Appendix

A. Pruning Schemes

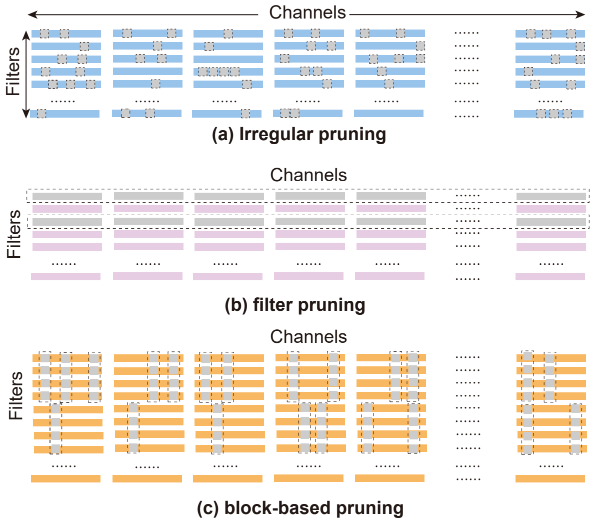

Figure A1. (a) Irregular pruning scheme, (b) filter pruning scheme, and (c) block-based pruning scheme.

We use Figure A1 to illustrate three representative pruning schemes i.e., irregular pruning [20, 19], filter pruning [28, 22], and block-based pruning [12, 29], where grey represents the pruned weights and colors are for remaining non-zero weights in the GEMM matrix for a convolutional layer. The irregular pruning scheme in Figure A1 (a) prunes weights at arbitrary locations, and it can achieve a high pruning rate with negligible accuracy loss, but is not compatible with data-parallel executions on computing devices. The filter pruning scheme in Figure A1 (b) prunes whole filters in the GEMM matrix, and it preserves regularity on the pruned models, but suffers from significant accuracy loss. In the block-based pruning scheme in Figure A1 (c), weights are partitioned into blocks with the same size. Figure A1 (c) illustrates an example block size of $(4, 1)$. In the block-based pruning, weights in the same block are either pruned to zeros or remaining all together. The block-based pruning can achieve negligible accuracy loss as the irregular pruning does, but are far more friendly for inference execution on computing devices.

B. Edge Devices Performance

We test the inference speed of a single compressed model and the overall framework on a mobile device (Samsung Galaxy S10 smartphone) with Qualcomm Adreno 640 GPU [26]. For the compressed expert model with 4X compression rate (or 8X), the computation number (FLOPs) is reduced by 2.4X (or 4.5X for 8X compression), and the inference latency is decreased by 2.2X (or 3.8X for 8X compression). As the experts in MoL can run in parallel, the overall framework latency is reduced by 1.8X (or 3.1X for 8X compressed model) compared with the uncompressed model.

C. Additional Results

Table A1. **Test accuracy (%) of expert models (in original size) on Tiny-ImageNet dataset under various attack methods.**

	M_{clean}	M_{ℓ_1}	M_{ℓ_2}	M_{ℓ_∞}	M_{fog}	M_{snow}	M_{RT}
clean accuracy	**61.02**	54.46	49.05	42.75	10.03	41.24	53.95
PGD ℓ_1 (ϵ = 12)	0.92	**36.39**	38.99	31.56	0.00	2.38	0.82
PGD ℓ_2 (ϵ = 1.0)	0.05	17.44	**27.90**	22.07	0.00	0.05	0.05
PGD ℓ_∞ (ϵ = 8/255)	0.00	1.18	9.80	**21.88**	0.00	0.00	0.00
Fog ($\{t, light\} = \{0.4, 1.2\}$)	4.00	3.26	2.35	2.04	**56.49**	5.14	1.75
Snow ($darkness = 8$)	3.97	3.19	2.63	2.11	1.45	**18.53**	2.49
Rotation ($max = 180°$)	25.69	19.69	16.03	15.56	3.01	13.17	**52.29**

Table A2. **Test accuracy (%) of pruned expert models** (4× **pruning rate leading to 1/4 non-zero parameters) on CIFAR-10 dataset under various attack methods.** The three numbers in each column are from irregular pruning (I), filter pruning (F), and block-based pruning (B), respectively. All attack settings are as same as Table 1.

	M_{clean}	M_{ℓ_1}	M_{ℓ_2}	M_{ℓ_∞}	M_{fog}	M_{snow}	M_{RT}
	I/F/B	I/F/B	I/F/B	I/F/B	I/F/B	I/F/B	I/F/B
clean	**95.5/93.9/95.4**	83.8/81.3/83.6	85.3/82.1/84.3	86.9/82.9/86.3	70.6/71.3/72.4	90.9/85.9/88.5	86.6/81.9/84.4
PGD ℓ_1	2.2/0.6/2.4	**55.6/52.8/55.1**	55.4/51.8/55.3	49.3/47.7/49.9	2.3/2.1/2.7	4.2/3.7/4.3	0.0/0.0 /2.0
PGD ℓ_2	0.0/ 0.0/0.0	45.0/ 42.2/44.3	**51.1/47.9/49.8**	33.0/ 32.3/33.4	0.4/ 0.3/0.6	0.4/0.3/0.5	0.0/0.0 /0.1
PGD ℓ_∞	0.0/ 0.0/0.0	31.3/ 30.6/31.6	42.7/39.1 /42.8	**49.9/45.9/50.1**	0.3/0.1 /0.4	0.3/ 0.1/0.3	0.0/0.0 /0.3
Fog	18.0/ 16.7/18.1	15.2/ 14.6/15.4	10.6/11.6 / 10.4	10.3/9.15/11.3	**89.1/74.7/88.9**	29.2/27.9/28.7	13.2/12.5 /13.6
Snow	30.0/ 28.7/30.2	24.9/ 24.8/ 25.1	24.9/ 25.1/ 25.4	25.9/ 25.2/26.0	18.1/16.15 /29.1	**72.1/63.5/69.6**	14.7/15.9 /15.2
Rotation	29.8/27.1 /31.5	28.7/ 28.4/ 29.2	29.4/29.9/30.2	30.7/ 29.7/ 30.5	22.2/ 21.1/23.2	31.8/30.5 /31.6	**82.0/70.8/79.7**

Table A3. **Test accuracy (%) of expert models (pruned into 1/4 of the original size) on Tiny-ImageNet dataset under various attack methods.** The three numbers in each column are from irregular pruning (I), filter pruning (F), and block-based pruning (B), respectively. Same attack settings are used as Table A1.

Tiny-ImageNet	M_{clean}	M_{ℓ_1}	M_{ℓ_2}	M_{ℓ_∞}	M_{fog}	M_{snow}	M_{RT}
	I/F/B	I/F/B	I/F/B	I/F/B	I/F/B	I/F/B	I/F/B
clean	**60.9/55.5/61.2**	52.5/48.9/54.8	45.7/42.2/47.7	41.2/36.7/43.5	14.7/12.5/14.7	44.0/32.1/40.3	52.0/45.9/52.7
PGD ℓ_1	0.9/0.3/1.4	**34.6/32.7/38.3**	36.0/33.1/38.5	30.4/27.2/32.2	0.0/0.0/0.0	5.1/1.9/5.2	0.6/0.3/1.1
PGD ℓ_2	0.1/0.0/0.0	16.3/16.3/20.4	**26.1/24.3/28.3**	20.3/19.7/22.6	0.0/0.0/0.0	0.2/0.1/0.1	0.0/0.0/0.0
PGD ℓ_∞	0.0/0.0/0.0	1.1/1.3/1.8	9.0/9.0/10.8	**21.0/17.0/21.5**	0.0/0.0/0.0	0.0/0.0/0.0	0.0/0.0/0.0
Fog	3.8/3.7/4.5	2.7/3.6/3.6	1.8/2.9/2.3	1.7/2.1/2.3	**40.1/24.2/37.0**	4.4/3.1/4.3	1.9/2.4/2.3
Snow	3.9/3.3/4.1	2.6/3.1/2.9	2.6/2.3/2.9	1.9/2.2/2.1	1.9/1.6/2.1	**6.3/5.5/6.3**	2.4/2.0/2.6
Rotation	23.9/18.9/23.2	18.4/16.9/19.4	15.5/14.5/16.9	14.0/13.5/15.0	4.2/3.8/4.2	13.0/8.2/9.4	**48.0/39.9/49.7**

Table A4. **Test accuracy (%) of expert models (pruned into 1/8 of original size) on CIFAR-10 dataset under various attack methods.** The three numbers in each cell are from irregular pruning (I), filter pruning (F), and block-based pruning (B), respectively. Same attack settings are used as Table 1.

	M_{clean}	M_{ℓ_1}	M_{ℓ_2}	M_{ℓ_∞}	M_{fog}	M_{snow}	M_{RT}
	I/F/B	I/F/B	I/F/B	I/F/B	I/F/B	I/F/B	I/F/B
clean	**95.5/91.8/95.5**	83.4/77.3/82.7	84.7/76.5/84.1	86.1/75.6/85.7	71.0/66.7/74.7	90.2/75.4/88.2	83.9/75.1/85.8
PGD ℓ_1	2.1/1.9/2.4	**55.4/49.7/54.0**	54.3/48.7/54.7	30.7/29.2/30.0	3.3/1.0/2.3	4.4/1.3/4.0	0.3/0.1/1.5
PGD ℓ_2	0.0/0.0/0.0	44.8/40.4/43.7	**49.8/44.5/49.3**	33.0/31.2/34.2	0.2/0.1/0.4	0.3/0.2/0.5	0.0/0.0/0.2
PGD ℓ_∞	0.0/0.0/0.0	31.7/28.6/30.6	41.7/36.0/41.9	**49.9/41.2/49.0**	0.1/0.0/0.4	0.1/0.1/0.4	0.0/0.0/0.2
Fog	16.6/13.4/19.7	15.7/16.5/15.6	11.9/12.2/9.8	12.7/12.0/13.5	**89.5/68.7/89.6**	28.1/31.9/28.6	12.4/11.7/12.7
Snow	28.9/26.2/29.2	24.9/23.2/24.5	25.0/20.0/25.2	25.1/24.3/24.8	19.9/18.9/27.2	**65.5/55.4/66.9**	15.6/14.8/15.0
Rotation	27.1/24.8/30.6	29.6/28.3/29.9	28.6/27.1/29.9	28.8/28.1/28.6	21.1/19.5/24.5	30.4/27.7/31.3	**78.3/69.8/80.6**

Table A5. Test accuracy (%) comparisons on CIFAR-10 for (a) defending against only the adversarial perturbations and (b) defending against both adversarial and natural perturbations. The proposed *MoL* system at different sizes (original size, 1/4 of original size, and 1/8 of original size) are compared with baselines i.e., MAX [45], AVG [45], and MSD [33]. Detailed attack evaluation performance is listed under each perturbation type.

		(a) Defense against only adversarial perturbations						(b) Defense against both adversarial and natural perturbations				
		MAX	*AVG*	*MSD*	***MoL***	***MoL***$_{1/4}$	***MoL***$_{1/8}$	*MAX*	*AVG*	***MoL***	***MoL***$_{1/4}$	***MoL***$_{1/8}$
clean accuracy		74.30	72.04	71.18	**87.11**	84.94	84.23	70.25	74.05	**86.08**	84.12	82.42
ℓ_1 Attacks (ϵ = 16)	PGD [31]	47.69	49.08	51.40	**55.71**	54.04	53.37	45.84	41.48	**59.18**	57.46	52.26
	Gaussian Noise	33.01	32.20	34.65	33.80	35.26	**36.08**	18.20	22.25	29.20	29.80	**30.00**
	Salt&Pepper [39]	39.02	41.20	38.16	45.00	49.50	**52.03**	18.00	22.00	40.48	46.04	**46.50**
ℓ_2 Attacks (ϵ = 1.0)	PGD [31]	46.54	45.35	43.75	**54.21**	52.45	49.52	35.34	33.23	**55.25**	50.78	49.58
	FGSM [17]	50.00	52.50	45.60	**56.50**	55.50	53.00	38.52	38.77	**58.00**	54.00	53.15
	DNN attack [40]	28.30	31.40	24.90	41.80	45.50	**47.00**	11.40	12.20	33.20	**41.50**	37.80
	Deepfool [35]	43.80	44.50	46.10	**55.80**	53.50	43.60	46.30	38.40	52.70	**57.50**	54.00
	RayS [6]	46.46	49.60	45.80	56.24	60.00	**59.48**	44.34	34.35	54.95	56.87	**60.87**
ℓ_∞ Attacks (ϵ = 8/255)	PGD [31]	40.27	38.40	42.66	**52.01**	46.44	45.87	35.07	24.08	**50.25**	47.18	44.62
	FGSM [17]	43.00	42.50	44.50	**55.02**	54.00	47.40	41.00	30.70	51.50	**52.00**	47.34
	Deepfool [35]	35.50	42.00	43.00	**53.10**	46.50	42.30	38.40	40.20	43.50	**51.50**	47.80
	RayS [6]	38.91	29.82	31.64	**50.71**	51.58	44.91	27.41	26.82	**44.31**	44.60	50.64
Natural Perturbation Attacks	Fog [36]	-	-	-	-	-	-	25.70	24.50	29.30	**30.25**	27.01
	Snow [36]	-	-	-	-	-	-	31.04	27.77	**39.79**	36.92	32.90
	Rotation [15]	-	-	-	-	-	-	32.87	33.59	43.76	**48.82**	44.55
All attack [33]		33.00	32.20	36.40	**44.20**	40.00	39.00	31.40	30.80	**43.80**	40.40	38.60

Table A6. Test accuracy (%) comparisons on Tiny-ImageNet for (a) defending against only the adversarial perturbations and (b) defending against both adversarial and natural perturbations. The proposed *MoL* system at different sizes (original size and 1/4 of original size) are compared with baselines i.e., MAX [45], AVG [45], and MSD [33].

	(a) Defense against only adversarial perturbations					(b) Defense against both adversarial and natural perturbations			
	MAX	*AVG*	*MSD*	***MoL***	***MoL***$_{1/4}$	*MAX*	*AVG*	***MoL***	***MoL***$_{1/4}$
clean accuracy	21.86	23.58	28.70	**50.87**	46.23	8.24	18.48	**49.96**	44.77
ℓ_1 Attacks (ϵ = 12)	14.38	17.66	14.56	**25.81**	22.07	6.37	10.18	**24.63**	20.15
ℓ_2 Attacks (ϵ = 1.0)	16.20	17.54	19.26	**27.96**	24.55	3.80	6.92	**23.78**	18.68
ℓ_∞ Attacks (ϵ = 8/255)	16.62	15.36	14.79	**23.65**	19.98	4.23	6.78	**19.21**	16.35
Natural Perturbation Attacks	-	-	-	-	-	2.21	4.83	12.62	**13.62**