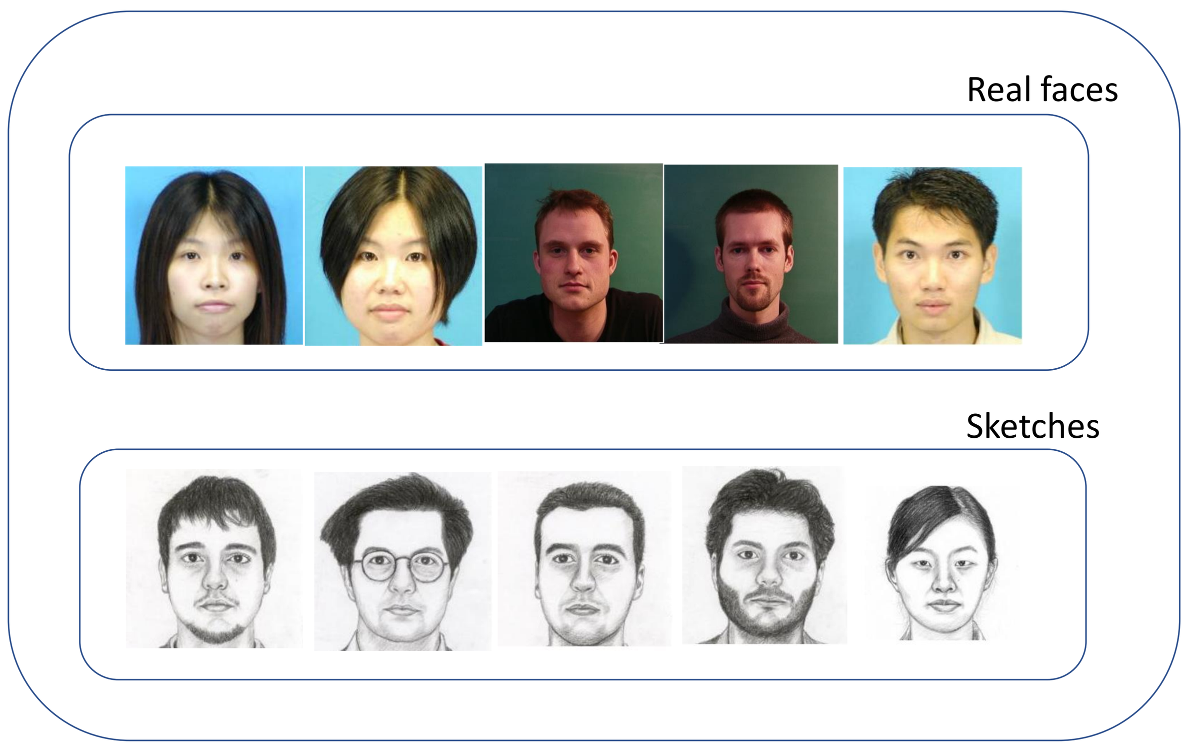
SFU
Simon Fraser University

Face sketch generation using Generative Attentional Networks

PROBLEM

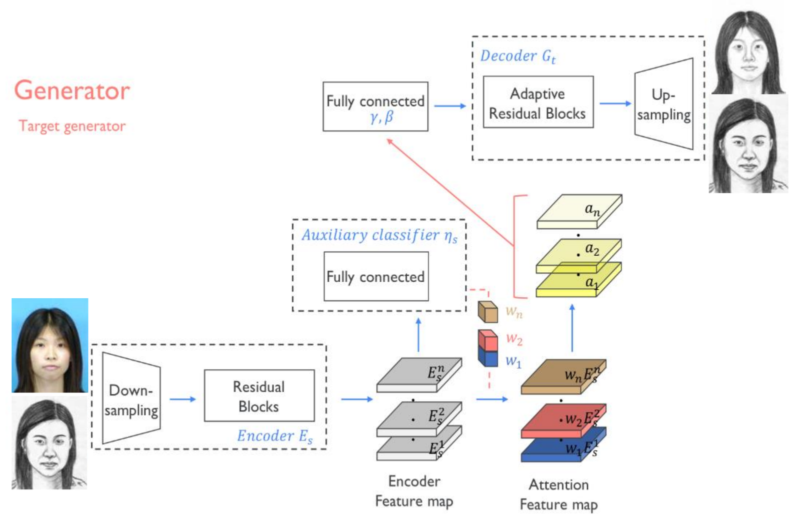
Generative Adversarial Network (GANs) has been extensively used in machine learning community to create new faces, scenes, and applying style transfers. In this project, we modified two GANs based architecture to generate realistic face sketches, given an input domain of cropped photos.

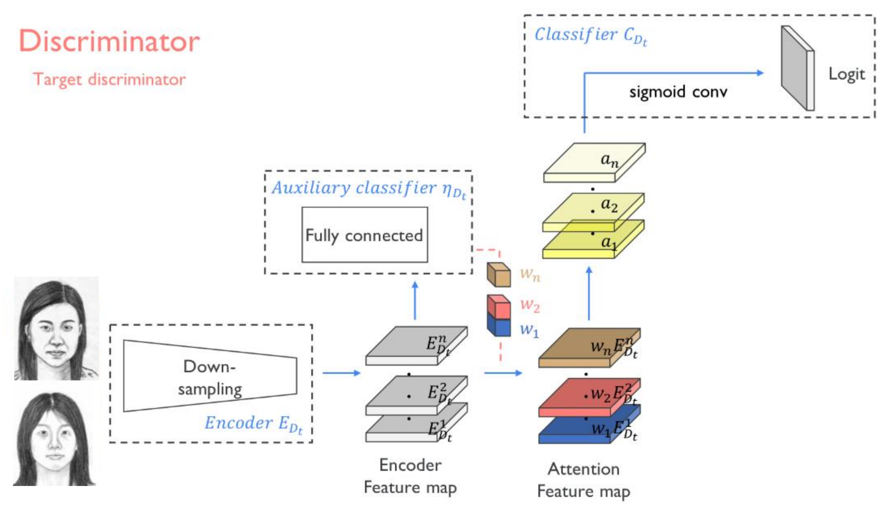


Contributions:

- Introduced edge loss term to reduce "adding glasses" problem.
- Compiled a dataset for unsupervised training of face sketch generation.

Unsupervised Generative Attentional Networks





Major differences from a conventional GAN:

Attention module (auxiliary classifier): $\eta_s(x) = \sigma(\Sigma_k w_s^k \Sigma_{ij} E_s^{k_{ij}}(x))$

Adaptive Layer-Instance Normalization:

$$\hat{a_I} = \frac{a-\mu_I}{\sqrt{\sigma_I^2+\epsilon}}, \hat{a_L} = \frac{a-\mu_L}{\sqrt{\sigma_L^2+\epsilon}},$$

$$AdaLIN(a,\gamma,\beta) = \gamma\cdot(\rho\cdot\hat{a_I}+(1-\rho)\cdot\hat{a_L})+\beta,$$

$$\rho \leftarrow clip_{[0,1]}(\rho-\tau\Delta\rho),$$

Loss function

$$\min_{G_{s\to t},G_{t\to s},\eta_s,\eta_t}\ \max_{D_s,D_t,\eta D_s,\eta D_t} \lambda_1 L_{gan}+\lambda_2 L_{cycle}+\lambda_3 L_{identity}+\lambda_4 L_{cam}+\lambda_5 L_{edge}$$

Encourage alignment of edges between a face and sketch

Where $G_{s\to t}$ and D_t can improve

Color distributions of input and output images should be similar

Translations of image X_s to X_t and back should preserve image

Loss details

$$L_{cycle}^{s\to t} = \mathbb{E}_{x\sim X_s}[|x - G_{t\to s}(G_{s\to t}(x)))|_1]$$

$$L_{identity}^{s\to t} = \mathbb{E}_{x\sim X_t}[|x - G_{s\to t}(x)|_1]$$

$$L_{cam}^{s\to t} = -(\mathbb{E}_{x\sim X_s}[log(\eta_s(x))] + \mathbb{E}_{x\sim X_t}[log(1-\eta_s(x))])$$

$$L_{cam}^{D_t} = \mathbb{E}_{x\sim X_t}[(\eta D_t(x))^2] + \mathbb{E}_{x\sim X_s}[log(1-\eta_{D_t}(G_{s\to t}(x))^2]$$

Training details:

First 100 epochs: $\lambda_1 = 5, \lambda_2 = 10, \lambda_3 = 10, \lambda_4 = 1000, \lambda_5 = N/A$
100-400 epochs: $\lambda_1 = 2, \lambda_2 = 10, \lambda_3 = 10, \lambda_4 = 1000, \lambda_5 = 0.5$

Total time taken: 48 hours on a GTX 1080Ti

Training and Results

Dataset: For training and testing, we used the Chinese University of Hong Kong's Face Sketch Database to get images for sketches. For images of faces, we used the Chicago Face Dataset and the IMM Face dataset. For training, we used around 500 images and 150 images for testing.

Experiment: We used the data to train our proposed network and used results trained on CycleGAN as a baseline to compare our results with. We obtain quantitative results using a testing set from CUHK which has faces and corresponding sketches made by an artist.

Quantitative results :

	CycleGan	U GAT IT (proposed loss function)
Structural Similarity (SSIM)	0.556	0.62

Qualitative results : Face

U-GAT-IT prediction

CycleGAN prediction

Conclusion: U GAT IT performs much better than CycleGAN. It does not add unnecessary artefacts, and preserves the shape and expression of the face. The higher SSIM backs up the claim.

References

[1] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017.
[2] J.Kim, M.Kim, H.Kang, and K.Lee, "U-GAT-IT: Unsupervised Generative Attentional Networks with Adaptive Layer-Instance Normalization for Image-to-Image Translation", arXiv:1907.10830 [cs], Jul. 2019.
[3] 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