Generating Representative Samples for Few-Shot Classification
Supplementary Material

Jingyi Xu
Stony Brook University
jingyixu@cs.stonybrook.edu

Hieu Le*
Amazon Robotics
ahieu@amazon.com

1. Using a deeper network architecture for the decoder

The decoder of our proposed VAE plays a vital role in our framework as it maps the latent space of the VAE and the semantic embedding to the visual feature embedding space. We use a network with two fully-connected (FC) layers for the decoder in our main setting. We experiment with a deeper network where we add an FC layer with 4096 hidden units and a LeakyReLU [5] layer to the decoder. Table 1 summarizes the results. Using a deeper network degrades the performance of our model under both 1-shot and 5-shot settings for both *mini*ImageNet [10] and *tiered*ImageNet [12].

	*mini*ImageNet		*tiered*ImageNet	
Decoder	1-shot	5-shot	1-shot	5-shot
2-FC Layers (Main paper)	72.79 ± 0.19	80.70 ± 0.16	74.21 ± 0.24	84.17 ± 0.18
3-FC Layers	71.68 ± 0.20	80.32 ± 0.16	73.84 ± 0.24	84.05 ± 0.25

Table 1. **Few-shot classification performance of our method using different network architecture for the decoder**. In our main setting, we use as our decoder a network with two fully-connected (FC) layers. "3-FC Layers" denotes the setting where we add an FC layer with 4096 hidden units and a LeakyReLU layer to the decoder. The performance degrades for both 1-shot and 5-shot settings with a deeper network.

2. Sample visualization

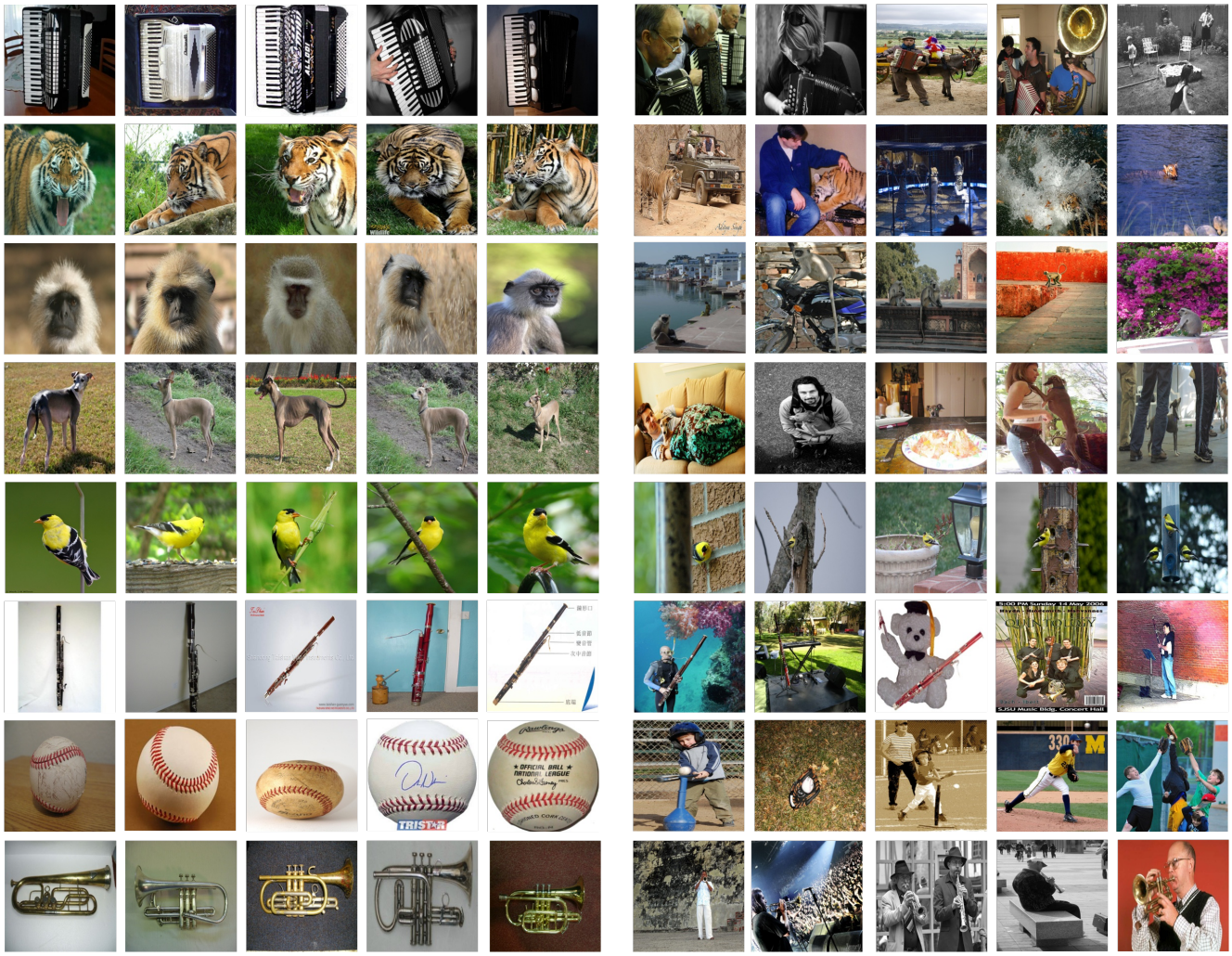
In Figure 1, we provide additional visualization of some representative samples and non-representative samples based on the representativeness probability computed via our method. The samples on the left panel are images with high probabilities. These images mostly contain the main object of the category and are easy to recognize. On the contrary, the samples on the right panel are those with small probabilities. They contain various class-unrelated objects and can lead to noisy features for constructing class prototypes.

3. Analysis of base class prototypes with the proposed sample selection method

Our feature selection method filters out non-representative samples from the base classes before training our VAE model. As shown in the main paper - Figure 3, our method performs better as the representativeness threshold increases. A higher threshold means we select samples that are more representative, resulting in a less amount of training data points. The second column of Table 2 shows the percentages of selected samples via our method for 10 classes of the *mini*ImageNet dataset. Here we use a common threshold of 0.9 for all classes. Note that the average number of available training data for each class is 600. Our method achieves state-of-the-art few-shot classification performance when using only a small fraction of these training data. For example, we only use 42 images from class "Wok" and 91 from class "Jellyfish" to train our VAE model.

We observe a correlation between the numbers of selected samples and the distances between the estimated prototypes and the ground truth prototypes of the base classes. We first estimate a *ground truth* prototype for each base class using all available features. The prototype is computed as the mean of all features. We then train a baseline VAE model using all available training data from the base classes to compare with our model trained using only the representative data. For each

*Work done outside of Amazon



Representative Non-representative

Figure 1. **Examples of representative samples (left) and non-representative samples (right)**. We visualize 5 images with high probabilities and 5 images with small probabilities computed via our proposed method for 8 classes from *tiered*ImageNet dataset.

class, we obtain two prototypes using the generated features from the baseline model and our VAE model. $\mathcal{D}_{\text{all-data}}$ denotes the distances between the prototypes estimated using the baseline model and the ground truth prototypes. $\mathcal{D}_{\text{selected-data}}$ denotes the distances for our proposed model. As can be seen from the last column of Table 2, our VAE model trained with only representative samples approximates better the ground truth prototypes. Moreover, the improvements are more pronounced for classes with small amounts of training samples.

4. 1-shot classification accuracy with different support images

We observe that performance of few-shot learning methods heavily depends on the representativeness of the support samples. For example, Figure 2 shows the 5-way 1-shot accuracy of the Meta-Baseline method [3] and our method. Here we fix the 5 classes used for evaluation and experiment with 5 different support images for one of the 5 classes. These support images have different L_2 distances to the mean feature of the class (*i.e.*, ground truth prototype), ranging from 0.5 to 0.7. A smaller value means the support feature is more representative. As can be seen, the performance of both Meta-Baseline and our method decreases dramatically when the representativeness of the support sample decreases.

Class name	% selected data	$\mathcal{D}_{\text{all-data}} \rightarrow \mathcal{D}_{\text{selected-data}}$ ($\downarrow$ improvement)
Wok	7.0%	0.68 $\rightarrow$ 0.51 ($\downarrow$ 0.17)
Parallel bars	4.3%	0.67 $\rightarrow$ 0.52 ($\downarrow$ 0.15)
Green Mamba	6.7%	0.68 $\rightarrow$ 0.54 ($\downarrow$ 0.14)
Bolete	6.0%	0.61 $\rightarrow$ 0.50 ($\downarrow$ 0.11)
Boxer	5.8%	0.76 $\rightarrow$ 0.64 ($\downarrow$ 0.12)
Jellyfish	15.2%	0.66 $\rightarrow$ 0.62 ($\downarrow$ 0.04)
Dugong	15.7%	0.69 $\rightarrow$ 0.64 ($\downarrow$ 0.05)
Spider web	17.3%	0.64 $\rightarrow$ 0.59 ($\downarrow$ 0.05)
Snorkel	13.7%	0.68 $\rightarrow$ 0.64 ($\downarrow$ 0.04)
Hair Slide	13.5%	0.50 $\rightarrow$ 0.44 ($\downarrow$ 0.06)

Table 2. **Percentages of representative samples.** We show the percentages of representative samples for 10 classes of the *tiered*ImageNet dataset, selected via our sample selection method. The VAE model trained only with these representative data estimates better the ground truth prototypes of the base classes.$\mathcal{D}_{\text{all-data}}$ denotes the distances between the prototypes estimated using the VAE model trained with all data and the ground truth prototypes. $\mathcal{D}_{\text{selected-data}}$ denotes the distances between the prototypes estimated using the VAE model trained with only the selected data and the ground truth prototypes.

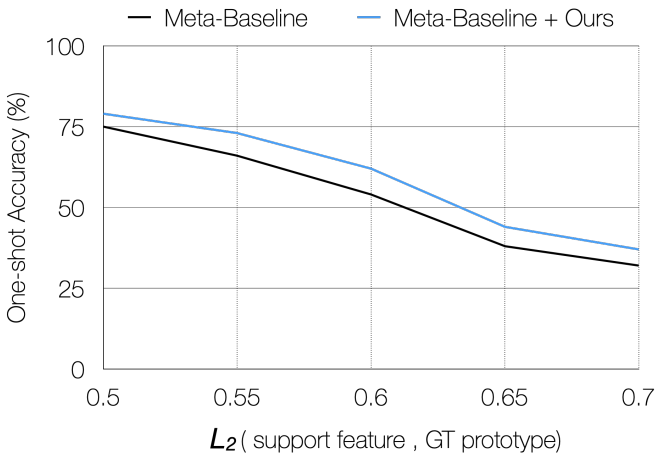


Figure 2. **1-shot, 5-way classification accuracy with different support images.** FSL methods heavily depend on the representativeness of the support samples. The figure shows the 1-shot, 5-way accuracy of the Meta-Baseline method [3] and our method. Here we fix the 5 classes used for evaluation and experiment with 5 different support images for one of the 5 classes. These support images have different L_2 distances to the mean feature of the class (*i.e.*, ground truth prototype), ranging from 0.5 to 0.7. A smaller value means the support feature is more representative. As can be seen, the performance of both Meta-Baseline and our method decreases dramatically when the representativeness of the support sample decreases.

5. Zero-shot Performance

We show the effectiveness of our generated features without any few-shot sample given. In this case, the prototype is obtained by only taking the mean of the features generated by our VAE models. As shown in Table 3, using our 0-shot prototypes outperforms the 1-shot prototypes estimated from the real sample.

	Meta-Baseline [3]	ProtoNet [16]
1-shot	63.17 ± 0.23	62.39
0-shot with SVAE	66.42 ± 0.22	65.47 ± 0.41
0-shot with R-SVAE	**66.76 ± 0.21**	**69.23 ± 0.39**

Table 3. **Zero-shot classification accuracy.** We construct class prototypes using only generated features from SVAE and R-SVAE. "1-shot" indicates the performance of the baseline methods in the 1-shot setting using support features.

6. Performance on CIFAR-FS and FC-100

In Table 4, we provide the performance of our method on two additional FSL datasets - CIFAR-FS and FC-100. On these both datasets, our method improves the Meta-Baseline method by large margins.

	Dataset	1-shot	5-shot
Meta-Baseline	CIFAR-FS	64.96 ± 0.51	75.85 ± 0.40
Meta-Baseline + SVAE		72.07 ± 0.45	77.18 ± 0.39
Meta-Baseline + R-SVAE		**73.25 ± 0.44**	**78.89 ± 0.37**
Meta-Baseline	FC-100	41.31 ± 0.42	51.84 ± 0.40
Meta-Baseline + SVAE		45.65 ± 0.40	54.37 ± 0.40
Meta-Baseline + R-SVAE		**45.75 ± 0.40**	**54.44 ± 0.40**

Table 4. 1-shot and 5-shot classification accuracy on CIFAR-FS and FC-100.

7. Comparison with methods using semantic information.

In Table 5, we compare our method with the FSL methods using semantics information including TriNet [2], CFA [6], FSLKT [9], and AM3 [14]. We did not include the results of ProtoComNet since it is under transductive setting.

	backbone	1-shot	5-shot
TriNet	ResNet18	58.12 ± 1.37	76.92 ± 0.69
CFA	ResNet18	58.5 ± 0.8	76.6 ± 0.6
FSLKT	ConvNet(128F)	64.42 ± 0.72	74.16 ± 0.56
AM3	ResNet12	65.30 ± 0.49	78.10 ± 0.36
Ours	ResNet12	**74.84 ± 0.23**	**83.28 ± 0.40**

Table 5. Comparison to prior semantic-based methods on *mini*ImageNet.

8. Ablation study for the sample selection method.

We compare our sample selection method based on Gaussian distribution with other methods including herding [11] and K-means selection [1] in Table 6. The experiment is conducted on the *mini*ImageNet dataset.

	1-shot	5-shot
Baseline	69.96 ± 0.21	79.92 ± 0.16
Herding	72.14 ± 0.20	80.48 ± 0.16
K-means selection	72.31 ± 0.20	80.55 ± 0.16
Ours (Gaussian)	**72.79 ± 0.19**	**80.70 ± 0.16**

Table 6. 1-shot and 5-shot classification accuracy on *mini*ImageNet using different clustering methods.

9. Comparison with augmentation-based methods

We show the results of our method in comparison with state-of-the-art augmentation-based methods in Table 7. These methods include MetaGAN [17], AFHN [8], Delta-Encoder [13], IDeMe-Net [4], MABAS [7] and DC [15].

	backbone	1-shot	5-shot
MetaGAN	ResNet18	52.71 ± 0.64	68.63 ± 0.67
AFHN	ResNet18	62.38 ± 0.72	78.16 ± 0.56
Delta-Encoder	ResNet18	59.90	69.70
IDeMe-Net	ResNet10	59.14 ± 0.86	74.63 ± 0.74
MABAS	ResNet12	65.08± 0.86	82.70 ± 0.54
DC	WRN28	68.57 ± 0.55	82.88 ± 0.42
Ours	ResNet12	**74.84 ± 0.23**	**83.28 ± 0.40**

Table 7. Comparison to prior augmentation-based methods on *mini*ImageNet.

References

[1] Ernie Chang, Xiaoyu Shen, Hui-Syuan Yeh, and Vera Demberg. On training instance selection for few-shot neural text generation. *ArXiv*, abs/2107.03176, 2021. 4

[2] Wei-Yu Chen, Yen-Cheng Liu, Zsolt Kira, Yu-Chiang Frank Wang, and Jia-Bin Huang. A closer look at few-shot classification. In *International Conference on Machine Learning(ICML)*, 2019. 4

[3] Yinbo Chen, Zhuang Liu, Huijuan Xu, Trevor Darrell, and Xiaolong Wang. Meta-baseline: Exploring simple meta-learning for few-shot learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9062–9071, 2021. 2, 3, 4

[4] Z. Chen, Yanwei Fu, Yu-Xiong Wang, Lin Ma, Wei Liu, and Martial Hebert. Image deformation meta-networks for one-shot learning. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8672–8681, 2019. 4

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *ICCV*, 2015. 1

[6] Ping Hu, Ximeng Sun, Kate Saenko, and Stan Sclaroff. Weakly-supervised compositional featureaggregation for few-shot recognition. volume abs/1906.04833, 2019. 4

[7] Jaekyeom Kim, Hyoungseok Kim, and Gunhee Kim. Model-agnostic boundary-adversarial sampling for test-time generalization in few-shot learning. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16*, pages 599–617. Springer, 2020. 4

[8] Kai Li, Yulun Zhang, Kunpeng Li, and Yun Fu. Adversarial feature hallucination networks for few-shot learning. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 4

[9] Zhimao Peng, Zechao Li, Junge Zhang, Yan Li, Guo-Jun Qi, and Jinhui Tang. Few-shot image recognition with knowledge transfer. pages 441–449, 2019. 4

[10] Sachin Ravi and H. Larochelle. Optimization as a model for few-shot learning. In *ICLR*, 2017. 1

[11] Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, G. Sperl, and Christoph H. Lampert. icarl: Incremental classifier and representation learning. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5533–5542, 2017. 4

[12] Mengye Ren, Eleni Triantafillou, Sachin Ravi, Jake Snell, Kevin Swersky, Joshua B. Tenenbaum, H. Larochelle, and Richard S. Zemel. Meta-learning for semi-supervised few-shot classification. *ArXiv*, abs/1803.00676, 2018. 1

[13] Eli Schwartz, Leonid Karlinsky, Joseph Shtok, Sivan Harary, Mattias Marder, Rogerio Feris, Abhishek Kumar, Raja Giryes, and Alex M. Bronstein. Delta-encoder: an effective sample synthesis method for few-shot object recognition. In *Advances in Neural Information Processing Systems (NeurIPS)*, June 2018. 4

[14] Chen Xing, Negar Rostamzadeh, Boris N. Oreshkin, and Pedro H. O. Pinheiro. Adaptive cross-modal few-shot learning. In *NeurIPS*, 2019. 4

[15] Shuo Yang, Lu Liu, and Min Xu. Free lunch for few-shot learning: Distribution calibration. volume abs/2101.06395, 2021. 4

[16] Han-Jia Ye, Hexiang Hu, D. Zhan, and Fei Sha. Few-shot learning via embedding adaptation with set-to-set functions. pages 8805–8814, 2020. 4

[17] Ruixiang ZHANG, Tong Che, Zoubin Ghahramani, and Yangqiu Song Yoshua Bengi and. Metagan: An adversarial approach to few-shot learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2018. 4