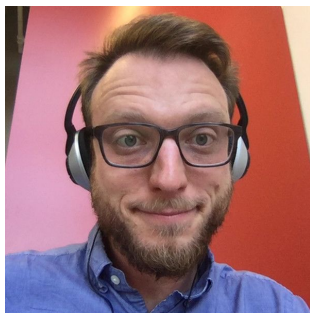


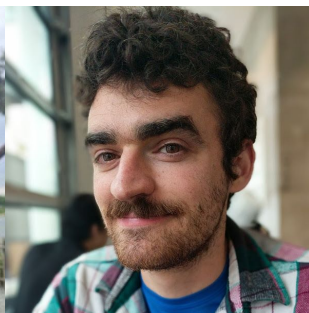
Realistic Evaluation of Deep Semi-Supervised Learning Algorithms



Avital Oliver*



Augustus Odena*



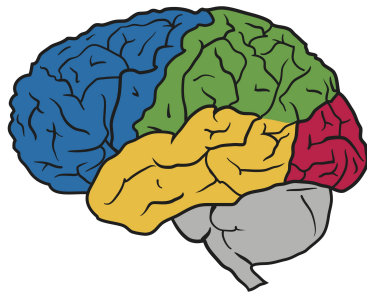
Colin Raffel*

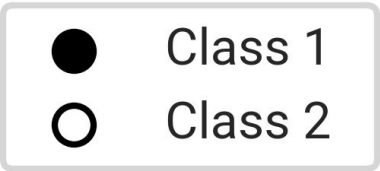
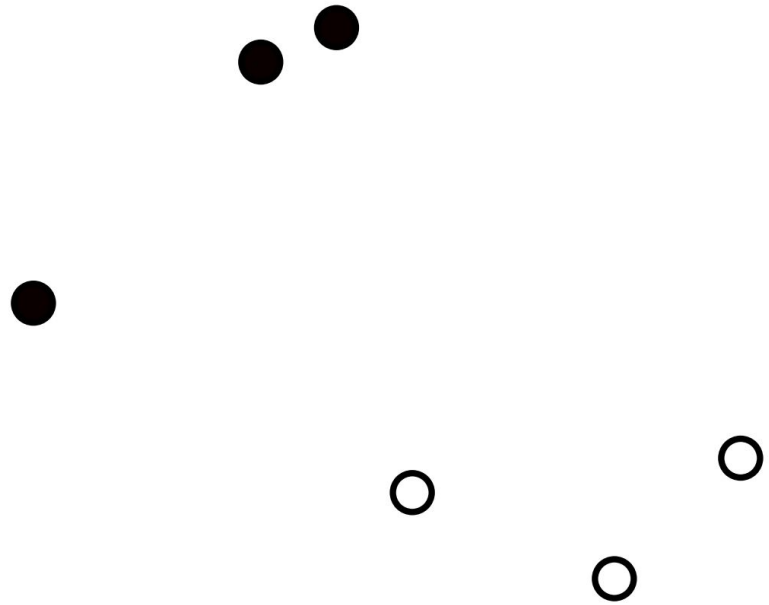


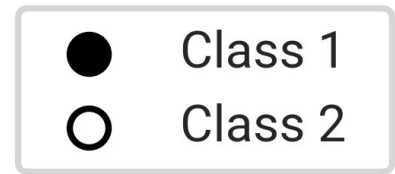
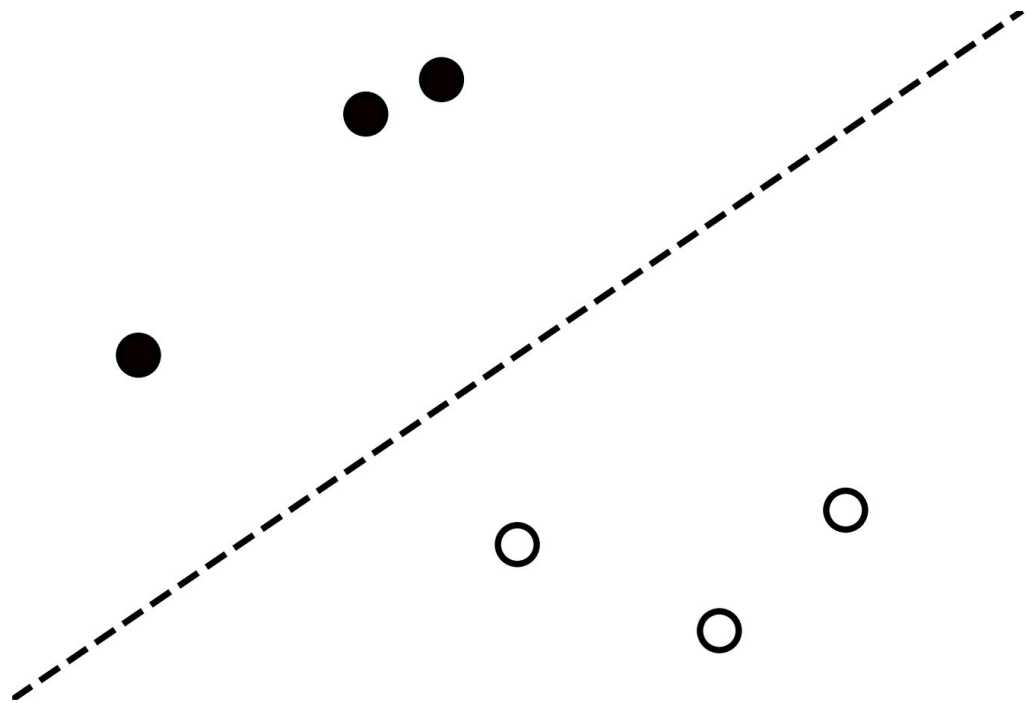
Ekin Dogus Cubuk

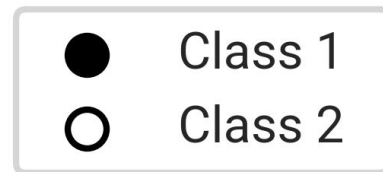
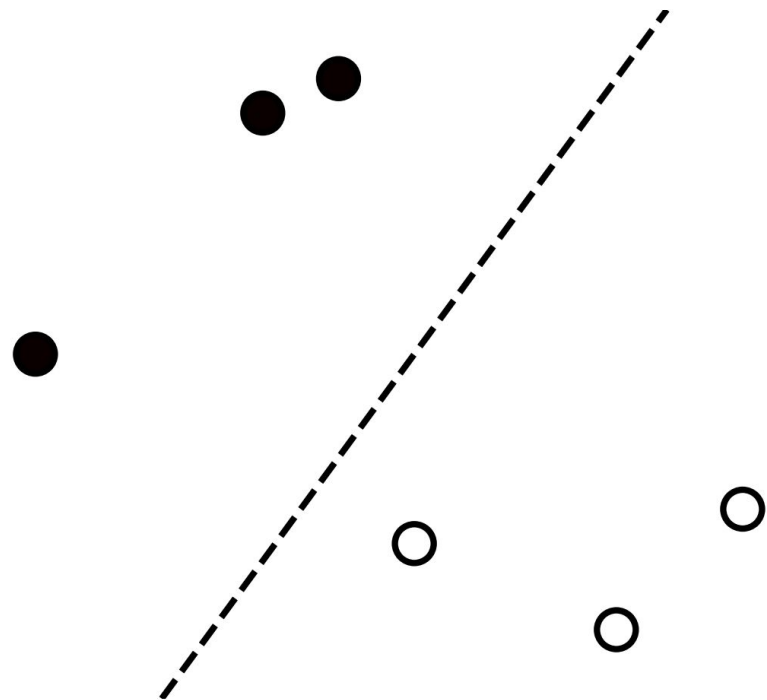


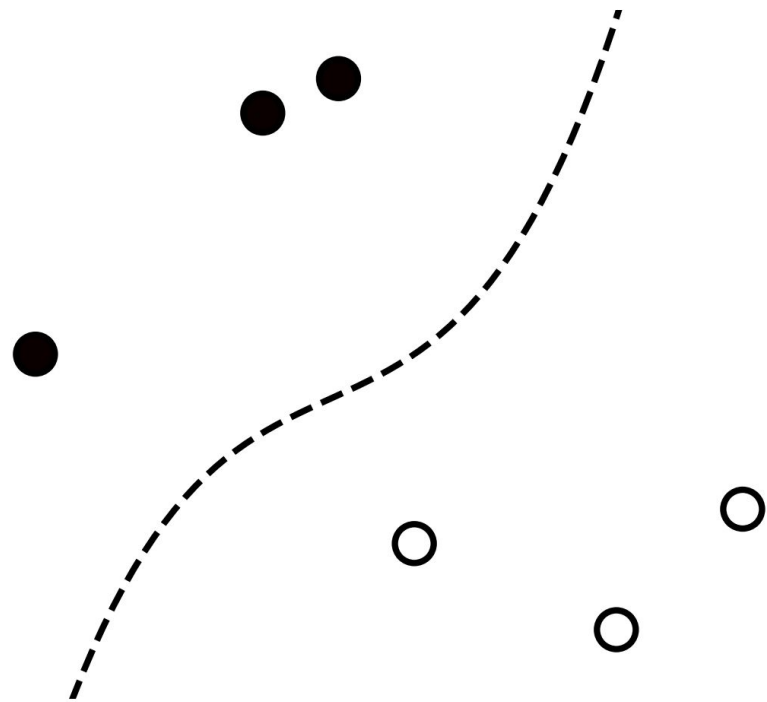
Ian Goodfellow

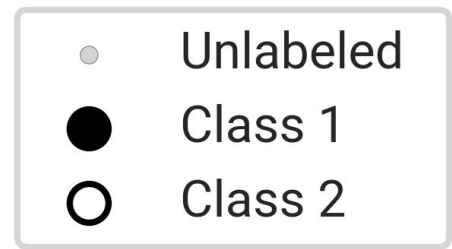
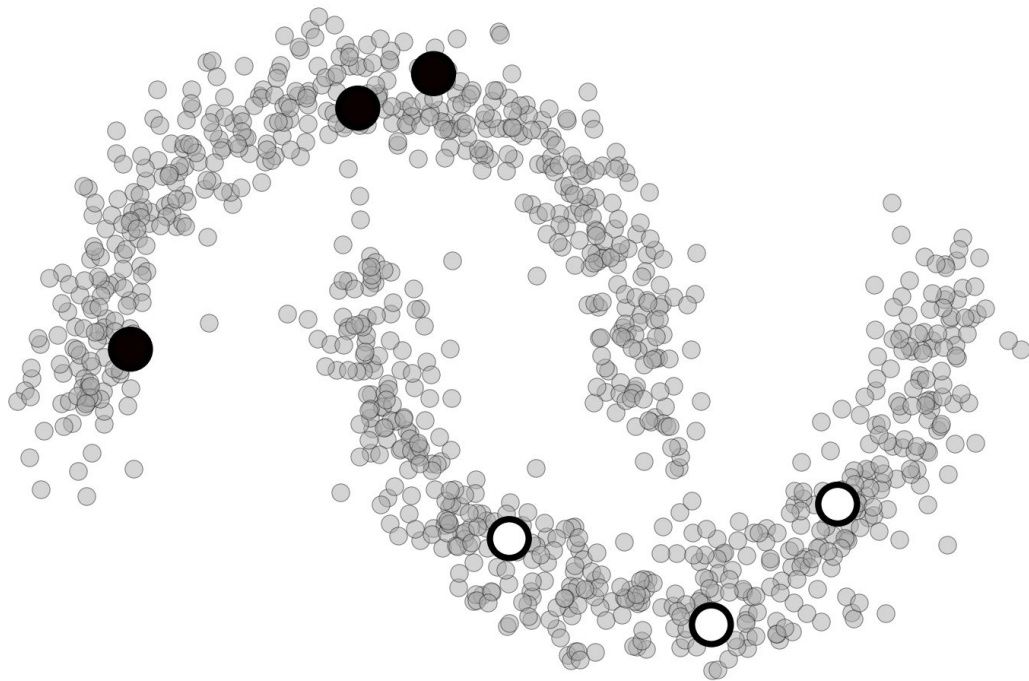


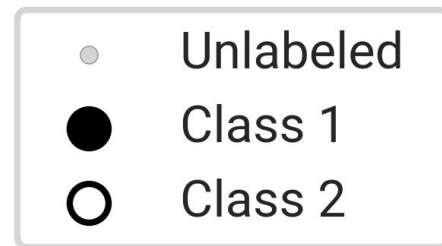
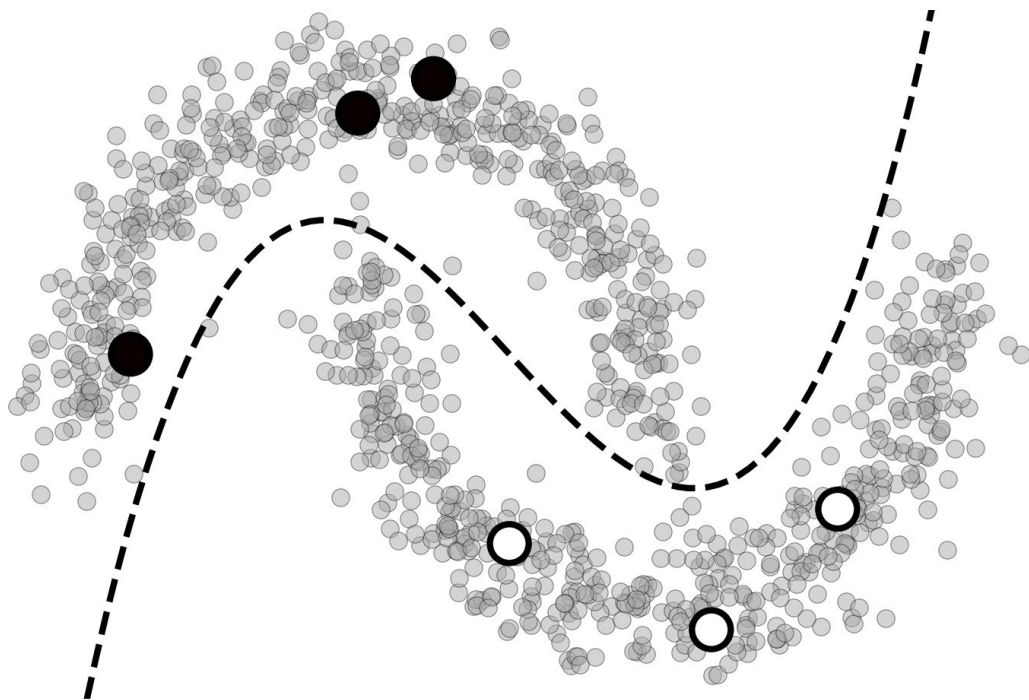












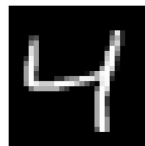
Label: 5



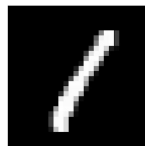
Label: 0



Label: 4



Label: 1



Label: 9



Label: 2



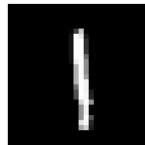
Label: 1



Label: 3



Label: 1



Label: 4



Label: 3



Label: 5



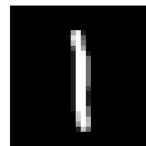
Label: 3



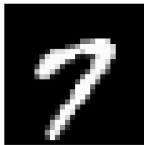
Label: 6



Label: 1



Label: 7



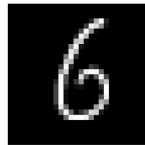
Label: 2



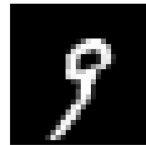
Label: 8



Label: 6



Label: 9



Label: 4



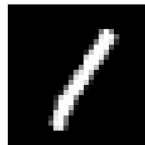
Label: 0



Label: 9



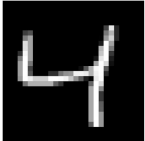
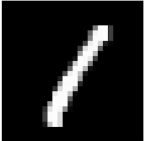
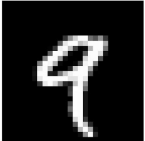

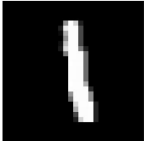

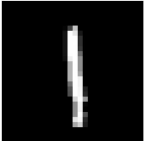
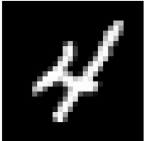

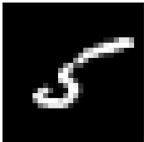


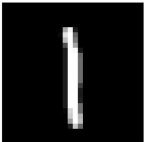
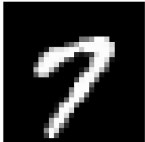

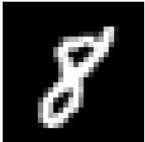
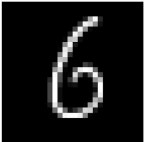
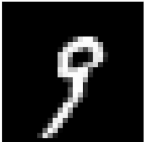



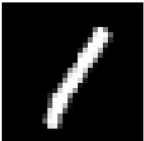
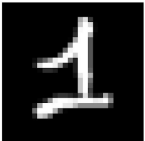


Label: 1



Label: 1



Label: 5 	Label: 0 	Label: 4 	Label: 1 	Label: 9 
Unlabeled 	Unlabeled 	Unlabeled 	Unlabeled 	Unlabeled 
Unlabeled 	Unlabeled 	Unlabeled 	Unlabeled 	Unlabeled 
Unlabeled 	Unlabeled 	Unlabeled 	Unlabeled 	Unlabeled 
Unlabeled 	Unlabeled 	Unlabeled 	Unlabeled 	Unlabeled 

Label: 5



Label: 0



Label: 4



Label: 1



Label: 9



VS.

Label: 5



Label: 0



Label: 4



Label: 1



Label: 9



Unlabeled



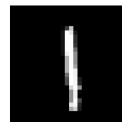
Unlabeled



Unlabeled



Unlabeled



Unlabeled



Unlabeled



Unlabeled



Unlabeled



Unlabeled



Unlabeled



Unlabeled



Unlabeled



Unlabeled



Unlabeled



Unlabeled



Unlabeled



Unlabeled



Unlabeled



Unlabeled



Unlabeled



Method	CIFAR-10 4000 Labels	SVHN 1000 Labels
Supervised	20.26 \pm 0.38%	12.83 \pm 0.47%
Π -Model	16.37 \pm 0.63%	7.19 \pm 0.27%
Mean Teacher	15.87 \pm 0.28%	5.65 \pm 0.47%
VAT	13.86 \pm 0.27%	5.63 \pm 0.20%
VAT + EntMin	13.13 \pm 0.39%	5.35 \pm 0.19%
Pseudo-Label	17.78 \pm 0.57%	7.62 \pm 0.29%

Error rates for our reimplementation of different semi-supervised learning methods.

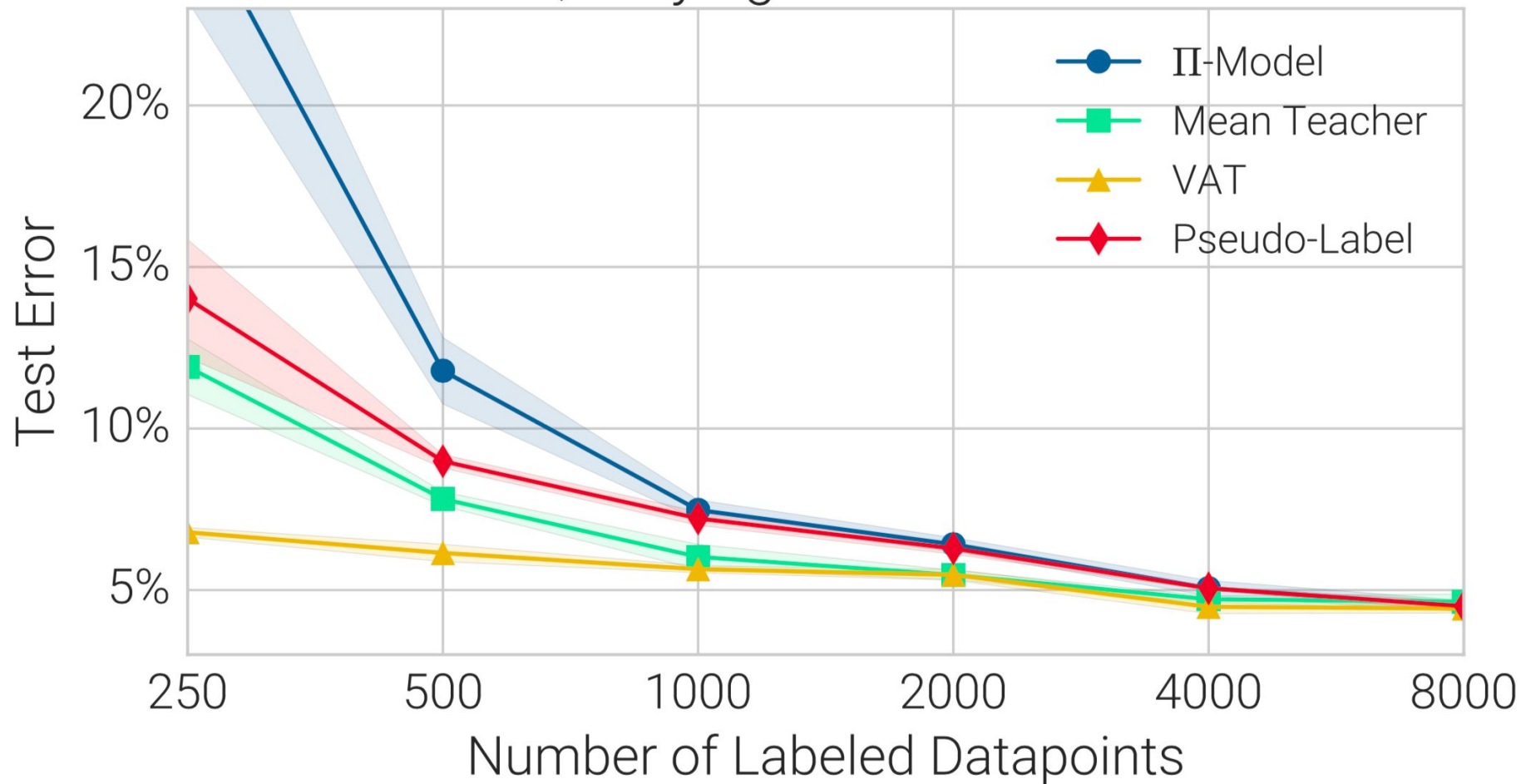
Method	CIFAR-10 4000 Labels	SVHN 1000 Labels
Π -Model [32]	34.85% \rightarrow 12.36%	19.30% \rightarrow 4.80%
Π -Model [46]	13.60% \rightarrow 11.29%	–
Π -Model (ours)	20.26% \rightarrow 16.37%	12.83% \rightarrow 7.19%
Mean Teacher [50]	20.66% \rightarrow 12.31%	12.32% \rightarrow 3.95%
Mean Teacher (ours)	20.26% \rightarrow 15.87%	12.83% \rightarrow 5.65%

Reported change in error rate from fully-supervised (no unlabeled data) to semi-supervised learning.

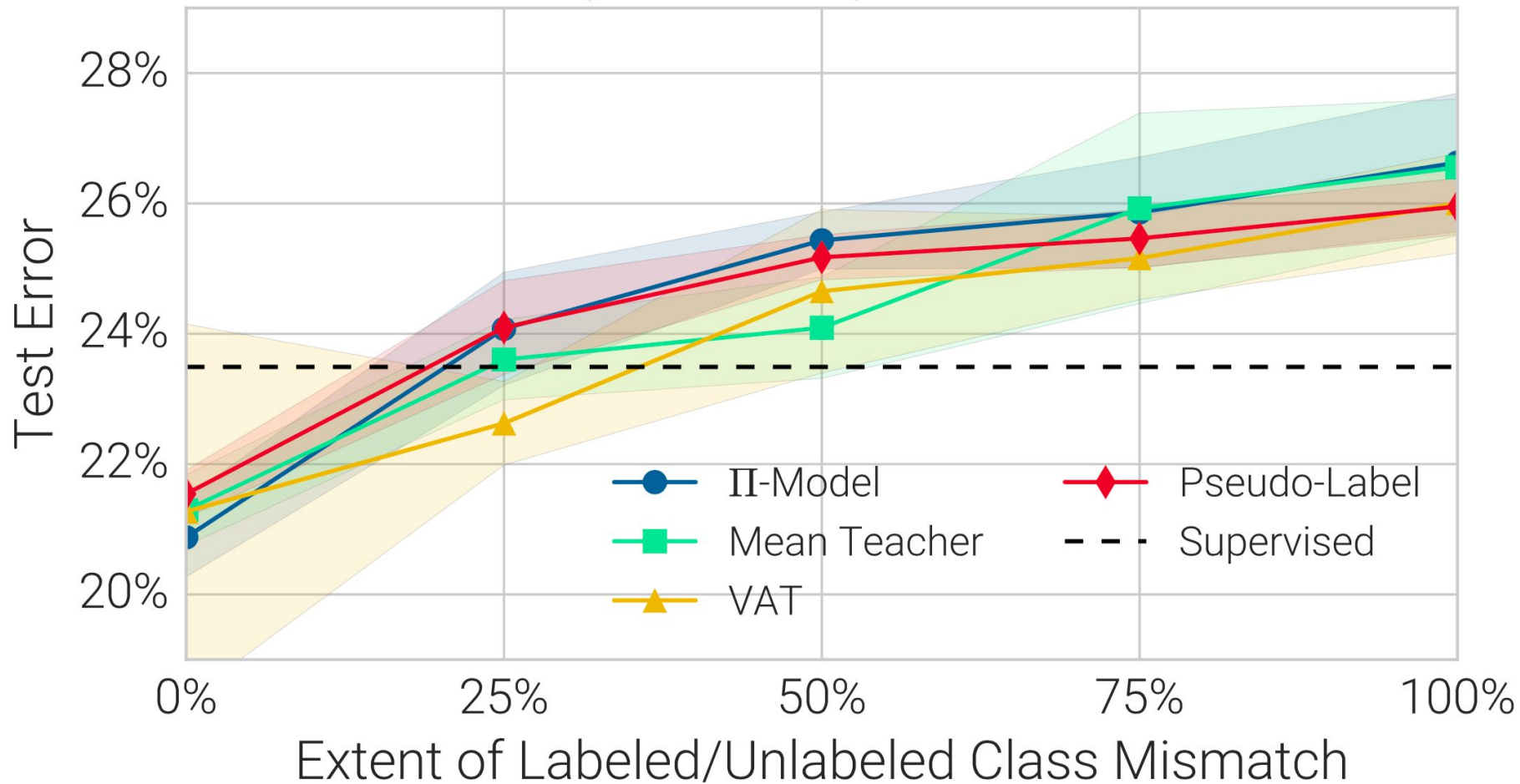
Method	CIFAR-10 4000 Labels
VAT with Entropy Minimization	13.13%
ImageNet \rightarrow CIFAR-10	12.09%
ImageNet \rightarrow CIFAR-10 (no overlap)	12.91%

Comparison of error rates for the best-performing semi-supervised learning method and transfer learning.

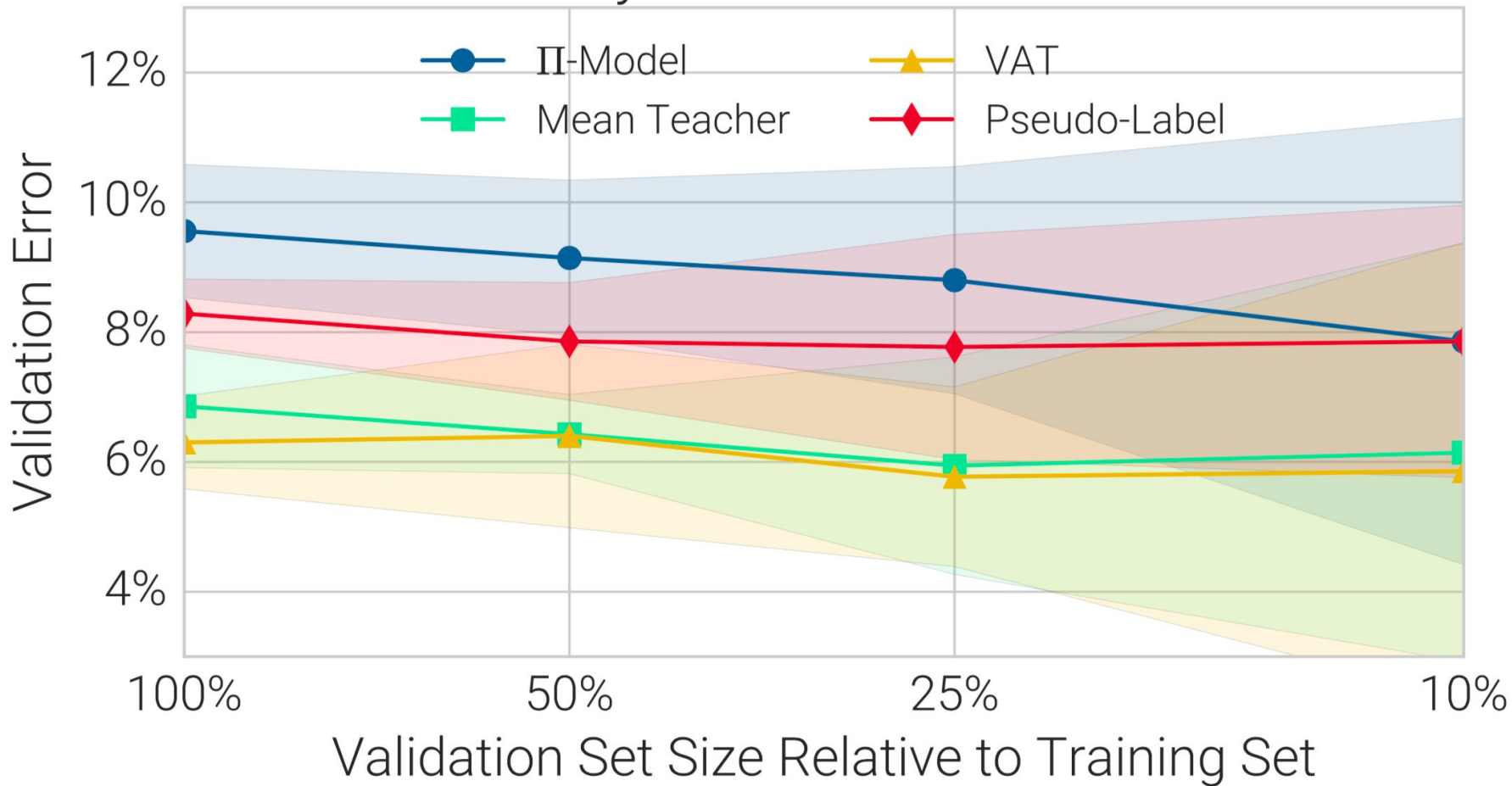
SVHN, Varying Number of Labels

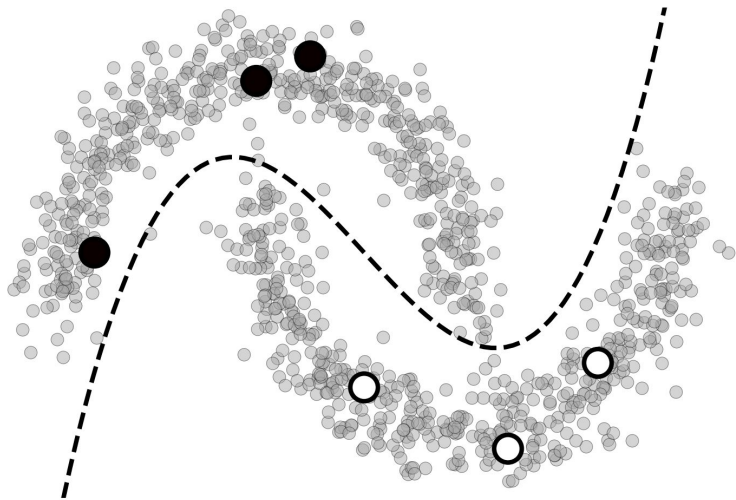


CIFAR-10, 6 classes, 400 Labels Each



Variability in Small Validation Sets





Paper: <https://arxiv.org/abs/1804.09170>

Code: <https://github.com/brain-research/realistic-ssl-evaluation/>

Poster: Today, 10:45 AM - 12:45 PM, Room 210 & 230 AB **#95**