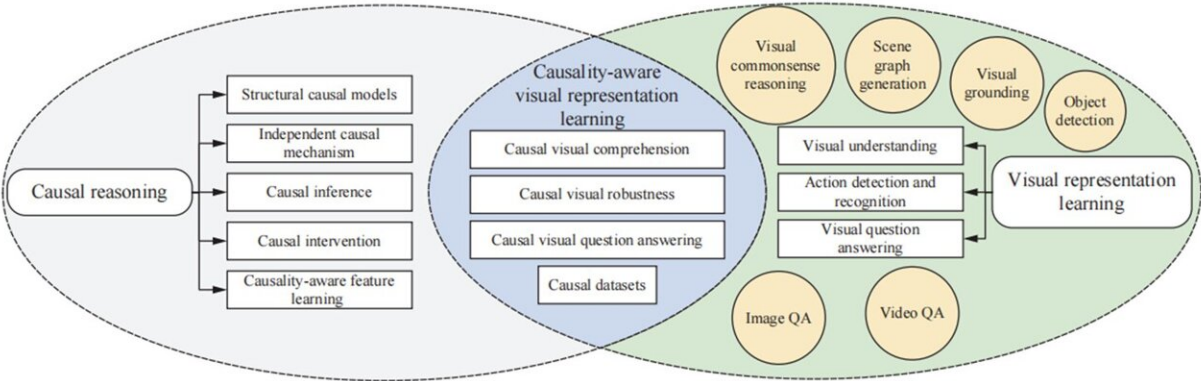


# Causal reasoning meets visual representation learning: A prospective study

November 22 2023



Overview of the structure of this paper, including the discussion of related methods, datasets, challenges, and the relations among causal reasoning, visual representation learning, and their integration. Credit: Beijing Zhongke Journal Publishing Co.

With the emergence of huge amounts of heterogeneous multi-modal data—including images, videos, texts/languages, audios, and multi-sensor data—deep learning-based methods have shown promising performance for various computer vision and machine learning tasks, such as visual comprehension, video understanding, visual-linguistic analysis, and multi-modal fusion.

However, existing methods rely heavily upon fitting the data

distributions and tend to capture the spurious correlations from different modalities, and thus fail to learn the essential causal relations behind the multi-modal knowledge, which has good generalization and cognitive abilities.

Inspired by the fact that most of the data in computer vision society are independent and identically distributed (i.i.d.), a substantial body of literature has adopted data augmentation, pre-training, self-supervision, and novel architectures to improve the robustness of the state-of-the-art deep neural network architectures. However, it has been argued that such strategies only learn correlation-based patterns (statistical dependencies) from data and may not generalize well without the guarantee of the i.i.d setting.

Due to its powerful ability to uncover the underlying structural knowledge about data-generating processes that allow interventions to generalize well across different tasks and environments, causal reasoning offers a promising alternative to correlation learning.

Recently, causal reasoning has attracted increasing attention in myriad high-impact domains within computer vision and machine learning, such as interpretable deep learning, causal feature selection, visual comprehension, visual robustness, visual question answering, and video understanding. A common challenge of these causal methods is how to build a strong cognitive model that can fully discover causality and spatial-temporal relations.

In their paper, the researchers aim to provide a comprehensive overview of causal reasoning for visual representation learning, attracting attention, encouraging discussions, and bringing to the forefront the urgency of developing novel causality-guided visual representation learning methods.

Although there are some surveys about causal reasoning, these works are intended for general representation learning tasks such as deconfounding, out-of-distribution (OOD) generalization, and debiasing.

The work is [published](#) in the journal *Machine Intelligence Research*.

Uniquely, this paper focuses on the systematic and comprehensive survey of related works, [datasets](#), insights, future challenges and opportunities for causal reasoning, visual representation learning, and their integration. To present the review more concisely and clearly, this paper selects and cites related works by considering their sources, publication years, impact, and the coverage of different aspects of the topic surveyed in this paper.

Overall, the main contributions of this work are as follows.

Firstly, this paper presents the basic concepts of causality, the structural causal model (SCM), the independent causal mechanism (ICM) principle, causal inference, and causal intervention. Then, based on the analysis, the paper further gives some directions for conducting causal reasoning on visual representation learning tasks. This paper may be the first that proposes potential research directions for causal visual representation learning.

Secondly, a prospective review is introduced to systematically and structurally evaluate existing works according to their efforts in the above-pointed directions for conducting causal visual representation learning more efficiently. Researchers focus on the relation between visual representation learning and causal reasoning and provide a better understanding of why and how existing causal reasoning methods can be helpful in visual representation learning, as well as providing inspiration for future research and studies.

Thirdly, the new paper explores and discusses future research areas and open problems related to using causal reasoning methods to tackle visual representation learning. This can encourage and support the broadening and deepening of research in related fields.

Section 2 provides the preliminaries, which include five parts. Its first part is the basic concepts of causality. Causal learning is different from statistical learning, which aims to discover causal relationships beyond statistical relations. Learning causality requires machine learning methods not only to predict the outcome of i.i.d. experiments but also to reason from a causal perspective.

The second part is the SCM, which considers the formulation of a causality style. The third part is the ICM principle, which describes the independence of causal mechanisms. The fourth part is causal inference, whose purpose is to estimate the outcome shift (or effect) of different treatments. The last part is causal intervention, which aims to capture the causal effects of interventions (i.e., variables) and take advantage of causal relations in datasets to improve model performance and generalization ability.

Traditional feature learning methods usually learn the spurious correlation introduced by confounders. This will reduce the robustness of models and make models hard to generalize across domains. Causal reasoning, a learning paradigm that reveals the real causality from the outcome, overcomes the essential defect of correlation learning and learns robust, reusable, and reliable features.

In Section 3, researchers review the recent representative causal reasoning methods for general feature learning, which mainly consist of three main paradigms: 1) structural causal model (SCM) embedded, 2) applying causal intervention/counterfactual, and 3) Markov boundary (MB) based feature selection.

Visual representation learning has made great progress in recent years, and can utilize spatial or/and temporal information to complete specific tasks, including visual understanding (object detection, scene graph generation, visual grounding, visual commonsense reasoning), action detection and recognition, and visual question answering, etc.

In Section 4, researchers introduce these representative visual learning tasks and discuss the existing challenges and necessity of applying causal reasoning to visual representation learning.

According to the above-discussed visual representation learning methods, the current machine learning, especially representation learning, faces several challenges: 1) lack of interpretability, 2) poor generalization ability, and 3) over-reliance on correlations of data distribution. Causal reasoning offers a promising alternative to address these challenges.

The discovery of causality helps uncover the causal mechanism behind the data, allowing the machine to understand better why and to make decisions through intervention or counterfactual reasoning.

In Section 5, researchers summarize some recent approaches for causal visual representation learning. Visual representation learning is an emerging research topic and has appeared since the 2020s. The related tasks can be roughly categorized into several main aspects: 1) causal visual understanding, 2) causal visual robustness, and 3) causal visual question answering. In this section, researchers discuss these three representative causal visual representations of learning tasks.

Correlation-based models may perform well in existing datasets, not because these models have a strong reasoning capability, but because these datasets cannot fully support the evaluation of the models' reasoning capability. Spurious correlations in these datasets can be

exploited by the model to cheat, which means that the model just concentrates on superficial correlation learning, not real causal reasoning, only approximating the distribution of the dataset.

For example, in the VQA v1.0 dataset for the VQA task, the model simply answers "yes" when seeing the question "Do you see a ...", which will achieve nearly 90% accuracy. Due to this shortcoming in current datasets, researchers need to build benchmarks that can evaluate the true causal reasoning capability of models.

In Section 6, researchers use image question-answering benchmarks and video question-answering benchmarks as examples to analyze the current research situation of related causal reasoning datasets and give some future directions.

Section 7 proposes and discusses some future research directions. Causal reasoning with visual representation learning has a variety of applications. Modeling causal reasoning for a variety of tasks can achieve a better perception of the real world. In this section, researchers introduce the applications from five aspects: image/video analysis, explainable artificial intelligence, recommendation system, human-computer dialog and interaction, and crowd intelligence analysis.

They also discuss how causal reasoning benefits various real-world applications.

Some researchers have successfully implemented causal reasoning for visual representation learning to discover causality and visual relations. However, causal reasoning for visual representation learning is still in its infancy stage, and many issues remain unsolved. Therefore, Section 8 highlights several possible research directions and open problems to inspire further extensive and in-depth research on this topic.

Potential research directions for causal visual representation learning can be summarized as:

1. more reasonable causal relation modeling
2. more precise approximation of intervention distributions
3. more proper counterfactual synthesizing process
4. large-scale benchmarks and evaluation pipeline

This paper provides a comprehensive survey on causal reasoning for visual representation learning. Researchers hope that this survey can help attract attention, encourage discussions, and bring to the forefront the urgency of developing novel causal [reasoning](#) methods, publicly available benchmarks, and consensus-building standards for reliable visual representation learning and related real-world applications more efficiently.

**More information:** Yang Liu et al, Causal Reasoning Meets Visual Representation Learning: A Prospective Study, *Machine Intelligence Research* (2022). [DOI: 10.1007/s11633-022-1362-z](https://doi.org/10.1007/s11633-022-1362-z)

Provided by Beijing Zhongke Journal Publishing Co.

Citation: Causal reasoning meets visual representation learning: A prospective study (2023, November 22) retrieved 29 April 2024 from <https://techxplore.com/news/2023-11-causal-visual-representation-prospective.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.