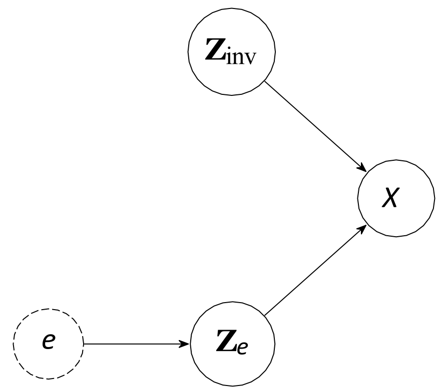
Unsupervised Representation Learning - an Invariant Risk Minimization Perspective

Yotam Norman Ron Meir

Technion - Israel Institute of Technology

Motivation

- Given different data sources (environments) $e \in \mathcal{E}$
- For each environment $X_e \sim P^e_X$.
- Previous work focused on supervised settings $(X, Y)_e \sim P^e_{X,Y}$.



- Question: Given a generative model, can invariant features $\mathbf{Z}_{\text{inv}}$ be extracted?

Unsupervised IRM Optimization Problem

- Optimization objective using maximum likelihood
- Constraining the solution to one with invariant featurization $\hat{\mathbf{Z}}_{\text{inv}} = \varphi(X)$:

$$\max_{\theta} \sum_{e \in \mathcal{E}_{train}} \log P^e_\theta(X, \phi(X))$$
$$\text{s.t. } P^i_\theta(\phi(X)) = P^j_\theta(\phi(X)) \quad \forall i, j \in \mathcal{E}_{train}$$

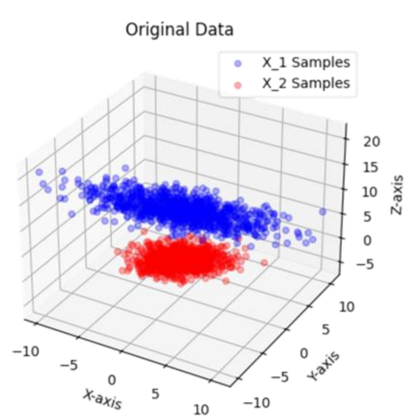
- The solution is designed to achieve the best reconstruction of the data based on invariant features

PICA Algorithm

- Gaussian linear case
- Similar to PCA, consider

$$\max_{u} \sum_{e \in \mathcal{E}} u^\top \Sigma^e_x u$$
$$\text{s.t. } \quad \|u\|_2^2 = 1, \quad u^\top \Sigma^i_x u = u^\top \Sigma^j_x u \quad \forall i, j \in \mathcal{E}_{\text{train}}$$

- We propose the "Principal Invariant Components Analysis" (PICA) algorithm
- Dimensionality reduction by eliminating "Environmental dimensions"

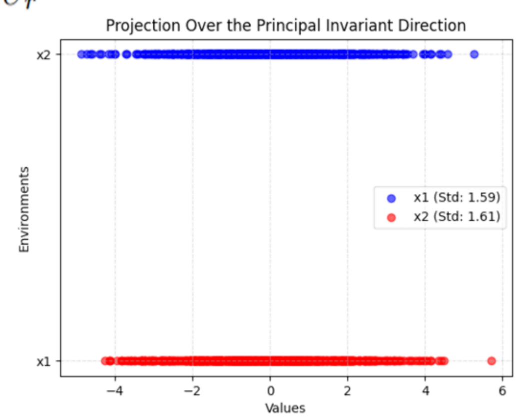


Algorithm 1 PICA

Input: Covariance matrices Σ^1_x, Σ^2_x, desired reduced invariant dimension $d_r \leq d_I$
Output: d_r invariant principal directions U_r

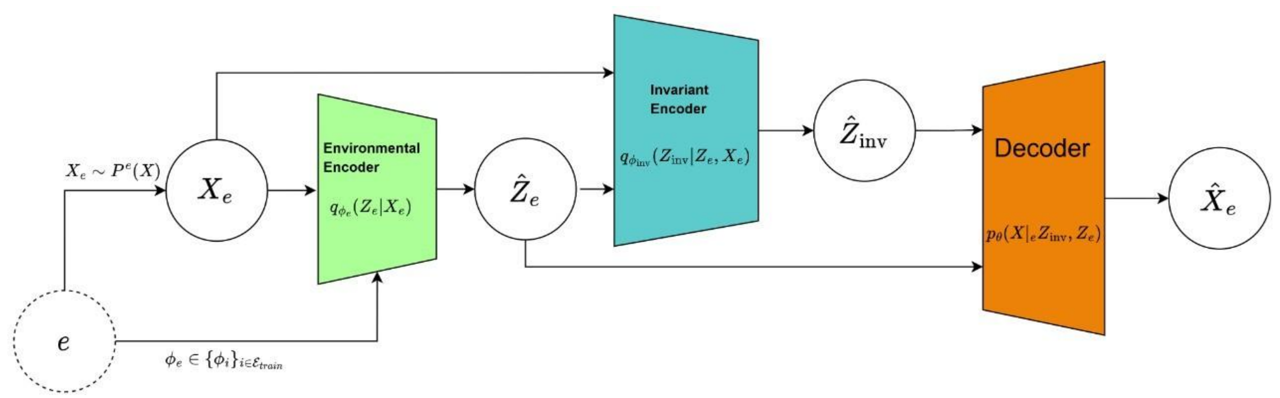
$U_r = \emptyset$
find: $U = \{u \in \mathbb{R}^n | (\Sigma^1_x - \Sigma^2_x)u = 0\}$

for i = 1 to d_r **do**
  $u^* = \max_{u \in U} u^\top (\Sigma^1_x + \Sigma^2_x) u$
  $U_r = U_r \cup \{u^*\}$
  $U = U \setminus \{u^*\}$
end for
End Procedure

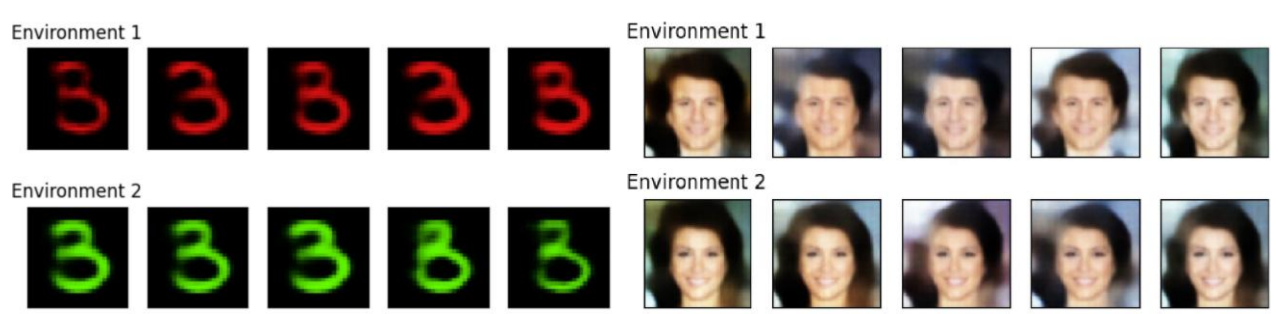


VIAE Algorithm

- "Variational Invariant Auto Encoder", proposed algorithm to solve the UIRM problem.
- Based on the VAE architecture
- Specially designed to recover the assumed generative model



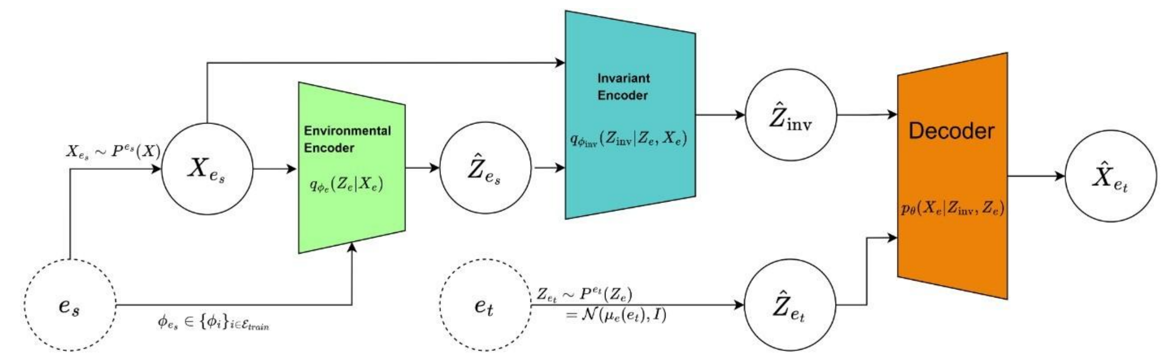
- Sampling Z_{inv} and Z_e independently
- Generates the same sample in different environments



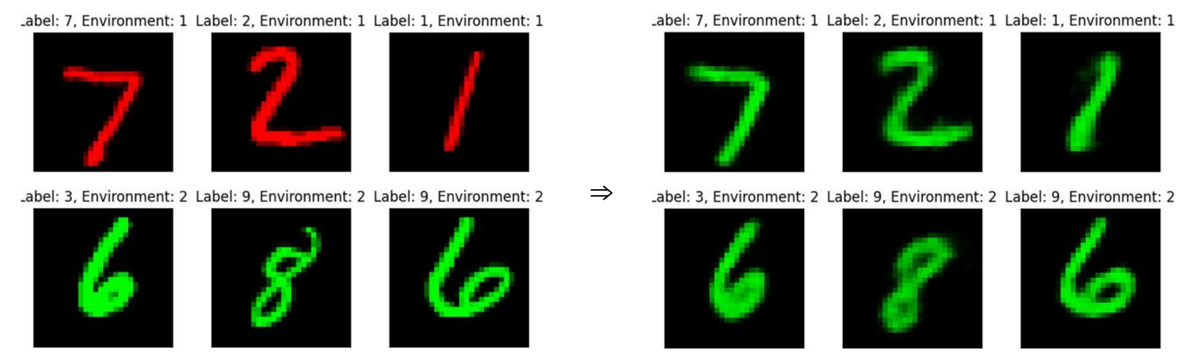
SCMNIST CelebA

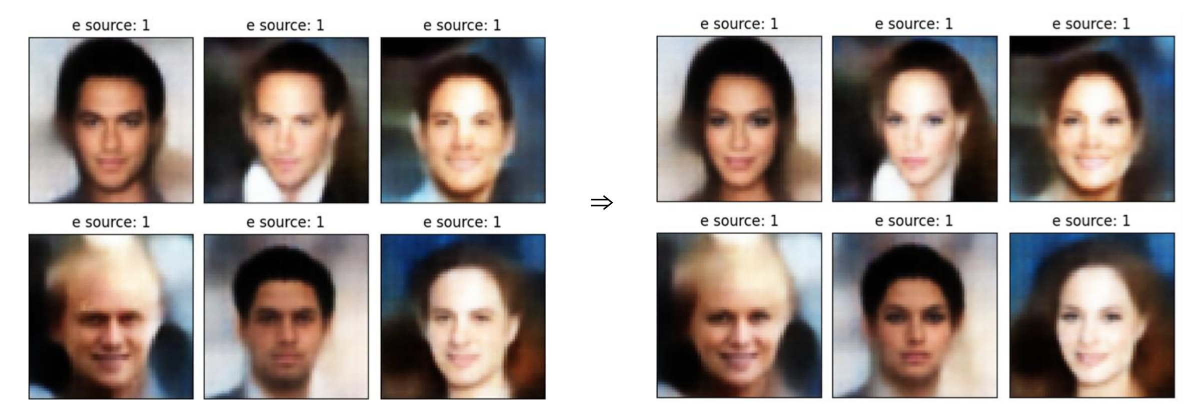
Environment Transfer

- Can we generate an invariant sample?
- If there was only one environment, could we consider the UIRM problem as solved?
- VAIE enables us to "transfer" a sample from one environment to the other



- Use the Encoders to recover the Invariant features, then combine them with generated environmental features
- The outcome from the decoder is a "transferred" sample





Conclusions & Future Work

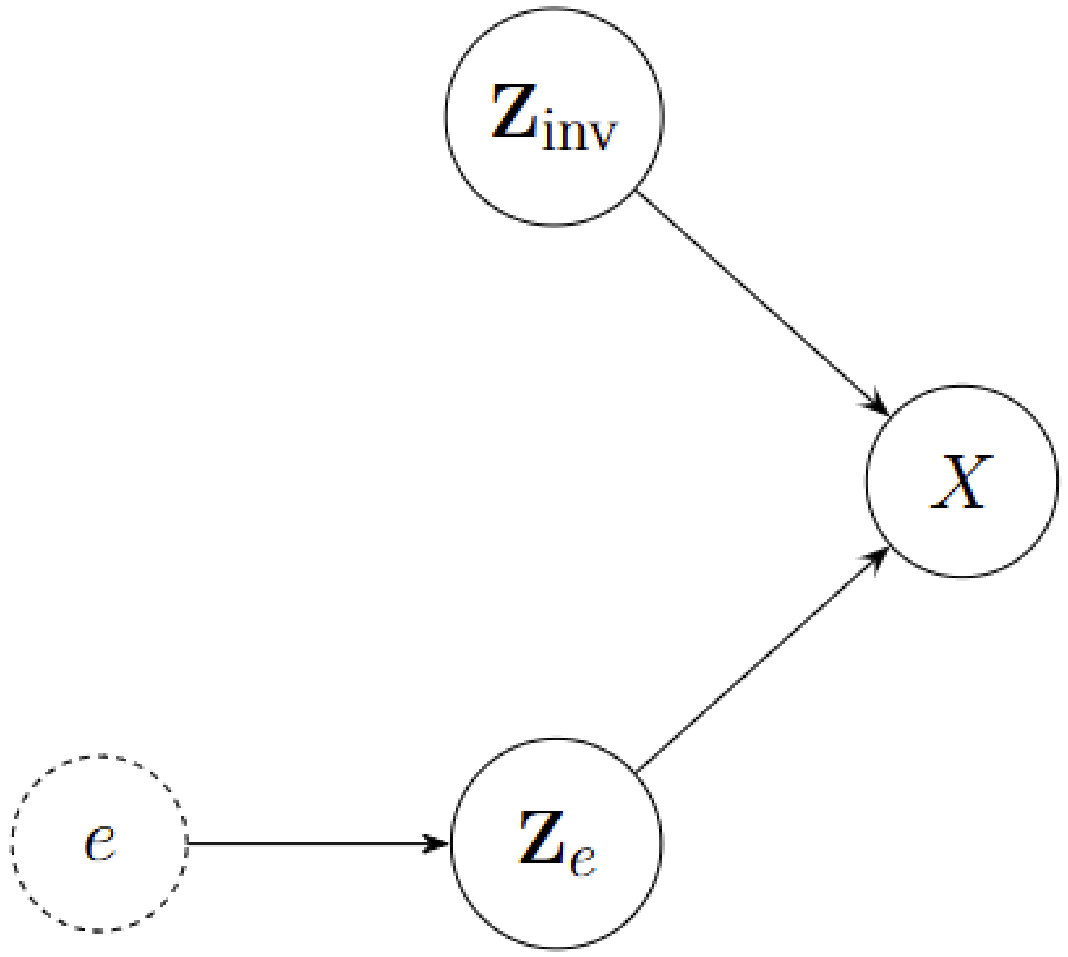
- We offer an unsupervised approach for the IRM OOD problem.
- VIAE shows promising results in tackling this problem.
- Future work: Can environment transfer work for unknown environments in the zero-shot case?

Yotam Norman Ron Meir

Technion - Israel Institute of Technology

Motivation

- Given different data sources (environments) $e \in \mathcal{E}$.
- For each environment $X_e \sim P_X^e$.
- Previous work focused on supervised settings $(X, Y)_e \sim P_{X,Y}^e$.



- **Question:** Given a generative model, can invariant features $\mathbf{Z}_{\text{inv}}$ be extracted?

Yotam Norman Ron Meir
Technion - Israel Institute of Technology

Unsupervised IRM Optimization Problem

- Optimization objective using maximum likelihood
- Constraining the solution to one with invariant featurization $\hat{\mathbf{Z}}_{\text{inv}} = \phi(X)$:

$$\max_{\theta} \sum_{e \in \mathcal{E}_{train}} \log P_{\theta}^{e}(X, \phi(X))$$
$$\text{s.t. } P_{\theta}^{i}(\phi(X)) = P_{\theta}^{j}(\phi(X)) \quad \forall i, j \in \mathcal{E}_{train}$$

Yotam Norman Ron Meir
Technion - Israel Institute of Technology

PICA Algorithm

- Gaussian linear case
- Similar to PCA, consider

$$\max_{u} \sum_{e \in \mathcal{E}} u^\top \Sigma_x^e u$$
$$\text{s.t.} \quad ||u||_2^2 = 1, \quad u^\top \Sigma_x^i u = u^\top \Sigma_x^j u \ \ \forall i, j \in \mathcal{E}_{\text{train}}.$$

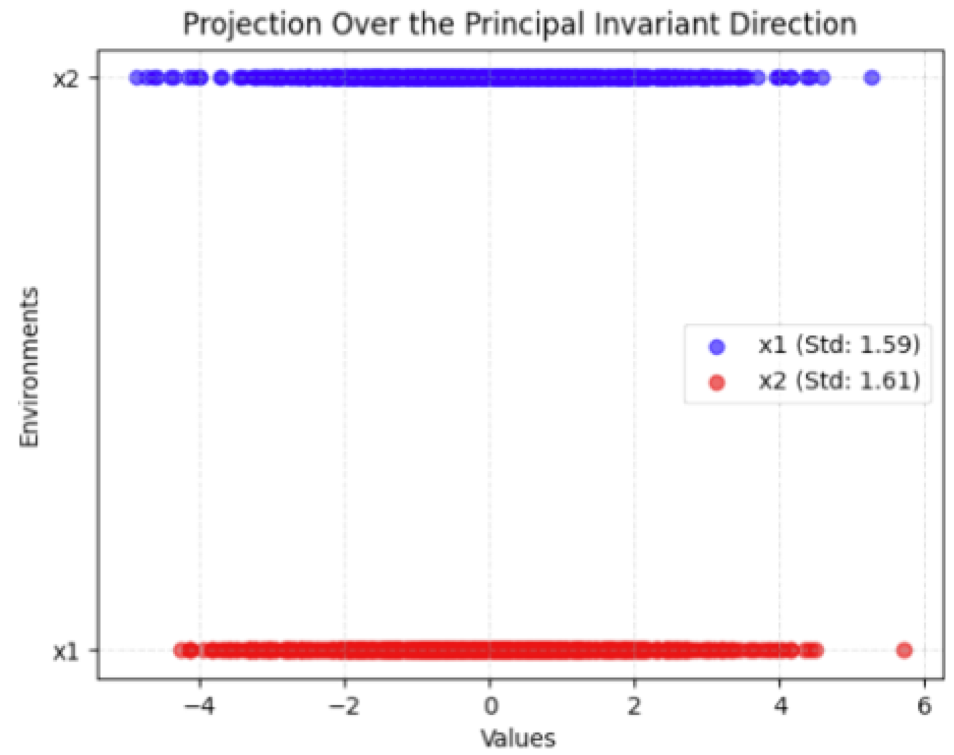
- We propose the "Principal Invariant Components Analysis" (PICA) algorithm
- Dimensionality reduction by eliminating "Environmental dimensions"

Algorithm 1 PICA

Input: Covariance matrices Σ_x^1, Σ_x^2. desired reduced invariant dimension $d_r \leq d_I$
Output: d_r invariant principal directions U_r

$U_r = \emptyset$
find: $U = \{u \in \mathbb{R}^n | (\Sigma_x^1 - \Sigma_x^2)u = 0\}$

for i = 1 to d_r **do**
  $u^* = \max_{u \in U} \ u^\top \left(\Sigma_x^1 + \Sigma_x^2\right) u$
  $U_r = U_r \cup \{u^*\}$
  $U = U \setminus \{u^*\}$
end for
End Procedure



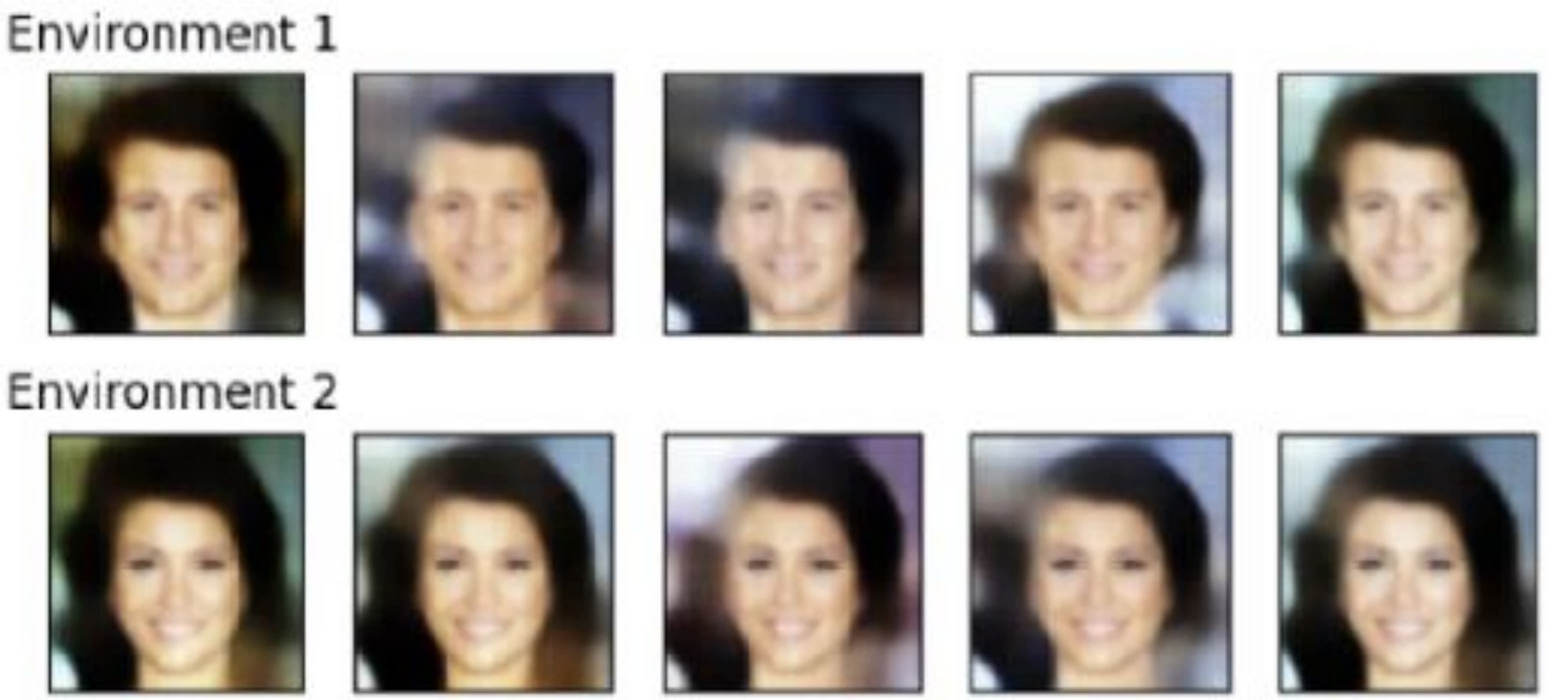
Yotam Norman Ron Meir

Technion - Israel Institute of Technology

VIAE Algorithm

- Sampling Z_{inv} and Z_e independently
- Generates the same sample in different environments

SCMNIST



CelebA

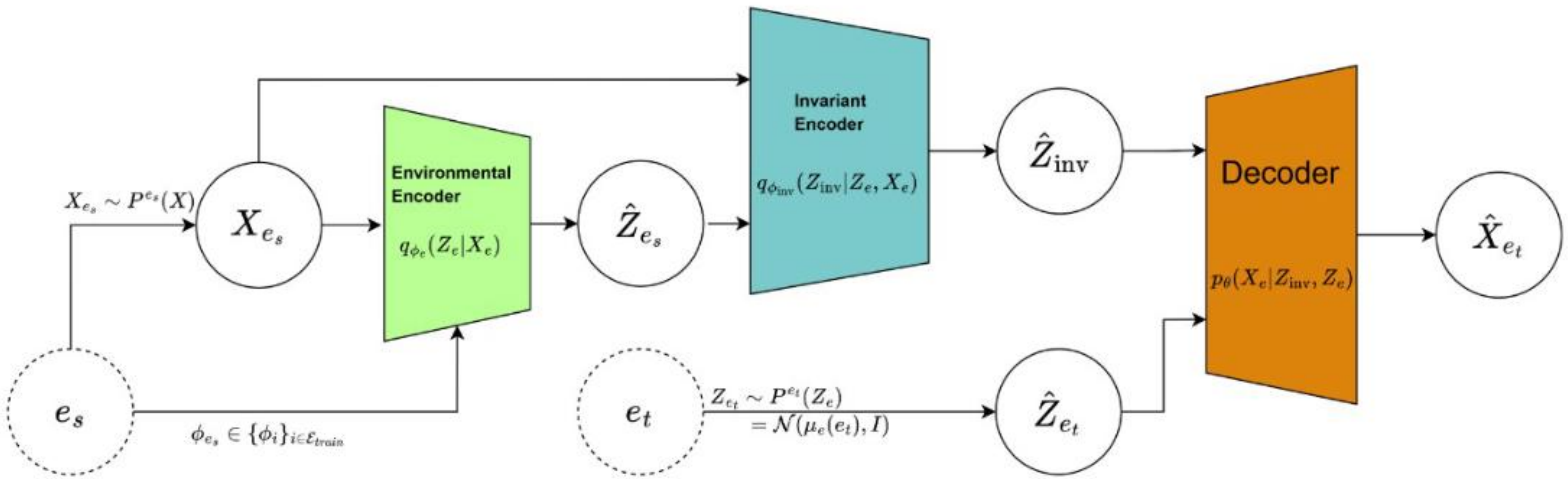
Unsupervised Representation Learning - an Invariant Risk Minimization Perspective

Yotam Norman Ron Meir

Technion - Israel Institute of Technology

Environment Transfer

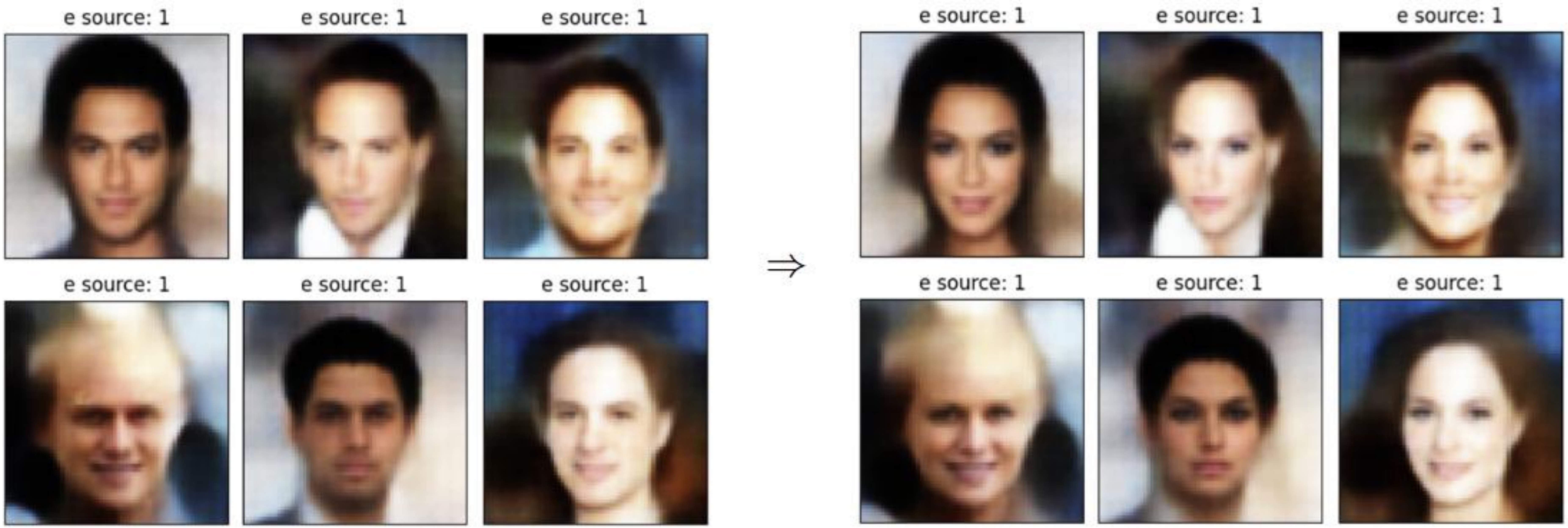
- If there was only one environment, could we consider the UIRM problem as solved?
- VIAE enables us to “transfer” a sample from one environment to the other



- Use the Encoders to recover the Invariant features, then combine them with generated environmental features
- The outcome from the decoder is a “transferred” sample

Yotam Norman Ron Meir
Technion - Israel Institute of Technology

Environment Transfer



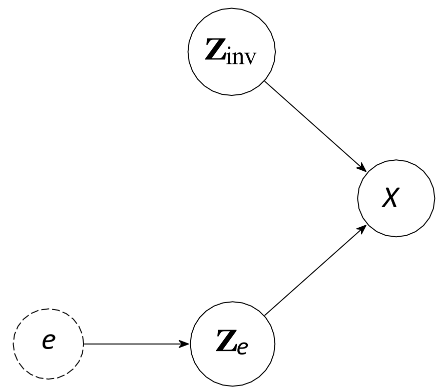
Unsupervised Representation Learning - an Invariant Risk Minimization Perspective

Yotam Norman   Ron Meir

Technion - Israel Institute of Technology

Motivation

- Given different data sources (environments) $e \in \mathcal{E}$
- For each environment $X_e \sim P_X^e$.
- Previous work focused on supervised settings $(X, Y)_e \sim P_{X,Y}^e$.



- Question: Given a generative model, can invariant features $\mathbf{Z}_{\text{inv}}$ be extracted?

Unsupervised IRM Optimization Problem

- Optimization objective using maximum likelihood
- Constraining the solution to one with invariant featurization $\hat{\mathbf{Z}}_{\text{inv}} = \varphi(X)$:

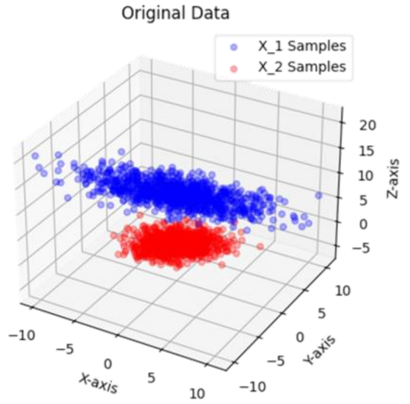
$$\max_{\theta} \sum_{e \in \mathcal{E}_{train}} \log P_\theta^e(X, \phi(X))$$
$$\text{s.t. } P_\theta^i(\phi(X)) = P_\theta^j(\phi(X)) \quad \forall i, j \in \mathcal{E}_{train}$$

- The solution is designed to achieve the best reconstruction of the data based on invariant features

PICA Algorithm

- Gaussian linear case
- Similar to PCA, consider

$$\max_{u} \sum_{e \in \mathcal{E}} u^\top \Sigma_x^e u$$
$$\text{s.t. } \quad ||u||_2^2 = 1, \quad u^\top \Sigma_x^i u = u^\top \Sigma_x^j u \quad \forall i, j \in \mathcal{E}_{\text{train}}$$



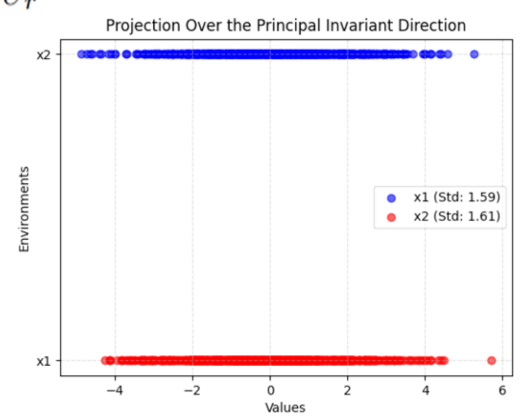
- We propose the "Principal Invariant Components Analysis" (PICA) algorithm
- Dimensionality reduction by eliminating "Environmental dimensions"

Algorithm 1 PICA

Input: Covariance matrices Σ_x^1, Σ_x^2. desired reduced invariant dimension $d_r \le d_I$
Output: d_r invariant principal directions U_r

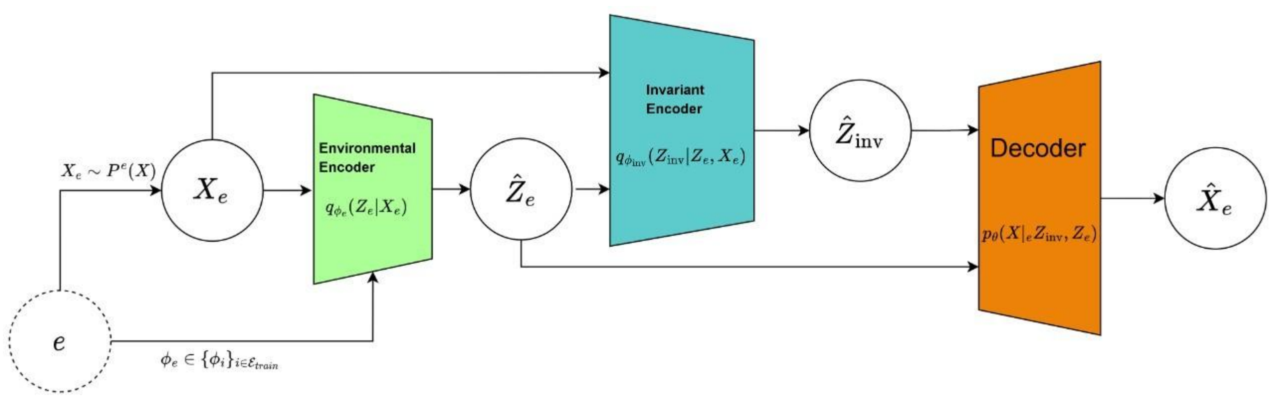
$U_r = \emptyset$
find: $U = \{u \in \mathbb{R}^n | (\Sigma_x^1 - \Sigma_x^2)u = 0\}$

for i = 1 to d_r **do**
  $u^* = \max_{u \in U} \; u^\top (\Sigma_x^1 + \Sigma_x^2) u$
  $U_r = U_r \cup \{u^*\}$
  $U = U \setminus \{u^*\}$
end for
End Procedure

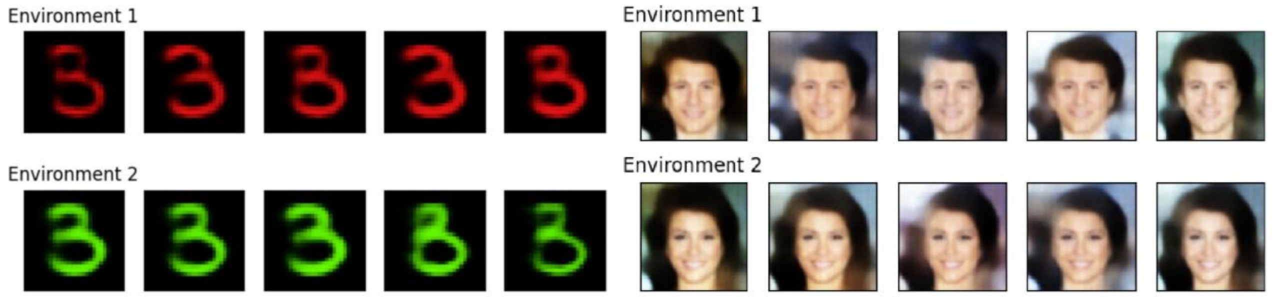


VIAE Algorithm

- "Variational Invariant Auto Encoder", proposed algorithm to solve the UIRM problem.
- Based on the VAE architecture
- Specially designed to recover the assumed generative model



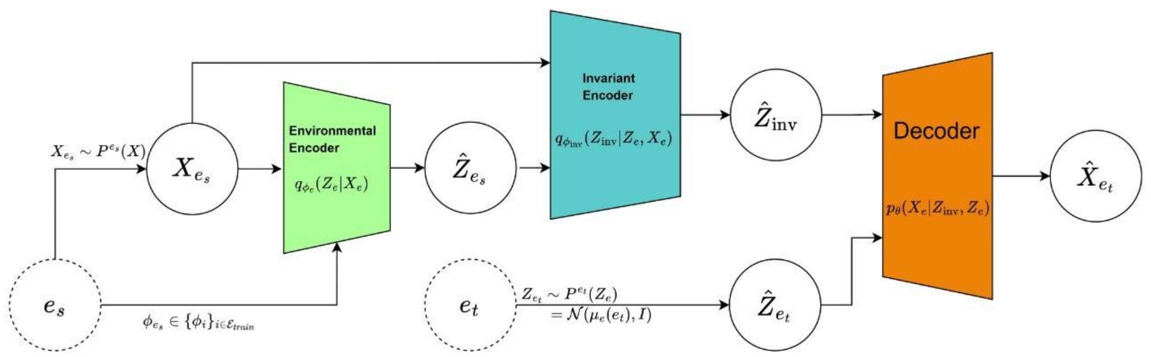
- Sampling Z_{inv} and Z_e independently
- Generates the same sample in different environments



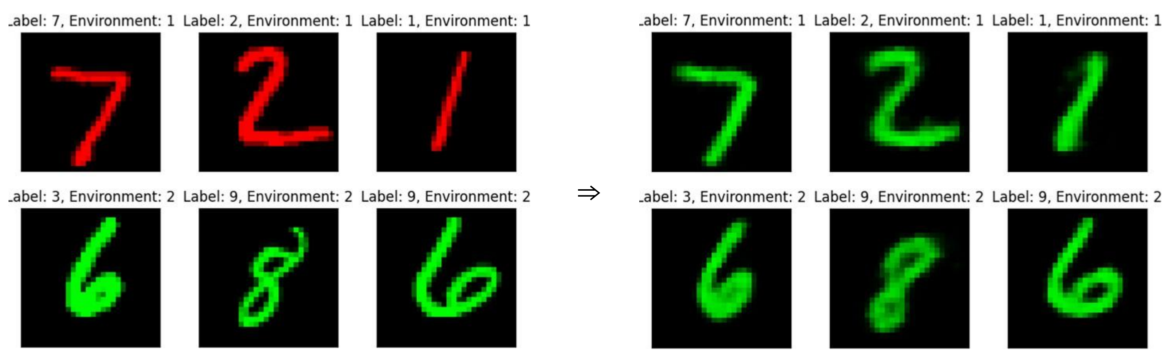
SCMNIST   CelebA

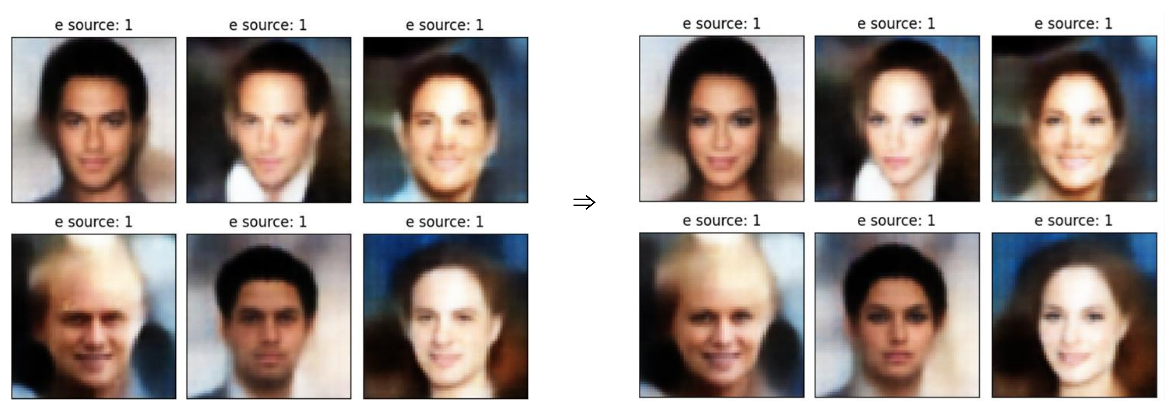
Environment Transfer

- Can we generate an invariant sample?
- If there was only one environment, could we consider the UIRM problem as solved?
- VAIE enables us to "transfer" a sample from one environment to the other



- Use the Encoders to recover the Invariant features, then combine them with generated environmental features
- The outcome from the decoder is a "transferred" sample





Conclusions & Future Work

- We offer an unsupervised approach for the IRM OOD problem.
- VIAE shows promising results in tackling this problem.
- Future work: Can environment transfer work for unknown environments in the zero-shot case?