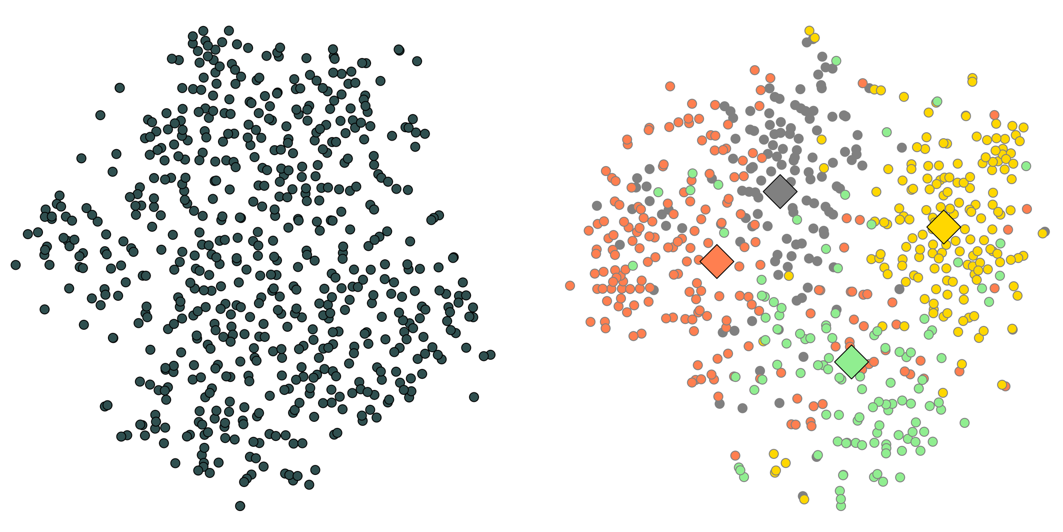
Mixture-based Feature Space Learning for Few-shot Image Classification

2021 ICCV OCTOBER 11-17 VIRTUAL

Arman Afrasiyabi, Jean-François Lalonde, Christian Gagné
arman.afrasiyabi.1@ulaval.ca, {jean-francois.lalonde, christian.gagne}@gel.ulaval.ca
Université Laval, Institut intelligence et données, Mila

Motivation

- We introduce Mixture-based Feature Space Learning (MixtFSL) for obtaining a rich and robust feature representation
- our MixtFSL aims to learn a multimodal representation for the base classes

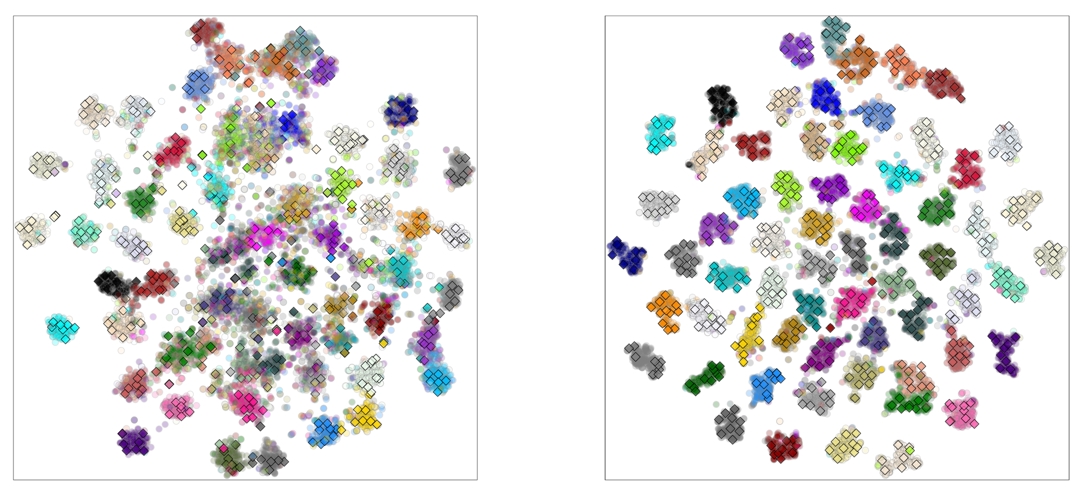


(a) without MixtFSL (b) our MixtFSL

- The idea is to learn both the *representation* and the *mixture model* jointly in an online manner

MixtFSL

- We present a robust two-stage scheme for training such a model.
- The training is done end-to-end in a fully differentiable fashion, without the need for an offline clustering method.

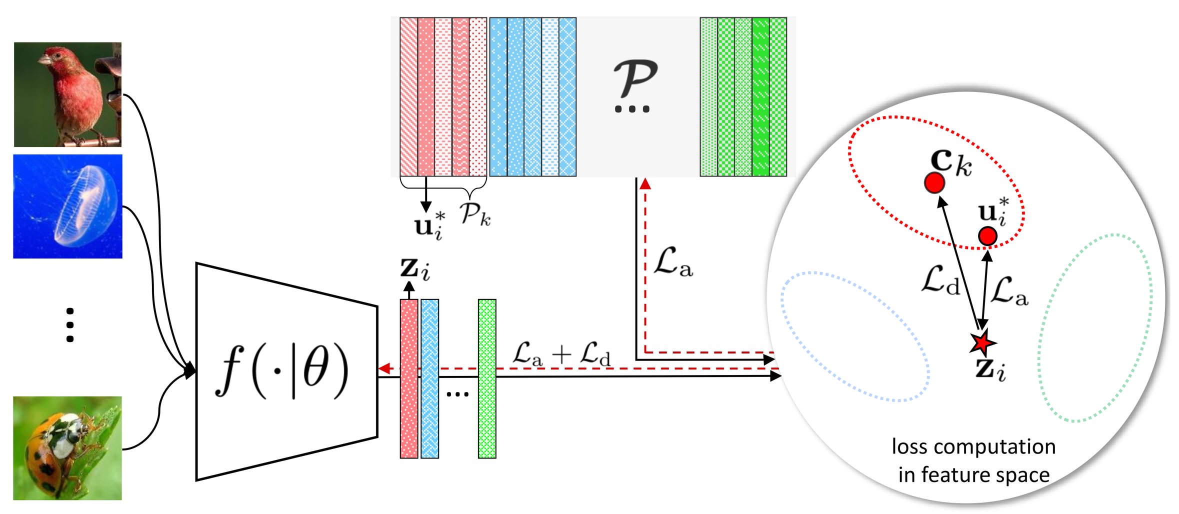


(a) initial training (b) progressive following

- We demonstrate, through an extensive experiments on four standard datasets and using four backbones, that our MixtFSL outperforms the state of the art in most of the cases tested.

Initial Training

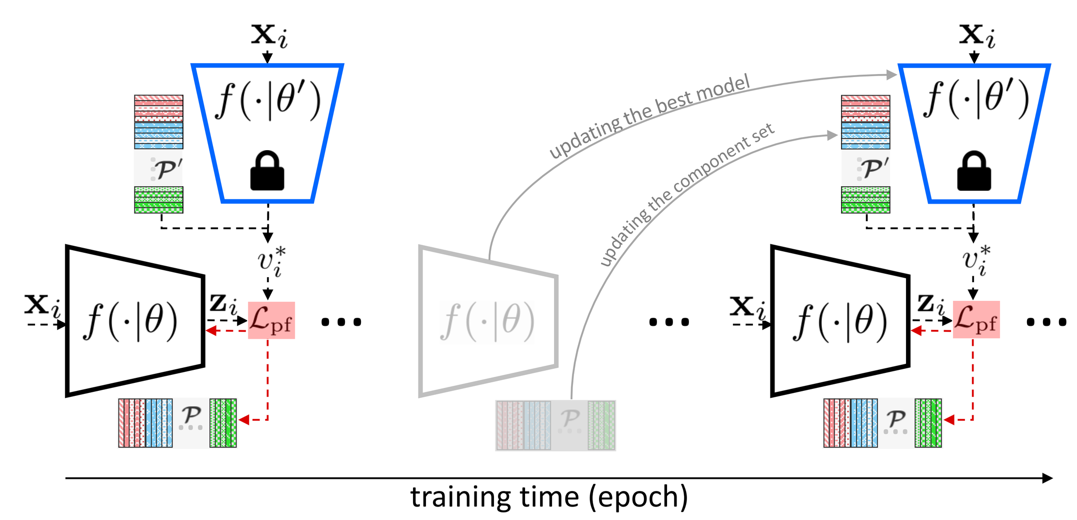
- The initial training of $f(\cdot|\theta)$ and the learnable mixture model $\mathcal{P}$ from the base class set $\mathcal{X}^b$ is illustrated in figure bellow.



- Model parameters are updated using two losses: the "assignment" loss $\mathcal{L}_a$, which updates both the feature extractor and the mixture model such that feature vectors are assigned to their nearest mixture component; and the "diversity" loss $\mathcal{L}_d$, which updates the feature extractor to diversify the selection of components for a given class.

Progressive Following

- The progressive following stage that aim to break the complex dynamic of simultaneously determining nearest components while training the representation $f(\cdot|\theta)$ and mixture $\mathcal{P}$. The approach is shown in bellow.
- Using the "prime" notation (θ' and $\mathcal{P}'$ to specify the best feature extractor parameters and mixture component so far, resp.), the approach starts by taking a copy of $f(\cdot|\theta')$ and $\mathcal{P}'$, and by using them to determine the nearest component of each training instance:



Results: miniImageNet

- Evaluations of our MixFSL on *mini*-ImageNet using different Conv4 and ResNet12.

Table 1. Evaluation on miniImageNet in 5-way. Bold/blue is best/second, and ± is the 95% confidence intervals in 600 episodes.

Method	Backbone	1-shot	5-shot
ProtoNet [64]	Conv4	49.42 ± 0.78	68.20 ± 0.66
MAML [19]	Conv4	48.07 ± 1.75	63.15 ± 0.91
RelationNet [67]	Conv4	50.44 ± 0.82	65.32 ± 0.70
Baseline++ [8]	Conv4	48.24 ± 0.75	66.43 ± 0.63
IMP [2]	Conv4	49.60 ± 0.80	68.10 ± 0.80
MemoryNetwork [5]	Conv4	**53.37** ± 0.48	66.97 ± 0.35
Arcmax [1]	Conv4	51.90 ±0.79	69.07 ± 0.59
Neg-Margin [41]	Conv4	52.84 ±0.76	70.41 ±0.66
MixtFSL (ours)	Conv4	52.82 ±0.63	**70.67** ±0.57
DNS [62]	RN-12	62.64 ±0.66	78.83 ±0.45
Var.FSL [87]	RN-12	61.23 ±0.26	77.69 ±0.17
MTL [66]	RN-12	61.20 ±1.80	75.50 ±0.80
SNAIL [46]	RN-12	55.71 ±0.99	68.88 ±0.92
AdaResNet [48]	RN-12	56.88 ±0.62	71.94 ±0.57
TADAM [49]	RN-12	58.50 ±0.30	76.70 ±0.30
MetaOptNet [37]	RN-12	62.64 ±0.61	78.63 ±0.46
Simple [69]	RN-12	62.02 ±0.63	79.64 ±0.44
TapNet [83]	RN-12	61.65 ±0.15	76.36 ±0.10
Neg-Margin [41]	RN-12	63.85 ±0.76	81.57 ±0.56
MixtFSL (ours)	RN-12	**63.98** ±0.79	**82.04** ±0.49

tieredImageNet and FC100

- Evaluations of our MixFSL on tieredImageNet using different ResNet12 and ResNet18.

Table 2. Evaluation on tieredImageNet and FC100 in 5-way classification. Bold/blue is best/second best, and ± indicates the 95% confidence intervals over 600 episodes.

	Method	Backbone	1-shot	5-shot
tieredImageNet	DNS [62]	RN-12	66.22 ±0.75	82.79 ±0.48
	MetaOptNet [37]	RN-12	65.99 ±0.72	81.56 ±0.53
	Simple [69]	RN-12	69.74 ±0.72	84.41 ±0.55
	TapNet [83]	RN-12	63.08 ±0.15	80.26 ±0.12
	Arcmax* [1]	RN-12	68.02 ±0.61	83.99 ±0.62
	MixtFSL (ours)	RN-12	**70.97** ±1.03	**86.16** ±0.67
	Arcmax [1]	RN-18	65.08 ±0.19	83.67 ±0.51
	ProtoNet [64]	RN-18	61.23 ±0.77	80.00 ±0.55
	MixtFSL (ours)	RN-18	**68.61** ±0.91	**84.08** ±0.55
FC100	TADAM [49]	RN-12	40.1 ± 0.40	56.1 ± 0.40
	MetaOptNet [37]	RN-12	41.1 ± 0.60	55.5 ± 0.60
	ProtoNet† [64]	RN-12	37.5 ± 0.60	52.5 ± 0.60
	MTL [66]	RN-12	43.6 ± 1.80	55.4 ± 0.90
	MixtFSL (ours)	RN-12	**44.89** ±0.63	**60.70** ±0.67
	Arcmax [1]	RN-18	40.84 ± 0.71	57.02 ± 0.63
	MixtFSL (ours)	RN-18	**41.50** ±0.67	**58.39** ±0.62

*our implementation †taken from [37]

CUB and Cross-Domain

- Evaluations of our MixtFSL on CUB in object recognition and cross-domain adaptation using ResNet18.

Table 3. Fine-grained and on cross-domain from miniImageNet to CUB evaluation in 5-way using ResNet-18. Bold/blue is best/second, and ± is the 95% confidence intervals on 600 episodes.

	CUB		miniIN→CUB
Method	1-shot	5-shot	5-shot
GNN-LFT◇ [70]	51.51 ±0.8	73.11 ±0.7	–
Robust-20 [13]	58.67 ±0.7	75.62 ±0.5	–
RelationNet‡ [67]	67.59 ±1.0	82.75 ±0.6	57.71 ±0.7
MAML‡ [18]	68.42 ±1.0	83.47 ±0.6	51.34 ±0.7
ProtoNet‡ [64]	71.88 ±0.9	86.64 ±0.5	62.02 ±0.7
Baseline++ [8]	67.02 ±0.9	83.58 ±0.5	64.38 ±0.9
Arcmax [1]	71.37 ±0.9	85.74 ±0.5	64.93 ±1.0
Neg-Margin [41]	72.66 ±0.9	**89.40** ±0.4	67.03 ±0.8
MixtFSL (ours)	**73.94** ±1.1	86.01 ±0.5	**68.77** ±0.9

‡taken from [68] ◇backbone is ResNet-10

An Extension of MixtFSL

- Two changes are necessary to adapt our MixtFSL to exploit the "centroid alignment" of [1].
- First, we employ the learned mixture model $\mathcal{P}$ to find the related base classes.
- Second, they used a classification layer $\mathbf{W}$ in $c(\mathbf{x}|\mathbf{W}) \equiv \mathbf{W}^\top f(\mathbf{x}|\theta)$ (followed by softmax).

Table 6. Comparison of our MixtFSL with alignment (MixtFSL-Align) in 5-way classification. Here, bold is the best performance.

	Method	Backbone	1-shot	5-shot
miniIN	Cent. Align.* [1]	RN-12	63.44 ±0.67	80.96 ±0.61
	MixtFSL-Align. (ours)	RN-12	**64.38** ±0.73	**82.45** ±0.62
	Cent. Align.* [1]	RN-18	59.85 ±0.67	80.62 ±0.72
	MixtFSL-Align. (ours)	RN-18	**60.44** ±1.02	**81.76** ±0.74
tieredIN	Cent. Align.* [1]	RN-12	71.08 ±0.93	86.32 ±0.66
	MixtFSL-Align. (ours)	RN-12	**71.83** ±0.99	**88.20** ±0.55
	Cent. Align.* [1]	RN-18	69.18 ±0.86	**85.97** ±0.51
	MixtFSL-Align. (ours)	RN-18	**69.82** ±0.81	85.57 ±0.60

* our implementation

Acknowledgement

This project was supported by funding from NSERC-Canada, Mitacs, Prompt-Québec. Project page: `https://lvsn.github.io/MixtFSL/`