A Additional Ablation Studies

LSR can be optionally applied on the student's one-hot ground truth labels. However, we observed that this did not provide performance gain, so we did not include it in our final design. Table 1 provides additional ablation study of the effect of LSR on the performance. As can be seen, although LSR alone improves performance, its effect is not significant in our framework.

MT	U	LSR	SLS	Acc_i	Acc
×	×	×	×	71.80	75.12
×	×	√	×	72.45	81.81
×	×	×	√	80.93	85.33
√	×	√	×	81.78	85.16
√	×	×	√	82.56	86.82
√	√	√	×	81.94	85.43
√	√	×	√	**83.18**	**86.94**
√	√	√	√	82.01	86.82

Table 1: Accuracy of the different components of our model with and without LSR in RAF-DB with 30% symmetric noise.

In Section 4.3 of the paper, we show ablation studies for RAF-DB with 30% symmetric noise. In Tables 2 and 3 we report more results for asymmetric noise and other noise rates. Please refer to Section 4.3 for more details.

MT	U	SLS	Sym (50%)	Sym (70%)	Asym (30%)
×	×	×	62.71	47.78	70.50
√	×	×	77.54	62.61	81.61
×	√	×	76.33	62.71	79.40
×	×	√	75.78	57.50	80.90
√	√	×	78.52	63.13	81.61
√	×	√	78.98	66.19	81.84
√	√	√	**80.44**	**71.77**	**82.69**

Table 2: Accuracy of the different components of our model with 50% symmetric noise, 70% symmetric noise, and 30% asymmetric noise in RAF-DB.

B Numerical Results Comparing with SOTA Methods

In Section 4.4 of the paper, we plot the comparison results between SOFT and the state-of-the-art methods with different rates of symmetric (Fig. 4) and asymmetric (Fig. 5(a)) noise. Here, we provide the numerical results in Table 4 for symmetric noise and in Table 5 for asymmetric noise. Please refer to Section 4.4 for the analysis of the results.

Smooth method	Teacher	Ins-aware	Non-zero	Sym (50%)	Sym (70%)	Asym (30%)
$\times$	$\times$	$\times$	$\times$	78.52	63.13	81.6
LSR	$\times$	$\times$	$\times$	78.74	61.93	81.84
LSR*	$\surd$	$\times$	$\times$	78.91	64.76	81.84
SLS(0)	$\surd$	$\surd$	$\times$	80.18	70.57	81.91
SLS	$\surd$	$\surd$	$\surd$	**80.44**	**71.77**	**82.69**

Table 3: The effects of different design components for SLS with 50% symmetric noise, 70% symmetric noise, and 30% asymmetric noise in RAF-DB.

Method	Noise(%)	RAF-DB	AffectNet
Baseline	10	80.43±0.72	57.21±0.31
SCN[2]	10	81.92±0.69	58.48±0.62
DMUE[1]	10	83.19±0.83	61.21±0.36
SOFT	10	**88.93±0.13**	**61.31±0.08**
Baseline	20	78.01±0.29	56.21±0.31
SCN[2]	20	80.02±0.32	56.98±0.28
DMUE[1]	20	81.02±0.69	59.06±0.34
SOFT	20	**88.09±0.06**	**60.99±0.17**
Baseline	30	75.12±0.78	52.67±0.45
SCN[2]	30	77.46±0.64	55.04±0.54
DMUE[1]	30	79.41±0.74	56.88±0.56
SOFT	30	**86.94±0.21**	**59.64±0.06**
Baseline	50	62.71±0.45	51.29±0.36
SCN[2]	50	71.54±0.67	43.09±0.58
DMUE[1]	50	72.43±0.74	56.06±0.49
SOFT	50	**80.44±0.17**	**57.27±0.22**
Baseline	70	47.78±0.47	43.76±0.89
SCN[2]	70	44.95±1.49	-
DMUE[1]	70	45.27±0.64	-
SOFT	70	**71.77±0.18**	**49.00±0.10**

Table 4: Accuracy on RAF-DB and AffectNet with injected symmetrical noise.

Method	10%	20%	30%	40%	50%
DMUE[1]	74.35±0.42	73.44±0.34	65.06±0.46	52.48±0.52	41.81±0.57
SOFT	**87.54±0.09**	**85.56±0.19**	**82.69±0.18**	**69.17±0.21**	**49.39±0.23**

Table 5: Accuracy on RAF-DB with different rates of injected asymmetric noise. SOFT consistently beats the current state-of-the-art, DMUE.

References

1. She, J., Hu, Y., Shi, H., Wang, J., Shen, Q., Mei, T.: Dive into ambiguity: Latent distribution mining and pairwise uncertainty estimation for facial expression recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6248–6257 (2021)
2. Wang, K., Peng, X., Yang, J., Lu, S., Qiao, Y.: Suppressing uncertainties for large-scale facial expression recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6897–6906 (2020)