Seminar in AI (Master)

Realistic Evaluation of Deep Semi-Supervised Learning Algorithms [1]

Pseudo-Label : The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks [2]

Jaechan Park / k11850713

Supervisor - Dr. Andreu Vall

JKU

[1] Oliver et al. NeurIPS 2018 Proceedings. https://arxiv.org/abs/1804.09170
[2] Dong-Hyun Lee, ICML 2013 Workshop : Challenges in Representation Learning (WREPL)

Motivation

- **Deep Neural Network - supervised learning method**
  - highly accurate and trustworthy method
  - sufficient large labeled dataset required
  - most real-world dataset is unlabeled
- **Lots of human effort**
- **Expensive**
- **beneficial to leverage unlabeled dataset in supervised learning method**

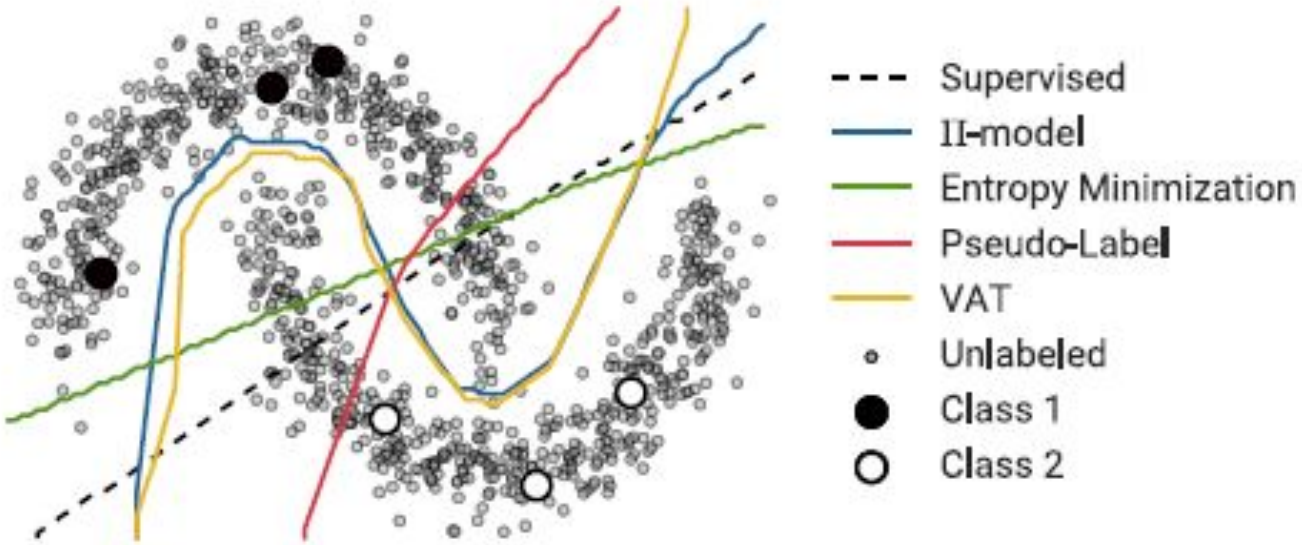
→ **Semi-Supervised Learning (SSL)**

Suggestive Evaluation Methodologies

- Question - applicable to “real-world” settings?
- Proposes better approaches to measure applicability
  - A shared implementation (standardized)
  - High-quality fully-supervised baseline
  - Compare to “Transfer” Learning
  - Reflect on class distribution mismatch
  - Different size of labeled and unlabeled data
  - practical small validation set

SSL Methods - Decision Boundary

- “two moons” dataset
- A MLP with three hidden layers / 10 ReLU units
- Omit “Mean Teacher” and “Temporal Ensembling”



Experiments

- A common image classification model
  - WRN-28-2 (ResNet with depth 28 and width 2, batch normalization and leakyReLU nonlinearities)
  - Adam optimizer
- Dataset
  - SVHN - converted floating point values [-1,1]
  - data augmentation - random translation by up to 2 pixels
  - 65,932 train / 7,325 validation (64,932 unlabeled images)
  - CIFAR-10 - global contrast and Zero component analysis (ZCA) normalization applied
  - data augmentation - random horizontal flipping, translation by up to 2 pixels, Gaussian input noise with standard deviation 0.15
  - 45,000 train / 5,000 validation (41,000 unlabeled images)
- Trained all network for 500,000 updates / 100 batch size / no early stopping

Experiments

- Reproduction - CIFAR-10 (4,000 labeled) and SVHN (1,000 labeled)
  - at the lowest validation error

Dataset	# Labels	Supervised	Π-Model	Mean Teacher	VAT	VAT+EntMin	Pseudo-Label
CIFAR-10	4000	20.26±.38%	16.37±.63%	15.87±.28%	13.86±.27%	13.13±.39%	17.78±.57%
SVHN	1000	12.83±.47%	7.19±.27%	5.65±.47%	5.63±.20%	5.35±.19%	7.62±.29%

- Change of error rate (fully-supervised → semi-supervised)

Method	CIFAR-10 4000 Labels	SVHN 1000 Labels
Π-Model[3]	34.85% → 12.36%	19.30% → 4.80%
Π-Model[4]	13.60% → 11.29%	-
Π-Model	20.26% → 16.37%	12.83% → 7.19%
Mean Teacher [5]	20.66% → 12.31%	12.32% → 3.95%
Mean Teacher	20.26% → 15.87%	12.83% → 5.65%

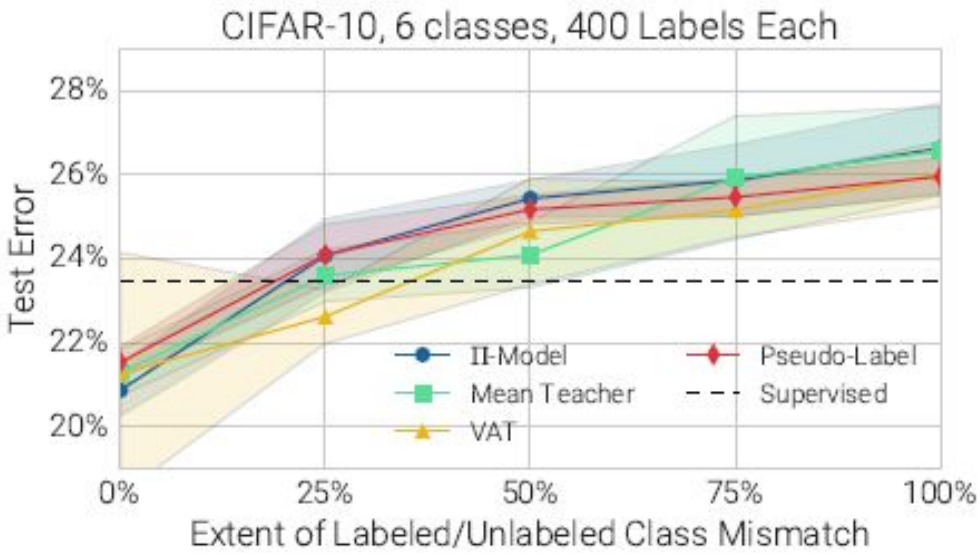
[3]Laine, Samuli and Aila, Timo. Temporal ensembling for semi-supervised learning. In Fifth International Conference on Learning Representations, 2017.
[4] Sajjadi, Mehdi, Javanmardi, Mehran, and Tasdizen, Tolga. Regularization with stochastic transformations and perturbations for deep semi-supervised learning. In Advances in Neural Information Processing Systems, 2016.
[5]Tarvainen, Antti and Valpola, Harri. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. Advances in Neural Information Processing Systems, 2017.

Experiments

- Error rate using SSL and transfer learning
  - pre-trained by ImageNet downsampled to 32x32
  - challenge to achieve a convincing results with SVHN dataset

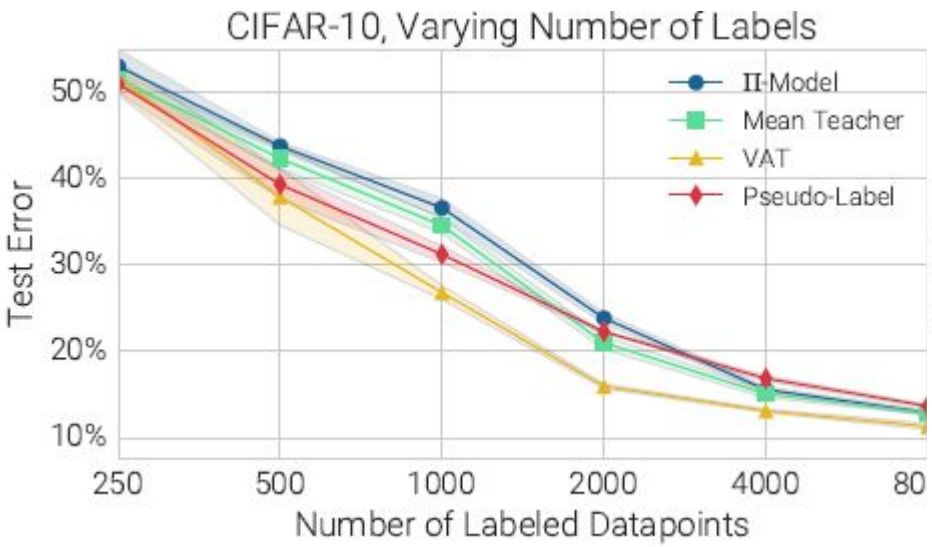
Method	CIFAR-10 (4000 Labels)
VAT with Entropy Minimization	13.13%
ImageNet → CIFAR-10	12.09%
ImageNet → CIFAR-10 (no overlap)	12.91%

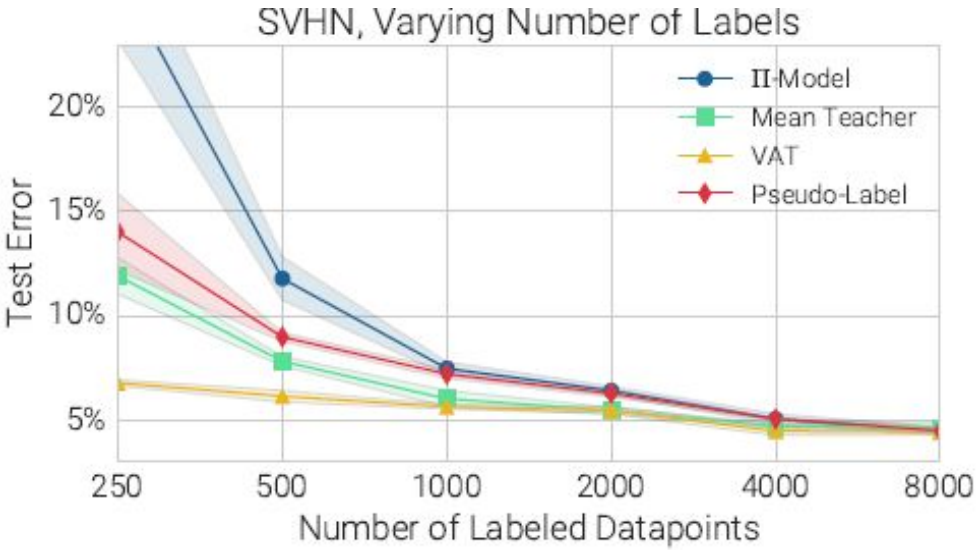
- Class Distribution Mismatch
  - 6 classes - labeled
  - max 4 classes mismatch



Experiments

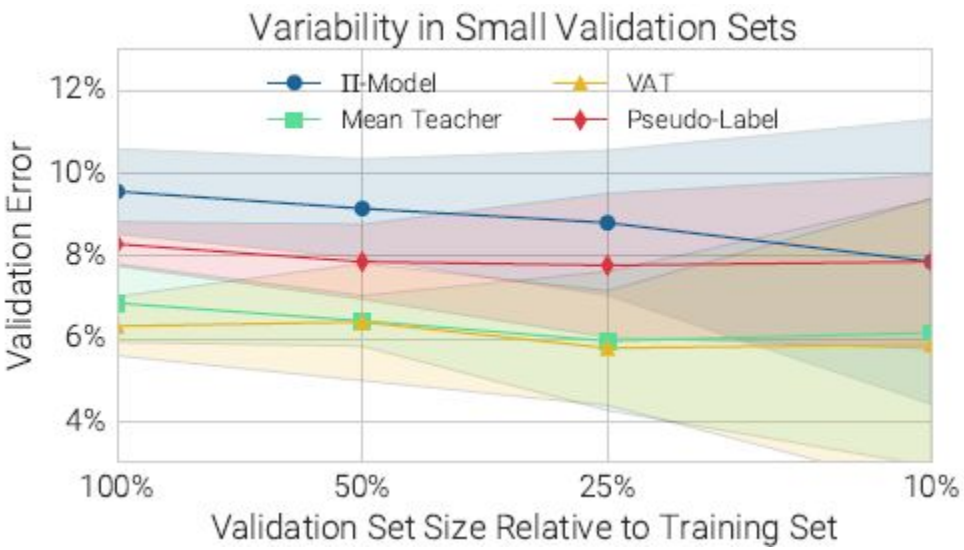
- Vary number of labeled data samples

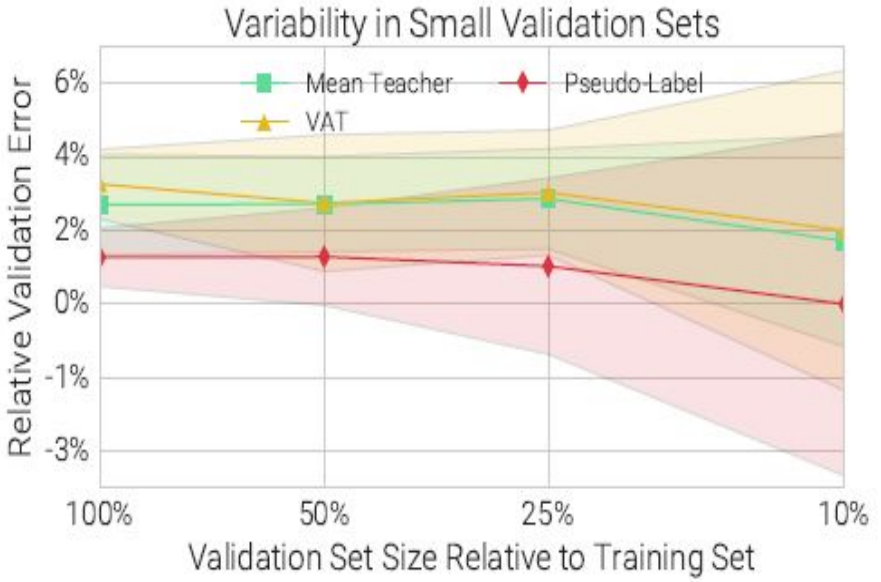




Experiments

- Set varying size of validation sets of SVHN
  - same size (100%) to the size of training set
  - mean and standard deviation - difference in validation error (each SSL method and Π-model





SSL Method : Pseudo-Label

- (most) Deep Neural Networks - mainly consists of two phases
  - unsupervised pre-training
  - fine-tuning (supervised)
- The two phases can be applied to semi-supervised learning fashion
- “Pseudo-Label” - a simpler technique of semi-supervised learning
- Picking up the class based on the maximum predicted probability of every weight update
- Train in using a supervised learning method with labeled and unlabeled data at the same time
- Use Denoising Auto-Encoder (DAE) and Dropout for the sake of performance improvement
- In effect equivalent to Entropy Regularization
- Minimized entropy for unlabeled data → reduce the overlapping of class probability distribution

SSL Method : Pseudo-Label

- Enable to assign the class based on maximum predicted probability for each unlabeled sample
- Assigned “fake” labels of unlabeled samples behavior as “true” labels in supervised learning

$$y_i' = \begin{cases} 1 & if\ i = argmax_{i'} f_{i'}(x) \\ 0 & otherwise \end{cases}$$

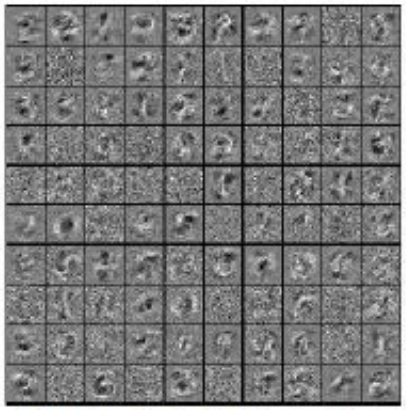
Experiment

- Dataset - MNIST
- For comparison, set the same semi-supervised learning setting [5]
- Set different labeled training set to 100,600,1000, and 3000
- 1000 validation set separately for hyperparameter setting
- 1 hidden layer
- rectified linear unit (ReLU) for hidden unit / 5000 units
- sigmoid unit for output unit

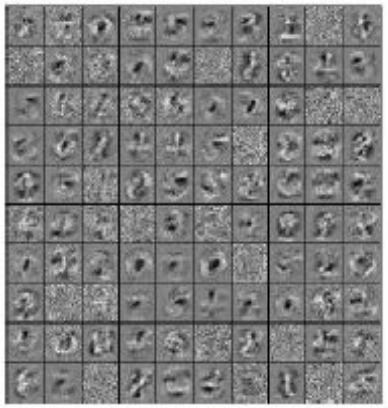
[5] Rifai, S., Dauphin, Y., Vincent, P., Bengio, Y., and Muller, X. The manifold tangent classifier. Advances in Neural Information Processing Systems, 2011, 24: 2294-2302, Weston, J., Ratle, F., and Collobert, R. Deep learning via semi-supervised embedding. In Proceedings of the 25th international conference on Machine learning. ACM, 2008. p. 1168-1175

Results

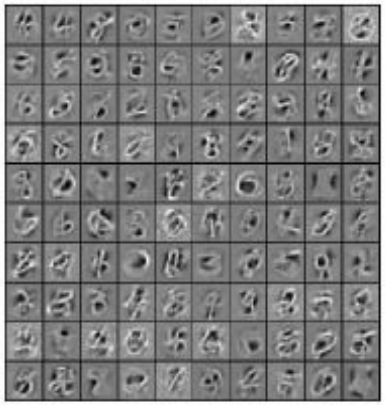
- 600 MNIST labeled data - supervised training



without unlabeled data (dropNN)



with unlabeled data and Pseudo-Label (+PL)



using unsupervised pre-training with DAE (+PL+DAE)

Results

Method	100	600	1000	3000
NN	25.81	11.44	10.7	6.04
SVM	23.44	8.85	7.77	4.21
CNN	22.98	7.68	6.45	3.35
TSVM	16.81	6.16	5.38	3.45
DBN-$_{R}$NCA	-	8.7	-	3.3
$_{EMBED}$NN	16.86	5.97	5.73	3.59
CAE	13.47	6.3	4.77	3.22
MTC	12.03	5.13	3.64	**2.57**
$_{DROP}$NN	21.89	8.57	6.59	3.72
+PL	16.15	5.03	4.30	2.80
+PL+DAE	**10.49**	**4.01**	**3.46**	2.69

Discussion

- **Realistic Evaluation of Deep Semi-Supervised Learning Algorithm**
  - Still, limited ability for direct comparison with the past work
- **Pseudo-Label : The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks**
  - In the conclusion section, there is no discussion about the results between Pseudo-label and Pseudo-Label with DAE

Thank you!