# ROAD: Robust Unsupervised Domain Adaptation with Noisy Labels

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

Hongyuan Zhu
hongyuanzhu.cn@gmail.com
Institute for Infocomm Research ($I^2R$),
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-cs/ROAD.

*Dezhong Peng is also with Chengdu Ruibei Yingte Information Technology Co., Ltd.
†Corresponding author.


*MM '23, October 29–November 3, 2023, Ottawa, Canada*

ACM ISBN 979-8-4007-0108-5/23/10…$15.00
https://doi.org/10.1145/3581783.3612296

## CCS CONCEPTS

• **Computing methodologies** → **Object recognition**; *Image representations.*

## KEYWORDS

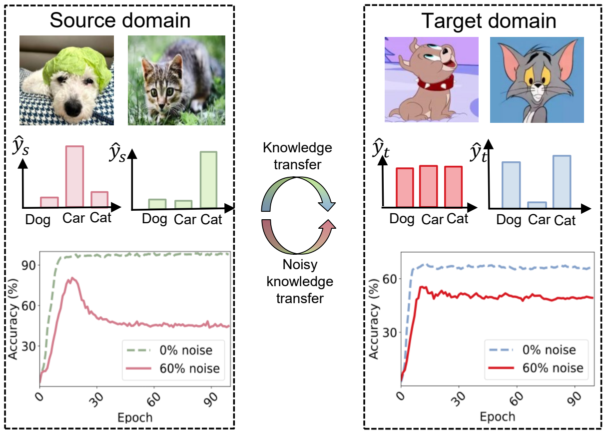
Unsupervised domain adaptation, Learning with noisy labels, Image classification, Self-adaptive weighted scheme

**ACM Reference Format:**
Yanglin Feng, Hongyuan Zhu, Dezhong Peng, Xi Peng, and Peng Hu. 2023. ROAD: Robust Unsupervised Domain Adaptation with Noisy Labels. In *Proceedings of the 31st ACM International Conference on Multimedia (MM '23), October 29-November 3, 2023, Ottawa, ON, Canada.* ACM, New York, NY, USA, 10 pages. 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.




(a) Source domain (b) Target domain

**Figure 1: The challenges encountered in unsupervised domain adaptation with noisy labels in the source domain. The upper part presents a comparison of prediction examples specific to each scenario, while the lower part depicts the classification performance during training. Notably, the presence of noisy labels in the source domain leads to the network model that gradually overfits noisy labels, resulting in noisy knowledge. This noisy knowledge transfer to the target domain amplifies decision uncertainty, leading to a degradation of domain adaptation performance 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, $\mathcal{I}_s = \{\mathcal{X}_s, \mathcal{Y}_s\} = \{\boldsymbol{x}_i^s, y_i^s\}_{i=1}^{N_s}$ represents the source domain data with noisy labels, where $N_s$ represents the number of samples in the source domain, $\boldsymbol{x}_i^s$ and $y_j^s \in \{1, 2, \cdots, C\}$ respectively are the $i^{th}$ sample and its class label which is potentially corrupted. Similarly, $\mathcal{I}_t = \{\boldsymbol{x}_j^t\}_{j=1}^{N_t}$ represents the target domain data, where $\boldsymbol{x}_j^t$ denotes the $i^{th}$ sample, and $N_t$ denotes the number of samples in the target domain. We also define $\{\hat{\mathcal{Y}}_s, \hat{\mathcal{Y}}_t\} = \{\hat{y}_i^s, \hat{y}_j^t\}_{i=1,j=1}^{N_s,N_t} = \{[\hat{y}_{i1}^s, \cdots, \hat{y}_{iC}^s], [\hat{y}_{j1}^t, \cdots, \hat{y}_{jC}^t]\}_{i=1,j=1}^{N_s,N_t}$ as the predictions of the domain-shared network $f = [f_1, \cdots, f_C]$, $f_i$ is the $i^{th}$ class output of the domain-shared network, $\hat{y}_{ij}^s = \frac{\exp(f_j(\boldsymbol{x}_i^s)/\tau)}{\sum_k^C \exp(f_k(\boldsymbol{x}_i^s)/\tau)}$ and $\hat{y}_{ij}^t = \frac{\exp(f_j(\boldsymbol{x}_i^t)/\tau)}{\sum_k^C \exp(f_k(\boldsymbol{x}_i^t)/\tau)}$ 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:

$$\mathcal{L}_{warm-up} = \frac{1}{N_s} \sum_{i}^{N_s} cetp_i^s, \tag{1}$$

where $cetp_i^s = -\sum_j^C y_{ij}^s \log \hat{y}_{ij}^s$ 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 noisy labels, enabling our ROAD to effectively address unsupervised domain adaptation under the influence of DNI. The overall training objective function is defined as follows:

$$\mathcal{L} = \mathcal{L}_{rwce} + \mu \mathcal{L}_{da}, \tag{2}$$

where $\mathcal{L}_{rwce}$ denotes the loss function employed by RSWL (see Equation (6)), $\mathcal{L}_{da}$ 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

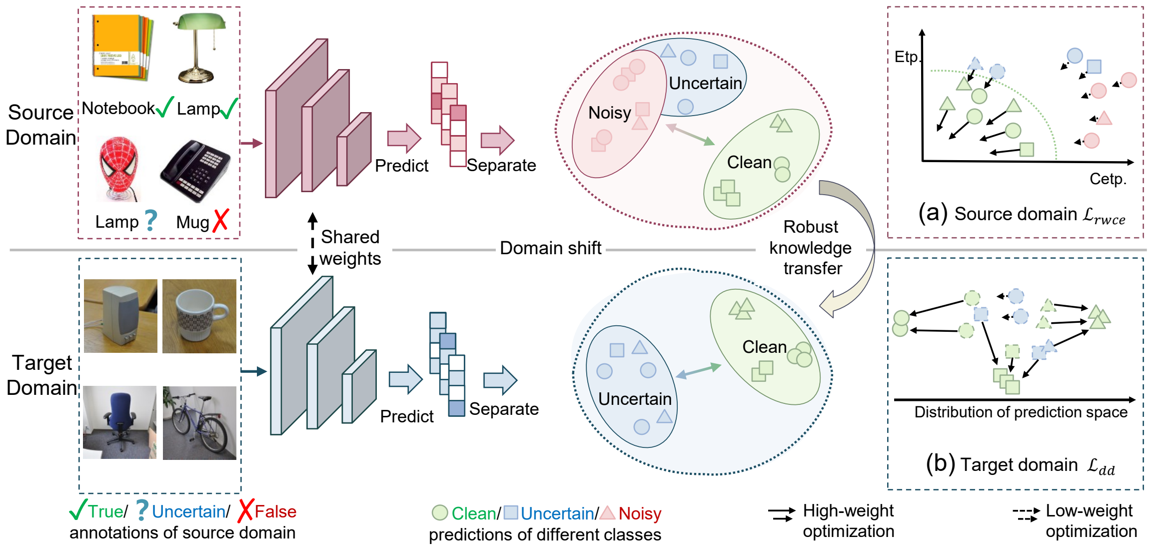



**Figure 2: The pipeline of our Robust Unsupervised Domain Adaptation Framework (ROAD). The data of different domains are fed into the domain-shared network to compute their soft class predictions. Then the Robust Self-adaptive Weighted Learning mechanism (RSWL) and Robust Domain-adapted Prediction Learning mechanism (RDPL) are conducted on the predictions of source and target domains, respectively. Specifically, RSWL employs a self-adaptive weighted scheme upon the mutual information (i.e. cross-entropy) between predictions and labels and self-information (i.e. entropy) of predictions to make the model focus on reliable samples. Furthermore, RDPL attempts to minimize the weighted decision uncertainty of predictions in the target domain, resulting in the robust transfer of discrimination knowledge against dual noisy information (DNI).**

Equation (2) by using stochastic gradient descent. The following sections elaborate on the two components of ROAD.

### 3.3 Robust Self-adaptive Weighted Learning

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, can be leveraged to estimate the reliability of samples in the source domain. We employ the mutual information between the predictions and the ground-truth labels, specifically the cross-entropy, to quantify the degree of approximation of each sample's prediction to the ground-truth labels. This leads to the construction of reliability for the source-domain samples, denoted as $V_i^{cetp}$, as follows:

$$V_i^{cetp} = 1 - \frac{e^{cetp_i^s} - e^{-cetp_i^s}}{e^{cetp_i^s} + e^{-cetp_i^s}}, \tag{3}$$

where $V_i^{cetp}$ represents the reliability of the $i^{th}$ sample, while $cetp_i^s$ denotes the cross-entropy calculated from the $i^{th}$ sample in the source domain.

Additionally, considering the memorization effect of DNNs [1], where DNNs provide more confident predictions for clean samples after a warm-up period, we leverage the self-information of the predictions, i.e., entropy, to estimate the confidence of predictions for each sample in the source domain. The level of confidence, denoted as $V_i^{etp}$, could be quantified as follows:

$$V_i^{etp} = \frac{B \cdot e^{\alpha(-etp_i^s+1)}}{\sum_j^B e^{(-etp_j^s+1)}}, \tag{4}$$

where $V_i^{etp}$ represents the prediction confidence of the $i^{th}$ sample in the source domain, $B$ denotes the size of the mini-batch, $etp_i^s = -\sum_j^C \hat{y}_{ij}^s \log \hat{y}_{ij}^s$ is the entropy function, and $\alpha$ is a trade-off parameter.

Consequently, we combine the reliability $V^{cetp}$ and the confidence $V^{etp}$ 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:

$$\mathcal{L}_{rwce} = -\frac{\sum_i^{N_s} \left( V_i^{etp} V_i^{cetp} \sum_j^C (y_{ij}^s \log \hat{y}_{ij}^s) \right)}{\sum_j^{N_s} V_j^{etp} V_j^{cetp}}. \tag{5}$$

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:

$$T_i = \frac{1}{C}\sum_{j}^{C}\frac{\sum_{k\neq j}^{C}\hat{y}_{ij}^{t}\hat{y}_{ik}^{t}}{\sum_{l}^{C}\hat{y}_{ij}^{t}\hat{y}_{il}^{t}}. \tag{6}$$

The objective is to minimize the average of the decision uncertainty term $T$ for all target samples, which facilitates obtaining domain-adapted discriminative predictions in the target domain. Thus, we define the vanilla domain-adaptation loss $\mathcal{L}_{da}^{'}$ as follows:

$$\mathcal{L}_{da}^{'} = \frac{1}{N_t}\sum_{i}^{N_t} T_i. \tag{7}$$

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:

$$W_i^{etp} = \frac{B\cdot e^{\alpha(-etp_i^t+1)}}{\sum_j^B e^{(-etp_j^t+1)}}, \tag{8}$$

where $etp_i^t = -\sum_j^C \hat{y}_{ij}^t \log \hat{y}_{ij}^t$ is the entropy function.

Based on the above, we exploit the prediction confidence $W^{etp}$ to weigh the uncertainty of each sample, constructing the loss function $\mathcal{L}_{da}$ for our RDPL, as shown below:

$$\mathcal{L}_{da} = \frac{\sum_i^{N_t}(W_i^{etp}T_i)}{\sum_j^{N_t} W_j^{etp}}. \tag{9}$$

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**, **Ar→Pr**, **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 <u>underlined</u>.**

| Method | Office-31 [34] | | | | | Office-Home [37] | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0% | 20% | 40% | 60% | 80% | 0% | 20% | 40% | 60% | 80% |
| ResNet-50 [13] | 76.1 | 63.7 | 53.7 | 37.3 | 19.2 | 46.1 | 39.7 | 30.0 | 22.4 | 12.3 |
| DAN [24] | 80.4 | 71.3 | 61.5 | 48.5 | 26.5 | 56.3 | 41.9 | 30.9 | 25.1 | 14.1 |
| DANN [7] | 82.2 | 69.3 | 56.3 | 43.3 | 22.2 | 57.6 | 42.0 | 29.8 | 22.6 | 10.5 |
| AFN [40] | 85.7 | 80.8 | 71.3 | 62.8 | 41.2 | 67.3 | 63.2 | 59.2 | 46.9 | 28.4 |
| CDAN [26] | 87.7 | 76.6 | 66.5 | 52.2 | 28.2 | 63.8 | 50.4 | 38.9 | 28.4 | 14.2 |
| TCM [46] | **89.7** | 81.2 | 74.2 | 61.5 | 37.8 | 71.1 | 49.8 | 39.6 | 31.2 | 16.9 |
| DMAL [16] | 86.6 | 80.0 | 70.9 | 59.6 | 35.7 | 66.1 | 47.2 | 43.7 | 38.8 | 14.7 |
| CAF [39] | 88.6 | 76.7 | 64.8 | 52.1 | 32.7 | 69.0 | 42.4 | 34.5 | 28.4 | 17.7 |
| MCC [17] | 89.4 | 74.5 | 69.9 | 62.6 | <u>44.4</u> | <u>74.2</u> | 55.7 | 49.9 | 43.1 | 30.8 |
| CGDM [6] | 88.4 | <u>82.1</u> | <u>77.9</u> | <u>67.4</u> | 39.0 | 68.5 | 62.1 | 58.0 | 46.6 | 20.3 |
| SENTRY [32] | 87.3 | 81.9 | 74.9 | 66.0 | 35.3 | 72.3 | 63.3 | 59.8 | 49.8 | 31.6 |
| SHOT [21] | 88.6 | 82.0 | 75.6 | 63.5 | 29.9 | 71.8 | <u>65.8</u> | <u>61.3</u> | <u>51.8</u> | <u>33.8</u> |
| CoUDA [48] | 83.0 | 59.8 | 50.4 | 35.5 | 16.0 | 42.1 | 35.6 | 31.5 | 20.7 | 12.9 |
| Ours | <u>89.6</u> | **85.6** | **82.0** | **78.1** | **60.9** | **76.0** | **67.5** | **64.2** | **56.7** | **47.2** |

**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 <u>underlined</u>.**

| Method | Asymmetric | | | | Symmetric | | | |
|---|---|---|---|---|---|---|---|---|
| | 0% | 10% | 20% | 40% | 20% | 40% | 60% | 80% |
| ResNet-101 [13] | 52.4 | 51.6 | 46.1 | 39.7 | 36.6 | 31.9 | 17.2 | 14.2 |
| DAN [24] | 61.1 | 43.3 | 39.6 | 33.1 | 37.7 | 33.9 | 27.6 | 19.1 |
| DANN [7] | 59.2 | 47.1 | 43.9 | 45.3 | 49.6 | 43.3 | 33.7 | 18.3 |
| AFN [40] | 76.1 | 53.9 | 46.0 | 42.6 | 44.8 | 36.8 | 31.5 | 22.0 |
| CDAN [26] | 66.8 | 58.0 | 52.2 | 52.3 | 64.0 | 51.4 | 43.0 | 19.2 |
| TCM [46] | 68.4 | 65.7 | 62.4 | 52.7 | 66.7 | <u>64.6</u> | <u>55.8</u> | 30.2 |
| DMAL [16] | 60.0 | 41.1 | 36.0 | 31.2 | 35.8 | 27.3 | 28.1 | 13.9 |
| CAF [39] | <u>80.3</u> | 62.0 | 62.0 | 49.0 | 54.1 | 50.3 | 50.8 | <u>32.8</u> |
| MCC [17] | 78.8 | **70.8** | 68.7 | <u>65.6</u> | 67.9 | 58.7 | 29.7 | 22.9 |
| CGDM [6] | **82.3** | 68.2 | 61.6 | 55.2 | 66.0 | 60.4 | 42.2 | 30.8 |
| SENTRY [6] | 69.2 | 66.9 | 64.0 | 60.2 | 52.7 | 53.0 | 40.1 | 21.3 |
| SHOT [21] | 74.6 | <u>69.5</u> | <u>68.9</u> | 62.0 | <u>70.5</u> | 59.0 | 45.1 | 22.7 |
| CoUDA [48] | 45.7 | 38.8 | 33.9 | 33.0 | 38.4 | 36.1 | 30.5 | 21.4 |
| Ours | **82.3** | 69.2 | **69.0** | **65.8** | **72.4** | **70.3** | **69.8** | **61.6** |

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., $\mathcal{L}_{rwce}$ and $\mathcal{L}_{da}$), as well as each component within the losses (i.e., $V^{etp}$, $V^{cetp}$, $W^{etp}$ 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 $\mathcal{L}_{rwce}$, 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 $\mathcal{L}_{da}$ with the vanilla loss of RDPL $\mathcal{L}_{da}^{'}$ (i.e., $\mathcal{L}_{da}$ w/o $W^{ept}$) results in performance degradation, which demonstrates the significance of $W^{ept}$ in enabling RDPL to mitigate the learning of unreliable discrimination knowledge propagated from the source domain to the target domain. 3) Substituting $\mathcal{L}_{rwce}$ with CE leads to a remarkable drop in domain-adaptation performance, underscoring the robustness of our $\mathcal{L}_{rwce}$, which assigns self-adaptive weights to clean and noisy data, particularly for challenging tasks (e.g., **D**→**A** and **W**→**A**).

Overall, the ablation study demonstrates the effectiveness of each component in our framework and emphasizes the crucial role of $\mathcal{L}_{da}$, $\mathcal{L}_{rwce}$ and $W^{ept}$ in addressing UDA with noisy labels.

## 4.4 Parameter Analysis

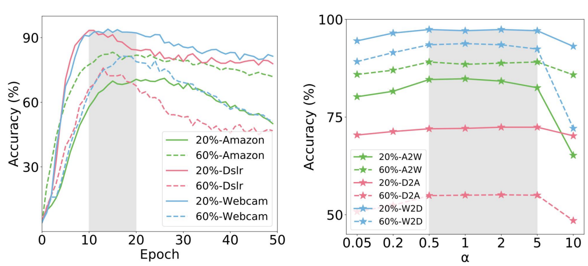
In order to evaluate the sensitivity of our ROAD to different hyper-parameter settings, namely $N_w$ and $\alpha$, we plot the accuracy scores for the parameter analysis on Office-31, as shown in Figure 3. Without loss of generality, we consider both low (20%) and high (60%) noise levels in all experiments. From Figure 3a, it could be observed that selecting $N_w$ in the range of 10 to 20 training epochs,

**Table 3: Category-wide performance comparison under extremely high noise rates (40% asymmetric label noise and 80% symmetric label noise) in T → V transfer task of the VisDA-2017 datasets. We abbreviate Plane, Bicycle, Bus, Car, Horse, Knife, Motorcycle, Person, Plant, Skateboard, Train, and Truck to PL, BC, BS, CR, HO, KN, MC, PS, PLT, SB, TR, and TK, respectively. The highest accuracy is shown in bold and the second highest accuracy is underlined.**

| Method | Asymmetric(40%) | | | | | | | | | | | | | Symmetric(80%) | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | PL | BC | BS | CR | HO | KN | MC | PS | PLT | SB | TR | TK | Avg. | PL | BC | BS | CR | HO | KN | MC | PS | PLT | SB | TR | TK | Avg. |
| ResNet-101 [13] | 57.2 | 56.1 | 22.4 | 39.5 | 36.2 | 44.7 | 66.1 | 18.7 | 41.8 | 7.4 | 76.6 | 9.3 | 39.7 | 7.5 | 0.8 | 19.4 | 15.0 | 20.9 | **56.3** | 8.3 | 13.8 | 5.3 | 0.0 | 19.4 | 4.0 | 14.2 |
| DAN [24] | 56.1 | 51.2 | 20.8 | 26.7 | 48.1 | 41.8 | 30.0 | 2.5 | 38.1 | 30.7 | 35.8 | 14.8 | 33.1 | 28.1 | 19.5 | 53.2 | 2.3 | <u>39.5</u> | 3.9 | 21.2 | 5.3 | 26.7 | 17.3 | 11.9 | 0.0 | 19.1 |
| DANN [7] | 89.7 | 51.2 | 34.6 | 62.1 | 35.0 | 13.2 | 71.8 | 25.3 | 79.1 | 7.1 | 67.5 | 7.5 | 45.3 | 1.8 | 17.1 | 42.0 | 18.1 | 20.0 | <u>28.9</u> | 31.0 | 2.7 | 10.4 | 7.4 | 32.6 | 8.1 | 18.3 |
| AFN [40] | 58.7 | 3.7 | 13.8 | 55.7 | 79.4 | 30.7 | <u>89.3</u> | 15.7 | 67.5 | 10.6 | 76.8 | 9.0 | 42.6 | 21.3 | 25.6 | 31.4 | 12.7 | 34.5 | 3.7 | 35.5 | 10.3 | 4.1 | <u>28.8</u> | 50.8 | 5.0 | 22.0 |
| CDAN [26] | 89.5 | **80.5** | **74.1** | **66.1** | 66.3 | 29.0 | 36.6 | 52.0 | 82.1 | 11.1 | 22.1 | 18.6 | 52.3 | 8.7 | 17.1 | <u>61.7</u> | 15.3 | 31.2 | 13.2 | 26.8 | 4.0 | 14.9 | 7.4 | 15.8 | 14.0 | 19.2 |
| TCM [46] | 84.2 | 55.4 | 55.7 | 55.4 | 47.1 | 23.7 | 75.7 | 47.3 | 84.6 | 18.5 | 65.8 | 18.4 | 52.7 | 43.1 | <u>43.9</u> | 46.9 | 21.1 | 28.5 | 18.4 | 56.5 | 16.4 | 12.2 | 14.8 | 46.1 | 14.4 | 30.2 |
| DMAL [16] | 38.7 | 68.2 | 19.8 | 11.9 | 20.0 | 46.0 | 15.3 | 24.1 | 18.2 | 3.4 | **93.4** | 15.6 | 31.2 | 2.2 | 5.5 | 28.6 | 8.9 | 8.4 | 10.3 | 9.4 | 10.3 | 12.4 | 3.7 | 63.2 | 4.3 | 13.9 |
| CAF [39] | <u>93.4</u> | 78.1 | 48.1 | 42.9 | 60.3 | 31.6 | 32.4 | 41.3 | 68.7 | 7.4 | 63.6 | 19.8 | 49.0 | 42.1 | **61.0** | 54.3 | 31.1 | 47.5 | 21.1 | 25.4 | <u>30.7</u> | <u>30.0</u> | 11.1 | 25.1 | 13.9 | 32.8 |
| MCC [17] | 89.7 | 78.0 | 49.4 | 50.8 | 81.4 | **82.1** | 67.6 | 58.7 | 83.6 | 30.0 | 83.1 | 32.6 | <u>65.6</u> | 34.7 | 2.2 | 16.6 | 21.9 | 28.5 | 6.7 | 31.5 | 15.5 | 18.9 | 10.5 | 63.1 | <u>24.7</u> | 22.9 |
| CGDM [6] | 91.4 | 45.2 | 18.9 | 58.1 | 87.0 | <u>77.1</u> | 41.3 | **64.2** | 52.2 | **85.7** | 18.8 | 22.4 | 55.2 | <u>57.1</u> | 43.6 | 38.0 | 13.2 | 31.4 | 10.8 | 38.8 | 5.4 | 24.2 | 26.1 | **80.1** | 1.2 | <u>30.8</u> |
| SENTRY [32] | 87.7 | 61.0 | 4.9 | <u>64.0</u> | 80.0 | 60.5 | 71.7 | 52.0 | **86.6** | 29.6 | <u>88.3</u> | 36.1 | 60.2 | 8.8 | 14.6 | 22.2 | 23.2 | 5.0 | 7.9 | 66.2 | 28.0 | 10.4 | 18.5 | 39.0 | 11.6 | 21.3 |
| SHOT [21] | **96.2** | 43.9 | 24.7 | 54.2 | <u>87.2</u> | 65.7 | **90.1** | <u>62.7</u> | <u>85.1</u> | 14.8 | 75.3 | **44.2** | 62.0 | 5.3 | 9.8 | 58.0 | **57.6** | 3.8 | 13.2 | **90.1** | 10.7 | 7.5 | 3.7 | 1.4 | 11.6 | 22.7 |
| CoUDA [48] | 43.6 | 14.9 | 8.4 | 52.1 | 46.2 | 45.3 | 57.9 | 3.4 | 47.7 | 14.8 | 54.9 | 6.9 | 33.0 | 19.3 | 9.8 | 38.3 | 7.1 | 21.5 | **36.8** | 47.9 | 5.4 | 4.5 | 22.2 | 29.9 | 13.9 | 21.4 |
| Ours | 91.4 | <u>80.2</u> | <u>66.7</u> | 54.6 | **93.7** | 23.7 | 76.2 | 52.0 | 83.5 | <u>40.7</u> | 82.9 | <u>44.1</u> | **65.8** | **87.2** | 42.4 | **74.6** | <u>38.6</u> | **91.7** | 10.8 | <u>87.1</u> | **63.9** | **81.4** | **45.2** | <u>78.2</u> | **38.1** | **61.6** |

**Table 4: Ablation studies for ROAD on the Office-31 datasets with 0.6 symmetric noise. ✓ stands for use.**

| $\mathcal{L}_{rwce}$ | | $\mathcal{L}_{da}$ | | A→D | A→W | D→A | D→W | W→A | W→D | Avg. |
|---|---|---|---|---|---|---|---|---|---|---|
| $v^{etp}$ | $v^{cetp}$ | $w^{etp}$ | $T$ | | | | | | | |
| ✓ | ✓ | ✓ | ✓ | **84.6** | **84.8** | **55.1** | **85.2** | **64.9** | **93.8** | **78.1** |
| ✓ | ✓ | ✓ | | 75.2 | 75.8 | 23.5 | 74.7 | 56.9 | 87.8 | 65.7 |
| ✓ | ✓ | | ✓ | 78.3 | 79.1 | 48.5 | 84.1 | 63.2 | 91.7 | 76.3 |
| ✓ | | ✓ | ✓ | 78.9 | 78.4 | 39.4 | 70.8 | 57.4 | 83.3 | 68.0 |
| | ✓ | ✓ | ✓ | 80.9 | 81.3 | 51.7 | 83.2 | 62.7 | 92.8 | 75.4 |
| ✓ | | | | 73.8 | 74.9 | 21.1 | 73.9 | 55.8 | 84.9 | 64.1 |
| CE only | | ✓ | | 75.4 | 76.9 | 41.7 | 63.1 | 50.0 | 66.8 | 62.3 |
| w/o warm-up | | | | 4.2 | 4.5 | 3.5 | 4.7 | 4.3 | 10.6 | 5.3 |



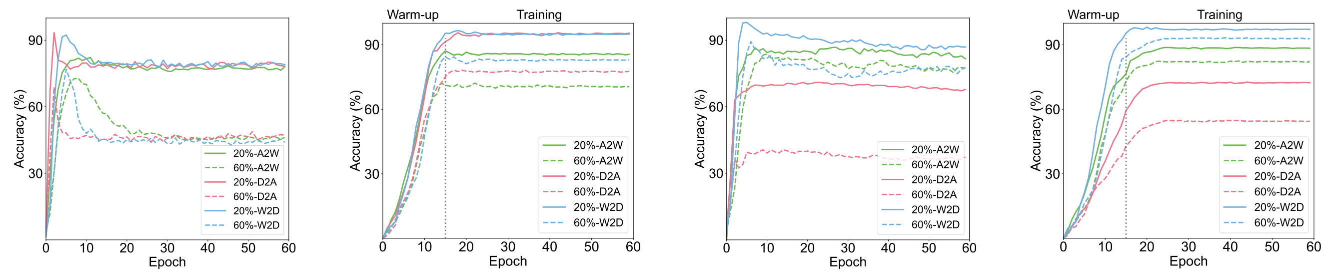

**(a) Comparison of candidates of $N_w$ (b) Comparison of candidates of $\alpha$**

**Figure 3: The image classification performance in terms of accuracy versus various candidates of $N_w$ and $\alpha$ on the Office-31. (a) shows the relationship between training epochs and classification performance in three domains (i.e., Amazon, Dslr, and Webcam), and (b) shows the relationship between $\alpha$ and performance in three transfer tasks (i.e., A→W, D→A, and W→D). The gray box shows the range of optimal choices.**

as suggested by the memorization effect of DNNs [1], leads to the most robust performance, which remains consistent across different datasets and noise rates. Furthermore, Figure 3b demonstrates that our ROAD achieves optimal performance when $\alpha$ is within the range of 0.5 to 5. This indicates that our framework strikes a balance between mitigating the adverse impact of samples with uncertain predictions and retaining reliable discriminative information, resulting in superior performance.

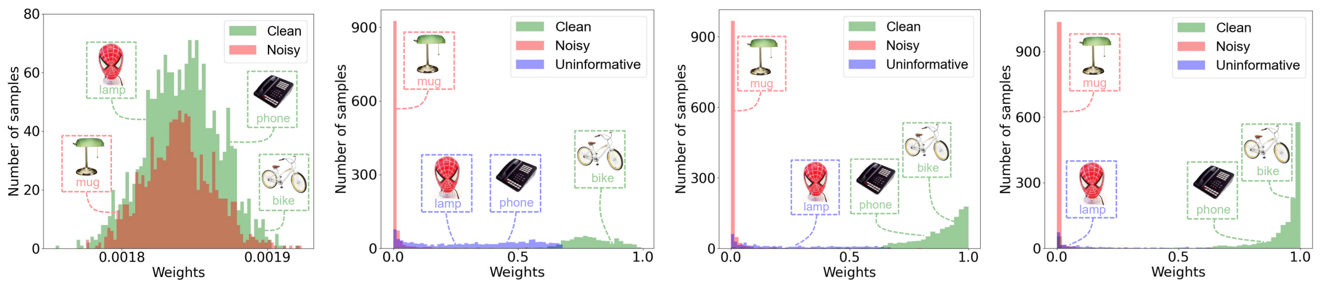
## 4.5 Visualization of Robustness Analysis

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 noisy and uncertain samples, thus effectively filtering out the noise present in the data. In the target domain, our ROAD can gradually reduce the decision uncertainty of predictions by assigning small weights to contaminated samples with high prediction entropy, thus enabling our model to achieve robust and domain-adapted predictions.

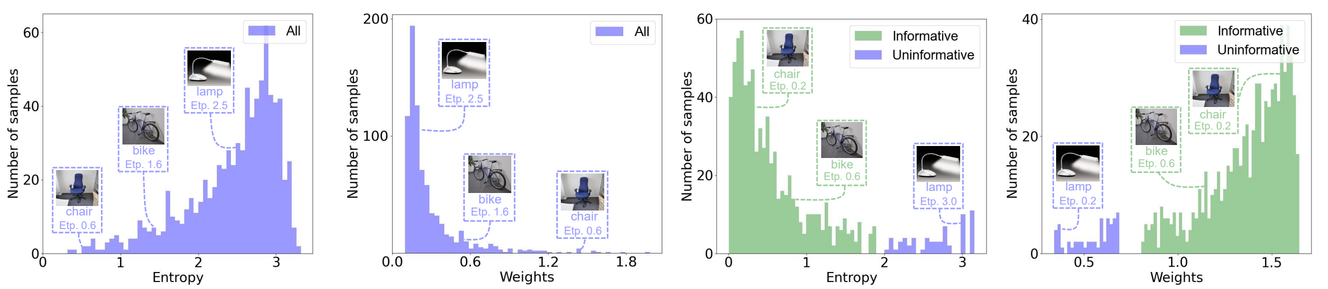



(a) Source performance of MCC (b) Source performance of ROAD (c) Target performance of MCC (d) Target performance of ROAD

**Figure 4: The unsupervised domain adaptation performance in terms of accuracy of MCC and our ROAD under 20% and 60% source noise on three transfer tasks (i.e., A→W, D→A, and W→D) of Office-31.**




(a) Source: weights before warm-up (b) Source: weights after warm-up (c) Source: weights during training (d) Source: weights after training




(e) Target: entropy after warm-up (f) Target: weights after warm-up (g) Target: entropy after training (h) Target: weights after training

**Figure 5: The weights or entropy of samples in different phases of our ROAD under 40% noise on A→W tasks of Office-31. The upper half shows the weighs of samples before warm-up, after warm-up, during training and after training. The bottom half shows the prediction entropy and weighs of samples after warm-up and after whole training.**

# 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 (2023YFQ0020, 2023YFG0033, 2023ZHCG0016, 2022YFQ0014, and 2022YFH0021), 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


[1] Devansh Arpit, Stanisław Jastrzębski, Nicolas Ballas, David Krueger, Emmanuel Bengio, Maxinder S Kanwal, Tegan Maharaj, Asja Fischer, Aaron Courville, Yoshua Bengio, et al. A closer look at memorization in deep networks. In *International conference on machine learning*, pages 233–242. PMLR, 2017.

[2] Alan Joseph Bekker and Jacob Goldberger. Training deep neural-networks based on unreliable labels. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2682–2686. IEEE, 2016.

[3] Xinlei Chen and Abhinav Gupta. Webly supervised learning of convolutional networks. In *Proceedings of the IEEE international conference on computer vision*, pages 1431–1439, 2015.

[4] Cheng Cheng, Xiaoyu Liu, Beitong Zhou, and Ye Yuan. Intelligent fault diagnosis with noisy labels via semi-supervised learning on industrial time series. *IEEE Transactions on Industrial Informatics*, 2023.

[5] Debasmit Das and CS George Lee. Graph matching and pseudo-label guided deep unsupervised domain adaptation. In *Artificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018, Proceedings, Part III 27*, pages 342–352. Springer, 2018.

[6] Zhekai Du, Jingjing Li, Hongzu Su, Lei Zhu, and Ke Lu. Cross-domain gradient discrepancy minimization for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3937–3946, 2021.

[7] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *The journal of machine learning research*, 17(1):2096–2030, 2016.

[8] Behnam Gholami, Vladimir Pavlovic, et al. Punda: Probabilistic unsupervised domain adaptation for knowledge transfer across visual categories. In *Proceedings of the IEEE international conference on computer vision*, pages 3581–3590, 2017.

[9] Aritra Ghosh, Himanshu Kumar, and P Shanti Sastry. Robust loss functions under label noise for deep neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31, 2017.

[10] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

[11] Arthur Gretton, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *The Journal of Machine Learning Research*, 13(1):723–773, 2012.

[12] Bo Han, Quanming Yao, Xingrui Yu, Gang Niu, Miao Xu, Weihua Hu, Ivor Tsang, and Masashi Sugiyama. Co-teaching: Robust training of deep neural networks with extremely noisy labels. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper/2018/file/a19744e268754fb0148b017647355b7b-Paper.pdf.

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[14] Peng Hu, Xi Peng, Hongyuan Zhu, Liangli Zhen, and Jie Lin. Learning cross-modal retrieval with noisy labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5403–5413, June 2021.

[15] Jiaxing Huang, Dayan Guan, Aoran Xiao, Shijian Lu, and Ling Shao. Category contrast for unsupervised domain adaptation in visual tasks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1203–1214, 2022.

[16] Jingke Huang, Ni Xiao, and Lei Zhang. Balancing transferability and discriminability for unsupervised domain adaptation. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[17] Ying Jin, Ximei Wang, Mingsheng Long, and Jianmin Wang. Minimum class confusion for versatile domain adaptation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXI 16*, pages 464–480. Springer, 2020.

[18] Junnan Li, Richard Socher, and Steven CH Hoi. Dividemix: Learning with noisy labels as semi-supervised learning. *arXiv preprint arXiv:2002.07394*, 2020.

[19] Yehao Li, Yingwei Pan, Ting Yao, and Tao Mei. Comprehending and ordering semantics for image captioning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17990–17999, 2022.

[20] Yuncheng Li, Jianchao Yang, Yale Song, Liangliang Cao, Jiebo Luo, and Li-Jia Li. Learning from noisy labels with distillation. In *Proceedings of the IEEE international conference on computer vision*, pages 1910–1918, 2017.

[21] Jian Liang, Dapeng Hu, and Jiashi Feng. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In *International Conference on Machine Learning*, pages 6028–6039. PMLR, 2020.

[22] Pengfei Liang, Wenhui Wang, Xiaoming Yuan, Siyuan Liu, Lijie Zhang, and Yiwei Cheng. Intelligent fault diagnosis of rolling bearing based on wavelet transform and improved resnet under noisy labels and environment. *Engineering Applications of Artificial Intelligence*, 115:105269, 2022.

[23] Tongliang Liu and Dacheng Tao. Classification with noisy labels by importance reweighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38 (3):447–461, 2016. doi: 10.1109/TPAMI.2015.2456899.

[24] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael Jordan. Learning transferable features with deep adaptation networks. In *International conference on machine learning*, pages 97–105. PMLR, 2015.

[25] Mingsheng Long, Han Zhu, Jianmin Wang, and Michael I Jordan. Deep transfer learning with joint adaptation networks. In *International conference on machine learning*, pages 2208–2217. PMLR, 2017.

[26] Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. Conditional adversarial domain adaptation. *Advances in neural information processing systems*, 31, 2018.

[27] Yawei Luo, Liang Zheng, Tao Guan, Junqing Yu, and Yi Yang. Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2507–2516, 2019.

[28] Xingjun Ma, Hanxun Huang, Yisen Wang, Simone Romano, Sarah Erfani, and James Bailey. Normalized loss functions for deep learning with noisy labels. In *International conference on machine learning*, pages 6543–6553. PMLR, 2020.

[29] Gregory P. Meyer and Niranjan Thakurdesai. Learning an uncertainty-aware object detector for autonomous driving. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 10521–10527, 2020. doi: 10.1109/IROS45743.2020.9341623.

[30] Nagarajan Natarajan, Inderjit S Dhillon, Pradeep K Ravikumar, and Ambuj Tewari. Learning with noisy labels. In C.J. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013. URL https://proceedings.neurips.cc/paper_files/paper/2013/file/3871bd64012152bfb53fdf04b401193f-Paper.pdf.

[31] Xingchao Peng, Ben Usman, Neela Kaushik, Judy Hoffman, Dequan Wang, and Kate Saenko. Visda: The visual domain adaptation challenge. *arXiv preprint arXiv:1710.06924*, 2017.

[32] Viraj Prabhu, Shivam Khare, Deeksha Kartik, and Judy Hoffman. Sentry: Selective entropy optimization via committee consistency for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8558–8567, 2021.

[33] Yang Qin, Dezhong Peng, Xi Peng, Xu Wang, and Peng Hu. Deep evidential learning with noisy correspondence for cross-modal retrieval. In *Proceedings of the 30th ACM International Conference on Multimedia*, MM '22, page 4948–4956, 2022. doi: 10.1145/3503161.3547922.

[34] Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell. Adapting visual category models to new domains. In *Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part IV 11*, pages 213–226. Springer, 2010.

[35] Yuan Sun, Zhenwen Ren, Peng Hu, Dezhong Peng, and Xu Wang. Hierarchical consensus hashing for cross-modal retrieval. *IEEE Transactions on Multimedia*, pages 1–14, 2023. doi: 10.1109/TMM.2023.3272169.

[36] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7167–7176, 2017.

[37] Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, and Sethuraman Panchanathan. Deep hashing network for unsupervised domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5018–5027, 2017.

[38] Yisen Wang, Xingjun Ma, Zaiyi Chen, Yuan Luo, Jinfeng Yi, and James Bailey. Symmetric cross entropy for robust learning with noisy labels. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

[39] 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.

[40] Ruijia Xu, Guanbin Li, Jihan Yang, and Liang Lin. Larger norm more transferable: An adaptive feature norm approach for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1426–1435, 2019.

[41] Tianyuan Xu, Xueliang Liu, Zhen Huang, Dan Guo, Richang Hong, and Meng Wang. Early-learning regularized contrastive learning for cross-modal retrieval with noisy labels. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 629–637, 2022.

[42] Hongliang Yan, Yukang Ding, Peihua Li, Qilong Wang, Yong Xu, and Wangmeng Zuo. Mind the class weight bias: Weighted maximum mean discrepancy for unsupervised domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2272–2281, 2017.

[43] Li Yi, Sheng Liu, Qi She, A Ian McLeod, and Boyu Wang. On learning contrastive representations for learning with noisy labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16682–16691, 2022.

[44] Donggeun Yoo, Namil Kim, Sunggyun Park, Anthony S Paek, and In So Kweon. Pixel-level domain transfer. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VIII 14*, pages 517–532. Springer, 2016.

[45] Quanzeng You, Hailin Jin, Zhaowen Wang, Chen Fang, and Jiebo Luo. Image captioning with semantic attention. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4651–4659, 2016.
[46] Zhongqi Yue, Qianru Sun, Xian-Sheng Hua, and Hanwang Zhang. Transporting causal mechanisms for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8599–8608, 2021.
[47] Jingyi Zhang, Jiaxing Huang, Zichen Tian, and Shijian Lu. Spectral unsupervised domain adaptation for visual recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9829–9840, 2022.
[48] Yifan Zhang, Ying Wei, Qingyao Wu, Peilin Zhao, Shuaicheng Niu, Junzhou Huang, and Mingkui Tan. Collaborative unsupervised domain adaptation for medical image diagnosis. *IEEE Transactions on Image Processing*, 29:7834–7844, 2020.
[49] Zhilu Zhang and Mert Sabuncu. Generalized cross entropy loss for training deep neural networks with noisy labels. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/file/f2925f97bc13ad2852a7a551802feea0-Paper.pdf.
[50] Zhilu Zhang and Mert Sabuncu. Generalized cross entropy loss for training deep neural networks with noisy labels. *Advances in neural information processing systems*, 31, 2018.
[51] Yongchun Zhu, Fuzhen Zhuang, Jindong Wang, Jingwu Chen, Zhiping Shi, Wenjuan Wu, and Qing He. Multi-representation adaptation network for cross-domain image classification. *Neural Networks*, 119:214–221, 2019.
[52] Yang Zou, Zhiding Yu, Xiaofeng Liu, B.V.K. Vijaya Kumar, and Jinsong Wang. Confidence regularized self-training. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.