IDEA: An Invariant Perspective for Efficient Domain Adaptive Image Retrieval

Haixin Wang[1], Hao Wu[2], Jinan Sun[1], Shikun Zhang[1], Chong Chen[1], Xian-Sheng Hua[3], Xiao Luo[4]

[1]Peking University [2]University of Science and Technology of China [3]Terminus Group [4]UCLA

NEURAL INFORMATION PROCESSING SYSTEMS

TL ; DR

In this paper, we introduce the Invariance-acquired Domain Adaptive Hashing (IDEA) model to address the challenges of unsupervised domain adaptive retrieval. IDEA distinguishes between causal and non-causal image effects, using causal features for generating discriminative hash codes enhanced by consistency learning, and employs a generative model for synthetic sample intervention, minimizing non-causal impact to achieve domain invariance. Comprehensive experiments validate IDEA's superiority over competitive baselines in cross-domain retrieval tasks.

Overview

Problem Definition

Given a source domain $\mathcal{D}^s = \{(\boldsymbol{x}_i^s, y_i^s)\}_{i=1}^{N_s}$ with N_s fully-labeled images and a target domain $\mathcal{D}^t = \{(\boldsymbol{x}_j^t)\}_{j=1}^{N_t}$ with unlabeled N_t images, both domain share a common label space $\mathcal{Y} = \{1, 2, \cdots, C\}$ despite potential distribution shifts. The objective is to develop a hashing-based retrieval model that projects an input image $\boldsymbol{x}$ onto a compact binary code $\boldsymbol{b} \in \{-1, 1\}^L$, where L represents the code length.

Contribution:

Problem Connection. We pioneer a novel perspective that connects invariant learning with domain adaptive hashing for efficient image retrieval.

Novel Methodology. Our method not only disentangles causal and non-causal features in each image following the principle of the information bottleneck, but also ensures hash codes are sufficiently invariant to the intervention of non-causal features.

High Performance. Comprehensive experiments across numerous datasets demonstrate that our method outperforms a range of competitive baselines in different settings.

Our Approach

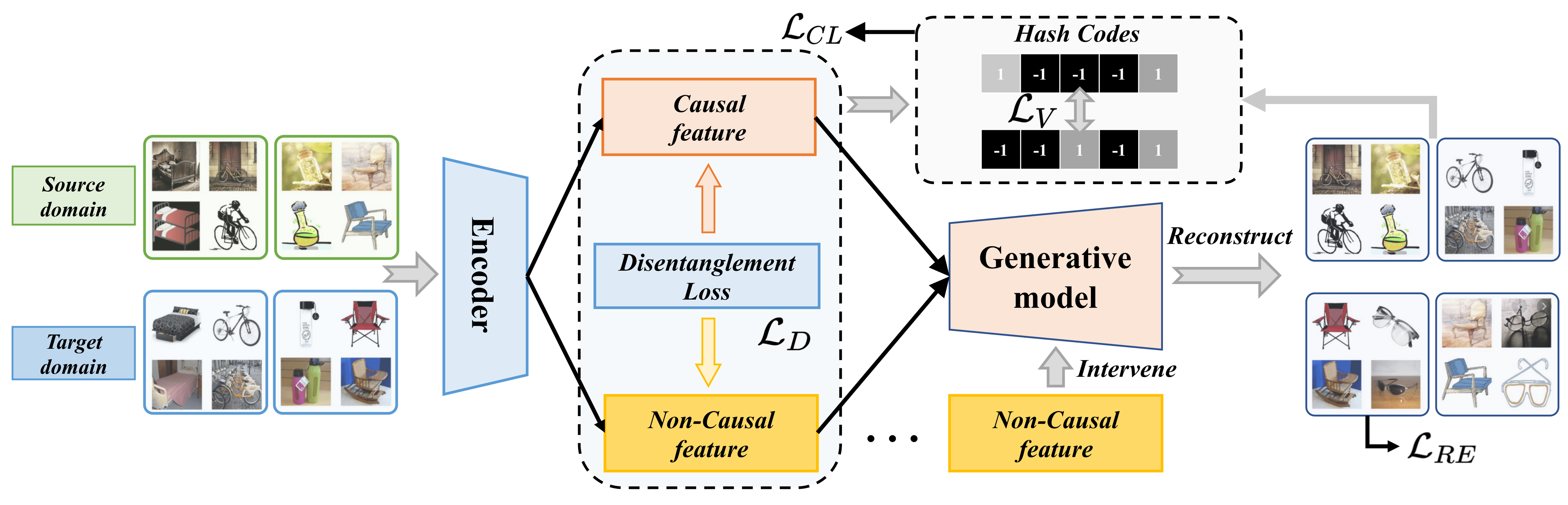

Figure 1. The framework of the proposed IDEA.

1. IDEA addresses unsupervised domain adaptive retrieval, separating images into causal and non-causal features using a Structural Causal Model (SCM) and information bottleneck principle to generate domain-invariant hash codes.

2. It employs consistency learning ($\mathcal{L}_{CL}$) and invariant learning under intervention ($\mathcal{L}_V$) to ensure the hash codes are discriminative and invariant to non-causal features.

3. The overall training objective combines causal feature disentanglement, consistency learning, image reconstruction, and invariance under intervention, formulated as:

$$\mathcal{L} = \mathcal{L}_D + \mathcal{L}_{CL} + \mathcal{L}_{RE} + \mathcal{L}_V \qquad (1)$$

Theoretical Analysis

We define:

$$\hat{I}(\boldsymbol{F}^n, \boldsymbol{Y}) = \mathbb{E}_{p(\boldsymbol{F}^n, \boldsymbol{Y})}\left[\log p(\boldsymbol{y} \mid \boldsymbol{f}^n)\right] - \mathbb{E}_{p(\boldsymbol{F}^n)}\mathbb{E}_{p(\boldsymbol{Y})}\left[\log p(\boldsymbol{y} \mid \boldsymbol{f}^n)\right] \qquad (2)$$

Then, we show that $\hat{I}(\boldsymbol{F}^n, \boldsymbol{Y})$ is an upper bound of $I(\boldsymbol{F}^n, \boldsymbol{Y})$. In formulation, we calculate their difference as follows:

$$\begin{aligned}
&\hat{I}(\boldsymbol{F}^n, \boldsymbol{Y}) - I(\boldsymbol{F}^n, \boldsymbol{Y}) \\
&= \mathbb{E}_{p(\boldsymbol{F}^n, \boldsymbol{Y})}\left[\log p(\boldsymbol{y} \mid \boldsymbol{f}^n)\right] - \mathbb{E}_{p(\boldsymbol{F}^n)}\mathbb{E}_{p(\boldsymbol{Y})}\left[\log p(\boldsymbol{y} \mid \boldsymbol{f}^n)\right] \\
&- \mathbb{E}_{p(\boldsymbol{F}^n, \boldsymbol{Y})}[\log p(\boldsymbol{y} \mid \boldsymbol{f}^n) - \log p(\boldsymbol{y})] \\
&= \mathbb{E}_{p(\boldsymbol{F}^n, \boldsymbol{Y})}[\log p(\boldsymbol{y})] - \mathbb{E}_{p(\boldsymbol{F}^n)}\mathbb{E}_{p(\boldsymbol{y})}[\log p(\boldsymbol{y} \mid \boldsymbol{f}^n)] \\
&= \mathbb{E}_{p(\boldsymbol{Y})}\left[\log p(\boldsymbol{y}) - \mathbb{E}_{p(\boldsymbol{F}^n)}[\log p(\boldsymbol{y} \mid \boldsymbol{f}^n)]\right] \\
&= \mathbb{E}_{p(\boldsymbol{Y})}\left[\log\left(\mathbb{E}_{p(\boldsymbol{F}^n)}[p(\boldsymbol{y} \mid \boldsymbol{f}^n)]\right) - \mathbb{E}_{p(\boldsymbol{F}^n)}[\log p(\boldsymbol{y} \mid \boldsymbol{f}^n)]\right] \\
&\geq 0 \text{ (Jensen's Inequality)},
\end{aligned} \qquad (3)$$

where the last inequality holds due to Jensen's Inequality with a convex function $\log(\cdot)$.

Results

Table 1. MAP performances on two bench-marking datasets with 64-bit hash codes.

	Office-Home						Office31						
Methods	P2R	C2R	R2A	R2P	R2C	A2R	A2D	A2W	W2D	D2A	W2A	D2W	Avg
Unsupervised Hashing Methods													
SH [62]	15.03	8.77	12.87	16.13	8.24	13.71	12.02	9.83	34.72	11.28	9.85	34.37	15.57
ITQ [17]	26.81	14.83	25.37	28.19	14.92	25.88	29.55	28.53	58.00	26.83	25.09	58.89	30.24
DSH [34]	8.49	5.47	9.67	8.26	5.28	9.69	16.66	15.09	39.24	16.33	13.58	41.07	15.74
LSH [16]	12.24	6.94	11.45	13.45	7.24	11.49	16.04	15.35	38.80	13.60	14.67	43.99	17.11
SGH [24]	24.51	13.62	22.53	25.73	13.51	22.93	24.98	22.47	53.94	22.17	20.52	56.36	26.94
OCH [35]	18.65	10.27	17.54	20.15	10.05	18.09	24.86	22.49	51.03	22.45	20.79	53.64	24.17
Transfer Hashing Methods													
ITQ+ [79]	17.61	9.55	14.25	-	-	-	17.99	15.00	42.29	-	-	-	19.45
LapITQ+ [79]	16.89	10.37	13.56	-	-	-	19.96	18.24	43.32	-	-	-	20.39
GTH-g [74]	20.00	10.99	18.28	21.95	11.68	19.05	23.08	21.20	49.38	19.52	17.41	50.14	23.56
DAPH [21]	27.20	15.29	27.35	28.19	15.29	26.37	32.80	28.66	60.71	28.66	27.59	64.11	31.85
PWCF [22]	34.03	24.22	28.95	34.44	18.42	34.57	39.78	34.86	67.94	35.12	35.01	72.91	38.35
DHLing [22]	48.47	30.81	38.68	45.24	25.15	43.30	41.96	45.10	75.23	42.89	41.74	79.91	46.54
PEACE [58]	53.04	38.72	42.68	54.39	28.36	45.97	46.69	48.89	78.82	46.91	46.95	83.18	51.22
Ours	**59.18**	**45.71**	**49.64**	**61.84**	**32.77**	**51.19**	**48.70**	**54.43**	**84.97**	**53.53**	**53.71**	**88.69**	**57.03**

Table 2. MAP performances on the Digits dataset with 64-bit hash codes.

	MNIST2USPS						USPS2MNIST						
Code Length	16	32	48	64	96	128	16	32	48	64	96	128	Avg
Unsupervised Hashing Methods													
SH [62]	15.56	13.67	13.80	13.45	13.35	12.95	15.59	14.35	14.22	13.57	12.92	12.96	13.87
ITQ [17]	13.05	15.57	18.54	20.12	23.12	23.89	13.69	17.51	20.40	20.30	22.79	24.59	19.46
DSH [34]	20.60	22.21	23.68	24.28	25.73	26.50	19.54	21.22	22.89	23.79	25.91	26.46	23.57
LSH [16]	12.40	13.54	15.89	16.01	18.54	20.44	12.76	14.86	14.77	16.89	16.32	19.67	16.01
SGH [24]	14.24	16.69	18.72	19.70	21.00	21.95	13.26	17.71	18.22	19.01	21.69	22.09	18.69
OCH [35]	13.73	17.22	19.59	20.18	20.66	23.34	15.51	17.75	18.97	21.50	21.27	23.68	19.45
Transfer Hashing Methods													
ITQ+ [79]	22.84	21.20	20.68	19.15	17.99	18.52	-	-	-	-	-	-	20.06
LapITQ+ [79]	24.26	24.03	23.76	24.59	23.33	22.73	-	-	-	-	-	-	23.78
GTH-g [74]	20.45	17.64	16.60	17.25	17.26	17.06	15.17	14.07	15.02	15.01	14.80	17.34	16.47
DAPH [21]	25.13	27.10	26.10	28.51	30.53	30.70	26.60	26.43	27.27	27.99	30.19	31.40	28.16
PWCF [22]	47.47	51.99	51.44	51.75	50.89	59.35	47.14	50.86	52.06	52.18	57.14	58.96	52.60
DHLing [22]	49.24	54.90	56.30	58.28	58.80	59.14	50.14	51.35	53.67	58.65	58.42	59.17	55.67
PEACE [58]	52.87	59.72	60.69	62.84	65.13	68.16	53.97	54.82	58.69	60.91	62.65	65.70	60.51
Ours	**58.89**	**64.48**	**65.72**	**67.48**	**70.24**	**74.34**	**60.99**	**61.47**	**65.45**	**67.97**	**69.72**	**72.31**	**66.59**

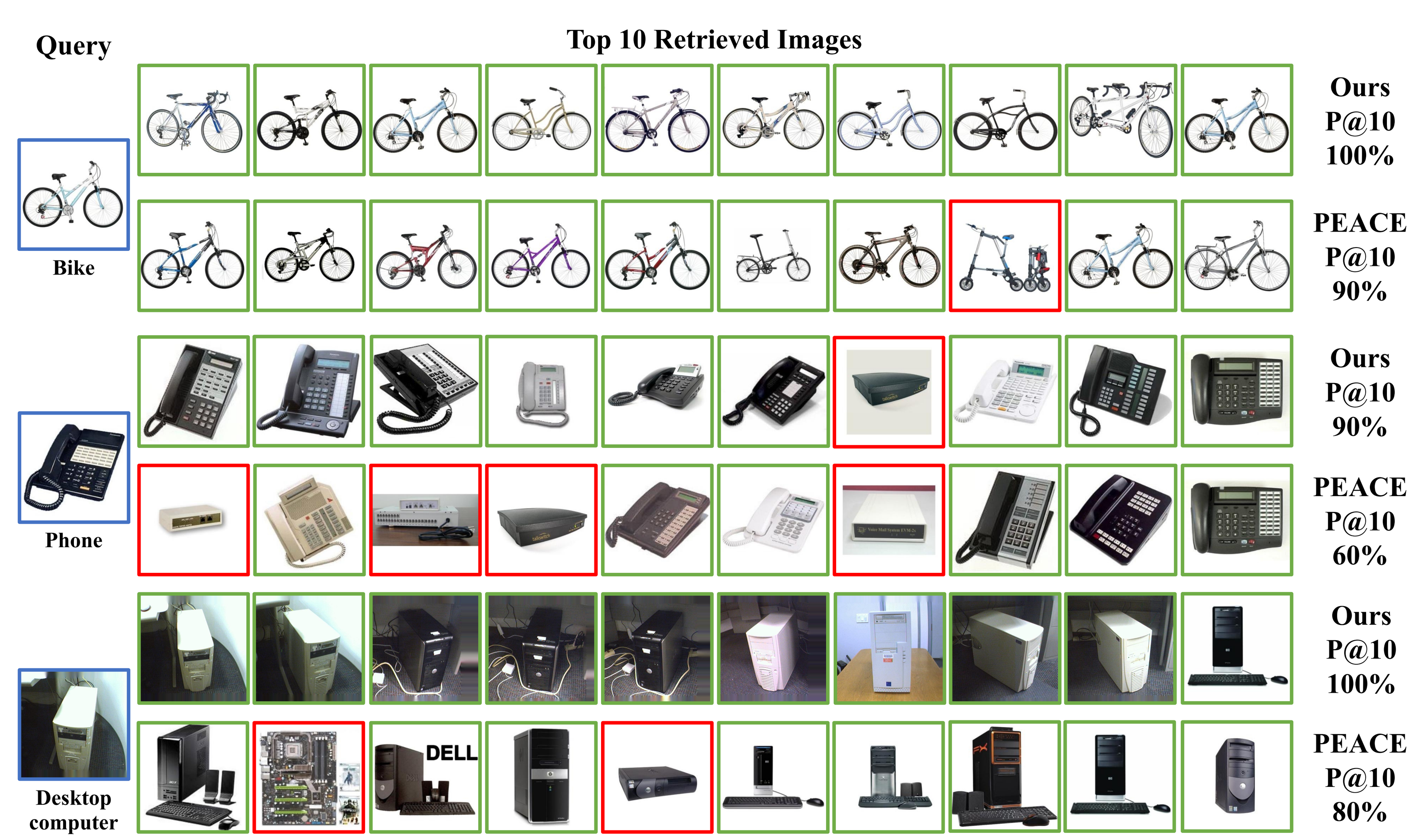

Figure 2. Top 10 Images and Precision@10 Examples on the Office-31 Dataset