Associative Alignment for Few-shot Image Classification Supplementary Material

Arman Afrasiyabi*, Jean-François Lalonde*, Christian Gagné*†

*Université Laval, †Canada CIFAR AI Chair, Mila
arman.afrasiyabi.1@ulaval.ca
{jflalonde,christian.gagne}@gel.ulaval.ca
https://lvsn.github.io/associative-alignment/

In this supplementary material, the following items are provided:

1. Validation error plot (sec. 1);
2. Ablation study on B (sec. 2);
3. Visualization (sec. 3);
4. More ways (sec. 4);
5. Comparison to no alignment (sec. 5);
6. Sensitivity to wrongly-related classes (sec. 6)
7. Ablation on the margin (sec. 7)

1 Validation error plot (refers to sec. 5.2)

Fig. 1 plots validation error after fine-tuning vs. the number of pre-training epochs. The "cosmax" function is used, with the entire network pre-trained on $\mathcal{X}^b$, and only the classification weights $\mathbf{W}$ fine-tuned on $\mathcal{X}^n$, as in [1]. The decrease in accuracy over the epochs (after 150 epoch for 1-shot) shows that pre-training should not be conducted for a fixed number of epochs.

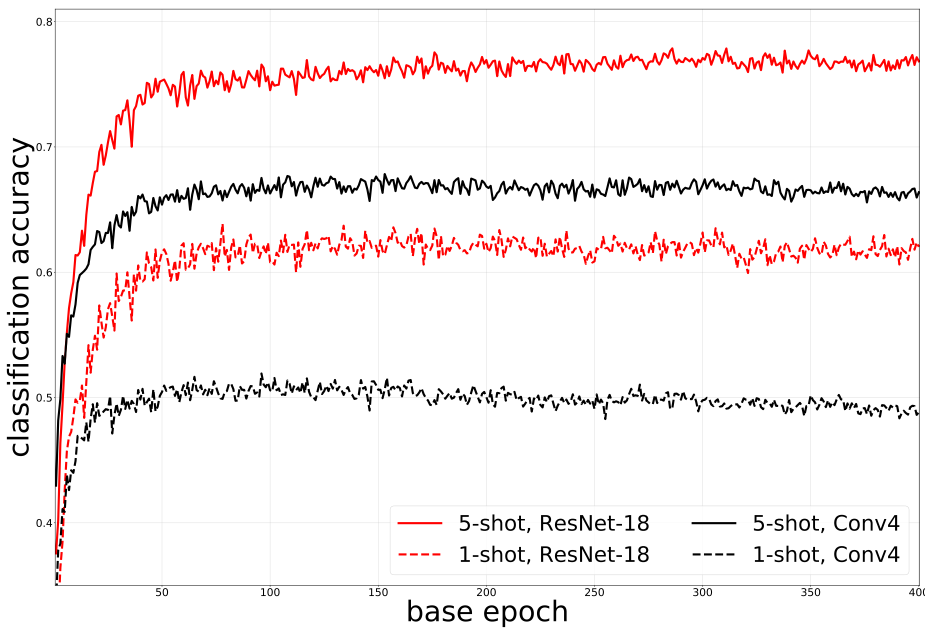

Fig. 1: Validation error after fine-tuning as a function of the number of pre-training base epochs on *mini*-ImageNet with the cosmax loss. Pre-training for a fixed number of iterations (here 400 as in [1]) may lead to overfitting the feature extraction on the base set. Each curve represents the average of 50 episodes.

2 Ablation study on B (refers to sec. 6.1)

Table 1 presents an ablation study for B, the number of related base classes selected for each novel class. We perform the study on few-shot image classification on the *mini*-ImageNet dataset using ResNet-18 backbone. Overall, better results are obtained with a larger value of B, except for the adversarial alignment method in the 5-shot scenario.

Table 1: Effect of three different number of related bases B on few-shot classification results on *mini*-ImageNet using ResNet-18 backbones. ± denotes the 95% confidence intervals over 600 episodes.

B	1-shot	5-shot
arcm.	58.07 ± 0.82	76.62 ± 0.58
1	55.76 ± 1.20	**79.34** ± 0.69
5	58.20 ± 1.14	78.65 ± 0.94
10	**58.84** ± 0.77	77.92 ± 0.82
12	58.79 ± 0.81	77.56 ± 0.85

(a) Adversarial alignment

B	1-shot	5-shot
arcm.	58.07 ± 0.82	76.62 ± 0.58
1	58.04 ± 0.98	77.54 ± 0.73
5	58.97 ± 1.06	79.14 ± 0.91
10	**59.88** ± 0.67	**80.23** ± 0.73
12	60.04 ± 0.77	80.18 ± 0.79

(b) Centroid alignment

3 Visualization of the alignment methods (refers to sec. 6.2)

Fig. 2 presents a 2D visualization of our adversarial and centroid alignment methods using t-SNE [2] on *mini*ImageNet (see sec. 6.1 for the dataset description) dataset in 5-shot 5-way scenario. While both methods achieve similar results with $B = 1$, the centroid method results yields more discriminative class separation compared to the adversarial method with $B = 10$. The multi-modalities

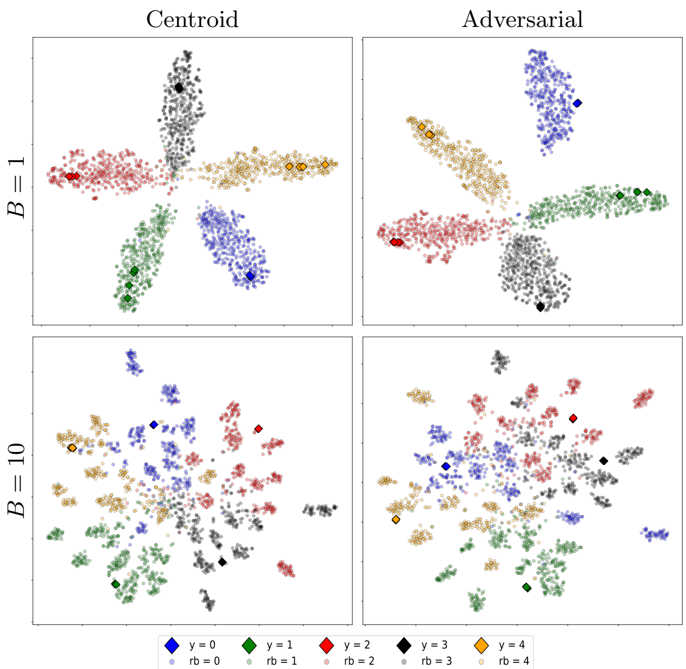

Fig. 2: Aligning novel and related base classes. Columns present centroid and adversarial distribution matching while the rows compare picking $B = 1$ and $B = 10$ related base classes for each novel class. We use t-SNE [2] to visualize the 512-dimensional feature space of ResNet-18 in 2D. Results are for 5-shot in a 5-way setting.

of certain base categories look inevitable and might indeed degrade the generalization performance compared to the single-mode case assumed by our centroid alignment strategy. We compute the percentage of classes for which our centroid alignment approach: 1) improves, 2) does not change, or 3) deteriorates performance compared to our strong baseline (using a fixed threshold of 1% on classification accuracy). In the 5-shot scenario using ResNet-18 on *mini*-ImageNet, our centroid alignment approach results in improvements for 69.8% of the classes (with 13.9% not changing, and 16.3% deteriorates).

4 More-way (refers to sec. 6.2)

We experiment with N-way, 5-shot experiment (for N = 5, 10, 20) to examine the effect of associative alignment on more-way using *mini*-ImageNet. As Table 2 presents, our associative alignment gains on the compared meta-learning and standard transfer learning methods. Specifically, we outperform the best of the compared method by 6.67%, 4.47%, 3.82% in 5-, 10-, and 20-way respectively. Note that we used 10, 5, 3 number of related base classes (B) 5-way, 10-way and 20-way respectively which corresponds to 60 classes out of all 64 base categories in *mini*-ImageNet.

Table 2: N-way 5-shot classification results on *mini*-ImageNet using ResNet-18 backbone. ± 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	**5-way**	**10-way**	**20-way**
meta-l.	MatchingNet‡ [5]	68.88 ± 0.69	52.27 ± 0.46	36.78 ± 0.25
	ProtoNet‡ [3]	73.68 ± 0.65	59.22 ± 0.44	44.96 ± 0.26
	RelationNet‡ [4]	69.83 ± 0.68	53.88 ± 0.48	39.17 ± 0.25
transfer-l.	softmax [1]	74.27 ± 0.63	55.00 ± 0.46	42.03 ± 0.25
	cosmax [1]	75.68 ± 0.63	63.40 ± 0.44	50.85 ± 0.25
	our baseline (sec. 5.1)	76.62 ± 0.58	62.95 ± 0.83	51.92 ± 1.02
	B	10	5	3
align.	adversarial	77.92 ± 0.82	64.87 ± 0.96	52.46 ± 0.99
	centroid	**80.35** ± 0.73	**68.17** ± 0.79	**54.67** ± 1.02

‡ implementation from [1]

5 Comparison to no alignment

Table 3 illustrates the effect of training the network using both novel and their related classes, but without the alignment losses. The results are shown in the "no alignment" row in table 3 below. Excluding the alignment loss slightly improves the accuracy compared to baseline by 0.82% and 0.24% in 1-shot and 5-shot using Conv4, respectively; however, it falls below the baseline by -2.13% and -2.34% in 1-shot and 5-shot using ResNet-18, respectively. In addition, except for the adversarial alignment in 1-shot using Conv4, both of the alignment strategies result in accuracy improvement in all of the scenarios, which shows the necessity of an alignment strategy.

Table 3: Evaluating the necessity of alignment loss. Few-shot classification results on *mini*-ImageNet using both Conv4 and ResNet-18 backbones. ± denotes the 95% confidence intervals over 600 episodes.

		Conv4		ResNet-18	
		1-shot	5-shot	1-shot	5-shot
	baseline	51.90 ± 0.79	69.07 ± 0.62	58.07 ± 0.82	76.62 ± 0.58
	no alignment	52.72 ± 0.79	69.31 ± 0.69	55.94 ± 0.88	74.28 ± 0.83
alignment	adversarial	52.13 ± 0.99	70.78 ± 0.60	58.84 ± 0.77	77.92 ± 0.82
	centroid	**53.14** ± 1.06	**71.45** ± 0.72	**59.88** ± 0.67	**80.35** ± 0.73

6 Sensitivity to wrongly-related classes

We also evaluate the sensitivity of the algorithm to the percentage of wrongly-related classes by replacing an increasing number of related base classes (selected by our algorithm) with random base classes instead (while keeping the total number of related base classes fixed to B=10). Results with the centroid alignment on *mini*-ImageNet and ResNet-18 are shown in table 4.

Small changes to the selected classes have little impact on performance showing the stability of our approach. Replacing 5 randomly-selected base classes with random ones still results in improved performance in the 5-shot scenario. Even if heuristic, our related base class selection algorithm results in much improved performance compared to the 0/10 case.

Table 4: Evaluating the sensitivity to wrongly-related classes. Few-shot classification results on *mini*-ImageNet using ResNet-18 backbone. ± denotes the 95% confidence intervals over 600 episodes.

selected / random	1-shot	5-shot
[paper] 10 / 0	**59.98** ± 0.7	**80.35** ± 0.7
9 / 1	59.74 ± 0.7	80.07 ± 0.9
8 / 2	59.77 ± 0.6	78.69 ± 0.8
5 / 5	58.36 ± 0.7	77.35 ± 0.8
0 / 10	56.72 ± 1.2	76.19 ± 0.8
[paper] baseline	58.07 ± 0.8	76.62 ± 0.6

7 Ablation on the margin m

We used episodic cross-validation to find the margin (m). In our experiments, we found that m needs to be adjusted according to the architectures rather than the datasets, which is likely due to its relation to the network learning capacity. An ablation for m on the *mini*-ImageNet validation set for the 5-way scenario is presented in table 5.

Table 5: ablation for margin (m) on the *mini*-ImageNet using ResNet-18 and Conv4 backbones. ± denotes the 95% confidence intervals over 600 episodes.

	Conv4		ResNet-18	
m	1-shot	5-shot	1-shot	5-shot
0.9	48.6	66.9	58.1	77.0
0.1	52.3	68.9	58.3	76.6
0.01	52.0	67.5	60.0	77.6

References

1. Chen, W.Y., Liu, Y.C., Kira, Z., Wang, Y.C.F., Huang, J.B.: A closer look at few-shot classification. arXiv preprint arXiv:1904.04232 (2019)
2. Maaten, L.v.d., Hinton, G.: Visualizing data using t-sne. Journal of Machine Learning Research (2008)
3. Snell, J., Swersky, K., Zemel, R.: Prototypical networks for few-shot learning. In: Advances in Neural Information Processing Systems (2017)
4. Sung, F., Yang, Y., Zhang, L., Xiang, T., Torr, P.H., Hospedales, T.M.: Learning to compare: Relation network for few-shot learning. In: The Conference on Computer Vision and Pattern Recognition (2018)
5. Vinyals, O., Blundell, C., Lillicrap, T., Wierstra, D., et al.: Matching networks for one shot learning. In: Advances in Neural Information Processing Systems (2016)