COMS W4775: Causal Inference: Project Report
Causal Physical Reasoning

Lovish Chum, Chengzhi Mao
UNI: lc3454, cm3797
Group 6

November 15, 2020

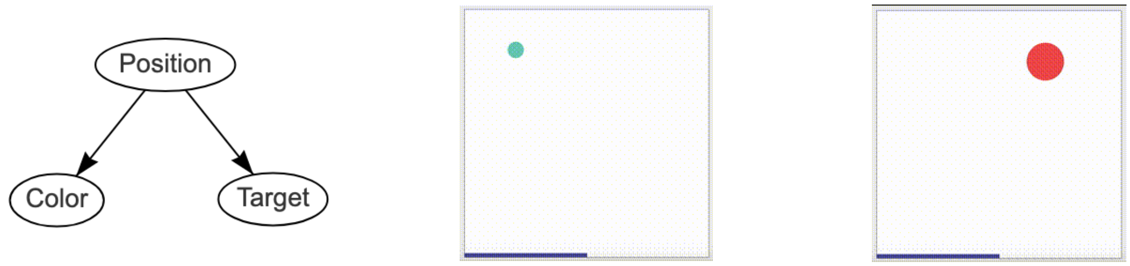
Causal Computer Vision

Visual perception is one of the key factors influencing human behavior. *Visual cause* of behavior is easily comprehensible to humans. Take an example of a bridge with cars on it; see Figure 1. Current computer vision algorithms can adequately answer questions like *What are the positions of the cars in the image ?*. However, the same techniques which are successful in answering these associational queries using a large amount of data fail to answer interventional questions. As an example, we can formulate questions like *What if we remove the car from the image ?* or *What if we change the color of the bridge in the image ?*, or more so *What if we remove the bridge from the image ?*

Moreover, empirical studies point that the lack of correct causal modeling in these techniques is a roadblock to generalization [1, 2, 3]. Earlier work addressing causality in a visual data include [4, 5]. While [5] target the causal relationships between objects in the scenes in observational perspective, [4] take the path of performing interventional experiments on the images.

Figure 1: Changing the color of a car or removing a car would not have significant consequences on rest of the image. On the other hand, removing the bridge will change the semantics of the image.

In this work, we come up with a simulated dataset for determining visual causes from a simulated video sequence. As determining visual causes in the "wild" is complicated, we simplify the problem to visual physical reasoning. Section 1 explains the dataset creation process in detail. Experiments pertaining to one particular task in the dataset are described in Section 2. Section 4 concludes the report and provides an outline to the future work. The dataset and corresponding code is available at this link.



(a) Causal graph $\mathcal{G}$ (b) Target met (c) Target not met

Figure 2: Causal graph dictates weather target (ball meeting the blue floor) will be met in future frames. In the contained in the dataset, green ball always meets the target but red ball always misses the target. For more examples, visit github.

1 Dataset Creation

We introduce Causal-PHYRE, a causal physical reasoning dataset. Build upon PHYRE[6], the dataset contains a set of tasks in the 2-D world. A task consists of an initial state and target. The initial state is pre-determined configuration of objects within the environment. The goal is a relation between two objects which need to be satisfied when the simulation ends. For all the tasks, we follow the relation touching, in concurrence with PHYRE.

In our case, each task is pre-determined based on a causal graph as illustrated in Figure 2. Furthermore, every task has an associated goal state (eg. *make the ball touch the blue floor* etc.) which we call *target* and an initial state in which the goal may or may not be satisfied. For the example in Figure 2, it is intuitive for us that color is not the visual cause of the ball meeting the blue floor. Instead, the position of the ball is what matters in determining whether the ball meets the target. Causal-PHYRE is designed keeping in mind the following points:-

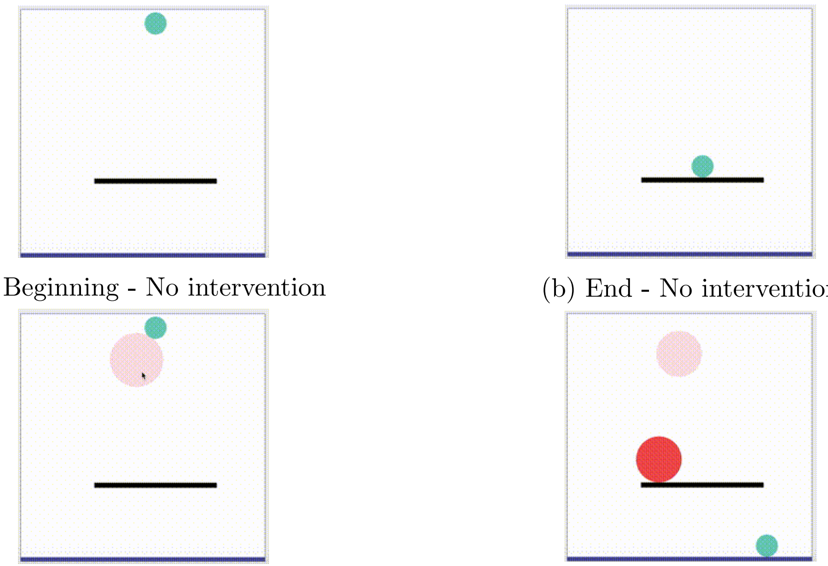
- **Spurious Correlations** - The causal graphs for tasks are created in a manner that a spurious correlation exists between target and another entity. This may be *color*, *size*, *shape* or *position* of an object. Ideally, we want to have a model which avoids the correlate and determines the cause of the target.
- **Identifiable Cause** - Cause of the target (ie. effect) is visually identifiable and included in the graph $\mathcal{G}$.
- **Interventions** - The user can intervene on the target by placing another object in the environment or changing the color or position of an existing object. An example intervention is depicted in Figure 3.

2 Simulations

2.1 Can Neural Network identify Causal Structure ?

We create simulations for experiment in Figure 2. We know that the ball meeting the blue floor is caused by the ball's horizontal position, but we make all the balls that are going to succeed green in color, and the balls which are going to fail in meeting the target red in color.

Theoretically, without knowing the causal graph, it is impossible to infer the correct causal structure from the observed distribution. We wonder how the model would perform by only relying on the observed



(a) Beginning - No intervention (b) End - No intervention

(c) Beginning - Intervention (d) End - Intervention

Figure 3: Intervention through placing a ball within the scene.

distribution. We first train the model with a neural network. Results in Fig. 4. Interestingly, the model does not totally fail if early stopped properly. The yellow test accuracy higher than 50% reveals the model finds some causality. It can achieve 72% accuracy in Fig 4, though after training long enough, it overfits and reduced to 50%, where all causality is lost.

The experiment reveals that although it is considered to be theoretically impossible to find the causal structure, the current neural network method can find partially detect the causal structure based on observed data only. This may explain the success of neural networks without training with the causal information explicitly, and it is also consistent with the finding of neural network overfitting and early stopping on real-world data.

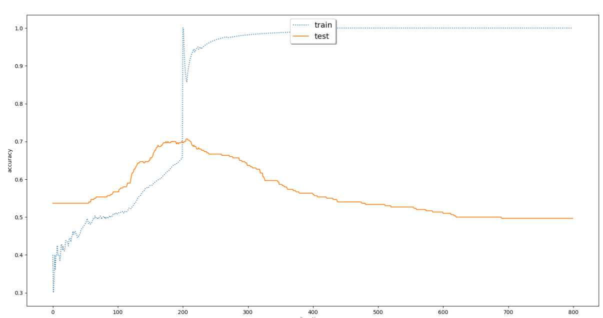
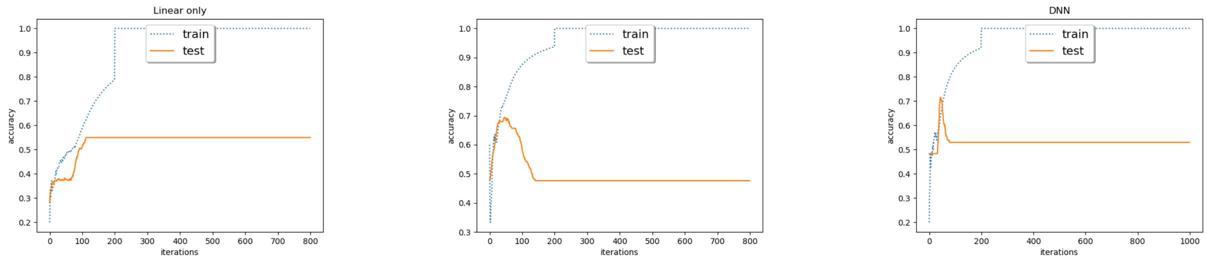


Figure 4: Can neural network model solve the bow graph? Theoretically, it is impossible. In practice, we find that it is partly possible. The yellow is the test accuracy.



(a) Linear (50% Accuracy) (b) Two Layers (69% Accuracy) (c) Five layers (72% Accuracy)

Figure 5: Model architecture in causal discovery. The results show that deep neural network biases the model towards causal learning, while a linear regression tends to find the trivial correlation instead of the causality.

2.2 Do Neural Network have Causal Bias ?

Subsection 2.1 reveals that it is possible to discover the part of the causal structure with a neural network. In this section, We investigate whether the architecture of the neural network matter.

We conduct experiments on linear neural network (linear regression), a two-layer, and a five-layer network. Results are shown in Figure 5. We find that using deeper architecture enables the model to learn partial causal structure, while a linear model tends to fail. This experiment explains the reason why neural network works well for the current application. We believe this is due to neural network introducing a causal inductive bias to the learning.

3 Intervention Method

3.1 Intervention for Causal Learning

As we learned in class, it is theoretically impossible to learn causal structure by only using the observed distribution as it violates the Pearl Causal Hierarchy(PCH). We introduce interventions, taking advantage of our physical simulation engine.

We apply the following algorithm:

Algorithm 1: Intervention-based Causal Discovery

initialize a neural network;
while *iteration* $<$ *epochs* **do**
  train neural network for the task;
  for *each attribution* **do**
    generation intervention value for that attribute,
    query groundtruth results from simulation engine,
    retrain the model
  end
end
Return model

Using the above algorithm, we find the model is able to learn the causality 99%. Figure 6 depicts the variation of the train and test accuracy with the iterations.

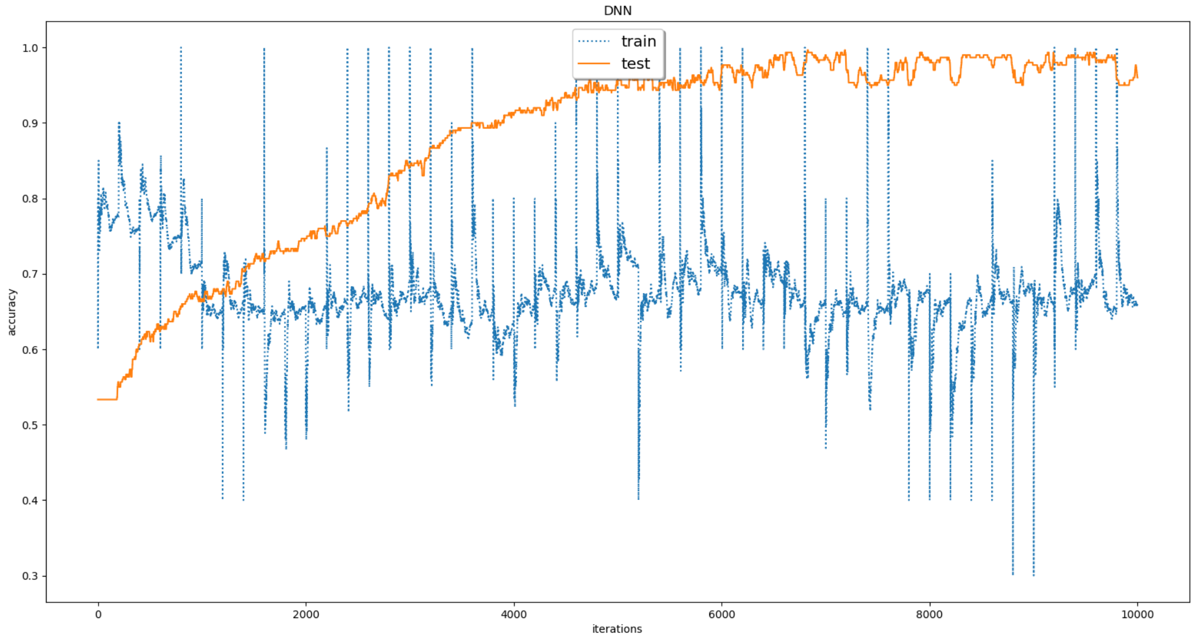


Figure 6: The model is able to learn the causality under intervention. The yellow is the test accuracy, which reveals the causality. The test accuracy approaches 100% as training converged, which shows the model is able to capture the causality.

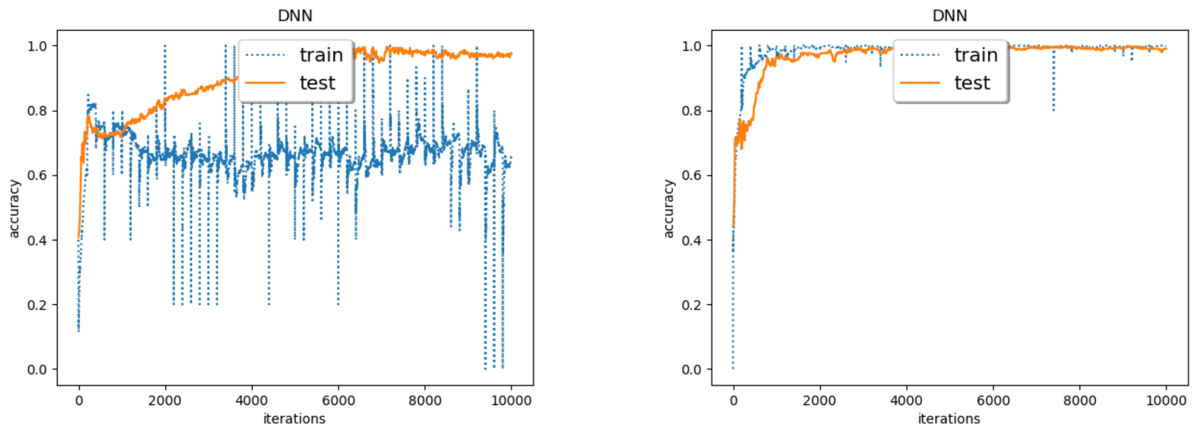
3.2 Gradient Based Intervention

We propose Gradient-based intervention algorithm to reduce the number of steps required to query the physical engine. Our intuition is that gradient suggests the direction which causes the model to change the output. Thus, we assume that this direction coincides with the cause of success (or meeting the target).

Algorithm 2: Gradient-based Intervention for Causal Discovery

```
initialize a neural network;
while iteration < epochs do
    train neural network for the task;
    Calculate the gradient for each of the input attribution;
    Pick up the attribute with the largest gradient;
    generation intervention value for that attribute,
    query groundtruth results from simulation engine,
    retrain the model
end
Return model
```

As shown in Figure 7, our gradient-based intervention strategy is 10 times more efficient than random intervention. As it is often expensive to do intervention experiment, our algorithm can be used to discover causality for the physical world.



(a) Random Intervention

(b) Gradient-based Intervention

Figure 7: The proposed gradient-based intervention speeds up the causal discovery of the model with less query to the physical engine.

4 Conclusions

In this work, we create a physical reasoning dataset to check weather the current methods are suitable for causal discovery. We conduct experiments for one particular task in the dataset. We discover that causality can be thought of as an inductive bias for current deep neural networks if we apply the early stopping strategy. By doing a random intervention, our model can learn the causality of the data. We also propose a gradient-based intervention, which reduces the number of intervenes required to learn the correct, even better causal relationship of the data.

Yet the baseline methods studied in this report are just a small step towards the goal of visual causal feature discovery. Our future work entails applying these methods to other tasks in the dataset. Also, the dataset can be improved by sampling tasks from a more complicated graph. Conducting experiments on a convoluted graph would test the limit of current methods.

References

[1] B. M. Lake, T. D. Ullman, J. B. Tenenbaum, and S. J. Gershman, "Building machines that learn and think like people," *Behavioral and brain sciences*, vol. 40, 2017.

[2] M. Edmonds, J. Kubricht, C. Summers, Y. Zhu, B. Rothrock, S.-C. Zhu, and H. Lu, "Human causal transfer: Challenges for deep reinforcement learning.," in *CogSci*, 2018.

[3] Y. Bengio, T. Deleu, N. Rahaman, R. Ke, S. Lachapelle, O. Bilaniuk, A. Goyal, and C. Pal, "A meta-transfer objective for learning to disentangle causal mechanisms," *arXiv preprint arXiv:1901.10912*, 2019.

[4] K. Chalupka, P. Perona, and F. Eberhardt, "Visual causal feature learning," *arXiv preprint arXiv:1412.2309*, 2014.

[5] D. Lopez-Paz, R. Nishihara, S. Chintala, B. Scholkopf, and L. Bottou, "Discovering causal signals in images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6979–6987, 2017.

[6] A. Bakhtin, L. van der Maaten, J. Johnson, L. Gustafson, and R. Girshick, "Phyre: A new benchmark for physical reasoning," in *Advances in Neural Information Processing Systems*, pp. 5083–5094, 2019.