

Label Calibration in Source Free Domain Adaptation

Shivangi Rai Rini Smita Thakur Kunal Jangid Vinod K Kurmi
Indian Institute of Science Education and Research Bhopal, India
{shivangir23, rinithakur, kunal24, vinodkk}@iiserb.ac.in

Abstract

Source-free domain adaptation (SFDA) utilizes a pre-trained source model with unlabeled target data. Self-supervised SFDA techniques generate pseudolabels from the pre-trained source model, but these pseudolabels often contain noise due to domain discrepancies between the source and target domains. Traditional self-supervised SFDA techniques rely on deterministic model predictions using the softmax function, leading to unreliable pseudolabels. In this work, we propose to introduce predictive uncertainty and softmax calibration for pseudolabel refinement using evidential deep learning. The Dirichlet prior is placed over the output of the target network to capture uncertainty using evidence with a single forward pass. Furthermore, softmax calibration solves the translation invariance problem to assist in learning with noisy labels. We incorporate a combination of evidential deep learning loss and information maximization loss with calibrated softmax in both prior and non-prior target knowledge SFDA settings. Extensive experimental analysis shows that our method outperforms other state-of-the-art methods on benchmark datasets. The code is available at <https://visdomlab.github.io/EKS/>.

1. Introduction

Deep learning has shown tremendous performance improvement on various computer vision tasks based on two assumptions (i) identical and independent distribution of training and test data (ii) availability of an enormous amount of labeled dataset [17]. However, in the real world setting there is always a constraint on labeled data availability, and there exists a domain discrepancy between training (source) and the test (target) dataset. Unsupervised domain adaptation (UDA) tackles this constraint by the inclusion of labeled source data and unlabeled target data, which effectively encounter the performance degradation of out-of-distribution sample. However, source data is not readily available during the adaptation stage due to privacy concerns and copyright issues. Therefore, source free domain adaptation (SFDA) [12] has access to only the source model

and unlabeled target data during adaptation stage. The principal methodologies of SFDA involve image style translation, domain-based reconstruction, self-training, and self-attention [14]. The self-training generates pseudolabels of the target and adapts the network with re-training.

The uncertainty quantification plays a pivotal role in pseudolabeling based SFDA methods. The predictive uncertainty comprises data, model and distributional uncertainty, which aids in selecting the appropriate high-confidence pixels of the unlabeled target data [8]. The pseudolabel refinement self-training SFDA methods are based on dropout based uncertainty thresholding, ensemble, Bayesian uncertainty matching, etc [13], [9], [6]. Bayesian modeling leverages moments approximation by learning posterior over weights whereas ensemble models utilize various independently trained networks for analytical estimation of the class probability distributions. These methods are computationally inefficient with memory and time constraints. Bayesian neural networks (BNN's) also suffer from the problem of intractable posterior inference, selection of weight prior, and sampling requirement at inference. The predictive uncertainty obtained by the UDA methods relies on the point estimate of the output, which fails to capture the miscalibration caused by the distribution shift between the source and target [31]. To overcome these problems, we propose dirichlet-based uncertainty calibration, which captures distributional uncertainty using evidential deep learning.

Evidential Deep Learning (EDL) quantifies distributional uncertainty using the Dempster-Shefer theory of evidence and subjective logic viewpoint [26]. In EDL, dirichlet prior is placed over the output predictions instead of network weights. Thus, it transforms the point estimate into the probability distribution over the simplex. Without the requirement for sampling, a grounded representation of both epistemic and aleatoric uncertainty can be learned by training a neural network to output the hyperparameters of the higher-order evidential distribution [2]. EDL computes predictive uncertainty with a single forward pass, unlike BNN's, thereby curtailing the computational requirements. Apart from distributional uncertainty, the other factor which

can lead to erroneous results in the SFDA setting is the translational invariance of softmax. The softmax operator for multi-class classification gives the same results for two samples if there exists a relative relationship between their logits. For example, the results of the softmax remain unaltered when the logits are added or subtracted by the particular constant value. It can lead to miscalibration in the SFDA setting with noisy pseudolabels. The problem of translational invariance can be solved with calibrated softmax and requisite evidential learning based loss as the conventional cross-entropy loss poses an issue of gradient shrinking with calibrated softmax [34].

- Inclusion of Dirichlet-based uncertainty calibration for pseudolabel refinement in both prior knowledge based SFDA setting (EKS) and on non-prior SFDA network (ES). In the prior setting, unary and binary bound is available for the unlabeled training target dataset.
- Introduction of a calibrated softmax function at the adaptation stage to reduce the effect of translation invariance. The network is trained with novel EDL loss incorporating the effect of distributional uncertainty and translation invariance.
- Analysis of the combination of EDL with information maximization (IM). Experiments validate our hypothesis that combining EDL with IM leads to better adaptation to the target domain.
- Extensive experiments are performed on benchmark datasets: DomainNet40, Office-Home, Office-31 and Digits datasets such as MNIST, SVHN, and USPS, which show performance enhancement over other state-of-the-art methods.

2. Related Work

2.1. Pseudolabeling in SFDA

Pseudolabeling is one of the important SFDA self-supervised self training method. The pseudolabeling approach is divided into three stages: prototype generation, pseudolabel assignment and pseudolabel filtering. DeepCluster [5] is the popular method to generate prototype of unlabeled target using K-means clustering. The pseudolabel assignment is usually based on the distance metric between pixel and prototype. The pseudolabels are inevitably noisy as the source model is used for its generation. The filtering or denoising of pseudolabels involve different uncertainty calibration techniques as given in [16], [18], [6], [24], [30]. DPL [6] generates pseudolabels by prototype estimation followed by class and pixel level uncertainty calibration. UPL-SFDA [30] quantify uncertainty of pseudolabels with duplication of pre-trained source model head multiple times with perturbations. The other SFDA setting involves loss reweighting with uncertainty calibrated reliable pseudola-

bel in self-supervised contrastive network. SHOT [15] is the benchmark SFDA work with self-supervised pseudolabeling and information maximization.

The recent settings of SFDA explores prior knowledge and active selection of target samples to incorporate more information about unlabeled target. In real world scenario, prior knowledge of target is available in terms of constraints on features, predictions, annotation, class proportion, etc. KSHOT [27] further modify the SHOT settings for the more realistic real world adaptation problem by inclusion of prior knowledge by putting constraints on unary and binary bound. POST [23] is prior guided self-training SFDA network for human pose estimation. It imposes prediction and feature consistency between the student and teacher for adaptation with human pose prior. Future research directions aim to reduce computational constraints due to the requirement of memory bank and Bayesian sampling.

2.2. Evidential Deep Learning

EDL explicitly models uncertainty on the basis of evidence for various computer vision tasks such as classification, regression, segmentation and object detection [2], [26], [28], [21]. It has shown unprecedented success in showing robustness against out-of-distribution samples and adversarial attacks. EDL leverages higher-order evidential prior over likelihood distribution to effectively capture aleatoric, epistemic and distributional uncertainty. The output of evidential network gives prediction along with evidence. EDL effectively differentiates between different sources of uncertainty as it can extricate (a) conflicting evidence (b) lack of evidence. In evidential classification networks, dirichlet prior is placed over multinomial distribution, whereas in regression networks, normal inverse gamma prior is placed over Gaussian distribution. The training methodology of EDL includes evidential losses in collaboration with the heuristic evidence regularization for uncertainty guidance [20]. In [4], EDL utilizes contrastive debiasing for open set action recognition problem to counter static bias and over-confident predictions. DEED [3] introduce EDL in multilabel classification target setting (multilabel MNIST) by mapping between evidential space and embedding space. The deep evidential fusion network performs multi-modal medical image classification [32].

Recently, federated evidential active learning has been introduced with both aleatoric and epistemic uncertainty calibration along with diversity relaxation strategy for image segmentation [7]. EDL has also been incorporated in continual learning setup to perform incremental object classification and out-of-distribution detection [1]. Although, EDL has been explored for classification, regression, federated learning, continual learning, meta-learning, its scope is not analyzed for SFDA settings.

3. Problem Statement

We address SFDA problem doing K class classification using the pre-trained source model (f_S) trained with labeled source data $\{x_i^s, y_i^s\}_{i=1}^{n_s}$ under the domain D_s , unlabeled target domain data $\{x_i^t\}_{i=1}^{n_t}$ under domain D_t with prior knowledge about target domain distribution (P_K). The inclusion of prior knowledge is done using unary bound (UB) and binary relationship (BR) of the unlabeled target data following the convention of [27]. Here, the objective of **Evidential KSHOT(EKS)** is to learn the target model (f_T) and infer $\{y_i^t\}_{i=1}^{n_t}$ using pre-trained source model and unlabeled target data with prior information and the objective of **Evidential SHOT(ES)** is to learn the target model (f_T) and infer $\{y_i^t\}_{i=1}^{n_t}$ using pre-trained source model (f_S) and unlabeled target data (D_t), under self-supervised setting without the use of prior knowledge.

4. Methodology

The source model (f_S) is trained using the labeled source data with label smoothing cross entropy objective. The adaptation of the target model relies on the self-training with pseudolabels generated from the source. We generate the pseudolabels using the prototype estimation and further refine them with prior knowledge as given in KSHOT [27]. The model is further adapted with EDL and calibrated IM as elaborated in Section 4.2 and 4.3. Figure 1 illustrates the proposed EKS method which demonstrates the evidential modeling through Dirichlet based calibration and its integration with information maximization.

4.1. Prior Knowledge

The pseudolabels obtained by the pre-trained source model are further refined by prior knowledge constraints of unary bound and binary relationship as given below:

1. **Binary Relationship:** It represents the relationship between two target classes.

$$\{(p_t^{(k_i)} - p_t^{(k_{i+1})} \geq 0) \mid i \in \{0, 1, \dots, K-2\}\}$$

2. **Unary Bound knowledge creation:**

$$\{a_k \cdot (1 - \sigma) \leq p_t^{(k)} \leq a_k \cdot (1 + \sigma) \mid k \in \mathcal{K}\}$$

where $\sigma \in \{0.0, 0.1, 0.5, 1.0, 2.0\}$, a_k is the empirical probability of the k^{th} class, $p_t^{(k)}$ is the class probability of the target distribution, K represents total number of classes, σ controls the degree of correctness of prior knowledge estimation.

As we keep increasing the value of σ from 0.0 to 2.0, we keep introducing discrepancy in the quality of prior knowledge which leads to significant difference between the distribution of pseudo labels and the ground truth. Due to lack

of knowledge under domain shifts, the performance of deterministic models deteriorates as they cannot accurately quantify the uncertainty associated with the new, unseen data distribution. Thus, we incorporate a Dirichlet based prediction calibration based on Evidential Deep Learning (EDL) along with calibrated information maximization for prior knowledge guided SFDA.

4.2. Dirichlet based uncertainty calibration for knowledge guided SFDA

Under domain shifts, the performance of deterministic models deteriorates as they cannot accurately quantify the uncertainty associated with the unseen data distribution of the target domain. Predictive uncertainty is important as it shows how confident a model is in its predictions, especially when there is domain shift. Moreover, self-supervised pseudo-labels require predictive uncertainty calibration for its refinement. The Dirichlet-based model predicts class probability by placing a dirichlet distribution over the class probabilities on target model, rather than treating prediction as single point estimate. The probability density function of ρ for given target sample x_i^t is expressed as:

$$p(\rho|x_i^t, \theta) = \text{Dir}(\rho|\alpha_i) = \begin{cases} \frac{\Gamma(\sum_{k=1}^K \alpha_{i_k})}{\prod_{k=1}^K \Gamma(\alpha_{i_k})} \prod_{k=1}^K \rho_k^{\alpha_{i_k}-1}, & \text{if } \rho \in \Delta^K \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where, α_i is the parameter of the Dirichlet distribution for x_i^t , $\Gamma(\cdot)$ is the gamma function, K is total number of classes and $\Delta^K = \left\{ \rho_i \mid \sum_{k=1}^K \rho_{i_k} = 1 \text{ and } \forall j \ 0 \leq \rho_{i_k} \leq 1 \right\}$ is a K -dimensional unit simplex, where ρ_i is the probability density function for the i^{th} sample and ρ_{i_k} denotes probability density function for sample i corresponding to class k . EDL modifies the loss function with evidence, which basically deduces the confidence of a particular prediction on the basis of the amount of support collected from data.

The predicted probability of target sample x_i^t for class k can be obtained as:

$$\begin{aligned} P(y = k|x_i^t, \theta) &= \int p(y = k|\rho_i) p(\rho_i|x_i^t, \theta) d\rho_i \\ &= \frac{\alpha_{i_k}}{\sum_{k=1}^K \alpha_{i_k}} \\ &= \frac{\exp(o_{i_k}) + \lambda}{\sum_{k=1}^K (\exp(o_{i_k}) + \lambda)}. \end{aligned} \quad (2)$$

here, $\alpha_{i_k} = \exp(o_{i_k}) + \lambda$, o_{i_k} is the observed logit for i^{th} sample corresponding to k^{th} class and λ is a constant.

While adapting the source trained model on the target domain, we minimize the negative log of the marginal likelihood (L_{nll}) to incorporate the discriminability in the

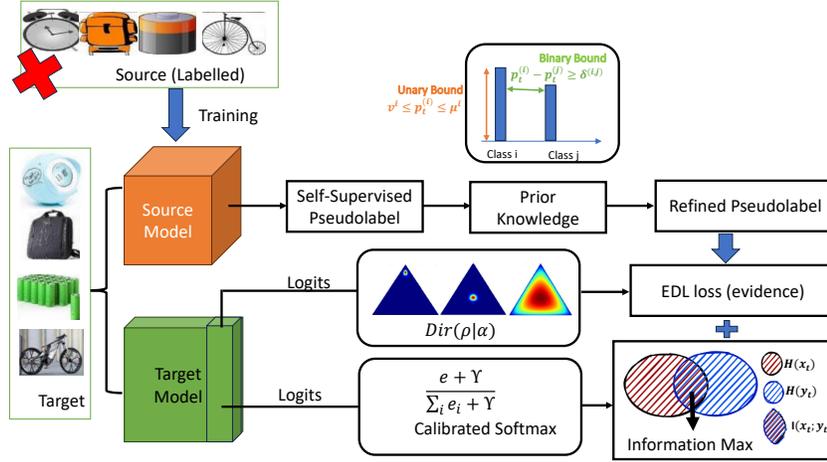


Figure 1. Overview of the proposed EKS method. The Dirichlet distribution and calibrated softmax is placed over logits of the target model followed by adaptation with EDL loss and IM loss in prior knowledge SFDA setting. In the proposed ES method, there is non-availability of prior knowledge, and self-supervised pseudolabels are directly used for EDL loss (best view in color).

adapted model. Discrimination-based loss (L_{nll}) is defined as:

$$\begin{aligned} \mathcal{L}_{nll} &= -\frac{1}{N} \sum_{i=1}^N \log P(y = k | x_i^t, \theta) \\ &= \frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K y_{ik} \left[\log \left(\sum_{j=1}^K \alpha_{ij} \right) - \log \alpha_{ik} \right] \end{aligned} \quad (3)$$

In Eq. 3, y_{ik} denotes the refined pseudolabels with prior knowledge.

Divergence-based loss (L_{kl}) minimizes the divergence between the predicted Dirichlet distribution and a target Dirichlet distribution. It reduces the impact of the evidence of the false labels for a target sample by reducing it to zero while adapting the model to the target domain.

$$\begin{aligned} \mathcal{L}_{kl} &= \frac{1}{NK} \sum_{i=1}^N D_{KL}(\text{Dir}(\rho_i | \tilde{\alpha}_i) \| \text{Dir}(\rho_i | \mathbf{1})) \\ &= \frac{1}{NK} \sum_{i=1}^N \left[\log \left(\frac{\Gamma(\sum_{j=1}^K \tilde{\alpha}_{ij})}{\Gamma(K) \prod_{j=1}^K \Gamma(\tilde{\alpha}_{ij})} \right) \right. \\ &\quad \left. + \sum_{k=1}^K (\tilde{\alpha}_{ik} - 1) \left(\psi(\tilde{\alpha}_{ik}) - \psi \left(\sum_{j=1}^K \tilde{\alpha}_{ij} \right) \right) \right] \end{aligned} \quad (4)$$

where, $\tilde{\alpha}_i = \mathbf{y}_i + (1 - \mathbf{y}_i) \odot \alpha_i$, is the Dirichlet parameter for the complementary labels, $\mathbf{1}$ is a vector of K ones, \odot is the element-wise multiplication operator and $\psi(\cdot)$ represents the digamma function.

Discrimination-based loss (L_{nll}) and Divergence-based loss (L_{kl}) together constitutes EDL loss (L_{edl}).

$$L_{edl} = L_{nll} + \beta * L_{kl} \quad (5)$$

where β is a hyperparameter, which balances the two losses.

4.3. Calibrated Softmax integration in Information Maximization

Entropy-based loss (L_{ent}) works towards reducing the uncertainty in each prediction by making the predictions confident.

$$L_{ent}(f_t; D_t) = -\mathbb{E}_{d_t \in D_t} \left[\sum_{k=1}^K \delta_k(f_t(d_t)) \log \delta_k(f_t(d_t)) \right] \quad (6)$$

where, D_t represents complete target domain data, d_t is the sample from the target domain, f_t is the target domain adapted model, δ_k represents standard softmax function, K is the total number of classes.

Divergence-based loss (L_{div}) penalizes the model when there is lesser diversity in predictions.

$$L_{div}(f_t; D_t) = \sum_{k=1}^K \hat{p}_k \log \hat{p}_k = D_{KL}(\hat{p}, K^{-1} \mathbf{1}_K) - \log K \quad (7)$$

where \hat{p}_k is the average class probability of the k^{th} class. L_{div} estimates the discrepancy between the empirical distributions of target predictions with uniform distribution, thereby encouraging the target network to make diverse predictions by avoiding bias towards a particular class.

But in cases where the source model is trained on a very small set of data or data from a very different distribution than the target domain, L_{ent} might push the model to make overconfident predictions as we go on minimizing the entropy. Standard softmax (Eq. 8) performs good on common-closed datasets with clean labels but it fails with erroneous results in the noisy label SFDA setting. Thus, we incorporate a calibrated softmax function (Eq. 9) to tackle the problem of overconfident prediction and also to do away with the translation invariance problem of standard softmax function. The standard softmax (δ_{ik}) and calibrated softmax ($\hat{\delta}_{ik}$) is given in Eq. 8 and 9, respectively.

$$\delta_{i_k} = \frac{e^{o_{i_k}}}{\sum_{j=1}^K e^{o_{i_j}}}, \quad k = 1, 2, \dots, K. \quad (8)$$

$$\hat{\delta}_{i_k} = \frac{e^{o_{i_k}} + \gamma}{\sum_{j=1}^K (e^{o_{i_j}} + \gamma)}, \quad k = 1, 2, \dots, K. \quad (9)$$

In Eq. 9, γ is a calibrated softmax constant, which is a hyperparameter. Calibrated Entropy-based loss (\hat{L}_{ent}) make use of calibrated softmax function ($\hat{\delta}_{i_k}$) rather than standard softmax function (δ_{i_k}). Similarly, we leverage calibrated softmax function into the Divergence-based loss, thereby getting (\hat{L}_{div}).

$$\hat{L}_{ent}(f_t; D_t) = -\mathbb{E}_{d_t \in D_t} \left[\sum_{k=1}^K \delta_k(f_t(d_t)) \log \delta_k(f_t(d_t)) \right] \quad (10)$$

$$\hat{L}_{div}(f_t; D_t) = \sum_{k=1}^K \hat{p}_{c_k} \log \hat{p}_{c_k} = D_{KL}(\hat{p}_c, K^{-1} \mathbf{1}_K) - \log K \quad (11)$$

where, \hat{p}_{c_k} is the average class probability for the k_{th} class after the integration of calibrated softmax Eq. (9).

Calibrated Entropy-based loss (\hat{L}_{ent}) and Calibrated Divergence-based loss (\hat{L}_{div}) together constitute Calibrated Information Maximization (IM) Loss (\hat{L}_{im}). It's inclusion into the model is shown in Figure 1.

$$\hat{L}_{im} = \hat{L}_{ent} + \hat{L}_{div} \quad (12)$$

4.4. Combining IM with EDL

When the domain shift is large, EDL might struggle to handle uncertainty due to lack of evidence in the target domain. Combining EDL loss (L_{edl}) and IM loss (\hat{L}_{im}), as demonstrated in Figure 1, can better handle significant domain shift issues. Total training loss (L_{total}) is defined as:

$$\begin{aligned} L_{total} &= w_1 L_{edl} + w_2 \hat{L}_{im} \\ &= w_1 (L_{nll} + \beta L_{kl}) + w_2 (\hat{L}_{ent} + \hat{L}_{div}) \end{aligned} \quad (13)$$

where w_1 and w_2 are the weights given to EDL loss (L_{edl}) and information maximization loss (\hat{L}_{im}) respectively, β is a hyperparameter to balance L_{nll} and L_{kl} . In EKS and ES, the value of w_1 and w_2 is set to 0.3 and 1.0 respectively. The pseudo-code of proposed EKS is given in Algorithm 1.

5. Experiments

Datasets: *Office-Home* [29] is a challenging benchmark, which comprises four different domains: Artistic (A), ClipArt (C), Product (P), and Real-World (R). It consists of 65 classes in each domain.

Domainnet40 [22] is a large UDA benchmark. It comprises four domains Clipart (C), Painting (P), Real (R), Sketch (S),

Algorithm 1: Evidential Deep Learning pseudolabel refinement in SFDA

Input: Pre-trained source model (f_S) is available trained with labeled source data $\{x_i^s, y_i^s\}_{i=1}^{n_s}$, Unlabeled target data $\{x_i^t\}_{i=1}^{n_t}$, Prior Knowledge P_K

Output: Output: Adapted target model f_T

```

1 for mini batch  $N$  of unlabeled samples
   $\{x_i^t\}_{i=1}^{n_t} \in D_t$  do
2   Prediction of pseudolabels:  $\hat{y}_t$ 
3   Refinement with prior knowledge:  $y_{ik}$ 
4   Modify target model with dirichlet prior on
   logits as per Eq. 1
5   Replace softmax with calibrated softmax as per
   Eq. 9
6   Train with EDL loss and calibrated IM loss as
   per Eq. 13
7 end

```

with each domain consisting of 40 classes.

Office31 [25] consists of three domains, Amazon(A), Webcam (W) and DSLR (D), with 31 classes in each domain.

Digits Dataset [11] MNIST (M), SVHN (S) and USPS (U) datasets, are the standard domain adaptation benchmarks for digits classification tasks. The three adaptation tasks are: (1) the source model trained on SVHN (S) dataset is adapted to the target domain MNIST(M), (2) the source model trained on MNIST(M) dataset is adapted to the target domain USPS (U), and (3) the source model trained on USPS (U) is adapted to target MNIST (M).

Implementation Details The experiments are performed on three benchmark datasets Domainnet40 [22], Office-Home [29], Office31 [25] and Digits [11] with the same hyperparameters as in the baseline [27]. ES/EKS report the mean of the three random seed values for each task to fairly compare our results with baseline. Pretrained ResNet-50 [10] network is backbone for performing domain wise adaptation on the above benchmark datasets. A Gurobi optimizer [19] is used for solving the rectification problem using prior knowledge. EKS/ES report the average classification accuracy over all the adaptation tasks of a particular dataset, for all values of σ in unary bound (UB), where, $\sigma \in \{0.0, 0.1, 0.5, 1.0, 2.0\}$ and binary relationship (BR) where, $\sigma = 1.0$, we use $\beta = 0.5$, as the balancing factor between L_{nll} and L_{kl} . The different values of γ are 1, 1.5, and 0.2 for Domainnet40, Office-Home and Office31 respectively for softmax calibration, based on the extensive experimentation. For Digits dataset, we use $\gamma = 1$, for all the experiments.

Table 1. Classification accuracies on Domainnet40

Method	κ	σ	C→S	C→P	C→R	R→S	R→C	R→P	S→C	S→P	S→R	P→S	P→C	P→R	Avg.
SHOT [15]	-	-	75.5	74	88.4	72.8	79.4	75.4	80.5	70.8	88.3	76.2	77.7	89.8	79.1
ES	-	-	75.51	77.69	90.16	72.13	79.23	80.03	79.50	79.09	88.17	76.30	75.42	89.71	80.24
KSHOT [27]	UB	0.0	77.0	76.4	91.5	75.3	83.6	77.5	80.2	70.3	89.7	76.3	82.3	91.7	81.0
EKS	UB	0.0	78.69	82.91	91.82	77.30	83.54	84.37	81.88	81.44	88.77	78.56	83.62	92.05	83.74
KSHOT [27]	BR	-	76.4	73.7	89.1	74.3	82.1	76.8	79.1	70.2	88.8	75.9	80.6	91.7	79.9
EKS	BR	-	77.34	80.90	91.08	74.16	83.27	83.52	80.28	79.90	87.75	76.88	80.96	91.53	82.30

Table 2. Classification accuracies on Office-Home

Method	κ	σ	A→C	A→P	A→R	C→A	C→P	C→R	P→A	P→C	P→R	R→A	R→P	R→C	Avg.
SHOT [15]	-	-	57.1	78.1	81.5	68.0	78.2	78.1	67.4	54.9	82.2	73.3	84.3	58.8	71.8
ES	-	-	56.81	78.64	82.05	68.18	77.82	79.07	68.11	54.01	82.02	73.20	84.06	58.26	71.85
KSHOT [27]	UB	0.0	58.2	80.0	82.9	71.1	80.3	80.7	71.3	56.8	83.2	75.5	86.6	60.3	73.9
EKS	UB	0.0	58.95	79.49	82.86	70.62	80.22	81.48	71.80	57.21	83.76	75.94	86.30	62.27	74.24
KSHOT [27]	BR	-	57.4	78.8	82.9	70.7	80	80.5	70.8	55.0	82.8	74.6	86.0	59.9	73.3
EKS	BR	-	58.46	78.99	83.27	71.41	80.33	80.84	71.10	56.74	83.39	75.48	86.33	60.90	73.94

6. Results and Analysis

6.1. Results

Tables 1, 2 and 4 lists results of **ES** and **EKS** on three benchmarks datasets Domainnet40, Office-Home and Office31. **EKS** and **ES** consistently achieve better classification accuracies on the target domain and also improved adaptation accuracy over KSHOT [27] and SHOT [15], respectively. It has been observed that the classification accuracy of **EKS** is significantly higher than **ES**, thereby highlighting the impact of prior knowledge even with evidential modeling. There is significant improvement on the tasks where domain gap is quite large. **EKS/ES** show consistent improvement in performance on both types of prior knowledge Unary Bound (UB) and Binary relationship (BR).

Results on Domainnet40: As listed in Table 1, for **UB** ($\sigma = 0.0$), **EKS** improves the performance by **+2.74 %** and for **BR**, it improves the performance by **+2.4 %** in comparison to baseline [27]. On the adaptation task (C→P), We observe a significant improvement of approximately **+6%**. **ES** achieves the improvement in performance by **+1.14%** in comparison to baseline SHOT [15].

Results on Office-Home: As listed in Table 2, for **UB** ($\sigma = 0.0$), **EKS** improves the performance by **+0.34%** and for **BR**, it improves the performance by **+0.64%** in comparison to baseline [27]. However, average class improvement is less, it is showing considerable improvement on certain domains as evident from Table 2. **ES** achieves the improvement in performance by **+0.05%** in comparison to baseline [15].

Results on Office31: As listed in Table 4, on Office31 dataset, **EKS** improves the performance of some challenging adaptation task like for A→D, by **+0.5%**. **ES** achieves the improvement in performance on A→W by **+1.35%**, on A→D by **+0.11%** and on W→A by **+1.09%**.

Results on Digits Dataset: As listed in Table 5, on Digits dataset, **EKS** improves the performance of some challenging adaptation task like for SVHN→MNIST, on various values of σ , such as on UB(0.0) by **+0.43%**, and for **BR**, it improves the performance by **+0.32%**. While adapting the model trained on USPS(U) to the target domain(M), consistent improvement in performance has been achieved through **EKS** over all values of σ .

Result of E-TransDA on Office31: We leverage our methodology into other SFDA technique (TransDA) and named it as E-TransDA. The result for the same is reported in the Table 6 on Office-31 dataset. We observe a consistent improvement in accuracy across all domains and in average adaptation accuracy as well, thus validating the generalizability of our approach.

6.2. Analysis

6.2.1 Calibration Analysis

An empirical analysis that quantifies how well the predicted confidence levels correspond to actual outcomes across the entire model, specifically, Expected Calibration Error (ECE) and Negative log likelihood (NLL) for **EKS** is evaluated. Table 3 shows the results of the ECE and NLL on Domainnet40 dataset.

6.2.2 Effect of combining IM with EDL

The analysis of the effectiveness of combining information maximization with EDL is done using the Office-Home dataset. Table 7 validate the result of the proposed framework. A comparison of classification accuracies is presented between the framework that combines EDL with IM (L_{total}) and the framework that solely uses EDL loss (L_{edl}), excluding IM loss from the training process.

Table 3. ECE and NLL on Domainnet40

Metric	Method	κ	C→S	C→P	C→R	R→S	R→C	R→P	S→C	S→P	S→R	P→S	P→C	P→R	Avg.
ECE	KSHOT [27]	BR	0.153	0.116	0.055	0.149	0.065	0.092	0.104	0.141	0.085	0.148	0.083	0.055	0.104
	EKS	BR	0.116	0.089	0.045	0.130	0.044	0.071	0.083	0.106	0.063	0.121	0.062	0.044	0.081
NLL	KSHOT [27]	BR	1.338	1.080	0.495	1.325	0.741	0.881	1.036	1.311	0.753	1.316	0.935	0.477	0.974
	EKS	BR	1.194	0.972	0.460	1.249	0.734	0.832	0.989	1.108	0.594	1.214	0.892	0.439	0.890

Table 4. Classification accuracies on OFFICE-31.

Method	κ	σ	A→W	A→D	W→A	W→D	D→A	D→W	Avg.
SHOT [15]	-	-	90.1	94.0	74.3	99.9	74.7	98.4	88.6
ES	-	-	91.45	94.11	75.39	99.27	74.47	97.36	88.67
KSHOT [27]	UB	0.0	98.5	97.6	76.2	99.8	75.0	99.0	91.0
EKS	UB	0.0	96.4	98.1	75.1	99.8	74.5	99.0	90.5
KSHOT [27]	BR	-	97.1	96.9	76.1	99.8	74.0	98.8	90.5
EKS	BR	-	96.5	97.1	75.9	99.8	74.3	98.7	90.4

Table 5. Classification accuracies on Digits dataset

Method	κ	σ	SVHN→MNIST	MNIST→USPS	USPS→MNIST
KSHOT [27]	UB	0.0	94.79	94.45	91.26
	UB	0.1	93.67	94.25	90.88
	UB	0.5	93.46	94.27	89.56
	UB	1.0	93.46	94.22	89.49
	UB	2.0	93.46	94.22	89.49
	BR	-	94.33	94.29	91.15
EKS	UB	0.0	95.22	94.60	91.78
	UB	0.1	93.82	94.59	91.25
	UB	0.5	93.36	94.40	89.80
	UB	1.0	93.36	94.47	89.77
	UB	2.0	93.36	94.50	89.77
	BR	-	94.65	94.57	91.40

Table 6. Classification accuracies on Office31 (E-TransDA)

Method	A→D	A→W	D→A	D→W	W→A	W→D	Avg.
TransDA [33]	97.20	95.00	73.70	99.30	79.30	99.60	90.68
E-TransDA (Ours)	97.59	96.48	74.51	98.87	80.09	100.00	91.26

EDL enhances the reliability of pseudo-labels by making confident predictions. In cases where there is a lack of evidence, EDL might lead the model to make overconfident predictions on noisy pseudo-labels, as it pushes the model to minimize entropy, which hampers the performance of the adapted model. IM complements EDL by its L_{div} component as it encourages diversified predictions among classes ensuring that the model does not become overconfident for a few classes. There is significant improvement in classification accuracies for various values of σ when combining EDL with IM, i.e., improvement of +1.94% on UB(0.0), +1.91% on UB(0.1) etc. and for BR, the improvement in performance by +2.32%. Thus, combination of EDL with IM achieve maximum increment in accuracy. Table 3 shows the results of the ECE and NLL on Domainnet40.

6.2.3 Integrating calibrated softmax function with IM

Table 8 validates the effectiveness of integrating the calibrated softmax function Eq. (9) instead of the standard softmax function Eq. (8). It reports comparison of classification accuracies on Office-Home dataset between standard softmax integration with IM and calibrated softmax integration

Table 7. Ablation study of IM loss on Office-Home dataset

κ	σ	EKS with IM	EKS w/o IM
UB	0.0	73.93	71.99
UB	0.1	73.71	71.80
UB	0.5	72.76	70.64
UB	1.0	72.19	70.00
UB	2.0	71.92	69.71
BR	-	73.62	71.30

Table 8. Ablation study of calibrated softmax based IM loss on Office Home

κ	σ	EKS(Standard Softmax)	EKS (calibrated softmax)
UB	0.0	73.93	74.24
UB	0.1	73.71	73.98
UB	0.5	72.76	72.72
UB	1.0	72.19	72.01
UB	2.0	71.92	71.66
BR	-	73.62	73.94

with IM. Calibrated softmax function mitigates the issue of translation invariance faced by the standard softmax function and predicts the probabilities by keeping in consideration the magnitude of the logits but the standard softmax only predicts the probabilities based on the relative relationships of the logits and does not consider the effect of the magnitude of the logits in the predictions. As reported in Table 8, there is consistent improvement in classification accuracies such as by +0.31% for UB(0.0) and by +0.32% for BR on Office-Home dataset with integration of the calibrated softmax function with the information maximization.

6.2.4 Effect of proposed EKS on various values of (σ)

As the value of σ grows from 0.1 to 2.0, the certainty of prior knowledge gets diluted more and more. Thus, with less knowledge, it becomes a challenge to improve the adaptation. Our proposed framework also performs better in these cases. The maximum increment in accuracy is at UB 0.1. The ablation study on classification accuracy at different value of bounds (σ) for Domainnet40, Office-Home and Office31 are given in Table 9, 10, 11 respectively.

6.2.5 Feature and Calibration Visualization

Figure 2a and 2b show the t-SNE visualization of features learned by the model when trained with KSHOT and EKS respectively, on task (P→C), while Figure 2c and 2d show the t-SNE visualization of features learned by the model when trained with KSHOT and EKS respectively on adaptation task (S→C). It is observed that adaptation of features

Table 9. Ablation study of classification accuracies on Office-Home where $\sigma \in \{0.1, 0.5, 1.0, 2.0\}$

Method	κ	σ	A→C	A→P	A→R	C→A	C→P	C→R	P→A	P→C	P→R	R→A	R→P	R→C	Avg.
KSHOT [27]	UB	0.1	58.1	79.2	83.2	70.4	80.0	80.7	71.4	56.5	83.0	75.6	86.0	60.8	73.7
	UB	0.5	57.4	79.1	82.1	69.4	78.1	79.5	69.3	55.2	81.8	74.0	85.1	60.2	72.6
	UB	1.0	57	79	82.1	68.6	77.8	79.3	68.4	55.1	81.7	73.5	84.8	59.3	72.2
	UB	2.0	56.4	78.7	82.1	68.3	77.8	79.3	67.9	54.2	81.7	73.3	84.8	58.7	71.9
EKS	UB	0.1	58.93	78.48	83.17	70.59	79.42	81.29	71.58	57.74	83.44	75.59	85.43	62.13	73.98
	UB	0.5	58.10	78.73	81.98	68.83	78.31	79.34	69.26	56.26	81.42	74.38	84.78	61.29	72.72
	UB	1.0	57.00	78.47	82.04	68.29	78.12	78.42	68.43	54.52	80.84	74.10	84.04	59.87	72.01
	UB	2.0	56.43	78.06	82.04	67.89	77.95	78.44	67.92	53.48	80.84	73.93	84.07	58.83	71.66

Table 10. Ablation study of classification accuracies on Domainnet40 where $\sigma \in \{0.0, 0.1, 0.5, 1.0, 2.0\}$

Method	κ	σ	C→S	C→P	C→R	R→S	R→C	R→P	S→C	S→P	S→R	P→S	P→C	P→R	Avg.
KSHOT [27]	UB	0.1	76.9	76.8	89.5	75.2	82.2	77.6	79.4	70.1	88.5	76.9	81.7	91.2	80.5
	UB	0.5	75.6	75.4	88.8	73.5	80.3	77.2	78.9	70.4	88.3	76.6	78.4	89.0	79.4
	UB	1.0	75.3	73.7	88.8	72.9	79.9	76.8	80.0	69.9	88	76.4	77.6	88.6	79.0
	UB	2.0	75.5	75.4	88.8	73.1	79.2	76.3	80.1	71.1	87.9	76.2	77.8	88.6	79.2
EKS	UB	0.1	78.88	82.49	91.34	76.78	83.69	83.97	81.06	81.42	88.03	78.17	82.36	91.65	83.32
	UB	0.5	76.31	79.55	90.62	73.84	81.87	83.19	80.03	80.84	87.81	77.20	77.84	90.66	81.65
	UB	1.0	75.60	79.04	90.49	73.23	79.60	82.86	79.72	80.44	87.46	76.73	75.60	90.26	80.92
	UB	2.0	75.28	78.79	90.47	73.02	79.19	82.39	79.56	79.55	87.44	76.35	75.43	90.27	80.64

Table 11. Ablation study of classification accuracies on Office-31 where, $\sigma \in \{0.0, 0.1, 0.5, 1.0, 2.0\}$

Method	κ	σ	A→W	A→D	W→A	W→D	D→A	D→W	Avg.
KSHOT [27]	UB	0.1	97.2	96.7	76.5	99.8	75.5	98.7	90.7
	UB	0.5	92.8	93.9	75.2	99.7	75.7	97.7	89.2
	UB	1.0	92.4	93.7	75.5	99.7	75.5	97.7	89.1
	UB	2.0	92.4	93.7	75.2	99.7	75.0	97.7	89.0
EKS	UB	0.1	96.4	96.6	75.8	99.8	75.3	98.3	90.4
	UB	0.5	93.2	94.5	75.0	99.5	75.3	97.4	89.2
	UB	1.0	91.4	94.4	74.8	99.5	74.6	97.4	88.7
	UB	2.0	91.3	94.4	74.8	99.5	74.7	97.4	88.7

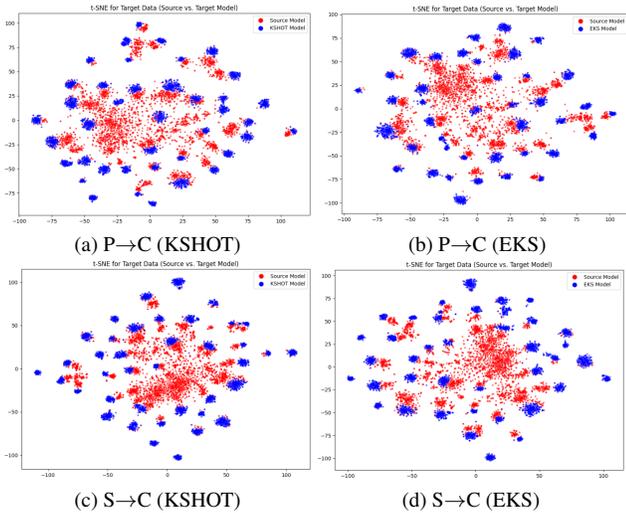


Figure 2. t-SNE plots for P→C and S→C on Domainnet40.

achieved through EKS is better than baseline KSHOT [27].

Calibration curve: Figure 3 shows the ECE curve for visualizing calibration, where the dashed line ($x = y$) represents ideal calibration. A curve closer to this line indicates better calibration. In Figure 3b, the EKS curve is closer to the dashed line compared to KSHOT, while Figure 3a shows

the ES curve is closer to the dashed line than SHOT, thus, validating our approach.

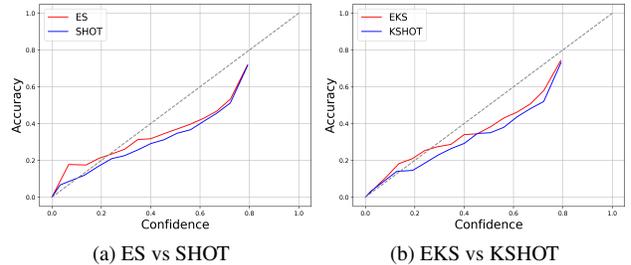


Figure 3. Expected Calibration Error (ECE) curve on the Domainnet40 dataset.

7. Conclusion

In this paper, SFDA has been addressed in two different settings: EKS (with prior knowledge) and ES (w/o prior knowledge). An evidence-based model training is implemented, assigning an input sample to the class with the strongest evidence. The effect of using a calibrated softmax function for predicting probability and its integration into information maximization has been explored. Extensive experiments prove that combining EDL loss with information maximization loss helps the model adapt better to the target domain.

Acknowledgment: Shivangi acknowledges the support of Visvesvaraya PhD fellowship by Ministry of Electronics and Information Technology (MeitY), Government of India. Rini acknowledges the financial support provided by the Department of Science and Technology, Government of India, through the WISE Post-Doctoral Fellowship Programme (Reference No. DST/WISE-PDF/ET-33/2023).

References

- [1] Eduardo Aguilar, Bogdan Raducanu, Petia Radeva, and Joost Van de Weijer. Continual evidential deep learning for out-of-distribution detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3444–3454, 2023. 2
- [2] Alexander Amini, Wilko Schwarting, Ava Soleimany, and Daniela Rus. Deep evidential regression. *Advances in neural information processing systems*, 33:14927–14937, 2020. 1, 2
- [3] Awais Ashfaq, Markus Lingman, Murat Sensoy, and Sławomir Nowaczyk. Deed: Deep evidential doctor. *Artificial Intelligence*, 325:104019, 2023. 2
- [4] Wentao Bao, Qi Yu, and Yu Kong. Evidential deep learning for open set action recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13349–13358, 2021. 2
- [5] Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. Deep clustering for unsupervised learning of visual features. In *Proceedings of the European conference on computer vision (ECCV)*, pages 132–149, 2018. 2
- [6] Cheng Chen, Quande Liu, Yueming Jin, Qi Dou, and Pheng-Ann Heng. Source-free domain adaptive fundus image segmentation with denoised pseudo-labeling. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24*, pages 225–235. Springer, 2021. 1, 2
- [7] Jiayi Chen, Benteng Ma, Hengfei Cui, and Yong Xia. Think twice before selection: Federated evidential active learning for medical image analysis with domain shifts. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11439–11449, 2024. 2
- [8] Danruo Deng, Guangyong Chen, Yang Yu, Furui Liu, and Pheng-Ann Heng. Uncertainty estimation by fisher information-based evidential deep learning. In *International Conference on Machine Learning*, pages 7596–7616. PMLR, 2023. 1
- [9] Ligong Han, Yang Zou, Ruijiang Gao, Lezi Wang, and Dimitris Metaxas. Unsupervised domain adaptation via calibrating uncertainties. In *CVPR Workshops*, volume 9, 2019. 1
- [10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 5
- [11] Judy Hoffman, Eric Tzeng, Taesung Park, Jun-Yan Zhu, Phillip Isola, Kate Saenko, Alexei Efros, and Trevor Darrell. Cycada: Cycle-consistent adversarial domain adaptation. In *International conference on machine learning*, pages 1989–1998. Pmlr, 2018. 5
- [12] Vinod K. Kurmi, Venkatesh K. Subramanian, and Vinay P. Namboodiri. Domain impression: A source data free domain adaptation method. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 615–625, January 2021. 1
- [13] JoonHo Lee and Gyemin Lee. Unsupervised domain adaptation based on the predictive uncertainty of models. *Neurocomputing*, 520:183–193, 2023. 1
- [14] Jingjing Li, Zhiqi Yu, Zhekai Du, Lei Zhu, and Heng Tao Shen. A comprehensive survey on source-free domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 1
- [15] Jian Liang, Dapeng Hu, and Jiashi Feng. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In *International conference on machine learning*, pages 6028–6039. PMLR, 2020. 2, 6, 7
- [16] Mattia Litrico, Alessio Del Bue, and Pietro Morerio. Guiding pseudo-labels with uncertainty estimation for source-free unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7640–7650, 2023. 2
- [17] Xiaofeng Liu, Chaehwa Yoo, Fangxu Xing, Hyejin Oh, Georges El Fakhri, Je-Won Kang, Jonghye Woo, et al. Deep unsupervised domain adaptation: A review of recent advances and perspectives. *APSIPA Transactions on Signal and Information Processing*, 11(1), 2022. 1
- [18] Zhihe Lu, Da Li, Yi-Zhe Song, Tao Xiang, and Timothy M Hospedales. Uncertainty-aware source-free domain adaptive semantic segmentation. *IEEE Transactions on Image Processing*, 2023. 2
- [19] Calvin Nau, Prashant Sankaran, and Katie McConky. Comparison of parameter tuning strategies for team orienteering problem (top) solved with gurobi. In *IIE Annual Conference. Proceedings*, pages 1–6. Institute of Industrial and Systems Engineers (IISE), 2022. 5
- [20] Deep Shankar Pandey and Qi Yu. Learn to accumulate evidence from all training samples: theory and practice. In *International Conference on Machine Learning*, pages 26963–26989. PMLR, 2023. 2
- [21] Younghyun Park, Wonjeong Choi, Soyeong Kim, Dong-Jun Han, and Jaekyun Moon. Active learning for object detection with evidential deep learning and hierarchical uncertainty aggregation. In *The Eleventh International Conference on Learning Representations*, 2023. 2
- [22] 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. 5
- [23] Dripta S Raychaudhuri, Calvin-Khang Ta, Arindam Dutta, Rohit Lal, and Amit K Roy-Chowdhury. Prior-guided source-free domain adaptation for human pose estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14996–15006, 2023. 2
- [24] Subhankar Roy, Martin Trapp, Andrea Pilzer, Juho Kannala, Nicu Sebe, Elisa Ricci, and Arno Solin. Uncertainty-guided source-free domain adaptation. In *European conference on computer vision*, pages 537–555. Springer, 2022. 2
- [25] Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell. Adapting visual category models to new domains. In *Computer Vision—ECCV 2010: 11th European Conference on*

Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part IV 11, pages 213–226. Springer, 2010. [5](#)

- [26] Murat Sensoy, Lance Kaplan, and Melih Kandemir. Evidential deep learning to quantify classification uncertainty. *Advances in neural information processing systems*, 31, 2018. [1](#), [2](#)
- [27] Tao Sun, Cheng Lu, and Haibin Ling. Prior knowledge guided unsupervised domain adaptation. In *European conference on computer vision*, pages 639–655. Springer, 2022. [2](#), [3](#), [5](#), [6](#), [7](#), [8](#)
- [28] Dennis Ulmer, Christian Hardmeier, and Jes Frellsen. Prior and posterior networks: A survey on evidential deep learning methods for uncertainty estimation. *arXiv preprint arXiv:2110.03051*, 2021. [2](#)
- [29] 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. [5](#)
- [30] Jianghao Wu, Guotai Wang, Ran Gu, Tao Lu, Yinan Chen, Wentao Zhu, Tom Vercauteren, Sébastien Ourselin, and Shaoting Zhang. Upl-sfda: Uncertainty-aware pseudo label guided source-free domain adaptation for medical image segmentation. *IEEE Transactions on Medical Imaging*, 42(12):3932–3943, 2023. [2](#)
- [31] Mixue Xie, Shuang Li, Rui Zhang, and Chi Harold Liu. Dirichlet-based uncertainty calibration for active domain adaptation. *arXiv preprint arXiv:2302.13824*, 2023. [1](#)
- [32] Shaoxun Xu, Yufei Chen, Chao Ma, and Xiaodong Yue. Deep evidential fusion network for medical image classification. *International Journal of Approximate Reasoning*, 150:188–198, 2022. [2](#)
- [33] Guanglei Yang, Hao Tang, Zhun Zhong, Mingli Ding, Ling Shao, Nicu Sebe, and Elisa Ricci. Transformer-based source-free domain adaptation. *arXiv:2105.14138*, 2021. [7](#)
- [34] Chen-Chen Zong, Ye-Wen Wang, Ming-Kun Xie, and Sheng-Jun Huang. Dirichlet-based prediction calibration for learning with noisy labels. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 17254–17262, 2024. [2](#)