Mixture-based Feature Space Learning for Few-shot Image Classification
Supplementary Material

Arman Afrasiyabi[⋆•], Jean-François Lalonde[⋆], Christian Gagné[⋆†•]
[⋆]Université Laval, [†]Canada CIFAR AI Chair, [•]Mila
https://lvsn.github.io/MixtFSL/

In this supplementary material, the following items are provided:

1. Ablation on the number of components N^k in the mixture model $\boldsymbol{\mathcal{P}}$ (sec. 1)
2. Dynamic of the training (sec. 2);
3. More ways ablation (sec. 3);
4. Ablation of the margin m (sec. 4);
5. Ablation of the temperature τ (sec. 5);
6. Visualization: from MixtFSL to MixtFSL-Alignment (sec. 6);

1. Ablation on the number of components N^k in the mixture model $\mathcal{P}$

Although our proposed MixtFSL automatically infers the number of per-class mixture components from data, we also ablate the initial size of mixture model N^k for each class to evaluate whether it has an impact on the final results. Table 1 presents 1- and 5-shot classification results on miniImageNet using ResNet-12 and ResNet-18 by initializing N^k to 5, 10, 15, and 20 components per class.

Initializing $N^k = 5$ results in lower classification accuracy compared to the higher N^k. We think this is possible due to the insufficient capacity of small mixture model $\mathcal{P}$ size. However, as long as N^k is sufficiently large (10, 15, 20), our approach is robust to this parameter and results do not change significantly as a function of N^k. Note that N^k cannot be set to an arbitrary high number due to memory limitations.

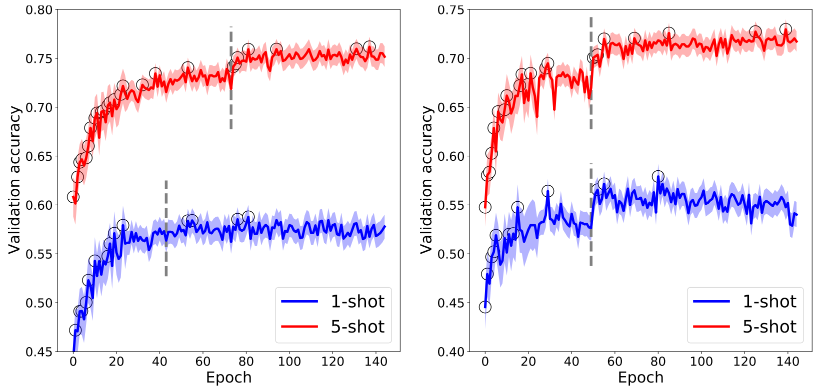
Table 1. Classification results on mini-ImageNet using ResNet-12 and ResNet-18 backbones as a function of the initial value for the number of components per class N^k. $\pm$ denotes the 95% confidence intervals over 300 episodes.

N^k	1-shot	5-shot
5	62.29 $\pm$1.08	78.85 $\pm$0.61
10	64.01 $\pm$0.79	81.87 $\pm$0.49
15	63.98 $\pm$0.79	82.04 $\pm$0.49
20	63.91 $\pm$0.80	82.05 $\pm$0.49

(a) ResNet-12

N^k	1-shot	5-shot
5	58.57 $\pm$1.09	76.44 $\pm$0.61
10	60.15 $\pm$0.80	77.71 $\pm$0.61
15	60.11 $\pm$0.73	77.76 $\pm$0.58
20	58.99 $\pm$0.81	77.77 $\pm$0.58

(b) ResNet-18



(a) ResNet-12 (b) ResNet-18

Figure 1. Validation accuracy of the first 150 epochs using ResNet-12 and ResNet-18 on miniImageNet. 1- and 5-shot scenarios are plotted using blue and red colors with their confidence intervals over 300 testing episodes of the validation set, respectively. The dashed vertical line is starting point of progressive following stage. The circles are the points when we update the best model.

2. Dynamic of the training

Fig. 1 evaluates the necessity of the two training stages (sec. 4 from the main paper) by showing the (episodic) validation accuracy during 150 epochs. The vertical dashed line indicates the transition between training stages. In most cases, the progressive following stage results in a validation accuracy gain.

3. More ways ablation

Table 2 presents more-way 5-shot comparison of our MixtFSL on miniImageNet using ResNet-18 and ResNet-12. Our MixtFSL gains 1.14% and 1.23% over the Pos-Margin [1] in 5-way and 20-way, respectively. Besides, MixtFSL gains 0.78% over Baseline++ [2] in 10-way.

We could not find "more-ways" results with the ResNet-12 backbone in the literature, but we provide our results here for potential future literature comparisons.

Table 2. N-way 5-shot classification results on mini-ImageNet using ResNet-18 and ResNet-12 backbones. $\pm$ denotes the 95% confidence intervals over 600 episodes. The best results prior this work is highlighted in blue, and the best results are presented in boldfaced.

Method	Backbone	**5-way**	**10-way**	**20-way**
MatchingNet‡ [7]	RN-18	68.88 ±0.69	52.27 ±0.46	36.78 ±0.25
ProtoNet‡ [5]	RN-18	73.68 ±0.65	59.22 ±0.44	44.96 ±0.26
RelationNet‡ [6]	RN-18	69.83 ±0.68	53.88 ±0.48	39.17 ±0.25
Baseline [2]	RN-18	74.27 ±0.63	55.00 ±0.46	42.03 ±0.25
Baseline++ [2]	RN-18	75.68 ±0.63	63.40 ±0.44	50.85 ±0.25
Pos-Margin [1]	RN-18	76.62 ±0.58	62.95 ±0.83	51.92 ±1.02
MixtFSL (ours)	RN-18	**77.76** ±0.58	**64.18** ±0.76	**53.15** ±0.71
MixtFSL (ours)	RN-12	82.04 ±0.49	68.26 ±0.71	55.41 ±0.71

‡ implementation from [2]

Table 3. Margin evaluation using miniImageNet in 5-way classification. Bold/blue is best/second best, and $\pm$ indicates the 95% confidence intervals over 600 episodes.

Method	**Backbone**	**1-shot**	**5-shot**
Neg-Margin* [3]	Conv4	51.81 $\pm$0.81	69.24 $\pm$0.59
ArcMax* [1]	Conv4	51.95 $\pm$0.80	69.05 $\pm$ 0.58
MixtFSL-Neg-Margin	Conv4	52.76 $\pm$0.67	**70.67** $\pm$0.57
MixtFSL-Pos-Margin	Conv4	**52.82** $\pm$0.63	70.30 $\pm$0.59
Neg-Margin* [3]	RN-12	61.90 $\pm$0.74	78.86 $\pm$0.53
ArcMax* [1]	RN-12	61.86 $\pm$0.71	78.55 $\pm$0.55
MixtFSL-Neg-Margin	RN-12	**63.98** $\pm$0.79	**82.04** $\pm$0.49
MixtFSL-Pos-Margin	RN-12	63.57 $\pm$0.00	81.70 $\pm$0.49
Neg-Margin* [3]	RN-18	59.15 $\pm$0.81	**78.41** $\pm$0.54
ArcMax* [1]	RN-18	58.42 $\pm$0.84	77.72 $\pm$0.51
MixtFSL-Neg-Margin	RN-18	**60.11** $\pm$0.73	77.76 $\pm$0.58
MixtFSL-Pos-Margin	RN-18	59.71 $\pm$0.76	77.59 $\pm$0.58
Neg-Margin* [3]	WRN	62.27 $\pm$0.90	80.52 $\pm$0.49
ArcMax* [1]	WRN	62.68 $\pm$0.76	80.54 $\pm$0.50
MixtFSL-Neg-Margin	WRN	63.18 $\pm$1.02	81.66 $\pm$0.60
MixtFSL-Pos-Margin	WRN	**64.31** $\pm$0.79	**81.63** $\pm$0.56

* our implementation

4. Ablation of the margin

As table 3 shows, a negative margin provides slightly better results than using a positive one, thus replicating the findings from Liu *et al*. [3], albeit with a more modest improvement than reported in their paper. We theorize that the differences between our results (in table 3) and theirs are due to slight differences in training setup (e.g., learning rate scheduling, same optimizer for base and novel classes). Nevertheless, the impact of the margin on our proposed MixtFSL approach is similar. We also note that in all cases except 5-shot on ResNet-18, our proposed MixtFSL yields significant improvements. Notably, MixtFSL provides classification improvements of 2.08% and 3.18% in 1-shot and 5-shot using ResNet-12.

The margin m in eq.1 (sec. 4.1) is ablated in Table 4 using the validation set of the miniImagNet dataset using ResNet-12 and ResNet-18. We experiment with both $m = 0.01$ to match Afrasiyabi *et al*. [1], and $m = -0.02$ to match Bin *et al*. [3].

Table 4. Margin m ablation on the miniImageNet using ResNet-12 and ResNet-18 backbones.

	ResNet-12		ResNet-18	
m	1-shot	5-shot	1-shot	5-shot
-0.02	61.85	80.38	60.57	79.04
+0.01	60.97	77.43	60.27	78.12

5. Ablation of the temperature τ

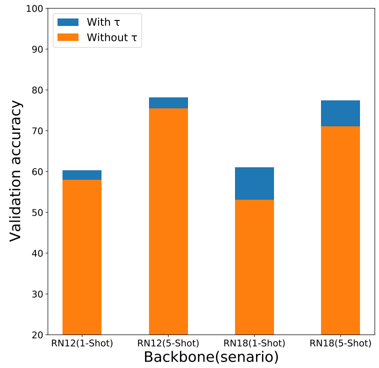
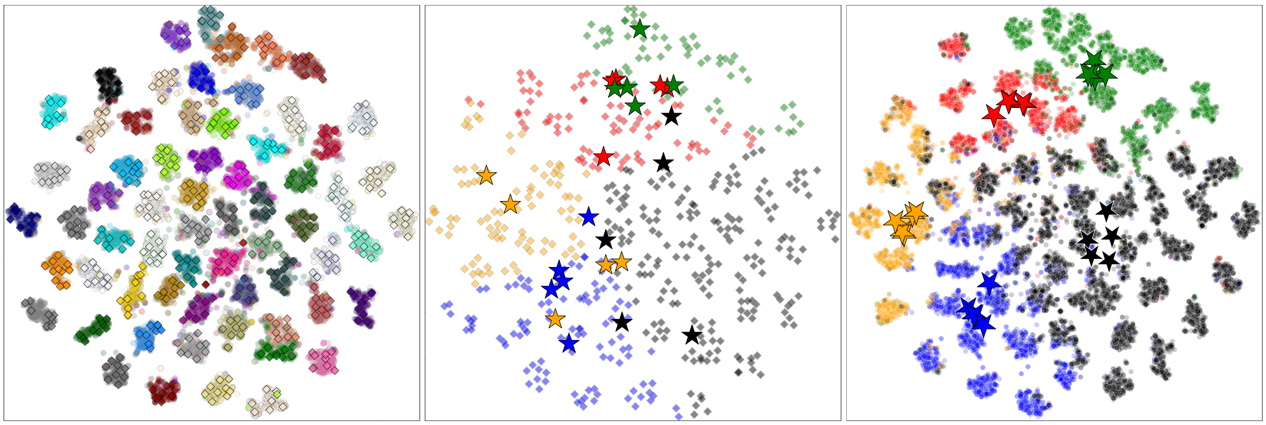


Figure 2. Effect of temperature τ on MixtFSL using ResNet-12 and -18 in 1- and 5-shot scenarios in miniImageNet's validation set. The orange bars are the classification results without temperature variable ($\tau = 1$), and the blue colored bars are the amount of classification gain by training the backbone with temperature variable ($\tau = 0.05$).

We ablate the effect of having a temperature variable τ in the initial training stage using the validation set. As fig. 2 presents, the validation set accuracy increases with the use of τ variable across the RN-12 and RN-18. Here, "without τ" corresponds to setting $\tau = 1$, and "with τ" to $\tau = 0.05$ (found on the validation set).

6. Visualization: from MixtFSL to MixtFSL-Alignment

Fig. 3 summarizes the visualization of embedding space from our mixture-based feature space learning (MixtFSL) to its centroid alignment extension (sec. 6.1 from the main paper). Fig. 3-(a) is a visualization of 200 base examples per class (circles) and the learned class mixture components (diamonds) after the progressive following training stage. Fig. 3-(b) presents the t-SNE visualization of novel class examples (stars) and related base detection (diamonds of the same color) using our proposed MixtFSL. Fig. 3-(c) presents the visualization of fine-tuning the centroid alignment of [1]. Here, the novel examples align to the center of their related bases.



(a) after progressive following (b) related-modes detection (c) after alignment

Figure 3. t-SNE [4] applied to the ResNet-12 base feature embedding. (a) learned base categories feature embedding (circles) and mixture components (diamonds) after the progressive following stages. (b) using 5-way (coded by color) novel example shown by stars to detect their related base classes with the learned mixture components shown by diamonds. (b) aligning the novel examples to the center of their related base classes without forgetting the base classes. Points are color-coded by related base and novel examples.

References

[1] Arman Afrasiyabi, Jean-François Lalonde, and Christian Gagné. Associative alignment for few-shot image classification. In *European Conference on Computer Vision*, 2020.

[2] Wei-Yu Chen, Yen-Cheng Liu, Zsolt Kira, Yu-Chiang Frank Wang, and Jia-Bin Huang. A closer look at few-shot classification. *arXiv preprint arXiv:1904.04232*, 2019.

[3] Bin Liu, Yue Cao, Yutong Lin, Qi Li, Zheng Zhang, Mingsheng Long, and Han Hu. Negative margin matters: Understanding margin in few-shot classification. In *European Conference on Computer Vision*, 2020.

[4] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 2008.

[5] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. In *Neural Information Processing Systems*, 2017.

[6] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip H.S. Torr, and Timothy M. Hospedales. Learning to compare: Relation network for few-shot learning. In *International Conference on Computer Vision and Pattern Recognition*, 2018.

[7] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. Matching networks for one shot learning. In *Neural Information Processing Systems*, 2016.