LOOKING BEYOND THE SURFACE WITH CONTRASTIVE LEARNING WITH ANTI-CONTRASTIVE REGULARIZATION (CLEAR)

Anonymous authors
Paper under double-blind review

ABSTRACT

Learning representations that are robust to superficial sources of variability is important to ensure such variability does not impact downstream tasks. For instance, in healthcare applications, we might like to learn features that are useful for identifying pathology, yet have similar distributions across diverse demographic groups, leading to more accurate and equitable diagnoses regardless of background or surface characteristics. More broadly, this capability can improve the generalizability of our representations by mitigating unwanted effects of variability not seen during training. In this work, we suppose that data representations can be semantically separated into two components: $content$ and $style$. The $content$ consists of information needed for downstream tasks – for example, it is predictive of the class label in a downstream classification problem – whereas the $style$ consists of attributes that are superficial in the sense that they are irrelevant to downstream tasks, yet may compromise performance due to associations observed in training data that do not generalize. Here we propose a weakly supervised framework, **C**ontrastive **LE**arning with **A**nti-contrastive **R**egularization (CLEAR), to effectively disentangle $content$ and $style$ in the latent space of a Variational Autoencoder (VAE). Our anti-contrastive penalty, which we call Pair Switching (PS), uses a novel label flipping approach to ensure content is recognized effectively and limited to the $content$ features. We perform experiments to quantitatively and qualitatively evaluate CLEAR-VAE across distinct data modalities. We then analyze the trade-off between disentanglement and ELBO, and the impact of various hyperparameters within our framework. Our results show that using disentangled representations from CLEAR-VAE, we can: (a) swap and interpolate $content$ and $style$ between any pair of samples, and (b) improve downstream classification performance in the presence of previously unseen combinations of $content$ and $style$.

1 INTRODUCTION

The information from an input can semantically separated into two parts: *content* and *style* (Mathieu et al., 2016; Bouchacourt et al., 2018; Hamaguchi et al., 2019). In a well-defined classification problem, where the ground truth outcome labels are provided, the *content* information directly relates to the outcome label. The *style* information, which may or may not correspond to a second set of labels in the dataset, is in principle irrelevant to the outcome of interest. However, the *style* can nevertheless influence a model's classification performance and generalizability due to spurious correlations observed in training data. In principle, such correlations can be avoided by collecting an arbitrarily large, representative dataset, but in practice they are common, for example due to 1. the rarity of a given outcome of interest; 2. heterogeneous *style* distributions across outcome groups, including due to biased outcomes; or 3. distribution shifts taking place between training and test time.

Several unsupervised disentangled representation learning methods have been developed in recent years (Higgins et al., 2017; Burgess et al., 2018; Kim & Mnih, 2018; Chen et al., 2018). These unsupervised methods achieve disentanglement through strong penalty on the KL regularization term in the ELBO, which factorizes the encoded representation. Consequently, each dimension is trained to be related to an independent feature in the modality. However, if we are interested in

using encoded information for downstream classification tasks, a key limitation is the difficulty of identifying the *content*-specific latent factors. The disentanglement can be directly achieved under a fully supervised method, where labels for both *content* and *style* are provided to the model, but it is inapplicable in a real-world scenario. In practice, however, we are unlikely to have a set of *style* labels that explicitly delineate unwanted sources of variability. Our aim is therefore to develop a weakly supervised learning method that not only learns *content* representations under the supervision of the outcome label, but also disentangles *style* from *content* without knowing the styles *a priori* or having a second set of labels indicating which of several styles is present. Instead, we use only a single set of labels, the so-called *content* labels, to form contrastive pairs solely involved in the training objective. However, ML-VAE still relies on grouping information (a weaker form of supervision compared to direct label supervision) to generate high-quality samples when implementing "accumulating evidence" during inference time. (Bouchacourt et al., 2018)

To effectively disentangle *content* and *style* under such conditions, we combine contrastive learning with a novel, anti-contrastive regularization term for the style features that we call Pair Switching (PS). We first apply a contrastive loss to ensure examples with matching labels have similar content representations, which prevents information about the style from being present among these features. However, in the absence of additional regularization, information about content will appear among the style features. To combat this, we then apply our PS term to encourage examples with *different* content labels to have similar style representations. This has two indirect, related effects. First, it prevents information about the content from appearing among the style features. Second, it encourages style features to have similar distributions across content labels.

This indirect matching of style distributions across content labels is most effective when there are multiple labels, each with partial representation of the full range of style variability. Indeed, our approach was motivated by circumstances where (a) style distributions differ between labels, yet (b) it is reasonable to suppose that style distributions would be consistent between labels in a large, representative, unbiased dataset. For example, many medical diagnoses are associated with demographic characteristics in observational datasets due to disparities in access to care, but it is reasonable to suppose that in an ideal, unbiased dataset, such associations would not exist.

In our framework, the VAE architecture does not require parallel branches of networks with shared weights, nor do we need pre-paired data for model training. Instead, VAE under Contrastive LEarning with Anti-contrastive Regularization (CLEAR-VAE) allows us to disentangle and recognize *content* and *style* from content labels alone – in other words, the standard supervised classification setup – by implicitly assuming that style distributions should be similar, on average, across each class.

We perform extensive experiments on different datasets to evaluate the performance of disentanglement in CLEAR-VAE and its efficacy in improving model generalizability. First, in our qualitative analysis, we present one-to-one "translations" between pairs of samples, where they exchange *content* or *style* with each other. We demonstrate that our approach achieves effective disentanglement, successfully extracting style characteristics that are known to us but unknown to the model through weakly supervised contrastive learning. Finally, we will quantitatively show CLEAR-VAE can improve a downstream model's generalizability on unseen combinations of *content* and *style*.

In summary, our contributions are as follows:

- Propose a new VAE-based framework to learn disentangled representations of content and style representation through weakly supervised contrastive regularization.
- Introduce Pair Switching (PS), a novel anti-contrastive penalty that enforces disentanglement by encouraging style features to have similar distributions across content labels.
- Demonstrate that CLEAR-VAE is able to disentangle *content* and *style* and recognize the semantic difference between them.
- Leverage CLEAR-VAE to enhance classification models' generalizability to previously unseen combinations of *style* and *content*.

2 Related work

Semantic Disentanglement in VAE. There are various frameworks to achieve the semantic disentanglement in VAE's latent representation. Mathieu et al. (2016) achieve the disentanglement

through adversarial training on the VAE decoder. Bouchacourt et al. (2018) organize samples into groups based on the ground truth labels within the mini-batches and extract group-level content representation. Hamaguchi et al. (2019) learn the *content* through similarity regularization. These methods are designed to semantically separate latent representations into two components: *content* and *style*. They provide evidence that an effective disentanglement can enhance model generalizability on unseen combinations of *content* and *style* (Mathieu et al., 2016; Bouchacourt et al., 2018; Hamaguchi et al., 2019). Different from Mathieu et al. (2016) and Hamaguchi et al. (2019), our method does not require data to be paired in advance. Bouchacourt et al. (2018) focusing on regularizing *content*, but we achieve semantic disentanglement by considering both *content* and *style*.

Contrastive Learning. Contrastive learning was originally developed as a form of supervised learning (Chopra et al., 2005; Koch et al., 2015). More recently, it has gain significant popularity in self-supervised learning (Oord et al., 2018; Frosst et al., 2019; He et al., 2020; Chen et al., 2020; Damrich et al., 2022). In these methods, positive and negative pairs are generated from the data, and the representations are learnt by calibrating the similarity or distance between the elements in each pair. A positive pair indicates the data belong to the same group, whereas a negative pair indicates the data come from different groups. In general, a contrastive learning method aims to discover an embedding space that effectively represents the data, where positive pairs are clustered together while negative pairs remain far apart.

Label Flipping. Label flipping refers to the technique of altering the ground truth labels during the model training stage. It is commonly adopted as "data poisoning attack" to trick the model into incorrect pattern recognition, and is commonly implemented to evaluate models' robustness when trained under adversarial settings (Xiao et al., 2012; Rosenfeld et al., 2020). Recently, there has been discussion around utilizing label flipping with contrastive learning. Li et al. (2022) propose a contrastive-learning framework that produces robust pre-trained representations under adversarial label flipping. Ren et al. (2021) implement label flipping in a more unusual way by reversing the labels for hard negative pairs to ease the contrastive optimization, since hard negative pairs can be semantically assumed to be positives. Different from the aforementioned approaches, we use label flipping to disconnect *style* latent variables and the content label. Specifically, we switch the labels of all contrastive pairs, turning positives into negatives and vice versa.

3 METHODS

3.1 CLEAR-VAE OVERVIEW

VAE, as a generative probabilistic model, consists of two components: an encoder $q_\phi(\boldsymbol{z}|\boldsymbol{x})$ that maps input $\boldsymbol{x}$ to $\mathcal{N}\big(\boldsymbol{z}; \boldsymbol{\mu}_\phi(\boldsymbol{x}), \boldsymbol{\Sigma}_\phi(\boldsymbol{x})\big)$ and a decoder $p_\theta(\boldsymbol{x}|\boldsymbol{z})$ that reconstructs $\boldsymbol{x}$ based on a stochastic latent representation $\boldsymbol{z}$ (Kingma, 2013). A standard VAE can be optimized under the Evidence Lower Bound (ELBO):

$$\mathcal{L}_{\text{VAE}} = \text{ELBO} = \mathbb{E}_{q_\phi}\left[\log p_\theta(\boldsymbol{x}|\boldsymbol{z})\right] - D_{\text{KL}}(q_\phi(\boldsymbol{z}|\boldsymbol{x})\|p(\boldsymbol{z})) \leq \log p_\theta(\boldsymbol{x}) \tag{1}$$

In Equation 1, the first term can be interpreted as the reconstruction loss: mean squared error if $\boldsymbol{x}$ is assumed to follow a Gaussian distribution, or binary cross-entropy if $\boldsymbol{x}$ is binary. Then second term regularizes the divergence between encoder and $p(\boldsymbol{z})$, the prior of $\boldsymbol{z}$, which is usually assumed to be $\mathcal{N}(0, \boldsymbol{I})$. The reparameterization trick is a key technique that separates the deterministic and stochastic parts of sampling from a latent distribution and allow the back-propagation to bypass the randomness in the probabilistic model Kingma (2013). β-VAE is a straightforward extension to learn disentangled representation from the data by multiplying the KL regularization term with a penalizing coefficient β (typically, $\beta > 1$) (Higgins et al., 2017). A stronger penalization can lead to a more factorized latent representation.

Figure 1 illustrates a schematic workflow of the framework. In CL, we assume the latent representation $\boldsymbol{z} = (\boldsymbol{z}^{(c)}, \boldsymbol{z}^{(s)})$ consists of two components $\boldsymbol{z}^{(c)}$ and $\boldsymbol{z}^{(s)}$, which denote the *content* latent variables and *style* latent variables, respectively. The probabilistic encoder extracts the variational representations of each component. In order to achieve disentanglement between the two components, we use contrastive regularization terms to encourage the separation and distinction between them. In the variational latent space, contrastive regularization will cluster the $\boldsymbol{z}^{(c)}$ representations together under the supervision of class labels, while enforcing the $\boldsymbol{z}^{(s)}$ representations to be as ambiguous as possible with respect to the class labels. Therefore, we optimize the CLEAR-VAE model

using the following loss function.

$$\mathcal{L} = \mathcal{L}_{\text{VAE}}(\beta) + \alpha\mathcal{L}^{(c)}_{\text{SNN}} + \alpha T(\mathcal{L}^{(s)}_{\text{SNN}}) \tag{2}$$

In Equation 2, $\mathcal{L}_{\text{VAE}}$ is the VAE objective function, and $\mathcal{L}_{\text{SNN}}$'s are the contrastive regularization terms. $T(\cdot)$ is the transformation, where we can apply PS, that converts contrastive regularization into anti-contrastive, thereby dissociating the *style* representation from the ground truth label of *content* (see more detail in Sec. 3.4). Additionally, β and α are the coefficients for KL regularization and contrastive regularization, respectively.

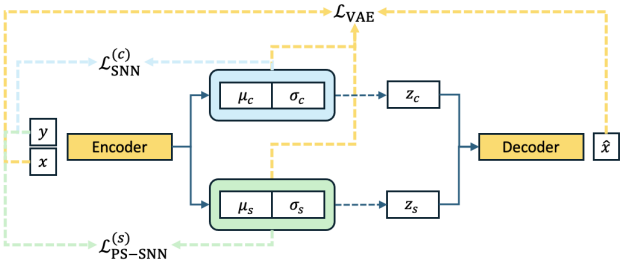


Figure 1: Schematic illustration of the CLEAR-VAE.

3.2 VAE LATENT FACTOR PARTITIONING

Given the partition $\boldsymbol{z} = (\boldsymbol{z}^{(c)}, \boldsymbol{z}^{(s)})$, we can reasonably assume they are independent both *a priori* and *a posteriori* conditional on $\boldsymbol{x}$. Therefore, we can decompose the KL regularization term into two pieces: one for $\boldsymbol{z}^{(c)}$ and the other for $\boldsymbol{z}^{(s)}$ (Appx. A.1). Consequently, the loss for a standard VAE with the latent feature partition can be succinctly written as

$$\mathcal{L}_{\text{VAE}} = -E_{q_\phi}\left[\log p_\theta(\boldsymbol{x}|\boldsymbol{z})\right] + D^{(c)}_{\text{KL}} + D^{(s)}_{\text{KL}} \tag{3}$$

Although we have a decomposition of VAE loss, $\mathcal{L}_{\text{VAE}}$ on its own is insufficient to teach the model to identify and differentiate $\boldsymbol{z}^{(c)}$ and $\boldsymbol{z}^{(s)}$. Thus, we have to incorporate contrastive regularization to enforce disentanglement of our latent representations. This regularization must achieve two goals. First, it has to encourage similarity in $\boldsymbol{z}^{(c)}$'s from samples from the same class. Second, there should be a significant gap between $I(\boldsymbol{z}^c; y)$ and $I(\boldsymbol{z}^s; y)$ where y is the label for *content*. Namely, $\boldsymbol{z}^{(c)}$ and $\boldsymbol{z}^{(s)}$ are disentangled.

3.3 SNN AS A CONTRASTIVE REGULARIZATION

The InfoNCE loss and NT-Xent loss have been commonly used as objectives in contrastive self-supervised learning, which allows only one positive pair per observation within a batch (Oord et al., 2018; Chen et al., 2020). IN contrast, the Soft Nearest Neighbors (SNN) loss, which is defined based on a distance metric rather than a similarity measure, allows each observation to be associated (*i.e.*, paired) with multiple other observations (Salakhutdinov & Hinton, 2007; Frosst et al., 2019). SNN may be viewed as the multi-positive-pair analogue of the NT-Xent loss when replacing $-\|\boldsymbol{z}_i, \boldsymbol{z}_j\|^2$ with the cosine similarity.

Therefore, we use a modified SNN loss to encourage $\boldsymbol{z}^{(c)}$ to capture content features in the data:

$$\mathcal{L}^{(c)}_{\text{SNN}} = \frac{1}{B}\sum_{i=1}^{B} -\log \frac{\sum_{i\neq j, j=1\ldots B} \mathbb{I}_{[y_i=y_j]} \exp\left\{f(p(\boldsymbol{z}^{(c)}_i|\boldsymbol{x}), p(\boldsymbol{z}^{(c)}_j|\boldsymbol{x})/\tau)\right\}}{\sum_{i\neq j, j=1\ldots B} \exp\left\{f(p(\boldsymbol{z}^{(c)}_i|\boldsymbol{x}), p(\boldsymbol{z}^{(c)}_j|\boldsymbol{x})/\tau\right\}} \tag{4}$$

where $p(\boldsymbol{z}^{(c)}_i|\boldsymbol{x}) = \mathcal{N}\left(\boldsymbol{z}^{(c)}_i \middle| \boldsymbol{\mu}^{(c)}(\boldsymbol{x}_i), \boldsymbol{\Sigma}^{(c)}(\boldsymbol{x}_i)\right)$ and $p(\boldsymbol{z}^{(c)}_j|\boldsymbol{x}) = \mathcal{N}\left(\boldsymbol{z}^{(c)}_j \middle| \boldsymbol{\mu}^{(c)}(\boldsymbol{x}_j), \boldsymbol{\Sigma}^{(c)}(\boldsymbol{x}_j)\right)$ are the variational encoded representation, and f is calculated between the two distributions. $f = g$ if g is a similarity measure (e.g. cosine similarity between $\boldsymbol{\mu}_\theta$'s), and $f = -g$ if g is a distance metric (e.g. Jeffrey divergence). Overall, f is positively associated with the similarity. Lastly, temperature τ control the model sensitivity to the differences in pairs.

3.4 PAIR SWITCHING AS ANTI-CONTRASTIVE REGULARIZATION FOR DISENTANGLEMENT

Apart from the using a contrastive regularization to learning meaningful represent, we also introduce an anti-contrastive regularization to (a) prevent content information from entering the style features, and (b) encourage style distributions to match across content labels. Therefore, the anti-contrastive regularization further encourage the disentanglement.

$$\text{pos}_i^{(c)} := \sum_{i \neq j, j=1\ldots B} \mathbb{I}_{[y_i = y_j]} \exp\left\{ f(p(\boldsymbol{z}_i^{(c)}|\boldsymbol{x}), p(\boldsymbol{z}_j^{(c)}|\boldsymbol{x})/\tau) \right\} \tag{5}$$

$$\text{neg}_i^{(c)} := \sum_{i \neq j, j=1\ldots B} \mathbb{I}_{[y_i \neq y_j]} \exp\left\{ f(p(\boldsymbol{z}_i^{(c)}|\boldsymbol{x}), p(\boldsymbol{z}_j^{(c)}|\boldsymbol{x})/\tau) \right\} \tag{6}$$

The SNN loss of the i-th individual can be abbreviated as $l_i^{(c)} = -\log \frac{\text{pos}_i^{(c)}}{\text{pos}_i^{(c)} + \text{neg}_i^{(c)}}$. Note that $l_i^{(c)}$ is always non-negative. The goal of minimizing it is to have $\text{pos}_i^{(c)} \gg \text{neg}_i^{(c)}$ so that we maximize the similarity within positive pairs and minimize it within negative pairs. If we minimize over $-\mathcal{L}_{\text{SNN}}^{(c)}$, we will encourage the representation to be ambiguous about the supervised label. Thus, to dissociate $\boldsymbol{z}^{(s)}$ with y, we can simply use the negative SNN loss applied to $\boldsymbol{z}_s$, which is $-\mathcal{L}_{\text{SNN}}^{(s)} = \frac{1}{B}\sum_{i=1}^{B} \log \frac{\text{pos}_i^{(s)}}{\text{pos}_i^{(s)} + \text{neg}_i^{(s)}}$. However, this introduces a negative value into the loss, thereby complicating the minimization of $\mathcal{L}_{\text{VAE}}$.

Alternatively, we can flip the label for positive and negative pairs in $\mathcal{L}_{\text{SNN}^{(s)}}$ to achieve the same goal of disconnecting $\boldsymbol{z}^{(s)}$ and y. Consequently, the regularization term for $\boldsymbol{z}^{(s)}$ becomes

$$T(-\mathcal{L}_{\text{SNN}}^{(s)}) = \frac{1}{B}\sum_{i=1}^{B} -\log \frac{\text{neg}_i^{(s)}}{\text{neg}_i^{(s)} + \text{pos}_i^{(s)}} \tag{7}$$

$$= \frac{1}{B}\sum_{i=1}^{B} -\log \frac{\sum_{i \neq j, j=1\ldots B} \mathbb{I}_{[y_i \neq y_j]} \exp(f(\boldsymbol{z}_i^{(c)}, \boldsymbol{z}_j^{(c)})/\tau)}{\sum_{i \neq j, j=1\ldots B} \exp(f(\boldsymbol{z}_i^{(c)}, \boldsymbol{z}_j^{(c)})/\tau)} \tag{8}$$

Note that this is always non-negative. Minimizing of $T(-\mathcal{L}_{\text{SNN}}^{(s)})$ will encourage the similarity within positive pairs to decrease, while increase similarity in negative pairs until we reach an equilibrium. This contradiction indicates an ambiguity between $\boldsymbol{z}_s$ and y. In the end, we disentangle $\boldsymbol{z}_c$ and $\boldsymbol{z}_s$.

4 MUTUAL INFORMATION GAP BETWEEN GROUPS

Chen et al. (2018) proposed a interpretable, classifier-free metric based on $I(z_j; y)/H(y)$, the normalized mutual information between a latent variable z_j and a ground truth factor y. The complete metric for a label y, known as the Mutual Information Gap (MIG), is defined as the difference between the top two latent variables with the highest normalized mutual information Chen et al. (2018). Whereas the original definition is an average over all possible ground truth labels, in our scenario, we consider MIG only for the class label y:

$$\text{MIG}(y) = \frac{1}{H(y)} \left(I(z^*; y) - \max_{z_j \neq z^*} I(z_j; y) \right) \tag{9}$$

The $\text{MIG}(y)$ measures the degree of disentanglement at level of individual latent variables. We adapt it to our specific scenario, where $\boldsymbol{z} = (\boldsymbol{z}^{(c)}, \boldsymbol{z}^{(s)})$, as follows:

$$\text{gMIG}(y) = \frac{1}{H(y)} \left(\frac{1}{d_c}\sum_{j=1}^{d_c} I(z_j^{(c)}; y) - \frac{1}{d_s}\sum_{j=1}^{d_s} I(z_j^{(s)}; y) \right) \tag{10}$$

We call this modified version the group mutual information gap (gMIG). It quantifies the gap between $\boldsymbol{z}_c$'s average association to y and $\boldsymbol{z}_s$'s average association to y. gMIG is bounded between -1 and 1.

5 EXPERIMENTAL SETUP

5.1 DATASETS

We evaluate the CLEAR-VAE framework on images and texts. Table 1 provides the definitions of *content* and *style* in each dataset, where $style$ convey both nontrivial and trivial signals. For each dataset, we perform two sets of experiments. First, we include all combinations of content and style in both training and testing sets to illustrate the semantic disentanglement. Secondly, to evaluate CLEAR-VAE encoder's generalizability on unseen *styles*, we use the labels for nontrivial style features to ensure that the *styles* in the training and testing sets differ in all classification tasks. More detailed dataset descriptions are given below.

Styled-MNIST. We enhance the MNIST dataset (LeCun et al., 1998) with noticeable style features to create a more discriminative and expressive feature space for *style*. We use the corruption transforming methods from Mu & Gilmer (2019). In our experiments, we randomly assign each digit to a transformation from the set {`identity/unchanged, stripe, zigzag, canny edge, tiny scaled, brightness`}. Figure 2 shows a batch of random samples in the dataset.

Figure 2: Styled-MNIST exemplar data.

CelebA. CelebA is a large-scale dataset of celebrity face images with 40 labeled attributes Liu et al. (2015). The dataset has a rich and diversified feature space. In our experiments, we treat the combination of gender and smile as the $content$ label. The rest of the attributes will be considered as *style* in our setting. In the classification experiments, we use hair color for the assignment of training and testing splits.

Amazon Product Reviews. We take a subset of the original Amazon Product Review dataset (Hou et al., 2024), where we randomly select 50,000 text reviews from product categories: {`all beauty, digital music, handmade product, health and personal care`}. We treat each review's rating as the *content* and its relevant product category as a *style* feature.

Table 1: *Content* and *style* in the experimental datasets. The **style** feature is utilized as the reference attribute to evaluate the quality of swapping and interpolation experiments and to split training and testing data for downstream classification tasks.

	modality	content label (y)	style
Styled-MNIST	image ($1 \times 28 \times 28$)	digit	**applied corruption** & handwriting strokes
CelebA	image ($3 \times 64 \times 64$)	gender $\times$ smiling	**hair color** & remaining characteristics
Amazon Product Review	text	ratings (sentiment)	**product category** & specific writing style

5.2 QUALITATIVE EVALUATION

Content and style swapping. We assess the quality of representation learning by manipulating $\boldsymbol{z}^{(c)}$ and $\boldsymbol{z}^{(s)}$ from testing samples. Specifically, we extract the latent representations $\boldsymbol{z}^{(c)}$ and $\boldsymbol{z}^{(s)}$ for both $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, swap either the content representation or style representation between different samples, and finally generate new images by decoding the resulting representations, which combine the content representation from one sample with the style representation from another. Bouchacourt et al. (2018) and Mathieu et al. (2016) called this qualitative analysis method "swapping". In the experimental results for image modalities, we present a grid of generated images. Each row corresponds to a fixed content representation $\boldsymbol{z}^{(c)}$, and each column corresponds to a fixed style representation $\boldsymbol{z}^{(s)}$. The image at coordinate (i, j) in the grid is generated using the content latent vector $\boldsymbol{z}_i^{(c)}$ and the style latent vector $\boldsymbol{z}_j^{(s)}$.

Interpolation. We also perform interpolation analysis on the latent representations to evaluate the quality of generated data. We generate sequences of images along the line segments between representations. To investigate disentanglement between *content* factors and *style* factors, we fix one representation and interpolate the other. For instance, when interpolating along the *style*, we generate images along the line segment between $(\boldsymbol{z}_i^{(c)}, \boldsymbol{z}_i^{(s)})$ and $(\boldsymbol{z}_i^{(c)}, \boldsymbol{z}_j^{(s)})$. When interpolating along *content*, we generate images along the line segment between $(\boldsymbol{z}_i^{(c)}, \boldsymbol{z}_i^{(s)})$ and $(\boldsymbol{z}_j^{(c)}, \boldsymbol{z}_i^{(s)})$.

Ablation Study. We perform an ablation study using the Styled-MNIST data to empirically verify the significance of contrastive and anti-contrastive regularization (Appx. A.2). We visualize the latent representations $\boldsymbol{\mu}^{(c)}$ and $\boldsymbol{\mu}^{(s)}$ in 2D space using t-SNE (Van der Maaten & Hinton, 2008).

5.3 Quantitative evaluation

In the classification experiments, the testing sets are all made of unseen combination of content and style. We design this setup to evaluate CLEAR-VAE generalizability. Let $\Omega_c = \{c_1, ..., c_p\}$ and $\Omega_s = \{s_1, ..., s_m\}$ be the set of contents and the set of styles, respectively. For each type of *content*, we only observe k styles in the training set and evaluate a model's discriminative performance using the other $m - k$ styles in the testing set. The training styles and testing styles are randomly assigned. Table 2 demonstrates one possible realization of the experiment setup. This setup is applied to all three datasets.

During the evaluation of a classification model, the testing labels are inaccessible. Therefore, we train an ML-VAE with accumulating evidence but avoid using it when calculating content representation during evaluation. To make fair comparisons between CLEAR-VAE, ML-VAE (without acc. ev.), and baseline models, the VAE encoders share similar architectures with the encoders in their baseline counterparts. Detailed architecture in provided in Appx. A.4. In another words, we replace the regular encoders with encoders trained under the two VAE frameworks. All datasets follow a multi-class classification setup, but we treat them as imbalanced classifications by framing each class as a one-vs-rest problem. Since the precision-recall curve focus more on the minority class, which are positively labeled in the one-vs-rest setting, we use the macro average of the one-vs-rest AUPR scores and top-1 accuracy the metrics for evaluation. Moreover, we are more interested in the relative improvement from the baseline models to CLEAR-VAE encoders.

For the text modality, we insert an MLP-based VAE into Facebook's pretrained Transformer model BART Lewis (2019). BART is the outer layer of the combined architecture, which calculates contextualized token embeddings and finally reconstruct the texts. The MLP VAE module, as the core of the structure, learns disentangled representations from the BART embeddings. In the CLEAR-VAE framework, contrastive regularization terms, $\mathcal{L}_{\text{SNN}}^{(c)}$ and $T(\mathcal{L}_{\text{SNN}}^{(s)})$, are applied to sequence embeddings calculated from `EOS` token's latent representations, while the KL regularization terms in $\mathcal{L}_{\text{VAE}}$ is applied to the entire contextualized embeddings.

	$k=1$		$k=2$		$\cdots$	$k=m-1$	
	train	test	train	test	$\cdots$	train	test
c_1	$\{s_1\}$	$\Omega_s \setminus \{s_1\}$	$\{s_3, s_5\}$	$\Omega_s \setminus \{s_3, s_5\}$	$\cdots$	$\Omega_s \setminus \{s_2\}$	$\{s_2\}$
c_2	$\{s_4\}$	$\Omega_s \setminus \{s_4\}$	$\{s_1, s_2\}$	$\Omega_s \setminus \{s_1, s_2\}$	$\cdots$	$\Omega_s \setminus \{s_1\}$	$\{s_1\}$
$\vdots$					$\cdots$		
c_p	$\{s_1\}$	$\Omega_s \setminus \{s_1\}$	$\{s_3, s_4\}$	$\Omega_s \setminus \{s_3, s_4\}$	$\cdots$	$\Omega_s \setminus \{s_m\}$	$\{s_m\}$

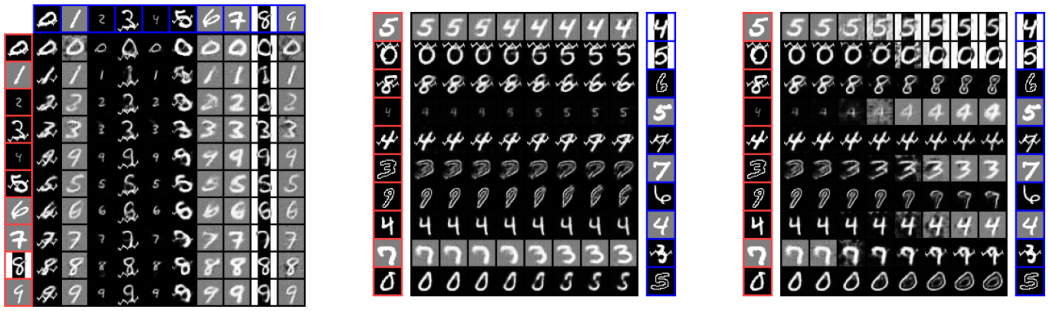
Table 2: Train and Test Results across Different Cycles for k Values

6 Result

6.1 Qualitative Analyses

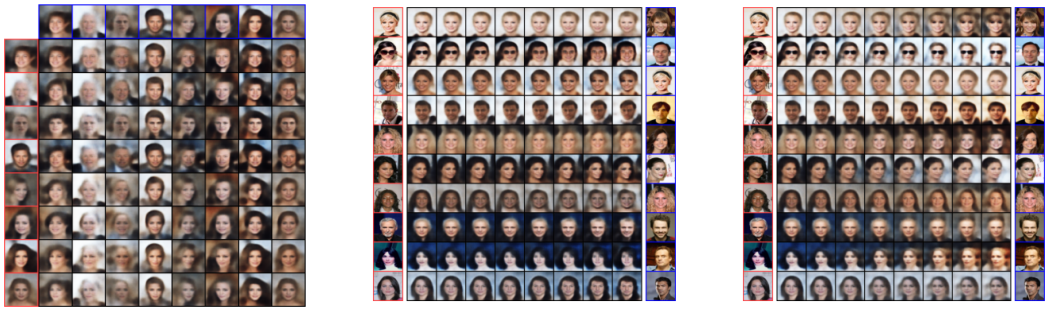
Figure 3 and 4 visualizes the swapping and interpolation experiment results from random samples in Styled-MNIST and CelebA datasets, respectively. In the swapping experiment results, the diagonal

images are the reconstructed samples, and the off-diagonal images are generated from the swapped representation. The identities for *content* is maintained within each row, and the identities for *style* is also consistent within each column. In the interpolation experiment results, either *content* factors or style factors from the source (left) samples are adjusted to match those of the target (right) samples, while keeping the other attribute unchanged.



(a) swapping experiment (b) interpolate along $\boldsymbol{z}^{(c)}$ (c) interpolate along $\boldsymbol{z}^{(s)}$

Figure 3: Swapping and interpolation experiments in Styled-MNIST. In panel (a), the column in red grids and the row in blue grids refer to the same set of test samples. Each row shares the same *content* while each column shares the same *style*. In panel (b), the interpolated images change its content from the source digits to the target digits, while the style is consistent during the interpolation. In panel (c), the interpolated images change its style from source styles to the target styles, while the digits' identities are fixed.



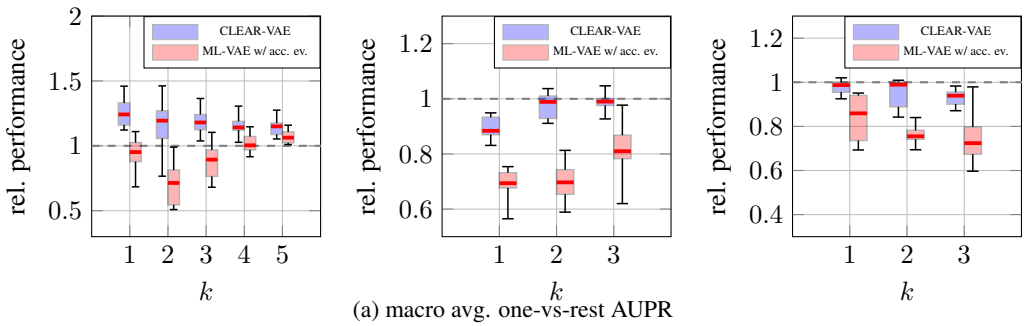
(a) swapping experiment (b) interpolate along $\boldsymbol{z}^{(c)}$ (c) interpolate along $\boldsymbol{z}^{(s)}$

Figure 4: Swapping and interpolation experiments in CelebA. In panel (a), gender and smiling (yes/no) are preserved within each row, and the hair colors are preserved within each column. In panel (b), the identities of the combination of gender and being smiling or not change from the source image to target image, while hair colors are preserved during the interpolation. In panel (c), on the contrary, the hair colors change from the source colors to the target colors, while the identities of gender and being smiling or not are preserved.

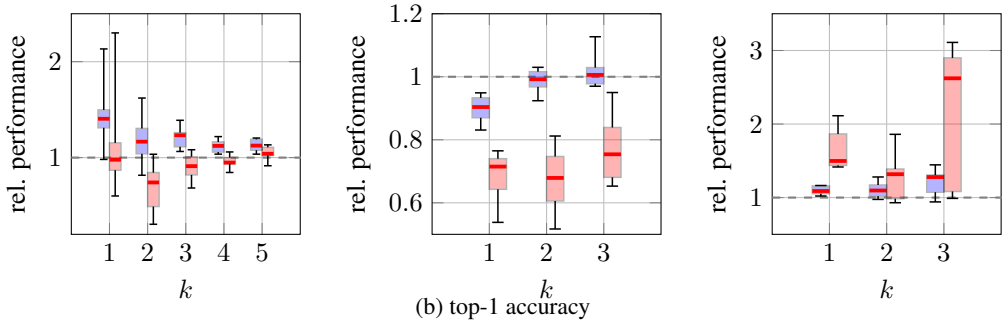
Based on the above analyses, we see that CLEAR-VAE, can learn the semantically disentangled latent representations from the data by solely utilizing contrastive pairs rather than conditioning on the specific ground truth *content* labels. CLEAR-VAE is a weakly supervised framework. It does not involve any labels for style attributes but still successfully extracts style characteristics unknown to the model. During the testing phase, it can perform sample-to-sample conversion through swapping and clustering without any extra conditional information. Moreover, Figure A1 shows that if the *style* feature space is clearly and easily recognizable, the latent representations for *styles* will also form clustering structure.

6.2 DOWNSTREAM IMBALANCED CLASSIFICATION

Since simply taking average and median performance scores can overlook the variability from the random train-test splits, we visualize the models' relative performances using the baseline as a reference. Figure 5 shows that CLEAR-VAE encoders in general can achieve better classification performance on unseen combination of *content* and *style*. It can semantically disentangle $\boldsymbol{z}^{(c)}$ and $\boldsymbol{z}^{(s)}$ from each other on the unseen combinations. As the baseline models' absolute performance increases as more styles are observed during the train time, we use relative performance to evaluate improvement of using disentangled representation. There exits variability in the relative improvements, because the training and testing splits are randomly generated. When k increases, the overlap between samples' style feature space increases, and the benefit of using disentangled representations stabilizes.



(a) macro avg. one-vs-rest AUPR



(b) top-1 accuracy

Figure 5: Relative performance on 10 randomly generated train-test-split for each dataset (column 1: Styled-MNIST; column 2: CelebA; column 3: Amazon Product Review) with k observed training styles. We take baseline CNN and BART performances as the references.

6.3 TRADE-OFFS BETWEEN MIG & ELBO IN CLEAR-VAE & ML-VAE

Let's first compare our CLEAR-VAE to ML-VAE. In our experiment, ML-VAE is trained with *accumulating evidence*. Accumulating group evidence require *content* labels for group-wise reparameterization, but the testing data are completely unlabeled in a classification or our weakly supervised setting. We are not able to using any information form groups in the test time. Thus, we train ML-VAE with accumulating evidence but test it without accumulating evidence.Quantitatively, CLEAR-VAE is consistently better than ML-VAE without test-time accumulating evidence. Qualitatively, CLEAR-VAE is better than ML-VAE even with accumulating evidence (Appx. A.5).

The β-VAE promotes disentanglement between a pair of individual latent variables when $\beta > 1$ (Higgins et al., 2017). If we consider the disentanglement between the blocks of latent variables, the individual latent variable's association will be diluted once we calculate group averages. Moreover, the model is incapable of collecting all content latent information in the designated partition $\boldsymbol{z}^{(c)}$. Therefore, we expect to see poor gMIG between the groups of latent variables, $\boldsymbol{z}^{(s)}$ and $\boldsymbol{z}^{(c)}$.

In contrast, CLEAR-VAE and ML-VAE are capable of achieving disentanglement between two groups of latent variables when $\beta < 1$. There is a competition between the individual-level disentanglement and semantic group-level disentanglement. The disentanglement is limited when we

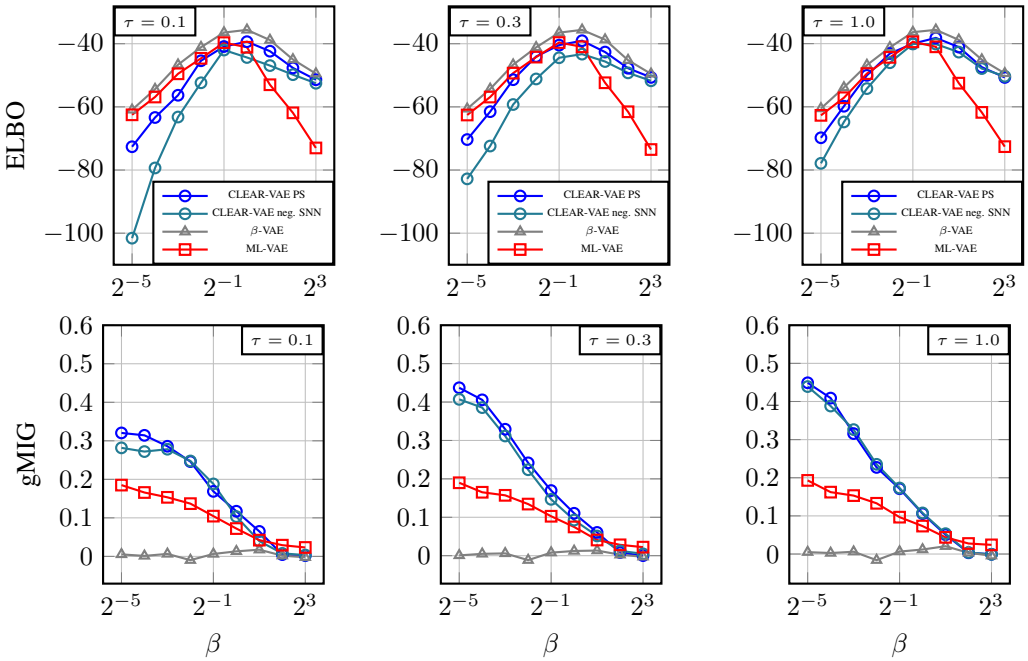


Figure 6: Investigation on the trade-off between ELBO & gMIG in different VAE models using Styled-MNIST as an examplar illustration. The legend is shared across all panels.

add the contrastive regularization terms without specifying β for the KL regularization. Disentanglement at the level of individual dimension has to be small to allow the contrastive regularization terms to be effective. However, the model cannot rely on arbitrarily small values of β in order to maintain a reasonable ELBO. Figure 6 visualizes this trade-off.

When temperature becomes larger, the difference between CLEAR-VAE with PS and CLEAR-VAE with negative SNN decreases. However, with a high temperature the model can be insensitive to important details (e.g. zigzag background vs. identity background). The choice of similarity or distance metrics used in $\mathcal{L}_{\text{SNN}}$ is another important hyperparameter. We empirically find that using cosine similarity $sim(\boldsymbol{\mu}^{(c)}, \boldsymbol{\mu}^{(s)})$ achieves the highest values of ELBO and gMIG compared to other distance metrics.

7 Conclusion

We propose CLEAR-VAE framework for learning semantically disentangled representations from observations share commonality in their content. CLEAR-VAE is weakly supervised in that it only relies on the contrastive pairs generated by the ground truth content labels. The swapping and interpolation experiments provide compelling evidence that CLEAR-VAE can disentangle and recognize the *content* and *style* features from the data and therefore enables controllable data generation. More importantly, this disentangled representation can help to address the issue of not observing all superficial variations during training, which can otherwise undermine the model's generalizability. Via our classification setup, we demonstrate CLEAR-VAE is able to enhance the downstream prediction on unseen combination of contents and styles. For future work, we wish to extend the contrastive learning with anti-contrastive regularization framework to other generative models. We also want to implement this technique to develop more reliable and equitable clinical decision support models that can accurately identify pathological features across diverse patient populations. Our code will be made publicly available.

REFERENCES

Diane Bouchacourt, Ryota Tomioka, and Sebastian Nowozin. Multi-level variational autoencoder: Learning disentangled representations from grouped observations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

Christopher P Burgess, Irina Higgins, Arka Pal, Loic Matthey, Nick Watters, Guillaume Desjardins, and Alexander Lerchner. Understanding disentangling in β-vae. *arXiv preprint arXiv:1804.03599*, 2018.

Ricky TQ Chen, Xuechen Li, Roger B Grosse, and David K Duvenaud. Isolating sources of disentanglement in variational autoencoders. *Advances in neural information processing systems*, 31, 2018.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pp. 1597–1607. PMLR, 2020.

Sumit Chopra, Raia Hadsell, and Yann LeCun. Learning a similarity metric discriminatively, with application to face verification. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pp. 539–546. IEEE, 2005.

Sebastian Damrich, Jan Niklas Böhm, Fred A Hamprecht, and Dmitry Kobak. From t-sne to umap with contrastive learning. *arXiv preprint arXiv:2206.01816*, 2022.

Nicholas Frosst, Nicolas Papernot, and Geoffrey Hinton. Analyzing and improving representations with the soft nearest neighbor loss. In *International conference on machine learning*, pp. 2012–2020. PMLR, 2019.

Ryuhei Hamaguchi, Ken Sakurada, and Ryosuke Nakamura. Rare event detection using disentangled representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9327–9335, 2019.

Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9729–9738, 2020.

Irina Higgins, Loic Matthey, Arka Pal, Christopher P Burgess, Xavier Glorot, Matthew M Botvinick, Shakir Mohamed, and Alexander Lerchner. beta-vae: Learning basic visual concepts with a constrained variational framework. *ICLR (Poster)*, 3, 2017.

Yupeng Hou, Jiacheng Li, Zhankui He, An Yan, Xiusi Chen, and Julian McAuley. Bridging language and items for retrieval and recommendation. *arXiv preprint arXiv:2403.03952*, 2024.

Hyunjik Kim and Andriy Mnih. Disentangling by factorising. In *International conference on machine learning*, pp. 2649–2658. PMLR, 2018.

Diederik P Kingma. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Gregory Koch, Richard Zemel, Ruslan Salakhutdinov, et al. Siamese neural networks for one-shot image recognition. In *ICML deep learning workshop*, volume 2, pp. 1–30. Lille, 2015.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

M Lewis. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461*, 2019.

Shikun Li, Xiaobo Xia, Shiming Ge, and Tongliang Liu. Selective-supervised contrastive learning with noisy labels. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 316–325, 2022.

Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of International Conference on Computer Vision (ICCV)*, December 2015.

Michael F Mathieu, Junbo Jake Zhao, Junbo Zhao, Aditya Ramesh, Pablo Sprechmann, and Yann LeCun. Disentangling factors of variation in deep representation using adversarial training. *Advances in neural information processing systems*, 29, 2016.

Norman Mu and Justin Gilmer. Mnist-c: A robustness benchmark for computer vision. *arXiv preprint arXiv:1906.02337*, 2019.

Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

Xuanchi Ren, Tao Yang, Yuwang Wang, and Wenjun Zeng. Learning disentangled representation by exploiting pretrained generative models: A contrastive learning view. *arXiv preprint arXiv:2102.10543*, 2021.

Elan Rosenfeld, Ezra Winston, Pradeep Ravikumar, and Zico Kolter. Certified robustness to label-flipping attacks via randomized smoothing. In *International Conference on Machine Learning*, pp. 8230–8241. PMLR, 2020.

Ruslan Salakhutdinov and Geoff Hinton. Learning a nonlinear embedding by preserving class neighbourhood structure. In *Artificial intelligence and statistics*, pp. 412–419. PMLR, 2007.

Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

Han Xiao, Huang Xiao, and Claudia Eckert. Adversarial label flips attack on support vector machines. In *ECAI 2012*, pp. 870–875. IOS Press, 2012.

A APPENDIX

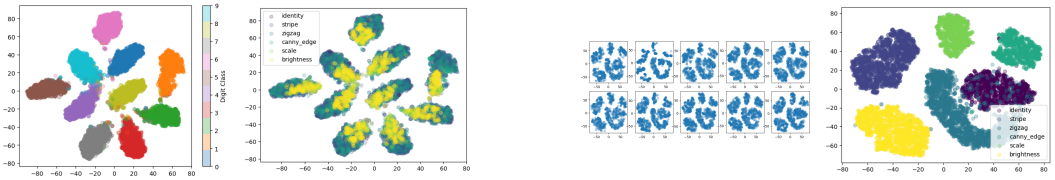
A.1 KL DECOMPOSITION

Assume $\boldsymbol{z}^{(c)}$ and $\boldsymbol{z}^{(s)}$ are independent *a priori* and *a posteriori*, the KL regularization term for VAE can be decomposed as follows.

$$
\begin{aligned}
D_{\mathrm{KL}}(q_\phi(\boldsymbol{z}|\boldsymbol{x})\|p(\boldsymbol{z})) &= \int \log \frac{q_\phi(\boldsymbol{z}^{(c)}, \boldsymbol{z}^{(s)}|\boldsymbol{x})}{p(\boldsymbol{z}^{(c)}, \boldsymbol{z}^{(s)})} q_\phi(\boldsymbol{z}^{(c)}, \boldsymbol{z}^{(s)}|\boldsymbol{x}) d\boldsymbol{z} \\
&= \iint \left(\log \frac{q_\phi(\boldsymbol{z}^{(c)}|\boldsymbol{x})}{p(\boldsymbol{z}^{(c)})} + \log \frac{q_\phi(\boldsymbol{z}^{(s)}|\boldsymbol{x})}{p(\boldsymbol{z}^{(s)})} \right) q_\phi(\boldsymbol{z}^{(c)}|\boldsymbol{x}) q_\phi(\boldsymbol{z}^{(s)}|\boldsymbol{x}) d\boldsymbol{z}^{(c)} d\boldsymbol{z}^{(s)} \\
&= D_{\mathrm{KL}} \left(q_\phi(\boldsymbol{z}^{(c)}|\boldsymbol{x}) \| p(\boldsymbol{z}^{(c)}) \right) + D_{\mathrm{KL}} \left(q_\phi(\boldsymbol{z}^{(s)}|\boldsymbol{x}) \| p(\boldsymbol{z}^{(s)}) \right)
\end{aligned}
$$

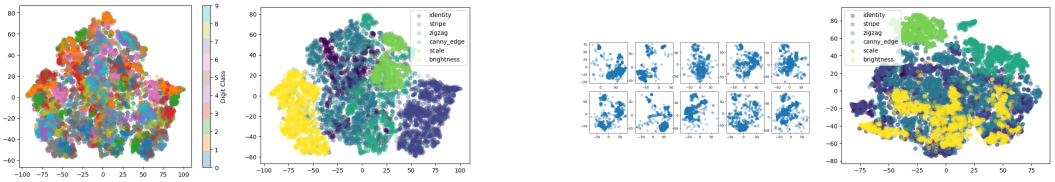
A.2 ABLATION STUDY

Here, we use the Styled-MNIST for as an exemplar illustration. Figure A1 visualizes clear clustering patterns in the 2D t-SNE space. Panel (a) shows the clustering of $\boldsymbol{\mu}^{(c)}$ colored by *content* and *style*, respectively. Panel (b) shows the $\boldsymbol{\mu}^{(s)}$ distributions stratified by *content* and the clustering of $\boldsymbol{\mu}^{(s)}$ colored by *style*. Without the contrastive and the anti-contrastive regularization, the representations of *content* and *style* are entangled in the latent space. When we remove both $\mathcal{L}^{(c)}_{\text{SNN}}$ and $T(\mathcal{L}^{(s)}_{\text{SNN}})$, a CLEAR-VAE will be reduced to a β-VAE. Figure A2 indicates that the model fails to disentangle *content* and *style*. If we remove $T(\mathcal{L}^{(s)}_{\text{SNN}})$, the CLEAR-VAE is still able to learn the representation of content and disentangle it from style to some extent. Figure A3 shows that it is sub-optimal and incapable of extracting details (e.g. zig-zag lines) in style features.



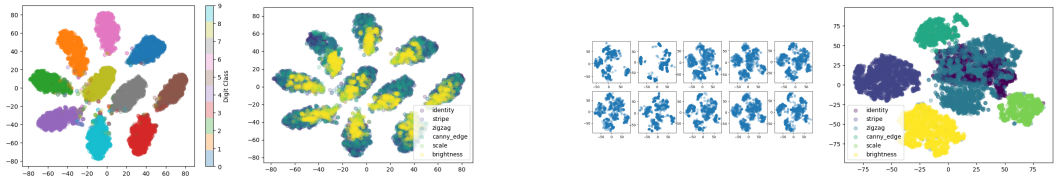
(a) $\boldsymbol{\mu}^{(c)}$ (b) $\boldsymbol{\mu}^{(s)}$

Figure A1: CLEAR-VAE-PS t-SNE visualization for test data *content* and *style* latent representations in styled-MNIST. The left figure of panel (b) stratifies $\boldsymbol{\mu}^{(s)}$ according to the digit, with the stratified plots arranged sequentially and in rows. Cosine similarity is used in $\mathcal{L}_{\text{SNN}}$. The hyperparameter configuration for the training objective function is: $d_{\boldsymbol{z}} = 16, \tau = 0.3, \beta = 1/8, \alpha = 100$, which achieves gMIG $= 0.305$, ELBO $= -54.2$



(a) $\boldsymbol{\mu}^{c}$ (b) $\boldsymbol{\mu}^{s}$

Figure A2: β-VAE t-SNE visualization for *content* and *style* latent representations in styled-MNIST. The hyperparameter configuration for the training objective function is: $d_{\boldsymbol{z}} = 16, \beta = 1/8$, which achieves gMIG $= 0$, ELBO $= -46.1$



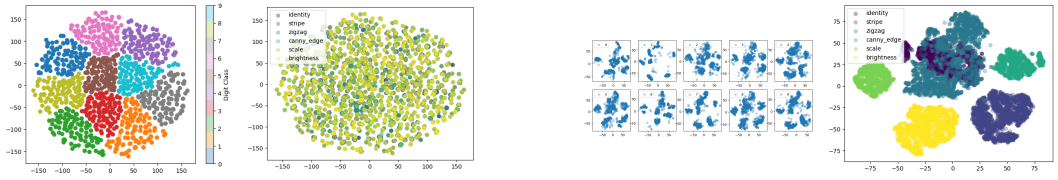
(a) $\boldsymbol{\mu}^c$ (b) $\boldsymbol{\mu}^s$

Figure A3: CLEAR-VAE (without regularizing $\mu^{(s)}$) t-SNE visualization for *content* and *style* latent representations in styled-MNIST. The hyperparameter configuration for the training objective function is: $d_{\boldsymbol{z}} = 16, \beta = 1/8, \alpha = 100$, which achieves gMIG $= 0.265$, ELBO $= -49.1$.

A.3 ML-VAE T-SNE PLOTS ON STYLED-MNIST

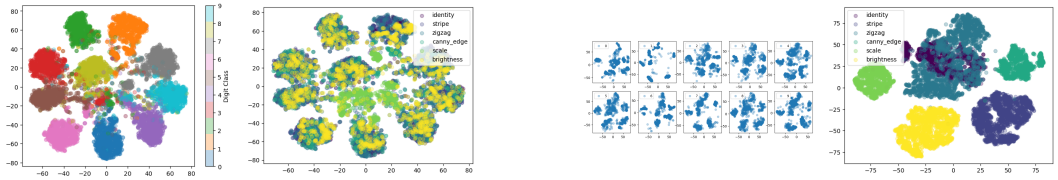
ML-VAE learn *content* representation from grouped data. Specifically, within each mini-batch, it organizes inputs into groups using the ground truth labels. ML-VAE with test-time accumulating evidence achieves better disentanglement, but $\boldsymbol{\mu}^{(c)}$'s have degenerated distributions. All *content* representations are mapped to a finite number of point masses in the latent space. Similar to CLEAR-VAE with the anti-contrastive regularization, both versions are incapable of differentiating zigzag-styled and identity-styled digits.

Test-time accumulating evidence will lead to degenerated $\boldsymbol{\mu}^{(c)}$ (Appx. A.3). The calculation is influenced by the test content label and thus leaks test content label's information to $\boldsymbol{z}^{(c)}$. Consequently, ML-VAE with test-time accumulated evidence is not appropriate for the quantitative evaluation in this setting. However, CLEAR-VAE and ML-VAE without test-time accumulating evidence do not rely on *content* label during test time.



(a) $\boldsymbol{\mu}^c$ (b) $\boldsymbol{\mu}^s$

Figure A4: ML-VAE with test-time accumulating evidence



(a) $\boldsymbol{\mu}^c$ (b) $\boldsymbol{\mu}^s$

Figure A5: ML-VAE without test-time accumulating evidence

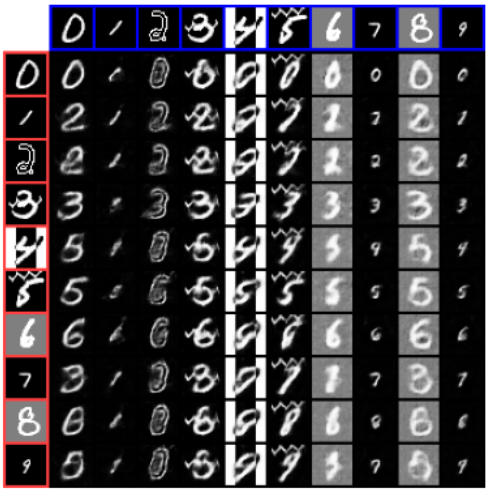
A.4 MODEL ARCHITECTURE

Image modalities. Encoder modules (including CLEAR-VAE encoder, ML-VAE encoder, and CNN feature extraction head) have a common structure: sequential blocks of Conv2D-BN2D-ReLU. Decoder modules (including CLEAR-VAE decoder and ML-VAE decoder) have the same structure: sequential blocks of ConvTranspose2D-BN2D-ReLU. For classification tasks, all methods employ 2-layer MLP classification heads with identical shapes.

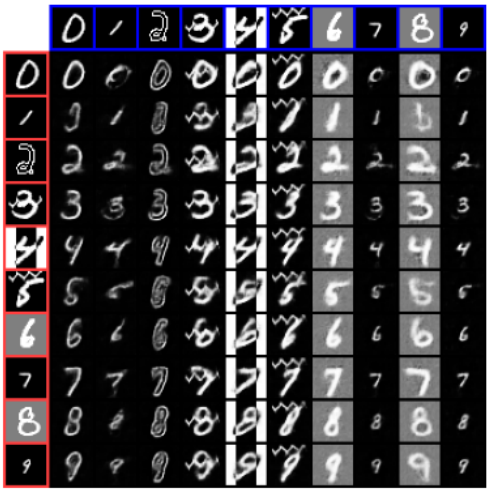
Text modality. Encoder modules in CLEAR-VAE and ML-VAE have a common combined structure: the BART encoder and a 2-layer MLP VAE encoder. Decoder modules (CLEAR-VAE encoder and ML-VAE encoder) have the same combined structure: a 2-layer MLP VAE decoder and the BART decoder. In classification experiments, all methods use 2-layer MLP classification heads in the same shape.

A.5 Image comparison between ML-VAE & CLEAR-VAE

Here we perform the swapping experiment for both ML-VAE (with acc. ev.) and CLEAR-VAE. Ideally, the result will have fixed content with varied styles in each row and a consistent style of varied content in each column. The results from CLEAR-VAE are comparable in quality to those from ML-VAE. Moreover, due to the nature of the method, CLEAR-VAE allows the content representations to have more variability.



(a) swapping experiment for ML-VAE



(b) swapping experiment for CLEAR-VAE

Figure A6: Swapping experiment performed on ML-VAE and CLEAR-VAE.

(a) swapping experiment for ML-VAE

(b) swapping experiment for CLEAR-VAE

Figure A7: Swapping experiment performed on ML-VAE and CLEAR-VAE.

A.6 INTERPOLATED EXAMPLES FOR AMAZON REVIEWS

Interpolating an `all beauty` product review from rating of 1 to rating of 5

Original Reviews:

- **(Source)** As others have posted, this product is not as advertised. I ordered twice, thinking I must have made a mistake the first time (I had just had my baby, so things were a little crazy right then). Nope. What I received was 17 packs of 64 wipes, when what I ordered was a pack of 3 larger refill packs and a dispenser container.
- **(Target)** I seriously love this product. I am always shopping for the perfect skincare items and this is one I will keep around. It makes my skin super soft and gives it a youthful glow. I have sensitive skin and it is not irritating at all. I would recommend wearing sunscreen after you use it because it can make your skin more sensitive to the sun.

Interpolated Sentences:

- As others have said, this product is not as advertised. I ordered twice, thinking I must have made a mistake the first time (I had just had my baby, so things were a little crazy right then). Nope. What I received was a pack of 6-pack of six-packs of six packs of six pack of sixpack of 6 packs of 6 pack of 12-pack with six-pack and six pack with six pack each. What the hell? What the heck?
- As far as I'm concerned, this product is not FDA-approved. I ordered this product from the FDA. I'm not sure why I ordered it from this product. I order it from Amazon.com. I thought it was for me. I had to order from this site. I have to get it from another company. I think it's because I ordered from this company. It's not FDA approved.
- As I said, I don't have a lot to say about this product. I just have to say that it's a lot of fun to me. I've never tried this product before. I have never tried it before. It's so easy to say no to it. I tried it once. I'm not sure if it's even fun to try it out. I really like it.I'm not a big fan of it.
- I'm not sure if this is a good idea or not, but it's a great idea to try it out and see if it works for you. I'm sure it will work for me. I've tried it out for myself. I love it. I think it's great. I really like it. It's awesome. I like it! I really love it! It's great! I'm so happy with it.I'm glad it's working for me too. I don't like it at all. I hate it when it's not working. I want it to be perfect. I mean, it's
- I have been using this product for over a year now. It's a great product. I love it. I use it every day. I really like it.I love it! I use this product.I use it everyday. I always use it when I'm sick. I need it to be perfect. I will use it for everything. I do not need it for anything. I don't need it at all. I just need it.

Interpolating a product review from `all beauty` and `health` while maintaining 5 for rating

Original Reviews:

- **(Source)** I love this mask. I can easily adjust it to keep my ears from getting sore. The integrated neck strap makes it easy to take on and off if you are going in and out of places where it is needed. The filter pocket is there, but pretty useless, so not a feature that makes much difference. Overall, this is a great cloth mask, adjustable, soft, comfortable, and convenient.
- **(Target)** The price and lack of a more complete description worried me when I purchased this. But I was happily surprised to find that it is quite a nice little bundle of incense for the price I paid. They are larger sticks and have a very woody smell with none of that annoying perfume-y smell. Not the smokiest stick I have ever burnt, which is good as well. It's a decent buy if you like cheap woodsy-smelling incense.

Interpolated Sentences:

- I love this mask. I can easily adjust it to keep my ears from getting sore. The integrated earbuds make it easy to take on and off if you are going in and out of places where it is needed. The filter is there, but pretty useless, so not a feature that makes much difference. Overall, this is a great choice for a new earbud. It's a great looking earphone earphone, so it's easy to use this earphone. This earphone is also great. This is great.
- I love this product. I love it. I can easily adjust it to my body temperature. I like it to be comfortable to wear it on my head. I also love it to take on and off if you are going to take it off if it is going to be hot or cold. It is comfortable to take off. The heat is there, but not too hot, so not a great deal that makes much difference. Overall, this is a great product. This product is great. This is great to wear on your head.This product is awesome. I really like it. It's great to use it on your
- I love this product. I love it. I like it so much. I can't wait to try it out on my own. I also love it to have it on my bed. It's so comfortable to wear it on and off if you are going to use it on someone else's bed. I think it is so comfortable. It is so easy to use. The only problem is that it is there is nothing on your bed. The bed is there. The mattress is there, too. It's a great place to sleep. I'm not sure if it'll be comfortable on my
- I love this product. I love it. I like it so much. I can't wait to try it out on my own. I also love it to be able to wear it on my face. I think it's great to have it on your face. It's so comfortable to wear on your skin. It is so comfortable. It's so soft to wear, so it is so natural to use it on you.It is so easy to use. I am so comfortable with it.I like it. This is my first time wearing it. It will be my last. I will be wearing it on
- I love this product. I love it. I like it so much. I can't wait to try it out on my own. I also love it for my kids. I have it for myself. It's great to have it on my bed. I think it's great for them. I really like it. It is great for me. It's great for my family. It makes me happy. It also makes me feel good. is great. is is great! is great is great, is great with kids.