Matching Feature Sets for Few-Shot Image Classification
Supplementary Materials

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

In this supplementary material, the following items are provided:

1. Dataset and backbone specifications(sec. 1);
2. Ablation with more ways and cross-domain results from miniImageNet $\mapsto$ CUB (sec. 2);
3. Visualizing mappers saliency (sec. 3);
4. Class structure in cluster (sec. 4);
5. Hausdorff distance ablation (sec. 5);

1. Dataset and backbone specifications

Table tab. 1 present the detailed specification of dataset, and and Table tab. 2 specifies the number of overall parameters in our set feature SetFeat extractor compared to the popular backbones used in the few-shot image classification literature.

Table 1. Specifications of miniImageNet, tieredImageNet and CUB.

Dataset	Number of examples	Source	Splits(train/val/test)	Split Reference
MiniImageNet	60,000	ImageNet† [6]	64/16/20	Vinyals *et al.* [9]
TieredImageNet	779,165	ImageNet† [6]	351/97/160	Ren *et al.* [5]
CUB	11,788	CUB-200-2011* [10]	100/50/50	Chen *et al.* [2]

† https://www.image-net.org/ * http://www.vision.caltech.edu/visipedia/CUB-200-2011.html

Table 2. Number of parameters for various backbones, compared with our SetFeat implementations (in blue). Blocks column illustrates the number of parameters in all the convolution layers. Mappers column shows the number of parameters in 10 employed mappers in SetFeat.

Backbone	Blocks	Mappers	Total
Conv4-64	0.113 M	–	0.113 M
SetFeat4-64	0.113 M	0.124 M	0.238 M
Conv4-512	1.591 M	–	1.591 M
SetFeat4-512	0.587 M	0.996 M	1.583 M
ResNet18	11.511 M	–	11.511 M
SetFeat12*	6.977 M	4.489 M	11.466 M
ResNet12	12.424 M	–	12.424 M
SetFeat12	7.447 M	4.902 M	12.349 M

2. Ablation with more ways and cross-domain results from miniImageNet $\mapsto$ CUB

Tab. 3 shows 5-way, 10-way, and 20-way comparisonds of SetFeat12* and SetFeat12 with ResNet18 and ResNet12, respectively. As illustrated in creftab:backboneparameters and mentioned in sec. 5.3 of the main paper, SetFeat12* (11.466M parameters) is the counterpart of ResNet18 (11.511M parameters).

Tab. 3 shows that SetFeat with the sum-min metric (eq. (5) from the main paper) achieves state-of-the-art results in 5-shot for all of 5-, 10- and 20-way classification. Notably, SetFeat12* and SetFeat12 gain 6.18% and 2.84% over MixtFSL [1] in 5-way, respectively. Additionally, last column of tab. 3 shows cross domain adaptation, where we pre-train our model on miniImageNet and test on the CUB dataset. Here, our SetFeat12* obtains the second best and is 0.92% below MixtFSL [1].

Table 3. N-way 5-shot classification results on miniImageNet using ResNet and SetFeat. $\pm$ denotes the 95% confidence intervals over 600 episodes. The best results prior to this work is highlighted in red, and the best results are presented in boldface.

Method	Backbone	miniImageNet 5-way	miniImageNet 10-way	miniImageNet 20-way	miniImageNet→CUB 5-way
MatchingNet‡ [9]	ResNet18	68.88 ±0.69	52.27 ±0.46	36.78 ±0.25	–
Neg-Margin‡ [4]	ResNet18	–	–	–	67.03 ±0.80
ProtoNet‡ [7]	ResNet18	73.68 ±0.65	59.22 ±0.44	44.96 ±0.26	62.02 ±0.70
RelationNet‡ [8]	ResNet18	69.83 ±0.68	53.88 ±0.48	39.17 ±0.25	57.71 ±0.70
Baseline [2]	ResNet18	74.27 ±0.63	55.00 ±0.46	42.03 ±0.25	65.57 ±0.25
Baseline++ [2]	ResNet18	75.68 ±0.63	63.40 ±0.44	50.85 ±0.25	64.38 ±0.90
Pos-Margin [1]	ResNet18	76.62 ±0.58	62.95 ±0.83	51.92 ±1.02	64.93 ±1.00
MixtFSL [1]	ResNet18	77.76 ±0.58	64.18 ±0.76	53.15 ±0.71	**68.77** ±0.90
Sum-min (ours)	SetFeat12*	**81.22**±0.45	**70.36**±0.46	**57.36**±0.36	67.85±0.70
MixtFSL [1]	ResNet12	82.04 ±0.49	68.26 ±0.71	55.41 ±0.71	–
Sum-min (ours)	SetFeat12	**82.71**±0.46	**71.10**±0.46	**57.97**±0.36	–

‡ implementation from [2]

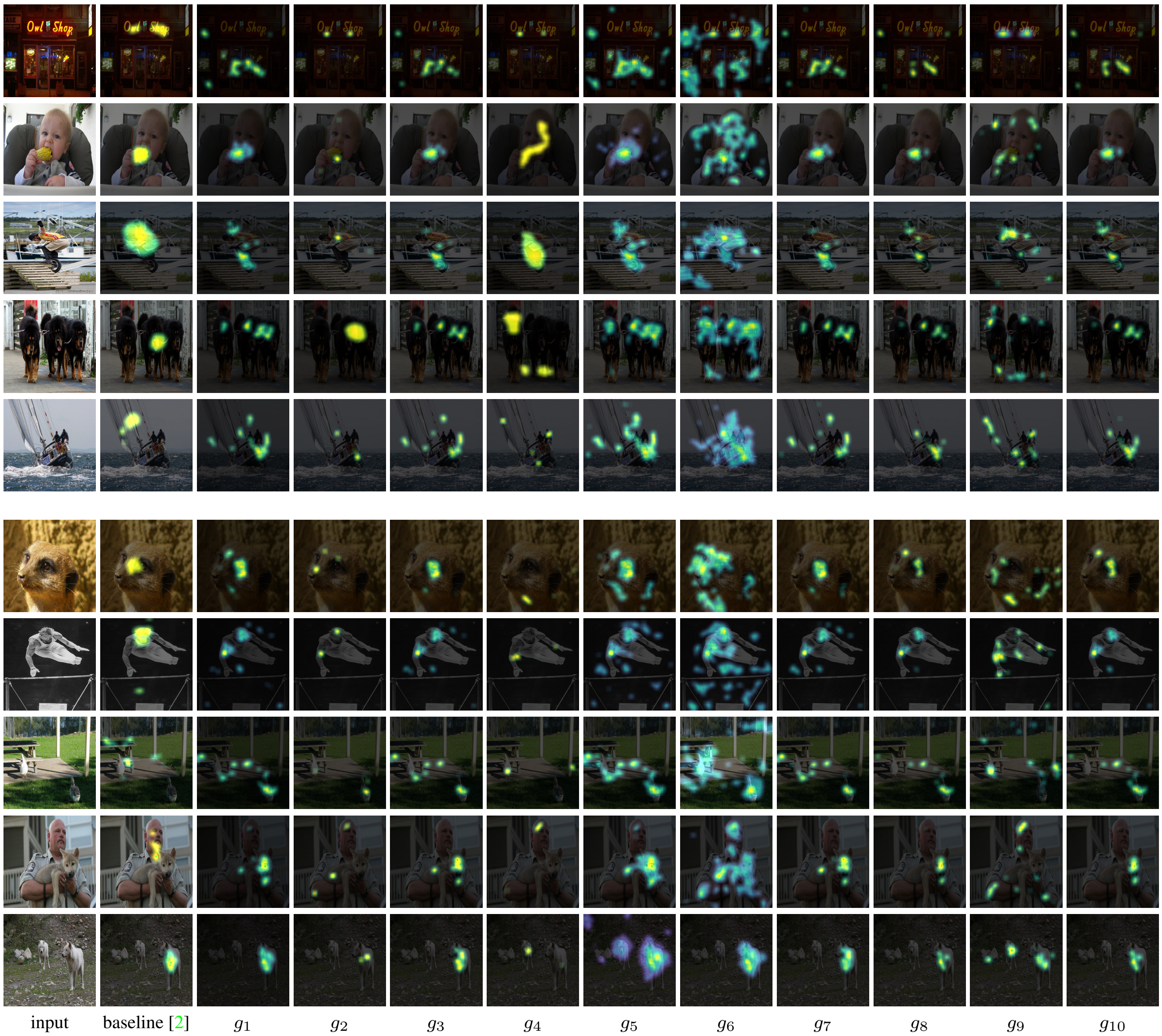

Figure 1. Gradient saliency maps after training SetFeat12 on miniImageNet. From left: input image, baseline [2] trained with ResNet12, and 10 different mappers from our SetFeat12 (g_i is the i-th mapper). The first five rows show examples from the training dataset, and the last five are from the validation set of miniImageNet.

3. Visualizing mappers saliency

Figs. 1 and 2 compare the gradient saliency maps of SetFeat12 and SetFeat4-64 using our sum-min metric with ResNet12 and Conv4-64 using "baseline" from [3]. Here SetFeat4-64 uses an FC-layer to compute mappers, while SetFeat12 uses a convolutional layer to do so. As shown in the figures, different mappers focus on different regions of the input image.

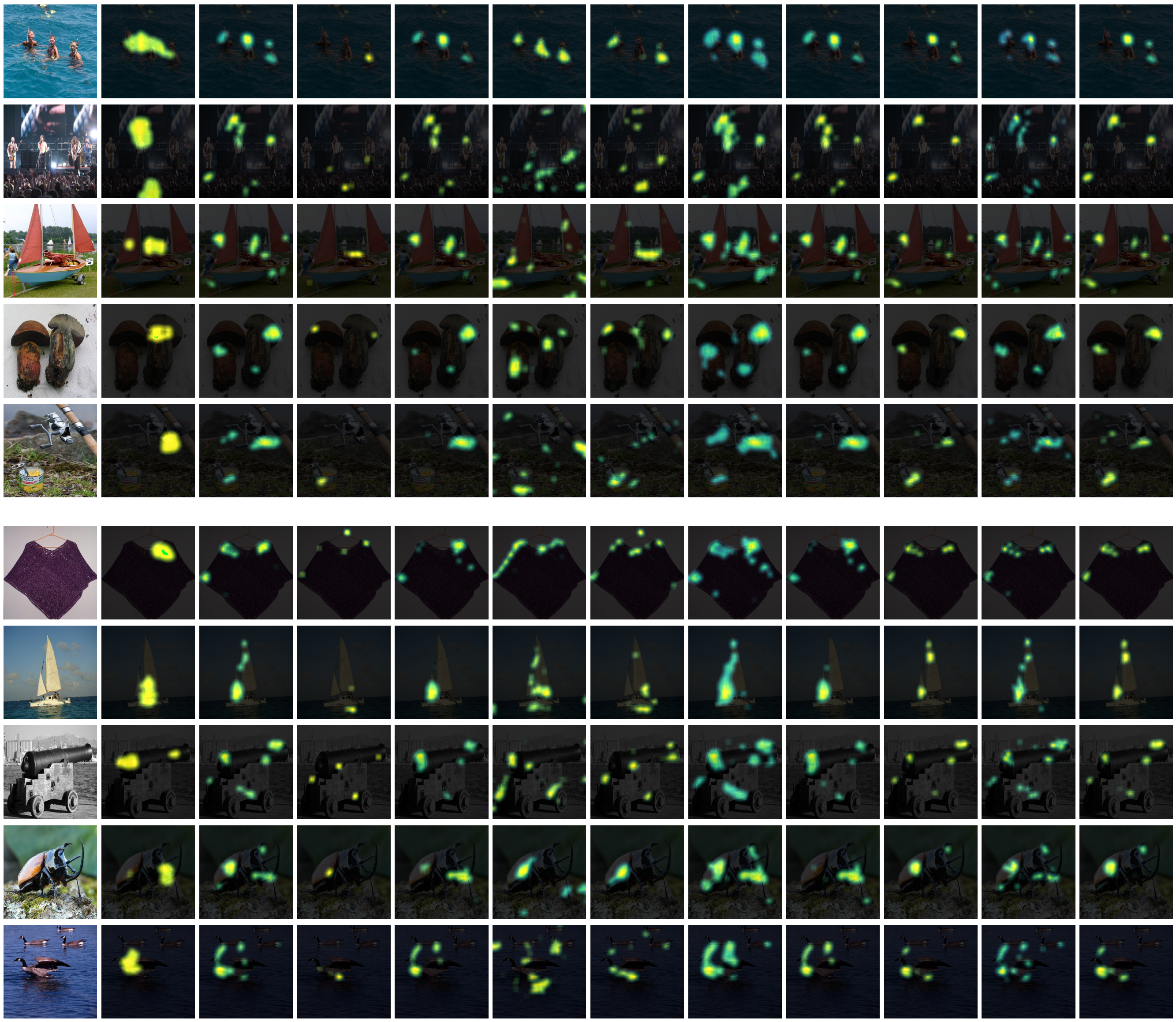

Figure 2. Gradient saliency maps after training SetFeat4-64 on miniImageNet. From left: input image, baseline [2] trained with Conv4-64, and 10 different mappers from our SetFeat4-64 (g_i is the i-th mapper). The first five rows show examples from the training dataset, and the last five are from the validation set of miniImageNet.

4. Class structure in cluster

Fig. 3 shows that tSNE for each mapper independently exhibits the expected class structure for both validation (top row) and train (bottom row) sets. Since tSNE applied over all mappers jointly **on the validation set in fig. 4 of paper**, the largest variation (across mappers) is captured.

5. Hausdorff distance ablation

Our matching feature set work can be extended to other set distances. Tab. 4 presents our method with Hausdorff (in blue) compared to our Sum-min for both miniIN and CUB.

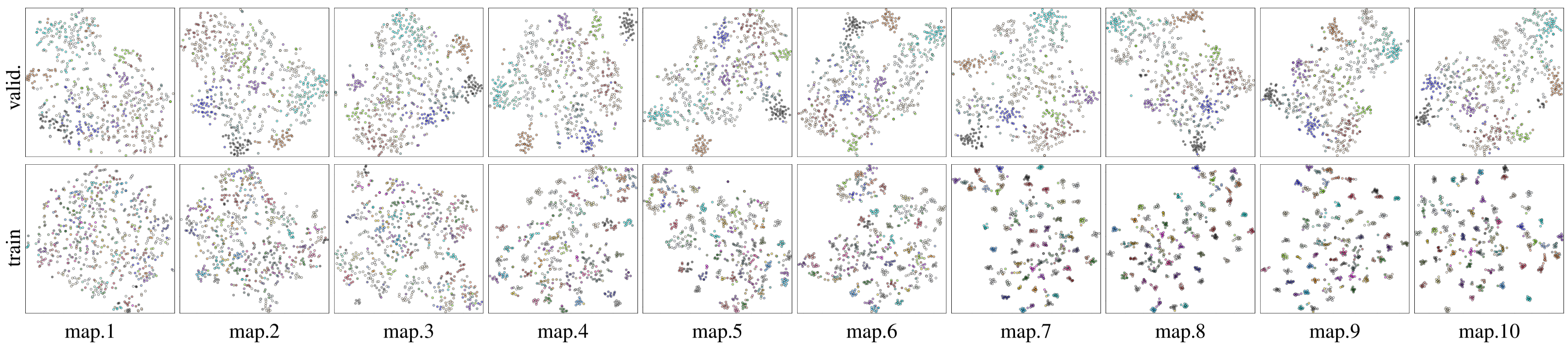

Figure 3. tSNE of miniIN’s 640 samples of 64 train and 16 valid. classes (color-coded) by SF12 mappers separately (columns).

Table 4. MiniIN (from Table 1) and CUB (from Table 3) by SF4-64 plus blue.

	config.	**1-shot**	**5-shot**
miIN	Sum-min	**57.18**	**73.67**
	Hausdorff	56.07	72.32
CUB	Sum-min	**72.09**	**87.05**
	Hausdorff	70.20	84.85

References

[1] Arman Afrasiyabi, Jean-François Lalonde, and Christian Gagné. Mixture-based feature space learning for few-shot image classification. *International Conference on Computer Vision*, 2021.

[2] Wei-Yu Chen, Yen-Cheng Liu, Zsolt Kira, Yu-Chiang Frank Wang, and Jia-Bin Huang. A closer look at few-shot classification. *International Conference on Learning Representations*, 2019.

[3] Zitian Chen, Yanwei Fu, Yu-Xiong Wang, Lin Ma, Wei Liu, and Martial Hebert. Image deformation meta-networks for one-shot learning. In *Conference on Computer Vision and Pattern Recognition*, 2019.

[4] 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.

[5] Mengye Ren, Eleni Triantafillou, Sachin Ravi, Jake Snell, Kevin Swersky, Joshua B Tenenbaum, Hugo Larochelle, and Richard S Zemel. Meta-learning for semi-supervised few-shot classification. *International Conference on Learning Representations*, 2018.

[6] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 2015.

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

[8] 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 *Conference on Computer Vision and Pattern Recognition*, 2018.

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

[10] Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. The Caltech-UCSD birds-200-2011 dataset, 2011.