Discretization-Induced Dirichlet Posterior for Robust Uncertainty Quantification on Regression

Xuanlong Yu[1,2], **Gianni Franchi**[2], **Jindong Gu**[3], **Emanuel Aldea**[1]

[1]SATIE, Paris-Saclay University
[2]U2IS, ENSTA Paris, Institut Polytechnique de Paris
[3]University of Oxford

Abstract

Uncertainty quantification is critical for deploying deep neural networks (DNNs) in real-world applications. An Auxiliary Uncertainty Estimator (AuxUE) is one of the most effective means to estimate the uncertainty of the main task prediction without modifying the main task model. To be considered robust, an AuxUE must be capable of maintaining its performance and triggering higher uncertainties while encountering Out-of-Distribution (OOD) inputs, i.e., to provide robust aleatoric and epistemic uncertainty. However, for vision regression tasks, current AuxUE designs are mainly adopted for aleatoric uncertainty estimates, and AuxUE robustness has not been explored. In this work, we propose a generalized AuxUE scheme for more robust uncertainty quantification on regression tasks. Concretely, to achieve a more robust aleatoric uncertainty estimation, different distribution assumptions are considered for heteroscedastic noise, and Laplace distribution is finally chosen to approximate the prediction error. For epistemic uncertainty, we propose a novel solution named Discretization-Induced Dirichlet pOsterior (DIDO), which models the Dirichlet posterior on the discretized prediction error. Extensive experiments on age estimation, monocular depth estimation, and super-resolution tasks show that our proposed method can provide robust uncertainty estimates in the face of noisy inputs and that it can be scalable to both image-level and pixel-wise tasks.

1 Introduction

Uncertainty quantification in deep learning has gained significant attention in recent years (Blundell et al. 2015; Kendall and Gal 2017; Lakshminarayanan, Pritzel, and Blundell 2017; Abdar et al. 2021). Deep Neural Networks (DNNs) frequently provide overconfident predictions and lack uncertainty estimates, especially for regression models outputting single point estimates, affecting the interpretability and credibility of the prediction results.

There are two types of uncertainty in DNNs: unavoidable aleatoric uncertainty caused by data noise, and reducible epistemic or knowledge uncertainty due to insufficient training data (Hüllermeier and Waegeman 2021; Kendall and Gal 2017; Malinin and Gales 2018). Disentangling and estimating them can better guide the decision-making based on DNN predictions. Many seminal methods (Blundell et al. 2015; Gal and Ghahramani 2016; Lakshminarayanan, Pritzel, and Blundell 2017; Kendall and Gal 2017; Wen, Tran, and Ba 2020; Franchi et al. 2022) have been proposed to capture these two types of uncertainty. However, these methods require extensive modifications to the underlying model structure or more computational cost. Furthermore, since DNNs are often designed as task-oriented, obtaining uncertainty estimates by changing the structure of DNNs might reduce main task performance.

As one of the most effective methods, Auxiliary Uncertainty Estimators (AuxUE) (Corbière et al. 2019; Yu, Franchi, and Aldea 2021; Jain et al. 2021; Corbière et al. 2021; Besnier et al. 2021; Upadhyay et al. 2022; Shen et al. 2023) aim to obtain uncertainty estimates without affecting the main task performance. AuxUEs are DNNs that rely on the main task models used for estimating the uncertainty of the main task prediction. They are trained using the input, output, or intermediate features of the *pre-trained* main task model. In practice, the model inputs can be distribution-shifted from the training set, such as samples disturbed by noise (Hendrycks and Dietterich 2019), or even Out-of-Distribution (OOD) data. The pre-trained main task models mainly exhibit aleatoric uncertainty in the outputs given the In-Distribution (ID) inputs. Meanwhile, higher epistemic uncertainty is expected to be raised when OOD data is fed. A robust AuxUE is required in this case to provide robust aleatoric uncertainty estimates when facing In-Distribution (ID) inputs and epistemic uncertainty estimates when encountering OOD inputs. This can help to make effective decisions under anomalies and uncertainty (Guo et al. 2022), such as in autonomous driving (Arnez et al. 2020). To achieve robustness, disentangling the two types of uncertainty becomes a prerequisite, aiding in improved epistemic uncertainty estimation and a more robust aleatoric uncertainty estimation solution.

For vision regression tasks, basic AuxUE addresses only aleatoric uncertainty estimation (Yu, Franchi, and Aldea 2021). Recent works (Upadhyay et al. 2022; Qu et al. 2022) aim to improve the generalization ability of the basic AuxUEs. In DEUP (Jain et al. 2021), the authors propose to add a density estimator based on normalizing flows (Rezende and Mohamed 2015) in the AuxUE, yet challenging to apply on pixel-wise vision tasks. In the current context, both the robustness analysis and modeling of epistemic uncertainty are

underexplored for vision regression problems.

To further explore robust aleatoric and epistemic uncertainty estimation in vision regression tasks, in this work, we propose a novel uncertainty quantification solution based on AuxUE. For estimating aleatoric uncertainty, we follow the approach of previous works such as (Nix and Weigend 1994; Kendall and Gal 2017; Yu, Franchi, and Aldea 2021; Upadhyay et al. 2022) and model the heteroscedastic noise using different distribution assumptions. For epistemic uncertainty quantification, we apply a discretization approach to the continuous prediction errors of the main task. This helps to mitigate the numerical impact of the training targets, which may be distributed in a long-tailed manner. With the discretized prediction errors, we propose parameterizing Dirichlet posterior (Sensoy, Kaplan, and Kandemir 2018; Charpentier, Zügner, and Günnemann 2020; Joo, Chung, and Seo 2020) for estimating epistemic uncertainty without relying on OOD data during the training process.

In summary, our contributions are as follows: (1) We propose a generalized AuxUE solution for aleatoric and epistemic uncertainty estimation; (2) We propose Discretization-Induced Dirichlet pOsterior (DIDO), a new epistemic uncertainty estimation strategy for regression, which, to the best of our knowledge, is the only existing work employing this distribution for regression; (3) We demonstrate that assuming the noise which affects the main task predictions to follow Laplace distribution can help AuxUE achieve a more robust aleatoric uncertainty estimation; (4) We propose a new evaluation strategy for the OOD analysis of pixel-wise regression tasks based on systematically non-annotated patterns. We show the robustness and scalability of the proposed generalized AuxUE and DIDO on the age estimation, super-resolution and monocular depth estimation tasks.

2 Related Works

Auxiliary uncertainty estimation Auxiliary uncertainty estimation strategies can be divided into two categories: unsupervised and supervised. For the former, Dropout layer injection (Mi et al. 2022; Gal and Ghahramani 2016) samples the network by forward propagations, and (Hornauer and Belagiannis 2022) proposed to use the gradients from the back-propagation. For the latter, AuxUEs are applied to obtain the uncertainty. In addition to regression-oriented ones presented in Section 1, we here introduce classification-oriented solutions. ConfidNet (Corbière et al. 2019) and KLoS (Corbière et al. 2021) learn the true class probability and evidence for the DNNs, respectively. Shen et al. (Shen et al. 2023) apply evidential classification (Joo, Chung, and Seo 2020) to their AuxUE. ObsNet (Besnier et al. 2021) uses adversarial noise to provide more abundant training targets in semantic segmentation task for their AuxUE.

Evidential deep learning and Dirichlet networks Evidential deep learning (Ulmer 2021) (EDL) is a modern application of the Dempster-Shafer Theory (Dempster 1968) to estimate epistemic uncertainty with single forward propagation. In classification tasks, EDL is usually formed as parameterizing a prior (Malinin and Gales 2018, 2019) or a posterior (Joo, Chung, and Seo 2020; Charpentier, Zügner, and Günnemann 2020; Charpentier et al. 2022; Sensoy, Kaplan, and Kandemir 2018) Dirichlet distribution. In regression, EDL estimates parameters of the conjugate prior of Gaussian distribution (Amini et al. 2020; Charpentier et al. 2022; Malinin et al. 2020). Multi-task learning is also applied to alleviate main task performance degradation (Oh and Shin 2022), yet using AuxUE will not affect main task performance. Therefore, we apply EDL to our AuxUE. Moreover, we are the first to apply the Dirichlet network to the regression tasks by discretizing the main task prediction errors.

Robustness of uncertainty estimation A robust uncertainty estimator should show stable performance when encountering images perturbed to varying degrees (Michaelis et al. 2019; Hendrycks and Dietterich 2019; Kamann and Rother 2021). Similar studies are applied to evaluate the robustness of uncertainty estimates (Yeo, Kar, and Zamir 2021; Franchi et al. 2022). Meanwhile, it should provide a higher uncertainty when facing OOD data, such as in classification tasks (Hendrycks and Gimpel 2017; Liang, Li, and Srikant 2018). In image-level regression, we can use the definition of OOD from image classification (Techapanurak and Okatani 2021) in, for example, age estimation task. But for pixel-wise regression tasks, the notion of OOD data is ill-defined. Typical OOD analysis estimates uncertainty on a different dataset than the training dataset (Charpentier et al. 2022). Yet, image patterns that are rarely assigned ground truth values in the training set can also be regarded as OOD. In this work, we also provide a new evaluation strategy for OOD patterns based on outdoor depth estimation to compensate for this experimental shortfall.

3 Method

We define a training dataset $\mathcal{D} = \{\mathbf{x}^{(i)}, y^{(i)}\}_i^N$ where N is the number of images. We consider that $\mathbf{x}, \mathbf{y}$ are drawn from a joint distribution $P(\mathbf{x}, \mathbf{y})$. A pipeline for the main task and auxiliary uncertainty estimation is shown in Fig. 1. We define a main task DNN $f_{\boldsymbol{\omega}}$ with trainable parameters $\boldsymbol{\omega}$ as shown in the blue area in Fig. 1. Similar to (Blundell et al. 2015), we view $f_{\boldsymbol{\omega}}$ as a probabilistic model $P(y|\mathbf{x}, \boldsymbol{\omega})$ which follows a Gaussian distribution $\mathcal{N}(y|\mu, \sigma^2)$ (Bishop and Nasrabadi 2006). The variable σ^2 represents the variance of the noise in the DNN's prediction, and the variable μ is the prediction $\hat{y} = f_{\boldsymbol{\omega}}(\mathbf{x})$ in this case. The noise is considered here to be homoscedastic as all data have the same noise. The parameter $\boldsymbol{\omega}$ is optimized by maximizing the log-likelihood: $\hat{\boldsymbol{\omega}} = \mathrm{argmax}_{\boldsymbol{\omega}} \log\left(P(\mathcal{D}|\boldsymbol{\omega})\right)$ which is often performed by minimizing Negative Log Likelihood (NLL) loss in practice. With the above-mentioned Gaussian assumption on $\hat{y}$, the NLL loss optimizes with the same objective as the Mean Square Error loss (Bishop and Nasrabadi 2006), thus, only the prediction goal y is considered, and the uncertainty modeling is absent in the main task model training objective.

AuxUE aims to obtain this missing uncertainty estimation without modifying $\hat{\boldsymbol{\omega}}$. We consider two DNNs σ_{Θ_1} and σ_{Θ_2} in our generalized AuxUE with parameters Θ_1 and Θ_2, i.e., the two DNNs in the orange area of the Fig. 1. σ_{Θ_1} is for estimating aleatoric uncertainty $\mathbf{u}_{\text{alea}}$, and σ_{Θ_2} is for estimating epistemic uncertainty $\mathbf{u}_{\text{epis}}$. The backbone of σ_{Θ_1}

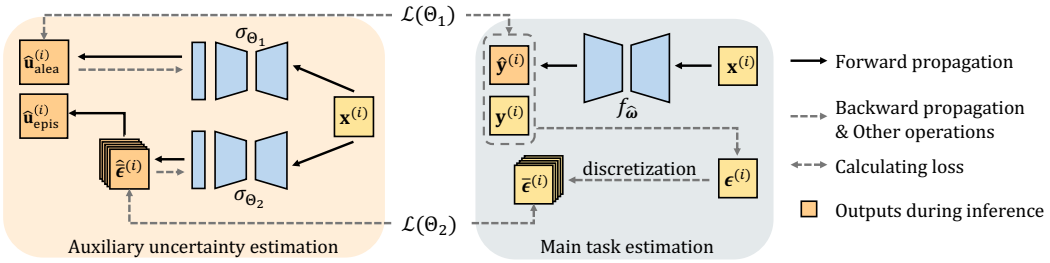

Figure 1: Pipeline of our proposed AuxUE solution. A generalized AuxUE is considered with two DNNs σ_{Θ_1} and σ_{Θ_2} for estimating aleatoric and epistemic uncertainty, respectively. Presented notations are consistent with and described in Section 3. The encoder parts of both DNNs can be shared, we compare the performance in Section 4.3. The input of AuxUE can be the input, output, or intermediate features of $f_{\hat{\boldsymbol{\omega}}}$, we here simplify it to the image $\mathbf{x}^{(i)}$ for brevity.

and σ_{Θ_2} are based on the basic AuxUEs such as ConfidNet (Corbière et al. 2019), BayesCap (Upadhyay et al. 2022) and SLURP (Yu, Franchi, and Aldea 2021) depending on the tasks. The input of AuxUE can be the input, output, or intermediate features of $f_{\hat{\boldsymbol{\omega}}}$ and it depends on the design of the basic AuxUEs, which is not the focus of this paper. We detail the inputs for different experiments in Supplementary material (Supp)[1] Section A.

3.1 Aleatoric Uncertainty Estimation on AuxUE

Based on the preliminaries of the settings, we now start with the first AuxUE σ_{Θ_1}, which addresses $\mathbf{u}_{\text{alea}}$ estimation problem as in SLURP and BayesCap.

We consider the data-dependent noise (Goldberg, Williams, and Bishop 1997; Bishop and Quazaz 1996; Nix and Weigend 1994) follows $\mathcal{N}(0, \boldsymbol{\sigma}^2)$. Then we use the DNN σ_{Θ_1} to estimate the heteroscedastic aleatoric uncertainty $\mathbf{u}_{\text{alea}}$ (Nix and Weigend 1994; Kendall and Gal 2017). $\widehat{\Theta}_1$ and the loss function $L(\Theta_1)$ are given by:

$$\widehat{\Theta}_1 = \underset{\Theta_1}{\operatorname{argmax}}\, P(\mathcal{D}|\hat{\boldsymbol{\omega}}, \Theta_1) = \underset{\Theta_1}{\operatorname{argmax}} \sum_{i=1}^{N} \log(P(y^{(i)}|\mathbf{x}^{(i)}, \hat{\boldsymbol{\omega}}, \Theta_1))$$

$$\mathcal{L}(\Theta_1) = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{2}\log(\sigma_{\Theta_1}(\mathbf{x}^{(i)})) + \frac{(y^{(i)} - f_{\hat{\boldsymbol{\omega}}}(\mathbf{x}^{(i)}))^2}{2\sigma_{\Theta_1}(\mathbf{x}^{(i)})}\right] \quad (1)$$

The top of the σ_{Θ_1} is an exponential or Softplus function to maintain the output non-negative. The aleatoric uncertainty estimation will be: $\hat{u}^{(i)}_{\text{alea}} = \sigma_{\Theta_1}(\mathbf{x}^{(i)})$. Minimizing $\mathcal{L}(\Theta_1)$ is also equivalent to making σ_{Θ_1} correctly predict the main task errors on the training set according to likelihood maximization. The errors set is denoted as $\boldsymbol{\epsilon} = \{\epsilon^{(i)}\}_{i=1}^{N} = \{(y^{(i)} - f_{\hat{\boldsymbol{\omega}}}(\mathbf{x}^{(i)}))^2\}_{i=1}^{N}$.

Given the fact that distribution assumption on the noise affecting $\hat{y}$ can be different than Gaussian, e.g., Laplacian (Marks et al. 1978) and Generalized Gaussian distribution (Nadarajah 2005; Upadhyay et al. 2022) also been considered in this work, the corresponding loss functions are provided in Supp Section B. The objective remains unchanged: employing AuxUE to estimate and predict the component associated with aleatoric uncertainty using various distribution assumptions. Perturbing input data in various ways with different types of noise makes it challenging to accurately identify the actual noise distribution. Relying on a single distribution assumption and loss function can affect the reliability of aleatoric uncertainty estimates. In Section 4.3, we assess the impact of different distribution assumptions and losses on the robustness of these estimates.

3.2 Epistemic Uncertainty Estimation on AuxUE

Modeling AuxUEs as formalized in Eq. 1 helps to estimate aleatoric uncertainty for $f_{\hat{\boldsymbol{\omega}}}$. Yet, taking this uncertainty prediction as an indicator for epistemic uncertainty is not methodologically grounded. Evidential learning is considered to be an effective uncertainty estimation approach (Ulmer 2021). In regression tasks, DNN estimates parameters of Gaussian distribution's conjugate prior, like Normal Inverse Gamma (NIG) distribution (Amini et al. 2020). The training will make the model fall back onto a NIG prior for the rare samples by attaching lower evidence to the samples with higher prediction errors using a regularization term in the loss function (Amini et al. 2020). Yet, long-tailed prediction errors in standard AuxUE lead to assigning high evidence to most data points, diminishing its ability to estimate epistemic uncertainty, as confirmed by our experiments.

In contrast to previous works, which consider the *numerical value* of the prediction errors for both aleatoric and epistemic uncertainty estimation, we disentangle them and apply discretization to mitigate numerical bias from long-tailed prediction errors. Specifically, σ_{Θ_1} focuses on aleatoric uncertainty considering the *numerical value* of prediction errors, while for epistemic uncertainty, σ_{Θ_2} will consider the *value-free categories* of the prediction errors. Specifically, we propose Discretization-Induced Dirichlet pOsterior (DIDO), involves discretizing prediction errors and estimating a Dirichlet posterior based on the discrete errors.

Discretization on prediction errors To mitigate numerical bias due to imbalanced data in our prediction error estimation, we employ a balanced discretization approach. Discretization is widely applied in classification approaches for regression (Yu, Franchi, and Aldea 2022). The popular discretization methods can be generally divided into hand-

[1]Refer to: https://arxiv.org/abs/2308.09065

crafted (Cao, Wu, and Shen 2017) and adaptive (Bhat, Alhashim, and Wonka 2021). The latter requires computationally expensive components like mini-ViT (Dosovitskiy et al. 2021) to extract global features. Thus, we discretize prediction errors in a handcrafted way.

For pixel-wise scenarios, discretization is applied using per-image prediction errors, and for other cases, such as image-level tasks and 1D signal estimation, we use per-dataset prediction errors. Details and demo-code can be found in Supp Section C.1 and C.2 respectively.

We divide the set of errors $\boldsymbol{\epsilon}$, denoted in Section 3.1, into K subsets, where the kth subset is represented by the subscript k. To do this, we sort the errors in ascending order and create a new set, denoted by $\boldsymbol{\epsilon}'$, with the same elements as $\boldsymbol{\epsilon}$. Then we divide $\boldsymbol{\epsilon}'$ into K subsets of equal size, represented by $\{\boldsymbol{\epsilon}_k\}_{k=1}^{K}$. Each error value $\epsilon^{(i)}$ is then replaced by the index of its corresponding subset $k \in [1, K]$, and transformed into a one-hot vector, denoted by $\bar{\boldsymbol{\epsilon}}^{(i)}$, as the final training target. Specifically, the one-hot vector is defined as:

$$\bar{\boldsymbol{\epsilon}}^{(i)} = [\bar{\epsilon}_1^{(i)} \dots \bar{\epsilon}_k^{(i)} \dots \bar{\epsilon}_K^{(i)}]^{\mathrm{T}} \in \mathbb{R}^K \tag{2}$$

where $\bar{\epsilon}_k^{(i)} = 1$ if $\epsilon^{(i)}$ belongs to the kth subset, and 0 otherwise. Each subset or bin represents a class of error severity. This process creates a new dataset, denoted by $\bar{\mathcal{D}} = \{\mathbf{x}^{(i)}, \bar{\boldsymbol{\epsilon}}^{(i)}\}_i^N$, consisting of discretized prediction errors represented as one-hot vectors, which serves for training the epistemic uncertainty estimator σ_{Θ_2}.

Modeling epistemic uncertainty using $\boldsymbol{\epsilon}$ in AuxUE In a Bayesian framework, given an input $\mathbf{x}$, the predictive uncertainty of a DNN is modeled by $P(y|\mathbf{x}, \mathcal{D})$. Since we have a trained main task DNN, and as proposed in (Malinin and Gales 2018), we assume a point-estimate $\hat{\boldsymbol{\omega}}$ of $\boldsymbol{\omega}$:

$$P(\boldsymbol{\omega}|\mathcal{D}) = \delta(\boldsymbol{\omega} - \hat{\boldsymbol{\omega}}) \rightarrow P(y|\mathbf{x}, \mathcal{D}) \approx P(y|\mathbf{x}, \hat{\boldsymbol{\omega}}) \tag{3}$$

with δ being the Dirac function.

We follow the previous assumption, i.e., the prediction is drawn from a Gaussian distribution $\mathcal{N}(y|\mu, \sigma^2)$ and according to (Amini et al. 2020), we denote $\boldsymbol{\alpha}$ as the parameters of the prior distributions of (μ, σ^2) and we have $P(\mu, \sigma^2|\boldsymbol{\alpha}, \hat{\boldsymbol{\omega}}) = P(\mu|\sigma^2, \boldsymbol{\alpha}, \hat{\boldsymbol{\omega}})P(\sigma^2|\boldsymbol{\alpha}, \boldsymbol{\omega}^*)$. After introducing $\boldsymbol{\alpha}$ and Eq. 3, we can approximate $P(y|\mathbf{x}, \mathcal{D})$ as:

$$\begin{aligned} P(y|\mathbf{x}, \mathcal{D}) &= \iint P(y|\mathbf{x}, \sigma^2)P(\sigma^2|\boldsymbol{\omega})P(\boldsymbol{\omega}|\mathcal{D})d\sigma^2 d\boldsymbol{\omega} \\ &= \int P(y|\mathbf{x}, \sigma^2)P(\sigma^2|\mathcal{D})d\sigma^2 \\ &\approx \int P(y|\mathbf{x}, \sigma^2)P(\sigma^2|\mathbf{x}, \boldsymbol{\alpha}, \hat{\boldsymbol{\omega}})d\sigma^2 \end{aligned} \tag{4}$$

Detailed derivation can be found in Supp Section C.3.

We can consider ϵ to be drawn from a continuous distribution parameterized by σ^2. The discrepancy in variances $P(\sigma^2|\mathcal{D})$ can describe epistemic uncertainty of the final prediction and the variational approach can be applied (Joo, Chung, and Seo 2020; Malinin and Gales 2018): $P(\sigma^2|\mathbf{x}, \boldsymbol{\alpha}, \hat{\boldsymbol{\omega}}) \approx P(\sigma^2|\mathcal{D})$. After discretization, we can transform the approximation to $P(\boldsymbol{\pi}|\mathbf{x}, \boldsymbol{\alpha}, \hat{\boldsymbol{\omega}}) \approx P(\boldsymbol{\pi}|\bar{\mathcal{D}})$, with $\bar{\mathcal{D}}$ defined above, $\boldsymbol{\pi}$ the parameters of a discrete distribution and $\boldsymbol{\alpha}$ re-defined as the prior distribution parameters of this discrete distribution. In the next section, we omit $\hat{\boldsymbol{\omega}}$ and $\mathbf{x}$ for the sake of brevity.

Dirichlet posterior for epistemic uncertainty According to the previous discussions on the epistemic uncertainty modeling and error discretization, we model Dirichlet posterior (Sensoy, Kaplan, and Kandemir 2018; Joo, Chung, and Seo 2020; Charpentier et al. 2022) on the discrete errors $\bar{\boldsymbol{\epsilon}}$ to achieve epistemic uncertainty on the main task.

Intuitively, we consider each one-hot prediction error $\bar{\boldsymbol{\epsilon}}^{(i)}$ to be drawn from a categorical distribution, and $\boldsymbol{\pi}^{(i)} = (\pi_1^{(i)}, \dots, \pi_K^{(i)})$ denotes the random variable over this distribution, where $\sum_{k=1}^{K} \pi_k^{(i)} = 1$ and $\pi_k^{(i)} \in [0, 1]$ for $k \in \{1, ..., K\}$. The conjugate prior of categorical distribution is a Dirichlet distribution:

$$P(\boldsymbol{\pi}^{(i)}|\boldsymbol{\alpha}^{(i)}) = \frac{\Gamma(S^{(i)})}{\prod_{k=1}^{K}\Gamma(\alpha_k^{(i)})} \prod_{k=1}^{K} {\pi_k^{(i)}}^{\alpha_k^{(i)}-1} \tag{5}$$

with $\Gamma(\cdot)$ the Gamma function, $\boldsymbol{\alpha}^{(i)}$ positive concentration parameters of Dirichlet distribution and $S^{(i)} = \sum_{k=1}^{K} \alpha_k^{(i)}$ the Dirichlet strength.

To get access to the epistemic uncertainty, the categorical posterior $P(\boldsymbol{\pi}|\bar{\mathcal{D}})$ is needed, yet it is untractable. Approximating $P(\boldsymbol{\pi}|\bar{\mathcal{D}})$ using Monte-Carlo sampling (Gal and Ghahramani 2016) or ensembles (Lakshminarayanan, Pritzel, and Blundell 2017) comes with an increased computational cost. Instead, we adopt a variational way to learn a Dirichlet distribution in Eq. 5 to approximate $P(\boldsymbol{\pi}|\bar{\mathcal{D}})$ as in (Joo, Chung, and Seo 2020). Here, σ_{Θ_2} outputs the concentration parameters $\boldsymbol{\alpha}$ of $P(\boldsymbol{\pi}|\boldsymbol{\alpha})$, and $\boldsymbol{\alpha}$ update according to the observed inputs. It can also be viewed as collecting the evidence $\boldsymbol{e}$ as a measure for supporting the classification decisions for each class (Sensoy, Kaplan, and Kandemir 2018), akin to estimating the Dirichlet posterior.

Since the numbers of data points are identical for each class in $\bar{\mathcal{D}}$, and no $\boldsymbol{e}^{(i)}$ output before training, we set the initial $\boldsymbol{\alpha}$ as $\mathbf{1}$ so that the Dirichlet concentration parameters can be formed as in (Sensoy, Kaplan, and Kandemir 2018; Charpentier, Zügner, and Günnemann 2020): $\boldsymbol{\alpha}^{(i)} = \boldsymbol{e}^{(i)} + \mathbf{1} = \sigma_{\Theta_2}(\mathbf{x}^{(i)}) + \mathbf{1}$, where $\boldsymbol{e}^{(i)}$ is given by an exponential function on the top of σ_{Θ_2}. Then we minimize the Kullback-Leibler (KL) divergence between the variational distribution $P(\boldsymbol{\pi}|\mathbf{x}, \Theta_2)$ and the true posterior $P(\boldsymbol{\pi}|\bar{\mathcal{D}})$ to achieve $\widehat{\Theta}_2$:

$$\begin{aligned} \widehat{\Theta}_2 &= \operatorname*{argmin}_{\Theta_2} \mathrm{KL}[P(\boldsymbol{\pi}|\mathbf{x}, \Theta_2)||P(\boldsymbol{\pi}|\bar{\mathcal{D}})] \\ &= \operatorname*{argmin}_{\Theta_2} \text{-}\mathbb{E}_{P(\boldsymbol{\pi}|\mathbf{x}, \Theta_2)}[\log P(\bar{\mathcal{D}}|\boldsymbol{\pi})] + \mathrm{KL}[P(\boldsymbol{\pi}|\mathbf{x}, \Theta_2)||P(\boldsymbol{\pi})] \end{aligned}$$

The loss function will be equivalent to minimizing the negative evidence lower bound (Jordan et al. 1999), considering the prior distribution $P(\boldsymbol{\pi})$ as $\mathrm{Dir}(\mathbf{1})$:

$$\begin{aligned} \mathcal{L}(\Theta_2) = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K}[\bar{\epsilon}_k^{(i)}(\psi(S^{(i)}) - \psi(\alpha_k^{(i)}))] \\ + \lambda \mathrm{KL}(\mathrm{Dir}(\boldsymbol{\alpha}^{(i)})||\mathrm{Dir}(\mathbf{1})) \end{aligned} \tag{6}$$

where ψ is the digamma function, λ is a positive hyperparameter for the regularization term and $\bar{\epsilon}$ is given by Eq. 2.

For measuring epistemic uncertainty, we consider using the spread in the Dirichlet distribution (Shen et al. 2023; Charpentier, Zügner, and Günnemann 2020), which is

shown in (Shen et al. 2023) to outperform other metrics, e.g. differential entropy. Specifically, the epistemic uncertainty is inversely proportional to the Dirichlet strength: $\hat{u}^{(i)}_{\text{epis}} = \sigma_{\widehat{\Theta}_2}(\mathbf{x}^{(i)}) = \frac{K}{S^{(i)}}$. The class corresponding to the maximum output from σ_{Θ_2} can also represent the aleatoric uncertainty. Yet, this is a rough estimate due to quantization errors and underperforming the other solutions. We provide the corresponding results in Supp Tab. A14. Overall, we take only σ_{Θ_1} output as the aleatoric uncertainty.

In conclusion, we propose a generalized AuxUE with two components, namely σ_{Θ_1} and σ_{Θ_2}, to quantify the uncertainty of main task model outputs. By assuming different distributions on heteroscedastic noise in training data (Section 3.1), σ_{Θ_1} is trained for aleatoric uncertainty estimation. Meanwhile, applying the proposed DIDO on σ_{Θ_2} and measuring the spread of the Dirichlet distribution (Section 3.2) aids in estimating epistemic uncertainty. Overall, we integrate the optimization for both uncertainty estimators, and the final loss for training the generalized AuxUE is:

$$\mathcal{L}_{\text{AuxUE}} = \mathcal{L}(\Theta_1) + \mathcal{L}(\Theta_2) \tag{7}$$

For $\mathcal{L}(\Theta_1)$, in addition to the Gaussian NLL, we will test other NLL loss functions according to different distribution assumptions in the experiment.

4 Experiments

In this section, we first show the feasibility of the proposed generalized AuxUE on toy examples. Then, we demonstrate the effectiveness of epistemic uncertainty estimation using the proposed DIDO on age estimation and monocular depth estimation (MDE) tasks, and investigate the robustness of aleatoric uncertainty estimation on MDE task. Due to page limitations, the experiments for an example of OOD detection in tabular data regression and the super-resolution task are provided in Supp Section A.2 and A.4 respectively.

In the result tables, the top two performing methods are highlighted in color. All the results are averaged by three runs. The shar.enc. and sep.enc. denote respectively shared-parameters for the encoders and separate encoders of σ_{Θ_1} and σ_{Θ_2} in the generalized AuxUE. For epistemic uncertainty, we compare our proposed method with the solutions based on modified main DNN: LDU (Franchi et al. 2022), Evidential learning (Evi.) (Amini et al. 2020; Joo, Chung, and Seo 2020) and Deep Ensembles (DEns.) (Lakshminarayanan, Pritzel, and Blundell 2017).

The detailed implementations and the main task performance for all experiments are provided in Supp Section A.

4.1 Toy Examples: Simple 1D Regression

We generate two toy datasets to illustrate uncertainty estimates given by our proposed AuxUE, as shown in Fig. 2. In both examples, a tight aleatoric uncertainty estimation is provided on training data areas. For epistemic uncertainty, in Fig. 2-A, DIDO provides small uncertainty until reaching the unknown inputs $x \notin [-3, 3]$. In Fig. 2-B, we report the 'in-between' uncertainty estimates (Foong et al. 2019). On the in-between part $x \in [-1, 3]$, DIDO can provide higher epistemic uncertainty than in training set regions

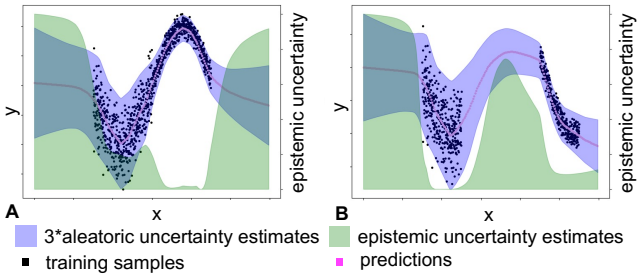

Figure 2: Results on toy examples. Aleatoric and epistemic uncertainty estimations given by our AuxUE are presented respectively as the uncertainty interval and degree (0-1).

$x \in [-3, -1]$ and $x \in [3, 5]$. In summary, the generalized AuxUE provides reliable uncertainty estimates in regions where training data is either present or absent.

4.2 Age Estimation and OOD Detection

Epistemic uncertainty estimation for age estimation is similar to one for classification problems but has rarely been discussed in previous works. We use (unmodified) official ResNet34 (He et al. 2016) checkpoints from Coral (Cao, Mirjalili, and Raschka 2020) as the main task models. Our AuxUE is applied in a ConfidNet (Corbière et al. 2019) style since it is more suitable for image-level tasks.

Evaluation settings and datasets We train the models on AFAD (Niu et al. 2016) training set and choose AFAD test set as the ID dataset for the OOD detection task. We take CIFAR10 (Krizhevsky, Hinton et al. 2009), SVHN (Netzer et al. 2011), MNIST (LeCun 1998), FashionMNIST (Xiao, Rasul, and Vollgraf 2017), Oxford-Pets (Parkhi et al. 2012) and Noise image generated by Pytorch (Paszke et al. 2019) (FakeData) as the OOD datasets. We employ the Areas Under the receiver operating Characteristic (***AUC***) and the Precision-Recall curve (***AUPR***) (higher is better for both) to evaluate OOD detection performance.

Results OOD detection results are shown in Tab. 1. DIDO performs the best on most datasets. On the Pets dataset, DIDO performs worse than DEns. and aleatoric uncertainty estimation head σ_{Θ_1}. We argue that images of pets provide features closer to facial information, resulting in higher evidence estimates given by DIDO. While σ_{Θ_1} performs better in this case, which can jointly make AuxUE a better uncertainty estimator. Overall, we consider that using generalized AuxUE with DIDO is an alternative that can better detect OOD inputs than ensembling-based solutions.

4.3 Monocular Depth Estimation Task

For the MDE task, we will evaluate both aleatoric and epistemic uncertainty estimation performance based on the AuxUE SLURP (Yu, Franchi, and Aldea 2021). Our generalized AuxUE is also constructed using SLURP as the backbone. We use BTS (Lee et al. 2019) as the main task model and KITTI (Geiger et al. 2013; Uhrig et al. 2017) Eigen-split (Eigen, Puhrsch, and Fergus 2014) training set for training both BTS and AuxUE models.

OOD sets	Metrics	Ours		Modified main DNN		
		σ_{Θ_1}	σ_{Θ_2}(DIDO)	LDU	Evi.	DEns.
CIFAR10	AUC ↑	96.0	**100**	95.2	50.0	99.2
	AUPR ↑	91.7	**100**	88.3	23.4	95.1
SVHN	AUC ↑	98.3	**100**	94.8	50.0	99.2
	AUPR ↑	98.1	**100**	93.2	44.3	97.8
MNIST	AUC ↑	97.8	**100**	97.6	50.0	99.6
	AUPR ↑	93.9	**100**	93.8	23.4	97.2
Fashion MNIST	AUC ↑	97.7	**100**	95.6	50.0	99.1
	AUPR ↑	94.0	**100**	89.3	23.4	93.8
Oxford Pets	AUC ↑	**82.9**	55.9	31.5	50.1	56.1
	AUPR ↑	**53.3**	23.9	12.5	18.5	21.3
Fake Data	AUC ↑	67.0	**80.8**	70.0	50.0	33.2
	AUPR ↑	59.7	**70.2**	58.8	49.5	37.8

Table 1: OOD detection results on Age estimation task. ID data is from the Asian Face Age Dataset (AFAD).

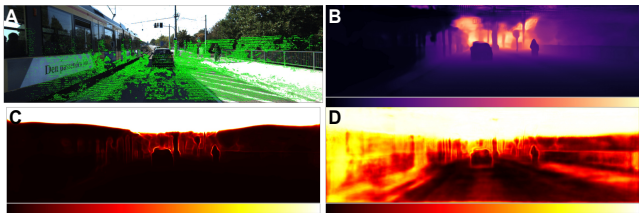

Figure 3: Illustrations on MDE. A: input image, green points represent pixels with depth ground truth; B: depth prediction; C and D: aleatoric and epistemic uncertainty estimations. In B, C, and D, brighter pixels correspond to higher values. The areas lacking depth ground truth, e.g. sky and tramway, are assigned high uncertainty using DIDO.

Aleatoric uncertainty estimation In this section, the goal is to analyze the fundamental performance and robustness of aleatoric uncertainty estimation under different distribution assumptions. We choose simple Gaussian (Sgau) (Nix and Weigend 1994), Laplacian (Lap), Generalized Gaussian (Ggau) (Upadhyay et al. 2022) and Normal-Inverse-Gamma (NIG) (Amini et al. 2020) distributions. We modify the loss functions and the head of the SLURP to output the desired parameters of the distributions.

Evaluation settings and datasets We first build Sparsification curves (SC) (Bruhn and Weickert 2006): we achieve predictive SC by computing the prediction error of the remaining pixels after removing a certain partition of pixels (5% in our experiment) each time according to the highest uncertainty estimations. We can also obtain an Oracle SC by removing the pixels according to the highest prediction errors. Then, we have the same metrics used in (Poggi et al. 2020): Area Under the Sparsification Error (***AUSE***, lower is better), and Area Under the Random Gain (***AURG***, higher is better). We choose absolute relative error (REL) and root mean square error (RMSE) as the prediction error metrics.

We generate KITTI-C from KITTI Eigen-split validation set using the code of ImageNet-C (Hendrycks and Dietterich 2019) to have different corruptions on the images to check the robustness of the uncertainty estimation solutions. We apply eighteen perturbations with five severities, including Gaussian noise, shot noise, etc., and take it along with the original KITTI for evaluation.

S	Metrics	Ggau	Sgau	NIG	Lap sep. enc. σ_{Θ_1}	Lap shar. enc. σ_{Θ_1}
0	AUSE-REL ↓	0.014	0.013	**0.012**	0.013	0.013
	AUSE-RMSE ↓	0.258	**0.202**	0.208	0.203	0.205
	AURG-REL ↑	0.023	0.023	**0.024**	0.023	0.023
	AURG-RMSE ↑	1.815	**1.871**	1.865	1.870	1.869
1	AUSE-REL ↓	0.021	0.019	**0.018**	0.019	**0.018**
	AUSE-RMSE ↓	0.482	**0.332**	0.335	0.336	**0.332**
	AURG-REL ↑	0.029	0.031	**0.032**	0.031	**0.032**
	AURG-RMSE ↑	2.215	**2.365**	2.362	2.361	**2.365**
2	AUSE-REL ↓	0.026	0.023	**0.022**	0.023	**0.022**
	AUSE-RMSE ↓	0.707	**0.463**	0.479	0.468	0.464
	AURG-REL ↑	0.035	**0.039**	**0.039**	0.038	**0.039**
	AURG-RMSE ↑	2.535	**2.779**	2.763	2.774	2.777
3	AUSE-REL ↓	0.036	**0.031**	**0.031**	**0.031**	**0.031**
	AUSE-RMSE ↓	1.176	0.737	0.806	**0.730**	0.749
	AURG-REL ↑	0.044	**0.049**	**0.049**	**0.049**	**0.049**
	AURG-RMSE ↑	2.862	3.301	3.232	**3.308**	3.289
4	AUSE-REL ↓	0.057	0.050	0.053	**0.049**	0.051
	AUSE-RMSE ↓	2.380	1.364	1.582	**1.268**	1.430
	AURG-REL ↑	0.051	0.058	0.054	**0.059**	0.056
	AURG-RMSE ↑	2.817	3.834	3.615	**3.929**	3.767
5	AUSE-REL ↓	0.082	0.064	0.069	**0.059**	0.066
	AUSE-RMSE ↓	3.878	2.043	2.414	**1.760**	2.157
	AURG-REL ↑	0.045	0.063	0.057	**0.067**	0.061
	AURG-RMSE ↑	2.377	4.213	3.842	**4.496**	4.098

Table 2: Aleatoric uncertainty estimation results on MDE. $S = 0$ represents original KITTI dataset and $S > 0$ represents KITTI-C datasets.

Results As shown in Tab. 2, the Laplace assumption is more robust when the severity increases, while Gaussian one works better when the noise severity is smaller. We also check the proposed generalized AuxUE with a shared encoder. It shows that the epistemic uncertainty estimation branch affects the robustness of aleatoric uncertainty estimation in this case, especially under stronger noise.

The next sections show epistemic uncertainty estimation results based on different methods. Furthermore, in Supp Tab. A15 and Tab. A16, we also verify whether aleatoric uncertainty methods based on different distribution assumptions can generalize to the OOD data, i.e., provide high uncertainty to the unseen patterns, even without explicitly modeling epistemic uncertainty.

Robustness under dataset change This experiment will explore the predictive uncertainty performance encountering the dataset change. Supervised MDE is an ill-posed problem that heavily depends on the training dataset. In our case, the main task model is trained on the KITTI dataset, so the model will output meaningless results on the indoor data, which should trigger a high uncertainty estimation. The results are shown in Tab. 3.

Evaluation settings and datasets We take ***AUC*** and ***AUPR*** as evaluation metrics. We take all the valid pixels from the KITTI validation set (ID) as the negative samples and the

Metrics	AuxUE with DIDO		Modified main DNN		
	Ours σ_{Θ_2} sep. enc.	Ours σ_{Θ_2} shar. enc.	LDU	Evi.	DEns.
AUC ↑	98.1	**98.4**	58.1	70.6	62.1
AUPR ↑	99.3	**99.4**	79.5	77.8	76.7

Table 3: Epistemic uncertainty estimation results encountering dataset change on MDE. The evaluation dataset here is NYU indoor depth dataset.

S	Metrics	AuxUE with DIDO		Modified main DNN		
		Ours σ_{Θ_2} sep. enc.	Ours σ_{Θ_2} shar. enc.	LDU	Evi.	DEns.
0	AUC ↑	**100.0**	99.9	96.5	76.7	93.5
	AUPR ↑	**100.0**	99.0	93.8	42.6	70.0
	Sky-All ↓	0.015	0.018	0.278	0.986	**0.005**
1	AUC ↑	**100.0**	99.9	96.3	69.7	92.8
	AUPR ↑	**99.9**	98.9	93.5	37.4	68.0
	Sky-All ↓	0.016	0.018	0.277	0.988	**0.005**
2	AUC ↑	**99.9**	**99.9**	95.9	65.4	92.3
	AUPR ↑	**99.8**	98.8	93.0	34.5	67.0
	Sky-All ↓	0.017	0.018	0.280	0.990	**0.005**
3	AUC ↑	**99.9**	99.7	95.9	62.3	91.6
	AUPR ↑	**99.7**	98.1	92.8	32.8	65.7
	Sky-All ↓	0.018	0.020	0.283	0.992	**0.005**
4	AUC ↑	**99.6**	99.5	96.1	58.8	91.8
	AUPR ↑	**99.1**	97.2	92.9	31.2	67.2
	Sky-All ↓	0.023	0.022	0.288	0.994	**0.005**
5	AUC ↑	98.5	**99.0**	96.5	58.5	92.2
	AUPR ↑	**97.1**	96.1	93.7	32.8	70.4
	Sky-All ↓	0.035	0.026	0.295	0.996	**0.005**

Table 4: Epistemic uncertainty estimation results encountering unseen pattern on MDE. The evaluation datasets here are KITTI Seg-Depth (S=0) and KITTI Seg-Depth-C (S>0).

valid pixels from the NYU (Nathan Silberman and Fergus 2012) validation set (OOD) as the positive samples.

Results Tab. 3 shows whether different uncertainty estimators can give correct indications facing the dataset change. The evidential learning method can provide competitive results, while our DIDO provides the best performance.

Robustness on unseen patterns during training This experiment focuses on how uncertainty estimators behave on unseen patterns during training. The unseen patterns are drawn from the same dataset distribution as the patterns used in training, and the outputs of the main task model for such patterns may be reasonable. Still, they cannot be evaluated and thus are unreliable. High uncertainty should be assigned to these predictions. Since this topic is rarely considered in MDE, we try to give a benchmark in this work.

Evaluation settings and datasets We select sky areas in KITTI as OOD patterns. This setting is based on the following reasons: due to the generalization ability of MDE DNNs, it is inappropriate to treat all pixels without ground truth as OOD. However, there is consistently no ground truth for the sky parts since LIDAR is used in depth acquisition. During training, sky patterns are masked and never seen by the DNNs (including the AuxUEs). Meanwhile, they are annotated in KITTI semantic segmentation dataset (Alhaija et al. 2018) (200 images), thus can be used for evaluation.

Three metrics are applied for evaluating OOD detection performance as shown in Tab. 4. ***AUC*** and ***AUPR***: we select 49 images that are not in the training set and have both depth and semantic segmentation annotations. For each image, we take the sky pixels as the positive class and the pixels with depth ground truth as the negative class. We use AUC and AUPR to assess the uncertainty estimation performance. Note that this metric does not guarantee that the uncertainty of the sky is the largest in the whole uncertainty map. Thus, we have ***Sky-All*** (lower is better): all 200 images with semantic segmentation annotations are selected for evaluation. The ground truth uncertainties are set as $\mathbf{1}$ for the sky areas. Then we normalize the predicted uncertainty, take the sky areas $\hat{\mathbf{u}}_{\text{sky}}$ from the whole uncertainty map and measure: $mean((\mathbf{1} - \hat{\mathbf{u}}_{\text{sky}})^2)$. For simplicity, we denote KITTI Seg-Depth for both evaluation datasets. We also generate a corruption dataset KITTI Seg-Depth-C using the same way in the aleatoric uncertainty estimation section.

Results Fig. 3 shows a qualitative example of typical uncertainty maps computed on KITTI images. More visualizations are presented in Supp Section E. In Tab. 4, Deep Ensembles can better assign consistent and higher uncertainty to the sky areas, but it is inadequate for identifying the ID and OOD areas. As outlined in Section 3.2, DIDO prioritizes rare patterns and then generalizes the uncertainty estimation ability to the unseen patterns. This results in assigning higher uncertainty to some few-shot pixels that have ground truth, making Sky-All results slightly worse. Yet, it can achieve a balanced performance on all the metrics, and at the same time, it maintains robust performance in the presence of noise.

4.4 Ablation Study

We conduct the ablation study in Supp Section D. **Hyperparameters.** We analyze the effect of the number of sets K defined in Section 3.2 for discretization and the regularization weight λ in Eq. 6. **Necessity of using AuxUE.** We also apply DIDO on the main task model to check impacts on main task performance. **Effectiveness of Dirichlet modeling.** We show the effectiveness of Dirichlet modeling instead of using the normal Categorical modeling based on the discretized prediction errors. For the former, we apply classical cross-entropy on the Softmax outputs given by the AuxUE.

5 Conclusion

In this paper, we propose a new solution for uncertainty quantification on regression problems based on a generalized AuxUE. We design and implement the experiments based on four different regression problems. By modeling heteroscedastic noise using Laplace distribution, the proposed AuxUE can achieve more robust aleatoric uncertainty. Meanwhile, the novel DIDO solution in our AuxUE can provide better epistemic uncertainty estimation performance on both image-level and pixel-wise tasks.

Acknowledgements

We acknowledge the support of the Saclay-IA computing platform. We also thank Mădălina Olteanu for the thought-provoking discussion for the article.

References

Abdar, M.; Pourpanah, F.; Hussain, S.; Rezazadegan, D.; Liu, L.; Ghavamzadeh, M.; Fieguth, P.; Cao, X.; Khosravi, A.; Acharya, U. R.; et al. 2021. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion*, 76: 243–297.

Alhaija, H.; Mustikovela, S.; Mescheder, L.; Geiger, A.; and Rother, C. 2018. Augmented Reality Meets Computer Vision: Efficient Data Generation for Urban Driving Scenes. *International Journal of Computer Vision (IJCV)*.

Amini, A.; Schwarting, W.; Soleimany, A.; and Rus, D. 2020. Deep evidential regression. *NeurIPS*.

Arnez, F.; Espinoza, H.; Radermacher, A.; and Terrier, F. 2020. A comparison of uncertainty estimation approaches in deep learning components for autonomous vehicle applications. *arXiv preprint arXiv:2006.15172*.

Besnier, V.; Bursuc, A.; Picard, D.; and Briot, A. 2021. Triggering Failures: Out-Of-Distribution detection by learning from local adversarial attacks in Semantic Segmentation. In *ICCV*.

Bhat, S. F.; Alhashim, I.; and Wonka, P. 2021. Adabins: Depth estimation using adaptive bins. In *CVPR*.

Bishop, C.; and Quazaz, C. 1996. Regression with input-dependent noise: A Bayesian treatment. *NeurIPS*.

Bishop, C. M.; and Nasrabadi, N. M. 2006. *Pattern recognition and machine learning*, volume 4. Springer.

Blundell, C.; Cornebise, J.; Kavukcuoglu, K.; and Wierstra, D. 2015. Weight uncertainty in neural network. In *ICML*.

Bruhn, A.; and Weickert, J. 2006. A confidence measure for variational optic flow methods. *Computational Imaging and Vision*, 31: 283.

Cao, W.; Mirjalili, V.; and Raschka, S. 2020. Rank consistent ordinal regression for neural networks with application to age estimation. *Pattern Recognition Letters*, 140: 325–331.

Cao, Y.; Wu, Z.; and Shen, C. 2017. Estimating depth from monocular images as classification using deep fully convolutional residual networks. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(11): 3174–3182.

Charpentier, B.; Borchert, O.; Zügner, D.; Geisler, S.; and Günnemann, S. 2022. Natural Posterior Network: Deep Bayesian Predictive Uncertainty for Exponential Family Distributions. In *ICLR*.

Charpentier, B.; Zügner, D.; and Günnemann, S. 2020. Posterior network: Uncertainty estimation without ood samples via density-based pseudo-counts. *NeurIPS*.

Corbière, C.; Lafon, M.; Thome, N.; Cord, M.; and Pérez, P. 2021. Beyond First-Order Uncertainty Estimation with Evidential Models for Open-World Recognition. In *ICML 2021 Workshop on Uncertainty and Robustness in Deep Learning*.

Corbière, C.; Thome, N.; Bar-Hen, A.; Cord, M.; and Pérez, P. 2019. Addressing failure prediction by learning model confidence. In *NeurIPS*.

Dempster, A. P. 1968. A generalization of Bayesian inference. *Journal of the Royal Statistical Society: Series B (Methodological)*, 30(2): 205–232.

Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. 2021. An image is worth 16x16 words: Transformers for image recognition at scale. *ICLR*.

Eigen, D.; Puhrsch, C.; and Fergus, R. 2014. Depth map prediction from a single image using a multi-scale deep network. *NeurIPS*.

Foong, A. Y.; Li, Y.; Hernández-Lobato, J. M.; and Turner, R. E. 2019. 'In-Between'Uncertainty in Bayesian Neural Networks. *arXiv preprint arXiv:1906.11537*.

Franchi, G.; Yu, X.; Bursuc, A.; Aldea, E.; Dubuisson, S.; and Filliat, D. 2022. Latent Discriminant deterministic Uncertainty. In *ECCV*.

Gal, Y.; and Ghahramani, Z. 2016. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *ICML*.

Geiger, A.; Lenz, P.; Stiller, C.; and Urtasun, R. 2013. Vision meets Robotics: The KITTI Dataset. *International Journal of Robotics Research (IJRR)*.

Goldberg, P.; Williams, C.; and Bishop, C. 1997. Regression with input-dependent noise: A Gaussian process treatment. *NeurIPS*.

Guo, Z.; Wan, Z.; Zhang, Q.; Zhao, X.; Chen, F.; Cho, J.-H.; Zhang, Q.; Kaplan, L. M.; Jeong, D. H.; and Jøsang, A. 2022. A Survey on Uncertainty Reasoning and Quantification for Decision Making: Belief Theory Meets Deep Learning. *arXiv preprint arXiv:2206.05675*.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *CVPR*.

Hendrycks, D.; and Dietterich, T. 2019. Benchmarking Neural Network Robustness to Common Corruptions and Perturbations. In *ICLR*.

Hendrycks, D.; and Gimpel, K. 2017. A Baseline for Detecting Misclassified and Out-of-Distribution Examples in Neural Networks. *ICLR*.

Hornauer, J.; and Belagiannis, V. 2022. Gradient-Based Uncertainty for Monocular Depth Estimation. In *ECCV*.

Hüllermeier, E.; and Waegeman, W. 2021. Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning*, 110: 457–506.

Jain, M.; Lahlou, S.; Nekoei, H.; Butoi, V.; Bertin, P.; Rector-Brooks, J.; Korablyov, M.; and Bengio, Y. 2021. Deup: Direct epistemic uncertainty prediction. *arXiv preprint arXiv:2102.08501*.

Joo, T.; Chung, U.; and Seo, M.-G. 2020. Being bayesian about categorical probability. In *ICML*.

Jordan, M. I.; Ghahramani, Z.; Jaakkola, T. S.; and Saul, L. K. 1999. An introduction to variational methods for graphical models. *Machine learning*, 37: 183–233.

Kamann, C.; and Rother, C. 2021. Benchmarking the robustness of semantic segmentation models with respect to common corruptions. *IJCV*.

Kendall, A.; and Gal, Y. 2017. What uncertainties do we need in bayesian deep learning for computer vision? In *NeurIPS*.

Krizhevsky, A.; Hinton, G.; et al. 2009. Learning multiple layers of features from tiny images. Technical report.

Lakshminarayanan, B.; Pritzel, A.; and Blundell, C. 2017. Simple and scalable predictive uncertainty estimation using deep ensembles. In *NeurIPS*.

LeCun, Y. 1998. The MNIST database of handwritten digits. *https://yann.lecun.com/exdb/mnist/*.

Lee, J. H.; Han, M.-K.; Ko, D. W.; and Suh, I. H. 2019. From big to small: Multi-scale local planar guidance for monocular depth estimation. *arXiv preprint arXiv:1907.10326*.

Liang, S.; Li, Y.; and Srikant, R. 2018. Enhancing The Reliability of Out-of-distribution Image Detection in Neural Networks. In *ICLR*.

Malinin, A.; Chervontsev, S.; Provilkov, I.; and Gales, M. 2020. Regression prior networks. *arXiv preprint arXiv:2006.11590*.

Malinin, A.; and Gales, M. 2018. Predictive uncertainty estimation via prior networks. *NeurIPS*.

Malinin, A.; and Gales, M. 2019. Reverse kl-divergence training of prior networks: Improved uncertainty and adversarial robustness. *NeurIPS*.

Marks, R. J.; Wise, G. L.; Haldeman, D. G.; and Whited, J. L. 1978. Detection in Laplace noise. *IEEE Transactions on Aerospace and Electronic Systems*, 866–872.

Mi, L.; Wang, H.; Tian, Y.; and Shavit, N. 2022. Training-Free Uncertainty Estimation for Dense Regression: Sensitivity as a Surrogate. In *AAAI*.

Michaelis, C.; Mitzkus, B.; Geirhos, R.; Rusak, E.; Bringmann, O.; Ecker, A. S.; Bethge, M.; and Brendel, W. 2019. Benchmarking robustness in object detection: Autonomous driving when winter is coming. *arXiv preprint arXiv:1907.07484*.

Nadarajah, S. 2005. A generalized normal distribution. *Journal of Applied statistics*, 32(7): 685–694.

Nathan Silberman, P. K., Derek Hoiem; and Fergus, R. 2012. Indoor Segmentation and Support Inference from RGBD Images. In *ECCV*.

Netzer, Y.; Wang, T.; Coates, A.; Bissacco, A.; Wu, B.; and Ng, A. Y. 2011. Reading Digits in Natural Images with Unsupervised Feature Learning. In *NeurIPS*.

Niu, Z.; Zhou, M.; Wang, L.; Gao, X.; and Hua, G. 2016. Ordinal Regression With Multiple Output CNN for Age Estimation. In *CVPR*.

Nix, D.; and Weigend, A. 1994. Estimating the mean and variance of the target probability distribution. In *ICNN*.

Oh, D.; and Shin, B. 2022. Improving evidential deep learning via multi-task learning. In *AAAI*.

Parkhi, O. M.; Vedaldi, A.; Zisserman, A.; and Jawahar, C. V. 2012. Cats and Dogs. In *CVPR*.

Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; Desmaison, A.; Kopf, A.; Yang, E.; DeVito, Z.; Raison, M.; Tejani, A.; Chilamkurthy, S.; Steiner, B.; Fang, L.; Bai, J.; and Chintala, S. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *NeurIPS*.

Poggi, M.; Aleotti, F.; Tosi, F.; and Mattoccia, S. 2020. On the uncertainty of self-supervised monocular depth estimation. In *CVPR*.

Qu, H.; Li, Y.; Foo, L. G.; Kuen, J.; Gu, J.; and Liu, J. 2022. Improving the reliability for confidence estimation. In *ECCV*.

Rezende, D.; and Mohamed, S. 2015. Variational inference with normalizing flows. In *ICML*.

Sensoy, M.; Kaplan, L.; and Kandemir, M. 2018. Evidential deep learning to quantify classification uncertainty. *NeurIPS*.

Shen, M.; Bu, Y.; Sattigeri, P.; Ghosh, S.; Das, S.; and Wornell, G. 2023. Post-hoc Uncertainty Learning using a Dirichlet Meta-Model. In *AAAI*.

Techapanurak, E.; and Okatani, T. 2021. Practical evaluation of out-of-distribution detection methods for image classification. *arXiv preprint arXiv:2101.02447*.

Uhrig, J.; Schneider, N.; Schneider, L.; Franke, U.; Brox, T.; and Geiger, A. 2017. Sparsity Invariant CNNs. In *3DV*.

Ulmer, D. 2021. A survey on evidential deep learning for single-pass uncertainty estimation. *arXiv preprint arXiv:2110.03051*.

Upadhyay, U.; Karthik, S.; Chen, Y.; Mancini, M.; and Akata, Z. 2022. BayesCap: Bayesian Identity Cap for Calibrated Uncertainty in Frozen Neural Networks. In *ECCV*.

Wen, Y.; Tran, D.; and Ba, J. 2020. BatchEnsemble: an alternative approach to efficient ensemble and lifelong learning. In *ICLR*.

Xiao, H.; Rasul, K.; and Vollgraf, R. 2017. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*.

Yeo, T.; Kar, O. F.; and Zamir, A. 2021. Robustness via cross-domain ensembles. In *CVPR*.

Yu, X.; Franchi, G.; and Aldea, E. 2021. SLURP: Side Learning Uncertainty for Regression Problems. In *BMVC*.

Yu, X.; Franchi, G.; and Aldea, E. 2022. On Monocular Depth Estimation and Uncertainty Quantification Using Classification Approaches for Regression. In *ICIP*.