Supplementary Materials for Towards Accurate Open-Set Recognition via Background-Class Regularization

Wonwoo Cho[1,2] and Jaegul Choo[1,2]

[1] KAIST AI
[2] Letsur Inc.
{wcho,jchoo}@kaist.ac.kr

A Algorithm of Our Proposed Method

For a better understanding of the training and inference processes of our method, we summarize the processes in Algorithm 1. In Section 3 of our main paper, we describe in-depth details about each step of the algorithm.

Algorithm 1 Training and inference steps of our method

– **Training** –
Given: DNN f_θ, Datasets $\mathcal{D}_t$ (known classes) and $\mathcal{D}_b$ (background classes)
Parameters: $\boldsymbol{\mu}_c$ (trainable) for $c \in \{1, \cdots, C\}$ and λ
while not converged **do**
  Sample $(\mathbf{x}^k, y) \in \mathcal{D}_t$ and $\mathbf{x}^b \in \mathcal{D}_b$
  for $c = 1$ **to** C **do**
    Compute $D_E^2(f_\theta(\mathbf{x}^k), \boldsymbol{\mu}_c)$ and $D_E^2(f_\theta(\mathbf{x}^b), \boldsymbol{\mu}_c)$
  end for
  Compute $\mathcal{L}_{cf} = \mathbb{E}_{(\mathbf{x}^k, y) \sim \mathcal{D}_t}[-\log P_d(y|\mathbf{x}^k)]$ based on Eq. (5)
  Compute $\mathcal{L}_{bg,k}$ and $\mathcal{L}_{bg,u}$ (See Section 3.4)
  $\mathcal{L} = \mathcal{L}_{cf} + \lambda(\mathcal{L}_{bg,k} + \mathcal{L}_{bg,u})$
  Update network parameter θ and $\boldsymbol{\mu}_c$ for all c
end while

– **Inference** –
Given: Trained f, $\boldsymbol{\mu}_c$ for all c, and a test sample $\mathbf{x}$ from $\mathcal{D}_{test}^k$ or $\mathcal{D}_{test}^u$
Parameters: A threshold τ for unknown-class rejection
for $c = 1$ **to** C **do**
  Compute $D_E^2(f_\theta(\mathbf{x}), \boldsymbol{\mu}_c)$ (Euclidean distance)
end for
if $\max_{c \in \{1, \cdots, C\}} -D_E^2(f_\theta(\mathbf{x}), \boldsymbol{\mu}_c) \geq \tau$ **then**
  $\widehat{y} = \arg\min_{c \in \{1, \cdots, C\}} D_E^2(f_\theta(\mathbf{x}), \boldsymbol{\mu}_c)$
else
  $\widehat{y} = C + 1$ ("unknown" class)
end if

B Previous Regularization Methods

Throughout our main paper, we compared the proposed approach to the following previous regularization methods proposed for Softmax classifiers.

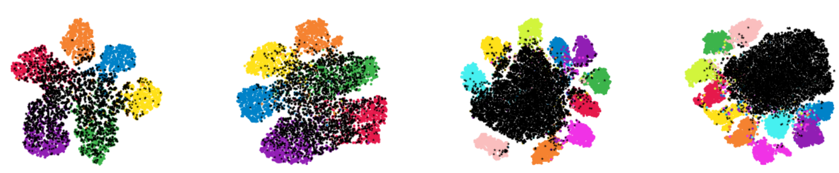
Objectosphere loss [2]. The authors proposed the entropic open-set and the objectosphere losses to regularize SoftMax classifiers. While the entropic open-set loss aims to maximize the information entropy of SoftMax probability for known unknown class (KUC) data, the objectosphere loss increases the gap between known known class (KKC) and KUC samples in terms of their latent feature magnitude. One can distinguish UUCs from KKCs by measuring the maximum value of $P(y|\mathbf{x})$ (probability) or $P(y|\mathbf{x}) \cdot ||f(\mathbf{x})||$ (probability and magnitude).

Uniformity loss [3]. Similar to [2], Hendrycks *et al.* also proposed to maximize the entropy of SoftMax probability for KUCs. UUCs can be distinguished from KKCs by measuring $\max_c P(y=c|\mathbf{x})$ or $-\sum_c P(y=c|\mathbf{x}) \log P(y=c|\mathbf{x})$.

Energy loss [5]. Recently, it was proposed to increase the energy gap between training and background samples, where the energy value can be computed by $-\log \sum_c \exp(\exp(\mathbf{w}_c^T f(\mathbf{x}) + b_c))$ in Eq. (1). One can measure the energy value of each input sample $\mathbf{x}$ to determine whether $\mathbf{x}$ is from KKCs or UUCs.

C Latent Feature Space Visualization

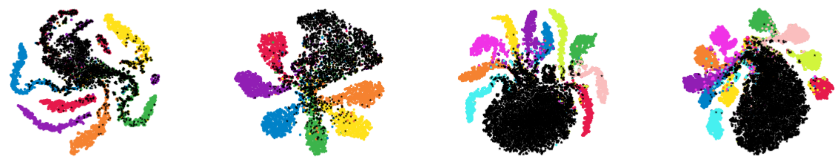
In this section, we split the 10 classes of CIFAR10 into 6 KKCs and 4 UUCs for Setting 1 (`S1`) experiments. For Setting 2 (`S2`), we used the CIFAR10 and the resized ImageNet (ImageNet-R) datasets as KKCs and UUCs, respectively. In the following, we compared Softmax and distance-based classifiers, and their regularized versions via t-SNE [6]. To regularize the Softmax and the distance-based classifiers, we used the uniformity [3] and our class-inclusion losses, respectively. In each t-SNE result, the black dots indicate UUC samples. For the other colors, each color represents a distinct class of KKCs.



(a) `S1` - Softmax (b) `S1` - Distance (c) `S2` - Softmax (d) `S2` - Distance

Fig. 1. t-SNE results of vanilla Softmax classifiers and distance-based classifiers.

Vanilla classifiers. Figure 1 depicts the t-SNE results of vanilla Softmax and distance-based classifiers in `S1` and `S2`. The figure implies that in comparison with `S2`, it would be more difficult to distinguish UUC samples from KKCs in `S1`.



(a) `S1` - Uniform (b) `S1` - Ours (c) `S2` - Uniform (d) `S2` - Ours

Fig. 2. t-SNE results of regularized Softmax classifiers and distance-based classifiers.

Background-class regularized classifiers. Figure 2 presents the t-SNE results of Softmax and distance-based classifiers trained with BCR ($\mathcal{L} = \mathcal{L}_{cf} + \lambda\mathcal{L}_{bg}$) in `S1` and `S2`. In comparison with Figure 1, Figure 2 shows that such regularization techniques assist classifiers to learn more effective latent feature representations in distinguishing UUCs from KKCs, especially in `S1`. Furthermore, the figure shows that the latent feature space regularized by the uniformity loss can yield inaccurate results in distance-based post-classification analysis.

D Training Details for Text Classification

We employed a simple GRU model [1], whose feature dimension is 128. Using the batch size of 64 for both KKCs and KUCs, we trained the model for 15 epochs via the Adam optimizer [4] with the initial learning rate of 0.01 and cosine annealing.

References

1. Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.: Learning phrase representations using rnn encoder-decoder for statistical machine translation. In: The 2014 Conference on Empirical Methods in Natural Language Processing (2014)
2. Dhamija, A.R., Günther, M., Boult, T.: Reducing network agnostophobia. Advances in Neural Information Processing Systems **31**, 9157–9168 (2018)
3. Hendrycks, D., Mazeika, M., Dietterich, T.: Deep anomaly detection with outlier exposure. In: International Conference on Learning Representations (2019)
4. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
5. Liu, W., Wang, X., Owens, J., Li, Y.: Energy-based out-of-distribution detection. Advances in Neural Information Processing Systems **33** (2020)
6. Van der Maaten, L., Hinton, G.: Visualizing data using t-sne. Journal of machine learning research **9**(11) (2008)