GroupifyVAE: from Group-based Definition to VAE-based Unsupervised Representation Disentanglement
Appendix

A. Preliminary

A.1. Group Theory Basics

Group: A set G together with a binary operation $\circ$ as: $G \times G :\to G$ satisfying the following conditions:

- Associativity: $\forall\, a, b, c \in G, s.t. (a \circ b) \circ c = a \circ (b \circ c)$
- Identity: $\exists\, e \in G,\ s.t. \forall\, a \in G, e \circ a = a \circ e = a$
- Inverse: $\forall\, a \in G, \exists\, a^{-1} \in G : a \circ a^{-1} = a^{-1} \circ a = e$

It is customary to represent a Group with a set G and the binary operation $\circ$ pair $(G, \circ)$. When the binary operation is clear,it is customary to represent Group $(G, \circ)$ as G. In this case, it is customary to use multiplication to represent the binary operation $\circ$, i.e., $a \circ b = ab,\ \forall\, a, b \in G$.

Abelian Group: A group $(G, \circ)$ satisfies the commutative condition: $\forall\, a, b \in G, a \circ b = b \circ a$

Subgroup: A Group $(G, \circ)$, A is a subset of G, if $(A, \circ)$ forms a Group, then A is the subgroup of G.

Equivalence Relation: A is a set, the subset R of set $A \times A = \{(a, b);\ \forall\, a, b \in A\}$ is a **Relation** on set A, and set R satisfying the following conditions: (It is customary represent $(a, b) \in R$ as $a \sim b$)

- Reflexive: $\forall\, a \in A, s.t. a \sim a$
- Symmetric: if $a \sim b$, then $b \sim a$
- Transitive: if $a \sim b$ and $b \sim c$, then $a \sim c$

$\forall\, a \in A$, we represent all the elements in A that equivalent to a as $\overline{a}$, i.e., $\overline{a} = \{b \in A;\ b \sim a\}$, $\overline{a}$ is **Equivalent Class**.

Isomorphism: $(G, \cdot), (G', \circ)$ are two Groups, mapping $f : G \to G'$ is a **Homomorphism** between G and G', if $\forall\, a, b \in G$, $f(a \cdot b) = f(a) \circ f(b)$, if mapping f is a bijection, then mapping f is a Isomorphism between G and G'.

Related Exmaples of Groups:
(i) ***Multiplicative group of integers modulo*** n: n is a positive integer, we define a Equivalence Relation on $\mathbb{Z}$ as $a \sim b \Leftrightarrow n|a - b$, i.e., $a = b(mod\ n)$. The relation divide $\mathbb{Z}$ into n Equivalent Classes: $\overline{0}, \overline{1}, \ldots, \overline{n-1}$, where $\overline{i}$ represents the Equivalent Class containing i, i.e., $\overline{i} = \{m \in \mathbb{Z} | m = i(mod\ n)\}$, let $Z_n = \{\overline{0}, \overline{1}, \ldots, \overline{n-1}\}$, Z_n with binary operation:$\overline{a} + \overline{b} = \overline{a + b}$ forms a Group, it is denoted by $\mathbb{Z}/n\mathbb{Z}$.
(ii) ***n-th root unity Group***: n is a positive integer. $C_n = \{e^{\frac{2\pi i a}{n}} | 0 \leq a \leq n-1\}$, then C_n with complex multiplication forms a Group. For mapping: $f : C_n \to (Z_n, +), e^{\frac{2\pi i a}{n}} \mapsto \overline{a}$, then

$$f\left(e^{\frac{2\pi i a}{n}} \cdot e^{\frac{2\pi i b}{n}}\right) = f\left(e^{\frac{2\pi i a + b}{n}}\right) = \overline{a + b} = \overline{a} + \overline{b} = f\left(e^{\frac{2\pi i a}{n}}\right) \cdot f\left(e^{\frac{2\pi i b}{n}}\right) \tag{1}$$

Therefore, f is a homomorphism, f is a bijection, and f is a isomorphism. In this paper, we designed two constrains to let the encoder and decoder forms a isomorphism mapping f.

Congruence Class: n is a positive integer, we define a Equivalence Relation on $\mathbb{Z}$ as $a \sim b \Leftrightarrow n|a - b$, i.e., $a = b(mod\ n)$. The relation divide Z into n Equivalent Classes: $\overline{0}, \overline{1}, \ldots, \overline{n-1}$, which are congruence classes and also the elements of the multiplicative group of integers modulo n.

Subgroup Generated by A: If A is a nonempty subset of the Group G, the **Set Generated by A**, denoted by $< A >$, is the set defined by $< A >= \{a \in G | a = a_i a_2 \dots a_n \text{with either} \, a_i \in A \; or \; a_i^{-1} \in A\}$, then the set $< A >$ is a subgroup of G.

Symmetry Group and Permutation Group: For a nonempty set Σ, the one-to-one mappings on itself σ called a **Permutation**, $S(\Sigma)$ denotes the set containing all the Permutations on Σ, $S(\Sigma)$ with mapping compound as binary operator forms a Group called Symmetry Group, the subgroup of $S(\Sigma)$ called Permutation Group. In particular, Symmetry Group of set $\{1, 2, \dots, n\}$ is denoted by S_n.

Group Action: Σ is a set, $S(\Sigma)$ is the Symmetry Group on Σ, homomorphism$\forall \, f : G \to S(\Sigma)$ called **permutation representation** on set Σ, then $f(g)$ is a Permutation on Σ called g action on Σ. $\forall \, a \in \Sigma$, define $ga = f(g)a$

Symmetry: A symmetry of a geometric figure is a rearrangement of the figure preserving the arrangement of its sides and vertices as well as its distances and angles.

Exmaple of symmetry: Take rectangle as an example in Figure 1, it is easy to see that a rotation of $180°$ or $360°$ are symmetries for returning a rectangle with the same orientation as the original one and the same relationship among the vertices. Similarly, A reflection either the vertical axis or the horizontal axis can also be seen to be a symmetry.

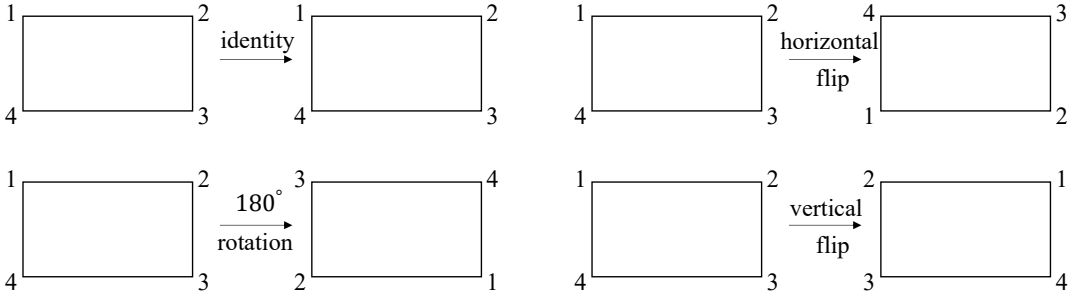


Figure 1. Rigid motions of a rectangle

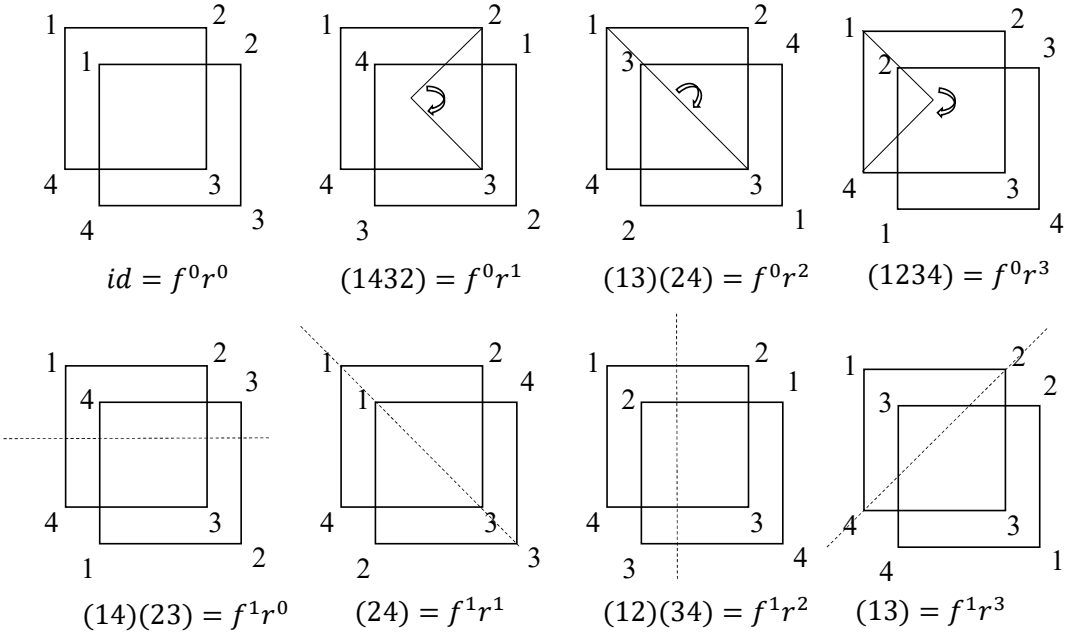


Figure 2. 4-th Dihedral Group (D_4): The groups of symmetries for square.

Rigid motion: A map from the plane to itself preserving the symmetry of an object is called a rigid motion. In particular, rigid motions of a regular n-sided polygon are permutations of the vertices.

nth Dihedral Group: The group of rigid motions of a regular n-sided polygon, which is denoted by D_n, $D_n \subseteq S_n$.

D_4 **as an example**: The vertices of a square are numbered by $\{1, 2, 3, 4\}$, which is in analog with image dataset, the rigid motions is in analog with actions in latent space. We often abbreviate permutation $1 \to 2, 2 \to 3, 3 \to 4, 4 \to 1$ as (1234). The elements of Group D_4 are shown in Figure 2, from which we know that all of the transformations are compounded by basic permutation horizontal flip f and rotate 90 degrees clockwise r. Please note that f and r are in analog with disentangled factors. What's more, The Group can be generated by these basic permutations: $D_4 =< f, r | f^2 = 1, r^4 = 1, fr = r^{-1}f >$. The constrains $f^2 = 1, r^4 = 1, fr = r^{-1}f$ is in analog with the group constrains in this paper.

B. Proof for Theorem 1

Proof. $\Rightarrow$) First, we show that when the existance of isomorphism $G \sim \Phi =< \varphi_1, \varphi_2, \ldots, \varphi_m >$ and $\varphi_i(f(w)) = f(\overline{w + c_i})$ will make the mapping c be equivariant between the actions on W and Z: $g \cdot c(w) = c(g \cdot w)$. Then, we proof that each Z_i is affected only by G_i.

Take $\forall\, g \in G$, $G = \mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z} \times \cdots \times \mathbb{Z}/n\mathbb{Z}$, here we assume $g = \overline{\sum_i k_i e_i}$. In this setting, the group action on W is: $g \cdot w = \overline{w + \sum_i k_i e_i}$, $\forall\, w \in W$. Similarly, the group actions on Z is: $g \cdot z = \overline{z + \sum_i k_i e_i}, \forall\, z \in Z$.

For the group $\Phi =< \varphi_1, \varphi_2, \ldots, \varphi_m >$, where $\varphi_i \in S\{f(w)\}$, $S\{f(w)\}$ is symmetry group of $\{f(w)\}$, the definition of φ_i is:

$$\varphi_i(f(w)) = b^{-1}(\overline{b(f(w)) + e_i}) = b^{-1}(\overline{c(w) + e_i}), \tag{2}$$

where $b^{-1} = d$ denotes the decoder, which can be regard as the invertion of encoder b. And the mapping c is defined as $c = b \circ f$. Assume there is a isomorphism $\tau : G \to \Phi$, we have $ker(\tau) = \{e_G\}$, e_G is the identity element in group G, and e_Φ is identity element in group Φ, so that $\tau(e_G) = e_\Phi$. Let $U = \{e_i \in G | 0 \le i \le m\}$, where e_i is the identity element of dimension i in G.

Assume $\exists i,\ 0 \le i \le m,\ s.t.\ \tau^{-1}(\varphi_i) = g$, $\forall g \in G$ but $g \notin U$, since τ is a bijection,

$$|\Phi| = | < \varphi_1, \varphi_2, \ldots, \varphi_m > | = | < e_1, e_2, \ldots, e_{i-1}, g, e_{i+1}, \ldots, e_m > | < |G|. \tag{3}$$

However, since τ is a isomorphism, we have $|\Phi| = |G|$, which is contradicts the above equation. Therefore, $\forall\, 0 \le i \le m, \tau^{-1}(\varphi_i) \in U$, $\tau(e_i) = \varphi_j$, by rearrange subscript of φ_i. May as well set $\tau(e_i) = \varphi_i$

Therefore, n times compose φ_i, since τ is a isomorphism, we have

$$\varphi_i^n = \tau(e_i)^n = \tau(ne_i) = \tau(e_G) = e_\Phi = id. \tag{4}$$

Recall that another condition: $\varphi_i(f(w)) = f(\overline{w + c_i})$, this condition with n times compose:

$$\varphi_i^n(f(w)) = f(\overline{w + nc_i}) = f(w), \tag{5}$$

which indicate that $c_i = (k+1)e_i$, if $k \neq 0$, then

$$Order(\varphi_i) = \frac{[n, k+1]}{k+1} < n, \tag{6}$$

which contradict with $order(\varphi_i) = n$. Therefore $k = 0$ (Please note that we do not consider the case of mutually exclusive Euler groups, if they are mutually exclusive, after redefining the generator, change the scale, this group is remain the same). Therefore in this case,

$$\varphi_i(f(w)) = f(\overline{w + e_i}), \Phi = \{\varphi_1^{\overline{k_1}} \varphi_2^{\overline{k_2}} \ldots \varphi_n^{\overline{k_n}} | k_i \in Z, \overline{k_i} = k_i \mod n\}. \tag{7}$$

We have

$$\forall\, g \in G, b^{-1}(g \cdot c(w)) = b^{-1}(\overline{\sum_i k_i e_i + c(w)}) = \varphi_1^{k_1} \varphi_2^{k_2} \ldots \varphi_n^{k_n}(f(w)) = f(\overline{w + \sum_i k_i e_i}) = f(g \cdot w), \tag{8}$$

then we have $g \cdot c(w) = b \circ f(g \cdot w) = c(g \cdot w)$, i.e. The map c is equivariant between the actions on W and Z.

Then we proof that each Z_i is affected only by G_i. $\forall\, g_i \in G_i$, for group action on $f(w)$

$$g_i \cdot f(w) = \varphi_i^{\overline{k}}(f(w)) = f(\overline{w + ke_i}) = b^{-1}(\overline{ke_i + c(w)}). \tag{9}$$

Based on the above equation, we have

$$g_i \cdot c(w) = b \circ b^{-1}(\overline{ke_i + c(w)}) = \overline{ke_i + c(w)}. \tag{10}$$

Therefore, Z_i is affected only by g_i for the decomposition $Z = Z_1 \times Z_2 \times \cdots \times Z_n$. Summarizing, the representation Z is disentangled with respect to G.

$\Leftarrow$) Review that definition of φ_i is

$$\varphi_i(f(w)) = b^{-1}(\overline{b(f(w)) + e_i}) = b^{-1}(\overline{c(w) + e_i}), \tag{11}$$

the mapping c is equivariant between the actions on W and Z: $g \cdot c(w) = c(g \cdot w)$, we have

$$\varphi_i(f(w)) = b^{-1}(\overline{c(w + e_i)}). \tag{12}$$

Therefore, $\varphi_i(f(w)) = f(\overline{w + c_i})$, $c = b \circ f$ is satisfied, and we define the mapping τ as:

$$\tau(\varphi) = g : \ \varphi(f(w)) = b^{-1}(\overline{c(w + g)}). \tag{13}$$

Because

$$\varphi_s(\varphi_t(f(w))) = \varphi_s(b^{-1}(\overline{c(w + g_s)})) = \varphi_s(f(w + g_s)) = b^{-1}(\overline{c(w) + g_s + g_t}), \tag{14}$$

and we have

$$\tau(\varphi_s \varphi_t) = g_s + g_t = \tau(\varphi_s) + \tau(\varphi_t). \tag{15}$$

Consequently, τ is a homomorphism. Due to $e_\Phi(f(w)) = f(w) = b^{-1}(c(w))$, $\tau(e_\Phi) = e_G$. Assume that there exist $\varphi_e \neq e_\Phi$, $s.t. \tau(\varphi_e) = e_G$, we have

$$\varphi_e(f(w)) = b^{-1}(c(w)) = f(w). \tag{16}$$

Therefore, $\varphi_e = id = e_\Phi$, and $\tau^{-1}(e_G) = \{e_\Phi\}$, τ is injective. We take $\forall\, g \in G, g = (\overline{k_1}, \overline{k_2}, \ldots, \overline{k_m})$, then

$$b^{-1}(c(w + g)) = b^{-1}(c(w + \overline{k_1}e_1 + \overline{k_2}e_2 + \cdots + \overline{k_m}e_m)) = \cdots = \varphi_1 \circ \cdots \circ \varphi_m(f(w)). \tag{17}$$

Moreover, $\varphi_1 \circ \cdots \circ \varphi_m \in \Phi$, τ is surjective, and τ is a isomorphism.

Q.E.D.

C. Proof for Theorem 2

Proof. The theorem is proofed in the following two steps, In the first step, we prove that the set generated by the transformation $\Phi =< \varphi_i | \varphi_i \varphi_j = \varphi_j \varphi_i, \varphi_i^n = e_\Phi = id, \forall\, 0 \leq i, j \leq m >$ forms a group under the compound operation. In the second step, we prove that the necessary and sufficient condition for the existence of isomorphism is that the Abel constraint and the Order constraint are satisfied at the same time.

Step 1: To verify that a set forms a group under a certain operation, we only need to verify that the elements in the set satisfied the following three requirements: 1. Associativity 2. Identity 3. Inverse. We take $\forall\, \varphi_i \in S\{f(w)\}$, $S\{f(w)\}$ is the symmetry group on images set $\{f(w)\}$. Therefore φ_i is a mapping: $\varphi_i : X \to X$, then $\forall\, \varphi_i, \varphi_j, \varphi_k \in \{\varphi_i, 0 \leq i \leq m\}$,

$$(\varphi_i \varphi_j)\varphi_k = \varphi_i(\varphi_j \varphi_k), \tag{18}$$

thus generators of Φ satisfy Associativity. However, the dose the element of Φ satisfy Associativity is unknown, we take $\forall\, \varphi_s, \varphi_t, \varphi_l \in \Phi$,

$$(\varphi_t \varphi_s)\varphi_l = \left(\prod_{s_i} \varphi_{s_i}^{\overline{k_{s_i}}} \prod_{t_i} \varphi_{t_i}^{\overline{k_{t_i}}}\right) \prod_{l_i} \varphi_{l_i}^{\overline{k_{l_i}}} = \prod_{s_i} \varphi_{s_i}^{\overline{k_{s_i}}} \left(\prod_{t_i} \varphi_{t_i}^{\overline{k_{t_i}}} \prod_{s_i} \varphi_{l_i}^{\overline{k_{l_i}}}\right) = \varphi_t(\varphi_s \varphi_l). \tag{19}$$

Therefore, the mapping set Φ satisfy Associativity under the compound operation. In the following part, we will find the unit element and inverse element of Φ, for the generator elements, $\forall\, 0 \leq i, j \leq m$, we have $\varphi_i^n \varphi_j = id \cdot \varphi_j = \varphi_j$, Therefore, for general elements we take $\forall\, \varphi \in \Phi, \varphi = \prod_i \varphi_i^{\overline{k_i}}$, we have:

$$\varphi_i^n \varphi = (id \cdot \varphi_j)\varphi_j^{\overline{k_j - 1}} \prod_{i\{i \neq j\}} \varphi_i^{\overline{k_i}} = \varphi_j^{\overline{k_j}} \prod_{i\{i \neq j\}} \varphi_i^{\overline{k_i}} = \prod_i \varphi_i^{\overline{k_i}} = \varphi, \tag{20}$$

and this implies that we have unit $e = \varphi_i^n = id$ in Φ. Identity is satisfied. For the generator element, $\forall\, 0 \leq i \leq m$, we have $\varphi_i^{\overline{k}}\varphi_i^{\overline{n-k}} = \varphi_i^{\overline{n}} = id = e, 0 \leq k \leq n$, for general element, $\forall\, \varphi \in \Phi, \varphi = \prod_i \varphi_i^{\overline{k_i}}$, we have:

$$\varphi \prod_i \varphi_i^{\overline{n-k_i}} = \prod_{i\{i\neq j\}} \varphi_i^{\overline{k_i}}(\varphi_j^{\overline{k_j}}\varphi_j^{\overline{n-k_j}}) \prod_{i\{i\neq j\}} \varphi_i^{\overline{n-k_i}} = \cdots = id, \tag{21}$$

which implies that for general element $\forall\, \varphi \in \Phi, \varphi = \prod_i \varphi_i^{\overline{k_i}}$, we have the inverse element $\varphi^{-1} = \prod_i \varphi_i^{\overline{n-k_i}}$ in Φ. Inverse is satisfied.

Summarizing, Φ is a group. Moreover, for generative element $\varphi_i, \varphi_j, \quad 0 \leq i, j \leq m$, we have $\varphi_i\varphi_j = \varphi_j\varphi_i$, for general element $\varphi_s, \varphi_t \in \Phi$,

$$\varphi_t\varphi_s = \prod_{s_i} \varphi_{s_i}^{\overline{k_{s_i}}} \prod_{t_i} \varphi_{t_i}^{\overline{k_{t_i}}} = \prod_{t_i} \varphi_{t_i}^{\overline{k_{t_i}}} \prod_{s_i} \varphi_{s_i}^{\overline{k_{s_i}}} = \varphi_s\varphi_t. \tag{22}$$

By pairwise exchange of generators, the $\varphi_{s_i}^{\overline{k_{s_i}}}$ is moved to the head of the equation one by one, then, equation 22 is obtained.

Summarizing, Φ is a Abelian group.

Q.E.D.

Step 2: In this step, we prove that the necessary and sufficient condition for the existence of isomorphism $\mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z} \times \ldots \mathbb{Z}/n\mathbb{Z} \sim < \varphi_1, \varphi_2, \ldots, \varphi_m >$ is that $0 \leq i, j \leq m$, $\varphi_i\varphi_j = \varphi_j\varphi_i$, and $0 \leq i \leq m$, $\varphi_i^n = e$ are satisfied simultaneously.

$\Rightarrow$)Assume the isomorphism is $h : G = \mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z} \times \ldots \mathbb{Z}/n\mathbb{Z} \to \Phi =< \varphi_1, \varphi_2, \ldots, \varphi_m >$, for the generator elements, $\forall\, \varphi_i, \varphi_j \in \{\varphi_i | 0 \leq i \leq m\}$, there are $g_s = \sum_{s_i} k_{s_i} e_{s_i}, g_t = \sum_{t_i} k_{t_i} e_{t_i} \in G, s.t.\ h(g_s) = \varphi_i, h(g_t) = \varphi_j$, since G is an Abelian group.

$$\varphi_i\varphi_j = h(g_s)h(g_t) = h(g_sg_t) = h(g_tg_s) = h(g_t)h(g_s) = \varphi_j\varphi_i. \tag{23}$$

For the generator elements in Φ, $0 \leq i \leq m, \varphi_i$, we have $g_l = \sum_{l_i} k_{l_i} e_{l_i} \in G, s.t. h(g_l) = \varphi_i$, compose itself n times,

$$\varphi_i^n = h(g_l)^n = h(ng_l) = h(\sum_{l_i} k_{l_i} n e_{l_i}) = h(\sum_{l_i} k_{l_i} e_G) = h(e_G) = e_\Phi. \tag{24}$$

In equation 24, $h(e_G) = e_\Phi$ is hold for h is a isomorphism, and $ker(h) = e_G$. Sufficiency is proven.

$\Leftarrow$) In the following part, we prove that when the two conditions are satisfied at the same time, the mapping τ we found is an isomorphism. Considering mapping $\tau : \Phi \to G$, The one-hot vector $(0, \ldots, \overline{1}, \ldots, 0) \in G$, that position i is 1 else is 0, is denoted by e_i. The definition of τ,

$$\begin{cases} \tau : \varphi_i \mapsto e_i; \\ \tau : \varphi_i\varphi_j \mapsto e_i + e_j. \end{cases} \tag{25}$$

For general elements in Φ, $\forall\, \phi_t, \phi_s \in \Phi$, where $\phi_t = \prod_{i_t} \varphi_{i_t}^{\overline{k_{i_t}}}, \phi_s = \prod_{i_s} \varphi_{i_s}^{\overline{k_{i_s}}}$, Note that ϕ_s is written in this form only if the set Φ is proved to be a group, which is hold by the two conditions.

$$\tau(\phi_t\phi_s) = \tau\left(\prod_{i_t} \varphi_{i_t}^{\overline{k_{i_t}}} \prod_{i_s} \varphi_{i_s}^{\overline{k_{i_s}}}\right) = \tau\left(\prod_{i_l} \varphi_{i_l}^{\overline{k_{i_l}}}\right) = \sum_{i_l} \overline{k_{i_l}} e_{i_l}, \tag{26}$$

and please note that the subscript can be merged since $\varphi_i\varphi_j = \varphi_j\varphi_i$, where i_l denotes the subscript after i_t, i_s merged. The summation on the right side of equation 26 is devided into two parts,

$$\sum_{i_l} \overline{k_{i_l}} e_{i_l} = \sum_{i_t} \overline{k_{i_t}} e_{i_t} + \sum_{i_s} \overline{k_{i_s}} e_{i_s} = \tau\left(\prod_{i_l} \varphi_{i_l}^{\overline{k_{i_l}}}\right) + \tau\left(\prod_{i_l} \varphi_{i_l}^{\overline{k_{i_l}}}\right) = \tau(\phi_t) + \tau(\phi_s). \tag{27}$$

Consequently, $\tau(\phi_t\phi_s) = \tau(\phi_t) + \tau(\phi_s)$, this implies that τ is a homomorphism. Then, only need to prove that the mapping τ is bijective. First we prove τ is injective, i.e. $ker(\tau) = \{e_G\}$, after that we prove τ is surjective, i.e. Find the inverse image of any element.

It's not hard to obtain $\tau(\phi) = \tau(e_\Phi \cdot \phi) = \tau(e_\Phi) + \tau(\phi)$, therefore $\tau(e_\Phi) = (0, \dots, 0) = e_G$, i.e. $e_\Phi \in \tau^{-1}(e_G)$, assume there is $\phi_e = \prod_{i_e} \varphi_{i_e}^{\overline{k_{i_e}}}, s.t. \tau(\phi_e) = e_G$

$$\begin{aligned} \phi_e &= \prod_{i_e} \varphi_{i_e}^{\overline{k_{i_e}}}, \tau(\phi_e) = \sum_{i_e} k_{i_e} e_{i_e} = e_G \Rightarrow k_{i_e} = l \cdot n; \\ \phi_e &= \prod_{i_e} \varphi_{i_e}^{\overline{k_{i_e}}} = \prod_{i_e} e_\Phi = e_\Phi, \end{aligned} \tag{28}$$

which means, $\tau^{-1}(e_G) = \{e_\Phi\}$, mapping τ is injective. In the following part, only need to prove that for any element in G, we can find the inverse image of it. We take $\forall\, g \in G, g = (\overline{k_1}, \overline{k_1}, \dots, \overline{k_m}) \in \mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z} \times \cdots \times \mathbb{Z}/n\mathbb{Z}$,

$$g = (\overline{k_1}, \dots, 0) + \cdots + (0, \dots, \overline{k_m}) = \sum_i \overline{k_i} e_i = \tau\left(\prod_i \varphi_i^{\overline{k_i}}\right). \tag{29}$$

Hence, $\tau^{-1}(g) = \prod_i \varphi_i^{\overline{k_i}} \in \Phi$, This implies that mapping τ is surjective. Since mapping τ is bijective and homomorphism, we conclude that τ is a isomorphism.

Q.E.D.

D. Proof for Theorem 3

Proof. Here we first introduce image space $X = \{x_1, x_2, \dots, x_N\}$ and representation space $Z = \{z_1, z_2, \dots, z_N\}$, the encoder $b : X \to Z$, and the decoder $d : Z \to X$, the definition $\varphi_i : X \to X, 0 \le i \le m$ is $\varphi_i(x) = d(a_i(\overline{b(x)}))$, where a_i is the group action of e_i on Z.

$\Rightarrow$) Obviously, we can obtain the expend form of $\varphi_j \circ \varphi_i$ from the above definition,

$$(\varphi_j \circ \varphi_i)(x) = \varphi_j(\varphi_i(x)) = \varphi_j(d(a_i(\overline{b(x)}))) = d(a_j(\overline{b(d(a_i(\overline{b(x)})))})). \tag{30}$$

For the condition: $\forall\, \varphi_i, \varphi_j \in \Phi, 0 \le i, j \le m$, we have $\varphi_i \varphi_j = \varphi_j \varphi_i$. The constrain on specific image can be obtained by

$$\varphi_i(\varphi_j(x)) = \varphi_j(\varphi_j(x)) \Rightarrow \varphi_i(\varphi_j(x)) - \varphi_j(\varphi_j(x)) = 0. \tag{31}$$

For the set of combinations of factors $C = \{(i, j) | \forall\, i, j \in I\}$, and the set containing images X, we have

$$\mathcal{L}_a = \sum_{x \in X} \sum_{(i,j) \in C} \|\varphi_i(\varphi_j(x)) - \varphi_j(\varphi_i(x))\| \text{ is optimized.} \tag{32}$$

The Abel loss is obtained. For the Order loss, we first consider n times compose of the same mapping φ_i,

$$\varphi_i^n = e_\Phi = id, \Rightarrow \varphi_i^{n-1} \varphi_i = \varphi_i \varphi_i^{n-1} = e_\Phi = id. \tag{33}$$

Therefore, $\varphi_i^{-1} = \varphi_i^{n-1}$. for a single image x, we have

$$\varphi_i(\varphi_i^{n-1}(x)) = \varphi_i(\varphi_i^{-1}(x)) = x \Rightarrow \varphi_i(\varphi_i^{-1}(x)) - x = 0. \tag{34}$$

Thus, for the set of factors I and the set containing all images X, we have

$$\sum_{x \in X} \sum_{i \in I} \|\varphi_i(\varphi_i^{-1}(x)) - x\| \text{ is optimized.} \tag{35}$$

For eliminating the bias of optimization, we optimize a symmetry form of the Order Loss, and we have

$$\mathcal{L}_o = \sum_{x \in X} \sum_{i \in I} (\|\varphi_i(\varphi_i^{-1}(x)) - x\| + \|\varphi_i^{-1}(\varphi_i(x)) - x\| \text{ is optimized.} \tag{36}$$

The Order Loss is obtained.

$\Leftarrow$)When the Abel Loss $\mathcal{L}_a$ is optimized, $\forall\, x \in X$, we have

$$\varphi_i(\varphi_j(x)) - \varphi_j(\varphi_i(x)) = 0 \Rightarrow \varphi_i(\varphi(x)) = \varphi_j(\varphi_i(x)). \tag{37}$$

Therefore, $\forall\, \varphi_i, \varphi_j \in \Phi, 0 \leq i, j \leq m$, we have $\varphi_i \varphi_j = \varphi_j \varphi_i$, we obtain the Abel constraint. When the Order Loss $\mathcal{L}_o$ is optimized, $\forall\, x \in X$, we have

$$\varphi_i(\varphi_i^{-1}(x)) - x = 0 \Rightarrow \varphi_i(\varphi_i^{-1}(x)) = x. \tag{38}$$

This implies that

$$\varphi_i^n = \varphi_i \circ \varphi_i^{n-1} = \varphi_i \circ \varphi_i^{-1} = e. \tag{39}$$

Therefore, we obtain the Order constraint. The Group constraints are satisfied.

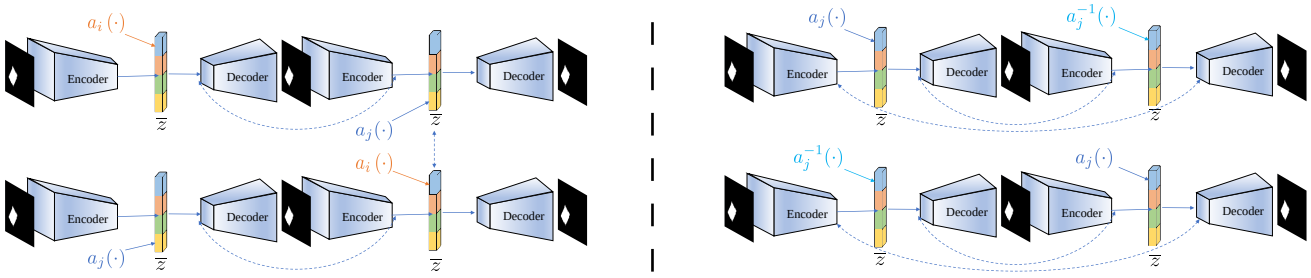
E. Details of Implementation

E.1. Abel Loss Details

As mentioned in the main paper, the Abel Loss of the groupifiable VAE-based model is as follows:

$$\mathcal{L}_a = \sum_{x \in X} \sum_{(i,j) \in C} \|\varphi_i(\varphi_j(x)) - \varphi_j(\varphi_i(x))\|, \tag{40}$$

where $\varphi_i(\varphi_j(x))$ represents the upper path of Figure 3 (a), and $\varphi_j(\varphi_i(x))$ represents the lower path of Figure 3 (a). For better optimization, we do not constrain such consistency on the reconstructed images but on their representations instead (straight dotted line in Figure 3 (a)). Besides, we constrain the consistency between the representations of intermediate images (curved dotted lines in Figure 3 (a)). a_i denotes the group action of e_i on Z, i.e., $a_i(\overline{z}) = \overline{z + e_i}$. We implement a_i by adding the action scale τ on the i-th dimension of z, then mapping it to $\overline{z}$ by the sine and cosine functions, so as to a_j.



(a) Abel Loss (b) Order Loss

Figure 3. Overview of the Isomorphism Loss. The Abel Loss constrains the exchangeability of Permutation Group Φ. The Order Loss constrains the cyclicity of Permutation Group Φ. Dotted line in the figure represent reconstruction loss. a_i, a_j, a_i^{-1} denotes group actions.

E.2. Order Loss Details

As mentioned in the main paper, the Order Loss of the groupified VAEs is as follows:

$$\mathcal{L}_o = \sum_{x \in X} \sum_{i \in I} \left(\|\varphi_i(\varphi_i^{-1}(x)) - x\| + \|\varphi_i^{-1}(\varphi_i(x)) - x\| \right), \tag{41}$$

where $\varphi_i(\varphi_i^{-1}(x))$ represents the lower path of Figure 3 (b), and $\varphi_i^{-1}(\varphi_i(x))$ represents the upper path of Figure 3 (b). Similar to the Abel Loss, we do not constrain such consistency on the reconstructed images for better optimization, but on their representations instead (long curved dotted line in Figure 3 (b)). Besides, we constrain the consistency between the representations of intermediate images (short curved dotted lines in Figure 3 (b)). a_j^{-1} denotes the group action of e_j^{-1} on Z, i.e., $a_j^{-1}(\overline{z}) = \overline{z + (n-1)e_j}$. We implement a_j^{-1} by adding the action scale $(N-1)\tau$ on the j-th dimension of z, then mapping it to $\overline{z}$ by the sine and cosine functions (we set $n = N$).

E.3. The set of factors I

For the given VAE-based models, there are two ways to get set I. (i) Assign some dimensions to I, e.g., set $I = \{0, 1, 2, 3, 4\}$, which is the first-5 of the dimensions, resulting in meaningful dimensions controllable. (ii) Dimensional KL divergence indicates the meaningful dimensions. We set $I = \{i | KL_i \geq T\}$, where T is a hyperparameter, which is set to 30 in our experiments.

F. Details of Experiments

F.1. Dataset details

In all the experiments, we resize the images to 64x64 resolution. We then introduce all the datasets used in our paper in details.

dSprites (Higgins et al., 2016): dSprites contains 737,280 binary 2D shapes (heart, oval and square) images with 4 ground truth factors: shape (3 values), scale (6 values), orientation (40 values), x-position (32 values), y-position (32 values). Then we introduce the variants of dSprites created by Locatello et al. (2019).

Color-dSprites, the shapes are colored with a random color.

Noisy-dSprites, we consider white-colored shapes on a noisy background.

Shapes3D (Kim & Mnih, 2018): Shapes3D contains 480,000 3D shapes images with 6 ground truth factors: shape (4 values), scale (8 values), orientation (15 values), floor color (10 values), wall color (15 values), object color (10 values).

Cars3D (Reed et al., 2015): This dataset consists of 183 car CAD models, each rendered from 24 azimuth directions and 4 elevations.

Table 1. Architecture of the encoder and decoder of VAEs. For original VAE, the dimension of input of the decoder is 10. For Groupified VAE, the dimension is 20.

Encoder	Decoder
Input: 64×64 $\times$ number of channels	Input: 10 or 20
4×4 conv, 32 ReLU, stride 2	FC, 256 ReLU
4×4 conv, 32 ReLU, stride 2	FC, 256 ReLU
4×4 conv, 32 ReLU, stride 2	FC, $4 \times 4 \times 32$ ReLU
4×4 conv, 32 ReLU, stride 2	4×4 deconv, 32 ReLU, stride 2
FC 256 ReLU	4×4 deconv, 32 ReLU, stride 2
FC 256 ReLU	4×4 deconv, 32 ReLU, stride 2
FC $2\times$ 10	4×4 deconv, number of channels, stride 2

F.2. Architecture for encoder and decoder

We follow Locatello et al. (2019) to use the same architecture of VAEs in all of the experiments as follows: the activation function used is ReLU except for the last layer of decoder, which is Sigmoid. As shown in Table 1. For the details of the Discriminator in FactorVAE, please refers to Table 3 (a) and (c).

Table 2. Hyperparameters and random seeds for every model.

Model	Parameter	Value
β-VAE	β	[10; 20; 30;]
AnnealedVAE	C	[10; 20; 30;]
	start	[3e4; 4e4; 5e4;]
	end	[2e4; 3e4; 4e4;]
FactorVAE	γ	[5; 10; 15]
β-TCVAE	β	[6; 9; 12]
	random seed	[1; 2; 3; 4; 5; 6; 7; 8; 9;]
	group	[True; False]

F.3. Experiment settings

We run different hyperparameters and random seeds for every VAE-based models implemented by Pytorch (Paszke et al., 2017). As shown in Table 2, for β-VAE, we assign 3 choices for β and 10 random seeds for both the original and groupified

Table 3. Shared hyperparameters in all experiments.

(a) Optimizer for Discriminator

Parameter	Values
Batch size	64
Optimizer	Adam
Adam: beta1	0.9
Adam: beta2	0.999
Adam: epsilon	1.00e-08
Adam: learning rate	0.0001

(b) General hyperparameters for VAE

Parameter	Values
Batch size	64
Latent space dimension	10
Optimizer	Adam
Adam: beta1	0.9
Adam: beta2	0.999
Adam: epsilon	1.00e-08
Adam: learning rate	0.0001
Decoder type	Bernoulli

(c) Architecture of Discriminator

Discriminator
FC, 1000 leaky ReLU
FC, 1000 leaky ReLU
FC, 1000 leaky ReLU
FC, 1000 leaky ReLU
FC, 1000 leaky ReLU
FC, 1000 leaky ReLU
FC, 2

VAEs: $3 \times 10 \times 2 = 60$ settings for each dataset. Similarly, we also assign 60 settings for FactorVAE and β-TCVAE. For AnnealVAE, we assign 3 choices for C_{max} and 3 choices for the start and end pair, also assign 10 random seeds. In summary, for all 5 datasets, we run $(((3 \times 10 \times 2) \times 3) + 3 \times 3 \times 10 \times 2) \times 5 = 1800$ models. For other hyperparameters, please refer to Table 3 (b).

G. More Qualitative Results and Score Distribution

G.1. Qualitative Results

The performance of original and groupified VAEs on all 5 datasets is shown in Table 4. Our method outperforms the original one on most of the datasets in terms of nearly all the metrics.

Table 4. Performance (mean ± variance) on different datasets and different models with different metrics. We evaluate β-VAE, AnnealVAE, FactorVAE, and β-TCVAE on dSprites, Cars3d, Shapes3d, Noisy-dSprites, and Color-dSprites for 1800 settings. These settings include different random seeds and hyperparameters.

dSprits	DCI		BetaVAE		MIG		FactorVAE	
	Original	Groupified	Original	Groupified	Original	Groupified	Original	Groupified
β-VAE	0.23 ± 0.10	$\mathbf{0.46 \pm 0.085}$	0.75 ± 0.083	$\mathbf{0.86 \pm 0.051}$	0.14 ± 0.097	$\mathbf{0.37 \pm 0.089}$	0.51 ± 0.098	$\mathbf{0.63 \pm 0.089}$
AnnealVAE	0.28 ± 0.10	$\mathbf{0.39 \pm 0.056}$	0.84 ± 0.050	$\mathbf{0.87 \pm 0.0067}$	0.23 ± 0.10	$\mathbf{0.34 \pm 0.061}$	$\mathbf{0.70} \pm 0.094$	$0.68 \pm \mathbf{0.058}$
FactorVAE	0.38 ± 0.10	$\mathbf{0.41 \pm 0.074}$	$\mathbf{0.89} \pm 0.040$	$\mathbf{0.89 \pm 0.020}$	0.27 ± 0.092	$\mathbf{0.31 \pm 0.061}$	0.74 ± 0.068	$\mathbf{0.75} \pm 0.075$
β-TCVAE	0.35 ± 0.065	$\mathbf{0.36} \pm 0.11$	0.86 ± 0.026	$\mathbf{0.861} \pm 0.038$	0.031 ± 0.17	$\mathbf{0.060 \pm 0.24}$	0.68 ± 0.098	$\mathbf{0.70} \pm 0.098$

Cars3d	DCI		BetaVAE		MIG		FactorVAE	
	Original	Groupified	Original	Groupified	Original	Groupified	Original	Groupified
β-VAE	0.18 ± 0.059	$\mathbf{0.24 \pm 0.041}$	$0.99 \pm 1.6e-3$	$\mathbf{1.0 \pm 0.0}$	0.071 ± 0.032	$\mathbf{0.11} \pm 0.032$	0.81 ± 0.066	$\mathbf{0.93 \pm 0.034}$
AnnealVAE	0.22 ± 0.046	$\mathbf{0.25 \pm 0.046}$	$\mathbf{0.99} \pm 4e-4$	$\mathbf{0.99 \pm 1.5e-4}$	0.074 ± 0.016	$\mathbf{0.10 \pm 0.014}$	0.82 ± 0.062	$\mathbf{0.87 \pm 0.028}$
FactorVAE	0.21 ± 0.054	$\mathbf{0.25 \pm 0.040}$	$0.99 \pm 1e-4$	$\mathbf{1.0 \pm 0.0}$	0.098 ± 0.027	$\mathbf{0.11} \pm 0.033$	0.90 ± 0.039	$\mathbf{0.93 \pm 0.034}$
β-TCVAE	0.24 ± 0.049	$\mathbf{0.26 \pm 0.046}$	$\mathbf{1.0 \pm 0.0}$	$\mathbf{1.0 \pm 0.0}$	0.10 ± 0.021	$\mathbf{0.11} \pm 0.033$	0.88 ± 0.040	$\mathbf{0.93 \pm 0.034}$

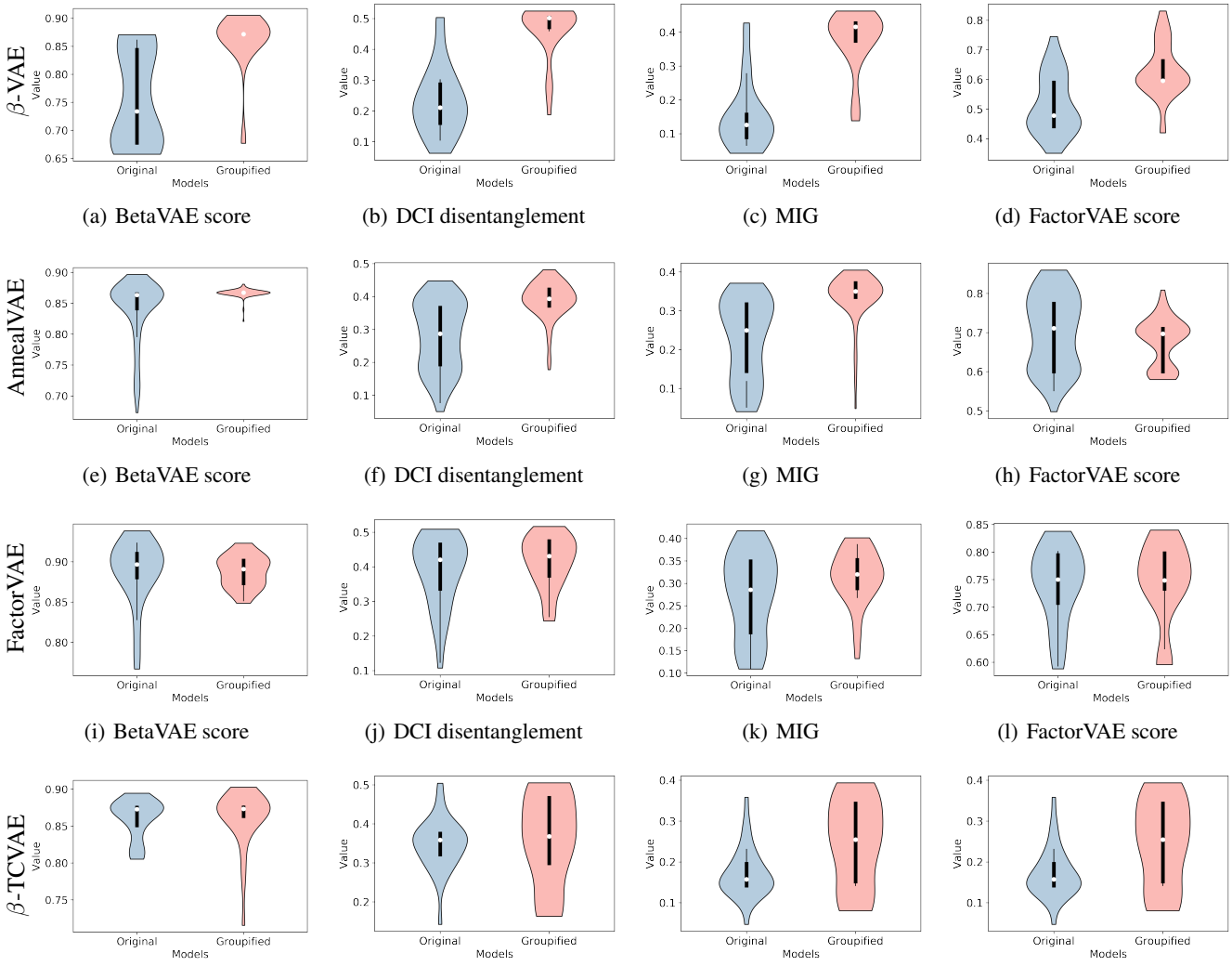
Noisy dSprits	DCI		betaVAE		MIG		FactorVAE	
	Original	Groupified	Original	Groupified	Original	Groupified	Original	Groupified
BetaVAE	0.056 ± 0.018	$\mathbf{0.087} \pm 0.051$	0.624 ± 0.090	$\mathbf{0.647 \pm 0.055}$	0.030 ± 0.022	$\mathbf{0.065} \pm 0.055$	0.355 ± 0.093	$\mathbf{0.407 \pm 0.071}$
Anneal VAE	0.053 ± 0.013	$\mathbf{0.060} \pm 0.022$	0.631 ± 0.036	$\mathbf{0.644 \pm 0.031}$	0.035 ± 0.027	$\mathbf{0.047} \pm 0.032$	0.434 ± 0.080	$\mathbf{0.481} \pm 0.087$
FactorVAE	$\mathbf{0.114} \pm 0.062$	$0.099 \pm \mathbf{0.057}$	0.682 ± 0.081	$\mathbf{0.684 \pm 0.070}$	$\mathbf{0.077} \pm 0.046$	$0.066 \pm \mathbf{0.046}$	0.437 ± 0.098	$\mathbf{0.468 \pm 0.098}$
BetaTCVAE	0.081 ± 0.036	$\mathbf{0.111} \pm 0.053$	0.605 ± 0.053	$\mathbf{0.635 \pm 0.050}$	0.040 ± 0.030	$\mathbf{0.068} \pm 0.042$	0.353 ± 0.091	$\mathbf{0.431} \pm 0.097$

Shapes3d	DCI		BetaVAE		MIG		FactorVAE	
	Original	Groupified	Original	Groupified	Original	Groupified	Original	Groupified
β-VAE	0.44 ± 0.176	$\mathbf{0.56 \pm 0.10}$	0.91 ± 0.072	$0.90 \pm \mathbf{0.045}$	0.28 ± 0.18	$\mathbf{0.42 \pm 0.15}$	$\mathbf{0.82} \pm 0.098$	$\mathbf{0.82 \pm 0.043}$
AnnealVAE	0.52 ± 0.051	$\mathbf{0.60} \pm 0.078$	$0.82 \pm \mathbf{0.076}$	$\mathbf{0.89} \pm 0.086$	0.48 ± 0.047	$\mathbf{0.50} \pm 0.052$	0.75 ± 0.074	$\mathbf{0.83 \pm 0.066}$
FactorVAE	0.47 ± 0.10	$\mathbf{0.49 \pm 0.065}$	$\mathbf{0.86 \pm 0.055}$	0.80 ± 0.075	0.33 ± 0.13	$\mathbf{0.43 \pm 0.11}$	$\mathbf{0.81 \pm 0.056}$	0.79 ± 0.066
β-TCVAE	0.66 ± 0.10	$\mathbf{0.72 \pm 0.061}$	$\mathbf{0.97 \pm 0.039}$	0.96 ± 0.042	0.40 ± 0.18	$\mathbf{0.47 \pm 0.090}$	0.89 ± 0.064	$\mathbf{0.90 \pm 0.046}$

Color dSprits	DCI		betaVAE		MIG		FactorVAE	
	Original	Groupified	Original	Groupified	Original	Groupified	Original	Groupified
BetaVAE	0.174 ± 0.097	$\mathbf{0.328} \pm 0.130$	0.798 ± 0.094	$\mathbf{0.844 \pm 0.050}$	0.103 ± 0.058	$\mathbf{0.243} \pm 0.118$	0.591 ± 0.148	$\mathbf{0.648 \pm 0.092}$
Anneal VAE	0.268 ± 0.103	$\mathbf{0.337} \pm 0.114$	0.843 ± 0.038	$\mathbf{0.856 \pm 0.031}$	0.219 ± 0.084	$\mathbf{0.252} \pm 0.104$	$\mathbf{0.718 \pm 0.065}$	0.692 ± 0.094
FactorVAE	0.294 ± 0.101	$\mathbf{0.322} \pm 0.104$	0.861 ± 0.038	$\mathbf{0.862 \pm 0.029}$	0.203 ± 0.080	$\mathbf{0.236} \pm 0.091$	$\mathbf{0.739 \pm 0.068}$	0.730 ± 0.080
BetaTCVAE	0.338 ± 0.052	$\mathbf{0.395} \pm 0.082$	0.876 ± 0.024	$\mathbf{0.881} \pm 0.031$	0.169 ± 0.040	$\mathbf{0.269} \pm 0.090$	0.711 ± 0.086	$\mathbf{0.786 \pm 0.050}$

G.2. Score Distribution

The detailed distribution of the performance is shown in this section (demonstrated by the Violin Plot (Hintze & Nelson, 1998)). The performance distributions on dSprits, Car3d, Noisy dSprites, Color dSprites and Shapes3d are shown in Figure 4, Figure 5, Figure 6, Figure 7 and Figure 8 respectively.



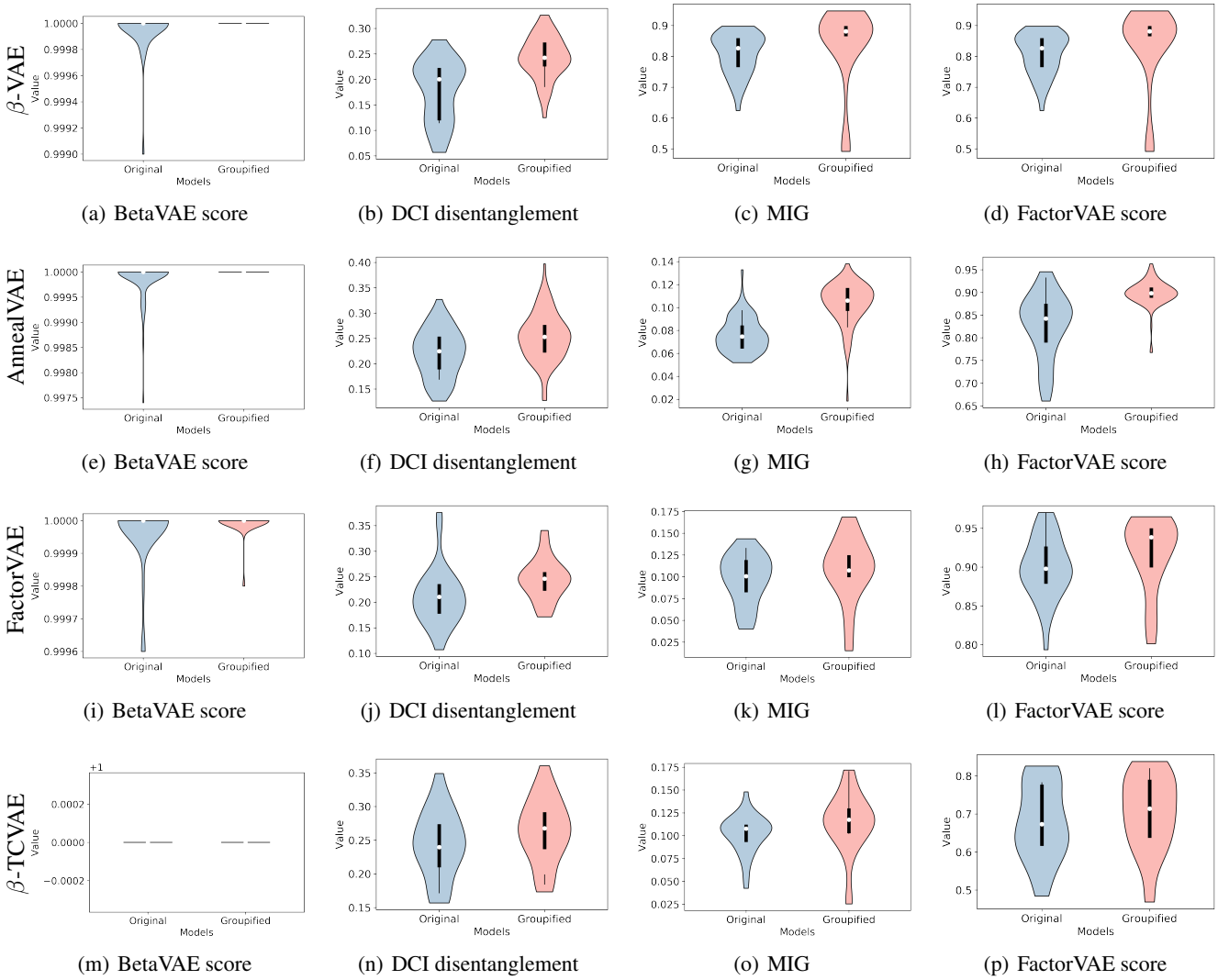
(a) BetaVAE score (b) DCI disentanglement (c) MIG (d) FactorVAE score

(e) BetaVAE score (f) DCI disentanglement (g) MIG (h) FactorVAE score

(i) BetaVAE score (j) DCI disentanglement (k) MIG (l) FactorVAE score

(m) BetaVAE score (n) DCI disentanglement (o) MIG (p) FactorVAE score

Figure 4. Performance distribution on dSprites. Variance is due to different hyperparameters and random seeds. We take four metrics into consideration: BetaVAE score, DCI disentanglement, MIG and FactorVAE score. We observe that groupified models outperform the original ones.



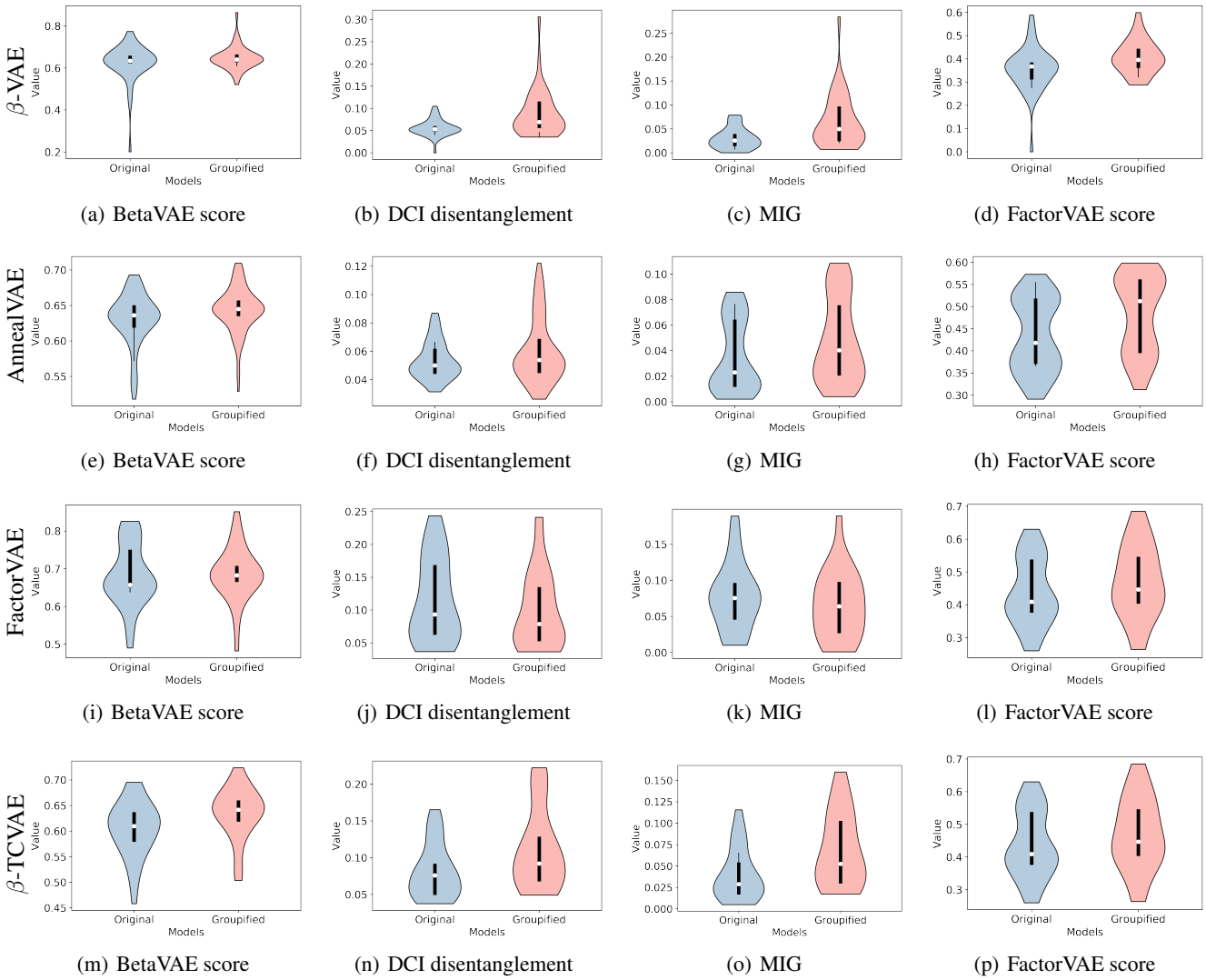
(a) BetaVAE score (b) DCI disentanglement (c) MIG (d) FactorVAE score

(e) BetaVAE score (f) DCI disentanglement (g) MIG (h) FactorVAE score

(i) BetaVAE score (j) DCI disentanglement (k) MIG (l) FactorVAE score

(m) BetaVAE score (n) DCI disentanglement (o) MIG (p) FactorVAE score

Figure 5. Performance distribution on Cars3d. Variance is due to different hyperparameters and random seeds. We take four metrics into consideration: BetaVAE score, DCI disentanglement, MIG and FactorVAE score. We observe that groupified models outperform the original ones.



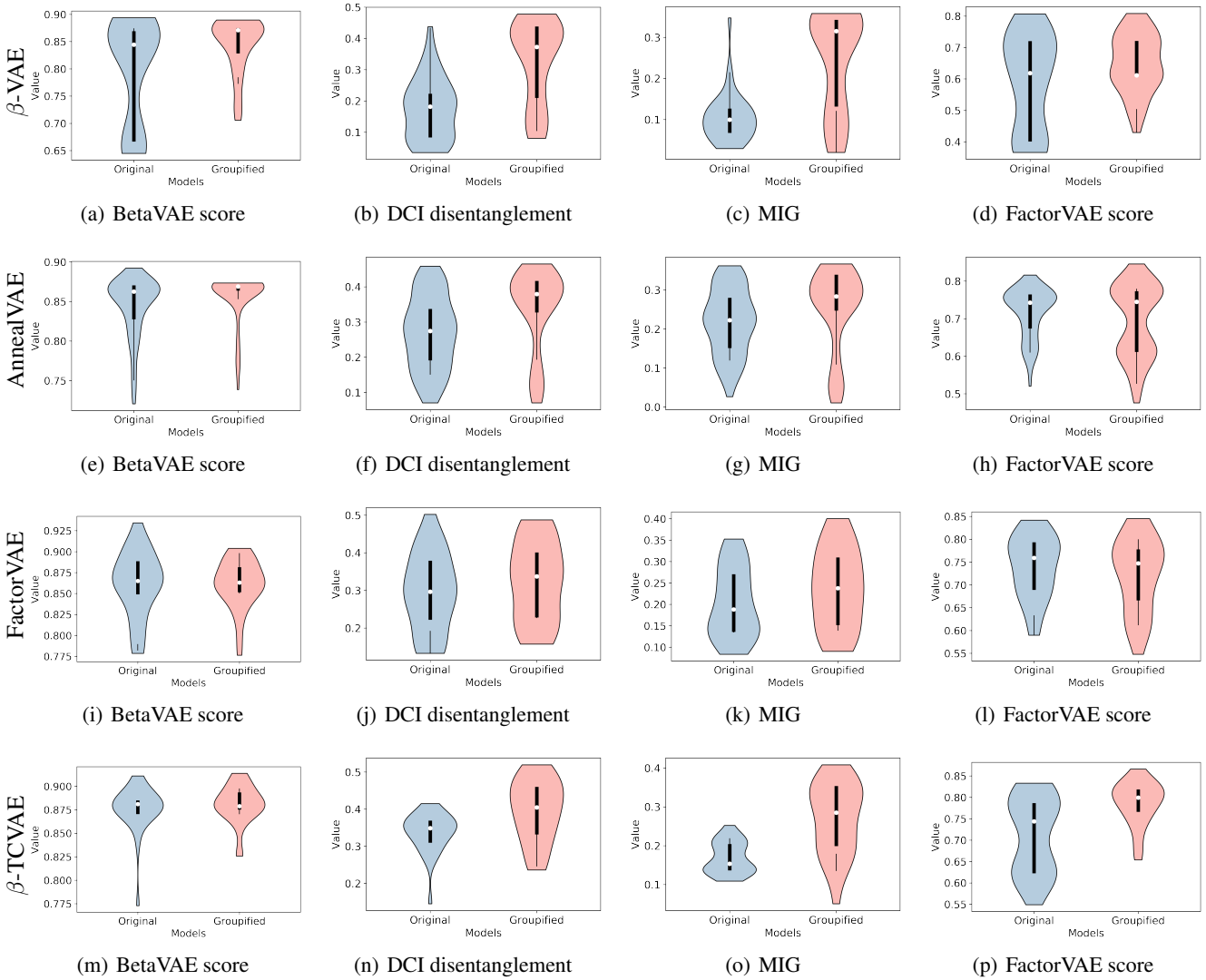
(a) BetaVAE score (b) DCI disentanglement (c) MIG (d) FactorVAE score

(e) BetaVAE score (f) DCI disentanglement (g) MIG (h) FactorVAE score

(i) BetaVAE score (j) DCI disentanglement (k) MIG (l) FactorVAE score

(m) BetaVAE score (n) DCI disentanglement (o) MIG (p) FactorVAE score

Figure 6. Performance distribution on Noisy dSprites. Variance is due to different hyperparameters and random seeds. We take four metrics into consideration: BetaVAE score, DCI disentanglement, MIG and FactorVAE score. We observe that groupified models outperform the original ones.



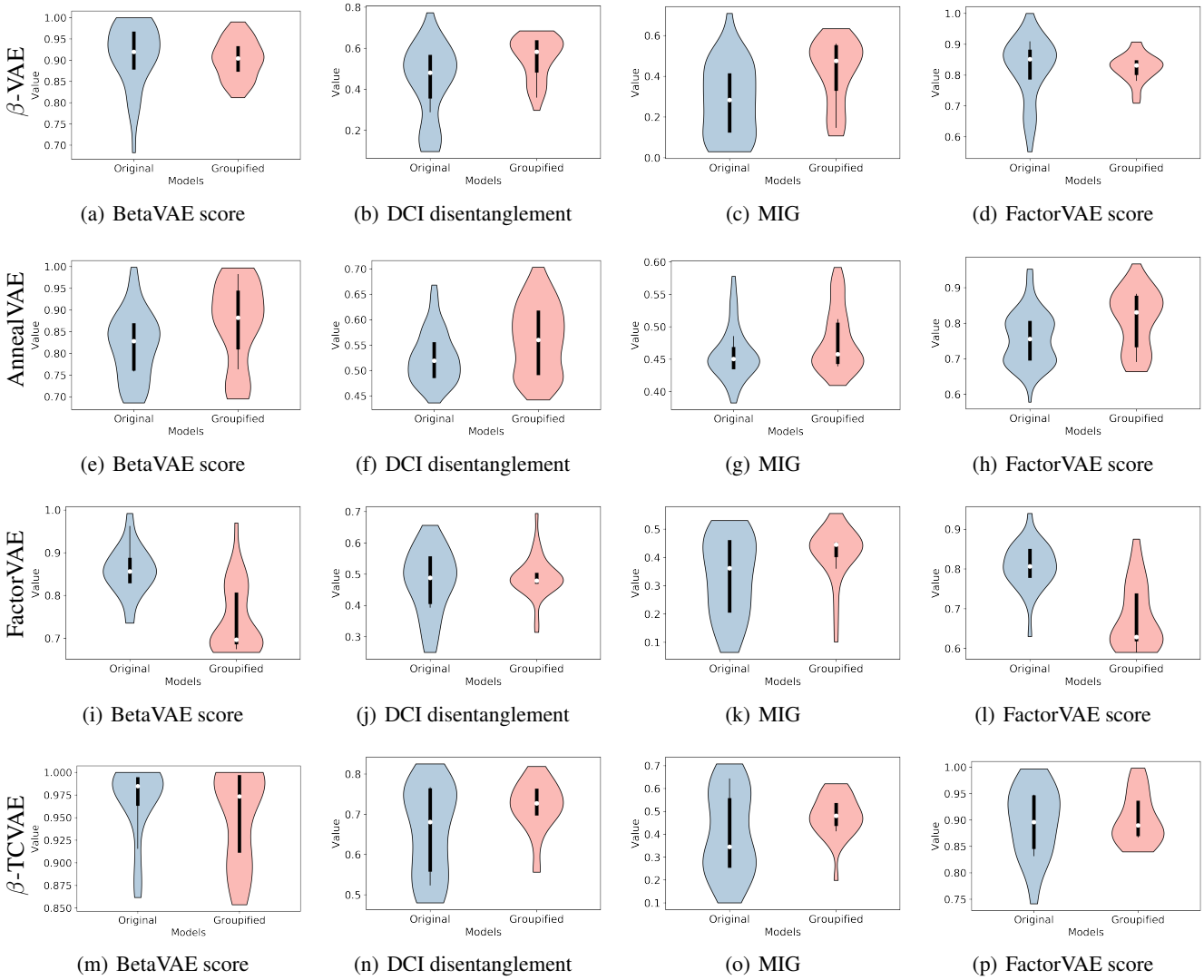
(a) BetaVAE score (b) DCI disentanglement (c) MIG (d) FactorVAE score

(e) BetaVAE score (f) DCI disentanglement (g) MIG (h) FactorVAE score

(i) BetaVAE score (j) DCI disentanglement (k) MIG (l) FactorVAE score

(m) BetaVAE score (n) DCI disentanglement (o) MIG (p) FactorVAE score

Figure 7. Performance distribution on Color dSprites. Variance is due to different hyperparameters and random seeds. We take four metrics into consideration: BetaVAE score, DCI disentanglement, MIG and FactorVAE score. We observe that groupified models outperform the original ones.



(a) BetaVAE score (b) DCI disentanglement (c) MIG (d) FactorVAE score

(e) BetaVAE score (f) DCI disentanglement (g) MIG (h) FactorVAE score

(i) BetaVAE score (j) DCI disentanglement (k) MIG (l) FactorVAE score

(m) BetaVAE score (n) DCI disentanglement (o) MIG (p) FactorVAE score

Figure 8. Performance distribution on Shapes3d. Variance is due to different hyperparameters and random seeds. We take four metrics into consideration: BetaVAE score, DCI disentanglement, MIG and FactorVAE score. We observe that groupified models outperform the original ones.

H. More Qualitative Results

We evaluate our methods qualitatively on two typical datasets: Cars3d and Shapes3d. We visualize the traversal results of orginal and groupified FactorVAE and β-TCVAE. For every factor, we traversal from 5 random representation. As shown in Figure 9 and Figure 10, the traversal results of groupified FactorVAE and β-TCVAE shows that these models learn less entangled representation on Shapes3D (e.g., Orientation of FactorVAE and Scale and Shape of β-TCVAE).

Similarly, as shown in Figure 11 and Figure 12, groupified FactorVAE and β-TCVAE archieve better disentanglement ability on Car3d (e.g., Rotation of FactorVAE and Yaw of β-TCVAE).

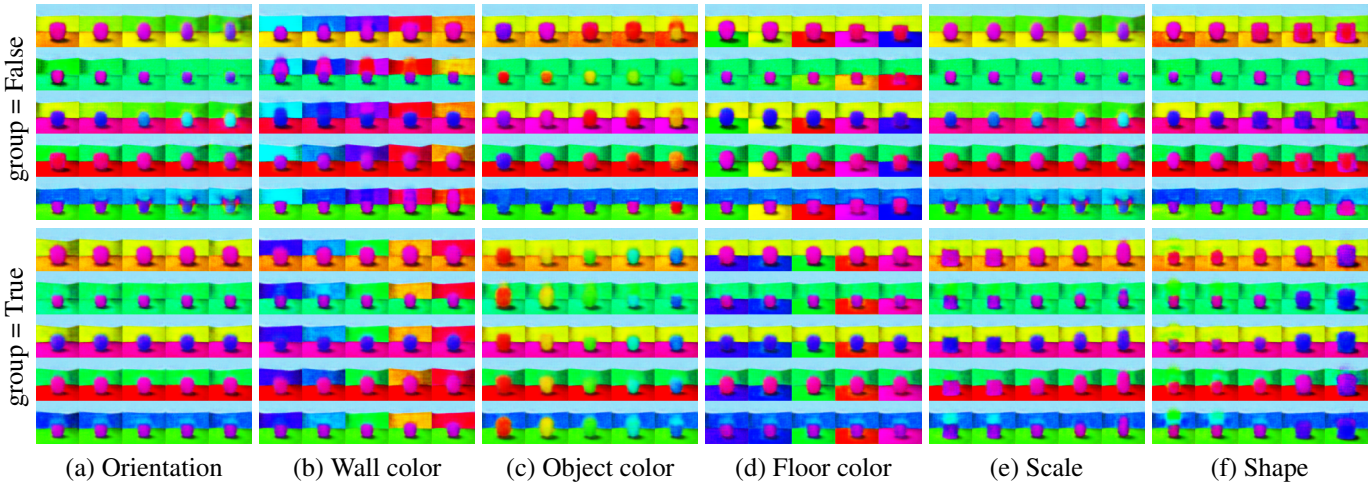


Figure 9. Learned latent variables using original and groupified FactorVAE on Shapes3d dataset. Traversal range is (-2, 2).

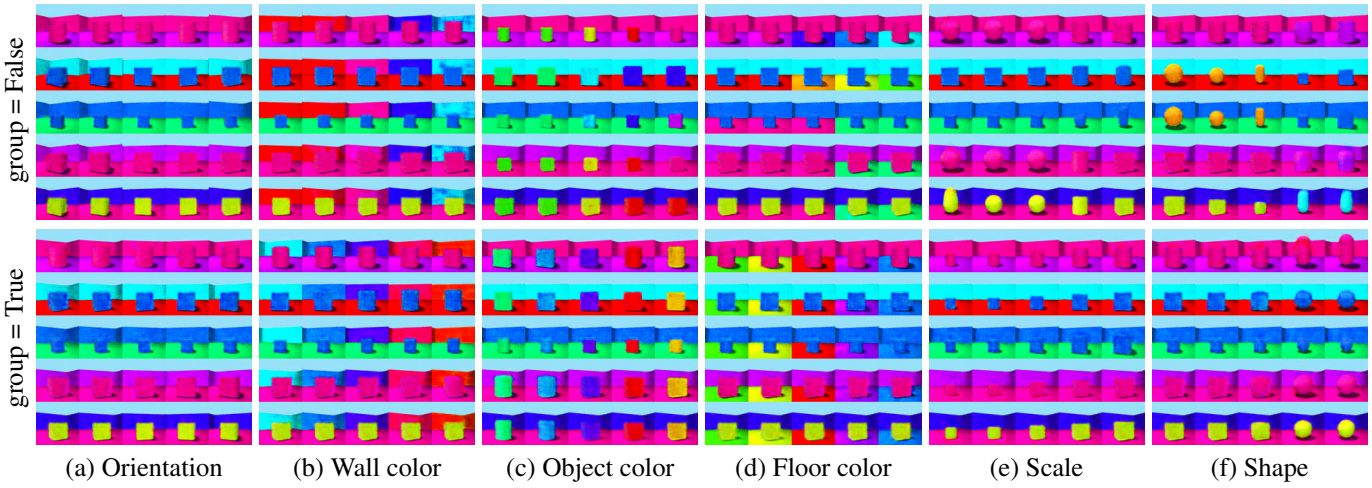


Figure 10. Learned latent variables using original and groupified β-TCVAE on Shapes3d dataset. Traversal range is (-2, 2).

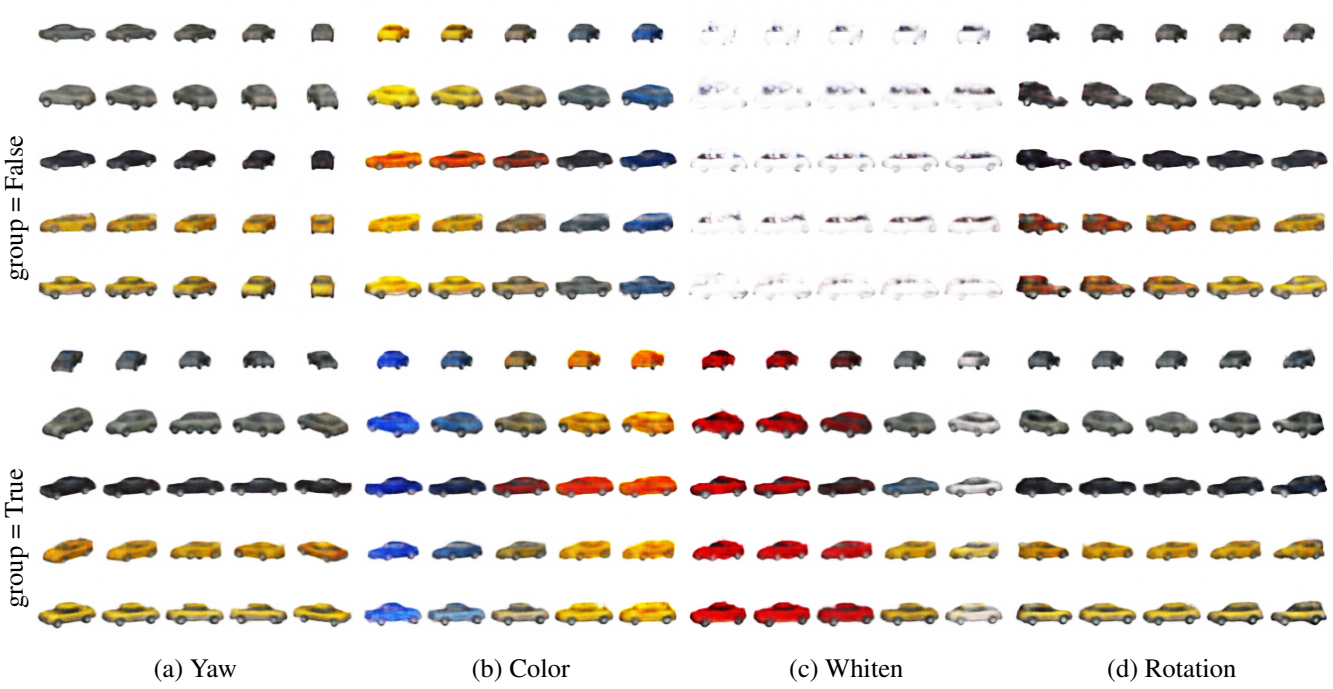


Figure 11. Learned latent variables using original and groupified FactorVAE on Car3d dataset. Traversal range is (-2, 2).

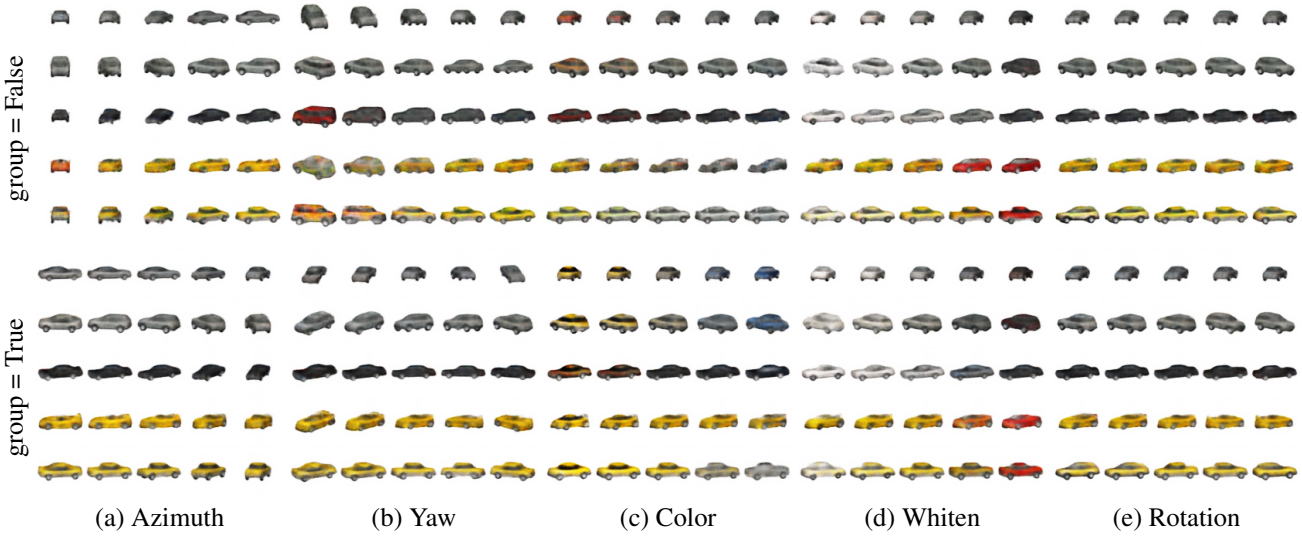
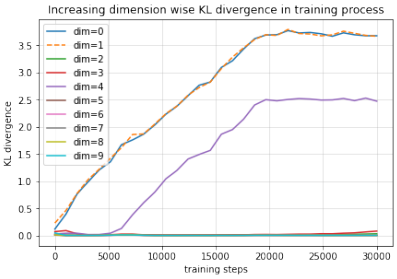


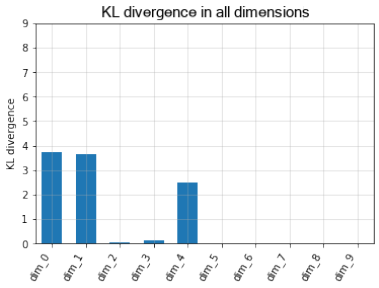
Figure 12. Learned latent variables using original and groupified β-TCVAE on Car3d dataset. Traversal range is (-2, 2).

I. More Meaningful Dimension Visualizations

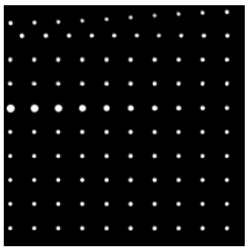
To illustrate that controllable dimensions in AnnealVAE is not an exception, we provide more visualizations of groupified AnnealVAEs. Groupified AnnealVAEs of different hyperparameters and random seeds are shown in Figure 13 to 21, suggesting that the dimensions of them are controllable.



(a) Dimension wised KL in training

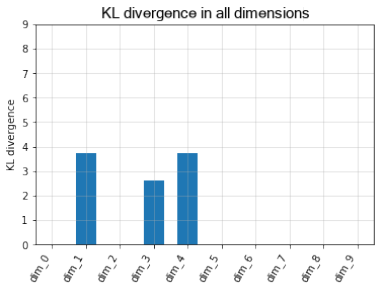


(b) Dimension wise KL in 30000 steps

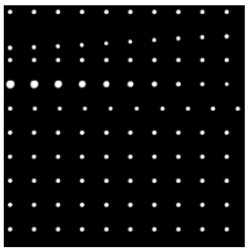


(c) Images traversal of dimensions

(d) Dimension wised KL in training

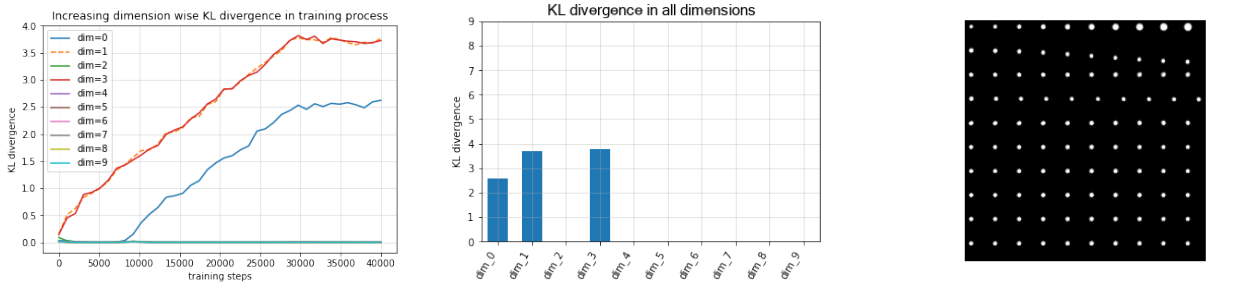


(e) Dimension wise KL in 30000 steps

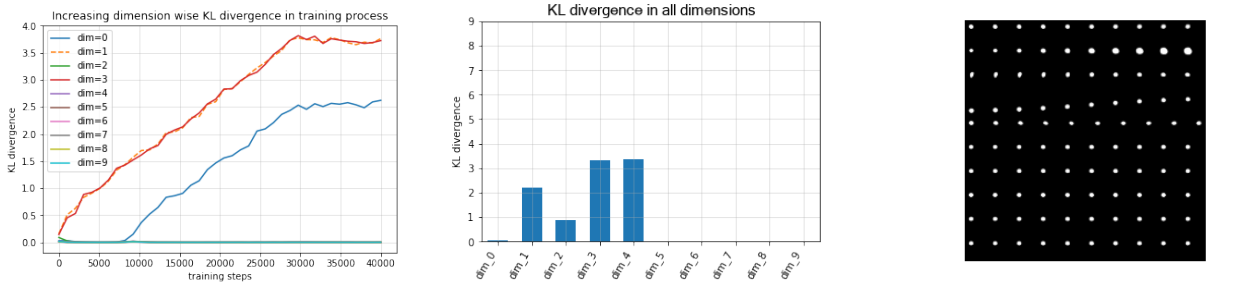


(f) Images traversal of dimensions

Figure 13. Meaningful dimensions visualization for $C_{max} = 10$, end = 30000 (different random seeds). The KL divergences of target dimension (0-4 dimension) increase one by one during training (a). The KL divergences in different dimensions are different amounts after training (b). As the the image traversal results (c) shows, the meaningful dimensions are learned in 0-4 dims. So as to (d), (e) and (f).

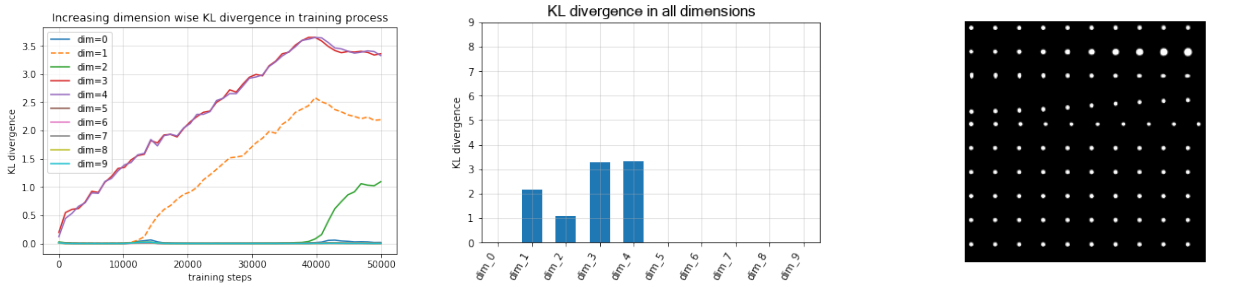


(a) Dimension wised KL in training (b) Dimension wise KL in 40000 steps (c) Images traversal of dimensions

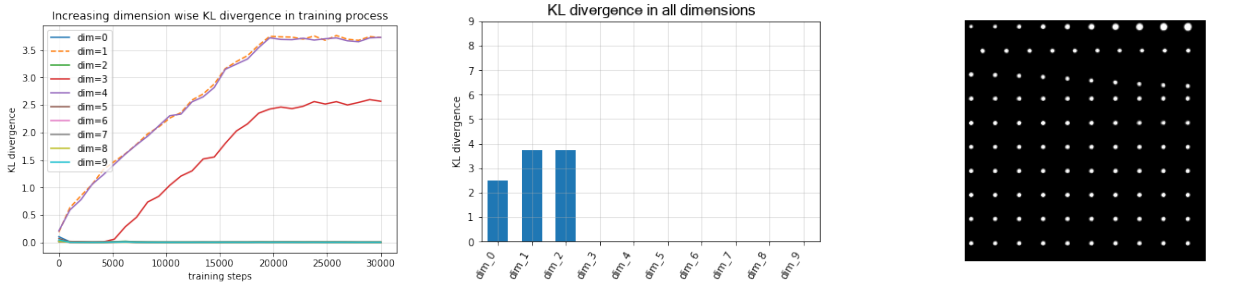


(d) Dimension wised KL in training (e) Dimension wise KL in 40000 steps (f) Images traversal of dimensions

Figure 14. Meaningful dimensions visualization for $C_{max} = 10$, end = 40000 (different random seeds). The KL divergences of target dimension (0-4 dimension) increase one by one during training (a). The KL divergences in different dimensions are different amounts after training (b). As the the image traversal results (c) shows, the meaningful dimensions are learned in 0-4 dims. So as to (d), (e) and (f).

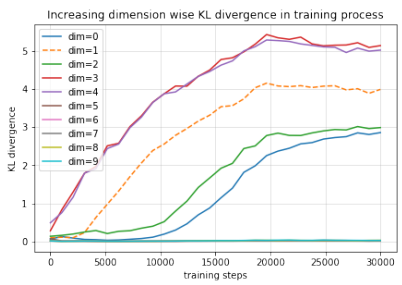


(a) Dimension wised KL in training (b) Dimension wise KL in 50000 steps (c) Images traversal of dimensions

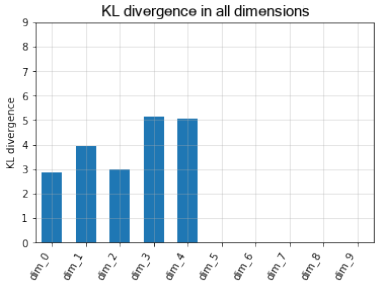


(d) Dimension wised KL in training (e) Dimension wise KL in 50000 steps (f) Images traversal of dimensions

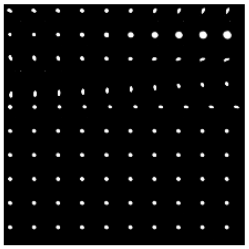
Figure 15. Meaningful dimensions visualization for $C_{max} = 10$, end = 50000 (different random seeds). The KL divergences of target dimension (0-4 dimension) increase one by one during training (a). The KL divergences in different dimensions are different amounts after training (b). As the the image traversal results (c) shows, the meaningful dimensions are learned in 0-4 dims. So as to (d), (e) and (f).



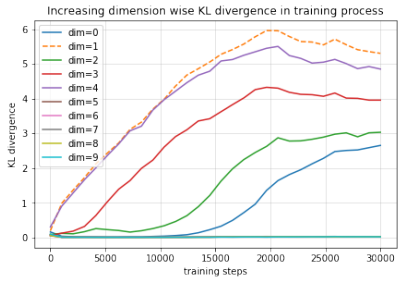
(a) Dimension wised KL in training



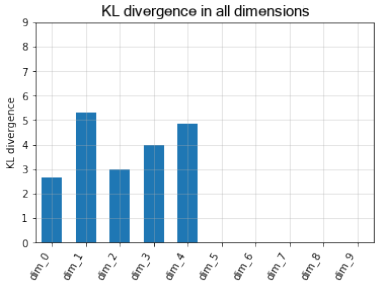
(b) Dimension wise KL in 50000 steps



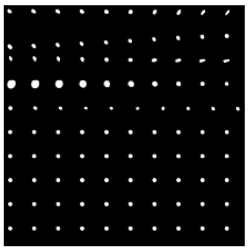
(c) Images traversal of dimensions



(d) Dimension wised KL in training



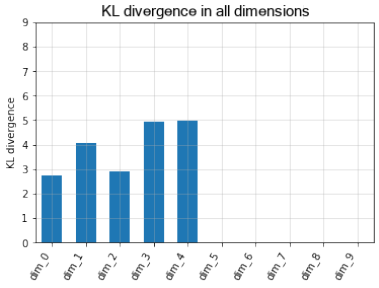
(e) Dimension wise KL in 50000 steps



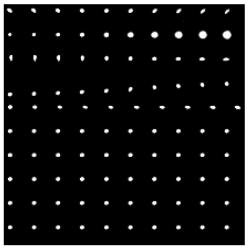
(f) Images traversal of dimensions

Figure 16. Meaningful dimensions visualization for $C_{max} = 20$, end = 30000 (different random seeds). The KL divergences of target dimension (0-4 dimension) increase one by one during training (a). The KL divergences in different dimensions are different amounts after training (b). As the the image traversal results (c) shows, the meaningful dimensions are learned in 0-4 dims. So as to (d), (e) and (f).

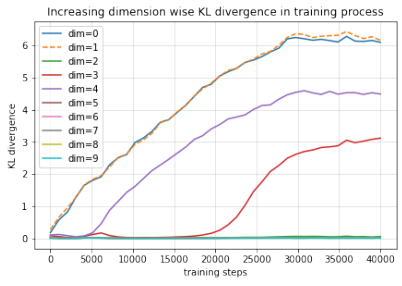
(a) Dimension wised KL in training



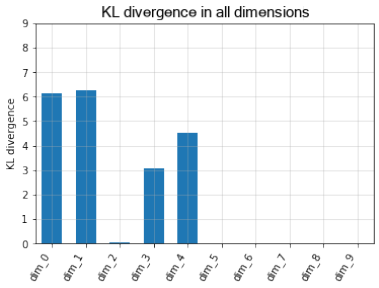
(b) Dimension wise KL in 50000 steps



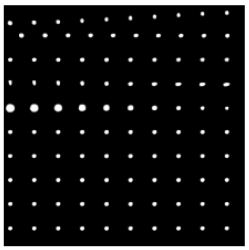
(c) Images traversal of dimensions



(d) Dimension wised KL in training

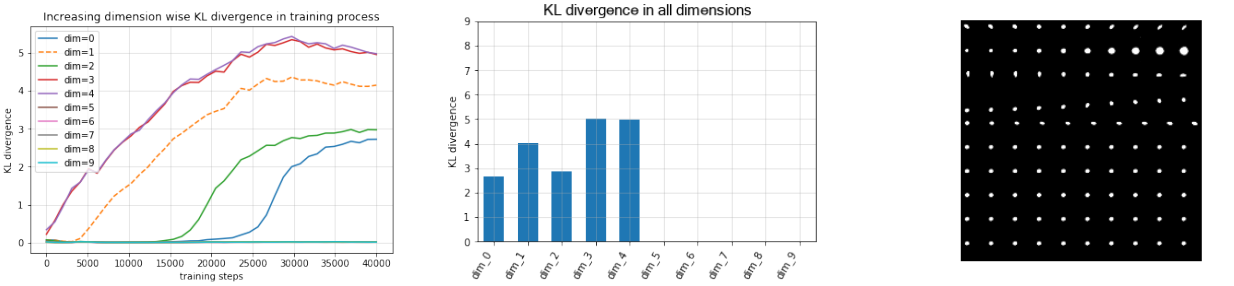


(e) Dimension wise KL in 50000 steps

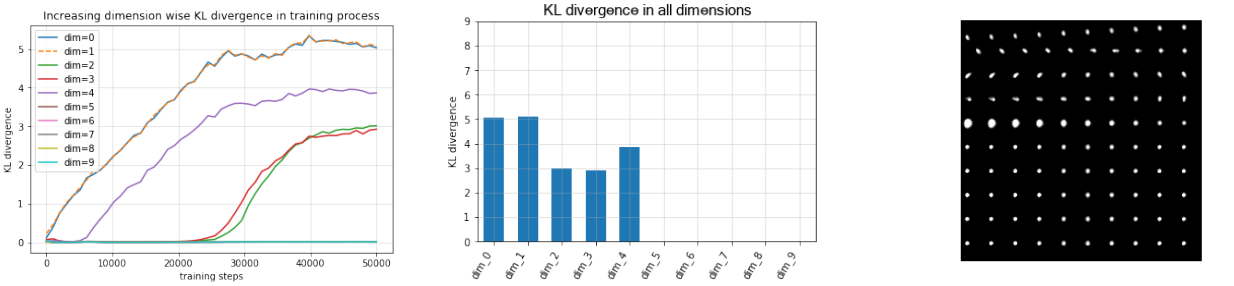


(f) Images traversal of dimensions

Figure 17. Meaningful dimensions visualization for $C_{max} = 20$, end = 40000 (different random seeds). The KL divergences of target dimension (0-4 dimension) increase one by one during training (a). The KL divergences in different dimensions are different amounts after training (b). As the the image traversal results (c) shows, the meaningful dimensions are learned in 0-4 dims. So as to (d), (e) and (f).

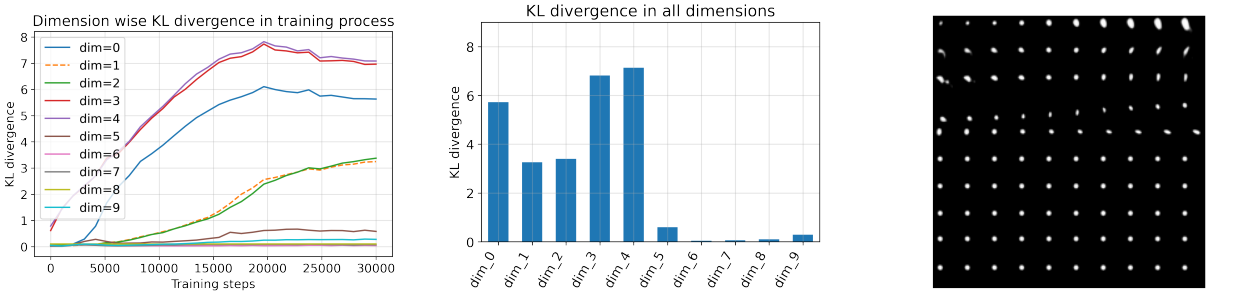


(a) Dimension wised KL in training (b) Dimension wise KL in 50000 steps (c) Images traversal of dimensions

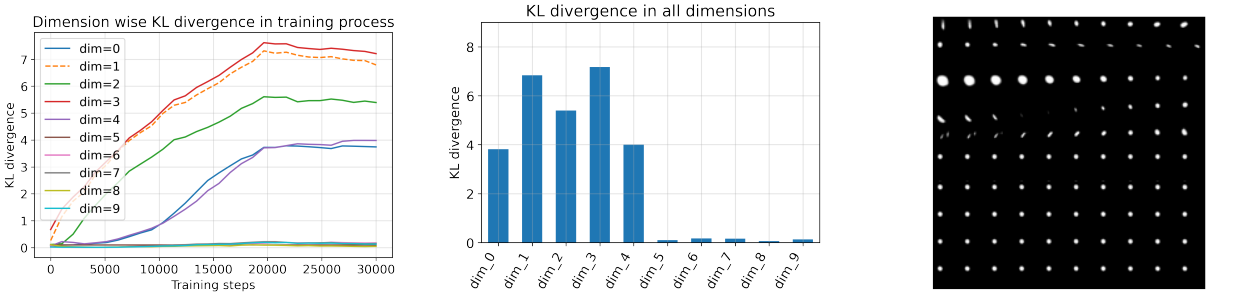


(d) Dimension wised KL in training (e) Dimension wise KL in 50000 steps (f) Images traversal of dimensions

Figure 18. Meaningful dimensions visualization for $C_{max} = 20$, end = 50000 (different random seeds). The KL divergences of target dimension (0-4 dimension) increase one by one during training (a). The KL divergences in different dimensions are different amounts after training (b). As the the image traversal results (c) shows, the meaningful dimensions are learned in 0-4 dims. So as to (d), (e) and (f).

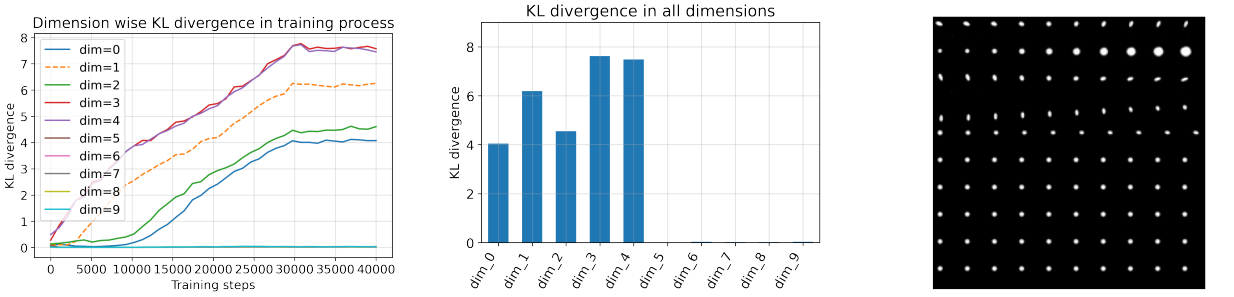


(a) Dimension wised KL in training (b) Dimension wise KL in 50000 steps (c) Images traversal of dimensions

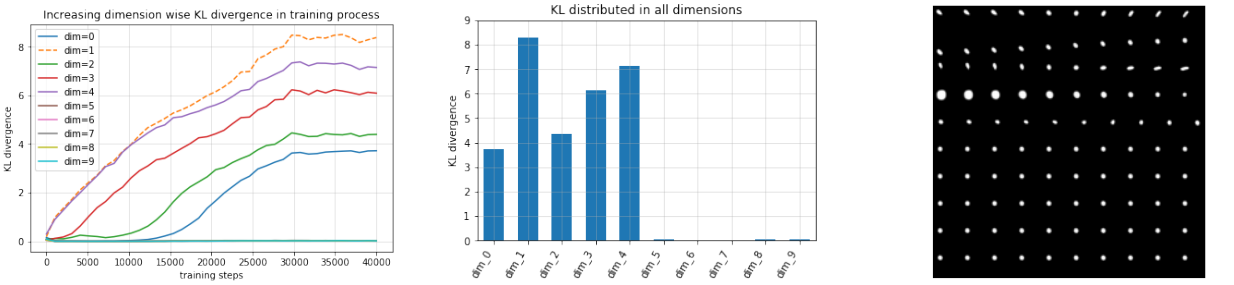


(d) Dimension wised KL in training (e) Dimension wise KL in 50000 steps (f) Images traversal of dimensions

Figure 19. Meaningful dimensions visualization for $C_{max} = 30$, end = 30000 (different random seeds). The KL divergences of target dimension (0-4 dimension) increase one by one during training (a). The KL divergences in different dimensions are different amounts after training (b). As the the image traversal results (c) shows, the meaningful dimensions are learned in 0-4 dims. So as to (d), (e) and (f).

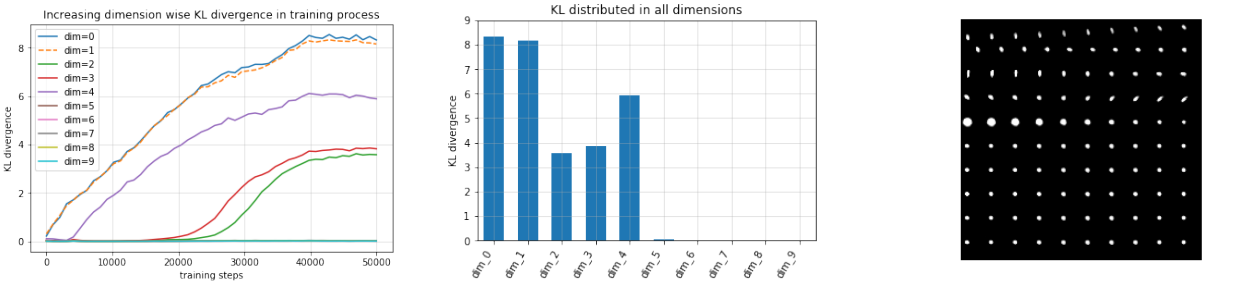


(a) Dimension wised KL in training (b) Dimension wise KL in 50000 steps (c) Images traversal of dimensions

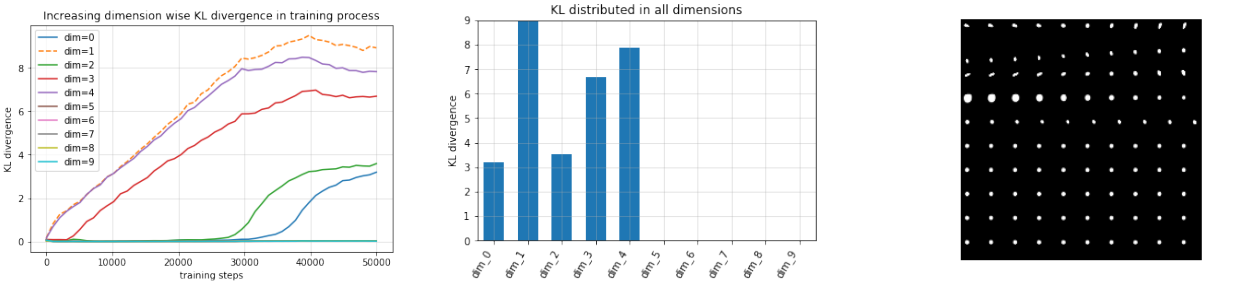


(d) Dimension wised KL in training (e) Dimension wise KL in 50000 steps (f) Images traversal of dimensions

Figure 20. Meaningful dimensions visualization for $C_{max} = 30$, end = 40000 (different random seeds). The KL divergences of target dimension (0-4 dimension) increase one by one during training (a). The KL divergences in different dimensions are different amounts after training (b). As the the image traversal results (c) shows, the meaningful dimensions are learned in 0-4 dims. So as to (d), (e) and (f).



(a) Dimension wised KL in training (b) Dimension wise KL in 50000 steps (c) Images traversal of dimensions

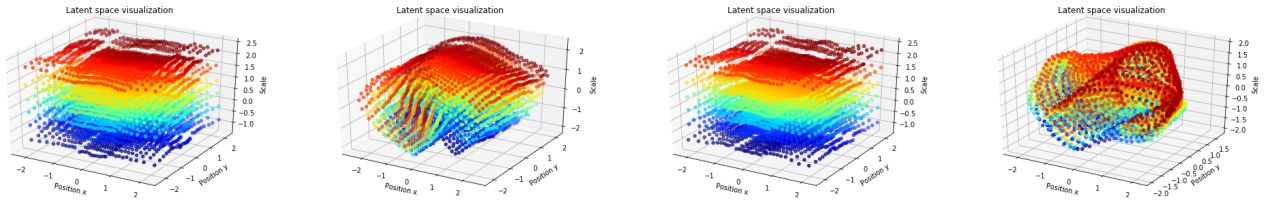


(d) Dimension wised KL in training (e) Dimension wise KL in 50000 steps (f) Images traversal of dimensions

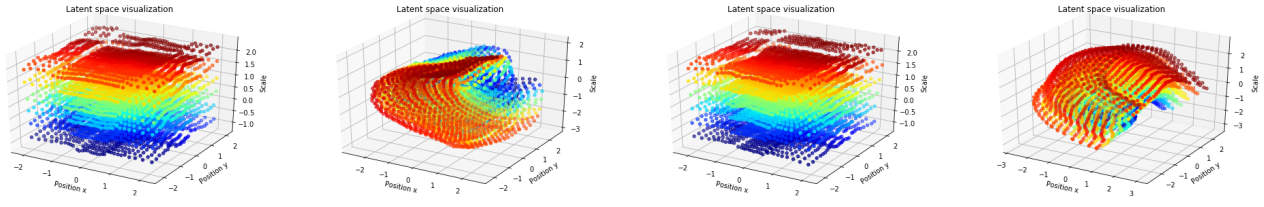
Figure 21. Meaningful dimensions visualization for $C_{max} = 30$, end = 50000 (different random seeds). The KL divergences of target dimension (0-4 dimension) increase one by one during training (a). The KL divergences in different dimensions are different amounts after training (b). As the the image traversal results (c) shows, the meaningful dimensions are learned in 0-4 dims. So as to (d), (e) and (f).

J. More Latent Space Visualizations

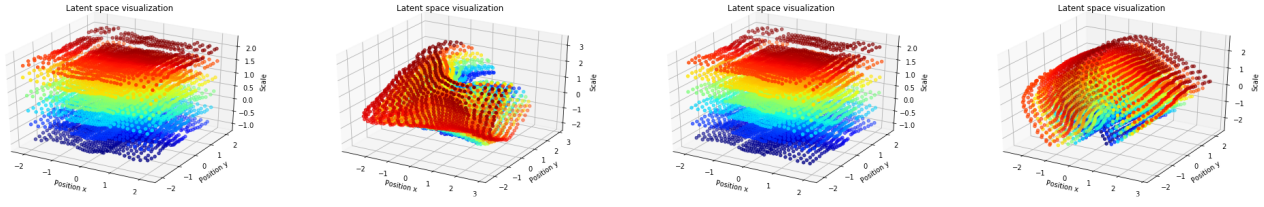
To illustrate that groupifying suppresses the latent space collapse in VAE-based methods is not an exception, we provide more visualizations of the latent space visualization of groupified VAEs. The original and groupified β-VAEs, AnnealVAEs, FactorVAEs and β-TCVAEs are shown in Figure 22, Figure 23, Figure 24 and Figure 25 respectively, which implies that the latent space of groupified VAEs are organized better than the original ones.



(a) $\beta = 10$, Groupified (b) $\beta = 10$, Original (c) $\beta = 10$, Groupified (d) $\beta = 10$, Original

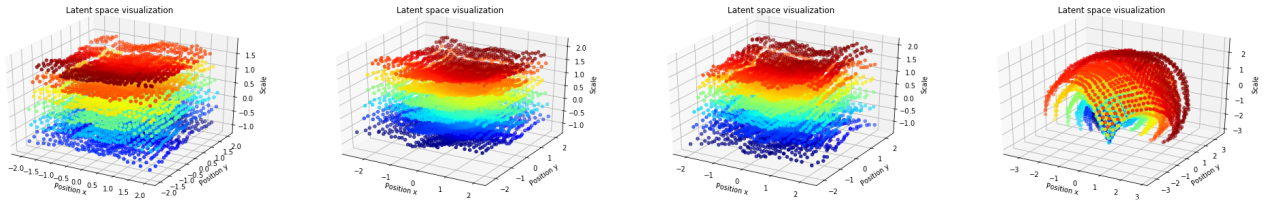


(e) $\beta = 20$, Groupified (f) $\beta = 20$, Original (g) $\beta = 20$, Groupified (h) $\beta = 20$, Original

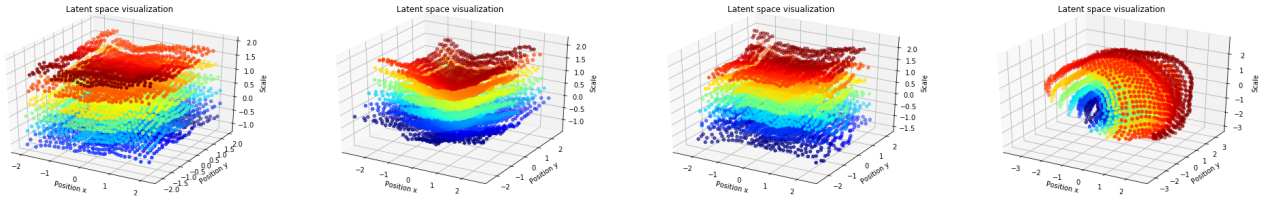


(i) $\beta = 30$, Groupified (j) $\beta = 30$, Original (k) $\beta = 30$, Groupified (l) $\beta = 30$, Original

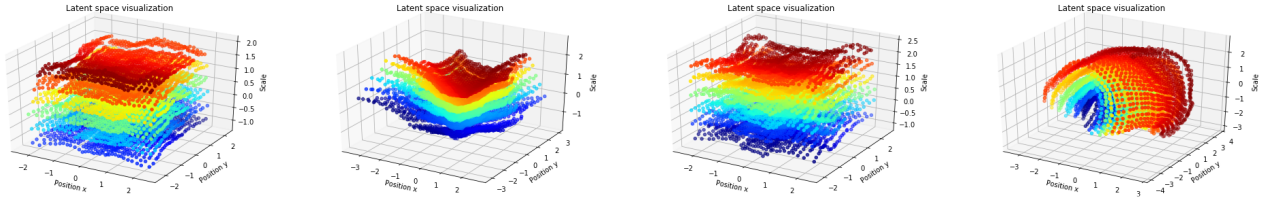
Figure 22. The latent space span by the groupified and original β-VAE. The models with same hyperparameter are trianed with different random seeds. The localization of each point is the disentangled representation of the corresponding image. And an ideal result is all the points form a cube and color variation is continuous. Higher hyper-parameter β results collapse of latent space. The collapse is suppressed by groupifying which lead to better disentanglement.



(a) $C_{max} = 10$, Groupified (b) $C_{max} = 10$, Original (c) $C_{max} = 10$, Groupified (d) $C_{max} = 10$, Original

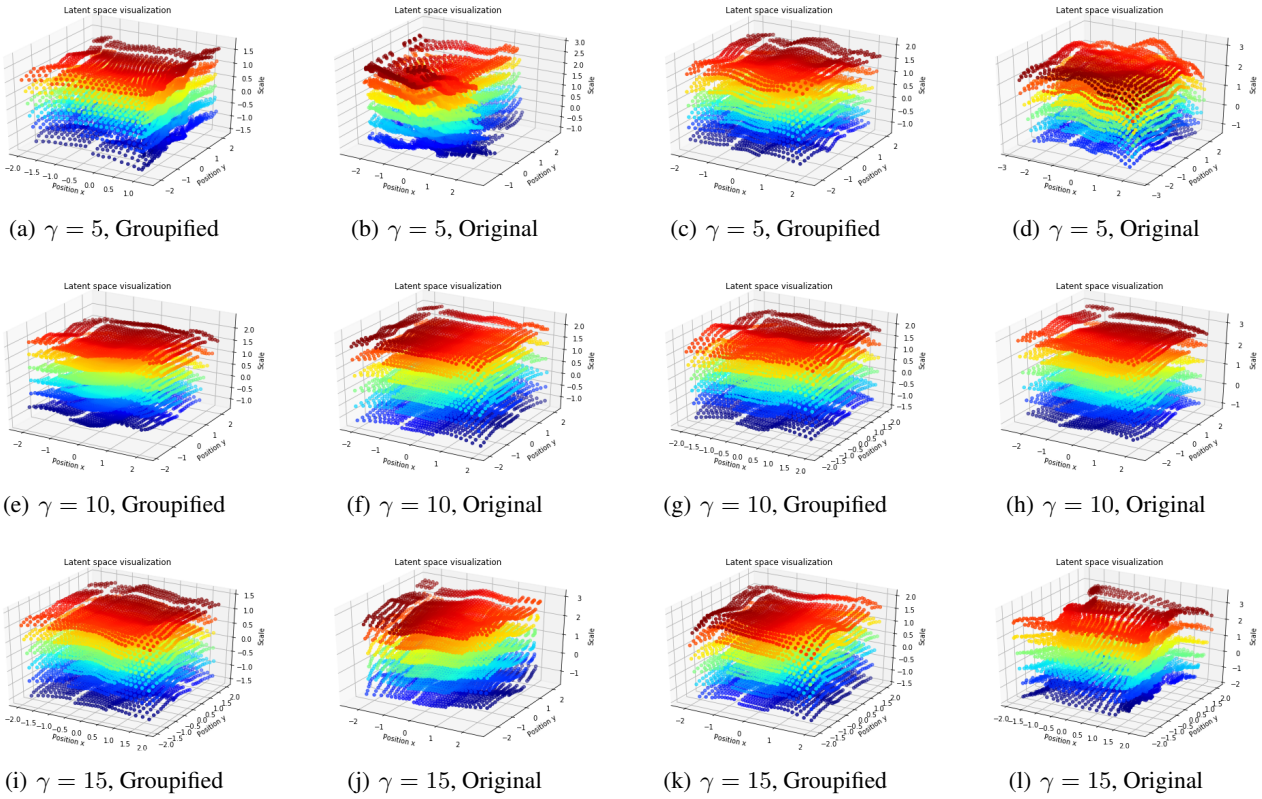


(e) $C_{max} = 20$, Groupified (f) $C_{max} = 20$, Original (g) $C_{max} = 20$, Groupified (h) $C_{max} = 20$, Original



(i) $C_{max} = 30$, Groupified (j) $C_{max} = 30$, Original (k) $C_{max} = 30$, Groupified (l) $C_{max} = 30$, Original

Figure 23. The latent space span by the groupified and original AnnealVAE. The models with same hyperparameter are trianed with different random seeds. The localization of each point is the disentangled representation of the corresponding image. And an ideal result is all the points form a cube and color variation is continuous. Higher hyper-parameter C_{max} results collapse of latent space. The collapse is suppressed by groupifying which lead to better disentanglement.

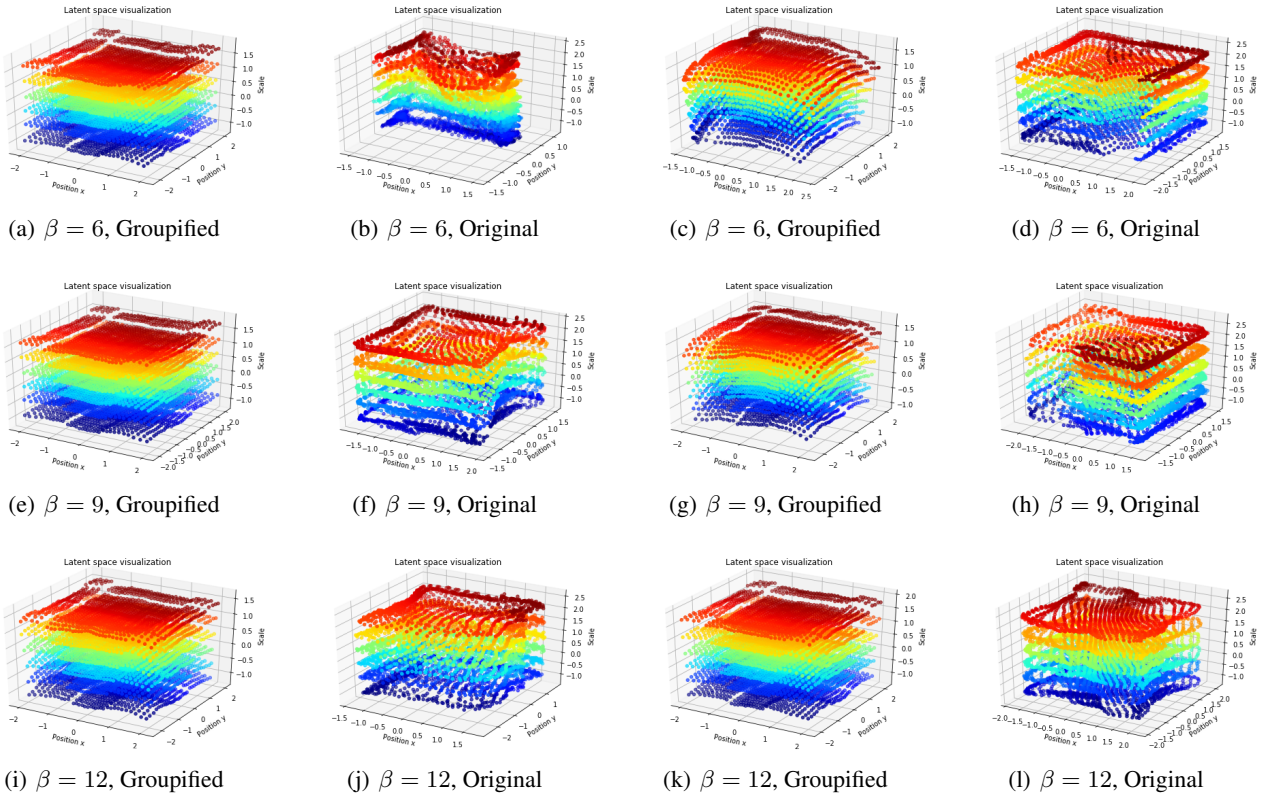


(a) $\gamma = 5$, Groupified (b) $\gamma = 5$, Original (c) $\gamma = 5$, Groupified (d) $\gamma = 5$, Original

(e) $\gamma = 10$, Groupified (f) $\gamma = 10$, Original (g) $\gamma = 10$, Groupified (h) $\gamma = 10$, Original

(i) $\gamma = 15$, Groupified (j) $\gamma = 15$, Original (k) $\gamma = 15$, Groupified (l) $\gamma = 15$, Original

Figure 24. The latent space span by the groupified and original FactorVAE. The models with same hyperparameter are trianed with different random seeds. The localization of each point is the disentangled representation of the corresponding image. And an ideal result is all the points form a cube and color variation is continuous. The collapse of the latent space is suppressed by groupifying which lead to better disentanglement.



(a) $\beta = 6$, Groupified (b) $\beta = 6$, Original (c) $\beta = 6$, Groupified (d) $\beta = 6$, Original

(e) $\beta = 9$, Groupified (f) $\beta = 9$, Original (g) $\beta = 9$, Groupified (h) $\beta = 9$, Original

(i) $\beta = 12$, Groupified (j) $\beta = 12$, Original (k) $\beta = 12$, Groupified (l) $\beta = 12$, Original

Figure 25. The latent space span by the groupified and original β-TCVAE. The models with same hyperparameter are trianed with different random seeds. The localization of each point is the disentangled representation of the corresponding image. And an ideal result is all the points form a cube and color variation is continuous. The collapse of the latent space is suppressed by groupifying which lead to better disentanglement.

References

Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., Mohamed, S., and Lerchner, A. beta-vae: Learning basic visual concepts with a constrained variational framework. 2016.

Hintze, J. L. and Nelson, R. D. Violin plots: a box plot-density trace synergism. *The American Statistician*, 52(2):181–184, 1998.

Kim, H. and Mnih, A. Disentangling by factorising. 2018.

Locatello, F., Bauer, S., Lucic, M., Raetsch, G., Gelly, S., Schölkopf, B., and Bachem, O. Challenging common assumptions in the unsupervised learning of disentangled representations. In *ICML*, pp. 4114–4124, 2019.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., and Lerer, A. Automatic differentiation in pytorch. 2017.

Reed, S. E., Zhang, Y., Zhang, Y., and Lee, H. Deep visual analogy-making. In *NeurIPS*, 2015.