методы повышения разрешения изображения	9
2.3. Описание архитектуры GAN-сети	13
2.4. Сценарии использования	14
ЗАКЛЮЧЕНИЕ	16
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	17

ВВЕДЕНИЕ

В ходе выполнения данной научной исследовательской работы предстоит подробно изучить предметную область выбранного на предыдущем этапе метода, проанализировать различные нейронные методы повышения разрешения изображения и описать архитектуру предполагаемого способа решения.

Были определены следующие задачи научной исследовательской работы:

- Изучение основ машинного обучения.
- Изучение нейронных методов повышения разрешения изображения.
- Описание архитектуры предполагаемого решения.

1.Обзор предметной области

1.1. Машинное обучение

Машинное обучение (Machine Learning) — обширный подраздел искусственного интеллекта, изучающий методы построения алгоритмов, способных обучаться. Различают два типа обучения. Обучение по прецедентам, или индуктивное обучение, основано на выявлении общих закономерностей по частным эмпирическим данным. Дедуктивное обучение предполагает формализацию знаний экспертов и их перенос в компьютер в виде базы знаний. Дедуктивное обучение принято относить к области экспертных систем, поэтому термины машинное обучение и обучение по прецедентам можно считать синонимами.

Машинное обучение находится на стыке математической статистики, методов оптимизации и классических математических дисциплин, но имеет также и собственную специфику, связанную с проблемами вычислительной эффективности и переобучения. Многие методы индуктивного обучения разрабатывались как альтернатива классическим статистическим подходам. Многие методы тесно связаны с извлечением информации и интеллектуальным анализом данных (Data Mining).

В основном, выделяют следующие основные типы задач:

- Обучение с учителем (supervised learning) — наиболее распространённый случай. Каждый прецедент представляет собой пару «объект, ответ». Требуется найти функциональную зависимость ответов от описаний объектов и построить алгоритм, принимающий на входе описание объекта и выдающий на выходе ответ. Функционал качества обычно определяется как средняя ошибка ответов, выданных алгоритмом, по всем объектам выборки.
- Обучение без учителя (unsupervised learning). В этом случае ответы не задаются, и требуется искать зависимости между объектами.

– Обучение с подкреплением (reinforcement learning). Роль объектов играют пары «ситуация, принятое решение», ответами являются значения функционала качества, характеризующего правильность принятых решений (реакцию среды). Как и в задачах прогнозирования, здесь существенную роль играет фактор времени. Примеры прикладных задач: формирование инвестиционных стратегий, автоматическое управление технологическими процессами, самообучение роботов, и т.д.

1.2. Нейронные сети

Искусственная нейронная сеть (ИНС) — математическая модель, а также её программное или аппаратное воплощение, построенная по принципу организации и функционирования биологических нейронных сетей — сетей нервных клеток живого организма. С точки зрения машинного обучения, нейронная сеть представляет собой частный случай методов распознавания образов, дискриминантного анализа, методов кластеризации и т. п.

У искусственных нейронов есть входы и выходы. Выходы нейронов связаны со входами других нейронов. Места, где они соединяются, называются синапсами. В синапсах проходящий сигнал может либо усилиться, либо ослабнуть. Каждый синапс может изменять степень усиления/ослабления сигнала. Когда синапсов очень много, можно подобрать такие их параметры, что входной сигнал будет преобразовываться в правильный выходной. На рисунке 1 представлена полная внутренняя модель искусственного нейрона.

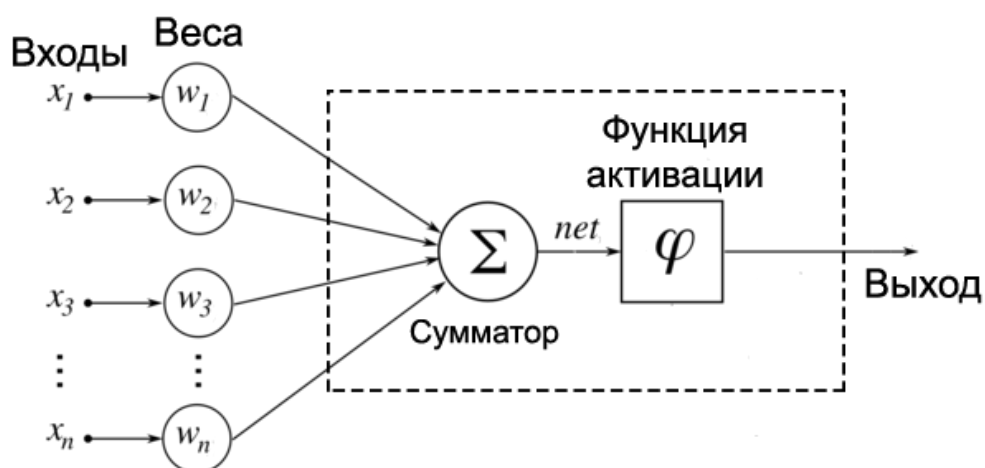


Рисунок 1 – Модель искусственного нейрона.

У каждого нейрона, в том числе и у искусственного, должны быть входы, через которые он принимает сигнал. Поступившие на входы сигналы умножаются на свои веса. Сигнал первого входа x_1 умножается на соответствующий этому входу вес w_1 . В итоге получаем $x_1 w_1$. И так до n -ого входа. В итоге на последнем входе получаем $x_n w_n$. Далее все произведения передаются в сумматор, который суммирует все входные сигналы, умноженные на соответствующие веса:

$$x_1 w_1 + x_2 w_2 + \dots + x_n w_n = \sum_{i=1}^n x_i w_i$$

Результатом работы сумматора является число, называемое взвешенной суммой. Просто так подавать взвешенную сумму на выход бессмысленно. Нейрон должен как-то обработать ее и сформировать адекватный выходной сигнал. Именно для этих целей и используют функцию активации, которая преобразует взвешенную сумму в такое число, которое и является выходом нейрона.

2. Описание предполагаемого способа решения задачи

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. 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 [2].

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 2 изображена архитектура свёрточной нейронной сети.

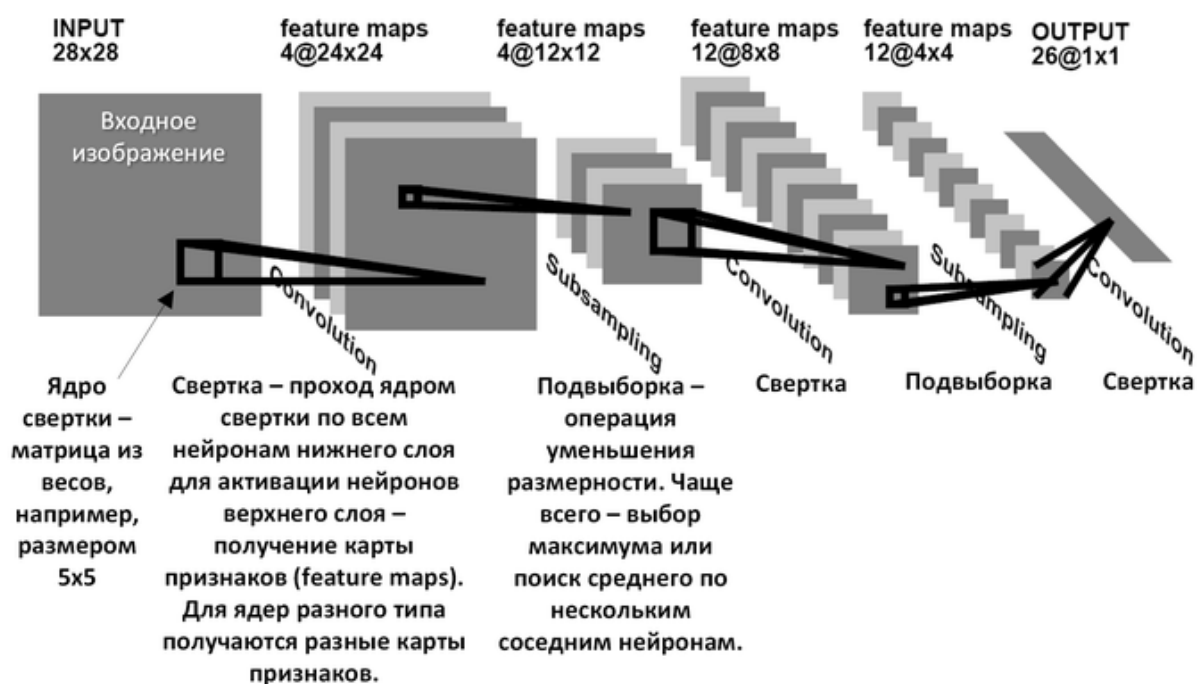


Рисунок 2 – Архитектура свёрточной нейронной сети.

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 [3] позволяет выучивать сложные распределения, а также генерировать выборки произвольного размера используя только вычислительно эффективный алгоритм обратного распространения ошибки для нейронных сетей.

Генеративно-сопоставительная 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-сопоставительную 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 3 изображена модель GAN сети.

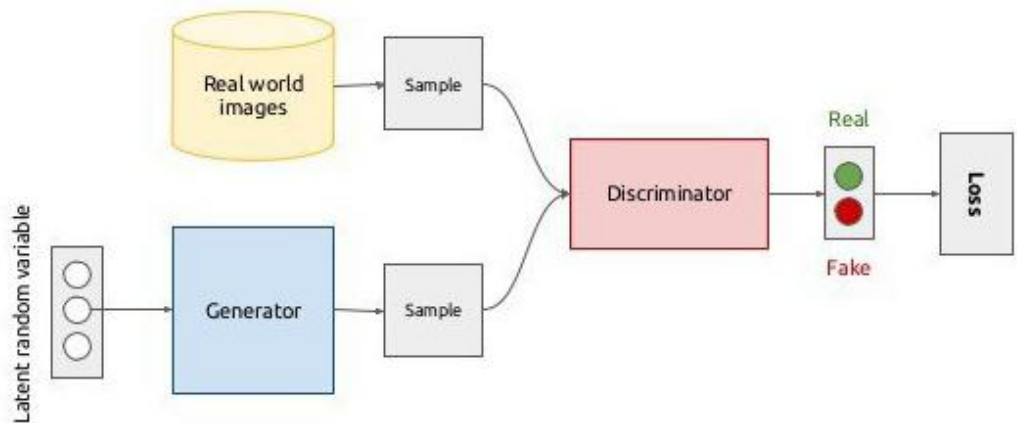


Рисунок 3 – Модель GAN сети.

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 $U[-1,1]$.

2.3. 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 4 представлена архитектура DCGAN сети, которая ляжет в основу разрабатываемого приложения.

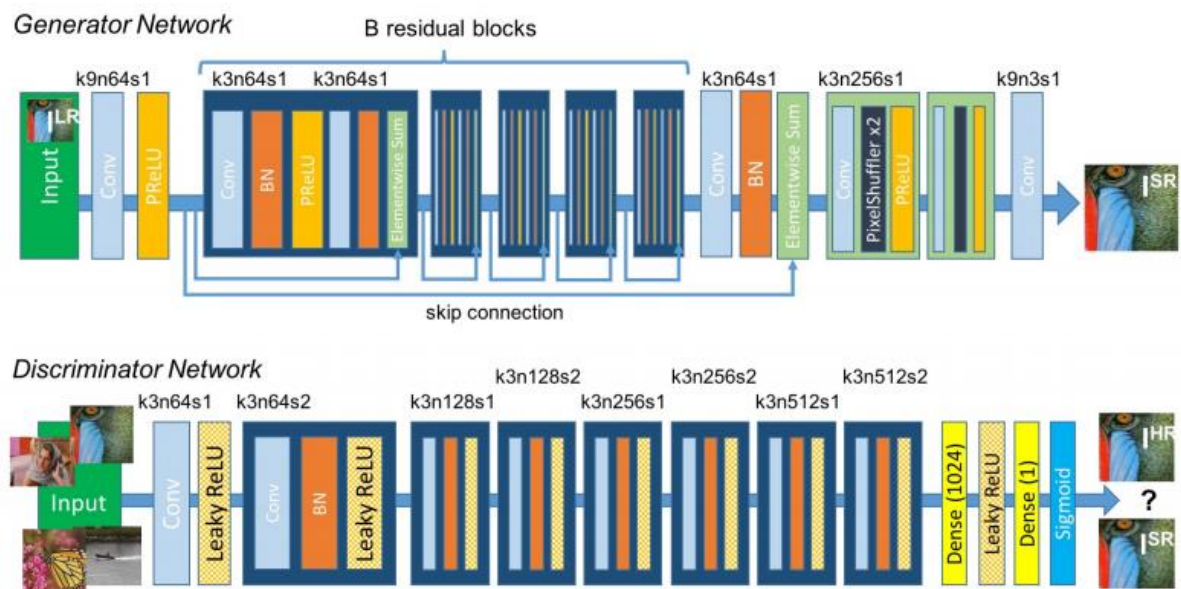


Рисунок 4 – Архитектура DCGAN сети.

2.4. Сценарии использования

В общем предполагается разработка приложения, в основу которого ляжет архитектура GAN-сети. На рисунке 5 представлен сценарий использования приложения.

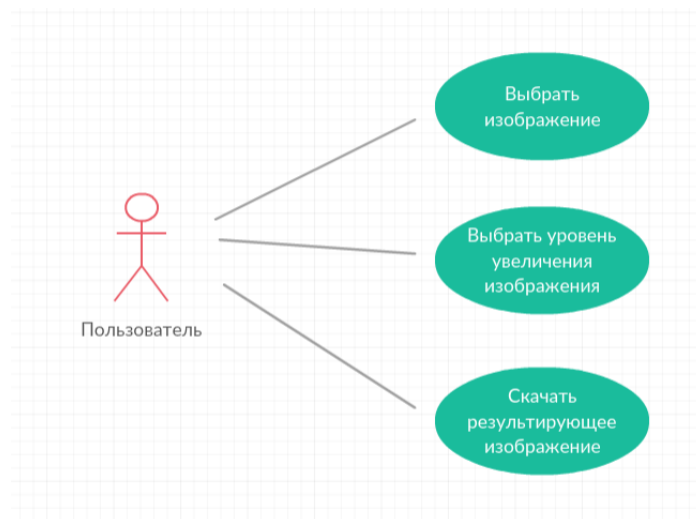


Рисунок 4 – Сценарий использования.

ЗАКЛЮЧЕНИЕ

В результате проведения научной исследовательской работы были выполнены следующие задачи:

- Изучены основы машинного обучения.
- Изучены нейронные методы повышения разрешения изображения.
- Описана архитектура предполагаемого решения.

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. J. Kim, J. K. Lee, and K. M. Lee, “Accurate image super-resolution using very deep convolutional networks,” arXiv preprint arXiv:1511.04587, 2015.
3. 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.