Domain-Specificity-inducing Transformers for Source-Free Domain Adaptation

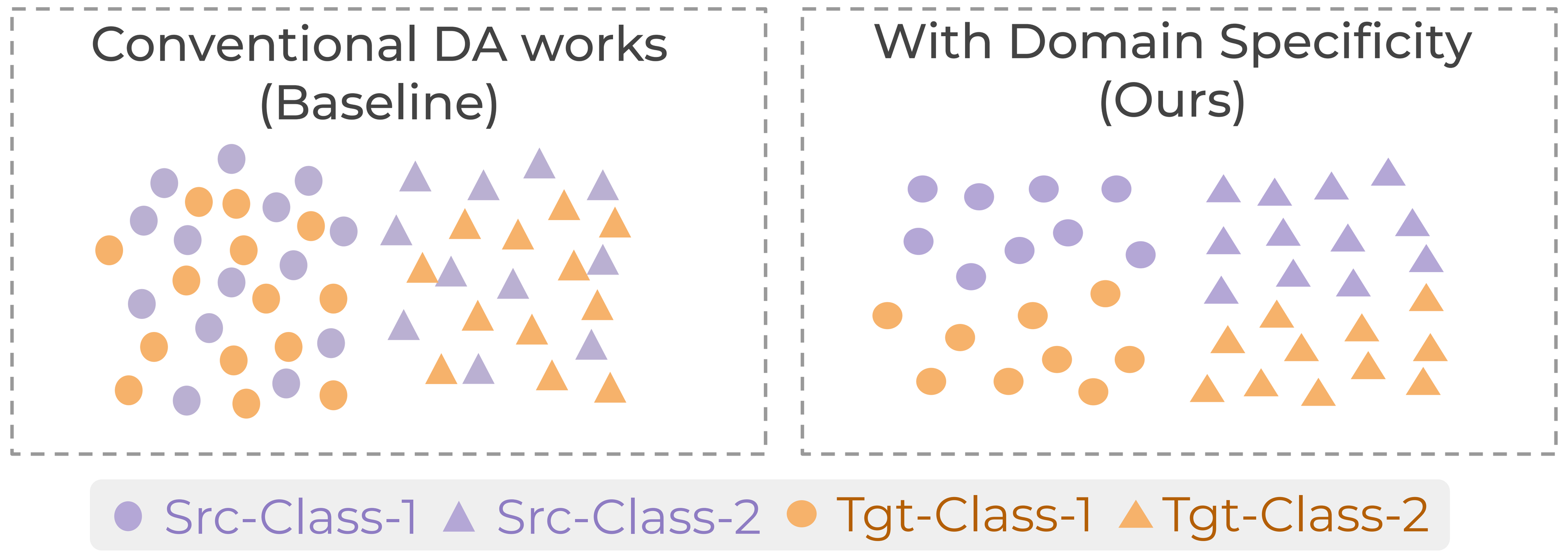
Sunandini Sanyal*, Ashish Ramayee Asokan*, Suvaansh Bhambri*,
Akshay Kulkarni, Jogendra Nath Kundu, R. Venkatesh Babu
Vision and AI Lab, Indian Institute of Science, Bangalore, India

Vision & AI Lab (VAL)

ICCV23 PARIS

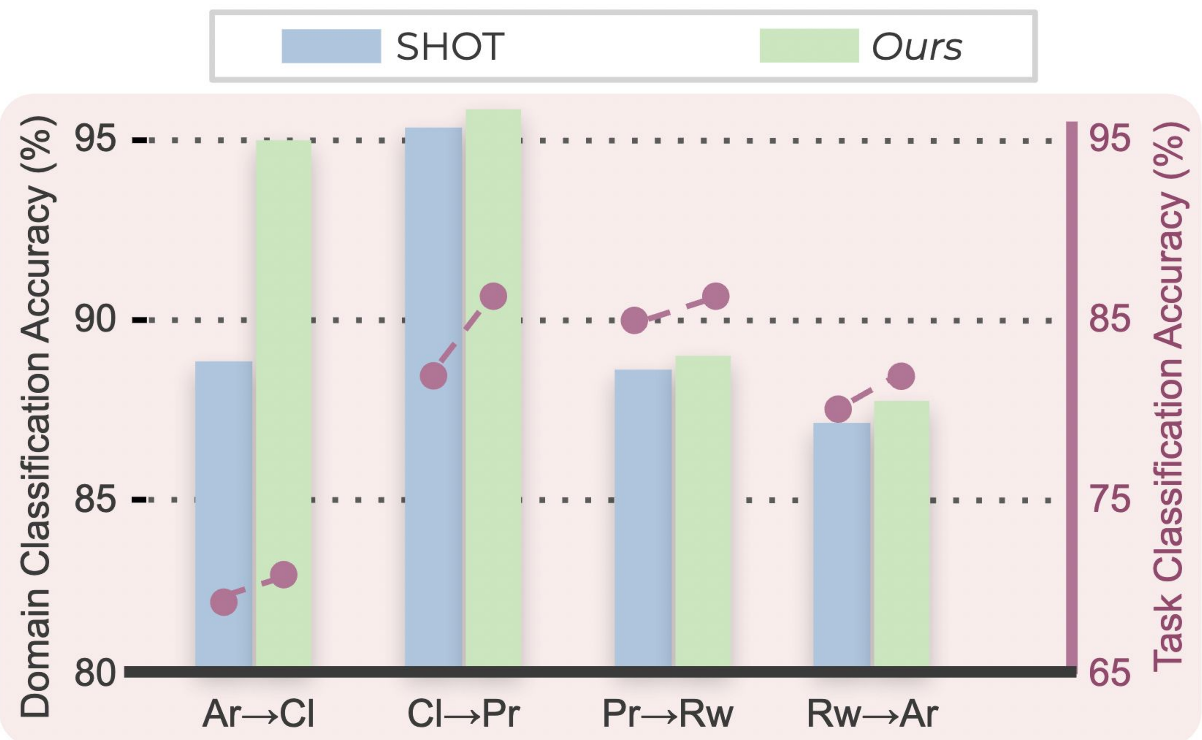
Introduction

- DA requires learning of two types of features:
  - Task-specific (TS) factors → goal task-oriented features that generalize across domains
  - Domain-specific (DS) factors → crucial in-domain characteristic knowledge
- Conventional DA methods focus on domain-invariant (DI) learning by aligning TS factors only
  - Limitations: Does not guarantee optimal performance as distance b/w the DI model and the support of a DS model is large.
- We motivate disentanglement and learning of both DS and TS factors for improved SFDA.



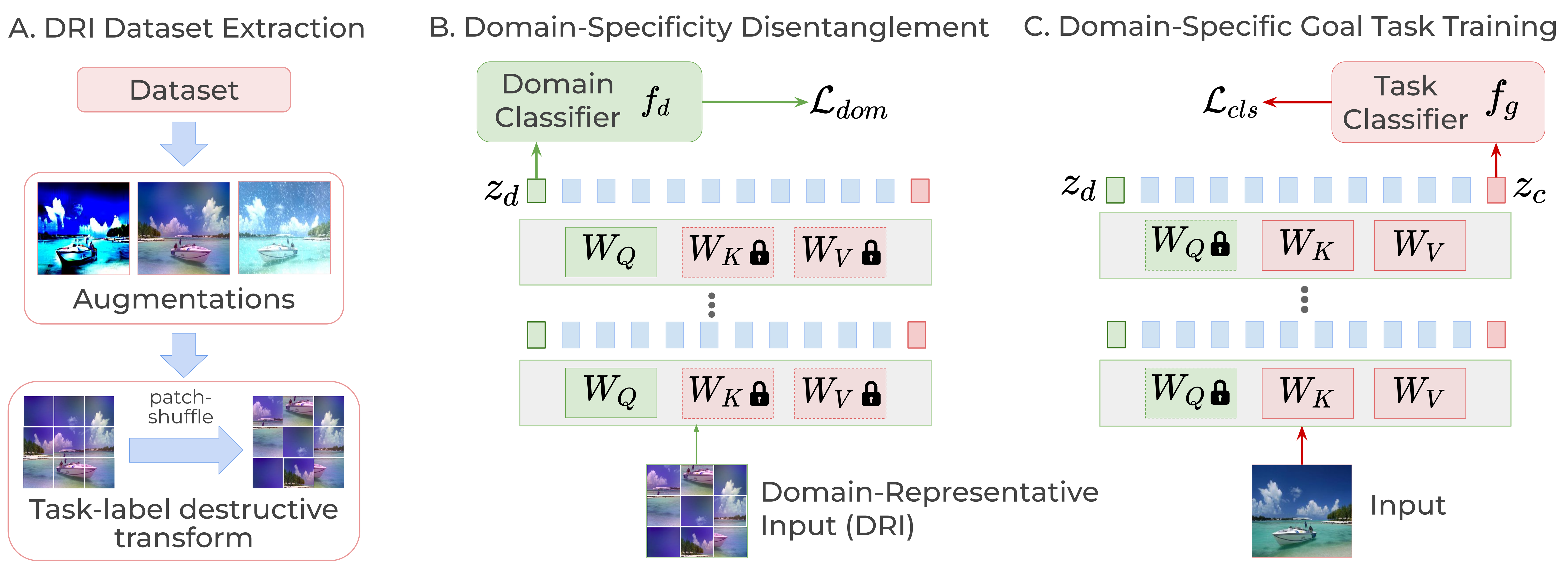
Key Insights

- Insight 1 → Domain-specificity leads to improved DA
  - An in-domain-trained model better represents domain-specific characteristics
- Insight 2 → Disentanglement of DS and TS factors enables better control over them



We demonstrate better source-target separation (blue, green) and better task-specificity (pink)

Approach : Domain-Specificity-inducing Transformer (DSiT)



Vendor side training

- Goal task training → Train W_K, W_V, z_c, and f_g using $\mathcal{L}_{cls}$ on the source domain

$$\min_{\theta_h \setminus \theta_Q, \theta_{f_g}} \mathbb{E}_{(x_s, y_s) \in \mathcal{D}_s} [\mathcal{L}_{cls}] \text{ where } \mathcal{L}_{cls} = \mathcal{L}_{ce}(f_g(z_c), y_c)$$

- Domain-specificity Disentanglement → Train W_Q, z_d, and f_d using $\mathcal{L}_{dom}$ on the source domain

$$\min_{\theta_Q, \theta_{f_d}} \mathbb{E}_{(x, y_d) \in \cup_i \mathcal{D}_s^{(i)}} [\mathcal{L}_{dom}] \text{ where } \mathcal{L}_{dom} = \mathcal{L}_{ce}(f_d(z_d), y_d)$$

Client side training

- Goal task training → Train W_K, W_V, z_c, and f_g using Information Maximization on the target domain

$$\min_{\theta_h \setminus \theta_Q, f_g} \mathbb{E}_{\mathcal{D}_t} [\mathcal{L}_{ent} + \mathcal{L}_{div}]$$

- Domain-specificity Disentanglement → Train W_Q, z_d, and f_d using $\mathcal{L}_{dom}$ on the target domain

$$\min_{\theta_Q, \theta_{f_d}} \mathbb{E}_{(x, y_d) \in \cup_i \mathcal{D}_s^{(i)}} [\mathcal{L}_{dom}] \text{ where } \mathcal{L}_{dom} = \mathcal{L}_{ce}(f_d(z_d), y_d)$$

Results on Single-source DA

Method	SF	OH	O-31	VisDA
TVT (WACV'23)	✗	83.5	93.8	83.9
SSRT-B (CVPR'22)	✗	85.4	93.5	88.7
CDTrans (ICLR'22)	✗	80.5	92.6	88.4
SHOT-B (ICML'20)	✓	78.1	90.5	82.8
DIPE (CVPR'22)	✓	78.2	91.7	86.3
Mixup (ICML'22)	✓	78.5	91.4	85.9
DSiT *(Ours)*	✓	**80.5**	**93.0**	**87.6**

Results on Multi-source DA (OH)

Method	SF	Avg.
Source-combine	✗	66.9
SImpAl (NeurIPS'20)	✗	72.2
SHOT (ICML'20)	✓	74.3
SHOT++ (TPAMI'21)	✓	75.7
CAiDA (NeurIPS'21)	✓	76.2
SHOT-B*	✓	83.7
DSiT-B (Ours)	✓	**84.7**

Results on Multi-target DA (OH)

Method	SF	Avg.
MT-MTDA (WACV'21)	✗	64.3
CDAN+DCL (NeurIPS'17)	✗	64.1
D-CGCT (CVPR'21)	✗	69.8
D-CGCT-B (CVPR'21)	✗	78.6
SHOT-B (ICML'20)	✓	76.4
DSiT *(Ours)*	✓	**78.3**

We demonstrate gains over SOTA methods across single-source, multi-source, and multi-target DA

Acknowledgements: This work was supported by a research grant from the Kotak IISc AI-ML Centre (KIAC)