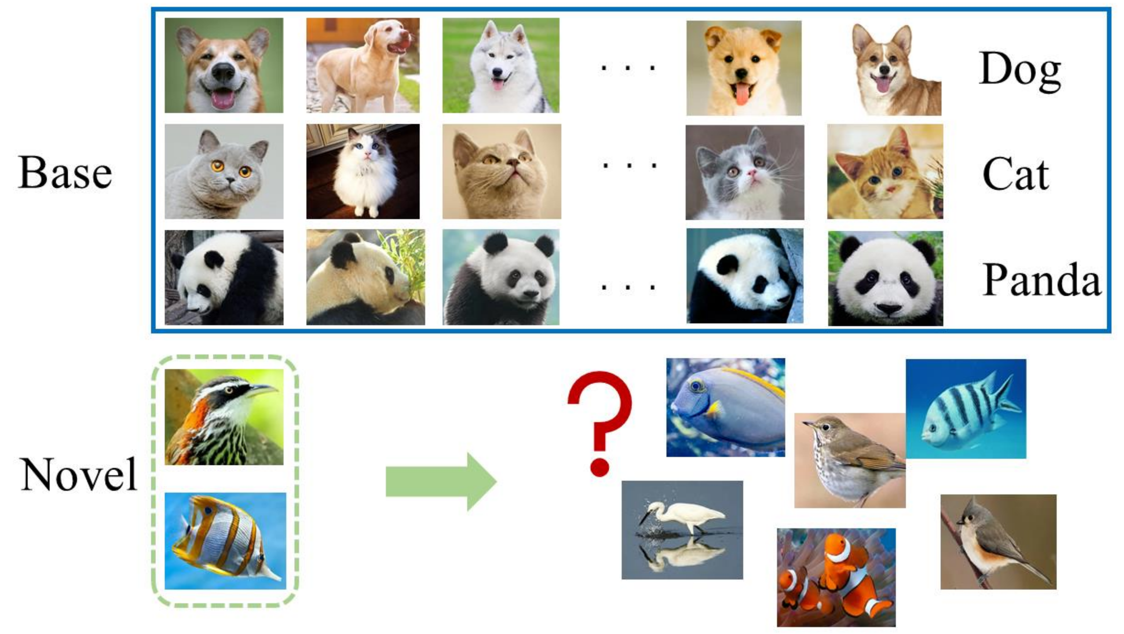
Instance-based Max-margin for Practical Few-shot Recognition

Minghao Fu, Ke Zhu*
National Key Laboratory for Novel Software Technology, Nanjing University, China
School of Artificial Intelligence, Nanjing University, China
fumh@lamda.nju.edu.cn, zhuk@lamda.nju.edu.cn

CVPR SEATTLE, WA JUNE 17-21, 2024

1. Introduction & Motivation

- Traditional few-shot learning evaluation relies on a split between *base and novel* sets, which is impractical.

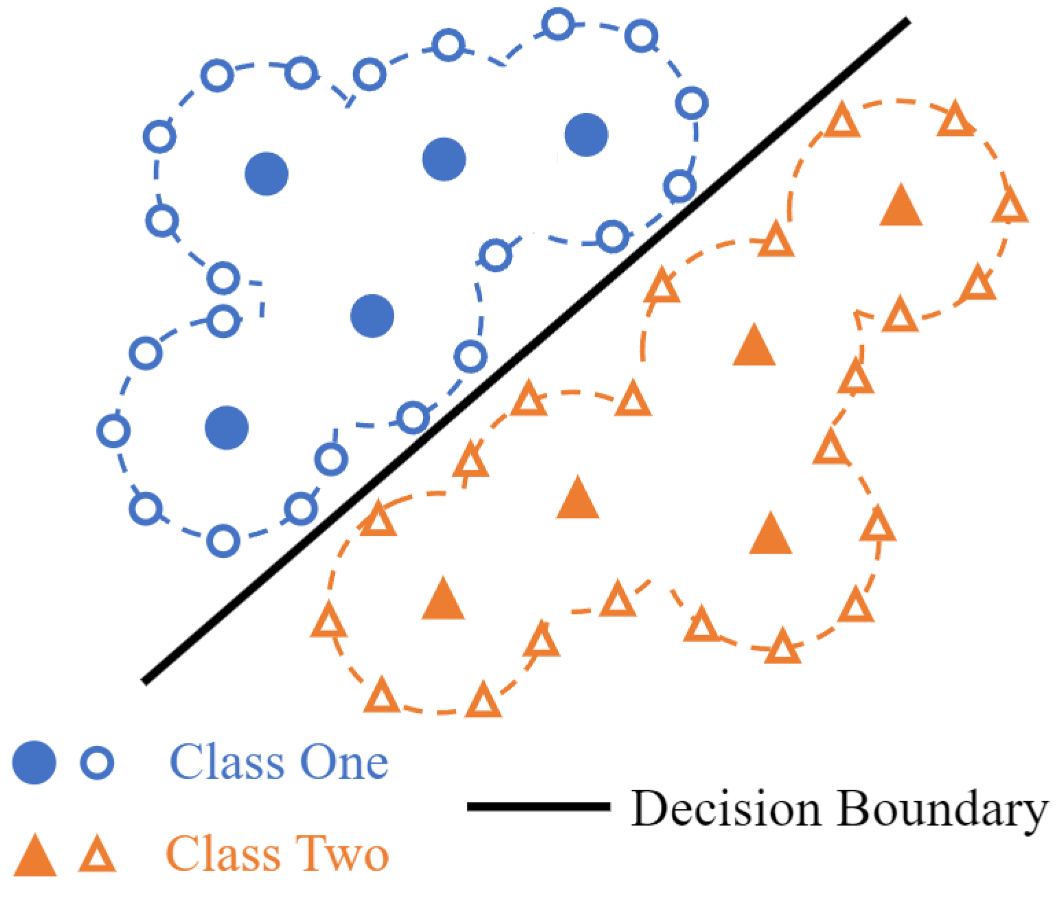


- Drawback:
  - Base and novel sets are semantically closely related.
  - Recognize a few categories.
  - Sophisticated evaluation procedure (≥500 runs).
- As a stark contrast, we human can:
  - Accumulate prior knowledge as both common-sense and domain knowledge from many diverse domains.
  - Learn many instead of a few novel concepts.

Our Solution

- **pFSL**: Practical Few-Shot Learning
  - ✓ Much more analogous to human prior knowledge.
    - Removing the base set.
    - Unsupervised pre-trained model based on big data.
    - Many-way FSL.
  - ✓ Much easier to evaluate.
    - Usually 3 runs is enough for evaluation.
  - ✓ Much more challenge than traditional FSL.
    - Traditional FSL is saturated to a certain extent, we believe this new setting can push the technical frontier further.

2. IbM2: Instance-based Max-margin



- Margin maximization in an *instance-based* manner
  - ✓ i.i.d. generate R virtual samples $z_{i,r}$ centered at z_i

$$z_{i,r} = z_i + \epsilon\delta_{i,r} \qquad \delta \sim N(0, I_d)$$

  - Virtual examples reside around a shell.

Gaussian Annulus Theorem: Almost all the probability of a high-dimensional spherical Gaussian with unit variance is concentrated in a thin annulus of width O(1) with radius √d when d is large.

  - Hypersphere of different examples can overlap.
  - ✓ **Maximize radius ϵ** meanwhile requires virtual examples to be correctly classified.

- How to maximize ϵ ?
  - Binary search based on training accuracy.
  - Slack trick: accuracy threshold T does not need to be 1 (~0.9).
- Ellipsoidal calibration
  - Isotropic noise sampling ignores the structural property of training example
  - An improved version:

$$z_{i,r} = z_i + \epsilon(s \odot \delta_{i,r})$$

Algorithm 1 Pseudo code for searching ϵ

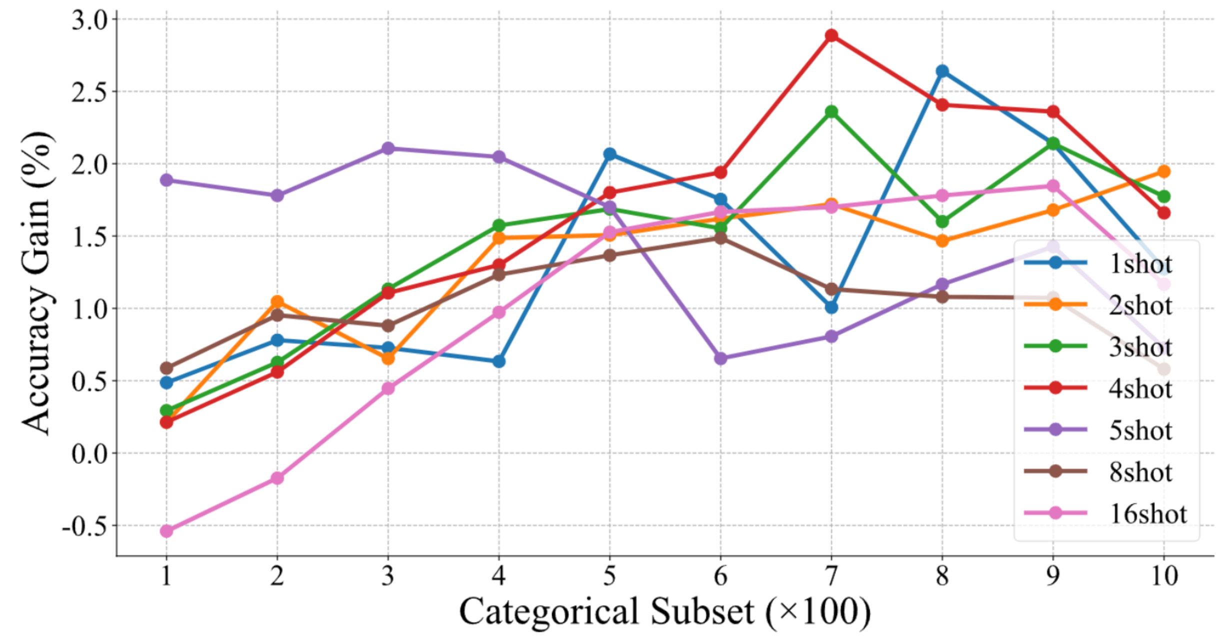
```
# Inputs:
#     x  : training features of a few-shot task
#     y  : training labels   of a few-shot task
#     R  : sampling times for an instance
#     T  : accuracy threshold for searching
# Outputs:
#     eps  : epsilon for sampling

left = 0.0
right = a large value
eps = right / 2

W = init_classifier
while True:
        acc = train_and_eval(W, x, y, eps, R)
        if acc > T:
                left = eps   # increase epsilon
        else:
                right = eps  # decrease epsilon
        eps = (left + right) / 2.0
        if right - left < 0.05:
                break
```

3. Experiments

Results on pFSL setting

Dataset	Pre-training Method	Backbone	IbM2	Shot per Class 1	2	3	4	5	8	16
ImageNet-1K	DINO	ViT-S/16		39.2 ± 0.3	49.2 ± 0.2	54.1 ± 0.4	56.7 ± 0.2	58.0 ± 0.1	60.4 ± 0.1	62.7 ± 0.1
			✓	39.2 ± 0.3	49.4 ± 0.3	**54.6 ± 0.4**	**57.6 ± 0.1**	**59.3 ± 0.1**	**62.4 ± 0.2**	**65.8 ± 0.1**
	MoCov3	ViT-S/16		32.7 ± 0.6	42.0 ± 0.2	46.9 ± 0.3	49.6 ± 0.4	51.0 ± 0.1	53.8 ± 0.1	56.6 ± 0.2
			✓	**33.9 ± 0.6**	**43.2 ± 0.2**	**48.4 ± 0.3**	**51.3 ± 0.3**	**52.8 ± 0.2**	**56.1 ± 0.1**	**59.8 ± 0.2**
	MSN	ViT-S/16		47.9 ± 0.1	56.2 ± 0.4	59.8 ± 0.3	61.6 ± 0.1	62.4 ± 0.2	64.4 ± 0.3	66.1 ± 0.1
			✓	47.8 ± 0.2	56.4 ± 0.4	**60.5 ± 0.2**	**62.5 ± 0.2**	**63.6 ± 0.2**	**66.0 ± 0.2**	**68.4 ± 0.0**
		ViT-B/4		53.2 ± 0.2	64.5 ± 0.4	68.9 ± 0.2	70.9 ± 0.2	72.0 ± 0.3	73.8 ± 0.1	75.0 ± 0.2
			✓	**54.0 ± 0.1**	64.9 ± 0.5	**69.4 ± 0.2**	**71.4 ± 0.1**	**72.7 ± 0.4**	**74.7 ± 0.0**	**76.4 ± 0.2**
		ViT-L/7		57.3 ± 0.4	66.5 ± 0.4	69.8 ± 0.5	71.6 ± 0.4	72.2 ± 0.2	73.8 ± 0.1	75.1 ± 0.1
			✓	57.7 ± 0.4	66.6 ± 0.5	70.1 ± 0.6	71.8 ± 0.4	72.6 ± 0.2	**74.3 ± 0.1**	**76.0 ± 0.0**
	SimCLR	ResNet50		21.4 ± 0.4	30.3 ± 0.1	36.1 ± 0.3	39.8 ± 0.2	42.0 ± 0.1	46.8 ± 0.1	51.9 ± 0.0
			✓	**23.6 ± 0.4**	**33.4 ± 0.2**	**39.0 ± 0.4**	**42.0 ± 0.3**	**44.2 ± 0.1**	**48.0 ± 0.0**	**52.7 ± 0.0**
	BYOL	ResNet50		26.5 ± 0.3	35.7 ± 0.2	41.5 ± 0.4	45.1 ± 0.2	47.2 ± 0.1	51.8 ± 0.1	57.1 ± 0.1
			✓	**27.5 ± 0.3**	**37.5 ± 0.1**	**43.3 ± 0.4**	**46.8 ± 0.2**	**49.1 ± 0.1**	**53.2 ± 0.1**	**58.0 ± 0.1**
CUB	DINO	ViT-S/16		35.4 ± 1.2	49.0 ± 0.5	56.8 ± 0.8	60.8 ± 0.7	65.2 ± 0.9	70.6 ± 0.9	75.9 ± 0.3
			✓	**36.2 ± 1.4**	**49.6 ± 0.6**	**57.4 ± 1.0**	**62.0 ± 0.6**	**66.4 ± 0.8**	**72.5 ± 0.8**	**79.0 ± 0.2**
	MSN	ViT-S/16		32.1 ± 1.6	45.0 ± 0.6	53.1 ± 0.6	56.7 ± 0.1	61.4 ± 0.5	67.3 ± 0.0	73.6 ± 0.4
			✓	**33.0 ± 1.4**	**45.8 ± 0.7**	53.2 ± 0.9	57.1 ± 0.4	**62.0 ± 1.0**	**68.4 ± 0.1**	**75.7 ± 0.2**
		ViT-L/7		34.9 ± 1.3	49.4 ± 0.4	58.8 ± 0.8	62.7 ± 0.9	67.2 ± 0.3	73.8 ± 0.8	80.4 ± 0.2
			✓	**37.5 ± 1.2**	**50.1 ± 0.5**	59.0 ± 0.8	62.6 ± 0.5	67.5 ± 0.6	73.9 ± 0.4	**81.0 ± 0.1**

In almost all cases,IbM2 benefited the few-shot learning and improved the top-1 accuracy.



On what classes can IbM2 help?

- IbM2 improves average per-class accuracy in almost every histogram bin.
- Recognition accuracy of more than 65% categories out of 1000 is improved in all shots.
- The more accurate a class is, the higher gains IbM2 can achieve over the baseline.

4. Contributions & Conclusions

- ✓ We propose a practical few-shot learning setting (pFSL): many-way (e.g., 200-way) recognition, uses an unsupervised pre-trained model, and has no base set.
- ✓ We propose IbM2, an instance-based max-margin algorithm, which suits few-shot, high dimensionality, and multiclass naturally.
- ✓ As shown by extensive experiments, IbM2 consistently improves both traditional FSL and the proposed pFSL.

First Author Second Author

LAMDA Learning And Mining from DatA