Supplementary Material for the Paper: Neural Characteristic Function Learning for Conditional Image Generation

Shengxi Li[1], Jialu Zhang[1], Yifei Li[1], Mai Xu*[1], Xin Deng[2], Li Li[1]

[1]School of Electronic and Information Engineering, Beihang University, Beijing, China
[2]School of Cyber Science and Technology, Beihang University, Beijing, China
{LiShengxi, JialuZhang, leafy, MaiXu, cindydeng, lili2005}@buaa.edu.cn

S1. Proofs of All Lemmas

We first list necessary equations from our manuscript, for the ease of following proofs in the supplementary material.

- $\mathcal{D}_{\mathcal{T}}(\mathcal{V}||\widetilde{\mathcal{V}})$ represents the CF discrepancy between two distributions and is formulated by

$$\begin{aligned}&\mathcal{D}^2_{\mathcal{T}}(\mathcal{V}||\widetilde{\mathcal{V}})\\&=\frac{1}{k}\sum_{i=1}^{k}\big(\Phi_{\mathcal{V}}(\mathbf{t}_i)-\Phi_{\widetilde{\mathcal{V}}}(\mathbf{t}_i)\big)\big(\Phi^*_{\mathcal{V}}(\mathbf{t}_i)-\Phi^*_{\widetilde{\mathcal{V}}}(\mathbf{t}_i)\big)\\&=\frac{1}{k}\sum_{i=1}^{k}\big(\Phi_{\mathcal{X},\mathcal{Y}}(\mathbf{t}_i)-\Phi_{\widetilde{\mathcal{X}},\widetilde{\mathcal{Y}}}(\mathbf{t}_i)\big)\big(\Phi^*_{\mathcal{X},\mathcal{Y}}(\mathbf{t}_i)-\Phi^*_{\widetilde{\mathcal{X}},\widetilde{\mathcal{Y}}}(\mathbf{t}_i)\big)\\&=\mathcal{D}^2_{\mathcal{T}}(\mathcal{X},\mathcal{Y}||\widetilde{\mathcal{X}},\widetilde{\mathcal{Y}}).\end{aligned}\tag{1}$$

- $\mathcal{L}(\mathcal{V}||\widetilde{\mathcal{V}})$ denotes the efficient metric in measuring the discrepancy between two distributions:

$$\mathcal{L}(\mathcal{V}||\widetilde{\mathcal{V}})=\max_{f}\mathcal{D}_{\mathcal{F}}(\mathcal{V}||\widetilde{\mathcal{V}}),\tag{2}$$

where $\mathcal{D}_{\mathcal{F}}(\mathcal{V}||\widetilde{\mathcal{V}})$ is formulated by

$$\mathcal{D}_{\mathcal{F}}(\mathcal{V}||\widetilde{\mathcal{V}})=\left(\frac{1}{k}\sum_{i=1}^{k}\big(\Phi^{f_i}_{\mathcal{V}}-\Phi^{f_i}_{\widetilde{\mathcal{V}}}\big)\big(\Phi^{f_i*}_{\mathcal{V}}-\Phi^{f_i*}_{\widetilde{\mathcal{V}}}\big)\right)^{\frac{1}{2}},\tag{3}$$

and

$$\Phi^{f_i}_{\mathcal{V}}=\mathbb{E}_{\mathcal{V}}[e^{jf_i(\mathbf{v})}]=\frac{1}{n}\sum_{i_v}^{n}e^{jf_i(\mathbf{v}_{i_v})}.\tag{4}$$

- Based on the proposed NCF network, the final loss function of our CCF-GAN can be decomposed by

$$\begin{aligned}&\mathcal{L}(\mathcal{X},\mathcal{Y}||\widetilde{\mathcal{X}},\widetilde{\mathcal{Y}})=\max_{f}\mathcal{D}_{\mathcal{F}}(\mathcal{X},\mathcal{Y}||\widetilde{\mathcal{X}},\widetilde{\mathcal{Y}})\\&=\left(\frac{1}{k}\sum_{i=1}^{k}\big(\Phi^{f_i}_{\mathcal{X},\mathcal{Y}}(\mathbf{t}^i_y)-\Phi^{f_i}_{\widetilde{\mathcal{X}},\widetilde{\mathcal{Y}}}(\mathbf{t}^i_y)\big)\big(\Phi^{f_i*}_{\mathcal{X},\mathcal{Y}}(\mathbf{t}^i_y)-\Phi^{f_i*}_{\widetilde{\mathcal{X}},\widetilde{\mathcal{Y}}}(\mathbf{t}^i_y)\big)\right)^{\frac{1}{2}},\end{aligned}\tag{5}$$

where ECF of the joint distribution has the following form:

$$\Phi^{f}_{\mathcal{X},\mathcal{Y}}(\mathbf{t}_y)=\frac{1}{n}\sum_{i_x=1}^{n}\sum_{i_y=1}^{c}e^{j\mathbf{t}_y^T\mathbf{y}_{i_y}}p(\mathbf{y}_{i_y}|\mathbf{x}_{i_x})e^{jf(\mathbf{x}_{i_x})}.\tag{6}$$

S1.1. Proof of Lemma 1

Lemma 1. *For any two random variables $\mathcal{V},\widetilde{\mathcal{V}}\in\mathbb{R}^d$, $\mathcal{L}(\mathcal{V}||\widetilde{\mathcal{V}})\geq\mathcal{D}_{\mathcal{T}}(\mathcal{V}||\widetilde{\mathcal{V}})$ for any $\mathcal{T}$, where $\mathcal{D}_{\mathcal{T}}(\mathcal{V}||\widetilde{\mathcal{V}})$ is defined in* (1).

Proof. Since $\Phi_{\mathcal{V}}(\mathbf{t}_i)$ is bounded by $|\Phi_{\mathcal{V}}(\mathbf{t}_i)|\leq 1$, $\mathcal{D}_{\mathcal{T}}(\mathcal{V}||\widetilde{\mathcal{V}})$ reaches its maximum at $\{\mathbf{t}^{\dagger}_i\}_{i=1}^k$. In other words, the set of samples $\{\mathbf{t}^{\dagger}_i\}_{i=1}^k$ maximally distinguish the CF of the distribution $\mathcal{V}$ from the other one $\widetilde{\mathcal{V}}$.

Furthermore, since f in $\mathcal{D}_{\mathcal{F}}(\mathcal{V}||\widetilde{\mathcal{V}})$ is built upon our NCF network in (3), it is able to fit any linear and non-linear functions. Thus, we choose the NCF function f to be a linear function of $f^{\dagger}_i(\mathbf{v})=(\mathbf{t}^{\dagger}_i)^T\mathbf{v}$. Then, we arrive at

$$\begin{aligned}\mathcal{L}(\mathcal{V}||\widetilde{\mathcal{V}})&=\max_{f}\mathcal{D}_{\mathcal{F}}(\mathcal{V}||\widetilde{\mathcal{V}})\\&\geq\mathcal{D}_{\mathcal{F}^{\dagger}}(\mathcal{V}||\widetilde{\mathcal{V}})=\mathcal{D}_{\mathcal{T}^{\dagger}}(\mathcal{V}||\widetilde{\mathcal{V}})\geq\mathcal{D}_{\mathcal{T}}(\mathcal{V}||\widetilde{\mathcal{V}}),\end{aligned}\tag{7}$$

where $\mathcal{F}^{\dagger}$ and $\mathcal{T}^{\dagger}$ represent the collections of linear projections $\{f^{\dagger}_i(\cdot)\}_{i=1}^k$ in (3) and maximally distinguishing $\{\mathbf{t}^{\dagger}_i\}_{i=1}^k$ in (1), respectively.

This completes the proof of Lemma 1. □

*Corresponding author

S1.2. Proof of Lemma 2

Lemma 2. *If $\mathcal{V}, \widetilde{\mathcal{V}} \in \mathbb{R}^d$ are two random variables, $\mathcal{L}(\mathcal{V}||\widetilde{\mathcal{V}})$ in* (2) *is a valid distance metric.*

Proof. A valid distance metric requires several important properties, namely, non-negativity, uniqueness, symmetry and triangle inequality, which are proved in the sequel.

Non-negativity: Non-negativity means $\mathcal{L}(\mathcal{V}||\widetilde{\mathcal{V}}) \geq 0$. By the definition, for each function f_i we have

$$\big(\Phi_{\mathcal{V}}^{f_i} - \Phi_{\widetilde{\mathcal{V}}}^{f_i}\big)\big(\Phi_{\mathcal{V}}^{f_i *} - \Phi_{\widetilde{\mathcal{V}}}^{f_i *}\big) \geq 0. \tag{8}$$

Thus, we have

$$\mathcal{D}_{\mathcal{F}}(\mathcal{V}||\widetilde{\mathcal{V}}) = \left(\frac{1}{k}\sum_{i=1}^{k}\big(\Phi_{\mathcal{V}}^{f_i} - \Phi_{\widetilde{\mathcal{V}}}^{f_i}\big)\big(\Phi_{\mathcal{V}}^{f_i *} - \Phi_{\widetilde{\mathcal{V}}}^{f_i *}\big)\right)^{\frac{1}{2}} \geq 0, \tag{9}$$

which proves $\mathcal{L}(\mathcal{V}||\widetilde{\mathcal{V}}) = \max_f \mathcal{D}_{\mathcal{F}}(\mathcal{V}||\widetilde{\mathcal{V}}) \geq 0$.

Uniqueness: Uniqueness means $\mathcal{L}(\mathcal{V}||\widetilde{\mathcal{V}}) = 0$ if and only if $\mathcal{V} = \widetilde{\mathcal{V}}$, which ensures that the metric $\mathcal{L}(\mathcal{V}||\widetilde{\mathcal{V}})$ outputs 0 only when the two distributions are the same.

- Sufficiency ($\{\mathcal{V} = \widetilde{\mathcal{V}}\} \Rightarrow \{\mathcal{L}(\mathcal{V}||\widetilde{\mathcal{V}}) = 0\}$): When $\mathcal{V} = \widetilde{\mathcal{V}}$, it is obvious to have $\Phi_{\mathcal{V}}^{f_i} = \Phi_{\widetilde{\mathcal{V}}}^{f_i}$ for any $f_i(\cdot)$. Therefore, we arrive at

$$\begin{aligned}\mathcal{L}(\mathcal{V}||\widetilde{\mathcal{V}}) &= \max_f \mathcal{D}_{\mathcal{F}}(\mathcal{V}||\widetilde{\mathcal{V}}) \\ &= \max_f \left(\frac{1}{k}\sum_{i=1}^{k}\big(\Phi_{\mathcal{V}}^{f_i} - \Phi_{\widetilde{\mathcal{V}}}^{f_i}\big)\big(\Phi_{\mathcal{V}}^{f_i *} - \Phi_{\widetilde{\mathcal{V}}}^{f_i *}\big)\right)^{\frac{1}{2}} = 0,\end{aligned} \tag{10}$$

  which proves the sufficiency.

- Necessity: According to Lemma 1, when $\mathcal{L}(\mathcal{V}||\widetilde{\mathcal{V}}) = 0$,

$$0 = \mathcal{L}(\mathcal{V}||\widetilde{\mathcal{V}}) \geq \mathcal{D}_{\mathcal{T}}(\mathcal{V}||\widetilde{\mathcal{V}}), \forall \mathcal{T}. \tag{11}$$

  This means $\big(\Phi_{\mathcal{V}}(\mathbf{t}_i) - \Phi_{\widetilde{\mathcal{V}}}(\mathbf{t}_i)\big)\big(\Phi_{\mathcal{V}}^*(\mathbf{t}_i) - \Phi_{\widetilde{\mathcal{V}}}^*(\mathbf{t}_i)\big) = 0$ for all $\mathbf{t}_i$. This way, $\Phi_{\mathcal{V}}(\mathbf{t}_i) = \Phi_{\widetilde{\mathcal{V}}}(\mathbf{t}_i)$ for all $\mathbf{t}_i$. Owing to the one-to-one correspondence between a random variable and its ECF, we have $\Phi_{\mathcal{V}}(\mathbf{t}) = \Phi_{\widetilde{\mathcal{V}}}(\mathbf{t})$ for all $\mathbf{t}$ if and only if $\mathcal{V} = \widetilde{\mathcal{V}}$. Therefore, $\mathcal{V} = \widetilde{\mathcal{V}}$ when $\mathcal{L}(\mathcal{V}||\widetilde{\mathcal{V}}) = 0$, which proves the necessity.

Symmetry: Symmetry means $\mathcal{L}(\mathcal{V}||\widetilde{\mathcal{V}}) = \mathcal{L}(\mathcal{V}||\widetilde{\mathcal{V}})$. It is obvious that

$$\begin{aligned}\mathcal{L}(\mathcal{V}||\widetilde{\mathcal{V}}) &= \max_f \mathcal{D}_{\mathcal{F}}(\mathcal{V}||\widetilde{\mathcal{V}}) \\ &= \max_f \left(\frac{1}{k}\sum_{i=1}^{k}\big(\Phi_{\mathcal{V}}^{f_i} - \Phi_{\widetilde{\mathcal{V}}}^{f_i}\big)\big(\Phi_{\mathcal{V}}^{f_i *} - \Phi_{\widetilde{\mathcal{V}}}^{f_i *}\big)\right)^{\frac{1}{2}} \\ &= \max_f \left(\frac{1}{k}\sum_{i=1}^{k}\big(\Phi_{\widetilde{\mathcal{V}}}^{f_i} - \Phi_{\mathcal{V}}^{f_i}\big)\big(\Phi_{\widetilde{\mathcal{V}}}^{f_i *} - \Phi_{\mathcal{V}}^{f_i *}\big)\right)^{\frac{1}{2}} \\ &= \max_f \mathcal{D}_{\mathcal{F}}(\widetilde{\mathcal{V}}||\mathcal{V}) = \mathcal{L}(\widetilde{\mathcal{V}}||\mathcal{V}),\end{aligned} \tag{12}$$

which proves the symmetry.

Triangle inequality: Triangle inequality ensures the relationship of $\mathcal{L}(\widetilde{\mathcal{V}}||\mathcal{V}) \leq \mathcal{L}(\mathcal{V}||\mathcal{Z}) + \mathcal{L}(\mathcal{Z}||\widetilde{\mathcal{V}})$ for any random variables $\mathcal{V}, \widetilde{\mathcal{V}}, \mathcal{Z}$, which ensures the smooth convergence when minimising the metric $\mathcal{L}(\widetilde{\mathcal{V}}||\mathcal{V})$. To prove the triangle inequality with the additional random variable $\mathcal{Z}$, we have

$$\begin{aligned}\mathcal{D}_{\mathcal{F}}(\mathcal{V}||\widetilde{\mathcal{V}}) &= \left(\frac{1}{k}\sum_{i=1}^{k}|\Phi_{\mathcal{V}}^{f_i} - \Phi_{\widetilde{\mathcal{V}}}^{f_i}|^2\right)^{\frac{1}{2}} \\ &= \left(\frac{1}{k}\sum_{i=1}^{k}|\Phi_{\mathcal{V}}^{f_i} - \Phi_{\mathcal{Z}}^{f_i} + \Phi_{\mathcal{Z}}^{f_i} - \Phi_{\widetilde{\mathcal{V}}}^{f_i}|^2\right)^{\frac{1}{2}} \\ &\leq \left(\frac{1}{k}\sum_{i=1}^{k}|\Phi_{\mathcal{V}}^{f_i} - \Phi_{\mathcal{Z}}^{f_i}|^2\right)^{\frac{1}{2}} + \left(\frac{1}{k}\sum_{i=1}^{k}|\Phi_{\mathcal{Z}}^{f_i} - \Phi_{\widetilde{\mathcal{V}}}^{f_i}|^2\right)^{\frac{1}{2}} \\ &= \mathcal{D}_{\mathcal{F}}(\mathcal{V}||\mathcal{Z}) + \mathcal{D}_{\mathcal{F}}(\mathcal{Z}||\widetilde{\mathcal{V}}),\end{aligned} \tag{13}$$

where the inequality holds by the Minkowski inequality. For the convenience, in (13), $|\cdot|^2$ represents the conjugate square of the complex-valued numbers, namely, $|c|^2 = c \cdot c^*$.

Furthermore, we assume that $\mathcal{D}_{\mathcal{F}}(\mathcal{V}||\widetilde{\mathcal{V}})$, $\mathcal{D}_{\mathcal{F}}(\mathcal{V}||\mathcal{Z})$ and $\mathcal{D}_{\mathcal{F}}(\mathcal{Z}||\widetilde{\mathcal{V}})$ reach the maximum at collections $\mathcal{F}^{\mathcal{V},\widetilde{\mathcal{V}}} = \{f_i^{\mathcal{V},\widetilde{\mathcal{V}}}\}_{i=1}^k$, $\mathcal{F}^{\mathcal{V},\mathcal{Z}} = \{f_i^{\mathcal{V},\mathcal{Z}}\}_{i=1}^k$ and $\mathcal{F}^{\mathcal{Z},\widetilde{\mathcal{V}}} = \{f_i^{\mathcal{Z},\widetilde{\mathcal{V}}}\}_{i=1}^k$, respectively. Based on (13), we have the following relationships

$$\begin{aligned}\max_f \mathcal{D}_{\mathcal{F}}(\mathcal{V}||\widetilde{\mathcal{V}}) &= \mathcal{D}_{\mathcal{F}^{\mathcal{V},\widetilde{\mathcal{V}}}}(\mathcal{V}||\widetilde{\mathcal{V}}) \\ &\leq \mathcal{D}_{\mathcal{F}^{\mathcal{V},\widetilde{\mathcal{V}}}}(\mathcal{V}||\mathcal{Z}) + \mathcal{D}_{\mathcal{F}^{\mathcal{V},\widetilde{\mathcal{V}}}}(\mathcal{Z}||\widetilde{\mathcal{V}}) \\ &\leq \mathcal{D}_{\mathcal{F}^{\mathcal{V},\mathcal{Z}}}(\mathcal{V}||\mathcal{Z}) + \mathcal{D}_{\mathcal{F}^{\mathcal{Z},\widetilde{\mathcal{V}}}}(\mathcal{Z}||\widetilde{\mathcal{V}}) \\ &= \max_f \mathcal{D}_{\mathcal{F}}(\mathcal{V}||\mathcal{Z}) + \max_f \mathcal{D}_{\mathcal{F}}(\mathcal{Z}||\widetilde{\mathcal{V}}).\end{aligned} \tag{14}$$

Thus, we arrive at

$$\mathcal{L}(\mathcal{V}||\widetilde{\mathcal{V}}) \leq \mathcal{L}(\mathcal{V}||\mathcal{Z}) + \mathcal{L}(\mathcal{Z}||\widetilde{\mathcal{V}}). \tag{15}$$

This completes the proof. □

For the completeness, we further provide in Proposition 1 the proof on the final loss function of our CCF-GAN, which employs the classified treatment on the data $\mathcal{X}$ and auxiliary $\mathcal{Y}$ cues.

Proposition 1. *For two sets of random variables* $\{\mathcal{X}, \mathcal{Y}\}$ *and* $\{\widetilde{\mathcal{X}}, \widetilde{\mathcal{Y}}\}$*,* $\mathcal{L}(\mathcal{X}, \mathcal{Y}||\widetilde{\mathcal{X}}, \widetilde{\mathcal{Y}})$ *in* (5) *is a valid distance metric.*

Proof. The non-negativity, symmetry and triangle inequality can be proved following the proof of Lemma 2, by substituting $\mathcal{V} = (\mathcal{X}, \mathcal{Y})$ and $\widetilde{\mathcal{V}} = (\widetilde{\mathcal{X}}, \widetilde{\mathcal{Y}})$. We next prove the remaining uniqueness in the following.

Uniqueness: Uniqueness means $\mathcal{L}(\mathcal{X}, \mathcal{Y}||\widetilde{\mathcal{X}}, \widetilde{\mathcal{Y}}) = 0$ if and only if $\mathcal{X} = \widetilde{\mathcal{X}}$ and $\mathcal{Y} = \widetilde{\mathcal{Y}}$, which ensures that the metric $\mathcal{L}(\mathcal{X}, \mathcal{Y}||\widetilde{\mathcal{X}}, \widetilde{\mathcal{Y}})$ outputs 0 only when the two sets of distributions are the same.

- Sufficiency ($\{\mathcal{X} = \widetilde{\mathcal{X}}, \mathcal{Y} = \widetilde{\mathcal{Y}}\} \Rightarrow \{\mathcal{L}(\mathcal{X}, \mathcal{Y}||\widetilde{\mathcal{X}}, \widetilde{\mathcal{Y}}) = 0\}$): When $\mathcal{X} = \widetilde{\mathcal{X}}$ and $\mathcal{Y} = \widetilde{\mathcal{Y}}$, it is obvious to have $\Phi^{f_i}_{\mathcal{X},\mathcal{Y}}(\mathbf{t}^i_y) = \Phi^{f_i}_{\widetilde{\mathcal{X}},\widetilde{\mathcal{Y}}}(\mathbf{t}^i_y)$ for any $f_i(\cdot)$. Therefore, we arrive at

$$\begin{aligned}&\mathcal{L}(\mathcal{X}, \mathcal{Y}||\widetilde{\mathcal{X}}, \widetilde{\mathcal{Y}}) = \max_f \mathcal{D}_{\mathcal{F}}(\mathcal{X}, \mathcal{Y}||\widetilde{\mathcal{X}}, \widetilde{\mathcal{Y}})\\ &= \left(\frac{1}{k}\sum_{i=1}^{k}\big(\Phi^{f_i}_{\mathcal{X},\mathcal{Y}}(\mathbf{t}^i_y) - \Phi^{f_i}_{\widetilde{\mathcal{X}},\widetilde{\mathcal{Y}}}(\mathbf{t}^i_y)\big)\big(\Phi^{f_i *}_{\mathcal{X},\mathcal{Y}}(\mathbf{t}^i_y) - \Phi^{f_i *}_{\widetilde{\mathcal{X}},\widetilde{\mathcal{Y}}}(\mathbf{t}^i_y)\big)\right)^{\frac{1}{2}}\\ &= 0,\end{aligned} \tag{16}$$

  which proves the sufficiency.

- Necessity: Since $\mathbf{t}_y$ is randomly sampled by fixed rules, we can extend Lemma 1 to $\mathcal{L}(\mathcal{X}, \mathcal{Y}||\widetilde{\mathcal{X}}, \widetilde{\mathcal{Y}}) \geq \mathcal{D}_{\mathcal{T}}(\mathcal{X}, \mathcal{Y}||\widetilde{\mathcal{X}}, \widetilde{\mathcal{Y}})$ for any $\mathcal{T}$. Thus, given $\mathcal{L}(\mathcal{X}, \mathcal{Y}||\widetilde{\mathcal{X}}, \widetilde{\mathcal{Y}}) = 0$, we have

$$\begin{aligned}0 = \mathcal{L}(\mathcal{X}, \mathcal{Y}||\widetilde{\mathcal{X}}, \widetilde{\mathcal{Y}}) &= \max_f \mathcal{D}_{\mathcal{F}}(\mathcal{X}, \mathcal{Y}||\widetilde{\mathcal{X}}, \widetilde{\mathcal{Y}})\\ &\geq \mathcal{D}_{\mathcal{T}}(\mathcal{X}, \mathcal{Y}||\widetilde{\mathcal{X}}, \widetilde{\mathcal{Y}}), \ \ \forall \mathcal{T}.\end{aligned} \tag{17}$$

  This means $\big(\Phi_{\mathcal{X},\mathcal{Y}}(\mathbf{t}_i) - \Phi_{\widetilde{\mathcal{X}},\widetilde{\mathcal{Y}}}(\mathbf{t}_i)\big)\big(\Phi^*_{\mathcal{X},\mathcal{Y}}(\mathbf{t}_i) - \Phi^*_{\widetilde{\mathcal{X}},\widetilde{\mathcal{Y}}}(\mathbf{t}_i)\big) = 0$ for all $\mathbf{t}_i$. This way, $\Phi_{\mathcal{X},\mathcal{Y}}(\mathbf{t}_i) = \Phi_{\widetilde{\mathcal{X}},\widetilde{\mathcal{Y}}}(\mathbf{t}_i)$ for all $\mathbf{t}_i$. Owing to the one-to-one correspondence between a random variable and its ECF, we have $\Phi_{\mathcal{X},\mathcal{Y}}(\mathbf{t}) = \Phi_{\widetilde{\mathcal{X}},\widetilde{\mathcal{Y}}}(\mathbf{t})$ for all $\mathbf{t}$ if and only if $(\mathcal{X}, \mathcal{Y}) = (\widetilde{\mathcal{X}}, \widetilde{\mathcal{Y}})$. Therefore, $\mathcal{X} = \widetilde{\mathcal{X}}$ and $\mathcal{Y} = \widetilde{\mathcal{Y}}$ when $\mathcal{L}(\mathcal{X}, \mathcal{Y}||\widetilde{\mathcal{X}}, \widetilde{\mathcal{Y}}) = 0$, which proves the necessity.

This completes the proof. □

S2. Additional Generation Results

S2.1. Conditional Generation and Interpolations on ImageNet Dataset

Regarding ImageNet dataset, additional conditional generation samples of resolution 128×128 are illustrated in Fig. S1 by proposed CCF-GAN, trained on the Pytorch BigGAN [1] platform, and each row represents one-class-conditioned generation. As can be seen from Fig. S1, our CCF-GAN achieves high-quality generation in various categories, and also retains sufficient diversity within each class. Furthermore, we provide interpolations in Fig. S2. The smooth interpolated images from the first column to the last column verify the desirable continuity of the latent space learnt by our CCF-GAN, instead of merely memorizing the dataset during training process.

S2.2. Comparison of Conditional Generation Samples of VGGFace_c1000 Dataset

The subjective comparison on VGGFace_c1000 dataset is shown in Fig. S3. All the methods were trained based on the BigGAN [1] platform. Each row in Fig. S3 presents the generation from the same class. Again, our CCF-GAN achieves the best performance for the conditional generation.

S2.3. Advancements under StudioGAN [6] Platform

Our CCF-GAN can achieve further improvements of conditional generation, based on the StudioGAN [6] platform, as reported in Table S1. As can be seen from this table, all methods almost witnessed improvements on FID values based on the StudioGAN [6] platform. However, our CCF-GAN still consistently achieves the best generation performances among all the state-of-the-art methods.

References

[1] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large-scale GAN training for high fidelity natural image synthesis. *arXiv preprint arXiv:1809.11096*, 2018.

[2] Mingming Gong, Yanwu Xu, Chunyuan Li, Kun Zhang, and Kayhan Batmanghelich. Twin auxilary classifiers gan. *Advances in neural information processing systems*, 32, 2019.

[3] Liang Hou, Qi Cao, Huawei Shen, Siyuan Pan, Xiaoshuang Li, and Xueqi Cheng. Conditional gans with auxiliary discriminative classifier. In *International Conference on Machine Learning*, pages 8888–8902. PMLR, 2022.

[4] Minguk Kang and Jaesik Park. Contragan: Contrastive learning for conditional image generation. *Advances in Neural Information Processing Systems*, 33:21357–21369, 2020.

[5] Minguk Kang, Woohyeon Shim, Minsu Cho, and Jaesik Park. Rebooting acgan: Auxiliary classifier gans with stable training. *Advances in Neural Information Processing Systems*, 34:23505–23518, 2021.

[6] MinGuk Kang, Joonghyuk Shin, and Jaesik Park. StudioGAN: A Taxonomy and Benchmark of GANs for Image Synthesis. *2206.09479 (arXiv)*, 2022.

[7] Augustus Odena, Christopher Olah, and Jonathon Shlens. Conditional image synthesis with auxiliary classifier gans. In *International conference on machine learning*, pages 2642–2651. PMLR, 2017.

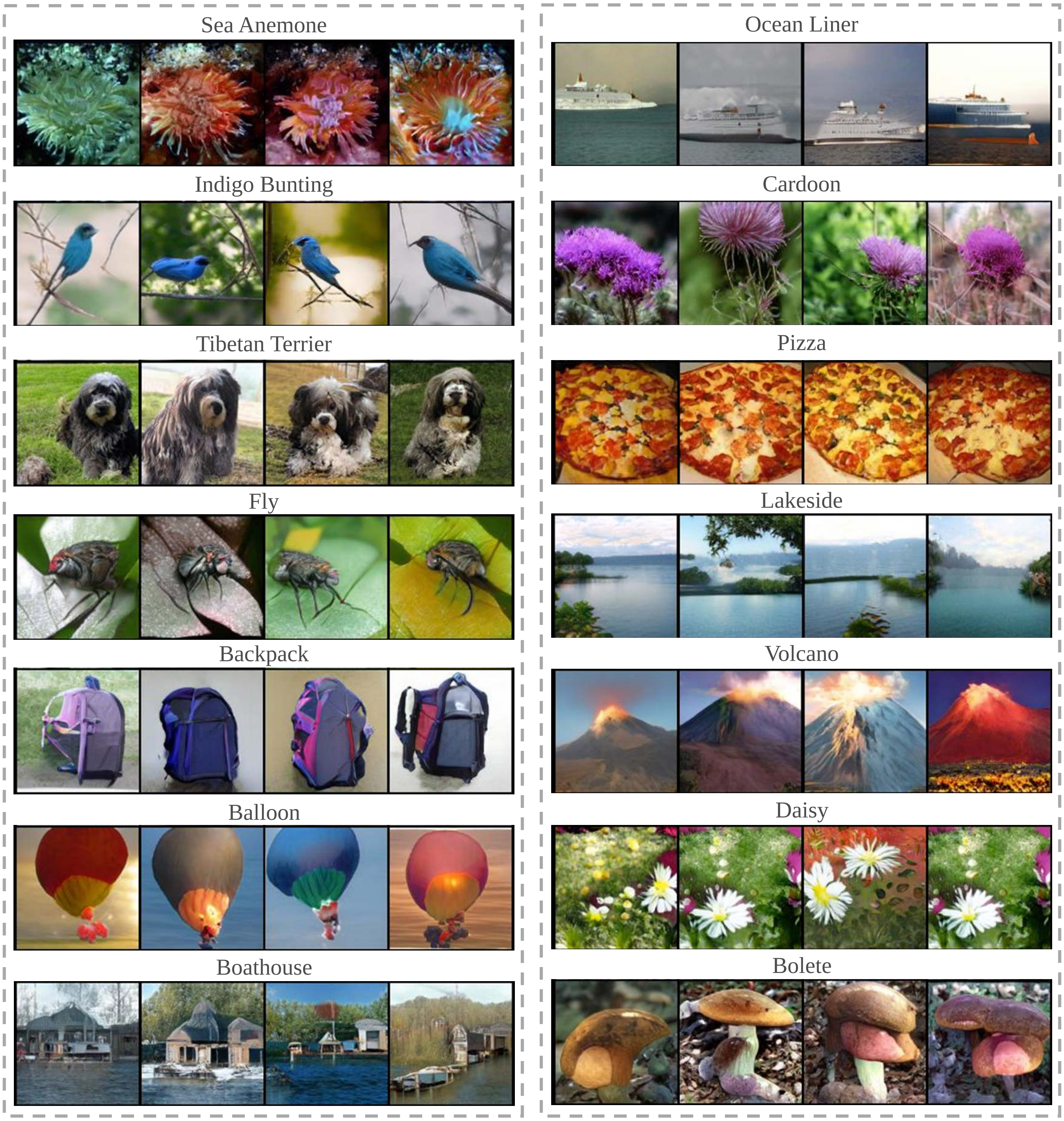

Figure S1: Conditional image generation on ImageNet dataset by the proposed CCF-GAN, trained based on the BigGAN [1] platform. Each row represents one class-conditioned generation.

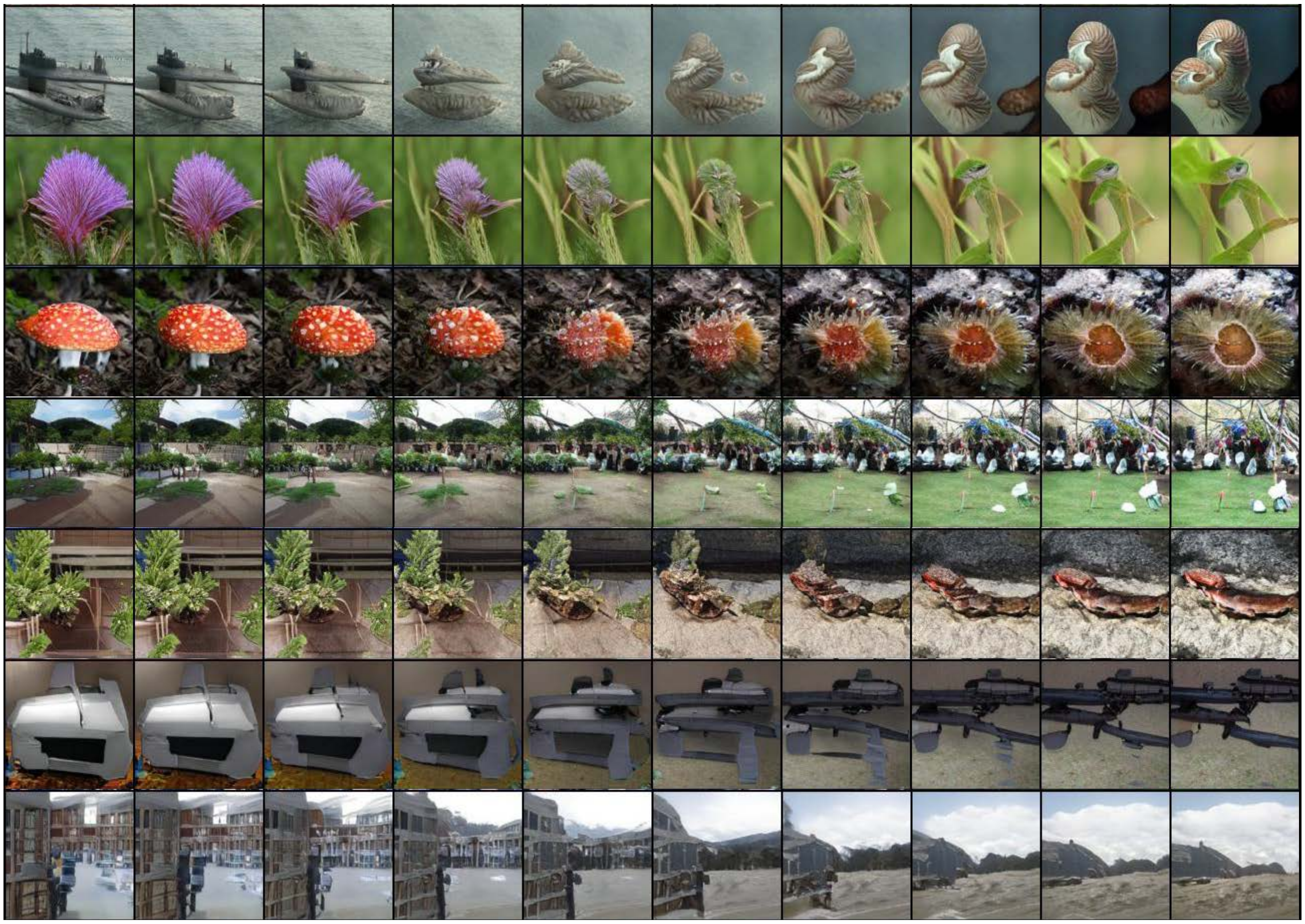

Figure S2: Interpolations of our CCF-GAN across different class labels of ImageNet dataset.

(a) ReACGAN

(b) ADCGAN

(c) CCF-GAN (Ours)

Figure S3: Comparison on conditional image generation on VGGFace_c1000 dataset. Each row represents one class-conditioned generation.

Table S1: Comparison on FID scores on CIFAR10, VGGFace_c1000, ImageNet datasets on Pytorch BigGAN [1] and StudioGAN [6] platforms. Symbol * denotes that the results are reported from [1], whereas † from [2], ‡ from [5], ** for [4] and †† for [3]. Otherwise, we ran the available codes by the corresponding default settings. We denote the best FID by red color and the second best by blue color.

Backbone	Dataset	BigGAN [1]	ContraGAN [4]	ACGAN [7]	TACGAN [2]	ReACGAN [5]	ADCGAN [3]	**CCF-GAN (Ours)**
Pytorch BigGAN [1]	CIFAR10	14.73*	10.60**	8.01	8.42	6.22	7.17	**6.08**
	VGGFace_c1000	24.07†	—	—	13.60†	6.47	7.94	**5.70**
	ImageNet	22.77†	19.44**	184.41†	23.75†	—	16.75††	**11.34**
StudioGAN [6]	CIFAR10	8.08‡	8.22‡	8.45‡	8.01‡	7.88‡	8.42††	**4.71**
	VGGFace_c1000	4.17	6.59	6.76	9.54	5.30	4.32	**3.91**
	ImageNet	16.36‡	25.16‡	25.35‡	—	13.98‡	11.65††	**11.11**