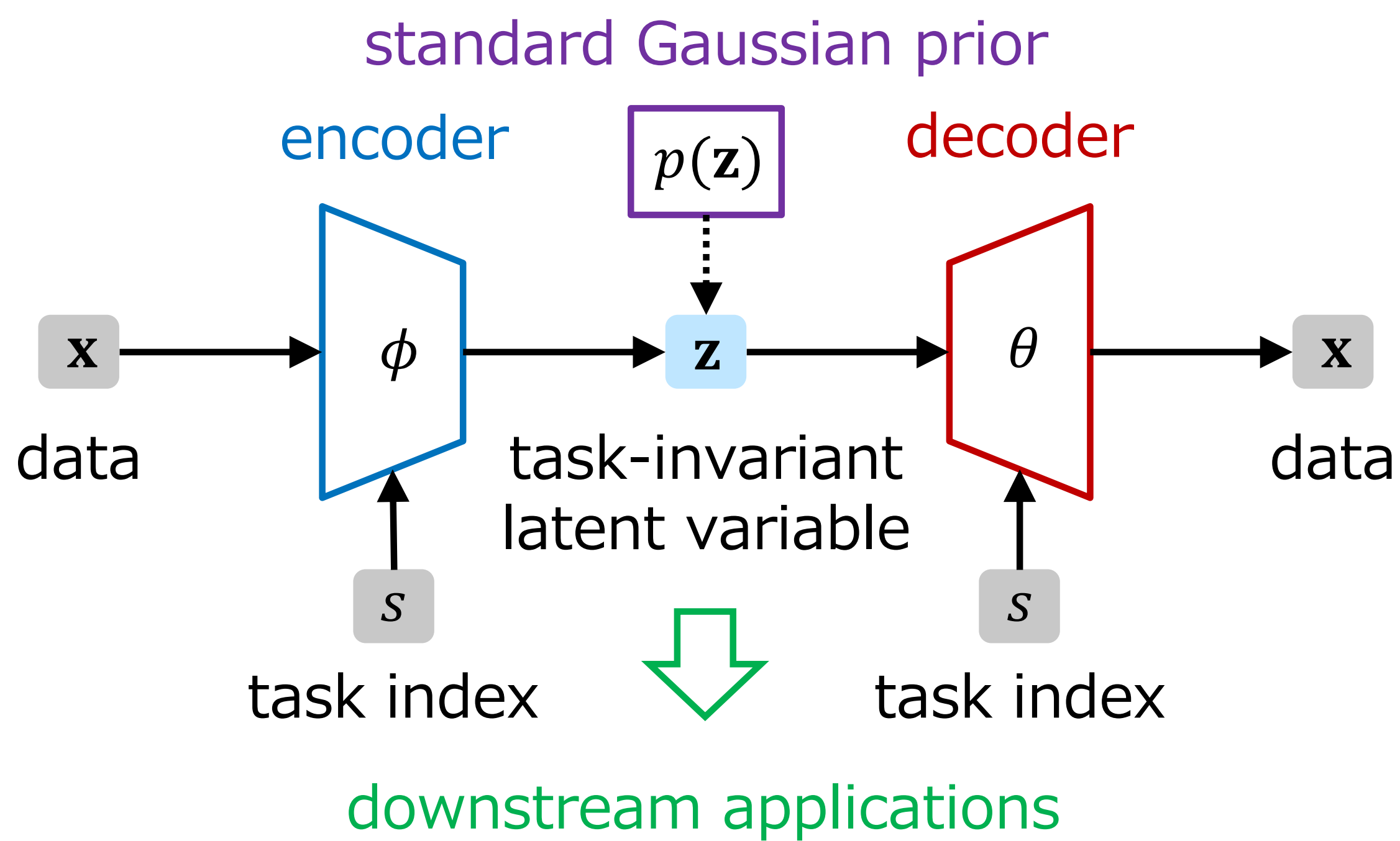


## 1. Conditional Variational Autoencoder

- Variational Autoencoder (VAE) は教師なし表現学習のための強力な潜在変数モデルだが、学習データ数が少ない時性能が低下することが知られている
- この問題を解決するため、複数タスクからタスク不変な潜在変数を学習する Conditional VAE (CVAE) が提案されている



- CVAEは、 $s$  が与えられた下での  $\mathbf{x}$  の条件付き確率を下記でモデル化する:

$$p_{\theta}(\mathbf{x}|\mathbf{z}, s) = \int \underbrace{p_{\theta}(\mathbf{x}|\mathbf{z}, s)}_{\text{decoder prior}} p(\mathbf{z}) d\mathbf{z} = \underbrace{\mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x}, s)}}_{\text{encoder}} \left[ \frac{p_{\theta}(\mathbf{x}|\mathbf{z}, s)p(\mathbf{z})}{q_{\phi}(\mathbf{z}|\mathbf{x}, s)} \right]$$

- CVAEは、下記の対数尤度の下界である変分下界 (evidence lower bound, ELBO) を最大化するように学習される:

$$\mathcal{F}_{\text{CVAE}}(\theta, \phi) = \mathbb{E}_{p_D(\mathbf{x}, s)q_{\phi}(\mathbf{z}|\mathbf{x}, s)} [\ln p_{\theta}(\mathbf{x}|\mathbf{z}, s)] - \mathbb{E}_{p_D(\mathbf{x}, s)} [D_{KL}(q_{\phi}(\mathbf{z}|\mathbf{x}, s) \| p(\mathbf{z}))]$$

データ分布

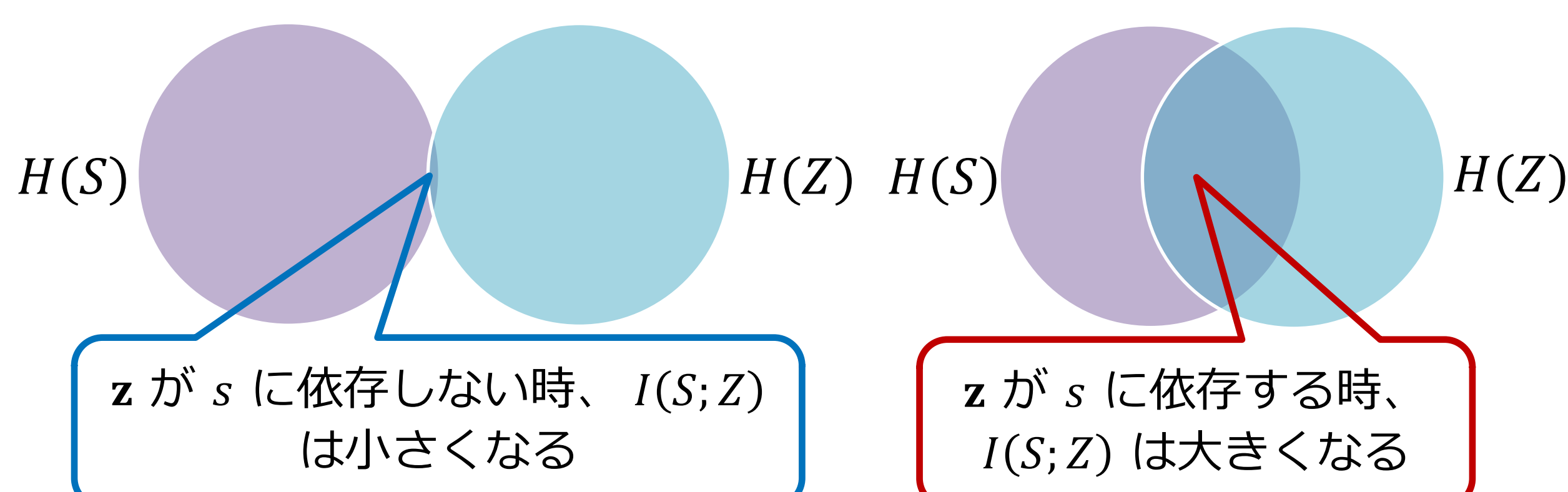
$= \mathcal{R}(\phi)$

## 2. 問題と貢献

- CVAEは、ある程度  $\mathbf{z}$  の  $s$  への依存性を減らせるが、多くの場合はこのタスク依存性が残ってしまう
- この研究の貢献は下記の3点:
  - タスク依存性の原因を調査し、**単純な事前分布**が原因の一つであることを明らかにした
  - タスク依存性を減らすという観点で、**最適な事前分布**を提案した
  - 提案手法で学習した表現が複数のタスク上で有効であることを、理論的・実験的に示した

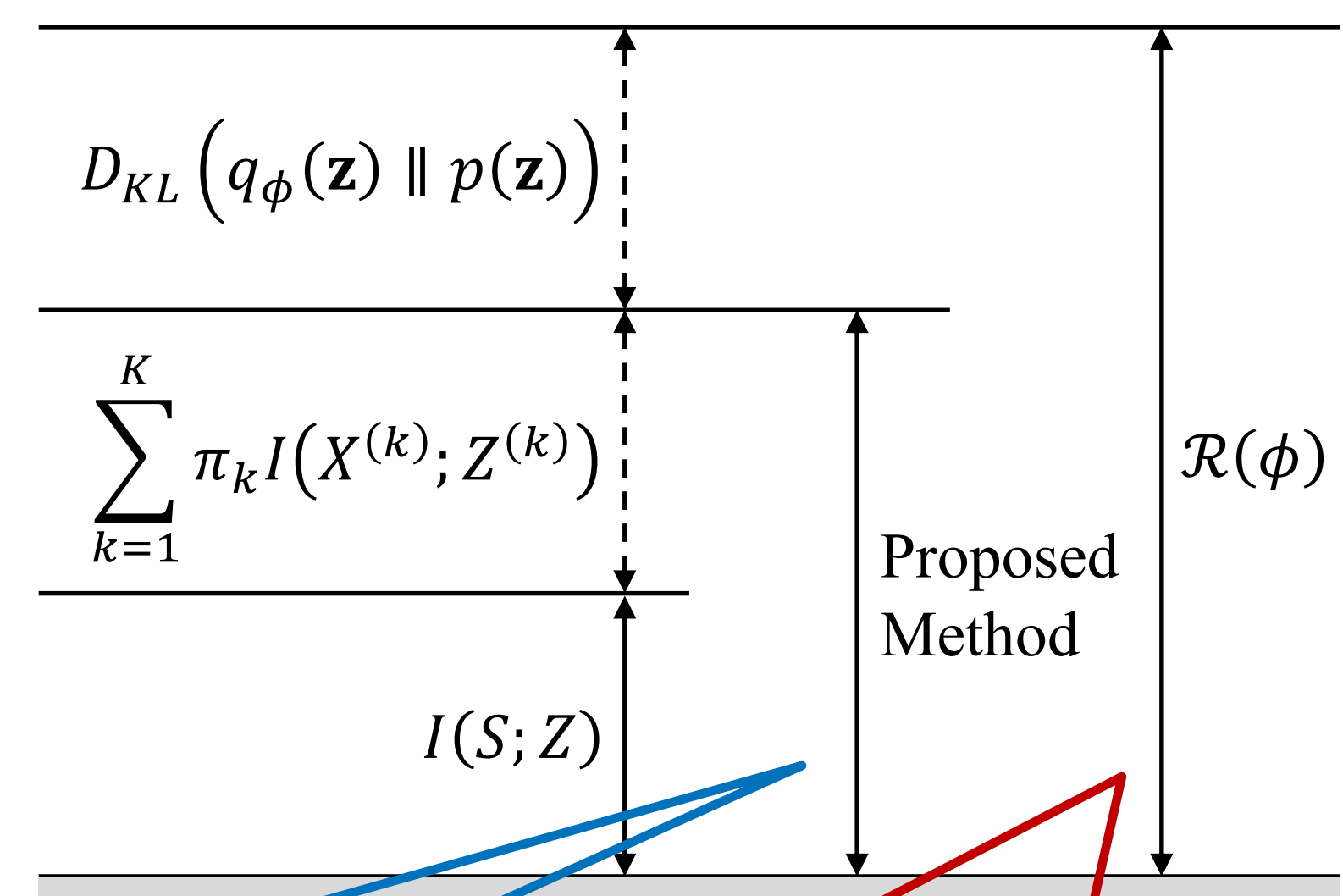
## 3. 相互情報量

- $\mathbf{z}$  の  $s$  への依存性の原因を調べるために、二つの確率変数の相互的な依存性を測定する相互情報量  $I(S; Z)$  を導入する



## 4a. 定理 1

- CVAEは、相互情報量  $I(S; Z)$  をその**ルーズ**な上界である  $\mathcal{R}(\phi)$  を通して最小化している



$p(\mathbf{z}) = q_{\phi}(\mathbf{z})$  のとき、 $\mathcal{R}(\phi)$  は  $I(S; Z)$  の**最もタイト**な上界となる

$D_{KL}(q_{\phi}(\mathbf{z}) \| p(\mathbf{z}))$  が通常大きい値を取るため、 $\mathcal{R}(\phi)$  は  $I(S; Z)$  の**ルーズ**な上界となっている

- つまり、単純な事前分布  $p(\mathbf{z})$  が**タスク依存性の原因の一つ**であり、 $q_{\phi}(\mathbf{z}) = \int q_{\phi}(\mathbf{z}|\mathbf{x}, s)p_D(\mathbf{x}, s)d\mathbf{x}$  がタスク依存性を減らす上での**最適な事前分布**である

## 4b. 定理 2

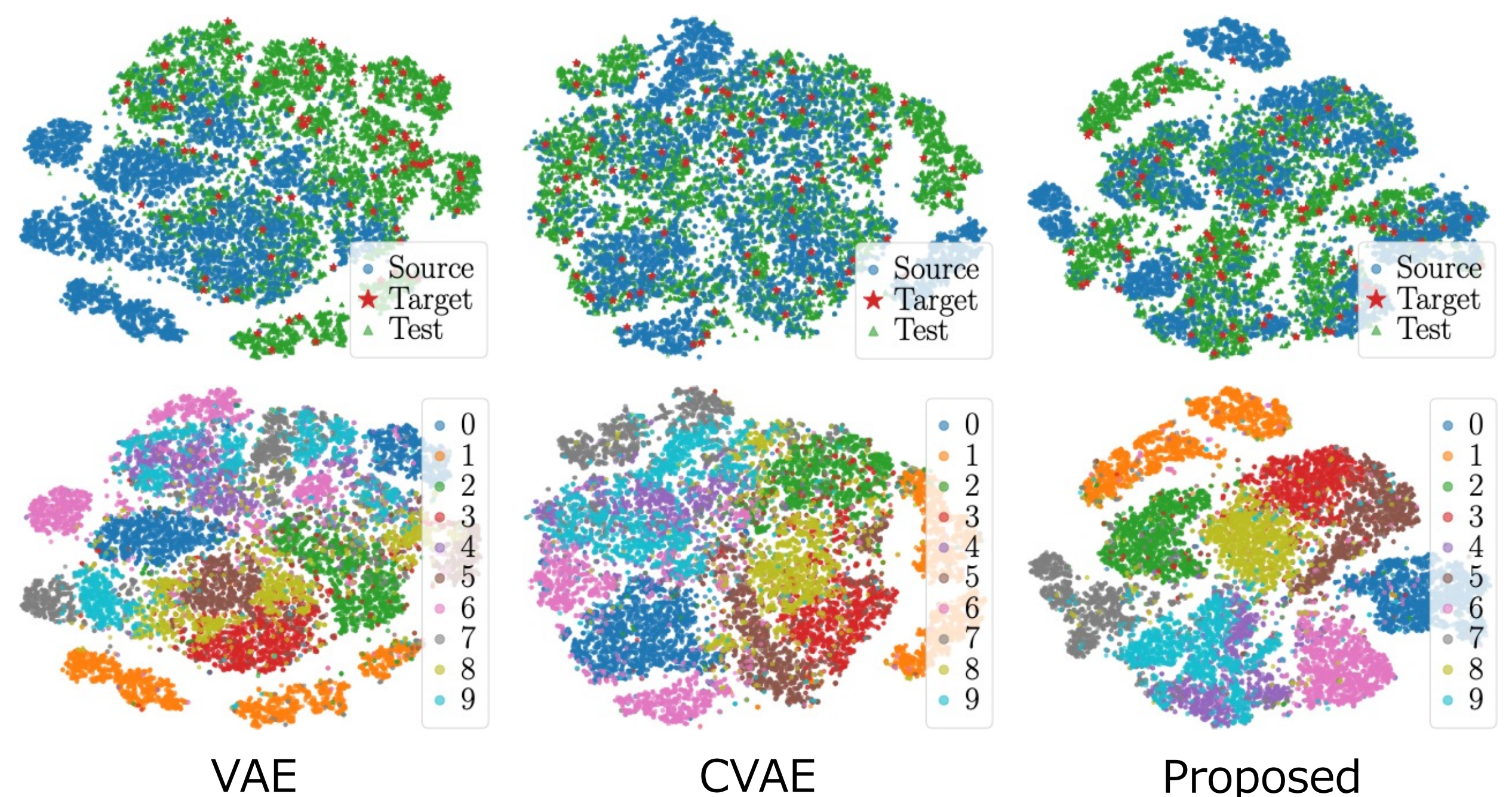
- 最適な事前分布を持つELBO  $\mathcal{F}_{\text{Proposed}}(\theta, \phi)$  は、オリジナルのELBO  $\mathcal{F}_{\text{CVAE}}(\theta, \phi)$  よりも、**常に大きい**か等しい:

$$\mathcal{F}_{\text{Proposed}}(\theta, \phi) = \mathcal{F}_{\text{CVAE}}(\theta, \phi) + D_{KL}(q_{\phi}(\mathbf{z}) \| p(\mathbf{z})) \geq \mathcal{F}_{\text{CVAE}}(\theta, \phi)$$

- つまり、 $\mathcal{F}_{\text{Proposed}}(\theta, \phi)$  は  $\mathcal{F}_{\text{CVAE}}(\theta, \phi)$  と比べて**より良い対数尤度の下界**となっており、良い表現獲得に貢献する

## 5. 実験

### USPS→MNIST における潜在変数の可視化



### 密度推定性能

	VAE	CVAE	Proposed
USPS→MNIST	$-163.25 \pm 2.15$	$-152.32 \pm 1.64$	<b><math>-149.08 \pm 0.86</math></b>
MNIST→USPS	$-235.23 \pm 1.54$	<b><math>-211.18 \pm 0.55</math></b>	$-212.11 \pm 1.48$
Synth→SVHN	$1146.04 \pm 35.65$	$1397.36 \pm 10.89$	<b><math>1430.27 \pm 11.44</math></b>
SVHN→Synth	$760.66 \pm 8.85$	$814.63 \pm 10.09$	<b><math>855.51 \pm 11.41</math></b>

### 下流タスク (クラス分類) の性能

	VAE	CVAE	Proposed
USPS→MNIST	$0.52 \pm 2.15$	$0.53 \pm 0.02$	<b><math>0.68 \pm 0.01</math></b>
MNIST→USPS	$0.64 \pm 0.01$	$0.67 \pm 0.01$	<b><math>0.74 \pm 0.02</math></b>
Synth→SVHN	$0.20 \pm 0.00$	<b><math>0.21 \pm 0.00</math></b>	$0.19 \pm 0.00$
SVHN→Synth	$0.25 \pm 0.01$	$0.25 \pm 0.00$	<b><math>0.26 \pm 0.00</math></b>