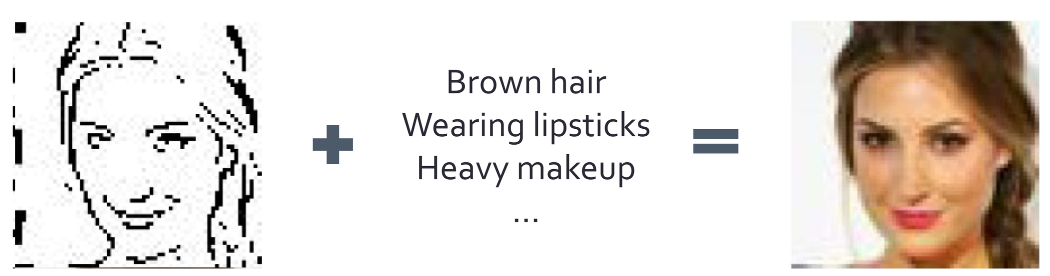
ATTRIBUTE-CONTROLLED FACE PHOTO SYNTHESIS FROM SIMPLE LINE DRAWING

Qi Guo, Ce Zhu, Zhiqiang Xia, Zhengtao Wang, Yipeng Liu
School of Electronic Engineering / Center for Robotics
University of Electronic Science and Technology of China (UESTC), Chengdu, China

ABSTRACT



- we propose a deep generative model to synthesize face photo from simple line drawing controlled by face attributes such as hair color and complexion.
  - In order to maximize the controllability of face attributes, an attribute-disentangled variational auto-encoder (AD-VAE) is firstly introduced to learn latent representations disentangled with respect to specified attributes.
  - Then we conduct photo synthesis from simple line drawing based on AD-VAE.
  - Regarding background and illumination as the style and human face as the content, we can also synthesize face photos with the target style of a style photo.

INTRODUCTION

- Face sketch-photo synthesis has been well developed in recent years for its widely application on law enforcement.

- As a rough kind of sketch, simple line drawing, merely containing the contour of human face, is easy to be obtained and modified for ordinary people, which also makes photo synthesis from simple line drawing a quite tough task.
- The prior works of face sketch-photo synthesis can be divided into two categories: traditional methods and deep learning methods.
  - traditional methods: based on an image patch dictionary of exemplars, time-consuming, poor generalization
  - deep learning methods: based on deep neural network , fast in testing stage due to its feed-forward framework, better generalization

METHOD

- Variational auto-encoder (VAE)
  - Given an input $x \in R^N$ and its corresponding latent variables $z \in R^M$, the basic structure of VAE consists of two networks: an encoder $q_\emptyset(z|x)$ (recognition model) to approximate posterior inference and a decoder $p_\theta(x|z)$ (generative model) to map the latent variables to data space.
  - max $L(x;\emptyset,\theta) = -D_{KL}(q_\emptyset(z|x)||p_\theta(z)) + E_{q_\emptyset(z|x)}[\log p_\theta(x|z)]$
  - $p_\theta(z) \sim N(0,I)$
  - $q_\emptyset(z|x)$ and $p_\theta(x|z)$ are multivariate Gaussian distributions parameterized by deep neural networks.

$$x \xrightarrow{q_\emptyset} z \xrightarrow{p_\theta} x$$

METHOD

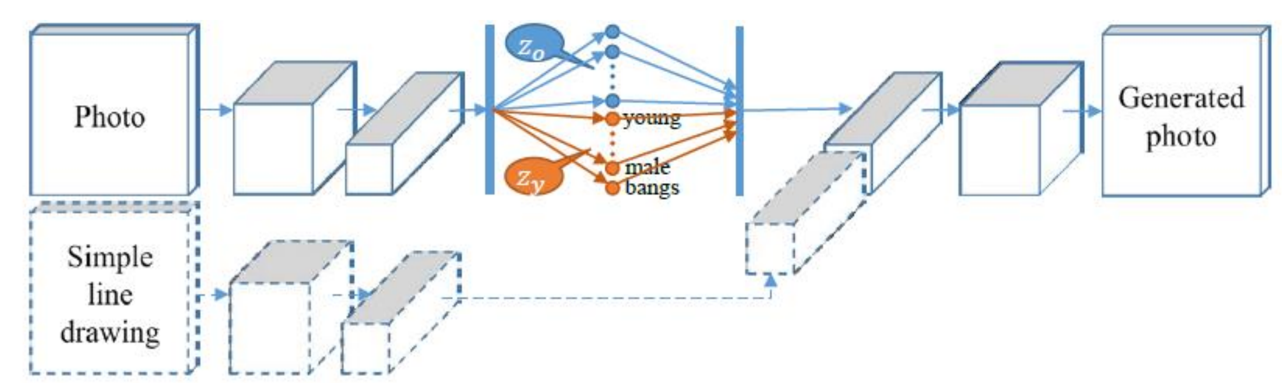
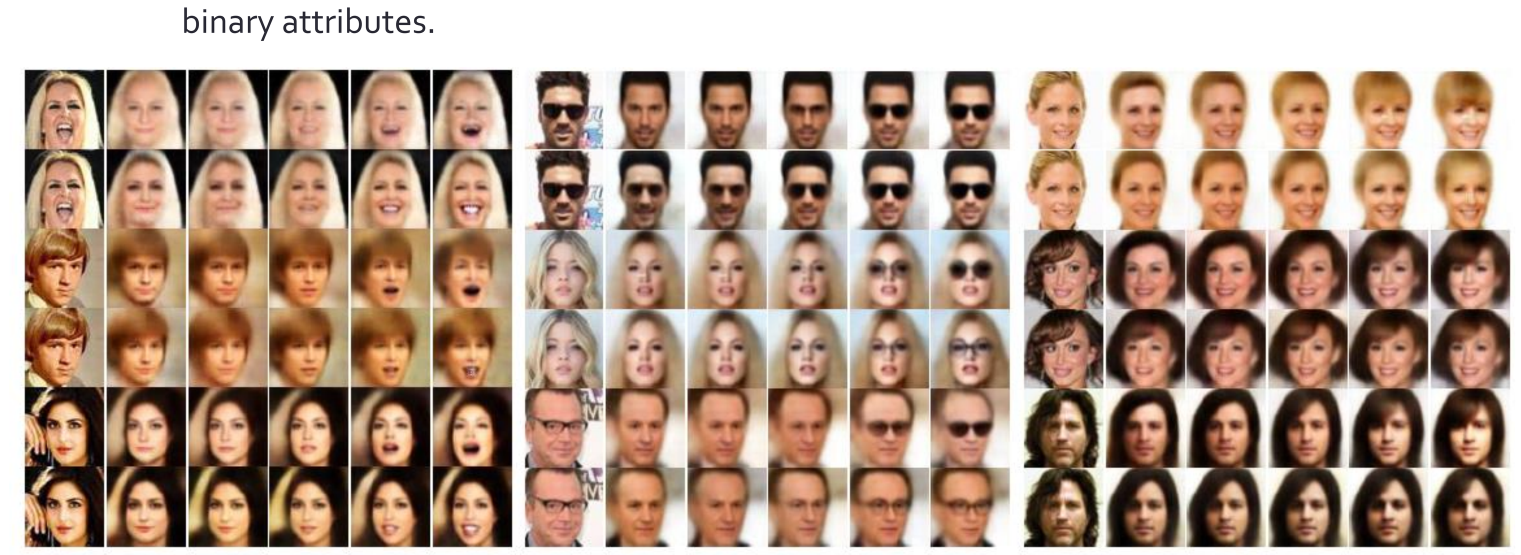


Fig. 1: Model architecture. Solid lines depict the basic structure of AD-VAE. Dashed lines show another channel of convolutional neural network which takes simple line drawing as input.

- Attribute-disentangled variational auto-encoder (AD-VAE)
  - As shown in Figure 1, the latent variables z is split into two parts: $z_y \in R^L$ that represent the variations of face attributes and $z_o \in R^K$ which capture remaining factors of variation, such as position and background .
  - $p_\theta(z_o) \sim N(0,I)$
  - $p_\theta\left(z_y^i|y^i\right) \sim N(y^i,\alpha)$
  - $$L(x,y;\emptyset,\theta) = -\sum_{i=1}^{L} \alpha^i D_{KL}\left(q_\emptyset(z_y^i)||p_\theta\left(z_y^i|y^i\right)\right) -\beta D_{KL}(q_\emptyset(z_o|x)\|p_\theta(z_o)) +E_{q_\emptyset(z_o|x)q_\emptyset(z_y|x)}[\log p_\theta(x|z_o,z_y)]$$
  - $q_\emptyset(z_o|x)$, $q_\emptyset(z_y|x)$ and $p_\theta(x|z_o,z_y)$ are multivariate Gaussian distributions parameterized by deep neural networks.
- Photo Synthesis from Simple Line Drawing
  - As shown in Figure 1, another channel of convolutional neural network is added to the AD-VAE which takes simple line drawing as input and its feature maps are concatenated to the decoder.
  - $$L(x,y,s;\emptyset,\theta) = -\sum_{i=1}^{L} \alpha^i D_{KL}\left(q_\emptyset(z_y^i)||p_\theta\left(z_y^i|y^i\right)\right) -\beta D_{KL}(q_\emptyset(z_o|x)\|p_\theta(z_o)) +E_{q_\emptyset(z_o|x)q_\emptyset(z_y|x)}[\log p_\theta(x|z_o,z_y,s)]$$

RESULTS

- Attribute Manipulation
  - In order to demonstrate the qualitative disentanglement with respect to the face attributes, we manipulate the attribute variables by varying desired attribute variable smoothly and keeping all other latent variables fixed.
  - We compare AD-VAE to attribute-conditioned variational auto-encoder (AC-VAE) [1].
  - For fair model comparison, we train both AC-VAE and AD-VAE on CelebA with 38 selected binary attributes.



(a) Mouth_slightly_open (b) Eyeglasses (c) Bangs

Fig. 2: Manipulating attribute variables. For each group, the first column is ground truth. For each ground truth, the first row is generated by AD-VAE and the second row is generated by AC-VAE.

RESULTS

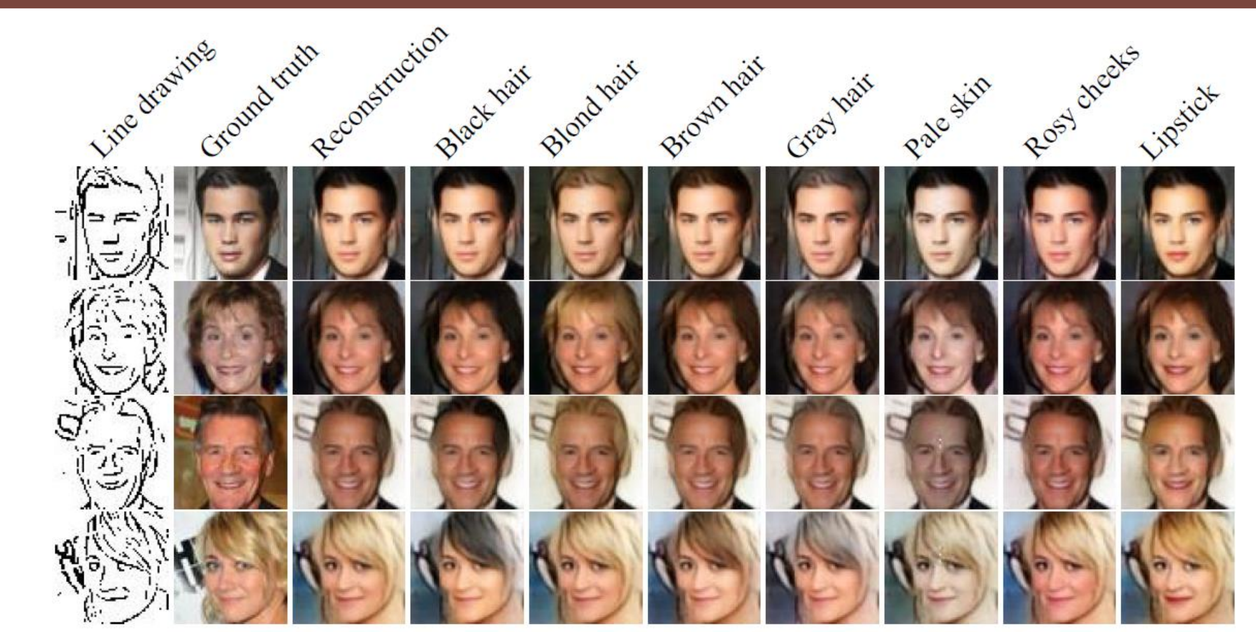


Fig. 3: Attribute controlled photo synthesis

- Photo Synthesis from Simple Line Drawing
  - As shown in Figure 3, controlled by face attributes, we can modify the hair color and complexion of the synthesized face photos.
  - In Figure 4, two types of simple line drawings with different stroke weights are used to generate face photos. Compared to the CSI [2] with content loss, our proposed method can synthesize more photorealistic and natural faces images even though the contours of human faces are not complete.
  - We exchange zo of several photos in ZJU-VIPA Line Drawing Face Database and CelebA dataset to synthesize face images with target styles in Figure 5.

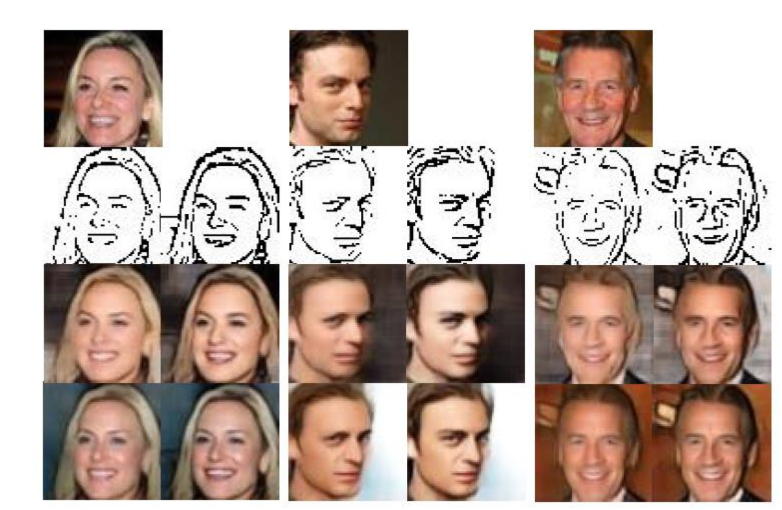


Fig. 4: Results comparison. From top to bottom: ground truth, two types of simple line drawings, CSI, our method

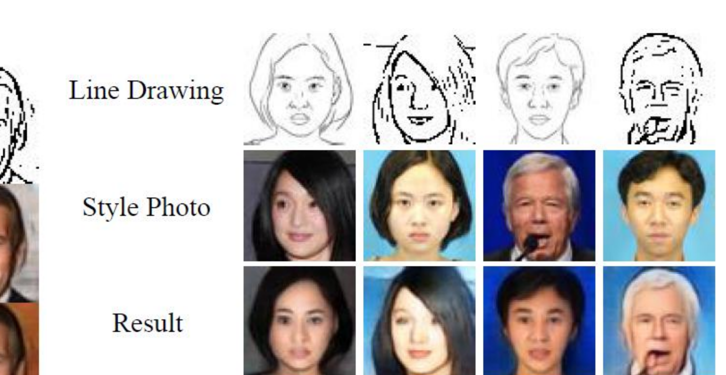


Fig. 5: Target style photo synthesis

CONCLUSION

- First, an attribute-disentangled variational autoencoder (AD-VAE) is introduced to disentangle variations of face attributes from other variations of face images.
- Then we synthesized face photo from simple line drawing based on AD-VAE.
- Experiments showed our proposed method could learn interpretable representations of face images and generate face images with rich details and desired attributes even though the simple line drawing is not complete.

REFERENCES

1. X.C. Yan, J.M. Yang, K. Sohn, and H.L Lee, "Attribute2image: Conditional image generation from visual attributes," in European Conference on Computer Vision. Springer, 2016, pp. 776–791.
2. Y. Güüçlütürk, U. Gülü, R. van Lier, and M. AJ. Van Gerven, "Convolutional sketch inversion," in European Conference on Computer Vision. Springer, 2016,pp. 810–824.