Group Equivariant Neural Networks : A gentle Introduction

September 17, 2021

Senad Beadini

Abstract

Group equivariant convolutional neural networks (G-CNNs) have been introduced by Cohen & Welling (2016). These new models can be considered as an evolution of convolutional nerual networks (CNNs). The key operation which G-CNNs use is G-Convolution; this is a new layer that make neural networks equivariant to new symmetries. In this paper, I'll provide a theoretical overview of G-CNNs, G-Convolution and finally I'll show the performance of G-CNNs vs CNNs on Fashion-MNIST and CIFAR-10 datasets.

1. Introduction

Convolution neural networks (CNNs) have become the gold standard for addressing sevaral problems in Artificial Intelligence (LeCun et al., 2015). CNNs have allowed to overcome the limitations of ancient models, in particular they bring the following important ideas into machine learning (Goodfellow et al., 2016): sparse interaction, parameters sharing and equivariant rappresentation.

Traditional neural networks use matrix multiplication as layer unit. This setting determines a huge number of parameters with just a few number of layers. On the contrary, convolutional networks typically have sparse wheights. This feature, in addition to significantly reducing parameters, makes CNNs capable of capturing common patterns in the data.

The convolution layer, embedded in CNNs, has a nice property called translation equivariance. In general we say that a function f is equivariant to a function g if $f(g(x)) = g(f(x))$. In the scenario of convolution, the function g can be any translation function of the input. A drawback is that convolution is not equivariant to other transformations/symmetries like rotations and reflections; for instance, rotating the image and then convolving is not the same as first convolving and then rotating the result. Thus, CNNs are not able to exploit directly other symmetries as well as translation. Cohen & Welling (2016) proposed a generalization of convolution *G-convolution* to overcome this issue.

Email: Senad Beadini <beadini.1754617@studenti.uniroma1.it>.

Deep Learning and Applied AI 2021, Sapienza University of Rome, 2nd semester a.y. 2020/2021.

2. Methods

In this section I have briefly discussed the math tools useful for defining *G-convolution*.

I recall that a group is a set with an binary function satisfying the properties of closure, associativity, identity, and invertibility. We say in general that, a symmetry of an object is a transformation of that object that leaves it unchanged. It can be shown that symmetries respect the group definition. The groups which enable the generalization of standarad convolution operation are two $p4$ and $p4m$ (Cohen & Welling, 2016).

2.1. p4 and p4m groups

The group $p4$ consists of all compositions of translations and rotations by 90 degrees. Insted, the group $p4m$ is composed by all compositions of 90-degree rotations, reflections, and translations. These groups can be parameterized as follows (Cohen & Welling, 2016) :

$$g(r,u,v) = \begin{bmatrix} \cos\frac{r\pi}{2} & -\sin\frac{r\pi}{2} & u \\ \sin\frac{r\pi}{2} & \cos\frac{r\pi}{2} & v \\ 0 & 0 & 1 \end{bmatrix} \tag{1}$$

$$g'(m,r,u,v) = \begin{bmatrix} (-1)^m\cos\frac{r\pi}{2} & (-1)^m\sin\frac{r\pi}{2} & u \\ \sin\frac{r\pi}{2} & \cos\frac{r\pi}{2} & v \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

where $g(r,u,v)$ parameterizes $p4$, $g'(m,r,u,v)$ parameterizes $p4m$, $0 \leq r < 4$ is the number of 90-degree rotations, $m \in 0,1$ the number of reflections, and $(u,v) \in \mathbb{Z}^2$. Thus the group operation is matrix multiplication for both groups, for instance, the group $p4$ acts on points in Z^2 (pixel coordinates) by multiplying the matrix $g(r,u,v)$ by the (homogenous) coordinate vector $x(u',v')$ of the point $(u',v') \in \mathbb{Z}^2$. Less formally, the matrix $g(r,u,v)$ rotates and translates a point (expressed as homogeneous coordinate vector) in pixel space via left-multiplication.

2.2. Group functions

Cohen & Welling (2016) introduce the following notation which express the transformations of the feature maps. In details, they express the transformation g acting on a set of feature maps as:

$$[L_g f](x) = [f * g^{-1}](x) = f(g^{-1}x) \tag{3}$$

This means that to obtain the value of the g-transformed feature map $L_g f$ at the point x, we need to see in the original feature map f at the point $g^{-1}x$.
If g represents a translation $t = (u, v) \in \mathbb{Z}^2$ then $g^{-1}x$ would be $x - t$.

3. G-convolution

We should say that formally the following are $correlation$s and not convolutions but the facts do not change. We can define the *G-convolution* between a pixel-image tensor with K-channel $f : \mathbb{Z}^2 \to \mathbb{R}^K$ and a filter $\phi : \mathbb{Z}^2 \to \mathbb{R}^k$ as :

$$[f \star \phi](g) = \sum_{y \in \mathbb{Z}^2} \sum_{k=1}^{K} f_k(y)\phi_k(g^{-1}y) \tag{4}$$

As in the previous section, let's consider G to be the translation group on $\mathbb{Z}^2$, we have $g^{-1}y = y - g$, therefore we obtain the standard definition of convolution. An important thing to say is that both the input image and the filter ϕ are functions of $\mathbb{Z}^2$,but the feature map $[f \star \phi]$ is a function on a group G. Hence, for all layers after the first, the filters ϕ must be functions on G, and the correlation operation needs to be write as follows :

$$[f \star \phi](g) = \sum_{h \in G} \sum_{k=1}^{K} f_k(h)\phi_k(g^{-1}h) \tag{5}$$

At the end, we can pool the feature map for each filter over the set of transformations, corresponding to average or max pooling over the specific group.
Visualizing feature maps on groups it's not so intuitive. In 1 there's an example of the 90 degree rotation *r* to a function on *p4*.

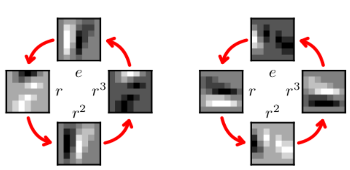

Figure 1. A p4 feature map and its rotation by r. In the right the result of the operation, a 90-degree rotation. (Cohen & Welling, 2016)

4. Experiment Result

Fashion-MNIST is composed by $60k$ Zalando's article images. Each example is a $28 * 28$ grayscale image associated with a label from 10 classes. CIFAR-10 contains $60k$ of $32 * 32$ color images in 10 different classes. The classes rappresent some common objects. The test set for both datasets contain $10k$ images and thus the training set $50k$.
Testing the effectiveness of G-CNNs, I have developed two models: a G-CNN and a standard CNN. Preciseley both models have 4 convolution and 2 fully conntected layers. Obviously, in G-CNN each convolutional layer is replaced with $p4m$ group convolution. The size of each layer was develop to have a comparable as much as possible number of parameters (roughly 2 million parameters). Since either Fashion-MNIST and CIFAR10 are dataset that do not contain samples rotated, I augmented the datasets with random rotations, this allows G-CNNs to exploit group equivariance.
For both models, I used adam with learning rate 0.001, the number of epochs is 30. In 1, we can see the result. According the tests G-CNNs are able to compete with CNNs and even though with some limitations in the tests (limited by a question of computational power) the accuracy is similary, slightly higher for G-CNNs. The code is available in the following link : Repository.

Model	Fashion-MNIST	CIFAR10
G-CNN	87.91%	64.18%
CNN	86.67%	57.37 %

Table 1. Accuracy on test set

5. Conclusion and Future work

In summary, I initially presented the problem of generalizing convolution to new symmetries. In addition, I briefly discussed the main tools that allowed the development of G-convolution. Finally, I tested the G-CNNs (using limited computing power, Colab only) showing that they perform at least like CNNs and even slightly better. According to the test carried out here, the new layer introduced in (Cohen & Welling, 2016) could be effectively used to replace the standard convolution layers in CNNs with the advantages of equivariance for $p4$ and $p4m$ group.
Group theory is a well know math field with a strong formalization and therefore the results proposed can be further generalized to new groups. For instance, G-CNNs for 3d groups is an interesting research. Another further generalizion was made recently by Romero et al. (2020) who introduce attention mechanism into G-CNNs. Each year there are new research results due to the fact that equivariance to symmetries is a fundamental requirement for *Intelligence*.

References

Cohen, T. and Welling, M. Group equivariant convolutional networks. In *International conference on machine learning*, pp. 2990–2999. PMLR, 2016.

Goodfellow, I., Bengio, Y., and Courville, A. *Deep learning*. MIT press, 2016.

LeCun, Y., Bengio, Y., and Hinton, G. Deep learning. *nature*, 521(7553):436–444, 2015.

Romero, D., Bekkers, E., Tomczak, J., and Hoogendoorn, M. Attentive group equivariant convolutional networks. In *International Conference on Machine Learning*, pp. 8188–8199. PMLR, 2020.