Roto-Translation Covariant Convolutional Networks for Medical Image Analysis

Erik J Bekkers[*,1], **Maxime W Lafarge**[*,2],
Mitko Veta[2], **Koen AJ Eppenhof**[2], **Josien PW Pluim**[2], **Remco Duits**[1]
Eindhoven University of Technology, [1]Department of Mathematics and Computer Science
and [2]Department of Biomedical Engineering, Eindhoven, The Netherlands
[*]Joint main authors - e.j.bekkers@tue.nl, m.w.lafarge@tue.nl

Abstract

We propose a framework for rotation and translation covariant deep learning using $SE(2)$ group convolutions. The group product of the special Euclidean motion group $SE(2)$ describes how a concatenation of two roto-translations results in a net roto-translation. We encode this geometric structure into convolutional neural networks (CNNs) via $SE(2)$ group convolutional layers. We introduce three layers: a *lifting layer* which lifts a 2D (vector valued) image to an $SE(2)$-image, i.e., 3D (vector valued) data whose domain is $SE(2)$; a *group convolution layer* from and to an $SE(2)$-image; and a *projection layer* from an $SE(2)$-image to a 2D image. The lifting and group convolution layers are $SE(2)$ *covariant* (the output roto-translates with the input). The final projection layer, a maximum intensity projection over rotations, makes the full CNN rotation *invariant*. We show with three different problems in histopathology, retinal imaging, and electron microscopy that with the proposed group CNNs, state-of-the-art performance can be achieved, without the need for data augmentation by rotation and with increased performance compared to standard CNNs that do rely on augmentation.

1 Introduction

Here we summarize the method and results of our recent [1] generalization of $\mathbb{R}^2$ convolutional neural networks (CNNs) to $SE(2)$ group CNNs (G-CNNs), in which the data lives on position orientation space and in which the convolution layers are defined in terms of representations of the special Euclidean motion group $SE(2)$. In essence this means that we replace the convolutions (with translations of a kernel) by $SE(2)$ group convolutions (with roto-translations of a kernel). The advantage of the proposed approach compared to standard $\mathbb{R}^2$ CNNs is that rotation covariance is encoded in the network design and does not have to be learned by the convolution kernels. E.g., a feature that may appear in the data under several orientations does not have to be learned for each orientation, but only once. As a result, there is no need for data augmentation by rotation and the kernel weights (that no longer need to learn rotation covariance) become available to increase the CNNs expressive capacity. Moreover, the proposed group convolution layers are compatible with standard CNN modules, allowing for easy integration in popular CNN designs.

A main objective of medical image analysis is to develop models that are invariant to the shape and appearance variability of the structures of interest, including their arbitrary orientations. Rotation-invariance is a desired property which our G-CNN framework generically deals with. We show state-of-the-art results with improvement over standard 2D CNNs on three different medical imaging tasks: mitosis detection in histopathology images, vessel segmentation in retinal images and cell boundary segmentation in electron microscopy (EM).

1st Conference on Medical Imaging with Deep Learning (MIDL 2018), Amsterdam, The Netherlands.

In relation to other approaches that incorporate rotation invariance/covariance in the network design, group convolution approaches [2, 3, 4, 5, 6] most naturally extend standard CNNs by replacing the convolution operators. In contrast to these G-CNN methods, we rely on kernel rotation via linear interpolation which simply appears in the CNN architecture as a (sparse) matrix-vector multiplication that maps a set of base weights to a full set of rotated kernels. As such our method allows to sample an arbitrary number of rotations (in contrast to $N = 4$ in [2] and [3]) and we do not have any constraints on the shape of the kernel (in contrast to steerable kernels [4] or separable [5] kernels).

2 $SE(2)$ convolutional neural networks

On $\mathbb{R}^2$ we define cross-correlation via inner products of translated kernels: $(k \star_{\mathbb{R}^2} f)(\mathbf{x}) := (\mathcal{T}_{\mathbf{x}} k, f)_{\mathbb{L}_2(\mathbb{R}^2)} := \int_{\mathbb{R}^2} k(\mathbf{x}' - \mathbf{x}) f(\mathbf{x}') \mathrm{d}\mathbf{x}'$,with $\mathcal{T}_{\mathbf{x}}$ the translation operator, the left-regular representation of the translation group $(\mathbb{R}^2, +)$. In the $SE(2)$ lifting layer we now simply replace translations of k by roto-translations via the $SE(2)$ representation $\mathcal{U}_g$, see e.g. in [1, Eq. (2)].

The $SE(2)$ lifting layer: Let $\underline{f}, \underline{k} : \mathbb{R}^2 \to \mathbb{R}^{N_c}$ be a vector valued 2D image and kernel (with N_c channels), with $\underline{f} = (f_1, \ldots, f_{N_c})$ and $\underline{k} = (k_1, \ldots, k_{N_c})$, then the group lifting correlations for vector valued images are defined by

$$(\underline{k} \tilde{\star} \underline{f})(g) := \sum_{c=1}^{N_c} (\mathcal{U}_g k_c, f_c)_{\mathbb{L}_2(\mathbb{R}^2)} = \sum_{c=1}^{N_c} \int_{\mathbb{R}^2} k_c(\mathbf{R}_\theta^{-1}(\mathbf{y} - \mathbf{x})) f_c(\mathbf{y}) \mathrm{d}\mathbf{y}. \tag{1}$$

These correlations *lift* 2D image data to data that lives on the 3D position orientation space $\mathbb{R}^2 \times S^1 \equiv SE(2)$. The *lifting layer* that maps from a vector image $\underline{f}^{(l-1)} : \mathbb{R}^2 \to \mathbb{R}^{N_{l-1}}$, with N_{l-1} channels at layer $l-1$, to an $SE(2)$ vector image $\underline{F}^{(l)} : SE(2) \to \mathbb{R}^{N_l}$ using a set of N_l kernels $\mathbf{k}^{(l)} := (\underline{k}_1^{(l)}, \ldots, \underline{k}_{N_l}^{(l)})$, each with N_{l-1} channels, is then defined by

$$\underline{F}^{(l)} = \mathbf{k}^{(l)} \tilde{\star} \underline{f}^{(l-1)} := \left(\begin{array}{ccc} \underline{k}_1^{(l)} \tilde{\star} \underline{f}^{(l-1)} & , \ \ldots \ , & \underline{k}_{N_l}^{(l)} \tilde{\star} \underline{f}^{(l-1)} \end{array} \right). \tag{2}$$

The $SE(2)$ group convolution layer: Let $\underline{F}, \underline{K} : SE(2) \to \mathbb{R}^{N_c}$ be a vector valued $SE(2)$ image and kernel, with $\underline{F} = (F_1, \ldots, F_{N_c})$ and $\underline{K} = (K_1, \ldots, K_{N_c})$, then the group correlations are

$$(\underline{K} \star \underline{F})(g) := \sum_{c=1}^{N_c} (\mathcal{L}_g K_c, F_c)_{\mathbb{L}_2(SE(2))} = \sum_{c=1}^{N_c} \int_{SE(2)} K_c(g^{-1} \cdot h) F_c(h) \mathrm{d}h, \tag{3}$$

with $\mathcal{L}_g$ the left-regular representation on $\mathbb{L}_2(SE(2))$ and with g^{-1} and $g \cdot h$ denoting resp. the group inverse and group product, see e.g. [1, Ch. 2.1]. Here $(K, F)_{\mathbb{L}_2(SE(2))} := \int_{SE(2)} K(h) F(h) \mathrm{d}h$ denotes the inner product on $\mathbb{L}_2(SE(2))$. A set of $SE(2)$ kernels $\mathbf{K}^{(l)} := (\underline{K}_1^{(l)}, \ldots, \underline{K}_{N_l}^{(l)})$ defines a group convolution layer, mapping from $\underline{F}^{(l-1)}$ with $N_{(l-1)}$ channels to $\underline{F}^{(l)}$ with $N_{(l)}$ channels, via

$$\underline{F}^{(l)} = \mathbf{K}^{(l)} \star \underline{F}^{(l-1)} := \left(\begin{array}{ccc} \underline{K}_1^{(l)} \star \underline{F}^{(l-1)} & , \ \ldots \ , & \underline{K}_{N_l}^{(l)} \star \underline{F}^{(l-1)} \end{array} \right). \tag{4}$$

The projection layer: Projects a multi-channel $SE(2)$ image back to $\mathbb{R}^2$ via

$$\underline{f}^{(l)}(\mathbf{x}) = \max_{\theta \in [0, 2\pi]} \underline{F}^{(l)}(\mathbf{x}, \theta). \tag{5}$$

3 Experiments, Results and Conclusion

Experiment: Based on the layers defined in Eqs. (2), (4) and (5) we define a straight forward G-CNN chain in which the first 4 layers are G-conv layers with 5x5 kernels, followed by a projection layer, and the last 2 layers are standard 1x1 2D conv layers. After each layer we apply batch normalization and a ReLU, see [1] for full details. We consider three different tasks in three different modalities: mitosis detection in histopathology images (validated on the AMIDA13 data using F_1-scores [7]), vessel segmentation in retinal images (validated on the DRIVE data using AUC values [8]) and cell boundary segmentation in EM (validated on the ISBI-EM data using the Rand metric [9]).

We consider the sampling of $SE(2)$ with $N \in \{1, 2, 4, 8, 16\}$ number of orientations. The 2D input is augmented at train and test time with transposed versions. For reference we also include transpose plus 90° rotation augmentation for the $N = 1$ experiment (which coincides with a standard 2D CNN) to show that these are not necessary in our $SE(2)$ networks for $N \geq 4$. Each experiment is repeated 3 times with random initialization and sampling to get an estimate of the mean and variance on the performance. For a fair comparison for different N the overall number of weights is matched. The number of "2D" activations in the last three layers is also matched. See [1, Table 1] for full details.

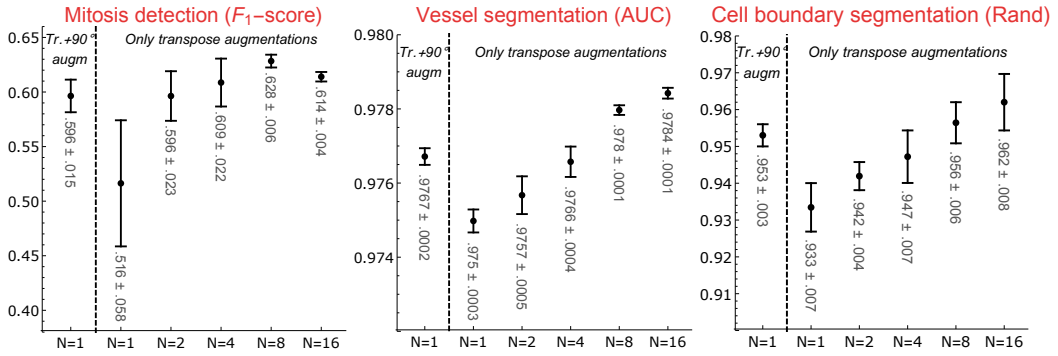

Figure 1: Mean results (± 1 std. dev.).

Results: See Fig. 1. In each experiment we see that the performance of the baseline with extra rotation augmentations is reached by the non-augmented G-CNNs for $N \geq 4$, and even is surpassed for $N \geq 8$. In the first two experiments we also observe that the variance on the output is reduced with increasing N. Our results on the public datasets match or improve upon the state of the art with the following scores: F_1-score=0.628 ± 0.006 for mitosis detection, AUC = 0.9784 ± 0.0001 for vessel segmentation, Rand = 0.962 ± 0.008 for cell boundary segmentation.

Conclusion: We conclude that it is beneficiary to include $SE(2)$ group convolution layers in CNN network design, as this avoids the need for rotation augmentation and it improves overall performance. In all three medical imaging problems we achieved state-of-the-art results with the same (basic) network design. Based on these results we expect that $SE(2)$ layers may lead to a further performance increase when embedded in more complex network designs, such as the popular UNets and ResNets.

References

[1] Bekkers, E., Lafarge, M., Veta, M., Eppenhof, K., Pluim, J., Duits, R.: Roto-translation covariant convolutional networks for medical image analysis. arXiv preprint arXiv:1804.03393 (2018)

[2] Cohen, T., Welling, M.: Group equivariant convolutional networks. In: Int. Conf. on Machine Learning. (2016) 2990–2999

[3] Dieleman, S., De Fauw, J., Kavukcuoglu, K.: Exploiting cyclic symmetry in convolutional neural networks. arXiv preprint arXiv:1602.02660 (2016)

[4] Weiler, M., Hamprecht, F.A., Storath, M.: Learning steerable filters for rotation equivariant cnns. arXiv preprint arXiv:1711.07289 (2017)

[5] Oyallon, E., Mallat, S., Sifre, L.: Generic deep networks with wavelet scattering. arXiv preprint arXiv:1312.5940 (2013)

[6] Bekkers, E.J., Loog, M., ter Haar Romeny, B.M., Duits, R.: Template matching via densities on the roto-translation group. IEEE tPAMI **40**(2) (2018) 452–466

[7] Veta, M., van Diest, P., Willems, S., et al.: Assessment of algorithms for mitosis detection in breast cancer histopathology images. MEDIA **20**(1) (2015) 237–248

[8] Staal, J., Abràmoff, M.D., Niemeijer, M., et al.: Ridge-based vessel segmentation in color images of the retina. IEEE TMI **23**(4) (2004) 501–509

[9] Arganda-Carreras, I., Turaga, S.C., et al.: Crowdsourcing the creation of image segmentation algorithms for connectomics. Front. in neuroanatomy **9** (2015) 142