MetaMax: Improved Open-Set Deep Neural Networks via Weibull Calibration: Supplementary Material

In this supplement, we provide additional experimental details on MetaMax. First, we describe the training procedure and show results that highlight the sensitivity of MetaMax to different hyperparameters. Next, we illustrate the performance differences among SoftMax, OpenMax, and MetaMax by analyzing their respective ROC curves. We also use scatter plots to illustrate the correlation between class activations and mean distances. Finally, we demonstrate the applicability of MetaMax by providing comparisons against SoftMax and OpenMax on DenseNet [2], ResNet [1], and VGGNet [5].

1. Training Parameters

In the first set of experiments where we showed the superiority of MetaMax over OpenMax, we trained DenseNet121 for 200 epochs with a batch size of 64 for all three datasets. DenseNet121 has a total of 6,960,006 parameters in its 121 layers. Training was performed using the Adam [3] optimization method. The learning rate decayed by a multiplicative factor γ at the end of each epoch. In the second set of experiments where we demonstrated the applicability of MetaMax on different classification networks, we conducted experiments using ResNet and VGGNet following the same experimental protocol.

2. Sensitivity to Hyperparameters

The hyperparameter q in Algorithm 1 is passed to the FitHigh function along with the obtained non-match scores. q controls the number of top non-match activations we choose to fit the per-class Weibull model during the inference step. We performed an experiment to show the sensitivity of MetaMax to this hyperparameter. To do this, we varied the range of q from 2 to 30. As shown in Table 1, the results are very robust to these changes.

q	**F1-Score** $\uparrow$	**AUROC** $\uparrow$
2	0.71107668	0.93882572
5	0.71114378	0.93882158
10	0.71109837	0.93880752
20	0.71141039	0.93878481
30	0.71089159	0.93872475

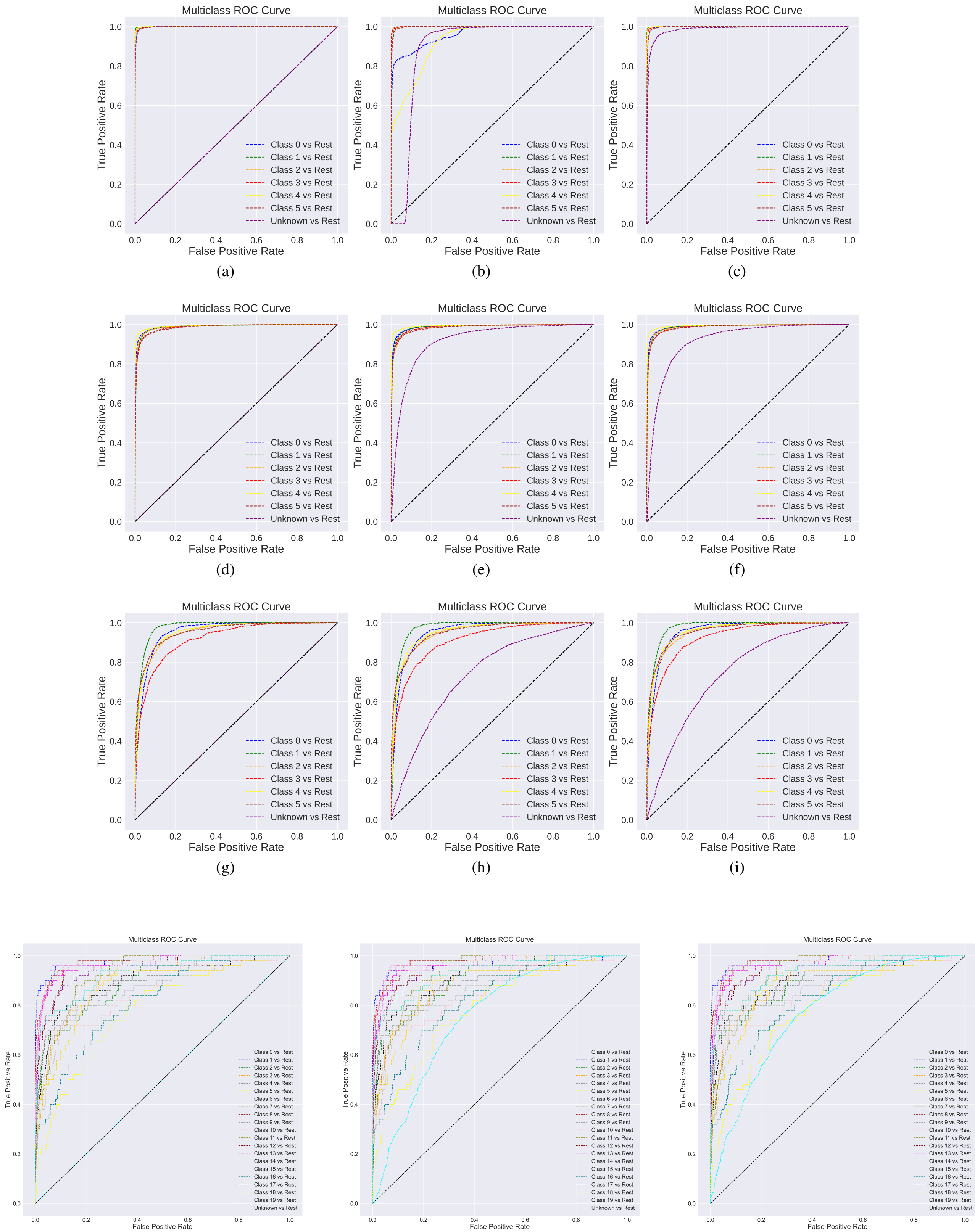
Table 1: The F1 and AUROC scores of MetaMax using DenseNet on CIFAR-10 with different values of the hyperparameter q.

3. ROC Curves

ROC curves demonstrating the magnitude of difficulty SoftMax, OpenMax, and MetaMax have with each dataset are shown in Fig. 1. In the first column of Fig. 1, the SoftMax activation function has steeper ROC curves for all classes except the unknown class. However, SoftMax's unknown ROC curve lies directly on top of the non-discrimination line, which dramatically hurts performance. OpenMax and MetaMax both have steep ROC curves for the unknown class. Nevertheless, MetaMax has a steeper curve when compared against OpenMax. In comparison to SoftMax and OpenMax, this aligns with the higher AUROC scores for MetaMax as shown in the main paper. The area under each class's ROC curve can be interpreted as being inversely related to the difficulty of picking a sample from that class against all other classes.

4. Scatter Plots

Intuitively, class activations offer a measure of similarity between a sample and its implicitly-stored mean activation vectors in the latent space. We present empirical evidence of this in Fig. 2 using the SVHN dataset. Fig. 2a shows a scatter plot for the activation vectors of class 4. The distance of each activation vector to the mean activation for class 4 is plotted on the y-axis. On the x-axis are the class activations for class 3. We can see that as the non-match activation increases, the distance from this sample to the class 4 descriptor also increases. We believe this correlation justifies the use of implicit over explicit class descriptors. Furthermore, it suggests that activations are highly correlated with the similarity between a sample and a class descriptor. In Fig. 2b we show a plot of class-4 class activations versus mean distances to the class-4 descriptor.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

Figure 1: The multiclass ROC curves for the (a-c) MNIST, (d-f) SVHN, (g-i) CIFAR10, and (j-l) TinyImageNet datasets. The left, middle, and right columns illustrate the ROC curves for SoftMax, OpenMax, and MetaMax, respectively.

(a) (b)

Figure 2: An illustration of the correlation between class activations and mean distances for (a) non-match and (b) match activation vectors. Both plots correspond to the activation vectors associated with class 4 in the SVHN [4] dataset. Figures (a) and (b) show plots of the activations of class 3 and 4, respectively, on the x-axis.

Method	MNIST	SVHN	CIFAR10	TinyImageNet
SoftMax	0.644	0.682	0.547	0.512
OpenMax	0.758	0.815	0.669	0.627
MetaMax (Ours)	**0.813**	**0.846**	**0.711**	**0.683**

Table 2: The F1-scores of MetaMax against other methods using DenseNet.

Method	MNIST	SVHN	CIFAR10	TinyImageNet
SoftMax	0.637	0.684	0.542	0.510
OpenMax	0.735	0.828	0.654	0.618
MetaMax (Ours)	**0.801**	**0.854**	**0.696**	**0.669**

Table 3: The F1-scores of MetaMax against other methods using ResNet.

Method	MNIST	SVHN	CIFAR10	TinyImageNet
SoftMax	0.623	0.671	0.528	0.496
OpenMax	0.695	0.793	0.607	0.583
MetaMax (Ours)	**0.752**	**0.817**	**0.654**	**0.618**

Table 4: The F1-scores of MetaMax against other methods using VGGNet.

5. ResNet and VGGNet Results

To verify the wide applicability of various classification networks in performing open-set recognition using MetaMax, we utilize and report the F1-scores on DenseNet (Table 2), ResNet (Table 3), and VGGNet (Table 4). We can see from these results that MetaMax consistently outperforms both OpenMax and the baseline network. This demonstrates the significance of our work in that MetaMax can be applied to any classification network, hence allowing it to operate under open-set conditions. Moreover, these results provide evidence that other open-set recognition methods can achieve performance gains using MetaMax alongside any backbone network.

References

[1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[2] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4700–4708, 2017.

[3] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[4] Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y Ng. Reading digits in natural images with unsupervised feature learning. In *Proceedings of the Conference on Neural Information Processing Systems Workshops*, 2011.

[5] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.