Non-Probabilistic Cosine Similarity Loss for Few-Shot Image Classification : Supplementary Material

Joonhyuk Kim[1]
juhkim@rit.kaist.ac.kr
Inug Yoon[1]
iuyoon@rit.kaist.ac.kr
Gyeong-Moon Park[2]
gmpark@etri.re.kr
Jong-Hwan Kim[1]
johkim@rit.kaist.ac.kr

[1] Korea Advanced Institute of Science and Technology (KAIST)
Daejeon, Republic of Korea
[2] Electronics and Telecommunications Research Institute (ETRI)
Daejeon, Republic of Korea

1 Architecture Details

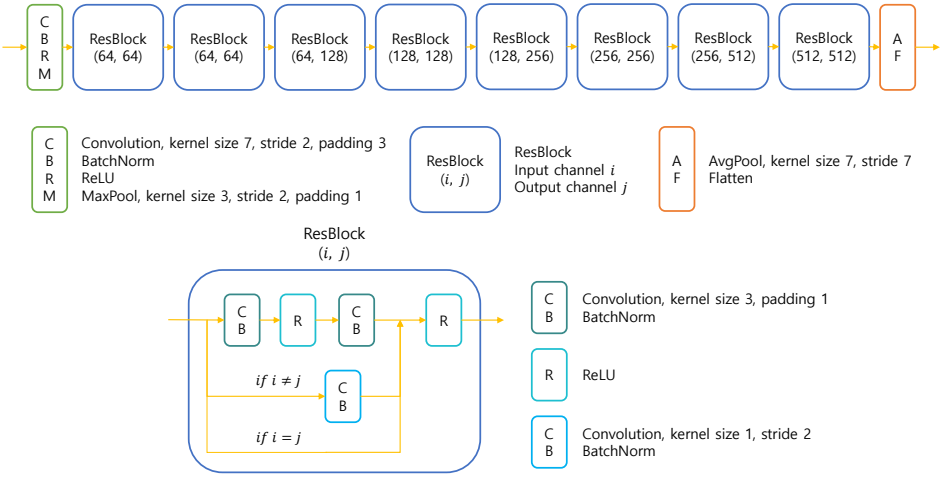

Figure 1: Architecture of ResNet-18

Figure 1 presents the detailed architecture of ResNet-18 structure. Specifically, a resized image with a size of $224 \times 224 \times 3$, is encoded to a 512-dimensional feature vector. It consists of seventeen convolution modules with 3×3 filters, and three 1×1 convolution modules.

2 Additional Experimental Results

2.1 Comparison with Other Losses for More-way and -shot Cases

Loss function	Architecture	5-way		
		1-shot (%)	5-shot (%)	10-shot (%)
CC	$\text{conv}(64)_{\times 2}$-$(128)_{\times 2}$	**52.78 ± 0.77**	68.76 ± 0.68	72.97 ± 0.57
AF		**52.62 ± 0.77**	68.82 ± 0.66	72.27 ± 0.58
CF		**52.32 ± 0.81**	69.15 ± 0.66	73.69 ± 0.58
NPC		**52.72 ± 0.86**	**70.79 ± 0.70**	**74.68 ± 0.59**
CC	ResNet-18	50.02 ± 0.84	66.60 ± 0.68	71.36 ± 0.64
AF		47.42 ± 0.79	63.64 ± 0.69	68.78 ± 0.66
CF		50.45 ± 0.83	66.96 ± 0.69	71.61 ± 0.63
NPC		**57.51 ± 0.85**	**75.37 ± 0.65**	**79.24 ± 0.61**

Loss function	Architecture	10-way		
		1-shot (%)	5-shot (%)	10-shot (%)
CC	$\text{conv}(64)_{\times 2}$-$(128)_{\times 2}$	**36.65 ± 0.47**	53.68 ± 0.43	58.73 ± 0.44
AF		35.90 ± 0.47	52.70 ± 0.44	57.86 ± 0.43
CF		**36.86 ± 0.49**	54.20 ± 0.44	59.49 ± 0.41
NPC		**36.66 ± 0.49**	**55.71 ± 0.45**	**61.30 ± 0.42**
CC	ResNet-18	34.86 ± 0.49	51.07 ± 0.44	56.53 ± 0.43
AF		32.12 ± 0.47	47.73 ± 0.41	53.44 ± 0.41
CF		34.89 ± 0.48	51.08 ± 0.44	56.48 ± 0.42
NPC		**40.93 ± 0.53**	**61.79 ± 0.44**	**67.25 ± 0.41**

Loss function	Architecture	20-way		
		1-shot (%)	5-shot (%)	10-shot (%)
CC	$\text{conv}(64)_{\times 2}$-$(128)_{\times 2}$	**24.43 ± 0.27**	39.38 ± 0.24	45.19 ± 0.21
AF		23.70 ± 0.25	38.52 ± 0.23	44.39± 0.23
CF		**24.71 ± 0.27**	40.18 ± 0.25	46.01 ± 0.24
NPC		**24.78 ± 0.26**	**42.10 ± 0.25**	**48.42 ± 0.23**
CC	ResNet-18	23.30 ± 0.27	37.03 ± 0.24	41.99 ± 0.22
AF		21.21 ± 0.27	33.88 ± 0.25	38.78 ± 0.23
CF		23.23 ± 0.27	36.64 ± 0.24	41.65 ± 0.22
NPC		**28.51 ± 0.31**	**48.30 ± 0.25**	**53.79 ± 0.24**

Table 1: The results of few-shot image classification on the test set of Mini-ImageNet dataset without fine-tuning. The results are average accuracies over 600 test episodes with the 95% confidence intervals.

Here, we report the additional experimental results for more-way and -shot cases. We conducted the experiments under the same conditions, but the loss functions for training. As can be seen from Table 1, when the feature extractor architecture deepened, the tendency to degrade the performance of the models trained with probability-based loss functions is maintained even under different-way and -shot conditions.

2.2 Comparison with Other Losses after Fine-tuning

We also did fine-tuning for the models trained with other loss functions, and compare with the one trained with NPC loss. Even after the fine-tuning stage, the model trained with

Loss function	Architecture	5-way		20-way	
		1-shot (%)	5-shot (%)	1-shot (%)	5-shot (%)
CC	ResNet-18	53.64 ± 0.83	73.76 ± 0.69	26.02 ± 0.30	44.95 ± 0.27
AF		51.32 ± 0.85	70.75 ± 0.71	23.49 ± 0.29	41.05 ± 0.26
CF		53.65 ± 0.87	74.08 ± 0.69	25.91 ± 0.30	44.86 ± 0.27
NPC		$\mathbf{60.98 \pm 0.87}$	$\mathbf{80.17 \pm 0.65}$	$\mathbf{32.37 \pm 0.33}$	$\mathbf{56.14 \pm 0.26}$

Table 2: The results of few-shot image classification on the test set of Mini-ImageNet dataset after fine-tuning. The results are average accuracies over 600 test episodes with the 95% confidence intervals.

NPC overwhelms other models not only for the 5-way case but also for the 20-way case (See Table 2). When the model is trained with NPC loss, the feature extractor is trained to draw the shared features across the novel categories; thus, it is easy to train it further to draw detailed features. However, probabilistic-based losses make the model overfitting on the base categories; thus, it is hard to bring detailed features of the novel categories within the same epochs.