

Разработка и исследование модели для генерации субъективно привлекательных лиц людей

М. А. Ким*, М. В. Хачумов*

** Кафедра математического моделирования и искусственного интеллекта,
Российский университет дружбы народов,
ул. Миклухо-Маклая, д.6, Москва, Россия, 117198*

Email: 1032201664@pfur.ru

Когнитивные способности человека позволяют осуществлять восприятие, анализ и внешнюю оценку собеседника. При этом объяснение субъективной оценки внешности нельзя назвать тривиальной задачей. Помимо известных общих закономерностей, таких как принципы геометрии лица, социальный и культурный опыт оценивающего, каждому человеку присущи комплексные индивидуальные факторы, влияющие на субъективное понимание красоты. Настоящая статья посвящена задаче построения модели машинного обучения, способной в соответствии с обратной связью от пользователя генерировать субъективно привлекательные лица людей. Предполагается, что такая модель поможет в изучении процесса формирования человеком оценки привлекательности лиц и выявлении индивидуальных факторов, влияющих на данный показатель. В работе рассмотрен сервис, предоставляющий пользователю последовательный выбор наиболее привлекательных изображений, созданных при помощи модели генеративно-состязательной сети (GAN). По окончании заданной длины последовательности изображений пользователю дается возможность оценить результирующие изображения на предмет привлекательности лиц по заданному показателю качества. Обучение нейронной сети производилось на датасете CelebA-HQ, который содержит 30000 1024×1024 изображений лиц знаменитостей. Для увеличения выборки пользователей сервис развернут в виде web-приложения и Telegram бота со свободным доступом, что позволяет собирать данные о сгенерированных результирующих изображениях. В статье даются выводы о качестве генерации изображений по мнению пользователей и о возможностях применения полученных результатов.

Ключевые слова: изображения лиц, визуализация, генеративные состязательные сети, машинное обучение.

1. Введение

Каким образом возможно оценить красоту? Человек способен мгновенно классифицировать изображение как привлекательное и даже выразить степень красоты заданного объекта в соответствии с определенной метрикой. Однако, в большинстве случаев, невозможно произвести количественную оценку избранного решения. Особенно остро данный аспект проявляется при анализе и генерации человеческих лиц.

Трудности исследования данного направления связаны в первую очередь с тем, что оценка привлекательности одного и того же объекта разными людьми может сильно различаться в виду, к примеру, возраста, культуры и пола оценивающего [1–3]. Безусловно, есть и общие закономерности, связывающие определенные черты лица и среднюю оценку его привлекательности, такие как его симметричность, приблизительно равная соотношению носа ко лбу и носа к подбородку [4]. Однако, в общем случае считается, что привлекательность – скорее субъективная, личностная характеристика, нежели объективный фактор, справедливый для большинства людей.

Создание визуально привлекательных изображений людей до сих пор является является сложной задачей в виду затруднительности описания концепции привлекательности. Большинство исследований в данной области направлены на выявление определенных закономерностей лица в соответствии с фотографией [5–7], однако, вероятно, подходы такого рода не позволяют объемлюще воспроизвести

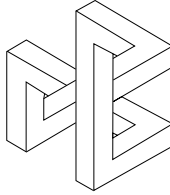


Рис. 1. Архитектура модели

факторы, влияющие на субъективную оценку в виду сложности человеческих суждений относительно привлекательности. Модели, создающие изображения на основе данных факторов, не способны полностью отразить эстетическое восприятие человека [8]. Это связано с тем, что генерация объектов ограничена заданными общими закономерностями и не учитывает субъективные предпочтения человека.

В данной статье представлен альтернативный метод созданию субъективно привлекательных изображений лиц людей на основе генеративно-состязательной нейронной сети (GAN) [9], обученной на датасете CelebA-HQ. Коррекция в генерации лиц в соответствии с предпочтениями человека производится по принципу исследования тензора скрытого состояния нейронной сети [10].

2. Построение модели

Предлагаемая архитектура модели представляет собой модификацию сверточной генеративно-состязательной сети глубокого обучения (DCGAN) [11]. DCGAN является прямым продолжением развития архитектуры GAN: модель также состоит из двух нейронных сетей, генератора и дискриминатора, обучаемых независимо друг от друга. Генератор создает изображения, основываясь на тензоре случайных чисел с плавающей точкой, генерируемых в соответствии с нормальным распределением. Дискриминатор дает оценку поступающим реальным и сгенерированным изображениям. Критерии качества для генератора и дискриминатора задаются согласно бинарной кроссэнтропии и представляются формулами (1), (2).

$$\text{loss_dis} = -\log(\text{real}) - \log(1 - \text{fake}) \quad (1)$$

$$\text{loss_gen} = -\log(\text{fake}), \quad (2)$$

где $\text{fake} \in [0, 1]$ — уверенность дискриминатора в генеративной природе изображения, $\text{real} \in [0, 1]$ — уверенность дискриминатора в том, что изображение является настоящим.

В представленных моделях генератор и дискриминатор соревнуются, что приводит к увеличению качества генерируемых изображений при увеличении числа эпох при условии обеспечения сходимости модели. Архитектура DCGAN изменяет классическую архитектуру GAN, добавляя сверточные слои Convolutional и Convolutional Transpose в генератор и дискриминатор. Данная модификация позволяет улучшить качество результирующих изображений при использовании множества скрытых слоев в нейронной сети.

Предлагаемая в статье модификация DCGAN содержит вектор скрытого состояния в генераторе, что позволяет создавать новые изображения на основе интерполяции векторов скрытого состояния целевых изображений, генерируя тем самым лица, совмещающие в себе черты целевых изображений. Архитектура решения представлена на рис. 1.

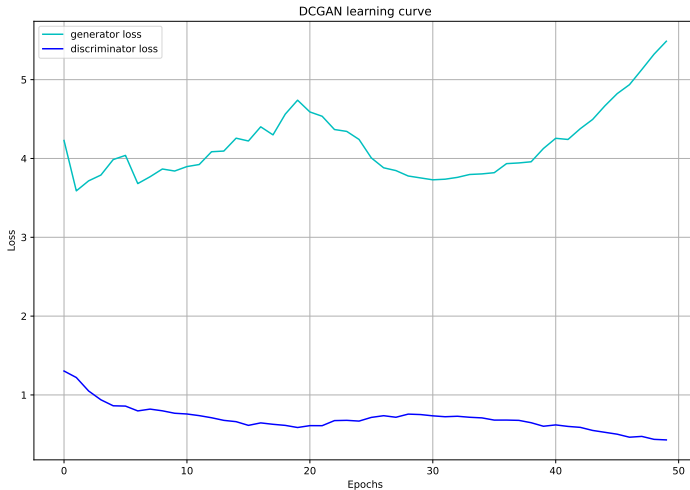


Рис. 2. Кривая обучения



Рис. 3. Примеры генерации лиц со случайным SEED

3. Обучение модели

Обучение модели производилось на датасете CelebA-HQ [12], который содержит 30000 изображений лиц знаменитостей с разрешением 1024×1024 px. Для обучения модели использовались исключительно изображения, представляющие из себя набор пикселей. Метаинформация, такая как именование различных признаков на изображении, не использовалась при обучении. Разрешение изображений было понижено до 128×128 px перед обучением. Обучение производилось на протяжении 50 эпох. Результирующая кривая обучения для генератора и дискриминатора представлена на рис. 2.

4. Внедрение и оценка модели

Примеры генерации лиц со случайным входным тензором показаны на рис. 3.



Рис. 4. Интерполяция векторов скрытого состояния

Результат исследования

Таблица 1

$\delta(\text{avg}(T), T^r)$	Число участников $n \in N$
≥ 2	2
1	9
0	3
-1	2
≤ -2	4

Пример интерполяции вектора скрытого состояния показан на рис. 4, где изображение слева получено путем интерполирования векторов скрытого состояния остальных изображений.

Гипотеза, выдвигаемая настоящей статьей, предполагает, что при выборе участником исследования нескольких субъективно привлекательных изображений и интерполировании их векторов скрытых состояний в новый тензор, при помощи данного тензора можно сгенерировать изображение, являющееся более привлекательным для участника исследования по сравнению с избранными ранее изображениями.

Для проверки гипотезы была выбрана стратегия, заключающаяся в последовательном показе набора из четырех случайно сгенерированных изображений три раза. В каждом наборе пользователю предлагалось выбрать наиболее привлекательное изображение и оценить степень его привлекательности по шкале от 1 до 10 (где 1 – наиболее отталкивающее возможное изображение, 10 – наиболее привлекательное возможное изображение). На основе векторов скрытых состояний трех полученных изображений строилось результирующее изображение, которое оценивалось пользователем по заданной выше метрике.

Для получения большего числа данных результирующая модель была внедрена в веб-сервис и telegram-бота, что позволило собрать статистику для подтверждения гипотезы. Сервисы были развернуты на серверах и предоставлены группе из 20 человек ($N = 20$), являющихся участниками исследования.

Результат эксперимента представлен в табл. 1, где $\text{avg}(T)$ – средняя оценка изображений, выбранных из наборов участником, T^r – оценка сгенерированного итогового изображения.

Результат исследования показывает, что 55% участников посчитали результирующее изображение более привлекательным, нежели выбранные из наборов. При этом 30% пользователей оценили итоговое изображение, как непривлекательное по сравнению с начальными.

При оценке результатов стоит брать во внимание тот факт, что построенная модель на базе DCGAN не способна генерировать реалистичные лица людей: пропорции не всегда соблюдены верно, изображения содержат артефакты. По данной

причине при проведении исследования его участниками отмечалось, что предоставленные изображения лиц не похожи на настоящей фотографии людей. В следствии этого, нельзя назвать выдвигаемую гипотезу доказанной, не смотря на тот факт, что исследование показывает определенный положительный результат.

Для последующего доказательства гипотезы следует использовать значительно более сложные предобученные модели машинного обучения по типу StyleGAN3 [13] и Stable Diffusion, что позволит генерировать действительно фотореалистичные изображения лиц людей, а также предоставит большую возможность в исследовании вектора скрытого состояния модели.

5. Заключение

В настоящей работе рассмотрен метод генерации субъективно привлекательных лиц людей путем интерполяции векторов скрытого состояния сверточной генеративно-состязательной сети глубокого обучения. Ранее данный метод не применялся в области создания человеческих лиц с использованием DCGAN. Дальнейшее исследование вектора скрытого состояния может помочь в выявление комплексных факторов, влияющих на субъективную оценку внешности человеком.

По результатам исследования можно сделать вывод в необходимости использования более сложных моделей нейронных сетей для решения поставленной проблемы в связи с неспособностью DCGAN генерировать фотореалистичные лица людей. Планируется проведение дальнейших исследований с использованием предобученных моделей StyleGAN3 и Stable Diffusion для улучшения качества генерации изображений.

Литература

1. M. R. Cunningham, A. R. Roberts, A. P. Barbee, P. B. Druen and C.-H. Wu. Their ideas of beauty are on the whole the same as ours. — Vol. 68, no. 2.
2. J. H. Langlois, J. M. Ritter, L. A. Roggman and L. S. Vaughn. Facial diversity and infant preferences for attractive faces. — Vol. 27, no. 1.
3. R. Thornhill and S. W. Gangestad. Facial attractiveness. — Vol. 3, no. 12. — P. 452-460.
4. K. Schmid, D. Marx and A. Samal. Computation of a face attractiveness index based on neoclassical canons, symmetry, and golden ratios. — Vol. 41, no. 8.
5. A. C. Little, B. C. Jones and L. M. DeBruine. Facial attractiveness: Evolutionary based research. — Vol. 366, no. 1571. — P. 1638-1659.
6. D. Perrett, K. A. May and S. Yoshikawa. Facial shape and judgements of female attractiveness. — Vol. 368. — P. 239-242.
7. J. Shi, A. Samal and D. Marx. How effective are landmarks and their geometry for face recognition?. — Vol. 102, no. 2. — P. 117-133.
8. M. Ibáñez-Berganza, A. Amico and V. Loreto. Subjectivity and complexity of facial attractiveness. — Vol. 9, no. 1. — P. 1-12.
9. I. Goodfellow et al. Generative adversarial nets. — P. 2672-2680.
10. Huiting Yang, Liangyu Chai, Qiang Wen, Shuang Zhao, Zixun Sun and Shengfeng He. Discovering interpretable latent space directions of gans beyond binary attributes. — Vol. 33.
11. A. Radford, L. Metz, S. Chintala, Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networ. — P. 31-38.
12. Tero Karras, Timo Aila, Samuli Laine, Jaakko Lehtinen [Электронный ресурс]. Papers With Code, 2021. См. URL: <https://paperswithcode.com/dataset/celeba-hq>
13. Tero Karras and Miika Aittala and Samuli Laine and Erik Härkönen and Janne Hellsten and Jaakko Lehtinen and Timo Aila [Электронный ресурс]. Alias-Free Generative Adversarial Networks (StyleGAN3)
14. Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, Björn Ommer. High-Resolution Image Synthesis with Latent Diffusion Models

Development and research of a model for generating subjectively attractive human faces

M. A. Kim*, M. V. Khachumov*

** Department of Information Technologies
Peoples' Friendship University of Russia
Miklukho-Maklaya str. 6, Moscow, 117198, Russia*

Email: 1032201664@pfur.ru

A person's cognitive capabilities allow for the perception, analysis, and external evaluation of others. At the same time, describing the subjective assessment of attractiveness cannot be considered a trivial task. In addition to well-established general patterns, including the principles of facial geometry and the social and cultural experiences of the evaluator, every person has complex, individual factors that influence their subjective understanding of attractiveness. This paper is focused on developing a machine learning model that can generate subjectively attractive human faces based on user feedback. It is hypothesized that such a model could be used to study the process of how a person forms their assessment of another's attractiveness and identify individual factors that contribute to this indicator. This study considers a service that offers users a consistent selection of the most appealing images generated using a generative adversarial network model. At the end of a specified length of sequence of images, users are given the opportunity to assess the resulting images based on the attractiveness of the faces using a given quality metric. The neural network has been trained on the CelebA-HQ dataset, which includes 30,000 images of celebrity faces with a resolution of 1024x1024 pixels. To increase the size of the user sample, the service has been deployed in the form of a web application and Telegram bot with free access. This allows users to collect data on the generated images. This article draws conclusions regarding the quality of the image generation based on user feedback and the potential applications of the results obtained.

Key words and phrases: face images, visualization, generative adversarial networks, machine learning.