Group Equivariant Convolutional Networks

January 5, 2023

Donato Francesco Pio Stanco 2027523

Abstract

In this paper will be introduced **Group equivariant Convolutional Neural Networks** (G-CNNs) (Cohen & Welling, 2016), that are a natural generalization of the traditional **convolutional neural networks** (CNNs) and those networks can be generalized to exploit larger groups of symmetries, including rotations and reflections.

1. Introduction

In a Convolutional Neural Network (CNN) we know that the Convolutional Layers are *translation equivariant* which means that applying the convolution on a translated input is the same thing to do the translation of the convolution of the input. In few words this means that the *symmetry* or *translation* is preserved in each layer, which makes it possible to exploit it not just in the first, but also in higher layers of the network. However, CNNs failed to make good use of these symmetries.

Group Equivariant Convolutional Neural Networks (G-CNNS) (Cohen & Welling, 2016) was proposed in 2016 as a generalization of CNNs, using **G-convolutions** to enjoy a substantially higher degree of weight sharing, which can be implemented with negligible computational overhead for discrete groups generated by translations, reflections and rotations. These networks are based on groups, a **group G** is a set that with a binary operation given two elements a,b $\in$ G, it produces another element *ab* and also it satisfies identitivity, associativity, closure and inverse.

We have that equivariance is a kind of important symmetry and in those architectures we have some layer ϕ that is defined as an action , which maps one representation to another and is structure preserving.

For G-spaces ϕ has to be equivariant if:

$$\phi(T_g x) = T'_g \phi(x)$$

Email: Donato Francesco Pio Stanco <stanco.2027523@studenti.uniroma1.it>.

Deep Learning and Applied AI 2022, Sapienza University of Rome, 2nd semester a.y. 2021/2022.

where T_g is transformation of g on x, then input the result to ϕ, while T'_g is a transformation of g on $\phi(x)$. Equivariance means that the above two transformations are equal.

The G-groups used are: the *symmetry group* that is the set of transformations that leaves the object equivariant (an exampe is the set of 2D integer translations $\mathbb{Z}^2$), the group *p4* that consists of translations and rotations by 90 degrees about any center of rotation in a square grid and the group *p4m* consists of all compositions of translations, mirror reflections, and rotations by 90 degrees about any center of rotation in the grid. Another important aspects are the functions of the group, described them is useful to understand how this kind of networks act. In few words the map from images to stacks of feature maps with channels in a CNN can be modeled as $f : \mathbb{Z}^2 \rightarrow \mathbb{R}^k$, the stack of feature maps returns a *K-dimensional* vector $f(x, y)$. A transformation acting on a set of feature maps is shown as below:

$$[L_g f](x) = [f \circ g^{-1}](x) = f(g^{-1}x)$$

This says that the operator L_g is an instantiation of the transformation operator T_g and we have:

$$L_g L_h = L_{gh}$$

So g represents a pure translation $t = (u,v) \in Z^2$, and $g^{-1}x$ simply means x-t. An example of feature map is provided in the following image:

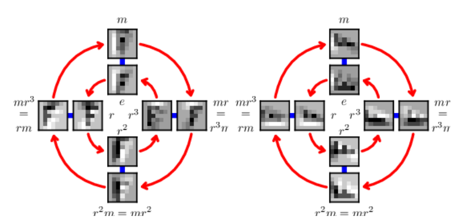

Figure 1. A p4m feature map and its rotation by r

Let's briefly describe how are done the G-CNNs. We have three layers which are **G- convolution**, **G-pooling** and **nonlinearity** and each one commutes with **G-transformation** of the domain of the image (Cohen & Welling, 2016).

2. Related Work

These new kind of convolutional neural networks have reached a big popularity in the years after the publication. They inspires some nice works, for example two good generalizations proposed by the same author were the Steerable CNNs (Cohen & Welling, 2017) and the Spherical CNNs (Cohen et al., 2018). Also this kind of networks have several applications in Medical Image Analysis with DenseNet (Li et al., 2020), breast tumor classification (Graham et al., 2019), 3D-GCNNs on Nodule classification (van Ginneken et al., 2010). These networks also were tested in many areas such as Trajectory Prediction with Equivariant Continous Convolution (Walters et al., 2021).

3. Experiments

In this section the results produced are used to shown the correctness of the ones presented by (Cohen & Welling, 2016). To confirm the experiments I decided to follow the pipeline proposed by the authors but also modified some part in order to make more long training and testing to achieve nice results.

3.1. Rotated MNIST

The model used for test this architecture on this dataset is like the pytorch example on the repository MNIST where the architecture of the network is the follow:

- the Convolutional and Droput Layers are replaced by four G-Convolutional layers
- the G-Convolutional layers use the **G-group p4**

For the **training** of the model the number of epochs is equal to **60** and the entire process takes around *37 minutes* of execution. For the results produced see section 4.

3.2. CIFAR 10

The model used for the test this architecture on this dataset is like the pytorch example on the repository CIFAR10 where the architecture of the network is the follow:

- all planar convolutions are replaced with **p4m group** convolution
- the number of filters in each convolutional layer was reduced by sqrt(8) to keep similar number of parameters
- the architecture used is the **ResNet18**

The adavantage of using **ResNets** is that these netowrks learn residual functions with reference to the layer inputs, instead of learning unreferenced functions. This approach makes it possible to train the network on thousands of layers without affecting performance (He et al., 2015).

For the **training** of the model the number of epochs is equal to **140** and the entire process takes *10 hours* of execution. For the results produced see section 4.

3.3. Code

The code of the project is in the following GitHub repository DeepLearningProject-GCNN.

4. Summary of the experiments

In this section are reported the results of the experiments for both the datasets.

Table 1. Performance comparison.

Results	Training Set	Test Set	Test Error
RotatedMNIST	99.0%	**99.0%**	**1.32%**
CIFAR10	100%	**94.23%**	**5.68%**

In the case of the *RotatedMNIST* after 60 epochs (specified in 3.1) the test error is equal to 1.32%, so the accuracy can be confirmed.

In the case of the *CIFAR10* after 140 epochs (specified in 3.2) the test error is equal to 5.68% that is a good result since the ResNet18 architecture is used and also with *planar convolutional layer* the test error is equal to **6.8%** so it is a very nice result.

Analyzing the results obtained from both tests, it can be seen that the performances of these particular convolutional neural networks are respected withe the ones presented in the literature (Cohen & Welling, 2016).

5. Future Works

These networks were widely studied and generalized in these years, we have an implementation for 3D data, medical analysis and some good generalization. Most of all the tasks that are in the literature are about image classification. As mentioned in (Li et al., 2020) this particular networks are used in medical analysis and a possible improvement of this in my opinion could be the combination of G-CNNs and classification or segmentation models according to the specific characteristics of lesions, tissues, structures due to the more complex symmetries in 3D images. Another task could be to use the GAN proposed in (Dey et al., 2021) to generate some new images for example a sketch image to a realistic one.

References

Cohen, T. S. and Welling, M. Group equivariant convolutional networks. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48*, ICML'16, pp. 2990–2999. JMLR.org, 2016.

Cohen, T. S. and Welling, M. Steerable CNNs. In *International Conference on Learning Representations*, 2017. URL https://openreview.net/forum?id=rJQKYt5ll.

Cohen, T. S., Geiger, M., Köhler, J., and Welling, M. Spherical CNNs. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=Hkbd5xZRb.

Dey, N., Chen, A., and Ghafurian, S. Group equivariant generative adversarial networks. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=rgFNuJHHXv.

Graham, S., Epstein, D., and Rajpoot, N. Rota-net: Rotation equivariant network for simultaneous gland and lumen segmentation in colon histology images. In Reyes-Aldasoro, C. C., Janowczyk, A., Veta, M., Bankhead, P., and Sirinukunwattana, K. (eds.), *Digital Pathology*, pp. 109–116, Cham, 2019. Springer International Publishing. ISBN 978-3-030-23937-4.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015. URL http://arxiv.org/abs/1512.03385.

Li, Y., Cao, G., and Cao, W. A dynamic group equivariant convolutional networks for medical image analysis. In *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 1056–1062, 2020. doi: 10.1109/BIBM49941.2020.9313601.

van Ginneken, B., Armato, S. G., de Hoop, B., van Amelsvoort-van de Vorst, S., Duindam, T., Niemeijer, M., Murphy, K., Schilham, A., Retico, A., Fantacci, M. E., Camarlinghi, N., Bagagli, F., Gori, I., Hara, T., Fujita, H., Gargano, G., Bellotti, R., Tangaro, S., Bolaños, L., Carlo, F. D., Cerello, P., Cristian Cheran, S., Lopez Torres, E., and Prokop, M. Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: The anode09 study. *Medical Image Analysis*, 14(6):707–722, 2010. ISSN 1361-8415. doi: https://doi.org/10.1016/j.media.2010.05.005. URL https://www.sciencedirect.com/science/article/pii/S1361841510000587.

Walters, R., Li, J., and Yu, R. Trajectory prediction using equivariant continuous convolution. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=J8_GttYLFgr.