

A Survey on Deep Domain Adaptation

Zhaoxin Li

School of Electrical and Computer Engineering, Nanfang College·Guangzhou, Guangzhou,
510900, Guangdong, China

Levi0328@outlook.com

Abstract. In the practical implementation of deep learning, challenges such as a heavy reliance on large datasets, high costs associated with data annotation, and significant computational resource consumption can lead to issues like incomplete datasets and inaccurate annotations. These factors negatively impact the model's generalization capabilities and overall performance. To address these concerns, recent research has increasingly focused on adaptive technologies. Domain adaptation seeks to transfer abundant labeled information from a source domain to an unlabeled target domain, thereby tackling the decline in machine learning model performance when faced with varying data distributions. By fine-tuning and optimizing models, it is possible to bridge gaps between different domains or scenarios effectively; this allows for leveraging knowledge from the source domain to better align with the characteristics of the target domain. Consequently, this approach reduces training and annotation costs while enhancing both accuracy and robustness in real-world applications. This review intends to thoroughly investigate the distinct contributions of two specific techniques: Joint Adversarial Domain Adaptation, and Deep Joint Distribution Optimal Transport for Unsupervised Domain Adaptation. Through this examination, we aim to uncover strategies for more effectively addressing challenges posed by limited data availability, high expenses, and substantial computing resource demands in deep learning by generating similar high-quality data that further minimizes discrepancies between source and target domains ultimately leading to marked improvements in model adaptability and generalization abilities while facilitating efficient deployment and performance enhancement of deep learning technologies across various practical application scenarios.

Keywords: Deep Domain Adaptation, Optimal Transport, Adversarial Learning.

1. Introduction

Deep learning [1-2] represents a category of machine learning techniques that employ multi-layer neural networks [3] to mimic the human brain's learning processes. By constructing a neural network with several hidden layers, computers can autonomously learn and extract abstract feature representations at multiple levels from vast amounts of data, effectively addressing complex challenges in pattern recognition, classification, and prediction. The essence of deep learning lies in its neural networks, which model intricate nonlinear relationships and execute various tasks through the collaborative functioning of nodes, layers, weights, biases, and activation functions. Despite achieving significant advancements across numerous critical domains, deep learning is still hindered by limitations related to data efficiency. Currently prevalent deep learning methods typically depend on extensive datasets that may require millions or even trillions of dollars for optimal performance. However, gathering and annotating such large volumes of data for each new task or domain proves not only costly but also practically challenging and sometimes nearly impossible to accomplish. This issue regarding data utilization efficiency [4] has emerged as a crucial factor restricting the widespread adoption of deep learning technologies across diverse application scenarios.

To tackle the challenge posed by data efficiency issues, an essential approach involves exploring the transfer capabilities inherent in deep learning [5]. Mobility a fundamental aspect of human cognition—enables us to leverage past experiences to swiftly adapt to novel situations even when faced with limited samples. In terms specific to deep learning contexts, migration refers to how deep neural networks can derive universally applicable feature representations from completed source tasks and subsequently transfer these optimized representations to relevant target tasks; this process

accelerates overall learning speed. Recent studies indicate that large-scale dataset-based deep models exhibit robust migration abilities across various downstream applications like visual detection and natural language processing. Consequently, migration has become a pivotal component in enhancing the efficiency associated with data usage in deep learning efforts. Alongside generalization capabilities as well as interpretability and robustness within models, it establishes a vital connection between machine intelligence and human cognitive processes.



Figure 1. Deep learning applications typically undergo two distinct phases: the first phase involves utilizing a Pre-Trained Model, which serves as the backbone network to learn from large-scale datasets and acquire transferable general knowledge. The second phase is known as Fine-Tuning, where model parameters are adjusted to align the acquired source knowledge with the target network for specific downstream tasks in the target domain. This stage generally requires a certain amount of labeled data relevant to the target field. In cases where there is an insufficient quantity of labeled data for the downstream task, it is possible to leverage source domain labeled data that comes from different distributions but pertains to similar tasks; this approach can enhance model performance on the target task through transfer learning techniques

In the entire lifecycle of deep learning, migration [5] plays a pivotal role. However, current research on migration primarily focuses on specific areas such as domain adaptation. As a fundamental component of transfer learning, its significance is evident, as it directly pertains to the widely studied issue of cross-domain knowledge transfer. Despite the depth of these studies, they fail to fully capture the comprehensive scope of migration within the broader framework of deep learning. Consequently, this review endeavors to explore domain barriers and bridge various fields within deep learning, offering a holistic reference for migration throughout the deep learning lifecycle. Given the breadth of the research area and space constraints, this review does not aim to cover all migration methods exhaustively. Instead, it emphasizes the core principles and key technologies of migration (Figure 1) and provides a concise overview of related extended literature. By doing so, we aspire to delineate the overarching landscape of migration in deep learning and offer valuable insights and guidance to researchers striving to enhance the efficiency of deep learning data utilization.

2. Domain Adaptation

Previous shallow domain adaptation (DA) methods [6] mainly focused on achieving knowledge transfer by either identifying shared features between the source and target domains or transforming source domain data to mimic the target domain's distribution. However, these approaches suffered from significant limitations, such as their reliance on manually designed feature transformation rules, difficulty in capturing high-level abstract information, and restricted adaptability when dealing with complex data distributions.

In contrast, recent domain adaptation techniques [7-8] have demonstrated numerous advantages. They automatically learn feature representations from data and facilitate knowledge transfer between different domains, thereby reducing the need for domain-specific knowledge. These techniques also harness the power of advanced models, like deep learning, which exhibit stronger data representation and generalization abilities. This enables them to effectively manage more complex data distributions

and cross-domain differences. Furthermore, by incorporating novel methods (Figure 2), such as adversarial learning [9], modern domain adaptation techniques achieve superior accuracy and robustness [10] across a range of tasks.

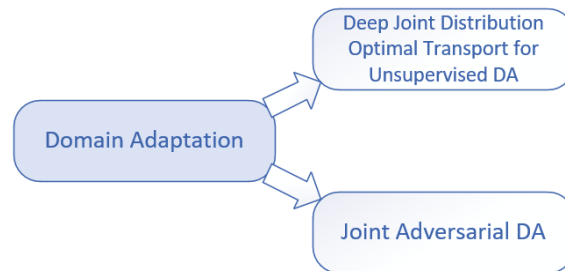


Figure 2. The survey is organized around two domain-adaptive technologies

2.1. Joint Adversarial Domain Adaptation

Adversarial learning within deep neural networks facilitates the extraction of transferable features that are applicable across various domains. Existing domain adaptation (DA) methods, which primarily concentrate on either domain-level or class-specific alignment, fall short in addressing extreme domain discrepancies common [11] in multimedia. Substantial differences can impair feature discriminability, while exclusive focus on class-level alignment may overlook the overall domain context. Our proposed JADA [12] framework integrates both aspects within a deep network, utilizing two minimax games to capture both global and category-specific information. This approach enhances cross-domain feature matching and improves prediction accuracy.

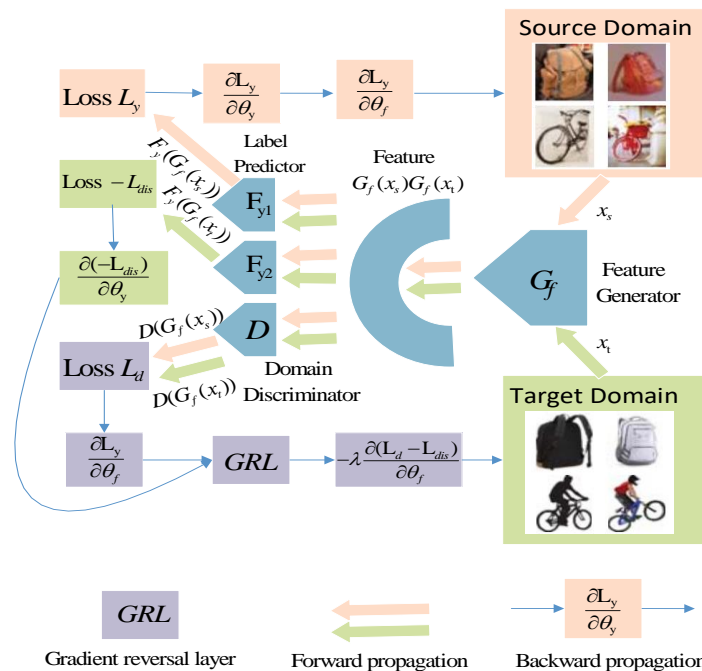


Figure 3. (Color recommended for best viewing.) An overview of the JADA architecture, where G_f and D denote the feature generator and domain discriminator, respectively; $G_f(\mathcal{X}_s)$ and $G_f(\mathcal{X}_t)$ represent the learned features for source and target data F_{y_1} and F_{y_2} are the two task-specific label predictors; L_y , $-L_{dis}$, and L_d signify the losses related to label predictions, discrepancy discrimination, and domain discrimination, respectively. GRL stands for gradient reversal layer, indicated by [symbol]; the red and green arrows illustrate the data flows for source and target data, respectively

2.1.1 Preliminaries

In DA problem, we denote the source domain $\mathcal{D}_s = \{(\mathcal{X}_{si}, \mathcal{Y}_{si})\}_{i=1}^{n_s}$ with n_s labeled samples, and the target domain $\mathcal{D}_t = \{(\mathcal{X}_{tj})\}_{j=1}^{n_t}$ with n_t unlabeled samples, where \mathcal{Y}_s is the corresponding label for \mathcal{X}_s . $n = n_s + n_t$ indicates the number of all the source and target samples. Since the two domains distributions are different, the goal of our method is to design a novel deep adversarial network that enables learning features $f = G_f(\mathcal{X})$ indistinguishable domain wisely as well as class-wisely, where G_f is the feature generator in JADA. Due to the distinct distributions of the two domains, our aim is to develop a unique deep adversarial network that facilitates the learning of features that are indistinguishable both at the domain and class levels. In this context, serves as the feature extractor within the JADA (Joint Adversarial Domain Adaptation) framework.

Adversarial-based domain adaptation faces three key technical challenges: (1) acquiring comprehensive domain knowledge to bridge the gap between the source and target domains; (2) discovering discriminative category structures within distributions to enable accurate class-specific alignment; and (3) Addressing the problem of mode collapse in adversarial training. These difficulties inspired our integrated adversarial domain adaptation strategy. In particular, within the JADA framework, we leverage the competitive interplay between the feature extractor and domain classifier to attain resilient domain-level correspondence. Furthermore, we introduce an adversarial dynamic between the feature generator and discrepancy discriminator, which guides the generation of target features that closely match the source class distributions. By integrating these two adversarial mechanisms into a streamlined deep framework, optimized through a three-player game, JADA enables seamless end-to-end training via standard back propagation.

2.1.2 Domain-wise Adversarial Learning

Adversarial training has demonstrated its efficacy in tackling diverse domain adaptation (DA) issues through a two-participant game aimed at acquiring domain-agnostic features. As depicted in Figure 3, the domain classifier D endeavors to precisely distinguish between samples from the source and target domains. Meanwhile, the feature extractor G_f is optimized to mislead the domain classifier, thus facilitating the acquisition of domain-invariant representations. In order to minimize the discrepancy in the common feature space, the parameters θ_d of D are optimized by minimizing the loss of domain discriminator while the parameters θ_f of G_f are trained to maximize the domain prediction loss. Here, we define $d_i = 1$ if $\mathcal{X}_i \in \mathcal{D}_s$, and $d_i = 0$ if $\mathcal{X}_i \in \mathcal{D}_t$. In addition, to guarantee the effectiveness of source model, the loss of the label predictors F_{y_1}, F_{y_2} on the annotated source data is also minimized. In JADA, the parameters of the two task-specific label predictors are denoted as $\theta_{y_1}, \theta_{y_2}$. Unlike the classical DANN, JADA features these two predictors, which work to align class-wise distributions across domains in an adversarial manner. As a result, the loss function for domain-wise alignment is formulated as follows:

$$\mathcal{J}_0(\theta_f, \theta_{y_1}, \theta_{y_2}, \theta_d) = \frac{1}{2n_s} \sum_{j=1}^2 \sum_{i=1}^{n_s} L_y(F_{y_j}(G_f(\mathcal{X}_{si})), \mathcal{Y}_{si}) - \frac{\lambda}{n} \sum_{\mathcal{X}_i \in (\mathcal{D}_s \cup \mathcal{D}_t)} L_d(G_d(G_f(\mathcal{X}_i)), d_i) \quad (1)$$

Here, λ serves as an adjustment coefficient to balance the loss $L_y(\cdot, \cdot)$ associated with label prediction and the loss $L_y(\cdot, \cdot)$ related to domain categorization. Within this study, the cross-entropy function is utilized for loss computation.

To extract global domain-invariant features, we are seeking the optimal parameters of $\hat{\theta}_f, \hat{\theta}_{y_1}, \hat{\theta}_{y_2}$ and $\hat{\theta}_d$, which will deliver a saddle point of Eq. (1) after the training convergence:

$$(\hat{\theta}_d) = \arg \max_{\theta_d} \mathcal{J}_0(\theta_f, \theta_{y_1}, \theta_{y_2}, \theta_d) \quad (2)$$

$$(\hat{\theta}_f, \hat{\theta}_{y_1}, \hat{\theta}_{y_2}) = \arg \min_{\theta_f, \theta_{y_1}, \theta_{y_2}} \mathcal{J}_0(\theta_f, \theta_{y_1}, \theta_{y_2}, \theta_d) \quad (3)$$

By this method, the acquired features f will become domain-invariant and encapsulate comprehensive domain-wide information, which has demonstrated its efficacy in addressing deep domain adaptation (DA) challenges.

2.1.3 Class-wise Adversarial Learning

In real-world multimedia domain adaptation scenarios, relying solely on domain-wise alignment can lead to a blurred distinction between source and target samples in the feature space [13], which undermines category discrimination, particularly for samples close to task-specific decision boundaries. To address this, we aim to align distribution information while preserving discriminative information. We utilize the decision boundaries of the source classifier as indicators of category structures. Inspired by previous work, we employ two source classifiers to identify target samples that lie near these boundaries, indicating their susceptibility to misclassification. The goal of JADA is to guide the feature extractor G_f to generate target features that closely align with the support of source classes, thereby achieving class-wise domain matching.

In Figure 2, we consider two source label predictors, F_{y_1} and F_{y_2} , which possess different characteristics and can be initialized differently to yield two distinct source classifiers capable of accurately classifying source data. When these predictors are used to forecast the labels of target data, target samples situated near decision boundaries are intuitively more likely to receive differing predictions from F_{y_1} and F_{y_2} .

In the context of class-wise adversarial learning, we strive to maximize the disagreement between F_{y_1} and F_{y_2} regarding their predictions for target samples. This methodology assists in recognizing target instances that fall beyond the scope of the source distribution. On the other hand, the function of the feature generator G_f is to reduce this discrepancy, thereby aligning the source and target distributions within individual classes. The interaction between the feature extractor and the two separate classifiers attains a state of balance through an adversarial training process.

Specifically, we denote F_{y_1} and F_{y_2} to construct a discrepancy discriminator, and the L1-norm of the disagreement between their predictions of target sample \mathcal{X}_t as discrepancy loss:

$$L_{dis}(\mathcal{P}_t^{y_1}, \mathcal{P}_t^{y_2}) = \|\mathcal{P}_t^{y_1} - \mathcal{P}_t^{y_2}\|_1 \quad (4)$$

Where $\mathcal{P}_t^{y_1}$ and $\mathcal{P}_t^{y_2}$ are probabilistic outputs of F_{y_1} and F_{y_2} for \mathcal{X}_t , respectively.

Therefore, we formulate the loss function of class-wise adversarial learning as:

$$\mathcal{J}_1(\theta_f, \theta_{y_1}, \theta_{y_2}) = \frac{1}{n_t} \sum_{i=1}^{n_t} L_{dis} \left(F_{y_1} \left(G_f(\mathcal{X}_{t_i}) \right), F_{y_2} \left(G_f(\mathcal{X}_{t_i}) \right) \right) \quad (5)$$

To achieve the class-wise adversarial goals, we want the optimal parameters $\hat{\theta}_f$, $\hat{\theta}_{y_1}$ and $\hat{\theta}_{y_2}$ to jointly satisfy:

$$(\hat{\theta}_{y_1}, \hat{\theta}_{y_2}) = \arg \max_{\theta_{y_1}, \theta_{y_2}} \mathcal{J}_1(\theta_f, \theta_{y_1}, \theta_{y_2}) \quad (6)$$

$$(\hat{\theta}_f) = \arg \min_{\theta_f} \mathcal{J}_1(\theta_f, \theta_{y_1}, \theta_{y_2}) \quad (7)$$

2.2. Deep Joint Distribution Optimal Transport for Unsupervised Domain Adaptation

Unsupervised domain adaptation (UDA) delves into situations where labeled data is available in the source domain, while the target domain remains unlabeled yet accessible during training. Initial methodologies emphasized aligning domains by projecting their respective data spaces onto each other, utilizing invariant features to bridge the gap in the original (input) space. As research evolved, these concepts were extended to deep learning frameworks, often involving weight sharing or reconstruction techniques. This progression saw the integration of Maximum Mean Discrepancy (MMD) [11] and association-based losses to further align source and target layers. A prominent development in this field has been the adoption of adversarial loss functions, which are crafted to

challenge Convolutional Neural Networks (CNNs) in distinguishing between source and target domain samples, thereby fostering the learning of domain-invariant representations.

Recently, adversarial techniques have been expanded to incorporate Generative Adversarial Networks (GANs) [14-16]. This includes using two GAN modules with shared weights to enforce similar activation distributions in image-to-image translation architectures, or fooling a GAN discriminator designed to differentiate between domains. These methods, based on adversarial image generation, utilize class conditioning or cycle consistency constraints to learn a discriminative embedding, where semantically similar images from both domains are mapped close to each other in the embedding space. Our proposed Deep JDOT approach adopts a shared embedding space for the two domains, similar to MMD-based methods, but introduces a unique authentication mechanism to the alignment process. Deep JDOT combines representation learning with discriminative learning, as the optimal transport coupling ensures distribution matching. Specifically, the JDOT class loss enables label propagation from source to target samples, and learning the coupling in the deep layers of the CNN enhances the authentication capability.

In summary, DeepJDOT not only aligns domains but also enhances authentication, addressing limitations of prior UDA methods that may sacrifice category authentication during domain alignment.

Optimal transport in domain adaptation. [17] Optimal transport techniques have been employed in domain adaptation to discover the transformation between domains, supported by solid theoretical foundations. In such methods, the coupling γ is leveraged to transfer (i.e. transform) source data points through a derived mapping, referred to as the barycenter mapping. Following this, a classification model is trained using the transformed source data representation. However, many of these methods are limited to handling small to medium-sized problems, as they depend on finding the exact solution to the optimal transport problem across all samples.

2.2.1 Deep Joint Distribution Optimal Transport (DeepJDOT)

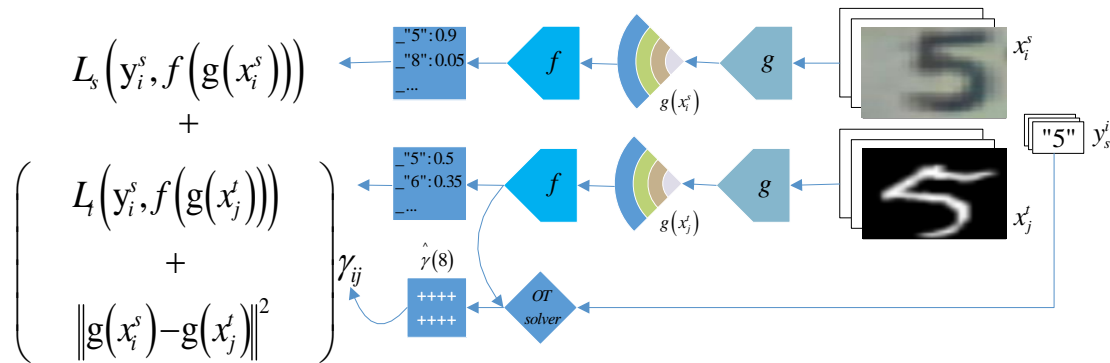


Figure 4. Overview of the proposed DeepJDOT method. While the structure of the feature extractor g and the classifier f are shared by both domains, they are represented twice to distinguish between the two domains. Both the latent representations and labels are used to compute per batch a coupling matrix γ that is used in the global loss function

The DeepJDOT model, illustrated in Figure 4, is composed of two parts: an embedding function $g: x \rightarrow z$, where the input is mapped into the latent space Z , and the classifier $f: z \rightarrow y$, this process projects the hidden space onto the label space within the target domain. The hidden space may correspond to any feature layer offered by the model, such as the second-to-last fully connected layer of a Convolutional Neural Network (CNN) in our instance. DeepJDOT jointly optimizes both this feature representation and the classification model to ensure robust performance on the target domain. The resolution to this challenge is attained by minimizing the subsequent loss function:

$$\min_{\gamma \in \Pi(\mu_s, \mu_t), f, g} \sum_i \sum_j \gamma_{ij} d(g(\mathbf{x}_i^s), \mathbf{y}_i^s; g(\mathbf{x}_j^t), f(g(\mathbf{x}_j^t))) \tag{8}$$

Where $d(g(\mathbf{x}_i^s), \mathbf{y}_i^s; g(\mathbf{x}_j^t), f(g(\mathbf{x}_j^t))) = \alpha \|g(\mathbf{x}_i^s) - g(\mathbf{x}_j^t)\|^2 + \lambda_t L(\mathbf{y}_i^s, f(g(\mathbf{x}_j^t)))$, the parameters α and λ_t serve as regulators in our approach. Similar to JDOT, the first part of the loss

evaluates the alignment of the embeddings between the source and target domains. In contrast, the second part concentrates on the classifier \mathcal{f} , which is trained on the target domain, and its agreement with the labels from the source domain. Although our method bears some resemblance to the JDOT formulation, a key difference is that DeepJDOT minimizes the Wasserstein distance between the joint distributions (embedding space paired with labels) within the CNN layers, instead of between the original input spaces. Given that deeper CNN layers capture both spatial and semantic details, we argue that they provide a more resilient representation of image content across domains, overcoming the limitations of raw features that are sensitive to changes in illumination, object pose, or relative positioning.

It is worth noting that the formulation presented in equation (9) solely relies on the classifier trained in the target domain. This approach prioritizes learning an effective classifier for the target domain while disregarding its performance on source samples. In recent research, this decline in source domain performance has been termed 'catastrophic forgetting.' To mitigate this issue, one can simply reintroduce the loss on the source domain into equation (9), resulting in the ultimate DeepJDOT objective:

$$\min_{\gamma, \mathcal{f}, \mathbf{g}} \frac{1}{n^s} \sum_i L_s(y_i^s, \mathcal{f}(\mathbf{g}(x_j^t))) + \sum_{i,j} \gamma_{i,j} \left(\alpha \|\mathbf{g}(x_i^s) - \mathbf{g}(x_j^t)\|^2 + \lambda_t L(y_i^s, \mathcal{f}(\mathbf{g}(x_j^t))) \right) \quad (9)$$

The final formulation represents the optimization problem addressed by DeepJDOT. Nevertheless, for substantial sample sizes, the requirement to compute a complete γ renders the problem computationally impractical, considering both memory and time complexity. In the subsequent section, we introduce an approximation technique based on stochastic optimization.

3. Experiment

3.1. Dataset

VisDA-2017 is a challenging synthetic-to-real dataset that serves as a benchmark for large-scale cross-domain object classification. It includes a comprehensive set of over 280,000 images across 12 categories, divided into training, validation, and test sets. We designate the training set, consisting of 152,397 synthetic images generated from 3D CAD models with varying angles and lighting, as the source domain. The target domain is represented by the validation set, which comprises 55,388 real images from MSCOCO.

3.2. Implementation

3.2.1 Joint Adversarial Domain Adaptation

Table 1. Accuracy (%) on VisDA-2017 for unsupervised DA (ResNet-101)

Method	plane	bcycl	bus	car	horse	knife	mcycl	person	plant	sktbrd	train	truck	Avg
Source Only	55.1	53.3	61.9	59.1	80.6	17.9	79.7	31.2	81.0	26.5	73.5	8.5	52.4
DAN	87.1	63.0	76.5	42.0	90.3	42.9	85.9	53.1	49.7	36.3	85.8	20.7	61.1
DANN	81.9	77.7	82.8	44.3	81.2	29.5	65.1	28.6	51.9	54.6	82.8	7.8	57.4
MCD	87.0	60.9	83.7	64.0	88.9	79.6	84.7	76.9	88.6	40.3	83.0	25.8	71.9
JADA	91.9	78.0	81.5	68.7	90.2	84.1	84.0	73.6	88.2	67.2	79.0	38.0	77.0

Note: The numbers in parentheses represent the disparity in accuracy between the original (unadapted) model and the modified (adapted) model. The results for CORAL and DAN are derived from the evaluation server⁶.

Experiment results on VisDA-2017: Table 1 presents the outcomes for the VisDA-2017 benchmark, demonstrating the efficacy of our model. In comparison to a source-only classifier directly applied to the target domain without adaptation, our model achieves higher accuracy across all object categories. Notably, domain-specific alignment methods fall short in categories like cars and trucks, indicating their limitations. Moreover, our approach significantly outperforms the best-

performing baseline method, MCD, highlighting the challenges of synthetic-to-real image adaptation. While MCD leverages task-specific decision boundaries, it may not fully address the complex domain and class-level discrepancies. Consequently, the need for a combined domain-wise and class-wise adaptation strategy becomes apparent, as it facilitates a more thorough alignment of distributions, tailored to the distinct characteristics of each domain (see Table 1).

3.2.2 Deep Joint Distribution Optimal Transport for Unsupervised Domain Adaptation

VisDA-2017:

Table 2. Performance of DeepJDOT on the VisDA 2017 classification challenge

Source Only	36.0	4.0	19.9	94.7	14.8	0.42	38.7	3.8	37.4	8.1	71.9	6.7	28.0
DeepCORAL	62.5	21.7	66.3	64.6	31.1	36.7	54.2	24.9	73.8	29.9	43.4	34.2	45.3 (19.0)
DAN	55.3	18.4	59.8	68.6	55.3	41.4	63.4	30.4	78.8	23.0	62.9	40.2	49.8 (19.5)
DeepJDOT	85.4	50.4	77.3	87.3	69.1	14.1	91.5	53.3	91.9	31.2	88.5	61.8	66.9 (38.9)

Model: In response to the complex adaptation challenges posed by VisDA, we use ResNet-50 as our base model. We modify its final layer by adding two Multi-Layer Perceptrons (MLPs) that sequentially map features to 512 hidden units and then to the 12 target classes. We first train this model on the source domain and then freeze it to extract source feature vectors. An identical version of this model is then adapted to the target domain. For computing the couplings, we use 4,096 samples from each domain. Additionally, we follow the data augmentation strategy outlined to further boost our model's performance.

Results: Table 2 displays the comparative results of DeepJDOT on the VisDA-2017 dataset, alongside benchmark methods such as DeepCORAL and DAN. Our submission, labeled as 'oatmil' on the evaluation platform, shows higher accuracy in most classes, except for the 'knife' category, compared to distribution alignment techniques like DeepCORAL and DAN. Notably, while DeepJDOT experiences negative transfer for the 'car' class, this is a common issue among current state-of-the-art methods, as indicated by the evaluation server's results. For a fair comparison, we also include (in parentheses in Table 4) the accuracy improvement achieved by our method over the source model when transferred to the target domain. At the time of publication, our approach ranked sixth in terms of average accuracy and third in terms of improvement from the source model to the target model. It is worth noting that the performance of DeepJDOT is dependent on the strength of the source model used; employing a larger Convolutional Neural Network (CNN) could potentially lead to even better results.

4. Conclusion

In **Joint Adversarial Domain Adaptation**, we propose a new method called Joint Adversarial Domain Adaptation (JADA) to address the challenge of domain adaptation. Unlike previous adversarial learning-based domain adaptation approaches, JADA simultaneously leverages both global domain-level information and fine-grained class-discriminative structures. It integrates the minimization of domain-specific and class-specific distribution discrepancies into a cohesive and synergistic adversarial learning framework. This unified approach makes JADA more robust to significant domain differences, facilitating a smoother transition between domains. Additionally, JADA is designed to be optimized end-to-end using back-propagation for efficient training. Experimental results demonstrate that JADA outperforms state-of-the-art methods on various standard cross-domain datasets.

In **Deep Joint Distribution Optimal Transport for Unsupervised Domain Adaptation**, We introduce the DeepJDOT model, an innovative technique designed to establish a shared latent representation space for both source and target distributions, while preserving their discriminative power. This is achieved by minimizing the divergence between the joint distributions of deep features and labels across the two domains, utilizing principles from optimal transport. To tackle this challenge, we develop an efficient stochastic algorithm that can be easily integrated into modern deep learning

frameworks. Importantly, despite its simplicity and compatibility, our approach outperforms current state-of-the-art methods in cross-domain digit classification and office-home adaptation tasks, and also achieves impressive results on the challenging VisDA-2017 adaptation benchmark.

References

- [1] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, 521 (7553), 436-444.
- [2] Junyi Chai, Hao Zeng, Anming Li, Eric W.T. Ngai, Deep learning in computer vision: A critical review of emerging techniques and application scenarios, *Machine Learning with Applications*, Volume 6, 2021, 100134, ISSN 2666-8270, <https://doi.org/10.1016/j.mlwa.2021.100134>.
- [3] Taye MM. Theoretical Understanding of Convolutional Neural Network: Concepts, Architectures, Applications, Future Directions. *Computation*. 2023; 11 (3): 52. <https://doi.org/10.3390/computation11030052>.
- [4] Sun, C., Shrivastava, A., Singh, S., & Gupta, A. (2017). Revisiting unreasonable effectiveness of data in deep learning era. In *Proceedings of the IEEE international conference on computer vision* (pp. 843-852).
- [5] Sünderhauf, N., Brock, O., Scheirer, W., Hadsell, R., Fox, D., Leitner, J., ... & Corke, P. (2018). The limits and potentials of deep learning for robotics. *The International journal of robotics research*, 37 (4-5), 405-420.
- [6] Mehrkanoon, S., Blaschko, M., & Suykens, J. (2018). Shallow and deep models for domain adaptation problems. *Proceedings ESANN 2018*, 291-299.
- [7] Wang, M., & Deng, W. (2018). Deep visual domain adaptation: A survey. *Neurocomputing*, 312, 135-153.
- [8] Zhao, S., Li, B., Xu, P., & Keutzer, K. (2020). Multi-source domain adaptation in the deep learning era: A systematic survey. *arxiv preprint arxiv: 2002.12169*.
- [9] Su, J. C., Tsai, Y. H., Sohn, K., Liu, B., Maji, S., & Chandraker, M. (2020). Active adversarial domain adaptation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 739-748).
- [10] Jang, H., & Tong, F. (2024). Improved modeling of human vision by incorporating robustness to blur in convolutional neural networks. *Nature Communications*, 15 (1), 1989.
- [11] Du, Z., Li, J., Su, H., Zhu, L., & Lu, K. (2021). Cross-domain gradient discrepancy minimization for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 3937-3946).
- [12] Li, S., Liu, C. H., Xie, B., Su, L., Ding, Z., & Huang, G. (2019). Joint adversarial domain adaptation. In *Proceedings of the 27th ACM International Conference on Multimedia* (pp. 729-737).
- [13] Li, Y., Zhang, D., Keuper, M., & Khoreva, A. (2024). Intra-& extra-source exemplar-based style synthesis for improved domain generalization. *International Journal of Computer Vision*, 132 (2), 446-465.
- [14] Aggarwal, A., Mittal, M., & Battineni, G. (2021). Generative adversarial network: An overview of theory and applications. *International Journal of Information Management Data Insights*, 1 (1), 100004.
- [15] Saxena, D., & Cao, J. (2021). Generative adversarial networks (GANs) challenges, solutions, and future directions. *ACM Computing Surveys (CSUR)*, 54 (3), 1-42.
- [16] Gui, J., Sun, Z., Wen, Y., Tao, D., & Ye, J. (2021). A review on generative adversarial networks: Algorithms, theory, and applications. *IEEE transactions on knowledge and data engineering*, 35 (4), 3313-3332.
- [17] Shuang Li, Mixue Xie, Kaixiong Gong, Chi Harold Liu, Yulin Wang, Wei Li; *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 11516-11525.