

## Supplementary Material: Noise as a Resource for Learning in Knowledge Distillation

Elahe Arani\*, Fahad Sarfraz\*, and Bahram Zonooz  
Advanced Research Lab, NavInfo Europe, Eindhoven, The Netherlands  
{elahe.arani, fahad.sarfraz}@navinfo.eu, bahram.zonooz@gmail.com

Table S1. PGD-20 robustness evaluation for Soft Randomization (SR) vs Gaussian Augmentation (GA). GA and SR indicate Gaussian Augmentation and Soft Randomization, respectively. We report the mean and 1 std of the worst-case PGD-20 attack with five random initializations for five different seeds.

$\sigma$	CIFAR-10		SVHN	
	GA	SR	GA	SR
0.01	0.00±0.00	<b>0.78±0.20</b>	0.08±0.02	<b>21.07±0.97</b>
0.02	0.00±0.00	<b>3.83±0.28</b>	0.21±0.06	<b>27.82±0.83</b>
0.03	0.02±0.01	<b>8.30±0.79</b>	0.52±0.08	<b>33.41±0.71</b>
0.04	0.12±0.03	<b>12.89±0.70</b>	1.09±0.23	<b>39.97±1.50</b>
0.05	0.38±0.04	<b>15.56±0.44</b>	1.99±0.13	<b>41.75±1.06</b>
0.1	4.20±0.38	<b>22.15±0.40</b>	10.82±1.46	<b>50.35±1.45</b>
0.2	12.19±0.29	<b>24.14±0.63</b>	25.94±1.05	<b>51.39±0.80</b>
0.3	17.64±0.24	<b>23.29±0.23</b>	30.97±0.19	<b>45.99±0.32</b>
0.4	20.98±0.16	<b>23.42±0.32</b>	34.34±0.32	<b>41.28±0.38</b>
0.5	22.34±0.33	<b>23.54±0.37</b>	37.78±0.60	<b>40.13±0.61</b>

Table S2. Generalization performance of Messy Collaboration on CIFAR-10 with varying noise rates trained on corrupted labels. We report the mean and 1 std for five different seeds.

Noise rate (r)	0	0.05	0.10	0.15	0.25	0.50
Teacher	95.11±0.00	92.61±0.00	89.37±0.00	86.56±0.00	78.94±0.00	55.53±0.00
Baseline	93.95±0.18	90.51±0.10	86.91±0.78	83.51±0.47	76.07±1.03	56.61±1.16
0	94.26±0.16	<b>92.67±0.14</b>	90.69±0.13	89.36±0.47	85.32±0.52	67.97±0.40
0.05	94.39±0.12	92.57±0.19	90.88±0.25	89.36±0.26	85.30±0.20	<b>68.75±0.78</b>
0.1	<b>94.41±0.21</b>	92.53±0.13	90.73±0.26	89.33±0.25	85.22±0.03	68.25±0.50
0.15	94.18±0.08	92.52±0.09	90.95±0.10	<b>89.54±0.17</b>	85.15±0.28	68.27±0.99
0.25	94.32±0.13	92.60±0.24	<b>90.97±0.12</b>	89.44±0.24	84.96±0.24	68.05±0.56
0.5	94.40±0.19	92.62±0.20	90.96±0.18	89.47±0.36	<b>85.46±0.36</b>	68.54±0.55

\*Equal contribution.

Table S3. Generalization performance of Messy Collaboration on SVHN with varying noise rates trained on corrupted labels. We report the mean and 1 std for five different seeds.

Noise rate (r)	0	0.05	0.10	0.15	0.25	0.50
Teacher	96.23±0.00	94.07±0.00	93.19±0.00	91.82±0.00	89.53±0.00	82.84±0.00
Baseline	96.11±0.10	93.92±0.18	92.78±0.38	91.58±0.18	89.01±0.59	81.56±1.13
0	<b>96.81±0.07</b>	96.29±0.13	95.94±0.12	95.61±0.14	<b>94.80±0.09</b>	<b>91.93±0.25</b>
0.05	96.77±0.16	96.30±0.05	95.98±0.11	95.61±0.08	94.79±0.13	91.74±0.24
0.1	96.72±0.14	<b>96.34±0.09</b>	95.95±0.09	95.63±0.13	94.73±0.11	91.74±0.12
0.15	96.80±0.13	96.31±0.07	<b>95.99±0.11</b>	95.65±0.12	94.77±0.17	91.73±0.09
0.25	96.77±0.03	96.31±0.07	95.93±0.14	<b>95.66±0.13</b>	94.67±0.15	91.52±0.12
0.5	96.81±0.13	96.27±0.07	95.90±0.08	95.59±0.08	94.59±0.10	91.05±0.13

Table S4. Out-of-distribution generalization performance on CINIC dataset for models trained with Messy Collaboration on CIFAR-10 with varying noise rates trained on corrupted labels. We report the mean and 1 std for five different seeds.

Teacher	70.23±0.00	66.22±0.00	62.66±0.00	59.07±0.00	53.63±0.00	36.75±0.00
Baseline	68.89±0.08	63.22±0.53	59.50±0.27	55.00±0.79	50.08±1.03	37.21±0.99
0	68.95±0.18	65.52±0.34	<b>63.83±0.14</b>	61.24±0.30	56.68±0.44	44.42±0.73
0.05	69.04±0.13	65.52±0.82	63.69±0.16	61.28±0.48	56.58±0.42	<b>45.15±0.21</b>
0.1	69.02±0.17	<b>65.97±0.38</b>	63.57±0.34	<b>61.68±0.39</b>	56.15±0.62	44.92±0.58
0.15	<b>69.20±0.15</b>	65.74±0.33	63.73±0.47	61.34±0.47	56.41±0.51	44.78±0.57
0.25	69.13±0.19	65.92±0.34	63.45±0.32	61.52±0.52	<b>56.74±0.53</b>	45.02±0.56
0.5	69.06±0.21	65.28±0.44	63.62±0.41	61.19±0.19	56.40±0.42	44.68±1.16