要点

- co-teaching+的优化
- confirmation bias 由cross update变成joint loss 解决
- disagreement变成agreement, 并且加入使两个网络相似的正则

	Decoupling	Co-teaching	Co-teaching+	JoCoR
small loss	✗	✓	✓	✓
cross update	✗	✓	✓	✗
joint training	✗	✗	✗	✓
disagreement	✓	✗	✓	✗
agreement	✗	✗	✗	✓

方法

Algorithm 1 JoCoR

Input: Network f with $\Theta = \{\Theta_1, \Theta_2\}$, learning rate η, fixed τ, epoch T_k and $T_{\max}$, iteration $I_{\max}$;

for $t = 1,2,\ldots,T_{\max}$ **do**
  Shuffle training set D;
  for $n = 1,\ldots,I_{\max}$ **do**
    Fetch mini-batch D_n from D;
    $\boldsymbol{p}_1 = f(\boldsymbol{x}, \Theta_1), \forall \boldsymbol{x} \in D_n$;
    $\boldsymbol{p}_2 = f(\boldsymbol{x}, \Theta_2), \forall \boldsymbol{x} \in D_n$;
    Calculate the joint loss ℓ by (1) using $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$;
    Obtain small-loss sets $\tilde{D}_n$ by (4) from D_n;
    Obtain L by (5) on $\tilde{D}_n$;
    Update $\Theta = \Theta - \eta \nabla L$;
  end for
  Update $R(t) = 1 - \min\left\{\frac{t}{T_k}\tau, \tau\right\}$
end for

Output: Θ_1 and Θ_2

loss function

$$\ell(x_i) = (1-\lambda) * \ell_{\text{sup}}(\boldsymbol{x}_i, y_i) + \lambda * \ell_{\text{con}}(\boldsymbol{x}_i) \qquad (1)$$

l_{sup}两个网络的CE loss之和

l_{con}为contrastive loss, 参考的依据为 agreement maximization principle：

```
不同的模型对wrong label的预测是很难相一致的
```

因此此正则项的目的为：wrong label的agreement error很大，不会被small loss采集

small loss selection

采用了R(t)，随着时间增长选的小样本逐渐减少

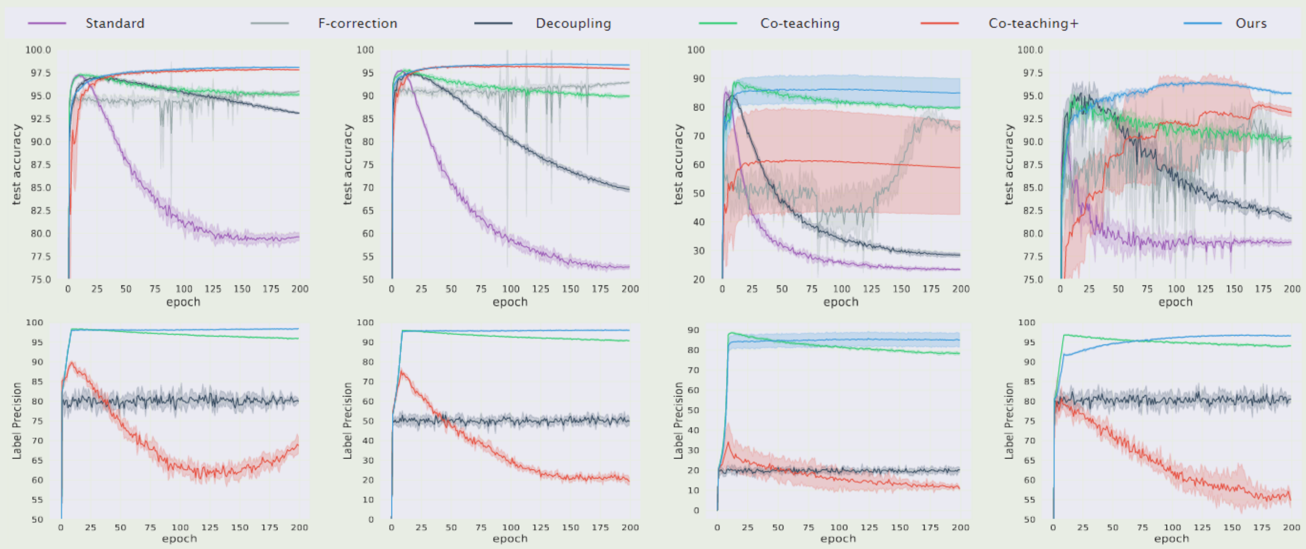
总结

1.confirmation bias：两个网络从相互挑小样本到联合损失函数，都避免error flow在一个网络上流动的情况

2.disagreement: 从两个网络更新disagreement area到共享一个agreement loss, 再配合small loss，都解决了每个样本都学习的问题

实验

t%e t%10e

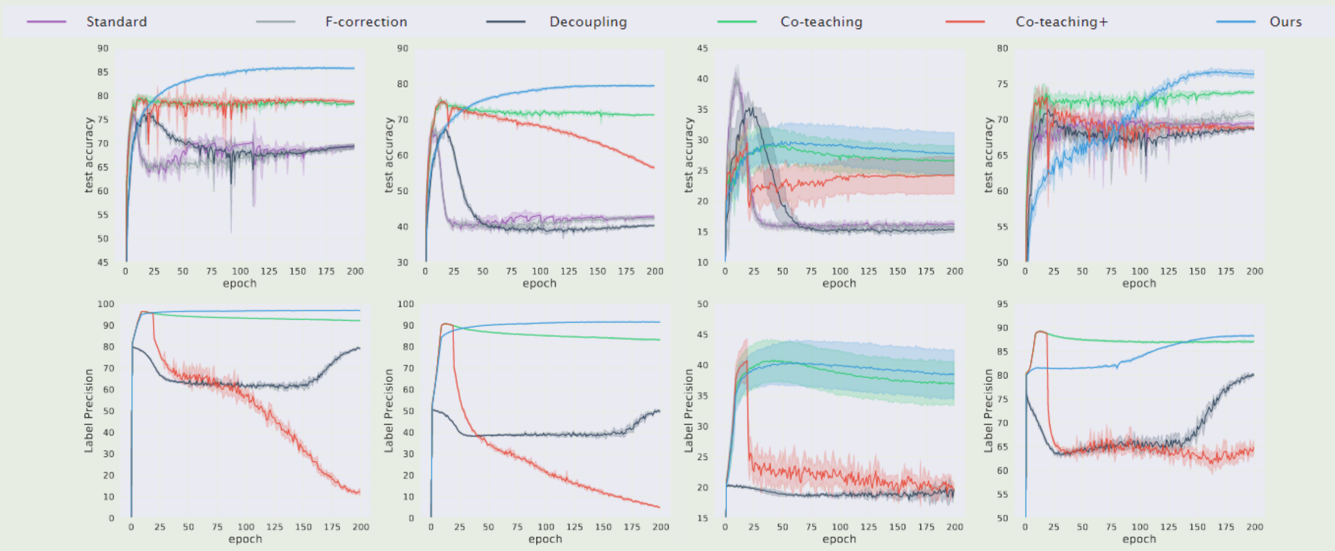


(a) Symmetry-20% (b) Symmetry-50% (c) Symmetry-80% (d) Asymmetry-40%

Figure 3. Results on MNIST dataset. Top: test accuracy(%) vs. epochs; bottom: label precision(%) vs. epochs.

Table 2. Average test accuracy (%) on *MNIST* over the last 10 epochs.

Flipping-Rate	Standard	F-correction	Decoupling	Co-teaching	Co-teaching+	JoCoR
Symmetry-20%	79.56 ± 0.44	95.38 ± 0.10	93.16 ± 0.11	95.10 ± 0.16	97.81 ± 0.03	**98.06** ± 0.04
Symmetry-50%	52.66 ± 0.43	92.74 ± 0.21	69.79 ± 0.52	89.82 ± 0.31	95.80 ± 0.09	**96.64** ± 0.12
Symmetry-80%	23.43 ± 0.31	72.96 ± 0.90	28.51 ± 0.65	79.73 ± 0.35	58.92 ± 14.73	**84.89** ± 4.55
Asymmetry-40%	79.00 ± 0.28	89.77 ± 0.96	81.84 ± 0.38	90.28 ± 0.27	93.28 ± 0.43	**95.24** ± 0.10

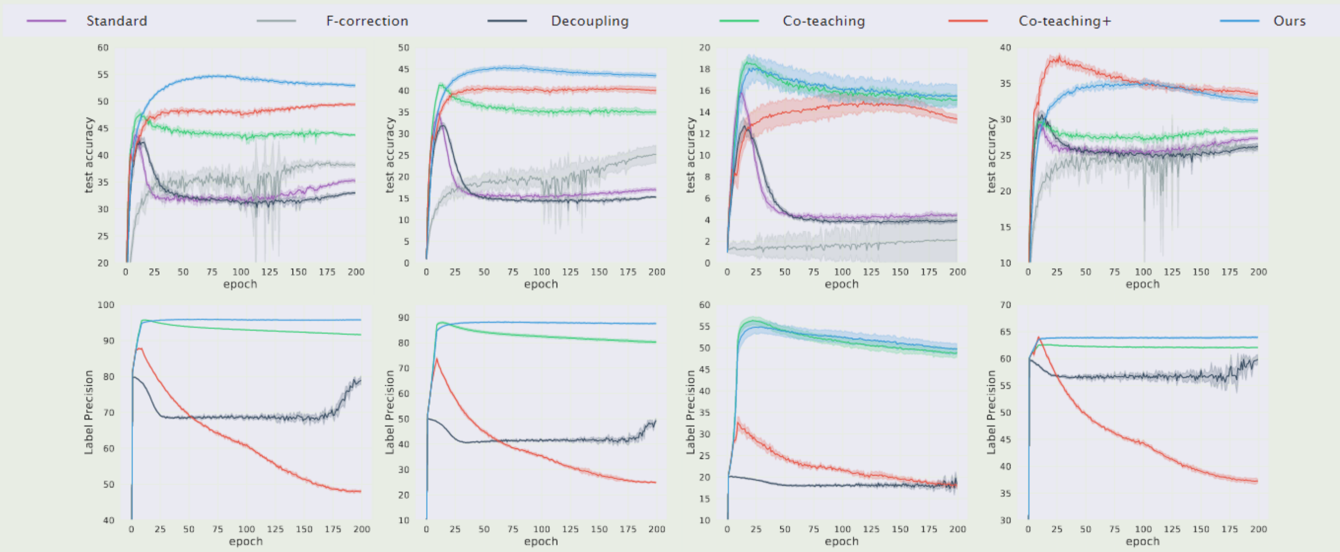


(a) Symmetry-20% (b) Symmetry-50% (c) Symmetry-80% (d) Asymmetry-40%

Figure 5. Results on CIFAR-10 dataset. Top: test accuracy(%) vs. epochs; bottom: label precision(%) vs. epochs.

Table 3. Average test accuracy (%) on *CIFAR-10* over the last 10 epochs.

Flipping-Rate	Standard	F-correction	Decoupling	Co-teaching	Co-teaching+	JoCoR
Symmetry-20%	69.18 ± 0.52	68.74 ± 0.20	69.32 ± 0.40	78.23 ± 0.27	78.71 ± 0.34	$\mathbf{85.73} \pm 0.19$
Symmetry-50%	42.71 ± 0.42	42.19 ± 0.60	40.22 ± 0.30	71.30 ± 0.13	57.05 ± 0.54	$\mathbf{79.41} \pm 0.25$
Symmetry-80%	16.24 ± 0.39	15.88 ± 0.42	15.31 ± 0.43	26.58 ± 2.22	24.19 ± 2.74	$\mathbf{27.78} \pm 3.06$
Asymmetry-40%	69.43 ± 0.33	70.60 ± 0.40	68.72 ± 0.30	73.78 ± 0.22	68.84 ± 0.20	$\mathbf{76.36} \pm 0.49$



(a) Symmetry-20% (b) Symmetry-50% (c) Symmetry-80% (d) Asymmetry-40%

Figure 6. Results on CIFAR-100 dataset. Top: test accuracy(%) vs. epochs; bottom: label precision(%) vs. epochs.

Table 4. Average test accuracy (%) on *CIFAR-100* over the last 10 epochs.

Flipping-Rate	Standard	F-correction	Decoupling	Co-teaching	Co-teaching+	JoCoR
Symmetry-20%	35.14 ± 0.44	37.95 ± 0.10	33.10 ± 0.12	43.73 ± 0.16	49.27 ± 0.03	$\mathbf{53.01} \pm 0.04$
Symmetry-50%	16.97 ± 0.40	24.98 ± 1.82	15.25 ± 0.20	34.96 ± 0.50	40.04 ± 0.70	$\mathbf{43.49} \pm 0.46$
Symmetry-80%	4.41 ± 0.14	2.10 ± 2.23	3.89 ± 0.16	15.15 ± 0.46	13.44 ± 0.37	$\mathbf{15.49} \pm 0.98$
Asymmetry-40%	27.29 ± 0.25	25.94 ± 0.44	26.11 ± 0.39	28.35 ± 0.25	$\mathbf{33.62} \pm 0.39$	32.70 ± 0.35

蒸馏实验

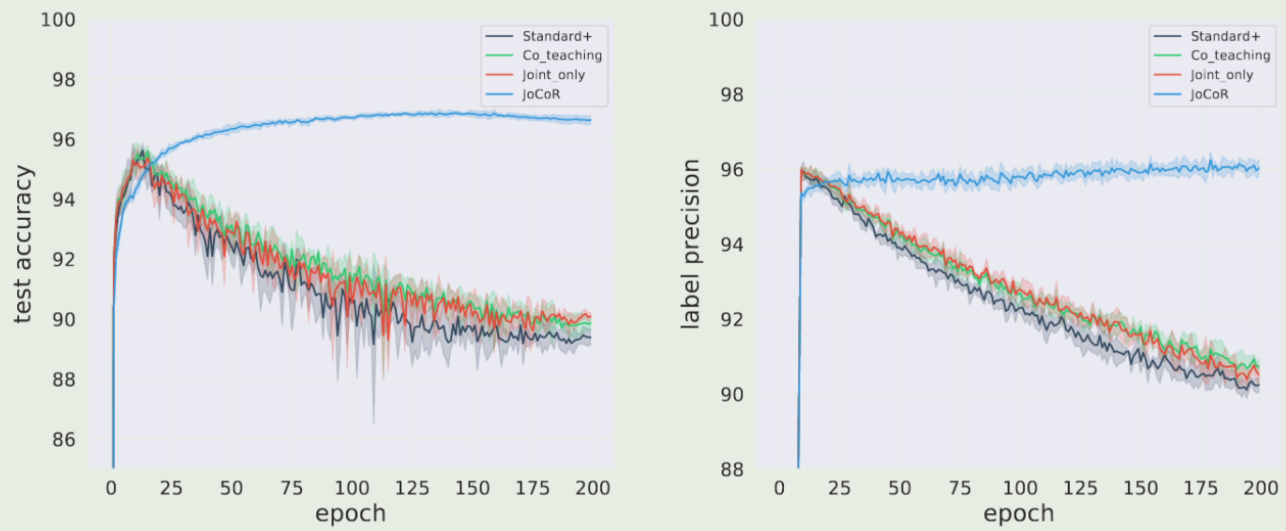


Figure 7. Results of ablation study on *MNIST*

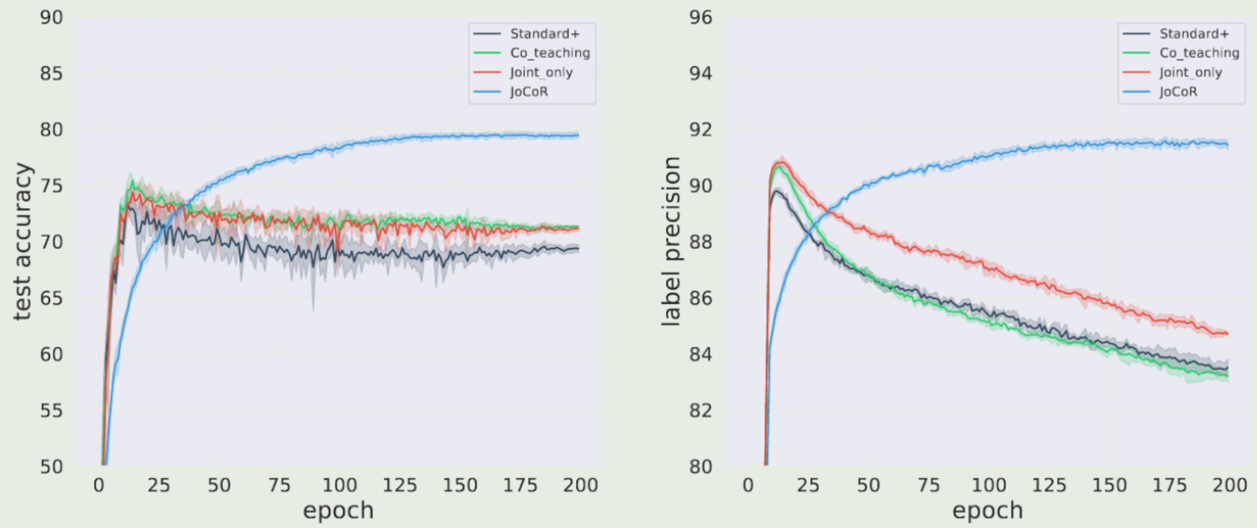


Figure 8. Results of ablation study on *CIFAR-10*

standard+: 带small loss pick的CE loss函数。他和co-teaching 和joint_only 对比证明要消除confirmation bias才能提高模型性能.

joint only: 不带正则项的损失函数。它和joCoR对比证明了运用agreement maximization是正确的.

co_teaching: 它和joint_only对比证明了 联合损失要强于cross update.