UCF CENTER FOR RESEARCH IN COMPUTER VISION

ICLR

IN DEFENSE OF PSEUDO-LABELING: AN UNCERTAINTY-AWARE PSEUDO-LABEL SELECTION FRAMEWORK FOR SEMI-SUPERVISED LEARNING

Mamshad Nayeem Rizve

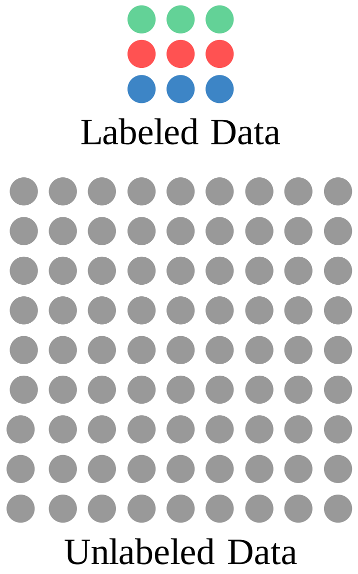
Kevin Duarte

Yogesh S Rawat

Mubarak Shah

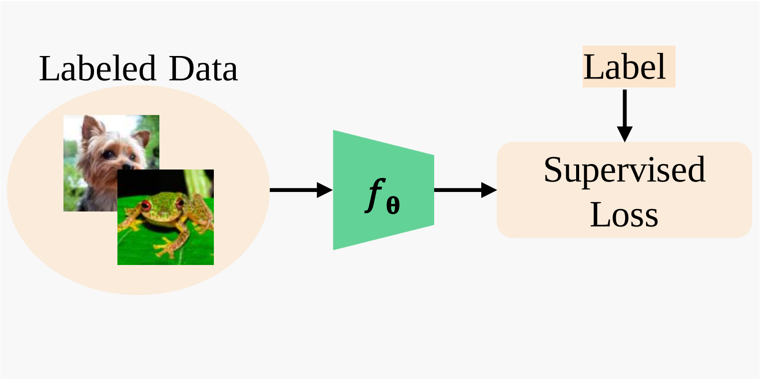
Semi-Supervised Learning

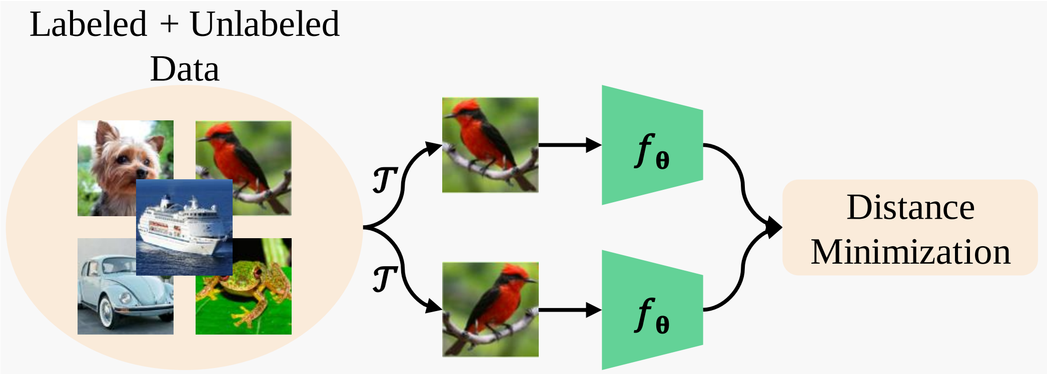
- Supervised learning relies on large labeled datasets
- Constructing large labeled datasets
  - expensive
  - time-consuming
- SSL leverages
  - a small amount of labeled data
  - a large amount of unlabeled data concurrently
- One of the fundamental problems in machine learning



Dominant SSL Approaches

- Consistency Regularization [1, 2, 3]
  - obtain perturbation/augmentation invariant output distribution
  - rely on domain-specific heavy data augmentation
  - limited applicability on domains which do not have a rich set of augmentations

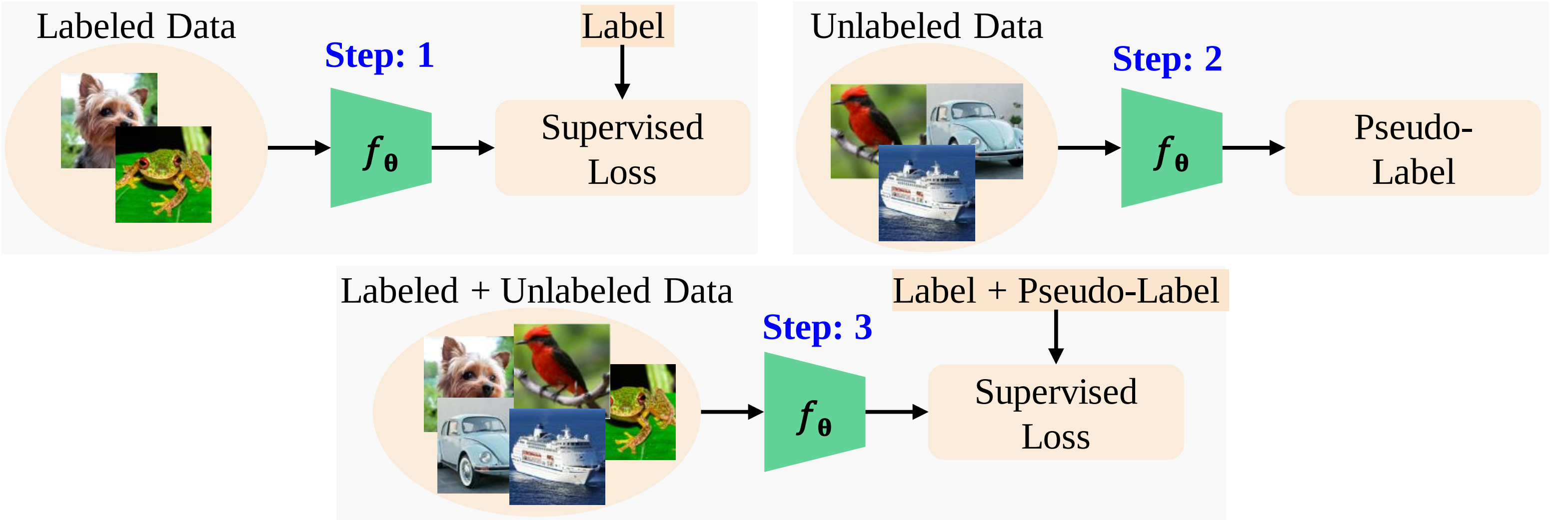




1. *Regularization with stochastic transformations and perturbations for deep semi-supervised learning*; Sajjadi et al.; Neurips 2016
2. *Interpolation consistency training for semi-supervised learning*; Verma et al.; IJCAI 2019
3. *Mixmatch: A holistic approach to semi-supervised learning*; Berthelot et al.; Neurips 2019

Dominant SSL Approaches

- Pseudo-Labeling [1]
  - generate pseudo-labels for unlabeled samples
  - does not require domain-specific data augmentation
  - performs poorly in comparison to consistency regularization



1. *Pseudo-label : The simple and efficient semi-supervised learning method for deep neural networks*; Lee; ICML Workshop 2013

Objective

Bridge the performance gap between

Pseudo-Labeling

and

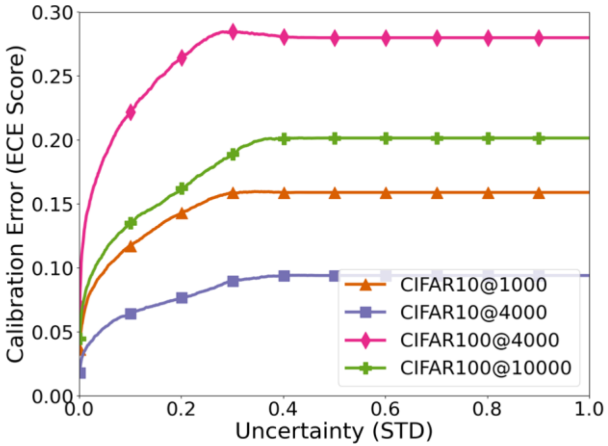
Consistency Regularization

Fundamental Issues with Pseudo-Labeling

- Training with small labeled set
  - leads to erroneous pseudo-label generation
  - many incorrect pseudo-labels leads to noisy training
- Incorrect pseudo-labels must be discarded
  - use high-confidence pseudo-labels for training

Pseudo-Label Selection

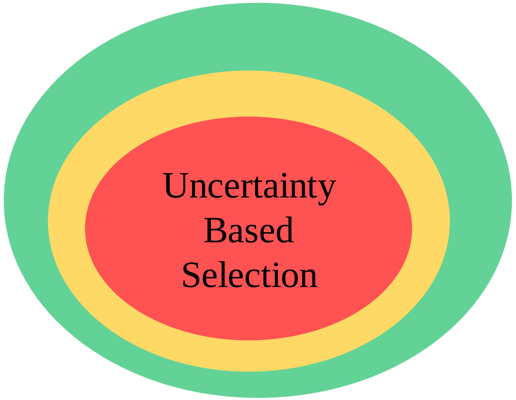
- Using highly confident pseudo-labels is insufficient
  - neural networks suffer from poor calibration [1]
  - many incorrect pseudo-labels are still selected
- Another interpretation of calibration
  - a notion of network's overall prediction uncertainty [2]
- We have empirically analyzed
  - the relationship between the calibration error and
  - individual output prediction uncertainties



1. *On Calibration of Modern Neural Networks*; Guo et al.; ICML 2017
2. *Simple and scalable predictive uncertainty estimation using deep ensembles*; Lakshminarayanan et al.; Neurips 2017

Pseudo-Label Selection

- Based on our observations
  - we select a subset of generated pseudo-labels with the following two criteria
    - the confidence of a prediction has to be high
    - the network has to be certain about the output prediction
- We call this method Uncertainty-Aware Pseudo-Label Selection (UPS)

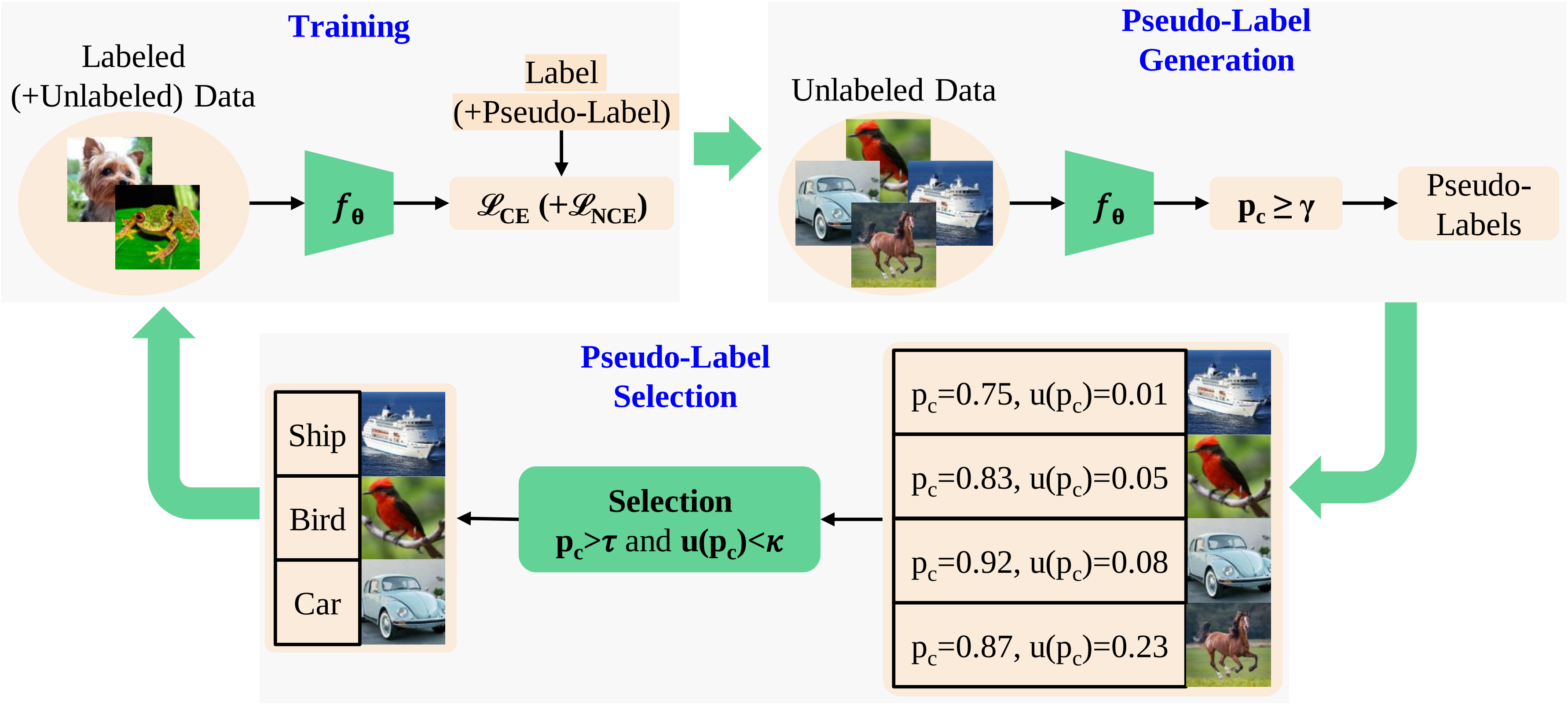


Pseudo-Label Selection

- Networks can be confident and certain about a class **not** being present
- We can use this information and select negative pseudo-labels
- This gives us two benefits:
  - negative learning for single label classification
  - allows for multi-label classification

Horse	✖
Dog	
Truck	✖
Cat	

UPS Method



Training
Labeled (+Unlabeled) Data
Label (+Pseudo-Label)
f_θ
$\mathscr{L}_{CE}$ (+$\mathscr{L}_{NCE}$)
Pseudo-Label Generation
Unlabeled Data
f_θ
$p_c \geq \gamma$
Pseudo-Labels
Pseudo-Label Selection
Ship
Bird
Car
Selection
$p_c > \tau$ and $u(p_c) < \kappa$
p_c=0.75, $u(p_c)$=0.01
p_c=0.83, $u(p_c)$=0.05
p_c=0.92, $u(p_c)$=0.08
p_c=0.87, $u(p_c)$=0.23

Results (CIFAR-10 and CIFAR-100)

Error rate (%) on the CIFAR-10 and CIFAR-100 test set:

Method	CIFAR-10		CIFAR-100	
	1000 labels	4000 labels	4000 labels	10000 labels
DeepLP†	22.02 ± 0.88	12.69 ± 0.29	46.20 ± 0.76	38.43 ± 1.88
TSSDL†	21.13 ± 1.17	10.90 ± 0.23	-	-
MT	19.04 ± 0.51	11.41 ± 0.25	45.36 ± 0.49	36.08 ± 0.51
MT + DeepLP	16.93 ± 0.70	10.61 ± 0.28	43.73 ± 0.20	35.92 ± 0.47
ICT	15.48 ± 0.78	7.29 ± 0.02	-	-
DualStudent	14.17 ± 0.38	8.89 ± 0.09	-	32.77 ± 0.24
R2-D2	-	-	-	32.87 ± 0.51
MixMatch	-	6.84	-	-
UPS†	$\mathbf{8.18 \pm 0.15}$	$\mathbf{6.39 \pm 0.02}$	$\mathbf{40.77 \pm 0.10}$	$\mathbf{32.00 \pm 0.49}$

Results (UCF-101 and Pascal VOC2007)

Accuracy (%) on the UCF-101 test set:

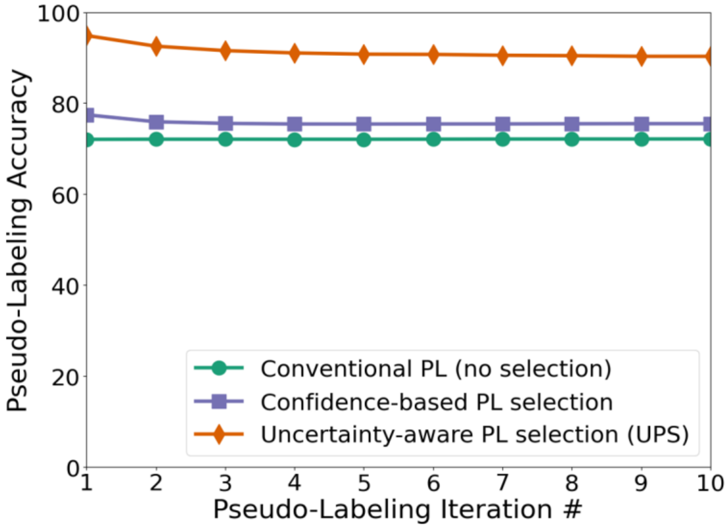
Method	20% labeled	50% labeled
Supervised	33.5	45.6
MT*	36.3	45.8
PL*	37.0	47.5
S4L*	37.7	47.9
UPS	**39.4**	**50.2**

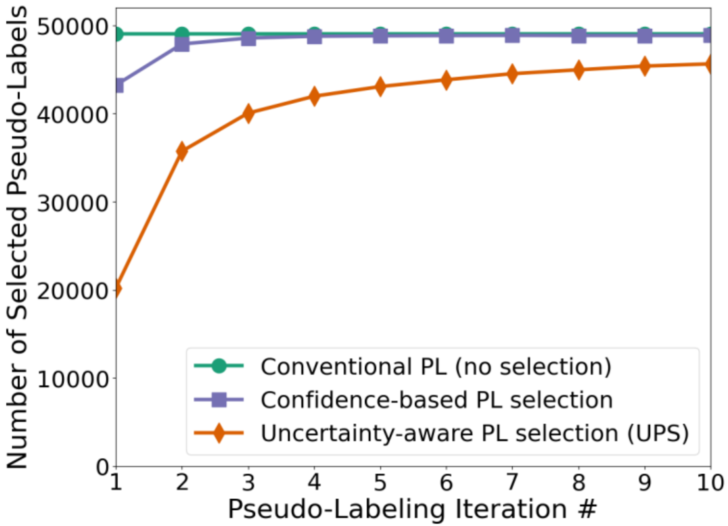
mAP scores on the Pascal VOC2007 test set:

Method	10% labeled	20% labeled
Supervised	18.36 ± 0.65	28.84 ± 1.68
PL	27.44 ± 0.55	34.84 ± 1.88
MixMatch	29.57 ± 0.78	37.02 ± 0.97
MT	32.55 ± 1.48	39.62 ± 1.66
UPS	$\mathbf{34.22 \pm 0.79}$	$\mathbf{40.34 \pm 0.08}$

Analysis (PL Accuracy)

UPS achieves higher pseudo-label accuracy while selecting similar number of pseudo-labels



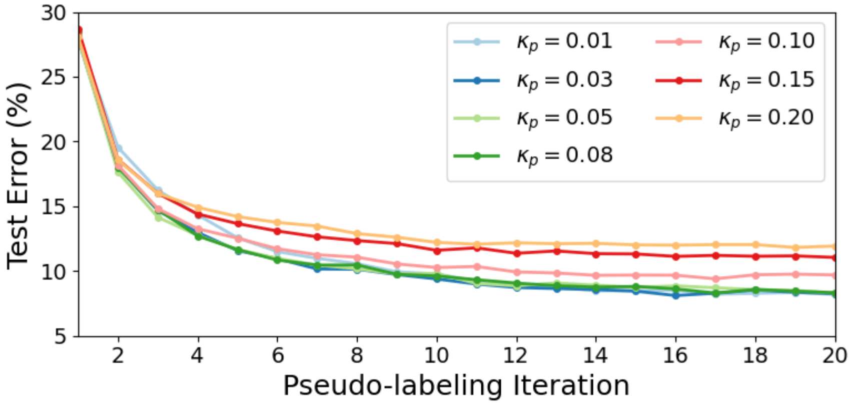


Analysis (Compatibility and Robustness)

UPS is compatible with most uncertainty estimation methods

Method	1000 labels	4000 labels
MC-Dropout	8.14	6.36
MC-SpatialDropout	8.28	6.60
MC-DropBlock	9.76	7.50
DataAug	8.28	6.72

UPS is robust to pseudo-label selection hyperparameters



Conclusion

- We have introduced UPS,
  - a simple and efficient framework for effective pseudo-labeling based SSL
- UPS competes with SOTA consistency regularization based methods
  - without inherently relying on strong data augmentation
- We are first to propose negative pseudo-labeling for SSL
- UPS is versatile:
  - it is domain agnostic
  - can be easily used for multi-label classification

Thank You