Forschungsarbeit

Titel: Mask-Guided Portrait Editing with Conditional GANs

Student: Prabhakar Kumar Panday, 3437515

Studiengang: Master's in Electrical Engineering
Prüfer: Prof. Dr.-Ing. Bin Yang
Betreuer: Mr. George Eskandar
Beginn: 18.10.2021

Portrait editing is a popular subject in photo manipulation, it is of great interest in the vision and graphics community due to its potential applications in movies, gaming, and others. Generative Adversarial Networks (GANs) [1] have made tremendous progress in synthesizing realistic faces [2], face aging [3], pose changing and attribute modification [4]. The three widely used approaches for portrait editing are: label-conditioned method [12], reference-guided method [13], and geometry-guided method [9]. This project will focus on geometry-guided technique using semantic facial mask as a shape guide for high-level facial component editing.

The motivation for this research project work is to use the facial mask of the reference image to guide image generation using conditional GANs. A face mask provides a good geometric constraint for facial features like nose, eyes, lips, skin, and hair; which helps in synthesizing realistic faces. Works based on face masks [5,6,7] achieve promising results, but they do not synthesize different faces, suffer from quality issues, like lack of fine details in skin, difficulty in dealing with hair, background blurring, and their diversity is limited to color or illumination. The goal of this research project is to build a framework based on conditional GANs [8] for portrait editing. This framework will be used to edit the facial features of a target image driven by a reference facial mask.

In a recent paper by, Pernuš et al. [11] the authors use facial masks to edit targeted regions of the images along with latent code optimization on GANs to generate high-resolution faces. The methodology used here fetches results in par with the state-of-the-art techniques and enables controlling size of individual facial components. However, this method edits the existing facial attributes and not fuse multiple faces together. Another technique SC-FEGAN, Jo et al. [14], uses free form face masks for portrait editing, requires very precise sketch and color as input. R-FACE [13] technique, uses image painting model as reference for controlling the structure of the face component.

This project will continue work based on the paper by Gu et al. [9], where the authors use a local embedding sub-network to learn five features of the reference facial image by using Variational Autoencoder (VAE) and then combine the learned component feature embedding and target facial image mask using a mask guided sub-network to generate the foreground face image. The authors are able to successfully synthesize diverse, high-quality, and controllable facial images from given masks, but the features of the output images are dominated by the target facial mask i.e., this method is not good at manipulating the target facial mask. To illustrate this shortcoming assume a target image with a narrow nose, if the reference face has a bigger nose, when copying it on the target image, the size of the nose in the output image is squeezed and the image appears unnatural. In a bid to solve this issue, in this research work each facial feature of the target image mask will be encoded and trained individually, which in-turn will allow combining multiple masks from both target and reference image.

The dataset to be used for training the model will be CelebA-HQ Dataset [10] which consists of 30,000 face images of celebs with their face masks which has 19 distinct facial features. In the first part of the project, a model will be built with only five facial features and later the number of facial features will be increased in steps based on the performance of the model.

The model will have four inputs, a source image x^s, a mask of the source image m^s, a target image x^t, and a mask of the target image m^t. A local embedding sub-network will be used to learn feature embedding for the input source image x^s using one VAE for each facial feature to encode embedding information for facial components. The mask guided generative sub-network then specifies the region of each embedded component feature and concatenates all features of the local components together with the target mask to generate the foreground face. Finally, the background fusing sub-network fuse the foreground face and the background to generate a natural facial image.

Once the framework is built, it will be trained on the dataset discussed in the previous section. For efficient training of the model, various loss functions will be implemented. The Local reconstruction loss (L_{local}) is defined as the MSE between the input and reconstructed instances to learn the feature embedding. The Global reconstruction loss (L_{global}) is the main loss function to compare the source image x^s with reconstructed image G.

Face parsing loss L_{GP} is used to generate sample face masks same as target mask, thus generating equivariant facial images. A pre-trained face parsing network, P_F is used to encourage the generated samples to have the same mask with the target mask.

$$L_{GP} = -E_{(x \sim Pr)}[\sum_{i,j} \log P(m^t_{i,j} | P_F(G(x^s, m^s, x^t, m^t))_i, j)]$$

Equation 1: Face parsing loss, where $m^t_{i,j}$ is the ground truth label of x^t located at (i,j). $P_F(G(x^s,m^s,x^t,m^t))_{i,j}$ is the predict pixel located at (i,j).

Overall loss function is the sum of individual loss functions [Equation 1].

$$L_{global} = \lambda_{local} . L_{local} + \lambda_{global} . L_{global} + \lambda_{GD} . L_{GD} + \lambda_{GP} . L_{GP}$$

Equation 2: Global Loss Function

The following section defines specific tasks to accomplish the above defined goals.

T1 Literature Research

T2 Build the Framework as in the research paper [9] and then improve it by adding face components to the mask-guided generative sub-network one by one.

T3 Training and testing the model on new training dataset

T4 Optimization of model and testing the performance

T5 Improve the model for new face components

T6 Evaluate Results

T7 Write Thesis

References:

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In Advances in neural information processing systems, pages 2672–2680, 2014.

[2] G. Antipov, M. Baccouche, and J.-L. Dugelay. Face aging with conditional generative adversarial networks. In Image Processing (ICIP), 2017 IEEE International Conference on, pages 2089–2093. IEEE, 2017.

[3] H. Yang, D. Huang, Y. Wang, and A. K. Jain. Learning face age progression: A pyramid architecture of gans. arXiv preprint arXiv:1711.10352, 2017.

[4] L. Tran, X. Yin, and X. Liu. Disentangled representation learning gan for pose-invariant face recognition. In CVPR, volume 3, page 7, 2017.

[5] Y. Shih, S. Paris, C. Barnes, W. T. Freeman, and F. Durand. Style transfer for headshot portraits. ACM Transactions on Graphics (TOG), 33(4):148, 2014.

[6] J. Fišer, O. Jamriška, D. Simons, E. Shechtman, J. Lu, P. Asente, M. Lukáč, and D. Sỳkora. Example-based syn- thesis of stylized facial animations. ACM Transactions on Graphics (TOG), 36(4):155, 2017.

[7] J.-Y. Zhu, R. Zhang, D. Pathak, T. Darrell, A. A. Efros, O. Wang, and E. Shechtman. Toward multimodal image-to- image translation. In Advances in Neural Information Pro-cessing Systems, pages 465–476, 2017.

[8] M. Mirza and S. Osindero. Conditional generative adversar-ial nets. arXiv preprint arXiv:1411.1784, 2014.

[9] Shuyang Gu, Jianmin Bao, Hao Yang, Dong Chen and Fang Wen, Mask-Guided Portrait Editing with Conditional GANs, arXiv:1905.10346, 2019

[10] Z. Liu, P. Luo, X. Wang, and X. Tang, “Deep Learning Face Attributes in the Wild,” in Proc. IEEE/CVF International Conference on Computer Vision, 2015.

[11] Pernuš, Martin, et al. “High Resolution Face Editing with Masked GAN Latent Code Optimization.” ArXiv:2103.11135 [Cs], July 2021. arXiv.org, http://arxiv.org/abs/2103.11135.

[12] Zhenliang He, Wangmeng Zuo, Meina Kan, Shiguang Shan, and Xilin Chen. AttGAN: Facial attribute editing by only changing what you want. IEEE TIP, 2019.

[13] Deng, Q.; Cao, J.; Liu, Y.; Chai, Z.; Li, Q.; and Sun, Z. 2020. Reference Guided Face Component Editing. arXiv preprint arXiv:2006.02051.

[14] Youngjoo Jo and Jongyoul Park (2019). SC-FEGAN: Face Editing Generative Adversarial Network with User's Sketch and Color. CoRR, abs/1902.06838.