Supplementary Material for Cross-modality Feature Fusion Network for Few-Shot 3D Point Cloud Classification

Minmin Yang*, Jiajing Chen*, Senem Velipasalar
Electrical Engineering and Computer Science Dept., Syracuse University, Syracuse, NY, USA
{myang47, jchen152, svelipas}@syr.edu

Our code is available at GitHub.

We provide the results of the additional experiments we performed on the sqMA module, and the visualization of the ablation study results below.

1. Additional Experiments on the sqMA Module

1.1. Using sqMA module in a different network

We further studied the effectiveness of the sqMA module by using it in a different network than our proposed one. More specifically, we added our proposed sqMA module to MetaOptNet [1]. The results and the improvement obtained are shown in Table 1. As can be seen, on both the ModelNet40-C and ScanObjectNN datasets, MetaOptNet with the sqMA module achieved higher accuracy in both 5-way 1-shot and 5-way 5-shot settings.

Method	ModelNet40-C		ScanObjectNN	
	5w-1s	5w-5s	5w-1s	5w-5s
MetaOptNet	69.50	82.93	52.30	67.26
MetaOptNet+sqMA	**70.00**	**84.24**	**54.09**	**68.96**

Table 1. Accuracy of the MetaOptNet before and after adding the sqMA module.

1.2. The Symmetric Structure of the sqMA Module

As shown in Fig. 1 (which is also on our main paper), the structure of the sqMA module is symmetric. More specifically, we update support features (F^s) and query features (F^q) by taking each other's features into consideration and sharing the weights of the fully-connected layers. We performed additional experiments on the ScanObjectNN dataset to study the importance of the symmetric structure of the sqMA module. In the first experiment, we used a single cross-attention module with query features as Query and support features as Key and Value. In the second experiment, we employed a cross-attention module with support features as Query and query features as Key and Value. The results in Table 2 show that sqMA is better than both of these approaches, which are denoted as Exp. 1 and Exp. 2, respectively. We believe that the symmetric structure of sqMA and the weight sharing across fully-connected layers can assist the mutual learning of query and support features.

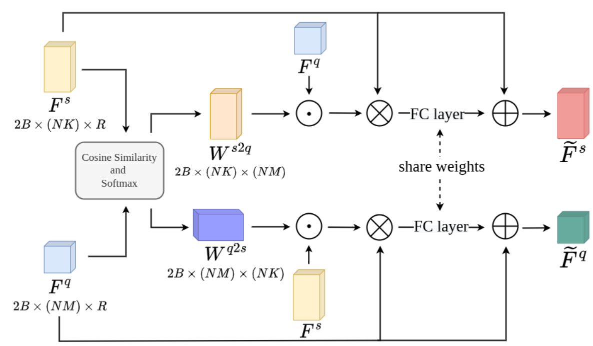


Figure 1. The proposed sqMA architecture. '$\odot$', '$\oplus$' and '$\otimes$' denote matrix multiplication, sum and concatenation, respectively.

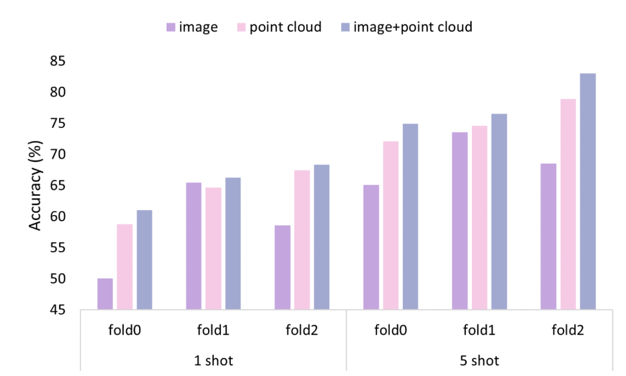
	Exp. 1	Exp. 2	sqMA
5w-1s	62.60	63.37	**65.26**
5w-5s	76.49	76.65	**78.14**

Table 2. Accuracy of three different attention modules on the ScanObjectNN dataset.

2. Visualization of Ablation Study Results

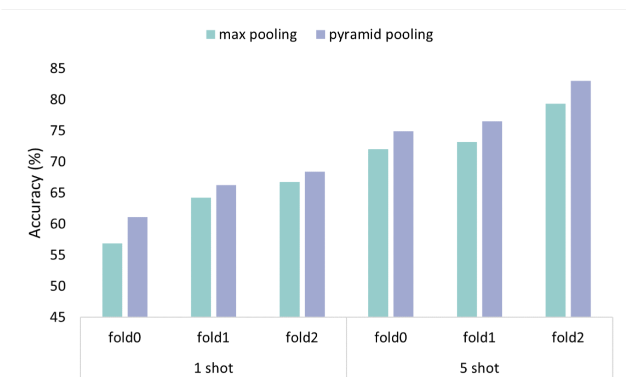
We conducted ablation study experiments to verify the effectiveness of (i) combining two data modalities, (ii) the sqMA module and (iii) the pyramid pooling operation in our main paper in Section 4.5. All the experiments are conducted on the ScanObjectNN dataset. Here, we further visualize the accuracy of different models with bar charts shown in Fig. 2. Figure 2 (a) shows that using depth images and point data together provides consistently better perfor-

*Both authors contributed equally. This work was supported in part by the National Science Foundation under Grant 1816732.



(a) Visualization of accuracy of using different input modalities.

(b) Visualization of accuracy of models with sqMA and without sqMA module.



(c) Visualization of accuracy of models using max pooling and pyramid pooling.

Figure 2. Ablation studies. Visualization of the accuracy for different model structures on the ScanObjectNN dataset.

mance than using single data modality in 1-shot and 5-shot settings. Figures 2 (b) and (c) show the effectiveness of the sqMA module and the pyramid pooling, respectively. More specifically, Figure 2 (b) shows the accuracy of our model with and without the sqMA module. Figure 2 (c) shows that pyramid pooling provides accuracy improvement in both 1-shot setting and 5-shot setting.

References

[1] Kwonjoon Lee, Subhransu Maji, Avinash Ravichandran, and Stefano Soatto. Meta-learning with differentiable convex optimization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10657–10665, 2019. 1