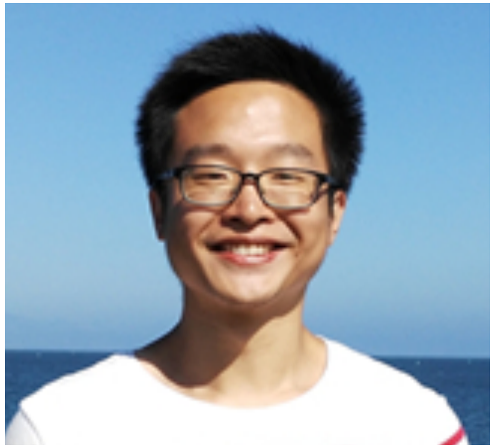
Paper ID: 4410

Self-Supervised Model Training and Selection for Disentangling GANs

Previous title:

InfoGAN-CR: Disentangling Generative Adversarial Networks with Contrastive Regularizers

Zinan Lin



CMU

Kiran Thekumparampil

UIUC

Giulia Fanti

CMU

Sewoong Oh

UW

Generative Adversarial Networks (GANs)

d input noise

z_1
z_2
…
z_d

Generator

k factors

- Hair color
- Rotation
- Background
- Bangs

- How do z_is control the factors?

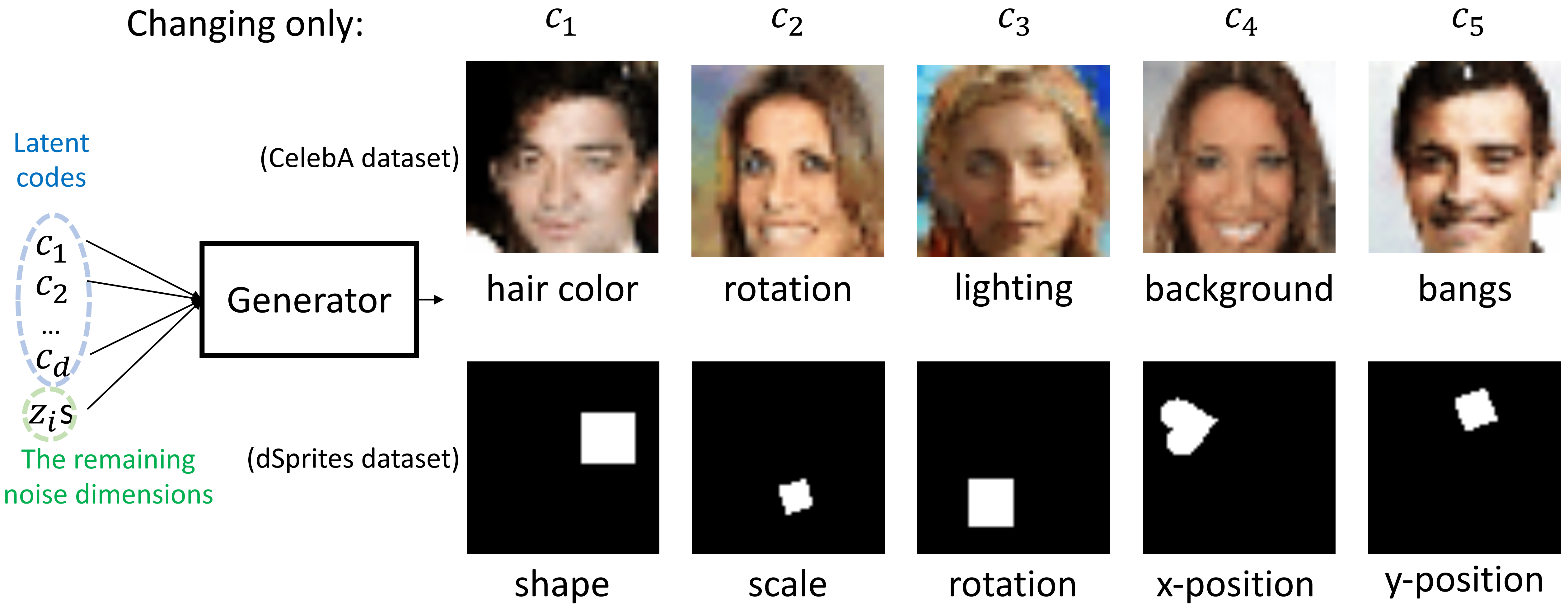
Vanilla GANs

z_1 → Factor_1
z_2 → Factor_2
… …
z_d → Factor_k

Disentangled GANs

$c_1 \rightarrow \text{Factor}_1$
$c_2 \rightarrow \text{Factor}_2$
… …
$c_k \rightarrow \text{Factor}_k$
z_is

Examples of Disentanglement



* CelebA example is generated by InfoGAN-CR. Dsprites example is synthetic for illustration.

Challenges in Learning Disentangled GANs

1. How to train disentangled GANs?

- GANs are good at generating high fidelity images,
  but are reported to be bad at disentanglement (e.g. compared with VAEs)

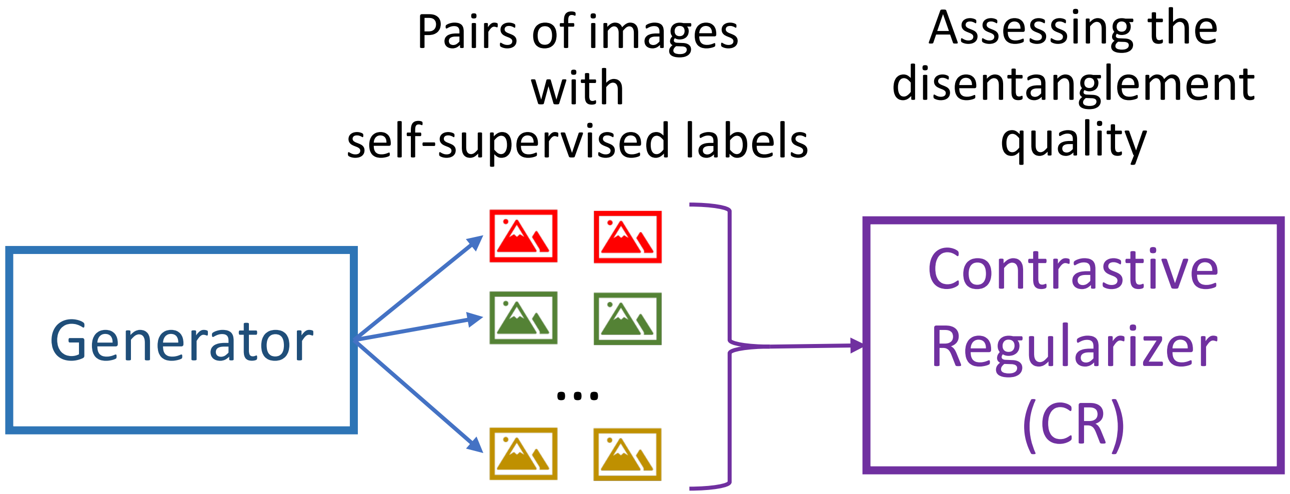
2. How to do *unsupervised* model selection?

- In practice, we don't have ground-truth factor labels
  for selecting the best disentangled models.

Our Solution: Self-Supervision

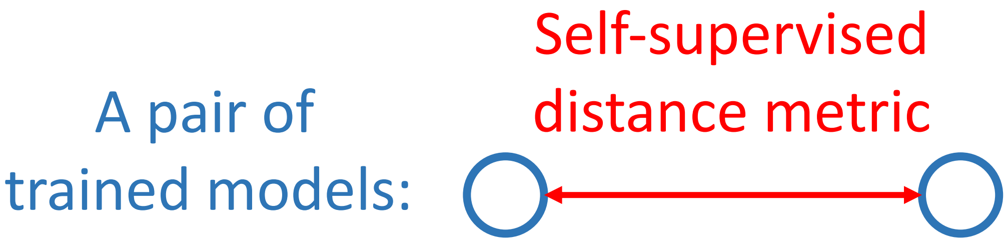
1. How to train disentangled GANs?

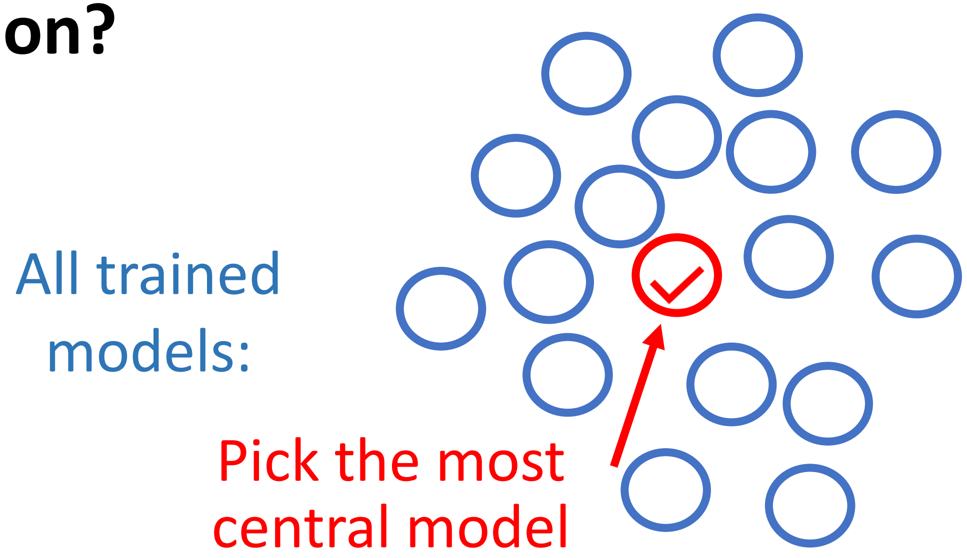
- InfoGAN-CR



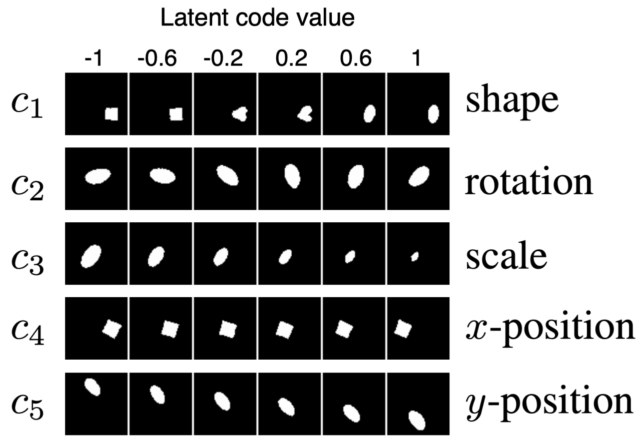
2. How to do *unsupervised* model selection?

- ModelCentrality





Benchmark dSprites Dataset



Supervised hyper-parameter selection

	Model	FactorVAE Metric ↑
VAE	VAE	0.63
	β-TCVAE	0.62
	HFVAE	0.63
	β-VAE	0.63
	CHyVAE	0.77
	FactorVAE	0.82
GAN	InfoGAN	0.59
	InfoGAN (modified)	0.83
	IB-GAN	0.80
	InfoGAN-CR	**0.90 ± 0.01**

Unsupervised model selection

InfoGAN-CR model selected with ModelCentrality	**0.92 ± 0.00**

Code & Paper & More results: https://github.com/fjxmlzn/InfoGAN-CR

Details

1. InfoGAN-CR (model training)
2. ModelCentrality (model selection)

Intuition of Contrastive Regularizer (CR)

- Use two latent codes $(c_1,\ldots,c_i,\ldots,c_k)$, $(c'_1,\ldots,c'_i,\ldots,c'_k)$ to generate a pair of images

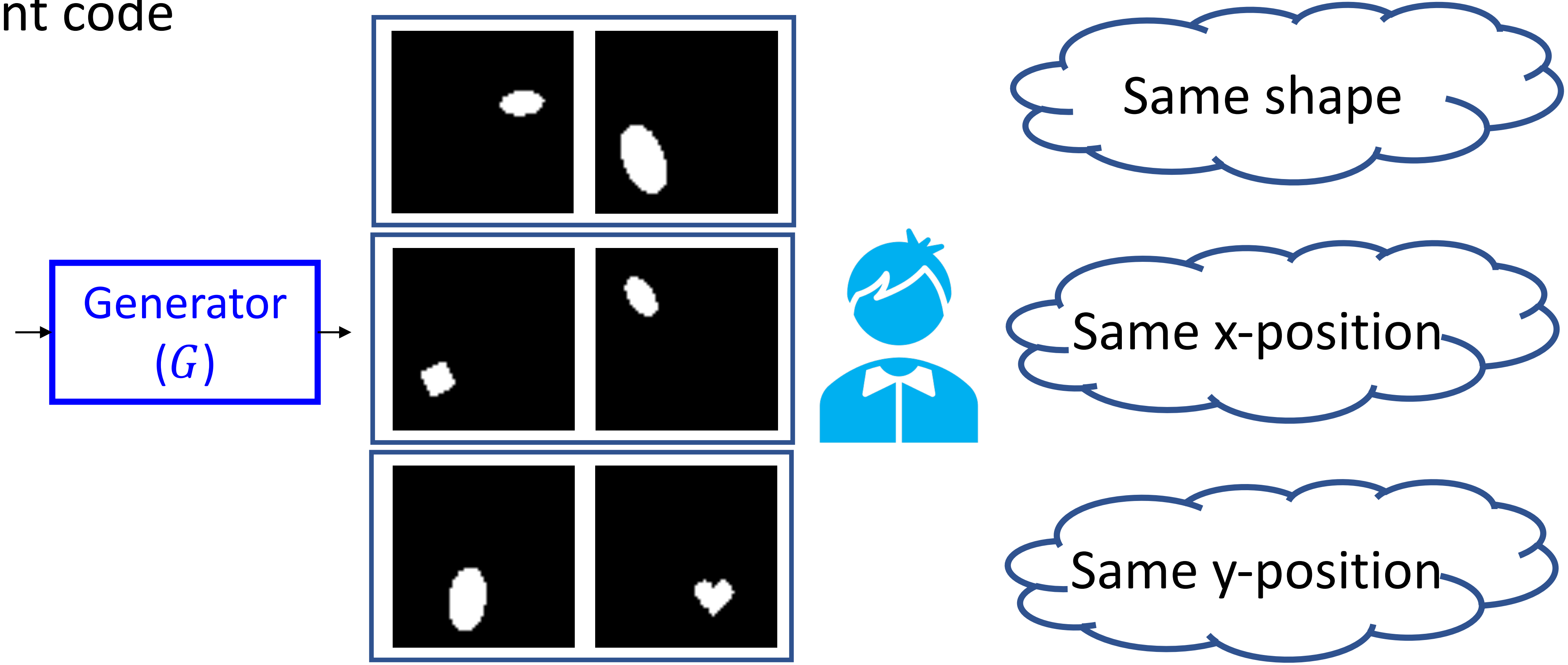
Equal

Same i-th latent code

$i = 1$

$i = 2$

$i = 3$



Intuition of Contrastive Regularizer (CR)

- Use two latent codes $(c_1,\ldots,c_i,\ldots,c_k)$, $(c'_1,\ldots,c'_i,\ldots,c'_k)$ to generate a pair of images

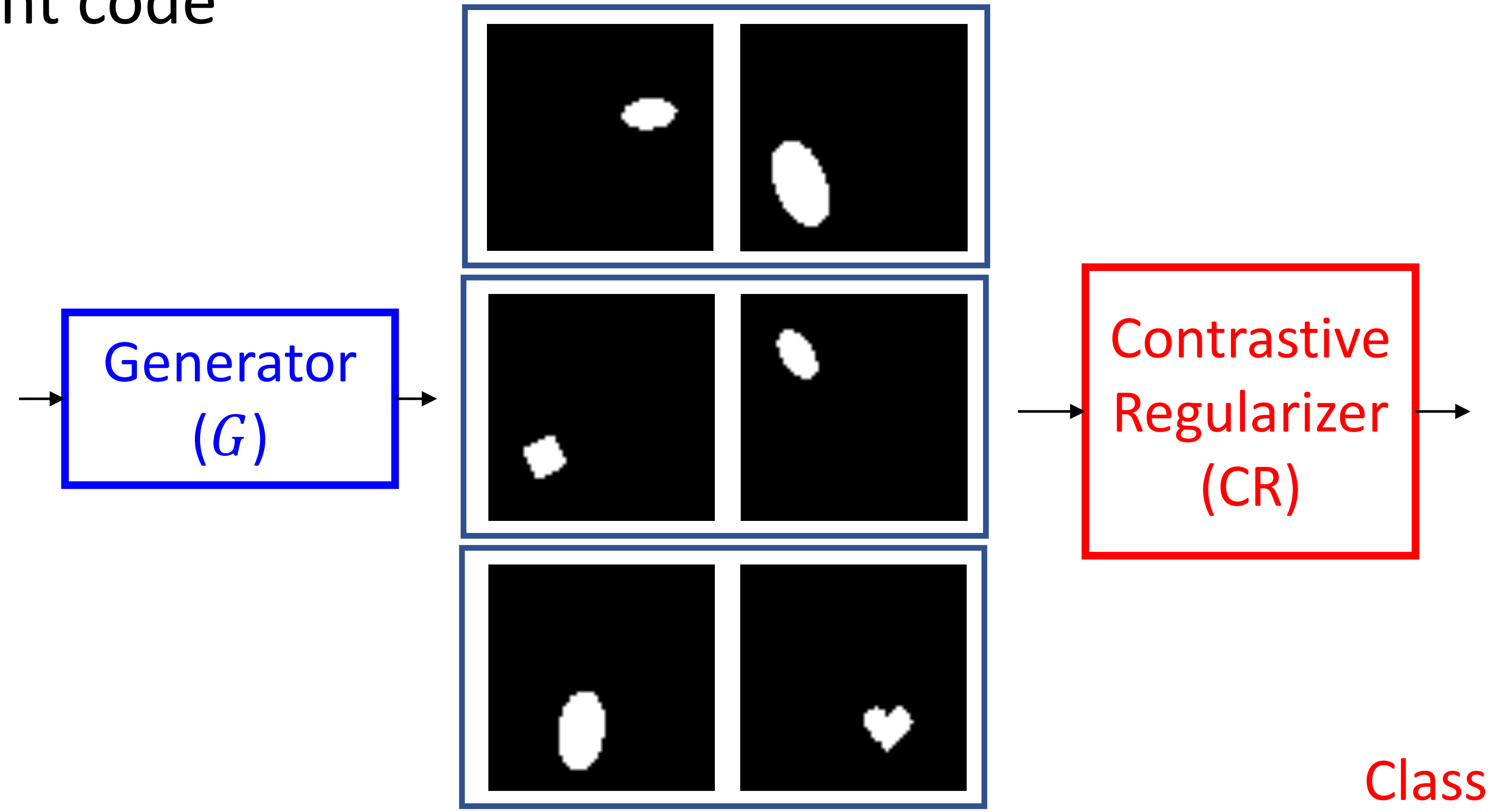
Equal

Same i-th latent code

$i = 1$

$i = 2$

$i = 3$



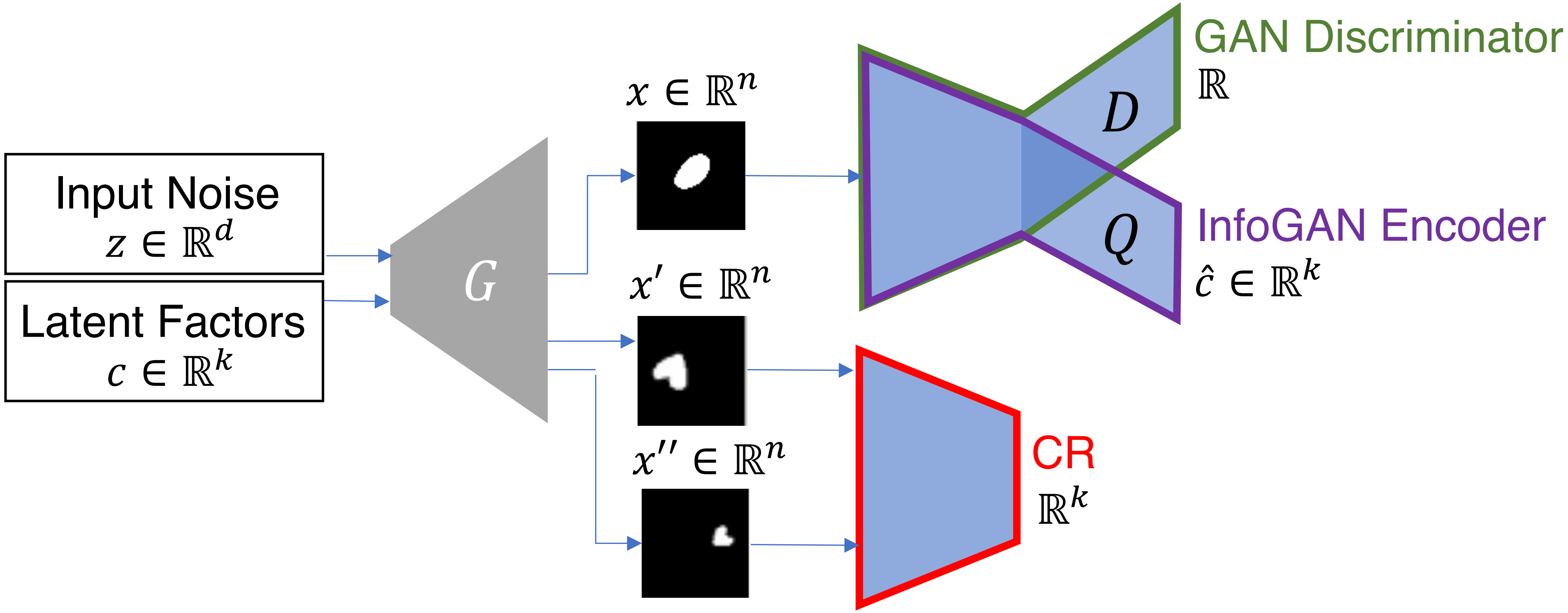
1

2

3

Classification task!

InfoGAN-CR



InfoGAN-CR loss: $$\min_{G,Q,H} \max_{D} \underbrace{L_{adv}(G,D) - \lambda I(G,Q)}_{\text{InfoGAN [1]}} - \alpha L_c(G,H)$$

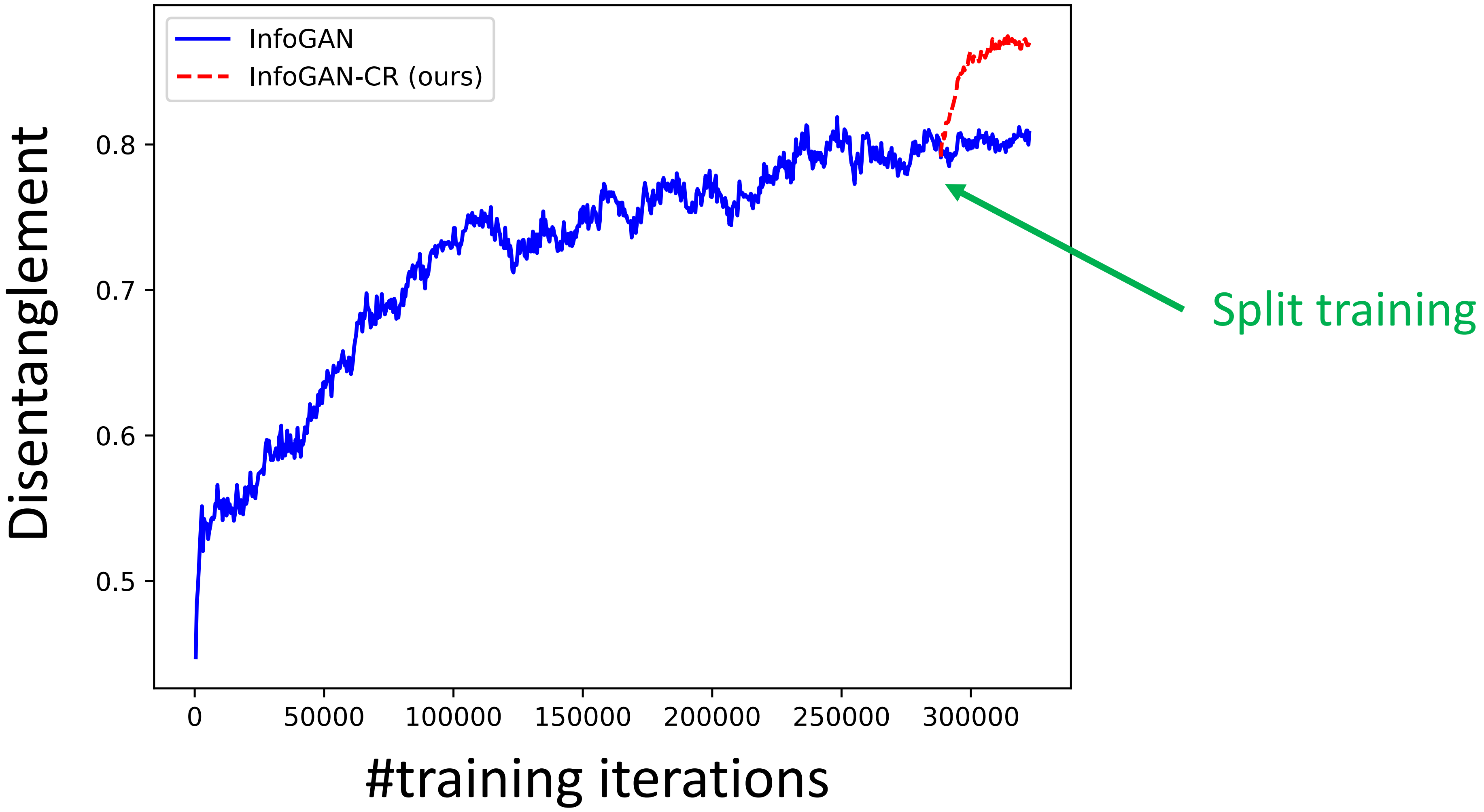
GAN's adversarial loss

Mutual info loss

classification accuracy of CR

[1] InfoGAN. Chen, X., Duan, Y., Houthooft, R., Schulman, J., Sutskever, I., & Abbeel, P. (NeurIPS 2016)

CR Achieves a Further Gain that Cannot be Gotten with InfoGAN Alone



(dSprites dataset)

Details

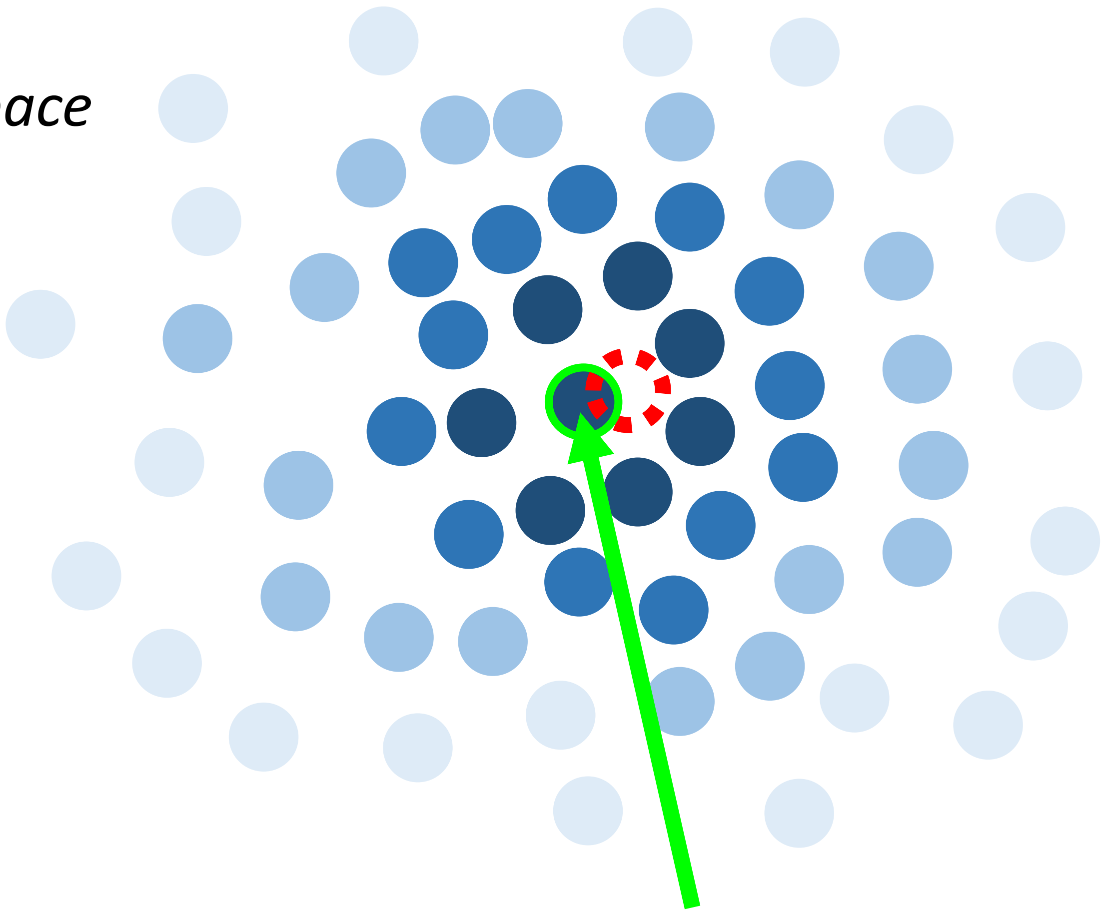
1. InfoGAN-CR (model training)

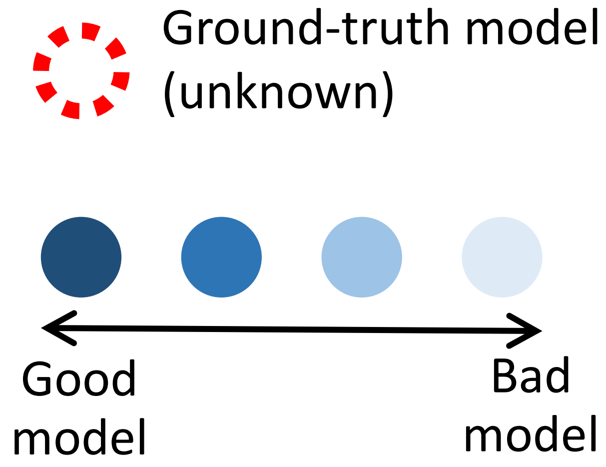
2. ModelCentrality (model selection)

Intuition of ModelCentrality

- Well-disentangled models are close to the (unknown) ground-truth model

Model space

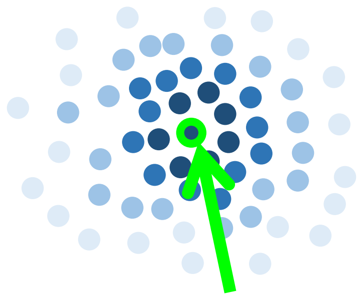




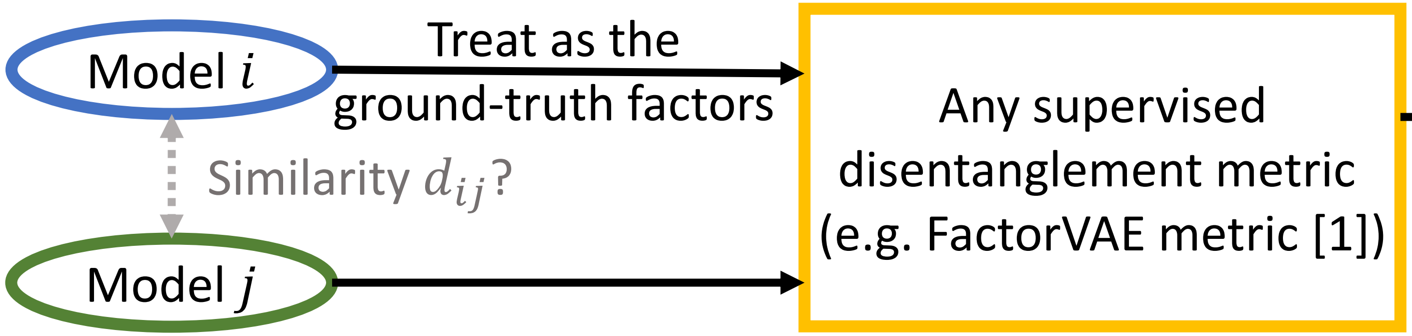
- Idea: pick the model at the <u>*"center"*</u>

ModelCentrality

- How to find the model at the *"center"*?



(1) Two models are close ⇔ Their latent codes control similar factors



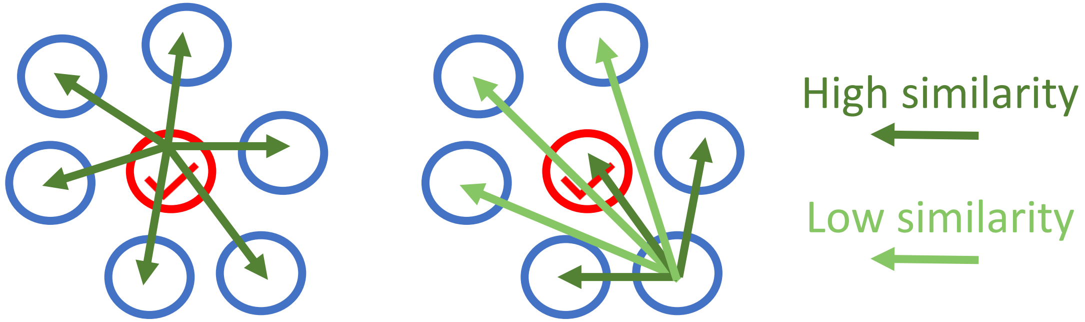
Asymmetric similarity c_{ij}

Similarity $d_{ij} = d_{ji} = (c_{ij} + c_{ji})/2$

(The higher, the more similar)

(2) *ModelCentrality score* $M_i = \frac{1}{N-1}\sum_{j \neq i} d_{ij}$

(3) Selection the model with the highest M_i



[1] Disentangling by Factorising. Kim, H., & Mnih, A. (ICML 2018)

Results of ModelCentrality

- Works well for GANs. Take **InfoGAN-CR** for example:

Method	**FactorVAE Metric ↑**
Supervised hyper-parameter selection	0.90
Unsupervised model selection with UDR Lasso [1]	0.86
Unsupervised model selection with UDR Spearman [1]	0.84
Unsupervised model selection with ModelCentrality	**0.92**

- Also works well for VAEs! Take **FactorVAE** for example:

Method	**FactorVAE Metric ↑**
Supervised hyper-parameter selection	0.83
Unsupervised model selection with UDR Lasso [1]	0.81
Unsupervised model selection with UDR Spearman [1]	0.79
Unsupervised model selection with ModelCentrality	**0.84**

[1] Unsupervised Model Selection for Variational Disentangled Representation Learning. Duan, S., Matthey, L., Saraiva, A., Watters, N., Burgess, C. P., Lerchner, A., & Higgins, I. (ICLR 2020).

Conclusion

- **InfoGAN-CR**
  - New disentangled GANs, achieving state-of-the-art results
- **ModelCentrality**
  - Unsupervised model selection approach for disentangled *GANs and VAEs*
- **InfoGAN-CR + ModelCentrality**
  - Unsupervised disentangled generative model training and selection package, achieving state-of-the-art results

Code & Paper & More results: https://github.com/fjxmlzn/InfoGAN-CR

Including:

- More theoretical & empirical analysis of InfoGAN & InfoGAN-CR
- Analysis of total correlation for GANs and CR for VAEs
- New challenging disentanglement datasets