Uncertain estimation based on prediction probability and feature prototypes for universal domain adaptation

Junsong Leng, Zhong Chen*, Guoyou Wang
State Key Laboratory of Multispectral Information Intelligent Processing Technology
School of Artificial Intelligence and Automation, Huazhong University of Science and Technology
Wuhan, China
ljs@hust.edu.cn, henpacked@hust.edu.cn, gywang@hust.edu.cn

***Abstract*—Universal domain adaptation (UniDA) aims to transfer knowledge from the source to the target domain without restrictions on label sets. The key challenge is to maintain robust classification for known classes while identifying unknown classes in the target domain. We propose a model based on uncertainty estimation to address this issue. It integrates a teacher–student network, domain adversarial training, and uncertainty estimation module. The student network provides pseudo-labels for the teacher network, with parameters updated via exponential moving average. Source features initialize class prototypes, which are updated during training. Finally, uncertainty estimation determines whether target samples belong to known classes. Experiments on multiple datasets show that our model outperforms existing baselines.**

***Index Terms*—Universal domain adaptation, uncertainty estimation, teacher–student network, domain adversarial training**

I. Introduction

Domain Adaptation (DA) aims to generalize knowledge learned from one domain to another, where the former is referred to as the source domain and the latter as the target domain. It has been widely applied to tasks such as image classification [1]–[6], object detection [7]–[13], and image segmentation [14]–[20]. In practical scenarios, however, labels in the target domain are usually unavailable, which has motivated extensive research on unsupervised domain adaptation (UDA).

Most existing DA studies assume that the source and target domains share an identical label space. Nevertheless, in real-world settings, the label set of the target domain is often unknown, and the target domain may contain classes that do not appear in the source domain. This implies that, in addition to shared classes, both the source and target domains may also include private classes. Such a setting is more realistic and encourages research on domain adaptation without prior knowledge of the target label space, namely universal domain adaptation (UniDA).

You [21] provided a clear definition of UniDA in their UDA work. UDA achieves domain alignment through domain adversarial training and identifies whether an image belongs to an unknown class using uncertainty scores based on domain similarity and entropy. DANCE [22] and DCC [23] argue that some previous methods neglect exploration of the intrinsic structure of the target domain and propose to distinguish known and unknown classes via self-supervised clustering. Similarly, Fu [24] leverage the complementary effects of entropy, consistency, and confidence to more clearly differentiate different levels of uncertainty, and demonstrate that classifier correction based on multiple classifiers can better distinguish known and unknown target samples.

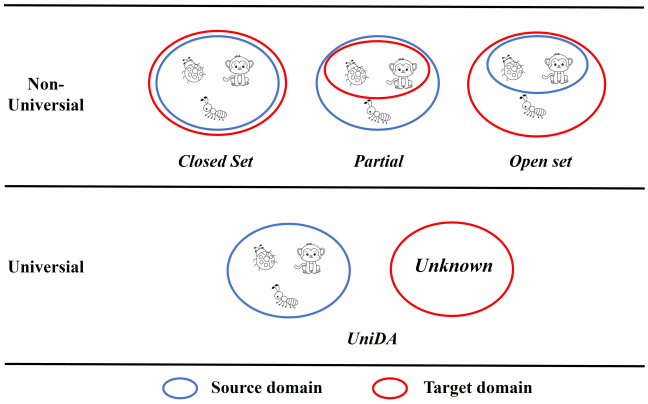

Fig. 1. Different DA categories

We believe that these methods, which exploit the internal structure of target-domain data and utilize uncertainty-based strategies to identify unknown classes, are both reasonable and effective. Inspired by these approaches, we adopt a pseudo-labeling strategy to provide supervision for target-domain data, encouraging the model to focus on target-domain feature characteristics. Furthermore, we employ an uncertainty estimation method based on prediction probability and feature prototypes to evaluate whether target-domain samples belong to known classes.

As illustrated in Fig. 1, DA tasks can be categorized into *non-universal* and *universal* settings. The non-universal DA setting includes closed-set [1], partial [25], [26], and open-set [27], [28] scenarios. We focus on UniDA to better meet practical needs.

Zhong Chen is corresponding author, this work is supported by the Civil Aerospace Technology Pre-research Project of China's 14th Five-Year Plan (Project No. D040404).

Based on the above considerations and practical settings, we design our model by taking into account the following task characteristics and requirements: (1) labels in the target domain are unavailable; (2) the label sets of the source and target domains are not necessarily identical; and (3) it is required to identify private label data in the target domain.

Our contributions are summarized as follows:

(1) To address the absence of target-domain labels in universal domain adaptation, we design a teacher–student network architecture in which the student network generates pseudo-labels for target-domain images, thereby providing additional supervisory signals. This enables the model to learn useful information from target-domain images.

(2) For target-domain image classification, we propose an uncertainty estimation method based on the predicted class probability and feature–prototype similarity to determine whether a target-domain image belongs to a known or an unknown class. Target-domain samples that satisfy the criterion are classified according to their predicted probabilities, while the remaining samples are assigned to the unknown class.

(3) We validate the effectiveness of the proposed method on multiple datasets, where our model consistently achieves the best performance compared with baseline approaches.

Once the paper is finalized, we will release all the code.

II. Related work

A. Different Domain Adaptation

Domain adaptation methods can be broadly categorized according to the assumptions imposed on the relationship between the source and target label spaces. Early studies mainly focus on *closed-set domain adaptation*, where the source and target domains share an identical label space. More relaxed settings have since been proposed to handle various forms of label space mismatch, including *partial domain adaptation* and *open-set domain adaptation*. Universal Domain Adaptation (UniDA) further generalizes these settings by allowing both the source and target domains to contain private classes in addition to a shared label subset.

Universal Domain Adaptation (UniDA) is a further extension of unsupervised domain adaptation, in which fewer assumptions are imposed on the label spaces of the source and target domains. Specifically, the source and target domains may share a common label set while also containing their own private classes. You [21] formally defined UniDA and showed that closed-set, partial, and open-set domain adaptation can all be viewed as special cases under UniDA .

In the closed-set setting, Deep Adaptation Network (DAN) [1] aligns source and target distributions within a deep convolutional network by embedding multiple task-specific layers into a reproducing kernel Hilbert space. It minimizes the multi-kernel maximum mean discrepancy (MK-MMD) to effectively reduce domain shift. By adopting a linear-time unbiased estimator, DAN achieves scalable training and strong performance on benchmarks such as Office-31. However, closed-set methods generally assume complete label space overlap and may suffer from severe negative transfer when this assumption is violated.

Partial domain adaptation relaxes this assumption by considering scenarios where the target label space is a subset of the source label space. IWAN [25] addresses this setting by introducing a two-discriminator adversarial framework. The first discriminator estimates sample importance weights to downweight source samples from outlier classes, while the second discriminator aligns the reweighted source distribution with the target. This design effectively suppresses negative transfer caused by irrelevant source classes. Similarly, SAN [26] proposes a selective adversarial network that employs multiple class-wise domain discriminators weighted by source classifier predictions. By selectively aligning only the shared classes and filtering out source-private classes, SAN further improves adaptation performance under partial domain shift.

Open-set domain adaptation considers the complementary scenario, where the target domain contains unknown classes absent from the source domain. Saito [27] propose an adversarial approach that trains a feature generator to either align target samples with known source classes or reject them as unknown based on a predefined threshold. This method enables effective separation of unknown target samples without requiring unknown-class annotations in the source domain, outperforming traditional distribution alignment approaches in open-set scenarios. Building upon this idea, Luo [28] propose Progressive Graph Learning, an end-to-end framework that integrates graph neural networks with episodic training to mitigate conditional shift. The method progressively identifies and rejects uncertain target samples as unknown while refining the classifier for shared classes.

Despite their effectiveness in specific settings, existing closed-set, partial, and open-set domain adaptation methods rely on restrictive assumptions about the label space relationship. UniDA provides a unified and more realistic formulation by simultaneously accounting for shared and private classes in both domains, motivating the development of more general and robust adaptation strategies.

B. Unknown Class Classification in UniDA

A core challenge in Universal Domain Adaptation (UniDA) lies in accurately identifying and classifying target samples belonging to unknown (private) classes, while simultaneously preserving reliable recognition of shared classes. Unlike conventional domain adaptation settings, UniDA requires not only domain alignment but also robust uncertainty estimation to distinguish shared-class samples from domain-private ones in the target domain.

Early UniDA methods primarily rely on uncertainty estimation derived from feature alignment and prediction behavior. You [21] provided the first systematic formulation of UniDA and proposed a domain-adversarial framework to align shared classes across domains. Unknown class identification is performed using uncertainty scores based on domain similarity and prediction entropy, under the assumption that unknown target samples are difficult to align and yield high

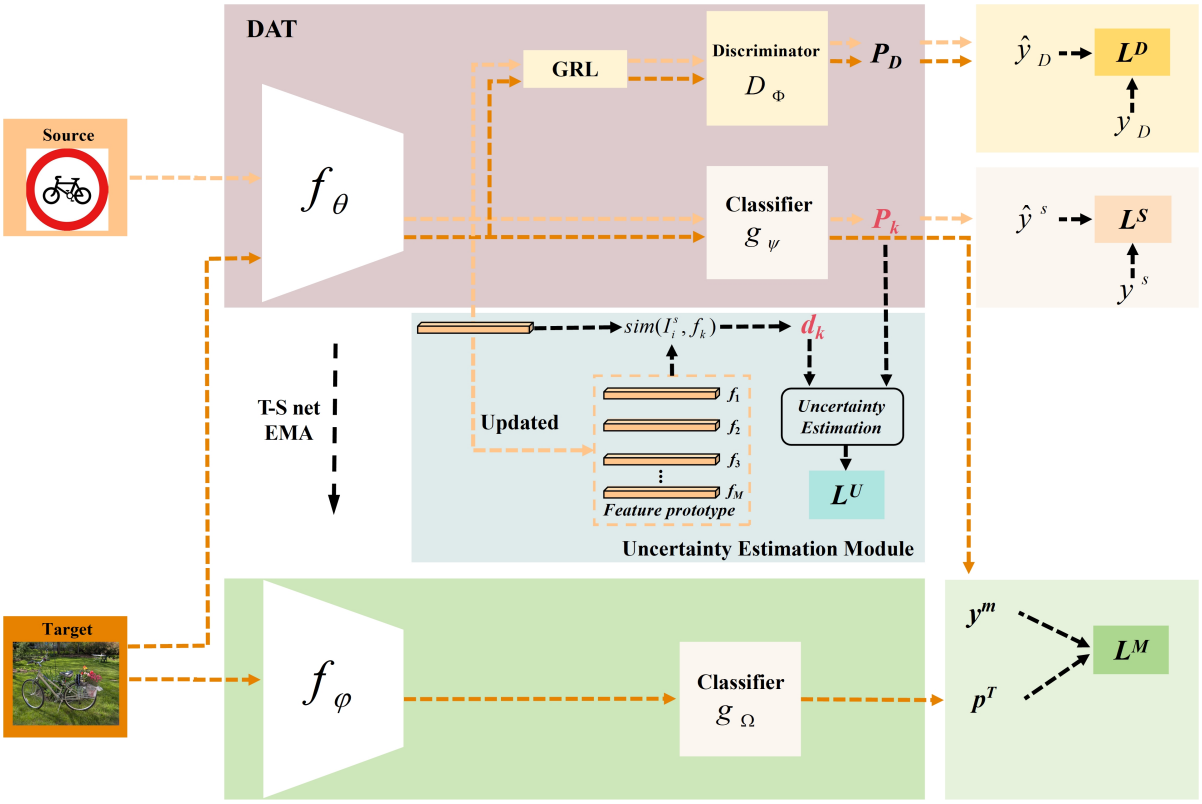

Fig. 2. The overall framework of the model. Consists of the teacher–student network, domain adversarial training, and uncertainty estimation module

uncertainty. Following this paradigm, Chen [29] introduced an UniDA framework that employs a cross-domain multi-sample contrastive loss based on mutual nearest neighbors to enhance shared-class matching. An evidence-based uncertainty measure is then used to distinguish known and unknown classes.

Subsequent works argue that uncertainty derived solely from global feature alignment may be insufficient, especially when target-private classes exhibit complex internal structures. DANCE [22] highlights that alignment-driven approaches often neglect the intrinsic data distribution of the target domain. To address this issue, DANCE leverages neighborhood clustering to capture target-domain structure, enabling more effective separation between shared and unknown classes. Similarly, Fu [24] exploits the complementary effects of entropy, prediction consistency, and confidence, and adopts multiple classifiers to model different degrees of uncertainty for improved unknown classes discrimination.

Beyond uncertainty-based scoring, several methods explicitly incorporate clustering or consensus mechanisms to identify shared semantic structures. DCC [23] proposes a domain-consensus clustering strategy that exploits shared knowledge at both semantic and sample levels, identifying shared classes via optimal cross-domain consistent clustering and treating inconsistent clusters as unknown classes. UniAM [30] further observes that many existing approaches emphasize global representations while overlooking discriminative local objects. It introduces a sparsely reconstructed attention mechanism to enhance common feature alignment (CFA) and target class separation (TCS), leading to improved unknown class identification.

More recently, UniDA has been extended to more challenging source-free scenarios. LEAD [31] and GLC [32] investigate universal domain adaptation without access to source domain data, relying on target-domain structure discovery and self-supervised signals to distinguish shared and unknown classes.

Despite notable progress, existing methods often rely on a single type of uncertainty cue, such as prediction entropy, alignment difficulty, or clustering inconsistency, which may be unreliable under severe domain shift or class overlap. Different from prior approaches, we propose a novel uncertainty estimation method that jointly leverages prediction confidence and feature–prototype similarity to assess sample categorization. By integrating semantic proximity with prediction reliability, our method provides a more robust criterion for unknown class classification and consistently improves performance across multiple benchmarks.

III. MOTHODOLOGY

A. Preliminary

In universal domain adaptation (UDA), we are given access to one labeled source domain and one unlabeled target domain. Let the input space be $\mathcal{X} \subset \mathbb{R}^d$ and the overall label space

be $\mathcal{Y} = \{1, 2, \ldots, K\}$. The source domain is denoted as $\mathcal{D}_s = \{(x_j^s, y_j^s)\}_{j=1}^{n_s}$, where $x_j^s \in \mathcal{X}$ is the j-th input sample in the source domain, $y_j^s \in \mathcal{Y}_S \subset \mathcal{Y}$ is the corresponding label, and n_s is the number of labeled samples in the source domain. The unlabeled target domain is represented as $\mathcal{D}_t = \{x_j^t\}_{j=1}^{n_t}$, where $x_j^t \in \mathcal{X}$ are drawn from a different distribution and n_t is the number of unlabeled samples in the target domain. The corresponding (unknown) label set of the target domain is denoted by $\mathcal{Y}_T \subset \mathcal{Y}$. We define the following sets to characterize label distribution across the source and target domains: the shared label set $\mathcal{Y}_C = \mathcal{Y}_S \cap \mathcal{Y}_T$, the source-private label set $\mathcal{Y}_{SP} = \mathcal{Y}_S \backslash \mathcal{Y}_T$, and the target-private label set $\mathcal{Y}_{TP} = \mathcal{Y}_T \setminus \mathcal{Y}_S$. We assume the source domain is drawn from a joint distribution $(x, y) \sim p(x, y)$, and the target domain from $(x, y) \sim q(x, y)$, with different marginals $p(x)$ and $q(x)$ reflecting domain shifts. The objective of UniDA is to learn a classifier $h : \mathcal{X} \to \mathcal{Y}$ using the labeled source data $\mathcal{D}_s$ and the unlabeled data $\mathcal{D}_t$ generalization well in the target domain. This paper proposes a model based on uncertainty estimation for UniDA. The model leverages techniques such as teacher-student networks, pseudo-labeling, and feature prototypes to effectively align the source and target domains while accurately classifying both the shared and private classes in the target domain. The model framework is illustrated in Fig. 2.

The model consists of the Domain adversarial training (DAT) module, teacher-student network (T-S net), and Uncertainty estimation module. DAT is used to address the problem of unaligned feature spaces. It consists of feature extractor f_θ, domain discriminator D_Φ and classifier g_ψ. In DAT, the feature extractor f_θ generates features for both source domain image x^s and target domain image x^t to deceive the domain discriminator D_Φ. When the domain discriminator D_Φ can not distinguish whether the image comes from the source or target domain, it can be considered that the feature spaces of the source and target domains are aligned. In order to encourage the model to learn the domain invariant features of the images, there is a gradient reversal layer (GRL) between the domain discriminator D_Φ and the feature extractor f_θ. The classifier g_ψ predicts label for the source domain image x^s. The target image x^t is input to the feature extractor f_θ, and the classifier g_ψ predicts label for x^t. x^t is also input to f_ϕ. g_Ω generates the pseudo-label p^T for x^t. The weights in the teacher-student network are updated through EMA.

B. Domain adversarial training

DAT is used to reduce domain shift. The loss function of DAT consists of source domain classification loss L^S and domain adversarial loss L^D. Define $h_\theta = f_\theta \circ g_\psi$. L^S is as follow:

$$\begin{aligned} L^S &= \frac{1}{n} \sum_{i=1}^{n} l(h_\theta(x_i^s), y_i^s) \\ l(\hat{y}, y) &= -\sum_{c=1}^{c} y_c \log \hat{y}_c \end{aligned} \tag{1}$$

$h_\theta(x_i^s)$ is the prediction for x_i^s. n denotes the number of samples in the batch. We use cross-entropy loss to calculate L^D. It express as follows:

$$L^D = \frac{1}{n} \sum_{i=1}^{n} [y_D \log \hat{y}_D] + \frac{1}{n} \sum_{j=1}^{n} [\log(1 - y_D) \log(1 - \hat{y}_D)] \tag{2}$$

where $y_D \in \{0, 1\}$ represents the ground-truth domain label, indicating whether a sample comes from the source domain ($y_D = 1$) or the target domain ($y_D = 0$).

C. Teacher-student network

To enhancing the model's ability to generate robust and discriminative features, the teacher-student network module is applied in model. The weight updating process is expressed as follows:

$$\varphi_{t+1} \leftarrow \alpha \varphi_t + (1 - \alpha)\theta_t \tag{3}$$

t represents the training step, and α is the smoothing factor. x_t is input to f_ϕ, and the classifier g_Ω predicts high quality pseudo-labels p^T for x_t.

$$p^T = \left[c = \arg\max g_\Omega(f_\varphi(x_t))\right] \tag{4}$$

$$y^m = g_\psi(f_\theta(x_t^m)) \tag{5}$$

L^M is expressed as follows:

$$L^M = \frac{1}{n} \sum_{i=1}^{n} q^T l(y_i^m, p^T) \tag{6}$$

q^T represents the quality estimation weighting. Since p^T of the image may be incorrect, especially noticeable during the initial stages of training, we introduce q^T in L^M. Expressed as follows:

$$q^T = \max g_\Omega(f_\varphi(x_t)) \tag{7}$$

D. Feature Prototype-based Similarity Calculation

The source domain images is input to f_θ to obtain their features before the training. The feature of images belonging to the same class are averaged to obtain the initialized feature prototypes $\{f_1, f_2, f_3, \ldots, f_M\}$, $M = |\mathcal{Y}_S|$. During training, the feature prototypes is updated by image feature I_j^s as follows:

$$f_j = \beta f_j + (1 - \beta) I_j^s \tag{8}$$

β is a hyperparameter. The maximum class prediction probability is as follows:

$$P_k = \arg \max_{j \in \{1,2,\ldots,M\}} p(y = j) \tag{9}$$

We then compute the cosine similarity between image feature I_j^s and the feature prototype f_k. The result is denoted as d_k, calculated as follows:

TABLE I
PERFORMANCE COMPARISON OF H-SCORE ON OFFICE-HOME FOR UNIDA

	A2C	A2P	A2R	C2A	C2P	C2R	P2A	P2C	P2R	R2A	R2C	R2P	Avg
DANCE [22]	26.7	11.3	18.0	33.2	12.5	14.3	41.6	39.9	33.3	16.3	27.1	25.9	25.0
OSBP [27]	39.6	45.1	46.2	45.7	45.2	46.8	45.3	40.5	45.8	45.1	41.6	46.9	44.5
DANN [33]	42.4	48.0	48.9	45.5	46.5	48.4	45.8	42.6	48.7	47.6	42.7	47.4	46.2
UAN [21]	51.6	51.7	54.3	61.7	57.6	61.9	50.4	47.6	61.5	62.9	52.6	65.2	56.6
CMU [24]	56.0	56.9	59.2	67.0	64.3	67.8	54.7	51.1	66.4	68.2	57.9	69.7	61.6
DCC [23]	58.0	54.1	58.0	74.6	70.6	77.5	64.3	**73.6**	75.0	81.0	**75.1**	80.4	70.1
OVANet [34]	62.8	75.5	78.6	70.7	68.8	75.0	71.3	58.6	80.5	76.1	64.1	78.9	71.7
TNT [29]	61.9	74.6	80.2	73.5	71.4	79.6	74.2	69.5	82.7	77.3	70.1	81.2	74.7
GLC [32]	64.3	78.2	**89.8**	63.1	81.7	**89.1**	**77.6**	54.2	**88.9**	80.7	54.2	85.9	75.7
SAN [35]	68.2	80.6	86.7	73.4	73.0	79.8	76.5	64.9	83.3	80.1	67.1	80.1	76.1
UniOT [36]	67.3	80.5	86.0	73.5	77.3	84.3	75.5	63.3	86.0	77.8	65.4	81.9	76.6
MLNet [37]	68.2	83.8	85.0	73.6	78.2	82.2	75.2	64.7	85.1	78.8	69.9	**83.9**	77.4
Ours	**70.2**	**84.1**	86.4	**75.3**	**82.2**	86.1	72.3	69.1	86.6	**81.3**	72.3	82.2	**79.0**

TABLE II
PERFORMANCE COMPARISON OF H-SCORE ON OFFICE-31 AND DOMAINNET DATASETS FOR UNIDA

Method	Office-31							DomainNet						
	A2D	A2W	D2A	D2W	W2A	W2D	AVG	P2R	P2S	R2P	R2S	S2P	S2R	AVG
DANN [33]	20.2	48.8	47.7	52.7	49.3	54.9	50.6	31.2	29.3	27.8	27.8	27.8	30.8	29.1
OSBP [27]	51.1	50.2	49.8	55.5	50.2	57.2	52.3	33.6	33.0	30.6	30.5	30.6	33.7	32.0
UAN [21]	59.7	58.6	60.1	70.6	60.3	71.4	63.5	41.9	43.6	39.0	39.0	38.7	43.7	41.0
CMU [24]	68.1	67.3	71.4	79.3	72.2	80.4	73.1	50.8	52.2	45.1	44.8	45.6	51.0	48.3
DANCE [22]	72.6	62.4	63.3	76.3	57.4	82.8	69.1	-	-	-	-	-	-	-
DCC [23]	88.5	78.5	70.2	79.3	75.9	88.6	80.2	56.9	50.3	43.7	44.9	43.3	56.2	49.2
TNT [29]	85.7	80.4	83.8	92.0	79.1	92.2	85.5	-	-	-	-	-	-	-
OVANet [34]	85.8	79.4	80.1	95.4	94.3	84.0	86.5	56.0	51.7	47.1	47.4	44.9	57.2	50.7
GLC [32]	81.5	84.5	**89.8**	90.4	88.4	92.3	87.8	**63.3**	50.5	54.9	50.9	**49.6**	61.3	55.1
UniOT [36]	87.0	88.5	88.4	98.8	87.6	96.6	91.2	59.3	47.8	51.8	46.8	48.3	58.3	52.1
MLNet [37]	**90.4**	**93.7**	89.7	96.2	88.4	**98.3**	92.8	-	-	-	-	-	-	-
Ours	90.2	92.6	88.1	**98.0**	**91.6**	96.9	**92.9**	59.5	**55.3**	**56.6**	**51.2**	48.3	**62.5**	**55.6**

$$d_k = \text{sim}(I_j^s, f_k)$$
$$\text{sim}(I_j^s, f_k) = \frac{I_j^s \cdot f_k}{\|I_j^s\| \|f_k\|} \tag{10}$$

E. Uncertainty estimation

In practice, each target sample falls into one of four cases:
(1) high P_k and high d_k (ideal known class)
(2) high P_k but low d_k (semantic shift)
(3) low P_k but high d_k (semantically close)
(4) low P_k and low d_k (typical unknown class).

For (1) and (2), high P_k dominates, so samples are regarded as known even with low similarity. In contrast, (3) and (4) are harder to separate, but (3) is semantically close to known classes. To address this, we design an uncertainty estimation method that learns a decision boundary between them. As follows:

$$e^{P_k - d_k} + d_k < \tau \tag{11}$$

In general, for cases (1) and (2), we use the exponential term to amplify the influence of. However, this amplification may cause the effect of to be overlooked, making samples in case (3) easily classified as unknown. Therefore, we add d_k as a penalty to balance the impact of the exponential term. In cases (3) and (4), the exponential term is close to 1, so in such situations, we linearly add d_k to give it more influence, allowing samples with high d_k to be more likely classified correctly as known.

Here, τ is a learnable parameter. If the number of samples in a batch that satisfy (11) is greater than or equal to $num = \text{batchsize}/4$, we update τ using binary cross-entropy loss. First, we compute the prediction logits:

$$y'_k = \tau - (e^{P_k - d_k} + d_k) \tag{12}$$

TABLE III
ABLATION EXPERIMENT FOR DIFFERENT COMPONENTS ON OFFICE-31

Method	A2D	A2W	D2A	D2W	W2A	W2D	AVG
w/o EMA	82.3	85.3	80.9	92.5	86.4	88.9	86.1
w/o DAT	87.3	90.5	86.3	95.9	89.9	94.3	90.7
w/o Uncertainty estimation	46.3	37.2	47.6	50.5	46.3	48.6	46.1
Ours	**90.2**	**92.6**	**88.1**	**98.0**	**91.6**	**96.9**	**92.9**

The true label is set to $y_{\text{unknown}} = 1$, and the BCE loss is calculated as:

$$L^U = \text{BCELoss}(y_{\text{unknown}}, y'_k) \tag{13}$$

During testing, the test image is input into the image encoder f_θ. P_k and d_k with the corresponding class are then computed. The label is predicted as follows:

$$label = \begin{cases} \text{unknown}, & \text{if } e^{P_k - d_k} + d_k < \tau \\ k, & \text{otherwise} \end{cases} \tag{14}$$

F. Optimization Objective Function

The overall optimization objective function is expressed as:

$$\min_{\theta,\tau} \max_{\phi} \Big(L^D(\theta, \phi) + L^S(\theta) + L^M(\theta) + L^U(\tau) \Big) \tag{15}$$

TABLE IV
ABLATION EXPERIMENT FOR DIFFERENT UNCERTAINTY ESTIMATION ON OFFICE-HOME

	H-score
$e^{(P_k - d_k)} + d_k$	79.0
$P_k + d_k$	73.3$_{(-5.7)}$
$e^{(P_k + d_k)}$	73.9$_{(-5.1)}$
$e^{(P_k - d_k)}$	70.3$_{(-8.7)}$
Only $P_k(\tau = 0.8)$	72.2$_{(-6.8)}$
Only $d_k(\tau = 0.8)$	54.8$_{(-24.2)}$

IV. EXPERIMENTS

A. datasets and evaluation protocols

a) datasets : We conduct experiments on 3 widely-used benchmark datasets for UniDA: **Office-31** [38] , **Office-Home** [39] and **DomainNet** [40], which cover a wide range of domain shifts in terms of visual appearance, data scale, and semantic diversity. **Office-31** is a classic benchmark for domain adaptation, consisting of 4,651 images from 31 object categories shared across three domains: Amazon (A), Webcam (W), and DSLR (D). The Amazon domain contains high-quality product images collected from online merchants, while the Webcam and DSLR domains consist of images captured by webcams and digital SLR cameras, respectively. **Office-Home** is a more challenging dataset designed to evaluate domain adaptation under larger domain gaps. It contains approximately 15,500 images from 65 categories across 4 domains: Art (A), Clipart (C), Product (P), and Real World (R). Compared to Office-31, Office-Home exhibits greater intra-class variation and more severe inter-domain discrepancies, making it a popular benchmark for domain adaptation settings. **DomainNet** is a large-scale and highly diverse benchmark for domain adaptation, comprising approximately 600,000 images from 345 categories over 6 domains: Clipart, Painting, Real, Sketch, Infograph, and Quickdraw. Due to its scale and complexity, DomainNet is commonly used to evaluate multi-source and generalized domain adaptation methods under realistic and challenging conditions. We do the same as others, conducting validation only in the 3 domains: painting (A), Real (R), and Sketch (S).

For Office31 and OfficeHome, we follow the label split approach of [36]. For DomainNetwe follow the label split method of [24]. For Office-31. We use the 10 classes shared by Office-31 [38] and Caltech-256 [41] as the common label set. Following alphabetical order, the next 10 classes are used as the source-private label set, and the remaining 11 classes are used as the target-private label set. For Office-Home. Following alphabetical order, the first 10 classes are selected as the common label set, the next 5 classes are used as the source-private label set, and the remaining classes are treated as the target-private label set. For DomainNet, Following the alphabetical order of class labels, the first 150 classes are selected as the common label set, the next 50 classes are treated as the source-private label set, and the remaining classes are regarded as the target-private label set.

b) Evaluation Protocols: We use the H-score, which comprehensively assesses the model's classification performance for both $\mathcal{Y}_C$ and $\mathcal{Y}_{TP}$. $Acc_{\mathcal{Y}_C}$ represents the classification accuracy for $\mathcal{Y}_C$, and $Acc_{\mathcal{Y}_{TP}}$ denotes the classification accuracy for $\mathcal{Y}_{TP}$. The H-score is calculated as:

$$\text{H-score} = \frac{2 \times Acc_{\mathcal{Y}_C} \times Acc_{\mathcal{Y}_{TP}}}{Acc_{\mathcal{Y}_C} + Acc_{\mathcal{Y}_{TP}}} \tag{16}$$

B. experiments details

We adopt ViT-B_16 as the backbone for both the teacher and student networks, initialized with ImageNet-21K pretrained weights. The model is optimized using stochastic gradient descent (SGD) with a learning rate of 0.002. A warm-up strategy is applied for the first epoch with a learning rate of 2×10^{-5}, followed by a cosine decay learning rate schedule. The batch size is set to 32. The initial value of τ is 2.1, and the

momentum coefficients α and β are both set to 0.999. Color-based data augmentation is employed, including adjustments of brightness, contrast, hue, and blur. All hyperparameters are determined through preliminary experiments. All experiments are conducted on a single NVIDIA RTX 4090 GPU with 24 GB memory.

C. experiments results

Our method is compared with current state-of-the-art (SOTA) baselines. All baseline results are either reported from the original papers or reproduced by us. Table I presents the experimental results on the Office-Home dataset, while Table II reports the results on the Office-31 and DomainNet datasets. Taking Table I as an example, A2C denotes the UniDA task from *Art* to *Clipart*, and AVG represents the average H-score over all UniDA tasks. Our model achieves state-of-the-art performance on Office-31, Office-Home, and DomainNet. Specifically, on Office-Home, our method attains an average H-score of 79.0, outperforming the second-best method MLNet [37] (77.4) by 1.6 . On Office-31, our model achieves an average H-score of 92.9, exceeding MLNet [37] (92.8) by 0.1. On DomainNet, our model achieves an average H-score of 55.6, surpassing GLC [32] (55.1) by 0.5. Overall, our method consistently achieves state-of-the-art results across all evaluated datasets, demonstrating its robustness and strong performance under diverse and challenging domain adaptation settings.

D. ablation experiment

a) Component Ablation Study.: We conduct component ablation experiments on the Office-31 dataset to investigate the impact of the EMA-based weight updating strategy, domain adversarial training (DAT), and the uncertainty estimation module on model performance. The experimental results are summarized in Table III. In the *w/o uncertainty estimation* setting, images are classified into known categories when $d_k > 0.8$, and directly assigned to the unknown class otherwise. In the *w/o EMA* setting, the model achieves an average H-score of 86.1, which is 6.8 points lower than our model proposed in this paper. Removing DAT (*w/o DAT*) results in an average H-score of 90.7, corresponding to a performance drop of 2.2 points. Notably, in the *w/o uncertainty estimation* setting, the model performance nearly collapses, yielding an average H-score of only 46.1, which is a substantial decrease of 46.8 points. This severe degradation arises because a fixed threshold ignores the sample-wise variability across different instances. Relying solely on d_k to determine whether a sample belongs to a known or unknown class is inherently unreliable. These results further validate the effectiveness and necessity of the proposed uncertainty estimation module.

b) Ablation Study on Uncertainty Estimation.: To further investigate the impact of different uncertainty estimation strategies based on P_k and d_k on model performance, we conduct a series of ablation studies. The evaluated uncertainty estimation methods include a simple linear formulation $P_k + d_k$, two exponential variants $e^{(P_k - d_k)}$ and $e^{(P_k + d_k)}$, as well as threshold-based baselines that rely solely on either P_k or d_k. All these methods can be regarded as variants of the proposed uncertainty estimation function $e^{(P_k - d_k)} + d_k$.

The ablation results on the Office-Home dataset are reported in TableIV. Our method achieves 79.0 in H-score. The simple linear formulation $P_k + d_k$ achieves an H-score of 73.3, which is 5.7 lower than our method. The uncertainty estimation strategy using exponential amplification $e^{(P_k + d_k)}$ obtains an H-score of 73.9, resulting in a performance drop of 5.1. Similarly, the method based on $e^{(P_k - d_k)}$ achieves an H-score of 70.3, which is 8.7 lower than our method. As analyzed in subection uncertainty estimation of methodology, the formulation $e^{(P_k - d_k)}$ fails to effectively distinguish between cases (3) and (4), leading to sub-optimal performance. Moreover, the classification strategies that rely solely on P_k or d_k yield H-scores of 72.2 and 54.8, respectively, with performance drops of 6.8 and 24.2. This result is expected, as relying on a single indicator is insufficient to accurately detect unknown classes. Consistent with the 4 cases analysis presented in subection uncertainty estimation of methodology, the proposed uncertainty estimation method is able to effectively separate unknown classes. Overall, the experimental results demonstrate that the proposed formulation $e^{(P_k - d_k)} + d_k$ significantly outperforms its variants, validating its effectiveness and robustness.

V. Conclusion

In this paper, we present an novel framework for Universal Domain Adaptation that jointly addresses domain alignment and unknown class identification. By integrating a teacher–student architecture with domain adversarial training, the proposed method enables stable semantic learning on unlabeled target data while mitigating domain shift. A key contribution of this work lies in a novel uncertainty estimation method that combines prediction confidence with feature–prototype similarity. Unlike prior approaches that rely on a single uncertainty cue, our formulation explicitly captures both classifier reliability and semantic proximity, allowing for a more robust separation between shared and target-private classes under severe domain shift. Extensive experiments on Office-31, Office-Home, and DomainNet demonstrate that the proposed method consistently achieves state-of-the-art performance. Comprehensive ablation studies further verify the effectiveness of each component and highlight the critical role of the proposed uncertainty estimation mechanism. We believe this work provides a principled and extensible perspective on UniDA, and offers a practical foundation for future research on generalized and open-world domain adaptation.

References

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

[2] Guoliang Kang, Lu Jiang, Yi Yang, and Alexander G Hauptmann, "Contrastive adaptation network for unsupervised domain adaptation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 4893–4902.

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

[4] Mingsheng Long, Han Zhu, Jianmin Wang, and Michael I Jordan, "Unsupervised domain adaptation with residual transfer networks," *Advances in neural information processing systems*, vol. 29, 2016.

[5] Shiqi Yang, Yaxing Wang, Joost Van De Weijer, Luis Herranz, and Shangling Jui, "Generalized source-free domain adaptation," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 8978–8987.

[6] Sicheng Zhao, Jing Jiang, Wenbo Tang, Jiankun Zhu, Hui Chen, Pengfei Xu, Björn W Schuller, Jianhua Tao, Hongxun Yao, and Guiguang Ding, "Multi-source multi-modal domain adaptation," *Information Fusion*, vol. 117, pp. 102862, 2025.

[7] Han-Kai Hsu, Chun-Han Yao, Yi-Hsuan Tsai, Wei-Chih Hung, Hung-Yu Tseng, Maneesh Singh, and Ming-Hsuan Yang, "Progressive domain adaptation for object detection," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2020, pp. 749–757.

[8] Naoto Inoue, Ryosuke Furuta, Toshihiko Yamasaki, and Kiyoharu Aizawa, "Cross-domain weakly-supervised object detection through progressive domain adaptation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 5001–5009.

[9] Poojan Oza, Vishwanath A Sindagi, Vishal M Patel, et al., "Unsupervised domain adaptation of object detectors: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 6, pp. 4018–4040, 2023.

[10] Dayan Guan, Jiaxing Huang, Aoran Xiao, Shijian Lu, and Yanpeng Cao, "Uncertainty-aware unsupervised domain adaptation in object detection," *IEEE Transactions on Multimedia*, vol. 24, pp. 2502–2514, 2021.

[11] Guofa Li, Zefeng Ji, and Xingda Qu, "Stepwise domain adaptation (sda) for object detection in autonomous vehicles using an adaptive centernet," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 10, pp. 17729–17743, 2022.

[12] Jinlong Li, Runsheng Xu, Xinyu Liu, Jin Ma, Baolu Li, Qin Zou, Jiaqi Ma, and Hongkai Yu, "Domain adaptation based object detection for autonomous driving in foggy and rainy weather," *IEEE Transactions on Intelligent Vehicles*, 2024.

[13] Wenzhang Zhou, Dawei Du, Libo Zhang, Tiejian Luo, and Yanjun Wu, "Multi-granularity alignment domain adaptation for object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 9581–9590.

[14] Yang Zhang, Philip David, and Boqing Gong, "Curriculum domain adaptation for semantic segmentation of urban scenes," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2020–2030.

[15] Yunsheng Li, Lu Yuan, and Nuno Vasconcelos, "Bidirectional learning for domain adaptation of semantic segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 6936–6945.

[16] Yuang Liu, Wei Zhang, and Jun Wang, "Source-free domain adaptation for semantic segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 1215–1224.

[17] Jingyi Xu, Weidong Yang, Lingdong Kong, Youquan Liu, Qingyuan Zhou, Rui Zhang, Zhijun Li, Wen-Ming Chen, and Ben Fei, "Visual foundation models boost cross-modal unsupervised domain adaptation for 3d semantic segmentation," *IEEE Transactions on Intelligent Transportation Systems*, 2025.

[18] Mathilde Bateson, Hoel Kervadec, Jose Dolz, Hervé Lombaert, and Ismail Ben Ayed, "Source-free domain adaptation for image segmentation," *Medical Image Analysis*, vol. 82, pp. 102617, 2022.

[19] Jin Hong, Yu-Dong Zhang, and Weitian Chen, "Source-free unsupervised domain adaptation for cross-modality abdominal multi-organ segmentation," *Knowledge-Based Systems*, vol. 250, pp. 109155, 2022.

[20] Huan Gao, Jichang Guo, Guoli Wang, and Qian Zhang, "Cross-domain correlation distillation for unsupervised domain adaptation in nighttime semantic segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 9913–9923.

[21] 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*, 2019, pp. 2720–2729.

[22] Kuniaki Saito, Donghyun Kim, Stan Sclaroff, and Kate Saenko, "Universal domain adaptation through self supervision," *Advances in neural information processing systems*, vol. 33, pp. 16282–16292, 2020.

[23] 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*, 2021, pp. 9757–9766.

[24] 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*. Springer, 2020, pp. 567–583.

[25] Jing Zhang, Zewei Ding, Wanqing Li, and Philip Ogunbona, "Importance weighted adversarial nets for partial domain adaptation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8156–8164.

[26] Zhangjie Cao, Mingsheng Long, Jianmin Wang, and Michael I Jordan, "Partial transfer learning with selective adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 2724–2732.

[27] Kuniaki Saito, Shohei Yamamoto, Yoshitaka Ushiku, and Tatsuya Harada, "Open set domain adaptation by backpropagation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 153–168.

[28] Yadan Luo, Zijian Wang, Zi Huang, and Mahsa Baktashmotlagh, "Progressive graph learning for open-set domain adaptation," in *International Conference on Machine Learning*. PMLR, 2020, pp. 6468–6478.

[29] Liang Chen, Yihang Lou, Jianzhong He, Tao Bai, and Minghua Deng, "Evidential neighborhood contrastive learning for universal domain adaptation," in *Proceedings of the AAAI conference on artificial intelligence*, 2022, vol. 36, pp. 6258–6267.

[30] Didi Zhu, Yinchuan Li, Junkun Yuan, Zexi Li, Kun Kuang, and Chao Wu, "Universal domain adaptation via compressive attention matching," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 6974–6985.

[31] Sanqing Qu, Tianpei Zou, Lianghua He, Florian Röhrbein, Alois Knoll, Guang Chen, and Changjun Jiang, "Lead: Learning decomposition for source-free universal domain adaptation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 23334–23343.

[32] Sanqing Qu, Tianpei Zou, Florian Röhrbein, Cewu Lu, Guang Chen, Dacheng Tao, and Changjun Jiang, "Upcycling models under domain and category shift," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 20019–20028.

[33] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario March, and Victor Lempitsky, "Domain-adversarial training of neural networks," *Journal of machine learning research*, vol. 17, no. 59, pp. 1–35, 2016.

[34] 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*, 2021, pp. 9000–9009.

[35] Zelin Zang, Lei Shang, Senqiao Yang, Fei Wang, Baigui Sun, Xuansong Xie, and Stan Z Li, "Boosting novel category discovery over domains with soft contrastive learning and all in one classifier," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 11858–11867.

[36] Wanxing Chang, Ye Shi, Hoang Tuan, and Jingya Wang, "Unified optimal transport framework for universal domain adaptation," *Advances in Neural Information Processing Systems*, vol. 35, pp. 29512–29524, 2022.

[37] Yanzuo Lu, Meng Shen, Andy J Ma, Xiaohua Xie, and Jian-Huang Lai, "Mlnet: Mutual learning network with neighborhood invariance for universal domain adaptation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024, vol. 38, pp. 3900–3908.

[38] Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell, "Adapting visual category models to new domains," in *Computer vision–ECCV 2010: 11th European conference on computer vision, Heraklion, Crete, Greece, September 5-11, 2010, proceedings, part iV 11*. Springer, 2010, pp. 213–226.

[39] 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*, 2017, pp. 5018–5027.

[40] 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*, 2019, pp. 1406–1415.

[41] Boqing Gong, Yuan Shi, Fei Sha, and Kristen Grauman, "Geodesic flow kernel for unsupervised domain adaptation," in *2012 IEEE conference on computer vision and pattern recognition*. IEEE, 2012, pp. 2066–2073.