Turbo: A Physical-Minded Approach to Generalized Autoencoders

Slava Voloshynovskiy

in collaboration with Guillaume Quétant, Vitaliy Kinach, Mariia Drozdova and Olga Taran




2023, H&N

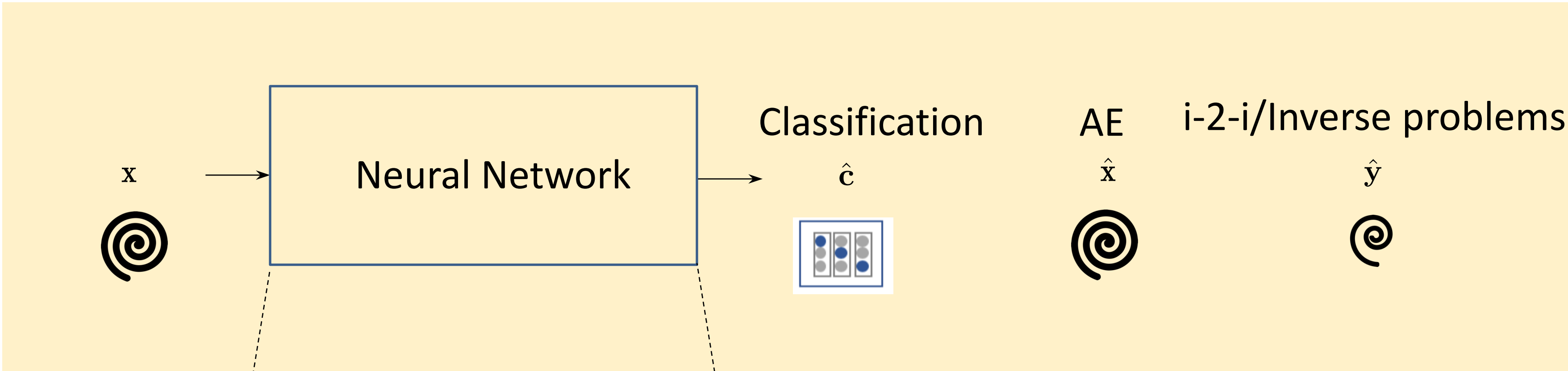
UNIVERSITÉ DE GENÈVE

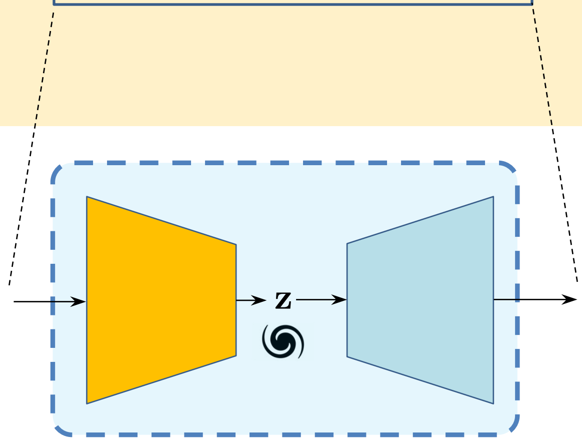
Agenda

- **What is information bottleneck (IBN)?**
- **Generalization of existing methods based on IBN**
  - VAE, InfoVAE, VAE/GAN, BIB-AE
- **Restrictions of IBN**
- **TURBO: physical-driven latent space AE**
- **Generalization based on TURBO**
  - AAE, SR-GAN, pix2pix, CycleGAN
- **Regression problems**
  - HEP translation
  - Hubble-to-Webb translation
  - Inverse problems in physics
- **Conclusions**

What is an information bottleneck (IBN)?

Given a neural network





Latent space can be:
- A single vector/tensor
- Hierarchical (Markov) vectors/tensors
- Multi-vector/Multi-tensor

Deterministic: $f_\phi(\mathbf{x})$ $g_\theta(\mathbf{z})$

Stochastic: $q_\phi(\mathbf{z}|\mathbf{x})$ $p_\theta(\bullet|\mathbf{z})$

FLOWS: $f_\phi(\mathbf{x})$ $g_\theta(\mathbf{z}) = f_\phi^{-1}(\mathbf{z})$

What is an information bottleneck (IBN)?

Definition of Information Bottleneck

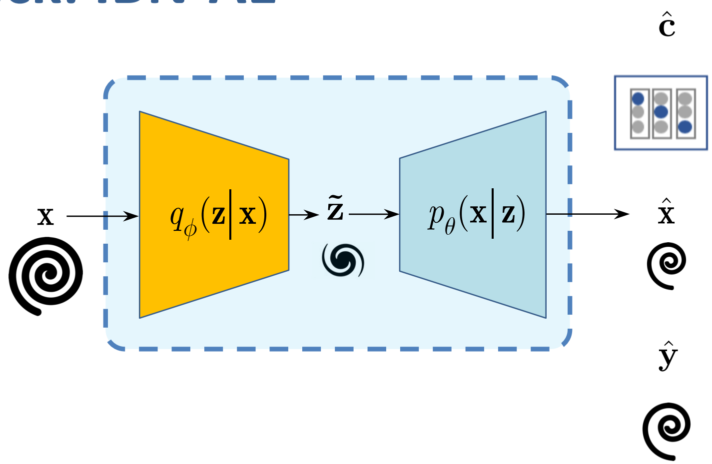
- The Information Bottleneck (IBN) theory is a framework for understanding the trade-off between the amount of information that is preserved in a representation and the amount of "compression" that is achieved

- The **IBN theory proposes** that a good representation is one that preserves **the most relevant information while discarding all irrelevant information for a targeted task**

Agenda

- **What is information bottleneck (IBN)?**
- **IBN based autoencoding**
- **Generalization of existing methods based on IBN**
  - VAE, InfoVAE, VAE/GAN, BIB-AE
- **Restrictions of IBN**
- **TURBO: physical-driven latent space**
- **Generalization based on TURBO**
  - AAE, SR-GAN, pix2pix, CycleGAN
- **Regression problems**
  - HEP translation
  - Hubble-to-Webb translation
  - Inverse problems in physics
- **Conclusions**
- **Open problems**

Information Bottleneck: IBN-AE



Lagrangian formulation

$$(\hat{\phi}, \hat{\theta}) = \underset{\phi,\theta}{\arg\min} \mathcal{L}_{\mathrm{IBN-AE}}(\phi, \theta)$$

$$\mathcal{L}_{\text{IBN-AE}}(\phi, \theta) = I_\phi(\mathbf{X}; \mathbf{Z}) - \beta I_{\phi,\theta}(\mathbf{Z}; \mathbf{X})$$

$$I_\phi(\mathbf{X}; \mathbf{Z}) = \mathbb{E}_{q_\phi(\mathbf{x},\mathbf{z})}\left[\log \frac{q_\phi(\mathbf{z}|\mathbf{x})}{\tilde{q}_\phi(\mathbf{z})}\right]$$

$$I_{\phi,\theta}(\mathbf{Z}; \mathbf{X}) = \mathbb{E}_{q_\phi(\mathbf{x},\mathbf{z})}\left[\log \frac{p_\theta(\mathbf{x}|\mathbf{z})}{p_{\mathbf{x}}(\mathbf{x})}\right]$$

Variational decomposition of terms

$$I_\phi(\mathbf{X};\mathbf{Z}) = \mathbb{E}_{q_\phi(\mathbf{x},\mathbf{z})}\left[\log \frac{q_\phi(\mathbf{z}|\mathbf{x})}{\tilde{q}_\phi(\mathbf{z})}\frac{p_\mathbf{z}(\mathbf{z})}{p_\mathbf{z}(\mathbf{z})}\right] = \underbrace{\mathbb{E}_{p_\mathbf{x}(\mathbf{x})}\left[D_{\mathrm{KL}}\left(q_\phi(\mathbf{z}|\mathbf{X}=\mathbf{x})\|p_\mathbf{z}(\mathbf{z})\right)\right]}_{\mathcal{D}_{\mathbf{z}|\mathbf{x}}} - \underbrace{D_{\mathrm{KL}}\left(\tilde{q}_\phi(\mathbf{z})\|p_\mathbf{z}(\mathbf{z})\right)}_{\mathcal{D}_{\tilde{\mathbf{z}}}}$$

$$I_\phi(\mathbf{Z};\mathbf{X}) = \mathbb{E}_{q_\phi(\mathbf{x},\mathbf{z})}\left[\log \frac{q_\phi(\mathbf{x}|\mathbf{z})}{p_\mathbf{x}(\mathbf{x})}\frac{p_\theta(\mathbf{x}|\mathbf{z})}{p_\theta(\mathbf{x}|\mathbf{z})}\right] = \underbrace{\mathbb{E}_{q_\phi(\mathbf{x},\mathbf{z})}\left[\log \frac{p_\theta(\mathbf{x}|\mathbf{z})}{p_\mathbf{x}(\mathbf{x})}\right]}_{I_{\phi,\theta}(\mathbf{Z};\mathbf{X})} + \underbrace{\mathbb{E}_{q_\phi(\mathbf{x},\mathbf{z})}\left[\frac{q_\phi(\mathbf{x}|\mathbf{z})}{p_\theta(\mathbf{x}|\mathbf{z})}\right]}_{D_{\mathrm{KL}}(q_\phi(\mathbf{x}|\mathbf{z})||p_\theta(\mathbf{x}|\mathbf{z}))\geq 0} \geq I_{\phi,\theta}(\mathbf{Z};\mathbf{X})$$

$$\begin{aligned} I_{\phi,\theta}(\mathbf{Z};\mathbf{X}) &= \mathbb{E}_{q_\phi(\mathbf{x},\mathbf{z})}\left[\log \frac{p_\theta(\mathbf{x}|\mathbf{z})}{p_\mathbf{x}(\mathbf{x})}\frac{\hat{p}_\theta(\mathbf{x})}{\hat{p}_\theta(\mathbf{x})}\right] \\ &= -\mathbb{E}_{p_\mathbf{x}(\mathbf{x})}\left[\log \hat{p}_\theta(\mathbf{x})\right] - \mathbb{E}_{p_\mathbf{x}(\mathbf{x})}\left[\log \frac{p_\mathbf{x}(\mathbf{x})}{\hat{p}_\theta(\mathbf{x})}\right] + \mathbb{E}_{p_\mathbf{x}(\mathbf{x})}\left[\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\log p_\theta(\mathbf{x}|\mathbf{z})\right]\right] \\ &= \underbrace{H\left(p_\mathbf{x}(\mathbf{x});\hat{p}_\theta(\mathbf{x})\right)}_{\geq 0} - D_{KL}\left(p_\mathbf{x}(\mathbf{x})\|\hat{p}_\theta(\mathbf{x})\right) - H_{\phi,\theta}(\mathbf{X}|\mathbf{Z}) \end{aligned}$$

$$I^{\mathrm{L}}_{\phi,\theta}(\mathbf{Z};\mathbf{X}) \triangleq \underbrace{-H_{\phi,\theta}(\mathbf{X}|\mathbf{Z})}_{\mathcal{L}(\mathbf{x},\hat{\mathbf{x}})} - \underbrace{D_{\mathrm{KL}}\left(p_\mathbf{x}(\mathbf{x})\|\hat{p}_\theta(\mathbf{x})\right)}_{\mathcal{D}_{\hat{\mathbf{x}}}}$$

$$p_\theta(\mathbf{x}|\mathbf{z}) \propto \exp\left(-\lambda\left\|\mathbf{x}-g_\theta(\mathbf{z})\right\|_1\right)$$

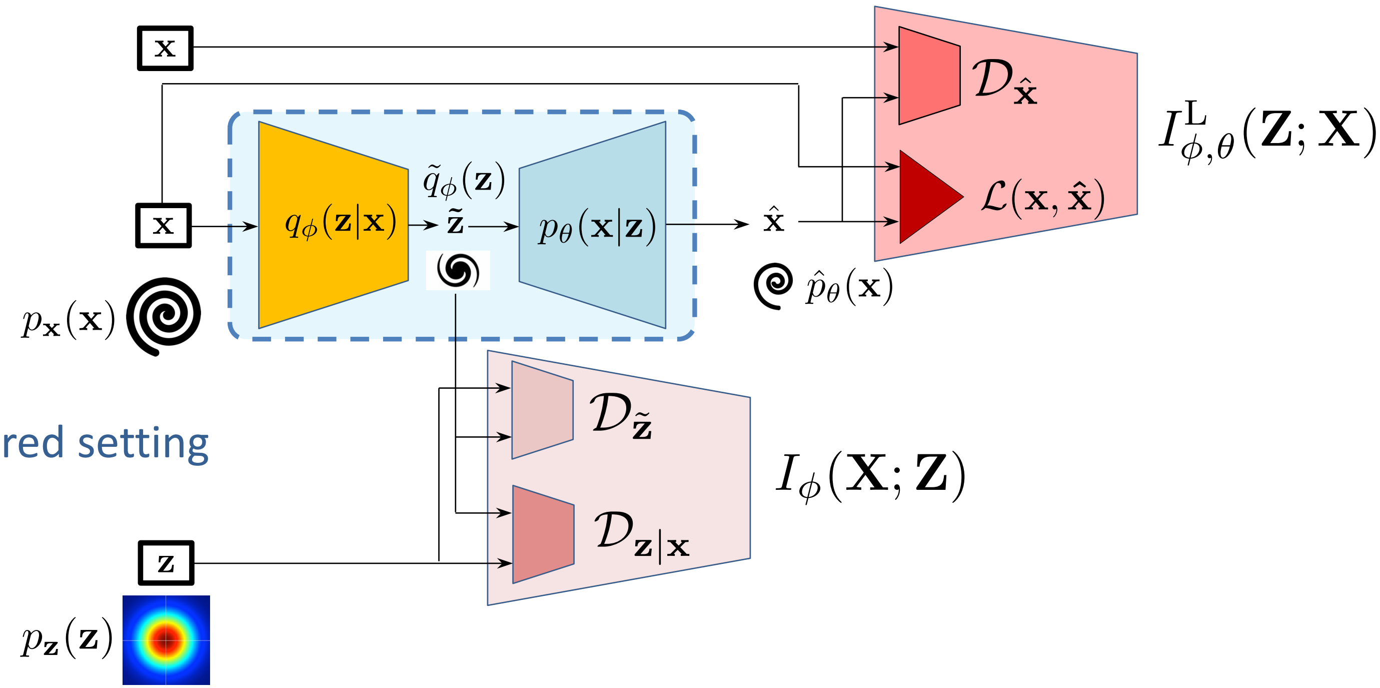
$$I^{\mathrm{L}}_{\phi,\theta}(\mathbf{Z};\mathbf{X}) = -\lambda\underbrace{\mathbb{E}_{p_\mathbf{x}(\mathbf{x})}\left[\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\|\mathbf{x}-g_\theta(\mathbf{z})\|_1\right]\right]}_{\mathcal{L}(\mathbf{x},\hat{\mathbf{x}})} - \underbrace{D_{\mathrm{KL}}\left(p_\mathbf{x}(\mathbf{x})\|\hat{p}_\theta(\mathbf{x})\right)}_{\mathcal{D}_{\hat{\mathbf{x}}}}$$

Bounded Information Bottleneck (BIB) Autoencoder [BIB-AE]

$$\mathcal{L}_{\text{BIB-AE}}(\phi,\theta) = I_{\phi}(\mathbf{X};\mathbf{Z}) - \beta I^{\text{L}}_{\theta,\phi}(\mathbf{Z};\mathbf{X})$$

$$I_{\phi}(\mathbf{X};\mathbf{Z}) = \mathbb{E}_{q_{\phi}(\mathbf{x},\mathbf{z})}\left[\log \frac{q_{\phi}(\mathbf{z}|\mathbf{x})}{\tilde{q}_{\phi}(\mathbf{z})} \frac{p_{\mathbf{z}}(\mathbf{z})}{p_{\mathbf{z}}(\mathbf{z})}\right] = \underbrace{\mathbb{E}_{p_{\mathbf{x}}(\mathbf{x})}\left[D_{\text{KL}}\left(q_{\phi}(\mathbf{z}|\mathbf{X}=\mathbf{x}) \| p_{\mathbf{z}}(\mathbf{z})\right)\right]}_{\mathcal{D}_{\mathbf{z}|\mathbf{x}}} - \underbrace{D_{\text{KL}}\left(\tilde{q}_{\phi}(\mathbf{z}) \| p_{\mathbf{z}}(\mathbf{z})\right)}_{\mathcal{D}_{\tilde{\mathbf{z}}}}$$

$$I^{\text{L}}_{\phi,\theta}(\mathbf{Z};\mathbf{X}) \triangleq \underbrace{-H_{\phi,\theta}(\mathbf{X}|\mathbf{Z})}_{\mathcal{L}(\mathbf{x},\hat{\mathbf{x}})} - \underbrace{D_{\text{KL}}\left(p_{\mathbf{x}}(\mathbf{x}) \| \hat{p}_{\theta}(\mathbf{x})\right)}_{\mathcal{D}_{\hat{\mathbf{x}}}}$$



Unpaired setting

Voloshynovskiy, Kondah, Rezaeifar, Taran, Hotolyak, Rezende, Information bottleneck through variational glasses, 2019.

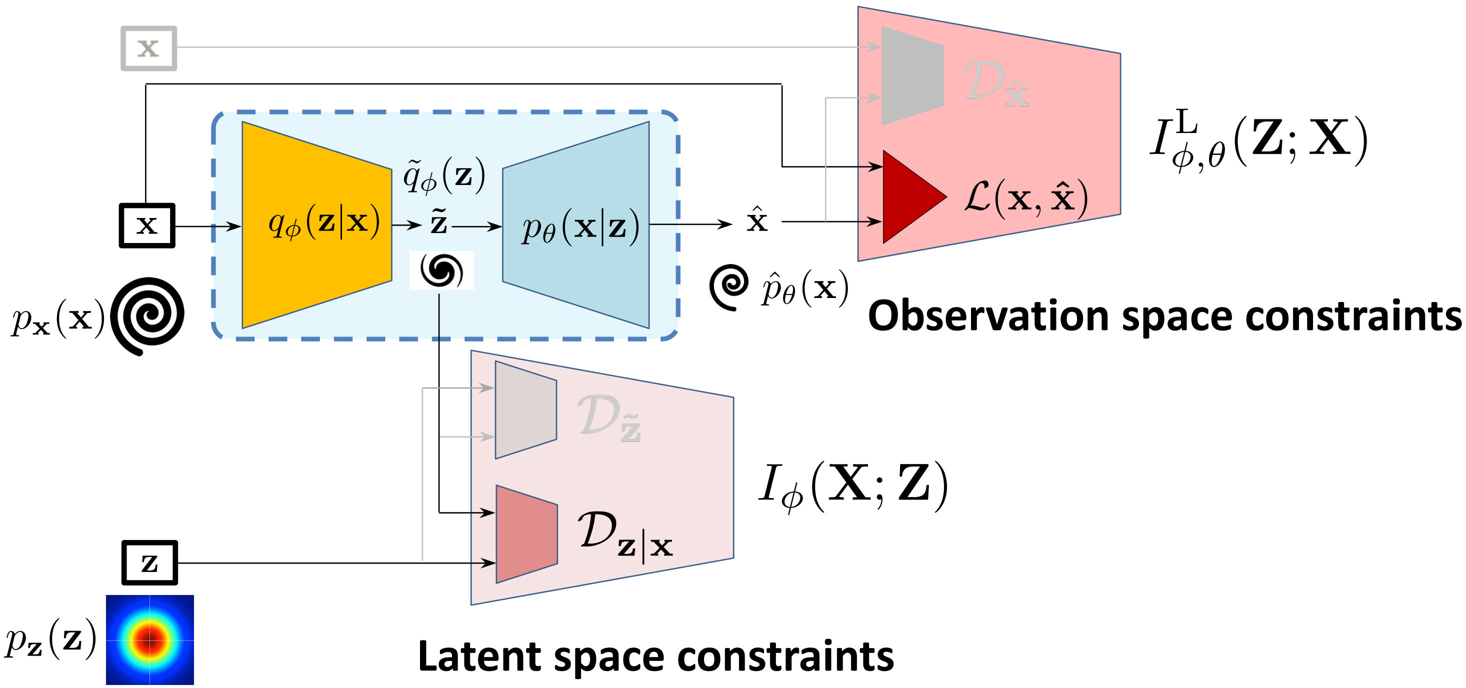
Agenda

- What is information bottleneck (IBN)?
- IBN based autoencoding
- **Generalization of existing methods based on IBN**
  - VAE, InfoVAE, VAE/GAN
- Restrictions of IBN
- TURBO: physical-driven latent space
- Generalization based on TURBO
  - AAE, SR-GAN, pix2pix, CycleGAN, Probabilistic AE
- Regression problems
  - HEP translation
  - Hubble-to-Webb translation
  - Inverse problems in physics
- Conclusions

IBN: generalization of existing schemes

VAE and $\beta -$VAE: Variational Autoencoder

$$\mathcal{L}_{\beta-\mathrm{VAE}}(\phi,\theta)=\underbrace{\mathbb{E}_{p_{\mathbf{x}}(\mathbf{x})}\left[D_{\mathrm{KL}}\left(q_{\phi}(\mathbf{z}|\mathbf{X}=\mathbf{x})\|p_{\mathbf{z}}(\mathbf{z})\right)\right]}_{\mathcal{D}_{\mathbf{z}|\mathbf{x}}}-\beta\,\underbrace{\mathbb{E}_{p_{\mathbf{x}}(\mathbf{x})}\left[\mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})}\left[\log p_{\theta}(\mathbf{x}|\mathbf{z})\right]\right]}_{\mathcal{L}(\mathbf{x},\hat{\mathbf{x}})}$$

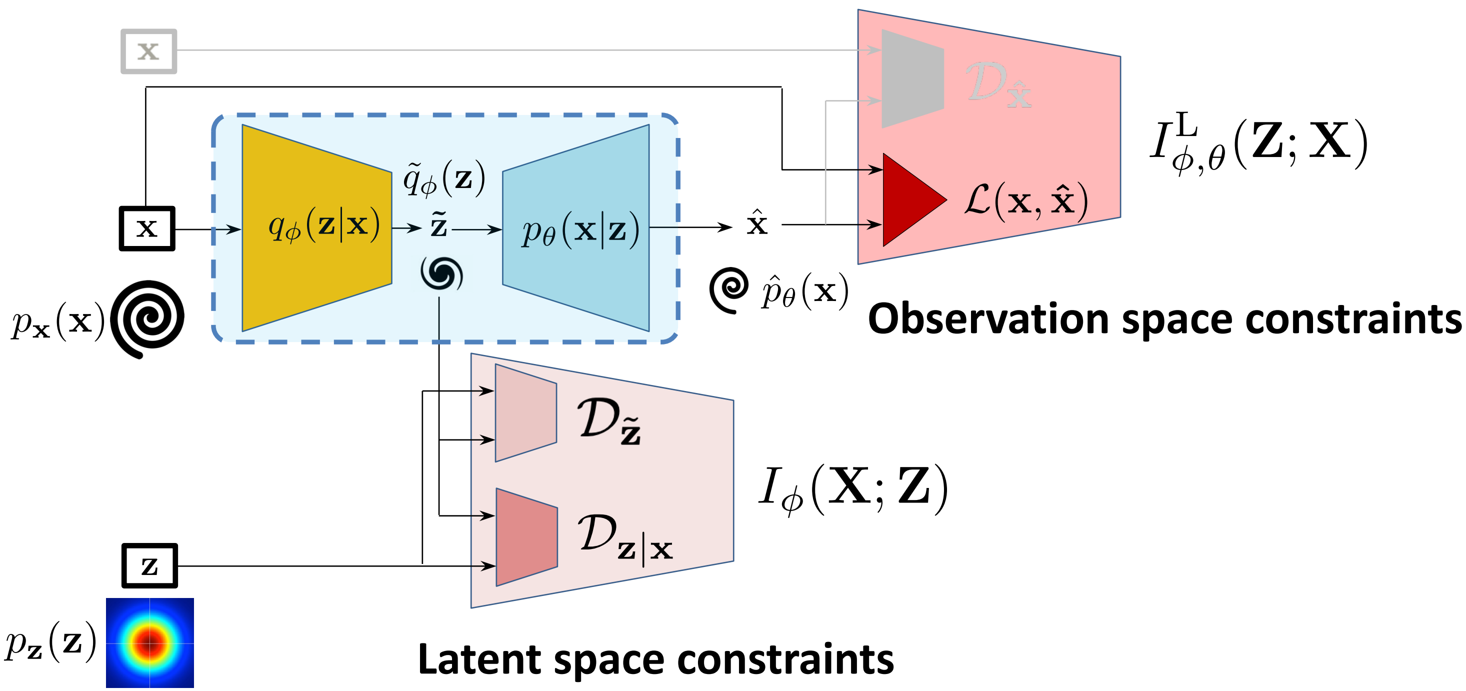


Kingma and Welling. Auto-encoding variational Bayes, 2014
Rezende, Mohamed, and Wierstra. Stochastic backpropagation and approximate inference in deep generative models., 2014
Higgins, Matthey, Pal, Burgess, Glorot, Botvinick, Mohamed, and Lerchner. beta-VAE: Learning basic visual concepts with a constrained variational framework, 2017

BIB: generalization of existing schemes

InfoVAE:

$$\mathcal{L}_{\text{InfoVAE}}(\phi,\theta) = I_\phi(\mathbf{X};\mathbf{Z}) - \beta \underbrace{\mathbb{E}_{p_{\mathbf{x}}(\mathbf{x})}\left[\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\log p_\theta(\mathbf{x}|\mathbf{z})\right]\right]}_{\mathcal{L}(\mathbf{x},\hat{\mathbf{x}})}$$

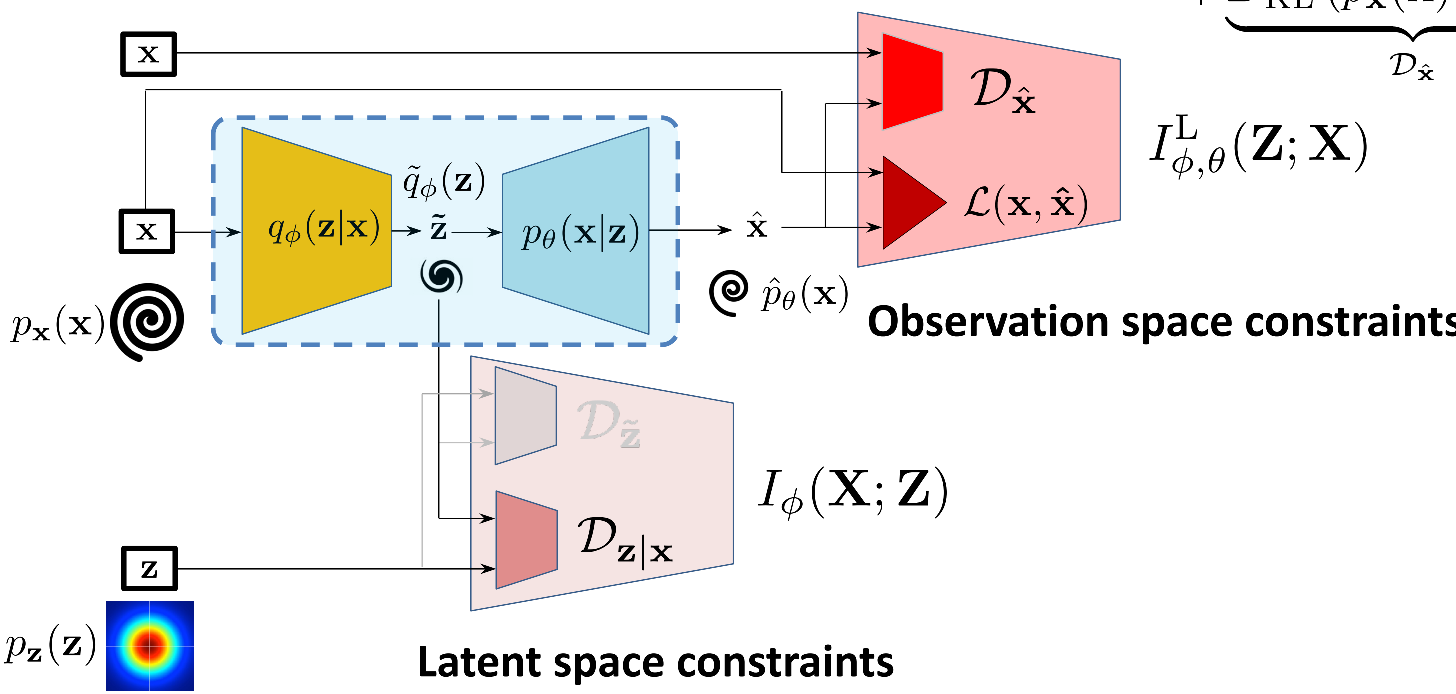


Zhao, Song, and Ermon. InfoVAE: Information maximizing variational autoencoders. 2017

BIB: generalization of existing schemes

VAE/GAN

$$\mathcal{L}_{\mathrm{VAE/GAN}}(\phi,\theta) = \underbrace{\mathbb{E}_{p_{\mathbf{x}}(\mathbf{x})}\left[D_{\mathrm{KL}}\left(q_\phi(\mathbf{z}|\mathbf{X}=\mathbf{x})\|p_{\mathbf{z}}(\mathbf{z})\right)\right]}_{\mathcal{D}_{\mathbf{z}|\mathbf{x}}} - \beta \underbrace{\mathbb{E}_{p_{\mathbf{x}}(\mathbf{x})}\left[\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\log p_\theta(\mathbf{x}|\mathbf{z})\right]\right]}_{\mathcal{L}(\mathbf{x},\hat{\mathbf{x}})} + \underbrace{D_{\mathrm{KL}}\left(p_{\mathbf{x}}(\mathbf{x})\|\hat{p}_\theta(\mathbf{x})\right)}_{\mathcal{D}_{\hat{\mathbf{x}}}}$$



Larsen, Sønderby, Larochelle, and Winther. Autoencoding beyond pixels using a learned similarity metric. 2015

Restrictions of IBN

- The information bottleneck (IB) theory posits that a neural network can be trained **to extract the most relevant information from its inputs** while **discarding task irrelevant information**

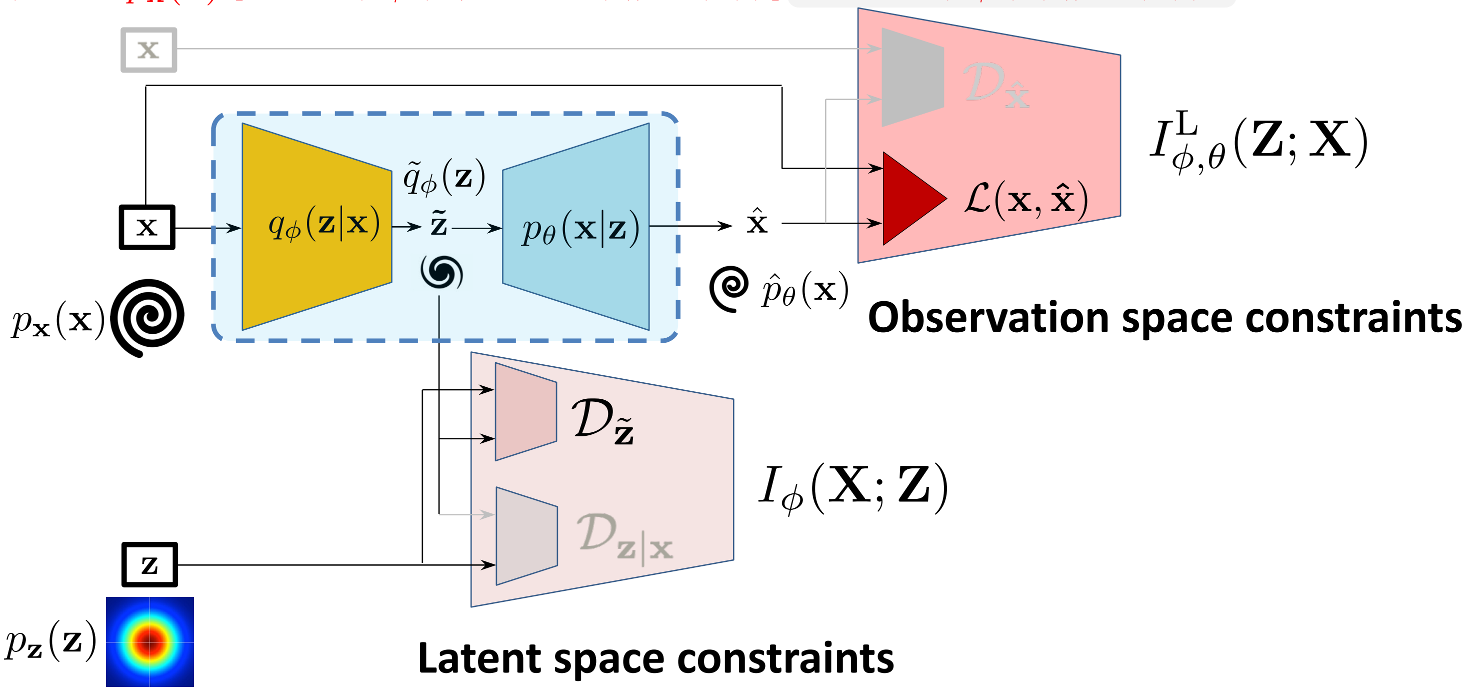
- However, it has **a number of restrictions**:
  - IBN does not have **any meaningful latent space** that would correspond to the physics of underlying phenomena
  - The latent space **does not correspond to typical physical observation or measurement models**
  - IBN **does not explain** systems such as AAE, CycleGAN, Probabilistic AE and many others
  - IBN **does not envision an optimization** of detectors, sensors and antennas as “physical encoders”

BIB: generalization of existing schemes

AAE: Adversarial Autoencoder – Not a case!

$$\mathcal{L}_{\mathrm{AAE}}(\phi,\theta) = D_{\mathrm{KL}}\left(\tilde{q}_\phi(\mathbf{z}) \| p_{\mathbf{z}}(\mathbf{z})\right) - \beta \mathbb{E}_{p_{\mathbf{x}}(\mathbf{x})}\left[\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\log p_\theta(\mathbf{x}|\mathbf{z})\right]\right]$$

$$I_\phi(\mathbf{X};\mathbf{Z}) = \mathbb{E}_{p_{\mathbf{x}}(\mathbf{x})}\left[D_{\mathrm{KL}}\left(q_\phi(\mathbf{z}|\mathbf{X}=\mathbf{x}) \| p_{\mathbf{z}}(\mathbf{z})\right)\right] - D_{\mathrm{KL}}\left(\tilde{q}_\phi(\mathbf{z}) \| p_{\mathbf{z}}(\mathbf{z})\right)$$



Makhzani, Shlens, Jaitly, Goodfellow, and Frey. Adversarial autoencoders, 2015

Agenda

- What is information bottleneck (IBN)?
- IBN based autoencoding
- Generalization of existing methods based on IBN
  - VAE, InfoVAE, VAE/GAN, BIB-AE
- Restrictions of IBN
- **TURBO: physical-driven latent space AE**
- Generalization based on TURBO
  - AAE, SR-GAN, pix2pix, CycleGAN, Probabilistic AE
- Regression problems
  - HEP translation
  - Hubble-to-Webb translation
  - Inverse problems in physics
- Conclusions

TURBO: generalized translation problem formulation

Main difference with IBN:

- Fundamental IBN task-irrelevance concept at the encoder is replaced by a concept of satisfaction of “relevance” to physical constraints on the latent space

Main consequences

- Impose meaningful physical priors on latent space
- Incorporate a fact the that data and latent space representation can be dependent
- Consider all options of paired, unpaired and partially paired data
- Consider two-way propagation of information (TURBO):
  - Encoding (generation) from both data and latent spaces
    - Link to CycleGAN-like architectures

TURBO: generalized translation problem formulation

IBN

$$(\hat{\phi}, \hat{\theta}) = \arg\min_{\phi,\theta} \mathcal{L}_{\mathrm{IBN-AE}}(\phi, \theta)$$

$$\mathcal{L}_{\text{IBN-AE}}(\phi, \theta) = I_{\phi}(\mathbf{X}; \mathbf{Z}) - \beta I_{\phi,\theta}(\mathbf{Z}; \mathbf{X})$$

TURBO

$$(\hat{\phi}, \hat{\theta}) = \arg\max_{\phi,\theta} \mathcal{L}^{\mathrm{Direct}}(\phi, \theta)$$

$$\mathcal{L}^{\mathrm{Direct}}(\phi, \theta) = \mathcal{I}^{\mathrm{z}}_{\phi}(\mathbf{X}; \mathbf{Z}) + \lambda_1 \mathcal{I}^{\mathbf{x}}_{\phi,\theta}(\mathbf{Z}; \mathbf{X})$$

Link between data and latent space

unpaired

$p_{\mathbf{x}}(\mathbf{x})$ $\quad p_{\mathbf{z}}(\mathbf{z})$

paired

$p(\mathbf{x}, \mathbf{z})$

Data encoding/generation

one-way

$$\mathbf{X} \underset{\text{encoder}}{\xrightarrow{q_{\phi}(\mathbf{z}|\mathbf{x})}} \tilde{\mathbf{Z}} \underset{\text{decoder}}{\xrightarrow{p_{\theta}(\mathbf{x}|\mathbf{z})}} \hat{\mathbf{X}}$$

two-way

$$\mathbf{X} \underset{\text{encoder}}{\xrightarrow{q_{\phi}(\mathbf{z}|\mathbf{x})}} \tilde{\mathbf{Z}} \underset{\text{decoder}}{\xrightarrow{p_{\theta}(\mathbf{x}|\mathbf{z})}} \hat{\mathbf{X}}$$

$$\mathbf{Z} \underset{\text{decoder}}{\xrightarrow{p_{\theta}(\mathbf{x}|\mathbf{z})}} \tilde{\mathbf{X}} \underset{\text{encoder}}{\xrightarrow{q_{\phi}(\mathbf{z}|\mathbf{x})}} \hat{\mathbf{Z}}$$

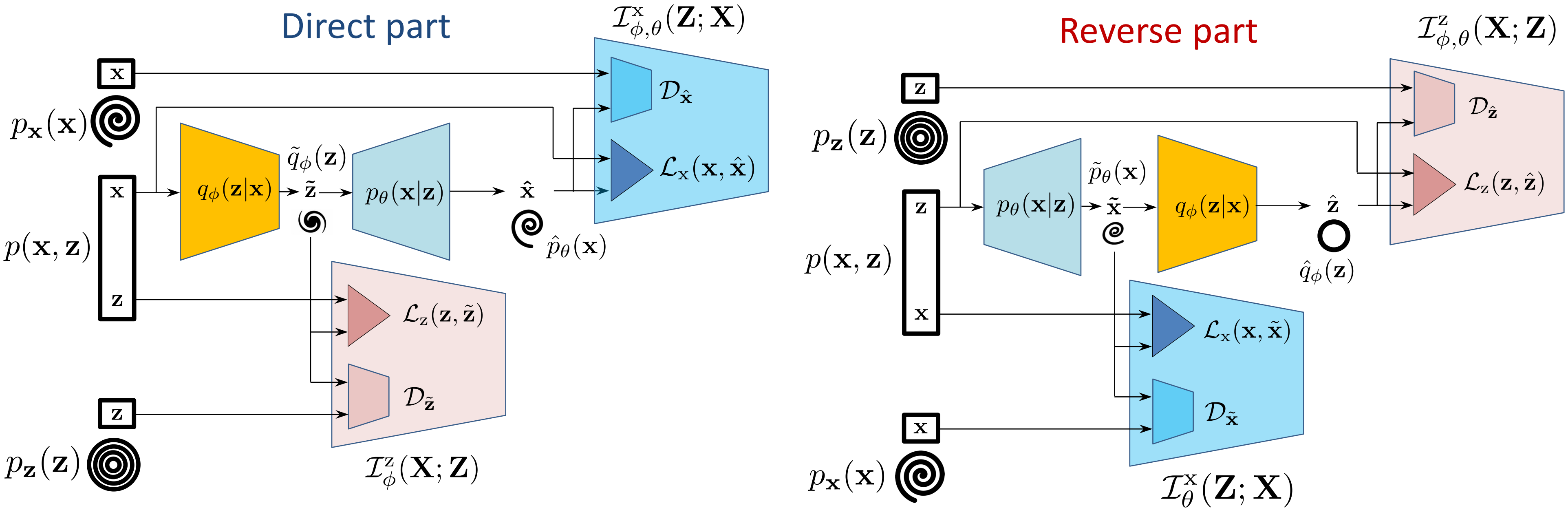
Type of latent space

“virtual” latent space

physically meanigful latent space

TURBO: generalized translation problem formulation

TURBO



Lagrangian formulation

$$(\hat{\phi},\hat{\theta}) = \arg\max_{\phi,\theta} \mathcal{L}_{\mathrm{TURBO}}(\phi,\theta)$$

$$\mathcal{L}_{\mathrm{TURBO}}(\phi,\theta) = \mathcal{L}^{\mathrm{Direct}}(\phi,\theta) + \alpha\mathcal{L}^{\mathrm{Reverse}}(\phi,\theta)$$

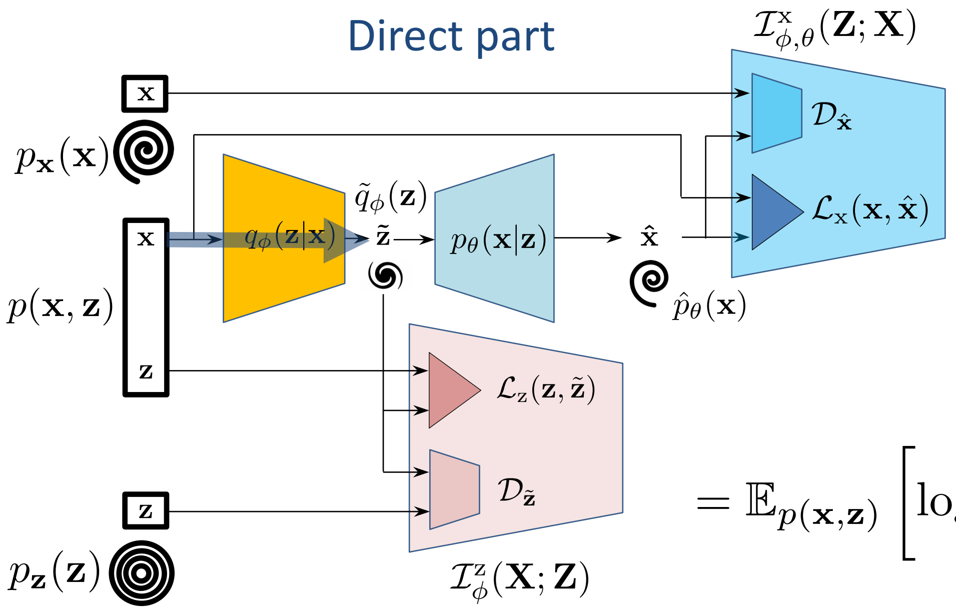
$$\mathcal{L}^{\mathrm{Direct}}(\phi,\theta) = \mathcal{I}^{\mathrm{z}}_{\phi}(\mathbf{X};\mathbf{Z}) + \lambda_1\mathcal{I}^{\mathrm{x}}_{\phi,\theta}(\mathbf{Z};\mathbf{X})$$

$$\mathcal{L}^{\mathrm{Reverse}}(\phi,\theta) = \mathcal{I}^{\mathrm{x}}_{\theta}(\mathbf{Z};\mathbf{X}) + \lambda_2\mathcal{I}^{\mathrm{z}}_{\phi,\theta}(\mathbf{X};\mathbf{Z})$$

G. Quétant, M. Drozdova, V. Kinakh, T. Golling, and S. Voloshynovskiy, "Turbo-Sim: a generalised generative model with a physical latent space." NeurIPS, ML4PhysicalSciences2021.

TURBO: generalized translation problem formulation

Direct part



Encoder loss

$$I(\mathbf{X};\mathbf{Z}) = \mathbb{E}_{p(\mathbf{x},\mathbf{z})}\left[\log \frac{p(\mathbf{z}|\mathbf{x})}{p_{\mathbf{z}}(\mathbf{z})}\frac{q_\phi(\mathbf{z}|\mathbf{x})}{q_\phi(\mathbf{z}|\mathbf{x})}\right]$$

$$\geq \mathbb{E}_{p(\mathbf{x},\mathbf{z})}\left[\log \frac{q_\phi(\mathbf{z}|\mathbf{x})}{p_{\mathbf{z}}(\mathbf{z})}\right]$$

$$= \mathbb{E}_{p(\mathbf{x},\mathbf{z})}\left[\log \frac{q_\phi(\mathbf{z}|\mathbf{x})}{p_{\mathbf{z}}(\mathbf{z})}\frac{\tilde{q}_\phi(\mathbf{z})}{\tilde{q}_\phi(\mathbf{z})}\right]$$

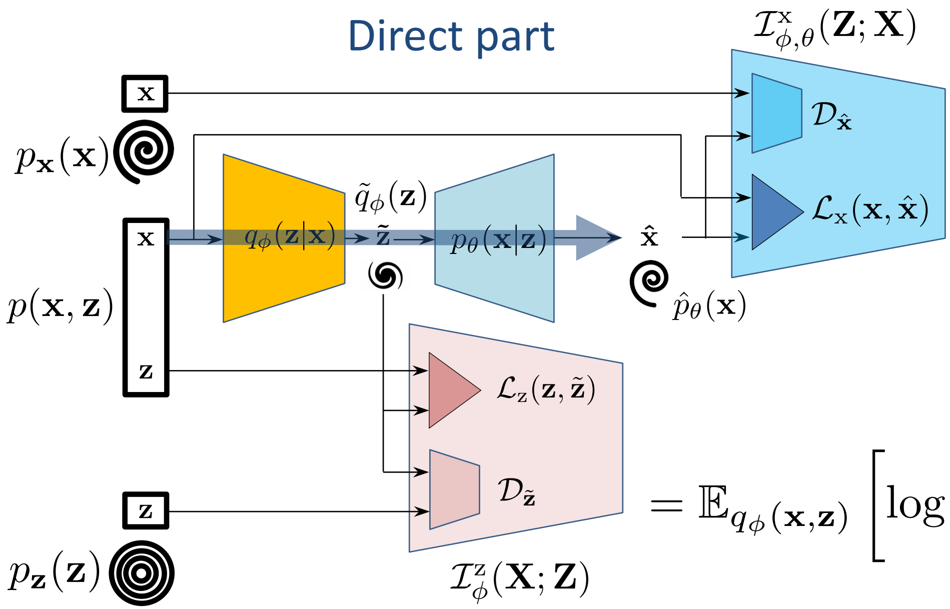
$$= \mathbb{E}_{p(\mathbf{x},\mathbf{z})}\left[\log q_\phi(\mathbf{z}|\mathbf{x})\right] - \mathbb{E}_{p(\mathbf{x},\mathbf{z})}\left[\log \frac{p_{\mathbf{z}}(\mathbf{z})}{\tilde{q}_\phi(\mathbf{z})}\right] - \mathbb{E}_{p(\mathbf{x},\mathbf{z})}\left[\log \tilde{q}_\phi(\mathbf{z})\right]$$

$$= \mathbb{E}_{p(\mathbf{x},\mathbf{z})}\left[\log q_\phi(\mathbf{z}|\mathbf{x})\right] - D_{\mathrm{KL}}\left(p_{\mathbf{z}}(\mathbf{z})\|\tilde{q}_\phi(\mathbf{z})\right) + H\left(p_{\mathbf{z}}(\mathbf{z});\tilde{q}_\phi(\mathbf{z})\right)$$

$$\geq \underbrace{\mathbb{E}_{p(\mathbf{x},\mathbf{z})}\left[\log q_\phi(\mathbf{z}|\mathbf{x})\right]}_{\mathcal{L}_{\mathbf{z}}(\mathbf{z},\tilde{\mathbf{z}})} - \underbrace{D_{\mathrm{KL}}\left(p_{\mathbf{z}}(\mathbf{z})\|\tilde{q}_\phi(\mathbf{z})\right)}_{\mathcal{D}_{\tilde{\mathbf{z}}}}$$

$$=: \mathcal{I}^{\mathrm{z}}_{\phi}(\mathbf{X};\mathbf{Z})$$

TURBO: generalized translation problem formulation



Decoder loss

$$I_\phi(\mathbf{Z};\mathbf{X}) = \mathbb{E}_{q_\phi(\mathbf{x},\mathbf{z})}\left[\log \frac{q_\phi(\mathbf{x}|\mathbf{z})}{p_{\mathbf{x}}(\mathbf{x})}\frac{p_\theta(\mathbf{x}|\mathbf{z})}{p_\theta(\mathbf{x}|\mathbf{z})}\right]$$

$$\geq \mathbb{E}_{q_\phi(\mathbf{x},\mathbf{z})}\left[\log \frac{p_\theta(\mathbf{x}|\mathbf{z})}{p_{\mathbf{x}}(\mathbf{x})}\right]$$

$$= \mathbb{E}_{q_\phi(\mathbf{x},\mathbf{z})}\left[\log \frac{p_\theta(\mathbf{x}|\mathbf{z})}{p_{\mathbf{x}}(\mathbf{x})}\frac{\hat{p}_\theta(\mathbf{x})}{\hat{p}_\theta(\mathbf{x})}\right]$$

$$= \mathbb{E}_{q_\phi(\mathbf{x},\mathbf{z})}\left[\log p_\theta(\mathbf{x}|\mathbf{z})\right] - \mathbb{E}_{q_\phi(\mathbf{x},\mathbf{z})}\left[\log \frac{p_{\mathbf{x}}(\mathbf{x})}{\hat{p}_\theta(\mathbf{x})}\right] - \mathbb{E}_{q_\phi(\mathbf{x},\mathbf{z})}\left[\log \hat{p}_\theta(\mathbf{x})\right]$$

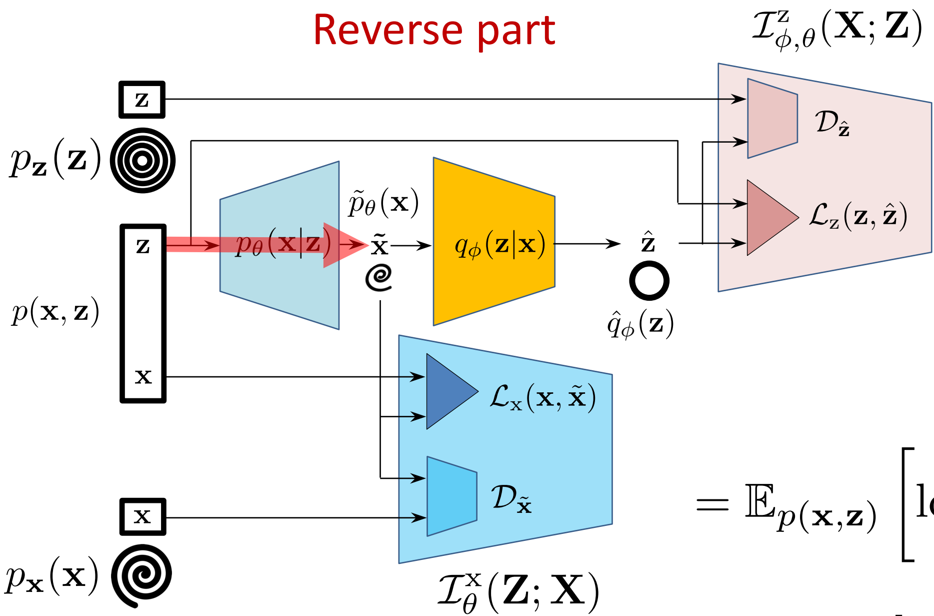
$$= \mathbb{E}_{q_\phi(\mathbf{x},\mathbf{z})}\left[\log p_\theta(\mathbf{x}|\mathbf{z})\right] - \mathbb{E}_{q_\phi(\mathbf{x},\mathbf{z})}\left[\log \frac{p_{\mathbf{x}}(\mathbf{x})}{\hat{p}_\theta(\mathbf{x})}\right] + H\left(p_{\mathbf{x}}(\mathbf{x});\hat{p}_\theta(\mathbf{x})\right)$$

$$\geq \mathbb{E}_{q_\phi(\mathbf{x},\mathbf{z})}\left[\log p_\theta(\mathbf{x}|\mathbf{z})\right] - \mathbb{E}_{q_\phi(\mathbf{x},\mathbf{z})}\left[\log \frac{p_{\mathbf{x}}(\mathbf{x})}{\hat{p}_\theta(\mathbf{x})}\right]$$

$$= \mathbb{E}_{p_{\mathbf{x}}(\mathbf{x})}\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\log p_\theta(\mathbf{x}|\mathbf{z})\right] - D_{\mathrm{KL}}\left(p_{\mathbf{x}}(\mathbf{x})\|\hat{p}_\theta(\mathbf{x})\right)$$

$$=: \mathcal{I}^{\mathbf{x}}_{\phi,\theta}(\mathbf{Z};\mathbf{X})$$

TURBO: generalized translation problem formulation



Decoder loss

$$I(\mathbf{X};\mathbf{Z}) = \mathbb{E}_{p(\mathbf{x},\mathbf{z})}\left[\log \frac{p(\mathbf{x}|\mathbf{z})}{p_{\mathbf{x}}(\mathbf{x})}\frac{p_\theta(\mathbf{x}|\mathbf{z})}{p_\theta(\mathbf{x}|\mathbf{z})}\right]$$

$$\geq \mathbb{E}_{p(\mathbf{x},\mathbf{z})}\left[\log \frac{p_\theta(\mathbf{x}|\mathbf{z})}{p_{\mathbf{x}}(\mathbf{x})}\right]$$

$$= \mathbb{E}_{p(\mathbf{x},\mathbf{z})}\left[\log \frac{p_\theta(\mathbf{x}|\mathbf{z})}{p_{\mathbf{x}}(\mathbf{x})}\frac{\tilde{p}_\theta(\mathbf{x})}{\tilde{p}_\theta(\mathbf{x})}\right]$$

$$= \mathbb{E}_{p(\mathbf{x},\mathbf{z})}\left[\log p_\theta(\mathbf{x}|\mathbf{z})\right] - \mathbb{E}_{p(\mathbf{x},\mathbf{z})}\left[\log \frac{p_{\mathbf{x}}(\mathbf{x})}{\tilde{p}_\theta(\mathbf{x})}\right] - \mathbb{E}_{p(\mathbf{x},\mathbf{z})}\left[\log \tilde{p}_\theta(\mathbf{x})\right]$$

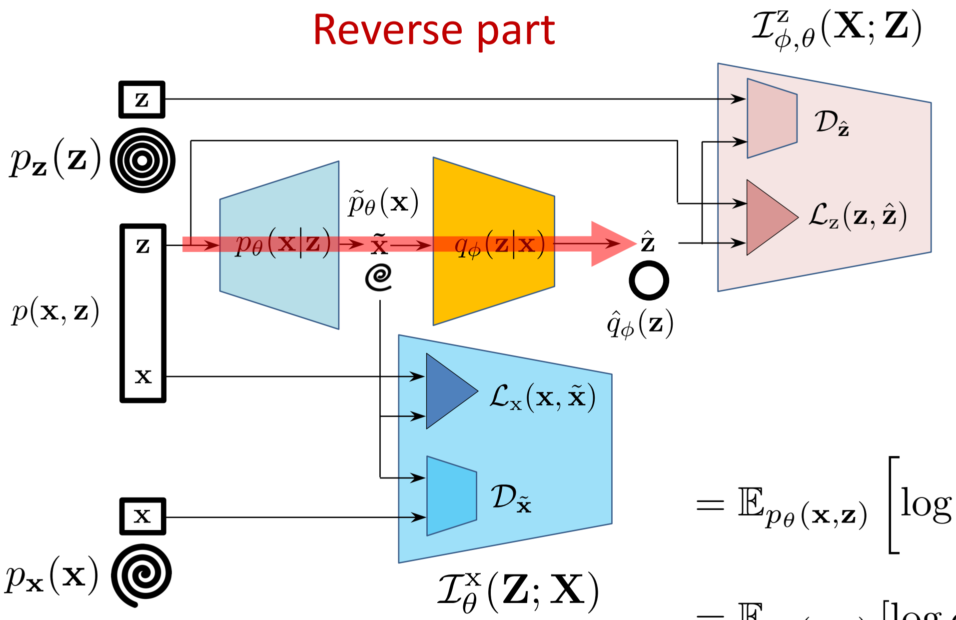
$$= \mathbb{E}_{p(\mathbf{x},\mathbf{z})}\left[\log p_\theta(\mathbf{x}|\mathbf{z})\right] - D_{\mathrm{KL}}\left(p_{\mathbf{x}}(\mathbf{x})\|\tilde{p}_\theta(\mathbf{x})\right) + H\left(p_{\mathbf{x}}(\mathbf{x});\tilde{p}_\theta(\mathbf{x})\right)$$

$$\geq \mathbb{E}_{p(\mathbf{x},\mathbf{z})}\left[\log p_\theta(\mathbf{x}|\mathbf{z})\right] - D_{\mathrm{KL}}\left(p_{\mathbf{x}}(\mathbf{x})\|\tilde{p}_\theta(\mathbf{x})\right)$$

$$=: \mathcal{I}^{x}_{\theta}(\mathbf{Z};\mathbf{X})$$

TURBO: generalized translation problem formulation

Reverse part



Enocder loss

$$I_\theta(\mathbf{X};\mathbf{Z}) = \mathbb{E}_{p_\theta(\mathbf{x},\mathbf{z})}\left[\log \frac{p_\theta(\mathbf{z}|\mathbf{x})}{p_{\mathbf{z}}(\mathbf{z})} \frac{q_\phi(\mathbf{z}|\mathbf{x})}{q_\phi(\mathbf{z}|\mathbf{x})}\right]$$

$$\geq \mathbb{E}_{p_\theta(\mathbf{x},\mathbf{z})}\left[\log \frac{q_\phi(\mathbf{z}|\mathbf{x})}{p_{\mathbf{z}}(\mathbf{z})}\right]$$

$$= \mathbb{E}_{p_\theta(\mathbf{x},\mathbf{z})}\left[\log \frac{q_\phi(\mathbf{z}|\mathbf{x})}{p_{\mathbf{z}}(\mathbf{z})} \frac{\hat{q}_\phi(\mathbf{z})}{\hat{q}_\phi(\mathbf{z})}\right]$$

$$= \mathbb{E}_{p_\theta(\mathbf{x},\mathbf{z})}\left[\log q_\phi(\mathbf{z}|\mathbf{x})\right] - \mathbb{E}_{p_\theta(\mathbf{x},\mathbf{z})}\left[\log \frac{p_{\mathbf{z}}(\mathbf{z})}{\hat{q}_\phi(\mathbf{z})}\right] - \mathbb{E}_{p_\theta(\mathbf{x},\mathbf{z})}\left[\log \hat{q}_\phi(\mathbf{z})\right]$$

$$= \mathbb{E}_{p_\theta(\mathbf{x},\mathbf{z})}\left[\log q_\phi(\mathbf{z}|\mathbf{x})\right] - \mathbb{E}_{p_\theta(\mathbf{x},\mathbf{z})}\left[\log \frac{p_{\mathbf{z}}(\mathbf{z})}{\hat{q}_\phi(\mathbf{z})}\right] + \mathbb{E}_{p_\theta(\mathbf{x}|\mathbf{z})}\left[H\left(p_{\mathbf{z}}(\mathbf{z});\hat{q}_\phi(\mathbf{z})\right)\right]$$

$$\geq \mathbb{E}_{p_\theta(\mathbf{x},\mathbf{z})}\left[\log q_\phi(\mathbf{z}|\mathbf{x})\right] - \mathbb{E}_{p_\theta(\mathbf{x},\mathbf{z})}\left[\log \frac{p_{\mathbf{z}}(\mathbf{z})}{\hat{q}_\phi(\mathbf{z})}\right]$$

$$= \mathbb{E}_{p_{\mathbf{z}}(\mathbf{z})}\mathbb{E}_{p_\theta(\mathbf{x}|\mathbf{z})}\left[\log q_\phi(\mathbf{z}|\mathbf{x})\right] - D_{\mathrm{KL}}\left(p_{\mathbf{z}}(\mathbf{z})\|\hat{q}_\phi(\mathbf{z})\right)$$

$$=: \mathcal{I}^{\mathbf{z}}_{\phi,\theta}(\mathbf{X};\mathbf{Z})$$

Agenda

- What is information bottleneck (IBN)?
- **IBN based autoencoding**
- Generalization of existing methods based on IBN
  - VAE, InfoVAE, VAE/GAN, BIB-AE
- Restrictions of IBN
- **TURBO:** physical-driven latent space
- **Generalization based on TURBO**
  - AAE, SR-GAN and pix2pix, CycleGAN
- Regression problems
  - HEP translation
  - Hubble-to-Webb translation
  - Inverse problems in physics
- Conclusions

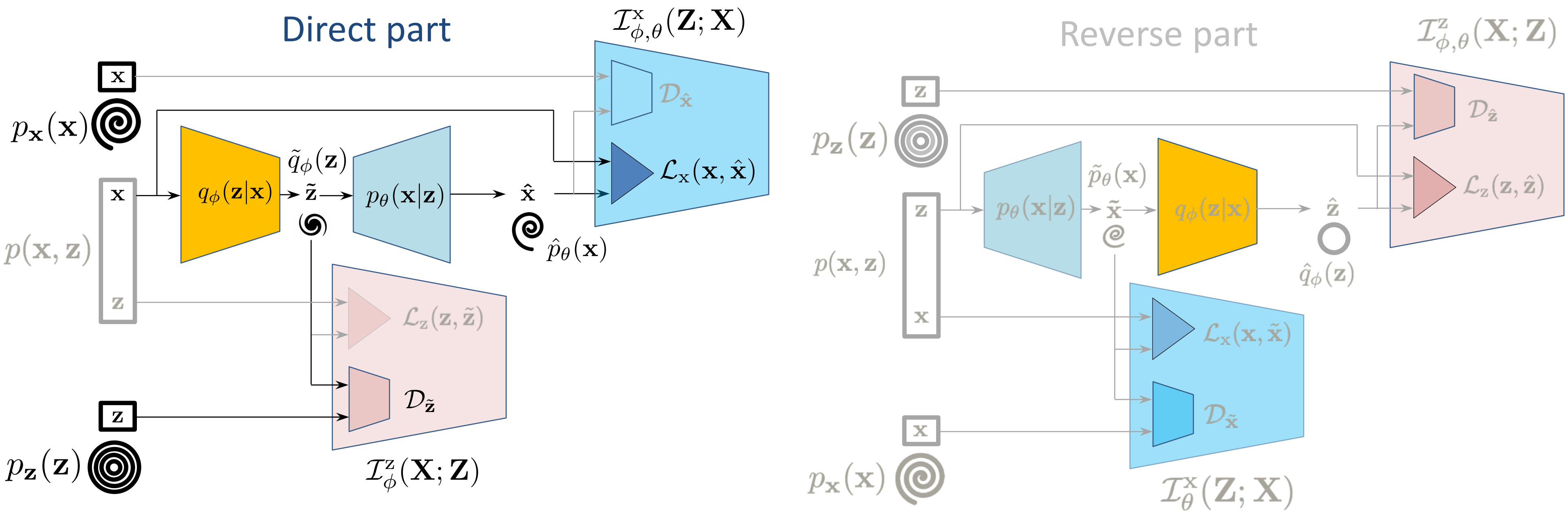
TURBO: generalization of existing schemes

AAE

$$\mathcal{L}_{\mathrm{AAE}}(\phi,\theta) = D_{\mathrm{KL}}\left(\tilde{q}_\phi(\mathbf{z}) \| p_{\mathbf{z}}(\mathbf{z})\right) - \beta \mathbb{E}_{p_{\mathbf{x}}(\mathbf{x})}\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\log p_\theta(\mathbf{x}|\mathbf{z})\right]$$

$$\mathcal{L}^{\mathrm{Direct}}(\phi,\theta) = D_{\mathrm{KL}}\left(p_{\mathbf{z}}(\mathbf{z}) \| \tilde{q}_\phi(\mathbf{z})\right) - \beta \mathbb{E}_{p_{\mathbf{x}}(\mathbf{x})}\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\log p_\theta(\mathbf{x}|\mathbf{z})\right]$$ in minimization form

TURBO



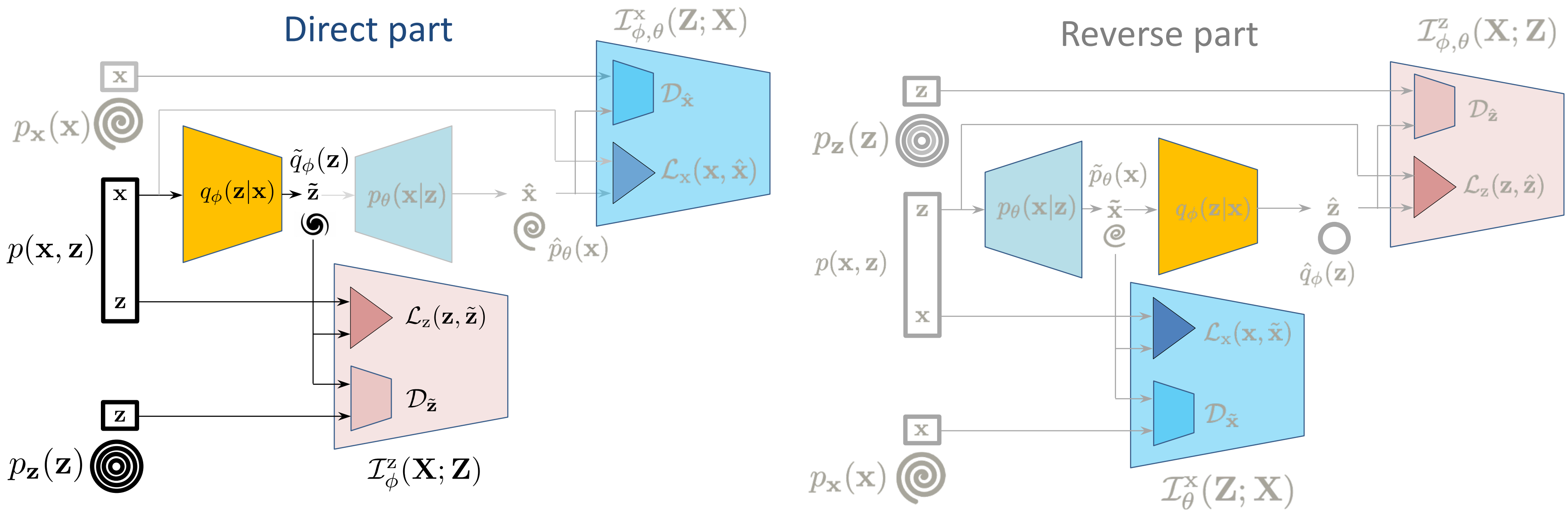
Makhzani, Shlens, Jaitly, Goodfellow, and Frey. Adversarial autoencoders, 2015

TURBO: generalization of existing schemes

Pix2Pix (paired setup) and SRGAN

$$\mathcal{L}_{\text{Pix 2 Pix}}(\theta) = \underbrace{\mathbb{E}_{p(\mathbf{x},\mathbf{z})}\left[\log q_\phi(\mathbf{z}|\mathbf{x})\right]}_{\mathcal{L}_{\mathbf{z}}(\mathbf{z},\tilde{\mathbf{z}})} - \underbrace{D_{\text{KL}}\left(\tilde{q}_\phi(\mathbf{z})\|p_{\mathbf{z}}(\mathbf{z})\right)}_{\mathcal{D}_{\tilde{\mathbf{z}}}}$$

TURBO



Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros, Image-to-Image Translation with Conditional Adversarial Networks, CVPR, 2017

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z. and Shi, W., Photo-realistic single image super-resolution using a generative adversarial network. CVPR 2017

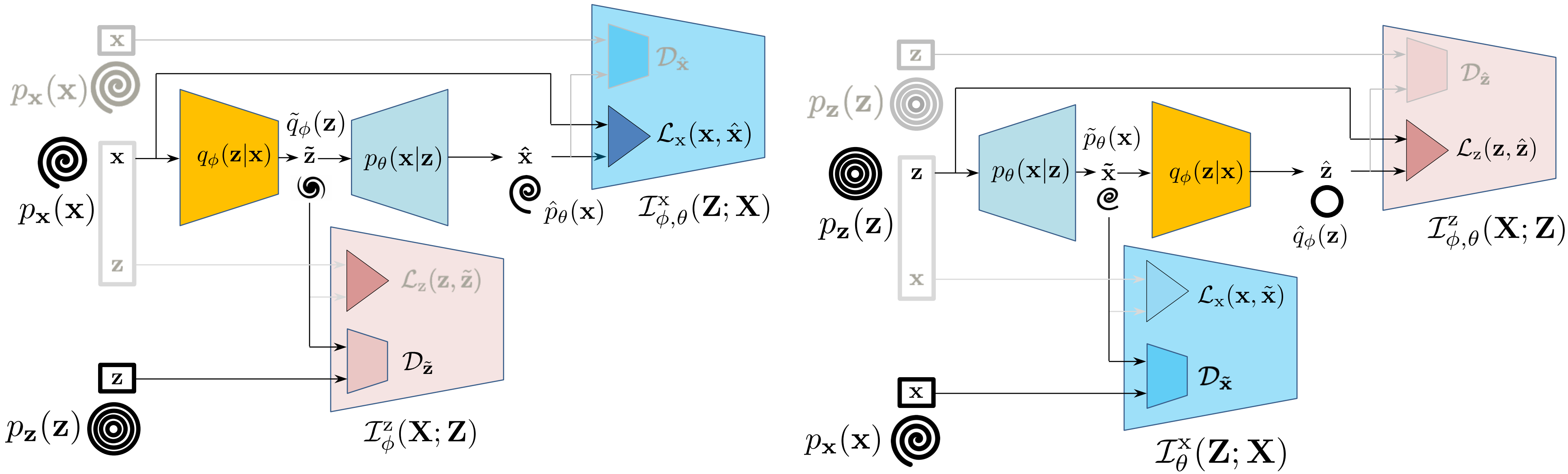
TURBO: generalization of existing schemes

CycleGAN (unpaired setup)

$$\mathcal{L}_{\text{CycleGAN}}(\phi,\theta) = -\underbrace{D_{\text{KL}}\left(p_{\mathbf{z}}(\mathbf{z})\|\tilde{q}_{\phi}(\mathbf{z})\right)}_{\mathcal{D}_{\tilde{\mathbf{z}}}} + \underbrace{\mathbb{E}_{p_{\mathbf{x}}(\mathbf{x})}\mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})}\left[\log p_{\theta}(\mathbf{x}|\mathbf{z})\right]}_{\mathcal{L}_{\mathbf{x}}(\mathbf{x},\hat{\mathbf{x}})}$$

$$-\underbrace{D_{\text{KL}}\left(p_{\mathbf{x}}(\mathbf{x})\|\tilde{p}_{\theta}(\mathbf{x})\right)}_{\mathcal{D}_{\tilde{\mathbf{x}}}} + \underbrace{\mathbb{E}_{p_{\mathbf{z}}(\mathbf{z})}\mathbb{E}_{p_{\theta}(\mathbf{x}|\mathbf{z})}\left[\log q_{\phi}(\mathbf{z}|\mathbf{x})\right]}_{\mathcal{L}_{\mathbf{z}}(\mathbf{z},\hat{\mathbf{z}})}$$

TURBO



Jun-Yan Zhu*, Taesung Park*, Phillip Isola, and Alexei A. Efros. "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks", ICCV 2017

Agenda

- **What is information bottleneck (IBN)?**
- **IBN based autoencoding**
- **Generalization of existing methods based on IBN**
  - VAE, InfoVAE, VAE/GAN, BIB-AE
- **Restrictions of IBN**
- **TURBO: physical-driven latent space**
- **Generalization based on TURBO**
  - AAE, SR-GAN, pix2pix, CycleGAN, Probabilistic AE
- **Regression problems**
  - HEP translation
  - Hubble-to-Webb translation
  - Inverse problems in physics
- **Conclusions**
- **Open problems**

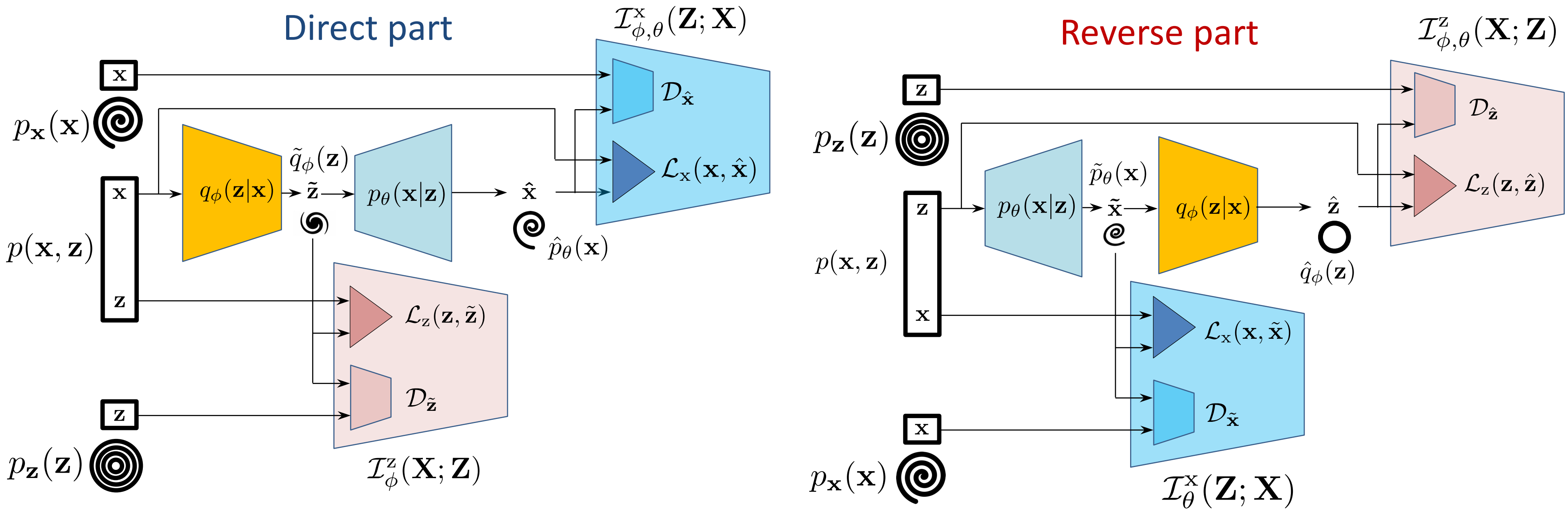
Agenda

- What is information bottleneck (IBN)?
- IBN based autoencoding
- Generalization of existing methods based on IBN
  - VAE, InfoVAE, VAE/GAN, BIB-AE
- Restrictions of IBN
- TURBO: physical-driven latent space
- Generalization based on TURBO
  - AAE, SR-GAN, pix2pix, CycleGAN, Probabilistic AE
- **Regression problems**
  - **HEP translation**
  - Hubble-to-Webb translation
  - Inverse problems in physics
- Conclusions
- Open problems

Regression problems $\mathbf{z} \mapsto \hat{\mathbf{x}}$

HEP translation problem

TURBO



Physical meaninful latent space

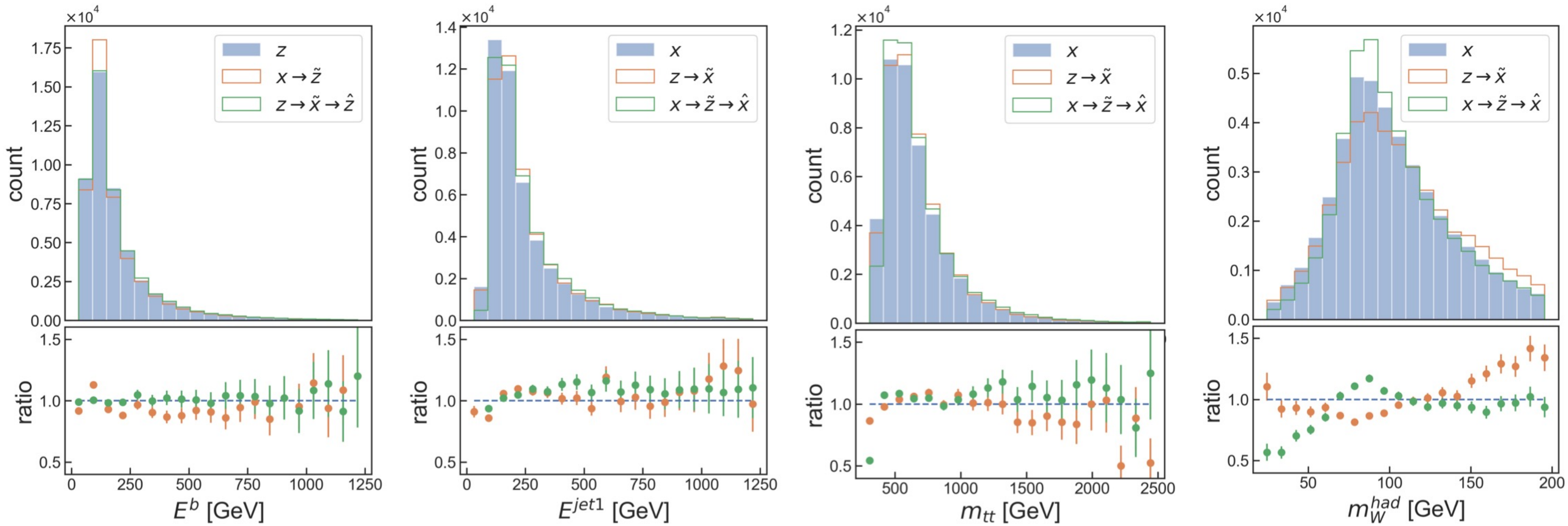
Z is the theory space, i.e. right after the collision, before any interaction with the detector

X is the experiment space, i.e. after reconstructing the detector signal

G. Quétant, M. Drozdova, V. Kinakh, T. Golling, and S. Voloshynovskiy, "Turbo-Sim: a generalised generative model with a physical latent space." NeurIPS, ML4PhysicalSciences2021.

Regression problems $\mathbf{z} \mapsto \hat{\mathbf{x}}$

HEP translation problem



Conclusions:

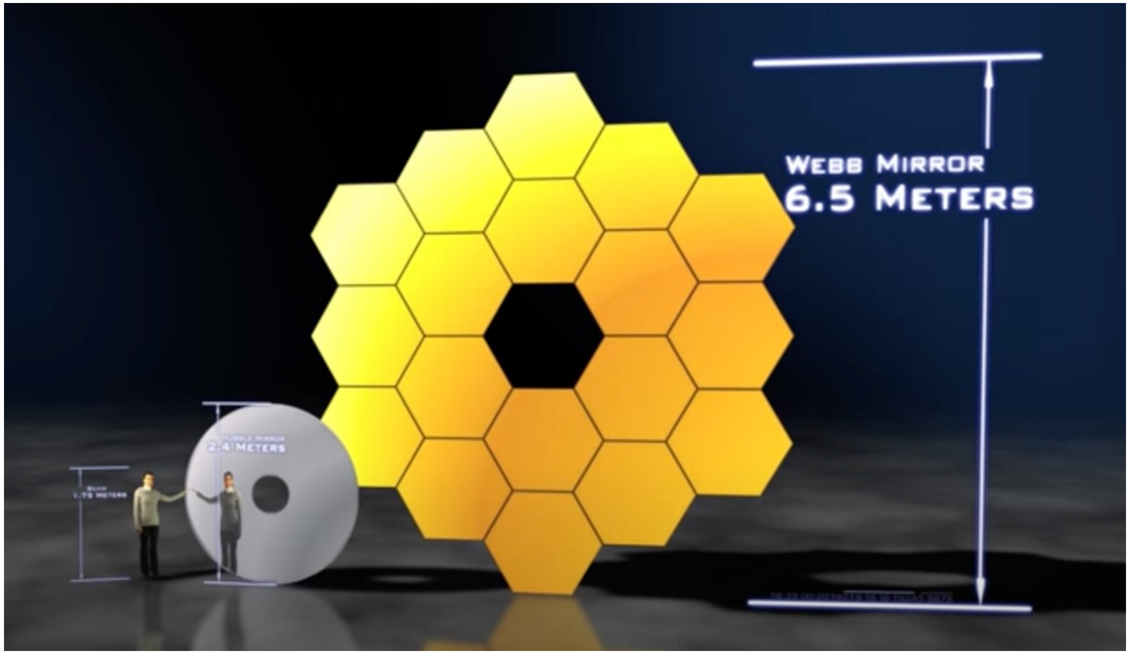
- Much sense in maximising mutual information since X and Z are very correlated
- Competitive with state-of-the-art, outperforming it in some tasks
- Trained for both generation and inference at the same time

G. Quétant, M. Drozdova, V. Kinakh, T. Golling, and S. Voloshynovskiy, "Turbo-Sim: a generalised generative model with a physical latent space." NeurIPS, ML4PhysicalSciences2021.

Agenda

- What is information bottleneck (IBN)?
- IBN based autoencoding
- Generalization of existing methods based on IBN
  - VAE, InfoVAE, VAE/GAN, BIB-AE
- Restrictions of IBN
- TURBO: physical-driven latent space
- Generalization based on TURBO
  - AAE, SR-GAN, pix2pix, CycleGAN, Probabilistic AE
- **Regression problems**
  - HEP translation
  - Hubble-to-Webb translation
  - Inverse problems in physics
- Conclusions

Hubble-to-Webb translation problem



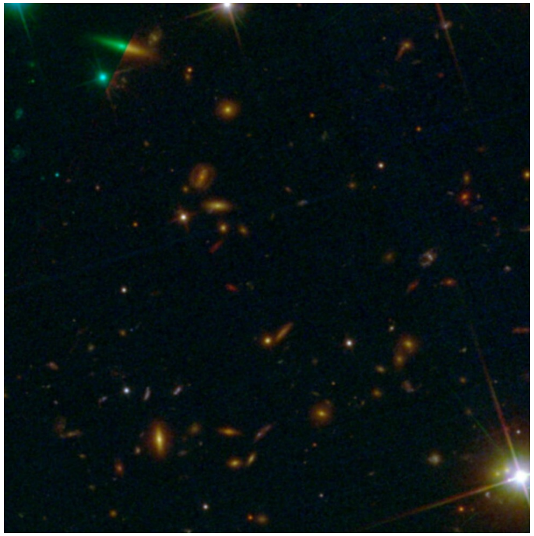
Apr. 24, 1990 Dec. 25, 2021

- Different size of mirrors: 6,5 m Webb vs 2,2 m of Hubble
- Different bands
  - Hubble: ultraviolet light, visible light and a small slice of infrared
  - Webb: optimized for infrared but can see red, orange, and gold visible light.
- Different resolutions, sensitivities and captures different phenomena

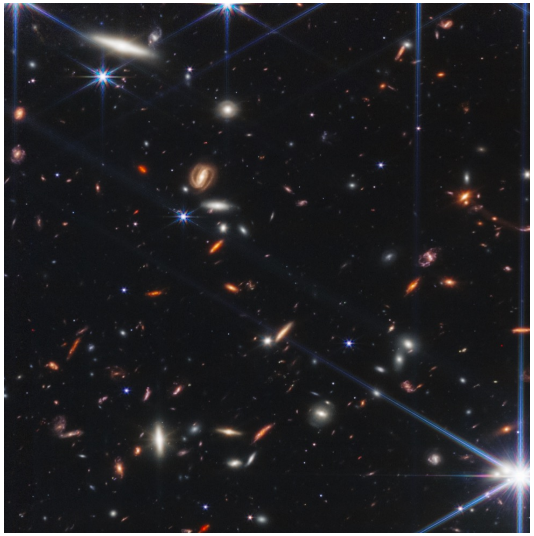
https://www.jwst.nasa.gov/content/about/comparisonWebbVsHubble.html

Hubble-to-Web translation problem

Results of prediction



Hubble

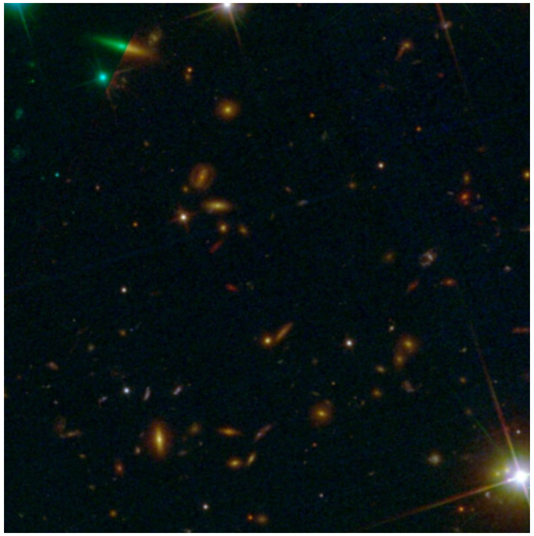


Webb

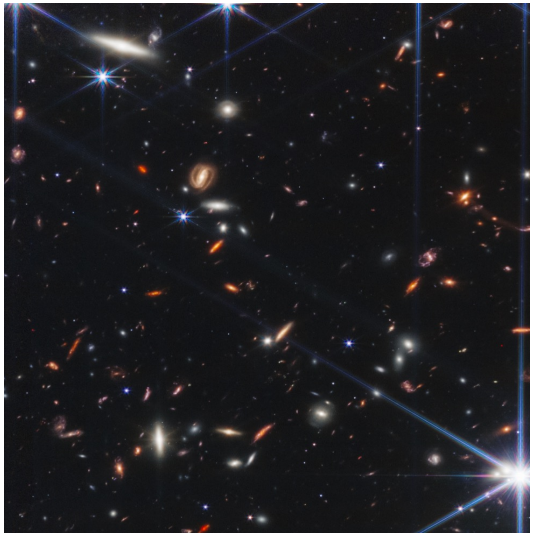
Predicted Webb

Hubble-to-Web translation problem

Results of prediction



Hubble



Webb

Predicted Webb

Agenda

- What is information bottleneck (IBN)?
- IBN based autoencoding
- Generalization of existing methods based on IBN
  - VAE, InfoVAE, VAE/GAN, BIB-AE
- Restrictions of IBN
- TURBO: physical-driven latent space
- Generalization based on TURBO
  - AAE, SR-GAN, pix2pix, CycleGAN, Probabilistic AE
- **Regression problems**
  - HEP translation
  - Hubble-to-Webb translation
  - Inverse problems in physics
- Conclusions
- Open problems

Regression problems $\mathbf{z} \mapsto \hat{\mathbf{x}}$

Inverse problems

a common basis for most of imaging problems

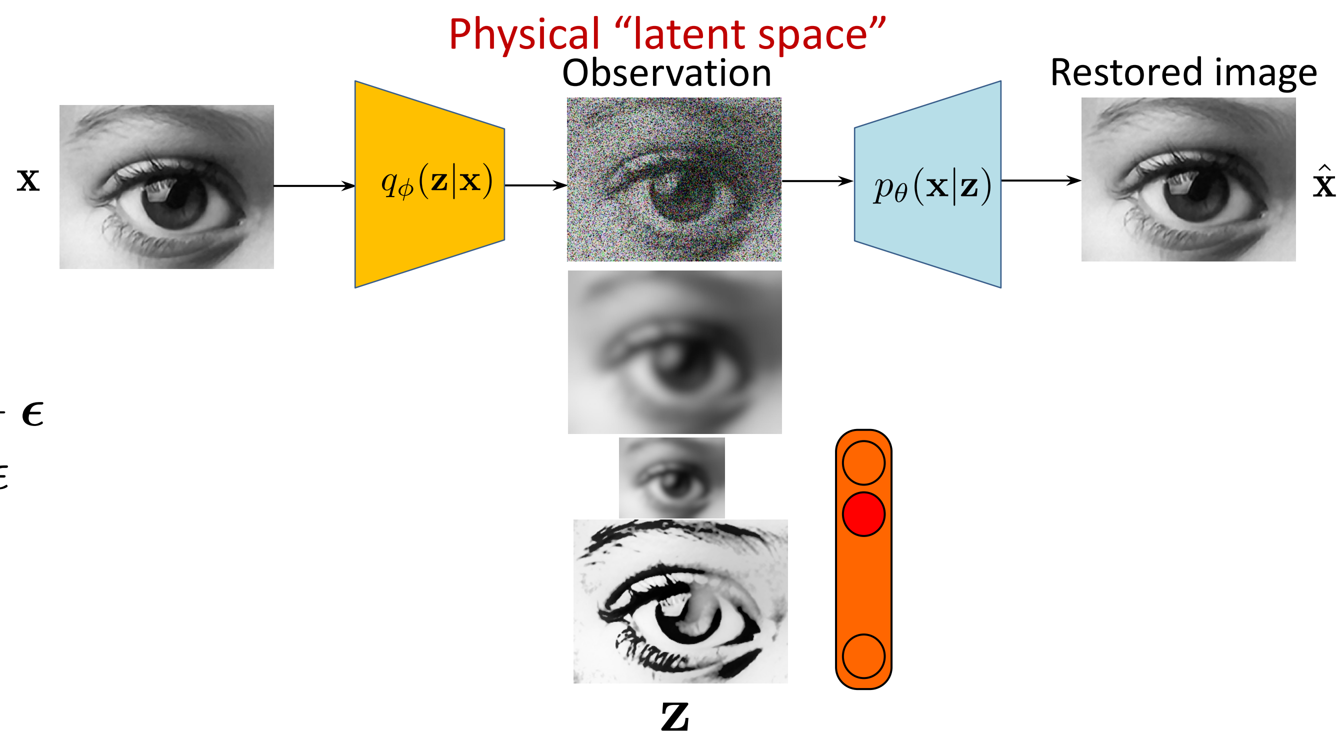
$$\mathbf{z} = f_\phi(\mathbf{x}, \boldsymbol{\epsilon})$$

- Known
- Unknown

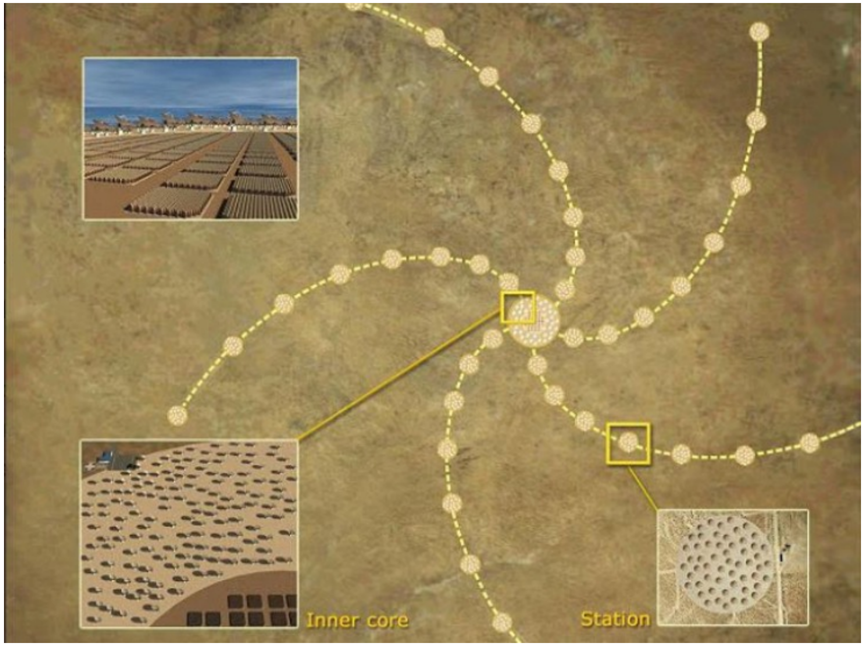
- Sampling: fMRI (k-space), arrays (uv-plane)
- Compressive sensing
- Learnable compressive sampling
- Denoising
- Restoration and reconstruction
- Superresolution
- Inpainting

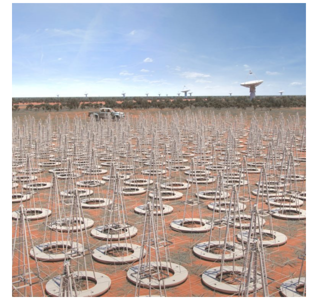
$$\mathbf{z} = P_\phi \Psi \mathbf{x} + \boldsymbol{\epsilon}$$

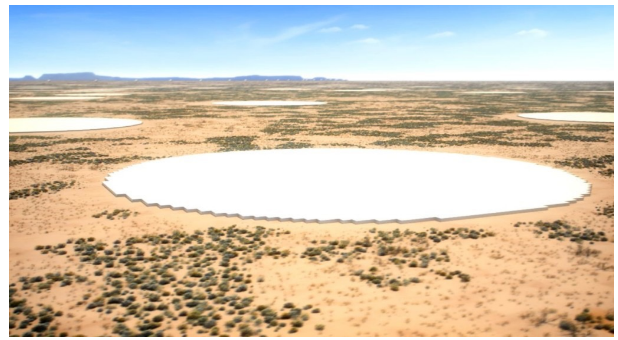
$$\mathbf{z} = A_\phi \mathbf{x} + \boldsymbol{\epsilon}$$



Regression problems $\mathbf{z} \mapsto \hat{\mathbf{x}}$







Square Kilometer Array (SKA) – imaging tool of the 21st century:

- Huge amount of data (expected data are about 1 PB per day)
- Problems with Big Data:
  - Reconstruction (where? and how?)
  - Intercontinental data exchange
  - Storage
  - Analytics and Science

Similar problems in fMRI, CT, computational photography, etc.

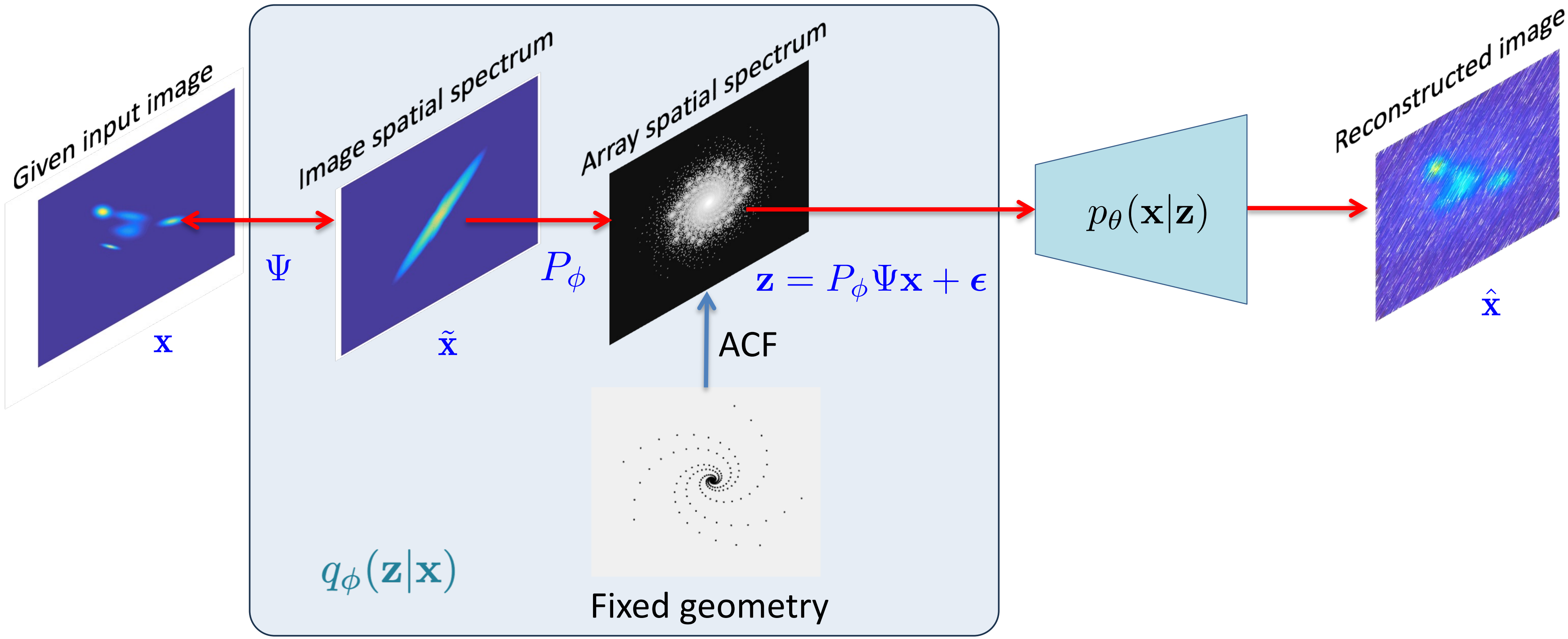
Information Bottleneck: regression problems $\mathbf{z} \mapsto \hat{\mathbf{x}}$

$\mathbf{z} = P_\phi \Psi \mathbf{x} + \epsilon$

$\mathbf{z} = A_\phi \mathbf{x} + \epsilon$

Ψ Fourier transform operator

P_ϕ Sampling operator



Problems:

- Proper image priors $\Omega(\mathbf{x})$
- Joint optimization of sampling operator P_ϕ (encoder $q_\phi(\mathbf{z}|\mathbf{x})$) and decoder $p_\theta(\mathbf{x}|\mathbf{z})$.

Agenda

- **What is information bottleneck (IBN)?**
- **IBN based autoencoding**
- **Generalization of existing methods based on IBN**
  - VAE, InfoVAE, VAE/GAN, BIB-AE
- **Restrictions of IBN**
- **TURBO: physical-driven latent space**
- **Generalization based on TURBO**
  - AAE, SR-GAN, pix2pix, CycleGAN, Probabilistic AE
- **Regression problems**
  - HEP translation
  - Hubble-to-Webb translation
  - Inverse problems in physics
- **Conclusions**

Conclusions

- We considered IBN in the variational formulation
- IBN is a useful tool for the analysis and generalization of existing schemes but it has its own restrictions
- TURBO can fulfill the gap in physical applications where data should have some meaningful latent space
- TURBO can generalize schemes not governed by IBN and envision new architectures
- **Not covered problems:**
  - Extension to Probabilistic AE (latent space with FLOW)
  - Extension to Diffusion-type models (latent space with Markov chain)
  - Extension to score based models (linking MAP with physical models)

Conclusions

- You can find more details about Turbo

entropy

MDPI

Article

TURBO: The Swiss Knife of Auto-Encoders

Guillaume Quétant *, Yury Belousov, Vitaliy Kinakh and Slava Voloshynovskiy *

Centre Universitaire d'Informatique, Université de Genève, Route de Drize 7, CH-1227 Carouge, Switzerland; yury.belousov@unige.ch (Y.B.); vitaliy.kinakh@unige.ch (V.K.)
* Correspondence: guillaume.quetant@unige.ch (G.Q.); svolos@unige.ch (S.V.)

Abstract: We present a novel information-theoretic framework, termed as TURBO, designed to systematically analyse and generalise auto-encoding methods. We start by examining the principles of information bottleneck and bottleneck-based networks in the auto-encoding setting and identifying their inherent limitations, which become more prominent for data with multiple relevant, physics-related representations. The TURBO framework is then introduced, providing a comprehensive derivation of its core concept consisting of the maximisation of mutual information between various data representations expressed in two directions reflecting the information flows. We illustrate that numerous prevalent neural network models are encompassed within this framework. The paper underscores the insufficiency of the information bottleneck concept in elucidating all such models, thereby establishing TURBO as a preferable theoretical reference. The introduction of TURBO contributes to a richer understanding of data representation and the structure of neural network models, enabling more efficient and versatile applications.

Keywords: information bottleneck; TURBO; generalisation; auto-encoder; variational approximation; lower bound; mutual information; physical latent space; representations; Kullback–Leibler divergence

https://www.mdpi.com/1099-4300/25/10/1471

Thank you!