Bridging the Gap: Differentially Private Equivariant Deep Learning for Medical Image Analysis

Florian A. Hölzl, Daniel Rueckert, Georgios Kaissis
{florian.hoelzl, daniel.rueckert, g.kaissis}@tum.de
Institute for Artificial Intelligence in Medicine
Technical University of Munich

Abstract

Machine learning with formal privacy-preserving techniques like Differential Privacy (DP) allows one to derive valuable insights from sensitive medical imaging data while promising to protect patient privacy, but it usually comes at a sharp privacy-utility trade-off. In this work, we propose to use steerable equivariant convolutional networks for medical image analysis with DP. Their improved feature quality and parameter efficiency yield remarkable accuracy gains, narrowing the privacy-utility gap.

Keywords: Differential Privacy, Equivariant Convolutions, Steerable Kernels

1. Introduction

Deep learning-based medical imaging analysis has so-far been unable to fully leverage the advances in other fields of computer vision because many advanced applications require large datasets to obtain sufficient accuracy for diagnostic use. In medicine, procuring such datasets is problematic, as patient privacy mandates minimising the amount of data collected and/or shared. Privacy-enhancing technologies allow one to derive insights from sensitive datasets while protecting the individuals, and represent the best chance to date to incentivize data sharing in an ethical and responsible manner. Deep learning with Differential Privacy (DP) (Dwork and Roth, 2014; Abadi et al., 2016), a gold-standard technique for privacy preservation, enables analysts to train predictive models which can be shared for diagnostic purposes while offering formal guarantees about how much information can be extracted from their representations. However, DP leads to sharp utility trade-offs, as gradients are norm-bounded and noised, while noise scales proportionally to the number of model parameters, making training of large models disproportionately difficult. Thus, advanced deep learning techniques, able to learn robust and generalizable features under DP must be developed (Tramèr and Boneh, 2020). Equivariant CNN (ECNN) architectures (to Euclidean transformations) (Cohen and Welling, 2016, 2017) exhibit greatly increased data efficiency, improved generalization and high parameter efficiency, especially in domains with high degrees of intra-image symmetry such as medical imaging. So far, no works have investigated the training of ECNNs under DP, even though their characteristics render them highly attractive for this use-case. In this paper, we initiate an empirical investigation into DP ECNNs using steerable kernels. Our results demonstrate that ECNNs have superior performance with a narrowed privacy-utility gap, even at smaller model sizes. We show that these beneficial attributes are a result of lower gradient sparsity, improved feature extraction characteristics and better model calibration.

© Florian A. Hölzl, Daniel Rueckert and Georgios Kaissis.

Dataset	Data Modality	Group	No Augmentation		Augmentation	
			Non-DP	DP	Non-DP	DP
MedMNIST	Blood Cell Microscope	$\{e\}$	94.05%	89.19%	96.52%	90.61%
		C_1	93.52%	89.98%	95.57%	90.47%
		C_4	95.35%	91.30%	96.67%	91.39%
		C_8	95.32%	86.56%	96.30%	92.39%
		C_{16}	95.74%	91.48%	96.16%	91.67%
		$O(2)$	96.13%	92.51%	96.71%	93.73%
	Dermatoscope	$\{e\}$	76.79%	71.78%	78.48%	72.41%
		C_1	75.77%	72.84%	76.48%	72.27%
		C_4	75.18%	73.25%	77.84%	74.17%
		C_8	76.48%	72.79%	77.58%	66.88%
		C_{16}	76.86%	72.53%	77.57%	70.64%
		$O(2)$	77.24%	73.29%	77.98%	72.45%
	Colon Pathology	$\{e\}$	83.53%	80.11%	85.75%	79.41%
		C_1	82.35%	83.57%	82.22%	83.68%
		C_4	83.97%	86.89%	83.43%	83.52%
		C_8	85.11%	88.13%	85.23%	88.75%
		C_{16}	85.88%	89.14%	85.38%	88.94%
		$O(2)$	82.12%	77.91%	80.87%	80.64%

Table 1: Validation accuracies of the ResNet-9 for a model layout of 8-16-32 channels.

2. Results

For all experiments, we utilized the ResNet-9 architecture and privacy parameters ($\varepsilon = 7.42, \delta = 10^{-5}$) of Klause et al. (2022), adapted for equivariant training, and the MedMNIST dataset (Yang et al., 2021). Steerable ECNNs were implemented with regular representations and group pooling for the cyclic groups C_N, with N, the number of discrete rotations, restricted to $N/2$ after the first residual layer (Weiler and Cesa, 2019), and with induced irreducible representations of frequency 1, induced gated non-linearities and induced norm-pooling for the $O(2)$ group. All results are summarized in Table 1.

Superior DP classification performance: Steerable ECNNs generally outperformed their non-equivariant counterparts when training with DP, most notably on the *colon pathology* dataset, where an accuracy increase of $\approx 9\%$ was observed. The $O(2)$ group performed best on the *blood cell* and *dermatology* datasets, indicating that the benefit of higher equivariance outweighed the additional noise required due to the larger model size ($240k$ vs. $90k$ parameters for C_{16}). Out of the cyclic groups, the C_4 group performed especially well, indicating a good trade-off between equivariance and model size. Augmentations (performed as in Hoffer et al. (2020)) –on average– increased accuracy by 0.56% for DP ECNNs but decreased performance by 0.15% for non-equivariant models.

Reduced private/non-private accuracy gap: The DP validation accuracy was –on average– 5.5% lower for conventional CNNs compared to non-private training. ECNNs diminished this gap, with a reduction of only 2.8%, approaching the non-DP baseline set by Yang et al. (2021) on MedMNIST using much larger architectures like ResNet-50.

Improved accuracy despite smaller model size: DP ECNNs not only outperformed larger conventional CNNs, but even larger ECNNs. For instance, the $35k$ parameter C_4 DP ECNN outperformed both the much larger $2,5M$ parameter conventional CNN but also the $2,3M$ parameter DP ECNN (Figure 1 left).

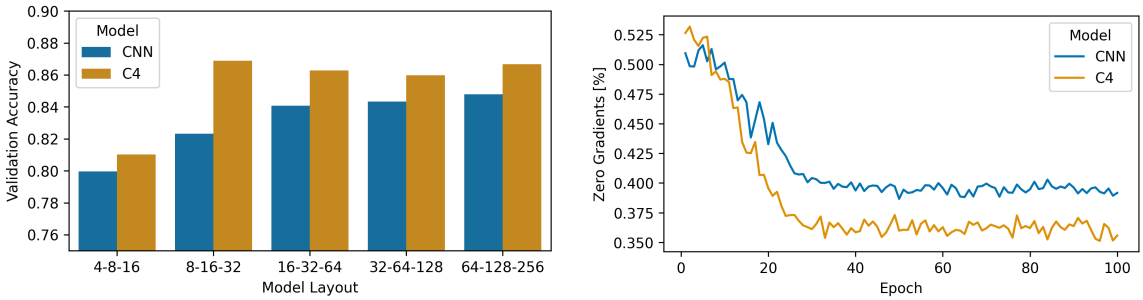

Figure 1: Results on the *colon pathology* dataset for different DP conventional and C_4 equivariant model widths (left). $\ell^0_{10^{-5}}$ gradient sparsity during training (right).

Result Interpretation: DP ECNNs had substantially lower gradient sparsity as measured by the ℓ^0_ϵ-norm, leading to the inclusion of information about negative classes as discussed in Berrada et al. (2018), exemplarily shown for the $35k$ parameter C_4 DP ECNN in Figure 1 right. DP ECNNs also exhibited better model calibration (e.g. 41.17% lower Brier score in the aforementioned model). Moreover, a more robust feature extraction by the steerable kernels was observed, indicated by a higher magnitude filter impulse response (FIR) and an improved co-location of the FIR with the input features used for prediction (Figure 2). These findings corroborate the higher sample and feature efficiency of ECNNs even in the DP setting, explaining their superior performance.

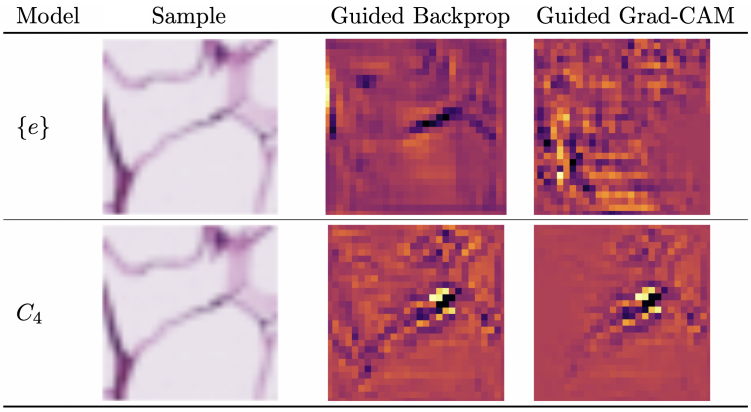

Figure 2: Guided Backprop/Grad-CAM of the final model layer. *Colon pathology* dataset.

3. Conclusion

Our preliminary findings indicate that DP ECNNs can help close the privacy-utility gap in medical imaging tasks through the beneficial characteristics of equivariant convolutions and the resulting feature, sample and parameter efficiency. In future work, we intend to investigate a larger variety of datasets and models in medical imaging and beyond.

References

Martín Abadi, Andy Chu, Ian J. Goodfellow, H. B. McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep Learning with Differential Privacy. *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 2016.

Leonard Berrada, Andrew Zisserman, and M Pawan Kumar. Smooth Loss Functions for Deep Top-k Classification. In *International Conference on Learning Representations*, 2018.

Taco Cohen and Max Welling. Group Equivariant Convolutional Networks. In *International Conference of Machine Learning*, 2016.

Taco Cohen and Max Welling. Steerable CNNs. In *International Conference on Learning Representations*, 2017.

Cynthia Dwork and Aaron Roth. The Algorithmic Foundations of Differential Privacy. *Found. Trends Theor. Comput. Sci.*, 9:211–407, 2014.

Elad Hoffer, Tal Ben-Nun, Itay Hubara, Niv Giladi, Torsten Hoefler, and Daniel Soudry. Augment Your Batch: Improving Generalization Through Instance Repetition. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8126–8135, 2020.

Helena Klause, Alexander Ziller, Daniel Rueckert, Kerstin Hammernik, and Georgios Kaissis. Differentially private training of residual networks with scale normalisation. In *ICML Theory and Practice of Differential Privacy Workshop*, 2022.

Florian Tramèr and Dan Boneh. Differentially Private Learning Needs Better Features (or Much More Data). In *International Conference on Learning Representations*, 2020.

Maurice Weiler and Gabriele Cesa. General E(2)-Equivariant Steerable CNNs. *Advances in Neural Information Processing Systems*, 32, 2019.

Jiancheng Yang, Rui Shi, D. Wei, Zequan Liu, Lin Zhao, Bilian Ke, Hanspeter Pfister, and Bingbing Ni. MedMNIST v2: A Large-Scale Lightweight Benchmark for 2D and 3D Biomedical Image Classification. *ArXiv*, abs/2110.14795, 2021.