UNIVERSITÉ LAVAL

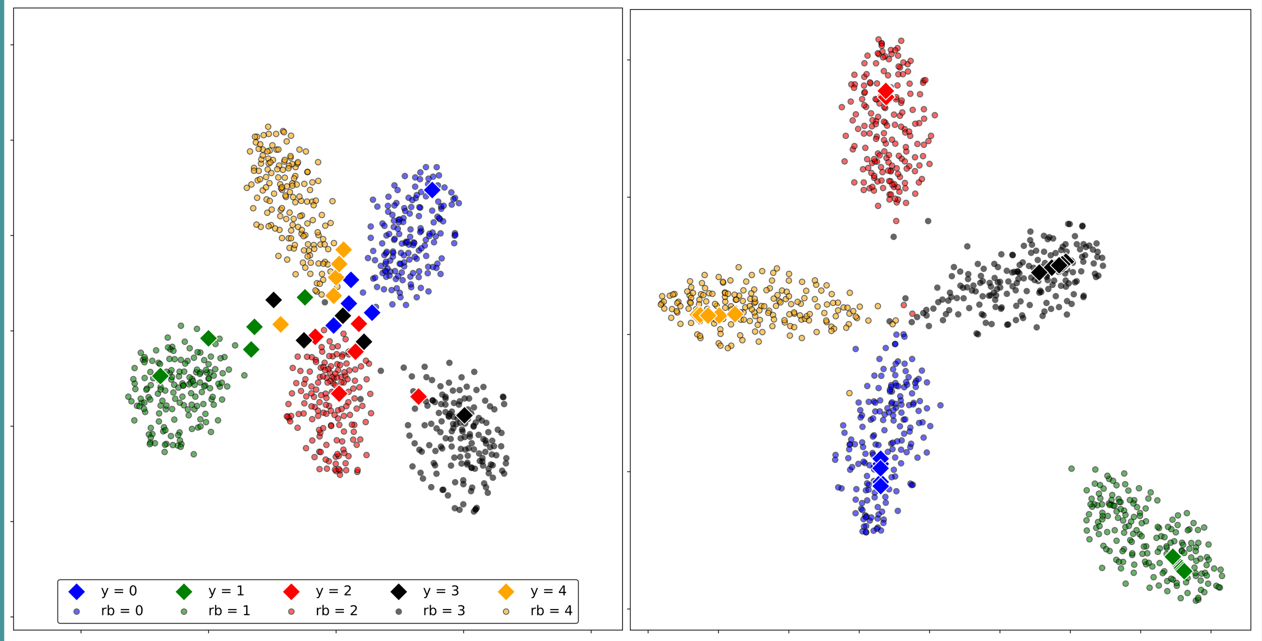
Associative Alignment for Few-shot Image Classification

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

Motivation

- We aim at preventing overfitting without restricting the learning capabilities
- Our approach relies on the transfer learning, but it consists of two stages: i) detecting related bases, ii) associative alignment.
- Fine-tuning directly on related base could confuse the network (see fig.(a))



(a) before alignment (b) after alignment

- We aim to *align*, in feature space, the novel examples with the related base samples (fig.(b))

Detecting Related Bases

- Our detecting algorithm associates B base classes to each novel class
- train the both of the network and classifcation layer using base categories
- adapt the last layer using novel categories
- the B base classes with the highest score for a given novel class are kept as the related base

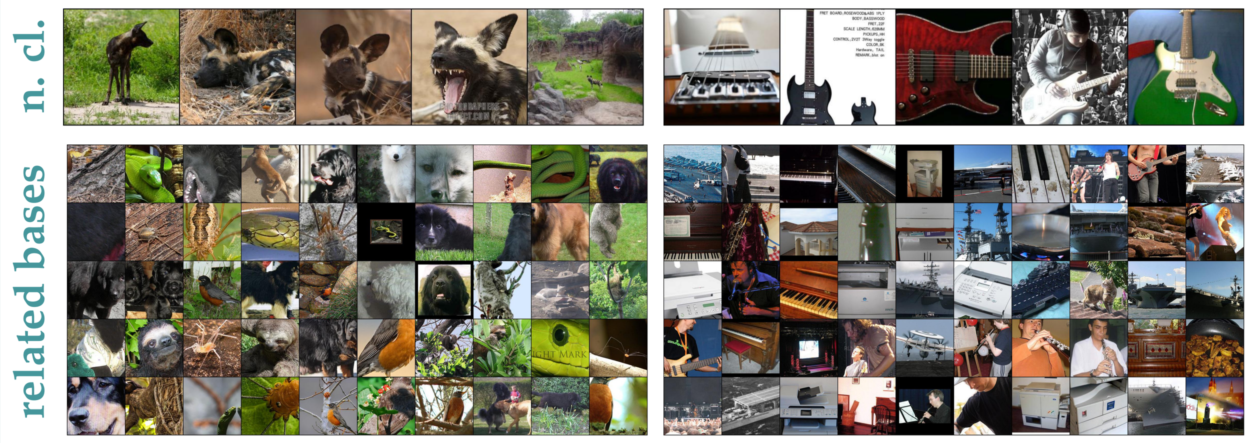
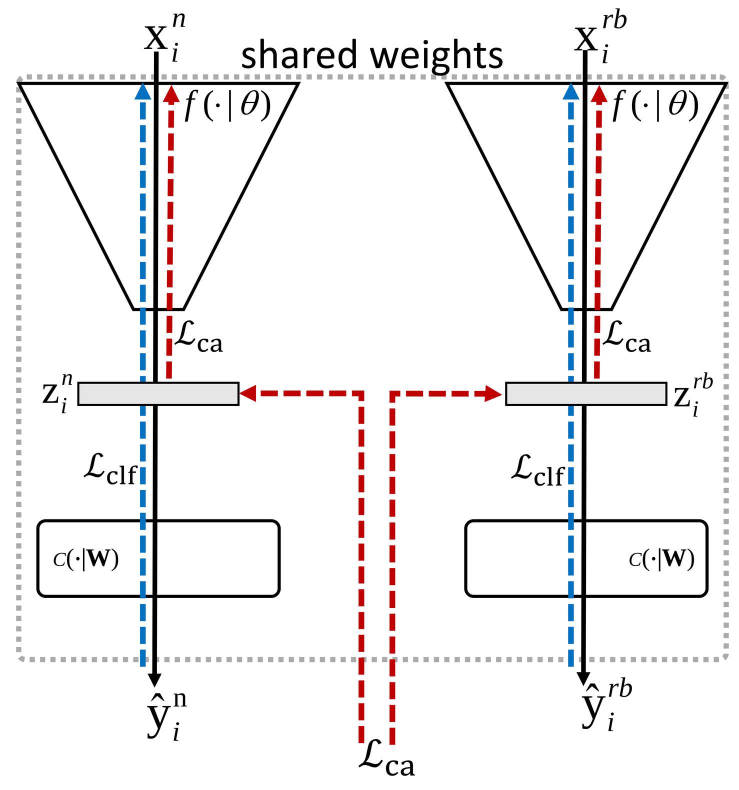


figure above shows two examples of novel categories in 5-shot, 5-way, the result of the detection algorithm as related bases for $B = 10$.

Centroid Alignment

- Our centroid alignment explicitly push the novel classes towards the center of their related bases
- first we compute the centroid of each novel classes $\boldsymbol{\mu}_i$ using parameterized network $f(.|\theta)$
- given novel class $\mathcal{X}^n$ examples and their related bases$\mathcal{X}^{rb}$, we define the centroid alignment loss:

$$\mathcal{L}_{\mathrm{ca}}(\mathcal{X}^n) = \frac{-1}{N} \sum_{i=1}^{K^n} \sum_{(\mathbf{x}_j,\cdot)\in\mathcal{X}_i^n} \log \frac{\exp[-\|f(\mathbf{x}_j|\theta) - \boldsymbol{\mu}_i\|_2^2]}{\sum_{k=1}^{K^n}\exp[-\|f(\mathbf{x}_j|\theta) - \boldsymbol{\mu}_k\|_2^2]}$$

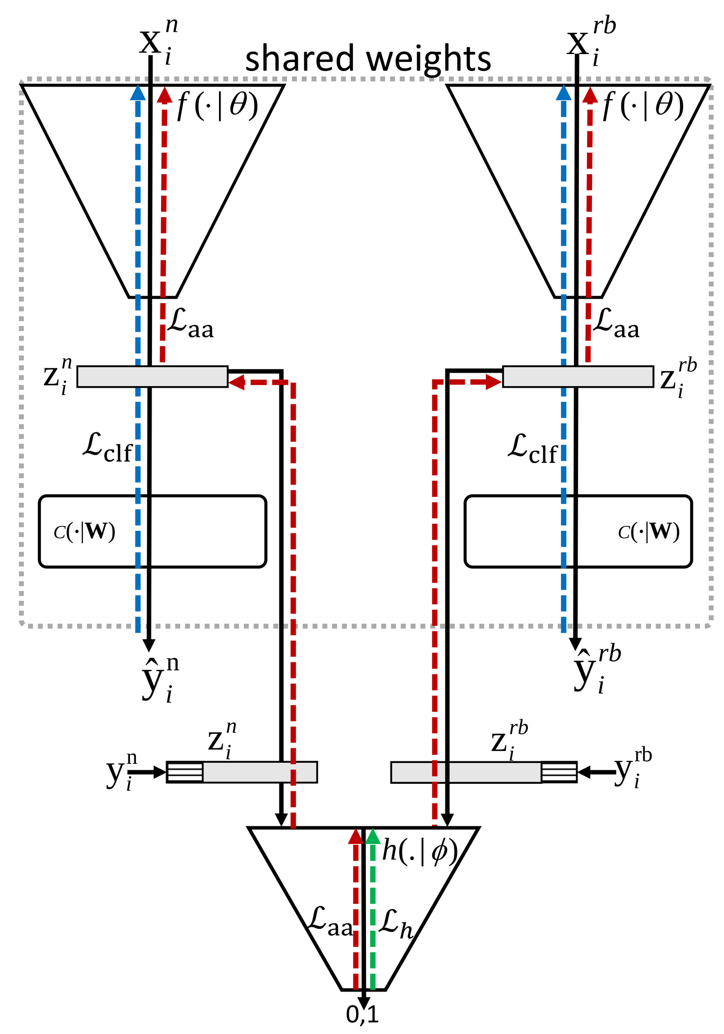


- where, N is the number of all examples

Adversarial Alignment

- Here, a parameterized critic network $h(\cdot|\phi)$ is conditioned by the concatenation of the feature embedding of either a novel exampl $\mathbf{x}_i^n$ and one of its related class example $\mathbf{x}_j^{rb}$, along with y_i^n
- after updating the critic network, the encoder parameters θ are updated using the following loss

$$\mathcal{L}_{\mathrm{aa}}(\mathcal{X}^n) = \frac{1}{K^n} \sum_{(\mathbf{x}_i^n, y_i^n)\in\mathcal{X}^n} h\left([f(\mathbf{x}_i^n|\theta)\ y_i^n]|\phi\right) .$$



Results: Object Recognition Evaluation

- Tables bellow presents evaluations of the proposed strategies on *mini*-ImageNet and tieredImageNet using a ResNet-18 and WRN-28-10.

	Method	*mini*-ImageNet 1-shot	*mini*-ImageNet 5-shot	tieredImageNet 1-shot	tieredImageNet 5-shot
ResNet-18	ProtoNet‡ [19]	54.16 ± 0.82	73.68 ± 0.65	61.23 ± 0.77	80.00 ± 0.55
	SNAIL [13]	55.71 ± 0.99	68.88 ± 0.92	–	–
	MTL [20]	61.20 ± 1.80	75.50 ± 0.80	–	–
	VariationalFSL [24]	61.23 ± 0.26	77.69 ± 0.17	–	–
	MetaOptNet [11]	**62.64** ± 0.61	78.63 ± 0.46	65.99 ± 0.72	81.56 ± 0.53
	our baseline (arcmax)	58.07 ± 0.82	76.62 ± 0.58	65.08 ± 0.19	83.67 ± 0.51
	adversarial alignment	58.84 ± 0.77	77.92 ± 0.82	66.44 ± 0.61	85.12 ± 0.53
	centroid alignment	59.88 ± 0.67	**80.35** ± 0.73	**69.29** ± 0.56	**85.97** ± 0.49

	Method	*mini*-ImageNet 1-shot	*mini*-ImageNet 5-shot	tieredImageNet 1-shot	tieredImageNet 5-shot
WRN-28-10	LEO [16]	61.76 ± 0.08	77.59 ± 0.12	66.33 ± 0.09	81.44 ± 0.12
	wDAE [9]	61.07 ± 0.15	76.75 ± 0.11	68.18 ± 0.16	83.09 ± 0.12
	CC+rot [7]	62.93 ± 0.45	79.87 ± 0.33	70.53 ± 0.51	84.98 ± 0.36
	Robust-dist++ [16]	63.28 ± 0.62	81.17 ± 0.43	–	–
	Transductive-ft [4]	65.73 ± 0.68	78.40 ± 0.52	73.34 ± 0.71	85.50 ± 0.50
	our baseline (arcmax)	63.28 ±0.71	78.31 ±0.57	68.47 ±0.86	84.11 ±0.65
	adversarial alignment	64.79 ±0.93	82.02 ±0.88	73.87 ±0.76	84.95 ±0.59
	centroid alignment	**65.92** ± 0.60	**82.85** ± 0.55	**74.40** ± 0.68	**86.61** ±0.59

Evaluation of 5-way classification accuracy, with ± indicating the 95% confidence intervals over 600 episodes. The best result is boldfaced, while the best result *prior to this work* is highlighted in blue. Here,“–” indicates when a paper does not report results in the corresponding scenario.

Results: Cross-Domain

- Table bellow presents the evaluation on cross-domain from *mini*-ImageNet to CUB-200-2011 (CUB).

Method	1-shot	5-shot	10-shot
ProtoNet‡ [23]	–	62.02 ± 0.70	–
MAML‡ [6]	–	51.34 ± 0.72	–
Diverse 20 [5]	–	66.17 ± 0.73	–
cosmax† [2]	43.06 ± 1.01	64.38 ± 0.86	67.56±0.77
our baseline (arcmax)	45.60 ± 0.94	64.93 ± 0.95	68.95±0.78
adversarial	44.37 ± 0.94	70.80 ± 0.83	79.63 ±0.71
centroid	46.85 ± 0.75	70.37 ± 1.02	**79.98** ±0.80

The best result is boldfaced, while the best result *prior to this work* is highlighted in blue.

This project was supported by funding from NSERC-Canada, Mitacs, Prompt-Québec, and E Machine Learning. Project page: https://lvsn.github.io/associative-alignment/