2019

Solid Spherical Energy (SSE) CNNs for Efficient 3D Medical Image Analysis

Vincent Andrearczyk
University of Applied Sciences Western Switzerland

Valentin Oreiller
University of Applied Sciences Western Switzerland

Julien Fageot
University of Applied Sciences Western Switzerland

See next page for additional authors

Recommended Citation

Andrearczyk, V., Oreiller, V., Fageot, J., Montet, X. & Depeursinge, A. (2019). Solid spherical energy (SSE) CNNS for efficient 3D medical image analysis. *IMVIP 2019: Irish Machine Vision & Image Processing*, Technological University Dublin, Dublin, Ireland, August 28-30. doi: 10.21427/ccjb-2504

Solid Spherical Energy (SSE) CNNs for Efficient 3D Medical Image Analysis

Vincent Andrearczyk[1], Valentin Oreiller[1,2], Julien Fageot[1,3], Xavier Montet[4], and Adrien Depeursinge[1,2]

[1]*Institute of Information Systems, University of Applied Sciences Western Switzerland (HES-SO), Sierre, Switzerland*
[2]*Centre Hospitalier Universitaire Vaudois (CHUV), Lausanne, Switzerland*
[3]*Harvard School of Engineering and Applied Sciences, Cambridge, MA, USA*
[4]*Hopitaux Universitaires de Genève (HUG), Geneva, Switzerland*

July 15, 2019

Abstract

Invariance to local rotation, to differentiate from the global rotation of images and objects, is required in various texture analysis problems. It has led to several breakthrough methods such as local binary patterns, maximum response and steerable filterbanks. In particular, textures in medical images often exhibit local structures at arbitrary orientations. Locally Rotation Invariant (LRI) Convolutional Neural Networks (CNN) were recently proposed using 3D steerable filters to combine LRI with Directional Sensitivity (DS). The steerability avoids the expensive cost of convolutions with rotated kernels and comes with a parametric representation that results in a drastic reduction of the number of trainable parameters. Yet, the potential bottleneck (memory and computation) of this approach lies in the necessity to recombine responses for a set of predefined discretized orientations. In this paper, we propose to calculate invariants from the responses to the set of spherical harmonics projected onto 3D kernels in the form of a lightweight Solid Spherical Energy (SSE) CNN. It offers a compromise between the high kernel specificity of the LRI-CNN and a low memory/operations requirement. The computational gain is evaluated on 3D synthetic and pulmonary nodule classification experiments. The performance of the proposed approach is compared with steerable LRI-CNNs and standard 3D CNNs, showing competitive results with the state of the art.

Keywords: Convolutional neural network, local rotation invariance, spherical harmonics, spectrum

1 Introduction

The translation equivariance of the convolution operator is greatly exploited in Convolutional Neural Networks (CNNs), as particularly adapted to grid-like data analysis. Rotation equivariance/invariance, originally approximated by data-augmentation in CNNs training, has recently been extensively studied using group theory [Andrearczyk and Depeursinge, 2018, Cohen and Welling, 2016, Winkels and Cohen, 2018]. The Group equivariant CNNs (G-CNN) uses rotated versions of the filters with right angle rotations in 2D [Cohen and Welling, 2016] or 3D [Winkels and Cohen, 2018]. Rotation invariance is then obtained by pooling across orientations. This network architecture with both spatial and orientation weight sharing showed promising results in the analysis of pulmonary nodule detection [Winkels and Cohen, 2018]. Results on the 3D textures classification [Andrearczyk and Depeursinge, 2018], however, suggested the use of a finer rotation sampling. Note that G-CNNs are adapted to equivariance with respect to *finite* subgroups of the rotation group. While an arbitrary sampling of rotations can be used in 2D [Bekkers et al., 2018], the number of 3D finite rotation groups

is limited [Winkels and Cohen, 2018]. Steerable filters were used in 2D CNNs [Weiler et al., 2018b] [Worrall et al., 2017] to obtain rotation invariance from an arbitrary number of rotations without the need of convolving the input with all rotated version of the filters. Recently, 3D steerable CNNs were developed in [Weiler et al., 2018a] with general architectures to implement the global equivariance to rotations. Besides, a related work was proposed in [Eickenberg et al., 2017], where spherical harmonics energy is used as a source of rotation invariance in the context of scattering transform.

In the above approaches, global rotation equivariance is maintained throughout the network, while invariance is obtained by pooling over the orientations after the last convolution layer. This global rotation invariance is fundamental in various computer vision applications such as objects recognition with arbitrary orientations. Some images, however, are composed of textures with well-defined local structures at arbitrary orientations. Computed Tomography (CT) and Magnetic Resonance Imaging (MRI), for instance, may exhibit diverse tissue alterations [Gatenby et al., 2013] with specific visual signatures including blobs and intersecting surfaces. These local low-level patterns are generally characterized by discriminative directional properties (e.g. composed of edges) and are repeated at arbitrary orientations (see example in Figure 1). These two properties

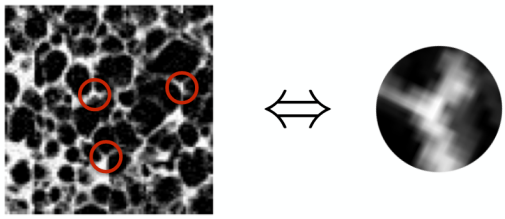

Figure 1: Examples of directional local patterns (collagen junctions) occurring at various arbitrary orientation in lung CT images.

require using image operators that can combine Directional Sensitivity (DS) with Local Rotation Invariance (LRI) as defined in [Andrearczyk et al., 2019]. While local operators sensitive to directional features are required to avoid mixing blobs, edges and ridges, DS is often antagonist to the notion of rotation invariance. A convolutional operator, for instance, is equivariant to rotations if and only if the filter is isotropic, thus insensitive to directional patterns. To this end, more complex operators have been designed to combine LRI and DS including MR8 [Varma and Zisserman, 2005], Local Binary Patterns (LBP) [Ojala et al., 2002], 3D Riesz wavelets [Dicente Cid et al., 2017] and Spherical Harmonic (SH) invariants [Depeursinge et al., 2018].

The work in [Andrearczyk et al., 2019] exploited the steerability of SHs to obtain a 3D CNN architecture which combines DS and LRI. This is achieved with an arbitrarily fine rotation sampling and controlled operator support by steering convolutional responses to a set of SHs. The rotation invariance is then realized by a max-pooling across all the orientations. In this paper, we propose an alternative to this approach by calculating invariants in the form of Solid Spherical Energy (SSE) obtained after projecting local data neighborhoods onto solid SH kernels. As opposed to steerable kernels, the SSE invariants obviate the need for SH response recombinations for a set of discretized orientations as performed in [Andrearczyk et al., 2019], which results in a drastic reduction in terms of GPU memory and operations. Note that it is also an advantage over the 3D G-CNN approach that requires the computation of response maps for every rotated versions of the filters, before performing a max-pooling operation over the orientations. The proposed 3D SSE-CNNs can be trained end to end, learning various radial supports to the SHs that extract informative features characterizing the analyzed 3D textures.

2 Methods

Images are described as analog functions from $\mathbb{R}^3$ to $\mathbb{R}$. We use spherical coordinates with $\rho \geq 0$ the radius, $\theta \in [0, \pi]$ the elevation angle, and $\phi \in [0, 2\pi)$ the azimuthal angle. Working over a continuum allows us to consider the complete space of $3D$ rotations $SO(3)$ when rotating an image I. We discuss the practical discretization of our continuous-domain framework in Section 2.2.

2.1 Invariants from Solid Spherical Energy

We aim at defining locally rotation invariant texture operators $\mathscr{G}$ that are able to capture the local directional information from the image $I : \mathbb{R}^3 \to \mathbb{R}$ [Depeursinge and Fageot, 2018]. Our strategy is to use SHs for their ability to produce spectral invariants when combined within a fixed degree n, referred to as Solid Spherical Energy (SSE) in the following. This allows combining LRI and DS, which are known to be antagonist in simple linear convolutional designs [Andrearczyk et al., 2019, Depeursinge et al., 2018].

Spherical harmonics. The family of SHs is denoted by $(Y_{n,m})$ where $n \geq 0$ is the degree and $m \in \{-n, \cdots, n\}$ is the order. The formal definition can be found for instance in [Driscoll and Healy, 1994]. For our purpose, it is sufficient to know that the SHs form an orthonormal basis of the square-integrable functions on the $2D$ sphere. Moreover, for any rotation $\mathrm{R} \in SO(3)$, the rotated SH $Y_{n,m}(R\cdot)$ can be expressed as a linear combination of the other SHs of the same degree n as

$$Y_{n,m}(\mathrm{R}\cdot) = \sum_{m'=-n}^{n} \mathrm{D}_{n,R}[m, m'] \, Y_{n,m'}. \tag{1}$$

The matrix $\mathrm{D}_{n,\mathrm{R}} \in \mathbb{C}^{(2n+1)\times(2n+1)}$ is called a Wigner D-matrix. It can be computed in closed form and is unitary [Varshalovich et al., 1988]. In particular, for any vector $\boldsymbol{c} \in \mathbb{C}^{2n+1}$, we have that $\|\mathrm{D}_{n,\mathrm{R}}\boldsymbol{c}\| = \|\boldsymbol{c}\|$ (norm preservation). We consider finitely many degrees, $N \geq 0$ being the maximal degree (equivalent to a maximal frequency), resulting in $\sum_{n=0}^{N}(2n+1) = (N+1)^2$ SHs of degree up to N.

Radial profiles and the SSE response maps. The SHs $(Y_{n,m})$ are defined on the $2D$ sphere and then modulated by a trainable radial support h to produce a $3D$ filter $hY_{n,m} = h(\rho)Y_{n,m}(\theta, \phi)$. After convolution with the image I, the responses $I * hY_{n,m}$ with $m = -n, \cdots, n$ contain the spectral information of degree n, which is used to define the texture operator $\mathscr{G}_n$ as

$$\mathscr{G}_n\{I\}(\boldsymbol{x}) = \sum_{m=-n}^{n} |(I * hY_{n,m})(\boldsymbol{x})|^2. \tag{2}$$

Let us study the desirable properties of $\mathscr{G}_n$ in the following. At a fixed spatial position $\boldsymbol{x} \in \mathbb{R}^3$, the projection $(I * hY_{n,m})(\boldsymbol{x}) = \langle hY_{n,m}, I(\boldsymbol{x} - \cdot)\rangle$ measures the correlation[1] of $hY_{n,m}$ with I at $\boldsymbol{x}$.

Eq. (2) defines an operator with perfect rotational invariance (see Proposition 1 below), while being sensitive to directional information via spherical frequencies of degree $n > 0$. We call $\mathscr{G}_n\{I\}$ the *SSE response map of degree n* of I.

Invariance of the SSE response maps. We have the following invariance properties of $\mathscr{G}_n$. For an in-depth discussion about the invariance and equivariance properties of texture operators, see [Depeursinge and Fageot, 2018].

Proposition 1 *The SSE response maps are equivariant to rotations and translations as follows*

$$\mathscr{G}_n\{I(\cdot - \boldsymbol{x}_0)\} = \mathscr{G}_n\{I\}(\cdot - \boldsymbol{x}_0) \quad \textit{for any } \boldsymbol{x}_0 \in \mathbb{R}^3, \tag{3}$$

$$\mathscr{G}_n\{I(\mathrm{R}_0\cdot)\} = \mathscr{G}_n\{I\}(\mathrm{R}_0\cdot) \quad \textit{for any } \mathrm{R}_0 \in SO(3). \tag{4}$$

In addition, $\mathscr{G}_n$ is locally rotation invariant in the sense that, if $h(\rho) = 0$ for $\rho > \rho_0$, then $\mathscr{G}_n\{J\}(\boldsymbol{x}) = \mathscr{G}_n\{I\}(\boldsymbol{x})$ if J is obtained from I by locally rotating an object of radius less than ρ_0 around $\boldsymbol{x}$.

Proof 1 *Given a filter g (e.g. $g = hY_{n,m}$), from the relation $(I(\cdot - \boldsymbol{x}_0) * g)(\boldsymbol{x}) = (I * g)(\boldsymbol{x} - \boldsymbol{x}_0)$, we deduce that $\mathscr{G}_n\{I(\cdot - \boldsymbol{x}_0)\}(\boldsymbol{x}) = \sum_{m=-n}^{n} |(I * hY_{n,m})(\boldsymbol{x} - \boldsymbol{x}_0)|^2 = \mathscr{G}_n\{I\}(\boldsymbol{x} - \boldsymbol{x}_0)$, which is (3). Now, using that $(I(\mathrm{R}_0\cdot) * g)(\boldsymbol{x}) =$*

[1] Strictly speaking, it is the correlation of $hY_{n,m}$ with $I(-\cdot)$, I being real valued.

$(I * g(\mathrm{R}_0^{-1}\cdot))(\mathrm{R}_0\boldsymbol{x})$ *and* (1)*, we have*

$$\begin{aligned}\mathscr{G}_n\{I(\mathrm{R}_0\cdot)\}(\boldsymbol{x}) &= \sum_{m=-n}^{n} |(I * (hY_{n,m}(\mathrm{R}_0^{-1}\cdot)))(\mathrm{R}_0\boldsymbol{x})|^2 \\ &= \sum_{m=-n}^{n} \left| \sum_{m'=-n}^{n} \mathrm{D}_{n,\mathrm{R}_0^{-1}}[m,m'](I * hY_{n,m'})(\mathrm{R}_0\boldsymbol{x}) \right|^2 . \end{aligned} \tag{5}$$

Now, the Wigner D-matrix being norm-preserving, we have that $\sum_m |c_m|^2 = \sum_m |\sum_{m'} \mathrm{D}_{n,\mathrm{R}_0^{-1}}[m,m']c_{m'}|^2$ *for any* $\boldsymbol{c} = (c_{-n},\ldots,c_n)$*. Applying this relation to* $c_m = (I * hY_{n,m})(\mathrm{R}_0\boldsymbol{x})$*, we deduce* (4) *from* (5)*. Finally, as developed in [Andrearczyk et al., 2019], the LRI is a consequence of the equivariance to rotations and translations.*

Radial sensitivity and learning. We used a single radial profile h in the above. One can refine the radial sensitivity of the method by defining SH response maps $\mathscr{G}_{n,i}\{I\}$ for various channels $i \in \{1\cdots C\}$ associated with various radial profiles h_i. Typically, we use h_i with a spatial support located in the area $G = \{\boldsymbol{x} \in \mathbb{R}^3, \|\boldsymbol{x}\| \leq \rho_0\}$, where $\rho_0 > 0$ is fixed, giving C texture operators for each degree n that are locally rotation invariant in the sense of Proposition 1. In total, the method yields $C \cdot (N+1)$ response maps. In practice, the radial profiles are learned via backpropagation to optimize the considered image processing task. The specificity of the method is, therefore, to use the handcrafted spectrum-based framework of [Depeursinge et al., 2018] in a learning framework similar to [Andrearczyk et al., 2019].

2.2 Discretization

The radial profiles h_i, and consequently the modulated SHs $h_i Y_{n,m}$, have a compact spherical support $G = \{\boldsymbol{x} \in \mathbb{R}^3, \|\boldsymbol{x}\| \leq \rho_0\}$. For each output channel i, we consider the voxelization of the radial profile $h_i(\rho)$ as in [Andrearczyk et al., 2019]. 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. The values of the $h_i Y_{n,m}$ over the continuum is deduced from the discretization using linear interpolation as detailed in [Andrearczyk et al., 2019]. As discussed in [Andrearczyk et al., 2019], the maximal frequency N cannot be taken arbitrarily large once the radial profiles are voxelized. In practice, one considers small values of N that are adapted to the kernel size.

3 Experimental Setup

3.1 Datasets

The first dataset was developed in [Andrearczyk et al., 2019] and contains 1000 synthetic $32 \times 32 \times 32$ texture volumes of two classes. The textures are designed by placing binary segments and crosses at random orientations and locations with overlap. The classes differ by the proportion of patterns and variability is introduced by the overlapping patterns, the linear interpolation of the 3D rotations and the varying densities. 800 volumes are used for training and the remaining 200 for testing.

The second dataset used in [Andrearczyk et al., 2019] is a subsample of the American National Lung Screening Trial (NLST) annotated by radiologists [Martin et al., 2019]. It includes 485 pulmonary nodules, 244 benign and 241 malignant, from distinct patients in CT. The input volumes, originally ranging from $16 \times 16 \times 16$ to $128 \times 128 \times 128$, are centered on the tumor and padded or cropped to the size $64 \times 64 \times 64$. Balanced training and test splits are used with 392 and 93 volumes respectively. The Hounsfield units are clipped in the range $[-1000, 400]$, then standardized with zero mean and unit variance.

3.2 Network architecture

We follow the architectures and hyperparameters of the LRI-CNNs in [Andrearczyk et al., 2019] with a first SSE convolution layer, replacing the steering and orientation max-pooling by the invariant calculation. Spatial global average pooling is then used to aggregate the LRI responses. For the nodule experiment, the responses

are aggregated inside the nodule masks instead of across the entire feature maps. The number of trainable parameters is reduced as compared to the LRI-CNN in [Andrearczyk et al., 2019] as there are no coefficients for each n, m pair (see Section 4). Finally, the aggregated responses are connected to dense layers. For the synthetic experiment, a single dense layer with softmax activation and a cross-entropy loss. For the second, more complex experiment, an intermediate dense layer with 128 neurons is used before the same final softmax layer. ReLU activations are employed and the networks are trained using Adam optimizer. More details on the reproduced filters sizes and training settings can be found in [Andrearczyk et al., 2019]. For the synthetic experiment the following parameters are used: 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 this network as SSE-CNN and compare it with LRI-CNN with responses steered at M orientations (best results only reported with $M = 72$ and $M = 24$) and with the Z3-CNN (standard 3D CNN).

4 Results and Discussions

Results of the proposed SSE-CNN on the synthetic dataset for varying values of the maximal degree N are reported and compared with the state of the art methods (Z3-CNN and LRI-CNN) in Fig. 2. Similarly, results in the pulmonary nodules classification (NLST) are reported in Fig. 3.

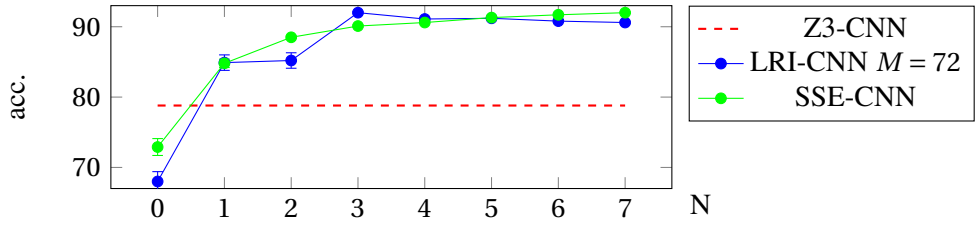

Figure 2: Average accuracy (%) and standard error (10 repetitions) on the synthetic dataset with varying values of N and comparison with the state of the art LRI-CNN with $M = 72$ orientations (best reported results) and Z3-CNN.

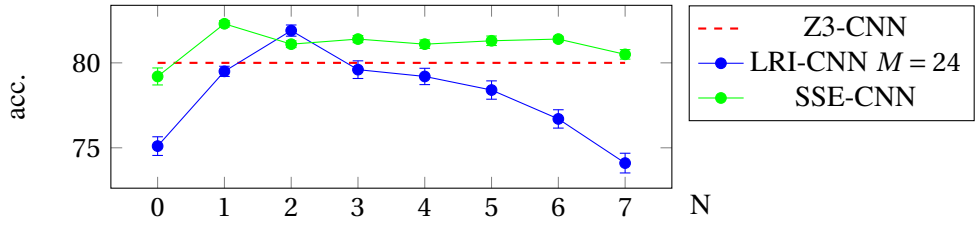

Figure 3: Average accuracy (%) and standard error (10 repetitions) on the pulmonary nodule dataset with varying values of N and comparison with the state of the art LRI-CNN with $M = 24$ orientations (best reported results) and Z3-CNN.

The results in Figures 2 and 3 show that the proposed SSE-CNN performs as good or better than the LRI-CNN. This suggests that the full parametric representation and the steering are not always necessary and that the SSE is sufficient and maybe simpler to learn on these datasets. From a theoretic perspective, when compared to using steerability in the LRI-CNN, the use of degree-wise norms in the SSE-CNN discards inter-degree phases and mixes intra-degree responses. However, this information does not seem to be crucial for the two discrimination tasks considered in this paper. In future work, we will investigate more robust invariants such as the bispectrum [Kakarala and Mao, 2010].

Increasing the maximal spherical frequency N helps on the synthetic dataset (Fig. 2) as it contains well defined directional patterns (lines and crosses). To investigate whether the performance gain obtained with SHs of increasing degrees differs from a simple increase of parameters and associated feature maps, we trained our SSE-CNN with $N = 0$ and more output channels ($C = 8$ instead of $C = 2$). This setup relies on a number of output feature maps that is equal to the SSE-CNN with $N = 3$ in Fig. 2. Yet, the accuracy of the former

Table 1: Computational and parameters comparison with the state of the art on the synthetic dataset with the setup of Fig. 2. The computational time is measured in seconds for 1000 iterations trained on a Tesla K80 GPU.

model	Z3-CNN	LRI-CNN								SSE-CNN							
N	-	0	1	2	3	4	5	6	7	0	1	2	3	4	5	6	7
# param.	694	24	30	40	54	72	94	120	184	22	28	34	40	46	52	58	64
time	28	56	63	88	100	126	180	211	221	30	33	54	64	82	122	145	161

is $81.5\%_{\pm 3.2}$ versus $90.1\%_{\pm 1.5}$ for the latter, which highlights the relevance of SSE extracted at various degrees n. Besides, in order to show the importance of the proposed LRI design, we compare it to the Z3-CNN with right-angle rotation data-augmentation. Our SSE-CNN largely outperforms the latter > 90% with $N > 2$ versus 84.0% on the synthetic dataset.

The drop of accuracy in the NLST dataset (Figure 3) of the LRI-CNN [Andrearczyk et al., 2019] with large frequencies may be due to its complexity together with the limited amount of data. The simplified design of the SSE-CNN likely acts as a regularizer explaining the large differences in performance.

The number of trainable parameters and the computational times are reported in Table 1. The number of parameters is largely reduced as compared to the standard Z3-CNN and LRI-CNN [Andrearczyk et al., 2019]. We can differentiate two types of parameter reduction, namely the weight sharing across rotations and the parametric representation. The weight sharing is present, among others, in the G-CNN [Cohen and Welling, 2016, Winkels and Cohen, 2018] where the same kernels account for multiple orientations. Similarly, the LRI-CNN [Andrearczyk et al., 2019] shares trainable radial profiles and harmonic coefficients to obtain responses to rotated kernels. The parametric representation, on the other hand, is used in [Andrearczyk et al., 2019, Weiler et al., 2018b] by learning a combination of basis filters instead of every voxel of the kernels. Our approach goes one step further as we do not learn an explicit full parametric representation of the kernels: we rather only calculate the norm of the SHs responses, resulting in a large reduction of the number of parameters. One bottleneck with 3D rotation invariant CNNs such as 3D G-CNN [Winkels and Cohen, 2018] and LRI-CNN [Andrearczyk et al., 2019] is the GPU memory usage when obtaining 3D response maps at all orientations M before orientation pooling. This memory consumption is largely reduced in the proposed implementation by computing invariants on the SHs responses rather than calculating responses at all orientations. For a similar reason, the time consumption is lower than the LRI-CNN as shown in Table 1. Note that the implementation only uses existing TensorFlow functions and the computation could be further improved by efficient parallelization and CUDA programming such as used in the Z3-CNN.

5 Conclusions and Future Work

In conclusion, we showed the benefit of using invariants from SSE responses to combine LRI and DS in an efficient lightweight architecture. The main benefit over the steerable [Andrearczyk et al., 2019] and G-CNN [Winkels and Cohen, 2018] approaches is the lower computation and memory cost: our method does not require to compute responses at all desired orientations. Furthermore, it requires less parameters which could make this approach more suited to medical image analysis where data are scarce and models should generalize with few training samples. Experiments on 3D synthetic textures and scans of pulmonary nodules showed competitive results with the steerable approach and a large improvement from a standard 3D CNN with or without data-augmentation. In a future work, we will explore deeper architectures with LRI and potentially more discriminative invariants as mentioned in Section 4

Acknowledgments

This work was supported by the Swiss National Science Foundation (grants 205320_179069 and P2ELP2_181759).

References

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

[Andrearczyk et al., 2019] Andrearczyk, V., Fageot, J., V., O., X., M., and Depeursinge, A. (2019). Exploring local rotation invariance in 3D CNNs with steerable filters. In *International Conference on Medical Imaging with Deep Learning*.

[Bekkers et al., 2018] Bekkers, E. J., Lafarge, M. W., Veta, M., Eppenhof, K. A., Pluim, J. P., and Duits, R. (2018). Roto-translation covariant convolutional networks for medical image analysis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*.

[Cohen and Welling, 2016] Cohen, T. and Welling, M. (2016). Group equivariant convolutional networks. In *International Conference on Machine Learning*.

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

[Depeursinge et al., 2018] Depeursinge, A., Fageot, J., Andrearczyk, V., Ward, J., and Unser, M. (2018). Rotation invariance and directional sensitivity: Spherical harmonics versus radiomics features. In *Workshop on Machine Learning in Medical Imaging*.

[Dicente Cid et al., 2017] Dicente Cid, Y., Müller, H., Platon, A., Poletti, P., and Depeursinge, A. (2017). 3-D solid texture classification using locally-oriented wavelet transforms. *IEEE Transactions on Image Processing*, 26(4):1899–1910.

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

[Eickenberg et al., 2017] Eickenberg, M., Exarchakis, G., Hirn, M., and Mallat, S. (2017). Solid harmonic wavelet scattering: Predicting quantum molecular energy from invariant descriptors of 3d electronic densities. In *Advances in Neural Information Processing Systems*, pages 6540–6549.

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

[Kakarala and Mao, 2010] Kakarala, R. and Mao, D. (2010). A theory of phase-sensitive rotation invariance with spherical harmonic and moment-based representations. In *Conference on Computer Vision and Pattern Recognition*.

[Martin et al., 2019] Martin, S. P., Hofmeister, J., Burgmeister, S., Orso, S., Mili, N., Guerrier, S., Victoria-Fesser, M. P., Soccal, P. M., Triponez, F., Karenovics, W., Mach, N., Depeursinge, A., Becker, C. D., Rampinelli, C., Summers, P., Müller, H., and Montet, X. (2019). Identification of malignant lung nodules and reduction in false-positive findings by augmented intelligence: A radiomic study based on the NLST dataset. *Journal of Clinical Oncology, submitted*.

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

[Varma and Zisserman, 2005] Varma, M. and Zisserman, A. (2005). A statistical approach to texture classification from single images. *International Journal of Computer Vision*, 62(1-2):61–81.

[Varshalovich et al., 1988] Varshalovich, D., Moskalev, A., and Khersonskii, V. (1988). *Quantum theory of angular momentum*. World Scientific.

[Weiler et al., 2018a] Weiler, M., Geiger, M., Welling, M., Boomsma, W., and Cohen, T. (2018a). 3D steerable CNNs: Learning rotationally equivariant features in volumetric data. In *Neural Information Processing Systems*.

[Weiler et al., 2018b] Weiler, M., Hamprecht, F., and Storath, M. (2018b). Learning steerable filters for rotation equivariant CNNs. In *Computer Vision and Pattern Recognition*.

[Winkels and Cohen, 2018] Winkels, M. and Cohen, T. (2018). 3D G-CNNs for pulmonary nodule detection. In *International Conference on Medical Imaging with Deep Learning*.

[Worrall et al., 2017] Worrall, D., Garbin, S. J., Turmukhambetov, D., and Brostow, G. (2017). Harmonic networks: Deep translation and rotation equivariance. In *Computer Vision and Pattern Recognition*.