Journal of Sociology and Ethnology (2021)
Clausius Scientific Press, Canada

DOI: 10.23977/jsoce.2021.030503
ISSN 2616-2318 Vol. 3 Num. 5

Research on Face Attribute Editing Method Based on Deep Neural Network

Caoshuai Kang

School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China

Keywords: Face attribute editing, Deep neural network, Deep learning, Transfer learning

Abstract: With the development of computer vision technology, the application of face attributes editing has been expanded upon the real scene. Face attribute editing aims to change one or more attributes of an image, such as hair color, skin color, age, etc., and keep other attributes unchanged. The key of attribute editing is to maintain the high quality and accuracy of the target attribute image. Most methods focus on editing with or without certain attributes, and lack of adding or deleting attributes for specific templates, such as bangs with specific shapes. On the other hand, when multi-attribute is edited at the same time, the existing methods are difficult to decouple multi-attribute in feature space, which leads to problems in multi-attribute editing mode There are many defects, such as artifact, face irrelevant deformation, background change and so on. Aiming at these problems, this paper studies the face attribute editing method based on deep neural network.

1. Introduction

As an intuitive semantic feature, face attributes describe the visual features of facial images that human beings can understand, such as hair color, bangs and glasses. As an important feature of face, face attribute has its unique role and contribution in many real application scenarios, such as face verification [1], face detection , face recognition [2] and facial expression recognition. Before the advent of deep learning, face attribute editing algorithms were mainly divided into physical model-based reasoning methods and face prototype based evolution methods. The physical model reasoning method uses real face data to model, trying to find the physical change law of face. At present, the existing face attribute editing framework mainly relies on the generation model, including the generation of confrontation network [3] and self encoder. In the face attribute editing algorithm, template attributes editing is an important application part, which is the necessity of more accurate and personalized face customization. It is not easy to solve the problem of face visual attributes recognition, especially in uncontrollable scenes, we need to solve the correlation and difference between attributes, as well as the influence of illumination, posture, occlusion and other external factors[4].

2. Joint Training of Face Attributes and Face Recognition

2.1 Face Attribute Analysis

For the analysis of face attribute categories, there are many classification methods of different angles, and attribute tags are often very useful. In general, it can be divided into global attribute and local attribute, as well as internal attribute and external attribute. Global attributes: describe the features of the whole face, such as age, gender, skin color, race, etc; Local attribute: it describes the local features of human face, mainly including features of facial features, such as thick lips and lees Nose, etc; Internal attributes: describe the biological characteristics of human face, which are invariant in different environments, such as gender and facial features. These internal attribute features are closely related to identity, which can provide some auxiliary information on face recognition and can be used for joint learning. External attribute: describes the additional information on human face in wearing, such as wearing glasses, necklaces and so on. The information is not necessarily related to the identity information on human face. An ideal face recognition system needs to be robust to these external attributes.

2.2 Multi Label Representation of Face Attributes

In face multi-attribute recognition, the corresponding labels of training images are often integrated into multi label forms. The common expression forms include label cod, multi label regression and joint multi classification. Tag cod is to encode the possible combination of multiple attributes (for example, the Asian male high nose is defined as 000, the Asian male short nose is defined as 001, etc.) to transform the multi-attribute problem with a single attribute multi classification problem after cod[5]. The significant disadvantage of label cod is that the number of samples of each cod category is very small, and it is difficult to train neural network. In addition, label cod does not consider the relationship between multiple attributes, so it is not used frequently.

2.3 Attribute Recognition Based on Multi Objective Learning

The main difficulties of using multi-objective learning of attribute recognition as additional information about multi task supervision training for face recognition. The data sets with both attribute annotation and identity annotation are very rare, so it is necessary to transfer the attribute and identity information. For some attributes, there is the problem of unbalanced distribution of training samples, the distribution of training samples is unbalanced

3. Face Generation Based on Attribute Information

3.1 Common Methods of Face Image Generation

After the rise of Generation Adversarial Network (GAN) in academia, generating more realistic face images has become a hot issue. The methods of editing face attributes to generate face images can be divided into two categories: methods based on data migration and methods based on automatic encoding and decoding. The method based on data migration divides the face image with a certain attribute and the image without it into two categories. The expression information on the two categories is transferred to the real image addition and removal of existing attributes. The method based on automatic encoding and decoding is to learn the depth cod of face image by encoder, and then use the given attribute features as conditions to generate the corresponding attribute face image by decoder.

3.2 Pixel Level Reconstruction Constraints

In the training process, in order to obtain better representation of the generated image at both the pixel level and the feature level, there are four constraints: (1) the generated image and the input image are as close as possible at the pixel level under the same attribute conditions; (2) the generated image and the input image are as close as possible under the same attribute conditions, Feature level as much as possible; (3) the generated image with attribute information should meet the condition of attribute classification; (4)the generated image with attribute information should be as close as possible to the real image.

4. Evaluating Indicator

4.1 Subjective Quality Evaluation

The purpose of face attributes editing task is to provide a good visual effect on changing most of the details of the input image. Therefore, subjective quality evaluation is the most reliable and direct way. Subjective quality evaluation includes the analysis of image fidelity, naturalness, intelligibility and other dimensions. Limited by the high dimension of the information contained in the image, it is difficult to capture all the details of the image simply relying on the fixed mathematical model and indicators.

Human evaluation indicators can capture the gap between the quality of image generation and give the comparison of advantages and disadvantages by virtue of the high intelligence of human brain. Generally, subjective quality evaluation uses average opinion score to judge image quality.

4.2 Objective Quality Evaluation

The quantitative quality evaluation is predicted by a fixed mathematical model, and the quantitative result is not affected by the difference between input image. In order to get a meaningful score of the evaluation index of generative countermeasure network, some necessary conditions are needed.

According to the FID index, the vectors in the same data field are all subject to a specific distribution, the real image data set and the generated image data set have different distributions, and the distribution has multi-dimensional variables. The goal of generating countermeasure network is to constrain the two distributions to be the same as much as possible.

5. Conclusion

In this paper, some problems of the current face attribute editing method are discussed and studied. Face attribute editing technology can edit the specific attribute area of face as accurately as possible. With the development of deep learning technology in recent years, The face based on deep learning and generating confrontation network still needs editing technology. Through intelligent operation, it avoids the tedious operation of many artists in the past image editing, saves a lot of labor costs, and has broad application prospects.

References

[1] Y. Taigman, M. Yang, M. Ranzato and L. Wolf(2014). DeepFace: Closing the Gap to Human-Level Performance in Face Verification. 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 2014, pp:1701-1708.

[2] F. Schroff, D. Kalenichenko and J. Philbin(2015). FaceNet: A unified embedding for face recognition and clustering. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, pp:815-823.

[3] Goodfellow I, Pouget-Abadie J, Mirza M(2014). Generative adversarial nets. Advances in neural information processing systems[C]. 2014: 2672-2680.
[4] Zhu J Y, Park T, Isola P, et al(2014). Unpaired image-to-image translation using cycle-consistent adversarial networks. Proceedings of the IEEE international conference on computer vision[C]. 2017: 2223-2232.
[5] Wu P W, Lin Y J, Chang C H(2019). Relgan: Multi-domain image-to-image translation via relative attributes. Proceedings of the IEEE International Conference on Computer Vision[C]. 2019: 5914-5922.