



# Roto-translation equivariant convolutional networks: Application to histopathology image analysis

Maxime W. Lafarge<sup>a,\*</sup>, Erik J. Bekkers<sup>b</sup>, Josien P.W. Pluim<sup>a</sup>, Remco Duits<sup>b</sup>, Mitko Veta<sup>a</sup>

<sup>a</sup> Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven, the Netherlands

<sup>b</sup> Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, the Netherlands

## ARTICLE INFO

### Article history:

Received 4 December 2019

Revised 3 June 2020

Accepted 14 August 2020

Available online 31 October 2020

### Keywords:

Group convolutional neural network

Roto-translation equivariance

Computational pathology

Mitosis detection

Tumor detection

Nuclei segmentation

## ABSTRACT

Rotation-invariance is a desired property of machine-learning models for medical image analysis and in particular for computational pathology applications. We propose a framework to encode the geometric structure of the special Euclidean motion group  $SE(2)$  in convolutional networks to yield translation and rotation equivariance via the introduction of  $SE(2)$ -group convolution layers. This structure enables models to learn feature representations with a discretized orientation dimension that guarantees that their outputs are invariant under a discrete set of rotations.

Conventional approaches for rotation invariance rely mostly on data augmentation, but this does not guarantee the robustness of the output when the input is rotated. At that, trained conventional CNNs may require test-time rotation augmentation to reach their full capability.

This study is focused on histopathology image analysis applications for which it is desirable that the arbitrary global orientation information of the imaged tissues is not captured by the machine learning models. The proposed framework is evaluated on three different histopathology image analysis tasks (mitosis detection, nuclei segmentation and tumor detection). We present a comparative analysis for each problem and show that consistent increase of performances can be achieved when using the proposed framework.

© 2020 The Author(s). Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

## 1. Introduction

Invariance to irrelevant factors of variability is a desirable property of machine learning models, in particular for medical image analysis problems for which models are expected to generalize to unseen shapes, appearances, or to arbitrary orientations. For example, histopathology image analysis problems require processing a digital slide of a stained specimen whose global orientation is strictly arbitrary. Indeed, in the preparation workflow of histology slides, resection of the tissue is done arbitrarily and local structures within the section can have any three-dimensional orientation. In this context, models whose output varies with the orientation of the input constitute a source of uncertainty. The output of such image analysis systems should be rotation invariant, meaning that the output of a model should not change when its input is rotated.

Convolutional Neural Networks (CNNs) are the method of choice to solve complex image analysis tasks, in part due to the translation co-variance induced by trainable  $\mathbb{R}^2$  convolution operators. In theory, this structure allows CNNs to learn features in any orientation given sufficient capacity. For example, if a specific edge detector is a relevant filter for the task at hand, it is expected that the CNN learns this filter in all possible directions. Typical solutions to obtain rotation invariance consist in augmenting the dataset by generating additional randomly rotated samples, with the expectation that the model will learn the relevant features that are artificially observed under these additional orientations. Although data augmentation is a way to induce and encourage an invariance prior, such approaches do not guarantee conventional CNNs to be rotation-invariant. Furthermore, with such approaches it is common practice to average predictions of the trained model on a set of rotated inputs at test time: this can increase the robustness of the model, however it comes at the cost of a computational overhead.

We propose to replace convolutions in  $\mathbb{R}^2$  by group convolutions using representations of the special Euclidean motion group

\* Corresponding author.

E-mail address: [m.w.lafarge@tue.nl](mailto:m.w.lafarge@tue.nl) (M.W. Lafarge).

$SE(2)$  (roto-translation of a kernel) so as to explicitly encode the orientation of the learned features. This structure ensures that the learned representation is co-variant/equivariant with the orientation of the input for rotations that lay on the pixel grid and to some extent for rotations that are out of the pixel grid. This equivariance property implies that an oriented feature of interest will get extracted independently of the spatial orientation of the input. We achieve orientation encoding at resolution levels higher than 90-degree via bi-linear interpolation of the  $SE(2)$  convolution kernels. Finally rotation invariance can be achieved via a projection operation with respect to the encoded orientation of the learned representation.

**Contributions** This work builds upon our previous work presented at the MICCAI conference 2018 (Bekkers et al., 2018a). In addition to a more detailed description of the proposed framework, we now present a comparative analysis of models with different angular discretization levels of the  $SE(2)$ -image representations. Here we focus on three types of histopathology image analysis problems (mitosis detection, nuclei segmentation and tumor detection), for which we conduct experiments on popular and realistic benchmark datasets. With this we also show that the  $SE(2)$ -image representations can be integrated in other classical CNN architectures such as U-net (Ronneberger et al., 2015). Finally, in a new series of in-depth experimental analyses we show an increased robustness of the proposed group convolutional neural networks (G-CNNs) compared to standard CNNs with respect to rotational variations in the data. This includes a quantitative and qualitative assessment of rotational invariance of the trained networks, as well as a data regime analysis in which we investigate the effect of increased angular resolution when the data availability is reduced.

## 2. Rotation invariance, related work, and contributions

### 2.1. Rotation invariance via G-CNNs

We distinguish between invariance and equivariance/covariance as follows. An artificial neural network (NN) is invariant with respect to certain transformations when the output of the network does not change under transformations on the input. We call a NN equivariant, or covariant<sup>1</sup>, when the output transforms in a predictable way when the input is transformed (we formalize this statement in Section 3.2). The property of equivariance guarantees that no information is lost when the input is transformed. Standard CNNs are equivariant to translations: if the input is translated the output translates accordingly and we do not need to worry about learning how to deal with translated inputs. It turns out that *group convolution layers* are the only type of linear NN layers that are guaranteed to be equivariant (see e.g. (Bekkers, 2019, Thm. 1)) and that the standard convolution layer is a special case that is translation equivariant. In this paper, we construct  $SE(2)$  equivariant group convolution layers and with it build G-CNNs with which we solve problems in histopathology that require rotation invariance.

Nowadays, rotation invariance is often still dealt with via data augmentations. In such an approach the data is rotated during training time while keeping the target label fixed, thereby aiming for the network to learn how to classify input samples regardless of their orientation. Downsides of this approach are that 1) valuable network capacity is spent on learning geometric behavior at the cost of descriptive representation learning, 2) rotation invariance is not guaranteed, and 3) augmentation only captures geometric invariance globally. G-CNNs solve these problems by hard-

coding geometric structure into the network architecture such that 1) geometric behavior does not have to be learned, 2) rotation invariance is guaranteed by construction, and 3) each group convolution layer achieves local equivariance on its own, so that global equivariance is still obtained when the layers are stacked.

The local-to-global equivariance property means that G-CNNs recognize both low-level features (e.g. edges), mid-level features (e.g. individual cells), and high-level features (e.g. tissue structure) independent of their orientations. In this paper we experimentally show that  $SE(2)$  equivariant G-CNNs indeed solve all three aforementioned problems and that in fact the added geometric structures leads to networks that significantly outperform classical CNNs trained with data-augmentation.

### 2.2. Related work on G-CNNs

#### 2.2.1. G-CNN methods

In the seminal work by Cohen and Welling (2016) a framework is proposed for group equivariant CNNs. In G-CNNs, the convolution operator is redefined in terms of actions of a transformation group, and by consistent use of the group structure (rules for concatenating transformations) equivariance is ensured. They showed a significant performance gain of G-CNNs over classical CNNs, however, the practical applicability was limited to discrete transformation groups that leave the pixel grid intact (s.a. 90° rotations and reflections). Subsequent work in the field focused on expanding the class of transformation groups that are suitable for G-CNNs by:

1. Working with a grid that has more symmetries than the standard Cartesian grid (Hooeboom et al., 2018).
2. Expanding convolution kernels in a special basis, tailored to the transformation group of interest, that enables to build steerable CNNs (Worrall et al., 2017)
3. Relying on interpolation methods to transform kernels (Bekkers et al. (2018a), or relying on analytic basis functions and sample the transformed kernels at arbitrary resolution (Weiler et al., 2017; Bekkers et al., 2018b).

Extensions to 3D transformation groups are described in Worrall and Brostow (2018), Winkels and Cohen (2019), Weiler et al. (2018), Andrearczyk et al. (2019), generalization to equivariance beyond roto-translations are described in Bekkers (2019), Worrall and Welling (2019), extension to spherical data are described in Cohen et al. (2018a), Kondor and Trivedi (2018), Thomas et al. (2018), Esteves et al. (2018a), and additional theoretical results and further generalizations of G-CNNs are described in Cohen et al. (2018b), Kondor and Trivedi (2018), Cohen et al. (2019). Applications of G-CNN methods in medical image analysis are discussed below in Section 2.2.4.

Although the first of the above generalizations elegantly enables an exact implementation of G-CNNs of roto-translations with a finer resolution than the 90° rotation angles of Cohen and Welling (2016), it is a very specific approach that does not generalize well to other groups. The second approach does not require to sample transformed kernels at all, but works exclusively by manipulations of basis coefficients in a similar way as standard 2D convolutions (and translations) can be described in the Fourier domain. This approach however requires careful bookkeeping of the coefficients, only optimizes over kernels expressible by the basis, and the choice for non-linear activation functions is limited. In this paper we rely on the third approach. We build upon our previous work (Bekkers et al., 2018a) and use bi-linear interpolation to efficiently transform (unconstrained) convolution kernels. This allows us to build  $SE(2)$  equivariant G-CNNs at arbitrary angular resolutions.

<sup>1</sup> Terminology changes between fields of study (mathematics, physics, machine learning) and often refer to the same. Following custom in machine learning research we will use the term equivariance.

### 2.2.2. Rotation equivariant machine learning

Prior, and in parallel, to the above discussed G-CNN methods, group convolution methods for pattern recognition have been proposed that, at the time, were not regarded as G-CNNs or not treated in the full generality of (end-to-end) deep learning. E.g., Gens and Domingos (2014) redefine the convolution operator and construct sparse (approximative) group convolution layers that are used to build what they called deep symmetry networks. Scattering convolution networks, as proposed by Mallat (2012), involve a concatenation of separable group convolutions with well-designed hand-crafted filters followed by the modulus as activation function. Other examples are orientation score based template matching (Bekkers et al., 2015), cyclic symmetry networks (Dieleman et al., 2016), oriented response networks (Zhou et al., 2017), and vector field networks (Marcos et al., 2017), which can all be considered instances of roto-translation equivariant G-CNNs.

Other techniques that focus on equivariance properties of CNNs work via transformations on input feature maps, rather than transformations of convolution kernels as in G-CNNs, and are closely related to spatial transformer networks (Jaderberg et al., 2015). These methods include warped CNNs (Henriques and Vedaldi, 2017), polar transformer networks (Esteves et al., 2018b), and equivariant transformer networks (Tai et al., 2019). Although these methods describe elegant and efficient ways for achieving (global) equivariance, they often break translation equivariance and local symmetries as the transformations act globally on the whole inputs.

### 2.2.3. Group theory in medical image analysis

Equivariance constraints and group theory take a prominent position in the mathematical foundations of classical image analysis, e.g., in scale space and wavelet theory. In medical image analysis, group theoretical algorithms enable to respect natural equivariance constraints and deal with context and the complex geometries that are abundant in medical images. Examples of group theoretical techniques, closely related to G-CNNs, are orientation score (Duits et al., 2007; Janssen et al., 2018) methods such as crossing preserving vessel enhancement based on gauge theory on Lie groups (Franken and Duits, 2009; Hannink et al., 2014; Duits et al., 2016), vessel and nerve fiber enhancement (in diffusion imaging) via group convolutions with Gaussian (derivative) kernels (Duits and Franken, 2011; Zhang et al., 2015; Portegies et al., 2015), and anatomical landmark recognition via group convolutions (Bekkers, 2019). In other, non-convolutional methods in medical image analysis, group theory provides a powerful tool to deal with symmetries and geometric structure, such as in statistical shape atlases (Hefny et al., 2015), shape matching (Hou et al., 2018), registration (Arsigny et al., 2006; Ashburner, 2007) and in general in statistics on non-Euclidean data structures (Pennec et al., 2019). Following this successful line of geometry driven methods in medical image analysis, we propose in this paper to rely on G-CNNs to solve tasks in histopathology in an end-to-end learning setting.

### 2.2.4. G-CNNs in medical image analysis

For many medical image analysis tasks, the location, reflection or orientation of objects of interest should not affect the output of the developed models. Although typical solutions rely on data augmentation, several studies investigated G-CNNs in the context of medical image analysis to leverage this prior into building equivariant models that outperform classical CNNs.

In Winkels and Cohen (2018), Winkels and Cohen (2019), Andrearczyk et al. (2019), G-CNNs were used to detect pulmonary nodules in CT scans. G-CNNs were also investigated for segmentation tasks in dermoscopy images (Li et al., 2018), retinal images (Bekkers et al., 2018a) and microscopy images (Bekkers et al., 2018a; Chidester et al., 2019a; Graham et al., 2019). Chidester et al. (2019b) proposed a variation of G-CNNs for the

classification of sub-cellular protein localization in microscopy images.

Rotation-equivariant models have shown to be particularly efficient for problems in histopathology images, at cell level for mitosis detection (Bekkers et al., 2018a), nuclei segmentation (Chidester et al., 2019a), and at higher tissue levels for tumor detection in lymph node sections (Veeling et al., 2018) and gland-lumen segmentation in colon histology images (Graham et al., 2019).

## 3. Material and methods

We evaluate the proposed framework on three relevant histopathology image analysis tasks: mitosis detection, nuclei classification, and patch-based tumor detection. In this section, we first describe the benchmark datasets corresponding to the analysis tasks, that we used to train and evaluate the models. We then describe the relationship between the proposed framework and group theory, and our proposed implementation via bi-linear interpolation of rotated convolution kernels.

### 3.1. Datasets

We chose three popular benchmark datasets of hematoxylin-eosin stained histological slides, in order to assess the performances of the proposed framework and its variants in a controlled and reproducible setup. We chose datasets for which objects of interest are observed at different scales, thus covering a range of problems that are typically addressed in histopathology image analysis. In these datasets, we assume that the orientation of the objects of interest is irrelevant for the classification task. Therefore we hypothesize that any bias in the orientation information captured by a non-rotation-invariant CNN could be reflected in its performance on the selected benchmarks. This hypothesis will be experimentally confirmed in Section 5.

**Mitosis detection** We used the public dataset *AMIDA13* (Veta et al., 2015) that consists of high power-field (HPF) images (resolution  $\sim 0.25 \mu\text{m}/\text{px}$ ) from 23 breast cancer cases. Eight cases (458 mitotic figures) were used to train the models and four cases (92 mitoses) for validation. Evaluation is performed on a test set of 11 independent cases (533 mitoses), following the evaluation procedure of the *AMIDA13* challenge, for details see (Veta et al., 2015).

**Multi-organ nuclei segmentation** We used the subset of the public multi-organ dataset introduced by Kumar et al. (2017), that consists of 24 HPF images (resolution  $\sim 0.25 \mu\text{m}/\text{px}$ ), selected from WSIs of four different tissue types (Breast, Liver, Kidney and Prostate), provided by *The Cancer Genome Atlas* (Network et al., 2012), associated with mask annotations of nucleus instances. We used the balanced dataset split proposed in Lafarge et al. (2019):  $4 \times 3$  HPF images for training (7337 nuclei),  $4 \times 1$  HPF images for validation (1474 nuclei) and  $4 \times 2$  HPF images for testing (4130 nuclei). Given the high staining variability of the dataset, all the images were stain normalized using the method described in Macenko et al. (2009).

**Patch-based tumor detection** We used the public *PCam* dataset introduced by Veeling et al. (2018), that consists of 327,680 image patches (resolution  $\sim 1 \mu\text{m}/\text{px}$ ), selected from WSIs of lymph node sections derived from the *Camelyon16* Challenge (Ehteshami Bejnordi et al., 2017). The patches are balanced across the two classes (benign or malignant), based on the tumor area provided in Ehteshami Bejnordi et al. (2017), and we used the dataset split proposed by Veeling et al. (2018).

**Data regime analysis** In order to study the behavior of the compared models when data availability is reduced, we analyzed the

performances under different data regimes, by using reduced versions of the training sets. We constructed:

- Three variations of the mitosis dataset by sequentially removing two cases out of the original eight.
- Two variations of the nuclei dataset by sequentially removing one HPF image per organ out of the original three HPF images per organ.
- Four variations of the patch-based tumor dataset by randomly removing 25%, 50%, 75% and 90% in each class-subset of the training data.

### 3.2. Group representation in CNNs

#### 3.2.1. The roto-translation group $SE(2)$

A group is a mathematical structure that consists of a set  $G$ , for example a collection of transformations, together with a binary operator  $\cdot$  called the group product that satisfies four fundamental properties: *Closure*: For all  $h, g \in G$  we have  $h \cdot g \in G$ ; *Identity*: There exists an identity element  $e$ ; *Inverse*: for each  $g \in G$  there exists an inverse element  $g^{-1} \in G$  such that  $g^{-1} \cdot g = g \cdot g^{-1} = e$ ; and *Associativity*: For each  $g, h, i \in G$  we have  $(g \cdot h) \cdot i = g \cdot (h \cdot i)$ .

The group product essentially describes how two consecutive transformations, e.g. by  $g, h \in G$ , result in a single net transformation  $(g \cdot h) \in G$ . Here, we consider the group of roto-translations, denoted<sup>2</sup> by  $SE(2) = \mathbb{R}^2 \rtimes SO(2)$ , which consists of the set of all planar translations (in  $\mathbb{R}^2$ ) and rotations (in  $SO(2)$ ), together with the group product given by

$$g \cdot g' = (\mathbf{x}, \mathbf{R}_\theta) \cdot (\mathbf{x}', \mathbf{R}_{\theta'}) = (\mathbf{R}_\theta \mathbf{x}' + \mathbf{x}, \mathbf{R}_{\theta+\theta'}), \quad (1)$$

with group elements  $g = (\mathbf{x}, \theta)$ ,  $g' = (\mathbf{x}', \theta') \in SE(2)$ , with translations  $\mathbf{x}, \mathbf{x}'$  and planar rotations by  $\theta, \theta'$ . The group acts on the space of positions and orientations  $\mathbb{R}^2 \times S^1$  via

$$g \cdot (\mathbf{x}', \theta') = (\mathbf{R}_\theta \mathbf{x}' + \mathbf{x}, \theta + \theta').$$

Since  $(\mathbf{x}, \mathbf{R}_\theta) \cdot (\mathbf{0}, 0) = (\mathbf{x}, \theta)$ , we can identify the group  $SE(2)$  with the space of positions and orientations  $\mathbb{R}^2 \times S^1$ . As such we will often write  $g = (\mathbf{x}, \theta)$ , instead of  $(\mathbf{x}, \mathbf{R}_\theta)$ . Note that  $g^{-1} = (-\mathbf{R}_\theta^{-1} \mathbf{x}, -\theta)$  since  $g \cdot g^{-1} = g^{-1} \cdot g = (\mathbf{0}, 0)$ .

#### 3.2.2. Group representations

The structure of the group can be mapped to other mathematical objects (such as 2D images) via representations. Representations of a group  $G$  are linear transformations  $\mathcal{R}_g : \mathbb{L}_2(X) \rightarrow \mathbb{L}_2(X)$ , parameterized by group elements  $g \in G$  that transform vectors, e.g. signals/images  $f \in \mathbb{L}_2(X)$  on a space  $X$ , and which share the group structure via

$$(\mathcal{R}_g \circ \mathcal{R}_h)(f) = \mathcal{R}_{gh}(f), \quad \text{with } g, h \in G.$$

We use different symbols for the representations of  $SE(2)$  on different type of data structures. In particular, we write  $\mathcal{R} = \mathcal{U}$  for the left-regular representation of  $SE(2)$  on 2D images  $f \in \mathbb{L}_2(\mathbb{R}^2)$ , and it is given by

$$(\mathcal{U}_g f)(\mathbf{x}') = f(\mathbf{R}_\theta^{-1}(\mathbf{x}' - \mathbf{x})), \quad (2)$$

with  $g = (\mathbf{x}, \theta) \in SE(2)$ ,  $\mathbf{x}' \in \mathbb{R}^2$ . It corresponds to a roto-translation of the image. We write  $\mathcal{R} = \mathcal{L}$  for the left-regular representation on functions  $F \in \mathbb{L}_2(SE(2))$  on  $SE(2)$ , which we refer to as  $SE(2)$ -images, and it is given by

$$(\mathcal{L}_g F)(g') = F(g^{-1} \cdot g') = F(\mathbf{R}_\theta^{-1}(\mathbf{x}' - \mathbf{x}), \theta' - \theta), \quad (3)$$

with  $g = (\mathbf{x}, \theta)$ ,  $g' = (\mathbf{x}', \theta') \in SE(2)$ . In Section 3.3 we define the G-CNN layers in terms of these representations.

<sup>2</sup> It is the semi-direct product (denoted by  $\rtimes$ ) of the group of planar translations  $\mathbb{R}^2$  and rotations  $SO(2)$ , i.e., it is not the direct product since the rotation part acts on the translations in (1) in the group product of  $SE(2)$ .

#### 3.2.3. Equivariance

Given the above definitions, we can formalize the notation of equivariance. An operator  $\Phi : \mathbb{L}_2(X) \rightarrow \mathbb{L}_2(Y)$  is equivariant with respect to a group  $G$  if

$$\Phi(\mathcal{R}_g(f)) = \mathcal{R}'_g(\Phi(f)), \quad (4)$$

with  $\mathcal{R}_g$  and  $\mathcal{R}'_g$  representations of  $G$  on respectively functions the domains  $X$  and  $Y$ . I.e., if we transform the input by  $\mathcal{R}_g$ , then we know that the output transforms via  $\mathcal{R}'_g$ . To ensure that we maintain the equivariance property (4) of linear operators  $\Phi$  it is required that we define such  $\Phi$  in terms of representations of  $G$ , that is, via group convolutions (see e.g. (Bekkers, 2019, Thm. 1), (Duits, 2005, Thm. 21), or (Cohen et al., 2018b, Thm. 6.1)).

### 3.3. $SE(2)$ group convolutional network layers

#### 3.3.1. Notation and 2D convolution layers

In the following we denote the space of multi-channel feature maps on a domain  $X$  by  $(\mathbb{L}_2(X))^N$ , with  $N$  the number of channels. The feature maps themselves are denoted by  $\underline{f} = (f_1, \dots, f_N)$ , with each channel  $f_i \in \mathbb{L}_2(X)$ . The inner product between such feature maps on  $X$  is denoted by

$$(\underline{k}, \underline{f})_{(\mathbb{L}_2(X))^N} := \sum_{c=1}^N (k_c, f_c)_{\mathbb{L}_2(X)}$$

with  $(k, f)_{\mathbb{L}_2(X)} = \int_X k(\mathbf{x}') f(\mathbf{x}') d\mathbf{x}'$  the standard inner product between real-valued functions on  $X$ . Then, with these notations we note that the classical 2D cross-correlation<sup>3</sup> operator can be defined in terms of inner products of input feature map  $\underline{f}$  with translated convolution kernels  $\underline{k}$  via

$$(\underline{k} \star_{\mathbb{R}^2} \underline{f})(\mathbf{x}) := (\mathcal{T}_{\mathbf{x}} \underline{k}, \underline{f})_{(\mathbb{L}_2(\mathbb{R}^2))^N} = \sum_{c=1}^N \int_{\mathbb{R}^2} k_c(\mathbf{x}' - \mathbf{x}) f_c(\mathbf{x}') d\mathbf{x}', \quad (5)$$

with  $\mathcal{T}_{\mathbf{x}}$  the translation operator, the left-regular representation of the translation group  $(\mathbb{R}^2, +)$ . It is well known that convolution layers  $\Phi$ , mapping between 2D feature maps (i.e. functions on  $X = Y = \mathbb{R}^2$ ), are equivariant with respect to translations. I.e. in Eq. (4) we let  $\mathcal{R}'_g = \mathcal{R}_g = \mathcal{T}_{\mathbf{x}}$  be the left-regular representation of the translation group with  $g = (\mathbf{x}) \in \mathbb{R}^2$ .

#### 3.3.2. Roto-translation equivariant convolution layers

Next we define two types of convolution layers that are equivariant with respect to roto-translations. We do so simply by replacing the translation operator in Eq. (5) with a representation of  $SE(2)$ . When the input is a 2D feature map  $\underline{f} \in (\mathbb{L}_2(\mathbb{R}^2))^N$  we need to rely on the representation  $\mathcal{U}_g$  of  $SE(2)$  on 2D images, and define the **lifting correlation**:

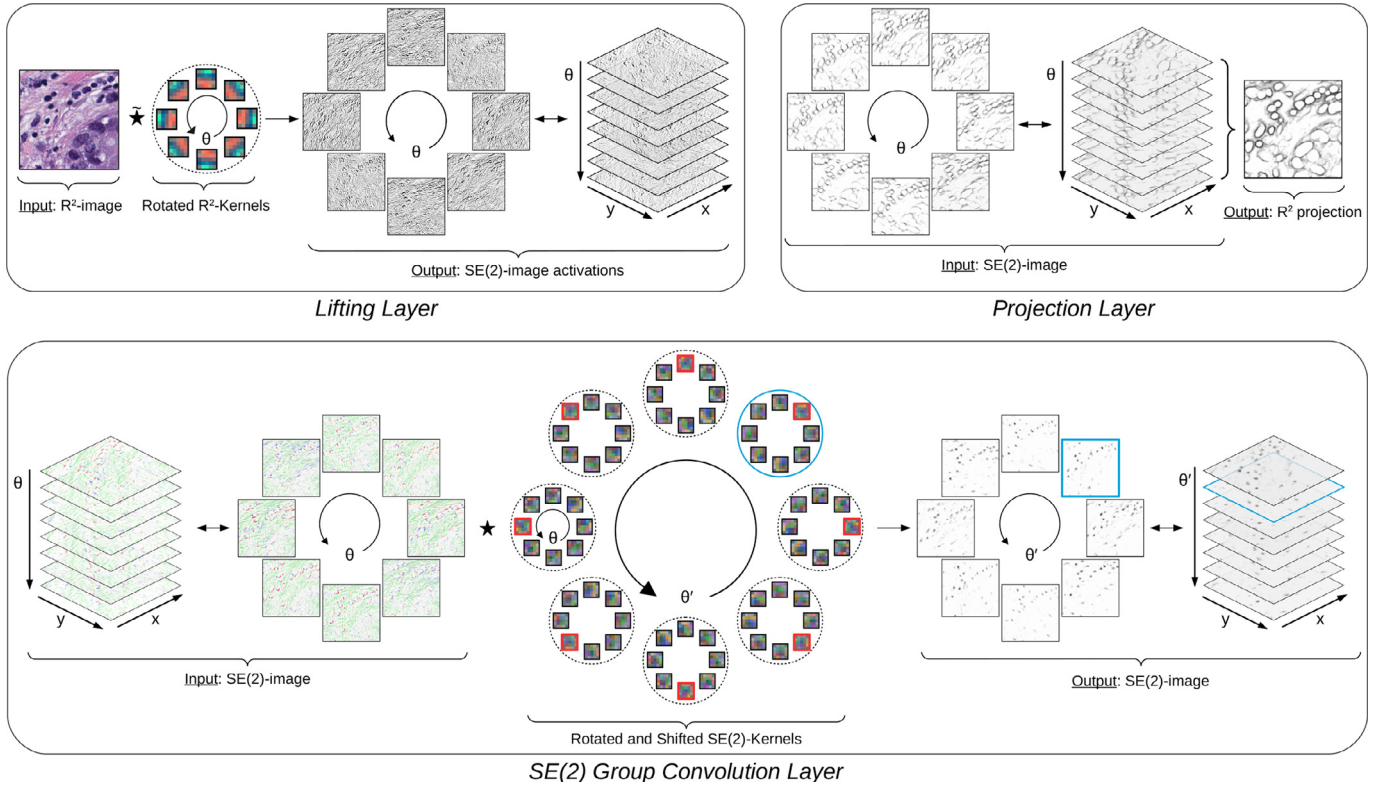
$$(\underline{k} \star \underline{f})(g) := (\mathcal{U}_g \underline{k}, \underline{f})_{(\mathbb{L}_2(\mathbb{R}^2))^N} = \sum_{c=1}^N \int_{\mathbb{R}^2} k_c(\mathbf{R}_\theta^{-1}(\mathbf{x}' - \mathbf{x})) f_c(\mathbf{x}') d\mathbf{x}'. \quad (6)$$

These correlations *lift* 2D image data to data that lives on the 3D position orientation space  $\mathbb{R}^2 \times S^1 \equiv SE(2)$  by matching convolution kernels under all possible translations and rotations.

We define the **lifting layer**, recall Fig. 1, as an operator  $\tilde{\Phi}^{(l)} : (\mathbb{L}_2(\mathbb{R}^2))^{N_{l-1}} \rightarrow (\mathbb{L}_2(SE(2)))^{N_l}$  that maps a 2D feature map  $\underline{f}^{(l-1)} \in (\mathbb{L}_2(\mathbb{R}^2))^{N_{l-1}}$  with  $N_{l-1}$  channels to an  $SE(2)$  feature map  $\underline{F}^{(l)} \in$

<sup>3</sup> In CNNs one can take a convolution or a cross-correlation viewpoint and since these operators simply relate via a kernel reflection, the terminology is often used interchangeably. We take the second viewpoint, our G-CNNs are implemented using cross-correlations.





**Fig. 1.** Illustration of the three types of layers investigated in our G-CNNs. The *lifting layer* uses a set of rotated kernels in  $\mathbb{R}^2$  to output an activation map that is an image on  $SE(2)$ . The  *$SE(2)$  group convolution layer* applies a *shift-twist convolution* via a set of rotated-and-shifted kernels in  $SE(2)$  to output a  $SE(2)$ -image activation map (red border highlights the kernel transformation, cyan border highlights the output of a  $SE(2)$  kernel). The *projection layer* transforms an input  $SE(2)$ -image onto  $\mathbb{R}^2$  via a rotation-invariant operation (pixel-wise maximum projection is used here). A 3-channel input is shown for the  $SE(2)$  group convolution layer and 1-channel outputs are shown for all the layers: this is done for illustrative purposes but more channels are used in practice. The example images used for the examples are extracted from a trained nuclei segmentation model with a 8-fold discretization of  $SE(2)$ . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$(\mathbb{L}_2(SE(2)))^{N_l}$  with  $N_l$  channels via lifting correlations with a collection of  $N_l$  kernels, denoted with  $\mathbf{k}^{(l)} := (k_1^{(l)}, \dots, k_{N_l}^{(l)})$ , each kernel with  $N_{l-1}$  channels, via

$$\underline{F}^{(l)} = \tilde{\Phi}^{(l)}(\underline{f}^{(l-1)}) := \mathbf{k}^{(l)} \star \underline{f}^{(l-1)}, \quad (7)$$

where we overload the  $\star$  symbol defined in Eq. (6) to also denote the lifting correlation between a set of convolution kernels and a vector valued feature map via  $\mathbf{k}^{(l)} \star \underline{f}^{(l-1)} := \left( k_1^{(l)} \star \underline{f}^{(l-1)}, \dots, k_{N_l}^{(l)} \star \underline{f}^{(l-1)} \right)$ . Note that such operators are equivariant with respect to roto-translations when in (4) we let  $\mathcal{T}_g = \mathcal{U}_g$  and  $\mathcal{T}_g' = \mathcal{L}_g$  be the representations of  $SE(2)$  given respectively in (2) and (3), indeed  $\tilde{\Phi}^{(l)}(\mathcal{U}_g \underline{f}^{(l-1)}) = \mathcal{L}_g \tilde{\Phi}^{(l)}(\underline{f}^{(l-1)})$ .

The lifting layer thus generates higher-dimensional feature maps on the space of roto-translations. An  $SE(2)$  equivariant layer that takes such feature maps as input is then again obtained by taking inner products of the input feature map  $\underline{F}$  with (3D) roto-translated convolution kernels  $\underline{K}$ , where the kernels are transformed by application of the representation  $\mathcal{L}_g$  of  $SE(2)$  on  $\mathbb{L}_2(SE(2))$ . **Group correlations** are then defined as

$$\begin{aligned} (\underline{K} \star \underline{F})(g) &:= \sum_{c=1}^{N_c} (\mathcal{L}_g K_c, F_c)_{\mathbb{L}_2(SE(2))} \\ &= \sum_{c=1}^{N_c} \int_{SE(2)} K_c(g^{-1} \cdot g') F_c(g') dg'. \end{aligned} \quad (8)$$

Note here, that a rotation of an  $SE(2)$  convolution kernel is obtained via a shift-twist, a planar rotation and shift along the  $\theta$ -axis,

see Eq. (3) and Fig. 1. The convolution kernels  $\underline{K}$  are 3-dimensional and they assign weights to activations at positions and orientations relative to a central position and orientation (relative to  $g \in SE(2)$ ). A set of  $SE(2)$  kernels  $\mathbf{K}^{(l)} := (K_1^{(l)}, \dots, K_{N_l}^{(l)})$  then defines a **group convolution layer**, which we denote with  $\Phi^{(l)}$ , and which maps from  $SE(2)$  feature maps  $\underline{F}^{(l-1)}$  at layer  $l-1$ , with  $N_{l-1}$  channels, to  $SE(2)$ -feature maps  $\underline{F}^{(l)}$  at layer  $l$ , with  $N_l$  channels, via

$$\underline{F}^{(l)} = \Phi^{(l)}(\underline{F}^{(l-1)}) := \mathbf{K}^{(l)} \star \underline{F}^{(l-1)}, \quad (9)$$

where we overload the group correlation symbol  $\star$ , defined in (8), to also denote correlation between a set of convolution kernels and a vector valued feature map on  $SE(2)$  via  $\mathbf{K}^{(l)} \star \underline{F}^{(l-1)} := \left( K_1^{(l)} \star \underline{F}^{(l-1)}, \dots, K_{N_l}^{(l)} \star \underline{F}^{(l-1)} \right)$ .

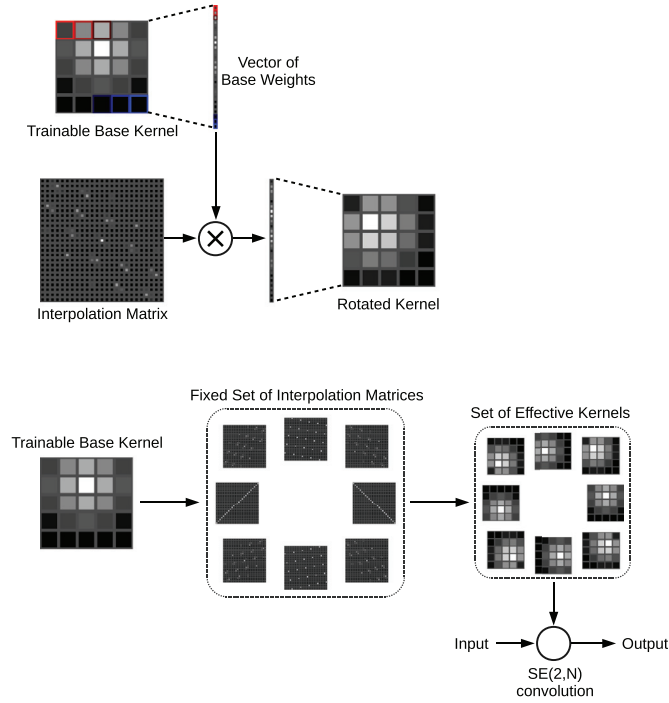
Finally, we define the **projection layer** as the operator that projects a multi-channel  $SE(2)$  feature map back to  $\mathbb{R}^2$  via

$$\underline{f}^{(l)}(\mathbf{x}) = \mathcal{P}(\underline{F}^{(l)})(\mathbf{x}) := \text{mean}_{\theta \in [0, 2\pi)} \underline{F}^{(l)}(\mathbf{x}, \theta). \quad (10)$$

Here we define the projection layer as taking the mean over the orientation axis, however, we note that any permutation invariant operator (on the  $\theta$ -axis) could be used to ensure local rotation invariance, such as e.g. the commonly used max operator (Cohen and Welling, 2016; Bekkers et al., 2018a).

### 3.4. Discretized $SE(2, N)$ group convolutional network

Discretized 2D images are supported on a bounded subset of  $\mathbb{Z}^2 \subset \mathbb{R}^2$  and the kernels live on a spatially rectangular grid of size  $n \times n$  in  $\mathbb{Z}^2$ , with  $n$  the kernel size. We discretize the group



**Fig. 2.** Illustration of the process generating a rotated set of effective kernels from a trainable vector of base weights via the introduction of fixed interpolation matrix in the computational pipeline.

$SE(2, N) := \mathbb{R}^2 \rtimes SO(2, N)$ , with the space of 2D rotations in  $SO(2)$  sampled with  $N$  rotation angles  $\theta_i = \frac{2\pi}{N}i$ , with  $i = 0, \dots, N-1$ .

The discrete lifting kernels  $\mathbf{k}^{(l)}$  at layer  $l$ , are used to map a 2D input image with  $N_{l-1}$  channels to an  $SE(2, N)$ -image with  $N_l$  channels, and thus have a shape of  $n \times n \times N_{l-1} \times N_l$  (the discretization of  $\mathbf{k}^{(l)}$  is illustrated in Fig. 1 as a set of  $n$  rotated  $\mathbb{R}^2$  kernels, distributed on a circle). Likewise, the  $SE(2, N)$  kernels  $\mathbf{K}^{(l)}$  have a shape of  $n \times n \times N \times N_{l-1} \times N_l$ .

The lifting and group convolution layers require rotating the spatial part of the kernels and shift along the  $\theta$ -axis for the  $SE(2)$ -kernels. We obtain the rotated spatial parts of each kernel via bi-linear interpolation. The discretization of a single lifting kernel  $k_{i,j}^{(l)}$  and its  $N$  rotated versions is illustrated in the top-left part of Fig. 1. The discretization of a single group correlation kernel  $K_{i,j}^{(l)}$  and its  $N$  rotated and  $\theta$ -shifted versions is illustrated in the bottom part of Fig. 1.

In order to construct the rotated sets of effective kernels  $\mathbf{k}^{(l)}$  or  $\mathbf{K}^{(l)}$  we rely on bi-linear interpolation. We first define a set of trainable vectors containing base weights that are used to generate rotated versions of the same base 2D kernel via bi-linear interpolation. We implemented this rotation process in the computational pipeline via the definition of non-trainable interpolation matrices, each coding for a rotation step, and the introduction of respective matrix multiplication operations. This process is illustrated in Fig. 2.

Although these sets of rotated kernels are used in the computational pipeline, only the base weights are updated during the network optimization. By construction, the effective kernels are differentiable with respect to their base weight, enabling their update in back-propagation of gradients (since the matrix multiplication operation is differentiable).

## 4. Experiments

In this section, we present the G-CNN architectures that we build using the layers defined in Section 3.3 and we describe the

experiments that we used to analyze and validate them. In the construction of the G-CNNs we adhere to the following principle of group equivariant architecture design.

**G-CNN design principle** A sequence of layers starting with a lifting layer (Eq. (7)) and followed by one or more group convolution layers (Eq. (9)), possibly intertwined with point-wise nonlinearities, results in the encoding of roto-translation equivariant feature maps. If such a block is followed by a projection layer (Eq. (10)) then the entire block results in a encoding of features that is guaranteed to be rotationally invariant. Our implementation of the G-CNN layers is available at <https://github.com/tueimage/se2cnn>.

### 4.1. Applications and model architectures

For each task introduced in Section 3.1 we conducted two experiments: first, we trained a set of variations of a baseline CNN, by changing the orientation sampling level  $N$  of their  $SE(2, N)$  layers, while keeping the total number of weights of each model approximately the same. Second, we trained each model with the reduced data regime counterparts of the training sets introduced in Section 3.1. For each task we opted for versions of straight-forward architectures with a low number of parameters that were in-line with methods reported in the literature. This way, we propose new G-CNN baselines that facilitate comparative experiments and that can be extended to more sophisticated architectures.

**Mitosis detection** We used the mitosis classification model originally described in Bekkers et al. (2018a) as a baseline: a 6-layer CNN with three down-sampling steps, such that the overall receptive field is of size  $68 \times 68$ .

We designed the G-CNN variants of this baseline described in Table 1, by replacing the first convolution layer by a lifting layer, replacing the following convolution layers by group convolution layers and inserting a projection layer before the last fully connected layer.

The models were trained with batches of size 64 balanced across classes. Non-mitosis class patches were sampled based on a hard negative mining procedure (Cireřan et al., 2013) using a first baseline model trained with random negative patches. The models were trained to minimize the cross-entropy of the binary-class predictions.

**Nuclei segmentation** For the nuclei segmentation task, we opted for a 7-layer U-net that corresponds to two spatial down/up-sampling operations with an overall receptive field of size  $44 \times 44$ . The sequence of operations defining this G-CNN architecture is given in the first column of Table 2.

The label associated with each input image is a 3-class mask corresponding to the foreground, background and border of the nuclei it contains (these masks can then be used to retrieve an individual nucleus using a segmentation procedure such as described in Section 5).

The models were trained with batches of size 16 balanced across patients, to minimize the class-weighted cross-entropy of the softmax activated output maps corresponding to the three target masks.

**Tumor detection** The baseline architecture we used for the tumor detection model is a 6-layer CNN with three down-sampling steps, such that the overall receptive field is of size  $88 \times 88$  (see Table 3 for the detailed architecture).

The models were trained with batches of size 64 balanced across classes. We refined both classes by running a hard negative mining procedure (Cireřan et al., 2013) using a first baseline model trained with the original dataset of the benchmark. The models were trained to minimize the cross-entropy of the binary-class predictions.

**Table 1**

Architecture of the investigated G-CNN models for mitosis detection. The left-most column indicates the operations applied in each layer. *Max. Proj.* indicates the projection operation on  $\mathbb{R}^2$ , achieved via maximum intensity projection along the orientations.

	<i>SE(2,N)</i> Groups Layers			
	N=1 ( $\mathbb{R}^2$ )	N=4 (p4)	N=8	N=16
Input	68×68×3			
Lifting Layer BN + ReLU MaxPool(2×2)	1×42×42×16(1040)	4×42×42×10(650)	8×42×42×8(520)	16×42×42×6(390)
Group Conv. BN + ReLU MaxPool(2×2)	1×14×14×16(5408)	4×14×14×10(8420)	8×14×14×8(10768)	16×14×14×6(12108)
Group Conv. BN + ReLU MaxPool(2×2)	1×5×5×16(5408)	4×5×5×10(8420)	8×5×5×8(10768)	16×5×5×6(12108)
Group Conv. BN + ReLU	1×1×1×64(21632)	4×1×1×16(13472)	8×1×1×8(10768)	16×1×1×4(8072)
Group Conv. BN + ReLU	1×1×1×16(1056)	4×1×1×16(1056)	8×1×1×16(1056)	16×1×1×16(1056)
Max. Proj.	1×1×16			
FC Layer Sigmoid	1×1×1(17)			
Total Weights	34561	32035	33897	33751

**Table 2**

Architecture and weight counting of the G-CNN models for patch-based tumor detection. The left-most column indicates the operations in each layer. *Concat(HL,x)* indicates the characteristic skip operation of the U-net architecture that consist in concatenating a centered crop of the output activation of the *x*th layer of the network. *Max. Proj.* indicates the projection operation on  $\mathbb{R}^2$ , achieved via maximum intensity projection along the orientations.

	<i>SE(2,N)</i> Groups Layers			
	N=1 ( $\mathbb{R}^2$ )	N=4 (p4)	N=8	N=16
Input	60×60×3			
Lifting Layer BN + ReLU MaxPool(2×2)	1×28×28×16(1040)	4×28×28×10(650)	8×28×28×8(520)	16×28×28×6(390)
Group Conv. BN + ReLU MaxPool(2×2)	1×12×12×16(5408)	4×12×12×10(8420)	8×12×12×8(10768)	16×12×12×6(12108)
Group Conv. BN + ReLU	1×8×8×16(5408)	4×8×8×10(8420)	8×8×8×8(10768)	16×8×8×6(12108)
Up-sampling Concat(HL,2) Group Conv. BN + ReLU	1×12×12×16(10784)	4×12×12×10(16820)	8×12×12×8(21520)	16×12×12×6(24204)
Up-sampling Concat(HL,1) Group Conv. BN + ReLU	1×20×20×64(43136)	4×20×20×16(26912)	8×20×20×8(21520)	16×20×20×4(16136)
Group Conv. BN + ReLU	1×20×20×16(1056)	4×20×20×16(1056)	8×20×20×16(1056)	16×20×20×16(1056)
Max. Proj.	20×20×16			
FC Layer Softmax	20×20×3(54)			
Total Weights	66886	62332	66206	66056

**Table 3**

Architecture and weight counting of the G-CNN models for patch-based tumor detection. The left-most column indicates the operations in each layer. *Mean. Proj.* indicates the projection operation on  $\mathbb{R}^2$ , achieved via mean intensity projection along the orientations.

Layers	<i>SE(2,N)</i> Groups			
	N=1 ( $\mathbb{R}^2$ )	N=4 (p4)	N=8	N=16
Input	88×88×3			
Lifting Layer BN + ReLU MaxPool(2×2)	1×42×42×32(2080)	4×42×42×19(1235)	8×42×42×14(910)	16×42×42×10(650)
Group Conv. BN + ReLU MaxPool(2×2)	1×19×19×32(21568)	4×19×19×19(30362)	8×19×19×14(32956)	16×19×19×10(33620)
Group Conv. BN + ReLU MaxPool(3×3)	1×5×5×32(21568)	4×5×5×19(30362)	8×5×5×14(32956)	16×5×5×10(33620)
Group Conv. BN + ReLU	1×1×1×64(43136)	4×1×1×16(25568)	8×1×1×8(18832)	16×1×1×4(13448)
Group Conv. BN + ReLU	1×1×1×16(1056)	4×1×1×16(1056)	8×1×1×16(1056)	16×1×1×16(1056)
Mean Proj.	1×1×16			
FC Layer Sigmoid	1×1×1(17)			
Total Weights	89425	88600	86727	82411

## 4.2. Implementation details

For all three baseline architectures, convolution kernels are of size  $5 \times 5$  with circular masking and fully connected layers are implemented as convolutional layers with kernels of shape  $1 \times 1$  to enable dense application (the resulting models can efficiently be applied on larger input sizes).

Batch Normalization (Ioffe and Szegedy, 2015) is used throughout the networks. Batch statistics are normally computed across batch and spatial dimensions of the activations, but we also included the orientation-axis of the *SE(2,N)*-image activation maps in the statistic computation to ensure their invariance with respect to the orientation of the input.

All models were trained with Stochastic Gradient Descent with momentum (learning rate 0.01, momentum 0.9) and a epoch-wise learning rate decay using a factor of 0.5 was applied. Training was stopped after convergence of the loss computed on the validation sets. All models were regularized with decoupled weight decay (coefficient  $5 \times 10^{-4}$ ). Baseline augmentation transformations

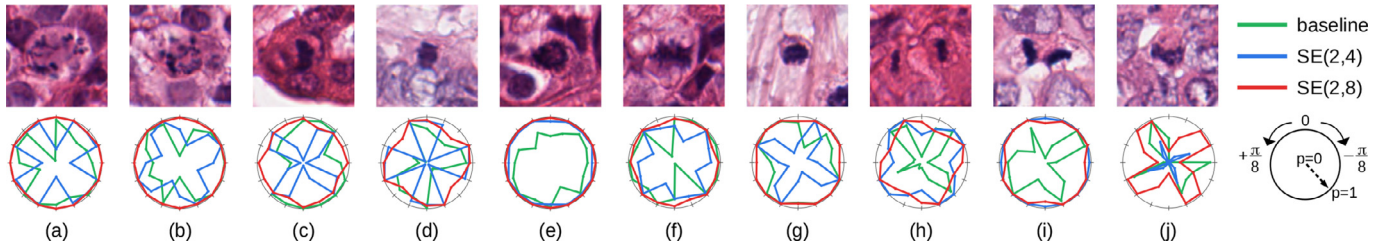
were applied to the training image patches (random spatial transposition, random 90-degree-wise rotation, random channel-wise brightness shifting).

## 4.3. Experiment: orientation sampling

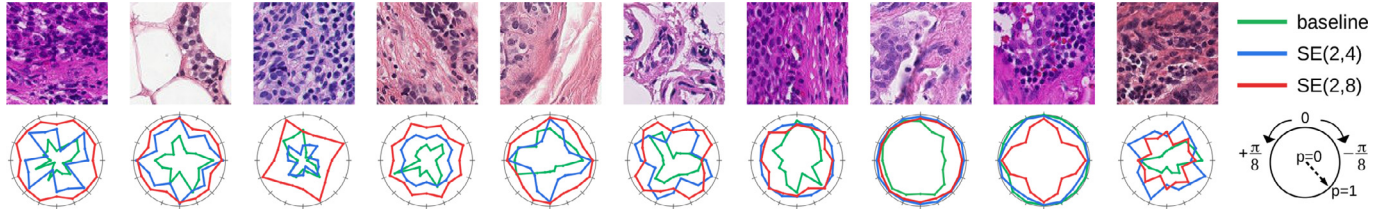
In order to assess the effect of using the proposed *SE(2,N)* G-CNN structure on the benchmark performances, we trained every model with  $N \in \{1, 4, 8, 16\}$ . In order to allow fair comparison we adjusted the number of channels in every layer involving *SE(2,N)*-image representation such that the total number of weights in the models stay close to the count of the corresponding baselines. The detailed distributions of the weights are shown in Tables 1–3: for each *SE(2,N)* group, the dimensions of the output of the layers are shown with the format  $N \times \text{Height} \times \text{Width} \times C$ , with *C* the number of output channels in the layer.

Each model was trained three times with random initialization seeds. We report the mean and standard deviation of the performances across three random initializations.





**Fig. 3.** Example of mitosis-centered image patches selected from the test set. Below each, polar plots show model predictions (distance from origin) as a function of the orientation of the input (angle coordinate) using steps of  $\pi/8$  rad. An ideal model would then produce a circle with maximum radius. Selected models are indicated with colors, and correspond to the best obtained models that were trained without reduced data regime over repeats (based on their  $F_1$ -score).



**Fig. 4.** Example of image patches selected from the test set of the PCam benchmark, for which pixels in the center area were classified as tumor tissue. Below each, polar plots show model predictions (distance from origin) as a function of the orientation of the input (angle coordinate) using steps of  $\pi/8$  rad. Selected models are indicated with colors, and correspond to the best obtained models that were trained without reduced data regime over repeats (based on their accuracy).

#### 4.4. Experiment: data regime experiments

In order to assess the effect of using the proposed  $SE(2,N)$  with varying sampling factor  $N$  when data is availability is reduced, we trained each model on the data-regime subsets presented in Section 3.1. Likewise, each model was trained three times with random initialization seeds so as to report the variability of the performances.

### 5. Results

This section summarizes the qualitative and quantitative results of the experiments we conducted. Each trained model was evaluated on the test set of its corresponding benchmark dataset based on standard performance metrics.

**Mitosis detection** For the mitosis detection task, models were densely applied on test images, followed by a smoothing operation before extracting all local maxima to be considered candidate detections. We computed the  $F_1$ -score of the set of detections using an operating point that is optimized on the validation set, as described in the scoring protocol used in (Veta et al., 2015).

**Nuclei segmentation** To quantify the performances of the nuclei segmentation model, generation of segmented candidate objects is obtained by following the protocol used in (Kumar et al., 2017; Lafarge et al., 2019). First, marker seeds are derived from thresholded foreground and background predictions, border predictions are used as the watershed energy landscape. Then, candidate objects that overlap the nuclei ground-truth masks by at least 50% of their area are considered hits, enabling object-level detection quantification to be calculated using the  $F_1$ -score. Thresholds to generate marker seeds were selected such that the  $F_1$ -score is maximized on the validation set.

**Patch-based tumor detection** To evaluate the tumor detection model, we computed the class probability of every patch of the test dataset and calculated the accuracy of the model given the ground-truth labels as in Veeling et al. (2018) after selection of the operating point that maximizes the accuracy on the validation set.

#### 5.1. Qualitative results

We qualitatively investigated the robustness of the prediction of different models to controlled rotations of the input. We see that

the model predictions can be very inconsistent for our best baseline model, in comparison to G-CNN models (see Figs. 3–5) in particular for cell or tissue morphologies that are typically asymmetric. For example, the mitotic figures (h) and (i) shown in Fig. 3 are in telophase (directed separation of the pair of chromosomes) and the variance of the prediction of the baseline model is higher for these cases (green curve) compared to the G-CNN models (blue and red curves). We also observe that for the  $SE(2,4)$  model, predictions that are obtained for an input image rotated with an angle below  $\pi/2$  rad also produce some variance, but present a  $\pi/2$  rad-period cyclic pattern.

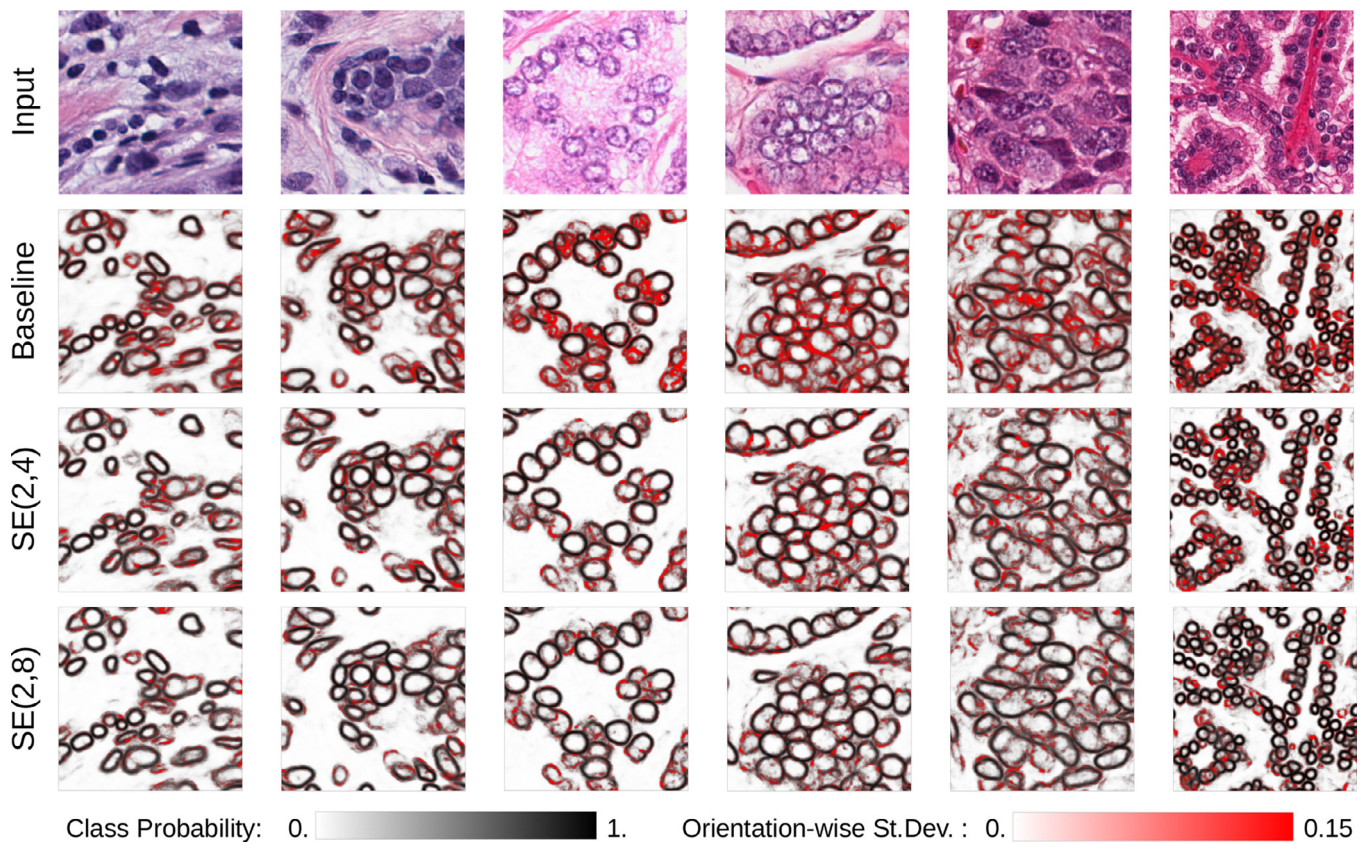
#### 5.2. Quantitative results

The performances of the trained models for both orientation sampling experiments and data regime experiments are summarized in the box plots of Figs. 6–8.

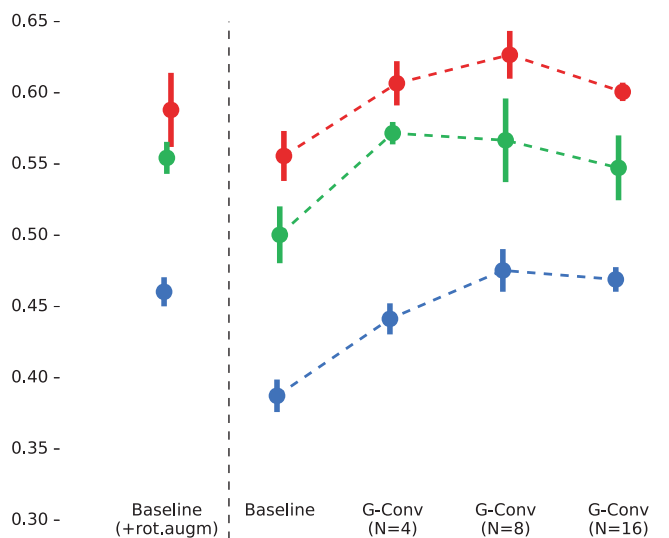
**Notes on absolute performances** For the mitosis detection benchmark, the best result we obtained is in line with the results previously reported in Lafarge et al. (2019) (best  $F_1$ -score of  $0.62 \pm 0.008$ ). For the PCam benchmark, the best result we obtained is in line with the results previously reported in Veeling et al. (2018) (best accuracy of 0.898). For the nuclei segmentation task, we note that the performances we achieved are significantly lower than the performances previously reported in the literature (on the same test set, Lafarge et al. (2019) reported a  $F_1$ -score of  $0.821 \pm 0.004$ ). We explain these difference by the strict constraints we imposed in the design of the baseline segmentation model of this study (lower receptive field, shallower network, lower weight capacity).

**Effect of orientation sampling** For all three studied tasks, we observed an increase of performance with the number of sampled orientations from  $N = 1$  to  $N = 8$ . For the full data regime of the mitosis detection experiments, the use of a  $SE(2,8)$ -G-CNN improves the  $F_1$ -score to  $0.626 \pm 0.015$  on average compared to  $0.556 \pm 0.016$  for the baseline model without test-time rotation augmentation (see Fig. 6). A similar increase of performances is observed for the nuclei segmentation experiments with an improvement of the  $F_1$ -score from  $0.754 \pm 0.006$  to  $0.771 \pm 0.06$  (see Fig. 7), and for the tumor detection experiments with an improvement of the accuracy from  $0.863 \pm 0.003$  to  $0.892 \pm 0.004$  (see Fig. 8).

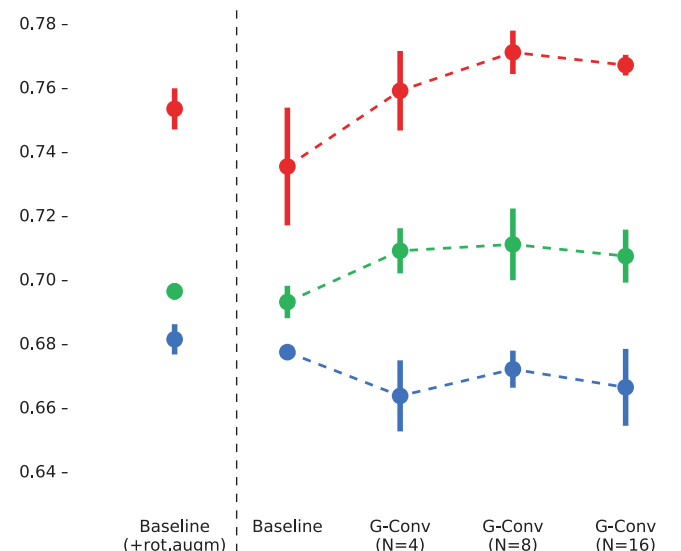




**Fig. 5.** Example of image patches selected from the test set of the nuclei segmentation benchmark (column 1-2: breast tissue, column 3-4: prostate tissue, column 5: kidney tissue, column 6: liver). For each image, and a selection of models, the raw predictions of the nucleus boundary class were computed and stored for the set of rotated inputs using steps of  $\pi/8$  rad. Predictions were re-aligned and their means were mapped to gray-scale and the standard deviations of the predictions were mapped to a white-to-red color scale. The overlap of these statistics is shown below each original image. Selected models are the best obtained models that were trained without reduced data regime over repeats (based on their  $F_1$ -score). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



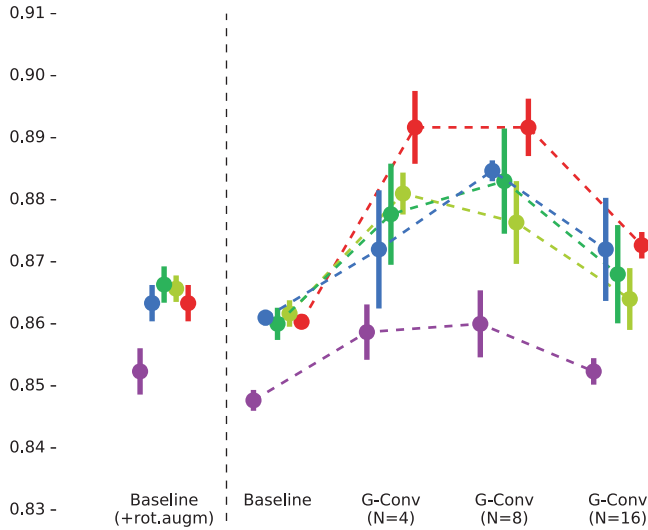
**Fig. 6.** Mean and Standard Deviation plots summarizing the  $F_1$ -score of the mitosis detection models. Mean  $\pm$  standard deviation is indicated. Color identifies the different data regime (red: 8 cases; green: 4 cases; blue: 2 cases). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 7.** Mean and Standard Deviation plots summarizing the  $F_1$ -score of the nuclei segmentation models. Mean  $\pm$  standard deviation is indicated. Color identifies the different data regime (red: 6 HPFs/organ; green: 4 HPFs/organ; blue: 2 HPFs/organ). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

We remark that the performances of the  $SE(2,4)$  G-CNN models are better than the baseline with test-time rotation augmentation as was previously reported in literature for similar tasks (Bekkers

et al., 2018a; Veeling et al., 2018). We also report that for all three tasks,  $SE(2,16)$  G-CNN models perform worse than the  $SE(2,8)$  G-CNN models.



**Fig. 8.** Mean and Standard Deviation plots summarizing the accuracy of the tumor detection models. Mean  $\pm$  standard deviation is indicated. Color identifies the different data regime (red: 100%; lime: 75%; green: 50%; blue: 25%; purple: 10%). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

**Effect of reduced data regime with orientation sampling** For all three tasks, we see a global consistent decrease of performances when less training data is available. In Fig. 8, the performances of the  $SE(2,4)$  and  $SE(2,8)$  G-CNN models trained with the 25%, 50% and 75% data regimes, are higher than for the baseline model at full data regime using test-time rotation augmentation. This reveals that under experimental conditions, data availability is not the only reason for limited performances since this experiment shows that the  $SE(2,N)$  G-CNN models enable achieving higher performances than the baseline models, even if less data is available.

## 6. Discussion and conclusions

The presented study investigated the effects of embedding the  $SE(2)$  group structure in CNNs, in the context of histopathology image analysis, across multiple controlled experimental setups.

The comparative analysis we conducted shows a consistent increase of performances for three different histopathology image analysis tasks when using the proposed  $SE(2,N)$  G-CNN architecture compared to conventional CNNs acting in  $\mathbb{R}^2$  evaluated with test-time rotation augmentation. This is in line with previously reported results when using G-CNNs with groups that lay on the pixel grid (p4, p4m) (Cohen and Welling, 2016; Veeling et al., 2018), but we also show that these performances can be surpassed when using groups with higher discretization levels of  $SE(2)$ .

This confirms that conventional  $\mathbb{R}^2$  CNNs struggle to learn a rotation equivariant representation based on data solely and that enforcing equivariant representation learning enables reaching higher performances. G-CNNs with  $SE(2,N)$  structure have the advantage to guarantee higher robustness to input orientation without requiring training-time or test-time rotation augmentation. Furthermore, the slight computational overhead for computing rotated convolutional operators and their gradient, at training time, can be canceled at test-time by computing and fixing all final oriented  $SE(2,N)$  kernels, resulting in a model that is computationally equivalent to conventional  $\mathbb{R}^2$  CNNs.

We show that these performances can be surpassed when using representations with higher angular resolution levels, as shown with experiments involving  $SE(2,8)$  G-CNNs and when the training data is of sufficient amount. This conclusion corroborates

the results we reported on other medical image analysis tasks (Bekkers et al., 2018a) and in studies that investigated models with rotated operators that lay outside of the pixel grid (Hoogeboom et al., 2018).

However, we also identified consistent lower performances for  $SE(2,16)$  G-CNNs compared to  $SE(2,8)$  G-CNNs at full data regime. We assume that this phenomenon is in part related to the model architectures we chose to enforce fixed model capacity, resulting in a number of channels in the representation of the  $SE(2,N)$  models being reduced when  $N$  increases. This reduced number of channels might affect the diversity of the features learned by the models, to the point that this limits their overall performances. Therefore, it appears there is a trade-off between performances and angular resolution at fixed capacity, further work would be necessary to confirm this hypothesis.

For the tumor detection task, we observed that the performances of the baseline models (with or without test-time rotation augmentation) reached a plateau, whatever the regime of available training data was among 25%, 50%, 75% or 100%. This indicates that in the conditions of the PCam dataset, the amount of available training data does not significantly influence the performances. However, the rotation-equivariant models were able to achieve better performances with increased data regime.

This behavior was not evidenced for the mitosis detection and nuclei segmentation experiments. We assume this result may be task-dependent or might be due to the fact that the plateau of performances observed for the tumor detection models was not reached yet for the two other tasks.

We qualitatively showed that in some cases, the predictions of conventional CNNs are inconsistent when inputs are rotated, whereas  $SE(2)$  G-CNNs show better stability in that sense. This suggests that the anisotropic learned features of conventional models only get activated when the input is observed in a specific orientation. On the shown examples (Section 5.1), the  $SE(2)$  models are more robust to the input orientation since their  $SE(2)$  structure guarantees the features to be expressed in multiple orientations. We also see that  $SE(2)$  models with a limited angular resolution can yet produce some variance for rotation angles lower than this resolution. This is also supported by the fact that higher performances were obtained for the experiments that compare  $SE(2,4)$  models to  $SE(2,8)$  models.

Still, variation of performances for these models was also observed when the input was rotated out of the pixel grid. We explain this limit from the approximation errors caused by two of the operators we used, and that have a weaker rotation equivariance property. First, the interpolation-based computation of the rotated kernels can cause small variations in the output when the input is rotated. Second, the pooling operators are not rotation equivariant by construction (since they lay on fixed down-sampled versions of the pixel grid), and so are another source of error.

In conclusion, we proposed a framework for  $SE(2)$  group-convolutional network and showed its advantages for histopathology image analysis tasks. This framework enables the learned models to be invariant to the natural roto-translational symmetry of histology images. We showed that G-CNNs models whose representation have a  $SE(2)$  structure yield better performances than conventional CNNs and our experiments suggest the ability of G-CNNs models to fully exploit the data amount of large datasets. Our results suggest the existence of a trade-off between network capacity and the chosen angular resolution of the  $SE(2,N)$  operators. We chose to experiment with light-weighted shallow model architectures in order to clearly show benefits of  $SE(2,N)$  equivariance: such light-weighted shallow model architectures allow for fair and transparent comparisons (where we control and fix the overall network capacity, see Tables 1–3). The proposed framework can also be applied to more heavy-weighted and deeper models



via the replacement of conventional  $\mathbb{R}^2$  convolutions by  $SE(2,N)$  convolutions, but this is beyond the scope of this article and is left for future work. Likewise, the use of more sophisticated data augmentation strategies that do not involve rotating the images can still be beneficial in practice. Other directions for future work include further analysis of the relationship between the newly introduced architecture-related hyper-parameters and their effect on model performances, as well as studying other prior structures that can improve model stability to other families of input transformations.

### Credit Author Statement

ML, EB, MV and JP contributed to the conception and design of the study. ML, EB and RD contributed to the mathematical underpinning of the framework. ML and EB implemented the roto-translation equivariant convolutional layers. ML implemented the experiments and wrote the first draft of the manuscript. All authors contributed to the result analysis, manuscript revision, read and approved the submitted version.

### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.media.2020.101849](https://doi.org/10.1016/j.media.2020.101849)

### CRedit authorship contribution statement

**Maxime W. Lafarge:** Conceptualization, Methodology, Software, Visualization, Writing - original draft, Writing - review & editing. **Erik J. Bekkers:** Conceptualization, Methodology, Software, Writing - review & editing. **Josien P.W. Pluim:** Conceptualization, Writing - review & editing. **Remco Duits:** Methodology, Writing - review & editing. **Mitko Veta:** Conceptualization, Writing - review & editing.

### References

- Andrearczyk, V., Fageot, J., Oreiller, V., Montet, X., Depeursinge, A., 2019. Exploring local rotation invariance in 3d CNNs with steerable filters. In: *MIDL*, pp. 15–26.
- Arsigny, V., Commowick, O., Pennec, X., Ayache, N., 2006. A log-euclidean framework for statistics on diffeomorphisms. In: *MICCAI*, pp. 924–931.
- Ashburner, J., 2007. A fast diffeomorphic image registration algorithm. *Neuroimage* 38, 95–113.
- Bekkers, E. J., 2019. B-Spline CNNs on Lie groups. *arXiv preprint arXiv:1909.12057*.
- Bekkers, E.J., Duits, R., Loog, M., 2015. Training of templates for object recognition in invertible orientation scores: application to optic nerve head detection in retinal images. In: *Energy Minimization Methods in Computer Vision and Pattern Recognition*, vol. 8932, pp. 464–477.
- Bekkers, E.J., Lafarge, M.W., Veta, M., Eppenhof, K.A., Pluim, J.P., Duits, R., 2018. Roto-translation covariant convolutional networks for medical image analysis. In: *MICCAI*, pp. 440–448.
- Bekkers, E.J., Loog, M., ter Haar Romeny, B.M., Duits, R., 2018. Template matching via densities on the roto-translation group. *tPAMI* 40, 452–466.
- Chidester, B., Ton, T.-V., Tran, M.-T., Ma, J., Do, M.N., 2019. Enhanced rotation-equivariant U-net for nuclear segmentation. *CVPR Workshops*.
- Chidester, B., Zhou, T., Do, M.N., Ma, J., 2019. Rotation equivariant and invariant neural networks for microscopy image analysis. *Bioinformatics* 35, i530–i537.
- Cireşan, D.C., Giusti, A., et al., 2013. Mitosis detection in breast cancer histology images with deep neural networks. In: *MICCAI*, pp. 411–418.
- Cohen, T., Geiger, M., Kähler, J., Welling, M., 2018. Spherical CNNs. *ICLR*.
- Cohen, T., Geiger, M., Weiler, M., 2018b. A general theory of equivariant CNNs on homogeneous spaces. *arXiv preprint arXiv:1811.02017*.
- Cohen, T., Welling, M., 2016. Group equivariant convolutional networks. In: *ICML*, pp. 2990–2999.
- Cohen, T.S., Weiler, M., Kicanaoglu, B., Welling, M., 2019. Gauge equivariant convolutional networks and the icosahedral cnn. *ICML*.
- Dieleman, S., De Fauw, J., Kavukcuoglu, K., 2016. Exploiting cyclic symmetry in convolutional neural networks. *arXiv preprint arXiv:1602.02660*.
- Duits, R., 2005. Perceptual organization in image analysis. Eindhoven University of Technology, the Netherlands Ph.D. thesis.
- Duits, R., Felsberg, M., Granlund, G., ter Haar Romeny, B., 2007. Image analysis and reconstruction using a wavelet transform constructed from a reducible representation of the Euclidean motion group. *Int. J. Comput. Vis.* 72, 79–102.
- Duits, R., Franken, E., 2011. Left-invariant diffusions on the space of positions and orientations and their application to crossing-preserving smoothing of HARDI images. *Int. J. Comput. Vis.* 92, 231–264.
- Duits, R., Janssen, M.H.J., Hannink, J., Sanguinetti, G.R., 2016. Locally adaptive frames in the roto-translation group and their applications in medical imaging. *J. Math. Imaging Vis.* 56, 367–402.
- Ehteshami Bejnordi, B., Veta, M., Johannes van Diest, P., van Ginneken, B., Karssemeijer, N., Litjens, G., van der Laak, J.A.W.M., the CAMELYON16 Consortium, 2017. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *Jama* 318, 2199–2210.
- Esteva, C., Allen-Blanchette, C., Makadia, A., Daniilidis, K., 2018. Learning  $SO(3)$  equivariant representations with spherical CNNs. *ECCV*.
- Esteva, C., Allen-Blanchette, C., Zhou, X., Daniilidis, K., 2018. Polar transformer networks. *ICLR*.
- Franken, E., Duits, R., 2009. Crossing-preserving coherence-enhancing diffusion on invertible orientation scores. *Int. J. Comput. Vis.* 85, 253.
- Gens, R., Domingos, P.M., 2014. Deep symmetry networks. In: *NeurIPS*, pp. 2537–2545.
- Graham, S., Epstein, D., Rajpoot, N., 2019. Rota-net: rotation equivariant network for simultaneous gland and lumen segmentation in colon histology images. In: *ECCV*, pp. 109–116.
- Hannink, J., Duits, R., Bekkers, E.J., 2014. Crossing-preserving multi-scale vesselness. In: *MICCAI*, pp. 603–610.
- Hefny, M.S., Okada, T., Hori, M., Sato, Y., Ellis, R.E., 2015. A liver atlas using the special euclidean group. In: *MICCAI*, pp. 238–245.
- Henriques, J.F., Vedaldi, A., 2017. Warped convolutions: efficient invariance to spatial transformations. In: *ICML*, pp. 1461–1469.
- Hoogeboom, E., W. T. Peters, J., Cohen, T., Welling, M., 2018. Hexaconv. *ICLR*.
- Hou, B., Miolane, N., Khanal, B., Lee, M.C., Alansary, A., McDonagh, S., Hajnal, J.V., Rueckert, D., Glocker, B., Kainz, B., 2018. Computing CNN loss and gradients for pose estimation with Riemannian geometry. In: *MICCAI*, pp. 756–764.
- Ioffe, S., Szegedy, C., 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: *ICML*, pp. 448–456.
- Jaderberg, M., Simonyan, K., Zisserman, A., others, 2015. Spatial transformer networks. In: *NeurIPS*, pp. 2017–2025.
- Janssen, M.M.H.J., Janssen, A.J.E.M., Bekkers, E.J., Olivn Bescs, J., Duits, R., 2018. Design and processing of invertible orientation scores of 3d images. *J. Math. Imaging Vis.* 1–32.
- Kondor, R., Trivedi, S., 2018. On the generalization of equivariance and convolution in neural networks to the action of compact groups. In: *ICML*, pp. 2747–2755.
- Kumar, N., Verma, R., Sharma, S., Bhargava, S., Vahadane, A., Sethi, A., 2017. A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE Trans. Med. Imaging* 36, 1550–1560.
- Lafarge, M., Pluim, J., Eppenhof, K., Veta, M., 2019. Learning domain-invariant representations of histological images. *Front. Med.* 6, 162.
- Li, X., Yu, L., Fu, C.-W., Heng, P.-A., 2018. Deeply supervised rotation equivariant network for lesion segmentation in dermoscopy images. In: *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis*. Springer, pp. 235–243.
- Macenko, M., Niethammer, M., Marron, J., Borland, D., Woosley, J.T., Guan, X., Schmitt, C., Thomas, N.E., 2009. A method for normalizing histology slides for quantitative analysis. In: *ISBI*, pp. 1107–1110.
- Mallat, S., 2012. Group invariant scattering. *Commun. Pure Appl. Math.* 65, 1331–1398.
- Marcos, D., Volpi, M., Komodakis, N., Tuia, D., 2017. Rotation equivariant vector field networks. In: *CVPR*, pp. 5048–5057.
- Network, C.G.A., et al., 2012. Comprehensive molecular portraits of human breast tumours. *Nature* 490, 61–70.
- Pennec, X., Sommer, S., Fletcher, T., 2019. Riemannian Geometric Statistics in Medical Image Analysis. Elsevier.
- Portegies, J.M., Fick, R.H.J., Sanguinetti, G.R., Meesters, S.P., Girard, G., Duits, R., 2015. Improving fiber alignment in HARDI by combining contextual PDE flow with constrained spherical deconvolution. *PLoS one* 10.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: convolutional networks for biomedical image segmentation. In: *MICCAI*, pp. 234–241.
- Tai, K. S., Bailis, P., Valiant, G., 2019. Equivariant transformer networks. *arXiv preprint arXiv:1901.11399*.
- Thomas, N., Smidt, T., Kearnes, S., Yang, L., Li, L., Kohlhoff, K., Riley, P., 2018. Tensor field networks: rotation- and translation-equivariant neural networks for 3d point clouds. *arXiv preprint arXiv:1802.08219*.
- Veeling, B.S., Linmans, J., Winkens, J., Cohen, T., Welling, M., 2018. Rotation equivariant CNNs for digital pathology. In: *MICCAI*, pp. 210–218.
- Veta, M., van Diest, P., Willems, S., et al., 2015. Assessment of algorithms for mitosis detection in breast cancer histopathology images. *Med. Image Anal.* 20, 237–248.
- Weiler, M., Geiger, M., Welling, M., Boomsma, W., Cohen, T., 2018. 3D steerable CNNs: learning rotationally equivariant features in volumetric data. In: *NeurIPS*, pp. 10381–10392.
- Weiler, M., Hamprecht, F.A., Storath, M., 2017. Learning steerable filters for rotation equivariant CNNs. *CVPR*.



- Winkels, M., Cohen, T.S., 2018. 3D G-CNNs for pulmonary nodule detection. MIDL.
- Winkels, M., Cohen, T.S., 2019. Pulmonary nodule detection in CT scans with equivariant cnns. *Med. Image Anal.* 55, 15–26.
- Worrall, D., Brostow, G., 2018. CubeNet: equivariance to 3d rotation and translation. In: *ECCV*, pp. 567–584.
- Worrall, D.E., Garbin, S.J., Turmukhambetov, D., Brostow, G.J., 2017. Harmonic networks: deep translation and rotation equivariance. In: *CVPR*, pp. 5028–5037.
- Worrall, D. E., Welling, M., 2019. Deep scale-spaces: equivariance over scale. *arXiv preprint arXiv:1905.11697*.
- Zhang, J., Bekkers, E.J., Abbasi-Sureshjani, S., Dashtbozorg, B., ter Haar Romeny, B.M., 2015. Robust and fast vessel segmentation via gaussian derivatives in orientation scores. In: *ICIAP*, pp. 537–547.
- Zhou, Y., Ye, Q., Qiu, Q., Jiao, J., 2017. Oriented response networks. In: *CVPR*, pp. 4961–4970.