DP-Mix: Mixup-based Data Augmentation for Differentially Private Learning

Wenxuan Bao[1], Francesco Pittaluga[2], Vijay Kumar B G[2], Vincent Bindschaedler[1]

[1]University of Florida, [2]NEC Labs America

Motivation

- Multiple sample data augmentation like Mixup boost machine learning performance but face challenges in Differential Private Machine Learning (DPML) due to sensitivity issues.
- Diffusion model is a powerful model to generate high-quality images, but it is not clear that how to fit them into DPML to enhance performance.
- We propose techniques: DP-MIX$_{self}$ and DP-MIX$_{diff}$ to apply Mixup and Diffusion model in DPML and show it surpasses prior SoTA **at no extra privacy cost.**

Background: Mixup Data Augmentation

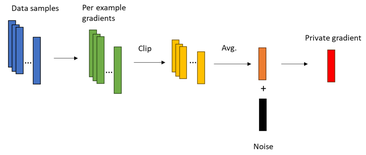
$$\begin{aligned}\hat{x} &= \lambda x_i + (1-\lambda)x_j,\\ \hat{y} &= \lambda y_i + (1-\lambda)y_j,\end{aligned}$$

where $\lambda \in [0,1]$ is a random number

Image source: https://hoya012.github.io/blog/Bag-of-Tricks-for-Image-Classification-with-Convolutional-Neural-Networks-Review/

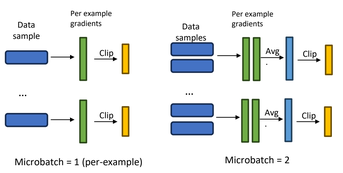
Background: DP-SGD[1]

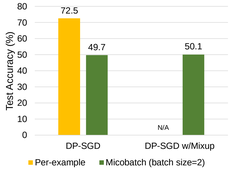
- Compute per example gradients.
- Clip them to norm ***C***.
- Average clipped gradients
- Add noise to average gradient.



Microbatch

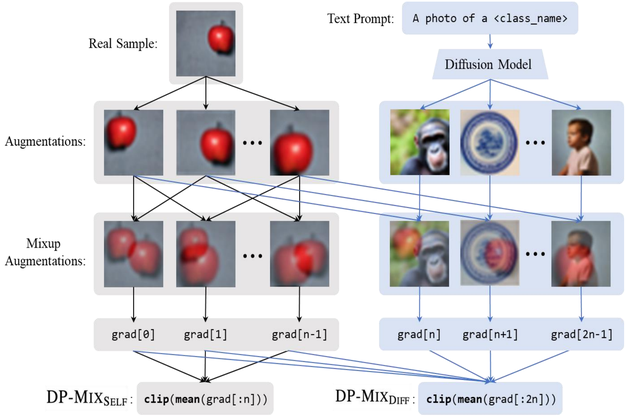
- Instead of using per-example gradient clipping, there is other method [2] that could clip the average of microbatch's gradients.
- Microbatch has a drawback: its sensitivity to adding or removing a sample is ***2C*** (as pointed out by Ponomareva et al. [3]), requiring more noise for the same privacy budget as per-example methods.
- To apply Mixup into DPML, the most straightforward method is to use Microbatch while our experiments show that it falls to improve performance with modest privacy budget (i.e, 8) as shown in the following bar figure.





Proposed Methods

- **DP-MIX$_{self}$** : Apply mixup to augmentations of real samples and then clip the average of those samples' gradients.
- **DP-MIX$_{diff}$** : Pretrain a private model on a public dataset using a Diffusion model. Generate diffusion samples with text prompts like "A photo of a <class name>". Mix these samples with augmented real samples and clip the average of their gradients



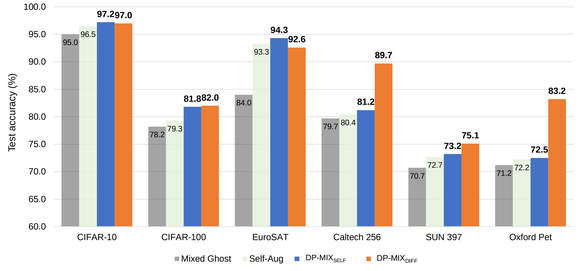
Our Contributions

1. We show empirically the straightforward application of Mixup i.e., using Microbatch, fails to improve model's performance.
2. We propose a technique called **DP-MIXself** to apply Mixup in DP-SGD by using Mixup to self-augmentations of one training sample. This method achieves **SoTA** performance for training from scratch and finetuning pre-trained models.
3. We also propose second technique called **DP-MIXdiff** to further enhance performance by using a text-to-image diffusion model to generate class-specific synthetic examples. We can mixup those diffusion samples with real training sample to achieve new **SoTA** performance **with no addition privacy cost.**

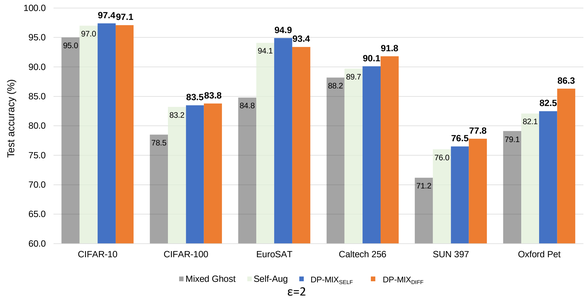
Results

Main results

- Our proposed **DP-MIX$_{self}$** and **DP-MIX$_{diff}$** achieve better results than prior SoTA. For Caltech 256 and Oxford Pet, we achieve about ***9%*** performance boost for ε=1.
- With privacy budget growth, performance boost decrease but still exists.



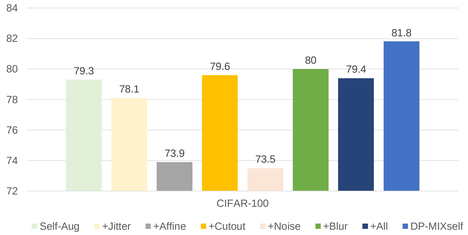
ε=1



ε=2

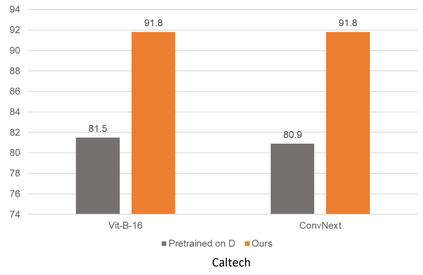
Ablation study | Other augmentations

- Other single-sample augmentations do **not** provide anywhere as much of an improvement as mixup



Ablation study | DP-Mix$_{diff}$ vs Pretraining with Diffusion Data

- Pre-training on sample diffusion models does **not** improve performance; mixing up training samples with them does.



Take aways and Future works

- We show how to apply **mixup** for DP training of ML models and demonstrate it surpasses the prior SoTA **at no extra privacy cost**.
- For future work, other multiple data augmentations could be applied to DPML too. We use public data to pretrain Diffusion model and private model but it is still an open question that how to use public data in the most efficient way for DPML.

Acknowledgments

This work was supported in part by the National Science Foundation under CNS-2055123. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] Abadi, Martin, et al. "Deep learning with differential privacy." Proceedings of the 2016 ACM SIGSAC conference on computer and communications security. 2016.

[2] McMahan, H. Brendan, et al. "A general approach to adding differential privacy to iterative training procedures." arXiv preprint arXiv:1812.06210 (2018).

[3] Ponomareva, Natalia, et al. "How to dp-fy ml: A practical guide to machine learning with differential privacy." Journal of Artificial Intelligence Research 77 (2023): 1113-1201.