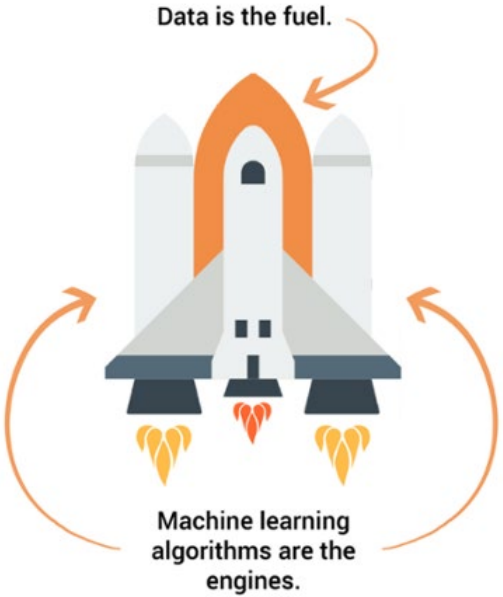
UCLA Samueli School of Engineering

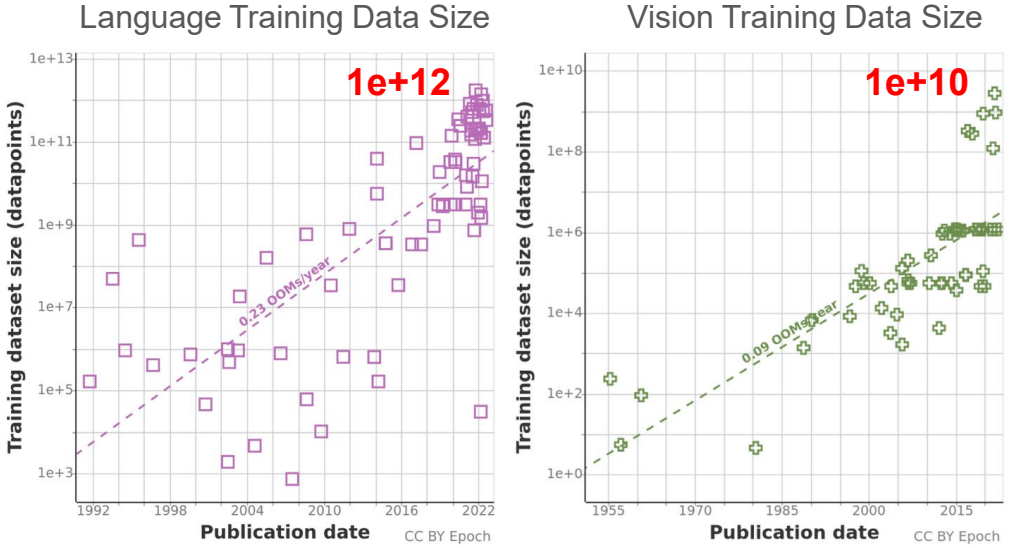
Efficient and Robust Deep Learning on Large Data

Yu Yang
PhD Student, Computer Science Department

This talk includes joint work with:
Tian Yu Liu, Eric Gan, Hao Kao,
Gintare Karolina Dziugaite,
Besmira Nushi, Hamid Palangi,
Baharan Mirzasoleiman (advisor)

Data is the new fuel!





Pablo Villalobos and Anson Ho (2022), "Trends in Training Dataset Sizes".

Problem 1: Large Data Makes Training Expensive

Example: ChatGPT is fine-tuned from **GPT-3**
Training **GPT-3** used 45TB data

Energy Consumption: 1,287 MWh
→ 17.8x average American yearly energy consumption!

CO2 Emission: 552 tons
→ same as driving a car from Earth to the Moon and then back!

Problem 2: Large Data is Vulnerable to Poisoning Attacks

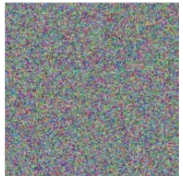
Large data is often crawled from the internet, thus it's vulnerable to data poisoning attacks:

Training:

Original

+

Imperceptible perturbation



Poison

A small set of poisons + A large set of clean data → Train → Poisoned model

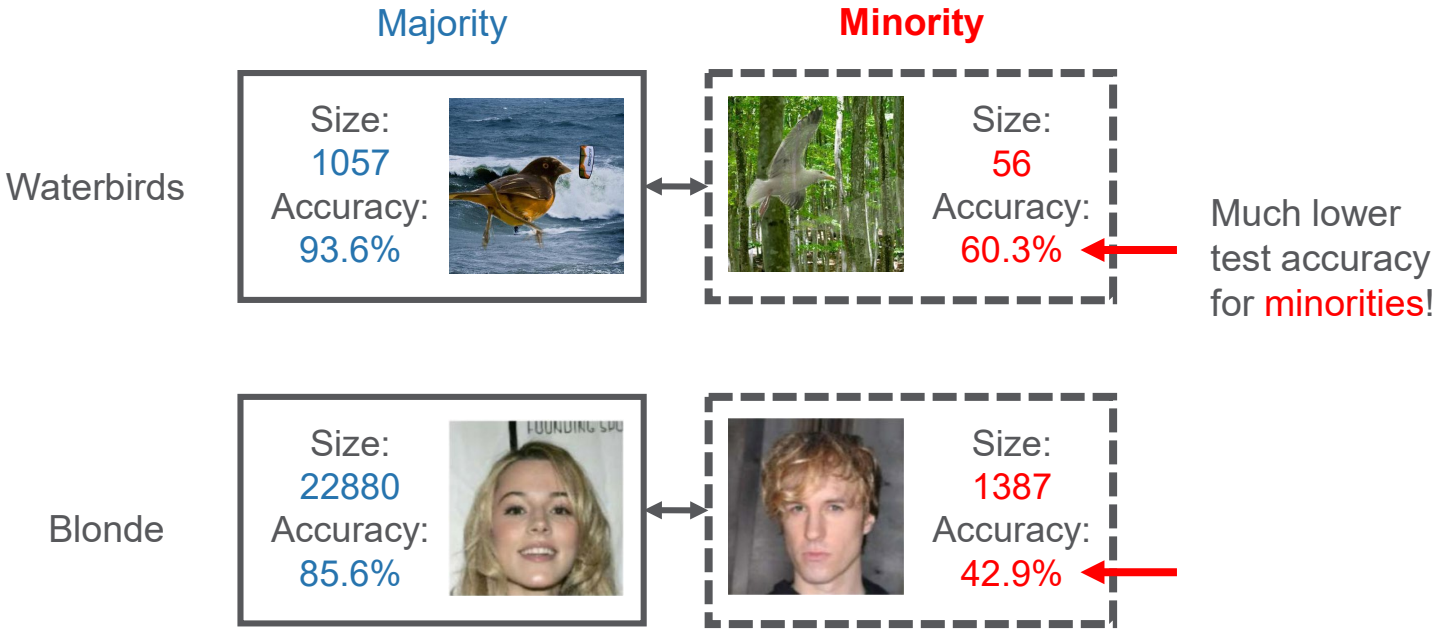
Testing:

Target

(Truck)

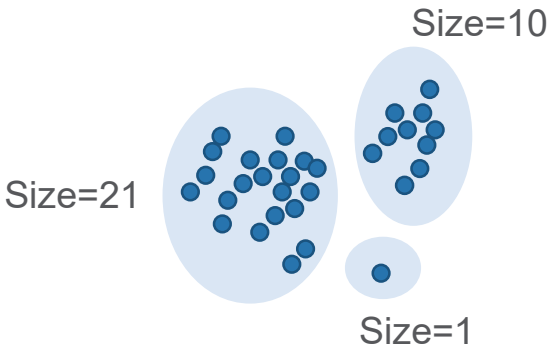
→ Poisoned model → Prediction: Bird

Problem 3: Large Real-world Data are Biased toward the Majority



My research addresses these problems by developing theoretically rigorous methods to improve efficiency and robustness of learning from large data

Gradient information Can Help Address the Above Problems!



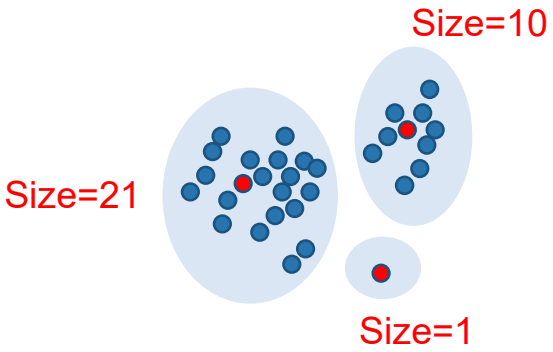
Clustering the training examples by their gradients gives lots of useful information.

This is a submodular problem:

$$F(S^*) = \sum_{i \in V} \min_{j \in S^*} \|\nabla f_i(w) - \nabla f_j(w)\| \leq \epsilon$$

very fast to solve with a greedy algorithm which guarantees a near-optimal solution!

1. Fast Training by Using Only Centers of Gradient Clusters

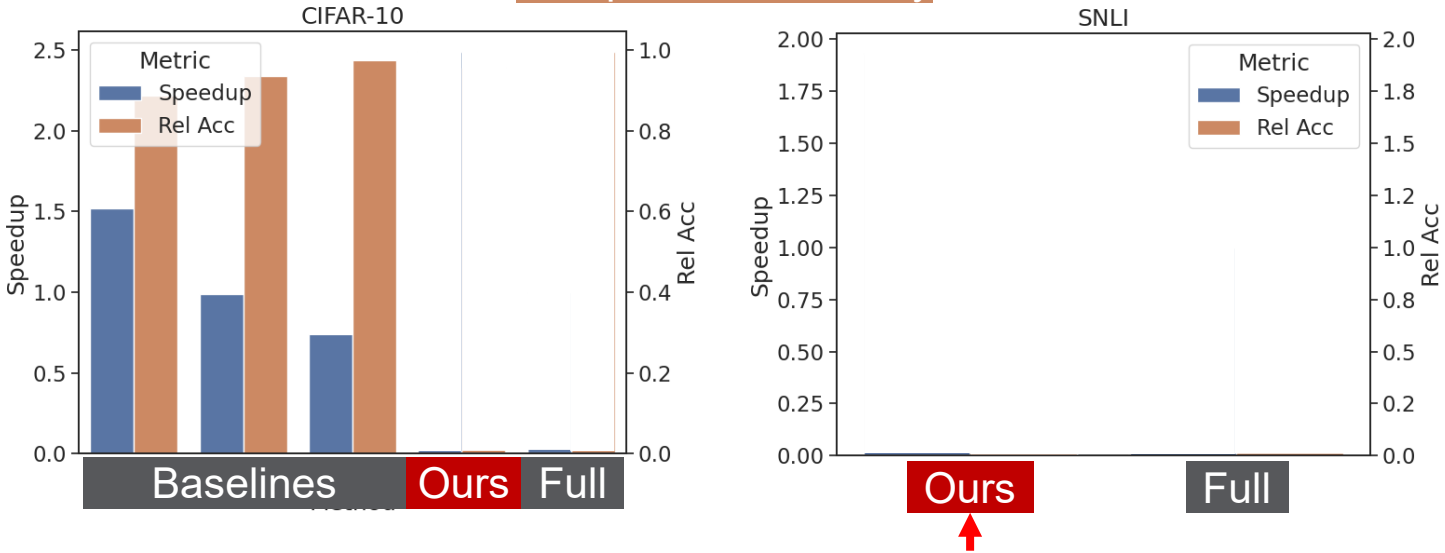


We use centers and sizes of clusters to estimate the gradients of their clusters to speed up the training.

1. Fast Training by Training Only Centers of Gradient Clusters

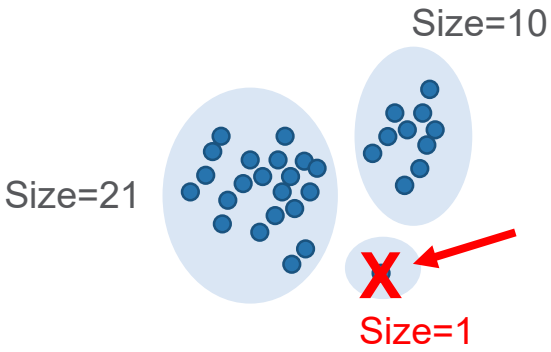
2x Faster

Comparable Accuracy



The only algorithm speeds up training on very large data and models with a theoretical guarantee!

2. Robust Training against Data Poisoning Attacks



We removes size-1 clusters which usually contain examples with outlier gradients to prevent data poisoning attacks.

2. Robust Training against Data Poisoning Attacks

Defend all kinds of attacks!

Attack	Senario	Undefended		Defended	
		Att Succ.↑	Test Acc.↑	Att Succ.↓	Test Acc.↑
Gradient Matching	from-scratch	45%	94.95%	1%	90.26%
Sleeper Agent (backdoor)	from-scratch	78.54%	94.42%	11.55%	88.28%
Bullseye Polytope	transfer	86%	94.69%	1%	94.80%
Feature Collision	transfer	40%	94.68%	0%	94.81%
Bullseye Polytope	finetune	80%	92.24%	0%	92.38%

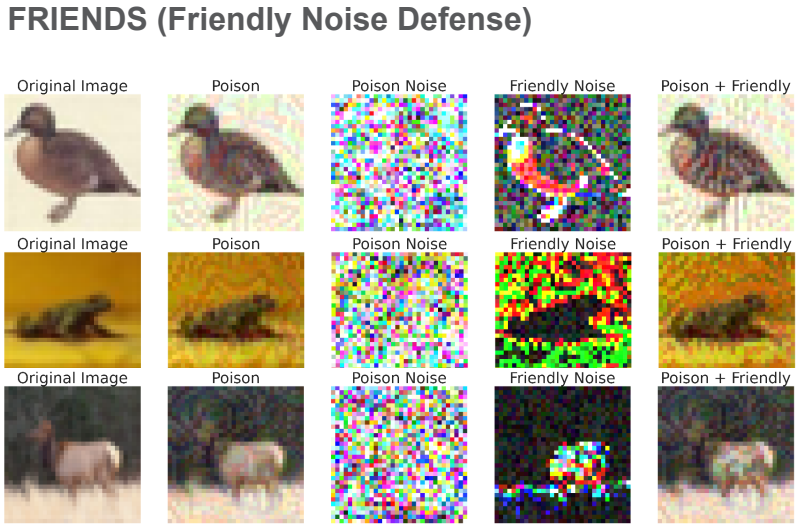
Breaks the attacks!

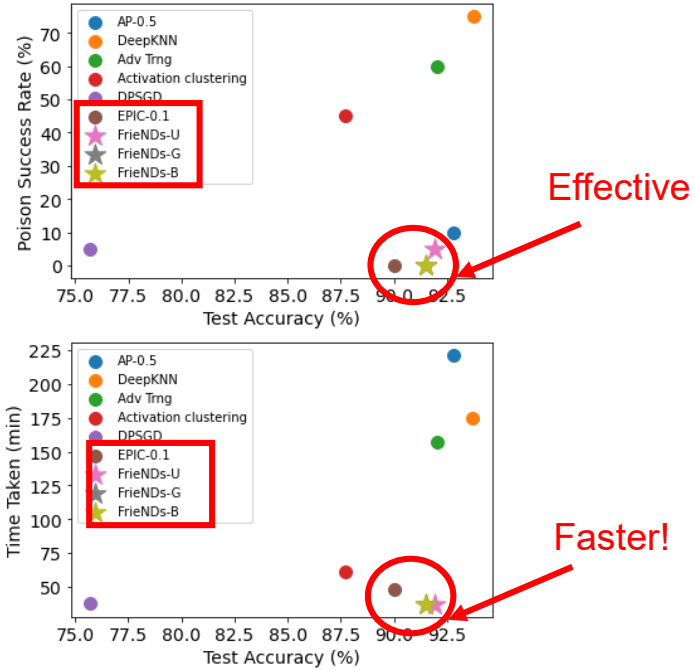
Theoretically guaranteed good generalization!

Epoch	Defense	Attack Succ.↓	Test Acc.↑	Time(hr:min)
40	None	25%	92.48%	00:15
40	DeepKNN (Peri et al., 2020)	21%	91.86%	02:25
40	Spectral Signatures (Tran et al., 2018)	17%	90.13%	00:40
40	Activation Clustering (Chen et al., 2019)	9%	84.20%	00:31
40	Diff. Priv. SGD (Hong et al., 2020)	2%	70.34%	00:16
40	Adv. Poisoning-0.25 (Geiping et al., 2021a)	4%	91.48%	01:53
40	Adv. Poisoning-0.5 (Geiping et al., 2021a)	1%	90.67%	02:02
40	Adv. Poisoning-0.75 (Geiping et al., 2021a)	0%	87.97%	02:26
40	EPIC-0.1 (Proposed)	2.7%±0.6%	90.92%±0.26%	00:22
40	Ours EPIC-0.2 (Proposed)	1.3%±0.6%	88.95%±0.08%	00:19
40	EPIC-0.3 (Proposed)	1.0%±0.0%	87.03%±0.11%	00:17

6x faster!

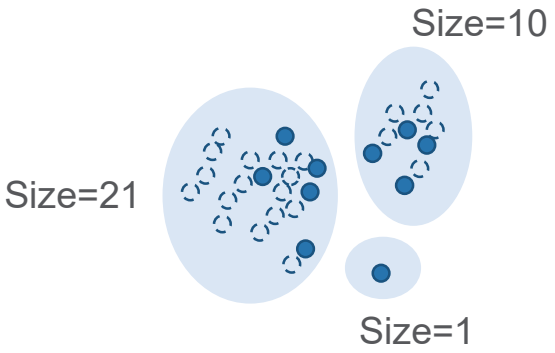
(Follow-up) Robust Training against Data Poisoning Attacks





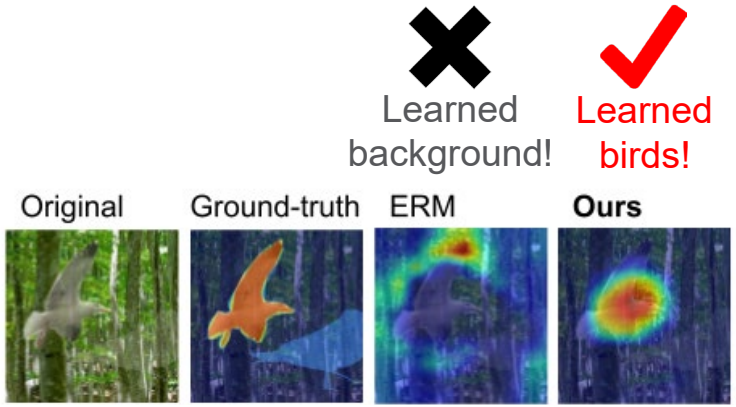
Friendly Noise against Adversarial Noise: A Powerful Defense against Data Poisoning Attacks. (NeurIPS 2022)

3. Improving Performance on Minority by Balancing the Gradient Clusters



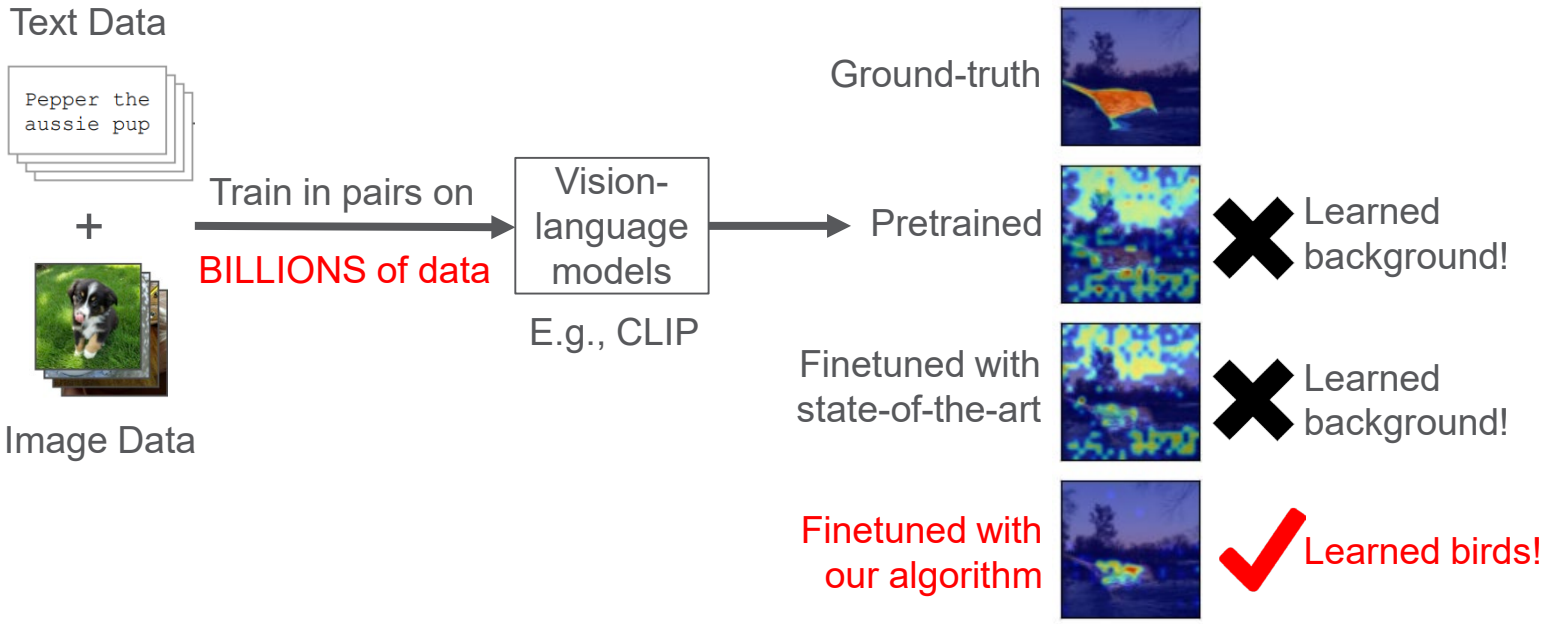
We make gradient clusters balanced in size to improve the performance on minorities.

3. Improving Performance on Minority by Balancing the Gradient Clusters



→ Improve the accuracy of minority by 20-40%!

(Follow-up) We Fix Large Vision-Language Models Too!



Takeaways

We used gradient clustering to solve the following major problems in deep learning:

Problem 1: Large Data Makes Training *Expensive!*
→ 2x speedup + comparable accuracy!

Problem 2: Large Data is *Vulnerable* to Poisoning Attacks*!*
→ ~0% attack success rate + 6x faster than other defenses!

Problem 3: Large Data are *Biased* toward the Majority!
→ Improves worst-group performance by 20-40%!

Thank you!