JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Paper tag: TUE-AM-368

CUDA: **C**onvolution-based **U**nlearnable **Da**tasets

Vinu Sankar Sadasivan[1]   Mahdi Soltanolkotabi[2]   Soheil Feizi[1]

University of Maryland 1856[1]   University of Southern California 1880[2]

@imVinusankars

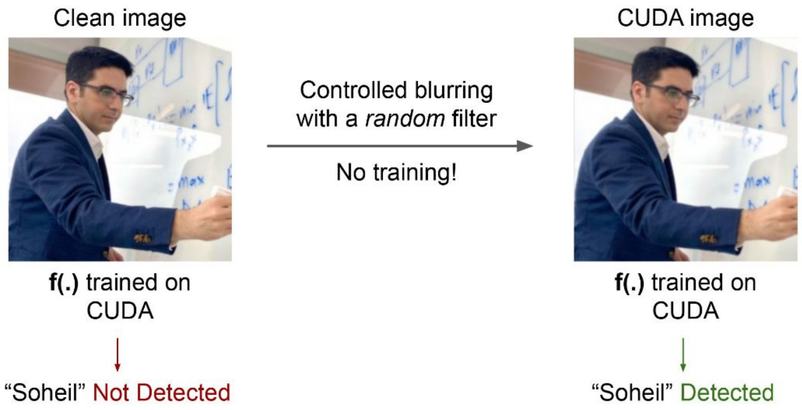
One-minute Pitch

Unlearnable images

- Attacker adds noise to training images
- Defender trains **f(.)** w/ poisoned data
- **f(.)** fails to classify clean data
- **f(.)** classifies poisoned data
- Privacy for facial recognition

CUDA

- Novel non-additive noise
- Controlled blurring creates shortcuts
- Model-free; 20-100x faster
- Very robust



Clean image

f(.) trained on CUDA

"Soheil" Not Detected

Controlled blurring with a *random* filter

No training!

CUDA image

f(.) trained on CUDA

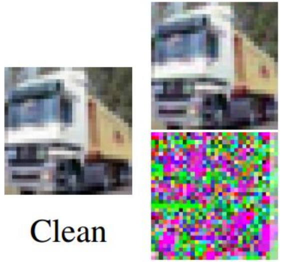
"Soheil" Detected

@imVinusankars

Previous Works Need Optimizations

- EM (Huang et al., ICLR 21) – min-min ⇦
- TAP (Fowl et al., NeurIPS 21) – min-max
- REM (Fu et al., ICLR 22) – min-min-max

$$\min_{\theta} \frac{1}{n} \sum_{i=1}^{n} \min_{\|\delta_i\| \le \rho_u} \ell(f'_{\theta}(x_i + \delta_i), y_i)$$



@imVinusankars

Previous Works Need Optimizations

- EM (Huang et al., ICLR 21) – min-min
- TAP (Fowl et al., NeurIPS 21) – min-max ⇦
- REM (Fu et al., ICLR 22) – min-min-max

$$\min_{\theta} \frac{1}{n} \sum_{i=1}^{n} \max_{||\delta_i|| \leq \rho_a} \ell(f_\theta(x_i + \delta_i), y_i)$$

– Expensive

– Not robust to AT

Previous Works Need Optimizations

- EM (Huang et al., ICLR 21) – min-min
- TAP (Fowl et al., NeurIPS 21) – min-max
- REM (Fu et al., ICLR 22) – min-min-max ⇦

$$\min_{\theta} \frac{1}{n} \sum_{i=1}^{n} \min_{\|\delta_i^u\| \le \rho_u} \max_{\|\delta_i^a\| \le \rho_a} \ell(f'_\theta(x_i + \delta_i^u + \delta_i^a), y_i)$$

– Expensive

– Not transferable to DenseNet-121

– Sensitive to AT hyperparameters

– REM breaks w/ grayscaling

CUDA is simple

- Clean – (x, y)
- Filter – s_y randomly generated for class y
- x’ = x□s_y where □ is convolution
- CUDA image – (x’ / MAX(x’), y)
- Network learns relation b/w s_y & y

$$s_y = \begin{pmatrix} z & z & z \\ z & 1 & z \\ z & z & z \end{pmatrix}_{k \times k\ (k = 3)}, \quad z \sim U(0, p_b)$$

– k is filter size

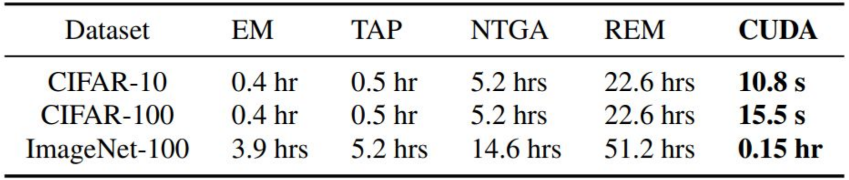
– p_b is blur parameter

– Controls blurring

@imVinusankars

CUDA is effective

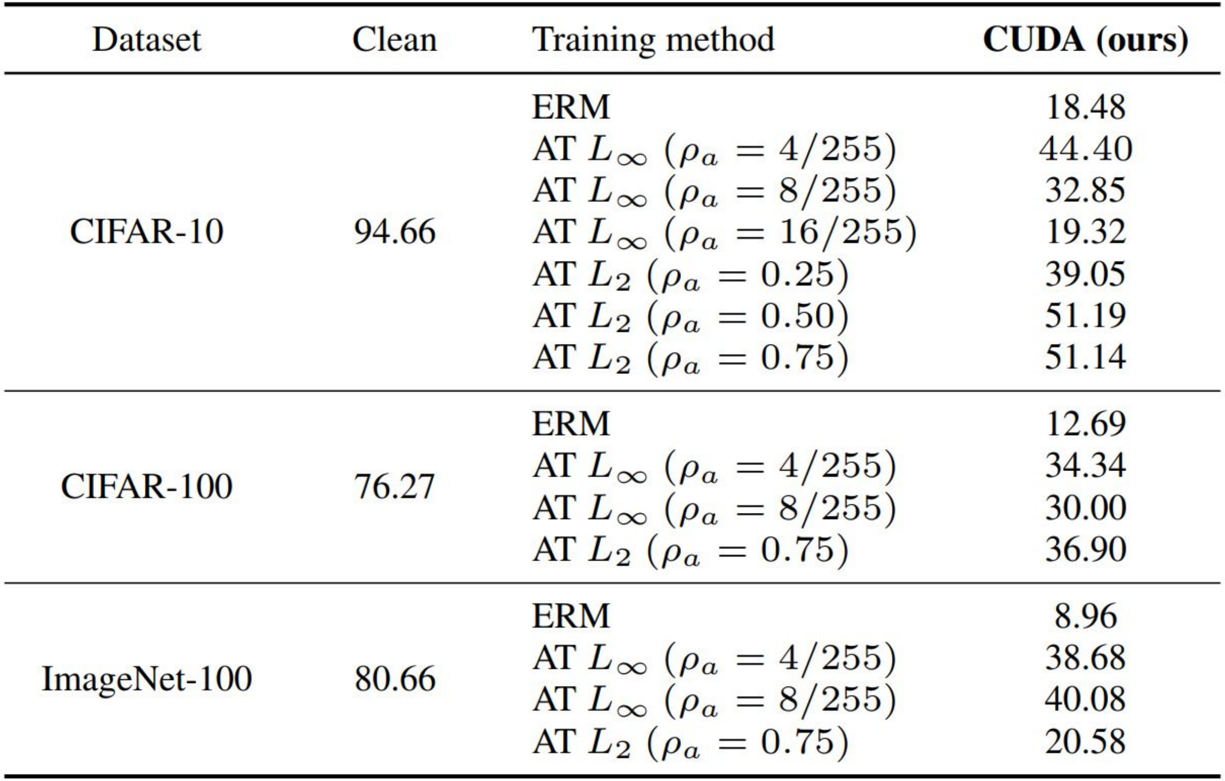
- Model-free; 20-100x faster
- Robust to AT, augs, adaptive defenses
- Transferable to different networks
- Theory



Dataset	EM	TAP	NTGA	REM	**CUDA**
CIFAR-10	0.4 hr	0.5 hr	5.2 hrs	22.6 hrs	**10.8 s**
CIFAR-100	0.4 hr	0.5 hr	5.2 hrs	22.6 hrs	**15.5 s**
ImageNet-100	3.9 hrs	5.2 hrs	14.6 hrs	51.2 hrs	**0.15 hr**

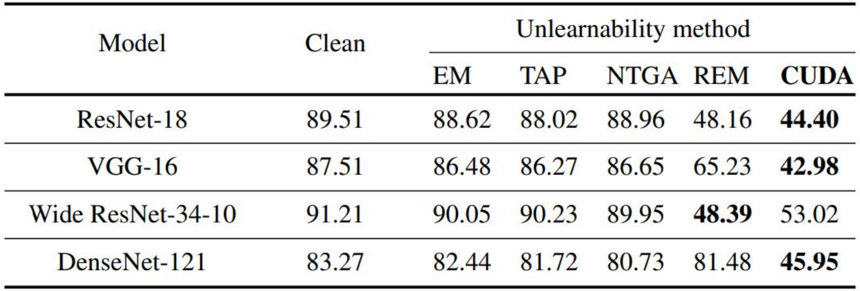
Theorem (Informal) Let D be a Gaussian mixture with two modes. P_D denotes optimal Bayesian classifier trained on D. $\tau_D(P)$ denote accuracy of P on D. For every clean D, $\exists$ D' s.t. $\tau_D(P_{D'}) < \tau_D(P_D)$.

CUDA is robust to AT

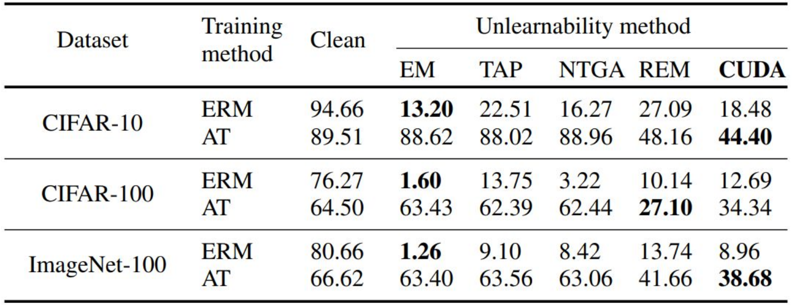


Dataset	Clean	Training method	**CUDA (ours)**
CIFAR-10	94.66	ERM	18.48
		AT L_∞ $(\rho_a = 4/255)$	44.40
		AT L_∞ $(\rho_a = 8/255)$	32.85
		AT L_∞ $(\rho_a = 16/255)$	19.32
		AT L_2 $(\rho_a = 0.25)$	39.05
		AT L_2 $(\rho_a = 0.50)$	51.19
		AT L_2 $(\rho_a = 0.75)$	51.14
CIFAR-100	76.27	ERM	12.69
		AT L_∞ $(\rho_a = 4/255)$	34.34
		AT L_∞ $(\rho_a = 8/255)$	30.00
		AT L_2 $(\rho_a = 0.75)$	36.90
ImageNet-100	80.66	ERM	8.96
		AT L_∞ $(\rho_a = 4/255)$	38.68
		AT L_∞ $(\rho_a = 8/255)$	40.08
		AT L_2 $(\rho_a = 0.75)$	20.58

CUDA is robust to networks & datasets



Model	Clean	Unlearnability method				
		EM	TAP	NTGA	REM	CUDA
ResNet-18	89.51	88.62	88.02	88.96	48.16	**44.40**
VGG-16	87.51	86.48	86.27	86.65	65.23	**42.98**
Wide ResNet-34-10	91.21	90.05	90.23	89.95	**48.39**	53.02
DenseNet-121	83.27	82.44	81.72	80.73	81.48	**45.95**

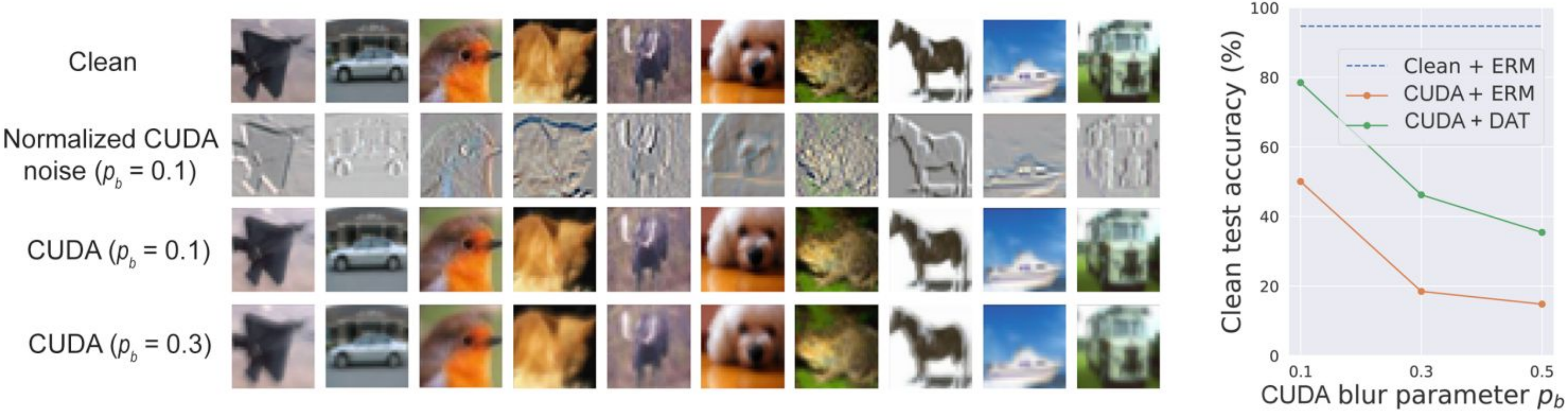


Dataset	Training method	Clean	Unlearnability method				
			EM	TAP	NTGA	REM	CUDA
CIFAR-10	ERM AT	94.66 89.51	**13.20** 88.62	22.51 88.02	16.27 88.96	27.09 48.16	18.48 **44.40**
CIFAR-100	ERM AT	76.27 64.50	**1.60** 63.43	13.75 62.39	3.22 62.44	10.14 **27.10**	12.69 34.34
ImageNet-100	ERM AT	80.66 66.62	**1.26** 63.40	9.10 63.56	8.42 63.06	13.74 41.66	8.96 **38.68**

Architectures	Resnet-18, VGG-16, Wide ResNet-34-10, DenseNet-121, MobileNet-V2, EfficientNet-V2-S, DeiT
Datasets	CIFAR-10, CIFAR-100, ImageNet-100, Tiny ImageNet

CUDA is robust to defenses

- Grayscaling, mixup, cutmix, cutout, auto augment, other regularizations, randomized smoothing
- Adaptive defense – Deconvolution-based Adversarial Training
- Stealth v/ Unlearnability trade-off



JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Paper tag: TUE-AM-368

CUDA: Convolution-based Unlearnable Datasets

Vinu Sankar Sadasivan[1]   Mahdi Soltanolkotabi[2]   Soheil Feizi[1]

1   2

@imVinusankars