Exploring local rotation invariance in 3D CNNs with steerable filters

Vincent Andrearczyk[1]
Julien Fageot[2]
Valentin Oreiller[1,3]
Xavier Montet[4]
Adrien Depeursinge[1,3]
[1] *University of Applied Sciences Western Switzerland (HES-SO), Sierre, Switzerland*
[2] *Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland*
[3] *Centre Hospitalier Universitaire Vaudois (CHUV), Lausanne, Switzerland*
[4] *Hopitaux Universitaires de Genève (HUG), Geneva, Switzerland*

Abstract

Locally Rotation Invariant (LRI) image analysis was shown to be fundamental in many applications and in particular in medical imaging where local structures of tissues occur at arbitrary rotations. LRI constituted the cornerstone of several breakthroughs in texture analysis, including Local Binary Patterns (LBP), Maximum Response 8 (MR8) and steerable filterbanks. Whereas globally rotation invariant Convolutional Neural Networks (CNN) were recently proposed, LRI was very little investigated in the context of deep learning. We use trainable 3D steerable filters in CNNs in order to obtain LRI with directional sensitivity, i.e. non-isotropic filters. Pooling across orientation channels after the first convolution layer releases the constraint on finite rotation groups as assumed in several recent works. Steerable filters are used to achieve a fine and efficient sampling of 3D rotations. We only convolve the input volume with a set of Spherical Harmonics (SHs) modulated by trainable radial supports and directly steer the responses, resulting in a drastic reduction of trainable parameters and of convolution operations, as well as avoiding approximations due to interpolation of rotated kernels. The proposed method is evaluated and compared to standard CNNs on 3D texture datasets including synthetic volumes with rotated patterns and pulmonary nodule classification in CT. The results show the importance of LRI in CNNs and the need for a fine rotation sampling.

Keywords: Local rotation invariance, convolutional neural network, steerable filters, 3D texture

1. Introduction

Convolutional Neural Networks (CNNs) have been used in various studies to analyze textures. Orderless pooling of feature maps is used to discard the overall shape and layout information and, thus, describe repetitive and diffuse texture patterns (Andrearczyk and Whelan, 2016; Cimpoi et al., 2016; Zhang et al., 2016). By construction, CNN architectures provide translation equivariance, which is particularly adapted to image analysis. This paper focuses on adding local rotation invariance in the CNN architecture, which is known to be crucial for biomedical applications (Depeursinge and Fageot, 2018).

Globally rotation equivariant/invariant CNNs have recently been extensively studied using group theory in order to propagate rotation equivariance throughout the network. The 2D Group equivariant CNNs (G-CNN) introduced in (Cohen and Welling, 2016) uses rotated convolutional filters with right angle rotations of the $p4$ symmetry group. Invariance is obtained by pooling across orientation channels after the last convolution layer. The G-CNN was recently extended to 3D images in (Winkels and Cohen, 2018) showing a performance increase in the analysis of pulmonary nodule detection. 3D G-CNNs were shown to improve classification of 3D textures (Andrearczyk and Depeursinge, 2018), yet the results motivated the use of a finer rotation sampling than right angle rotations from the Octahedral O group to capture realistic arbitrary 3D orientations of directional patterns. It is important to remark that G-CNNs are adapted to equivariance with respect to *finite* subgroups of the rotation group. In 2D, an arbitrary sampling of rotations can be used (Bekkers et al., 2018; Zhou et al., 2017) in a group equivariant approach, while the number of 3D finite rotation groups is restrained. The 2D harmonic network (Worrall et al., 2016) and 2D steerable CNN (Weiler et al., 2017) present similarities with the method proposed in this paper although in the 2D domain and not particularly designed for texture analysis. Finally, the 3D steerable CNNs (Weiler et al., 2018) are very general architectures that implement the global equivariance to rotation on the network, and the convolutional layer considered in this paper is covered by their characterization. As detailed below, we differ from their works by making an angular max pooling after the first convolution layer, which exploits the steerability of the filters, and more importantly, focuses on local invariances.

In the above approaches, global rotation equivariance is maintained all along the layers (see Fig. 1, left), and invariance is obtained by using orientation pooling at the end of the network after spatial average pooling. Global rotation invariance is fundamental in various applications. However, some images are composed of well-defined structures with arbitrary orientations. For instance, 3D textures observed in Computed Tomography (CT) and in Magnetic Resonance Imaging (MRI) exhibit diverse tissue alterations, including necrosis, angiogenesis, fibrosis, or cell proliferation (Gatenby et al., 2013). These alterations induce imaging signatures such as blobs, intersecting surfaces and curves. These local low-level patterns are characterized by discriminative directional properties and have arbitrary 3D orientations, which requires combining directional sensitivity with LRI. How-

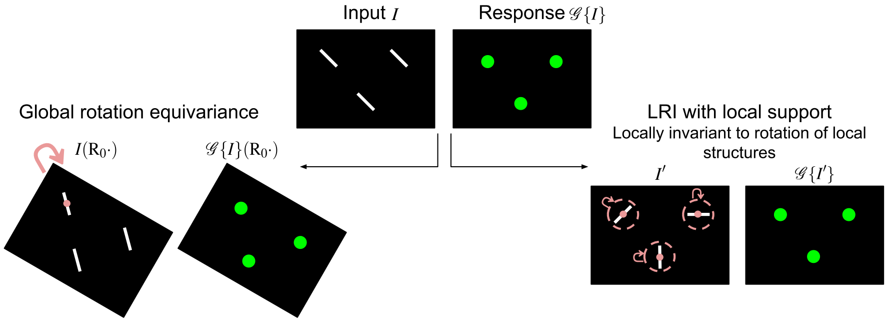

Input I
Response $\mathcal{G}\{I\}$
Global rotation equivariance
$I(\mathrm{R}_0\cdot)$
$\mathcal{G}\{I\}(\mathrm{R}_0\cdot)$
LRI with local support
Locally invariant to rotation of local structures
I'
$\mathcal{G}\{I'\}$

Figure 1: Illustration of global rotation equivariance and LRI in 2D. Rotating local structures (i.e. three white segments) in the input I results in the input I' on the right. The green dots illustrate the equivariant/invariant responses. Local and global rotations are shown in red and the local support G of the operator $\mathcal{G}$ (see Section 2.1) is represented as a dashed red line. It is worth noting that our CNN architecture will both present a global equivariance and a local invariance to rotations. Best viewed in color.

ever, rotation invariance is often antagonist with the will of being sensitive to directional features. The latter is required to avoid mixing blobs, edges and ridges. For instance, a spatial convolutional operator is equivariant to rotations if and only if the filter is isotropic, therefore insensitive to the directional features of the input signal. It follows that operators combining LRI and directional sensitivity require using more complex designs such as MR8 (Varma and Zisserman, 2005), local binary patterns (Ojala et al., 2002), 3D Riesz wavelets (Dicente Cid et al., 2017) and Spherical Harmonic (SH) invariants (Depeursinge et al., 2018) widely used in hand-crafted texture analysis (Depeursinge and Fageot, 2018).

In this paper, we exploit the steerability of SHs to obtain a CNN architecture which is both globally equivariant and locally invariant to rotations (see Fig. 1 for a 2D illustration). This is achieved with a fine rotation sampling and controlled operator support. The local support for the rotation invariance is set by the kernel size of the first layer. LRI is then obtained by pooling across orientations after this first layer. The implementation will be made publicly available.

2. Methods

We first introduce the framework in the continuous domain, hence voxel images, filters, and response maps are functions defined over the continuum $\mathbb{R}^3$. The discretization is then presented in Section 2.4. Spherical coordinates are defined as (ρ, θ, ϕ) with radius $\rho \geq 0$, elevation angle $\theta \in [0, \pi]$, and horizontal plane angle $\phi \in [0, 2\pi)$. The set of 3D rotations is denoted by $SO(3)$. A 3D rotation transformation R can be decomposed as three elementary rotations around the z, y' and z'' axes as $\mathrm{R} = \mathrm{R}_\alpha \mathrm{R}_\beta \mathrm{R}_\gamma$, with the (intrinsic) Euler angles $\alpha \in [0, 2\pi)$, $\beta \in [0, \pi]$, and $\gamma \in [0, 2\pi)$ respectively. We will use interchangeably R as a rotation transformation acting on $\mathbb{R}^3$ and on the two-dimensional sphere $\mathbb{S}^2$. Finally, the function $\boldsymbol{x} \mapsto f(\mathrm{R}\boldsymbol{x})$ is denoted by $f(\mathrm{R}\cdot)$.

2.1. Equivariant Local Texture Operators

We introduce the class of texture operators of interest that will be used in the first layer of our neural network. We consider a filter $f : \mathbb{R}^3 \to \mathbb{R}$, whose support G is assumed to be finite. For an image I and a position $\boldsymbol{x} \in \mathbb{R}^3$, we define the operator

$$\mathscr{G}\{I\}(\boldsymbol{x}) = \max_{\mathrm{R} \in SO(3)} |(I * f(\mathrm{R}\cdot))(\boldsymbol{x})|. \tag{1}$$

The operator combines a convolutional operator together with a max-pooling operation over the rotations R, and is an example of texture operator as presented in (Depeursinge and Fageot, 2018). Note that $(I * f(\mathrm{R}\cdot))(\boldsymbol{x})$ can be identified as a lifting group-convolution (Bekkers et al., 2018), followed by sub-group pooling (maximum over rotations). Then, $\mathscr{G}$ has the following properties:

- It is globally *equivariant to translations and rotations*, in the sense that, for any position $\boldsymbol{x}_0 \in \mathbb{R}^3$ and rotation $\mathrm{R}_0 \in SO(3)$,

  $$\mathscr{G}\{I(\cdot - \boldsymbol{x}_0)\} = \mathscr{G}\{I\}(\cdot - \boldsymbol{x}_0) \text{ and } \mathscr{G}\{I(\mathrm{R}_0\cdot)\} = \mathscr{G}\{I\}(\mathrm{R}_0\cdot). \tag{2}$$

  The proof is provided in Appendix A. In particular, if $\mathrm{R}_{\boldsymbol{x}_0}$ is a rotation around $\boldsymbol{x}_0 \in \mathbb{R}^3$, we have that $\mathscr{G}\{I(\mathrm{R}_{\boldsymbol{x}_0}\cdot)\} = \mathscr{G}\{I\}(\mathrm{R}_{\boldsymbol{x}_0}\cdot)$, as illustrated on the left part of Fig. 1.

- It is *local* in the sense that the filter f has a finite support $G = \{\boldsymbol{x} \in \mathbb{R}^3, \|\boldsymbol{x}\| \leq \rho_0\}$. As a consequence, $\mathscr{G}\{I\}(\boldsymbol{x})$ only depends on the values $I(\boldsymbol{y})$ for $\|\boldsymbol{y} - \boldsymbol{x}\| \leq \rho_0$.

The global equivariance to translations and rotations together with the locality create an invariance to local rotations (i.e. LRI) in the following sense: the rotation of an object or localized structure of interest in the image I around a position $\boldsymbol{x}$ does not affect the value of $\mathscr{G}\{I\}(\boldsymbol{x})$, as illustrated on the right part of Fig. 1.

2.2. Steerable Filters and Spherical Harmonics

Computing the texture operator (1) requires to maximize over any 3D rotation R for every position $\boldsymbol{x}$ of the image I, which can be computationally discouraging. To overcome this issue, we propose to use steerable filters, which have the advantage to allow for fast and efficient max pooling rotations (Chenouard and Unser, 2012; Fageot et al., 2018). A filter is steerable if any of its rotated version can be written as a linear combination of finitely many basis filters (Freeman and Adelson, 1991; Unser and Chenouard, 2013).

We consider filters f that are polar-separable, in the sense that they can be written as $f(\rho,\theta,\phi) = h(\rho)g(\theta,\phi)$ with $h:\mathbb{R}^+\to\mathbb{R}$ and $g:\mathbb{S}^2\to\mathbb{R}$. One can expand such steerable polar-separable filters in terms of the family of SHs $(Y_{n,m})_{n\geq 0,\ m\in\{-n\cdots n\}}$, where n is called the degree and m the order, and which form an orthonormal basis for square-integrable functions $g(\theta,\phi)$ on $\mathbb{S}^2$. We consider finitely many degrees, $N\geq 0$ being the maximal one. In particular, the number of elements of a SH family of maximum degree N is $\sum_{n=0}^{N}(2n+1) = (N+1)^2$. The definition of SHs can be found in Appendix B.

The general form of a polar-separable steerable filter with maximal degree $N\geq 0$ is

$$f(\rho,\theta,\phi) = h(\rho)g(\theta,\phi) = h(\rho)\sum_{n=0}^{N}\sum_{m=-n}^{n}\mathrm{C}_n[m]Y_{n,m}(\theta,\phi), \tag{3}$$

where $h(\rho)$ is the radial profile of f and the coefficients $\mathrm{C}_n[m]$ determine the angular profile $g(\theta,\phi)$. The condition of f being real is translated into the conditions that h itself is real and that the SH coefficients satisfy $\mathrm{C}_n[-m] = (-1)^m\overline{\mathrm{C}_n[m]}$ (see Appendix C).

For any $\mathrm{R}\in SO(3)$, the rotated version $Y_{n,m}(\mathrm{R}\cdot)$ of a SH can be expressed as

$$Y_{n,m}(\mathrm{R}\cdot) = \sum_{m'=-n}^{n}\mathrm{D}_{\mathrm{R},n}[m,m']Y_{n,m'}. \tag{4}$$

where the $\mathrm{D}_{\mathrm{R},n}\in\mathbb{C}^{(2n+1)\times(2n+1)}$ are the Wigner matrices (Varshalovich et al., 1988). Then, the steerable filter f can then be rotated efficiently with any $\mathrm{R}\in SO(3)$ to obtain a set of steered coefficients $\mathrm{C}_{\mathrm{R},n} = \mathrm{D}_{\mathrm{R},n}\mathrm{C}_n$ of $f(\mathrm{R}\cdot)$, with $\mathrm{C}_n = (\mathrm{C}_n[m])_{m\in\{-n,\dots,n\}}$. Then, the rotated filter $f(\mathrm{R}\cdot)$ is given by

$$f(\mathrm{R}\cdot)(\rho,\theta,\phi) = h(\rho)\sum_{n=0}^{N}\sum_{m=-n}^{n}\sum_{m'=-n}^{n}\mathrm{D}_{\mathrm{R},n}[m,m']\mathrm{C}_n[m']Y_{n,m}(\theta,\phi). \tag{5}$$

From (5), we see that any rotated version of f can be computed from the coefficients $(\mathrm{C}_n[m])_{0\leq n\leq N,-n\leq m\leq n}$.

2.3. 3D Steerable Convolution and Max Pooling

Exploiting (5), the convolutional operator $I*f(\mathrm{R}\cdot)$ in (1) is then computed as

$$I*f(\mathrm{R}\cdot) = \sum_{n=0}^{N}\sum_{m=-n}^{n}\left(\sum_{m'=-n}^{n}\mathrm{D}_{\mathrm{R},n}[m,m']\mathrm{C}_n[m']\right)(I*hY_{n,m}). \tag{6}$$

Therefore, one accesses the convolution with any rotated version of f by computing $\sum_n(2n+1) = (N+1)^2$ convolutions $(I * hY_{n,m})$, which we shall exploit for computing the response map $\mathscr{G}\{I\}$ of the texture operator (1). It is worth noting that the case $N = 0$ corresponds to filters f that are isotropic, i.e. $f(\mathrm{R}\cdot) = f$ for any $\mathrm{R} \in SO(3)$ (Depeursinge et al., 2018). As low degrees (e.g. $N = 1, 2$) are sufficient to construct small filters (see Section 2.4), the gain becomes substantial over a G-CNN approach for a fine sampling of orientations with a drastic reduction of the number of convolutions.

In practice, one has a set of steerable filters f_i of the form (3) with radial profiles h_i and coefficients $\mathrm{C}_{i,n}[m]$. The number of trainable parameters is reduced to the coefficients $\mathrm{C}_{i,n}[m]$, the radial profiles h_i and the biases (one scalar parameter per output channel i).

2.4. Discretization

The radial profiles h_i, and hence the filters f_i, have a compact spherical support $G = \{\boldsymbol{x} \in \mathbb{R}^3, \|\boldsymbol{x}\| \leq \rho_0\}$, where $\rho_0 > 0$ is fixed. For any i, we consider the voxelized version of the radial profile $h_i(\rho)$. The size of the support of the voxelized version is linked to the radius ρ_0 of the filter in the continuous domain and the level of voxelization. Due to the isotropic constraint, for a support of c^3 voxels, the number of trainable parameters for each h_i is then $\left\lceil \frac{(c-1)}{2} \times \sqrt{3} \right\rceil + 1$. The values of the filter $f_i(\rho, \theta, \phi)$ over the continuum is deduced from the discretization using linear interpolation (Fig. 2). Note that the corner effect (radial values for $\rho > \frac{(c-1)}{2}$) has a limited impact on the approximated rotation invariance and the freedom is maintained for these parameters to be set to zero during training.

The maximal frequency N cannot be taken arbitrarily large once the radial profiles are voxelized (Weiler et al., 2017). Indeed, the discretized filters f_i are defined over c^3 voxels, which imposes the upper bound $N \leq \rho_0 c/2$, where ρ_0 is the radius of the spherical support of h_i. This can be interpreted as the angular Nyquist frequency.

To sample the rotations, we uniformly sample points on the sphere using a triangulation method that iteratively splits octahedron faces to obtain the (α, β) Euler angles around z and y' respectively. We then sample the last angle γ around z'' uniformly between 0 and 2π. The Octahedral group, for instance, is obtained by sampling 6 points on the sphere (i.e. six (α, β) pairs) and four values of γ to obtain 24 right angle rotations. In this paper, we evaluate the following sets of rotations: single rotation, Klein's four rotations, octahedral 24 rotations and 72 rotations (18 points on the sphere and 4 values of γ). In the sequel, we denote by M the number of tested rotations.

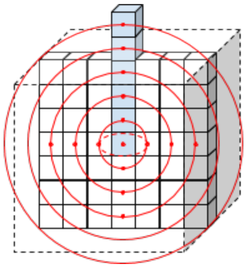
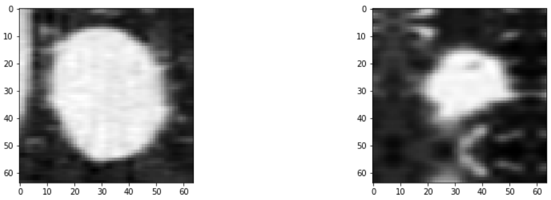


Figure 2: Illustration of a 2D slice of the isotropic radial profile h_i. The blue voxels represent the trainable parameters. The rest of the cube is linearly interpolated.

2.5. Datasets

We evaluate the proposed method on two experiments described in the following. In the first experiment, we built a dataset containing two classes of 500 synthetic volumes each. The volumes of size $32 \times 32 \times 32$ are generated by placing two $7 \times 7 \times 7$ patterns, namely a binary segment and a 2D cross with the same norm, at random 3D orientations and random locations with overlap. The number of patterns per volume is randomly set to $\lfloor d(\frac{s_v}{s_p})^3 \rfloor$, where s_v and s_p are the sizes of the volume and of the pattern respectively and the density d is in the range $[0.2, 0.4]$. The two classes vary by the proportion of the patterns, i.e. 10% segments with 90% crosses for the first class and vice versa for the second class. 800 volumes are used for training and the remaining 200 for testing. Despite the simplicity of this dataset, some variability is introduced by the overlapping patterns and the linear interpolation of the 3D rotations, making it challenging and more realistic.

The second dataset is a subsample of the American National Lung Screening Trial (NLST) that was annotated by radiologists at the University Hospitals of Geneva (HUG) (Martin et al., submitted). The dataset comprises 485 pulmonary nodules from distinct patients in CT, among which 244 were labeled benign and 241 malignant. We pad or crop the input volumes (originally ranging from $16 \times 16 \times 16$ to $128 \times 128 \times 128$) to the size $64 \times 64 \times 64$. We use the balanced training and test splits with 392 and 93 volumes respectively. Examples of 2D slices of the lung nodules are illustrated in Fig. 3. The Hounsfield units are clipped in the range $[-1000, 400]$, then normalized with zero mean and unit variance (using the training mean and variance).



(a) Benign nodule

(b) Malignant nodule

Figure 3: 2D slices from 3D volumes of benign and malignant pulmonary nodules.

2.6. Network Architecture

The first layer of the networks is the LRI layer (1). Global average spatial pooling is then used similarly to (Andrearczyk and Whelan, 2016). This pooling aggregates the locally invariant texture responses into a single scalar per feature map and is followed by fully connected layers. For the nodule experiment, we average the responses inside the nodule masks instead of across the entire feature maps. For the synthetic experiment, we connect directly the final softmax fully connected layer with a cross-entropy loss. For the second, more complex experiment, we use an intermediate fully connected layer with 128 neurons before the same final layer. Standard ReLU activations are employed. The networks are trained using Adam optimizer with $\beta_1 = 0.99$ and $\beta_2 = 0.9999$ and a batch size of 8. Other task-specific parameters are: for the synthetic experiment (kernel size $7 \times 7 \times 7$, stride 1, 2 filters and 50,000 iterations), for the nodule experiment (kernel size $9 \times 9 \times 9$, stride 2, 4 filters and 10,000 iterations).

We refer to the developed architecture as LRI-CNN and compare it to a network with the same architecture but with a standard 3D convolution layer, referred to as Z3-CNN.

2.7. Weights Initialization

The SHs are normalized to $\|Y_{n,m}\|_2 = 1$. The coefficients are then randomly initialized by a normal distribution with $Var[\mathrm{C}_{i,n}[m]] = \frac{2}{n_{in}(N+1)^2}$, where n_{in} is the number of input channels (generally 1), the radial profiles are initialized to $Var[h_i(\rho)] = 1$ and the biases to zero. This initialization is inspired by (He et al., 2015; Weiler et al., 2017) in order to avoid vanishing and exploding activations and gradients.

3. Experimental Results

The results for the synthetic experiment (3D textures of synthetic rotated patterns) are summarized in Table 1. Fig. 4 shows a comparison of standard 3D kernels (Z3-CNN) and SH parametric representations (LRI-CNN).

Table 1: Average accuracy (%) on the synthetic 3D local rotation dataset with $N = 2$.

model	**# orient. (M)**	**# filters**	**# param.**	**accuracy**$_{\pm\sigma}$
Z3-CNN	-	2	694	81.7$_{\pm 4.4}$
Z3-CNN	-	144	49,826	96.0$_{\pm 0.3}$
LRI-CNN	1	2	40	74.6$_{\pm 3.2}$
LRI-CNN	4	2	40	85.4$_{\pm 4.7}$
LRI-CNN	24	2	40	88.2$_{\pm 2.9}$
LRI-CNN	72	2	40	90.0$_{\pm 1.3}$

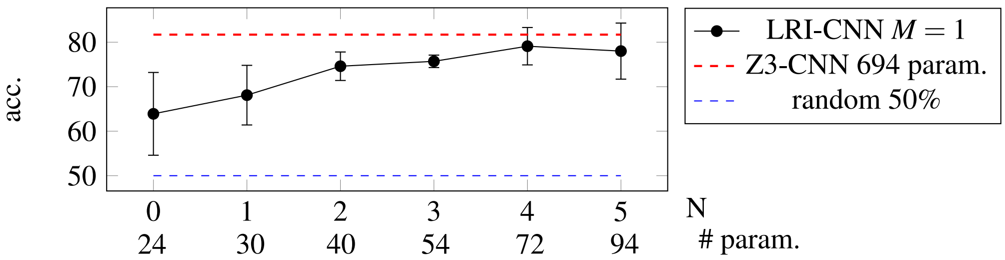


Figure 4: Comparison of standard 3D kernels (Z3-CNN) and SH parametric representation (LRI-CNN) with varying maximum degree N with a single orientation $M = 1$ (i.e. not using the steering capacity) on the synthetic 3D local rotation dataset. When $N \geq 2$, the performance of the SH parametric representation is very close to the Z3-CNN while using 15× fewer parameters.

Results for the nodule classification experiment on the pulmonary nodules classification (NLST) are summarized in Table 2. The results are averaged over 10 repetitions.

Table 2: Average accuracy (%) on the pulmonary nodule classification with $N = 2$.

model	**# orient. (M)**	**# filters**	**# param.**	**accuracy** $_{\pm\sigma}$
Z3-CNN	-	4	3,818	80.0$_{\pm 1.7}$
Z3-CNN	-	96	82,754	81.3$_{\pm 2.2}$
LRI-CNN	1	4	970	76.3$_{\pm 3.8}$
LRI-CNN	4	4	970	79.0 $_{\pm 3.0}$
LRI-CNN	24	4	970	**81.9**$_{\pm 3.3}$
LRI-CNN	72	4	970	80.7$_{\pm 7.4}$

4. Discussions and Conclusion

The results on the synthetic dataset (Table 1) show that increasing the number of orientation channels significantly improves the performance (74.6% with a single orientation vs 90.0% with 72 orientations) and outperforms a standard Z3-CNN with the same number of filters (81.7%). Despite the increased number of orientation channels, the number of trainable parameters remains extremely low (40 parameters). Note that using data augmentation with random rotations of the training samples would not help the Z3-CNN as its architecture is simply inappropriate for LRI and patterns are already present at many random orientations in the training set. Adding more filters to the Z3-CNN ($2 \times 72 = 144$ filters) allows to learn filters at different orientations and achieves 95.9% accuracy, at the heavy cost of parameters and convolution operations. As shown in Fig. 4 with a single orientation channel, i.e. without using the steering capacity, the degree $N = 0$ of the SH cannot differentiate well patterns (63.9% accuracy) as it is isotropic. The performance then increases with N and nearly reaches the standard Z3-CNN accuracy for $N \geq 2$ with a significantly lower number of parameters, underlining the compression power of the parametric SH representation.

Note that LRI can be obtained with a G-CNN implementation (Cohen and Welling, 2016) by pooling across orientation channels after the first layer, yet it is limited to M=24 and requires to convolve the input with every rotated filter.

The results on the pulmonary nodule classification experiment (see Table 2) confirm the importance of LRI and of the proposed approach in a real medical imaging application. Despite the lack of directional texture patterns in the data which may reduce the performance gain over the baseline, an increase in accuracy is obtained with the LRI-CNN as well as a reduction of trainable parameters by a factor of four.

In conclusion, we developed a 3D LRI convolutional network using steerable filters. The main benefits are the low number of trainable parameters, the limited number of convolutions as we only convolve with the limited set of SHs and steer the responses for an arbitrary number of rotations, and the exactness of the steerability, avoiding approximation for kernel rotations. The results on synthetic 3D textures and 3D pulmonary nodule classification confirmed the importance of LRI with directionally sensitive steerable filters and the compression power of the proposed approach. While the reduction of trainable parameters also comes with an increase of time and memory as compared with a standard 3D CNN, the proposed efficient and precise LRI design may be preferred over data augmentation or group equivariant networks in various scenarios. In future work, we will look into finding the maximum orientation responses and/or powerful invariant descriptors without recombining the responses for all orientations which is a current bottleneck for memory consumption on the GPUs. We will also explore the benefit and cost of using non-polar-separable filters.

Acknowledgments

This work was supported by the Swiss National Science Foundation (grant 205320_179069).

References

M. Abramowitz and I. Stegun. *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*, volume 55. Courier Corporation, 1964.

V. Andrearczyk and A. Depeursinge. Rotational 3D texture classification using group equivariant CNNs. *arXiv preprint arXiv:1810.06889*, 2018.

V. Andrearczyk and P.F. Whelan. Using filter banks in convolutional neural networks for texture classification. *Pattern Recognition Letters*, 84:63–69, 2016.

E. J. Bekkers, M. W. Lafarge, M. Veta, K. AJ. Eppenhof, J. PW Pluim, and R. Duits. Roto-translation covariant convolutional networks for medical image analysis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 440–448. Springer, 2018.

N. Chenouard and M. Unser. 3D steerable wavelets in practice. *IEEE Transactions on Image Processing*, 21(11):4522–4533, 2012.

M. Cimpoi, S. Maji, I Kokkinos, and A. Vedaldi. Deep filter banks for texture recognition, description, and segmentation. *International Journal of Computer Vision*, 118(1):65–94, 2016.

T.S. Cohen and M. Welling. Group equivariant convolutional networks. *CoRR*, abs/1602.07576, 2016.

A. Depeursinge and J. Fageot. Biomedical texture operators and aggregation functions: A methodological review and user's guide. In *Biomedical Texture Analysis*, pages 55–94. Elsevier, 2018.

A. Depeursinge, J. Fageot, V. Andrearczyk, J.P. Ward, and M. Unser. Rotation invariance and directional sensitivity: Spherical harmonics versus radiomics features. In *International Workshop on Machine Learning in Medical Imaging*, pages 107–115. Springer, 2018.

Y. Dicente Cid, H. Müller, A. Platon, P.A. Poletti, and A. Depeursinge. 3-D solid texture classification using locally-oriented wavelet transforms. *IEEE Transactions on Image Processing*, 26(4): 1899–1910, April 2017. doi: 10.1109/TIP.2017.2665041.

J. R. Driscoll and D. M. Healy. Computing Fourier Transforms and Convolutions on the 2-Sphere. *Advances in applied mathematics*, 15(2):202–250, 1994.

J. Fageot, V. Uhlmann, Zs. Püspöki, B. Beck, M. Unser, and A. Depeursinge. Principled design and implementation of steerable detectors. *arXiv preprint arXiv:1811.00863*, 2018.

W.T. Freeman and E.H. Adelson. The design and use of steerable filters. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 13(9):891–906, 1991.

R.A. Gatenby, O. Grove, and R.J. Gillies. Quantitative imaging in cancer evolution and ecology. *Radiology*, 269(1):8–14, 2013.

K. He, X. Zhang, S. Ren, and J. Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.

S. P. Martin, J. Hofmeister, S. Burgmeister, S. Orso, N. Mili, S. Guerrier, M. P. Victoria-Fesser, P. M. Soccal, F. Triponez, W. Karenovics, N. Mach, A. Depeursinge, C. D. Becker, C. Rampinelli, P. Summers, H. Müller, and X. Montet. Artificial intelligence in lung nodule classification: a radiomics solution. *European Respiratory Journal*, submitted.

T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray–scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, July 2002.

M. Unser and N. Chenouard. A Unifying Parametric Framework for 2D Steerable Wavelet Transforms. *SIAM Journal on Imaging Sciences*, 6(1):102–135, 2013.

M. Varma and A. Zisserman. A statistical approach to texture classification from single images. *International Journal of Computer Vision*, 62(1-2):61–81, 2005. ISSN 0920-5691. doi: 10.1007/s11263-005-4635-4.

D.A. Varshalovich, A.N. Moskalev, and V.K. Khersonskii. *Quantum theory of angular momentum*. World Scientific, 1988.

M. Weiler, F.A. Hamprecht, and M. Storath. Learning steerable filters for rotation equivariant CNNs. *arXiv preprint arXiv:1711.07289*, 2017.

M. Weiler, M. Geiger, M. Welling, W. Boomsma, and T.S. Cohen. 3D steerable CNNs: Learning rotationally equivariant features in volumetric data. *arXiv preprint arXiv:1807.02547*, 2018.

M. Winkels and T.S. Cohen. 3D G-CNNs for pulmonary nodule detection. *arXiv preprint arXiv:1804.04656*, 2018.

D.E. Worrall, S. J. Garbin, D. Turmukhambetov, and G.J. Brostow. Harmonic networks: Deep translation and rotation equivariance. *CoRR*, abs/1612.0, 2016. doi: 10.1109/CVPR.2017.758.

H. Zhang, J. Xue, and K. Dana. Deep TEN: Texture encoding network. *arXiv preprint arXiv:1612.02844*, 2016.

Yanzhao Zhou, Qixiang Ye, Qiang Qiu, and Jianbin Jiao. Oriented response networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 519–528, 2017.

Appendix A. Equivariant Texture Operator

We prove the following result.

Proposition 1 *A texture operator of the form* (1) *is equivariant to translations and rotations in the sense of* (2).

Proof The equivariance to translations uses that $(I(\cdot - \boldsymbol{x}_0) * g)(\boldsymbol{x}) = (I * g)(\boldsymbol{x} - \boldsymbol{x}_0)$. Applying this to $g = f(\mathrm{R}\cdot)$, we deduce

$$\mathscr{G}\{I(\cdot - \boldsymbol{x}_0)\}(\boldsymbol{x}) = \max_{\mathrm{R} \in SO(3)} |(I * f(\mathrm{R}\cdot))(\boldsymbol{x} - \boldsymbol{x}_0)| = \mathscr{G}\{I\}(\boldsymbol{x} - \boldsymbol{x}_0), \tag{7}$$

as expected. For the rotation, we use $(I(\mathrm{R}_0\cdot) * g)(\boldsymbol{x}) = (I * g(\mathrm{R}_0^{-1}\cdot))(\mathrm{R}_0\boldsymbol{x})$ applied to $g = f(\mathrm{R}\cdot)$, to deduce

$$\mathscr{G}\{I(\mathrm{R}_0\cdot)\}(\boldsymbol{x}) = \max_{\mathrm{R} \in SO(3)} |(I * f(\mathrm{R}\mathrm{R}_0^{-1}\cdot))(\mathrm{R}_0\boldsymbol{x})| = \max_{\mathrm{R} \in SO(3)} |(I * f(\mathrm{R}\cdot))(\mathrm{R}_0\boldsymbol{x})| = \mathscr{G}\{I\}(\mathrm{R}_0\boldsymbol{x}), \tag{8}$$

where the second equality simply exploits that $\mathrm{R}\mathrm{R}_0^{-1}$ describes the complete space $SO(3)$ of 3D rotations when R varies. ■

We remark that the equivariance to translations is simply due to the use of the convolution, while the equivariance to rotations requires the presence of the pooling over 3D rotations in (1).

Appendix B. Spherical Harmonics

The family of SHs is denoted by $(Y_{n,m})_{n \geq 0, m \in \{-n,\dots,n\}}$, where n is called the degree and m the order of $Y_{n,m}$. SHs form an orthonormal basis for square-integrable functions in the $2D$-sphere $\mathbb{S}^2$. They are defined as (Driscoll and Healy, 1994)

$$Y_{n,m}(\theta,\phi) = A_{n,m} P_{n,|m|}(\cos(\theta)) \mathrm{e}^{\mathrm{j}m\phi}, \tag{9}$$

with $A_{n,m} = (-1)^{(m+|m|)/2} \left(\frac{2n+1}{4\pi} \frac{(n-|m|)!}{(n+|m|)!} \right)^{1/2}$ a normalization constant and $P_{n,|m|}$ the associated Legendre polynomial given for $0 \leq m \leq n$ by (Abramowitz and Stegun, 1964).

$$P_{n,m}(x) := \frac{(-1)^m}{2^n n!} (1 - x^2)^{m/2} \frac{\mathrm{d}^{n+m}}{\mathrm{d}x^{n+m}} (x^2 - 1)^n. \tag{10}$$

Appendix C. Real Steerable Filters

A filter f is real if $\overline{f(\rho,\theta,\phi)} = f(\rho,\theta,\phi)$ for every (ρ,θ,ϕ). For filters given by (3), this means that

$$\overline{h(\rho)} \sum_{n,m} \overline{\mathrm{C}_n[m] Y_{n,m}(\theta,\phi)} = h(\rho) \sum_{n,m} \mathrm{C}_n[m] Y_{n,m}(\theta,\phi), \tag{11}$$

We use the symmetry of the spherical harmonics, $\overline{Y_{n,m}} = (-1)^m Y_{n,-m}$, on the left-hand side and change the sign of m on the right-hand side to get

$$\sum_{n,m} \overline{h(\rho) \mathrm{C}_n[m]} (-1)^m Y_{n,-m}(\theta,\phi) = \sum_{n,m} h(\rho) \mathrm{C}_n[-m] Y_{n,-m}(\theta,\phi), \tag{12}$$

The $Y_{n,m}$ being linearly independent, we deduce that the filter is real if and only if, for any ρ, n, m, $\overline{h(\rho)\mathrm{C}_n[m]}(-1)^m = h(\rho)\mathrm{C}_n[-m]$. By imposing that h is real, i.e., $\overline{h} = h$, we obtain the expected criterion on the $\mathrm{C}_n[m]$ coefficients, which is

$$\mathrm{C}_n[-m] = (-1)^m \overline{\mathrm{C}_n[m]}, \tag{13}$$