Capturing Label Characteristics in VAEs

ICLR 2021

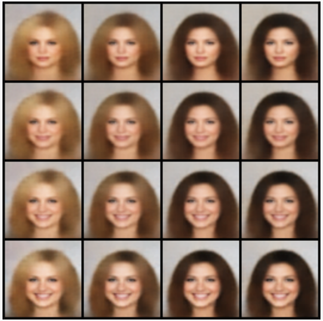
Tom Joy[1], Sebastian M. Schmon[1,2], Philip H. S. Torr[1], N. Siddharth⋆ [1,3] and Tom Rainforth⋆ [1]

March 25, 2021

[1]University of Oxford
[2]Improbable
[3]University of Edinburgh & The Alan Turing Institute

Representation Learning through VAEs

- VAEs learn representations of data.
- We want to capture characteristics in individual dimensions.
- Can alter characteristics without affecting others.
- Our model achieves this by leveraging label information.



Introducing label information

- Sometimes we have labels.
- A label contains the minimum amount of information about the characteristic.

Not Smiling

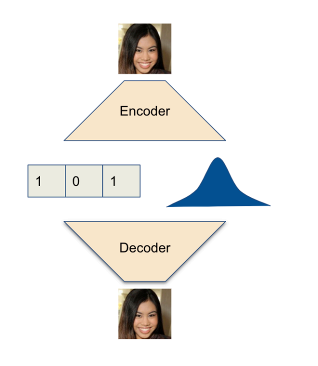
Smiling

Baseline

- Prior approaches treat the labels directly as latent values[1].
- Latents are partially-observed.

This is a bad idea

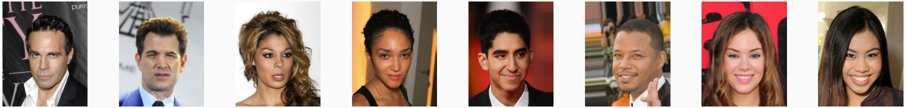
Placing the labels into the latent space destroys the representation.



[1] Diederik P. Kingma et al. "Semi-supervised learning with deep generative models". In: *NIPS* (2014).

How should we use labels?

- We want to capture more information than just the label itself.
- Traversing the latents should alter the characteristics.
- Latent space smooth needs to be smooth.



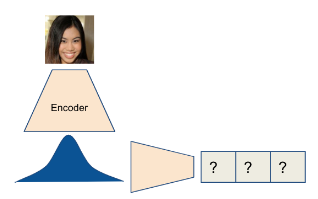
Less Smiling —— More Smiling

CCVAE - Inference model

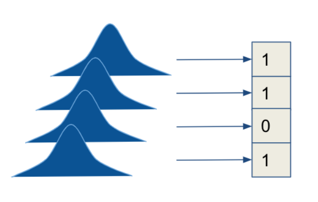
- Our model has an entirely smooth latent space.
- The labels are treated as auxiliary variables.
- Classification is performed on the latent space.

This is a good idea

The latent space now captures the information in the image.



- We classify a single label on a single latent.
- The latent space now has a latent for each label.



This is a good idea

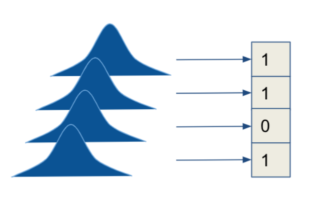
This is essential for disentanglement.

CCVAE - Disentangling the latent space

- Information about the characteristic needs to be present to classify it correctly.
- This forces the encoder to place the information in the correct latent.

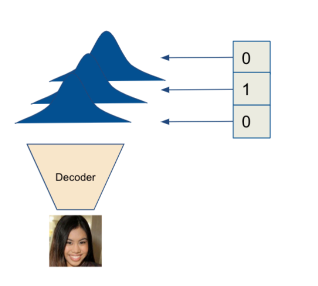
Disentanglement!

We can now capture more information than the labels and also disentangle them!



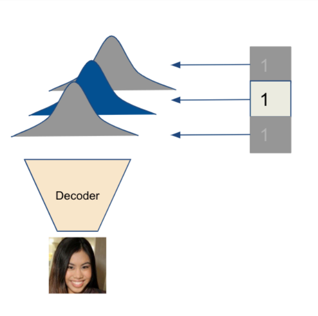
CCVAE - Generative model

- Each characteristic will have diversity.
- We want this diversity in conditional generations.
- We learn a distribution $p(z_i | y_i)$ for each latent.



CCVAE - Generative model

- We can now sample individual representations for individual characteristics.
- We can re-sample from one representation to gain diversity of that characteristic.

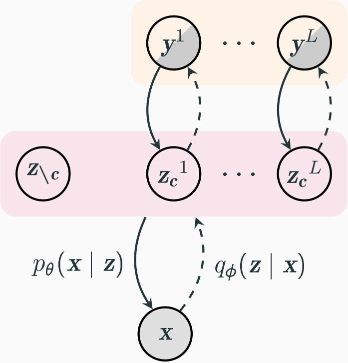


Diverse Generations

We can now generate diverse samples for each label, e.g. preserve everything but generate different glasses.

CCVAE - Graphical model

- We use the previous features in our graphical model.
- This enables us to capture more information than just the label.
- We also disentangle the latent space, allowing for traversals.



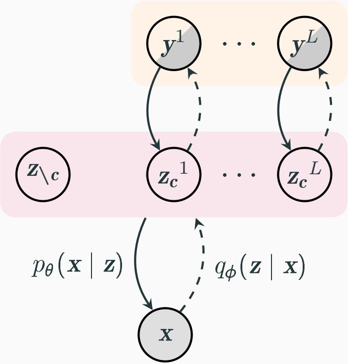
CCVAE - Supervised Objective

- Our objective represents the ELBO of the model.
- which contains a conditional prior $p_\psi(\boldsymbol{z} \mid \boldsymbol{y})$ and a classifier $q_{\varphi,\phi}(\boldsymbol{y} \mid \boldsymbol{x})$.

$$\mathcal{L}_{CCVAE}(\boldsymbol{x}, \boldsymbol{y}) = \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\left[\frac{q_\varphi(\boldsymbol{y} \mid \boldsymbol{z})}{q_{\varphi,\phi}(\boldsymbol{y} \mid \boldsymbol{x})} \log \frac{p_\theta(\boldsymbol{x} \mid \boldsymbol{z}) p_\psi(\boldsymbol{z} \mid \boldsymbol{y})}{q_\varphi(\boldsymbol{y} \mid \boldsymbol{z}) q_\phi(\boldsymbol{z} \mid \boldsymbol{x})}\right] + \log q_{\varphi,\phi}(\boldsymbol{y} \mid \boldsymbol{x}) + \log p(\boldsymbol{y}). \tag{1}$$

CCVAE - Semi-supervised Objective

- Our model can deal with the case without labels.
- Perform variational inference over the labels.
- Good performance is achieved at a supervision rate of 1/250.



CCVAE - Results

- Our model captures and disentangles characteristics.
- Permitting latent traversals and diverse conditional generations.
- Improves classification accuracy.

	CelebA			
Model	$f = 0.004$	$f = 0.06$	$f = 0.2$	$f = 1.0$
CCVAE	**0.832**	**0.862**	**0.878**	**0.900**
M2	0.794	0.862	0.877	0.893
DIVA	0.807	0.860	0.867	0.877
MVAE	0.793	0.828	0.847	0.864

Figure 1: Classification accuracies

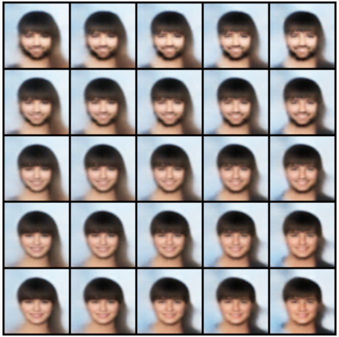


Figure 2: Latent Traversals

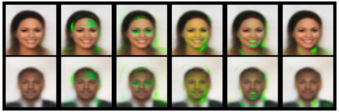


Figure 3: Variance of generations

Thank You

github.com/thwjoy/ccvae

Paper

Code