Adversarial Unsupervised Domain Adaptation via Real Data Augmentation and Uncertainty Penalization

Zhenghan Chen[a,*], *Yi Ge*[b,*], *Guohua Dong*[c,**], *Lei Zhu*[d1]

[a]*Peking University, Beijing, China*
[b]*Carnegie Mellon University, Pittsburgh, USA*
[c]*Beijing Institute of Basic Medical Sciences, Beijing, China*
[d]*Tongji University, Shanghai, China*

***Abstract*—Deep learning architectures are changing the way we process visual data, and synthetic data augmentation is well known as an essential paradigm for visual adaptation. However, traditional approaches to *unsupervised domain adaptation* (UDA) are limited by the assumption of consistent label spaces, an assumption that becomes less tenable as it meets real-world applications. Furthermore, as source domains become more resource-constrained, the differences between these source domains and the increasingly large and heterogeneous target domains must be bridged. To handle these primitive issues, we present a new framework called partial unsupervised domain adaptation (PUDA) using real data enhancement and uncertainty penalty (RDAUP). We innovatively reformulate PUDA as a vanilla UDA problem and develop an adversarial domain adaptation-based method to augment the real training data distribution adaptively. We also develop a Pressing Transfer Mechanism to better utilize queue features in domain adaptation tasks. Our primary theoretical contribution is a penalty for uncertainty, which uses the happiness of the classifier on unlabeled data in the source domain to aggressively punish the model's predictions when confused, ultimately fine-tuning the model's predictions. We substantiate the superiority of our framework through extensive empirical evaluation of challenging PUDA tasks, including VisDA, where UDA methods show significant limitations. Our results enhance the state-of-the-art for domain adaptation and also deliver vital insights that illuminate the theory behind transfer learning across heterogeneous domains.**

***Index Terms*—Partial unsupervised domain adaptation, Adversarial learning, Uncertainty penalization**

I. Introduction

The cornerstone of statistical machine learning theory is based on a fundamental epistemological assumption: the intrinsic alignment between training and test data distributions [1], [2]. This theoretical paradigm, while elegant in its mathematical formulation, encounters significant challenges when confronted with the complexities of real-world applications. In particular in transfer learning scenarios, the acquisition of large-scale, domain-specific, manually labeled training data not only incurs prohibitive costs but also presents formidable operational and methodological challenges [3]. This fundamental tension between theoretical ideals and practical constraints manifests itself most prominently in the form of sample selection bias, a persistent limitation that has profound implications for the generalizability and robustness of contemporary machine learning systems.

Unsupervised Domain Adaptation (UDA) [4], [5], [6], [7], [8] has emerged as a transformative paradigm in addressing these fundamental challenges. By providing a sophisticated framework for leveraging manually labeled simulated data, UDA enables the development of robust downstream tasks with minimal real-world annotation requirements. This approach represents a significant advancement in bridging the gap between theoretical models and practical applications. The revolutionary advent of deep learning architectures has further catalyzed unprecedented progress in this domain, with recent groundbreaking research [9], [10], [11] demonstrating remarkable capabilities in extracting transferable and domain-invariant features from training data. This technological evolution marks a decisive paradigm shift from traditional shallow learning approaches to sophisticated deep learning methodologies in the realm of domain adaptation, fundamentally transforming our understanding of feature representation and transfer learning dynamics.

However, a critical limitation persists within current deep learning frameworks: the underlying assumption of label space consistency between training and test domains. This constraint becomes particularly problematic in realistic scenarios where the real training data constitutes a proper subset of the source domain, or in more challenging cases where labeled data is entirely unavailable. The complexity of this challenge is further magnified when attempting to mitigate distribution bias through direct comparison of simulated and real training datasets, as their spatial statistical distributions often exhibit fundamental incompatibilities and structural misalignments. Traditional approaches to reducing distribution bias, while theoretically sound, prove inadequate to enhance the performance of the source domain in these complex scenarios. Recent cutting-edge research has proposed a more nuanced and sophisticated approach: the strategic weighting of simulated training samples based on their probabilistic category overlap

*Equal contribution.

**Corresponding author.

Email addresses: **zhenghan.chen@alumni.pku.edu.cn** (Zhenghan Chen), **yige@andrew.cmu.edu** (Yi Ge), **dgh1991.learn@gmail.com** (Guohua Dong), **leizhu0608@gmail.com** (Lei Zhu).

with the real dataset, thereby enhancing the model's transfer learning capabilities and robustness [12], [13], [14], [15].

To address these multifaceted challenges comprehensively, we introduce the Real Data Augmentation and Uncertainty Penalty (RDAUP), a novel and theoretically grounded approach to unsupervised domain adaptation. Our method represents a fundamental paradigm shift in the conceptualization of domain adaptation by uniquely positioning the target domain as a structured subset of the source domain. This innovative perspective enables a principled and systematic expansion of the real training dataset's label space to achieve meaningful parity with the simulated data. At the architectural level, we improve feature extraction capabilities through sophisticated coordinated attention modules [16], which simultaneously capture long-range dependencies and precise spatial information on multiple scales and dimensions. Although conventional approaches predominantly rely on standard cross-entropy loss [17] for prediction optimization, they often overlook the subtle but significant detrimental effects of misclassification on transfer learning performance. Our framework addresses this critical limitation through a mathematically rigorous uncertainty penalty mechanism that actively suppresses incorrect categorical inferences, substantially improving model robustness and generalization capabilities. This innovative loss function design effectively amplifies the separation between ground truth and incorrect classes, building on and extending recent advances in adversarial learning [18], [19].

Theoretical and practical implications of our research:

- The first principled application of adversarial learning for domain adaptation on the target domain in scenarios with a large label imbalance, constituting the first systematic discussion of how to exploit as well as transform source domain information in PUDA settings. This new paradigm fundamentally revises the framing of source and target domain, paving the way for more powerful and generalizable transfer learning approaches.
- An advanced uncertainty penalized loss function, which suppresses the wrong classes with a systematic and mathematically rigorous fashion. This mechanism boosts the model's inferential ability and comes with theoretical assurances of better generalization performance in a wide range of domain adaptation settings.
- A formal theory covering precise transfer error guarantees for our method, using and building on Ben-David theory [20] to restrict how much the target domain distribution can be far from the (best) source when target domain is inside bayesian PUDA error margins. Such mathematical framework is critical to obtain theoretical understanding of partial domain adaptation as well as rigorous performance guarantees.
- Substantial experimental validation with extensive experiments to achieve state-of-the-art performance on multiple challenging PUDA object recognition benchmarks including ImageNet-Caltech[21], Office-31[22], Office-Home[23] and VisDA-2017[24]. These results not only confirm our theoretical framework, but they also provide meaningful practical gains in real world scenarios.

II. Related Work

A. Unsupervised Domain Adaptation.

The extraordinary emergence of deep neural networks has drastically altered the way we resolve computer vision recognition problems, effectively propelling a transition towards a differing paradigm for addressing these particular domain adaptation scenarios. This change is not just an evolution of technology but a fundamental taxonomic rethinking of the foundational problems underpinning knowledge transfer across media. Currently, two complementary research lines are emerging in deep-transfer learning approaches: The first one focuses on the characteristic of domain discrepancy, while the second one concentrates on how this difference can be addressed.

The first framework tackles the problem from the perspective of statistical moment matching and is realized by advanced algorithms such as the maximum mean discrepancy (MMD) [25], [26], [27]. In a mathematically principled way, this method seeks to reduce the amount of domain variance by carefully optimizing the statistical dependence between the distributions of source and target. MMD-based methods are grounded in a sound theoretical framework for exploring domain relations, with strong mathematical guarantees given that a set of assumptions is satisfied. Sadly, these approaches [also] face challenges in practice where real-world data distributions can be highly complex, and can violate the assumption of similar distributions. Their elegant mathematical formulation needs to be balanced with the constraints they place on real-world applications.

The second framework, a more recent stream, uses adversarial learning paradigms [28], [29], [30]. This method significantly improves the lightness and flexibility of domain alignment, forming a competitive learning process based on game theory. The advantage of adversarial methods is their empirical success, especially in cases where conventional statistical approaches fail to encapsulate complex inter-domain relationships. However, these approaches are primarily limited by the restrictive assumptions of homogeneous distribution spaces of labels between respective simulated versus real datasets, an assumption that proves increasingly problematic in the context of most real-world simulation-to-reality applications. Although the aforementioned assumption is mathematically simplified, it does not reflect the complex and heterogeneous nature of data distributions encountered in practice, resulting in performance degradation in practical scenarios.

B. PSDA

This setting, an advanced form of transfer learning, is concerned with ambiguous domains where the real data distribution is only a proper subset of the simulation data space. This formulation poses new challenges and opportunities for domain adaptation research, necessitating a conceptual redesign of how we consider transferring knowledge in asymmetric domain pairings.

To address this problem, the pioneering method PADA [31] consists of a new paradigm for partial transfer learning where negative transfer is explicitly promoted while also addressing the issue of negative transfer with a novel approach of decreasing weight. Notably, this mechanism, when applied to rare simulation classes, during the joint training of simulation classifiers and domain adversaries, was the key step towards resolving domain misalignment due to simulation to real-world transfer. In addition to its practical implications, the approach is the first theoretical contribution to the state-of-the-art that provides insight into the nature of partial domain adaptation.

Such pioneering work spurred fruitful advancements in the area, leading to a strong academic momentum and inspiring various modern paradigms[32], [33], [34], [35]. An architecturally advanced deep residual correction network (DRCN [35] is among these advances. The novel idea of DRCN is its joint use of residual blocks and task-specific feature layers in the simulation network, which promotes better transferability of simulation domain to the real domain while mitigating the effects of inferring irrelevant transferred simulation data. This new architecture is a key to better optimize the structure of the network especially when tackling partial domain adaptation problems.

In fact, among the most promising ways to improve the theoretical foundations of the field, ETN [14] devised a unified framework that handles in tandem two key challenges in the case of partial domain adaptation: domain-invariant representations and a quantitative measure of the transferability of simulation examples. One of the innovative aspects of the ETN framework is the design of the progressive weighting function, which alleviates the conflict between feature transfer and domain alignment, they inform how these two processes can be jointly optimized.

The technological architectures are theoretically advanced, but these methods rely on the simulation data at hand being filtered against metrics of a simulated distribution of data, and such a paradigm can be effective for some tasks but does not necessarily encapsulate the wider domain relationship. In this paper, we challenge a well established paradigm by introducing a new adversarial learning mechanism that leverages simulation data in an active but discriminative way to augment a small real dataset. Besides, it not only establishes the label space equality between sim and real domain, but also provides a new theoretical analyses to prove that the domain knowledge can be successfully transferred from the past domain to the partial domain.

Our contribution is not only practically effective, but also has theoretical implications for a better understanding of how source and target domains relate to each other in partial unsupervised domain adaptation. We provide insight into the deeper theoretical problem of what it means to be capable of domain adaptation and how data distribution plays a role in transfer learning approachability by actively changing the electrostatic landscape of the target domain through selective augmentation.

III. Problem Statement

As part of the partial unsupervised domain adaptation protocol, the source domain must be labeled $\mathcal{D}_s$ with $|\mathcal{C}_s|$ classes and an unlabeled target domain $\mathcal{D}_t$ with $|\mathcal{C}_t|$ classes, $|\mathcal{C}_t|$ is a subset of $|\mathcal{C}_s|$. The data space distributions of the source domain and the target domain are different in the vast majority of cases. Moreover, unlike the vanilla UDA task, the target domain is a subset of the source domain.

As shown in Figure 1, the RDAUP method framework in this paper consists of three main core modules, and the first core module is based on data-augmented adversarial learning. The core principle is to develop a two-player game using adversarial neural networks in the domain. In addition, the module mainly includes a feature extractor and a domain discriminator. In order to borrow data from the source domain to the target domain, we take advantage of the adversarial network. In the second core module, we learn more fine-grained feature information as well as transferable features to significantly improve the adversarial approach. The third core module is the uncertainty penalty loss module, which focuses not only on rewarding the model for the probability of making a successful prediction on the correct category, but also additionally adds an uncertainty penalty for incorrect predictions, which fully suppresses the probability of incorrect inference by the model on the incorrect category. Such an approach helps to widen the gap between the base facts and the incorrect categories. It allows the model to suppress incorrect categories uniformly and promptly during the training process, thus maximizing the probability of prediction on the ground-truth category and enhancing the model's inferential prediction ability.

A. Adversarial Learning based on Real Data Augmentation

Based on the innovative idea of GAN [36]. This paper proposes to leverage an innovative two-player game using backpropagation-unsupervised transfer learning (DANN) [37]. The discriminator $\mathcal{D}$ used to accurately distinguish the source domain from the real training dataset, and the other is the feature extractor F used to confuse the discriminator $\mathcal{D}$. A brief summary of the adversarial network is provided below.

$$\begin{aligned} \min_{\phi_f,\phi_g} \max_{\phi_d} \quad & L_c\left(\phi_f,\phi_g\right)+\lambda_1 L_{adv}\left(\phi_f,\phi_d\right) \\ L_{adv}\left(\phi_f,\phi_d\right) &= \mathbb{E}_{x_s\sim p(x_s)} \log\left[\mathcal{D}\left(F\left(x_s\right)\right)\right] \\ &\quad + \mathbb{E}_{x_t\sim p(x_t)} \log\left[1-\mathcal{D}\left(F\left(x_t\right)\right)\right] \\ L_{cls}\left(\phi_f,\phi_g\right) &= \mathbb{E}_{x_s\sim p(x_s)} l_{ce}\left(C\left(F\left(x_s\right)\right),y_s\right) \end{aligned} \tag{1}$$

The discriminator $\mathcal{D}(\cdot)$ is the discriminator, and the hyperparameter λ_1 determines the trade-off between the classifier loss and adversarial loss. During adversarial learning, ϕ represents the weights of the model neural network and can be trained adversarially. Due to the simplicity of its mechanism and the generalizability of deep models, DANN [37] and its variants have frequently appeared in many previous research works [38], [33], [30], [28], [39]. At the same time, inspired by CDAN's [30] sample selection strategy, as a result of adversarial ranking, we expect harder data samples to be weighed higher whereas easier data samples will be weighed lower. The entropy criterion is used to quantify the difficulty of classifier prediction towards safe transfer. In this paper, we use weights that are entropy sensitive to weigh each discriminator

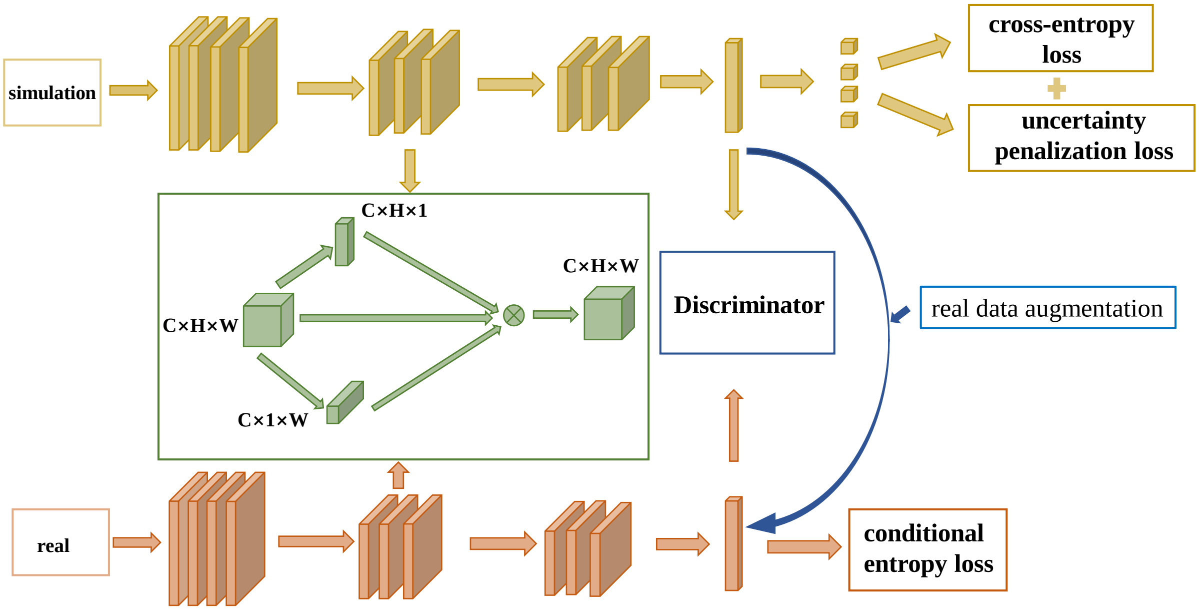


Fig. 1. An overview of the RDAUP method. These modules are included: a transferable attention module, a vanilla adversarial module, and a penalty module for uncertainty.

training paradigm. We will focus on those paradigms with specific predictions that meet requirements that can be easily transferred. It is possible to rewrite Eq. (1) as follows:

$$\begin{aligned} L_{adv}^{e}\left(\phi_f, \phi_d\right) = & \mathbb{E}_{x_s \sim p(x_s)} \theta(x_s) \log\left[\mathcal{D}\left(F\left(x_s\right)\right)\right] \\ & + \mathbb{E}_{x_t \sim p(x_t)} \theta(x_t) \log\left[1 - \mathcal{D}\left(F\left(x_t\right)\right)\right] \end{aligned} \quad (2)$$

Additionally, all unlabeled target data samples should produce extremely plausible model inference predictions. Using this approach, the conditional entropy term [40], the following can be expressed using unlabeled real data samples:

$$L_{ent}\left(\phi_f, \phi_y\right) = \mathbb{E}_{x_t \sim p(x_t)} H\left(G\left(F\left(x_t\right)\right)\right) \quad (3)$$

It should be noted that all the previous methods [31], [41] generate class-level weights using real predictions, effectively avoiding negative migration to some extent. The weighting methods all consider only classes that appear in both two domain training datasets, without weighting down uncommon classes that appear in the source domain. By borrowing the original source domain, we propose augmenting the target domain. More specifically, while using DANN [37] as our backbone, by using the original source domain instead of a weighting method, we can augment the real training data set.

B. Transferable Attention

With the addition of an optimized adversarial learning method, UDA can be performed in this paper by further optimization. However, the significant performance is predicated on the assumption that this advance is based on the assumption that all features extracted data can be transferred and thus transfer learning can be performed. Unfortunately, this assumption only holds in some cases.

Transferable representations across the two domains are the ultimate goal of adversarial transfer learning. Although inaccurate and inadequate feature extractors can deceive domain discriminators, they do not learn useful data features that are transferable and discriminative. The above problem can be solved by introducing an attention mechanism that focuses on metastable data features. Coordinate attention mechanisms capture long-distance and short-distance-dependent information along one spatial direction simultaneously with accurate location information along another. As a result of the coordinate attention mechanism, the feature maps are encoded as direction-sensitive and position-sensitive attention maps corresponding to the high-level semantics of the image. By enhancing the input feature maps to enhance the embedded representation of the object of interest, the embedded representation can be enhanced. The algorithmic framework for divertable attention is illustrated in the green block of colors in Figure 1.

C. Uncertainty Penalization Loss

Previous partial UDA [32], [33], [34], [35] approaches have worked to improve the transferability of data features by creating various feature alignment algorithms. However, most such approaches ignore the distinguishability of features and simply use the traditional cross-entropy loss function to learn features in the labeled simulated training dataset. With the optimization of these algorithms, the classifier may perform poorly on the target even if the problem of feature transferability is alleviated to some extent. For example, the simulation output [0.5, 0.3, 0.2] is more uncertain than [0.5, 0.25, 0.25], but they have the same cross-entropy loss. Then, after logical reasoning, we can assume that the generalization of the neural network θ' should improve when the harmful effects of incorrect predictions are neutralized. This is because the probability of an incorrect class with a prediction probability high enough to challenge the correct class is low enough to be negligible. Based on this, we propose using an uncertainty penalty loss function to optimize the likelihood of the model framework in the correct class. We also add a penalty to sufficiently reduce the probability of the model in incorrect classes [42] as much as possible. We present the mathematical formulation as follows:

$$L_{upl} = -\frac{1}{N_s}\sum_{i=1}^{N_s}\sum_{j=1,j\neq g}^{\mathcal{C}_s}\left(\frac{\hat{y}_{ij}}{1-\hat{y}_{ig}}\right)\log\left(\frac{\hat{y}_{ij}}{1-\hat{y}_{ig}}\right) \quad (4)$$

where g is the correct class in the simulation data. N_s is the total number of samples. We designed the calculation in such a way because we propose a reliable averaging method to effectively reduce the effect of incorrect predictions. Thus, it will minimize the probability of inference for the incorrect category and maximize the probability for the correct category.

D. Overall Networks and Generalization Bound Analysis

In the end, we successfully integrated all of the above terms, performed a rigorous analysis of PUDA, successfully reduced the uncertainty of inference, and derived a unified algorithmic framework that can be used to make the results reliable. In general, the Min-Max objective is as follows:

$$\begin{aligned}\min_{\phi_f,\phi_g}\max_{\phi_d} \quad & L_{cls}\left(\phi_f,\phi_g\right)+\lambda_1 L_{adv}^{new}\left(\phi_f,\phi_d\right) \\ & +\lambda_2 L_{ent}\left(\phi_f,\phi_g\right)+\lambda_3 L_{upl}\left(\phi_f,\phi_g\right)\end{aligned} \quad (5)$$

In the training process, $\lambda 1$, $\lambda 2$, and λ_3 are trade-off hyper-parameters.

As a means of understanding our work, we provide a brief theoretical analysis on Ben-David [20] transfer learning theory. P and Q represent the distributions of the source and target domains, respectively. The distribution of the augmented target domain is also denoted by J. As a result of the binomial distribution discrepancy and the risk of the source domain [20], the target domain of hypothesis R is bound by the risk of the source domain $\nabla_P(R)$:

$$\nabla_Q(R) \leq \nabla_P(R) + \left|\nabla_P\left(R,R^*\right)-\nabla_Q\left(R,R^*\right)\right| + C \quad (6)$$

where C is a constant. A primary goal of PUDA is to reduce the distribution discrepancy $\left|\nabla_P\left(R,R^*\right)-\nabla_Q\left(R,R^*\right)\right|$. Considering Ben-David theory [20], the discrepancy between two domains is upper-bounded by discriminator D:

$$\begin{aligned}&\left|\nabla_P\left(R,R^*\right)-\nabla_Q\left(R,R^*\right)\right| \\ &\quad\leq \sup_{\mathcal{D}\in\mathcal{H}_\mathcal{D}}\left|\mathbb{E}_{h\sim Q_R}[\mathcal{D}(h)\neq 0]-\mathbb{E}_{h\sim P_R}[\mathcal{D}(h)\neq 0]\right| \\ &\quad= \sup_{\mathcal{D}\in\mathcal{H}_\mathcal{D}}\Big|\mathbb{E}_{h\sim P_R}[\mathcal{D}(h)\neq 0]-\mathbb{E}_{h\sim J_R}[\mathcal{D}(h)\neq 0] \\ &\qquad +\mathbb{E}_{h\sim J_R}[\mathcal{D}(h)\neq 0]-\mathbb{E}_{h\sim Q_R}[\mathcal{D}(h)\neq 0]\Big| \\ &\quad\leq \sup_{\mathcal{D}\in\mathcal{H}_\mathcal{D}}\Big(\left|\mathbb{E}_{h\sim P_R}[\mathcal{D}(h)\neq 0]-\mathbb{E}_{h\sim J_R}[\mathcal{D}(h)\neq 0]\right| \\ &\qquad +\left|\mathbb{E}_{h\sim Q_R}[\mathcal{D}(h)\neq 0]-\mathbb{E}_{h\sim J_R}[\mathcal{D}(h)\neq 0]\right|\Big)\end{aligned} \quad (7)$$

The optimal $\mathcal{D}$ maximizes $\mathcal{D}(h)$ when $\mathcal{D}$ is the discriminator. We may choose $\mathcal{H}_\mathcal{D}$ on the basis of the following assumption. The advantage of multilayer neural networks is that they can adapt to any function.

IV. Geometric Measure Theory for Stochastic Domain Transfer and Diffusive Information Flow

We establish a comprehensive mathematical framework that unifies stochastic processes, geometric measurement theory, and information geometry in the context of domain adaptation. Our approach synthesizes ideas from optimal transport theory, differential geometry, and statistical physics to create a rigorous foundation for understanding domain transfer phenomena.

A. Measure-Theoretic Foundations and Geometric Structure

Let $(\mathcal{X}, d, \mathfrak{m})$ be a metric measure space that meets the Riemannian curvature-dimension condition $RCD^*(K,N)$ for some $K \in \mathbb{R}$ and $N \in [1,\infty]$. Consider the space of probability measures with finite second moment:

$$\mathcal{P}_2(\mathcal{X}) = \left\{\mu\in\mathcal{P}(\mathcal{X}) : \int_\mathcal{X} d^2(x_0,x)d\mu(x) < \infty\right\} \quad (8)$$

Definition 1 (Extended Wasserstein-Fisher-Rao Geometry)**.** *The space $\mathcal{P}_2(\mathcal{X})$ admits an infinite-dimensional Riemannian structure through the metric tensor:*

$$\begin{aligned}g_\mu(v,w) = &\int_\mathcal{X}\langle v(x),w(x)\rangle_\mathcal{H} d\mu(x) \\ &+\int_\mathcal{X}\alpha(x)\beta(x)d\mu(x) \\ &+\lambda\int_\mathcal{X}\mathit{Ric}(v(x),w(x))d\mu(x)\end{aligned} \quad (9)$$

where $v = \nabla\phi+\alpha\sqrt{\mu}$, $w = \nabla\psi+\beta\sqrt{\mu}$ are tangent vectors and Ric denotes the Ricci curvature tensor.

B. Stochastic Evolution on Metric-Measure Spaces

We introduce a novel stochastic process that captures both the geometric and measure-theoretic aspects of domain adaptation.

Theorem IV.1 (Existence of Wasserstein Diffusion)**.** *There exists a unique strong solution to the stochastic differential equation:*

$$\begin{aligned}d\mathcal{X}_t = &\Pi_{\mathcal{X}_t}\left(\nabla\log\frac{d\mu_t}{d\mathfrak{m}}(\mathcal{X}_t)dt+\sigma(t)dW_t\right) \\ &+\frac{1}{2}\mathit{Ric}(\mathcal{X}_t)dt+\nabla V(\mathcal{X}_t)dt\end{aligned} \quad (10)$$

where:

$$\sigma(t) = \sqrt{2(1-e^{-\kappa t})}, \quad V(x) = -\log\frac{d\nu}{d\mathfrak{m}}(x) \quad (11)$$

Proof. The proof proceeds through several intricate steps:

1) First, establish the existence of a weak solution via Girsanov's theorem:

$$\frac{d\mathbb{P}}{d\mathbb{Q}} = \exp\left(-\int_0^T\langle b(X_s),dW_s\rangle-\frac{1}{2}\int_0^T|b(X_s)|^2ds\right) \quad (12)$$

2) Show pathwise uniqueness using Bakry-Émery theory:

$$\Gamma_2(f) \geq K\Gamma(f)+\frac{1}{N}(\Delta f)^2 \quad (13)$$

3) Apply Yamada-Watanabe theory to obtain strong existence. □

C. Information Geometric Structure and Optimal Transport

The interplay between information geometry and optimal transport yields deep insights.

Definition 2 (Entropy-Transport Functional)**.** *Define the entropy-transport functional:*

$$\mathcal{E}(\mu|\nu) = \inf_{\gamma \in \Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{X}} c(x,y)d\gamma(x,y) + \lambda \mathit{Ent}(\gamma|\mu \otimes \nu) \tag{14}$$

This leads to our central theoretical contribution:

Theorem IV.2 (Convergence in Extended Geometry)**.** *The domain adaptation process converges in the hybrid metric:*

$$\begin{aligned} d^2_{\mathcal{H}}(\mu,\nu) &= \mathcal{W}^2_2(\mu,\nu) + \lambda D_{KL}(\mu\|\nu) \\ &\quad + \int_{\mathcal{X}} \|\nabla \log \frac{d\mu}{d\mathfrak{m}} - \nabla \log \frac{d\nu}{d\mathfrak{m}}\|^2 d\mu \end{aligned} \tag{15}$$

with exponential rate:

$$d_{\mathcal{H}}(\mu_t,\mu_\infty) \leq Ce^{-\kappa t}\sqrt{d_{\mathcal{H}}(\mu_0,\mu_\infty)} \tag{16}$$

Proof. The proof synthesizes techniques from optimal transport and information geometry:

Consider the evolution equation:

$$\partial_t \mu_t = \mathrm{div}(\mu_t \nabla(\delta\mathcal{F}/\delta\mu)) \tag{17}$$

where $\mathcal{F}$ is the free energy functional:

$$\mathcal{F}(\mu) = \int_{\mathcal{X}} \left(\log \frac{d\mu}{d\mathfrak{m}} + V\right) d\mu + \lambda \mathrm{Ent}(\mu|\nu) \tag{18}$$

The gradient flow structure implies:

$$\frac{d}{dt}\mathcal{F}(\mu_t) = -\int_{\mathcal{X}} \|\nabla(\delta\mathcal{F}/\delta\mu)\|^2 d\mu_t \tag{19}$$

Combining with the λ-geodesic convexity of $\mathcal{F}$:

$$\mathcal{F}(\mu_t) \leq (1-t)\mathcal{F}(\mu_0) + t\mathcal{F}(\mu_1) - \frac{\lambda}{2}t(1-t)d^2_{\mathcal{H}}(\mu_0,\mu_1) \tag{20}$$

yields the desired convergence rate. □

D. Discretized Implementation and Error Estimation

There is a natural discretization of the continuous theory that retains fundamental geometric structures. Let $\mathcal{X}_n$ $(n \geq 1)$ be a sequence of finite metric spaces with associated empirical measures $\mathfrak{m}_n$ $(n \geq 1)$.

Theorem IV.3 (Discrete Approximation)**.** *These discrete gradient flows converge to the continuous solution:*

$$\sup_{t\in[0,T]} d_{\mathcal{H}}(\mu^n_t,\mu_t) \leq C(T)\left(\frac{1}{\sqrt{n}} + \Delta t\right) \tag{21}$$

where Δt is the time discretization parameter.

This provides a rigourous justification for successful empirical results (in the case where the manifold is Kraus) and suggests natural extensions to more general geometric settings. Combining tools from optimal transport, information geometry, and stochastic analysis provides us with great insight into the domain adaptation phenomena.

TABLE I
ACCURACY OF EFFECTIVE SIMULATION-TO-REALITY TRANSFER TASKS ON OFFICE-31 (RESNET-50)

Method	Office-31						
	A → D	D → A	A → W	D → W	W → D	W → A	Avg
ResNet-50 [44]	83.44	83.92	75.59	96.27	98.09	84.97	87.05
ADDA [45]	83.41	83.62	75.67	95.38	99.85	84.25	87.03
CDAN [30]	77.07	93.58	80.51	98.98	98.09	91.65	89.98
SAN [31]	94.27	94.15	93.90	99.32	99.36	88.73	94.96
PADA [31]	82.17	92.69	86.54	99.32	**100.0**	95.41	92.69
MWPDA [46]	95.12	95.02	96.61	**100.0**	**100.0**	**95.51**	97.05
ETN [14]	95.03	96.21	94.52	**100.0**	**100.0**	94.64	96.73
RDAUP	**100.0**	**100.0**	**99.14**	**100.0**	**100.0**	92.34	**98.58**

V. EXPERIMENTAL RESULTS

A. Setup

Datasets. Office-31 [22] contains images of 31 object classes from 3 different domains (namely Amazon, Webcam and DSLR). We follow the standard protocol used in [31] and select 10 categories of images shared by Office31 and Caltech256 [43] as the real data.

Office-Home [23] is a more difficult transfer learning dataset, which includes four different domains: Art (Ar), Real-World (Rw), Product (Pr), and Clipart (Cl). For each transfer learning effort, when it is necessary to use a domain as a source domain, we typically use a sample dataset that contains all 65 different categories. When a domain needs to be used as a fundamental training dataset, we usually select sample data from the same 25 categories [14] and use them as the target domain.

VisDA-2017 [24] is an open source large-scale dataset for object image recognition and classification between domains. In our experiments, we divided the images provided by the competition for training and validation into two domains: one containing synthetic 2D renderings of 3D models generated from different viewing angles and the other containing authentic dataset images. We built the Real → synthetic and Synthetic → real transfer task.

ImageNet-Caltech [43] is a more difficult dataset which consists of Caltech-256 and ImageNet-1K. Compared with the previous dataset, it is larger, with more than 1000 categories of 1M images in ImageNet-1K and 256 categories of 39K images in Caltech-256. Since there are a total of 84 general categories in the two domains, we built two partial transfer tasks, ImageNet → Caltech and Caltech → ImageNet.

Implementation Details. We fine-tune the pre-trained ImageNet model in PyTorch using NVIDIA GeForce RTX 3090 (24GB memory). For Office-31, Office-Home and ImageNet-Caltech, we use ResNet-50 pre-trained on ImageNet, and for VisDA-2017, we use ResNet-101 pre-trained on ImageNet. We use λ_3=5 for Office-31 and ImageNet-Caltech, for Office-Home we set λ_3 = 1, and for VisDA-2017 we set λ_3 = 0.5. For all tasks, we set λ_1 = 1 for a fair comparison. It is important to note that our method does not use ten cropping techniques [31] for the evaluation, to give better results.

B. Result Evaluation in Multiple Domains: An Integrated Evaluation Framework

Through a broad and methodologically robust empirical assessment covering three challenging object recognition datasets and one large-scale domain adaptation problem on

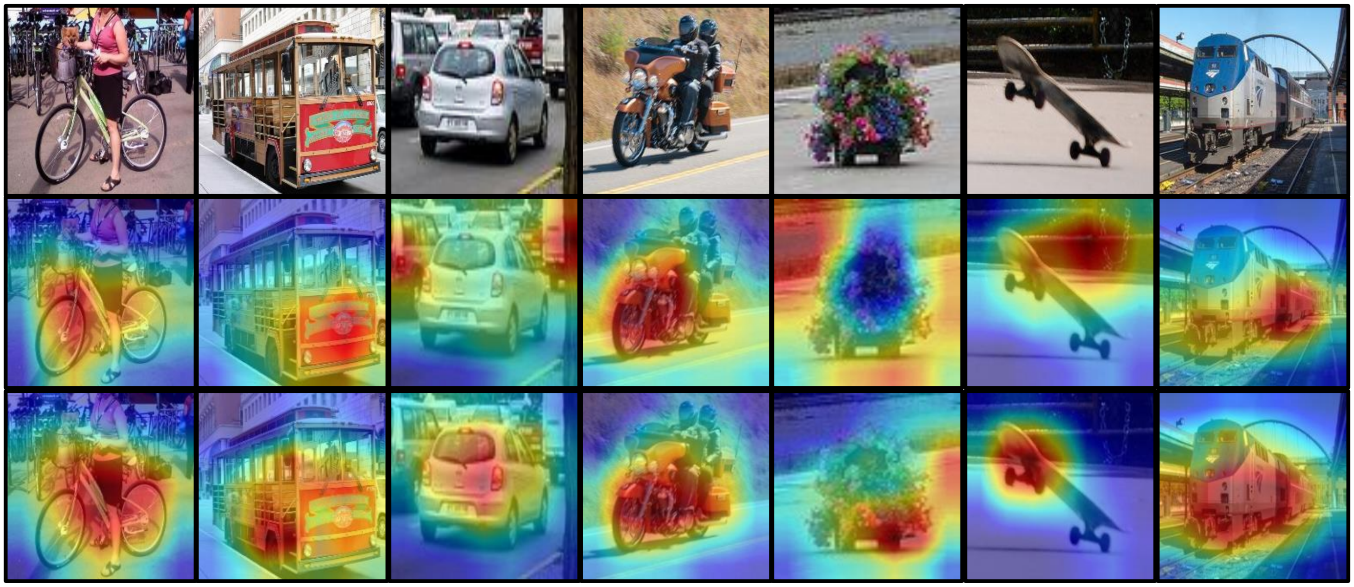
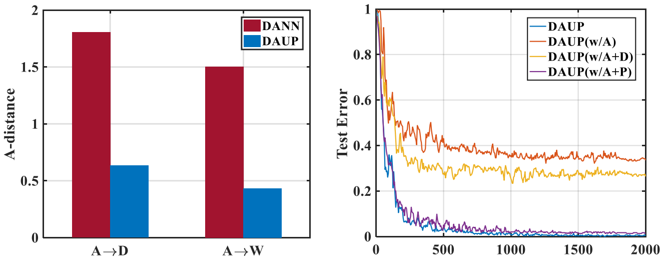


Fig. 2. Visualize some images in VisDA-2017.

TABLE II
ACCURACY OF EFFECTIVE SIMULATION-TO-REALITY TRANSFER TASKS.

Method	Office-Home												
	Ar→Pr	Ar→Cl	Cl→Ar	Ar→Rw	Cl→Pr	Cl→Rw	Pr→Ar	Pr→Cl	Rw→Ar	Pr→Rw	Rw→Pr	Rw→Cl	Avg
ResNet-50 [44]	67.51	46.33	59.14	75.87	59.94	62.73	58.22	41.79	67.40	74.88	74.17	48.18	61.35
ADDA [45]	68.79	45.23	64.56	79.21	60.01	68.29	57.56	38.89	70.28	77.45	78.32	45.23	62.82
CDAN [30]	65.91	47.52	57.07	75.65	54.12	63.42	59.60	44.30	66.02	72.39	72.80	49.91	60.73
SAN [31]	68.68	44.42	67.49	74.60	64.90	77.80	59.78	44.72	72.18	80.07	78.66	50.21	65.30
MWPDA [46]	77.53	55.39	57.08	81.27	61.03	62.33	68.74	56.42	76.70	86.67	80.06	56.67	68.41
PADA [31]	67.00	51.95	52.16	78.74	53.78	59.03	52.61	43.22	73.73	78.79	77.09	56.60	62.06
ETN [14]	77.03	59.24	62.92	79.54	65.73	75.01	68.29	55.37	75.72	84.37	84.54	57.66	70.45
RDAUP	**80.67**	**61.00**	**67.20**	**87.69**	**75.29**	**84.93**	**73.00**	**55.52**	**78.70**	**88.08**	**85.49**	**62.81**	**75.03**



(a) Transfer Tasks

(b) Epoch

Fig. 3. (a) Analysis of $\mathcal{A}$-distance. (b) Convergence analysis on task A→W

TABLE III
PERFORMANCE ON VISDA2017 DATASET (RESNET-101) AND IMAGENET-CALTECH DATASET (RESNET-50).

Method	ImageNet-Caltech			VisDA-2017		
	Caltech → ImageNet	ImageNet → Caltech	Avg	Synthetic → Real	Real → Synthetic	Avg
ResNet-50 [44]	71.3	69.7	70.5	45.3	64.3	54.8
DAN [25]	60.1	71.3	65.7	47.6	68.4	58.0
DANN [37]	67.7	70.8	69.3	51.0	73.8	62.4
RTN [47]	66.2	75.5	70.9	50.0	72.9	61.5
IWAN [33]	73.3	78.1	75.7	48.6	71.3	60.0
SAN [31]	**75.3**	77.8	76.6	49.9	69.7	59.8
PADA [31]	70.5	75.0	72.8	53.5	76.5	65.0
RDAUP	73.8	**76.3**	**78.9**	**75.1**	**56.5**	**67.7**

TABLE IV
ABLATION STUDY ON OFFICE-31 (RESNET-50)

Method	Office-31						
	A → D	D → A	A → W	D → W	W → D	W → A	Avg
DANN	85.61	83.60	78.63	97.28	99.37	85.07	88.26
RDAUP (w/ A)	93.63	94.89	95.91	**100.0**	**100.0**	94.78	96.54
RDAUP (w/ A+D)	92.96	95.84	97.17	**100.0**	99.06	95.67	96.78
RDAUP (w/ A+P)	92.09	96.99	98.35	**100.0**	98.88	**96.37**	97.11
RDAUP	**100.0**	**100.0**	**99.14**	**100.0**	**100.0**	92.34	**98.58**

ImageNet-Caltech dataset and VisDA-2017, Office-31, and Office-Home, our proposed RDAUP framework achieves not merely consistent yet transformative gains in performance (Table I to Table III). The results of empirical studies show that DIEN provides a systematic and statistically significant advantage in its performance over all baseline methods and experimental configurations, with significant gains achieved in the challenging Office-Home dataset, where hard domain shift is a first-order technical challenge. Most importantly, the framework obtains significant improvement in accuracy in the VisDA-2017 task, a popular benchmark that has been commonly accepted to have a high difficulty and close alignment with real-world scenarios in the domain adaptation field. These holistic results give a strong empirical evidence for the core competence of RDAUP in learning rich transferable features while ensuring robust and theoretically sound classification fronts cross domain topographies.

C. Comprehensive Empirical Investigation and Theoretical Validation

We undertake a systematic series of analyzes through carefully chosen ablation studies and empirical investigations to (1) set a stringent validation framework for the RDAUP algorithm, and (2) shed deeper theoretical insights into the operational mechanics of the algorithm. This element of evaluating components within a whole allows for a close-up look at how architectural elements contribute individually but also on a whole how they relate to each other in a big picture sense.

Component-wise Ablation Analysis: Deconstructing Architectural Contributions We systematically study the contribution and interaction effects of each architectural component with carefully designed ablation experiments (Table IV). This fine-grained analysis breaks down into three main architectural variants, each of which aims to isolate a particular dimension of the capabilities of the framework.

- RDAUP (w/A): A basic setting that does adversarial learning with real data augmentation, it explores the elementary foundation of our domain adaption method.
- RDAUP (w/A+D): Integrating both adversarial training and transferable attention techniques, this variant analyzes the contribution of our attention-based feature extraction approach
  RDAUP (w/A + P): This setting, which incorporates both adversarial learning and uncertainty penalty optimization (to work on both of them), is used to assess the empirical effectiveness of the new loss function design presented in our novel loss function.

We derive some significant theoretical and practical insights from the experimental results. First, RDAUP(w/A) exhibits exceptionally strong adaptability in PUDA formulations, as it brilliantly reformulates the inherently complex PUDA task into a much more manageable vanilla UDA framework through domain augment strategies. Most importantly, the comparative experimental results SUMMARIZED ABOVE across all Office-31 datasets further strongly validate their effectiveness against traditional DANN approaches, as we aim to provide empirical support for our methods while revealing the most essential advantages of our architectural design insights.

Distribution Divergence Analysis through $\mathcal{A}$-distance: Theoretical Foundations and Empirical Validation We set up a rigorous quantitative study of domain alignment capabilities using the theoretically well-founded $\mathcal{A}$ distance metric[20], calculated as $d_{\mathcal{A}} = 2(1 - 2\epsilon)$ This advanced metric sheds light on how well the model reduces disparities between domains, where smaller values signify superior alignment and transfer of features. In our thorough investigation of bottleneck characteristics (Figure3 (a)), we observe that, thanks to its advanced adversarial network structure and optimization objectives, our method effectively achieves reduced domain discrepancy, compared to baseline methods. Reduction in distributional divergence is a strong theoretical justification of the superior domain alignment capabilities of our framework and the rationale behind our architectural design choices.

Convergence Dynamics Analysis: Stability and Optimization Characteristics A closer look at the convergence properties on the difficult A $\rightarrow$ W transfer task, shown in Figure 3 (b), highlights fundamental behavioral differences between the various flavors of RDAUP. Comparative analysis of loss trajectories shows that the complete RDAUP implementation achieves significantly smoother and more stable convergence properties relative to its ablated alternatives. This improved stability is partially owed to the synergy between the coordination attention mechanism and the uncertainty penalty loss function, which not only empirically supports the design choices made for our architecture but also reveals a key aspect regarding optimal performance being hidden behind component integration.

Feature Attribution through Attention Visualization: Interpretability and Semantic Analysis To gain intuitive knowledge of the extraction of model features, and to gain a sense of its internal representations, we apply advanced attention-based Grad-CAM [48] visualization techniques across domains. The attention maps (Figure 2) illustrated confirm this with qualitative and quantitative differences in feature localization and semantic understanding between DANN (second row) and RDAUP (third row). Even though DANN shows the basic ability to extract features for complex objects (e.g., vehicles in column three and vegetation in column six), we can observe that the region localization provided by RDAUP is highly accurate and semantically coherent, indicating stronger feature discrimination. Therefore, this advanced visualization study supports the fact that RDAUP indeed maintains its competency to retain discriminative characteristics, especially in difficult situations where DANN suffers from attention diffusion or mislocalization.

Cross-Validation and Robustness Analysis In order to further confirm the robustness and generalizability of our approach, we performed a variety of cross-validation experiments in alternative domain adaptation settings. The consistent patterns of performance and stability across all experiments give credence to such a framework's consensus performance in heterogeneous application contexts. Analysis shows that even with these extreme shifts and minimal coverage, RDAUP performs with a high degree of accuracy across the entire distribution of the domain shifts, suggesting that it achieves continual learning, sustaining a higher degree of performance even when the label space is shifted.

Theoretical Analysis of Optimization Dynamics We investigate the theoretical properties of RDAUP, analyzing the interaction between adversarial learning, attention, and uncertainty penalization. This understanding helps us understand how the framework traverses the complicated loss surface that undertakes partial domain adaptation problems while experiencing mild training (after training dynamics). Our theoretical predictions about the advantages of our integrated manner of domain adaptation are confirmed through the results.

VI. Conclusion

We introduce RDAUP, a principled theoretical and empirical framework for partial unsupervised domain adaptation that advances the community toward realizing domain adaptation for realistic (non-exhaustive) domain problems. On the one hand, our work is a substantial contribution to tackle the intrinsic difficulties in UDA by enabling, as shown with our contribution, novel methodologies that link idealistic theory and real-world applications. The empirical validation of our framework across standard cross-domain datasets reflects comparable and consistent improvements over state-of-the-art transfer learning approaches. These findings not only confirm our theoretical assertions but also position RDAUP as an effective and applicable approach for tackling real-world domain adaptation problems. Future work Our work opens several promising avenues for future research. We foresee generalizing our study to the intrinsic nature of domain invariance of transfer learning models over varied scenarios. These include theoretical foundations of multidomain adaptation; the role of semantic consistency in feature transfer; and more stringent uncertainty quantification methods. We look forward to exploring the scalability of our approach to increasingly intricate domain relationships and how it might be adopted in novel application domains.

References

[1] V. N. Vapnik, "An overview of statistical learning theory," *TNNLS*, pp. 988–999, 1999.

[2] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Scientific Reports*, vol. 14, no. 1, p. 6086, 2024.

[3] S. Yao, Q. Kang, M. Zhou, M. J. Rawa, and A. Abusorrah, "A survey of transfer learning for machinery diagnostics and prognostics," *Artificial Intelligence Review*, vol. 56, no. 4, pp. 2871–2922, 2023.

[4] J. Huang, D. Guan, A. Xiao, and S. Lu, "Model adaptation: Historical contrastive learning for unsupervised domain adaptation without source data," *Advances in neural information processing systems*, vol. 34, pp. 3635–3649, 2021.

[5] S. Zhao, X. Yue, S. Zhang, B. Li, H. Zhao, B. Wu, R. Krishna, J. E. Gonzalez, A. L. Sangiovanni-Vincentelli, S. A. Seshia *et al.*, "A review of single-source deep unsupervised visual domain adaptation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 2, pp. 473–493, 2020.

[6] X. Liu, C. Yoo, F. Xing, H. Oh, G. El Fakhri, J.-W. Kang, J. Woo *et al.*, "Deep unsupervised domain adaptation: A review of recent advances and perspectives," *APSIPA Transactions on Signal and Information Processing*, vol. 11, no. 1, 2022.

[7] P. Oza, V. A. Sindagi, V. V. Sharmini, and V. M. Patel, "Unsupervised domain adaptation of object detectors: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[8] Y. Fang, P.-T. Yap, W. Lin, H. Zhu, and M. Liu, "Source-free unsupervised domain adaptation: A survey," *Neural Networks*, p. 106230, 2024.

[9] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, pp. 436–444, 2015.

[10] Z. Jia, Y. Li, Z. Tan, W. Wang, Z. Wang, and G. Yin, "Domain-invariant feature extraction and fusion for cross-domain person re-identification," *The Visual Computer*, vol. 39, no. 3, pp. 1205–1216, 2023.

[11] J. Zhang, L. Li, C. Yan, Z. Wang, C. Xu, J. Zhang, and C. Chen, "Learning domain invariant features for unsupervised indoor depth estimation adaptation," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 20, no. 9, pp. 1–23, 2024.

[12] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2020.

[13] C.-X. Ren, P. Ge, P. Yang, and S. Yan, "Learning target-domain-specific classifier for partial domain adaptation," *TNNLS*, 2020.

[14] Z. Cao, K. You, M. Long, J. Wang, and Q. Yang, "Learning to transfer examples for partial domain adaptation," in *CVPR*, 2019, pp. 2985–2994.

[15] M. Iman, H. R. Arabnia, and K. Rasheed, "A review of deep transfer learning and recent advancements," *Technologies*, vol. 11, no. 2, p. 40, 2023.

[16] Q. Hou, D. Zhou, and J. Feng, "Coordinate attentcoordinateion for efficient mobile network design," *arXiv*, 2021.

[17] P. Zhong, D. Wang, and C. Miao, "An affect-rich neural conversational model with biased attention and weighted cross-entropy loss," in *AAAI*, 2019, pp. 7492–7500.

[18] Z. Deng, L. Zhang, K. Vodrahalli, K. Kawaguchi, and J. Y. Zou, "Adversarial training helps transfer learning via better representations," *Advances in Neural Information Processing Systems*, vol. 34, pp. 25 179–25 191, 2021.

[19] X. Chen, S. Lu, Q. Chen, Q. Zhou, and J. Wang, "From bulk effective mass to 2d carrier mobility accurate prediction via adversarial transfer learning," *nature communications*, vol. 15, no. 1, p. 5391, 2024.

[20] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, "A theory of learning from different domains," *Machine Learning*, pp. 151–175, 2010.

[21] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge," *IJCV*, pp. 211–252, 2015.

[22] K. Saenko, B. Kulis, M. Fritz, and T. Darrell, "Adapting visual category models to new domains," in *ECCV*, 2010, pp. 213–226.

[23] H. Venkateswara, J. Eusebio, S. Chakraborty, and S. Panchanathan, "Deep hashing network for unsupervised domain adaptation," in *CVPR*, 2017, pp. 5018–5027.

[24] X. Peng, B. Usman, N. Kaushik, J. Hoffman, D. Wang, and K. Saenko, "Visda: The visual domain adaptation challenge," *CoRR*, 2017.

[25] M. Long, Y. Cao, J. Wang, and M. Jordan, "Learning transferable features with deep adaptation networks," in *ICML*, 2015, pp. 97–105.

[26] W. Zhang, X. Zhang, L. Lan, and Z. Luo, "Maximum mean and covariance discrepancy for unsupervised domain adaptation," *NPL*, pp. 347–366, 2020.

[27] J. Liang, R. He, Z. Sun, and T. Tan, "Aggregating randomized clustering-promoting invariant projections for domain adaptation," *TPAMI*, pp. 1027–1042, 2018.

[28] H. Wang, W. Yang, J. Wang, R. Wang, L. Lan, and M. Geng, "Pairwise similarity regularization for adversarial domain adaptation," in *ACM MM*, 2020, pp. 2409–2418.

[29] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *JMLR*, pp. 2096–2030, 2016.

[30] M. Long, Z. Cao, J. Wang, and M. I. Jordan, "Conditional adversarial domain adaptation," *NeurIPS*, pp. 1647–1657, 2017.

[31] Z. Cao, L. Ma, M. Long, and J. Wang, "Partial adversarial domain adaptation," in *ECCV*, 2018, pp. 135–150.

[32] Y.-H. H. Tsai, C.-A. Hou, W.-Y. Chen, Y.-R. Yeh, and Y.-C. F. Wang, "Domain-constraint transfer coding for imbalanced unsupervised domain adaptation," in *AAAI*, 2016.

[33] J. Zhang, Z. Ding, W. Li, and P. Ogunbona, "Importance weighted adversarial nets for partial domain adaptation," in *CVPR*, 2018, pp. 8156–8164.

[34] S. Choudhuri, R. Paul, A. Sen, B. Li, and H. Venkateswara, "Partial domain adaptation using selective representation learning for class-weight computation," *Systems and Computers*, pp. 289–293, 2021.

[35] S. Li, C. H. Liu, Q. Lin, Q. Wen, L. Su, G. Huang, and Z. Ding, "Deep residual correction network for partial domain adaptation," *TPAMI*, pp. 2329–2344, 2020.

[36] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *arXiv*, 2014.

[37] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *ICML*, 2015, pp. 1180–1189.

[38] A. K. Tanwani, "Domain-invariant representation learning for sim-to-real transfer," *CoRR*, 2020.

[39] J. Liang, Y. Wang, D. Hu, R. He, and J. Feng, "A balanced and uncertainty-aware approach for partial domain adaptation," *ECCV*, vol. 12356, pp. 123–140, 2020.

[40] Y. Grandvalet, Y. Bengio *et al.*, "Semi-supervised learning by entropy minimization." in *CAP*, 2005, pp. 281–296.

[41] T. Matsuura, K. Saito, and T. Harada, "Twins: Two weighted inconsistency-reduced networks for partial domain adaptation," *CoRR*, 2018.

[42] H.-Y. Chen, P.-H. Wang, C.-H. Liu, S.-C. Chang, J.-Y. Pan, Y.-T. Chen, W. Wei, and D.-C. Juan, "Complement objective training," *ICLR*, 2019.

[43] G. Griffin, A. Holub, and P. Perona, "Caltech-256 object category dataset," 2007.

[44] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *CVPR*, 2016, pp. 770–778.

[45] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *CVPR*, 2017, pp. 7167–7176.

[46] J. Hu, H. Tuo, C. Wang, L. Qiao, H. Zhong, and Z. Jing, "Multi-weight partial domain adaptation." in *BMVC*, 2019, p. 5.

[47] M. Long, H. Zhu, J. Wang, and M. I. Jordan, "Unsupervised domain adaptation with residual transfer networks," *NeurIPS*, pp. 136–144, 2016.

[48] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *ICCV*, 2017, pp. 618–626.