Санкт-Петербургский государственный университет

Кафедра компьютерного моделирования и многопроцессорных систем

**Голиков Матвей Алексеевич**

**Выпускная квалификационная работа бакалавра**

**Применение алгоритмов машинного обучения в промышленном дизайне**

Направление 02.03.02

«Фундаментальная информатика и информационные технологии»

ООП «Программирование и информационные технологии»

Научный руководитель,

PhD,

доцент

Корхов В. В.

Санкт-Петербург

2019

# Содержание

[Введение 3](#_Toc10150916)

[Постановка задачи 4](#_Toc10150917)

[Обзор литературы 6](#_Toc10150918)

[Глава 1. Обзор области 7](#_Toc10150919)

[1.1. Применение компьютерных технологий в промышленном дизайне 7](#_Toc10150920)

[1.2. Модели эстетического оценивания 7](#_Toc10150921)

[1.3. Генеративный дизайн 7](#_Toc10150922)

[1.4. Эстетическое оценивание объектов и изображений 8](#_Toc10150923)

[Глава 2. Выбор метода для исследования 10](#_Toc10150924)

[2.1. Вариационный автокодировщик 10](#_Toc10150925)

[Глава 3. Выбор и подготовка данных 15](#_Toc10150926)

[3.1. Выбор наборов данных 15](#_Toc10150927)

[3.2. Подготовка данных 16](#_Toc10150928)

[Глава 4. Исследование 18](#_Toc10150929)

[4.1. Применение VAE к задаче генерации цифр 18](#_Toc10150930)

[4.2. Применение VAE к задаче генерации лиц 19](#_Toc10150931)

[4.3. Исследование генерации красивых цифр 19](#_Toc10150932)

[Вывод 23](#_Toc10150933)

[Заключение 24](#_Toc10150934)

[Список литературы 25](#_Toc10150935)

Введение

Современные достижения в компьютерной технике позволяют обрабатывать немыслимое количество собранных данных. Ранее были предложены методы для обработки информации, позволяющие в упрощенном виде моделировать работу человеческого мозга. Такие модели получили название Искусственные Нейронные Сети(ИНС).

Преимуществами данного подхода являются обширность решаемых задач, устойчивость к шумам входных данных, адаптация к изменениям, отказоустойчивость, сверхбыстрое взаимодействие.[1]

Несмотря на все перечисленные преимущества, даже спустя более 15 лет истории использования ИНС, в этой области до сих пор не разрешены многие вопросы, начиная от причин получения таких результатов, до способов применения ИНС к различным задачам.

В последнее время ИНС широко применяется в промышленных системах, выполняя задачи, на которые ранее тратили десятки часов инженеры и дизайнеры. Правильное применение алгоритмов машинного обучения помогают не только облегчить работу, но и зачастую полностью переложить ее на компьютер.

Постановка задачи

В настоящий момент существуют публикации и системы, задачей которых является оптимизация конструкций в промышленности используя ИНС. Проблема заключается в том, что все они предполагают наличие заранее определенной эстетической формы, а оптимизация ведется только по внутренней структуре или поверхности этой формы [2,3]. Для использования полного потенциала ИНС на них также следует возложить задачу генерации и эстетических форм, что приводит нас к логически полному процессу создания объекта в промышленном дизайне: генерации полностью нового, оптимизированного объекта. В идеальном представлении должен существовать такой программный инструмент, позволяющий генерировать некие эстетические формы или объекты, которые в дальнейшем могут быть использованы для оптимизации уже готовыми решениями в области генеративного дизайна.

Проблемы подхода, связанного с созданием новых объектов заключается в их правильной оценке с точки зрения дизайна, так как субъективное ощущение чего-то красивого не всегда возможно описать математическими уравнениями. Но появление работ, связанных с оценками эстетичности картинок[4], красоты лиц[5,6] и качества фотографий[7] на основе алгоритмов машинного обучения позволяют задуматься о появлении систем, позволяющих не только оценивать, но и создавать объекты, основываясь на оценках эстетичности исходных данных.

Для начала работы над такими системами, необходимо:

* исследовать решения из смежных областей;
* подобрать наиболее подходящий метод машинного обучения;
* подобрать наборы данных, позволяющие легко провести анализ после исследования;
* провести эксперименты, которые позволят наглядно определить, возможно ли перенести ту же степень красоты с объекта одного набора данных на сгенерированный объект из другого набора.

Обзор литературы

Информация, предоставленная в [1] дает краткий экскурс по основным понятиям Искусственных Нейронных сетей.

Источники [2, 3, 4, 7, 9] использовались для обзора исследовательской области, а также подтолкнули на идею использования методов, в которых присутствует скрытое пространство.

[5, 6, 11, 14] – полезные источники наборов данных, так же в статьях присутствует аналитическая и практическая часть, позволяющая углубиться в исследования оценки эстетичности изображений.

Статьи [8, 10] позволили более точно понять о положении компьютерных технологий в промышленном дизайне.

Статья [12] послужила прочной теоретической базой в исследовании. В ней содержится подробное описание и принципы работы использованного метода.

Ресурс [13] включает в себе невероятный обзор вариационных автокодировщиков, предоставленный материал очень сильно помог при проведении исследования и позволил быстрее освоиться в использованных алгоритмах.

[15, 16] – официальные ссылки на использованные инструменты в ходе работы.

Глава 1. Обзор области

* 1. Применение компьютерных технологий в промышленном дизайне

Идея применения не только вычислительных возможностей компьютеров появилась еще в прошлом веке, компьютеры стали использоваться при проектировании, создании музыки, анимации, создании промышленных продуктов. Причем область их применения не ограничивалась проектированием, некоторые функции человека по созданию новых эстетических форм, генерации новых картин, а также помощников в создании вариантов дизайна переложили на компьютеры.

В стандартном варианте для промышленном дизайне используются Системы Автоматизированного Проектирования(САПР), которые позволяют в полной мере проводить все работы по разработке и проектированию проектов любой сложности, предоставляя более широкие возможности к визуализации, упрощая работу инженеров и дизайнеров. Наиболее популярными системами на рынке являются AutoCAD, AutoDesk 3DS Max, ArchiCAD. [8]

1.2. Модели эстетического оценивания

Следующим шагом в применении компьютерных технологий стали работы, посвященные построению моделей, позволяющих оценить эстетичность создаваемого объекта по заранее известным правилам. Данный подход можно назвать прародителем генеративного дизайна, так как, по сути, проводятся те же этапы создания объекта. А за основу правил берутся усредненные данные уже готовых объектов, что отдаленно напоминает машинное обучение. Одним из примеров таких моделей может является система BRIDGER [9].

1.3. Генеративный дизайн

Генеративный дизайн как новый подход к построению готовых образцов для дальнейшей их оценки дизайнерами очень перспективен, так сама его модель позволяет внедрить на нескольких этапов машинное обучение. По мнению автора статьи [10], основными методами в генеративном дизайне мы выделяются:

* Установка параметров;
* Построение алгоритмов и правил;
* Программирование процессов;
* Варьирование результатов моделирования;
* Визуализация процессов.

В настоящий момент генеративный дизайн повсеместно используются в промышленных целях, позволяя задавать определенные правила, по которым должен строиться новый объект, машина может создавать продукт, который будет превосходить такой же продукт, созданный человеком, не только по заданным критериям, но и скорости его создания.

1.4. Эстетическое оценивание объектов и изображений

На данный момент собрано множество наборов данных, включающих в себя оценки эстетической привлекательности изображений или объектов на них. В качестве примера могут служить:

* SCUT-FBP [5]: датасет из 500 азиатских женских лиц, для каждого сопоставляется значение от 1 до 5, где 1 – совсем не привлекательная, 5 – очень привлекательная.
* AVA: A large-scale database for aesthetic visual analysis [11]: набор из более чем 250000 картин с большим разнообразием метаданных, включая оценку эстетичности и 60 семантических признаков и тд.

В работе [5] были применены различные алгоритмы машинного обучения для оценки красоты женщины на фотографии, результат, полученный в ходе этого исследования подтверждает теорию о том, что компьютер может оценить эстетическую составляющую объекта на изображении в рамках одного и того же датасета, без переноса “восприятия” на другие объекты.

На основе датасета AVA и других, в которых возможна эстетическая оценка, существуют работы, показывающие хорошую точность в определении красоты изображения. Наиболее интересным кажется подход, основанный на отображении исходного изображения в некоторое пространство. Проведя анализ этого пространства, можно оценить красивость по удаленности подаваемых данных от эстетически привлекательных. Анализируя это пространство, мы можем получить оценку эстетической привлекательности подаваемого изображения, исходя из его близости к другим объектам в “эстетическом” пространстве [4,7].

Глава 2. Выбор метода для исследования

2.1. Вариационный автокодировщик

Основываясь на опыте работ по оцениванию эстетичности объектов, картин и тд., наиболее подходящим для решения это задачи будет алгоритм, который переводит исходные данные в некоторое пространство, а после это может восстанавливать (декодировать) по вектору из этого пространства объекты. Лучшим решением для исследования, по моему мнению, будет Вариационный автокодировщик (VAE). Для этого рассмотрим математическую составляющую[12]:

– скрытые переменные

– данные.

Правая часть выражения – моделирование новых объектов.

— вероятность изображения конкретный цифры,

— распределение параметров в скрытом пространстве, найденных моделью,

— вероятность появление картинки при условии Z, то есть при конкретных параметрах из скрытого пространства.

Для того, чтобы создавать новые объекты нужно ввести функцию, назовем ее . Прибавляя шум , распишем вероятность появления картинки при условии :

То есть – генерирующая функция, которая должна как можно точнее воссоздать объекты.

Пусть у нас есть детерминированное семейство функций с параметром : , но если – случайный вектор, а – зафиксированный, то – случайная переменная в X. Таким образом, нам нужно найти такое , при которых мы сможем сэмплировать из так, чтобы c большой вероятностью была похоже на экземпляр из . Получается, что нам нужно максимизировать вероятность для каждого X в тренировочной выборке:

|  |  |
| --- | --- |
|  | (1) |

Здесь заменено на распределение , что позволило ввести явную зависимость от по закону полной вероятности.

По методу максимального правдоподобия, данная модель будет генерировать объекты максимального похожие на те, что присутствовали в тренировочных данных. В VAE выходное распределение Гауссово, то есть

Для того, чтобы решить уравнение (1) нужно обозначить две проблемы:

* Какой смысл кроется за скрытыми параметрами z и как их определить;
* Как решить интеграл по .

Первая проблема решается  
заменой на , где – многослойная нейронная сеть.

Вторая проблема заключается в максимизации (1), где . Если найти вычислимую формулу для и взять от нее градиент, то тогда мы можем оптимизировать модель использую стохастический градиентный спуск. Проблема заключается в том, что нам нужно сэмплировать очень больше количество и вычислять приближенное значение но для получения точного значения понадобится слишком большое значение . Так же еще одна проблема будет заключаться в оценке похожести сгенерированного и исходного объекта, так как малейшее попиксельное отклонение сильно влияет на результат метрики похожести. Конечно, данную проблему можно решить более совершенной метрикой, но ее выбор для каждого класса задач или наборов данных очень трудоемкий процесс.

Поэтому в VAE допускают, что почти для всех , значение близко к 0. И поэтому не влияет оценку . Поэтому для вычисления берут только те , которые с наибольшей вероятностью сгенерируют . Для этого вводят новое распределение , принимая на вход , на выходе получим распределение по , которые с наибольше вероятностью сгенерируют . Это распределение позволяет вычислять намного проще.

Для начала строится связь между и :

Записывается расстояние Кульбака-Леблейра для и

Применяя правило Байса к , в это уравнение вводится и

Домножая на -1, переставляя получаем:

|  |  |
| --- | --- |
|  | (2) |

В уравнении (2) – любое распределение, поэтому стоит выбирать , зависящее от , что сделает очень малым.

В левой части (2) находится , который нужно максимизировать плюс ошибка от распределения , которая должна почти равняться 0 при довольно большой емкости , правую сторону нужно оптимизировать стохастическим градиентным спуском (СГС).

Для применения СГС нужно определить , обычно в качестве этого распределения выбирают , где и – произвольные функции с параметром , который узнается из данных. На практике пара этих функций реализуется нейронными сетями, а является диагональной матрицей, что позволяет вычислять правую часть формулы (2).

Слагаемое , с учетом упрощения   может быть вычислено как:

Последнее, что нужно определить для применения СГС: распространение ошибки.

Для начала берутся случайные значения  и подаются в для получения приближенного значения ]. Так как мы не можем использовать обратное распространение ошибки через случайные значения, из-за этого необходимо использовать “reparametrization trick”, схема его использования показана на рисунке 1.

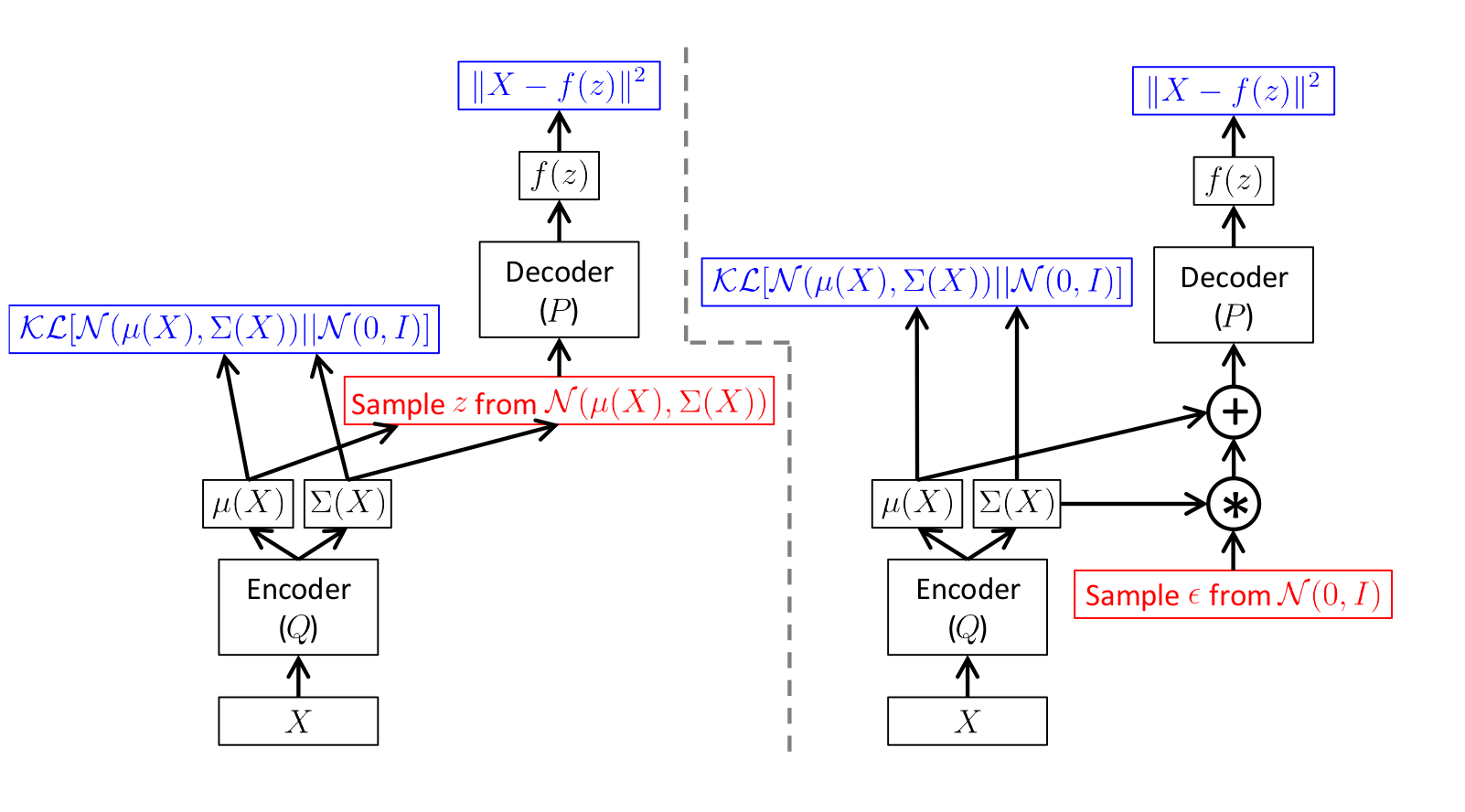


Рис. 1 Схема VAE, справа - с “reparametrization trick”, а слева без. Красные блоки – сэмплирование, синие – вычисление ошибки.

“reparametrization trick” – это решение, реализующее перенос сэмлирования на внешний слой. Зная – среднюю ковариацию и мы можем сэмплировать из , сначала сэмплировав из и потом вычисляя вектор . Таким образом, финальное выражание от которого мы берем градиент:

*“После того как мы обучили такой вариационный автоэнкодер, декодер становится полноправной генеративной моделью. По сути и энкодер-то нужен в основном для того, чтобы обучить декодер отдельно быть генеративной моделью.”* [13]

Глава 3. Выбор и подготовка данных

3.1. Выбор наборов данных

Так как оценка эстетичности – чисто субъективна, то необходимо подобрать такой датасет, по которому можно будет наиболее точно определить, является ли сгенерированный объект красивым. Так же стоит учесть, что эти объекты одного набора данных будет сравнить между собой чтобы понять “эстетическое” пространство и отличить, какой из двух является эстетичнее. Один из датасетов должен иметь количественную оценку красоты всех его объектов.

Таким образом, сформировав требования к наборам данных:

* Легкая интерпретация и оценка красоты существующих и сгенерированных объектов;
* Данные должны быть размечены на предмет их эстетической привлекательности;
* Возможность привести данные к одинаковому формату.

Исходя из этого, были выбраны следующие два датасета:

1. SCUT-FBP, содержащий в себе 500 фотографий лиц азиатских девушек разной размерности. Иллюстрация небольшого количества фотографий приведена на рисунке 2.



Рис. 2 Пример лиц из SCUT-FBP набора данных

1. В качестве набора, из которого мы будем генерировать новые объекты, выбран датасет [14], состоящий из 56443 различных шрифтов для каждого из 64 символов (английский алфавит с заглавными и строчными буквами и 10 цифр), размерность изображения 64\*64 в градациях серого. Для проведения исследования была выбрана цифра 5, так как по изгибу, наклону, наличию пропусков и обхвату пикселей это одна из наиболее сложных цифр. Иллюстрация небольшого количества шрифтов приведена на рисунке 3.



Рис. 3 Пример цифры 5 для 10 разных шрифтов

* 1. Подготовка данных

Так как смысл исследования заключается в вопросе, возможно ли на основе понятия красоты одного объекта перенести его на другой, то данные, которые будут подаваться в энкодер для кодирования в скрытое пространство нужно привести к общему виду. В нашем случае необходимо сжать все изображения лиц до размера и типа данных пятерок, то есть к виду n\*64\*64.

Так как цифры пять представлены в виде градаций серого, то было принято решение сделать лица монохромными.

Результатом данного преобразования получился датасет, проиллюстрированный на рисунке 4.

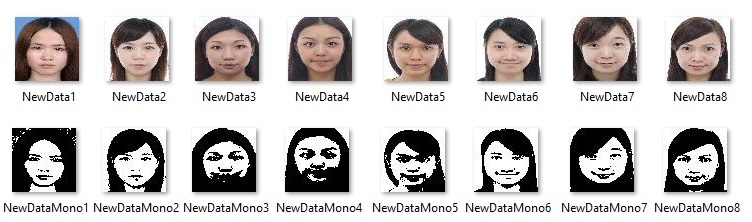


Рис. 4 Пример подготовленных данных

Глава 4. Исследование

4.1. Применение VAE к задаче генерации цифр

В качестве языка программирования для исследования был выбран Python [15]. Для реализации Вариационного автоэнкодера использовалась библиотека Keras [16].

Так как для исследования необходимо визуализировать результаты, а также вручную подбирать параметры вектора скрытого пространства, было принято решение установить размерность скрытого пространства равную 2. Из-за этого может наблюдаться очень размытые сгенерированные объекты, так как перевод из пространства 64\*64 в двумерное не может не сопровождаться потерями в качестве генерируемых изображений.

На рисунке 5 представлены результаты генерации цифр:



Рис. 5 Распределение пятерок в скрытом пространстве

4.2. Применение VAE к задаче генерации лиц

По своей сути, данный алгоритм ничем не отличается от применённого в предыдущем параграфе, но результатом генерации буду женские лица в градациях серого(рисунок 6)

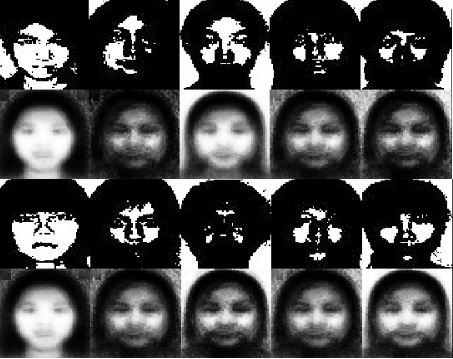


Рис. 6 Исходные и восстановленные женские лица

Как упоминалось в предыдущем параграфе, из-за слишком малой размерности скрытого пространства, изображения получается очень смазанными.

Более того, из-за очень маленькой выборки (500 лиц) не удастся очень точно обучить энкодер и декодер, но так как суть данной работы заключается в исследовании машинного понимания красивого, а не генерация новых лиц как таковых, то это не должно повлиять на результаты работы.

4.3. Исследование генерации красивых цифр

Идея заключается в том, чтобы подать красивое и некрасивое лицо в энкодер цифры 5 и попытаться обнаружить связь между эстетичностью лица и эстетичностью получающейся цифры. На рисунке 7 представлена цифра 5, “сгенерированная из лица” с оценкой красоты в 3.2 по шкале от 1 до 5.

На рисунках 8 и 9 представлены сгенерированные цифры из лиц, которые имеют весомое различие как по внешнему виду, так и по выставленной им оценке. Несмотря на это, уверенно говорить о наличии связи не совсем корректно, так как при переводе в единый формат данных теряется цветовой составляющая картинки, что может влиять на полученный результат. Исходя из полученных результатов можно сделать вывод, что сильнее всего на положение в пространстве сейчас играет роль количество пикселей, имеющих высокую яркость. Но если попытаться интерпретировать это в термины оценки человека, то чем светлее человек (ровный тон кожи, светлые глаза, отсутствие морщин) тем более тонкая и яркая пятерка будет возникать при генерации.

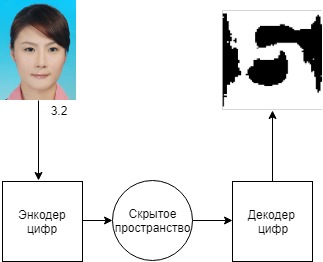


Рис. 7 Модель проведения эксперимента

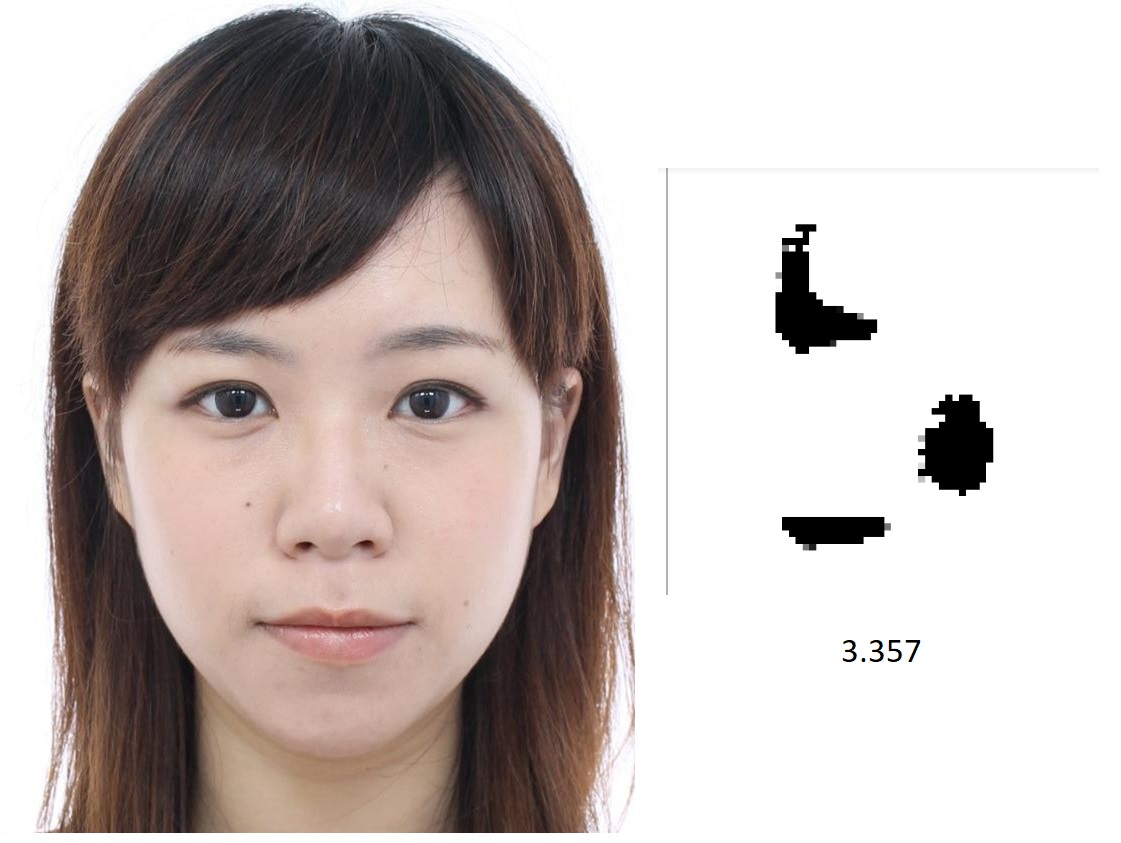


Рис.8 сгенерированная цифра для лица со скорингом 3.357



Рис.9 сгенерированная цифра для лица со скорингом 22

4.3. Объединение датасетов с использованием cVAE

cVAE – conditional VAE, это условный вариационный автокодировщик, почти ничем не отличающийся от обычного VAE, но в этом случае на вход подается еще и label набора данных.

В исследуемом случае label для цифр 5 будет 5, а для лиц – 1.

Данный подход должен разграничить наборы данных между собой, но с другой стороны, для двух наборов данных будет существовать 1 энкодер и 1 декодер.

Идея исследования почти идентичная: необходимо проверить на соответствие красоту подаваемого лица и сгенерированную цифру.



Рис. 10 Сгенерированная цифра из энкодера объединённых датасетов со скорингом 2.757

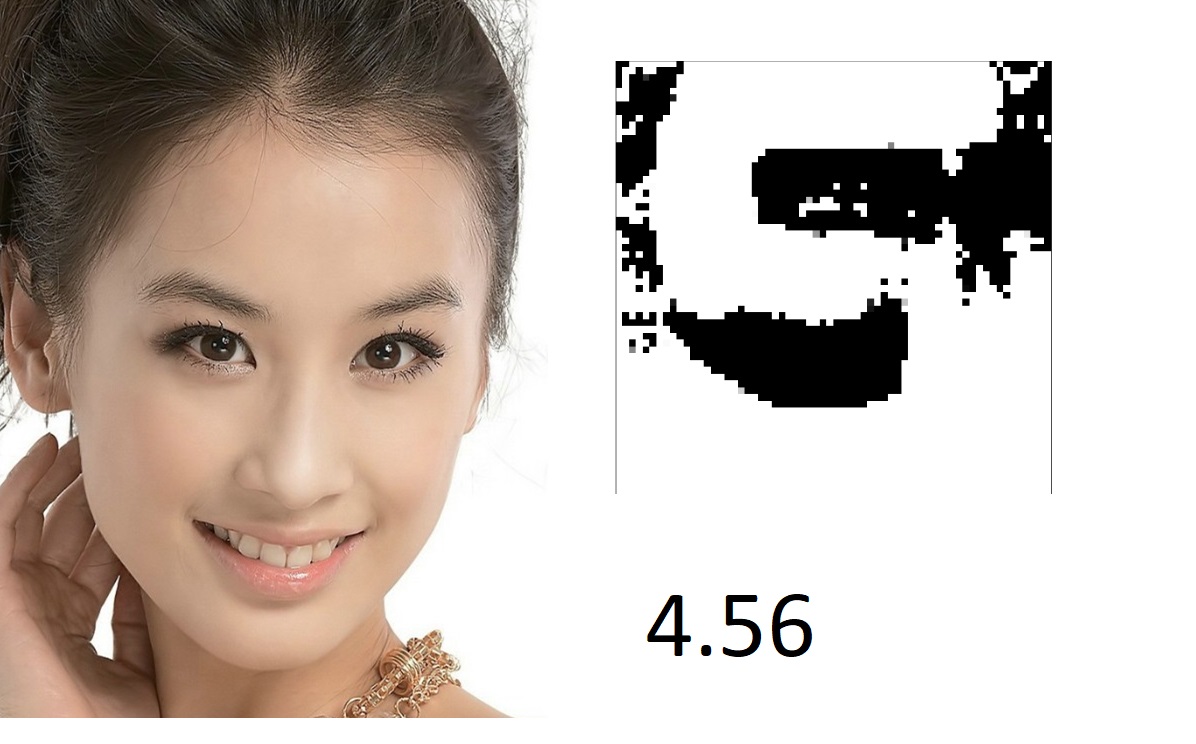


Рис. 11 Сгенерированная цифра из энкодера объединённых датасетов со скорингом 4.56



Рис. 12 Сгенерированная цифра из энкодера объединённых датасетов со скорингом 2.8

На рисунках 10, 11 и 12 предоставлены результаты работы вариационного автокодировщика с объединенными датасетами. К сожалению, из-за большой разнородности данных и недостаточного объема обучающей выборки для лиц почти невозможно дать интерпретацию .

Вывод

Несмотря на то, что результаты тяжело интерпретируемы, в дальнейшем исследовании и усовершенствовании метода возможно перенести понятие красоты с одного объекта на другой.

Вполне вероятно, что появление новых наборов данных и использование других подходов позволит в будущем разработать систему, которая самостоятельно будет генерировать полностью готовые эстетические формы.

Мое исследование показало, что идея оценки привлекательности объектов разной природы имеет все шансы на дальнейшее развитие.

Заключение

В данной работы были рассмотрены основные подходы к индустриальному дизайну, начиная от простейших методов, заканчивая алгоритмами, использующими машинное обучение. Так же был проведен анализ литературы, основной целью которых было оценивание эстетической составляющих объектов на изображении при заранее известных оценках на тренировочных данных.

В результате обзора области, было выявлено, что современные подходы к индустриальному дизайну не включают в себя генерацию красивых объектов какой-либо природы, либо опираются на заранее известные, составленные экспертами правила. Исходя из этого факта был подобран метод машинного обучения, наборы данных и проведено исследование, позволяющее понять, возможно ли переносить понятие красивого с одного объекта на другой в рамках подобранного алгоритма.

Список литературы

1. В целом об ИНС // URL: https://neuralnet.info/chapter/введение/#В-целом-об-ИНС (дата обращения: 30.05.2019).
2. Deepak K. Gupta, Marco Barink, Yulia Galagan, and Matthijs Langelaar Integrated Front–Rear-Grid Optimization of Free-Form Solar Cells // IEEE JOURNAL OF PHOTOVOLTAICS, VOL. 7, NO. 1, JANUARY 2017 294-302.
3. Deepak K.Gupta, Matthijs Langelaar, Marco Barink, Fred van Keulen  Optimizing front metallization patterns: Efficiency with aesthetics in free-form solar cells // [Renewable Energy](https://www.sciencedirect.com/science/journal/09601481) Volume 86, February 2016, Pages 1332-1339
4. Katharina Schwarz, Patrick Wieschollek, Hendrik P. A. Lensch Will People Like Your Image? Learning the Aesthetic Space // 4 Dec 2017.
5. Duorui Xie, Lingyu Liang, Lianwen Jin, Jie Xu, Mengru Li SCUT-FBP: A Benchmark Dataset for Facial Beauty Perception // 2015.
6. Douglas Gray, Kai Yu, Wei Xu, and Yihong Gong Predicting Facial Beauty without Landmarks // 2012.
7. Understanding Aesthetics with Deep Learning by Appu Shaji // URL: https://devblogs.nvidia.com/understanding-aesthetics-deep-learning/ (дата обращения: 30.05.2019).
8. Сейтказиева Назгул Cалбаровна, Ниязмаметова Сахинур Абликимовна Использование информационных технологий в процессе дизайн-проектирования // Academy. 2017. №10 (25). URL: https://cyberleninka.ru/article/n/ispolzovanie-informatsionnyh-tehnologiy-v-protsesse-dizayn-proektirovaniya (дата обращения: 30.05.2019).
9. Yoram Reich A model of aesthetic judgment in design // Artificial Intelligence in Engineering 8 (1993) 141-153
10. Ризаева А.Д. ГЕНЕРАТИВНЫЙ ДИЗАЙН: ПРОГРАММИРОВАНИЕ, КАК НОВЫЙ ИНСТРУМЕНТ ДЕЯТЕЛЬНОСТИ ДИЗАЙНЕРА // Материалы VII Международной студенческой научной конференции «Студенческий научный форум»
11. Naila Murray, Luca Marchesotti, Florent Perronnin AVA: A large-scale database for aesthetic visual analysis // 2012 IEEE Conference on Computer Vision and Pattern Recognition. 2012.
12. Carl Doersch Tutorial on Variational Autoencoders // 13 Aug 2016.
13. Автоэнкодеры в Keras, Часть 3: Вариационные автоэнкодеры (VAE) // URL: https://habr.com/ru/post/331552/ (дата обращения: 30.05.2019).
14. Analyzing 50k fonts using deep neural networks // URL: https://erikbern.com/2016/01/21/analyzing-50k-fonts-using-deep-neural-networks.html (дата обращения: 30.05.2019).
15. Python // URL: https://www.python.org/ (дата обращения: 30.05.2019).
16. Keras: The Python Deep Learning library // URL: https://keras.io/ (дата обращения: 30.05.2019).