FY2025/2025年度

Machine Intelligence for Medical Engineering Team

Tatsuya Harada

医用機械知能チーム　原田 達也

RIKEN Center for Advanced Intelligence Project

A Theory of Learning Unified Model via Knowledge Integration from Label Space Varying Domains

D. Zhang, T. Westfechtel, T. Harada. CVPR2025

1. Problem Setting

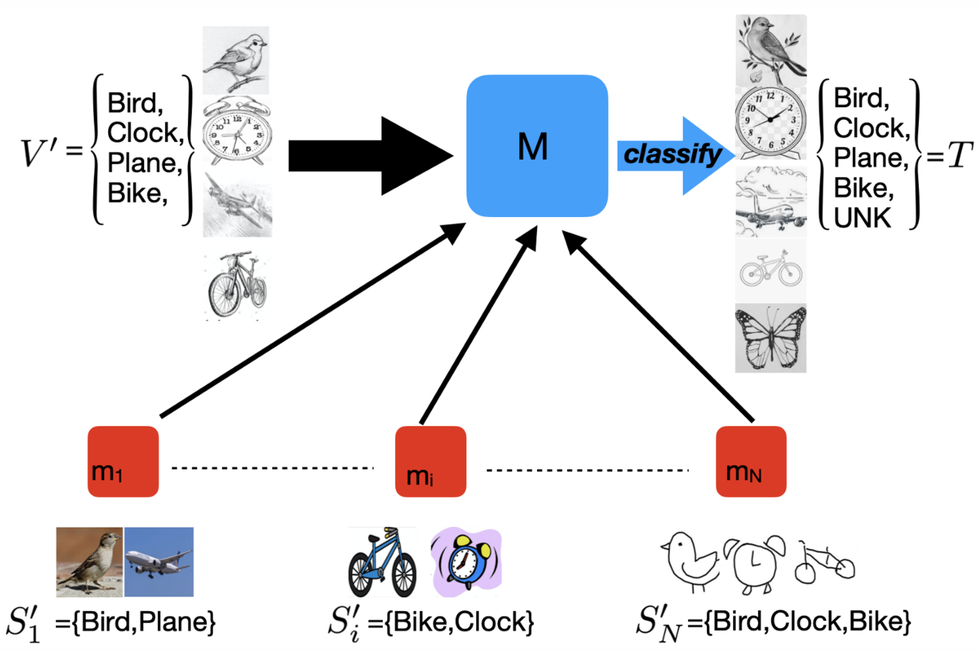

Figure 1: Given multiple models learned from different source domains with varying label space $\{S'_i : \mathcal{X} \times (\mathcal{Y}'_i \subseteq \mathcal{Y}')\}_{i=1}^N$, a few-shot labeled target domain $V' : \mathcal{X} \times (\mathcal{Y}' = \{1, ..., K-1\})$ and an unlabeled target domain including unseen categories $T : \mathcal{X} \times (\mathcal{Y} = \{1, ..., K\})$, the goal is to build a target classifier $h : \mathcal{X} \to \mathcal{Y}$ than can both recognize known and unknown classes.

2. Learning Theory

Preliminary 1 Let ϵ denote l_1 loss, for $\forall h \in \mathcal{H} : \mathcal{X} \to \{0, 1\}$, the following holds:

$$\epsilon_T(h) \leq \epsilon_S(h) + d_{\mathcal{H}\Delta\mathcal{H}}(S, T)/2 + \lambda_{S,T},$$
$$\lambda_{S,T} = \min_{h^* \in \mathcal{H}}[\epsilon_S(h^*) + \epsilon_T(h^*)], \text{(assumed small)}$$

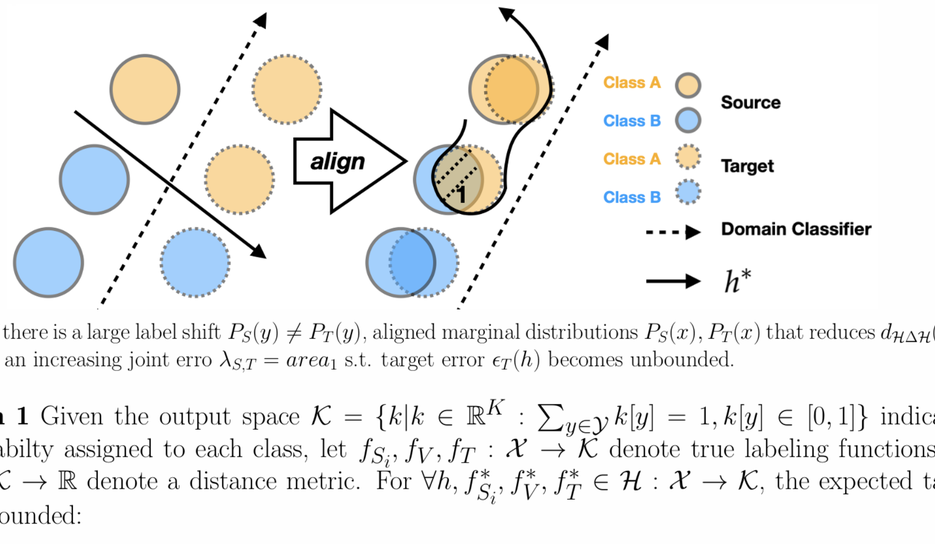

Figure 2: if there is a large label shift $P_S(y) \neq P_T(y)$, aligned marginal distributions $P_S(x), P_T(x)$ that reduces $d_{\mathcal{H}\Delta\mathcal{H}}(S, T)$ can lead to an increasing joint erro $\lambda_{S,T} = area_1$ s.t. target error $\epsilon_T(h)$ becomes unbounded.

Theorem 1 Given the output space $\mathcal{K} = \{k | k \in \mathbb{R}^K : \sum_{y \in \mathcal{Y}} k[y] = 1, k[y] \in [0, 1]\}$ indicating the probabilty assigned to each class, let $f_{S_i}, f_V, f_T : \mathcal{X} \to \mathcal{K}$ denote true labeling functions and $\epsilon : \mathcal{K} \times \mathcal{K} \to \mathbb{R}$ denote a distance metric. For $\forall h, f^*_{S_i}, f^*_V, f^*_T \in \mathcal{H} : \mathcal{X} \to \mathcal{K}$, the expected target error is bounded:

$$\begin{aligned} 2\epsilon_T(h) &\leq \epsilon_V(h) + \sum\nolimits_i \alpha_i U_i(h), \quad s.t. \quad \sum\nolimits_i \alpha_i = 1 \\ &= \epsilon_V(h) + \sum\nolimits_i \alpha_i [\epsilon_{S_i}(h) + 2 \underbrace{D_{S_i,V,T}(f^*_{S_i}, f^*_V, f^*_T, h)}_{\text{domain discrepancy}} + 2 \underbrace{\theta_i}_{\text{deviation}}], \end{aligned} \tag{1}$$

Assumption 1 $\exists f^*_{S_i}, f^*_V, f^*_T$ s.t. the empirical deviation from true labeling functions on finite data $\sum_i \alpha_i \hat{\theta}_i \to 0$.

Remark 1 Theorem 1 has a bounded joint error as $\min_h[\epsilon_V(h) + U_i(h)] \geq \lambda_{S_i,T} + \lambda_{V,T}$ to address increasing joint error in marginal distribution alignment due to large label shift. Assumption 1 is more feasible in large domain shift than $\hat{\lambda} \to 0$, allowing wider application in real-world problems.

3.Inference in Varying Label Space

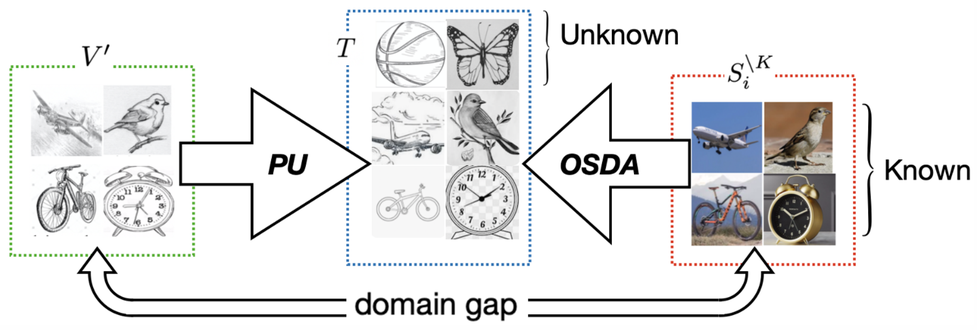

Figure 3: OSDA can be considered as PU with covariate shift where $P_{S_i}(x|y) \neq P_T(x|y)$ for $y \in \mathcal{Y}'$.

Definition 1 (Unknown Predictive Discrepancy) Let $\upsilon : \mathcal{K} \times \mathcal{K} \to \mathbb{R}$ denote the Unknown Predictive Discrepancy to measure the disagreement between K-th outputs of $f, f' : \mathcal{X} \to \mathcal{K}$.

Assumption 2 Let $S_i^k = P_{S_i}(x|y = k), T^k = P_T(x|y = k)$ denote class conditional distributions, $S_i^{\setminus K} = P_{S_i}(x|y \neq K), T' = P_T(x|y \neq K)$ indicate incomplete domains excluding unknown class K. $\exists g : \mathcal{X} \to \mathcal{Z}$ s.t. $P_{S^K}(z) = P_{T^K}(z), P_{S_i^{\setminus K}}(z) = P_{T'}(z)$.

Lemma 1 Let $\sum_{k=1}^K \pi^k_{S_i} = 1, \sum_{k=1}^K \pi^k_T = 1$ denote label distributions. Given Assumption 2, $\forall f : \mathcal{Z} \to \mathcal{K}$, the expectation over S can be estimated by $S_i^{\setminus K}$ and υ with a mild condition that $\pi^K_{S_i} = \pi^K_T = 1 - \beta$:

$$\epsilon_{S_i}(h; g) = \beta[\epsilon_{S_i^{\setminus K}}(h; g) - \upsilon_{S_i^{\setminus K}}(h; g)] + \upsilon_T(h; g) \tag{2}$$

Remark 2 Expectation regarding "missing classes" in $S_i^{\setminus K} : \mathcal{X} \times \mathcal{Y}'$ (includes categories not presenting in $S'_i : \mathcal{X} \times \mathcal{Y}'_i \subseteq \mathcal{Y}'$) is estimated with $V' : \mathcal{X} \times \mathcal{Y}'$.

4. Towards Source-Free Knowledge Transfer

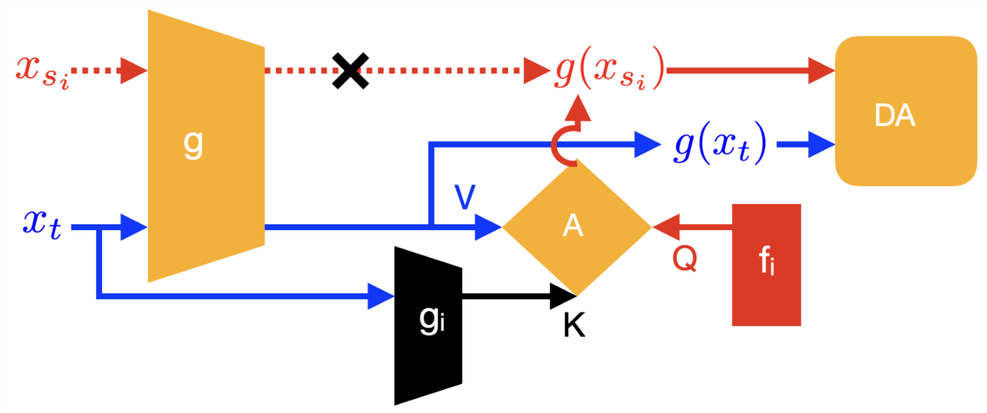

Figure 4: Given pre-trained source models consisting of black-box feature extractors $\{g_i : \mathcal{X} \to \mathcal{Z}_i \subseteq \mathbb{R}^{F_i}\}_{i=1}^N$ and visible Bayesian classifiers $\{(f_i; \mu_i, \sigma_i) : \mathcal{Z}_i \to \mathcal{K}\}_{i=1}^N$, the attention-based feature generation (AFG) module produces pseudo source features using a weighted average of unlabeled target features.

Pretrained source features $g_i(S'_i)$ of each class can be approximated by weight samples from f_i as

$$g_i(S'_i) = \begin{pmatrix} g_i(x^1_{s_i}) \\ \vdots \\ g_i(x^{\|\mathcal{Y}'_i\|}_{s_i}) \end{pmatrix} := \mu_i + \sigma_i \odot \begin{pmatrix} \zeta^1_i \\ \vdots \\ \zeta^{\|\mathcal{Y}'_i\|}_i \end{pmatrix}, \zeta^j_i \sim \mathcal{N}(0, I) \text{ with size } \|\mathcal{Y}'_i\| \times F_i$$

Traning source features $g(S'_i)$ are generated with the attention mechanism based on the distance in space $\mathcal{Z}'_i$ by learning query and key mapping functions $w_{q_i}, w_{k_i} : \mathcal{Z}_i \to \mathcal{Z}'_i \subseteq \mathbb{R}^{F'_i}$ contrastively to group similar features while pushing different ones away,

$$g(S'_i) = \text{softmax}(\frac{w_{q_i}(g_i(S'_i)) \cdot w_{k_i}(g_i(T'))^\mathsf{T}}{\sqrt{F'_i}}) g(T') \tag{3}$$

5. Experiment

METHOD	TYPE	→Clipart		→Product		→RealWorld		→Art		Avg.	
		1-shot	3-shot	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot
OSBP	Source-Combine	60.4	62.6	70.1	72.3	69.7	68.3	60.7	64.3	65.2	66.9
PGL		59.0	61.8	67.7	69.9	66.7	68.9	61.2	64.0	63.7	66.2
ANNA		65.8	67.7	71.0	73.4	70.3	70.3	61.0	63.7	67.0	68.8
PUJE		65.8	71.7	73.3	74.2	75.0	78.1	65.5	67.3	69.9	72.8
MOSDANET	Multi-Source	61.5	65.9	70.0	73.8	71.4	69.6	61.6	63.6	66.1	68.2
HyMOS		56.6	64.4	64.4	67.3	66.2	68.4	59.0	62.2	61.6	65.6
UM		68.0	72.1	79.0	83.0	79.4	80.8	67.7	70.3	**73.5**	**76.6**
MPU*	Source-Free	46.3	54.4	59.7	66.3	57.8	60.2	58.3	62.5	55.5	60.9
OSBP*		44.5	56.5	55.6	65.1	59.3	64.3	55.6	59.9	53.8	61.5
PUJE*		52.2	58.4	. 65.0	70.3	66.2	70.0	58.7	62.7	60.5	65.4
UM+AFG		61.1	66.0	77.0	80.1	72.0	78.8	60.3	64.6	**67.6**	**72.4**
METHOD	TYPE	→Clipart		→Painting		→Real		→Sketch		Avg.	
		1-shot	3-shot	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot
OSBP	Source-Combine	54.2	57.4	49.8	53.1	62.6	64.0	49.5	50.1	54.0	56.2
PGL		59.8	62.0	59.4	61.4	67.4	69.4	59.7	61.2	61.6	63.5
ANNA		55.6	61.5	53.6	54.3	67.5	66.5	57.9	58.1	58.7	60.1
PUJE		64.4	66.2	59.8	61.7	67.7	69.3	61.2	64.2	63.3	65.4
MOSDANET	Multi-Source	56.4	55.3	55.6	58.2	68.5	69.8	54.1	54.9	58.7	59.6
HyMOS		53.0	54.4	54.1	56.0	65.1	67.4	56.3	57.1	57.1	58.7
UM		70.3	71.5	66.0	68.8	75.1	78.5	66.1	69.5	**69.4**	**72.1**
MPU*	Source-Free	54.5	57.6	55.0	60.1	62.4	66.4	48.4	52.9	55.1	59.3
MOSDANET*		58.1	60.5	54.3	59.3	63.2	62.5	49.4	54.3	56.3	59.2
PUJE*		60.5	62.2	55.3	61.4	64.0	67.8	53.1	56.2	58.2	61.9
UM+AFG		64.8	69.7	60.0	64.2	67.6	73.4	60.0	64.8	**63.1**	**68.0**

Table 1: HOS (%) of ResNet-50 model fine-tuned on Office-Home & DomainNet dataset under 1-shot/3-shot settings (HOS=2(OS* × UNK)/(OS* + UNK); OS* denotes normalized accuracy for the known class only).

METHOD	TYPE	BACKBONE	Office-Home						DomainNet						Avg.		
			→Art			→Product			→Painting			→Real					
			UNK	OS*	HOS	UNK	OS*	HOS	UNK	OS*	HOS	UNK	OS*	HOS	UNK	OS*	HOS
UM	Multi-Source	ResNet-50	72.8	63.3	67.7	78.7	79.3	79.0	77.9	57.3	66.0	74.9	75.2	75.1	76.1	**68.8**	72.0
UM+AFG	Source-Free	ResNet-50	66.1	55.5	60.3	83.3	71.6	77.0	63.9	56.5	60.0	76.6	60.6	67.6	72.5	61.1	66.2
UM+AFG		ViT-16	77.7	59.8	67.5	87.6	80.9	84.1	68.5	57.4	62.5	82.8	73.2	77.7	**79.2**	67.8	**73.0**

Table 2: HOS (%) of ViT-B/16 model fine-tuned on Office-Home & DomainNet dataset under 1-shot setting.

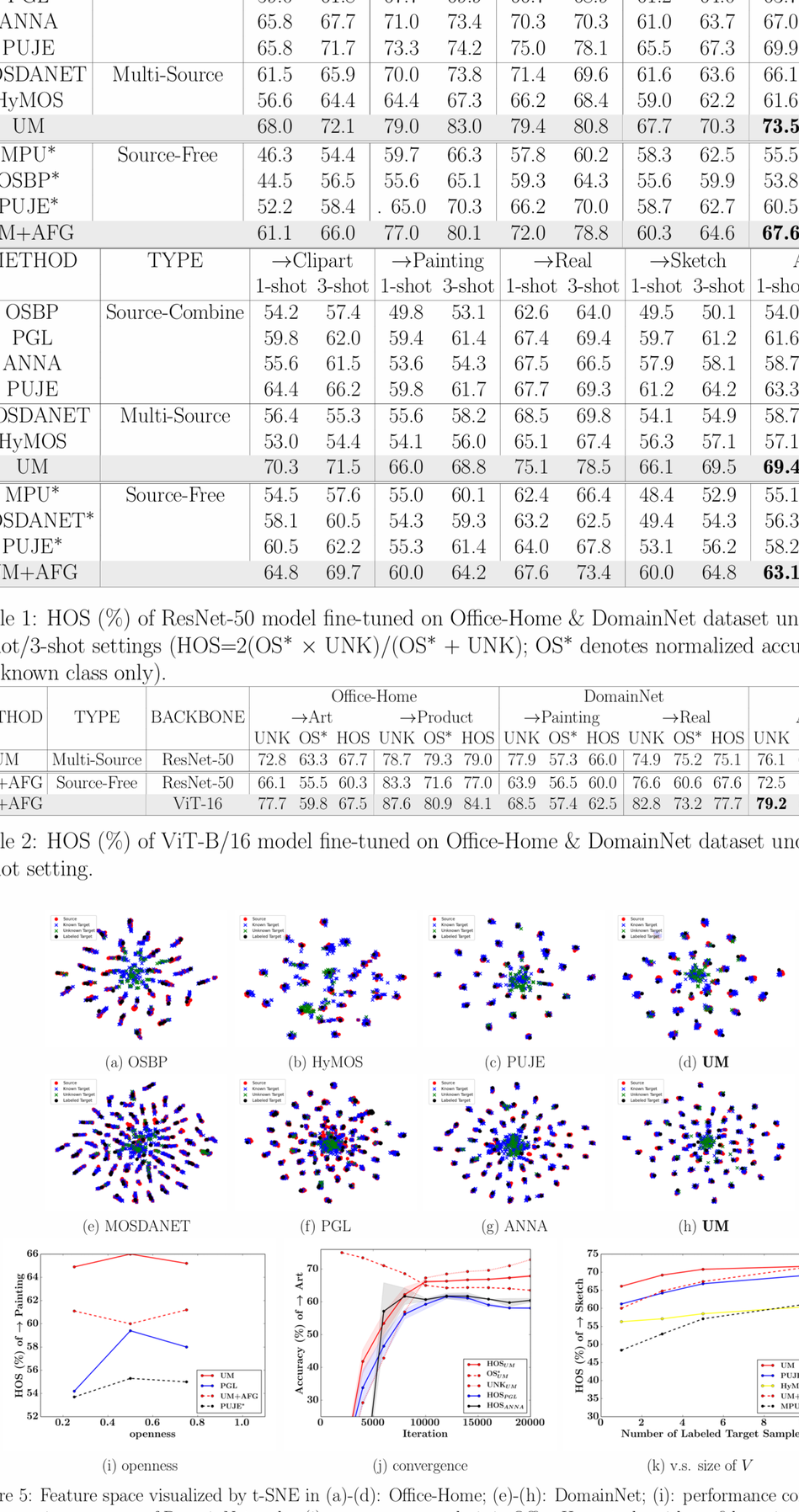

Figure 5: Feature space visualized by t-SNE in (a)-(d): Office-Home; (e)-(h): DomainNet; (i): performance comparisons w.r.t. varying openness of DomainNet tasks; (j): convergence analysis in Office-Home tasks with confidence intervals; (k): accuracy v.s. the number of labeled target samples in DomainNet tasks.