Universal Domain Adaptation from Foundation Models: A Baseline Study

Bin Deng and Kui Jia

South China University of Technology
eebindeng@mail.scut.edu.cn

November 6, 2023

Abstract

Foundation models (e.g., CLIP or DINOv2) have shown their impressive learning and transfer capabilities in a wide range of visual tasks, by training on a large corpus of data and adapting to specific downstream tasks. It is, however, interesting that foundation models have not been fully explored for universal domain adaptation (UniDA), which is to learn models using labeled data in a source domain and unlabeled data in a target one, such that the learned models can successfully adapt to the target data. In this paper, we make comprehensive empirical studies of state-of-the-art UniDA methods using foundation models. We first observe that, unlike fine-tuning from ImageNet pre-trained models, as previous methods do, fine-tuning from foundation models yields significantly poorer results, sometimes even worse than training from scratch. While freezing the backbones, we demonstrate that although the foundation models greatly improve the performance of the baseline method that trains the models on the source data alone, existing UniDA methods generally fail to improve over the baseline. This suggests that new research efforts are very necessary for UniDA using foundation models. Based on these findings, we introduce *CLIP distillation*, a parameter-free method specifically designed to distill target knowledge from CLIP models. The core of our *CLIP distillation* lies in a self-calibration technique for automatic temperature scaling, a feature that significantly enhances the baseline's out-class detection capability. Although simple, our method outperforms previous approaches in most benchmark tasks, excelling in evaluation metrics including H-score/H^3-score and the newly proposed universal classification rate (UCR) metric. We hope that our investigation and the proposed simple framework can serve as a strong baseline to facilitate future studies in this field. The code is available at `https://github.com/szubing/uniood`.

1 Introduction

A foundational goal of machine visual is to develop a model that can be applied to data from different distributions. With the emergence of many large-scale pre-trained models such as CLIP [28], ALIGN [18], and DINOv2 [24], significant progress has been made recently towards achieving this goal. These "foundation models" [1] often exhibit significantly greater robustness to various benchmark distribution shifts compared to standard training models. Taking image classification as an example, both CLIP and a standard ResNet50 achieve an accuracy of 76% on ImageNet, but CLIP has shown significant improvements with an accuracy increase of 6% on ImageNetV2 and an accuracy increase of 35% on ImageNet Sketch [28]. Due to the powerful ability of these foundation models, techniques for applying them on downstream applications are increasingly important. Indeed, the research community has spent significant effort during the past few years to improving the fine-tuning of these models for various downstream tasks, including few-shot classification [21], out-of-distribution (OOD) detection [6], and OOD generalization [37, 19], among others.

Surprisingly, universal domain adaptation (UniDA) [38], one of the practical applications that aims to adapt to one specific target domain without any restriction on label sets, has not been thoroughly explored to date under the powerful foundation models. This paper aims to fill this gap by initially assessing the performance of state-of-the-art UniDA methods when applied to foundation models. Through comprehensive experiments, we conclude several interesting findings. First of all, unlike fine-tuning from pre-trained models on ImageNet, fine-tuning from foundation models yields significantly poorer results, sometimes even worse than training from scratch. We then freeze the foundation models and focus solely on updating the classifier head. In this scenario, all methods achieved substantial improvements over their prior results that are reliant on ImageNet pre-trained models. However, the performance gap between the Source Only (SO) baseline and state-of-the-art (SOTA) methods has notably narrowed, rendering them largely comparable across various benchmark tasks. These findings suggest that new research efforts are very necessary for UniDA using foundation models.

Based on our empirical observations, we present *CLIP distillation*, a parameter-free technique designed to distill knowledge from CLIP models. The core of our *CLIP distillation* method revolves around a self-calibration approach that automatically adjusts temperature scaling. This feature significantly enhances the baseline model's ability to detect out-class samples. Additionally, we introduce a new evaluation metric for UniDA that is threshold- and ratio-free, making it suitable for methods that do not consider threshold effects. Despite its simplicity, our method demonstrates exceptional robustness and efficacy in various task scenarios, spanning open-partial, open, closed, and partial UniDA settings. It excels in both the established H-score/H^3-score metrics and the novel UCR metric. This straightforward approach sets a new standard for UniDA using foundation models, providing a solid baseline for future research in this field.

The main contributions of this paper are summarized as follows:

- To the best of our knowledge, we are the first to tackle the UniDA challenge and conduct comprehensive studies into existing methods when applied to foundation models. Our findings underscore the urgent need for further research for UniDA using these powerful foundation models.

- We propose *CLIP distillation* for UniDA, which sets a new baseline for adaptation from foundation models. Our method includes a self-calibration technique for automatic temperature scaling, rendering *CLIP distillation* parameter-free and robust across diverse task settings.

- We propose a novel evaluation metric for UniDA, the Universal Classification Rate (UCR), which is insensitive to threshold and ratio considerations. Additionally, in order to facilitate rigorous and replicable experimentation in UniDA, we have developed and made available the UniOOD framework. UniOOD simplifies the incorporation of new datasets and algorithms through a few lines of code, ensuring fairer comparisons between various methods.

2 Related works

Universal domain adaptation. Different from the traditional domain adaptation (DA) problem, which assumes all labels in the target domain are identical to the source domain, universal domain adaptation (UniDA) [38] assumes that there is no prior knowledge about the label relationship between source and target domains. Due to the existence of labels shift in UniDA, classical DA methods of adversarial adaptation such as DANN [9] often suffer from negative transfer. To address this problem, UAN [38] and CMU [8] use sample-level uncertainty criteria to assign weights for each sample before adversarial alignment. In addition to adversarial adaptation, self-training or self-supervised-based methods usually have better performance due to the exploiting of discriminative representation on the target domain. Among these, DANCE [30] uses self-supervised neighborhood clustering to learn the target data structure; DCC [20] exploits cross-domain consensus knowledge to discover discriminative clusters of both domains; MATHS [3] designs a contrastive learning scheme to nearest neighbors for feature alignment; OVANet [31] proposes to train a one-vs-all classifier for each class and applies entropy minimization to target samples during adaptation; and more recently, UniOT [2] uses optimal transport criteria to select more confident clusters to target samples for self-training. However, all of these methods are evaluated solely using models pre-trained in ImageNet. In this paper, we compare against the most state-of-the-art methods under the foundation models. To our knowledge, these methods are DANCE, OVANet, and UniOT, which we detail in Section 4.1 respectively. We show that there exists a strong baseline that can be competitive with or outperform the more complex methods listed above when using foundation models.

Adaptation of foundation models. The exceptional performance of foundation models in traditional vision tasks has led to a growing interest in developing more effective adaptive methods. In addition to adopting linear probing [15], full fine-tuning [23], or zero shot [28] to the backbone models, many new strategies or methods have been proposed. For example, prompt learning based methods [41, 40, 12] propose to learn better prompts under the language-vision models. CLIP-Adapter [10] and Tip-Adapter [39] are going to construct additional light models for efficient fine-tuning while freezing the backbone models. Surgical fine-tuning [19] suggests selectively fine-tuning a subset of layers based on different types of distribution shift. WiSE-FT [37] proposes to enhance the model robustness by integrating the zero-shot model and the fine-tuning model. And more recently, cross-model adaptation [21] shows the most powerful few-shot ability to CLIP based models by incorporating multi-modalities as training samples for ensemble training. In this paper, different from all these methods that aim to adapt models for closed-set classification task, we exploit to adapt for UniDA problem. We also show how effective would be if these representative methods are directly applied for the UniDA tasks.

Related subfields. UniDA is also closely related to open-set recognition (OSR) [32] and out-of-distribution (OOD) detection [16]. OSR extends the closed-set classification to a more realistic open-set classification, where test samples may come from domains of unknown classes. This setting is very similar to UniDA but it assumes that there exists no domain shift and that one can not access the target domain during training. OOD detection, on the other hand, focuses on detecting the out-class samples only. In theory, a recent work of [7] unifies OSR and OOD detection into the same framework and shows that the loss criterion must be carefully designed otherwise it may face an intractable learning problem. In this paper, we are inspired by these works and introduce a similar evaluation metric of UCR for UniDA. Although many methods of OSR and OOD detection have been proposed during the past few years, a recent empirical study by Vaze et al. [34] shows that a good closed-set classifier can be competitive with or even superior to previous complex methods. These findings align with our results on UniDA under the foundation models.

3 Problem formulation

In UniDA, we are provided with a source domain dataset $\mathcal{D}^s = \{(\mathbf{x}_i^s, y_i^s)\}_{i=1}^{n_s}$ consisting of n_s samples, where the i-th sample $\mathbf{x}_i^s \in \mathbb{R}^d$ is a d dimensional vector and $y_i^s \in \mathcal{Y}^s$ is the associated label. Additionally, we have a target domain dataset $\mathcal{D}^t = \{(\mathbf{x}_i^t)\}_{i=1}^{n_t}$, which contains n_t unlabeled samples from the same d-dimensional space. Samples in the source and target domains are drawn from their respective distributions, $\mathcal{D}^s \sim \pi_s(X^s, Y^s)$ and $\mathcal{D}^t \sim \pi_t(X^t, Y^t)$. We represent the collection of labels in the source domain as $\mathcal{Y}^s$ and in the target domain as $\mathcal{Y}^t$. Let $\mathcal{Y}^{st} = \mathcal{Y}^s \cap \mathcal{Y}^t$ be the domain-shared label set and $\mathcal{Y}^{t/s} = \mathcal{Y}^t \setminus \mathcal{Y}^s$ be the target-private label set. Similarly, $\mathcal{Y}^{s/t}$ is the set of source-private labels. In UniDA, we make no assumptions about $\mathcal{Y}^t$. Hence, $\mathcal{Y}^{st}$ and $\mathcal{Y}^{t/s}$ are also unknown. For convenience, we refer to target samples belonging to $\mathcal{Y}^{st}$ (known classes) as in-class samples $\mathcal{D}_{in}^t$ and those belonging to $\mathcal{Y}^{t/s}$ (unknown classes) as out-class samples $\mathcal{D}_{out}^t$.

The learning task of UniDA can be converted as two subtasks of in-class discrimination and out-class detection. Such objectives could be implemented by a unified framework as: (1) learning a scoring function $s : \mathbb{R}^d \to \mathbb{R}$ for out-class detection and (2) learning a classifier $f : \mathbb{R}^d \to \mathbb{R}^{|\mathcal{Y}^s|}$ for in-class discrimination. The scoring function s assigns a score to each sample, which reflects the uncertainty level regarding it being an out-class sample. A higher score indicates a higher likelihood of belonging to the in-class category. UniDA methods require a threshold value for the scoring function s to distinguish between out-class and in-class samples. This threshold can either be learned automatically or set manually.

Method type	Methods	Source	Target	Classifier	Scoring rule	Threshold value
Baseline	Source Only (SO)	✓	✗	softmax	negative entropy	$-\log(\lvert\mathcal{Y}^s\rvert)/2$
UniDA SOTAs	DANCE [30]	✓	✓	softmax	negative entropy	$-\log(\lvert\mathcal{Y}^s\rvert)/2$
	OVANet [31]	✓	✓	softmax	binary softmax prob.	1/2
	UniOT [2]	✓	✓	OT	maximum OT mass	$1/(n^t + T)$
CLIP adaptations	WiSE-FT [37]	✓	✗	softmax	negative entropy	$-\log(\lvert\mathcal{Y}^s\rvert)/2$
	CLIP cross-model [21]	✓	✗	softmax	negative entropy	$-\log(\lvert\mathcal{Y}^s\rvert)/2$
	CLIP zero-shot [28]	✗	✗	NN	maximum logit	-
Ours	CLIP distillation	✗	✓	softmax	negative entropy	$-\log(\lvert\mathcal{Y}^s\rvert)/2$

Table 1: A brief introduction of different methods. Previous approaches are categorized into three groups: SO, the baseline method that trains models solely on source data; DANCE [30], OVANet [31], and UniOT [2], state-of-the-art methods designed explicitly for the UniDA task; and WiSE-FT [37], CLIP cross-model [21], and CLIP zero-shot [28], three SOTA methods for CLIP model adaptation. In this paper, we introduce CLIP distillation, as detailed in Section 5. Note that our method uses the source data for confidence calibration, not for training.

Methods	ImageNet-pretrained			Random initialization			DINOv2-pretrained			CLIP-pretrained		
	H-score	H^3-score	UCR	H-score	H^3-score	UCR	H-score	H^3-score	UCR	H-score	H^3-score	UCR
Fine-tuning backbone												
SO	58.16	63.73	52.31	8.01	1.45	6.78	0.54	0.46	7.16	2.15	1.95	8.46
DANCE[30]	42.36	49.17	28.36	0.82	1.02	5.81	0.42	0.44	5.94	0.46	0.58	5.98
OVANet[31]	38.27	45.60	53.63	1.55	1.23	7.36	6.65	0.82	5.51	1.06	1.08	4.40
UniOT[2]	71.23	70.93	65.52	11.95	2.17	9.44	7.56	2.13	5.98	15.02	2.67	8.42
Freeze backbone												
SO	56.69	62.19	52.90	14.16	1.73	5.09	57.63	65.16	53.12	70.30	73.58	67.28
DANCE[30]	42.59	50.07	28.12	10.08	1.67	6.40	44.79	53.58	34.53	67.79	71.73	54.17
OVANet[31]	56.36	62.75	67.77	6.55	1.58	4.30	57.91	65.40	48.86	56.36	62.75	67.77
UniOT[2]	68.26	67.16	62.25	11.24	1.39	7.59	62.73	67.12	52.27	75.87	74.81	67.58

Table 2: Comparison results using ViT-B [33] backbone with various pre-trained models and fine-tuning modes. Results are conducted on the VisDA dataset in the open-partial UniDA setting.

Typically, the learning classifier $f = h \circ \phi$ comprises a feature extractor ϕ and a classifier head h. Prior research in UniDA primarily focuses on fine-tuning ϕ using ImageNet pre-trained backbones. In our study, we aim to explore the training of f and the scoring function s using foundation models such as CLIP backbones.

4 Empirical analysis of UniDA methods with foundation models

4.1 UniDA methods review

We briefly review some representative state-of-the-art (SOTA) methods for comparison: DANCE [30], OVANet [31], and UniOT [2]. Notably, certain other methods are not included in the comparison table due to their inferior performance when compared to these approaches. For a concise overview of these methods, see Table 1.

Source Only (SO, baseline). The Source Only (SO) method involves standard cross-entropy loss training on the source data alone. In inference, the softmax classifier f is employed for predictions, and the scoring function s is constructed based on the entropy of the softmax output probabilities, following the approach used in DANCE.

DANCE[30]. This approach not only utilizes standard training on the source data but also introduces a self-supervised loss for target feature clustering and an entropy separation loss to either align target features with the source domain or classify them as unknown classes. Following the training, a softmax classifier f is learned based on the similarities between the target features and the source prototypes. The scoring function s is calculated as the negative entropy of the softmax output. The threshold for out-class detection is set as $-\log(|\mathcal{Y}^s|)/2$.

OVANet[31]. This approach introduces a one-vs-all network for tackling the UniDA task,

Methods	Resnet50					DINOv2					CLIP				
	Office	OH	VD	DN	Avg	Office	OH	VD	DN	Avg	Office	OH	VD	DN	Avg
H-score															
SO	66.34	54.39	30.33	39.2	47.56	89.6	82.77	57.53	68.38	74.57	91.98	84.52	69.85	61.49	76.96
DANCE[30]	80.32	39.06	2.78	26.91	37.27	**91.93**	84.38	53.89	68.74	74.73	**94.7**	89.01	71.9	60.53	79.03
OVANet[31]	83.33	71.68	44.57	49.57	62.29	86.51	76.83	**58.03**	55.76	69.28	93.36	85.42	59.47	70.7	77.24
UniOT[2]	**84.37**	**75.97**	**54.48**	**50.88**	**66.42**	89.16	**87.54**	56.6	**69.86**	**75.79**	92.32	**89.45**	**79.1**	**71.42**	**83.07**
H^3-score															
SO	65.55	56.87	34.92	42.16	49.87	88.74	82.81	65.03	70.9	**76.87**	89.95	82.7	74.24	64.65	77.88
DANCE[30]	71.57	44.0	4.04	31.18	37.7	**90.3**	83.88	61.84	**71.14**	76.79	**91.64**	85.6	75.77	63.94	79.24
OVANet[31]	76.8	67.96	45.91	**50.34**	60.25	86.64	78.65	**65.45**	60.26	72.75	90.8	83.33	66.07	**71.1**	77.82
UniOT[2]	**77.33**	**70.49**	**46.01**	47.91	**60.43**	87.68	**85.5**	61.94	70.43	76.39	89.07	**87.09**	**77.69**	69.9	**80.94**
UCR															
SO	81.21	65.18	24.59	31.04	50.5	**91.2**	**84.46**	**50.36**	61.45	**71.87**	93.98	86.89	63.46	63.19	76.88
DANCE[30]	84.47	69.34	**44.13**	32.71	57.66	87.32	83.32	41.33	**63.52**	68.87	95.17	**90.33**	57.78	**64.88**	77.04
OVANet[31]	81.38	67.83	36.26	34.43	54.97	89.96	82.03	46.48	58.6	69.27	**95.36**	88.18	68.57	64.3	**79.1**
UniOT[2]	**84.74**	**73.65**	41.29	**34.52**	**58.55**	86.63	84.41	44.15	57.85	68.26	90.62	88.85	**72.22**	62.88	78.64

Table 3: Comparison results between the baseline (Source Only (SO)) and SOTAs using different backbones in the open-partial UniDA setting.

Methods	Office				OfficeHome				VisDA				DomainNet				Avg
	(10/10)	(10/0)	(31/0)	(10/21)	(10/5)	(15/0)	(65/0)	(25/40)	(6/3)	(6/0)	(12/0)	(6/6)	(150/50)	(150/0)	(345/0)	(150/195)	
H-score																	
SO	89.6	92.29	**83.99**	90.5	82.77	81.76	66.72	64.41	57.53	64.49	42.45	38.06	68.38	70.36	51.66	50.53	68.47
DANCE[30]	**91.93**	**94.98**	81.67	80.32	84.38	81.93	64.28	57.03	53.89	65.26	34.87	26.49	68.74	70.51	51.59	49.3	66.07
OVANet[31]	86.51	88.7	82.17	**92.02**	76.83	76.2	**70.6**	**71.29**	**58.03**	62.44	**56.91**	**61.51**	55.76	57.49	**58.92**	**58.51**	**69.62**
UniOT[2]	89.16	94.52	65.44	41.04	**87.54**	**85.56**	55.81	38.55	56.6	**71.56**	39.31	29.62	**69.86**	**72.64**	54.0	45.08	62.27
H^3-score																	
SO	88.74	90.73	**83.99**	90.5	82.81	82.15	66.72	64.41	65.03	66.31	42.45	38.06	70.9	72.2	51.66	50.53	69.2
DANCE[30]	**90.3**	**92.48**	81.67	80.32	83.88	82.22	64.28	57.03	61.84	66.85	34.87	26.49	**71.14**	**72.3**	51.59	49.3	66.66
OVANet[31]	86.64	88.38	82.17	**92.02**	78.65	78.24	**70.6**	**71.29**	**65.45**	64.83	**56.91**	**61.51**	60.26	61.67	**58.92**	**58.51**	**71.0**
UniOT[2]	87.68	91.9	65.44	41.04	**85.5**	**84.11**	55.81	38.55	61.94	**68.67**	39.31	29.62	70.43	72.17	54.0	45.08	61.95
UCR																	
SO	**91.2**	93.93	90.11	94.59	**84.46**	**83.82**	81.9	81.84	**50.36**	59.14	66.98	66.93	61.45	64.0	68.94	68.74	**75.52**
DANCE[30]	87.32	94.24	86.97	85.2	83.32	82.04	80.1	75.76	41.33	53.03	52.94	44.51	**63.52**	**65.85**	**69.73**	68.6	70.9
OVANet[31]	89.96	92.53	90.1	**94.64**	82.03	81.37	81.86	**81.84**	46.48	53.6	**67.0**	**66.98**	58.6	61.29	68.97	**68.74**	74.12
UniOT[2]	86.63	**95.51**	**91.02**	59.84	84.41	82.5	**82.93**	57.78	44.15	**60.66**	64.22	40.67	57.85	63.61	69.08	60.38	68.83

Table 4: Comparison results between the baseline (Source Only (SO)) and SOTAs using DINOv2 backbone in four UniDA settings (open-partial, open, closed, partial).

consisting of a standard source classifier f and $|\mathcal{Y}^s|$ binary softmax classifiers. Throughout training, each binary classifier is responsible for distinguishing source samples in its respective class from those in other classes. Once trained, the scoring function s is constructed based on the output probability of a chosen binary classifier. The threshold for out-class detection is set at 0.5.
UniOT[2]. Similar to DANCE, this approach also employs self-supervised learning to achieve target clustering and alignment. It does so by constructing source prototypes and numerous target prototypes. However, what sets it apart is its use of Optimal Transport (OT) to select confident target samples and prototypes to achieve its objective. Finally, the classifier f makes predictions based on the outcomes of the optimal transport process between target features and source prototypes. The scoring function s is created based on the maximum OT probability of target samples. The threshold $(1/(n^t + T))$ for identifying out-class samples is dynamically adjusted according to T, which is task-specific.

4.2 Key observations and suggestions

Fine-tuning backbone or not? Table 2 presents a performance comparison among three pre-trained models: ImageNet-pretrained, DINOv2-pretrained, and CLIP-pretrained. We also compare these models to training from scratch, where the backbone is initialized randomly. It is interesting to note that, in contrast to fine-tuning from the ImageNet pre-trained model, fine-tuning from foundation models (DINOv2 or CLIP) often yields significantly poorer results and, in some cases, performs even worse than training from scratch. While keeping the backbones frozen, results of the baseline method (SO) based on foundation models exhibit improvement compared to that based on the ImageNet pre-trained model. This improvement is particularly significant when using CLIP models. Building upon this observation, we maintain a frozen backbone and only update parameters in other modules when using

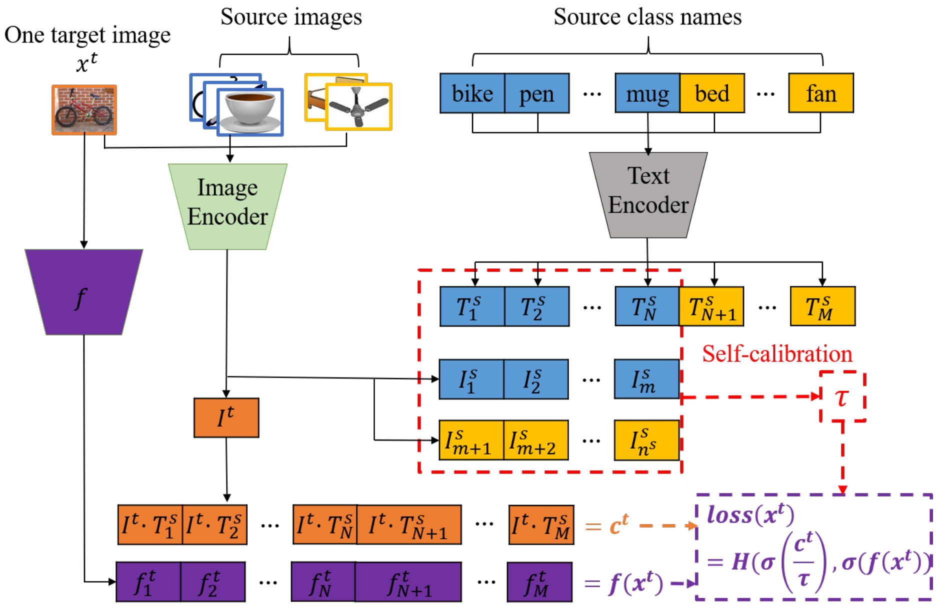
One target image
x^t
Source images
Source class names
bike
pen
···
mug
bed
···
fan
Image Encoder
Text Encoder
f
T_1^s T_2^s ··· T_N^s T_{N+1}^s ··· T_M^s
Self-calibration
I_1^s I_2^s ··· I_m^s
τ
I^t
I_{m+1}^s I_{m+2}^s ··· $I_{n^s}^s$
$I^t \cdot T_1^s$ $I^t \cdot T_2^s$ ··· $I^t \cdot T_N^s$ $I^t \cdot T_{N+1}^s$ ··· $I^t \cdot T_M^s$ $= c^t$
f_1^t f_2^t ··· f_N^t f_{N+1}^t ··· f_M^t $= f(x^t)$
$loss(x^t)$
$= H(\sigma(\frac{c^t}{\tau}), \sigma(f(x^t)))$

Figure 1: Training pipeline of *CLIP distillation* for UniDA.

foundation models.

Why using foundation models? Previous UniDA studies have only verified their results using the ImageNet pre-trained Resnet50 model. To demonstrate how these methods perform when using foundation models as their backbones, we present the comparison in Table 3. It is evident that when employing foundation models as backbones, all methods exhibit a significant improvement in performance. It indicates that the high-level features learned in foundation models are more robust than those in ImageNet pre-trained models.

Which learning algorithm is best? Table 4 presents a comparative analysis of the baseline and state-of-the-art (SOTA) methods using the DINOv2 foundation model as the backbone. To provide a comprehensive evaluation, we report results across four distinct UniDA scenarios: open-partial, open, closed, and partial, spanning four different UniDA benchmarks, namely, Office, OfficeHome, VisDA, and DomainNet. In terms of H-score/H^3-score metrics, UniOT excels in the open-partial and open settings, while the OVANet outperforms other methods in the closed and partial settings. It is noteworthy that when considering the UCR metric, all methods demonstrate similar performance across all tasks. Overall, when averaging over all tasks and all metrics, no single method consistently outperforms the others. This indicates that existing UniDA methods generally fail to improve over the baseline.

DINOv2 or CLIP? We selected two of the most powerful visual foundation models, ViT-L/14@336px from CLIP [28] and DINOv2 (dinov2_vitl14) [24], for comparison. The results presented in Tables 3, 4, and 5 demonstrate that CLIP models exhibit greater effectiveness than DINOv2 models in UniDA tasks. This observation has motivated us to develop a UniDA method using CLIP foundation models.

In conclusion, we have demonstrated that a powerful backbone is crucial for UniDA tasks. Our findings emphasize the need for further research in the UniDA field, particularly when utilizing these more powerful foundation models.

5 Proposed method

Based on the above observations, we are motivated to adapt from the most powerful CLIP models and propose *CLIP distillation* method. Let $f : \mathbb{R}^d \rightarrow \mathbb{R}^{|\mathcal{Y}^s|}$ be the classifier before softmax layer and σ be the softmax function, then the loss of the CLIP distillation to each target example $\mathbf{x}^t$ is:

$$\text{loss}(\mathbf{x}^t) = H(\sigma(\mathbf{c}^t/\tau), \sigma(f(\mathbf{x}^t))), \quad (1)$$

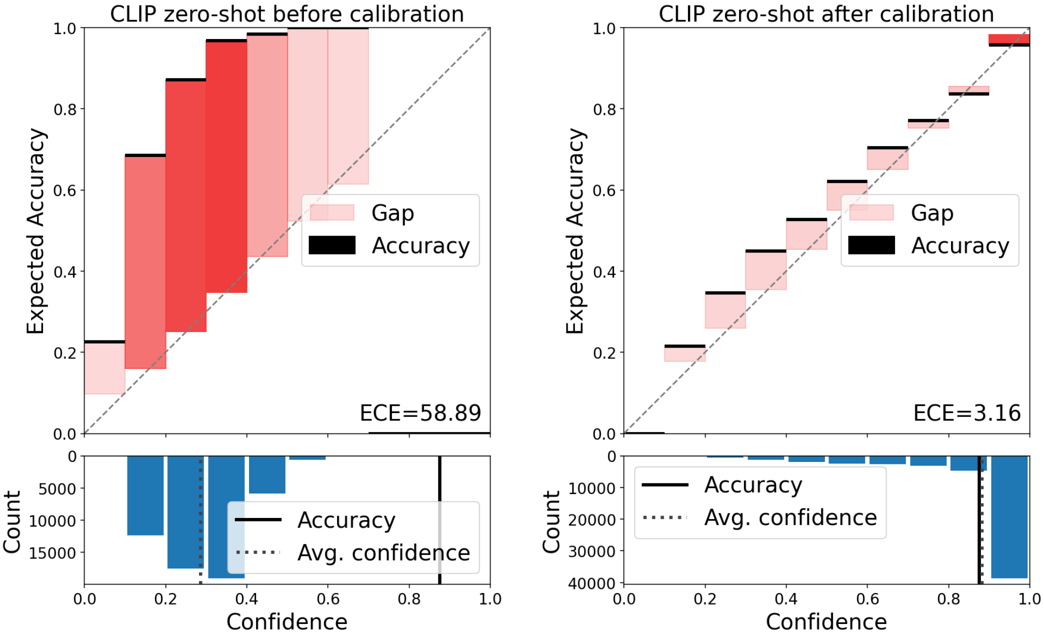

Figure 2: Reliability diagrams(top) and confidence histograms (bottom) for CLIP zero-shot model before and after calibration on VisDA dataset.

where $\mathbf{c}^t$ is the output logit of $\mathbf{x}^t$ in the CLIP zero-shot model, H is the cross-entropy, and τ is the scaling temperature, which is automatically learned using the proposed self-calibration method (5.2). The training framework for *CLIP distillation* is illustrated in Figure 1.

5.1 Motivations

Why distillation? We employ distillation [17] to address UniDA using the CLIP model for two main reasons. First, distillation provides a straightforward means to learn from the powerful CLIP model without updating its parameters, thus preventing the risk of destabilizing the CLIP model. Second, given that the CLIP model already yields promising results on the closed-set out-of-distribution (OOD) data [28], distillation from target data resembles a self-training technique, which is theoretically well-grounded [36].
Why scaling the logits? In UniDA tasks, however, the objective is not solely closed-set classification but also demands effective out-class detection. We argue that a well-calibrated model plays a pivotal role in achieving this goal. To gain a clearer perspective on our argument, we illustrate two reliability diagrams [4] comparing the CLIP zero-shot model before and after calibration in Figure 2. As depicted in the figure, without logit scaling, the CLIP zero-shot method tends to classify most samples with low confidence, even if they are classified correctly. This would readily lead to misidentifying the majority of in-class samples as out-class ones, causing a decrease in in-class classification performance. After calibration through temperature scaling, significantly improved confidence estimates can be observed, resulting in a better trustworth prediction system. Therefore, we scale the logits to ensure the model's proper calibration and enhance its performance in both out-class detection and in-class discrimination [13].

5.2 Learning temperature scaling by source confidence calibration

Confidence calibration by temperature scaling faces a challenge for UniDA tasks since we do not have prior knowledge about the target categories. To address this challenge, we propose to learn using the source data. We evenly divide the source data into two parts by class. The first part of the samples is treated as in-class samples for IID calibration, while the second part of the samples is treated as out-class samples for OOD calibration.

IID calibration. Given a ground truth joint distribution $\pi_{in}(X, Y) = \pi_{in}(Y|X)\pi_{in}(X)$, the expected calibration error (ECE) for a prediction model is defined as

$$\mathbb{E}_{\hat{P}}[|\mathbb{P}(\hat{Y} = Y|\hat{P} = p) - p|], \tag{2}$$

where $\hat{Y}$ is a class prediction and $\hat{P}$ is its associated confidence, i.e. probability of correctness. ECE could be approximated by partitioning predictions into K equally-spaced bins (similar to the reliability diagrams, see Figure 2) and taking a weight average of the bins' accuracy/confidence difference [22]. Specifically,

$$\text{ECE}_{in} = \sum_{k=1}^{K} \frac{|B_k|}{n_{in}} |\text{acc}(B_k) - \text{conf}(B_k)|, \tag{3}$$

where n_{in} is the number of in-class samples.

OOD calibration. For out-class samples $\{x_i\}_{i=1}^{n_{out}}$, which do not belong to any specific predicted category, our objective is to maintain a uniform distribution of their output class probabilities:

$$\text{ECE}_{out} = \frac{1}{n_{out}} \sum_{i=1}^{n_{out}} |\text{conf}(x_i) - \frac{1}{N}|, \tag{4}$$

where N is the number of in-class categories.

Negative log likelihood. As a standard measure of probabilistic model's quality [14], netagive log likelihood (NLL) is widely used in the context of deep learning, which is also known as the cross entropy loss. Given the known ground truth of in-class samples and a prediction probabilistic model $\hat{\pi}_{in}(Y|X)$, NLL is defined as:

$$\text{NLL}_{in} = -\sum_{i=1}^{n_{in}} \log(\hat{\pi}_{in}(y_i|x_i)) \tag{5}$$

NLL_{in} is minimized if and only if $\hat{\pi}_{in}(Y|X)$ recovers the ground truth conditional distribution $\pi_{in}(Y|X)$.

Our overall objective of learning temperature scaling is then written as

$$\tau_{opt} = \arg\min_{\tau} \text{ECE}_{in} + \text{ECE}_{out} + \text{NLL}_{in}. \tag{6}$$

6 Experiments

In this section, we conduct a comparative analysis of our method to demonstrate its robustness and effectiveness in tackling UniDA. We employ CLIP as the backbone due to its superior performance, as outlined in Section 4.2.

6.1 Datasets and experimental setup

Dataset. We train the above methods on the standard benchmark datasets for UniDA: Office [29], OfficeHome (OH) [35], VisDA (VD) [26], and DomainNet (DN) [25]. Office has 31 categories and three domains: Amazon (A), DSLR (D), and Webcam (W). OfficeHome contains 65 categories and four domains: Art (A), Clipart (C), Product (P), and Real-World (R) images. VisDA is a synthetic-to-real dataset with 12 categories in total. DomainNet is the largest dataset, including 345 categories and six domains, where three domains – Painting (P), Real (R), and Sketch (S) – are used in experiments following previous work [31, 2]. For each dataset, we further split the total categories into three disjoint parts – common categories $\mathcal{Y}^{st}$, source private categories $\mathcal{Y}^{s/t}$, and target private categories $\mathcal{Y}^{t/s}$ – to consist of the source and target domains. For a more comprehensive study, we assign each dataset with four different class splits: open-partial, open, closed, partial, following [30]. The different classes splits result in different running tasks, and each split setting, denoted ($|\mathcal{Y}^{st}|/|\mathcal{Y}^{s/t}|$), is shown in each table.

Methods	Office				OfficeHome				VisDA				DomainNet				Avg
	(10/10)	(10/0)	(31/0)	(10/21)	(10/5)	(15/0)	(65/0)	(25/40)	(6/3)	(6/0)	(12/0)	(6/6)	(150/50)	(150/0)	(345/0)	(150/195)	
								H-score									
SO	91.98	91.87	80.22	89.79	84.52	82.05	58.12	58.31	69.85	75.79	55.31	57.19	61.49	65.63	38.27	35.88	68.52
DANCE[30]	**94.7**	96.09	75.76	66.83	89.01	83.95	55.42	46.63	71.9	74.3	58.08	49.5	60.53	65.24	37.56	30.92	66.03
OVANet[31]	93.36	91.16	74.64	87.53	85.42	80.29	64.65	65.92	59.47	39.27	43.55	42.58	70.7	72.4	57.22	55.86	67.75
UniOT[2]	92.32	**96.48**	59.95	41.31	**89.45**	**86.64**	59.27	43.6	79.1	**83.08**	71.62	62.03	71.42	73.21	**63.72**	55.18	70.52
WiSE-FT[37]	82.34	94.07	47.87	53.57	79.37	73.44	13.64	16.56	62.68	72.21	30.05	27.4	3.74	7.92	0.3	0.29	41.59
CLIP cross-model[21]	93.04	93.59	83.21	92.55	86.2	84.26	62.65	63.14	77.69	81.66	62.08	67.98	61.98	67.1	36.2	34.06	71.71
CLIP distillation ($\tau = 1$)	0.0	0.07	0.0	0.0	0.17	1.31	0.0	0.0	0.0	0.05	0.0	0.0	0.0	0.0	0.0	0.0	0.1
CLIP distillation (Ours)	87.46	91.84	83.34	94.32	87.37	85.37	77.76	79.52	84.73	82.24	74.03	81.83	**73.48**	**74.64**	60.16	65.89	80.25
CLIP distillation (Ours, fixed model)	86.74	91.89	**83.8**	**94.39**	86.4	84.77	**80.58**	**80.81**	**84.74**	82.26	**76.08**	**83.12**	72.37	74.22	**74.93**	**75.22**	**82.02**
								H^3-score									
SO	89.95	89.79	80.22	89.79	82.7	81.14	58.12	58.31	74.24	76.9	55.31	57.19	64.65	67.5	38.27	35.88	68.75
DANCE[30]	**91.64**	92.4	75.76	66.83	85.6	82.43	55.42	46.63	75.77	75.86	58.08	49.5	63.94	67.22	37.56	30.92	65.97
OVANet[31]	90.8	89.42	74.64	87.53	83.33	79.94	64.65	65.92	66.07	47.2	43.55	42.58	71.1	72.1	57.22	55.86	68.24
UniOT[2]	89.07	**93.24**	59.95	41.31	**87.09**	**85.16**	59.27	43.6	77.69	78.17	71.62	62.03	69.9	70.83	**63.72**	55.18	69.24
WiSE-FT[37]	83.4	91.19	47.87	53.57	79.4	75.32	13.64	16.56	68.68	74.4	30.05	27.4	5.46	11.21	0.3	0.29	42.42
CLIP cross-model[21]	90.56	90.87	83.21	92.55	83.73	82.54	62.65	63.14	79.96	80.82	62.08	67.98	65.02	68.54	36.2	34.06	71.49
CLIP distillation ($\tau = 1$)	0.0	0.11	0.0	0.0	0.25	1.93	0.0	0.0	0.0	0.08	0.0	0.0	0.0	0.0	0.0	0.0	0.15
CLIP distillation (Ours)	86.9	89.74	83.34	94.32	84.39	83.18	77.76	79.52	84.8	81.2	74.03	81.83	**73.0**	**73.6**	60.16	65.89	79.6
CLIP distillation (Ours, fixed model)	86.45	89.77	**83.8**	**94.39**	83.73	82.74	**80.58**	**80.81**	**84.8**	**81.21**	**76.08**	**83.12**	72.08	73.18	**74.93**	**75.22**	**81.43**
								UCR									
SO	93.98	94.95	91.44	96.99	86.89	84.53	83.55	84.44	63.46	71.17	76.66	79.51	63.19	66.01	71.26	70.78	79.93
DANCE[30]	95.17	97.09	87.69	81.1	90.33	86.76	81.74	75.63	57.78	63.45	67.86	56.4	64.88	68.37	71.66	67.6	75.84
OVANet[31]	95.36	95.8	91.4	96.84	88.18	85.72	83.52	84.37	68.57	66.45	76.68	79.58	64.3	66.94	71.37	70.94	80.38
UniOT[2]	90.62	97.2	92.14	55.94	88.85	85.59	85.5	63.61	72.22	78.8	84.79	74.78	62.88	67.33	73.85	67.99	77.63
WiSE-FT[37]	95.27	96.33	92.3	97.57	90.77	89.28	87.28	88.44	70.83	77.88	81.45	84.43	68.72	71.66	75.74	75.77	83.98
CLIP cross-model[21]	**95.38**	96.18	**93.24**	**97.58**	89.71	87.82	86.95	87.97	73.22	79.06	81.15	83.76	68.81	71.53	75.57	75.55	83.97
CLIP zero-shot[28]	90.1	97.68	87.69	96.61	90.21	89.67	89.08	89.43	78.6	82.86	87.56	88.1	70.78	73.34	79.48	79.87	85.69
CLIP distillation ($\tau = 1$)	92.46	97.75	87.68	96.61	92.91	**91.71**	89.08	89.41	80.9	85.75	87.56	88.1	69.2	72.81	79.48	79.88	86.33
CLIP distillation (Ours)	93.76	**97.92**	87.88	96.61	92.91	91.49	**89.71**	**89.91**	**82.59**	**86.39**	**88.11**	**88.81**	73.08	74.93	**80.33**	**82.03**	**87.28**
CLIP distillation (Ours, fixed model)	93.39	97.91	87.69	96.61	**93.02**	**91.73**	89.08	89.43	82.38	86.37	87.56	88.1	**74.86**	**77.21**	79.49	79.87	87.17

Table 5: Comparison results between existing methods and the proposed method using CLIP backbone in four UniDA settings (open-partial, open, closed, partial).

The detail information about these four datasets and the class-split settings are presented in Appendix.

Implementation. For fair comparison between different methods, we implement UniOOD, a code framework to streamline rigorous and reproducible experiments in UniDA. By using the UniOOD framework, all methods are run under the same learning setting. The initial learning rate is set to 0.01 for all new layers and 0.001 for pre-trained backbone if it is fine-tuned and decays using the cosine schedule rule with a warmup of 50 iterations. We use SGD optimizer with momentum 0.9 and the batch size is set to 32 for each domain. The number of training iterations are set to 5000, 10000, or 20000 based on the scale of the training data, which is detailed in Appendix. We report results of the last checkpoint due to the absence of validation data and average them among three random runs. Due to space constraints, we provide the average results for each split setting, while the detailed results for individual tasks can be found in the appendix. Hyperparameters for previous methods follow their official codes. We do not use any data augmentation during training for fair comparison to different methods, which may be different from previous works.

6.2 Evaluation and discussion

As Universal Domain Adaptation (UniDA) encompasses a dual objective, it aims to not only reject samples from unknown classes $\mathcal{Y}^{t/s}$ but also accurately classify samples from the correct classes $\mathcal{Y}^{st}$. This makes the evaluation of the UniDA method more complex. There are various evaluation metrics designed to handle the unknown classes $\mathcal{Y}^{t/s}$ in different ways. However, each of these metrics has certain drawbacks, which we discuss in detail for each of them.

6.2.1 Hard out-class detection criteria

Average class accuracy: The initial metric used to evaluate UniDA is the average class accuracy, calculated over a total of $|\mathcal{Y}^{st}| + 1$ classes, including all unknown classes $\mathcal{Y}^{t/s}$ grouped together as a superclass [38]. The drawback of this metric is that it is highly sensitive to the number of shared classes $\mathcal{Y}^{st}$. Having a significant number of shared classes would undeniably render the detection of unknown classes trivial. However, in such scenarios, there is a possibility that the number of out-class samples exceeds the number of in-class samples by multiple folds. One may argue that a simplified weighted accuracy might solve this issue, but we lack prior knowledge to determine the appropriate weights for the different classes.

H-score: The H-score is later proposed to balance the importance of detecting samples outside the class and classifying in the class samples [8]. H-score also includes all unknown classes as a superclass but calculates the harmonic mean of the average classes accuracy on known classes (acc$_{in}$) and the accuracy on the superclass (acc$_{out}$), i.e., H-score = 2·acc$_{in}$·acc$_{out}$/(acc$_{in}$+acc$_{out}$). This metric is more reasonable than average class accuracy introduced above, but also has a significant bias to the ratios between the numbers of in-class and out-class samples. Due to the lack of prior knowledge to target data, these ratios may diverse in different tasks. It is usually impossible to handle all tasks of different ratios in order to have a fair evaluation between different methods.

H^3-score: In addition to the dual objective of UniDA, the quality of clustering for target private samples is introduced in [2] as an additional objective to facilitate the discovery of target private classes. The H^3-score is calculated as 3/((1/acc$_{in}$)+(1/acc$_{out}$)+(1/NMI)), where Normalized Mutual Information (NMI) is the widely used metric for clustering. While H^3-score provides a more comprehensive evaluation, incorporating additional NMI into UniDA is beyond the scope of the current study. Furthermore, H^3-score faces similar challenges as those encountered with H-score.

It is worth noting that in scenarios where the target data lacks unknown classes, the H-score and H^3-score metrics lose their applicability and degenerate into acc$_{in}$.

6.2.2 Soft out-class detection criteria

The criterias mentioned above require us to classify a sample as either out-class or in-class, which means that we have to set a threshold for out-class detection. In this paper, we are motivated from the field of open set recognition (OSR) (Open Set Classification Rate (OSCR) [5] and the Detection and Identification Rate (DIR)[27]) and introduce a new UniDA evaluation metric, which is threshold- and ratio-free. However, unlike the OSR task, which assumes the absence of source private classes and the presence of target private classes, UniDA is more flexible and does not impose such strict constraints. Therefore, we adapt these metrics to accommodate various UniDA scenarios, introducing a new metric called Universal Classification Rate (UCR).

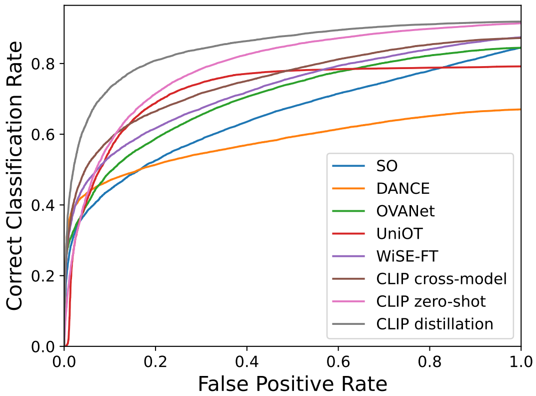

Figure 3: (CCR vs FPR) curve.

Universal classification rate (UCR). To calculate UCR, we compute a pair of Correct Classification Rate (CCR) and False Positive Rate (FPR) by varying the scoring threshold θ. CCR assesses the proportion of correctly classified in-class samples from $\mathcal{D}^t_{in}$, and FPR quantifies the fraction of out-class samples from $\mathcal{D}^t_{out}$ that are incorrectly detected.

$$\begin{aligned}\mathrm{CCR}(\theta) &= \frac{|\{\mathbf{x}|\mathbf{x}\in\mathcal{D}^t_{in}\wedge f(\mathbf{x})=\mathrm{label}(\mathbf{x})\wedge s(\mathbf{x})>\theta\}|}{|\mathcal{D}^t_{in}|}\\ \mathrm{FPR}(\theta) &= \frac{|\{\mathbf{x}|\mathbf{x}\in\mathcal{D}^t_{out}\wedge s(\mathbf{x})>\theta\}|}{|\mathcal{D}^t_{out}|}.\end{aligned} \tag{7}$$

Then, the UCR is calculated as

$$\text{UCR} = \begin{cases} \text{Area Under the (CCR vs FPR) Curve}, & \text{if } |\mathcal{D}^t_{out}| > 0 \\ \text{CCR}(-\infty), & \text{if } |\mathcal{D}^t_{out}| = 0 \end{cases} \tag{8}$$

where, CCR($-\infty$) is identical to the closed-set classification accuracy on $\mathcal{D}^t_{in}$. Figure 3 shows an example illustration to the (CCR vs FPR) curve on VisDA task under the (6/3) setting. The distinction between UCR and AUROC lies in the replacement of the true positive rate (TPR) with the correct classification rate (CCR) in UCR. In contrast to previous evaluation metrics, UCR does not rely on thresholds or ratios, making it an additional criterion that does not consider the threshold effects.

6.3 Comparison with SOTA UniDA methods

Table 5 presents the comparative results between our method and existing state-of-the-art (SOTA) UniDA approaches across three distinct evaluation metrics. The results encompass four distinct UniDA settings, namely open-partial, open, closed, and partial settings, denoted as different class splits represented as ($|\mathcal{Y}^{st}|/|\mathcal{Y}^{s/t}|$) within the table. Regarding the H-score and H^3-score metrics, it is evident that our method and the leading state-of-the-art approach are on par with each other in the open-partial and open settings. However, our method exhibits a substantial improvement (>10%) in six out of eight tasks in the closed and partial settings. In terms of the UCR metric, our method significantly outperforms state-of-the-art UniDA methods in three out of four datasets: OfficeHome, VisDA, and DomainNet, across all four settings. In general, our method exhibits superior robustness across various settings and establishes a new state-of-the-art on UniDA benchmarks, excelling in both the H-score/H^3-score metrics and the UCR metric.

6.4 Comparison with SOTA CLIP-adaptation methods

Recall that our focus is on developing a UniDA method based on foundation models like CLIP. Therefore, we also provide comparisons with some state-of-the-art (SOTA) adaptation methods that leverage CLIP models, even though they were not originally designed for the UniDA task. These methods include CLIP zero-shot (baseline) [28], WiSE-FT [37], and CLIP cross-model [21]. WiSE-FT is a new fine-tuning method for improving robustness by ensembling the weights of the zero-shot and fine-tuned models. CLIP cross-model is a recent study introduced by Lin et al. [21], which has demonstrated the most remarkable few-shot capability to date by leveraging cross-model information. However, as all these methods can not directly be used for UniDA, we construct a scoring function s following the SO method except for CLIP zero-shot, as illustrated in Table 1. While these methods have displayed remarkable enhancements in closed-set robustness benchmarks like ImageNet, they frequently exhibit lower performance than the SOTA UniDA methods when evaluated using the H-score/H^3-score metric on UniDA benchmarks, as shown in Table 5. However, it is noteworthy that all these adaptation methods consistently outperform the SOTA UniDA methods when considering the UCR metric. Our method maintains its position as the most powerful performer in terms of both H-score/H^3-score and UCR evaluation metrics.

6.5 Analysis and ablation study

Temperature scaling is necessary. To demonstrate the effectiveness of our temperature scaling, we report the results of the CLIP distillation methods when setting $\tau = 1$, as shown in Table 5. It's apparent that without temperature scaling, CLIP distillation struggles to distinguish samples between in-class and out-class categories, leading to nearly zero performance on the H-score/H^3-score metrics, though its UCR results are marginally lower compared to those with appropriate scaling. This demonstrates the necessity of temperature scaling and the superiority of our self-calibration method.

Distillation helps improve UCR but not H-score. Comparing the results between CLIP distillation with a model that gets updated and one with a fixed model (Table 5), we conclude

Methods	Office				OfficeHome				VisDA				DomainNet				Avg
	(10/10)	(10/0)	(31/0)	(10/21)	(10/5)	(15/0)	(65/0)	(25/40)	(6/3)	(6/0)	(12/0)	(6/6)	(150/50)	(150/0)	(345/0)	(150/195)	
H-score																	
w/o IID calibration	**90.31**	76.33	77.52	89.3	**87.61**	84.32	70.47	72.47	82.26	58.09	57.84	62.03	72.46	73.69	57.21	61.43	73.33
w/o NLL calibration	88.52	73.42	74.77	86.78	87.26	83.54	64.86	67.27	78.74	54.6	54.8	58.1	71.39	72.82	55.02	58.54	70.65
w/o OOD calibration	53.55	72.84	**86.96**	**96.54**	74.54	74.21	**83.42**	**86.5**	78.86	73.11	**78.16**	**86.44**	72.97	73.83	**61.85**	**68.96**	76.42
Ours	87.46	**91.84**	83.34	94.32	87.37	**85.37**	77.76	79.52	**84.73**	**82.24**	74.03	81.83	**73.48**	**74.64**	60.16	65.89	**80.25**
H^3-score																	
w/o IID calibration	**88.74**	79.13	77.52	89.3	**84.68**	82.64	70.47	72.47	83.13	63.75	57.84	62.03	72.34	72.98	57.21	61.43	73.48
w/o NLL calibration	87.5	76.96	74.77	86.78	84.49	82.15	64.86	67.27	80.7	60.91	54.8	58.1	71.61	72.41	55.02	58.54	71.05
w/o OOD calibration	58.74	76.26	**86.96**	**96.54**	75.44	75.27	**83.42**	**86.5**	80.78	75.01	**78.16**	**86.44**	72.64	73.06	**61.85**	**68.96**	77.25
Ours	86.9	**89.74**	83.34	94.32	84.39	**83.18**	77.76	79.52	**84.8**	**81.2**	74.03	81.83	**73.0**	**73.6**	60.16	65.89	**79.6**
UCR																	
w/o IID calibration	93.72	97.88	87.76	96.61	93.16	91.82	89.51	89.77	82.3	86.16	87.63	88.19	73.99	75.92	80.28	81.8	87.28
w/o NLL calibration	93.65	97.88	87.74	96.61	**93.17**	**91.88**	89.43	89.75	82.18	86.15	87.61	88.16	**74.17**	**76.16**	80.24	81.7	87.28
w/o OOD calibration	90.6	96.13	**88.18**	**96.67**	87.97	85.86	**89.99**	**90.15**	**83.41**	84.53	**88.56**	**89.67**	71.72	73.13	**80.39**	**82.24**	86.2
Ours	**93.76**	**97.92**	87.88	96.61	92.91	91.49	89.71	89.91	82.59	**86.39**	88.11	88.81	73.08	74.93	80.33	82.03	**87.28**

Table 6: Ablation studies.

		ViT-B/16					ViT-L/14					ViT-L/14@336px				
		Office	OH	VD	DN	Avg	Office	OH	VD	DN	Avg	Office	OH	VD	DN	Avg
open-partial	H-score	83.36	84.66	82.11	66.16	79.07	87.19	87.04	84.66	72.54	82.86	87.46	87.37	84.73	73.48	83.26
	H^3-score	81.59	80.22	81.79	67.09	77.67	85.25	84.1	85.04	72.36	81.69	86.9	84.39	84.8	73.0	82.27
	UCR	81.46	90.4	82.11	64.14	79.53	92.05	92.33	82.58	71.99	84.74	93.76	92.91	82.59	73.08	85.58
open	H-score	84.22	81.41	84.05	67.81	79.37	90.76	85.14	80.93	73.64	82.62	91.84	85.37	82.24	74.64	83.52
	H^3-score	83.78	78.36	81.43	68.11	77.92	89.4	82.97	80.42	73.1	81.47	89.74	83.18	81.2	73.6	81.93
	UCR	93.3	87.8	86.37	66.41	83.47	97.71	90.83	86.11	73.87	87.13	97.92	91.49	86.39	74.93	87.68
closed	H-score/H^3-score	73.51	69.37	70.11	52.58	66.39	82.0	76.95	68.83	58.82	71.65	83.34	77.76	74.03	60.16	73.82
	UCR	79.25	83.89	87.31	74.02	81.12	87.35	89.17	87.56	79.58	85.91	87.88	89.71	88.11	80.33	86.51
partial	H-score/H^3-score	85.58	74.64	77.57	57.41	73.8	93.31	79.32	75.98	64.49	78.28	94.32	79.52	81.83	65.89	80.39
	UCR	86.56	86.54	88.29	76.58	84.49	95.89	89.45	88.08	81.34	88.69	96.61	89.91	88.81	82.03	89.34

Table 7: Results of CLIP distillation based on different CLIP models.

that while distillation for updating models slightly enhances the UCR, it does not show a consistent improvement in the H-score/H^3-score. Nevertheless, distillation provides other advantages when applied to a smaller model.

Each calibration loss plays a key role. We conducted ablation studies to assess the significance of each calibration loss component in our method, namely ECE_{in}, ECE_{out}, and NLL_{in}, corresponding to IID calibration, OOD calibration, and NLL calibration, respectively. The results of this analysis are presented in Table 6. It is evident that the absence of either IID or NLL calibration leads to a substantial decrease in performance in the closed and partial settings. Conversely, the lack of OOD calibration affects the results in the open-partial and open settings. In summary, each calibration loss contributes significantly to the calibration process.

Our method is stable on different CLIP models. We present the results of our method when executed on various CLIP models, as outlined in Table 7. The findings reveal that our method exhibits stability when deployed on different backbones, with a modest decrease in performance for smaller models.

7 Conclusion, limitations and future work

In this paper, inspired by the robustness of large-scale pre-trained models to distribution shifts, we set out to develop a UniDA method utilizing these foundation models. We initially conducted comprehensive experiments to evaluate how the existing state-of-the-art UniDA methods perform when applied to foundation models. Our analysis of the results revealed several noteworthy findings, indicating the necessity for further research in the context of UniDA with foundation models. As a response to these insights, we introduced a straightforward method involving target data distillation, which establishes a new state-of-the-art in UniDA using CLIP models. The significant improvements over previous results demonstrate the promising potential of employing foundation models for UniDA tasks. We hope that our investigation and the introduction of this straightforward framework can act as a robust baseline, thus promoting future research in this domain.

Our work has certain limitations. For instance, we focused on freezing the encoder when using foundation models due to the subpar results observed in full fine-tuning. However, recent studies have introduced new techniques for improving full fine-tuning with these models, such as the fine-tuning pre-trained methods [11] and surgical fine-tuning [19]. We did not explore these techniques due to the substantial computational resources required, leaving this as a potential avenue for future research. Furthermore, our method does not incorporate source data during the training process. We anticipate that future work can enhance our approach by leveraging information from the source data to further improve its performance.

A Experimental setup details

Dataset: We provide detail information about four datasets – Office [29], OfficeHome (OH) [35], VisDA (VD) [26], and DomainNet (DN) [25] – in Table A1.

Office (31 categories)	Domains	Amazon (A)	DSLR (D)	Webcam (W)	-
	Number of Samples	2817	498	795	-
OfficeHome (65 categories)	Domains	Art (A)	Clipart (C)	Product (P)	RealWorld (R)
	Number of Samples	2427	4365	4439	4357
VisDA (12 categories)	Domains	Syn (S)	Real (R)	-	-
	Number of Samples	152397	55388	-	-
DomainNet (345 categories)	Domains	Painting (P)	Real (R)	Sketch (S)	-
	Number of Samples	50416	120906	48212	-

Table A1: Datasets information.

Classes split settings: The total categories of each dataset are split into the three disjoint parts – common categories $\mathcal{Y}^{st}$, source private categories $\mathcal{Y}^{s/t}$, and target private categories $\mathcal{Y}^{t/s}$ – to consist source and target domains. Since $|\mathcal{Y}^{st}| + |\mathcal{Y}^{s/t}| + |\mathcal{Y}^{t/s}|$ is fixed, we name each split setting as ($|\mathcal{Y}^{st}|/|\mathcal{Y}^{s/t}|$). The split settings for different datasets are shown in Table A2, following previous setting protocols [30]. Note that we only assign two split settings to DomainNet is because of the absence of samples in some categories in the Painting domain.

Datasets	Split settings			
	open-partial	open	closed	partial
Office	(10/10)	(10/0)	(31/0)	(10/21)
OfficeHome	(10/5)	(15/0)	(65/0)	(25/40)
VisDA	(6/3)	(6/0)	(12/0)	(6/6)
DomainNet	(150/50)	(150/0)	(345/0)	(150/195)

Table A2: Classes split settings on four datasets.

Number of training iterations: The maximum number of training iterations for the model is determined based on the scale of the training dataset. It is set to either 5000, 10000, or 20000 for different task settings, as indicated in Table A3.

Datasets	Split settings			
	open-partial	open	closed	partial
Office	5000	5000	10000	10000
OfficeHome	5000	5000	10000	10000
VisDA	10000	10000	20000	20000
DomainNet	10000	10000	20000	20000

Table A3: Training iterations on different task settings.

Text template using for CLIP zero-shot method: We follow the ensemble text templates in [21] for CLIP zero-shot method, which include 180 templates. Each class prototype is calculated as the mean vector of the 180 corresponding text encoding vectors.
Compute description: Our computing resource is a single GPU of NVIDIA GeForce RTX 3090 with 32 Intel(R) Xeon(R) Silver 4215R CPU @ 3.20GHz.
Existing codes used: To fair comparison to different methods, we build a code farmework – UniOOD, which integrates many previous methods. All codes to implement previous methods are directly copied from their official codes:

DANCE [30]: `https://github.com/VisionLearningGroup/DANCE`;
OVANet [31]: `https://github.com/VisionLearningGroup/OVANet`;
UniOT [2]: `https://github.com/changwxx/UniOT-for-UniDA`;

WiSE-FT [37]: `https://github.com/mlfoundations/wise-ft`;

CLIP cross-model [21]: `https://github.com/linzhiqiu/cross_modal_adaptation`.

The use of DINOv2 [24] and CLIP models [28] follows `https://github.com/facebookresearch/dinov2` and `https://github.com/openai/CLIP` respectively.

B Detail experimental results

Methods	A2D	A2W	D2A	D2W	W2A	W2D	Avg
H-score							
SO	54.48±3.1	53.47±1.48	79.11±0.25	76.59±1.0	73.56±1.42	60.82±1.84	66.34
DANCE	77.87±0.94	75.97±0.92	77.0±1.74	90.92±2.59	72.37±2.69	87.79±2.16	80.32
OVANet	**82.95**±0.56	**77.61**±1.19	67.91±2.24	**94.99**±0.3	81.25±0.37	**95.26**±0.27	83.33
UniOT	77.09±1.11	76.73±0.57	**86.44**±0.42	91.22±0.79	**85.45**±0.53	89.3±0.52	**84.37**
H^3-score							
SO	60.18±2.76	59.35±1.0	66.71±0.44	75.07±0.61	65.44±0.49	66.54±1.46	65.55
DANCE	73.08±1.55	70.89±0.66	64.03±1.56	79.95±1.89	62.11±1.63	79.35±1.81	71.57
OVANet	**80.45**±0.67	**76.76**±1.04	59.72±1.45	**87.29**±1.24	67.26±0.37	**89.29**±0.93	76.8
UniOT	76.88±2.23	75.02±0.64	**69.41**±1.86	86.17±1.65	**68.86**±1.4	87.66±0.74	**77.33**
UCR							
SO	69.72±1.45	62.83±2.31	79.54±0.62	94.19±1.27	82.78±0.9	98.2±0.22	81.21
DANCE	**79.04**±2.5	**79.86**±0.87	82.61±0.48	93.21±1.29	81.68±0.72	90.42±2.18	84.47
OVANet	71.79±0.64	65.18±1.03	73.98±2.03	**97.3**±0.67	81.46±0.44	**98.54**±0.14	81.38
UniOT	72.66±2.24	72.81±2.18	**87.12**±0.98	94.74±0.75	**87.51**±0.5	93.57±2.33	**84.73**

Table B1: Office: ResNet50 & (10/10) setting

Methods	A2D	A2W	D2A	D2W	W2A	W2D	Avg
H-score							
SO	89.62±0.15	82.87±0.1	**89.1**±0.36	94.71±0.4	88.53±0.59	92.74±0.56	89.6
DANCE	**91.38**±0.26	**86.19**±0.91	87.33±1.86	**99.56**±0.18	**89.14**±2.89	**97.96**±0.42	**91.93**
OVANet	82.41±5.27	81.95±1.12	88.62±0.31	90.59±4.13	87.74±2.5	87.75±2.88	86.51
UniOT	82.89±0.24	84.82±2.02	88.29±0.3	94.79±1.23	87.1±0.94	97.09±0.61	89.16
H^3-score							
SO	89.26±0.1	86.9±0.07	**85.05**±0.22	95.23±0.27	84.7±0.36	91.3±0.36	88.74
DANCE	**90.41**±0.17	**89.31**±0.65	83.96±1.14	**98.45**±0.12	**85.05**±1.77	94.6±0.26	**90.3**
OVANet	84.28±3.73	86.22±0.82	84.76±0.19	92.37±2.9	84.2±1.55	87.99±1.94	86.64
UniOT	85.77±1.1	87.65±2.43	81.74±1.02	94.21±0.58	80.72±0.27	**95.98**±0.45	87.68
UCR							
SO	**93.29**±0.26	**85.18**±0.16	83.24±0.33	99.07±0.51	**86.48**±0.62	**99.91**±0.0	**91.2**
DANCE	84.39±0.16	79.17±0.45	78.92±5.46	**100.0**±0.0	81.83±5.82	99.61±0.24	87.32
OVANet	92.04±2.44	83.05±0.68	83.01±1.05	96.28±2.62	86.31±1.18	99.05±0.27	89.96
UniOT	77.59±2.95	82.07±2.75	**83.98**±1.7	94.6±1.0	82.89±2.59	98.67±0.19	86.63

Table B2: Office: DINOv2 & (10/10) setting

Methods	A2D	A2W	D2A	D2W	W2A	W2D	Avg
H-score							
SO	92.71±0.13	89.3±0.19	90.08±0.21	93.95±0.17	88.16±0.22	97.69±0.2	91.98
DANCE	**96.02**±0.2	90.18±0.83	93.68±2.56	**98.64**±0.21	90.23±2.67	**99.42**±0.24	**94.69**
OVANet	93.82±0.58	88.88±0.63	92.3±2.09	97.63±0.0	89.35±0.21	98.16±0.0	93.36
UniOT	84.59±1.64	**92.24**±1.23	**94.5**±1.58	94.84±1.79	**94.76**±0.95	92.99±0.61	92.32
WiSE-FT	77.89±0.42	70.19±0.1	80.83±0.22	92.68±0.17	76.75±0.33	95.7±0.37	82.34
CLIP cross-model	94.24±0.39	89.82±0.32	92.34±0.11	93.65±0.3	92.23±0.05	95.99±0.24	93.05
CLIP distillation (Ours)	91.47±0.0	85.83±0.0	86.67±0.03	82.74±0.0	87.66±0.0	90.37±0.0	87.46
H^3-score							
SO	92.41±0.09	91.54±0.13	83.22±0.12	94.74±0.12	82.12±0.13	95.65±0.13	89.95
DANCE	**94.58**±0.13	**92.15**±0.58	**85.22**±1.4	**97.87**±0.14	83.29±1.53	**96.75**±0.15	**91.64**
OVANet	93.14±0.38	91.24±0.44	84.47±1.17	97.2±0.0	82.81±0.12	95.95±0.0	90.8
UniOT	87.14±0.7	91.12±0.48	84.31±1.23	94.03±2.03	84.27±1.17	93.56±0.34	89.07
WiSE-FT	82.04±0.31	77.18±0.08	77.74±0.14	93.88±0.11	75.18±0.21	94.37±0.24	83.4
CLIP cross-model	93.41±0.25	91.9±0.22	84.5±0.06	94.54±0.2	**84.43**±0.03	94.55±0.16	90.56
CLIP distillation (Ours)	91.59±0.0	89.08±0.0	81.25±0.02	86.83±0.0	81.83±0.0	90.84±0.0	86.9
UCR							
SO	88.41±0.08	90.78±0.01	93.09±0.24	98.55±0.09	93.27±0.15	99.8±0.01	93.98
DANCE	**95.96**±0.3	93.03±1.31	92.92±3.15	**99.61**±0.03	89.5±2.19	**99.99**±0.0	95.17
OVANet	91.39±0.11	92.25±0.45	94.92±0.2	99.18±0.02	94.46±0.26	99.94±0.01	95.36
UniOT	76.31±2.68	90.53±0.4	93.58±1.57	95.05±0.95	94.31±1.24	93.94±0.64	90.62
WiSE-FT	91.82±0.07	92.29±0.05	94.53±0.1	98.52±0.04	95.0±0.08	99.48±0.3	95.27
CLIP cross-model	90.86±0.17	92.18±0.07	**95.49**±0.01	98.55±0.01	**95.42**±0.01	99.76±0.0	**95.38**
CLIP zero-shot	91.26±0.0	89.87±0.0	89.17±0.0	89.87±0.0	89.17±0.0	91.26±0.0	90.1
CLIP distillation (Ours)	93.69±0.0	**93.14**±0.0	94.48±0.0	93.14±0.0	94.47±0.0	93.62±0.0	93.76

Table B3: Office: CLIP & (10/10) setting

Methods	A2C	A2P	A2R	C2A	C2P	C2R	P2A	P2C	P2R	R2A	R2C	R2P	Avg
H-score													
SO	50.35±0.25	50.87±0.31	55.44±0.59	59.31±1.19	49.15±0.57	57.57±1.09	62.01±0.48	50.21±0.68	56.83±0.41	56.43±0.12	52.31±0.3	52.19±0.41	54.39
DANCE	39.64±3.7	38.23±5.91	38.69±3.66	38.55±3.4	13.22±0.27	37.21±0.43	51.73±2.72	43.89±1.31	43.2±1.78	29.33±4.01	44.47±3.19	50.62±0.88	39.06
OVANet	58.01±0.64	78.91±0.18	82.15±0.56	69.4±0.63	68.1±0.22	76.41±0.08	71.98±0.58	56.77±0.23	81.72±0.11	**77.94**±0.55	58.91±0.21	79.81±0.46	71.68
UniOT	**66.13**±0.97	**80.42**±0.7	**84.56**±0.54	**72.79**±0.2	**76.59**±1.66	**82.42**±0.96	**75.82**±0.96	**65.87**±0.76	**85.07**±1.14	76.61±0.3	**64.8**±1.23	**80.6**±0.44	**75.97**
H^3-score													
SO	51.29±0.3	57.77±0.29	60.11±0.52	57.45±0.76	55.1±0.45	60.21±0.78	60.58±0.38	50.49±0.28	61.05±0.32	57.72±0.08	52.1±0.12	58.55±0.33	56.87
DANCE	42.95±3.02	45.61±5.5	45.11±3.32	43.56±2.91	18.22±0.34	43.86±0.39	54.24±2.11	46.3±0.88	49.38±1.61	35.1±3.77	46.67±2.44	56.98±0.71	44.0
OVANet	53.85±0.3	77.3±0.23	77.29±0.11	65.06±0.44	68.91±0.25	72.69±0.18	66.75±0.54	53.51±0.29	76.83±0.02	**70.97**±0.53	55.14±0.38	77.25±0.29	67.96
UniOT	**59.59**±1.0	**78.39**±0.47	**78.43**±0.36	**65.85**±0.31	**74.71**±1.07	**75.91**±0.74	**69.21**±1.03	**58.94**±0.6	**79.31**±1.22	68.95±0.19	**58.33**±0.81	**78.23**±0.41	**70.49**
UCR													
SO	38.15±0.75	74.76±0.35	89.28±0.45	61.78±1.34	63.97±0.77	78.44±1.84	62.67±0.39	37.08±0.14	85.66±0.41	69.25±0.91	39.61±0.22	81.54±0.92	65.18
DANCE	**53.53**±0.77	72.81±0.82	80.43±1.28	70.95±0.91	69.87±1.65	78.64±1.24	**73.94**±0.38	**50.25**±0.52	82.14±0.77	69.4±2.15	**52.06**±0.82	78.11±0.51	69.34
OVANet	42.81±0.9	77.99±0.1	89.85±0.29	65.89±0.97	65.1±0.36	79.33±0.74	65.7±1.32	39.84±0.03	87.6±0.43	74.79±1.27	42.57±0.35	82.55±0.44	67.83
UniOT	50.72±1.15	**83.24**±1.46	**92.86**±0.45	**72.98**±1.91	**77.33**±1.93	**89.54**±1.01	68.89±2.22	47.06±1.21	**89.51**±0.93	**75.29**±1.35	51.6±1.26	**84.75**±0.71	**73.65**

Table B4: OfficeHome: ResNet50 & (10/5) setting

Methods	A2C	A2P	A2R	C2A	C2P	C2R	P2A	P2C	P2R	R2A	R2C	R2P	Avg
H-score													
SO	73.23±0.2	86.77±0.25	88.11±0.18	85.08±0.15	86.37±0.09	89.68±0.16	81.99±0.24	71.77±0.25	85.24±0.22	83.15±0.27	75.66±0.23	86.15±0.15	82.77
DANCE	79.49±0.68	84.7±0.29	89.82±0.58	85.41±0.22	90.3±0.58	91.36±0.12	80.89±0.64	70.7±1.14	85.46±0.47	85.01±0.24	77.72±0.29	91.74±0.33	84.38
OVANet	72.18±0.25	83.0±0.17	82.62±0.45	75.83±0.57	77.15±0.77	72.46±1.4	82.0±0.15	56.85±0.33	85.45±0.17	82.2±0.8	66.38±0.12	85.79±0.22	76.83
UniOT	**82.07**±0.74	**92.04**±1.57	**93.06**±0.43	**86.88**±0.37	**90.87**±2.11	**91.93**±0.18	**83.8**±0.52	**77.44**±1.85	**90.22**±0.49	**85.65**±0.61	**81.24**±0.59	**95.29**±0.26	**87.54**
H^3-score													
SO	74.52±0.14	87.67±0.17	87.28±0.11	83.16±0.09	87.4±0.06	88.31±0.1	81.17±0.16	73.5±0.18	85.38±0.15	81.93±0.17	76.18±0.15	87.25±0.1	82.81
DANCE	78.72±0.44	86.25±0.2	88.39±0.38	**83.37**±0.14	90.04±0.39	89.38±0.08	80.45±0.42	72.75±0.81	85.54±0.31	**83.12**±0.15	77.56±0.19	90.99±0.21	83.88
OVANet	73.79±0.18	85.07±0.12	83.61±0.31	77.04±0.39	80.88±0.56	76.39±1.03	81.18±0.1	62.34±0.27	85.53±0.11	81.31±0.52	69.65±0.09	87.0±0.15	78.65
UniOT	**80.22**±0.57	**90.95**±0.57	**90.42**±0.38	83.08±0.48	**90.1**±1.54	**89.63**±0.39	**81.57**±0.29	**77.13**±1.42	**88.16**±0.38	82.32±0.23	**79.47**±0.34	**92.94**±0.21	**85.5**
UCR													
SO	66.85±0.34	93.05±0.15	**95.01**±0.1	**86.34**±0.13	88.9±0.11	93.43±0.05	**82.68**±0.76	65.57±0.54	89.16±0.25	**85.73**±0.24	71.16±0.12	95.63±0.11	**84.46**
DANCE	**77.77**±0.75	85.58±0.49	93.59±0.31	82.59±0.23	87.63±0.37	**94.52**±0.04	72.09±0.98	64.88±2.32	86.04±0.52	83.27±1.01	**75.84**±0.77	96.03±0.11	83.32
OVANet	61.81±0.28	89.89±0.3	94.48±0.14	82.45±0.34	85.53±0.23	92.8±0.15	81.14±0.6	61.55±0.85	**91.35**±0.41	84.82±0.2	64.6±0.06	93.92±0.04	82.03
UniOT	72.91±1.26	**93.71**±1.57	94.84±0.59	77.37±1.08	**93.32**±2.18	93.75±0.94	80.53±1.39	**67.28**±1.81	90.8±0.79	78.69±1.21	71.56±0.56	**98.2**±0.05	84.41

Table B5: OfficeHome: DINOv2 & (10/5) setting

Methods	A2C	A2P	A2R	C2A	C2P	C2R	P2A	P2C	P2R	R2A	R2C	R2P	Avg
H-score													
SO	75.73±0.08	84.45±0.07	88.31±0.02	86.28±0.16	88.14±0.23	91.1±0.12	81.52±0.14	78.42±0.1	89.53±0.13	85.3±0.07	81.68±0.04	83.82±0.09	84.52
DANCE	82.2±0.08	**94.02**±0.01	90.2±0.06	86.63±0.16	**93.72**±0.16	**93.79**±0.02	85.35±0.21	84.47±0.09	91.7±0.11	86.44±0.11	83.97±0.05	**95.64**±0.39	89.01
OVANet	80.55±0.4	91.84±0.08	90.74±0.27	86.07±0.17	90.52±0.11	91.78±0.12	75.64±0.67	68.54±0.71	90.29±0.12	85.74±0.2	80.56±0.31	92.75±0.77	85.42
UniOT	**88.28**±0.33	92.03±0.56	**93.39**±0.31	87.61±0.56	90.19±0.44	92.09±0.82	85.99±0.37	**86.08**±0.36	92.04±0.7	86.87±0.13	**86.47**±0.43	92.35±0.57	**89.45**
WiSE-FT	68.39±0.04	89.68±0.01	90.9±0.04	69.46±0.27	87.43±0.14	87.14±0.2	68.7±0.45	59.92±0.34	85.8±0.05	78.64±0.26	73.97±0.29	92.43±0.11	79.37
CLIP cross-model	78.37±0.06	85.06±0.05	89.18±0.12	88.98±0.08	88.09±0.24	91.52±0.1	85.91±0.05	82.6±0.11	91.59±0.07	86.8±0.17	82.09±0.05	84.24±0.09	86.2
CLIP distillation (Ours)	83.7±0.0	82.21±0.02	88.65±0.03	**89.97**±0.03	83.45±0.02	89.62±0.01	**89.22**±0.07	84.43±0.01	**94.1**±0.0	**90.23**±0.0	84.75±0.01	88.14±0.01	87.37
H3-score													
SO	74.95±0.05	86.46±0.05	87.22±0.01	79.88±0.09	89.0±0.16	89.02±0.08	77.1±0.09	76.69±0.06	88.01±0.09	79.31±0.04	78.74±0.02	86.02±0.06	82.7
DANCE	79.06±0.05	**92.91**±0.01	88.44±0.04	80.08±0.09	**92.72**±0.11	**90.71**±0.01	79.34±0.12	80.45±0.06	89.4±0.07	79.97±0.06	80.14±0.03	**93.96**±0.25	85.6
OVANet	78.03±0.25	91.48±0.05	88.79±0.17	79.76±0.1	90.61±0.07	89.45±0.08	73.5±0.42	70.1±0.5	88.5±0.07	79.57±0.11	78.04±0.19	92.08±0.51	83.33
UniOT	**83.27**±0.32	91.93±0.37	**91.63**±0.23	**83.12**±0.53	90.97±0.37	90.32±0.58	**82.57**±0.81	**82.11**±0.1	90.76±0.35	**84.06**±0.56	**82.16**±0.2	92.22±0.12	**87.09**
WiSE-FT	69.99±0.03	90.04±0.01	88.89±0.02	69.49±0.18	88.52±0.09	86.46±0.13	68.98±0.3	63.84±0.26	85.58±0.03	75.36±0.16	73.79±0.19	91.87±0.07	79.4
CLIP cross-model	76.66±0.04	86.88±0.03	87.79±0.08	81.4±0.05	88.97±0.16	89.29±0.06	79.67±0.03	79.31±0.07	89.33±0.04	80.17±0.09	78.99±0.03	86.31±0.06	83.73
CLIP distillation (Ours)	79.98±0.0	84.88±0.01	87.44±0.02	81.95±0.02	85.76±0.01	88.07±0.01	81.54±0.04	80.42±0.0	**90.9**±0.0	82.1±0.0	80.62±0.01	89.0±0.01	84.39
UCR													
SO	71.07±0.09	91.59±0.09	95.69±0.06	90.87±0.12	93.28±0.11	95.59±0.06	76.68±0.1	76.23±0.13	94.01±0.04	86.36±0.06	76.35±0.19	95.01±0.08	86.89
DANCE	80.3±0.07	**96.38**±0.07	95.85±0.04	90.33±0.07	96.25±0.08	96.87±0.04	80.12±0.12	**85.02**±0.17	95.72±0.06	85.97±0.11	82.83±0.12	**98.33**±0.03	90.33
OVANet	74.67±0.25	94.7±0.11	96.66±0.08	91.45±0.12	95.04±0.09	95.75±0.11	74.86±0.48	80.2±0.12	95.05±0.03	86.19±0.25	77.77±0.15	95.88±0.4	88.18
UniOT	**83.66**±1.53	95.93±0.71	97.09±0.12	85.3±0.44	94.79±1.04	95.52±0.62	79.53±3.62	81.36±0.86	95.12±0.1	79.36±2.17	82.07±1.19	96.52±0.38	88.85
WiSE-FT	76.68±0.05	94.9±0.03	97.41±0.03	93.97±0.08	**96.57**±0.11	97.54±0.04	87.12±0.31	80.76±0.03	96.53±0.02	92.18±0.1	78.95±0.13	96.6±0.01	90.77
CLIP cross-model	74.37±0.12	93.56±0.2	96.71±0.04	92.74±0.18	95.33±0.1	96.82±0.04	86.83±0.01	80.21±0.07	96.41±0.01	89.66±0.05	77.93±0.05	95.92±0.08	89.71
CLIP zero-shot	80.3±0.0	94.82±0.0	94.29±0.0	91.45±0.0	94.82±0.0	94.29±0.0	91.45±0.0	80.3±0.0	94.29±0.0	91.45±0.0	80.3±0.0	94.82±0.0	90.21
CLIP distillation (Ours)	82.3±0.0	96.0±0.0	**97.57**±0.0	**94.74**±0.0	96.24±0.0	**97.68**±0.0	**94.8**±0.0	82.94±0.0	**98.06**±0.0	**94.96**±0.0	**82.88**±0.0	96.71±0.0	**92.91**

Table B6: OfficeHome: CLIP & (10/5) setting

Methods	P2R	P2S	R2P	R2S	S2P	S2R	Avg
H-score							
SO	42.24±0.13	40.84±0.22	45.31±0.43	41.1±0.12	28.05±0.24	37.67±0.47	39.2
DANCE	20.09±0.26	25.86±0.92	34.6±1.19	41.83±0.42	18.68±0.49	20.42±0.29	26.91
OVANet	55.29±0.08	45.35±0.2	**52.17**±0.17	44.69±0.05	44.62±0.33	55.32±0.06	49.57
UniOT	**56.75**±0.11	**47.4**±0.21	51.72±0.23	**47.23**±0.24	**46.17**±0.23	**56.02**±0.28	**50.88**
H^3-score							
SO	47.97±0.11	40.92±0.13	47.43±0.29	41.54±0.08	31.99±0.21	43.08±0.39	42.16
DANCE	26.22±0.31	29.49±0.86	38.98±1.04	42.53±0.33	23.42±0.54	26.47±0.31	31.18
OVANet	**58.24**±0.07	**44.2**±0.15	**52.12**±0.12	**43.67**±0.08	**46.09**±0.23	**57.72**±0.1	**50.34**
UniOT	55.7±0.12	42.39±0.18	47.78±0.28	42.51±0.18	44.28±0.17	54.78±0.29	47.91
UCR							
SO	43.03±0.23	25.57±0.21	36.19±0.13	26.03±0.15	21.6±0.02	33.85±0.31	31.04
DANCE	42.56±0.13	23.88±0.58	35.89±0.58	**29.33**±0.36	25.61±0.12	38.99±0.09	32.71
OVANet	43.03±0.11	27.57±0.38	**38.18**±0.02	27.65±0.26	**30.79**±0.29	39.34±0.19	34.43
UniOT	**43.41**±0.15	**27.72**±0.42	36.16±0.06	28.57±0.34	29.8±0.09	**41.48**±0.43	**34.52**

Table B7: DomainNet: ResNet50 & (150/50) setting

Methods	P2R	P2S	R2P	R2S	S2P	S2R	Avg
H-score							
SO	71.74±0.06	68.13±0.08	67.37±0.1	67.5±0.17	64.85±0.13	70.71±0.07	68.38
DANCE	71.64±0.12	68.86±0.17	67.64±0.05	68.76±0.15	65.19±0.08	70.36±0.18	68.74
OVANet	51.66±0.49	66.05±0.02	66.69±0.19	64.24±0.1	53.8±0.41	32.11±0.45	55.76
UniOT	**73.44**±0.28	**68.89**±0.16	**67.94**±0.07	**70.16**±0.18	**65.6**±0.43	**73.1**±0.04	**69.86**
H^3-score							
SO	**76.17**±0.04	67.99±0.05	70.07±0.08	67.57±0.11	68.22±0.1	75.39±0.05	70.9
DANCE	76.09±0.09	**68.47**±0.11	**70.26**±0.04	**68.4**±0.1	**68.48**±0.06	75.13±0.14	**71.14**
OVANet	59.74±0.44	66.59±0.02	69.58±0.14	65.36±0.07	59.63±0.34	40.65±0.49	60.26
UniOT	75.69±0.18	66.74±0.08	69.12±0.04	68.12±0.2	67.24±0.26	**75.67**±0.11	70.43
UCR							
SO	67.3±0.13	57.35±0.15	60.28±0.09	57.72±0.26	58.33±0.09	67.72±0.04	61.45
DANCE	**67.98**±0.22	**60.11**±0.14	**62.85**±0.07	**61.64**±0.08	**60.58**±0.2	**67.96**±0.22	**63.52**
OVANet	64.04±0.05	53.42±0.09	58.34±0.14	53.91±0.12	56.9±0.2	64.97±0.08	58.6
UniOT	64.18±0.33	53.61±0.19	54.84±0.16	56.26±0.42	52.74±0.45	65.45±0.18	57.85

Table B8: DomainNet: DINOv2 & (150/50) setting

Methods	P2R	P2S	R2P	R2S	S2P	S2R	Avg
H-score							
SO	67.35±0.07	57.55±0.04	58.17±0.12	61.14±0.19	53.01±0.14	71.72±0.06	61.49
DANCE	67.6±0.07	57.33±0.06	55.57±0.25	59.65±0.06	52.09±0.08	70.94±0.12	60.53
OVANet	74.87±0.15	69.55±0.11	67.97±0.27	70.41±0.17	65.73±0.09	75.65±0.11	70.7
UniOT	74.67±0.39	**69.56**±0.19	**69.32**±0.47	71.38±0.21	67.15±0.52	76.42±0.16	71.42
WiSE-FT	5.79±0.04	1.51±0.04	4.24±0.14	3.71±0.03	1.77±0.04	5.45±0.09	3.74
CLIP cross-model	69.19±0.1	58.18±0.14	57.11±0.03	60.85±0.18	53.13±0.16	73.43±0.05	61.98
CLIP distillation (Ours)	**80.17**±0.02	**72.1**±0.06	67.64±0.04	**71.65**±0.04	**68.4**±0.03	**80.9**±0.01	**73.48**
H^3-score							
SO	72.52±0.06	58.72±0.03	61.83±0.09	61.17±0.13	57.84±0.11	75.83±0.04	64.65
DANCE	72.71±0.05	58.57±0.04	59.84±0.19	60.16±0.04	57.11±0.06	75.25±0.09	63.94
OVANet	78.15±0.11	66.53±0.07	**68.86**±0.18	67.06±0.1	67.31±0.06	78.72±0.08	71.11
UniOT	74.89±0.11	65.78±0.12	67.97±0.34	67.65±0.09	66.15±0.36	76.95±0.25	69.9
WiSE-FT	8.4±0.05	2.24±0.06	6.18±0.2	5.41±0.05	2.63±0.06	7.92±0.13	5.46
CLIP cross-model	73.93±0.08	59.16±0.1	61.03±0.02	60.97±0.12	57.94±0.12	77.1±0.04	65.02
CLIP distillation (Ours)	**81.92**±0.02	**68.07**±0.04	68.64±0.02	**67.8**±0.03	**69.15**±0.02	**82.43**±0.01	**73.0**
UCR							
SO	66.02±0.08	58.9±0.15	62.84±0.14	62.92±0.1	56.69±0.15	71.78±0.18	63.19
DANCE	67.79±0.1	60.3±0.19	62.77±0.05	63.68±0.09	62.86±0.15	71.88±0.16	64.88
OVANet	68.0±0.2	59.14±0.18	64.61±0.13	62.38±0.11	58.54±0.19	73.13±0.18	64.3
UniOT	68.37±0.31	57.98±0.38	59.43±0.4	61.62±0.05	57.71±0.52	72.16±0.22	62.88
WiSE-FT	73.53±0.11	63.93±0.18	66.78±0.07	66.38±0.1	64.47±0.08	77.24±0.12	68.72
CLIP cross-model	74.17±0.11	63.87±0.05	66.87±0.1	66.39±0.09	63.48±0.19	78.06±0.06	68.81
CLIP zero-shot	79.43±0.0	65.78±0.0	67.12±0.0	65.78±0.0	67.12±0.0	79.43±0.0	70.78
CLIP distillation (Ours)	**80.95**±0.01	**67.02**±0.01	**70.8**±0.01	**67.9**±0.02	**70.38**±0.01	**81.41**±0.01	**73.08**

Table B9: DomainNet: CLIP & (150/50) setting

Methods	A2D	A2W	D2A	D2W	W2A	W2D	Avg
H-score							
SO	95.4±0.19	93.49±0.03	88.88±0.21	93.46±0.71	90.11±0.29	92.37±0.82	92.29
DANCE	**96.5**±0.14	**96.13**±0.21	89.95±0.61	97.92±0.64	90.76±0.46	98.61±0.8	**94.98**
OVANet	88.37±5.41	92.18±0.39	83.36±2.25	94.15±1.29	85.06±2.35	89.07±4.13	88.7
UniOT	92.97±2.72	94.82±1.3	**90.52**±0.62	**97.97**±0.41	**90.96**±0.59	**99.9**±0.14	94.52
H^3-score							
SO	94.9±0.12	93.11±0.02	84.83±0.13	93.09±0.47	85.57±0.17	92.87±0.55	90.73
DANCE	**95.62**±0.09	**94.84**±0.14	**85.47**±0.37	96.0±0.41	**85.96**±0.28	96.99±0.52	**92.48**
OVANet	90.07±3.8	92.24±0.26	81.38±1.42	93.55±0.85	82.46±1.47	90.59±2.86	88.38
UniOT	92.52±1.89	94.19±0.7	84.88±0.73	**96.83**±0.14	85.51±0.96	**97.49**±0.29	91.9
UCR							
SO	95.54±0.42	94.09±0.34	84.6±0.12	99.48±0.03	89.88±0.22	99.96±0.0	93.93
DANCE	**97.23**±0.02	**95.83**±0.48	86.24±1.26	**99.71**±0.34	86.52±1.07	99.94±0.09	94.25
OVANet	93.18±1.39	93.55±0.41	83.22±0.74	97.97±1.0	88.57±1.37	98.66±0.34	92.53
UniOT	96.74±2.16	94.86±0.71	**90.98**±2.56	99.08±0.37	**91.39**±1.31	**100.0**±0.0	**95.51**

Table B10: Office: DINOv2 & (10/0) setting

Methods	A2D	A2W	D2A	D2W	W2A	W2D	Avg
H-score							
SO	92.04±0.19	92.01±0.13	91.71±0.27	95.51±0.09	89.77±0.33	90.17±0.43	91.87
DANCE	93.15±0.21	93.08±0.05	**96.74**±0.06	**98.74**±0.05	**95.81**±1.5	**99.0**±0.89	96.09
OVANet	93.3±1.17	90.3±0.84	82.93±3.14	97.48±0.11	85.37±0.83	97.58±1.61	91.16
UniOT	**95.11**±1.4	**95.92**±0.42	95.45±0.71	98.36±1.04	95.68±0.94	98.35±0.17	**96.48**
WiSE-FT	92.77±0.0	90.99±0.0	92.78±0.37	97.87±0.08	92.1±0.23	97.93±0.2	94.07
CLIP cross-model	93.51±0.16	92.19±0.05	93.85±0.01	94.97±0.0	92.47±0.01	94.54±0.08	93.59
CLIP distillation (Ours)	89.75±0.0	91.91±0.0	92.68±0.0	93.6±0.0	91.99±0.01	91.11±0.12	91.84
H^3-score							
SO	91.55±0.12	91.87±0.08	86.0±0.16	94.17±0.06	84.85±0.2	90.32±0.29	89.79
DANCE	92.29±0.14	92.58±0.04	**88.89**±0.03	96.24±0.03	**88.36**±0.85	96.04±0.56	92.4
OVANet	92.38±0.76	90.73±0.57	80.64±2.0	95.44±0.07	82.18±0.51	95.13±1.03	89.42
UniOT	**94.19**±0.85	**95.24**±0.09	88.09±0.89	**96.52**±0.34	88.22±0.27	**97.18**±0.13	**93.24**
WiSE-FT	92.04±0.0	91.19±0.0	86.62±0.22	95.69±0.05	86.22±0.14	95.36±0.13	91.19
CLIP cross-model	92.52±0.1	91.99±0.03	87.24±0.01	93.82±0.0	86.44±0.01	93.19±0.05	90.87
CLIP distillation (Ours)	90.04±0.0	91.81±0.0	86.57±0.0	92.92±0.0	86.16±0.01	90.94±0.08	89.74
UCR							
SO	88.49±0.1	91.81±0.21	95.35±0.18	99.32±0.03	94.83±0.18	99.88±0.02	94.95
DANCE	94.74±1.76	94.2±0.04	**97.14**±0.01	99.58±0.0	**96.94**±0.27	**99.97**±0.02	97.09
OVANet	91.89±0.35	94.05±0.29	95.31±0.41	99.42±0.02	94.91±0.22	99.24±1.05	95.8
UniOT	93.27±1.82	97.54±1.0	96.7±0.12	**99.61**±0.23	96.65±0.07	99.44±0.4	97.2
WiSE-FT	91.21±0.07	95.19±0.07	96.26±0.09	99.45±0.02	96.13±0.07	99.74±0.3	96.33
CLIP cross-model	90.77±0.27	94.46±0.33	96.53±0.0	99.35±0.0	96.13±0.05	99.85±0.0	96.18
CLIP zero-shot	98.69±0.0	98.98±0.0	95.38±0.0	98.98±0.0	95.38±0.0	98.69±0.0	97.68
CLIP distillation (Ours)	**98.74**±0.0	**99.01**±0.0	96.0±0.0	99.03±0.0	96.0±0.0	98.73±0.0	**97.92**

Table B11: Office: CLIP & (10/0) setting

Methods	A2C	A2P	A2R	C2A	C2P	C2R	P2A	P2C	P2R	R2A	R2C	R2P	Avg
H-score													
SO	74.9±0.17	85.61±0.19	87.6±0.12	82.45±0.19	86.13±0.09	86.4±0.12	76.33±0.39	71.41±0.16	83.55±0.18	83.04±0.39	77.68±0.14	85.98±0.16	81.76
DANCE	78.61±0.67	81.8±0.17	89.25±0.32	82.14±0.23	88.13±0.53	87.27±0.06	69.35±1.04	70.36±0.8	83.09±0.15	84.47±0.09	77.83±0.15	90.9±0.26	81.93
OVANet	73.36±0.11	82.65±0.11	81.65±0.47	74.49±0.3	77.98±0.67	71.38±1.27	77.13±0.35	57.0±0.44	84.02±0.37	80.74±0.49	68.85±0.04	85.1±0.19	76.2
UniOT	**80.59**±0.42	**87.84**±1.19	**91.77**±0.37	**84.98**±0.41	**89.74**±1.81	**88.21**±0.21	**80.2**±0.75	**76.9**±1.57	**86.74**±0.71	**86.03**±0.52	**80.99**±0.49	**92.69**±0.28	**85.56**
H^3-score													
SO	75.67±0.12	86.88±0.13	86.95±0.08	81.47±0.12	87.24±0.06	86.16±0.08	77.38±0.27	73.25±0.11	84.25±0.12	81.86±0.25	77.53±0.09	87.13±0.11	82.15
DANCE	78.14±0.44	84.22±0.12	88.02±0.2	81.27±0.15	88.59±0.36	**86.73**±0.04	72.45±0.75	72.51±0.56	83.93±0.1	**82.77**±0.06	77.63±0.1	90.44±0.17	82.22
OVANet	74.61±0.07	84.83±0.08	82.95±0.33	76.11±0.21	81.48±0.49	75.58±0.94	77.93±0.24	62.45±0.35	84.57±0.25	80.35±0.32	71.43±0.03	86.53±0.13	78.23
UniOT	**79.21**±0.27	**87.85**±0.96	**89.4**±0.06	**82.14**±0.23	**89.48**±1.22	86.72±0.37	**79.05**±0.56	**76.51**±1.07	**86.22**±0.36	82.32±0.21	**79.14**±0.41	**91.27**±0.26	**84.11**
UCR													
SO	71.4±0.28	**90.81**±0.07	**93.32**±0.13	**83.73**±0.09	**89.93**±0.05	89.52±0.06	77.62±0.3	67.28±0.16	88.04±0.47	**84.91**±0.29	74.81±0.03	94.45±0.17	**83.82**
DANCE	**78.01**±0.76	82.09±0.33	92.71±0.23	81.39±0.14	87.58±0.48	89.31±0.07	66.1±0.29	67.3±0.93	83.99±0.25	84.73±1.35	**76.47**±0.51	**94.86**±0.1	82.05
OVANet	67.01±0.23	88.64±0.14	92.25±0.2	79.75±0.16	86.79±0.12	87.95±0.22	76.09±0.39	63.18±0.48	**89.32**±0.34	83.57±0.26	68.92±0.07	92.92±0.1	81.37
UniOT	72.79±0.54	81.47±0.6	93.0±0.29	79.55±0.75	89.06±2.96	**90.59**±0.24	**80.06**±0.82	**69.68**±1.47	87.11±0.73	82.88±0.58	74.0±1.36	89.78±0.65	82.5

Table B12: OfficeHome: DINOv2 & (15/0) setting

Methods	A2C	A2P	A2R	C2A	C2P	C2R	P2A	P2C	P2R	R2A	R2C	R2P	Avg
H-score													
SO	75.41±0.09	83.2±0.05	86.19±0.14	81.69±0.12	87.15±0.21	86.87±0.08	75.93±0.28	76.58±0.06	86.49±0.21	83.02±0.09	78.68±0.02	83.39±0.08	82.05
DANCE	80.03±0.04	89.48±0.04	87.88±0.1	80.8±0.1	89.69±0.17	88.31±0.09	73.42±0.06	76.28±0.14	86.18±0.16	83.27±0.04	80.29±0.05	91.72±0.25	83.95
OVANet	78.42±0.21	84.47±0.43	88.3±0.27	80.71±0.41	84.84±0.45	86.53±0.11	64.85±0.7	60.25±0.53	84.74±0.22	82.66±0.17	77.01±0.2	90.66±0.49	80.29
UniOT	**84.19**±0.27	**90.91**±0.95	**90.48**±0.58	83.97±0.52	**91.71**±0.97	88.47±0.82	80.06±1.1	**82.36**±0.62	87.92±0.26	85.26±0.45	**82.58**±0.95	**91.76**±0.63	**86.64**
WiSE-FT	64.63±0.04	82.12±0.16	84.11±0.03	60.82±0.31	82.93±0.25	79.64±0.26	59.95±0.38	54.5±0.31	80.59±0.1	71.29±0.23	69.77±0.27	90.9±0.22	73.44
CLIP cross-model	77.8±0.06	84.46±0.06	87.64±0.08	84.68±0.1	87.19±0.23	88.59±0.09	81.77±0.11	81.85±0.1	89.5±0.09	84.61±0.09	79.31±0.15	83.66±0.1	84.26
CLIP distillation (Ours)	81.47±0.02	81.69±0.01	88.05±0.01	**87.23**±0.0	82.95±0.02	**88.81**±0.02	**85.3**±0.0	81.03±0.01	**92.45**±0.02	**86.82**±0.0	81.27±0.01	87.33±0.02	85.37
H^3-score													
SO	74.74±0.06	85.58±0.04	85.83±0.09	77.2±0.07	88.32±0.14	86.28±0.05	73.68±0.17	75.51±0.04	86.03±0.14	77.99±0.05	76.86±0.02	85.71±0.06	81.14
DANCE	77.71±0.02	89.91±0.02	86.95±0.06	76.67±0.06	90.04±0.12	87.22±0.06	72.09±0.04	75.31±0.09	85.82±0.11	78.13±0.02	77.87±0.03	91.4±0.17	82.43
OVANet	76.69±0.13	86.47±0.3	87.22±0.18	76.61±0.24	86.73±0.31	86.06±0.07	66.34±0.49	64.09±0.4	84.87±0.15	77.77±0.1	75.78±0.13	90.69±0.33	79.94
UniOT	**80.45**±0.38	**91.01**±0.54	**89.65**±0.24	**80.66**±0.8	**91.68**±0.64	**87.73**±0.51	78.7±0.74	**79.83**±0.34	87.94±0.34	**82.5**±0.48	**79.68**±0.48	**92.04**±0.43	**85.16**
WiSE-FT	67.32±0.03	84.82±0.11	84.45±0.02	63.48±0.22	85.39±0.18	81.39±0.18	62.84±0.28	59.62±0.24	82.05±0.07	70.7±0.15	70.96±0.19	90.86±0.15	75.32
CLIP cross-model	76.29±0.04	86.47±0.04	86.79±0.06	78.96±0.06	88.35±0.16	87.41±0.06	77.25±0.06	78.84±0.06	88.0±0.06	78.92±0.05	77.26±0.09	85.9±0.07	82.54
CLIP distillation (Ours)	78.61±0.01	84.51±0.01	87.05±0.01	80.42±0.0	85.4±0.01	87.55±0.01	**79.32**±0.0	78.34±0.0	**89.87**±0.02	80.19±0.0	78.49±0.0	88.45±0.01	83.18
UCR													
SO	73.39±0.06	89.19±0.07	92.23±0.02	87.93±0.05	92.11±0.15	90.87±0.18	72.67±0.26	74.93±0.13	89.21±0.07	81.48±0.04	76.85±0.16	93.48±0.07	84.53
DANCE	80.41±0.04	91.01±0.22	92.63±0.08	88.15±0.09	93.8±0.05	91.57±0.1	75.77±0.02	78.93±0.19	91.13±0.28	83.35±0.21	80.17±0.1	94.22±0.05	86.76
OVANet	76.65±0.18	92.89±0.06	93.55±0.15	87.56±0.1	93.41±0.2	90.89±0.26	71.32±0.23	77.95±0.06	90.35±0.07	81.78±0.22	78.15±0.1	94.09±0.34	85.72
UniOT	79.4±0.54	90.87±2.05	94.6±0.51	83.25±0.88	93.25±0.34	90.9±0.92	76.83±4.88	75.45±1.72	89.5±0.32	80.12±1.47	78.79±0.85	94.17±0.73	85.59
WiSE-FT	78.15±0.02	93.32±0.04	**95.79**±0.03	92.46±0.05	**95.31**±0.15	95.16±0.06	83.67±0.19	80.21±0.02	93.99±0.06	88.57±0.13	79.71±0.09	**95.0**±0.0	89.28
CLIP cross-model	75.95±0.1	91.07±0.13	93.99±0.06	90.54±0.18	93.73±0.1	93.36±0.09	83.61±0.07	79.94±0.09	93.65±0.08	85.53±0.08	78.29±0.04	94.15±0.03	87.82
CLIP zero-shot	80.42±0.0	93.53±0.0	93.71±0.0	91.03±0.0	93.53±0.0	93.71±0.0	91.03±0.0	80.42±0.0	93.71±0.0	91.03±0.0	80.42±0.0	93.53±0.0	89.67
CLIP distillation (Ours)	**81.47**±0.0	**93.84**±0.0	**96.13**±0.0	**93.17**±0.0	94.15±0.0	**96.32**±0.0	**93.56**±0.06	**82.15**±0.04	**96.81**±0.04	**93.39**±0.0	**82.08**±0.0	94.82±0.0	**91.49**

Table B13: OfficeHome: CLIP & (15/0) setting

Methods	P2R	P2S	R2P	R2S	S2P	S2R	Avg
H-score							
SO	73.95±0.01	70.14±0.02	68.96±0.11	69.83±0.05	66.51±0.03	72.74±0.02	70.36
DANCE	74.04±0.09	70.33±0.05	68.79±0.06	70.63±0.07	67.03±0.11	72.22±0.14	70.51
OVANet	54.47±0.44	67.84±0.01	67.47±0.02	66.35±0.17	54.74±0.2	34.07±0.72	57.49
UniOT	**77.1**±0.28	**71.91**±0.58	**70.0**±0.24	**72.78**±0.22	**68.73**±0.09	**75.3**±0.43	**72.64**
H^3-score							
SO	77.38±0.01	70.06±0.01	**70.57**±0.08	69.86±0.03	68.85±0.02	76.48±0.01	72.2
DANCE	77.44±0.07	**70.19**±0.03	70.45±0.04	70.39±0.04	**69.22**±0.07	76.1±0.1	**72.3**
OVANet	61.92±0.38	68.51±0.01	69.53±0.01	67.49±0.12	59.95±0.16	42.6±0.75	61.67
UniOT	**77.62**±0.21	69.74±0.31	69.75±0.15	**70.58**±0.04	68.8±0.05	**76.51**±0.3	72.17
UCR							
SO	70.72±0.02	59.82±0.13	61.89±0.08	60.51±0.15	60.66±0.12	70.4±0.02	64.0
DANCE	**71.43**±0.01	**62.13**±0.17	**63.72**±0.13	**63.89**±0.03	**63.24**±0.23	70.69±0.07	**65.85**
OVANet	68.15±0.11	55.82±0.17	59.94±0.16	56.97±0.11	58.49±0.14	68.39±0.15	61.29
UniOT	70.9±0.21	59.07±0.27	60.27±0.08	61.74±0.25	58.76±0.23	**70.9**±0.6	63.61

Table B14: DomainNet: DINOv2 & (150/0) setting

Methods	P2R	P2S	R2P	R2S	S2P	S2R	Avg
H-score							
SO	72.15±0.06	62.77±0.1	62.5±0.19	65.55±0.09	56.61±0.05	74.19±0.03	65.63
DANCE	72.2±0.07	62.16±0.2	60.81±0.06	65.86±0.16	56.69±0.01	73.75±0.02	65.25
OVANet	77.09±0.17	71.64±0.04	69.57±0.13	72.51±0.1	67.04±0.11	76.58±0.11	72.41
UniOT	76.78±0.08	70.35±0.31	**71.84**±0.05	**73.56**±0.15	**69.32**±0.36	77.43±0.19	73.21
WiSE-FT	11.76±0.11	4.52±0.09	8.84±0.11	7.73±0.07	3.77±0.08	10.9±0.13	7.92
CLIP cross-model	74.9±0.1	64.16±0.03	62.16±0.06	65.96±0.05	58.29±0.04	77.11±0.08	67.1
CLIP distillation (Ours)	**81.65**±0.01	**73.37**±0.04	68.39±0.02	73.08±0.01	69.15±0.02	**82.23**±0.01	**74.64**
H^3-score							
SO	75.38±0.04	63.2±0.07	64.39±0.14	65.05±0.06	60.09±0.04	76.86±0.02	67.5
DANCE	75.42±0.05	62.78±0.14	63.18±0.04	65.25±0.11	60.15±0.01	76.54±0.02	67.22
OVANet	78.9±0.12	68.92±0.03	**69.22**±0.09	69.46±0.06	67.53±0.07	78.55±0.08	72.1
UniOT	76.08±0.08	67.15±0.42	68.52±0.1	69.74±0.09	66.52±0.07	76.99±0.19	70.83
WiSE-FT	16.48±0.15	6.55±0.13	12.46±0.15	10.93±0.1	5.51±0.12	15.34±0.17	11.21
CLIP cross-model	77.36±0.07	64.13±0.02	64.15±0.04	65.32±0.03	61.35±0.03	78.92±0.05	68.54
CLIP distillation (Ours)	**82.03**±0.01	**69.98**±0.02	68.44±0.02	**69.81**±0.01	**68.95**±0.01	**82.42**±0.01	**73.61**
UCR							
SO	70.57±0.02	62.09±0.09	64.72±0.15	65.34±0.17	59.2±0.21	74.13±0.04	66.01
DANCE	72.14±0.06	63.86±0.13	65.17±0.13	67.27±0.02	66.4±0.05	75.36±0.03	68.37
OVANet	72.32±0.16	62.88±0.06	65.72±0.18	65.36±0.02	60.15±0.14	75.22±0.1	66.94
UniOT	73.72±0.17	61.89±0.29	64.05±0.18	65.77±0.19	62.91±0.19	75.66±0.22	67.33
WiSE-FT	77.55±0.03	67.24±0.08	68.76±0.1	68.97±0.16	67.45±0.11	79.97±0.02	71.66
CLIP cross-model	78.0±0.09	67.6±0.15	68.62±0.16	68.86±0.05	65.94±0.11	80.14±0.08	71.53
CLIP zero-shot	82.0±0.0	68.37±0.0	69.64±0.0	68.37±0.0	69.64±0.0	82.0±0.0	73.34
CLIP distillation (Ours)	**82.86**±0.02	**68.89**±0.01	**72.53**±0.03	**69.92**±0.02	**72.03**±0.02	**83.35**±0.02	**74.93**

Table B15: DomainNet: CLIP & (150/0) setting

Methods	A2D	A2W	D2A	D2W	W2A	W2D	Avg
H-score/H^3-score							
SO	85.49±0.06	83.35±0.18	**67.23**±0.53	**99.38**±0.1	**68.51**±0.33	**100.0**±0.0	**83.99**
DANCE	85.23±0.48	**83.49**±0.21	60.16±3.33	99.28±0.0	62.17±2.37	99.69±0.0	81.67
OVANet	**88.03**±0.22	83.27±0.34	61.73±1.68	96.0±0.08	64.01±0.49	100.0±0.0	82.17
UniOT	63.33±0.94	59.42±1.83	51.9±0.94	79.87±0.64	52.81±0.52	85.29±3.81	65.44
UCR							
SO	**91.97**±0.16	92.49±0.16	77.77±0.36	99.54±0.12	78.89±0.17	**100.0**±0.0	90.11
DANCE	89.49±0.53	90.69±0.31	70.1±2.32	**99.62**±0.0	72.06±1.39	99.87±0.09	86.97
OVANet	91.97±0.16	92.49±0.06	77.78±0.28	99.54±0.12	78.8±0.28	100.0±0.0	90.1
UniOT	91.63±1.23	**92.79**±0.57	**81.24**±0.48	98.7±0.24	**81.74**±0.26	100.0±0.0	**91.02**

Table B16: Office: DINOv2 & (31/0) setting

Methods	A2D	A2W	D2A	D2W	W2A	W2D	Avg
H-score/H^3-score							
SO	76.99±0.54	71.98±0.1	68.54±0.19	97.64±0.15	66.94±0.25	99.26±0.0	80.22
DANCE	73.29±1.1	63.14±0.16	62.46±1.66	**98.4**±0.0	57.63±3.83	**99.64**±0.32	75.76
OVANet	72.5±0.73	64.35±0.3	58.02±1.13	96.26±0.05	58.07±0.66	98.67±0.0	74.64
UniOT	58.03±2.06	56.8±1.14	53.8±0.89	69.88±1.5	53.23±0.88	67.96±3.44	59.95
WiSE-FT	31.89±0.31	25.87±0.19	39.04±0.14	74.62±0.26	36.65±0.29	79.13±0.09	47.87
CLIP cross-model	79.87±0.45	77.27±0.38	73.99±0.08	97.46±0.0	72.03±0.05	98.65±0.12	83.21
CLIP distillation (Ours)	**81.97**±0.0	**81.82**±0.0	**82.12**±0.03	85.66±0.0	**82.41**±0.03	86.07±0.0	**83.34**
UCR							
SO	93.37±0.28	**95.14**±0.26	81.03±0.08	99.66±0.06	79.65±0.09	99.8±0.0	91.44
DANCE	89.56±0.28	85.7±0.57	76.3±0.32	**99.75**±0.0	74.81±0.92	**100.0**±0.0	87.69
OVANet	93.31±0.25	94.93±0.16	81.04±0.09	99.66±0.06	79.64±0.15	99.8±0.0	91.4
UniOT	92.44±0.96	94.17±0.99	84.17±0.56	98.78±0.62	84.32±0.3	98.93±0.34	92.14
WiSE-FT	93.24±0.19	92.91±0.42	84.76±0.07	98.99±0.0	84.2±0.1	99.67±0.09	92.3
CLIP cross-model	**94.98**±0.0	94.8±0.16	85.56±0.04	99.75±0.0	84.52±0.09	99.8±0.0	**93.23**
CLIP zero-shot	88.15±0.0	89.18±0.0	85.73±0.0	89.18±0.0	85.73±0.0	88.15±0.0	87.69
CLIP distillation (Ours)	88.15±0.0	89.31±0.0	**86.0**±0.03	89.43±0.0	**86.01**±0.05	88.35±0.0	87.88

Table B17: Office: CLIP & (31/0) setting

Methods	A2C	A2P	A2R	C2A	C2P	C2R	P2A	P2C	P2R	R2A	R2C	R2P	Avg
H-score/H^3-score													
SO	**59.36**±0.1	74.0±0.07	77.72±0.03	56.65±0.05	71.85±0.04	68.49±0.09	52.6±0.16	**51.14**±0.18	73.53±0.05	69.0±0.06	**59.9**±0.14	86.39±0.03	66.72
DANCE	56.97±0.31	67.65±0.62	78.01±0.1	54.79±0.41	69.48±1.04	67.41±0.2	46.53±0.36	44.19±0.81	71.55±0.41	68.64±0.32	59.15±0.17	**86.97**±0.29	64.28
OVANet	54.35±0.18	**78.3**±0.09	**83.43**±0.13	**74.66**±0.19	**82.64**±0.08	**84.98**±0.21	**61.55**±0.26	39.19±0.14	**80.72**±0.14	**70.67**±0.37	51.02±0.22	85.63±0.3	**70.59**
UniOT	50.15±0.13	55.39±0.71	63.4±0.4	52.6±0.6	57.93±0.62	59.99±0.61	45.47±0.49	43.98±0.22	59.39±0.21	58.06±1.11	51.69±0.32	71.62±0.03	55.81
UCR													
SO	72.02±0.03	**86.92**±0.04	88.95±0.04	**82.24**±0.19	86.94±0.07	87.58±0.11	**76.03**±0.15	67.13±0.22	86.8±0.08	84.05±0.09	71.87±0.19	92.3±0.03	81.9
DANCE	72.06±0.41	81.96±0.27	88.46±0.18	78.73±0.16	83.08±0.05	85.43±0.11	68.82±0.36	65.96±0.07	84.9±0.21	83.6±0.07	75.91±0.21	92.27±0.29	80.1
OVANet	71.94±0.17	86.84±0.05	89.0±0.01	82.04±0.13	86.99±0.11	87.54±0.14	75.96±0.26	67.03±0.29	86.73±0.06	84.16±0.11	71.72±0.09	92.36±0.07	81.86
UniOT	**75.71**±0.61	85.66±0.28	**89.66**±0.35	81.13±0.59	**87.53**±0.35	**88.73**±0.14	72.76±0.72	**70.88**±0.58	**87.83**±0.31	**84.45**±0.61	**77.89**±0.5	**92.98**±0.56	**82.93**

Table B18: OfficeHome: DINOv2 & (65/0) setting

Methods	A2C	A2P	A2R	C2A	C2P	C2R	P2A	P2C	P2R	R2A	R2C	R2P	Avg
H-score/H^3-score													
SO	42.18±0.02	65.78±0.1	67.64±0.01	46.06±0.18	68.16±0.16	64.49±0.08	42.68±0.12	37.5±0.07	71.0±0.07	58.19±0.19	48.05±0.09	85.69±0.05	58.12
DANCE	39.45±0.12	59.51±1.16	68.27±0.05	45.45±0.17	58.08±0.31	63.0±0.1	36.2±0.56	35.31±0.1	68.49±0.29	58.1±0.47	47.07±0.06	86.12±0.02	55.42
OVANet	53.14±0.16	67.63±0.26	77.34±0.07	61.55±0.37	67.28±0.31	73.51±0.21	50.26±0.2	40.58±0.43	75.23±0.19	68.72±0.14	53.28±0.23	87.32±0.2	64.65
UniOT	52.4±0.23	54.54±0.9	64.91±0.76	61.16±0.31	66.6±1.57	72.71±0.73	47.04±0.66	47.5±0.23	59.05±0.61	60.14±0.58	54.32±0.22	70.81±1.06	59.27
WiSE-FT	5.98±0.11	14.86±0.05	17.47±0.04	3.65±0.12	9.79±0.11	9.0±0.13	9.35±0.07	5.21±0.08	26.2±0.02	14.31±0.11	8.7±0.18	39.13±0.07	13.64
CLIP cross-model	44.97±0.19	73.26±0.06	72.75±0.08	52.76±0.25	77.72±0.06	71.95±0.09	47.85±0.1	41.82±0.04	73.98±0.09	58.98±0.03	49.1±0.09	86.71±0.03	62.65
CLIP distillation (Ours)	**64.4**±0.01	**92.28**±0.03	**90.14**±0.02	**75.48**±0.05	**92.17**±0.02	**89.63**±0.04	**67.03**±0.04	**55.98**±0.01	**84.68**±0.03	**71.18**±0.05	**59.4**±0.01	**90.77**±0.01	**77.76**
UCR													
SO	72.07±0.11	87.95±0.04	90.96±0.04	82.19±0.19	88.87±0.06	89.79±0.08	77.21±0.15	69.28±0.27	89.85±0.09	85.55±0.02	75.01±0.05	93.86±0.07	83.55
DANCE	71.78±0.19	82.62±0.23	90.46±0.01	81.33±0.17	86.07±0.11	88.24±0.15	72.35±0.34	69.52±0.17	87.54±0.08	84.74±0.1	74.08±0.08	92.16±0.08	81.74
OVANet	72.07±0.12	87.99±0.08	90.98±0.06	82.01±0.14	88.78±0.14	89.79±0.05	77.27±0.07	69.14±0.09	89.75±0.08	85.51±0.16	75.22±0.3	93.77±0.2	83.52
UniOT	75.48±0.86	91.03±0.24	92.17±0.23	84.97±0.91	90.74±0.41	90.18±0.09	79.06±1.0	73.81±0.14	90.98±0.13	85.87±0.29	77.72±0.3	93.97±0.39	85.5
WiSE-FT	76.3±0.08	92.93±0.03	93.61±0.03	87.76±0.09	92.36±0.07	93.02±0.02	82.98±0.12	74.78±0.08	92.4±0.04	88.74±0.19	77.6±0.08	94.86±0.02	87.28
CLIP cross-model	75.67±0.08	93.52±0.07	93.33±0.01	86.24±0.09	92.14±0.02	92.33±0.1	82.75±0.07	74.43±0.12	92.39±0.05	88.31±0.13	77.15±0.12	**95.17**±0.01	86.95
CLIP zero-shot	77.69±0.0	94.32±0.0	94.51±0.0	89.82±0.0	94.32±0.0	94.51±0.0	89.82±0.0	77.69±0.0	94.51±0.0	89.82±0.0	77.69±0.0	94.32±0.0	89.08
CLIP distillation (Ours)	**79.06**±0.0	**94.74**±0.03	**95.0**±0.0	**90.48**±0.0	**94.72**±0.01	**94.92**±0.05	**90.19**±0.0	**78.78**±0.03	**94.84**±0.0	**90.44**±0.0	**78.81**±0.03	94.55±0.0	**89.71**

Table B19: OfficeHome: CLIP & (65/0) setting

Methods	P2R	P2S	R2P	R2S	S2P	S2R	Avg
H-score/H^3-score							
SO	56.91±0.08	48.36±0.05	50.97±0.07	50.08±0.08	49.25±0.07	54.4±0.07	51.66
DANCE	56.7±0.08	48.93±0.18	50.6±0.05	50.78±0.2	49.03±0.08	53.51±0.31	51.59
OVANet	**71.99**±0.05	48.15±0.19	**52.21**±0.1	45.3±0.12	**60.43**±0.02	**75.43**±0.03	**58.92**
UniOT	59.05±0.17	**52.93**±0.09	50.64±0.42	**53.44**±0.14	50.33±0.28	57.59±0.22	54.0
UCR							
SO	76.11±0.13	62.71±0.04	67.28±0.08	62.65±0.13	66.89±0.03	78.02±0.06	68.94
DANCE	75.62±0.12	**64.66**±0.09	**68.6**±0.12	64.89±0.12	**67.52**±0.04	77.09±0.06	**69.73**
OVANet	**76.16**±0.02	62.71±0.07	67.28±0.07	62.75±0.03	66.9±0.03	**78.04**±0.05	68.97
UniOT	75.8±0.09	64.49±0.12	66.91±0.1	**65.24**±0.08	65.56±0.06	76.46±0.06	69.08

Table B20: DomainNet: DINOv2 & (345/0) setting

Methods	P2R	P2S	R2P	R2S	S2P	S2R	Avg
H-score/H^3-score							
SO	47.32±0.07	31.92±0.03	33.24±0.06	35.42±0.06	30.24±0.04	51.46±0.04	38.27
DANCE	47.09±0.03	32.1±0.11	31.62±0.24	33.88±0.14	29.89±0.11	50.78±0.04	37.56
OVANet	65.92±0.1	51.21±0.11	49.47±0.07	54.56±0.05	51.87±0.19	70.27±0.09	57.22
UniOT	71.17±0.32	**60.36**±0.17	**57.09**±0.35	**63.79**±0.17	**56.98**±0.15	72.93±0.07	**63.72**
WiSE-FT	0.56±0.01	0.09±0.0	0.47±0.03	0.25±0.01	0.11±0.01	0.3±0.01	0.3
CLIP cross-model	46.33±0.11	29.96±0.09	29.69±0.12	32.42±0.1	28.35±0.06	50.46±0.06	36.2
CLIP distillation (Ours)	**76.53**±0.0	55.58±0.02	48.36±0.02	55.01±0.02	49.0±0.03	**76.48**±0.01	60.16
UCR							
SO	76.06±0.08	64.06±0.23	69.55±0.18	68.6±0.04	67.67±0.13	81.64±0.08	71.26
DANCE	75.6±0.11	66.47±0.13	70.01±0.03	68.07±0.03	69.35±0.12	80.44±0.06	71.66
OVANet	76.15±0.06	64.36±0.23	69.65±0.09	68.78±0.05	67.66±0.14	81.63±0.05	71.37
UniOT	79.76±0.04	68.2±0.13	71.57±0.2	71.12±0.1	69.99±0.04	82.45±0.08	73.85
WiSE-FT	82.2±0.08	69.95±0.11	72.67±0.08	71.79±0.02	72.33±0.11	85.51±0.02	75.74
CLIP cross-model	82.4±0.01	69.93±0.16	72.48±0.1	71.33±0.02	71.61±0.06	85.66±0.04	75.57
CLIP zero-shot	88.5±0.0	74.54±0.0	75.27±0.0	74.65±0.0	75.27±0.0	88.68±0.0	79.48
CLIP distillation (Ours)	**89.06**±0.01	**75.39**±0.01	**76.4**±0.0	**75.48**±0.01	**76.39**±0.01	**89.24**±0.01	**80.33**

Table B21: DomainNet: CLIP & (345/0) setting

Methods	A2D	A2W	D2A	D2W	W2A	W2D	Avg
H-score/H^3-score							
SO	91.71±0.21	88.29±0.46	80.88±0.76	99.53±0.33	82.58±0.19	**100.0**±0.0	90.5
DANCE	70.75±4.51	79.7±3.77	65.66±0.08	**100.0**±0.0	65.83±0.26	100.0±0.0	80.32
OVANet	**95.5**±0.0	**93.82**±0.37	**82.73**±0.15	97.49±0.14	**82.6**±0.21	100.0±0.0	**92.02**
UniOT	37.2±4.35	36.71±1.82	40.67±0.63	43.64±1.61	39.16±1.08	48.85±3.19	41.04
UCR							
SO	**97.03**±0.3	93.56±0.28	86.95±0.3	99.44±0.32	**90.54**±0.34	**100.0**±0.0	94.59
DANCE	79.83±2.35	87.34±0.7	71.75±0.44	**100.0**±0.0	72.3±0.47	100.0±0.0	85.2
OVANet	97.03±0.3	**93.79**±0.32	**87.09**±0.38	99.44±0.32	90.5±0.31	100.0±0.0	**94.64**
UniOT	49.04±2.9	50.85±1.73	61.93±0.34	66.1±1.27	60.65±0.23	70.49±2.62	59.84

Table B22: Office: DINOv2 & (10/21) setting

Methods	A2D	A2W	D2A	D2W	W2A	W2D	Avg
H-score/H^3-score							
SO	86.4±0.5	83.72±0.24	89.13±0.52	96.43±0.22	84.28±0.55	**98.75**±0.0	89.79
DANCE	65.23±6.21	43.33±4.71	57.13±8.0	90.0±0.0	47.77±7.84	97.5±0.0	66.83
OVANet	91.42±0.2	77.48±2.74	81.96±4.4	95.79±0.11	81.63±4.31	96.92±0.0	87.53
UniOT	43.36±1.88	36.25±2.27	39.29±1.02	42.39±0.83	37.23±2.93	49.34±1.09	41.31
WiSE-FT	39.89±0.39	28.34±0.15	50.65±0.36	76.36±0.18	43.01±0.44	83.15±0.39	53.57
CLIP cross-model	89.62±0.2	87.18±0.18	93.65±0.13	**96.64**±0.0	91.03±0.13	97.19±0.2	92.55
CLIP distillation (Ours)	**92.87**±0.0	**91.56**±0.0	**95.81**±0.0	94.81±0.0	**95.81**±0.0	95.05±0.0	**94.32**
UCR							
SO	95.12±0.6	97.63±0.28	95.27±0.05	99.44±0.16	94.5±0.13	**100.0**±0.0	96.99
DANCE	76.01±5.34	71.41±1.31	76.55±2.69	91.53±0.0	71.71±3.41	99.36±0.0	81.09
OVANet	94.69±0.6	97.18±0.16	95.27±0.05	99.44±0.16	94.43±0.18	100.0±0.0	96.83
UniOT	47.56±2.1	48.14±1.73	61.97±0.63	63.73±2.0	58.21±1.8	56.05±0.52	55.94
WiSE-FT	**96.18**±0.0	**97.74**±0.16	96.31±0.05	**99.66**±0.0	95.93±0.09	99.58±0.3	97.57
CLIP cross-model	96.18±0.0	97.63±0.28	96.35±0.0	99.66±0.0	95.65±0.05	100.0±0.0	**97.58**
CLIP zero-shot	95.54±0.0	97.63±0.0	**96.66**±0.0	97.63±0.0	**96.66**±0.0	95.54±0.0	96.61
CLIP distillation (Ours)	95.54±0.0	97.63±0.0	96.66±0.0	97.63±0.0	96.66±0.0	95.54±0.0	96.61

Table B23: Office: CLIP & (10/21) setting

Methods	A2C	A2P	A2R	C2A	C2P	C2R	P2A	P2C	P2R	R2A	R2C	R2P	Avg
H-score/H^3-score													
SO	**61.02**±0.08	74.29±0.1	79.21±0.1	50.62±0.23	61.51±0.16	62.24±0.14	49.47±0.18	**50.7**±0.39	70.06±0.08	69.75±0.23	**61.87**±0.37	82.23±0.14	64.41
DANCE	56.34±1.01	57.05±1.2	79.52±1.6	47.43±3.36	50.13±1.26	57.75±0.51	35.09±0.43	34.65±0.42	63.59±1.81	66.5±0.6	56.41±0.11	79.9±1.89	57.03
OVANet	58.72±0.04	**81.24**±0.61	**87.33**±0.36	**72.12**±0.39	**77.08**±0.33	**81.4**±0.33	**63.93**±0.24	41.44±0.48	**80.21**±0.48	**74.46**±0.25	54.17±0.31	**83.33**±0.22	**71.29**
UniOT	43.53±0.88	40.65±0.59	39.34±1.4	34.31±2.27	37.81±1.26	33.29±0.73	32.87±0.75	34.05±0.74	37.52±0.5	44.56±1.09	42.73±0.26	41.94±1.3	38.55
UCR													
SO	75.88±0.05	**88.85**±0.16	91.2±0.03	**80.44**±0.49	80.04±0.07	**84.58**±0.18	**76.55**±0.3	67.72±0.41	**85.86**±0.09	85.49±0.15	**74.43**±0.1	91.07±0.21	**81.84**
DANCE	73.25±0.7	78.3±1.02	90.78±0.73	72.3±1.6	72.32±1.08	78.67±0.46	58.95±0.4	56.78±0.56	80.58±0.79	82.58±0.37	73.97±0.05	90.68±0.26	75.76
OVANet	**76.06**±0.13	88.85±0.08	**91.22**±0.09	80.1±0.04	**80.24**±0.14	84.28±0.27	76.55±0.38	**67.8**±0.42	85.83±0.13	**85.58**±0.07	74.39±0.17	**91.2**±0.16	81.84
UniOT	53.89±0.47	58.39±0.71	74.6±0.59	46.71±0.95	55.11±0.29	61.31±2.03	40.96±0.86	44.14±1.4	66.24±0.36	65.2±0.83	55.56±0.43	71.3±0.99	57.78

Table B24: OfficeHome: DINOv2 & (25/40) setting

Methods	A2C	A2P	A2R	C2A	C2P	C2R	P2A	P2C	P2R	R2A	R2C	R2P	Avg
H-score/H^3-score													
SO	44.33±0.19	65.34±0.35	71.4±0.35	44.78±0.1	59.45±0.47	63.91±0.17	46.61±0.26	41.37±0.14	72.02±0.12	59.08±0.17	53.28±0.1	78.2±0.12	58.31
DANCE	34.87±0.17	41.08±1.78	74.03±1.62	31.93±1.16	34.31±0.22	54.84±2.92	25.85±1.81	21.6±0.81	63.45±2.65	57.55±0.35	45.7±0.79	74.36±0.06	46.63
OVANet	60.09±0.43	70.91±0.3	81.93±0.39	60.59±0.48	61.88±1.41	72.19±0.26	52.46±0.48	45.8±1.0	76.8±0.16	68.38±0.23	58.47±0.2	81.48±0.52	65.92
UniOT	44.17±0.19	43.91±0.93	39.98±1.07	47.21±0.58	41.66±2.14	47.94±3.9	37.16±1.5	44.8±0.48	39.38±1.05	48.13±0.6	43.92±0.44	44.92±2.23	43.6
WiSE-FT	9.8±0.04	17.69±0.16	22.75±0.08	6.1±0.04	11.27±0.08	13.4±0.28	11.26±0.09	7.26±0.14	29.79±0.09	17.6±0.12	13.3±0.28	38.55±0.19	16.56
CLIP cross-model	48.18±0.04	70.76±0.25	75.41±0.24	51.32±0.17	70.49±0.36	70.68±0.13	52.32±0.19	48.38±0.12	75.23±0.09	59.69±0.2	55.33±0.29	79.92±0.18	63.14
CLIP distillation (Ours)	**70.86**±0.04	**89.85**±0.0	**92.26**±0.0	**77.75**±0.05	**89.45**±0.02	**91.75**±0.03	**69.17**±0.03	**61.13**±0.02	**86.02**±0.03	**72.81**±0.0	**64.91**±0.02	**88.26**±0.03	**79.52**
UCR													
SO	79.38±0.32	87.73±0.24	91.94±0.14	81.14±0.38	81.94±0.21	87.7±0.33	78.6±0.3	76.66±0.22	91.33±0.09	84.82±0.16	81.93±0.16	90.05±0.1	84.44
DANCE	69.39±0.34	75.16±1.63	90.8±0.4	72.39±0.6	70.55±0.52	82.33±0.83	62.63±0.81	61.95±0.7	83.45±0.18	80.38±0.11	74.07±0.35	84.48±0.09	75.63
OVANet	79.5±0.12	87.45±0.12	91.81±0.09	80.96±0.3	81.89±0.03	87.61±0.29	78.51±0.0	76.42±0.15	91.22±0.08	84.73±0.17	82.41±0.2	89.9±0.44	84.37
UniOT	57.03±1.24	58.43±0.76	75.43±1.72	64.19±3.19	68.33±2.08	79.72±2.56	44.14±0.74	55.84±1.23	69.87±1.01	63.7±0.87	59.86±0.61	66.76±0.11	63.61
WiSE-FT	**82.93**±0.21	91.69±0.19	93.82±0.12	88.03±0.23	87.58±0.23	91.9±0.13	84.21±0.2	82.07±0.16	93.34±0.03	88.77±0.35	**84.82**±0.14	92.12±0.03	88.44
CLIP cross-model	82.37±0.15	**92.57**±0.12	93.12±0.16	85.8±0.09	86.87±0.1	90.34±0.05	84.27±0.04	81.83±0.32	93.58±0.03	87.82±0.09	84.3±0.05	**92.75**±0.12	87.97
CLIP zero-shot	81.61±0.0	91.2±0.0	94.53±0.0	90.36±0.0	91.2±0.0	94.53±0.0	90.36±0.0	81.61±0.0	94.53±0.0	90.36±0.0	81.61±0.0	91.2±0.0	89.43
CLIP distillation (Ours)	82.73±0.03	91.6±0.0	**94.92**±0.0	**90.97**±0.04	**91.6**±0.05	**94.92**±0.0	**90.54**±0.0	**82.27**±0.0	**94.81**±0.0	**90.63**±0.0	82.33±0.0	91.58±0.03	**89.91**

Table B25: OfficeHome: CLIP & (25/40) setting

Methods	P2R	P2S	R2P	R2S	S2P	S2R	Avg
H-score/H^3-score							
SO	54.81±0.08	49.95±0.11	50.3±0.28	50.85±0.13	45.97±0.05	51.3±0.02	50.53
DANCE	53.56±0.24	**51.09**±0.37	49.55±0.48	**50.97**±0.82	45.88±0.27	44.77±0.81	49.3
OVANet	**70.74**±0.07	49.81±0.17	**52.21**±0.11	45.79±0.16	**59.03**±0.03	**73.48**±0.07	**58.51**
UniOT	46.39±0.55	47.44±0.16	44.6±0.28	48.12±0.55	40.24±0.3	43.68±0.36	45.08
UCR							
SO	73.96±0.08	63.8±0.13	66.12±0.22	65.89±0.35	67.25±0.02	**75.39**±0.05	68.73
DANCE	73.47±0.13	**66.02**±0.1	**66.62**±0.5	**66.0**±0.82	**67.49**±0.27	72.03±0.19	68.61
OVANet	**74.01**±0.05	63.81±0.08	66.26±0.05	65.84±0.07	67.14±0.03	75.38±0.03	**68.74**
UniOT	66.45±0.17	59.91±0.09	57.66±0.25	60.4±0.34	54.19±0.51	63.7±0.06	60.38

Table B26: DomainNet: DINOv2 & (150/195) setting

Methods	P2R	P2S	R2P	R2S	S2P	S2R	Avg
H-score/H^3-score							
SO	44.52±0.09	31.86±0.16	32.39±0.02	34.58±0.12	25.71±0.11	46.24±0.26	35.88
DANCE	42.13±0.74	28.66±0.23	25.54±0.33	27.51±0.41	21.75±0.42	39.93±0.71	30.92
OVANet	65.01±0.07	51.31±0.26	48.81±0.22	54.76±0.22	47.84±0.25	67.4±0.19	55.85
UniOT	60.68±0.59	54.95±0.29	50.63±0.31	56.8±0.05	47.35±0.97	60.69±0.56	55.18
WiSE-FT	0.57±0.01	0.06±0.01	0.52±0.04	0.32±0.02	0.07±0.01	0.2±0.01	0.29
CLIP cross-model	43.44±0.12	29.73±0.18	28.59±0.05	31.68±0.23	24.92±0.05	46.01±0.03	34.06
CLIP distillation (Ours)	**79.95**±0.01	**63.34**±0.02	**54.64**±0.01	**62.03**±0.01	**55.87**±0.05	**79.52**±0.01	**65.89**
UCR							
SO	74.92±0.05	63.6±0.4	68.76±0.24	70.77±0.11	67.52±0.11	79.14±0.15	70.78
DANCE	72.51±0.15	63.94±0.37	64.22±0.24	64.88±0.31	64.98±0.06	75.05±0.17	67.6
OVANet	74.9±0.05	64.19±0.25	68.84±0.13	71.06±0.07	67.52±0.27	79.11±0.08	70.94
UniOT	75.12±0.17	64.5±0.34	64.31±0.24	67.24±0.32	61.14±0.89	75.63±0.27	67.99
WiSE-FT	81.23±0.08	70.46±0.23	72.08±0.11	74.16±0.09	72.95±0.09	83.74±0.02	75.77
CLIP cross-model	81.34±0.08	70.44±0.21	72.06±0.14	73.51±0.06	71.97±0.19	84.0±0.11	75.55
CLIP zero-shot	87.99±0.0	76.44±0.0	75.2±0.0	76.41±0.0	75.2±0.0	87.98±0.0	79.87
CLIP distillation (Ours)	**89.82**±0.0	**78.62**±0.01	**77.68**±0.01	**78.48**±0.01	**77.76**±0.01	**89.8**±0.01	**82.03**

Table B27: DomainNet: CLIP & (150/195) setting

References

[1] Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*, 2021.

[2] Wanxing Chang, Ye Shi, Hoang Tuan, and Jingya Wang. Unified optimal transport framework for universal domain adaptation. In *Advances in Neural Information Processing Systems*, 2022.

[3] Liang Chen, Qianjin Du, Yihang Lou, Jianzhong He, Tao Bai, and Minghua Deng. Mutual nearest neighbor contrast and hybrid prototype self-training for universal domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022.

[4] Morris H DeGroot and Stephen E Fienberg. The comparison and evaluation of forecasters. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 32(1-2):12–22, 1983.

[5] Akshay Raj Dhamija, Manuel Günther, and Terrance Boult. Reducing network agnostophobia. In *Advances in Neural Information Processing Systems*, 2018.

[6] Sepideh Esmaeilpour, Bing Liu, Eric Robertson, and Lei Shu. Zero-shot out-of-distribution detection based on the pre-trained model clip. In *Proceedings of the AAAI conference on artificial intelligence*, 2022.

[7] Zhen Fang, Yixuan Li, Jie Lu, Jiahua Dong, Bo Han, and Feng Liu. Is out-of-distribution detection learnable? In *Advances in Neural Information Processing Systems*, 2022.

[8] Bo Fu, Zhangjie Cao, Mingsheng Long, and Jianmin Wang. Learning to detect open classes for universal domain adaptation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XV 16*, pages 567–583. Springer, 2020.

[9] 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.

[10] Peng Gao, Shijie Geng, Renrui Zhang, Teli Ma, Rongyao Fang, Yongfeng Zhang, Hongsheng Li, and Yu Qiao. Clip-adapter: Better vision-language models with feature adapters. *arXiv preprint arXiv:2110.04544*, 2021.

[11] Sachin Goyal, Ananya Kumar, Sankalp Garg, Zico Kolter, and Aditi Raghunathan. Finetune like you pretrain: Improved finetuning of zero-shot vision models, 2022.

[12] Jindong Gu, Ahmad Beirami, Xuezhi Wang, Alex Beutel, Philip Torr, and Yao Qin. Towards robust prompts on vision-language models. *arXiv preprint arXiv:2304.08479*, 2023.

[13] Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q Weinberger. On calibration of modern neural networks. In *International conference on machine learning*, pages 1321–1330. PMLR, 2017.

[14] Trevor Hastie, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009.

[15] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16000–16009, 2022.

[16] Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. In *International Conference on Learning Representations*, 2017.

[17] Geoffrey Hinton, Oriol Vinyals, and Jeffrey Dean. Distilling the knowledge in a neural network. In *NIPS Deep Learning and Representation Learning Workshop*, 2015.

[18] Chao Jia, Yinfei Yang, Ye Xia, Yi-Ting Chen, Zarana Parekh, Hieu Pham, Quoc Le, Yun-Hsuan Sung, Zhen Li, and Tom Duerig. Scaling up visual and vision-language representation learning with noisy text supervision. In *International Conference on Machine Learning*, pages 4904–4916. PMLR, 2021.

[19] Yoonho Lee, Annie S Chen, Fahim Tajwar, Ananya Kumar, Huaxiu Yao, Percy Liang, and Chelsea Finn. Surgical fine-tuning improves adaptation to distribution shifts. In *The Eleventh International Conference on Learning Representations*, 2023.

[20] Guangrui Li, Guoliang Kang, Yi Zhu, Yunchao Wei, and Yi Yang. Domain consensus clustering for universal domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9757–9766, 2021.

[21] Zhiqiu Lin, Samuel Yu, Zhiyi Kuang, Deepak Pathak, and Deva Ramanan. Multimodality helps unimodality: Cross-modal few-shot learning with multimodal models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19325–19337, 2023.

[22] Mahdi Pakdaman Naeini, Gregory Cooper, and Milos Hauskrecht. Obtaining well calibrated probabilities using bayesian binning. In *Proceedings of the AAAI conference on artificial intelligence*, 2015.

[23] Behnam Neyshabur, Hanie Sedghi, and Chiyuan Zhang. What is being transferred in transfer learning? In *Advances in neural information processing systems*, 2020.

[24] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, Mahmoud Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Hervé Jegou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. Dinov2: Learning robust visual features without supervision, 2023.

[25] Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment matching for multi-source domain adaptation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1406–1415, 2019.

[26] 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.

[27] P. Jonathon Phillips, Patrick Grother, and Ross Micheals. Evaluation methods in face recognition. In *Handbook of Face Recognition*, pages 551–574. Springer London, London, 2011.

[28] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021.

[29] Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell. Adapting visual category models to new domains. In *European Conference on Computer Vision*, pages 213–226. Springer, 2010.

[30] Kuniaki Saito, Donghyun Kim, Stan Sclaroff, and Kate Saenko. Universal domain adaptation through self supervision. In *Advances in neural information processing systems*, 2020.

[31] Kuniaki Saito and Kate Saenko. Ovanet: One-vs-all network for universal domain adaptation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9000–9009, 2021.

[32] Walter J. Scheirer, Anderson de Rezende Rocha, Archana Sapkota, and Terrance E. Boult. Toward open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(7):1757–1772, 2013.

[33] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, 2017.

[34] Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Open-set recognition: A good closed-set classifier is all you need. In *International Conference on Learning Representations*, 2022.

[35] 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.

[36] Colin Wei, Kendrick Shen, Yining Chen, and Tengyu Ma. Theoretical analysis of self-training with deep networks on unlabeled data. In *International Conference on Learning Representations*, 2021.

[37] Mitchell Wortsman, Gabriel Ilharco, Jong Wook Kim, Mike Li, Simon Kornblith, Rebecca Roelofs, Raphael Gontijo Lopes, Hannaneh Hajishirzi, Ali Farhadi, Hongseok Namkoong, et al. Robust fine-tuning of zero-shot models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7959–7971, 2022.

[38] Kaichao You, Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. Universal domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2720–2729, 2019.

[39] Renrui Zhang, Wei Zhang, Rongyao Fang, Peng Gao, Kunchang Li, Jifeng Dai, Yu Qiao, and Hongsheng Li. Tip-adapter: Training-free adaption of clip for few-shot classification. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXV*, pages 493–510. Springer, 2022.

[40] Kaiyang Zhou, Jingkang Yang, Chen Change Loy, and Ziwei Liu. Conditional prompt learning for vision-language models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16816–16825, 2022.

[41] Kaiyang Zhou, Jingkang Yang, Chen Change Loy, and Ziwei Liu. Learning to prompt for vision-language models. *International Journal of Computer Vision*, 130(9):2337–2348, 2022.