UNIVERSITÉ DE GENÈVE

Fast indexing based on the latent space representation

Slava Voloshynovskiy

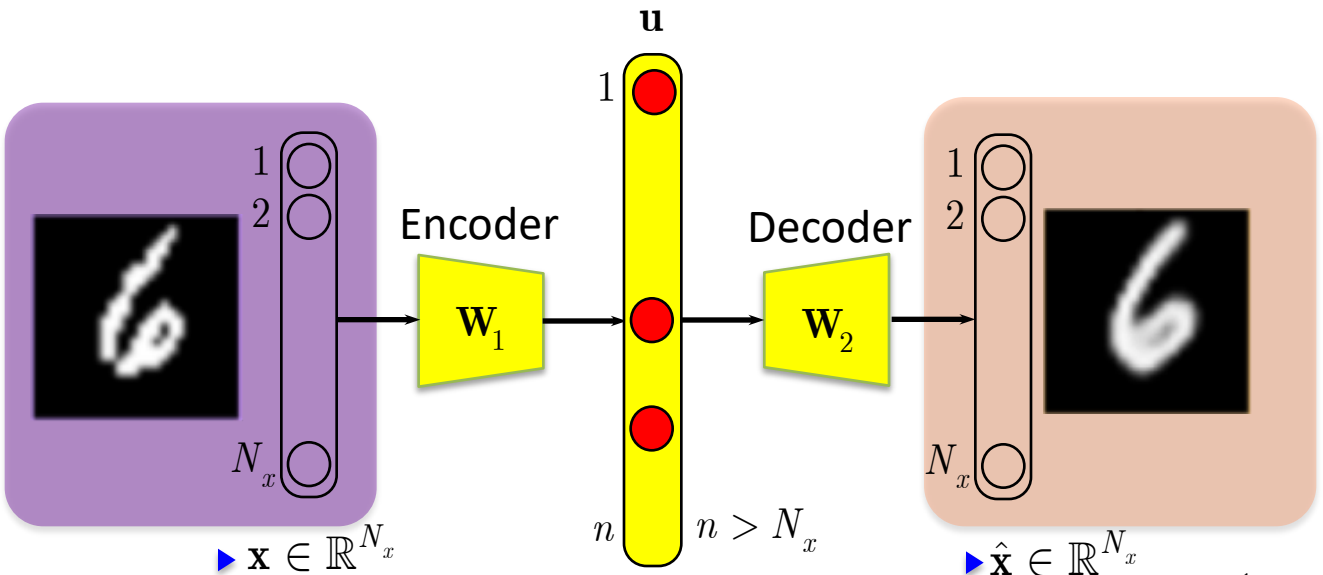
University of Geneva

Switzerland

2019, Paris

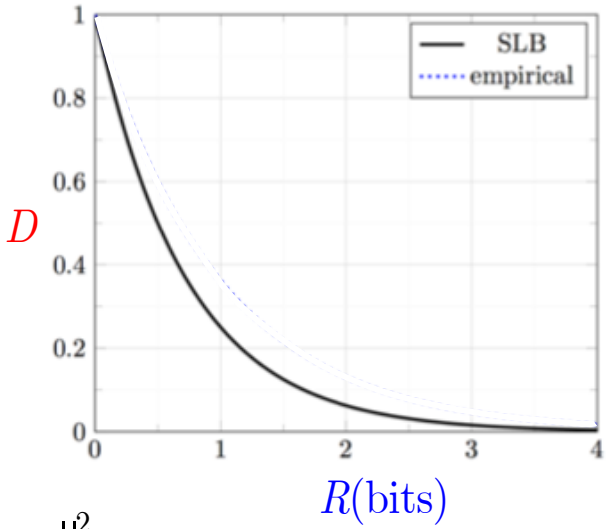
SIP stochastic information processing

Basic framework



- $\mathbf{x} \in \mathbb{R}^{N_x}$
- $\mathbf{x} \sim p(\mathbf{x})$
- $\mathbf{u} \in \{-1, 0, +1\}^n$
- $\|\mathbf{u}\|_0 \leq s$
- Rate: $R = \frac{1}{n} \log_2 \left(\binom{n}{s} 2^s \right)$
- $\hat{\mathbf{x}} \in \mathbb{R}^{N_x}$
- Distortion: $\frac{1}{N_x} \|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 \leq D$

Rate-distortion function



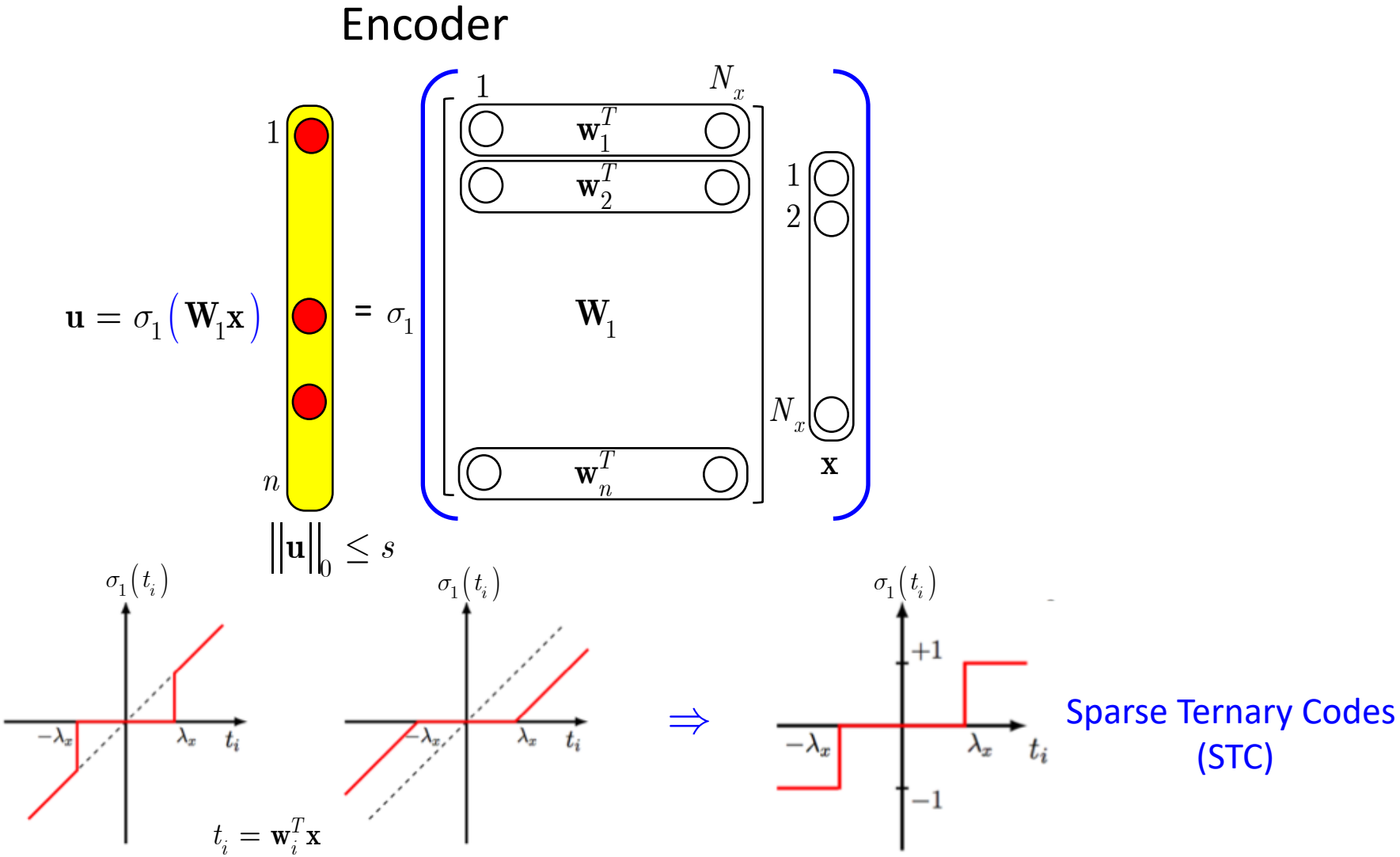
Remark: Shannon rate-distortion

- $\mathbf{u} \in \{0, +1\}^{N_x}$
- $\|\mathbf{u}\|_0 = 1$

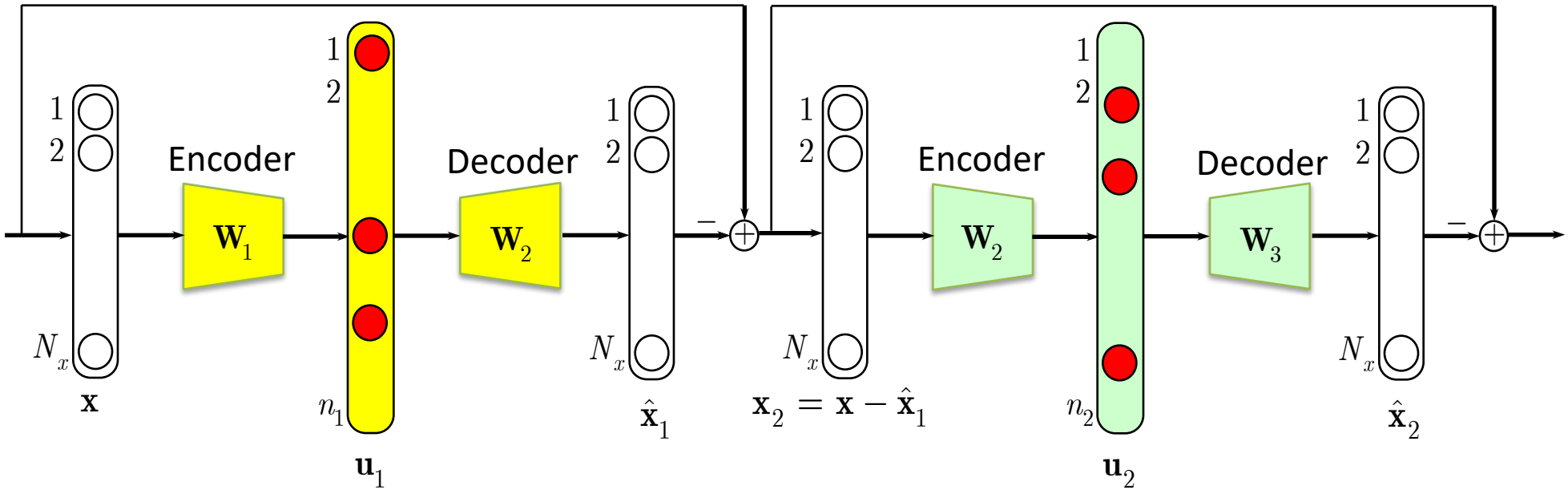
Impractical result

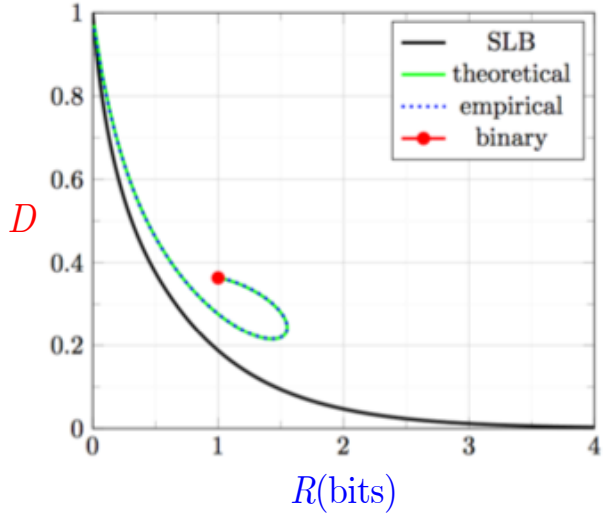
- $N_x \to \infty$
- $n \sim 2^{N_x R}$ codebook in exponential in N_x
- $p(\mathbf{x})$ source distribution should be known

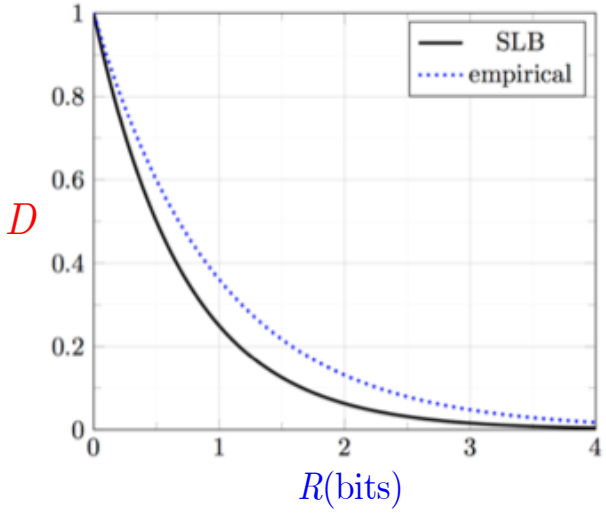
Basic framework



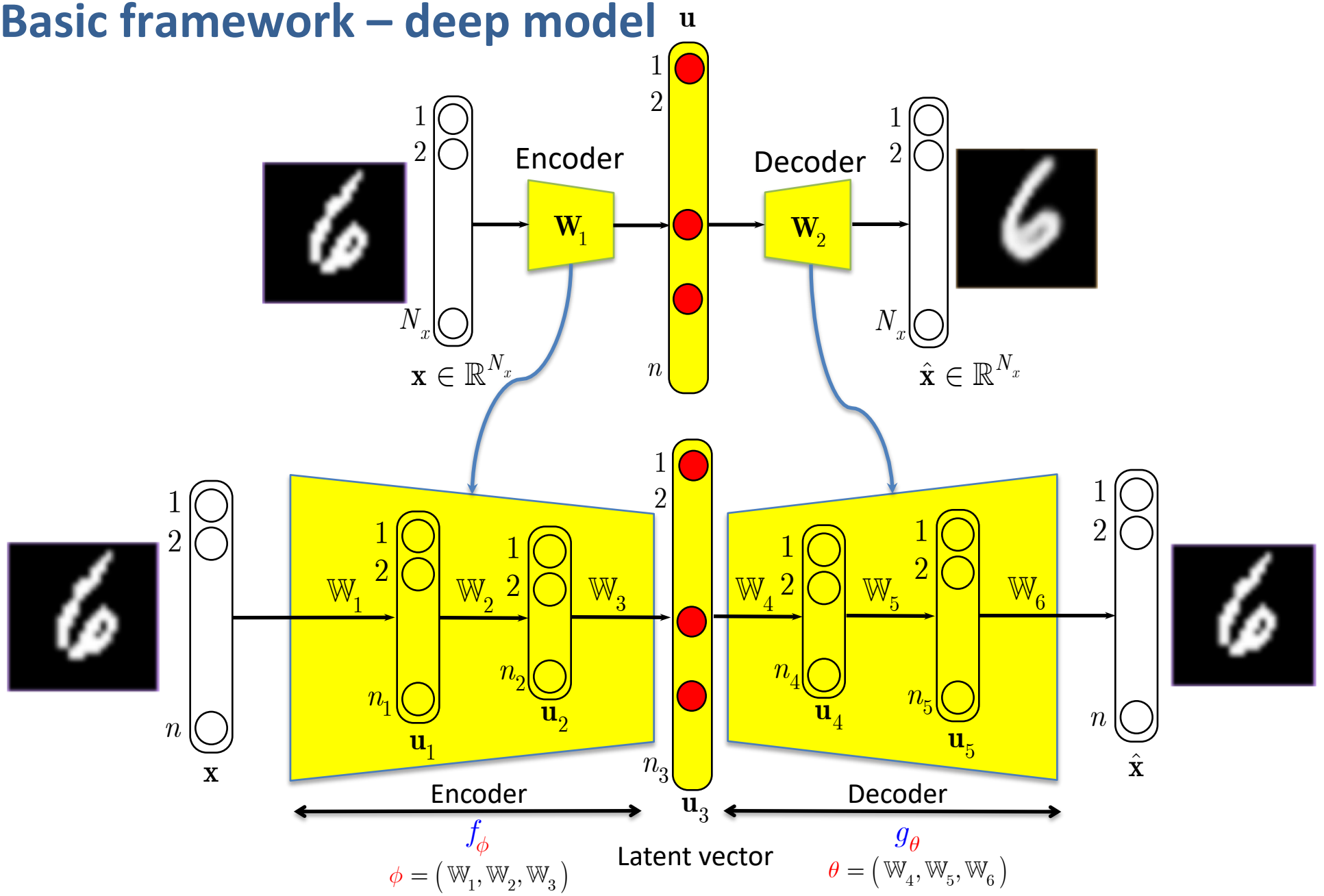
Basic framework – successive refinement







Basic framework – deep model



Applications

- Regression problems
  - Denoising, superresolution, compression
- Search and indexing problems
  - Physical object security
- Classification based on distributed VAEs

Applications

- Regression problems
  - Denoising, superresolution, compression, novelty detection
- Search and indexing problems
  - Physical object security
- Classification based on distributed VAEs

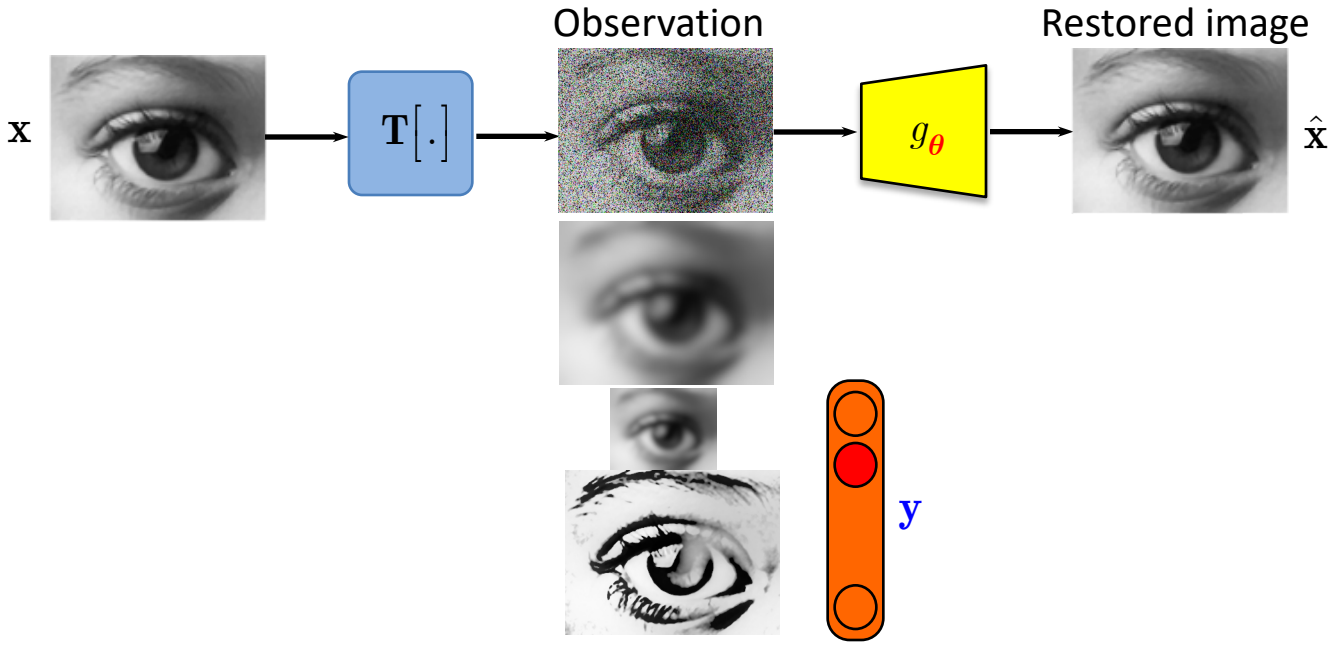
Regression problems

a common basis
for most of imaging problems

$$\mathbf{y} = \mathbf{T}\left[\mathbf{x}\right] + \mathbf{z}$$

- Known
- Unknown

- Sampling
- Compressive sensing
- Learnable compressive sampling
- Denoising
- Restoration
- Compression
- Superresolution
- Inpainting



Regression problems

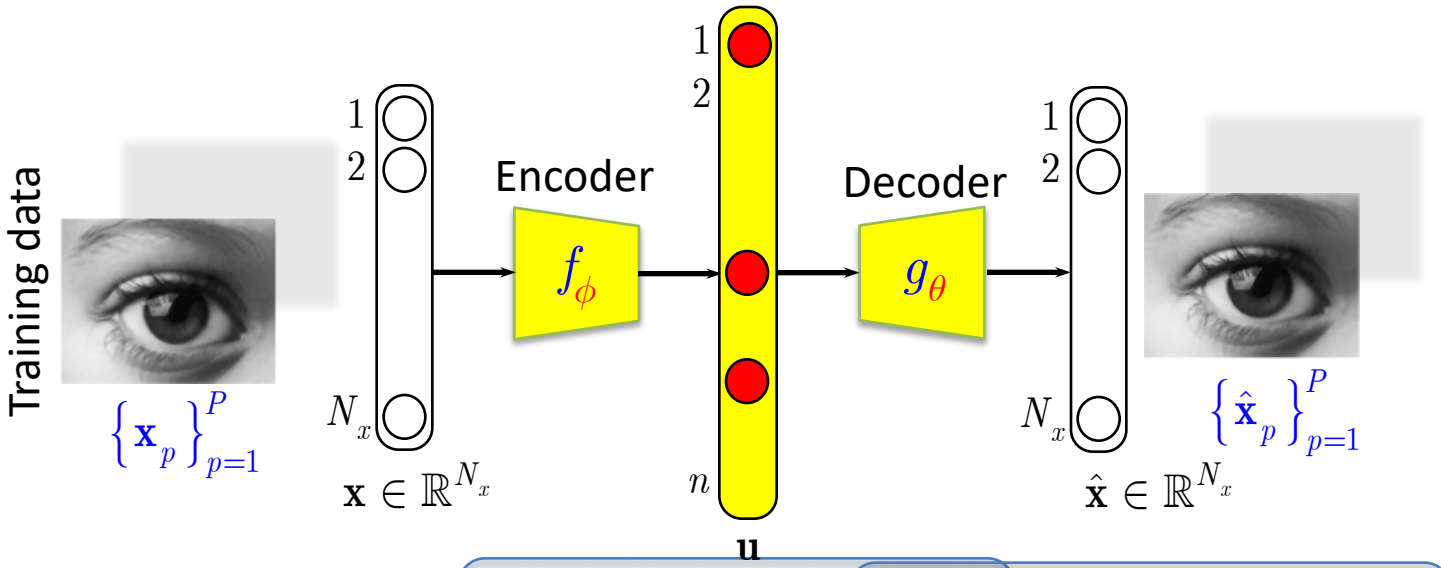
Smoothness of solution, local correlations ...

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \left\| \mathbf{y} - \mathbf{T}\left[\mathbf{x}\right] \right\|_2^2 + \lambda \Omega\left(\mathbf{x}\right)$$

$p(\mathbf{x})$

$\Rightarrow$ often unknown
$\Rightarrow$ very difficult to describe analytically
$\Rightarrow$ defined solely based on human expertise
$\Rightarrow$ domain or application specific

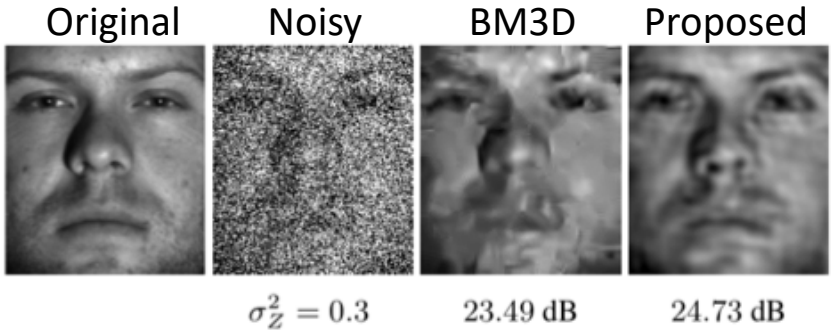
$$\Omega\left(\mathbf{x}\right) = -\ln p\left(\mathbf{x}\right)$$



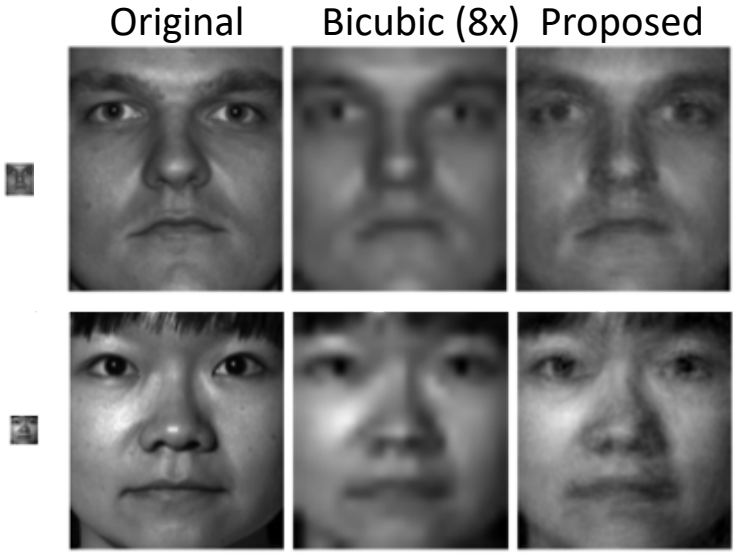
$$\left(\hat{\phi}, \hat{\theta}\right) = \arg\min_{\phi,\theta} \sum_{p=1}^{P} \left\| \left(\mathbf{u}_p - f_\phi\left(\mathbf{x}_p\right)\right) \right\|_2^2 + \lambda \Omega\left(\mathbf{u}_p\right) + \beta \left\| \left(\mathbf{x}_p - g_\theta\left(\mathbf{u}_p\right)\right) \right\|_2^2 + \lambda_1 \Omega\left(\phi\right) + \lambda_2 \Omega\left(\theta\right)$$

Regression problems

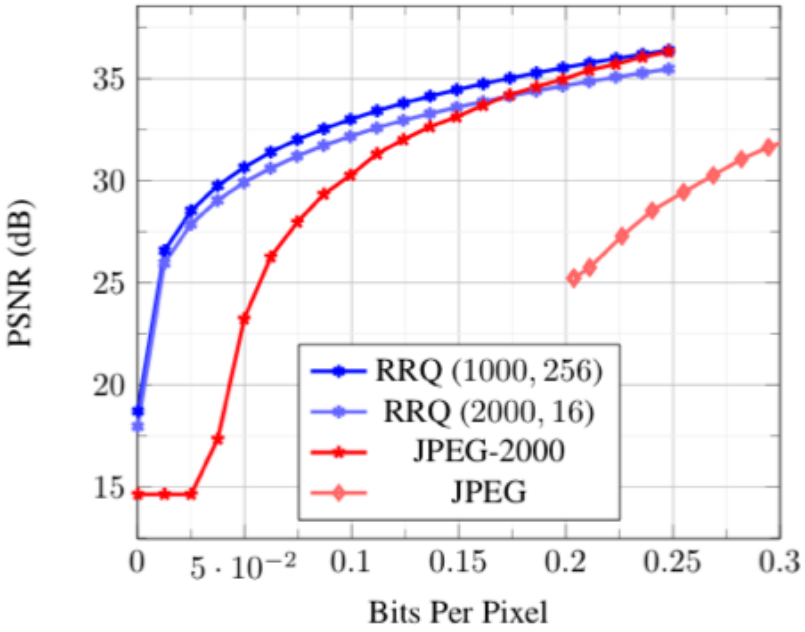
Denoising



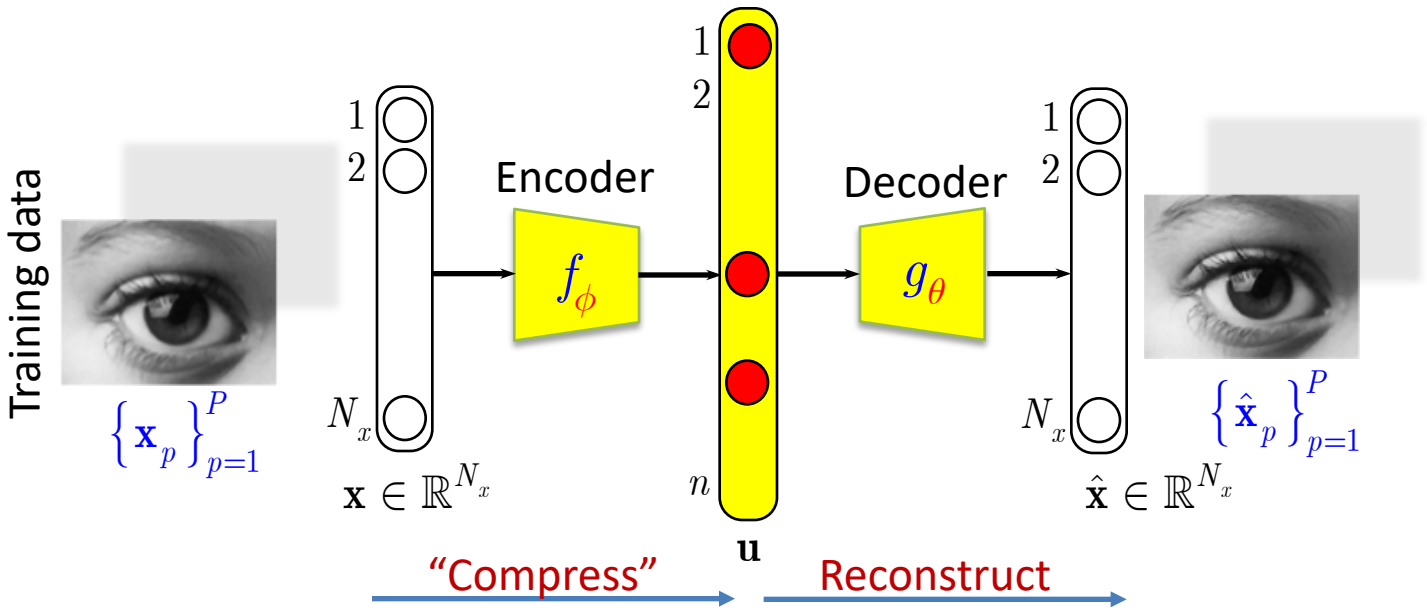
Superresolution (single image)



Lossy image compression (domain adapted compression)



Link to Information Bottleneck formulation

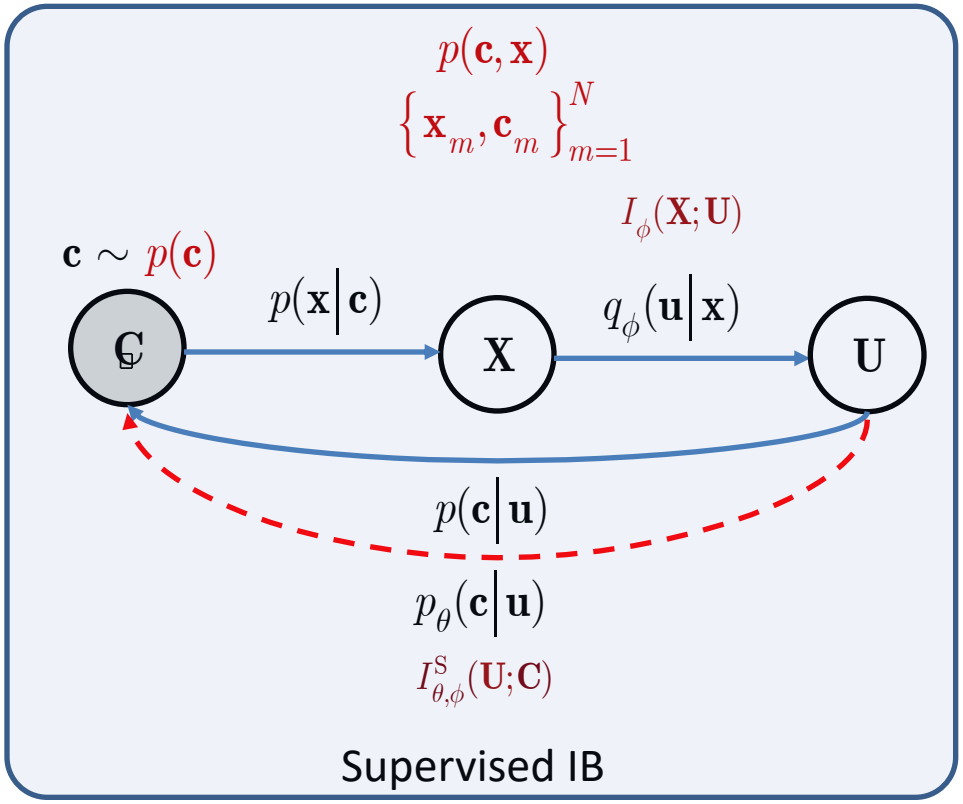


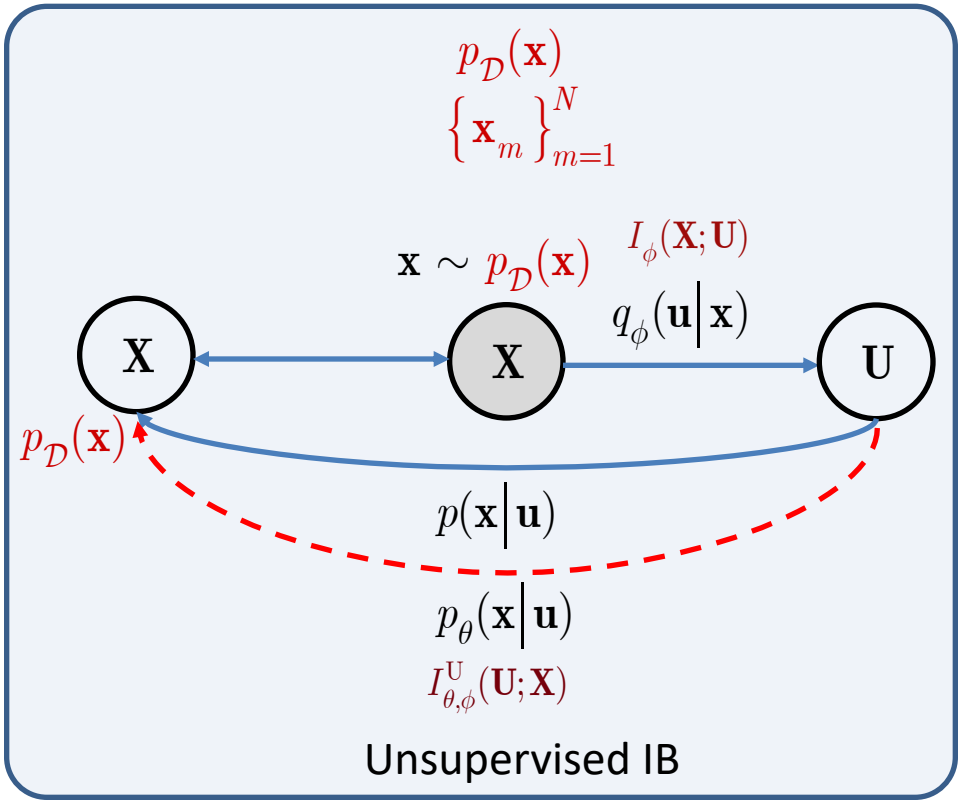
Information Bottleneck for AE formulation

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

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

Link to Information Bottleneck formulation





$$\mathcal{L}_{\mathrm{S}}(\phi,\theta) = I_\phi(\mathbf{X};\mathbf{U}) - \beta I_{\phi,\theta}(\mathbf{U};\mathbf{C})$$

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

Link to Information Bottleneck formulation

Bounded Information Bottleneck AE (BIB-AE)

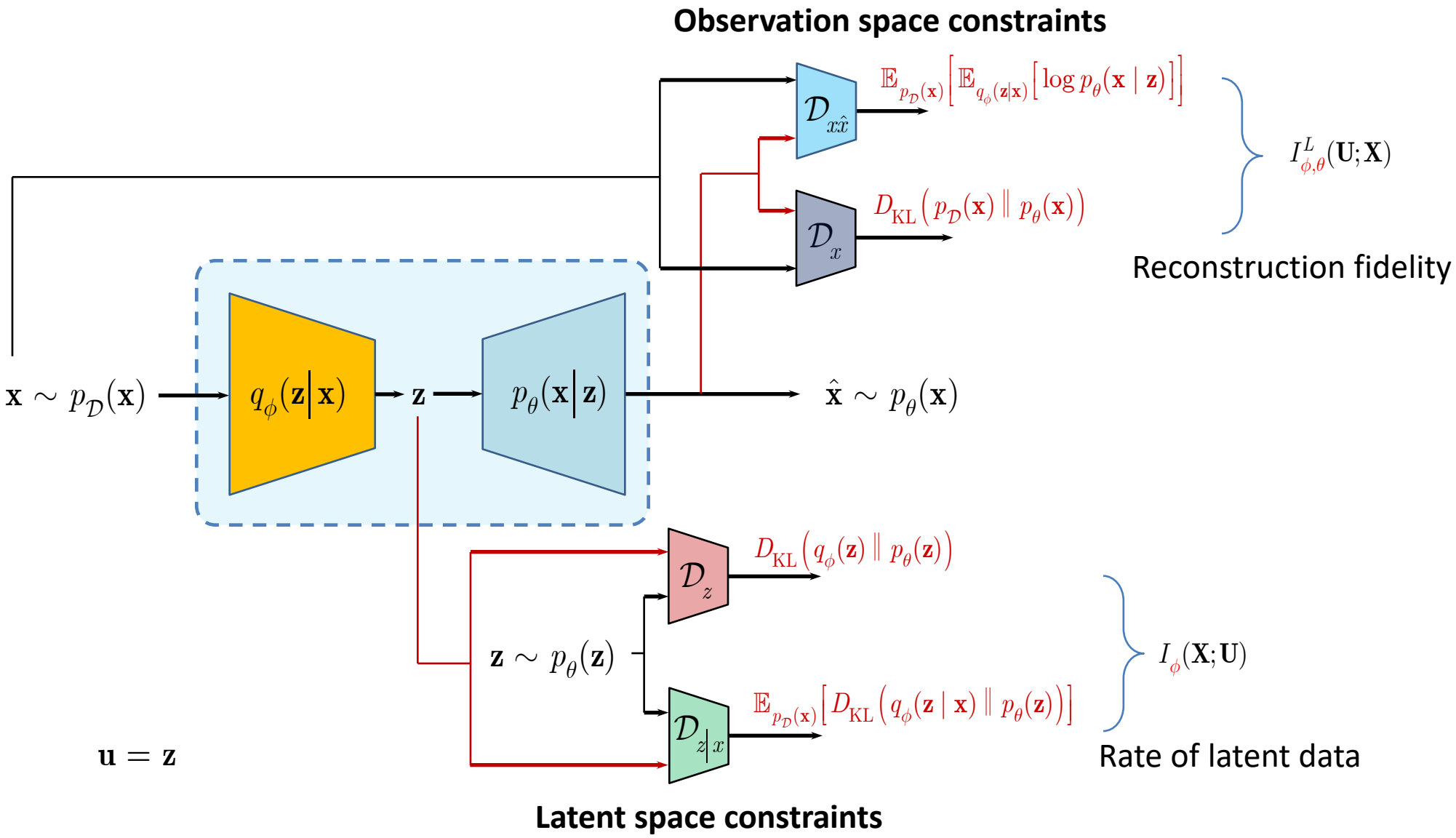
$$\mathcal{L}_{\text{BIB-AE}}(\phi,\theta) = I_{\phi}(\mathbf{X};\mathbf{U}) - \beta I^{L}_{\phi,\theta}(\mathbf{U};\mathbf{X})$$

where $I^{L}_{\phi,\theta}(\mathbf{U};\mathbf{X}) \leq I_{\phi,\theta}(\mathbf{U};\mathbf{X})$

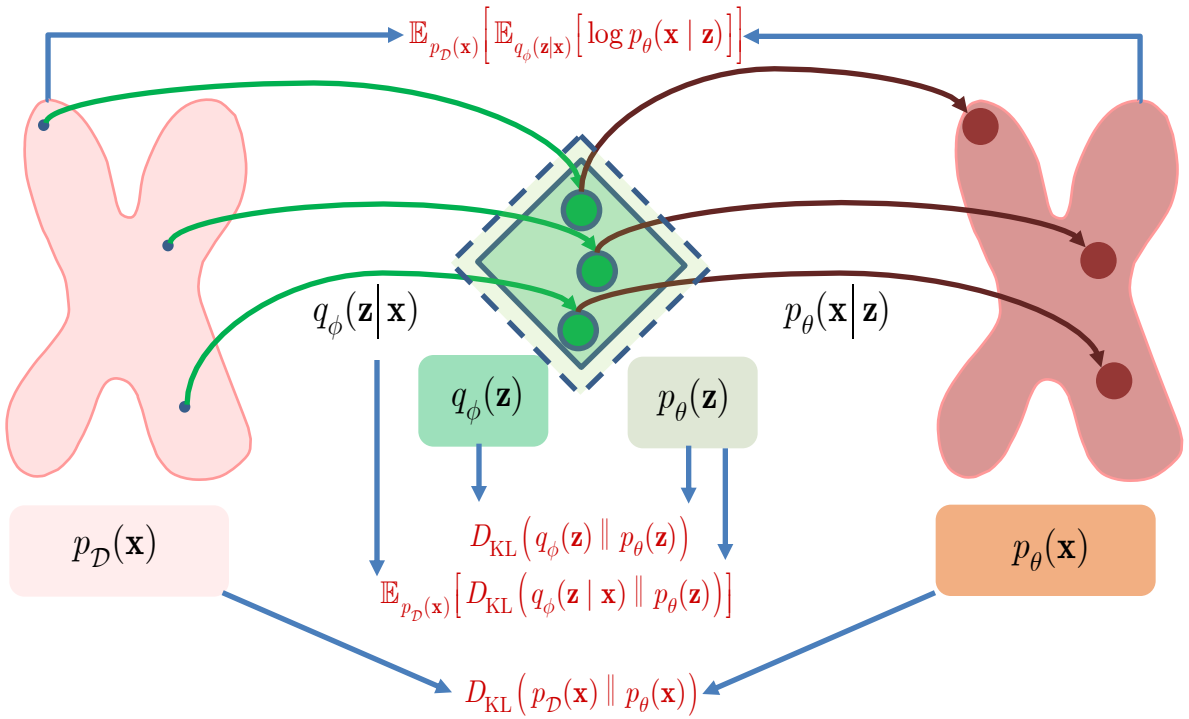
$$I_{\phi}(\mathbf{X};\mathbf{U}) = \underbrace{\mathbb{E}_{p_{\mathcal{D}}(\mathbf{x})}\Big[D_{\text{KL}}\Big(q_{\phi}(\mathbf{z}\mid\mathbf{x}) \parallel p_{\theta}(\mathbf{z})\Big)\Big]}_{\text{A}} - \underbrace{D_{\text{KL}}\Big(q_{\phi}(\mathbf{z}) \parallel p_{\theta}(\mathbf{z})\Big)}_{\text{B}}$$

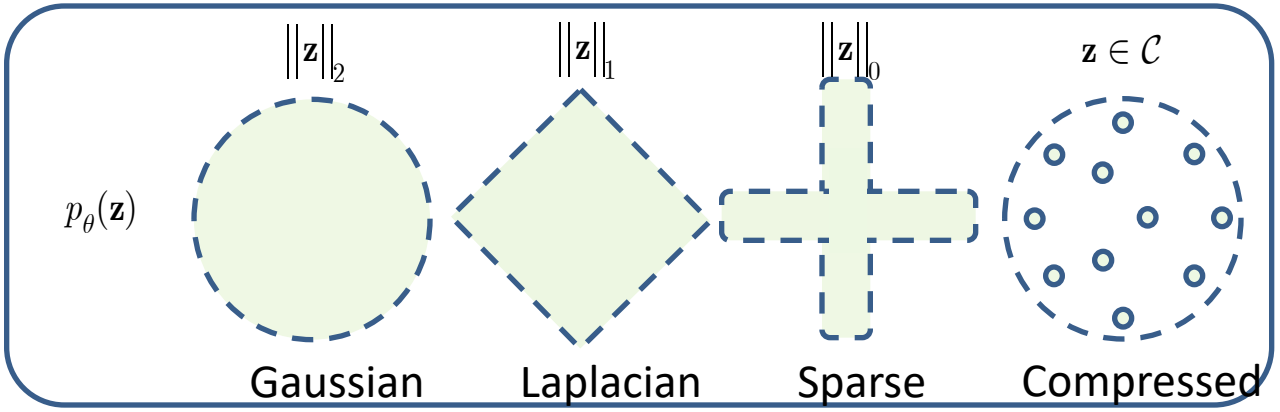
$$I^{L}_{\phi,\theta}(\mathbf{U};\mathbf{X}) = \underbrace{\mathbb{E}_{p_{\mathcal{D}}(\mathbf{x})}\Big[\mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})}\Big[\log p_{\theta}(\mathbf{x}\mid\mathbf{z})\Big]\Big]}_{\text{C}} - \underbrace{D_{\text{KL}}\Big(p_{\mathcal{D}}(\mathbf{x}) \parallel p_{\theta}(\mathbf{x})\Big)}_{\text{D}}$$

Link to Information Bottleneck formulation



Link to Information Bottleneck formulation

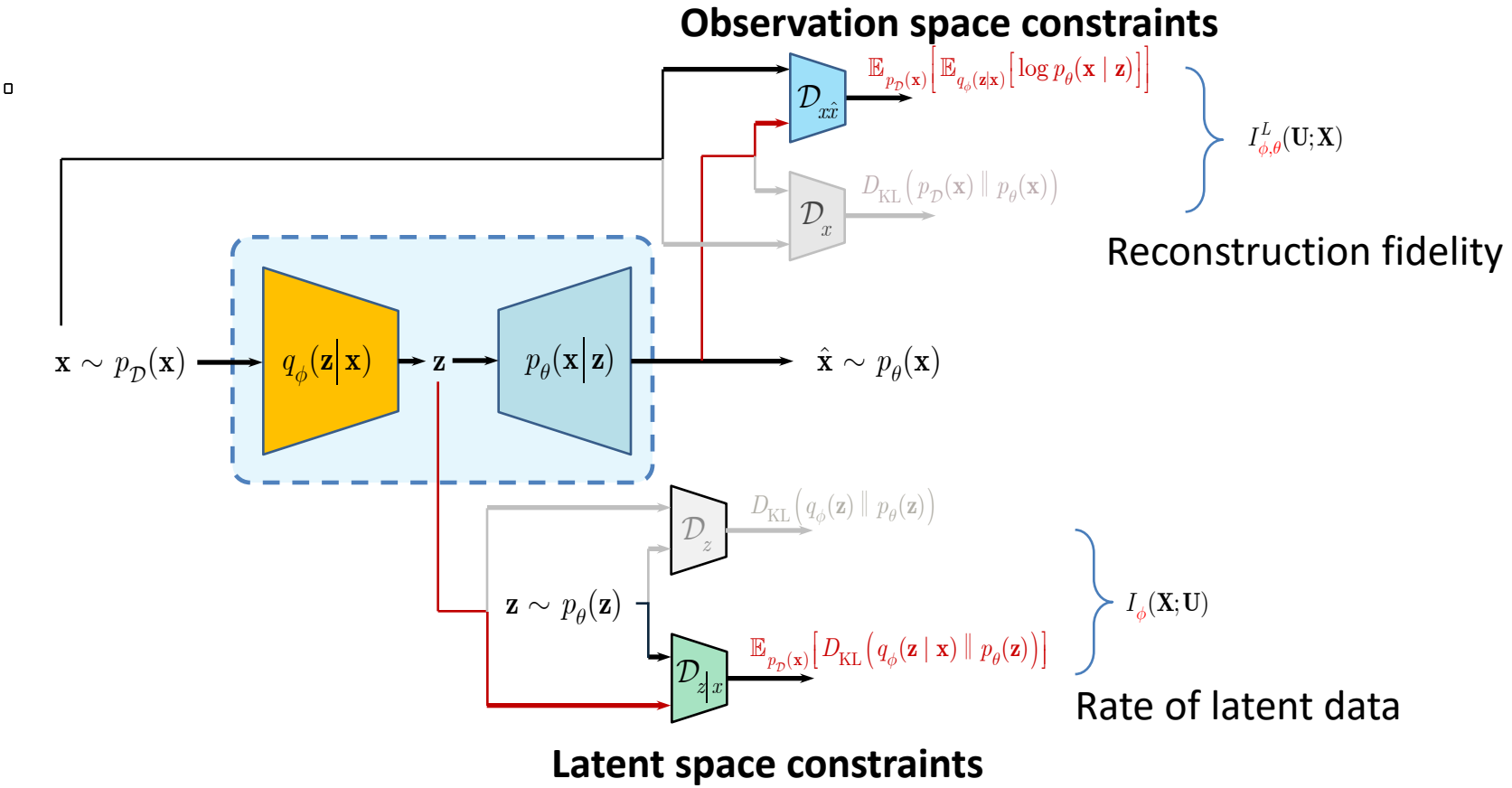




Link to Information Bottleneck formulation

VAE and $\beta-$VAE: Variational Autoencoder

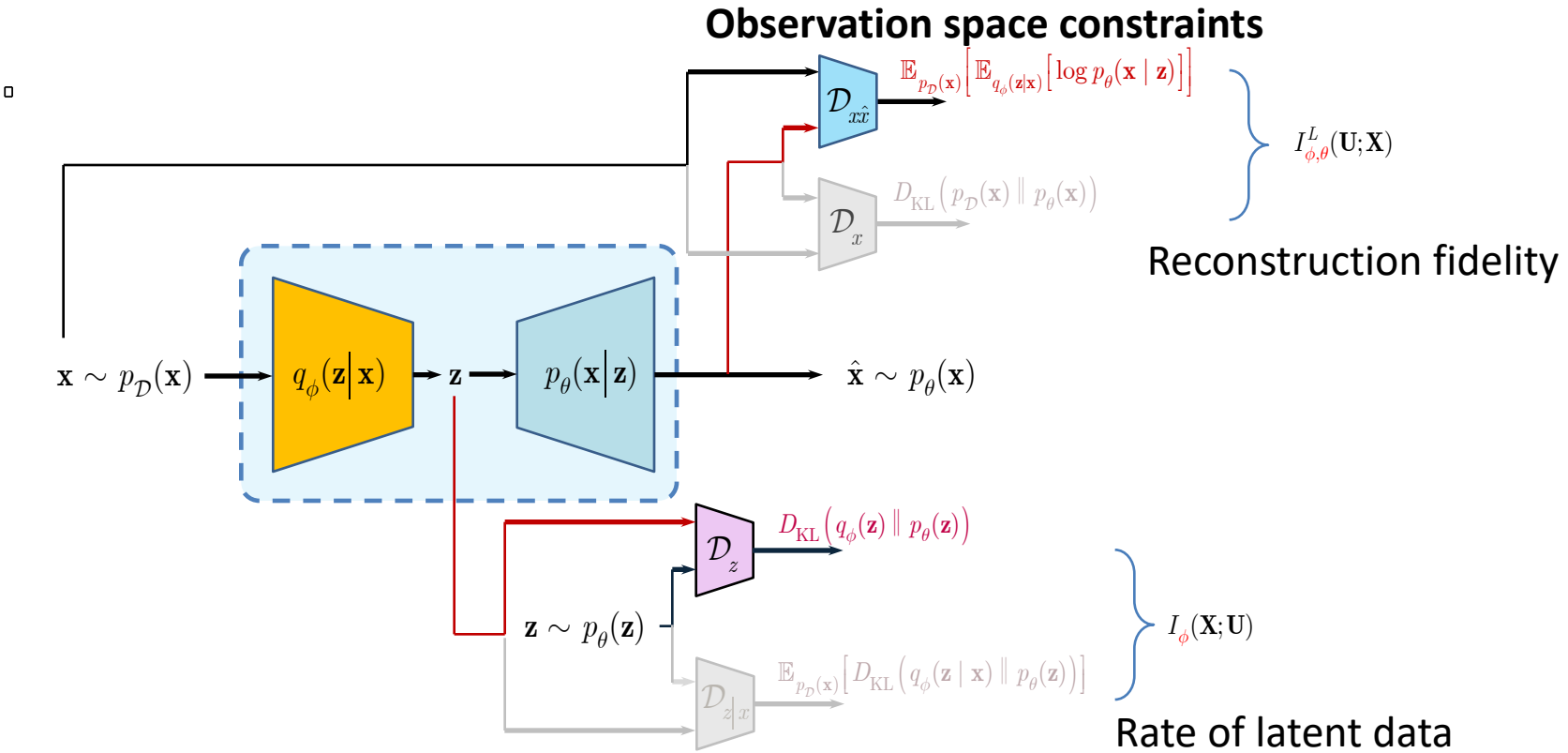
$$\mathcal{L}_{\text{VAE}}(\phi,\theta) = \mathbb{E}_{p_{\mathcal{D}}(\mathbf{x})}\Big[D_{\text{KL}}\Big(q_\phi(\mathbf{z}\mid\mathbf{x})\parallel p_\theta(\mathbf{z})\Big)\Big] - \beta\mathbb{E}_{p_{\mathcal{D}}(\mathbf{x})}\Big[\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\Big[\log p_\theta(\mathbf{x}\mid\mathbf{z})\Big]\Big]$$



Link to Information Bottleneck formulation

AAE: Adversarial Autoencoder

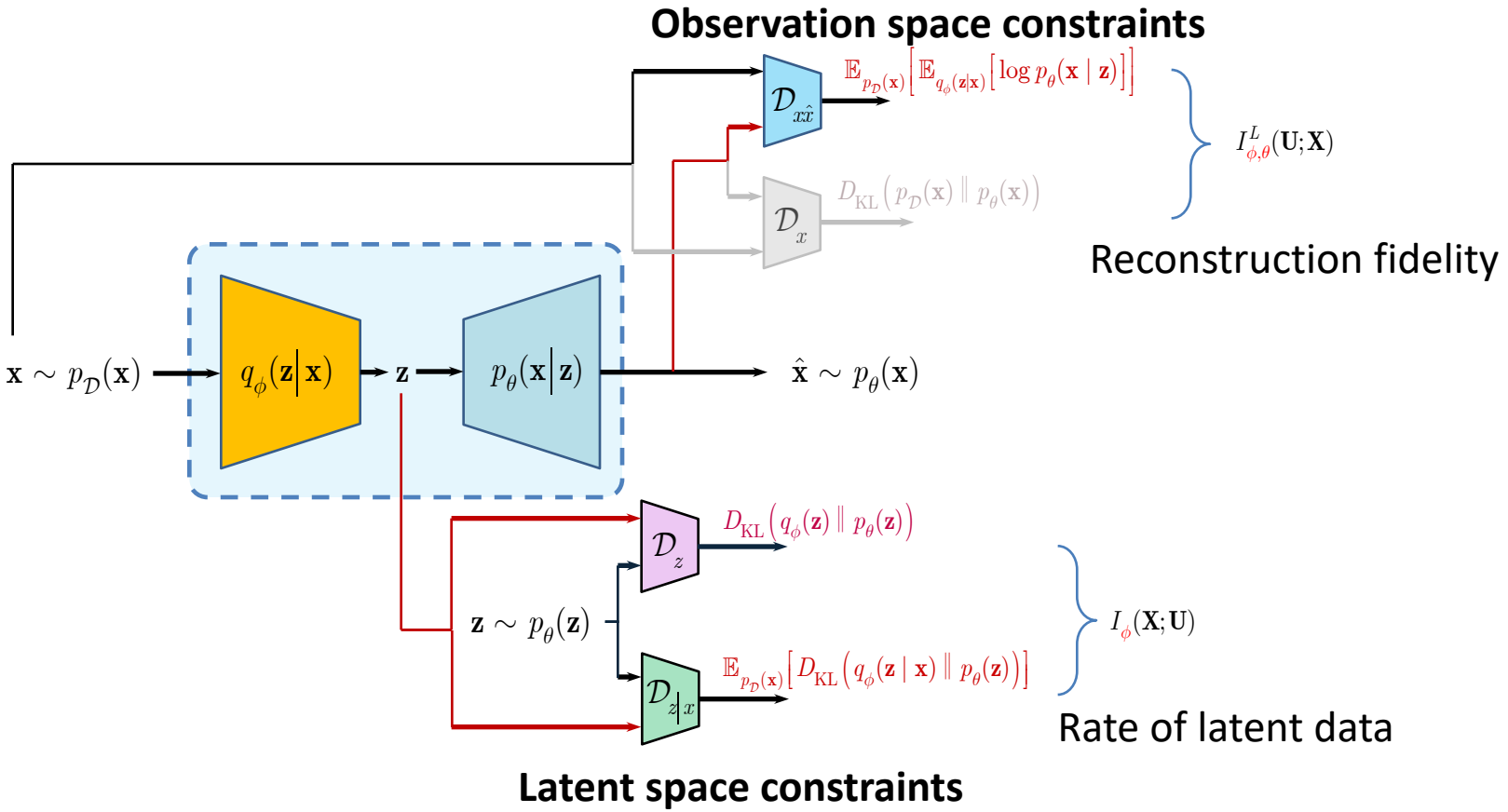
$$\mathcal{L}_{\text{AAE}}(\phi,\theta) = \mathbb{E}_{p_{\mathcal{D}}(\mathbf{x})}\Big[D_{\text{KL}}\big(q_\phi(\mathbf{z}) \parallel p_\theta(\mathbf{z})\big)\Big] - \beta\mathbb{E}_{p_{\mathcal{D}}(\mathbf{x})}\Big[\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\big[\log p_\theta(\mathbf{x} \mid \mathbf{z})\big]\Big]$$



Link to Information Bottleneck formulation

InfoVAE:

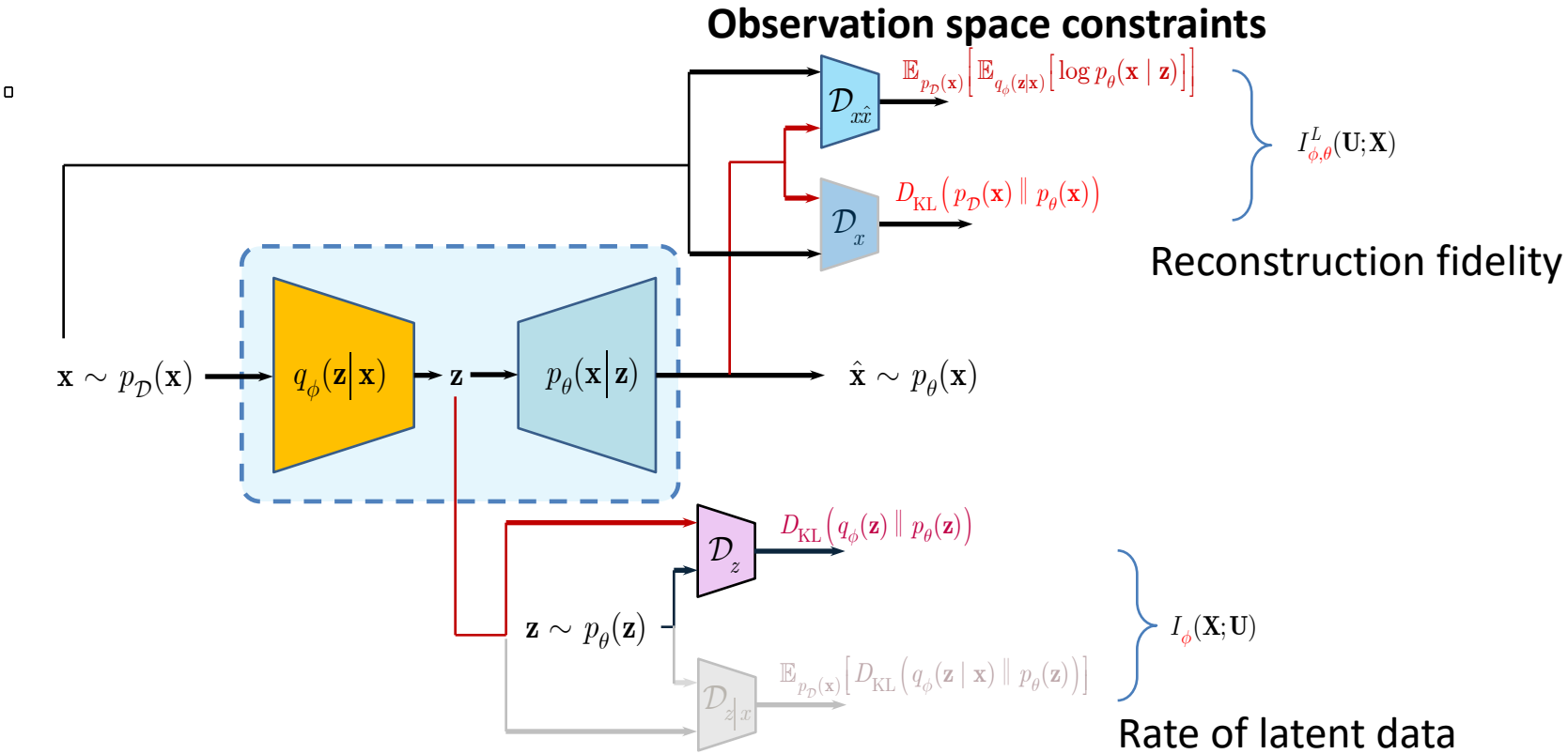
$$\mathcal{L}_{\text{InfoVAE}}(\phi,\theta) = I_{\phi}(\mathbf{X};\mathbf{U}) - \beta\mathbb{E}_{p_{\mathcal{D}}(\mathbf{x})}\Big[\mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})}\big[\log p_{\theta}(\mathbf{x}\mid\mathbf{z})\big]\Big]$$



Information Bottleneck based Novelty Detection

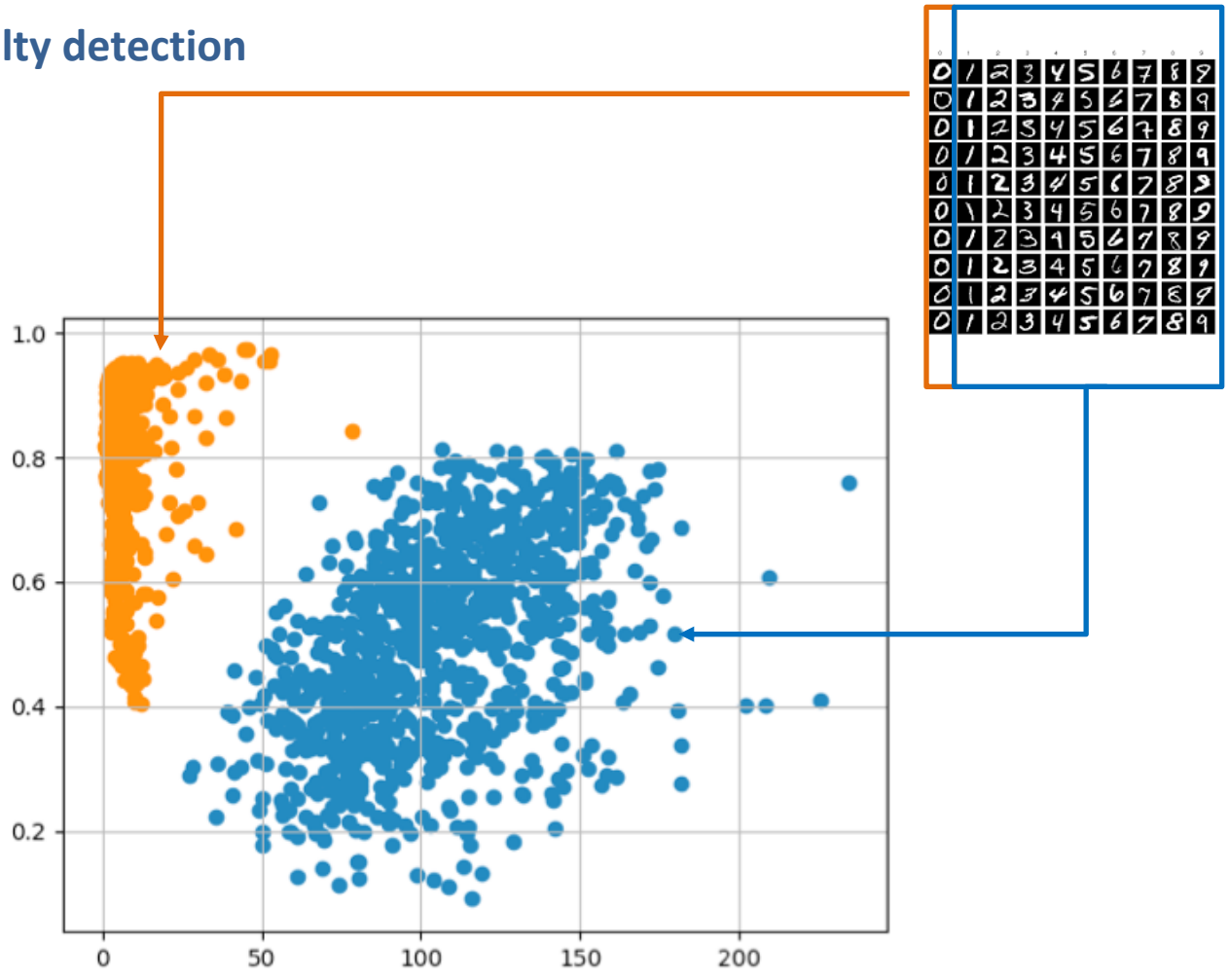
ND-AAE

$$\mathcal{L}_{ND-\mathrm{AAE}}(\phi,\theta) = \mathbb{E}_{p_{\mathcal{D}}(\mathbf{x})}\Big[D_{\mathrm{KL}}\big(q_{\phi}(\mathbf{z}) \parallel p_{\theta}(\mathbf{z})\big)\Big] - \beta\Big[\mathbb{E}_{p_{\mathcal{D}}(\mathbf{x})}\Big[\mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})}\big[\log p_{\theta}(\mathbf{x} \mid \mathbf{z})\big]\Big] + D_{\mathrm{KL}}\big(p_{\mathcal{D}}(\mathbf{x}) \parallel p_{\theta}(\mathbf{x})\big)\Big]$$



Link to Information Bottleneck formulation

Link to novelty detection



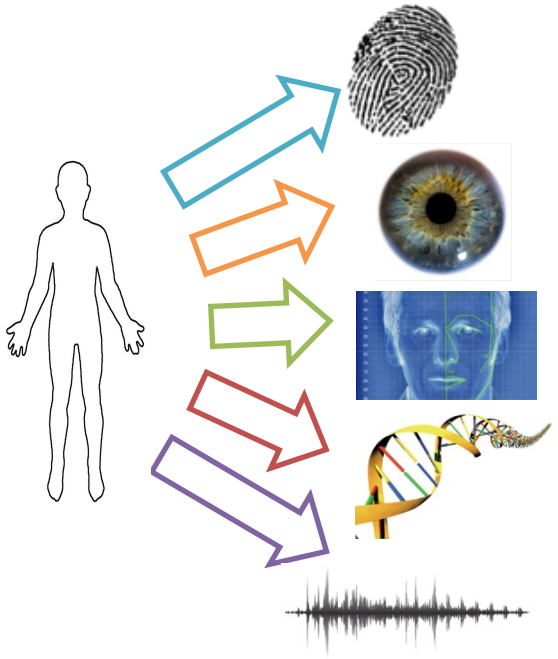
Applications

- Regression problems
  - Denoising, superresolution, compression, novelty detection
- **Search and indexing problems**
  - **Physical object security**
- Adversarial machine learning
  - Countermeasures

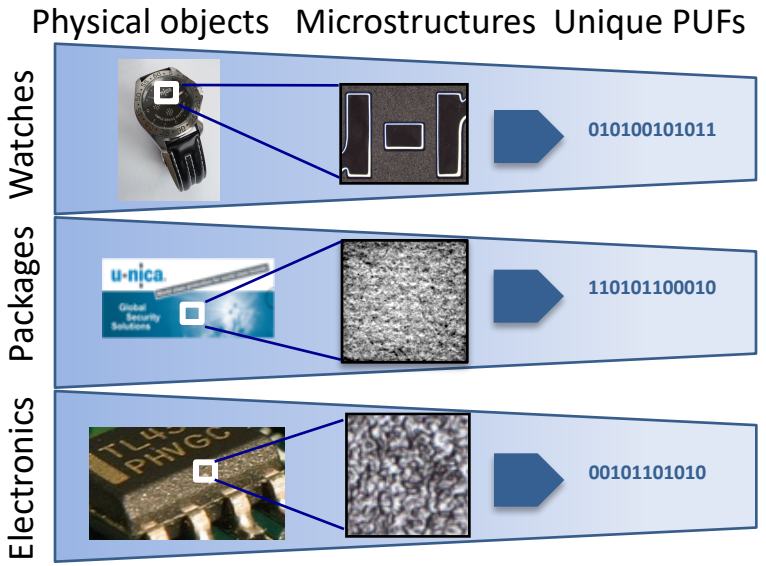
Object identification

Intuition behind physical uncloneable functions (PUFs)

Humans = biometrics



Physical Objects



All physical objects are unique like humans

Object identification

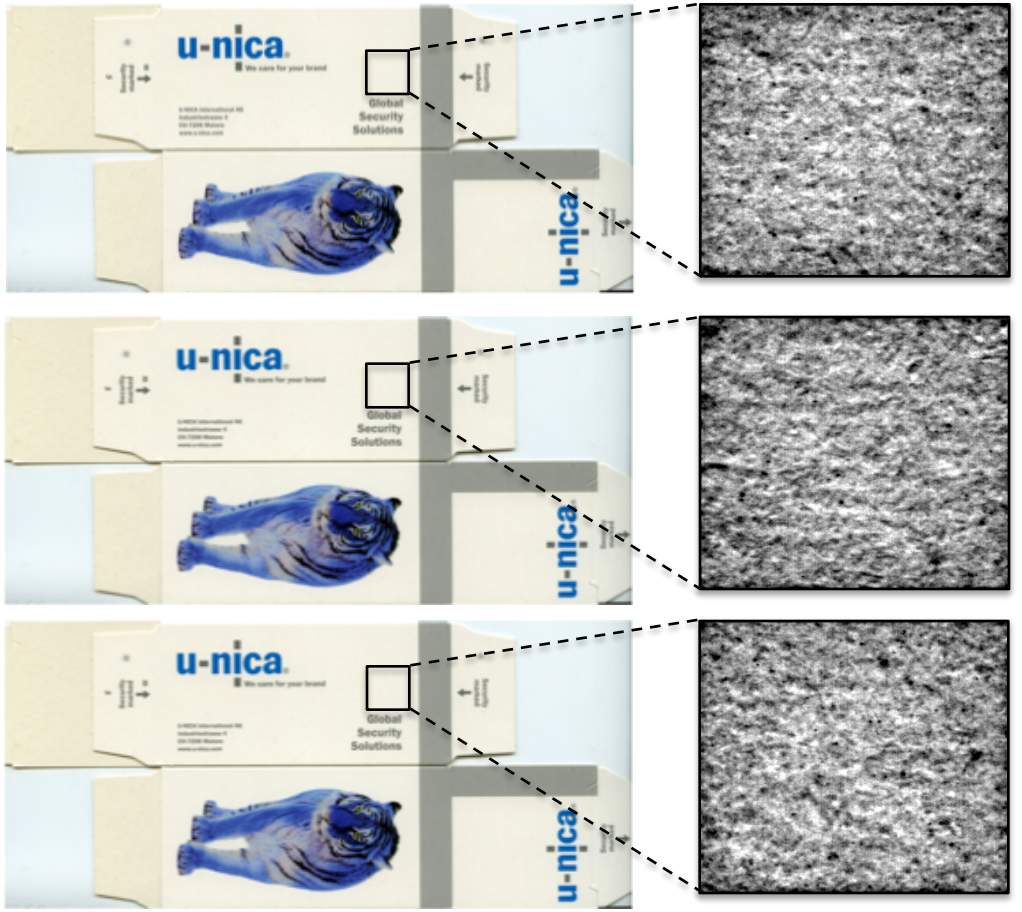
Paper microstructures =PUFs

Package 1

Package 2

.
.
.

Package M

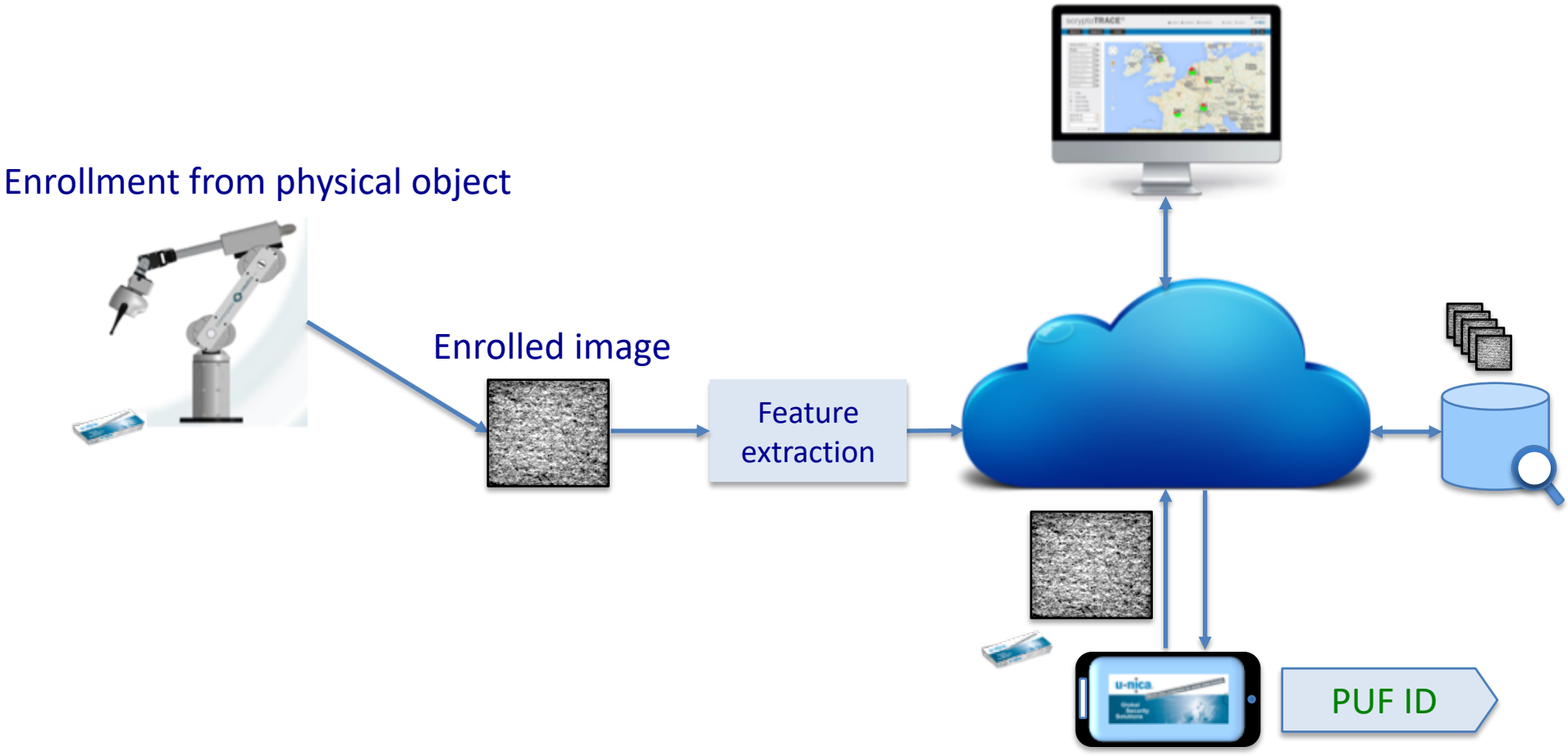


Individually unique PUFs

= unique identifier for Track&Trace

Visibly packages look identical

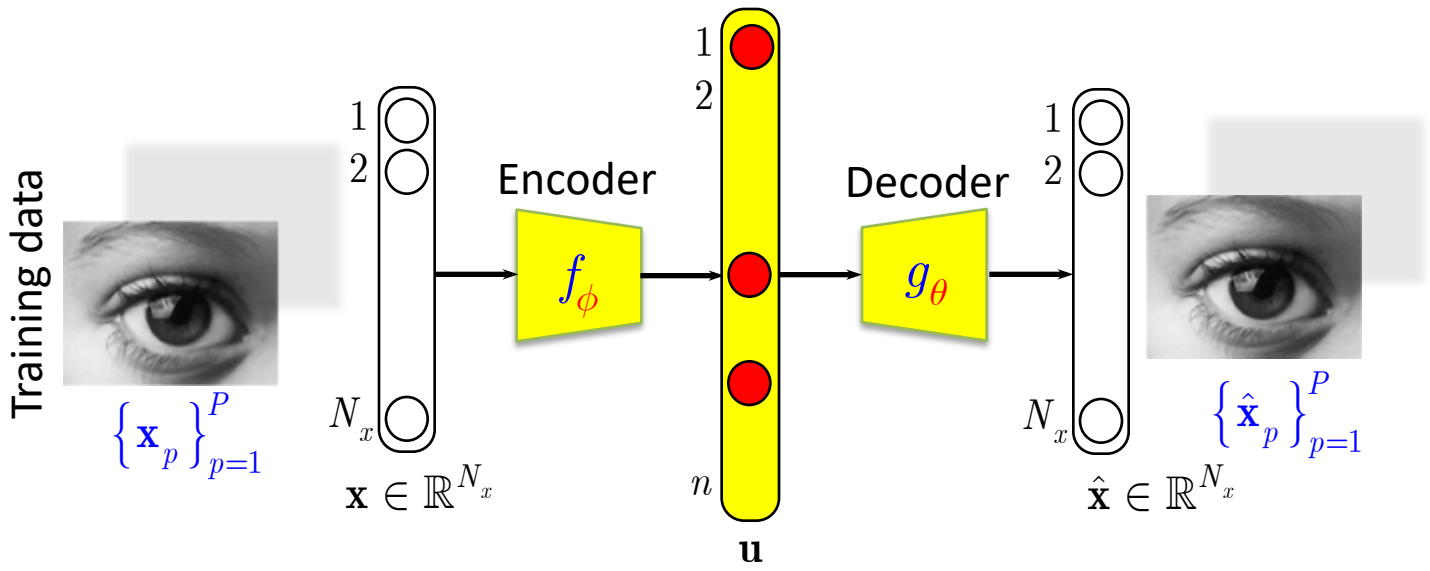
Object identification



Open issue:

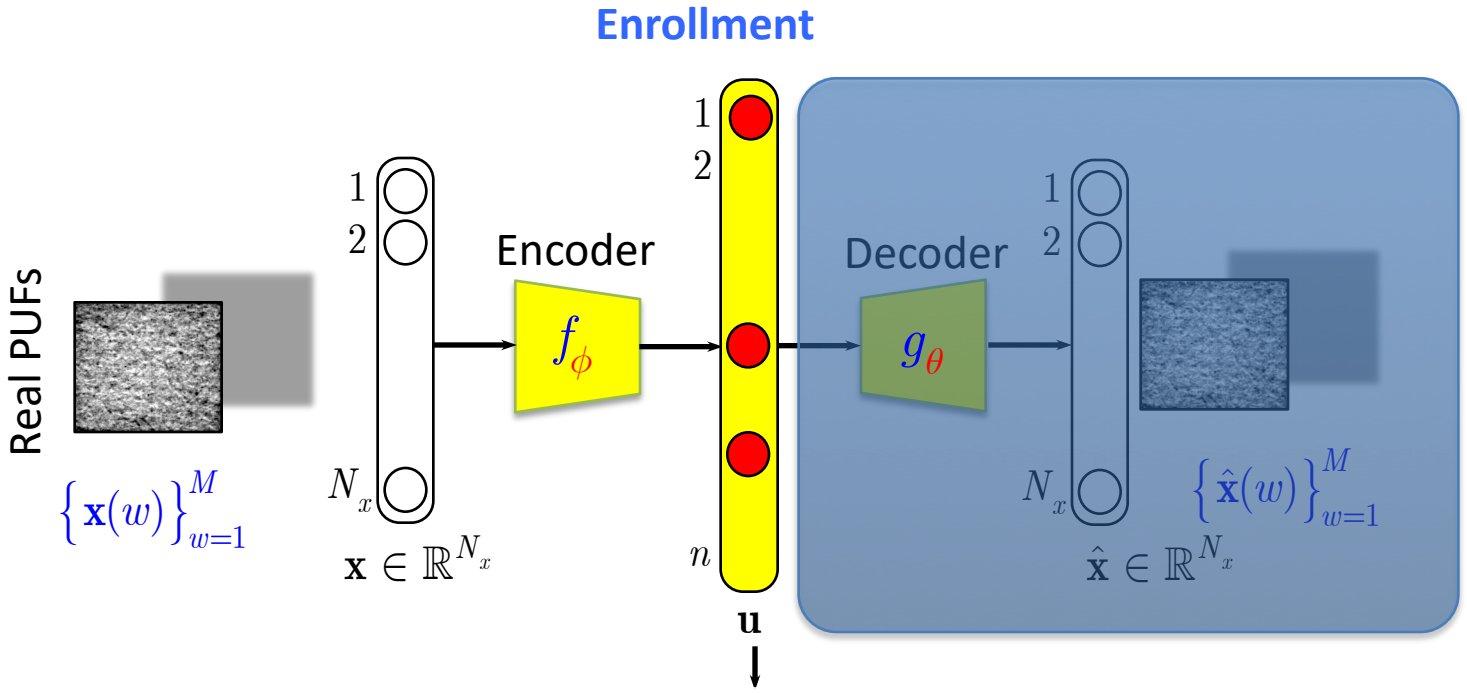
Big Data (millions of objects with high-dimensional features)

Fast search based on STC



Open issue:

Big Data (millions of objects with high-dimensional features)



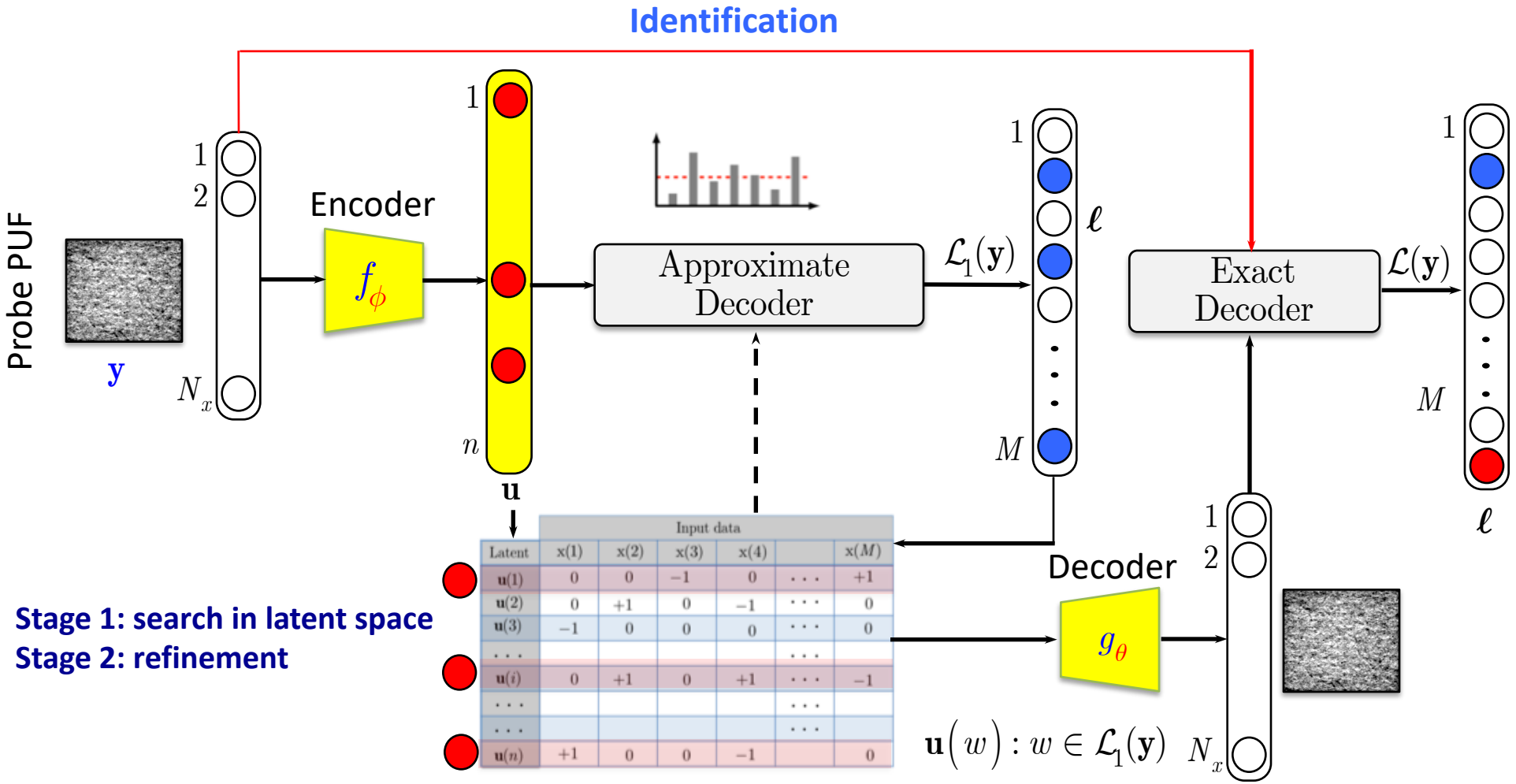
	Input data					
Latent	x(1)	x(2)	x(3)	x(4)		x(M)
u(1)	0	0	−1	0	· · ·	+1
u(2)	0	+1	0	−1	· · ·	0
u(3)	−1	0	0	0	· · ·	0
· · ·					· · ·	
u(i)	0	+1	0	+1	· · ·	−1
· · ·					· · ·	
· · ·					· · ·	
u(n)	+1	0	0	−1		0

Very sparse representation

- efficient storage
- suitable for fast indexing

Open issue:

Big Data (millions of objects with high-dimensional features)

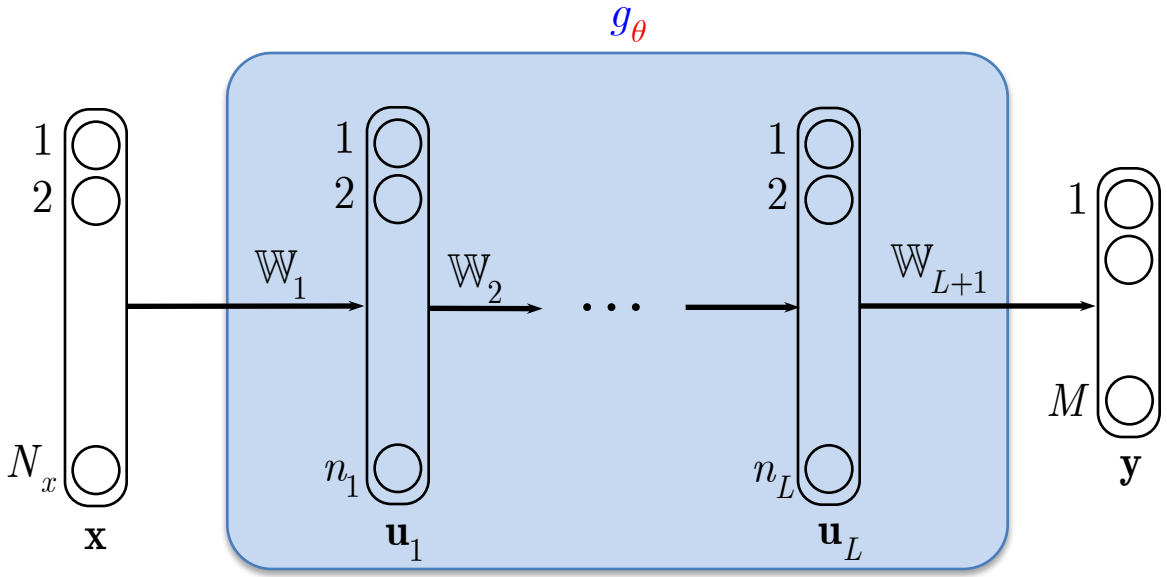


Applications

- Regression problems
  - Denoising, superresolution, compression
  - Learnable compressive sampling
- Search and indexing problems
  - Physical object security
- **Classification based on distributed VAEs**

Modern "framework" of classification

End-to-end training of classifiers



Classifier parameters

$$\theta = \left(\mathbb{W}_1, \mathbb{W}_2, \cdots \mathbb{W}_{L+1} \right)$$

Labels of classes

$$w \in \mathcal{W} = \left\{ 1, 2, \cdots, M \right\}$$
$$p \in \mathcal{P}_S = \left\{ 1, 2, \cdots, P_S \right\}$$

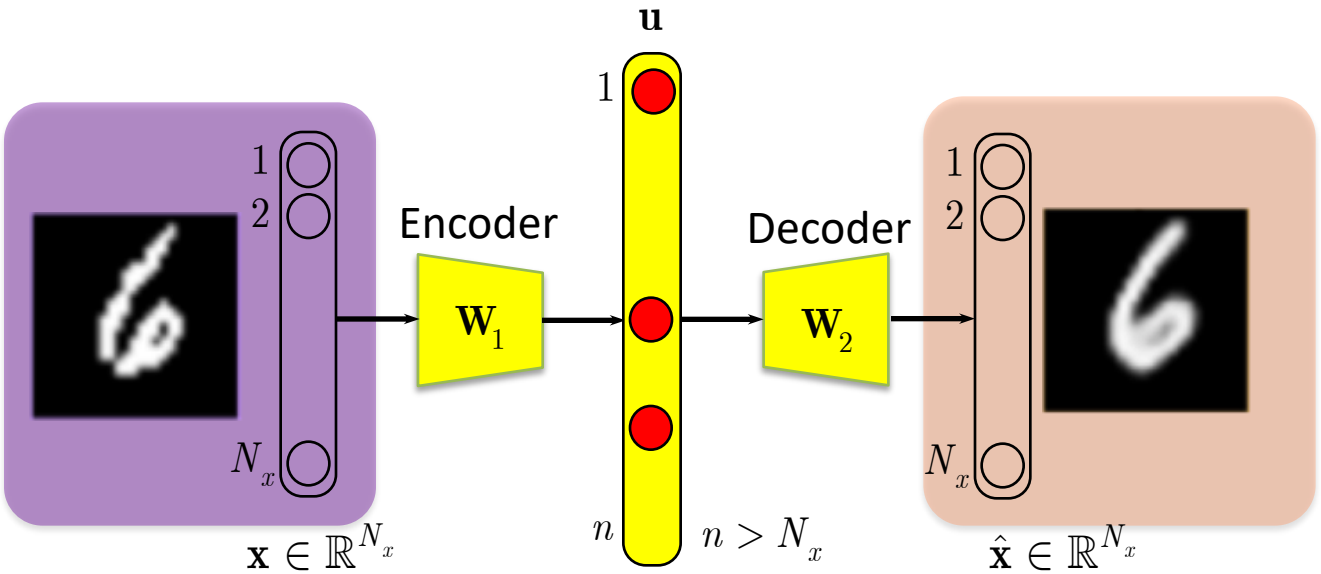
$$\hat{\theta} = \arg\min_{\theta} \sum_{w=1}^{M} \sum_{p=1}^{P_S} L\left(\mathbf{y}_p(w), g_\theta\left(\mathbf{x}_p(w) \right) \right) + \lambda \Omega\left(\theta \right)$$

$$g_\theta\left(\mathbf{x} \right) = \sigma_{L+1}\left(\mathbb{W}_{L+1} \ldots \sigma_1\left(\mathbb{W}_1 \mathbf{x} \right) \right)$$

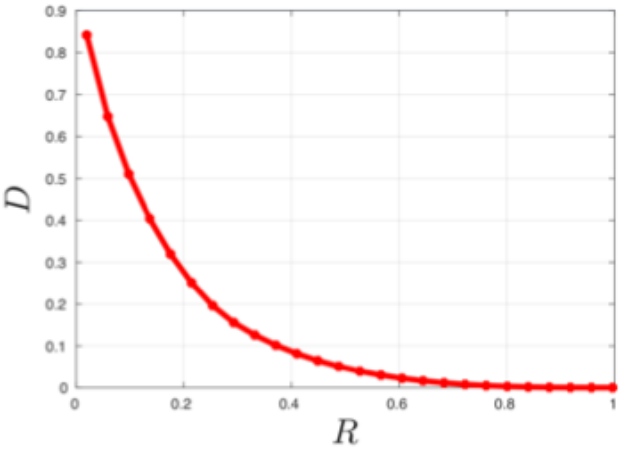
+ Fully "automated procedure" but:

- **a lot of training data are needed**
- **highly supervised**
- **"physics" of training process and nature of learned features are poorly understood**
- **vulnerable to adversarial attacks**

Proposed classification paradigm: classification by compression



Rate-distortion function



$$\mathbf{u} \in \{-1, 0, +1\}^{N_x}$$

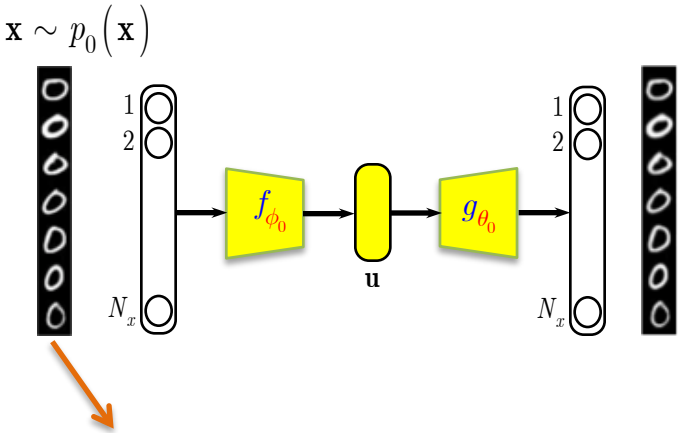
$$\|\mathbf{u}\|_0 \leq s$$

Distortion: $\frac{1}{N_x}\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 \leq D$

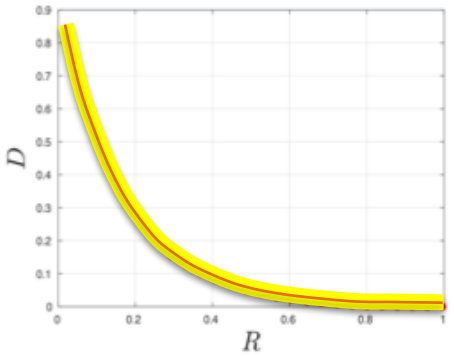
Rate: $R = \frac{1}{n}\log_2\left(\binom{n}{s}2^s\right)$

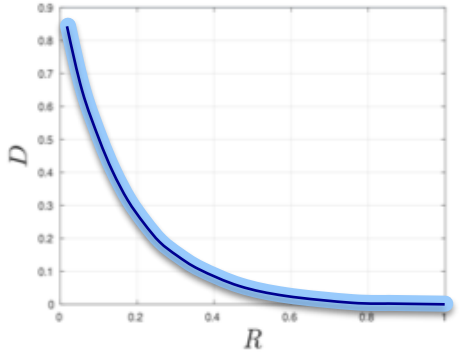
Compression as a "matched filter" for a given class

Proposed classification paradigm: classification by compression

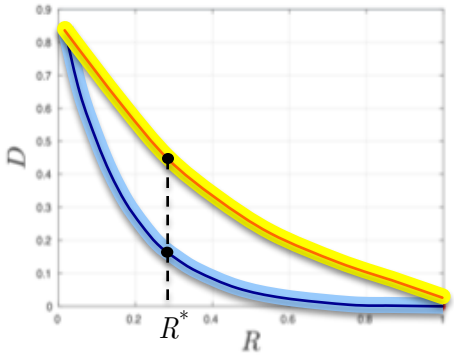


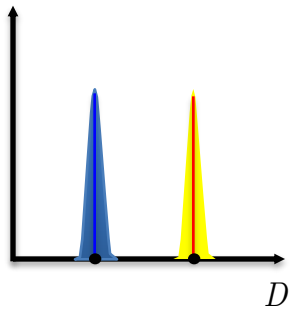
Matched case



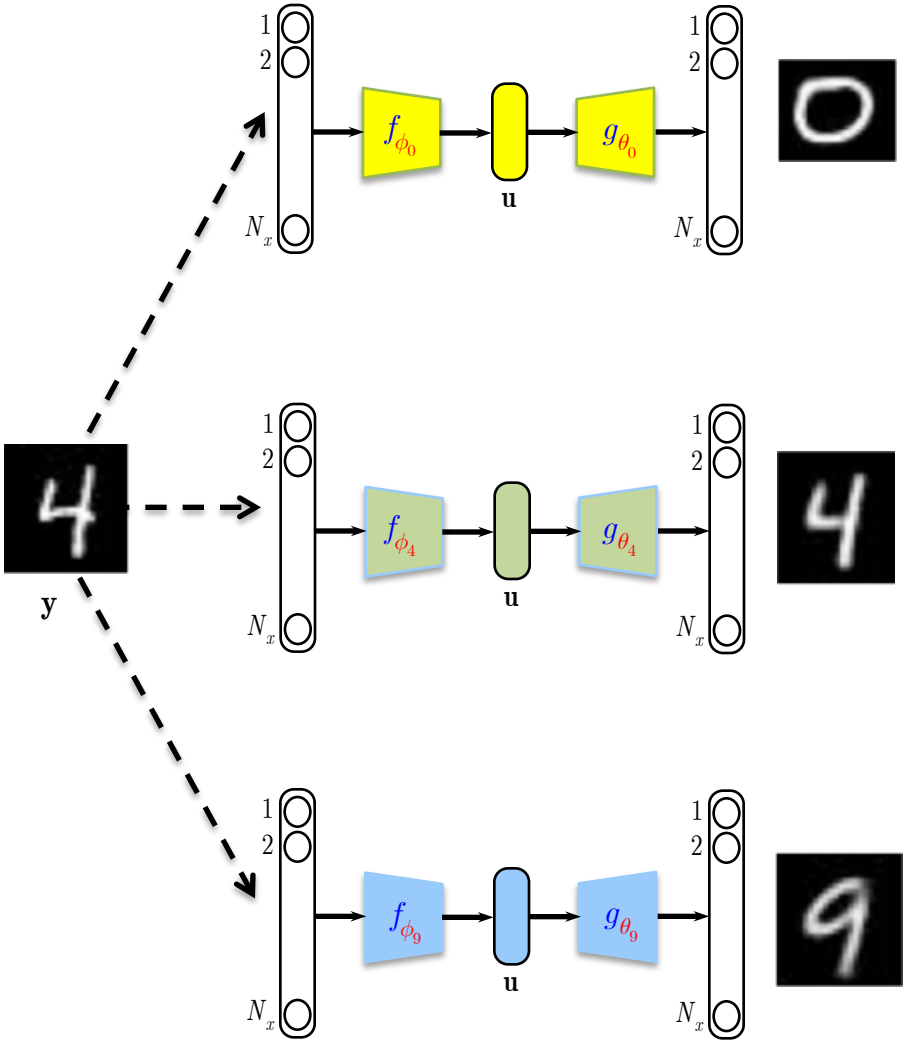


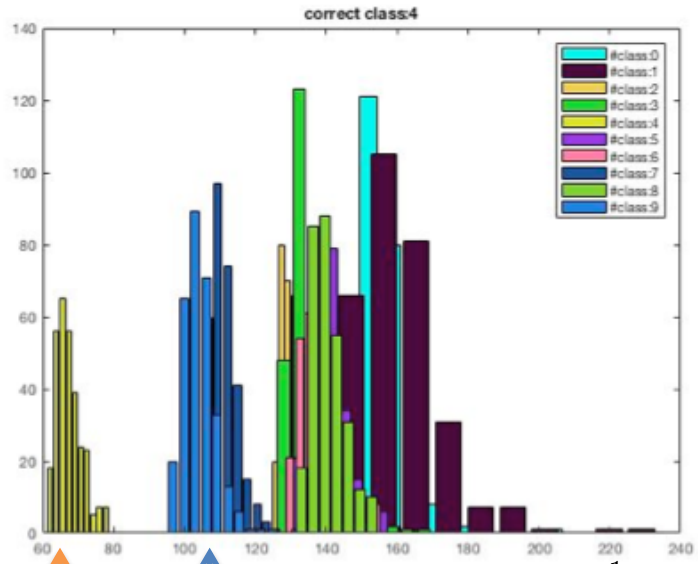
Mismatched case



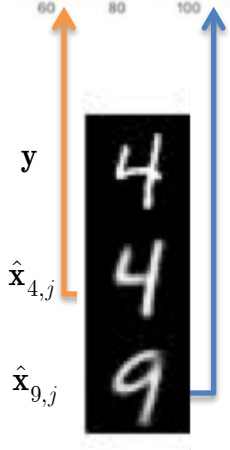


Proposed classification paradigm: classification by compression



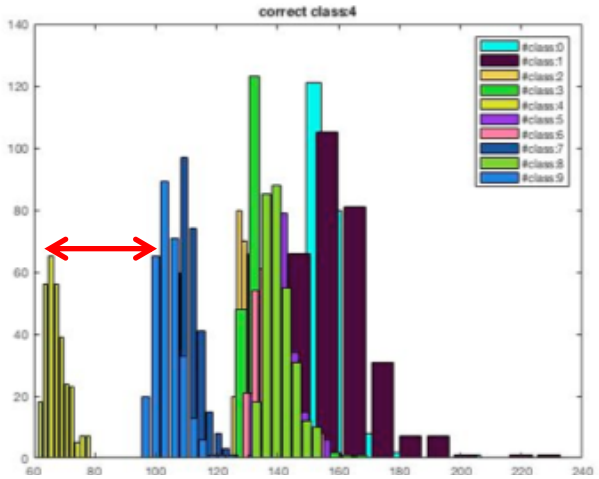


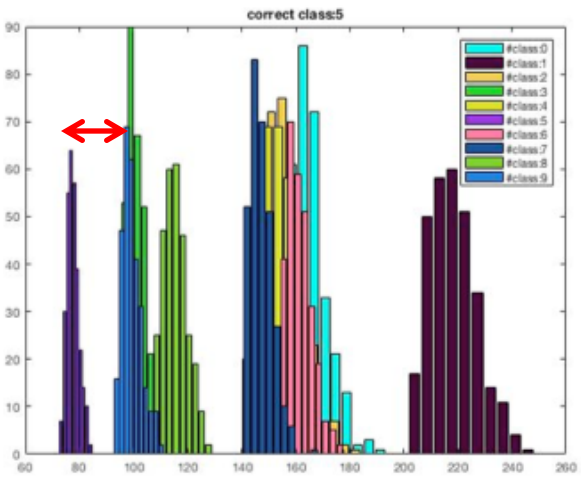
$$D = \frac{1}{N_x}\left\|\mathbf{y} - \hat{\mathbf{x}}_{w,j}\right\|_2$$

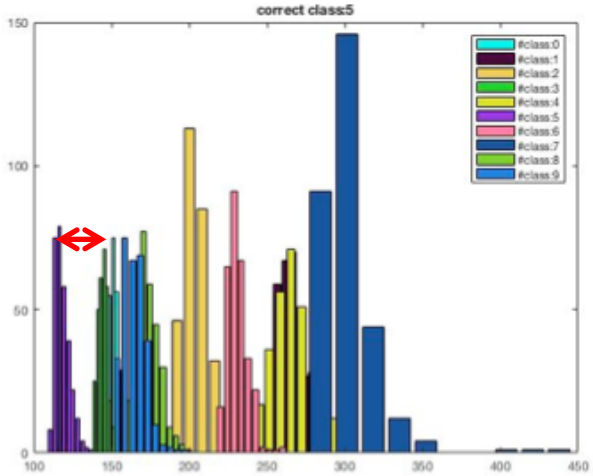


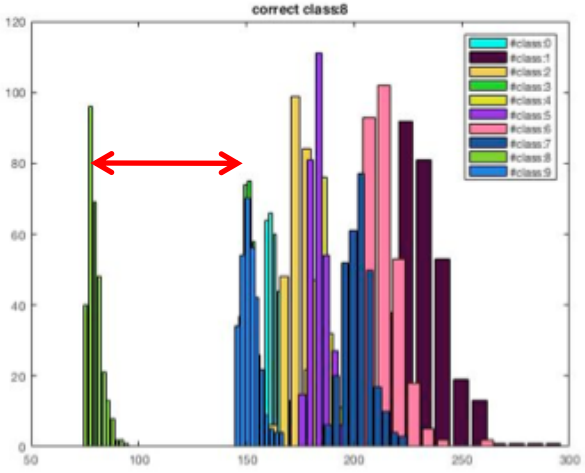
Proposed classification paradigm: classification by compression

Recognition: correct cases



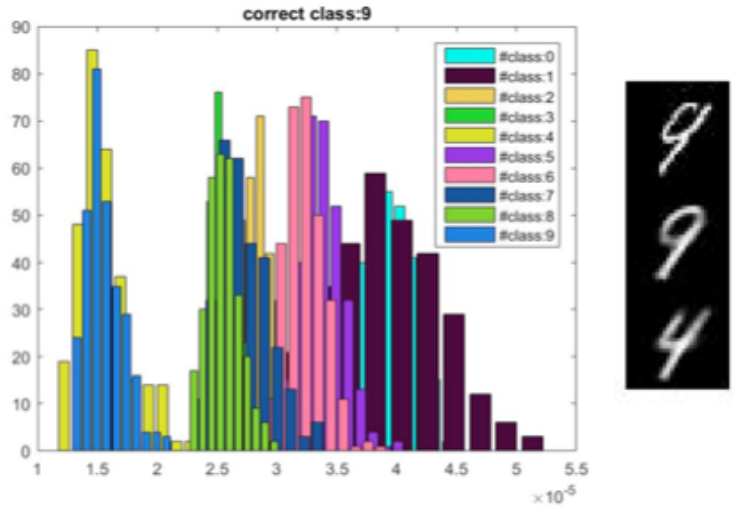


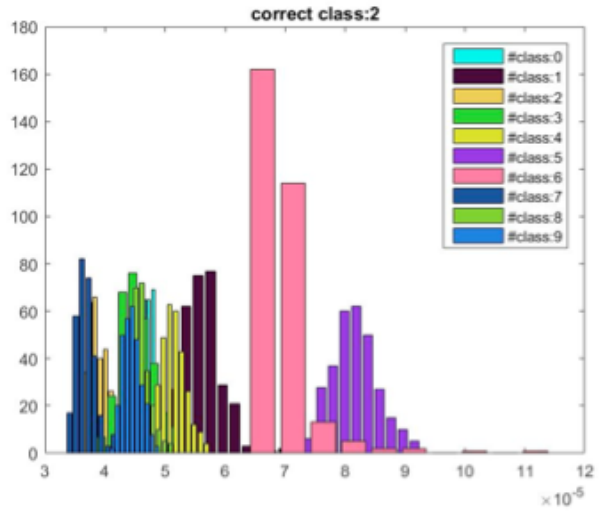




Proposed classification paradigm: classification by compression

Recognition: erroneous cases

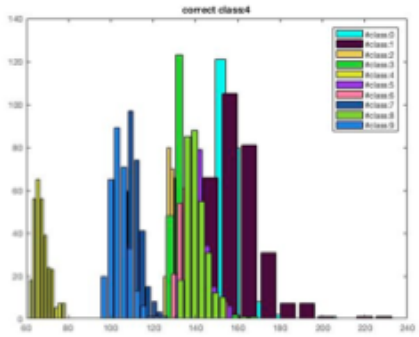




Proposed classification paradigm: classification by compression

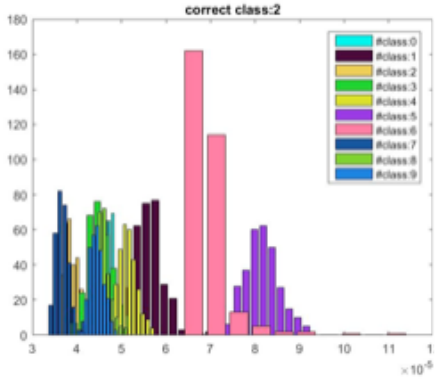
Rejection of “suspicious” and adversarial samples

Recognition: correct cases

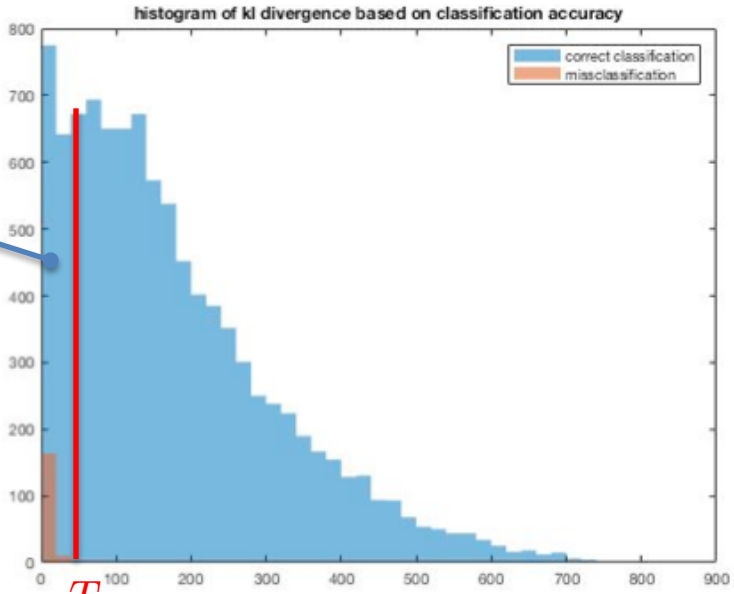


$D_{KL} \geq T$

Recognition: erroneous cases



$D_{KL} < T$



Proposed classification paradigm: classification by compression

Recognition: rejection of "abnormal cases"

Tab. 2: VAE results after 40 epochs of training

Thr.	$dim_z = 20$		$dim_z = 15$		$dim_z = 10$	
	Acc.	P_{miss}	Acc.	P_{miss}	Acc.	P_{miss}
1e-100	98.04	0	98.21	0	98.04	0.006
1e-50	98.04	0	98.21	2e-4	98.03	0.011
1e-10	98.04	0	98.21	9e-4	97.99	0.03
1	98.88	0.012	98.96	0.010	98.43	0.059
5	99.42	0.038	99.45	0.030	98.97	0.082
10	99.68	0.068	99.69	0.048	99.17	0.099
15	99.79	0.097	99.77	0.063	99.25	0.112
20	99.82	0.125	99.82	0.077	99.33	0.123
no rejection	98.05	-	98.21	-	98.05	-

Trust in results:
rejecting just 7.7%, one obtains 0.18% classification error

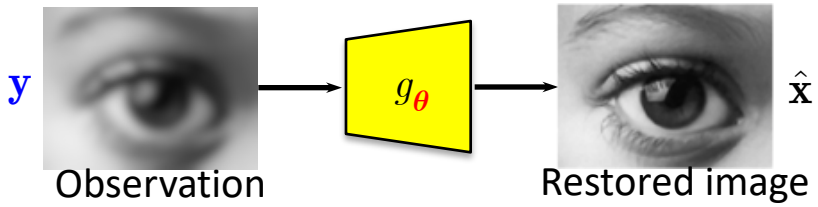
Interpretability of results:
A clear and meaningful interpretation of learned features

Summary

Regression

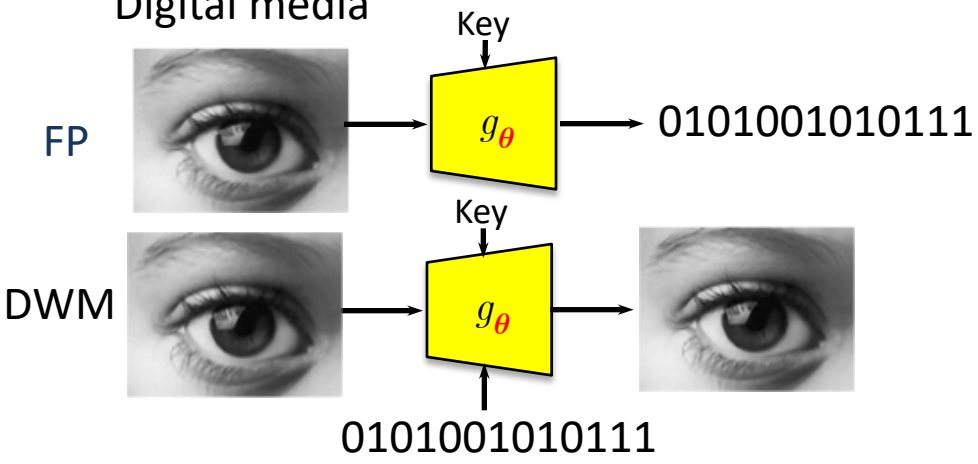
restoration
denoising
superresolution
compressive sensing

$$\mathbf{y} = \mathbf{T}[\mathbf{x}] + \mathbf{z}$$



Security

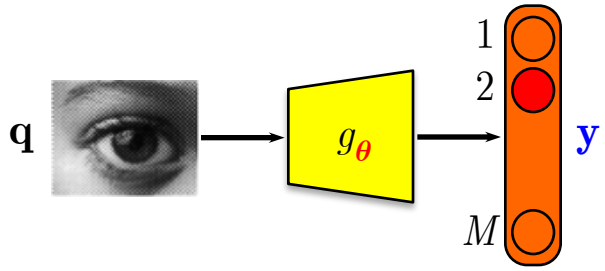
Physical objects and humans
Digital media



Classification

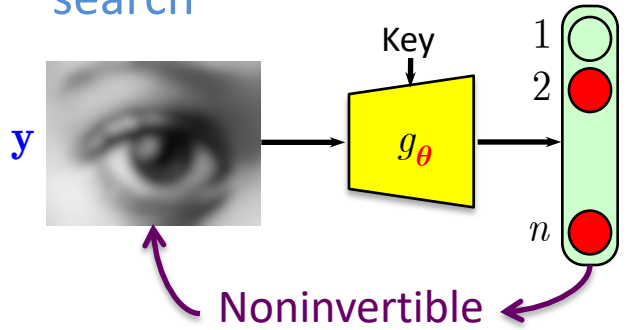
$$\{\mathbf{y}_p(w), \mathbf{x}_p(w)\}$$

$$w \in \{1, 2, \cdots M\}, p = 1, 2, \cdots, P$$



Privacy preserving applications

- identification
- authentication
- indexing
- search



Seniors

Dr. S.Ferdowsi

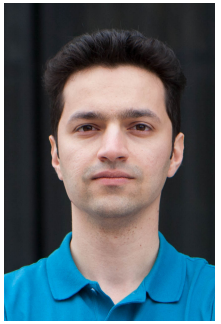
Dr. T. Holotyak

PhD students

S. Rezaeifar

O. Taran

D. Ullmann



B. Razeghi

S. Bonev


Thank You!