Continual Evidential Deep Learning for Out-of-Distribution Detection

Eduardo Aguilar[123]*, Bogdan Raducanu[2], Petia Radeva[23], Joost Van de Weijer[2]

[1]Dept. de Ingeniería de Sistemas y Computación, Universidad Católica del Norte, Antofagasta, Chile
[2]Computer Vision Center, Universitat Autònoma de Barcelona, Barcelona, Spain
[3]Dept. de Matemàtiques i Informàtica, Universitat de Barcelona, Barcelona, Spain

`eduardo.aguilar@ucn.cl`, `petia.ivanova@ub.edu`, `{bogdan,joost}@cvc.uab.es`

Abstract

Uncertainty-based deep learning models have attracted a great deal of interest for their ability to provide accurate and reliable predictions. Evidential deep learning stands out achieving remarkable performance in detecting out-of-distribution (OOD) data with a single deterministic neural network. Motivated by this fact, in this paper we propose the integration of an evidential deep learning method into a continual learning framework in order to perform simultaneously incremental object classification and OOD detection. Moreover, we analyze the ability of vacuity and dissonance to differentiate between in-distribution data belonging to old classes and OOD data. The proposed method [1], called CEDL, is evaluated on CIFAR-100 considering two settings consisting of 5 and 10 tasks, respectively. From the obtained results, we could appreciate that the proposed method, in addition to provide comparable results in object classification with respect to the baseline, largely outperforms OOD detection compared to several posthoc methods on three evaluation metrics: AUROC, AUPR and FPR95.

1. Introduction

Deep learning has reached amazing accuracy in a vast number of applications such as image segmentation [23], image retrieval [5], image generation [47], 3D image reconstruction [30], etc. sometimes outperforming humans in either strategy games [27] or medical image diagnostics [61]. However, for trustworthy AI methods, it is crucial that the developed algorithms are able to assess the uncertainty of their estimations. Therefore, the field of uncertainty estimation in neural networks has seen a growing research interest [1], including methods based on Bayesian Neural Networks [36], Deep Ensembles [44] and Single Deterministic Neural Networks [52].

---

*Corresponding author.
[1]Code will be made available at `https://github.com/Eaaguilart/cedl`.

Among the methods to estimate the uncertainty of neural networks, *deep evidential learning* [48] stands out because of its ability to pinpoint the various sources of uncertainty: the method can distinguish between a lack of confidence (as measured by *vacuity*) and conflicting evidence (as measured by *dissonance*) [60, 16]. Within this approach, the network's output is used to set the parameters of a Dirichlet distribution on the class probabilities. These parameters can then be used to assess the uncertainty of the network. The method has been successfully evaluated for the task of OOD detection [48]. However, the method assumes that all the available data is jointly available and can therefore perform the training under the i.i.d. assumption. This assumption is not realistic for many real-world applications, where data arrives sequentially, presenting a distribution shift over time [7, 39]. In this context, the capability of a model to detect OOD samples has become a necessity in order to guarantee the robustness and safety of these systems.

On the other hand, continual learning is concerned with the sequential learning from non-stationary data (the data is provided as a sequence of *tasks*), where the learner has to balance the trade-off between the acquisition of new knowledge while consolidating the previous one, a constraint known as *plasticity-stability dilemma* [40]. The main problem of continual learning is represented by *catastrophic forgetting* [24], which refers to a significant drop in performance on previous tasks when learning a new task. Generally, three main strategies exist to address this highly undesirable effect: regularization methods [24, 29], rehearsal methods [45, 35], and parameter isolation methods [49]. These methods are mainly used in combination with a standard cross-entropy loss for the classification task, and when evaluating, a closed-world protocol is applied (meaning evaluation is restricted to classes seen during training). Evaluating these methods within an open-world assumption has only recently received some attention [2, 20, 19]. However, to the best of our knowledge, *continual evidential deep learning* has not yet been explored in depth.

In this paper, we investigate to what extent an Evidential

Deep Learning method embedded in a Continual Learning framework allows to detect OOD data and to differentiate it with respect to in-distribution data from old classes. This way, we could better handle different types of data during the learning process and also improve the robustness of the model against OOD data. Our main contributions are the following:

- We are the first to integrate evidential deep learning into a continual learning approach to simultaneously perform incremental object classification and OOD detection. To this end, we propose a new loss function that combines cross-entropy loss, knowledge distillation and evidential KL-divergence.

- We analyze the ability of vacuity and dissonance to differentiate in-distribution data belonging to old classes from OOD data. Interestingly, we find that vacuity provides a good measure for OOD data detection, while dissonance struggles to distinguish in-distribution (previous classes) data from unseen classes.

- With the proposed CEDL method, we have largely outperformed several state-of-the-art posthoc methods used for OOD detection.

2. Related work

We will briefly describe the main related works in uncertainty estimation and continual learning.

Uncertainty Estimation. Uncertainty occurs mainly in two types: epistemic uncertainty, produced by the lack of training data; and aleatoric uncertainty, related to the random noise produced during the sample collection [18]. To estimate uncertainty, there are several deep neural methods, designed mainly for classification and regression problems [9], such as: Bayesian methods, which learn a distribution over the weights [4]; Ensemble methods, which arise from their interpretation of Bayesian methods as an ensemble of thin networks with shared weights [28]; Test-time augmentation methods, which perform several forward passes by randomly modifying the input data to perform prediction and estimate uncertainty [53]; and singles deterministic methods, such as those based on evidence theory [48].

Uncertainty-aware deep learning models have been successfully developed in the past in a continual learning framework, but with different objectives with respect to this work. Considering that uncertainty is a natural way to identify what to remember and what to change as we continually learn, various works have proposed approaches [42, 8, 34] to approximate Bayesian learning with the goal of mitigating catastrophic forgetting by exploiting uncertainty to regularize the parameters of neural networks. On the other hand, single deterministic methods capable of estimating

uncertainty have also been proposed as an effective and efficient alternative to Bayesian methods [54, 14]. In [54], the performance degradation resulting from the distribution shift that occurs due to the limited number of stored samples from the previous task was analyzed. For this purpose, they propose a method based on the Diritchet Priority Network [38] to model the distribution shift and minimize the bias introduced during the rehearsal process. Recently, an Evidential deep learning approach was proposed for the class-based incremental semantic segmentation problem [14]. Here the authors propose to use the estimated uncertainty as a probability measure for the background class, justifying that this class shifts after each increment (past and future classes are correlated). In our work we are going to adopt and evaluate an Evidential deep learning approach to carry out the OOD detection problem.

Continual Learning. There are several surveys on continual learning, focussing on task-incremental learning [7], class-incremental learning (CIL) [39], online continual learning [37], and categorizing underlying biological motivation [43]. We will here shortly focus on the main strategies developed in continual learning that we apply within the context of this paper.

Regularization is one of the main strategies to prevent forgetting. Weight regularization methods compute a prior importance for all parameters in the network [24, 33, 58]. This prior is then applied to regularize the weights when training new tasks. On the other hand, functional regularization methods [29, 15] apply a regularization on the network output. This regularization is in general based on knowledge distillation [13]. Another popular strategy is data rehearsal, where a small set of data from previous tasks is stored in a buffer, and replayed during the training of new tasks [35, 45]. In general, this method is very efficient to counter forgetting, however, it does increase the memory requirements of the system. For settings, where due to privacy concerns or legislation, the storing of data is prohibited, pseudo-rehearsal methods [50, 55] have been developed which use a generator to replay data from previous tasks.

In this paper, we propose a method for continual evidential learning that incorporates both regularization and rehearsal to counter catastrophic forgetting.

Continual Out-of-Distribution Detection. OOD has been extensively studied for some time in the classical machine learning setting (when all the available training data is provided in advance) [12, 31, 51]. A comprehensive survey on OOD detection could be found in [57] and recently a generalized benchmark has been proposed [56] in order to unify the evaluation methodologies of several existing methods.

However, the study of OOD in the more realistic context of continual learning only recently started to receive

attention from the research community, where most of the work adopted the CIL setting. One of the first works who pioneered this research direction is presented in [2], where they established benchmarks for the continual OOD detection problem. Additionally, their analysis provided new insights on how potentially forgotten samples are treated by the OOD detection methods.

A novel approach is proposed in [20] where they use exemplars to learn a new and independent classifier that is able to detect OOD samples for each task. In [10] they detect OOD samples by correcting output bias towards new classes and enhancing output confidence difference based on task discriminativeness. In [19] an unified approach for both class-incremental and task-incremental settings is proposed, using a dual mechanism: (i) task masking to overcome catastrophic forgetting, and (ii) a learning method for building a model for each task based on OOD detection. A different approach is adopted in [46], where they introduce a self-supervised continual novelty detector, which builds incrementally a statistical model over the space of intermediate features produced by a deep network, and utilizes feature reconstruction errors as uncertainty scores to guide the detection of novel samples.

In [22], the authors presented a theoretical study according to which an improved CIL performance is strongly correlated with a good OOD detector. This theoretical study has been further extended in [21], where they show that CIL is indeed learnable.

3. Methodology

In this section we briefly introduce the evidential deep learning method and then explain our integration of it into a continual learning approach.

3.1. Evidential Deep Learning

Deep Learning methods for object recognition often use softmax activation in the output layers to perform classification and provide a measure of confidence. However, it has been demonstrated that the softmax outputs are biased to the data used for training, providing a high probability even for wrong predictions [41]. To provide a reliable prediction, it is desirable to apply Deep Learning methods that are able to quantify the uncertainty in the predictions. *Evidential Deep Learning* (EDL) [48] is a method developed for this purpose and is able to provide prediction uncertainty using a single deterministic neural network. EDL quantifies belief masses and uncertainty using subjective logic theory, which formalizes belief assignments over a discernment framework as a Dirichlet distribution.

In the evidential framework, it is assumed that the class probability for a sample i is drawn from a prior Dirichlet distribution, given by:

$$\text{Dir}(p, \alpha) = \frac{\Gamma(S)}{\prod_{c=1}^{C} \Gamma(\alpha_c)} \prod_{c=1}^{C} p_i^{\alpha_c - 1}; \alpha_c > 0, \quad (1)$$

where p is a probability mass function, C is the number of classes, $\alpha = \{\alpha_1, \ldots, \alpha_c\}$ are the Dirichlet parameters, $\Gamma(.)$ is the gamma function and $S = \sum_{c=1}^{C} \alpha_c$ is referred to as the Dirichlet strength. As for α_c, it is calculated as the sum of the evidence of the c-th class (e_c) with one. The evidence is learned by the model and obtained from the logit layer after applying on it a non-negative function such as ReLU [48] or Exponential [3].

The training of an EDL model for classification is very similar to that of a classical neural network. The main difference is the activation used in the output layer, where softmax is replaced by a non-negative activation. Then, the model can be trained by minimizing the evidential cross-entropy loss (L_{ECE}) to form the multi-nomial opinions (learn the evidence) for C-class classification of a given sample i as a Dirichlet distribution. The L_{ECE} is formally defined as follows [48, 3]:

$$L_{ECE} = \sum_{j=1}^{C} y_{ij}(log(S_i) - log(\alpha_{ij})) \quad (2)$$

where y_i is the one-hot vector encoding of the ground-truth label of the i-th sample.

Note that when the neural network finds shared patterns between different classes, it may generate some evidence about incorrect labels and thus minimize the loss function, L_{ECE}. To avoid this behavior, the authors in [48] propose a loss function, L_{EKL} which acts as a regularization with the objective that the total evidence decreases to zero for a sample, if it cannot be correctly classified. The L_{EKL}, can be defined as follows:

$$L_{EKL} = \text{KL}[\text{Dir}(p_i|\tilde{\alpha}_i)||Dir(p_i|\mathbf{1})] =$$
$$log\left(\frac{\Gamma(\sum_{c=1}^{C} \tilde{\alpha}_{ic})}{\Gamma(C)\prod_{c=1}^{C} \Gamma(\tilde{\alpha}_{ic})}\right) +$$
$$\sum_{c=1}^{C} (\tilde{\alpha}_{ic} - 1)\left[F(\tilde{\alpha}_{ic}) - F(\sum_{j=1}^{C} \tilde{\alpha}_{ij})\right] \quad (3)$$

where $\text{KL}[.||.]$ corresponds to the Kullback-Leibler divergence, $\tilde{\alpha}_i$ are the Dirichlet parameters after removal of the non-misleading evidence and $F(.)$ is the digamma function.

After training the model, uncertainty and class prediction can be obtained. In evidential theory, there are several measures of uncertainty [17, 60], among them are: *Vacuity*, measuring the lack of evidence; and *Dissonance*, measuring the conflict of evidences. These uncertainty measures are detailed as follows:
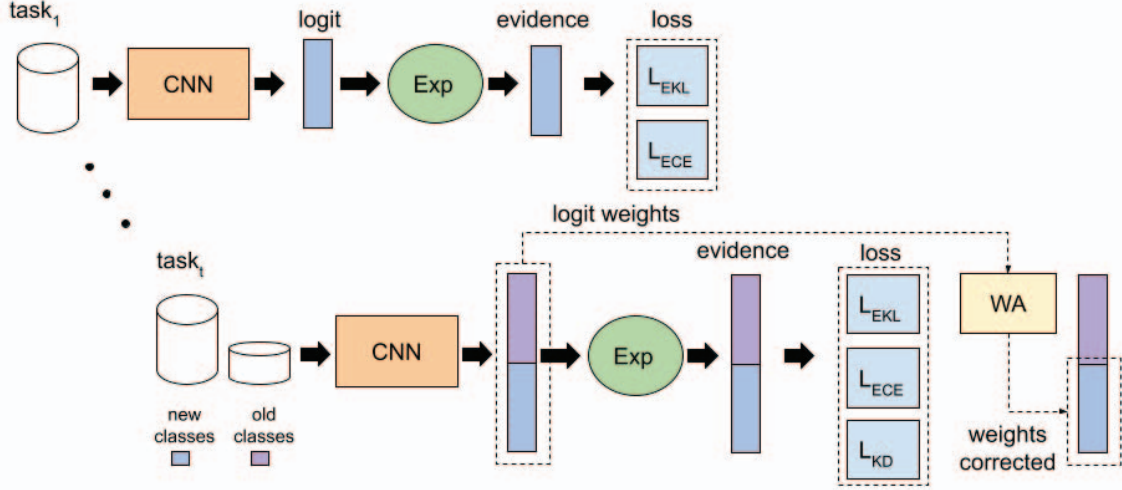
$$\text{Vac}(\alpha) = \frac{C}{S} \quad (4)$$

Figure 1: Overview of the proposed Continual Evidential Deep Learning approach.

$$\text{Diss}(\alpha) = \sum_{c=1}^{C} b_c \frac{\sum_{i \neq c} b_i \text{Bal}(b_i, b_c)}{\sum_{i \neq c} b_i} \tag{5}$$

$$\text{Bal}(b_i, b_c) = \begin{cases} 1 - \frac{|b_i - b_c|}{b_i + b_c} & \text{if } b_i b_c \neq 0 \\ 0 & \text{else} \end{cases}$$

where $b = \frac{e}{S}$ is the belief mass.

3.2. Continual Evidential Deep Learning

The proposed Continual EDL approach (see Fig. 1) integrates the EDL method with the continual deep learning strategies for exemplar rehearsal and knowledge distillation to counter forgetting. Our method first applies rehearsal of data and knowledge distillation to transfer knowledge from the old model to the new one. Then motivated by Zhao et al. [59], in a second phase, taking into account that the model still has a prediction bias towards the new classes, a Weight Aligning (WA) method is applied to correct the model trained in the first phase.

To integrate both methods, several changes in the Continual Deep Learning architecture and loss function are necessary. Specifically, inspired by [48], the softmax activation is replaced by an Exponential (Exp) activation that appears more stable than the ReLU proposed in the original EDL implementation. Moreover, the minimization of the loss function, L_{ECE} considering the evidence obtained after applying Exp activation seems similar to the typical softmax cross-entropy minimization used in classical neural networks, which ensures comparable results in terms of accuracy, but with extra information about the confidence of the prediction. As for the loss function, L_{EKL} is added to preserve the model's ability to capture uncertainty. The proposed loss function is defined as follows:

$$L = \lambda_1 L_{ECE} + \lambda_2 L_{EKL} + \lambda_3 L_{KD} \tag{6}$$

where λ_1, λ_2 and λ_3 weight the importance of each loss function and L_{KD} is the knowledge distillation loss defined as:

$$L_{KD} = \text{KL}[p^s(\tau)||p^t(\tau)] =$$
$$\sum_{c=1}^{C_{old}} p_c^t(\tau) \times (log(p_c^t(\tau)) - log(p_c^s(\tau))) \tag{7}$$

with a τ parameter that acts as a temperature.

The L_{KD} is incorporated to the loss function, L from the second task training. It is interesting to note that the proposed loss function contemplates two regularizations, one to maintain the knowledge acquired by the old model, (L_{KD}) and the other - to avoid providing low uncertainty in misclassified data (L_{EKL}). We argue that since L_{KD} allows transferring knowledge about evidence of old classes, in a continual learning approach L_{EKL} should focus only on the new classes. Therefore, the L_{EKL} for the t-th task is redefined as follows:

$$L_{EKL} = \text{KL}[\text{Dir}(p_{it}|\tilde{\alpha_{it}})||Dir(p_{it}|\mathbf{1})] \tag{8}$$

where p_{it} and α_{it} are the probability mass function and Dirichlet parameters related to the C_{new} new classes for task t and sample i, respectively.

Finally, in the same way as [10], we apply a bias correction (BC) method to reduce the output value towards biased classes. For this purpose, during inference time, the output logits are normalized based on the norm of the weight vectors in the classifier corresponding to each learned class.

4. Validation

In this section, we describe the experimental setting, the evaluation procedure and metrics to validate the proposed method.

4.1. Experimental Settings

We follow the protocol for OOD detection evaluation used in [10] to compare the proposed approach with existing posthoc methods such as: MSP [12], ODIN [31], Energy [32], Entropy [26] and MSP_BC_CE [10]. These posthoc methods are applied on top of the baseline model corresponding to the continual learning approach published in [59]. For all experiments, ResNet32 [11] is used as backbone and trained for 120 epochs with a batch size of 128, for each incremental step. Optimization is performed using SGD with an initial learning rate of $1e - 1$ that is reduced following a cosine annealing schedule, a momentum of 0.9 and a weight decay of $5e - 4$.

The methods are evaluated on CIFAR-100 [25], a widely used dataset in continual learning. The order of the classes is shuffled according to [54]. Then, the dataset is divided in 5 and 10 tasks with 20 and 10 base classes and incremental step of the same size, respectively.

As for rehearsal, we use a growing buffer that stores 20 samples per class, herded as [45] did. Regarding the knowledge distillation, a temperature of $T = 2$ is used. With respect to the loss function, the weighting parameters were set empirically to ensure a classification performance close to the baseline. Specifically, $\lambda_1 = 0.5$ and $\lambda_2 = 0.5$ for the first task, and $\lambda_1 = 0.45$, $\lambda_2 = 0.5$ and $\lambda_3 = 0.05$ for the remaining tasks. Finally, to avoid overfitting during training, the RandAugment [6] method is used for data augmentation.

All experiments were performed using PyTorch library on a computer with a NVIDIA RTX 2080 TI graphics card.

4.2. Evaluation

Detecting OOD in an unsupervised continual learning scenario involves distinguishing between unlabeled data from previously learned tasks (in-distribution or IND) and data from the new tasks (OOD) [10]. On the other hand, knowing whether new samples belong to old (IND_f), current (IND_c) or unseen classes (OOD) may be convenient in order to treat them differently during learning in a continuous learning framework [2]. Based on this, we evaluated OOD detection performance before each incremental learning step as follows:

- IND vs. OOD: Ability to discriminate samples of previously learned classes with respect to samples of unseen classes.
- IND_c vs. OOD: Ability to discriminate samples of current classes with respect to samples of unseen classes. Useful in determining how catastrophic forgetting affects OOD detection.
- IND_f vs. OOD: Ability to discriminate between samples of old classes with respect to unseen classes.

	Step size 10		Step size 20	
Method	ACA	AIA	ACA	AIA
Baseline	50.96%	63.87%	56.36%	68.02%
CEDL	51.07%	63.16%	55.61%	66.27%

Table 1: Classification results in terms of ACA and AIA.

- IND_c vs. IND_f: Ability to discriminate samples of recently learned classes with respect to old classes.

Metrics. For comparative purposes, we selected the most commonly used metrics to assess both classification and OOD detection. In the case of classification, the average classification accuracy (ACA) corresponds to the accuracy of all classes seen after the t-th task; and the average incremental accuracy (AIA) corresponds to the average accuracy over all tasks. In the case of OOD detection, OOD is the negative class and the rest are considered as positive classes, with the exception of the IND_c vs IND_f evaluation, where IND_f is treated as negative class. With this in mind, in general, OOD detection performance is calculated as a binary classification problem using the area under the receiver operating characteristic curve (AUROC), the area under the accuracy-recall curve (AUPR) and the false positive rate at the 95% of recall (FPR95). A higher value of AUROC and AUPR, and a lower value of FPR95, indicates better detection performance.

5. Results

In this section, the experimental results related to the detection of OOD data and the analysis of two uncertainty measures, vacuity and dissonance, are presented.

5.1. Continual OOD detection

In this work we perform OOD detection after each incremental step, considering two settings of 5 and 10 tasks, respectively. Although this work focuses primarily on the OOD detection problem and not on object classification, as we can see in Table 1, our approach provides comparable or slightly inferior results with respect to the baseline in terms of ACA and AIA on the test data of the classes seen. But, at the same time, it also provides the prediction's uncertainty without requiring additional computational resources.

Regarding the detection of OOD data, in our proposal we used the estimated uncertainty related to the prediction of each sample to perform this detection. Specifically through the vacuity measure, as the performance vastly outperforms the dissonance measure in detecting OOD (see Table 2). High vacuity means higher probability for OOD data and low vacuity is related to IND data. The obtained results are reported in terms of FPR95, AUPR and ROC for the pro-

	Step size 10			Step size 20		
Method	AUROC ↑	AUPR ↑	FPR95 ↓	AUROC ↑	AUPR ↑	FPR95 ↓
MSP [12]	61.14%	86.62%	91.91%	66.34%	81.87%	90.09%
ODIN [31]	60.23%	86.25%	92.43%	65.39%	81.33%	90.73%
Energy Score [32]	60.85%	86.10%	91.29%	65.03%	80.46%	90.76%
Entropy [26]	61.72%	86.76%	91.05%	66.71%	82.07%	89.74%
MSP-BC-CE [10]	62.47%	87.11%	91.14%	67.18%	82.14%	90.00%
CEDL (dissonance)	55.74%	84.60%	95.76%	57.04%	77.40%	95.96%
CEDL (vacuity)	**64.44%**	**87.24%**	**85.87%**	**72.42%**	**84.69%**	**76.09%**

Table 2: Average AUROC, AUPR and FPR95 over all tasks with step size of 10 and 20. Best results are in bold.

Evaluation	Task	CEDL (vac.)	MSP-BC-CE
IND_f vs. OOD	2	85.70%	93.15%
	3	91.10%	94.35%
	4	89.75%	92.60%
IND_c vs. OOD	1	52.05%	85.65%
	2	51.85%	83.25%
	3	70.15%	85.60%
	4	73.15%	82.45%
IND_c vs. IND_f	2	73.55%	88.40%
	3	80.20%	88.63%
	4	84.45%	87.80%
	5	91.39%	90.81%

Table 3: Performance evaluation in terms of FPR95 between different subsets of data (IND_c, IND_f and OOD).

posed method and also for the state-of-the-art postdoc methods. The evaluation is performed following the protocol proposed in [10], where after each task the OOD detection is performed taking into account that the test data corresponding to the learned classes are considered as IND data and the unseen classes as OOD data. The average of the results of all tasks for each metric and for two incremental step sizes is shown in Table 2 for all methods. Here, we highlight the effectiveness of the proposed method, achieving a wide performance improvement in all metrics with respect to all state-of-the-art methods. Particularly, a noticeable improvement is observed for the incremental step size of 20, where a difference of 14% of FPR95 is achieved with respect to the second-best model (MSP-BC-CE), which means that our method is able to misclassify less OOD data as IND when recall is high. In addition, with respect to the AUROC and AUPR metrics, the improvement is approximately 5.0% and 2.5%, respectively.

Motivated by [2], the IND data is divided into IND_c and IND_f to analyze the ability of the methods to distinguish between data belonging to old classes (IND_f), current classes (IND_c) and unseen classes (OOD). In Table 3, we present

the results of the best and second best method in terms of FPR95 for the IND_f vs. OOD, IND_c vs. OOD and IND_c vs. IND_f evaluations. It can be seen from the table that it is very difficult to differentiate between the data belonging to old classes with respect to data from current or unseen classes. However, in the first tasks our proposal offers a promising performance (less than 85%) for detecting IND_f data. Moreover, we can observe that the proposed method loses its good ability to distinguish between different subsets of data as the training progresses. Despite this, the proposed method is still better than MSP-BC-CE. In the case of IND vs. OOD evaluation, we observe that our proposal provides a great performance improvement with respect to MSP-BC-CE, in some tasks over 30% difference. This performance declines when all INDs are considered, highlighting that quantified uncertainty suffers from the catastrophic forgetting of old classes.

5.2. Vacuity and dissonance analysis

Figure 2 shows how vacuity (top) and dissonance (bottom) change after training each task. In this experiment we use a 5-task split. Here, the resulting model after training on each task is used to quantify the uncertainty in the test data for all tasks. As expected, the uncertainty of the test data belonging to the classes of the current task is lower than that of the rest, as measured by both vacuity and dissonance. In addition, it is observed that the vacuity of the data from old classes and current classes is lower than unseen classes. However, as the training progresses, the vacuity in the data of the old classes tends to be higher and thus closer to OOD. As for dissonance, conflicting evidence is expected in the data for the old classes (mostly conflicting with the unseen classes), but in our case this only occurs in a small part of the test data. This can be seen in the model generated after task 4, where the dissonance of the old classes tends to be higher than that of the unseen ones. In most cases, the dissonance is very low.

Further analysis regarding vacuity and dissonance can be performed by analyzing Figure 3. In every plot, we report the average uncertainty for the test data. In the upper

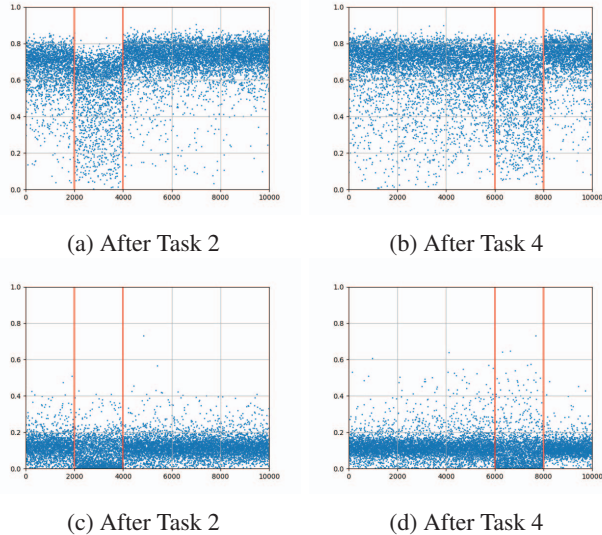| (a) After Task 2 | (b) After Task 4 |
| (c) After Task 2 | (d) After Task 4 |

Figure 2: Vacuity (top) and dissonance (bottom) estimated for the proposed method on all test data after training of the model with task 2 and task 4. The red vertical line divides the test data in IND data (left side) with respect to the OOD data (right side). The data are grouped by old classes (left), current classes (center) and unseen classes (right).

graphs, each line represents the average uncertainty provided by the model trained on the i-th task for the test data for each of the tasks, separately. From that we can observe the average uncertainty with respect to the data belonging to the old (left side), current (in the task id) and unseen (right side) classes. For example, for the model trained after the second task (task id = 2), we can observe that the vacuity is low for the data belonging to the same task, high for the data belonging to unseen classes and in between for old classes. As for dissonance, for the same task, we can see that it is low for the data corresponding to the current task, high for the old classes and for the unseen classes a little lower than for the old ones. Both behaviors remain the same for all tasks and highlight the ability of the proposed model to differentiate between IND and OOD data. On the other hand, regarding the bottom graphs the line represents the uncertainty provided by the i-th model after training the i-th task for the j-th task data. In this case, we can observe how the vacuity and dissonance increase after each incremental step having the maximum difference between the current classes and the immediately unseen classes.

Considering vacuity tends to be lower for IND data than OOD and dissonance higher for IND data than OOD, we evaluate whether by combining both measures of uncertainty it would be possible to better discriminate between data belonging to old classes with respect to the unseen ones. For this, we analyze the performance in OOD detection by the proposed combination of uncertainty metrics

($CU = \beta \times$ vacuity $+ (1 - \beta) \times (1 -$ dissonance$)$) for different values of $beta$ in the range $[0 - 1]$ (see Figure 4). Results are presented using incremental steps of size 10 and 20, where OOD detection is performed with respect to the test data from the previous and next tasks. The box plots show the mean and median in green and yellow lines, respectively. Here we can observe that vacuity provides much better performance than dissonance for the evaluation of IND_f vs OOD. However, for some particular values of β (e.g, $\beta = 0.3$), it is possible to improve the good performance achieved using only vacuity. We argue that this occurs, because some particular samples provide a conflicting evidence when the model is trained with new data, which can be detected with the dissonance measure and helps the model decide whether the data correspond to unseen classes or already seen by the model.

6. Conclusions

In this paper, we proposed a new method, called CEDL, which integrates a deep evidential learning approach into a continual learning framework to perform OOD detection based on uncertainty estimation. For this purpose, two uncertainty measures were analyzed: vacuity and dissonance. From the experimental results, we conclude that vacuity largely outperforms the state-of-the-art for OOD detection on the CIFAR-100 dataset in both settings of 5 and 10 sequential tasks, respectively. In addition, it was observed that vacuity tends to be lower and dissonance higher in the old class data relative to the unseen class data. Thus, the proposed CEDL method was able to retain evidence on old class data, but evidence on current classes conflicted with the old ones. Despite the latter, both measures could be used to differentiate between old and unseen class data. Although vacuity alone provides a good measure for OOD data detection, we also found that the combination of vacuity and dissonance can further the OOD detection. Finally, it can also be observed that the ability of the model to distinguish between different subsets of data (IND_c, IND_f and OOD) decreases after each incremental step. For this reason, in future work, we will explore improving the model's ability to cope with catastrophic forgetting in order to preserve the quality of uncertainty quantification and thus detection.
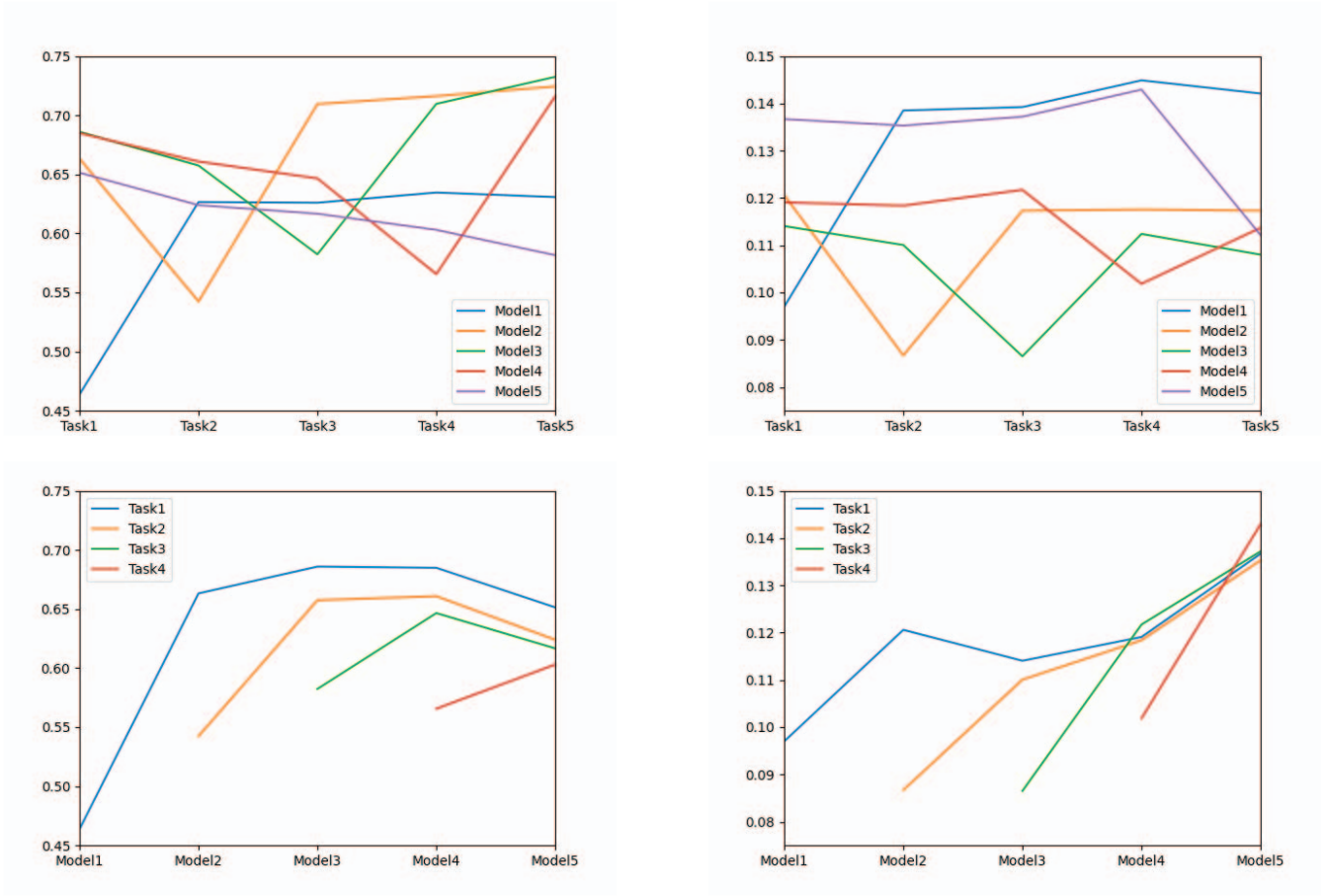
Figure 3: Average of Vacuity (left) and Dissonance (right) obtained by the proposed method. The i-th model corresponds to the resulting model after training with the j-th task. In the upper graphs, each line represents the results of the i-th model for the test data of the j-th task. In contrast, in the lower graphs, each line corresponds to the results of the i-th task for the j-th model.
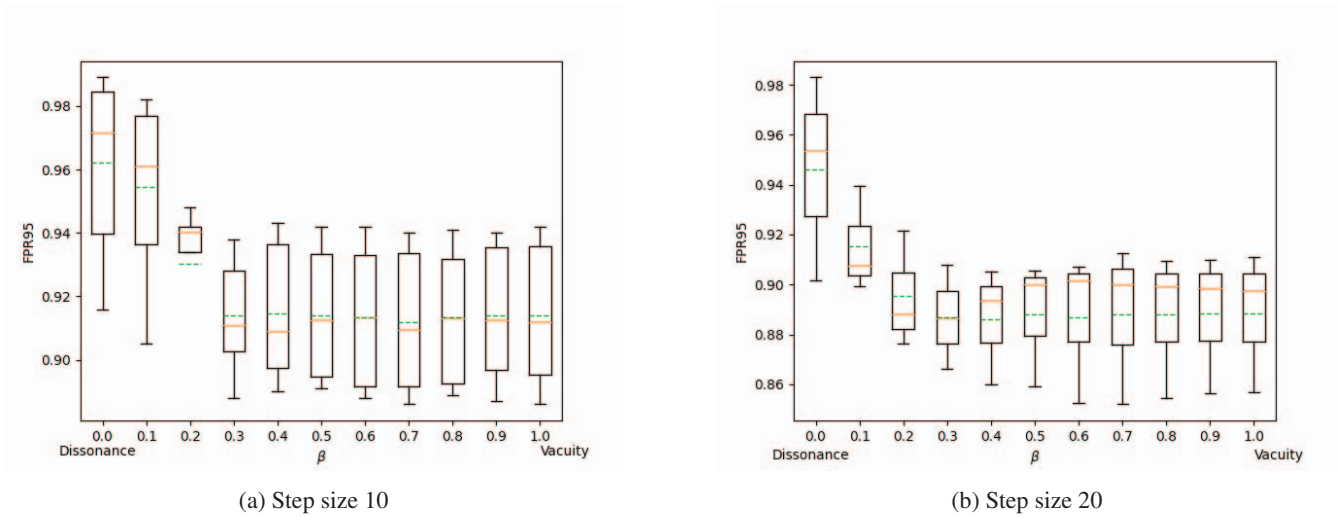

(a) Step size 10

(b) Step size 20

Figure 4: Box plot illustrating the results obtained when dissonance and vacuity are combined to perform detection between IND_f data and OOD data. The green and yellow lines correspond to the mean and median yield achieved in the evaluation after each increment, respectively.

References

[1] Moloud Abdar, Farhad Pourpanah, Sadiq Hussain, Dana Rezazadegan, and Li et al. Liu. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion*, 76:243—-297, 2021. 1

[2] Rahaf Aljundi, Daniel Olmeda Reino, Nikolay Chumerin, and Richard E Turner. Continual novelty detection. In *Conference on Lifelong Learning Agents*, pages 1004–1025. PMLR, 2022. 1, 3, 5, 6

[3] 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. 3

[4] 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. 2

[5] Pinaki Nath Chowdhury, Ayan Kumar Bhunia, Aneeshan Sain, Subhadeep Koley, Tao Xiang, and Yi-Zhe Song. What can human sketches do for object detection? In *Proceedings of the IEEE/CVF Computer Vision and Pattern Recognition*, pages 15083–15094, 2023. 1

[6] Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 702–703, 2020. 5

[7] Matthias De Lange, Rahaf Aljundi, Marc Masana, Sarah Parisot, Xu Jia, Aleš Leonardis, Gregory Slabaugh, and Tinne Tuytelaars. A continual learning survey: Defying forgetting in classification tasks. *IEEE transactions on pattern analysis and machine intelligence*, 44(7):3366–3385, 2021. 1, 2

[8] Sayna Ebrahimi, Mohamed Elhoseiny, Trevor Darrell, and Marcus Rohrbach. Uncertainty-guided continual learning with bayesian neural networks. In *International Conference on Learning Representations*, 2019. 2

[9] Jakob Gawlikowski, Cedrique Rovile Njieutcheu 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. 2

[10] Jiangpeng He and Fengqing Zhu. Out-of-distribution detection in unsupervised continual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3850–3855, 2022. 3, 4, 5, 6

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 5

[12] Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. In *International Conference on Learning Representations*, 2016. 2, 5, 6

[13] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. In *NIPS Deep Learning Workshop*, 2014. 2

[14] Karl Holmquist, Lena Klasén, and Michael Felsberg. Evidential deep learning for class-incremental semantic segmentation. In *Scandinavian Conference on Image Analysis*, pages 32–48. Springer, 2023. 2

[15] Saihui Hou, Xinyu Pan, Chen Change Loy, Zilei Wang, and Dahua Lin. Learning a unified classifier incrementally via rebalancing. In *International Conference on Computer Vision*, 2019. 2

[16] Yibo Hu, Yuzhe Ou, Xujiang Zhao, Jin-Hee Cho, and Feng Chen. Multidimensional uncertainty-aware evidential neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 7815–7822, 2021. 1

[17] Audun Josang, Jin-Hee Cho, and Feng Chen. Uncertainty characteristics of subjective opinions. In *2018 21st International Conference on Information Fusion (FUSION)*, pages 1998–2005. IEEE, 2018. 3

[18] Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? *Advances in neural information processing systems*, 30, 2017. 2

[19] Gyuhak Kim, Sepideh Esmaeilpour, Changnan Xiao, and Bing Liu. Continual learning based on ood detection and task masking. In *Proceedings of the IEEE/CVF International Conference on Computer Vision, Workshops*, 2022. 1, 3

[20] Gyuhak Kim, Zixuan Ke, and Bing Liu. A multi-head model for continual learning via out-of-distribution replay. In *Conference on Lifelong Learning Agents*, 2022. 1, 3

[21] Gyuhak Kim, Changnan Xiao, Tatsuya Konishi, and Bing Liu. Learnability and algorithm for continual learning. In *International Conference on Machine Learning*, pages 16877–16896, 2023. 3

[22] Gyuhak Kim Kim, Changnan Xiao, Tatsuya Konishiy, Zixuan Ke, and Bing Liu. A theoretical study on solving continual learning. In *Advances in neural information processing systems*, 2022. 3

[23] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. Segment anything. *arxiv preprint arXiv:2304.02643*, 2023. 1

[24] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526, 2017. 1, 2

[25] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009. 5

[26] Johnson Kuan and Jonas Mueller. Back to the basics: Revisiting out-of-distribution detection baselines. In *ICML Workshop on Principles of Distribution Shift*, 2022. 5, 6

[27] Okyu Kwon. Very simple statistical evidence that alphago has exceeded human limits in playing go game. *arxiv preprint arXiv:2002.11107*, 2020. 1

[28] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in neural information processing systems*, 30, 2017. 2

[29] Zhizhong Li and Derek Hoiem. Learning without forgetting. *IEEE transactions on pattern analysis and machine intelligence*, 40(12):2935–2947, 2017. 1, 2

[30] Zhaoshuo Li, Thomas Müller, Alex Evans, Russell H. Taylor, Mathias Unberath, Ming-Yu Liu, and Chen-Hsuan Lin. Neuralangelo: High-fidelity neural surface reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8456–8465, 2023. 1

[31] Shiyu Liang, Yixuan Li, and R Srikant. Enhancing the reliability of out-of-distribution image detection in neural networks. In *International Conference on Learning Representations*, 2018. 2, 5, 6

[32] Weitang Liu, Xiaoyun Wang, John Owens, and Yixuan Li. Energy-based out-of-distribution detection. *Advances in neural information processing systems*, 33:21464–21475, 2020. 5, 6

[33] Xialei Liu, Marc Masana, Luis Herranz, Joost Van de Weijer, Antonio M Lopez, and Andrew D Bagdanov. Rotate your networks: Better weight consolidation and less catastrophic forgetting. In *Int. Conf. on Pattern Recognit.*, 2018. 2

[34] Noel Loo, Siddharth Swaroop, and Richard E Turner. Generalized variational continual learning. In *International Conference on Learning Representations*, 2020. 2

[35] David Lopez-Paz and Marc'Aurelio Ranzato. Gradient episodic memory for continual learning. *Advances in neural information processing systems*, 30, 2017. 1, 2

[36] Wesley J. Maddox, Timur Garipov, Pavel Izmailov, Dmitry Vetrov, and Andrew Gordon Wilson. A simple baseline for bayesian uncertainty in deep learning. In *Advances in neural information processing systems*, 2019. 1

[37] Zheda Mai, Ruiwen Li, Jihwan Jeong, David Quispe, Hyunwoo Kim, and Scott Sanner. Online continual learning in image classification: An empirical survey. *Neurocomputing*, 469:28–51, 2022. 2

[38] Andrey Malinin and Mark Gales. Predictive uncertainty estimation via prior networks. *Advances in neural information processing systems*, 31, 2018. 2

[39] Marc Masana, Xialei Liu, Bartłomiej Twardowski, Mikel Menta, Andrew D Bagdanov, and Joost Van De Weijer. Class-incremental learning: survey and performance evaluation on image classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5):5513–5533, 2022. 1, 2

[40] Martial Mermillod, Aurélia Bugaiska, and Patrick Bonin. The stability-plasticity dilemma: investigating the continuum from catastrophic forgetting to age-limited learning effects. *Frontiers in Psychology*, 2013. 1

[41] Anh Nguyen, Jason Yosinski, and Jeff Clune. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 427–436, 2015. 3

[42] Cuong V Nguyen, Yingzhen Li, Thang D Bui, and Richard E Turner. Variational continual learning. In *International Conference on Learning Representations*, 2018. 2

[43] German I Parisi, Ronald Kemker, Jose L Part, Christopher Kanan, and Stefan Wermter. Continual lifelong learning with neural networks: A review. *Neural networks*, 113:54–71, 2019. 2

[44] Rahul Rahaman et al. Uncertainty quantification and deep ensembles. *Advances in Neural Information Processing Systems*, 34:20063–20075, 2021. 1

[45] Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, and Christoph H Lampert. icarl: Incremental classifier and representation learning. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 2001–2010, 2017. 1, 2, 5

[46] Amanda Rios, Nilesh Ahuja, Ibrahima Ndiour, Utku Genc, Laurent Itti, and Omesh Tickoo. incdfm: Incremental deep feature modeling for continual novelty detection. In *European Conference on Computer Vision*, 2022. 3

[47] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10684–10695, 2022. 1

[48] Murat Sensoy, Lance Kaplan, and Melih Kandemir. Evidential deep learning to quantify classification uncertainty. *Advances in neural information processing systems*, 31, 2018. 1, 2, 3, 4

[49] Joan Serra, Didac Suris, Marius Miron, and Alexandros Karatzoglou. Overcoming catastrophic forgetting with hard attention to the task. In *International conference on machine learning*, pages 4548–4557. PMLR, 2018. 1

[50] Hanul Shin, Jung Kwon Lee, Jaehong Kim, and Jiwon Kim. Continual learning with deep generative replay. In *Advances in Neural Information Processing Systems*, 2017. 2

[51] Jihoon Tack, Sangwoo Mo, Jongheon Jeong, and Jinwoo Shin. Csi: Novelty detection via contrastive learning on distributionally shifted instances. In *Advances in Neural Information Processing Systems*, 2020. 2

[52] 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. 1

[53] Guotai Wang, Wenqi Li, Michael Aertsen, Jan Deprest, Sébastien Ourselin, and Tom Vercauteren. Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks. *Neurocomputing*, 338:34–45, 2019. 2

[54] Felix Wiewel, Alexander Bartler, and Bin Yang. Dirichlet prior networks for continual learning. In *2022 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2022. 2, 5

[55] Chenshen Wu, Luis Herranz, Xialei Liu, Yaxing Wang, Joost van de Weijer, and Bogdan Raducanu. Memory replay GANs: learning to generate images from new categories without forgetting. In *Advances in Neural Information Processing Systems*, 2018. 2

[56] Jingkang Yang, Pengyun Wang, and Dejian et al. Zou. Openood: Benchmarking generalized out-of-distribution detection. In *Advances in Neural Information Processing Systems*, 2022. 2

[57] Jingkang Yang, Kaiyang Zhou, Yixuan Li, and Ziwei Liu. Generalized out-of-distribution detection: A survey. *arXiv preprint arXiv:2110.11334*, 2022. 2

[58] Friedemann Zenke, Ben Poole, and Surya Ganguli. Continual learning through synaptic intelligence. In *Proc. Int. Conf. Mach. Learn.*, 2017. 2

[59] Bowen Zhao, Xi Xiao, Guojun Gan, Bin Zhang, and Shu-Tao Xia. Maintaining discrimination and fairness in class incremental learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13208–13217, 2020. 4, 5

[60] Xujiang Zhao, Yuzhe Ou, Lance Kaplan, Feng Chen, and Jin-Hee Cho. Quantifying classification uncertainty using regularized evidential neural networks. In *AAAI FSS*, 2019. 1, 3

[61] Wenying Zhou, Yang Yang, and Cheng et al. Yu. Ensembled deep learning model outperforms human experts in diagnosing biliary atresia from sonographic gallbladder images. *Nature Communications*, 12(1259), 2021. 1