Improving facial attraction in video

Maycon Prado Rocha Silva
State University of Campinas
Brazil, Campinas, SP
mayconrochasilva@gmail.com

José Mario De Martino
State University of Campinas
Brazil, Campinas, SP
martino@fee.unicamp.br

ABSTRACT

The face plays an important role both socially and culturally and has been extensively studied especially in investigations on perception. It is accepted that an attractive face tends to draw and keep the attention of the observer for a longer time. Drawing and keeping the attention is an important issue that can be beneficial in a variety of applications, including advertising, journalism, and education. In this article, we present a fully automated process to improve the attractiveness of faces in images and video. Our approach automatically identifies points of interest on the face and measures the distances between them, fusing the use of classifiers searches the database of reference face images deemed to be attractive to identify the pattern of points of interest more adequate to improve the attractiveness. The modified points of interest are projected in real-time onto a three-dimensional face mesh to support the consistent transformation of the face in a video sequence. In addition to the geometric transformation, texture is also automatically smoothed through a smoothing mask and weighted sum of textures. The process as a whole enables the improving of attractiveness not only in images but also in videos in real time.

Keywords

Video processing, computer vision, artificial intelligence, image processing, attractiveness improvement, texture processing, geometric transformation.

1. INTRODUCTION

Beauty today is desired by many individuals. Past studies[Sla00][Lan87] demonstrate that people who are considered more attractive usually tend to be perceived as more competent and, as a consequence, to be more favored in life than the less attractive ones. A beautiful person is naturally more attractive and holds the attention of a spectator more closely, favoring, for example, advertising, journalism and also teaching. Industry has explored the trend looking for ways to highlight and improve the attractiveness of people. In this context, the face has a high relevance and consequently has been extensively worked on to become more attractive, both by definite means and by computational corrections.

The contributions given by the area of computer graphics have played an important role in assisting facial enhancement, removing and smoothing faults or imperfections. Sophisticated applications have been widely used to remove skin irregularities and measure.

Researches [Let15][Ley08][Eis06][Guo09] on improving facial attractiveness show results very close to the natural. The main forms are through geometric processing, overlapping textures and smoothing masks. In these researches, all the techniques presented use aided processing, which depends on point of interest marking, or on image positioning and/or specific equipment. This makes the use more limited to a studio and can be applicable only to static images.

This article presents an approach for improving facial attractiveness in videos, where, in real time, face identification, mapping of points of interest, identification of a similar individual, geometric processing and texture smoothing occurs.

2. RELATED WORK

Leite and De Martino [Let15] present a study to improve facial attractiveness in two-dimensional front images using geometric transformation and texture smoothing. The results validated in the research indicate success in raising the attractiveness of processed samples. The mentioned limitations are: manual marking of points of interest of the face, neutral expression only and manual identification of models.

Another work presents an application launched in 2016 by an autonomous group of

Russian developers the FaceApp [Lon17] (Figure 1). Capable of modifying images in a dynamic way, such as: changing expression, rejuvenating, adding props, changing gender, changing hair, among others.

Figure 1. Images generated using FaceApp

Similarly, the Meitu application [Con19] (Figure 2) features functions that add effects that make photos more attractive, such as: smoothing the skin, facial reduction, and makeup application.

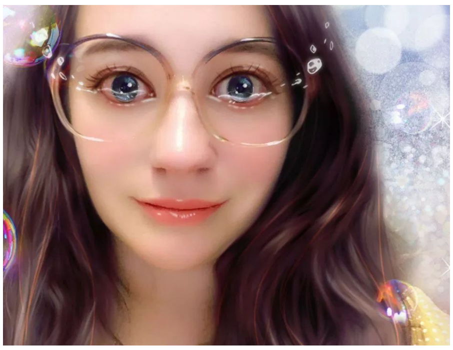


Figure 2. Image generated using Meitu

The developers of FaceApp and Meitu applications cite that machine learning and computer vision techniques have been used through free source libraries such as TensorFlow [Aba16] and OpenCV [Bran00].

The disadvantage of both FaceApp and Meitu is the application only in pre-defined models and the impossibility of extending the use in videos.

In 2017, using real-time video processing, Snapchat added in its mobile application a functionality capable of transforming face components into different configurations such as face change, zombie transformation and tracking of three-dimensional models that can be added according to the position of the face detected in the video.

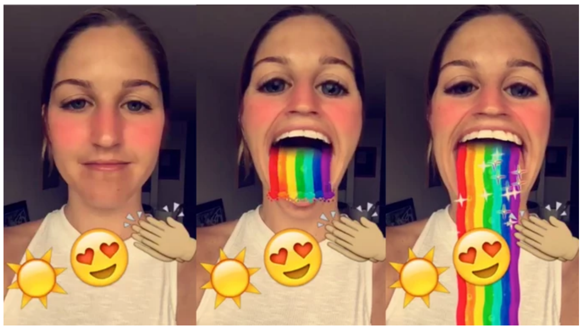


Figure 3. Images generated using Snapchat

Our work resembles the results of FaceApp and Snapchat [Cne17](Figure 3) applications in aspects of Face Detection, geometry transformation and processing of three-dimensional models . However, it is directed with the theme of improving attractiveness by adding techniques presented in Leite's [Let15] work. Expand than focus on video processing and automatic detection of point of interest.

[Fan2016] present a method using a default mesh based on the local chin shape that has great effect on facial attractiveness, and [Cho2012] replace facial features dynamically on videos, these two approaches, different this work, uses different model templates not based on common distances.

[XI2016] present the temporal-spatial-smooth warping (TSSW) method to achieve a high temporal coherence for video face editing. TSSW is based on two observations: (1) the control lattices are critical for generating warping surfaces and achieving the temporal coherence between consecutive video frames, and (2) the temporal coherence and spatial smoothness of the control lattices can be simultaneously and effectively preserved. The framework proposed can be used to improve the work presented in this article combined with the model similarity detection .

[Zha2017] and [Dia2019] combine different ways to make facial beautiful in static images with expressive results using models with no similar facial features distances.

3. PROPOSED SOLUTION

Tatiane and De Martino [Let15] presents a study to improve facial attractiveness in front images in two dimensions. The results were presented in obtaining a more attractive face exploring warping technique. However, some significant points are highlighted for continuity of new works, as follows:

- Automatic detection of points of interest.

- Projection of points of interest in three-dimensional coordinates.

- Video processing;

These points served as an inspiration for the work presented in this article, and the evolution made addresses continuity. That is, the results extend the technique presented by [Let15] in complement with the automatic detection of points in a static image and three-dimensional projection. The process covers the concepts of automatic detection of points of interest as well as the real-time identification of a basic model for transformation. The methodology is organized in stages, according to Figure 4.

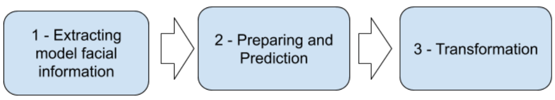


Figure 4. The stages of the method

3.1 Model facial information

In this stage a reference base is formed with information from individuals considered attractive. As the final result of the stage, a database of information of characteristics of attractive individuals is formed. The experiment used a list of 200 people known as the 100 most attractive women and the 100 most attractive men judged by the general public in 2014 by Maxim magazine [Max14]. The list was used for the construction of a base of three-dimensional points of interest meshes, where each model had its distances from its points extracted and stored for later use. Figure 5 present this stage overall.

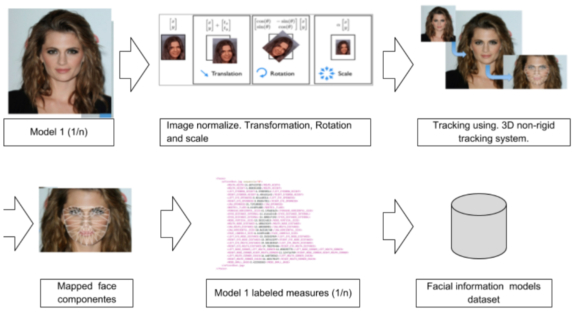


Figure 5. Extracting facial information from a face mode

For each individual, a set of images are collected in different positions and facial expressions. The first iteration with the images is to keep the eyes in a horizontal alignment rearranging using transformation and scale matrix as Figure 6.

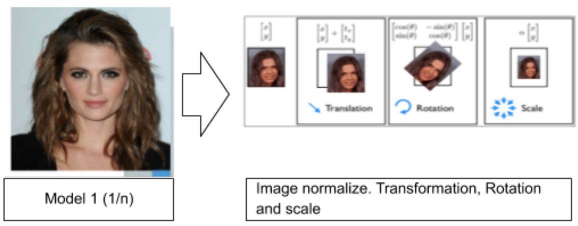
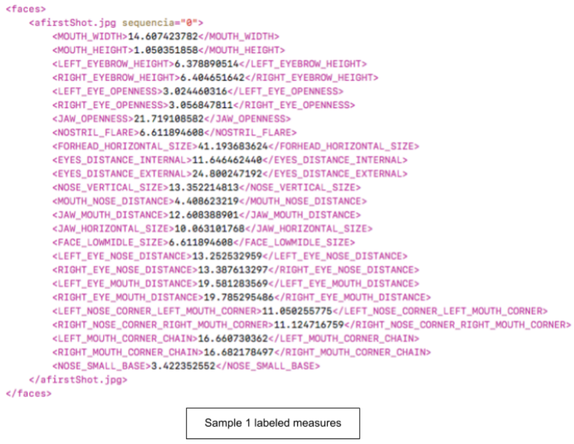


Figure 6. Image normalize process

In the second step, techniques of face Detection and extraction of points and distances are applied for each image to obtain characteristics of the individual. To extract the points of interest in two dimensions, the Face Tracker, developed by Sarigh, was used, based on the Haar-like pattern recognition process [Sar12], (see Figure 7). The Posit [Sar12] technique was used to estimate the points in three dimensions.

ISBN 978-80-86943-38-1



```
<faces>
    <afirstShot.jpg sequencia="0">
        <MOUTH_WIDTH>14.607423782</MOUTH_WIDTH>
        <MOUTH_HEIGHT>1.050351858</MOUTH_HEIGHT>
        <LEFT_EYEBROW_HEIGHT>6.378890514</LEFT_EYEBROW_HEIGHT>
        <RIGHT_EYEBROW_HEIGHT>6.404651642</RIGHT_EYEBROW_HEIGHT>
        <LEFT_EYE_OPENNESS>3.024460316</LEFT_EYE_OPENNESS>
        <RIGHT_EYE_OPENNESS>3.056847811</RIGHT_EYE_OPENNESS>
        <JAW_OPENNESS>21.719108582</JAW_OPENNESS>
        <NOSTRIL_FLARE>6.611894608</NOSTRIL_FLARE>
        <FORHEAD_HORIZONTAL_SIZE>41.193683624</FORHEAD_HORIZONTAL_SIZE>
        <EYES_DISTANCE_INTERNAL>11.646462440</EYES_DISTANCE_INTERNAL>
        <EYES_DISTANCE_EXTERNAL>24.800247192</EYES_DISTANCE_EXTERNAL>
        <NOSE_VERTICAL_SIZE>13.352214813</NOSE_VERTICAL_SIZE>
        <MOUTH_NOSE_DISTANCE>4.408623219</MOUTH_NOSE_DISTANCE>
        <JAW_MOUTH_DISTANCE>12.608388901</JAW_MOUTH_DISTANCE>
        <JAW_HORIZONTAL_SIZE>10.063101768</JAW_HORIZONTAL_SIZE>
        <FACE_LOWMIDLE_SIZE>6.611894608</FACE_LOWMIDLE_SIZE>
        <LEFT_EYE_NOSE_DISTANCE>13.252532959</LEFT_EYE_NOSE_DISTANCE>
        <RIGHT_EYE_NOSE_DISTANCE>13.387613297</RIGHT_EYE_NOSE_DISTANCE>
        <LEFT_EYE_MOUTH_DISTANCE>19.581283569</LEFT_EYE_MOUTH_DISTANCE>
        <RIGHT_EYE_MOUTH_DISTANCE>19.785295486</RIGHT_EYE_MOUTH_DISTANCE>
        <LEFT_NOSE_CORNER_LEFT_MOUTH_CORNER>11.050255775</LEFT_NOSE_CORNER_LEFT_MOUTH_CORNER>
        <RIGHT_NOSE_CORNER_RIGHT_MOUTH_CORNER>11.124716759</RIGHT_NOSE_CORNER_RIGHT_MOUTH_CORNER>
        <LEFT_MOUTH_CORNER_CHAIN>16.660730362</LEFT_MOUTH_CORNER_CHAIN>
        <RIGHT_MOUTH_CORNER_CHAIN>16.682178497</RIGHT_MOUTH_CORNER_CHAIN>
        <NOSE_SMALL_BASE>3.422352552</NOSE_SMALL_BASE>
    </afirstShot.jpg>
</faces>
```


Figure 7. Labeled measures sample

The points of interest found, (see Figure 8) as example model 1 with maped face components and labeled, were triangulated according to the application of The Delaunay algorithm described by [Gui92] , thus constituting a mesh of points of interest.

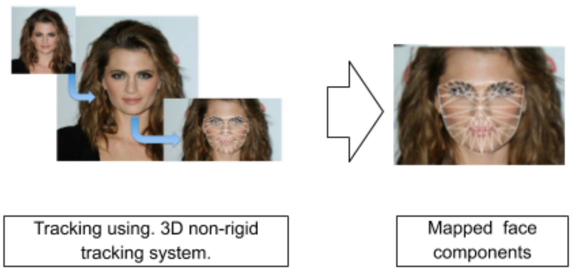


Figure 8. Mapped face components

Each extracted meshes dimensions are stored for use in the later stages.

3.2 Preparing and Prediction

The goal of this stage is to identify a similarity of any individual to some other of the reference basis of individuals considered attractive. In a video with some individual, sample images are extracted to identify the characteristics. For each image, techniques of face detection and extraction of points and distances are applied to obtain characteristics of the individual. The characteristics of the individual are compared with the characteristics of each individual which form the reference basis for information on characteristics of attractive individuals. The objective of the comparison is to identify the set of characteristics that most closely resembles the individual who wishes to apply the improvement of attractiveness. At the end of the stage a more similar and attractive face model is identified for the individual concerned.

In our implementation (Figure 9), the processing for comparison of information was done using KNN (K Nearest Neighbor) and SVM (Support Vector Machines) [Zha2006] data classifiers. The classifier's input for training is the file of meshes extracted from the model facial information stage.

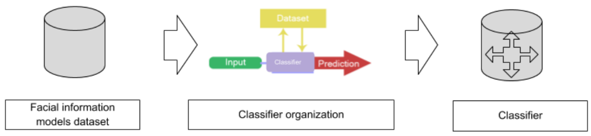


Figure 9. Classifier organization

The prediction (Figure 10) aims to answer which model most resembles a mesh of points of interest from any sample. For this purpose, the classifier has as input the points and distances of the face, and as output the indication of the individual that has the most similarity with the input face. The classifier has been configured in advance using the basis of attractive features.

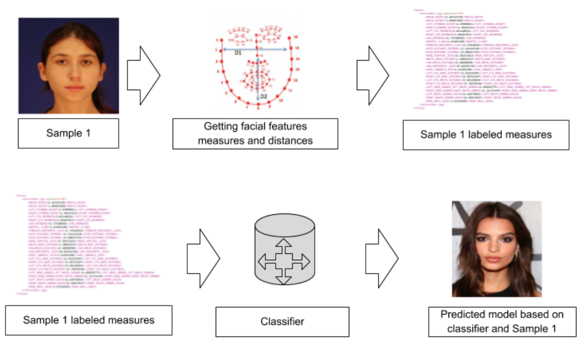


Figure 10. Prediction process

3.3. Transformation

Once the set of characteristics with greatest attractiveness and similarity with the face to be enhanced is identified, a geometric transformation is applied to all the video frames whose face is to be enhanced. As a result (see Figure 11 below), the individual has the same size and shape of components of the face (mouth, nose and eyes) according to the face model with more similar and attractive features. In our implementation, we applied the technique of warping points in image and blending textures, based on the work [Guo09],

ISBN 978-80-86943-38-1

[Lon17], [Con19] and [Cne17] that use so-called points of control to perform the distortion of the images, which were transformed to the points of the individual according to the model characteristics more similar and attractive identified.

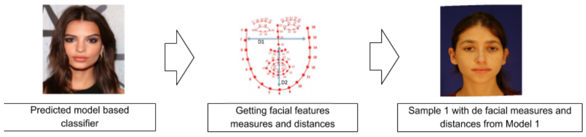


Figure 11. Geometric transformation process

For gaining scale in frame-by-frame processing of video and combine textures the process uses shape extraction based on the mesh of points of interest. The technique is based on the work of Saragih [Sar12] the process was evolved to perform the extraction of a the 3d mesh from a face identified in a video. Once the face is detected, it can be observed that its edges form a geometric shape around the mesh, it is made through the connection of the points and the texturing extracted from the frame in the delimited area. The geometric shape with the applied texture is called mesh (see Figure 12).

Figure 12. Mesh extract

4. EXPERIMENTAL RESULTS

Some results of our approach are presented in Table 1 which shows original videos in second column and processed videos in the third column.

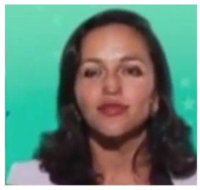
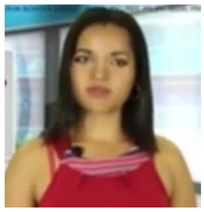
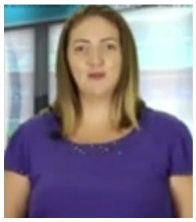
Model	**Unmodified video**	**Modified video**
Model 1	 https://youtu.be/H3OYtewsxoc	 https://youtu.be/dG9DvDjQAzI
Model 2	 https://youtu.be/h6aeP3WcW9E	 https://youtu.be/HWLe9_hqdQE
Model 3	 https://youtu.be/sHW49lc9-QU	 https://youtu.be/a-9upUQubQM
Model 4	 https://youtu.be/Ws15_nvSkGM	 https://youtu.be/izeIvMX7S6A
Model 5	 https://youtu.be/H7Fsuu_ExDs	 https://youtu.be/_mWUwarekvQ

Table 1 – unmodified x modified vídeos samples and links.

ISBN 978-80-86943-38-1

The videos was produced using a mac book pro late 2011, using 8gb of memory and a intel core i5 processor. The time of process to each frame 640x360 resolution took in average 3,1s.

An assessment was made to identify whether there was a gain in attractiveness in the processed videos.

The evaluation process was done through a question form and applied individually to a group of people in a working environment. The selection of people was random and no group received special attention. The applied form was online and the answers stored in a database. The questions contained in the form were divided into two parts, being the first to identify gender and age, and the second to collect attractiveness notes given individually for a set of demonstrated videos.

The environment used was a closed room to preserve silence and avoid any external distraction. The site was equipped with a projection screen and artificial light. In this environment the total of responses collected was 96.

The application of the form was conducted by a host who followed the same dynamic, addressing person to person, inviting to participate in the evaluation process. Each guest evaluator was directed by the host to the room prepared for the application of the form. The host was responsible for narrating each question and collecting the answer. Below are the first two questions related to the first part of the questionnaire:

-What is your age range?

-What is your gender?

The subsequent questions were directed by the presenter as follows: "watch the video and answer the question below".

A total of five videos were presented, and for each, the host asked: "As for the person seen in video 1, on a scale of 1 to 7 being 1 for an unattractive person and 7 for a very attractive person, classify:" the videos presented contained a shot of a random person expressing himself in speech to the camera. The total amount of videos used throughout the process were 10, 5 being processed by the attractiveness improvement technique and the other 5 not processed. The 10 videos were distributed in 2 forms, namely A and B:

of the 5 videos of form A, 3 were processed by the attractiveness improvement technique and the 2 others were not. Form B was assembled with the same videos; however, processing was applied to the 2 videos that did not receive processing on Form A.

The application forms to applicants were alternated. For half of the group of applicants form a was submitted and for the other half Form B. The answers were gathered in a single response base and the results presented in Table 2 and in the graphic of figure 13.

Model	**Sum of notes when modified video**	**Sum of notes when unmodified video**
Model 1	137	95
Model 2	140	125
Model 3	127	11
Model 4	115	110
Model 5	140	104

Table 2 – Sum of notes assigned modified videos x not modified

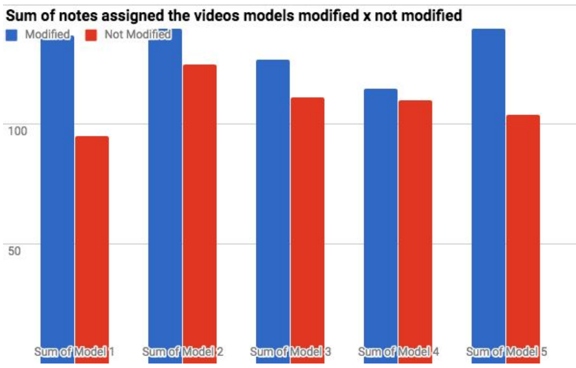


Figure 13. Graph result of the evaluation

5. CONCLUSION

The study presented demonstrates the use of computer vision and image transformation in video processing in favor of improving facial attractiveness. The automatic detection of points of interest along with the projection of three-dimensional forms provided the advance. Limitations found can be evolved using complementary technologies. The inaccuracy of the detection of facial points automatically can be combined with the use of three-dimensional cameras for better measurement avoiding flickers in de detection and inappropriate delimitation of features that cause a unexpected morphing transformation. Detection of mood and images of models in different expressions can also contribute to a better result.

 ISBN 978-80-86943-38-1

Increasing the sample base of models with different face patterns can smooth out distortions given a face pattern incompatibility between base model and target model.

6. REFERENCES

[Let15] Leite, T. S.; De Martino, J. M. Improving the attractiveness of faces in images. Proceedings of the 28th Conference on Graphics, Patterns and Images - Sibgrapi 2015, Salvador, Bahia, Brazil,.

[Max14] Maxim Media. In: REVISTA. 2014. Available at: <http://www.maxim.com/women/2014-hot-100-post>. Access: 26 jun. 2018.

[Ley08] Leyand, Tommer et al. Data-driven enhancement of facial attractiveness.ACM Transactions on Graphics (TOG), v. 27, n. 3, p. 38, 2008.

[Eis06] Eisenthal, Yael; DROR, Gideon; RUPPIN, Eytan. Facial attractiveness: Beauty and the machine. Neural Computation, v. 18, n. 1, p. 119-142, 2006.

[Guo09] Guo, Dong; SIM, Terence. Digital face makeup by example. In: Computer Vision and Pattern Recognition, CVPR 2009. IEEE Conference on. IEEE, p. 73-79. 2009.

[Lon17] Lomas, Natasha et al. FaceApp uses neural networks for photorealistic selfie tweaks. 2017. Available at: <https://techcrunch.com/2017/02/08/faceapp-uses-neural-networks-for-photorealist

[Con19] Conger, Kate et al. The cost of hot selfie app Meitu? A healthy dose of your personal info. 2017. Available at: <https://techcrunch.com/2017/01/19/meitu-app-collects-personal-data/>. Access: 10 mar. 2019.

[Cne17] Constine, Josh In: Now Snapchat has "Filter Games". 2017. Available at: <https://techcrunch.com/2016/12/23/snapchat-games/>. Access: 10 mar. 2019.

[Bran00] Bradski, Gary. "The opencv library." Dr Dobb's J. Software Tools 25 (2000): 120-125.

[Aba16] Abadi, Martín, et al. "Tensorflow: A system for large-scale machine learning." 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16). 2016.

[Sar12] Saragih, J., et al. Non-Rigid Face Tracking. Mastering OpenCV with Practical Computer Vision Projects, p. 189-233, 2012.

[Sla00] Slater, A., Bremner, G., Johnson, S. P., Sherwood, P., Hayes, R., & Brown, E. Newborn infants' preference for attractive faces: The role of internal and external facial features. Infancy, 1(2), 265-274, 2000.

[Lan87] Langlois, Judith H. et al. Infant preferences for attractive faces: Rudiments of a stereotype?. Developmental psychology, v. 23, n. 3, p. 363, 1987.

[Dia2019]Diamant, N., Zadok, D., Baskin, C., Schwartz, E., & Bronstein, A. M. Beholder-GAN: Generation and Beautification of Facial Images with Conditioning on Their Beauty Level. 2019.

[Gui92]Guibas, L. J., Knuth, D. E., & Sharir, M, Randomized incremental construction of Delaunay and Voronoi diagrams. Algorithmica, 7(1-6), p. 381-413, 1992.

[Fan2016] Fan, X., Chai, Z., Feng, Y., Wang, Y., Wang, S., & Luo, Z.. An efficient mesh-based face beautifier on mobile devices. Neurocomputing, 172, 134-142. 2016.

[Cho2012] Chou, J. K., Yang, C. K., & Gong, S. D.. Face-off: automatic alteration of facial features. Multimedia Tools and Applications, 56(3), 569-596. 2012.

[Dia2019] Diamant, N., Zadok, D., Baskin, C., Schwartz, E., & Bronstein, A. M.. Beholder-GAN: Generation and Beautification of Facial Images with Conditioning on Their Beauty Level. 2019.

[Zha2017]Zhang, L., Zhang, D., Sun, M. M., & Chen, F. M. (2017). Facial beauty analysis based on geometric feature: Toward attractiveness assessment application. Expert Systems with Applications, 82, 252-265. 2017.

[Zha2006] Zhang, H., Berg, A. C., Maire, M., & Malik, J. (2006). SVM-KNN: Discriminative nearest neighbor classification for visual category recognition. In 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06) (Vol. 2, pp. 2126-2136). IEEE.2006

 ISBN 978-80-86943-38-1