

DNA: Domain Generalization with Diversified Neural Averaging

Xu Chu, Yujie Jin, Wenwu Zhu, Yasha Wang, Xin Wang, Shanghang Zhang, Hong Mei

ICML | 2022

Thirty-ninth International
Conference on Machine Learning

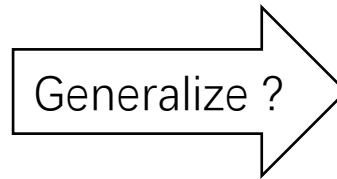


Background

- **Goal of Domain Generalization (DG)**
 - learn classifiers that generalize on the unseen target domain



Training distribution (source domain) \mathcal{T}



Testing distribution (target domain) \mathcal{S}

- **Motivation**

- empirical risk minimization is error-prone to domain shift
- domain shift: testing distribution $\mathcal{T} \neq$ training distribution \mathcal{S}
 - E.g, various weather conditions for training and testing in autonomous driving tasks

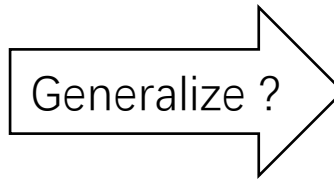
Background

- **Goal of Domain Generalization (DG)**

- learn classifiers that generalize on the unseen target domain by training on the source domain



Training distribution (source domain) \mathcal{T}



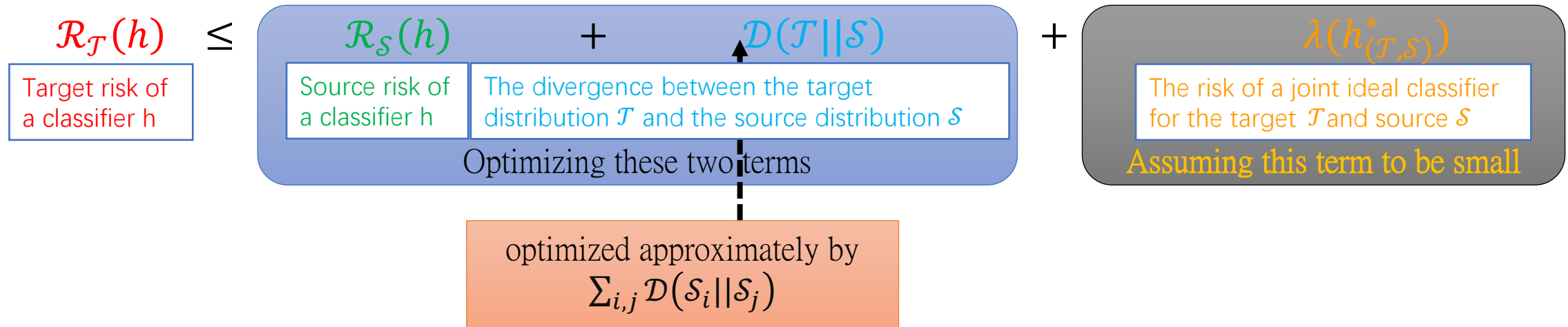
Testing distribution (target domain) \mathcal{S}

- **Motivation**

- empirical risk minimization is error-prone to domain shift
- domain shift: testing distribution $\mathcal{T} \neq$ training distribution \mathcal{S}
 - E.g, various weather conditions for training and testing in autonomous driving tasks

Limitation of Three-term-trade-off

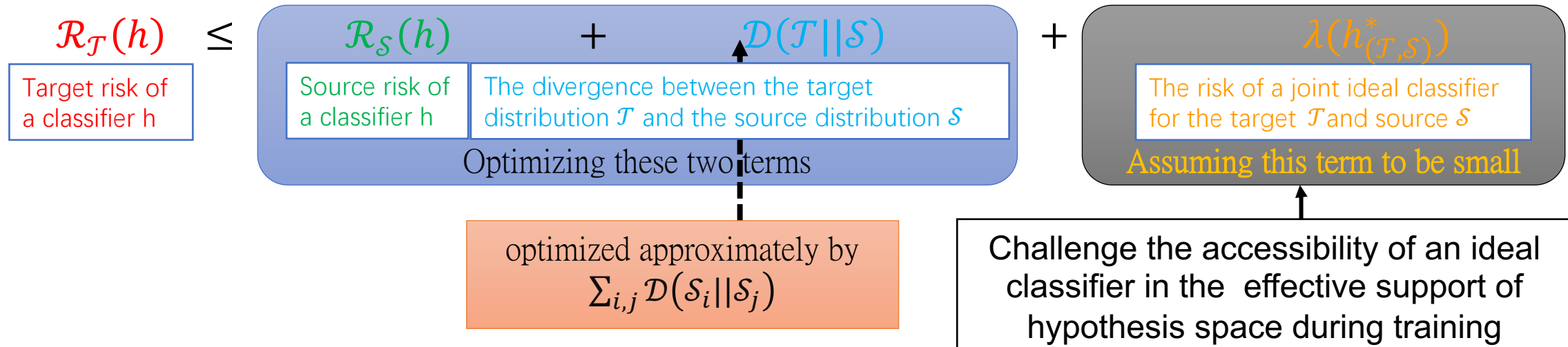
- A pervasive theme of target risk decomposition



- Memorization on the seen training domain
 - only seen source-discriminative features are memorized,
 - target-discriminative features could be ignored

Limitation of Three-term-trade-off

- A pervasive theme of target risk decomposition



- **Memorization on the seen training domain**

- only seen source-discriminative features are memorized,
- target-discriminative features could be ignored

Implying a bad balance of the three-term-trade-off

Our solution

- **bridging ensemble and DG for easier target risk control**
 - Ensemble: Enlarging the effective support of the hypothesis space without the cost of introducing more parameters
- **contribution summary**
 - Propose a novel pruned Jensen-Shannon divergence (PJS) loss, which connects the generalization risk and a pruned Jensen-Shannon divergence
 - Propose upper bounds on the target risk of classifier ensembles
 - Propose a practical diversified neural averaging (DNA) algorithm optimizing the PAC-Bayesian bound

Our solution

- bridging ensemble and DG for easier target risk control
 - Ensemble: Enlarging the effective support of the hypothesis space without the cost of introducing more parameters
- **contribution summary**
 - Propose a novel pruned Jensen-Shannon divergence (PJS) loss, which connects the generalization risk and a pruned Jensen-Shannon divergence

Definition 1 (pruned Jensen-Shannon loss)

Let \mathbb{A} be the support of the underlying (empirical) distribution where the realization (\mathbf{x}, \mathbf{y}) is drawn from. The pruned Jensen-Shannon (PJS) loss $\ell_{PJS} : \mathbb{X} \times \mathbb{I} \times \mathbb{H} \rightarrow \mathbb{R}^+$ is, $\ell_{PJS}(\mathbf{x}, \mathbf{y}, \mathbf{h}) = (\log \frac{2}{h^*+1} + h^* \log \frac{2h^*}{h^*+1}) 1_{\{(\mathbf{x}, \mathbf{y}) \in \mathbb{A}\}}$, where $h^* \in (0, 1)$ is the vector component of \mathbf{h} that corresponds to the correct class, and $1_{\{\cdot\}}$ is the indicator function.

Definition 2 (pruned Jensen-Shannon divergence).

The pruned Jensen-Shannon (PJS) divergence $D_{PJS}(\mathcal{P} \parallel \Upsilon) : M_+^\lambda \times M_+^\lambda \rightarrow \mathbb{R}$ from a measure Υ to a measure \mathcal{P} is defined by an integral over the support set $\mathbb{A}_{\mathcal{P}}$ of measure \mathcal{P} , that is $D_{PJS}(\mathcal{P} \parallel \Upsilon) = \int_{\mathbb{A}_{\mathcal{P}}} p \log \frac{2p}{p+v} + v \log \frac{2v}{p+v} d\lambda$.

Proposition 1 (connecting risk and PJS divergence)

Provided a domain $\mathcal{P} \in M_+(\mathbb{X} \times \mathbb{I})$, assume the existence of the density function over the input space \mathbb{X} , denoted by $p^{\mathcal{P}}(\mathbf{x})$. Then $p^{\mathcal{P}}(\mathbf{x})\mathbf{h}$ induces a (set of) measure(s) $\Upsilon^{\mathcal{P}} \in M_+(\mathbb{X} \times \mathbb{Y})$ whose density over $\mathbb{X} \times \mathbb{Y}$ is $p^{\mathcal{P}}(\mathbf{x})\mathbf{h}$. For a classifier \mathbf{h} and a domain $\mathcal{P} \in M_+(\mathbb{X} \times \mathbb{I})$. Let $\Upsilon^{\mathcal{P}}$ be the induced measure. Assuming $p^{\mathcal{P}}(\mathbf{y}|\mathbf{x}) \in \{0, 1\}$, then

$$R_{\mathcal{P}}^{\ell_{PJS}}(\mathbf{h}) = \mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim \mathcal{P}} [\ell_{PJS}(\mathbf{x}, \mathbf{y}, \mathbf{h})] = D_{PJS}(\mathcal{P} \parallel \Upsilon^{\mathcal{P}}). \quad (1)$$

Our solution

- **contribution summary**

- Propose a novel pruned Jensen-Shanno the generalization risk and a pruned Jer

- Propose upper bounds on the (target) risk of classifier ensembles

Theorem 1 (inequalities related to the ρ -ensemble)

For any $\rho \in M_+(\mathbb{H})$, any measure \mathcal{P} , take $\mathbb{D}_{\mathcal{P}} = \mathbb{E}_{\mathcal{P}}[\text{Var}_{\rho}(\sqrt{\ell_{PJS}})]$, then there is

$$\sqrt{R_{\mathcal{P}}^{\ell_{PJS}}(\rho)} \leq \sqrt{\mathbb{E}_{\rho}[R_{\mathcal{P}}^{\ell_{PJS}}(\mathbf{h})] - \mathbb{D}_{\mathcal{P}}} \leq \mathbb{E}_{\rho}[\sqrt{R_{\mathcal{P}}^{\ell_{PJS}}(\mathbf{h})}]. \quad (3)$$

Moreover, $\mathbb{D}_{\mathcal{P}} > 0$ is a necessary condition for the second inequality becoming strict. If $\frac{\ell_{PJS}(\mathbf{x}_1, \mathbf{y}, \mathbf{h})}{\ell_{PJS}(\mathbf{x}_2, \mathbf{y}, \mathbf{h})}$ is varying on (a non-zero measured subset of) the set $\{(\mathbf{x}, \mathbf{y}, \mathbf{h}) : \text{Var}_{\rho}(\sqrt{\ell_{PJS}}) > 0\}$ for $\mathbf{x}_1 \neq \mathbf{x}_2$, then $\mathbb{D}_{\mathcal{P}} > 0$ is a sufficient condition for the second inequality being strict.

Theorem 2 (PAC-Bayesian generalization upper bound)

For a fixed source domain \mathcal{P} and a fixed target domain \mathcal{Q} . Suppose that π is a prior over \mathbb{H} , which is independent of draws of source realizations $D^n = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^n \stackrel{i.i.d.}{\sim} \mathcal{P}^n$. Then for any $c > 0$, $\rho \in M_+(\mathbb{H})$, and any $\delta \in (0, 1)$, with probability over $1 - \delta$

$$\sqrt{R_{\mathcal{Q}}^{\ell_{PJS}}(\rho)} \leq 2\sqrt{2D_{JS}(\mathcal{P} \parallel \mathcal{Q})} + \sqrt{\mathbb{E}_{\rho}[\hat{R}_{\mathcal{P}}^{\ell_{PJS}}(\mathbf{h})] - \hat{\mathbb{D}}_{\mathcal{P}} + \frac{2D_{KL}(\rho \parallel \pi) + \log \frac{1}{\delta} + \Psi_{\mathcal{P}, \pi}^{\ell_{PJS}}(c, n)}{cn}}, \quad (6)$$

where $\log \mathbb{E}_{\pi^2} \mathbb{E}_{\mathcal{P}^n} [e^{cn \mathbb{E}_{\mathcal{P}} \sqrt{\ell(\mathbf{h}') \mathbb{E}_{\mathcal{P}} \sqrt{\ell(\mathbf{h})} - \hat{\mathbb{D}}_{\mathcal{P}} \sqrt{\ell(\mathbf{h}') \hat{\mathbb{D}}_{\mathcal{P}} \sqrt{\ell(\mathbf{h})}}}] = \Psi_{\mathcal{P}, \pi}^{\ell}(c, n)$ is constant w.r.t. ρ for fixed c, n, π, ℓ , and δ .

Corollary 1 (target risk upper bound of the ensembles)

Given a fixed source domain \mathcal{P} and a target domain \mathcal{Q} , for any measure $\rho \in M_+(\mathbb{H})$ on hypothesis space \mathbb{H} ,

$$\text{sqrt } R_{\mathcal{Q}}^{\ell_{PJS}}(\rho) \leq \sqrt{\mathbb{E}_{\rho}[R_{\mathcal{P}}^{\ell_{PJS}}(\mathbf{h})] - \mathbb{D}_{\mathcal{P}}} + 2\sqrt{2D_{JS}(\mathcal{P} \parallel \mathcal{Q})}. \quad (4)$$

Corollary 2 (joint risk upper bound of the ensembles)

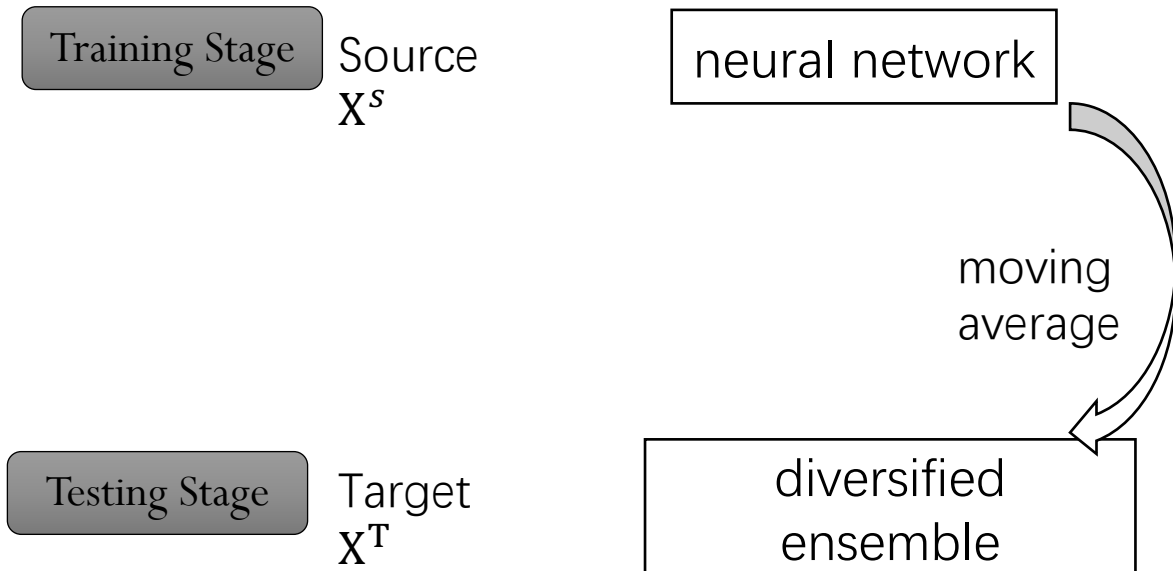
Given a fixed source domain \mathcal{P} and a target domain \mathcal{Q} , for any measure $\rho \in M_+(\mathbb{H})$ on the hypothesis space \mathbb{H} ,

$$\begin{aligned} \sqrt{R_{\mathcal{P}}^{\ell_{PJS}}(\rho)} + \sqrt{R_{\mathcal{Q}}^{\ell_{PJS}}(\rho)} &\leq \sqrt{\mathbb{E}_{\rho}[R_{\mathcal{P}}^{\ell_{PJS}}(\mathbf{h})] - \mathbb{D}_{\mathcal{P}}} + \sqrt{\mathbb{E}_{\rho}[R_{\mathcal{Q}}^{\ell_{PJS}}(\mathbf{h})] - \mathbb{D}_{\mathcal{Q}}} \\ &\leq \mathbb{E}_{\rho}[\sqrt{R_{\mathcal{P}}^{\ell_{PJS}}(\mathbf{h})}] + \mathbb{E}_{\rho}[\sqrt{R_{\mathcal{Q}}^{\ell_{PJS}}(\mathbf{h})}]. \end{aligned} \quad (5)$$

Our solution

• contribution summary

- Propose a novel pruned Jensen-Shar the generalization risk and a pruned
- Propose upper bounds on the (target
- Propose a practical diversified neural averaging (DNA) algorithm optimizing the PAC-Bayesian bound

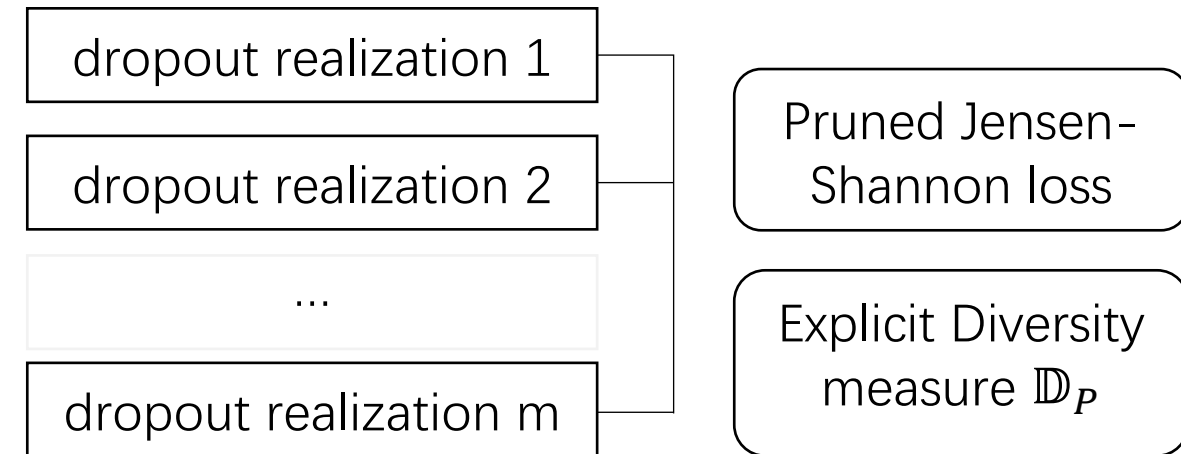


Theorem 2 (PAC-Bayesian generalization upper bound)

For a fixed source domain \mathcal{P} and a fixed target domain \mathcal{Q} . Suppose that π is a prior over \mathbb{H} , which is independent of draws of source realizations $D^n = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^n \stackrel{i.i.d.}{\sim} \mathcal{P}^n$. Then for any $c > 0$, $\rho \in M_+(\mathbb{H})$, and any $\delta \in (0, 1)$, with probability over $1 - \delta$

$$\sqrt{R_{\mathcal{Q}}^{\ell_{PJS}}(\rho)} \leq 2\sqrt{2D_{JS}(\mathcal{P} \parallel \mathcal{Q})} + \sqrt{\mathbb{E}_{\rho}[\hat{R}_{\mathcal{P}}^{\ell_{PJS}}(\mathbf{h})] - \hat{\mathbb{D}}_{\mathcal{P}} + \frac{2D_{KL}(\rho \parallel \pi) + \log \frac{1}{\delta} + \Psi_{\mathcal{P}, \pi}^{\ell_{PJS}}(c, n)}{cn}}, \quad (6)$$

where $\log \mathbb{E}_{\pi^2} \mathbb{E}_{\mathcal{P}^n} [e^{cn \mathbb{E}_{\mathcal{P}} \sqrt{\ell(\mathbf{h}')} \mathbb{E}_{\mathcal{P}} \sqrt{\ell(\mathbf{h})} - \hat{\mathbb{D}}_{\mathcal{P}} \sqrt{\ell(\mathbf{h}')} \hat{\mathbb{D}}_{\mathcal{P}} \sqrt{\ell(\mathbf{h})}}] = \Psi_{\mathcal{P}, \pi}^{\ell}(c, n)$ is constant w.r.t. ρ for fixed c, n, π, ℓ , and δ .



Prediction

Experimental evaluation

Method	PACS	VLCS	OfficeHome	TerraIncognita	DomainNet	Avg
ERM (Vapnik, 1999)	85.5 ± 0.2	77.5 ± 0.4	66.5 ± 0.3	46.1 ± 1.8	40.9 ± 0.1	63.3
IRM (Arjovsky et al., 2019)	83.5 ± 0.8	78.5 ± 0.5	64.3 ± 2.2	47.6 ± 0.8	33.9 ± 2.8	61.6
DRO (Sagawa et al., 2020)	84.4 ± 0.8	76.7 ± 0.6	66.0 ± 0.7	43.2 ± 1.1	33.3 ± 0.2	60.7
MMD (Li et al., 2018c)	84.6 ± 0.5	77.5 ± 0.9	66.3 ± 0.1	42.2 ± 1.6	23.4 ± 9.5	58.8
DANN (Ganin et al., 2016)	83.6 ± 0.4	78.6 ± 0.4	65.9 ± 0.6	46.7 ± 0.5	38.3 ± 0.1	62.6
CDANN (Li et al., 2018d)	82.6 ± 0.9	77.5 ± 0.1	65.8 ± 1.3	45.8 ± 1.6	38.3 ± 0.3	62.0
MTL (Blanchard et al., 2021)	84.6 ± 0.5	77.2 ± 0.4	66.4 ± 0.5	45.6 ± 1.2	40.6 ± 0.1	62.9
ARM (Zhang et al., 2020)	85.1 ± 0.4	77.6 ± 0.3	64.8 ± 0.3	45.5 ± 0.3	35.5 ± 0.2	61.7
VREx (Krueger et al., 2021)	84.9 ± 0.6	78.3 ± 0.2	66.4 ± 0.6	46.4 ± 0.6	33.6 ± 2.9	61.9
RSC (Huang et al., 2020)	85.2 ± 0.9	77.1 ± 0.5	65.5 ± 0.9	46.6 ± 1.0	38.9 ± 0.5	62.7
Mixup (Wang et al., 2020b)	84.6 ± 0.6	77.4 ± 0.6	68.1 ± 0.3	47.9 ± 0.8	39.2 ± 0.1	63.4
MLDG (Li et al., 2018a)	84.9 ± 1.0	77.2 ± 0.4	66.8 ± 0.6	47.7 ± 0.9	41.2 ± 0.1	63.6
SagNet (Nam et al., 2021)	86.3 ± 0.2	77.8 ± 0.5	68.1 ± 0.1	48.6 ± 1.0	40.3 ± 0.1	64.2
CORAL (Sun & Saenko, 2016)	86.2 ± 0.3	78.8 ± 0.6	68.7 ± 0.3	47.6 ± 1.0	41.5 ± 0.1	64.5
mDSDI (Bui et al., 2021)	86.2 ± 0.2	79.0 ± 0.3	69.2 ± 0.4	48.1 ± 1.4	42.8 ± 0.1	65.1
SWAD (Cha et al., 2021)	88.1 ± 0.4	79.1 ± 0.4	70.6 ± 0.3	50.0 ± 0.4	46.5 ± 0.2	66.9
DNA (ours)	88.4 ± 0.1	79.0 ± 0.1	71.2 ± 0.1	52.2 ± 0.4	47.2 ± 0.1	67.6

- With similar computational cost, the proposed DNA achieves competitive performance
- With lower computational cost, the proposed DNA achieves competitive performance compared with a common ensemble strategy
- The training process of the DNA (b) is more stable compared to the other variants, indicating the ease of model selection

Experimental evaluation

Method	PACS	VLCS	OfficeHome	TerraIncognita	DomainNet	Avg
ERM (Vapnik, 1999)	85.5 \pm 0.2	77.5 \pm 0.4	66.5 \pm 0.3	46.1 \pm 1.8	40.9 \pm 0.1	63.3
IRM (Arjovsky et al., 2019)	83.5 \pm 0.8	78.5 \pm 0.5	64.3 \pm 2.2	47.6 \pm 0.8	33.9 \pm 2.8	61.6
DRO (Sagawa et al., 2020)	84.4 \pm 0.8	76.7 \pm 0.6	66.0 \pm 0.7	43.2 \pm 1.1	33.3 \pm 0.2	60.7
MMD (Li et al., 2018c)	84.6 \pm 0.5	77.5 \pm 0.9	66.3 \pm 0.1	42.2 \pm 1.6	23.4 \pm 9.5	58.8
DANN (Ganin et al., 2016)	83.6 \pm 0.4	78.6 \pm 0.4	65.9 \pm 0.6	46.7 \pm 0.5	38.3 \pm 0.1	62.6
CDANN (Li et al., 2018d)	82.6 \pm 0.9	77.5 \pm 0.1	65.8 \pm 1.3	45.8 \pm 1.6	38.3 \pm 0.3	62.0
MTL (Blanchard et al., 2021)	84.6 \pm 0.5	77.2 \pm 0.4	66.4 \pm 0.5	45.6 \pm 1.2	40.6 \pm 0.1	62.9
ARM (Zhang et al., 2020)	85.1 \pm 0.4	77.6 \pm 0.3	64.8 \pm 0.3	45.5 \pm 0.3	35.5 \pm 0.2	61.7
VREx (Krueger et al., 2021)	84.9 \pm 0.6	78.3 \pm 0.2	66.4 \pm 0.6	46.4 \pm 0.6	33.6 \pm 2.9	61.9
RSC (Huang et al., 2020)	85.2 \pm 0.9	77.1 \pm 0.5	65.5 \pm 0.9	46.6 \pm 1.0	38.9 \pm 0.5	62.7
Mixup (Wang et al., 2020b)	84.6 \pm 0.6	77.4 \pm 0.6	68.1 \pm 0.3	47.9 \pm 0.8	39.2 \pm 0.1	63.4
MLDG (Li et al., 2018a)	84.9 \pm 1.0	77.2 \pm 0.4	66.8 \pm 0.6	47.7 \pm 0.9	41.2 \pm 0.1	63.6
SagNet (Nam et al., 2021)	86.3 \pm 0.2	77.8 \pm 0.5	68.1 \pm 0.1	48.6 \pm 1.0	40.3 \pm 0.1	64.2
CORAL (Sun & Saenko, 2016)	86.2 \pm 0.3	78.8 \pm 0.6	68.7 \pm 0.3	47.6 \pm 1.0	41.5 \pm 0.1	64.5
mDSDI (Bui et al., 2021)	86.2 \pm 0.2	79.0 \pm 0.3	69.2 \pm 0.4	48.1 \pm 1.4	42.8 \pm 0.1	65.1
SWAD (Cha et al., 2021)	88.1 \pm 0.4	79.1 \pm 0.4	70.6 \pm 0.3	50.0 \pm 0.4	46.5 \pm 0.2	66.9
DNA (ours)	88.4 \pm 0.1	79.0 \pm 0.1	71.2 \pm 0.1	52.2 \pm 0.4	47.2 \pm 0.1	67.6

- With similar computational cost, the proposed DNA achieves competitive performance
- With lower computational cost, the proposed DNA achieves competitive performance compared with the common ensemble strategy
- The training process of the DNA (b) is more stable compared to the other variants, indicating the ease of model selection

	OfficeHome					TerraIncognita				
	A	C	P	R	Avg	L100	L38	L43	L46	Avg
K=5	64.6 \pm 0.3	56.1 \pm 0.4	77.8 \pm 0.2	79.5 \pm 0.3	69.5	53.0 \pm 1.4	42.0 \pm 0.2	58.2 \pm 0.5	42.6 \pm 1.6	49.0
K=10	65.9 \pm 0.3	56.6 \pm 0.1	78.5 \pm 0.2	80.0 \pm 0.1	70.3	55.9 \pm 1.3	43.9 \pm 1.2	60.0 \pm 0.4	42.8 \pm 1.0	50.6
K=20	66.4 \pm 0.1	57.0 \pm 0.2	78.8 \pm 0.1	80.5 \pm 0.1	70.7	57.1 \pm 1.2	43.7 \pm 0.9	60.0 \pm 0.4	43.1 \pm 1.0	51.0
DNA	67.7 \pm 0.2	57.7 \pm 0.3	78.9 \pm 0.2	80.5 \pm 0.2	71.2	56.8 \pm 1.2	47.0 \pm 0.9	61.0 \pm 0.5	44.0 \pm 1.0	52.2

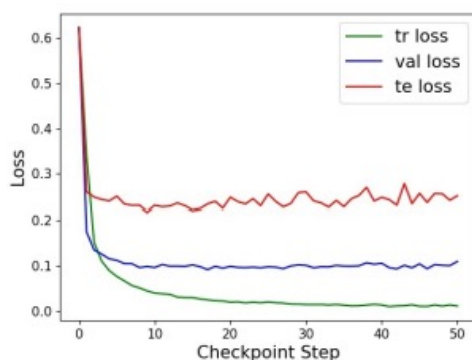
Thanks!

chu_xu@tsinghua.edu.cn

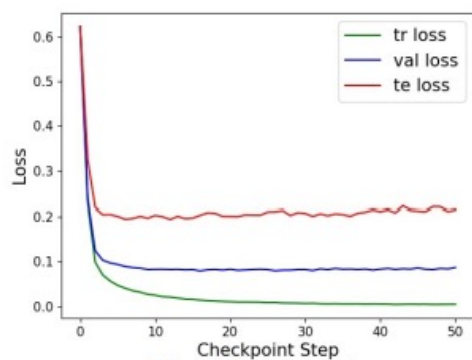
Experimental evaluation

Method	PACS	VLCS	OfficeHome	TerraIncognita	DomainNet	Avg
ERM (Vapnik, 1999)	85.5 \pm 0.2	77.5 \pm 0.4	66.5 \pm 0.3	46.1 \pm 1.8	40.9 \pm 0.1	63.3
IRM (Arjovsky et al., 2019)	83.5 \pm 0.8	78.5 \pm 0.5	64.3 \pm 2.2	47.6 \pm 0.8	33.9 \pm 2.8	61.6
DRO (Sagawa et al., 2020)	84.4 \pm 0.8	76.7 \pm 0.6	66.0 \pm 0.7	43.2 \pm 1.1	33.3 \pm 0.2	60.7
MMD (Li et al., 2018c)	84.6 \pm 0.5	77.5 \pm 0.9	66.3 \pm 0.1	42.2 \pm 1.6	23.4 \pm 9.5	58.8
DANN (Ganin et al., 2016)	83.6 \pm 0.4	78.6 \pm 0.4	65.9 \pm 0.6	46.7 \pm 0.5	38.3 \pm 0.1	62.6
CDANN (Li et al., 2018d)	82.6 \pm 0.9	77.5 \pm 0.1	65.8 \pm 1.3	45.8 \pm 1.6	38.3 \pm 0.3	62.0
MTL (Blanchard et al., 2021)	84.6 \pm 0.5	77.2 \pm 0.4	66.4 \pm 0.5	45.6 \pm 1.2	40.6 \pm 0.1	62.9
ARM (Zhang et al., 2020)	85.1 \pm 0.4	77.6 \pm 0.3	64.8 \pm 0.3	45.5 \pm 0.3	35.5 \pm 0.2	61.7
VREx (Krueger et al., 2021)	84.9 \pm 0.6	78.3 \pm 0.2	66.4 \pm 0.6	46.4 \pm 0.6	33.6 \pm 2.9	61.9
RSC (Huang et al., 2020)	85.2 \pm 0.9	77.1 \pm 0.5	65.5 \pm 0.9	46.6 \pm 1.0	38.9 \pm 0.5	62.7
Mixup (Wang et al., 2020b)	84.6 \pm 0.6	77.4 \pm 0.6	68.1 \pm 0.3	47.9 \pm 0.8	39.2 \pm 0.1	63.4
MLDG (Li et al., 2018a)	84.9 \pm 1.0	77.2 \pm 0.4	66.8 \pm 0.6	47.7 \pm 0.9	41.2 \pm 0.1	63.6
SagNet (Nam et al., 2021)	86.3 \pm 0.2	77.8 \pm 0.5	68.1 \pm 0.1	48.6 \pm 1.0	40.3 \pm 0.1	64.2
CORAL (Sun & Saenko, 2016)	86.2 \pm 0.3	78.8 \pm 0.6	68.7 \pm 0.3	47.6 \pm 1.0	41.5 \pm 0.1	64.5
mDSDI (Bui et al., 2021)	86.2 \pm 0.2	79.0 \pm 0.3	69.2 \pm 0.4	48.1 \pm 1.4	42.8 \pm 0.1	65.1
SWAD (Cha et al., 2021)	88.1 \pm 0.4	79.1 \pm 0.4	70.6 \pm 0.3	50.0 \pm 0.4	46.5 \pm 0.2	66.9
DNA (ours)	88.4 \pm 0.1	79.0 \pm 0.1	71.2 \pm 0.1	52.2 \pm 0.4	47.2 \pm 0.1	67.6

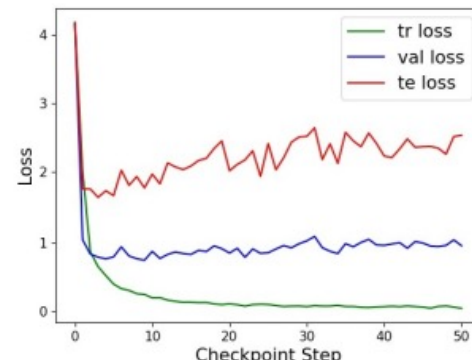
- With similar computational cost, the proposed DNA achieves competitive performance
- With lower computational cost, the proposed DNA achieves competitive performance compared with a common ensemble strategy
- The training process of the DNA (Fig. b) is more stable compared to the other variants, indicating the ease of model selection



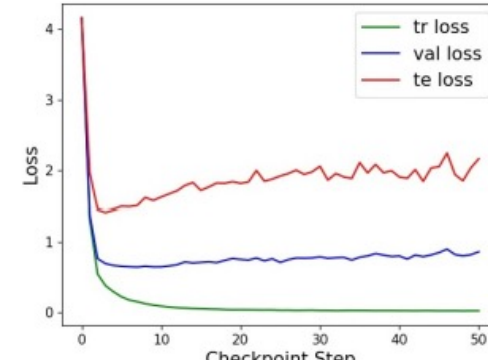
(a) PJS(single)



(b) PJS(SWA)



(c) CE(single)



(d) CE(SWA)