Weekly Study Report

A Look at ICLR 2025 Submission List

Wang Ma

2024-10-18

Electrical, Computer, and Systems Engineering Department
Rensselaer Polytechnic Institute

Outline

1. Denoising Diffusion Causal Discovery

1.1.1 Structural Equations Models

For Linear SEM:

$$X = XW + E, \quad \min_{W} \frac{1}{2n} \left\| X - XW \right\|_F^2 + \lambda_1 \left\| W \right\|_1 + \lambda_2 \left\| W \right\|_2.$$

For nonlinearity:

$$X = f(X; W) + E = f_2(f_1(X)W) + E.$$

1.1.2 Continuous DAG Constraint

$$h(W) = \mathrm{tr}\left(e^{W \circ W}\right) - d$$

1.1.3 Denoising Diffusion Probabilistic Models

Updating Rule:

$$\begin{aligned} x_t &= \sqrt{1 - \beta_t} x_{t-1} + \sqrt{\beta_t} z_{t-1} \\ &= \sqrt{\bar{\alpha_t}} x_0 + \sqrt{1 - \bar{a_t}} z \end{aligned}$$

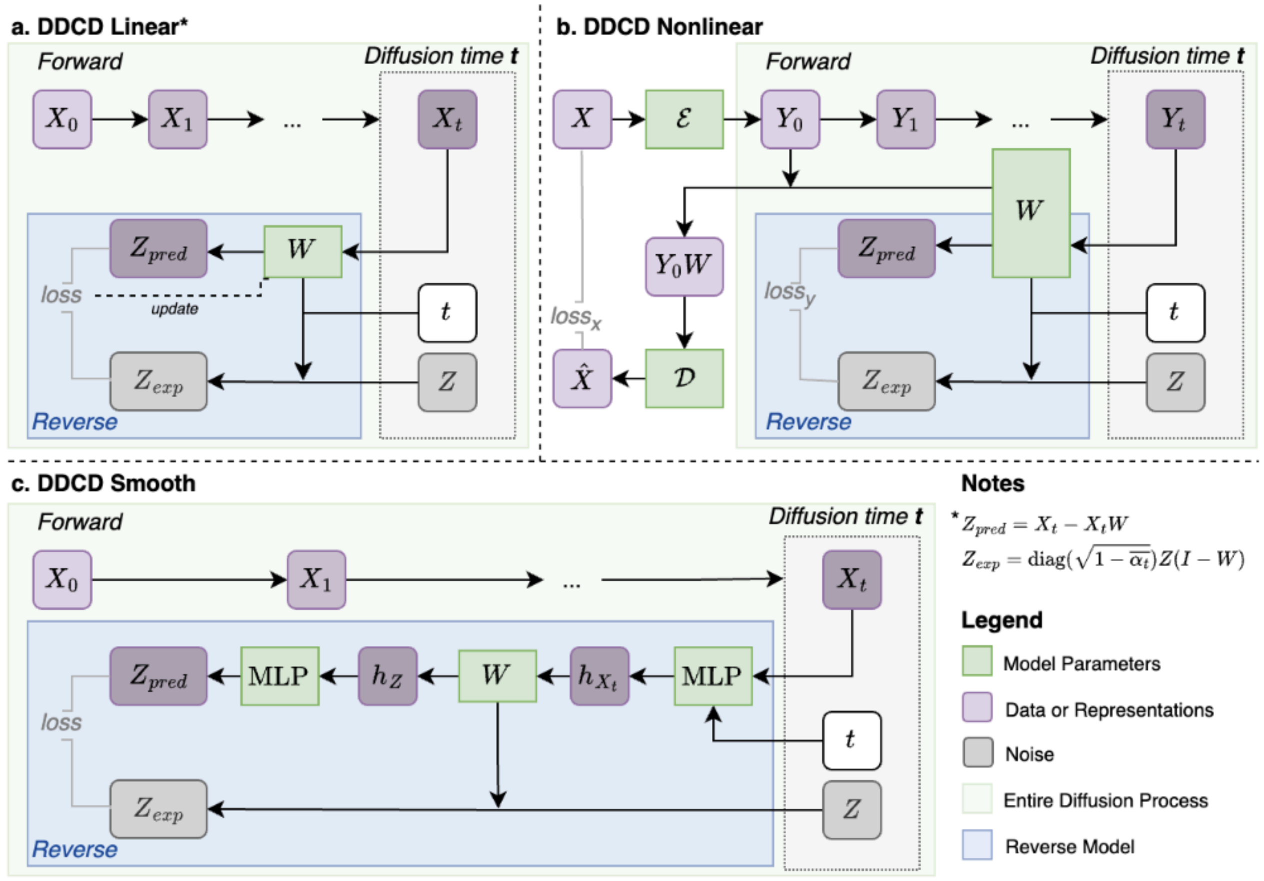


Figure 1: Model architectures of proposed models in this paper.

1.2.1 Denoising Diffusion Models for Linear SEMS

$$\min_W \frac{1}{2n} \|\underbrace{(X_t - X_t W)}_{\text{predicted noise}} - \underbrace{diag\left(\sqrt{1-\bar{\alpha_t}}\right) Z(I - W)}_{\text{expected noise}}\|_F^2 + \lambda_1 \|W\|_1 + \lambda_2 \|W\|_2$$

Theorem 1. *For linear SEMs, the objective functions in Equation 8 and Equation 2 are equivalent.*

Proof. Consider the case when each sample in $\boldsymbol{X}$ is perturbed using the forward diffusion process in 7. The perturbed observational data could be written as Equation 9. Here we use $\boldsymbol{t}$ to denote the different diffusion time steps for all samples in $\boldsymbol{X}$ and the diag() operator scales each row of $\boldsymbol{X}_0$ and $\boldsymbol{Z}$ accordingly based on the diffusion schedule.

$$\boldsymbol{X_t} = \text{diag}(\sqrt{\overline{\alpha_t}})\boldsymbol{X_0} + \text{diag}(\sqrt{1-\overline{\alpha_t}})\boldsymbol{Z}, \tag{9}$$

With Equations 9 and 1, we can easily develop Equations 10-12

$$\boldsymbol{X_t W} = \text{diag}(\sqrt{\overline{\alpha_t}})\boldsymbol{X_0 W} + \text{diag}(\sqrt{1-\overline{\alpha_t}})\boldsymbol{ZW}, \tag{10}$$

$$\boldsymbol{X_t} - \boldsymbol{X_t W} = \text{diag}(\sqrt{\overline{\alpha_t}})(\boldsymbol{X_0} - \boldsymbol{X_0 W}) + \text{diag}(\sqrt{1-\overline{\alpha_t}})(\boldsymbol{Z} - \boldsymbol{ZW}) \tag{11}$$

$$\text{diag}(\sqrt{\overline{\alpha_t}})(\boldsymbol{X_0} - \boldsymbol{X_0 W}) = (\boldsymbol{X_t} - \boldsymbol{X_t W}) - \text{diag}(\sqrt{1-\overline{\alpha_t}})\boldsymbol{Z}(\boldsymbol{I} - \boldsymbol{W}) \tag{12}$$

1.2.2 Denoising Diffusion Models for Nonlinear SEMs

$$Y = f_1(X) + E_1$$

$$X = f_2(YW) + E_2$$

$$Y = YW + E_3$$

$$\min_W \frac{1}{2n} \left\| \underbrace{(X - f_2(f_1(X)W))}_{\text{reconstruction error}} + \underbrace{(Y_t - Y_t W)}_{\text{predicted noise}} - \underbrace{diag\left(\sqrt{1-\bar{\alpha_t}}\right) Z(I - W)}_{\text{expected noise}} \right\|_F^2$$

$$+\lambda_1 \|W\|_1 + \lambda_2 \|W\|_2$$

2. Gradient Based Causal Discovery with Diffusion Model

2.1 Representing SCM by Diffusion Process

A linear SCM:

$$X = AX + Z = (I - A)^{-1} Z.$$

Nonlinear extension:

$$X = g\big((I - A)^{-1} f(Z)\big)$$

$$g^{-1}(X) = Ag^{-1}(X) + f(Z)$$

Forward Process:

$$Z = f^{-1}\big((I - A) g^{-1}(X)\big)$$

$$X \to \text{diffusion} \to (I - A) \to \text{diffusion} \to Z$$

Reverse Process:

$$Z \to \underbrace{\text{reverse}}_{f} \to (I - A)^{-1} \to \underbrace{\text{reverse}}_{g} \to \bar{X}$$

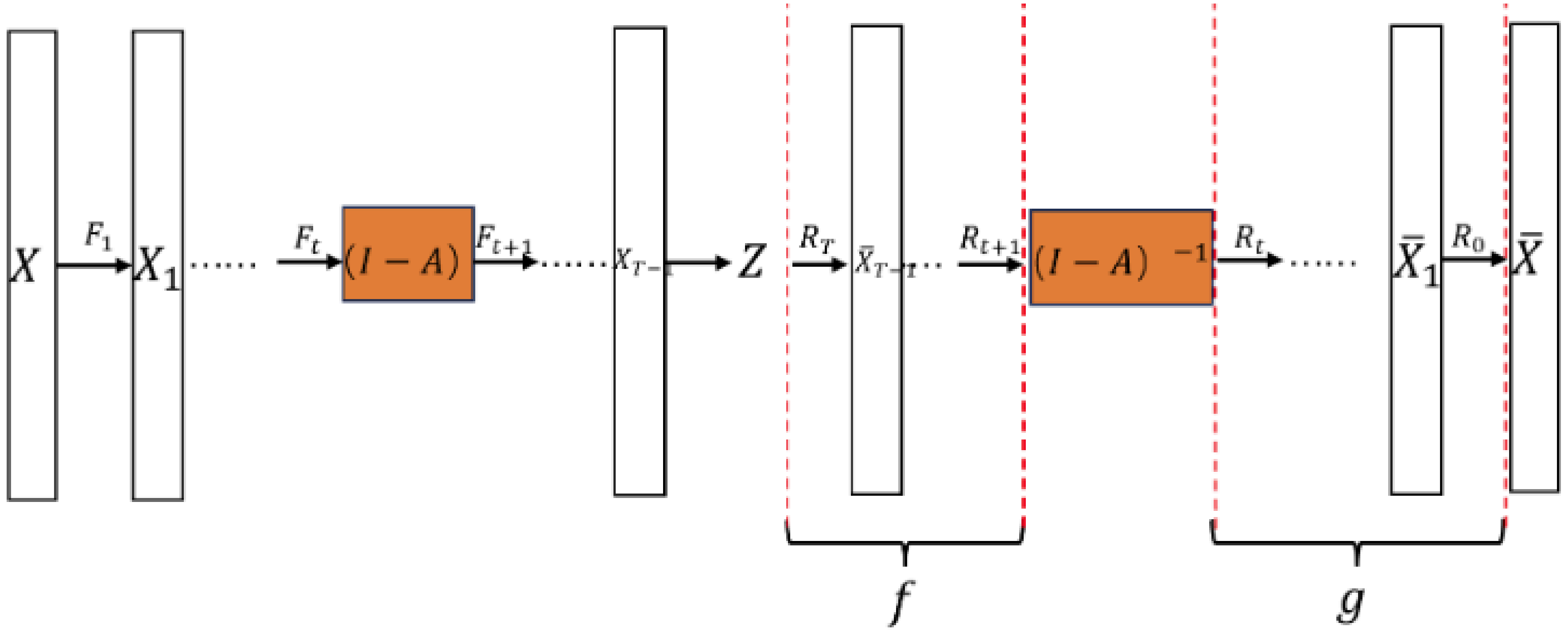


Figure 1: Illustration of the diffusion based causal modeling.

2.1 Representing SCM by Diffusion Process

Objective:

$$\min_{A,\Theta} L(A,\Theta) + \lambda h(A),$$

where $L(A,\Theta)$ is the diffusion ELBO loss and $h(A)$ is the DAG constraint.

3. Previous work about Diffusion X Causal Discovery

ICLR 2023: Diffusion Models for Causal Discovery via Topological Ordering

1. **Lemma:** Leaf node j satisfies $\mathrm{Var}_X\left[H_{j,j}(\log p(x))\right] = 0$.
2. In Diffusion models, we leverage the fact the the trained model ε_θ approximates the score $\nabla_{x_j} \log p(x)$ of the data.
3. Thus we will approximate the score's Jacobian via diffusion training:

$$H_{i,j} \log p(x) \approx \nabla_{i,j} \varepsilon_\theta(x, t).$$

4. Uncertainty - related work

Title: *Uncertainty modeling for fine-tuned implicit functions*, for applications in implicit neural representation like NeRF.

Algorithm 1 Dropsembles

▷ *Task A*

Require: : $\mathcal{D}_A$

1: $\hat{p}(\theta|\mathcal{D}_A) \leftarrow$ Train f_θ on $\mathcal{D}_A$ with dropout

▷ *Task B*

Require: : $\mathcal{D}_B, \hat{p}(\theta|\mathcal{D}_A)$

2: **for** $m = 1$ to M **do**

3: $\quad \theta^m_{init} \leftarrow$ Sample a thinned network initialized from $\hat{p}(\theta|\mathcal{D}_A)$

4: $\quad \hat{\theta}^m \leftarrow \arg\min_\theta \hat{R}_{\mathcal{D}_B}(\theta)$ ▷ Train thinned network on $\mathcal{D}_B$

5: **end for**

6: Obtain predictions and uncertainty estimates $\leftarrow$ Ensemble $\{\hat{\theta}^m\}_{m\in[M]}$

4.2 Detecting Discrepancies between AI-Generated and Natural Images using Uncertainty

Detecting Discrepancies Between AI-Generated and Natural Images Using Uncertainty

Anonymous authors
Paper under double-blind review

Abstract

In this work, we propose a novel approach for detecting AI-generated images by leveraging predictive uncertainty to mitigate misuse and associated risks. The motivation arises from the fundamental assumption regarding the distributional discrepancy between natural and AI-generated images. **The feasibility of distinguishing natural images from AI-generated ones is grounded in the distribution discrepancy between them**. Predictive uncertainty offers an effective approach for capturing distribution shifts, thereby providing insights into detecting AI-generated images. Namely, as the distribution shift between training and testing data increases, model performance typically degrades, often accompanied by increased predictive uncertainty. Therefore, we propose to employ predictive uncertainty to reflect the discrepancies between AI-generated and natural images. In this context, the challenge lies in ensuring that the model has been trained over sufficient natural images to avoid the risk of determining the distribution of natural images as that of generated images. We propose to leverage large-scale pre-trained models to calculate the uncertainty as the score for detecting AI-generated images. This leads to a simple yet effective method for detecting AI-generated images using large-scale vision models: images that induce high uncertainty are identified as AI-generated. Comprehensive experiments across multiple benchmarks demonstrate the effectiveness of our method.

Simply a weight-space perturbation method.

5. Passing by the Diffusion Models

Representation Alignment for Generation: Training Diffusion Transformers Is Easier Than You Think

Sihyun Yu[1] **Sangkyung Kwak**[1] **Huiwon Jang**[1] **Jongheon Jeong**[2]
Jonathan Huang[3] **Jinwoo Shin**[1*] **Saining Xie**[4*]
[1]KAIST [2]Korea University [3]Scaled Foundations [4]New York University

available at Arxiv on Oct 9

5. Passing by the Diffusion Models

ONE STEP DIFFUSION VIA SHORTCUT MODELS

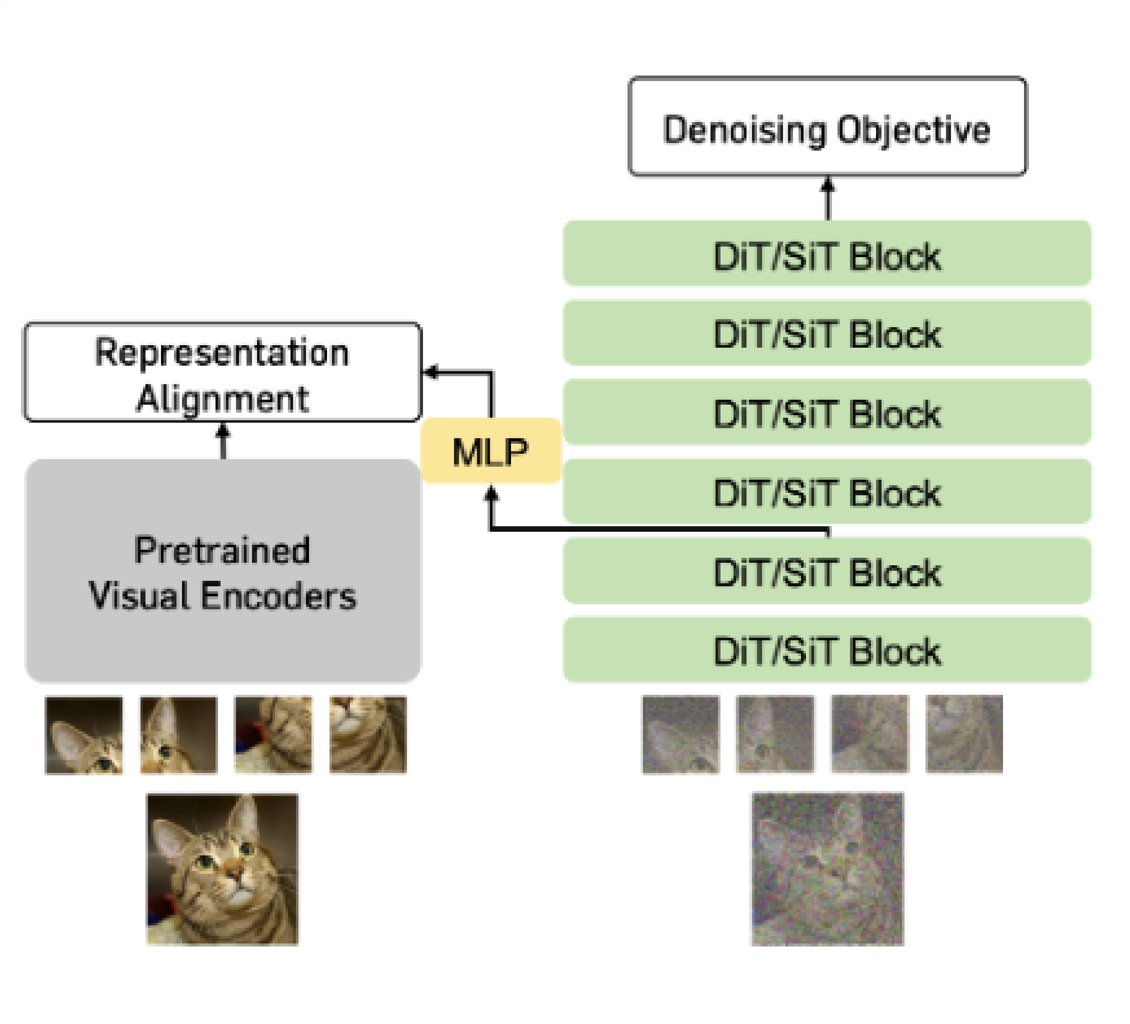
Kevin Frans
UC Berkeley
kvfrans@berkeley.edu

Danijar Hafner
UC Berkeley

Sergey Levine
UC Berkeley

Pieter Abbeel
UC Berkeley

avaiblable at Arxiv on Oct 16


Denoising Objective
DiT/SiT Block
DiT/SiT Block
DiT/SiT Block
DiT/SiT Block
DiT/SiT Block
DiT/SiT Block
Representation Alignment
MLP
Pretrained Visual Encoders

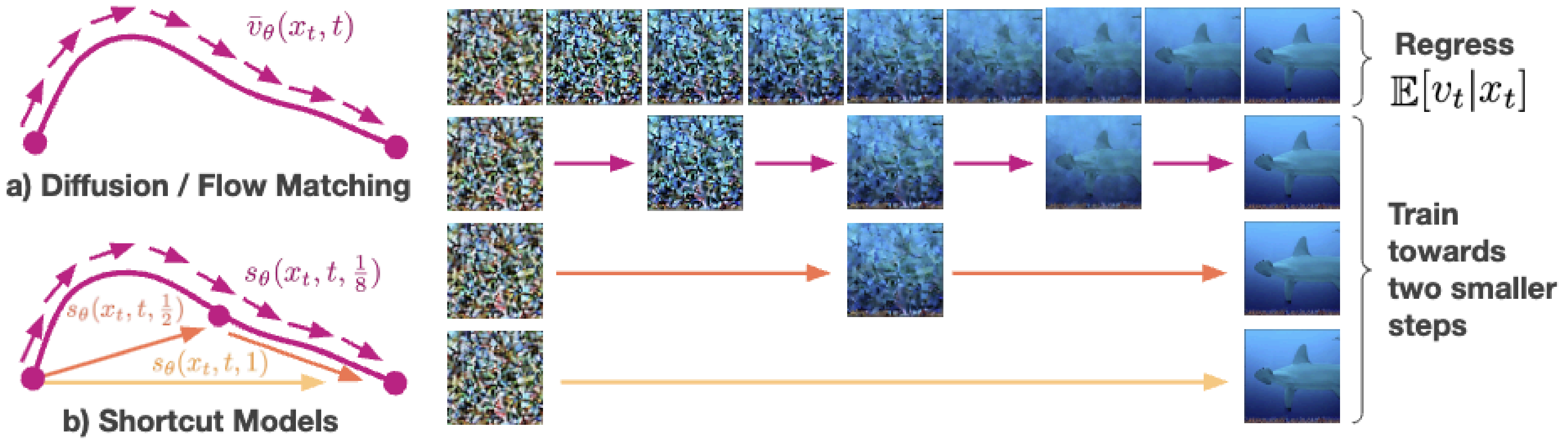


Figure 3: **Overview of shortcut model training.** At $d \approx 0$, the shortcut objective is equivalent to the flow-matching objective, and can be trained by regressing onto empirical $\mathbb{E}[v_t|x_t]$ samples. Targets for larger d shortcuts are constructed by concatenating a sequence of two $d/2$ shortcuts. Both objectives can be trained jointly; shortcut models do not require a two-stage procedure or discretization schedule.

5.2 Shortcut Models

$$\mathcal{L}^{\mathrm{S}}(\theta) = E_{x_0 \sim \mathcal{N},\, x_1 \sim D,\, (t,d) \sim p(t,d)} \Big[\underbrace{\| s_\theta(x_t, t, 0) - (x_1 - x_0) \|^2}_{\text{Flow-Matching}} + \underbrace{\| s_\theta(x_t, t, 2d) - s_{\text{target}} \|^2}_{\text{Self-Consistency}} \Big], \tag{5}$$

$$\text{where} \quad s_{\text{target}} = s_\theta(x_t, t, d)/2 + s_\theta(x'_{t+d}, t, d)/2 \quad \text{and} \quad x'_{t+d} = x_t + s_\theta(x_t, t, d)d.$$

6. Plan for Next Week

6. Plan for Next Week

1. Explore relative works from NeurIPS 2024 Accepted List and ICLR 2025 Submmision List
2. Getting familiar with and running the code of other Bayesian Causal Discover Methods, such as, DiBS, BCD Nets, MC^3...
3. Further reading about Generative Model related Uncertainty/Causal Discovery