

## SUPPLEMENTARY MATERIAL FOR ADVERSARIAL META-LEARNING

## TREND OF AVERAGE LOSS &amp; TOP-1 ACCURACY FOR 5-WAY 5-SHOT LEARNING ON MINIIMAGENET

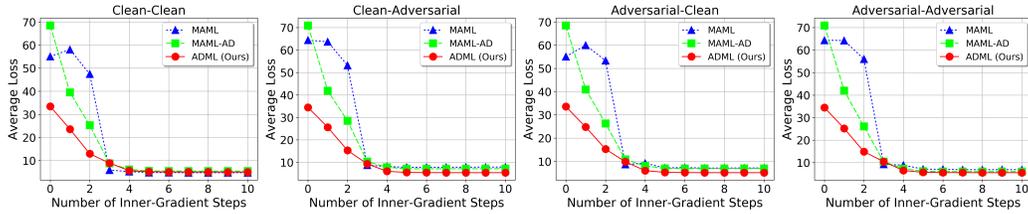


Figure 4: Average loss over the gradient update step for 5-way 5-shot learning on MiniImageNet

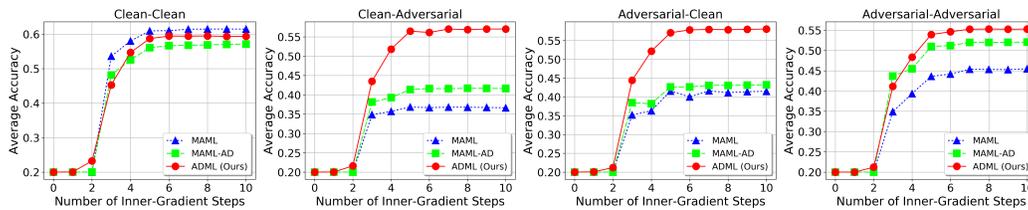


Figure 5: Top-1 accuracy over the gradient update step for 5-way 5-shot learning on MiniImageNet

## MODEL SPECIFICATION

For both datasets (i.e, MiniImageNet and CIFAR100), we followed the architecture used by Finn et al. (2017) for image embedding, which contains four  $3 \times 3$  convolutional blocks with batch normalizations, ReLU activations and  $2 \times 2$  max-poolings. *Note that this model is the threatened by the aforementioned adversarial attacks.*

## CIFAR100 WITH FGSM ATTACK

Note that the maximum perturbations adopted in FGSM Attack are 2 and 0.2.

Table 3: Average classification accuracies on CIFAR100 with FSGM Attack (5-way, 1-shot)

Method	Backbone	Meta-testing	$\epsilon = 2$		$\epsilon = 0.2$	
			Clean	Adversarial	Clean	Adversarial
MAML	32-32-32-32	Clean	57.67±1.76%	26.40±1.55%	57.67±1.76%	43.30±1.68%
		Adversarial	28.13±1.56%	28.23±1.64%	43.03±1.76%	39.00±1.70%
MAML-AD	32-32-32-32	Clean	52.70±1.89%	36.20±1.65%	52.70±1.89%	39.17±1.82%
		Adversarial	37.27±1.72%	41.67±1.86%	37.80±1.70%	37.60±1.78%
Matching Nets	64-64-64-64	Clean	47.94±0.56%	25.06±0.36%	47.68±0.52%	39.03±0.51%
		Adversarial	24.82±0.46%	27.72±0.43%	40.08±0.57%	37.79±0.44%
Relation Nets	64-96-128-256	Clean	58.68±0.92%	31.11±0.93%	58.72±0.90%	45.03±0.76%
		Adversarial	30.85±0.92%	30.52±0.59%	45.85±1.01%	41.40±0.80%
R2D2 (64C)	64-64-64-64	Clean	59.76±2.04%	26.07±1.00%	59.76±2.04%	35.53±1.47%
		Adversarial	27.20±1.52%	31.51±1.07%	43.63±2.17%	37.10±1.43%
R2D2	96-192-384-512	Clean	<b>60.52±2.01%</b>	26.56±0.93%	<b>60.52±2.01%</b>	36.71±1.45%
		Adversarial	27.90±1.61%	31.94±1.26%	43.64±2.21%	37.73±1.47%
ADML (Ours)	32-32-32-32	Clean	55.70±2.00%	<b>50.90±1.84%</b>	55.70±2.00%	<b>49.30±1.76%</b>
		Adversarial	<b>54.50±1.69%</b>	<b>50.60±1.83%</b>	<b>52.90±1.92%</b>	<b>45.00±1.79%</b>

Table 4: Average classification accuracies on CIFAR100 with FSGM Attack (5-way, 5-shot)

Method	Backbone	Meta-testing	$\epsilon = 2$		$\epsilon = 0.2$	
			Clean	Adversarial	Clean	Adversarial
MAML	32-32-32-32	Clean	74.03±0.89%	31.29±0.78%	74.03±0.89%	54.15±1.00%
		40%	65.69±0.92%	36.14±0.84%	68.99±0.94%	55.79±0.98%
		Adversarial	33.34±0.90%	43.66±0.86%	59.08±1.00%	53.93±0.96%
MAML-AD	32-32-32-32	Clean	67.71±0.96%	44.61±0.90%	67.73±0.96%	56.07±0.95%
		40%	64.85±0.90%	53.59±0.88%	65.93±0.93%	57.96±0.93%
		Adversarial	48.37±0.99%	58.92±0.97%	59.45±1.00%	56.33±0.98%
Matching Nets	64-64-64-64	Clean	62.95±0.46%	28.14±0.37%	62.58±0.49%	47.14±0.45%
		40%	54.39±0.48%	28.64±0.36%	57.86±0.48%	47.01±0.48%
		Adversarial	29.40±0.44%	32.77±0.42%	53.34±0.52%	46.50±0.46%
Relation Nets	64-96-128-256	Clean	75.52±0.66%	35.37±0.55%	75.22±0.70%	55.75±0.68%
		40%	66.85±0.79%	36.70±0.54%	68.67±0.80%	55.33±0.69%
		Adversarial	40.46±0.88%	39.82±0.57%	60.52±0.82%	55.50±0.69%
R2D2 (64C)	64-64-64-64	Clean	76.09±1.54%	27.83±1.10%	76.09±1.54%	38.77±1.74%
		40%	69.19±1.53%	38.00±1.21%	71.39±1.54%	50.96±1.55%
		Adversarial	35.14±1.75%	43.21±1.19%	58.99±1.75%	52.20±1.62%
R2D2	96-192-384-512	Clean	<b>76.29±1.44%</b>	29.53±1.10%	<b>76.29±1.44%</b>	40.28±1.66%
		40%	<b>69.53±1.47%</b>	39.32±1.11%	<b>71.68±1.49%</b>	52.04±1.56%
		Adversarial	35.79±1.60%	43.28±1.24%	58.85±1.89%	52.87±1.48%
ADML (Ours)	32-32-32-32	Clean	69.90±0.88%	<b>65.68±0.87%</b>	69.90±0.88%	<b>66.72±0.90%</b>
		40%	67.61±0.93%	<b>62.83±0.88%</b>	69.52±0.88%	<b>63.53±0.93%</b>
		Adversarial	<b>65.26±0.98%</b>	<b>64.18±0.86%</b>	<b>66.81±0.95%</b>	<b>66.33±0.84%</b>

MINIIMAGENET WITH FFGSM ATTACK

Note that the maximum perturbations adopted in FFGSM Attack are 2, 1 and 0.5, and the step size is set to 10/255.

Table 5: Average classification accuracies on MiniImageNet with FFGSM Attack (5-way, 1-shot)

Method	Backbone	Meta-testing	$\epsilon = 2$		$\epsilon = 1$		$\epsilon = 0.5$	
			Clean	Adversarial	Clean	Adversarial	Clean	Adversarial
MAML	32-32-32-32	Clean	48.47±1.77%	24.90±1.39%	48.47±1.77%	30.73±1.64%	48.47±1.77%	39.76±1.85%
		Adversarial	27.23±1.61%	23.73±1.47%	33.97±1.72%	29.20±1.67%	42.73±1.85%	38.63±1.78%
MAML-AD	32-32-32-32	Clean	40.63±1.69%	23.03±0.92%	42.90±1.88%	33.13±1.67%	42.43±1.80%	38.47±1.75%
		Adversarial	27.87±1.23%	28.70±1.65%	33.50±1.61%	33.47±1.67%	40.87±1.82%	39.23±1.82%
Matching Nets	64-64-64-64	Clean	43.87±0.41%	25.63±0.36%	43.87±0.41%	32.57±0.50%	43.87±0.41%	35.54±0.44%
		Adversarial	26.14±0.41%	28.96±0.39%	34.22±0.48%	33.92±0.41%	34.99±0.38%	36.23±0.43%
Relation Nets	64-96-128-256	Clean	49.67±0.85%	25.53±0.46%	49.67±0.85%	32.64±0.59%	49.67±0.85%	42.06±0.76%
		Adversarial	26.94±0.82%	24.51±0.47%	34.07±0.88%	29.99±0.61%	42.76±0.90%	39.37±0.75%
R2D2 (64C)	64-64-64-64	Clean	49.52±1.70%	20.51±0.32%	49.52±1.70%	22.06±0.85%	49.52±1.70%	32.08±1.60%
		Adversarial	24.46±1.37%	24.29±0.85%	29.71±1.53%	27.15±0.89%	37.86±1.76%	34.63±1.26%
R2D2	96-192-384-512	Clean	<b>51.80±1.70%</b>	20.06±0.26%	<b>51.80±1.70%</b>	21.19±0.54%	<b>51.80±1.70%</b>	31.91±1.42%
		Adversarial	22.68±1.35%	24.14±0.92%	26.98±1.48%	26.94±0.94%	40.07±1.79%	34.26±1.33%
ADML (Ours)	32-32-32-32	Clean	42.20±1.82%	<b>33.30±1.87%</b>	48.60±1.91%	<b>38.80±1.75%</b>	48.80±1.94%	<b>44.20±1.61%</b>
		Adversarial	<b>37.20±1.65%</b>	<b>31.00±1.71%</b>	<b>40.90±1.84%</b>	<b>35.70±1.78%</b>	<b>45.10±1.78%</b>	<b>39.70±1.91%</b>

Table 6: Average classification accuracies on MiniImageNet with FFGSM Attack (5-way, 5-shot)

Method	Backbone	Meta-testing	$\epsilon = 2$		$\epsilon = 1$		$\epsilon = 0.5$	
			Clean	Adversarial	Clean	Adversarial	Clean	Adversarial
MAML	32-32-32-32	Clean	61.45±0.91%	30.46±0.70%	61.46±0.91%	40.91±0.87%	61.46±0.91%	53.58±0.93%
		40%	54.29±0.94%	30.90±0.74%	57.77±0.92%	40.67±0.87%	59.85±0.88%	53.38±0.96%
		Adversarial	32.91±0.85%	31.90±0.87%	43.30±0.89%	41.18±0.96%	55.06±0.93%	51.63±1.01%
MAML-AD	32-32-32-32	Clean	57.55±0.98%	39.27±0.93%	58.74±0.94%	49.53±0.95%	59.63±0.95%	56.01±0.98%
		40%	55.38±0.93%	40.98±0.88%	57.35±0.91%	49.89±0.92%	59.39±0.94%	55.83±0.94%
		Adversarial	41.19±0.89%	41.17±0.96%	51.51±0.89%	49.27±0.94%	56.97±0.97%	54.42±0.99%
Matching Nets	64-64-64-64	Clean	55.99±0.47%	30.85±0.45%	55.99±0.47%	41.28±0.42%	55.99±0.47%	48.45±0.44%
		40%	46.32±0.41%	32.77±0.52%	51.66±0.45%	45.53±0.46%	53.27±0.40%	49.61±0.41%
		Adversarial	32.65±0.41%	33.56±0.50%	45.30±0.45%	46.74±0.52%	48.98±0.43%	50.30±0.45%
Relation Nets	64-96-128-256	Clean	63.85±0.73%	30.35±0.50%	63.85±0.73%	41.74±0.62%	63.85±0.73%	54.73±0.68%
		40%	53.54±0.83%	28.89±0.50%	56.15±0.78%	38.75±0.65%	60.59±0.74%	52.59±0.71%
		Adversarial	37.34±0.77%	31.31±0.53%	49.51±0.75%	42.63±0.66%	58.90±0.70%	54.67±0.71%
R2D2 (64C)	64-64-64-64	Clean	65.48±1.35%	20.81±0.45%	65.48±1.35%	23.44±0.95%	65.48±1.35%	38.97±1.62%
		40%	59.56±1.58%	28.71±0.98%	60.27±1.59%	36.07±1.40%	61.87±1.58%	48.33±1.50%
		Adversarial	30.83±1.65%	29.80±0.99%	40.40±1.68%	38.64±1.37%	51.03±1.56%	50.04±1.46%
R2D2	96-192-384-512	Clean	<b>68.42±1.28%</b>	20.61±0.42%	<b>68.42±1.28%</b>	22.97±0.86%	<b>68.42±1.28%</b>	39.53±1.78%
		40%	<b>60.68±1.48%</b>	26.90±1.05%	<b>61.14±1.60%</b>	33.47±1.14%	<b>63.77±1.64%</b>	48.47±1.51%
		Adversarial	26.50±1.37%	29.11±1.02%	35.25±1.49%	38.00±1.22%	51.75±1.56%	50.94±1.50%
ADML (Ours)	32-32-32-32	Clean	58.68±0.94%	<b>47.22±0.91%</b>	60.22±0.95%	<b>52.06±1.00%</b>	62.07±0.83%	<b>58.36±0.98%</b>
		40%	55.97±0.84%	<b>43.69±0.93%</b>	58.07±0.87%	<b>51.68±0.89%</b>	61.29±0.88%	<b>57.61±0.92%</b>
		Adversarial	<b>50.98±0.94%</b>	<b>45.04±0.92%</b>	<b>56.56±0.91%</b>	<b>52.44±0.97%</b>	<b>60.20±0.95%</b>	<b>60.25±0.98%</b>

## MINIIMAGENET WITH RFGSM ATTACK

Note that the maximum perturbations adopted in RFGSM Attack are 0.4, 0.2 and 0.1, the step size is set to 8/255 and the number of steps is 5.

Table 7: Average classification accuracies on MiniImageNet with RFGSM Attack (5-way, 1-shot)

Method	Backbone	Meta-testing	$\epsilon = 0.4$		$\epsilon = 0.2$		$\epsilon = 0.1$	
			Clean	Adversarial	Clean	Adversarial	Clean	Adversarial
MAML	32-32-32-32	Clean	48.47±1.77%	22.13±1.27%	48.47±1.77%	31.73±1.70%	48.47±1.77%	45.13±1.82%
		Adversarial	21.10±1.51%	22.17±1.46%	35.03±1.74%	30.00±1.73%	42.90±1.73%	41.23±1.89%
MAML-AD	32-32-32-32	Clean	41.53±1.80%	23.13±1.09%	42.53±1.87%	34.93±1.55%	43.07±1.86%	33.60±1.48%
		Adversarial	22.67±1.08%	32.07±1.69%	35.40±1.62%	37.27±1.72%	37.03±1.65%	38.63±1.70%
Matching Nets	64-64-64-64	Clean	43.87±0.41%	23.63±0.53%	43.87±0.41%	31.53±0.46%	43.87±0.41%	36.32±0.41%
		Adversarial	25.55±0.56%	24.61±0.49%	33.37±0.42%	34.70±0.39%	36.59±0.46%	35.88±0.51%
Relation Nets	64-96-128-256	Clean	49.67±0.85%	22.53±0.39%	49.67±0.85%	33.07±0.59%	49.67±0.85%	44.69±0.78%
		Adversarial	23.38±0.75%	22.45±0.40%	34.30±0.88%	31.04±0.63%	45.07±0.88%	42.12±0.78%
R2D2 (64C)	64-64-64-64	Clean	49.52±1.70%	20.34±0.35%	49.52±1.70%	22.78±1.99%	49.52±1.70%	38.24±1.66%
		Adversarial	22.08±1.50%	24.21±0.78%	30.29±1.74%	29.68±1.23%	42.11±1.73%	39.42±1.34%
R2D2	96-192-384-512	Clean	<b>51.80±1.70%</b>	20.12±0.31%	<b>51.80±1.70%</b>	22.27±0.81%	<b>51.80±1.70%</b>	39.23±1.56%
		Adversarial	20.94±1.31%	22.91±0.72%	28.82±1.59%	29.23±1.04%	44.37±1.88%	38.95±1.36%
ADML (Ours)	32-32-32-32	Clean	43.20±1.74%	<b>33.10±1.80%</b>	44.50±1.88%	<b>42.70±1.68%</b>	48.90±1.64%	<b>45.31±1.68%</b>
		Adversarial	<b>33.10±1.57%</b>	<b>36.90±1.80%</b>	<b>41.90±1.79%</b>	<b>40.60±1.84%</b>	<b>45.70±1.68%</b>	<b>42.60±1.84%</b>

Table 8: Average classification accuracies on MiniImageNet with RFGSM Attack (5-way, 5-shot)

Method	Backbone	Meta-testing	$\epsilon = 0.4$		$\epsilon = 0.2$		$\epsilon = 0.1$	
			Clean	Adversarial	Clean	Adversarial	Clean	Adversarial
MAML	32-32-32-32	Clean	61.45±0.91%	25.65±0.62%	61.46±0.91%	40.14±0.88%	61.45±0.91%	55.62±0.97%
		40%	52.51±0.93%	24.80±0.69%	57.11±0.91%	42.15±0.91%	60.63±0.91%	56.29±0.96%
		Adversarial	26.92±0.74%	27.31±0.71%	44.26±0.96%	42.27±0.91%	58.11±0.99%	54.54±0.98%
MAML-AD	32-32-32-32	Clean	56.31±0.99%	27.11±0.70%	57.52±0.99%	47.70±0.97%	58.85±0.95%	54.21±0.92%
		40%	50.73±0.95%	36.37±0.84%	57.06±0.90%	51.36±0.98%	58.89±0.91%	55.92±0.93%
		Adversarial	29.69±0.77%	46.85±0.93%	50.68±0.92%	52.72±0.94%	55.77±0.97%	54.85±0.97%
Matching Nets	64-64-64-64	Clean	55.99±0.47%	24.11±0.56%	55.99±0.47%	38.53±0.41%	55.99±0.47%	52.13±0.51%
		40%	47.53±0.46%	26.81±0.50%	52.26±0.44%	36.27±0.50%	54.83±0.41%	54.00±0.46%
		Adversarial	25.33±0.52%	28.45±0.55%	40.52±0.39%	36.61±0.42%	52.92±0.46%	53.33±0.48%
Relation Nets	64-96-128-256	Clean	63.85±0.73%	24.75±0.42%	63.85±0.73%	40.59±0.64%	63.85±0.73%	56.94±0.70%
		40%	52.29±0.86%	24.24±0.44%	56.17±0.81%	39.00±0.63%	61.36±0.74%	55.98±0.73%
		Adversarial	28.69±0.74%	27.60±0.47%	48.05±0.88%	43.20±0.67%	60.38±0.77%	56.11±0.72%
R2D2 (64C)	64-64-64-64	Clean	65.48±1.35%	20.38±0.31%	65.48±1.35%	24.89±1.10%	65.48±1.35%	48.47±1.56%
		40%	59.35±1.57%	26.89±0.89%	60.35±1.50%	38.63±1.40%	61.83±1.53%	53.38±1.51%
		Adversarial	27.11±1.36%	28.80±0.98%	38.69±1.69%	42.28±1.27%	44.72±1.75%	54.53±1.50%
R2D2	96-192-384-512	Clean	<b>68.42±1.28%</b>	20.29±0.26%	<b>68.42±1.28%</b>	24.36±0.92%	<b>68.42±1.28%</b>	49.64±1.68%
		40%	<b>60.04±1.53%</b>	24.20±0.75%	<b>62.15±1.54%</b>	35.73±1.27%	<b>64.54±1.61%</b>	54.33±1.62%
		Adversarial	22.57±1.33%	27.13±0.86%	36.22±1.64%	40.39±1.14%	56.92±1.69%	55.49±1.51%
ADML (Ours)	32-32-32-32	Clean	56.93±0.89%	<b>43.62±0.93%</b>	60.64±1.03%	<b>53.54±0.93%</b>	62.40±0.91%	<b>58.33±0.95%</b>
		40%	55.09±0.88%	<b>40.54±0.91%</b>	58.43±0.90%	<b>53.83±0.93%</b>	62.37±0.88%	<b>58.08±0.94%</b>
		Adversarial	<b>49.36±0.87%</b>	<b>47.78±0.83%</b>	<b>57.70±0.92%</b>	<b>56.77±0.89%</b>	<b>61.03±0.93%</b>	<b>58.44±0.92%</b>

## MINIIMAGENET WITH RPGD ATTACK

Note that the maximum perturbations adopted in RPGD Attack are 1.6, 0.8 and 0.4, the step size is set to  $2/255$  and the number of steps is 40.

Table 9: Average classification accuracies on MiniImageNet with RPGD Attack (5-way, 1-shot)

Method	Backbone	Meta-testing	$\epsilon = 1.6$		$\epsilon = 0.8$		$\epsilon = 0.4$	
			Clean	Adversarial	Clean	Adversarial	Clean	Adversarial
MAML	32-32-32-32	Clean	48.47±1.77%	25.07±1.44%	48.47±1.77%	32.13±1.67%	48.47±1.77%	41.37±1.78%
		Adversarial	27.97±1.59%	24.77±1.54%	36.30±1.77%	30.57±1.74%	43.20±1.82%	39.57±1.76%
MAML-AD	32-32-32-32	Clean	41.03±1.74%	24.03±0.87%	42.27±1.85%	33.87±1.59%	42.63±1.87%	36.63±1.67%
		Adversarial	29.93±1.49%	29.83±1.67%	35.37±1.64%	35.97±1.76%	37.57±1.65%	38.07±1.77%
Matching Nets	64-64-64-64	Clean	43.87±0.41%	26.36±0.43%	43.87±0.41%	31.08±0.40%	43.87±0.41%	35.26±0.56%
		Adversarial	29.36±0.36%	28.34±0.39%	33.52±0.58%	30.02±0.52%	36.60±0.51%	33.75±0.48%
Relation Nets	64-96-128-256	Clean	49.67±0.85%	26.77±0.47%	49.67±0.85%	34.99±0.64%	49.67±0.85%	43.67±0.78%
		Adversarial	28.19±0.80%	25.32±0.48%	36.34±0.88%	32.19±0.63%	44.18±0.87%	41.01±0.77%
R2D2 (64C)	64-64-64-64	Clean	49.52±1.70%	20.58±0.39%	49.52±1.70%	23.46±0.96%	49.52±1.70%	34.92±1.62%
		Adversarial	24.90±1.43%	25.30±0.96%	31.70±1.67%	28.27±1.21%	40.52±1.74%	37.49±1.28%
R2D2	96-192-384-512	Clean	<b>51.80±1.70%</b>	20.19±0.32%	<b>51.80±1.70%</b>	22.80±0.82%	<b>51.80±1.70%</b>	34.93±1.56%
		Adversarial	22.04±1.26%	25.30±0.97%	30.63±1.63%	28.00±1.04%	42.80±1.85%	37.18±1.44%
ADML (Ours)	32-32-32-32	Clean	45.20±1.85%	<b>40.00±1.86%</b>	43.90±1.82%	<b>40.50±1.72%</b>	48.90±1.82%	<b>44.52±1.96%</b>
		Adversarial	<b>37.90±1.77%</b>	<b>34.90±1.74%</b>	<b>42.60±1.64%</b>	<b>39.30±1.83%</b>	<b>45.40±1.92%</b>	<b>41.50±1.83%</b>

Table 10: Average classification accuracies on MiniImageNet with RPGD Attack (5-way, 5-shot)

Method	Backbone	Meta-testing	$\epsilon = 1.6$		$\epsilon = 0.8$		$\epsilon = 0.4$	
			Clean	Adversarial	Clean	Adversarial	Clean	Adversarial
MAML	32-32-32-32	Clean	61.45±0.91%	32.24±0.78%	61.45±0.91%	43.41±0.92%	61.45±0.91%	54.96±0.89%
		40%	55.39±0.90%	31.55±0.78%	57.41±0.95%	43.97±1.00%	59.99±0.90%	54.89±0.94%
		Adversarial	34.48±0.83%	33.69±0.83%	46.48±0.87%	44.67±0.92%	57.59±0.91%	54.05±0.93%
MAML-AD	32-32-32-32	Clean	56.31±0.98%	42.80±0.88%	58.56±0.93%	51.43±0.97%	60.06±0.95%	56.18±0.93%
		40%	53.99±0.94%	43.58±0.88%	57.23±0.97%	51.61±1.00%	58.35±0.96%	56.05±0.94%
		Adversarial	46.41±0.90%	43.90±0.90%	53.87±0.93%	52.68±0.92%	57.33±0.91%	55.93±0.92%
Matching Nets	64-64-64-64	Clean	55.99±0.47%	33.50±0.42%	55.99±0.47%	44.11±0.36%	55.99±0.47%	53.31±0.45%
		40%	49.33±0.45%	34.58±0.40%	51.56±0.42%	44.89±0.39%	54.02±0.47%	53.78±0.53%
		Adversarial	35.02±0.40%	34.67±0.39%	43.08±0.40%	45.84±0.42%	52.15±0.49%	51.36±0.44%
Relation Nets	64-96-128-256	Clean	63.85±0.73%	32.15±0.54%	63.85±0.73%	44.52±0.64%	63.85±0.73%	56.53±0.68%
		40%	54.06±0.82%	30.69±0.54%	57.20±0.80%	41.96±0.70%	58.53±0.74%	54.91±0.75%
		Adversarial	39.64±0.79%	34.20±0.57%	51.91±0.74%	46.04±0.66%	59.12±0.70%	56.13±0.70%
R2D2 (64C)	64-64-64-64	Clean	65.48±1.35%	21.06±0.55%	65.48±1.35%	26.08±1.14%	65.48±1.35%	43.64±1.54%
		40%	59.79±1.53%	30.65±1.07%	60.54±1.59%	39.01±1.48%	62.36±1.61%	37.49±1.28%
		Adversarial	33.70±1.60%	32.25±1.25%	42.28±1.64%	42.32±1.40%	54.35±1.70%	53.65±1.35%
R2D2	96-192-384-512	Clean	<b>68.42±1.28%</b>	20.85±0.53%	<b>68.42±1.28%</b>	25.35±1.13%	<b>68.42±1.28%</b>	44.63±1.70%
		40%	<b>60.66±1.56%</b>	28.77±1.01%	<b>61.48±1.60%</b>	36.32±1.31%	<b>63.15±1.59%</b>	52.01±1.52%
		Adversarial	27.59±1.39%	31.16±1.11%	40.29±1.51%	41.16±1.34%	55.50±1.63%	54.00±1.50%
ADML (Ours)	32-32-32-32	Clean	58.40±0.94%	<b>48.62±0.97%</b>	59.94±0.86%	<b>54.62±1.01%</b>	61.54±0.97%	<b>57.20±0.94%</b>
		40%	56.28±0.94%	<b>45.83±0.90%</b>	58.27±0.92%	<b>53.60±0.99%</b>	60.68±0.93%	<b>56.33±0.98%</b>
		Adversarial	<b>52.28±0.88%</b>	<b>45.82±0.94%</b>	<b>57.36±0.96%</b>	<b>54.39±0.96%</b>	<b>59.68±0.92%</b>	<b>57.16±1.00%</b>