Equivariance versus Augmentation for Spherical Images

Jan E. Gerken[1,2,3] Oscar Carlsson[1] Hampus Linander[4] Fredrik Ohlsson[5] Christoffer Petersson[6] Daniel Persson[1]

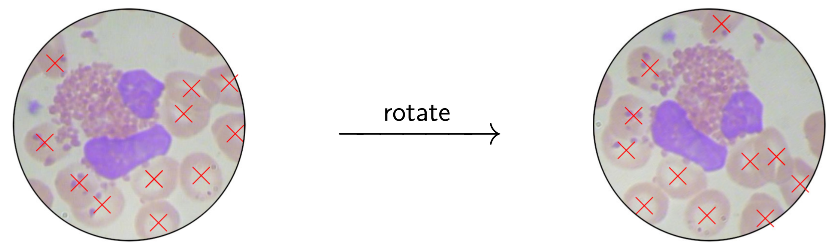
1: Department of Mathematical Sciences, Chalmers University of Technology, 2: Machine Learning Group at Berlin Institute of Technology, 3: Berlin Institute for the Foundations of Learning and Data (BIFOLD), 4: Department of Physics, University of Gothenburg, 5: Department of Mathematics and Mathematical Statistics, Umeå University, 6: Zenseact, Gothenburg

Abstract

We analyze the role of rotational equivariance in CNNs applied to spherical images. We compare the performance of the group equivariant networks known as S2CNNs and standard non-equivariant CNNs trained with an increasing amount of data augmentation. Our models are trained and evaluated on single or multiple items from the MNIST or FashionMNIST dataset projected onto the sphere. For the *invariant* task of image classification, we find that by considerably increasing the amount of data augmentation and the size of the networks, it is possible for the standard CNNs to reach at least the same performance as the equivariant network. In contrast, for the *equivariant* task of semantic segmentation, the non-equivariant networks are consistently outperformed by the equivariant networks with significantly fewer parameters. We also analyze and compare the inference latency and training times of the different networks, enabling detailed tradeoff considerations between equivariant architectures and data augmentation for practical problems. Our code is publicly available.

Symmetries in ML

- Many machine learning problems have an inherent symmetry, e.g. in medical imaging



Equivariance

- Equivariant neural networks build the symmetries of the problem into the network architecture
- A network is equivariant with respect to a symmetry group G iff

$$\mathcal{N}(T_g x) = T'_g \mathcal{N}(x) \qquad \forall g \in G$$

- Widely used in applications such as quantum chemistry, medical imaging, ...
- Equivariant networks require a specialized architecture
- Equivariance is guaranteed to hold exactly

Data augmentation

- Enlarge training set by adding transformed data points
- Increases training time
- No need for specialized architecture, thus easy to implement
- No guarantee for exact equivariance

Dataset

- MNIST digits projected onto the sphere with classification labels and segmentation masks

Main problem

Analyze trade-off between data augmentation and equivariance for invariant and equivariant tasks.

Key takeaway

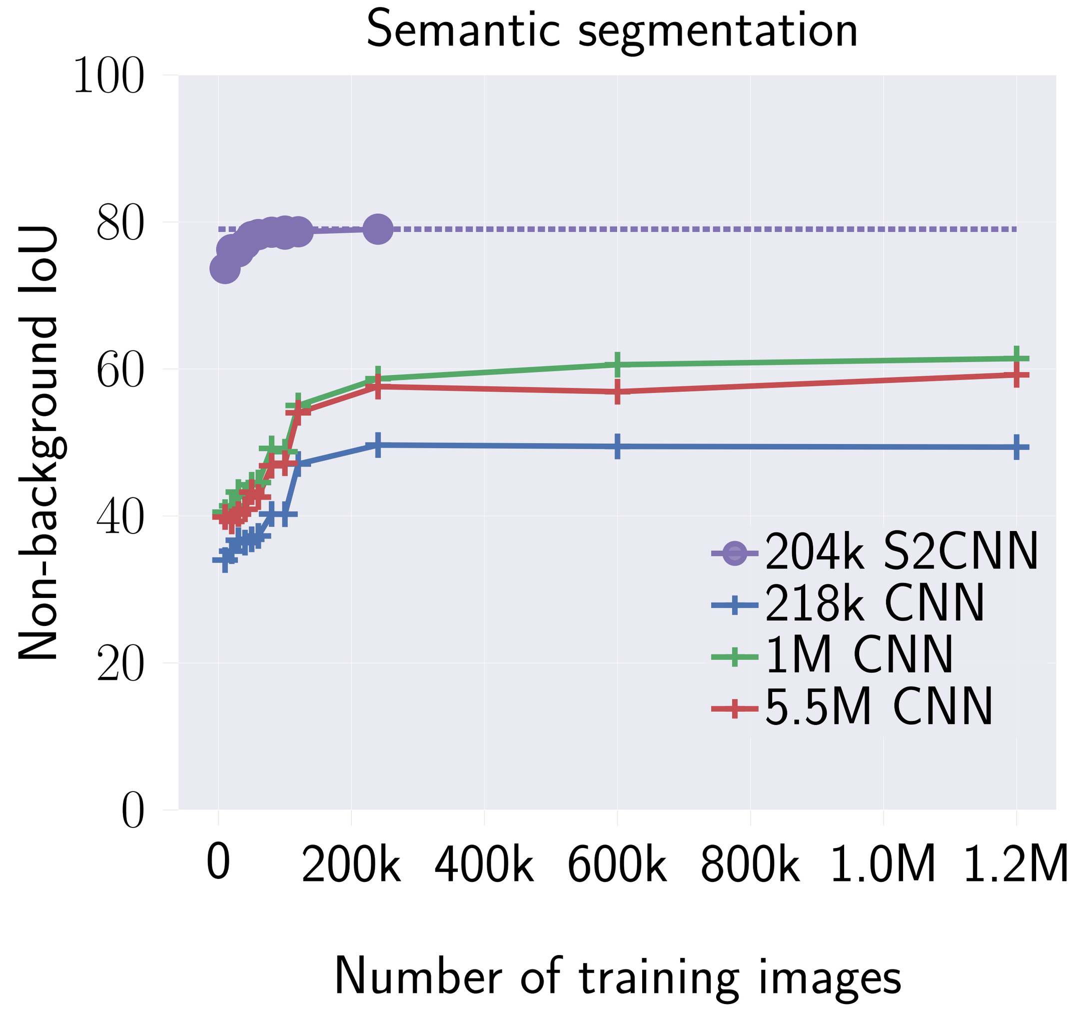
For equivariant tasks, the performance of non-equivariant networks trained with data augmentation saturates well below the performance of much smaller equivariant models.

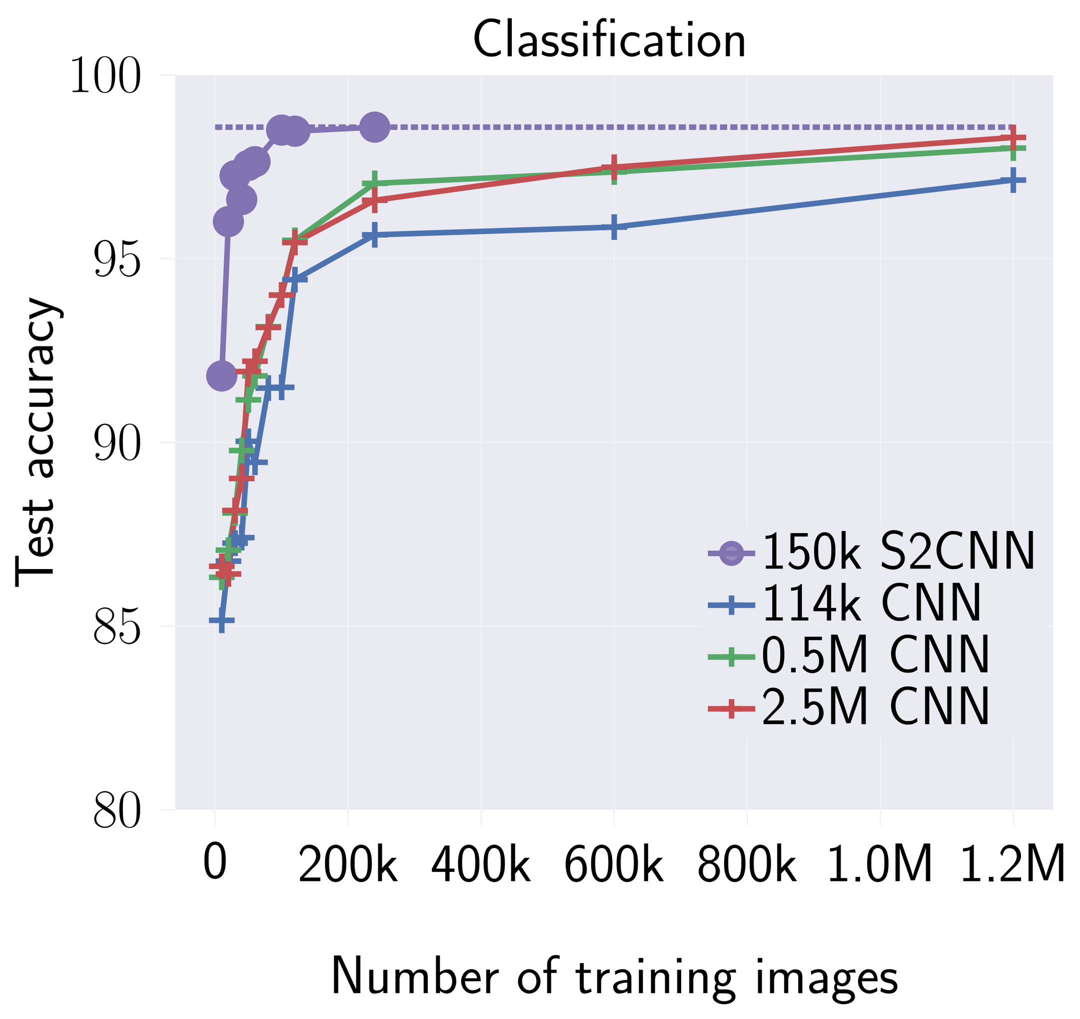
For invariant tasks, the performance of larger non-equivariant networks trained with data augmentation can reach the performance of equivariant networks, although the non-equivariant networks take longer to train.

Results

Performance

- Numbers in model names refer to the number of trainable parameters





Throughput and training times

Due to the specialized architecture the equivariant model has much higher latency at similar parameter counts than the non-equivariant models.

Model	Latency (ms)	Throughput (N/s)
204k S2CNN	111 ± 0.6	9.0 ± 0.04
200k CNN	5.93 ± 0.24	169 ± 5.8

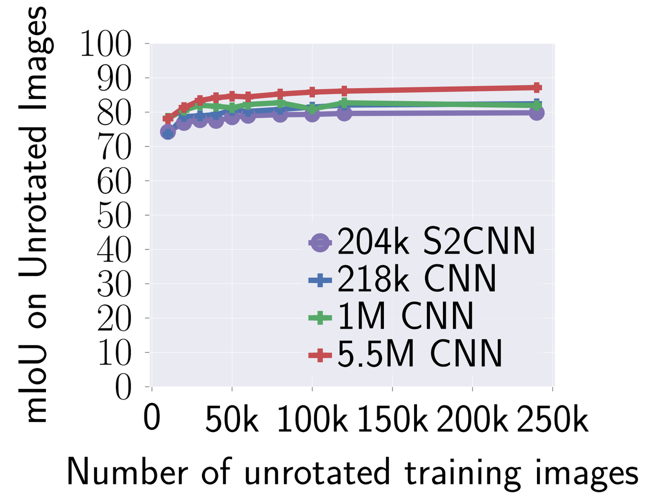
At matched accuracy the total training time for the non-equivariant model trained with data augmentation is much higher than the training time for the equivariant model trained without data augmentation.

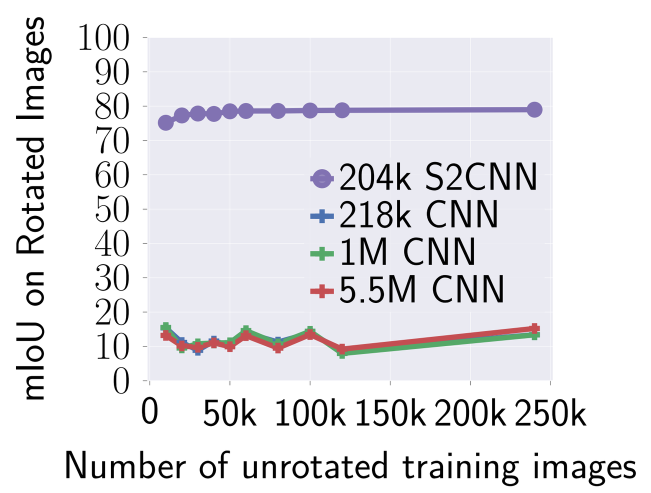
Model	Accuracy	Training time
150k S2CNN	97.64%	15h
5M CNN	97.49%	26h

Further results

Rotated vs non-rotated test images

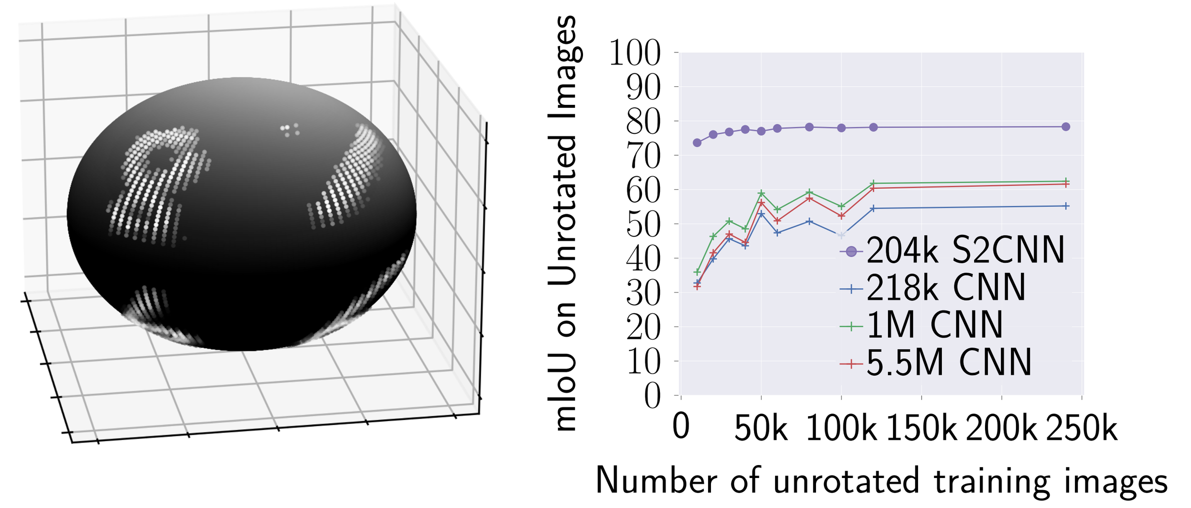
When training is performed only on unrotated images, the non-equivariant models outperform the equivariant models on unrotated test data. On rotated test data, the non-equivariant performance deteriorates whereas the equivariant performance is unaffected.





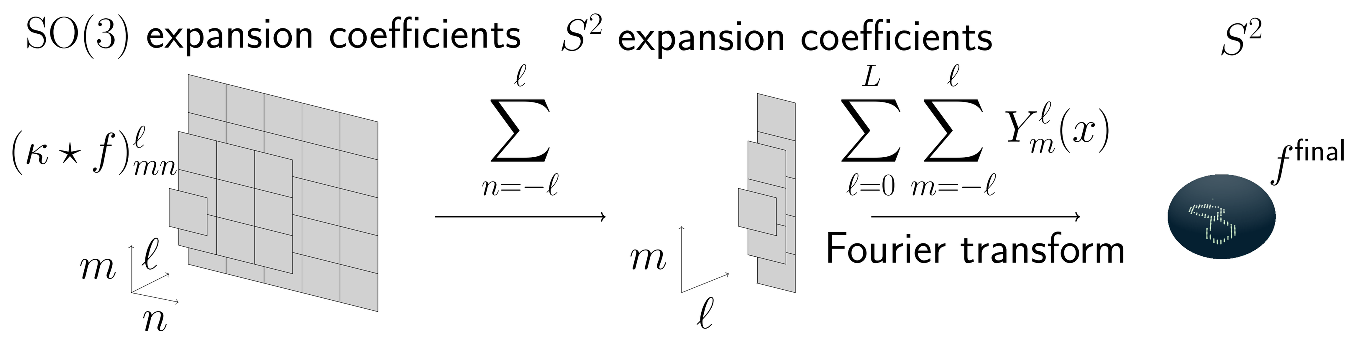
Multiple digits

Similar results hold for semantic segmentation with four digits projected onto the sphere.



New equivariant S2CNN layer for semantic segmentation

We added a layer to the S2CNN architecture (Cohen et al., 2018) reducing feature maps on $\mathrm{SO}(3)$ to feature maps on S^2 for semantic segmentation.



Poster presented at ICML-2022 in Baltimore.

Cohen, T. S., Geiger, M., Köhler, J., and Welling, M. Spherical CNNs. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net, 2018.