Deep learning for face image editing

Alasdair Newson
alasdair.newson@telecom-paris.fr

LTCI day

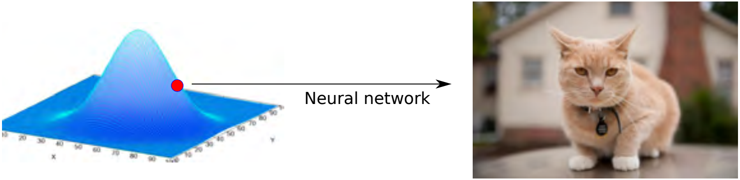
Introduction

Introduction

- **Image editing** is a common, critical and time-consuming task in such domains as film post-production
  - Movie industry often employs hundreds of digital artists for editing
- In particular, **facial images**, are of great importance
  - Controlling facial attributes : hair, smile, glasses etc


- For classical methods, this is a challenging task
  - Requires prior knowledge/model about facial images

- Since 2014, **deep generative models** have seen an explosion of research actvity
- Deep generative models : deep neural networks
  - Input : random vector
  - Output : random image (or data)



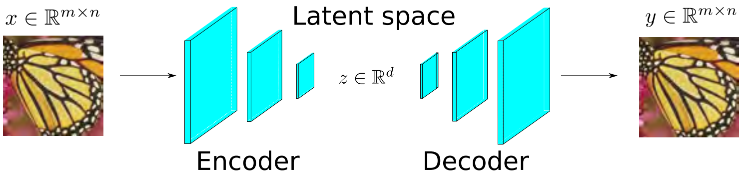
Probabilistic model in latent space

Synthesis of random image

- These rely on the concept of a **latent space**

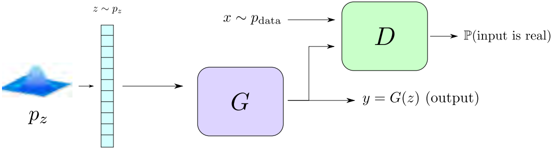
Introduction

- The latent space is (almost always) a **smaller dimensional** space



- More **compact**, better properties. Can be used for image editing/restoration/analysis

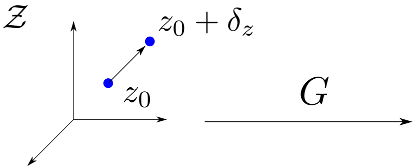
- Most common deep generative networks: Variational Autoencoders*, **Generative Adversarial Networks**† (GANs)



- We work with the recent **StyleGAN**‡; produces high-quality results

* ***Auto-Encoding Variational Bayes***, *D. P. Kingma, M. Welling, arXiv preprint arXiv:1312.6114, 2013*
† ***Generative Adversarial Nets***, *Goodfellow et al, NIPS 2014*
‡ ***Analyzing and improving the image quality of stylegan***, *Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., and Aila, T. CVPR 2020*

- Main idea of our work: **navigate** in the latent space to **achieve an editing goal**



$G(z_0)$

$G(z_0 + \delta_z)$

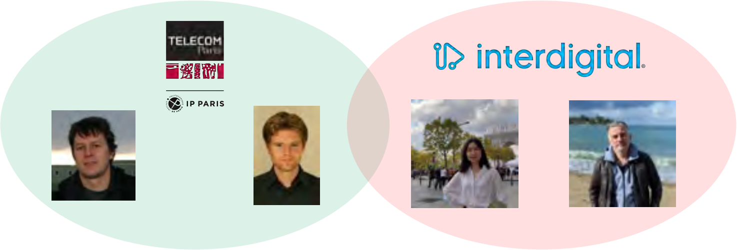
- Previous work on this subject[†,‡]

[*] ***A latent transformer for disentangled face editing in images and videos***, *Yao, X., Newson, A., Gousseau, Y., and Hellier, P., ICCV 2021*

[†] ***Interfacegan: Interpreting the Disentangled Face Representation Learned by Gans***, *Shen, Y., Yang, C., Tang, X. and Zhou, B, PAMI, 2020*

[‡] ***Ganspace: Discovering interpretable gan controls***, *Harkonen, E., Hertzmann, A., Lehtinen, J. and Paris, S., NIPS 2020*

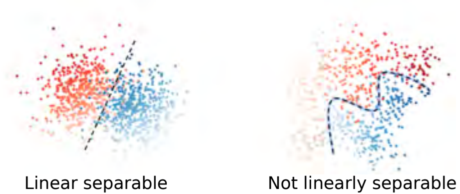
- PhD work of **Xu Yao**, CIFRE PhD with Interdigital
- Published in ICCV 2021*



Collaborators

* ***A latent transformer for disentangled face editing in images and videos***, *Yao, X., Newson, A., Gousseau, Y., and Hellier, P., ICCV 2021*

- Previous approaches* often suppose that facial attributes are **linearly separable** in the latent space



- This is a limiting hypothesis. Thus, we remove this it in this work
- We train a neural network to navigate in StyleGAN's latent space $\mathcal{W}$

$$w_1 = w_0 + \alpha T(w_0), \tag{1}$$

* ***Interfacegan: Interpreting the Disentangled Face Representation Learned by Gans***, *Shen, Y., Yang, C., Tang, X. and Zhou, B, PAMI, 2020*

Face Editing with Generative Models

- We minimise the following loss functions:

$$\mathcal{L}_{\mathsf{cls}}(w) = -y_k \log(p_k) - (1-y_k)\log(1-p_k), \tag{2}$$

- $p_k = C(T_k(w))_k$: probability of the target attribute k
- $y_k \in \{0,1\}$: desired label

Face Editing with Generative Models

- We minimise the following loss functions:

$$\mathcal{L}_{\mathsf{cls}}(w) = -y_k \ \mathsf{log} \ (p_k) - (1 - y_k) \ \mathsf{log} \ (1 - p_k), \tag{2}$$

- $p_k = C(T_k(w))_k$: probability of the target attribute k
- $y_k \in \{0, 1\}$: desired label

$$\mathcal{L}_{\mathsf{attr}}(w) = \sum_{i \neq k} (1 - \gamma_{i,k}) \|p_i - C(w)_i\|_2^2, \tag{3}$$

- $\gamma_{i,k}$: absolute correlation between a_i and target attribute a_k
  - Mesured on the database

Face Editing with Generative Models

- We minimise the following loss functions:

$$\mathcal{L}_{\text{cls}}(w) = -y_k \text{ log } (p_k) - (1 - y_k) \text{ log } (1 - p_k), \tag{2}$$

- $p_k = C(T_k(w))_k$: probability of the target attribute k
- $y_k \in \{0, 1\}$: desired label

$$\mathcal{L}_{\text{attr}}(w) = \sum_{i \neq k} (1 - \gamma_{i,k}) \|p_i - C(w)_i\|_2^2, \tag{3}$$

- $\gamma_{i,k}$: absolute correlation between a_i and target attribute a_k
  - Mesured on the database

$$\mathcal{L}_{\text{rec}}(w) = \|T(w) - w\|_2^2. \tag{4}$$

- Do not move too far from original latent code: **maintain identity**

Final loss

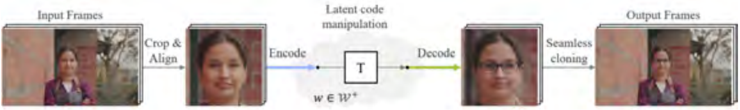
$$\mathcal{L} = \mathbb{E}_w \left[\mathcal{L}_{\text{cls}}(w) + \lambda_{\text{attr}} \mathcal{L}_{\text{attr}}(w) + \lambda_{\text{rec}} \mathcal{L}_{\text{rec}}(w) \right], \tag{5}$$

- For more efficient training, we train a **classifier in the latent space**

Final loss

$$\mathcal{L} = \mathbb{E}_w \left[\mathcal{L}_{\text{cls}}(w) + \lambda_{\text{attr}} \mathcal{L}_{\text{attr}}(w) + \lambda_{\text{rec}} \mathcal{L}_{\text{rec}}(w) \right], \tag{5}$$

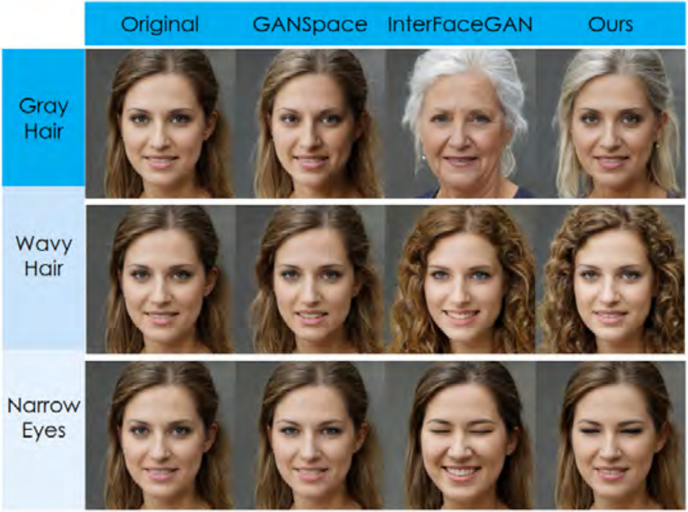
- For more efficient training, we train a **classifier in the latent space**
- We also propose a **pipeline to edit videos**
  1. Preprocessing: landmark detection
  2. Image editing: Encode*, manipulate, generate
  3. Seamless cloning†



* ***Encoding in style: a Stylegan Encoder for Image-to-Image Translation***, *Richardson, E., Alaluf, Y., Patashnik, O., Nitzan, Y., Azar, Y., Shapiro, S., and Cohen-Or, D., CVPR, 2021*

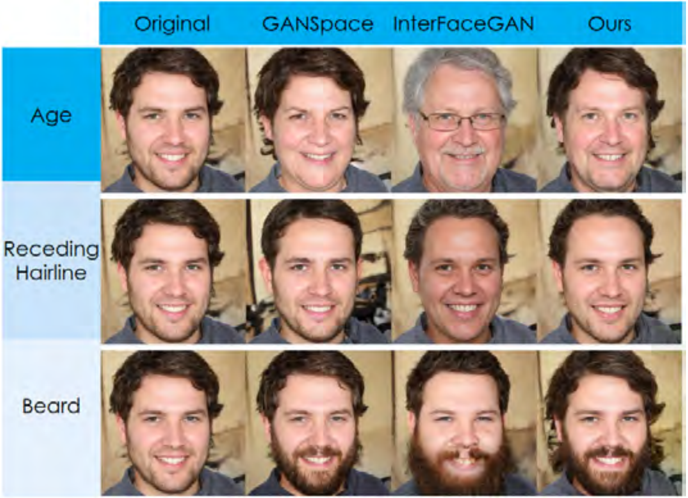
† ***Poisson image editing***, *Pérez, P., Gangnet, M., and Blake, A., Siggraph 2003*

Image attribute editing results



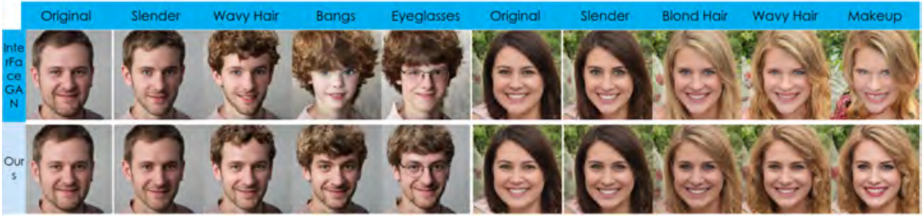
Face Editing with Generative Models

Image attribute editing results



Face Editing with Generative Models

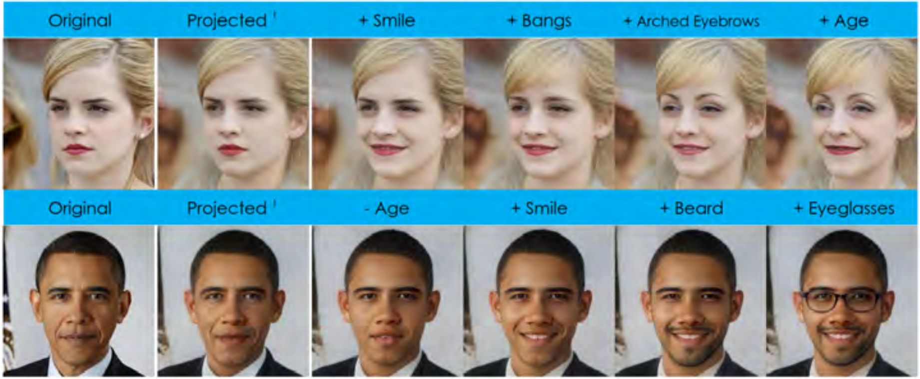
Sequential editing results: synthetic images



- Our approach preserves identity and other attributes better

Face Editing with Generative Models

Sequential editing results: real images



Conclusion

Summary

- We proposed a latent space-based face attribute editing algorithm*
  - Uses the latent space of **StyleGAN**
- State-of-the art, **disentangled**, results on **real** (non-synthetic) results, **identity preservation**
- **Video pipeline** to edit videos

Limitations and future work

- A **different network** needs to be trained for each attribute
  - Better to try and modify the latent space itself
- The **non-linearity** of the latent space is still an open question
- More work needed on stability and robustness in videos

* ***A latent transformer for disentangled face editing in images and videos***, *Yao, X., Newson, A., Gousseau, Y., and Hellier, P., ICCV 2021*

References

- **A latent transformer for disentangled face editing in images and videos**, Yao, X., Newson, A., Gousseau, Y., and Hellier, P., ICCV 2021

- **Interfacegan: Interpreting the Disentangled Face Representation Learned by Gans**, Shen, Y., Yang, C., Tang, X. and Zhou, B, PAMI, 2020
- **Ganspace: Discovering interpretable gan controls**, Härkönen, E., Hertzmann, A., Lehtinen, J., and Paris, S., NIPS 2020
- **Analyzing and improving the image quality of stylegan**, Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., and Aila, T. CVPR 2020