EnLPF: Entropy Low Pass Filter, Edge Effect Elimination in Out-of-Distribution Detection

A Preprint

Butian Xiong*
School of Science and Engineering
Chinese University of Hong Kong, Shenzhen
Shenzhen, Guangdong, China
120090584@link.cuhk.edu.cn
Hajun Lee
School of Data Science
Chinese University of Hong Kong, Shenzhen
Shenzhen, Guangdong, China
117010437@link.cuhk.edu.cn

December 25, 2022

Abstract

Detecting out-of-distribution (OOD) inputs is a central challenge for safely deploying machine learning models in the real world. Existing solutions for scene recognition usually combine the semantic information with the original classification model. While it usually ignores out-of-distribution samples in semantic segmentation. The current study provides a method to solve a classification out of distribution problem using a segmentation mask generated by entropy. Then we use an entropy mask to prevent misinformation in the segmentation method. Finally, we use a low pass filter to eliminate the edge effect on the entropy map. We achieve more than two per cent improvement on each of the datasets, including MIT67, SUNRGBD, and Places365 compared to the original SOTA method.

Keywords Out-of-Distributiom · Scene Recognition · Low Pass Filter · Fourier Transformation · Entropy

1 Introduction

Out-of-distribution (OOD) describes a situation when the classes of test data is not entirely overlapping with the original training dataset's classes. OOD detection is vital for safely deploy machine learning models to open-world situations. To eliminate this issue, multiple different OOD detection method has mergedYang et al. [2022]Wei et al. [2022]Huang et al. [2021]Sun et al. [2022]Zaeemzadeh et al. [2021]Sun et al. [2021]Yang et al. [2021]Du et al. [2022]Huang and Li [2021]Liu et al. [2020]Jung et al. [2021]. Most of the state of the art methods for OOD detection in scene recognition usually combine the semantic information with the original classification methodYang et al. [2021]López-Cifuentes et al. [2020]Miao et al. [2021]. However, the assumption those methods make is the object semantic detect are usually in distribution. After an extensive experiment, we found out that there are many objects in the semantic model that are actually out of distribution. Therefore, the semantic model sometimes may pass the wrong information to the classification model.

To solve the problem mentioned above, we used an entropy-based method that can be served as a score to help us to mask the semantic model's output. However, we find out that the edge between two objects will also have a very high entropy. In other words, edges will also be classified as out of distribution as figure 1 shows.

*Responsible for code implementation, algorithm design, presentation, writing the paper

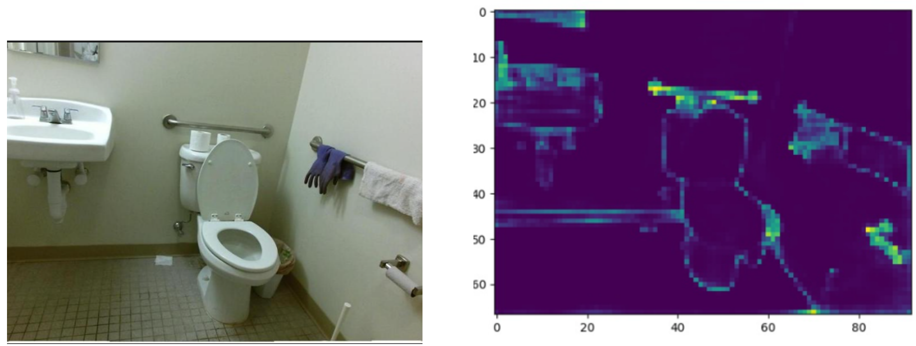

Figure 1: (a) Original input for OOD detection, the gloves, toilet paper, and iron tube are out of distribution (b) The entropy map that represent which pixel has lowest confidence. Despite the fact that iron tube, gloves and toilet paper can be easily regarded as OOD, the edges between toilet and wall can also be regarded as out of distribution

To eliminate this problem issue, we provide a pipeline that can reduce the edge effect. In a nutshell, our method mainly includes the following three parts. First, transform the semantic probability result to an entropy map. Second, apply a strict low pass filter to the original entropy map and do the normalization. Thirdly, using thresholding to mask corresponding semantic output.

2 Related Work

Out of Distribution Detection. OOD detection aims to distinguish test images that come from different classes compared to training datasetHendrycks and Gimpel [2016]. The basic method for OOD detection is Max Softmax Probability(MSP)Hendrycks and Gimpel [2016]. To amplify the difference of softmax output for different classes, an energy-based method that adjust the temperature of the softmax function also came into beenLiu et al. [2020]. On the other hand, one can also use the output that does not come through the softmax. layerHendrycks et al. [2019]. To simplify the threshold selection procedure, standardization methods are also considered usefulJung et al. [2021]. Instead of considering output on the logits layer, some also considered clustering on the feature spacesZaeemzadeh et al. [2021]. Those methods mentioned above only use training data without importing new data.

Despite some works importing a small amount of out-of-distribution data into the training procedureYu and Aizawa [2019]Li and Vasconcelos [2020]. Some fascinating methods import human-synthesized data as out-of-distribution data, which can dramatically improve the accuracy of OOD detectionDu et al. [2022].

As for scene recognition, despite using substantial computing resources and extremely large training datasetSingh et al. [2022], an easier yet reliable method is by combining multi-modal information such as semantic and classification modalMiao et al. [2021]Huang and Li [2021].

3 Problem Setup

3.1 Close samples

The close samples and far samples are usually hard to define. But it is easy to understand by listing some examples. Relative samples include dogs, cats, lions, tigers, and so on. Far models include satellite pictures of the pacific ocean, x-ray pictures of human chest, human-written numbers and son on. In the current study, we are dealing with close OOD samples. That means the in-distribution classes are similar to the OOD samples.

3.2 No External Data

In our experiment, we do not provide extra dataset to our classification model.

3.3 Training Procedure

We are considering a multi-classes classification problem with K classes. For each sample in training dataset, it has a pair of data $(x_i, y_i)_{i=1}^{N}$, where N is the sample size of training datatset size. Note that $y_i \in 1...K$ The parameters of model is θ, during training we have:

$$\theta_c(\theta_r(x_i), \theta_s(x_i)) = p_i \tag{1}$$

$$\mathcal{L}_{\mathcal{CE}}(p_i, y_i) \tag{2}$$

Whereθ_c is the classifier of the model, θ_s is the semantic segmentation model, and θ_s is the backbone of a classification model. $\mathcal{L}_{\mathcal{CE}}$ is a cross entropy loss function.

3.4 OOD Detection

After obtaining the reliable model, we start to make the inference. We will select the sample with low max probability as an out-of-distribution class. if $max(\theta(x_i)) < thre$, then we will regard this sample as out of distribution. Function max will give the max value of the input vector. As for the value of there, it is a hyperparameter that can be manually adjusted.

4 Method

During this section, we introduce our method, entropy low pass filter to make the semantic segmentation model generate more reliable information to interpreted.

4.1 Semantic Map

At the end of the semantic segmentation model, there will be an output for each pixel.

$$argmax(\theta_s(x_i)) = S_i \quad S_i \in R^{H\times W} \tag{3}$$

$$\theta_s(x_i) = SP_i \quad SP_i \in R^{H\times W\times O} \tag{4}$$

Where x_i represents the i^{th} sample of the test dataset. As for the argmax function, it will return the argument position whose value is the max among others. S_i is the semantic map of the i^{th} sample of test dataset. H and W are actually two parameters that define the hight and width of the semantic map.

But, if we do not directly apply the argmax function, we will get a distribution for different pixel. Therefore, O represents the number of object classes that can be identified by semantic segmentation model.

4.2 Entropy Map

Denote $SP_i(x, y)$ represents the distribution of the pixel at coordinate (x,y). To simplify the equation, we denote SP_i as q Then the entropy value Ent can be given as:

$$Ent(SP_i(x, y)) = Ent(q) = -\sum_{i=1}^{O} q_i ln q_i \tag{5}$$

After applying this to all pixel of the semantic map, we will have an entropy map, which can be denoted as $Ent(SP_i)$ or simpy written as Ent_i

4.3 Edge Effect Elimination

To eliminate the edge, we first apply a low pass filter to original entropy map.

$$F(k, l) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} Ent(m, n) e^{-j2\pi(\frac{k}{M}m + \frac{l}{N}n)} \tag{6}$$

$$F_{low} = F(k, l) * G(x, y) \tag{7}$$

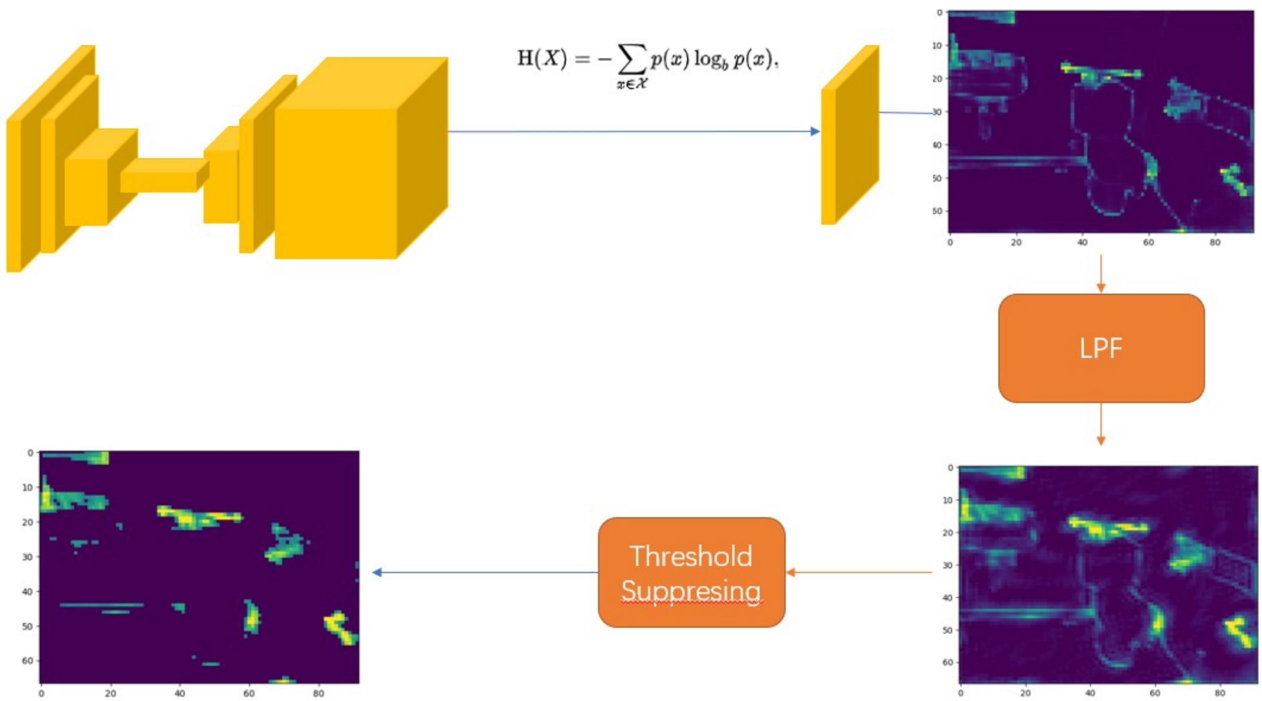

Figure 2: First, we transform the semantic segmentation output to an entropy map. After that, we apply a low pass filter to the origin entropy map, we use a threshold selection to mask all the parts that have a low entropy value

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{8}$$

$$R(Ent_i) = T(\mathcal{F}_{inverse}(F_{low})) \tag{9}$$

First we will apply discreate Fourier transform to our original entropy map, to get a frequency domain function, denote as as F(k,l),then we apply a Gaussian filter to the frequency domain and then do the inverse Fourier transform. The Gaussian low pass filter is denoted as G(x,y). Where F_{low} denote the image passing through the low pass filter. And after that, the T function which will threshoulding the inverse Fourier transform result obtained from F_{low}.

The whole procedure can be represented in the figure 2

5 Experiment and Result

In this section, we verified the effect of our EntLPF method.

5.1 Experiment Setup

In Distribution Dataset In this experiment, we use SUNRGB-DSong et al. [2015] as in distribution dataset. SUNRGB-D dataset is usually a very hard dataset whose classification accuracy range from 55 per cent to 65 per cent for the different backbone. The size of the different datasets may vary. However, we use a transformation function to resize the input picture to 224*224.

Out of Distribution Dataset In this experiment, we use several scene recognition datasets, including MIT67Quattoni and Torralba [2009], Place365-14Zhou et al. [2017a], and SUN397Xiao et al. [2010] as OOD dataset. We blend the test dataset of in distribution and OOD dataset to test our model's sensitivity.

Table 1: Places365-14 out of distribution result

Method	FPR@95 ↓	Detection Error ↓	AUC. ↑
EnLPF(ours)	**22.4**	**35.0**	**70.6**
Maxlogit Hendrycks et al. [2019] ICML2022	30.9	39.1	64.6
Std-Maxlogit Jung et al. [2021] ICCV2021	25.5	37.4	66.7
Ent-Max Macêdo et al. [2021] TNNLS2021	37.4	43.1	60.2
KL-Prototype	32.4	40.4	63.5

Table 2: MIT67 out of distribution result

Method	FPR@95 ↓	Detection Error ↓	AUC. ↑
Maxlogit EnLPF(Ours)	**16.3**	**31.6**	**73.9**
Maxlogit Hendrycks et al. [2019] ICML2022	22.3	35.0	70.1
Std-Maxlogit Jung et al. [2021] ICCV2021	21.8	35.2	69.1
Ent-Max Macêdo et al. [2021] TNNLS2021	31.4	39.6	64.6
KL-Prototype	32.4	40.5	63.5

5.2 Evaluation Matrix

We evaluate the performance of OOD detection by measuring the following metrics: (1) the false positive rate (FPR95) of OOD examples when the true positive rate of in-distribution examples is 95%; (2) the area under the receiver operating characteristic curve (AUROC); and (3) the detection error

5.3 Training Details

We use a ResNet-based semantic segmentation model together with an attention module called OFAMMiao et al. [2021] to train on the ADE20KZhou et al. [2017b] for 20 epochs.

As for the classification model, we first pre-train our model on the places365 dataset with the hyper-parameters illustrated below. We use the cosine learning scheduler with warm-up epochs set as 20, weight decay set as 0.3, base learning rate as 0.02, warm-up learning rates 5×10^{-6}. After training for 100 epochs, we then do the transfer learning in the distribution dataset with same hyperparameters except we freeze training of backbone for 60 epochs. We conduct pretrained of segmentation model and classification model on 8 NVIDIA GeForce RTX 2080Ti and implement all methods with default parameters using PyTorch.

5.4 result

The result we obtained are illustrated in table1 and table 2. We set our in-distribution dataset as SUNRGBD we set ood dataset as Places365-14, which means we only extract 14 classes from the standard places365 dataset. The result in the table shown here states that our method outperforms better than the current SOTA method. However, some experiment on other datasets shows the contrasting result—usually, our method performs well on an experiment whose in-distribution dataset is extremely over-fitting.

6 Discussion

During this section, we will explain why our method works and why in some situations, our method fails. Besides that, we also provide a possible future work that can be done to improve our OOD detection accuracy.

6.1 EnLPF

The reason why our method entropy low pass filter works is that we mask the misinformation from the semantic segmentation model. On the other hand, the procedure of the low pass filter makes the low confidence part more apparent, as for the edge part, it will be ignored. In this circumstance, we usually will output a better result. The procedure during the thresholding is by normalizing and setting a threshold. This procedure will force the low confident part to be masked as an unknow object. Therefore, the following classification model will know which part of the segmentation result should be trusted.

6.2 Failure Cases

There are a lot of failure cases that exist, but we do not show them in this paper since their failure is far before it can show results, as shown in the table above. The first kind of failure is that the segmentation modal is hard to combine with classification features even though they are intuitively understandable. There are a lot of datasets that cannot converge during training when simply combining the segmentation features and classification features. The accuracy of classification is around 10% for 100 epochs.

The second failure case is that even though some datasets converged during training, the classification accuracy is much lower than just using one classification model. For instance, the classification accuracy for just using a classification model is around 80%—the result of two models is usually around 20%.

The third case is actually not a failure. One can quickly notice that the model we have is hugely imbalanced, the segmentation path is much more complex than the classification path. On the other hand, due to the calculation limitation, we can not deploy this model online.

6.3 Furture Work

Although I am not an expert in dealing with the multi-modal problem, I know there are several tricks to alleviate this issue. For instance, re-weightingNgiam et al. [2011] and others.

A more reliable method may need two separate model, one for detecting OOD, and the other used to detect classification. So that we do not lose accuracy on classification when we improve the OOD detection result.

And the most crucial experiment is that we must try to pass both the entropy map and the semantic map together to test a result.

7 Conclusion

In this paper, we propose an entropy low pass filter for scene OOD detection. We first train our semantic model in a standard way on ADE20K. We also train the classification model on Places365 standard dataset. Then we use the training semantic segmentation model to obtain a semantic map and entropy map. Then we apply our low pass filter to the entropy map using Fourier Transformation. After that, we use a thresholding skill to mask the pixel on the semantic map with low confidence score. This score is obtained from an entropy map which already passed the thresholding. Finally, we concatenate the semantic map with features obtained from the classification model. Then we will use the post-processing skills to detect which sample is OOD. Our method performs better than the current SOTA if the training on both modals converges. However, it leaves a lot of questions, including why some datasets converge and others do not.

References

Nianzu Yang, Kaipeng Zeng, Qitian Wu, Xiaosong Jia, and Junchi Yan. Learning substructure invariance for out-of-distribution molecular representations. In *Advances in Neural Information Processing Systems*, 2022.

Hongxin Wei, Renchunzi Xie, Hao Cheng, Lei Feng, Bo An, and Yixuan Li. Mitigating neural network overconfidence with logit normalization. *arXiv preprint arXiv:2205.09310*, 2022.

Rui Huang, Andrew Geng, and Yixuan Li. On the importance of gradients for detecting distributional shifts in the wild. *Advances in Neural Information Processing Systems*, 34:677–689, 2021.

Yiyou Sun, Yifei Ming, Xiaojin Zhu, and Yixuan Li. Out-of-distribution detection with deep nearest neighbors. *arXiv preprint arXiv:2204.06507*, 2022.

Alireza Zaeemzadeh, Niccolo Bisagno, Zeno Sambugaro, Nicola Conci, Nazanin Rahnavard, and Mubarak Shah. Out-of-distribution detection using union of 1-dimensional subspaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9452–9461, 2021.

Yiyou Sun, Chuan Guo, and Yixuan Li. React: Out-of-distribution detection with rectified activations. *Advances in Neural Information Processing Systems*, 34:144–157, 2021.

Jingkang Yang, Haoqi Wang, Litong Feng, Xiaopeng Yan, Huabin Zheng, Wayne Zhang, and Ziwei Liu. Semantically coherent out-of-distribution detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8301–8309, 2021.

Xuefeng Du, Zhaoning Wang, Mu Cai, and Yixuan Li. Vos: Learning what you don't know by virtual outlier synthesis. *arXiv preprint arXiv:2202.01197*, 2022.

Rui Huang and Yixuan Li. Mos: Towards scaling out-of-distribution detection for large semantic space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8710–8719, 2021.

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.

Sanghun Jung, Jungsoo Lee, Daehoon Gwak, Sungha Choi, and Jaegul Choo. Standardized max logits: A simple yet effective approach for identifying unexpected road obstacles in urban-scene segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15425–15434, 2021.

Alejandro López-Cifuentes, Marcos Escudero-Viñolo, Jesús Bescós, and Álvaro García-Martín. Semantic-aware scene recognition. *Pattern Recognition*, 102:107256, 2020.

Bo Miao, Liguang Zhou, Ajmal Saeed Mian, Tin Lun Lam, and Yangsheng Xu. Object-to-scene: Learning to transfer object knowledge to indoor scene recognition. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2069–2075. IEEE, 2021.

Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. *arXiv preprint arXiv:1610.02136*, 2016.

Dan Hendrycks, Steven Basart, Mantas Mazeika, Mohammadreza Mostajabi, Jacob Steinhardt, and Dawn Song. Scaling out-of-distribution detection for real-world settings. *arXiv preprint arXiv:1911.11132*, 2019.

Qing Yu and Kiyoharu Aizawa. Unsupervised out-of-distribution detection by maximum classifier discrepancy. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9518–9526, 2019.

Yi Li and Nuno Vasconcelos. Background data resampling for outlier-aware classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13218–13227, 2020.

Mannat Singh, Laura Gustafson, Aaron Adcock, Vinicius de Freitas Reis, Bugra Gedik, Raj Prateek Kosaraju, Dhruv Mahajan, Ross Girshick, Piotr Dollár, and Laurens van der Maaten. Revisiting weakly supervised pre-training of visual perception models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 804–814, 2022.

Shuran Song, Samuel P Lichtenberg, and Jianxiong Xiao. Sun rgb-d: A rgb-d scene understanding benchmark suite. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 567–576, 2015.

Ariadna Quattoni and Antonio Torralba. Recognizing indoor scenes. In *2009 IEEE conference on computer vision and pattern recognition*, pages 413–420. IEEE, 2009.

Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. Places: A 10 million image database for scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017a.

Jianxiong Xiao, James Hays, Krista A Ehinger, Aude Oliva, and Antonio Torralba. Sun database: Large-scale scene recognition from abbey to zoo. In *2010 IEEE computer society conference on computer vision and pattern recognition*, pages 3485–3492. IEEE, 2010.

Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso, and Antonio Torralba. Scene parsing through ade20k dataset. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 633–641, 2017b.

David Macêdo, Tsang Ing Ren, Cleber Zanchettin, Adriano LI Oliveira, and Teresa Ludermir. Entropic out-of-distribution detection. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2021.

Jiquan Ngiam, Aditya Khosla, Mingyu Kim, Juhan Nam, Honglak Lee, and Andrew Y Ng. Multimodal deep learning. In *ICML*, 2011.