When Does Self-Supervision Improve Few-Shot Learning?

Jong-Chyi Su, Subhransu Maji, Bharath Hariharan

ECCV 2020

https://arxiv.org/pdf/1910.03560.pdf

Outline

- Investigate the role of Self-Supervision Learning (SSL) in Few-Shot Learning (FSL)
- Findings
  - SSL improves the few-shot learner even when the datasets are **small** and **only** use images within the dataset
  - SSL can hurt when the distribution of images using for meta-learning and SSL are different
- Propose a simple approach to classify unlabeled data to pick data from similar domains

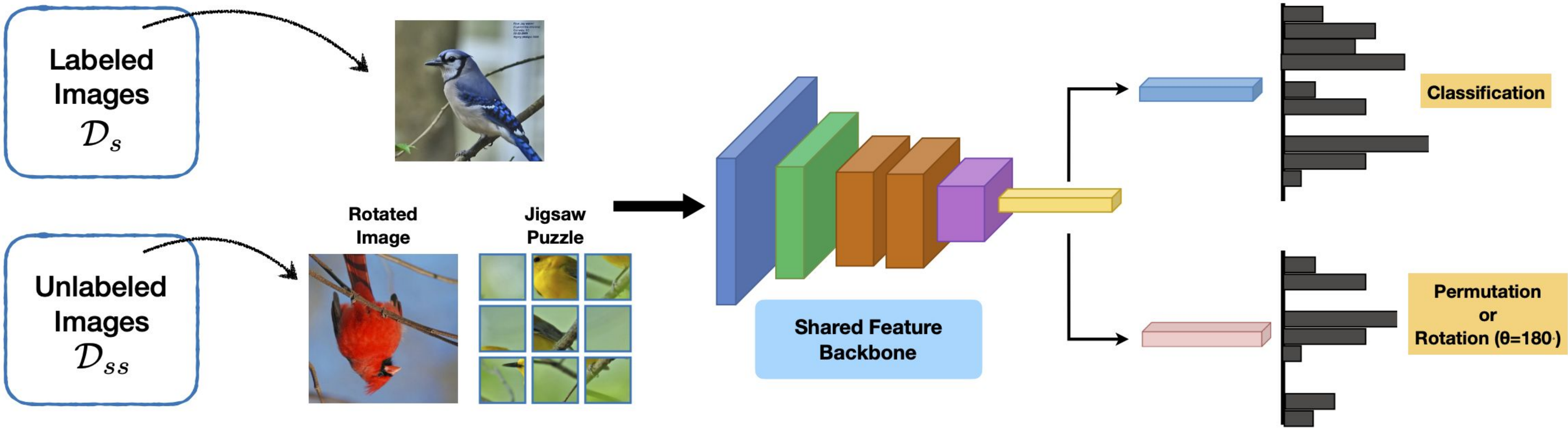


Fig. 1: **Combining supervised and self-supervised losses for few-shot learning.** Self-supervised tasks such as jigsaw puzzle or rotation prediction act as a data-dependent regularizer for the shared feature backbone. Our work investigates how the performance on the *target task domain* ($\mathcal{D}_s$) is impacted by the choice of the *domain used for self-supervision* ($\mathcal{D}_{ss}$).

Few-shot with Self-supervision

Few-shot loss

$$\mathcal{L}_s := \sum_{(x_i, y_i) \in \mathcal{D}_s} \ell\big(g \circ f(x_i), y_i\big) + \mathcal{R}(f, g).$$

Self-supervision loss

$$\mathcal{L}_{ss} := \sum_{x_i \in \mathcal{D}_{ss}} \ell\big(h \circ f(\hat{x}_i), \hat{y}_i\big).$$

Data: (x_i, y_i)

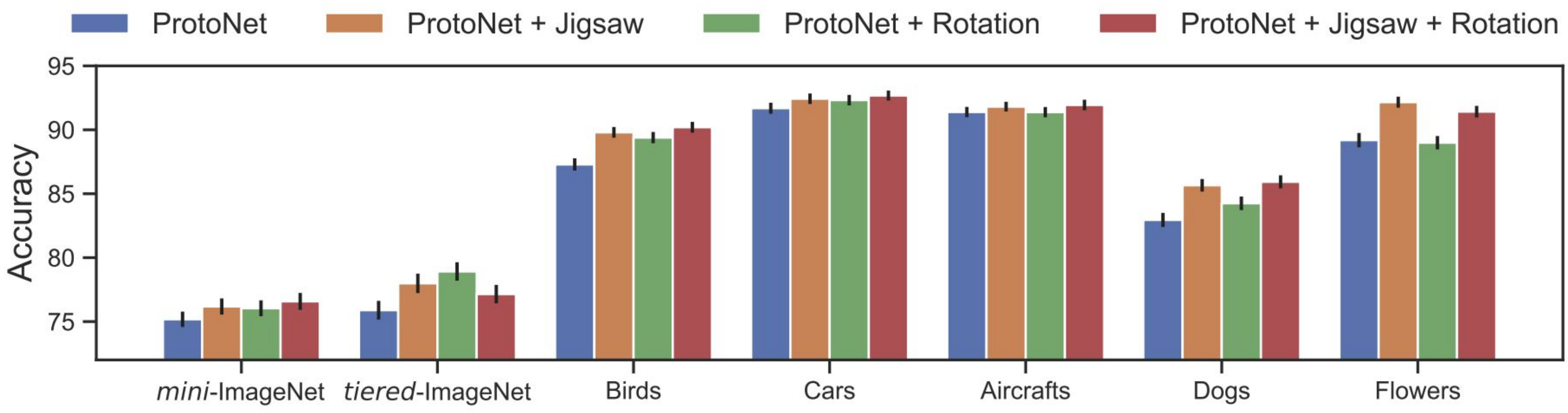
Augmented data: (x_hat_i, y_hat_i)

Encoder: f

FSL classifier: g

Self-supervision

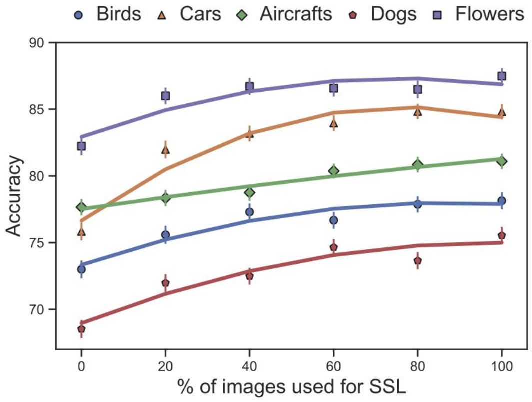
- Jigsaw puzzle task loss:
    - The input x is tiled into 3x3 regions, then permuted to obtain x_hat
    - The label y_hat is the index of the permutation.
    - The y_hat is then grouped by hamming distance to control the difficulty of the task
- Rotation task loss:
    - The image is rotated by angle of 0, 90, 180, 270 degree



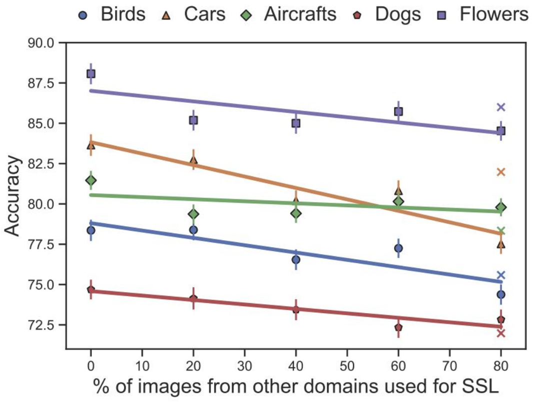
Loss	Birds	Cars	Aircrafts	Dogs	Flowers
	5-way 5-shot				
Softmax	81.5±0.5	87.7±0.5	89.2±0.4	77.6±0.6	91.0±0.5
Softmax + Jigsaw	83.9±0.5	90.6±0.5	89.6±0.4	77.8±0.6	91.1±0.5
MAML	81.2±0.7	86.9±0.6	88.8±0.5	77.3±0.7	79.0±0.9
MAML + Jigsaw	81.1±0.7	89.0±0.5	89.1±0.5	77.3±0.7	82.6±0.7
ProtoNet	87.3±0.5	91.7±0.4	91.4±0.4	83.0±0.6	89.2±0.6
ProtoNet + Jigsaw	**89.8±0.4**	**92.4±0.4**	**91.8±0.4**	**85.7±0.5**	**92.2±0.4**

Domain shift

(a) SSL and FSL are trained on the same domain
(b) SSL and FSL are trained on different domains



(a) Effect of number of images on SSL.



(b) Effect of domain shift on SSL.

Data selection for SSL

- Using a “domain weighted” model to select the top images based on a domain classifier.
- Binary classifier with positive sample from training set and negative from unlabeled data
-

- No SSL: FSL only on 20% dataset (lower bound)
- SSL 20% dataset: FSL + SSL on 20% dataset
- SSL pool random: randomly select from the pool
- SSL pool weight: using binary classifier
- Oracle: FSL on 20% and SSL on 100% dataset