

# Evidential Uncertainty Quantification: A Variance-Based Perspective

Ruxiao Duan<sup>1</sup>, Brian Caffo<sup>1</sup>, Harrison X. Bai<sup>2</sup>, Haris I. Sair<sup>2</sup>, Craig Jones<sup>1</sup>

<sup>1</sup>Johns Hopkins University

<sup>2</sup>Johns Hopkins University School of Medicine

{rduan6, bcaffo1, hbai7, craigj}@jhu.edu, hsair1@jhmi.edu

## Abstract

*Uncertainty quantification of deep neural networks has become an active field of research and plays a crucial role in various downstream tasks such as active learning. Recent advances in evidential deep learning shed light on the direct quantification of aleatoric and epistemic uncertainties with a single forward pass of the model. Most traditional approaches adopt an entropy-based method to derive evidential uncertainty in classification, quantifying uncertainty at the sample level. However, the variance-based method that has been widely applied in regression problems is seldom used in the classification setting. In this work, we adapt the variance-based approach from regression to classification, quantifying classification uncertainty at the class level. The variance decomposition technique in regression is extended to class covariance decomposition in classification based on the law of total covariance, and the class correlation is also derived from the covariance. Experiments on cross-domain datasets are conducted to illustrate that the variance-based approach not only results in similar accuracy as the entropy-based one in active domain adaptation but also brings information about class-wise uncertainties as well as between-class correlations. The code is available at <https://github.com/KerryDRX/EvidentialADA>. This alternative means of evidential uncertainty quantification will give researchers more options when class uncertainties and correlations are important in their applications.*

## 1. Introduction

Recent decades have witnessed significant advances in deep learning, which has been employed in various domains including computer vision and natural language processing. Despite the great success of Deep Neural Networks (DNNs), their application in safety-critical tasks is still highly restricted because of their lack of transparency [30], vulnerability to domain shifts [22] and adversarial attacks [28, 33, 35], and incapability of reliably calibrating prediction uncertainties [4]. Therefore, uncertainty quan-

tification of DNNs has attracted great attention more recently, and many approaches [6, 11, 17, 21, 27, 38, 41, 46] have been devised to estimate predictive uncertainties.

Evidential Deep Learning (EDL) [32] brings insights into the direct estimation of model prediction uncertainties with a single forward propagation, and fundamental theories have been established for both classification [21, 32] and regression [1]. Learning is formulated as a process of evidence acquisition, and each training sample adds support to the higher-order evidential distribution [1]. Compared to many other alternatives of uncertainty quantification methods, EDL does not modify model architecture or incur additional computation [12]. In addition, EDL is also able to disentangle uncertainties from different origins: uncertainty due to the conflict of evidence and uncertainty due to the lack of evidence. The former is known as data uncertainty or aleatoric uncertainty (AU), which refers to the intrinsic randomness of the data as a result of natural data complexity and cannot be decreased by training with more data. The latter is known as model uncertainty or epistemic uncertainty (EU), which originates from the lack of knowledge about the data mainly due to data distribution mismatch and can be reduced by increasing training data.

The approach to separate total uncertainty into AU and EU varies in regression and classification. In regression problems [1, 36, 40], the total uncertainty is quantified as predictive variance, which can be decomposed based on the law of total variance [19]. In classification problems [21, 23, 43], the total uncertainty is quantified as predictive entropy, which can be decomposed based on the definition of mutual information [3]. These two types of methods are respectively referred to as variance-based approach and entropy-based approach in this paper.

Although the entropy-based approach for evidential uncertainty quantification has demonstrated promising results [18, 23, 43], the variance-based method is typically not considered in classification settings. Since variance is quantified for a single variable only and is not directly applicable to a probability distribution, entropy has been regarded as the better option to quantify the total uncertainty of a single

classification sample. Consequently, classification uncertainty quantification has been limited to the sample level. However, the variance-based uncertainty quantification introduced in this paper can be applied to each class individually, estimating uncertainty at the class level.

Active domain adaptation (ADA), the combination of active learning (AL) and domain adaptation (DA), is selected as the objective task to evaluate the variance-based uncertainty quantification approach in this work. DNNs are known to suffer from insufficient annotated data and domain shifts, thus AL and DA have been intensively studied in recent years to alleviate these two problems. AL identifies valuable data from the unlabeled pool to annotate, giving the model the best performance with the fewest annotations [29], but a sampling heuristic for data selection is required, where uncertainty quantification can play an important role. DA mitigates performance degradation when the model is trained on data from one distribution (source domain) but evaluated on data from another distribution (target domain), and a recent study has shown the effectiveness of uncertainty-guided training in unsupervised DA [43]. Therefore, the quality of uncertainty quantification can be directly reflected by the performance of ADA. Dirichlet-based Uncertainty Calibration (DUC) [43], an EDL-based ADA framework, has been proposed and achieved state-of-the-art (SOTA) performance on multiple ADA datasets.

**Contributions:** 1) While the previous entropy-based approach quantifies EDL uncertainties at the sample level only, we derive a variance-based approach to achieve sample-level and class-level uncertainty quantification. 2) We introduce an EDL class covariance calculation method, based on which between-class correlations can be obtained and highly similar class pairs can be identified. 3) We devise a simultaneous certainty and uncertainty sampling strategy in active learning to boost the model performance under large domain shift. 4) Experiments on cross-domain image classification datasets are conducted to illustrate the effectiveness of our variance-based approach in ADA, with SOTA performance. Class-level uncertainty quantification and class correlation analysis are also demonstrated.

## 2. Related Work

**Evidential deep learning** [1, 21, 32] treats model learning as an evidence-acquisition process, in which each training sample introduces new evidence to a learned higher-order evidential distribution. The evidential model predicts the parameters of the evidential distribution, then aleatoric and epistemic uncertainties can be directly derived from the distribution parameters. EDL has been applied in various fields such as active learning [14], domain adaptation [43], semi-supervised learning [45], open set action recognition [5], reinforcement learning [7], object detection [23], depth estimation [1], stereo matching [40], molecular property

prediction [36], and medical image segmentation [18].

As for uncertainty quantification in EDL, typically the entropy-based approach is adopted in classification [21, 23, 43, 43, 45], while the variance-based approach is used in regression [1, 36, 40]. A recent work [15] explores the possibility of applying the variance-based approach in a brain segmentation problem, but the quality of uncertainty is not being tested, thus the reliability of this approach in downstream tasks, such as active learning which requires precise uncertainty estimation, remains unclear.

**Active domain adaptation** is a particular AL problem in which labeled and unlabeled data are drawn from different underlying distributions, thus the representativeness of the target domain (*i.e.*, targetness) is a crucial factor that has to be considered compared to general active learning settings. AADA [37] and TQS [10] employ a domain discriminator to quantify targetness, but the learning of the discriminator is not related to the classifier, leading to the selection of less useful samples. CLUE [26] and DBAL [8] measure targetness by clustering, but redundant sampling becomes a concern as the interrelationship among unlabeled target data is not taken into account. The domain discriminator and clustering applied in these approaches also lead to extra computational expenses. EADA [42] and SDM-AG [44] attempt to mitigate domain shift by picking samples based on an auxiliary loss function, but uncertainty quantification is still based on point estimate predictions, which are error-prone on the target dataset. DUC [43] leverages evidential uncertainties for model training and sample selection, achieving SOTA performance on multiple ADA datasets.

## 3. Evidential Uncertainty Quantification

### 3.1. Previous Work

**Deep learning classification.** In a traditional  $C$ -class classification problem, let  $\mathbf{y}$  denote the one-hot vector encoding the ground truth class  $c_g \in \{1, 2, \dots, C\}$  of input data  $\mathbf{x}$ , with  $y_{c_g} = 1$  and  $y_c = 0$  for all  $c \neq c_g$ . A DNN uses softmax activation to convert the prediction logits into a class probability vector  $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_C]^\top$ , in which  $\mu_c$  is the probability that input  $\mathbf{x}$  belongs to class  $c$ ,  $\mu_c \in [0, 1]$  for all  $c$ , and  $\sum_{c=1}^C \mu_c = 1$ . Hence,

$$\mathbf{y} \sim \text{Cat}(\boldsymbol{\mu}) \quad (1)$$

where  $\text{Cat}(\cdot)$  represents the categorical distribution.

**Evidential deep learning classification.** EDL [21, 32] considers the class probability vector  $\boldsymbol{\mu}$  as a random vector that follows a Dirichlet (Dir) distribution, *i.e.*,

$$\boldsymbol{\mu} \sim \text{Dir}(\boldsymbol{\alpha}) \quad (2)$$

in which  $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_C]^\top$  is a vector of Dirichlet parameters with  $\alpha_c > 0$  for all  $c$ . The probability density

function of Dirichlet distribution is given by

$$D(\boldsymbol{\mu}|\boldsymbol{\alpha}) = \frac{\Gamma(\alpha_0)}{\prod_{c=1}^C \Gamma(\alpha_c)} \prod_{c=1}^C \mu_c^{\alpha_c-1} \quad (3)$$

in which  $\Gamma(\cdot)$  denotes *Gamma* function and  $\alpha_0 = \sum_{c=1}^C \alpha_c$  is the Dirichlet strength. Following the Dirichlet distribution, the expected probability vector is  $\bar{\boldsymbol{\mu}} := \mathbb{E}[\boldsymbol{\mu}] = \frac{\boldsymbol{\alpha}}{\alpha_0}$ , and the expected probability of class  $c$  is  $\bar{\mu}_c := \mathbb{E}[\mu_c] = \frac{\alpha_c}{\alpha_0}$ .

**Evidential neural network.** Unlike a DNN which predicts class probabilities  $\boldsymbol{\mu}$ , an evidential neural network (ENN) predicts Dirichlet parameters  $\boldsymbol{\alpha}$  instead. Therefore, to construct an ENN, it suffices to replace the softmax activation of a DNN with some function that outputs positive values only (*e.g.*, exponential function) to accommodate the constraint of  $\boldsymbol{\alpha}$ , and the internal model architecture can be preserved. The final predicted class  $c_p$  of an ENN is the class with maximum expected probability, *i.e.*,

$$c_p = \arg \max_c \bar{\mu}_c = \arg \max_c \alpha_c \quad (4)$$

**Evidential model training.** An ENN can be optimized via the loss function  $\mathcal{L}^{EDL}$ , which consists of a negative log likelihood term  $\mathcal{L}^{NLL}$  and a KL-divergence term  $\mathcal{L}^{KL}$  for regularization, with  $\lambda_{reg}$  being the regularization coefficient.

$$\mathcal{L}^{EDL} = \mathcal{L}^{NLL} + \lambda_{reg} \mathcal{L}^{KL} \quad (5)$$

The final expressions of  $\mathcal{L}^{NLL}$  and  $\mathcal{L}^{KL}$  are respectively

$$\mathcal{L}^{NLL} = \sum_{c=1}^C y_c (\log(\alpha_0) - \log(\alpha_c)) \quad (6)$$

$$\begin{aligned} \mathcal{L}^{KL} = & \log \left( \frac{\Gamma(\sum_{c=1}^C \tilde{\alpha}_c)}{\Gamma(C) \prod_{c=1}^C \Gamma(\tilde{\alpha}_c)} \right) \\ & + \sum_{c=1}^C (\tilde{\alpha}_c - 1) \left( \psi(\tilde{\alpha}_c) - \psi \left( \sum_{c=1}^C \tilde{\alpha}_c \right) \right) \end{aligned} \quad (7)$$

in which  $\tilde{\alpha}_c = y_c + (1 - y_c)\alpha_c$  and  $\psi(\cdot)$  is *digamma* function. The derivations are detailed in [32].

**Entropy-based classification uncertainty.** Previous approaches [18,23,43] adopt an entropy-based method to estimate classification uncertainty in EDL, quantifying the total uncertainty  $U$  of a sample  $\boldsymbol{x}$  as the Shannon entropy of the expected class probabilities of  $\boldsymbol{x}$ . An information theory approach is employed to decompose  $U$  into aleatoric uncertainty  $U^{alea}$  and epistemic uncertainty  $U^{epis}$ , *i.e.*,  $U = U^{alea} + U^{epis}$ . The final expressions of the entropy-based

uncertainties are

$$U = - \sum_{c=1}^C \bar{\mu}_c \log \bar{\mu}_c \quad (8)$$

$$U^{alea} = \sum_{c=1}^C \bar{\mu}_c (\psi(\alpha_0 + 1) - \psi(\alpha_c + 1)) \quad (9)$$

$$U^{epis} = - \sum_{c=1}^C \bar{\mu}_c (\log \bar{\mu}_c + \psi(\alpha_0 + 1) - \psi(\alpha_c + 1)) \quad (10)$$

The derivations can be found in [21].

**Variance-based regression uncertainty.** Deep evidential regression [1] has been proposed for regression uncertainty quantification. Formally, it is assumed that the prediction target  $y$  is a continuous random variable following a Gaussian distribution, whose mean and variance jointly follow a Normal Inverse-Gamma (NIG) distribution:

$$y \sim \mathcal{N}(\mu, \sigma^2) \quad (11)$$

$$\mu, \sigma^2 \sim \text{NIG}(\gamma, \nu, \alpha, \beta) \quad (12)$$

in which  $\gamma \in \mathbb{R}, \nu > 0, \alpha > 1, \beta > 0$ . Thus a regression ENN predicts the NIG distribution parameters.

In regression, the total uncertainty  $U$  is quantified as the variance of prediction target  $y$ . Based on the law of total variance [19], the total uncertainty can be decomposed into aleatoric and epistemic components:

$$\underbrace{\text{Var}[y]}_{\text{total}} = \underbrace{\mathbb{E}[\text{Var}[y|\mu, \sigma^2]]}_{\text{aleatoric}} + \underbrace{\text{Var}[\mathbb{E}[y|\mu, \sigma^2]]}_{\text{epistemic}} \quad (13)$$

According to [1], the regression uncertainties are derived as

$$U = \frac{(\nu + 1)\beta}{\nu(\alpha - 1)} \quad (14)$$

$$U^{alea} = \frac{\beta}{\alpha - 1} \quad (15)$$

$$U^{epis} = \frac{\beta}{\nu(\alpha - 1)} \quad (16)$$

### 3.2. Variance-based Classification Uncertainty

**Motivation.** A weakness of the traditional entropy-based approach is that classification uncertainties  $U$ ,  $U^{alea}$ , and  $U^{epis}$  can only be quantified up to the sample level (Eq. (8), Eq. (9), Eq. (10)). That is to say, we can at best derive uncertainties associated with an entire single sample  $\boldsymbol{x}$ . Imagine a scenario in which the predictive uncertainty of a vehicle image  $\boldsymbol{x}$  is high because the model is not able to distinguish between two classes on this image: automobile and truck. For most other categories, such as cat and dog, the model is certain that the image does not belong to any of them, thus those classes do not contribute much to the total

uncertainty. The entropy-based approach does not provide any information about these class-level uncertainties, which can sometimes be crucial in downstream applications.

**From regression to classification.** In EDL, classification and regression problems share the same underlying assumption: the prediction target is supposed to follow a bi-level probability distribution (Eq. (1) and Eq. (2) for classification; Eq. (11) and Eq. (12) for regression). Therefore, the uncertainty quantification and decomposition approach in regression (Eq. (13)) should also be applicable in classification. However, regression has a scalar target  $y$  while classification has a vector target  $\mathbf{y}$ . Since variance is not directly calculable for a multidimensional vector, we extend the variance in Eq. (13) to covariance to adapt to the multidimensionality of  $\mathbf{y}$ .

**Covariance quantification and decomposition.** The covariance matrix of a classification target label  $\mathbf{y}$  is defined as

$$\text{Cov}[\mathbf{y}] := \mathbb{E}[(\mathbf{y} - \mathbb{E}[\mathbf{y}])(\mathbf{y} - \mathbb{E}[\mathbf{y}])^\top] \quad (17)$$

which can be decomposed into aleatoric and epistemic components based on the law of total covariance [19]:

$$\underbrace{\text{Cov}[\mathbf{y}]}_{\text{total}} = \underbrace{\mathbb{E}[\text{Cov}[\mathbf{y}|\boldsymbol{\mu}]]}_{\text{aleatoric}} + \underbrace{\text{Cov}[\mathbb{E}[\mathbf{y}|\boldsymbol{\mu}]]}_{\text{epistemic}} \quad (18)$$

We define the aleatoric and epistemic components of the covariance matrix  $\text{Cov}[\mathbf{y}]$  as aleatoric covariance matrix  $\text{Cov}[\mathbf{y}]^{\text{alea}}$  and epistemic covariance matrix  $\text{Cov}[\mathbf{y}]^{\text{epis}}$ . Following the distribution assumptions Eq. (1) and Eq. (2), the covariance matrices can be derived as follows:

$$\text{Cov}[\mathbf{y}] = \text{Diag}(\bar{\boldsymbol{\mu}}) - \bar{\boldsymbol{\mu}}\bar{\boldsymbol{\mu}}^\top \quad (19)$$

$$\text{Cov}[\mathbf{y}]^{\text{alea}} = \frac{\alpha_0}{\alpha_0 + 1} (\text{Diag}(\bar{\boldsymbol{\mu}}) - \bar{\boldsymbol{\mu}}\bar{\boldsymbol{\mu}}^\top) \quad (20)$$

$$\text{Cov}[\mathbf{y}]^{\text{epis}} = \frac{1}{\alpha_0 + 1} (\text{Diag}(\bar{\boldsymbol{\mu}}) - \bar{\boldsymbol{\mu}}\bar{\boldsymbol{\mu}}^\top) \quad (21)$$

in which  $\text{Diag}(\cdot)$  represents a diagonal matrix with the specified vector on its diagonal. The derivations are provided in our supplementary material.

Similar to regression (Eq. (15) and Eq. (16)), the aleatoric and epistemic components are proportional:  $\frac{\text{Cov}[\mathbf{y}]^{\text{alea}}}{\text{Cov}[\mathbf{y}]^{\text{epis}}} = \alpha_0$ , which is reasonable since as the model gains more evidence, the Dirichlet strength  $\alpha_0$  tends to increase, suppressing the model uncertainty.

**Class-wise uncertainty quantification.** With the covariance matrices, the variance-based evidential uncertainties of each class can be directly retrieved from the diagonal:

$$U_c = \text{Cov}[\mathbf{y}]_{c,c} = \bar{\mu}_c(1 - \bar{\mu}_c) \quad (22)$$

$$U_c^{\text{alea}} = \text{Cov}[\mathbf{y}]_{c,c}^{\text{alea}} = \frac{\alpha_0}{\alpha_0 + 1} \bar{\mu}_c(1 - \bar{\mu}_c) \quad (23)$$

$$U_c^{\text{epis}} = \text{Cov}[\mathbf{y}]_{c,c}^{\text{epis}} = \frac{1}{\alpha_0 + 1} \bar{\mu}_c(1 - \bar{\mu}_c) \quad (24)$$

in which  $U_c$ ,  $U_c^{\text{alea}}$ , and  $U_c^{\text{epis}}$  are the total, aleatoric, and epistemic uncertainties associated with class  $c$  of data  $\mathbf{x}$ .

**Sample-wise uncertainty quantification.** Sample-wise uncertainty can be obtained by simply aggregating the class-wise uncertainties over all the classes:

$$U = \sum_{c=1}^C U_c = 1 - \sum_{c=1}^C \bar{\mu}_c^2 \quad (25)$$

$$U^{\text{alea}} = \sum_{c=1}^C U_c^{\text{alea}} = \frac{\alpha_0}{\alpha_0 + 1} \left( 1 - \sum_{c=1}^C \bar{\mu}_c^2 \right) \quad (26)$$

$$U^{\text{epis}} = \sum_{c=1}^C U_c^{\text{epis}} = \frac{1}{\alpha_0 + 1} \left( 1 - \sum_{c=1}^C \bar{\mu}_c^2 \right) \quad (27)$$

in which  $U$ ,  $U^{\text{alea}}$ , and  $U^{\text{epis}}$  are the total, aleatoric, and epistemic uncertainties associated with data  $\mathbf{x}$ .

**Correlation quantification.** The class correlation matrix can be directly obtained from the covariance matrix by definition:

$$\text{Corr}[\mathbf{y}] := \frac{\text{Cov}[\mathbf{y}]}{\sigma(\mathbf{y})\sigma(\mathbf{y})^\top} \quad (28)$$

in which  $\sigma(\mathbf{y}) = \sqrt{\text{diag}(\text{Cov}[\mathbf{y}])}$  represents the class standard deviation and  $\text{diag}(\cdot)$  denotes the diagonal of a matrix. The off-diagonal entries of the correlation matrix provide between-class correlation information, *e.g.*,  $\text{Corr}[\mathbf{y}]_{i,j}$  is the correlation between class  $i$  and class  $j$  ( $i \neq j$ ) on the sample  $\mathbf{x}$ . This gives a measure of how different classes correlate with each other in a given sample.

**Comparison with entropy-based approach.** Entropy-based and variance-based approaches adopt the same way to construct and train the ENN but quantify uncertainties in different manners. Both methods can quantify uncertainties of a sample  $\mathbf{x}$  as a whole (Eq. (8), Eq. (9), Eq. (10) for entropy-based; Eq. (25), Eq. (26), Eq. (27) for variance-based). However, our variance-based approach further provides uncertainties of each single class  $c$  of sample  $\mathbf{x}$  (Eq. (22), Eq. (23), Eq. (24)) and yields the between-class correlation of each pair of classes (Eq. (28)).

**From theory to application.** The variance-based classification EDL theories can be applied following three steps: 1) Choose an appropriate classification DNN model and turn it into an ENN by changing the output activation from softmax to any function that produces only positive values (*e.g.*, exponential function). 2) Train the ENN by Eq. (5). 3) Given a test sample  $\mathbf{x}$  during inference, the ENN predicts the Dirichlet parameters  $\boldsymbol{\alpha}$ . Then the class prediction is given by Eq. (4), the associated predictive uncertainties can be quantified via Eq. (22)-Eq. (27), and the class correlations can be calculated with Eq. (28).

## 4. Active Domain Adaptation

### 4.1. Previous Work

**Problem formulation.** ADA assumes that there is a source domain dataset  $\mathcal{S} = \{(\mathbf{x}_i^s, \mathbf{y}_i^s)\}_{i=1}^{n_s}$  and a target domain dataset  $\mathcal{T} = \{\mathbf{x}_i^t\}_{i=1}^{n_t}$ . The two domains follow different data distributions but share the same label space. Target data are separated into a labeled subset  $\mathcal{T}^l$  and an unlabeled subset  $\mathcal{T}^u$ , i.e.,  $\mathcal{T} = \mathcal{T}^l \cup \mathcal{T}^u$ . Initially, all target data are unlabeled,  $\mathcal{T}^l = \emptyset$  and  $\mathcal{T}^u = \mathcal{T}$ .

In active selection round  $k \in \{1, 2, \dots, K\}$ ,  $b_k^u$  samples in  $\mathcal{T}^u$  are selected by some sampling strategy to query labels from the oracle. These selected data are then removed from  $\mathcal{T}^u$  and added to  $\mathcal{T}^l$  with their labels. A total budget of  $B$  determines the maximum number of samples that can be labeled, and typically,  $B = \sum_{k=1}^K b_k^u$ .

**DUC algorithm.** DUC [43] is an EDL-based active learning framework that has achieved SOTA performance in ADA. Its active sampling approach employs a two-step selection technique based on two types of evidential uncertainties. In each active round  $k$ , the data from  $\mathcal{T}^u$  are first sorted by their EU. The top  $\kappa b_k^u$  uncertain samples are passed to the second step and sorted by their AU. The most uncertain  $b_k^u$  samples in the second step are finally sent to the oracle to annotate.

**DUC model training.** In addition to supervised training of EDL, unsupervised uncertainty guidance (UG) [43] is utilized in model optimization. By minimizing the uncertainties of unlabeled targets  $\mathcal{T}^u$ , domain shift can be greatly reduced. The unsupervised loss for a sample  $\mathbf{x}$  is designed as the combination of its AU and EU:

$$\mathcal{L}^{UG} = \lambda_a U^{alea} + \lambda_e U^{epis} \quad (29)$$

The total loss is the sum of supervised loss (EDL) and unsupervised loss (UG).

$$\mathcal{L} = \sum_{\mathbf{x}_i \in \mathcal{S} \cup \mathcal{T}^l} \mathcal{L}^{EDL}(\mathbf{x}_i) + \sum_{\mathbf{x}_i \in \mathcal{T}^u} \mathcal{L}^{UG}(\mathbf{x}_i) \quad (30)$$

### 4.2. Certainty Sampling

**Motivation.** In this paper, we propose simultaneous certainty and uncertainty sampling in active learning to exploit the value of confident samples. Previous AL and ADA algorithms mainly focus on picking the most uncertain samples to annotate, ignoring the value of certain samples in the dataset. However, if wisely utilized, certain samples can boost the model performance: 1) As the model is certain about those data, their predictions are very likely to be correct. Thus adding them to the labeled pool can provide additional information for supervised learning. 2) Removing certain samples from the unlabeled pool effectively reduces the search space for uncertain samples. Thus for uncertainty

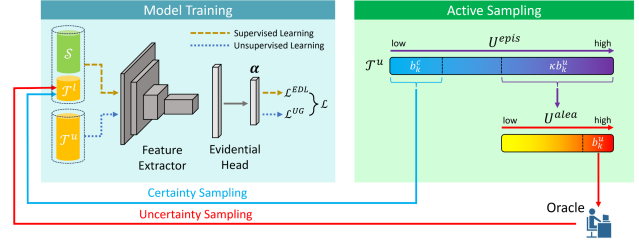


Figure 1. DUC framework with simultaneous certainty and uncertainty sampling. In active sampling round  $k$ , the unlabeled target data in  $\mathcal{T}^u$  are first sorted by their epistemic uncertainty  $U^{epis}$ . The  $b_k^c$  samples with the least  $U^{epis}$  are moved from  $\mathcal{T}^u$  to labeled target dataset  $\mathcal{T}^l$  with their pseudo labels, and the  $\kappa b_k^u$  samples with the most  $U^{epis}$  are then sorted by their aleatoric uncertainty  $U^{alea}$ . The  $b_k^u$  samples with the most  $U^{alea}$  are sent to the oracle to query labels, after which these samples are moved to  $\mathcal{T}^l$ .

sampling in the current and subsequent active rounds, computation is reduced and the likelihood of discovering informative samples is increased.

**DUC with certainty sampling.** As illustrated in Fig. 1, certainty sampling can be incorporated into DUC in its first step of active selection: when sorting  $\mathcal{T}^u$  by their EU, we select the top  $b_k^c$  samples with the least EU as the certain samples and add them directly to  $\mathcal{T}^l$  with their predictions as pseudo labels. Note that this EU-based certainty sampling does not incur extra computation, as  $\mathcal{T}^u$  is not sorted one more time using another metric. Though applied here in the DUC framework, this idea of certainty sampling is generally applicable in any AL algorithm.

**Class-balanced certainty sampling.** We introduce a variant of certainty sampling to address the concern that the selected certain samples might be biased towards simple classes. To avoid training set imbalance, when picking  $b_k^c$  most certain samples in round  $k$ , we first divide  $\mathcal{T}^u$  into  $C$  classes according to their predictions and then select about  $\frac{b_k^c}{C}$  most certain samples from each class. The class-balanced selection is vital if class difficulties vary greatly.

## 5. Experiments

### 5.1. Datasets

Experiments are based on two multi-domain image classification datasets: Office-Home [39] and Visda-2017 [25].

Office-Home [39] is a set of 15,588 images categorized into 65 classes that are collected in typical office and home settings. Images are drawn from 4 different domains: Art (A), Clipart (C), Product (P), and Real World (R).

Visda-2017 [25] contains synthetic (S) and real (R) images categorized into 12 classes. We use the training set (152,397 synthetic images) as source data and the validation set (55,388 real images) as target data.

AL Method	Office-Home														Average
	Visda-2017 S→R	A→C	A→P	A→R	C→A	C→P	C→R	P→A	P→C	P→R	R→A	R→C	R→P		
Source-only	44.7±0.1	42.1	66.3	73.3	50.7	59.0	62.6	51.9	37.9	71.2	65.2	42.6	76.6	58.3	
CoreSet [31]	81.9±0.3	51.8	72.6	75.9	58.3	68.5	70.1	58.8	48.8	75.2	69.0	52.7	80.0	65.1	
Random	78.1±0.6	52.5	74.3	77.4	56.3	69.7	68.9	57.7	50.9	75.8	70.0	54.6	81.3	65.8	
BvSB [16]	81.3±0.4	56.3	78.6	79.3	58.1	74.0	70.9	59.5	52.6	77.2	71.2	56.4	84.5	68.2	
WAAL [34]	83.9±0.4	55.7	77.1	79.3	61.1	74.7	72.6	60.1	52.1	78.1	70.1	56.6	82.5	68.3	
Entropy	82.7±0.3	58.0	78.4	79.1	60.5	73.0	72.6	60.4	54.2	77.9	71.3	58.0	83.6	68.9	
BADGE [2]	84.3±0.3	58.2	79.7	79.9	61.5	74.6	72.9	61.5	56.0	78.3	71.4	60.9	84.2	69.9	
AADA [37]	80.8±0.4	56.6	78.1	79.0	58.5	73.7	71.0	60.1	53.1	77.0	70.6	57.0	84.5	68.3	
DBAL [8]	82.6±0.3	58.7	77.3	79.2	61.7	73.8	73.3	62.6	54.5	78.1	72.4	59.9	84.3	69.6	
TQS [10]	83.1±0.4	58.6	81.1	81.5	61.1	76.1	73.3	61.2	54.7	79.7	73.4	58.9	86.1	70.5	
CLUE [26]	85.2±0.4	58.0	79.3	80.9	68.8	77.5	76.7	66.3	57.9	81.4	75.6	60.8	86.3	72.5	
EADA [42]	88.3±0.1	63.6	84.4	83.5	70.7	83.7	80.5	73.0	63.5	85.2	78.4	65.4	88.6	76.7	
DUC-ENT w/o CS [43]	89.3±0.1	66.2±0.9	85.1±0.3	84.7±0.2	71.1±1.0	84.0±0.4	81.5±0.0	72.0±0.8	67.0±0.8	85.6±0.2	79.9±0.4	71.0±0.5	89.5±0.5	78.1±0.2	
DUC-VAR w/o CS	89.4±0.1	66.2±0.4	85.0±0.4	85.0±0.2	70.8±0.4	85.1±0.3	81.7±0.2	71.7±0.1	67.0±0.3	85.6±0.5	80.1±0.3	70.9±0.3	89.4±0.4	78.2±0.1	
DUC-ENT w/ CS	89.3±0.2	<b>69.0</b> ±0.7	85.3±0.9	<b>85.3</b> ±0.1	73.0±0.3	85.1±0.4	82.6±0.1	73.9±0.3	68.5±0.3	<b>86.0</b> ±0.1	<b>80.6</b> ±0.5	<b>72.5</b> ±0.3	89.5±0.3	79.3±0.2	
DUC-VAR w/ CS	<b>89.5</b> ±0.1	68.3±0.4	<b>85.9</b> ±0.2	85.2±0.1	<b>73.8</b> ±0.0	<b>85.6</b> ±0.4	<b>83.4</b> ±0.5	<b>74.8</b> ±0.9	<b>68.9</b> ±0.5	85.7±0.3	80.2±0.0	71.7±0.9	<b>89.7</b> ±0.3	<b>79.5</b> ±0.2	

Table 1. Classification accuracy (%) of ResNet-50 on Office-Home and Visda-2017 with a budget of 5% target data. DUC-ENT: DUC with entropy-based uncertainty quantification. DUC-VAR: DUC with variance-based uncertainty quantification. CS: certainty sampling.

## 5.2. Experimental Setup

We adopt most experimental settings of [43] to get comparable results. There are  $K = 5$  active selection rounds and a budget of  $B = 5\%|\mathcal{T}|$  in total. In active round  $k \in \{1, 2, \dots, K\}$ ,  $\kappa = 10$ , and  $b_k^u = B/5 = 1\%|\mathcal{T}|$  uncertain samples and  $b_k^c = \kappa\%|\mathcal{T}|$  certain samples are selected from the unlabeled pool  $\mathcal{T}^u$ . We gradually increase the certainty sampling ratio because as training progresses, the model can give increasingly more correct predictions.

ENN is constructed using ResNet-50 [13] pre-trained on ImageNet [9] as the backbone and exponential function as the output activation. The model is trained for 50 epochs on Office-Home and 40 epochs on Visda-2017, and active sampling is performed in epochs 10, 12, 14, 16, and 18. SGD optimizer with batch size 32, momentum 0.9, and weight decay of 0.001 and the learning rate scheduling policy in [20] are used for model training. The initial learning rates for Office-Home and Visda-2017 are 0.004 and 0.001, respectively. In the loss function,  $\lambda_a = 0.05$ . For entropy-based method,  $\lambda_e = 1$ . For variance-based method,  $\lambda_e = 50$  for Office-Home and  $\lambda_e = 10$  for Visda-2017. The regularization coefficient  $\lambda_{reg} = \frac{1}{C}$  to mitigate the influence of the number of classes  $C$ .

For each domain transition, we independently run the experiment three times, each with a randomly selected seed, and report the mean and standard deviation of the three final accuracies. The experiments are implemented with PyTorch [24] on an NVIDIA GeForce RTX 4090 Laptop GPU.

## 5.3. Results

### 5.3.1 Classification Accuracy

The accuracy comparison of DUC with entropy-based and variance-based evidential uncertainties (denoted by DUC-ENT and DUC-VAR respectively) on the two datasets is illustrated in Tab. 1, together with the results of previous AL algorithms.

Source	UG	US	CS	A→C
✓	-	-	-	46.0
✓	✓	-	-	52.8
✓	-	✓	-	62.1
✓	✓	✓	-	66.2
✓	✓	✓	✓	68.3

Table 2. Ablation study on Office-Home A→C domain transition. Source: training by source data. UG: uncertainty guidance. US: uncertainty sampling. CS: certainty sampling.

The variance-based approach achieves slightly better average accuracy than the entropy-based one on the two datasets and is also more stable when certainty sampling is not involved, as the average standard deviation of DUC-VAR on 12 domains of Office-Home (0.32) is significantly lower than that of DUC-ENT (0.50).

With certainty sampling, the average accuracy of Office-Home is improved by more than 1%, but the effect of certainty sampling on transitions with little domain shift is negligible (e.g., on Visda-2017 S→R and Office-Home R→P). However, it is noteworthy that certainty sampling never harms the model performance in any domain transition using either DUC-ENT or DUC-VAR.

### 5.3.2 Ablation Study

The three main components of the improved DUC framework with variance-based uncertainty quantification are uncertainty guidance (UG), uncertainty sampling (US), and certainty sampling (CS). The contributions of the three parts are analyzed through an ablation study on the Office-Home dataset Art to Clipart domain transition (Tab. 2) since this transition suffers from the greatest domain shift and makes the effect of each component more evident.

**Uncertainty guidance.** It can be noticed that UG training always improves accuracy on the target domain, with or without active US. Even in the unsupervised domain adap-



Uncertainty	Quantification	Office-Home														Average
		Visda-2017	A→C	A→P	A→R	C→A	C→P	C→R	P→A	P→C	P→R	R→A	R→C	R→P		
AU	ENT	75.6±0.2	78.9±0.1	80.8±0.5	84.3±0.1	83.4±0.2	83.3±0.4	83.2±0.7	<b>83.4±0.1</b>	80.1±0.1	85.7±0.1	85.7±0.2	<b>81.3±0.5</b>	85.8±0.0	83.0±0.2	
	VAR	75.3±0.1	<b>79.5±0.2</b>	<b>81.5±0.2</b>	<b>85.0±0.0</b>	<b>83.5±0.4</b>	<b>83.4±0.1</b>	<b>83.6±0.1</b>	83.3±0.2	<b>80.7±0.3</b>	<b>86.0±0.2</b>	<b>86.3±0.2</b>	80.0±0.2	<b>86.5±0.0</b>	<b>83.3±0.1</b>	
EU	ENT	73.8±0.1	77.8±0.0	79.5±1.0	83.1±0.3	81.8±0.1	82.2±0.4	81.8±0.7	81.9±0.1	78.8±0.0	84.4±0.1	83.7±0.3	80.1±0.6	84.5±0.0	81.7±0.2	
	VAR	73.8±0.0	78.5±0.0	80.5±0.2	84.2±0.1	82.3±0.6	81.8±0.0	82.0±0.2	81.8±0.0	79.2±0.4	84.7±0.1	84.9±0.3	78.6±0.2	85.5±0.1	82.0±0.0	

Table 3. AUROC (%) of misclassification detection on Office-Home and Visda-2017 based on uncertainties in epoch 10. AU: aleatoric uncertainty. EU: epistemic uncertainty. ENT: entropy-based uncertainty quantification. VAR: variance-based uncertainty quantification.

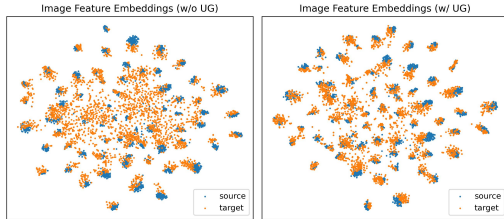


Figure 2. The t-SNE visualization of image feature embeddings derived by the model trained without (left) and with (right) variance-based uncertainty guidance on Office-Home dataset A→C domain transition. No active sampling is performed.

tation setting where no target samples are labeled, UG improves accuracy on the target domain by a large margin (6.8%), meaning that minimizing target evidential uncertainty serves as good guidance to model training.

A qualitative study on the feature space (Fig. 2) demonstrates that UG training significantly mitigates the domain shift from source to target. When no guidance is provided, target features shift greatly from source ones and tend to be widely spread, incapable of forming into class-wise clusters. However, with the guidance of variance-based evidential uncertainties, the target features appear visually closer to the source ones and learn to group into small clusters. The learned features of target samples from different classes are far more distinguishable, even if no target labels are provided to them at this stage.

**Certainty sampling.** A vital prerequisite for CS to work is that the certain sample selection criterion should be a strong indicator of prediction correctness. Therefore, whether low epistemic uncertainty indicates a high likelihood of prediction correctness has to be examined. The average prediction accuracy of all the selected class-balanced certain samples is 94.2% in Office-Home and 97.2% in Visda-2017. Since a total of 15% target data are selected as certain samples, prediction mistakes are almost inevitable, but an overall accuracy of more than 94% is still sufficient to improve the model performance under most circumstances.

### 5.3.3 Evidential Uncertainty Analysis

#### Evidential uncertainty for misclassification detection.

Apparently, it is preferable that before active selection, most samples chosen in US are incorrectly predicted and most

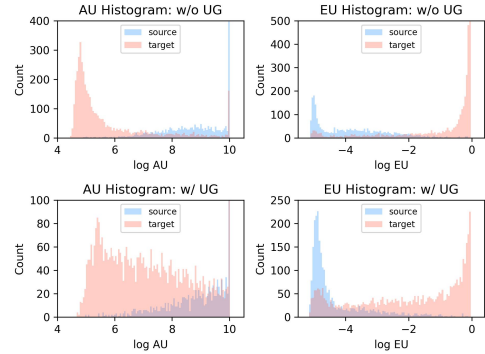


Figure 3. Uncertainty histograms of Office-Home source (Art) and target (Clipart) domains. The model is trained on source data only (top row) and with UG (bottom row) without active sampling. Uncertainty is shown in the logarithmic scale for elegance.

samples chosen in CS are correctly predicted, so that we can maximize the performance gain from a limited labeling budget without introducing excessive wrong pseudo labels in training. Both these two desires require a strong relationship between evidential uncertainty and misclassification.

This relationship can be explicitly measured by treating the uncertainty estimation as a binary classification problem, in which the prediction target is whether the sample is misclassified or not and the prediction score is the evidential uncertainty [21]. In this way, AUROC can be quantified to evaluate how evidential uncertainty is related to the likelihood of giving a wrong prediction. The mean and standard deviation of AUROC scores from the three runs of DUC are given in Tab. 3. The scores are calculated in epoch 10, right before the first active sampling round. It can be observed that variance-based evidential uncertainties have a higher AUROC score than the entropy-based one on most domain transitions in Office-Home, but the entropy-based aleatoric uncertainty is a better metric to detect misclassification than the variance-based one in Visda-2017.

**Evidential uncertainties across domains.** To investigate whether variance-based epistemic uncertainty is a good measure of targetness and the effect of UG training on domain-level uncertainties, experiments are conducted to obtain the uncertainty histograms demonstrated in Fig. 3.

The target domain has, in general, a higher epistemic uncertainty when UG is not performed, indicating that epis-

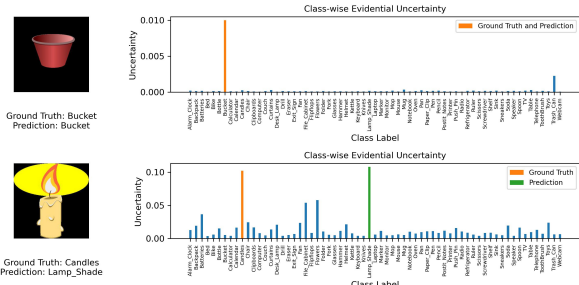


Figure 4. Class-level evidential uncertainties of two example Clipart images from Office-Home A→C domain transition.

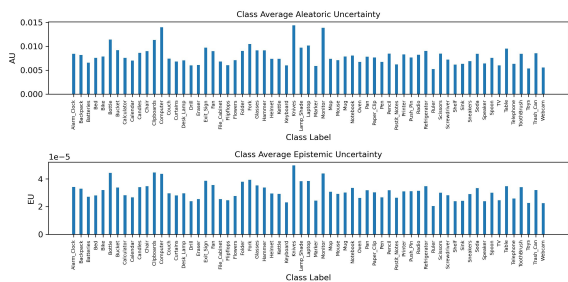


Figure 5. Office-Home Clipart domain class-level evidential uncertainties under A→C domain transition.

temic uncertainty can serve as the domain discriminator to some extent. After the model is trained with target uncertainty guidance, epistemic uncertainties of some samples from both domains are reduced, but the difference between the two distributions is still significant.

The aleatoric uncertainty, however, is much smaller in target than in source. When no target label is provided, the model does not have sufficient information to precisely estimate the intrinsic data uncertainty and tends to give relatively low aleatoric uncertainties to the target samples.

**Class-level Evidential Uncertainties.** Based on Eq. (22)-Eq. (24), uncertainties can be quantified at the class level for each image in the target domain, showing which classes the model is most uncertain about. Fig. 4 gives two examples with their class-wise uncertainties. By aggregating all the domain samples, average class-wise evidential uncertainties can be quantified for the entire domain, as shown in Fig. 5. This information is useful in applications that require uncertainty estimation at the class level.

### 5.3.4 Between-class Correlation Analysis

The correlation matrix in Eq. (28) can be calculated for each individual image sample  $\mathbf{x}$ . To investigate the correlation between a pair of classes  $(i, j)$  in a given dataset, we take the average of  $\text{Corr}[\mathbf{y}]_{i,j}$  of all the samples from class  $i$  and class  $j$ . This gives the dataset-level class correlation.

Rank	Class $i$	Class $j$	Correlation
1	Computer	Monitor	-0.364
2	Bucket	Trash Can	-0.239
3	Desk Lamp	Lamp Shade	-0.226
4	Knives	Scissors	-0.213

Table 4. The top 4 most correlated classes in the Office-Home Clipart domain, obtained from A→C domain transition.

Samples from classes other than  $i$  and  $j$  are excluded because most of them do not contain information about the two classes. Note that when class labels are unavailable, predictions can be used as a proxy for ground truth classes.

Here, we assume the labels are available and check the class correlations in the Office-Home Clipart dataset (Tab. 4). A relatively large negative correlation indicates that the model tends to make more mistakes in distinguishing the two classes. The average correlations are generally not too strong because for many samples, the model is still able to discriminate between the two categories. However, the relative ranking of pair correlations is still informative, as highly similar pairs, such as computer and monitor, can be successfully identified using our approach.

## 6. Applications and Limitations

The EDL-based uncertainty quantification framework is widely applicable in scenarios where uncertainty estimation is crucial. In active learning, the uncertainty metric helps to identify informative samples, which can be annotated to improve model performance. In medical image classification and segmentation, uncertainty indicates to what extent the diagnosis can be trusted. In autonomous driving, uncertainty given by the object detector is pivotal for the vehicle to make driving decisions to ensure safety.

However, the EDL approach is only applicable in the supervised setting, where the ground truth targets are available for model training. For unsupervised learning, EDL is not capable of deriving uncertainties. Extending EDL to unsupervised settings is a possible direction of future research.

## 7. Conclusion

Inspired by the theory of deep evidential regression, we introduce a variance-based approach to quantify evidential uncertainties in deep learning classification problems. Compared with the traditional entropy-based approach, our method derives not only sample-level but also class-level uncertainties. In addition, we show that between-class correlations can be calculated in the EDL setting to identify highly similar class pairs. We also propose certainty sampling to boost active learning performance under large domain shift. This variance-based approach will enable the development of more advanced algorithms based on class-level evidential uncertainties and correlation information.



## References

- [1] Alexander Amini, Wilko Schwarting, Ava Soleimany, and Daniela Rus. Deep evidential regression. *Advances in Neural Information Processing Systems*, 33:14927–14937, 2020.
- [2] Jordan T Ash, Chicheng Zhang, Akshay Krishnamurthy, John Langford, and Alekh Agarwal. Deep batch active learning by diverse, uncertain gradient lower bounds. *arXiv preprint arXiv:1906.03671*, 2019.
- [3] Robert B Ash. *Information theory*. Courier Corporation, 2012.
- [4] Murat Seckin Ayhan and Philipp Berens. Test-time data augmentation for estimation of heteroscedastic aleatoric uncertainty in deep neural networks. In *Medical Imaging with Deep Learning*, 2018.
- [5] 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.
- [6] Charles Blundell, Julien Cornebise, Koray Kavukcuoglu, and Daan Wierstra. Weight uncertainty in neural network. In *International conference on machine learning*, pages 1613–1622. PMLR, 2015.
- [7] Bertrand Charpentier, Ransalu Senanayake, Mykel Kochenderfer, and Stephan Günnemann. Disentangling epistemic and aleatoric uncertainty in reinforcement learning. *arXiv preprint arXiv:2206.01558*, 2022.
- [8] Antoine de Mathelin, Francois Deheeger, Mathilde Mougeot, and Nicolas Vayatis. Discrepancy-based active learning for domain adaptation. *arXiv preprint arXiv:2103.03757*, 2021.
- [9] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.
- [10] Bo Fu, Zhangjie Cao, Jianmin Wang, and Mingsheng Long. Transferable query selection for active domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7272–7281, 2021.
- [11] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059. PMLR, 2016.
- [12] Jakob Gawlikowski, Cedrique Rovile Njjeutcheu Tassi, Mohsin Ali, Jongseok Lee, Matthias Humt, Jianxiang Feng, Anna Kruspe, Rudolph Triebel, Peter Jung, Ribana Roscher, et al. A survey of uncertainty in deep neural networks. *arXiv preprint arXiv:2107.03342*, 2021.
- [13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [14] Patrick Hemmer, Niklas Kühl, and Jakob Schöffer. Deal: Deep evidential active learning for image classification. *Deep Learning Applications, Volume 3*, pages 171–192, 2022.
- [15] Craig K Jones, Guoqing Wang, Vivek Yedavalli, and Haris Sair. Direct quantification of epistemic and aleatoric uncertainty in 3d u-net segmentation. *Journal of Medical Imaging*, 9(3):034002–034002, 2022.
- [16] Ajay J Joshi, Fatih Porikli, and Nikolaos Papanikolopoulos. Multi-class active learning for image classification. In *2009 IEEE conference on computer vision and pattern recognition*, pages 2372–2379. IEEE, 2009.
- [17] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in neural information processing systems*, 30, 2017.
- [18] Hao Li, Yang Nan, Javier Del Ser, and Guang Yang. Region-based evidential deep learning to quantify uncertainty and improve robustness of brain tumor segmentation. *Neural Computing and Applications*, pages 1–15, 2022.
- [19] Michel Loève. *Probability theory*. Courier Dover Publications, 2017.
- [20] Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. Conditional adversarial domain adaptation. *Advances in neural information processing systems*, 31, 2018.
- [21] Andrey Malinin and Mark Gales. Predictive uncertainty estimation via prior networks. *Advances in neural information processing systems*, 31, 2018.
- [22] Yaniv Ovadia, Emily Fertig, Jie Ren, Zachary Nado, David Sculley, Sebastian Nowozin, Joshua Dillon, Balaji Lakshminarayanan, and Jasper Snoek. Can you trust your model’s uncertainty? evaluating predictive uncertainty under dataset shift. *Advances in neural information processing systems*, 32, 2019.
- [23] 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.
- [24] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.
- [25] 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.
- [26] Viraj Prabhu, Arjun Chandrasekaran, Kate Saenko, and Judy Hoffman. Active domain adaptation via clustering uncertainty-weighted embeddings. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8505–8514, 2021.
- [27] Tiago Ramalho and Miguel Miranda. Density estimation in representation space to predict model uncertainty. In *Engineering Dependable and Secure Machine Learning Systems: Third International Workshop, EDSMLS 2020, New York City, NY, USA, February 7, 2020, Revised Selected Papers 3*, pages 84–96. Springer, 2020.
- [28] Amrith Rawat, Martin Wistuba, and Maria-Irina Nicolae. Harnessing model uncertainty for detecting adversarial examples. In *NIPS Workshop on Bayesian Deep Learning*, 2017.

- [29] Pengzhen Ren, Yun Xiao, Xiaojun Chang, Po-Yao Huang, Zhihui Li, Brij B Gupta, Xiaojiang Chen, and Xin Wang. A survey of deep active learning. *ACM computing surveys (CSUR)*, 54(9):1–40, 2021.
- [30] Abhijit Guha Roy, Sailesh Conjeti, Nassir Navab, Christian Wachinger, Alzheimer’s Disease Neuroimaging Initiative, et al. Bayesian quicknat: Model uncertainty in deep whole-brain segmentation for structure-wise quality control. *NeuroImage*, 195:11–22, 2019.
- [31] Ozan Sener and Silvio Savarese. Active learning for convolutional neural networks: A core-set approach. *arXiv preprint arXiv:1708.00489*, 2017.
- [32] Murat Sensoy, Lance Kaplan, and Melih Kandemir. Evidential deep learning to quantify classification uncertainty. *Advances in neural information processing systems*, 31, 2018.
- [33] Alexandru Constantin Serban, Erik Poll, and Joost Visser. Adversarial examples—a complete characterisation of the phenomenon. *arXiv preprint arXiv:1810.01185*, 2018.
- [34] Changjian Shui, Fan Zhou, Christian Gagné, and Boyu Wang. Deep active learning: Unified and principled method for query and training. In *International Conference on Artificial Intelligence and Statistics*, pages 1308–1318. PMLR, 2020.
- [35] Lewis Smith and Yarin Gal. Understanding measures of uncertainty for adversarial example detection. *arXiv preprint arXiv:1803.08533*, 2018.
- [36] Ava P Soleimany, Alexander Amini, Samuel Goldman, Daniela Rus, Sangeeta N Bhatia, and Connor W Coley. Evidential deep learning for guided molecular property prediction and discovery. *ACS central science*, 7(8):1356–1367, 2021.
- [37] Jong-Chyi Su, Yi-Hsuan Tsai, Kihyuk Sohn, Buyu Liu, Subhansu Maji, and Manmohan Chandraker. Active adversarial domain adaptation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 739–748, 2020.
- [38] Joost Van Amersfoort, Lewis Smith, Yee Whye Teh, and Yarin Gal. Uncertainty estimation using a single deep deterministic neural network. In *International conference on machine learning*, pages 9690–9700. PMLR, 2020.
- [39] Hemant 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.
- [40] Chen Wang, Xiang Wang, Jiawei Zhang, Liang Zhang, Xiao Bai, Xin Ning, Jun Zhou, and Edwin Hancock. Uncertainty estimation for stereo matching based on evidential deep learning. *Pattern Recognition*, 124:108498, 2022.
- [41] Qingyang Wu, He Li, Lexin Li, and Zhou Yu. Quantifying intrinsic uncertainty in classification via deep dirichlet mixture networks. *arXiv preprint arXiv:1906.04450*, 2019.
- [42] Binhui Xie, Longhui Yuan, Shuang Li, Chi Harold Liu, Xinjing Cheng, and Guoren Wang. Active learning for domain adaptation: An energy-based approach. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 8708–8716, 2022.
- [43] Mixue Xie, Shuang Li, Rui Zhang, and Chi Harold Liu. Dirichlet-based uncertainty calibration for active domain adaptation. *arXiv preprint arXiv:2302.13824*, 2023.
- [44] Ming Xie, Yuxi Li, Yabiao Wang, Zekun Luo, Zhenye Gan, Zhongyi Sun, Mingmin Chi, Chengjie Wang, and Pei Wang. Learning distinctive margin toward active domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7993–8002, 2022.
- [45] Xujiang Zhao, Feng Chen, Shu Hu, and Jin-Hee Cho. Uncertainty aware semi-supervised learning on graph data. *Advances in Neural Information Processing Systems*, 33:12827–12836, 2020.
- [46] Xujiang Zhao, Yuzhe Ou, Lance Kaplan, Feng Chen, and Jin-Hee Cho. Quantifying classification uncertainty using regularized evidential neural networks. *arXiv preprint arXiv:1910.06864*, 2019.