

# Scale-Equivariant Deep Learning for 3D Data

Thomas Wimmer<sup>1</sup>, Vladimir Golkov<sup>1,2</sup>, Hoai Nam Dang<sup>3</sup>, Moritz Zaiss<sup>3</sup>,  
Andreas Maier<sup>3</sup>, and Daniel Cremers<sup>1,2</sup>

<sup>1</sup> Technical University of Munich

<sup>2</sup> Munich Center for Machine Learning

<sup>3</sup> Friedrich-Alexander University Erlangen-Nuremberg

thomas.m.wimmer@tum.de, vladimir.golkov@tum.de

**Abstract.** The ability of convolutional neural networks (CNNs) to recognize objects regardless of their position in the image is due to the translation-equivariance of the convolutional operation. Group-equivariant CNNs transfer this equivariance to other transformations of the input. Dealing appropriately with objects and object parts of different scale is challenging, and scale can vary for multiple reasons such as the underlying object size or the resolution of the imaging modality. In this paper, we propose a scale-equivariant convolutional network layer for three-dimensional data that guarantees scale-equivariance in 3D CNNs. Scale-equivariance lifts the burden of having to learn each possible scale separately, allowing the neural network to focus on higher-level learning goals, which leads to better results and better data-efficiency. We provide an overview of the theoretical foundations and scientific work on scale-equivariant neural networks in the two-dimensional domain. We then transfer the concepts from 2D to the three-dimensional space and create a scale-equivariant convolutional layer for 3D data. Using the proposed scale-equivariant layer, we create a scale-equivariant U-Net for medical image segmentation and compare it with a non-scale-equivariant baseline method. Our experiments demonstrate the effectiveness of the proposed method in achieving scale-equivariance for 3D medical image analysis.<sup>4</sup>

**Keywords:** Scale-Equivariance · Data Efficiency · Segmentation

## 1 Introduction

One of the greatest advantages of convolutional neural networks (CNNs) is their *equivariance* under spatial shifts (translations), i.e. a mathematically guaranteed ability to recognise objects and object parts at any positions they might appear at in images. Recent methods additionally provide equivariance under other transformations of the input such as rotation or scaling, in order to guarantee that features get detected well at different orientations or sizes. Scale-equivariance, i.e. guaranteed detection of features across various scales/sizes, is a beneficial

<sup>4</sup> We publish our code using PyTorch for further research and application:

<https://github.com/wimmerth/scale-equivariant-3d-convnet>

property of neural networks because an object or object part can have different sizes in different images, and because combining datasets with different image resolutions, for example in medical imaging, allows for constructing richer, more informative training datasets than if large parts of the data are left out. In practice, scale-equivariance often leads to better results [24, 25, 35].

In deep learning, data augmentation is often used as a means to improve the recognition of scaled features. However, data augmentation does not guarantee equivariance, it merely tries to approximate it, and imposes an additional burden on the neural network to learn each possible scale of each feature separately. In many cases, results of such learned equivariance are worse than with guaranteed equivariance [3, 6].

Therefore, intensive research has been conducted in recent years on the development of scale-equivariant neural networks [16, 25]. So far, these methods have been limited to the two-dimensional case. Since the detection of features of different scales is an important aspect in 3D machine learning tasks as well, in the present work we extend the concept of scale-equivariance to three dimensions. We propose novel scale-equivariant neural network layers for 3D data, including convolutions, normalization and pooling, that can be used instead of usual 3D neural network layers to achieve scale-equivariance. Medical image analysis, such as brain tumor segmentation, has been shown to benefit from aggressive data augmentation to approximate scale-equivariance [12] and as 3D data in many applications are available at various image resolutions, it is of high relevance to extend scale-equivariance to the 3D setting.

Scale-equivariant neural networks have been shown to outperform other neural networks especially in the low data regime [24, 35], which is particularly interesting for 3D applications such as MRI and datasets of rare diseases, as there is usually much less training data available than for example photographs in the two-dimensional case. In this paper, we first present the theoretical foundations for scale-equivariant convolutions and review previous works in this field. We lift the concept of scale-equivariant convolutions to the three-dimensional space and evaluate the performance of the proposed layers through a series of experiments based on the brain tumor segmentation task. To demonstrate the efficacy of our approach, we evaluate it against a baseline method based on standard convolutions which are only translation-equivariant.

## 2 Related Work

When image features can appear at a variety of sizes and locations, a primitive approach is to train a neural network that is neither equivariant nor is specifically designed to easily approximate equivariance. Instead, primitive methods tediously learn to approximate equivariance, and their training dataset must contain differently transformed features. That dataset might be obtained by data augmentation (random transformations) if the original dataset lacks such diversity. The additional burden of learning every feature at every possible size distracts

the training from the main learning goals, and in practice yields suboptimal results.

Better results can be achieved by methods that are designed to slightly facilitate the learning of approximate equivariance. An example is to use a branch of a neural network that decides how to transform (for example scale) the input before passing it to another neural-network branch, thus facilitating (but not guaranteeing) equivariance only for features that scale jointly but not independently [10, 13]. Another example are capsule networks, which encourage the separation of visual features and their poses (for example scale) in the latent representations by computing deeper features in ways that benefit from such a separation [23]. Other examples are scale-dependent pooling of latent features [33], or the addition of downsampling and/or upsampling branches in the network [4].

The best results are usually obtained by methods that mathematically guarantee equivariance. Examples include translation-equivariance that is achieved by using convolutional networks, equivariance under 2D rotations and translations [6, 8, 29], 3D rotations and translations [17, 21, 26, 28], or 2D scalings and translations [9, 14–16, 18, 22, 24, 25, 31, 32, 35]. Special cases of scale-equivariant neural networks use invariant rather than more generally equivariant layers not only as the last layer but also throughout [9, 15], which does not allow information about the relative scale of different features to be used to compute deeper features. Another special case keeps information about different scales separated until the last network layer [14, 16, 32], with a similar effect.

Neural networks achieve scale-equivariance by using differently scaled versions of convolutional filters (or, equivalently, of the feature maps). Recent works reduce discretization artifacts during scaling by representing filters using certain truncated bases, for example Hermite polynomials with Gaussian envelopes [25, 35], Gaussian derivatives [14, 16], or radial harmonics [9, 22]. In the medical domain, scale-equivariant neural networks have so far successfully been applied for histopathology image segmentation [34] and 2D MRI reconstruction [11].

The present paper extends the theory of scale-equivariant neural networks to 3D. We base our method on the work by [25] for the 2D case as it achieved state-of-the-art performance while being the most flexible approach.

### 3 Methods

In this section, we present the theory behind scale-equivariant convolutions and use them to propose a novel scale-equivariant layer for 3D data. We furthermore propose a novel scale-equivariant 3D U-Net that can be used in tasks like medical image segmentation.

A *group*  $(G, \cdot)$  is defined as a set  $G$  closed under an associative binary operation  $\cdot : G \times G \rightarrow G$ , with an identity element  $e \in G$ , and where every element has an inverse also in the set. A group  $(G, \cdot)$  is often simply denoted by  $G$ , with the binary operation  $\cdot$  implied. A mapping  $\Phi : \mathcal{X} \rightarrow \mathcal{Y}$  is *equivariant* under actions of the group  $G$  when

$$\Phi(L_g[f]) = L'_g[\Phi(f)] \quad \forall f \in \mathcal{X} \quad \forall g \in G, \quad (1)$$

where  $f$  is the input of  $\Phi$ , for example a medical image or latent feature map, and  $L_g$  and  $L'_g$  are the actions of  $g \in G$  on  $\mathcal{X}$  and  $\mathcal{Y}$ , respectively. When  $L'_g$  is the identity for all  $g \in G$ , then  $\Phi$  is called invariant under actions of the group  $G$ .

The *scaling group* can be defined as  $H = (\mathbb{R}_{>0}, \cdot)$ , i.e. consisting of positive scaling factors and with multiplication  $\cdot$  as the binary operation. A scale transformation  $L_s$  of a  $d$ -dimensional real-valued image  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  (for example a 3D MRI volume with  $d = 3$ ) by a scaling factor  $s \in H$  (i.e.  $s \in \mathbb{R}_{>0}$ ) can be formalized as a scale transformation of image coordinates  $x \in \mathbb{R}^d$  and can then be formulated as  $[L_s[f]](x) = f(s^{-1}x)$ . We refer to the transformation as upscale if  $s > 1$  and downscale if  $s < 1$ .

We want our neural network to be equivariant under scalings and translations. By combining the scaling group  $H$  with the group  $T$  of translations using the semi-direct product, we obtain the *group of scalings and translations*  $HT = \{(s, t) \mid s \in H, t \in T\}$  with  $s \in \mathbb{R}_{>0}$ ,  $t \in \mathbb{R}^d$ , sometimes denoted as  $HT \cong \mathbb{R}_{>0} \ltimes \mathbb{R}^d$  [31], with the identity element  $(1, 0)$ , the binary operation

$$(s_2, t_2) \cdot (s_1, t_1) = (s_2 s_1, s_2 t_1 + t_2), \quad (2)$$

and the inverse  $(s, t)^{-1} = (s^{-1}, -s^{-1}t)$ . As can be seen in this definition, the order of applying scaling and translation matters, i.e.  $(s, t) = (s, 0) \cdot (1, t) \neq (1, t) \cdot (s, 0)$ .

The group actions  $L_{(s,t)}$  of  $(s, t) \in HT$  on functions  $f : \mathbb{R}^d \rightarrow \mathbb{R}^C$  (input images) and on functions  $h : HT \rightarrow \mathbb{R}^C$  (latent feature maps) are defined as follows:

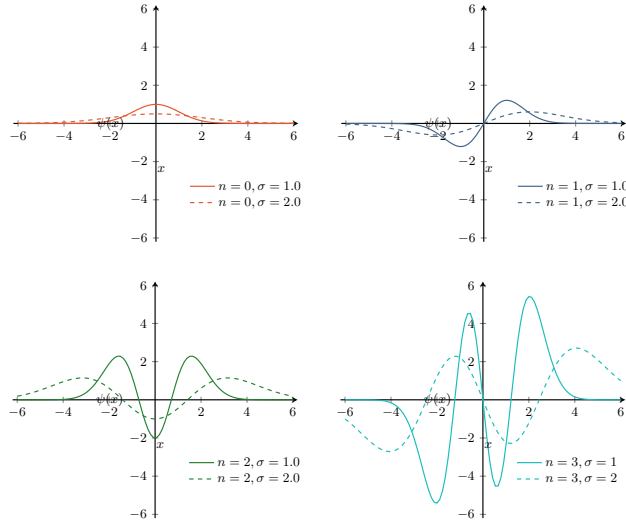
$$\begin{aligned} L_{(s,t)}[f](x) &= f(s^{-1}(x - t)), \\ L_{(s,t)}[h](s', t') &= h((s, t)^{-1}(s', t')) = h(s^{-1}s', s^{-1}(t' - t)). \end{aligned} \quad (3)$$

A *group-convolution*  $\star_G$  [6, 7, 25], using a locally compact group  $G$ , of a function  $f : X \rightarrow \mathbb{R}^C$ , for example a medical image (e.g.  $X = \mathbb{Z}^3$ ) or a latent feature map ( $X = G$ ), with a function  $\psi : X \rightarrow \mathbb{R}^C$  (e.g. a convolutional filter), where  $C \in \mathbb{N}$  is referred to as the number of *channels* of the input  $f$ , is defined as

$$[f \star_G \psi](g) = \int_{x \in X} f(x)[L_g[\psi]](x) d\mu(x), \quad (4)$$

where  $g \in G$  with its according group action  $L_g$ , and  $\mu(x)$  is the Haar measure. Thus, the output of a group convolution is a feature map defined on the group  $G$ . When that feature map is used as an input to a subsequent group convolution, the filters must also be defined on  $G$ , because the input and filter in a group convolution are defined on the same space.

Group-convolutional network layers generalize group convolutions from having several input channels (which we included in the definition of group convolutions above) to additionally having several output channels. This is achieved by performing a group convolution of the input with each of several filters, each yielding a separate output channel. Several channels are used in neural networks to disentangle different features. The filters in these layers are represented as a weighted sum of fixed (truncated-)basis functions. The weights in that sum are the trainable parameters of the layer.

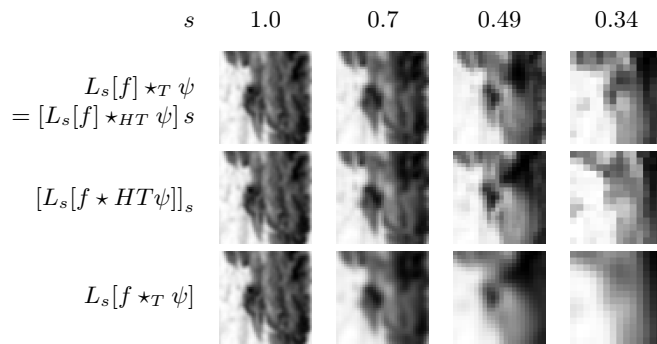


**Fig. 1.** Basis functions defined as Hermite polynomials  $H_n$  (of different degrees  $n$ ) with Gaussian envelopes, scaled with different scales  $\sigma$ . The kernel basis for three-dimensional scale-equivariant convolutions is formed from the multiplication of three oriented basis functions (oriented in the  $x$ ,  $y$ , and  $z$  directions) with equal or different degrees of Hermite polynomials.

Note that when the group  $G$  is the group  $T = (\mathbb{R}^d, +)$  of translations, with addition  $+$  as the binary operation, and the group action  $L_t$  of a translation  $t \in T$  is  $L_t[f](x) = f(x - t)$ , then Eq. (4) is the well-known standard convolution.

Group convolution using the (continuous) scaling group computes a value for each of the infinitely many elements of the group. In order to make the computational and output memory requirements finite, usually a discrete subgroup of the continuous group is used, and the subgroup is additionally truncated to a semigroup if it is still infinite [3, 25]. A discrete subgroup of the scaling group can be constructed as  $\{\dots, s^{-1}, 1, s^1, s^2, \dots\}$  with a base scale  $s$ , for example  $s = 0.9$ , and truncated for example to  $\{1, s, s^2, s^3\}$ . The truncation breaks the equivariance under scales beyond the truncation boundary, but is still locally correct, i.e. guarantees equivariance under discrete scales within the truncated discretized group of scales [31].

We propose scale-equivariant 3D convolutional layers by performing group-convolutions using the discretized truncated version of the group. For filters, truncated bases based on Hermite polynomials with a Gaussian envelope (see Fig. 1) were shown to work the best in practice for scale-equivariant deep learning due to reduced interpolation artifacts during scaling [25]. We generalize the construction of such bases to 3D. We multiply combinations of 1D Hermite polynomials of increasing order (with the maximum order being a hyperparameter) along each of the three dimensions and apply 3D Gaussian envelopes.



**Fig. 2.** Comparison of 2D slices of the 4D/3D output of scale-equivariant and non-scale-equivariant convolution. Scaling  $L_s$  of the input  $f$  followed by scale-equivariant convolution  $\star_{HT}$  with a filter  $\psi$  (first row) yields almost the same result as scale-equivariant convolution followed by scaling and selection  $[\cdot]_s$  of the respective “scale slice” along the scale dimension of the feature map on  $HT$  (second row). This shows that  $\star_{HT}$  is scale-equivariant apart from small interpolation artifacts. On the other hand, the ordinary (i.e. non-scale-equivariant) convolution  $\star_T$  followed by scaling (third row) yields a different result, and thus is not scale-equivariant. Note that due to the three-dimensional nature of the data, the 2D slices are not just scaled versions of each other but also strongly influenced by values of neighbouring slices when scaled, depending on the scale.

We used replicate-padding along the “scale dimension” of feature maps on the group because this technique is a good compromise between the truncation of the group and the imprecision of equivariance due to padding [35]. We also create additional scale-equivariant layers, namely the addition of bias terms, pointwise nonlinearities, max- and average-pooling over subgroups (e.g. the group of scalings), and normalization (batch normalization, instance normalization) for the 3D setting, analogously to the 2D setting [25]. Neural networks consisting of combinations of these layers are equivariant as well [6]. For segmentation tasks, the last layer of a scale-equivariant network is a pooling layer over the scale-dimension.

Finally, we propose scale-equivariant transposed convolutions (upconvolutions). Transposed convolutions have been shown to be equivariant in the group-convolution setting in general but have so far only been used in the rotation-equivariant setting [30]. Transposed convolutions can be helpful in creating CNN architectures for medical image segmentation, such as the U-Net [5].

## 4 Experiments

### 4.1 Experimental Setup

We perform brain tumor segmentation on the BraTS 2020 dataset. The inputs are four MRI contrasts (T1w, T1w with gadolinium, T2w, and T2w-FLAIR) with

image registration applied. The output targets are voxel-wise annotations of three different tumor classes (Gd-enhancing tumor, peritumoral edema, and necrotic and non-enhancing tumor core) obtained through annotations of up to four raters and approved by neuroradiologists [1, 2, 19]. In our experiments, we restricted the output target to binary labels representing healthy tissue or tumors. As we did not have access to the validation and test set of the BraTS challenge, our full dataset consists of 369 samples of which we used up to 250 for the training of our models. We applied instance normalization to the samples for feature scaling, which performed slightly better than other feature scaling methods. Since training-data augmentation by random scaling has been shown [25] to be beneficial also for scale-equivariant methods (possibly due to introducing intermediate scales not present in the discretized group, and increasing training-data diversity due to interpolation artifacts), we study the effect of scaling training samples with a random scale between 0.7 and 1.0 in every training step in additional dedicated experiments.

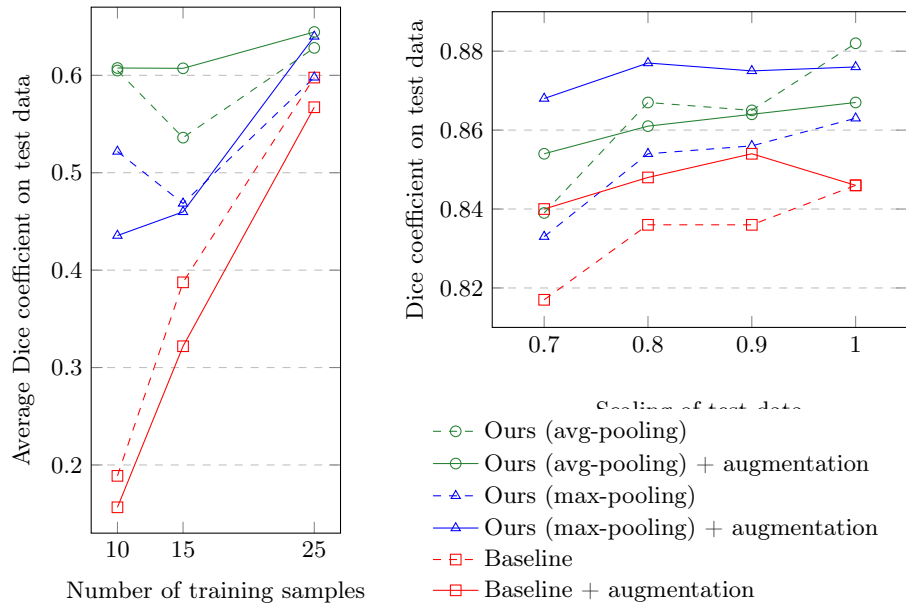
We base our model on the U-Net architecture [5] using four downsampling and upsampling blocks (with 4, 8, 16, and 32 channels from top to bottom). Down- and upsampling are performed using strided and transposed scale-equivariant convolutional layers, respectively (kernel size 5, stride 2), each followed by two scale-equivariant convolutional layers, and residual connections are used within blocks. Skip-connections are used between blocks of the same resolution except for the uppermost layer. The truncated scaling group used in the scale-equivariant network is  $\{0.9^0, 0.9^1, 0.9^2, 0.9^3\} = \{1.0, 0.9, 0.81, 0.729\}$  and the kernel basis consists of 27 basis functions, constructed as described in Section 3. Experiments were carried out to compare max- vs. avg-pooling over the scaling group in the last layer.

We compare our model to a baseline using the same architecture but with 16, 32, 64 and 128 channels in the respective network blocks. The network width and depth, learning rate and other hyperparameters were tuned using manual search over a wide range separately for the baseline method and the scale-equivariant methods. Training was performed for up to 160 epochs, depending on the size of the training set, using an NVIDIA RTX 8000 GPU, up to 5 GB VRAM and 10 GB RAM. The training lasted about 2 hours on average.

The loss used is a sum of the soft Dice loss [20] and the binary cross-entropy, and the Adam optimizer with a learning rate of 0.01 was used for training with an exponential learning rate decay.

## 4.2 Results and Discussion

Our proposed method benefits from its inherent scale-equivariance which is robust against scale changes as can be seen in Fig. 2. It outperformed the baseline on all scalings of the test data (see Fig. 3). The scale-equivariant network using average-pooling and no data augmentation reached a Dice score of  $0.882 \pm 0.104$  on the (non-scaled) test set, and thus outperforms the method using max-pooling ( $0.876 \pm 0.110$ ) and the non-scale-equivariant baseline method ( $0.846 \pm 0.128$ ) that are using training-data augmentation (see Tab. 1).

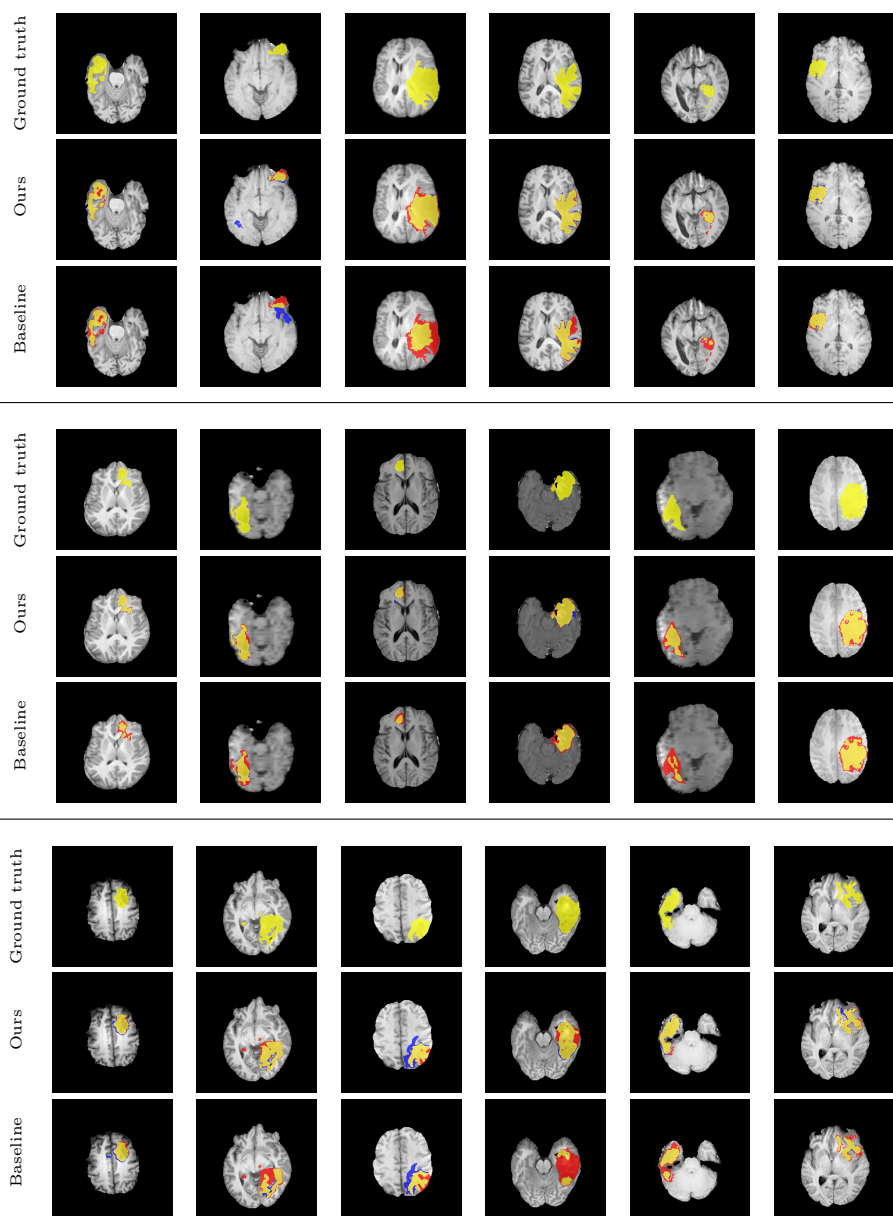


**Fig. 3.** Comparison of methods in terms of quality of brain tumor segmentation. (Left:) Performance comparison when trained on less training data. The Dice scores are averaged over all test data scalings  $\{0.7, 0.8, 0.9, 1.0\}$ ; results are similar for each individual scaling factor. Scale-equivariant neural networks considerably outperform the baseline method when trained on only few samples. (Right:) Generalization of trained methods to scaled test data. The scale-equivariant models consistently outperform the baseline method.

Training on an augmented dataset increased the performance of the baseline model in dealing with scaled versions of the test data. Data augmentation proved to be beneficial for the scale-equivariant method as well, as it stabilizes training and improves the networks’ capability to deal with interpolation artifacts introduced through artificial scaling of the data. This is evidenced by the better performance of the models on scaled test data (see Fig. 3) and is consistent with observations for scale-equivariant methods in the two-dimensional domain [25]. Our method trained with data augmentation and using max-pooling outperforms the baseline trained with and without data augmentation on all scalings of the test data. Visualizations of the segmentation results are shown in Fig. 4.

In a separate experiment, we evaluate the data efficiency of the proposed method. The scale-equivariant method was less affected by a reduction in training data than the baseline method: for a small number of training data with only 10 or 15 training samples, the proposed method performed up to three times better than the baseline method. Thus, it is a valuable tool in the medical setting, where training data is often limited due to high data-acquisition costs, privacy, or rare diseases.





**Fig. 4.** Visualization of the ground truth segmentation and the predictions of the scale-equivariant method and the baseline method. True positives are shown in yellow, false positives in blue, and false negatives in red. Our proposed method generally performs better even with complex lesion shapes.

**Table 1.** Quality of brain tumor segmentation using the scale-equivariant neural network (ours) and baseline neural network, both trained with data augmentation. The scale-equivariant neural network generalizes to a broad range of scales and outperforms the baseline on all scales of test data.

Scaling of test data	1.0		0.9		0.8		0.7	
Method	Ours	Baseline	Ours	Baseline	Ours	Baseline	Ours	Baseline
Dice coefficient	<b>0.876</b>	0.846	<b>0.875</b>	0.854	<b>0.877</b>	0.848	<b>0.868</b>	0.840
	$\pm$ <b>0.110</b>	$\pm$ 0.128	$\pm$ <b>0.115</b>	$\pm$ 0.131	$\pm$ <b>0.099</b>	$\pm$ 0.140	$\pm$ <b>0.103</b>	$\pm$ 0.154
Balanced accuracy	<b>0.940</b>	0.916	<b>0.944</b>	0.923	<b>0.933</b>	0.920	<b>0.920</b>	0.919
	$\pm$ <b>0.055</b>	$\pm$ 0.078	$\pm$ <b>0.053</b>	$\pm$ 0.072	$\pm$ <b>0.060</b>	$\pm$ 0.078	$\pm$ <b>0.070</b>	$\pm$ 0.085

## 5 Conclusions

Our method shows strong results and consistently outperforms the baseline. Max- and average-pooling yield slightly different results depending on the context. Average-pooling uses information from various scales simultaneously and thus can use “fractal (multi-scale) properties” [27] of the image. Due to relying on several scales at once, these image properties get easily destroyed by interpolation artifacts, or move beyond the truncation boundary of the scale dimension (see Section 3), when scaling the training data or test data. This might explain why average-pooling trained without data augmentation outperforms all other methods on unscaled test data, but is more negatively affected by scaling of test data. On the other hand, max-pooling is not affected at all by artifacts at the truncation boundary if they have smaller magnitude than the maximal values selected by max-pooling.

We propose a range of scale-equivariant neural network layers that can be used to analyse medical images, but can also be applied in other fields with 3D voxelized data. The formulation of scale-equivariant networks could further be extended to point cloud data or other 3D data representations. Our proposed method outperformed the non-scale-equivariant baseline and showed its efficiency in a low-resource setting, which can be especially helpful in the medical area.

## References

1. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific data* **4**(1), 1–13 (2017)
2. Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., Shinohara, R.T., Berger, C., Ha, S.M., Rozycki, M., et al.: Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. *arXiv preprint arXiv:1811.02629* (2018)
3. Bekkers, E.J.: B-spline CNNs on Lie groups. In: *International Conference on Learning Representations* (2019)

4. Chen, Y., Fan, H., Xu, B., Yan, Z., Kalantidis, Y., Rohrbach, M., Yan, S., Feng, J.: Drop an octave: Reducing spatial redundancy in convolutional neural networks with octave convolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3435–3444 (2019)
5. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-net: learning dense volumetric segmentation from sparse annotation. In: International conference on medical image computing and computer-assisted intervention. pp. 424–432. Springer (2016)
6. Cohen, T., Welling, M.: Group equivariant convolutional networks. In: International conference on machine learning. pp. 2990–2999. PMLR (2016)
7. Cohen, T., et al.: Equivariant convolutional networks. Ph.D. thesis (2021)
8. Cohen, T.S., Welling, M.: Steerable CNNs. In: International Conference on Learning Representations (2016)
9. Ghosh, R., Gupta, A.K.: Scale steerable filters for locally scale-invariant convolutional neural networks. arXiv preprint arXiv:1906.03861 (2019)
10. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 580–587 (2014)
11. Gunel, B., Sahiner, A., Desai, A.D., Chaudhari, A.S., Vasanawala, S., Pilanci, M., Pauly, J.: Scale-equivariant unrolled neural networks for data-efficient accelerated MRI reconstruction. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part VI. pp. 737–747. Springer (2022)
12. Isensee, F., Jäger, P.F., Full, P.M., Vollmuth, P., Maier-Hein, K.H.: nnU-Net for brain tumor segmentation. In: International MICCAI Brainlesion Workshop. pp. 118–132. Springer (2020)
13. Jaderberg, M., Simonyan, K., Zisserman, A., et al.: Spatial transformer networks. *Advances in neural information processing systems* **28**, 2017–2025 (2015)
14. Jansson, Y., Lindeberg, T.: Scale-invariant scale-channel networks: Deep networks that generalise to previously unseen scales. *Journal of Mathematical Imaging and Vision* **64**(5), 506–536 (2022)
15. Kanazawa, A., Sharma, A., Jacobs, D.: Locally scale-invariant convolutional neural networks. *Deep Learning and Representation Learning Workshop: Neural Information Processing System* (2014)
16. Lindeberg, T.: Scale-covariant and scale-invariant Gaussian derivative networks. In: International Conference on Scale Space and Variational Methods in Computer Vision. pp. 3–14. Springer (2021)
17. Liu, R., Lauze, F., Bekkers, E., Erleben, K., Darkner, S.: Group convolutional neural networks for DWI segmentation. In: *Geometric Deep Learning in Medical Image Analysis*. pp. 96–106. PMLR (2022)
18. Marcos, D., Kellenberger, B., Lobry, S., Tuia, D.: Scale equivariance in CNNs with vector fields (2018)
19. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE transactions on medical imaging* **34**(10), 1993–2024 (2014)
20. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision (3DV). pp. 565–571. IEEE (2016)

21. Müller, P., Golkov, V., Tomassini, V., Cremers, D.: Rotation-equivariant deep learning for diffusion MRI. In: International Society for Magnetic Resonance in Medicine (ISMRM) Annual Meeting (2021)
22. Naderi, H., Goli, L., Kasaei, S.: Scale equivariant CNNs with scale steerable filters. In: 2020 International Conference on Machine Vision and Image Processing (MVIP). pp. 1–5. IEEE (2020)
23. Sabour, S., Frosst, N., Hinton, G.E.: Dynamic routing between capsules. *Advances in neural information processing systems* **30** (2017)
24. Sosnovik, I., Moskalev, A., Smeulders, A.: Disco: accurate discrete scale convolutions (2021)
25. Sosnovik, I., Szmaja, M., Smeulders, A.: Scale-equivariant steerable networks. In: International Conference on Learning Representations (2019)
26. Thomas, N., Smidt, T., Kearnes, S., Yang, L., Li, L., Kohlhoff, K., Riley, P.: Tensor field networks: Rotation-and translation-equivariant neural networks for 3D point clouds. *arXiv preprint arXiv:1802.08219* (2018)
27. Weibel, E.R.: Fractal geometry: a design principle for living organisms. *American Journal of Physiology-Lung Cellular and Molecular Physiology* **261**(6), L361–L369 (1991)
28. 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)
29. 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)
30. Winkens, J., Linmans, J., Veeling, B.S., Cohen, T.S., Welling, M.: Improved semantic segmentation for histopathology using rotation equivariant convolutional networks (2018)
31. Worrall, D., Welling, M.: Deep scale-spaces: Equivariance over scale. *Advances in Neural Information Processing Systems* **32** (2019)
32. Xu, Y., Xiao, T., Zhang, J., Yang, K., Zhang, Z.: Scale-invariant convolutional neural networks. *arXiv preprint arXiv:1411.6369* (2014)
33. Yang, F., Choi, W., Lin, Y.: Exploit all the layers: Fast and accurate CNN object detector with scale dependent pooling and cascaded rejection classifiers. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 2129–2137 (2016)
34. Yang, Y., Dasmahapatra, S., Mahmoodi, S.: Scale-equivariant UNet for histopathology image segmentation. In: *Geometric Deep Learning in Medical Image Analysis* (2022)
35. Zhu, W., Qiu, Q., Calderbank, R., Sapiro, G., Cheng, X.: Scaling-translation-equivariant networks with decomposed convolutional filters. *Journal of machine learning research* **23**(68), 1–45 (2022)