

UNIVERSITÄT DES SAARLANDES

Integrating Over the Unobserved

Bayesian Reasoning with Structured Latent Variables

Isabel Valera

Saarland University

Bayes AI workshop, Edinburgh, Scotland UK

27.10.2026

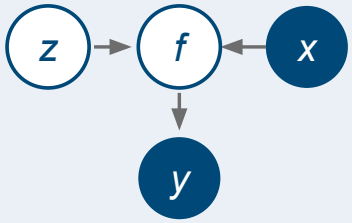
erc

ellis

Much of What We Want to Reason About Is Unobserved

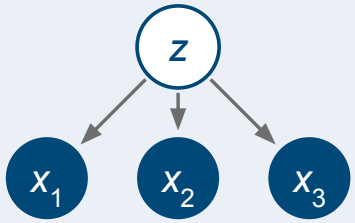
UNIVERSITÄT DES SAARLANDES

Each is a latent variable we never observe directly, so we must infer and integrate out.



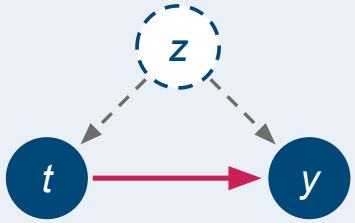
Functions

What underlying function generated sparse observations?



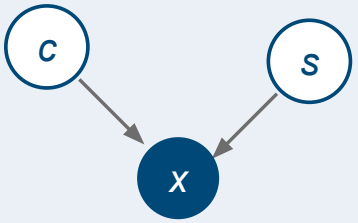
Shared factors

What information is common across modalities?



Hidden causes

How can we generate worlds we have never observed?



Concepts

How can we steer what a deep generative model generates?

TODAY's talk: Bayesian Inference for Deep Generative Models

UNIVERSITÄT DES SAARLANDES

Bayesian inference

Priors $p(z)$

Structure latent variables by encoding assumptions on the latent structure

Likelihoods $p(x \mid z)$

Modelling assumptions on data generation

Posteriors $p(z \mid x)$

Update beliefs and quantify uncertainty over latents

+

Deep generative models

Expressiveness

Model complex, high-dimensional data distributions

Flexibility

Learn rich latent representations of real-world data

Scalability

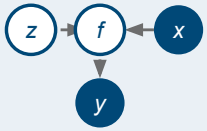
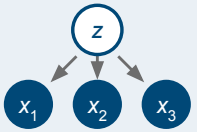
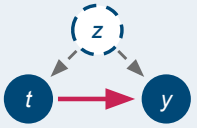
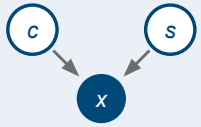
Efficient training and amortized inference

*How can we leverage structured Bayesian reasoning to **learn, reason about, and control latent structure in deep generative models**?*

Beyond Density Estimation: From Representation to Control

UNIVERSITÄT DES SAARLANDES

What can we achieve by learning and reasoning over structured latent variables?

1	**Represent** VAMoH · LDMI · TV-INRs	Infer distributions over continuous functions Latents encode **function representations**	
2	**Integrate** HELVAE · Hölder++	Pool beliefs from multiple modalities Latents encode **shared information**	
3	**Reason** Causal NF · DeCaFlow	Marginalize hidden confounders for causal queries Latents encode **hidden causes**	
4	**Control** VHCB	Steer generation by manipulating latent concepts Latents encode **human-interpretable concepts**	

UNIVERSITÄT DES SAARLANDES

Can we learn distributions over complex neural functions by performing Bayesian inference in a low-dimensional latent space?

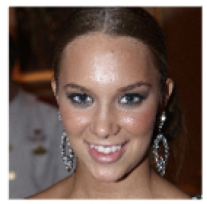
1 **Represent** — 2 Integrate — 3 Reason — 4 Control

GOAL: From Observations to Distributions over Functions

UNIVERSITÄT DES SAARLANDES

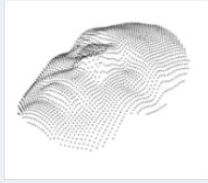
How can we learn expressive distributions of continuous functions underlying sparse observations?

SPATIAL



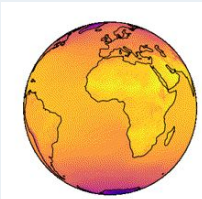
2D images

$f : \mathbb{R}^2 \to \mathbb{R}^3$



3D shapes

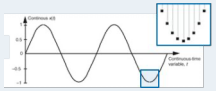
$f : \mathbb{R}^3 \to \{0, 1\}$



Climate

$f : \mathbb{S}^2 \to \mathbb{R}$

TEMPORAL



Time series

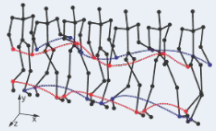
$f : \mathbb{R} \to \mathbb{R}^d$



Audio

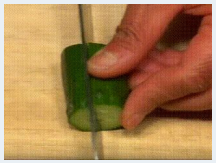
$f : \mathbb{R} \to \mathbb{R}$

SPATIO-TEMPORAL



Motion

$f : \mathbb{R} \to \mathbb{R}^{3J}$



Video

$f : \mathbb{R}^3 \to \mathbb{R}^3$

Usually these signals are continuous functions but observed via discretized data such as pixels, voxels or fixed sampling rates. Instead, we model the continuous function itself and query it at any coordinate.

HOW: From Inference over Latents to Distributions over Functions

UNIVERSITÄT DES SAARLANDES

LATENT

$z \sim p(z)$

low-dimensional

HYPERNETWORK

$\theta = g_\phi(z)$

outputs INR weights

INR

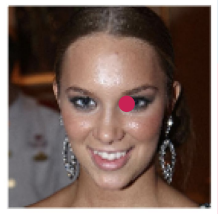
$y = f_\theta(x)$

continuous function at any coordinate x

DATA

coordinate $x = (x_1, x_2)$

value $y = (r, g, b)$



Parameterizing continuous functions with NNs

The NN parametrizes a continuous function over the inputs, e.g., pixel coordinates

INR = Implicit Neural Representation

Hypernetwork = NN that parametrizes another NN, usually with significantly lower no. of parameters

HOW: From Inference over Latents to Distributions over Functions

UNIVERSITÄT DES SAARLANDES

LATENT

$z \sim p(z)$

low-dimensional

HYPERNETWORK

$\theta = g_\phi(z)$

outputs INR weights

INR

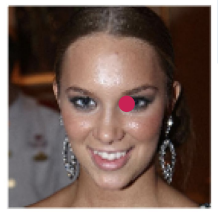
$y = f_\theta(x)$

continuous function at any coordinate x

DATA

coordinate $x = (x_1, x_2)$

value $y = (r, g, b)$



Parameterizing distribution over functions

Reason in a low-dimensional latent space, not over millions of network weights.

distribution over z → distribution on INR weights → distribution over continuous functions

HOW: From Inference over Latents to Distributions over Functions

UNIVERSITÄT DES SAARLANDES

LATENT

$$z \sim p(z)$$

low-dimensional

→

HYPERNETWORK

$$\theta = g_\phi(z)$$

outputs INR weights

→

INR

$$y = f_\theta(x)$$

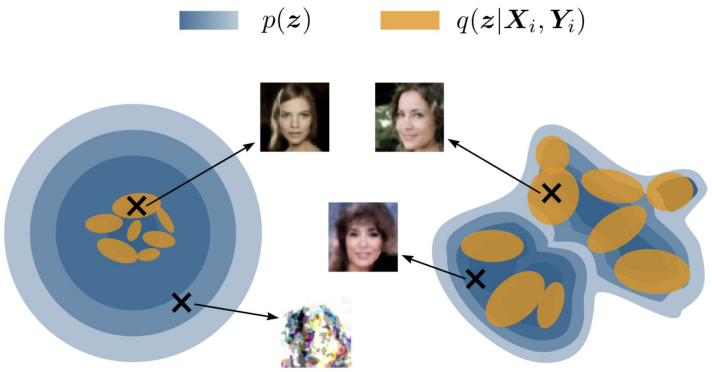
continuous function at any coordinate x

→

DATA

coordinate $x = (x_1, x_2)$

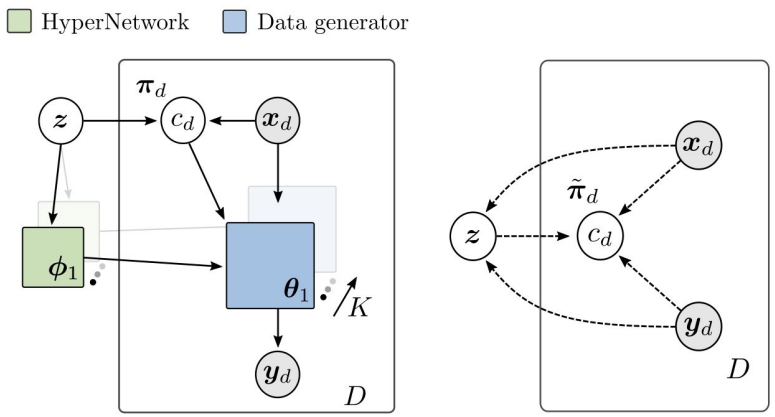
value $y = (r, g, b)$



Posterior over completions

Given a few observed points, e.g., pixels, the posterior q(z | points) induces a posterior distribution over the whole function.

VAMoH: A Variational Mixture of HyperGenerators



(a) Generative model (b) Inference model

PRIOR

$p_\psi(z)$

Planar normalizing flow: a flexible prior that avoids the prior hole

LIKELIHOOD

$\theta_k = g_{\phi_k}(z)$

Mixture of K hypergenerators (MLPs); per coordinates $c_d \sim \text{Cat}(\pi_d(x_d, z))$

POSTERIOR

$q_\gamma(z \mid X, Y)$

PointConv encoder on any set of observed points, no pre-imputation

$$p(Y, C, z \mid X) = p_\psi(z) \prod_{d=1}^{D} \prod_{k=1}^{K} \left[\pi_{dk}\, p_{\theta_k}(y_d \mid x_d)\right]^{[c_d=k]}$$

RESULTS: Conditional Generation in a Single Forward Pass

UNIVERSITÄT DES SAARLANDES

Image completion on unseen data



(a) Missing a patch

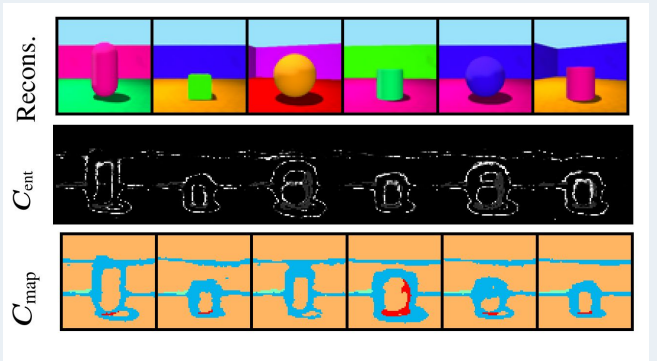


(b) Missing half of the image

2.3–3.6×
faster than Functa (competing metalearning approach) with 3 gradient steps

7–11×
faster than Functa with 10 gradient steps

Interpretable mixture



Rows: reconstruction, posterior entropy of c_d, MAP assignment. One hypergenerator fills colours; the others handle borders.

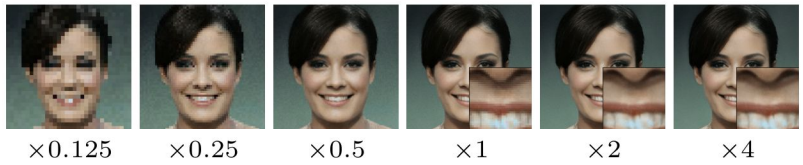
Beyond VAMoH: Scaling and Conditioning Function Distributions

UNIVERSITÄT DES SAARLANDES

SCALING EXPRESSIVENESS

Hyper-Transforming LDMs

LDMI · ICML 2025



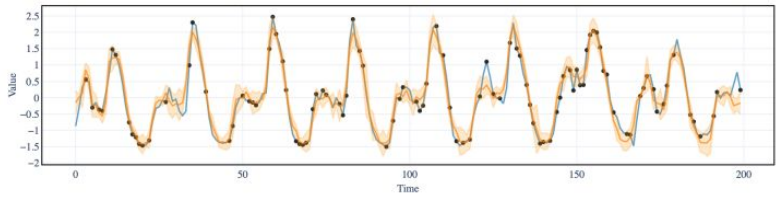
Transformer hypernetworks generate INR weights from latent variables; latent diffusion provides a flexible generative prior.

→ **Scalable function generation and reuse of pretrained latent spaces.**

INDIVIDUALIZING INFERENCE

Temporal Variational INRs

TV-INRs · TMLR 2026



(b) Imputation task for Electricity dataset $L = 200$, $\tau = 0.5$.

Covariate-conditioned latent-variable models represent distributions over continuous-time functions.

→ **Efficient individualized imputation and forecasting from irregular observations.**

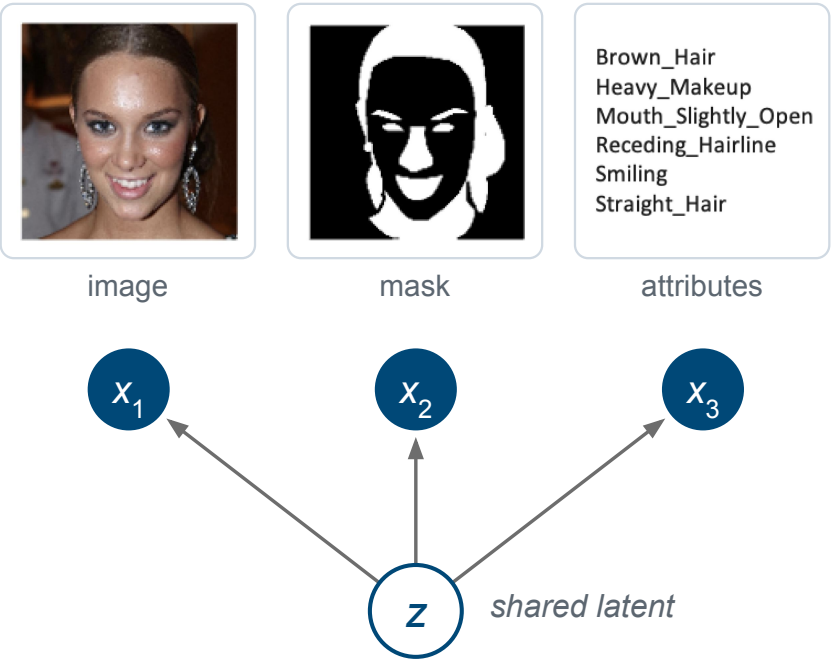
INTEGRATE: Pooling Information Across Modalities

UNIVERSITÄT DES SAARLANDES

How should we combine different posterior beliefs about the same latent variables?

1 Represent — 2 **Integrate** — 3 Reason — 4 Control

GOAL: From Single to Multimodal Learning



1 **Shared representations**

z captures what the modalities have in common

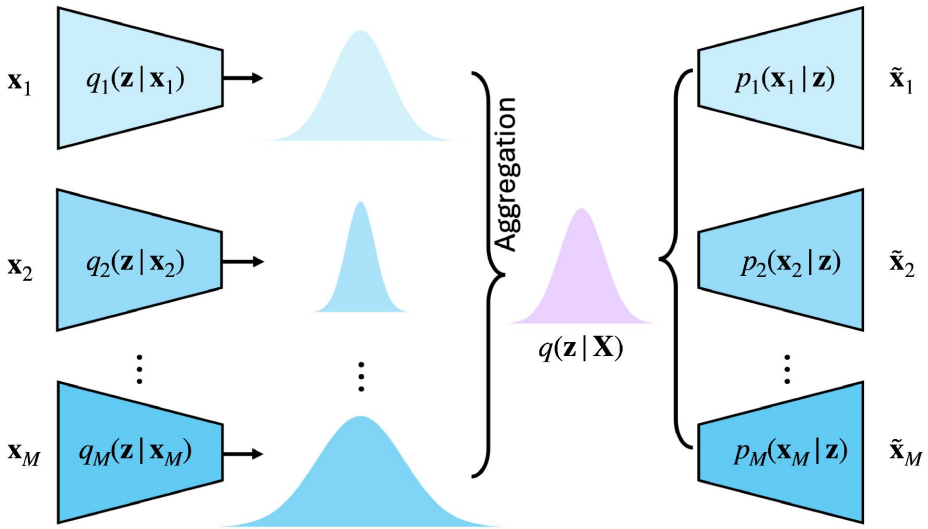
2 **Cross-modal generation**

Generate a coherent mask from the image alone

3 **Missing modalities**

Infer z and generate whatever modality is absent

Multimodal VAEs: One Shared Latent, Many Views



Unimodal experts $q_j(z \mid x_j)$ are aggregated into $q(z \mid X)$

GENERATIVE MODEL

$$p_\theta(X, z) = p(z) \prod_{j=1}^{M} p_{\theta_j}(x_j \mid z)$$

INFERENCE MODEL

$$q_\phi(z \mid X) = \mathrm{Agg}\big(\{q_{\phi_j}(z \mid x_j)\}_{j=1}^{M}\big)$$

Each encoder is an expert with its own belief about z. How do we aggregate them?

Two competing criteria:

generative quality (realism, diversity) vs coherence (semantic consistency across modalities)

HOW: Aggregation as Probabilistic Opinion Pooling

Aggregate experts to minimize an α-divergence between the individual and aggregated distributions:

$$q(z) = \arg\min_{\varphi \in \mathcal{P}} \sum_{j=1}^{M} \lambda_j \, D_\alpha(q_j \,\|\, \varphi) \qquad \Longrightarrow \qquad q(z) \propto \Big(\sum_{j=1}^{M} \lambda_j \, q_j(z)^\alpha \Big)^{1/\alpha}$$

$\alpha \to 0$

Reverse KL

Asymmetric, unbounded

Product of experts

$$q(z \mid X) \propto \prod_{j=1}^{M} q_j(z \mid x_j)$$

✓ high quality × low coherence

$\alpha = 1$

Forward KL

Asymmetric, unbounded

Mixture of experts

$$q(z \mid X) = \frac{1}{M} \sum_{j=1}^{M} q_j(z \mid x_j)$$

× low quality ✓ high coherence

HOW: Aggregation as Probabilistic Opinion Pooling

Aggregate experts to minimize an α-divergence between the individual and aggregated distributions:

$$q(z) = \arg\min_{\varphi \in \mathcal{P}} \sum_{j=1}^{M} \lambda_j \, D_\alpha(q_j \,\|\, \varphi) \qquad \Longrightarrow \qquad q(z) \propto \Big(\sum_{j=1}^{M} \lambda_j \, q_j(z)^\alpha \Big)^{1/\alpha}$$

$\alpha \to 0$

Reverse KL

Asymmetric, unbounded

Product of experts

$$q(z \mid X) \propto \prod_{j=1}^{M} q_j(z \mid x_j)$$

✓ high quality × low coherence

$\alpha = 0.5$

Symmetric α-divergence

Symmetric, bounded

Hölder pooling

✓ high quality

✓ high coherence

$\alpha = 1$

Forward KL

Asymmetric, unbounded

Mixture of experts

$$q(z \mid X) = \frac{1}{M} \sum_{j=1}^{M} q_j(z \mid x_j)$$

× low quality ✓ high coherence

Hellinger vs. Hölder++ Aggregation

UNIVERSITÄT DES SAARLANDES

EXACT SYMMETRIC HÖLDER POOLING: $$q(z \mid X) = \sum_{j=1}^{M} \pi_j \, q_j(z \mid x_j) \; + \; \sum_{i<j} \pi_{ij} \, q_{ij}^{(1/2)}(z \mid x_i, x_j), \qquad \pi_{ij} \propto S_{ij}$$

1 **HELVAE**

AISTATS 2026

Moment-matching Gaussian approximation of Hölder pooling with α = 0.5. No sub-sampling during training.

2 **Hölder**

exact pooling

The pooled posterior is a **mixture of unimodal and pairwise components**, weighted by the overlap S_{ij} of experts i and j.

Hellinger vs. Hölder++ Aggregation

UNIVERSITÄT DES SAARLANDES

EXACT SYMMETRIC HÖLDER POOLING: $$q(z \mid X) = \sum_{j=1}^{M} \pi_j \, q_j(z \mid x_j) \; + \; \sum_{i<j} \pi_{ij} \, q_{ij}^{(1/2)}(z \mid x_i, x_j), \qquad \pi_{ij} \propto S_{ij}$$

1 HELVAE

AISTATS 2026

Moment-matching Gaussian approximation of Hölder pooling with α = 0.5. No sub-sampling during training.

2 Hölder

exact pooling

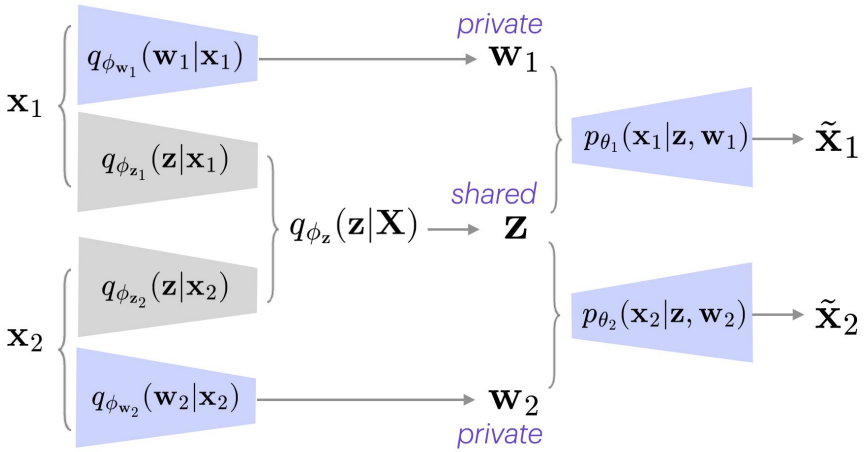
The pooled posterior is a mixture of unimodal and pairwise components, weighted by the overlap S_{ij} of experts i and j.

3 Hölder+ / ++

ICML 2026

Shared z **plus private** w_j per modality (+), and **hierarchical inference** of w_j given z (++), to keep z informative.

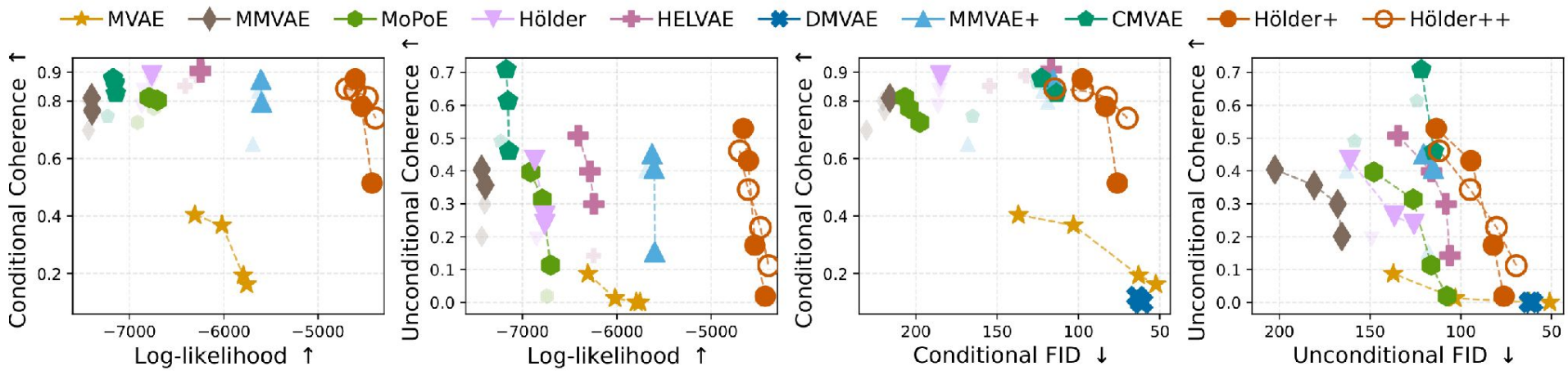
Private + shared latents



HIERARCHICAL INFERENCE (HÖLDER++)

$$q(z, W \mid X) = q(z \mid X) \prod_{j=1}^{M} q(w_j \mid x_j, z)$$

RESULTS: The Best Quality–Coherence Trade-off



Optimal region

Upper right in every panel: high coherence with high quality (likelihood or low FID)

Main result

HELVAE improves trade-off over single shared representation models.

Hölder+ and Hölder++ reach the best trade-off using private-shared latents

UNIVERSITÄT DES SAARLANDES

How can we learn a generative model from observational data that allows us to simulate interventions, and thus unseen worlds?

1 Represent — 2 Integrate — 3 **Reason** — 4 Control

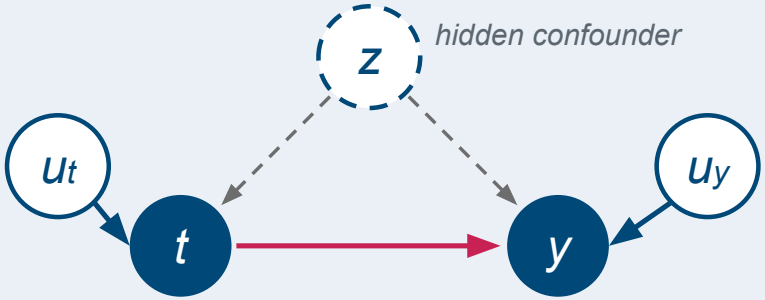
Causal Generative Models in One Slide

UNIVERSITÄT DES SAARLANDES

Causal generative model as a substitute of the true SCM generating the data
for efficient and accurate causal inference

SCM = STRUCTURAL CAUSAL MODEL

$$x_i := f_i\big(\mathrm{pa}(i),\, u_i,\, z\big)$$



$$p(y \mid t) \;\neq\; p(y \mid \mathrm{do}(t))$$

Seeing is not doing when z is unobserved

Pearl's ladder of causation

3 **Counterfactual**
What would have happened?
$p\big(y^{\mathrm{cf}} \mid \mathrm{do}(t^{\mathrm{cf}}),\, x^{\mathrm{f}}\big)$

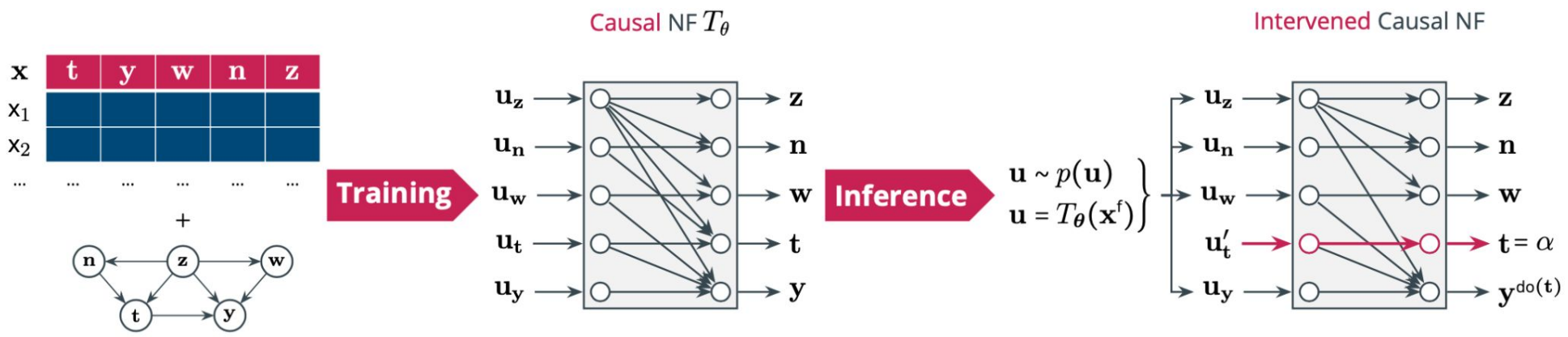
2 **Interventional**
What happens if we fix t?
$p\big(y \mid \mathrm{do}(t = \alpha)\big)$

1 **Observational**
What do we see?
$p(y \mid t)$

A causal generative model samples from all three rungs.

GOAL: From Observations to Interventions

How can we learn a generative model from observational data that allows us to simulate causal interventions?

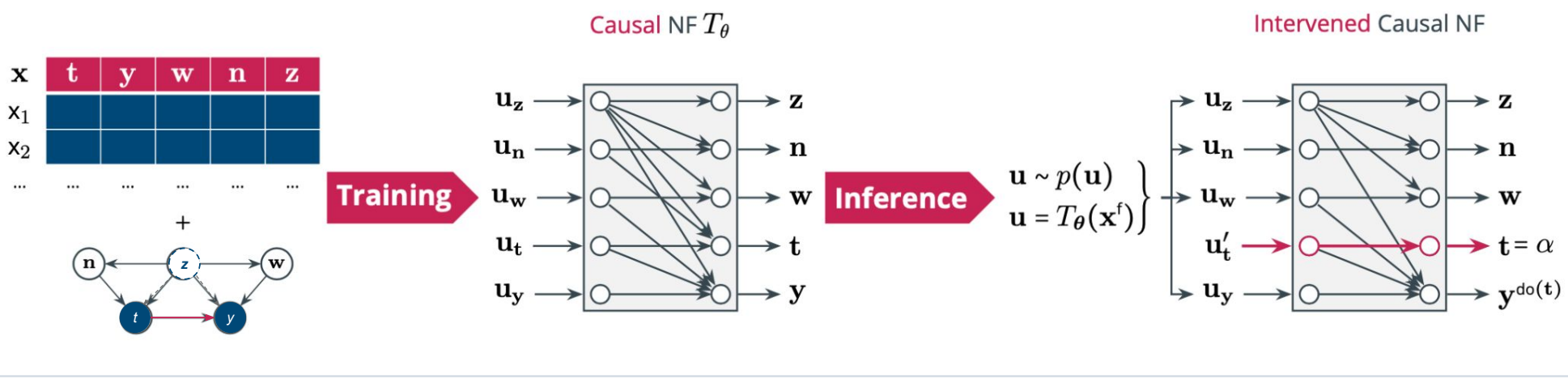


$$\mathcal{G} = \mathcal{G}_\theta \;\; \& \;\; p_{\mathcal{M}}(x) = p_\theta(x) \quad \Longrightarrow \quad p_{\mathcal{M}}\big(y \mid \mathrm{do}(t = \alpha)\big) = p_\theta\big(y \mid \mathrm{do}(t = \alpha)\big)$$

Matching the data under the right graph is enough to answer interventional queries, provided there are no hidden confounders.

GOAL: From Observations to Interventions

How can we learn a generative model from observational data that allows us to simulate causal interventions?

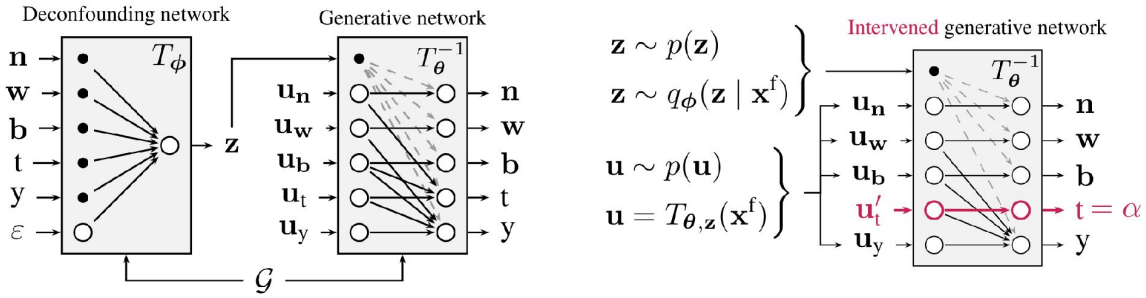


$$\mathcal{G} = \mathcal{G}_\theta \;\&\; p_{\mathcal{M}}(x) = p_\theta(x) \quad \not\Rightarrow \quad p_{\mathcal{M}}\big(y \mid \mathrm{do}(t=\alpha)\big) = p_\theta\big(y \mid \mathrm{do}(t=\alpha)\big)$$

Matching the data under the right graph is **NOT enough** to answer interventional queries, **when hidden confounding.**

HOW: DeCaFlow, a Deconfounding Causal Generative Model

What we need infer the hidden confounders and integrate over them, while keeping the causal structure



(a) DeCaFlow training computations.

(b) Computation of causal queries.

TRAINING OBJECTIVE

$$\mathcal{L}(\theta, \phi) = \mathbb{E}_{q_\phi}\left[\log p_\theta(x \mid z)\right] - \mathrm{KL}\left[q_\phi(z \mid x) \,\|\, p(z)\right]$$

Trained once per dataset, from observational data and the causal graph.

PRIOR

$p(z) = \mathcal{N}(0, I)$

Over the hidden confounders

LIKELIHOOD

$p_\theta(x \mid z) = \prod_i p_\theta\left(x_i \mid \mathrm{pa}(i), z\right)$

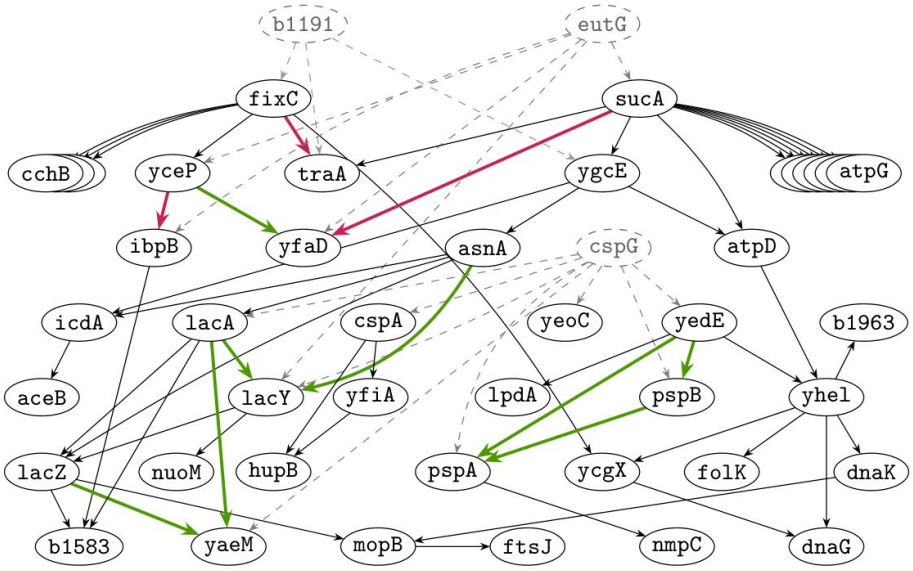
Conditional causal NF that respects the graph

POSTERIOR

$q_\phi\left(z \mid \mathrm{ch}(z) \cup \mathrm{pa}(\mathrm{ch}(z))\right)$

Deconfounding network: a conditional NF encoding causal structure

From One Generative Model to Many Causal Queries



Ecoli70: 43 observed genes, 3 hidden confounders (dashed). Green: hidden-confounded effects DeCaFlow can identify.

1 **Fit observational data**

with a DeCaFlow that encodes the **causal graph**

2 **Check proxy-identifiability**

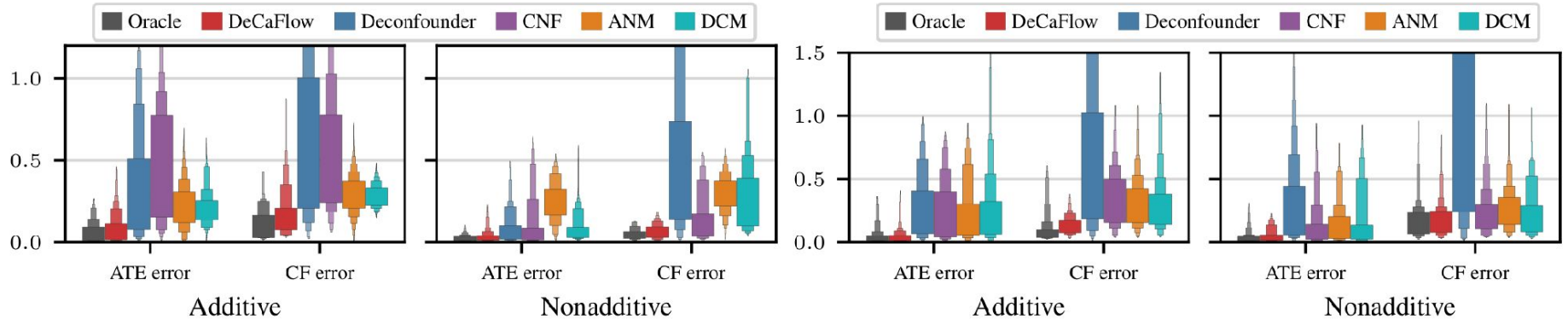
with the provided the **causal graph**

3 **Estimate any identifiable query**

$p(y \mid do(t))$ or $p(y \mid do(t), x^f)$

Train once per dataset: estimate query for any identifiable treatment–outcome pair.

RESULTS: Close to an Oracle That Observes the Confounders



(a) Sachs' dataset.

(b) Ecoli70 dataset.

43 · 3

observed variables and hidden confounders in Ecoli70, one training run

≈ oracle

ATE and counterfactual error on Sachs and Ecoli70; better than CNF, ANM and DCM

0.16 → 0.005

MMD between sensitive groups on law-school data (RMSE 1.41 → 1.60)

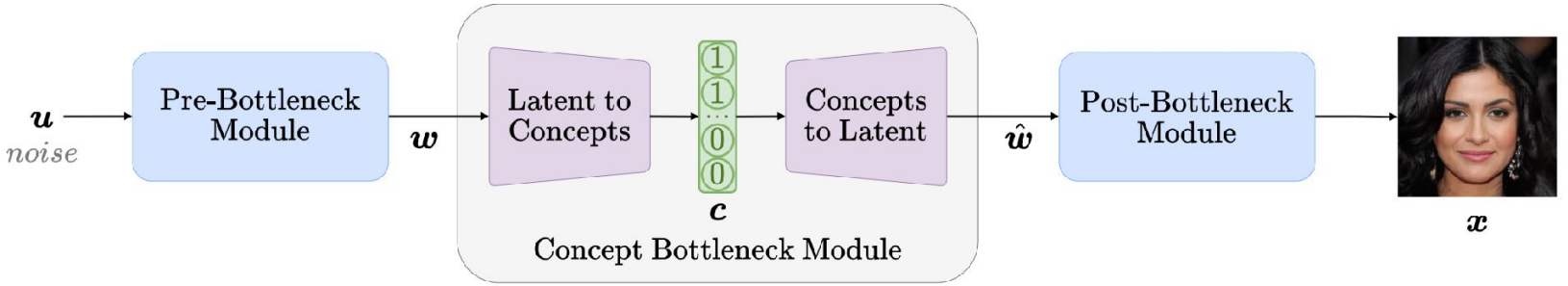
UNIVERSITÄT DES SAARLANDES

How can Bayesian inference help us learn human-interpretable latent variables that we can manipulate to control generation?

1 Represent — 2 Integrate — 3 Reason — 4 **Control**

GOAL: From Latent Representations to Controllable Concepts

How can we control what a generative model generates?



c = [**Black hair** = 1, **Smiling** = 1, **Mouth open = 0**, **Young** = 1]

c = [**Black hair** = 1, **Smiling** = 1, **Mouth open = 1**, **Young** = 1]

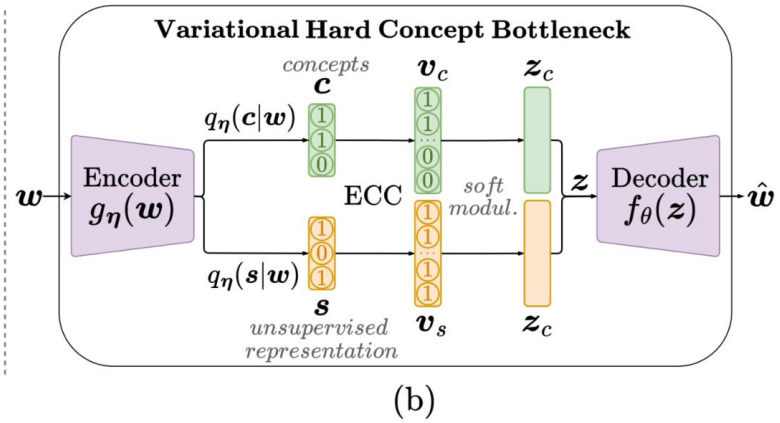
× **Soft, deterministic concepts**

Concept probabilities leak extra information, so interventions can fail to change the output

× **No generative process over concepts**

Steering is limited to editing existing samples; we cannot sample from a concept configuration

HOW: A Probabilistic Hard Concept Bottleneck



(b)

- **Hard concepts** mitigate leakage
- **Probabilistic** layer: generate directly from a concept configuration

PRIOR

$c \in \{0,1\}^K,\ s \in \{0,1\}^L$

Binary concepts and a binary side channel

LIKELIHOOD

$p_\theta(w \mid z),\ z = [z_c; z_s]$

Decoder on smoothed, error-correction-coded bits

POSTERIOR

$q_\eta(c \mid w),\ q_\eta(s \mid w)$

Robust concept inference via error-correcting codes

CONTROL

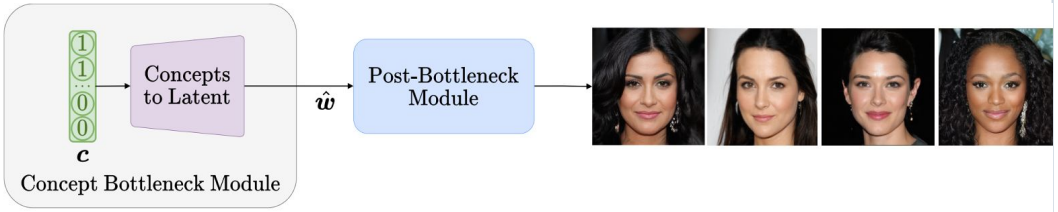
$c_k : 0 \to 1$

Set concept bits to generate or to intervene

RESULTS: Generating from a Concept Configuration

UNIVERSITÄT DES SAARLANDES

Steerable generation

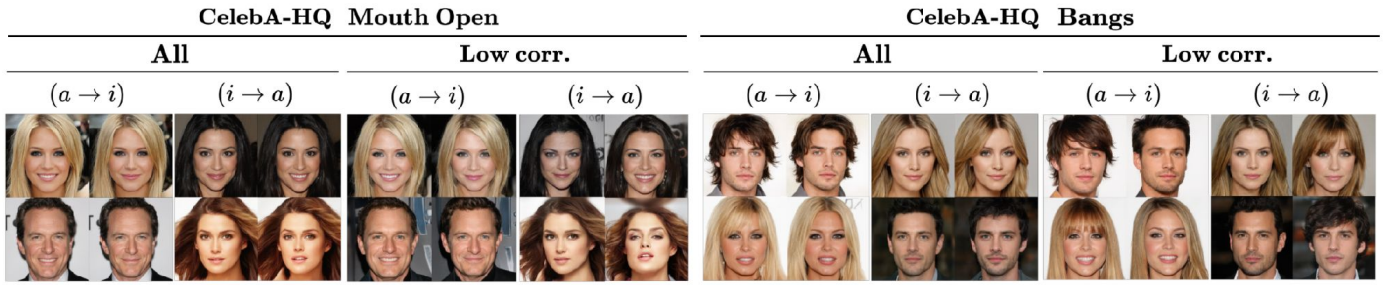


C = [‘Black Hair’ = 1, ‘Smiling’ = 1, ‘Mouth Open’ = 0, ‘Young’ = 1, ‘Female’ = 1]

0.87 → 0.55

target accuracy, from concept patterns seen in training to uniformly random concept sets

Out-of-distribution concept combinations and biases of the pretrained generator limit how far we can steer.



0.77 → 0.42

target accuracy from high- to low-correlation concept set

LIMITATIONS: Do Manipulations Behave Predictably?

UNIVERSITÄT DES SAARLANDES

How do we guarantee that concept manipulations have predictable effects, and that unrelated attributes remain unchanged?

Task activate the concept ‘mustache’



- Low-correlation concept set: the intervention mostly changes the target
- All 40 concepts: correlations and imbalance in the data leak into the intervention

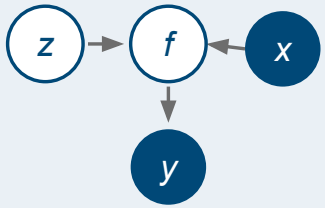
Results with denoising diffusion models and more in the paper

Beyond Density Estimation: What Does Bayesian Reasoning Enable?

UNIVERSITÄT DES SAARLANDES

Bayesian inference over structured latent variables enables
increasingly rich capabilities in deep generative models.

1 · REPRESENT

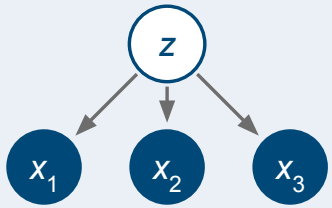


Functions

What underlying function generated sparse observations?

Latents encode **function variability**

2 · INTEGRATE

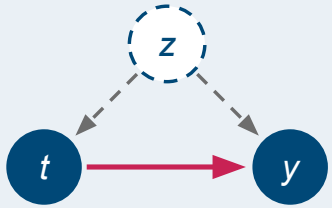


Shared factors

What information is common across modalities?

Latents encode **shared information**

3 · REASON

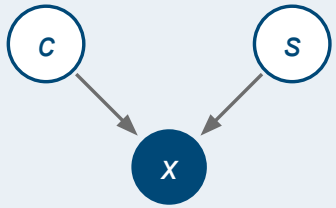


Hidden causes

How can we generate worlds we have never observed?

Latents encode **hidden causes**

4 · CONTROL



Concepts

How can we steer what a deep generative model generates?

Latents encode **interpretable attributes**

Beyond Inference: Identifiability and Reliable Control

UNIVERSITÄT DES SAARLANDES

When can we trust what inferred latent variables represent, and what happens when we intervene them?

z Latent-variable identifiability

Under what assumptions can we recover the underlying latent factors from observed data in a way that they are interpretable to humans?

Supervision | Structural assumptions

z' Reliable interventions

How can we manipulate these factors and guarantee predictable effects on the generated data, while keeping unrelated factors unchanged?

Causal semantics | Interventions

IDENTIFIABILITY GUARANTEES



UNIVERSITÄT DES SAARLANDES

Thank You!

Questions welcome

JOINT WORK WITH

1 · Represent

Batuhan Koyuncu, Pablo Sánchez-Martín, Ignacio Peis, Pablo M. Olmos, Jes Frellsen, Rachael DeVries, Ole Winther

2 · Integrate

Huyen Vo, María Martínez-García

3 · Reason

Alejandro Almodóvar, Adrián Javaloy, Juan Parras, Santiago Zazo, Pablo Sánchez-Martín

4 · Control

María Martínez-García, Ricardo Vazquez Alvarez, Alejandro Lancho, Pablo M. Olmos







BACKUP: References

UNIVERSITÄT DES SAARLANDES

REPRESENT

B. Koyuncu, P. Sánchez-Martín, I. Peis, P. M. Olmos, I. Valera. Variational Mixture of HyperGenerators for Learning Distributions over Functions. ICML 2023.

I. Peis, B. Koyuncu, I. Valera, J. Frellsen. Hyper-Transforming Latent Diffusion Models. ICML 2025.

B. Koyuncu, R. DeVries, O. Winther, I. Valera. Temporal Variational Implicit Neural Representations. TMLR 2026.

INTEGRATE

H. Vo, I. Valera. Hellinger Multimodal Variational Autoencoders. AISTATS 2026.

H. Vo, M. Martínez-García, I. Valera. Hölder++: Improving the Quality-Coherence Trade-off in Multimodal VAEs. ICML 2026.

REASON

A. Javaloy, P. Sánchez-Martín, I. Valera. Causal Normalizing Flows: from Theory to Practice. NeurIPS 2023.

A. Almodóvar, A. Javaloy, J. Parras, S. Zazo, I. Valera. DeCaFlow: A Deconfounding Causal Generative Model. NeurIPS 2025.

CONTROL

M. Martínez-García, R. Vazquez Alvarez, A. Lancho, P. M. Olmos, I. Valera. A Probabilistic Hard Concept Bottleneck for Steerable Generative Models. ICLR 2026.