

408 **A Notation and definitions**

Notation	Definition
$H(p, q)$	Cross-entropy between “target” distribution p and “predicted” distribution q
x	A labeled example, used as input to a model
p	A (one-hot) label
L	The number of possible label classes (the dimensionality of p)
\mathcal{X}	A batch of labeled examples and their labels
\mathcal{X}'	A batch of processed labeled examples produced by MixMatch
u	An unlabeled example, used as input to a model
q	A guessed label distribution for an unlabeled example
\mathcal{U}	A batch of unlabeled examples
409 \mathcal{U}'	A batch of processed unlabeled examples with their label guesses produced by MixMatch
θ	The model’s parameters
$p_{\text{model}}(y \mid x; \theta)$	The model’s predicted distribution over classes
$\text{Augment}(x)$	A stochastic data augmentation function that returns a modified version of x . For example, $\text{Augment}(\cdot)$ could implement randomly shifting an input image, or implement adding a perturbation sampled from a Gaussian distribution to x .
$\lambda_{\mathcal{U}}$	A hyper-parameter weighting the contribution of the unlabeled examples to the training loss
α	Hyperparameter for the Beta distribution used in MixUp
T	Temperature parameter for sharpening used in MixMatch
K	Number of augmentations used when guessing labels in MixMatch

B Tabular results

B.1 CIFAR-10

Training the same model with supervised learning on the entire 50000-example training set achieved an error rate of 4.13%.

Methods/Labels	250	500	1000	2000	4000
PiModel	53.02 ± 2.05	41.82 ± 1.52	31.53 ± 0.98	23.07 ± 0.66	17.41 ± 0.37
PseudoLabel	49.98 ± 1.17	40.55 ± 1.70	30.91 ± 1.73	21.96 ± 0.42	16.21 ± 0.11
Mixup	47.43 ± 0.92	36.17 ± 1.36	25.72 ± 0.66	18.14 ± 1.06	13.15 ± 0.20
VAT	36.03 ± 2.82	26.11 ± 1.52	18.68 ± 0.40	14.40 ± 0.15	11.05 ± 0.31
MeanTeacher	47.32 ± 4.71	42.01 ± 5.86	17.32 ± 4.00	12.17 ± 0.22	10.36 ± 0.25
MixMatch	11.08 ± 0.87	9.65 ± 0.94	7.75 ± 0.32	7.03 ± 0.15	6.24 ± 0.06

Table 5: Error rate (%) for CIFAR10.

B.2 SVHN

Training the same model with supervised learning on the entire 73257-example training set achieved an error rate of 2.59%.

Methods/Labels	250	500	1000	2000	4000
PiModel	17.65 ± 0.27	11.44 ± 0.39	8.60 ± 0.18	6.94 ± 0.27	5.57 ± 0.14
PseudoLabel	21.16 ± 0.88	14.35 ± 0.37	10.19 ± 0.41	7.54 ± 0.27	5.71 ± 0.07
Mixup	39.97 ± 1.89	29.62 ± 1.54	16.79 ± 0.63	10.47 ± 0.48	7.96 ± 0.14
VAT	8.41 ± 1.01	7.44 ± 0.79	5.98 ± 0.21	4.85 ± 0.23	4.20 ± 0.15
MeanTeacher	6.45 ± 2.43	3.82 ± 0.17	3.75 ± 0.10	3.51 ± 0.09	3.39 ± 0.11
MixMatch	3.78 ± 0.26	3.64 ± 0.46	3.27 ± 0.31	3.04 ± 0.13	2.89 ± 0.06

Table 6: Error rate (%) for SVHN.

417 B.3 SVHN+Extra

418 Training the same model with supervised learning on the entire 604388-example training set achieved
 419 an error rate of 1.71%.

Methods/Labels	250	500	1000	2000	4000
PiModel	13.71 ± 0.32	10.78 ± 0.59	8.81 ± 0.33	7.07 ± 0.19	5.70 ± 0.13
PseudoLabel	17.71 ± 0.78	12.58 ± 0.59	9.28 ± 0.38	7.20 ± 0.18	5.56 ± 0.27
Mixup	33.03 ± 1.29	24.52 ± 0.59	14.05 ± 0.79	9.06 ± 0.55	7.27 ± 0.12
VAT	7.44 ± 1.38	7.37 ± 0.82	6.15 ± 0.53	4.99 ± 0.30	4.27 ± 0.30
MeanTeacher	2.77 ± 0.10	2.75 ± 0.07	2.69 ± 0.08	2.60 ± 0.04	2.54 ± 0.03
MixMatch	2.22 ± 0.08	2.17 ± 0.07	2.18 ± 0.06	2.12 ± 0.03	2.07 ± 0.05

Table 7: Error rate (%) for SVHN+Extra.

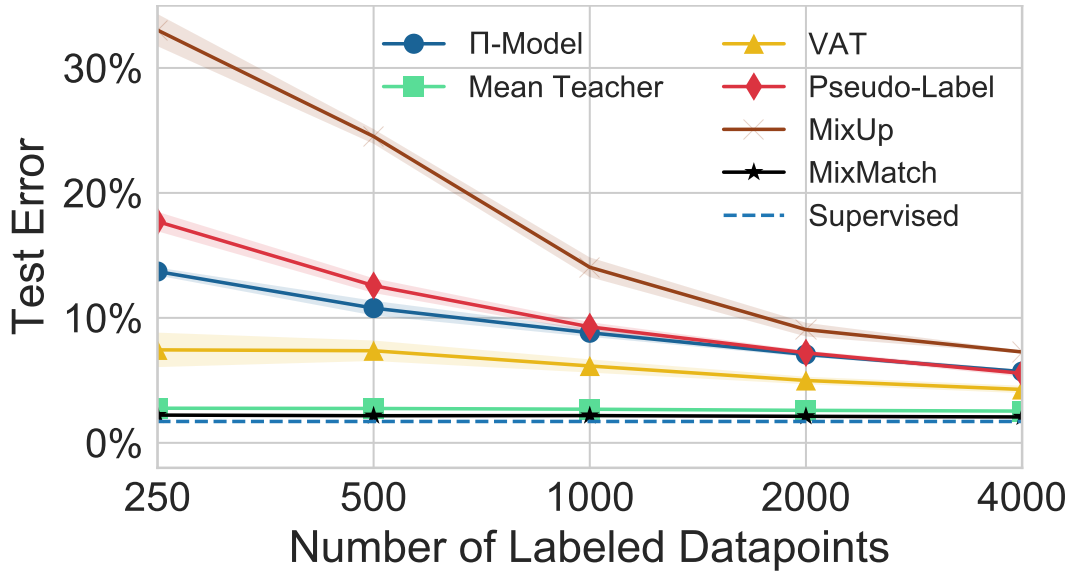


Figure 4: Error rate comparison of MixMatch to baseline methods on SVHN+Extra for a varying number of labels. With 250 examples we reach nearly the state of the art compared to supervised training for this model.