Bachelorarbeit / Masterarbeit

Titel: DEUTSCH / ENGLISH

Student: Prabhakar Kumar Panday, 3437515

Studiengang: Master’s in Electrical Engineering
Prüfer: Prof. Dr. Andreas Bulling
Betreuer: Florian Strohm
Beginn: 01.10.2021

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 other applications. Generative Adversarial Networks (GANs)[1] have made tremendous progress in synthesizing realistic faces[2], face aging[3], pose changing and attribute modification[4]. However, these existing approaches still suffer from quality issues, like lack of fine details in skin, difficulty in dealing with hair, background blurring so on.

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 synthesize realistic faces. Some works based on face masks[5,6,7] achieve promising results, but they do not synthesize different faces 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.

This project will continue work based on the paper[9], where the authors use a local embedding sub-networks to learn five features of the reference facial image by using VAEs 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[fig 1]. The authors are able to successfully synthesize diverse, high-quality, and controllable facial images from given masks, but the features of the output image is dominated by the target facial mask. To illustrate this shortcoming let us take 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 project work, each facial feature of the target mask will be individually trained and then embedded with the 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.

The model will have four inputs, a source image x^s, the mask of source image m^s, a target image x^t, and the mask of target image m^t,[fig 1]. A local embedding sub-network will be used to learn feature embedding for the input source image x^s using one variational auto-encoder(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 the efficient training of the model, various loss functions are used. Local reconstruction loss (L_{local}) is MSE loss between the input and reconstructed instances to learn the feature embedding. 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} and L_{GD} are used to generate mask equivariant facial images. Overall loss function is the sum of individual loss function.

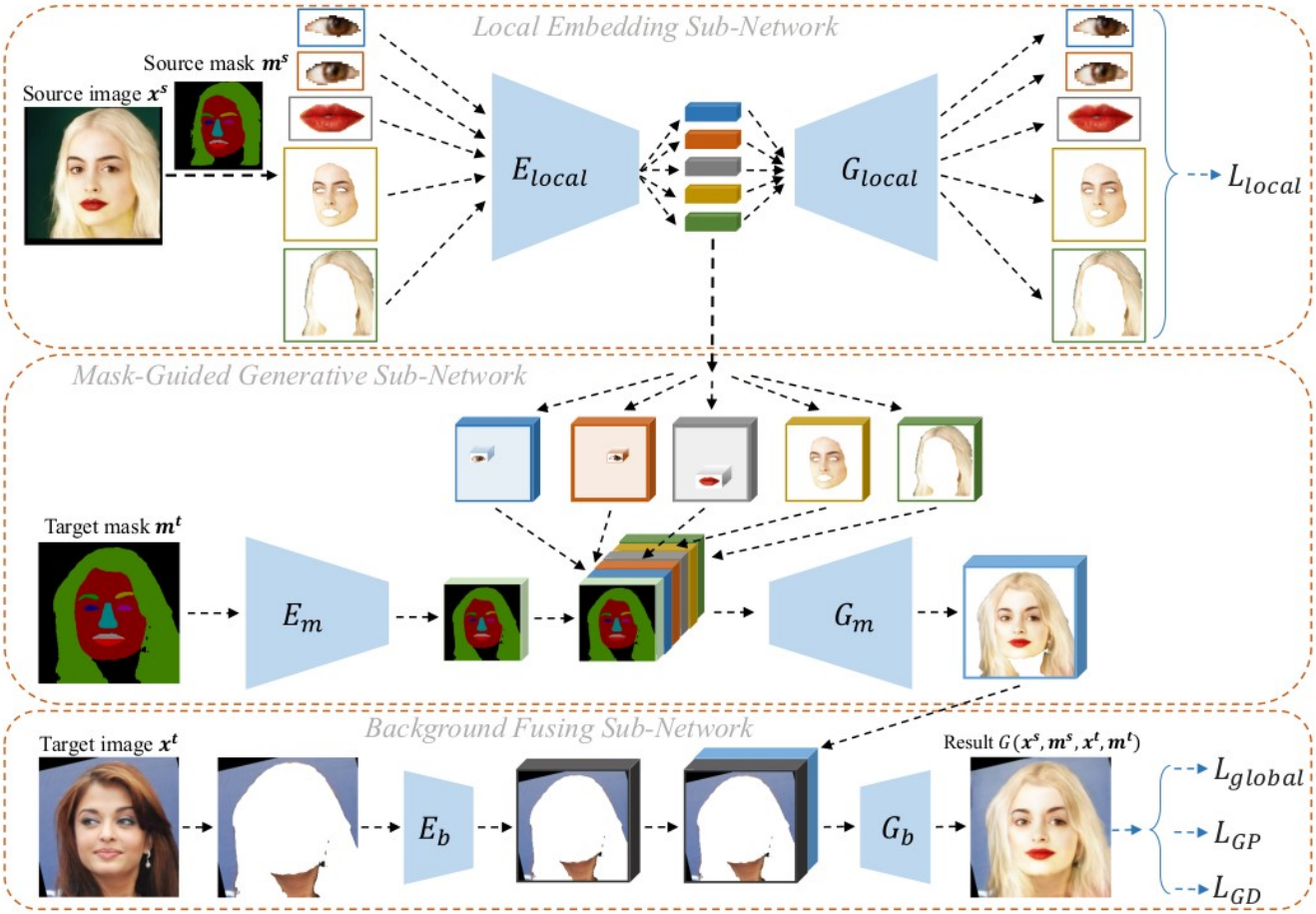
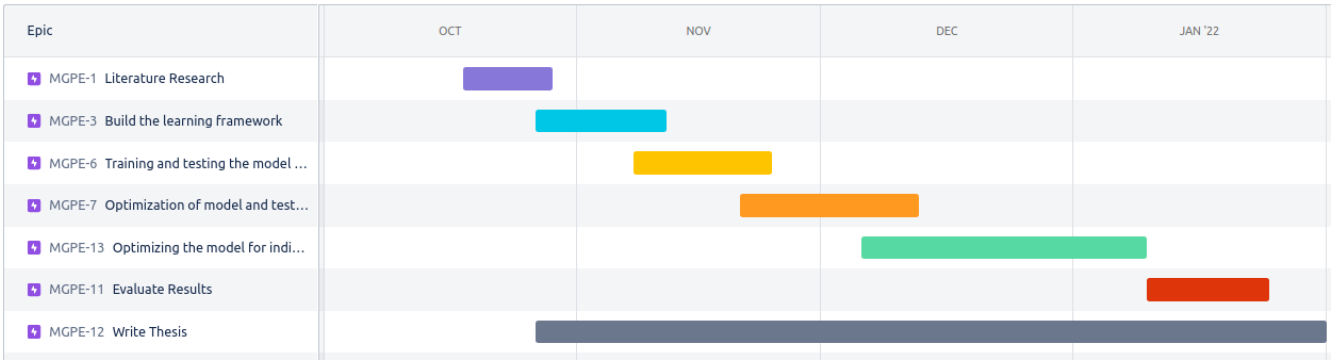


Figure 1: The framework for mask-guided portrait editing. It contains three parts: local embedding sub-network, mask guided generative sub-network, and background fusing sub-network. Local embedding sub-network learns the feature embedding of the local components of the source image. Mask guided sub-network combines the learned component feature embedding and mask to generate the foreground face image. Background fusing sub-network generates the final result from the foreground face and the background. The loss functions are drawn with the blue dashed lines.

In the following we define specific tasks to accomplish the above defined goals.

T1 Literature Research
T2 Build the Framework
T3 Training and testing the model on new training dataset
T4 Optimization of model and testing the performance
T5 Optimizing the model for individual face components
T6 Evaluate Results
T7 Write Thesis



[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.