VAE variants

Hao Dong

Peking University

PEKING UNIVERSITY 北京大學

VAE variants

- Convolutional VAE
- Conditional VAE
- β-VAE
- IWAE
- Ladder VAE
- Progressive + Fade-in VAE
- VAE in speech
- Temporal Difference VAE (TD-VAE)

Representation learning

Hierarchical representation learning

Temporal representation learning

VAE variants

北京大學 PEKING UNIVERSITY

- Convolutional VAE
- Conditional VAE

Representation learning
- IWAE
- β-VAE

Hierarchical representation learning
- Ladder VAE
- Progressive + Fade-in VAE

Temporal representation learning
- VAE in speech
- Temporal Difference VAE (TD-VAE)

Convolutional Variational Autoencoder

- **Limitations of vanilla VAE**
  - The size of weight of fully connected layer == input size x output size
  - If VAE uses fully connected layers only, will lead to curse of dimensionality when the input dimension is large (e.g., image).

- **Solution**

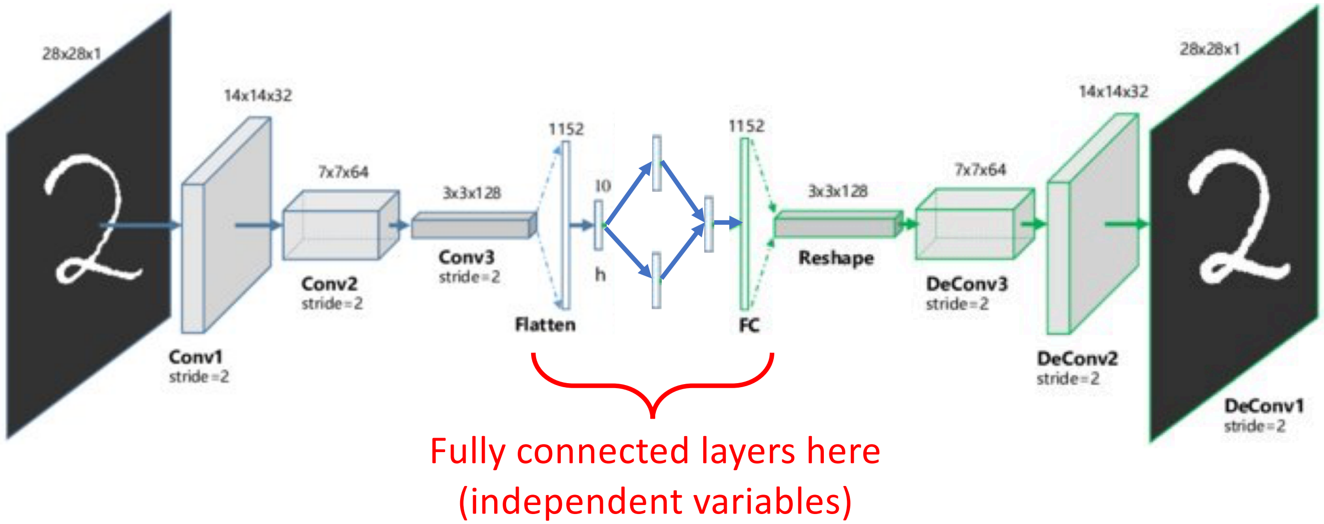


Image is modified from: Deep Clustering with Convolutional Autoencoder. NIPS 2017.

VAE variants

- Convolutional VAE
- Conditional VAE

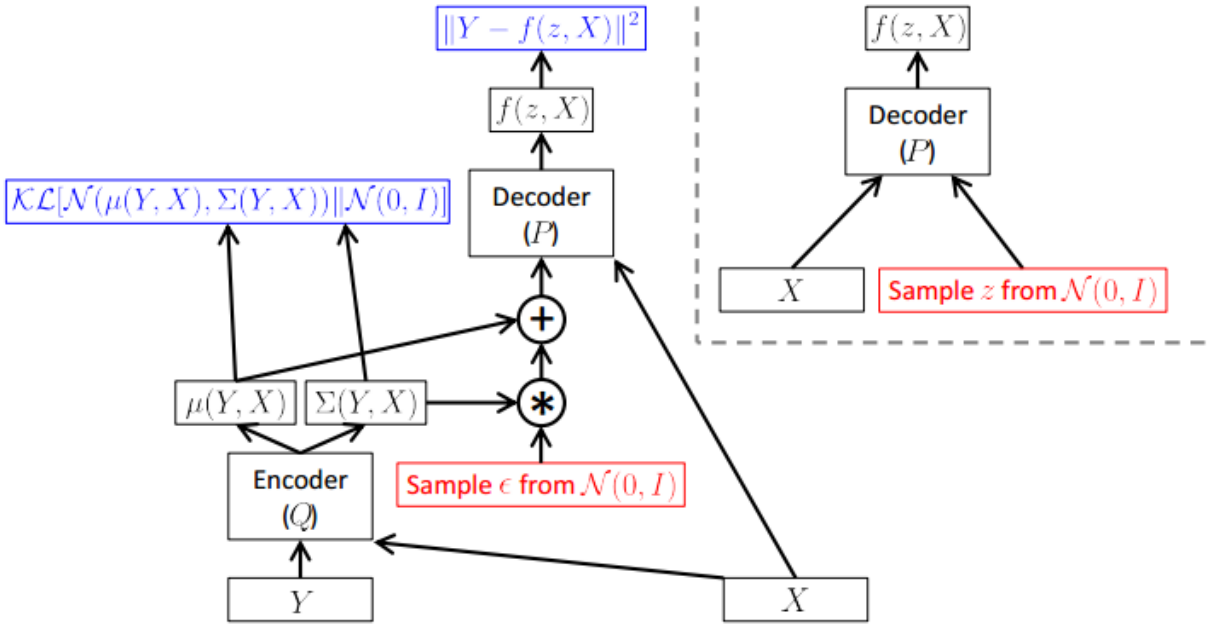
Representation learning
- β-VAE
- IWAE

Hierarchical representation learning
- Ladder VAE
- Progressive + Fade-in VAE

Temporal representation learning
- VAE in speech
- Temporal Difference VAE (TD-VAE)

Conditional Variational Autoencoder

- **Train and inference with labeled data.**



Learning structured output representation using deep conditional generative models. NeurIPS 2015.

Recap: Variational Autoencoder

- **Recap: Setting up the objective**
  - Maximize P(X)
  - Set Q(z) to be an arbitrary distribution

$$\mathcal{D}\left[Q(z) \| P(z|X)\right] = E_{z\sim Q}\left[\log Q(z) - \log P(z|X)\right]$$

$$\mathcal{D}\left[Q(z) \| P(z|X)\right] = E_{z\sim Q}\left[\log Q(z) - \log P(X|z) - \log P(z)\right] + \log P(X)$$

$$\log P(X) - \mathcal{D}\left[Q(z) \| P(z|X)\right] = E_{z\sim Q}\left[\log P(X|z)\right] - \mathcal{D}\left[Q(z) \| P(z)\right]$$

$$\boxed{\log P(X) - \mathcal{D}\left[Q(z|X) \| P(z|X)\right] = E_{z\sim Q}\left[\log P(X|z)\right] - \mathcal{D}\left[Q(z|X) \| P(z)\right]}$$

Auto-Encoding Variational Bayes. Diederik P. Kingma, Max Welling. ICLR 2013

Recap: Variational Autoencoder

PEKING UNIVERSITY

- **Recap: Setting up the objective**

$$\log P(X) - \mathcal{D}\left[\underbrace{Q(z|X)}_{\text{encoder}} \| \underbrace{P(z|X)}_{\text{ideal}}\right] = \boxed{E_{z\sim Q}\left[\underbrace{\log P(X|z)}_{\text{reconstruction}}\right] - \underbrace{\mathcal{D}\left[Q(z|X)\|P(z)\right]}_{\text{KLD}}}$$

Auto-Encoding Variational Bayes. Diederik P. Kingma, Max Welling. ICLR 2013

Conditional Variational Autoencoder

- **Setting up the objective with labels**
  - Maximize P(Y|X)
  - Set Q(z) to be an arbitrary distribution

$$\mathcal{D}\left[Q(z|Y,X)\|P(z|Y,X)\right] = E_{z\sim Q(\cdot|Y,X)}\left[\log Q(z|Y,X) - \log P(z|Y,X)\right]$$

$$\mathcal{D}\left[Q(z|Y,X)\|P(z|Y,X)\right] = \\ E_{z\sim Q(\cdot|Y,X)}\left[\log Q(z|Y,X) - \log P(Y|z,X) - \log P(z|X)\right] + \log P(Y|X)$$

$$\mathcal{D}\left[Q(z|Y,X)\|P(z|Y,X)\right] = \\ E_{z\sim Q(\cdot|Y,X)}\left[\log Q(z|Y,X) - \log P(Y|z,X) - \log P(z|X)\right] + \log P(Y|X)$$

$$\boxed{\log P(Y|X) - \mathcal{D}\left[Q(z|Y,X)\|P(z|Y,X)\right] = \\ E_{z\sim Q(\cdot|Y,X)}\left[\log P(Y|z,X)\right] - \mathcal{D}\left[Q(z|Y,X)\|P(z|X)\right]}$$

Learning structured output representation using deep conditional generative models. NeurIPS 2015.

Conditional Variational Autoencoder

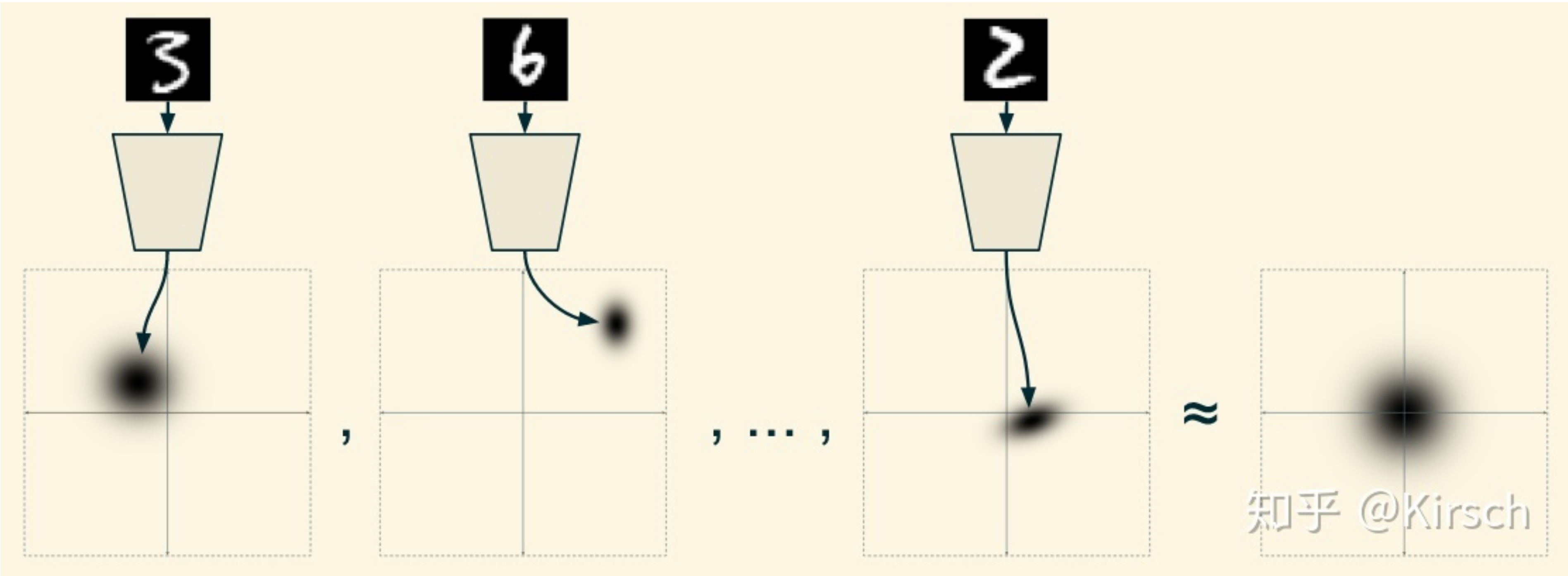
- **Setting up the objective**

$$\log P(Y|X) - \mathcal{D}\left[\overset{\text{encoder}}{Q(z|Y,X)} \,\|\, \overset{\text{ideal}}{P(z|Y,X)}\right] = \boxed{\underbrace{E_{z\sim Q(\cdot|Y,X)}\left[\log P(Y|z,X)\right]}_{\text{reconstruction}} - \underbrace{\mathcal{D}\left[Q(z|Y,X)\,\|\,P(z|X)\right]}_{\text{KLD}}}$$

Learning structured output representation using deep conditional generative models. NeurIPS 2015.

Conditional Variational Autoencoder

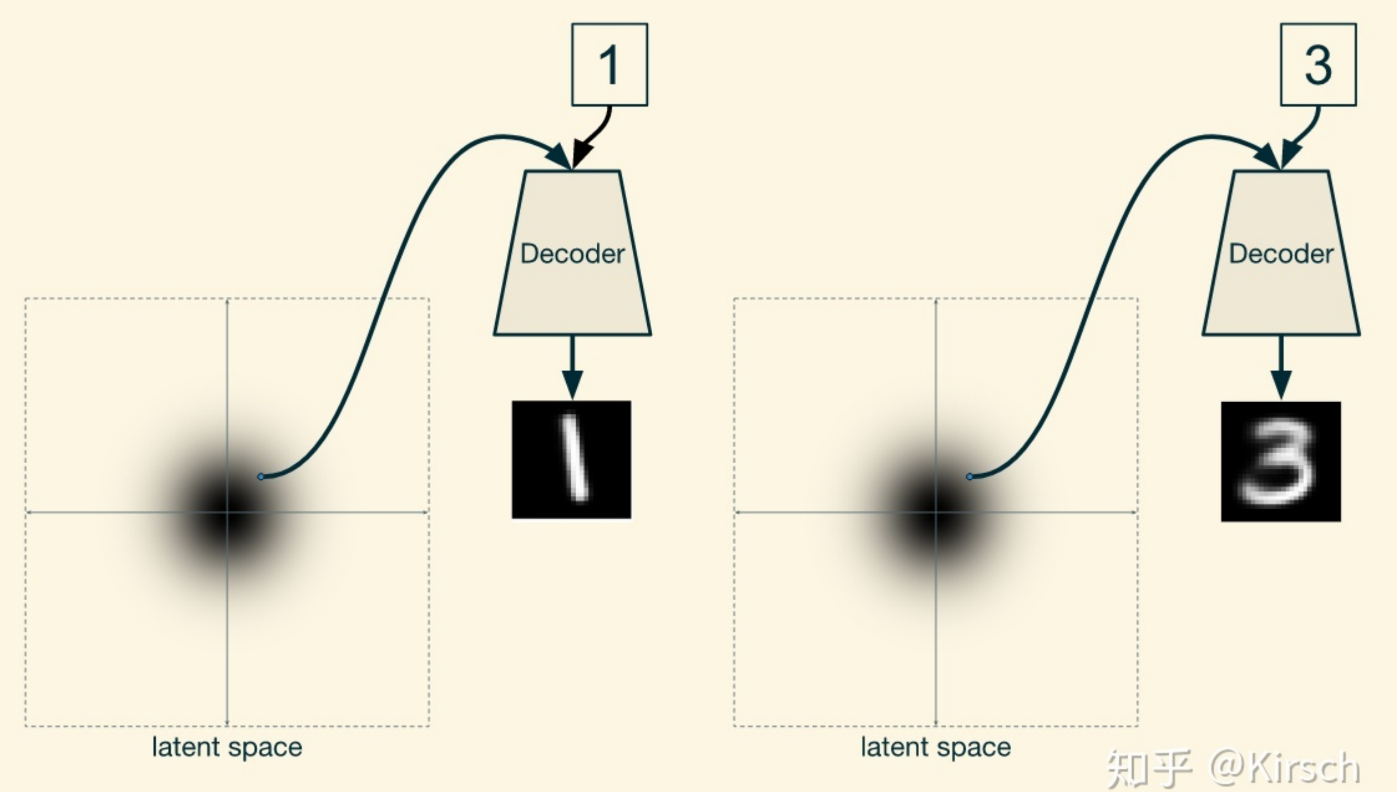
- **Train and inference without labeled data i.e., vanilla VAE**



Learning structured output representation using deep conditional generative models. NeurIPS 2015.

Conditional Variational Autoencoder

- **Train and inference with labeled data.**

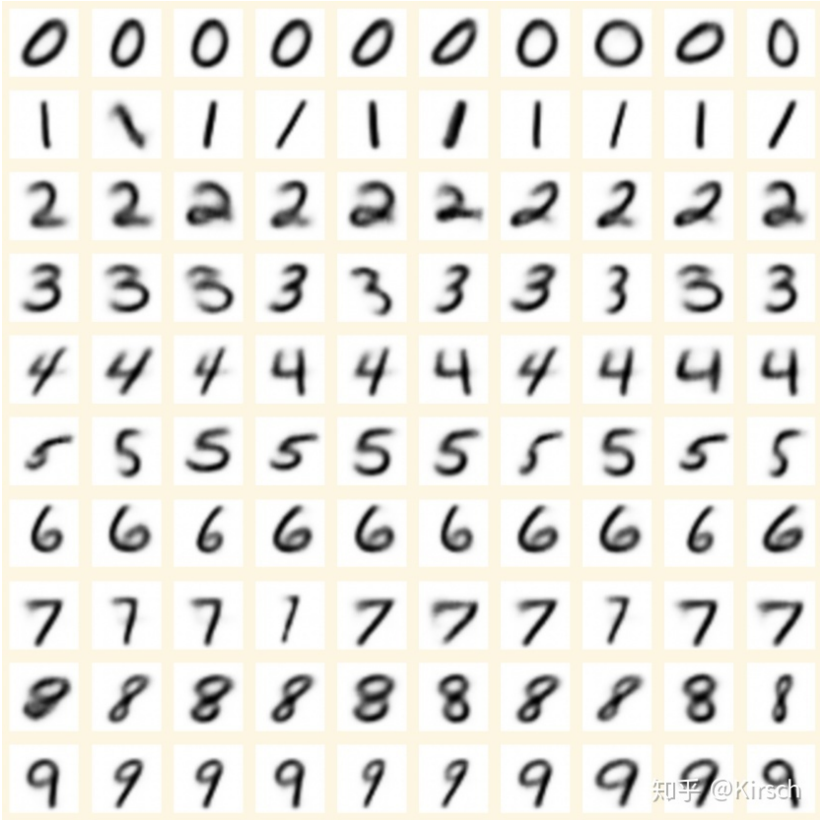


A variational autoencoder generating images according to given labels

Learning structured output representation using deep conditional generative models. NeurIPS 2015.

Conditional Variational Autoencoder

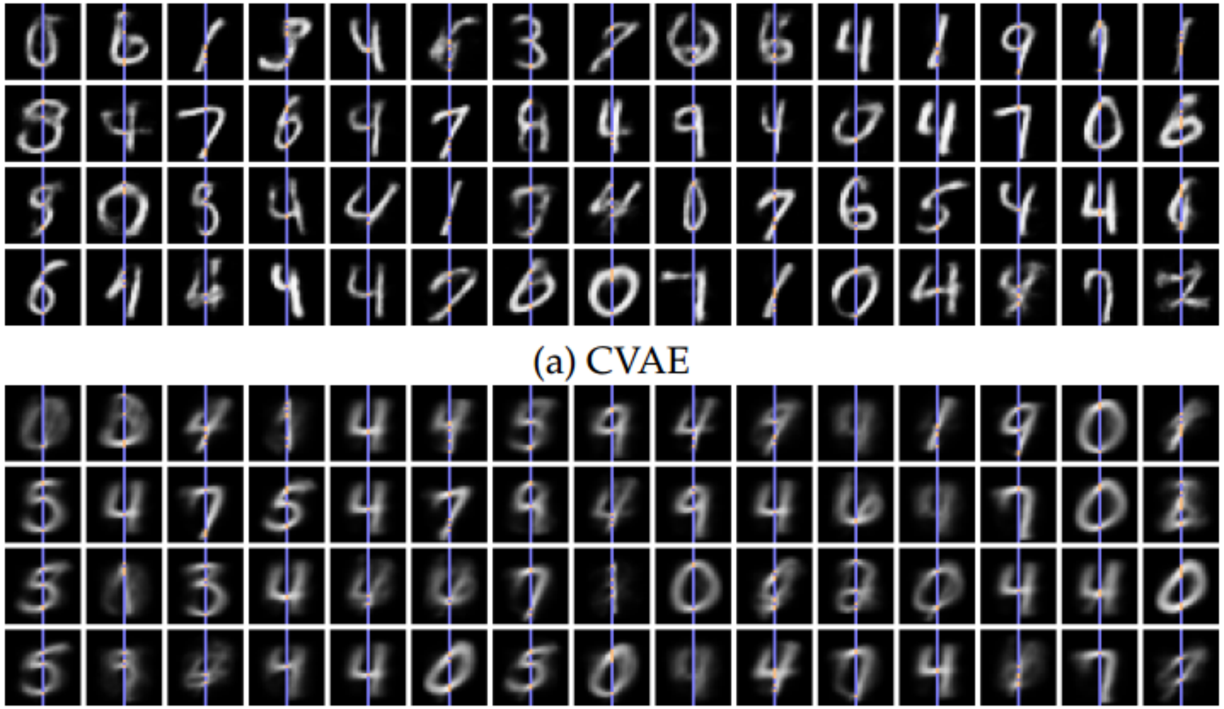
- **Train and inference with labeled data.**



Learning structured output representation using deep conditional generative models. NeurIPS 2015.

Conditional Variational Autoencoder

- **Train and inference with labeled data.**



(a) CVAE

(b) Regressor

Learning structured output representation using deep conditional generative models. NeurIPS 2015.

VAE variants

Representation learning

Hierarchical representation learning

Temporal representation learning

- Convolutional VAE
- Conditional VAE
- β-VAE
- IWAE
- Ladder VAE
- Progressive + Fade-in VAE
- VAE in speech
- Temporal Difference VAE (TD-VAE)

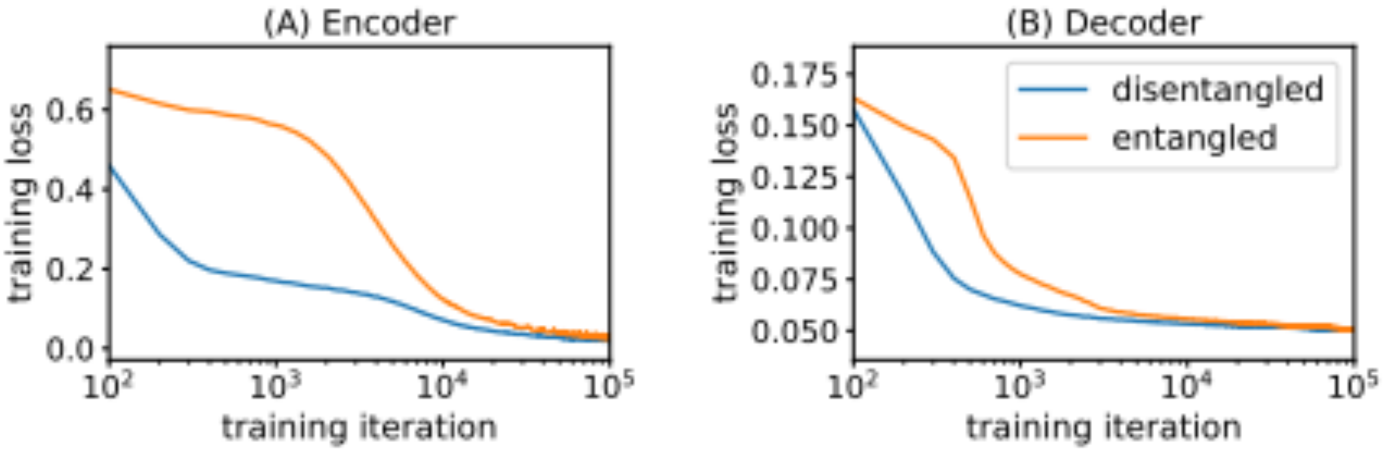
Before we start

- **Disentangled / Factorized representation**
  - Each variable in the inferred latent representation is only sensitive to one single generative factor and relatively invariant to other factors
  - Good interpretability and easy generalization to a variety of tasks

β-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework. Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. ICLR 2017.

Before we start

- **Unsupervised hierarchical representation learning**



Rethinking Style and Content Disentanglement in Variational Autoencoders. ICLR 2018 Workshops.

β-VAE

- **Unsupervised representation learning**
  - Augment the original VAE framework with a single hyper-parameter β that modulates the learning constraints
  - Impose a limit on the capacity of the latent information channel

β-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework. Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. ICLR 2017.

β-VAE

$$\max_{\phi,\theta} \mathbb{E}_{\mathbf{x}\sim\mathcal{D}}[\mathbb{E}_{\mathbf{z}\sim q_\phi(\mathbf{z}|\mathbf{x})} \log p_\theta(\mathbf{x}|\mathbf{z})]$$

$$\text{subject to } D_{\mathrm{KL}}(q_\phi(\mathbf{z}|\mathbf{x}) \| p_\theta(\mathbf{z})) < \delta$$

β-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework. Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. ICLR 2017.

β-VAE

$$
\begin{aligned}
\mathcal{F}(\theta, \phi, \beta) &= \mathbb{E}_{\mathbf{z}\sim q_\phi(\mathbf{z}|\mathbf{x})} \log p_\theta(\mathbf{x}|\mathbf{z}) - \beta(D_{\mathrm{KL}}(q_\phi(\mathbf{z}|\mathbf{x})||p_\theta(\mathbf{z})) - \delta) \\
&= \mathbb{E}_{\mathbf{z}\sim q_\phi(\mathbf{z}|\mathbf{x})} \log p_\theta(\mathbf{x}|\mathbf{z}) - \beta D_{\mathrm{KL}}(q_\phi(\mathbf{z}|\mathbf{x})||p_\theta(\mathbf{z})) + \beta\delta \\
&\geq \mathbb{E}_{\mathbf{z}\sim q_\phi(\mathbf{z}|\mathbf{x})} \log p_\theta(\mathbf{x}|\mathbf{z}) - \beta D_{\mathrm{KL}}(q_\phi(\mathbf{z}|\mathbf{x})||p_\theta(\mathbf{z})) \qquad \text{; Because } \beta,\delta \geq 0
\end{aligned}
$$

β-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework. Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. ICLR 2017.

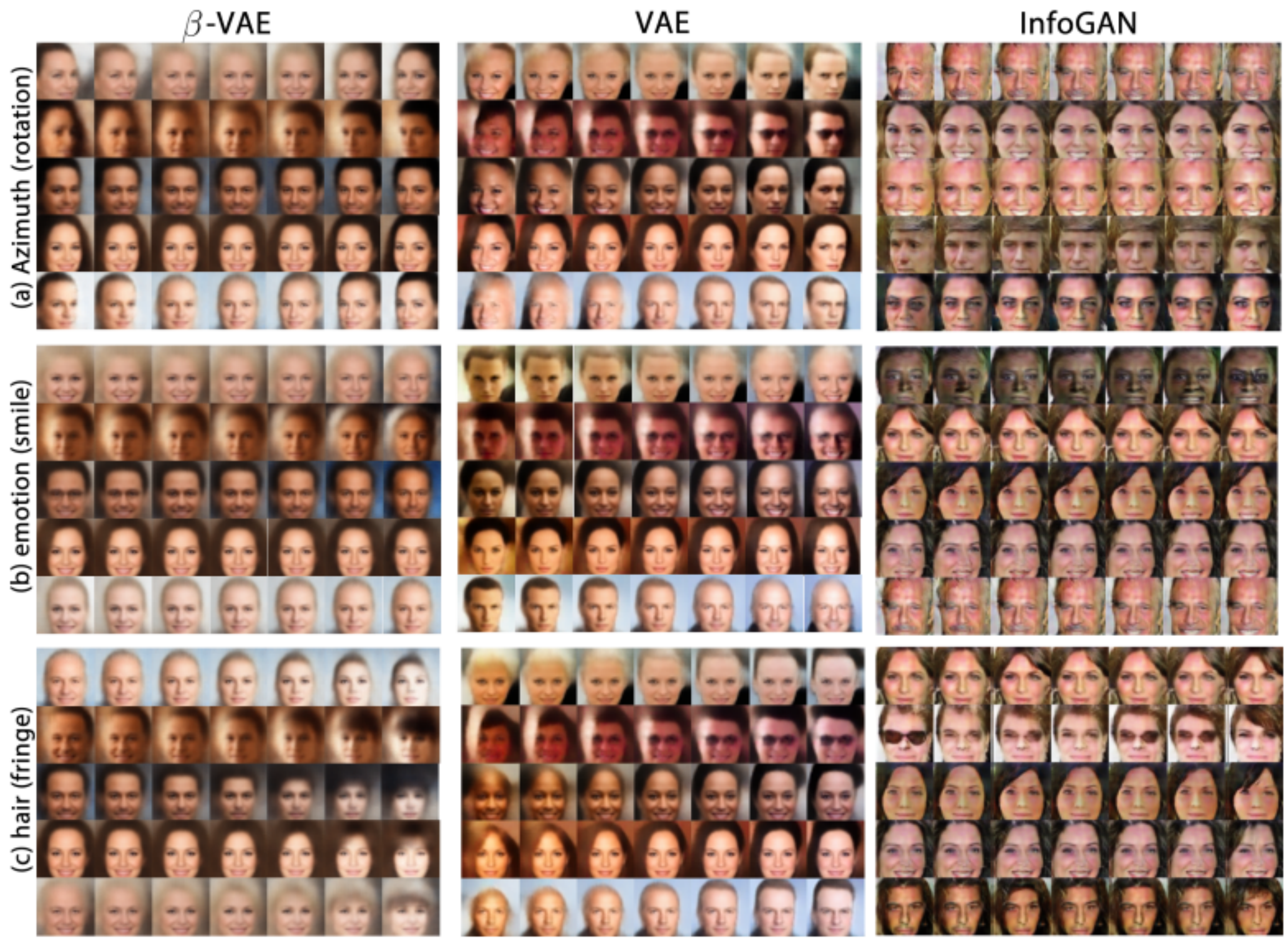
β-VAE

北京大學 PEKING UNIVERSITY

$$L_{\mathrm{BETA}}(\phi, \beta) = -\mathbb{E}_{\mathbf{z} \sim q_\phi(\mathbf{z}|\mathbf{x})} \log p_\theta(\mathbf{x}|\mathbf{z}) + \beta D_{\mathrm{KL}}\left(q_\phi(\mathbf{z}|\mathbf{x}) \| p_\theta(\mathbf{z})\right)$$

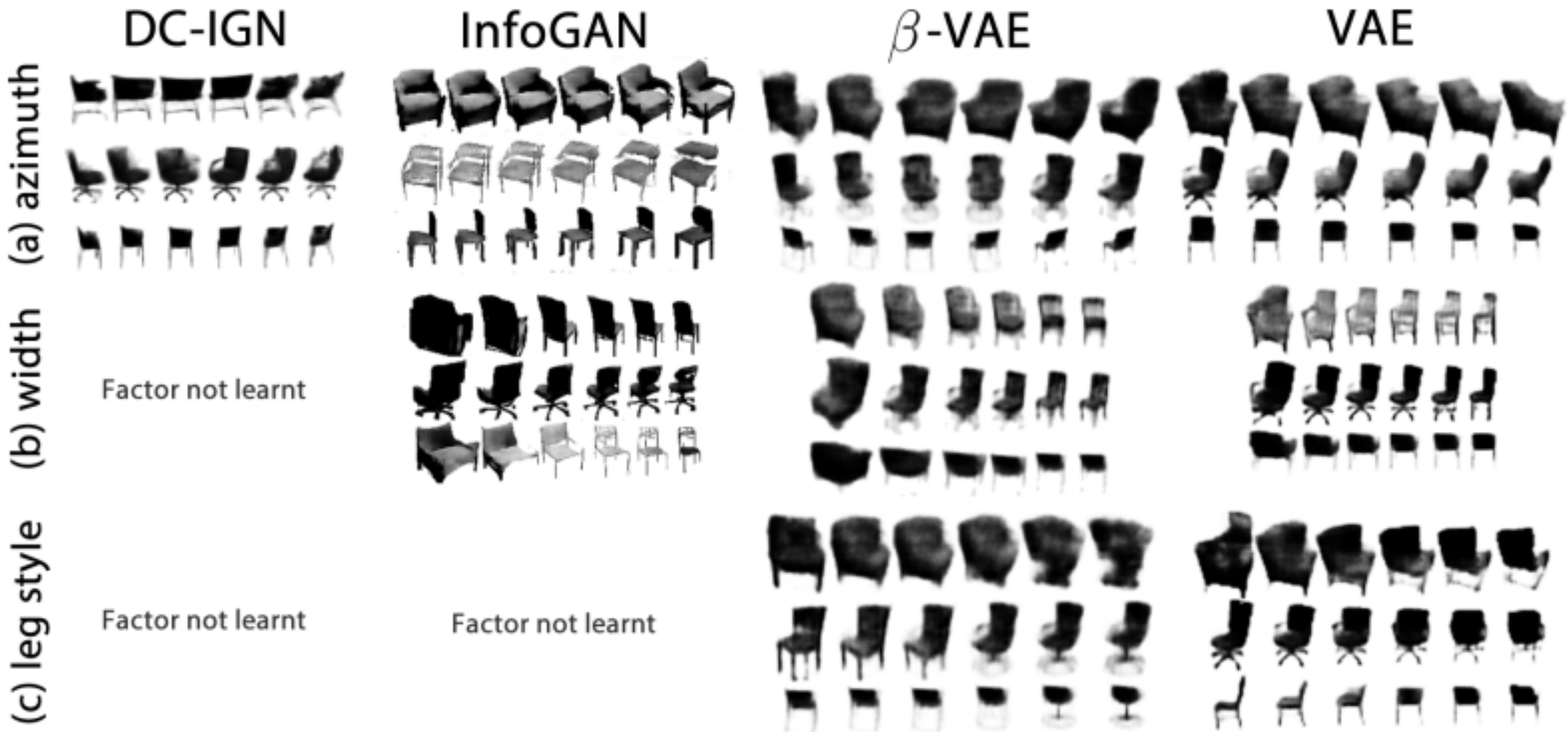
β-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework. Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. ICLR 2017.

β-VAE



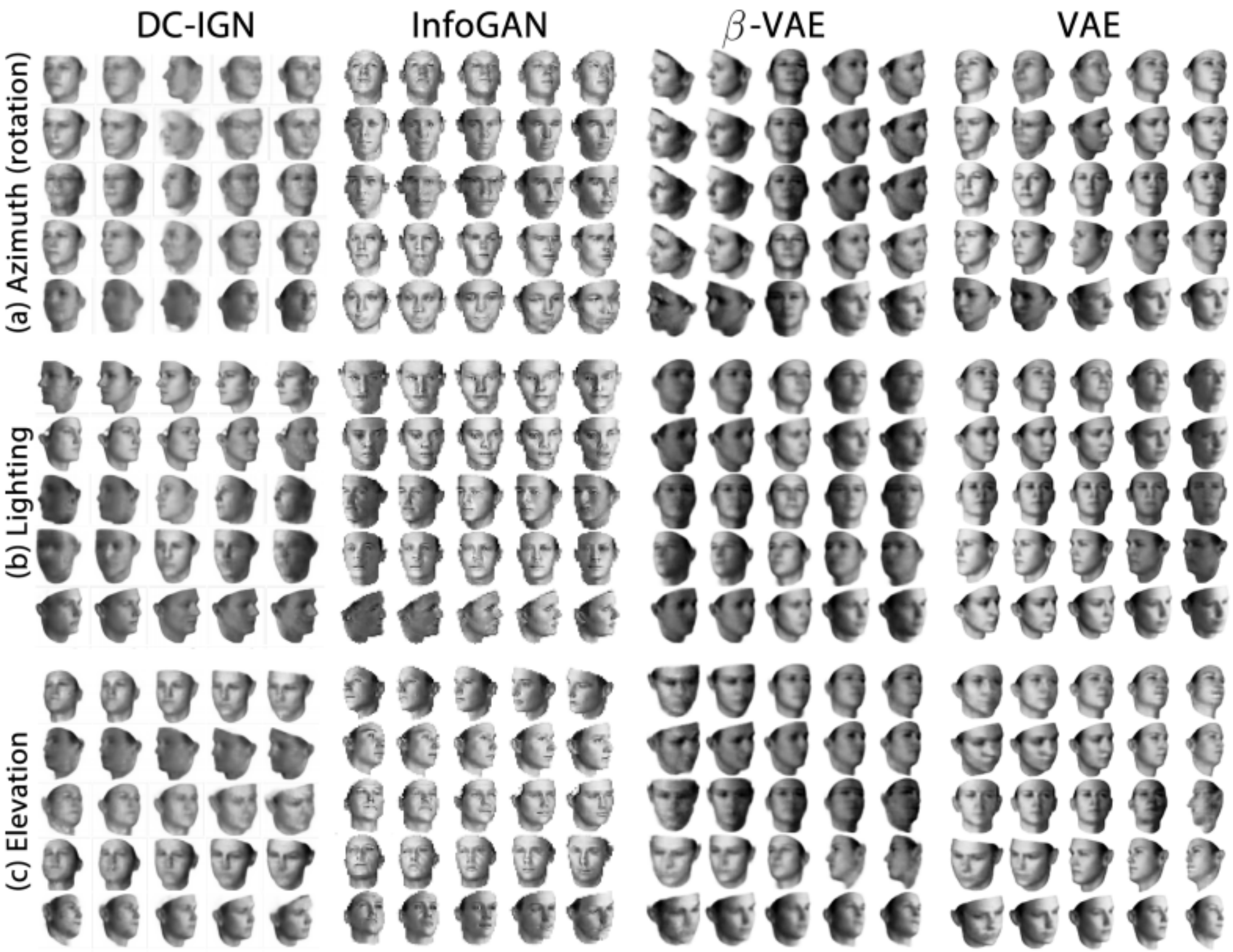
β-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework. Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. ICLR 2017.

β-VAE



β-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework. Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. ICLR 2017.

β-VAE



β-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework. Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. ICLR 2017.

β-VAE

- **Discussion: It is really unsupervised?**
  - It is unsupervised/self-supervised learning, because it does not need any label data
  - It is not fully unsupervised learning, it works because of the inductive bias of the neural network model, the hierarchical design introduces prior knowledge about the data

β-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework. Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. ICLR 2017.

VAE variants

Representation learning

Hierarchical representation learning

Temporal representation learning

- Convolutional VAE
- Conditional VAE
- β-VAE
- IWAE
- Ladder VAE
- Progressive + Fade-in VAE
- VAE in speech
- Temporal Difference VAE (TD-VAE)

IWAE

- **Optimize a tighter lower bound than VAE**
  - VAE just optimizes a lower bound of log P(X)

$$\log P(X) - \mathcal{D}\left[\underbrace{Q(z|X)}_{\text{encoder}} \| \underbrace{P(z|X)}_{\text{ideal}}\right] = \boxed{\underbrace{E_{z\sim Q}\left[\log P(X|z)\right]}_{\text{reconstruction}} - \underbrace{\mathcal{D}\left[Q(z|X) \| P(z)\right]}_{\text{KLD}}}$$

Importance Weighted Autoencoders. ICLR 2016.

IWAE

- **Optimize a tighter lower bound than VAE**

$$\log p(x) = \log \int p(x,z)dz = \log \int \frac{p(x,z)}{q(z|x)} q(z|x)dz = \log E_{q(z|x)}[\frac{p(x,z)}{q(z|x)}]$$

$$\log E_{q(z|x)}[\frac{p(x,z)}{q(z|x)}] = \log E_{z_1,z_2,\ldots,z_k \sim q(z|x)}[\frac{1}{k}\sum_{i=1}^{k}\frac{p(x,z_i)}{q(z_i|x)}]$$

$$L_k(x) = E_{z_1,z_2,\ldots,z_k \sim q(z|x)}[\log \frac{1}{k}\sum_{i=1}^{k}\frac{p(x,z_i)}{q(z_i|x)}] \leq \log E_{z_1,z_2,\ldots,z_k \sim q(z|x)}[\frac{1}{k}\sum_{i=1}^{k}\frac{p(x,z_i)}{q(z_i|x)}] = \boxed{\log p(x)}$$

Importance Weighted Autoencoders. ICLR 2016.

IWAE

$$ELBO(\theta) = \mathbf{E}_q[\log p(x, z)] - \mathbf{E}_q[\log q_\theta(z \mid x)]$$

VAE 的loss： $E_{z\sim q(z|x)}[\log \frac{p(x,z)}{q(z|x)}]$

而IWAE的loss： $E_{z_1,z_2,\ldots,z_k\sim q(z|x)}[\log \frac{1}{k}\sum_{i=1}^{k}\frac{p(x,z_i)}{q(z_i|x)}]$

Importance Weighted Autoencoders. ICLR 2016.

IWAE

- **Why "Importance weighted"**

$$\nabla_\theta \; E_{z_1, z_2, \ldots, z_k \sim q(z|x,\theta)}[\log \frac{1}{k} \sum_{i=1}^{k} \frac{p(x, z_i|\theta)}{q(z_i|x,\theta)}] = E_{z_1, z_2, \ldots, z_k \sim q(z|x,\theta)}[\nabla_\theta \; \log \frac{1}{k} \sum_{i=1}^{k} w_i]$$

where $w_i = \dfrac{p(x, z_i|\theta)}{q(z_i|x,\theta)}$

VAE: $\dfrac{1}{k} \sum_{i=1}^{k} \nabla_\theta \log w_i$

IWAE: $\sum_{i=1}^{k} \tilde{w_i} \nabla_\theta \; \log w_i$

Importance Weighted Autoencoders. ICLR 2016.

VAE variants

Representation learning
- Convolutional VAE
- Conditional VAE
- β-VAE
- IWAE

Hierarchical representation learning
- Ladder VAE
- Progressive + Fade-in VAE

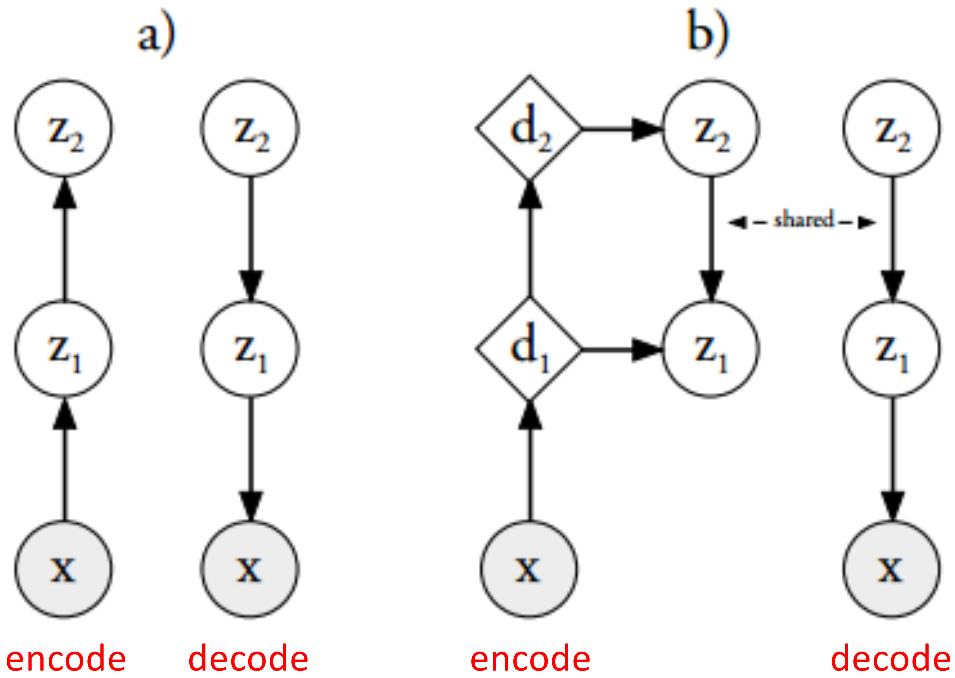
Temporal representation learning
- VAE in speech
- Temporal Difference VAE (TD-VAE)

Ladder VAE

- **To learn hierarchical latent representation**
- **Deep models with several layers of dependent stochastic variables are difficult to train**
  - Limiting the improvements obtained using these highly expressive models

LVAE: Ladder Variational Autoencoder. Casper Kaae Sønderby, Tapani Raiko, Lars Maaløe, Søren Kaae Sønderby, and Ole Winther. NIPS 2016.

Ladder VAE



Hierarchical VAE

Ladder VAE

LVAE: Ladder Variational Autoencoder. Casper Kaae Sønderby, Tapani Raiko, Lars Maaløe, Søren Kaae Sønderby, and Ole Winther. NIPS 2016.

Ladder VAE

$$\mathcal{L}(\theta, \phi; \mathbf{x})_{WU} = -\beta KL(q_\phi(z|x)||p_\theta(\mathbf{z})) + E_{q_\phi(z|x)}\left[\log p_\theta(\mathbf{x}|\mathbf{z})\right]$$

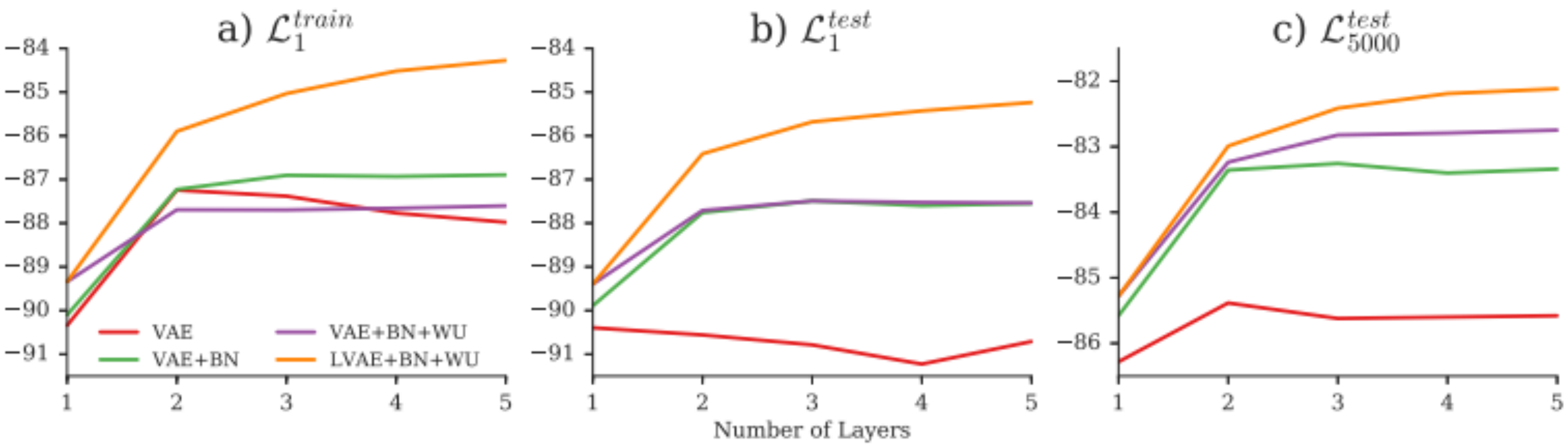


Figure 3: MNIST log-likelihood values for VAEs and the LVAE model with different number of latent layers, Batch-normalization (*BN*) and Warm-up (*WU*). a) Train log-likelihood, b) test log-likelihood and c) test log-likelihood with 5000 importance samples.

LVAE: Ladder Variational Autoencoder. Casper Kaae Sønderby, Tapani Raiko, Lars Maaløe, Søren Kaae Sønderby, and Ole Winther. NIPS 2016.

Ladder VAE

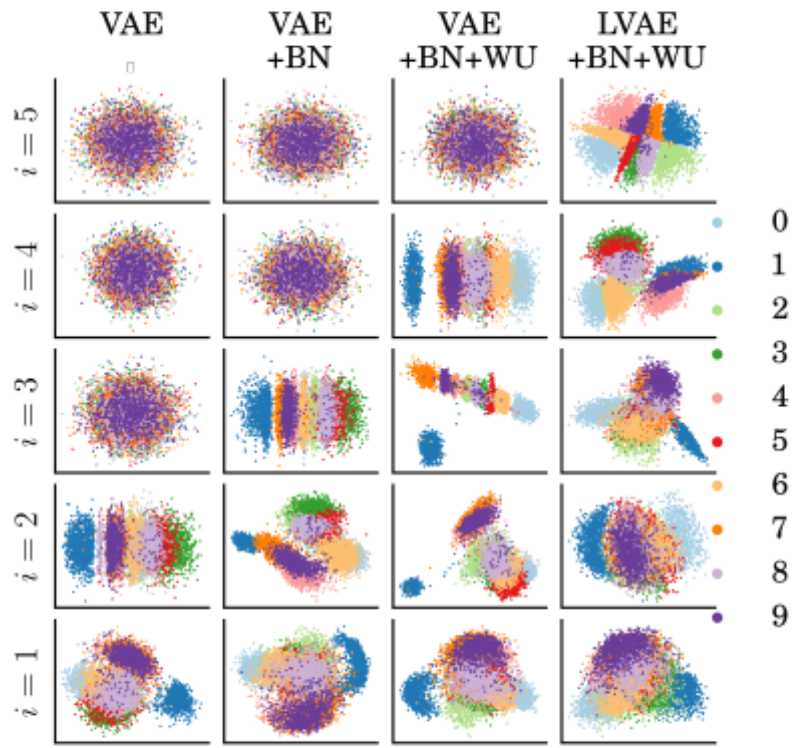
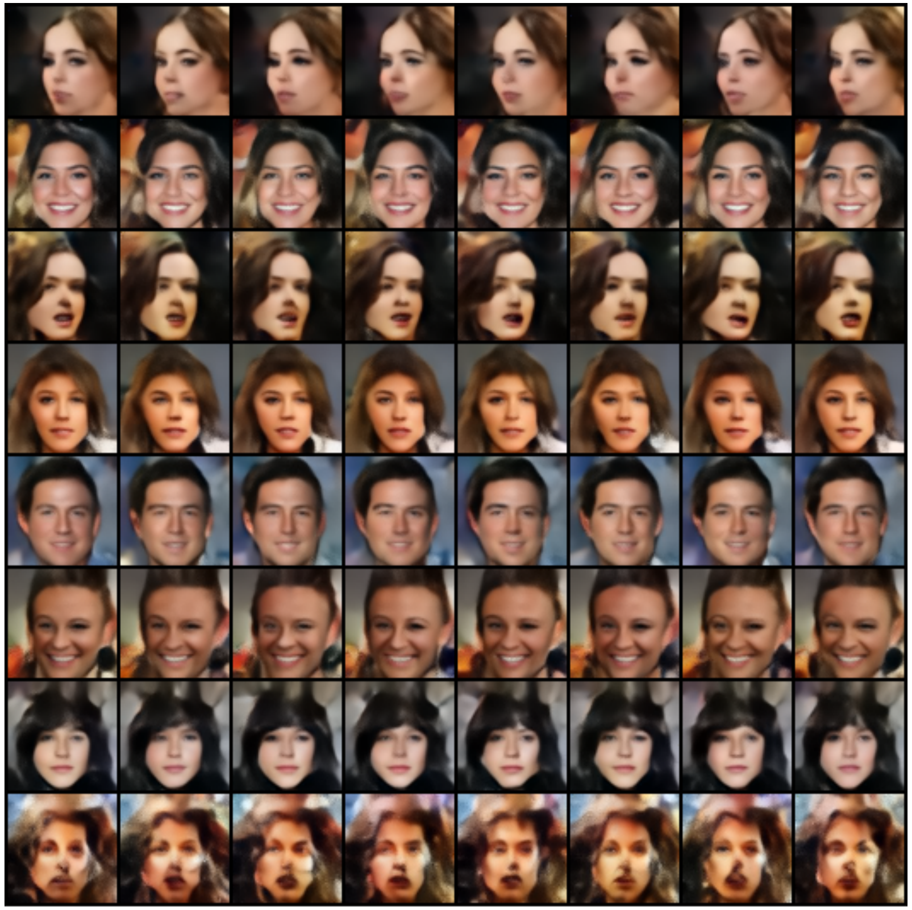


Figure 6: PCA-plots of samples from $q(z_i|z_{i-1})$ for 5-layer VAE and LVAE models trained on MNIST. Color-coded according to true class label

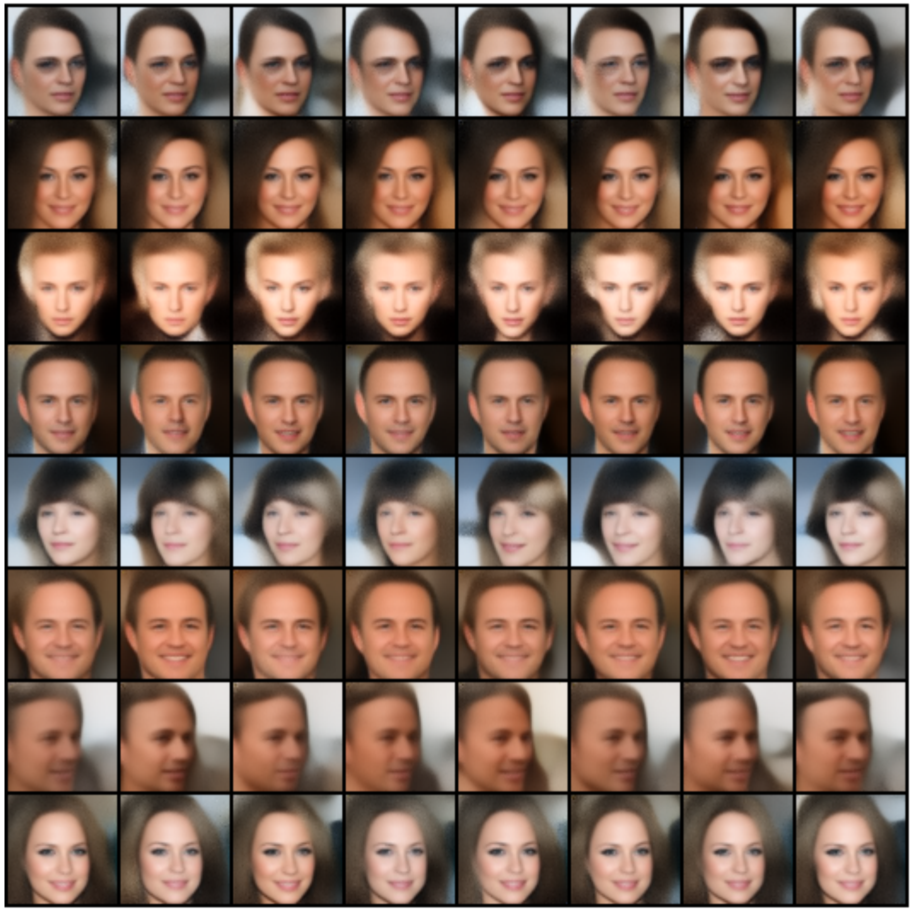
LVAE: Ladder Variational Autoencoder. Casper Kaae Sønderby, Tapani Raiko, Lars Maaløe, Søren Kaae Sønderby, and Ole Winther. NIPS 2016.

Ladder VAE



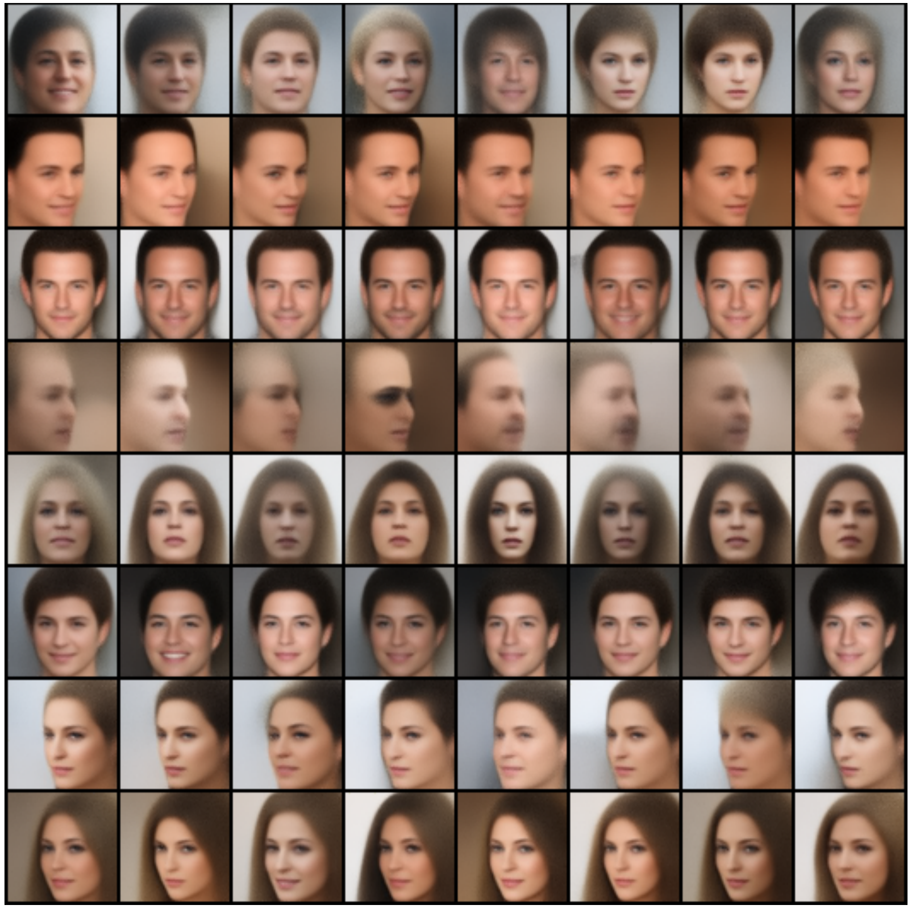
LVAE: Ladder Variational Autoencoder. Casper Kaae Sønderby, Tapani Raiko, Lars Maaløe, Søren Kaae Sønderby, and Ole Winther. NIPS 2016.

Ladder VAE



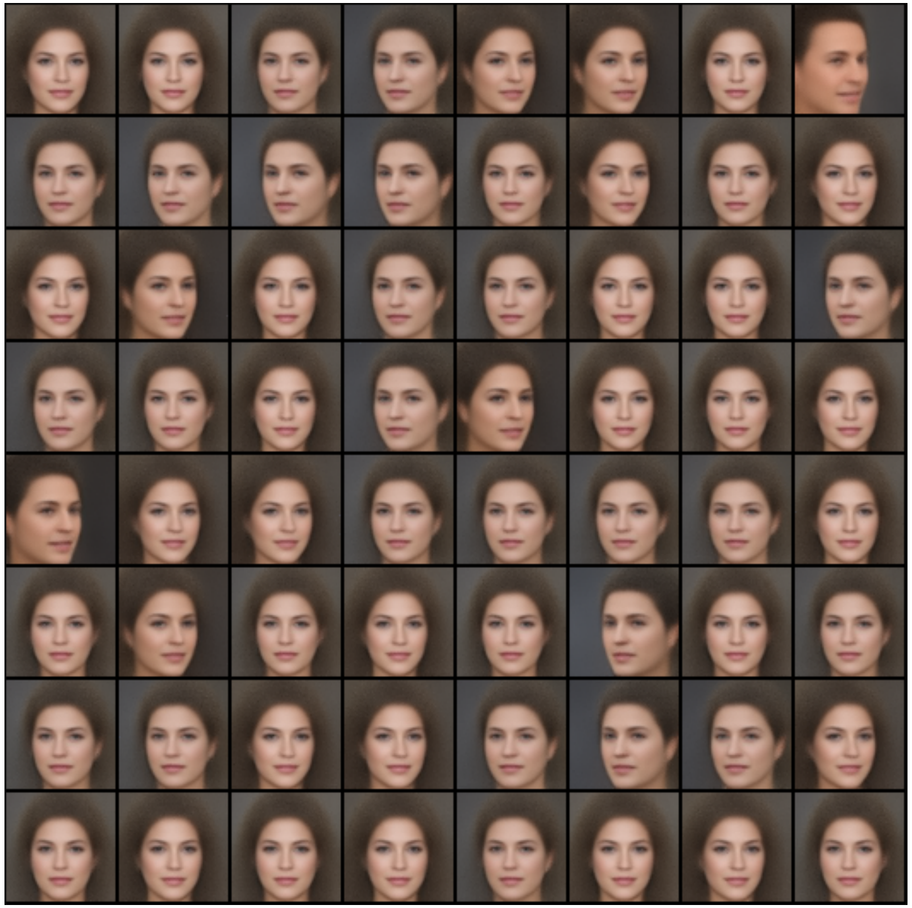
LVAE: Ladder Variational Autoencoder. Casper Kaae Sønderby, Tapani Raiko, Lars Maaløe, Søren Kaae Sønderby, and Ole Winther. NIPS 2016.

Ladder VAE



LVAE: Ladder Variational Autoencoder. Casper Kaae Sønderby, Tapani Raiko, Lars Maaløe, Søren Kaae Sønderby, and Ole Winther. NIPS 2016.

Ladder VAE



LVAE: Ladder Variational Autoencoder. Casper Kaae Sønderby, Tapani Raiko, Lars Maaløe, Søren Kaae Sønderby, and Ole Winther. NIPS 2016.

VAE variants

北京大學
PEKING UNIVERSITY

- Convolutional VAE
- Conditional VAE
- β-VAE
- IWAE
- Ladder VAE
- Progressive + Fade-in VAE
- VAE in speech
- Temporal Difference VAE (TD-VAE)

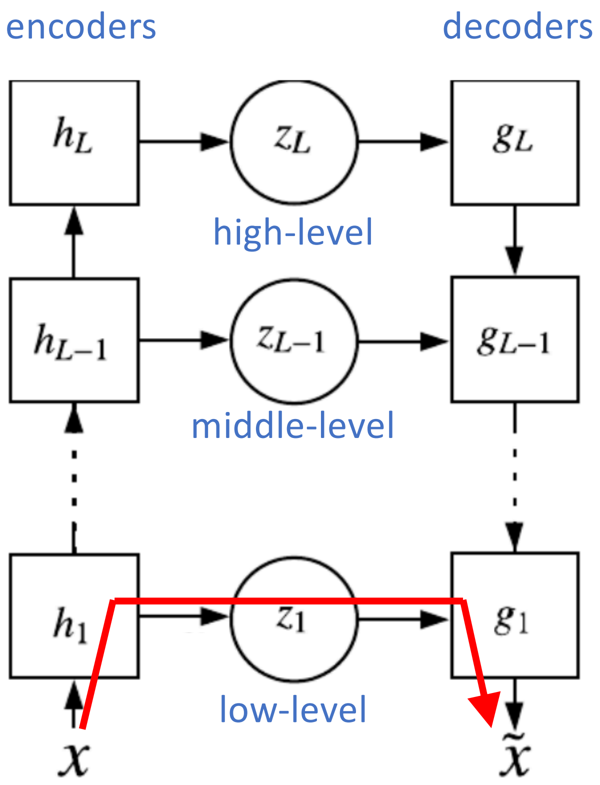
Representation learning: β-VAE, IWAE

Hierarchical representation learning: Ladder VAE, Progressive + Fade-in VAE

Temporal representation learning: VAE in speech, Temporal Difference VAE (TD-VAE)

Progressive + Fade-in VAE

- **Discussion**

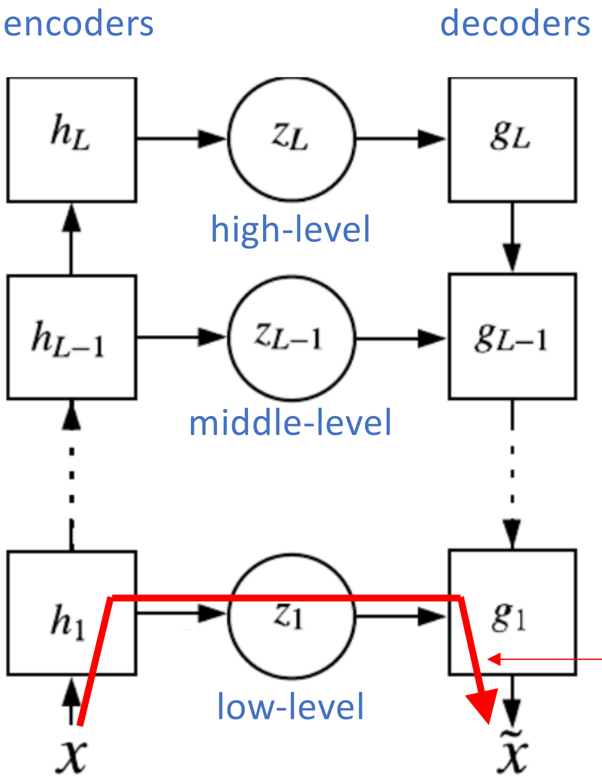


Can we directly train a hierarchical VAE with ladder structure like that?

Progressive Learning and Disentanglement of Hierarchical Representations. ICLR 2020.

Progressive + Fade-in VAE

- **Discussion**



$$
\begin{aligned}
\mathrm{E}_{q(x)}[KL(q(z|x)||p(z))] &= \int\int q(x)q(z|x)log\frac{q(z|x)}{p(z)}dxdz \\
&= \int\int q(x,z)log\frac{q(z|x)q(x)q(z)}{p(z)q(x)p(z)}dxdz \\
&= \int\int [q(x,z)log\frac{q(x,z)}{q(x)q(z)} + q(x,z)log\frac{q(z)}{p(z)}]dxdz \\
&= MI_{q(x,z)}(x,z) + \int\int q(x,z)log\frac{q(z)}{p(z)}dxdz \\
&= MI_{q(x,z)}(x,z) + \int q(z)log\frac{q(z)}{p(z)}dz \\
&= \boxed{MI_{q(x,z)}(x,z)} + KL(q(z)||p(z))
\end{aligned}
$$

Information **SHORTCUT** problem

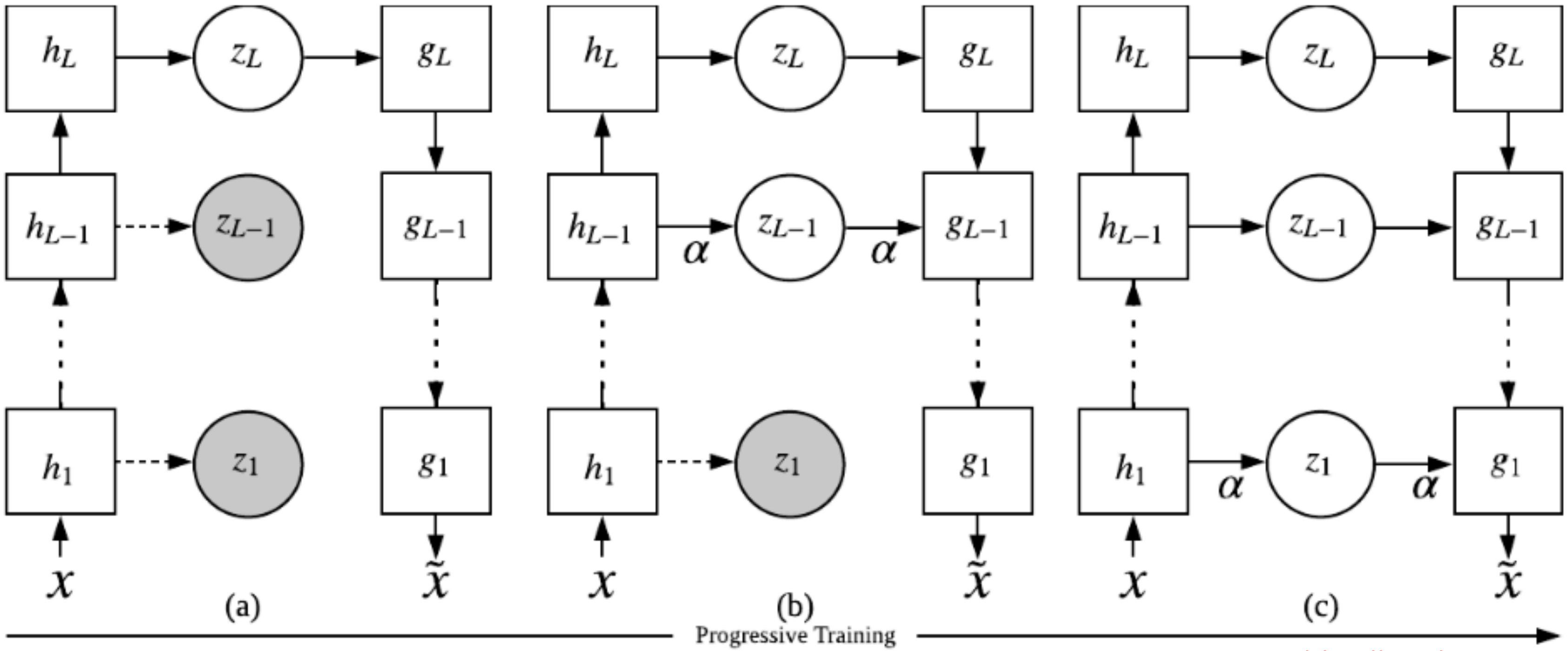
all information go through the low-level path,
other paths are ignored.
model is lazy…

Progressive Learning and Disentanglement of Hierarchical Representations. ICLR 2020.

Progressive + Fade-in VAE

- **Progressive + Fade-in**

α increase from 0 to 1 gradually



Stage 1: enable high-level path

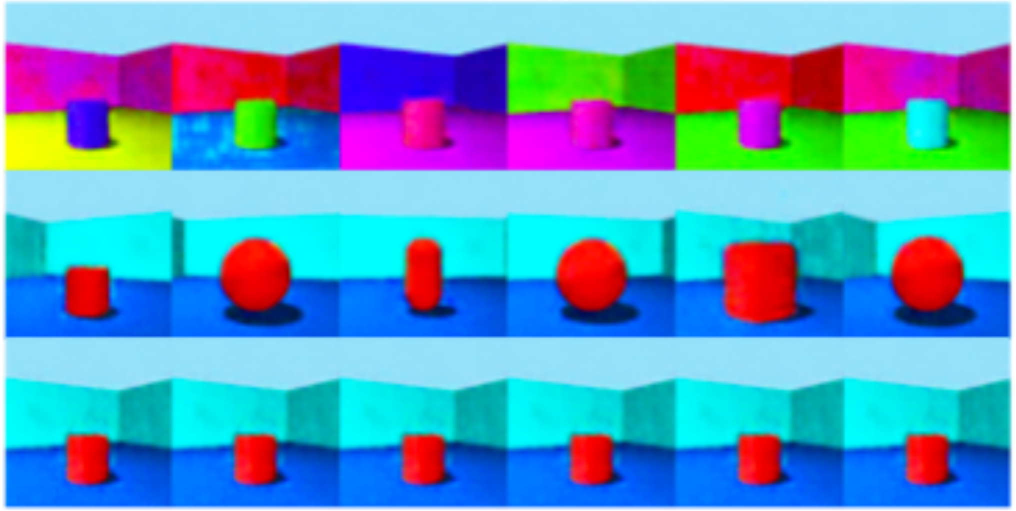
Stage 2: enable high-level and middle paths

Stage 3: enable all paths …

Progressive Learning and Disentanglement of Hierarchical Representations. ICLR 2020.

Progressive + Fade-in VAE

- **Results**



z_3 High-level: background and foreground colors

z_2 Middle-level: shape

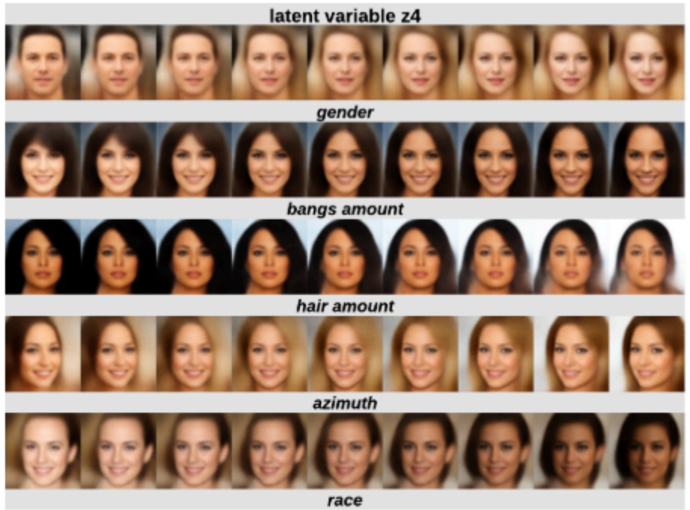
z_1 Low-level: …

Progressive Learning and Disentanglement of Hierarchical Representations. ICLR 2020.

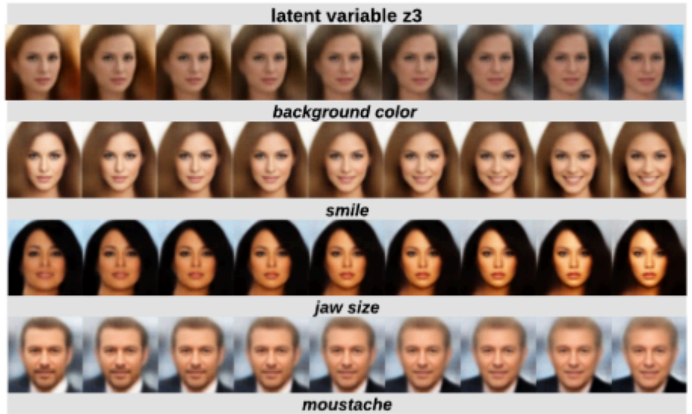
Progressive + Fade-in VAE

- **Results**

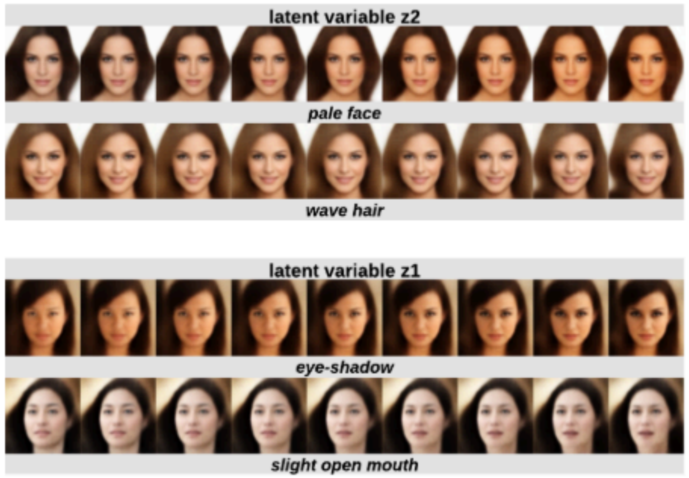
High-level



Middle-level



Low-level



Progressive Learning and Disentanglement of Hierarchical Representations. ICLR 2020.

VAE variants

- Convolutional VAE
- Conditional VAE

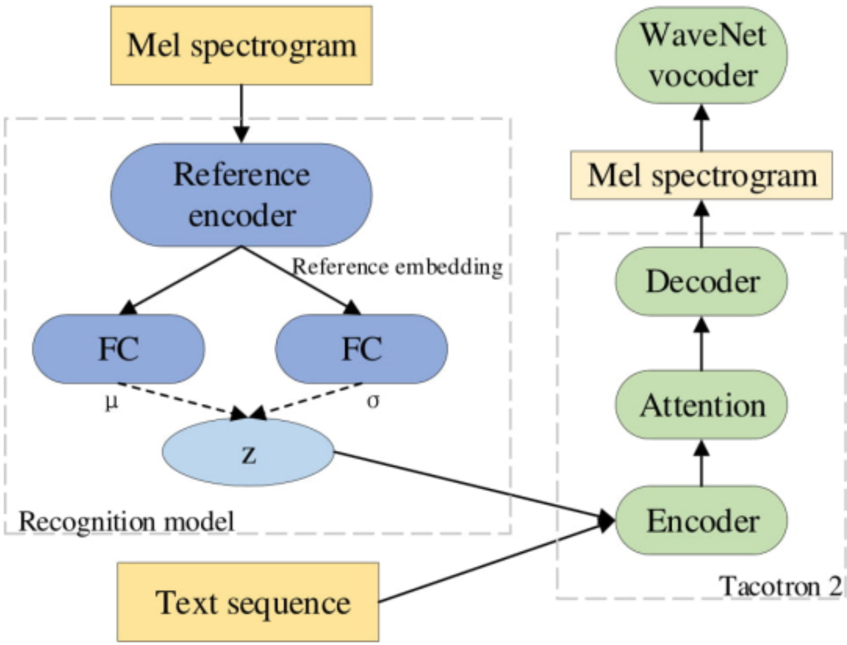
Representation learning
- β-VAE
- IWAE

Hierarchical representation learning
- Ladder VAE
- Progressive + Fade-in VAE

Temporal representation learning
- VAE in speech
- Temporal Difference VAE (TD-VAE)

VAE in speech

- **Learning latent representations for style control and transfer in end-to-end speech synthesis**
- **RNN as encoder**



Learning latent representations for style control and transfer in end-to-end speech synthesis. ICASSP 2019.

VAE variants

PEKING UNIVERSITY

- Convolutional VAE
- Conditional VAE

Representation learning
- β-VAE
- IWAE

Hierarchical representation learning
- Ladder VAE
- Progressive + Fade-in VAE

Temporal representation learning
- VAE in speech
- Temporal Difference VAE (TD-VAE)

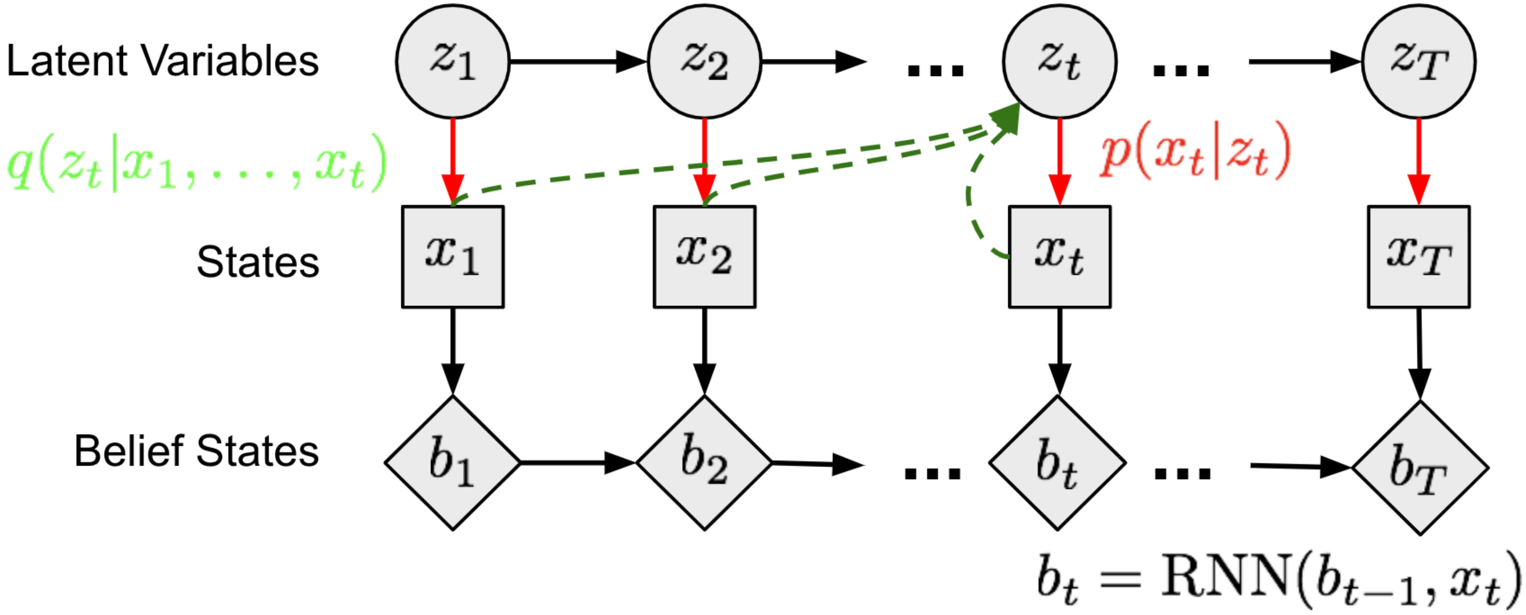
TD-VAE

- **To model temporal information**

Temporal Difference VAE. ICLR 2019.

TD-VAE

- **State-space model as a Markov Chain model**



Temporal Difference VAE. ICLR 2019.

TD-VAE

$$b_t = belief(x_1, \ldots, x_t) = belief(b_{t-1}, x_t) \qquad b_t = \mathrm{RNN}(b_{t-1}, x_t)$$

$$p(x_{t+1}, \ldots, x_T | x_1, \ldots, x_t) \approx p(x_{t+1}, \ldots, x_T | b_t)$$

Temporal Difference VAE. ICLR 2019.

TD-VAE

$$
\begin{aligned}
\log p(x) &\geq \log p(x) - D_{\mathrm{KL}}(q(z|x) \| p(z|x)) \\
&= \mathbb{E}_{z\sim q} \log p(x|z) - D_{\mathrm{KL}}(q(z|x) \| p(z)) \\
&= \mathbb{E}_{z\sim q} \log p(x|z) - \mathbb{E}_{z\sim q} \log \frac{q(z|x)}{p(z)} \\
&= \mathbb{E}_{z\sim q} [\log p(x|z) - \log q(z|x) + \log p(z)] \\
&= \mathbb{E}_{z\sim q} [\log p(x, z) - \log q(z|x)] \\
\log p(x) &\geq \mathbb{E}_{z\sim q} [\log p(x, z) - \log q(z|x)]
\end{aligned}
$$

Temporal Difference VAE. ICLR 2019.

TD-VAE

$$\begin{aligned}
&\log p(x_t|x_{<t}) \\
&\geq \mathbb{E}_{(z_{t-1},z_t)\sim q}[\log p(x_t, z_{t-1}, z_t|x_{<t}) - \log q(z_{t-1}, z_t|x_{\leq t})] \\
&\geq \mathbb{E}_{(z_{t-1},z_t)\sim q}[\log p(x_t|z_{t-1}, z_t, x_{<t}) + \log p(z_{t-1}, z_t|x_{<t}) - \log q(z_{t-1}, z_t|x_{\leq t})] \\
&\geq \mathbb{E}_{(z_{t-1},z_t)\sim q}[\log p(x_t|z_t) + \log p(z_{t-1}|x_{<t}) + \log p(z_t|z_{t-1}) - \log q(z_{t-1}, z_t|x_{\leq t})] \\
&\geq \mathbb{E}_{(z_{t-1},z_t)\sim q}[\log p(x_t|z_t) + \log p(z_{t-1}|x_{<t}) + \log p(z_t|z_{t-1}) - \log q(z_t|x_{\leq t}) - \log q(z_{t-1}|z_t, x_{\leq t})]
\end{aligned}$$

Notice two things:

- The red terms can be ignored according to Markov assumptions.
- The blue term is expanded according to Markov assumptions.
- The green term is expanded to include an one-step prediction back to the past as a smoothing distribution.

Temporal Difference VAE. ICLR 2019.

TD-VAE

$$\log p(x_t|x_{<t}) \geq \mathbb{E}_{(z_{t-1},z_t)\sim q}[\log p(x_t|z_t) + \log p(z_{t-1}|x_{<t}) + \log p(z_t|z_{t-1}) - \log q(z_t|x_{\leq t}) - \log q(z_{t-1}|z_t, x_{\leq t})]$$

Precisely, there are four types of distributions to learn:

1. $p_D(.)$ is the **decoder** distribution:
   - $p(x_t \mid z_t)$ is the encoder by the common definition;
   - $p(x_t \mid z_t) \to p_D(x_t \mid z_t)$;
2. $p_T(.)$ is the **transition** distribution:
   - $p(z_t \mid z_{t-1})$ captures the sequential dependency between latent variables;
   - $p(z_t \mid z_{t-1}) \to p_T(z_t \mid z_{t-1})$;
3. $p_B(.)$ is the **belief** distribution:
   - Both $p(z_{t-1} \mid x_{<t})$ and $q(z_t \mid x_{\leq t})$ can use the belief states to predict the latent variables;
   - $p(z_{t-1} \mid x_{<t}) \to p_B(z_{t-1} \mid b_{t-1})$;
   - $q(z_t \mid x_{\leq t}) \to p_B(z_t \mid b_t)$;
4. $p_S(.)$ is the **smoothing** distribution:
   - The back-to-past smoothing term $q(z_{t-1} \mid z_t, x_{\leq t})$ can be rewritten to be dependent of belief states too;
   - $q(z_{t-1} \mid z_t, x_{\leq t}) \to p_S(z_{t-1} \mid z_t, b_{t-1}, b_t)$;

Temporal Difference VAE. ICLR 2019.

TD-VAE

To incorporate the idea of jumpy prediction, the sequential ELBO has to not only work on $t, t+1$, but also two distant timestamp $t_1 < t_2$. Here is the final TD-VAE objective function to maximize:

$$J_{t_1,t_2} = \mathbb{E}[\log p_D(x_{t_2}|z_{t_2}) + \log p_B(z_{t_1}|b_{t_1}) + \log p_T(z_{t_2}|z_{t_1}) - \log p_B(z_{t_2}|b_{t_2}) - \log p_S(z_{t_1}|z_{t_2}, b_{t_1}, b_{t_2})]$$

Temporal Difference VAE. ICLR 2019.

Summary

Representation learning

Hierarchical representation learning

Temporal representation learning

- Convolutional VAE
- Conditional VAE
- β-VAE
- IWAE
- Ladder VAE
- Progressive + Fade-in VAE
- VAE in speech
- Temporal Difference VAE (TD-VAE)

Thanks