Shanshan HUANG, Yuanhao WANG, Zhili GONG, Jun LIAO, Shu WANG, Li LIU, 2024. Controllable image generation based on causal representation learning. *Frontiers of Information Technology & Electronic Engineering*, 25(1):135-148.
https://doi.org/10.1631/FITEE.2300303

Controllable image generation based on causal representation learning

Key words: Image generation; Controllable image editing; Causal structure learning; Causal representation learning

Corresponding author: Li LIU
E-mail: dcsliuli@cqu.edu.cn
ORCID: https://orcid.org/0000-0002-4776-5292

Motivation

1. Existing conditional image generation methods require independent sampling of labels, ignoring the dependencies between the labels. The generated image can be affected by spurious correlations, which can lead to irrational results.
2. In practical applications, the semantically meaningful factors of interest are not independent but causally related, and a critical aspect often overlooked.
3. Causal representation learning methods for controllable image generation have shown promising results. However, the quality of generated images is suboptimal, and these methods require a predefined causal structure among the attributes.

Method

A causal controllable image generation method is proposed by integrating causal structure learning (CSL) with bi-directional generative adversarial networks (GANs).

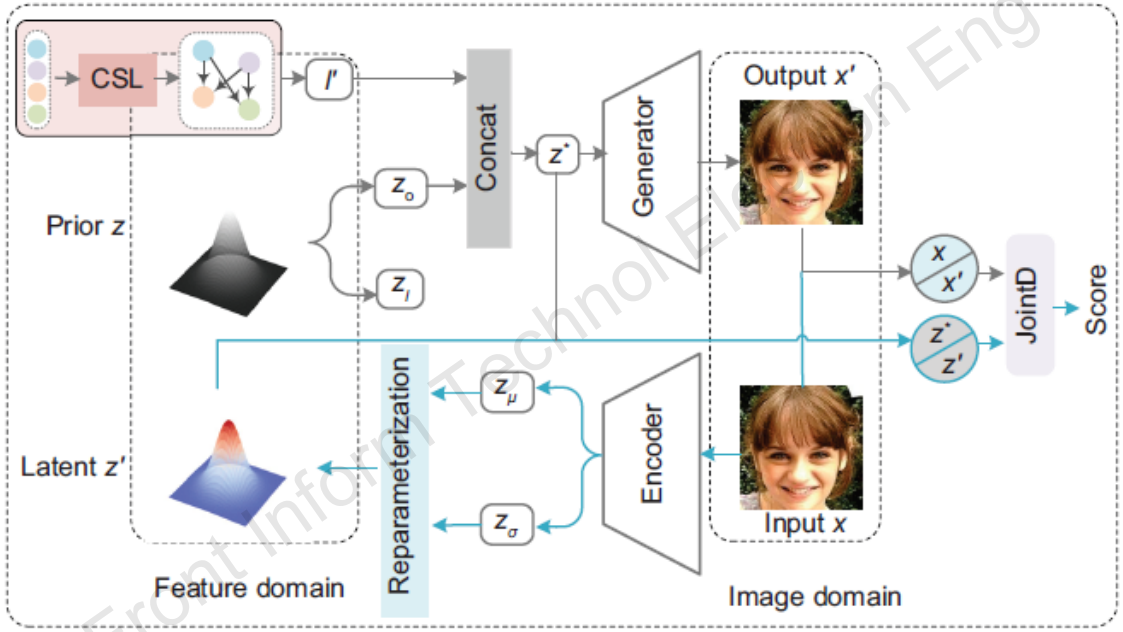

Fig. 1 Overall workflow of our framework. In the training phase, our model first uses the CSL module to learn causality from the image labels, thereby constraining the underlying representation $z' = E(x)$ obtained from the encoder to be consistent with the learned causal relationships. Then, the causal representation l', which encodes causal relationships between factors of interest in image generation, is used to replace a portion z_l of the prior distribution z, which is then concatenated with the other factors z_o needed for image generation to obtain a new latent representation z^*. This representation is then fed into the generator G to obtain the generated image $x' = G(z^*)$. Note that JointD is trained alternatively with the generator G and encoder E. In the inference phase, we can achieve causal controllability of an image attribute by modifying the value of a latent representation along a specific dimension, i.e., causal intervention

Method

CSL module:

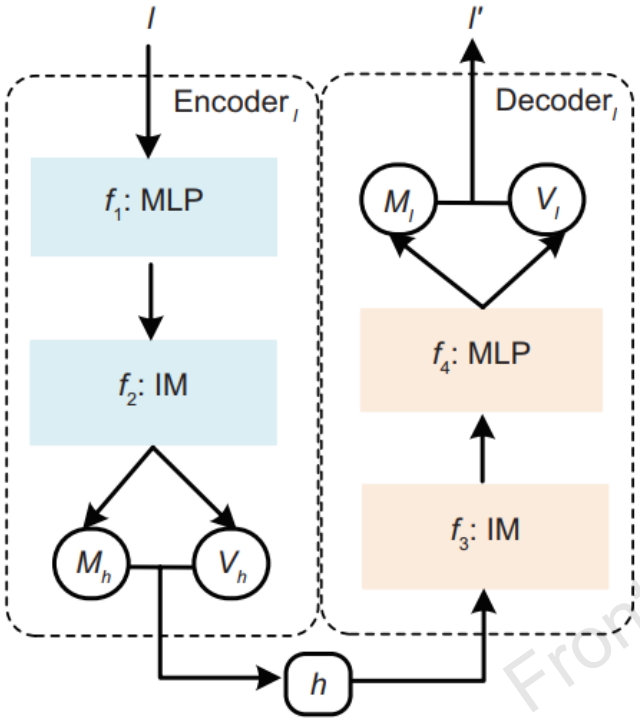

Fig. 2 Workflow of the CSL module, which consists of two components: Encoder$_l$ and Decoder$_l$. Encoder$_l$ acts as an inference model designed to encode the input attribute l as the latent posterior h. Decoder$_l$ acts as the causal generative model for reconstructing the input attributes

Optimization objective of the CSL module:

$$\mathcal{L}_\nu(A,\varphi,\kappa) = \mathcal{L}_\mathrm{r} + r_l + \kappa c(A) + \frac{\nu}{2}|c(A)|^2, \qquad (15)$$

$$\mathcal{L}_\mathrm{r} = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{d} l_{ij}\log(M_l)_{ij}.$$

$$r_l = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{d}(M_h)_{ij}^2.$$

$$c(A): \mathrm{tr}[(I+\alpha A\circ A)^m] - m = 0,$$

Method

Image generation module:

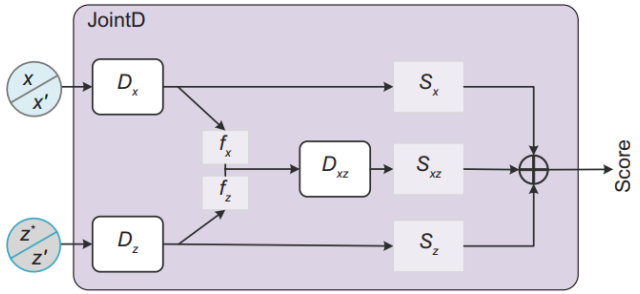

Fig. 3 Workflow of JointD

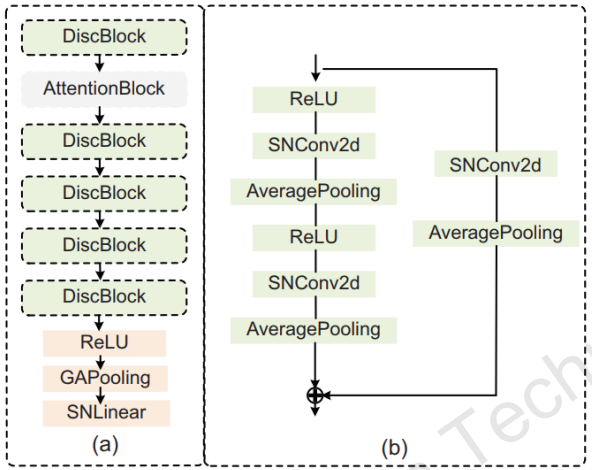

Fig. 5 Detail of discriminator D_x: (a) discriminator; (b) DiscBlock

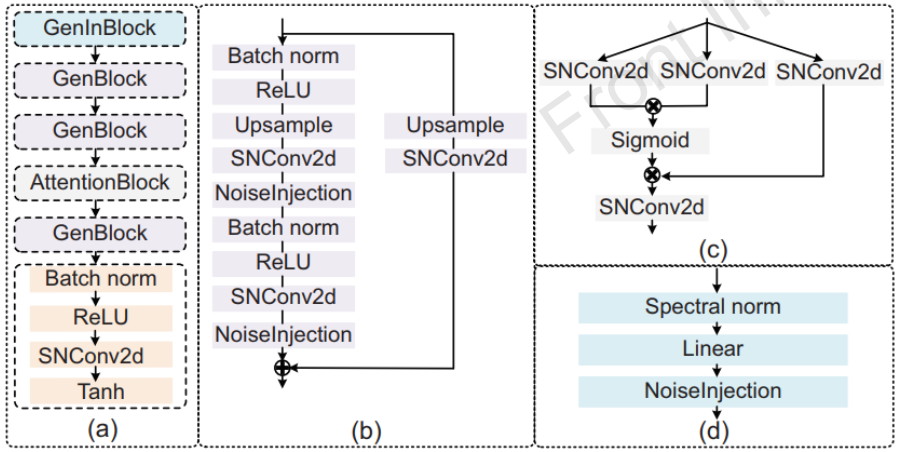

Fig. 4 Details of the generator: (a) generator; (b) GenBlock; (c) AttentionBlock; (d) GenInBlock

Optimization objective of the image generation module:

$$\mathcal{L}_{G,E} = \mathbb{E}_{x\sim p_x, z\sim p_e(z|x)}[l_{G,E}(x, E(x))] - \mathbb{E}_{z\sim p_z, x\sim p_g(x|z)}[l_{G,E}(G(z^*), z^*)] + \lambda \mathcal{L}_{\text{sup}}(E), \tag{16}$$

$$\mathcal{L}_{\text{sup}}(E) = \mathbb{E}_{x,l}[l_s(E; x, l)],$$

$$\mathcal{L}_D = \mathbb{E}_{x\sim p_x, z\sim p_e(z|x)}[l_D(x, E(x), 1)] + \mathbb{E}_{z\sim p_z, x\sim p_g(x|z)}[l_D(G(z^*), z^*, -1)], \tag{17}$$

$$l_{G,E}(x, z) = s_x(x) + s_z(z) + s_{xz}(x, z)$$

$$l_D(x, z, \tau) = h(\tau s_x) + h(\tau s_z) + h(\tau s_{xz}),\ \tau \in \{-1, 1\}$$

$$\begin{cases} s_x(x) = \theta_x^{\mathrm{T}} D_x^{\Theta}(x), \\ s_z(z) = \theta_z^{\mathrm{T}} D_z^{\Theta}(z), \\ s_{xz}(x, z) = \theta_{xz}^{\mathrm{T}} D_{xz}^{\Theta}(D_x^{\Theta}(x) + D_z^{\Theta}(z)), \end{cases}$$

Experiments

Visual results on the CelebA dataset

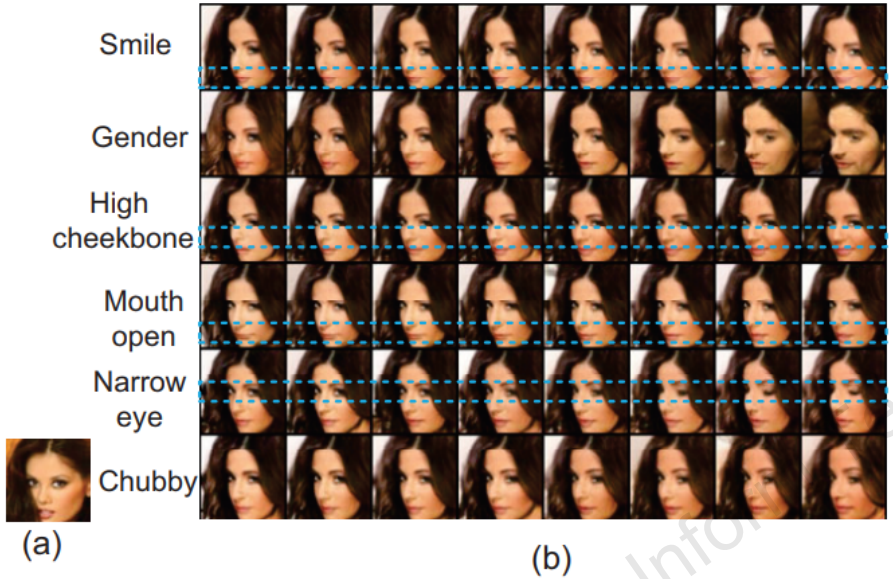

Fig. 6 Results of CCIG in causal controllable generation on the CelebA (smile) dataset: (a) source image; (b) results. For each row, we change only one latent factor and fix all other latent factors. By changing the cause factor (smile or gender), we observe corresponding changes in the effect factors (mouth open and narrow eye). Conversely, by changing the effect factor (mouth open, high cheekbone, narrow eye, or chubby), the reconstruction can become a counterfactual image, while the cause factor remains unchanged

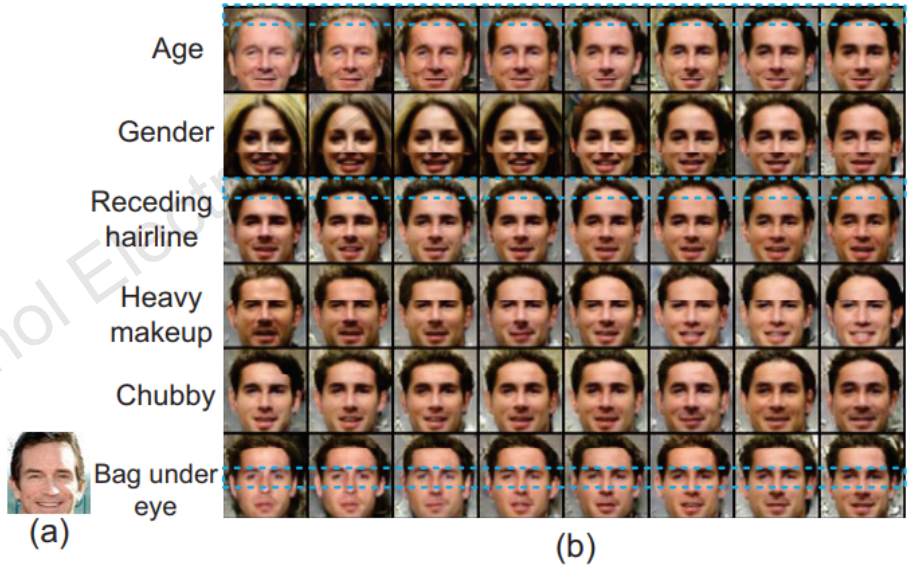

Fig. 7 Results of CCIG in causal controllable generation on the CelebA (age) dataset: (a) source image; (b) results. For each row, we change only one latent factor and fix all other latent factors. By changing the cause factor (age or gender), we observe corresponding changes in the effect factors (receding hairline and bag under eye). Conversely, by changing the effect factor (receding hairline, heavy makeup, chubby, or bag under eye), the reconstruction can become a counterfactual image, while the cause factor remains unchanged

Experiments

Performance comparison

Table 1 Performance comparison with the baseline methods

Method	IS_MEAN	IS_STD	FID	MOS*	MOS**
CausalVAE	1.3659	0.0694	284.41	4.16	4.52
CMGAN	1.3392	0.0412	218.12	3.52	3.86
CausalGAN	1.8314	**0.0157**	144.16	3.20	4.15
DEAR	2.0140	0.0232	97.16	4.58	4.03
CCIG (L2)	2.0360	0.0214	96.89	4.43	4.60
CCIG (CE)	**2.1461**	0.0347	**96.71**	**4.66**	**4.62**

* Image quality; ** controllability of image editing. FID: Fréchet inception distance; MOS: mean opinion score. IS_MEAN and IS_STD represent the mean and standard deviation of the inception score (IS), respectively. Best results are in bold

Table 2 Results of ablation experiments

Method	IS_MEAN	IS_STD	FID	MOS*	MOS**
CCIG_GNA	**2.296**	0.0740	104.63	4.58	4.33
CCIG_DNA	2.1107	0.0649	100.77	4.40	4.37
CCIG_DNs_{xz}	1.9482	0.0482	102.36	4.33	4.43
CCIG (CE)	2.1461	**0.0347**	**96.71**	**4.66**	**4.62**

* Image quality; ** controllability of image editing. FID: Fréchet inception distance; MOS: mean opinion score. IS_MEAN and IS_STD represent the mean and standard deviation of the inception score (IS), respectively. Best results are in bold

Conclusions

1. We propose a novel causal controllable image generation (CCIG) framework, which combines a CSL module with an image generation module (IGM), to learn causal graphs of image attributes, and the learned causal graphs are used to constrain the latent representations to understand the image generation mechanism.
2. We propose a bi-directional generative model for image generation and representation learning, which combines an encoder, a generator, and a joint discriminator (JointD) to improve the model representation learning capability.
3. We instantiate the proposed framework and conduct extensive experiments over a public dataset, CelebA, verifying the effectiveness and rationality of CCIG.

Shanshan HUANG is a doctoral candidate at the School of Big Data and Software, Chongqing University. She received her MS degree in Software Engineering from the School of Software, Yunnan University, in 2021. Her research interests include deep learning, computer vision and image processing, and causal learning.

Li LIU is currently a professor with the School of Big Data and Software at Chongqing University. He received his BS degree from Lanzhou University, Lanzhou, in 2003, and his PhD degree from the Université Paris-Sud XI in 2008. His research interests include causal learning, human-computer interaction technology, behavioral cognitive recognition, computer vision, etc. He has published nearly 200 papers in AAAI, TSE, PR, and top academic journals and conferences at home and abroad, more than 70 papers of which have been searched by SCI, more than 80 papers have been recommended by CCF, and 2 papers have been highly cited by ESI; the google citation rate is nearly 3000, and the SCI citation rate is nearly 1000; he has been authorized to hold 10 patents, and has won 1 provincial and ministerial level award.