ROAD: Robust Unsupervised Domain Adaptation with Noisy Labels

Yanglin Feng
fcyzff@163.com
Sichuan University
Chengdu, China

Hongyuan Zhu
hongyuanzhu.cn@gmail.com
Institute for Infocomm Research (I²R),
A’STAR
Singapore

Dezhong Peng∗
pengdz@scu.edu.cn
Sichuan University
Chengdu, China

Xi Peng
pengx.gm@gmail.com
Sichuan University
Chengdu, China

Peng Hu†
penghu.ml@gmail.com
Sichuan University
Chengdu, China

ABSTRACT
In recent years, Unsupervised Domain Adaptation (UDA) has emerged as a popular technique for transferring knowledge from a labeled source domain to an unlabeled target domain. However, almost all of the existing approaches implicitly assume that the source domain is correctly labeled, which is expensive or even impossible to satisfy in open-world applications due to ubiquitous imperfect annotations (i.e., noisy labels). In this paper, we reveal that noisy labels interfere with learning from the source domain, thus leading to noisy knowledge being transferred from the source domain to the target domain, termed Dual Noisy Information (DNI). To address this issue, we propose a robust unsupervised domain adaptation framework (ROAD), which prevents the network model from overfitting noisy labels to capture accurate discrimination knowledge for domain adaptation. Specifically, a Robust Adaptive Weighted Learning mechanism (RSWL) is proposed to adaptively assign weights to each sample based on its reliability to enforce the model to focus more on reliable samples and less on unreliable samples, thereby mining robust discrimination knowledge against noisy labels in the source domain. In order to prevent noisy knowledge from misleading domain adaptation, we present a Robust Domain-adapted Prediction Learning mechanism (RDPL) to reduce the weighted decision uncertainty of predictions in the target domain, thus ensuring the accurate knowledge of source domain transfer into the target domain, rather than uncertain knowledge from noise impact. Comprehensive experiments are conducted on three widely-used UDA benchmarks to demonstrate the effectiveness and robustness of our ROAD against noisy labels by comparing it with 13 state-of-the-art methods. Code is available at https://github.com/penghu-es/ROAD.

1 INTRODUCTION
Recent years have witnessed the success of deep learning in pushing forward the rapid development of multimedia applications, such as image captioning [19, 45], cross-modal retrieval [33, 35], etc. However, deep neural networks (DNNs) heavily rely on large labeled datasets and often struggle to generalize well to data from different domains, making them impractical for a new unlabeled domain in open-world scenarios. To overcome the challenge, Unsupervised Domain Adaptation (UDA) techniques are proposed to transfer knowledge from a labeled source domain to an unlabeled target domain in the presence of domain shift in data distribution.

While domain shift and label scarcity are significant obstacles in UDA, practical scenarios also involve additional challenges, one of which this paper mainly focuses on, namely noisy labels. To be specific, source domain data also frequently encounter various types of interference and noise, including but not limited to image motion blur, tailing, and sensor noise. These factors pose challenges for data annotation in various applications like medical image classification [48], industrial inspection [4, 22], and autonomous driving [29], resulting in the presence of certain label noise in the datasets. Moreover, the high cost of expert annotation and time drives more and more people to choose open-source annotation platforms and crowdsourcing for labeling, which can easily introduce noisy labels into the source data. Intuitively, noisy labels will inevitably mislead deep models to learn from the source domain, leading to performance degradation in the target domain.

CCS CONCEPTS

• Computing methodologies → Object recognition; Image representations.

KEYWORDS

Unsupervised domain adaptation, Learning with noisy labels, Image classification, Self-adaptive weighted scheme

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

MM ’23, October 29–November 3, 2023, Ottawa, ON, Canada.
© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-8-4007-0108-5/23/10...
https://doi.org/10.1145/3581783.3612296

1 INTRODUCTION
Recent years have witnessed the success of deep learning in pushing forward the rapid development of multimedia applications, such as image captioning [19, 45], cross-modal retrieval [33, 35], etc. However, deep neural networks (DNNs) heavily rely on large labeled datasets and often struggle to generalize well to data from different domains, making them impractical for a new unlabeled domain in open-world scenarios. To overcome the challenge, Unsupervised Domain Adaptation (UDA) techniques are proposed to transfer knowledge from a labeled source domain to an unlabeled target domain in the presence of domain shift in data distribution.

While domain shift and label scarcity are significant obstacles in UDA, practical scenarios also involve additional challenges, one of which this paper mainly focuses on, namely noisy labels. To be specific, source domain data also frequently encounter various types of interference and noise, including but not limited to image motion blur, tailing, and sensor noise. These factors pose challenges for data annotation in various applications like medical image classification [48], industrial inspection [4, 22], and autonomous driving [29], resulting in the presence of certain label noise in the datasets. Moreover, the high cost of expert annotation and time drives more and more people to choose open-source annotation platforms and crowdsourcing for labeling, which can easily introduce noisy labels into the source data. Intuitively, noisy labels will inevitably mislead deep models to learn from the source domain, leading to performance degradation in the target domain.

CCS CONCEPTS

• Computing methodologies → Object recognition; Image representations.

KEYWORDS

Unsupervised domain adaptation, Learning with noisy labels, Image classification, Self-adaptive weighted scheme

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

MM ’23, October 29–November 3, 2023, Ottawa, ON, Canada.
© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-8-4007-0108-5/23/10...
https://doi.org/10.1145/3581783.3612296

1 INTRODUCTION
Recent years have witnessed the success of deep learning in pushing forward the rapid development of multimedia applications, such as image captioning [19, 45], cross-modal retrieval [33, 35], etc. However, deep neural networks (DNNs) heavily rely on large labeled datasets and often struggle to generalize well to data from different domains, making them impractical for a new unlabeled domain in open-world scenarios. To overcome the challenge, Unsupervised Domain Adaptation (UDA) techniques are proposed to transfer knowledge from a labeled source domain to an unlabeled target domain in the presence of domain shift in data distribution.

While domain shift and label scarcity are significant obstacles in UDA, practical scenarios also involve additional challenges, one of which this paper mainly focuses on, namely noisy labels. To be specific, source domain data also frequently encounter various types of interference and noise, including but not limited to image motion blur, tailing, and sensor noise. These factors pose challenges for data annotation in various applications like medical image classification [48], industrial inspection [4, 22], and autonomous driving [29], resulting in the presence of certain label noise in the datasets. Moreover, the high cost of expert annotation and time drives more and more people to choose open-source annotation platforms and crowdsourcing for labeling, which can easily introduce noisy labels into the source data. Intuitively, noisy labels will inevitably mislead deep models to learn from the source domain, leading to performance degradation in the target domain.

CCS CONCEPTS

• Computing methodologies → Object recognition; Image representations.

KEYWORDS

Unsupervised domain adaptation, Learning with noisy labels, Image classification, Self-adaptive weighted scheme

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

MM ’23, October 29–November 3, 2023, Ottawa, ON, Canada.
© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-8-4007-0108-5/23/10...
https://doi.org/10.1145/3581783.3612296

1 INTRODUCTION
Recent years have witnessed the success of deep learning in pushing forward the rapid development of multimedia applications, such as image captioning [19, 45], cross-modal retrieval [33, 35], etc. However, deep neural networks (DNNs) heavily rely on large labeled datasets and often struggle to generalize well to data from different domains, making them impractical for a new unlabeled domain in open-world scenarios. To overcome the challenge, Unsupervised Domain Adaptation (UDA) techniques are proposed to transfer knowledge from a labeled source domain to an unlabeled target domain in the presence of domain shift in data distribution.

While domain shift and label scarcity are significant obstacles in UDA, practical scenarios also involve additional challenges, one of which this paper mainly focuses on, namely noisy labels. To be specific, source domain data also frequently encounter various types of interference and noise, including but not limited to image motion blur, tailing, and sensor noise. These factors pose challenges for data annotation in various applications like medical image classification [48], industrial inspection [4, 22], and autonomous driving [29], resulting in the presence of certain label noise in the datasets. Moreover, the high cost of expert annotation and time drives more and more people to choose open-source annotation platforms and crowdsourcing for labeling, which can easily introduce noisy labels into the source data. Intuitively, noisy labels will inevitably mislead deep models to learn from the source domain, leading to performance degradation in the target domain.
Therefore, we have to consider not only domain shifts but also noisy labels in UDA, which however is less touched so far.

Existing studies on learning with noisy labels [30, 38, 49] demonstrate that DNNs tend to overfit the corrupted labels, resulting in performance degradation. Similarly, the presence of noisy labels within the source domain inevitably leads to the overfitting issue during training on that domain, resulting in the transfer of noisy knowledge to target domains. Specifically, the presence of noisy knowledge amplifies prediction uncertainty/noise in the target domain, hindering the learning of correct discrimination from target domains and consequently decreasing the performance of domain adaptation. We refer to this cascading adverse effect caused by noisy labels in the source domain as Dual Noisy Information (DNI), which is depicted schematically in Figure 1. Notably, DNI poses a greater challenge than learning with noisy labels in a single domain, as it requires simultaneous mitigation of noisy information across different domains.

To tackle this challenge, this paper proposes a robust unsupervised domain adaptation framework (ROAD) to convey the discrimination knowledge from a source domain to target domains as shown in Figure 2, which consists of a novel Robust Self-adaptive Weighted Learning mechanism (RSWL) and a Robust Domain-adapted Prediction Learning mechanism (RDPL). To be specific, RSWL is proposed to alleviate the negative impact of noisy labels in the source domain by dynamically assigning weights to each sample based on its reliability. This adaptive weighting scheme aims to pay more attention to reliable samples and less to unreliable ones in the source domain. To achieve this, our RSWL divides the noisy data into clean and noisy (aka reliable and unreliable) partitions by mutual information between predictions and ground-truth labels and self-information of predictions in the source domain based on the memorization effect of Deep Neural Networks (DNNs) [1]. By adaptively assigning higher learning weights to reliable samples and lower weights to unreliable samples, RSWL mitigates the impact of noisy labels, enabling the learning of robust discrimination knowledge in the source domain. On the other hand, RDPL is presented to reduce weighted prediction uncertainty in the target domain, thereby facilitating the robust transfer of discrimination knowledge from the source domain to the target domain. To achieve this, RDPL assigns higher weights to samples that contain significant self-information of predictions, thus making the model focus on minimizing decision uncertainty within similar categories for target domains, instead of chaotic uncertainty obtained from noise impact. Consequently, thanks to RSWL and RDPL, our ROAD could simultaneously reduce the adverse impacts of both noisy labels and noisy predictions (i.e., DNI) to achieve a robust transfer of discrimination knowledge.

Overall, this paper makes four main contributions:

- We reveal and summarize the existence of cascading adverse effects caused by noisy labels in the source domain for unsupervised domain adaptation, termed Dual Noisy Information (DNI). To solve DNI, we propose a robust unsupervised domain adaptation framework (ROAD) to capture accurate discrimination knowledge for robust domain adaptation. To the best of our knowledge, this work could be the first study on the problem.
- To mitigate the impact of noisy labels in the source domain, a novel Robust Self-adaptive Weighted Learning mechanism (RSWL) is proposed to dynamically assign learning weights for each sample based on its reliability to enforce the model focus on reliable samples, thereby mining robust discrimination knowledge against noisy labels in the source domain.
- To prevent noisy knowledge from misleading the domain-adaptive learning on target domains, a Robust Domain-adapted Prediction Learning mechanism (RDPL) is proposed to minimize the weighted prediction uncertainty in the target domain, thus preferentially facilitating the robust transfer of reliable discrimination knowledge from the source domain, instead of chaotic knowledge from noise impact.
- We experimentally demonstrate the robustness and effectiveness of our ROAD against symmetric/asymmetric noisy labels in the source domain. Our method shows an impressive performance in three multi-domain image datasets, outperforming the state-of-the-art UDA methods without bells and whistles.

2 RELATED WORKS

2.1 Unsupervised Domain Adaptation

Over the past decade, numerous transferable methods have been proposed to tackle the issue of performance degradation in deep neural networks (DNNs) caused by inherent domain shifts between labeled source and unlabeled target domains. These methods could be roughly grouped into three categories: 1) Discrepancy-based methods treat cross-domain data as distinct distributions and aim to mitigate domain shifts by minimizing the differences between
each pair of distributions. To achieve this, they utilize various metrics to measure the dissimilarity between distributions, such as Maximum Mean Discrepancy (MMD) [11, 24] and its variants [25, 42, 51]. By minimizing the discrepancies, these methods attempt to align the distributions across distinct domains. 2) Adversarial-based methods tackle the problem of domain shifts by introducing adversarial learning in the training process, such as adversarial-based methods [7, 27, 36], fine-grained generating adversarial-based methods [10, 44, 46, 47], etc. 3) Self-supervision-based methods address the issue of domain shifts by leveraging pre-trained DNNs on labeled source domain data to obtain pseudo-labels in the target domain. The pseudo-labels are seen as noisy labels and then used to retrain the DNNs using the robust learning paradigm [5, 8, 15, 52]. However, they heavily rely on the well-labeled source domain, which can lead to overfitting issues in the presence of noisy labels.

2.2 Learning with Noisy Labels

To address the problem of ubiquitous imperfect annotations during training, various methods have been proposed to mitigate the adverse effect of noisy labels and improve the robustness of DNNs. These works could be roughly categorized into three classes: 1) Model-oriented methods focus on designing different DNN architectures that explicitly model the transformation matrices of label noise to capture the underlying patterns of label noise, thus enabling the network to learn effectively in the presence of noisy labels [2, 3]. 2) Sample-oriented methods leverage a small set with perfectly labeled data to learn priori clean discrimination to reweight samples [23] or refurbish labels [12, 18, 20], thus achieving robust learning against noisy labels. However, acquiring additional well-labeled data is also expensive or even impossible in some real-world applications, which directly limits the applicability of these methods. 3) Loss-oriented methods primarily focus on designing robust optimization objectives [9, 14, 28, 50] or regularization techniques [41, 43] that prevent DNNs from being corrupted by noisy labels during training. Although these loss functions are concise and theoretically sound in preventing DNNs from overfitting to noisy labels, they are designed specifically for single-domain learning, ignoring the adverse effect of dual noisy information.

3 THE PROPOSED METHOD

3.1 Problem Formulation

To ensure clarity and facilitate understanding, we begin by providing definitions for the notations used in this paper. Throughout the paper, boldface uppercase letters, boldface lowercase letters, and general uppercase letters represent matrices, vectors, and scalars, respectively. Let \( \mathcal{D} = \{ \mathcal{I}_s, \mathcal{I}_t \} \) denote a \( C \)-category source-target domain dataset. Here, \( \mathcal{I}_s = \{ X_s, Y_s \} = \{ x_{si}, y_{si} \}^{N_t}_{i=1} \) represents the source domain data with noisy labels, where \( N_t \) denotes the number of samples in the source domain, \( x_{si} \) and \( y_{si} \in \{ 1, 2, \cdots, C \} \) respectively are the \( i^{th} \) sample and its class label which is potentially corrupted. Similarly, \( \mathcal{I}_t = \{ X_t, Y_t \} = \{ x_{ti}, y_{ti} \}^{N_t}_{i=1} \) represents the target domain data, where \( x_{ti} \) denotes the \( i^{th} \) sample, and \( N_t \) denotes the number of samples in the target domain. We also define \( \{ \bar{Y}_s, \bar{Y}_t \} = \{ [\hat{y}^s_{ij}], [\hat{y}^t_{ij}] \}^{N_t}_{i=1, j=1} \) as the predictions of the domain-shared network \( f = [f_c, \cdots, f_c] \), \( f_i \) is the \( i^{th} \) class output of the domain-shared network, \( \bar{y}^s_{ij} = \frac{\exp(f_i(x^s_i)/\tau)}{\sum_k \exp(f_k(x^s_i)/\tau)} \) and \( \bar{y}^t_{ij} = \frac{\exp(f_i(x^t_i)/r)}{\sum_k \exp(f_k(x^t_i)/r)} \) are the prediction probability of \( i^{th} \) sample for \( j^{th} \) class in the source and target domain, respectively. \( \tau \) is the temperature parameter.

In unsupervised domain adaptation (UDA), the aforementioned corrupted label information in the source domain is inevitably introduced explicitly or implicitly into the cross-domain knowledge transfer, as illustrated in Figure 1. Consequently, the noise information in the source and target domains jointly and progressively dominates the training process, causing the networks to overfit the noisy labels in the source domain and resulting in a degradation in domain adaptation performance. This cascading adverse effect caused by noisy labels in the source domain, namely Dual Noisy Information (DNI), simultaneously corrupts the learning in both the source and target domains. It is evident that DNI poses greater challenges compared to traditional learning methods with single-domain noisy labels.

3.2 Overview of Method

To address the challenges posed by unsupervised domain adaptation in the presence of DNI, we propose a robust unsupervised domain adaptation framework (ROAD), which consists of two phases: the warm-up phase and the main training phase.

In the warm-up phase, we leverage the memorization effect of DNNs [1] and train the domain-shared network for a brief period of \( N_w \) epochs to obtain an initial reliable discriminative model. The objective function during this phase is defined as follows:

\[
L_{\text{warm-up}} = \frac{1}{N_s} \sum_{i=1}^{N_s} \text{cetp}^s_i, \quad (1)
\]

where \( \text{cetp}^s_i = -\sum_j y^s_{ij} \log \hat{y}^s_{ij} \) represents the general cross-entropy loss calculated for the \( i^{th} \) sample in the source domain.

In the main training phase, building upon the memorization effect of DNNs obtained from the warm-up phase, our ROAD incorporates two mechanisms: the Robust Self-adaptive Weighted Learning mechanism (RSWL) and the Robust Domain-adapted Prediction Learning mechanism (RDPL). Specifically, RSWL is designed to mitigate the negative impact of label noise in the source domain to learn discrimination knowledge robustly. Meanwhile, RDPL is employed to eliminate the decision uncertainty brought by domain adaptation and noisy knowledge, thereby facilitating reliable knowledge transfer from the source domain to the target domain. The joint utilization of RSWL and RDPL establishes a defense against domain noise labels, enabling our ROAD to effectively address unsupervised domain adaptation under the influence of DNI. The overall training objective function is defined as follows:

\[
L = L_{\text{wrec}} + \mu L_{\text{da}}, \quad (2)
\]

where \( L_{\text{wrec}} \) denotes the loss function employed by RSWL (see Equation (6)). \( L_{\text{da}} \) represents the loss function utilized by RDPL (see Equation (9)). \( \mu \) is the trade-off parameter controlling their relative importance. The optimization of our ROAD involves minimizing...
In order to mitigate the negative effects of noisy labels on predictions in the source domain, we propose a mechanism called Robust Self-adaptive Weighted Learning (RSWL) to prioritize reliable samples and de-emphasize unreliable ones. To achieve this, we first estimate the reliability of each sample by calculating the mutual information between predictions and labels, and subsequently assigning dynamic learning weights to each sample based on its reliability.

Recent works reveal that deep neural networks (DNNs) are inclined to fit simple patterns [1]. Specifically, after a short warm-up period, DNNs tend to provide correct predictions for simple (i.e., clean) samples that are close to the ground-truth labels, while incorrect predictions for challenging (i.e., noisy) ones that deviate significantly from the erroneous labels. This phenomenon, aka the memorization effect of DNNs [1], could be quantified as follows:

\[
V_{i}^{\text{etp}} = \frac{B \cdot e^{C} \cdot e^{C (1-e^{-C (\log \hat{y}_{i} - \log \bar{y}_{i})})}}{\sum_{j} e^{C} \cdot e^{C (1-e^{-C (\log \hat{y}_{i} - \log \bar{y}_{i})})}},
\]

where \(V_{i}^{\text{etp}}\) represents the prediction confidence of the \(i^{th}\) sample in the source domain, \(B\) denotes the size of the mini-batch, \(\log \hat{y}_{i} - \log \bar{y}_{i}\) is the entropy function, and \(C\) is a trade-off parameter.

Consequently, we combine the reliability \(V_{i}^{\text{etp}}\) and the confidence \(V_{i}^{\text{etp}}\) to weigh the cross-entropy loss for each sample. This adaptive weighting scheme emphasizes reliable and confident samples while de-emphasizes the unreliable and unconfident samples, resulting in the following weighted cross-entropy loss:

\[
L_{\text{wce}} = -\frac{\sum_{i} V_{i}^{\text{etp}} V_{i}^{\text{etp}} C \left( \log \hat{y}_{i} - \log \bar{y}_{i} \right)}{\sum_{i} V_{i}^{\text{etp}} V_{i}^{\text{etp}}}. \tag{5}
\]

This loss dynamically assigns weights to each sample based on the reliability and confidence of its prediction during discriminant learning in the source domain. Hence, it effectively prevents noisy
samples from dominating the training process in the source domain through the adaptive weighting scheme.

3.4 Robust Domain-adapted Prediction Learning

Although our RSWL could effectively address the issue of noisy labels in the source domain, achieving robust domain-adapted predictions remains a fundamental challenge in UDA with noisy labels. Additionally, blindly focusing on robustness and performance in the source domain while ignoring domain shifts may not lead to a corresponding improvement in the target domain.

Recent UDA methods suggest that domain shifts result in uncertain decisions of source-only networks between similar categories in the target domain [21]. This observation indicates that domain-adapted discrimination is influenced by the decision uncertainty in the target domain. To tackle this challenge, our Robust Domain-adapted Prediction Learning mechanism (RDPL), inspired by [17], models the degree of decision uncertainty for each target sample using the following equation:

$$T_i = \frac{1}{C} \sum_{j}^{C} \frac{\sum_{k \neq j}^{C} \hat{y}_{ik}^{t} \hat{y}_{jk}^{t}}{\sum_{k \neq j}^{C} \hat{y}_{ik}^{t} \hat{y}_{jk}^{t}}. \quad (6)$$

The objective is to minimize the average of the decision uncertainty term $T_i$ for all target samples, which facilitates obtaining domain-adapted discriminative predictions in the target domain. Thus, we define the vanilla domain-adaptation loss $L_{da}'$ as follows:

$$L_{da}' = \frac{1}{N_t} \sum_{i}^{N_t} T_i. \quad (7)$$

Unfortunately, the aforementioned decision uncertainty is not solely caused by domain shifts for DNI. The propagation of noisy knowledge from the source domain will result in chaotic predictions for same samples, significantly amplifying the decision uncertainty. However, eliminating the decision uncertainty of these contaminated samples could lead to an influx of unreliable discrimination knowledge, thereby misleading the domain adaptation process.

To address this challenge, we quantify the prediction confidence of each sample in the target domain using the self-information of predictions, i.e., the entropy. This quantification is similar to RSWL, and it is defined as follows:

$$W_{i}^{etp} = B \cdot e^{(\text{etp}_{i}^{t+1})}, \quad (8)$$

where $\text{etp}_{i}^{t} = -\sum_{j}^{C} \hat{y}_{ij}^{t} \log \hat{y}_{ij}^{t}$ is the entropy function.

Based on the above, we exploit the prediction confidence $W_{i}^{etp}$ to weigh the uncertainty of each sample, constructing the loss function $L_{da}$ for our RDPL, as shown below:

$$L_{da} = \frac{\sum_{i}^{N_t} (W_{i}^{etp} T_i)}{\sum_{j}^{N_t} W_{j}^{etp}}. \quad (9)$$

By minimizing this weighted decision uncertainty loss, we enforce the model focus on reducing the decision uncertainty of the samples with high confidence to alleviate the adverse effects of noisy knowledge, thus ensuring the robust transfer of discrimination knowledge to the target domain.

4 EXPERIMENTS

To thoroughly evaluate the effectiveness of our ROAD against dual noisy information (DNI), we conducted extensive comparative experiments on three widely used multi-domain datasets for image domain adaptation: Office-31 [34], Office-Home [37] and VisDA-2017 [31].

4.1 Experimental Settings

All experiments are performed on GeForce RTX 3090 GPUs, and all the reported quantitative results are the average of three runs for all the methods. Here is a brief introduction to the multi-domain datasets used in the experiments:

Office-31 [34]: This dataset is a widely-used multi-domain image dataset for image domain adaptation, consisting of 4,652 natural images from 31 categories. It comprises three image domains: Amazon (A), Webcam (W), and Dslr (D) with 2,817, 795, and 498 images, respectively. To enable bias-free evaluation, we evaluate all methods on 6 transfer tasks: A→D, D→W, D→A, D→W, W→A and W→D.

Office-Home [37]: This dataset is a larger multi-domain image dataset for image domain adaptation, containing 65 categories. It includes four image domains: Art (Ar), Clipart (Cl), Product (Pr), and Real World (Rw). Compared to Office-31, Office-Home presents a greater challenge due to more differences between different domains and a larger number of categories. To enable unbiased evaluation, we evaluate all methods on all 12 transfer tasks: Ar→Cl, Pr→Ar, Ar→Rw, Cl→Ar, Cl→Pr, Cl→Rw, Pr→Ar, Pr→Cl, Pr→Rw, Rw→Ar, Rw→Cl and Rw→Pr.

VisDA-2017 [31]: This dataset consists of a large volume of data belonging to 12 categories. It contains two image domains: 152,397 synthetic images (T) and 55,388 natural images (V). We build a transfer task: T→V as in [6, 17].

In the experiments, we compared our ROAD with 13 state-of-the-art methods, including ResNet50/101 [13], DAN [24], DANN [7], AFN [40], CDAN [26], TCM [46], DMAL [16], CAF [39], MCC [17], CGDM [6], SENTRY [32], SHOT [21], and CoUDA [48]. For comprehensive evaluations, two types of synthetic noise are adopted for comparison in our experiments, i.e., symmetric and asymmetric noisy labels. To be specific, we conducted extensive comparison experiments using symmetric noisy labels on the three datasets, and asymmetric noisy labels on VisDA-2017. In addition, we followed the established practices of previous works [17, 21] for selecting backbone networks and configuring the training and evaluation settings. To investigate the impact of different noise levels, we set the symmetric noise rates to 0.2, 0.4, 0.6, and 0.8, while the asymmetric noise rates were set to 0.1, 0.2, and 0.4.

4.2 Comparison with the State-of-the-Art Methods

We conducted extensive UDA experiments for image classification on three datasets to evaluate the performance of our ROAD in comparison to the 13 baselines. Due to space limitations, we present the average experimental results of all domain adaptation tasks on
Table 1: Performance comparison in terms of average accuracy (%) under the symmetric noise rates of 0, 0.2, 0.4, 0.6, and 0.8 on the Office-31 and Office-Home datasets. The highest accuracy is shown in bold and the second highest accuracy is underlined.

<table>
<thead>
<tr>
<th>Method</th>
<th>Office-31 [34]</th>
<th>Office-Home [37]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0% 20% 40% 60% 80%</td>
<td>0% 20% 40% 60% 80%</td>
</tr>
<tr>
<td>ResNet-50 [13]</td>
<td>76.1 63.7 53.7 37.3 19.2</td>
<td>46.1 39.7 30.0 22.4 12.3</td>
</tr>
<tr>
<td>DAN [24]</td>
<td>80.4 71.3 61.5 48.5 26.5</td>
<td>56.3 41.9 30.9 25.1 14.1</td>
</tr>
<tr>
<td>DANN [7]</td>
<td>82.2 69.3 56.3 43.3 22.2</td>
<td>57.6 42.0 29.8 22.6 10.5</td>
</tr>
<tr>
<td>APN [40]</td>
<td>85.7 80.8 71.3 62.8 41.2</td>
<td>67.3 63.2 59.2 46.9 28.4</td>
</tr>
<tr>
<td>CDAN [26]</td>
<td>87.7 76.6 66.5 52.2 28.2</td>
<td>63.8 50.4 38.9 28.4 14.2</td>
</tr>
<tr>
<td>TCM [46]</td>
<td>89.7 81.2 74.2 61.5 37.8</td>
<td>71.1 49.8 39.6 31.2 16.9</td>
</tr>
<tr>
<td>DML [16]</td>
<td>86.6 80.0 70.9 59.6 35.7</td>
<td>66.1 47.2 43.7 38.8 14.7</td>
</tr>
<tr>
<td>CAF [39]</td>
<td>88.6 76.7 64.8 52.1 32.7</td>
<td>69.0 42.4 34.5 28.4 17.7</td>
</tr>
<tr>
<td>MCC [17]</td>
<td>89.4 74.5 69.9 62.6 44.4</td>
<td>74.2 55.7 49.9 43.1 30.8</td>
</tr>
<tr>
<td>CGDM [6]</td>
<td>88.4 82.1 77.9 67.4 39.0</td>
<td>68.5 62.1 58.0 46.6 20.3</td>
</tr>
<tr>
<td>SENTRY [32]</td>
<td>87.3 81.9 74.9 66.0 35.3</td>
<td>72.3 63.3 59.8 49.8 31.6</td>
</tr>
<tr>
<td>SHOT [21]</td>
<td>88.6 82.0 75.6 63.5 29.9</td>
<td>71.8 65.8 61.3 51.8 33.8</td>
</tr>
<tr>
<td>CoUDA [48]</td>
<td>83.0 59.8 50.4 35.5 16.0</td>
<td>62.1 35.6 31.5 20.7 12.9</td>
</tr>
<tr>
<td>Ours</td>
<td>89.6 85.6 82.0 78.1 60.9</td>
<td>76.0 67.5 64.2 56.7 47.2</td>
</tr>
</tbody>
</table>

Table 2: Performance comparison in terms of average accuracy (%) under the symmetric noise rates of 0.2, 0.4, 0.6, 0.8 and asymmetric noise rates of 0, 0.1, 0.2, 0.4 on the VisDA-2017 dataset. The highest accuracy is shown in bold and the second highest accuracy is underlined.

<table>
<thead>
<tr>
<th>Method</th>
<th>Asymmetric</th>
<th>Symmetric</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0% 10% 20% 40%</td>
<td>0% 20% 40% 60% 80%</td>
</tr>
<tr>
<td>ResNet-101 [13]</td>
<td>52.4 51.6 46.1 39.7</td>
<td>36.6 31.9 17.2 14.2</td>
</tr>
<tr>
<td>DAN [24]</td>
<td>61.1 43.3 39.6 33.1</td>
<td>37.7 33.9 27.6 19.1</td>
</tr>
<tr>
<td>DANN [7]</td>
<td>59.2 47.1 43.9 45.3</td>
<td>49.6 43.3 33.7 18.3</td>
</tr>
<tr>
<td>APN [40]</td>
<td>76.1 53.9 46.0 34.7</td>
<td>44.8 36.9 31.5 22.0</td>
</tr>
<tr>
<td>CDAN [26]</td>
<td>66.8 58.0 52.2 52.3</td>
<td>64.0 51.4 43.0 19.2</td>
</tr>
<tr>
<td>TCM [46]</td>
<td>68.4 65.7 62.4 52.7</td>
<td>66.7 64.6 55.8 30.2</td>
</tr>
<tr>
<td>DML [16]</td>
<td>60.0 41.1 36.0 31.2</td>
<td>35.8 27.3 28.1 13.9</td>
</tr>
<tr>
<td>CAF [39]</td>
<td>80.3 62.0 62.0 49.0</td>
<td>54.1 50.3 50.8 32.8</td>
</tr>
<tr>
<td>MCC [17]</td>
<td>78.8 70.8 68.7 65.6</td>
<td>67.9 58.7 29.7 22.9</td>
</tr>
<tr>
<td>CGDM [6]</td>
<td>82.3 68.2 61.6 55.2</td>
<td>66.0 60.4 42.2 30.8</td>
</tr>
<tr>
<td>SENTRY [6]</td>
<td>69.2 66.9 64.0 60.2</td>
<td>52.7 53.0 40.1 21.3</td>
</tr>
<tr>
<td>SHOT [21]</td>
<td>74.6 69.5 68.9 62.0</td>
<td>70.5 59.0 45.1 22.7</td>
</tr>
<tr>
<td>CoUDA [48]</td>
<td>45.7 38.8 33.9 33.0</td>
<td>38.4 36.1 30.5 21.4</td>
</tr>
<tr>
<td>Ours</td>
<td>82.3 69.2 69.0 65.8</td>
<td>72.4 70.3 69.8 61.6</td>
</tr>
</tbody>
</table>

Each dataset as shown in Tables 1 and 2. Specifically, we report the experimental results for the Office-31 and Office-Home datasets under symmetric label noise in Table 1. The experimental results for the VisDA-2017 dataset under both symmetric and asymmetric label noise are shown in Table 2. Furthermore, the category-wide experimental results for the VisDA-2017 dataset with high noise rates (80% symmetric label noise and 40% asymmetric label noise) are presented in Table 3. From these results, we could draw the following observations:

1) The Dual Noisy Information (DNI) caused by source-domain noisy labels significantly degrades the domain-adaptation performance of each UDA baseline. Under high noise rates and challenging tasks, the performance degradation is so severe that some methods even perform worse than Resnet50/101 without any domain adaptation design.

2) In the face of chaotic synthetic symmetric noisy labels, our ROAD shows excellent robustness. Especially, our method could achieve 61.6% in terms of accuracy under 80% noise on the large-scale VisDA-2017 dataset, which is higher than the second-highest method CAF by 28.8%.

3) Despite the challenging asymmetric noise, which introduces highly disorienting class conditional noise that will weaken the memorization effect of DNNs, our ROAD still achieves superior robustness.

4) Although some methods (e.g., CGDN and SENTRY) maintain considerable average performance in the presence of high-level noise, as shown in Table 3, they suffer from unbalanced classification predictions and tend to exclude challenging categories due to the influence of noisy labels. In contrast, our ROAD achieves a well-balanced performance across all categories.

4.3 Ablation Study

In this section, we conduct an ablation study to investigate the contribution of each proposed loss (i.e., \( L_{\text{wce}} \) and \( L_{\text{da}} \), as well as each component within the losses (i.e., \( V^{\text{etp}}, V^{\text{celp}}, W^{\text{etp}} \) and \( T \)), and the warm-up phase to UDA with noisy labels in the source domain. We perform comparison experiments on six UDA tasks using the Office-31 dataset to thoroughly evaluate the contribution of ablating each component independently from our framework. In particular, to comprehensively study the robustness of \( L_{\text{wce}} \), we replace it with a general Cross-Entropy loss (CE) instead of detaching it. The results are presented in Table 4. From the table, one could make the following observation: 1) ROAD with/without any component leads to an improvement/drop in domain-adaptation performance, respectively, demonstrating the contribution of each component to our framework. 2) Replacing \( L_{\text{da}} \) with the vanilla loss of RDPL \( L_{\text{da}}' \) (i.e., \( L_{\text{da}} \) w/o \( W^{\text{etp}} \)) results in performance degradation, which demonstrates the significance of \( W^{\text{etp}} \) in enabling RDPL to mitigate the learning of unreliable discrimination knowledge propagated from the source domain to the target domain. 3) Substituting \( L_{\text{wce}} \) with CE leads to a remarkable drop in domain-adaptation performance, underscoring the robustness of our \( L_{\text{wce}} \), which assigns self-adaptive weights to clean and noisy data, particularly for challenging tasks (e.g., \( D_{\text{->A}} \) and \( W_{\text{->A}} \)).

Overall, the ablation study demonstrates the effectiveness of each component in our framework and emphasizes the crucial role of \( L_{\text{da}} \), \( L_{\text{wce}} \) and \( W^{\text{etp}} \) in addressing UDA with noisy labels.

4.4 Parameter Analysis

In order to evaluate the sensitivity of our ROAD to different hyperparameter settings, namely \( N_{\text{sp}} \) and \( \alpha \), we plot the accuracy scores for the parameter analysis on Office-31, as shown in Figure 3. Without loss of generality, we consider both low (20%) and high (60%) noise levels in all experiments. From Figure 3a, it could be observed that selecting \( N_{\text{sp}} \) in the range of 10 to 20 training epochs,
Table 3: Ablation studies for ROAD on the Office-31 datasets with 0.6 symmetric noise. ✓ stands for use.

<table>
<thead>
<tr>
<th>Method</th>
<th>Asymmetric(40%)</th>
<th>Symmetric(80%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PL BC BS CR HO KN MC PS PLT SB TR TK</td>
<td>Avg.</td>
</tr>
<tr>
<td>ResNet-101 [13]</td>
<td>57.2 56.1 22.4 39.5 36.2 44.7 66.1 18.7 41.8 7.4 76.6 9.3</td>
<td>39.7</td>
</tr>
<tr>
<td>DAN [24]</td>
<td>56.1 51.2 20.8 26.7 48.1 41.8 30.0 2.5 38.1 30.7 35.8 14.8</td>
<td>33.1</td>
</tr>
<tr>
<td>DANN [7]</td>
<td>89.7 51.2 36.2 61.5 35.0 13.2 71.8 25.3 79.1 7.1 67.5 7.5</td>
<td>45.3</td>
</tr>
<tr>
<td>AFN [40]</td>
<td>58.7 3.7 13.8 55.7 39.0 30.7 89.3 15.7 67.5 10.6 6.8 9.0</td>
<td>42.6</td>
</tr>
<tr>
<td>CDAN [26]</td>
<td>89.5 80.57 41.6 66.3 29.0 36.6 52.0 82.1 11.1 22.1 18.6</td>
<td>52.3</td>
</tr>
<tr>
<td>TCM [46]</td>
<td>84.2 55.4 55.7 54.7 47.1 23.7 75.7 47.3 84.6 8.5 65.8 18.4</td>
<td>52.7</td>
</tr>
<tr>
<td>DML [16]</td>
<td>38.7 68.2 19.8 11.9 20.0 46.0 15.3 24.1 18.2 3.4 93.4 15.6</td>
<td>31.2</td>
</tr>
<tr>
<td>CAF [39]</td>
<td>93.4 78.1 48.1 42.9 60.3 31.6 32.4 41.3 68.7 7.4 63.6 19.8</td>
<td>49.0</td>
</tr>
<tr>
<td>MCC [17]</td>
<td>89.7 78.0 49.0 54.8 81.4 82.1 67.6 58.7 83.6 30.0 83.1 32.6</td>
<td>65.6</td>
</tr>
<tr>
<td>CGDM [6]</td>
<td>91.4 45.2 18.9 58.1 87.0 71.1 41.3 64.2 52.2 85.7 18.8 22.4</td>
<td>55.2</td>
</tr>
<tr>
<td>SENTRY [32]</td>
<td>87.7 61.0 4.9 64.0 80.0 65.5 71.7 52.0 86.6 29.6 88.3 36.1</td>
<td>60.2</td>
</tr>
<tr>
<td>SHOT [21]</td>
<td>96.2 43.9 24.7 54.2 67.2 65.7 90.1 62.7 85.1 14.8 75.3 44.2</td>
<td>62.0</td>
</tr>
<tr>
<td>CoUDA [48]</td>
<td>43.6 14.9 8.4 52.1 46.2 45.3 37.9 3.4 47.7 14.8 54.9 6.9</td>
<td>33.0</td>
</tr>
</tbody>
</table>

Ours 91.4 80.2 66.7 54.6 93.7 23.7 76.2 52.0 83.5 40.7 82.9 44.1 65.8 87.2 42.4 74.6 38.6 91.7 10.8 87.1 63.9 81.4 45.2 78.2 38.1 161.6

Table 4: Comparison of candidates of 𝑁𝛼 with 0.6 symmetric noise. ✓ stands for use.

<table>
<thead>
<tr>
<th>Method</th>
<th>𝑋𝑤</th>
<th>𝑋𝑑</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>𝑉𝑒𝑤</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>𝑉𝑟𝑤</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>𝑉𝑟𝑑</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>𝑇</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>𝑉𝑒𝑑</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>𝑋𝑤</th>
<th>𝑋𝑑</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>CE only</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>w/o warm-up</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

4.5 Visualization of Robustness Analysis

To provide a comprehensive understanding of the robustness exhibited by our ROAD, we conduct visualization experiments for UDA with noisy labels on the Office-31 dataset. Firstly, we present a performance comparison between our ROAD and the MCC [17] in both the source and target domains throughout the learning process, as shown in Figure 4. Additionally, to shed light on the reasons behind the robustness and superior performance of our ROAD, we visualize the weights assigned to different phases, as illustrated in Figure 5. From the experimental results, the following observations can be drawn: 1) Throughout the whole training process, our ROAD mitigates the negative impact of noisy labels in the source domain without overfitting the dual noisy information. As a result, it maintains superior and robust performance, while MCC suffers from performance degradation in both domains. 2) As the learning proceeds, our ROAD gradually learns to distinguish between clean and noisy samples in the source domain by assigning small weights to contaminated samples with uncertain predictions and retaining reliable discriminative information, resulting in superior performance.
5 CONCLUSION

In this paper, we address a less-touched challenge of unsupervised domain adaptation (UDA) in open-world scenarios, where noisy labels in the source domain interfere with the learning process and propagate adverse effects to the target domain. To overcome the challenge, we propose a novel UDA framework, named ROAD, to learn discriminative and domain-adapted predictions robustly, which consists of two key mechanisms: the Robust Self-adaptive Weighted Learning mechanism (RSWL) and the Robust Domain-adapted Prediction Learning mechanism (RDPL). Specifically, RSWL is employed to adaptively assign weights to samples, enabling our model to focus on reliable samples for robust discrimination knowledge. In addition, RDPL enhances the transferability of discrimination knowledge from the source to the target domain by reducing the decision uncertainty of domain-adaptive predictions in the target domain. To demonstrate the robustness and effectiveness of our ROAD, we conduct comprehensive experiments by comparing it against 13 state-of-the-art UDA methods on three datasets, considering scenarios with and without synthetic noisy labels.

ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China (U21B2040, U19A2078, 62176171, 61971296, and 62102274), Sichuan Science and Technology Planning Project (2023YFG0008, 2023YFG0033, 2023ZHC0016, 2022YFQ0014, and 2022YFQ0021), Chengdu Science and Technology Project (2023-XT00-00004-GX), Fundamental Research Funds for the Central Universities (SCU2022JG002 and YJ202140), A*STAR AME Programmatic Funding A18A2b0046, RobotHTPO Seed Fund under Project C211518008 and EDB Space Technology Development Grant under Project S22-19016-STDP.
REFERENCES


[34] Binhui Xie, Shuang Li, Fangrui Lv, Chi Harold Liu, Guoren Wang, and Dapeng Wu. A collaborative alignment framework of transferable knowledge extraction for unsupervised domain adaptation. IEEE Transactions on Knowledge and Data Engineering, 2022.


