OOD in causal perspective: from domain generalization to plankton image analysis

Authors: Shiyang Yan

Presenter: Shiyang Yan

Shiyang.Yan@inria.fr

April 20, 2023

Overview

Introduction to Domain Generalization

- Domain Generalization (DG) is a challenging transfer learning task, which aims to train a recognition model in several known domains and test it in an unknown target domain.
- In this presentation, we analyze the DG problem from a viewpoint of causal inference.

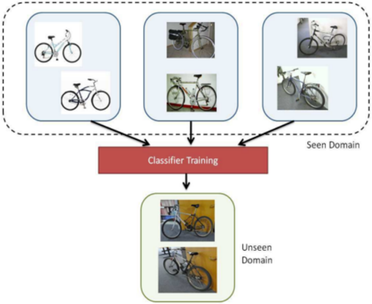

Figure: An illustration of the domain generalization task.

Background(Causality)

- Ladder of Causation: Causality provides two distinct layers on top of traditional statistics.
- (Pearl et al.) characterizes this relationship as the Ladder of Causality.
- The Ladder has three levels: Seeing (statistics, the current prediction model), Doing (intervention, what we target in this presentation), and Imagining (counterfactuals).
- Confounder is an existing problem in the current prediction model that could be solved by intervention: is a variable that influences both the cause and effect, causing a spurious association.

Background(Back-door Adjustment)

- Intervention: Back-door adjustment and Front-door adjustment (use mediator, like attention).
- drug(cause) $<-->$ (age, gender) $<-->$ health (effect)
- Back-door adjustment: Determine confounder C (age, gender) drive both X (a drug, the cause) and A (health, the effect).
- Find units with the same values for C (same age, same gender), but different values for X, and compute the difference in A. If there is a difference in A between these units, it should be due to X, and not due to C. **We will use this technique in the back-door adjustment for domain generalization.**
- Here C, the confounder is measurable.
- This is a kind of intervention since we control the value of the confounder to find the true cause.

Causal structure in DG

- In DG, we denote the inputs as causes X.
- We denote the actual output of the category classifier as A, the effects.
- A is also termed as the potential outcome, and the remaining challenge is how to only use the training data to estimate it effectively.
- Usually, we would use Monte Carlo estimation to approximate A as $E[A|X = x]$, for example, usual deep net training with stochastic gradient descent (SGD) is an example of Monte Carlo estimation.
- The unobserved confounder C, which affects the input images x and the potential outcome A.
- Both the **background** and **domain** of the images affect both the causes X, and the outcomes A, making them the confounder C.

$$\begin{aligned} X &= f_{generation}(N, C), \\ A &= f_{recognition}(X, C). \end{aligned} \tag{1}$$

Causal Inference in DG

- The Monte Carlo estimation of a recognition model is biased, which is expressed as $E[A|X=x] \neq A$.
- In causal theory, the backdoor adjustment, which requires observing and measuring all the confounders C and associating them with each data sample, is expressed as $E[Y|X=x, C=c] = Y$.
- The confounder C is required to be measurable in a data sample level and be processed together with the input data x.
- Measuring the confounders in a data sample level is extremely hard.
- The deconfounder theory (Wang and Blei, 2019) approximate the proxy confounder via factorization of multiple causes, more formally: $P(x_1, ..., x_N|Z=z) = \prod_n P(x_n|Z=z)$, which means that the z is a good substitute confounder if generative learning can factorize and represent the distribution effectively.

Methods

- As we observe, modelling the DG process with the original deconfounder theory will bring a neglected variable D, which is the domain, in the final causal diagram, making the computation incorrect.

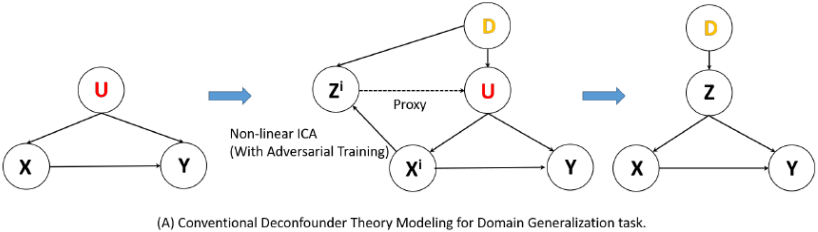

Figure: The causal diagram in our scheme.

Methods: A solution

- Factorization in our case, only makes the $\{Z_i, i = 1, ..., N\}$ mutually independent of each other, and cannot eliminate the dependence link between Z and D.

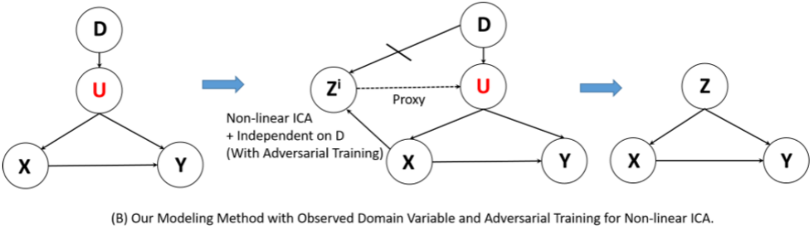

Figure: A solution proposed.

- We would like a restricted factorized model to make the deconfounder theory condition well satisfied, with a new restriction that the derived factors are invariant to D.

Methods: system diagram

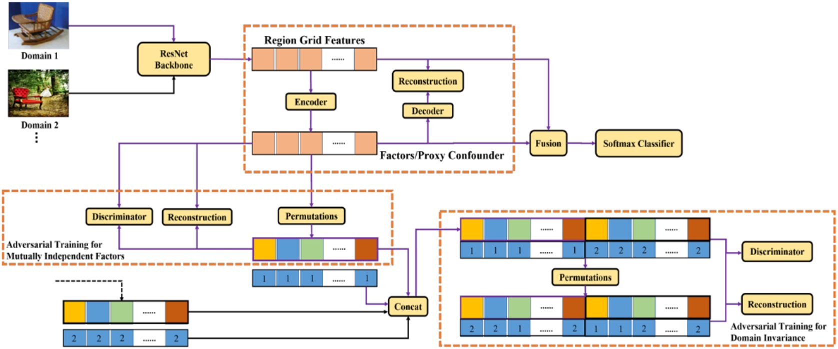

Figure: We achieve permutation invariant via adversarial learning, to fulfil the domain-invariant Non-linear ICA factors.
We fully utilize the 1d convolutions for all the transformations.

Methods: Explanations

- Once we obtain the domain-invariant proxy confounder, we apply the example weighting (attention mechanism) to perform the **back-door** intervention:

$$
\begin{aligned}
&\alpha = X \odot Z, \\
&sim = \exp\left(\frac{\alpha_i}{\sum_{j=1}^{d} \alpha_j}\right), \\
&F_c = Z + sim * X,
\end{aligned}
\tag{2}
$$

change the causes while keeping the confounder unchanged.

- Where α is the dot product of the region features and the proxy confounders, which is subsequently processed via a Softmax, and elementally multiplied with the image features to form F_c, i.e., the feature of confounders.
- The final representation feature is expressed as $F_f = F_c + X$.

Methods: Training objectives

- The training loss is expressed as:

$$L = Loss_{classify} + \lambda Loss_{AdvICA} \tag{3}$$

Numerical Results

Table 1. Comparative assessment of ADVICA on the VLCS dataset.

Methods	Caltech101	LabelMe	Sun09	VOC2007	Avg
Mixup (Yan et al., 2020)	98.3	64.8	72.1	74.3	77.4
MLDG (Li et al., 2018a)	97.4	65.2	71.0	75.3	77.2
MMD (Li et al., 2018b)	97.7	64.0	72.8	75.3	77.5
CDANN (Li et al., 2018c)	97.1	65.1	70.7	77.1	77.5
MTL (Blanchard et al., 2021)	97.8	64.3	71.5	75.3	77.2
SagNet (Nam et al., 2021)	97.9	64.5	71.4	**77.5**	77.8
ARM (Zhang et al., 2021)	**98.7**	63.6	71.3	76.7	77.6
VREx (Krueger et al., 2021)	98.4	64.4	**74.1**	76.2	78.3
RSC (Huang et al., 2020)	97.9	62.5	72.3	75.6	77.1
ADVICA	98.6	**65.1**	**74.1**	76.9	**79.2**

Table 2. Comparative assessment of ADVICA on the PACS dataset.

Methods	Art-painting	Sketch	Cartoon	Photo	Avg
MetaReg (Balaji et al., 2018)	87.20	70.30	79.20	**97.60**	83.60
ERM (Gulrajani & Lopez-Paz, 2021)	88.1	78.0	79.1	97.8	85.7
MASF (Dou et al., 2019)	82.89	72.29	80.49	95.01	82.67
DSON (Seo et al., 2020)	87.04	**82.90**	80.62	95.99	**86.64**
ADVICA	**87.42**	76.33	**83.61**	95.15	85.62

Table 3. Comparative assessment of ADVICA on the Office-Home dataset.

Methods	Art	Clipart	Product	Real World	Avg
MMD-AAE (Saito et al., 2018)	56.5	47.3	72.1	74.8	62.7
CCSA (Motiian et al., 2017)	59.9	49.9	74.1	75.7	64.9
JiGen (Carlucci et al., 2019)	53.0	47.5	71.5	72.8	61.2
CrossGrad (Shankar et al., 2018)	58.4	49.4	73.9	75.8	64.4
FAR (Jin et al., 2020)	61.4	**52.9**	74.5	75.4	66.0
ADVICA	**65.8**	52.3	**75.7**	**78.8**	**68.2**

Visualization

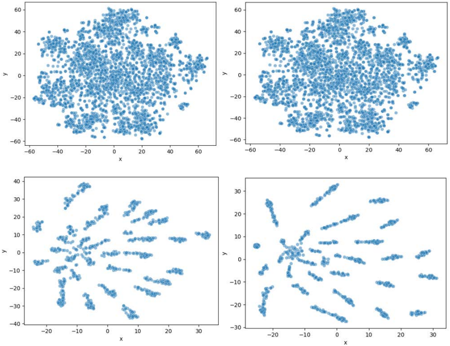

Figure: Quality of the X representation on the VOC2007 (Left) and LabelMe (Right) domains (VLCS benchmark) using a t-SNE visualization. Top: representation learned by ResNet-50. Bottom: representation learned by ADVICA.

Identifiability

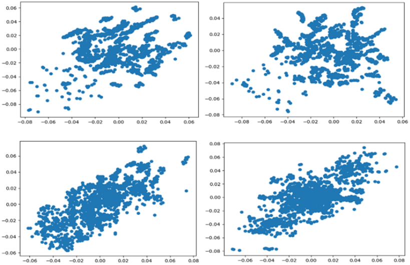

Figure: **We also improve the identifiability of the nonlinear ICA by using the auxiliary domain labels (to make the factors domain-invariant).** Given a pair of images (x_i, x_j), given their associated z_i and z_j (respectively z_i' and z_j' in another run), we see that $||z_i - z_j|| \approx ||z_i' - z_j'||$, in other words, Z and Z' are equivalent up to an isometry (Fig. 1, associating to each pair (i, j) the point with coordinates $(||z_i - z_j||, ||z_i' - z_j'||)$ show that points are close to the diagonal, comparatively to (Brakel & Bengio, 2017)). We do not set priors on Z. Comparing two runs on the VLCS dataset (Q4). Top: after (Brakel & Bengio, 2017). Bottom: AdvICA.

Plankton image analysis

- Monitoring plankton populations in situ are fundamental to preserving the aquatic ecosystem.
- Plankton microorganisms are in fact susceptible to minor environmental perturbations, that can reflect consequent morphological and dynamical modifications.
- Plankton is a collection of aquatic microorganisms floating passively in the water. It plays a big role in the marine ecosystem.

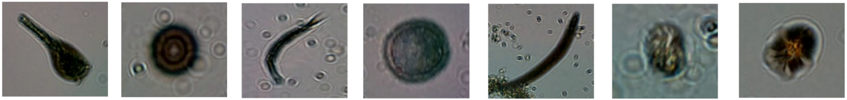

Figure: Some examples of Plankton Images.

OOD in plankton analysis

- OOD in plankton image analysis: anomaly detection and zero-shot learning (different to DG at the first glance, but they are all OOD problems).
- Anomaly detection is required as we need to pre-process the collected plankton images from the natural environment.
- Zero-shot recognition is to facilitate the new species detection, fulfilling the recognition with limited training data but able to recognize large-amount of plankton species.
- These two tasks require causal deconfounding as the confounder deteriorates the recognition accuracy of the classifier.
- The confounder factor in plankton images includes the **background, measurement bias, and possible style of the image**, which should be measured properly.

Questions?

Questions?

Thank you.