

Reinforced Domain Selection for Continuous Domain Adaptation

Hanbing Liu, Huaize Tang, Yanru Wu, Yang Li[†] and Xiao-Ping Zhang

Shenzhen Key Laboratory of Ubiquitous Data Enabling, Tsinghua Shenzhen International Graduate School, Tsinghua University
{liuhb24,thz21,wu-yr21}@mails.tsinghua.edu.cn, yangli@sz.tsinghua.edu.cn, xpzhang@ieee.org

Abstract—Continuous Domain Adaptation (CDA) effectively bridges significant domain shifts by progressively adapting from the source domain through intermediate domains to the target domain. However, selecting intermediate domains without explicit metadata remains a substantial challenge that has not been extensively explored in existing studies. To tackle this issue, we propose a novel framework that combines reinforcement learning with feature disentanglement to conduct domain path selection in an unsupervised CDA setting. Our approach introduces an innovative unsupervised reward mechanism that leverages the distances between latent domain embeddings to facilitate the identification of optimal transfer paths. Furthermore, by disentangling features, our method facilitates the calculation of unsupervised rewards using domain-specific features and promotes domain adaptation by aligning domain-invariant features. This integrated strategy is designed to simultaneously optimize transfer paths and target task performance, enhancing the effectiveness of domain adaptation processes. Extensive empirical evaluations on datasets such as *Rotated MNIST* and *ADNI* demonstrate substantial improvements in prediction accuracy and domain selection efficiency, establishing our method’s superiority over traditional CDA approaches.

Index Terms—continuous domain adaptation, domain selection, reinforcement learning, feature disentanglement, unsupervised reward mechanism

I. INTRODUCTION

Domain shift is a common challenge in many real-life applications [1]. Continuous Domain Adaptation (CDA) strategically mitigates the challenge of significant domain shifts by seamlessly transferring from the source domain through various intermediate domains to the target domain [2], [3]. Traditional CDA methods such as self-training, pseudo-labeling, adversarial algorithms, and optimal transport have advanced significantly [4]–[14]. However, the dynamic selection of intermediate domains without explicit metadata remains a complex problem [15]. To address this issue, novel methodologies have been proposed to enhance domain sorting and minimize errors. For instance, [16] introduces a progressive domain discriminator in their work, enabling the generation of domain sequences without pre-defined domain indexes. Furthermore, [7] introduced a Wasserstein-based transfer curriculum that strategically sorts intermediate domains using Wasserstein distance, reducing cumulative errors through a multi-path strategy. Nonetheless, these methods mainly focus on sorting available domains. The resulting sequence of domains, a.k.a. transfer path, is not guaranteed to be optimal: sparser path would lead to larger gaps between subsequent domains, prone to negative transfer, while denser path requires longer training time and is more likely to accumulate errors. Hence, it is necessary to incorporate dynamic transfer path learning during the adaptation process to incorporate only the essential domains.

Learning the optimal path among multiple intermediate domains poses a significant combinatorial optimization challenge due to its dynamic nature [17]. In response, we propose a novel Reinforcement

Learning (RL) strategy for dynamically selecting intermediate domains during the CDA process as illustrated in Figure 1(a). Existing methods combining RL with continual domain adaptation, such as [17]–[20], typically focus on dynamically expanding network models, which significantly increases computational demands. Diverging from these approaches, we are inspired by [21], which utilized policy gradient methods for training data selection in supervised settings. However, their reliance on target labels for reward calculation is not viable in unsupervised settings. Additionally, the omission of domain continuity limits its applicability to the CDA challenge.

Therefore, we propose a surrogate reward function based on the distance between feature distributions of different domains to learn the optimal domain selection policy without supervision. Furthermore, features are inherently high-dimensional, which requires longer training time for the system to converge to an optimal policy [22]. We leverage the finding in CDA that task-related information is typically invariant to domain shifts [23], [24]. By disentangling features into domain-invariant and domain-specific components, our approach not only learns a domain-agnostic model but also enhances the accuracy of domain shift estimations through the application of low-dimensional latent domain embeddings. To the best of our knowledge, this study is the first to integrate reinforcement learning with feature disentanglement to tackle the domain selection challenge in CDA scenario. Extensive evaluations across handwritten digits and medical image classification datasets demonstrate our approach’s superiority, enhancing prediction accuracy and path selection strategy over traditional CDA methods. The key contributions of this study are outlined as follows:

- 1) **RL-based intermediate domain selection:** We formulate the problem of dynamically selecting intermediate domains using RL with feature disentanglement to optimize the transfer path, focusing on simultaneous optimization of the transfer path and prediction outcomes.
- 2) **Novel reward mechanism:** Our novel domain selection policy employs an unsupervised reward mechanism based on the distance between latent domain embeddings, enhancing strategy precision.
- 3) **Disentangled domain embedding:** We separate domain-specific from domain-invariant features to improve the extraction of transferable features for domain adaptation, while enabling more precise domain shift estimations through low-dimensional embeddings.

II. METHODOLOGY

A. Problem Definition

This research investigates a CDA setting involving multiple unlabeled auxiliary domains accompanying a labeled source domain. The domain indices are defined as $\mathcal{G} = \mathcal{G}_s \cup \mathcal{G}_t \cup \mathcal{G}_i$, categorizing

[†] Corresponding author.

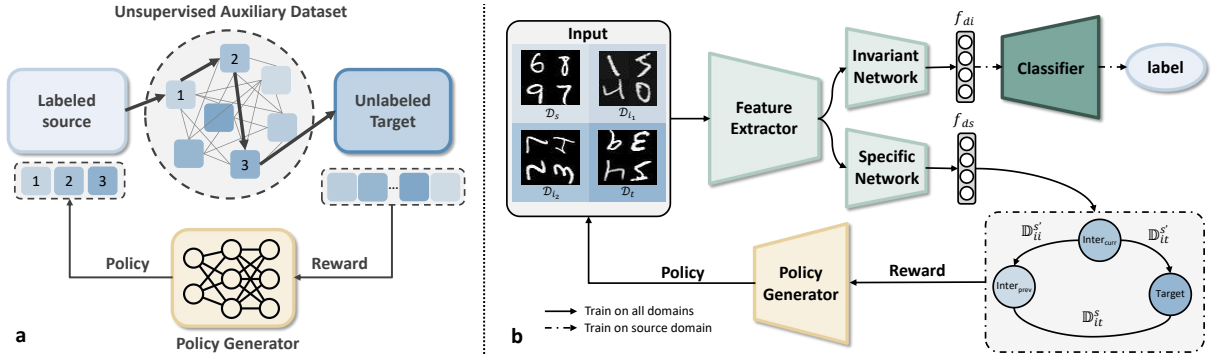


Fig. 1. **Overview and framework of our method.** (a) Overview of Continual Domain Adaptation using Reinforcement Learning. Our approach employs a policy generator to devise strategies for intermediate domains, thus establishing an optimal transfer path. (b) Framework of the Proposed Method. Input from the source, target, and intermediate domains is processed by a feature extractor to derive common features. Subsequently, a dual-network system isolates domain-invariant and domain-specific features. The domain-specific features from various domains are then evaluated based on their distances to calculate rewards, which assist the policy generator in formulating policies for each intermediate domain.

each domain into source (\mathcal{G}_s), target (\mathcal{G}_t), and intermediate (\mathcal{G}_i). The source domain \mathcal{D}_s consists of labeled tuples $(\mathbf{x}_j^s, y_j^s, g_j^s)$, where \mathbf{x}_j^s represents feature vectors, y_j^s are the labels, and g_j^s indicates the domain index for each $j = 1, \dots, n_s$. The unlabeled target domain, denoted as \mathcal{D}_t , comprises tuples (\mathbf{x}_j^t, g_j^t) for $j = 1, \dots, n_t$. Similarly, the unlabeled intermediate domain \mathcal{D}_{i_k} includes tuples (\mathbf{x}_j^i, g_j^i) for $j = 1, \dots, n_{i_k}$ and $k = 1, \dots, K$, which means there are K intermediate domains. \mathbf{x}_j^t and \mathbf{x}_j^i are represented as feature vectors, indexed by $g_j^t \in \mathcal{G}_t$ and $g_j^i \in \mathcal{G}_i$. This reinforced domain selection problem in CDA can be formulated by its transfer path $h = (\mathcal{G}_{h_1}, \mathcal{G}_{h_2}, \dots, \mathcal{G}_{h_L})$ where $L \leq K$ and $\{h_l\}_{l=1}^L$ is the domain id of the l -th intermediate domain in the path. The primary objective of this research is to derive an optimal path \hat{h} , such that by transferring along \hat{h} during training, we accurately predict the labels $(y_j^t)_{j=1}^{n_t}$ for the feature vectors in the unlabeled target domain \mathcal{D}_t , utilizing both the labeled data from the source domain and the structural insights gleaned from the intermediate domains.

B. Method Framework

We present a novel approach that integrates reinforcement learning with feature disentanglement to address the challenges of domain selection in CDA, as depicted in Figure 1(b). Our methodology utilizes a dataset consisting of a labeled source domain, multiple unlabeled intermediate domains, and a target domain. A feature extractor F isolates common features across these domains, which are subsequently processed by an invariant network I and a specific network S . The invariant network extracts domain-invariant features f_{di} , while the specific network focuses on domain-specific features f_{ds} , with mutual information ensuring the independence of these components. Subsequently, f_{ds} from intermediate domains feed into a policy generation network P , which formulates a selection strategy via a policy gradient algorithm based on the discrepancies between the domains. During inference, the trained networks \hat{F} , \hat{I} , and classifier \hat{C} collaborate to predict labels in the target domain.

C. Feature Disentanglement

To enhance feature alignment and decomposition, we employ a feature extractor F , along with an Invariant network I and a Specific network S . Initially, all domains are input into the feature extractor to derive common features. These are subsequently handled by networks I and S to extract domain-invariant and domain-specific features, respectively, ensuring targeted feature isolation for subsequent analysis. The features are categorized into two types: domain-invariant

feature f_{di} and domain-specific feature f_{ds} . These features are further divided into three components for each domain type: f_{di}^s, f_{di}^i , and f_{di}^t which represent the domain-invariant features of the source, intermediate, and target domains, respectively. Similarly, f_{ds}^s, f_{ds}^i , and f_{ds}^t correspond to the domain-specific features. The features f_{di} and f_{ds} are derived by modulating Mutual Information (MI) [25], a fundamental metric quantifying the dependency between two random variables. Given the substantial computational complexity of calculating MI directly, we utilize the Mutual Information Neural Estimator (MINE) [26], which provides a practical approach to estimate MI from n i.i.d. samples using a neural network T_θ . In our neural network model, the lower bound of mutual information estimation is implemented as

$$I(X; Z) = \frac{1}{n} \sum_{i=1}^n T_\theta(\mathbf{x}^{(i)}, \mathbf{z}^{(i)}) - \log \left(\frac{1}{n} \sum_{i=1}^n e^{T_\theta(\mathbf{x}^{(i)}, \bar{\mathbf{z}}^{(i)})} \right), \quad (1)$$

where $(\mathbf{x}^{(i)}, \mathbf{z}^{(i)})$ represent samples from the joint distribution and $\bar{\mathbf{z}}^{(i)}$ are sampled from the marginal distribution. This neural network-based methodology offers a scalable solution for estimating MI in extensive datasets.

The loss functions for the invariant and specific networks, \mathcal{L}_{mi} and \mathcal{L}_{ms} respectively, are defined as follows:

$$\mathcal{L}_{mi} = -[I(f_{di}^s; f_{di}^i) + I(f_{di}^s; f_{di}^t) + I(f_{di}^i; f_{di}^t)], \quad (2)$$

$$\mathcal{L}_{ms} = I(f_{ds}^s; f_{ds}^i) + I(f_{ds}^s; f_{ds}^t) + I(f_{ds}^i; f_{ds}^t), \quad (3)$$

where I denotes the mutual information estimation as defined in Equation 1. The invariant network strives to maximize the pairwise mutual information, while the specific network aims to minimize it. Subsequently, the unified domain-invariant feature f_{di}^s from the source domain is classified by the classifier C , utilizing cross-entropy loss \mathcal{L}_{ce} for supervised learning:

$$\mathcal{L}_{ce} = -\mathbb{E} \left[\sum_{m=1}^M \mathbb{1}[m = y^s] \log(C(f_{di}^s)) \right], \quad (4)$$

where M denotes the number of categories in a classification problem. By isolating both unified and distinctive features, we improve the efficiency of transfer processes and support the development of policy generation.

Algorithm 1: Joint training algorithm

Input: Source domain \mathcal{D}_s , Target domain \mathcal{D}_t , Intermediate domains $\mathcal{D}_{i_k}, k = 1, \dots, K$, transfer path h , Feature Extractor F , Invariant network I , Specific network S , and Policy Generator P .

Output: Well-trained $\hat{F}, \hat{I}, \hat{S}$, and \hat{P} .

```
1 Initialize:  $F, I, S, P \sim$  Xavier initialization
2 for  $e \leftarrow 1, \dots$  do
3   for  $k \leftarrow 1, \dots, K$  do
4     Shuffle intermediate domain set
5     for  $\mathcal{D}_{i_k}$  in intermediate domain set do
6       Feature Disentanglement
7       update  $F, I, C$  using Equation 2 + 4
8       update  $F, S$  using Equation 3
9       Policy Generation
10      update transfer path  $h$  according to action  $a$ 
11      compute one-step reward using Equation 5
12    Cumulative Reward
13    compute cumulative reward using Equation 6
14  Policy Gradient Ascent
15  compute the mean of the above cumulative reward
16  update  $P$  using Equation 8
```

D. Policy Generation

We employ Reinforcement Learning (RL) to generate policies for each intermediate domain. The components of this RL framework are defined as follows.

State: The state s is defined by the domain-specific features. Specifically, s at time T is represented as $s_T = (\Phi_{T-1}^i, \Phi_T^i, \Phi_T^t)$, where Φ denotes the domain-specific features f_{ds} . Here, Φ_{T-1}^i corresponds to the domain-specific features of the previously selected intermediate domain, Φ_T^i to the current intermediate domain, and Φ_T^t to the target domain. The initial state, Φ_{T-1}^i is set to be the representation of the source domain, denoted as Φ_{T-1}^s .

Action: The action a is a binary decision that can take values of either 0 or 1, which determines whether the current intermediate domain is selected during the policy generation process. The action a is determined probabilistically to be 1 with probability p and 0 with probability $1 - p$, where p is the output probability from the policy network. This binary decision constitutes the policy $\pi(a|s)$. After each action, the policy network updates the domain-specific feature representation Φ_T^i , resulting in a transition of the state s to s' .

Reward: The reward $R_T(s, a, s')$ is generated through an unsupervised reward mechanism that measures the distance between domain-specific features across various domains. This reward is calculated based on future states and is defined as follows:

$$R_T = \begin{cases} 2\mathbb{D}_{it}^s - \mathbb{D}_{ii}^{s'} - \mathbb{D}_{it}^{s'} & \text{if } a = 1, \left(\mathbb{D}_{ii}^{s'} < \mathbb{D}_{it}^s \right) \text{ and } \left(\mathbb{D}_{it}^{s'} < \mathbb{D}_{it}^s \right) \\ -inf & \text{if } a = 1, \left(\mathbb{D}_{ii}^{s'} \geq \mathbb{D}_{it}^s \right) \text{ and } \left(\mathbb{D}_{it}^{s'} \geq \mathbb{D}_{it}^s \right) \\ 0 & \text{if } a = 0. \end{cases} \quad (5)$$

where s' denotes the subsequent state. The terms \mathbb{D}_{it} and \mathbb{D}_{ii} represent the distances $d(\Phi_T^i, \Phi_T^t)$ and $d(\Phi_T^i, \Phi_{T-1}^i)$, respectively, which measure the distance between the current intermediate domain and the target domain, and the distance between the current and previously selected intermediate domains. This reward strategy prior-

itizes selecting domains that shorten the transfer distance, specifically those where the distances to both the previously selected intermediate domain and the target domain are smaller than the distance between the previously selected intermediate domain and the target domain. The distance function d is defined using the Wasserstein distance [27]. This metric is essential for establishing generalization bounds in domain adaptation and is highly effective in reducing domain shift within the Wasserstein framework [7], [28], [29].

In this study, we employ a neural network-based policy generator within the RL framework, utilizing the classical policy gradient method [30]. This method models the policy parametrically as $\pi(a|s; \theta)$, optimized through gradient ascent to maximize the expected value of the value function, $J(\theta) = \mathbb{E}_s[V(s; \theta)]$. The state value function $V(s; \theta)$ and the cumulative reward function $Q(s, a)$ are defined as follows:

$$V(s; \theta) = \sum_a \pi(a|s; \theta) \cdot Q(s, a), \quad (6)$$
$$Q(s, a) = \mathbb{E}_{a \sim \pi(\cdot|s; \theta)} \left[\sum_{k=T}^{\infty} \gamma^{k-T} R_k(s, a, s') \right],$$

where γ denotes the discount factor, and $R(s, a, s')$ represents the single-step reward as Equation 5. Policy updates are conducted using gradient ascent:

$$\theta \leftarrow \theta + \tau \nabla_{\theta} J(\theta), \quad (7)$$
$$\nabla_{\theta} J(\theta) = \mathbb{E}_{a \sim \pi(\cdot|s; \theta)} [\nabla_{\theta} \log \pi(a|s; \theta) \cdot Q(s, a)],$$

which adjusts the likelihood of actions based on their reward. By incorporating Equation 5 and Equation 6 into Equation 7, we can derive the specific form of $\nabla_{\theta} J(\theta)$,

$$\nabla_{\theta} J(\theta) = a \nabla_{\theta} \log p \sum_{k=T}^{\infty} \gamma^{k-T} R_k(s, a, s') + (1 - a) \nabla_{\theta} \log(1 - p) \sum_{k=T}^{\infty} \gamma^{k-T} R_k(s, a, s') \quad (8)$$

The policy gradient method leverages the advantages of reinforcement learning, allowing for dynamic adjustments to the policy in response to changes in domain-specific features. We adopt a joint training approach for both the feature extractor and the policy generator, as outlined in Algorithm 1, to achieve simultaneous optimization of the transfer path and prediction outcomes.

III. EXPERIMENT RESULTS

A. Data Description

We utilize two datasets in our study. The Rotated MNIST dataset [31] expands upon the original MNIST by featuring images of digits rotated between 0 to 180 degrees, organized into 11 domains, with each domain containing 1,000 samples. The Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset [32] provides MRI images to support Alzheimer's research. It comprises a source domain with individuals aged 50-70, including 190 samples, and several intermediate domains for ages 70-92, each with 50 samples, which are utilized for classifying into five disease categories.

B. Quantitative and Qualitative Results

We have compared classical Continuous Domain Adaptation (CDA) methods with those that incorporate Reinforcement Learning (RL) to address challenges within CDA, setting the number of intermediate domains at four. As demonstrated in Table I, our approach

TABLE I
ACCURACY FOR TWO DATASETS OF DIFFERENT ALGORITHMS

Category	Method	ROT MNIST (\uparrow)	ADNI (\uparrow)
CDA	EAML [33]	70.4	68.3
	AGST [8]	76.2	57.3
	Gradual ST [9]	87.9	64.5
	CDOT [15]	75.6	82.6
	CMA [10]	65.3	55.4
	W-MPOT [7]	89.1	88.3
	CIDA [6]	85.7	73.6
CDA with RL	SDG [21]	89.8	80.5
	GRADIENT [18]	90.2	88.4
	CLEAS [19]	92.8	86.1
	RCL [17]	91.5	89.3
	Ours	93.4	90.5

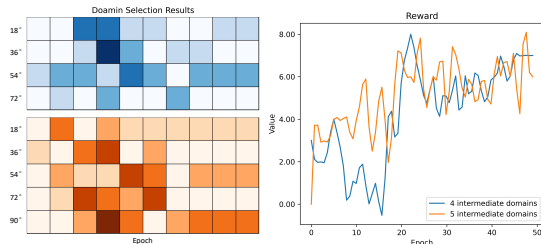


Fig. 2. **Reinforced Domain selection results.** The y-axis of the left figure represents various intermediate domains, each identified by a specific rotation angle ranging from the source domain at 0 degrees to the target domain, which is the last intermediate domain incremented by an additional 18 degrees. The x-axis samples every five epochs. A color gradient from light to dark illustrates the order of domain selection throughout the experiment, with white denoting domains that were not selected. The blue and yellow curves in the right figure represent setups with four and five intermediate domains, respectively, consistent with the configurations shown in the left figure.

achieves the highest accuracy on both datasets, with improvements of 0.6% and 1.2% on Rotated MNIST and ADNI, respectively. The greater gains on the ADNI dataset, which is higher-dimensional and more complex than MNIST, underscore our method’s capability to effectively select paths and enhance performance in complex settings. Furthermore, the results demonstrate that RL significantly enhances CDA effectiveness, as evident from the superior performance in the lower half of the table.

Figure 2 demonstrates that as rewards increase, domain selection becomes more strategic, evidenced by a systematic increase in the angles of the selected domains. Upon reward stabilization, the model selectively omits intermediate domains with minor shifts or redundant details to minimize errors. The optimal paths for configurations with four and five intermediate domains are respectively $\hat{h} = (18, 54)$ and $\hat{h} = (18, 54, 90)$. Subsequently, Figure 3 shows that the specific network effectively captures distinct features from each domain, which exhibit a continuity that facilitates continual domain transfer. Conversely, the invariant network extracts domain-invariant features, facilitating a unified feature distribution across various domains. This highlights the effectiveness of mutual information in ensuring consistent

TABLE II
ABLATION STUDY FOR THE NUMBER OF INTERMEDIATE DOMAIN POOL

K	Accuracy
2	92.8
3	93.2
4	93.4
5	93.8
6	94.1
7	93.5
8	93.2

feature representation, regardless of domain variations.

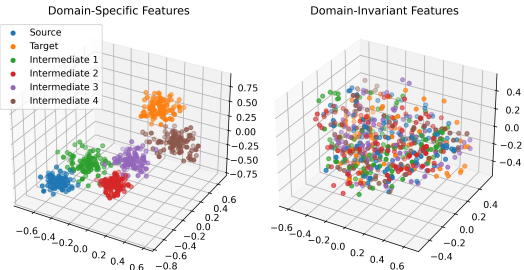


Fig. 3. **Visualizations of the domain-specific and domain-invariant features.** This image was generated using the t-SNE [34] method and visualized by reducing the dimensions to three. Different colors represent different domains.

TABLE III
ABLATION STUDY FOR THE MODEL STRUCTURE

Classification Model	Feature Disentanglement	Policy Generation	Accuracy
✓	✗	✗	50.8
✓	✓	✗	81.5
✓	✓	✓	93.4

C. Ablation Study

We conducted an ablation study on the Rotated MNIST dataset to evaluate the impact of feature disentanglement and policy generation within our CDA framework. As depicted in Table III, incorporating these components increased the prediction accuracy to 93.4%, highlighting their critical role in improving model performance. Additionally, we investigated the impact of varying the number of intermediate domain pool, denoted by K (Table II). Contrary to typical CDA challenges, our method, enhanced by the policy generation mechanism, maintained high accuracy even with an increased number of intermediate domains, demonstrating its robust ability to manage domain transfer effectively.

IV. CONCLUSION

Our study addresses the challenge of dynamic domain selection in Continuous Domain Adaptation by joint Reinforcement Learning and feature disentanglement, simultaneously optimizing the transfer path and improving prediction outcomes. Our domain selection policy, driven by an unsupervised reward mechanism based on distances between latent domain embeddings and learned through a policy gradient algorithm, significantly enhances strategic precision. By distinguishing domain-specific from domain-invariant features, our approach improves the extraction of transferable features vital for effective domain adaptation and enables more precise estimations of domain shifts using low-dimensional embeddings. Empirical validation on datasets like Rotated MNIST and ADNI confirms our method’s superiority, surpassing traditional CDA approaches with improved prediction accuracy and more efficient path selection.

ACKNOWLEDGEMENTS

This work is supported in part by the Natural Science Foundation of China (Grant 62371270), Shenzhen Key Laboratory of Ubiquitous Data Enabling (No.ZDSYS20220527171406015), and Tsinghua Shenzhen International Graduate School-Shenzhen Pengrui Endowed Professorship Scheme of Shenzhen Pengrui Foundation.

REFERENCES

- [1] H. Chen, Y. Song, L.-R. Dai, I. McLoughlin, and L. Liu, "Self-supervised representation learning for unsupervised anomalous sound detection under domain shift," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 471–475.
- [2] M. Zhang, H. Marklund, N. Dhawan, A. Gupta, S. Levine, and C. Finn, "Adaptive risk minimization: Learning to adapt to domain shift," *Advances in Neural Information Processing Systems*, vol. 34, pp. 23 664–23 678, 2021.
- [3] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.
- [4] J. Moon, D. Das, and C. G. Lee, "Multi-step online unsupervised domain adaptation," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 41 172–41 576.
- [5] Y. Xu, Z. Jiang, A. Men, Y. Liu, and Q. Chen, "Delving into the continuous domain adaptation," in *Proceedings of the 30th ACM International Conference on Multimedia*, 2022, pp. 6039–6049.
- [6] H. Wang, H. He, and D. Katabi, "Continuously indexed domain adaptation," in *Proceedings of the 37th International Conference on Machine Learning*, 2020, pp. 9898–9907.
- [7] H. Liu, J. Wang, X. Zhang, Y. Guo, and Y. Li, "Enhancing continuous domain adaptation with multi-path transfer curriculum," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 2024, pp. 286–298.
- [8] S. Zhou, L. Wang, S. Zhang, Z. Wang, and W. Zhu, "Active gradual domain adaptation: Dataset and approach," *IEEE Transactions on Multimedia*, vol. 24, pp. 1210–1220, 2022.
- [9] A. Kumar, T. Ma, and P. Liang, "Understanding self-training for gradual domain adaptation," in *International Conference on Machine Learning*. PMLR, 2020, pp. 5468–5479.
- [10] J. Hoffman, T. Darrell, and K. Saenko, "Continuous manifold based adaptation for evolving visual domains," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 867–874.
- [11] J. Liang, R. He, Z. Sun, and T. Tan, "Distant supervised centroid shift: A simple and efficient approach to visual domain adaptation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2975–2984.
- [12] J. Liang, D. Hu, and J. Feng, "Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation," in *International Conference on Machine Learning*. PMLR, 2020, pp. 6028–6039.
- [13] A. Bitarafan, M. S. Baghshah, and M. Gheisari, "Incremental evolving domain adaptation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 8, pp. 2128–2141, 2016.
- [14] E. Tzinis, Y. Adi, V. K. Ithapu, B. Xu, and A. Kumar, "Continual self-training with bootstrapped remixing for speech enhancement," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 6947–6951.
- [15] G. Ortiz-Jiménez, M. El Gheche, E. Simou, H. P. Maretic, and P. Frossard, "Forward-backward splitting for optimal transport based problems," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 5405–5409.
- [16] H.-Y. Chen and W.-L. Chao, "Gradual domain adaptation without indexed intermediate domains," *Advances in Neural Information Processing Systems*, vol. 34, pp. 8201–8214, 2021.
- [17] J. Xu and Z. Zhu, "Reinforced continual learning," *Advances in neural information processing systems*, vol. 31, 2018.
- [18] P. Huang, M. Xu, J. Zhu, L. Shi, F. Fang, and D. Zhao, "Curriculum reinforcement learning using optimal transport via gradual domain adaptation," *Advances in Neural Information Processing Systems*, vol. 35, pp. 10 656–10 670, 2022.
- [19] Q. Gao, Z. Luo, D. Klabjan, and F. Zhang, "Efficient architecture search for continual learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 11, pp. 8555–8565, 2022.
- [20] L. Chen, C. Chang, Z. Chen, B. Tan, M. Gašić, and K. Yu, "Policy adaptation for deep reinforcement learning-based dialogue management," in *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2018, pp. 6074–6078.
- [21] M. Liu, Y. Song, H. Zou, and T. Zhang, "Reinforced training data selection for domain adaptation," in *Proceedings of the 57th annual meeting of the association for computational linguistics*, 2019, pp. 1957–1968.
- [22] R. Mowakeaa, S.-J. Kim, and D. K. Emge, "Kernal-based lifelong policy gradient reinforcement learning," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 3500–3504.
- [23] Q. Lao, X. Jiang, M. Havasi, and Y. Bengio, "Continuous domain adaptation with variational domain-agnostic feature replay," *arXiv preprint arXiv:2003.04382*, 2020.
- [24] X. Peng, Z. Huang, X. Sun, and K. Saenko, "Domain agnostic learning with disentangled representations," in *International conference on machine learning*. PMLR, 2019, pp. 5102–5112.
- [25] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, vol. 69, no. 6, p. 066138, 2004.
- [26] M. I. Belghazi, A. Baratin, S. Rajeshwar, S. Ozair, Y. Bengio, A. Courville, and D. Hjelm, "Mutual information neural estimation," in *International conference on machine learning*. PMLR, 2018, pp. 531–540.
- [27] C. Villani and C. Villani, "The wasserstein distances," *Optimal Transport: Old and New*, pp. 93–111, 2009.
- [28] J. Shen, Y. Qu, W. Zhang, and Y. Yu, "Wasserstein distance guided representation learning for domain adaptation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.
- [29] I. Redko, E. Morvant, A. Habrard, M. Sebban, and Y. Bennani, "A survey on domain adaptation theory: learning bounds and theoretical guarantees," *arXiv preprint arXiv:2004.11829*, 2020.
- [30] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller, "Deterministic policy gradient algorithms," in *International conference on machine learning*. Pmlr, 2014, pp. 387–395.
- [31] T. Nishida, T. Endo, and Y. Kawaguchi, "Zero-shot domain adaptation of anomalous samples for semi-supervised anomaly detection," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.
- [32] R. C. Petersen, P. S. Aisen, L. A. Beckett, M. C. Donohue, A. C. Gamst, D. J. Harvey, C. R. Jack, W. J. Jagust, L. M. Shaw, A. W. Toga *et al.*, "Alzheimer's disease neuroimaging initiative (adni): clinical characterization," *Neurology*, vol. 74, no. 3, pp. 201–209, 2010.
- [33] H. Liu, M. Long, J. Wang, and Y. Wang, "Learning to adapt to evolving domains," *Advances in neural information processing systems*, vol. 33, pp. 22 338–22 348, 2020.
- [34] C. Séjourné, R. Couillet, and P. Comon, "A large-dimensional analysis of symmetric sne," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 2970–2974.