# Regular SE(3) Group Convolutions for Volumetric Medical Image Analysis

Thijs P. Kuipers[0009−0007−7198−2856] and Erik J. Bekkers[0000−0003−4418−2160]

Amsterdam Machine Learning Lab, Informatics Institute, University of Amsterdam

**Abstract.** Regular group convolutional neural networks (G-CNNs) have been shown to increase model performance and improve equivariance to different geometrical symmetries. This work addresses the problem of SE(3), i.e., roto-translation equivariance, on volumetric data. Volumetric image data is prevalent in many medical settings. Motivated by the recent work on separable group convolutions, we devise a SE(3) group convolution kernel separated into a continuous SO(3) (rotation) kernel and a spatial kernel. We approximate equivariance to the continuous setting by sampling uniform SO(3) grids. Our continuous SO(3) kernel is parameterized via RBF interpolation on similarly uniform grids. We demonstrate the advantages of our approach in volumetric medical image analysis. Our SE(3) equivariant models consistently outperform CNNs and regular discrete G-CNNs on challenging medical classification tasks and show significantly improved generalization capabilities. Our approach achieves up to a 16.5% gain in accuracy over regular CNNs.

**Keywords:** geometric deep learning · equivariance · group convolution · SE(3) · volumetric data.

## 1 Introduction

Invariance to geometrical transformations has been long sought-after in the field of machine learning [6, 12]. The strength of equipping models with inductive biases to these transformations was shown by the introduction of convolutional neural networks (CNNs) [13]. Following the success of CNNs, [7] generalized the convolution operator to commute with geometric transformation groups other than translations, introducing group-convolutional neural networks (G-CNNs), which have been shown to outperform conventional CNNs [19, 20, 2, 11].

Early G-CNNs were mainly concerned with operating on 2D inputs. With the increase in computing power, G-CNNs were extended to 3D G-CNNs. Volumetric data is prevalent in many medical settings, such as in analyzing protein structures [15] and medical image analysis [5, 21, 26, 27]. Equivariance to symmetries such as scaling and rotations is essential as these symmetries often naturally occur in volumetric data. Equivariance to the group of 3D rotations, SO(3), remains a non-trivial challenge for current approaches due to its complex structure and non-commutative properties [19].

An important consideration regarding 3D convolutions that operate on volumetric data is overfitting. Due to the dense geometric structure in volumetric data and the high parameter count in 3D convolution kernels, 3D convolutions are highly susceptible to overfitting [14]. G-CNNs have been shown to improve generalization compared to CNNs [23, 20]. However, G-CNNs operating on discrete subgroups can exhibit overfitting to these discrete subgroups [3], failing to obtain equivariance on the full continuous group. This effect is amplified for 3D G-CNNs, limiting their improved generalization capabilities.

**Contributions.** In this work, we introduce regular continuous group convolutions equivariant to SE(3), the group of roto-translations. Motivated by the work on separable group convolutions [11], we separate our SE(3) kernel in a continuous SO(3) and a spatial convolution kernel. We randomly sample discrete equidistant SO(3) grids to approximate the continuous group integral. The continuous SO(3) kernels are parameterized via radial basis function (RBF) interpolation on a similarly equidistantly spaced grid. We evaluate our method on several challenging volumetric medical image classification tasks from the MedM-NIST [26, 27] dataset. Our approach consistently outperforms regular CNNs and discrete SE(3) subgroup equivariant G-CNNs and shows significantly improved generalization capabilities. To this end, this work offers the following contributions.

1. We introduce separable regular SE(3) equivariant group convolutions that generalize to the continuous setting using RBF interpolation and randomly sampling equidistant SO(3) grids.
2. We show the advantages of our approach on volumetric medical image classification tasks over regular CNNs and discrete subgroup equivariant G-CNNs, achieving up to a 16.5% gain in accuracy over regular CNNs.
3. Our approach generalizes to $SE(n)$ and requires no additional hyperparameters beyond setting the kernel and sample resolutions.
4. We publish our SE(3) equivariant group convolutions and codebase for designing custom regular group convolutions as a Python package.[1]

**Paper Outline.** The remainder of this paper is structured as follows. Section 2 provides an overview of current research in group convolutions. Section 3 introduces the group convolution theory and presents our approach to SE(3) equivariant group convolutions. Section 4 presents our experiments and an evaluation of our results. We give our concluding remarks in Section 5.

## 2    Literature Overview

Since the introduction of the group convolutional neural network (G-CNN), research in G-CNNs has grown in popularity due to their improved performance

---

[1] Our codebase can be accessed at: https://github.com/ThijsKuipers1995/gconv.

and equivariant properties over regular CNNs. Work on G-CNNs operating on volumetric image data has primarily been focused on the 3D roto-translation group SE(3) [22, 20, 19]. CubeNet was the first introduced 3D G-CNN, operating on the rotational symmetries of the 3D cube [22]. The approach presented in [20] similarly works with discrete subsets of SE(3). These approaches are not fully equivariant to SE(3). Steerable 3D G-CNNs construct kernels through a linear combination of spherical harmonic functions, obtaining full SE(3) equivariance [19]. Other approaches that are fully SE(3) equivariant are the Tensor-Field-Network [18] and N-Body networks [17]. However, these operate on point clouds instead of 3D volumes.

## 3    Separable SE($n$) Equivariant Group Convolutions

This work introduces separable SE(3) equivariant group convolutions. However, our framework generalizes to SE($n$). Hence, we will describe it as such. Section 3.1 presents a brief overview of the regular SE($n$) group convolution. Section 3.2 introduces our approach for applying this formulation to the continuous domain.

### 3.1    Regular Group Convolutions

The traditional convolution operates on spatial signals, i.e., signals defined on $\mathbb{R}^n$. Intuitively, one signal (the kernel) is slid across the other signal. That is, a translation is applied to the kernel. From a group-theoretic perspective, this can be viewed as performing the group action from the translation group. The convolution operator can then be formulated in terms of the group action. By commuting to a group action, the group convolution produces an output signal that is equivariant to the transformation imposed by the corresponding group.

**The SE($n$) Group Convolution Operator.** Instead of operating on signals defined on $\mathbb{R}^n$, SE($n$)-convolutions operate on signals defined on the group $SE(n) = \mathbb{R}^n \rtimes SO(n)$. Given an $n$-dimensional rotation matrix $\mathbf{R}$, and SE(3)-signals $f$ and $k$, the SE($n$) group convolution is defined as follows:

$$(f *_{\text{group}} k)(\mathbf{x}, \mathbf{R}) = \int_{\mathbb{R}^n} \int_{SO(n)} k\left(\mathbf{R}^{-1}(\tilde{\mathbf{x}} - \mathbf{x}), \mathbf{R}^{-1}\tilde{\mathbf{R}}\right) f(\tilde{\mathbf{x}}, \tilde{\mathbf{R}}) \mathrm{d}\tilde{\mathbf{R}} \mathrm{d}\tilde{\mathbf{x}}. \quad (1)$$

**The Lifting Convolution Operator.** Input data is usually not defined on SE($n$). Volumetric images are defined on $\mathbb{R}^3$. Hence, the input signal should be lifted to SE($n$). This is achieved via a lifting convolution, which accepts a signal $f$ defined on $\mathbb{R}^n$ and applies a kernel $k$ defined on SO($n$), resulting in an output signal on SE($n$). The lifting convolution is defined as follows:

$$(f *_{\text{lifting}} k)(\mathbf{x}) = \int_{\mathbb{R}^n} k(\mathbf{R}^{-1}(\tilde{\mathbf{x}} - \mathbf{x})) f(\tilde{\mathbf{x}}) \mathrm{d}\tilde{\mathbf{x}}. \quad (2)$$

The lifting convolution can be seen as a specific case of group convolution where the input is implicitly defined on the identity group element.

### 3.2    Separable SE($n$) Group Convolution

The group convolution in Equation 1 can be separated into a convolution over $\mathbf{R}$ followed by a convolution over $\mathbb{R}^n$ by assuming $k_{\mathrm{SE}(n)}(\mathbf{x}, \mathbf{R}) = k_{\mathrm{SO}(n)}(\mathbf{R})k_{\mathbb{R}^n}(\mathbf{x})$. This improves performance and significantly reduces computation time [11].

**The Separable SE($n$) Kernel.** Let $i$ and $o$ denote in the input and output channel indices, respectively. We separate the SE(3) kernel as follows:

$$k_{\mathrm{SE}(n)}^{io}(\mathbf{x}, \mathbf{R}) = k_{\mathrm{SO}(n)}^{io}(\mathbf{R})k_{\mathbb{R}^n}^{o}(\mathbf{x}). \tag{3}$$

Here, $k_{\mathrm{SO}(n)}$ performs the channel mixing, after which a depth-wise separable spatial convolution is performed. This choice of separation is not unique. The channel mixing could be separated from the SO($n$) kernel. However, this has been shown to hurt model performance [11].

**Discretizing the Continuous SO($n$) Integral.** The continuous group integral over SO($n$) in Equation 1 can be discretized by summing over a discrete SO($n$) grid. By randomly sampling the grid elements, the continuous group integral can be approximated [24]. However, randomly sampled kernels may not capture the entirety of the group manifold. This will result in a noisy estimate. Therefore, we constrain our grids to be uniform, i.e., grid elements are spaced equidistantly. Similarly to the authors of [2], we use a repulsion model to generate SO($n$) grids of arbitrary resolution.

**Continuous SO($n$) Kernel with Radial Basis Function Interpolation.** The continuous SO($n$) kernel is parameterized via a similarly discrete SO($n$) uniform grid. Each grid element $\mathbf{R}_i$ has corresponding learnable parameters $\mathbf{k}_i$. We use radial basis function (RBF) interpolation to evaluate sampled grid elements. Given a grid of resolution $N$, the continuous kernel $k_{\mathrm{SO}(n)}$ is evaluated for any $\mathbf{R}$ as:

$$k_{\mathrm{SO}(n)}(\mathbf{R}) = \sum_{i=1}^{N} a_{d,\psi}(\mathbf{R}, \mathbf{R}_i)\mathbf{k}_i. \tag{4}$$

Here, $a_{d,\psi}(\mathbf{R}, \mathbf{R}_i)$ represents the RBF interpolation coefficient of $\mathbf{R}$ corresponding to $\mathbf{R}_i$ obtained using Gaussian RBF $\psi$ and Riemannian distance $d$. The uniformity constraint on the grid allows us to scale $\psi$ to the grid resolution dynamically. This ensures that the kernel is smooth and makes our approach hyperparameter-free.

## 4    Experiments and Evaluation

In this section, we present our results and evaluation. Section 4.1 introduces our experimental setup. Our results on MedMNIST are presented in Sections 4.2 and 4.3. Section 4.4 offers a deeper look into the generalization performance. Directions for future work based on our results are suggested in Section 4.5.

### 4.1 Evaluation Methodology

From here on, we refer to our approach as the SE(3)-CNN. We evaluate the SE(3)-CNNs for different group kernel resolutions. The sample and kernel resolutions are kept equal. We use a regular CNN as our baseline model. We also compare discrete SE(3) subgroup equivariant G-CNNs. K-CNN and T-CNN are equivariant to the 180 and 90-degree rotational symmetries, containing 4 and 12 group elements, respectively.

All models use the same ResNet [8] architecture consisting of an initial convolution layer, two residual blocks with two convolution layers each, and a final linear classification layer. Batch normalization is applied after the first convolution layer. In the residual blocks, we use instance normalization instead. Max spatial pooling with a resolution of $2 \times 2 \times 2$ is applied after the first residual block. Global pooling is applied before the final linear layer to produce SE(3) invariant feature descriptors. The first layer maps to 32 channels. The residual blocks map to 32 and 64 channels, respectively. For the G-CNNs, the first convolution layer is a lifting convolution, and the remainders are group convolutions. All spatial kernels have a resolution of $7 \times 7 \times 7$. Increasing the group kernel resolution increases the number of parameters. Hence, a second baseline CNN with twice the number of channels is included. The number of parameters of the models is presented in Table 1.

We evaluate the degree of SE(3) equivariance obtained by the SE(3)-CNNs on OrganMNIST3D [4, 25] and rotated OrganMNIST3D. For rotated Organ-MNIST3D, samples in the test set are randomly rotated. We further evaluate FractureMNIST3D [9], NoduleMNIST3D [1], AdrenalMNIST3D [27], and SynapseMNIST3D [27] from the MedMNIST dataset [26, 27]. These volumetric image datasets form an interesting benchmark for SE(3) equivariant methods, as they naturally contain both isotropic and anisotropic features. All input data has a single channel with a resolution of $28 \times 28 \times 28$. Each model is trained for 100 epochs with a batch size of 32 and a learning rate of $1 \times 10^{-4}$ using the Adam [10] optimizer on an NVIDIA A100 GPU. The results are averaged over three training runs with differing seeds.

### 4.2 SE(3) Equivariance Performance

Table 1 shows the accuracies and accuracy drops obtained by the evaluated models on the OrganMNIST3D test set and rotated test set. The decrease in accuracy is calculated as the percentage of the difference between the test scores on the test set and the rotated test set. Both baselines suffer from a high accuracy drop. This is expected, as these models are not equivariant to SE(3). K-CNN and T-CNN fare better. Due to its higher SO(3) kernel resolution, T-CNN outperforms K-CNN. However, these methods do not generalize to the SE(3) group. The SE(3)-CNNs obtain significantly lower drops in accuracy, showing their improved generalization to SE(3). The SE(3)-CNNs at sample resolutions of 12 and 16 also reach higher accuracies than both baseline models. As the sample resolution increases, performance on the standard test set shows a more pronounced

Table 1: Number of parameters, computational performance, accuracies, and drop in accuracy on test scores between OrganMNIST3D and rotated OrganMNIST3D. Computational performance is measured during training. The highest accuracy and lowest error are shown in **bold**.

| Model | Baseline | Baseline big | SE(3)-CNN | | | | | K-CNN | T-CNN |
|---|---|---|---|---|---|---|---|---|---|
| Sample res. | - | - | 4 | 6 | 8 | 12 | 16 | 4 | 12 |
| # parameters | 89k | 200k | 80k | 96k | 111k | 142k | 172k | 80k | 142k |
| Seconds/epoch | 3.37 | 5.83 | 9.67 | 14.43 | 18.48 | 27.31 | 36.37 | 9.70 | 27.75 |
| Memory (GB) | 3.34 | 4,80 | 6.89 | 9.188 | 12.38 | 15.75 | 20.86 | 6.70 | 15.52 |
| Accuracy | 0.545 | 0.697 | 0.655 | 0.681 | 0.688 | 0.703 | 0.698 | 0.633 | **0.722** |
| Rotated Acc. | 0.207 | 0.264 | 0.581 | 0.593 | 0.592 | 0.608 | **0.628** | 0.327 | 0.511 |
| Drop in Acc. % | 62.15 | 62.09 | 11.26 | 12.91 | 14.07 | 13.60 | **10.09** | 48.27 | 29.27 |

increase than on the rotated test set. This results in a slight increase in accuracy drop. At a sample resolution of 16, accuracy on the standard test set decreases while the highest accuracy is obtained on the rotated test set. A high degree of SE(3) equivariance seems disadvantageous on OrganMNIST3D. This would also explain why T-CNN achieved the highest accuracy on the standard test set, as this model generalizes less to SE(3). OrganMNIST3D contains samples aligned to the abdominal window, resulting in high isotropy. This reduces the advantages of SE(3) equivariance.

### 4.3   Performance on MedMNIST

The accuracies obtained on FractureMNIST3D, NoduleMNIST3D, AdrenalMNIST3D, and SynapseMNIST3D are reported in Table 2. The SE(3)-CNNs obtain the highest accuracies on all datasets. On FreactureMNIST3D, the highest accuracy is achieved by the SE(3)-CNN (12). Both K-CNN and T-CNN achieve an accuracy very similar to the baseline models. The baseline-big model slightly outperforms both K-CNN and T-CNN. On NoduleMNIST3D, SE(3)-CNN (6) and SE(3)-CNN (8) achieve the highest accuracy, with SE(3)-CNN (12) performing only slightly lower. K-CNN and T-CNN outperform both baseline models. On AdrenalMNIST3D, the differences in accuracy between all models are the lowest. SE(3)-CNN (16) obtains the highest accuracy, whereas the baseline model obtains the lowest. The baseline-big model outperforms K-CNN and TCNN. On SynapseMNIST3D, we again observe a significant difference in performance between the SE(3)-CNNs and the other models. SE(3)-CNN (6) obtained the highest performance. T-CNN outperforms K-CNN and both baseline models. However, the baseline-big model outperforms K-CNN. On NoduleMNIST3D and AdrenalMNIST3D, only a slight performance gain is achieved by the SE(3)-CNNs. This is likely due to the isotropy of the samples in these datasets. In these cases, SE(3) equivariance is less beneficial. In contrast, FractureMNIST3D

Table 2: Accuracies and SE(3)-CNN performance gain in percentage points (p.p.) over the CNN baselines on FractureMNIST3D, NoduleMNIST3D, AdrenalM-NIST3D, and SynapseMNIST3D. Sample resolution in parenthesis behind the model name. Standard deviation in parenthesis behind the accuracies. The highest accuracy is indicated in **bold**.

| Model | Fracture | Nodule | Adrenal | Synapse |
|---|---|---|---|---|
| Baseline | 0.450 ($\pm$0.033) | 0.834 ($\pm$0.019) | 0.780 ($\pm$0.006) | 0.694 ($\pm$0.006) |
| Baseline-big | 0.499 ($\pm$0.032) | 0.846 ($\pm$0.010) | 0.806 ($\pm$0.022) | 0.720 ($\pm$0.019) |
| SE(3)-CNN (4) | 0.588 ($\pm$0.029) | 0.869 ($\pm$0.008) | 0.800 ($\pm$0.006) | 0.865 ($\pm$0.010) |
| SE(3)-CNN (6) | 0.617 ($\pm$0.013) | **0.875** ($\pm$0.008) | 0.804 ($\pm$0.002) | 0.870 ($\pm$0.008) |
| SE(3)-CNN (8) | 0.615 ($\pm$0.002) | **0.875** ($\pm$0.014) | 0.815 ($\pm$0.015) | **0.885** ($\pm$0.007) |
| SE(3)-CNN (12) | **0.621** ($\pm$0.002) | 0.873 ($\pm$0.005) | 0.814 ($\pm$0.010) | 0.858 ($\pm$0.028) |
| SE(3)-CNN (16) | 0.604 ($\pm$0.012) | 0.858 ($\pm$0.011) | **0.832** ($\pm$0.005) | 0.869 ($\pm$0.023) |
| K-CNN (4) | 0.486 ($\pm$0.012) | 0.859 ($\pm$0.013) | 0.798 ($\pm$0.011) | 0.709 ($\pm$0.009) |
| T-CNN (12) | 0.490 ($\pm$0.036) | 0.862 ($\pm$0.006) | 0.800 ($\pm$0.011) | 0.777 ($\pm$0.021) |
| Gain in p.p. | 12.2 | 2.9 | 2.6 | 16.5 |



(a) Train scores on SynapseMNIST3D.          (b) Test scores on SynapseMNIST3D.

Fig. 1: Accuracy scores of the baseline models, the SE(3)-CNN (4) and (16) models, and the discrete SE(3) subgroup models on (a) the train set and (b) the test set of SynapseMNIST3D.

and SynapseMNIST3D are more anisotropic, resulting in significant performance gains of up to 16.5%.

## 4.4   Model Generalization

The scores obtained on both the train set and test sets of SynapseMNIST3D in Figures 1a and 1b, respectively. We observed similar behavior on all datasets. Figure 1 shows a stark difference between the SE(3)-CNNs and the other models. The baselines and K-CNN and T-CNN converge after a few epochs on the

train set. SE(3)-CNN (16) requires all 100 epochs to converge on the train set. SE(3)-CNN (4) does not converge within 100 epochs. This improvement during the training window is also observed in the test scores. This suggests that SE(3)-CNNs suffer less from overfitting, which results in improved model generalization. K-CNN and T-CNN behave similarly to the baselines. We hypothesize that this results from the weight-sharing that occurs during the RBF interpolation. We do observe a higher variance in scores of the SE(3)-CNNs, which we attribute to the random nature of the convolution kernels.

### 4.5   Future Work

With an increase in the sample resolution, a better approximation to SE(3) equivariance is achieved. However, we observe that this does not necessarily improve model performance, e.g., in the case of isotropic features. This could indicate that the equivariance constraint is too strict. We could extend our approach to learn *partial* equivariance. Rather than sampling on the entire SO(3) manifold, each group convolution layer could learn sampling in specific regions. This suggests a compelling extension of our work, as learning partial invariance has shown to increase model performance [16].

## 5   Conclusion

This work proposed an SE(3) equivariant separable G-CNN. Equivariance is achieved by sampling uniform kernels on a continuous function over SO(3) using RBF interpolation. Our approach requires no additional hyper-parameters compared to CNNs. Hence, our SE(3) equivariant layers can replace regular convolution layers. Our approach consistently outperforms CNNs and discrete subgroup equivariant G-CNNs on challenging medical image classification tasks. We showed that 3D CNNs and discrete subgroup equivariant G-CNNs suffer from overfitting. We showed significantly improved generalization capabilities of our approach. In conclusion, we have demonstrated the advantages of equivariant methods in medical image analysis that naturally deal with rotation symmetries. The simplicity of our approach increases the accessibility of these methods, making them available to a broader audience.

## References

[1]   Samuel G Armato III et al. "The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans". In: *Medical physics* 38.2 (2011), pp. 915–931.

[2]   Erik J Bekkers. "B-Spline {CNN}s on Lie groups". In: *International Conference on Learning Representations*. 2020. URL: https://openreview.net/forum?id=H1gBhkBFDH.

[3]   Erik J Bekkers et al. "Roto-translation covariant convolutional networks for medical image analysis". In: *International conference on medical image computing and computer-assisted intervention*. Springer. 2018, pp. 440–448.

[4]   Patrick Bilic et al. "The liver tumor segmentation benchmark (lits)". In: *Medical Image Analysis* 84 (2023), p. 102680.

[5]   Luca Bogoni et al. "Impact of a computer-aided detection (CAD) system integrated into a picture archiving and communication system (PACS) on reader sensitivity and efficiency for the detection of lung nodules in thoracic CT exams". In: *Journal of digital imaging* 25.6 (2012), pp. 771–781.

[6]   Taco Cohen. "Learning transformation groups and their invariants". PhD thesis. PhD thesis, University of Amsterdam, 2013.

[7]   Taco Cohen and Max Welling. "Group equivariant convolutional networks". In: *International conference on machine learning*. PMLR. 2016, pp. 2990–2999.

[8]   Kaiming He et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[9]   Liang Jin et al. "Deep-learning-assisted detection and segmentation of rib fractures from CT scans: Development and validation of FracNet". In: *EBioMedicine* 62 (2020), p. 103106.

[10]  Diederik P Kingma and Jimmy Ba. "Adam: A method for stochastic optimization". In: *3rd International Conference for Learning Representations*. 2015.

[11]  David M Knigge, David W Romero, and Erik J Bekkers. "Exploiting redundancy: Separable group convolutional networks on lie groups". In: *International Conference on Machine Learning*. PMLR. 2022, pp. 11359–11386.

[12]  Risi Kondor and Shubhendu Trivedi. "On the generalization of equivariance and convolution in neural networks to the action of compact groups". In: *International Conference on Machine Learning*. PMLR. 2018, pp. 2747–2755.

[13]  Yann LeCun, Yoshua Bengio, et al. "Convolutional networks for images, speech, and time series". In: *The handbook of brain theory and neural networks* 3361.10 (1995), p. 1995.

[14]  Charles R Qi et al. "Volumetric and multi-view cnns for object classification on 3d data". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 5648–5656.

[15]  Nicolas Renaud et al. "DeepRank: a deep learning framework for data mining 3D protein-protein interfaces". In: *Nature communications* 12.1 (2021), pp. 1–8.

[16]  David W. Romero and Suhas Lohit. *Learning Equivariances and Partial Equivariances From Data*. 2022. URL: https://openreview.net/forum?id=jFfRcKVut98.

[17]  Laurent Sifre and Stéphane Mallat. "Rotation, scaling and deformation invariant scattering for texture discrimination". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2013, pp. 1233–1240.

[18]  Nathaniel Thomas et al. "Tensor field networks: Rotation-and translation-equivariant neural networks for 3d point clouds". In: *arXiv preprint arXiv:1802.08219* (2018).

[19]  Maurice Weiler et al. "3d steerable cnns: Learning rotationally equivariant features in volumetric data". In: *Advances in Neural Information Processing Systems* 31 (2018).

[20]  Marysia Winkels and Taco S Cohen. "3D G-CNNs for pulmonary nodule detection". In: *arXiv preprint arXiv:1804.04656* (2018).

[21]  Marysia Winkels and Taco S. Cohen. "3D G-CNNs for Pulmonary Nodule Detection". In: *Medical Imaging with Deep Learning*. 2018. URL: https://openreview.net/forum?id=H1sdHFiif.

[22]  Daniel Worrall and Gabriel Brostow. "Cubenet: Equivariance to 3d rotation and translation". In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018, pp. 567–584.

[23]  Daniel E Worrall et al. "Harmonic networks: Deep translation and rotation equivariance". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, pp. 5028–5037.

[24]  Wenxuan Wu, Zhongang Qi, and Li Fuxin. "Pointconv: Deep convolutional networks on 3d point clouds". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 9621–9630.

[25]  Xuanang Xu et al. "Efficient multiple organ localization in CT image using 3D region proposal network". In: *IEEE transactions on medical imaging* 38.8 (2019), pp. 1885–1898.

[26]  Jiancheng Yang, Rui Shi, and Bingbing Ni. "MedMNIST Classification Decathlon: A Lightweight AutoML Benchmark for Medical Image Analysis". In: *IEEE 18th International Symposium on Biomedical Imaging (ISBI)*. 2021, pp. 191–195.

[27]  Jiancheng Yang et al. "MedMNIST v2-A large-scale lightweight benchmark for 2D and 3D biomedical image classification". In: *Scientific Data* 10.1 (2023), p. 41.