Table 1: Results. P/C is prediction / constraint accuracy in %.

(a) Class-Similarity constraint.

	Fashion-MNIST			CIFAR-10		
	P	C	λ	P	C	λ
Baseline	77.55	84.31	–	79.06	48.65	–
DL2	77.88	89.15	0.6	78.55	52.21	0.4
I_{G}	63.46	91.30	3.0	77.65	81.56	1.2
I_{KD}	75.39	80.59	0.8	78.82	72.94	0.6
I_{LK}	64.64	97.28	4.0	76.06	87.75	6.0
I_{GG}	67.25	95.54	3.0	74.83	88.67	10.0
I_{RC}	76.79	92.56	0.8	79.14	80.51	0.8
$(I_{\mathrm{RC}})_{s=9}$	77.06	93.63	0.8	78.30	80.87	0.8
$(I_{\mathrm{RC}})_{\phi=x^2}$	**76.88**	**95.94**	1.0	**78.31**	**90.74**	1.6
I_{YG}	74.14	80.15	1.0	77.81	73.19	0.8

(b) Group constraint.

	GTSRB		
	P	C	λ
Baseline	89.93	49.97	–
DL2	**88.16**	**77.25**	7.0
T_{G}	88.38	74.20	5.0
T_{LK}	85.26	78.43	5.0
T_{RC}	86.52	77.56	5.0
T_{YG}	87.47	76.47	5.0

Table 2: Average epoch train times in seconds.

(a) Class-Similarity constraint.

	Fashion-MNIST	CIFAR-10
Baseline	0.6 s	3.4 s
DL2	0.6 s	3.4 s
I_{G}	0.6 s	3.8 s
I_{KD}	0.7 s	4.0 s
I_{LK}	0.7 s	4.0 s
I_{GG}	0.7 s	4.0 s
I_{RC}	0.7 s	4.0 s
$(I_{\mathrm{RC}})_{s=9}$	0.7 s	4.1 s
$(I_{\mathrm{RC}})_{\phi=x^2}$	0.7 s	4.1 s
I_{YG}	0.7 s	4.1 s

(b) Group constraint.

	GTSRB
Baseline	1.1 s
DL2	1.1 s
T_{G}	1.1 s
$T_{\mathrm{ŁK}}$	1.1 s
T_{RC}	1.1 s
T_{YG}	1.2 s