Orthogonal Rotation Invariant Moments for Equivariant Convolutional Neural Networks

Jaspreet Singh, Chandan Singh
Department of Computer Science, Punjabi University, Patiala -147002, Punjab, India

Abstract

The standard convolutional neural networks (CNNs) are inherently equivariant to translation due to the convolutional operation. In order to achieve the equivariance to other affine geometric transformations such as rotation and reflection, the group-equivariant CNNs (G-CNNs) are proposed which are equivariant to the transformations defined by the group. In this work, we use Zernike moments which belongs to the class of orthogonal rotation invariant moments to achieve global equivaiance with respect to rotation, reflection and translation. The Zernike moment layer is embedded before the fully connected layers of CNNs and G-CNNs. The new proposed Zernike moment based CNN and G-CNN models achieve high recognition rates on the rotated MNIST dataset as compared to standard CNN (Z2CNN) and G-CNN (P4CNN).

1. Introduction

The convolutional layers of the CNN are equivariant to translation which means shifting the original image and then feeding through the network is similar to first feeding the original input image and then shifting the feature maps [1]. However, the standard CNNs are not equivariant to other affine geometric transformations such as rotation and reflection. Invariance or equivariance with respect to different geometric transformations is one of the highly desirable properties of the deep learning models, especially for the task of image classification. As a result, novel classes of CNNs are proposed in order to encode the geometric structures within the architectures of CNNs. Chen and Welling [2] proposed a new class of convolutional neural networks known as group-equivariant CNN (G-CNNs) based on the theory of groups. The convolutional layer of the standard CNN is replaced with group convolutional layer in G-CNN. It turns out that the convolutional layer of CNN is special case of group convolutional layer and the group convolutional layers are the only layers in the linear neural networks that are guaranteed to be equivariant. The key feature of G-CNNs is its equivariance with respect to the transformations defined by the special group. However, the G-CNNs are equivariant to discrete transformations that leave the pixel grid intact (e.g., $90°$-rotations, translations and reflections). Worall et al. [3] proposed the Harmonic networks (H-Nets) by restricting the filters of the convolutional layers to be from the circular harmonic family. H-Nets hard-bake patch-wise $360°$ rotational equivariance into deep image representation. Hoogeboom et al. [4] proposed the HexaConvs which has 6-fold rotational symmetry as compared to the 4-fold rotational symmetry of G-CNNs[2] which allows more parameter sharing. The proposed HexaConvs outperforms the planar CNNs and G-CNNs. Bekkers [5] proposed a modular framework for G-CNNs for arbitrary Lie groups in order to overcome the limitations of standard G-CNN[2] which is practically applicable to only either discrete groups or continuous compact groups. In another effort Bekkers et al. [1] proposed the *SE(2)* group convolutional which proves that a concatenation of two roto-translations results in a net roto-translation. The three new layers are introduced to achieve fully equivariance throughout the CNN: a lifting layer, group convolutional layer and a projection layer. As a result he proposed *SE(2)* equivariant G-CNN is not only equivariant to orientations in the input data that lay on the pixel gird but also to orientations that are out of the pixel grid. The key features of *SE(2)* G-CNNs are: i) it learns the geometric structures into the network architecture and ii) equivariance is guaranteed [1]. Sabour et al. [6] proposed the capsule neural networks which are inspired from the part of brain, called primary visual cortex, responsible for recognition. Each capsule in the capsule network is a group of convolutional neurons and the dynamic routing algorithm is developed for learning between the primary and digit capsules. Capsule networks are equivariant to complex global transformations.

2. Proposed Method

Let $f(r,\theta)$ be a 2-D function in the continuous polar domain. The Zernike moments (ZMs) of the function $f(r,\theta)$ of order p and repetition q over the unit disk are defined as follows[7]:

$$ZM_{p,q}(f) = \frac{p+1}{\pi}\int_{0}^{2\pi}\int_{0}^{1} f(r,\theta)R_{p,q}(r)e^{-iq\theta}rdrd\theta, \qquad (1)$$

where $i=\sqrt{-1}, p \in Z^{+}, |q| \le p, p-|q| = even$, and $R_{p,q}(r)$ is the radial kernel function which is defined as follows:

$$R_{p,q}(r) = \sum_{k=0}^{\frac{p-|q|}{2}} \frac{(-1)^k (p-k)!}{k!\left(\frac{p+|q|}{2}-k\right)!\left(\frac{p-|q|}{2}-k\right)!} r^{p-2k}. \qquad (2)$$

Since the kernel function $R_{p,q}(r)$ is orthogonal and complete over the unit disk, the function $f(r,\theta)$ can be reconstructed as follows:

$$\hat{f}(r,\theta) = \sum_{p=0}^{p_{max}} \sum_{q=-p_{max}}^{p_{max}} ZM_{p,q}(f)R_{p,q}(r)e^{iq\theta}, \qquad (3)$$

where p_{max} is the maximum order of moments which could be set to any positive integer number. The higher is the value of p_{max}, the closer will be $\hat{f}(r,\theta)$ to $f(r,\theta)$.

In the case of standard CNNs, some numbers of fully connected layers are applied to combine the information from the various filter responses. These fully connected layers do not maintain rotation equivariance or invariance properties. For G-CNNs, convolution and downsampling are applied until the spatial dimensions are eliminated and the resulting feature map becomes merely a feature vector equal to the size of the number of filters in the convolution layer. Instead of directly feeding the filter responses to fully connected layers for CNNs and downsampling filter responses to a feature vector for G-CNNs. In this work we compute the magnitude of ZMs over the feature maps before fully connected layers in order to make the filter responses of the last convolutional layer invariant to rotation which is shown in Fig. 1.

3. Experimental Results

The proposed model architectures Z2CNN+ZMs and P4CNN+ZMs are evaluated using rotated MNIST dataset which has been utilized as a benchmarking dataset for the previous works on rotation invariance. The rotated MNIST dataset consists of total 60,000 images which are divided into train, validation and test datasets. The number of images in the training, validation and testing datasets are 10000, 2000 and 50000, respectively. The test error (%) achieved by the existing Z2CNN (standard CNN) and P4CNN (G-CNN) and proposed Z2CNN+ZMs and P4CNN+ZMs are shown in Table 1. It can be clearly observed from the table that the proposed models outperforms over the existing models.

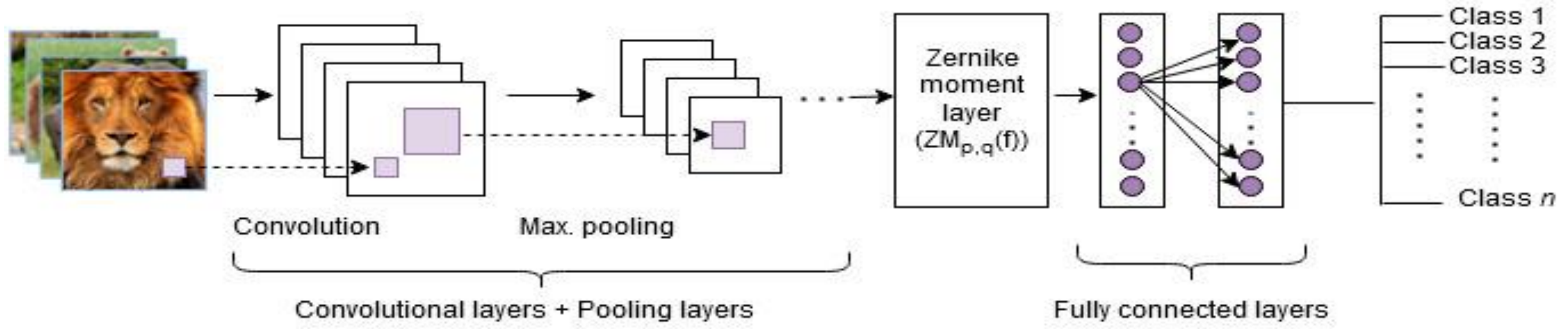

Figure 1. Proposed model architecture for Z2CNN+ZMs and P4CNN+ZMs.

Table 1. Comparison of existing Z2CNN and P4CNN, and proposed Z2CNN+ZMs and P4CNN+ZMs models on rotated MNIST dataset.

Model		Test Error (%)
Existing	Z2CNN	5.03
	P4CNN	2.28
Proposed	Z2CNN+ZMs	3.98
	P4CNN+ZMs	1.78

References

[1]. Bekkers, E. J., Lafarge, M. W., Veta, M., Eppenhof, K. A., Pluim, J. P., & Duits, R. (2018, September). Roto-translation covariant convolutional networks for medical image analysis. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 440-448). Springer, Cham.

[2]. Cohen, T., & Welling, M. (2016, June). Group equivariant convolutional networks. In International conference on machine learning (pp. 2990-2999).

[3]. Worrall, D. E., Garbin, S. J., Turmukhambetov, D., & Brostow, G. J. (2017). Harmonic networks: Deep translation and rotation equivariance. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 5028-5037).

[4]. Hoogeboom, E., Peters, J. W., Cohen, T. S., & Welling, M. (2018). Hexaconv. arXiv preprint arXiv:1803.02108.

[5]. Bekkers, E. J. (2019). B-spline cnns on lie groups. arXiv preprint arXiv:1909.12057.

[6]. Sabour, S., Frosst, N., & Hinton, G. E. (2017). Dynamic routing between capsules. In Advances in neural information processing systems (pp. 3856-3866).

[7]. Singh, C., Walia, E., & Upneja, R. (2013). Accurate calculation of Zernike moments. Information Sciences, 233, 255-275.