



ARTICLE

EDTM: Efficient Domain Transition for Multi-Source Domain Adaptation

Mangyu Lee^{1,#}, Jaekyun Jeong^{2,#}, Yun Wook Choo³, Keejun Han⁴ and Jungeun Kim^{2,*}

¹Department of Computer Science and Engineering, Kongju National University, Cheonan, Republic of Korea

²Department of Electrical and Computer Engineering, Inha University, Incheon, Republic of Korea

³Department of Civil and Environmental Engineering, Kongju National University, Cheonan, Republic of Korea

⁴School of Computer Engineering, Hansung University, Seoul, Republic of Korea

*Corresponding Author: Jungeun Kim. Email: jekim@inha.ac.kr

#These authors contributed equally to this work

Received: 10 October 2025; Accepted: 22 January 2026; Published: 26 February 2026

ABSTRACT: Domain adaptation aims to reduce the distribution gap between the training data (source domain) and the target data. This enables effective predictions even for domains not seen during training. However, most conventional domain adaptation methods assume a single source domain, making them less suitable for modern deep learning settings that rely on diverse and large-scale datasets. To address this limitation, recent research has focused on Multi-Source Domain Adaptation (MSDA), which aims to learn effectively from multiple source domains. In this paper, we propose Efficient Domain Transition for Multi-source (EDTM), a novel and efficient framework designed to tackle two major challenges in existing MSDA approaches: (1) integrating knowledge across different source domains and (2) aligning label distributions between source and target domains. EDTM leverages an ensemble-based classifier expert mechanism to enhance the contribution of source domains that are more similar to the target domain. To further stabilize the learning process and improve performance, we incorporate imitation learning into the training of the target model. In addition, Maximum Classifier Discrepancy (MCD) is employed to align class-wise label distributions between the source and target domains. Experiments were conducted using Digits-Five, one of the most representative benchmark datasets for MSDA. The results show that EDTM consistently outperforms existing methods in terms of average classification accuracy. Notably, EDTM achieved significantly higher performance on target domains such as Modified National Institute of Standards and Technology with blended background images (MNIST-M) and Street View House Numbers (SVHN) datasets, demonstrating enhanced generalization compared to baseline approaches. Furthermore, an ablation study analyzing the contribution of each loss component validated the effectiveness of the framework, highlighting the importance of each module in achieving optimal performance.

KEYWORDS: Multi-source domain adaptation; imitation learning; maximum classifier discrepancy; ensemble based classifier; EDTM

1 Introduction

In recent years, deep learning has achieved significant advancements in various fields such as image recognition, speech recognition, and natural language processing through training on large-scale datasets [1]. To effectively train on such datasets, the data collection process plays a crucial role. However, while data are generated at an extremely rapid rate, accurate labeling requires considerable time and cost, which hinders the advancement of deep learning models [2]. Consequently, addressing the challenge of data labeling has emerged as an essential task for the continued development of deep learning models.

Traditional deep learning models are typically effective only within the domain in which they have been trained. Even when the same labels are shared across datasets, differences in data distributions can significantly hinder predictive performance. This limitation necessitates re-labeling and retraining models for each new but similar dataset, which is both time-consuming and inefficient. While transfer learning can alleviate part of this burden by leveraging labeled target data to adapt a pretrained model, its applicability is limited in scenarios where target labels are unavailable. Domain adaptation addresses this issue by mitigating the distribution discrepancy between source and target data without relying on labeled target samples. By effectively integrating the characteristics of both domains, the model can extract domain-invariant feature representations, leading to consistent and robust predictions. Due to these capabilities, domain adaptation has emerged as a promising approach to reduce the cost and effort associated with data labeling, especially in settings where labeled data in the target domain are scarce or entirely absent.

With the increasing diversity of data and the emergence of domain-specific variations even within the same type of data, recent trends in deep learning have shifted toward leveraging diverse training data more effectively. Utilizing such a variety of data sources can improve model generalization, enable the acquisition of broader knowledge, and help prevent overfitting. One approach that applies this idea to domain adaptation is Multi-source Domain Adaptation (MSDA), which involves two key challenges: integrating knowledge from multiple source domains and aligning this knowledge with the target domain.

Traditional MSDA methods primarily focus on effectively mixing or aggregating the characteristics of multiple source domains, often relying on adversarial GAN-based frameworks to align the joint source–target distributions. While these adversarial approaches have demonstrated promising results, the min–max optimization tends to emphasize global distribution matching and may overlook more fine-grained, label-wise discrepancies across domains. Motivated by this limitation, our work adopts Maximum Classifier Discrepancy (MCD) to perform a more delicate and class-aware alignment process. By leveraging classifier-induced discrepancies rather than adversarial signals, MCD enables the model to capture subtle structural differences that GAN-based approaches may fail to address, thereby enabling more precise domain integration.

Although numerous studies have proposed methods to tackle both source integration and target alignment, solving these tasks simultaneously in a stable and effective manner remains a significant challenge. To address these issues, this paper introduces three complementary approaches.

The first method is the Ensemble-based Classifier Expert, which assigns weights to multiple source domain models using a domain classification model to generate a single integrated classifier [3]. Source domains that are more similar to the target domain contribute more significantly to the final prediction, thereby improving adaptation effectiveness.

The second method is Imitation Learning, where the target classifier is trained to mimic the output of the trained source ensemble model [3]. This enables the target model not only to learn from source-domain supervision but also to incorporate pseudo-label guidance for target data, enhancing prediction stability and adaptation quality.

The final method is Maximum Classifier Discrepancy (MCD), which considers label-wise data distributions during adaptation [4]. Unlike methods that align only the overall distribution, MCD adjusts discrepancies across individual labels, allowing for fine-grained and class-aware domain alignment.

By incorporating these three methods, we propose Efficient Domain Transition for Multi-source Domain Adaptation (EDTM), which effectively integrates knowledge from multiple source domains and aligns label-wise distributions between source and target domains. The proposed EDTM framework offers the following contributions:

- Developed a model that prioritizes source domains more similar to the target domain, enhancing the transferability of knowledge.
- Improved learning stability and prediction efficiency by training the target model to imitate the source ensemble's outputs while adapting to target-specific characteristics.
- Achieved more precise domain integration through label-wise alignment using MCD.
- Conducted extensive experiments on benchmark MSDA datasets, demonstrating that EDTM consistently outperforms existing methods.

2 Related Work

2.1 Domain Adaptation

Many domain adaptation studies have been conducted to effectively align data distributions. Among them, numerous works have focused on domain adaptation using adversarial learning techniques. Ganin et al. [5] proposed Domain Adversarial Neural Networks (DANN), which align the distributions of two domains using a Gradient Reversal Layer. Tzeng et al. [6] introduced Adversarial Discriminative Domain Adaptation (ADDA), which leverages the Generative Adversarial Networks (GAN) algorithm to align distributions across domains. Saito et al. [4] highlighted a limitation of conventional adversarial learning methods, which align data without considering label-wise distributions. To address this, they proposed Maximum Classifier Discrepancy (MCD), a method that aligns domains while accounting for label-wise distribution differences.

2.2 Multi-Source Domain Adaptation

Multi-source domain adaptation has been actively studied to address the diversity of source domains and the distribution shift between source and target domains. Many works focus on handling source domain diversity by computing similarities between domains and integrating knowledge through weighted contributions. Zhao et al. [7] proposed Multi-source Distilling Domain Adaptation (MDDA), which calculates the similarity between target and source domains based on distance, assigns weights accordingly, and predicts data using an ensemble approach. Wang et al. [8] utilized statistical distance metrics such as Maximum Mean Discrepancy to measure feature distribution differences between each source domain and the target domain. These differences were then used as weights, allowing more similar source domains to contribute more significantly to target domain prediction. They proposed Learning to Combine for Multi-Source Domain Adaptation (LtC-MSDA).

Nguyen et al. [3] proposed Student-Teacher Ensemble Multi-source Domain Adaptation (STEM), which learns from diverse source domains using an Ensemble-based Teacher Expert and transfers this knowledge to the target model through Imitation Learning. This approach enhances prediction accuracy by aligning domain distributions using a GAN-based method.

3 Methodology

This section describes the EDTM framework and its key components: Ensemble-based Classifier Expert, Imitation Learning, and MCD. Finally, the training process of EDTM is introduced.

3.1 EDTM Framework

Fig. 1 illustrates the overall architecture of the proposed EDTM framework. The framework utilizes data collected from multiple source domains (K domains), denoted as $\mathcal{X}_s = \{\mathcal{X}_s^k\}_{k=1}^K$, along with target domain data \mathcal{X}_t .

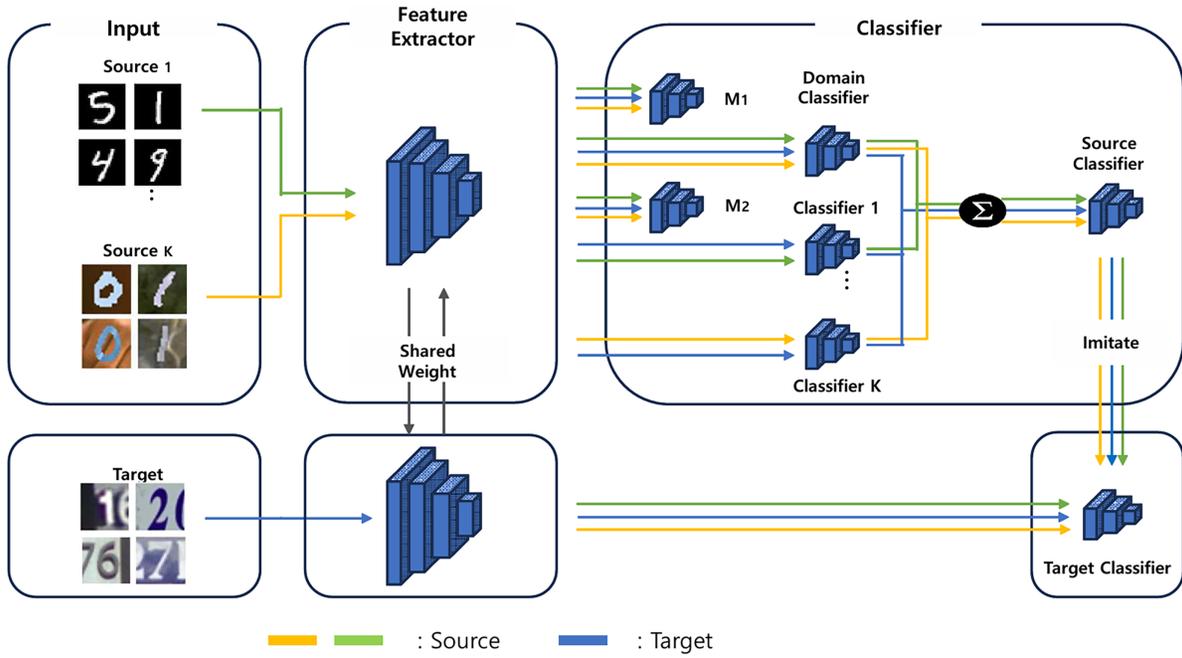


Figure 1: Overall architecture of the proposed EDTM framework. The model extracts shared features from multiple source and target domains using a common feature extractor. Source-specific classifiers and a domain discriminator are used to form a unified source classifier, which guides the target classifier through imitation learning. Additionally, MCD classifiers assist in learning domain-invariant and class-discriminative representations.

For clarity, all loss functions in this section are computed over mini-batches. We denote $\mathcal{B}_k^s \subset X_k^s$ as the mini-batch sampled from the k th source domain and $\mathcal{B}^t \subset X^t$ as the mini-batch sampled from the target domain. Unless otherwise stated, all summations are taken over samples within a mini-batch, and each loss term represents the batch-averaged loss.

First, both source and target data are fed into a shared feature extractor F , which maps them into a common representation space. For each source domain, a label classifier $\{C_k\}_{k=1}^K$ is trained to predict class labels, and a domain discriminator D is trained to distinguish the domain of each input.

The trained classifiers $\{C_k\}_{k=1}^K$ and discriminator D are then integrated to form a unified source classifier C_s , which reflects information from all source domains. Subsequently, the target classifier C_t is trained to align with the predictions of C_s , thereby adapting to the target domain and improving its classification performance.

In addition, two MCD classifiers M_1 and M_2 are introduced to perform label prediction on the entire source dataset. These classifiers guide the feature extractor F to learn domain-invariant and harmonized representations across different domains.

Through this architecture and training strategy, EDTM effectively combines information from multiple source domains and learns domain-aware, label-discriminative features tailored to the target domain.

3.2 Ensemble-Based Classifier Expert

The Ensemble-based Classifier Expert integrates the classifiers trained on each source domain, $\{C_k\}_{k=1}^K$, to construct C^s . First, the feature extractor (F) extracts features from the data X^s . Then, $\{C_k\}_{k=1}^K$ and the domain classifier (D) are trained. The trained D predicts the domain membership of the data and analyzes the features of X^t to determine which domain the given data belongs to.

This prediction result is used as a weight for each classifier $\{C_k\}_{k=1}^K$ when constructing C^s . This weight-based integration process ensures that information from source domains similar to the target domain is effectively reflected. Using this method, we integrate knowledge from each source domain to construct C^s . The construction process of C^s is defined as:

$$C^s(x) = \sum_{k=1}^K w_k(x) C_k(x), \quad (1)$$

where $w_k(x)$ is the weight assigned to each classifier C_k , determined by the domain classifier D [3]. The weights $\{w_k(x)\}_{k=1}^K$ form a probability distribution over the source domains, satisfying $\sum_{k=1}^K w_k(x) = 1$, which enables normalized and interpretable contributions from each source classifier.

Compared to a single classifier trained on an individual source domain, the ensemble-based formulation reduces domain-specific bias by aggregating complementary knowledge from multiple sources and alleviates negative transfer by down-weighting less relevant domains, leading to more robust predictions on the target domain.

Compared to using a single classifier trained on a specific source domain, the ensemble-based formulation reduces domain-specific bias by aggregating complementary knowledge from multiple sources. Moreover, the adaptive weighting mechanism mitigates negative transfer by suppressing the influence of less relevant source domains, while preventing overfitting to a single source domain. As a result, the proposed ensemble classifier provides a more robust and stable decision boundary for target-domain samples.

3.3 Imitation Learning

Imitation Learning trains C^t to mimic C^s , thereby leveraging the combined knowledge from multiple source domains. To achieve this, the prediction difference between C^s and C^t is measured, and C^t is trained to minimize this difference. The prediction discrepancy between C^s and C^t is computed as:

$$L_{imi}(x) = \sum_{n=1}^N C^s(F(x))^n \log C^t(F(x))^n, \quad (2)$$

where N represents the total number of labels, and $C^s(F(x))^n$ and $C^t(F(x))^n$ denote the probabilities that the input data x belongs to the n -th label. This approach enables C^t to learn from C^s even in the absence of labeled data. By applying this method, we trained C^t not only to mimic C^s for X^s but also for X^t . As a result, C^t is further refined to reflect the characteristics of the target domain [3].

3.4 Maximum Classifier Discrepancy

MCD effectively aligns data distributions across domains while considering label distributions. First, the feature extractor (F) extracts features from X^s and X^t . Then, two classifiers, M_1 and M_2 , are trained to classify labels based on features extracted from the source data. Through this training, M_1 and M_2 achieve stable prediction performance on the source domain. However, due to the randomness in the training process, the decision boundaries of the two classifiers may differ. This discrepancy increases the likelihood that the models will produce different predictions when encountering data with distributions different from the source domain. In particular, if target domain data are located near the decision boundaries, prediction inconsistency becomes more pronounced.

To leverage this property, the extracted features of X^t are fed into the trained classifiers M_1 and M_2 , which output the probability distributions $p_1(X^t)$ and $p_2(X^t)$ over the labels. The difference between $p_1(X^t)$

and $p_2(X^t)$ is then calculated to measure the prediction discrepancy between the two classifiers, denoted as $d(p_1(X^t), p_2(X^t))$. The prediction discrepancy is computed as follows:

$$d(p_1(x), p_2(x)) = \frac{1}{N} \sum_{n=1}^N |p_1(x)^n - p_2(x)^n|, \quad (3)$$

where $p_1(x)^n$ and $p_2(x)^n$ represent the probabilities that the input data x belongs to the n -th label. To align the label distributions of the target and source domains, F is trained to minimize $d(p_1(X^t), p_2(X^t))$. Conversely, M_1 and M_2 are trained to maximize $d(p_1(X^t), p_2(X^t))$, enabling them to better detect label distribution differences between target and source data. Through this process, domain alignment is achieved while accounting for label-wise distribution differences between the source and target domains [4].

3.5 Overall Training

EDTM trains by sampling mini-batches x^s and x^t from the source data X^s and target data X^t in each epoch. In every epoch, the loss functions $\{L_{C_k}\}_{k=1}^K$ and L_D are first computed using the sampled mini-batches for $\{C_k\}_{k=1}^K$ and D . Through this process, $\{C_k\}_{k=1}^K$ are trained to accurately classify the labels of the source domain data, while D is trained to correctly classify the domain of each source sample. The loss functions L_{C_k} and L_D are computed as follows:

$$L_{C_k} = \sum_{x_k^s \in X_k^s} y_k^s \log C_k(F(x_k^s)), \quad (4)$$

where this loss encourages each source classifier C_k to correctly predict class labels using features extracted by F .

$$L_D = \sum_{x_k^s \in X_k^s} d_k^s \log D(F(x_k^s)). \quad (5)$$

This loss enables the domain classifier D to learn discriminative representations for identifying the originating source domain of each sample. Where y_k^s represents the ground-truth label of x_k^s , and d_k^s denotes the domain label of x_k^s . Next, C^s is constructed using $\{C_k\}_{k=1}^K$ and D . To ensure that C^t mimics the output of C^s , the loss function L_I for C^t is computed as follows:

$$L_I = L_{imi}(x^s) + L_{imi}(x^t). \quad (6)$$

This imitation loss enforces output-level consistency between C^s and C^t for both source and target samples. Additionally, to ensure that M_1 and M_2 effectively classify the labels of x^s , the loss function L_M is computed as:

$$L_M = \sum_{k=1}^2 \sum_{x^s \in X^s} y^s \log M_k(F(x^s)). \quad (7)$$

This loss guides both classifiers M_1 and M_2 to learn reliable decision boundaries based on labeled source data. The network components, including F , $\{C_k\}_{k=1}^K$, C^t , D , M_1 , and M_2 , are then jointly trained to minimize the sum of all computed loss functions:

$$L_D + L_I + L_M + \sum_{k=1}^K L_{C_k}. \quad (8)$$

This joint optimization step aligns feature extraction, classification, domain discrimination, and imitation objectives within a unified training framework.

After this initial training phase, M_1 and M_2 are retrained. First, the trained M_1 and M_2 are used to compute $d(p_1(x^t), p_2(x^t))$. Then, to enhance their ability to effectively detect target data, they are trained to maximize $d(p_1(x^t), p_2(x^t))$. The corresponding loss function, L_{MCD} , is defined as follows:

$$L_{MCD} = L_M - d(p_1(x^t), p_2(x^t)). \quad (9)$$

This objective encourages the two classifiers to produce maximally discrepant predictions on target samples, facilitating the identification of target-domain uncertainty. Finally, F undergoes retraining. The following loss function ensures that F is trained to minimize $d(p_1(x^t), p_2(x^t))$:

$$L_F = d(p_1(x^t), p_2(x^t)). \quad (10)$$

By minimizing this discrepancy, the feature extractor F learns target-invariant representations that align the predictions of M_1 and M_2 . This training process is repeated for γ iterations. Algorithm 1 presents the pseudocode for the proposed EDTM training process.

Algorithm 1: Training procedure for EDTM

Input: Source domains $\{(X^s, Y^s)\}_{s=1}^K$, Target domain X^t , Hyperparameters $\alpha, \beta, \delta, \gamma$

Output: Trained F , $\{C_k\}_{k=1}^K$, C^t , D , M_1 , M_2

```

1 for epoch in epochs do
2   Sample mini-batches of  $\{(X^s, Y^s)\}_{s=1}^K, X^t$ ;
3   for k in K do
4     Compute  $L_{C_k}$ ;
5   end
6   Compute  $L_D$ ;
7   Construct  $C^s$  using  $\{C_k\}_{k=1}^K$  and  $D$ ;
8   Compute  $L_I, L_M$ ;
9   Update  $F, \{C_k\}_{k=1}^K, C^t, D, M_1, M_2$  by minimizing Eq. (8);
10  Compute  $L_{MCD}$ ;
11  Update  $M_1, M_2$  by minimizing  $L_{MCD}$ ;
12  for repeat in  $\gamma$  do
13    Compute  $L_F$ ;
14    Update  $F$  by minimizing  $L_F$ ;
15  end
16 end

```

4 Experiments

This section describes the datasets used in the experiments and the performance comparison between existing MSDA methods and the proposed EDTM. Finally, the experimental results are presented.

4.1 Datasets

The datasets used for benchmarking are Digits-Five, Office-Caltech, and Office-31, which are representative benchmarks for evaluating the performance of domain adaptation algorithms. Digits-Five consists of Modified National Institute of Standards and Technology (MNIST) (mt) [9], Modified MNIST with

blended background images (MNIST-M) (*mm*) [5], Street View House Numbers (SVHN) (*sv*) [10], Synthetic Digits (SYN) (*sy*) [5], and United States Postal Service (USPS) (*up*) [11], each containing class labels from 0 to 9. Since the classes are the same across datasets but the domains vary, this collection serves as an effective benchmark for evaluating MSDA performance.

Office-Caltech consists of four domains: Amazon, Webcam, DSLR, and Caltech, which share ten common object categories and are commonly represented using pre-extracted features from ResNet-101. This dataset configuration is based on the Office+Caltech-10 benchmark [12]. Office-31 contains three domains: Amazon, Webcam, and DSLR, with thirty-one object categories, providing a more diverse set of classes compared to Office-Caltech. The Office-31 dataset was originally introduced for visual domain adaptation research [13]. These datasets capture variations in image acquisition settings, backgrounds, and viewpoints, making them effective for assessing the robustness of domain adaptation methods in real-world scenarios. In our approach, one dataset is selected as the target domain, and the remaining datasets are used as source domains for training.

For datasets that do not provide a predefined train/test split, we randomly sample 20% of the data as the test set to ensure a consistent evaluation protocol across benchmarks. By performing cross-validation across each dataset, we assess the generalizability of the model's performance.

4.2 Experimental Settings

To evaluate the performance of MSDA and verify the effectiveness of the proposed EDTM, we followed three experimental protocols widely adopted in previous studies. The protocols are as follows.

Single-best refers to the approach where adaptation to the target domain is performed using each of the four source domains individually, and the result with the highest accuracy is reported. This represents the traditional single-source domain adaptation method.

Source-combine is an approach in which data from multiple source domains are combined into one, and adaptation is performed based on the integrated data. This setting evaluates performance when information from diverse domains is merged.

Multi-source utilizes MSDA techniques to effectively learn from multiple source domains and transfer the knowledge to the target domain.

All experiments were repeated three times, and the average accuracy was used as the final evaluation metric. The common training settings were as follows: the learning rate was set to 2×10^{-4} , and the batch size was 16. The Adam optimizer was used with its first-moment decay parameter $\beta_1 = 0.5$. In Adam, the β parameters control the exponential decay rates for the moving averages of gradients and squared gradients, thereby stabilizing updates and smoothing noisy gradient signals during training [14]. Following prior work, we set $\gamma = 3$ for EDTM [4]. All experiments were conducted on a system equipped with an Intel i5-13500H CPU and an NVIDIA RTX 2080 GPU.

In addition, Table 1 summarizes the architecture used in our experiments on the *Digits-Five* dataset. The “Feature Extractor” describes the shared feature-learning backbone, the “Classifier” specifies the architecture of each individual classifier used for domain-specific prediction, and the “Domain Discriminator” corresponds to the network structure employed for domain classification. This table provides detailed layer configurations to support model interpretability and reproducibility.

Table 1: Model architecture for the feature extractor, classifier, and domain discriminator.

Module	Layer	Details/Parameters
Feature Extractor	Instance Normalization	–
	Conv2D	64 filters, 3 × 3 kernel, stride = 1, activation = ReLU
	Conv2D	64 filters, 3 × 3 kernel, stride = 1, activation = ReLU
	Conv2D	64 filters, 3 × 3 kernel, stride = 1, activation = ReLU
	MaxPool	pool size = (2, 2)
	Dropout	enabled during training
	Conv2D	64 filters, 3 × 3 kernel, stride = 1, activation = ReLU
	Conv2D	64 filters, 3 × 3 kernel, stride = 1, activation = ReLU
	Conv2D	64 filters, 3 × 3 kernel, stride = 1, activation = ReLU
	MaxPool	pool size = (2, 2)
Classifier	Dropout	enabled during training
	Conv2D	64 filters, 3 × 3 kernel, stride = 1, activation = ReLU
	Conv2D	64 filters, 3 × 3 kernel, stride = 1, activation = ReLU
	Conv2D	64 filters, 3 × 3 kernel, stride = 1, activation = ReLU
	Global Average Pooling	enabled
Domain Discriminator	Dense	10 output classes, activation = None
	Dense	100 units, activation = ReLU
	Dense	4 units, activation = None

4.3 Experiment Results on Digits-Five

Table 2 provides a detailed comparative analysis of domain adaptation methods evaluated on the *Digits-Five* benchmark, revealing distinct behavioral characteristics across the *Single-best*, *Source-combine*, and *Multi-source* settings.

Table 2: Experiment result on the Digits-Five dataset.

Model	mt	mm	sv	sy	up	Avg
Single-best						
Source-only	98.8	53.0	81.3	93.0	84.6	82.1
GAN [15]	99.0	84.4	81.9	94.1	90.8	90.0
DANN [5]	97.6	71.3	63.5	85.4	92.3	82.0
ADDA [6]	97.9	71.6	75.5	86.5	92.8	84.8
MCD [4]	99.1	74.3	89.5	45.3	98.0	81.2
Source-combine						
Source-only	98.6	73.6	87.0	96.1	94.6	89.9
GAN [15]	96.1	64.0	84.2	93.6	90.0	85.5
DANN [5]	97.9	70.8	68.5	87.4	93.5	83.6
ADDA [6]	97.9	72.3	75.0	86.7	93.1	85.0

(Continued)

Table 2 (continued)

Model	mt	mm	sv	sy	up	Avg
MCD [4]	96.2	72.5	78.9	87.5	95.3	86.1
Multi-source						
MDAN [16]	98.0	69.5	69.2	87.4	92.4	83.3
MDDA [7]	98.8	78.6	89.3	89.7	93.9	88.1
LtC-MSDA [8]	99.0	85.6	83.2	93.0	98.3	91.8
STEM [3]	99.5	87.2	87.4	97.0	97.5	93.7
MIAN [17]	98.0	84.0	88.0	93.0	96.0	92.0
OT MSDA [18]	99.2	85.3	88.6	95.0	99.1	93.4
EDTM (ours)	99.4	97.8	92.1	98.4	97.8	97.1

In the *Single-best* setting, each model is trained using a single source domain that yields the highest average performance. Under this setting, the GAN-based method achieves the best average accuracy among baseline approaches (90.0%), exhibiting strong performance on the *mt*, *sy*, and *up* domains. Nevertheless, its performance drops noticeably on the *mm* and *sv* domains (84.4% and 81.9%), indicating limited robustness to domain-specific visual discrepancies. Similar instability is observed in other single-source methods such as DANN, ADDA, and MCD. In particular, DANN and ADDA struggle significantly on the visually challenging *mm* and *sv* domains, while MCD suffers a severe performance collapse on the *sy* domain despite maintaining high accuracy on other targets. These results highlight a fundamental limitation of single-source adaptation: its effectiveness heavily depends on the latent compatibility between the selected source and the target domain. However, in unsupervised domain adaptation, such compatibility cannot be determined in advance due to the absence of target labels, which often leads to unstable optimization or insufficient feature alignment, as clearly evidenced by the *sy*-domain failure of MCD.

In the *Source-combine* setting, all source domains are jointly used for training, leading to improved generalization in many cases. The Source-only baseline, for example, shows a substantial increase in average accuracy to 89.9%, confirming that exposure to diverse source distributions is beneficial for learning more transferable representations. However, this improvement is not universal across all methods. Notably, GAN experiences a significant performance degradation compared to its Single-best counterpart, particularly on the *mm* domain, and MCD also exhibits reduced accuracy on specific targets such as *sv*. These observations indicate that naive aggregation of source domains can introduce negative transfer, where incompatible or conflicting knowledge from different sources interferes with effective feature learning. As a result, indiscriminate mixing of heterogeneous domains may obscure discriminative structures and lead to performance degradation, even relative to carefully selected single-source training.

In contrast, the *Multi-source* setting demonstrates clear advantages by explicitly modeling relationships among multiple source domains. Early MSDA methods such as MDAN and MDDA show moderate improvements but still exhibit weaknesses on challenging domains. More advanced approaches, including MIAN, OT MSDA, and STEM, achieve stronger and more stable performance by selectively weighting or aligning source contributions. MIAN attains an average accuracy of 92.0% with relatively balanced performance across all target domains. OT MSDA further improves the overall accuracy to 93.44%, achieving exceptionally high results on the *mm* and *up* domains (99.2% and 99.1%), which demonstrates the effectiveness of weighted domain alignment strategies. STEM achieves an average accuracy of 93.7% and serves as a strong benchmark by modeling domain divergence from an information-theoretic perspective.

Compared to STEM, the proposed EDTM achieves substantial and consistent performance gains across all target domains. In particular, EDTM significantly outperforms STEM on the most challenging domains, achieving 97.8% on *mm*, 92.1% on *sv*, and 97.8% on *up*. These improvements correspond to gains of +10.6%, +4.7%, and +0.3% over STEM, respectively. With an overall average accuracy of 97.1%, EDTM surpasses the previous best method by a large margin and establishes a new state-of-the-art on the Digits-Five benchmark. The results indicate that EDTM effectively mitigates negative transfer by selectively integrating complementary knowledge from multiple heterogeneous sources through expert-based learning and discrepancy-guided training. Consequently, EDTM achieves robust generalization even under severe domain shifts and pronounced style mismatches.

4.4 Experiment Results on Office-Series Datasets

Tables 3 and 4 report the performance comparison between the proposed EDTM and the state-of-the-art MSDA baseline, STEM, on the Office-series datasets. Although both datasets share the same high-level objective, they exhibit different degrees of domain heterogeneity, enabling an analysis of not only absolute accuracy but also the stability of adaptation across domains under practical experimental constraints.

Table 3: Experiment result on the Office-Caltech dataset.

Model	webcam	dslr	caltech	amazon	Avg
STEM [3]	99.7	100.0	93.5	96.3	97.4
EDTM (ours)	99.7	100.0	94.6	96.3	97.6

Table 4: Experiment result on the Office-31 dataset.

Model	webcam	dslr	amazon	Avg
STEM [3]	71.6	83.1	–	77.4
EDTM (ours)	80.3	82.7	–	81.5

On the *Office-Caltech* dataset, both methods achieve near-saturated performance on the visually similar and relatively low-variance domains (*webcam* and *dslr*), reaching 99.7%–100.0%. Under such conditions, the primary performance differentiator becomes the most challenging domain, *caltech*, whose visual characteristics deviate more substantially from the remaining sources. EDTM improves the *caltech* accuracy from 93.5% (STEM) to 94.6%, corresponding to a +1.1% gain, and consequently increases the overall average accuracy from 97.4% to 97.6%. An examination of per-domain performance gaps further indicates that domain heterogeneity still affects adaptation: EDTM exhibits a 5.4% gap between *caltech* (94.6%) and the saturated *dslr* domain (100.0%), whereas STEM shows a larger gap of 6.5%. This result suggests that even when most domains are relatively easy, effective multi-source adaptation requires selectively integrating complementary source knowledge while avoiding domination by over-confident or overly similar sources. Therefore, the Office-Caltech results emphasize that MSDA performance should be evaluated not only by average accuracy but also by how effectively residual errors on the most domain-shifted targets are reduced.

The *Office-31* dataset represents a more challenging adaptation scenario due to larger cross-domain distribution shifts and more pronounced appearance variations. In our experimental setting, reliable domain adaptation to the *amazon* domain was not achieved, and thus results for this domain are not reported. This limitation reflects the inherent difficulty of the *amazon* domain, which exhibits substantial visual

and semantic discrepancies from other sources under the considered MSDA configuration. Consequently, the analysis focuses on domains where stable adaptation behavior can be meaningfully assessed. Within this setting, EDTM significantly outperforms STEM on the *webcam* domain, achieving 80.3% compared to 71.6% (+8.7%). This improvement leads to an increase in the reported average accuracy from 77.4% to 81.5% (+4.1%). The substantial gain on the *webcam* domain is particularly informative, as visually constrained domains tend to amplify the risk of negative transfer when heterogeneous sources are combined indiscriminately. The observed improvement indicates that EDTM more effectively suppresses harmful inter-source interference and produces a more reliable alignment for the target domain. From the perspective of unsupervised domain adaptation, this robustness is critical because the optimal source–target compatibility cannot be identified in the absence of target labels. As a result, stability under domain heterogeneity, rather than sensitivity to a specific favorable source pairing, becomes a key criterion for robust MSDA.

Overall, the Office-series results demonstrate that EDTM consistently improves upon a strong MSDA baseline across both a near-saturated benchmark (Office-Caltech) and a more heterogeneous benchmark (Office-31), even under practical domain adaptation constraints. The performance gains are concentrated on the most domain-shifted and challenging targets (e.g., *caltech* in Office-Caltech and *webcam* in Office-31), supporting the claim that EDTM more effectively mitigates negative transfer and selectively integrates complementary source knowledge, leading to more stable and reliable multi-source domain adaptation.

4.5 Feature Visualization

To analyze the distributional differences between domains and to evaluate how effectively EDTM reduces this discrepancy, we visualized the learned feature representations using the t-SNE algorithm [19]. Following this methodology, we compare the feature distributions before and after domain adaptation on the Digits-Five benchmark. In this experiment, MNIST-M is used as the target domain, while the remaining datasets—MNIST, SVHN, SYN, and USPS—are utilized as source domains.

Fig. 2 illustrates the t-SNE results for both the Source-only model and the proposed EDTM. In the Source-only model (left), the source (red) and target (blue) samples are heavily entangled without forming clear class-wise clusters. The target MNIST-M features, in particular, appear densely concentrated near the center, indicating that the model fails to extract meaningful representations due to the strong domain shift introduced by texture and color perturbations.

In contrast, the EDTM results (right) show that samples from all source domains and the target domain form distinguishable and well-aligned clusters. The target features closely follow the structure of the source clusters, demonstrating that EDTM effectively integrates information from multiple source domains and reduces the domain discrepancy. This alignment confirms that EDTM successfully transfers discriminative knowledge to the target domain, enabling class-consistent feature extraction even without access to target labels.

4.6 Ablation Study

We conducted an ablation study to evaluate the contribution of each loss function in EDTM to the overall model performance. Table 5 presents the variation in accuracy as each loss component is incrementally incorporated.

Initially, employing only the loss functions L_D and L_{C_k} from the Ensemble-based Classifier Expert resulted in an accuracy of 87.5%. This outcome indicates that the model effectively integrates knowledge from multiple source domains. However, this configuration alone was insufficient to fully mitigate the distributional shift between the source and target domains.

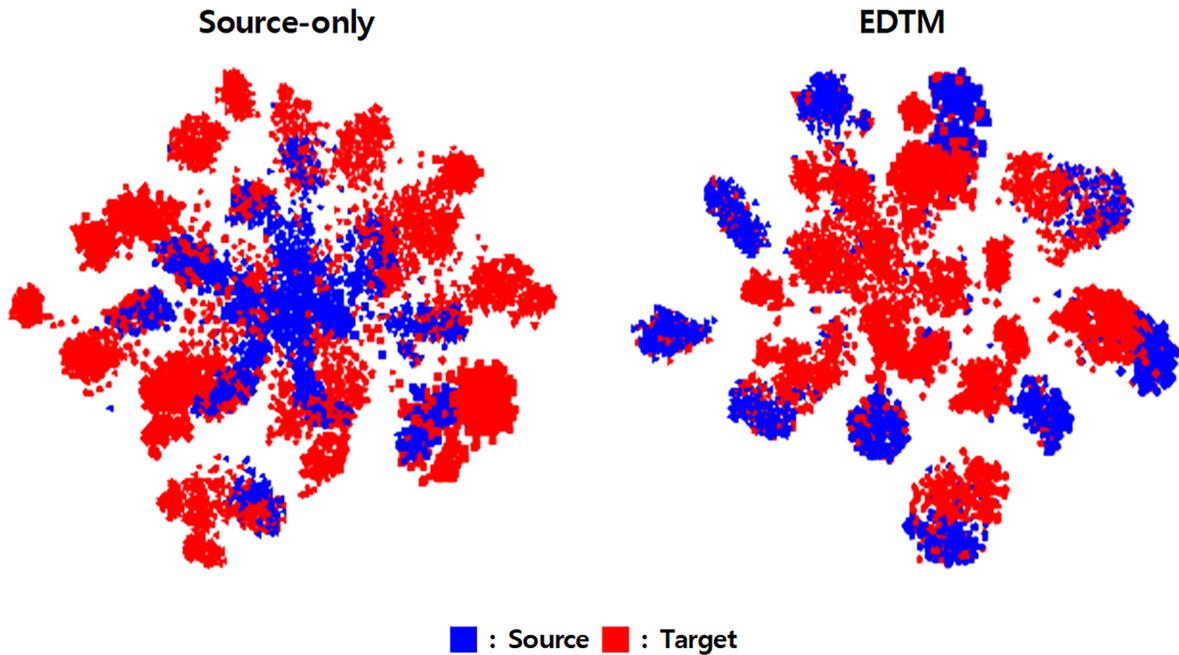


Figure 2: Feature visualization using t-SNE on the Digits-Five benchmark. Source domains (red): MNIST, SVHN, SYN, USPS; Target domain (blue): MNIST-M. Left: Source-only model showing severe domain mismatch. Right: EDTM achieving clear class-wise alignment across domains.

Table 5: Ablation study on EDTM.

L_D, L_{C_k}	L_I	L_M, L_{MCD}, L_F	Avg
✓			87.5
✓	✓		93.3
✓		✓	93.4
✓	✓	✓	97.1

The inclusion of the imitation learning loss L_I led to a substantial improvement in accuracy, reaching 93.3%. This result highlights the pivotal role of imitation learning in adapting effectively to the target domain.

Similarly, augmenting the base configuration (L_D and L_{C_k}) with the MCD-based loss functions L_M , L_{MCD} , and L_F further enhanced performance to 93.4%. This finding suggests that these components are effective in aligning the label-wise distributions between the source and target domains.

Finally, incorporating all proposed loss functions yielded the highest accuracy of 97.1%. These results demonstrate that each component contributes effectively to model performance, while jointly addressing different challenges in a complementary manner, thereby generating stronger learning synergy and enabling effective integration of the distributions between the source and target data.

5 Conclusion

In this paper, we propose the EDTM framework, which effectively addresses two key processes: integrating knowledge from diverse source domains and applying it to the target domain. Specifically, EDTM incorporates three main components. First, the Ensemble-based Classifier Expert assigns greater importance to source domains that are more similar to the target domain, allowing the model to emphasize

more transferable knowledge during training. Second, Imitation Learning enables the model to adapt more effectively to the target domain by learning target-specific patterns based on pseudo-labels generated from the ensemble expert. Third, the MCD-based loss refines the alignment between the source and target domains by adjusting the label-wise distribution, ensuring fine-grained class-level adaptation.

These components are jointly optimized within EDTM, resulting in a model that robustly generalizes across domain shifts. Experimental results on the *Digits-Five* benchmark demonstrate that EDTM achieves state-of-the-art performance with an average accuracy of 97.1%, outperforming STEM by 3.8%. The improvements are especially notable on challenging domains such as *mm* and *sv*, where EDTM substantially surpasses previous approaches. Experiments on the Office-series datasets further confirm the robustness of the proposed method, with EDTM achieving the highest or near-highest accuracy across all evaluation settings.

The feature visualization using t-SNE highlights the effectiveness of domain alignment. Whereas the Source-only baseline shows scattered and ambiguous clusters for target samples, EDTM produces clear and well-separated feature distributions, indicating successful knowledge transfer and improved discriminability. The ablation study further verifies the contribution of each component. Starting with only L_D and L_{C_k} , the model achieved 87.5% accuracy. Adding Imitation Learning increased accuracy to 93.3%, while incorporating the MCD-based losses resulted in 93.4%. Combining all proposed losses yielded the best performance of 97.1%, demonstrating their complementary roles.

While MSDA naturally introduces more components than single-source adaptation, prior studies have consistently shown that leveraging multiple source domains can significantly reduce domain bias and improve generalization robustness [3,7,8]. Furthermore, EDTM maintains a computational structure with a time complexity of $\mathcal{O}(n)$ per iteration, as all major operations scale linearly with the number of samples used during training. This efficiency makes EDTM suitable for realistic scenarios where data originate from multiple sensors, environments, or acquisition conditions, and where lightweight yet effective multi-domain adaptation is required.

Recent MSDA research has increasingly explored transformer-based architectures and self-supervised representation learning, aiming to capture richer cross-domain dependencies and to reduce reliance on labeled source data. Although EDTM employs a classifier-driven design rather than transformer encoders or contrastive self-supervision, its lightweight structure offers a complementary direction that emphasizes stability, interpretability, and efficiency. Integrating EDTM's ensemble-driven domain weighting with transformer-based feature extractors or self-supervised pretraining may further enhance adaptation performance, suggesting a promising avenue for future exploration.

Despite its strong empirical performance, EDTM has limitations that could be addressed in future work. The framework relies on feature extractor quality and assumes that source domains contain sufficiently diverse knowledge to guide adaptation, which may be restrictive in extremely heterogeneous domain settings. Moreover, EDTM does not explicitly model long-range cross-domain interactions that transformer-based MSDA frameworks can capture. Future studies may investigate hybrid architectures that combine EDTM with transformer-based feature backbones, incorporate self-supervised objectives for improved representation learning, or extend EDTM to large-scale, high-resolution, or open-set domain adaptation scenarios.

Beyond these technical directions, an important next step is to evaluate EDTM on a broader collection of benchmark datasets and to validate its scalability in more diverse environments. In addition, applying EDTM to real-world data collected from practical industrial or manufacturing processes represents a promising future direction. Such extensions would not only demonstrate the robustness of EDTM in operational

settings but also provide valuable insights into how MSDA techniques can be deployed in real applications requiring cross-domain generalization.

In summary, EDTM effectively combines ensemble learning, imitation learning, and distribution alignment into a computationally efficient MSDA framework. Strong performance across the *Digits-Five* and Office-series benchmarks demonstrates that EDTM not only surpasses existing methods but also provides a well-balanced approach between performance and system complexity, while offering a foundation that can be further enhanced in future transformer-based, self-supervised, or real-world MSDA research directions.

Acknowledgement: Not applicable.

Funding Statement: This work was partly supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2024-00406320) and by the Institute of Information & Communications Technology Planning & Evaluation (IITP)-Innovative Human Resource Development for Local Intellectualization Program Grant funded by the Korea government (MSIT) (IITP-2026-RS-2023-00259678).

Author Contributions: Conceptualization, Jaekyun Jeong and Jungeun Kim; data curation, Jaekyun Jeong and Mangyu Lee; formal analysis, Jaekyun Jeong and Mangyu Lee; investigation, Jaekyun Jeong and Mangyu Lee; methodology, Jaekyun Jeong and Jungeun Kim; validation, Jaekyun Jeong, Yun Wook Choo, Keejun Han and Jungeun Kim; writing—original draft, Jaekyun Jeong and Jungeun Kim; writing—review and editing, Jaekyun Jeong, Mangyu Lee, Yun Wook Choo, Keejun Han and Jungeun Kim. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The data supporting the findings of this study are openly available from the official websites of the publicly available benchmark datasets used in this study. No new datasets were created in this study. The experiments were conducted using existing publicly available datasets, including MNIST (<http://yann.lecun.com/exdb/mnist/>), MNIST-M (<https://github.com/pumpikano/tf-dann>), SVHN (<http://ufldl.stanford.edu/housenumbers/>), USPS (<https://iee-dataport.org/documents/usps-handwritten-digits>), Synthetic Digits (SYN) (<https://github.com/pumpikano/tf-dann>), Office-31 (<https://www.cc.gatech.edu/~judy/domainadapt/>), and Office-Caltech (<https://www.cc.gatech.edu/~judy/domainadapt/>).

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst.* 2012;25:84–90. doi:10.1145/3065386.
2. Quionero-Candela J, Sugiyama M, Schwaighofer A, Lawrence ND. *Dataset shift in machine learning.* Cambridge, MA, USA: MIT Press; 2009.
3. Nguyen V-A, Nguyen T, Le T, Tran QH, Phung D. STEM: an approach to multi-source domain adaptation with guarantees. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision.* Piscataway, NJ, USA: IEEE; 2021. p. 9352–63. doi:10.1109/ICCV48922.2021.00922.
4. Saito K, Watanabe K, Ushiku Y, Harada T. Maximum classifier discrepancy for unsupervised domain adaptation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* Piscataway, NJ, USA: IEEE; 2018. p. 3723–32. doi:10.1109/CVPR.2018.00392.
5. Ganin Y, Ustinova E, Ajakan H, Germain P, Larochelle H, Laviolette F, et al. Domain-adversarial training of neural networks. *J Mach Learn Res.* 2016;17(59):1–35. doi:10.1007/978-3-319-58347-1_10.
6. Tzeng E, Hoffman J, Saenko K, Darrell T. Adversarial discriminative domain adaptation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* Piscataway, NJ, USA: IEEE; 2017. p. 7167–76. doi:10.1109/CVPR.2017.316.

7. Zhao S, Wang G, Zhang S, Gu Y, Li Y, Song Z, et al. Multi-source distilling domain adaptation. Proc AAAI Conf Artif Intell. 2020;34(7):12975–83. doi:10.1609/aaai.v34i07.6997.
8. Wang H, Xu M, Ni B, Zhang W. Learning to combine: knowledge aggregation for multi-source domain adaptation. In: Computer Vision-ECCV 2020: European Conference on Computer Vision. Cham, Switzerland: Springer; 2020. p. 727–44. doi:10.1007/978-3-030-58598-3_43.
9. LeCun Y, Cortes C, Burges C. The MNIST database of handwritten digits [Internet]. 1998 [cited 2026 Jan 10]. Available from: <http://yann.lecun.com/exdb/mnist/>.
10. Netzer Y, Wang T, Coates A, Bissacco A, Wu B, Ng AY. Reading digits in natural images with unsupervised feature learning. NIPS Workshop Deep Learn Unsupervised Feature Learn. 2011;2011(5):7. doi:10.1109/icdar.2011.95.
11. Hull JJ. A database for handwritten text recognition research. IEEE Trans Pattern Anal Mach Intell. 1994;16(5):550–4. doi:10.1109/34.291440.
12. Gong B, Shi Y, Sha F, Grauman K. Geodesic flow kernel for unsupervised domain adaptation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE; 2012. p. 2066–73. doi:10.1109/CVPR.2012.6247911.
13. Shekhar S, Patel VM, Nasrabadi NM, Chellappa R. Analysis of visual domain adaptation with sample transfer. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Piscataway, NJ, USA: IEEE; 2013. p. 354–61.
14. Kingma DP, Ba J. Adam: a method for stochastic optimization. arXiv:1412.6980. 2014.
15. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. arXiv:1406.2661. 2014.
16. Zhao H, Zhang S, Wu G, Moura JMF, Costeira JP, Gordon GJ. Adversarial multiple source domain adaptation. 2018 [cited 2026 Jan 10]. Available from: <https://papers.neurips.cc/paper/8075-adversarial-multiple-source-domain-adaptation.pdf>.
17. Park GY, Lee SW. Information-theoretic regularization for multi-source domain adaptation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway, NJ, USA: IEEE; 2021. p. 9214–23. doi:10.1109/ICCV48922.2021.00908.
18. Hu B, Wang J. A weighted multi-source domain adaptation approach for surface defect detection. IET Image Process. 2022;16(8):2210–8. doi:10.1049/ipr2.12484.
19. Van der Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res. 2008;9:2579–605.