Paper Presentation: Deep Structural Causal Models for Tractable Counterfactual Inference

Chang Yan from group 9: Predicting Hemorrhage Related Outcomes with CT Volumetry for Traumatic Hemothorax

https://arxiv.org/abs/2006.06485

2021/4/13

Pawlowski, N., Castro, D. C., & Glocker, B. (2020). Deep structural causal models for tractable counterfactual inference. arXiv preprint arXiv:2006.06485.

Recap of our project

We are developing deep-learning based algorithms to perform 3D segmentation on CT scans, and predict hemothorax volume as voxel count accordingly. Also, the 3D segmentation result helps human operator to assess the quality of volume prediction.

Reason for choosing the paper

01 Deep Structural Causal Models are novel inventions aiming to improve the performance of deep neural networks on image prediction

02 One of their experiments was to predict MRI scans, which has some similarity to our task

03 Their project is open-source and uses PyTorch and Pyro based algorithms

04 The DSCM focuses on the causality, which provides connection between parameters in deep networks and the actual events happening

05 The paper was first published in less than 1 years ago, reflecting novel and cutting-edge developments of deep learning

Author's problems on current DL models

01 DL is known to be susceptible to learning spurious correlations

02 DL tend to amplify biases

03 DL is exceptionally vulnerable to changes in the input distribution

Author's problems on current SCM models

01 SCMs are typically employed with simple linear mechanisms

02 works well for scalar variables and can be useful for decision making, but is not flexible enough to model higher-dimensional data such as images

Author's goals

01 Develop a general framework for building structural causal models (SCMs) with deep learning components, called DSCMs, to solve the problems above

02 Model counterfactual inference that is missing from existing deep causal learning methods

Significance of study

01 Causal DL models could be capable of learning relationships from complex high-dimensional data

02 By explicitly modelling causal relationships and acknowledging the difference between causation and correlation, causality becomes a natural field of study for improving the transparency, fairness, and robustness of DL based systems

03 The tractable inference of deep counterfactuals enables novel research avenues that aim to study causal reasoning on a per instance rather than population level

Background information

Pearl's ladder of causation

- Association

describes reasoning about passively observed data. Correlations in the data

"What are the odds that I observe. . . ?"

- Intervention

concerns interactions with the environment. It requires knowledge beyond just observations

"What happens if I do. . . ?"

- Counterfactuals

hypothetical scenarios. Counterfactual is the generative processes to imagine alternative outcomes for individual data points

"What if I had done A instead of B?"

Background information

Structural causal models and how them fulfill the ladder of causation

$$\mathfrak{G} := (\mathbf{S}, P(\boldsymbol{\epsilon})) \qquad \mathbf{S} = (f_1, \ldots, f_K)$$

f_k is their structural assignments

$$x_k := f_k(\epsilon_k; \mathbf{pa}_k) \qquad P(\boldsymbol{\epsilon}) = \prod_{k=1}^{K} P(\epsilon_k)$$

x_k is the events, ε_k is the exogenous noise, pa_k is the set of direct causes of x_k
P(ε) is the joint distribution over mutually independent exogenous noise variables

- Association

Embedded in this model

- Intervention

do(x_k := a). disconnect x_k with its parents and change structural assignment f_k. Possible of changing both S and P(ε).

- Counterfactuals

hypothetical retrospective interventions: 'What would x_i have been if x_j were different, given that we observed x?' Only change S, not P(ε)

Background information

Do mathematically in 3 steps

- Abduction:

Predict the 'state of the world' (the exogenous noise, ε) that is compatible with the observations, x, i.e. infer $P(\varepsilon|x)$

- Action:

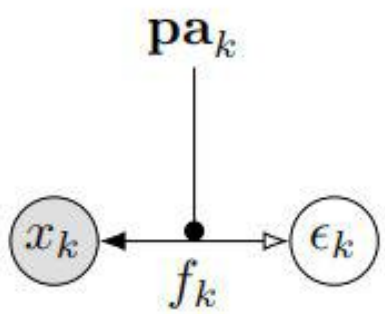
Perform an intervention (e.g. $do(x_k := x_k')$) corresponding to the desired manipulation, resulting in a modified SCM $G' = (S', P(\varepsilon|x))$

- Prediction:

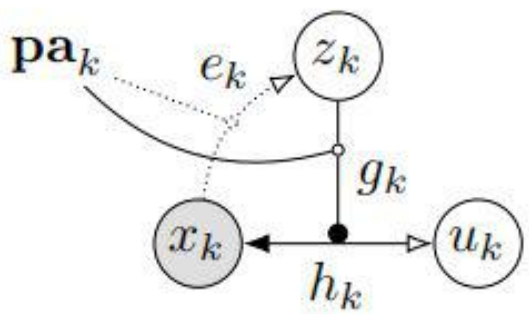
Compute the quantity of interest based on the distribution entailed by the new counterfactual SCM as P(x).

Author's work and implenmetations

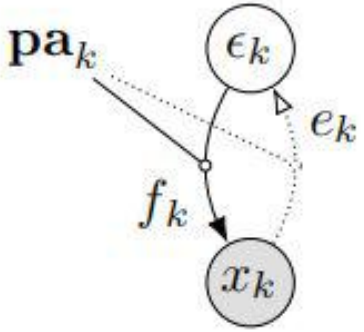
They use recent advances in normalizing flows and variational inference to model mechanisms for composable DSCMs that enable tractable counterfactual inference.



(a) Invertible explicit likelihood



(b) Amortised explicit likelihood



(c) Amortised implicit likelihood

Flow based model Variation approximation Not used

Here, white arrows indicates abductive direction, and black arrows indicates generative direction. Dotted lines are amortized variational approximation. f_k is the forward model, e_k is an encoder that amortizes abduction in non-invertible mechanisms, g_k is a 'high-level' non-invertible branch (e. g. a probabilistic decoder), and h_k is a 'low-level' invertible mapping (e.g. reparametrization)

Author's work and implenmetations

Deep counterfactual inference algorithm they implemented:

- Abduction:

Use the trained encoder ej to approximate ε_j
And calculate approximated P(ε|x, pak)

- Action:

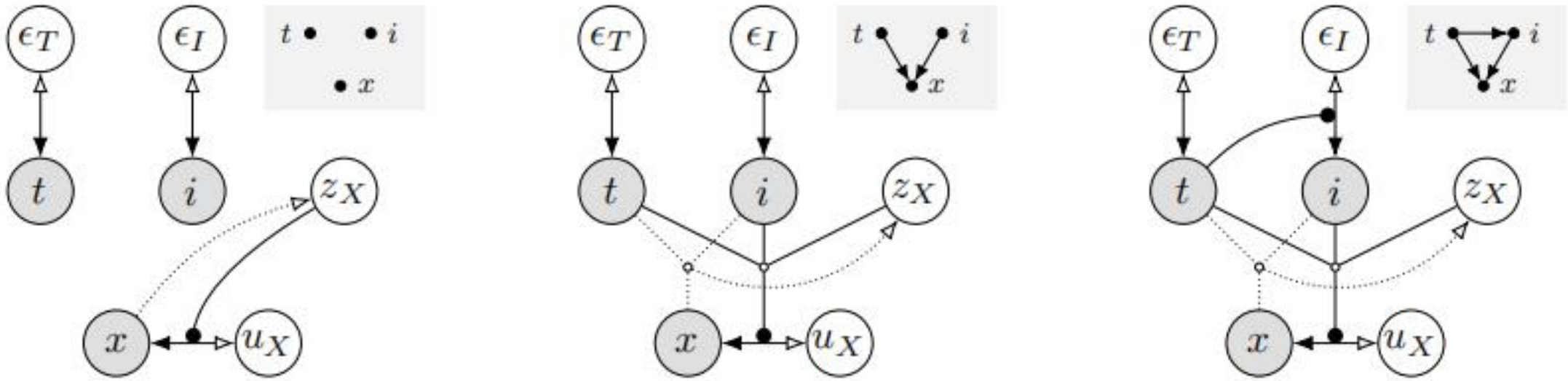
Replace x_k by either a constant $x_k := x_k'$ or by surrogate mechanism $x_k := f_k'(\varepsilon_k, pa_k)$ resulting in a modified SCM G' = (S', P(ε|x))

- Prediction:

First approximate the counterfactual distribution using Monte Carlo method.
Then sample from the distribution using uncorrelated Gaussian decoder for images.

Authors' experiments and results

Case Study I: Morpho-MNIST

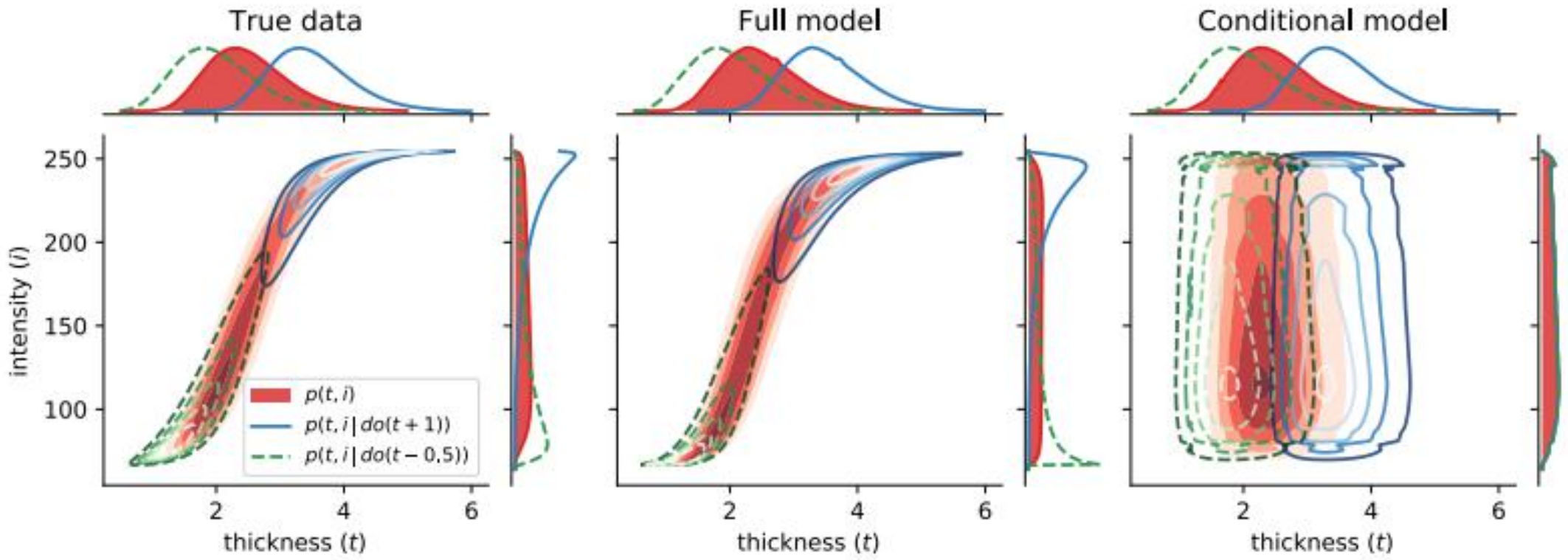


(a) Independent (b) Conditional (c) Full

They test three different models in a synthetic dataset based on MNIST digits, where they defined stroke thickness to cause the brightness of each digit: thicker digits are thicker digits are brighter whereas thinner digits are dimmer.

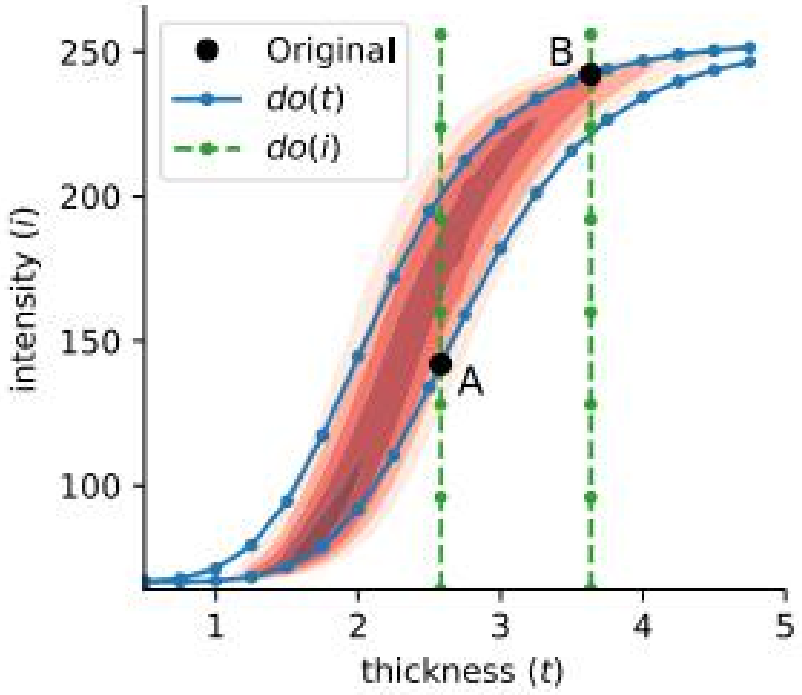
Authors' experiments and results

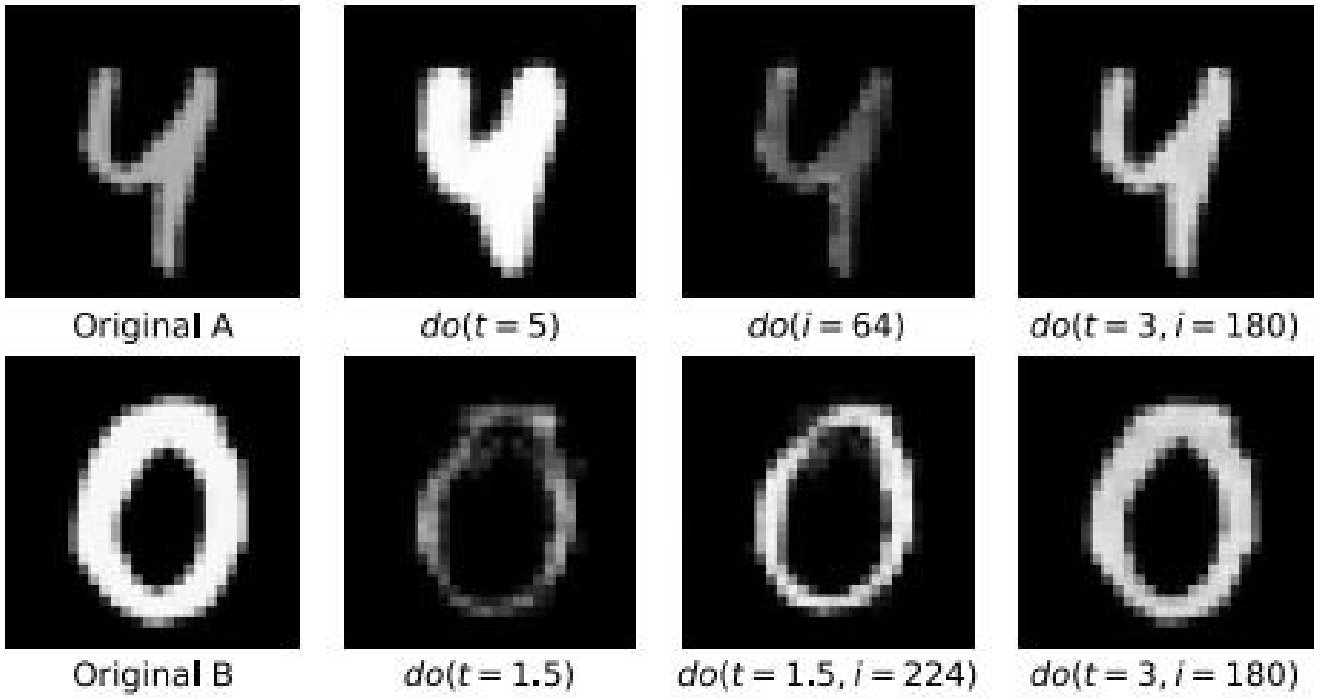
Case Study I: Morpho-MNIST



Authors' experiments and results

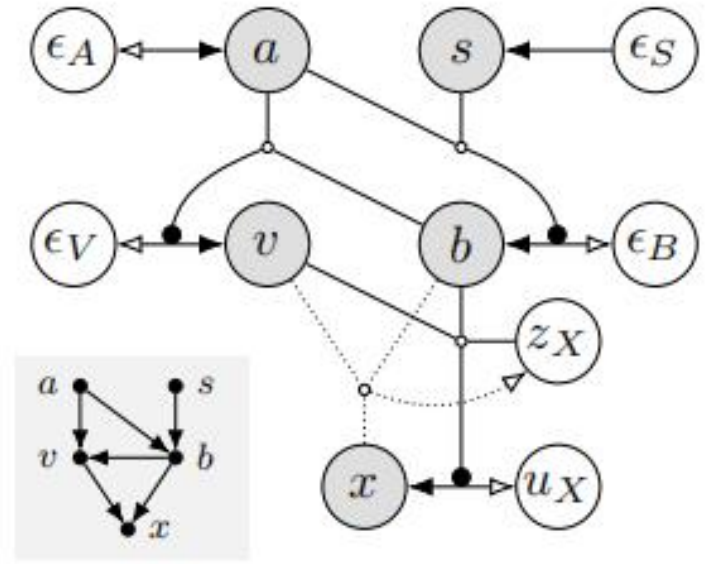
Case Study I: Morpho-MNIST



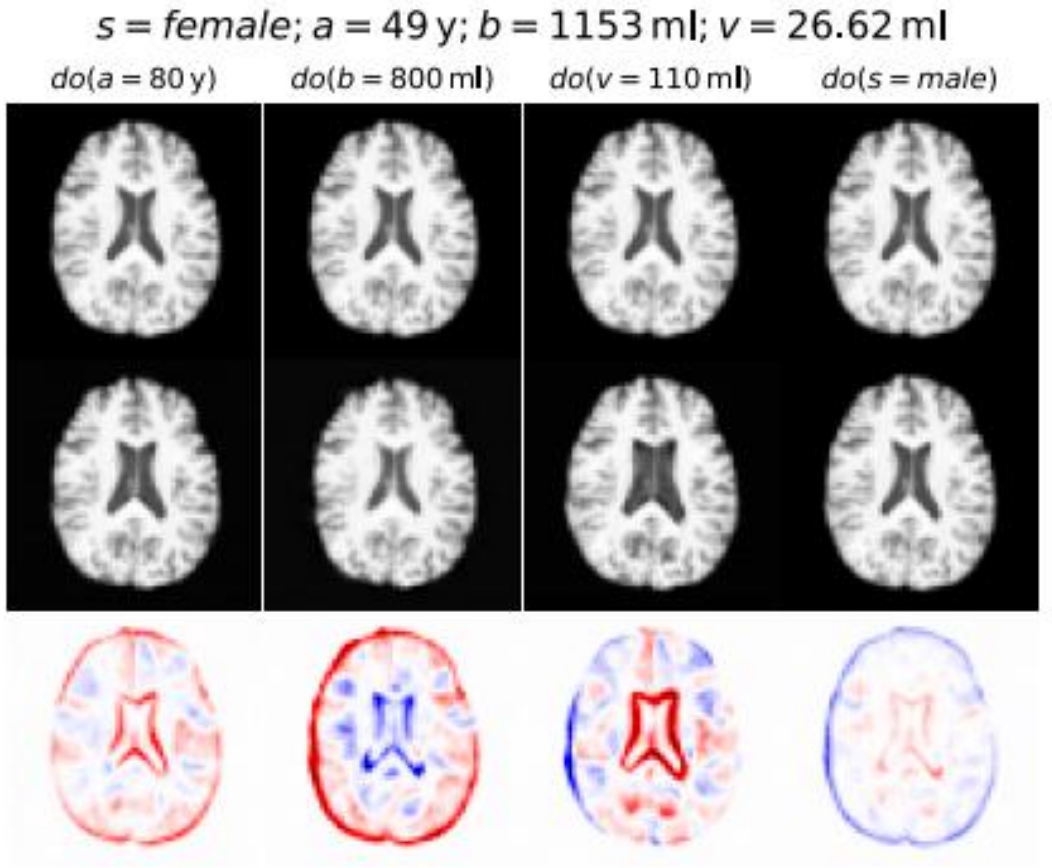


Authors' experiments and results

Case Study 2: Brain MRI



(a) Computational graph



(b) Original image, counterfactuals, and difference maps

Importance and relevance to me:

1. It provides me with a deep insight with SCMs and how they can be combined with and improve DL.

2. The causal inference they developed could possibility be added to our model to improve explainability and assess confidence.

Good points they did:

1. A novel process of integrating SCM and DL with sufficient mathematical basis, and also proposed specific mathematical ways to model counterfactual inference, which other studies fail to achieve.

2. Used casual graphs to explain their model designs.

3. Provided a set of pictures to show the affect of changing each variables, easy to understand.

Criticisms:

1. Although they properly defined the mathematical basis, they did not talk much into the actual structure of the neural networks they design. However, the deep network structure itself is also crucial to the performance.
2. What's worse, their code is completely undocumented, and badly structured.
3. They actually used over 10 different decoders in their code, but in their paper they only talked about the Gaussian decoder, not the others.

Criticisms:

4. When assessing the casual relationship in the brain MRI model, they have 4 variables and only change one at a time.

5. On experiment one, when the i variable is neither direct nor indirect cause of t, changing i does not change t at all. However, on experiment two, the s variable is also neither direct nor indirect cause of v, but apparently changing s has some affect on v. They did not say anything about why there is a difference.

Criticisms:

6. Most of their assessment of causality in results are based on qualitative analysis, not quantitative. It would be better if they can say something mathematically about how accurate their prediction is. They only showed that the DSCM can model causality, but not tested if those modeled relationships are actually correct and how they compared to ground truth.

Possible next steps

1. Develop a way to use DSCM to discover implicit causalities, instead of only modelling assumed causalities.

2. Add a more robust and mathematical way to assess the quality of prediction.

Thank You