

RESEARCH ARTICLE

Boosting Out-of-Distribution Image Detection With Epistemic Uncertainty

DOKWAN OH¹, DAEHYUN JI¹, OHMIN KWON², AND YOONSUK HYUN³¹Samsung Advanced Institute of Technology, Samsung Electronics, Suwon 16678, South Korea²School of Electrical Engineering, Chungbuk National University, Cheongju 28644, South Korea³Department of Mathematics, Inha University, Incheon 22212, South Korea

Corresponding author: Yoonsuk Hyun (yshyun21@inha.ac.kr)

This work was supported in part by the National Research Foundation of Korea (NRF) Grant (NRF-2021R1C1C1008929) funded by the Korean Government, and in part by the Inha University Research Grant. The work of Ohmin Kwon was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant NRF-2020R1A6A1A12047945.


ABSTRACT Modern deep neural networks are known to generate over-confident class predictions even for unseen samples. However, safety-critical applications are required to understand examples that differ from the training distribution. For example, an autonomous vehicle must return a instant refusal feedback when encountering an unexpected situation. The voice secretary should re-ask the user for a command that was not understood to prevent malfunction. In this paper, we propose an out-of-distribution sample detection algorithm using *Uncertainty-based Additive Fast Gradient Sign Method (UA-FGSM)*, which uses Monte Carlo (MC) dropout during backpropagation. The proposed uncertainty-based method forces in-distribution sample predictions to be more over-confident and out-of-distribution sample predictions to be less over-confident in the pre-trained model. This boosts the discrimination between the in-distribution and out-of-distribution samples. In addition, we further boost this difference by continuously accumulating uncertainty-based gradients. Our method uses the inherent epistemic uncertainty of the pre-trained model. Therefore, the proposed algorithm does not require knowledge of the domain of the in-distribution dataset and works by simple pre-processing of the already trained model without any re-training. We demonstrate its effectiveness using diverse network architectures on various popular image datasets and noisy settings.

INDEX TERMS Deep learning, epistemic uncertainty, fast gradient sign method, out of distribution sample detection.

I. INTRODUCTION

One of the most intriguing problems in machine learning is determining whether the given input belongs to the same distribution as the training data. This is crucial for ensuring the safety and reliability of the trained algorithm, especially in safety-critical applications. This is known as an out-of-distribution (OOD) detection problem. Recent studies on OOD sample detection have focused on the confidence scores generated by a pre-trained model. In particular, the maximum values of the confidence scores from all classes can help determine whether the query input data represent

an in-distribution sample. These approaches are based on the observation that the predictions of OOD samples are highly overconfident about incorrect or unknown classes but relatively less overconfident comparing to in-distribution samples. Thus, a change in confidence score caused by even a small perturbation will be different for both cases. In this study, we aim to improve the OOD sample detection performance by proposing a framework that can further amplify the above-mentioned phenomenon. The key idea is to capture the epistemic uncertainty that the pre-trained model already possesses. Epistemic uncertainty refers to uncertainty in the process model. This is mainly caused by a lack of data and knowledge, and modeling epistemic uncertainty can contribute to the development of a

The associate editor coordinating the review of this manuscript and approving it for publication was Nagarajan Raghavan .

rigorous algorithm. We focused on the usefulness of epistemic uncertainty in the OOD detection problem. In addition, the epistemic uncertainty of the OOD sample, which is not used in the training phase, is greater than that of the in-distribution sample. By superimposing these two characteristics, i.e., the relative confidence difference observation mentioned above and the proposed relative uncertainty difference assumption, we obtained the ability to further distinguish between the in-distribution and OOD samples. Both are implicit but important pieces of information that can be extracted from the pre-trained model.

For precise detection, simply observing the epistemic uncertainty of the given sample is insufficient. To combine the change in confidence score with uncertainty, the perturbation method can be made dependent on the uncertainty.

The contributions of this study are as follows:

- We introduce a novel image generation algorithm with an uncertainty-based gradient method. The method uses Monte Carlo (MC) dropout for gradient calculation with respect to the input image to apply epistemic uncertainty.
- Furthermore, we propose an additive method of multiple gradients obtained by the MC dropout from a single image to further increase the discrimination of in-distribution and OOD samples.
- Finally, we demonstrate and interpret the performance of the proposed method on state-of-the-art network architectures under diverse combinations of in-distribution and OOD dataset pairs.

II. RELATED WORK

A. OOD SAMPLE DETECTOR

Hendrycks et al. [1] first proposed a method for detecting OOD examples using numerous datasets. They explained that the softmax classifier of a well-trained neural network tends to be able to distinguish between in-distribution and OOD samples. This study is referred to as the baseline in this area. Liang et al. [2] (Odin) reported that using temperature scaling in the softmax function allows the OOD sample to be distinguished well. They explained the influence of temperature scaling through mathematical analysis using Taylor series expansion. Additionally, they proposed a method to broaden the gap between the two distributions by applying a small perturbation to the input. This perturbation was calculated using the fast gradient sign method, which was designed for adversarial attacks [6]. Lee et al. [3] exploited the leverage of generative adversarial networks (GANs) to virtually generate OOD examples for the given in-distribution samples. They trained GANs and distinguished in-distribution and OOD samples simultaneously on a single network. They proposed that joint training loss additionally minimizes the Kullback–Leibler (KL) divergence from the predictive distribution of OOD samples to be closer to the uniform distributions to make less confident predictions. The confident classifier improves the GAN, which further improves the performance of the confident classifier as the training progresses. Vyas et al. [4] designed a margin-based loss term, termed

cross-entropy loss of in-distribution samples, which maintains a margin between the OOD and in-distribution samples. Based on the margin-based loss that seeks to maintain a margin m , they proposed a method to train an ensemble of classifiers. In addition, they defined their own OOD sample detection score, which expects a higher detection score for in-distribution samples than for OOD samples. Lee et al. [5] proposed a method for detecting OOD samples, which applies to any pre-trained softmax neural classifier. The main idea of the method was to measure the probability density of OOD samples in the feature space of a neural network using a generative classifier based on Mahalanobis distance. They obtained class conditional Gaussian distributions with respect to the features of the neural network, which resulted in a confidence score based on the Mahalanobis distance. To further improve the performance, they used ensemble methods such as weighted averaging of all confidence scores in the neural network. They trained a logistic regression detector using validation samples to calculate the weights for each layer. Our proposed method focuses on simple pre-processing techniques such as those used by Hendrycks et al. [1] and Liang et al. [2]. Our method does not require information about in-distribution samples or heavy computations such as network re-training or ensembles.

B. ADVERSARIAL ATTACK

There are various approaches for developing OOD sample detectors that are inspired by the basic idea of adversarial attacks [6]. Goodfellow et al. [6] first proposed a method, called the Fast Gradient Sign Method (FGSM), to decrease the softmax score for the true label by adding a small perturbation to the image, forcing the neural network to make incorrect predictions. The FGSM calculates the gradient of the cost function $J(\mathbf{x}, y)$ with respect to the input to the neural network. The adversarial examples \mathbf{x}^{adv} are generated using the following equation:

$$\mathbf{x}^{adv} = \mathbf{x} + \varepsilon \cdot \text{sign}(\nabla_{\mathbf{x}} J(\mathbf{x}, y)). \quad (1)$$

Here, $J(\mathbf{x}, y)$ is the cost function of the trained model, $\nabla_{\mathbf{x}}$ denotes the gradient of the model with respect to a sample x with the correct label y , and ε denotes the parameter that adjusts the magnitude of the perturbation. Recent studies have used other variations of the FGSM, as discussed below. Iterative methods (Kurakin et al. [7]) iteratively apply a fast gradient multiple times with a small step size α . The iterative version of the FGSM (I-FGSM) can be expressed as:

$$\mathbf{x}_0^{adv} = \mathbf{x}, \quad \mathbf{x}_{t+1}^{adv} = \mathbf{x}_t^{adv} + \alpha \cdot \text{sign}(\nabla_{\mathbf{x}} J(\mathbf{x}_t, y)). \quad (2)$$

To satisfy the $L1$ (or $L2$) bound in the adversarial image, one can clip \mathbf{x}_t^{adv} into the ε vicinity of \mathbf{x} or simply set $\alpha = \varepsilon/T$, where T denotes the number of iterations. It has been shown that iterative methods are stronger white-box adversaries than one-step methods at the cost of worse transferability. We use this attack mechanism in a reverse manner to determine the effect of attention on the targeted label direction. In addition, similar to I-FGSM, we were able to achieve a boosting effect

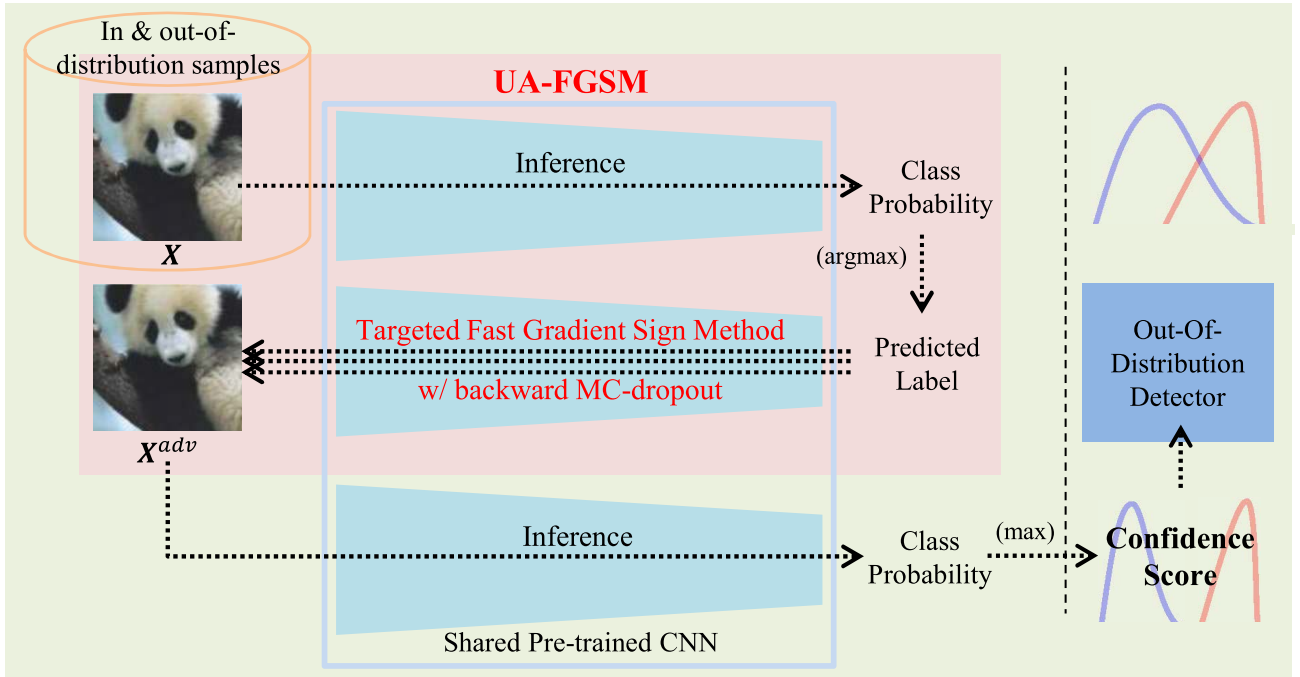


FIGURE 1. Overall structure of the proposed OOD sample detector. It is performed with only two forward paths and one backward path, without any training process.

on detection performance by incorporating the concept of iterative operations into our framework.

III. UA-FGSM: UNCERTAINTY-BASED ADDITIVE FAST GRADIENT SIGN METHOD

The sample perturbation method in adversarial attacks enlarges the gap between samples of different types. We adopted this idea to solve the OOD detection problem. The OOD sample is more likely to be unstable than the in-distribution sample, and a small change would lead to an incorrect decision. One of our key suggestions is the use of reverse perturbation. Instead of using the sign for computed gradients, we used the opposite sign for perturbation. This reverse-way perturbation acts as attention for an in-distribution sample, and as a distortion for an OOD sample. Another critical strategy is to apply uncertainty during the perturbation process. In addition to the simple gradient-based perturbation, we applied MC dropout [15] to extract epistemic uncertainty. In the case of an OOD sample, high epistemic uncertainty makes \mathbf{x}^{adv} more diverse in the less overconfident direction. The suggested equation for the UA-FGSM is as follows:

$$\mathbf{x}^{adv} = \mathbf{x} - \varepsilon \cdot \sum_{mc} \text{sign}(\nabla_{\mathbf{x}_{mc}} J(\mathbf{x}_{mc}, y)), \quad (3)$$

where \mathbf{x}_{mc} is a variable in which the same image \mathbf{x} is repeatedly concatenated mc (dropout trial) times on the mini-batch axis, and $\nabla_{\mathbf{x}_{mc}}$ are gradients with respect to \mathbf{x}_{mc} using backward MC dropout. In other words, UA-FGSM is a procedure in which several gradients reflecting epistemic

uncertainty are accumulated from the pre-trained model in the same image. Fig. 1 shows the overall structure of the proposed OOD sample detector. It is performed with only two forward paths and one backward path, without any training process. Backward MC dropout was implemented using a mini-batch, and several uncertainty-based gradients were obtained by performing only one backward path. Thus, the suggested method requires very little computation time compared with the method that requires retraining for its accuracy.

Fig. 2 shows the iterative accumulation of uncertainty-based additive gradients as the number of dropout trials (mc) increases. Compared with the same dropout trial (mc), the texture structure of the gradient is preserved in the in-distribution sample, but quickly disappears in the OOD sample. This phenomenon enhances the gap between the perturbed images from the in-distribution and OOD samples. At the same time, this provides an effect of preventing the drop in accuracy as the dropout trial (mc) increases, as shown in Figs. 5c and 5f.

The detailed procedure of the suggested UA-FGSM is described in Fig. 3. We expect to see an effect of attention by applying a small perturbation as opposed to an adversarial attack, as in Odin [2]. The class with the highest class probability, which is determined to be correct, increases further through the summation of the gradient. At this time, the effect of attention is more in the in-distribution sample than in the OOD sample (for example, the airplane class). On the other hand, the number of class probabilities that are transferred to the remaining classes is higher in the OOD sample than in the in-distribution sample. This makes the predictions for

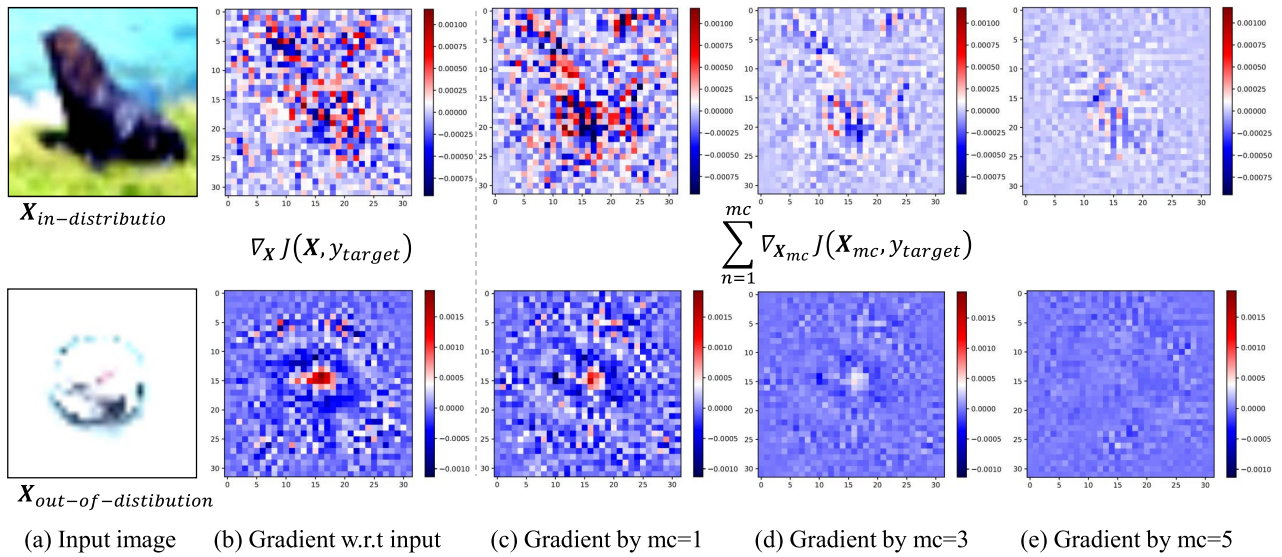


FIGURE 2. Examples of uncertainty-based additive gradients: (a) in-distribution and OOD samples, (b) gradient of FGM, and (c, d, e) changes in additive gradient as dropout trial (mc) increases. As we can observe from (c, d, e), the gradients are averaging out as the dropout trial (mc) increases.

in-distribution samples more over-confident than that of OOD samples, resulting in higher detection performance. A more detailed procedure is presented in Algorithm 1.

IV. EXPERIMENTS

A. SETTINGS

For experimental comparison, we followed all the settings done in the work by Odin [2].

1) PRE-TRAINED MODELS

We have trained the following three famous models: MobileNet-v2 [8], ShuffleNet-v2 [9], and DenseNet-BC100 [10]. The accuracy and size of the models are listed in Table 1. The model network did not use dropout during training for 300 epochs, with a batch size of 64, momentum of 0.9, and weight decay of $1e-4$. The learning rate started at 0.1 and was dropped by a factor of 10 at 50% and 75% of the training progress, respectively.

TABLE 1. Test accuracy and model size on CIFAR datasets.

Architecture	CIFAR-10	CIFAR-100
ShuffleNet-v2	92.53%	72.57%
	(1.26M)	(1.36M)
MobileNet-v2	93.4%	75.36%
	(2.30M)	(2.41M)
DenseNet-BC100	95.43%	77.06%
	(0.77M)	(0.80M)

2) OUT-OF-DISTRIBUTION DATASET

Test images from the CIFAR-10 and CIFAR-100 [11] datasets can be viewed as in-distribution (positive) examples.

The following five datasets were used for OOD (negative) examples: (1) TinyImageNet: TinyImageNet comprises a subset of ImageNet [12]. Two versions of the dataset were used by either randomly cropping or downsampling each image to 32×32 . (2) LSUN: Large-scale scene understanding [13] dataset can also be used in a similar manner as TinyImageNet by crop and resize. (3) iSUN: iSUN [14] was used after downsampling each image to a pixel size of 32×32 . (4) Gaussian and uniform noises were synthesized with a pixel size of 32×32 . All datasets contained 10,000 samples, except for iSUN (8,925 samples). As in Odin, 1,000 examples from the datasets were used as a set for parameter tuning, and the remaining examples were used for testing.

3) EVALUATION METRICS

In Table 2, results of the comparative experiments that were conducted with Odin [2] in various configurations are listed. For evaluation, the following metrics were used: (1) TNR at 95% TPR: This can be interpreted as the probability that a negative (OOD) example is well classified as negative when the true-positive rate is as high as 95%. (2) Detection Accuracy: This is defined as $0.5(1-FPR) + 0.5(TPR)$, where we assume that both the positive and negative examples have equal probability of appearing in the test set. (3) AUROC: Area under the receiver operating characteristic curve. (4) AUPR: AUPR is the area under the precision-recall curve. The larger the above metrics, the better the performance.

B. RESULTS

1) COMPARISON WITH ODIN [2]

In Fig. 4a, we show the ROC curves when ShuffleNet-v2 was evaluated on CIFAR-100 (positive) images against the LSUN (negative) examples. The blue curve corresponds to the ROC curve when using the baseline method [2], whereas the red

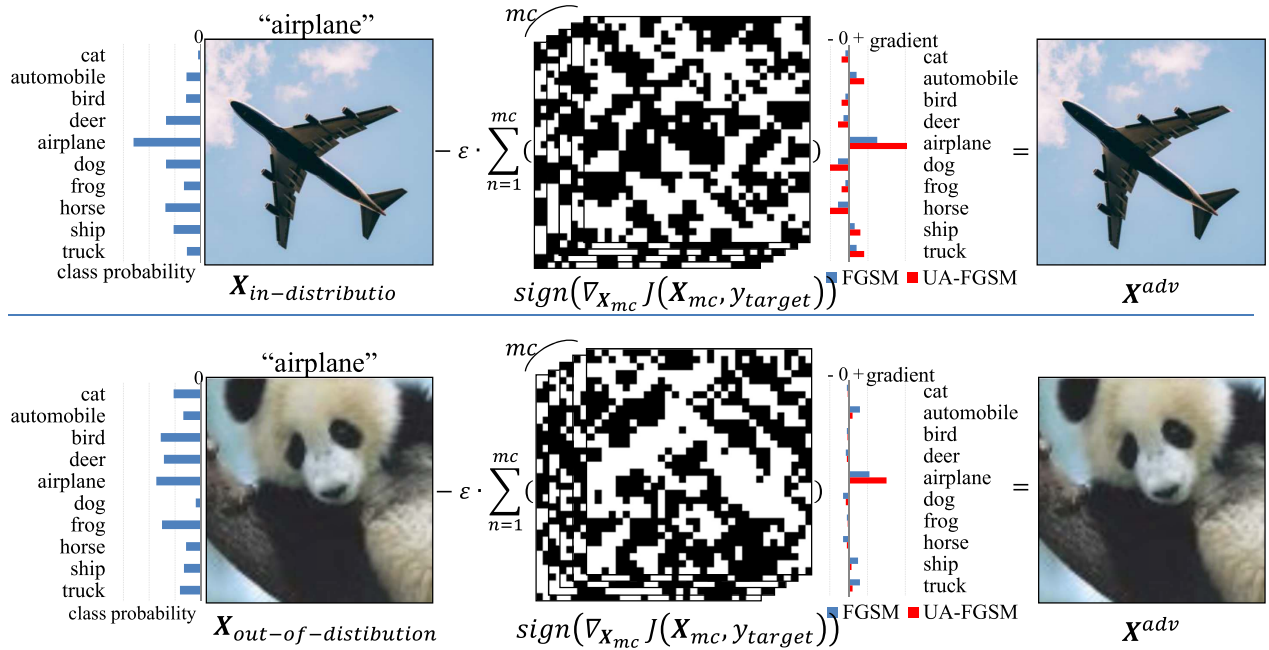


FIGURE 3. Illustration of the UA-FGSM procedure.

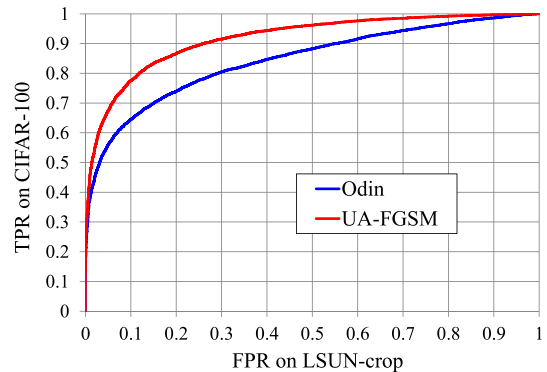
Algorithm 1 OOD Image Detection by UA-FGSM

Input: Test image \mathbf{x} , pre-trained model F , number of minibatch for dropout trial mc , and scale of perturbation ε

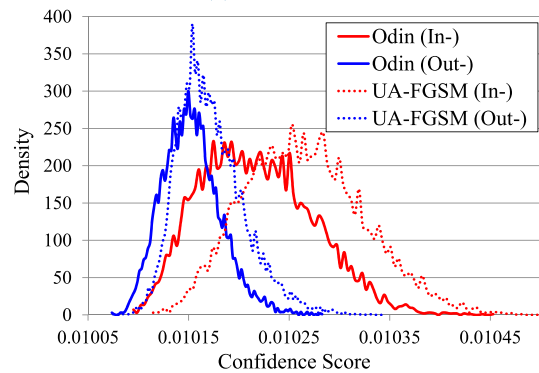
Output: Confidence score for OOD image detector \mathbf{C}

- 1: **procedure** UA-FGSM
- 2: Compute class prediction: $\mathbf{P} \leftarrow F(\mathbf{x})$
- 3: Predict target label: $y \leftarrow \text{argmax}(P)$
- 4: Concatenate same image \mathbf{x} along mc minibatch axis
 : $\mathbf{x}_{mc} \leftarrow \text{concat}(\mathbf{x})$
- 5: Compute gradients w.r.t. \mathbf{x}_{mc} using MC dropout
 : $\mathbf{G}_{mc} \leftarrow \nabla_{\mathbf{x}_{mc}} J(\mathbf{x}_{mc}, y)$
- 6: $\mathbf{O} \leftarrow \{0\}$
- 7: **for** $n \leftarrow 1$ to mc **do**
- 8: Accumulate sign of gradients
 : $\mathbf{O} \leftarrow \mathbf{O} + \text{sign}(\mathbf{G}_n)$
- 9: **end for**
- 10: **end procedure**
- 11: Generate a perturbed image: $\mathbf{x}^{adv} \leftarrow \mathbf{x} - \varepsilon \cdot \mathbf{O}$
- 12:
- 13: Compute new class prediction: $\mathbf{P}' \leftarrow F(\mathbf{x}^{adv})$
- 14: Computing the confidence score: $\mathbf{C} \leftarrow \text{max}(\mathbf{P}')$
- 15: **return** \mathbf{C}

curve corresponds to that of the proposed method. A large gap was observed between the blue and red ROC curves. Fig. 4b shows the distribution of the confidence scores for the entire dataset obtained in the same experimental environment as the ROC curve. Compared with Odin, the in-distribution samples become overconfident overall and OOD samples become less overconfident. This directly leads to an improved detection performance by making both distributions more separable.



(a) ROC Curve



(b) Density of confidence score

FIGURE 4. Comparison of the obtained results with baseline (Odin [2]).

2) CHOOSING PARAMETERS

The scale of perturbation (ε) of algorithms 1 was set to 0.0014. We did not perform any additional parameter

TABLE 2. Distinguishing in-distribution and OOD test set for image classification. In this table, for every detection metric, a larger value indicates better performance, which is highlighted in bold.

Out-of-distribution dataset	TNR (95% TPR)		Detection Accuracy		AUROC			AUPR In		AUPR Out	
	Hendrycks [1] / Odin [2] / Ours										
MobileNet-v2 CIFAR-10	TinyImageNet (crop)	71.7/ 86.7/ 88.3	84.9/ 92.5/ 92.7	94.9/ 97.7/ 97.9	97.5/ 98.0/ 98.1	93.7/ 97.4/ 97.6					
	TinyImageNet (resize)	70.1/ 89.9/ 92.1	83.1/ 93.1/ 93.9	97.0/ 98.2/ 98.5	97.4/ 98.4/ 98.6	93.9/ 98.1/ 98.3					
	LSUN (crop)	67.6/ 80.8/ 84.3	84.3/ 90.9/ 91.4	95.3/ 96.8/ 97.2	96.3/ 97.2/ 97.5	94.3/ 96.3/ 97.0					
	LSUN (resize)	76.3/ 93.2/ 94.9	85.5/ 94.2/ 95.0	95.9/ 98.7/ 98.9	97.3/ 98.8/ 99.0	94.9/ 98.6/ 98.8					
	iSUN	74.4/ 93.0/ 94.7	85.1/ 94.2/ 94.9	96.3/ 98.7/ 98.9	97.7/ 98.8/ 99.0	95.1/ 98.6/ 98.8					
	Uniform	65.3/ 68.9/ 98.8	91.0/ 94.0/ 96.9	94.9/ 95.8/ 98.5	95.1/ 97.3/ 99.0	88.8/ 90.3/ 96.6					
	Gaussian	90.5/ 99.9/ 100.0	93.1/ 97.7/ 99.1	96.8/ 98.9/ 99.7	96.8/ 99.3/ 99.8	95.2/ 97.7/ 99.4					
MobileNet-v2 CIFAR-100	TinyImageNet (crop)	33.8/ 55.9/ 62.5	72.1/ 84.2/ 85.7	84.1/ 92.0/ 93.3	85.5/ 93.1/ 94.1	80.3/ 90.9/ 92.2					
	TinyImageNet (resize)	24.3/ 47.5/ 55.8	70.1/ 81.7/ 84.2	77.6/ 89.9/ 91.9	80.0/ 91.3/ 93.1	75.5/ 88.4/ 90.6					
	LSUN (crop)	20.7/ 40.2/ 58.1	63.8/ 80.9/ 84.6	76.8/ 88.5/ 92.2	81.1/ 90.3/ 93.3	72.1/ 86.2/ 91.1					
	LSUN (resize)	28.9/ 46.9/ 55.2	66.3/ 81.5/ 83.8	76.9/ 89.7/ 91.7	79.3/ 91.2/ 92.9	73.9/ 88.0/ 90.2					
	iSUN	30.3/ 44.9/ 54.1	70.0/ 81.2/ 83.6	76.4/ 89.3/ 91.4	80.8/ 90.8/ 92.6	73.3/ 87.5/ 89.8					
	Uniform	13.3/ 16.1/ 87.7	80.3/ 92.1/ 95.7	80.3/ 92.3/ 96.8	85.9/ 95.5/ 98.1	59.9/ 82.2/ 91.8					
	Gaussian	0.0/ 0.0/ 0.0	72.1/ 81.3/ 87.7	55.5/ 72.6/ 83.9	78.7/ 84.1/ 90.8	40.3/ 58.0/ 68.9					
ShuffleNet-v2 CIFAR-10	TinyImageNet (crop)	71.8/ 85.8/ 87.8	86.6/ 92.2/ 92.6	96.8/ 97.5/ 97.8	97.7/ 97.9/ 98.1	95.8/ 97.1/ 97.5					
	TinyImageNet (resize)	71.9/ 86.6/ 89.5	83.3/ 92.2/ 93.0	97.0/ 97.6/ 98.0	97.9/ 97.8/ 98.1	95.8/ 97.3/ 97.8					
	LSUN (crop)	69.4/ 77.3/ 81.0	85.4/ 90.3/ 91.0	95.8/ 96.3/ 96.8	96.7/ 96.9/ 97.2	94.3/ 95.7/ 96.4					
	LSUN (resize)	82.6/ 92.6/ 94.7	87.8/ 94.4/ 95.0	97.3/ 98.5/ 98.8	98.7/ 98.7/ 99.0	96.0/ 98.2/ 98.7					
	iSUN	81.7/ 91.0/ 93.0	88.4/ 93.5/ 94.2	97.7/ 98.3/ 98.6	98.6/ 98.5/ 98.8	96.9/ 98.1/ 98.5					
	Uniform	6.0/ 6.7/ 49.7	88.0/ 89.0/ 93.0	89.1/ 89.5/ 94.6	93.5/ 93.6/ 96.7	77.7/ 78.0/ 87.5					
	Gaussian	89.7/ 90.0/ 99.7	94.9/ 95.4/ 97.6	97.0/ 97.3/ 99.0	98.1/ 98.3/ 99.4	93.8/ 94.1/ 98.0					
ShuffleNet-v2 CIFAR-100	TinyImageNet (crop)	33.4/ 42.5/ 47.0	73.3/ 79.0/ 80.4	83.3/ 87.1/ 88.6	84.4/ 88.4/ 89.6	80.9/ 85.7/ 87.5					
	TinyImageNet (resize)	27.7/ 38.1/ 43.6	71.2/ 77.2/ 79.1	79.8/ 85.4/ 87.4	80.2/ 86.9/ 88.6	77.7/ 83.9/ 86.3					
	LSUN (crop)	20.3/ 27.1/ 36.6	68.8/ 77.6/ 80.4	78.8/ 84.5/ 87.8	83.2/ 87.2/ 89.7	73.3/ 80.5/ 84.7					
	LSUN (resize)	24.4/ 34.7/ 40.3	70.9/ 76.1/ 77.9	78.1/ 84.2/ 86.2	80.8/ 85.8/ 87.5	75.5/ 82.5/ 84.9					
	iSUN	27.5/ 35.1/ 41.5	70.1/ 76.6/ 78.5	77.9/ 84.6/ 86.7	81.0/ 86.0/ 87.7	69.9/ 82.8/ 85.3					
	Uniform	2.3/ 98.4/ 100.0	48.4/ 96.8/ 98.2	89.3/ 98.7/ 99.5	95.5/ 99.1/ 99.7	84.4/ 97.6/ 99.3					
	Gaussian	3.7/ 91.3/ 98.9	48.9/ 95.0/ 97.0	88.8/ 97.7/ 99.1	95.4/ 98.4/ 99.4	83.3/ 95.6/ 98.6					
DenseNet100 CIFAR-10	TinyImageNet (crop)	61.7/ 92.0 / 91.7	78.1/ 94.1 / 93.6	94.7/ 98.5 / 98.3	96.5/ 98.6 / 98.4	92.8/ 98.3 / 98.1					
	TinyImageNet (resize)	54.6/ 87.9/ 89.7	76.1/ 92.7/ 93.1	93.5/ 97.8/ 97.9	95.3/ 98.0/ 98.2	91.4/ 97.5/ 97.7					
	LSUN (crop)	58.9/ 89.3 / 88.5	76.8/ 92.8 / 92.2	94.0/ 97.9 / 97.5	96.1/ 97.9 / 97.2	93.0/ 97.9 / 97.6					
	LSUN (resize)	65.5/ 95.6/ 96.5	80.7/ 95.4/ 95.8	96.0/ 99.0/ 99.1	96.4/ 99.1/ 99.2	93.8/ 98.8/ 98.9					
	iSUN	54.9/ 93.4/ 94.5	78.9/ 94.5/ 95.0	94.9/ 98.6/ 98.8	95.9/ 98.8/ 99.0	93.3/ 98.5/ 98.7					
	Uniform	23.0/ 32.8/ 77.1	75.7/ 89.3/ 93.3	90.1/ 92.5/ 96.3	92.9/ 95.1/ 97.6	78.3/ 84.8/ 92.4					
	Gaussian	3.8/ 8.8/ 45.3	77.3/ 84.8/ 90.7	85.2/ 87.0/ 93.6	90.2/ 91.5/ 95.9	72.5/ 76.0/ 87.0					
DenseNet100 CIFAR-100	TinyImageNet (crop)	13.1/ 55.2/ 56.1	57.7/ 82.7/ 83.0	75.2/ 90.8/ 91.0	80.0/ 91.5/ 91.7	84.1/ 90.1/ 90.1					
	TinyImageNet (resize)	16.7/ 49.4/ 52.0	60.8/ 80.4/ 81.2	69.9/ 88.4/ 89.2	68.9/ 88.8/ 89.6	64.8/ 87.9/ 88.4					
	LSUN (crop)	13.3/ 41.4/ 54.1	53.0/ 79.9/ 82.9	76.6/ 87.6/ 90.8	77.3/ 88.9/ 91.7	60.1/ 86.0/ 90.0					
	LSUN (resize)	15.9/ 42.6/ 45.3	68.1/ 79.5/ 80.4	66.1/ 87.5/ 88.2	68.7/ 88.6/ 89.2	64.3/ 86.2/ 86.7					
	iSUN	16.0/ 43.4/ 45.9	62.3/ 79.4/ 80.3	69.4/ 87.3/ 88.2	67.0/ 88.1/ 89.0	65.3/ 86.1/ 86.6					
	Uniform	0.0/ 0.0/ 0.1	33.3/ 72.7/ 79.9	50.1/ 65.6/ 77.9	66.8/ 77.9/ 86.1	40.7/ 53.4/ 63.6					
	Gaussian	0.0/ 0.2/ 14.5	35.7/ 82.9/ 89.5	66.3/ 81.8/ 90.8	76.5/ 88.8/ 94.3	50.4/ 67.7/ 80.5					

searches on the value found by Odin and used it as is. In Table 2, the dropout trial (mc) and dropout ratio are set to 2 and 1% in all experiments.

3) MAIN RESULTS

The comparison of all detection metrics with Odin is summarized in Table 2. We observed improved performance across most neural network architectures and dataset pairs. In particular, there was a significant improvement in CIFAR-10 and noisy samples, although a slight improvement was observed in CIFAR-10.

V. DISCUSSION

A. EFFECTS OF PARAMETERS

In this subsection, we observe the changes in detection metric and accuracy with changes in the dropout trial (mc) and dropout ratio. Figs. 5a, 5b, and 5c show the trend graphs of the obtained detection metric and accuracy drop obtained when the architecture was MobileNet-v2, in-distribution dataset was CIFAR-100, and OOD dataset was LSUN. Figs. 5d, 5e, and 5f were obtained using a combination of ShuffleNet-v2, CIFAR-100, and iSUN. The result from Odin, which is the leftmost point of the black circle curve,

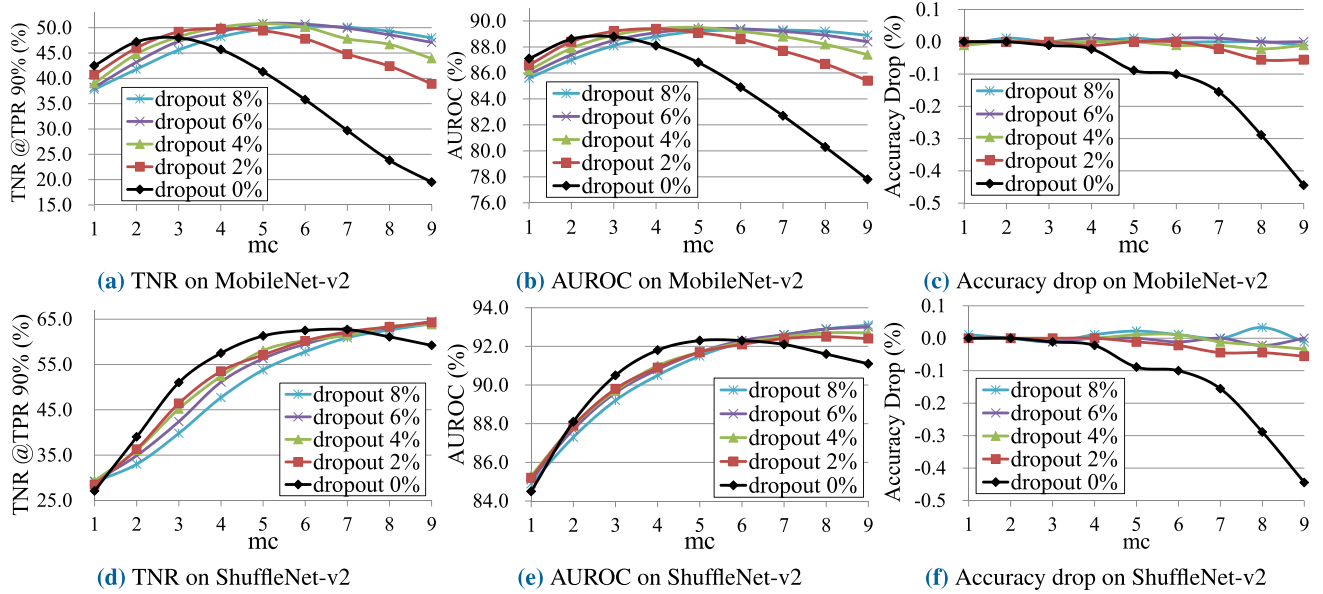


FIGURE 5. Effect of dropout trial (*mc* of x-axis) and dropout ratio (in legend) on TNR, AUROC and accuracy drop.

is reproduced equivalently when the dropout trial (*mc*) is one and dropout ratio is 0%. When the dropout trial (*mc*) is 1 in Figs. 5a and 5b, our performance is lower than that of Odin because of the ambiguity of uncertainty due to MC dropout. However, as the dropout trial (*mc*) increased, the performance improved. Because the gradient of Odin is deterministic, as the perturbation of Odin accumulates, the perturbed image of Odin got distorted and the performance dropped. However, the gradients in our proposed method exhibited stochastic characteristics. Owing to these characteristics, the more they accumulate, the more likely they are to average out and leave only stochastic frequently occurring information. This phenomenon can be clearly observed in Fig. 2, which leads to an improvement in the out-of-distribution sample detection performance, without any drop in the accuracy against Odin, as shown in Figs. 5c and 5f.

We applied MC dropout at the end of each basic building block unit of the network architecture. For example, ShuffleNet-v2 consists of a basic unit and a unit for spatial down-sampling, whereas DenseNet-BC100 consists of a basic unit and a bottleneck unit. In contrast, MobileNet-v2 consists of only one basic unit. Different network architectures have different structures and number of building blocks; therefore, the influence of MC dropout is also different. This means that the degree of effectiveness of the dropout trial (*mc*) and dropout ratio may vary for each architecture and in-distribution dataset. Therefore, Figs. 5a and 5d, and 5b and 5e have different optimal points.

B. ANALYSIS ON UA-FGSM

1) EFFECT OF EPISTEMIC UNCERTAINTY

The sum of the distances between the logit of the target class $f_y(\mathbf{x})$ used for gradient calculation and the logits of the

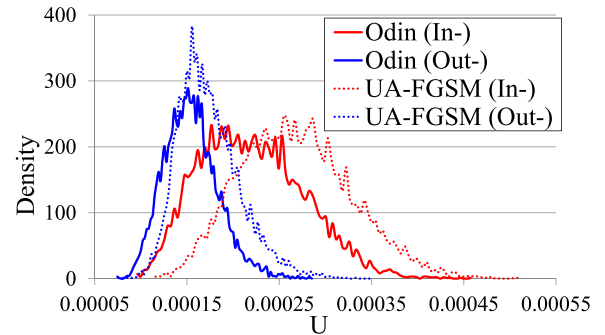


FIGURE 6. Density of $U(\mathbf{x})$ from (4).

remaining classes $f_i(\mathbf{x})$ can be defined as follows:

$$U(\mathbf{x}) = \frac{1}{N-1} \sum_{i \neq y} [f_y(\mathbf{x}) - f_i(\mathbf{x})], \quad (4)$$

where N denotes the number of classes. We aim to understand how epistemic uncertainty affects the confidence score through $U(\mathbf{x})$. Fig. 6 shows the density of (4). The gradient in the in-distribution sample is likely to be calculated as the target for the same class, whereas the gradient in the OOD sample is likely to be distributed to various classes owing to its epistemic uncertainty. This phenomenon becomes more apparent when the prediction is not confident, that is, when the value of $U(\mathbf{x})$ is low. In the case of the in-distribution sample, the proposed UA-FGSM (red dotted line) generally shifts Odin's distribution (red solid line) to the right (i.e., the predictions are over-confident) independent of the value of $U(\mathbf{x})$. However, in the case of an OOD sample in the low $U(\mathbf{x})$ region, where $U(\mathbf{x})$ is less than 0.00015 in Fig. 6, we can observe that the UA-FGSM (blue dotted line) shifts

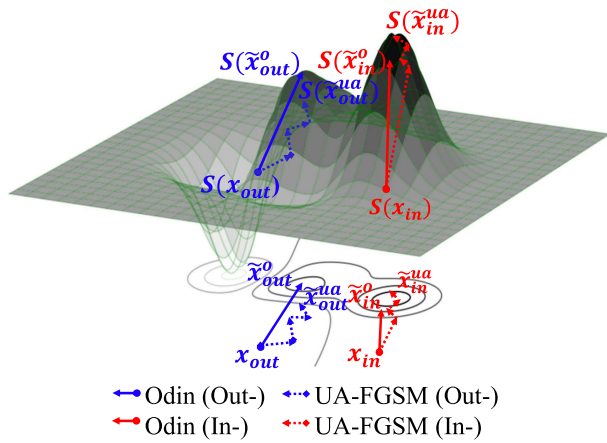


FIGURE 7. Illustration of the effect of additive way.

Odin’s distribution (blue solid line) to the left, making it relatively less over-confident. This effect of epistemic uncertainty makes it possible to separate the in-distribution and OOD samples and intuitively improves the detection performance.

2) EFFECT OF ADDITIVE WAY

The scale of perturbation of Odin is a very sensitive parameter that significantly affects the detection performance. Therefore, tuning it empirically is time-consuming. The effects of the proposed additive method on the gradient domain are illustrated in Fig. 7. The upper index of all points is Odin, denoted by “o” with a solid line, and UA-FGSM is denoted by “ua” with a dashed line. $S(\mathbf{x})$ denotes the softmax scores before each detection method is applied and $S(\tilde{\mathbf{x}})$ is the softmax value of the calculated confidence score. In the case of an in-distribution sample \mathbf{x}_{in} (red), Odin attempted to reach the optimum point by manual tuning ϵ . It is sometimes inadequate and overflowing. However, our method tends to approach global optimum through several steps, even though it is not an optimization method. In the case of the OOD sample \mathbf{x}_{out} (blue), it operated contrarily to the above interpretation. Odin attempted to predict the incorrect answer with over-confidence. This lead to reaching a local optimum with a large confidence value.

However, our method added gradients to relatively diverse directions compared to Odin. This prevented the prediction of incorrect answers from becoming over-confident and therefore, did not converge at a local optimum. This lead to a relatively low confidence value and improved the discrimination ability. Furthermore, owing to the addition effect of stochastic gradients, in our approach, as the dropout trial (mc) accumulates, the gradient in the image domain approaches zero, as shown in Fig. 2. This method has the advantage of being highly robust in terms of accuracy degradation. Moreover, an interesting phenomenon is observed in Figs. 5c and 5f. When the dropout trial (mc) is increased, the detection performance and accuracy sometimes improve at the same time.

C. DIFFERENCE WITH PREVIOUS WORK

In this study, we proposed a method to further boost the OOD sample detection capability through MC dropout in gradient computation for the first time. More importantly, our main contribution is to improve the existing gradient-based perturbation used by Liang et al. [2], which makes it difficult to preserve natural accuracy, in an additive way. Therefore, we proposed a secure OOD sample detection framework without any accuracy drop, even when aggressive perturbation is applied.

The iterative FGSM (I-FGSM) proposed by Kurakin et al. [7] is highly cited in the field of adversarial attacks and defense. However, this technique has never been used as a reverse-way concept in the OOD sample detection field that is targeted in this study. In addition, [7] proposed a method that finds a solution iteratively according to time t , as can be seen by its name and the formula in (2). Therefore, the computation time for the method in [7] was proportional to the number of iterations. In general, a value of 10 determined empirically is mainly used as the ratio in the field of adversarial attacks or defense. In contrast, our proposed UA-FGSM is a method that finds a solution by simultaneously adding mc -trials calculated as mini-batch operations in just one iteration. This means that when I-FGSM is used instead of UA-FGSM, it takes approximately 10 times the computational time of the latter.

In addition, when I-FGSM was applied in this study, the perturbed sample became closer to the target class as time t increased. This does not guarantee that the perturbed samples do not cross the decision boundaries between the classes. Therefore, while the proposed UA-FGSM was observed to prevent the accuracy drop, as shown in Figs. 5c and 5f, the I-FGSM did not guarantee a drop in accuracy as time t increased.

For GTX 1080, Cuda 10.0, and Cudnn 9.0, the average execution time for 1,000 samples was measured to be approximately 45 s and 48 s for [2] and the proposed method, respectively. Because the MC-dropout operation is implemented for application to a mini-batch, it can be observed that there is no significant difference from [2]. In contrast, [7] took approximately 410 s when t was set to 10.

D. SENSITIVITY OF ALGORITHM

We measured the sensitivity of the algorithm using ten different random seeds, and the results were identical. The optimal MC dropout ratio used in the experiment was 1%, which is probably the reason for no change.

VI. CONCLUSION

In this paper, we propose a simple and effective method for improving OOD sample detection performance. We propose UA-FGSM to capture the epistemic uncertainty from only the pre-trained model. In addition, the accumulation of uncertainty-based gradients further improves the detection performance. Our method is only a simple pre-processing, which does not require any information about the

in-distribution and OOD datasets. In addition, our approach does not require heavy computation to retrain or ensemble the pre-trained model. Finally, we verified the usefulness of the proposed method using various model architectures and datasets and attempted to understand the effectiveness of the proposed method through various analyses of the experimental results.

VII. FURTHER WORK

In a certain combination of the network architecture and noise dataset, we observed that the TNR was extremely small. For example, when we measured the performance in an experimental environment where the in-distribution samples were from CIFAR-100, OOD samples were Gaussian, and the network was MobileNet-v2, the detection accuracy was high (87.7%), while the TNR value was zero, as given in Table 2. More specifically, the distribution of the confidence score on Gaussian noise may be higher than that of CIFAR-100, implying that MobileNet-v2 trained with CIFAR-100 may provide highly over-confident predictions for Gaussian noise samples. We continue to study and analyze this phenomenon in depth to make it work properly. Based on this analysis, we would like to improve the current algorithm to achieve better performance and robustness.

REFERENCES

- [1] D. Hendrycks and K. Gimpel, "A baseline for detecting misclassified and out-of-distribution examples in neural networks," in *Proc. ICLR*. Toulon, France, 2017. [Online]. Available: <https://openreview.net/forum?id=Hkg4TI9xl>
- [2] S. Liang, Y. Li, and R. Srikant, "Enhancing the reliability of out-of-distribution image detection in neural networks," in *Proc. ICLR*. Vancouver, BC, Canada, 2018. [Online]. Available: <https://openreview.net/forum?id=H1VGkIxRZ>
- [3] K. Lee, H. Lee, K. Lee, and J. Shin, "Training confidence-calibrated classifiers for detecting out-of-distribution samples," in *Proc. ICLR*. Vancouver, BC, Canada, 2018. [Online]. Available: <https://openreview.net/forum?id=ryiAv2xAZ>
- [4] A. Vyas, N. Jammalamadaka, X. Zhu, D. Das, B. Kaul, and T. Willke, "Out-of-distribution detection using an ensemble of self supervised leave-out classifiers," in *Proc. ECCV*. Munich, Germany, Sep. 2018, pp. 550–564.
- [5] K. Lee, K. Lee, H. Lee, and J. Shin, "A simple unified framework for detecting out-of-distribution samples and adversarial attacks," in *Proc. Neurips*. Montreal, QC, Canada, 2018, pp. 7167–7177.
- [6] I. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," in *Proc. ICLR*. San Diego, CA, USA, 2015. [Online]. Available: <https://arxiv.org/abs/1412.6572>
- [7] A. Kurakin, I. Goodfellow, and S. Bengio, "Adversarial machine learning at scale," in *Proc. ICLR*. Toulon, France, 2017. [Online]. Available: <https://openreview.net/forum?id=BJm4T4Kgx>
- [8] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.
- [9] N. Ma, X. Zhang, H. Zheng, and J. Sun, "ShuffleNet V2: Practical guidelines for efficient CNN architecture design," in *Proc. ECCV*. Munich, Germany, Sep. 2018, pp. 116–131.
- [10] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.
- [11] A. Krizhevsky. (2009). *Learning Multiple Layers of Features From Tiny Images*. Accessed: Jan. 1, 2022. [Online]. Available: <https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf>
- [12] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
- [13] F. Yu, A. Seff, Y. Zhang, S. Song, T. Funkhouser, and J. Xiao, "LSUN: Construction of a large-scale image dataset using deep learning with humans in the loop," 2015. *arXiv:1506.03365*.
- [14] P. Xu, K. A. Ehinger, Y. Zhang, A. Finkelstein, S. R. Kulkarni, and J. Xiao, "TurkerGaze: Crowdsourcing saliency with webcam based eye tracking," 2015. *arXiv:1504.06755*.
- [15] A. Kendall and Y. Gal, "What uncertainties do we need in Bayesian deep learning for computer vision," in *Proc. Neurips*. Long Beach, CA, USA, 2017, pp. 5574–5584.
- [16] M. Andrey and M. Gales, "Predictive uncertainty estimation via prior networks," in *Proc. Neurips*. Montreal, QC, Canada, 2018, pp. 7047–7058.
- [17] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," in *Proc. Neurips*. Long Beach, CA, USA, 2017, pp. 6402–6413.
- [18] N. Carlini and D. Wagner, "Adversarial examples are not easily detected: Bypassing ten detection methods," in *Proc. 10th ACM Workshop Artif. Intell. Secur.*, Nov. 2017, pp. 3–14.
- [19] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proc. ICML (Proceedings of Machine Learning Research)*, New York, NY, USA, vol. 48, 2016, pp. 1050–1059. [Online]. Available: <https://proceedings.mlr.press/>
- [20] Y. Gal, R. Islam, and Z. Ghahramani, "Deep Bayesian active learning with image data," in *Proc. ICML (Proceedings of Machine Learning Research)*, Sydney, NSW, Australia, vol. 70, 2017, pp. 1183–1192. [Online]. Available: <https://proceedings.mlr.press/>
- [21] Y. Mu, W. Liu, X. Liu, and W. Fan, "Stochastic gradient made stable: A manifold propagation approach for large-scale optimization," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 2, pp. 458–471, Feb. 2017, doi: [10.1109/TKDE.2016.2604302](https://doi.org/10.1109/TKDE.2016.2604302).
- [22] T. Mensink, J. Verbeek, F. Perronnin, and G. Csurka, "Distance-based image classification: Generalizing to new classes at near-zero cost," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2624–2637, Nov. 2013, doi: [10.1109/TPAMI.2013.83](https://doi.org/10.1109/TPAMI.2013.83).
- [23] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2012.
- [24] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: <http://www.deeplearningbook.org>
- [25] A. Kendall, "Geometry and uncertainty in deep learning for computer vision," Ph.D. dissertation, Dept. Eng., Univ. Cambridge, Cambridge, U.K., 2018.
- [26] Y. Gal, "Uncertainty in deep learning," Ph.D. dissertation, Dept. Eng., Univ. Cambridge, Cambridge, U.K., 2016.
- [27] C. Chow, "On optimum recognition error and reject tradeoff," *IEEE Trans. Inf. Theory*, vol. IT-16, no. 1, pp. 41–46, Jan. 1970, doi: [10.1109/TIT.1970.1054406](https://doi.org/10.1109/TIT.1970.1054406).
- [28] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *J. Mach. Learn. Res.*, vol. 12, pp. 2121–2159, Feb. 2011.
- [29] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, "Learning hierarchical features for scene labeling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1915–1929, Aug. 2013, doi: [10.1109/TPAMI.2012.231](https://doi.org/10.1109/TPAMI.2012.231).
- [30] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006, doi: [10.1016/j.patrec.2005.10.010](https://doi.org/10.1016/j.patrec.2005.10.010).
- [31] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, Jan. 2013, doi: [10.1109/TPAMI.2012.59](https://doi.org/10.1109/TPAMI.2012.59).
- [32] B. M. Lake, R. Salakhutdinov, and J. B. Tenenbaum, "Human-level concept learning through probabilistic program induction," *Science*, vol. 350, no. 6266, pp. 1332–1338, 2015, doi: [10.1126/science.aab3050](https://doi.org/10.1126/science.aab3050).
- [33] Z. Zeng, R. Kang, M. Wen, and E. Zio, "A model-based reliability metric considering aleatory and epistemic uncertainty," *IEEE Access*, vol. 5, pp. 15505–15515, 2017, doi: [10.1109/ACCESS.2017.2733839](https://doi.org/10.1109/ACCESS.2017.2733839).
- [34] H. D. Kabir, A. Khosravi, M. A. Hosen, and S. Nahavandi, "Neural network-based uncertainty quantification: A survey of methodologies and applications," *IEEE Access*, vol. 6, pp. 36218–36234, 2018, doi: [10.1109/ACCESS.2018.2836917](https://doi.org/10.1109/ACCESS.2018.2836917).
- [35] V. Freschi, S. Delpriori, E. Lattanzi, and A. Bogliolo, "Bootstrap based uncertainty propagation for data quality estimation in crowd-sensing systems," *IEEE Access*, vol. 5, pp. 1146–1155, 2017, doi: [10.1109/ACCESS.2017.2651942](https://doi.org/10.1109/ACCESS.2017.2651942).

- [36] K. Xiao, J. Zhao, Y. He, C. Li, and W. Cheng, "Abnormal behavior detection scheme of UAV using recurrent neural networks," *IEEE Access*, vol. 7, pp. 110293–110305, 2019, doi: [10.1109/ACCESS.2019.2934188](https://doi.org/10.1109/ACCESS.2019.2934188).
- [37] M. Lin, T. Wang, X.-B. Li, J. Liu, Y. Wang, Y. Zhu, and W.-P. Wang, "An uncertainty-incorporated approach to predict the winner in StarCraft II using neural processes," *IEEE Access*, vol. 7, pp. 101609–101619, 2019, doi: [10.1109/ACCESS.2019.2930581](https://doi.org/10.1109/ACCESS.2019.2930581).
- [38] W. Alhakami, A. Alharbi, S. Bourouis, R. Alroobaea, and N. Bouguila, "Network anomaly intrusion detection using a nonparametric Bayesian approach and feature selection," *IEEE Access*, vol. 7, pp. 52181–52190, 2019, doi: [10.1109/ACCESS.2019.2912115](https://doi.org/10.1109/ACCESS.2019.2912115).
- [39] B. Hussain, Q. Du, S. Zhang, A. Imran, and M. A. Imran, "Mobile edge computing-based data-driven deep learning framework for anomaly detection," *IEEE Access*, vol. 7, pp. 137656–137667, 2019, doi: [10.1109/ACCESS.2019.2942485](https://doi.org/10.1109/ACCESS.2019.2942485).
- [40] W. Yu, Z. Ding, C. Hu, and H. Liu, "Knowledge reused outlier detection," *IEEE Access*, vol. 7, pp. 43763–43772, 2019, doi: [10.1109/ACCESS.2019.2906644](https://doi.org/10.1109/ACCESS.2019.2906644).
- [41] S. U. Jan, S. Ahmed, V. Shakhov, and I. Koo, "Toward a lightweight intrusion detection system for the Internet of Things," *IEEE Access*, vol. 7, pp. 42450–42471, 2019, doi: [10.1109/ACCESS.2019.2907965](https://doi.org/10.1109/ACCESS.2019.2907965).



DOKWAN OH received the B.S. degree in electronic engineering and the Ph.D. degree in electrical and electronic engineering from KAIST, Daejeon, South Korea, in 2002 and 2009, respectively. He was a Visiting Researcher with UC Irvine, University of Pennsylvania, and RIKEN, in 2002, 2005, and 2010, respectively. He was a Senior Engineer with Health and Medical Equipment Business, Samsung Electronics, Suwon, South Korea, from 2011 to 2015. He was a Principal Researcher with the Samsung Advanced Institute of Technology (SAIT), Suwon, from 2015 to 2017. He has been a Visiting Scholar at the MILA Quebec AI Institute, Montreal, QC, Canada, from 2018 to 2020. He is currently a Principal Researcher with SAIT, South Korea. His current research interests include machine learning, deep learning, and computer vision.



DAEHYUN JI received the B.S. degree in electronic engineering from Kyungbuk National University, Daegu, South Korea, in 2003, and the Ph.D. degree in electrical and electronic Engineering from POSTECH, Pohang, South Korea, in 2009. He was a Senior Software Engineer with the Mobile Division, Samsung Electronics, Suwon, South Korea, from 2009 to 2015. He was a Research Master and a Principal Researcher with the Samsung Advanced Institute of Technology (SAIT), Suwon, from 2015 to 2017. He was a Visiting Scholar at MILA Quebec AI Institute, Montreal, QC, Canada, from 2018 to 2020. He is currently a Research Master and a Principal Researcher with Samsung Advanced Institute of Technology (SAIT), Suwon. His current research interests include deep learning, computer vision, ADAS and autonomous driving, nonlinear control, and robotics. He has published over 30 international papers in these areas.



OHMIN KWON received the B.S. degree in electronic engineering from Kyungbuk National University, Daegu, South Korea, in 1997, and the Ph.D. degree in electrical and electronic engineering from POSTECH, Pohang, South Korea, in 2004. He was a Senior Researcher with the Mechatronics Center, Samsung Heavy Industries, Daejeon, from 2004 to 2006. He was a Visiting Scholar at Texas A&M University, College Station, TX, USA, from 2017 to 2018. He is currently a Professor with the School of Electrical Engineering, Chungbuk National University, Cheongju, South Korea. His current research interests include time-delay systems, cellular neural networks, robust control and filtering, large-scale systems, secure communication through synchronization between two chaotic systems, complex dynamical networks, multiagent systems, and sampled data control. He has published over 170 international papers in these areas. He was named "One of the Highly Cited Researchers in the field of Mathematics," in 2015, 2016, and 2017. He currently serves as an Associate Editor for *Neural Networks*, *International Journal of Control, Automation and Systems*, *Journal of Institute of Control, Robotics and Systems*, and *Journal of Applied Mathematics and Informatics*.



YOONSUK HYUN received the B.S. degree in mathematics from Seoul National University, Seoul, South Korea, in 2002, and the Ph.D. degree in mathematics from the MIT, Cambridge, USA, in 2011. He was a Research Fellow at KIAS, from 2011 to 2015, and a Senior Researcher at the Samsung Advanced Institute of Technology (SAIT), from 2015 to 2020. He has been working as an Assistant Professor with the Department of Mathematics, Inha University, Incheon, South Korea, since 2020. His current research interests include machine learning, deep learning, and computer vision.

...