ISSN 1512-1232

UDC-004.8

Constructing Convolutional Neural Networks with 90 Degree Rotational Equivariance and Invariance

Teimuraz Saghinadze

Muskhelishvili Institute of Computational Mathematics of the Georgian Technical University,
4 Grigol Peradze str., Tbilisi, 0159, Georgia
saghinadze.teimuraz@gtu.ge

Abstract
It's a well-known fact that the convolutional layer has the property of translational equivariance. However, it's non-obvious how to expand the symmetry group associated with the said layer. Employing key definitions adopted in deep geometric learning, we construct the set of filters that induce 90-degree rotational equivariance without modifying the convolutional operator. This work is primarily intended as a theoretical exercise, beginning with a predefined symmetric group in mind and producing a convolutional layer with the desired equivariance.

Keywords: *convolutional neural networks, symmetry, equivariance, invariance.*

1. Introduction

The widespread adoption of CNNs has led to the development of various architectures applicable to numerous computer vision tasks and beyond [1,2]. Convolutional layers possess the property of translational equivariance. Most CNN networks are designed in a manner where the input of shape (C, H, W) (channel, height, width) is transformed into an intermediate output of shape $(C_i, 1, 1)$. If all the functions used up to this point were translationally equivariant, the entire network would be translationally invariant.

Building upon this observation, one can introduce additional symmetries to CNNs. For instance, it is reasonable to assume that a rotated, horizontally, or vertically flipped image of a person still conveys the same information. The conventional approach to addressing this is via data augmentation [3]. Though it is challenging to measure this precisely in practice, this strategy basically encourages the network to acquire weights that are more or less insensitive to such augmentations. [4].

Alternatively, one can introduce specific inductive biases into the network [5,6]. In the following sections, we propose a simplified approach that does not alter the convolution operator but imposes constraints on the filters. Although the results for the horizontal and vertical symmetries of an image are analogous, we will concentrate on the procedure for the 90-degree rotation function because it is a bit more involved.

2. Key definitions

Since the notation is quite varied among authors, we introduce some basic definitions here.

We take $X \in \mathbb{R}^{(H+1)\times(W+1)}$ and $K \in \mathbb{R}^{(n+1)\times(n+1)}$, where X can be understood as a grid of gray-scale pixels and K as filter (kernel) of convolution. For simplicity of notation, we'll range indices i, j of matrix X over $i \in \{0,1,\dots,H\}$ and $j \in \{0,1,\dots,W\}$, similarly u, v range over $\{0,1,\dots,n\}$, $n \leq H$ and $n \leq W$.

ISSN 1512-1232

Definition 2.1:

$$\mathbb{M}_n \coloneqq \{x \in \mathbb{R}^{H\times W}: n \leq H \text{ and } n \leq W\}$$

Definition 2.2:
A convolution of matrix X with filter K can be expressed as,

$$[K \otimes X]_{i',j'} = \sum_{u=0}^{n}\sum_{v=0}^{n} K_{u,v}X_{i'+u,j'+v}$$

where i' and j' range over $\{0,1,\cdots,H-n\}$ and $\{0,1,\cdots,W-n\}$ respectively.
Some remarks:

- Definition 2.2 is a definition of convolution that is used in CNN-s and doesn't agree with the mathematical definition of convolution.
- Order of K and X with respect to $[\,.\,\otimes\,.\,]$ matter.
- To the best of our knowledge to specify domain and codomain of $[\,.\,\otimes\,.\,]$ one needs to understand it as family of maps indexed by the size of filter K namely $[\,.\,\otimes_n\,.\,]: \mathbb{R}^{n\times n} \times \mathbb{M}_n \to \mathbb{M}_1$. When index is omitted, it's implicitly inferred from dimensions of K.

Definition 2.3. 90° counterclockwise rotation of a matrix (a digital image) can be represented via a map $R: \mathbb{M}_1 \to \mathbb{M}_1$ such that following holds.

$$X \in \mathbb{R}^{H\times W} \to R(X) \in \mathbb{R}^{W\times H};$$
$$(X)_{j,i} = X_{i,W-j} \text{ where } i \in \{0,1,\ldots,H\} \text{ and } j \in \{0,1,\ldots,W\}.$$

3. Properties of 90^0 rotation function

Observation 3.1. One can simply verify following statements:

a) $\forall i \in \{0,1,\ldots H\}\left(\forall j \in \{0,1,\ldots,W\}\left(R(X)_{j,i} = X_{i,W-j}\right)\right)$ if and only if $\forall i \in \{0,1,\ldots H\}\left(\forall j \in \{0,1,\ldots,W\}\left(\left(R(X)_{W-j,i} = X_{i,j}\right)\right)\right)$.

b) R is a bijection.

c) $R^2(X)_{i,j} = X_{H-i,W-j}$.

d) $R^3(X)_{j,i} = X_{H-i,j}$.

e) $R^4 = \mathrm{id}_{\mathbb{M}_1}$.

f) $(R^k)^{-1} = R^{4-k}$ where $k \in \{1,2,3,4\}$.

g) $\{\mathrm{id}_{M_{\mathbb{1}}}, R, R^2, R^3\}$ with function composition is a group.

Proof. a)

($\Rightarrow$): substituting j with $W-j$ in the left-hand side results in desired identity.

($\Leftarrow$): given that $R(X)_{W-j,i} = X_{i,j}$ holds re-index using $\bar{j} = W-j$ and $W-j,i$ to obtain identity on the left-hand side.

Proof. b)

R **is injective.** To show this fact let's assume that it's not, then there are X and Y such that $X \neq Y$ and $R(X) = R(Y)$. If X and Y don't agree on their dimensions, $R(X)$ and $R(Y)$ will have different dimensions as well, leading to contradiction. If X and Y agree on dimensions then $X \neq Y$ entails that there's at least one pair of indices i_0, j_0 such that $X_{i_0,j_0} \neq Y_{i_0,j_0}$, but then using a) $R(X)_{W-j,i} = X_{i,j} \neq Y_{i,j} = R(Y)_{W-j,i}$ which leads to contradiction again.

ISSN 1512-1232

R **is surjective.** Given any matrix $Y \in \mathbb{M}_1$ one can construct $X \in \mathbb{M}_1$ such that $X_{i,j} = Y_{W-j,i}$ for every appropriate index, hence $R(X) = Y$. Hence R is surjective.

Proofs of c), d), e) and f) follow from repeated application of definition of R.

Proof. g)

From a) it follows that R is a member of $\mathrm{Auto}(\mathbb{M}_1)$, hence $\langle R \rangle$ a is group generated by R. From e) it at most of order 4. R^1 and R^3 clearly be identity maps since it exchanges dimension of rectangular matrix. Neither $R^{(2)}$ is an identity map since if we take a particular matrix $X \in \mathbb{R}^{(H+1)\times(W+1)}$ such that $X_{i,j} = \delta_{0,i}\delta_{0,j}$ and either H or W differs from 0, then $R^2(X) \neq X$.

QED

Observation 3.2. Following five statements are equivalent:

a) $K = R(K)$.
b) $R(K) = R^2(K)$.
c) $R^2(K) = R^3(K)$.
d) $R^3(K) = K$.
e) $K = R^{-1}(K)$.

Proof.

Since R is a function, most arguments are trivial.

- b) follows from a).
- c) from b).
- d) follows from c) and observation 1.e.
- follows from d) and observation 1.e.
- is the same statement as d), based on observation 1.f.

QED

We'll also need following two simple observations:

Observation 3.3.

$$R([K \otimes X])_{i,j} = \sum_{u=0}^{n}\sum_{v=0}^{n} K_{u,v} X_{j+u,W-n-i+v}.$$

Proof.

This follows from first applying definition of R followed definition of convolution.

$$R([K \otimes X])_{i,j} = [K \otimes X]_{j,W-n-i} = \sum_{u=0}^{n}\sum_{v=0}^{n} K_{u,v} X_{j+u,W-n-i+v}.$$

QED

Observation 3.4.

$$[K \otimes R(X)]_{i,j} = \sum_{u=0}^{n}\sum_{v=0}^{n} K_{n-v,u} X_{j+u,W-n-i+v}.$$

Proof.

$$[K \otimes R(X)]_{i,j} = \sum_{u=0}^{n}\sum_{v=0}^{n} K_{u,v} R(X)_{i+u,j+v} = \sum_{u=0}^{n}\sum_{v=0}^{n} K_{u,v} X_{j+v,W-(i+u)}.$$

ISSN 1512-1232

(interchange order of summation and rename u to v and v to u)

$$\sum_{u=0}^{n}\sum_{v=0}^{n} K_{u,v}X_{j+v,W-(i+u)} = \sum_{u=0}^{n}\sum_{v=0}^{n} K_{v,u}X_{j+u,W-(i+v)}.$$

(reversing the order of summation with respect to v)

$$\sum_{u=0}^{n}\sum_{v=0}^{n} K_{v,u}X_{j+u,W-(i+v)} = \sum_{u=0}^{n}\sum_{v=0}^{n} K_{n-v,u}X_{j+u,W-(i+(n-v))} = \sum_{u=0}^{n}\sum_{v=0}^{n} K_{n-v,u}X_{j+u,W-n-i+v}.$$

QED

Lemma 3.1.

$$\exists n_0 \leq n\big(K \in \mathbb{R}^{n_0\times n_0} \land K = R^{-1}(K)\big) \leftrightarrow \forall X \in \mathbb{M}_n(R([K \otimes X]) = [K \otimes R(X)]).$$

Proof.

($\Rightarrow$): If $\forall u, v \in \{0,1,\cdots,n\}\big(K_{u,v} = K_{n-v,u}\big)$ for some n then given an $X \in \mathbb{M}_n$

$$R([K \otimes X])_{i,j} = \sum_{u=0}^{n}\sum_{v=0}^{n} K_{u,v}X_{j+u,W-n-i+v} = \sum_{u=0}^{n}\sum_{v=0}^{n} K_{n-v,u}X_{j+u,W-n-i+v} = [K \otimes R(X)]_{i,j}.$$

($\Leftarrow$): To show the other direction we'll show the contrapositive.

$$\exists u, v \in \{0,1,\cdots,n\}\big(K_{u,v} \neq K_{n-v,u}\big) \Rightarrow \exists X \in \mathbb{M}_n(R([K \otimes X]) \neq [K \otimes R(X)]).$$

Let's say u_0 and v_0 are such indices that $K_{u_0,v_0} \neq K_{n-v_0,u_0}$.

Let $X \in \mathbb{R}^{(n+1)\times(n+1)}$ (of same size as K) such that $X_{i,j} = \delta_{u_0,i}\delta_{v_0,j}$.

Then $R([K \otimes X]) = K_{u_0,v_0})$ and $([K \otimes R(X)] = K_{n-v_0,u_0}$. Hence $R([K \otimes X] \neq [K \otimes R(X)])$.

QED

Corollary 3.1.

$$\forall X \in \mathbb{M}_n(R([K \otimes X]) = [K \otimes R(X)]) \leftrightarrow K \in \{\, k \in \mathbb{R}^{n_0\times n_0} |\, n_0 \leq n, R(k) = k\}.$$

4. Final remarks

For rotational symmetries, a filter is necessarily a square matrix. But for horizontally and vertically symmetric convolutions it can be any rectangular matrix. Following the same reasoning presented in previous sections one can derive $K = H(K)$ and $K = V(K)$ are necessary and sufficient conditions for convolution to be equivariant with respect to H and V respectively, where H reversed order columns and V reverses the order of rows.

A kernel size greater than 3x3 is needed if one wants solely rotational symmetry and no horizontal or vertical symmetries; otherwise, such kernels, by construction, will necessarily have horizontal and vertical symmetries.

While implementing such a rotationally equivariant kernel, other parts of the network should satisfy the same criterion. But one can check that the most widely used components like elementwise activation functions, batch norm [7], and residual connections [2] satisfy these conditions.

Acknowledgement. The author was supported by the European Union's Grant Project GAIN (grant agreement no.101078950)

ISSN 1512-1232

References:

[1]. Krizhevsky, A., Sutskever, E., and Hinton, G. E. 2017. ImageNet classification with deep convolutional neural networks. Commun. ACM 60, 6 (June 2017), 84–90. https://doi.org/10.1145/3065386

[2]. He K., Zhang X., Ren S., and Sun J., "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

[3]. Shorten, C., Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. *J Big Data* **6**, 60 (2019). https://doi.org/10.1186/s40537-019-0197-0

[4]. Bronstein, M. M., Bruna, J., Cohen, T., & Veličković, P. (2021). Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*.

[5]. Cohen, T. & Welling, M. (2016). Group Equivariant Convolutional Networks. Proceedings of the 33rd International Conference on Machine Learning, in Proceedings of Machine Learning Research 48:2990-2999 Available from https://proceedings.mlr.press/v48/cohenc16.html.

[6]. Romero D., Bekkers E., Tomczak E., Hoogendoorn M., “Attentive Group Equivariant Convolutional Networks”, *Proceedings of the 37th International Conference on Machine Learning*, PMLR 119:8188-8199, 2020.

[7]. Ioffe, S. and Szegedy, C. (2015) Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. ICML'15: Proceedings of the 32nd International Conference on International Conference on Machine Learning, 2015, 448-456.

Article received: 2023-12-12