Information-theoretic stochastic contrastive conditional GAN: InfoSCC-GAN

Vitaliy Kinakh[1*], **Mariia Drozdova**[1,2], **Guillaume Quétant**[1,2],
Tobias Golling[2] **& Slava Voloshynovskiy**[1*]
[1]Department of Computer Science
[2]Department of Particle Physics
University of Geneva
Switzerland
`{vitaliy.kinakh,svolos}@unige.ch`

Abstract

Conditional generation is a subclass of generative problems where the output of the generation is conditioned by the attribute information. In this paper, we present a stochastic contrastive conditional generative adversarial network (InfoSCC-GAN) with an explorable latent space. The InfoSCC-GAN architecture is based on an unsupervised contrastive encoder built on the InfoNCE paradigm, an attribute classifier and an EigenGAN generator. We propose a novel training method, based on generator regularization using external or internal attributes every n-th iteration, using a pre-trained contrastive encoder and a pre-trained classifier. The proposed InfoSCC-GAN is derived based on an information-theoretic formulation of mutual information maximization between input data and latent space representation as well as latent space and generated data. Thus, we demonstrate a link between the training objective functions and the above information-theoretic formulation. The experimental results show that InfoSCC-GAN outperforms the "vanilla" EigenGAN in the image generation on AFHQ and CelebA datasets. In addition, we investigate the impact of discriminator architectures and loss functions by performing ablation studies. Finally, we demonstrate that thanks to the EigenGAN generator, the proposed framework enjoys a stochastic generation in contrast to vanilla deterministic GANs yet with the independent training of encoder, classifier, and generator in contrast to existing frameworks. Code, experimental results, and demos are available online at github.com/vkinakh/InfoSCC-GAN.

1 Introduction

In this paper, we present a new information-theoretic stochastic contrastive conditional generative adversarial network InfoSCC-GAN. The proposed approach is based on the stochastic generative model EigenGAN [1] with explorable latent spaces, independent contrastive encoder and independent classifier for class label regularization. The EigenGAN baseline generator ensures that the model is truly stochastic. In contrast to other conditional generation methods, our model is based on an independent contrastive encoder and attribute classifier. By using them, we avoid a complex and joint procedure of encoder and classifier training, when the model does not produce realistic images in the early iterations. Also, since we use the encoder pre-trained on the real data, we ensure that it properly contrasts real data and avoids contrasting poorly generated data.

We provide an information-theoretical problem formulation of the proposed model in Section 2.

[*] V. Kinakh and S. Voloshynovskiy are corresponding authors.

35th Conference on Neural Information Processing Systems (NeurIPS 2021), Sydney, Australia.

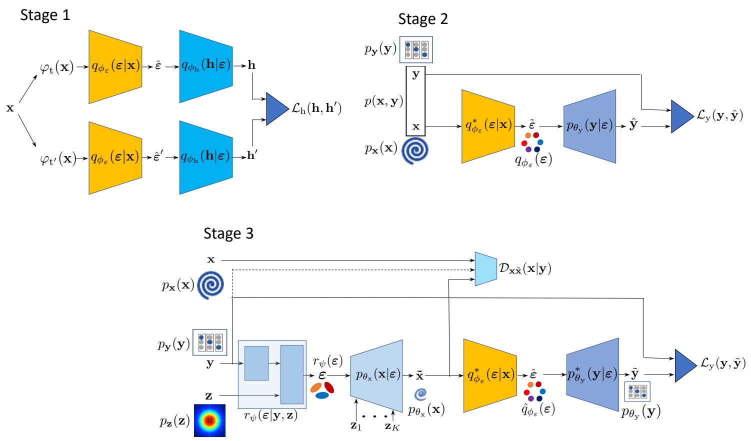
Stage 1
$\mathbf{x}$
$\varphi_{\mathrm{t}}(\mathbf{x})$
$q_{\phi_\varepsilon}(\boldsymbol{\varepsilon}|\mathbf{x})$
$\hat{\varepsilon}$
$q_{\phi_{\mathrm{h}}}(\mathbf{h}|\boldsymbol{\varepsilon})$
$\mathbf{h}$
$\mathcal{L}_{\mathrm{h}}(\mathbf{h}, \mathbf{h}')$
$\varphi_{\mathrm{t}'}(\mathbf{x})$
$q_{\phi_\varepsilon}(\boldsymbol{\varepsilon}|\mathbf{x})$
$\hat{\varepsilon}'$
$q_{\phi_{\mathrm{h}}}(\mathbf{h}|\boldsymbol{\varepsilon})$
$\mathbf{h}'$
Stage 2
$p_{\mathbf{y}}(\mathbf{y})$
$\mathbf{y}$
$p(\mathbf{x}, \mathbf{y})$
$\mathbf{x}$
$p_{\mathbf{x}}(\mathbf{x})$
$q^*_{\phi_\varepsilon}(\boldsymbol{\varepsilon}|\mathbf{x})$
$\tilde{\varepsilon}$
$q_{\phi_\varepsilon}(\boldsymbol{\varepsilon})$
$p_{\theta_{\mathrm{y}}}(\mathbf{y}|\boldsymbol{\varepsilon})$
$\hat{\mathbf{y}}$
$\mathcal{L}_{\mathrm{y}}(\mathbf{y}, \hat{\mathbf{y}})$
Stage 3
$\mathbf{x}$
$p_{\mathbf{x}}(\mathbf{x})$
$\mathcal{D}_{\mathbf{x}\tilde{\mathbf{x}}}(\mathbf{x}|\mathbf{y})$
$p_{\mathbf{y}}(\mathbf{y})$
$\mathbf{y}$
$\mathbf{z}$
$p_{\mathbf{z}}(\mathbf{z})$
$r_\psi(\boldsymbol{\varepsilon}|\mathbf{y}, \mathbf{z})$
$r_\psi(\boldsymbol{\varepsilon})$
$\boldsymbol{\varepsilon}$
$p_{\theta_{\mathrm{x}}}(\mathbf{x}|\boldsymbol{\varepsilon})$
$\mathbf{z}_1$
$\mathbf{z}_K$
$\tilde{\mathbf{x}}$
$p_{\theta_{\mathrm{x}}}(\mathbf{x})$
$q^*_{\phi_\varepsilon}(\boldsymbol{\varepsilon}|\mathbf{x})$
$\hat{\varepsilon}$
$\hat{q}_{\phi_\varepsilon}(\boldsymbol{\varepsilon})$
$p^*_{\theta_{\mathrm{y}}}(\mathbf{y}|\boldsymbol{\varepsilon})$
$\tilde{\mathbf{y}}$
$p_{\theta_{\mathrm{y}}}(\mathbf{y})$
$\mathcal{L}_{\mathrm{y}}(\mathbf{y}, \tilde{\mathbf{y}})$

Figure 1: The proposed InfoSCC-GAN. Stage 1. Contrastive encoder training. Stage 2. Classifier training. Stage 3. Contrastive generator training.

We summarize our contributions in this paper as follows: (i) we proposed a novel stochastic contrastive conditional generative adversarial network (InfoSCC-GAN) for stochastic conditional image generation with explorable latent space. It is based on the EigenGAN generator, an independent contrastive encoder and an independent attributes' classifier; (ii) we introduce a novel classification regularization technique, which is based on updating the model each n-th iteration and updating the generator using adversarial and classification loss separately; (iii) we provide the information-theoretic interpretation of the proposed system; (iv) we perform the ablation studies to determine the contribution of each part of the model to the overall performance.

2 Information-theoretic formulation

The training stages of InfoSCC-GAN are shown in Fig. 1 and explained below.

2.1 The training of the encoder (stage 1)

The encoder training is based on the maximization problem:

$$\hat{\boldsymbol{\phi}_{\boldsymbol{\varepsilon}}} = \underset{\boldsymbol{\phi}_{\boldsymbol{\varepsilon}}}{\operatorname{argmax}} \, I_{\boldsymbol{\phi}_{\boldsymbol{\varepsilon}}}(\mathbf{X}; \mathbf{E}), \tag{1}$$

where $I_{\boldsymbol{\phi}_\varepsilon}(\mathbf{X}; \mathbf{E}) = \mathbb{E}_{p(\mathbf{x}, \boldsymbol{\varepsilon})} \left[\log \frac{q_{\boldsymbol{\phi}_\varepsilon}(\boldsymbol{\varepsilon}|\mathbf{x})}{q_{\boldsymbol{\phi}_\varepsilon}(\boldsymbol{\varepsilon})} \right]$, where $q_{\phi_\varepsilon}(\boldsymbol{\varepsilon}|\mathbf{x})$ denotes the encoder and $q_{\phi_\varepsilon}(\boldsymbol{\varepsilon})$ - the marginal latent space distribution.

In the framework of contrastive learning, (1) is maximized based on the infoNCE framework [2]. In the practical implementation, one can use the approach similar to SimCLR[3], where the inner product between the positive pairs created from the augmented views of the same image is maximized and the inner product between the negative pairs based on different images is minimized[2]. Alternatively, one can use other approaches to learn the representation $\boldsymbol{\varepsilon}$ such as BYOL [4], Barlow Twins [5], etc. without loose of generality of the proposed approach. It should be pointed out that the encoder is trained independently from the decoder in the scope of the considered setup.

[2] The SimCLR training is based on the maximization $I_{\boldsymbol{\phi}_\varepsilon, \phi_n}(\mathbf{X}; \mathbf{H})$, but since $I_{\boldsymbol{\phi}_\varepsilon, \phi_n}(\mathbf{X}; \mathbf{H}) < I_{\boldsymbol{\phi}_\varepsilon}(\mathbf{X}; \mathbf{E})$ one could lower bound (1).

2.2 The training of the class attribute classifier (stage 2)

The class attribute classifier training is based on the maximization problem:

$$\hat{\boldsymbol{\theta}}_{\mathrm{y}} = \underset{\boldsymbol{\theta}_{\mathrm{y}}}{\operatorname{argmax}} \, I_{\boldsymbol{\phi}_{\varepsilon}^{*},\boldsymbol{\theta}_{\mathrm{y}}}(\mathbf{Y};\mathbf{E}), \tag{2}$$

where $I_{\boldsymbol{\phi}_{\varepsilon}^{*},\boldsymbol{\theta}_{\mathrm{y}}}(\mathbf{Y};\mathbf{E}) = H(\mathbf{Y}) - H_{\boldsymbol{\phi}_{\varepsilon}^{*},\boldsymbol{\theta}_{\mathrm{y}}}(\mathbf{Y}|\mathbf{E})$ and $H(\mathbf{Y}) = -\mathbb{E}_{p_{\mathbf{y}}(\mathbf{y})} \log p_{\mathbf{y}}(\mathbf{y})$ is the conditional entropy of $\mathbf{Y}$ and the conditional entropy is defined as $H_{\boldsymbol{\phi}_{\varepsilon}^{*},\boldsymbol{\theta}_{\mathrm{y}}}(\mathbf{Y}|\mathbf{E}) = -\mathbb{E}_{p_{\mathbf{x}}(\mathbf{x})}\left[\mathbb{E}_{q_{\boldsymbol{\phi}_{\varepsilon}^{*}}(\boldsymbol{\varepsilon}|\mathbf{x})}\left[\log p_{\boldsymbol{\theta}_{\mathrm{y}}}(\mathbf{y}|\boldsymbol{\varepsilon})\right]\right]$. Since $H(\mathbf{Y})$ is independent of the parameters of the encoder and classifier, (2) reduces to the lower bound minimization:

$$\hat{\boldsymbol{\theta}}_{\mathrm{y}} = \underset{\boldsymbol{\theta}_{\mathrm{y}}}{\operatorname{argmin}} \, H_{\boldsymbol{\phi}_{\varepsilon}^{*},\boldsymbol{\theta}_{\mathrm{y}}}(\mathbf{Y}|\mathbf{E}), \tag{3}$$

that under the categorical conditional distribution $p_{\boldsymbol{\theta}_{\mathrm{y}}}(\mathbf{y}|\boldsymbol{\varepsilon})$ can be expressed as the categorical cross entropy $\mathcal{L}_{\mathrm{y}}(\mathbf{y},\hat{\mathbf{y}})$.

2.3 The training of the decoder, i.e., the mapper and generator (stage 3)

The decoder is trained first to maximize the mutual information between the class attributes $\tilde{\mathbf{y}}$ predicted from the generated images and true class attributes $\mathbf{y}$:

$$(\hat{\boldsymbol{\theta}}_{\mathrm{x}}, \hat{\boldsymbol{\psi}}) = \underset{\boldsymbol{\theta}_{\mathrm{x}},\boldsymbol{\psi}}{\operatorname{argmax}} \, I_{\boldsymbol{\psi},\boldsymbol{\theta}_{\mathrm{x}},\boldsymbol{\phi}_{\varepsilon}^{*},\boldsymbol{\theta}_{\mathrm{y}}^{*}}(\mathbf{Y};\mathbf{E}), \tag{4}$$

where $I_{\boldsymbol{\psi},\boldsymbol{\theta}_{\mathrm{x}},\boldsymbol{\phi}_{\varepsilon}^{*},\boldsymbol{\theta}_{\mathrm{y}}^{*}}(\mathbf{Y};\mathbf{E}) = H(\mathbf{Y}) - H_{\boldsymbol{\psi},\boldsymbol{\theta}_{\mathrm{x}},\boldsymbol{\phi}_{\varepsilon}^{*},\boldsymbol{\theta}_{\mathrm{y}}^{*}}(\mathbf{Y}|\mathbf{E})$ and $H(\mathbf{Y}) = -\mathbb{E}_{p_{\mathbf{y}}(\mathbf{y})} \log p_{\mathbf{y}}(\mathbf{y})$ and the conditional entropy is defined as $H_{\boldsymbol{\psi},\boldsymbol{\theta}_{\mathrm{x}},\boldsymbol{\phi}_{\varepsilon}^{*},\boldsymbol{\theta}_{\mathrm{y}}^{*}}(\mathbf{Y}|\mathbf{E}) = -\mathbb{E}_{p_{\mathbf{y}}(\mathbf{y})}\left[\mathbb{E}_{p_{\mathbf{z}}(\mathbf{z})}\left[\mathbb{E}_{r_{\boldsymbol{\psi}}(\boldsymbol{\varepsilon}|\mathbf{y},\mathbf{z})}\left[\mathbb{E}_{p_{\boldsymbol{\theta}_{\mathrm{x}}}(\mathbf{x}|\boldsymbol{\varepsilon})}\left[\mathbb{E}_{q_{\boldsymbol{\phi}_{\varepsilon}^{*}}(\boldsymbol{\varepsilon}|\mathbf{x})}\left[\log p_{\boldsymbol{\theta}_{\mathrm{y}}^{*}}(\mathbf{y}|\boldsymbol{\varepsilon})\right]\right]\right]\right]\right]$, $p_{\theta_{\mathrm{y}}}(\mathbf{y}|\boldsymbol{\varepsilon})$ corresponds to the classifier and $q_{\phi_{\varepsilon}}^{*}(\boldsymbol{\varepsilon}|\mathbf{x})$ denotes the pre-trained encoder. Since $H(\mathbf{Y})$ is independent of the parameters of the encoder and classifier, (4) reduces to the lower bound minimization:

$$(\hat{\boldsymbol{\theta}}_{\mathrm{x}}, \hat{\boldsymbol{\psi}}) = \underset{\boldsymbol{\theta}_{\mathrm{x}},\boldsymbol{\psi}}{\operatorname{argmin}} \, H_{\boldsymbol{\psi},\boldsymbol{\theta}_{\mathrm{x}},\boldsymbol{\phi}_{\varepsilon}^{*},\boldsymbol{\theta}_{\mathrm{y}}^{*}}(\mathbf{Y}|\mathbf{E}), \tag{5}$$

that under the categorical conditional distribution $p_{\boldsymbol{\theta}_{y}}(\mathbf{y}|\boldsymbol{\varepsilon})$ can be expressed as the categorical cross entropy $\mathcal{L}_{\mathrm{y}}(\mathbf{y},\tilde{\mathbf{y}})$.

Finally, the decoder should produce samples that follow the distribution of training data $p_{\mathbf{x}}(\mathbf{x})$ that corresponds to the maximization of mutual information:

$$(\hat{\boldsymbol{\theta}}_{\mathrm{x}}, \hat{\boldsymbol{\psi}}) = \underset{\boldsymbol{\theta}_{\mathrm{x}},\boldsymbol{\psi}}{\operatorname{argmax}} \, I_{\boldsymbol{\psi},\boldsymbol{\theta}_{\mathrm{x}}}(\mathbf{X};\mathbf{E}), \tag{6}$$

where $I_{\boldsymbol{\psi},\boldsymbol{\theta}_{\mathrm{x}}}(\mathbf{X};\mathbf{E}) = \mathbb{E}_{p_{\mathbf{x}}(\mathbf{x})}\left[\mathbb{E}_{p_{\mathbf{y}}(\mathbf{y})}\left[\mathbb{E}_{p_{\mathbf{z}}(\mathbf{z})}\left[\mathbb{E}_{r_{\boldsymbol{\psi}}(\boldsymbol{\varepsilon}|\mathbf{y},\mathbf{z})}\left[\mathbb{E}_{p_{\boldsymbol{\theta}_{\mathrm{x}}}(\mathbf{x}|\boldsymbol{\varepsilon})}\left[\log \frac{p_{\boldsymbol{\theta}_{\mathrm{x}}}(\mathbf{x}|\boldsymbol{\varepsilon})}{p_{\mathbf{x}}(\mathbf{x})}\right]\right]\right]\right]\right] = \mathbb{E}_{p_{\mathbf{y}}(\mathbf{y})}\left[\mathbb{E}_{p_{\mathbf{z}}(\mathbf{z})}\left[\mathbb{E}_{r_{\boldsymbol{\psi}}(\boldsymbol{\varepsilon}|\mathbf{y},\mathbf{z})}\left[\mathbb{D}_{\mathrm{KL}}(p_{\boldsymbol{\theta}_{\mathrm{x}}}(\mathbf{x}|\mathbf{E}=\boldsymbol{\varepsilon})||p_{\boldsymbol{\theta}_{\mathrm{x}}}(\mathbf{x}))\right]\right]\right] - \mathbb{D}_{\mathrm{KL}}(p_{\mathbf{x}}(\mathbf{x})||p_{\boldsymbol{\theta}_{\mathrm{x}}}(\mathbf{x}))$, where $p_{\theta_{\mathrm{x}}}(\mathbf{x})$ denotes the distribution of generated samples $\tilde{\mathbf{x}}$. Since $\mathbb{D}_{\mathrm{KL}}(p_{\boldsymbol{\theta}_{\mathrm{x}}}(\mathbf{x}|\mathbf{E}=\boldsymbol{\varepsilon})||p_{\boldsymbol{\theta}_{\mathrm{x}}}(\mathbf{x})) \geq 0$, the maximization of the above mutual information reduces to the minimization:

$$(\hat{\boldsymbol{\theta}}_{\mathrm{x}}, \hat{\boldsymbol{\psi}}) = \underset{\boldsymbol{\theta}_{\mathrm{x}},\boldsymbol{\psi}}{\operatorname{argmin}} \, \mathbb{D}_{\mathrm{KL}}(p_{\mathbf{x}}(\mathbf{x})||p_{\boldsymbol{\theta}_{\mathrm{x}}}(\mathbf{x})). \tag{7}$$

The above discriminator is denoted as $\mathcal{D}_{\mathbf{x}\tilde{\mathbf{x}}}(\mathbf{x})$. At the same time, one can also envision the discriminator conditioned on the attribute class $\mathbf{y}$, e.g., $\mathcal{D}_{\mathbf{x}\tilde{\mathbf{x}}}(\mathbf{x} \mid \mathbf{y})$ that is implemented as a set of discriminators for each subset of generated and original samples defined by the class attributes $\mathbf{y}$.

3 Experiments

In this section, we describe the generation experiments. For the evaluation, we use 3 metrics: Fréchet inception distance (FID) [6], inception score (IS) [7] and Chamfer distance [8]. Since Chamfer

Table 1: Conditional generation results on CelebA dataset with 5 selected attributes.

FID ↓	IS ↑	**Attribute Control Accuracy ↑**				
		Bald	Eyeglasses	Mustache	Wearing Hat	Wearing Necktie
27.84	9.91	93.27%	99.88%	95.68%	94.62%	98.62%

Table 2: Conditional generation results on CelebA dataset with 10 selected attributes.

FID↓	IS↑	**Attribute Control Accuracy↑**				
		Bald	Black Hair	Blond Hair	Brown Hair	Double Chin
32.39	9.04	89.74%	89.61%	86.86%	85.55%	84.82%
		Eyeglasses	Gray Hair	Mustache	Wearing Hat	Wearing Necktie
		99.6%	81.71%	92.27%	92.83%	89.26%

distance works in low dimensional spaces, we compute features of the real and generated image by the pre-trained encoder, then compute the 3D t-SNEs of these features, which are used to compute the Chamfer distance. We perform ablation studies on AFHQ dataset. To determine whether the conditional generated images obey the needed attributes, we use attribute control accuracy. The attribute control accuracy is computed as the percentage of the images for which the output of the attribute classifier is the same as an input attribute. The attribute control accuracy measures how good the generator is at conditional generation.

3.1 EigenGAN

We compare the proposed InfoSCC-GAN with the original EigenGAN [1] on the AFHQ dataset. Our model is based on the same generator while using different inputs and conditional regularization. In the current setup, EigenGAN has 6 layers each with 6 dimensions that are used for interpretable and controllable features exploration. The original EigenGAN achieves FID score of **29.02** and IS of **8.52** after 200000 training iterations on AFHQ dataset using global discriminator and Hinge loss [9]. The EigenGAN does not allow for interpretable feature exploration for the wild animal images. It can be explained by the imbalance since the "wild" animals class includes multiple distinct subclasses such as tiger, lion, fox, and wolf, which are not semantically close.

3.2 Conditional generation

We achieve the best FID score of **11.59**, IS of **11.06** and Chamfer distance **3645** using the InfoSCC-GAN approach after 200000 training iterations using Patch discriminator [10] and LSGAN [11] loss on AFHQ dataset. In the current setup, we have 6 layers with 6 explorable dimensions. The results on CelebA dataset with 5, 10 and 15 attribute labels are presented in Tables 1, 2, 3.

4 Ablation studies

In this section, we describe the ablation studies we have performed on the type of discriminator and the discriminator loss.

Table 3: Conditional generation results on CelebA dataset with 15 selected attributes.

FID↓	IS↑	**Attribute Control Accuracy↑**				
		Bald	Blurry	Chubby	Double Chin	Eyeglasses
34.97	8.87	83.6%	96.46%	80.1%	95.74%	98.11%
		Goatee	Gray Hair	Mustache	Narrow Eyes	Pale Skin
		89.09%	90.78%	87.64%	74.22%	86.91%
		Receding Hairline	Rosy Chicks	Sideburns	Wearing Hat	Wearing Necktie
		86.46%	78.88%	74.9%	97.64%	94.87%

Table 4: Discriminator ablation studies.

Discriminator	Loss	FID ↓	IS ↑	Chamfer distance ↓
Global	Hinge	13.08	10.71	4030
Global	Non saturating	25.62	10.33	28595
Global	LSGAN	29.02	9.89	45583
Patch	Hinge	15.95	10.51	7327
Patch	Non saturating	14.83	10.21	5114
Patch	LSGAN	**11.59**	**11.06**	**3645**

4.1 Discriminator ablation studies

In this section, we describe the discriminator and loss ablation studies. We compare two discriminators: global discriminator and patch discriminator. The global discriminator outputs one value that is the probability of the image being real. The architecture of the global discriminator is inspired by the EigenGAN paper. The patch discriminator outputs a tensor of values that represent the probability of the image patch being real, the architecture of the patch discriminator is inspired by the pix2pix GAN [10]. We compare these discriminators in combination with discriminator losses: Hinge loss, non-saturating loss and LSGAN. The results of the studies are presented in Table. 4. For all of the discriminators and losses, used in the study, the attribute control accuracy is in the range of 99-100%.

5 Conclusions

In this paper, we propose a novel stochastic contrastive conditional GAN InfoSCC-GAN, which produces stochastic conditional image generation with an explorable latent space. We provide the information-theoretical formulation of the proposed system. Unlike other contrastive image generation approaches, our method is truly a stochastic generator, that is controlled by the class attributes and by the set of stochastic parameters. We apply a novel training methodology based on using a pre-trained unsupervised contrastive encoder and a pre-trained classifier with every n-th iteration using a classification regularization. We propose an information-theoretical interpretation of the proposed system. We propose a novel attribute selection approach based on clustering embeddings computed using an encoder. The proposed model outperforms "vanilla" EigenGAN on AFHQ dataset, while it also provides conditional image generation. We have performed ablations studies to determine the best setup for conditional image generation. Finally, we have performed experiments on AFHQ and CelebA datasets.

Acknowledgments and Disclosure of Funding

This research was partially funded by the SNF Sinergia project (CRSII5-193716) Robust deep density models for high-energy particle physics and solar flare analysis (RODEM).

References

[1] Zhenliang He, Meina Kan, and S. Shan. Eigengan: Layer-wise eigen-learning for gans. *ArXiv*, abs/2104.12476, 2021.

[2] Aäron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *ArXiv*, abs/1807.03748, 2018.

[3] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey E. Hinton. A simple framework for contrastive learning of visual representations. *ArXiv*, abs/2002.05709, 2020.

[4] Jean-Bastien Grill, Florian Strub, Florent Altch'e, C. Tallec, Pierre H. Richemond, Elena Buchatskaya, Carl Doersch, B. A. Pires, Z. Guo, M. G. Azar, Bilal Piot, K. Kavukcuoglu, R. Munos, and Michal Valko. Bootstrap your own latent: A new approach to self-supervised learning. *ArXiv*, abs/2006.07733, 2020.

[5] J. Zbontar, L. Jing, Ishan Misra, Y. LeCun, and Stéphane Deny. Barlow twins: Self-supervised learning via redundancy reduction. In *ICML*, 2021.

[6] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *NIPS*, 2017.

[7] Tim Salimans, I. Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. In *NIPS*, 2016.

[8] Nikhila Ravi, Jeremy Reizenstein, David Novotny, Taylor Gordon, Wan-Yen Lo, Justin Johnson, and Georgia Gkioxari. Accelerating 3d deep learning with pytorch3d. *arXiv:2007.08501*, 2020.

[9] Jae Hyun Lim and J. C. Ye. Geometric gan. *ArXiv*, abs/1705.02894, 2017.

[10] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5967–5976, 2017.

[11] Xudong Mao, Qing Li, Haoran Xie, Raymond Y. K. Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks. *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2813–2821, 2017.