A Bispectral 3D U-Net for Rotation Robustness in Medical Segmentation

Arthur Chevalley[1,2], Valentin Oreiller[2], Julien Fageot[2], John O. Prior[1], Vincent Andrearczyk[2], and Adrien Depeursinge[1,2]

[1] Nuclear Medicine and Molecular Imaging Department, Lausanne University Hospital (CHUV), Lausanne, Switzerland
[2] Institute of Informatics, HES-SO Valais-Wallis University of Applied Sciences and Arts Western Switzerland, Sierre, Switzerland

Abstract. Segmentation models achieved expert-level performance in a large variety of medical applications. However, their robustness to rotations, crucial for clinical use, is rarely discussed with the risk of discarding subtle but diagnostically relevant structures appearing in a wide range of positions and rotations. In this work, we investigate the robustness to rotations of a standard 3D nnU-Net in the context of two segmentation tasks: the hippocampus in MRI and the pulmonary airway system in CT. In addition, we introduce a 3D Locally Rotation Invariant (LRI) operator based on the bispectrum to achieve high robustness to input rotations. It is compared to a standard nnU-Net, a nnU-Net with extended rotational data augmentation and XEdgeConv, a state-of-the-art approach for RI. While all models performed similarly regarding the Dice score for right-angle rotations, the Bispectral U-Net outperformed other designs in the context of finer and more realistic rotations. Furthermore, the Bispectral U-Net and the XEdgeConv were more stable w.r.t. input rotation, i.e. the predictions are significantly more consistent across input rotations. Important inconsistencies of the nnU-Net were observed for lung airway segmentation, suggesting potential risks for clinical use.

Keywords: Local Rotation Invariance · Robust 3D Segmentation · Convolutional Bispectral Network · Deep Learning · Medical Image Analysis

1 Introduction

Convolutional Neural Networks (CNN) are currently the workhorse for many medical image analysis tasks. These models must reach high performance and reliability for segmentation as errors can lead to severe clinical consequences. The nnU-Net [7] introduced a self-adapting framework reaching state-of-the-art performances. However, this framework relies solely on data augmentation to achieve rotation robustness, which may not be sufficient for accurately contouring biomedical structures appearing at a wide range of orientations, with global as well as local image rotations. Nevertheless, the impact of input rotations on the model's performance has been little studied to date. In this work, we first evaluate the performance robustness and rotation stability of the nnU-Net. We then propose a 3D extension of a 2D Rotation Invariant (RI) segmentation model based on the bispectral operator [12]. The operator complexity, projection over the spherical harmonics rather than the circular ones, and the range of possible rotations (parameterized by 3 angles in 3D) make this step quite challenging. We finally compare the performance and stability of these models with another RI model, XEdgeConv, on two segmentation tasks. Code will be made available.

2 Related Work

CNN's robustness to input rotations, a key property for various medical tasks, has seen limited investigation. Generally, rotational robustness is achieved with heavy data augmentation which does not pledge invariance. To address this, invariant or equivariant convolution layers were proposed. One strategy is to design group-equivariant convolution layers relying on rotated and reflected duplicates of all kernels to achieve equivariance [20, 1]. Other developed specific parametric and/or steerable kernels [18] to create SE(3) equivariant networks using Circular [22] or Spherical Harmonics [4].

For texture classification, handcrafted Locally Rotation Invariant (LRI) operators achieved excellent results, e.g. local binary patterns [6] or steerable detectors [5, 16]. Those operators have also been incorporated in 3D CNNs using SHs [2, 11]. RI 3D point cloud segmentation was proposed [26, 14, 15, 8] but only a few works focused on invariant segmentation of 3D images.

CubeNet [21] uses the group convolution in 3D to create rotation and translation equivariant CNNs using Klein's four- and tetrahedral-group, yielding a *Roto-translational group-convolution*. 3D-UCaps [10] proposed a 3D capsules pathway in addition to a standard CNN. However, the capsules do not fully encode rotation invariance as they reported similar segmentation performances robustness compared to a standard U-Net. XEdgeConv [17] uses recent advances in graph neural networks to construct a kernels from translation and permutation invariant graphs. Finally, [12] implemented a 2D LRI U-Net based on the bispectrum operator, which we extend to 3D image segmentation in this work.

3 Methods

This section presents the mathematical background for the proposed Bispectral U-Net for 3D images. By embedding the bispectrum operator in a convolutional layer, it is convenient to compare the effect of various convolution strategies for a given network architecture, e.g. a *standard* 3D convolution for the nnU-Net model or another RI method such as XEdgeConv [17].

3.1 Notations

The 3D images considered are defined as functions $I(\boldsymbol{x}) \in L_2(\mathbb{R}^3)$ where $I(\boldsymbol{x})$ is the image intensity at the location $\boldsymbol{x} = (x_1, x_2, x_3) \in \mathbb{R}^3$. The spherical coordinates are defined as (ρ, θ, ϕ) where $\rho \geq 0$ is the radius, $\theta \in [0, \pi]$ the elevation angle and $\phi \in [0, 2\pi[$ the horizontal plane angle. On $\mathbb{R}^3$, the unit sphere is defined as $\mathbb{S}^2 = \{\boldsymbol{x} \in \mathbb{R}^3 : ||\boldsymbol{x}||_2 = 1\}$. Functions on the sphere are given as $f \in L_2(\mathbb{S}^2)$ using the spherical coordinates. The inner-product for $f, g \in L_2(\mathbb{S}^2)$ is defined by $\langle f, g\rangle_{L_2(\mathbb{S}^2)} = \int_0^\pi \int_0^{2\pi} f(\theta,\phi)g(\theta,\phi)\sin(\theta)\mathrm{d}\phi\mathrm{d}\theta$. The triangle function is defined as $\mathrm{tri}(\boldsymbol{x}) = 1 - |\boldsymbol{x}|$ if $|\boldsymbol{x}| < 1$ and $\mathrm{tri}(\boldsymbol{x}) = 0$ otherwise. Finally, the Kronecker product is denoted by $\otimes$ and the Hermitian transpose by †.

3.2 Bispectrum Operators

In addition to the global rotation and translation equivariance properties provided by any LRI operator [2], the bispectrum operator is sensitive to directional patterns and complete, i.e. functions with identical bispectrum are rotated versions of each other [9]. As this work focuses on analysing rotation robustness, we only recall essential definitions as extended details are provided in [11].

The bispectrum is computed using the Fourier transform on the sphere relying on Spherical Harmonics (SH). The family of SHs, made of functions Y_n^m for a degree $n \in \mathbb{N}$ and order m with $-n \leq m \leq n$, is known to form an orthonormal basis of $L_2(\mathbb{S}^2)$. Therefore, any function $f \in L_2(\mathbb{S}^2)$ can be projected onto the SH basis following the inner product

$$F_n^m = \langle f, Y_n^m\rangle_{L_2(\mathbb{S}^2)}, \tag{1}$$

in which case $f = \sum_{n\geq 0}\sum_{-n\leq m\leq n} F_n^m Y_n^m$. The Spherical Fourier (SF) vector for a given degree n is grouping the coefficients of all orders m as

$$\mathcal{F}_n = [F_n^{-n} .. F_n^0 .. F_n^n]. \tag{2}$$

Finally, following [19], a bispectrum coefficient of degree l, $|n-n'| \leq l \leq n+n'$, $n, n' \geq 0$, of any function $f \in L_2(\mathbb{S}^2)$ obtained with the operator $\mathcal{B}$ is defined as

$$b_{n,n'}^l(f) = [\mathcal{F}_n \otimes \mathcal{F}_{n'}]C_{nn'}\tilde{\mathcal{F}_l}^\dagger = \mathcal{B}\{\mathcal{F}_n, \mathcal{F}_{n'}, \mathcal{F}_l\}, \tag{3}$$

where $\mathcal{F}_n \otimes \mathcal{F}_{n'}$ is a $1 \times (2n+1)(2n'+1)$ vector and $C_{nn'}$ is the $(2n+1)(2n'+1) \times (2n+1)(2n'+1)$ Clebsh-Gordan matrix containing the namesake coefficients. $\tilde{\mathcal{F}_l}$ contains the SF vector of degree l. It is zero-padded to match $C_{nn'}$ size and allows to *select* only the rows associated with the l^{th} degree.

Bispectrum Operators for 3D Images While Eq. (2) applies to functions defined on the sphere, this work is interested in its application to 3D images. This requires extending SH bases to 3D volumes. The so-called Solid SHs are created by multiplying SHs with compactly supported radial profiles $h_n(\rho)$ [11]. Solid SHs of degree n and order m evaluated on the Cartesian grid are defined as

$$\kappa_n^m(\boldsymbol{x}) = \kappa_n^m(\rho,\theta,\phi) = h_n(\rho)Y_n^m(\theta,\phi). \tag{4}$$

From this equation, SF maps can be created for each degree by convolving the image with the solid SHs as

$$\mathcal{F}_n(\boldsymbol{x}) = [(I * \kappa_n^m)(\boldsymbol{x})]_{m=-n}^{m=n}. \tag{5}$$

Note the slight abuse of notation as $\mathcal{F}_n(\boldsymbol{x})$ is the local projection, around $\boldsymbol{x}$, of the image to a function on the sphere, further projected onto the SHs basis [11]. Finally, for any $I \in L_2(\mathbb{R}^3)$ and $\boldsymbol{x} \in \mathbb{R}^3$, the bispectrum image operator of degree $n, n' \geq 0$ and $|n - n'| \leq l \leq n + n'$ can be created from Eq. (3) and (5) as

$$\mathcal{G}_{n,n',l}\{I\}(\boldsymbol{x}) = \mathcal{B}\{\mathcal{F}_n(\boldsymbol{x}), \mathcal{F}_{n'}(\boldsymbol{x}), \mathcal{F}_l(\boldsymbol{x})\}. \tag{6}$$

This operator inherits the invariance properties (e.g. LRI) of the bispectrum [11].

3.3 Bispectral LRI Layer Implementation

The bispectrum operator implementation requires setting the maximal SH decomposition degree $N \geq 0$. The computed coefficients are restrained to limit the computational cost, i.e. complexity of $\mathcal{O}(N^3)$. In addition, N is limited by the discretization of the solid SH on a cubic kernel, similarly to a Nyquist frequency, by $N \leq \frac{\pi c}{4}$ where c is the kernel size [2]. As stated in [11], only the components satisfying $0 \leq n \leq n'$ and $0 \leq n + n' \leq N$ are kept as $b^l_{n,n'}(f)$ and $b^l_{n',n}(f)$ are proportional independently of f. In addition, only the indices where the sum $n + n' + l$ is even are kept as the coefficients are observed to be zero otherwise.

The implementation of the solid SHs is done in multiple steps. Firstly, the radial profiles $h_n(\rho)$ are constructed as linear combinations of radial functions $\psi_j(\rho)$, Eq. (7) LHS. For this work, the radial functions were set to $\psi_j(\rho) = \text{tri}(\rho - j)$. The radial profiles are then evaluated on a Cartesian grid for discretization, as in [2], and normalised. Finally, they are multiplied with the SHs to create the kernel Eq. (7) RHS as

$$h_n^{i,o}(\rho) = \sum_{j=0}^{J} w_{n,j}^{i,o} \psi_j(\rho) \xrightarrow{Eq.~(4)} \kappa_{n,m}^{i,o} = \Big(\sum_{j=0}^{J} w_{n,j}^{i,o} \psi_j(\rho)\Big) Y_n^m(\theta, \phi), \tag{7}$$

where $w_{n,j}^{i,o}$ are the trainable parameters of the model. J corresponds to the number of radial profiles, i.e. half the kernel size. The indices i and o iterate over $[1, ..., C_{in}]$ and $[1, ..., C_{out}]$ representing the number of input and output channels of the layer. Each bispectral convolution is performed in four steps. First, the feature maps, i.e. SF maps, are computed as a convolution

$$\mathcal{F}_n^o(\boldsymbol{x}) = \sum_{i=1}^{C_{in}} [(y_i * \kappa_{n,m}^{i,o})(\boldsymbol{x})]_{m=-n}^{m=n}, \tag{8}$$

with y_i the i^{th} channel of the previous feature maps and $\kappa_n^{i,o}$ the kernel described in Eq. (7). The indices i and o iterate over all input and output channels. The second step is to compute Eq. (6), using the SF maps of each degree $\{\mathcal{F}_n^o(\boldsymbol{x}), \mathcal{F}_{n'}^o(\boldsymbol{x}), \mathcal{F}_l^o(\boldsymbol{x})\}$. We consider the LRI output feature maps computed at each layer via the multichannel bispectrum operator

$$\mathcal{G}_{n,n',l}^o\{\boldsymbol{y}\}(\boldsymbol{x}) = [\mathcal{F}_n^o(\boldsymbol{x}) \otimes \mathcal{F}_{n'}^o(\boldsymbol{x})] C_{nn'} \tilde{\mathcal{F}_l^o(\boldsymbol{x})}^{\dagger}, \tag{9}$$

where $\boldsymbol{y}(\boldsymbol{x}) = [y_1(\boldsymbol{x}), \ldots, y_{C_{in}}(\boldsymbol{x})]$ and $\mathcal{F}_n^o(\boldsymbol{x})$ is given by Eq. (8). Note that stricly speaking, the operator Eq. (9) is not a bispectrum operator in the sense of Eq. (6) due to the sum in Eq. (8). It, however, inherits the bispectrum operator's equivariance properties, which we demonstrate in Supplementary material 2.

A non-linearity, $\sigma(x) = \text{sign}(x)\log(1 + x)$, is applied to avoid vanishing and exploding gradients due to the bispectral feature maps sizes. Finally, a bias is added and a ReLU function is applied to the features maps before the final $1 \times 1 \times 1$ convolution to reduce the number of output channels to C_{out}.

3.4 Datasets

The evaluation was conducted on two datasets pre-processed using nnU-Net [7] pipeline. The first one is the HippoCampus (HC) segmentation task of the *Medical Segmentation Decathlon* dataset[1] [3]. This dataset consists of 260 3D Magnetic Resonance Imaging (MRI) volumes with two classes contoured, i.e. the HC head as well as the union of its body and tail. 80% of the data (208 images) are used for training and the remaining 20% (52 images) for testing. The training set is split in four folds with 75% for training (156 images) and the remaining 52 images for validation. A second dataset, the Airway Tree Modeling 2022 (ATM22)[2] [24, 13, 25, 27, 23] dataset was also selected as it may be more prone to rotation sensibility. Its main task is to segment the pulmonary airway tree on Computed Tomography (CT) images. The first pre-processing step is to resample images to have an isotropic sampling, discarding images with a spacing difference larger than 0.5, resulting in 288 usable images. 75% (215 images) are used for training while the remaining 25% (73 images) are selected for testing. A subset of 22 test images were randomly selected to limit computational load when applying test time image rotations required to evaluate robustness (see Section 3.6). The training set is split again in two folds with 156 training and 59 validation images. For both datasets, the test results are the averaged model's output of all folds producing a more robust estimate of the model performances.

3.5 Network Details

The network architecture was generated using nnU-Net framework [7]. The encoder path was composed of four modules each containing two convolutional layers with a kernel size of 3. Every layer was constituted of a convolution followed by batch normalisation and LeakyRELU, with a negative slope of 0.01. Depending on the model tested, i.e. either standard convolution, XEdgeConv or bispectral, the convolution type was selected accordingly in each of those layers. Between modules, a max pooling layer, with a stride and kernel of 2 was used to reduce the dimensionality. The decoder path comprised three modules preceded by a trilinear upsampling. Finally, the prediction was computed with a $1 \times 1 \times 1$ standard convolution and a softmax activation and binarized with argmax. For

[1] http://medicaldecathlon.com/, July 2024.
[2] https://atm22.grand-challenge.org/, July 2024.

HC, a patch size of 40^3 was used based on nnU-Net implementation [7] whereas for ATM22, a patch size of 56^3 was selected based on GPUs' memory limitation. Even though this patch size is rather low to segment the whole airway tree, our main interest is to investigate the rotation stability and not the absolute segmentation performance. Compared to other models, the number of base features of the Bispectral U-Net is reduced to eight to fit the available memory. The network was trained using Dice and cross-entropy losses on an NVIDIA V100 for HC[3] and A100 for ATM22[4]. The Bispectral U-Net used Adam optimization with a decaying learning rate starting at 1e-3. The maximum number of epochs was set to 100 for the HC and 50 for ATM22 as the models reached a plateau. Two other models were selected for comparison. nnU-Net [7] as a baseline and XEdgeConv [17], a state-of-the-art model regarding rotation stability in medical images. Both methods were trained with their default parameters, stochastic gradient descent and learning rate decay starting at 1e-2. PyTorch 2.1.1 was used for all models. All three models used the standard nnU-Net data augmentation during training, including rotation randomly sampled between $\pm 30°$ for each axis. However, as the nnU-Net's robustness to rotation solely relies on augmentation, a fourth model, referred to as nnU-Net Extended, was trained with an extended rotational augmentation range of $\pm 180°$ for each axis.

3.6 Metrics and Evaluation

The rotational robustness and stability of each model are assessed by feeding multiple rotations of every test image and comparing the results between rotations for two tests. The first, referred to as *performance robustness* test, shows the segmentation performance of the model via the standard Dice Similarity Coefficient (DSC) between the rotated image with the rotated ground truth. The second, referred to as the *rotational stability* test, directly evaluates the model's stability by comparing the prediction of each rotated image with the non-rotated prediction. First, the network is fed with multiple rotations of the same image. Then, the inverse rotation is applied to each output probabilities map before being compared to the non-rotated probability map. Finally, the Root Mean Square Error (RMSE) between them measures consistency across rotations, e.g. a perfectly equivariant network will have an RMSE of zero. A set of rotations must be selected as testing every angle is not feasible. The first set comprises the 24 right-angle rotations according to Euler intrinsic angles $zx'z''$. However, since patient orientation is often controlled in medical images, i.e. a brain MRI or lung CT are unlikely to be upside down, right-angle rotations could be irrelevant since simple pre-processing steps could align all images. A set of realistic rotations is created by uniformly sampling 13 spherical coordinates within a cone of 45 degrees angle, i.e. $\phi \in [-45°, 45°]$ and $\theta \in [0, 360°]$, referred to as *Cone*[5]. Only rotations along z and x' are tested as z'' rotations induce limited variations for the right-angle test.

[3] An epoch was computed in ≈ 15 minutes using up to 27Gb for a batch size of two.
[4] An epoch was computed in ≈ 45 minutes using up to 70Gb for a batch size of one.
[5] A spline interpolation of order three was used when executing those rotations.

4 Results

The right-angle rotations distributions are shown in Fig. 1 a) and c) while the *Cone* rotations distributions are shown in Fig. 1 b) and d). The classic nnU-Net is also included as the training range of rotations includes the testing rotations. The means of each distribution are reported in Table 1. Qualitative comparisons

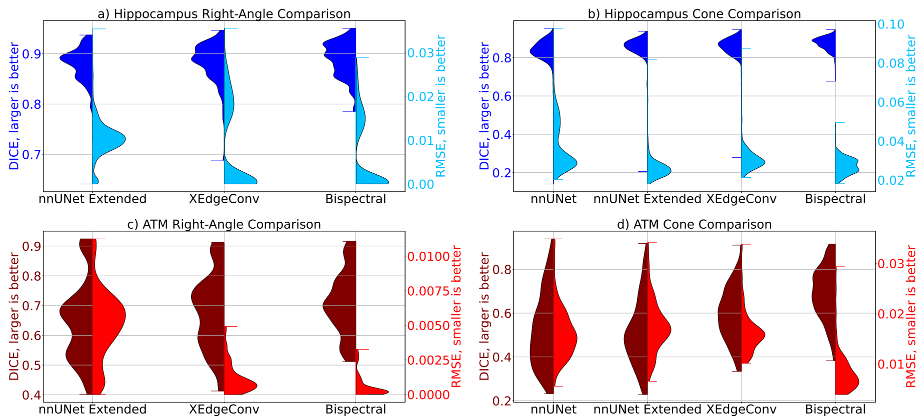

Fig. 1. DICE performance robustness of each model is shown on the darker side of each violin plot while the lighter side shows the RMSE rotation stability. The violin plots are created using the metrics score for every image and every rotation. Sub-figures a), b) are the HippoCampus (HC) dataset with both classes aggregated in a single distribution. Sub-fig. c), d) shows Airway Tree Modeling (ATM22) distributions.

are provided in Supplementary material 1. For right-angle rotations on the HC, all models except Bispectral U-Net and XEdgeConv second class ($p = 0.06$), are performing significantly different from each other in terms of performance robustness ($p \leq 0.015$). For rotational stability, Bispectral U-Net achieved significantly lower RMSE from all models ($p \leq 0.016$), while nnU-Net Extended and XEdgeConv are not significantly different ($p = 0.10$) for the second class. For *cone* rotations, all performance distributions except XEdgeConv and nnU-Net Extended (Class 1 $p = 0.67$, class 2 $p = 0.49$) are significantly different ($p = 0.02$). Regarding stability for realistic rotations, only the first class distributions of Bispectral U-Net and nnU-Net Extended ($p = 0.25$) are not significantly different. For extreme rotations on ATM22, nnU-Net Extended and XEdgeConv show no significant difference in performance ($p = 0.058$), but the Bispectral U-Net differs significantly from both (nnU-Net Extended $p = 0.0005$ and XEdgeConv $p = 0.0055$). All RMSE distributions have significantly different null p-values. For moderate rotations, only nnU-Net and nnU-Net Extended performance distributions are not different ($p = 0.39$). For rotational stability, only XEdgeConv and nnU-Net Extended are not significantly different ($p = 0.82$).

Table 1. HC and ATM22 performance and stability metric means for the experiments with either 24 right-angle or 13 *Cone* rotations. Best results are highlighted in bold. Note that nnUNet results are not included in the table as they are worse than its extended version and compromise the table's readability. For the HC, both class's performances are regrouped in a single distribution.

Model	performance robustness (DSC)				rotational stability (RMSE)			
	Right-Angles		Cone		Right-Angles		Cone	
	HC	ATM22	HC	ATM22	HC	ATM22	HC	ATM22
Ext. nnU-Net	87.88%	64.93%	85.17%	50.97%	10.23e-3	4.96e-3	**2.64e-4**	1.74e-2
XEdge	88.38%	**67.42%**	85.21%	58.95%	8.13e-3	1.18e-3	3.09e-4	1.61e-2
Ours	**88.93%**	64.40%	**87.21%**	**68.89%**	**6.74e-3**	**0.75e-3**	2.72e-4	**0.85e-2**

5 Discussions and Conclusions

In this work, we investigated segmentation network stability w.r.t rotations. This question has seen limited investigations even though medical images contain complex structures in a wide range of rotations. We first investigated nnU-Net stability to right-angle rotations. Note that the nnU-Net was trained with a wider range of rotations than the default ($\pm 180°$ instead of the basic $\pm 30°$). We then evaluated the same rotations on two RI networks trained without the extended rotation augmentation. One state-of-the-art approach, XEdgeConv, and the proposed 3D Bispectral U-Net based on its 2D alternative [12]. When comparing the three networks for extreme rotations, clear benefits from the invariant networks can be observed w.r.t rotation stability while preserving the segmentation performance. Both invariant networks share very close performances with slightly better stability for the Bispectral U-Net. However, for smaller and more realistic rotations (i.e. *Cone*), the standard and extended nnU-Nets showed moderate performance robustness and rotational stability, with potentially serious clinical consequences. In this context, the Bispectral U-Net achieved significantly higher performance robustness and rotational stability for both datasets, even more so for the ATM22 dataset.

When looking at the prediction confidence of different models, available in the Supplementary material 1, both nnUNet models have much more voxels with lower confidence, i.e. greener, compared to other models. In addition, in some extreme cases, the models even detect the positive class outside of the main structure. When looking at right-angle rotations, the confidence of each model is higher, except for the normal nnUNet.

The distinct observed trends between extreme and *Cone* rotations could be due to (i) more complex directional structures in ATM22 being more sensitive to smaller rotations and (ii) the induced effect of the rotations on the images is less diverse with right-angle rotations as they contain trivial symmetries. The difference with XEdgeConv can be explained as they use a maximum pooling across neighbouring voxels, which could lose information for smaller rotations.

Our tests also show that the nnU-Net Extended is very competitive with RI networks regarding rotation stability for extreme rotations.

Our network shows several limitations. The bispectrum coefficients require large memory to be computed during training, as well as testing, thus limiting maximum decomposition degree, kernel size and patch size. However, such a bispectral model could be more than beneficial for datasets with smaller volumes and further computational optimisation. In addition, the training is much longer than other approaches even if the number of parameters is significantly lower as the bispectrum generates large matrices. Similarly, during inference, the prediction also requires an important memory and a rather long time to compute all the bispectrum coefficients. Nevertheless, we could successfully apply it to two real-world clinical applications. Finally, when subject to right-angle rotations, the Bispectral U-Net did not show a generalised performance increase compared to the state-of-the-art model. Further improvements would be to conduct a hyper-parameters search and test the network with larger kernel sizes and maximal degrees to extract more information. Similarly, the effect of smaller rotations should be more thoroughly investigated to know when a Bispectral U-Net would be preferred.

Acknowledgments. This work was partially funded by the Swiss National Science Foundation (SNSF) with the projects 205320_219430 and 205320_179069, the Swiss Cancer Research foundation with the project TARGET (KFS-5549-02-2022-R), and the Hasler Foundation with the project MSxplain number 21042.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Andrearczyk, V., Depeursinge, A.: Rotational 3D texture classification using group equivariant CNNs. arXiv preprint arXiv:1810.06889 (2018)
2. Andrearczyk, V., Fageot, J., Oreiller, V., Montet, X., Depeursinge, A.: Local rotation invariance in 3D CNNs. Medical image analysis **65**, 101756 (2020)
3. Antonelli, M., Reinke, A., Bakas, S., Farahani, K., Kopp-Schneider, A., Landman, B.A., Litjens, G., Menze, B., Ronneberger, O., Summers, R.M., et al.: The Medical Segmentation Decathlon. Nature communications **13**(1), 4128 (2022)
4. Esteves, C., Allen-Blanchette, C., Makadia, A., Daniilidis, K.: Learning SO(3) Equivariant Representations with Spherical CNNs. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 52–68 (2018)
5. Fageot, J., Uhlmann, V., Püspöki, Z., Beck, B., Unser, M., Depeursinge, A.: Principled design and implementation of steerable detectors. IEEE Transactions on Image Processing **30**, 4465–4478 (2021)
6. Hadid, A.: The local binary pattern approach and its applications to face analysis. In: 2008 First Workshops on Image Processing Theory, Tools and Applications. pp. 1–9. IEEE (2008)
7. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods **18**(2), 203–211 (2021)
8. Jiang, M., Wu, Y., Zhao, T., Zhao, Z., Lu, C.: PointSIFT: A SIFT-like Network Module for 3D Point Cloud Semantic Segmentation. arXiv preprint arXiv:1807.00652 (2018)
9. Kakarala, R.: Completeness of bispectrum on compact groups. arXiv preprint arXiv:0902.0196 **1** (2009)
10. Nguyen, T., Hua, B.S., Le, N.: 3D-UCaps: 3D Capsules Unet for Volumetric Image Segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24. pp. 548–558. Springer (2021)
11. Oreiller, V., Andrearczyk, V., Fageot, J., Prior, J.O., Depeursinge, A.: 3D solid spherical bispectrum CNNs for biomedical texture analysis. arXiv preprint arXiv:2004.13371 (2020)
12. Oreiller, V., Fageot, J., Andrearczyk, V., Prior, J.O., Depeursinge, A.: Robust Multi-Organ Nucleus Segmentation Using a Locally Rotation Invariant Bispectral U-Net. In: International Conference on Medical Imaging with Deep Learning. pp. 929–943. PMLR (2022)
13. Qin, Y., Chen, M., Zheng, H., Gu, Y., Shen, M., Yang, J., Huang, X., Zhu, Y.M., Yang, G.Z.: AirwayNet: A Voxel-Connectivity Aware Approach for Accurate Airway Segmentation Using Convolutional Neural Networks. In: International conference on medical image computing and computer-assisted intervention. pp. 212–220. Springer (2019)
14. Rao, Y., Lu, J., Zhou, J.: Spherical Fractal Convolutional Neural Networks for Point Cloud Recognition. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 452–460 (2019)
15. Sun, X., Lian, Z., Xiao, J.: SRINet: Learning Strictly Rotation-Invariant Representations for Point Cloud Classification and Segmentation. In: Proceedings of the 27th ACM international conference on multimedia. pp. 980–988 (2019)
16. Unser, M., Chenouard, N.: A unifying parametric framework for 2D steerable wavelet transforms. SIAM Journal on Imaging Sciences **6**(1), 102–135 (2013)

17. Weihsbach, C., Hansen, L., Heinrich, M.: XEdgeConv: Leveraging graph convolutions for efficient, permutation- and rotation-invariant dense 3D medical image segmentation. In: Geometric Deep Learning in Medical Image Analysis. pp. 61–71. PMLR (2022)
18. Weiler, M., Geiger, M., Welling, M., Boomsma, W., Cohen, T.S.: 3D Steerable CNNs: Learning rotationally equivariant features in volumetric data. Advances in Neural Information Processing Systems **31** (2018)
19. Weiler, M., Hamprecht, F.A., Storath, M.: Learning Steerable Filters for Rotation Equivariant CNNs. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 849–858 (2018)
20. Winkels, M., Cohen, T.S.: Pulmonary nodule detection in CT scans with equivariant CNNs. Medical image analysis **55**, 15–26 (2019)
21. Worrall, D., Brostow, G.: CubeNet: Equivariance to 3d rotation and translation. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 567–584 (2018)
22. Worrall, D.E., Garbin, S.J., Turmukhambetov, D., Brostow, G.J.: Harmonic networks: Deep translation and rotation equivariance. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 5028–5037 (2017)
23. Yu, W., Zheng, H., Zhang, M., Zhang, H., Sun, J., Yang, J.: BREAK: Bronchi Reconstruction by gEodesic transformation And sKeleton embedding. In: 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI). pp. 1–5. IEEE (2022)
24. Zhang, M., Wu, Y., Zhang, H., Qin, Y., Zheng, H., Tang, W., Arnold, C., Pei, C., Yu, P., Nan, Y., et al.: Multi-site, Multi-domain Airway Tree Modeling. Medical Image Analysis **90**, 102957 (2023)
25. Zhang, M., Zhang, H., Yang, G.Z., Gu, Y.: CFDA: Collaborative Feature Disentanglement and Augmentation for Pulmonary Airway Tree Modeling of COVID-19 CTs. In: International conference on medical image computing and computer-assisted intervention. pp. 506–516. Springer (2022)
26. Zhang, Z., Hua, B.S., Rosen, D.W., Yeung, S.K.: Rotation Invariant Convolutions for 3D Point Clouds Deep Learning. In: 2019 International conference on 3d vision (3DV). pp. 204–213. IEEE (2019)
27. Zheng, H., Qin, Y., Gu, Y., Xie, F., Yang, J., Sun, J., Yang, G.Z.: Alleviating Class-wise Gradient Imbalance for Pulmonary Airway Segmentation. IEEE transactions on medical imaging **40**(9), 2452–2462 (2021)